Addressing Label Scarcity and Domain Shift in Medical Image Segmentation

Suruchi Kumari and Pravendra Singh*

Indian Institute of Technology Roorkee

Abstract. Limited labeled data and domain shifts present significant challenges for accurate medical image segmentation. Semi-supervised learning (SSL) and unsupervised domain adaptation (UDA) methods address these challenges individually. Existing SSL methods do not perform well in UDA scenarios, and vice versa. We observe that excelling in SSL requires effective learning from limited labeled data while avoiding overfitting, whereas in UDA, the domain gap must be effectively reduced. To design a novel unified framework that tackles both the scarcity of labeled data and domain shift, it is essential to address both objectives. To accomplish this, we introduce Wavelet Frequency Exchange (WFE), which decomposes encoder features into low and high-frequency components and exchanges high-frequency features between labeled and unlabeled data. WFE provides two key benefits: it disrupts overfitting by preventing the model from memorizing details from limited labeled data in SSL, and it reduces the domain gap in UDA. To improve the representation of exchanged features, we propose a Learnable Parametric Feature Network (LPFN), which includes downsampling and upsampling blocks. These blocks include Parametric Spline (PS) layers, which map the relationships between the exchanged features using a spline function. Evaluations on two publicly available medical datasets demonstrate the effectiveness of our method.

Keywords: Semi-supervised learning · Unsupervised domain adaptation · Wavelet Frequency Exchange · Medical image segmentation.

1 Introduction

Deep neural networks have significantly advanced the field of medical image analysis (MIA). Most leading techniques in MIA rely on supervised learning, which requires labeled data. However, labeling medical data is both costly due to the need for expert manual annotation and time-consuming, especially for tasks such as segmentation that require per-pixel labeling [5, 18]. On the other hand, obtaining unlabeled data is more feasible. As a result, semi-supervised learning (SSL) has emerged as a valuable paradigm, capable of achieving performance comparable to supervised learning at a lower cost. In SSL, a small amount of labeled data is used alongside a large volume of unlabeled data [10,

---

* Corresponding author: pravendra.singh@cs.iitr.ac.in

11]. Several SSL techniques have been developed for medical image segmentation. Pseudo-labeling [4, 20] and consistency regularization methods [1] are commonly applied in this context. Bai et al. [1] use bidirectional copy-pasting between labeled and unlabeled data to effectively enhance learning from labeled samples. VCLIPSeg [13] integrates CLIP embeddings in a voxel-wise manner to learn semantic relationships among pixels. However, these methods assume that labeled and unlabeled data come from the same distribution. In practice, medical images collected from different medical centers using various scanning systems often exhibit significant domain shifts [12]. As a consequence, current approaches have limited applicability in real-world scenarios.

To address the challenges of domain shift, researchers have increasingly adopted unsupervised domain adaptation (UDA) techniques. UDA aims to bridge the gap between labeled (source domain) and unlabeled data (target domain) [15]. Previously, SSL and UDA were addressed as distinct challenges. However, since both frameworks involve learning from both labeled and unlabeled data, it is natural to develop a unified solution that effectively manages both scenarios. Existing SSL approaches are not well-suited for UDA, and vice versa [23, 21]. Developing an architecture that performs effectively well in both situations is a substantial challenge. Wang et al. [23] developed a framework for volumetric SSL that addresses both SSL and UDA tasks through aggregation and decoupling framework.

To design a unified framework that effectively handles both scarcity of labeled data and domain shift, we must address two key challenges: (1) Overfitting with limited labeled data: SSL relies on a small set of labeled data. However, with limited ground truths available, models tend to overfit to this labeled set, hindering generalization. (2) Minimizing the domain gap in UDA: Reducing the domain gap is a non-trivial task, especially when there is a significant domain shift, such as between MRI and CT modalities. If we can develop a framework that simultaneously tackles both challenges, we will achieve a unified solution that enhances performance in both SSL and UDA settings.

To achieve both objectives, we introduce Wavelet Frequency Exchange (WFE). WFE decomposes the encoder features into low (LF) and high-frequency (HF) coefficients using discrete wavelet transform (DWT). The Inverse Discrete Wavelet Transform (IDWT) is then used to reconstruct the spatial features from their wavelet coefficients. During the reconstruction step, the low-frequency coefficients are preserved, while the high-frequency coefficients are exchanged between the labeled and unlabeled features. In this way, WFE effectively mixes the high-frequency features of the labeled and unlabeled data. To further improve the representation of exchanged features, we propose the learnable Parametric Feature Network (LPFN), which includes LPFN downsampling (LPFN-DB) and upsampling (LPFN-UB) blocks. LPFN-DB efficiently captures relationships between the exchanged features by first converting the features into patches, then modeling these relationships using Parametric Spline (PS) layers. During downsampling, it progressively captures features at different scales, while LPFN-UB

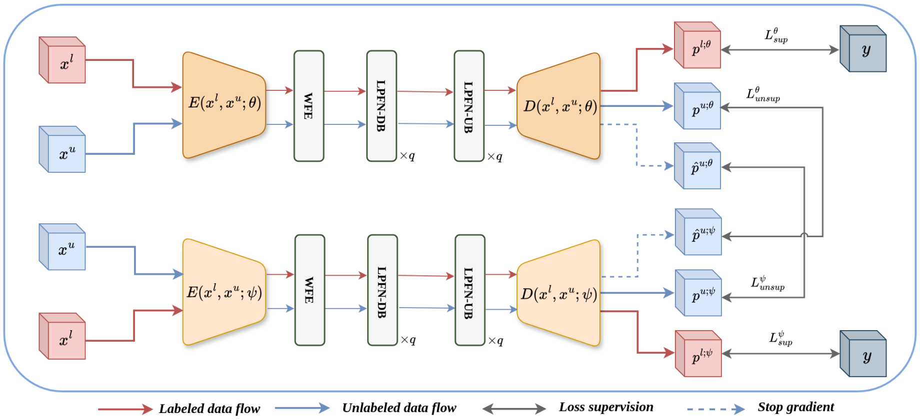

Fig. 1: Overview of our proposed framework: We utilize a co-training framework to learn from both labeled and unlabeled data. Each model processes images that have undergone distinct augmentation methods.

reconstructs spatial details using PS layers. This strategy enhances the overall representation of the exchanged features.

Our proposed approach achieves the first objective by incorporating high-level features from unlabeled data into the labeled data. Overfitting often occurs in SSL when the model memorizes specific details of the limited labeled data. By exchanging high-frequency components, we disrupt this direct memorization of intricate details, forcing the model to focus on broader, more generalizable patterns rather than on the specific features of the labeled data. We achieve the second objective by leveraging the fact that high-frequency features often capture domain-specific details, such as edges, textures, and fine structures. Swapping these features forces the model to learn representations that are consistent across both domains. As a result, the model becomes less sensitive to the specific characteristics of either domain and more focused on shared, generalizable features.

2 Methodology

2.1 Preliminaries

In this problem, we are provided with two datasets: labeled data $D_l = \{(x_i^l, y_i)\}_{i=1}^N$ and unlabeled data $D_u = \{x_i^u\}_{i=1}^M$. Here, N indicates the total number of labeled images, while M denotes the total number of unlabeled images. Each image x_i belongs to $\mathbb{R}^{H \times W \times D}$, and each corresponding label y_i is represented in $\mathbb{R}^{H \times W \times D \times C}$, where C is the number of possible classes. In SSL, both labeled and unlabeled data originate from the same domain, whereas in UDA, the labeled data and unlabeled data come from different domains. Our objective is to develop a deep learning model $f(.)$ that effectively utilizes both labeled and unlabeled data to address both scenarios in a unified approach.

2.2 Learning Framework and Loss Functions

Our method consists of two models with the same architecture, $f(x^l, x^u, \theta)$ and $f(x^l, x^u, \psi)$. To enhance diversity between the models, each receives images that have undergone distinct augmentation methods. Both models generate predictions for labeled and unlabeled data and are trained using both supervised and unsupervised losses, as shown in Figure 1. For both kinds of losses, we employ conventional dice and cross-entropy loss functions. The supervised loss for both models is formulated as follows:

$$L^{\theta}_{\text{sup}} = \frac{1}{N}\sum_{i=1}^{N} \mathcal{L}_{\text{DiceCE}}(p_i^{l;\theta}, y_i), \quad L^{\psi}_{\text{sup}} = \frac{1}{N}\sum_{i=1}^{N} \mathcal{L}_{\text{DiceCE}}(p_i^{l;\psi}, y_i) \tag{1}$$

where $\mathcal{L}_{\text{DiceCE}}(x, y) = \frac{1}{2}\left(\mathcal{L}_{\text{dice}}(x, y) + \mathcal{L}_{\text{CE}}(x, y)\right)$ represents the Dice loss and cross-entropy loss, respectively. $p^{l;\theta}$ and $p^{l;\psi}$ are the predictions for labeled data from the models $f(\cdot;\theta)$ and $f(\cdot;\psi)$, respectively.

For unlabeled data we utilize the cross-pseudo supervision [4] approach where each model generates pseudo-labels for the unlabeled data and then uses these pseudo-labels to provide supervision to the other model. The prediction of the model $f(\cdot;\theta)$ for an unlabeled image is denoted by $p^{u;\theta}$, and the pseudo-label it produces is given by $\hat{p}^{u;\theta} = \text{argmax}(p^{u;\theta})$, where argmax returns the index of the maximum value. Similarly, the prediction of the model $f(\cdot;\psi)$ for unlabeled data is denoted by $p^{u;\psi}$, and its pseudo-label is $\hat{p}^{u;\psi}$. The unsupervised loss for both models is formulated as follows:

$$L^{\theta}_{\text{unsup}} = \frac{1}{M}\sum_{i=1}^{M} \mathcal{L}_{\text{DiceCE}}(p_i^{u;\theta}, \hat{p}_i^{u;\psi}), \quad L^{\psi}_{\text{unsup}} = \frac{1}{M}\sum_{i=1}^{M} \mathcal{L}_{\text{DiceCE}}(p_i^{u;\psi}, \hat{p}_i^{u;\theta}) \tag{2}$$

2.3 Framework Overview

Our model is based on an encoder-decoder convolutional neural network architecture. Specifically, the function $f(\cdot)$ can be decomposed into two components: $E(\cdot)$, the encoder, and $D(\cdot)$, the decoder. The encoder $E(\cdot)$ is responsible for extracting hierarchical features from the input data, while the decoder $D(\cdot)$ reconstructs the output from the compressed features produced by the encoder. The feature maps generated by the encoder's last convolutional block for labeled data, denoted as z^l, and for unlabeled data, denoted as z^u, are utilized in the Wavelet Frequency Exchange (WFE) process, as shown in Figure 2.

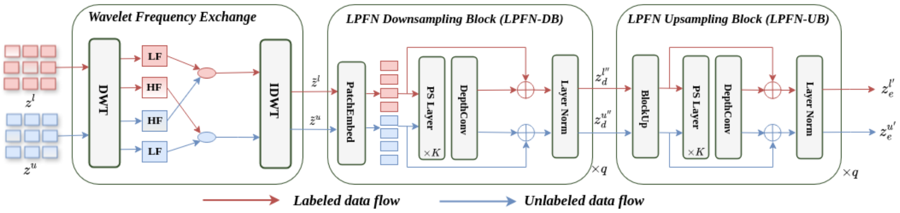

Fig. 2: Detailed architecture of our framework

2.4 Wavelet-Based Frequency Exchange

WFE uses the discrete wavelet transform (DWT) with the Haar wavelet to decompose the encoder features into low- and high-frequency coefficients. For a 3D feature z, this process yields eight distinct frequency coefficients, as illustrated below:

$$LLL,\ \{LLH, LHL, LHH, HLL, HLH, HHL, HHH\} = \mathrm{DWT}(z) \tag{3}$$

The low-frequency component LLL represents the smoothed, low-frequency approximation of the 3D features, capturing general patterns and structures. In contrast, the remaining seven high-frequency (HF) coefficients capture detailed information across various dimension combinations, emphasizing finer features and edges in the data. First, the labeled and unlabeled features, z^l and z^u, are passed through the DWT function, as shown in Equations 4. This process decomposes each feature into its LF and HF coefficients. The IDWT is then used to reconstruct the spatial features from their wavelet coefficients. During the reconstruction step, the LF coefficients are preserved, while the HF coefficients of the labeled features are exchanged with those of the unlabeled features, and vice versa, as illustrated in Equations 5.

$$LF^l, HF^l = \mathrm{DWT}\left(z^l\right), \qquad LF^u, HF^u = \mathrm{DWT}\left(z^u\right) \tag{4}$$

$$\tilde{z}^l = \mathrm{IDWT}\left(LF^l, HF^u\right), \qquad \tilde{z}^u = \mathrm{IDWT}\left(LF^u, HF^l\right) \tag{5}$$

The modified labeled $\tilde{z}^l$ and unlabeled features $\tilde{z}^u$ are then passed to the LPFN translation block for further processing.

2.5 Learnable Parametric Feature Network

The Learnable Parametric Feature Network (LPFN) consists of two LPFN blocks: the LPFN Downsampling Block (LPFN-DB) and the LPFN Upsampling Block (LPFN-UB). LPFN-DB performs spatial downsampling of the feature maps, reducing their resolution while preserving essential information. LPFN-UB increases the spatial resolution of the feature maps, aiding in the reconstruction of high-resolution outputs. For brevity, we have omitted the labeled (l) and unlabeled (u) superscripts from this point onward.

LPFN Downsampling Block: LPFN-DB consists of three major components: the Patch Embedding operation (PatchEmbed); Parametric Spline (PS) layer; and the Depthwise Convolution layer (DepthConv). The PatchEmbed operation reshapes the wavelet-exchanged features $\tilde{z}$ into sequences of flattened 2D patches. Each patch is then linearly projected to a new space, which can be effectively implemented as a convolution operation with a kernel size matching the patch size [26]. By dividing an image into patches, the model can focus on capturing local patterns and textures, which aids in learning fine-grained details. The output obtained from PatchEmbed operation z_d is passed to PS layer, as shown in Equation 6.

The Parametric Spline (PS) layer in the LPFN network transforms input features by using a combination of two components: a residual function and a spline-based function. More specific details of the parametric spline layer are provided in Section 2.6. The output of PS layer is passed to DepthConv module. DepthConv [2] is utilized to perform convolution independently on each channel of the feature maps. Consequently, the number of parameters and computations is significantly reduced. The overall output of the LPFN-DB is formalized as follows:

$$z_d = \text{PatchEmbed}(\tilde{z}) \tag{6}$$

$$z'_d = \text{DepthConv}(\text{PS}(z_d), \ldots, K) \tag{7}$$

$$z''_d = z_d + z'_d \tag{8}$$

The Equation 8 represents the residual connection, where the output of the PatchEmbed operation z_d is added to the output of the combined PS layers and DepthConv outputs z'_d. This residual connection helps preserve the original information and facilitates better gradient flow during training. Finally, z''_d is passed to Layer Normalization (LN) to stabilize and accelerate training, as shown in Figure 2. The symbol K represents the number of PS layers. We will obtain two sets of features from the LPFN-DB block: labeled features $z^{l''}_d$ and unlabeled features $z^{u''}_d$.

LPFN Upsampling Block: The LPFN-UB consists three major components: the upsampling operation Upsample ; the PS layer; the DepthConv layer (DepthConv).

$$z_e = \text{BlockUp}(z''_d) \tag{9}$$

$$z'_e = \text{LN}\left(\text{DepthConv}(\text{PS}(z_e), \ldots, K)\right) \tag{10}$$

The Upsample operation (BlockUp) enlarges the feature map z''_d to a higher resolution. The skip connections are also integrated in this block, it introduces high-resolution details from earlier layers of LPFN-DB. Finally, the LPFN layer

Table 1: Comparison of the proposed approach with other approaches on the LA dataset using 5 and 10 percent labeled data.

Method	Scans used		Metrics				Scans used		Metrics			
	Labeled	Unlabeled	Dice↑	Jaccard↑	95HD↓	ASD↓	Labeled	Unlabeled	Dice↑	Jaccard↑	95HD↓	ASD↓
VNet	4(5%)	0	52.55	39.60	47.05	9.87	8(10%)	0	82.74	71.72	13.35	3.26
UA-MT [28]	4(5%)	76(95%)	82.26	70.98	13.71	3.82	8(10%)	72(90%)	87.79	78.39	8.68	2.12
DTC [17]			81.25	69.33	14.90	3.99			87.51	78.17	8.23	2.36
SASSNet [14]			81.60	69.63	16.16	3.58			87.54	78.05	9.84	2.59
MC-Net [25]			83.59	72.36	14.07	2.70			87.62	78.25	10.03	1.82
SS-Net [24]			86.33	76.15	9.97	2.31			88.55	79.62	7.49	1.90
BCP [1]			88.02	78.72	7.90	2.15			89.62	81.31	6.81	1.76
MLRPL [19]			-	-	-	-			89.86	81.68	6.91	1.85
VCLIPSeg [13]			88.51	79.49	10.48	2.28			90.59	82.87	6.22	1.61
Genericssl [23]			89.93	81.82	5.25	1.86			90.31	82.40	5.55	1.64
Ours			90.62	82.63	5.16	1.64			91.53	84.40	4.82	1.63

and DepthConv operation, followed by Layer Normalization, process the feature map to capture significant features and stabilize training. We will obtain two sets of features after the LPFN-UB block: labeled features $z_e^{l'}$ and unlabeled features $z_e^{u'}$, as shown in Figure 2.

2.6 Parametric Spline Layer

For an input x, the parametric spline function is expressed as:

$$PS(x) = (\Phi_{K-1} \circ \Phi_{K-2} \circ \cdots \circ \Phi_0)(x)$$

In this structure, each layer applies a transformation, Φ_k, to the input x_k to generate the output for the next layer, x_{k+1}, which can be written as:

$$x_{k+1} = \Phi_l(x_k) = \begin{pmatrix} \phi_{k,1,1}() & \cdots & \phi_{k,1,n_k}() \\ \vdots & \ddots & \vdots \\ \phi_{k,n_{k+1},1}() & \cdots & \phi_{k,n_{k+1},n_k}() \end{pmatrix} x_k$$

Each activation function ϕ is represented by a spline [16], enabling a flexible and expressive modeling capability for the inputs.

Inspired by [16], we utilize parametric spline function with B-splines. The approach involves incorporating residual activation functions, where the activation function $\phi(x)$ is represented as:

$$\phi(x) = w_b b(x) + w_s \text{spline}(x)$$

Specifically, $b(x)$ is chosen as $\text{silu}(x) = \frac{x}{1+e^{-x}}$, while $\text{spline}(x)$ is defined as a linear combination of B-splines with trainable coefficients c_i, as shown below:

$$\text{spline}(x) = \sum_i c_i B_i(x)$$

In our method, we further employ exponential moving average (EMA) on the weights w_b, w_s and c_i, for smoother gradient flow.

Table 2: Comparison of the proposed method with other approaches on the MMWHS dataset for the MR to CT and CT to MR setting.

	MR To CT						CT To MR					
Method	Dice ↑					ASD ↓	Dice ↑					ASD ↓
	AA	LAC	LVC	MYO	Avg	Avg	AA	LAC	LVC	MYO	Avg	Avg
PnP-AdaNet [6]	74.0	68.9	61.9	50.8	63.9	12.8	43.7	68.9	61.9	50.8	63.9	8.9
AdaOutput [22]	65.2	76.6	54.4	43.6	59.9	9.6	60.8	39.8	71.5	35.5	51.9	5.7
CycleGAN [29]	73.8	75.7	52.3	28.7	57.6	10.8	64.3	30.7	65.0	43.0	50.7	6.6
CyCADA [8]	72.9	77.0	62.4	45.3	64.4	9.4	60.5	44.0	77.6	47.9	57.5	7.9
SIFA [3]	81.3	79.5	73.8	61.6	74.1	7.0	65.3	62.3	78.9	47.3	63.4	5.7
DSAN [7]	79.9	84.8	82.8	66.5	78.5	5.9	71.3	66.2	76.2	52.1	66.5	5.4
LMISA-3D [9]	84.5	82.8	88.6	70.1	81.5	2.3	60.7	72.4	86.2	64.1	70.8	3.6
Genericssl [23]	93.2	89.5	91.7	86.2	90.1	1.7	62.8	87.4	61.3	74.1	71.4	7.9
Ours	92.3	93.5	91.7	92.6	92.5	1.4	69.8	77.3	73.9	76.6	74.4	6.1

3 Experiments

Datasets. We conducted experiments on a semi-supervised benchmark dataset, namely Left Atrium [27], and a domain adaptation dataset, namely MMWHS [30], under two scenarios: CT to MR and MR to CT. The LA dataset consists of 100 scans, with 80 scans used for training and 20 for evaluation. The labeled dataset is used for validation during training to select the optimal model. The MMWHS dataset contains CT and MRI images of the same anatomical regions, with 20 images for each modality.

Comparison with SOTA Methods for SSL and UDA To thoroughly assess our approach in tackling SSL, we compare it against several state-of-the-art (SOTA) methods, as presented in Tables 1. Our method consistently achieves superior performance across different SSL datasets. Likewise, for UDA, we evaluate our approach against SOTA techniques. As shown in Tables 2, our method surpasses existing approaches.

Architecture Ablation. In this section, we highlight the significance of different architectures. The first row in Table 3 presents the results of the co-training approach, as illustrated in Figure 1, which excludes the WFE and LPFN blocks. We then incorporate the WFE, with the features obtained from this stage being processed through a single convolution block before being fed into the decoder *WFE-a*. Next, we implement the *WFE-ConvBlock*, which utilizes a convolutional translation block similar to the LPFN translation block, except that the LPFN layers are replaced with convolutional layers. Finally, we present the results achieved with the LPFN translation block, which shows a 19.04 % improvement over the co-training framework.

4 Conclusion

In this work, we propose a novel unified approach to address the challenges of limited labeled data and domain shift. Our approach introduces the WFE

Table 3: Ablation study on using different architectures in our approach over the MMWHS dataset in the MR-to-CT setting.

Co-training	WFE-a	WFE-ConvBlock	WFE-LPFN Block	Dice ↑	ASD ↓
✓				73.49	6.72
✓	✓			80.16	4.32
✓		✓		87.32	2.92
✓			✓	92.53	1.41

block, which decomposes encoder features into low and high-frequency coefficients and exchanges HF coefficients between labeled and unlabeled data. To effectively process these exchanged features, we introduce LPFN, which include both downsampling and upsampling LPFN blocks. These blocks capture non-linear patterns within the data using LPFN layers, which are formed by stacking LPFNs with trainable activation functions and weights updated via exponential moving averages. Our method mitigates overfitting by preventing the model from memorizing specific details in SSL and reduces the domain gap in UDA, resulting in improved performance in both scenarios.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Bai, Y., Chen, D., Li, Q., Shen, W., Wang, Y.: Bidirectional copy-paste for semi-supervised medical image segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11514–11524 (2023)
2. Cao, J., Li, Y., Sun, M., Chen, Y., Lischinski, D., Cohen-Or, D., Chen, B., Tu, C.: Do-conv: Depthwise over-parameterized convolutional layer. IEEE Transactions on Image Processing **31**, 3726–3736 (2022)
3. Chen, C., Dou, Q., Chen, H., Qin, J., Heng, P.A.: Synergistic image and feature adaptation: Towards cross-modality domain adaptation for medical image segmentation. In: Proceedings of the AAAI conference on artificial intelligence. vol. 33, pp. 865–872 (2019)
4. Chen, X., Yuan, Y., Zeng, G., Wang, J.: Semi-supervised semantic segmentation with cross pseudo supervision. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 2613–2622 (2021)
5. Chen, X., Wang, X., Zhang, K., Fung, K.M., Thai, T.C., Moore, K., Mannel, R.S., Liu, H., Zheng, B., Qiu, Y.: Recent advances and clinical applications of deep learning in medical image analysis. Medical image analysis **79**, 102444 (2022)
6. Dou, Q., Ouyang, C., Chen, C., Chen, H., Glocker, B., Zhuang, X., Heng, P.A.: Pnp-adanet: Plug-and-play adversarial domain adaptation network at unpaired cross-modality cardiac segmentation. IEEE Access **7**, 99065–99076 (2019)
7. Han, X., Qi, L., Yu, Q., Zhou, Z., Zheng, Y., Shi, Y., Gao, Y.: Deep symmetric adaptation network for cross-modality medical image segmentation. IEEE transactions on medical imaging **41**(1), 121–132 (2021)

8. Hoffman, J., Tzeng, E., Park, T., Zhu, J.Y., Isola, P., Saenko, K., Efros, A., Darrell, T.: Cycada: Cycle-consistent adversarial domain adaptation. In: International conference on machine learning. pp. 1989–1998. Pmlr (2018)
9. Jafari, M., Francis, S., Garibaldi, J.M., Chen, X.: Lmisa: A lightweight multi-modality image segmentation network via domain adaptation using gradient magnitude and shape constraint. Medical Image Analysis **81**, 102536 (2022)
10. Jiao, R., Zhang, Y., Ding, L., Xue, B., Zhang, J., Cai, R., Jin, C.: Learning with limited annotations: a survey on deep semi-supervised learning for medical image segmentation. Computers in Biology and Medicine p. 107840 (2023)
11. Kumari, S., Singh, P.: Data efficient deep learning for medical image analysis: A survey. arXiv preprint arXiv:2310.06557 (2023)
12. Kumari, S., Singh, P.: Deep learning for unsupervised domain adaptation in medical imaging: Recent advancements and future perspectives. Computers in Biology and Medicine p. 107912 (2023)
13. Li, L., Lian, S., Luo, Z., Wang, B., Li, S.: Vclipseg: Voxel-wise clip-enhanced model for semi-supervised medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 692–701. Springer (2024)
14. Li, S., Zhang, C., He, X.: Shape-aware semi-supervised 3d semantic segmentation for medical images. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part I 23. pp. 552–561. Springer (2020)
15. Liu, X., Yoo, C., Xing, F., Oh, H., El Fakhri, G., Kang, J.W., Woo, J., et al.: Deep unsupervised domain adaptation: A review of recent advances and perspectives. APSIPA Transactions on Signal and Information Processing **11**(1) (2022)
16. Liu, Z., Wang, Y., Vaidya, S., Ruehle, F., Halverson, J., Soljačić, M., Hou, T.Y., Tegmark, M.: Kan: Kolmogorov-arnold networks. arXiv preprint arXiv:2404.19756 (2024)
17. Luo, X., Chen, J., Song, T., Wang, G.: Semi-supervised medical image segmentation through dual-task consistency. In: Proceedings of the AAAI conference on artificial intelligence. vol. 35, pp. 8801–8809 (2021)
18. Shen, W., Peng, Z., Wang, X., Wang, H., Cen, J., Jiang, D., Xie, L., Yang, X., Tian, Q.: A survey on label-efficient deep image segmentation: Bridging the gap between weak supervision and dense prediction. IEEE transactions on pattern analysis and machine intelligence **45**(8), 9284–9305 (2023)
19. Su, J., Luo, Z., Lian, S., Lin, D., Li, S.: Mutual learning with reliable pseudo label for semi-supervised medical image segmentation. Medical Image Analysis p. 103111 (2024)
20. Suruchi, K., Singh, P.: Leveraging fixed and dynamic pseudo-labels in cross-supervision framework for semi-supervised medical image segmentation. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 3491–3501 (2025)
21. Suruchi, K., Singh, P.: A unified degradation-robust approach to ssl and uda for 3d medical images. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 39, pp. 4428–4436 (2025)
22. Tsai, Y.H., Hung, W.C., Schulter, S., Sohn, K., Yang, M.H., Chandraker, M.: Learning to adapt structured output space for semantic segmentation. In: CVPR. pp. 7472–7481 (2018)
23. Wang, H., Li, X.: Towards generic semi-supervised framework for volumetric medical image segmentation. Advances in Neural Information Processing Systems **36** (2024)

24. Wu, Y., Wu, Z., Wu, Q., Ge, Z., Cai, J.: Exploring smoothness and class-separation for semi-supervised medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 34–43. Springer (2022)
25. Wu, Y., Xu, M., Ge, Z., Cai, J., Zhang, L.: Semi-supervised left atrium segmentation with mutual consistency training. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part II 24. pp. 297–306. Springer (2021)
26. Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J.M., Luo, P.: Segformer: Simple and efficient design for semantic segmentation with transformers. Advances in neural information processing systems **34**, 12077–12090 (2021)
27. Xiong, Z., Xia, Q., Hu, Z., Huang, N., Bian, C., Zheng, Y., Vesal, S., Ravikumar, N., Maier, A., Yang, X., et al.: A global benchmark of algorithms for segmenting the left atrium from late gadolinium-enhanced cardiac magnetic resonance imaging. Medical image analysis **67**, 101832 (2021)
28. Yu, L., Wang, S., Li, X., Fu, C.W., Heng, P.A.: Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II 22. pp. 605–613. Springer (2019)
29. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision. pp. 2223–2232 (2017)
30. Zhuang, X., Shen, J.: Multi-scale patch and multi-modality atlases for whole heart segmentation of mri. Medical image analysis **31**, 77–87 (2016)