Imperial College London

Department of Earth Science and Engineering

Computer vision for carbonate core classification

Harriet Louise Dawson

Submitted in part fulfilment of the requirements for the degree of

Doctor of Philosophy

Declarations

Declaration of originality

I declare that this thesis, submitted for the degree of Doctor of Philosophy at Imperial College London and Associateship of the Royal School of Mines, is original research that has been completed by me, and carried out under the supervision of Professor Cédric M. John and Professor Olivier Dubrule. All published and unpublished material used in this thesis has been appropriately referenced. None of this work has previously been submitted to Imperial College London or any other academic institution for a degree or diploma, or for any other qualification.

Harriet L. Dawson

1st March 2024

Copyright declaration

Abstract

Core data is a fundamentally important resource in characterising the geological subsurface. Globally, cores provide a record of diverse past geologic, climatic, and biological processes that have operated over multiple timescales. Leveraging insights from core images presents a pivotal challenge due to the largely subjective nature of traditional core description methods, inherent complexities and variability in geological data, and the paucity of significantly large, labelled datasets. Deep learning presents a data-driven approach to deriving predictive models from observational data.

With the evolution of machine learning, and growing awareness of the importance of big data, new automated workflows using deep learning algorithms present alternative approaches to the extensive time and labour requirements of traditional, manual core descriptions. Focusing solely on carbonate rocks, this thesis aims to establish foundations for applying deep learning to core interpretation. First, convolutional neural networks (CNNs) are used to classify carbonate textures from core images, optimising a workflow for lithological prediction across datasets of different magnitudes. To address the limiting factor of insufficient data, Generative Adversarial Networks (GANs) are applied to generate synthetic core images, thus filling sampling gaps and reducing overfitting. The performance of synthetic data is then compared with traditional data augmentation methods. Lastly, CNN-based object detection algorithms are applied to identify and quantify multiple types of carbonate grain, with results benchmarked against human interpretations to evaluate the efficacy of the method.

This approach leverages computer vision models, such as CNNs and GANs, to enable streamlined geological interpretation and data-driven insights. As geoscience data availability increases through new datasets and the digitalisation of historical records, these methods can be further developed. The frameworks proposed in this thesis offer a pathway for advancing deep learning-based carbonate classification and have the potential for broader application to cores from diverse formations and lithologies.

Acknowledgements

First and foremost, I would like to thank my supervisors Cédric John and Olivier Dubrule. Cédric, thank you for the confidence and trust you have placed on me since day one. For your patience and support in guiding me through to the end, and for helping me to find and secure my dream job. Olivier, thank you for support and guidance at the start of my PhD journey, and for always stopping to have a chat and brainstorm in the corridors of the RSM, even in your retirement.

Thank you to all my other mentors and collaborators for useful discussions and support over the years. I would like to highlight five people in particular. Firstly, thank you to Al Fraser and Peter Allison who trusted me to do this PhD and have continued to support me and follow my journey. It is always a pleasure to sit down to have a chat and learn from both of your extensive knowledge. To Gary Hampson, thank you for all the opportunities you have sent my way and for always being a smiling face. I apologise for all the eel theories. Lastly, thanks to Phillipa 'Sentinel-1' Mason and Alan 'Dead Plant Man' Spencer. Thank you for trusting me to learn from you and teach alongside you. I have so many fond memories of our times on board the good ship Spain. Be sure to take care of Brian for me, this is Madame Whiplash signing off.

Thank you to the wonderful friends and colleagues I've met along the way:

> To all those involved in the NERC Centre for Doctoral Training in Oil & Gas, – to Lorna Morrow, John Underhill, and Anna Clark, for their leadership and organisation. To all those in the 2018 cohort, including my fellow Imperial reps, Michael and Denis, and, in particular, my fellow baddies – Lucinda, Jonathan, Ash, Alex, Emma, and Amy, for making the CDT such an incredible experience. I hope to dance with you all again at another ceilidh in the not-too-distant future.
>
> To members of the John Lab (formerly known at the Carbonate Research Group) past and present – Rob, Mahmoud, Saira, Xuan, Simon, John, Qi, Adhipa, Waleed, Zayad, Marta, Claire, and Sarah. I must give a special shout out to Sarah, my best friend, who I am eternally grateful to have met. I would also consider Sarah my unofficial mentor, my guardian PhD student who helped to get me going and keep my feet on the ground, always answered my questions no matter how silly, and who I could always rely on to rant to when things weren't going quite so well. I have so many amazing memories of our time together from field trips, conferences, holidays, coffee mornings, and the various pubs around Victoria Station. I know we will always remain close no matter what the future holds.
>
> To the Department of Earth Science and Engineering community – there are so many awesome people in this department that have made this journey possible. To David, gracias por ser parte de mi vida – ha sido un placer compartir este camino contigo y que seas mi amigo. To Harry, thanks for all the

memes and casual drinks. To Anissa, Emma, Lizzie, and Victoria, thank you for all the chats and cups of tea.

To my PhD wildcats – Evie, Sinéad, Catherine, Tom, Alex, Bailey, Ruth, Jenny, Graham, Sian, and Cecily. Doing a PhD in the midst of a global pandemic is hard. I am so thankful to have met such a wonderful group of friends, who are always ready to hype you up, have an exceptional sense of humour, love the occasional conspiracy theory, and will always, *always* be keen for a glass of wine (or more…). You are all fantastic people and rock doctors. I think we made coach proud.

To my pup, Busby. Perhaps it is silly to write a paragraph for my dog who will never read this, but I truly am thankful for your intuition well beyond your one-year-old brain, for hugging me whenever you sensed I was stressed (which, to be fair, was a lot), and for getting me outside every day. This thesis is written in the font Garamond, in honour of your (unofficial) middle name.

To my fiancé, Jamie. Thank you for somehow tolerating me over all these years, for always putting my work first with unconditional patience, love, and support, and for picking me up and pushing me forward whenever I doubted myself. You have always been there, encouraging me with a 'yes' or 'you've got this'. Having you as my best friend and number one cheerleader is such a blessing, because I know I will always have someone in my corner rooting for me. This thesis simply would not exist without you. Here's to the future.

And finally, I must thank my incredible parents, to whom I owe a debt of gratitude that words cannot fully express. You have always supported my ambitions and encouraged me to pursue my dreams. You have made countless sacrifices to make sure you could provide me with that all I ever needed, and so much more. I would not be where I am today without you. Meur ras.

Dedication

To my mum, the strongest woman I know.

‘I am never afraid of what I know’.

Anna Sewell, Black Beauty

Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1 Research problem and rationale

Cores provide valuable insights into rock formations, structural features, and sedimentary sequences, offering geologists a window into the past that spans millions of years. Descriptive data derived from observed features forms the basis of many geological practices, particularly those relating to core analysis. However, the interpretation of observational data in geology is not without its challenges. Subjectivity, ambiguity, and the inherent limitations of traditional analytical methods can often hinder the reliability and efficiency of geological analyses. Previous studies have shown significant human biases in description, as exemplified for carbonate rock facies where the same Dunham classes (Dunham, 1962; Embry and Klovan, 1971) were classified using different textural names (Lokier and Al Junaibi, 2016). As a case example, a synthetic floatstone was classified by 64 experienced carbonate sedimentologists as a floatstone (42.2% of participants), a wackestone (23.4%), a rudstone (10.9%), and packstone (9.4%), and 'other' (14.1%). Despite nearly 60% of the participants misclassifying this texture, they overwhelmingly (>70%) stated that this was an 'easy' rock to classify (Lokier and Al Junaibi, 2016). The subjectivity of human observations results in heterogenous attribution of facies names to the same rocks, despite the use of a common classification scheme. This uncertainty limits the usefulness of descriptive data.

In recent years, the advent of machine learning has emerged as a promising solution to address these issues, with the potential to revolutionise the way geologists analyse and interpret data from diverse types of images. By leveraging advanced algorithms and computational techniques, machine learning offers geoscientists a way to enhance the accuracy, consistency, and scalability of geological analyses, ultimately advancing our understanding of Earth's past, present, and future.

The aim of this study is to overcome the barrier of reliable and fast quantitative descriptive data by investigating the ability of deep learning algorithms to correctly identify the texture and constituents of carbonate rock cores in an automated fashion, using multiple types of artificial neural networks. The goal of this project is to build a system that can reliably extract information from digital core images and build an objective software framework for interpretation. Application of these advancing digital technologies has the potential to revolutionise this field of geology, with more accurate and uniform facies descriptions, faster analysis times, and reduction of natural biases. Within the energy industry, application of this approach could lead to a standardised description of facies at the basin scale and an enhanced characterisation of porosity and permeability relationships and, hence, reservoir quality based on lithofacies. This would result in an improved sequence stratigraphic framework, and, by extension, reduced risk associated with the development of carbonate fields. Within academia, applications would be wide ranging and more varied. This could include, but is not limited to, a better understanding of palaeoenvironmental reconstructions, sequence stratigraphy, and diagenetic processes.

The project is divided into three main areas of research. The first focuses on the application of image classification models to the textural recognition and classification of digital carbonate core images. The second proposes the use of synthetic data generation methods to produce additional data closely resembling real samples in data-limited situations. The final part of my research investigates the application of object detection algorithms to extract a range of skeletal components from digital carbonate core images, with the aim of identifying and quantifying these extracted features to enable further interpretations. Through the development of these research methods, I will illustrate the transformative potential of computer vision models in increasing confidence, reliability, and speed in the classification of carbonate lithologies. This will, in turn, facilitate the development of more accurate and consistent depositional models, and drive innovation in the geosciences.

1.2 Core strengths and values

1.2.1 Core acquisition

Extracted from boreholes during drilling operations, cores offer valuable insights into subsurface formations. Cores are obtained as continuous cylindrical samples of rocks at specified intervals using specialised coring drills, recovering samples at depths down to hundreds of metres (Jutzeler *et al.*, 2014). These processes can be achieved through standard coring systems or specialised coring systems, tailored for specific conditions such as directional or horizontal drilling, preservation of original core conditions (e.g., pressure or fluid saturation), and for unconsolidated materials. The lengths and diameters of recovered cores will vary depending on the coring system employed. For example, in Integrated Ocean Drilling Program (IODP) expeditions, core samples are usually retrieved in 9.5m long cores with a diameter of 6.5cm when using the advanced piston corer (APC), extended core barrel (XCB), or rotary core barrel (RCB) systems (Betzler *et al.*, 2017). Post-drilling wireline methods can also allow for retrieval of additional small sidewall cores from zones of interest within the wellbore (Saucier and Sincock, 2022).

During core recovery and handling procedures, cores may be susceptible to physical damage or alteration, including loss of reservoir fluids and invasion of drilling fluid (Kalhor Mohammadi *et al.*, 2023). The extent of such damage depends on various factors such as rock type, rock properties, and the nature of drilling fluids (Gamal *et al.*, 2020). Mechanical stresses during core recovery and transportation, especially in unconsolidated or fractured formations, can induce fracturing in cores. To mitigate such damage and ensure mechanical stability during core transportation, specialised core barrel materials and liners are used to facilitate core removal, while recovering small core sections helps prevent compaction (Ubani *et al.*, 2012).

Oriented coring is another valuable technique wherein cores are extracted while preserving the *in-situ* position relative to the north direction. This method proves beneficial in determining structural properties of the subsurface, including the strike and direction of fractures and beddings. Oriented cores facilitate the analysis of directional parameters, such as permeability (Ma *et al.*, 2022). Achieving oriented coring involves the use of special oriented core barrels to label and continuously record orientation data during core extraction. While this method is considered the most accurate and direct, it can also be considerably more costly. However, alternative indirect methods for core reorientation exist, varying in cost and accuracy, such as matching core structures to diameter logs or seismic data, employing palaeomagnetic techniques, or comparing core images to image logs from Formation MicroScanner (FMS) (MacLeod *et al.*, 1994).

1.2.2 Core archives

Recovered cores are cut into smaller sections, typically around 1.5m length, and stored in trays. These sections are arranged in multiple rows within each tray, with depths labelled to denote their respective positions, progressing from shallow depths to deeper ones (Figure 1.1). The length of each section of core, including spacers representing intervals of no recovery, is entered into a database as "curated length", which often differs by several centimetres from the recovered length measured on the core receiving platform (Sager *et al.*, 2023).

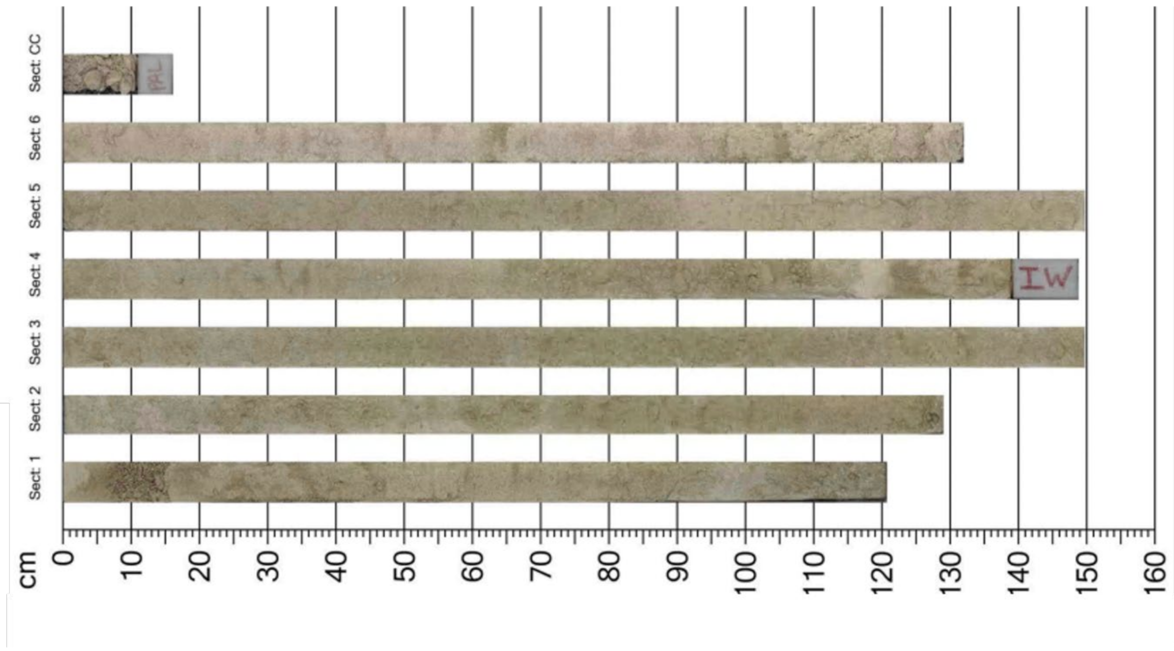

Figure 1.1: Example of low-resolution whole core photos from the Janus database for the International Ocean Discovery Program Expeditions 1-312.

Whole core photographs captured from a distance using a handheld camera can often suffer from a lack detail and perspective distortions (Gegg and Gegg, 2023). However, these may be improved with advanced imaging setups (e.g., USBR, 1998). At present, the best results are achieved through line scan imaging. Cores are steadily moved by an electric motor past a fixed camera, which captures sequential images. These images are then stitched together automatically, yielding a full core image (e.g., Geotek Limited, 2017).

Recently, there has been a significant shift towards archiving cores in digital libraries. Digital archiving offers many advantages over physical storage, including loss, damage, or alteration of cores, while facilitating convenient remote access for flexible evaluations (Shreeve, 2024). Modern core scanning technology, capable of producing high-resolution images at rates of 5–40 pixels/mm and scanning hundreds of metres of core per day, has transformed core imaging into a substantial big data opportunity (DMT GeoTec, 1996; DMT GmbH, 2000). These scans capture crucial visual features, facilitating identification of core lithology, texture, stratigraphy, and structural features. Consequently, the next frontier in core archive innovation lies in establishing modern image-based core analysis methodologies.

1.2.3 The importance of cores in the digital age

Core characteristics hold significance across many fields, including geoscience, geotechnical engineering, and mining. As illustrated in Figure 1.2, cores can provide detailed information about lithological characteristics, mineralogy, sedimentary structures, and fossil content, enabling geologists to reconstruct past environments and interpret depositional environments and geological events. Unlike other sampling methods, such as well logs or seismic surveys, coring material is considered a direct record of the geological strata, allowing detailed investigation of geochemical, microbiological, palaeomagnetic, and petrophysical properties, such as permeability and porosity (Fossen, 2024). Cores serve as essential tools for understanding sedimentary deposition, diagenetic processes, structural geology, and palaeoenvironmental conditions. Furthermore, cores are crucial for resource exploration and reservoir characterisation in industries such as petroleum geoscience, mining, and groundwater studies. By analysing cores, geologists can assess reservoir quality, identify hydrocarbon-bearing formations, and evaluate the potential for mineral deposits.

Cores serve as crucial benchmarks for calibrating well log and seismic data, offering the most accurate representation of underground formations (Ashena and Thonhauser, 2018). Modern technology facilitates the digital storage of retrieved cores, enabling their representation in 2-D or 3-D formats (Tiwari *et al.*, 2017). These digital records encapsulate essential visual properties of cores, including lithologies, textures, and geological structures (Betlem *et al.*, 2020). The trend toward digital archiving of cores is on the rise due to its advantages in long-term storage safety and enhanced accessibility compared to physical storage methods (Shreeve, 2024). With advancements in scanning technology and digital archiving, a significant big data opportunity is presented for revolutionising core analysis methodologies. The integration of advanced machine learning techniques with core images enables the creation of automated digital tools for geologists, poised to enhance geological description and interpretation processes. Similar computer vision technologies have already been adopted in other fields, e.g., autonomous driving (e.g., Nataprawira *et al.*, 2021), smart cities (e.g., Osman, 2019), and healthcare (e.g., Obermeyer and Emanuel, 2016; Aggarwal *et al.*, 2021). This raises the question; can these technologies be translated to core description, and to what extent?

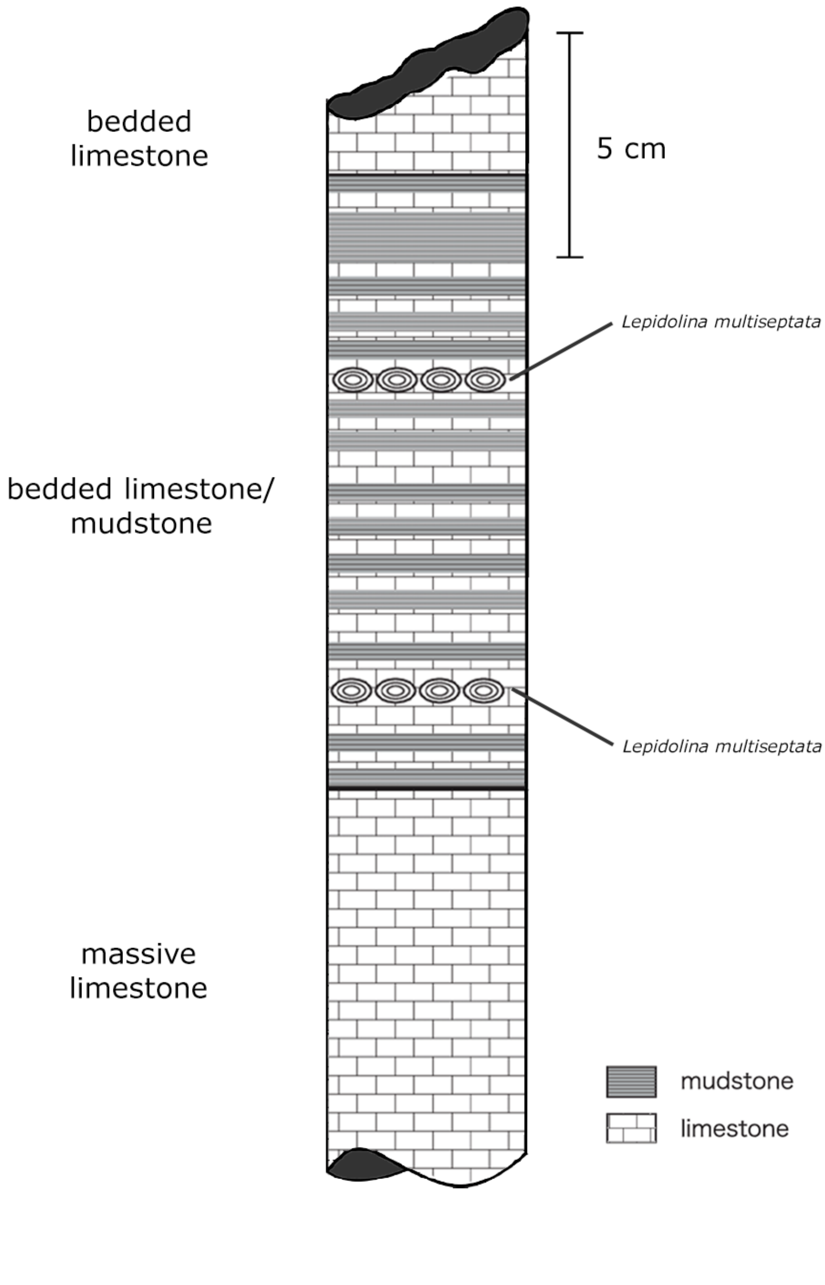

Figure 1.2: Schematic of a whole core section showing different lithologies, structural information, and biostratigraphy.

1.2.4 Big data in geoscience

The emergence of big data signifies the vast volume of digital information and the rapidly increasing availability and diversity of data (Chen *et al.*, 2020). Big data, whether in the form of images, videos, or text, is characterised by its volume, velocity, variety, veracity, value, and variability (Figure 1.3), though its definition remains open-ended (Wamba *et al.*, 2015). This concept holds substantial potential in automating processes and developing data-driven systems, thereby influencing various domains including medicine (e.g., Obermeyer and Emanuel, 2016), marketing (e.g., Saura, 2021) and the technology industry (e.g., Davenport and Dyché, 2013; Barchiesi *et al.*, 2021). The rise of big data will undoubtedly play a role in shaping the future of geoscience research, despite the geological community being comparatively slow to embrace the big data technologies that are transforming other sciences (Dramsch, 2020). In the context of cores, big data is anticipated to be generated through the modern core management practices and the growing digitisation of core archives outlined above.

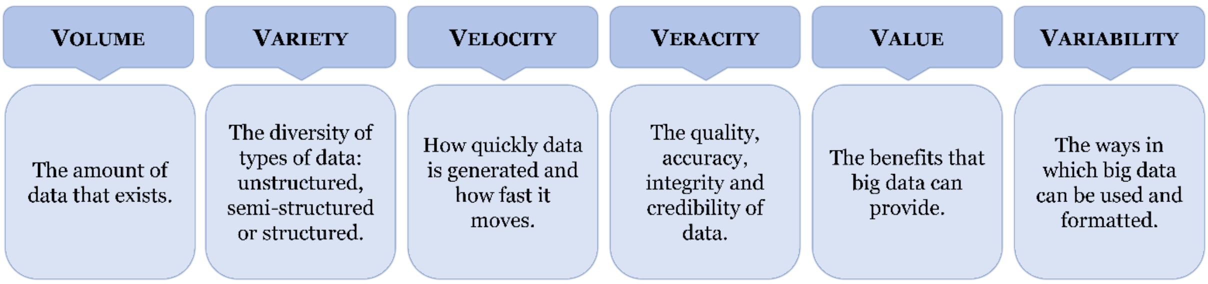

Figure 1.3: The 6 Vs of big data.

To fully leverage the potential of this data, it necessitates management through more sophisticated intelligent methods than the traditional analytical methods. In big data analysis, the emphasis shifts from merely assessing the data to utilising it for future predictions and automating processes (Davenport and Dyché, 2013). Unlike traditional data analysis, big data demands advanced analytical techniques to handle its vast volume and extract its true value. Deep learning has emerged as an ideal approach. Deep learning models, such as CNNs, possess the capability to effectively process large datasets by extracting meaningful insights and learning representations from them (Goodfellow *et al.*, 2016). Advancements in high-performance computing power have allowed big data and these models to become more practicable than ever before (LeCun *et al.*, 2015).

Adopting these big data techniques across the subdisciplines of geosciences holds the potential to drive cost reduction and streamline processes through automation (Qi, 2020). While previous efforts have focused on leveraging data from sources like well logs with traditional machine learning algorithms to address challenges in petroleum and mining sectors (e.g., Valentín *et al.*, 2019), the integration of contemporary deep learning methods in the literature is comparatively limited. Moreover, there exists a gap in applications to core data, particularly core images, despite the proliferation of digital core image data. Given the rise in popularity of digital archiving superseding traditional storage methods, conventional handling and management of such digital data will prove inefficient. Application of advanced machine learning methods emerges as a promising avenue to better exploit this wealth of data.

1.3 Machine learning in geoscience

Over the past decade, the rise of artificial intelligence (AI), both in theory and in practice, has revolutionised the field of computer science. In short, AI is a hugely popular topic of broad and current interest, and its popularity, and the hype surrounding it, is growing all the time. Machine learning, a subset of AI, is becoming an increasingly prevalent alternative to traditional data analysis methods and has already achieved significant successes across a wide range of disciplines. Fuelled by the recent advancements in computational power and an explosion in data availability, machine learning is gaining more and more traction in many knowledge engineering areas such as classification, clustering, and regression (e.g., Fayyad *et al.,* 1996; Boutaba *et al.,* 2018; Carvalho *et al.,* 2019). Due to their effectiveness in identifying and modelling non-linear processes, artificial neural networks have been able to excel in several computer vision tasks. Examples of current successful application areas include image processing (e.g., Egmont-Petersen *et al.,* 2002; LeCun *et al.,* 2015; He *et al.,* 2016), facial recognition (e.g., Fasel and Luettin, 2003; Sun *et al.,* 2015; Hossain *et al.*, 2023), speech recognition (e.g., Deng *et al.,* 2013; Abel-Hamid *et al.,* 2014; Izbassarova *et al.,* 2020), medical imaging and diagnosis (e.g., Erickson *et al.,* 2018; Chinnam *et al.*, 2022), and plant disease detection (e.g., Sladojevic *et al.,* 2016; Golhani *et al.,* 2018).

The use of machine learning in geoscience has been rapidly growing for nearly two decades; however, progression has been relatively uneven and, despite increasing popularity within the geoscientific community, practitioners were comparatively slower to engage with these recent advancements (Bergen *et al.,* 2019; Pires de Lima *et al.*, 2019a). The adoption of deep learning can now be seen across all aspects of geoscience and its subdisciplines (Dramsch, 2020). Remote sensing was an early adopter (e.g., Lary *et al.*, 2016; Maxwell *et al.*, 2018), with geomorphology (e.g., Valentine and Kalnins, 2016; Goldstein *et al.*, 2019), hydrogeology (e.g., Shen, 2018; Sit *et al.*, 2020), solid Earth geophysics (e.g., Bergen *et al.,* 2019; Yu and Ma, 2021), seismology and seismic interpretation (e.g., Wang *et al.*, 2018; Kong *et al.*, 2019; Xie *et al.*, 2020), and geochemistry (e.g., Zuo *et al.*, 2019; He *et al.*, 2022) closely following suite. Climate science has received thorough assessments of the potential uses of machine learning methods for modelling, engineering, and mitigation to address the problem (e.g., Karpatne and Kumar, 2017; Huntingford *et al.*, 2019; Rolnick *et al.*, 2022). Current machine learning applications within the geoscientific community include, but are not limited to, seismic facies classification (e.g., Meldahl *et al*., 2001; West *et al*., 2002; Chevitarese *et al.,* 2018), lithofacies classification from wireline logs (e.g., Bestagini *et al.,* 2017; Halotel *et al.,* 2020), classification of volcanic ash (e.g., Shoji *et al.,* 2018), analysis of seismicity (Kortström *et al*., 2016; DeVries *et al*., 2018; Perol *et al*., 2018), and inverse problems (e.g. Mosser *et al.,* 2020).

These initial uses of machine learning have typically been limited in scope, but recent efforts to promote the use of benchmark geoscientific datasets in global and local competitions have the potential to drive deeper, broader, and more collaborative efforts e.g. the SEG Machine Learning Facies Classification Contest (Hall, 2016; Hall and Hall, 2017) and the TGS Salt Identification Challenge (Kaggle, 2019). Despite this, many existing datasets from long-established sources still remain largely unexplored; partly due to the lack of properly labelled data and also due to the inefficiency of traditional analytical techniques, which are typically resource intensive (Klump and Robertson, 2015). One of the key advantages of using machine learning for geoscience is the ability to process large volumes of multidimensional data, making these analytical techniques ultimately more time and cost effective. It is anticipated that the application of machine learning will see dramatic progress in the automation of complex prediction tasks, the solution of inverse problems through multiscale modelling, and the discovery of new or alternative patterns and relationships that are not readily visible to humans (Karpatne *et al.,* 2018; Mosser *et al.,* 2020). This project aims to help solve the data under-utilisation problem by implementing machine learning techniques for facies classification. In doing so, this should provide more efficient and reliable analytical methods for both industry and academia.

1.3.1 Challenges for machine learning in geoscience

Despite previous successes, there are several characteristics of geoscience processes and applications that may continue to prove challenging to traditional machine learning algorithms. Firstly, some challenges may arise from the nature of geoscience, since geological processes are part of a tightly coupled, complex web of causes and effects which span from the nanoscale to planetary scale, over timescales of microseconds to billions of years. The spatial and temporal extent of geological phenomena can be difficult for classical machine learning algorithms as it is assumed that observed variables are independent and identically distributed (Klump and Robertson, 2015). The Earth system, and the geological data that we obtain from it, contains numerous potential variables, all of which may impact each other and must, therefore, be considered simultaneously. The complex and highly dimensional nature of these data generates the need to scale existing machine learning methods or develop novel algorithms. One such approach that could help to solve the challenge of separating large-scale and small-scale processes is manifold learning, which focuses on non-linear dimensionality reduction (Lawrence, 2012).

The geological processes we observe can often display a high degree of spatial and temporal variability, leading to a rich heterogeneity among collected data. For instance, due to varying climatic conditions and depositional environments across different geographies, the characteristics of geoscience variables may vary significantly between locations (Karpatne *et al.,* 2018). Many of the phenomena we study are also unique or occur infrequently. Identifying and characterising classes from geoscience data is, therefore, challenging due to the comparatively low number of data samples and the skew between classes.

Although many geoscience applications may involve large amounts of data, a common problem is the paucity of ground truth, labelled datasets. This is, in part, due to the laborious and expensive nature of geological data collection (Klump and Robertson, 2015). Datasets may also contain noise and missing values; while many variables cannot be measured directly, but only inferred from observations or model simulations. Many geological applications rely on an iterative process of collecting different samples and subjectively applying descriptive geological interpretations to known instances. While data availability in geosciences is increasing, it is still comparatively small and sparse given the complexity of the phenomena under study. In machine learning, the limited representative training samples may result in poor performance of algorithms due to e.g., overfitting, where the model is overly complex relative to the small dataset. Hence, there is a need for machine learning techniques that can account for smaller datasets, such as transfer learning (Pan and Yang, 2010), or the construction of synthetic datasets through simulations (Ebert-Uphoff and Deng, 2017) or frameworks, e.g., generative adversarial neural networks (GANs) (Mosser *et al.,* 2018; Zhang *et al.*, 2019).

1.3.2 Machine learning directions in geoscience

Novel research is required to develop new machine learning based approaches that can incorporate a broad knowledge of geological processes and use this information effectively, ensuring theoretically consistent results. As discussed above, geoscience phenomena are characterised by heterogeneous, non-linear, multiscale and highly dimensional processes. The available data is typically intermittent, multiresolution, contains considerable sources of uncertainty, and is relatively sparse in the given context. Therefore, in order to help constrain the learning of the complex relationships in geoscientific applications and efficiently guide knowledge discovery, the small sample size of available datasets must be supplemented by the underlying scientific principles that govern these processes. We need novel approaches that can leverage the recent advances in big data driven research with methods that exploit the domain knowledge of the theoretical frameworks under study.

Over the past few decades, we have seen the progression and many successes of powerful and complex machine learning models, exemplified by the recent surge of deep learning algorithms. However, many of the best-performing models in different domains, including deep neural networks for image recognition (e.g. Krizhevsky *et al.,* 2012), are viewed as complicated, "black box" models that do not disclose the details by which predictions are being made (Koh and Liang, 2017). This purported nature of neural networks can act as a barrier to the adoption of machine learning in applications where interpretability is essential, such as in geoscience applications, where interest lies in both predictive modelling and scientific understanding. Therefore, a significant area for research is the design of interpretive machine learning approaches that integrate both domain knowledge and causal inference, thereby enabling the algorithms to be understood by scientists and related to long-standing geoscientific theories and models.

Machine learning presents a data-driven approach to deriving predictive models from observational data. Owing to the spatiotemporal nature of geoscience data, the challenges faced in geoscience share some similarities with challenges in computer vision and natural speech processing, where deep learning has seen significant progress using architectures such as CNNs and recurrent neural networks (RNNs). Given the ability of these deep learning frameworks to automatically extract relevant features from data, there is the potential for a wide range of advanced machine learning methods to be effectively applied to extensive geoscience problems. Machine learning methods have already shown great successes and capacity in several specific geoscience applications, but considerable research challenges remain before those methods can be widely and easily applied to other areas of geoscience. Moreover, the challenges presented by these applications could drive the development of new machine learning algorithms suited to more complex datasets. The effective classification of geoscience data could also benefit from recent advances in machine learning e.g. transfer learning, where a model trained on a previous task with abundant training data is partially retrained on a smaller dataset to improve prediction performance (Pan and Yang, 2010). Transfer learning is still in the relatively early stages of development and deployment, so many opportunities exist for novel machine learning research. New approaches incorporating theory-guided learning will need to be developed, where an understanding of underlying theories will assist algorithms in handling complex data, providing alternative views of integrated information and enabling fundamentally new forms of autonomy, reasoning, and learning.

Previous classification tasks have focused largely on interpreting data from seismic or from wireline logs (e.g. Hall, 2016; Chen and Zeng, 2018; Halotel *et al.,* 2020), as opposed to core images. This project builds on a pilot study by the Carbonate Research Group at Imperial (John and Kanagandran, 2019), which tested the effectiveness of neural networks in recognising carbonate rock facies using the Dunham classification scheme. Using high-resolution core images from ODP Leg 194, the CNN model achieved 89.2% accuracy across seven classes. The results from this study show great potential for the use of machine learning in carbonate classification from digital images. Ghiasi-Freez *et al.* (2014) also presented an automated approach to the Dunham classification through basic image processing and segmentation techniques applied to a small sample of thin section images, achieving 81.6% accuracy. This PhD project proposes a new approach to carbonate core classification though theory-guided machine learning, where a combination of image classification, object detection, and deep generative algorithms will aim to improve overall interpretational accuracy, as well as reducing subjectivity and interpretation time. This has huge potential to transform the way in which we approach this field of geology.

1.4 Research questions

1.4.1 Aims

Having emphasised the importance of leveraging insights from core data, the overarching goal of this thesis is to improve our ability to objectively classify core images, with more accurate and uniform descriptions and reduction of analysis times.

From this we formulate the following research questions:

> **Research Question 1**: Can we successfully leverage transfer learning of convolutional neural networks (CNNs) to train a classifier that will generalise well on unseen data, given the heterogenous nature of geological images?
>
> **Research Question 2**: Can we condition deep generative models to generate synthetic carbonate textures representative of original core image data?
>
> **Research Question 3**: What are the comparative advantages and limitations of using synthetic data generated by Generative Adversarial Networks (GANs) versus traditional data augmentation techniques in improving the performance of CNNs in geological image classification?
>
> **Research Question 4**: What is the effectiveness of employing object detection algorithms to extract features, primarily different grains, from core images?

These questions form the main aims of the thesis and will be more explicitly addressed in the discussion. In this thesis, geological examples will focus on cores of Cenozoic carbonates from the North-Eastern Australian margin and the Maldives archipelago, where good preservation of carbonate platforms has enabled detailed studies of relative sea level fluctuations (Isern *et al.*, 2002; Betzler *et al.*, 2017). However, in the future, it is anticipated that the findings of this research will contribute to the widespread application of machine learning algorithms to aid geological interpretation, with the developed frameworks being extended to a wider implementation across a range of spatial and temporal scales.

1.4.2 Objectives

The following objectives were undertaken to address the questions outlined above. These objectives are separated based on the research chapters. The chapters are associated with the questions they help to address.

The key objectives for Research Question 1 are:

1.1. Compile core image datasets with orders of magnitude difference in data quantity to evaluate the impact of dataset size on classification performance.

1.2. Train different CNN architectures to classify carbonate core images according to the modified Dunham classification (Dunham, 1962; Embry and Klovan, 1971).

1.3. Evaluate the prediction performance for each CNN architecture based on metrics for mean per class accuracy, precision, recall, model complexity, computational complexity, training time, and classification speed.

1.4. Evaluate the impact of limited data on the generalisation ability of CNNs.

1.5. Establish an optimised workflow for lithological prediction on various sized datasets.

The key objectives of Research Question 2 are:

2.1. Use different GAN architectures to generate carbonate textures from a small training dataset of core images.

2.2. Evaluate the likeness of the synthetic images as representations of realistic carbonate textures using the Fréchet Inception Distance (FID).

2.3. Determine whether the synthetic images generated by the GAN would be correctly classified by a CNN classier trained for carbonate textures.

The key objectives of Research Question 3 are:

3.1. Artificially increase the size of a small dataset using various traditional augmentation techniques such as geometric transformations, photometric transformations, noise injection, and filtering.

3.2. Artificially increase the size of a small dataset through the addition of synthetic images generated by a GAN.

3.3. Perform a comparative evaluation of the traditional and synthetic data augmentation techniques to improve CNN classification performance.

3.4. Evaluate the impact of the different data augmentation techniques to reduce overfitting on a complex geological dataset.

The key objectives of Research Question 4 are:

4.1. Compile an annotated dataset of skeletal components in carbonate core images.

4.2. Use different CNN-based object detection algorithms to identify constituent grains in carbonate core images.
4.3. Compare the performance of one-stage and two-stage detectors using average precision, mean average precision (mAP) and inference time.
4.4. Incorporate error analysis by visualising intermediate feature representations.
4.5. Perform a direct performance comparison with benchmark human classification to indicate the degree of agreement between machine learning and human identification.

1.5 Thesis outline

The outline of this thesis is articulated around a novel workflow for carbonate core classification. This workflow is concerned with the preparation of core images, experimenting with different deep learning models, and comparing the automatic analysis to human-based results. Moreover, the thesis investigates the feasibility of incorporating deep learning workflows and concludes on both the capabilities and limitations of the used methods.

This thesis is divided into six chapters. This chapter forms the introduction to the thesis and sets out the background, Chapter 2 describes the data used in the project, Chapters 3-5 detail the main results and fulfil the objectives of the project, which are then discussed and brought together to answer the original aims in Chapter 6. Each research chapter can be read as a standalone paper but, together, these chapters collectively address the aims and objectives outlined above.

Chapter 2: Digital core image data used in this study

This chapter provides a general overview of the training data used for the subsequent research chapters. The concept of digital core images, image processing, and image analysis are introduced. A brief overview of the geological context of each study area is also provided.

Chapter 3: Application of convolutional neural networks (CNNs) for textural carbonate core classification

This chapter addresses Research Question 1 and outlines a study aiming to enhance geological classification practices by leveraging transfer learning of top-performing CNNs on carbonate core images. As these methods are typically tested using limited datasets (e.g., n = 996 in Kim (2022)), a new and much larger dataset of carbonate core images is compiled that allows for direct performance comparison on datasets of different orders of magnitude. These datasets are used to evaluate the efficacy of the different CNN architectures on complex geological data, and a workflow is presented for automated core classification which successfully predicts Dunham textures. The results of Chapter 3 also indicate the importance of dataset size, as models trained on smaller datasets show a tendency to overfit.

This research chapter was published in **Computers & Geosciences** in January 2023. All of this published work was conceptualised, completed, and written by me, with supervision, review and editing by my supervisors and co-authors Professor Cédric M. John and Professor Oliver Dubrule.

Chapter 4: Generative adversarial networks (GANs) for synthetic carbonate core images

Chapter 4 of this thesis addresses Research Questions 2 and 3 and is motivated by the significance of data scarcity in the geosciences. However, Chapter 4 is also motivated by observations and conclusions made in Chapter 3 that shows how using transfer learning on smaller datasets could be prone to overfitting. This chapter introduces methods for generating synthetic carbonate textures using GANs to address the challenge of obtaining large-scale annotated datasets in the geoscience domain. The generated images are utilised for synthetic data augmentation with the aim of enhancing the performance of CNNs in core image classification. Demonstrated on a limited dataset of core images, the study evaluates two GAN architectures, Deep Convolutional GAN (DCGAN) and StyleGAN2, for synthesising high-quality carbonate textures across seven classes from the modified Dunham Classification (Dunham, 1962; Embry and Klovan, 1971). Furthermore, a CNN is trained on augmented datasets from both traditional data augmentation and synthetic data and compared for reducing overfitting. The findings of Chapter 4 demonstrate the potential of synthetic data augmentation to improve geological classification tasks and suggest applicability to other geological core images.

Chapter 5: Object detection algorithms to identify skeletal components in carbonate cores

This chapter tackles Research Question 4 by addressing the challenge of identifying constituent grains in carbonate rocks, which traditionally demands specialised expertise. Leveraging recent advancements in deep learning-based object detection algorithms, the study explores their application in classifying multiple types of carbonate grains within carbonate core images. By employing transfer learning with pre-trained weights, both one-stage (YOLO v5) and two-stage (Faster R-CNN) detectors are evaluated. Despite the prevalent use of one-stage detectors, the study finds the two-stage detector exhibits the most robust performance for the geological data in question. Additionally, the study extends the approach to provide a direct performance comparison with benchmark human interpretation on images not included in the training set.

This research chapter was published in **Marine and Petroleum Geology** in September 2024. All of this published work was conceptualised, completed, and written by me, with supervision, review and editing by my supervisors and co-author Professor Cédric M. John.

Chapter 6: Discussion and conclusions

Chapter 6 is a discussion of the contributions presented in the individual research chapters, highlighting some of the challenges and opportunities of using deep learning models for core descriptions, and places the methods into a larger context of using computer vision for solving geological image analysis problems. Additionally, it offers reflections and suggestions for future research and development.

Final conclusions of this thesis are presented at the end of Chapter 6.

The research chapters in this thesis use either 'we/our' pronouns or the passive voice, in recognition of co-authors who appear, or who will appear, on the published version of each chapter. In the case of each research chapter, I confirm that I conceptualised and developed the research, conducted the data preparation and data analysis, produced the figures and tables, and wrote the manuscripts. The contributions of co-authors Professor Cédric M. John (Chapters 3–5) and Professor Olivier Dubrule (Chapter 3) included PhD supervision, as well as review and editing of manuscripts, i.e., the research chapters.

All data used in this thesis are available in the public domain and we have made the code used for training, as well as example pre-trained models, available as additional supporting material[1]. The public dataset of high-resolution carbonate core scans made available by the John Lab may serve as benchmark cases to study the application of deep learning to geological core interpretation.

[1] www.john-lab.org/software-and-data/

Chapter 2: Digital core image data used in this study[2]

2.1 Introduction

For all research chapters in this thesis, the datasets used are the derivatives of digital core images of carbonate samples collected from ODP and IODP expeditions. Three expeditions were chosen due to the principal targeting of carbonate bodies; ODP Leg 133 (Davies *et al.*, 1991) and ODP Leg 194 (Isern *et al.*, 2002) off north-eastern Australia, and IODP Expedition 359 in the Maldives archipelago (Betzler *et al.*, 2017). For the subsequent datasets, only high-resolution core images scanned to Tag Image File Format (TIFF) were used. Due to predominantly lossless compression, TIFF files retain the detail and colour depth of images and scans (Murray and vanRyper, 1996).

High-resolution imaging of core sections has been routine since ODP Leg 198, and the digital core scanners used during ocean drilling expeditions currently capture approximately 10-20 pixels per millimetre of core (Wilkens *et al.*, 2009). For the cores drilled during ODP Legs 133 and 194, which predate the routine digital imaging, the scans were performed by C.M. John in 2007 using a prototype line scanner at the Gulf Coast Repository, which is now onboard the JOIDES Resolution.

The digital core images are colour images using the RGB colour model, meaning they are composed of three colour channels: red (R), green (G), and blue (B). These colour channels are each represented as a separate matrix, and the combination of these matrices forms a tensor encoding the image. Each pixel in a colour image holds three intensity values, i.e. one for each of the red, green, and blue channels. The intensity of each channel ranges from 0 to 255, allowing for nearly 16.8 million possible colour variations (Rittner *et al.*, 2010).

Digital core scan images of a 1.5 metre section are approximately 31,600 pixels x 1,750 pixels; however, only the sediment core section itself is required for the purposes of these studies, which is on the order of 29,900 pixels x 1,375 pixels within the image. These individual images can provide some of the highest resolution examples of core properties freely available at present.

Carbonates vary greatly with environment and geologic age, so here we provide the geological context of the cores studied in subsequent research chapters.

---

[2] A version of some information presented in this chapter has been published in **Computers & Geosciences**: **Dawson, H.L.**, Dubrule, O., John, C.M. (2023) Impact of dataset size and convolutional neural network architecture on transfer learning for carbonate rock classification. Computers & Geosciences 171, 105284, doi: 10.1016/j.cageo.2022.105284

2.2 Geological setting of the north-eastern Australian margin

During Ocean Drilling Program (ODP) Leg 133 and Leg 194, conducted on the Queensland and Marion Plateaus off the Great Barrier Reef in northeast Australia, the depositional history of a carbonate-dominated shelf was documented (Figure 2.1) (Isern *et al.*, 2002). Sedimentation commenced on this subsiding basement post-Cretaceous rifting, followed by a transgression during the Oligocene–early Miocene period. Subsequently, two major carbonate platforms, the Northern and Southern Marion Platforms (NMP and SMP, respectively), emerged, reaching maturity in the middle and late Miocene epochs, after which both edifices drowned (Isern *et al.*, 2002). While the NMP was buried under a thin layer of deeper water deposits post-drowning, the SMP remains discernible today with moderate relief above the surrounding seafloor.

Nowadays, the plateau surface is found in a water depth ranging between 100–500 metres (Figure 2.1) (Davies *et al.,* 1991). The sedimentation patterns are strongly influenced by ocean currents, which have led to local and periodic winnowing, resulting in condensed or absent sediments in certain areas and sedimentation concentration in others. Drilling activities during ODP Leg 194 uncovered a series of unconformities marked by the presence of submarine hardgrounds at multiple sites, signifying significant hiatuses in the sedimentary record. These surfaces are also identifiable in seismic data (Figures 2.2 and 2.3), as well as in petrographic, biostratigraphic, and logging data (Isern *et al.*, 2002; Isern *et al.,* 2004). With a combination of seismic stratigraphy and biostratigraphy, these unconformities and the associated hiatuses were dated and contextualised within the overall depositional/stratigraphic framework (Isern *et al.,* 2002).

2.2.1 Geological setting of the Marion Plateau

The Marion Plateau, situated off the northeastern coast of Australia within the Great Barrier Reef Marine Park, served as the location for ODP Leg 194, which targeted eight sites (1192–1199), each penetrating approximately 500–700 metres of sediment from depths ranging between 304 and 419 metres (Figure 2.1). These sites were strategically positioned along two seismic lines (Figures 2.2 and 2.3), intersecting the margins of the NMP and SMP, two Miocene carbonate platforms. The palaeotopography between these platforms has been filled by latest Miocene-Holocene slope and hemipelagic-drift deposits. The SMP, in particular, forms a more ocean-facing promontory compared to the NMP, potentially influencing biota, geochemistry, and diagenesis (Ehrenberg and Dickson, 2003).

ODP Leg 194 was a continuation to ODP Leg 133, which explored nine sites on the Queensland and Marion plateaus in 1990 (Figure 2.1; Davies *et al.*, 1991; Pigram *et al.*, 1992). The primary goal of Leg 194 was to gauge the magnitude of a significant eustatic sea-level fall in the late middle Miocene. Estimates derived from the Leg 194 data range from 86 metres (Isern *et al.*, 2002) to 50 metres (± 5.0 metres) (John *et al.*, 2004). Additional objectives included the examination of carbonate depositional facies and paleoenvironmental controls, dating and characterisation of the main stratigraphic units identified on seismic lines, and exploration of fluid flow, diagenesis, and petrophysical properties (Isern *et al.*, 2001; 2002).

The carbonate platforms of the Marion Plateau exhibit flat tops and steep, asymmetric slopes reminiscent of rimmed tropical platforms dominated by coral reefs, green algae, and ooids, as seen in the Bahamas. However, the biotic assemblages of the Marion platforms primarily comprise red algae, bryozoans, mollusks, and benthic foraminifera, with corals and green algae being less prevalent or absent. Ooids are also notably absent, suggesting a heterozoan association indicative of either cool (subtropical) water temperatures or specific conditions like elevated nutrient supply (Carannante *et al.*, 1988; James and Clarke, 1997; Pomar, 2001; Pomar *et al.*, 2004). Instead of reefal facies, coarse bioclastic debris forms the primary substance and support for the Marion platforms. The influence of oceanic currents during platform growth is inferred from asymmetric slope geometries, sculptured drift deposits, fragmental sediments, and current-swept phosphatic hardgrounds (Isern *et al.*, 2002; Glenn and Kronen, 1993). As discussed by Pomar (2001) and Pomar *et al.*, (2004), the platforms' tops likely reached depths limited by wave base at considerably greater water depths (possibly 40–100 metres) than typical of photozoan platforms.

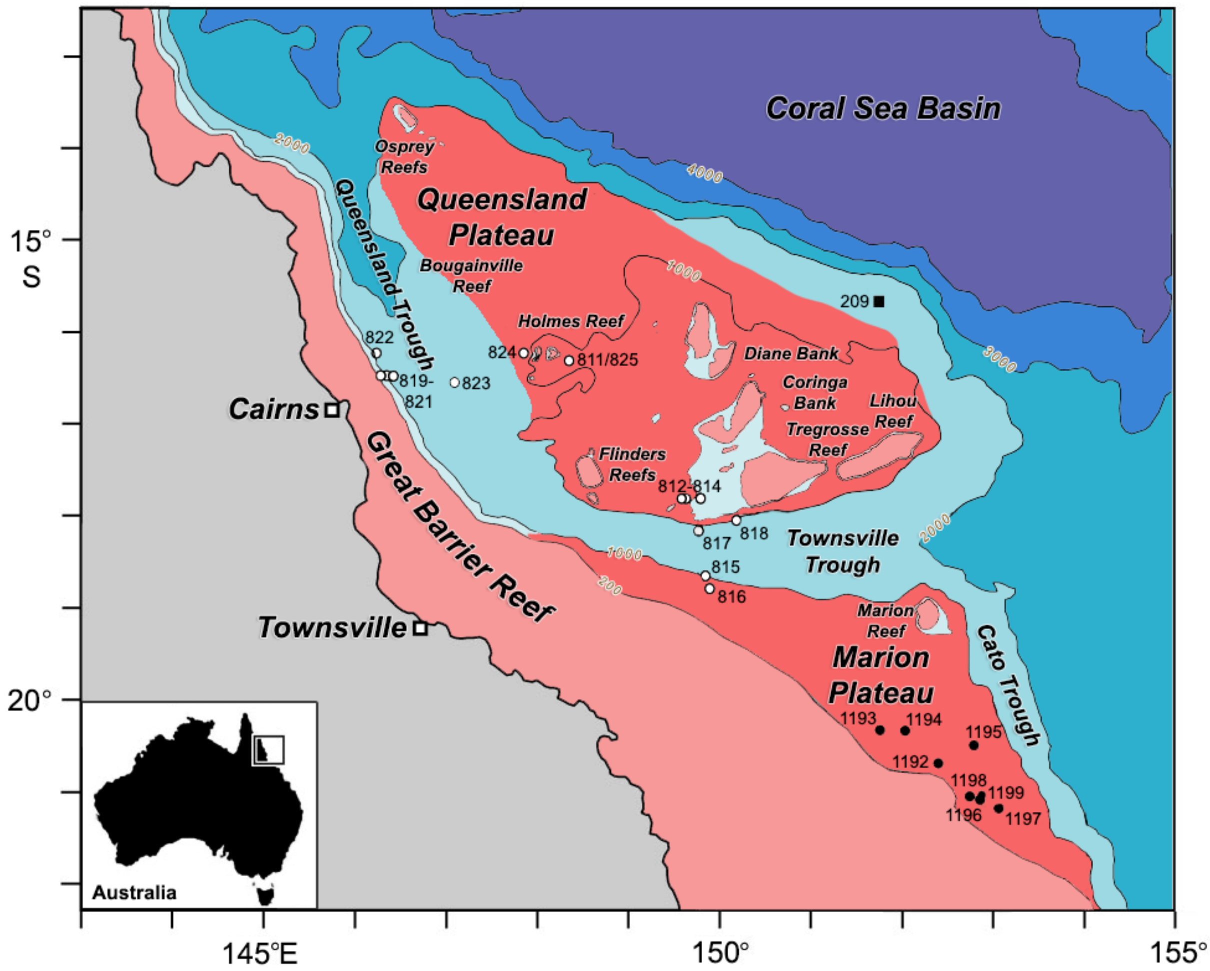

Figure 2.1: Map showing the position of the Queensland and Marion Plateaus relative to bathymetry (in metres) and main physiographic features offshore north-eastern Australia (modified from Heck *et al.,* 2007). Marked are the positions of DSDP Site 209 (solid square), ODP Leg 133 sites (open circles), and ODP Leg 194 sites (solid circles). Areas of modern reef development on the Queensland and Marion Plateaus, as well as the extensive area of the Great Barrier Reef, are indicated in light pink. Inset map shows the location of the margin relative to Australia.

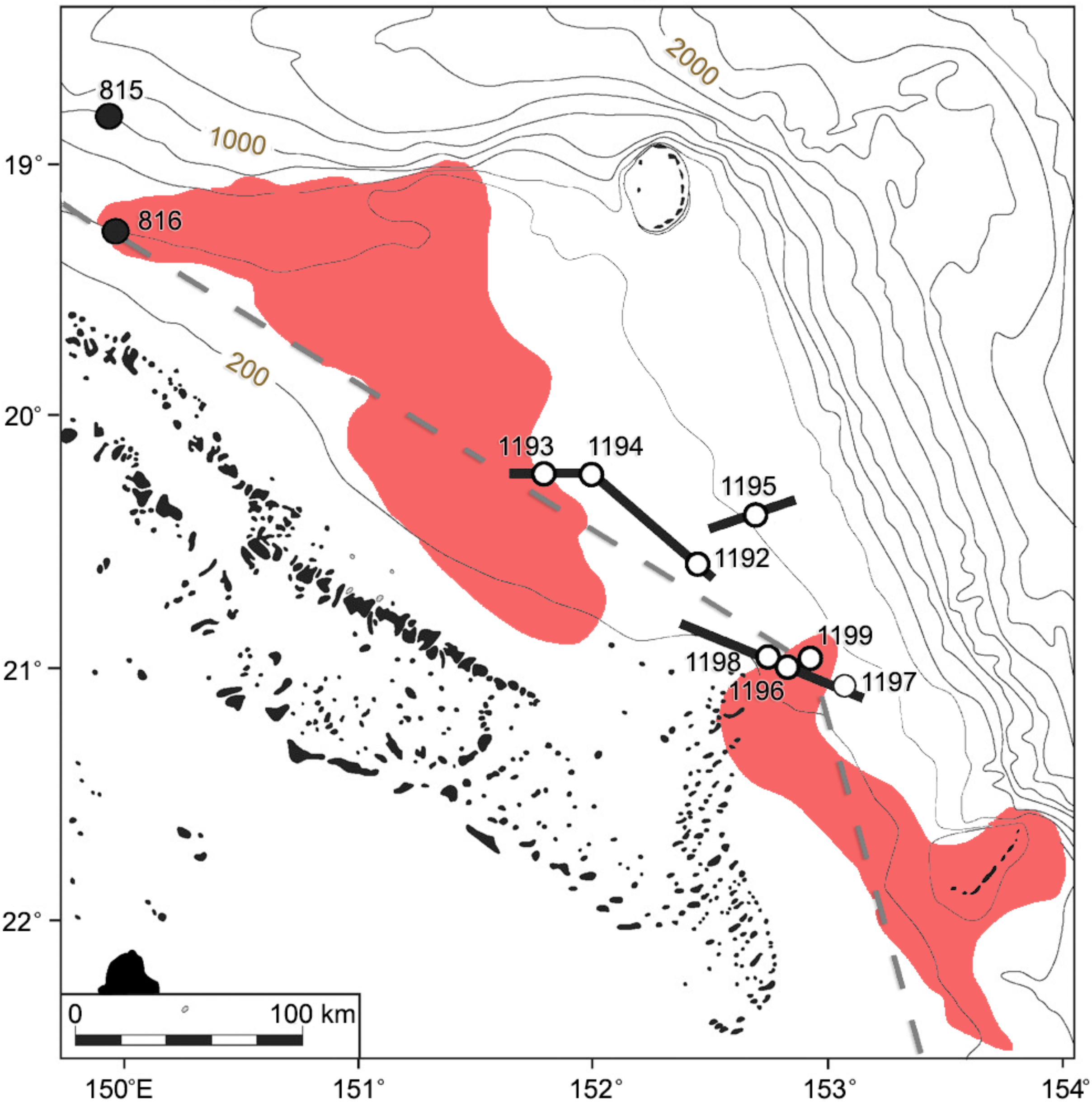

Figure 2.2: Map showing the estimated extent of the Marion Plateau carbonate platforms (NMP and SMP), indicated by the red shaded areas (modified from Ehrenberg *et al.*, 2006). Marked are the positions of ODP Leg 194 sites (solid circles) and two ODP Leg 133 sites (open circles). Seismic lines in Figure 2.3 are shown as thick lines connecting Leg 194 sites. Dashed grey line denotes the boundary of the Great Barrier Reef Marine Park. Bathymetric contours are given in metres, with intervals of 200m.

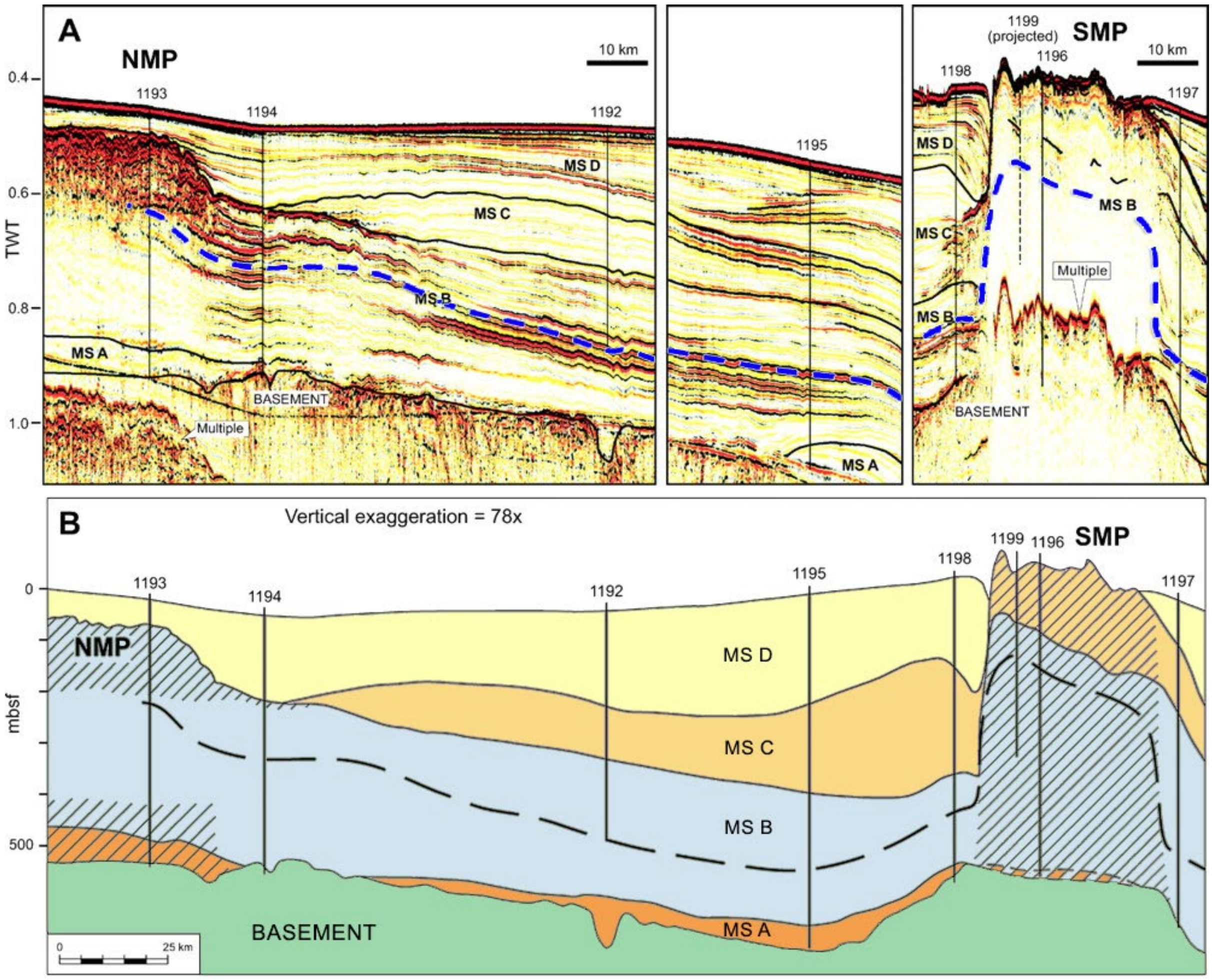

Figure 2.3: Seismic lines and schematic cross section of the two transects across the Northern Marion Platform (NMP) and the Southern Marion Platform (SMP) with the ODP Leg 194 sites and the megasequences (MS) A-D (modified from Isern *et al.*, 2002).

2.2.2 Stratigraphy and biostratigraphic age constraints

The sediments of the Marion Plateau rest atop an irregular basement topography composed of variably altered basaltic volcanic rocks, likely formed during Late Jurassic-Early Cretaceous rifting of the Coral Sea region. Subsequent gradual subsidence of this passive continental margin initiated a marine transgression in the mid-Oligocene period, followed by the development of carbonate platforms throughout the Miocene era, during which the paleolatitude decreased from approximately 32 to 24° S (presently around 20° S; Davies *et al.*, 1989; Hall, 2002).

The stratigraphy of the Marion platforms is delineated based on four megasequences identified from seismic data (Figure 2.3) (Davies *et al.*, 1991; Isern *et al.*, 2002; 2004) and lithological units characterised through shipboard core descriptions. The megasequences are denoted as A through D, with A being the oldest and

D the youngest. Seismic data indicates that the Marion platforms correspond to Megasequences B and C (Figure 2.3), established from seismic imaging of the off-platform slope and drift sediments. Biostratigraphic analysis of the off-platform sites indicates that Megasequences B and C date back to the early to late Miocene period (approximately 22–5 Ma; Isern *et al.*, 2002). However, correlating the seismically layered off-platform sites with the more transparent platform sites entails significant uncertainties due to the poor seismic imaging of platform internal structure (Figure 2.3) and the inferior biostratigraphic controls for the neritic biota of the platform sites compared to the deeper-water biota of the off-platform sites.

2.3 Geological setting of the Maldives archipelago

The Maldives archipelago, situated in the central equatorial Indian Ocean, forms an isolated tropical carbonate platform as part of the Chagos-Laccadive Ridge, southwest of India (Figure 2.4). Characterised by a north-south orientation, a double row of atolls encircles the Inner Sea of the Maldives (Figure 2.4). Inter-atoll channels, deepening toward the Indian Ocean, separate the atolls from each other. The Inner Sea comprises a basin with depths reaching up to 550 metres. The Maldives' carbonate deposition began in the Eocene, devoid of any terrigenous input, and has continued since then (Aubert and Droxler, 1992; Purdy and Bertram, 1993).

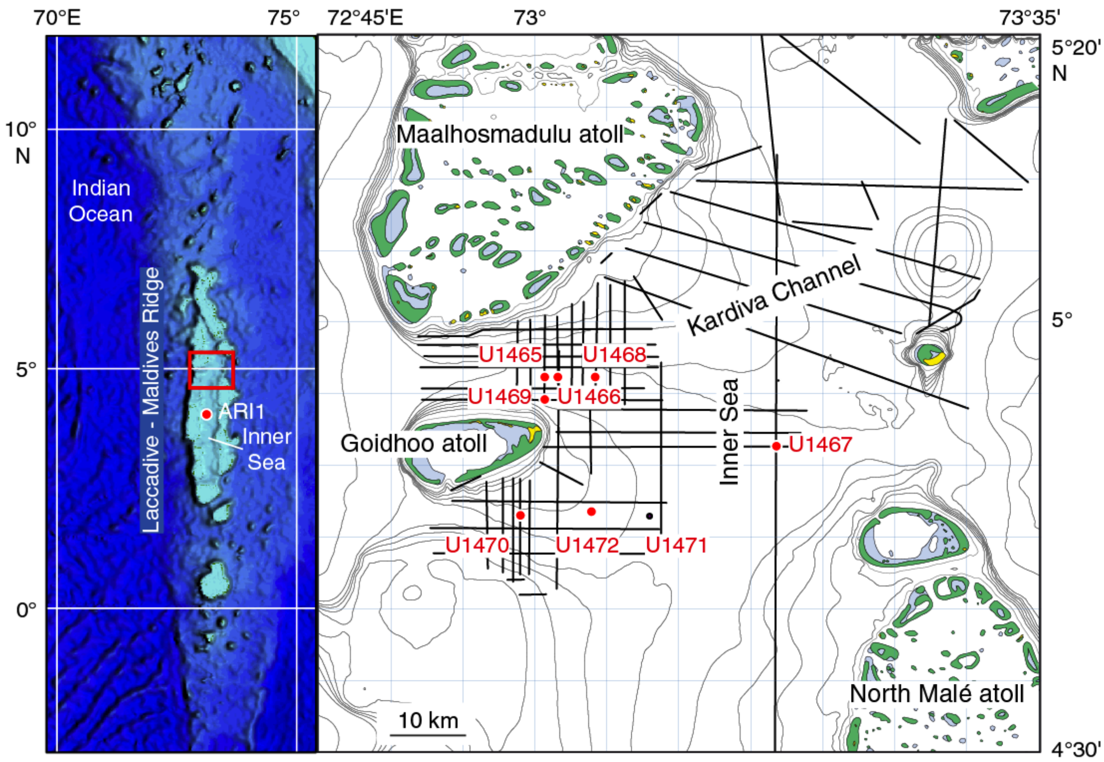

Figure 2.4: Location map of Expedition 359 sites in the Inner Sea of the Maldives (from Betzler *et al.*, 2017).

The modern Maldives archipelago consists of approximately 1,200 smaller atolls positioned near or slightly above the sea surface. These atolls are encircled by discontinuous marginal rims, which delineate lagoons with depths ranging from 50 to 60 metres. Deep passages interrupt these rims, facilitating strong currents within the lagoons that influence sediment redistribution and the development of patch reefs (Ciarapica and Passeri, 1993; Betzler *et al.*, 2015). The marginal reefs are predominantly composed of robust-branching corals and coralline algae, while the lagoonal reefs exhibit domal corals along with detrital sand and rubble facies (Gischler *et al.*, 2008). Muddy sediments are confined to atoll areas sheltered by large marginal reefs (Ciarapica and Passeri, 1993; Gischler, 2006; Betzler *et al.*, 2015).

The oceanward margins of the Maldives archipelago generally exhibit steep inclinations, with dips ranging from 20° to 30°, extending to depths of 2,000 metres. On the Inner Seaside, terraced atoll slopes exhibit similar dip angles but reach depths of only 150 metres, where the gradient lessens abruptly (Fürstenau *et al.*, 2010). The Inner Sea is characterised by periplatform ooze deposition, which locally accumulates into sediment drift bodies (Figure 2.5) (Droxler *et al.*, 1990; Malone *et al.*, 1990; Betzler *et al.*, 2009, 2013a, 2013b; Paul *et al.*, 2012; Lüdmann *et al.*, 2013).

The climatic and oceanographic conditions of the Maldives are governed by the seasonally reversing Indian monsoon system (Tomczak and Godfrey, 2003). During the Northern Hemisphere summer (April–November), southwestern winds prevail, while northeastern winds dominate the winter months (December–March). These winds generate westward currents in winter and eastward currents in summer. Inter-seasonally, equatorial westerlies in the Indian Ocean establish strong eastward-flowing surface currents with velocities reaching up to 1.3 m/s. These currents maintain relatively high velocities, even at depths exceeding 200 metres (Tomczak and Godfrey, 2003). Within the passages of modern atolls, surface currents may reach speeds of up to 2 m/s, contributing to winnowing processes and the formation of hard bottoms in passages and lagoons (Ciarapica and Passeri, 1993; Gischler, 2006; Betzler *et al.*, 2015).

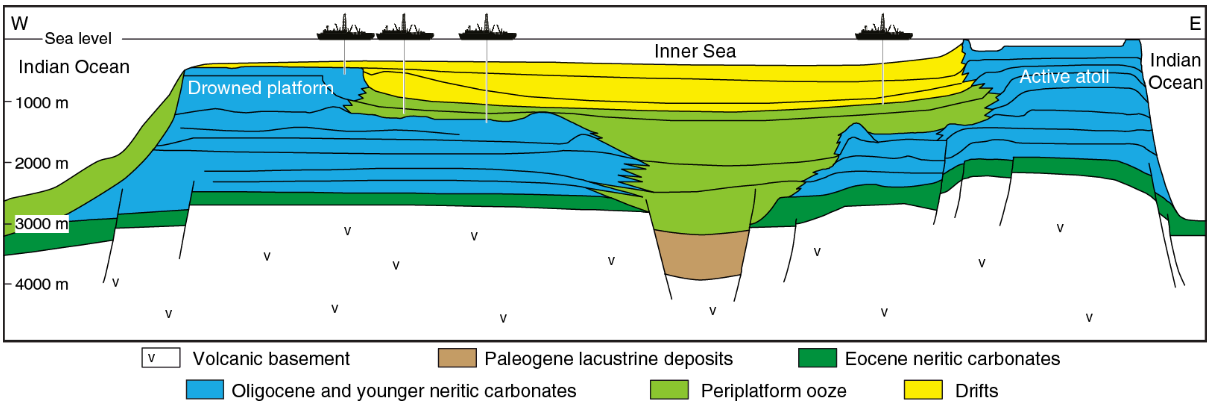

Figure 2.5: Maldives carbonate platform stratigraphy (from Betzler *et al.*, 2017).

The Maldives formed on a lower Palaeogene (60–50 Ma) volcanic basement (Duncan and Hargraves, 1990) (Figure 2.5). Based on previous deep core data, the long-term subsidence rate of the Maldives is estimated to be approximately 0.03–0.04 mm/year (Aubert and Droxler, 1996; Belopolsky and Droxler, 2004). However, sedimentological evidence from the Rasdhoo atoll suggests a substantially faster subsidence rate since the last interglacial period, with a maximum rate of 0.15 mm/year over the past 135,000 years (Gischler *et al.*, 2008).

2.4 Coring conventions

The identification and numbering of sites, holes, cores, and samples adhere to the standard procedure established by the International Ocean Discovery Program (IODP). A complete curatorial identifier for a sample includes details such as the expedition, site, hole, core number, core type, section number, archive or working half (if applicable), piece number (for hard rocks), and the interval in centimetres measured from the top of the core section. For instance, the sample identification "359-U1465A-2H-5 (Piece 1, 4–7 cm)" denotes a 3 cm sample from Piece 1 taken between 4 and 7 cm below the top of Section 5 of Core 2 in Hole A at Site U1465 during Expedition 359 (Figure 2.6). The letter preceding the hole number, such as "U," indicates that the hole was drilled by the US implementing organization (USIO) platform, the JOIDES Resolution. The drilling system used for obtaining a core is denoted in the sample identifiers as follows:

- R = rotary core barrel (RCB);
- H = advanced piston corer (APC);
- X = extended core barrel (XCB).

Core intervals are determined by factors such as the length of the drill string, the seafloor depth, and the distance the core barrel is advanced; these are reported in DSF (drill string feet). The length of the core is the sum of the lengths of its sections. Unconsolidated to partially consolidated sediment cores retrieved from depths of less than a few hundred meters below the seafloor typically expand upon recovery (usually by a few percent to as much as 15%), resulting in a discrepancy between the recovered interval and the cored interval. Additionally, a coring gap often occurs between cores, leading to differences between DSF and core depth below seafloor (CSF) depths for a stratigraphic interval. Moreover, when core recovery exceeds 100% of the cored interval, a sample from the bottom of one core will have a CSF depth deeper than that of a sample from the top of the subsequent core, resulting in overlapping data associated with the two core intervals.

In cases of incomplete core recovery, all cored material is assumed to originate from the top of the drilled interval as a continuous section for curation purposes. However, the true depth interval within the cored interval remains unknown, introducing sampling uncertainty in age-depth analysis or when correlating core data with downhole logging data.

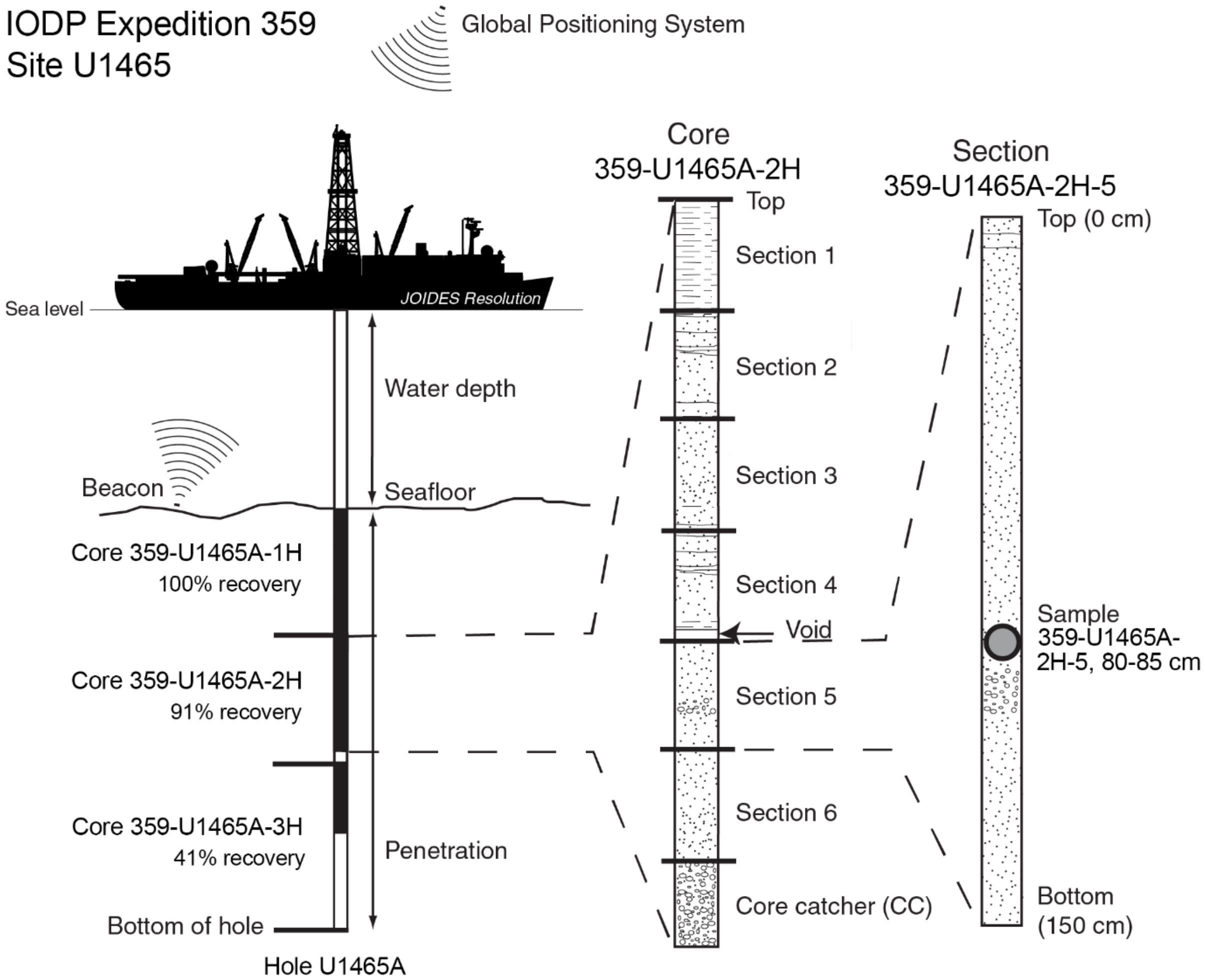

Figure 2.6: IODP conventions for naming sites, holes, cores, and samples (modified from Gallagher *et al.,* 2017).

Chapter 3: Application of Convolutional Neural Networks (CNNs) for Textural Carbonate Core Classification[3]

ABSTRACT

Modern geological practices, in both industry and academia, rely largely on a legacy of observational data at a range of scales. However, widespread ambiguities in the petrographic description of rock facies reduce the reliability of descriptive data. Previous studies have demonstrated potential for the use of convolutional neural networks (CNNs) in the classification of facies from digital images; however, it remains to be determined which of the available CNN architectures performs best for a geological classification task. We evaluate the ability of top performing CNNs from the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) to classify carbonate core images using transfer learning, a method where a pre-trained model is adapted to a new task, systematically developing a performance comparison between these architectures on a complex geological dataset. Three datasets with orders of magnitude difference in data quantity (7,000-104,000 samples) were collected containing images across seven classes from the modified Dunham Classification for carbonate rocks. Following training of nine different CNNs of four architectures on these datasets, we find the Inception-v3 architecture to be most suited to this classification task, achieving 92% accuracy when trained on the larger dataset. Furthermore, we show that even when using transfer learning the size of the dataset plays a key role in the performance of the models, with those trained on the smaller datasets showing a strong tendency to overfit. This has direct implications for the application of deep learning in geosciences as many papers currently published use very small datasets of less than 5,000 samples. Application of the framework developed in this research could aid the future of deep learning based carbonate classification, with further potential to be easily modified to suit the classification of cores originating from different formations and lithologies.

3.1 Introduction

Descriptions of observed geological features play a large role in driving applications and research within the Earth Sciences; for example, when facies descriptions of core data are used to derive regional

[3] A version of this chapter has been published in **Computers & Geosciences**:
Dawson, H.L., Dubrule, O., John, C.M. (2023) Impact of dataset size and convolutional neural network architecture on transfer learning for carbonate rock classification. Computers & Geosciences 171, 105284. doi:10.1016/j.cageo.2022.105284

stratigraphic trends, or as the basis for petrophysical classifications (Hull *et al.*, 2015; John and Kanagandran, 2019). The modified Dunham classification (Dunham, 1962; Embry and Klovan, 1971) is acknowledged as one of the most commonly used classification schemes for the systematic description of carbonate rocks. However, recent studies have shown that even with the well-established and clearly defined divisions of the scheme, experienced sedimentologists may often classify similar facies using different textural names (Lokier and Al Junaibi, 2016). Deep learning presents a data-driven approach to deriving predictive models from observational data.

The use of machine learning and deep learning in geoscience is rapidly growing; however, progression has been relatively uneven and, despite increasing popularity within the geoscientific community, practitioners have been comparatively slow to engage with these recent advancements (Bergen *et al.*, 2019; Reichstein *et al.*, 2019). Previous classification tasks have focused largely on interpreting data from seismic or from wireline logs (e.g., Hall, 2016; Qian *et al.*, 2018; Saporetti *et al.*, 2018; Halotel *et al.*, 2020). A pilot study in the John Lab research group, formally the Carbonate Research Group, (John and Kanagandran, 2019) tested the effectiveness of neural networks in recognising carbonate rock facies according to the Dunham classification scheme. Using high-resolution core images from Ocean Drilling Program (ODP) Leg 194, the convolutional neural network (CNN) model achieved 89.2% accuracy across seven classes, based on a supervised learning approach with labelled data. The results from this study show great potential for the use of deep learning in carbonate classification from digital images. Several studies now have applied CNN models to broad lithofacies classifications in both core and thin section images (e.g., Baraboshkin *et al.*, 2020; Pires de Lima *et al.*, 2020; Koeshidayatullah *et al.*, 2020a; Chawshin *et al.*, 2021; Germay *et al.*, 2023). However, it remains to be determined which of the available CNN architectures performs best for the classification of carbonate core images.

This chapter evaluates the ability of a selection of the top performing CNNs from the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) to classify carbonate core images, systematically developing a performance comparison between these architectures on a complex geological dataset. This deep learning approach is applied to three datasets with orders of magnitude difference in data quantity (a 104k dataset, a 42k dataset, and a 7k dataset), highlighting the impact of dataset size on the performance and reliable application of the models. The findings of this research serve as a necessary first step for further research into training deep learning algorithms for carbonate textural classifications. Application of these advanced digital technologies may have the potential to revolutionise descriptive geology, with more accurate and uniform facies descriptions, faster analysis times, and reduction of natural biases.

3.1.1 Background

Machine learning has been explored in various subdisciplines of the Earth Sciences for nearly 60 years (e.g., Preston and Henderson, 1964; Wickman, 1965; Krumbein and Dacey, 1969), with a marked uptake in successful applications over the past three decades (e.g., Baldwin *et al.*, 1990; Brown *et al.*, 2000; de Matos *et al.*, 2007; Pires de Lima *et al.*, 2019). Current deep learning applications within the geoscientific community include, but are not limited to, seismic facies classification (e.g. West *et al.*, 2002; Chevitarese *et al.*, 2018), lithofacies classification from wireline logs (e.g. Bestagini *et al.*, 2017; Halotel *et al.*, 2020), volcanic ash detection (e.g. Shoji *et al.*, 2018; Torrisi *et al.*, 2021), bioturbation identification (e.g. Timmer *et al.*, 2021), glacial varve detection (e.g. Fabijanska *et al.*, 2020), seismology (e.g. Kortström *et al.*, 2016; Wu *et al.*, 2018; Mousavi *et al.*, 2019), and inverse problems (e.g. Mosser *et al.*, 2020).

Despite the successes of these approaches, many existing datasets from long-established sources remain largely unexplored. This is partly due to the lack of properly labelled data and the inefficiency of traditional analytical techniques, which are typically resource intensive (Oikonomou *et al.*, 2017). Deep learning offers a promising solution to this challenge by enabling the efficient processing of large volumes of multidimensional data, thereby making these analytical workflows significantly more time and cost effective. The application of machine learning is expected to drive substantial progress in the automation of complex prediction tasks, the solution of inverse problems through multiscale modelling, and the discovery of new or alternative patterns and relationships that are not readily visible to humans (Karpatne *et al.*, 2019; Mosser *et al.*, 2020). Recent efforts to promote the use of benchmark geoscientific datasets in global and local competitions have the potential to drive deeper, broader, and more collaborative efforts e.g. the SEG Machine Learning Facies Classification Contest (Hall, 2016; Hall and Hall, 2017), the TGS Salt Identification Challenge (Kaggle, 2019) and the FORCE 2020 Machine Learning Competition (Bormann *et al.*, 2020).

CNNs have dominated computer vision research in recent years, with deep CNNs becoming one of the most widely adopted methods for image classification across various domains (Nguyen *et al.*, 2018). Their suitability for these tasks lies in their architecture, which is specifically designed to automatically and adaptively learn spatial hierarchies of features from images. Unlike traditional algorithms that require handcrafted feature extraction, CNNs use convolutional layers to detect low-level patterns (e.g., edges and textures) in the initial layers and progressively learn more complex structures (e.g., shapes and objects) in deeper layers. This hierarchical feature extraction capability makes CNNs highly effective for recognising and classifying objects, even in complex, high-dimensional image datasets (Yamashita et al., 2018). With an increase in depth of a neural network, performance is indeed improved, but this also results in an increase in computational resources and training time. Training CNNs from scratch, by creating a new model and training it entirely with raw data, typically requires datasets containing millions of images across thousands of classes, such as the ImageNet dataset (http://www.image-net.org/; Deng *et al.*, 2009). Such extensive data is essential for the model to generalise well, but this requirement can be a significant barrier in domains with limited labelled data.

Although many geoscience applications may involve large amounts of data, a common problem is the paucity of datasets with ground-truth labels. This is, in part, due to the laborious and expensive nature of geological data collection (Klump and Robertson, 2015). Datasets may also contain noise and missing values; while many variables cannot be measured directly, but only inferred from observations. Many geological applications rely on an iterative process of collecting different samples and subjectively applying descriptive geological interpretations to known instances. While data availability in geosciences is increasing, it is still comparatively small and sparse given the complexity of the phenomena under study. In deep learning, limited representative training samples may result in poor performance of algorithms, foften due to overfitting. Overfitting occurs when a model is too complex relative to the dataset, capturing noise as well as the general underlying patterns (Kuhn and Johnson, 2013). Model complexity is typically measured by the number of trainable parameters, which is determined by the architecture of the neural network, including the number of layers, nodes per layer, and convolutional filters in CNNs.

Although it is possible to adjust model architecture to reduce complexity and mitigate overfitting, there are inherent challenges in doing so. In image-based applications, for example, the number of parameters is directly influenced by the size and resolution of the input images (Ranjan *et al.*, 2023). High-resolution images significantly increase the number of parameters in a model because larger images require more variables to process their detailed features (Thambawita *et al.*, 2021). When working with small datasets, the total number of available images is often too low to effectively train such a high-parameter model. This mismatch between the large number of parameters and the limited amount of training data makes it challenging to design a model that is both complex enough to capture the important features of the data and simple enough to avoid overfitting.

One promising method of applying deep CNNs to limited datasets and reducing training time is transfer learning (e.g. Pan and Yang, 2010; Shin *et al.*, 2016; Tan *et al.*, 2018). This approach utilises a pre-trained neural network to act as a feature extractor for a smaller dataset. The core idea is that CNNs trained on significantly large and diverse image datasets, such as ImageNet, learn features in their early layers that are "generic" — that is, broadly applicable to various types of image datasets (Yosinski *et al.*, 2014). For example, early layers often detect simple patterns like edges and textures, while deeper layers identify more domain-specific features. These generic features can then be reused for different tasks.

However, the assumption that these features are generic requires careful consideration. The term "generic" implies that the extracted features are not overly specialised to the original dataset and can generalise to new datasets with similar properties. While many studies (e.g., Yosinski et al., 2014) provide evidence supporting this generalisation, it is not guaranteed in all cases. The degree of "genericity" often depends on the similarity between the source dataset used for pre-training (e.g., ImageNet) and the target dataset (e.g., core images). Measuring this generality is challenging but can be approximated by evaluating the transfer performance on the target dataset (Zhuang *et al.*, 2020). If performance is high without extensive fine-tuning, the pre-trained features are likely sufficiently generic.

Several publications have successfully applied machine learning, deep learning, and transfer learning approaches to image-based lithological classifications of well logs, borehole images, thin sections, microtomographic (micro-CT) images, Formation Micro-Imager logs (FMI), and core images (Table 3.1). Though CNN-based lithological classifications were investigated by multiple approaches, many limitations and requirements remain. Among these limitations, we cite:

(1) The majority of studies are based on limited datasets containing under 10,000 data points. Of the studies presented in Table 3.1, 71% used datasets of under 10,000 data points. Of those, 88% are under 5,000 data points and a further 44% contain ≤1,000 data. It is not examined within the reports, what effect these limited data may have on the generalisation of the models presented to further unseen data.

(2) Whilst high classification results are reported, many of the accuracies and F1-scores provided are either based on training data, or it is unclear whether these results were achieved on training and validation data or on test data. Since the test set is a sample of unseen data that is used to provide an unbiased measure of the final model performance; by not evaluating their proposed models on a test set, these studies do not investigate how well they will generalise when new data are introduced.

(3) Most efforts towards automated lithological predictions are based on well log data, including sonic, neutron, resistivity, gamma, and density logs, while the associated core from these logs is often overlooked. Despite capturing distinct rock properties, traditional well log data are one-dimensional and leverage limited information on lithologic characteristics. Core analysis is a key pillar in precise lithological classification, with the majority of well log studies assigning ground-truth labels based on interpretations made through the related cores. More work is needed to extract meaningful insights from complex core image data.

(4) Existing works fail to consider the trade-off between classification performance and computational complexity of the models. This quantitative metric can determine the suitability of the models for real-world applications in real-time, where the computational cost must satisfy the computational budget of the equipment.

(5) Many advances in automatic lithological interpretation employ private data that are limited to a specific well site, reservoir, or depositional setting. It is, therefore, difficult to apply the trained models to other data, greatly limiting the practicality for more general challenges.

Table 3.1: Summary of techniques and comparison between studies for lithological classification. The table is divided into groups based on data type used in the studies and organised in chronological order.

Author(s) and Year	Dataset	Number of Classes	Number of Data Points	Method	Model/ Architecture	Result
			CORE IMAGES			
Thomas *et al.*, 2011	Mixed Lithology Core Images	4	315	Machine Learning	Nearest Neighbour	94.29% *(Accuracy)*
Insua *et al.*, 2015	Carbonate Core Images	6	3,852	Machine Learning	Random Forest	79% *(Accuracy)*
Ivchenko *et al.*, 2018	Mixed Lithology Core Images	4	800	Deep Learning	Custom CNN	88.7% *(Accuracy)*
John and Kanagandran, 2019	Carbonate Core Images	7	35,197	Transfer Learning	Inception-v3	89.2% *(Accuracy)*
Mandal and Rezaee, 2019	Mixed Lithology Core Image and Well Log Data	9	4,149	Machine Learning	Custom ANN	80% *(Accuracy)*
Miller *et al.*, 2019	Mixed Lithofacies Core Images	6	min: 120 max: 240[4]	Deep Learning	Custom CNN	98% *(Accuracy)*
Pires de Lima *et al.*, 2019a	Carbonate (Limestone) Core Images	5	285	Transfer Learning	MobileNet-V2	100% *(Accuracy)*
Pires de Lima *et al.*, 2019b	Carbonate-Siliciclastic Core Images	11	7,039	Transfer Learning	ResNetV2	95% *(Accuracy)*
Baraboshkin *et al.*, 2020	Mixed Lithology Core Images	6	85,600	Transfer Learning	GoogLeNet	57% *(F1-Score)*
Alzubaidi *et al.*, 2021	Carbonate-Siliciclastic Core Images	4	76,500	Transfer Learning	ResNeXt-50	93.12% *(Accuracy)*
Chawshin *et al.*, 2021	Mixed Lithofacies Core CT Scans	20	225,524	Deep Learning	Custom CNN	51.6% *(F1-Score)*
Martin *et al.*, 2021	Mixed Lithology Core Images (using Colour Channel Logs)	5	~65,900[5]	Supervised Machine Learning	Custom Model	69% *(Accuracy)*

[4] The authors do not provide the size of the dataset, but instead report 20 to 40 images per class.
[5] The authors do not provide the size of the dataset, but instead report 659m of labelled core. This estimate is based on the assumption of centimetre-scale labelling.

Zhang *et al.*, 2021a	Mixed Lithology Core Images	3	32,023	Deep Learning	Custom CNN	82% (*Accuracy*)
Falivene *et al.*, 2022	Mixed Lithofacies Core Images	6	34,855	Transfer Learning	SegNet (with VGG16 weights)	80% (*Accuracy*)
Fu *et al.*, 2022	Mixed Lithology Core Images	10	15,000	Transfer Learning	ResNeSt-50	99.6% (*Accuracy*)
Kim, 2022	Carbonate Core Images and Well Log Data	4	996	Deep Learning	Custom CNN	75.2% (*Accuracy*)
Solum *et al.*, 2022	Mixed Lithofacies Core Images	7	~3,667[6]	Transfer Learning	SegNet (with VGG16 weights)	79.4% (*Accuracy*)
Germay *et al.*, 2023	Mixed Lithology Core Images and "Hard Data" (e.g., XRF, Grain Size)	27	479	Transfer Learning	EfficientNet	99% (*Accuracy*)
Larson *et al.*, 2023	Carbonate-Siliciclastic Core Images and XRF Data	8	2,121	Deep Learning	Multilayer Perceptron (MLP)	88.4% (*Precision*)
THIN SECTION IMAGES						
Jobe *et al.*, 2018	Carbonate Thin Section Images	6	~9,000	Transfer Learning and Fine-Tuning	Inception-v3	83% (*Accuracy*)
Koeshidayatullah *et al.*, 2020a	Carbonate Thin Section Images	6	~13,000	Transfer Learning and Fine-Tuning	Inception-ResNet v2	92% (*F1-Score*)
Pires de Lima *et al.*, 2020	Siltstone Thin Section Images	5	5,515	Transfer Learning and Fine-Tuning	ResNet50 Inception-v3	96% (*Accuracy*)
OTHER IMAGES						
Zhang *et al.*, 2017	Mixed Lithology Borehole Images	3	1,500	Deep Learning	Custom CNN	95%
dos Anjos *et al.*, 2021	Carbonate Rock Micro-CT Images	3	6,000	Deep Learning	Custom CNN	81.33% (*Accuracy*)

[6] The authors do not provide the size of the dataset, but instead report >1,100m of labelled core split into 0.3m segments.

Jiang *et al.*, 2021	Borehole-Resistivity Images	2	5,156 (cleaned dataset)	Deep Learning	Custom CNN	84.7% *(Accuracy)*
Tian *et al.*, 2023a	Shale Lithofacies FMI Images	5	1,362	Machine Learning	RF with Gray Level Co-Occurrence Matrix	76% (*Accuracy*)
			WELL LOG DATA			
Li and Anderson-Sprecher, 2006	Mixed Lithology Well Logs	5	593	Machine Learning	Naïve Bayes Classifier	72% *(Accuracy)*
Qi and Carr, 2006	Carbonate Well Logs	6	2,023	Machine Learning	Custom ANN	85% *(Accuracy)*
Dubois *et al.*, 2007	Carbonate-Siliciclastic Well Logs	8	3,647	Machine Learning	Custom ANN	78% (*Accuracy*)
Al-Anazi and Gates, 2010	Clastic Well Logs	3	865	Machine Learning	Support Vector Machine (SVM)	72% (*Accuracy*)
Gifford and Agah, 2010	Carbonate-Siliciclastic Well Logs	8	2,274	Machine Learning	Multi-Agent Collaborative Learning	84.5% (*Accuracy*)
Tang *et al.*, 2011	Carbonate Well Logs	8	~1,000	Deep Learning	Probabilistic Neural Network (PNN)	~60% *(Accuracy)*
Wang and Carr, 2012	Shale Lithofacies Well Logs	7	182	Machine Learning	Custom ANN	85.71% (*Accuracy*)
Bhattacharya *et al.*, 2016	Shale Lithofacies Well Logs	7	698	Machine Learning	SVM	87.3% *(Accuracy)*
Hall, 2016	Carbonate-Siliciclastic Well Logs	9	4,149	Machine Learning	SVM	43% (*F1-Score*)
Al-Mudhafar, 2017	Mixed Lithology Well Logs	7	669	Deep Learning	PNN	95.81% (*F1-Score*)
Bestagini *et al.*, 2017	Carbonate-Siliciclastic Well Logs	9	4,149	Machine Learning	Gradient Boosting Classifier	61% (*F1-Score*)
Tschannen *et al.*, 2017	Carbonate-Siliciclastic Well Logs	9	800	Transfer Learning(?)	GoogLeNet	57.4% (*F1-Score*)
Xie *et al.*, 2018	Mixed Lithology Well Logs	8	915	Machine Learning	Gradient Tree Boosting	82.2% (*F1-Score*)
Dell'Aversana, 2019	Mixed Lithology Well Logs	6	~21,000	Machine Learning	Random Forest (RF)	99% *(F1-Score)*

He *et al.*, 2019	Mixed Lithology Well Logs	10	12,194	Machine Learning	Multilayer Perceptron (MLP)	83% *(Accuracy)*
Imamverdiyev and Sukhostat, 2019	Mixed Lithology Well Logs	9	4,149	Deep Learning	Custom 1D-CNN	93.2% *(Accuracy)*
Valentín *et al.*, 2019	Mixed Lithology Well Logs	4	187,605	Deep Learning	Custom CNN	81.45% *(Accuracy)*
Bressan *et al*,, 2020	Mixed Lithology Well Logs	4	24,504	Machine Learning	MLP	85.08% *(Accuracy)*
He *et al.*, 2020	Mixed Lithology Well Logs	7	462	Deep Learning	Custom Deep Neural Network (DNN)	60.1% *(F1-Score)*
Hossain *et al.*, 2020	Mixed Lithology Well Logs	5	2,747	Machine Learning	Rough Set Theory	98% *(Accuracy)*
Lan *et al.*, 2021	Carbonate Well Logs	5	1,260	Semi-Supervised Machine Learning	PU-Learning	85.80% *(Accuracy)*
Merembayev *et al.*, 2021	Mixed Lithology Well Logs	5	59,423	Machine Learning	RF	97.7% *(Accuracy)*
Saporetti *et al.*, 2021	Carbonate-Siliciclastic Well Logs	7	40	Deep Learning	Custom ANN	73.6% *(Accuracy)*
Al-Mudhafar *et al.*, 2022	Carbonate Well Logs	4	399	Machine Learning	Extreme Gradient Boosting (XGBoost)	98% *(Accuracy)*
Li *et al.*, 2022	Mixed Lithofacies Well Logs	7	1,238	Machine Learning	gcELM	93.62% *(Accuracy)*
Liu and Liu, 2022	Mixed Lithofacies Well Logs	3	3,490	Deep Learning	Hybrid CNN-LSTM	87.3% *(Accuracy)*
Wood, 2022	Carbonate-Siliciclastic Well Logs	9	4,149	Machine Learning	Support Vector Classifier (SVC)	77.2% *(Accuracy)*
Zheng *et al.*, 2022	Mixed Lithofacies Well Logs	9	109,894	Machine Learning	XGBoost	80% *(Accuracy)*
Dong *et al.*, 2023	Mixed Lithology Well Logs	8	915	Machine Learning	Ensemble Model: SVM + AdaBoost + LightGBM	84.9% *(Accuracy)*
Hou *et al.*, 2023	Shale Lithofacies Well Logs	6	*Not specified*	Machine Learning	RF	86.6% *(F1-Score)*

Mubarak and Koeshidayatullah, 2023	Mixed Lithology Well Logs	4	~5,000,000	Machine Learning	AutoML	98% (*F1-Score*)
Ren *et al.*, 2023	Mixed Lithology Well Logs	4	1,305	Machine Learning	Fuzzy Decision Tree	85.5% (*Accuracy*)

Based on the results from the previous studies (Table 3.1) and owing to the relatively small size of geological datasets, transfer learning was identified as the best method for the image classification tasks in this work. This study will demonstrate how deep learning algorithms can be used to elicit information from digital carbonate core images to improve the reliability and objectivity of classifications. It will also demonstrate that despite the clear benefit of transfer learning, dataset size plays an important role, and many geological datasets are still considered too small to train an algorithm that generalises well to unseen data.

This limitation points to a broader challenge within the field of artificial intelligence: the reliance of deep neural networks on large, diverse datasets to achieve generalisability. Neural networks excel at identifying patterns in the data they are trained on, but their ability to generalise is often constrained by the representativeness and scale of the training data (Samek *et al.*, 2021). This raises a fundamental question about the scalability of AI solutions: if training data is limited in quantity or scope, the resulting models may be overly specific, reducing their usefulness across broader applications. In domains like geoscience, where datasets are inherently smaller and more specialised, this issue may be particularly pronounced.

These challenges suggest that the current AI boom may encounter limitations as the focus shifts from well-resourced domains with extensive datasets, such as those based on internet-scale data, to fields with more constrained resources. Developing strategies to address these limitations will be crucial for ensuring that deep learning remains a viable tool across diverse scientific and industrial domains.

All data used in this study, as is the case for the rest of this thesis, are available in the public domain[7]. As additional supporting material, we have provided a code that can be used to easily to perform transfer learning for geological images using any of the built-in TensforFlow-Keras image classification models used in this chapter[8].

Carbonate classification schemes

A consistent classification of rocks and sediments is crucial for ensuring effective communication and dissemination throughout the industrial and academic communities (Scholle and Umer-Scholle, 2003). There have been many proposed classification schemes for the description of carbonate rocks, which are

[7] https://web.iodp.tamu.edu/LORE/
[8] https://github.com/johnlab-research/dawson-facies-2022

intended to determine composition and, subsequently, depositional setting (e.g. Bramkamp and Powers, 1958; Folk, 1959, 1962; Dunham, 1962; Leighton and Pendexter, 1962; Nelson *et al.*, 1962; Todd, 1966; Embry and Klovan, 1971; Wright, 1992). Of these, the classification schemes by Folk (1959, 1962) and Dunham (1962), with modifications from Embry and Klovan (1971), have been the most widely adopted in industry and academia (Lokier and Al Junaibi, 2016).

The Folk (1959, 1962) classification is a textural scale that encompasses grain size, grain composition, packing, roundness, and sorting. A petrographic microscope in a research setting is, therefore, more applicable for concise naming according to the Folk classification. The modified Dunham classification is primarily textural in nature and is based on the characterisation of the overall support framework (Figures 3.1 and 3.2). This makes the scheme relatively easy to use in the field or when looking at cores or thin sections. The modified Dunham classification is primarily used in this project as it yields more relevant information on the process and environment of carbonate deposition. It should be noted that, for this study, we have defined seven principal names; choosing to omit the crystalline carbonate class from Dunham (1962) (Figure 3.1) and the three subclasses of boundstone from Embry and Klovan (1971) (Figure 3.2), instead adopting the original boundstone class from Dunham (Figure 3.3):

Mudstone = mud-supported fabric with less than 10% grains.

Wackestone = mud-supported fabric with more than 10% grains.

Packstone = grain-supported fabric containing intergranular mud.

Grainstone = grain-supported fabric containing no mud.

Floatstone = matrix-supported fabric, with at least 10% of grains ≥2 mm.

Rudstone = grain-supported fabric, with at least 10% of grains ≥2 mm.

Boundstone = original components organically bound during deposition.

Within these definitions, the 'grain' component is >30 μm mean diameter (i.e., visible to the eye or determinable with a hand lens) and the 'mud' component is <30 μm (Embry and Klovan, 1971). The matrix within floatstone and rudstone classes may range from mudstone to grainstone. For example, a sample containing rhodoliths of 15 mm diameter with grains in contact with one another and having a grainstone matrix would be classed as a rudstone.

Strictly speaking, the Dunham (1962) textural classification applies solely to the depositional textures of lithified material; however, we have extended the Dunham classifications to include unlithified and partially lithified carbonate sediments because the textures of the carbonate rocks and the sediments recovered in the cores are very similar. No distinction is made in this study between different levels of lithification, with all rocks and sediments classed solely on a textural basis.

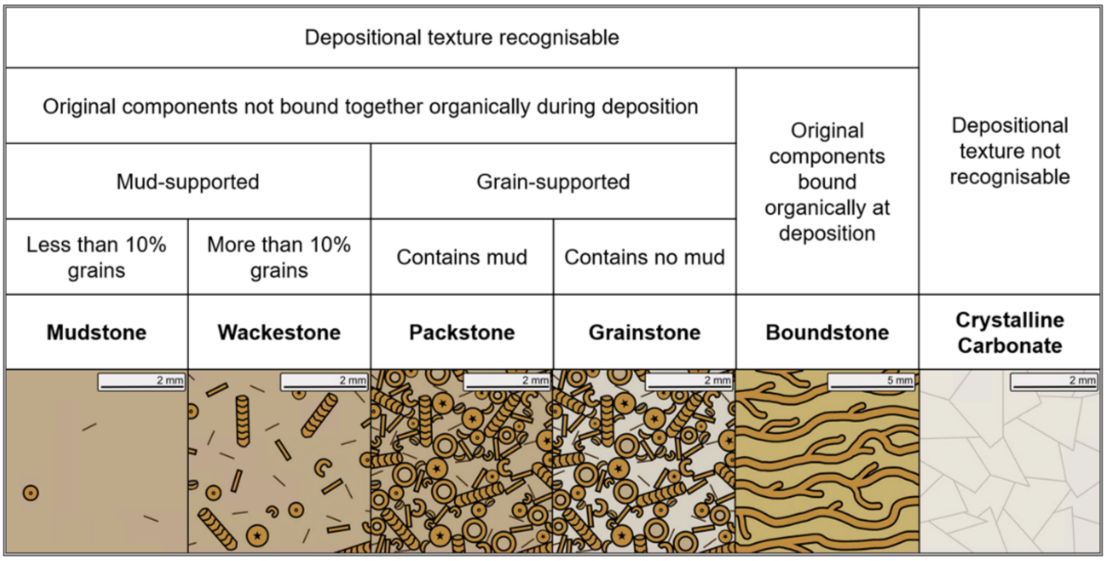

Figure 3.1: The Dunham classification of carbonate rocks according to their depositional texture (after Dunham, 1962).

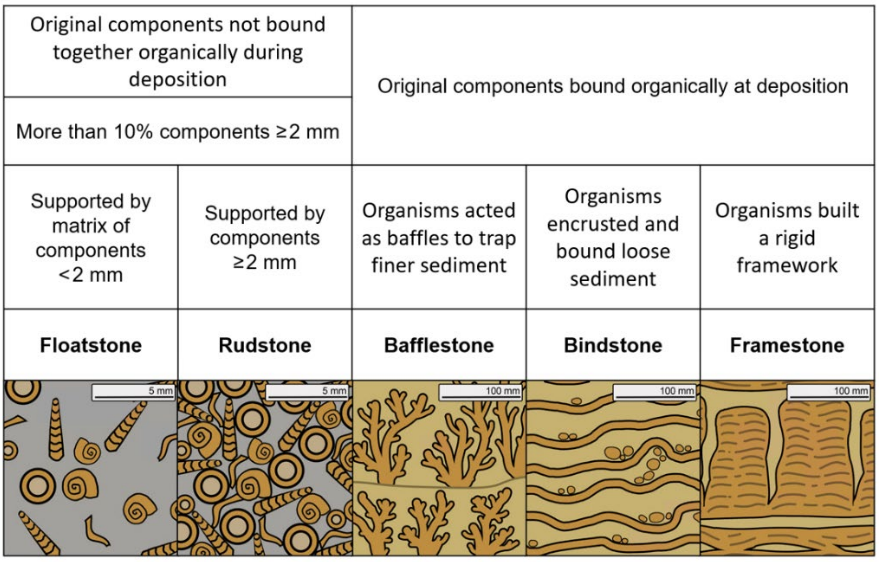

Figure 3.2: Summary of the major modifications to the Dunham (1962) classification provided by the Embry and Klovan (1971) scheme (after Embry and Klovan, 1971). For coarse-grained biogenic limestones (with constituent grains >2mm), the terms 'floatstone' and 'rudstone' replace 'wackestone' and 'packstone', respectively. The 'boundstone' class from Dunham is further subdivided into three subclasses based on the mode by which the sediment was organically-bound. These terms have been widely adopted by studies on reefs, bioherms, and other biogenic carbonates.

Allochthonous						Autochthonous
Original components not bound together organically during deposition						Original components bound organically at deposition
Less than 10% components ≥2 mm				More than 10% components ≥2 mm		
Mud-supported		Grain-supported		Supported by matrix of components <2 mm	Supported by components ≥2 mm	
Less than 10% grains	More than 10% grains	Contains mud	Contains no mud			
Mudstone	**Wackestone**	**Packstone**	**Grainstone**	**Floatstone**	**Rudstone**	**Boundstone**

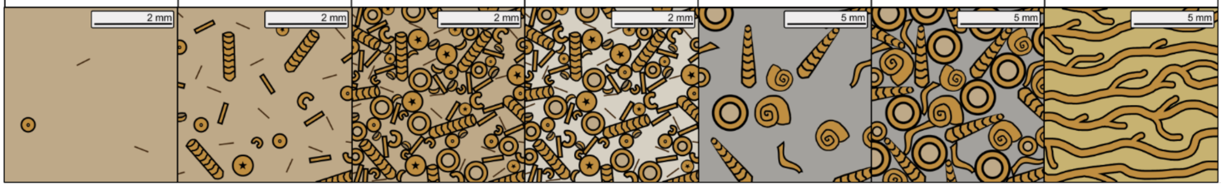

Figure 3.3: The modified Dunham classification scheme showing the seven classes used in this research (after Dunham, 1962; Embry and Klovan, 1971).

3.2 Methodology

3.2.1 Computing environment

All training and evaluation stages were completed using Python version 3.8 (https://www.python.org) with the *TensorFlow* (version 2.0) (Abadi *et al.*, 2016), *tensorflow-keras* backend (version 2.4) (Chollet, 2015), *FastAI* (Howard and Gugger, 2020), *Scikit-learn* (version 0.22) (Pedregosa *et al.*, 2011), *NumPy* (version 1.17) (Oliphant, 2006), *OpenCV* (version 4.1.2) (Bradski, 2000) and *Matplotlib* (version 2.2.4) (Hunter, 2007) libraries. In a transfer learning approach to image classification, the listed libraries play complementary roles, supporting tasks from data preparation and model training to evaluation and visualisation. Details of the contributions from each library are given below. The deep learning frameworks were implemented using the CPU of a MacBook Pro 2019 model with a 2.4GHz 8-Core Intel Core i9 processor and 64GB of RAM, running the macOS Catalina 10.15.7 operating system.

TensorFlow

TensorFlow is a powerful deep learning framework that support training and deploying machine learning models. In transfer learning, TensorFlow provides robust tools for loading pre-trained models from its ModelHub. These models, trained on large datasets like ImageNet, offer generalisable features that can be fine-tuned for specific tasks. TensorFlow excels in computational efficiency, leveraging GPU acceleration and an optimised computational graph to handle large-scale training operations (Abadi *et al.*, 2016). This makes it ideal for adapting complex architectures to the needs of custom image datasets.

TensorFlow-Keras

Keras, which serves as TensorFlow's high-level API, simplifies the process of designing and modifying neural networks. It is particularly valuable in transfer learning because it allows developers to seamlessly load and modify pre-trained models, freeze the feature extraction layers, and append new fully connected layers for classification. Keras also offers built-in utilities for data preprocessing and evaluation, streamlining tasks that would otherwise require significant manual coding (Chollet, 2015).

FastAI

FastAI is another library that can significantly enhance the transfer learning workflow. Known for its simplicity and focus on best practices, FastAI automates many aspects of model training, such as selecting optimal hyperparameters and implementing advanced training strategies like progressive resizing and discriminative learning rates. These features enable researchers to improve model performance on custom datasets (Howard and Gugger, 2020).

Scikit-learn

Scikit-learn, a versatile machine learning library, supports transfer learning in areas beyond neural network training. It provides tools for data preprocessing, such as splitting datasets into training, validation, and test sets, as well as dimensionality reduction techniques, like principal component analysis (PCA) or t-distributed stochastic neighbour embedding (t-SNE), for feature visualisation or selection. Additionally, Scikit-learn facilitates the evaluation of classification models through metrics such as precision, recall, and confusion matrices (Pedregosa *et al.*, 2011). These features make it a critical component for analysing the performance of a transfer learning model.

NumPy

NumPy underpins the computational foundation of the entire ecosystem. As a library for numerical computing, NumPy enables efficient manipulation of large arrays and matrices, which are fundamental for handling image datasets and model inputs. It supports data transformations, such as reshaping image arrays to fit specific model requirements, and serves as the computational backbone for many other libraries, including TensorFlow and OpenCV (Oliphant, 2006).

OpenCV

OpenCV specialises in image and video processing, making it useful for preparing datasets for transfer learning. With OpenCV, researchers can perform image preprocessing tasks such as resizing, cropping, and normalising images, as well as applying various data augmentation techniques. OpenCV also aids in extracting and visualising image features, enhancing the understanding of patterns and relationships in the data (Bradski, 2000). Its ability to handle real-time image processing adds further utility in dynamic applications.

Matplotlib

Finally, Matplotlib provides essential tools for data visualisation. In transfer learning, it is often used to plot training and validation curves, such as accuracy and loss over epochs, helping to monitor model performance during training. It also facilitates the visualisation of datasets, augmented images, and predictions, offering insights into how the model processes and classifies images. Furthermore, Matplotlib is commonly used to generate confusion matrices and other graphical representations that summarise evaluation metrics, enabling clear communication of results (Hunter, 2007).

Together, these libraries form a cohesive toolkit for transfer learning in image classification. TensorFlow and Keras handle the core deep learning tasks, while FastAI optimises and simplifies training workflows. Scikit-learn and NumPy provide essential support for preprocessing and numerical operations, and OpenCV ensures efficient image-specific preprocessing and manipulation. Matplotlib facilitates monitoring and detailed visualisation of training processes and results. Each library works together to ensure that the transfer learning pipeline is both efficient and effective.

3.2.2 Data preparation

The dataset for this study, as discussed in-depth in Chapter 2, has been created using high-resolution core images scanned to TIFF format from the upper to distal carbonate slope transects drilled during ODP Leg 133 (Davies *et al.*, 1991) and ODP Leg 194 (Isern *et al.*, 2002) in the carbonate platforms and troughs of north-eastern Australia, and the platform interior carbonates drilled during IODP Expedition 359 in the Maldives archipelago (Betzler *et al.*, 2017). A total of 588 cores were used, with a total thickness of approximately 391.15m, comprising 84.58m from ODP Leg 133, 133.22m from ODP Leg 194, and

173.35m from IODP Expedition 359 (Table 3.2). A full list of cores used is provided in Appendix A (Table A.1). High-resolution imaging of core sections has been routine since ODP Leg 198, and the digital core scanners used during ocean drilling expeditions currently capture approximately 10-20 pixels per millimetre of core (Wilkens *et al.*, 2009). For the cores drilled during ODP Legs 133 and 194, the scans were performed by C.M. John in 2007 using a prototype line scanner at the Gulf Coast Repository, which is now onboard the JOIDES Resolution.

The digital core images are colour images using the RGB colour model, meaning they are composed of three colour channels: red (R), green (G), and blue (B). These colour channels are each represented as a separate matrix, and the combination of these matrices forms a tensor encoding the image. Digital core scan images of a 1.5 metre section are approximately 31,600 pixels x 1,750 pixels; however, only the sediment core section itself is required here, which is on the order of 29,900 pixels x 1,375 pixels within the image. These individual images can provide some of the highest resolution examples of core properties available at present.

Table 3.2: Summary of the number of images and total core length (in metres) used for each Dunham class in the image classification task.

	ODP Leg 133		ODP Leg 194		IODP Expedition 359		Total Images	Total Core (m)
	No. of Images	Core Length (m)	No. of Images	Core Length (m)	No. of Images	Core Length (m)		
boundstone	1,563	5.86	1,108	4.16	918	3.44	3,589	13.46
floatstone	1,062	3.98	7,415	27.81	1,206	4.52	9,683	36.31
grainstone	2,066	7.75	8,372	31.40	6,051	22.69	16,489	61.83
mudstone	9,640	36.15	5,779	21.67	3,555	13.33	18,974	71.15
packstone	2,570	9.64	1,528	5.73	19,058	71.47	23,156	86.84
rudstone	3,021	11.33	4,377	16.41	4,939	18.52	12,337	46.26
wackestone	2,632	9.87	6,946	26.05	10,500	39.38	20,078	75.29
							104,306	**391.15**

Image resizing

Many deep learning algorithms, including CNNs, require all images in a dataset to be resized to a unified dimension. To ensure compatibility with the original dimensions of the pre-trained architectures and to leverage the natural-image features learned by these networks, the core images used in this work were cropped to 224 x 224 pixels or 299 × 299 pixels, according to the network input size. This was performed using the slicing tool in Adobe Photoshop 2020 (version 21) (Adobe Inc., 2020). Downscaling images can cause loss of resolution, so each core image was sliced into multiple smaller images to preserve the original features of the carbonate textures. These images were then saved to JPEG format for compatibility with TensorFlow. Digital core scan images of a 1.5 metre section are approximately 31,600 pixels x 1,750 pixels; however, only about 29,900 pixels x 1,375 pixels within this image represent the core itself. From each full core scan, a maximum of 400 smaller images (e.g., 299 x 299 pixels) can be produced, depending on core recovery. The size of the datasets referred to in this work corresponds to the number of these smaller images derived from the core scans.

Following the core descriptions provided by the original expedition scientists as a base for the classifications, each core was analysed for the identifiable textural classes, and any boundaries were marked to delineate areas of uniform texture. The scan was then sliced according to each section of uniform texture within the core image. This approach minimised the risk sampling multiple textures within a single image, which could otherwise introduce noise into the dataset. However, it should be acknowledged that this methodology involves subjective judgement during the classification of textures, as it relies on the interpretations of the original expedition scientists and the lead author of this study. These subjective inputs may introduce bias into the dataset, which could propagate through the CNN structure and influence model performance. To mitigate this potential bias, classifications from multiple experts have been included where possible.

To ensure the networks were extracting and learning features from the cores as opposed to other articles, such as Styrofoam inserts and core liners, all images were manually inspected and any images containing less than 70% core after slicing were removed from the dataset. This created a dataset containing 104,306 images across seven classes from the modified Dunham Classification (Dunham, 1962; Embry and Klovan, 1971): mudstone, packstone, wackestone, grainstone, floatstone, rudstone, and boundstone, as defined in Section 3.1.1 (Table 3.2; Figures 3.3 and 3.4).

To the best of our knowledge, this is the largest carbonate core image dataset used for image classification to date. In geoscience applications, however, there are many instances where only limited training data may be available. Therefore, in this study, three datasets with orders of magnitude difference in data quantity were created, to allow further comparison of model performance across datasets of different sizes: the original 104k dataset, a 42k dataset, and a 7k dataset (Figure 3.5).

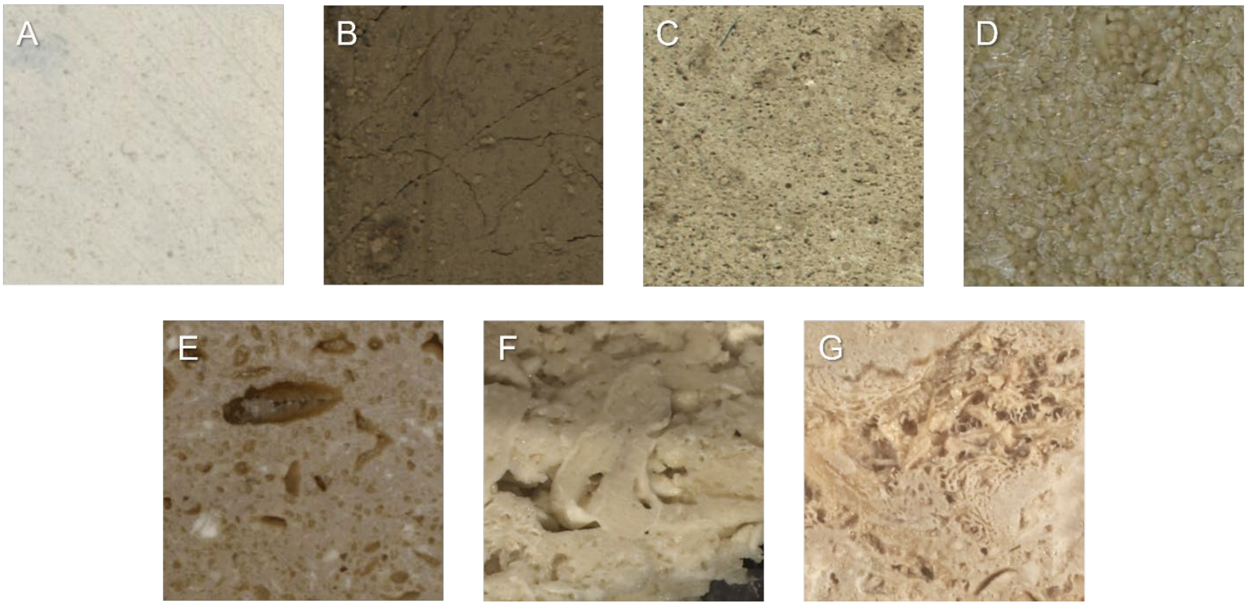

Figure 3.4: Sample dataset images, selected through random sampling: (a) mudstone, (b) wackestone, (c) packstone, (d) grainstone, (e) floatstone, (f) rudstone, (g) boundstone.

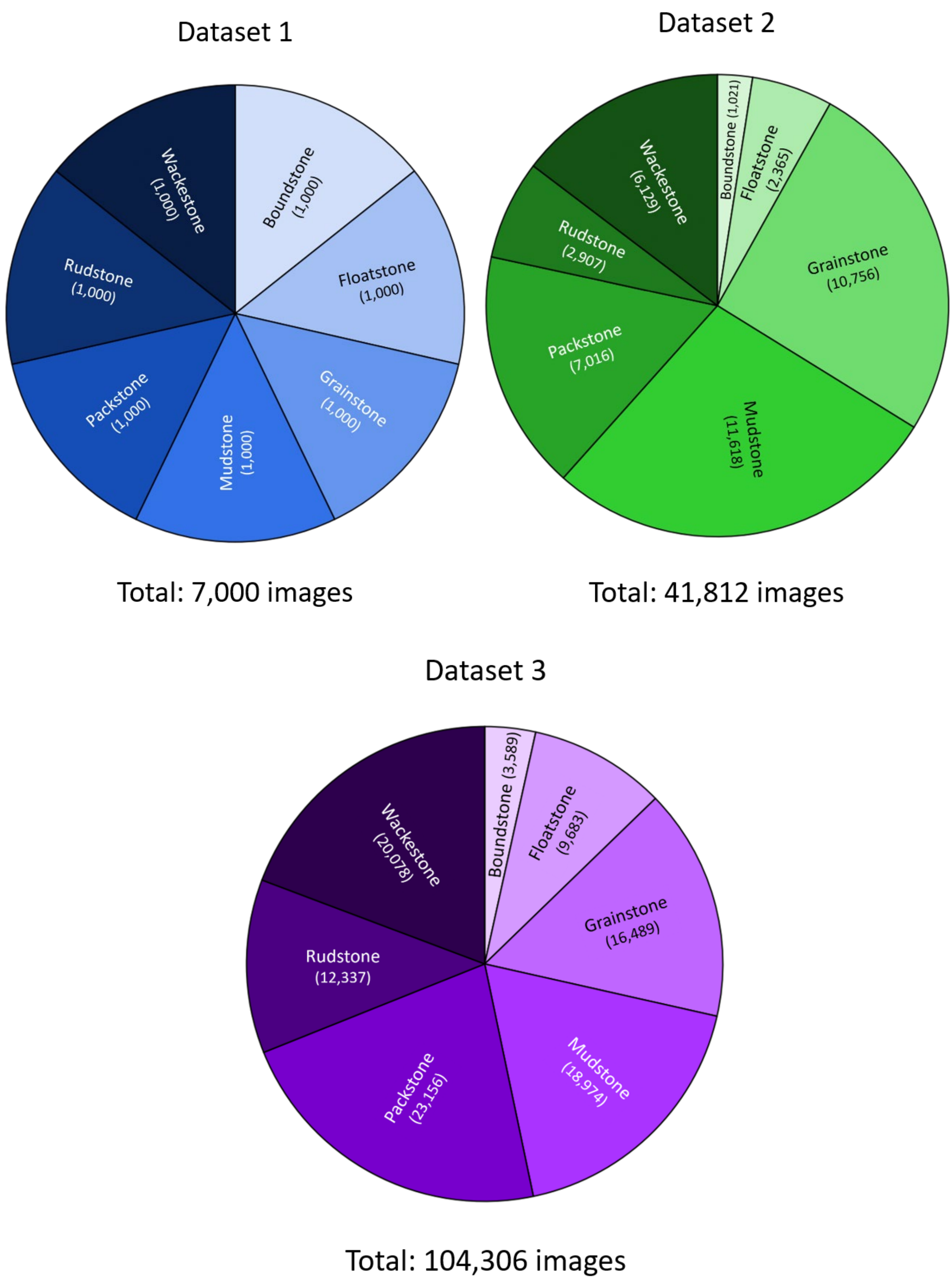

Figure 3.5: Distribution of core images for each class across the three datasets.

Dataset splits

Prediction reliability is one of the main concerns in the performance evaluation of supervised deep learning algorithms (Consonni *et al.*, 2010; Alsina *et al.*, 2017). In this study, each dataset was split into 80% training data and 20% test data. We further used the training data with cross validation to create 5 folds of 80% training data and 20% validation data (i.e., validation set = 16% of the total dataset, Figure 3.6). Since both the 48k and 104k datasets have a significant class imbalance, we used stratified k-fold cross-validation (k=5), where the percentage of samples for each class is maintained in every fold, ensuring the class distribution in each fold matched the distribution in the complete training dataset.

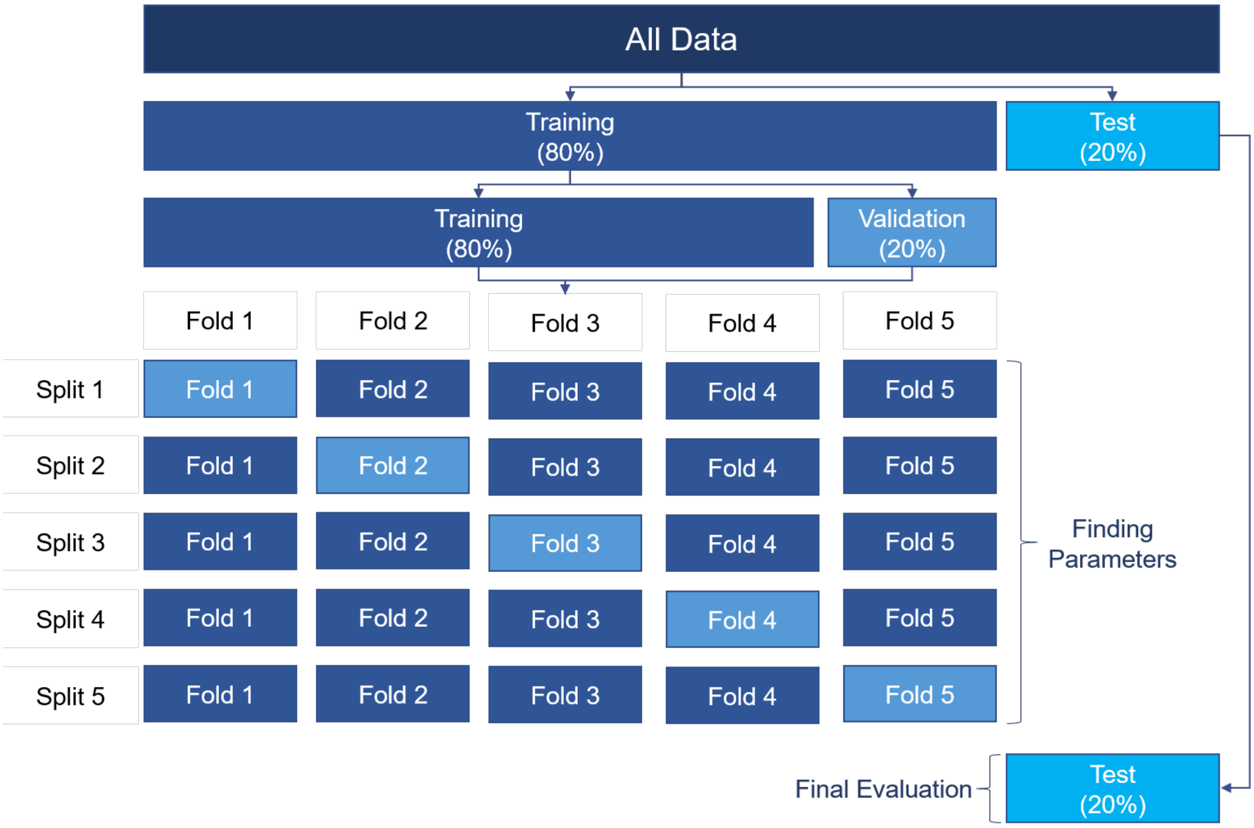

Figure 3.6: Stratified k-fold cross validation (k = 5) used in this study, where each dataset is divided into training, validation and test data.

3.2.3 Model training

To classify carbonate core images according to the modified Dunham Classification (Figure 3.3; Dunham, 1962; Embry and Klovan, 1971), we use the learned parameters of CNN architectures pre-trained on the ImageNet dataset. ImageNet, first developed in 2009, is a large-scale hierarchical image database designed for general image classification tasks (Deng et al., 2009). It provides a robust starting point for training deep learning models, as its dataset contains over 14 million labelled images spanning thousands of classes. While the pre-trained weights from ImageNet are not directly tailored to classify carbonate core images, they serve as a foundation for feature extraction.

The general image classification capabilities of the ImageNet-trained models allow them to recognise patterns such as edges, textures, and shapes that are present in a wide variety of images. In our study, these capabilities were leveraged by taking the pre-trained models and modifying them to suit the specific task of carbonate core classification. Specifically, the final layer of each CNN, responsible for performing classification, was replaced with a new classifier designed to classify images based on textural observations from carbonate cores (Figure 3.7). This approach combines the robust feature extraction capabilities of the pre-trained CNNs with the task-specific requirements of our application, thereby enabling the networks to classify core images effectively.

Nine different CNNs representing four different architectures—DenseNet (Huang et al., 2017a), Inception-v3 (Szegedy et al., 2016), ResNet (He et al., 2016), and VGG (Simonyan and Zisserman, 2014)—were trained on our carbonate core datasets (Table 3.3). These architectures were selected for their established performance in image classification tasks and their varying approaches to learning features from data. For example, ResNet employs skip connections to mitigate the vanishing gradient problem in very deep networks (He et al., 2016), while DenseNet uses dense connections to improve feature propagation (Huang et al., 2017a). Inception-v3 optimises computational efficiency through mixed convolutions (Szegedy et al., 2016), and VGG employs a simpler but highly effective architecture of stacked convolutional layers (Simonyan and Zisserman, 2014). By testing these different architectures, we aim to determine which is best suited for the classification of carbonate cores. Full details of these architectures can be found in their original references.

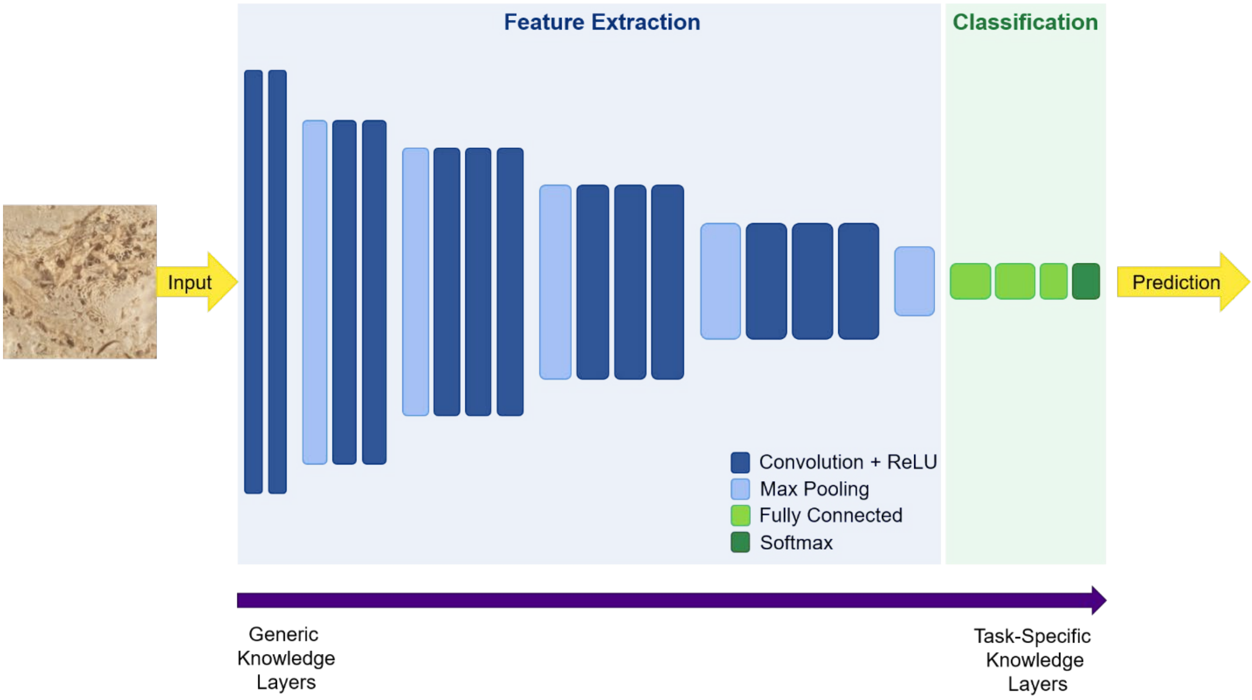

Figure 3.7: Schematic representation of the CNN parts used in transfer learning (based on the VGG16 architecture). The feature extraction part of a CNN (shown in blue) is composed of convolutional and pooling layers and is used as the convolutional base. The main purpose of the convolutional base is to generate features from images, such as edges, lines and curves in the lower layers and shapes in the upper layers. The classification part of a CNN (shown in green) is usually composed of fully connected and Softmax layers. The classifier predicts the class of the input image based on task-specific features.

Table 3.3: Summary of models used in this comparison.

Model	Size	Parameters	Depth	Image Input Size
DenseNet121	33 MB	8,062,504	121	224 x 224
DenseNet169	57 MB	14,307,880	169	224 x 224
DenseNet201	80 MB	20,242,984	201	224 x 224
Inception-v3	92 MB	23,851,784	48	299 x 299
ResNet50	98 MB	25,636,712	50	224 x 224
ResNet101	171 MB	44,707,176	101	224 x 224
ResNet152	232 MB	60,419,944	152	224 x 224
VGG16	528 MB	138,357,544	16	224 x 224
VGG19	549 MB	143,667,240	19	224 x 224

The new classification head was implemented with the following architecture: an average pooling layer; a fully connected layer with 1,024 hidden units and rectified linear unit (ReLU) activation; a dropout rate of 0.2; and a final fully connected Softmax layer. These specific parameter values were chosen based on a combination of established best practices in deep learning literature and preliminary experimentation to optimise performance for the carbonate core classification task. The average pooling layer reduces spatial dimensions while retaining important features, the fully connected layer with ReLU activation introduces non-linearity to model complex patterns, and the dropout rate mitigates overfitting by randomly disabling a fraction of neurons during training. The Softmax layer ensures the output is normalised into probabilities across the classification categories.

Using the fine-tuning method, each CNN was trained for a total of 25 epochs. In this context, an epoch refers to one complete pass through the entire training dataset during the learning process. During the initial 15 epochs, only the classification head was trained while the rest of the pre-trained network weights were frozen, meaning they remained unchanged.. Subsequently, the uppermost layers of the networks were unfrozen, allowing the models to be trained as a whole for the remaining 10 epochs. This fine-tuning strategy was adopted to gradually adapt the learned higher-order feature representations of the pre-trained networks to the specific task of carbonate core classification.

This approach is informed by both theoretical understanding and prior research on transfer learning, which demonstrates that freezing the base layers of a pre-trained model can help to initially prevent overfitting on small datasets and allow the classification head to specialise. Unfreezing and training the upper layers later enables the model to refine the learned features for the new task (Yosinski et al., 2014; Pan and Yang, 2010). These methods are widely used in computer vision applications and were selected here based on their proven efficacy in similar contexts.

Model training in deep learning involves minimising the difference between the predicted labels and the true labels of the training dataset. This is achieved through iterative adjustments to the weights and biases within the network layers—specifically, the kernels of the convolutional layers and the fully connected layers (Yamashita *et al.*, 2018). During this process, network performance is continuously evaluated using a cost function, which computes the difference between the predicted output and the actual labels. This evaluation occurs through forward propagation, where the input data is passed through the network to generate predictions (Richard and Lippmann, 1991; Yamashita *et al.*, 2018).

In the case of multi-class classification tasks, a common approach is to use cross-entropy loss as the cost function (Ho and Wookey, 2019). Cross-entropy measures the difference between the true probability distribution (i.e., the true class labels) and the predicted probability distribution for each sample. It quantifies how well the model output probability distribution matches the true distribution. To produce the final class prediction for each sample, the Softmax function is applied to the output of the cross-entropy loss. The

Softmax function converts the raw output scores of the model into probabilities by exponentiating each score and normalising them, ensuring that the sum of all probabilities equals one. This allows for a probabilistic interpretation, where the class with the highest probability is selected as the predicted class.

The learnable parameters of the network (i.e., the weights and biases) are updated iteratively according to the loss value through backpropagation and optimisation algorithms, such as gradient descent This process adjusts the parameters in order to minimise the loss and improve model accuracy over time. In this study, we evaluated the performance of two commonly used gradient descent algorithms: stochastic gradient descent (SGD) with momentum (Qian, 1999) and Adam (Kingma and Ba, 2014).

Using the one-cycle policy to improve training, reduce overfitting and allow the network to converge faster (Smith, 2017; 2018), the optimal learning rate was determined to be between 1e - 2 and 1e - 3 for the first 15 epochs and between 1e - 5 and 1e - 6 for the rest of the training. The one-cycle policy is a learning rate scheduling technique that adjusts the learning rate dynamically during training. Initially, the learning rate starts at a very low value, then increases linearly to a maximum value over a portion of the training process. After reaching the peak value, the learning rate decreases back to a lower value for the rest of the training. This approach has been shown to improve training speed and model performance by allowing the model to learn effectively at different stages of the training process (Smith, 2017; 2018).

The learning rate is a hyperparameter that controls the step size at each iteration while moving towards a minimum of the loss function. In this study, it was measured in terms of the rate of change of the weights during training, with the values expressed in powers of 10 (i.e., 1e-2, 1e-3, etc.), where these are typical representations for small learning rates in scientific notation (e.g., 1e-2 is 0.01). The units are dimensionless since they correspond to the magnitude of the update to the model weights.

Batch size refers to the number of training examples used in one forward/backward pass of the network. In this study, the batch size was held constant at 32 for all models. This choice helps balance computational efficiency and model generalization. Smaller batch sizes can lead to noisy gradient estimates, while larger batch sizes can require more memory and may lead to slower convergence. A batch size of 32 has been found to be effective in many deep learning applications, providing a good trade-off between training time and performance. Training was repeated for a total of five times and predictions were averaged, since CNN performance can exhibit random minor fluctuations.

3.2.4 Evaluation

During training, the performance of the networks was evaluated by monitoring the changes in training and validation losses with each epoch. Following the completion of training, evaluation metrics (accuracy, precision and recall) and the confusion matrices (e.g. Ruuska *et al.*, 2018; Skansi, 2018) were used to compare the prediction performances of the different CNNs on the image classes of the carbonate core test dataset. Accuracy measures the ratio between the number of correctly predicted classifications and the total number of samples classified:

$$\text{Overall Accuracy: ACC} = \frac{\text{TN} + \text{TP}}{\text{TN} + \text{TP} + \text{FN} + \text{FP}} \tag{3.1}$$

whereby TN is the number of true negative cases, TP is the number of true positive cases, FN is the number of false negatives, and FP is the number of false positives. Since overall accuracy makes no distinction between classes, with correctly predicted classifications for each class being treated equally, this could be viewed as a misleading metric. This is particularly true in the case of significant class imbalances, where classes containing more samples will dominate the statistic. For this reason, we use multiclass averaging to report the mean per class accuracy (MPCA), given as the unweighted mean of the sum of the accuracies for each independent class:

$$\text{Mean Per Class Accuracy: MPCA} = \frac{\sum \text{ACC of Each Class}}{\text{Number of Classes}} \tag{3.2}$$

Precision calculates the proportion of positive identifications that were correct i.e., the probability that the predicted class of the carbonate facies actually belongs to that class:

$$\text{Precision: P} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3.3}$$

Recall calculates the proportion of actual positives that were correctly identified:

$$\text{Recall: R} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3.4}$$

We also consider how computational cost impacts classification performance by considering the model complexity, computational complexity, training time and classification speed of each architecture. By analysing the relationships between these metrics, we can further develop our understanding of the best practices for applying these models to geological images. Following Becker *et al.* (2021), model complexity is defined as the number of trainable parameters of the neural network. Computational complexity is defined in terms of floating-point operations (FLOPs) required by the model for a single forward pass.

3.3 Results

In most cases, for Dataset 42k and Dataset 104k, the best results were achieved using the Adam optimizer. For Dataset 7k, the SGD with momentum produced markedly better results for all but one of the models (Table 3.4).

3.2.1 Performance evaluation for overall accuracy

On the large dataset (Dataset 104k), within the individual architectures, the deeper CNNs generally achieved a higher MPCA than shallower networks (Table 3.4, Figure 3.8). When considering the MPCA against the computational complexity of the architectures, within the VGG and DenseNet architectures, we clearly see an increase in MPCA with increasing depth and complexity. However, increased computational complexity does not always translate directly into a proportional increase in classification performance. A weak negative correlation between the MPCA and the FLOPs is present with an R2 value of 0.635. This is corroborated by the highest accuracy (91.59%) being obtained by Inception-v3, which shows comparatively lower model and computational complexities (Figure 3.8). The results observed for the larger dataset are similar to those for Dataset 42k (Figure 3.9), where Inception-v3 again achieves the highest accuracy (89.61%).

For the small dataset (Dataset 7k), we see a clear increase in MPCA with increasing computational complexity for a single forward pass (Figure 3.10). This is evidenced by the highest accuracy (85.11%) being achieved by VGG19, the most computationally complex architecture evaluated (20 GFLOPs). Those architectures with lower complexity achieve a lower MPCA e.g., DenseNet201 requires only 4 GFLOPs, but obtains the lowest MPCA (66.70%).

As is expected, within the individual architectures, we see the larger and more complex networks requiring a greater training time. For Dataset 7k, there is clear trade-off existing between faster training and higher recognition performance, with VGG19 achieving 85.11% accuracy with the longest training time of 219 minutes (Figure 3.11). By contrast, ResNet50 produces the second lowest accuracy of 70.16% after the shortest training time of 56 minutes. Despite the trend seen in the smaller dataset, for Datasets 42k and 104k, the highest MPCAs are achieved by Inception-v3, 89.61% and 91.59% respectively, whilst requiring a middle-of-the-range training time, 146 minutes and 194 minutes respectively (Figures 3.12 and 3.13). After this point, there is a general decrease in MPCA, suggesting these models may be training for too long and, as a result, are starting to overfit the training data.

Table 3.4: Performance evaluation of classification accuracy during the prediction stage. The results for both optimizers under each dataset are shown for comparison. The numbers given in bold represent the best MPCA for each dataset when compared to other classification models.

	Mean Per Class Accuracy (MPCA)					
	Dataset 7k		**Dataset 42k**		**Dataset 104k**	
Model	**SGDM**	**Adam**	**SGDM**	**Adam**	**SGDM**	**Adam**
DenseNet121	74.48	65.64	83.24	85.16	89.29	90.10
DenseNet169	77.37	67.12	84.47	88.94	**91.22**	90.76
DenseNet201	66.70	67.23	82.06	87.77	90.74	90.91
Inception-v3	73.12	65.99	**89.13**	**89.61**	91.14	**91.59**
ResNet50	70.16	61.73	86.05	86.78	88.46	89.49
ResNet101	71.92	63.18	86.53	87.09	89.07	90.03
ResNet152	73.44	65.52	87.79	87.73	89.62	88.91
VGG16	81.56	70.06	78.36	79.62	85.38	86.95
VGG19	**85.11**	**71.31**	81.17	81.63	87.93	88.26

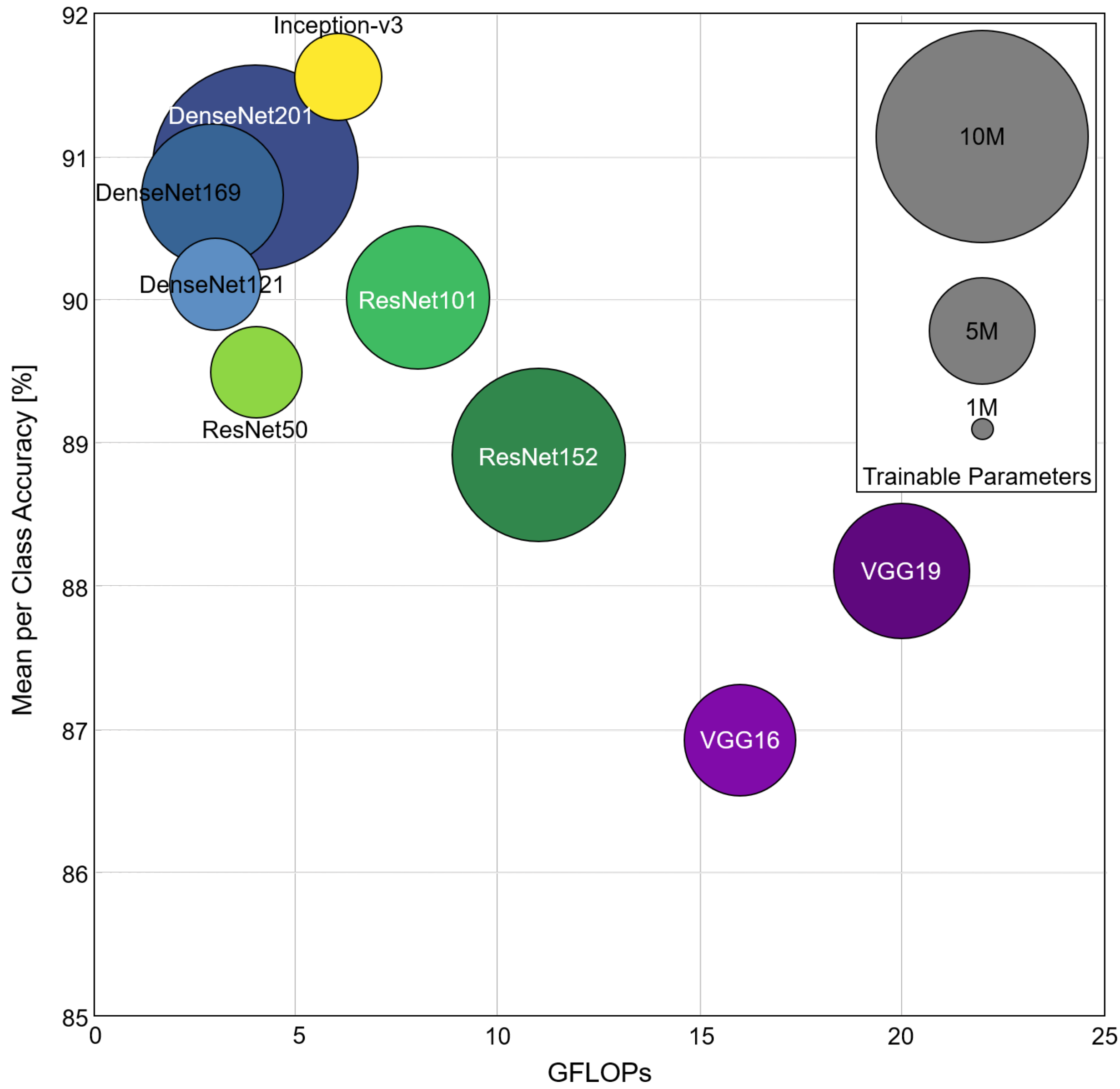

Figure 3.8: Average accuracy vs computational complexity vs model complexity for Dataset 104k. The CNN architectures are compared based on classification performance (MPCA on the y-axis), computational complexity (GFLOPs on the x-axis), and model complexity (the number of trainable parameters, proportional to circle size).

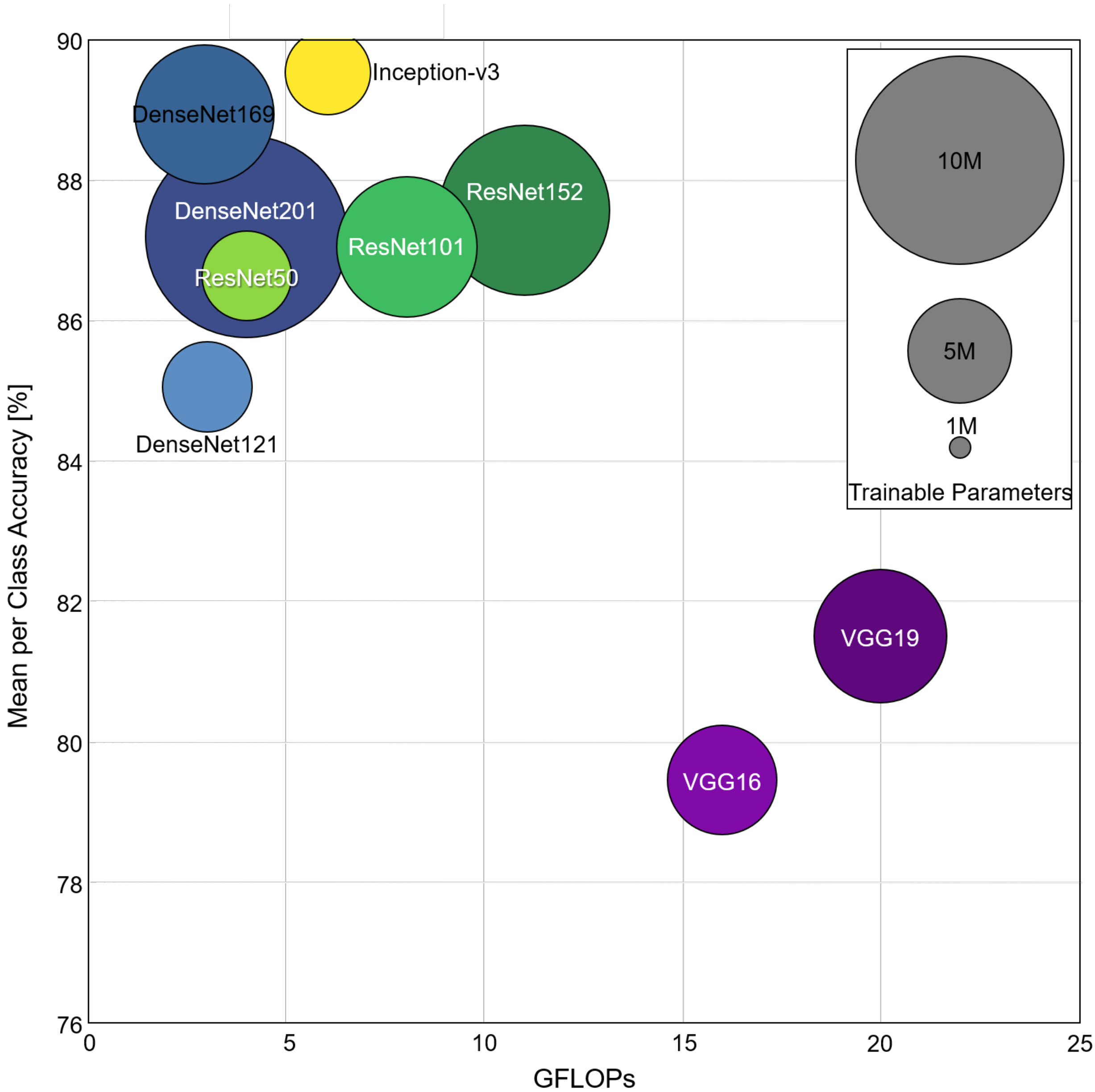

Figure 3.9: Average accuracy vs computational complexity vs model complexity for Dataset 42k. The CNN architectures are compared based on classification performance (MPCA on the y-axis), computational complexity (GFLOPs on the x-axis), and model complexity (the number of trainable parameters, proportional to circle size).

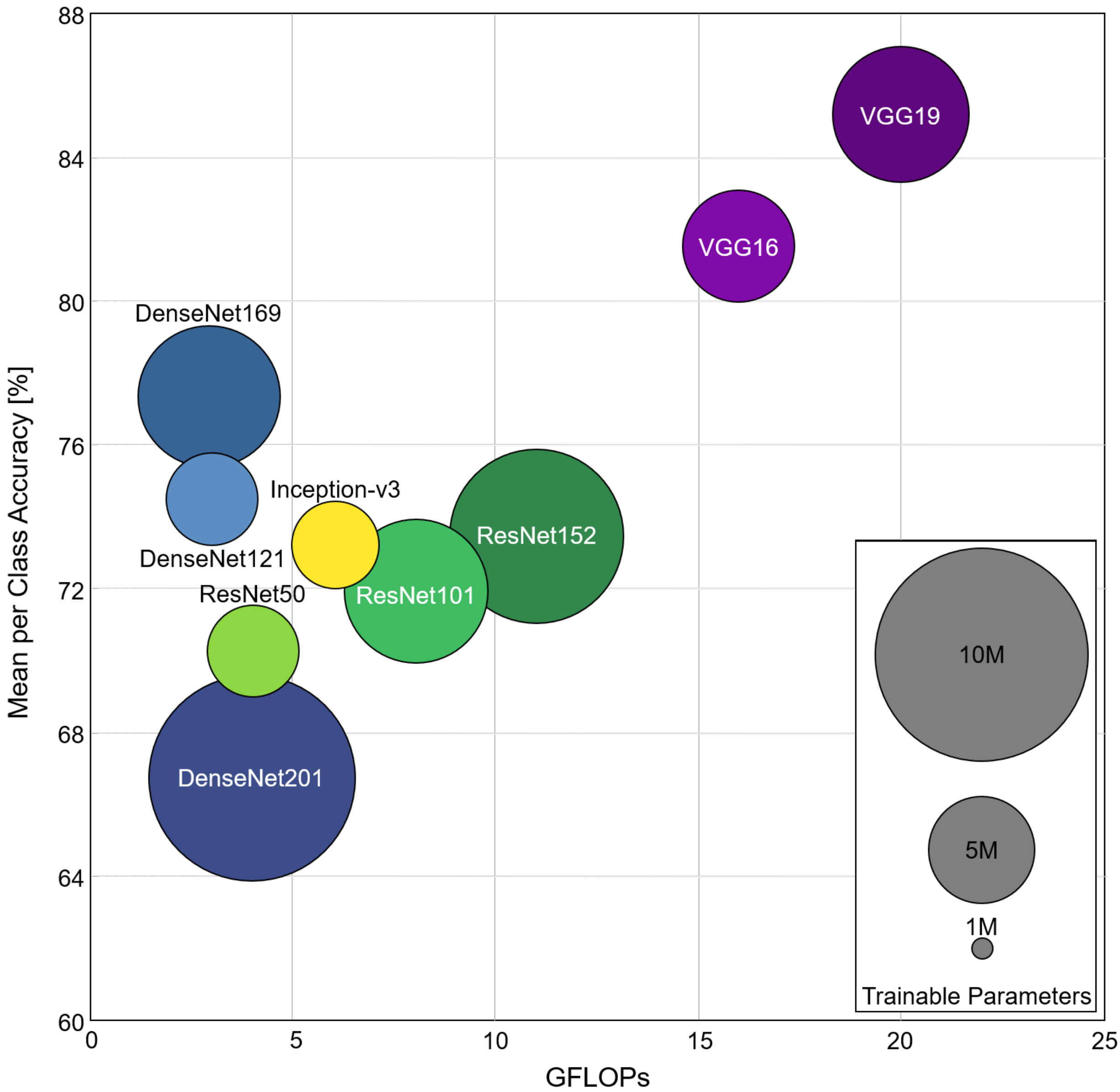

Figure 3.10: Average accuracy vs computational complexity vs model complexity for Dataset 7k. The CNN architectures are compared based on classification performance (MPCA on the y-axis), computational complexity (GFLOPs on the x-axis), and model complexity (the number of trainable parameters, proportional to circle size).

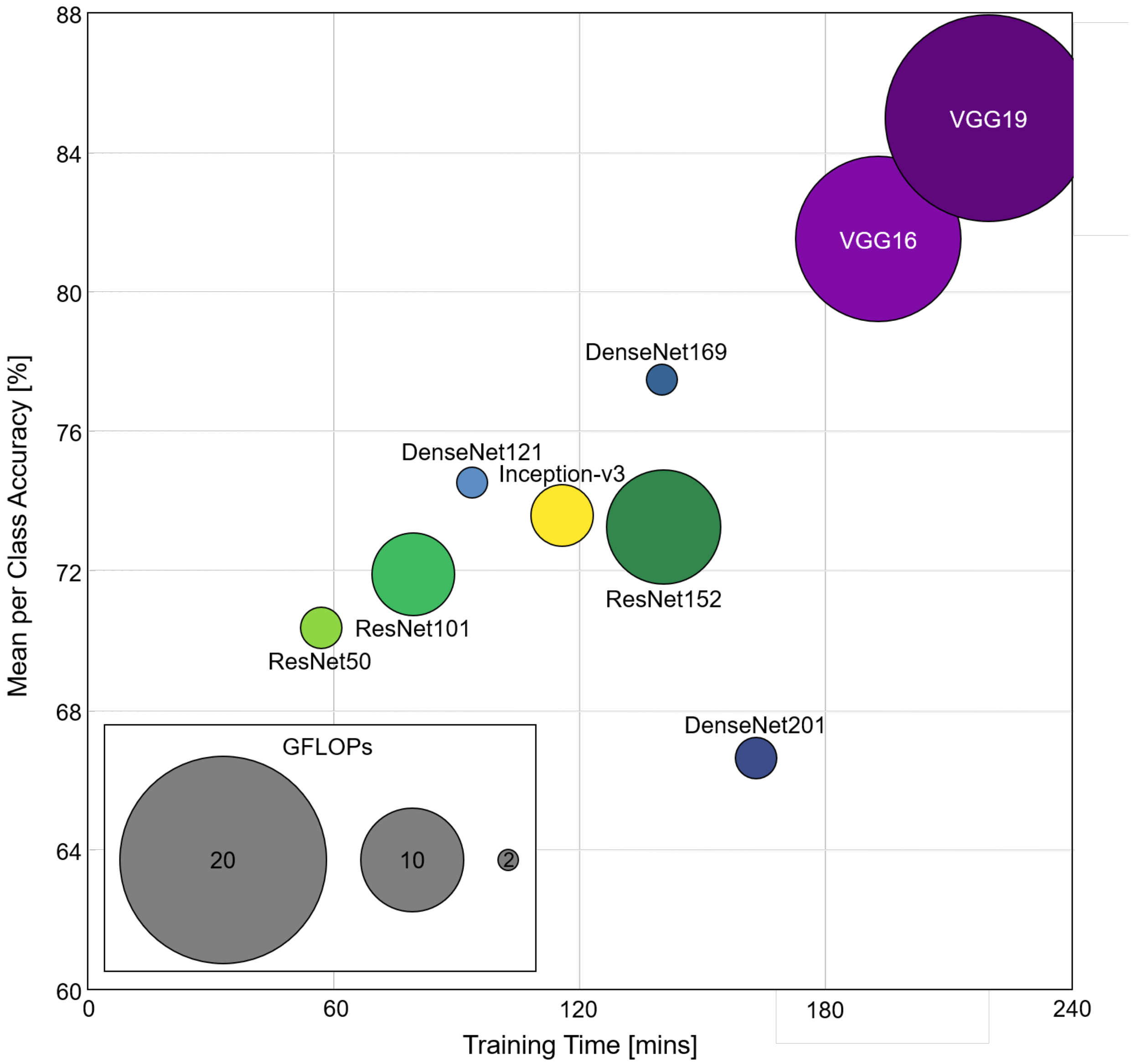

Figure 3.11: Average accuracy vs training time vs computational complexity for Dataset 7k. The CNN architectures are compared based on classification performance (MPCA on the y-axis), training time (given in minutes on the x-axis), and computational complexity (GFLOPs, proportional to circle size).

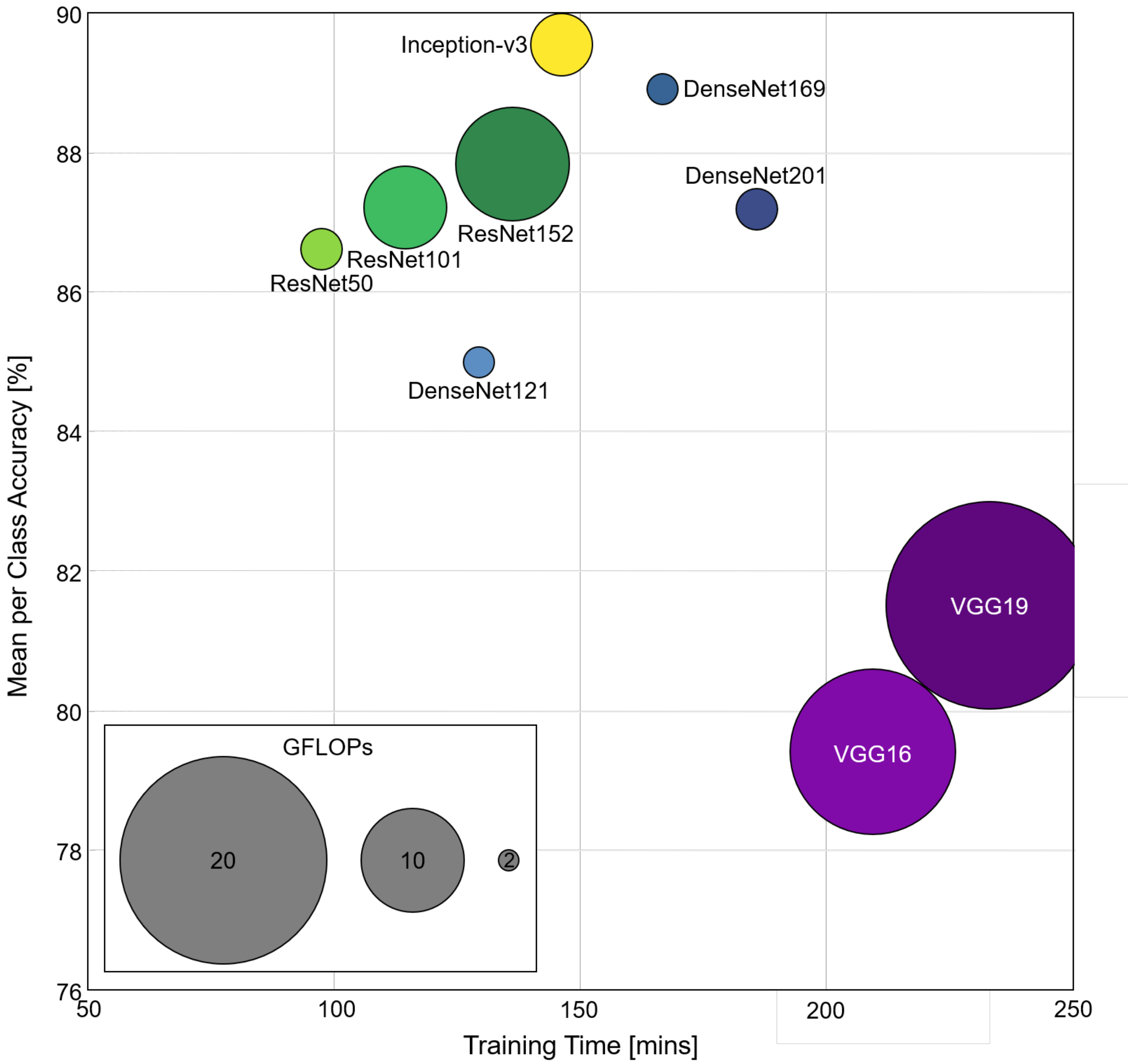

Figure 3.12: Average accuracy vs training time vs computational complexity for Dataset 42k. The CNN architectures are compared based on classification performance (MPCA on the y-axis), training time (given in minutes on the x-axis), and computational complexity (GFLOPs, proportional to circle size).

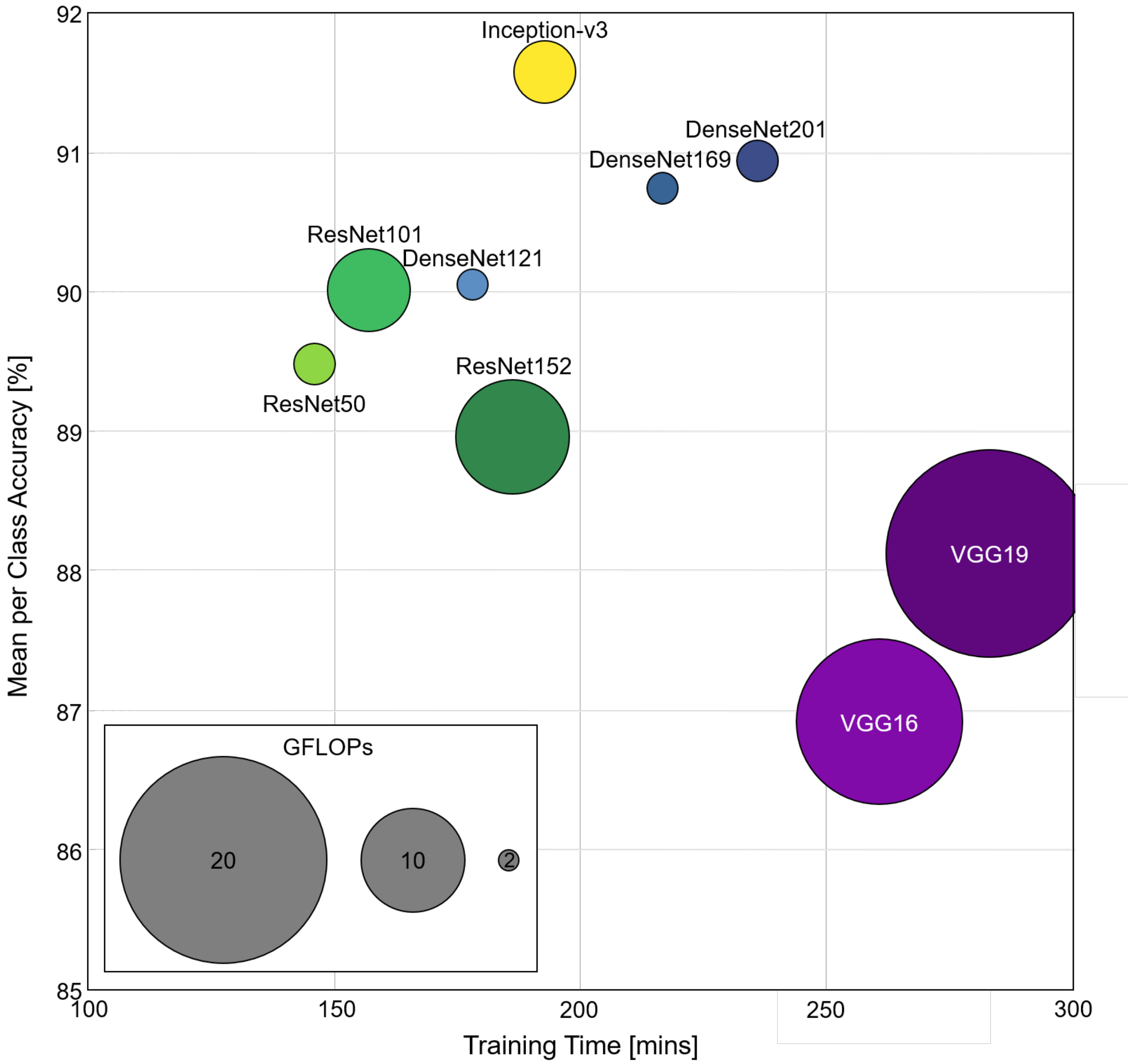

Figure 3.13: Average accuracy vs training time vs computational complexity for Dataset 104k. The CNN architectures are compared based on classification performance (MPCA on the y-axis), training time (given in minutes on the x-axis), and computational complexity (GFLOPs, proportional to circle size).

3.2.2 Performance evaluation for each class

We observe the top performing classifier (Inception-v3, Dataset104k) predicts all of the classes (Dunham facies) with recall values above 0.86. The maximum precision observed was in the grainstone class (0.96) for DenseNet201 trained on Dataset 104k, and the best recall was in the mudstone class (0.97) for Inception-v3 trained on Dataset 104k (Figures 3.14 and 3.15). The grainstone and mudstone classes represent opposite ends of the ≤2mm fraction of the Dunham classification scale. Grainstone is the grain-supported texture that lacks mud, while mudstone is the finest mud-supported texture, containing less than 10% grains (Figure 3.3). These distinct differences in texture and grain content of these two classes may make them easier to differentiate from the other classes, leading to the observed high precision for grainstone and high recall for mudstone.

Across all architectures and datasets, the lowest precision is observed in the floatstone class (0.62) (Figure 3.14), indicating that the model frequently misclassified other classes as floatstone, leading to a higher number of false positives. This could be due to the floatstone class being visually or texturally similar to other classes, or insufficient representation in the dataset, making it harder for the model to distinguish correctly. Similarly, wackestone had the lowest recall (0.61) across all architectures (Figure 3.15). This suggests that the model had difficulty identifying all true instances of wackestone, likely due to its position in the middle of the scale, where borderline cases may share features with both mudstone and packstone. Floatstone and wackestone are both intermediate classes that are mud or matrix supported and contain more than 10% grains, with floatstone being the coarser (≥2mm) grained equivalent of wackestone. These intermediate classes are likely more prone to misclassification because their textures may overlap with adjacent classes, making it harder for the model to distinguish them accurately.

The highest precision and recall values are achieved in five out of the seven classes by Inception-v3. It should be noted, that for the small dataset, since each class has only 1,000 images, misclassifications are likely to have more of an impact on measured performance than those occurring in Dataset 42k and 104k, due to the reduced number of samples available for training and validation. This can lead to higher variance in performance metrics.

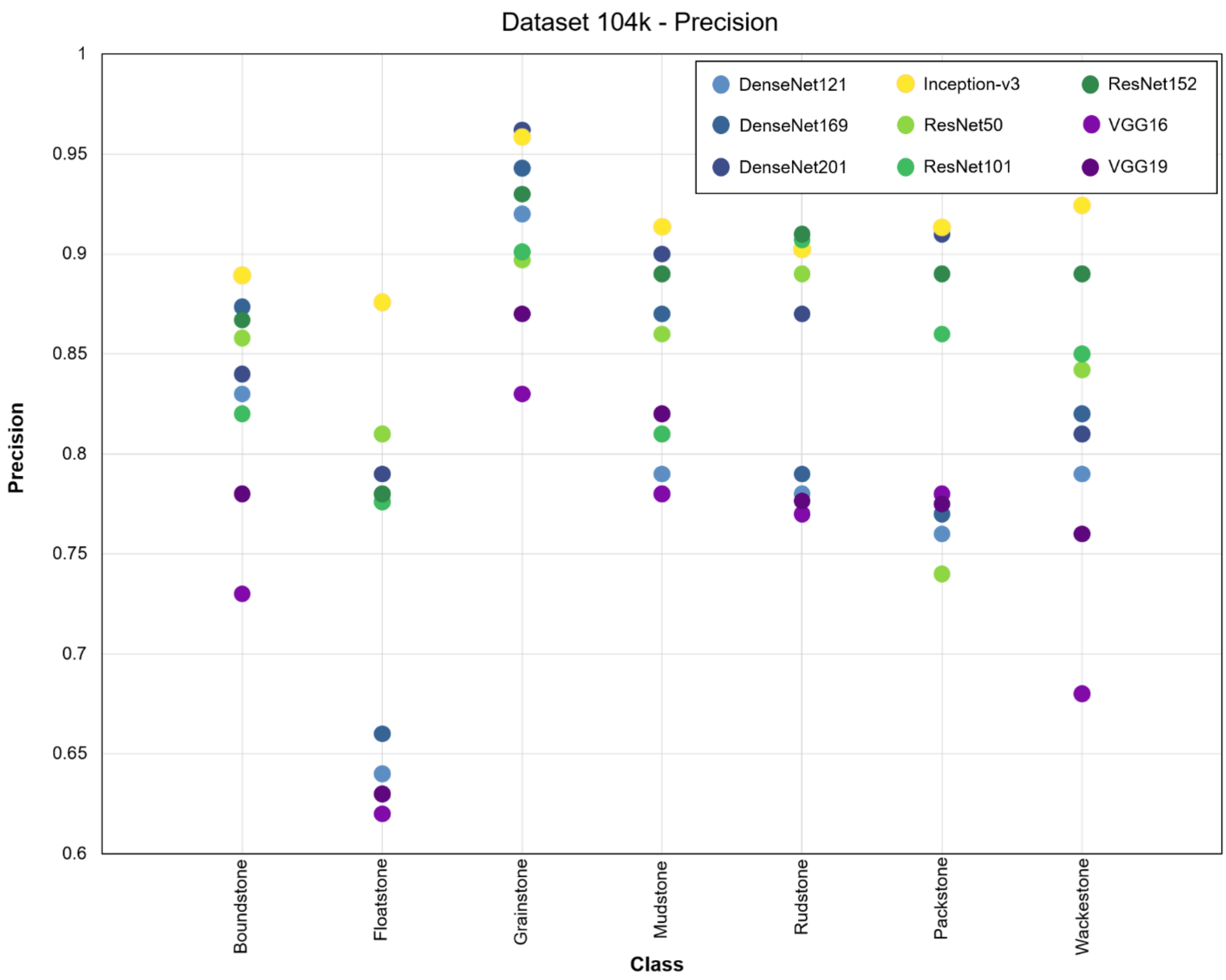

Figure 3.14: Precision graph for Dataset 104k.

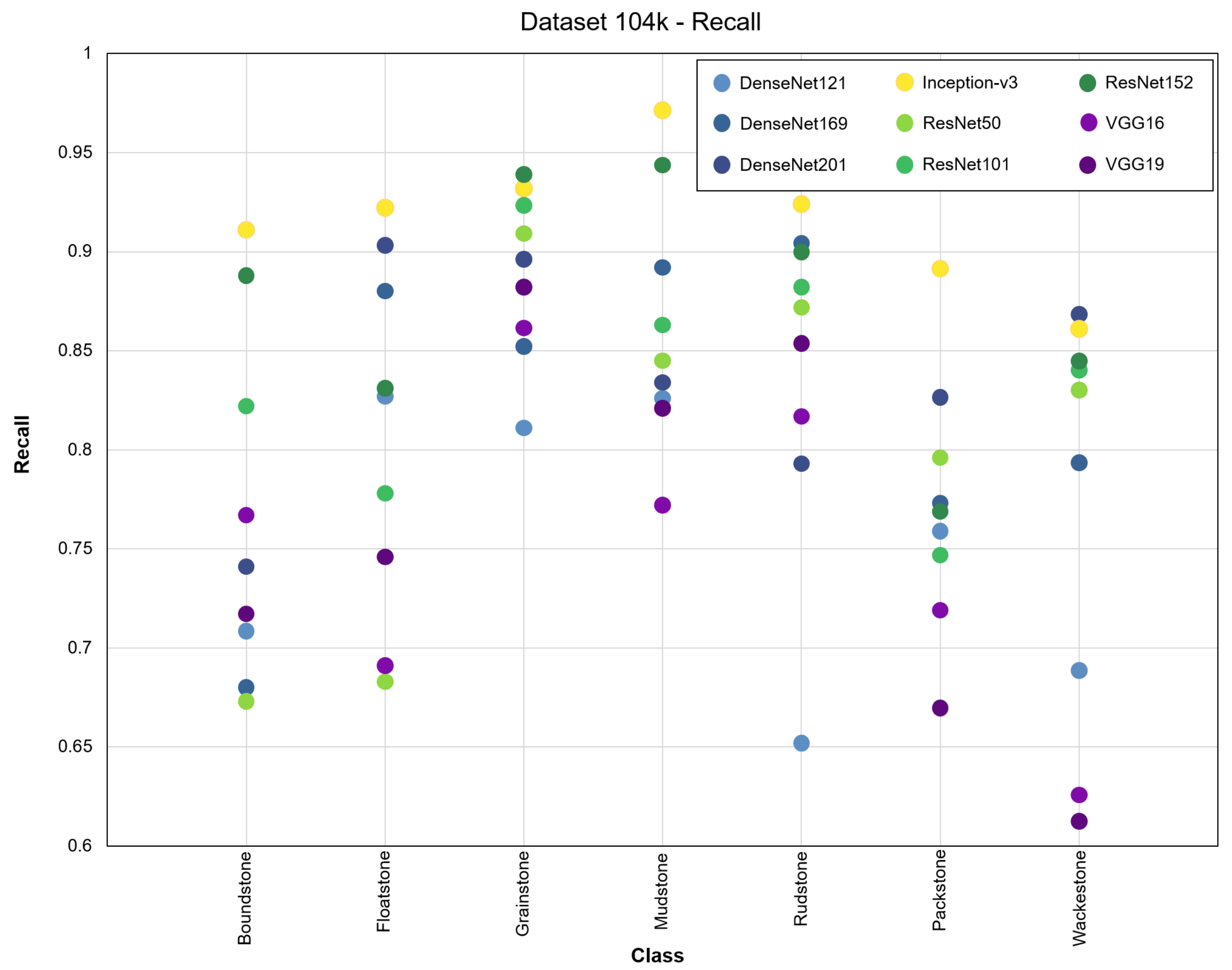

Figure 3.15: Recall graph for Dataset 104k.

3.4 Discussion

3.4.1 Comparison with human performance and best model selection

The most successful model (Inception-v3, Dataset 104k) achieved an accuracy of nearly 92% (Table 3.4), also outperforming the other architectures in terms of precision, recall and computing time and resources. This significantly exceeds results achieved in previous carbonate classification studies, namely the 79% accuracy of carbonates facies prediction using random forest from physical properties (Insua *et al.*, 2014) and the 89% accuracy achieved in the pilot neural network study (John and Kanagandran, 2019). While these results indicate that deep CNNs can provide a significant improvement on existing classifications, achieving a similar or better accuracy than experienced geologists (Lokier and Al Junaibi, 2016), the transferability of these results to other contexts remains uncertain. The performance observed here is specific to the dataset and the classification task at hand, which focuses on textural carbonate rock facies. Factors such as dataset size, diversity, and the specific characteristics of the rock types could influence how well these models perform on different datasets or geological contexts. Nevertheless, the results suggest that deep CNNs, with proper training and fine-tuning, have the potential to significantly outperform traditional methods in carbonate classification and may offer improvements in other similar geological classification tasks, especially with the observed 250x gain in description speed (Figure 3.16).

The average classification speed of a person on the ImageNet dataset is estimated to be around 50 images per minute, which equates to $0.8\dot{3}$ images per second (Markoff, 2012). It should be noted that the classification task from which this estimate was achieved, whilst relatively complex due to the large number of possible categories, is based on non-specialist images, and as such this estimate should be considered an upper bounding limit of the speed a person might be able to classify geological images. In comparison, an automated classification system based on the top performing model (Inception-v3, Dataset 104k) has a classification speed of 1,179 images per second, calculated by inverting the inference time in which the test set (20,860 images) was classified. Even the model with the slowest classification speed on this dataset (VGG19) is able to classify 256 images per second (Figure 3.16). In this manner, application of an automated system would allow for many more metres of core to be interpreted with lower human input and effort. As such, the role of a specialist geologist could be reallocated from lower-level descriptive tasks to QA/QC of automated classification, interpreting more complex images or exploring the broader meaning and implications of the data.

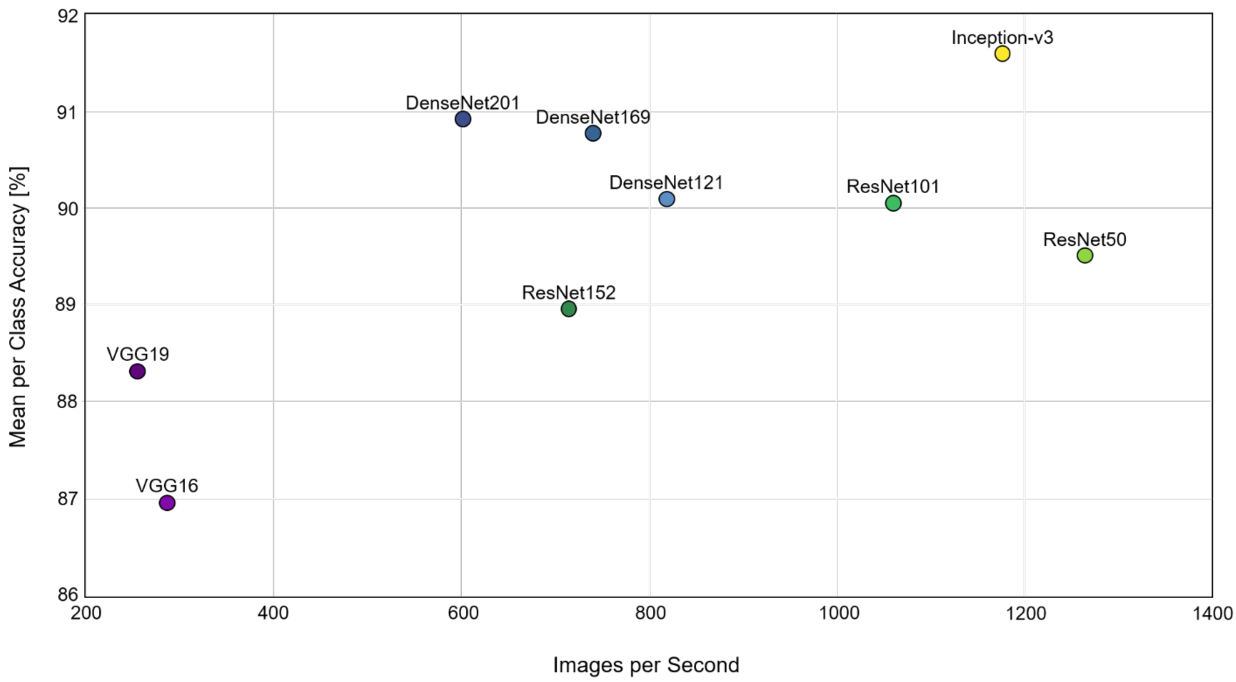

Figure 3.16: MPCA vs images per second for architectures trained on Dataset 104k.

The use of CNNs with either fewer layers or lower computational complexity draws advantages in shorter training times and lesser hardware requirements compared to their deeper, more complex counterparts. The overall training process can be facilitated by these shorter training times as this could enable the integration and implementation of improvement methods such as "human in the loop" annotations; for example, where training is supervised by an expert geologist, who may assess any misclassifications showing high losses above a defined threshold in the validation set for incorrect labels, helping to reduce label noise. This could also allow for images with increased resolution, since there is potential for more images to be processed in a shorter training time, meaning larger images could be cropped into multiple smaller divisions rather than being downscaled. This could be of particular relevance for carbonate classification according to the Dunham classification, where specific divisions of the scheme are based on components measuring 2mm or less; thus, requiring higher resolutions of input images, since crucial features could be lost due to downscaling.

As previously mentioned, the availability of large, labelled datasets is a major limitation for the application of deep learning in the geosciences. The comparisons of CNN architectures and dataset sizes in this chapter are intended to aid anyone wanting to apply an image classification model to their own geological dataset. Therefore, we draw recommendations for each dataset of different magnitude, so the most suitable model may be found for each. Comparing the results for Datasets 42k and 104k, we recognise similar trends in accuracy, precision, recall and computational resources. We would, therefore, again recommend the Inception-v3 architecture for datasets of this magnitude. Considering the smaller dataset (Dataset 7k), a clear trade off exists between computational complexity and higher classification performance (Figure 3.11). Therefore, for a smaller magnitude dataset, despite requiring more training time and computing resources, we would recommend the use of VGG19 due to the significant difference in accuracy achieved by this model over the other architectures evaluated. The transfer learning approach is seen to significantly reduce training times, with each model training completed within 5 hours. For comparison, the custom CNN model used in Zhang *et al.* (2021a) had a training time of around 12 hours.

3.4.2 Effect of dataset size on prediction performance

There is an increase in prediction performance with increasing magnitude of the dataset size (Table 3.4) suggesting that dataset size affects prediction accuracy, even in transfer learning with CNNs. This is an important finding, as the majority of current studies use deep learning on datasets with under 10,000 samples (Table 3.1). When trained on the smaller dataset (Dataset 7k), although the Inception-v3 and ResNet architectures produce high training accuracies (>90%), much lower test accuracies are achieved (Table 3.4), suggesting likely overfitting, i.e. that the models are unable to generalise well to unseen data. Soekhoe *et al.* (2016) found that for smaller target datasets, when using a fine-tuning approach, it is possible to reduce overfitting by leaving more layers frozen in the base model. By contrast, for large datasets, models can be improved through unfreezing and training more layers. Our results support this dichotomy, with the CNNs trained on Dataset 104k proving to be much more robust to overfitting.

Data augmentation is one way to artificially enlarge a training dataset, and reduce overfitting, by creating new images from existing training data that belong to the same class as the original image (Shorten and Khoshgoftaar, 2019). One of the most popular methods of data augmentation is the application of traditional affine and elastic transformations. This involves generating new images by applying procedures such as rotations, reflections, shifting, distortions, cropping, scaling, or colour shift. The '*tensorflow-Keras'* deep learning library in Python provides an easy way to incorporate augmentation using the '*ImageDataGenerator*' class (Keras Library); and the open-source Python library '*Albumentations*' (Buslaev *et al.*, 2020) offers a highly diverse set of more complex image transform operations that are optimised for different computer vision tasks, including image classification. Data augmentation techniques can, ultimately, help to expand a limited dataset, reduce overfitting, and improve the robustness of a model. However, it is important to ensure that the transformations applied to the images do not alter them in such a way that they are no longer recognisable by a carbonate geologist as, or a realistic representation of, carbonate rocks.

It can be argued that traditional data augmentation techniques generate limited new data and little additional information can be gained from these relatively simple modifications to the original images (Motamed *et al.*, 2021). While it is important to the robustness of the model, observing and classifying the same image in different states (e.g., rotation through 90º) may not have a large impact on the overall generalisability of the model. Synthetic data from generative models may provide more diversity in a dataset by producing new data samples to further improve the training process and increase model robustness and generalisation (Strelcenia and Prakoonwit, 2023). Among the broad family of generative methods, one of the most popular models are Generative Adversarial Networks (GANs), which take random noise from latent space and generate new data instances that mimic the feature distribution of the original dataset (Goodfellow *et al.*, 2014; Wang *et al.*, 2023a).

This study has highlighted that even when using transfer learning the size of the dataset plays a key role in the performance of the models, with those trained on the smaller datasets showing a strong tendency to overfit. Our results do, however, indicate that transferring features learned from a large source dataset to a much smaller target dataset still adds value by improving prediction performance and reducing risk of overfitting. It is shown that classification performance can continue to improve as more data is provided to the model; however, the capacity of the model needs to be regulated to support such data increases. Ultimately, there may be a point of diminishing returns where more data will not provide more insight into how to better improve classifications. For future applications applying transfer learning to geological images, we would recommend a minimum dataset size of 100,000 images which, in the case of limited samples, could be achieved through applied data augmentation and synthetic data techniques.

3.4.3 Understanding sources of error and model biases

One of the major aims for this project is to reduce the subjectivity of carbonate facies classifications. Confusion matrix analysis highlights some specific errors affecting individual classes, which may be arising from sources of biases. We observe the main areas of misclassification are between the floatstone and rudstone classes, the boundstone and rudstone classes, the wackestone and packstone classes, and the mudstone and wackestone classes (Figure 3.17). This suggests that the error is occurring between facies that are adjacent in the modified Dunham classification. Therefore, the trained networks errors are similar to errors a geologist could make. Furthermore, the results show that the CNN misclassified images of floatstone for rudstone and wackestone for packstone more frequently than the reverse, thus more frequently misclassifying matrix-supported textures as grain-supported textures. Despite this small saliency bias, the CNN performs more consistently and more accurately than the human classification results in the study by Lokier and Al Junaibi (2016).

The causes of this bias in the model are likely to stem from three possible sources. Firstly, the bias could have been introduced through the human labelling of the core images. Saliency is a known cognitive-psychological, ecological and evolutionary bias in humans (Korteling *et al.*, 2018). As implemented in this study, when analysing a new image, a CNN will generate a set of probabilities that this image belongs to each of the learned microfacies. Since the current dataset consists solely of images classified by the lead author, it is plausible that an element of bias could have been introduced into the algorithm training and, therefore, the final model. We acknowledge that this cannot be completely eliminated as a source of inaccuracy in this study. There is a possibility to reduce interpreter bias in future workings through the compilation of datasets consisting of multiple classifications made by many individuals through projects developed on online crowding-sourcing services. Citizen science already contributes to a variety of scientific fields, including astronomy, environmental science, ecology, medicine, and seismology (Theobald *et al.*, 2015; Haklay *et al.*, 2018). There are now thousands of projects collecting and processing data annually, with many contributing to peer-reviewed articles (Kelling *et al.*, 2019; SciStarter.org, 2020). When properly designed, conducted, and evaluated, these projects can generate high-quality data to solve diverse problems.

Secondly, it is possible that the CNN architecture itself is biased towards salient features. Therefore, a cause of the error could be through image data encoding before classification. Where a geologist would rely upon a defined set of features and measurements to classify carbonate facies according to the modified Dunham scheme, the CNNs operate with no prior knowledge of specific attributes and, thus, perform the classifications based solely on image characteristics. Taking the misclassifications between mudstone and wackestone, for example, the feature vectors in an image of a mudstone would be sparse as mudstones tend to be relatively featureless. Hence, classification would be based on a sparse feature vector for this class. However, mudstones and wackestones are both mud-supported classes, with a subtle threshold difference at 10% grains. For borderline cases e.g., grainier mudstones containing 5-9% grains, the grains would become a significant feature, and the CNN may place larger weights on these features, thus leading to a misclassification as a wackestone.

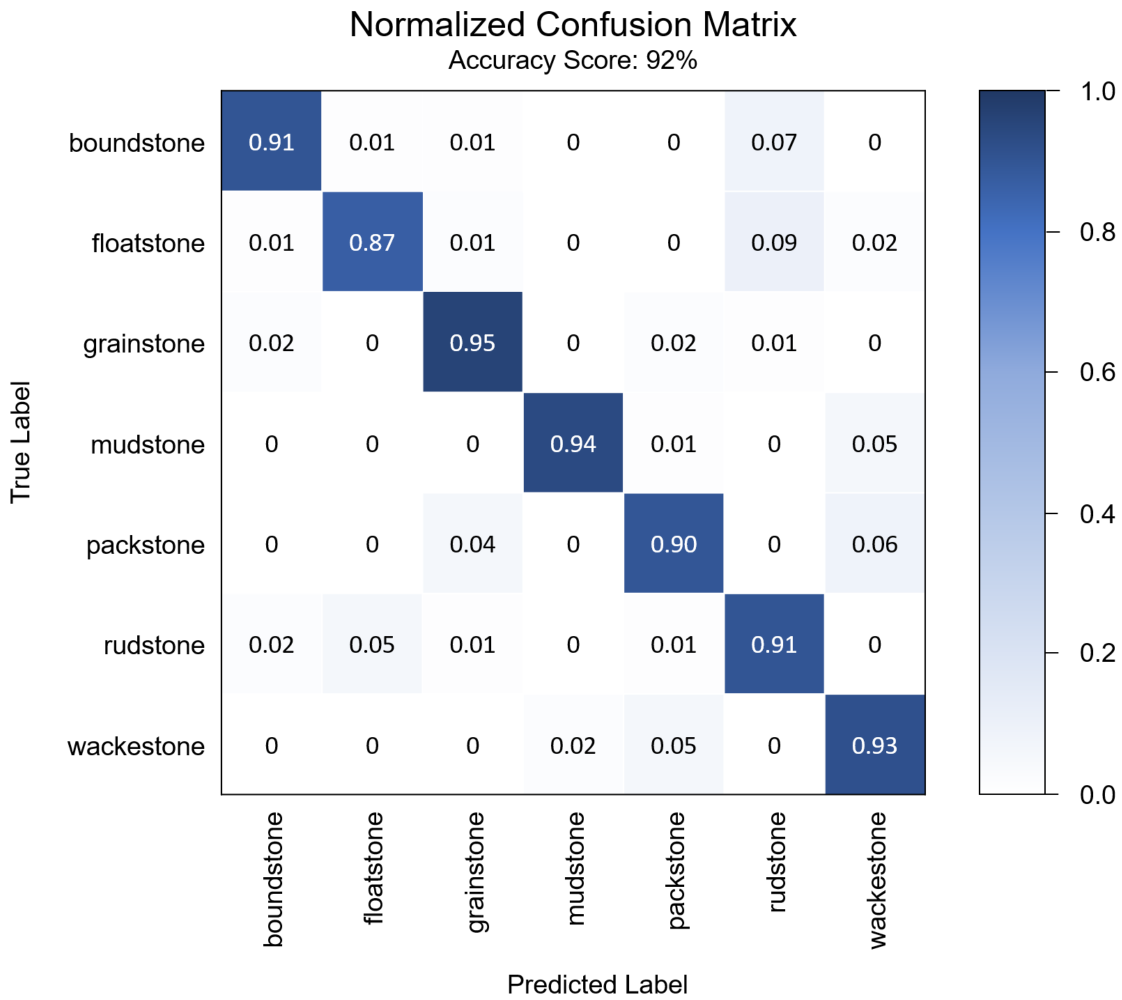

Figure 3.17: Normalised confusion matrix for the Inception-v3 classifier on Dataset 104k. Average model accuracy = 92% for 7 facies.

However, it is also worth considering whether the CNN methods are hinting at a deeper issue: the classification of a continuum of rock characteristics into a small number of discrete classes. Many rock textures, especially those close to the threshold values (e.g., mudstone vs. wackestone), exist on a spectrum, and classifying them into distinct categories could inherently introduce ambiguity. The issue may not lie solely with the ability of the CNN to distinguish between classes but rather with the conceptualisation of rock facies as discrete classes in the first place. In such cases, the CNN might be struggling with this artificial binning of continuous properties into rigid categories, which could lead to errors for classes that have overlap in their characteristics. To address this issue, introducing more labelled images of borderline cases to the dataset could likely further improve accuracy by providing the model with more representative examples of the spectrum of rock features. Additionally, having multiple individuals classify these images could highlight any differences in interpretation, further helping to refine the classification process. From this, it is logical to suggest that the bias may also be due to limited or imbalanced data. Considering the confusion matrix in Figure 3.18, which is based on the balanced dataset, we observe that the misclassifications are again arising in the adjacent Dunham classes; floatstone and rudstone, boundstone and rudstone, wackestone and packstone, mudstone and wackestone, and, additionally, grainstone and packstone. To this effect, it should be noted that the balanced dataset is considerably smaller in magnitude, so we cannot rule out that the error here could, in part, be due to lower sampling.

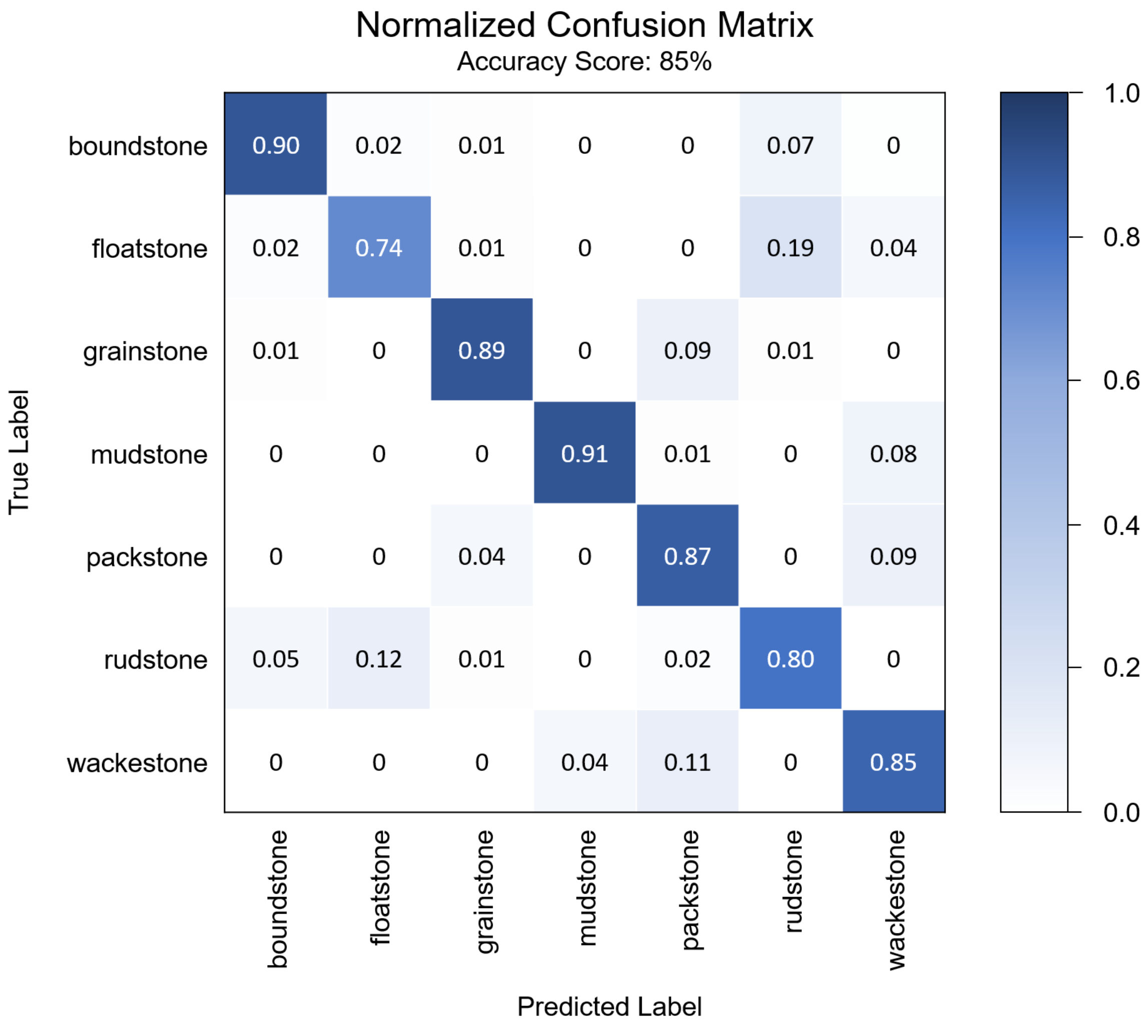

Figure 3.18: Normalised confusion matrix for the VGG19 classifier on Dataset 7k. Average model accuracy = 85% for 7 facies.

3.5 Conclusions

This study builds on a new approach to carbonate core classification though deep learning, where image classification algorithms aim to improve overall interpretation accuracy, as well as reducing subjectivity and interpretation time. This chapter evaluated the ability of nine different CNNs of four architectures to classify carbonate core images, systematically developing a performance comparison between these architectures on three complex geological datasets with orders of magnitude difference in data quantity. Following a transfer learning and fine-tuning approach, we used the learned parameters from the pre-trained CNN architectures to classify the core images according to the modified Dunham Classification, where classes are defined based on textural observations (Dunham, 1962; Embry and Klovan, 1971).

The results show great potential for the use of deep learning in automated carbonate classification from digital core images, where high level performance was achieved across all models, even with limited and unbalanced datasets. The highest overall accuracy of 92% was achieved by the Inception-v3 architecture when trained on the larger (104k) dataset. When considering all evaluation metrics presented, for textural carbonate core classification, we find the Inception-v3 architecture to be the most suitable model for medium to large datasets, and the VGG19 architecture to be the most suitable for smaller datasets. Furthermore, we have shown that even when using transfer learning the size of the dataset plays a key role in the performance of the models, with those trained on the smaller datasets showing a strong tendency to overfit.

The use of machine learning in geoscience is increasingly recognised as a promising tool, but its applications are still developing, and the adoption of machine learning techniques remains limited in many subfields. While the growth of interest in applying deep learning methods to geological problems is evident, it is still unclear how widely and effectively these methods can be integrated into practical geoscience workflows. One significant challenge in improving the prediction performance of CNNs for geological applications is the lack of large, labelled training datasets. However, as data availability in geosciences continues to increase, with more recent data being generated and the digitalisation of existing datasets from long-established sources, the potential for improving deep learning models grows. Ultimately, the development and application of the framework laid out in this study could aid the future of deep learning-based carbonate classification and could be easily modified for the classification of cores originating from different lithologies and formations, paving the way for future developments in this area.

Chapter 4: Generative Adversarial Networks (GANs) for Synthetic Carbonate Core Images

ABSTRACT

Deep learning techniques have brought significant advances to a broad spectrum of computer vision applications, largely through the use of extensive annotated datasets. However, for the geoscience domain, acquiring such datasets poses considerable challenges. This chapter introduces methods for generating carbonate textures using Generative Adversarial Networks (GANs) and demonstrates how these synthetic images can enhance convolutional neural network (CNN) performance in core image classification through data augmentation. The study focuses on a limited dataset of 7,000 carbonate core images (5,600 training images) and compares two GAN architectures: Deep Convolutional GAN (DCGAN) and StyleGAN2. These models are used to generate high-quality carbonate textures representing seven classes from the modified Dunham Classification. The quality of the synthetic textures is evaluated using visualisation assessments and class validation with a top-performing CNN. Additionally, a CNN is trained using both traditional data augmentation techniques and synthetic data, developing a performance comparison for reducing overfitting on a complex geological dataset. StyleGAN2 with adaptive discriminator augmentation (ADA) outperforms DCGAN, achieving a Fréchet Inception Distance (FID) of 6.79 on the carbonate core dataset. Using only traditional data augmentation, the CNN achieved a sensitivity of 71.4% and specificity of 74.2%. By incorporating synthetic data from the StyleGAN2-ADA model into the training set, sensitivity increased to 76.7% and specificity to 78.1%. These findings suggest that robust models can be developed even with small or imbalanced datasets. The synthetic data augmentation approach outlined here offers potential for broader application to other geological image datasets, contributing to improved classification performance in geoscience.

4.1 Introduction

In recent years, f machine learning technologies have been evolving rapidly, resulting in an active uptake of applications within the Earth Sciences and garnering significant interest from industry, academia and popular culture (Karpatne *et al.*, 2018; Koroteev and Tekic, 2021; Sircar *et al.*, 2021). This is, in part, due to rapid developments in the field of computer vision, particularly with breakthroughs in deep learning algorithms that have enabled computers to better discover and learn complicated patterns in data. Such breakthroughs have been fuelled by the increase in available computing power, access to big data, and userfriendly software frameworks, allowing neural networks to be developed that are deeper than ever before. These models have achieved state-of-the-art performances on large benchmark datasets, such as ImageNet (Deng *et al.*, 2009) and CIFAR-100 (Krizhevsky and Hinton, 2009) and have even surpassed human performance on the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) (Russakovsky *et al.*, 2015; Lundervold and Lundervold, 2019). However, performance on these challenges does not necessarily translate well to many real-world applications, due to limited representative training data.

Even when using transfer learning, training a deep learning model on insufficient data can result in overfitting, where the model is overly complex relative to the size of the dataset and fails to generalise the features and patterns present in the training dataset (Dawson *et al.*, 2023). Overfitting remains a major issue in machine learning and there has been an extensive range of literature dedicated to developing regularisation strategies, such as dropout and weight decay, for addressing this problem (e.g., Srivastava *et al.*, 2014; Schmidt *et al.*, 2019; Hosseini *et al.*, 2020; Montesinos López *et al.*, 2022). These methods aim to simplify models, thus preventing them from learning irrelevant patterns in the data. Data augmentation, the focus of this chapter, relates to techniques that aim to improve model performance by artificially increasing the size and quality of training datasets through the generation of new and diverse instances (Konidaris *et al.*, 2019; Shorten and Khoshgoftaar, 2019).

One of the major challenges limiting the wider successful adoption of machine learning-based methods in geoscientific applications is the requirement to train models with real geological data. Although many operations may involve large amounts of data, acquiring datasets with ground-truth labels is challenging due, in part, to the laborious and expensive nature of geological data collection. In many cases, this data is also rarely made publicly available, often due to legal restrictions preventing sharing between operating companies and countries. Synthetic data offers a promising solution to this challenge. By generating artificial data that is representative of geological textures without being location-specific, synthetic datasets can be used for training deep learning models.

Generative Adversarial Networks (GANs) (Goodfellow *et al.*, 2014) are a leading method for synthetic data generation, offering the ability to produce realistic and diverse samples. GANs have demonstrated remarkable success across various domains, from augmenting synthetic medical imaging datasets (e.g., Motamed *et al.*, 2021) to fluid flow prediction (e.g., Siddani *et al.*, 2021).). By enriching datasets with synthetic examples, GANs can help to mitigate challenges related to data scarcity and privacy concerns. Here, we present a comparison of two popular GAN architectures, Deep Convolutional GAN (DCGAN) (Radford *et al.*, 2015) and StyleGAN2-ADA (Karras *et* al., 2020), for the generation of synthetic carbonate textures to enhance CNN performance in core image classifications.

DCGAN is an early and widely adopted GAN architecture designed to stabilise training and improve the quality of generated images by leveraging deep convolutional layers. It replaces fully connected layers with convolutional operations, enabling it to learn hierarchical spatial features, which are critical for generating realistic textures (Radford et al., 2015). However, DCGANs can struggle to generate high-resolution images and may be prone to mode collapse, where the generator produces a limited variety of samples (Jain *et al.*, 2022).

StyleGAN2 introduces several improvements over the earlier GAN architectures. It introduces a style-based design, which separates the generation process into distinct stages, enabling finer control over features such as texture, structure, and detail. This design improves the quality and diversity of generated samples, making it particularly suitable for creating high-resolution images with complex patterns (Chai *et al.*, 2024). These enhancements make StyleGAN2 well-suited for geoscientific applications, where datasets are often limited and highly specialised.

The synthetic textures generated by these models are then used to augment the training data for convolutional neural networks (CNNs) in a core image classification task. Using the 7k carbonate core dataset from Dawson *et al.*, 2023 (Chapter 3), we then evaluate the effectiveness of StyleGAN2 compared to traditional data augmentation techniques. This chapter provides a glimpse into the transformative role of GANs in the realm of synthetic data augmentation, offering a powerful solution to address data-driven challenges in deep learning applications.

4.1.1 Background

The widespread adoption of data augmentation in deep learning is evident across many fields (e.g., Hussain *et al.*, 2018; Taylor and Nitschke, 2018; Chlap *et al.*, 2021; Nanni *et al.*, 2021; Lalitha and Latha, 2022). Within these studies, the traditional augmentation techniques most applied involve rotations, reflections, brightness and contrast adjustments, colour space shifts, scaling, and applying filters, such as blurring, sharpening and edge detection (Figure 4.1). The primary objective of these transformations is to generate new images that preserve the original semantic content while introducing diversity into the dataset. The semantic content of an image refers to the meaningful information it conveys, such as the objects, patterns, or features critical for understanding or recognising what the image represents (Machajdik & Hanbury, 2010). For instance, in the case of geological images, semantic content might include the specific rock textures, mineral patterns, or structural features used for classification or analysis.

Despite advantages in improving the learning and generalisation capabilities of neural networks, data augmentation is not without limitations. Typically, augmentation techniques are constrained to minor image alterations, as more substantial changes may compromise the semantic content, thereby disrupting the integrity of the data. Moreover, the suitability of augmentation methods will vary across different problems. For example, geological images, such as those of cores or thin sections, would require more conservative augmentations due to established conventions and standards, e.g., the classification of sediments and sedimentary rocks according to the British Geological Survey Rock Classification Scheme (Hallsworth and Knox, 1999). Additionally, traditional augmentation operates on a per-image basis, lacking the ability to leverage information from the broader dataset to enrich the generated samples.

Deep learning-based augmentation techniques can automatically learn the features and patterns within images to generate realistic synthetic data, enhancing the robustness and generalisation capabilities of models while reducing the chances of overfitting (Figure 4.1). By synthesising entirely new training samples, these methods offer a powerful alternative for addressing data scarcity in computer vision (Mumuni and Mumuni, 2022). Synthetic data generation is particularly beneficial when the available data for a specific task is insufficient, difficult to obtain, or of poor quality. Additionally, synthetic datasets can be tailored to provide more diverse and enriched representations suited to specific tasks. In recent years, variants of GANs have emerged as the most widely used deep learning models for generating synthetic image data (Chlap *et al.*, 2021).

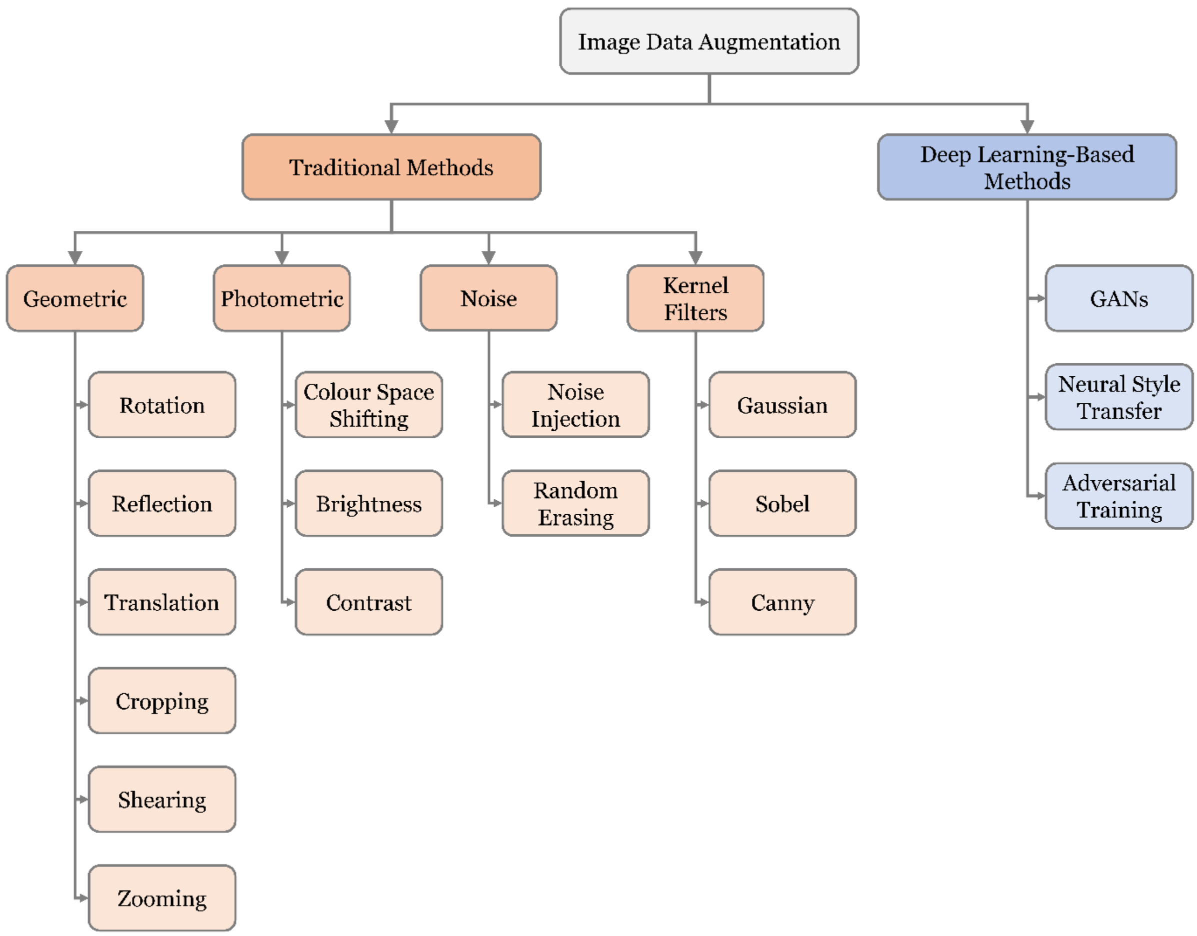

Figure 4.1: Data augmentation taxonomy.

Originally introduced by Goodfellow et al. in 2014, GANs employ an adversarial training framework in which two competing neural networks—the generator and the discriminator—are trained simultaneously in a two-player minimax game. The generator creates new data samples, while the discriminator evaluates these samples for authenticity, determining whether or not they belong to the actual training dataset (Figure 4.2). The generator begins by producing images from random noise vectors, increasing the possibility of producing a broad range of outputs that resemble the original dataset This is of particular value in fields where data are limited and challenging to collect (Frid-Adar *et al.*, 2018; Röglin *et al.*, 2022; Chan *et al.*, 2023).

Within the geosciences, GANs have gained attention in recent years for a variety of applications with a focus on deployment in the petroleum industry (e.g., Shams *et al.*, 2020; Zhang and Lin, 2020; Tang *et al.*, 2021; Klyuchnikov *et al.*, 2022). A large portion of this research has centred on reconstruction of porous media and fluid flow prediction (e.g., Mosser *et al.*, 2017; Siddani *et al.*, 2021; Volkhonskiy *et al.*, 2022; Yin *et al.*, 2023; Zhang *et al.*, 2023a). With regards to lithofacies prediction, most previous studies have explored various data sources, such as seismic (Picetti *et al.*, 2019; Liu *et al.*, 2020a), thin section (Nanjo and Tanaka, 2020; Rubo *et al.*, 2023), micro-computed tomography (micro-CT) (Niu *et al.*, 2020; Bizhani *et al.*, 2022), and wellbore data (Klyuchnikov *et al.*, 2022).

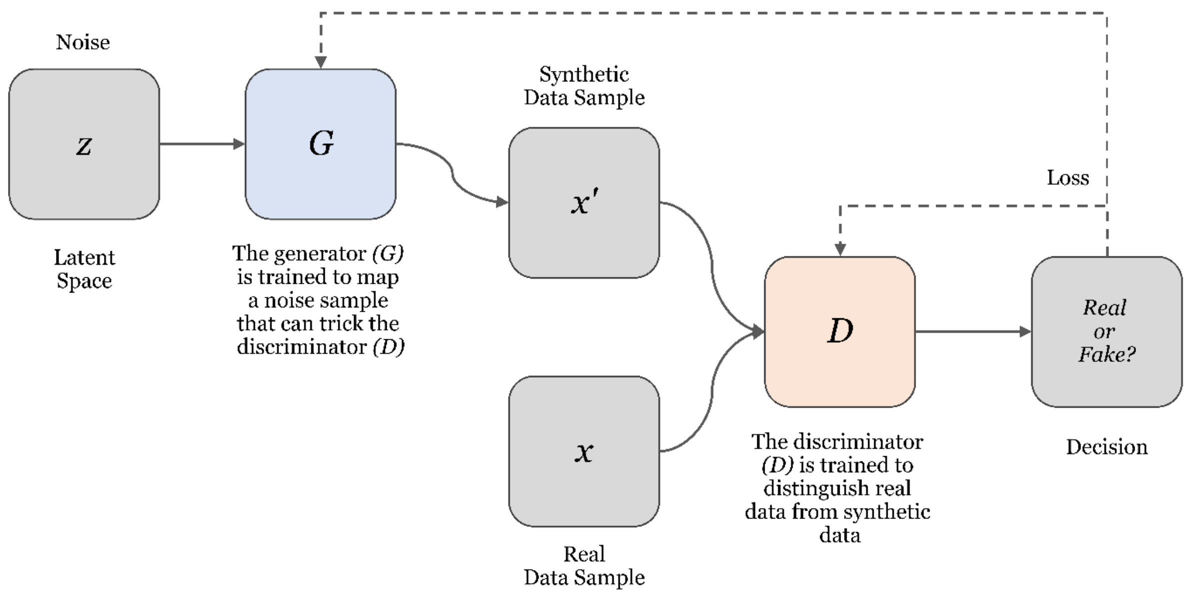

Figure 4.2: Basic structure of a Generative Adversarial Network (GAN). The two models that are learned during the training process are the generator (*G*) and the discriminator (*D*). They are generally implemented by multi-layer neural networks consisting of convolutional and/or fully connected layers. Modified from Creswell *et al.*, 2018.

The implementation of GANs for geological core images has seen limited exploration, and despite their potential for addressing data scarcity in deep learning tasks. One of the major limiting factors for applying GANs to core images is the requirement for a significantly large dataset, as demonstrated in the original GAN implementation with the MNIST (LeCun *et al.*, 1998), Toronto Face Database (TFD) (Susskind *et al.*, 2010), and CIFAR-10 (Krizhevsky *et al.*, 2014) datasets containing 70,000, 112,234 and 60,000 images, respectively. For some GAN architectures, the generator has also been shown to overfit when using datasets of around 1,000 images (Feng *et al.*, 2021).

To address these challenges, this chapter compares two GAN architectures, DCGAN and StyleGAN2, for the generation of synthetic carbonate textures. DCGAN, with its simple and robust convolutional architecture, was chosen for its foundational role in establishing GANs as effective tools for image synthesis tasks. StyleGAN2, recognised for its ability to generate highly realistic and diverse images, was selected due to its advancements in style-based architecture, which enhance control over image synthesis and improve visual quality. This comparative approach leverages the 7k carbonate core dataset from Dawson *et al.*, 2023 (Chapter 3), to explore how these architectures can address the dataset size limitations commonly associated with core images. We then compare synthetic data augmentation using GANs with traditional data augmentation methods to evaluate their respective impacts on improving the performance of CNNs for carbonate core classification. The ultimate objectives of this work are:

(1) Develop a novel application of GANs to generate synthetic carbonate textures from a small original dataset.
(2) Assess the quality of the synthetic images through common evaluation metrics and visual inspection.
(3) Validate the classes of generated textures by deploying a trained carbonate core classification model on the synthetic dataset.
(4) Compare the impact of synthetic data vs traditional data augmentation methods for improving CNN classification performance.

By demonstrating the capability of GANs to generate high-quality synthetic data, we further aim to highlight their potential to advance geological image generation tasks. The findings will underscore how generative models can support the application of deep learning algorithms to complex geological datasets, paving the way for broader adoption in geoscientific research.

4.2 Methodology

4.2.1 Data preparation

We use the 7k image dataset from Chapter 3 (Dawson *et al.*, 2023), which contains 7,000 sliced carbonate core images divided equally across seven classes from the modified Dunham Classification (Dunham, 1962; Embry and Klovan, 1971): mudstone, packstone, wackestone, grainstone, floatstone, rudstone, and boundstone, as defined earlier in Section 3.1.1 (Figures 3.3 and 4.3). Each class contains 1,000 images. The original dataset was created with three-colour channel (RGB) squared images with sides of 299 × 299 pixels. In order to draw comparison of classification performance after various augmentations, the training and testing dataset splits from the original implementation were kept the same, resulting in a training dataset of 5,600 core images across the seven classes.

Data preprocessing was carried out using standard image libraries included in the StyleGAN2 framework (Karras *et al.*, 2020), OpenCV (Bradski, 2000), Numpy (Harris *et al.*, 2020), Pillow (Clark, 2015), and PyTorch (Paszke *et al.*, 2019). Following the specifications of the StyleGAN2 architecture, input images were required to have square dimensions that are powers of two (i.e., 32 × 32, 256 × 256, 512 × 512 pixels). The original images were downsized to 256 × 256 pixels to satisfy the StyleGAN2-ADA requirements while retaining a high resolution to preserve the essential textural features of the core dataset (Karras *et al.*, 2020). The synthetic images were later upscaled to 299 × 299 pixels, so as to meet the input requirements for the CNN classification model.

For the DCGAN application (Radford *et al.*, 2015), the model received the 256 × 256 pixels images as input. To minimise computational overhead when generating larger images, synthetic images were initially created at 64 × 64 pixels These were later then upscaled to 299 × 299 pixels through bilinear interpolation to meet the input requirements for the CNN classification model.

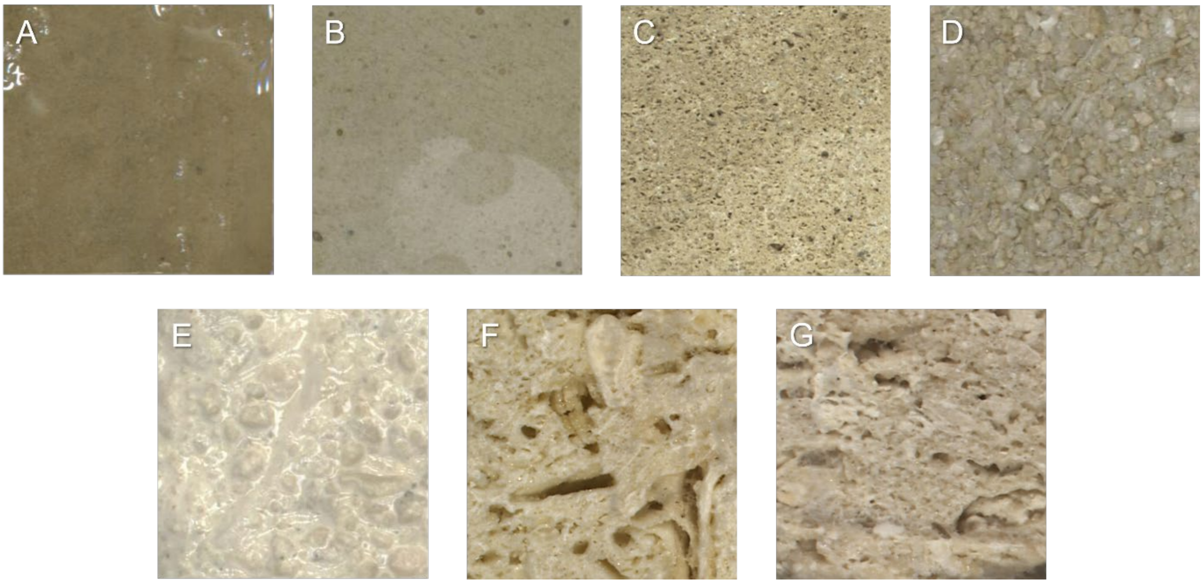

Figure 4.3: Sample original dataset images: (a) mudstone, (b) wackestone, (c) packstone, (d) grainstone, (e) floatstone, (f) rudstone, (g) boundstone.

4.2.2 Experiment design

While the test dataset was left unaltered, the following transformations were applied to the training dataset:

(1) **Traditional augmentations**: including geometric, photometric, noise and kernel filter methods.

(2) **Synthetic image generation**: using GANs to create new images to enrich the original training dataset with realistic synthetic samples.

After each transformation, a CNN was trained using transfer learning and evaluated on the untouched test dataset (Figure 4.4). For the CNN model, we use the Inception-v3 model (Szegedy *et al.*, 2016). Although the VGG-16 architecture was concluded to be the best performing architecture when using transfer learning on a small geological dataset of <10,000 images (Dawson *et al.*, 2023), by applying data augmentation techniques we can increase the size of the original dataset into the 10,000 – 100,000s. When using a dataset of <10,000 images, the Inception-v3 model was seen to overfit. Therefore, using Inception-v3 as the base classification model provides a robust test of whether the implemented augmentation methods can mitigate overfitting and improve classification performance.

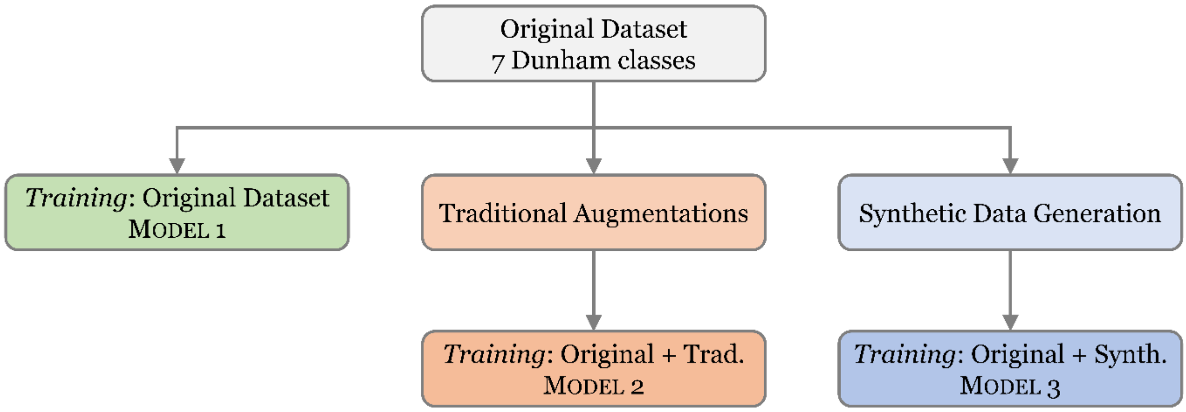

Figure 4.4: Workflow for training the CNN models after each data augmentation application. Abbreviations: Trad., traditional augmentations; Synth., synthetic images generated by GANs.

4.2.3 Generative adversarial networks for core image synthesis

StyleGAN2

StyleGAN2 was chosen as the primary generative model for its ability to deliver state-of-the-art results on high-resolution images and its robustness when working with small datasets containing fewer than 1,000 images (Karras et al., 2020). The experiments were conducted using the StyleGAN2 configuration of the official StyleGAN3 repository[9] (Karras *et al.*, 2021). While most default parameters provided by the implementation were retained, the following modifications were made to optimise performance for the dataset. The β_1 parameter in the Adam optimiser was adjusted to 0.9 to improve convergence stability by reducing oscillations during training. Mixed precision was disabled to ensure numerical accuracy, particularly critical for small datasets where minor inconsistencies could impact training outcomes. A hyperparameter search was not conducted due to the decision to use established, well-tested parameters from the repository for consistency and reproducibility.

To evaluate the effects of transfer learning and data augmentation on image synthesis, an ablation study was conducted. In this context, an ablation study systematically isolates and tests individual components or configurations of a model to determine their impact on performance (Islam *et al.*, 2024). The following experimental designs were implemented:

(1) **Baseline**: Training was initiated from scratch with all StyleGAN2 augmentations disabled.
(2) **Pre-trained**: Training began using pre-trained StyleGAN2 weights from the Flickr-Faces-HQ (FFHQ) dataset, with augmentations disabled.

[9] https://github.com/NVlabs/stylegan3.

(3) **Augmented**: Training was conducted from scratch with mirroring (horizontal flipping) and adaptive discriminator augmentation (ADA) enabled.
(4) **Pre-trained and Augmented**: Training utilised pre-trained FFHQ StyleGAN2 weights with mirroring and ADA enabled.

All experiments were conducted on the 7k carbonate core dataset. Training ran for 6,250 ticks, where a tick represents a fixed number of training steps in the StyleGAN2 framework. Each tick corresponds to processing a batch of images through both the generator and discriminator, with metrics computed and model weights saved every 50 ticks. Each of the experiments took approximately 4.5 to 5 days to complete for the 256 × 256 pixels dataset on the NVIDIA Pascal P100 GPU. To ensure robust and reliable results, each configuration was repeated five times.

Deep Convolutional GAN (DCGAN)

We consider a second GAN architecture, DCGAN (Radford *et al.*, 2015), that has been widely used in synthetic image reconstruction to draw a performance comparison of the generative models. The generator in this model employs a series of fractionally-strided convolutional layers that progressively transform an input noise-vector into a tensor matching the shape of the training data. Fractionally-strided convolutional layers, also known as transposed convolutional layers, are used to increase the spatial resolution of feature maps. They perform the opposite operation of standard convolutional layers, expanding the dimensions of the input instead of reducing them. This allows the generator to progressively up-sample the noise vector, adding detail and structure at each step until the output resembles the target data (Meng *et al.*, 2021). The structures of the generator and the discriminator used in this study are provided in Appendix B (Figures B.1 and B.2).

The DCGAN model uses the minimax loss as described in the original GAN paper (Goodfellow *et al.*, 2014). For this study, the DCGAN implementation was adapted using the TorchGAN framework (Pal and Das, 2021), which is designed for handling 2D image data. Unlike some other GAN architectures, DCGAN does not support the generation of class specific labels. As a result, synthetic images were generated separately for each class in the dataset to ensure that the output aligned with the class distribution of the original training data.

4.2.4 Evaluation metrics for synthetic images

Fréchet Inception Distance (FID)

To evaluate the quality of the generated images, we used the Fréchet Inception Distance (FID) , a widely recognised metric for assessing the performance of GANs in natural imaging. The FID measures the similarity between two sets of images—real and generated—by comparing their statistical distributions. Specifically, it calculates the Fréchet distance between two multivariate Gaussians distributions. These distributions are derived from the feature representations of the images extracted from the coding layer of an Inception network pre-trained on the ImageNet dataset (Heusel *et al.*, 2017).

In more practical terms, the FID evaluates how close the generated images are to the real ones in terms of both visual realism and diversity. It does this by analysing patterns in a high-dimensional feature space rather than solely on pixel-level similarities. Lower FID scores indicate that the synthetic images more closely resemble the real dataset, demonstrating better quality and variety (DuMont Schütte *et al.,* 2021).

The interpretation of FID scores varies depending on the specific dataset and application context. A FID score of 0 would indicate perfect similarity between the real and generated datasets, meaning that the GAN has generated synthetic images indistinguishable from the real ones (Benny *et al.,* 2021). For natural image datasets, such as CIFAR-10 (Krizhevsky et al., 2014) or CelebA (Liu *et al.*, 2015), and complex high-resolution datasets, such as ImageNet (Krizhevsky *et al.*, 2012), achieving lower FID scores is more challenging as the characteristics of the data can make it harder to accurately model their distributions. For example, unconditional generation implementations of StyleGAN2 and DCGAN achieved FID scores of 11.07 (Zhao *et al.*, 2020) and 15.86 (Zhao *et al.*, 2021), respectively. In contrast, the conditional generation implementation of StyleGAN-XL (Sauer et al., 2022), a model specifically tailored for large diverse datasets, achieved FID scores of 1.85 on CIFAR-10 and 2.30 on ImageNet (256 x 256). However, it is important to note that while lower FID scores indicate better performance, the specific threshold for what constitutes a "low" score is relative and will vary based on the dataset and the specific application. For example, if one GAN achieves a FID of 15 and another scores 10 on the same dataset, the latter is deemed better, even if neither score is objectively low. Therefore, FID scores should be interpreted in the context of the specific dataset and compared against baseline scores or other models evaluated under similar conditions.

The FID offers several advantages of the FID that make it an effective evaluation metric. First, it can correlate reasonably well with human perception due to its focus on high-level features that capture semantic content. This aligns similarly to how humans typically perceive quality and realism (Heusel *et al.*, 2017; Borji, 2022). It should be acknowledged, however, that human perception involves subjective and contextual factors, which the FID cannot capture. Second, it is sensitive to distortions, allowing it to detect subtle artefacts or inconsistencies in the generated images (Borji, 2022). Finally, it is computationally efficient, requiring a relatively manageable number of samples to provide robust evaluations (Heusel *et al.*, 2017; Borji, 2019). Given these strengths, the FID was chosen as a key metric to quantify the realism and diversity of the carbonate core textures generated by the GAN models.

CNN Class Validation

Using the Inception-v3 architecture trained on the 104k dataset from Chapter 3 (92% accuracy) (Dawson *et al.*, 2023), we assess whether the synthetic images generated by the GANs would be correctly classified by the classifier model. For each class, we check if the predicted label matches the class label of synthetic image and evaluate the test with precision and recall, as follows.

Precision calculates the proportion of positive identifications that were correct i.e., the probability that the predicted class of the carbonate facies actually belongs to that class:

$$\text{Precision: P} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{4.1}$$

whereby TP is the number of true positive cases and FP is the number of false positives.

Recall calculates the proportion of actual positives that were correctly identified:

$$\text{Recall: R} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4.2}$$

whereby FN is the number of false negatives.

4.2.5 Traditional augmentation techniques

Traditional augmentation techniques involve applying basic transformations to images, usually by either remapping pixel positions or manipulating intensity values to produce a new image. Each transformation is applied to an individual image from the existing dataset, and the modified image is added back into the dataset, increasing its overall size. While relatively straightforward to implement, this approach can effectively enhance the performance of a trained model by diversifying the training data. Here, we provide an overview of these methods and how they were applied in this study.

To apply the transformations to the images, we used both the standard tf.image package (Figure 4.5) and the more complex Albumentations (Buslaev *et al.*, 2020) (Figure 4.6) library. A total of 29 transformations were applied to the 7k dataset, creating an augmented dataset of 168,000 images. These transformations included geometric adjustments, intensity manipulations, and filtering techniques. It should be noted that the transformations applied to the original images did not alter them in such a way that they were no longer realistic representations of carbonate rocks.

To ensure the augmented images remained realistic representations of carbonate rocks, criteria such as consistency with domain-specific conventions, maintaining key textural features, and visual inspection by experts were used. Transformations that risked altering the semantic content or geological relevance of the images were excluded. This ensured that the augmented dataset remained both diverse and representative of real-world carbonate textures.

Geometric transformations

Geometric transformations are among the most commonly used basic augmentation techniques. These typically involve affine transformations such as translating, rotating, and reflecting (Figure 4.5). Affine transformations are a set of linear mapping operations that preserve points, straight lines, and planes. They allow for basic geometric manipulations of an image while maintaining the spatial relationships between points (Burger *et al.,* 2009). This means that parallel lines remain parallel, and relative proportions between points are preserved. The transformation parameters can be predefined or can be randomly sampled to introduce variability. Despite their simplicity, these methods , have been shown to significantly enhance performance in many computer vision tasks (Wong *et al.*, 2016; Mumuni and Mumuni, 2022). Due to their effectiveness and ease of implementation, geometric transformations are often applied as a preliminary step before introducing more advanced augmentation methods.

Photometric transformations

Unlike geometric transformations, photometric augmentation methodsmodify the pixel intensity values of images while maintaining their spatial structure (Khalifa *et al.*, 2022). Pixel intensity values refer to the numerical values assigned to each pixel in an image that determine its colour and brightness. These values can vary depending on the colour model used (e.g., greyscale or RGB), but the principle is the same; they represent the visual characteristics of the image.

In this study, we used photometric transformation techniques to alter the composition of the RGB channels using specified functions, such as adjusting the brightness or contrast, or through gamma correction, which adjusts overall brightness while preserving details in shadows or highlights (Figure 4.5). The most used photometric data augmentation approaches in the Albumentations library include colour jittering, colour shift and image enhancement techniques (Figure 4.6). Colour jittering ("*ColorJitter*", Figure 4.6) randomly modifies the intensities in individual RGB channels, simulating variations in lighting or colour balance. Colour shift ("*RGBShift*", Figure 4.6) adds or subtracts a fixed value from one or more colour channels, shifting the overall colour tone of the image (Buslaev et al., 2020) These techniques contribute to dataset variability by altering visual properties without compromising the structural integrity of the image.

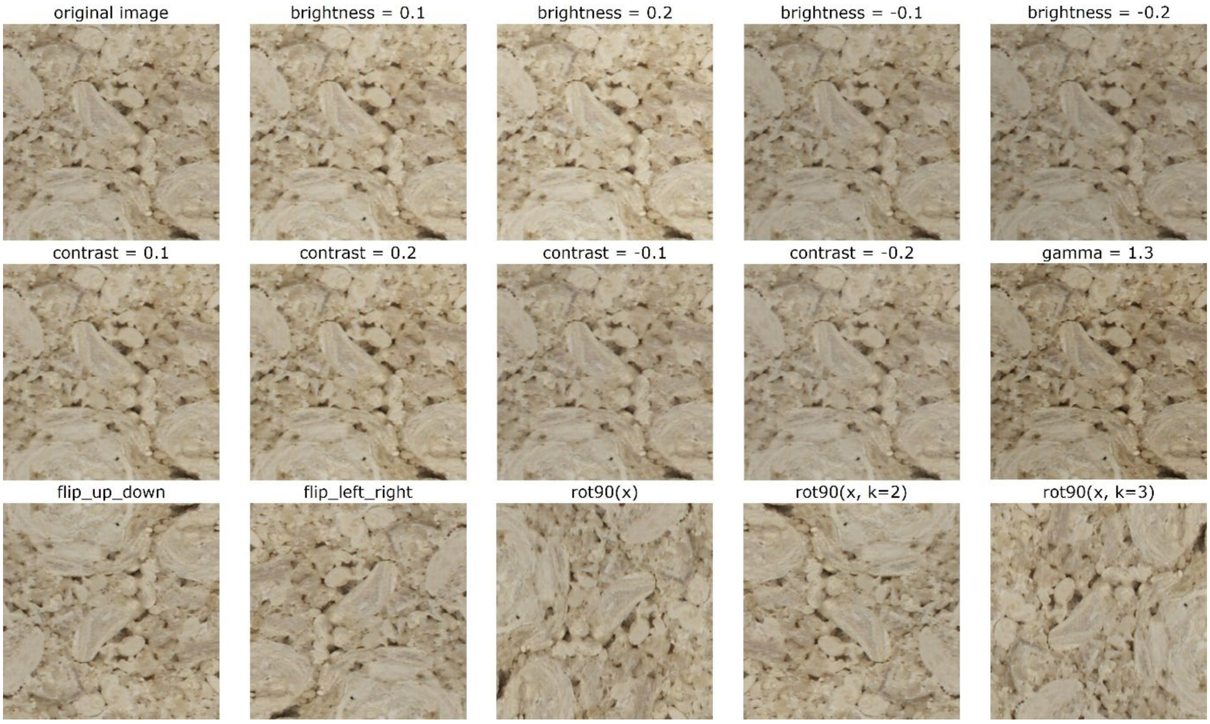

Figure 4.5: Examples of commonly used basic data augmentation techniques from the tf.image package. An original instance from the rudstone class is shown in the top left. Abbreviations: rot90, rotate image(s) by 90 degrees.

Noise injection

Noise injection is a widely used augmentation strategy that aims to simulate realistic image noise. Gaussian noise injection, the most common approach modifies pixel intensities by adding values sampled from a Gaussian distribution (Figure 4.6). This technique can improve model robustness by training it to handle noisy input conditions.

Filtering

Filtering involves applying convolutional operations to modify pixel intensities based on their neighbours. A convolutional kernel is passed across the image, enabling transformations such as sharpening, blurring, or smoothing. S Additionally, edge detection filters can enhance and emphasise object edges within images (Figure 4.6). These augmented outputs diversify the dataset while retaining the essential features needed for effective model training.

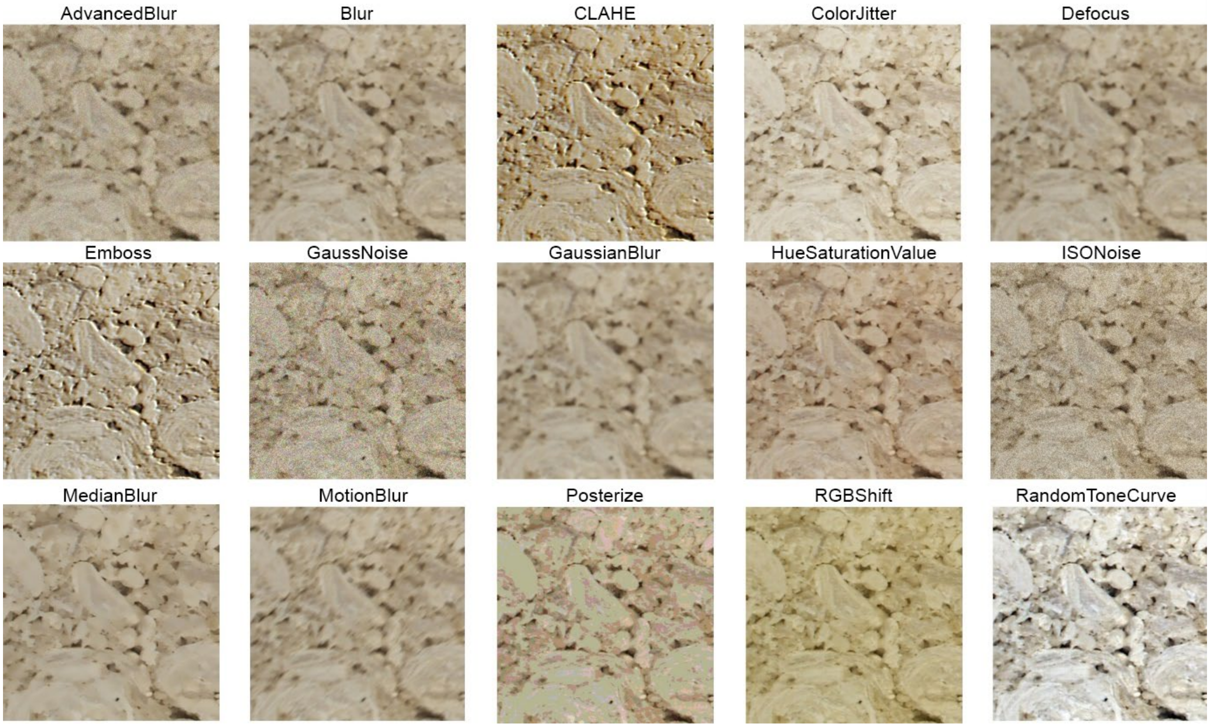

Figure 4.6: Examples of the photometric, noise and kernel filter data augmentation techniques used in this study from the Albumentations library. Abbreviations: CLAHE, Contrast Limited Adaptive Histogram Equalisation; GaussNoise, Gaussian noise; ISONoise, random noise simulating high ISO (light sensitivity) settings.

4.3 Results

4.3.1 Evaluation of the synthetic carbonate core images generated by DCGAN

The DCGAN architecture failed to produce realistic samples after training for 1,000 epochs, where an epoch refers to a complete pass through the entire training dataset by the model during training. Despite indications of training stability with a discriminator loss of 0.689 and a generator loss of 0.738, the generated images lacked realism. Upon visual inspection, it was clear that the synthetic images produced were not representative of a carbonate texture (Figure 4.7). Although light and dark shades and potential shapes were present, there were no clear distinguishable textures that would enable class identification. It is also clear that some checkerboard artifacts remain (Odena *et al.*, 2016). The synthetic images produced were square images with sides of 64 x 64 pixels. Upscaling of these images to 256 x 256 pixels to match both the size of the original dataset and the CNN input size would have caused further loss of resolution. Therefore, it was decided not to include these images in any of the subsequent tests.

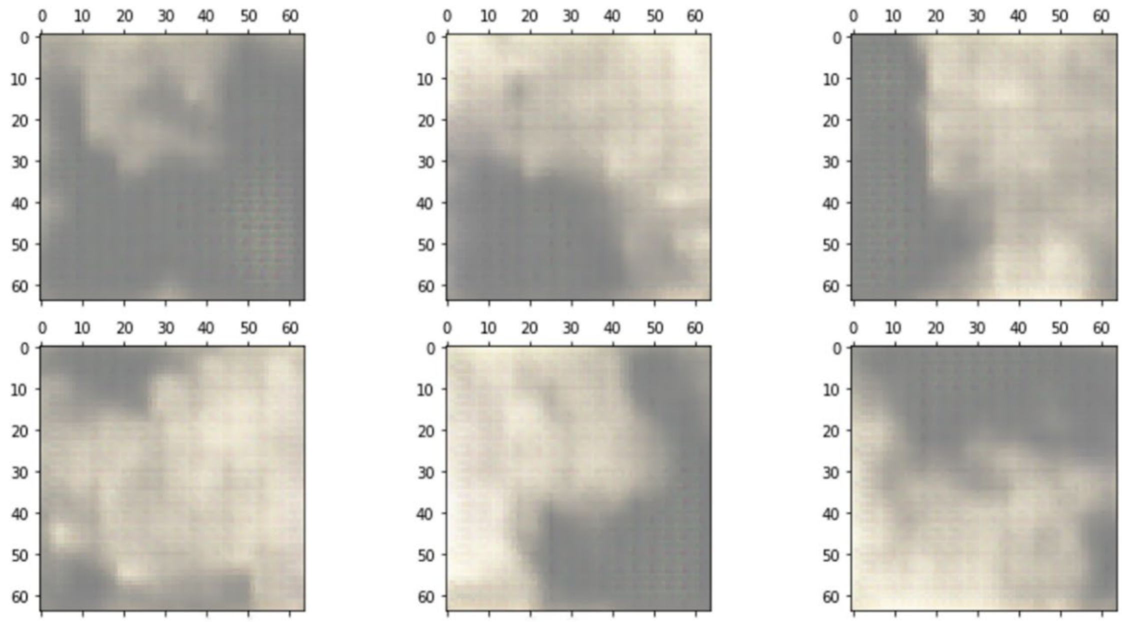

Figure 4.7: Examples of synthetic boundstone images generated by DCGAN after 1,000 epochs.

4.3.2 Evaluation of the synthetic carbonate core images generated by StyleGAN2

The average FID scores achieved by StyleGAN2 were 14.62, 10.58, 11.04, and 6.79, for Experiments 1 through 4, respectively (Table 4.1). These FID scores, while dataset-specific, are considered a good benchmark for this study. As FID evaluations have not been previously conducted on a carbonate core dataset, these results establish an initial standard for future comparisons. For general context, a FID score below 20 is often regarded as strong performance on complex datasets (Lucic *et al.*, 2018). Thus, the scores achieved, particularly the 6.79 observed in Experiment 4, suggest that the generated images closely resemble real carbonate textures.

The results achieved demonstrate that transfer learning and data augmentation were both effective tools for reducing overfitting on a limited geological data. Individually, each method improved upon the FID score of the Baseline (1) experiment by around 25-28% (95% confidence). The most significant improvement to the FID score, a reduction of ~53% (95% confidence), was achieved when transfer learning and augmentations (Experiment 4) were combined (Table 4.1). These reductions in FID scores indicate enhanced realism and diversity in the generated images. Table 4.1 highlights a clear correlation between lower FID scores and higher average precision and recall values. This suggests that as the FID score decreases, synthetic images become increasingly easier for the CNN classifier to interpret correctly. This indicates that the out-of-the-box StyleGAN2 architecture can produce realistic, diverse geological core images and achieve stable performance without extensive hyperparameter tuning. Additionally, data augmentation and transfer learning were seen to have helped reduce the generator loss and stabilise training (Figure 4.8).

Generator loss measures how well the generator in a GAN performs at creating realistic images that can deceive the discriminator. In a GAN, the generator attempts to create synthetic images, while the discriminator attempts to distinguish between real and synthetic images. The loss value reflects how well these objectives are balanced: lower generator loss indicates that the generator is producing more convincing images. Generally, loss in machine learning refers to a function that quantifies the difference between the predicted output and the desired output. In the context of GANs, generator loss decreases as the quality of synthetic images improves, although excessively low loss could indicate overfitting or mode collapse, where the generator produces only limited variations of a single image.

Figure 4.8 illustrates the average generator loss during training for the StyleGAN2 model under the four experimental setups. The Baseline configuration (Experiment 1) exhibits the highest and most unstable generator loss throughout training, indicating limited optimisation and difficulties in improving generator performance. In contrast, the Pre-trained setup (Experiment 2) starts with a lower initial loss due to the pre-trained weights and maintains consistently lower loss values compared to the baseline, demonstrating the benefits of pre-training. The Augmented configuration (Experiment 3) further reduces the generator loss, suggesting that data augmentation improves the quality of the generated samples and enhances training stability. Notably, the Pre-trained and Augmented setup (Experiment 4) achieves the lowest and most stable loss values across all iterations. This demonstrates the effectiveness of combining pre-training and augmentation in facilitating convergence and improving the ability of the generator to approximate the real data distribution. Overall, these results highlight the progressive improvements achieved through augmentation, pre-training, and their combination, with the Pre-trained and Augmented (Experiment 4) configuration being the most effective.

From the results in Table 4.1 and Figure 4.8, it is clear that the pretrained StyleGAN2-ADA produces the best results for synthetic core image generation. Visual inspection of the generated images showed reduced noise and greater accuracy in replicating fine textural details (Figure 4.9). Visualising the images also helps to show if and where mode collapse has occured, whereby the generator produces only limited variations of a single image. In our case, we see that a diverse selection of carbonate textures has been produced (Figures 4.9-4.13).

Table 4.1: Performance evaluation of the StyleGAN2 models. Precision and recall values were based on the classification results of the Inception-v3 classifier model trained on the 104k dataset from Dawson *et al.*, 2023. The numbers given in bold represent the best result for each evaluation metric when compared to the other models. Data bars are added for each metric to visually highlight the trend between lower FID scores and higher precision and recall values, illustrating the relationship between improved synthetic image quality and better classification performance.

Experiment	FID	Precision	Recall
Baseline (1)	14.62	0.59	0.54
Pre-trained (2)	10.58	0.79	0.72
Augmented (3)	11.04	0.76	0.68
Pre-trained + Augmented (4)	**6.79**	**0.82**	**0.86**

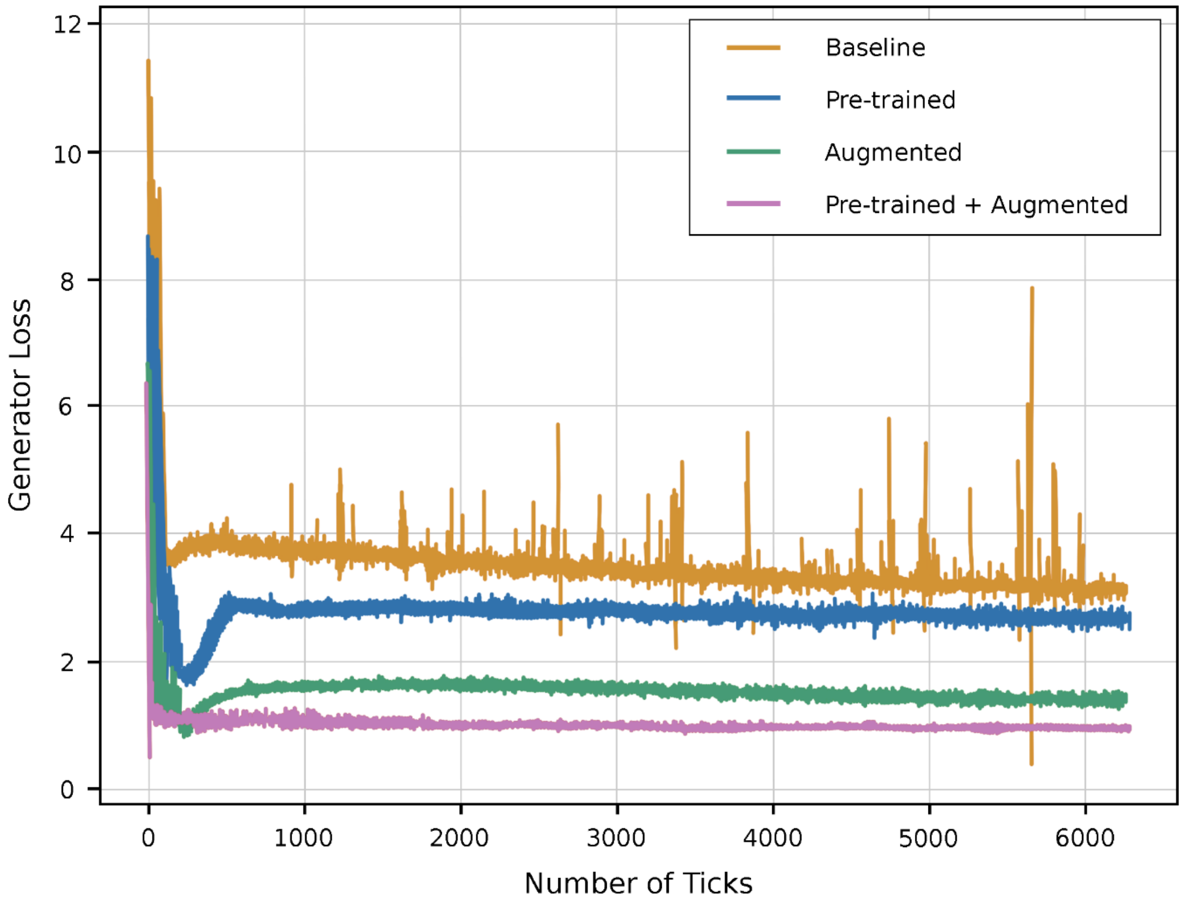

Figure 4.8: Average generator loss across training for StyleGAN2 under four experimental setups: Baseline, Pre-trained, Augmented, and Pre-trained + Augmented. The Pre-trained + Augmented setup demonstrates the lowest and most stable generator loss, highlighting the effectiveness of combining pre-training and data augmentation in improving model performance.

We note, however, the presence of artefacts in some images: some have line artefacts (Figure 4.10), some contain background features not present in the original image (Figure 4.11), some wet samples have generated line scanner reflections (Figure 4.12), and some reconstructions contain generated objects (Figure 4.13). These artefacts may arise from the generator misinterpreting certain features during training or attempting to generalise patterns in the data. Although these additional generated artefacts and objects may represent background images that may occur in a dataset, it is not clear if these would confuse a classifier based on textural divisions. It was decided to include these images in the augmented dataset for the subsequent experiments in Section 4.3.2, following the hypothesis that the presence of additional artefacts might enhance model robustness by exposing it to a wider range of potential image variability. By doing so, the ability of the model to generalise to real-world geological data, which may include imperfections or extraneous features, is likely to improve. This approach underscores the importance of balancing synthetic image fidelity with practical considerations for training robust classifiers.

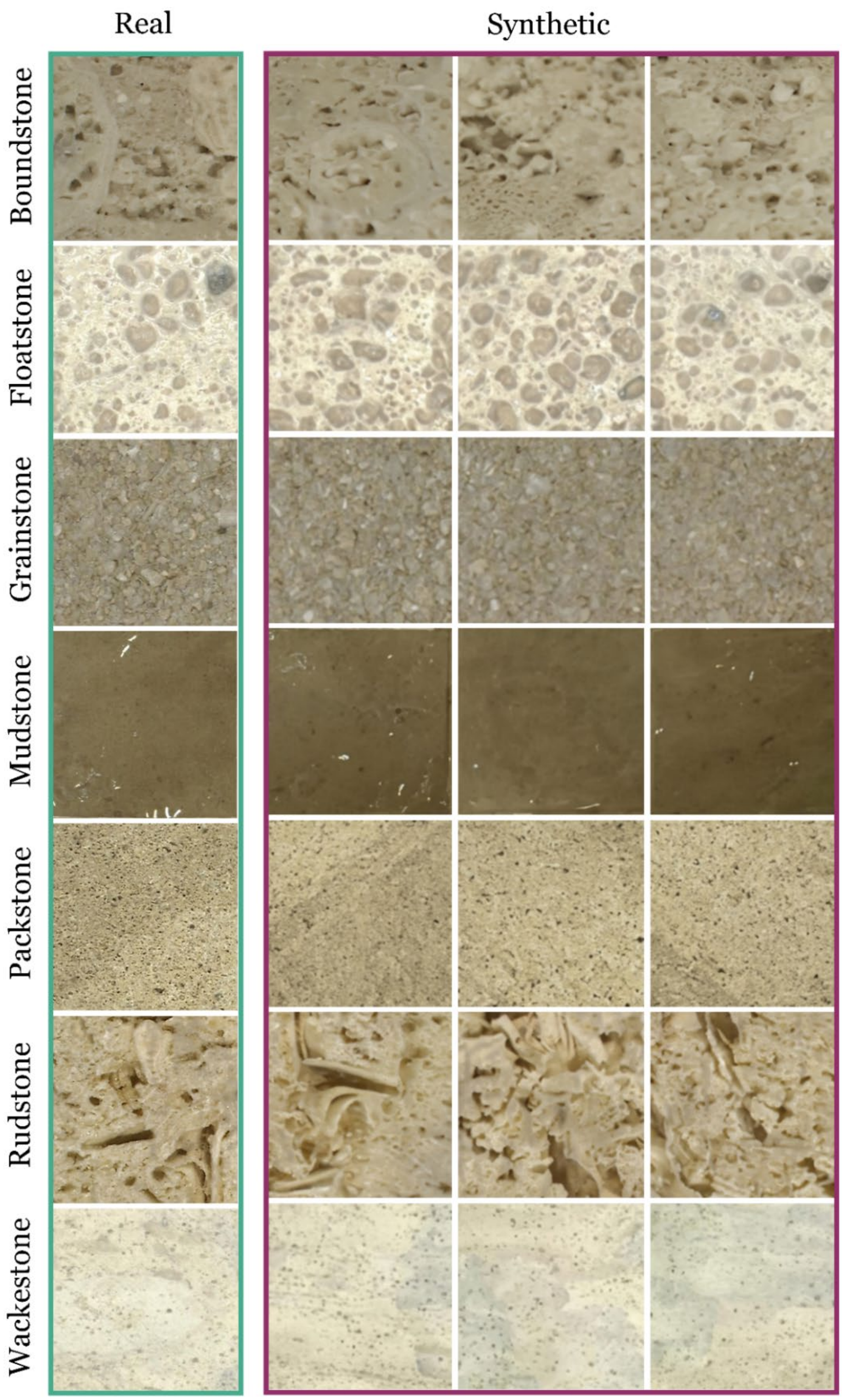

Figure 4.9: Real and synthetic images of carbonate textures. First column depicts real images from the original dataset. Second, third and fourth columns correspond to synthetic images generated by the GAN. Original images were taken from the following cores: Boundstone, 133-816A-15X-1; Floatstone, 359-U1465C-24X-1; Grainstone, 359-U1465A-4H-4; Mudstone, 359-U1466A-24X-2; Packstone, 133-815A-46X-1; Rudstone, 133-816B-2R-1; Wackestone, 359-U1471E-41R-2.

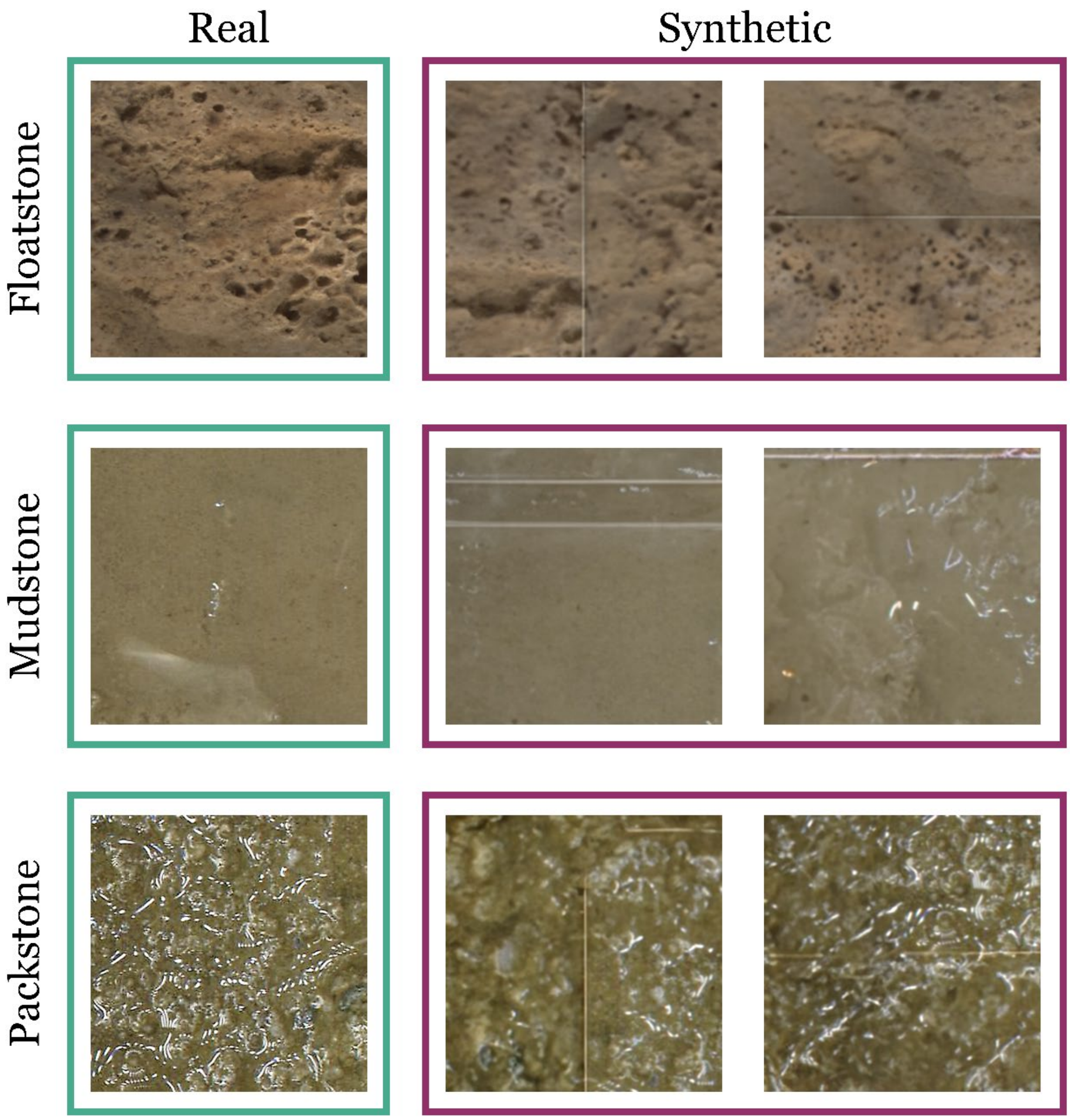

Figure 4.10: Generated images with line artifacts. Left images in the green boxes show original core textures, images in the purple boxes to the right show synthetic images. Original images are taken from the following cores: Floatstone, 133-812B-10R-1; Mudstone, 359-U1466A-24X-2; Packstone, 359-U1471A-2H-1.

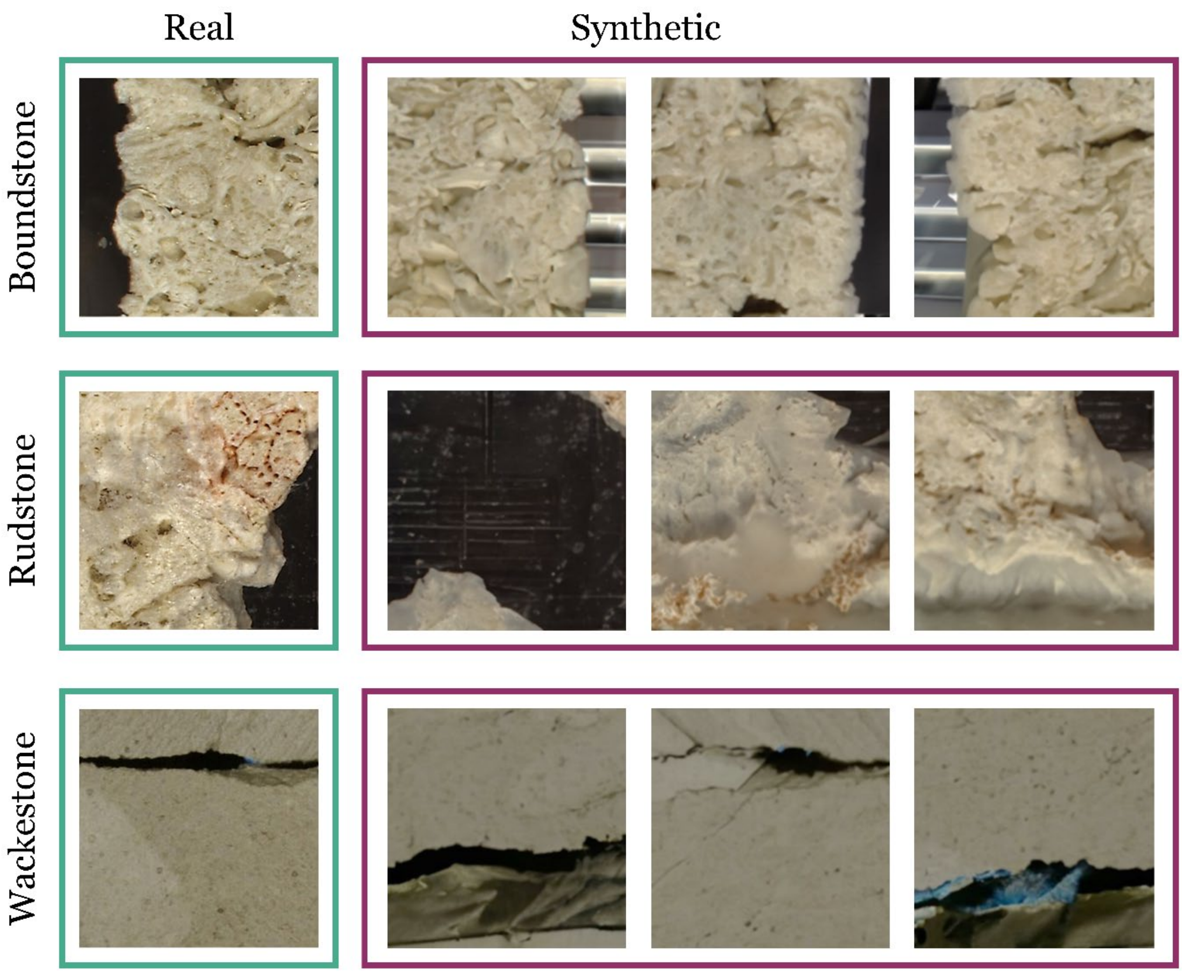

Figure 4.11: Generated images with background features. Left images in the green boxes show original core textures, images in the purple boxes to the right show synthetic images. Synthetic boundstone images show examples of generated background features. Synthetic rudstone images shows examples of generated lines in the background/core liner. Synthetic wackestone images show examples of generated core liner. Original images are taken from the following cores: Boundstone, 133-816C-1R-4; Rudstone, 133-816B-2R-1; Wackestone, 359-U1471E-16R-1.

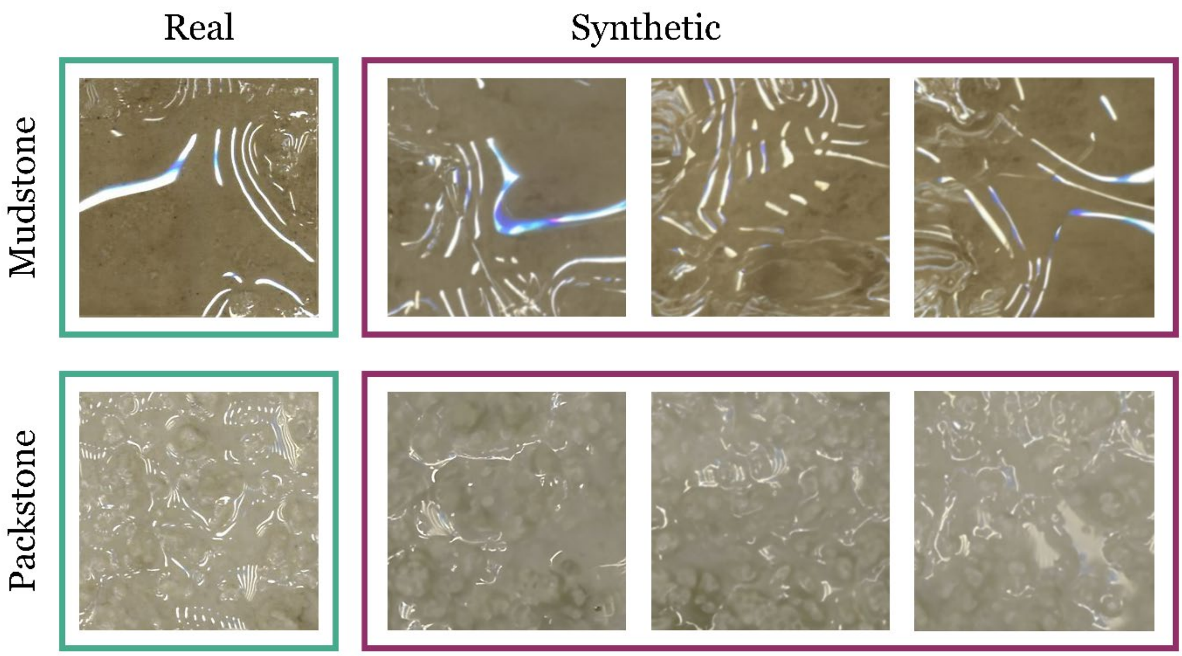

Figure 4.12: Generated images of wet samples with line scanner reflections. Left images in the green boxes show original core textures, images in the purple boxes to the right show synthetic images. Original images are taken from the following cores: Mudstone, 359-U1466A-24X-2; Packstone, 359-U1468-29F-2.

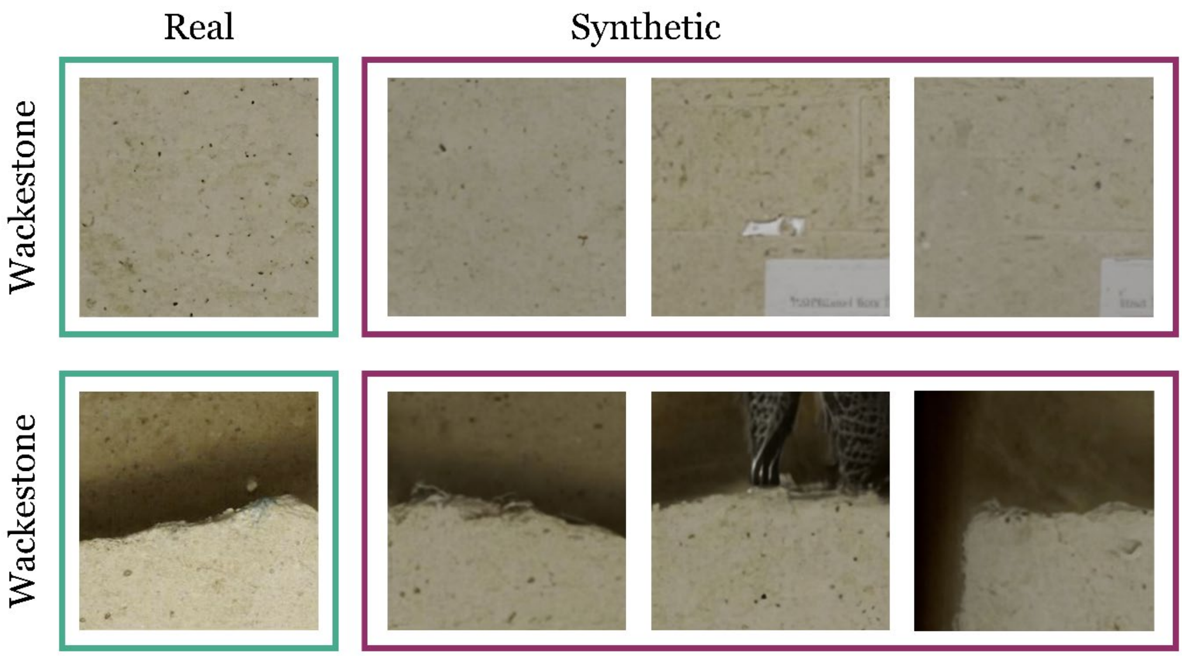

Figure 4.13: Generated images containing objects. Left images in the green boxes show original core textures, images in the purple boxes to the right show synthetic images. The top row show examples of generated labels along with artefact lines. The bottom row shows an example of a generated mesh-like texture. Original images are taken from the following cores: Wackestone, 359-U1471E-16R-1.

4.3.3 Comparative evaluation of data augmentation techniques

We compared the ability of the synthetic data generated using StyleGAN2-ADA and traditional augmentation methods to improve CNN classification performance. The Inception-v3 network was trained on the original 7k dataset, to provide a baseline for classification performance.

To assess the effect of GAN-based augmentation, synthetic core images were generated using StyleGAN2-ADA and combined with the original, creating a mixed synthetic dataset of 22,400 training images (Model 3, Table 4.2). For the traditional augmentation approach, basic image augmentation techniques were applied during the preprocessing step of training, expanding the dataset to 168,000 augmented images (Model 2, Table 4.2). Both approaches significantly enhanced image diversity o while preserving essential features and details, leading to improved CNN classification performance when used for training (Table 4.2).

Traditional augmentation (Model 2) resulted in a 4.2% increase in sensitivity and an 8.8% increase in specificity, compared to the baseline. Using the GAN-based augmentation (Model 3) yielded even greater improvements, with sensitivity increasing by 9.5% and specificity by 12.7% (Table 4.2). These improvements are significant, as these relatively large percentage increases in sensitivity and specificity are considered noteworthy in the context of deep learning, where performance gains can often be marginal at best. Despite the smaller dataset compared to Model 2, Model 3 achieves the highest sensitivity (76.7%) and specificity (78.1%), demonstrating the utility of high-quality synthetic data in improving classification tasks. These results suggest that the synthetic samples generated by StyleGAN2-ADA provide greater diversity and realistic variability compared to traditional augmentation, leading to better model performance even with fewer training images.

Table 4.2: Performance evaluation of classification models trained on augmented datasets. Model 1 is trained on the original dataset, Model 2 is trained on the original data + traditional augmentation, Model 3 is trained on original dataset + synthetic data (generated by StyleGAN2-ADA).

Classification Model	Training Images	Sensitivity (%)	Specificity (%)
Model 1	5,600	67.2	65.4
Model 2	168,000	71.4	74.2
Model 3	22,400	76.7	78.1

The training curves (Figures 4.14 and 4.15) provide a performance comparison between models trained on the original dataset and the dataset augmented with synthetic images. Models trained solely on the original dataset demonstrate a clear gap between training and test curves in both accuracy and loss plots, indicating a tendency to overfit as the model performs significantly better on the training data than the test data. In contrast, models trained with the original dataset augmented with synthetic images achieve higher test accuracy, lower test loss, and exhibit a reduced gap between training and test curves in both plots. This suggests that the addition of synthetic data improves the generalisation ability of the model, allowing it to perform more consistently across both training and unseen test data.

For the synthetic dataset, training accuracy and loss values are seen to be lower than those achieved in training using the original data. However, this is to be expected. The mixed synthetic dataset, with 22,400 images, contains significantly more examples than the original dataset of 7,000 images, introducing greater diversity and complexity in the data. This forces the model to generalise across a wider range of patterns and prevents it from simply memorising the training examples. The smaller original dataset allows for the model to more easily memorise specific patterns and overfit to the training data, resulting in higher training accuracy but at the expense of generalisation. Additionally, the inclusion of synthetic images generated by StyleGAN2-ADA introduces variations that might not exist in the original dataset, further challenging the model to adapt to a broader variety of features. As a result, the model may take longer to converge and exhibit higher training loss, reflecting the increased difficulty of fitting a more diverse dataset. While neither model has fully converged within the given epochs, the training trends suggest that extending the training duration with early stopping could further refine performance (Figures 4.14 and 4.15).

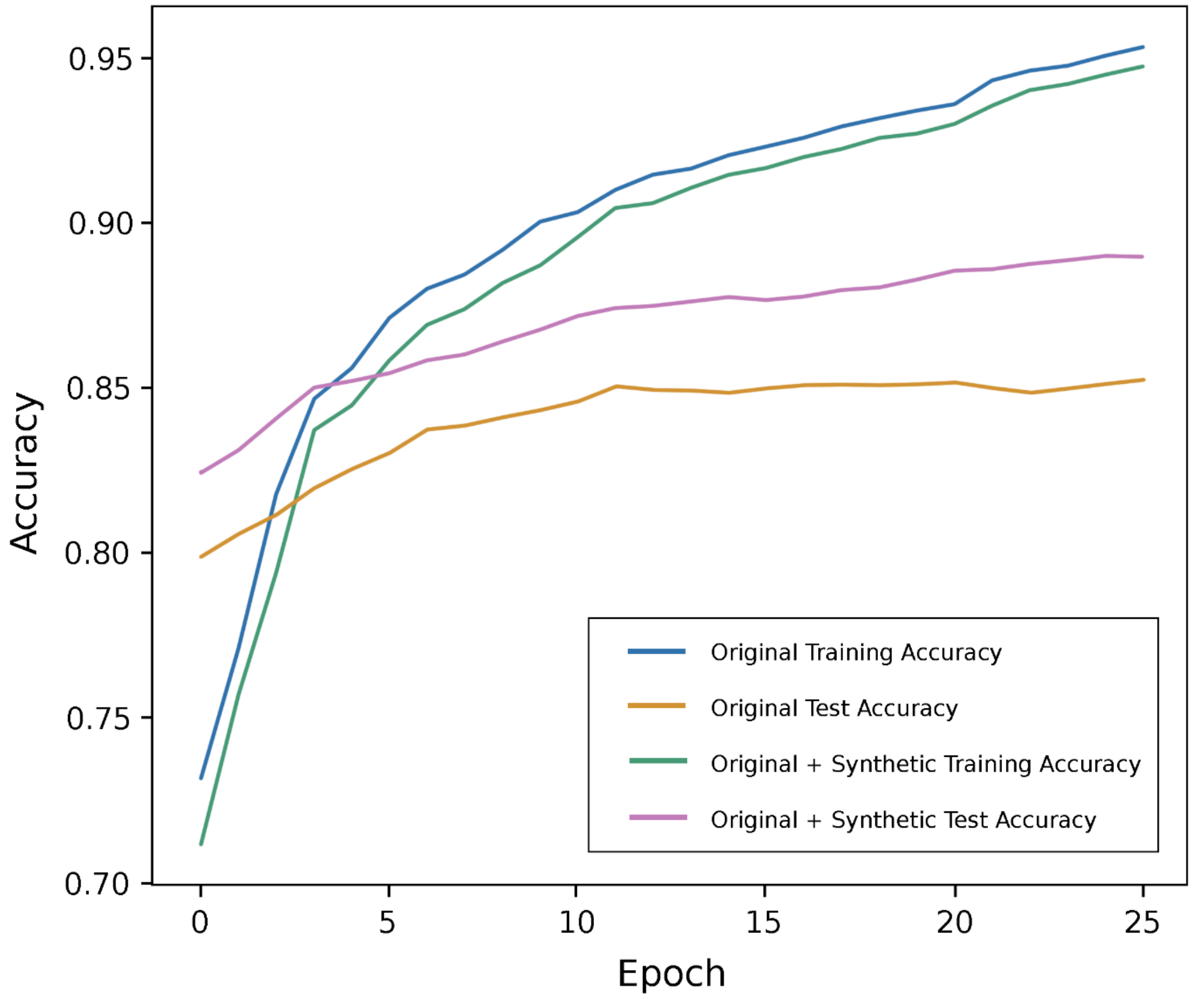

Figure 4.14: Accuracy learning curves over 25 epochs for models trained on the original dataset and the dataset augmented with synthetic images. Models trained with synthetic data demonstrate improved test accuracy and reduced overfitting compared to models trained solely on the original dataset.

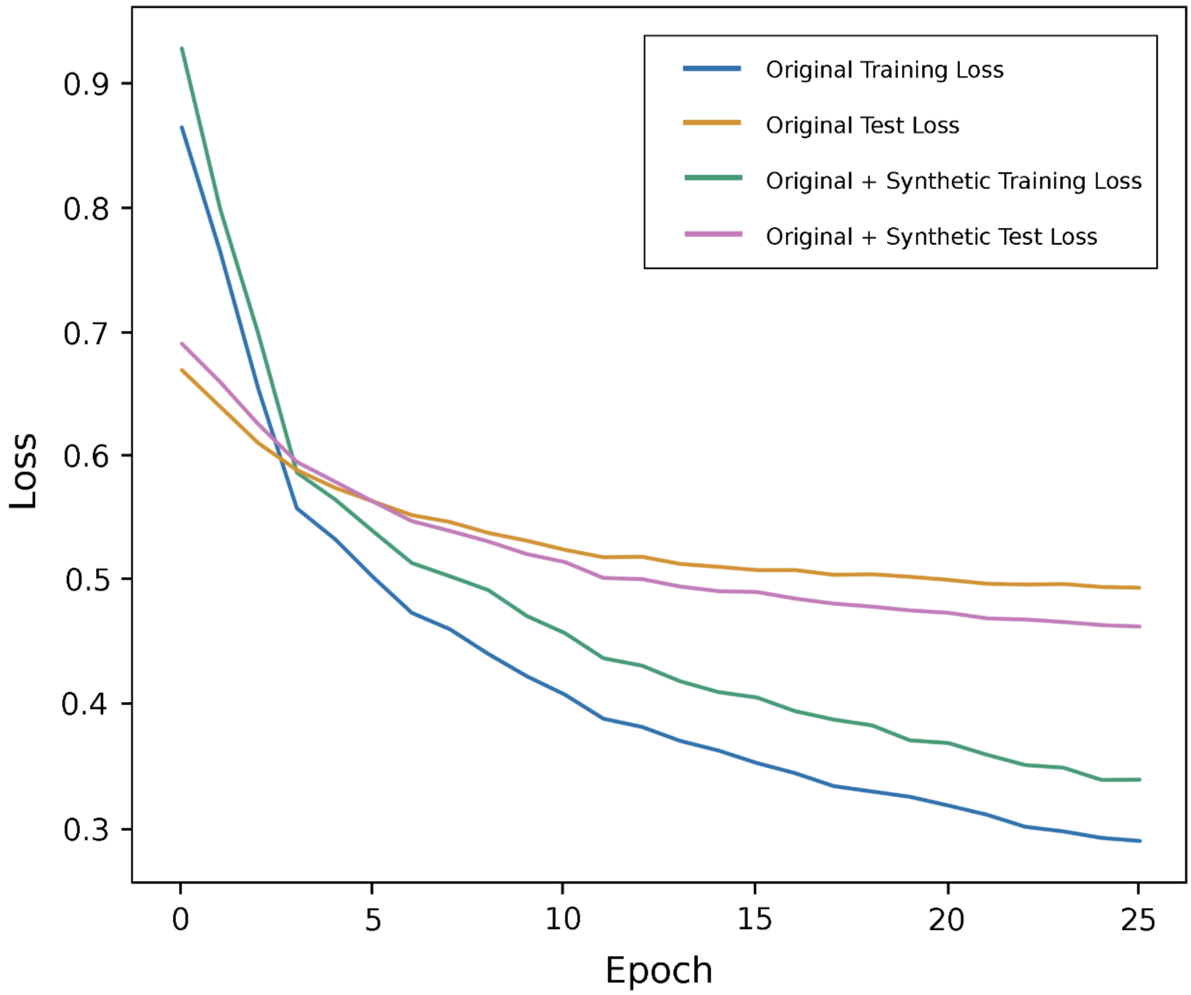

Figure 4.15: Loss learning curves over 25 epochs for models trained on the original dataset and the dataset augmented with synthetic images. Models trained with synthetic data demonstrate reduced test loss and decreased overfitting compared to models trained solely on the original dataset.

4.4 Discussion

4.4.1 Leveraging GANs for geological image augmentation

Since their introduction in 2014, GANs have gained prominence across a variety of domains, most notably in the synthesis of images, videos, and text. Their capacity in generating realistic and diverse instances of data has positioned them as a valuable tool in research for advancing machine learning and artificial intelligence. In the context of geological research, GANs hold particular promise due to the scarcity of high-quality, labelled geological images for training and analysis. This study explored the feasibility of using GANs to generate synthetic core images, as well as assessing their quality and utility in improving classification model performance.

To address the challenge posed by limited datasets for textural carbonate core classification models, the StyleGAN2-ADA network was employed to generate synthetic images representing Dunham textures. Separate StyleGAN2-ADA models were trained for each Dunham texture class, enabling the creation of a more extensive and diverse dataset compared to the original. Visual comparisons between the real and synthetic images (Figure 4.9) demonstrate that the GAN-generated images are of acceptable quality, , making them credible substitutes for augmenting the original dataset. While these synthetic images appear realistic to the human eye, additional evaluation metrics confirmed their effectiveness as training data for carbonate classification models (Figures 4.14 and 4.15). For each Dunham class, a separate StyleGAN2-ADA model was trained using the real images of that class. Overall, the use of StyleGAN2-ADA enabled us to create a bigger and diverse dataset by augmenting the original dataset. This approach demonstrated a clear potential to train more robust geological classification models compared to models trained exclusively on the original small dataset.

Unlike traditional augmentation methods, which rely on geometric and photometric transformations to modify existing images, GAN models like StyleGAN2-ADAlearn the underlying statistical distribution of the training data. In our case, this means the GAN tries to understand the patterns, features, and relationships in the images of carbonate cores it is trained on. In doing so, the GAN captures what makes each type of Dunham texture unique. This includes the visual characteristics such as the arrangement, shape, and colour patterns that define each textural class. Once the GAN understands these patterns, it can generate new artificial core images that look like they belong to the equivalent Dunham class, even though they are entirely synthetic. Traditional image augmentations, while proven effective in enhancing dataset size and variability, are inherently constrained by the transformations applied and cannot match the diversity introduced by GANs (Frid-Adar et al., 2018; Shorten and Khoshgoftaar, 2019).

The results in Table 4.2 demonstrate that the classification model trained on the GAN-augmented dataset (Model 3) outperformed the model trained on the dataset expanded through traditional augmentation techniques (Model 2), even though the latter contained significantly more images. This highlights the quality and diversity of the images generated by StyleGAN2-ADA, which contributed to a significant improvement in CNN sensitivity and specificity. Notably, sensitivity increased by 9.5% and specificity by 12.7%, both of which are substantial gains in the context of deep learning, where even small percentage gains are often challenging to achieve. These results underscore the impact of synthetic data on model performance and demonstrate the effectiveness of this augmentation approach.

The success of the StyleGAN2-ADA model can likely be attributed to its use of adaptive discriminator augmentation (ADA). This mechanism helps to prevent generator overfitting, a common limitation in GAN training, while maintaining high-quality image generation. The ADA technique allows StyleGAN2-ADA to generate realistic images even with small training datasets (Karras et al., 2020). This makes it a better fit for geological image generation compared to models like DCGAN, which failed to produce useful data augmentations in our study. Consequently, StyleGAN2-ADA emerges as the more suitable choice for geological image generation, effectively addressing the challenges posed by small datasets. The workflow presented here provides a foundation for generating geological datasets. By leveraging advanced techniques, such as style-mixing within the StyleGAN2 models, it could be possible to produce more tailored generative models for specific applications, such as the synthesis of carbonate constituents or other geologically relevant features. Style-mixing works by blending features from multiple latent codes during the generation process, allowing for greater diversity and control in the synthetic images produced. In a GAN, a latent code is essentially a blueprint that determines what a generated image will look like. By combining latent codes at different stages of the generation process, style-mixing allows the GAN to merge characteristics from multiple sources. For instance, one latent code might influence the overall structure of the image, such as the general arrangement of textures, while another might affect finer details, like the appearance of specific patterns or colours. This blending results in images that combine features from different training examples in unique ways (Angermann *et al.,* 2024). This opens new avenues for developing high-quality synthetic datasets for geological research.

4.4.2 Limitations of DCGAN for geological data augmentation

The failure of DCGAN to produce useful data augmentations in this study highlights several key limitations of this early GAN architecture. While DCGAN was a pioneering architecture in GAN development, its design lacks many of the advanced techniques that newer GAN architectures, such as StyleGAN2, employ to address common challenges in image synthesis.

One critical shortcoming of DCGAN is its tendency to overfit when trained on small datasets. In this study, the available dataset of carbonate textures was relatively limited, and without adaptive mechanisms to counteract overfitting, the discriminator in the DCGAN framework likely became overly powerful, quickly learning the training data and leaving the generator unable to improve (Kim and Park, 2023). This imbalance would have made it difficult for the generator to learn effectively, resulting in synthetic images that lacked diversity or realism.

Furthermore, DCGAN is prone to mode collapse, a phenomenon where the generator repeatedly produces only a narrow range of outputs, failing to capture the full statistical diversity present in the original dataset (Jain *et al.*, 2022). For geological datasets, such as the carbonate textures studied here, capturing a wide range of variations is critical to creating a robust and useful synthetically augmented dataset. The inability of DCGAN to produce such diversity further limits its utility in this context.

In addition to these challenges, DCGAN is also known to produce checkerboard artefacts (Figure 4.7), a problem arising from uneven overlap of kernel strides during up-sampling in the generator (Odena *et al.*, 2016). These artefacts compromise and degrade the visual quality of the synthetic images, reducing their utility as reliable augmentations and making them less suitable for downstream applications, such as training CNN classifiers. Unlike newer architectures, DCGAN does not include mechanisms to address or minimise these artefacts, further reducing its effectiveness for generating high-quality synthetic data.

The architectural simplicity of DCGAN limits its ability to handle the complex and nuanced features present in geological datasets. Additionally, DCGAN lacks the fine-grained control over image generation found in more modern architectures, such as StyleGAN2 (Yates *et al.*, 2022). Without the ability to separately manipulate elements like texture, structure, and style, DCGAN struggles to capture and replicate the features characteristic of geological images. The absence of progressive growing or other stabilising techniques further compounds these issues, leading to inconsistent results during training. This limitation becomes particularly problematic when trying to produce realistic and detailed augmentations of complex textures like those seen in carbonate cores.

In contrast, StyleGAN2-ADA incorporates features specifically designed to address these challenges, including mechanisms to reduce overfitting, improve diversity, and control image quality. The failure of DCGAN in this study serves as a reminder of the ongoing evolution in GAN architectures and the critical need to select models tailored to the specific demands of the dataset and application.

4.4.3 Limitations of StyleGAN2 for geological data augmentation

While StyleGAN2-ADA demonstrated higher quality performance compared to DCGAN and traditional augmentation techniques in generating synthetic carbonate textures, it is not without limitations. A key issue observed in this study was the presence of artefacts in the generated images (Figures 4.10-4.13). These artefacts, such as line artefacts, additional background features not present in the original dataset, and scanner-like reflections in wet samples, raise questions about the fidelity of the synthetic images and their ability to accurately represent carbonate textures.

The generated artefacts likely stem from difficulties in learning the intricate statistical distribution of the training data. Geological textures, such as those in carbonate cores, are highly complex and nuanced, often requiring fine-grained modelling of features like grain structure, porosity, and cementation. While StyleGAN2 does well to capture much of this complexity, it may inadvertently introduce patterns or distortions that are not representative of real-world geological samples. The presence of these artefacts could potentially impact on downstream tasks, such as training classification models. Although including artefact-laden images may improve model robustness by exposing classifiers to a wider range of variations, there is also a risk that the presence of unrealistic features could bias the classifier or reduce its sensitivity to genuine textural distinctions.

Furthermore, the relatively small training dataset may have also contributed to the limitations observed here. Even with adaptive discriminator augmentation (ADA) to mitigate overfitting, the quality and diversity of the generated images are inherently tied to the richness of the training data. A limited dataset constrains the ability of the model to generalise and increases the likelihood of generating artefacts or potentially missing certain textural variations altogether.

Finally, the computational demands of StyleGAN2 represent a practical limitation. Training the model to generate high-resolution geological images requires significant resources, including extensive GPU memory and processing power. This constraint may limit accessibility to researchers.

4.4.4 Recommendations for future work

The success of this study demonstrates the capacity of GAN models, such as StyleGAN2-ADA, as tools for generating synthetic datasets for use with other machine learning algorithms in geological research workflows. While these results show significant improvements in classification performance, future research should aim to validate and refine this approach further. Evaluating alternative generative architectures, including more recent models like StyleGAN3 (Karras et al., 2021), could yield even higher-quality synthetic datasets while addressing challenges such as generated artefacts. These artefacts, including line patterns, extraneous reflections, and unrealistic background features, suggest the need for improved architectural design or post-processing techniques to further enhance the fidelity of synthetic data.

Building on the foundation established here, future studies could explore using the pre-trained weights of this model to develop more domain-specific dataset generators. Such models could focus on creating datasets tailored to unique geological features, enabling more specialised applications. Feature modification of core images and exploration of latent space are exciting avenues for visualising and analysing geological variables, such as changes in composition or texture, by manipulating specific attributes. This approach could further demonstrate the flexibility and utility of GANs in geoscience workflows.

The incorporation of emerging technologies, such as image-to-image translation, provides another promising direction. For instance, transforming FMS images into synthetic core images could help to further bridge gaps in data availability. The use of super-resolution techniques (Niu et al., 2020) could up-sample lower-resolution core datasets available from other sources, making them more suitable for modern machine learning applications. Moreover, associating latent vectors with well-log and geochemical data could expand the practical applications of GANs, allowing for the generation of core images based on compositional or environmental variables. This capability could be particularly valuable in reservoir analysis, enabling synthetic image generation that aligns closely with real-world data trends.

Finally, enhancing the explainability of GANs is crucial for understanding how these models organise and structure data within latent space. This knowledge could offer insights into the geological features encoded within the synthetic data and guide more targeted model development. As GAN technology evolves, these advancements could unlock new possibilities for augmenting datasets and improving the performance of machine learning algorithms in geoscience.

4.5 Conclusions

This study explored the application of generative adversarial networks (GANs) for generating synthetic carbonate core images based on real-world examples. We demonstrated that the synthetic images generated by the StyleGAN2-ADA architecture closely replicate key characteristic features of carbonate textures, as confirmed through evaluations using Fréchet Inception Distance (FID) scores, visual analysis, and validation with a trained classification model. These results highlight the potential of GAN-generated datasets to supplement limited geological image datasets, ultimately improving classification model performance.

The results are encouraging, particularly given the substantial improvements in sensitivity and specificity observed when using GAN-augmented datasets. The ability to generate realistic and diverse images efficiently from a trained model opens exciting opportunities for advancing geological research and improving the robustness of machine learning models. While the initial training process is computationally demanding, the efficient reuse of trained models can help to ensure long-term practicality.

However, challenges remain, particularly regarding the presence of artefacts in some generated images, such as line artefacts, scanner-like reflections, and non-representative background features. These artefacts, while not detrimental to the overall success of the method, highlight areas for improvement to further enhance the fidelity and utility of synthetic datasets. Addressing these limitations could involve exploring alternative architectures, such as StyleGAN3 or diffusion models, that are specifically designed to mitigate artefacts and improve image quality.

Future work should build on these promising results by exploring advanced techniques to further refine the generation process. This could include incorporating grayscale and multi-channel training images, leveraging style-mixing for greater diversity, and investigating the latent space of the GAN model to control and modify specific features in synthetic images. These enhancements could improve the realism of generated images and also provide a powerful tool for visualising and analysing geological variables.

As GAN research continues to advance, developments in training stability and consistency are expected, which will further improve the reliability and applicability of these models. The strong results achieved in this study provide a solid foundation for future work and demonstrate the potential for GAN-generated datasets to transform geological research and associated machine learning applications.

Chapter 5: Object Detection Algorithms to Identify Skeletal Components in Carbonate Cores[10,11]

ABSTRACT

Identification of constituent grains in carbonate rocks requires specialist experience to distinguish between skeletal grains that change through geological ages, preserved in differing alteration stages, and cut in random orientations across core sections. Recent studies have demonstrated the effectiveness of machine learning in classifying lithofacies from thin section, core, and seismic images, with faster analysis times and reduction of natural biases. In this study, we explore the application and limitations of convolutional neural network (CNN) based object detection frameworks to identify and quantify multiple types of carbonate grains within close-up core images of carbonate lithologies. We compiled nearly 400 images of high-resolution core images from three ODP and IODP expeditions. Over 9,000 individual carbonate components of 11 different classes were manually labelled from this dataset. Using pre-trained weights, a transfer learning approach was applied to evaluate one-stage (YOLO v5) and two-stage (Faster R-CNN) detectors under different feature extractors (CSP-Darknet53 and ResNet50-FPN, respectively). Despite the current popularity of one-stage detectors, our results show Faster R-CNN with ResNet50-FPN backbone provides the most robust performance, achieving 0.73 mean average precision (mAP). Furthermore, we extend the approach by deploying the trained model to two ODP sites from Leg 194 that were not part of the training set (ODP Sites 1196 and 1199), providing a performance comparison with benchmark human interpretation.

5.1 Introduction

Core data play a fundamental part in the reconstruction of past geologic, climatic, and biological conditions which, in turn, further our understanding of the Earth's past, present and future. Most carbonate constituent grains are biogenic in origin and are usually formed close to their site of deposition. While the extent to which these grains are transported from their site of formation remains uncertain, they may

[10] A version of this chapter was presented at the **EGU General Assembly 2023**:
Dawson, H.L, John, C.M. (2023) Deep learning based identification of carbonate rock components in core images. EGU General Assembly 2023, Vienna, Austria, 24-28 Apr 2023, EGU23-2388. doi: 10.5194/egusphere-egu23-2388.
[11] A version of this chapter has been published in **Marine and Petroleum Geology**:
Dawson, H.L, John, C.M. (2024) Object Detection Algorithms to Identify Skeletal Components in Carbonate Cores. Marine and Petroleum Geology, 106965. doi: 10.1016/j.marpetgeo.2024.106965.

provide ecological information about their formative environment, as well as biostratigraphic information (Cook *et al.*, 1983). Characterisation of these framework components in carbonate rocks can greatly enhance core observations through more detailed classifications, aiding in the interpretation of depositional environments and marine ecosystem evolution, and the establishment of complex post-depositional and diagenetic histories (Memon *et al.*, 2023). This is important for a wide range of industrial and academic applications including, but not limited to, energy and mineral resource exploration, global carbonate cycling, and palaeoenvironmental reconstruction (e.g., Flügel, 2004; Ridgwell and Zeebe, 2005; Payne *et al.*, 2006; Utami *et al.*, 2018; Al-Ramadan *et al.*, 2019; Koeshidayatullah *et al.*, 2020b).

Carbonate sedimentology is complicated and identifying constituents requires specialist experience that is often done on a qualitative rather than quantitative basis, given the time requirements. Difficulty in the identification of carbonate skeletal grains arises from a multitude of factors, largely due to the diverse nature of carbonate-producing organisms with a wide variety of shell morphologies, which are cut in random orientations across core sections (Tucker and Wright, 2009). The assemblages of these organisms change with environment as well as with geologic time (Figure 5.1). Where the types of grains change but not the energy of the environment of deposition, this may lead to co-evolution in shell shape and carbonate textures. This is further complicated by diagenetic alteration which, due to the presence of unstable aragonite and high-Mg calcite, is common and extensive in many carbonate grains (Scholle and Ulmer-Scholle, 2003). Collectively, these factors highlight the intricacies involved in differentiating and characterising carbonate components; a task that is typically labour-intensive, costly, and requires a substantial in-depth knowledge of carbonate sedimentology and palaeontology. The qualitative nature of this approach, and reliance on human interpretation, introduces the potential for bias and subjectivity, in turn limiting consistency and reproducibility (Lokier and Al Junaibi, 2016).

Recent studies have leveraged computer vision algorithms, particularly convolutional neural networks (CNNs), for automated descriptions of cores, thin sections, well logs, and seismic data (e.g., Huang *et al.*, 2017b; John and Kanagandran, 2019; Deng *et al.*, 2021; Chai *et al.*, 2022; Zheng *et al.*, 2022; Dawson *et al.*, 2023). Previous efforts have primarily focused on machine learning applications for lithofacies classification, such as that presented in Chapter 3 (Dawson *et al.*, 2023), with few attempts at the identification and characterisation of grains or features within digital rock images. Several authors have presented object detection-based approaches towards recognising microfossils and cement grains and calculating porosity in thin sections (Koeshidayatullah *et al.*, 2020a; Yu *et al.*, 2021; Bonté *et al.*, 2022; Bouziat *et al.*, 2022). In other fields outside of geoscience, such technologies have been an important step towards autonomous driving systems (e.g., Khan *et al.*, 2023), crime prevention (e.g., Saikia *et al.*, 2017), planetary surface mapping (e.g. Bickel *et al.*, 2020; Rubanenko *et al.*, 2021), and disaster response (e.g., Pi *et al.*, 2020). The efficacy of this methodology for core component analysis, however, remains to be tested.

In this chapter, we develop an object detection model for simultaneously locating and identifying multiple types of constituent grains within carbonate core images. Such application, when combined with the automated textural classification presented in Chapter 3, could more closely imitate human interpretations by providing textural classes along with qualifiers and descriptors. Using a transfer learning approach with pre-trained weights, the performances of one-stage and two-stage detectors under different feature extractors are evaluated to assess the most effective approach. Finally, we deploy the trained model to ODP Leg 194 Sites 1196 and 1199 (Isern *et al.*, 2002), thereby extending the approach to quantify the abundances of biogenic components and developing a performance comparison with human interpretation.

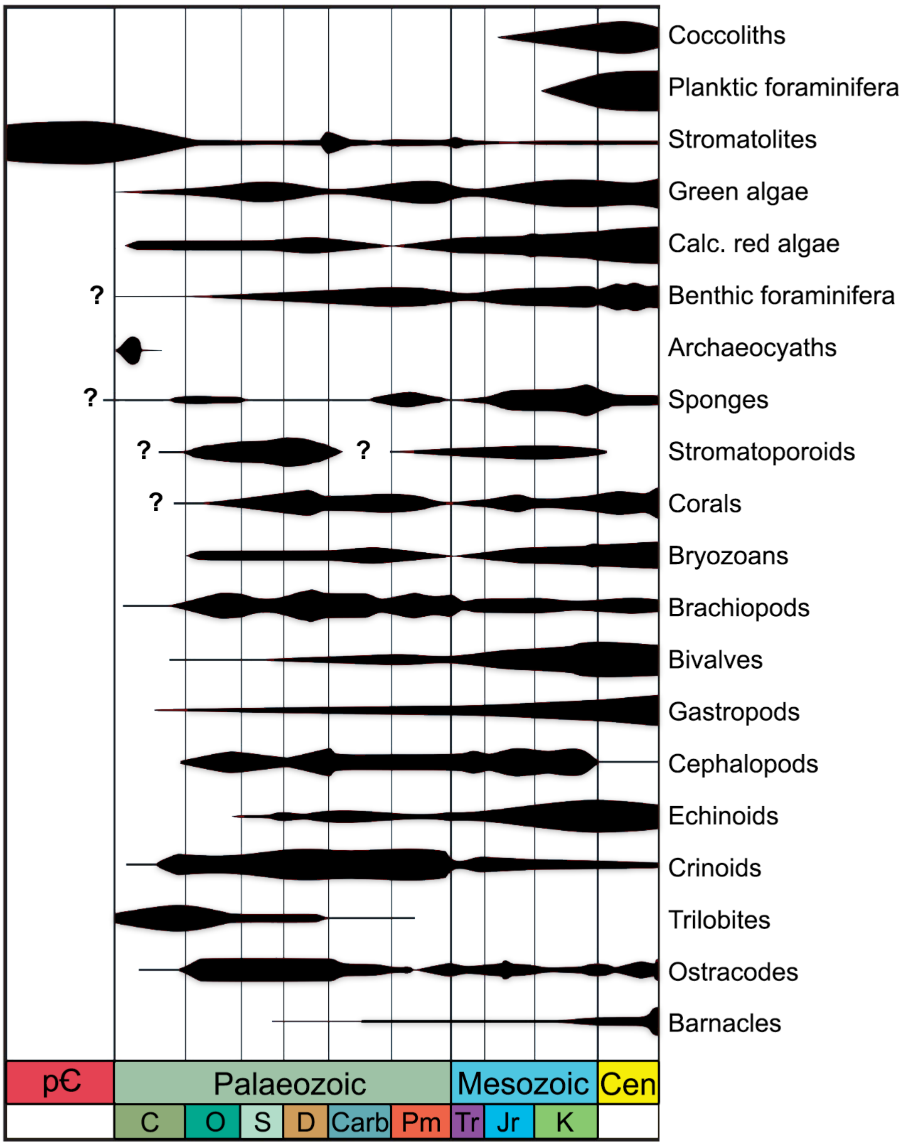

Figure 5.1: Summary of the major groups of marine carbonate-producing organisms showing fluctuations in approximate diversity and abundance through time (modified from Scholle and Ulmer-Scholle, 2003). Although based on approximations, it provides a generic overview of the types of organisms likely to occur in carbonate rocks of any age.

5.1.1 Background

With the rise of autonomous driving systems, intelligent healthcare monitoring, smart video surveillance, and anomaly detection, the demand for fast and accurate object detection models has grown significantly (e.g., Li *et al.*, 2018; Gupta *et al.*, 2021; Chen *et al.*, 2022; Lobo *et al.*, 2023). Modern deep learning algorithms have risen to meet this demand, offering high-quality image analysis capabilities, including image classification and object detection (Alzubaidi et al., 2021). These models allow for the identification and classification of multiple objects within a single image, as well as the localisation of each occurrence through the drawing of bounding boxes, which are defined by coordinates indicating the size and position of each object (Diwan *et al.*, 2023).

Object detection has significantly evolved over the past two decades, with progress typically separated into two distinct phases: traditional methods and deep learning-based methods (Table 5.1) (Kaur and Singh, 2023). For traditional methods, the object detection pipeline is usually divided into three stages (Zhao *et al.*, 2019):

1) **Region selection** – as the different objects within an image may occur in different positions with varying sizes and aspect ratios, the whole image is inspected using a multiscale sliding window approach. Despite proposing all possible object locations, the large number of candidate windows generated often results in many irrelevant regions and is computationally expensive.

2) **Feature extraction** – after identifying potential object regions through the sliding window approach, each region is analysed further to extract visual features that provide robust and meaningful representations., Techniques such as Haar-like (Lienhart and Maydt, 2002), Scale Invariant Feature Transform (SIFT) (Lowe, 2004), and Histogram of Oriented Gradients (HOG) detector (Dalal and Triggs, 2005) are commonly used. However, there is difficulty in creating a comprehensive feature descriptor that can correctly describe all objects, due to wide variations in appearance, perspective, lighting, and background conditions.

3) **Classification** – target objects need to be identified using a classifier, such as Support Vector Machine (SVM) (Cortes and Vapnik, 1995), AdaBoost (Freund and Schapire, 1997) or Deformable Part-based Model (DPM) (Felzenszwalb *et al.*, 2008). In doing so, the representations are made more hierarchical and informative for visual recognition.

The traditional algorithms presented in Table 5.1 represent important milestones in the history of object detection, each contributing novel approaches to address specific challenges. Viola-Jones detectors, Haar-like features, and HOG detectors were particularly influential in the early 2000s, providing effective solutions for real-time and accurate object detection. SIFT and Speeded Up Robust Features (SURF), on the other hand, focused on robust feature extraction, while DPM introduced deformable part-based modelling for improved object representation.

Table 5.1: Milestones of object detection algorithms over the past 20 years.

Publication	Algorithm	Phase	Milestone
Viola and Jones, 2001	Viola Jones (VJ) Detectors	Traditional	First real-time detection of human faces without any constraints, using a "cascade" architecture of increasingly more complex classifiers
Lienhart and Maydt, 2002	Haar-like	Traditional	Used rectangular filters to compute feature responses based on differences between adjacent rectangular regions.
Lowe, 2004	Scale Invariant Feature Transform (SIFT)	Traditional	Transformed image data into scale-invariant coordinates relative to local features.
Dalal and Triggs, 2005	Histogram of Oriented Gradients (HOG) Detector	Traditional	Counted occurrences of gradient orientation in the localised portion of an image, using magnitude and angle of the gradient to compute features.
Bay *et al*, 2006	Speeded Up Robust Features (SURF)	Traditional	Provided a faster alternative to SIFT by using Integral images and Hessian matrix approximations.
Felzenszwalb *et al.*, 2008	Deformable Part-based Model (DPM)	Traditional	Modelled object parts with flexible templates, allowing for deformation and variability.
Girshick *et al.*, 2014	Region-based Convolutional Neural Networks (R-CNN)	Deep Learning (*Two-Stage*)	Used selective search for generating region proposals, combined with deep learning for accurate object localization and classification.
He *et al.*, 2014	Spatial Pyramid Pooling Networks (SPPNet)	Deep Learning (*Two-Stage*)	Introduced Spatial Pyramid Pooling, which allowed for the use of images of arbitrary sizes by adapting to a fixed-size representation.
Girshick *et al.*, 2015	Fast R-CNN	Deep Learning (*Two-Stage*)	Improved the efficiency of R-CNN by introducing a Region of Interest (ROI) pooling layer.
Ren *et al.*, 2015; 2017	Faster R-CNN	Deep Learning (*Two-Stage*)	Introduced a Region Proposal Network (RPN) to generate region proposals, eliminating the need for external region proposal methods.
Liu *et al.*, 2016	Single Shot MultiBox Detector (SSD)	Deep Learning (*One-Stage*)	Pioneered the concept of predicting multiple bounding boxes of different scales and aspect ratios in a single forward pass.

Dai *et al.*, 2016	Region-based Fully Convolutional Neural Networks (R-FCN)	Deep Learning (*Two-Stage*)	Constructed position-sensitive score maps, which divide the regions of interest into a spatial grid.
Redmon *et al.*, 2016	You Only Look Once (YOLO)	Deep Learning (*One-Stage*)	Presented a unified approach by dividing the image into a grid and predicting bounding boxes and class probabilities simultaneously.
He *et al.*, 2017	Mask R-CNN	Deep Learning (*Two-Stage*)	Extended Faster R-CNN to include an additional branch for pixel-wise segmentation, enabling simultaneous object detection and instance segmentation.
Lin *et al.*, 2017a	Feature Pyramid Networks (FPN)	Deep Learning (*Two-Stage*)	Addressed the challenge of detecting objects at different scales by introducing a top-down architecture for feature pyramids.
Lin *et al.*, 2017b	RetinaNet	Deep Learning (*One-Stage*)	Introduced focal loss to address the class imbalance problem and combined a FPN with a one-stage detection framework.
Redmon and Farhadi, 2017	YOLOv2 (YOLO9000)	Deep Learning (*One-Stage*)	Extended YOLO with iterative improvements, including BatchNorm, higher resolution, and anchor boxes, to handle objects at different scales.
Redmon and Farhadi, 2018	YOLOv3	Deep Learning (*One-Stage*)	Added an objectness score to bounding box prediction, added connections to the backbone network layers, and made predictions at three separate levels of granularity to improve performance on smaller objects.
Bochkovskiy *et al.*, 2020	YOLOv4	Deep Learning (*One-Stage*)	Introduced a series of optimizations, including improved feature aggregation, PANet, and Mish activation function.
Jocher, 2020	YOLOv5	Deep Learning (*One-Stage*)	Focused on simplicity, ease of use, and performance, using a PyTorch training and deployment framework for improved flexibility and user adoption.
Pramanik *et al.*, 2021	Granulated R-CNN (G-RCNN)	Deep Learning (*Two-Stage*)	Generated ROIs using the granulation over the pooling feature map.
Jocher *et al.*, 2023	YOLOv8	Deep Learning (*One-Stage*)	Presented an anchor-free model that directly predicts the centre of an object instead of the offset from a known anchor box.

While traditional object detection methods are relatively mature and have been seen to achieve state-of-the-art results, there are inherent shortcomings. Firstly, the sliding window-based region selection strategy is inefficient and inaccurate due to the high computational complexity and high window redundancy. Secondly, the extraction of low-level vision feature representations does not effectively bridge the semantic gap (Zhao *et al.*, 2022; Wang *et al.*, 2023b).

The advent of deep learning-based methods has led to significantly higher object detection performance on benchmark datasets, such as PASCAL Visual Object Classes (VOC) (Everingham *et al.*, 2010), with mean Average Precision (mAP) increasing from approximately 30-50% for traditional methods like HOG-DPM to over 70-80% for deep learning models like Faster R-CNN and YOLOv3 (Ren et al., 2015; 2017; Redmon and Farhadi, 2018).Since the proposal of region-based convolutional neural networks (R-CNN) by Girshick *et al.* in 2014, the development trend has shifted to incorporate CNNs, leading to continually improved accuracy for various complex applications and improved detection speed towards real-time object detection (Xiao *et al.*, 2020).

Over the past decade, many novel deep learning-based object detection models have been successfully applied to image data for a range of applications (Inbar *et al.*, 2023). The pipelines for these methods are generally thought of as either of two approaches: one-stage and two-stage detectors (Table 5.1; Figure 5.2). It is normally regarded that two-stage detectors may achieve higher accuracy, but with a high computational cost than one-stage detectors (e.g., Carranza-García *et al.*, 2021).

The two-stage methods divide the detection process into region proposal and classification stages (Figure 5.2). These frameworks first propose several candidate regions using reference boxes, known as anchors. These regions of interest (ROIs) are then classified, and their localisation refined. In R-CNN, an external selective search is used for generating region proposals that are then fed to a CNN for classification and bounding box regression (Girshick et al., 2014). Further developments within the R-CNN family of detectors produced the Faster R-CNN model, in which feature maps produced by a CNN backbone are shared between both region proposal and classification stages, reducing the cost of externally generating proposals and resulting in a significant improvement in efficiency (Ren *et al.*, 2017). Faster R-CNN has inspired many subsequent works that have tried to further improve detection accuracy with different approaches, such as designing alternative backbones that can obtain more semantic representations or proposing modifications to the existing framework. The two-stage approach can achieve more reliable detections, but with an increase in computational requirements.

In one-stage methods, object detection is viewed as a regression problem where classification and localisation are performed directly in a single feed-forward fully convolutional network, thus achieving faster inference rates (Figure 5.2). Single Shot MultiBox Detector (SSD) (Liu *et al.*, 2016) and You Only

Look Once (YOLO) (Redmon *et al.*, 2016) were among the first proposals for a simpler, unified architecture, without requiring per-ROI computations. However, the problem of extreme class imbalance between foreground and background in images prevented these models to detection accuracies of approximately 31.2% average precision (AP) for SSD and 33.0% AP for YOLO on COCO test-dev (Lin *et al.*, 2014; Lu *et al.*, 2020). RetinaNet attempted to address this problem by introducing a focal loss function and incorporating a Feature Pyramid Network (FPN) architecture to efficiently detect objects at multiple scales, achieving a higher detection accuracy of 40.8% AP on COCO test-dev (Lin *et al.*, 2017b). YOLOv3 employed a series of optimisations, including a DarkNet-53 backbone and multiple detection scales, to better balance speed and accuracy (Redmon and Farhadi, 2018).

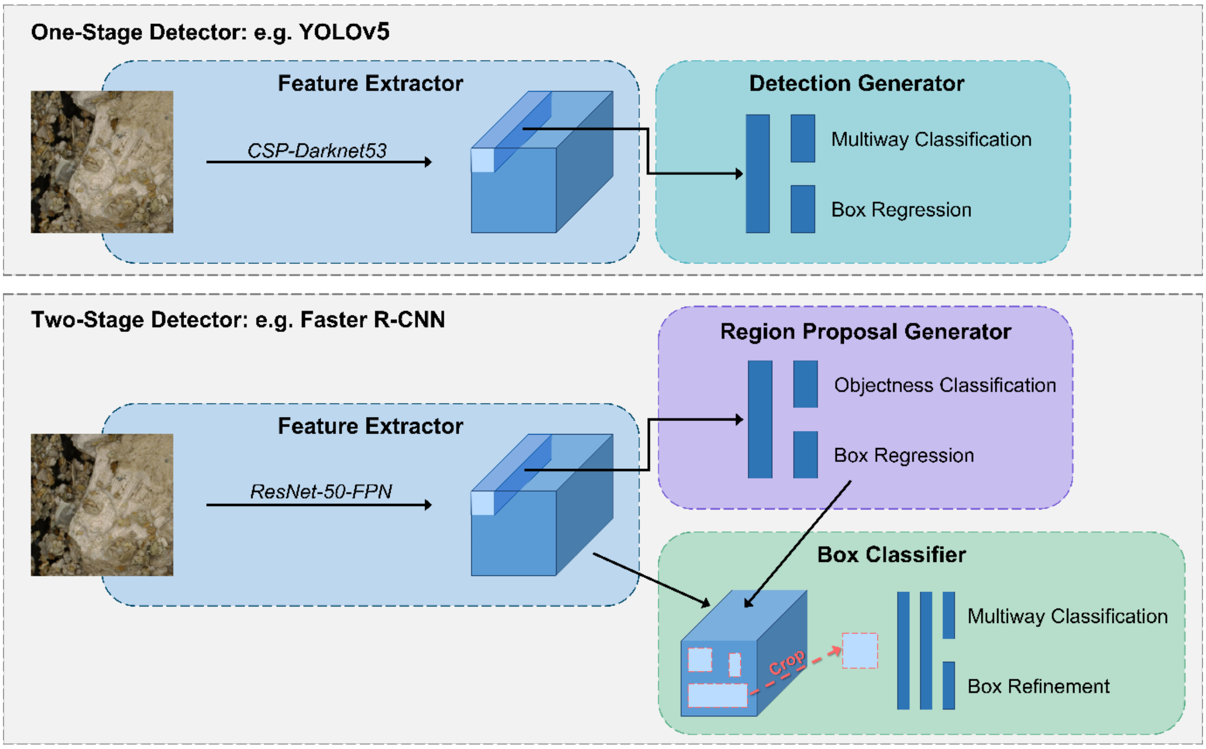

Figure 5.2: Deep learning objection detection meta-architectures for one-stage and two-stage detectors (modified from Huang *et al.*, 2017b). "Objectness" measures membership to a set of object classes vs. background.

More recently, the design of anchor-free models has been a popular approach in the development of one-stage detectors. Anchor-free frameworks, such as YOLOv8, directly predict the centre of an object instead of the offset from a known anchor box (Jocher *et al.*, 2023). These simple networks can be more flexible, but difficulties may arise when detecting objects in cluttered scenes or when objects are partially occluded. In general, when compared to their two-stage counterparts, one-stage detectors detection methods have achieved slightly lower accuracies, with many models reporting mean average precision (mAP) values in the range of 50–60% on COCO datasets, compared to 70–80% mAP achieved by two-stage methods like Faster R-CNN. However, one-stage detectors are often favoured for their higher detection speed, with many models capable of real-time detection (Soviany and Ionescu, 2018).

Object detection for geoscience

Object detection is important for a wide range of geoscience-related applications. From analysis of the literature, it is apparent that many of the current studies are based around geospatial data in remote sensing applications including, but not limited to, land cover mapping (e.g., Zhang *et al.*, 2020), sinkhole detection (e.g., Yavariabdi *et al.*, 2023), target detection (e.g., Zhou *et al.*, 2016; Xu and Wu, 2020), and landslide detection (e.g., Cheng *et al.*, 2021). For seismic data, work has been done for detection seismic objects, such as faults, gas chimneys, salt domes and 4D anomalies (e.g., De Groot *et al.*, 2004; Xiong *et al.*, 2018; Wei *et al.*, 2022), and for seismic exploration (e.g., Tian *et al.*, 2023b), signal detection for seismic cross-correlation (e.g., Zhao *et al.*, 2021), and horizon interpretation (e.g., Wu *et al.*, 2019). Deep learning-based object detection algorithms have also been applied to geochemical data (e.g., Zhang *et al.*, 2021b), planetary data for surface mapping (e.g., Bickel *et al.*, 2020; Rubanenko *et al.*, 2021) and target detection (e.g., Yang and Cai, 2021; Bechtold *et al.*, 2023), wireline logs (e.g., Wei *et al.*, 2019; Hong *et al.*, 2021), and outcrop photographs (e.g., Zhenlong *et al.*, 2023).

Despite these successes, there has been limited application to core data, where research has been largely focused on the identification and localisation of fractures (e.g., Alzubaidi *et al.*, 2022; 2023) or basic rock typing and lithological classification (e.g., Panferov *et al.*, 2020; Alzubaidi *et al.*, 2021). Several authors have also proposed rock detection models which can identify and classify different lithologies of hand specimens against a white or neutral background (Liu *et al.*, 2020b; Xu *et al.*, 2021; Xu *et al.*, 2022). With regards to the identification of grains and fossils, current studies are limited to X-ray micro-tomography (MicroCT) images (e.g., Carvalho *et al.*, 2020) and thin section images where internal structures are visible to aid detection (e.g., Koeshidayatullah *et al.*, 2020a; Liu and Song, 2020; Yu *et al.*, 2021; Bonté *et al.*, 2022; Bouziat *et al.*, 2022) or neutral background images of hand specimens where the models are assessed on their ability to identify the different fossil types, rather than identify and locate within a physical rock (e.g., Wang *et al.*, 2022; Liu *et al.*, 2023). While these hand specimen datasets can provide valuable insights for training models, particularly in capturing the full 3D geometry of fossils that are often sectioned in thin sections or core images, they may be less useful for practical applications. This limitation arises because specimens in these collections are typically well-preserved, clean, and isolated, representing idealised examples that are not often encountered in real-world geological contexts. In contrast, fossils found in the field or in cores are often incomplete, obscured by matrix material, or partially eroded, presenting additional challenges that these models must address to perform effectively in practical scenarios.

Such technologies have been important step towards automated lithological and taxonomic interpretations, which elicits the question – can this technology be translated to core analysis applications? To date, there has been no attempt to apply deep learning-based object detection methods to fossil and grain identification and classification in core image data. Here, the aim is to develop an object detection algorithm that identifies carbonate constituent grains and provides an accurate point count. With this technology, carbonate components could be located and characterised on a per-grain basis in an automated way. For this study, we compare performance on core data of two of the best-known detection methods based on deep learning: YOLOv5, as the one-stage detector, and Faster R-CNN, as the two-stage detector. YOLOv5 was chosen over YOLOv8 due to the nature of the core images, since the newer version may be less effective in the presence of partial occlusion, multiple objects, and cluttered scenes.

5.2 Methodology

5.2.1 Computing environment

All training, evaluation and inference stages were completed using the PyTorch (v1.7.1) deep learning framework (Paszke *et al.*, 2019) on the publicly accessible cloud service Google Colab Pro[12], running Python 3, with access to V100 or A100 Nvidia GPUs and high-RAM runtimes (≥20GB).

5.2.2 Data preparation

The dataset for this study, as discussed in detail in Chapter 2, has been created using high-resolution core images scans in TIFF format from ODP Leg 133 (Davies *et al.*, 1991) and ODP Leg 194 (Isern *et al.*, 2002) offshore north-eastern Australia, and IODP Expedition 359 in the Maldives archipelago (Betzler *et al.*, 2017). The information from the digital core images is represented by a 3D tensor with separate red, green, and blue (RGB) channels. Digital core scan images of a 1.5 metre section are approximately 31,600 pixels

[12] https://colab.research.google.com/

x 1,750 pixels; however, only the sediment core section itself is required here, which is on the order of 29,900 pixels x 1,375 pixels within the image (Dawson *et al.*, 2023).

Following the core descriptions provided by the original expedition scientists and the locations of close-up core images, regions of interest within the core sections were identified. From this, the raw core scan images were cropped to sub-images of 640 x 640 pixels, creating a dataset of 381 images for this study. It should be noted that no images of cores taken from ODP Leg 194 Sites 1196 and 1199 are included in the dataset, as these are later used for testing model deployment and generalisation on unseen data.

Data annotation

All images were manually annotated with bounding boxes around carbonate components, along with corresponding class labels, using the Annotation Toolbox on the online platform "Supervisely" (Supervisely, 2023). The quality of this manual annotation process plays a significant role in determining the performance of the subsequent automated detection and classification processes. Any inaccuracies or inconsistencies in the annotations could propagate through the model training pipeline, potentially reducing accuracy and reliability. While no formal assessment of annotation quality was conducted in this study, measures such as cross-verification by multiple annotators or benchmarking against expert geologist input could be implemented in future work to enhance annotation consistency and reliability.

In total, 9,048 individual carbonate components were labelled into 11 classes (Figure 5.3; Table 5.2): aggregate grains, bivalve, bryozoan, coral, echinoderm, foraminifera ("foram"), gastropod, non-skeletal grains (including clasts, peloids, ooids and other coated grains), red algae, serpulid, and skeletal grains (where it was not possible to identify to class level). No distinction was made between benthic and planktonic foraminifers.

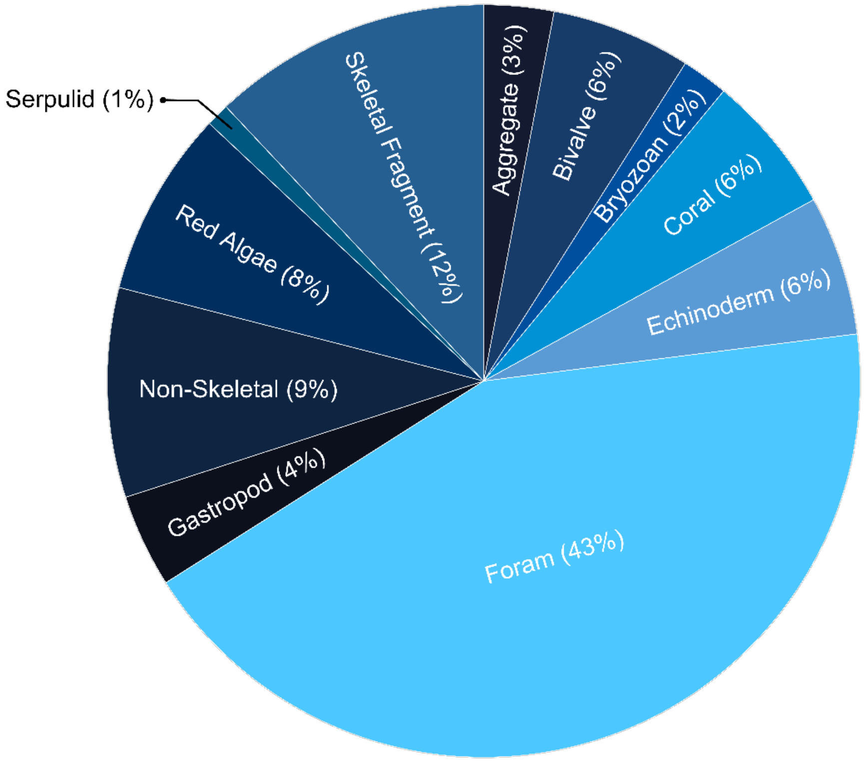

Figure 5.3: Distribution of object instances in the dataset.

Table 5.2: Object overview for each class. Abbreviations: Min, minimum; Max, maximum; Av., average; px, pixels.

Class Name	Object Count	Min Area (%)	Max Area (%)	Av. Area (%)	Min Height (px)	Min Height (%)	Max Height (px)	Max Height (%)	Av. Height (px)	Av. Height (%)	Min Width (px)	Min Width (%)	Max Width (px)	Max Width (%)	Av. Width (px)	Av. Width (%)
aggregate	233	0.04	6.09	0.25	11	1.72	208	32.5	28.91	4.52	10	1.56	120	18.75	29.73	4.65
bivalve	580	0.03	27.49	0.93	4	0.62	342	53.44	48.51	7.58	8	1.25	629	98.28	58.44	9.13
bryozoan	184	0.05	12.56	0.62	10	1.56	310	48.44	44.57	6.96	15	2.34	178	27.81	45.43	7.1
coral	540	0.11	100	8.02	18	2.81	640	100	130.21	20.35	20	3.12	640	100	143.43	22.41
echinoderm	501	0.03	3.03	0.27	7	1.09	128	20	29.77	4.65	5	0.78	194	30.31	35.12	5.49
foram	3935	0	6.16	0.05	2	0.31	129	20.16	13.44	2.1	4	0.62	250	39.06	13.48	2.11
gastropod	379	0.03	11.94	0.68	10	1.56	209	32.66	46.94	7.33	12	1.88	241	37.66	46.52	7.27
non-skeletal	752	0	77.5	0.33	5	0.78	496	77.5	23.49	3.67	4	0.62	640	100	23.41	3.66
red-algae	726	0.02	96.88	2.39	8	1.25	639	99.84	68.2	10.66	9	1.41	621	97.03	72.05	11.26
serpulid	92	0.06	16.29	1.3	6	0.94	284	44.38	60.63	9.47	14	2.19	235	36.72	69.55	10.87
skeletal	1126	0.02	19.04	0.58	8	1.25	339	52.97	43.08	6.73	9	1.41	230	35.94	44.79	7

Dataset splits

The dataset was split into three subsets using the following proportions: 70% for training, 10% for validation and 20% for testing (Table 5.3).

Table 5.3: Number of images and labels in each data subset.

	Images	Labels
Train	267	6,289
Val	38	995
Test	76	1,764

5.2.3 Detection algorithms

We compare the performances of two deep learning based object detection models: YOLOv5 (Jocher, 2020), a one-stage detector, and Faster R-CNN (Ren *et al.*, 2017), a two-stage detector.

YOLOv5

In comparison to the previous YOLO releases, the YOLOv5 architectures (Jocher, 2020) are more flexible to use and offer improved accuracy, faster calculated speed, and smaller model size (Jiang *et al.*, 2022). Since there are multiple YOLOv5 variants of differing model sizes, we selected one of the smaller models, YOLOv5s, for this study. Whilst in nearly all cases, the larger models, like YOLOv5l and YOLOv5x, will likely produce better results, they have considerably more parameters, require more CUDA memory to train, more data to avoid overfitting, and are slower to run (Table 5.4).

Table 5.4: Pre-trained checkpoints for YOLOv5 models.

Model	COCO mAP	Speed V100 (ms)	Params (M)	Weights Size (MB)
YOLOv5n	28.0	6.3	1.9	4
YOLOv5s	37.4	6.4	7.2	14
YOLOv5m	45.4	8.2	21.2	41
YOLOv5l	49.0	10.1	46.5	89
YOLOv5x	50.7	12.1	86.7	166

Based on the results of Table 5.4, we conclude that YOLOv5s exhibits the best trade-off between detection accuracy, detection rate and model compactness. Hence, YOLOv5s was chosen as the one-stage detection model in this work.

The architecture of YOLOv5 (Figure 5.4) consists of three main parts (Aldughayfiq *et al.*, 2023):

1) **Backbone**: focuses on feature extraction, learning hierarchical representations of the input images.
2) **Neck**: enhances the features by incorporating multiscale information, allowing the model to better handle objects of different sizes.
3) **Head**: generates final predictions, including bounding boxes, objectness scores, and class probabilities, based on the processed features from the backbone and neck.

The backbone, which serves as the feature extraction network, consists of three main modules: the CBS (Convolutional, BatchNorm, and SiLu) module, the C3 (CSPNet with three convolutional layers) module, and the SPPF (Spatial Pyramidal Pooling Fusion) module. The CBS module is a fundamental building block in the YOLOv5 backbone, combining three essential operations in each unit: convolutional operations, which apply convolutional filters to the input feature map to capture spatial patterns; batch normalization, which normalises the output of the convolutional operation, contributing to stable training and faster convergence; and Sigmoid Linear Unit (SiLu) activation operations, a smooth and non-monotonic function that help to mitigate the vanishing gradient problem (Zhang *et al.*, 2023b). The C3 residual network module incorporates the Cross-Stage Partial (CSP) to facilitate the integration of feature maps from different stages of the network. The three convolutional layers extract hierarchical features from the input data, capturing both high-level and low-level information and enhancing complex pattern recognition. The SPPF module uses spatial pyramidal pooling to capture and combine multiscale features from the input feature map, which improves detection of objects with diverse spatial characteristics.

The neck of YOLOv5 comprises a combination of Feature Pyramid Network (FPN) and Path Aggregation Network (PANet). Introduced by Lin *et al.* (2017), the FPN structure is used to capture and convey semantic information through the integration of feature maps from different network layers, enabling the model to effectively handle objects of various scales effectively. The bottom-up structure of PANet (Liu *et al.*, 2018) complements the FPN by facilitating the flow of strong localisation features from lower layers to higher layers. The FPN+PAN structure enhances the feature fusion process, allowing the model to adapt to objects of various sizes and complexities.

The head is responsible for making final predictions based on the features extracted by the backbone and neck. Through multiscale target detection, the head processes these feature maps and generates predictions for bounding box coordinates, class probabilities, and confidence scores.

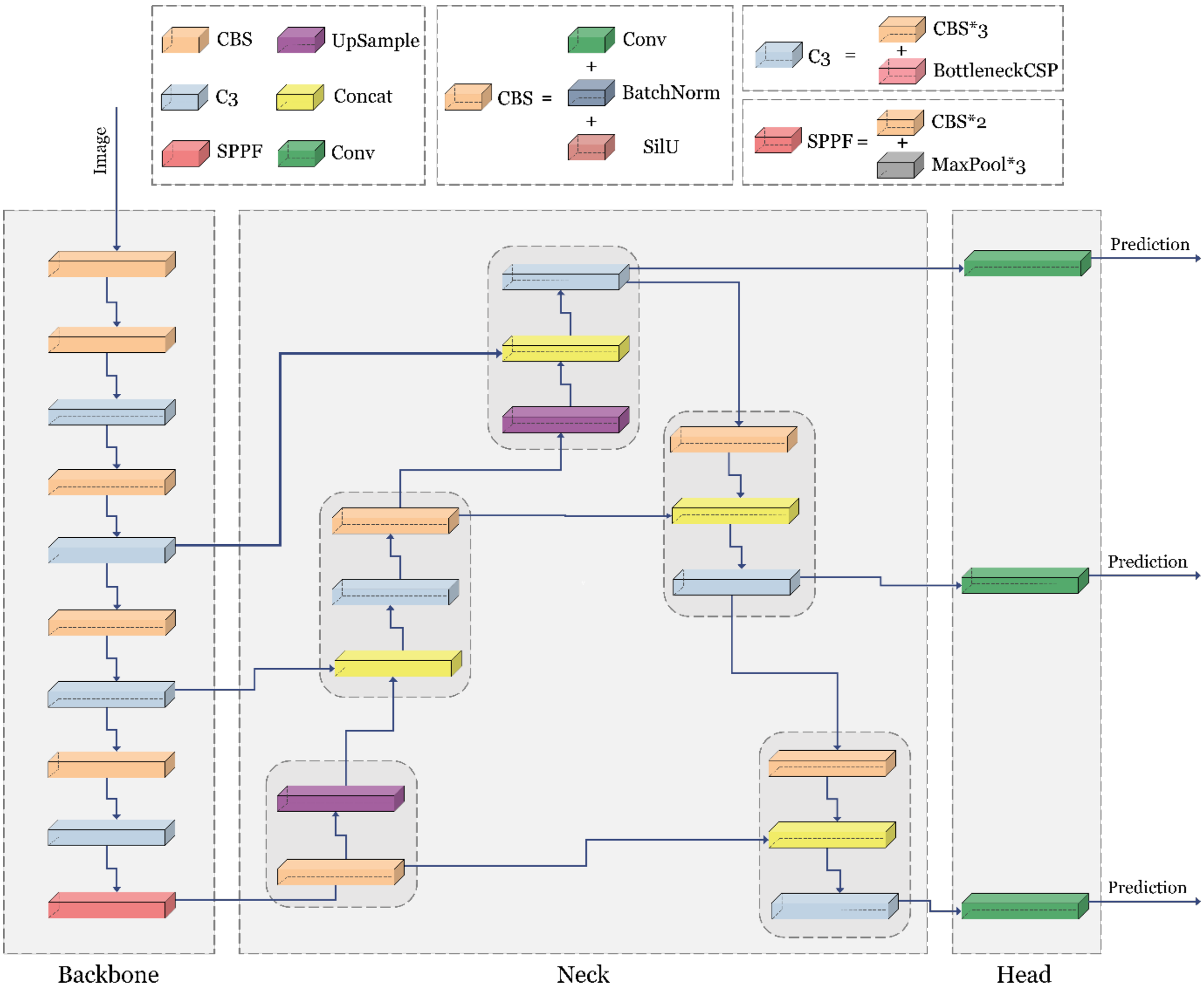

Figure 5.4: Image pipeline through YOLOv5 architecture.

Faster R-CNN

Of the R-CNN family, Faster R-CNN is one of the most widely used (Singh *et al.*, 2021). The architecture of Faster R-CNN (Figure 5.5) is, at the conceptual level, composed of three neural networks (Wang and Liu, 2020):

1) **Feature Network**: the backbone CNN that acts as a feature extractor**.**
2) **Region Proposal Network (RPN)**: generates region proposals, which are candidate regions that may contain objects.
3) **Detection Network**: performs classification and regression of bounding box coordinates.

The input image is first passed through the backbone CNN, which acts as an efficient feature extractor to form a feature map. The role of the RPN module is to generate candidate bounding boxes, which are then fed into the detection network. To generate the region proposals, a sliding window passes over the output convolutional feature map. At each sliding position, three sizes, 128^2, 256^2, and 512^2, and three aspect ratios, 1:1, 1:2, and 2:1, are defined, generating $k = 9$ anchors for each point in the feature map (Figure 5.6). For a convolutional feature map of size $W \times H$, there are $W\ H\ k$ anchors generated in total.

For each region proposal, a feature vector is extracted and then fed to two sibling fully connected (FC) layers – a box-classification layer and a box-regression layer (Figure 5.6). The box-classification layer represents a binary classifier that generates an 'objectness' score for each region proposal, i.e., the probability that each region contains an object or background. The box-regression layer returns a 4-D vector defining the bounding box coordinates of the region (Middya *et al.*, 2021). This 4-D vector specifies the bounding box in the form of $[x_{min}, y_{min}, x_{max}, y_{max}]$ representing the corners of the bounding box in 2D image space.

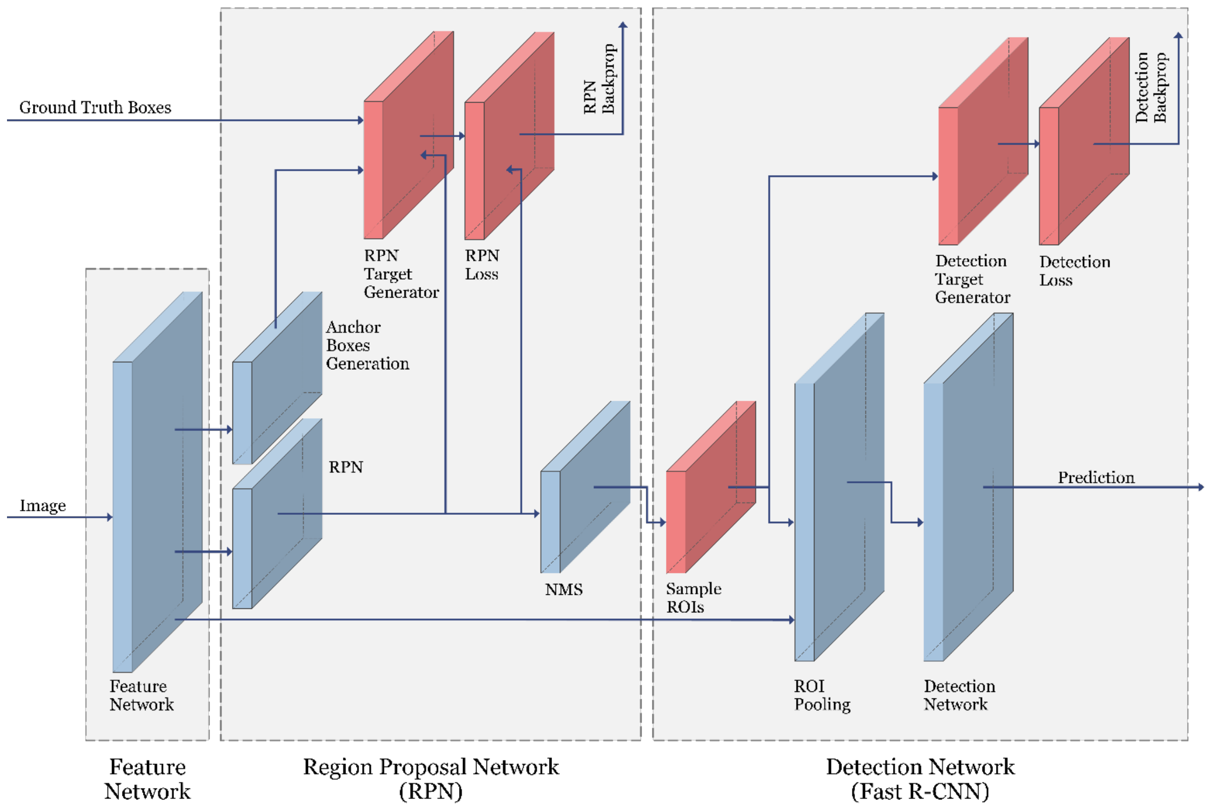

Figure 5.5: Image pipeline through Faster R-CNN architecture. The red coloured blocks are only active during training. Abbreviations: NMS, non-maximum suppression; ROI, region of interest.

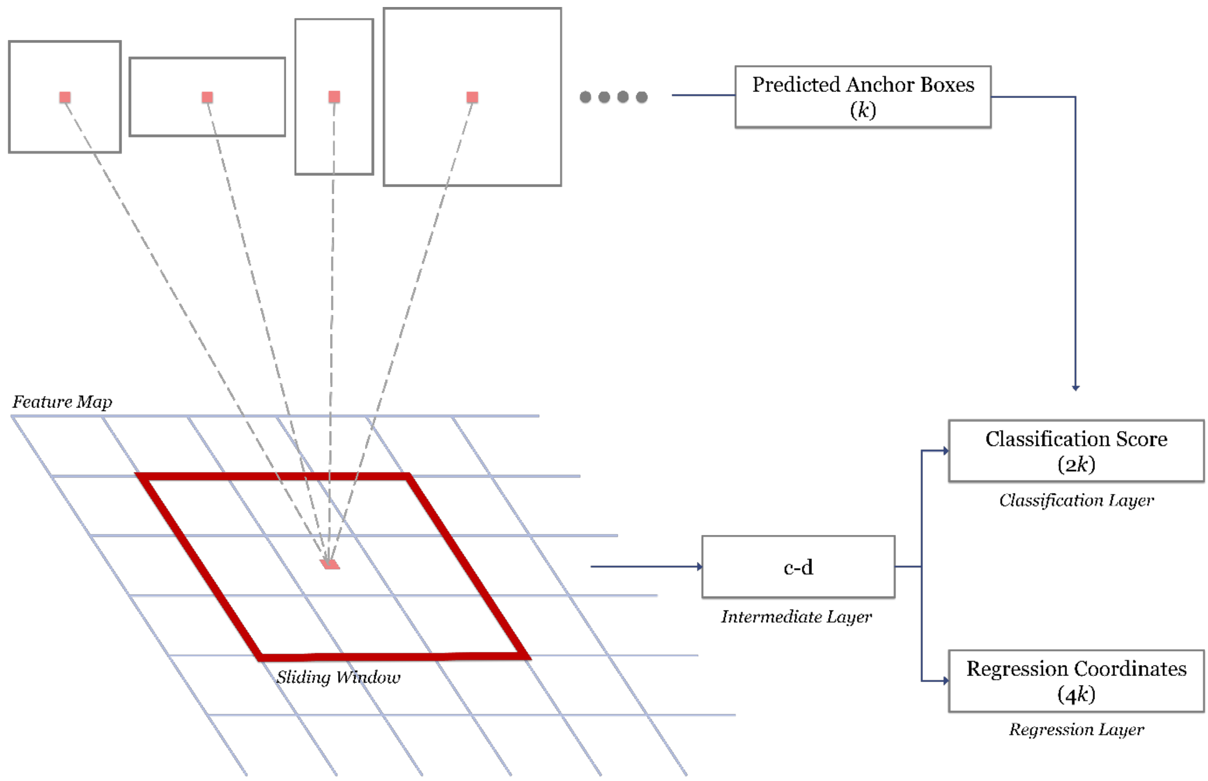

Figure 5.6: Illustration of the mini-network in Region Proposal Network (RPN) at a single position (modified from Ren *et al.*, 2017). The mini-network operates in a sliding window fashion, meaning that the fully connected layers are shared across all spatial locations. In this context, the sliding window refers to a small fixed-size region (sub-window) that moves across the convolutional feature map, one step at a time. At each position, the sub-window (or receptive field) acts as the input to the neural network, and the network processes this region to generate proposals for potential bounding boxes. The pink square within the sliding window typically represents the central point or anchor for that sub-window, around which region proposals are generated. For each anchor, the network predicts a set of bounding boxes with different sizes and aspect ratios (e.g., 128^2, 256^2, and 512^2, and aspect ratios 1:1, 1:2, 2:1). These bounding boxes define regions of interest around the anchor, which are then classified as containing an object or background and further refined by the box-regression layer. By sliding the window over the entire feature map, the RPN efficiently generates region proposals for the entire image.

For the detection network, Fast R-CNN (Girshick, 2015) is used as the region-based object detection CNN that utilises the proposals generated by the RPN. The region of interest (ROI) pooling layer extracts small fixed-size feature maps from the region proposals, allowing the model to focus on relevant spatial information for each proposed region (Chen *et al.*, 2023). The features from the ROI pooling layer are then fed into the fully connected layers in the classification head. These layers are responsible for assigning class labels to the proposed regions and refining the coordinates of the bounding boxes. Non-maximum suppression is applied to filter out redundant bounding box predictions and select the most confident detections. The remaining bounding boxes, along with their associated class labels and confidence scores, constitute the final predictions of the Faster R-CNN model.

5.2.4 Model training

For training from scratch, Jocher (2020) recommends a dataset with ≥ 1,500 images per class and ≥ 10,000 instances (labelled objects) per class. Therefore, due to the small size of the core dataset, transfer learning was identified as best approach for the object detection tasks in this study. Both models were trained using pre-trained weights on the Microsoft Common Objects in Context (COCO) dataset (Lin *et al.*, 2014), and the detection heads were fine-tuned towards carbonate component representation learning by adjusting the parameters of the final layers of the model. Specifically, the pre-trained convolutional layers, which were initially trained on a large, general dataset, were kept frozen during the fine-tuning process. The last few layers, which are responsible for detecting specific object classes, were replaced with new layers tailored to identify carbonate components in the images. During the fine-tuning process, these new detection heads were trained using the core dataset to adapt the model to carbonate components, ensuring that the model could learn the relevant features specific to this task, such as texture, shape, and other visual properties of carbonate components. During training and inference stages, all images were scaled to 640 x 640 pixels and auto-orientation of pixel data was applied with exchangeable image file format (EXIF)-orientation stripping.

YOLOv5

In this work, we use the YOLOv5s version with CSP-Darknet53 backbone, pre-trained on COCO 2017 dataset. The network was trained with a batch size of 32 for 300 epochs, using the Stochastic Gradient Descent (SGD) optimiser with momentum 0.937, initial learning rate of 0.01 and a weight decay of 0.0005. The batch size of 32 was selected as a balance between model training speed and memory constraints, as it provides stable gradients without exceeding GPU memory limits. Larger batch sizes can accelerate training, but smaller ones may improve generalisation. A momentum of 0.937 was chosen based on recommendations from prior works, as it effectively accelerates convergence in the relevant direction while damping oscillations. The initial learning rate of 0.01 was chosen as it is commonly effective in deep learning models for object detection tasks, providing an optimal rate for updates without overshooting the optimal minimum. A weight decay of 0.0005 was used as a regularisation technique to prevent overfitting by penalizing large weights, following the approach in the original YOLOv5 implementation. Additionally, the Lambda Learning Rate scheduler was used gradually reduce the learning rate during training, as proposed in the original implementation, which helps the model converge faster and perform better. The rest of the hyperparameters were kept as the default values, including those for data augmentation.

Faster R-CNN

We use the Faster R-CNN model architecture with a ResNet50-FPN backbone, pre-trained on COCO 2017 dataset. The network was trained for 300 epochs with a batch size of 32. SGD was used as the optimiser, with an initial learning rate of 0.01 and weight decay of 0.0005. A Reduce Learning Rate on Plateau scheduler with a default patience of 10 epochs and a factor of 0.1 was used. For data augmentation, the Albumentations library (Buslaev *et al.*, 2020) was used to apply horizontal flip (0.5), HSV (Hue, Saturation, and Value) augmentation (hue = 0.015, saturation = 0.7, value = 0.4), image scale (0.5), and translation (0.1). The rest of the hyperparameters remained as the default values.

5.2.5 Evaluation metrics

During training, the performance of the detectors was evaluated by monitoring the changes in training and validation losses with each epoch. Training loss refers to the error calculated on the training set, while validation loss measures the error on a separate validation set that is not used for training. Following completion of training, performance evaluation was carried out on the test set of the core dataset using two primary metrics, namely computational cost and mean average precision (mAP), which is based on the sub-metrics Intersection over Union (IoU), precision and recall.

The IoU measures the degree of overlap between the ground truth bounding box (which outlines the true location of the object) and the predicted bounding box coordinates (the model's prediction of where the object is located). Specifically, the IoU is calculated as the area of overlap between the two boxes divided by the area of their union:

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}} \tag{5.1}$$

The area of overlap is simply the region where both boxes coincide, while the area of union refers to the total area covered by both boxes combined. The higher the IoU, the more accurate the predicted bounding box is compared to the ground truth. A threshold of 0.5 for the IoU is commonly used in many object detection benchmarks, such as the PASCAL VOC challenge, and it represents a balance between strict and lenient detection performance. If the IoU exceeds the threshold value of 0.5, the prediction is considered a True Positive. A prediction would be considered a False Positive if the IoU does not exceed the threshold value, if there is no overlap between the true and predicted boxes, or in the case of a misclassification. If the network fails to find an object, the prediction is considered a False Negative (Figure 5.7).

Precision calculates the proportion of positive detections that were correct:

$$\text{Precision: P} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{5.2}$$

whereby TP is the number of true positive cases, and FP is the number of false positives.

Recall calculates the proportion of true positives that were correctly identified, i.e., the ability of the model to make correct detections out of all ground truth bounding boxes:

$$\text{Recall: R} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{5.3}$$

whereby FN is the number of false negatives.

Average Precision (AP) is defined as the area under the precision-recall (PR) curve, thus summarising the PR curve to one scalar value:

$$\text{Average Precison (AP)} = \int_{r\,=\,0}^{1} \text{P(R)}\, d\text{R} \tag{5.4}$$

When both precision and recall are high, AP is high, and when either is low, the AP is low across a range of confidence threshold values. The range for AP is given between 0 and 1.

Mean Average Precision (mAP) is calculated by taking the mean of AP across all the classes under consideration:

$$\text{mAP} = \frac{1}{n}\sum_{k=1}^{k\,=\,n} \text{AP}_k \tag{5.5}$$

where AP_k is the AP of class k, and n is the number of classes.

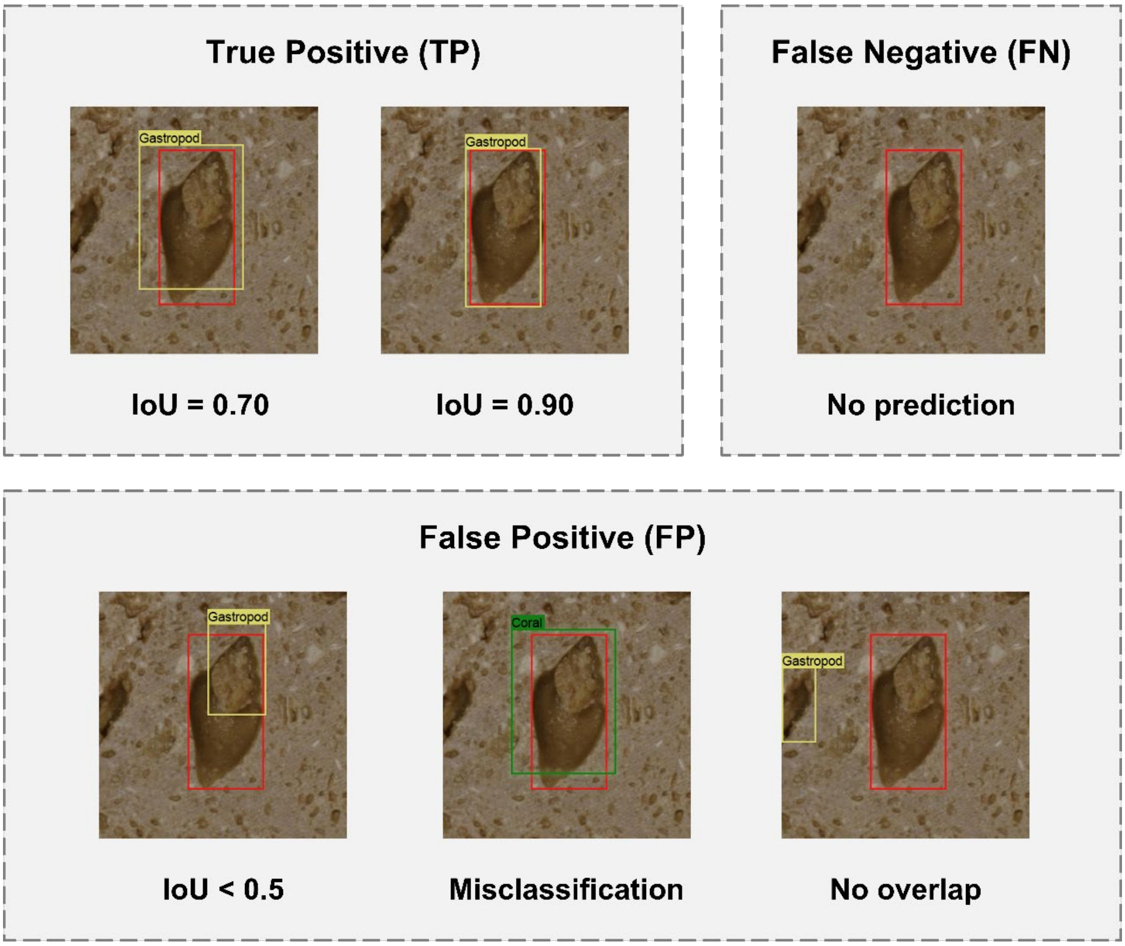

Figure 5.7: Example predictions with an Intersection over Union (IoU) threshold value of 0.5. The threshold is used to determine whether a prediction is true positive (TP), false positive (FP) or false negative (FN). The red bounding box represents the ground truth label. The example model predictions are shown by the yellow (gastropod) and green (coral) boxes.

5.2.6 Feature map visualisation

Feature maps, particularly those of convolutional layers, represent the outputs of feature extraction from convolutional kernels. They encompass both low-dimensional features like brightness and texture, as well as high-dimensional features that may be challenging for human interpretation. While lower layers typically capture simple, interpretable features, such as edges or textures, higher layers tend to capture more abstract patterns that may be task-specific, such as object shapes or spatial relationships. These high-dimensional features could represent meaningful aspects of the data that are crucial for model performance, though they are often difficult for humans to interpret directly. However, there is a risk that they may also capture noise or artefacts, particularly when the dataset is small or imbalanced. Therefore, exploring and interpreting these features through visualisation or saliency maps could provide insights into whether they truly reflect relevant patterns or are simply overfitting to the dataset. In this study, we analysed changes in low-dimensional features within different hidden layer feature maps, and then correlated these differences with the performance of the object detection networks.

For visualisation, we selected four convolution kernels from four convolutional layers as focal points. Given our approach of examining feature map variations to aid interpretation of network-related principles, we opted for four feature layers spanning different stages of the feature extraction module across the two object detection networks. As the number of convolution kernels in each convolutional layer is different, up to a maximum of 1024, we used stratified sampling to extract four convolution kernels per layer, with the total pixel values as indices for visualisation. First, we calculated the total pixel values for each convolution kernel feature map according to Equation 5.6 and sorted them in descending order.

$$\text{Pixel value} = \sum_{i=1}^{N}\sum_{j=1}^{M} \text{value}_i^j \tag{5.6}$$

where N is the width (number of columns) and M is the length (number of rows) of the feature map, (i, j) is the position of the pixel, and value is the value of the pixel. Each feature map is a 2D matrix, where each element represents the activation value of a specific pixel.

The pixel valuation calculation (Equation 5.6) calculates the total pixel value across a grid of pixels and is typically used for summarising or aggregating the pixel values in a feature map. This captures overall intensity or information from a specific convolution kernels output. Summing the pixel values of the feature map generated by a kernel helps identify which parts of the image are most activated by that kernel, giving insight into its importance or contribution to the detection task. The number of kernels in each layer was then divided into four equal parts, Finally, we selected the first image from each part for visualisation. Given that images are transmitted in tensor form within the network, the process of visualising feature maps involved extracting the tensor of the target convolution kernel and outputting it as a two-dimensional image.

5.3 Results

Two different object detection frameworks were compared, including both one-stage (YOLOv5) and two-stage (Faster R-CNN) detectors. Performance is measured by the average precision on the IoU with a threshold of 0.5 (AP_{50}) for each class, mean average precision (mAP), and the inference time required to perform a prediction (Tables 5.5, 5.6; Figure 5.8).

Table 5.5: Performance evaluation of object detection frameworks during the prediction stage.

Detection Pipeline	Backbone	Parameters[13]	Mean Average Precision (mAP)	Average Inference Time (ms/image)
YOLOv5	CSP-Darknet53	7.2M	0.59	48.3
Faster R-CNN	ResNet50-FPN	41.8M	0.73	729.1

Table 5.6: Performance evaluation of object detection frameworks for each class.

	Class	Detection Pipeline	
		YOLOv5	**Faster R-CNN**
Average Precision (AP_{50})	**Aggregate**	0.31	0.52
	Bivalve	0.74	0.89
	Bryozoan	0.58	0.66
	Coral	0.62	0.74
	Echinoderm	0.48	0.63
	Foraminifera	0.29	0.59
	Gastropod	0.90	0.95
	Non-Skeletal	0.56	0.68
	Red Algae	0.75	0.91
	Serpulid	0.53	0.64
	Skeletal Fragment	0.72	0.77
Mean Average Precision (mAP)		**0.59**	**0.73**

[13] Parameters refers to the number of parameters, i.e., weights and biases, in the model. This is calculated by summing the number of elements in each layer of the model. This value may differ from that given when running inference e.g., for YOLOv5s, the model is fused in test.py, detect.py and export.py, eliminating nn.BatchNorm2d() layers and slightly reducing parameter count.

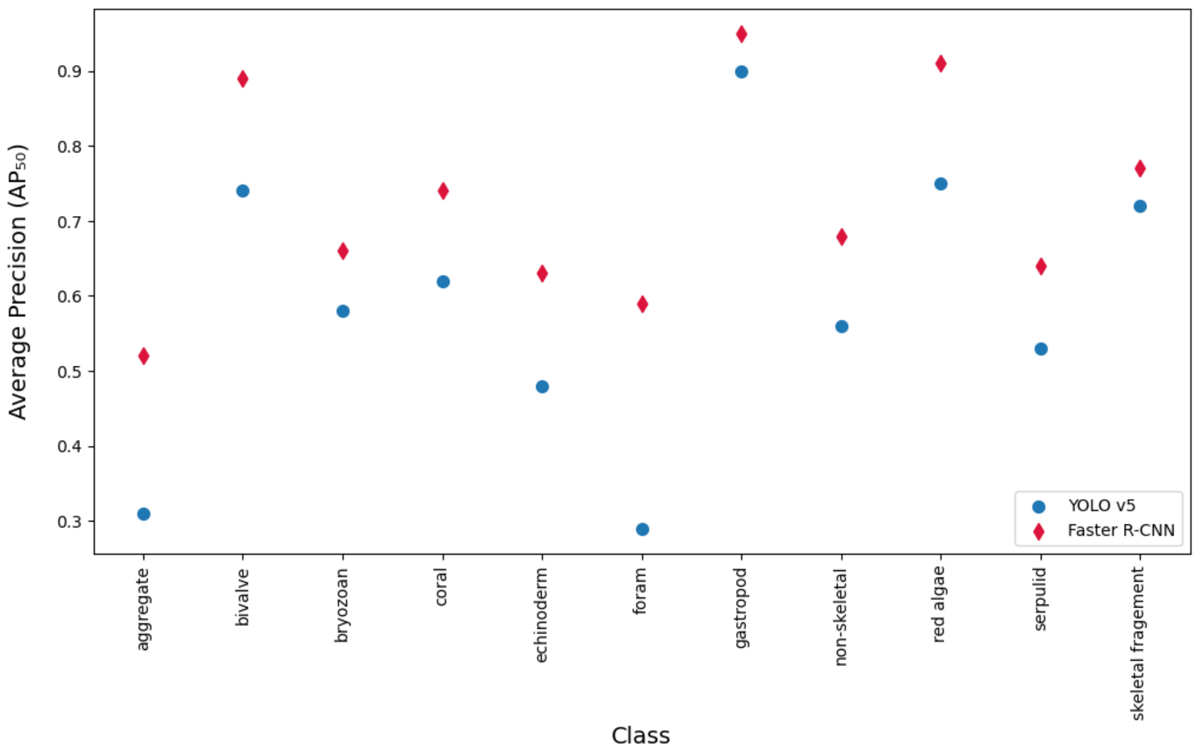

Figure 5.8: Average precision (AP_{50}) of object detection frameworks for each class.

For YOLOv5, the AP_{50} varies significantly between classes, ranging from a low of 0.29 for foraminifera to a high of 0.90 for gastropods, with a mAP of 0.59 (Table 5.6; Figure 5.8). YOLOv5 exhibits the fastest detection speed among the evaluated networks, with anaverage inference time of 48.3 ms per image (Table 5.5). For Faster R-CNN, the highest AP_{50} is observed from gastropods (0.95) whilst the lowest is for aggregates (0.52) (Table 5.6; Figure 5.8). Despite achieving the highest AP_{50} in all studied classes and a mAP (0.73), the inference time of Faster-RCNN is substantially slower at 729.1 ms per image when predicting different objects in a single image (Table 5.5, 5.6; Figure 5.8).

The variability in detection performance across classes warrants further exploration. The superior performance of both networks on gastropods may be attributed to their distinctive shapes, well-defined boundaries, and relatively larger sizes in the images (Table 5.2), which may make them easier to identify. Similarly, red algae and bivalves also perform well, likely due to their larger size and their unique and recognisable morphological features.

Conversely, classes such as foraminifera and aggregates show much lower average precision. Foraminifera are often small and intricate, which may be challenging for the networks to distinguish, particularly when contrast is suboptimal. Aggregates, on the other hand, present a different challenge. They tend to lack consistent shape or features, appearing as amorphous clusters that may overlap or blend with the surrounding matrix, making them difficult to localise and classify reliably. These factors may contribute to higher misclassification rates and greater overlap with background elements, leading to reduced AP_{50} values.

The variability observed in intermediate classes, such as echinoderms, bryozoans, and serpulids, may stem from a combination of factors, including variability in their morphological features and possible similarities in appearance with other classes. Additionally, the networks may struggle when certain objects are partially occluded, poorly illuminated, or present at varying scales within the dataset.

The disparity in mAP between YOLOv5 and Faster R-CNN highlights the trade-off between speed and accuracy. The lower AP_{50} scores achieved by YOLOv5 across all classes could be attributed to its one-stage design, optimised for speed rather than precision. Conversely, the two-stage approach of Faster-RCNN enables more precise region proposals and refined classification, though at the cost of computational efficiency. This suggests that the choice of detection framework should consider the specific application requirements, balancing the need for real-time performance against the demand for high precision.

5.4 Discussion

The most successful model, Faster-RCNN, achieved a mAP of 0.73, outperforming YOLOv5 in terms of AP for all classes. However, this detection accuracy was at the expense of a significantly slower inference time (729.1 ms/image). This is consistent with results achieved in previous object detection studies for component identification in thin section images (Koeshidayatullah *et al.*, 2020a).

The foraminifer class was one of the worst performing classes, showing a relatively poor AP in all models (0.29–0.59). This result is counterintuitive since foraminifera are the most common class in the dataset (3,865 instances; 42.72%). We suspect that the poor performance of this class may instead be due to the size of the instances, with the bounding box areas constituting an average of 0.05% of the total image, with an average height of 13.44px and an average width of 13.48px (Table 5.2). In contrast, the gastropod class, which showed the best metrics in all models with an AP of 0.90–0.95, makes up only 4.14% of the dataset (375 instances) and has an average bounding box area of 0.68% of the entire image, with an average height of 46.94px and an average width of 37.66 pixels (Table 5.2). This suggests that smaller objects (<35px) may present a challenging and significant issue for the object detection models examined here.

Examining different examples of the performance of this framework in detecting carbonate grains reveal areas of strengths and weaknesses. In Figure 5.9, a lithified core, we see that the confident predictions of foraminifera are incorrect and are, in fact, holes or molds within the core itself. This could suggest that where the model is predicting higher than expected occurrences of foraminifera, these may be misclassifications of other features. It is important to note here that while many molds could indeed be large foraminifera that were dissolved, there is no distinction for holes or molds within the dataset classes and so these could look close to foraminifera. The gastropod identified with high confidence shows a clear turreted structure. The high precision of this class may be related to the distinct structure of gastropod shells compared to those of other components. For red algae, detection is generally good, however the bounding boxes do not always cover the full extent of the occurrence.

In Figure 5.10, an unlithified foraminifera-rich packstone, we observe misclassifications of aggregate and skeletal grains. The most interesting of this is the skeletal misclassifications, which are reflections in the wet sample of the LED lighting used by the line scanner. Although incorrect, the shapes formed by the reflections are similar to those of skeletal fragments. Background images containing reflections in wet samples could be introduced to the training dataset to improve robustness against this in future. Only bounding boxes with a confidence threshold >60% are displayed here, showing that although the model has identified many occurrences of foraminifera within the sample, the majority have been missed or fall below the threshold.

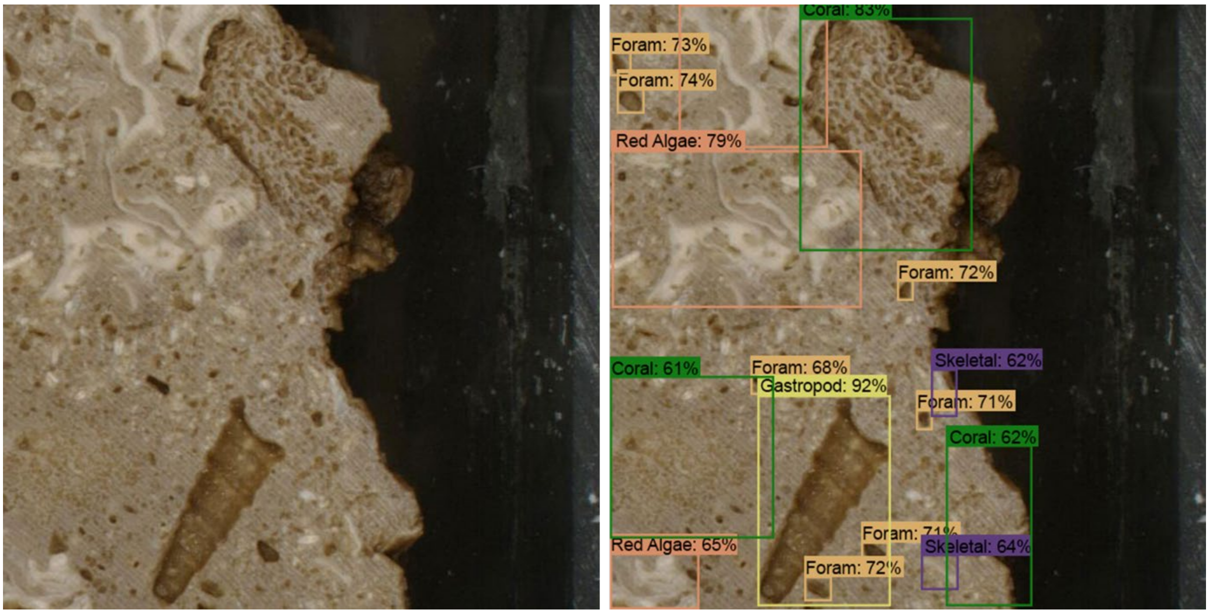

Figure 5.9: Example of predicted carbonate components for each identified object in Core 359-U1469A-17R-1 using Faster R-CNN. Only bounding boxes with a confidence threshold >60% are displayed for ease of interpretation.

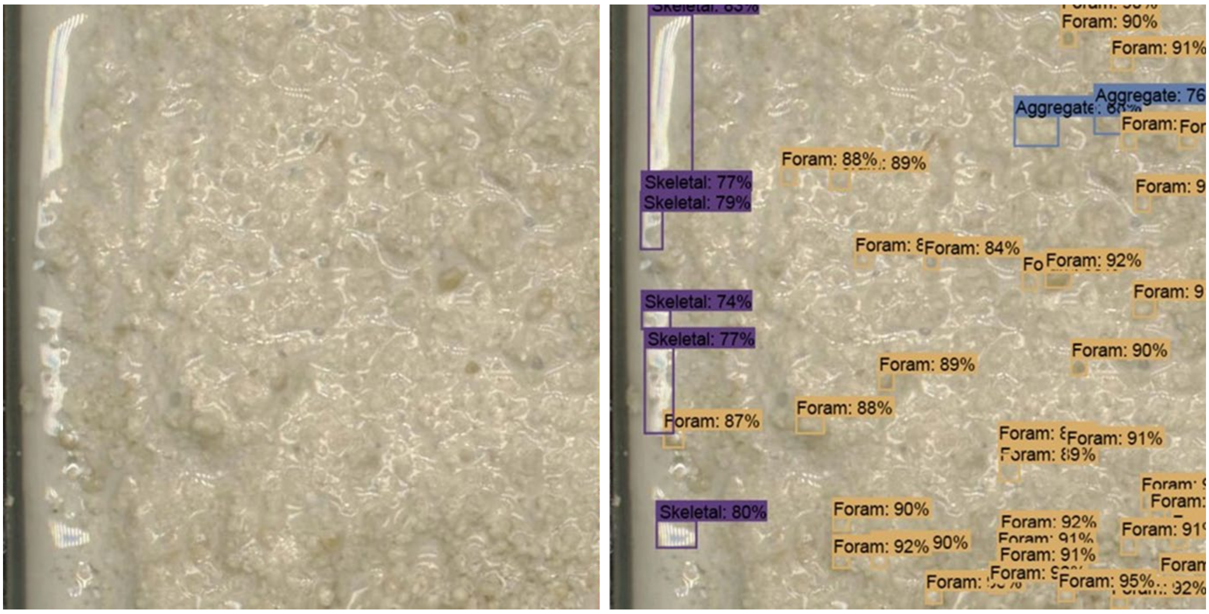

Figure 5.10: Example of predicted carbonate components for each identified object in Core 359-U1470A-17H-1 using Faster R-CNN. Only bounding boxes with a confidence threshold >60% are displayed.

In Figure 5.11, a partially lithified sample, we see that again many foraminifera occurrences are missing. This appears to be where the grains are rotated to a more laterally orientated view, suggesting that the model is sensitive to distinct morphologies in a frontal view. This could be addressed in future by introducing more examples of these grains in different orientations into the training set. Whilst a significant part of the coral has been identified, the model fails to include the entire structure within the bounding box, missing much of the lower structure. This suggests that the model may be more sensitive to distinct textures like that of the growth lines shown in the bounding box of Figure 5.11 and the tabulae identified in the upper coral bounding box of Figure 5.9. The implied dependence of the network on distinct identifying textures introduces a significant, though not surprising, limitation to the generalisation ability of the model. If the model were to be deployed on samples from different locations, depositional environments, or geologic time, it is highly likely that the morphological variability of the grains, shells, or fossils in those samples could differ greatly from that of the training dataset. This variability might stem from evolutionary changes, differences in sedimentary processes, or regional ecological factors, resulting in forms that the model has not been trained to recognise. Furthermore, in samples subjected to a higher degree of diagenesis, where original textures and internal structures are largely lost, the preserved outlines may be the only remaining discernible features. These outlines could lack the detailed information the network has learned to associate with specific classes, leading to reduced performance or misclassification.

This limitation highlights a broader challenge in applying deep learning models to geological datasets: the inherent diversity and complexity of geological materials. Unlike standard object detection tasks where objects (e.g., cars or animals) have relatively consistent appearances, geological samples exhibit vast heterogeneity influenced by natural processes over millions of years. This highlights the need for training datasets that are representative of the full range of variability expected in deployment scenarios, as well as the potential value of transfer learning or domain adaptation techniques. These approaches could enable the model to adjust to new environments or conditions by leveraging knowledge from related domains while minimising the need for extensive retraining.

Additionally, it suggests that for applications requiring high generalisability, a focus on developing models capable of interpreting fundamental geometric features rather than texture-dependent cues may be more effective. For example, incorporating shape analysis, edge detection, or multi-scale feature extraction methods could help mitigate this dependency on specific textures, enhancing model robustness to variations in depositional settings or diagenetic overprints.

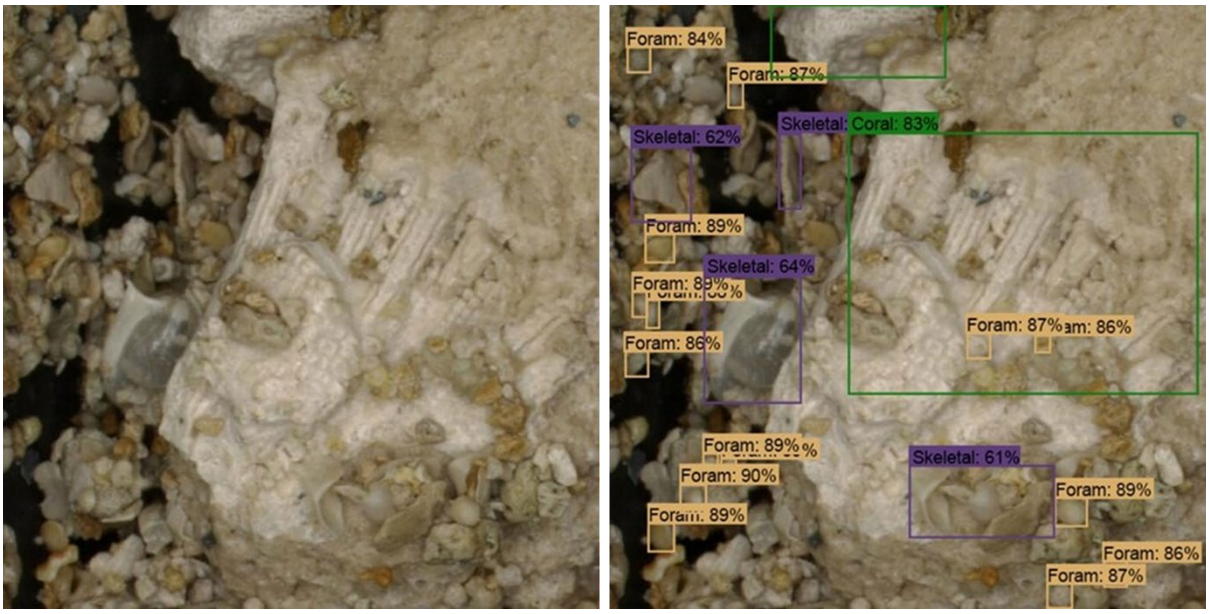

Figure 5.11: Example of predicted carbonate components for each identified object in Core 359-U1465C-8F-cc using Faster R-CNN. Only bounding boxes with a confidence threshold >60% are displayed.

5.4.1 Error analysis

Poor performance within the foraminifer class can be analysed further by considering the feature maps for each model (Figure 5.12). By visualising the intermediate feature representations across the different detectors, we can gain a better understanding of how the results were derived from the input data patterns.

It should be noted that the proposed visualisation method here only provides a preliminary analysis and explanation of the causes of errors in network performance from the feature map perspective.

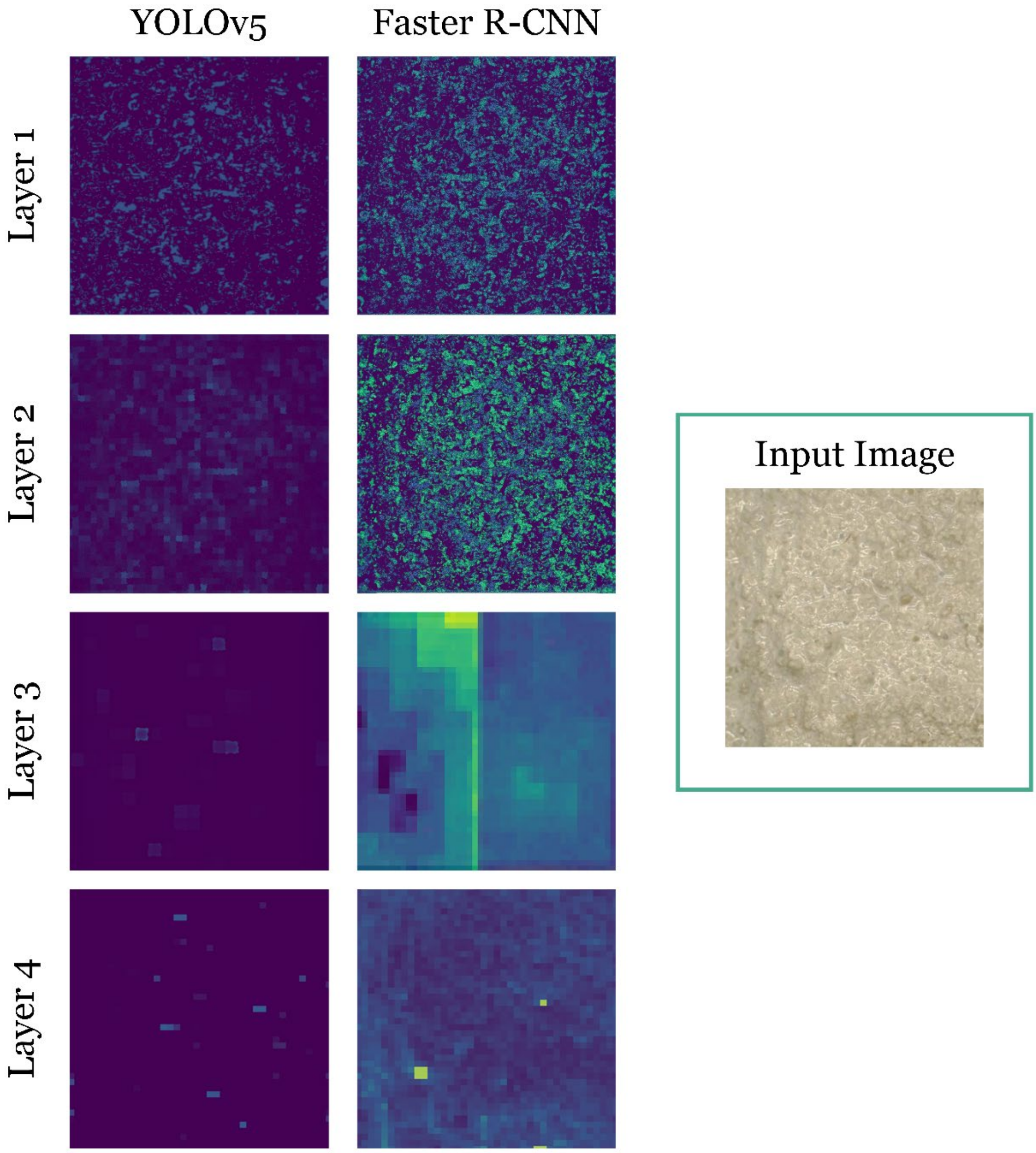

Figure 5.12: Visualisation of selected feature maps for YOLOv5 and Faster R-CNN of a foraminifera-rich packstone (Core 359-U1470A-15H-1). Following the 'viridis' colour map, the darker purple colourways indicate areas of lower importance, increasing through blues and greens to bright yellow, which indicates greater influence. Bright areas are the 'activated' regions, meaning the convolutional filter detected the pattern it was looking for.

The colour maps in the feature map visualisations represent the activation levels of different regions in the input image as processed by the convolutional layers of YOLOv5 and Faster R-CNN (Figure 5.12). These activations highlight areas that the network considers relevant based on its learned filters. Comparison of the image pixel values shows that the brightness and texture of the targets in Faster-RCNN are more visible to the naked eye, indicating a higher contrast with the background. In the visualisation for YOLOv5, we see a decrease in the overall pixel value, where the contrast between the target features and the background is greatly decreased to the point that the target patterns are almost invisible (Figure 5.12). The edge, brightness, and texture features are more obvious in Faster-RCNN, suggesting that the network can extract more convolution kernels of target information. This may partly explain why the detection effect of Faster-RCNN is significantly better than that of YOLOv5.

The challenge lies in visually connecting these abstract feature maps to specific details in the input image, particularly in deeper layers. For instance, while Layer 1 maps show relatively fine-grained activations that may align with edges or brightness variations in the input image, the activations in Layer 3, especially for Faster R-CNN, become more abstract and less clearly tied to any recognisable patterns. The yellow and green regions in the maps indicate higher activation, but it is unclear whether these regions correspond directly to foraminifera or are responses to unrelated textures or noise. The progression through layers demonstrates how the networks move from low-level feature detection (edges and textures) to high-level representations, but this abstraction complicates visual correlation.

For YOLOv5, we observe that the model is not sensitive to much of the image and very few features are extracted overall. We see from this that the majority of foraminifera occurrences may be missed altogether by the network. The Faster-RCNN architecture is more sensitive to occurrences, since we can observe more green and yellow areas compared to YOLOv5 (Figure 5.12). This suggests that Faster R-CNN extracts more features from the input image. However, this sensitivity does not necessarily translate to accuracy and the confidence and influence of the majority of the highlighted features are not strongly affecting the extracted visual pattern. Therefore, the model may miss instances that occur within the image, and we would expect lower recall values (Figure 5.13). The activations may highlight features that are irrelevant or unrelated to foraminifera, resulting in false positives. Without further validation, it is difficult to confirm whether the activated regions genuinely correspond to the presence of foraminifera or if the network is responding to other textures or background elements.

To address these challenges and improve validation, several additional methods could be employed. First, overlaying the feature maps onto the input image with ground-truth annotations (i.e., expert-marked locations of foraminifera) would allow a direct comparison between the activations and known patterns. This could provide a measure of how well the network is identifying relevant features. Second, techniques like Gradient-weighted Class Activation Mapping (Grad-CAM) could be used to produce heatmaps that show which parts of the input image contribute most to the classification decisions of the network (Selvaraju *et al.*, 2017). Unlike raw feature maps, Grad-CAM highlights regions of the input image that influence network predictions, implying whether these regions correspond to foraminifera. Third, incorporating background-only images (i.e., images with no objects) into the dataset during training could help the network learn to differentiate between meaningful patterns and noise, potentially reducing False Positives (FP). In the context of a carbonate core, background images could be those containing other artefacts that are not of interest, such as the ruler/scale, Styrofoam inserts, or the black background, or those containing no constituent grains, such as a mudstone. We recommend that adding around 0-10% background images could help to reduce FPs. For reference, the Microsoft COCO dataset has 1,000 background images, 1% of the total (Lin *et al.*, 2014). No labels are required for background images. Lastly, quantitative evaluations, such as recall and precision metrics computed against ground-truth labels, provide objective measures of model performance, complementing the qualitative insights from feature map visualisations (Figure 5.13). Together, these approaches could help to clarify the relationship between the feature maps, the input image, and network sensitivity, while providing a robust assessment of model accuracy and generalisability. For Faster-RCNN, the majority of classes present an AP_{50} of over 0.60, with only the aggregate and foraminifera classes falling below this (0.52 and 0.59, respectively) (Table 5.6; Figure 5.8). Precision-recall (PR) curves are graphical representations that can be used to evaluate the performance of classification models and are particularly useful in scenarios where the distribution of classes is imbalanced. Considering the PR curve in Figure 5.13, we observe the recall, i.e., completeness of predictions, is notably lower, suggesting that the model is actually missing instances for all classes.

Expanding on the earlier definitions in section 5.2.5, false negatives occur when true instances of the target class, such as foraminifera in this case, are overlooked by the model and incorrectly classified as background or another class. This is of particular concern in applications where missing key features could have significant consequences, such as in geological interpretations where undetected features might misrepresent classifications. Lower recall values highlight this issue, suggesting that the model struggles to detect all true instances within the dataset (Figure 5.13). A potential cause of this is imbalanced class distribution, where some classes are underrepresented compared to others. This imbalance could bias the model towards predicting the dominant classes, leading to an increased rate of false negatives.

On the other hand, false positives occur when the model incorrectly identifies background or unrelated textures as instances of the target class. This results in an inflated precision score, as it gives the illusion of fewer incorrect predictions relative to the correct positive predictions. False positives might stem from noisy data or features in the background that resemble the target class. In geological contexts, this could lead to overclassification of textures as being foraminifera-rich, skewing interpretations of the rock record. The issue of false positives could also be exacerbated by the lack of background images in the dataset, which limits the ability of the model to learn to distinguish between target and non-target features.

The PR curve shown in Figure 5.13 offers a detailed visualisation of the trade-off between precision and recall for different classes identified by the Faster-RCNN model with a 0.5 confidence threshold. This curve is particularly useful in evaluating model performance when dealing with imbalanced datasets, as it illustrates how well the model captures positive instances (recall) without incorrectly labelling negative instances as positive (precision).

Each line in the PR curve corresponds to a distinct class. The ideal performance would see a line maintaining high precision while achieving high recall (i.e., closer to the top-right corner of the plot). However, the variability among the lines indicates that the model performs differently across classes, with some classes achieving higher precision and recall than others. For instance, the "gastropod" class exhibits relatively better performance, with a significant region maintaining high precision and recall, whereas classes such as "aggregate" and "serpulid" show a sharp decline in precision as recall increases. This suggests that the model struggles to generalise well across all classes, potentially due to differences in feature complexity, class imbalance, or overlapping characteristics in the dataset.

The drop in precision for certain classes at higher recall values highlights the presence of false positives. As the model attempts to capture more instances of the target class, it also begins to incorrectly classify non-target instances, reducing precision. Conversely, the lower recall values for some classes imply that the model misses true positive instances, leading to false negatives. This indicates that the model has become more conservative in making positive predictions, leading to higher precision but lower recall. This may be related to the nature of the core samples, where the orientation of cut through components may vary greatly, leading to the model not identifying occurrences which differ from those learned in training set. The discussion of Figure 5.13 ties into the earlier points about the challenges of the model with false negatives and false positives. The variability in performance across classes indicates a need for further refinement in the training process. Strategies such as increasing the representation of underperforming classes through data augmentation or synthetic data generation could help balance performance. Additionally, incorporating negative examples, such as background-only images or those without foraminifera, would train the model to better recognise and ignore irrelevant features. Adjusting the decision threshold for classification could also help to balance precision and recall, depending on the specific application requirements (e.g., prioritising high recall in cases where missing instances is critical). By addressing the underlying causes of false negatives and positives, model performance could be significantly improved, enhancing both its reliability and applicability to geological datasets.

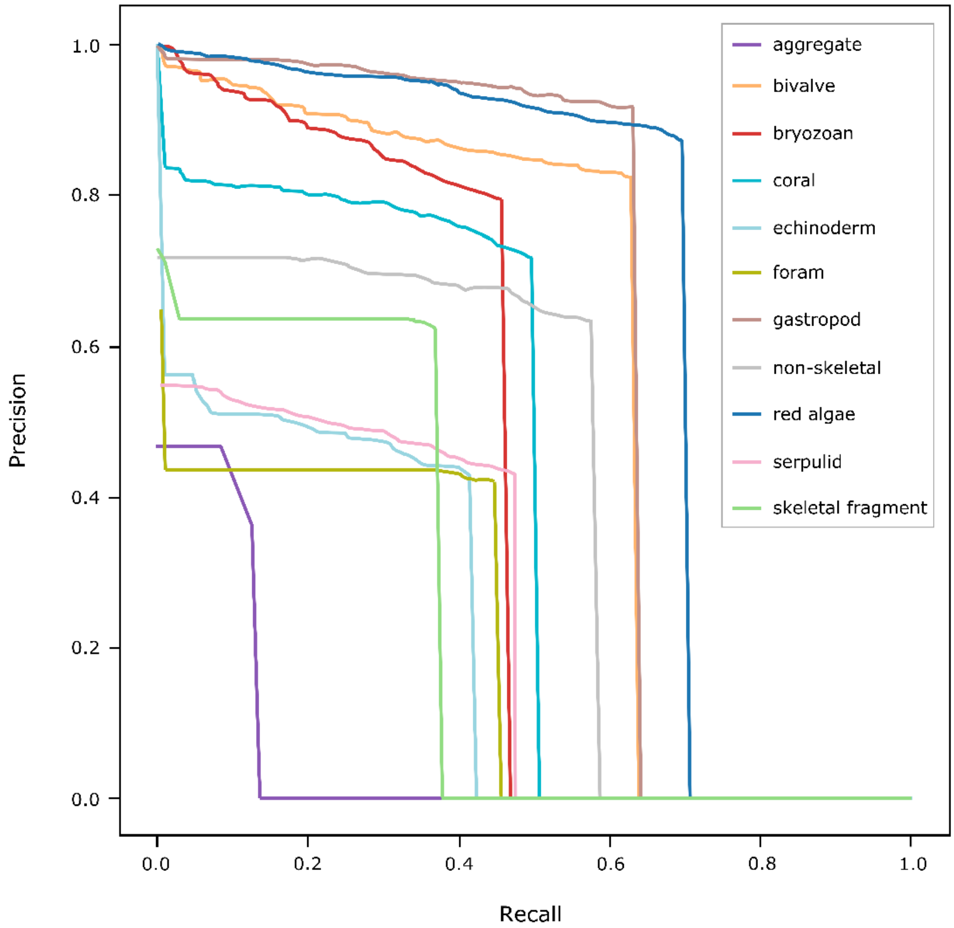

Figure 5.13: Precision-recall curve for Faster-RCNN with a 0.5 confidence threshold.

5.4.2 Comparison with human performance

To demonstrate the effectiveness of our best performing model, we draw comparison of human and machine learning (ML) performances on the same geologic material to help evaluate the algorithm success rates and provide an understanding of ML performance relative to that of a human expert. Conesa *et al.*, (2005) carried out an extensive biosedimentary and palaeoenvironmental study of the carbonate cores drilled at ODP sites 1196 and 1199 on the Southern Marion Plateau, North-Eastern Australia. Their work provides detailed percentage counts of biogenic components and fossil assemblages observed from core intervals and thin section samples. We deployed the trained Faster R-CNN model on to unseen data from Sites 1196 and 1199, which had a total inference time of just over 3 hours. As mentioned previously, it was ensured that no images of cores taken from these drill sites were included in the original dataset.

To draw direct visual comparison, we overlay the results of the model on the original microfacies biogenic components vs lithostratigraphic columns provided by the authors (Conesa *et al.*, 2005) (Figures 5.14 and

5.15). The data tables used to construct Figures 5.14 and 5.15 are provided in Appendix C. To enable direct comparison with the trained classes of the model, from Conesa *et al.* (2015) we have grouped the "small benthic foraminifers", "larger benthic foraminifers" and "planktonic foraminifers" into a singular "foraminifers" column, and the "red algae" and "Geniculate red algae" into a singular "red algae" column. The "aggregate", "non-skeletal" and "skeletal" classes from our dataset are not included in this comparison, as they do not feature in the original study by Conesa *et al.* (2015). The fill colours within plots indicate the degree of agreement between human and AI identification: overlap in quantification is indicated by blue, black indicates model results lower than human count, and yellow indicates model results exceeding human count (Figures 5.14 and 5.15). It should be noted that for the "green algae" column, there is no estimate provided by the ML model as this was not included in the training classes.

The general trend of the fossil assemblages throughout the cores is seen to be visibly similar between the ML and human frequencies (Figures 5.14 and 5.15). We see an initial abundance of diversity in the lower assemblages (Site 1196: Subunit IIIA; Site 1199 Subunit IIB) and general scattered occurrence of corals, suggestive of an inner-platform environment with patch reefs. The spike in coral frequency observed at Site 1196: Subunit IC is indicative of a coral reef environment. In the upper assemblages (Sites 1196 and 1199: Subunits IB–IA), we see a dominance of red algal facies that are devoid of other biogenic components, suggestive of an outer-platform environment.

Despite differences in the absolute count, for nearly all groups the overall trends are the same and these estimates are achieved in a fraction of the time needed for a human to do the same work. Areas for improvement are apparent in the echinoderm and foraminifera classes. For both Sites 1196 and 1199, the echinoderm estimate of the ML model is notably lower than the human estimate. For the foraminifera class, the ML estimate is considerably higher than that of the human expert. This may be related to the earlier observation of the model misclassifying other small core features as foraminifera.

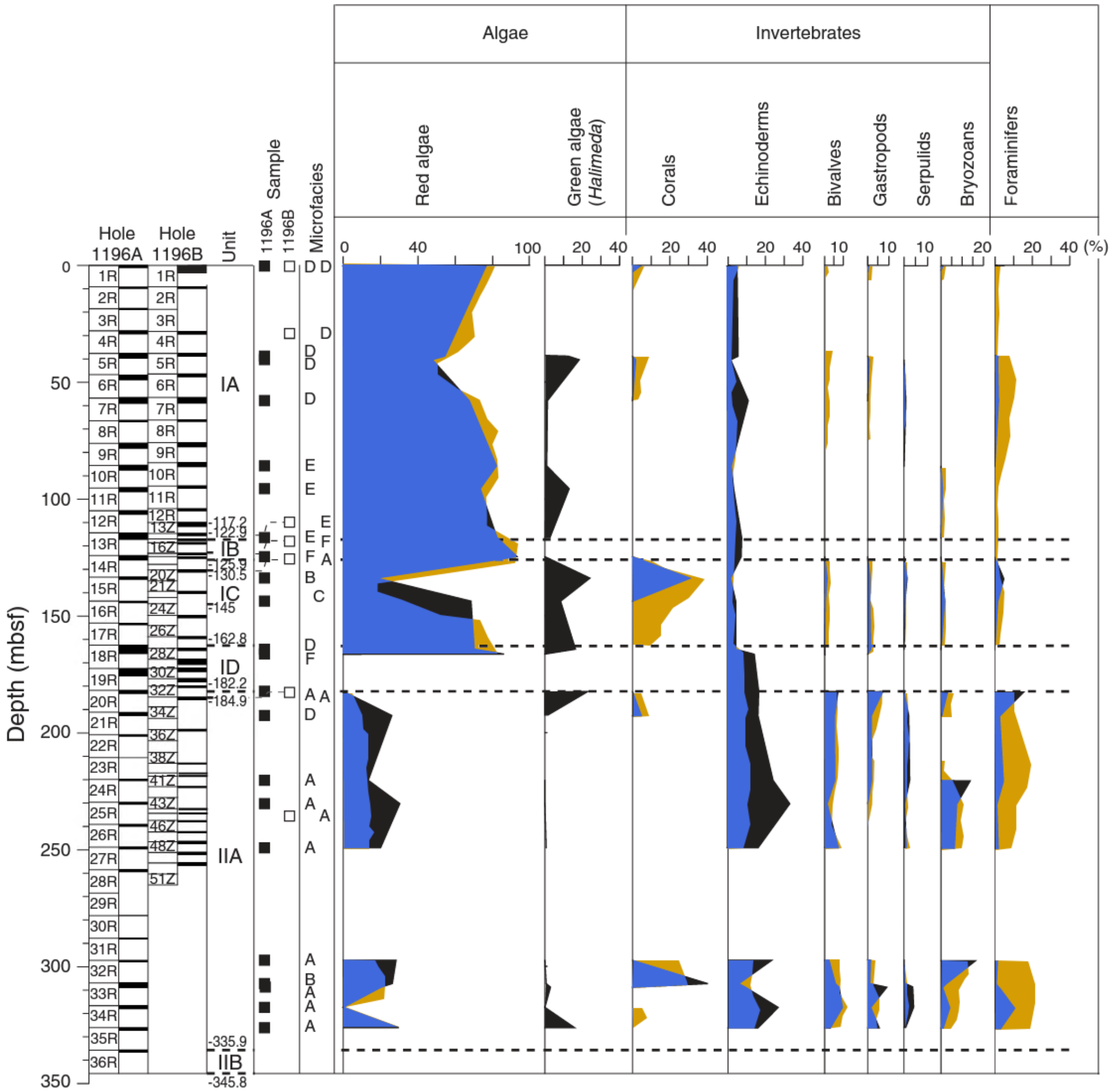

Figure 5.14: Performance comparison of Faster R-CNN with human interpretation from ODP Leg 194 Site 1196 (after Conesa *et al.*, 2005). Frequencies of biogenic components are shown against the Site 1196 lithostratigraphic column. Blue = overlap; Yellow = higher ML estimate; Black = higher human estimate.

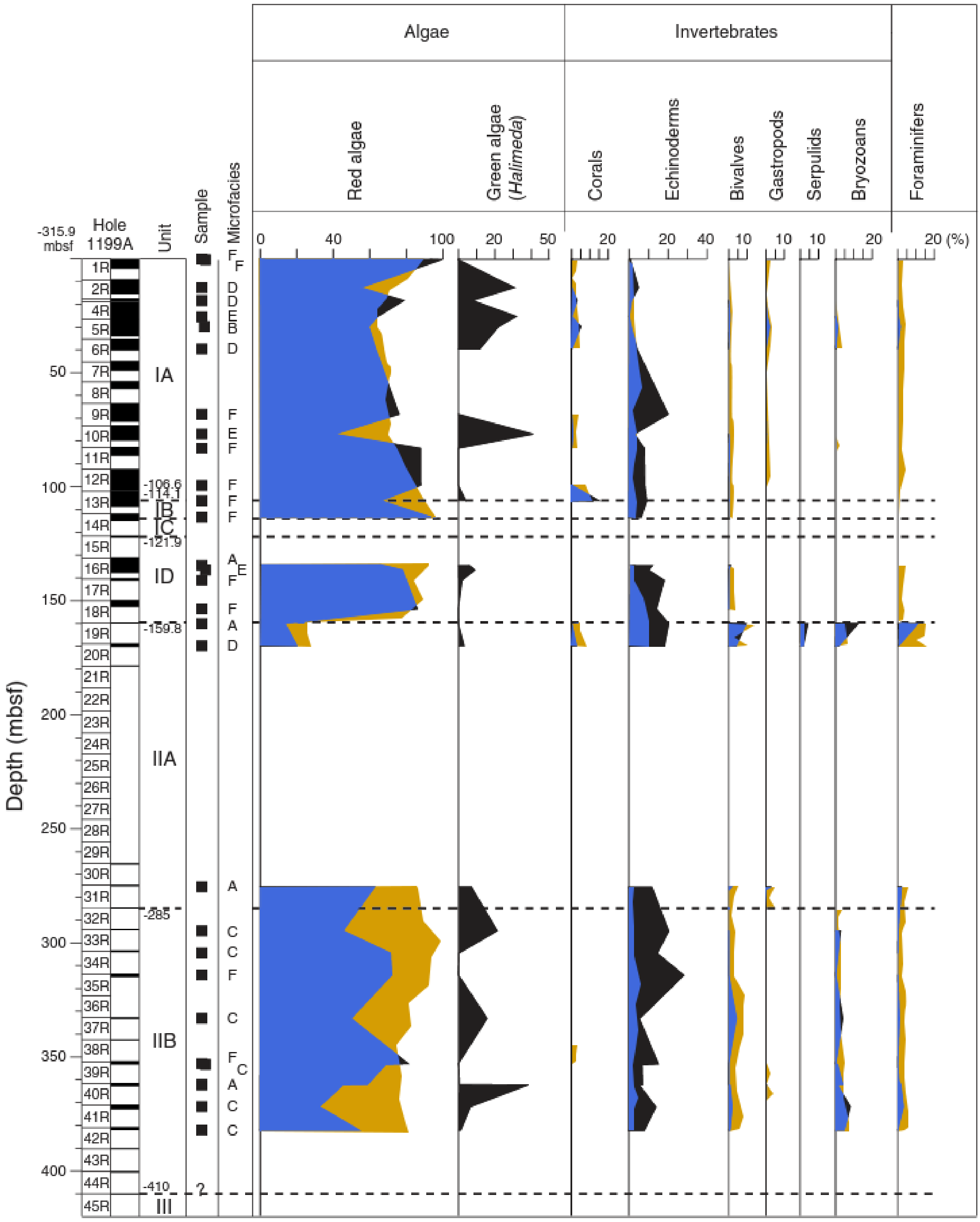

Figure 5.15: Performance comparison of Faster R-CNN with human interpretation from ODP Leg 194 Site 1199 (after Conesa *et al.*, 2005). Frequencies of biogenic components are shown against the Site 1199 lithostratigraphic column. Blue = overlap; Yellow = higher ML estimate; Black = higher human estimate.

5.4.3 Recommendations for future work

The performance of any supervised deep learning model that relies on labelled data will ultimately be limited by the quality of the annotations provided by humans. In some cases, the classification of certain features, such as distinguishing between different shell morphologies could be seen to be ambiguous and subjective, and thus may limit the result of the models. Having multiple domain expert annotators can provide a valuable approach to increasing the accuracy and reliability of the annotations. By comparing and contrasting the annotations of multiple experts, any discrepancies or inconsistencies in the data can be identified and corrected. This can help to ensure that the resulting model is trained on high-quality data that accurately reflects the complexities of the domain.

Deep learning-based methods are generally data hungry when performing a narrow task. These datasets typically comprise tens of millions of images, divided into hundreds of classes. Depending on the applications, collecting this quantity of data for geological applications, including core images, can be very difficult to achieve. Therefore, different algorithm optimisations may be necessary to better train the networks. Network optimisation is often achieved by applying various hyperparameters fine-tuning approaches such as weight initialisation, batch normalisation, and regularisation (Goodfellow et al., 2016).

Both one-stage and two-stage detectors show substantial potential in the task of component identification in carbonate core images. The main advantage of deep learning-based object detection here is the ability to accurately locate and identify the different grains individually. In addition, this method presents the potential for future real-time carbonate interpretations with improved accuracy and speed. However, there remain several key areas for improvement.

Mislabelling of classes may persist even when model loss is low. This issue can arise due to the complex nature of carbonate constituents and any diagenetic alteration present. While grains typically present as well-defined objects conducive to bounding box delineation, some less distinguishable morphological features may not be so easily isolated by bounding boxes. Further expanding the annotated training dataset to encompass these particular classes, along with other carbonate grains from geologic timescales, holds promise for refining the learning process, enhancing model performance, and fostering more accurate predictions. Additionally, a hierarchical classification system for the skeletal grains (e.g., at class or genus level) would be essential for applications in carbonate biostratigraphy, as well as further improving interpretations of depositional environments. Moreover, given the stratigraphic ranges of certain carbonate grains, particularly skeletal ones, a comprehensive training dataset spanning all stratigraphic ages would be required for developing a tool applicable to diverse temporal intervals.

Detection of smaller carbonate grains (<35px) presents challenges, as they are either undetected or are detected with low accuracy (<60% confidence) during the object detection task. This issue is consistent with similar problems encountered in using object detection networks for recognising real-world objects (Hu *et al.*, 2018). The root of this problem lies in the iterative combination of low-level features during the convolution process and the inherent bias of the model, which tends to prioritise larger objects within an image. To address this challenge, Hu *et al.* (2018) proposed a solution involving multiscale feature extraction from various convolution layers and integrating these features into a unified 1-D vector.

Developing on from object detection, future work should investigate the application of either supervised or unsupervised semantic and instance segmentation techniques (e.g., mask R–CNN; He *et al.*, 2017 or U-Net; Ronneberger *et al.*, 2015). This progression could offer a more reliable interpretation and potentially enhance the characterisation of grains in carbonate rocks. Additionally, in scenarios where expanding the size of the training dataset is not feasible, an alternative strategy to generate a significantly larger and statistically significant dataset is by leveraging Generative Adversarial Networks (GANs). As presented in Chapter 4, applications of GANs for synthetic data can improve the performance of the algorithms and provide more close-to-real world examples for the model to learn. Further development of these methods applied to object detection may hold promise for future investigations.

5.5 Conclusions

Carbonate sedimentology is complicated. Recent advancements in computer vision and deep learning, particularly convolutional neural networks (CNNs), have shown promise for developing tools to automate aspects of this process. This work has contributed to this emerging field by exploring the application and limitations of CNN-based object detection frameworks in identifying and quantifying multiple types of carbonate grains within core images.

This study is the first comparison of human and deep-learning carbonate biogenic component identification. From this comparison, we can see that the object detection method is capable of producing a frequency distribution that is similar in trend to that of the expert interpretation from Conesa *et al.* (2005). In the comparison between one-stage and two-stage detectors, the highest detection accuracies are obtained from the Faster R–CNN framework, with a mean average precision (mAP) of 0.73. While this evaluation is relative to human detection—a process that naturally involves some subjectivity—it demonstrates the ability of the network to replicate and, in some cases, refine the outcomes of traditional identification methods. This highlights the potential of Faster R–CNN to enhance detection accuracy while also underscoring the importance of further validation to address the inherent limitations of the human-based benchmark. The detection speed of Faster R-CNN was notably slower than that of YOLOv5. Depending on the nature of analysis, the one-stage framework may also be useful to perform real-time fully automated carbonate core analysis, e.g., simultaneously during line scanning of the core. However, for the presented application, the detection speed is not really an issue, and more importance is placed on the detection accuracy. Overall, these findings demonstrate the future potential of CNN-based object detection frameworks in automating the identification and quantification of carbonate grains in core images, providing a valuable tool for geologists and researchers in the field.

Chapter 6: Discussion and conclusions

6.1 General synthesis

Geological interpretations must grapple with the inherent complexity and variability of geological systems, introducing a level of uncertainty into the analyses. Data derived from geological sources, whether from surface outcrops or subsurface cores and well logs, often exhibit ambiguity, allowing for multiple valid interpretations. Variations in expertise, perspectives, and biases among geoscientists can lead to differing conclusions drawn from the same dataset. This subjectivity significantly contributes to uncertainty in geological descriptions and interpretations. As data availability continuously grows, it is important to process it with high speed and precision to further improve geological models. To address these challenges, in this thesis I sought to develop new approaches and frameworks towards automated carbonate core descriptions. To achieve this, I compiled novel datasets including textural carbonate core images and annotated carbonate grains. With these datasets, I exemplified a systematic approach to successfully using convolutional neural networks (CNNs) for textural classification using transfer learning on datasets of various sizes (Chapter 3) and I developed new frameworks to generate synthetic core images (Chapter 4) and to identify and quantify skeletal grains (Chapter 5). Collectively, the work presented in this thesis satisfies the aims and objectives outlined in Section 1.4 of the Introduction.

Four custom image datasets, all carbonate core images, were compiled specifically for the computer vision models (Objectives 1.1 and 4.1). Each dataset had a distinct purpose, serving a specific role within and across the three computer vision methodologies presented and the associated experiments. In each of the three research chapters, a comprehensive methodology delineating the pre-processing steps is provided. The deep learning models trained on these datasets were aimed at extracting a range of geological features including, but not limited to, textures, lithologies (Objective 1.2), and constituent grain types (Objective 4.2) from the core images (Chapters 3 and 5). The integration of synthetic carbonate textures with the original training dataset of core images (Chapter 4) has the potential to significantly enhance the accuracy of textural carbonate classification, even with limited data. By combining generated images with real core images, a CNN image classification model can more precisely predict Dunham classes based on the distinctive features present in each image. It was observed that by incorporating synthetic data, the sensitivity and specificity of the model increased from 67.2% and 65.4%, respectively, to 76.7% and 78.1%, respectively (Table 4.3). These improvements are particularly significant, as such substantial percentage increases in sensitivity and specificity are noteworthy in the field of deep learning, where performance gains are often incremental at best. This data augmentation approach not only enhances interpretative quality but can also improve robustness and generalisation ability by providing additional new data samples (Objectives 3.3 and 3.4).

Through the experiments conducted with the custom datasets developed in this thesis, initial evidence for the potential of computer vision in carbonate core classification has been presented. While the results demonstrate promise, particularly in terms of the ability to detect and classify carbonate textures and certain fossil types, challenges related to textural dependence, class imbalance, model performance across different depositional environments suggest that further refinement and testing are necessary. The findings indicate that while computer vision may be a valuable tool in this context, its full efficacy will depend on addressing these limitations and improving the ability of the models to generalise to more diverse datasets.

Across all three computer vision methods, as evidenced by the results presented in the individual research chapters, the variability and richness of the image datasets used for training, validation, and testing directly influenced model performance (Objective 1.4). For example, using images that had undergone data augmentation transformations, such as variations to brightness and contrast emulating images that may be captured under varying lighting conditions, contributed to a more diverse and extensive training set, resulting in improved predictions and generalisation capabilities (Table 4.3) (Objectives 3.3 and 3.4).

Across all three computer vision methods, as evidenced by the results presented in the individual research chapters, the variability and richness of the image datasets used for training, validation, and testing directly influenced model performance (Objective 1.4). For example, using images that had undergone data augmentation transformations, such as variations to brightness and contrast emulating images captured under varying lighting conditions, contributed to a more diverse and extensive training set, which enhanced predictions and generalisation capabilities (Table 4.3) (Objectives 3.3 and 3.4). However, the use of GAN-generated images, while offering the advantage of diversifying the dataset with synthetic examples, introduced some artefacts that may have posed challenges for model learning. These artefacts could negatively affect model performance, but they also provided an opportunity for the model to learn to differentiate between target and non-target features, potentially improving its robustness to variations and anomalies in real-world data.

The main aim of Research Question 1 was to develop a CNN-based transfer learning approach to core classification. This was addressed in Chapter 3 where, using three datasets of different sizes, I trained multiple different CNN architectures to classify carbonate textures according to the divisions of the modification Dunham classification scheme (Dunham, 1962; Embry and Klovan, 1971) (Objectives 1.1 and 1.2). The ability of each CNN architecture to classify the carbonate core images was then evaluated, systematically developing a performance comparison between these architectures on a complex geological dataset (Objective 1.3). In doing so, the impact of dataset size on the performance and reliable application of the models was highlighted, with the models trained on the small (7k) dataset showing a tendency to overfit (Objective 1.4). This is an important finding, as many current studies use deep learning algorithms on datasets with under 10,000 samples (Table 3.1). The highest overall accuracy of 92% was achieved by the Inception-v3 architecture when trained on the larger (104k) dataset with an 80:20 train-test split. When considering all evaluation metrics presented (Objective 1.4), for textural carbonate core classification, the Inception-v3 architecture was found to be the most suitable model for medium to large datasets, and the VGG19 architecture to be the most suitable for smaller datasets (Figure 6.1) (Objective 1.5). The results of Chapter 3 show great potential for the use of deep learning in automated carbonate classification from digital core images, where high level performance was achieved across all models, even in the case of limited and unbalanced datasets.

Given the importance of dataset size highlighted in Chapter 3, the main aim of Research Question 3 was to evaluate synthetic and traditional data augmentation approaches in improving CNN classification performance on a small dataset. This was, in part, addressed by the work conducted for Research Question 2, which aimed to condition generative models to create synthetic carbonate textures that were representative of real core images. I addressed these aims in Chapter 4 by using generative adversarial network (GAN) architectures to generate synthetic images from the training set of the 7k dataset (Objective 2.1). By evaluating Fréchet Inception Distance (FID) scores, image morphological features and validating the classes with a trained classification model, it was shown that GANs were able to generate synthetic images that could match key characteristic features observed in original carbonate textures (Objectives 2.2 and 2.3). The pretrained StyleGAN2-ADA model was able to produce the best examples for synthetic core image generation with a FID of 6.79 (Table 4.1). Visual inspection of the generated images did reveal the presence of artifacts in some images. However, it was decided not to exclude these images from the augmented dataset as they could potentially further improve model robustness (Objective 3.2).

To evaluate data augmentation techniques, two new datasets based on augmentations to the 7k dataset in Chapter 3 were created, with one containing the original training data plus augmented images from traditional transformations (Figure 4.1), and one containing the original training data plus the synthetic images generated by the pretrained StyleGAN2-ADA model (Objectives 3.1 and 3.2). The ability of these augmentation approaches to improve CNN classification performance was then evaluated, with both methods showing a significant increase in prediction ability and a reduction in overfitting (Objectives 3.3 and 3.4). Using traditional augmentations, sensitivity and specificity increased by 4.2% and 8.8%, respectively. Using the GAN-based augmentation, sensitivity and specificity increased by 9.5% and 12.7%, respectively (Table 4.3) (Objective 3.3). This novel application of GANs for the generation of synthetic carbonate core images shows significant potential for addressing the issues associated with limited or unbalanced data. Going forward, future work should aim to further validate the use of generative models as a data augmentation tool in geoscience workflows. I revisit this topic in Section 6.2 where I outline how future work could address this research challenge.

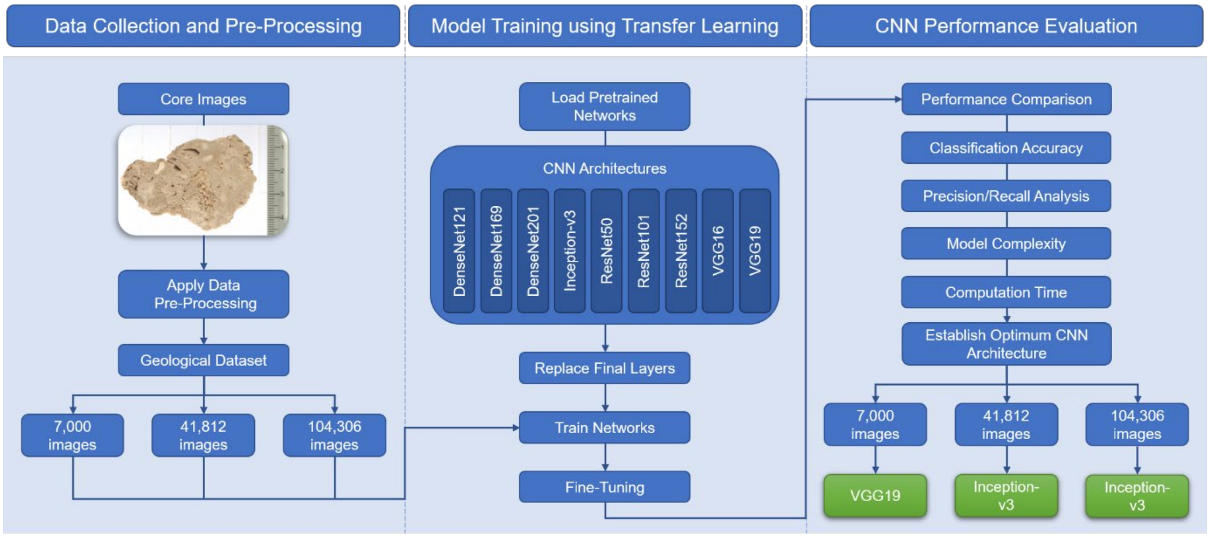

Figure 6.1: The carbonate core classification workflow proposed in this thesis.

Finally, the main aim of Research Question 4 was to develop an object detection approach to the identification and quantification of constituent grains in core images and I addressed this in Chapter 5. This aim was primarily motivated by the importance of characterising framework components in core observations of carbonate rocks to, ultimately, aid in the interpretation of depositional environments and marine ecosystem evolution, and the establishment of complex post-depositional and diagenetic histories (Memon *et al.*, 2023). To address this aim, an annotated dataset of carbonate constituent grains was created from the core images (Objective 4.1). Using pre-trained weights, a transfer learning approach was employed to evaluate one-stage (YOLOv5) and two-stage (Faster R-CNN) detectors under different feature extractors. Despite the current popularity of one-stage detectors in the wider computer science community, the results of Chapter 5 revealed the Faster R-CNN with ResNet50-FPN backbone to provide the most robust performance, achieving 0.73 mean average precision (mAP) (Objectives 4.2 and 4.3). Despite superior detection accuracy, detection speed was notably slower compared to the one-stage YOLOv5 framework (Tables 5.4 and 5.5). This means the two-stage detector would be less convenient for many real-time applications but for this core analysis application, it was concluded that the accuracy of detection is a more crucial factor, as is also the case for the medical domain (e.g., Yang and Yu, 2021; Sahin *et al.*, 2023; Xu *et al.*, 2023) (Objective 4.3). By visualising the intermediate feature representations across the different detectors, it was noted that the edge, brightness, and texture features were more obvious in Faster-RCNN than in YOLOv5, suggesting that the network can extract more convolution kernels of target information (Objective 4.4). This may partly explain why the detection effect of Faster-RCNN was significantly better than that of YOLOv5.

To evaluate the performance of the best-performing model presented in Chapter 5, comparisons were made between the results of the deep learning algorithm and human expert classifications on the same geological material. This comparison, based on the work of Conesa et al. (2005), who conducted a detailed biosedimentary and palaeoenvironmental analysis of carbonate cores from ODP sites 1196 and 1199 in the Southern Marion Plateau, provided insights into the success rates of the machine learning algorithm relative to human performance (Objective 4.5). The general patterns in fossil assemblage distributions across the cores, as identified by both the human expert and the machine learning model, showed visually similar trends, particularly in the relative abundance of different fossil groups (Figures 5.14 and 5.15). In this context, trend refers to the overarching pattern of frequency distributions, such as the relative proportion of each fossil type across the core samples.

However, despite these similarities, some misclassification errors were present in the model output. To more rigorously assess the degree of agreement between human and machine classification, further statistical analysis, such as correlation coefficients or Kappa statistics, could be performed to quantify this similarity and evaluate whether it is statistically significant. While the results demonstrate that the machine learning model can replicate general trends in fossil distributions, the accuracy of the model in terms of exact counts and classification accuracy needs further refinement.

Although the algorithm provides a much faster alternative to human analysis, any claim of "exemplifying efficacy" would require a more critical consideration. The deep learning model, while promising, still needs improvements in accuracy before it can fully complement or replace human experts in practical geological analyses. Nevertheless, it serves as a useful tool for quickly identifying large-scale patterns and trends, offering significant potential for future work in geosciences.

As a whole, the three research chapters in this thesis present new approaches and frameworks that enable us to gain sophisticated and quantitative insights into the importance of deep learning as a data-driven approach to deriving predictive models from observational data. Importantly, the work presented in each of these research chapters complements one another and, further, opens doors to a number of new research trajectories. Overall, this thesis introduces a novel methodology for addressing interpretational biases in geological practices, accomplished through the development of deep learning models capable of reliably extracting information from digital core images for objective interpretations.

6.2 Challenges for computer vision in geoscience

The three research chapters in this thesis can be read as standalone studies but, together, they present a deep learning-based approach to carbonate core description. There are a series of challenges which are common to some, if not all, chapters, and here I discuss a selection of these in detail.

One of the most prevalent challenges encountered in this thesis, impacting the performance of all the computer vision models presented, is the availability and variability of data. Lack of data is acknowledged as one of the biggest and most common challenges when building custom computer vision models (Aldoseri *et al.*, 2023). In the case of insufficient data, training a model to effectively identify and classify complex, rare, and domain-specific objects becomes challenging. Training of the comparatively smaller dataset in Chapter 3 demonstrated the significant impact on prediction performance. The citations for the publication from Chapter 3 extend well beyond Earth and Planetary Sciences into machine learning applications for e.g., molecular biology, chemical engineering, medicine, agricultural and biological sciences, materials science, mathematics, and decision science, demonstrating the widespread impact of this research to the broader computer science community. The challenge of insufficient data has been partially addressed in Chapter 4 through the incorporation of synthetic data generated by GANs and the implementation of standard data augmentation techniques within the training dataset.

Data quality represents another pertinent concern in the realm of computer vision. Despite the availability of data, its quality may not always meet the requisite standards, thereby potentially hindering the performance of a computer vision model. Instances of poor image quality, such as blurriness, low contrast, or the presence of artifacts, can pose challenges for accurately classifying or identifying objects of interest. The training datasets for this thesis were compiled using high-resolution scanned images, ensuring that image quality would meet the required standards of the models. To improve the robustness of the models to lower image quality, which may be encountered in future deployment on new data, data augmentation techniques, such as those to introduce noise or blurring to images, can be employed.

Certain domains, such as geoscience, may pose unique challenges that necessitate tailored approaches when developing custom computer vision models. For instance, in the case of core images, accurately extracting geological features and managing complex images or subtle variations in texture may require the deployment of multiple models. Moreover, the scale of geological features could present a distinct challenge for the computer vision models examined in this thesis. Although not directly addressed in this thesis, the integration of the Feature Pyramid Network (FPN) in the object detection models for Chapter 5 could help to accommodate the variability in scales of geological data. All images in the training datasets for this thesis were pre-processed to be of the same scale, but wider application of the developed frameworks would need to account for images with diverse scale variations. Additionally, the annotations of grains for the object detection task would need to embed scale information within the models.

Annotation and labelling of images pose another significant challenge, especially if handling extensive datasets or complex annotations. Maintaining consistency and accuracy throughout the labelling and annotation process is crucial to ensure the production of high-quality training data. Considerable time and effort were dedicated to image labelling and data preparation and pre-processing, constituting a fundamental contribution to the thesis. Specifically, the process involved manually annotating images with bounding boxes and class labels, followed by extensive pre-processing to ensure consistent formatting and alignment. This stage took approximately 6 months for each research chapter, including time spent performing quality checks to ensure that the data was correctly annotated and formatted for model training. The diverse nature of geological formations introduces further complexities, particularly when differentiating between visually similar rock lithologies of varying sizes and morphologies.

Additionally, the involvement of expert geologists is imperative to validate the work of annotators, thereby guaranteeing the provision of precise information for incorporation into the computer vision models. Having multiple domain expert annotators can also provide a valuable approach to increasing the accuracy and reliability of the annotations. By comparing the annotations of multiple experts, any discrepancies or inconsistencies in the data can be identified and corrected. This can help to ensure that the resulting model is trained on high-quality data that accurately reflects the complexities of the domain. While expert geologists were not directly involved in multiple rounds of annotation for this work, the visual core descriptions from the original expedition scientists were used as a guide for classifications in Chapter 3, helping to ensure data quality. This provided expert-level insight into the labelling, though, as noted, multiple expert annotators were not employed to compare interpretations and resolve discrepancies. Nevertheless, this approach still contributed valuable insights for training the model on high-quality data, reflective of the complexities of the geological domain.

Model architecture plays a pivotal role in determining the performance and predictive capabilities of computer vision models. Choosing the most appropriate architecture often entails experimentation with various models and hyperparameters, presenting a considerable challenging effort. Chapters 3-5 of this thesis exemplify this notion, where a significant portion of the results presented therein revolve around identifying the most suitable configuration for each model to address the specific geological challenges at hand. The process involved evaluating different image classification, object detection, and GAN architectures to determine which performed best for each deep learning application to carbonate core images. For example, hyperparameters such as batch size, learning rate, and optimiser were adjusted to assess their influence on accuracy and inference times. Additionally, different data augmentation techniques, such as brightness and contrast adjustments, were tested to improve model generalisation in the context of variable lighting and image quality. The broad significance of this challenge to computer vision applications is again exemplified by the citations for the publication from Chapter 3, which extend well beyond the domain of geoscience. These citations demonstrate how the issues addressed in this thesis resonate with broader applications in fields where accurate, efficient image analysis is critical.

Finally, the training of the computer vision models, particularly the generative models in Chapter 4, proved to be computationally demanding, presenting a challenge due to limited resources. Transfer learning approaches were used to help mitigate this but, in certain cases, model training spanned up to a week, impeding swift prototyping and iteration. Computational intensity depended on factors such as dataset size, the high resolution of the images, and the complexity of the architectures used. Consequently, the computational resources needed for training and evaluating the models varied and can incur significant computational costs. While the metrics used in this thesis are mainly inspired by those best suited to core descriptions, additional criteria as in Chapter 3, such as the time to sample a new realisation, may be of great importance in practical applications. Although this thesis exclusively evaluated GANs as a generative model for synthetic data, selection of the model or training formulation should also be considered as a hyperparameter, necessitating further studies to determine the optimal model based on the relevant application metrics. These studies could use approaches such Bayesian optimisation for hyperparameter tuning of deep neural networks (e.g., Bergstra et al., 2011; Snoek et al., 2012), or automated machine learning methods (Auto-ML) (Hutter et al., 2019), including neural architecture search (NAS) (Elsken et al., 2018).

6.3 Recommendations for future work

With the growing interest in automatization and data recognition, machine learning and deep learning algorithms will continue to play an increasingly important role in the geosciences. In this thesis, I developed new approaches which advance our ability to objectively describe and quantify digital carbonate core images. Importantly, I have identified a number of future research avenues which could directly build on this work and are important next steps to further advance our understanding.

The image datasets and deep learning frameworks implemented in this thesis can be established as a benchmark for future research endeavours in geological applications of computer vision. These models could serve as a baseline against which the performance of alternative algorithms can be compared. Moreover, the custom datasets could be refined by incorporating additional lithologies across geologic timescales or introducing further variations of the existing classes to enrich variability and elevate model robustness. As evidenced by the findings in Chapters 3 and 4, training a classification model with a more extensive and diverse dataset can yield notably enhanced prediction results, as well as improved generalisation ability. The study conducted in Chapter 4 demonstrated that GANs are capable of generating realistic carbonate textures that can be used to artificially increase the size of a dataset. This approach could be extended to address the issue of class imbalance, whereby synthetic data instances could be generated for lower sampled classes. This has the potential to significantly improve the performance of the algorithms by providing more close-to-real world examples for the model to learn from.

As this thesis mainly focused on lithological classification, sedimentary structures with the samples, such as crossbedding, lamination, or bioturbation, were not considered. Future development of the frameworks presented here could involve the inclusion of structures, allowing refinement of the lithofacies classification based on more complex sedimentary structures. Comparable with the approach presented by Hall and

Govert (2016), through the identification of bedding and lamination planes, the azimuth and dip of such planes could be used for the calculation of true thicknesses. For carbonate rocks, this could help to better infer sedimentation rates, gradual changes in environmental conditions, and cyclic sedimentation patterns (Flügel and Munnecke, 2010).

Another concept could involve leveraging the outlined computer vision models and workflows to extract geological features from 3-D scanner and photogrammetry data, rather than solely on 2D images. Integrating the additional 3-D data into these workflows could potentially enhance the accuracy of model predictions. This is attributed to the spatial features offered by photogrammetry data, which could facilitate a more explicit consideration of the scale of the geological features.

This thesis focused on exploring the feasibility of extracting features from carbonate core images only. In addition to the methods presented in this thesis, future work could also focus on simulating 3-D rock properties models (e.g., porosity and permeability) at in-situ reservoir conditions by rigorously integrating data acquired at smaller scale, such as using micro-CT scanners, with x-ray computed tomography (XCT) images of cores using appropriate upscaling methods. Since the workflow presented in this thesis is aimed at improving our ability to objectively interpret core images, the frameworks developed work to capture fine lithological changes present in the cores. Since these approaches could help to reduce uncertainty in reservoir models, future work should, therefore, focus on the integration of such high-resolution core images with well log data or image logs from electrical imaging tools, such as FMS (Formation MicroScanner) and FMI (Fullbore Formation MicroImager), to further enrich the core analysis. The validity of an integrated system is demonstrated by the practicality of employing machine learning for lithological classification from well logs (e.g., Hall, 2016; Valentín *et al.*, 2019; Mubarak and Koeshidayatullah, 2023) and exploring the potential of combined colour and texture characteristics of core data with data logs for lithological analysis (Gonzalez *et al.*, 2020). Understanding how to integrate these data is crucial for preserving essential information across various model scales and preventing the dismissal of potentially crucial data points as outliers, particularly when they are significant at relevant scales. This could be useful not only in the oil and gas industry, but also for carbon capture and storage or geothermal projects.

Finally, the workflow developed in this thesis could be extended to other types of lithologies, such as siliciclastic systems or volcanic rocks from geothermal reservoirs. In addition to the recent and ongoing advances in technology, with the availability of new XCT scanners enabling the acquisition of higher resolution images in 3-D, the workflow could also be adapted to mineral exploration cores.

6.4 Conclusions

In this thesis, I have developed a core description workflow that leverages the current advancements in computational power, utilising large datasets of digital core images and deep learning algorithms for computer vision. This approach aims to enhance the efficiency and accuracy of carbonate core classification, building on established methods while integrating modern computational techniques to address challenges specific to the geoscience domain. The purpose of this research was to lay the foundations for a deep learning approach to digital core interpretation that can revolutionise outdated routines and reduce the associated time, effort, and costs of such methodologies. Central to this endeavour was the use of convolution neural networks (CNNs) for image classification and object detection to analyse high resolution scans of carbonate cores, a research area currently underexplored compared to other data modalities, such as well logs and seismic data. The main investigations were extracting information on lithological textures and biogenic components, and addressing the use of synthetic data in geological studies, as summarised below:

Chapter 3 investigated the ability of different deep CNN architectures to classify carbonate core images, thereby systematically developing a performance comparison between the architectures on three complex geological datasets with orders of magnitude difference in data quantity. Following a transfer learning approach, the learned parameters from the pre-trained CNN architectures were used to classify the core images according to the modified Dunham Classification, where classes are defined based on textural observations (Dunham, 1962; Embry and Klovan, 1971). The results demonstrated the great potential of deep learning in automated carbonate classification from digital core images, where high level performance was achieved across all models, even with limited and unbalanced datasets. The highest overall accuracy of 91.59% was achieved by the Inception-v3 architecture, when trained on the larger dataset of 104,306 images. For textural carbonate core classification on datasets of varying sizes, the Inception-v3 architecture was found to be the most suitable model for medium to large datasets, and the VGG19 architecture to be the most suitable for smaller datasets. Furthermore, it was shown that even when using transfer learning the size of the dataset plays a key role in the performance of the models, with those trained on the smaller datasets showing a strong tendency to overfit.

Chapter 4 explored the application of generative adversarial networks (GANs) for synthetic image generation of carbonate textures, as a means of artificially increasing dataset size. Through the evaluation of Fréchet Inception Distance (FID) scores, image morphological features and validating the classes with a trained classification model, it was shown that the synthetic images generated by the StyleGAN2-ADA model were able to match key characteristic features of the carbonate textures in the original training dataset. It was also demonstrated that these generated images could be used for data augmentation to improve the performance of CNNs for core image classification. Furthermore, when compared to traditional data augmentation techniques, the addition of synthetic data to the original training dataset was shown to better reduce overfitting.

Chapter 5 focused on the application of CNN-based object detection algorithms to identify and quantify skeletal components in carbonate core images. The use of object detection in addition to the image classification task of Chapter 3 presents a robust method that is closer to the human approach to carbonate description. The results demonstrated the promising potential of deep learning in automating core description and interpretation, where the best performing model achieved 73% mean average precision (mAP) even with limited training data. Although the highest detection accuracy was obtained by the two-stage Faster R-CNN framework, detection speed was notably slower compared to the one-stage YOLO-v5 framework. However, for this core analysis application, the accuracy of detection is more crucial. For future wider applications, the one-stage detector may be more appropriate for real-time analyses.

The investigations and findings of this thesis achieved the key aims outlined in Section 1.4.1. The successful extraction of different rock features from core images, including textural lithologies in Chapter 3 and carbonate grains in Chapter 5, indicated the promising potential of CNNs in core image analysis. In addition, this thesis explored different approaches to improve model performance in the case of a limited dataset and proposed that synthetic data augmentation generally works better than traditional data augmentation procedures (e.g., rotation, reflection, and noise addition). Furthermore, this thesis discussed the various future pathways that could be considered to further advance the application of deep learning to solve core analysis and geoscience-related problems, in general. Compiling further datasets across geologic timescales and different lithologies will serve as a necessary step towards improving the prediction capability and the ability of the models presented to generalise and perform a more robust fully automated carbonate core description. Overall, this thesis makes a valuable contribution to the holistic understanding of the importance of deep learning in geoscience, particularly in carbonate sedimentology and core interpretation, and lays a foundation for further research and development to accelerate this field.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur. M., et al. (2016). Tensorflow: a system for large-scale machine learning. 12^{th} Symposium on Operating Systems Design and Implementation (OSDI), 265-283.

Abdel-Hamid, O., Mohamed, A.R., Jiang, H., Deng, L., Penn, G., Yu, D. (2014). Convolutional Neural Networks for Speech Recognition. IEEE/ACM Transactions on Audio, Speech, and Language Processing 22(10), 1533-1545. doi: 10.1109/TASLP.2014.2339736.

Adobe Inc. (2020). Adobe Photoshop 2020. Available from https://www.adobe.com/uk/products/photoshop.html.

Aggarwal, R., Sounderajah, V., Martin, G., Ting, D.S., Karthikesalingam, A., King, D., Ashrafian, H., Darzi, A., (2021). Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis. NPJ digital medicine, 4(65). doi: 10.1038/s41746-021-00438-z.

Al-Anazi, A., Gates, I.D. (2010). On the Capability of Support Vector Machines to Classify Lithology from Well Logs. Natural Resources Research 19, 125-139. doi: 10.1007/s11053-010-9118-9.

Aldughayfiq, B., Ashfaq, F., Jhanjhi, N. Z., Humayun, M. (2023). YOLOv5-FPN: a robust framework for multi-sized cell counting in fluorescence images. Diagnostics 13(13), 2280. doi: 10.3390/diagnostics13132280.

Al-Mudhafar, W.J. (2017). Integrating well log interpretations for lithofacies classification and permeability modeling through advanced machine learning algorithms. Journal of Petroleum Exploration and Production Technology 7, 1023-1033. doi: 10.1007/s13202-017-0360-0.

Al-Mudhafar, W.J., Abbas, M.A., Wood, D.A. (2022). Performance evaluation of boosting machine learning algorithms for lithofacies classification in heterogeneous carbonate reservoirs. Marine and Petroleum Geology 145(105886), 1-21. doi: 10.1016/j.marpetgeo.2022.105886.

Al-Ramadan, K., Koeshidayatullah, A., Cantrell, D., Swart, P. K. (2020). Impact of basin architecture on diagenesis and dolomitization in a fault-bounded carbonate platform: outcrop analogue of a pre-salt carbonate reservoir, Red Sea rift, NW Saudi Arabia. Petroleum Geoscience 26(3), 448-461. doi: 10.1144/petgeo2018-125.

Alsina, E.F., Chica, M., Trawinski, K., Regattieri, A. (2017). On the use of machine learning methods to predict component reliability from data-driven industrial case studies. The International Journal of Advanced Manufacturing Technology 94, 2419-2433.

Alzubaidi, F. Mostaghimi, P., Swietojanski, P., Clark, S.R., Armstrong, R.T. (2021). Automated lithology classification from drill core images using convolutional neural networks. Journal of Petroleum Science and Engineering 197, 107933. doi: 10.1016/j.petrol.2020.107933.

Alzubaidi, F., Makuluni, P., Clark, S. R., Lie, J. E., Mostaghimi, P., Armstrong, R. T. (2022). Automatic fracture detection and characterization from unwrapped drill-core images using mask R–CNN. Journal of Petroleum Science and Engineering 208, 109471. doi: 10.1016/j.petrol.2021.109471.

Alzubaidi, F., Nalinakumar, H., Clark, S.R, Lie, J.E., Mostaghimi, P., Armstrong, R.T. (2023). Improved Fracture Segmentation from Unwrapped Drill-Core Images Using an Innovative Two-Stage Segmentation Approach. Mathematical Geosciences 55, 829-857. doi: 10.1007/s11004-023-10053-1.

Alzubaidi, L., Zhang, J., Humaidi, A.J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M.A., Al-Amidie, M., Farhan, L. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. Journal of big Data 8,(53) 1-74. doi: 10.1186/s40537-021-00444-8.

Angermann, C., Bereiter-Payr, J., Stock, K., Degenhart, G., Haltmeier, M. (2024). Three-Dimensional Bone-Image Synthesis with Generative Adversarial Networks. Journal of Imaging 10(12), 318. doi: 10.3390/jimaging10120318.

Ashena, R., Thonhauser, G. (2018). Coring methods and systems. Berlin: Springer, XXVI, 243. doi: 10.1007/978-3-319-77733-7.

Aubert, O., Droxler, A.W. (1992). General Cenozoic evolution of the Maldives carbonate system (equatorial Indian Ocean). Bulletin Des Centres de Recherches Exploration-Production Elf-Aquitaine 16(1):113-136.

Baldwin, J.L., Bateman, R.M., Wheatley, C.L. (1990). Application of a neural network to the problem of mineral identification from well logs. The Log Analyst 3, 279-293.

Baraboshkin, E.E., Ismailova, L.S., Orlov, D.M., Zhukovskaya, E.A., Kalmykov, G.A., Khotylev, O.V., Baraboshkin, E.Y., Koroteev, D.A. (2020). Deep convolutions for in-depth automated rock typing. Computers & Geosciences 135, 104330.

Barchiesi, M. A., Colladon, A. F. (2021). Big data and big values: When companies need to rethink themselves. Journal of Business Research, 129, 714-722. doi: 10.1016/j.jbusres.2019.10.046.

Bay, H., Tuytelaars, T., Van Gool, L. (2006). SURF: Speeded Up Robust Features. In: Leonardis, A., Bischof, H., Pinz, A. (eds) Computer Vision – ECCV 2006, Lecture Notes in Computer Science 3951. Springer, Berlin, Heidelberg. doi: 10.1007/11744023_32.

Bechtold, A., Paar, G., Garolla, F., Nowak, R., Fritz, L., Traxler, C., Sidla, O., Koeberl, C. (2023). Planetary scientific target detection via deep learning: A case study for finding shatter cones in Mars rover images. Meteoritics & Planetary Science 58(9), 1274-1286. doi: 10.1111/maps.14054.

Becker, B., Vaccari, M., Prescott, M., Grobler, T. (2021). CNN Architecture Comparison for Radio Galaxy Classification. Monthly Notices of the Royal Astronomical Society 503(2), 1828-1846.

Belopolsky, A.V., and Droxler, A.W. (2004). Seismic expressions of prograding carbonate bank margins: middle Miocene, Maldives, Indian Ocean. In: Eberli, G.P., Masaferro, J.L., and Sarg, J.F. (Eds.), Seismic Imaging of Carbonate Reservoirs and Systems. AAPG Memoir 81, 267-290.

Benny, Y., Galanti, T., Benaim, S., Wolf, L. (2021). Evaluation metrics for conditional image generation. International Journal of Computer Vision 129, 1712-1731. doi: 0.1007/s11263-020-01424-w.

Bergen, K.J., Johnson, A.P., de Hoop, M.V., Beroza, G.C. (2019). Machine learning for data-driven discovery in solid Earth geoscience. Science 363, 1299. doi: 10.1126/science.aau0323.

Bestagini, P., Lipari, V. Tubaro, S. (2017). A machine learning approach to facies classification using well logs. SEG Technical Program Expanded Abstracts 2017, 2137-2142. doi: 10.1190/segam2017-17729805.1.

Betlem, P., Birchall, T., Ogata, K., Park, J., Skurtveit, E., Senger, K. (2020). Digital drill core models: structure-from-motion as a tool for the characterisation, orientation, and digital archiving of drill core samples. Remote Sensing 12(2), 330. doi: 10.3390/rs12020330.

Betzler, C., Hübscher, C., Lindhorst, S., Reijmer, J.J.G., Römer, M., Droxler, A.W., Fürstenau, J., Lüdmann, T. (2009). Monsoon-induced partial carbonate platform drowning (Maldives, Indian Ocean). Geology 37(10), 867-870. doi: 10.1130/G25702A.1.

Betzler, C., Fürstenau, J., Lüdmann, T., Hübscher, C., Lindhorst, S., Paul, A., Reijmer, J.J.G., Droxler, A.W., (2013a). Sea-level and ocean-current control on carbonate-platform growth, Maldives, Indian Ocean. Basin Research 25(2), 172-196. doi: 10.1111/j.1365-2117.2012.00554.x.

Betzler, C., Lüdmann, T., Hübscher, C., and Fürstenau, J. (2013b). Current and sea-level signals in periplatform ooze (Neogene, Maldives, Indian Ocean). Sedimentary Geology 290, 126-137. doi: 10.1016/j.sedgeo.2013.03.011.

Betzler, C., Lindhorst, S., Lüdmann, T., Weiss, B., Wunsch, M., Braga, J.C. (2015). The leaking bucket of a Maldives atoll: implications for the understanding of carbonate platform drowning. Marine Geology 366, 16-33. doi: 10.1016/j.margeo.2015.04.009.

Betzler, C., Eberli, G.P., Alvarez Zarikian, C.A., Alonso-García, M., Bialik, O.M., Blättler, C.L., Guo, J.A., Haffen, S., Horozal, S., Inoue, M., Jovane, L., Kroon, D., Lanci, L., Laya, J.C., Ling Hui Mee, A., Lüdmann, T., Nakakuni, M., Nath, B.N., Niino, K., Petruny, L.M., Pratiwi, S.D., Reijmer, J.J.G., Reolid, J., Slagle, A.L., Sloss, C.R., Su, X., Swart, P.K., Wright, J.D., Yao, Z., Young, J.R. (2017). Expedition 359 Summary. In: Betzler, C., Eberli, G.P., Alvarez Zarikian, C.A., and the Expedition 359 Scientists, Maldives Monsoon and Sea Level. Proceedings of the International Ocean Discovery Program, 359: College Station, TX (International Ocean Discovery Program). doi: 10.14379/iodp.proc.359.101.2017.

Bhattacharya, S., Carr, T.R., Pal, M. (2016). Comparison of supervised and unsupervised approaches for mudstone lithofacies classification: Case studies from the Bakken and Mahantango-Marcellus Shale, USA. Journal of Natural Gas Science and Engineering 33, 1119-1133. doi: 10.1016/j.jngse.2016.04.055.

Bickel, V. T., Conway, S. J., Tesson, P. A., Manconi, A., Loew, S., Mall, U. (2020). Deep learning-driven detection and mapping of rockfalls on Mars. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 13, 2831-2841. doi: 10.1109/JSTARS.2020.2991588.

Bizhani, M., Ardakani, O. H., Little, E. (2022). Reconstructing high fidelity digital rock images using deep convolutional neural networks. Scientific Reports 12, 4264. doi: 10.1038/s41598-022-08170-8.

Bochkovskiy, A., Wang, C. Y., Liao, H. Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934.

Bonté, D., Bouziat, A., Ferraille, M., Lechevallier, A., Koroko, A., Rohais, S. (2022). Deep Learning Technologies to Assist the Processing of Geological Data: a Support to Optimize E&P Workflows? In: EAGE Conference on Digital Innovation for a Sustainable Future 2022(1) 1-5, European Association of Geoscientists & Engineers. doi: 10.3997/2214-4609.202272019.

Borji, A. (2019). Pros and cons of GAN evaluation measures. Computer Vision and Image Understanding, 179, 41-65. doi: 10.1016/j.cviu.2018.10.009.

Bormann, P., Aursand, P., Dilib, F., Manral, S., Dischington, P. (2020). FORCE 2020 Well log and lithofacies dataset for machine learning competition. Dataset on Zenodo. doi: 10.5281/ZENODO.4351156.

Boutaba, R., Salahuddin, M.A., Limam, N., Ayoubi, S., Shahriar, N., Estrada-Solano, F., Caicedo, O.M. (2018). A comprehensive survey on machine learning for networking: evolution, applications and research opportunities. Journal of Internet Services and Applications 9(16), 1-99. doi: 10.1186/s13174-018-0087-2.

Bouziat, A., Lechevallier, A., Koroko, A., Lecomte, J., Feraille, M., Rohais, S., Desroziers, S. (2022). Assisted interpretation of thin sections and core samples with Deep Learning workflows. In: Second EAGE Digitalization Conference and Exhibition 2022(1) 1-5, European Association of Geoscientists & Engineers. doi: 10.3997/2214-4609.202239019

Bradski, G. (2000). The OpenCV library. Dr. Dobb's Journal: Software Tools for the Professional Programmer 120, 122-125.

Brampkamp, R.A., Powers, R.W. (1958). Classification of Arabian Carbonate Rocks. Bulletin of The Geological Society of America 69(10), 1305-1318. doi: 10.1130/0016-7606(1958)69[1305:COACR]2.0.CO;2.

Bressan, T.S., Kehl de Souza, M., Girelli, T.J., Chemale Junior, F. (2020). Evaluation of machine learning methods for lithology classification using geophysical data. Computers & Geosciences 139, 104475. doi: 10.1016/j.cageo.2020.104475.

Brown, W.M., Gedeon, T.D., Groves, D.I., Barnes, R.G. (2000). Artificial neural networks: A new method for mineral prospectivity mapping. Australian Journal of Earth Sciences 47(4), 757-770. doi: 10.1046/j.1440-0952.2000.00807.

Burger, W., Burge, M. J. (2009). Geometric operations. In: Principles of Digital Image Processing: Core Algorithms, London, 191-254. doi: 10.1007/978-1-84800-195-4_10.

Buslaev A, Iglovikov VI, Khvedchenya E, Parinov A, Druzhinin M, Kalinin AA. (2020). Albumentations: Fast and Flexible Image Augmentations. Information 11(2), 125. doi: 10.3390/info11020125.

Carannante, G., Esteban, M., Milliman, J.D., Simone, L. (1988). Carbonate lithofacies as paleolatitude indicators: problems and limitations. Sedimentary Geology 60(1-4), 333-346. doi: 10.1016/0037-0738(88)90128-5.

Carranza-García, M., Torres-Mateo, J., Lara-Benítez, P., García-Gutiérrez, J. (2021). On the Performance of One-Stage and Two-Stage Object Detectors in Autonomous Vehicles Using Camera Data. Remote Sensing 13(1), 89. doi: 10.3390/rs13010089.

Carvalho, D.V., Pereira, E.M., Cardoso, J.S. (2019). Machine Learning Interpretability: A Survey on Methods and Metrics. Electronics 8(8), 832-866. doi: 10.3390/electronics8080832.

Carvalho, L. E., Fauth, G., Fauth, S. B., Krahl, G., Moreira, A. C., Fernandes, C. P., Von Wangenheim, A. (2020). Automated microfossil identification and segmentation using a deep learning approach. Marine Micropaleontology 158, 101890. doi: 10.1016/j.marmicro.2020.101890.

Chai, X., Nie, W., Lin, K., Tang, G., Yang, T., Yu, J., Cao, W. (2022). An open-source package for deep-learning-based seismic facies classification: benchmarking experiments on the SEG 2020 open data. IEEE Transactions on Geoscience and Remote Sensing 60, 1-19. doi: 10.1109/TGRS.2022.3144666.

Chai, P., Hou, L., Zhang, G., Tushar, Q., Zou, Y. (2024). Generative adversarial networks in construction applications. Automation in Construction 159, 105265. doi: 10.1016/j.autcon.2024.105265.

Chan, W. H., Fung, B. S., Tsang, D. H., Lo, I. M. (2023). A freshwater algae classification system based on machine learning with StyleGAN2-ADA augmentation for limited and imbalanced datasets. Water Research 243, 120409. doi: 10.1016/j.watres.2023.120409.

Chawshin, K., Berg, C.F., Varagnolo, D., Lopez, O. (2021). Lithology classification of whole core CT scans using convolutional neural networks. SN Applied Sciences 3, 668. doi: 10.1007/s42452-021-04656-8.

Chen, J., Zeng, Y. (2018). Application of Machine Learning in Rock Facies Classification with Physics-Motivated Feature Augmentation. arXiv preprint arXiv:1808.09856.

Chen, L., Wang, L., Miao, J., Gao, H., Zhang, Y., Yao, Y., He, J. (2020). Review of the application of big data and artificial intelligence in geology. In Journal of Physics: Conference Series 1684(1), 012007. IOP Publishing.

Chen, Y. Y., Lin, Y. H., Hu, Y. C., Hsia, C. H., Lian, Y. A., Jhong, S. Y. (2022). Distributed real-time object detection based on edge-cloud collaboration for smart video surveillance applications. IEEE Access 10, 93745-93759. doi: 10.1109/ACCESS.2022.3203053.

Chen, W., Li, Y., Tian, Z., Zhang, F. (2023). 2D and 3D object detection algorithms from images: A Survey. Array 19, 100305. doi: 10.1016/j.array.2023.100305.

Cheng, L., Li, J., Duan, P., Wang, M. (2021). A small attentional YOLO model for landslide detection from satellite remote sensing images. Landslides 18(8), 2751-2765. doi: 10.1007/s10346-021-01694-6.

Chevitarese, D.S., Szwarcman, D., e Silva, R.G., & Brazil, E.V. (2018). Deep learning applied to seismic facies classification: A methodology for training. In: Saint Petersburg 2018 European Association of Geoscientists & Engineers, Conference Proceedings – EAGE Innovations in Geosciences, 1-5. doi: 10.3997/2214-4609.201800237.

Chinnam, S.K.R., Sistla, V., Kolli, V.K.K. (2022). Multimodal attention-gated cascaded U-Net model for automatic brain tumor detection and segmentation. Biomedical Signal Processing and Control 78, 103907. doi: 10.1016/j.bspc.2022.103907.

Chlap, P., Min, H., Vandenberg, N., Dowling, J., Holloway, L., Haworth, A. (2021). A review of medical image data augmentation techniques for deep learning applications. Journal of Medical Imaging and Radiation Oncology 65(5), 545-563. doi: 10.1111/1754-9485.13261.

Chollet, F. (2015). Keras. https://github.com/fchollet/keras.

Ciarapica, G., Passeri, L. (1993). An overview of the Maldivian coral reefs in Felidu and North Malé atoll (Indian Ocean): platform drowning by ecological crises. Facies 28(1), 33-65. doi: 10.1007/BF02539727.

Clark, A., (2015). Pillow (PIL Fork) Documentation, readthedocs. Available at: https://buildmedia.readthedocs.org/media/pdf/pillow/latest/pillow.pdf.

Conesa, G.R.A., Favre, E., Münch, P., Dalmasso, H., Chaix, C. (2005). Biosedimentary and paleoenvironmental evolution of the Southern Marion Platform from the middle to late Miocene (northeast Australia, ODP Leg 194, Sites 1196 and 1199). In: Anselmetti, F.S., Isern, A.R., Blum, P., and Betzler, C. (Eds.), Proceedings of the Ocean Drilling Program: Scientific Results 194: College Station, TX (Ocean Drilling Program), 1-38. doi: 10.2973/odp.proc.sr.194.005.2005.

Consonni, V., Ballabio, D., Todeschini, R. (2010). Evaluation of model predictive ability by external validation techniques. Journal of Chemometrics 24, 194-201. doi: 10.1002/cem.1290.

Cook, H. E., Hine, A. C., Mullins, H. T. (1983). Platform margin and deep water carbonates (Vol. 12). Society of Economic Paleontologists and Mineralogists.

Cortes, C., Vapnik, V. (1995). Support vector machine. Machine Learning 20(3), 273-297.

Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., Bharath, A. A. (2018). Generative adversarial networks: An overview. IEEE Signal Processing Magazine 35(1), 53-65. doi: 10.1109/MSP.2017.2765202.

Dai, J., Li, Y., He, K., Sun, J. (2016). R-FCN: object detection via region-based fully convolutional networks. Advances in Neural Information Processing Systems 29, 379-387.

Dalal, N., Triggs, B (2005). Histograms of oriented gradients for human detection. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 2005, 886-893. doi: 10.1109/CVPR.2005.177.

Davenport, T.H., Dyché, J. (2013). Big data in big companies. International Institute for Analytics, 3(1-31).

Davies, P.J., Symonds, P.A., Feary, D.A., and Pigram, C.J. (1989). The evolution of the carbonate platforms of northeast Australia. In: Crevello, P.D., Wilson, J.L., Sarg, J.F., Read, J.F. (Eds.), Controls on Carbonate Platform and Basin Development. Special Publications of SEPM 44, 233-258.

Davies, P.J., McKenzie, J.A., Palmer-Julson, A., *et al.* (1991). Principal results and summary. In: Davies, P.J., McKenzie, J.A., Palmer-Julson, A., *et al.*, Proceedings of the Ocean Drilling Program, Initial Reports 133, 73-134.

Dawson, H.L., Dubrule, O., John, C.M. (2023). Impact of dataset size and convolutional neural network architecture on transfer learning for carbonate rock classification. Computers & Geosciences 171, 105284. doi: 10.1016/j.cageo.2022.105284.

Dawson, H.L, John, C.M. (2024) Object Detection Algorithms to Identify Skeletal Components in Carbonate Cores. Marine and Petroleum Geology, 106965. doi: 10.1016/j.marpetgeo.2024.106965.

De Groot, P., Ligtenberg, H., Oldenziel, T., Connolly, D., Meldahl, P. (2004). Examples of multi-attribute, neural network-based seismic object detection. Geological Society, London, Memoirs 29(1), 335-338. doi: 10.1144/GSL.MEM.2004.029.01.31.

Dell'Aversana, P. (2019). Comparison of different Machine Learning algorithms for lithofacies classification from well logs. Bollettino di Geofisica Teorica ed Applicata 60(1), 69-80. doi: 10.4430/bgta0256

de Matos, M.C., Osorio, P.L.. Johann. P.R. (2007). Unsupervised seismic facies analysis using wavelet transform and self-organizing maps. Geophysics 72(1), 9-21. doi: 10.1190/1.2392789.

Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L. (2009). ImageNet: a large-scale hierarchical image database. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 248-255. doi: 10.1109/CVPR.2009.5206848.

Deng, L., Hinton, G., Kingsbury, B. (2013). New types of deep neural network learning for speech recognition and related applications: an overview. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, 8599-8603. doi: 10.1109/ICASSP.2013.6639344.

Deng, T., Xu, C., Lang, X., Doveton, J. (2021). Diagenetic facies classification in the Arbuckle formation using deep neural networks. Mathematical Geosciences 53(7), 1491-1512. doi: 10.1007/s11004-021-09918-0.

DeVries, P.M., Viégas, F., Wattenberg, M., Meade, B.J. (2018). Deep learning of aftershock patterns following large earthquakes. Nature 560(7720), 632-634. doi: 10.1038/s41586-018-0438-y.

Diwan, T., Anirudh, G., Tembhurne, J. V. (2023). Object detection using YOLO: Challenges, architectural successors, datasets and applications. Multimedia Tools and Applications 82(6), 9243-9275. doi: 10.1007/s11042-022-13644-y.

DMT GeoTec, (1996). DMT CoreScan Color User's Manual: Acquisition and Evaluation Software: Essen, Germany (Deutsche Montan Technologie GmbH).

DMT GmbH, (2000). DMT CoreScan Color Acquisition Software Digicore and Hardware Information: Essen, Germany (Deutsche Montan Technologie GmbH).

Dong, S-Q., Sun, Y-M., Xu, T., Zeng, L-B., Du, X-Y., Yang, X., Liang, Y., (2023). How to improve machine learning models for lithofacies identification by practical and novel ensemble strategy and principles. Petroleum Science 20(2), 733-752. doi: 10.1016/j.petsci.2022.09.006.

dos Anjos, C.E., Avila, M.R., Vasconcelos, A.G., Neta, A.M.P., Medeiros, L.C., Evsukoff, A.G., Surmas, R., Landau, L. (2021). Deep learning for lithological classification of carbonate rock micro-CT images. Computational Geosciences 25, 971-983. doi: 10.1007/s10596-021-10033-6.

Dramsch, J. S. (2020). 70 years of machine learning in geoscience in review. Advances in Geophysics, 61, 1-55. doi: 10.1016/bs.agph.2020.08.002.

Droxler, A.W., Haddad, G.A., Mucciarone, D.A., Cullen, J.L. (1990). Pliocene–Pleistocene aragonite cyclic variations in Holes 714A and 716B (The Maldives) compared with Hole 633A (The Bahamas): records of climate-induced CaCO3 preservation at intermediate water depths. In: Duncan, R.A., Backman, J., Peterson, L.C., et al., Proceedings of the Ocean Drilling Program, Scientific Results, 115: College Station, TX (Ocean Drilling Program), 539-577. doi: 10.2973/odp.proc.sr.115.179.1990.

Dubois, M.K., Bohling, G.C., Chakrabarti, S. (2007). Comparison of four approaches to a rock facies classification problem. Computers & Geosciences 33(5), 599-617.

DuMont Schütte, A., Hetzel, J., Gatidis, S., Hepp, T., Dietz, B., Bauer, S., Schwab, P. (2021). Overcoming barriers to data sharing with medical image generation: a comprehensive evaluation. NPJ digital medicine 4(1), 141. doi: 10.1038/s41746-021-00507-3.

Dunham, R.J. (1962). Classification of carbonate rocks according to depositional texture. In: Classification of Carbonate Rocks (Ed. Ham, W.E.), American Association of Petroleum Geologists Memoirs 1, 108-121.

Ebert-Uphoff, I., Deng, Y. (2017). Causal discovery in the geosciences—Using synthetic data to learn how to interpret results. Computers & Geosciences 99, 50-60. doi: 10.1016/j.cageo.2016.10.008.

Egmont-Petersen, M., de Ridder, D., Handels, H. (2002). Image processing with neural networks—a review. Pattern Recognition 35(10), 2279-2301. doi: 10.1016/S0031-3203(01)00178-9.

Ehrenberg, S. N., Dickson, J. A. D. (2003). Miocene carbonate platforms of the Marion Plateau, offshore NE Australia: contrasting styles of cathodoluminescence, geochemistry, dolomitization, architecture, and biota related to differing degrees of platform isolation. In: 12th Bathurst Meeting, International Conference of Carbonate Sedimentology, Durham, 8-10.

Ehrenberg, S.N., McAthur, J.M., Thirlwall, M.F. (2006). Growth, Demise, and Dolomitization of Miocene Carbonate Platforms on the Marion Plateau, Offshore NE Australia. Journal of Sedimentary Research 76(1), 91-116. doi: 10.2110/jsr.2006.06.

Embry, A.F., Klovan, J.E. (1971). A Late Devonian reef tract on Northeastern Banks Island, NWT. Bulletin of Canadian Petroleum Geology 19, 730-781.

Erickson, B.J., Korfiatis, P., Kline, T.L., Akkus, Z., Philbrick, K., Weston, A.D. (2018). Deep Learning in Radiology: Does One Size Fit All? Journal of the American College of Radiology 15(3), 521-526. doi: 10.1016/j.jacr.2017.12.027.

Everingham, M., Van Gool, L., Williams, C. K., Winn, J., Zisserman, A. (2010). The PASCAL Visual Object Classes (VOC) Challenge. International Journal of Computer Vision 88, 303-338.

Fabijanska, A., Feder, A., Ridge, J. (2020). DeepVarveNet: Automatic detection of glacial varves with deep neural networks. Computers & Geosciences 144, 104584. doi: 10.1016/j.cageo.2020.104584.

Falivene, O., Auchter, N.C., Pires de Lima, R., Kleipool, L., Solum, J.G., Zarian, P., Clark, R.W., Espejo, I. (2022). Lithofacies identification in cores using deep learning segmentation and the role of geoscientists: Turbidite deposits (Gulf of Mexico and North Sea). AAPG Bulletin 106(7), 1357-1372. doi: 10.1306/03112221015.

Fasel, B., Luettin, J. (2003). Automatic facial expression analysis: a survey. Pattern recognition 36(1), 259-275. doi: 10.1016/S0031-3203(02)00052-3.

Fayyad, U., Piatetsky-Shapiro, G., Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. AI Magazine 17(3), 37-54. doi: 10.1609/aimag.v17i3.1230.

Felzenszwalb, P., McAllester, D., Ramanan, D. (2008). A discriminatively trained, multiscale, deformable part model. In: 2008 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Anchorage, AK, USA, 1-8. doi: 10.1109/CVPR.2008.4587597.

Feng, Q., Guo, C., Benitez-Quiroz, F., Martinez, A. M. (2021). When do GANs replicate? On the choice of dataset size. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) 2021, 6701-6710.

Flügel, E. (2004). Microfacies of Carbonate Rocks: Analysis, Interpretation and Application. Springer Berlin Heidelberg. doi: 10.1007/978-3-662-08726-8.

Folk, R.L. (1959). Practical petrographic classification of limestones. AAPG Bulletin 43, 1-38.

Folk, R.L. (1962). Spectral subdivision of limestone types. In: Classification of Carbonate Rocks (Ed. Ham, W.E.), American Association of Petroleum Geologists Memoirs 1, 62-84.

Fossen, H. (2024). Structural core observations in a siliciclastic reservoir-scale framework. Geological Society, London, Special Publications, 527(1), SP527-2022. doi: 10.1144/SP527-2022-27.

Freund, Y., Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences 55(1), 119-139. doi: 10.1006/jcss.1997.1504.

Frid-Adar, M., Diamant, I., Klang, E., Amitai, M., Goldberger, J., Greenspan, H. (2018). GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. Neurocomputing 321, 321-331. doi: 10.1016/j.neucom.2018.09.013.

Fu, D., Su, C., Wang, W., Yuan, R. (2022). Deep learning based lithology classification of drill core images. PLoS One 17(7), e0270826. doi: 10.1371/journal.pone.0270826.

Fürstenau, J., Lindhorst, S., Betzler, C., Hübscher, C. (2010). Submerged reef terraces of the Maldives (Indian Ocean). Geo-Marine Letters 30(5), 511-515. doi: 10.1007/s00367-009-0174-2.

Gallagher, S. J., Fulthorpe, C. S., Bogus, K., Auer, G., Baranwal, S., Castañeda, I. S., Christensen, B. A., De Vleeschouwer, D., Franco, D. R., Groeneveld, J., Gurnis, M., Haller, C., He, Y., Henderiks, J., Himmler, T., Ishiwa, T., Iwatani, H., Jatiningrum, R. S., Kominz, M. A., Korpanty, C. A., Lee, E. Y., Levin, E., Mamo, B. L., McGregor, H. V., McHugh, C. M., Petrick, B. F., Potts, D. C., Rastegar Lari, A., Renema, W., Reuning, L., Takayanagi, H., Zhang, W. (2017). Site U1459. In: Indonesian Throughflow, edited by: Gallagher, S. J., Fulthorpe, C. S., Bogus, K., and the Expedition 356 Scientists, Proceedings of the International Ocean Discovery Program 356, College Station, TX, doi: 10.14379/iodp.proc.356.104.2017.

Gamal, H., Elkatatny, S., Adebayo, A., Bageri, B. (2020). Effect of exposure time on the compressive strength and formation damage of sandstone while drilling horizontal wells. Journal of Petroleum Science and Engineering 195, 107590. doi: 10.1016/j.petrol.2020.107590.

Gegg, L., Gegg, J. (2023). Poor Man's Line Scan–a simple tool for the acquisition of high-resolution, undistorted drill core photos. Scientific Drilling, 32, 55-59. doi: 10.5194/sd-32-55-2023.

Geotek Limited (2017). Geotek Core Imaging System (CIS): High resolution linescan photographs for immediate and archival core record applications. Available: http://www.geotek.co.uk/wp-content/uploads/2017/06/MSCL-CIS.pdf.

Germay, C, Lhomme, T., Perneder, L. (2023). High-resolution core data and machine learning schemes applied to rock facies classification. In: Neal, A., Ashton, M., Williams, L.S., Dee, S.J., Dodd, T.J.H., Marshall, J.D. (eds) Core Values: the Role of Core in Twenty-first Century Reservoir Characterization. Geological Society, London, Special Publications (527), doi: 10.1144/SP527-2021-193.

Ghiasi-Freez, J., Honarmand-Fard, S., Ziaii, M. (2014). The automated Dunham classification of carbonate rocks through image processing and an intelligent model. Petroleum Science and Technology 32(1), 100-107. doi: 10.1080/10916466.2011.576369.

Gifford, C.M., Agah, A. (2010). Collaborative multi-agent rock facies classification from wireline well log data. Engineering Applications of Artificial Intelligence 23(7), 1158-1172. doi: 10.1016/j.engappai.2010.02.004.

Girshick, R., Donahue, J., Darrell, T., Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 580-587.

Girshick, R. (2015). Fast R-CNN. 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 1440-1448. doi: 10.1109/ICCV.2015.169.

Gischler, E. (2006). Sedimentation on Rasdhoo and Ari atolls, Maldives, Indian Ocean. Facies 52(3), 341-360. doi: 10.1007/s10347-005-0031-3.

Gischler, E., Hudson, J.H., Pisera, A. (2008). Late Quaternary reef growth and sea level in the Maldives (Indian Ocean). Marine Geology 250(1–2), 104-113. doi: 10.1016/j.margeo.2008.01.004.

Glenn, C.R., Kronen Jr, J.D., Symonds, P.A., Wei, W., Kroon, D. (1993). High-resolution sequence stratigraphy, condensed sections, and flooding events off the Great Barrier Reef: 0-1.5 Ma. In: Proceedings of ODP 133, Scientific Results, 353-364. Ocean Drilling Program College Station, TX.

Goldstein, E. B., Coco, G., Plant, N. G. (2019). A review of machine learning applications to coastal sediment transport and morphodynamics. Earth-Science Reviews 194, 97-108. doi: 10.1016/j.earscirev.2019.04.022.

Golhani, K., Balasundram, S.K., Vadamalai, G., Pradhan, B. (2018). A review of neural networks in plant disease detection using hyperspectral data. Information Processing in Agriculture 5(3), 345-371. doi: 10.1016/j.inpa.2018.05.002.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., (2014). Generative Adversarial Nets. Advances in Neural Information Processing Systems 27.

Goodfellow, I., Bengio, Y., Courville, A. (2016). Deep learning. MIT Press.

Gupta, A., Anpalagan, A., Guan, L., Khwaja, A. S. (2021). Deep learning for object detection and scene perception in self-driving cars: Survey, challenges, and open issues. Array 10, 100057. doi: 10.1016/j.array.2021.100057.

Haklay M., Mazumdar S., Wardlaw J. (2018). Citizen Science for Observing and Understanding the Earth. In: Mathieu PP., Aubrecht C. (eds) Earth Observation Open Science and Innovation. ISSI Scientific Report Series 15, 69-88. doi: 10.1007/978-3-319-65633-5_4.

Hall, R. (2002). Cenozoic geological and plate tectonic evolution of SE Asia and the SW Pacific: computer-based reconstructions, model and animations. Journal of Asian Earth Sciences 20(4), 353-431. doi: 10.1016/S1367-9120(01)00069-4.

Hall, B. (2016). Facies classification using machine learning. The Leading Edge 35(10), 906-909. doi: 10.1190/tle35100906.1.

Hall, M., Hall, B. (2017). Distributed collaborative prediction: Results of the machine learning contest. The Leading Edge 36(3), 267-269. doi: 10.1190/tle36030267.1.

Hallsworth, C. R., Knox, R. (1999). BGS Rock Classification Scheme Volume 3. Classification of Sediments and Sedimentary Rocks. British Geological Survey Research Report RR/99/03, Keyworth, UK, 1-44.

Halotel, J., Demyanov, V., Gardiner, A. (2020). Value of geologically derived features in machine learning facies classification. Mathematical Geosciences 52, 5-29. doi: 10.1007/s11004-019-09838-0.

Harris, C.R., Millman, K.J., Van Der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., Kern, R. (2020). Array programming with NumPy. Nature 585(7825), 357-362. doi: 10.1038/s41586-020-2649-2.

He, J., La Croix, A.D., Wang, J., Ding, W., Underschultz, J.R. (2019). Using neural networks and the Markov Chain approach for facies analysis and prediction from well logs in the Precipice Sandstone and Evergreen Formation, Surat Basin, Australia. Marine and Petroleum Geology 101, 410-427. doi: 10.1016/j.marpetgeo.2018.12.022.

He, K., Zhang, X., Ren, S., Sun, J. (2014). Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *In*: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds) Computer Vision – ECCV 2014. Lecture Notes in Computer Science 8691. Springer, Cham. doi: 10.1007/978-3-319-10578-9_23.

He, K., Gkioxari, G., Dollár, P., Girshick, R. (2017). Mask R-CNN. Proceedings of the IEEE International Conference on Computer Vision (ICCV) 2017, 2961-2969.

He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition, 770-778.

He, M., Gu, H., Wan, H. (2020). Log interpretation for lithology and fluid identification using deep neural network combined with MAHAKIL in a tight sandstone reservoir. Journal of Petroleum Science and Engineering 194, 107498. doi: 10.1016/j.petrol.2020.107498.

He, Y., Zhou, Y., Wen, T., Zhang, S., Huang, F., Zou, X., Ma, X., Zhu, Y. (2022). A review of machine learning in geochemistry and cosmochemistry: Method improvements and applications. Applied Geochemistry 140, 105273. doi: 10.1016/j.apgeochem.2022.105273.

Heck, P.R., Frank, M., Anselmetti, F.S., Kubik, P.W. (2007). Origin and age of submarine ferromanganese hardgrounds from the Marion Plateau, offshore northeast Australia. In Anselmetti, F.S., Isern, A.R., Blum, P., and Betzler, C. (Eds.), 2006, Proc. ODP, Sci. Results, 194: College Station, TX (Ocean Drilling Program), 1-22. doi: 10.2973/odp.proc.sr.194.008.2007.

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S. (2017). GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. Advances in Neural Information Processing Systems, 30 (NIPS 2017).

Ho, Y., Wookey, S. (2019). The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling. IEEE Access, 8, 4806-4813. doi: 10.1109/ACCESS.2019.2962617.

Hong, Z., Yan, G. L., Cui, X. L., Gao, G., Hu, G. M. (2021). A novel approach to the automatic classification of wireline log-predicted sedimentary microfacies based on object detection. Applied Geophysics, 1-16. doi: 10.1007/s11770-021-0962-7.

Hossain, S., Umer, S., Rout, R. K., Tanveer, M. (2023). Fine-grained image analysis for facial expression recognition using deep convolutional neural networks with bilinear pooling. Applied Soft Computing 134, 109997. doi: 10.1016/j.asoc.2023.109997.

Hossain, T.M., Watada, J. Aziz, I.A., Hermana, M. (2020). Machine Learning in Electrofacies Classification and Subsurface Lithology Interpretation: A Rough Set Theory Approach. Applied Sciences 10(17), 5940. doi: 10.3390/app10175940.

Hosseini, M., Powell, M., Collins, J., Callahan-Flintoft, C., Jones, W., Bowman, H., Wyble, B. (2020). I tried a bunch of things: The dangers of unexpected overfitting in classification of brain data. Neuroscience & Biobehavioral Reviews 119, 456-467. doi: 10.1016/j.neubiorev.2020.09.036.

Hou, M., Xiao, Y., Lei, Z., Yang, Z, Lou, Y., Liu, Y. (2023). Machine Learning Algorithms for Lithofacies Classification of the Gulong Shale from the Songliao Basin, China. Energies 16(6), 2581. doi: 10.3390/en16062581.

Howard, J., Gugger, S. (2020). Fastai: a layered API for deep learning. Information 11(2), 108. doi: 10.3390/info11020108

Hu, G. X., Yang, Z., Hu, L., Huang, L., Han, J. M. (2018). Small object detection with multiscale features. International Journal of Digital Multimedia Broadcasting, 2018. doi: 10.1155/2018/4546896.

Huang, G., Liu, Z., Weinberger, K.Q. (2017a). Densely Connected Convolutional Networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2261-2269.

Huang, L., Dong, X., Clee, T. E. (2017b) A scalable deep learning platform for identifying geologic features from seismic attributes. The Leading Edge, 36(3), 249-256. doi: 10.1190/tle36030249.1.

Hull, D., Chapman, P., Miller, D., Ingraham, D., Fritz, N., Kernan, N. (2015). Regional Eagle Ford Modeling: Integrating Facies, Rock Properties, and Stratigraphy to Understand Geologic and Reservoir Characteristics. Unconventional Resources Technology Conference, San Antonio, Texas. doi: 10.15530/urtec-2015-2173648.

Hunter, J.D. (2007). Matplotlib: A 2D graphics environment. Computing in Science and Engineering 9(3), 90-95.

Huntingford, C., Jeffers, E. S., Bonsall, M.B., Christensen, H.M., Lees, T., & Yang, H. (2019). Machine learning and artificial intelligence to aid climate change research and preparedness. Environmental Research Letters, 14(12), 124007. doi: 10.1088/1748-9326/ab4e55.

Hussain, Z., Gimenez, F., Yi, D., Rubin, D. (2018). Differential data augmentation techniques for medical imaging classification tasks. In American Medical Informatics Association (AMIA) Annual Symposium Proceedings 2017. 979-984.

Imamverdiyev, Y., Sukhostat, L. (2019). Lithological facies classification using deep convolutional neural network. Journal of Petroleum Science and Engineering 174, 216-228.

Inbar, O., Shahar, M., Gidron, J., Cohen, I., Menashe, O., Avisar, D., (2023). Analyzing the secondary wastewater-treatment process using Faster R-CNN and YOLOv5 object detection algorithms. Journal of Cleaner Production 416, 137913. doi: 10.1016/j.jclepro.2023.137913.

Insua, T.L., Hamel, L., Moran, K., Anderson, L.M., Webster, J.M. (2014). Advanced classification of carbonate sediments based on physical properties. Sedimentology 62(2), 590-606. doi: 10.1111/sed.12168.

Isern, A.R., Anselmetti, F.S., Blum, P., and the Leg 194 Scientific Party (2001). ODP Leg 194: sea level magnitudes recorded by continental margin sequences on the Marion Plateau, northeast Australia. JOIDES Journal 27(2), 7-11.

Isern, A.R., Anselmetti, F.S., Blum, P., et al. (2002). Leg 194 Summary. Proceedings of the Ocean Drilling Program, Initial Reports 194, 1-88.

Isern, A.R., Anselmetti, F.S., Blum, P. (2004). A Neogene carbonate platform, slope, and shelf edifice shaped by sea level and ocean currents, Marion Plateau (northeast Australia). In: Eberli, G.P., Massaferro, J.L., and Sarg, J.F. (Eds.), Seismic imaging of carbonate reservoirs and systems. AAPG Memoir 81, 291-307.

Islam, T., Hafiz, M. S., Jim, J. R., Kabir, M. M., Mridha, M. F. (2024). A systematic review of deep learning data augmentation in medical imaging: Recent advances and future research directions. Healthcare Analytics 5, 100340. doi: 10.1016/j.health.2024.100340.

Ivchenko, A.V., Baraboshkin, E.E., Ismailova, L.S., Orlov, D.M., Koroteev, D.A., Baraboshkin, E.Y. (2018).Core photo lithological interpretation based on computer analyses. IEEE Northwest Russia Conference on Mathematical Methods in Engineering and Technology, 425-428.

Izbassarova A., Duisembay A. James A.P. (2020). Speech Recognition Application Using Deep Learning Neural Network. In: James A. (eds) Deep Learning Classifiers with Memristive Networks: Modeling and Optimization in Science and Technologies 14, 69-79. doi: 10.1007/978-3-030-14524-8_5.

Jain, S., Seth, G., Paruthi, A., Soni, U., Kumar, G. (2022). Synthetic data augmentation for surface defect detection and classification using deep learning. Journal of Intelligent Manufacturing 33, 1007-1020. doi: 10.1007/s10845-020-01710-x.

James, N. P., Clarke, J. A. (1997). Cool-water carbonates. SEPM Society for Sedimentary Geology 56. doi: 10.2110/pec.97.56.

Jiang, J., Xu, R., James, S.C., Xu, C. (2021). Deep-Learning-Based Vuggy Facies Identification from Borehole Images. SPE Reservoir Evaluation & Engineering 24(01), 250-261.

Jiang, P., Ergu, D., Liu, F., Cai, Y., Ma, B. (2022). A Review of YOLO Algorithm Developments. Procedia Computer Science 199, 1066-1073. doi: 10.1016/j.procs.2022.01.135.

Jobe, T.D., Vital-Brazil, E., Khait, M. (2018). Geological Feature Prediction Using Image-Based Machine Learning. Petrophysics 59(6), 750-760.

Jocher, G. (2020). Ultralytics YOLOv5 (7.0) [Computer software]. doi: 10.5281/zenodo.3908559. https://github.com/ultralytics/yolov5

Jocher, G., Chaurasia, A., Qiu, J. (2023). Ultralytics YOLO (Version 8.0.0) [Computer software]. https://github.com/ultralytics/ultralytics.

John, C.M., Kanagandran, S. (2019). AI to improve the reliability and reproducibility of descriptive data: a case study using convolutional neural networks to recognize carbonate facies in cores. AAPG Annual Convention and Exhibition, San Antonio, TX.

John, C.M., Karner, G.D., Mutti, M. (2004). δ18O and Marion Plateau backstripping: combining two approaches to constrain late middle Miocene eustatic amplitude. Geology 32, 829-832. doi: 10.1130/G20580.1.

Jutzeler, M., White, J. D., Talling, P. J., McCanta, M., Morgan, S., Le Friant, A., Ishizuka, O. (2014). Coring disturbances in IODP piston cores with implications for offshore record of volcanic events and the Missoula megafloods. Geochemistry, Geophysics, Geosystems, 15(9), 3572-3590. doi: 10.1002/2014GC005447.

Kaggle (2019). TGS Salt Identification Challenge. https://www.kaggle.com/c/tgs-salt-identification-challenge.

Kalhor Mohammadi, M., Riahi, S., Boek, E. S. (2023). An insight review on formation damage induced by drilling fluids. Reviews in Chemical Engineering 39(3), 387-415. doi: 10.1515/revce-2020-0106.

Karpatne, A., Kumar, V. (2017). Big data in climate: Opportunities and challenges for machine learning. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 21-22. doi: 10.1145/3097983.3105810

Karpatne, A., Ebert-Uphoff, I., Ravela, S., Babaie, H. A., & Kumar, V. (2018). Machine learning for the geosciences: Challenges and opportunities. IEEE Transactions on Knowledge and Data Engineering 31(8), 1544-1554. doi: 10.1109/TKDE.2018.2861006.

Karras, T., Aittala, M., Hellsten, J., Laine, S., Lehtinen, J., Aila, T. (2020). Training generative adversarial networks with limited data. Advances in Neural Information Processing Systems 33 (NeurIPS 2020), 12104-12114.

Karras, T., Aittala, M., Laine, S., Härkönen, E., Hellsten, J., Lehtinen, J., & Aila, T. (2021). Alias-free generative adversarial networks. Advances in Neural Information Processing Systems 34 (NeurIPS 2021), 852-863.

Kaur, R., Singh, S. (2023). A comprehensive review of object detection with deep learning. Digital Signal Processing 132, 103812. doi: 10.1016/j.dsp.2022.103812.

Kelling, S., Johnston, A., Bonn, A., Fink, D., Ruiz-Gutierrez, V., Bonney, R., Fernandez, M., Hochachka, W.M., Julliard, R., Kraemer, R., Guralnick, R. (2019). Using Semistructured Surveys to Improve Citizen Science Data for Monitoring Biodiversity. BioScience 69(3), 170-179. doi: 10.1093/biosci/biz010.

Khalifa, N. E., Loey, M., Mirjalili, S. (2022). A comprehensive survey of recent trends in deep learning for digital images augmentation. Artificial Intelligence Review 55(3), 2351-2377. doi: 10.1007/s10462-021-10066-4.

Khan, S. A., Lee, H. J., Lim, H. (2023). Enhancing Object Detection in Self-Driving Cars Using a Hybrid Approach. Electronics 12(13), 2768. doi: 10.3390/electronics12132768.

Kim, J. (2022). Lithofacies classification integrating conventional approaches and machine learning technique. Journal of Natural Gas Science and Engineering 100, 104500. doi: 10.1016/j.jngse.2022.104500.

Kim, J., Park, H. (2023). Limited Discriminator GAN using explainable AI model for overfitting problem. ICT Express 9(2), 241-246. doi: 10.1016/j.icte.2021.12.014.

Kingma, D.P., Ba, J. (2014). Adam: A Method for Stochastic Optimization. arXiv preprint arXiv:1412.6980.

Klump, J.F., Robertson, J. (2015).Synthetic geology - Exploring the "what if?" in geology. AGU Fall Meeting Abstracts, American Geophysical Union, Fall Meeting 2015, IN51A-1783.

Koeshidayatullah, A., Morsilli, M., Lehrmann, D.J., Al-Ramadan, K., Payne, J.L. (2020a). Fully automated carbonate petrography using deep convolutional neural networks. Marine and Petroleum Geology 122, 104687. doi: 10.1016/j.marpetgeo.2020.104687.

Koeshidayatullah, A., Corlett, H., Stacey, J., Swart, P.K., Boyce, A., Robertson, H., Whitaker, F. Hollis, C. (2020b). Evaluating new fault-controlled hydrothermal dolomitization models: Insights from the Cambrian Dolomite, Western Canadian Sedimentary Basin. Sedimentology 67(6), 2945-2973. doi: 10.1111/sed.12729.

Klyuchnikov, N., Ismailova, L., Kovalev, D., Safonov, S., Koroteev, D. (2022). Generative Adversarial Networks for synthetic wellbore data: Expert perception vs mathematical metrics. Journal of Petroleum Science and Engineering 211, 110106. doi: 10.1016/j.petrol.2022.110106.

Koh, P. W., Liang, P. (2017). Understanding black-box predictions via influence functions. In: International Conference on Machine Learning, PMLR 70, 1885-1894.

Kong, Q., Trugman, D.T., Ross, Z.E., Bianco, M.J., Meade, B.J., Gerstoft, P. (2019). Machine learning in seismology: Turning data into insights. Seismological Research Letters 90(1), 3-14. doi: 10.1785/0220180259.

Konidaris, F., Tagaris, T., Sdraka, M., Stafylopatis, A. (2019). Generative Adversarial Networks as an Advanced Data Augmentation Technique for MRI Data. Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2019), 48-59. doi: 10.5220/0007363900480059.

Koroteev, D., Tekic, Z. (2021). Artificial intelligence in oil and gas upstream: Trends, challenges, and scenarios for the future. Energy and AI 3, 100041. doi: 10.1016/j.egyai.2020.100041.

Korteling, J.E., Brouwer, A-M., Toet, A. (2018). A Neural Network Framework for Cognitive Bias. Frontiers in Psychology 9(1561). doi: 10.3389/fpsyg.2018.01561.

Kortström, J., Uski, M., Tiira, T. (2016). Automatic classification of seismic events within a regional seismograph network. Computers & Geosciences 87, 22-30. doi: 10.1016/j.cageo.2015.11.006.

Krizhevsky, A., Hinton, G. (2009). Learning multiple layers of features from tiny images.

Krizhevsky, A., Nair, V., Hinton, G. (2014). The CIFAR-10 dataset. Online: http://www. cs. toronto. edu/kriz/cifar.html, 55(5).

Krizhevsky, A., Sutskever, I., Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems 25 (NIPS 2012), 1-9.

Krumbein, W.C., Dacey, M.F. (1969). Markov chains and embedded Markov chains in geology. Journal of the International Association for Mathematical Geology 1(1), 79-96.

Kuhn, M., Johnson, K. (2013). Over-Fitting and Model Tuning. *In*: Applied Predictive Modeling. Springer, New York, NY. doi: 10.1007/978-1-4614-6849-3_4.

Lalitha, V., Latha, B. (2022). A review on remote sensing imagery augmentation using deep learning. Materials Today: Proceedings 62, 4772-4778. doi: 10.1016/j.matpr.2022.03.341.

Lan, X., Zou, C., Kang, Z., Wu, X. (2021). Log facies identification in carbonate reservoirs using multiclass semi-supervised learning strategy. Fuel 302, 121145. doi: 10.1016/j.fuel.2021.121145.

Larson, T.E., Loucks, R.G., Sivil, J.E., Hattori, K.E., Zahm, C.K. (2023). Machine learning classification of Austin Chalk chemofacies from high-resolution x-ray fluorescence core characterization. AAPG Bulletin 107(6), 907-927. doi: 10.1306/09232220095.

Lary, D. J., Alavi, A. H., Gandomi, A. H., Walker, A. L. (2016). Machine learning in geosciences and remote sensing. Geoscience Frontiers 7(1), 3-10. doi: 10.1016/j.gsf.2015.07.003.

Lawrence, N. D. (2012). A unifying probabilistic perspective for spectral dimensionality reduction: Insights and new models. Journal of Machine Learning Research 13, 1609-1638.

LeCun, Y., Bengio, Y., Hinton, G. (2015). Deep learning. Nature 521(7553), 436-444. doi: 10.1038/nature14539.

LeCun, Y., Bottou, L., Bengio, Y., Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE 86, 2278-2324. doi: 10.1109/5.726791.

Leighton, M.W., Pendexter, C. (1962). Carbonate rock types. In: Classification of Carbonate Rocks (Ed. Ham, W.E.), American Association of Petroleum Geologists Memoirs 1, 33-60.

Li, S., Zhou, K., Zhao, L., Xu, Q., Liu, J. (2022). An improved lithology identification approach based on representation enhancement by logging feature decomposition, selection and transformation. Journal of Petroleum Science and Engineering 209, 109842. doi: 10.1016/j.petrol.2021.109842.

Li, X., Li, W., Liu, B., Liu, Q., Yu, N. (2018). Object-Oriented Anomaly Detection in Surveillance Videos. In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 2018, 1907-1911. doi: 10.1109/ICASSP.2018.8461422.

Li, Y., Anderson-Sprecher, R. (2006). Facies identification from well logs: A comparison of discriminant analysis and naïve Bayes classifier. Journal of Petroleum Science and Engineering 53(3-4), 149-157. doi: 10.1016/j.petrol.2006.06.001.

Lienhart, R., Maydt, J. (2002). An extended set of Haar-like features for rapid object detection. In: Proceedings. International Conference on Image Processing, Rochester, NY, USA, 2002, I-I. doi: 10.1109/ICIP.2002.1038171.

Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C. L. (2014). Microsoft COCO: Common Objects in Context. In: Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13, 740-755. Springer International Publishing. https://cocodataset.org/#home.

Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S. (2017a). Feature pyramid networks for object detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (Honolulu, HI), 2117-2125. doi: 10.1109/CVPR.2017.106.

Lin, T. Y., Goyal, P., Girshick, R., He, K., Dollár, P. (2017b). Focal loss for dense object detection. Proceedings of the IEEE International Conference on Computer Vision (ICCV) 2017, 2980-2988.

Liu, J-J, Liu, J-C. (2022). Integrating deep learning and logging data analytics for lithofacies classification and 3D modeling of tight sandstone reservoirs. Geoscience Frontiers 13(1), 101311. doi: doi.org/10.1016/j.gsf.2021.101311.

Liu, M., Jervis, M., Li, W., Nivlet, P. (2020a). Seismic facies classification using supervised convolutional neural networks and semisupervised generative adversarial networks. Geophysics 85(4), O47-O58. doi: 10.1190/geo2019-0627.1.

Liu, S., Qi, L., Qin, H., Shi, J., Jia, J. (2018). Path aggregation network for instance segmentation. In: Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition, 8759–8768.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., Berg, A. C. (2016). SSD: Single shot multibox detector. Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14 2016, Proceedings, Part I 14, 21-37. doi: 10.1007/978-3-319-46448-0_2.

Liu, X., Jiang, S., Wu, R., Shu, W., Hou, J., Sun, Y., Sun, J., Chu, D., Wu, Y. and Song, H. (2023). Automatic taxonomic identification based on the Fossil Image Dataset (> 415,000 images) and deep convolutional neural networks. Paleobiology 49(1), pp.1-22. doi: 10.1017/pab.2022.14.

Liu, X., Song, H. (2020). Automatic identification of fossils and abiotic grains during carbonate microfacies analysis using deep convolutional neural networks. Sedimentary Geology 410, 105790. doi: 10.1016/j.sedgeo.2020.105790.

Liu, X., Wang, H., Jing, H., Shao, A., Wang, L. (2020b). Research on intelligent identification of rock types based on faster R-CNN method. IEEE Access 8, 21804-21812. doi: 10.1109/ACCESS.2020.2968515.

Liu, Z., Luo, P., Wang, X., Tang, X. (2015). Deep learning face attributes in the wild. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV'15), 3730-3738.

Lobo, P., Vilaça, J. L., Torres, H., Oliveira, B., Simões, A. (2023). Smart scan of medical device displays to integrate with a mHealth application. Heliyon 9(6), e16297. doi: 10.1016/j.heliyon.2023.e16297.

Lokier, S.W., Al Junaibi, M. (2016). The petrographic description of carbonate facies: are we all speaking the same language? Sedimentology 63, 1843-1885. doi: 10.1111/sed.12293.

Lowe, D.G. (2004). Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision 60, 91-110. doi: 10.1023/B:VISI.0000029664.99615.94.

Lu, X., Li, Q., Li, B., Yan, J. (2020). Mimicdet: Bridging the gap between one-stage and two-stage object detection. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, JM. (eds) Computer Vision – ECCV 2020. ECCV 2020. Lecture Notes in Computer Science 12359, 541-557. doi: 10.1007/978-3-030-58568-6_32.

Lucic, M., Kurach, K., Michalski, M., Gelly, S., Bousquet, O. (2018). Are gans created equal? a large-scale study. Advances in Neural Information Processing Systems 31, (NeurIPS 2018), 1-10.

Lüdmann, T., Kalvelage, C., Betzler, C., Fürstenau, J., Hübscher, C. (2013). The Maldives, a giant isolated carbonate platform dominated by bottom currents. Marine and Petroleum Geology 43, 326-340. doi: 10.1016/j.marpetgeo.2013.01.004.

Lundervold, A. S., Lundervold, A. (2019). An overview of deep learning in medical imaging focusing on MRI. Zeitschrift für Medizinische Physik 29(2), 102-127. doi: 10.1016/j.zemedi.2018.11.002.

Ma, C., Tan, G., Li, X., Xu, J., Chen, J. (2022). Core Orientation Technology Based on Drilling Trajectory Projection and Its Application in In Situ Stress Measurement of the Deepest Shaft in China. Minerals 12(5), 521. doi: 10.3390/min12050521.

Machajdik, J., Hanbury, A. (2010). Affective image classification using features inspired by psychology and art theory. In: Proceedings of the 18th ACM international conference on Multimedia, October 2010, 83-92.

MacLeod, C. J., Parson, L. M., Sager, W. W. (1994). Reorientation of cores using the Formation MicroScanner and Borehole Televiewer: application to structural and paleomagnetic studies with the Ocean Drilling Program. Proceedings of the Ocean Drilling Program, 135 Scientific Results.

Malone, M.J., Baker, P.A., Burns, S.J., Swart, P.K. (1990). Geochemistry of periplatform carbonate sediments, Leg 115, Site 716 (Maldives archipelago, Indian Ocean). In: Duncan, R.A., Backman, J., Peterson, L.C., et al., Proceedings of the Ocean Drilling Program, Scientific Results, 115: College Station, TX (Ocean Drilling Program), 647-659. doi: 10.2973/odp.proc.sr.115.184.1990.

Mandal, P.P., Rezaee, R. (2019). Facies classification with different machine learning algorithm – An efficient artificial intelligence technique for improved classification. ASEG Extended Abstracts 2019, 1-6. doi: 10.1080/22020586.2019.12072918.

Markoff, J. (2012). Seeking a Better Way to Find Web Images. New York Times, 20 November 2012. https://www.nytimes.com/2012/11/20/science/for-web-images-creating-new-technology-to-seek-and-find.html.

Martin, T., Meyer, R., Jobe, Z. (2021). Centimeter-Scale Lithology and Facies Prediction in Cored Wells Using Machine Learning. Frontiers in Earth Science 9(659611), 1-18. doi: 10.3389/feart.2021.659611.

Maxwell, A.E., Warner, T.A., Fang, F. (2018). Implementation of machine-learning classification in remote sensing: An applied review. International Journal of Remote Sensing 39(9), 2784-2817. doi: 10.1080/01431161.2018.1433343.

Meldahl, P., Heggland, R., Bril, B., de Groot, P. (2001). Identifying faults and gas chimneys using multiattributes and neural networks. The Leading Edge 20(5), 474-482. doi: 10.1190/1.1438976.

Memon, F.H., Tunio, A.H., Memon, K.R., Mahesar, A.A., Abbas, G. (2023). Unveiling the Diagenetic and Mineralogical Impact on the Carbonate Formation of the Indus Basin, Pakistan: Implications for Reservoir Characterization and Quality Assessment. Minerals 13(12), 1474. doi: 10.3390/min13121474.

Meng, Y., Kuppannagari, S., Kannan, R., Prasanna, V. (2021). How to avoid zero-spacing in fractionally-strided convolution? a hardware-algorithm co-design methodology. 2021 IEEE 28th International Conference on High Performance Computing, Data, and Analytics (HiPC), Bengaluru, India, 81-90. doi: 10.1109/HiPC53243.2021.00022.

Merembayev, T., Kumangaliyev, D., Bekbauov, B., Amanbek, Y. (2021). A Comparison of Machine Learning Algorithms in Predicting Lithofacies: Case Studies from Norway and Kazakhstan. Energies 14(7), 1896. doi: 10.3390/en14071896.

Middya, A. I., Chattopadhyay, D., Roy, S. (2021). Garbage detection and classification using faster-RCNN with inception-V2. 2021 IEEE 18th India Council International Conference (INDICON), Guwahati, India, 1-6. doi: 10.1109/INDICON52576.2021.9691547.

Miller, R.S., Rhodes, S., Khosla, D., Nino, F. (2019). Application of Artificial Intelligence for Depositional Facies Recognition—Permian Basin. Proceedings, Society of Exploration Geophysicists Unconventional Resources Technology Conference (URTeC), Denver, Colorado (2019), 4410-4415. doi: 10.15530/urtec-2019-193.

Montesinos López, O. A., Montesinos López, A., Crossa, J. (2022). Overfitting, model tuning, and evaluation of prediction performance. In: Multivariate statistical machine learning methods for genomic prediction, Cham: Springer International Publishing, 109-139. doi: 10.1007/978-3-030-89010-0_4.

Mosser, L., Dubrule, O., Blunt, M. J. (2017). Reconstruction of three-dimensional porous media using generative adversarial neural networks. Physical Review E 96(4), 043309. doi: 10.1103/PhysRevE.96.043309.

Mosser, L., Dubrule, O., Blunt, M. J. (2018). Stochastic reconstruction of an oolitic limestone by generative adversarial networks. Transport in Porous Media 125(1), 81-103. doi: 10.1007/s11242-018-1039-9.

Mosser, L., Dubrule, O., Blunt, M. J. (2020). Stochastic seismic waveform inversion using generative adversarial networks as a geological prior. Mathematical Geosciences 52, 53-79. doi: 10.1007/s11004-019-09832-6.

Motamed, S., Rogalla, P., Khalvati, F. (2021). Data augmentation using Generative Adversarial Networks (GANs) for GAN-based detection of Pneumonia and COVID-19 in chest X-ray images. Informatics in Medicine Unlocked 27, 100779. doi: 10.1016/j.imu.2021.100779.

Mousavi, S.M., Zhu, W., Sheng, Y., Beroza, G.C. (2019). CRED: A Deep Residual Network of Convolutional and Recurrent Units for Earthquake Signal Detection. Scientific Reports 9 (10267). doi: 10.1038/s41598-019-45748-1.

Mubarak, Y., Koeshidayatullah, A. (2023). Hierarchical automated machine learning (AutoML) for advanced unconventional reservoir characterization. Scientific Reports 13, 13812. doi: 10.1038/s41598-023-40904-0.

Mumuni, A., Mumuni, F. (2022). Data augmentation: A comprehensive survey of modern approaches. Array 16, 100258. doi: 10.1016/j.array.2022.100258.

Murray, J.D., vanRyper, W. (1996). Encyclopedia of Graphics File Formats (Second ed.). O'Reilly. ISBN 1-56592-161-5.

Nanjo, T., Tanaka, S. (2020). Carbonate lithology identification with generative adversarial networks. In: International Petroleum Technology Conference (IPTC) 2020, Dhahran, Kingdom of Saudi Arabia, January 2020. doi: 10.2523/IPTC-20226-ABSTRACT.

Nanni, L., Paci, M., Brahnam, S., Lumini, A. (2021). Comparison of different image data augmentation approaches. Journal of Imaging, 7(12), 254. doi: 10.3390/jimaging7120254.

Nataprawira, J., Gu, Y., Goncharenko, I., Kamijo, S. (2021). Pedestrian detection using multispectral images and a deep neural network. Sensors 21(7), 2536. doi: 10.3390/s21072536.

Nelson, H.F., Brown, C.W., Brineman, J.H. (1962). Skeletal limestone classification. In: Classification of Carbonate Rocks (Ed. Ham, W.E.), American Association of Petroleum Geologists Memoirs 1, 224-252.

Nguyen, L.D., Lin, D., Lin, Z., Cao, J. (2018). Deep CNNs for microscopic image classification by exploiting transfer learning and feature concatenation. 2018 IEEE International Symposium on Circuits and Systems (ISCAS), Florence, 1-5.

Niu, Y., Wang, Y. D., Mostaghimi, P., Swietojanski, P., Armstrong, R. T. (2020). An innovative application of generative adversarial networks for physically accurate rock images with an unprecedented field of view. Geophysical Research Letters 47(23), e2020GL089029. doi: 10.1029/2020GL089029.

Obermeyer, Z., Emanuel, E. J. (2016). Predicting the future—big data, machine learning, and clinical medicine. The New England Journal of Medicine 375(13), 1216. doi: 10.1056/NEJMp1606181.

Odena, A., Dumoulin, V., Olah, C. (2016). Deconvolution and checkerboard artifacts. Distill 1(10), e3. doi: 10.23915/distill.00003.

Oikonomou, D., Alaei, B., Larsen, E., Jackson, C.A-L. (2017). Machine Learning in Petroleum Geoscience: Constructing EarthNET. NGF Winter Conference, Oslo, Norway, January 2017.

Oliphant, T.E. (2006). A guide to NumPy. Methods 1, 85.

Osman, A. M. S. (2019). A novel big data analytics framework for smart cities. Future Generation Computer Systems 91, 620-633. doi: 10.1016/j.future.2018.06.046.

Pal, A., Das, A. (2019). TorchGAN: a flexible framework for GAN training and evaluation. arXiv preprint arXiv:1909.03410. doi: 10.48550/arXiv.1909.03410.

Pan, S. J., Yang, Q. (2010). A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering 22(10), 1345-1359.

Panferov, V., Tailakov, D., Donets, A. (2020). Recognition of Rocks Lithology on the Images of Core Samples. 2020 Science and Artificial Intelligence Conference (S.A.I.ence), Novosibirsk, Russia, 54-57. doi: 10.1109/S.A.I.ence50533.2020.9303197.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A. (2019). Pytorch: An imperative style, high-performance deep learning library. Advances in Neural Information Processing Systems 32 (NeurIPS 2019).

Paul, A., Reijmer, J.J.G., Fürstenau, J., Kinkel, H., Betzler, C. (2012). Relationship between late Pleistocene sea-level variations, carbonate platform morphology and aragonite production (Maldives, Indian Ocean). Sedimentology 59(5), 1540-1658. doi: 10.1111/j.1365-3091.2011.01319.x.

Payne, J.L., Lehrmann, D.J., Wei, D., Knoll, A.H. (2006). The pattern and timing of biotic recovery from the end-Permian extinction on the Great Bank of Guizhou, Guizhou Province, China, PALAIOS 21(1), 63-85. doi: 10.2110/palo.2005.p05-12p.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research 12, 2825-2830.

Perol, T., Gharbi, M., Denolle, M. (2018). Convolutional neural network for earthquake detection and location. Science Advances 4(2), e1700578. doi: 10.1126/sciadv.1700578.

Pi, Y., Nath, N. D., Behzadan, A. H. (2020). Convolutional neural networks for object detection in aerial imagery for disaster response and recovery. Advanced Engineering Informatics, 43, 101009. doi: 10.1016/j.aei.2019.101009.

Picetti, F., Lipari, V., Bestagini, P., Tubaro, S. (2019). Seismic image processing through the generative adversarial network. Interpretation 7(3), SF15-SF26. doi: 10.1190/INT-2018-0232.1.

Pigram, C.J., Davies, P.J., Feary, D.A., Symonds, P.A. (1992). Absolute magnitude of the second-order middle to late, Miocene sea-level fall, Marion Plateau, northeast Australia. Geology 20(9), 858-862. doi: 10.1130/0091-7613(1992)020<0858:AMOTSO>2.3.CO;2.

Pires de Lima, R., Bonar, A.L., Coronado, D.D., Marfurt, K., Nicholson, C. (2019a). Deep convolutional neural networks as a geological image classification tool. SEPM Publications, The Sedimentary Record 17(2), 4- 9. doi: 10.2110/sedred.2019.2.4.

Pires de Lima, R., Suriamin, F., Marfurt, K.J., Pranter, M.J. (2019b). Convolutional neural networks as aid in core lithofacies classification. Interpretation 7, SF27-SF40. doi: 10.1190/INT-2018-0245.1.

Pires de Lima, R., Duarte, D., Nicholson, C., Slatt, R., Marfurt, K.J. (2020). Petrographic microfacies classification with deep convolutional neural networks. Computers & Geosciences 142, 104481.

Pomar, L. (2001). Types of carbonate platforms: a genetic approach. Basin Research 13(3), 313-334. doi: 10.1046/j.0950-091x.2001.00152.x.

Pomar, L., Brandano, M., Westphal, H. (2004). Environmental factors influencing skeletal grain sediment associations: a critical review of Miocene examples from the western Mediterranean. Sedimentology 51(3), 627-651. doi: 10.1111/j.1365-3091.2004.00640.x.

Pramanik, A., Pal, S.K., Maiti, J. Mitra, P. (2021). Granulated RCNN and Multi-Class Deep SORT for Multi-Object Detection and Tracking. IEEE Transactions on Emerging Topics in Computational Intelligence 6(1), 171-181. doi: 10.1109/TETCI.2020.3041019.

Preston, F.W., Henderson, J.H. (1964). Fourier Series Characterization of Cyclic Sediments for Stratigraphic Correlation. Symposium on cyclic sedimentation: Kansas Geological Survey, Bulletin 169, 415-425.

Purdy, E.G., Bertram, G.T. (1993). Carbonate Concepts from the Maldives, Indian Ocean. AAPG Studies in Geology 34, 7-55.

Qi, C. C. (2020). Big data management in the mining industry. International Journal of Minerals, Metallurgy and Materials 27, 131-139. doi: 10.1007/s12613-019-1937-z.

Qi, L., Carr, T.R. (2006). Neural network prediction of carbonate lithofacies from well logs, Big Bow and Sand Arroyo Creek fields, Southwest Kansas. Computers & Geosciences 32(7), 947-964. doi: 10.1016/j.cageo.2005.10.020.

Qian, N. (1999). On the momentum term in gradient descent learning algorithms. Neural Networks 12, 145-151.

Qian, F., Yin, M., Liu, X.Y., Wang, Y.J., Lu, C., Hu, G.M. (2018). Unsupervised seismic facies analysis via deep convolutional autoencoders. Geophysics 83(3), A39-A43.

Radford, A., Metz, L., Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434. doi: 10.48550/arXiv.1511.06434.

Ranjan, R., Tsukuda, S., Good, C. (2023). Effects of image data quality on a convolutional neural network trained in-tank fish detection model for recirculating aquaculture systems. Computers and Electronics in Agriculture 205, 107644. doi: 10.1016/j.compag.2023.107644.

Redmon, J., Divvala, S., Girshick, R., Farhadi, A. (2016). You only look once: Unified, real-time object detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2016, 779-788.

Redmon, J., Farhadi, A. (2017). YOLO9000: better, faster, stronger. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2017, 7263-7271.

Redmon, J., Farhadi, A. (2018). YOLOv3: An incremental improvement. arXiv preprint arXiv:1804.02767. doi: 10.48550/arXiv.1804.02767.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., Prabhat (2019). Deep learning and process understanding for data-driven Earth system science. Nature 566, 195-204.

Ren, Q., Zhang, H., Zhang, D., Zhao, X. (2023). Lithology identification using principal component analysis and particle swarm optimization fuzzy decision tree. Journal of Petroleum Science and Engineering 220(A), 111233. doi: 10.1016/j.petrol.2022.111233.

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. Advances in Neural Information Processing systems 28, 91-99.

Ren, S., He, K., Girshick, R, Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence 39(6), 1137-1149. doi: 10.1109/TPAMI.2016.2577031.

Richard, M. D., Lippmann, R. P. (1991). Neural network classifiers estimate Bayesian *a posteriori* probabilities. Neural computation, 3(4), 461-483. doi: 10.1162/neco.1991.3.4.461.

Ridgwell, A., Zeebe, R.E. (2005). The role of the global carbonate cycle in the regulation and evolution of the Earth system. Earth and Planetary Science Letters 234(3-4), 299-315. doi: 10.1016/j.epsl.2005.03.006.

Rittner, L., Flores, F.C., Lotufo, R.A. (2010). A tensorial framework for color images. Pattern Recognition Letters 31(4), 277-296. doi: 10.1016/j.patrec.2009.09.030.

Röglin, J., Ziegeler, K., Kube, J., König, F., Hermann, K. G., Ortmann, S. (2022). Improving classification results on a small medical dataset using a GAN; An outlook for dealing with rare disease datasets. Frontiers in Computer Science 4, 858874. doi: 10.3389/fcomp.2022.858874.

Rolnick, D., Donti, P.L., Kaack, L.H., Kochanski, K., Lacoste, A., Sankaran, K., Ross, A.S., Milojevic-Dupont, N., Jaques, N., Waldman-Brown, A., Luccioni, A.S. (2022). Tackling climate change with machine learning. ACM Computing Surveys (CSUR) 55(2), 1-96. doi: 10.1145/3485128.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18 (pp. 234-241). Springer International Publishing. doi: 10.1007/978-3-319-24574-4_28.

Rosebrock, A. (2019). Fine-tuning with Keras and Deep Learning. https://www.pyimagesearch.com/2019/06/03/fine-tuning-with-keras-and-deep-learning/.

Rubanenko, L., Pérez-López, S., Schull, J., Lapôtre, M. G. (2021). Automatic detection and segmentation of barchan dunes on Mars and Earth using a convolutional neural network. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 14, 9364-9371. doi: 10.1109/JSTARS.2021.3109900.

Rubo, R. A., Michelon, M. F., de Carvalho Carneiro, C. (2023). Carbonate lithofacies classification in optical microscopy: a data-centric approach using augmentation and GAN synthetic images. Earth Science Informatics 16(1), 617-635. doi: 10.1007/s12145-022-00901-9.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C. (2015). ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision 115, 211-252. doi: 10.1007/s11263-015-0816-y.

Ruuska, S., Hämäläinen, W., Kajava, S., Mughal, M., Matilainen, P., Mononen, J. (2018). Evaluation of the confusion matrix method in the validation of an automated system for measuring feeding behaviour of cattle. Behavioural Processes 148, 56-62. doi: 10.1016/j.beproc.2018.01.004.

Sager, W., Hoernle, K., Höfig, T.W., Avery, A.J., Bhutani, R., Carvallo, C.A., Class, C., Dai, Y., Dalla Valle, G., Del Gaudio, A.V., Fielding, S., Gaastra, K.M., Han, S., Heaton, D., Homrighausen, S., Kubota, Y., Li, C.-F., Nelson, W.R., Petrou, E., Potter, K.E., Pujatti, S., Scholpp, J., Shervais, J.W., Thoram, S., Tikoo-Schantz, S.M., Tshiningayamwe, M., Wang, X.-J., Widdowson, M., 2023. Expedition 391 methods. In Sager, W., Hoernle, K., Höfig, T.W., and the Expedition 391 Scientists, Walvis Ridge Hotspot. Proceedings of the International Ocean Discovery Program, 391: College Station, TX (International Ocean Discovery Program). doi: 10.14379/iodp.proc.391.102.2023.

Sahin, M. E., Ulutas, H., Yuce, E., Erkoc, M. F. (2023). Detection and classification of COVID-19 by using faster R-CNN and mask R-CNN on CT images. Neural Computing and Applications 35(18), 13597-13611. doi: 10.1007/s00521-023-08450-y.

Saikia, S., Fidalgo, E., Alegre, E., Fernández-Robles, L. (2017). Object Detection for Crime Scene Evidence Analysis Using Deep Learning. *In*: Battiato, S., Gallo, G., Schettini, R., Stanco, F. (eds) Image Analysis and Processing - ICIAP 2017. Lecture Notes in Computer Science 10485. Springer, Cham. doi: 10.1007/978-3-319-68548-9_2.

Samek, W., Montavon, G., Lapuschkin, S., Anders, C. J., Müller, K. R. (2021). Explaining deep neural networks and beyond: A review of methods and applications. Proceedings of the IEEE 109(3), 247-278. doi: 10.1109/JPROC.2021.3060483.

Saporetti, C.M., da Fonseca, L.G., Pereira, E., de Oliveira, L.C. (2018). Machine learning approaches for petrographic classification of carbonate-siliciclastic rocks using well logs and textural information. Journal of Applied Geophysics 155, 217-225.

Saporetti, C.M., Goliatt, L., Pereira, E. (2021). Neural network boosted with differential evolution for lithology identification based on well logs information. Earth Science Informations 14, 133-140. doi: 10.1007/s12145-020-00533-x.

Saucier, D., Sincock, K. (2024). An Operator view on trends in core acquisition. Geological Society, London, Special Publications, 527(1), SP527-2022. doi: 10.1144/sp527-2022-62.

Sauer, A., Schwarz, K., Geiger, A. (2022). StyleGAN-XL: Scaling StyleGAN to large diverse datasets. In: ACM SIGGRAPH 2022 Conference Proceedings 49, 1-10. doi: 10.1145/3528233.3530738.

Saura, J. R. (2021). Using data sciences in digital marketing: Framework, methods, and performance metrics. Journal of Innovation & Knowledge 6(2), 92-102. doi: 10.1016/j.jik.2020.08.001.

Schmidt, J., Marques, M. R., Botti, S., Marques, M. A. (2019). Recent advances and applications of machine learning in solid-state materials science. npj Computational Materials 5(83). doi: 10.1038/s41524-019-0221-0.

Scholle, P.A., Ulmer-Scholle, D.S. (2003). Carbonate Classification. A Color Guide to the Petrography of Carbonate Rocks: Grains, textures, porosity, diagenesis, AAPG Memoir 77, 284-292.

SciStarter.org (2020). Project Finder. SciStarter. Available from: https://scistarter.org/finder.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual Explanations From Deep Networks via Gradient-Based Localization. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 618-626.

Shams, R., Masihi, M., Boozarjomehry, R. B., Blunt, M. J. (2020). Coupled generative adversarial and auto-encoder neural networks to reconstruct three-dimensional multi-scale porous media. Journal of Petroleum Science and Engineering 186, 106794. doi: 10.1016/j.petrol.2019.106794.

Shen, C. (2018). A transdisciplinary review of deep learning research and its relevance for water resources scientists. Water Resources Research 54(11), 8558-8593. doi: 10.1029/2018WR022643.

Shin, H.C., Roth, H.R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., Summers, R.M. (2016). Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Transactions on Medical Imaging 35(5), 1285-1298.

Shoji, D., Noguchi, R., Otsuki, S., Hino, H. (2018). Classification of volcanic ash particles using a convolutional neural network and probability. Scientific Reports 8(1), 8111. doi: 10.1038/s41598-018-26200-2.

Shorten, C., Khoshgoftaar, T.M. (2019). A survey on Image Data Augmentation for Deep Learning. Journal of Big Data 6, 60. doi: 10.1186/s40537-019-0197-0.

Shreeve, J. W. (2024). The use of a Multi-Sensor Core Scanner Workflow as the Backbone of a Digital Core Repository. Geological Society, London, Special Publications, 527(1), SP527-2022. doi: 10.1144/SP527-2022-65.

Siddani, B., Balachandar, S., Moore, W. C., Yang, Y., Fang, R. (2021). Machine learning for physics-informed generation of dispersed multiphase flow using generative adversarial networks. Theoretical and Computational Fluid Dynamics 35, 807-830. doi: 10.1007/s00162-021-00593-9.

Simonyan, K., Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. International Conference on Learning Representations, 1-14.

Singh, S., Ahuja, U., Kumar, M., Kumar, K., & Sachdeva, M. (2021). Face mask detection using YOLOv3 and faster R-CNN models: COVID-19 environment. Multimedia Tools and Applications 80, 19753-19768. doi: 10.1007/s11042-021-10711-8.

Sircar, A., Yadav, K., Rayavarapu, K., Bist, N., Oza, H. (2021). Application of machine learning and artificial intelligence in oil and gas industry. Petroleum Research 6(4), 379-391. doi: 10.1016/j.ptlrs.2021.05.009.

Sit, M., Demiray, B. Z., Xiang, Z., Ewing, G. J., Sermet, Y., Demir, I. (2020). A comprehensive review of deep learning applications in hydrology and water resources. Water Science and Technology 82(12), 2635-2670. doi: 10.2166/wst.2020.369.

Skansi, S. (2018). Introduction to deep learning: from logical calculus to artificial intelligence. Springer, Berlin, 191 pp. doi: 10.1007/978-3-319-73004-2.

Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., Stefanovic, D. (2016). Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification. Computational Intelligence and Neuroscience 2016(6), 3289801. doi: 10.1155/2016/3289801.

Smith, L.N. (2017). Cyclical Learning Rates for Training Neural Networks. 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), 464-472. doi: 10.1109/WACV.2017.58

Smith, L.N. (2018). A Disciplined Approach to Neural Network Hyper-Parameters: Part 1--learning Rate, Batch Size, Momentum, and Weight Decay. arXiv preprint arXiv:1803.09820.

Soekhoe, D., van der Putten, P., Plaat, A. (2016). On the Impact of Data Set Size in Transfer Learning Using Deep Neural Networks. In: Boström H., Knobbe A., Soares C., Papapetrou P. (eds) Advances in Intelligent Data Analysis XV. IDA 2016. Lecture Notes in Computer Science 9897. Springer, Cham. doi: 10.1007/978-3-319-46349-0_5.

Solum, J.G., Auchter, N.C., Falivene, O., Cilona, A., Kleipool, L., Zarian, P. (2022). Accelerating core characterization and interpretation through deep learning with an application to legacy data sets. Interpretation 10(3), SE71-SE83. doi: 10.1190/INT-2021-0189.1.

Soviany, P., Ionescu, R. T. (2018). Optimizing the Trade-Off between Single-Stage and Two-Stage Deep Object Detectors using Image Difficulty Prediction. 20th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), Timisoara, Romania, 2018, 209-214. doi: 10.1109/SYNASC.2018.00041.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. Journal of Machine Learning Research 15, 1929-1958.

Strelcenia, E.; Prakoonwit, S. (2023). A Survey on GAN Techniques for Data Augmentation to Address the Imbalanced Data Issues in Credit Card Fraud Detection. Machine Learning & Knowledge Extraction 5, 304-329. doi: 10.3390/make5010019.

Sun, Y., Wang, W., Tang, X. (2015). Deeply learned face representations are sparse, selective, and robust. IEEE Conference on Computer Vision and Pattern Recognition, 2892-2900.

Supervisely (2023). Supervisely Computer Vision platform [Computer Vision Tools]. In: Supervisely Ecosystem. https://supervisely.com

Susskind, J. M., Anderson, A. K., Hinton, G. E. (2010). The Toronto Dace Database. Department of Computer Science, University of Toronto, Toronto, ON, Canada, Technical Report, 3, 29.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.B. (2016). Rethinking the Inception Architecture for Computer Vision. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2818-2826. doi: 10.1109/CVPR.2016.308.

Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., Liu, C. (2018). A Survey on Deep Transfer Learning. In: Kůrková V., Manolopoulos Y., Hammer B., Iliadis L., Maglogiannis I. (eds) Artificial Neural Networks and Machine Learning – ICANN 2018. ICANN 2018. Lecture Notes in Computer Science, Springer, Cham. 11141. doi: 10.1007/978-3-030-01424-7_27.

Tang, J., Fan, B., Xiao, L., Tian, S., Zhang, F., Zhang, L., Weitz, D. (2021). A new ensemble machine-learning framework for searching sweet spots in shale reservoirs. SPE Journal 26(01), 482-497. doi: 10.2118/204224-PA.

Tang, H., Toomey, N., Meddaugh, W.S. (2011). Using an Artificial-Neural-Network Method To Predict Carbonate Well Log Facies Successfully. SPE Reservoir Evaluation & Engineering 14(01), 35-44. doi: 10.2118/123988-PA.

Taylor, L., Nitschke, L. (2018). Improving Deep Learning with Generic Data Augmentation. IEEE Symposium Series on Computational Intelligence (SSCI), Bangalore, India, 2018, 1542-1547. doi: 10.1109/SSCI.2018.8628742.

Thambawita, V., Strümke, I., Hicks, S. A., Halvorsen, P., Parasa, S., Riegler, M. A. (2021). Impact of image resolution on deep learning performance in endoscopy image classification: An experimental study using a large dataset of endoscopic images. Diagnostics 11(12), 2183. doi: 10.3390/diagnostics11122183.

Theobald, E.J., Ettinger, A.K., Burgess, H.K., DeBey, L.B., Schmidt, N.R., Froehlich, H.E., Wagner, C., HilleRisLambers, J., Tewksbury, J., Harsch, M.A., Parrish, J.K. (2015). Global change and local solutions: Tapping the unrealized potential of citizen science for biodiversity research. Biological Conservation 181, 236-244.

Thomas, A., Rider, M., Curtis, A., MacArthur, A. (2011). Automated lithology extraction from core photographs. First Break 29(6), 103-109. doi: 10.3997/1365-2397.29.6.51281.

Tian, M., Tan, M., Wang, M. (2023a). Identification of Shale Lithofacies from FMI Images and ECS Logs Using Machine Learning with GLCM Features. Processes 11(10), 2982. doi: 10.3390/pr11102982.

Tian, X., Lu, W., Li, Y., Liu, J., Zhong, M., Pan, H., Jiang, B. (2023b). Improved Seismic Residual Diffracted Multiple Suppression Method Based on Object Detection and Image Segmentation. IEEE Transactions on Geoscience and Remote Sensing 61(5901213), 1-13. doi: 10.1109/TGRS.2023.3234568.

Timmer, E., Knudson, C., Gingras, M. (2021). Applying deep learning for identifying bioturbation from core photographs. AAPG Bulletin 105(4), 631-638. doi: 10.1306/0819201905.

Tiwari, S., Mishra, S., Srihariprasad, G., Vyas, D., Warhade, A., Nikalje, D., Bartakke, V., Mahesh, B., Tembhurnikar, P., Roy, S., (2017). High resolution core scan facility at BGRL-MoES, Karad, India. Journal of the Geological Society of India 90 ,795-797. doi: 10.1007/s12594-017-0793-6.

Todd, T.W. (1966). Petrogenetic classification of carbonate rocks. Journal of Sedimentary Petrology 36, 317-340.

Tomczak, M., Godfrey, J.S. (2003). Regional Oceanography: An Introduction (2nd ed.): Delhi (Daya Publishing House).

Torrisi, F., Folzani, F., Corradino, C., Amato, E., Del Negro, C. (2021). Detecting Volcanic Ash Plume Components from Space using Machine Learning Techniques. AGU 2021 Fall Meeting. doi: 10.1002/essoar.10509947.1.

Tschannen, V., Delescluse, M., Rodriguez, M., Keuper, J. (2017). Facies classification from well logs using an inception convolutional network. arXiv preprint.

Tucker, M.E., Wright, V.P. (2009). Carbonate sedimentology. John Wiley & Sons. doi: 10.1002/9781444314175.

Ubani, C. E., Adeboye, Y. B., Oriji, A. B. (2012). Advances in coring and core analysis for reservoir formation evaluation. Petroleum & Coal 54(1), 42-51.

USBR (1998). Guidelines for Core Logging, USBR Engineering Geology Field Manual, 2nd edition, vol. I, United States Department of the Interior Bureau of Reclamation, 432, 294-312.

Utami, D.W., Reuning, L., Cahyarini, SY. (2018). Satellite- and field-based facies mapping of isolated carbonate platforms from the Kepulauan Seribu Complex, Indonesia. The Depositional Record 4, 255-273. doi: 10.1002/dep2.47.

Valentín, M.B., Bom, C.R., Coelho, J.M., Correia, M.D., de Albuquerque, M.P., de Albuquerque, M.P., Faria, E.L. (2019). A deep residual convolutional neural network for automatic lithological facies identification in Brazilian pre-salt oilfield wellbore image logs. Journal of Petroleum Science and Engineering 179, 474-503. doi: 10.1016/j.petrol.2019.04.030.

Valentine, A., Kalnins, L. (2016). An introduction to learning algorithms and potential applications in geomorphometry and earth surface dynamics. Earth Surface Dynamics 4(2), 445-460. doi: 10.5194/esurf-4-445-2016.

Viola, P., Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR) 2001 1, I-I.

Volkhonskiy, D., Muravleva, E., Sudakov, O., Orlov, D., Burnaev, E., Koroteev, D., Belozerov, B., Krutko, V. (2022). Generative adversarial networks for reconstruction of three-dimensional porous media from two-dimensional slices. Physical Review E 105(2), 025304. doi: 10.1103/PhysRevE.105.025304.

Wamba, S. F., Akter, S., Edwards, A., Chopin, G., Gnanzou, D. (2015). How 'big data' can make big impact: Findings from a systematic review and a longitudinal case study. International Journal of Production Economics 165, 234-246. doi: 10.1016/j.ijpe.2014.12.031.

Wang, G., Carr, T.R. (2012). Marcellus Shale Lithofacies Prediction by Multiclass Neural Network Classification in the Appalachian Basin. Mathematical Geosciences 44, 975-1004. doi: 10.1007/s11004-012-9421-6.

Wang, R., Bashyam, V., Yang, Z., Yu, F., Tassopoulou, V., Chintapalli, S.S., Skampardoni, I., Sreepada, L.P., Sahoo, D., Nikita, K., Abdulkadir, A., Wen, J., Davatzikos, C. (2023a). Applications of generative adversarial networks in neuroimaging and clinical neuroscience. NeuroImage 269, 119898. doi: 10.1016/j.neuroimage.2023.119898.

Wang, H., Li, C., Zhang, Z., Kershaw, S., Holmer, L.E., Zhang, Y., Wei, K., Liu, P. (2022). Fossil brachiopod identification using a new deep convolutional neural network. Gondwana Research 105, 290-298. doi: 10.1016/j.gr.2021.09.011.

Wang, J. Z., Zhao, S., Wu, C., Adams, R. B., Newman, M. G., Shafir, T., Tsachor, R. (2023b). Unlocking the Emotional World of Visual Media: An Overview of the Science, Research, and Impact of Understanding Emotion. Proceedings of the IEEE 111(10), 1236-1286. doi: 10.1109/JPROC.2023.3273517.

Wang, K., Liu, M. (2021). A feature-optimized Faster regional convolutional neural network for complex background objects detection. IET Image Processing 15(2), 378-392. doi: 10.1049/ipr2.12028.

Wang, Z., Di, H., Shafiq, M. A., Alaudah, Y., AlRegib, G. (2018). Successful leveraging of image processing and machine learning in seismic structural interpretation: A review. The Leading Edge 37(6), 451-461. doi: 10.1190/tle37060451.1.

Wei, Z., Hu, H., Zhou, H. W., Lau, A. (2019). Characterizing rock facies using machine learning algorithm based on a convolutional neural network and data padding strategy. Pure and Applied Geophysics 176, 3593-3605. doi: 10.1007/s00024-019-02152-0.

Wei, X. L., Zhang, C. X., Kim, S. W., Jing, K. L., Wang, Y. J., Xu, S., Xie, Z. Z. (2022). Seismic fault detection using convolutional neural networks with focal loss. Computers & Geosciences 158, 104968. doi: 10.1016/j.cageo.2021.104968.

West, B. P., May, S. R., Eastwood, J. E., Rossen, C. (2002). Interactive seismic facies classification using textural attributes and neural networks. The Leading Edge 21(10), 1042-1049. doi: 10.1190/1.1518444.

Wickman, F.E. (1965). Repose period patterns of volcanoes, 5. General discussion and a tentative stochastic model. Arkiv for Mineralogi Och Geologi 4(5), 351-367.

Wilkens, R.H., Niklis, N., Frazer, M. (2009). Data report: digital core images as data: an example from IODP Expedition 303. In: Channell, J.E.T., Kanamatsu, T., Sato, T., Stein, R., Alvarez Zarikian, C.A., Malone, M.J., and the Expedition 303/306 Scientists, Proceedings of the Integrated Ocean Drilling Program 303/306, 1-16. doi: 10.2204/iodp.proc.303306.201.2009.

Wood, D.A. (2022). Carbonate / siliciclastic lithofacies classification aided by well-log derivative, volatility and sequence boundary attributes combined with machine learning. Earth Science Informatics 15, 1699-1721. doi: 10.1007/s12145-022-00829-0

Wong, S. C., Gatt, A., Stamatescu, V., McDonnell, M. D. (2016). Understanding data augmentation for classification: when to warp?. In: 2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA). doi: 10.1109/DICTA.2016.7797091.

Wright, V.P. (1992). A revised classification of limestones. Sedimentary Geology 76, 177-185.

Wu, H., Zhang, B., Lin, T., Cao, D., Lou, Y. (2019). Semiautomated seismic horizon interpretation using the encoder-decoder convolutional neural network. Geophysics 84(6), B403-B417. doi: 10.1190/geo2018-0672.1.

Wu, X., Liang, L., Shi, Y., Fomel, S. (2018). FaultSeg3D: Using synthetics to train an end-to-end convolutional neural network for 3D seismic fault segmentation. Geophysics 84(3), IM35-IM43. doi: 10.1190/geo2018-0646.1.

Xiao, Y., Tian, Z., Yu, J., Zhang, Y., Liu, S., Du, S., Lan, X. (2020). A review of object detection based on deep learning. Multimedia Tools and Applications 79, 23729-23791. doi: 10.1007/s11042-020-08976-6.

Xie, Y., Zhu, C., Zhou, W., Li, Z., Liu, X., Tu, M. (2018). Evaluation of machine learning methods for formation lithology identification: A comparison of tuning processes and model performances. Journal of Petroleum Science and Engineering 160, 182-193. doi: 10.1016/j.petrol.2017.10.028.

Xie, Y., Ebad Sichani, M., Padgett, J.E., DesRoches, R. (2020). The promise of implementing machine learning in earthquake engineering: A state-of-the-art review. Earthquake Spectra 36(4), 1769-1801. doi: 10.1177/8755293020919419.

Xiong, W., Ji, X., Ma, Y., Wang, Y., AlBinHassan, N. M., Ali, M. N., Luo, Y. (2018). Seismic fault detection with convolutional neural network. Geophysics 83(5), O97-O103. doi: 10.1190/geo2017-0666.1.

Xu, D., Wu, Y. (2020). MRFF-YOLO: a multi-receptive fields fusion network for remote sensing target detection. Remote Sensing 12(19), 3118. doi: 10.3390/rs12193118.

Xu, J., Ren, H., Cai, S., Zhang, X. (2023). An improved faster R-CNN algorithm for assisted detection of lung nodules. Computers in Biology and Medicine 153, 106470. doi: 10.1016/j.compbiomed.2022.106470.

Xu, Z., Ma, W., Lin, P., Shi, H., Pan, D. Liu, T. (2021). Deep learning of rock images for intelligent lithology identification. Computers & Geosciences 154, 104799. doi: 10.1016/j.cageo.2021.104799.

Xu, Z., Shi, H., Lin, P., Ma, W. (2022). Intelligent on-site lithology identification based on deep learning of rock images and elemental data. IEEE Geoscience and Remote Sensing Letters 19(6511205), 1-5. doi: 10.1109/LGRS.2022.3179623.

Yamashita, R., Nishio, M., Do, R.K.G., Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. Insights Imaging 9, 611-629. doi: 10.1007/s13244-018-0639-9.

Yang, S., Cai, Z. (2021). High-resolution feature pyramid network for automatic Crater detection on Mars. IEEE Transactions on Geoscience and Remote Sensing 60, 1-12. doi: 10.1109/TGRS.2021.3104925.

Yang, R., Yu, Y. (2021) Artificial convolutional neural network in object detection and semantic segmentation for medical imaging analysis. Frontiers in Oncology 11, 638182. doi: 10.3389/fonc.2021.638182.

Yates, M., Hart, G., Houghton, R., Torres, M. T., Pound, M. (2022). Evaluation of synthetic aerial imagery using unconditional generative adversarial networks. ISPRS Journal of Photogrammetry and Remote Sensing 190, 231-251. doi: 10.1016/j.isprsjprs.2022.06.010.

Yavariabdi, A., Kusetogullari, H., Orhan, O., Uray, E., Demir, V., Celik, T., Mendi, E. (2023). SinkholeNet: A novel RGB-slope sinkhole dataset and deep weakly-supervised learning framework for sinkhole classification and localization. The Egyptian Journal of Remote Sensing and Space Sciences 26(4), 966-973. doi: 10.1016/j.ejrs.2023.10.006.

Yin, R., Teng, Q., Wu, X., Zhang, F., Xiong, S. (2023). Three-dimensional reconstruction of granular porous media based on deep generative models. Physical Review E 108(5), 055303. doi: 10.1103/PhysRevE.108.055303.

Yosinski, J., Clune, J., Bengio, Y., Lipson, H., Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N.D., Weinberger, K.Q. (2014). How transferable are features in deep neural networks? Proceedings of the 27th International Conference on Neural Information Processing Systems (NIPS'14) 2, 3320-3328.

Yu, S., Ma, J. (2021). Deep learning for geophysics: Current and future trends. Reviews of Geophysics 59(3), e2021RG000742. doi: 10.1029/2021RG000742.

Yu, X., Ye, K., Du, C., Gong, H., Ma., Z. (2021). Microscopic recognition of micro fossils in carbonate rocks based on convolutional neural network. 石油实验地质 (Petroleum Geology & Experiment) 43(5), 880-885. doi: 10.11781/sysydz202105880.

Zhang, J., Ambrose, W., Xie, W. (2021a). Applying convolutional neural networks to identify lithofacies of large-n cores from the Permian Basin and Gulf of Mexico: The importance of the quantity and quality of training data. Marine and Petroleum Geology 133, 105307. doi: 10.1016/j.marpetgeo.2021.105307.

Zhang, P.Y., Sun, J.M., Jiang, Y.J., Gao, J.S. (2017). Deep Learning Method for Lithology Identification from Borehole Images. Paper presented at the EAGE Conference and Exhibition, Paris, France, 12-15 June.

Zhang, T., Hu, G., Yang, Y., Du, Y. (2023a). A Super-Resolution Reconstruction Method for Shale Based on Generative Adversarial Network. Transport in Porous Media 150(2), 383-426. doi: 10.1007/s11242-023-02016-1.

Zhang, T. F., Tilke, P., Dupont, E., Zhu, L. C., Liang, L., Bailey, W. (2019). Generating geologically realistic 3D reservoir facies models using deep learning of sedimentary architecture with generative adversarial networks. Petroleum Science 16, 541-549. 10.1007/s12182-019-0328-4.

Zhang, Z., Lin, Y. (2020). Data-driven seismic waveform inversion: A study on the robustness and generalization. IEEE Transactions on Geoscience and Remote sensing 58(10), 6900-6913. doi: 10.1109/TGRS.2020.2977635.

Zhang, C., Zuo, R., Xiong, Y. (2021b) Detection of the multivariate geochemical anomalies associated with mineralization using a deep convolutional neural network and a pixel-pair feature method. Applied Geochemistry 130, 104994. doi: 10.1016/j.apgeochem.2021.104994.

Zhang, X., Han, L., Han, L., Zhu, L. (2020) How well do deep learning-based methods for land cover classification and object detection perform on high resolution remote sensing imagery? Remote Sensing 12(3), 417. doi: 10.3390/rs12030417.

Zhang, Y., Shen, S., Xu, S. (2023b) Strip steel surface defect detection based on lightweight YOLOv5. Frontiers in Neurorobotics 17, 1263739. doi: 10.3389/fnbot.2023.1263739.

Zhao, S., Yao, X., Yang, J., Jia, G., Ding, G., Chua, T.S., Schuller, B.W., Keutzer, K. (2022). Affective image content analysis: Two decades review and new perspectives. IEEE Transactions on Pattern Analysis and Machine Intelligence 44(10), 6729-6751. doi: 10.1109/TPAMI.2021.3094362.

Zhao, Y., Niu, F., Zhang, Z., Li, X., Chen, J., Yang, J. (2021). Signal detection and enhancement for seismic crosscorrelation using the wavelet-domain Kalman filter. Surveys in Geophysics 42, 43-67. doi: 10.1007/s10712-020-09620-6.

Zhao, Z.-Q., Zheng, P., Xu, S.-T., Wu, X. (2019). Object Detection With Deep Learning: A Review. IEEE Transactions on Neural Networks and Learning Systems 30(11), 3212-3232, doi: 10.1109/TNNLS.2018.2876865.

Zheng, D., Hou, M., Chen, A., Zhong, H., Qi, Z,, Ren, Q., You, J., Wang, H., Ma, C. (2022). Application of machine learning in the identification of fluvial-lacustrine lithofacies from well logs: A case study from Sichuan Basin, China. Journal of Petroleum Science and Engineering 215(A), 110610. doi: 10.1016/j.petrol.2022.110610.

Zhenlong, H., Jikang, W., Jinrong, S., Xinwei, L., Wentian, Z. (2023). Intelligent Lithology Identification Methods for Rock Images Based on Object Detection. Natural Resources Research 32, 2965-2980. doi: 10.1007/s11053-023-10271-8.

Zhou, P., Cheng, G., Liu, Z., Bu, S., Hu, X. (2016). Weakly supervised target detection in remote sensing images based on transferred deep features and negative bootstrapping. Multidimensional Systems and Signal Processing 27, 925-944. doi: 10.1007/s11045-015-0370-3.

Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., He, Q. (2020). A comprehensive survey on transfer learning. Proceedings of the IEEE 109(1), 43-76. doi: 10.1109/JPROC.2020.3004555.

Zuo, R., Xiong, Y., Wang, J., Carranza, E.J.M. (2019). Deep learning and its application in geochemical mapping. Earth-Science Reviews 192, 1-14. doi: 10.1016/j.earscirev.2019.02.023.

Appendix A: List of cores and assigned classes for image classification

Table A.1. List of all cores used in dataset 104k, along with the corresponding class assignments. Core IDs for a sample are given as the full curatorial identifier according to the IODP naming convention and consist of the following information: expedition, site, hole, core number, core type, section number. Where Core IDs appear multiple times with different Dunham classes, this reflects the presence of multiple textural classes within the core.

Core ID	Dunham Class
133-811A-19H-2	floatstone
133-811A-19H-2	packstone
133-811A-20H-1	floatstone
133-811A-3H-3	packstone
133-811A-3H-5	floatstone
133-811B-16X-1	rudstone
133-811B-18X-1	grainstone
133-811B-18X-2	grainstone
133-811B-19X-1	grainstone
133-811C-3H-2	floatstone
133-811C-4H-3	floatstone
133-812A-12X-1	packstone
133-812A-18X-1	floatstone
133-812A-20X-1	packstone
133-812A-5X-1	grainstone
133-812B-10R-1	floatstone
133-812B-11R-1	floatstone
133-812B-12R-1	floatstone
133-812B-13R-1	floatstone
133-812B-2W-1	grainstone
133-812B-2W-2	grainstone
133-812B-3R-1	grainstone
133-812B-4R-1	wackestone
133-812B-5R-1	wackestone
133-812B-6R-1	grainstone
133-812B-7R-1	rudstone
133-812B-9R-1	wackestone
133-812C-12H-4	packstone
133-812C-12H-5	packstone
133-812C-12H-6	packstone
133-813A-21H-5	packstone
133-813A-22X-1	grainstone

133-814A-14H-3	packstone
133-814A-14H-4	packstone
133-814A-15H-1	packstone
133-814A-16H-2	packstone
133-814A-23X-1	mudstone
133-814A-23X-2	mudstone
133-814A-24X-1	mudstone
133-814A-24X-2	mudstone
133-814A-24X-3	mudstone
133-814A-26X-1	mudstone
133-814A-26X-2	mudstone
133-814A-26X-3	mudstone
133-814A-26X-4	mudstone
133-814A-26X-5	mudstone
133-814A-27X-1	mudstone
133-814A-27X-2	mudstone
133-814A-27X-3	mudstone
133-814A-27X-4	mudstone
133-814A-28X-1	mudstone
133-814A-28X-2	mudstone
133-814A-28X-3	mudstone
133-814A-29X-4	floatstone
133-815A-46X-1	packstone
133-815A-46X-CC	packstone
133-815A-47X-1	packstone
133-816A-10H-7	mudstone
133-816A-10H-CC	mudstone
133-816A-11H-1	rudstone
133-816A-11H-2	floatstone
133-816A-11H-CC	floatstone
133-816A-12X-1	floatstone
133-816A-13X-CC	floatstone
133-816A-14X-1	floatstone
133-816A-15X-1	boundstone
133-816B-1R-1	mudstone
133-816B-1R-2	mudstone
133-816B-1R-CC	floatstone
133-816B-2R-1	floatstone
133-816B-2R-1	rudstone
133-816B-3R-1	rudstone
133-816B-4R-1	rudstone
133-816B-7R-1	rudstone
133-816B-8R-1	rudstone
133-816B-8R-2	rudstone
133-816C-10R-1	boundstone

133-816C-11R-1	boundstone
133-816C-11R-2	boundstone
133-816C-11R-3	boundstone
133-816C-12R-1	boundstone
133-816C-13R-1	boundstone
133-816C-1R-4	floatstone
133-816C-4R-1	boundstone
133-816C-4R-1	floatstone
133-816C-5R-1	boundstone
133-816C-6R-1	boundstone
133-816C-8R-1	boundstone
133-816C-8R-2	boundstone
133-816C-9R-1	boundstone
133-817D-19R-1	wackestone
133-817D-23R-1	boundstone
133-817D-23R-1	rudstone
133-817D-26R-1	packstone
133-817D-27R-1	rudstone
133-817D-30R-1	wackestone
133-817D-33R-1	wackestone
133-819A-25X-2	wackestone
133-819A-27X-2	wackestone
133-819A-27X-3	wackestone
133-820A-13H-1	mudstone
133-820A-13H-2	mudstone
133-820A-13H-3	mudstone
133-820A-14H-1	mudstone
133-820A-14H-3	mudstone
133-820A-15H-6	mudstone
133-820A-20X-1	mudstone
133-820A-28X-1	mudstone
133-820A-6H-1	mudstone
133-820A-6H-2	mudstone
133-820A-6H-3	mudstone
133-821A-16H-3	mudstone
133-821A-16H-4	mudstone
133-821A-18X-1	mudstone
133-821A-21X-3	wackestone
133-821A-24X-1	mudstone
133-821A-30X-1	mudstone
133-821A-33X-3	mudstone
133-821A-42X-1	mudstone
133-821B-17V-2	floatstone
133-821B-17V-3	mudstone
133-821B-20X-1	mudstone

133-823A-6H-4	floatstone
133-823B-15X-1	rudstone
133-823B-20X-2	floatstone
133-823B-20X-3	floatstone
133-823B-20X-3	rudstone
133-823B-22X-6	rudstone
133-823B-62X-2	mudstone
133-823B-6H-4	mudstone
133-823C-10R-CC	floatstone
133-823C-16R-2	rudstone
133-823C-20R-2	rudstone
133-823C-7R-6	floatstone
133-824A-10X-1	rudstone
133-824A-14H-1	rudstone
133-824A-21X-1	mudstone
133-824A-28X-CC	floatstone
133-824A-28X-CC	rudstone
133-824A-29X-CC	floatstone
133-824A-29X-CC	rudstone
133-824A-30X-CC	floatstone
133-824A-30X-CC	rudstone
133-824A-32X-CC	rudstone
133-824A-36X-CC	rudstone
133-824A-5H-1	rudstone
133-824A-5H-6	rudstone
133-824C-11R-CC	floatstone
133-824C-11R-CC	rudstone
133-824C-2R-CC	floatstone
133-824C-5R-CC	floatstone
133-824C-6R-CC	floatstone
133-824C-7R-CC	floatstone
133-824C-8R-1	floatstone
133-826A-11R-1	rudstone
133-826A-14R-1	rudstone
133-826A-16R-1	rudstone
133-826A-4R-1	boundstone
133-826A-4R-1	rudstone
133-826A-5R-1	rudstone
194-1192A-17H-1	mudstone
194-1192A-18H-1	mudstone
194-1192A-19H-1	mudstone
194-1192A-19H-2	mudstone
194-1192A-20H-1	mudstone
194-1192A-21H-1	mudstone
194-1192A-22H-3	mudstone

194-1192A-25H-5	wackestone
194-1192A-26H-1	wackestone
194-1192A-26H-3	packstone
194-1192A-27H-1	wackestone
194-1192A-2H-1	packstone
194-1192A-3H-1	grainstone
194-1192A-5H-3	packstone
194-1192A-6H-3	wackestone
194-1192B-1H-1	grainstone
194-1192B-21X-2	grainstone
194-1192B-6H-2	wackestone
194-1192B-8H-1	wackestone
194-1192B-9X-2	grainstone
194-1193A-10X-1	boundstone
194-1193A-18X-1	floatstone
194-1193A-1H-1	grainstone
194-1193A-27X-1	rudstone
194-1193A-29X-1	boundstone
194-1193A-31X-2	mudstone
194-1193A-32X-1	mudstone
194-1193A-33X-2	floatstone
194-1193A-39X-1	floatstone
194-1193A-41X-2	grainstone
194-1193A-43X-1	mudstone
194-1193A-44X-1	mudstone
194-1193A-45X-4	mudstone
194-1193A-60X-1	grainstone
194-1193A-60X-2	grainstone
194-1193A-60X-3	grainstone
194-1193A-63X-1	grainstone
194-1193A-79X-1	grainstone
194-1193A-81X-1	grainstone
194-1193B-1R-2	rudstone
194-1193B-2R-1	rudstone
194-1193B-6Z-1	floatstone
194-1193C-4R-2	boundstone
194-1193C-5R-1	grainstone
194-1193C-5R-2	grainstone
194-1193C-5R-3	grainstone
194-1194A-10H-1	wackestone
194-1194A-13H-1	wackestone
194-1194A-16X-1	floatstone
194-1194A-1H-1	grainstone
194-1194A-6H-1	mudstone
194-1194A-7H-4	mudstone

194-1194A-8H-1	mudstone
194-1194A-9H-1	mudstone
194-1194B-11R-1-CC	packstone
194-1194B-18R-1	grainstone
194-1194B-19R-1	grainstone
194-1194B-20R-1	grainstone
194-1194B-21R-1	grainstone
194-1194B-22R-1	grainstone
194-1194B-23R-1	grainstone
194-1194B-25R-1	grainstone
194-1194B-25R-2	grainstone
194-1194B-25R-3	grainstone
194-1194B-25R-4	grainstone
194-1194B-25R-5	grainstone
194-1194B-3R-1	floatstone
194-1194B-4R-1	floatstone
194-1195A-2H-2	grainstone
194-1195A-2H-3	grainstone
194-1195A-3H-3	grainstone
194-1195B-19H-1	wackestone
194-1195B-21H-1	wackestone
194-1195B-30X-1	wackestone
194-1195B-36X-1	mudstone
194-1195B-37X-1	mudstone
194-1195B-38X-1	mudstone
194-1195B-45X-2	wackestone
194-1195B-46X-1	wackestone
194-1196A-10R-1	floatstone
194-1196A-11R-1	floatstone
194-1196A-12R-1	floatstone
194-1196A-13R-1	floatstone
194-1196A-13R-2	floatstone
194-1196A-14R-1	floatstone
194-1196A-14R-2	boundstone
194-1196A-15R-1	boundstone
194-1196A-16R-1	floatstone
194-1196A-17R-1	floatstone
194-1196A-18R-1	rudstone
194-1196A-18R-2	rudstone
194-1196A-19R-1	floatstone
194-1196A-19R-2	floatstone
194-1196A-1R-1	rudstone
194-1196A-20R-1	rudstone
194-1196A-21R-1	floatstone
194-1196A-25R-1	floatstone

194-1196A-2R-1	floatstone
194-1196A-32R-1	floatstone
194-1196A-33R-1	boundstone
194-1196A-33R-1	floatstone
194-1196A-33R-2	floatstone
194-1196A-34R-1	floatstone
194-1196A-35R-1	floatstone
194-1196A-36R-1	boundstone
194-1196A-37R-1	boundstone
194-1196A-3R-1	floatstone
194-1196A-41R-1	floatstone
194-1196A-42R-1	floatstone
194-1196A-44R-1	floatstone
194-1196A-48R-1	floatstone
194-1196A-4R-1	floatstone
194-1196A-53R-1	floatstone
194-1196A-53R-3	floatstone
194-1196A-53R-4	boundstone
194-1196A-54R-1	rudstone
194-1196A-57R-2	floatstone
194-1196A-57R-3	floatstone
194-1196A-58R-1	floatstone
194-1196A-59R-1	rudstone
194-1196A-5R-1	floatstone
194-1196A-5R-1	rudstone
194-1196A-61R-1	floatstone
194-1196A-66R-1	rudstone
194-1196A-66R-2	rudstone
194-1196A-66R-3	rudstone
194-1196A-6R-1	rudstone
194-1196A-70R-3	floatstone
194-1196A-7R-1	rudstone
194-1196A-7R-2	floatstone
194-1196A-8R-1	floatstone
194-1196A-9R-1	rudstone
194-1196B-10R-1	floatstone
194-1196B-11R-1	floatstone
194-1196B-12R-1	floatstone
194-1196B-13Z-1	rudstone
194-1196B-14Z-1	rudstone
194-1196B-14Z-2	rudstone
194-1196B-15Z-1	floatstone
194-1196B-16Z-1	rudstone
194-1196B-1R-2	floatstone
194-1196B-22Z-1	boundstone

194-1196B-29Z-1	rudstone
194-1196B-29Z-2	rudstone
194-1196B-2R-1	floatstone
194-1196B-30Z-1	rudstone
194-1196B-31Z-1	rudstone
194-1196B-4R-1	floatstone
194-1196B-5R-1	floatstone
194-1196B-6R-1	floatstone
194-1196B-7R-1	floatstone
194-1196B-8R-1	floatstone
194-1196B-9R-1	floatstone
194-1197B-44R-1	wackestone
194-1197B-44R-2	wackestone
194-1197B-46R-2	wackestone
194-1197B-47R-1	mudstone
194-1197B-53R-5	rudstone
194-1197B-57R-1	mudstone
194-1197B-64R-1	floatstone
194-1198A-16H-1	wackestone
194-1198A-16H-2	mudstone
194-1198A-17H-1	wackestone
194-1198A-17H-7	mudstone
194-1198A-1H-1	wackestone
194-1198A-20H-1	wackestone
194-1198A-22R-4	mudstone
194-1198A-22R-5	mudstone
194-1198A-23R-1	mudstone
194-1198A-2H-6	wackestone
194-1198A-4H-2	mudstone
194-1198A-6H-1	wackestone
194-1198A-7H-1	wackestone
194-1198B-10R-1	floatstone
194-1198B-33R-2	floatstone
194-1199A-10R-1	floatstone
194-1199A-10R-2	floatstone
194-1199A-11R-1	floatstone
194-1199A-12R-5	floatstone
194-1199A-13R-1	floatstone
194-1199A-14R-1	floatstone
194-1199A-14R-3	boundstone
194-1199A-15R-1	boundstone
194-1199A-16R-1	boundstone
194-1199A-16R-1	floatstone
194-1199A-17R-1	floatstone
194-1199A-18R-1	boundstone

194-1199A-18R-1	floatstone
194-1199A-18R-2	floatstone
194-1199A-1R-1	floatstone
194-1199A-1R-2	floatstone
194-1199A-1R-3	floatstone
194-1199A-2R-1	floatstone
194-1199A-2R-2	floatstone
194-1199A-2R-3	floatstone
194-1199A-2R-4	floatstone
194-1199A-39R-1	floatstone
194-1199A-3R-1	floatstone
194-1199A-40R-1	floatstone
194-1199A-42R-1	floatstone
194-1199A-4R-1	floatstone
194-1199A-4R-2	floatstone
194-1199A-4R-3	floatstone
194-1199A-4R-4	floatstone
194-1199A-5R-1	floatstone
194-1199A-5R-2	floatstone
194-1199A-6R-1	floatstone
194-1199A-7R-1	floatstone
194-1199A-8R-1	floatstone
194-1199A-9R-2	floatstone
359-U1465A-10H-CC	floatstone
359-U1465A-11X-CC	floatstone
359-U1465A-1H-1	floatstone
359-U1465A-1H-1	grainstone
359-U1465A-1H-2	floatstone
359-U1465A-1H-2	grainstone
359-U1465A-1H-3	grainstone
359-U1465A-1H-4	grainstone
359-U1465A-1H-5	grainstone
359-U1465A-1H-CC	grainstone
359-U1465A-2H-1	rudstone
359-U1465A-4H-1	grainstone
359-U1465A-4H-2	grainstone
359-U1465A-4H-3	grainstone
359-U1465A-4H-4	grainstone
359-U1465A-4H-5	grainstone
359-U1465B-12R-1	grainstone
359-U1465B-13R-1	grainstone
359-U1465B-13R-1	rudstone
359-U1465B-14R-1	grainstone
359-U1465B-14R-1	rudstone
359-U1465B-16R-1	rudstone

359-U1465B-17R-1	rudstone
359-U1465B-18R-1	floatstone
359-U1465B-18R-1	grainstone
359-U1465B-19R-1	packstone
359-U1465B-20R-1	packstone
359-U1465B-4R-1	rudstone
359-U1465B-5R-1	rudstone
359-U1465B-8R-1	grainstone
359-U1465B-9R-1	grainstone
359-U1465C-10X-1	boundstone
359-U1465C-11X-1	boundstone
359-U1465C-11X-1	mudstone
359-U1465C-12X-1	boundstone
359-U1465C-13X-CC	floatstone
359-U1465C-15F-2	floatstone
359-U1465C-15F-CC	floatstone
359-U1465C-16X-CC	floatstone
359-U1465C-17X-CC	floatstone
359-U1465C-17X-CC	mudstone
359-U1465C-20X-1	mudstone
359-U1465C-21X-1	mudstone
359-U1465C-22X-CC	packstone
359-U1465C-23X-1	mudstone
359-U1465C-24X-1	floatstone
359-U1465C-25X-1	floatstone
359-U1465C-3F-3	packstone
359-U1465C-8F-1	rudstone
359-U1466A-10H-1	rudstone
359-U1466A-10H-2	packstone
359-U1466A-14H-5	packstone
359-U1466A-18X-1	wackestone
359-U1466A-19F-1	wackestone
359-U1466A-19F-2	wackestone
359-U1466A-19H-1	rudstone
359-U1466A-20F-2	wackestone
359-U1466A-21H-3	wackestone
359-U1466A-22H-5	wackestone
359-U1466A-24X-2	mudstone
359-U1466A-26F-2	packstone
359-U1466A-27F-2	packstone
359-U1466A-30F-2	wackestone
359-U1466A-31F-2	wackestone
359-U1466A-31F-3	wackestone
359-U1466A-31F-CC	wackestone
359-U1466A-32F-3	wackestone

359-U1466A-33F-2	wackestone
359-U1466A-38X-2	packstone
359-U1466A-3H-2	packstone
359-U1466A-3H-3	packstone
359-U1466A-43F-2	packstone
359-U1466A-44X-2	grainstone
359-U1466A-44X-3	grainstone
359-U1466A-44X-CC	grainstone
359-U1466A-47F-2	packstone
359-U1466A-4H-1	rudstone
359-U1466A-5H-2	packstone
359-U1466A-5H-3	packstone
359-U1466A-8H-3	packstone
359-U1466A-9H-6	packstone
359-U1466B-13R-1	packstone
359-U1466B-2R-1	packstone
359-U1466B-47R-2	packstone
359-U1466B-55R-1	mudstone
359-U1466B-55R-2	mudstone
359-U1466B-55R-CC	mudstone
359-U1466B-56R-1	mudstone
359-U1466B-57R-3	mudstone
359-U1467A-1H-1	mudstone
359-U1467A-2H-1	mudstone
359-U1467B-24H-4	mudstone
359-U1467B-2H-7	mudstone
359-U1467B-4H-4	mudstone
359-U1468A-100X-1	mudstone
359-U1468A-102X-1	mudstone
359-U1468A-102X-2	mudstone
359-U1468A-103X-1	mudstone
359-U1468A-103X-2	mudstone
359-U1468A-103X-3	mudstone
359-U1468A-106X-2	floatstone
359-U1468A-106X-3	floatstone
359-U1468A-109X-2	floatstone
359-U1468A-109X-CC	floatstone
359-U1468A-10H-1	rudstone
359-U1468A-10H-3	rudstone
359-U1468A-10H-4	grainstone
359-U1468A-10H-5	packstone
359-U1468A-11H-1	rudstone
359-U1468A-22F-1	packstone
359-U1468A-29F-2	packstone
359-U1468A-31F-1	packstone

359-U1468A-31F-2	packstone
359-U1468A-34F-1	wackestone
359-U1468A-37F-2	mudstone
359-U1468A-37F-3	mudstone
359-U1468A-3H-3	grainstone
359-U1468A-3H-5	packstone
359-U1468A-49F-2	packstone
359-U1468A-51F-1	packstone
359-U1468A-53X-1	packstone
359-U1468A-58X-CC	mudstone
359-U1468A-59X-2	packstone
359-U1468A-5H-2	grainstone
359-U1468A-5H-3	packstone
359-U1468A-60X-CC	wackestone
359-U1468A-61X-2	packstone
359-U1468A-67X-2	packstone
359-U1468A-68X-2	wackestone
359-U1468A-69X-4	wackestone
359-U1468A-6H-4	rudstone
359-U1468A-6H-5	rudstone
359-U1468A-70X-3	wackestone
359-U1468A-7H-1	rudstone
359-U1468A-7H-2	rudstone
359-U1468A-7H-3	rudstone
359-U1468A-7H-4	grainstone
359-U1468A-7H-5	rudstone
359-U1468A-88X-3	wackestone
359-U1468A-8H-2	rudstone
359-U1468A-8H-3	rudstone
359-U1468A-8H-4	grainstone
359-U1468A-8H-5	rudstone
359-U1468A-94X-1	floatstone
359-U1468A-98X-1	mudstone
359-U1468A-98X-2	mudstone
359-U1468A-9H-1	grainstone
359-U1469A-17R-1	floatstone
359-U1469A-17R-2	floatstone
359-U1469A-17R-CC	floatstone
359-U1469B-2R-CC	rudstone
359-U1469B-2R-CC-A	rudstone
359-U1469B-3R-CC	rudstone
359-U1469B-4R-CC	rudstone
359-U1470A-10H-3	packstone
359-U1470A-11H-2	packstone
359-U1470A-12H-3	packstone

359-U1470A-17H-1	grainstone
359-U1470A-17H-2	grainstone
359-U1470A-17H-3	floatstone
359-U1470A-17H-3	grainstone
359-U1470A-1H-1	grainstone
359-U1470A-20F-2	rudstone
359-U1470A-21F-1	floatstone
359-U1470A-21F-1	grainstone
359-U1470A-21F-2	floatstone
359-U1470A-21F-2	grainstone
359-U1470A-2H-2	grainstone
359-U1470A-2H-3	grainstone
359-U1470A-3H-1	grainstone
359-U1470A-3H-2	grainstone
359-U1470A-4H-1	grainstone
359-U1470A-4H-2	grainstone
359-U1470A-4H-5	packstone
359-U1470A-5H-3	grainstone
359-U1470A-5H-3	packstone
359-U1470A-5H-4	grainstone
359-U1470A-8H-3	packstone
359-U1470B-12R-CC	floatstone
359-U1470B-13R-CC	floatstone
359-U1470B-14R-CC	floatstone
359-U1470B-16R-CC	floatstone
359-U1470B-18R-1	boundstone
359-U1470B-18R-1	floatstone
359-U1470B-18R-1	grainstone
359-U1470B-18R-CC	boundstone
359-U1470B-18R-CC	floatstone
359-U1470B-2R-CC	boundstone
359-U1470B-4R-CC	boundstone
359-U1470B-5R-1	boundstone
359-U1470B-5R-1	grainstone
359-U1470B-5R-2	boundstone
359-U1470B-5R-2	rudstone
359-U1470B-5R-3	boundstone
359-U1470B-5R-3	rudstone
359-U1470B-5R-CC	boundstone
359-U1470B-6R-1	boundstone
359-U1470B-7R-1	boundstone
359-U1470B-7R-1	floatstone
359-U1470B-8R-CC	floatstone
359-U1471A-12H-3	packstone
359-U1471A-13H-6	packstone

359-U1471A-14H-6	packstone
359-U1471A-15H-1	wackestone
359-U1471A-16H-2	wackestone
359-U1471A-17H-3	wackestone
359-U1471A-18H-1	wackestone
359-U1471A-19H-2	wackestone
359-U1471A-1H-1	packstone
359-U1471A-1H-2	packstone
359-U1471A-20H-2	wackestone
359-U1471A-23H-2	packstone
359-U1471A-28H-6	packstone
359-U1471A-29H-5	packstone
359-U1471A-2H-1	packstone
359-U1471A-30H-3	packstone
359-U1471A-31H-2	packstone
359-U1471A-54X-CC	wackestone
359-U1471A-74X-1	wackestone
359-U1471A-75X-2	wackestone
359-U1471A-78X-3	wackestone
359-U1471A-79X-2	wackestone
359-U1471C-6H-5	wackestone
359-U1471C-7H-1	wackestone
359-U1471C-8H-2	wackestone
359-U1471E-11R-3	wackestone
359-U1471E-16R-1	wackestone
359-U1471E-16R-2	wackestone
359-U1471E-17R-1	wackestone
359-U1471E-17R-2	wackestone
359-U1471E-17R-3	wackestone
359-U1471E-18R-2	wackestone
359-U1471E-19R-1	wackestone
359-U1471E-37R-1	wackestone
359-U1471E-38R-2	wackestone
359-U1471E-40R-2	wackestone
359-U1471E-41R-2	wackestone
359-U1471E-42R-2	wackestone
359-U1471E-9R-2	wackestone
359-U1472A-13H-3	wackestone
359-U1472A-15H-1	wackestone
359-U1472A-15H-3	wackestone
359-U1472A-15H-7	wackestone
359-U1472A-16H-4	wackestone
359-U1472A-18H-3	wackestone
359-U1472A-19H-2	wackestone
359-U1472A-4H-5	packstone

359-U1472A-5H-2	packstone
359-U1472A-6H-3	packstone
359-U1472A-7H-1	floatstone

Appendix B: Structure of DCGAN

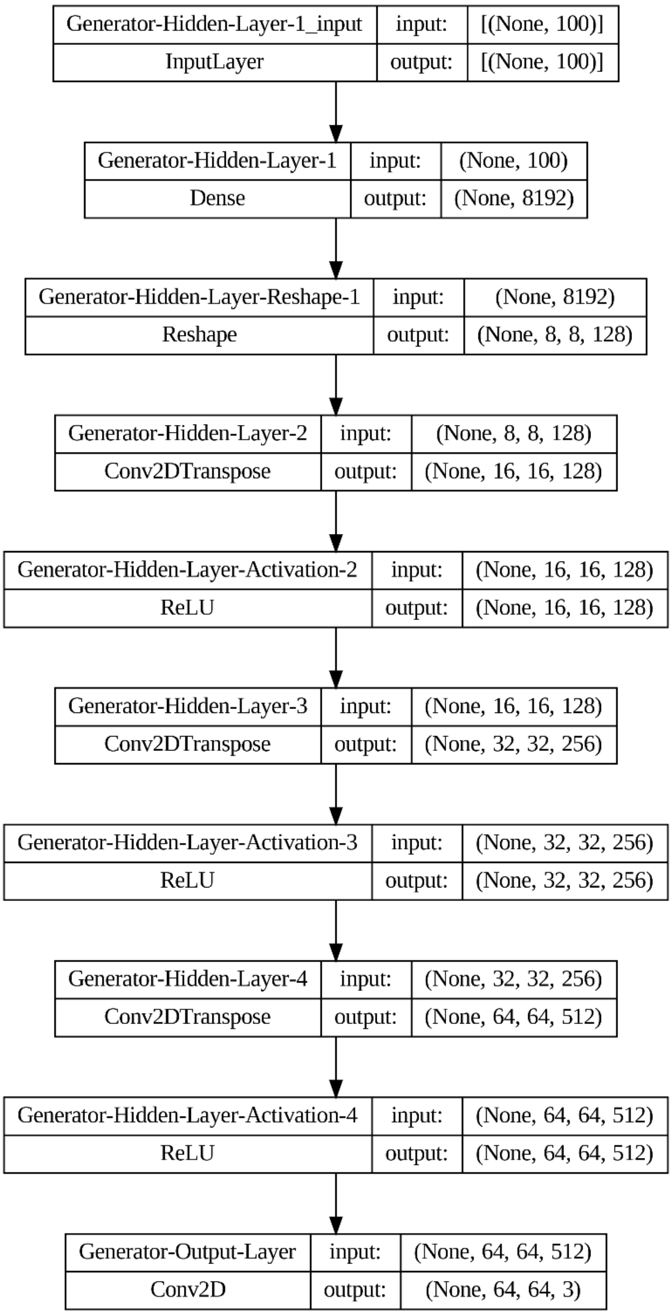

Figure B.1: Structure of the generator in the DCGAN model. The generator starts with a dense layer that transforms the 100-dimensional latent space into a feature map, followed by a reshaping layer to form an initial 8 × 8 × 128 tensor. Three subsequent transposed convolutional layers progressively upscale this tensor to 16 × 16, 32 × 32, and 64 × 64 resolutions, with ReLU activations ensuring non-linearity at each stage. The final output layer uses a tanh activation to produce a 64 × 64 × 3 RGB image with pixel values normalized between -1 and 1.

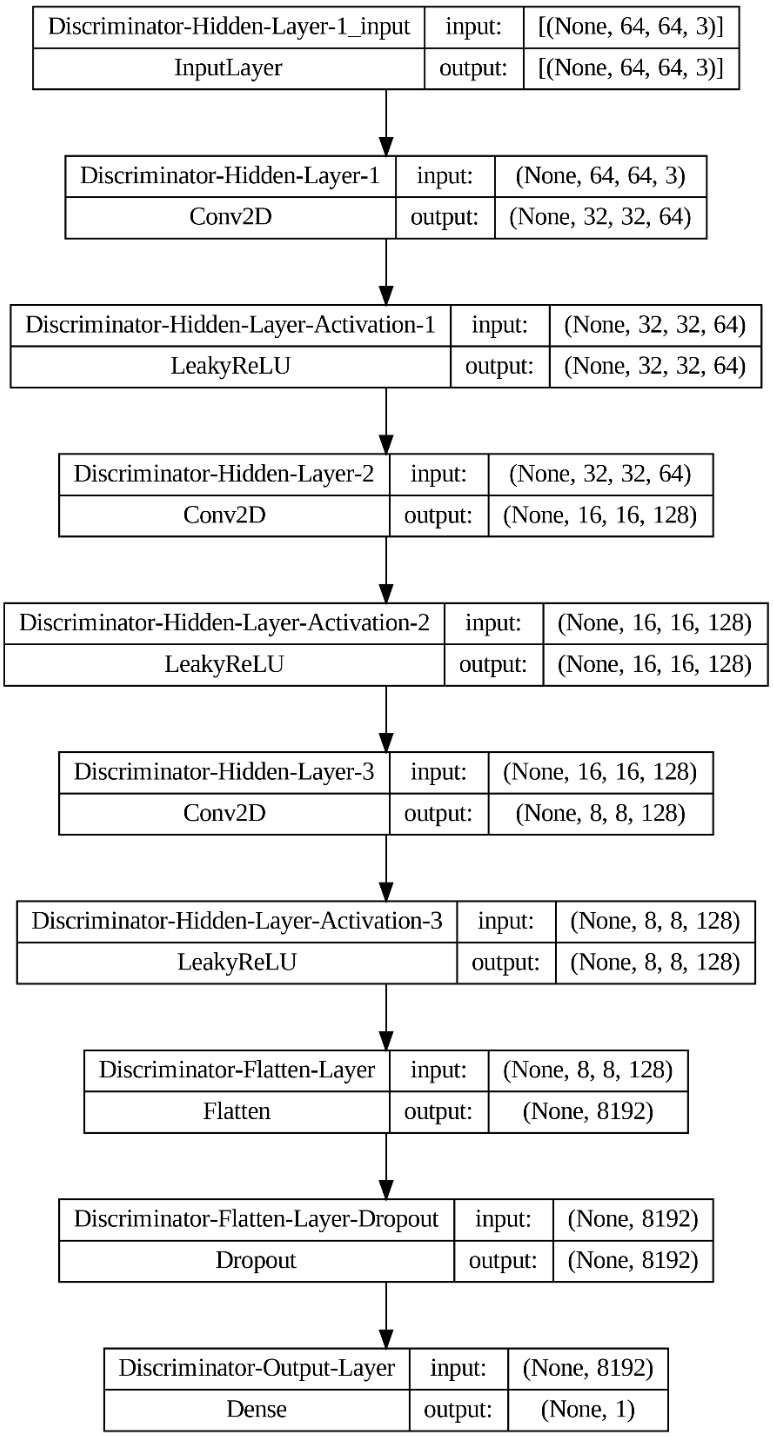

Figure B.2: Structure of the discriminator in the DCGAN model. The discriminator processes 64 × 64 × 3 RGB input images through three convolutional layers with strides of 2 × 2, reducing spatial dimensions while extracting features. Each convolutional layer is followed by a LeakyReLU activation function to allow small gradients for negative inputs, aiding in convergence. A flattening layer and dropout are applied to reduce overfitting, followed by a dense output layer with a sigmoid activation to classify images as real or synthetic.

Appendix C: Data tables for biogenic component frequency graphs

Table C.1: Percentage values of bioclasts from human interpretation for Hole 1196 (from Conesa *et al.*, 2005).

				Algae		Invertebrates						
Hole	**Core**	**Section**	**Interval**	**Red algae**	**Green algae (*Halimeda*)**	**Corals**	**Echinoderms**	**Bivalves**	**Gastropods**	**Serpulids**	**Bryozoans**	**Foraminifers**
1196A	**1R**	**1**	**35-40**	77.5	0.4	4.1	4.8		0.4		0.8	
1196B	**1R**	**3**	**1-6**	29.4	29.4	5.1	4.4	1.6	0.8	0.4	0.6	8.8
1196A	**2R**	**1**										
1196A	**3R**	**1**										
1196B	**4R**	**1**	**28-30**	33.3	14.3		5	0.7			5	7.7
1196A	**5R**	**1**	**70-72**	54.4	12.4		5.4					1.7
1196A	**5R**	**2**	**74-77**	52.4	18.4	1.7	1.4		0.7			1.4
1196A	**6R**	**1**										
1196A	**7R**	**1**	**106-109**	67.4	1.4		10.8			0.8	0.4	1.8
1196A	**8R**	**1**										
1196A	**9R**	**1**										
1196A	**10R**	**1**	**8-12**	82	1		1.7					
1196A	**11R**	**1**	**24-28**	73.3	13		3				0.7	
1196A	**12R**	**1**										
1196A	**13R**	**2**	**38-41**	82.3	2.7		7.3					
1196B	**14Z**	**1**	**68-73**	81.1	3.1		7.4	0.8			0.4	
1196A	**14R**	**1**	**42-46**	93			7					
1196B	**17Z**	**1**	**54-57**	78.5			19.5					
1196B	**20Z**	**1**	**16-20**	82.7	7.3		6				1	
1196A	**15R**	**1**	**0-3**	17.8	24.1	30.8	1.3	1.3	1.1	1.3	0.8	4.3
1196A	**16R**	**1**	**25-27**	68.4	8.4		4.1	0.4	1	0.4	1.7	1.4
1196A	**17R**	**1**										
1196A	**18R**	**2**	**15-18**	70.1	16.4		4.1		1.4			
1196A	**18R**	**3**	**75-78**	86			14					
1196A	**19R**	**1**										
1196A	**20R**	**1**	**45-47**				16.4	7.1	7.1		3.1	13.7
1196B	**33Z**	**1**	**3-5**	56			24				3	1
1196A	**21R**	**1**	**119-121**	26	3	4.7	16	5.3	1.7	2	0.3	2.3
1196A	**22R**	**1**										
1196A	**23R**	**1**										
1196A	**24R**	**1**	**13-17**	13.4			24.1	5.4	1.8	2.4	13.8	4.1
1196A	**25R**	**1**	**50-52**	30.3			33	1.3		0.3	7.4	1
1196B	**46Z**	**1**	**14-17**	28			24	5.7	0.4		13.4	2.7
1196A	**27R**	**1**	**7-10**	20	0.7		16	7.3		1	6.3	1.7
1196A	**28R**	**1**										
1196A	**29R**	**1**										
1196A	**30R**	**1**										
1196A	**31R**	**1**										
1196A	**32R**	**1**	**11-14**	28.3			23.3	2.3	1.3		16.3	1.3
1196A	**33R**	**1**	**48-53**	26.3	0.7	40.3	5.7	5.7	1	0.3	2.3	0.7
1196A	**33R**	**2**	**57-60**	20.3	3		9.3	7.7	9.3	3.7	0.7	2.3
1196A	**34R**	**1**	**96-99**				26.8	9.5	1.4	4.1	4.1	9.4
1196A	**35R**	**1**	**1-5**	29.7	16		16	3	5.4	0.7	1	2.7

Table C.2: Percentage values of bioclasts from Faster R-CNN for Hole 1196.

			Algae		Invertebrates						
Hole	**Core**	**Section**	**Red algae**	**Green algae (*Halimeda*)**	**Corals**	**Echinoderms**	**Bivalves**	**Gastropods**	**Serpulids**	**Bryozoans**	**Foraminifers**
1196A	**1R**	**1**	80.2	0	5.1	5	2	2.5	0	2	3
1196B	**1R**	**3**									
1196A	**2R**	**1**	76	0	1	3	0.5	0	0	0	2
1196A	**3R**	**1**	69	0	0	2.7	0	0	0	0	2.5
1196B	**4R**	**1**	70	0	0	2.4	0	0	0	0	2
1196A	**5R**	**1**	62.5	0	0	2.2	4.9	2.5	0	0	1.8
1196A	**5R**	**2**	50	0	9	2	3	2.3	0.4	0	7.5
1196A	**6R**	**1**	51	0	4	5	2	2	0.8	0	11
1196A	**7R**	**1**	72.5	0	5	1.8	3	1.4	0.3	0	9.7
1196A	**8R**	**1**	76	0	0	10	4	1	0.6	0	7.6
1196B	**8R**	**1**	80.5	0	0	9	2	0.6	0	0	7.8
1196A	**9R**	**1**	81	0	0	4.8	2.5	0	0	0	8.2
1196A	**10R**	**1**	82	0	0	4.7	0	0	0	2.2	4.8
1196A	**11R**	**1**	81	0	0	3.1	0	0	0	2.4	2
1196A	**12R**	**1**	76	0	0	2.9	0	0	0	1	1
1196B	**13Z**	**1**	76	0	0	4.9	0	0	0	2	2
1196A	**13R**	**2**	94	0	0	3.5	0	0	0	0	2.2
1196B	**14Z**	**1**									
1196A	**14R**	**1**	90.8	0	1	2.4	1.8	2	0.4	0.8	0.8
1196B	**17Z**	**1**									
1196B	**20Z**	**1**									
1196A	**15R**	**1**	18	0	38	3	4	2	2	2	4
1196A	**16R**	**1**	19	0	30	2.7	2.5	1	0.4	2.4	5
1196B	**24Z**	**1**	50	0	22	4.2	3	6	0.4	2.5	4.8
1196A	**17R**	**1**	52	0	15	4	2	5.8	0.2	2.4	3
1196A	**18R**	**2**	77.4	0	9.8	2.5	2	2.5	0	2	2.2
1196A	**18R**	**3**	82.5	0	0	8	0	1.5	0	0	0
1196A	**19R**	**1**									
1196A	**20R**	**1**	5	0	5	9	5.5	7.1	2	6	10
1196B	**33Z**	**1**	8	0	6.5	11	7	7	2.5	4.8	9.9
1196A	**21R**	**1**	10	0	9	9	6	5.1	2	5.1	10.4
1196B	**36Z**	**1**	12	0	0	8.8	6	4.6	2.4	0	13
1196A	**22R**	**1**	14	0	0	9	7.5	2.5	2.5	0	15.2
1196A	**23R**	**1**	13.5	0	0	12	7	4	1.8	2	19
1196B	**41Z**	**1**	12	0	0	12	8.5	3.5	1.6	1	18
1196A	**24R**	**1**	13.5	0	0	12.2	6.5	2.6	1.8	7	16.5
1196A	**25R**	**1**	14.5	0	0	9.9	4.8	1.8	2.5	10.2	11.5
1196B	**46Z**	**1**	16	0	0	11	4	0	0.6	9.8	12
1196A	**26R**	**1**	14	0	0	12	3	0	0	10	11.5
1196B	**47R**	**1**	17	0	0	10.5	6	0	0.6	10.4	10.2
1196B	**48R**	**1**	14	0	0	9.5	7.5	0	2.3	10	9
1196A	**27R**	**1**	13.9	0	0	8	9	0	2.7	9.8	8
1196A	**28R**	**1**									
1196A	**29R**	**1**									
1196A	**30R**	**1**									
1196A	**31R**	**1**									
1196A	**32R**	**1**	18	0	25	14	7.9	4.8	0	12.4	17.5
1196A	**33R**	**1**	22	0	28	11.4	8	2.6	1	8	19
1196A	**33R**	**2**	21	0	0	14.6	7.5	5.2	1	9	21
1196A	**34R**	**1**	2	0	5.1	17	12	5.5	2.5	8	20.5
1196A	**35R**	**1**	29	0	7	14.4	8	4.8	0	4.9	19

Table C.3: Percentage values of bioclasts from human interpretation for Hole 1199 (from Conesa *et al.*, 2005).

Hole	Core	Section	Interval	Algae		Invertebrates						Foraminifers
				Red algae	Green algae (*Halimeda*)	Corals	Echinoderms	Bivalves	Gastropods	Serpulids	Bryozoans	
1199A	1R	1	3-6	93.5			5.4	0.4				
1199A	1R	1	32-34	99.3			0.7					
1199A	2R	3	20-23	54.8	30.8		4.8					
1199A	3R	1	77-81	79.4	7.7	3	1.3					
1199A	4R	6	138-40	64.4	31.7			0.7	0.7			0.4
1199A	5R	2	127-131	59.4	22.1	5.1	1.1	0.4	2.1		1.1	1.4
1199A	6R	3	69-72	63.3	11.7		4					
1199A	7R	1										
1199A	8R	1										
1199A	9R	4	12-16	76			20					
1199A	10R	3	58-60	40.7	41	0.7	2.7		0.3			
1199A	11R	1	16-20	87.9			7.8	0.5				
1199A	12R	5	93-95	92			8					
1199A	13R	3	115-119	66	4	15	9					
1199A	14R	1	137-142	93.7			6.3					
1199A	15R	1										
1199A	16R	3	40-44	69	6		12	1			1	
1199A	16R	4	91-96	78	9		10			3		
1199A	17R	1	23-26	80	2		18					
1199A	18R	3	21-26	86			14					
1199A	19R	1	0-4	13.6			20.3	9.6		4.3	12.3	11
1199A	20R	1	17-22	20.1	3.1	2.7	18.1	4.1		1.7	1.7	0.4
1199A	21R	1										
1199A	22R	1										
1199A	23R	1										
1199A	24R	1										
1199A	25R	1										
1199A	26R	1										
1199A	27R	1										
1199A	28R	1										
1199A	29R	1										
1199A	30R	1										
1199A	31R	1	0-3	65	7		11.7	1.7	3		0.3	2
1199A	32R	1										
1199A	33R	1	11-13	47.6	21.3		20.3				2.7	0.3
1199A	34R	1	28-31	71.3	0.3		14.4				1.3	
1199A	35R	1	13-17	72			28					
1199A	36R	1										
1199A	37R	1	2-6	51	15.7		5.3	4.3			4	1
1199A	38R	1										
1199A	39R	1	56-60	81			15					
1199A	39R	1	91-94	70	0.3		6.7				1	
1199A	40R	1	9-13	60.4	0.7		7			4		0.7
1199A	40R	1	27-31	46.3	38.3		5.3	0.7	0.3		1	1.7
1199A	41R	1	19-22	35.4	6.7		13.7	2			8	3
1199A	42R	1	83-85	55.3	1.3		7.7	0.7			4.7	
1199A	43R	1										
1199A	44R	1										
1199A	45R	1										

Table C.4: Percentage values of bioclasts from Faster R-CNN for Hole 1199.

			Algae		Invertebrates						
Hole	**Core**	**Section**	**Red algae**	**Green algae (*Halimeda*)**	**Corals**	**Echinoderms**	**Bivalves**	**Gastropods**	**Serpulids**	**Bryozoans**	**Foraminifers**
1199A	**1R**	**1**	89	0	3.5	0	0	2.5	0	0	2.8
1199A	**1R**	**2**	81	0	3	2	0.4	2	0	0	2.5
1199A	**2R**	**3**	71	0	0	2.1	0.5	0.8	0	0	2.3
1199A	**3R**	**1**	69	0	2.5	2.4	1	0	0	0	2.5
1199A	**4R**	**6**	64	0	3	2.5	1.6	0.4	0	0	3
1199A	**5R**	**2**	64.5	0	4	2.6	1.2	3	0	1.5	4.8
1199A	**6R**	**3**	67	0	4.5	3	0.8	2.6	0	4	3.8
1199A	**7R**	**1**	69.6		0	4.8	1.2	0.2	0	0	3.6
1199A	**7R**	**2**	72		0	5.1	2	0	0	0	3
1199A	**8R**	**1**	71		0	7	2.5	0	0	0	2.8
1199A	**9R**	**4**	68	0	4.5	2	2.2	0.4	0	0	2.7
1199A	**10R**	**3**	72	0	2	3.5	3	0.8	0	0	2.7
1199A	**11R**	**1**	70.5	0	0	4.4	2.8	1.4	0	1.4	1.8
1199A	**12R**	**5**	79.8	0	7	2.6	2.4	2	0	0	4.8
1199A	**13R**	**3**	85	0	10	2.7	2.3	0	0	0	0
1199A	**14R**	**1**	96	0	0	3.8	0	0	0	0	0
1199A	**15R**	**1**									
1199A	**16R**	**3**	93	0	0	3	0	0	0	0	0
1199A	**16R**	**4**	89	0	0	2.4	3	0	0	0	4.7
1199A	**17R**	**1**	84	0	0	2.5	2.8	0	0	0	3
1199A	**18R**	**1**	88		0	7.5	3.2	0	0	0	1
1199A	**18R**	**3**	77	0	0	9.8	0	0	0	0	4
1199A	**19R**	**1**	25	0	4.5	10	14.1	0	1.8	4.9	15.1
1199A	**20R**	**1**	27	0	8	9.8	10	0	1.8	6	14.9
1199A	**21R**	**1**									
1199A	**22R**	**1**									
1199A	**23R**	**1**									
1199A	**24R**	**1**									
1199A	**25R**	**1**									
1199A	**26R**	**1**									
1199A	**27R**	**1**									
1199A	**28R**	**1**									
1199A	**29R**	**1**									
1199A	**30R**	**1**									
1199A	**31R**	**1**	84.3	0	0	1.9	4.9	4	0	0	4.9
1199A	**32R**	**1**	87.5	0	0	1.8	1.2	3.8	0	2.6	3
1199A	**33R**	**1**	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
1199A	**34R**	**1**	91	0	0	2.2	2.5	0	0	2.3	2
1199A	**35R**	**1**	89	0	0	4.7	2	0	0	2.5	1.8
1199A	**36R**	**1**	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
1199A	**37R**	**1**	82	0	0	4.8	8	0	0	2.4	2.6
1199A	**38R**	**1**	75	0	3.5	4	7.8	0	0	3.8	4.8
1199A	**39R**	**1**	76	0	2.5	2.3	4.5	2	0	5	3
1199A	**40R**	**1**	77	0	0	2.4	4.8	0	0	4.6	4.8
1199A	**41R**	**1**	75.8	0	0	4.7	7.6	2.4	0	4.8	4.7
1199A	**42R**	**1**	80.5	0	0	2	4.9	0	0	7	4.8
1199A	**43R**	**1**									
1199A	**44R**	**1**									
1199A	**45R**	**1**									