

Unsupervised Single-Source Domain Generalization for Robust Quantification of Lymphatic Perfusion

No Author Given

No Email Given

Abstract. Infants and young children with Fontan circulation, resulting from life-saving heart surgery, experience altered hemodynamics that can lead to severe, life-threatening complications, often driven by congestion in the lymphatic system. These complications typically become more apparent as the patients grow older. Early detection and management of these complications require precise assessment of lymphatic perfusion through segmentation of lymphatic fluids. However, manual evaluation is time-intensive and subject to variability, underscoring the need for automated, reliable segmentation in clinical practice. Yet, domain shifts arising from variations in MRI acquisition protocols (*e.g.*, BLADE versus SPACE sequences) and scanner differences present significant challenges for current segmentation models. In this work, we propose the use of Causality-Inspired Single-Source Domain Generalization (CISDG) to develop a robust and accurate segmentation network for lymphatic perfusion patterns across diverse imaging domains. Using a dataset of T₂-weighted MR images from 71 patients, we demonstrate that our CISDG model outperforms conventional segmentation networks, including nnUNet, in both source and target domains. Our results indicate that the proposed method not only generalizes effectively across domain shifts but also holds promise for enhancing diagnostic efficiency and reliability in clinical settings.

1 Introduction

Congenital univentricular heart defects affect approximately 3 in 10,000 live births [1], necessitating a series of surgical interventions within the first years of life and resulting in the Fontan circulation. The significantly altered hemodynamics in patients with Fontan circulation, frequently lead to severe complications such as protein-losing enteropathy or plastic bronchitis, which are associated with congestion within the lymphatic system [2]. For early detection and effective management of these complications, regular monitoring through imaging of lymphatic perfusions is crucial. In this context a curated dataset was established by *Anonymised* and *Anonymised*, it includes T₂-weighted routine MRI scans of 71 Fontan patients, treated at the University Hospital in *Anonymised* between 2007 and 2020. We revised this dataset and further refined it through additional manual annotations, which were performed in collaboration with a pediatric cardiologist.

Domain shifts result from variations in imaging devices, protocols, and settings often occurring even within one institution. These present a significant challenge in medical image segmentation and can lead to performance decreases in the segmentation ability when insufficient data from the target domain is available during training. How domain shifts can impede a model’s performance can be seen in the variations between T₂-

weighted MRI images acquired with a BLADE sequence and with a SPACE sequence. Although the nnUNet by [3] performs well on the BLADE sequences within the dataset, it exhibits a significant performance drop when confronted with SPACE sequences. This appears due to the limited sample size of SPACE images and the intensity differences between the BLADE and SPACE images. However, domain shifts are not limited to differences between sequences, they can also occur in other contexts where imaging conditions vary. This highlights the necessity for a model that robustly and accurately segments lymphatic perfusions in images from various domains, without additional training images, which are often scarce due to privacy concerns, costs, and the labor-intensiveness of manual annotation. To find a solution that robustly segments lymphatic perfusion patterns across different domains, we explored the Causality-Inspired Single-Source Domain Generalization (CISDG) approach by [4]. Unlike multi-source domain generalization or conventional data augmentation, this approach does not require any target domain data during training and exclusively uses data from one domain for training. It focuses on learning domain-invariant features, that can be applied to unseen target domains. This approach has demonstrated promising results in segmenting abdominal cross-modality, cardiac cross-sequence, and prostate cross-site datasets. Nevertheless, the authors could not find an application in the context of lymphatic perfusion pattern segmentation, as proposed in this paper. The CISDG network outperformed the nnUNet on the limited target domain data available to us and achieved slightly better results than the nnUNet on the source domain. We propose extending the evaluation of the network across other domains, however, the results are providing a promising direction for future research.

2 Materials and Methods

Dataset. Since 2007, the University Hospital in *Anonymised* has implemented a specialized protocol for managing univentricular heart defects. This protocol incorporates a T_2 -weighted MR images with fat suppression, conducted six months post-Fontan completion, to evaluate the lymphatic perfusions.

By 2020 MR images had been acquired from 71 patients, forming the foundation for an underlying clinical trial, by *Anonymised*. For 67 patients, imaging was performed using the BLADE sequence, a motion-resistant, multi-shot turbo spin echo technique designed for breath-hold acquisition. BLADE incorporates intershot motion correction to effectively reduce artifacts

Tab. 1. Number of Volumes per Dataset subset for Training and for the full Dataset. Subset A contains only BLADE sequences, Subset B contains BLADE and SPACE.

Features		Subset A	Subset B	Dataset
Domain	BLADE	58	58	73
	SPACE	0	2	4
Region	Abdom.	5	5	7
	Full	48	50	63
	Thoracic	5	5	7
		0	0	4
Quality (n)	1	12	12	13
	2	27	27	36
	3	16	16	19
	4	3	3	5
Severity (n)	1	4	4	5
	2	23	24	27
	3	29	30	43
	4	1	1	1
		58	60	77

from sources such as intestinal peristalsis, residual respiratory movement, and cardiac pulsation. In the remaining four patients, the SPACE sequence was used—a single-slab, 3D turbo spin echo sequence that employs a slab-selective, variable excitation pulse to achieve high-resolution imaging of the target region. From these images, a curated dataset with pixel-wise manual annotations was curated by *Anonymised*, and serves as the foundational dataset for this paper. To enhance the dataset for this study, we conducted a comprehensive review and added detailed new annotations to improve data quality and segmentation accuracy.

For younger, smaller patients, a single scan is typically sufficient to capture the neck, chest, and abdomen. In taller, older patients, however, two separate scans are necessary: one covering the neck and chest, and the other focused on the abdomen. Within the foundational dataset from *Anonymised*, annotations in the thoracic scans were confined to thoracic lymphatic perfusions, while segmentations in the abdominal scans were limited to abdominal lymphatic perfusions. However, the abdominal or thoracic lymphatic perfusions might remain visible across these scans. Throughout the dataset refinement, precise pixel-wise annotations of these lymphatic perfusions were meticulously performed in collaboration with a pediatric cardiologist, utilizing *3D Slicer*, an open-source software tool. Given the similar high intensity of lymphatic perfusions and other tissues in T_2 -weighted images, it was crucial to consider anatomical structures to ensure correct identification and annotation. In certain images, the thoracic scan sufficiently covered the neck, chest, and abdomen. In such cases, comprehensive annotations were made within the thoracic scans, and the corresponding abdominal scans were exempt from training. Additionally, in the underlying database ratings in diverse categories were assigned to the images. In the dataset by *Anonymised*, each image received a severity rank on a scale from one to four, based on the assessments by two independent pediatric cardiologists, using the system introduced by [5, 6]. In this system, the thoracic, para-aortic, and portal-venous lymphatic perfusion patterns are individually ranked. For scans displaying multiple patterns, the mean of the individual scores was calculated. Furthermore, each scan was assigned a quality rating reaching from one, the best quality, to four, the lowest quality. Scans were also categorized by the domain and the annotated and displayed regions. Details on dataset distributions and the configuration of various data folds utilized during the training are provided in Tab. 1.

Causality-Inspired Single-Source Domain Generalization. In MR images, variations in image intensity occur due to differences in acquisition protocols and sequences. These acquisition shifts are particularly visible between images acquired with BLADE and SPACE sequences. Such differences pose significant challenges to standard segmentation models such as the nnUNet by [3]. To improve the model’s performance when exposed to acquisition shifts, [4] introduced a causality-inspired data augmentation approach. The workflow of the approach can be seen in figure 1, it is composed of two main techniques, global intensity non-linear augmentation (GIN) and interventional pseudo-correlation augmentation (IPA). GIN uses randomly weighted shallow convolutional networks to generate images with different appearances, thereby preventing the model from learning acquisition-related criteria. IPA autonomously resamples the appearance of possibly confounded objects, enhancing the model’s robustness against domain shifts. The efficiency of this model has been evaluated across several datasets, including an

abdominal cross-modality, a cardiac cross-sequence, and a prostate cross-site dataset, demonstrating superior performance to other state-of-the-art models. Considering the differences between BLADE and SPACE sequences as examples of acquisition shifts, and the robust performance of the model on diverse test data, we selected this model for our segmentation task. By using this advanced approach adapted for our dataset, we aim to address the challenges of domain shifts and enhance the accuracy of automated segmentation. To evaluate the model’s performance, we selected the 2D nnUNet by [3] as the state-of-the-art (SOTA) method. To ensure comparability, we adopted the same 2D configuration used in the domain generalization model. Furthermore, the 3D nnUNet did not yield improved performance in our previous work [*Anonymised*].

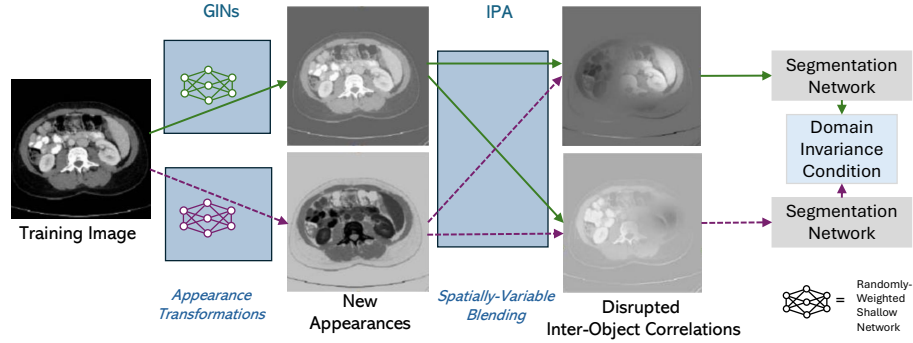


Fig. 1. Overview of the Causality-Inspired Single-Source Domain Generalization approach. During training, the input image is transformed using a Generative Invariant Network (GIN), producing a diverse set of appearances. These transformations are then refined through Invariant Perturbation Augmentation (IPA), which removes any residual inter-object perturbations. The resulting augmented images are used to train a segmentation network, equipping it to learn a domain-invariant representation for robust performance across diverse domains [4].

3 Results

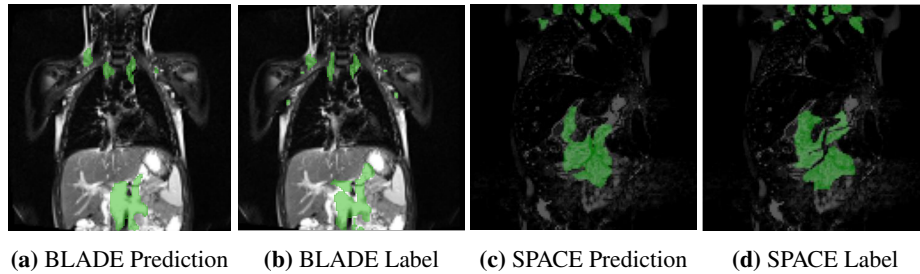


Fig. 2. Comparison of high-performing BLADE (Quality 2 and Severity 3) and SPACE (Quality 0 and Severity 3) full-body prediction and label by the CISDG model.

Figure 2 compares the label and the prediction by the CISDG model of an exemplary high-performing BLADE and SPACE scan. Tab. 2 shows the best checkpoint performance of the causality-inspired domain generalization (CISDG) network trained on Subset A compared to the performance of the 2d nnUNet trained on Subset A and B. On the BLADE sequence, the CISDG network marginally outperformed both nnUNet configurations by 5 % and 4 %. The nnUNet (A) failed to deliver satisfactory results on the SPACE images, whereas the nnUNet (B) showed 23 % better results. The CISDG model outperformed both nnUNets, by 13 % and 36 % on the SPACE domain. However, the CISDG network also experienced an 8 % inferior performance on the SPACE images compared to the BLADE images.

Tab. 2. Best checkpoint evaluation on the test set by T₂ sequence, best scores in bold, CISDG: causality-inspired single-source domain generalization network DSC: dice similarity coefficient, IoU: Intersection over Union, NSD: normalized surface distance

T ₂ Sequence	Measure	CISDG (A)	nnUNet (A)	nnUNet (B)
BLADE	DSC	0.45 ± 0.13	0.40 ± 0.17	0.41 ± 0.16
	IoU	0.30 ± 0.10	0.26 ± 0.14	0.27 ± 0.13
	NSD	0.68 ± 0.13	0.84 ± 0.06	0.85 ± 0.05
SPACE	DSC	0.37 ± 0.10	0.08 ± 0.09	0.24 ± 0.23
	IoU	0.23 ± 0.09	0.04 ± 0.05	0.16 ± 0.15
	NSD	0.49 ± 0.06	0.64 ± 0.06	0.66 ± 0.10

Additionally, the influence of the region, severity, and quality of the images on the model’s segmentation of the BLADE images was evaluated. The performance of the CISDG network, the nnUNet (A), and the nnUNet (B) were compared, the results are listed in Tab. 3. The nnUNet (A) achieved the best results on the thoracic images, followed by the abdominal images, and was the least accurate on whole-body images. Conversely, the CISDG model performed the best on whole-body images, followed by thoracic and abdominal images. In terms of quality, all networks demonstrated better performance on higher-quality images and performed the best on images with a severity rank of three. However, the CISDG model performed 15 % better than nnUNet (A) and 3 % better than nnUNet (B) on images with low severity, achieving similar results in this category as on images with severity rank three, and the nnUNet (A) performed 11 % better than the CISDG model and 3 % better than nnUNet (B) on images with quality four, even performing 16 % better than on images with quality 3. All models were least effective on images classified with a severity of two.

4 Discussion

For the BLADE sequence, the CISDG network demonstrated marginally superior performance over both nnUNets, indicating its robustness in handling images from the source domain. The CISDG model notably outperformed both nnUNets on the SPACE domain, despite being solely trained on BLADE sequences, this highlights the model’s generalization abilities across unknown domains. The nnUNet (A), which was only trained

Tab. 3. Average scores on the test set for the anatomical region, quality, and severity in the BLADE sequence, best scores in bold, CISDG: causality-inspired single-source domain generalization network DSC: dice similarity coefficient, IoU: Intersection over Union

Features		CISDG (A)		nnUNet (A)		nnUNet (B)	
		DSC	IoU	DSC	IoU	DSC	IoU
Region	Abdom.	0.39 $\pm$ 0.25	0.27 $\pm$ 0.20	0.46 $\pm$ 0.26	0.34 $\pm$ 0.23	0.45 $\pm$ 0.25	0.33 $\pm$ 0.22
	Full	0.46 $\pm$ 0.09	0.30 $\pm$ 0.08	0.37 $\pm$ 0.14	0.24 $\pm$ 0.11	0.39 $\pm$ 0.14	0.25 $\pm$ 0.10
	Thoracic	0.42 $\pm$ 0.00	0.27 $\pm$ 0.00	0.57 $\pm$ 0.00	0.40 $\pm$ 0.00	0.52 $\pm$ 0.00	0.35 $\pm$ 0.00
Quality	1	0.49 $\pm$ 0.00	0.33 $\pm$ 0.00	0.51 $\pm$ 0.00	0.34 $\pm$ 0.00	0.52 $\pm$ 0.00	0.35 $\pm$ 0.00
	2	0.49 $\pm$ 0.11	0.33 $\pm$ 0.09	0.45 $\pm$ 0.15	0.30 $\pm$ 0.13	0.45 $\pm$ 0.14	0.30 $\pm$ 0.12
	3	0.40 $\pm$ 0.05	0.25 $\pm$ 0.04	0.23 $\pm$ 0.12	0.14 $\pm$ 0.07	0.28 $\pm$ 0.13	0.17 $\pm$ 0.09
	4	0.28 $\pm$ 0.15	0.17 $\pm$ 0.10	0.39 $\pm$ 0.19	0.26 $\pm$ 0.14	0.36 $\pm$ 0.16	0.23 $\pm$ 0.12
Severity	1	0.47 $\pm$ 0.00	0.31 $\pm$ 0.00	0.32 $\pm$ 0.00	0.19 $\pm$ 0.00	0.44 $\pm$ 0.00	0.28 $\pm$ 0.00
	2	0.40 $\pm$ 0.14	0.26 $\pm$ 0.10	0.29 $\pm$ 0.16	0.18 $\pm$ 0.11	0.31 $\pm$ 0.15	0.19 $\pm$ 0.11
	3	0.47 $\pm$ 0.12	0.31 $\pm$ 0.10	0.49 $\pm$ 0.14	0.33 $\pm$ 0.12	0.48 $\pm$ 0.13	0.33 $\pm$ 0.11
	4	-	-	-	-	-	-

on BLADE images, failed on the SPACE domain, highlighting the limitations of the model and the necessity for a model with better generalization abilities. Incorporating a small number of SPACE samples into the training data, as in the data fold of the nnUNet (B), significantly improved the model’s performance, nevertheless, it did not reach the performance level of the CISDG network, highlighting that little data from the source domain can improve the performance but does not solve the problem of domain shifts.

A closer examination of the model’s performance on the BLADE images reveals that the CISDG model shows superior performance to both nnUNets on whole-body images, while both nnUNets perform better on thoracic and abdominal images and the nnUNets (A) achieved the best results in these categories. Due to the additional annotations, the employed dataset included only a limited number of thoracic images, while it included a considerable number of images depicting the entire body. The overall superior performance of the CISDG model over the nnUNets can be attributed to its ability to segment a greater number of whole-body images with greater precision. All networks improved with higher-quality images, indicating a connection between the quality and the performance. All models demonstrated the highest performance on images with severity level three, as the majority of training images were from category three the performance increase might not be directly related to the severity of the lymphatic perfusions but rather to the increased number of training samples. On top, the performance according to severity can not be fully assessed due to the very limited data and the absence of test data in the category of severity level four. The results highlight the great potential of the CISDG approach to enhance the generalization ability of a model. However, the sample size in the SPACE domain was very small, and further testing, with a larger sample size and data from other domains, would be necessary to fully assess the capabilities of the CISDG network in the context of lymphatic perfusion segmentation.

5 Conclusion

We demonstrated the effectiveness of the CISDG approach for robust lymphatic perfusion pattern segmentation in patients with congenital univentricular heart defects. Our findings indicate that the CISDG model outperforms traditional segmentation models, such as nnUNet, across diverse imaging protocols, including the SPACE and BLADE MRI sequences, highlighting its potential to address domain shifts. The model's ability to generalize without requiring target domain data provides a significant advantage in clinical settings where data from varied imaging protocols may be limited. Future work will focus on expanding this approach to larger and more diverse datasets, further validating CISDG's applicability and exploring its integration into routine diagnostic workflows to enhance diagnostic precision and reliability for patients with complex hemodynamic conditions.

References

1. Schwedler G, Lindinger A, Lange PE, Sax U, Olchvary J, Peters B et al. Frequency and spectrum of congenital heart defects among live births in Germany : a study of the Competence Network for Congenital Heart Defects. *Clinical research in cardiology : official journal of the German Cardiac Society*. 2011;100(12):1111–7.
2. Blum U, Meyer H, Beerbaum P, Peuster M. *Kompendium angeborene Herzfehler bei Kindern: Diagnose und Behandlung*. 2. Auflage. Berlin and Heidelberg: Springer, 2021.
3. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*. 2021;18(2):203–11.
4. Ouyang C, Chen C, Li S, Li Z, Qin C, Bai W et al. Causality-Inspired Single-Source Domain Generalization for Medical Image Segmentation. *IEEE transactions on medical imaging*. 2023;42(4):1095–106.
5. Schroeder C, Moosmann J, Cesnjevar R, Purbojo A, Rompel O, Dittrich S. A classification of abdominal lymphatic perfusion patterns after Fontan surgery. *European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery*. 2022;62(4).
6. Biko DM, DeWitt AG, Pinto EM, Morrison RE, Johnstone JA, Griffis H et al. MRI Evaluation of Lymphatic Abnormalities in the Neck and Thorax after Fontan Surgery: Relationship with Outcome. *Radiology*. 2019;291(3):774–80.