Materials & Design 272 (2026) 117298

Contents lists available at ScienceDirect

Materials & Design

journal homepage: www.elsevier.com/locate/matdes

StampFormer: A physics-guided material-geometry-coupled multimodal model for rapid prediction of physical fields in sheet metal stamping

Jiajie Luo[a,b], Mohamed Mohamed[a,f], Osama Hassan[a,c], Haosu Zhou[a], Yingxue Zhao[a], Haoran Li[a], Xinrun Li[b], Zhutao Shao[d], Yang Long[e], Jichun Li[b,*], Nan Li[a,*]

[a] Dyson School of Design Engineering, Imperial College London, London, SW7 2DB, UK
[b] School of Computing, Newcastle University, Newcastle upon Tyne, NE4 5TG, UK
[c] Department of Computing, Imperial College London, London, SW7 2DB, UK
[d] Multi-X Solution Limited, London, E6 2JA, UK
[e] Department of Computer Science, Durham University, Durham, DH1 3LE, UK
[f] Department of Mechanical Engineering, Faculty of Engineering, Helwan University, Helwan, 4034572, Egypt

HIGHLIGHTS

- StampFormer fuses material stress-strain curves with component geometry.
- Multiscale material injection improves forming–field prediction.
- The model predicts thinning, strain and displacement in under 1 s.
- Predicted fields agree closely with finite-element simulations.
- The surrogate supports rapid early–stage material–geometry screening.

ARTICLE INFO

Keywords:
Sheet metal forming
AI for science
Surrogate model
Real-time simulation
Multimodal learning

ABSTRACT

Traditional sheet metal forming relies on computationally expensive Finite Element Analysis (FEA), significantly prolonging design cycles. While surrogate models accelerate iteration, existing approaches either provide limited scalar outputs or ignore crucial material properties by focusing solely on geometry. To address this gap, we develop StampFormer, a physics-guided deep learning framework that simultaneously processes component geometry and material stress-strain responses to predict FEA outcomes. Its architecture employs a Material-Augmented Geometric Network (MAGN) for initial multimodal fusion, followed by a Hierarchical Material Embedding Injection Unit (HMEIU) that progressively integrates material features into an adapted Swin-UNet backbone. Evaluated on steel and aluminium crossmember stamping datasets, StampFormer predicts critical physical fields—including thinning, major/minor strain, plastic strain, and displacement—in under a second. Compared with the reference FEA, our model achieves a mean representative-maximum relative error below 8.2% for each of the four 2D fields in both material datasets and a masked displacement mean squared error below 0.71 mm^2. Within the investigated component family and fixed process conditions, this multimodal framework provides sub-second predictions for rapid early-stage material–geometry screening and design exploration.

1. Introduction

Sheet metal forming is crucial for producing complex parts in various sectors. This includes aerospace [1], medical devices [2], and, notably, automotive manufacturing [3]. As technology advances and performance demands increase, manufacturers have increasingly adopted advanced steels and aluminium alloys with high strength-to-weight ratios to produce high-performance and more fuel-efficient vehicles in recent years [4,5]. Despite the significant advantages these materials offer for vehicle lightweighting, it remains a challenge to design feasible and optimal parts with them [6]. This is because their forming processes often exhibit complex behaviours, such as thinning, limited ductility, and increased springback, which pose substantial challenges to

* Corresponding authors.
Email addresses: jichun.li@ncl.ac.uk (J. Li), n.li09@imperial.ac.uk (N. Li).

https://doi.org/10.1016/j.matdes.2026.117298
Received 23 June 2026; Received in revised form 29 September 2026; Accepted 17 October 2026
Available online 21 October 2026
0264-1275/© 2026 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

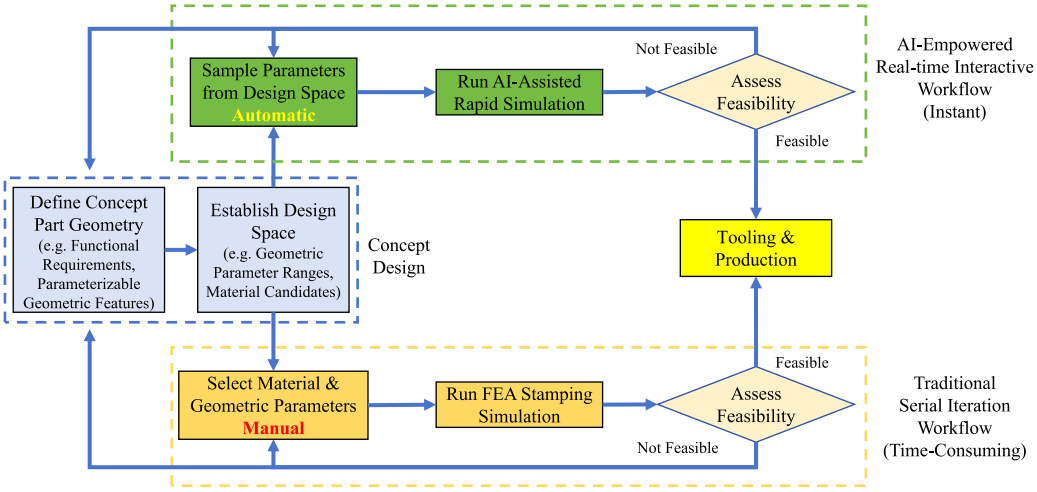


Fig. 1. Workflow of traditional sheet metal forming design and the proposed surrogate-assisted AI-empowered design process.

the design of manufacturable components [7]. To handle this complexity, it is essential to implement a rigorous and efficient design validation process. This step is crucial for ensuring that a part is manufacturable before investing in costly tooling and production lines [8].

As illustrated in the *Traditional Serial Iteration Workflow* section of Fig. 1, the conventional sheet-metal design process is constrained by a linear and repeated validation cycle [9]. In this workflow, designers first create a concept part and define its geometry, and engineers then select materials and geometric parameters manually based on their experience to run a feasibility check. Such feasibility analysis relies on complex Finite Element Analysis (FEA) simulations that require significant computing resources. FEA provides detailed predictions of how materials will flow, along with the potential failure regions [10]. This makes it an essential instrument for confirming a design before final production [7]. However, its steep computational cost prevents it from being an efficient tool for fast sheet metal design. In this cycle, each iteration introduces a substantial latency of hours or even days, severing the interactive link between design ideation and performance verification [11,12]. This inevitably lengthens the design cycle and pushes up the cost of development [3,13].

Advances in computing hardware have enabled Artificial Intelligence (AI) to fundamentally transform traditional engineering practices and boost productivity [14]. Consequently, AI applications in manufacturing have garnered significant academic attention [15]. Typical applications include predictive maintenance [16], supply chain optimisation [17], and smart quality inspection [18]—such as Kardovskyi et al.'s use of mask R-CNN for steel bar installation [19] and Senoner et al.'s explainable AI for semiconductor manufacturing [20]. Furthermore, AI-assisted design optimisation using surrogate models has recently emerged [3,12,21–25]. As shown in Fig. 1, this AI-empowered workflow rapidly simulates designs to accelerate the validation feedback loop, reducing the need for repeated time-consuming FEA. However, existing geometry-focused surrogates [22,26] often overlook the critical influence of material properties. In practice, even subtle material variations can significantly alter forming responses [27]. Addressing this material–geometry coupling gap is the focus of our work.

In this paper, we introduce StampFormer, a novel physics-guided deep learning framework designed to account for the critical impact of material properties in predicting physical fields (i.e., thinning, strain, and displacement) during sheet metal forming. In any real-world forming process, a part's final performance is co-determined by its geometry and the material used. Our method introduces a multimodal architecture that is, to the best of our knowledge, one of the first high-fidelity sheet-forming surrogates to explicitly combine a part's geometry and its material properties (i.e., stress-strain responses) to predict FEA outputs. Unlike geometry-centred image-based surrogates, StampFormer explicitly couples the material stress–strain response with the geometric representation and maintains this material information across multiple feature scales. This enables the model to account explicitly for material variations together with geometric variations within the investigated design domain. The conceptual distinction from existing surrogate approaches is further discussed in Section 2. The main contributions of this work are as follows:

1. By providing rapid feasibility predictions within the investigated design domain, the surrogate can reduce repetitive FEA evaluations and support faster early-stage material and geometric design exploration.
2. Our multimodal, image-based approach addresses the limitations of both scalar-based techniques, which provide limited spatial information, and conventional geometry-centred image-based surrogates, which do not explicitly incorporate the material stress–strain property as a variable.
3. The introduced architecture employs a dual-stage integration strategy. First, the Material-Augmented Geometric Network (MAGN) combines geometric inputs with material-specific context; subsequently, the Hierarchical Material Embedding Injection Unit (HMEIU) injects material embeddings at multiple hierarchical levels. Ablation experiments assess the contribution of material conditioning and the output-dependent effects of early fusion.

The rest of the paper is organised as follows. Section 2 reviews related work. Section 3 details the problem definition and how the data was prepared. Section 4 presents the proposed network architecture, while Section 5 details the experimental setup. Section 6 presents and discusses the results. Section 7 discusses engineering implications and metric choices, and Section 8 concludes the paper.

2. Related work

To address the workflow latency between the design and validation phases of traditional design cycles, the engineering field has increasingly turned to data-driven surrogate models. As depicted in the section *AI-Empowered Real-time Interactive Workflow* of Fig. 1, these models are trained to learn the complex input–output relationships of physical simulations in near real-time, thereby enabling rapid design iterations [28–30]. A relevant example can be found in structural mechanics,

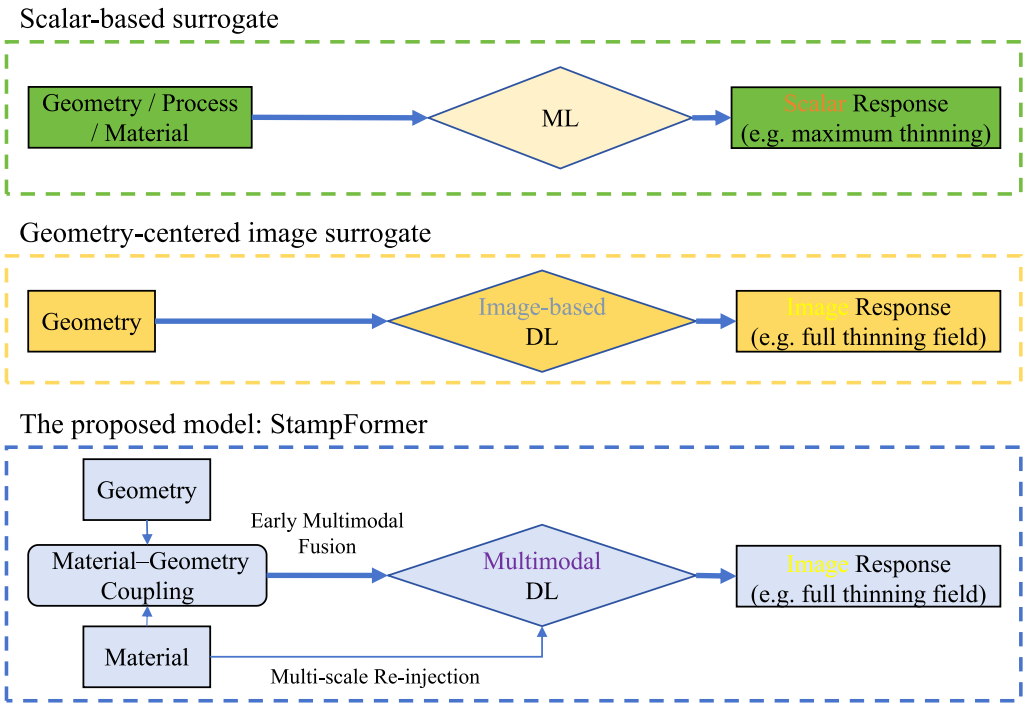


Fig. 2. Comparison of surrogate paradigms: scalar-based models predict discrete metrics from design variables, and geometry-centred surrogates infer spatial fields from geometry alone. StampFormer couples the material stress–strain response with geometry via early fusion and multi-scale re-injection to achieve full-field physical prediction.

where surrogate models are developed to accelerate FEA. By replacing complex, time-consuming simulations with a machine learning model, researchers can rapidly predict the dynamic response, such as the displacement of a truss structure, thereby preventing the need to rerun complex simulations [31]. Early data-driven surrogate models, like those in [32–34], were relatively simple. These approaches, which we refer to as scalar-based models, worked by regressing a few key design parameters against a single performance metric to simplify complex simulation processes.

While these scalar-based models—which predict a single, critical metric such as the maximum thinning rate—offer significant speed, they suffer from a major drawback: they provide only limited output information compared to image-based surrogate models [35–37]. For example, in the hot stamping of a vehicle component, key variables like geometry and processing conditions influence the final thinning field. Predicting the entire thinning field with image-based machine learning methods, rather than just predicting its maximum value, provides much more comprehensive information for engineers [3]. Similarly, Li et al. [11] used an image-based machine-learning method to construct a surrogate model for rapid prediction of crashworthiness performance in vehicle panel components. This approach provided engineers with richer information, such as precise locations, directly from the final predicted image-based outputs.

While existing image-based surrogate models can provide richer outputs than scalar-based surrogate models, they still have a major limitation: current architectures are primarily designed around geometric inputs and do not explicitly incorporate the full material stress–strain response as an additional input modality. As conceptually summarised in Fig. 2, scalar-based surrogate models typically map geometric, process, and material variables to individual scalar responses, while geometry-centred image-based models map geometric representations to spatial physical fields. In contrast, StampFormer introduces the material stress–strain response together with the component geometry, allowing the predicted physical field to depend jointly on material behaviour and geometric information. This distinction is particularly important when material variation itself is a design variable in sheet-metal forming.

Addressing the material–geometry coupling illustrated in Fig. 2 requires a suitable architectural foundation. The encoder–decoder structure, for example, is widely used for semantic segmentation, where the goal is to produce a dense, pixel-wise probability map, analogous to the physical fields in FEA simulations. Among the existing image-based FEA surrogate models, many have also been designed with an encoder–decoder structure [3,11,12,22,38]. As a pioneering encoder–decoder model, U-Net is widely employed in the field of medical image segmentation [39]. Over the past decade, encoder–decoder models for semantic segmentation have advanced considerably, undergoing multiple rounds of iteration [40] driven by the development of architectures such as Vision Transformer [41], Vision Mamba [42], and Swin Transformer [43]. A number of superior architectures, including TransUNet [44], UMamba [45], and SwinUNet [46], have since been introduced. Despite these architectural advancements, their application in manufacturing has yet to address the critical gap identified earlier: the need to integrate material properties with geometry. Fortunately, the recent emergence of multimodal learning—a subfield of AI focused on processing and fusing information from fundamentally different data types (such as 1D curves and 2D images)—provides the necessary techniques to close this gap. Building on this concept, StampFormer introduces early material–geometry coupling followed by multi-scale material re-injection, while retaining an encoder–decoder backbone for spatial full-field prediction. The detailed implementation of these mechanisms is presented in Section 4.

3. Data generation and pre-processing

StampFormer learns the relationship between component geometry, material constitutive response, and the resulting forming fields. The dataset therefore comprises three linked components: parameterised component geometries, material stress–strain curves, and the corresponding physical fields generated by finite-element analysis (FEA).

For each material–geometry combination included in the Design of Experiments (DoE), a separate PAM-STAMP simulation was performed to generate the target thinning, strain, and displacement fields. This section describes the generation of the material and geometry data and their pre-processing into the input and output formats required by StampFormer.

Throughout this manuscript, unless otherwise stated, the term "stress–strain curve" refers to the flow-stress–plastic-strain relation used in the FEA material model.

3.1. Material data generation

The reliability of StampFormer depends on adequate sampling of the constitutive material-response domain represented in the training data. In the present FEA formulation, material behaviour is prescribed using a lookup table of flow stress as a function of plastic strain, and StampFormer uses a discretised representation of the same response as its one-dimensional material input. The available baseline curves were therefore supplemented through controlled, bounded augmentation to improve the balance and sampling density of this constitutive-response domain. The purpose was not to create additional named commercial grades, but to provide a denser representation of the material responses relevant to the numerical forming simulations.

For a parent flow curve $\sigma_{f,i}(\varepsilon^p)$, a multiplicatively scaled curve can be expressed as

$$\sigma^{\text{aug}}_{f,i,k}(\varepsilon^p) = \alpha_{i,k}\sigma_{f,i}(\varepsilon^p), \tag{1}$$

where σ_f denotes flow stress, ε^p denotes plastic strain, and $\alpha_{i,k}$ is a positive scaling factor. This operation retains the plastic-strain coordinates and the characteristic shape of the parent hardening curve while varying its overall flow-stress level. Diversity in hardening behaviour is provided by the different parent curves, while the augmentation increases the sampling density around the response trajectories represented by these curves.

The initial steel dataset comprised 101 stress–strain curve records obtained from an existing material database. These records had been assigned to five dataset-specific material-response clusters containing 55, 10, 13, 3, and 20 curves, respectively, as shown in Fig. 3(a). Because the initial distribution was strongly imbalanced, a two-stage augmentation procedure was applied. In the first stage, 0, 28, 27, 33, and 11 additional curves were generated for the five clusters, respectively. This produced an intermediate dataset of 200 curves, with cluster sizes of 55, 38, 40, 36, and 31 (Fig. 3b), thereby reducing, although not eliminating, the original imbalance.

In the second stage, a further 65, 82, 80, 84, and 89 curves were generated for the respective clusters so that each cluster contained 120 curves. The final steel dataset therefore comprised 600 curves (Fig. 3(c)). The additional steel curves were generated within their respective material-response clusters by multiplying the flow-stress values of selected parent curves by positive scaling factors. In the final balancing stage, the scaling factors were selected within the interval [0.95, 1.10]. This bounded augmentation increased the coverage of initially under-represented regions while retaining the principal response characteristics of the source curves.

The aluminium dataset comprised 10 baseline stress–strain curve records, each for a different aluminium sheet material used in cold stamping, as shown in Fig. 4(a). Each baseline curve was scaled using 11 factors defined as

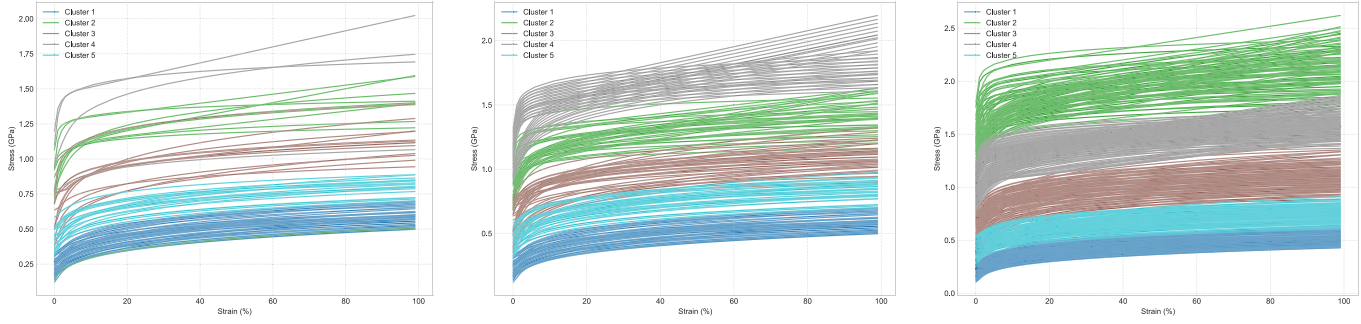
$$\alpha_k = 0.90 + 0.02k, \qquad k = 0, \ldots, 10, \tag{2}$$

corresponding to scaling factors from 0.90 to 1.10 at intervals of 0.02. Applying these 11 factors to each of the 10 baseline curves generated $10 \times 11 = 110$ aluminium curves in total (Fig. 4(b)). Because the factor $\alpha_k = 1.00$ was included, the final dataset contained the 10 original curves together with 100 scaled variants. The bounded $\pm 10\%$ range increased the local sampling density around the original material responses without introducing large extrapolations from their baseline constitutive behaviour.

The five material-response clusters were used as dataset-specific sampling strata to reduce imbalance and support distributed material selection in the DoE. Their labels were not supplied to StampFormer and are not intended to represent a common physical classification across the steel and aluminium datasets.

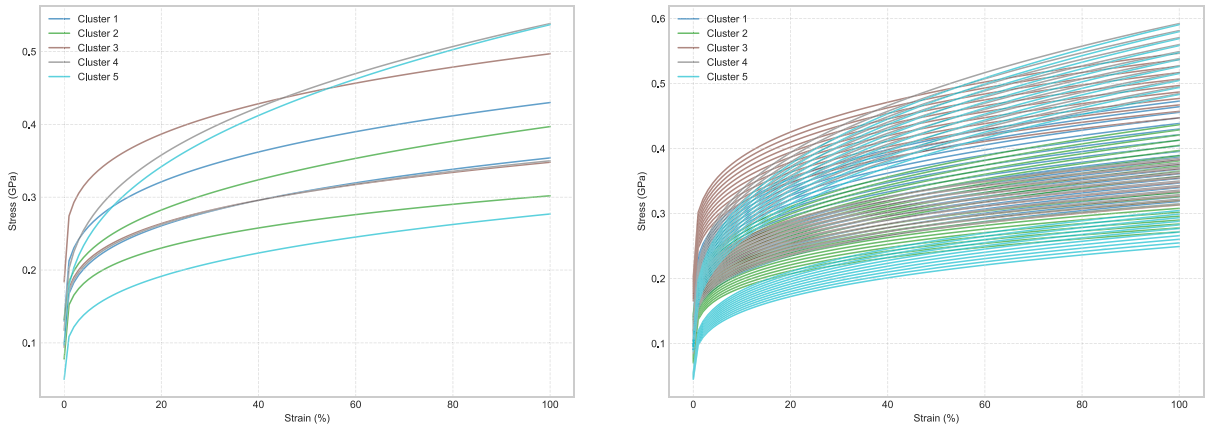
The augmented curves served as material-card inputs in the DoE, and the forming fields for each material–geometry pairing were generated through a separate PAM-STAMP simulation rather than being copied from an existing case. Under the fixed material-model and process assumptions adopted in this study, different grades, tempers, suppliers, or production batches may exhibit similar flow-stress–plastic-strain responses and consequently similar simulated forming behaviour. Conversely, nominally identical grades may exhibit variability in their measured material responses. Because StampFormer receives the numerical constitutive curve rather than a commercial material label, the relevant material input in the present framework is the flow-stress–plastic-strain response itself.

The augmented curves should therefore be interpreted as virtual constitutive samples that densify the response space represented by the baseline datasets, rather than as additional independent experimental measurements. This expanded response space enables StampFormer to learn how variations in flow-stress levels and the range of hardening



(a) Initial dataset (101 curves). (b) Intermediate dataset (200 curves). (c) Final balanced dataset (600 curves).

Fig. 3. Evolution of the steel stress–strain dataset: (a) the initial dataset of 101 curves, with cluster sizes of 55, 10, 13, 3, and 20; (b) the intermediate augmented dataset of 200 curves, with cluster sizes of 55, 38, 40, 36, and 31; and (c) the final balanced dataset of 600 curves, comprising 120 curves in each of the five material-response clusters. Throughout this manuscript, stress–strain curves refer to flow-stress–plastic-strain relations.



(a) Measured baseline dataset (10 curves). (b) Final augmented dataset (110 curves).

Fig. 4. Aluminium stress–strain datasets: (a) the 10 experimentally measured baseline curves, and (b) the final dataset of 110 curves obtained by applying 11 scaling factors from 0.90 to 1.10 at intervals of 0.02 to each baseline curve. The final dataset contains the 10 original curves and 100 scaled variants. Colours denote the dataset-specific material-response clusters used for distributed sampling in the DoE. (For interpretation of the references to colour in this Figure legend, the reader is referred to the web version of this article.)

behaviours represented in the dataset influence the predicted thinning, strain, and displacement fields.

To provide a consistent input format for StampFormer, all baseline and augmented stress–strain curves were interpolated and discretised onto a common grid of $T = 100$ uniformly distributed plastic-strain points. The corresponding flow-stress values form the material-input vector

$$\mathbf{s}_i = \left[\sigma_{f,i}\left(\varepsilon_1^p\right), \sigma_{f,i}\left(\varepsilon_2^p\right), \ldots, \sigma_{f,i}\left(\varepsilon_T^p\right)\right]^{\mathrm{T}} \in \mathbb{R}^{T\times 1}, \qquad T = 100. \tag{3}$$

The resulting sequence of $T = 100$ flow-stress values is used directly as the one-dimensional material input to the model. This interpolation standardises the input dimension and is distinct from the preceding material-response augmentation.

3.2. Geometric parameter generation

To create a comprehensive and representative dataset, the geometry was designed to capture the key deformation characteristics encountered in sheet-metal forming, while maintaining a sufficiently systematic parameterisation for studying the influence of geometric variations. The selected panel geometry incorporates representative automotive stamping features, including curved transitions, fillet radii, and a draft angle, which collectively introduce bending, stretching, and material flow around the forming radii. This geometry was therefore considered appropriate for evaluating the capability of the proposed method to predict the forming response over a range of practical geometric configurations.

As the selected panel geometry and the corresponding tooling/loading conditions are symmetric about the central plane, only half of the geometry was modelled in the numerical simulations. The symmetry boundary condition was applied along the symmetry plane, ensuring that the half-model reproduces the deformation behaviour of the complete panel while significantly reducing the computational cost and simulation time. This approach also enabled a larger number of simulations to be performed for systematic exploration of the geometric design space without compromising the physical representation of the forming process.

As illustrated in Fig. 5, the panel geometry includes several key geometric features, including the die radius R_1, punch radius R_2, and

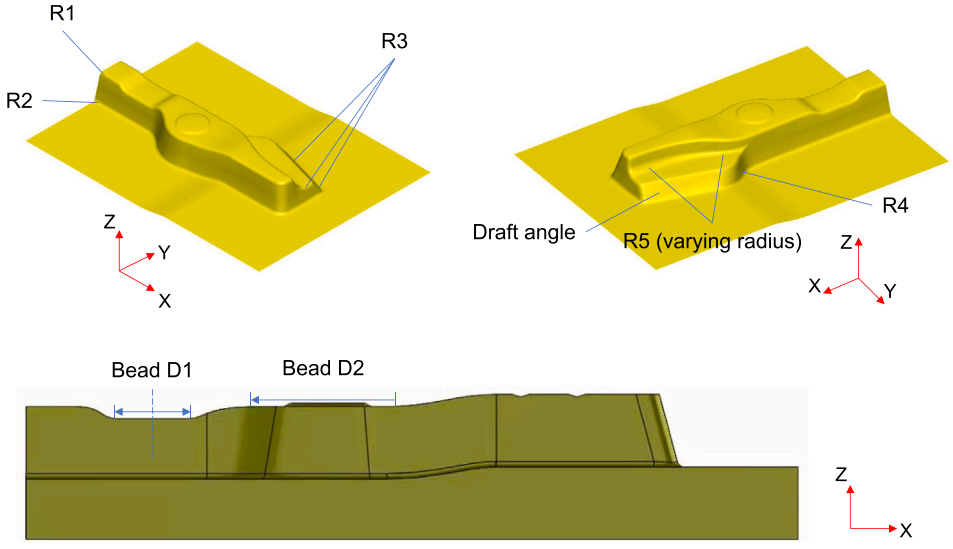


Fig. 5. The key geometric features of the proposed panel.

Table 1
Latin Hypercube sampling ranges for geometric parameters.

Parameter	Sampling Range
R_1 (mm)	5–10
R_2 (mm)	5–10
R_3 (mm)	5–10
R_4 (mm)	30–60
R_5 start (mm)	5–15
R_5 end (mm)	10–25
Bead D1 (mm)	30–60
Bead D2 (mm)	100–130
Draft angle (°)	35–60

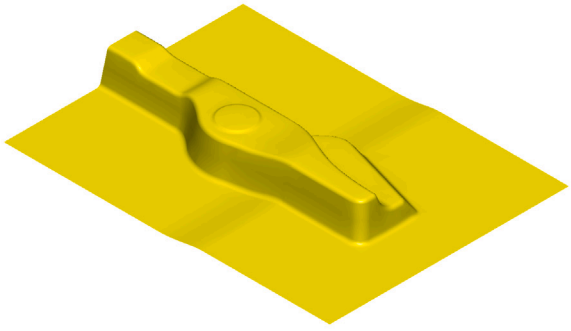
draft angle. Additional fillet radii, namely R_3, R_4, and R_5, were also introduced to represent variations in local geometric transitions. In particular, R_5 represents a variable fillet radius, for which the starting and ending radii were independently sampled from specified ranges. These geometric features were treated as variable parameters in the present study. Latin Hypercube Sampling (LHS) was then employed to generate 600 distinct combinations of the geometric parameters, resulting in 600 unique three-dimensional CAD geometries with approximately uniform coverage of the defined design space. The specific sampling ranges adopted for each geometric parameter are provided in Table 1.

3.3. Input and target image preparation

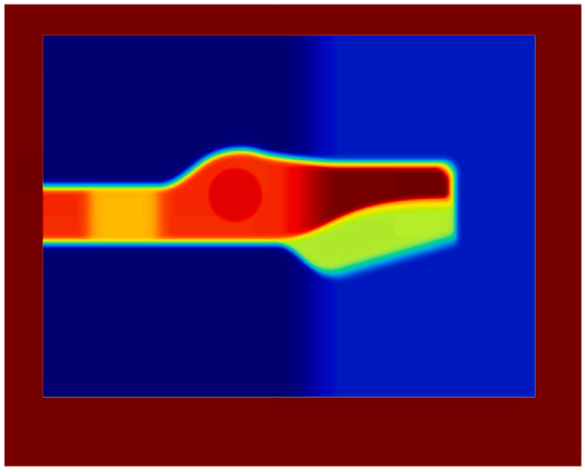
The StampFormer model accepts multimodal inputs: a 1D vector for the stress-strain curve and a 2D image representing the part's geometry. The model outputs a 2D image representing the predicted physical field. This output is a single-channel image for physical fields (e.g., thinning, plastic strain) plotted on the undeformed configuration or a 3-channel image for the displacement field. This necessitates the conversion of both the 3D CAD geometries and the 3D FEA simulation results into a uniform 2D image format.

The input geometry images are generated from the 3D CAD models. An information-preserving Z-plane projection is employed for the undercut-free panels considered here, as shown in Fig. 6(a) and (b). This projection relies on the fact that the geometry of a stamped component can be accurately represented as a single-valued height-map, without overlapping surfaces in the pressing direction. The out-of-plane height (Z-value) at each (X, Y) coordinate is recorded to create this 2D height-map. This height-map is then interpolated onto a uniform Cartesian grid with a resolution of 608×768 pixels, resulting in a single-channel image that serves as the 2D geometric input.

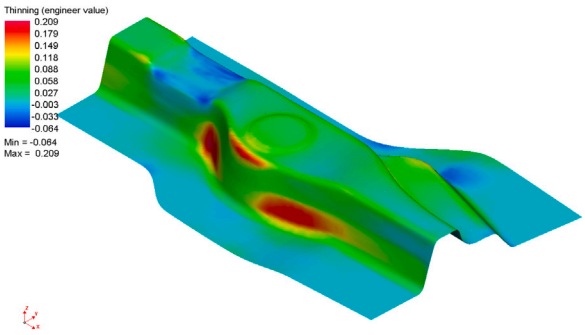
The target physical field images are generated from the FEA results. FEA simulations yield physical field data (e.g., thinning) on the final deformed 3D mesh as shown in Fig. 6(c). To ensure spatial alignment with the 2D input, these 3D results are mapped back to the initial undeformed 2D blank. Using the simulation's displacement field, each node on the deformed mesh is mapped to its original (X, Y) coordinates, and



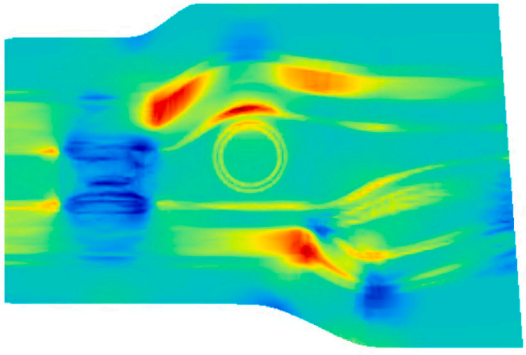
(a) 3D Die Geometry (Input)



(b) 2D Projected Height-map



(c) 3D Deformed Thinning Field (Target)



(d) 2D Mapped Thinning Field

Fig. 6. Illustration of the 3D-to-2D conversion process for model inputs and targets.

its corresponding physical field value is transferred. This process creates a 2D point cloud, which is then interpolated onto the same uniform grid, as illustrated in Fig. 6(d). For 2D physical fields such as thinning, major strain, minor strain, and plastic strain, this process yields a single-channel target image. For the 3D displacement field, the X, Y, and Z displacement components are similarly mapped back to the original grid, creating a 3-channel image that represents the 3D vector field.

3.4. Design of experiments

Building upon our prepared geometry and material data, we designed experiments to generate the target FEA simulation outcomes. This process created a comprehensive Design of Experiments (DoE) matrix. For the steel dataset, we randomly partitioned the 600 geometries into training, validation, and testing sets in a ratio of 0.8, 0.1, and 0.1. Each geometry was then paired with five materials drawn from five different groups, resulting in planned training, validation, and testing sets with 2400, 300, and 300 samples, respectively. A sample of this pairing is shown in Table 2. The same DoE methodology was applied to the aluminium dataset. Each material system was planned to contain 3000 simulations, giving 6000 planned simulations in total.

3.5. FEA simulation setup

The FEA simulations were employed to generate the training, validation, and test datasets for StampFormer. A representative automotive panel forming model was developed in PAM-STAMP (Fig. 7). The model integrates the tooling and blank geometries, the process setup, and the FEA solver controls. Tooling kinematics follow production practice: the die travels in the press direction towards a fixed punch, while a blank holder (binder) applies a prescribed blank-holder force to regulate sheet draw-in between the binder and die. All other parameters and settings, except for material type and geometry, were fixed, as specified in Table 3. The simulation matrix was generated automatically using integrated Python and PAM-STAMP scripts. Accordingly, the present dataset evaluates material and geometric variations under fixed process conditions; blank-holder force, friction coefficient, sheet thickness, and forming speed are not model input variables in the current study.

Table 2
A sample of the design of experiments matrix for the steel dataset. Note that for each distinct geometry, one material is systematically sampled from each of the five groups.

ID	Geometry ID	Material ID	Material Group
0	472	104	1
1	472	225	2
2	472	328	3
3	472	449	4
4	472	532	5
5	471	52	1
6	471	232	2
7	471	348	3
8	471	412	4
9	471	501	5

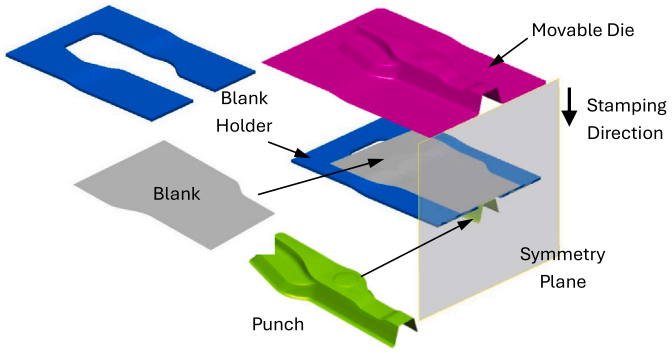


Fig. 7. The FEA model setup.

Table 3
The process parameters for the FEA model.

Model Parameter	Value/Definition
Blank Thickness	1.6 mm
Material Model	Isotropic Lookup Table
Tool Type	Feasibility Stamping (3-piece tools)
Forming Speed	250 mm/s (constant)
Blank Holding Force	800 kN
Friction Coefficient	0.12

Fig. 8 illustrates the overall simulation flow of the stamping process for the part, which is divided into three main stages: gravity loading, blank holding, and stamping, each representing a critical step in capturing the realistic behaviour of the material during forming. One example of the results obtained is the plastic strain distribution throughout the part, evaluated at the end of each stage, as illustrated in Fig. 9. Notably, due to occasional numerical non-convergence, only 5960 valid samples were successfully obtained and retained from the 6000 simulations planned in the DoE matrix.

4. Network architecture

In this study, we introduce StampFormer, a novel deep learning framework that provides fast and accurate predictions of complex physical fields in FEA. Throughout this section, let $I_{\text{geo}} \in \mathbb{R}^{H\times W\times 1}$ denote the input 2D geometry image, and let $S \in \mathbb{R}^{T\times 1}$ denote stress values sampled at T fixed strain levels representing the 1D stress-strain curve. As shown in Fig. 10, the proposed StampFormer is composed of three main components: MAGN, HMEIU, and the Physics-Guided Swin-UNet (PGSU).

4.1. Material-augmented geometric network

Inspired by heterogeneous data processing architectures, the MAGN in StampFormer is responsible for the initial multimodal fusion of the 2D part geometry and the 1D stress-strain curve. This robust multimodal integration is achieved through three key stages: material processing, material-aware feature modulation, and spatial feature integration.

For the material path, the 1D stress-strain curve $S \in \mathbb{R}^{T\times 1}$ is transformed into a dense representation. It is first linearly projected into a d_{mat}-dimensional space using a weight matrix $W_S \in \mathbb{R}^{1\times d_{\text{mat}}}$ and combined with a learnable positional embedding $E_{\text{pos,mat}} \in \mathbb{R}^{T\times d_{\text{mat}}}$. This sequence is then processed by a Transformer Encoder Layer to capture the intricate constitutive relationships within the material. Next, the output tokens from this Transformer are aggregated via average pooling (applied over the T token dimension) to produce a condensed, global material embedding $E_{\text{mat}} \in \mathbb{R}^{d_{\text{mat}}}$:

$$E_{\text{mat}} = \text{AvgPool}(\text{TransformerEncoder}(SW_S + E_{\text{pos,mat}})). \tag{4}$$

In the Material-Aware Feature Modulation (MAFM) stage, we map the global material characteristics into the spatial geometric domain. The input geometry is represented as a single-channel image $I_{\text{geo}} \in \mathbb{R}^{H\times W\times 1}$. To align the feature spaces, E_{mat} is passed through a multi-layer perceptron (MLP) to form a conditional material vector $\text{MLP}(E_{\text{mat}}) \in \mathbb{R}^{C}$, while I_{geo} is independently processed by a point-wise convolution (1×1 Conv) to extract its base spatial embeddings. The material vector is then spatially broadcast across the spatial dimensions and added to the geometric embeddings. This modulated feature is refined by a sequence of convolutional layers to yield an intermediate geometry-material fused feature map $F_{\text{fused}} \in \mathbb{R}^{H\times W\times C}$:

$$F_{\text{fused}} = \text{ConvLayers}(\text{Conv}_{1\times 1}(I_{\text{geo}}) + \mathcal{B}(\text{MLP}(E_{\text{mat}}))), \tag{5}$$

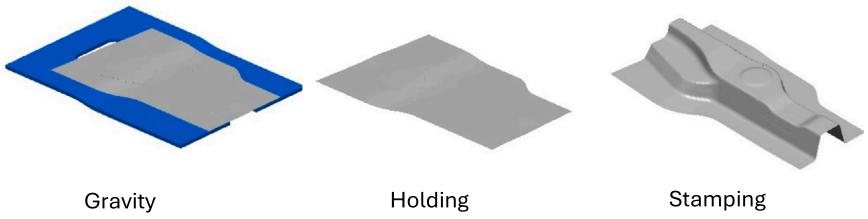


Fig. 8. FE simulation stages used in the forming workflow: gravity loading, blank holding, and stamping.

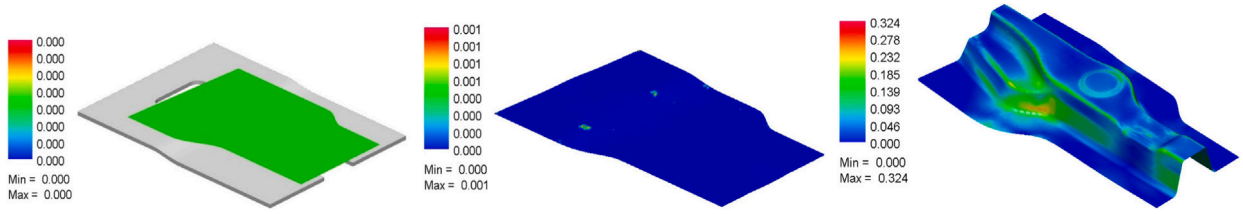


Fig. 9. FE simulation results showing the effective plastic strain distribution at different stages for one of the matrix cases.

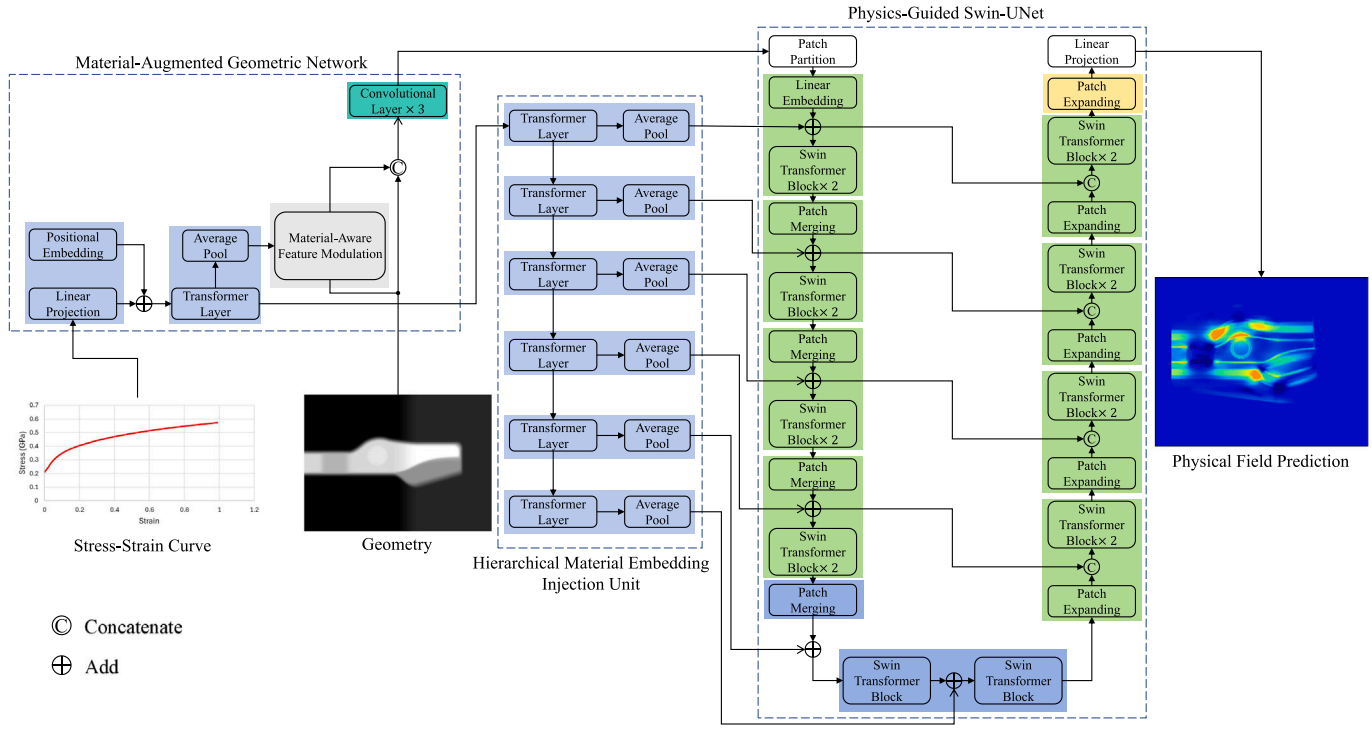


Fig. 10. Overall architecture of the StampFormer model. The framework consists of three main components: the MAGN for fusing geometry and material data, the HMEIU for multi-scale feature integration, and the PGSU backbone for predicting the final physical field.

where $\mathcal{B}(\cdot)$ denotes the spatial broadcasting operation that broadcasts the vector to $\mathbb{R}^{H\times W\times C}$. This guarantees that the geometric representations are uniformly conditioned by the macroscopic material properties.

Finally, in the spatial feature integration stage, we employ a concatenation-based skip connection to preserve pristine geometric boundary information. The initial raw geometry I_{geo} is concatenated with the fused feature F_{fused} along the channel dimension:

$$Z_{\text{concat}} = \text{Concat}(I_{\text{geo}}, F_{\text{fused}}). \tag{6}$$

This concatenated representation seamlessly bridges the explicit physical boundaries and the deep, material-conditioned features. Subsequently, Z_{concat} is processed through a final integration block (denoted as ConvInt) consisting of consecutive 3×3 convolutional layers coupled with ReLU activations. This yields the final, high-fidelity material-augmented feature map $X^{(0)} = \text{ConvInt}(Z_{\text{concat}})$, which is fed directly into the main Swin-UNet backbone.

4.2. Physics-guided Swin-UNet

As the main backbone of StampFormer, the PGSU is responsible for processing the fused multimodal feature map $X^{(0)}$ to predict the final physical field. It adopts the powerful Swin-UNet architecture, which follows a symmetric encoder–decoder structure built upon a series of Swin-Transformer blocks.

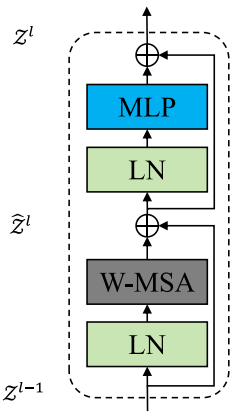


Fig. 11. The structure of a Swin Transformer block, which consists of Layer Normalisation (LN), Windowed Multi-head Self-Attention (W-MSA), and a multi-layer Perceptron (MLP), with residual connections.

As detailed in Fig. 11, the core of this architecture leverages the Window-based Multi-head Self-Attention (W-MSA) mechanism. The computation for a single Swin-Transformer block operating on an input feature map z^{l-1} is given by:

$$\hat{z}^l = \text{W-MSA}(\text{LN}(z^{l-1})) + z^{l-1}, \tag{7}$$

$$z^l = \text{MLP}(\text{LN}(\hat{z}^l)) + \hat{z}^l, \tag{8}$$

where LN denotes Layer Normalisation. By employing this window-based attention design, the network achieves robust feature extraction while maintaining high computational efficiency.

The encoder path takes the initial multimodal feature map $X^{(0)}$ as its input and consists of four hierarchical stages. Each stage comprises a pair of Swin-Transformer blocks followed by a Patch Merging layer, which reduces the spatial resolution while progressively increasing the feature dimension.

The physics-guided nature of this architecture stems from its explicit modelling of the multi-scale interaction between the two core physical inputs: geometry and material properties. The physics of forming dictates that global material characteristics (a 1D concept) influence local geometric deformation (a 2D field) at all scales. Therefore, our design integrates the HMEIU (detailed in Section 4.3) to hierarchically inject material information at each stage of the encoder, specifically after the spatial features have been downsampled by the Patch Merging layer. This allows the model to learn how global material characteristics affect both local, fine-grained features in shallow layers, and global, abstract features in deep layers.

The decoder path is symmetric, using Patch Expanding layers to upsample the features. Crucially, at each upsampling step, the decoder path brings back high-resolution, fine-grained features from the corresponding encoder stage via skip connections. These early features are fused with the abstract, semantically rich deep features from the deeper decoder layers. This fusion ensures the final prediction is both semantically rich and spatially precise.

4.3. Hierarchical material embedding injection unit

Instead of a single fusion at the MAGN, we introduce the HMEIU to ensure a deep, multi-scale integration of material embeddings throughout the network's main backbone, PGSU. It provides material embeddings of appropriate size at each stage of the feature extraction hierarchy, allowing the model to leverage material properties in conjunction with geometric features at various scales.

As shown in Fig. 10, the HMEIU operates as a parallel tower to the PGSU's down-sampling encoder. It takes the sequence of token embeddings from the initial stress-strain curves Transformer in the MAGN as its input, which is then progressively processed through a stack of dedicated Transformer Layers.

At each hierarchical level l of the PGSU encoder, the corresponding Transformer Layer within the HMEIU refines the stress-strain curves. Crucially, each of these dedicated Transformer layers is designed to output a feature dimension that aligns with the channel dimension of the corresponding PGSU encoder stage. The output tokens from this layer are then aggregated via average pooling to produce the level-specific material embedding, $E_{\text{mat}}^{(l)}$. As shown in Fig. 10, this embedding is injected into the main data stream of the PGSU after the Swin-Blocks and Patch Merging layer at each encoder stage. Let $X^{(l-1)}$ be the input to the l-th stage. The geometric features are first processed by the two consecutive Swin Transformer blocks and then downsampled by the Patch Merging layer to obtain the intermediate spatial feature $F^{(l)} \in \mathbb{R}^{N_l \times C_l}$:

$$F^{(l)} = \text{PatchMerging}(\text{SwinBlock}(\text{SwinBlock}(X^{(l-1)}))). \tag{9}$$

$E_{\text{mat}}^{(l)}$ is then broadcast and added to geometric features using an outer-product formulation to match dimensions:

$$X^{(l)} = F^{(l)} + \mathbf{1}_{N_l} \left(E_{\text{mat}}^{(l)}\right)^T, \tag{10}$$

where $\mathbf{1}_{N_l}(E_{\text{mat}}^{(l)})^T \in \mathbb{R}^{N_l \times C_l}$ and $\mathbf{1}_{N_l}$ represents an N_l-dimensional column vector of ones, explicitly denoting the spatial broadcasting of the material embedding across all geometric patches. This hierarchical injection mechanism ensures that as the geometric features become more abstract and complex during the downsampling process, the material representation is co-adapted and fused at each corresponding scale.

5. Experiments

5.1. Training setup

In this study, ten separate full StampFormer models were trained: one dedicated StampFormer model for each of the five target physical fields (thinning, major strain, minor strain, plastic strain, and displacement), for each of the two material datasets (steel and aluminium). All ten models were trained using a consistent experimental setup on their respective data, as described in Section 3. We employed the Adam optimiser for model training with standard parameters: beta parameters of $\beta_1 = 0.9$ and $\beta_2 = 0.999$ and an epsilon of $\epsilon = 10^{-8}$. The learning rate was managed by a step learning-rate scheduler, with an initial rate of 10^{-4}. The scheduler was configured to decrease the learning rate by a factor of $\gamma = 0.4$ every 100 epochs. The optimisation goal for each model was to minimize the Mean Squared Error (MSE) between its predicted physical field and the ground truth.

For the ablation comparison, two additional configurations were trained for each material–field task: a configuration without MAGN but retaining HMEIU (w/o MAGN), and a geometry-only configuration without either module. These are compared with the complete StampFormer architecture.

Considering the predicted field $\hat{Y} \in \mathbb{R}^{H \times W \times C}$ and the ground truth field $Y \in \mathbb{R}^{H \times W \times C}$ from the FEA simulation, the MSE loss is computed directly from the predicted and reference field values and is defined as:

$$\mathcal{L}_{\text{MSE}} = \frac{1}{H \cdot W \cdot C} \sum_{i=1}^{H} \sum_{j=1}^{W} \sum_{k=1}^{C} (Y_{ijk} - \hat{Y}_{ijk})^2. \tag{11}$$

Here, H and W are the height and width of the field, and C is the number of channels. Note that $C = 1$ for the four 2D fields, and $C = 3$ for the 3D displacement field.

During inference, the trained network is operated in evaluation mode with fixed model weights; therefore, identical preprocessed geometry and material inputs produce repeatable predictions apart from negligible hardware-level floating-point variation.

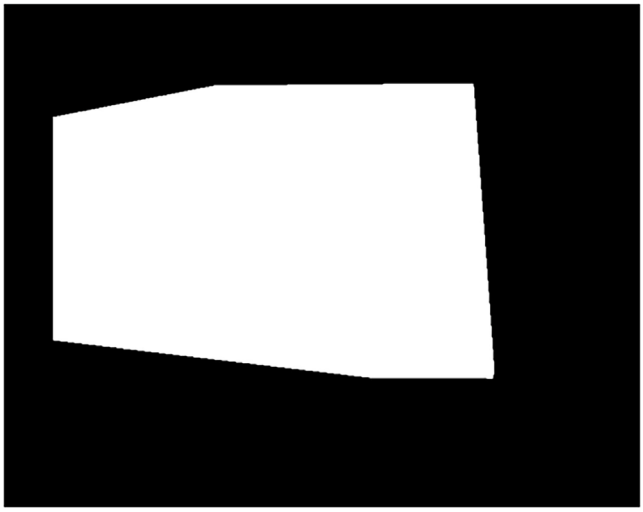


Fig. 12. Conservative mask used for the valid-domain MSE evaluation. The white region represents the physical evaluation domain, while the artificial padded background is excluded. (For interpretation of the references to colour in this Figure legend, the reader is referred to the web version of this article.)

5.2. Metrics

The training MSE loss is computed over the complete rectangular tensor, which contains zero padding. Hence, we introduce the masked MSE, which is evaluated over the valid physical domain to exclude the zero-padded background. Let $M_{ij} \in \{0,1\}$ denote the evaluation mask, where $M_{ij} = 1$ for the physical sheet region. The masked MSE is defined as

$$\mathrm{MSE}_{\mathrm{mask}} = \frac{\sum_{i=1}^{H}\sum_{j=1}^{W}\sum_{k=1}^{C} M_{ij}(Y_{ijk} - \hat{Y}_{ijk})^2}{C\sum_{i=1}^{H}\sum_{j=1}^{W} M_{ij}}. \tag{12}$$

Here, the mask used in our study is shown in Fig. 12. The mask represents the blank region in the stamping simulation, identified from the non-zero support of the mapped FEA data during preprocessing. The surrounding zero-padded background is excluded, and the same mask is applied to the reference and predicted fields for all compared configurations.

In addition to the masked MSE, we employ an engineering-oriented metric to evaluate the four 2D physical fields: thinning, major strain, minor strain, and plastic strain. For these fields, we focus on the accuracy of predicting the maximum value, as this is often of critical importance in forming analysis. To ensure stability and mitigate the effect of potential outliers, we define the representative maximum value as the average of the top 0.1% of values in a given field. Let $m_{0.1}(Y)$ and $m_{0.1}(\hat{Y})$ be the representative maximum values for the ground truth and predicted fields, respectively. We then calculate the Relative Error (RE) as:

$$\text{Relative Error } (\%) = 100 \times \frac{|m_{0.1}(Y) - m_{0.1}(\hat{Y})|}{m_{0.1}(Y)}. \tag{13}$$

The representative-maximum RE is therefore used as a complementary engineering-oriented metric together with the masked MSE, rather than as a replacement for full-field error evaluation.

To assess spatial localisation of critical thinning, we additionally measure the coverage of the ground-truth top 0.3% region by the predicted top 1% region. For each test sample, let Ω_{thin} denote the valid thinning evaluation domain, $G_{0.3}$ its pixels with the highest 0.3% of ground-truth thinning values, and P_1 its pixels with the highest 1% of predicted thinning values. Critical-region recall is defined as:

$$\mathrm{Recall}_{0.3\to1}(\%) = 100 \times \frac{|G_{0.3} \cap P_1|}{|G_{0.3}|}. \tag{14}$$

This metric measures coverage of the reference critical region within a fixed candidate inspection area of 1% of the valid domain. It complements the representative-maximum RE, which compares amplitudes rather than spatial coincidence. Recall is calculated separately for every test sample using the raw predictions.

For the 3D displacement field, we additionally evaluate the Mean Euclidean Displacement Error (MEDE) over the non-zero displacement region. Let $\mathbf{u}_p = [u_x, u_y, u_z]^T$ and $\hat{\mathbf{u}}_p$ denote the ground-truth and predicted displacement vectors at grid point p, respectively. The metric is defined as:

$$\mathrm{MEDE} = \frac{1}{|\Omega|}\sum_{p\in\Omega}\left\|\hat{\mathbf{u}}_p - \mathbf{u}_p\right\|_2, \tag{15}$$

where Ω denotes the non-zero displacement region. This metric represents the mean magnitude of the three-dimensional displacement prediction error and is expressed in millimetres.

In subsequent sections, a localised denoising step is applied exclusively to the visualisation of reconstructed 3D displacement surfaces to suppress boundary-padding artifacts. All primary quantitative results reported in Section 6 are computed directly from the raw model predictions without post-processing. The filtered results reported in Section 7.3 are used only to quantify the influence of this visualisation post-processing.

6. Results

In this section, we present a comprehensive evaluation of the StampFormer model. We begin with a detailed quantitative analysis of prediction accuracy using numerical metrics. The following subsections offer a visual analysis of the model's performance by presenting the predicted physical fields, which demonstrate qualitative agreement with the FEA ground truth.

6.1. Quantitative analysis

In this section, we perform a detailed quantitative analysis to evaluate the performance of StampFormer. To begin with, the masked MSE on the test set was calculated for both the steel and aluminium datasets across the five predicted physical fields. As defined in Section 5.2, this evaluation excludes the artificially zero-padded background and calculates the prediction error over the valid physical sheet region. The results are summarised in Table 4.

We recognise the limitations of using only MSE to evaluate model performance. Therefore, the field-specific RE and MEDE metrics defined in Section 5.2 are subsequently used to provide complementary engineering interpretation for the four 2D fields and the three-component displacement field, respectively. For the four 2D physical fields, performance is evaluated using the RE as defined in Eq. (13). We illustrate the distribution of RE values for the four 2D physical fields in Fig. 13. For the steel dataset, the mean RE is consistently low across all four fields, ranging from 4.5% to 7.2%, with the 95th percentile below 15% in most cases. These results indicate agreement with the representative-maximum strain/thinning amplitudes, while spatial coverage of critical thinning regions is evaluated separately in Section 6.2.

For the aluminium dataset, the mean RE ranges from 3.5% to 8.2%, with differences between the individual fields. For both materials, the error distributions were right-skewed, with a tail extending towards larger errors. However, the aluminium dataset exhibits heavier error tails, particularly for major strain and plastic strain. Further examination of the ground-truth representative maxima shows that the aluminium cases span a broader and, for several fields, higher range of critical-region responses than the steel cases, as shown in Fig. 14. Together with the smaller number of baseline aluminium stress–strain curves, this suggests that the heavier tails may be associated with sparser sampling of relatively uncommon extreme responses.

In addition, we present the distribution of the MEDE for the 3D displacement field in Fig. 15. For steel, the mean MEDE is 1.115 mm,

Table 4
Test-set masked MSE over the valid physical domain.

Dataset	Thinning	Major Strain	Minor Strain	Plastic Strain	Displacement (mm^2)
Steel	7.0×10^{-5}	1.02×10^{-4}	2.9×10^{-5}	1.91×10^{-4}	0.7001
Aluminium	1.15×10^{-4}	1.70×10^{-4}	2.4×10^{-5}	2.54×10^{-4}	0.4511

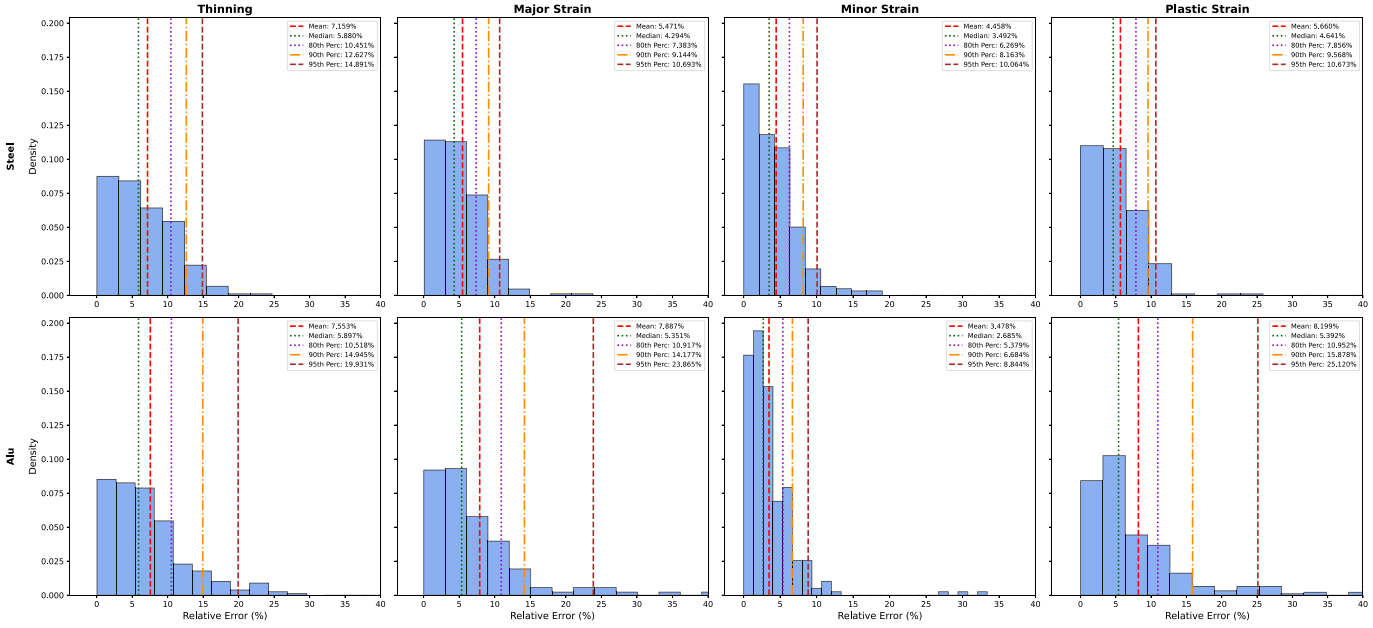


Fig. 13. Histograms showing the distribution of RE computed from the representative maximum, defined as the average of the top 0.1% of valid values in four 2D physical fields (i.e., thinning, major strain, minor strain and plastic strain). Additional lines are introduced to mark the mean, median, and various percentiles.

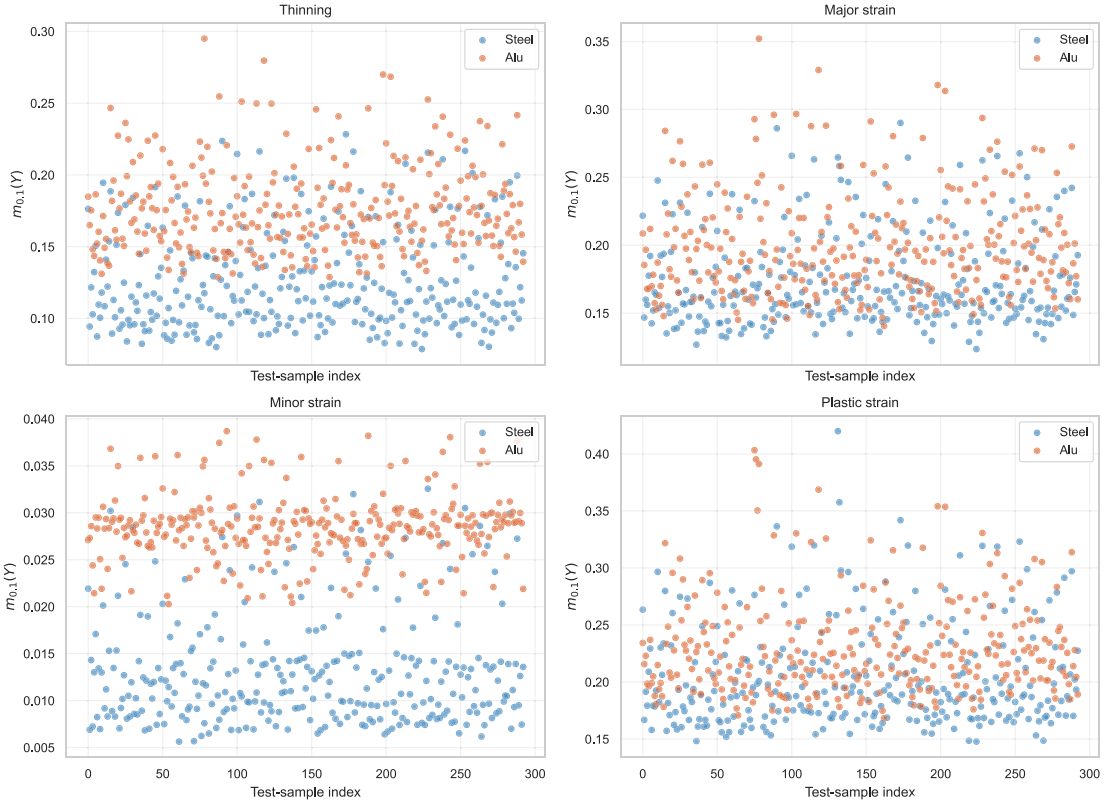


Fig. 14. Ground-truth representative maximum $m_{0.1}(Y)$ for the four 2D physical fields in the steel and aluminium test sets.

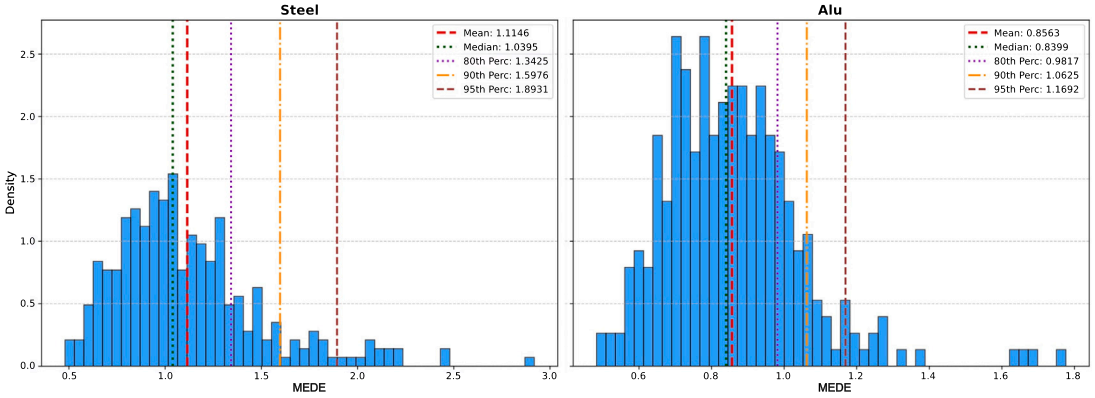


Fig. 15. Histograms showing the distribution of the mean Euclidean displacement error (MEDE) for the 3D displacement field. Additional lines indicate the mean, median, and selected percentiles.

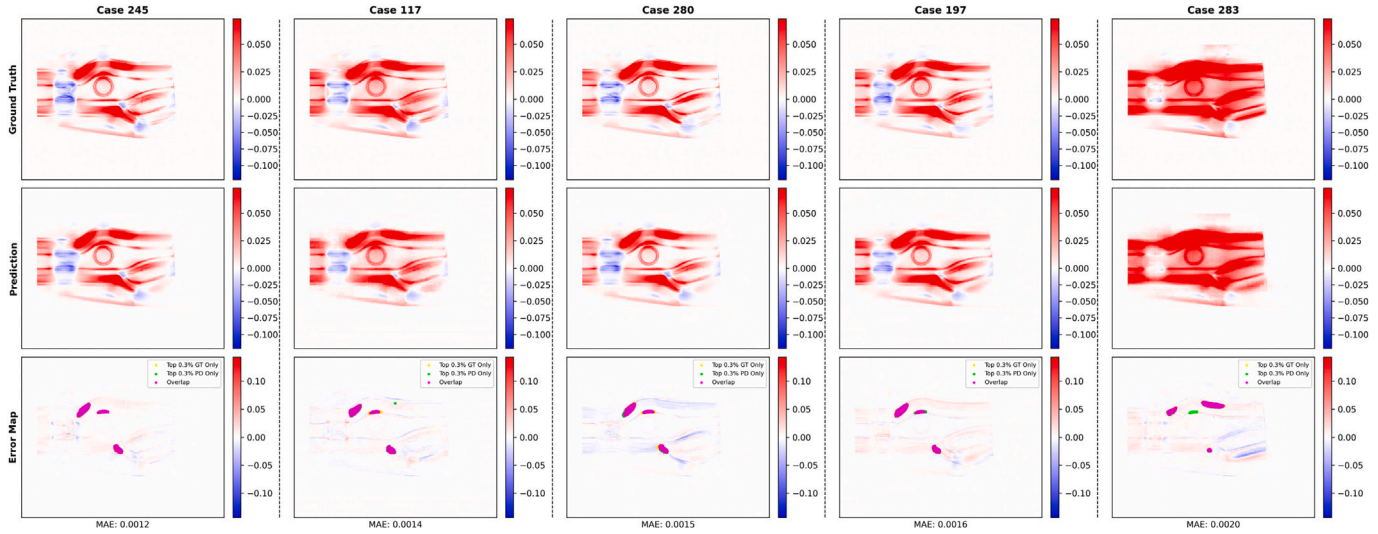


Fig. 16. Visualisation for thinning field prediction on five representative cases from the steel dataset. The top-value overlap map specifically visualizes the top 0.3% of values to compare the prediction of maximum thinning, highlighting regions of overlap, ground truth (GT) only, and prediction (PD) only (both regions use a top 0.3% threshold here; the representative-maximum RE uses a top 0.1%, and the separate recall analysis uses ground-truth top 0.3% and predicted top 1%). (For interpretation of the references to colour in this Figure legend, the reader is referred to the web version of this article.)

while for aluminium it is 0.856 mm. Unlike MSE, MEDE measures the mean magnitude of the three-dimensional displacement error over the non-zero displacement region and is therefore expressed directly in millimetres. It provides a complementary and physically intuitive interpretation of the displacement prediction error.

6.2. *Thinning field prediction*

In this section, we illustrate the results for the thinning field in Figs. 16 and 17. Each case is presented in three rows: the ground truth field from FEA, the corresponding prediction from our model, and a detailed top-value overlap map along with the prediction error. In industrial applications, accurate prediction of maximum thinning is important for identifying potential failure-prone regions. To better evaluate the prediction of maximum thinning, we visualize the regions with the highest values on the top-value overlap map. Specifically, we plot the top 0.3% of thinning values from both the ground truth (GT) and the prediction (PD). The top-value overlap map uses a colour-coded system to distinguish these regions: yellow highlights areas where high values are found only in the ground truth, green marks areas exclusive to the prediction, and magenta indicates where they overlap. The top-value overlap maps illustrate agreement in the locations of critical thinning in these selected cases; the following recall analysis extends the spatial assessment to the complete test sets.

To quantify critical-region coverage across the complete test set for each material, Fig. 18 reports the thinning recall defined in Section 5.2. The predicted top 1% region is treated as a candidate inspection area, and its coverage of the ground-truth top 0.3% region is evaluated. This represents an engineering screening scenario in which a somewhat broader candidate region is examined to reduce the risk of missing critical locations. Unlike the equal-fraction overlap maps above, this analysis evaluates coverage within a fixed inspection budget rather than exact agreement between equally sized regions. The median recall is approximately 73% for both steel and aluminium, indicating that the predicted candidate region captures a substantial proportion of the actual critical thinning region. The variation across cases also shows that this coverage is not uniform. These results complement the amplitude-based RE and support prioritisation of locations for further engineering assessment.

6.3. *Major-minor strain distribution prediction*

The predicted major and minor strain fields were further evaluated by plotting their major-minor strain distributions. These plots use the same axes as a forming limit diagram, with minor strain on

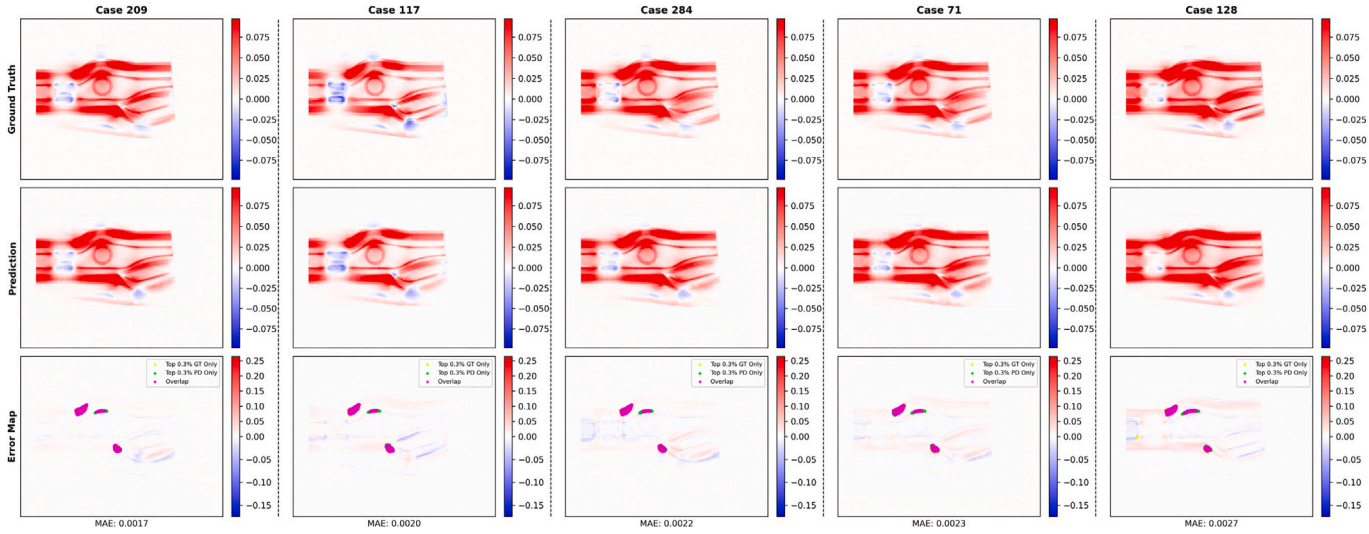


Fig. 17. Visualisation of thinning field prediction for five representative cases from the aluminium dataset. The top-value overlap map specifically visualizes the top 0.3% of values to compare the prediction of maximum thinning, highlighting regions of overlap, GT Only, and PD Only (both regions use a top 0.3% threshold here; the representative-maximum RE uses top 0.1%, and the separate recall analysis uses ground-truth top 0.3% and predicted top 1%). (For interpretation of the references to colour in this Figure legend, the reader is referred to the web version of this article.)

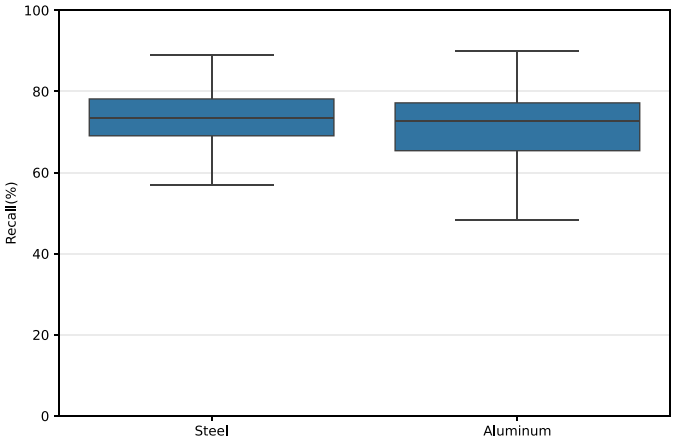


Fig. 18. Thinning critical-region recall across the complete steel and aluminium test sets, measuring the proportion of ground-truth top 0.3% of pixels covered by the predicted top 1% region within the valid evaluation domain. (For interpretation of the references to colour in this Figure legend, the reader is referred to the web version of this article.)

the horizontal axis and major strain on the vertical axis. Rather than complete FLDs, they are interpreted here as strain-state distributions. Figs. 19 and 20 compare the ground-truth and predicted strain-state point clouds for the steel and aluminium datasets. The overlays show that the model reproduces the overall shape, density and spread of the strain distributions, indicating that it captures the coupled major/minor strain response across the formed panel.

6.4. Displacement and plastic strain prediction

As shown in Figs. 21 and 22, we compare the predicted and ground-truth as-formed components, which are reconstructed from the displacement fields. On these surfaces, we superimpose the corresponding plastic strain fields. The top row in each figure shows the ground truth from the FEA simulation, while the bottom row shows our model's prediction. Because the model's predictions can be affected by the zero-padding values at the edges, some noise may appear. To enhance the clarity of the 3D renderings of the displacement fields, we applied a localised denoising process as a visual post-processing step. The reconstructed surfaces provide a spatial comparison between the predicted and FEA displacement responses together with the corresponding plastic-strain fields, confirming the model's capability to predict displacement and plastic strain with high fidelity. The denoising operation is used only to improve the readability of these reconstructed visualisations; the quantitative results reported above are calculated from the raw predictions.

7. Discussion

7.1. Engineering significance of the proposed method

In the traditional design cycle, a primary bottleneck is that time-consuming and computationally expensive FEA hinders rapid design iteration and optimisation. In this study, we introduced StampFormer, a surrogate model designed to predict complex physical fields in the sheet metal forming process with high accuracy and high speed. To quantify this, each FEA simulation in our dataset required several hours of computation to complete. By contrast, StampFormer provides predictions for physical fields in under a second on a single GPU. This acceleration, representing a speed-up of more than three orders of magnitude (from hours to sub-second), directly enables real-time feedback that was previously impossible.

Beyond the computational speed-up, StampFormer explicitly incorporates both geometry and the full material stress–strain response in an image-based forming surrogate. The present validation demonstrates within-family parametric generalization rather than unrestricted transfer to arbitrary component topologies. Its most immediate application is therefore rapid early-stage decision support within a predefined component family under fixed process conditions.

Representative use cases include comparing candidate materials, screening parameterised geometric variants, supporting design-space exploration and optimisation, and prioritising a smaller number of candidate designs for detailed FEA. In this role, StampFormer complements rather than replaces high-fidelity CAE and physical validation.

7.2. Discussion on the relative error metric

In Section 5.2, we defined a field-specific relative-error metric, which uses a representative maximum value, calculated from the average of the

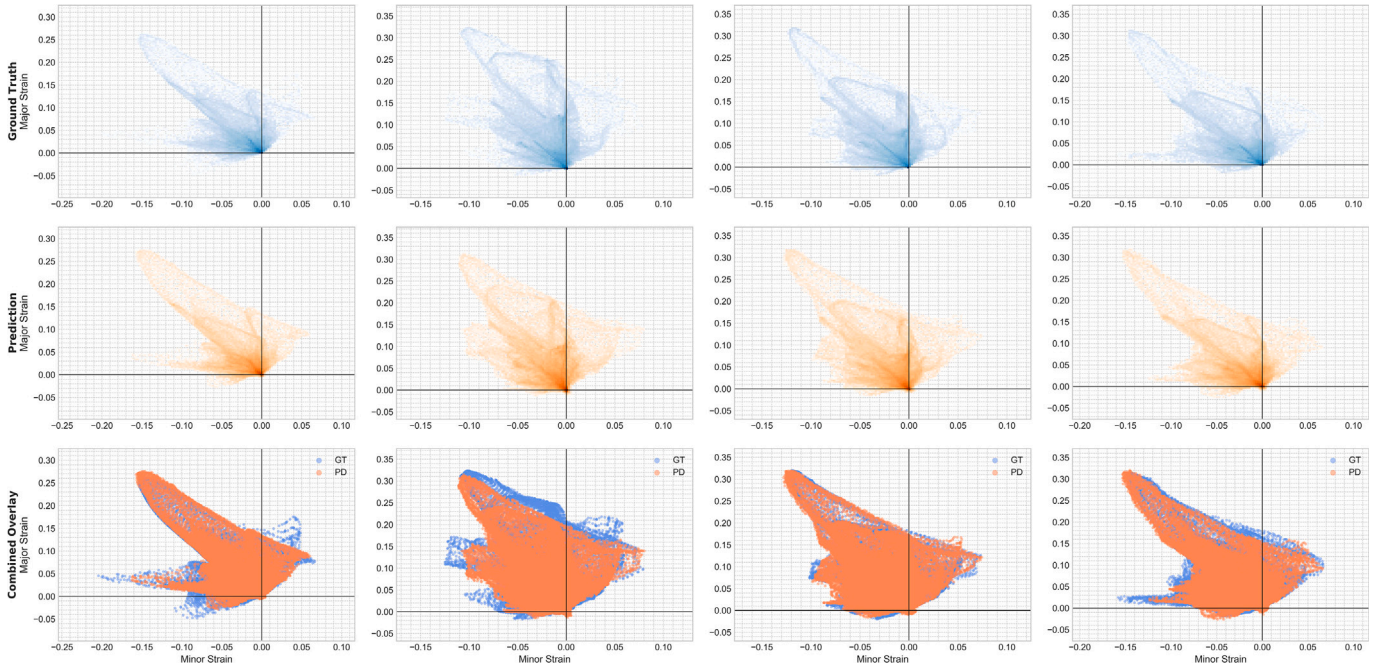


Fig. 19. Comparison of major-minor strain distributions from the steel dataset. The top two rows display the ground truth and predicted distributions, respectively. For direct comparison, the bottom row provides an overlay of both.

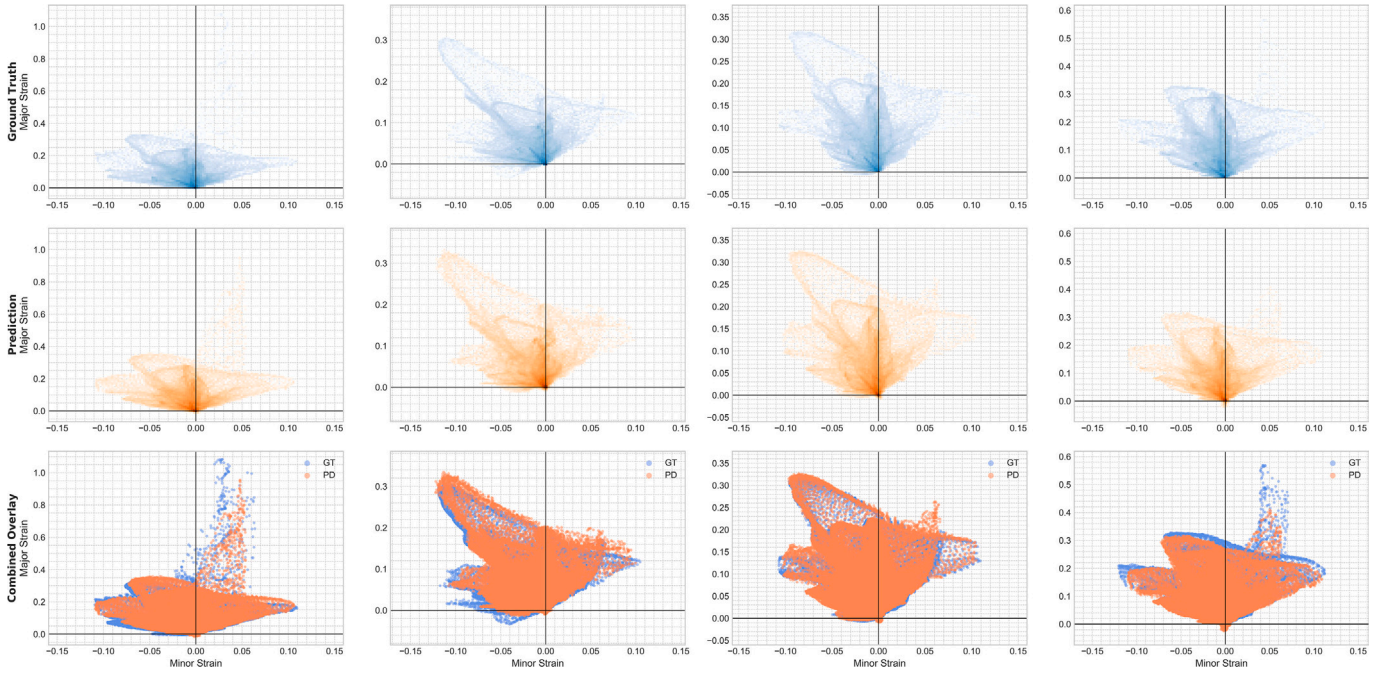


Fig. 20. Comparison of major-minor strain distributions from the aluminium dataset. The top two rows display the ground truth and predicted distributions, respectively. For direct comparison, the bottom row provides an overlay of both.

top 0.1% of values in a given field, rather than the absolute maximum value. This decision was made to enhance the numerical stability and robustness of our evaluation metric. As illustrated in Fig. 23, the distribution of values in a physical field like thinning can have a long tail, where a small number of nodes may exhibit extreme values due to numerical instability. Relying on a single maximum value can make the metric highly sensitive to these outliers. As a result, we average the top 0.1% of values to obtain a more stable and representative measure of the model's performance in the most critical regions, which is of greater practical importance in an industrial context.

7.3. Rationale for denoising displacement field predictions

As described in Section 5.2, denoising is used only for visualisation of the reconstructed 3D displacement surface. The rectangular tensor representation contains an artificially padded background, and

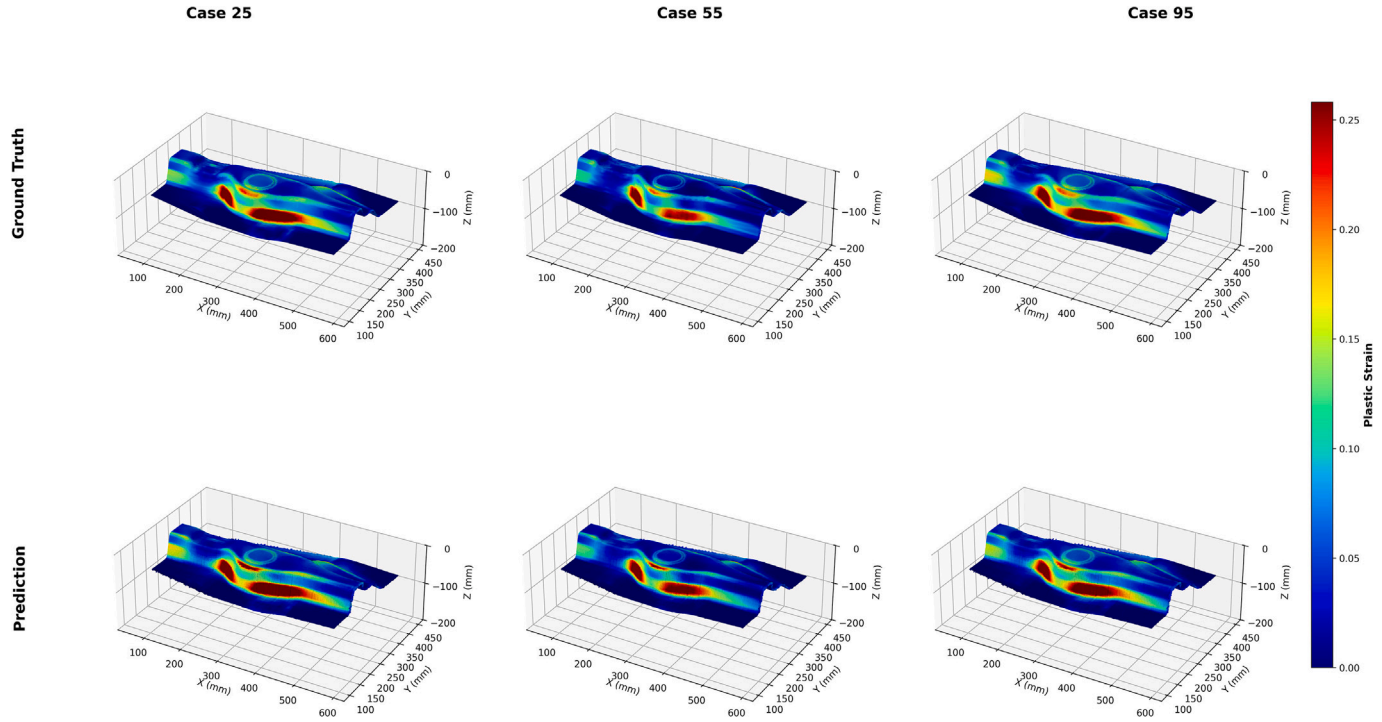


Fig. 21. Visualisation of displacement-reconstructed surfaces coloured by plastic strain from the steel dataset. The comparison between GT and PD highlights the model's accuracy in predicting the shape and plastic strain distribution of the final stamping part.

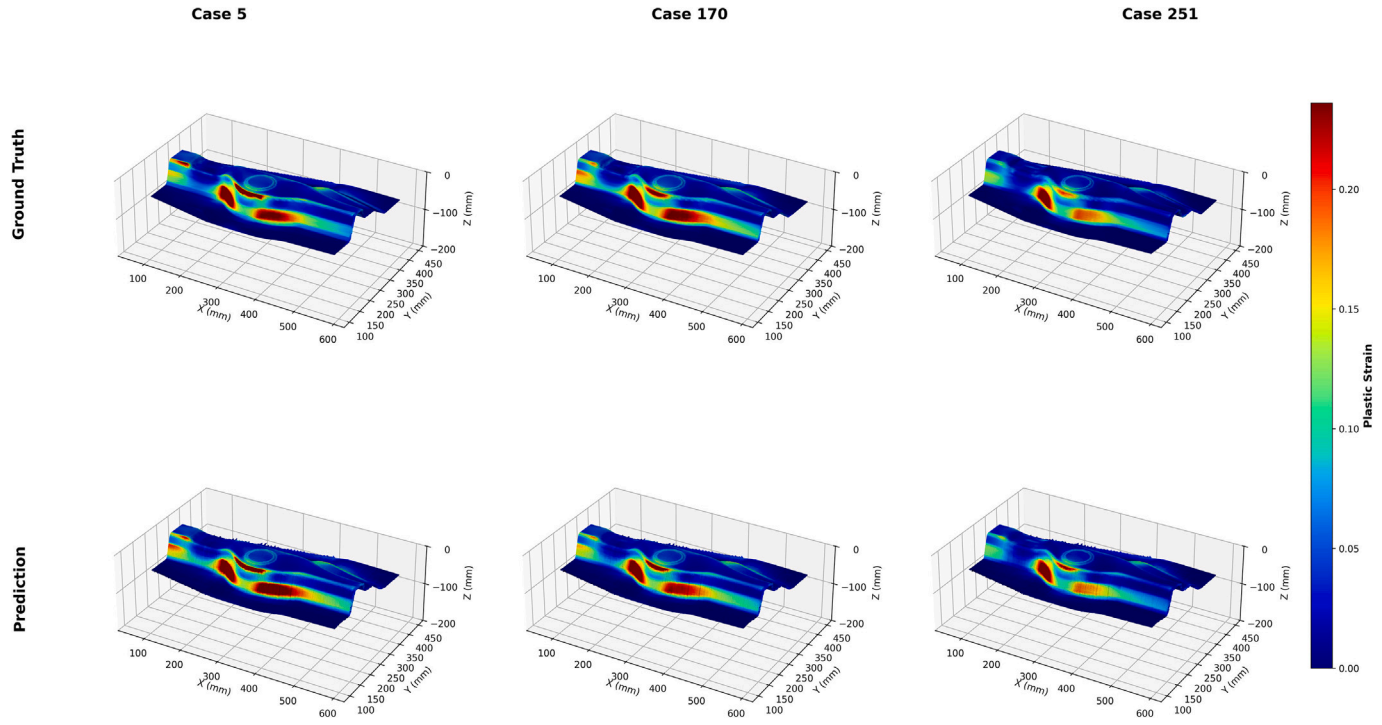


Fig. 22. Visualisation of displacement-reconstructed surfaces coloured by plastic strain from the aluminium dataset. The comparison between GT and PD highlights the model's accuracy in predicting the shape and plastic strain distribution of the final stamping part.

the discontinuity near this boundary can introduce fringe artifacts in the reconstructed surface (Fig. 24).

To distinguish boundary effects from interior prediction accuracy, we evaluated the raw displacement predictions over a boundary band comprising approximately 1% of the valid pixels and the remaining interior across all 293 test cases for each material. The masked-MSE expression in Section 5.2 is applied separately to each region, and the resulting sample-wise errors are averaged over the test set. As shown in Table 5, boundary MSE is approximately 5.04 and 11.79 times the interior MSE for steel and aluminium, respectively. This elevated boundary

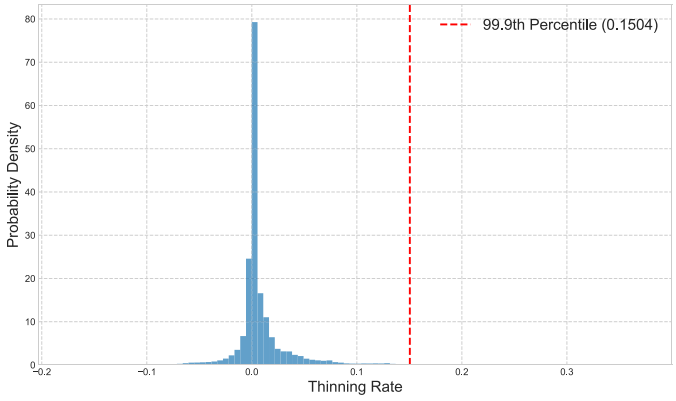


Fig. 23. The thinning field distribution of a random sample (where negative values indicate thickening), with a vertical line indicating the 99.9th percentile (top 0.1%) of values.

Table 5
Mean displacement MSE of the full model over the full valid domain, a boundary band comprising approximately 1% of the valid pixels, and the remaining interior (mm^2). Ratio denotes boundary MSE divided by interior MSE.

Material	All valid	Boundary	Interior	Ratio
Steel	0.7001	3.3947	0.6739	5.04
Aluminium	0.4511	4.8159	0.4086	11.79

error is consistent with the observed fringe artifacts, while the lower interior MSE indicates better prediction accuracy away from the boundary. Boundary pixels remain included in the primary masked MSE.

To quantify the influence of the visualisation filtering, we compared the masked displacement MSE before and after denoising. For steel, the MSE decreases from 0.7001 to 0.6925 mm^2, corresponding to a reduction of approximately 1.09%. For aluminium, it decreases from 0.4511 to 0.4456 mm^2, corresponding to a reduction of approximately 1.22%. These small changes indicate that the filtering has only a limited influence on the domain-averaged error and is used primarily to improve the readability of the reconstructed 3D visualisation. All primary quantitative results are based on the raw model predictions.

7.4. Ablation analysis of material conditioning

To examine the contribution of material conditioning, we compare three configurations: the complete StampFormer model (Full), a configuration without MAGN but retaining HMEIU (w/o MAGN), and a geometry-only configuration without either module. Table 6 reports the mean masked MSE over the 293 test cases for each material across all five predicted fields.

For all eight scalar-field tasks, covering four 2D fields for each material, the full model achieves the lowest mean masked MSE, followed by the no-MAGN and geometry-only configurations. Removing both material-conditioning modules increases the error to approximately 2.1–7.6 times that of the full model, approaching an order of magnitude for some tasks. This deterioration demonstrates that StampFormer effectively exploits constitutive information beyond what can be inferred from geometry alone. Adding MAGN to the HMEIU-containing configuration further reduces mean masked MSE by approximately 1.06%–3.47% across all eight tasks, supporting its contribution to the evaluated scalar-field predictions.

For displacement, the no-MAGN configuration achieves lower mean masked MSE than the full model (0.6696 versus 0.7001 mm^2 for steel and 0.3516 versus 0.4511 mm^2 for aluminium), although both outperform the geometry-only baseline. One possible explanation is an output-dependent interaction between early fusion and hierarchical material conditioning. Although MAGN retains the raw geometry through its concatenation-based skip connection, the additional fused features may affect how the backbone uses spatial information for displacement reconstruction. HMEIU provides hierarchical material conditioning without requiring this early fusion. These results motivate further investigation of fusion strategies tailored to different physical fields.

Table 6
Mean test-set masked MSE for the geometry-only, no-MAGN and full configurations. Displacement MSE is expressed in mm^2.

Material	Field	Geometry-only	w/o MAGN	Full
Steel	Thinning	5.3344×10^{-4}	7.1370×10^{-5}	6.9912×10^{-5}
	Major strain	6.8715×10^{-4}	1.0472×10^{-4}	1.0191×10^{-4}
	Minor strain	1.8284×10^{-4}	2.9422×10^{-5}	2.9111×10^{-5}
	Plastic strain	1.1660×10^{-3}	1.9466×10^{-4}	1.9118×10^{-4}
	Displacement (mm^2)	1.6140	0.6696	0.7001
Aluminium	Thinning	7.2375×10^{-4}	1.1824×10^{-4}	1.1545×10^{-4}
	Major strain	3.9938×10^{-4}	1.7628×10^{-4}	1.7019×10^{-4}
	Minor strain	1.2280×10^{-4}	2.4508×10^{-5}	2.3658×10^{-5}
	Plastic strain	5.3359×10^{-4}	2.6027×10^{-4}	2.5412×10^{-4}
	Displacement (mm^2)	0.9000	0.3516	0.4511

7.5. Limitations and future work

The present study has several applicability boundaries. First, the current validation is restricted to material and geometric variations within one parameterised crossmember component family; direct zero-shot transfer to substantially different component families has not been demonstrated. Extension to a new material system or geometry domain will therefore require representative data for adaptation and validation. The primary evaluation uses known material-card IDs on unseen geometries, so the specific material–geometry pairs in the test set are also unseen during training. Appendix A provides an additional held-out-material-ID evaluation for steel thinning, motivating further research into transfer to unseen materials and broader constitutive-response domains.

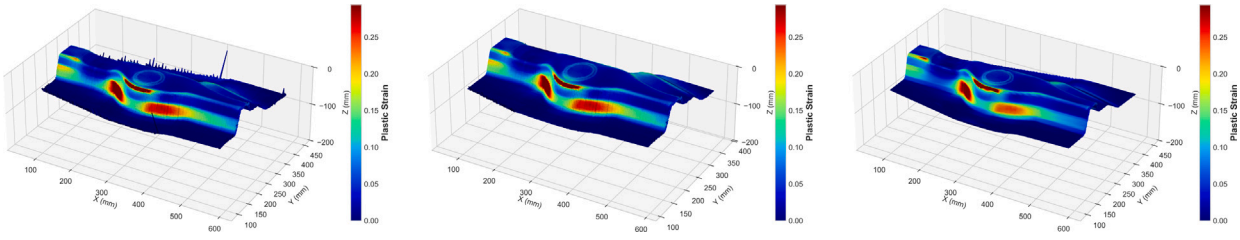
Second, process parameters including blank-holder force, friction coefficient, sheet thickness, and forming speed are fixed in the current dataset. StampFormer should therefore be interpreted as a material–geometry surrogate under these fixed process conditions rather than as a process-optimisation model. Incorporating process variables as an additional input modality is a relevant future extension.

Third, the surrogate is trained against PAM-STAMP simulation results and consequently learns the response of the underlying numerical model. Any systematic discrepancy between the CAE model and the physical forming process may therefore be inherited by the surrogate. For industrial deployment, the reference FEA model should first be calibrated and validated against physical forming trials.

Future work will investigate data-efficient adaptation and transfer-learning strategies for new materials and component families, incorporation of process variables, and further engineering and experimental validation.

8. Conclusion

This paper presents StampFormer, a physics-guided, material–geometry-coupled multimodal framework for rapid prediction of forming-response fields in sheet metal stamping. By integrating component geometry with material-response information, the model captures the influence of material behaviour on spatial deformation, thinning and



(a) Prediction before denoising (b) Prediction after denoising (c) Ground Truth

Fig. 24. Comparison of surfaces reconstructed from the displacement and coloured by plastic strain; denoising is used only for visualisation of the boundary artifacts caused by zero-padding in the training data.

strain distribution. The proposed architecture combines the Material-Augmented Geometric Network, the Hierarchical Material Embedding Injection Unit and a Physics-Guided Swin-UNet backbone to embed material information across multiple feature scales while preserving field-level spatial detail.

The results show that StampFormer reproduces the principal spatial patterns of the reference FEA fields across the investigated test cases. The model achieves a mean representative-maximum relative error below 8.2% for each of the four 2D forming fields in both material datasets and a masked displacement MSE below 0.71 mm^2, while generating predictions in under one second.

The primary validation demonstrates predictions for unseen geometries within a predefined crossmember family, using materials from the existing constitutive-response pool under fixed process conditions. Accordingly, StampFormer is primarily positioned as a rapid surrogate for early-stage material comparison, parametric geometric screening, and design-space exploration rather than unrestricted cross-family or process-variable generalisation. Extension to new component families, material systems, and variable process conditions remains an important direction for future work.

CRediT authorship contribution statement

Jiajie Luo: Writing – original draft, Software, Methodology, Conceptualization. **Mohamed Mohamed:** Writing – review & editing, Investigation, Data curation. **Osama Hassan:** Validation, Methodology. **Haosu Zhou:** Writing – review & editing, Software, Formal analysis. **Yingxue Zhao:** Writing – review & editing, Visualization, Formal analysis. **Haoran Li:** Writing – review & editing, Formal analysis. **Xinrun Li:** Methodology, Formal analysis. **Zhutao Shao:** Writing – review & editing, Resources, Funding acquisition. **Yang Long:** Writing – review & editing, Supervision, Methodology. **Jichun Li:** Writing – review & editing, Supervision, Funding acquisition, Formal analysis. **Nan Li:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Investigation, Funding acquisition, Conceptualization.

Declaration of Generative AI and AI-assisted technologies in the writing process

After completing the manuscript draft, the authors used Gemini and ChatGPT to assist with grammar, readability, and proofreading. These tools were not used to generate simulation data, design the experiments, implement the model, perform the analysis, or determine the scientific conclusions. All scientific content and interpretations were reviewed and verified by the authors, who take full responsibility for the publication.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank ESI Group for technical support with the PAM-STAMP software. This work was supported by the Innovate UK Smart Grant (Project Reference: 10083425) for Imperial College London and Multi-X Solution Limited, an Innovate UK KTP fund (10159356), and a start-up fund (OSR/0550/SASC/S022) for Dr Jichun Li from Newcastle University. For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this work.

Appendix A. Supplementary material-ID holdout evaluation

A.1. Material-ID overlap in the primary split

The primary evaluation uses non-overlapping geometry IDs across training, validation and testing, while material cards are drawn from the existing constitutive-response pool. Fig. A.25 audits the overlap of distinct material IDs in this original split. The steel training, validation and test subsets contain 600, 294 and 293 material IDs, respectively. Of these, 160 occur only in training, 147 occur in training and validation but not testing, 146 occur in training and testing but not validation, and 147 occur in all three subsets. For aluminium, all 110 material IDs occur in all three subsets. These counts refer to material cards rather than FEA cases. Thus, all test material IDs in the primary evaluation are represented in training, while their test geometries, and therefore the corresponding material–geometry pairs, are unseen during training.

A.2. Supplementary split and evaluation protocol

To examine predictions for material cards excluded from training, a focused steel-thinning experiment was conducted using a reduced training pool. This additional experiment is restricted to one material system and one output to limit the computational cost. The original geometry partition is retained: cases are filtered within their original subsets rather than transferred between them.

The reduced training subset retains original training cases associated with both the 160 train-only material IDs and the 147 IDs originally common to all three subsets, giving a pool of $160 + 147 = 307$ distinct training material IDs. To form the reduced training set, all original training cases associated with the 147 IDs originally shared by training and validation, but not testing, are removed from the original training set; their original validation cases form the supplementary validation subset. Likewise, all original training cases associated with the 146 IDs originally shared by training and testing, but not validation, are removed from the original training set when forming the reduced training set. Their 146 test cases form the held-out-material test subset. The remaining 147 test cases, whose material IDs occur in the reduced training pool, form the seen-material test subset. Thus, “seen” and “held-out” refer to material-card identity relative to the reduced training set; the geometries, and therefore the specific material–geometry pairs, in both test subsets remain unseen during training. The entire test set combines these two subsets,

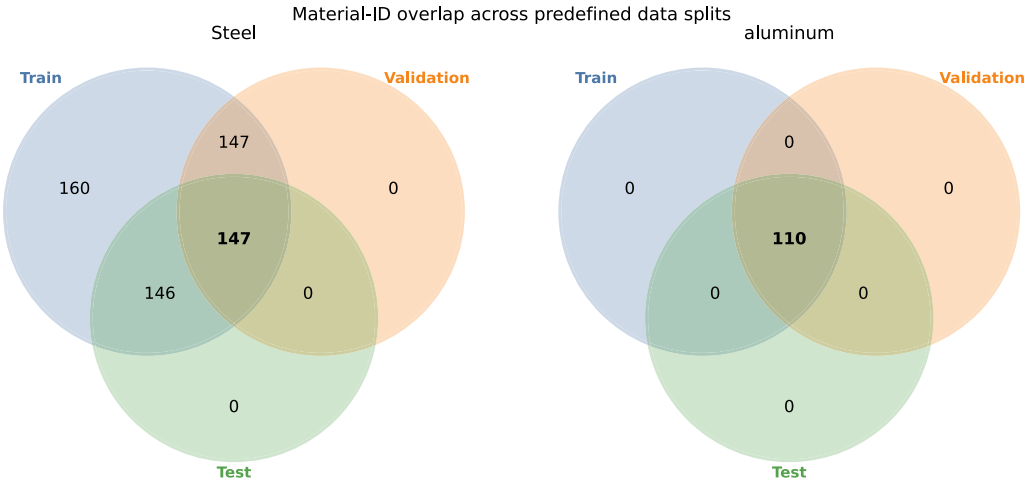


Fig. A.25. Distinct material-ID overlap in the original geometry-disjoint splits. Counts represent material cards, not FEA cases.

Table A.7
Steel-thinning performance in the supplementary material-ID holdout experiment. Seen-material and held-out-material test cases use material IDs included in and excluded from the reduced training set, respectively. The entire test set combines both subsets. Here, n denotes the number of test cases, MSE is evaluated over the valid masked domain, and P95 denotes the 95th percentile across cases.

Test subset	n	Masked MSE median	Masked MSE P95	RE median (%)	RE P95 (%)
Entire test set	293	4.3830×10^{-5}	1.1035×10^{-4}	6.59	15.97
Held-out-material cases	146	4.2980×10^{-5}	1.2227×10^{-4}	6.89	16.11
Seen-material cases	147	4.5816×10^{-5}	9.4797×10^{-5}	5.85	15.43

giving $146 + 147 = 293$ cases. The number of training material IDs is distinct from the number of retained material–geometry training cases.

Performance is evaluated using the masked MSE and representative-maximum RE defined in Section 5.2. Table A.7 reports their medians and 95th percentiles (P95) across the test cases, describing central performance and upper-tail error.

A.3. Results and interpretation

The held-out-material and seen-material test subsets have comparable masked-MSE medians, whereas the held-out-material subset has a higher P95 masked MSE. Relative to the seen-material subset, its median RE increases from 5.85% to 6.89%, and its P95 RE increases from 15.43% to 16.11%. These results indicate that the model can predict responses for material cards withheld from the reduced training pool, with less favourable upper-tail errors. Differences between the subsets may also reflect their geometry and response distributions, since the seen-material and held-out-material cases are not paired evaluations on identical geometries.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

[1] H. Blala, L. Lang, S. Khan, L. Li, S. Sijia, A. Guelailia, S.A. Slimane, S. Alexandrov, Forming challenges of small and complex fiber metal laminate parts in aerospace applications—a review, Int. J. Adv. Manuf. Technol. 126 (5) (2023) 2509–2543.
[2] W. Więckowski, M. Motyka, J. Adamus, P. Lacki, M. Dyner, Numerical and experimental analysis of titanium sheet forming for medical instrument parts, Materials 15 (5) (2022) 1735.
[3] H.R. Attar, H. Zhou, A. Foster, N. Li, Rapid feasibility assessment of components to be formed through hot stamping: a deep learning approach, J. Manuf. Process. 68 (2021) 1650–1671.
[4] A.K. Perka, M. John, U.B. Kuruveri, P.L. Menezes, Advanced high-strength steels for automotive applications: arc and laser welding process, properties, and challenges, Metals 12 (6) (2022) 1051.
[5] R. Chandel, N. Sharma, S.A. Bansal, A review on recent developments of aluminum-based hybrid composites for automotive applications, Emerg. Mater. 4 (5) (2021) 1243–1257.
[6] L. Hua, W. Zhang, H. Ma, Z. Hu, Investigation of formability, microstructures and post-forming mechanical properties of heat-treatable aluminum alloys subjected to pre-aged hardening warm forming, Int. J. Mach. Tools Manuf. 169 (2021) 103799.
[7] H.R. Attar, N. Li, A. Foster, A new design guideline development strategy for aluminium alloy corners formed through cold and hot stamping processes, Mater. Des. 207 (2021) 109856.
[8] S. Li, D. Zhou, A. Pan, Integrated lightweight optimisation design of wall thickness, material, and performance of automobile body side structure, Struct. Multidiscip. Optim. 67 (6) (2024) 95.
[9] I. Alawadhi, S. Ramnath, A. Bolar, Y. Fu, J.J. Shah, N. Zurbrugg, D. Detwiler, Structural design vs manufacturability costs of complex stamped components, in: Proc. Int. Des. Eng. Tech. Conf. Comput. Inf. Eng. Conf. (IDETC/CIE), vol. 88391, 2024, p. V005T05A008.
[10] D.J. Politis, A.D. Foster, N. Li, L. Wang, J. Lin, D. Szegda, Prediction of thinning behavior for complex-shaped, lightweight alloy panels formed through a hot stamping process, in: Proc. 2nd Int. Conf. Adv. High Strength Steel Press Hardening, World Scientific, 2016, pp. 395–401, https://doi.org/10.1142/9789813140622_0065
[11] H. Li, H. Zhou, N. Li, An integrated convolutional neural network-based surrogate model for crashworthiness performance prediction of hot-stamped vehicle panel components, in: MATEC Web Conf, vol. 401, 2024, 03013.
[12] H. Zhou, Y. Zhao, H. Li, T. Pfaff, N. Li, A multi-level graph-based surrogate model for real-time high-fidelity sheet forming simulations, Adv. Eng. Inform. 66 (2025) 103458.
[13] M. Cantamessa, F. Montagna, G. D'agnese, Data-driven innovation: challenges and insights of engineering design, Des. Sci. 10 (2024) e11.
[14] I.K. Nti, A.F. Adekoya, B.A. Weyori, O. Nyarko-Boateng, Applications of artificial intelligence in engineering and manufacturing: a systematic review, J. Intell. Manuf. 33 (6) (2022) 1581–1601.
[15] A.T.G. Tapeh, M.Z. Naser, Artificial intelligence, machine learning, and deep learning in structural engineering: a scientometrics review of trends and best practices, Arch. Comput. Methods Eng. 30 (1) (2023) 115–159.
[16] A. Ucar, M. Karakose, N. Kırımça, Artificial intelligence for predictive maintenance applications: key components, trustworthiness, and future trends, Appl. Sci. 14 (2) (2024) 898.
[17] N. Grover, AI-enabled supply chain optimization, Int. J. Adv. Res. Sci. Commun. Technol. (2025) 28–44.

[18] H. Taheri, A.S. Beni, Artificial intelligence, machine learning, and smart technologies for nondestructive evaluation, in: Handbook of Nondestructive Evaluation 4.0, Springer, 2025, pp. 853–881.
[19] Y. Kardovskyi, S. Moon, Artificial intelligence quality inspection of steel bars installation by integrating mask R-CNN and stereo vision, Autom. Constr. 130 (2021) 103850.
[20] J. Senoner, T. Netland, S. Feuerriegel, Using explainable artificial intelligence to improve process quality: evidence from semiconductor manufacturing, Manage. Sci. 68 (8) (2022) 5704–5723.
[21] S.J. Plathottam, A. Rzonca, R. Lakhnori, C.O. Iloeje, A review of artificial intelligence applications in manufacturing operations, J. Adv. Manuf. Process. 5 (3) (2023) e10159.
[22] H. Zhou, Q. Xu, Z. Nie, N. Li, A study on using image-based machine learning methods to develop surrogate models of stamp forming simulations, J. Manuf. Sci. Eng. 144 (2) (2022) 021012.
[23] H.R. Attar, A. Foster, N. Li, Development of a deep learning platform for sheet stamping geometry optimisation under manufacturing constraints, Eng. Appl. Artif. Intell. 123 (2023) 106295, https://doi.org/10.1016/j.engappai.2023.106295
[24] H.R. Attar, N. Li, Optimisation of deep drawn corners subject to hot stamping constraints using a novel deep-learning-based platform, IOP Conf. Ser.: Mater. Sci. Eng. 1238 (1) (2022) 012066, https://doi.org/10.1088/1757-899X/1238/1/012066
[25] H. Zhou, H. Li, Y. Zhao, P.R.N. Childs, N. Li, Image-based artificial intelligence-driven modelling for blank shape optimisation in sheet metal forming, Mater. Des. 256 (2025) 114302, https://doi.org/10.1016/j.matdes.2025.114302
[26] H. Attar, H. Zhou, N. Li, Deformation and thinning field prediction for HFQ formed panel components using convolutional neural networks, in: IOP Conf. Ser.: Mater. Sci. Eng, vol. 1157, IOP Publishing, 2021, 012079.
[27] Z. Ding, H.R. Attar, H. Wang, H. Liu, N. Li, Integrating convolutional neural network and constitutive model for rapid prediction of stress-strain curves in fibre reinforced polymers: a generalisable approach, Mater. Des. 241 (2024) 112849, https://doi.org/10.1016/j.matdes.2024.112849
[28] M. Cheng, X. Zhao, M. Dhimish, W. Qiu, S. Niu, A review of data-driven surrogate models for design optimisation of electric motors, IEEE Trans. Transp. Electrific. 10 (4) (2024) 8413–8431.
[29] R. Alizadeh, J.K. Allen, F. Mistree, Managing computational complexity using surrogate models: a critical review, Res. Eng. Des. 31 (3) (2020) 275–298.
[30] C. Ling, W. Kuo, M. Xie, An overview of adaptive-surrogate-model-assisted methods for reliability-based design optimisation, IEEE Trans. Reliab. 72 (3) (2022) 1243–1264.
[31] A. Hashemi, J. Jang, J. Beheshti, A machine learning-based surrogate finite element model for estimating dynamic response of mechanical systems, IEEE Access 11 (2023) 54509–54525.
[32] Y. Shi, Z. Lu, J. Zhou, E. Zio, A novel time-dependent system constraint boundary sampling technique for solving time-dependent reliability-based design optimisation problems, Comput. Methods Appl. Mech. Eng. 372 (2020) 113342.
[33] T. Hart-Rawung, J. Buhl, M. Bambach, A fast approach for optimisation of hot stamping based on machine learning of phase transformation kinetics, Procedia Manuf. 47 (2020) 707–712.
[34] D. Jankovič, M. Šimic, N. Herakovič, A data-driven simulation and Gaussian process regression model for hydraulic press condition diagnosis, Adv. Eng. Inform. 59 (2024) 102276.
[35] J. Kim, C. Lee, Prediction of turbulent heat transfer using convolutional neural networks, J. Fluid Mech. 882 (2020) A18.
[36] K. Ren, Y. Chew, Y.F. Zhang, J.Y.H. Fuh, G.J. Bi, Thermal field prediction for laser scanning paths in laser aided additive manufacturing by physics-based machine learning, Comput. Methods Appl. Mech. Eng. 362 (2020) 112734.
[37] Q. Tian, S. Guo, E. Melder, L. Bian, W.G. Guo, Deep learning-based data fusion method for in situ porosity detection in laser-based additive manufacturing, J. Manuf. Sci. Eng. 143 (4) (2021) 041011.
[38] Q. Xu, Z. Nie, H. Xu, H. Zhou, H.R. Attar, N. Li, F. Xie, X.-J. Liu, SuperMeshing: a new deep learning architecture for increasing the mesh density of physical fields in metal forming numerical simulation, J. Appl. Mech. 89 (1) (2022) 011002, https://doi.org/10.1115/1.4052195
[39] R. Azad, E.K. Aghdam, A. Rauland, Y. Jia, A.H. Avval, A. Bozorgpour, S. Karimijafarbigloo, J.P. Cohen, E. Adeli, D. Merhof, Medical image segmentation review: the success of U-Net, IEEE Trans. Pattern Anal. Mach. Intell. 46 (12) (2024) 10076–10095.
[40] Z. Luo, W. Yang, Y. Yuan, R. Gou, X. Li, Semantic segmentation of agricultural images: a survey, Inf. Process. Agric. 11 (2) (2024) 172–186.
[41] J.W. Kim, A.U. Khan, I. Banerjee, Systematic review of hybrid vision transformer architectures for radiological image analysis, J. Imaging Inform. Med. (2025) 1–15.
[42] X. Liu, C. Zhang, L. Zhang, Vision Mamba: a comprehensive survey and taxonomy, arxiv preprint arXiv:2405.04404, 2024.
[43] A. Kumar, S.P. Yadav, A. Kumar, An improved feature extraction algorithm for robust swin transformer model in high-dimensional medical image analysis, Comput. Biol. Med. 188 (2025) 109822.
[44] Z. Zhu, M. Sun, G. Qi, Y. Li, X. Gao, Y. Liu, Sparse dynamic volume TransUNet with multi-level edge fusion for brain tumor segmentation, Comput. Biol. Med. 172 (2024) 108284.
[45] D. Jain, R. Modzelewski, R. Hérault, C. Chatelain, E. Torfeh, S. Thureau, Differential-UMamba: rethinking tumor segmentation under limited data scenarios, arxiv preprint arXiv:2507.18177, 2025.
[46] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, M. Wang, Swin-unet: UNet-like pure transformer for medical image segmentation, in: Proc. Eur. Conf. Comput. vis. (ECCV), 2022, pp. 205–218.