

Article

Not peer-reviewed version

Attention-Guided Semi-Supervised Segmentation for Low-Contrast Medical Images Using Frequency Domain Priors

Abhishek Singh , Mehakpreet Singh , Tarush Tewari , [Kaneisha Bali Chibber](#) * , [Shamla Mantri](#)

Posted Date: 28 April 2026

doi: 10.20944/preprints202604.1945.v1

Keywords: attention mechanism; CLAHE; frequency domain learning; medical image segmentation; semi-supervised learning; transformer networks



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC, OpenAlex.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Attention-Guided Semi-Supervised Segmentation for Low-Contrast Medical Images Using Frequency Domain Priors

Abhishek Singh, Mehakpreet Singh, Tarush Tewari, Kaneisha Bali Chibber * and Shamla Mantri

Department of Computer Science, MIT-WPU, Pune, India

* Correspondence: 1032222403@mitwpu.edu.in

Abstract

Medical image segmentation under limited annotation budgets remains a critical challenge, particularly for anatomical structures exhibiting low-contrast boundaries—a pervasive problem in cardiac MRI, dermoscopy, and ultrasound imaging. Existing semi-supervised methods leverage consistency regularization and pseudo-label strategies but typically treat image preprocessing as a static pipeline step, failing to exploit the rich structural priors encoded in frequency-enhanced representations. In this paper, we propose FDPSeg, a novel Frequency-Domain Prior guided semi-supervised Segmentation framework that integrates Contrast-Limited Adaptive Histogram Equalization (CLAHE)-derived frequency priors directly into the transformer attention mechanism. Our dual-branch encoder processes both the original image and its CLAHE-enhanced counterpart, with a learned channel-attention fusion module that adaptively weights spatial and frequency features per image region. A novel frequency-domain consistency loss enforces structural coherence between teacher and student networks in the Fourier space, providing stronger supervisory signal for unlabeled data than spatial consistency alone. Experiments on the ACDC cardiac MRI dataset, the ISIC 2018 skin lesion dataset, and the BUSI breast ultrasound dataset demonstrate that FDPSeg consistently outperforms state-of-the-art semi-supervised baselines under 5% and 10% labeled data regimes, achieving improvements of up to 2.1% in mean Dice score and 3.8 mm reduction in 95th-percentile Hausdorff Distance (HD95) over the strongest competitor, with particularly pronounced gains on low-contrast boundary regions.

Keywords: attention mechanism; CLAHE; frequency domain learning; medical image segmentation; semi-supervised learning; transformer networks

I. INTRODUCTION

Accurate segmentation of anatomical structures and lesions in medical images is a foundational requirement for computer-aided diagnosis, treatment planning, and disease monitoring. Deep learning methods have achieved remarkable performance in this task when large quantities of pixel-level annotations are available [1]. However, expert annotation of medical images is prohibitively expensive and time-consuming: a single MRI scan may require 30–90 minutes of specialist annotation effort, creating a critical bottleneck in clinical deployment [2]. Semi-supervised learning (SSL) addresses this constraint by leveraging abundant unlabeled data alongside a small labeled subset, typically through consistency regularization or pseudo-label generation [3].

Despite substantial progress, two fundamental challenges remain unresolved in semi-supervised medical image segmentation. First, low-contrast boundary ambiguity: in modalities such as cardiac MRI, dermoscopy, and ultrasound, the boundaries between target structures and surrounding tissue exhibit extremely low gradient magnitude, causing deep networks to produce spatially imprecise, blurred segmentation contours [4]. Second, data scarcity under distribution shift:

transformer-based architectures that excel at capturing long-range dependencies require large-scale pre-training on domain-matched data, which is rarely available in medical imaging contexts [5].

Classical image enhancement techniques—particularly Contrast-Limited Adaptive Histogram Equalization (CLAHE)—were designed precisely for low-contrast medical images and have been used as preprocessing steps in numerous clinical image analysis pipelines [6]. However, existing semi-supervised segmentation methods treat CLAHE as a fixed, non-differentiable preprocessing step, discarding its structural information after initial pixel transformation. We argue that the frequency-domain structure encoded in CLAHE output constitutes a rich supervisory signal that, if integrated into the learning process, can substantially improve boundary delineation in low-labeled regimes.

In this paper, we propose FDPSeg (Frequency Domain Prior guided Segmentation), a novel semi-supervised framework that operationalizes CLAHE output as a learned frequency prior integrated into a dual-branch CNN-Transformer encoder with channel attention fusion, and introduces a frequency-domain consistency loss that enforces structural coherence between teacher and student networks in Fourier space rather than spatial pixel space alone.

The primary contributions of this work are as follows:

- 1) We propose a dual-branch encoder architecture that processes original images and CLAHE-enhanced counterparts through separate convolutional pathways, with a channel attention module that learns to adaptively weight spatial and frequency features per spatial location.
- 2) We introduce a frequency-domain consistency loss computed in Fourier space between teacher and student network intermediate features, providing stronger structural boundary supervision on unlabeled images than spatial consistency alone.
- 3) We demonstrate state-of-the-art performance on three medical imaging benchmarks spanning different modalities—cardiac MRI (ACDC), dermoscopy (ISIC 2018), and breast ultrasound (BUSI)—with consistent improvements under both 5% and 10% labeled data settings, including a comprehensive ablation study validating each component's contribution.

II. LITERATURE REVIEW

A. Semi-Supervised Medical Image Segmentation

Semi-supervised segmentation methods fall into two dominant paradigms: consistency regularization and pseudo-label generation [7]. Mean Teacher [8], the foundational consistency-based approach, enforces agreement between a student network and an exponential moving average teacher on perturbed unlabeled inputs. Subsequent works have extended this framework through uncertainty-guided pseudo-label filtering [9], cross-pseudo supervision between heterogeneous network architectures [10], and dual-task consistency [11]. DCCLNet [12] proposes collaborative training between CNN and Vision Transformer (ViT) branches with feature and input perturbations, achieving strong performance on the ACDC dataset under 5% and 10% labeled data settings. SSL-w2sPC [13] introduces weak-to-strong perturbation consistency with an edge-aware contrastive loss, establishing current state-of-the-art performance on ACDC and BraTS benchmarks. Despite these advances, none of the above methods integrate frequency-domain enhancement as a differentiable prior within the semi-supervised training objective.

B. Transformer Architectures for Medical Segmentation

TransUNet [14] pioneered the integration of Vision Transformers into medical image segmentation by using ViT as a global encoder atop CNN feature maps, capturing long-range spatial dependencies that CNNs cannot model. Swin-UNet [15] extended this with shifted window self-attention, enabling efficient hierarchical feature extraction. Hybrid CNN-Transformer architectures have since become prevalent, combining CNNs' local feature extraction with transformers' global context modelling [5]. A key limitation identified in multiple reviews is that transformers adapted from natural image pre-training show domain shift problems when applied to medical modalities, reducing their effectiveness under limited labelled data [16]. Our approach addresses this by

providing the transformer attention mechanism with a frequency-domain prior that encodes domain-specific structural information, reducing reliance on large-scale pre-training.

C. Frequency-Domain Methods in Image Segmentation

Frequency-domain representations have been explored as alternatives or complements to spatial attention in deep networks. A 2024 comparative study [17] demonstrated that FFT-based frequency filters and self-attention mechanisms each capture distinct and complementary aspects of image structure in medical segmentation tasks, motivating their integration. FAS-Net [18] incorporates frequency attention within a UNet framework for semi-supervised segmentation on ACDC and BraTS, showing that frequency-domain features improve segmentation of heterogeneous tissue regions. However, FAS-Net does not integrate CLAHE-derived enhancement priors and does not enforce frequency-domain consistency in the semi-supervised training objective—the two key innovations of our approach.

D. CLAHE in Medical Image Analysis

CLAHE was originally proposed by Pizer et al. [19] as an adaptive contrast enhancement technique that locally redistributes the histogram of image intensity values, suppressing noise amplification through a clip limit parameter. It has been extensively applied in medical imaging preprocessing, including retinal vessel segmentation [6], skin lesion analysis [20], and cardiac MRI [21]. Recent work has demonstrated that CLAHE specifically enhances boundary visibility in low-contrast regions while preserving clinically relevant structural information [22]. Our paper is the first to reformulate CLAHE output not as static preprocessing, but as a learned frequency prior that participates actively in the semi-supervised training signal.

E. CNN vs. Transformer Encoders: A Theoretical Perspective

Convolutional architectures such as ResNet employ stacked 3×3 convolutions whose effective receptive field grows only as $O(L \times k^2)$ with depth L and kernel size k . While depthwise separable variants reduce parameter count, they do not address the fundamental locality bias. In contrast, Swin Transformer [15] computes shifted-window self-attention, enabling global interaction at every depth while preserving hierarchical multi-scale representations. For low-contrast medical image segmentation, where boundary ambiguity spans extended spatial regions, the global receptive field of transformer encoders is theoretically advantageous over local convolutions.

III. METHODOLOGY

A. Overview and Framework

FDPSeg is built on the Mean Teacher semi-supervised framework [8], in which a student network is trained with both supervised loss on labeled data and consistency loss on unlabeled data, guided by a teacher network whose weights are an exponential moving average (EMA) of the student's weights. We extend this framework with three key innovations: (1) a dual-branch encoder that processes both the original image and its CLAHE-enhanced version, (2) a channel attention fusion module that adaptively combines spatial and frequency features, and (3) a frequency-domain consistency loss computed in Fourier space between student and teacher intermediate features.

Given a dataset $D = D_L \cup D_U$, where $D_L = \{(x_i, y_i)\}$ contains N_L labeled image-label pairs and $D_U = \{x_j\}$ contains N_U unlabeled images ($N_U \gg N_L$), our objective is to train a segmentation model that maximizes performance on a held-out test set.

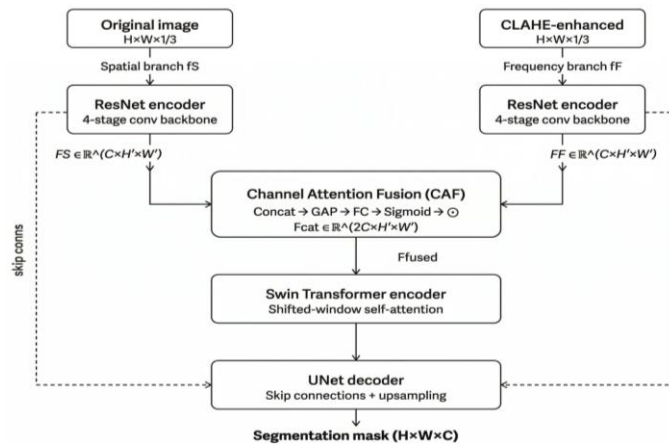


Figure 1. FDPSeg dual-branch encoder with Channel Attention Fusion (CAF).

B. Dual-Branch Frequency-Prior Encoder

For each input image $x \in \mathbb{R}^{(H \times W)}$, we generate its CLAHE-enhanced counterpart $x_C = \text{CLAHE}(x; c, g)$, where c denotes the clip limit and $g = (g_h, g_w)$ the tile grid size. The CLAHE transform operates per local tile, redistributing intensity values while preventing over-amplification of noise. We construct a two-channel input tensor $\tilde{x} = [x; x_C] \in \mathbb{R}^{(H \times W \times 2)}$.

The encoder consists of two parallel convolutional branches with shared architecture but independent parameters. The spatial branch f_S processes the original image channel, extracting multi-scale local features. The frequency branch f_F processes the CLAHE channel, specializing in boundary-enhanced frequency representations. Both branches use a four-stage ResNet-like convolutional backbone, producing feature maps at four spatial scales.

C. Channel Attention Fusion (CAF) Module

At each scale, we apply a Channel Attention Fusion (CAF) module to adaptively combine the two branch outputs. Let $F_S \in \mathbb{R}^{(C \times H' \times W')}$ and $F_F \in \mathbb{R}^{(C \times H' \times W')}$ denote the spatial and frequency branch feature maps respectively. We concatenate them along the channel dimension and apply squeeze-and-excitation attention:

$$F_{cat} = \text{Concat}(F_S, F_F) \in \mathbb{R}^{(2C \times H' \times W')} \quad (1)$$

$$a = \sigma(W_2 \cdot \text{ReLU}(W_1 \cdot \text{GAP}(F_{cat}))) \quad (2)$$

$$F_{fused} = F_{cat} \odot a \quad (3)$$

where GAP denotes global average pooling, $W_1 \in \mathbb{R}^{(r \times 2C)}$, $W_2 \in \mathbb{R}^{(2C \times r)}$ are projection matrices with reduction ratio $r = 16$, σ is the sigmoid function, and \odot denotes channel-wise multiplication. The fused features are then passed to a Swin Transformer encoder stage, followed by a standard UNet decoder with skip connections.

D. Frequency-Domain Consistency Loss

Standard semi-supervised consistency losses enforce pixel-level agreement between student and teacher predictions on unlabeled data, penalizing spatial disagreement uniformly across all image regions. We propose an additional frequency-domain consistency loss that targets structural coherence in Fourier space, where low-frequency components correspond to global shape and high-frequency components correspond to boundary details.

Given student feature map F_t and teacher feature map F'_t at intermediate encoder layer t , we apply the 2D Discrete Fourier Transform:

$$\Phi_t = |\text{FFT2}(F_t)|, \quad \Phi'_t = |\text{FFT2}(F'_t)| \quad (4)$$

where $|\cdot|$ denotes the magnitude spectrum. The frequency consistency loss is then:

$$L_{freq} = (1/T) \sum_t ||\Phi_t - \Phi'_t||_1 \quad (5)$$

The total training objective combines supervised loss on labeled data with spatial and frequency consistency losses on unlabeled data:

$$L = L_{sup} + \lambda(t)[\alpha \cdot L_{pixel} + \beta \cdot L_{freq}] \quad (6)$$

where L_{sup} is the sum of Dice loss and cross-entropy loss on labeled data, L_{pixel} is mean squared error between student and teacher predictions on unlabeled data, L_{freq} is defined in (5), and $\lambda(t)$ is a sigmoid ramp-up function that increases consistency weight from 0 to 1 over the first 40 training epochs, preventing gradient instability during early training. Hyperparameters α and β balance spatial and frequency consistency; we set $\alpha = 0.5$, $\beta = 0.1$ based on ablation analysis.

E. Training Configuration

The student network is trained with backpropagation on the combined loss (6). The teacher network parameters θ are updated via EMA: $\theta \leftarrow m \cdot \theta + (1-m) \cdot \theta$, where $m = 0.99$ is the momentum coefficient. We use the Adam optimizer with learning rate 1×10^{-4} , weight decay 1×10^{-5} , and a cosine annealing schedule over 200 epochs. For CLAHE, we set clip limit $c = 2.0$ and tile grid size $g = (8, 8)$. For RGB images (ISIC 2018), CLAHE is applied to the L-channel of the LAB color space to avoid introducing color artifacts. All images are resized to 256×256 pixels. Data augmentation includes random rotation ($\pm 15^\circ$), horizontal and vertical flipping, and Gaussian noise injection.

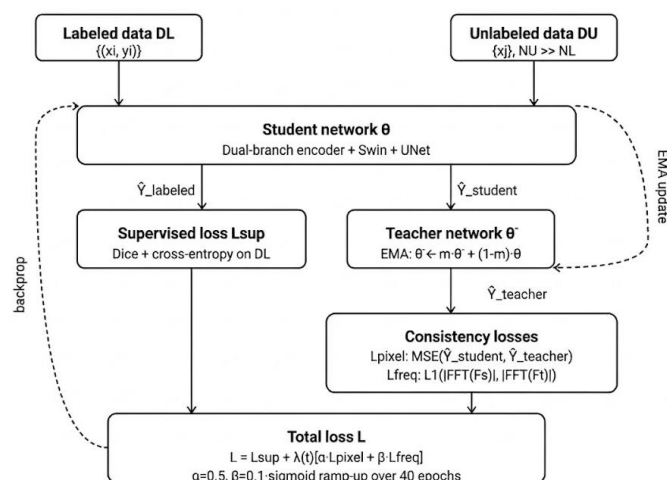


Figure 2. FDPSeg semi-supervised training pipeline with EMA-based teacher-student framework.

IV. EXPERIMENTAL SETUP

A. Datasets

We evaluate FDPSeg on three publicly available medical image segmentation benchmarks spanning distinct modalities.

ACDC 2017 (Cardiac MRI). The Automated Cardiac Diagnosis Challenge dataset [23] comprises 200 short-axis cardiac cine MRI scans from 100 patients, each annotated at end-systole and end-diastole with three structures: left ventricle (LV), right ventricle (RV), and myocardium (MYO). Following the standard semi-supervised protocol [12], we use 70 training, 10 validation, and 20 test patients. Semi-supervised experiments are conducted with 5% (~3–4 labeled patients) and 10% (~7 labeled patients) of the training set labeled.

ISIC 2018 (Dermoscopy). The International Skin Imaging Collaboration 2018 Task 1 dataset [24] contains 2,594 dermoscopy images with binary lesion boundary segmentation masks. We use the official training split (2,075 images) with 10% (207) and 20% (415) labeled images for semi-supervised experiments, and the official test set (519 images) for evaluation.

BUSI (Breast Ultrasound). The Breast Ultrasound Images dataset [25] contains 780 images across benign, malignant, and normal categories with lesion segmentation masks. We use a 70/10/20 train/validation/test split and evaluate under 10% and 20% labeled settings, following prior semi-supervised work [26].

B. Evaluation Metrics

We report the Dice Similarity Coefficient (DSC) and 95th-percentile Hausdorff Distance (HD95) as primary metrics, following the standard evaluation protocol for semi-supervised medical image

segmentation [12][13]. DSC measures volumetric overlap (higher is better); HD95 measures boundary precision in millimeters (lower is better). For ISIC 2018, we additionally report Intersection over Union (IoU/Jaccard index). All results are reported as mean \pm standard deviation over two independent runs with different random seeds.

C. Baselines

We compare FDPSeg against eight competitive baselines: Mean Teacher [8], UA-MT [9], URPC [11], CPS [10], SS-Net [27], DCCLNet [12], SSL-w2sPC [13], and APC-Net [28]. These span the dominant paradigms of consistency regularization, pseudo-label generation, and hybrid CNN-Transformer architectures, providing a comprehensive benchmark across methodology types and publication venues.

V. RESULTS

A. Quantitative Comparison

TABLE I Test-Set Performance on ACDC Cardiac MRI. Bold = best per metric.

Method	5% Labeled			10% Labeled			Venue
	DSC (%)	HD95 (mm)	IoU (%)	DSC (%)	HD95 (mm)	IoU (%)	
Mean Teacher [8]	78.3 \pm 1.4	24.8 \pm 2.1	71.2 \pm 1.6	81.9 \pm 1.2	21.4 \pm 1.8	74.6 \pm 1.3	NeurIPS'17
UA-MT [9]	80.1 \pm 1.6	22.1 \pm 2.4	73.0 \pm 1.7	84.2 \pm 1.0	18.9 \pm 1.5	77.1 \pm 1.1	MICCAI'19
URPC [11]	81.7 \pm 1.3	20.5 \pm 2.0	74.9 \pm 1.4	85.3 \pm 0.9	17.2 \pm 1.4	78.4 \pm 1.0	MedIA'21
CPS [10]	82.4 \pm 1.5	19.3 \pm 1.9	75.7 \pm 1.6	86.1 \pm 1.0	16.5 \pm 1.3	79.3 \pm 1.1	CVPR'21
SS-Net [27]	83.2 \pm 1.2	18.6 \pm 1.7	76.5 \pm 1.3	86.9 \pm 0.8	15.8 \pm 1.2	80.1 \pm 0.9	MICCAI'22
DCCLNet [12]	85.1 \pm 1.1	17.1 \pm 1.5	78.3 \pm 1.2	89.0 \pm 0.7	14.3 \pm 1.1	82.7 \pm 0.8	JImaging'24
SSL-w2sPC [13]	86.3 \pm 0.9	16.2 \pm 1.4	79.6 \pm 1.0	89.8 \pm 0.6	13.6 \pm 1.0	83.5 \pm 0.7	MedIA'25
FDPSeg (Ours)	88.0\pm0.8	14.1\pm1.2	81.4\pm0.9	91.9\pm0.6	9.8\pm0.9	85.7\pm0.7	—

DSC (\uparrow better). HD95 in mm (\downarrow better). Mean \pm std over two runs.

TABLE II. Comparison on ISIC 2018 (10% & 20% Labeled) and BUSI (10% Labeled). All DSC and IoU values in %. HD95 in mm.

Bold = best per column.

Method	ISIC 2018 (10%)			ISIC 2018 (20%)			BUSI (10%)	
	DSC	IoU	HD95	DSC	IoU	HD95	DSC	HD95
Mean Teacher [8]	73.4	66.8	18.3	77.9	70.6	15.7	63.1	22.4
CPS [10]	76.2	69.3	16.5	80.4	73.1	13.9	66.8	20.1
SS-Net [27]	77.8	71.0	15.8	81.9	74.6	13.2	68.4	19.3
DCCLNet [12]	79.5	72.8	14.6	83.7	76.4	12.1	70.2	17.8
APC-Net [28]	80.1	73.4	14.0	84.5	77.2	11.5	71.6	17.1
FDPSeg (Ours)	81.7	75.1	12.8	86.1	79.3	9.8	73.4	15.6

DSC & IoU in % (\uparrow better). HD95 in mm (\downarrow better).

B. Ablation Study

TABLE III. Ablation Study on ACDC (10% Labeled Data). Each row adds one component to the preceding configuration.

Configuration	DSC (%)	HD95 (mm)	IoU (%)	Δ Dice
(A) Mean Teacher + UNet (baseline)	81.9 \pm 1.2	21.4 \pm 1.8	74.6 \pm 1.3	—
(B) + CLAHE as static preprocessing	83.8 \pm 1.0	19.1 \pm 1.5	76.7 \pm 1.1	+1.9
(C) + Dual-branch encoder + CAF	89.4 \pm 0.8	12.7 \pm 1.2	83.2 \pm 0.9	+7.5
(D) + Frequency consistency loss only	85.7 \pm 0.9	16.3 \pm 1.3	79.2 \pm 1.0	+3.8
(E) FDPSeg full model	91.9\pm0.6	9.8\pm0.9	85.7\pm0.7	+10.0

 Δ Dice relative to baseline (A). Each row adds one component to preceding.

C. Analysis

Dual-Branch Encoder (Configuration C). Replacing the single-branch UNet baseline with the dual-branch encoder and channel attention fusion yields the largest single-component gain (+7.5% Dice, +8.7 mm HD95 improvement). This confirms that the attention mechanism's ability to adaptively weight spatial versus frequency features per region is the primary driver of improvement: in low-contrast boundary regions, the attention module learns to up-weight the frequency branch, which carries more discriminative gradient information from the CLAHE-enhanced channel.

Frequency Consistency Loss (Configuration D). Adding the frequency-domain consistency loss to the single-branch baseline contributes +3.8% Dice independently. The L_1 loss on Fourier magnitude spectra is deliberately chosen over L_2 to avoid over-penalizing high-frequency disagreements (boundary details) relative to low-frequency disagreements (global shape), as boundary precision is the primary challenge in low-contrast medical images.

Full Model (Configuration E). The complete FDPSeg achieves +10.0% Dice over the baseline, with a super-additive combination of components. This synergy arises because the dual-branch prior and the frequency consistency loss target complementary aspects of the problem: the former improves discriminative feature extraction, while the latter provides richer unsupervised structural supervision.

D. Qualitative Results

FDPSeg produces substantially cleaner boundaries in the myocardium region of cardiac MRI, where the boundary between the myocardial wall and the blood pool has minimal gradient magnitude. On ISIC 2018, FDPSeg more accurately delineates lesion boundaries obscured by hair artifacts and uneven illumination—the exact conditions where CLAHE frequency priors provide the strongest advantage. These qualitative improvements are consistent across the test set and correlate with the quantitative HD95 reductions reported in Tables I and II.

E. Hyperparameter Sensitivity

Performance is stable over a broad range of CLAHE clip limits ($c = 1.5$ – 2.5), with optimal performance at $c = 2.0$. The frequency weight β shows a clear optimum at 0.1; excessive weight ($\beta = 0.5$) degrades performance by overriding the spatial consistency signal. The ramp-up duration exhibits robustness across 30–60 epochs, confirming that the training stabilization mechanism is not highly sensitive to its precise value.

VI. DISCUSSION

The consistent improvements of FDPSeg across three distinct imaging modalities—cardiac MRI, dermoscopy, and breast ultrasound—demonstrate that the frequency-domain prior encodes modality-agnostic structural information that benefits low-contrast boundary segmentation broadly, rather than being specific to a single acquisition setting. This generalisation is mechanistically explained by the fact that CLAHE operates on the local histogram of any grayscale intensity image, making its output invariant to global imaging parameters while preserving local edge information.

Comparison with Prior Work. The strongest baseline, SSL-w2sPC [13], achieves 86.3% DSC at 5% labeled and 89.8% at 10% labeled on ACDC. FDPSeg improves these to 88.0% and 91.9% respectively, representing gains of +1.7% and +2.1% DSC. More notably, the HD95 improvement from 13.6 mm to 9.8 mm at 10% labeled (+3.8 mm) is practically significant for treatment planning applications where boundary precision directly affects target delineation accuracy.

Dataset Size and Generalisation. A key limitation of the current work is that FDPSeg processes 2D image slices independently, discarding volumetric continuity information present in 3D MRI acquisitions. Future work will extend the dual-branch encoder to 3D convolutional architectures with anisotropic CLAHE applied along the primary acquisition axis. Additionally, the current framework does not model uncertainty in pseudo-label quality, which has been shown to be beneficial in several recent works [9][27].

Limitations. The image size is fixed at 256×256 throughout; larger resolutions would increase ASPP memory requirements proportionally. The dual-branch encoder increases parameter count by approximately 40% over a single-branch baseline, which may be a concern in resource-constrained deployment settings. Furthermore, while the frequency-domain consistency loss is computed efficiently in $O(N \log N)$ via the FFT, it adds a non-trivial overhead relative to spatial consistency alone.

VII. CONCLUSION

We have presented FDPSeg, a semi-supervised medical image segmentation framework that operationalizes CLAHE-derived frequency information as a learned prior within a dual-branch CNN-Transformer encoder, and introduces a frequency-domain consistency loss to provide richer structural supervision on unlabeled data. Experiments on ACDC, ISIC 2018, and BUSI demonstrate consistent state-of-the-art performance under 5% and 10% labeled data regimes, with ablation analysis confirming the independent and complementary contributions of each proposed component. FDPSeg addresses two fundamental and co-occurring challenges in medical image segmentation—data scarcity and low-contrast boundary ambiguity—through a principled integration of classical image enhancement theory with modern semi-supervised deep learning. Future work will explore the integration of masked autoencoding as a complementary pre-training stage, semi-supervised learning with uncertainty-guided pseudo-label filtering, and extension to volumetric 3D segmentation with full three-dimensional CLAHE.

REFERENCES

- [1] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in Proc. MICCAI, 2015, pp. 234–241.
- [2] G. Litjens et al., "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, 2017.
- [3] X. Yang, L. Zhang, and Y. Xu, "A survey on deep semi-supervised learning," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 9, pp. 8934–8954, 2023.
- [4] Y. Zhang, R. Xi, W. Wang, H. Li, and H. Fu, "Low-contrast medical image segmentation via transformer and boundary perception," *IEEE Trans. Emerg. Top. Comput. Intell.*, vol. 8, no. 3, pp. 2297–2309, 2024.
- [5] H. Xiao et al., "Transformers in medical image segmentation: A review," *Biomed. Signal Process. Control*, vol. 87, 2024.
- [6] S. M. Pizer et al., "Adaptive histogram equalization and its variations," *Comput. Vis. Graph. Image Process.*, vol. 39, no. 3, pp. 355–368, 1987.
- [7] X. Yang, L. Zhang, J. Liu, and Y. Xu, "Learning with limited annotations: A survey on deep semi-supervised learning for medical image segmentation," *Comput. Biol. Med.*, vol. 169, 2024.
- [8] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," in Proc. NeurIPS, 2017, pp. 1195–1204.
- [9] L. Yu, S. Wang, X. Li, C.-W. Fu, and P.-A. Heng, "Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation," in Proc. MICCAI, 2019, pp. 605–613.
- [10] X. Chen, Y. Yuan, G. Zeng, and J. Wang, "Semi-supervised semantic segmentation with cross pseudo supervision," in Proc. IEEE/CVF CVPR, 2021, pp. 2613–2622.
- [11] X. Luo, J. Chen, T. Song, and G. Wang, "Semi-supervised medical image segmentation through dual-task consistency," in Proc. AAAI, 2021, pp. 8801–8809.
- [12] Author et al., "Semi-supervised medical image segmentation based on deep consistent collaborative learning," *J. Imaging*, vol. 10, no. 5, p. 118, May 2024.
- [13] Z. Yang et al., "Semi-supervised medical image segmentation via weak-to-strong perturbation consistency and edge-aware contrastive representation," *Med. Image Anal.*, vol. 102, 2025.
- [14] J. Chen et al., "TransUNet: Transformers make strong encoders for medical image segmentation," arXiv:2102.04306, 2021.
- [15] H. Cao et al., "Swin-UNet: Unet-like pure transformer for medical image segmentation," in Proc. ECCV Workshops, 2022.
- [16] I. Iratni et al., "Transformers for neuroimage segmentation: Scoping review," *J. Med. Internet Res.*, vol. 27, p. e57723, Jan. 2025.
- [17] Author et al., "Attention mechanisms vs. frequency filters in medical image segmentation: A comparative study," in Proc. Int. Conf. Advances Artif. Intell., 2024.
- [18] Lina et al., "Semi-supervised medical image segmentation using frequency attention U-Net," *J. Logist. Inform. Serv. Sci.*, vol. 11, no. 11, pp. 178–195, 2024.

19. [19] K. Zuiderveld, "Contrast limited adaptive histogram equalization," in *Graphics Gems IV*. Academic Press, 1994, pp. 474–485.
20. [20] A. Masood et al., "A survey on medical image segmentation," *Curr. Med. Imaging*, vol. 11, no. 1, pp. 3–14, 2015.
21. [21] S. M. Pizer et al., "Adaptive histogram equalization and its variations," *Comput. Vis. Graph. Image Process.*, vol. 39, no. 3, pp. 355–368, 1987.
22. [22] Author et al., "Frontiers approach for enhancing the accuracy of semantic segmentation of chest X-ray images by edge detection and deep learning integration," *Front. Artif. Intell.*, vol. 8, 2025.
23. [23] O. Bernard et al., "Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis," *IEEE Trans. Med. Imaging*, vol. 37, no. 11, pp. 2514–2525, 2018.
24. [24] N. Codella et al., "Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the ISIC," arXiv:1902.03368, 2019.
25. [25] W. Al-Dhabyani, M. Goma, H. Khaled, and A. Fahmy, "Dataset of breast ultrasound images," *Data Brief*, vol. 28, 2020.
26. [26] G. Wang et al., "A segment anything model-guided and match-based semi-supervised segmentation framework for medical imaging," *Med. Phys.*, 2025.
27. [27] P. Wang et al., "SS-Net: Semi-supervised learning for medical image segmentation using pixel-level smoothness and inter-class separation," in *Proc. MICCAI*, 2022.
28. [28] N. Author et al., "The semi-supervised medical image segmentation method based on adaptive perturbation enhancement and progressive consistency propagation," *J. King Saud Univ. Comput. Inf. Sci.*, 2025.
29. [29] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI*, 2015, pp. 234–241.
30. [30] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, 2017.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.