β-FFT: Nonlinear Interpolation and Differentiated Training Strategies for Semi-Supervised Medical Image Segmentation

Ming Hu[1,2] Jianfu Yin[1,2] Zhuangzhuang Ma[3*] Jianheng Ma[3*] Feiyu Zhu[1,2]
Bingbing Wu[1,2] Ya Wen[4] Meng Wu[5] Cong Hu[5,6†] Bingliang Hu[1†] Quan Wang[1†]
[1] Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences
[2] University of Chinese Academy of Sciences [3] Xidian University [4] Xi'an University of Technology
[5] Zhongnan Hospital of Wuhan University [6] The First Affiliated Hospital of Guangxi Medical University

Abstract

Co-training has achieved significant success in the field of semi-supervised learning(SSL); however, the homogenization phenomenon, which arises from multiple models tending towards similar decision boundaries, remains inadequately addressed. To tackle this issue, we propose a novel algorithm called β-FFT from the perspectives of data processing and training structure. In data processing, we apply diverse augmentations to input data and feed them into two sub-networks. To balance the training instability caused by different augmentations during consistency learning, we introduce a nonlinear interpolation technique based on the Fast Fourier Transform (FFT). By swapping low-frequency components between variously augmented images, this method not only generates smooth and diverse training samples that bridge different augmentations but also enhances the model's generalization capability while maintaining consistency learning stability. In training structure, we devise a differentiated training strategy to mitigate homogenization in co-training. Specifically, we use labeled data for additional training of one model within the co-training framework, while for unlabeled data, we employ linear interpolation based on the Beta(β) distribution as a regularization technique in additional training. This approach allows for more efficient utilization of limited labeled data and simultaneously improves the model's performance on unlabeled data, optimizing overall system performance. Code is available at the following link. https://github.com/Xi-Mu-Yu/beta-FFT.

1. Introduction

As manually annotating medical images such as CT, MRI, and pathology images is both costly and labor-intensive, this process becomes increasingly challenging for radiologists and other medical professionals as data volumes continue to grow, leading to scalability issues. Consequently, semi-supervised semantic segmentation has become particularly important in medical image analysis.

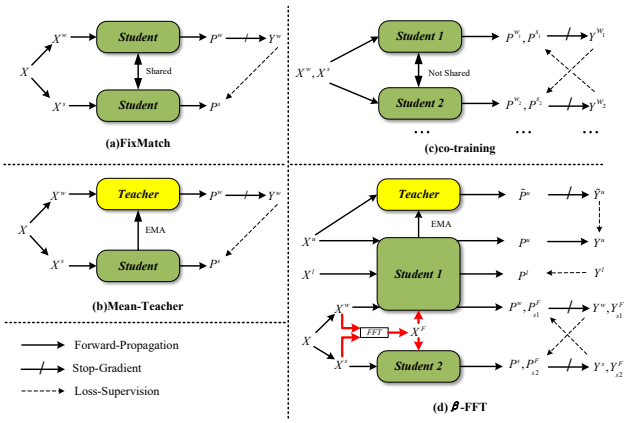

Figure 1. Illustrating the architectures for (a) FixMatch, (b) Mean-Teacher , (c) co-training , and (d) our approach β-FFT. X represents the dataset, P is the confidence map, Y denotes the labels encoded in one-hot format. More details in the method section.

The research on semi-supervised learning(SSL) began with the self-training method [11, 30], which enhances model learning by leveraging unlabeled data. Initially, researchers aimed to achieve this by self-generating labels, but this often led to unreliable results when handling unlabeled data[1]. To address this issue, researchers introduced consistency regularization methods[24, 31], usually applied by enforcing perturbations on unlabeled data in an online fashion, helping the model maintain stability when facing data variations.

As research progressed, data augmentation's role in bolstering model robustness has become evident. The **Fix-**

*Equal contribution
†Corresponding author.

Match method[32] optimizes semi-supervised learning by leveraging pseudo-labeling and consistency regularization on augmented unlabeled data (Fig.1a). Another notable approach in medical image semi-supervised segmentation is the **Mean Teacher** architecture[33], which employs an exponential moving average (EMA) of the student model's weights for the teacher network(Fig.1b). Inspired by this framework, various methods have been developed to enhance semi-supervised segmentation. For example, the UA-MT framework[45] utilizes uncertainty information to guide the student model in learning reliable targets. Verma et al.[34] introduced an **Interpolation Consistency Training Framework** to ensure that the predictions of interpolated unlabeled points align with actual data points. URPC[20] guarantees consistent predictions across scales. The method **BCP**[3] further enhances the Mean Teacher architecture by bi-directionally copying labeled and unlabeled data, allowing the unlabeled data to learn shared semantics from the labeled data and addressing the empirical mismatch in semi-supervised medical image segmentation.

While these approaches significantly enhance semi-supervised learning, they also pose challenges. The close coupling between teacher and student models[13] can hinder effective knowledge transfer, limiting the teacher's capacity to convey valuable insights. This coupling may lead to **confirmation bias**, causing models to excessively rely on existing biases and overlook potential new information. To overcome the limitations associated with the coupling in teacher-student models, researchers introduced the **co-training** method[13, 25]. Co-training leverages the complementary characteristics of multiple models to enhance knowledge sharing and transfer, enabling each model to gain additional knowledge from others, thus effectively improving the performance of SSL (Fig.1c).

Although co-training has achieved significant success in SSL, the risk of **model homogenization** remains a critical challenge[16]. Specifically, multiple models tend to converge to similar decision boundaries[1, 16], leading to homogenization. This phenomenon reduces the diversity of learned representations, thereby limiting the models' generalization capacity in semi-supervised settings. **This raises a crucial question: Can we introduce additional information or corrective mechanisms to maintain model accuracy while reducing homogenization between models in co-training, thereby enhancing overall performance?**

To mitigate the issue of model homogenization in co-training, we introduced improvements from two aspects: data processing and network training strategies.

Firstly, in terms of data processing, we incorporated diverse data augmentation techniques. By generating augmented samples with different intensities (including strong and weak augmentations) and training different subnetworks separately, we achieved model differentiation at the data level. Meanwhile, to prevent training instability caused by varying augmentation strengths, we employed the Fast Fourier Transform (FFT) to exchange low-frequency information between strongly and weakly augmented images, thereby creating new samples that lie between the two. This approach not only provides the model with new data perspectives but also effectively alleviates the instability caused by strong and weak augmentations.

Secondly, in terms of training strategies, we applied an additional training step to one of the subnetworks, while the other followed the original training process. We utilized labeled data for extra training on the selected model and introduced unlabeled data, generated through linear interpolation based on Beta distribution sampling, as a regularization term in the training process. This strategy not only helps maintain diversity in decision boundaries between the two models and reduces their homogenization but also further enhances the effectiveness of collaborative learning.We performed additional supervised training on Student Model 1 using labeled data, which not only enhances the performance of Student Model 1 but also helps to increase its independence. This approach aids in mitigating the confirmation bias between Student Model 1 and the Teacher network. To summarize, we make the following contributions:

1. **Nonlinear Interpolation Strategy**: By using the Fast Fourier Transform to exchange low-frequency information between weakly and strongly augmented images, we effectively mitigated the training instability caused by using different augmented data to address model homogenization. At the same time, this approach enriched the diversity of data samples and improved the model's generalization ability.
2. **Dehomogenization during training**: Student model 1 undergoes additional training, while Student model 2 does not. This differentiated training approach generates a unique loss for each model, effectively reducing homogenization between the two student models and preserving the diversity of their decision boundaries.

2. Related Work

2.1. Semi-Supervised Medical Image Segmentation

Previous methods can be broadly categorized into self-training methods[6, 49] and consistency regularization methods[8, 13]. Self-training algorithms are considered the fundamental prototype of pseudo-labeling methods[28], where a model is pre-trained on a labeled dataset and iteratively retrained or fine-tuned using predictions from unlabeled data. [23] matched these pseudo-labels by synthesizing new images rather than optimizing them. Within the framework of consistency regularization, [12] employed strong augmentation and weak augmentation to handle unlabeled data. Some studies have explored consistent data

transformations, such as patch shuffling data transformation [15], cut-and-paste augmentation [44], and copy-paste [3]. ABD[9] effectively integrates multiple perturbations through an adaptive bidirectional displacement mechanism, enhancing the quality of consistency learning. AD-MT[48] reduces confirmation bias and enhances model performance under limited labeled data by employing random periodic alternation and a counteracting disturbance module.

2.2. Frequency Domain Enhancement Techniques

Fourier domain processing techniques enhance model generalization, robustness, and adaptability in computer vision tasks, particularly for domain adaptation and data augmentation. Fourier Domain Adaptation (FDA) [43] replaces the low-frequency amplitude spectrum of source images with that of target images, enabling model adaptation to new domains while preserving structural information. The Fourier-based Domain Generalization Framework [41] systematically investigates the roles of amplitude and phase spectra in domain shifts, revealing that the amplitude spectrum captures domain-specific style information, while the phase spectrum retains structural content. This insight underpins frequency-based augmentation strategies for improved generalization. FreMix [40] performs frequency-based augmentation by mixing amplitude spectra of different images, thereby enhancing domain generalization.

2.3. Research Status of Homogenization

In the co-training framework of semi-supervised learning, the issue of model homogenization has become a core challenge that restricts performance improvement. To address this problem, researchers have proposed systematic solutions from three levels: data augmentation, model architecture, and training strategies. On the level of data augmentation, consistency training based on strong-weak augmentation combinations[9, 16, 27, 32] generates multi-view samples through differentiated disturbances, while the contrastive learning framework[14] further utilizes graph structures to dynamically allocate samples, enhancing complementarity among models. Recent work [26] also minimizes mutual information to constrain the independence of view features, reducing redundancy. On the model architecture level, heterogeneous network designs (such as combinations of CNN and Transformer[16, 22]) and model parameter diversification[10] are used to force models to focus on different feature patterns. On the training strategy level, dynamic optimization methods[39], asymmetric learning mechanisms (such as alternating training[48]) have been proven to effectively prevent model convergence.

3. Method

In semi-supervised segmentation, we aim to train a model using both labeled and unlabeled data. The labeled dataset $\mathcal{D}^l = \left\{(X_i^l, Y_i^l)\right\}_{i=1}^N$ contains N labeled images, where X_i^l is the image and Y_i^l is its corresponding segmentation label. The unlabeled dataset $\mathcal{D}^u = \left\{X_j^u\right\}_{j=1}^M$ consists of M unlabeled images, where X_j^u has no associated label. Typically, $N \ll M$, meaning the number of labeled images is much smaller than the unlabeled ones.

In our approach, we employ a single **teacher model** alongside two **student models**. The parameters of the teacher model are updated using an Exponential Moving Average (EMA) mechanism, specifically tuned based on the parameters of **Student Model 1**. The update process for the teacher model at each iteration can be expressed as:

$$\theta_T^{(t)} = \lambda\theta_T^{(t-1)} + (1-\lambda)\theta_{S1}^{(t)} \tag{1}$$

Here, $\theta_T^{(t)}$ denotes the parameters of the teacher model at the t-th iteration, $\theta_{S1}^{(t)}$ represents the parameters of Student Model 1 at the same iteration, and $\lambda \in [0, 1]$ serves as a smoothing factor that balances the influence of previous teacher parameters against those of Student Model 1.

3.1. Overview

1. **Data Augmentation and Teacher Network Training via Copy-Paste** We apply both weak and strong data augmentations to the data, enhancing its diversity through simple transformations and advanced techniques. Additionally, we utilize a Copy-Paste method to train a teacher network, ensuring high-quality pseudo-labels for the unlabeled data. This approach effectively improves the accuracy of the generated pseudo-labels.
2. **Nonlinear Interpolation:** To reduce homogenization in collaborative training and increase data diversity, we input data with varying degrees of enhancement (strong and weak) into different sub-models. We also exchange low-frequency components between weakly and strongly enhanced images to reduce the instability of collaborative training.
3. **Differentiated Training:** One model in the co-training framework undergoes additional training, while the other does not, reducing homogeneity between them.

3.2. Data Augmentation and Teacher Network Training via Copy-Paste

We begin by applying both strong(s) and weak(w) augmentations to the data. Weak augmentations include simple transformations such as rotation and flipping, while strong augmentations build on these with techniques like Cutout [7] and ColorJitter[42]. Inspired by BCP [3], we employ a Copy-Paste technique for further data enhancement. Specifically, the Copy-Paste process can be expressed as follows:

$$X_{w/s}^{in} = M \odot X_{w/s}^l + (1-M) \odot X_{w/s}^u, \tag{2}$$

$$X_{w/s}^{out} = M \odot X_{w/s}^u + (1-M) \odot X_{w/s}^l, \tag{3}$$

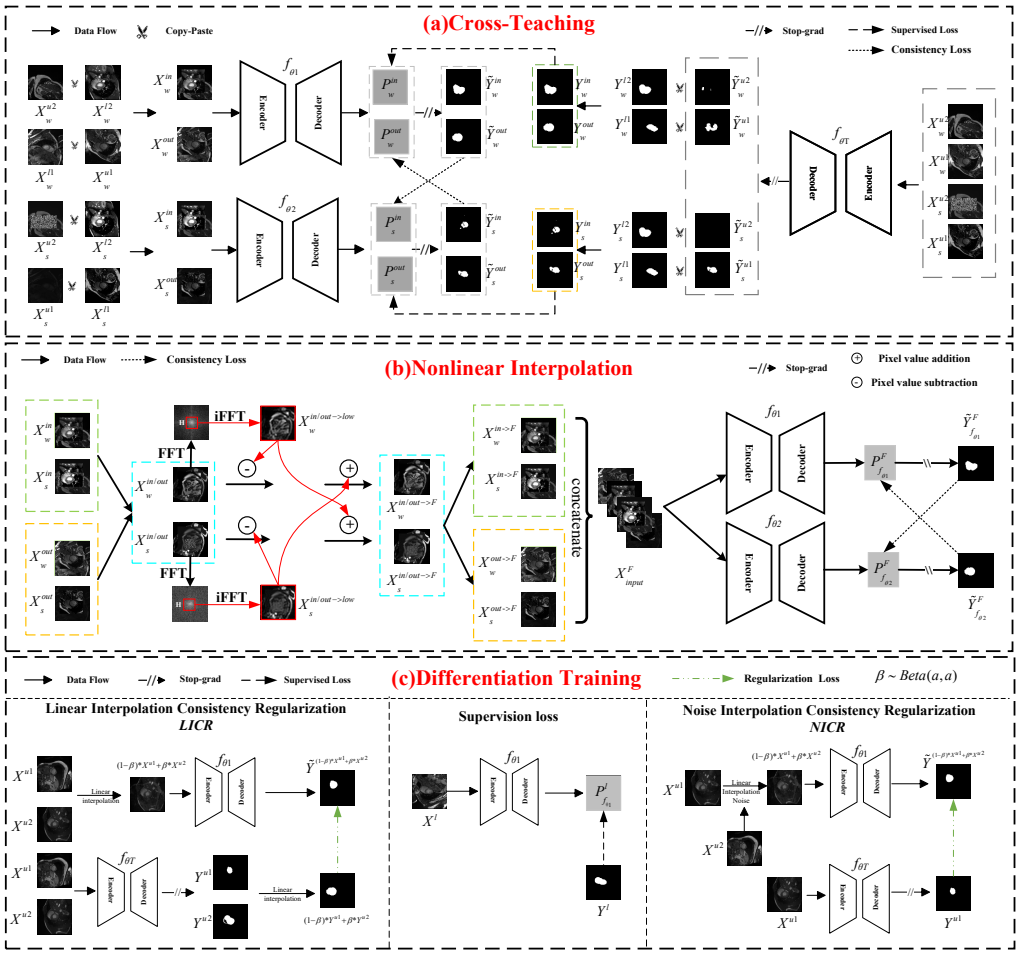

Figure 2. Overview of our β-FFT framework.In the figure, X represents the data, l represents labeled data, u represents unlabeled data, w denotes weak augmentation, and s denotes strong augmentation. P represents the confidence map obtained from the data through the model, $\tilde{Y}$ represents the class prediction map corresponding to the confidence map obtained through the model, and Y represents the corresponding ground truth labels.FFT represents the Fast Fourier Transform, and iFFT represents the Inverse Fast Fourier Transform.

Here, $X^l_{w/s}$ represents labeled data, and $X^u_{w/s}$ represents unlabeled data. We utilize a mask $M \in \{0,1\}^{W\times H}$ to perform the bidirectional copy-pasting operation, controlling the blending between the images. The mask defines a zero-value region of size $\eta H \times \eta W$, where $\eta \in (0,1)$ governs the proportion of the foreground region.

To ensure high-quality predictions for the unlabeled data, we pre-train a teacher model f_{θ_T} using two different labeled data, l_1 and l_2, as follows:

$$X^{Teacher}_{w/s} = M \odot X^{l_1}_{w/s} + (1-M) \odot X^{l_2}_{w/s}, \quad (4)$$

$$Y^{Teacher}_{w/s} = M \odot Y^{l_1}_{w/s} + (1-M) \odot Y^{l_2}_{w/s}. \quad (5)$$

Thus, the labels corresponding to the unlabeled data $X^u_{w/s}$ are given by:

$$Y^u_{w/s} = \underset{c\in C}{argmax}(f_{\theta_T}(X^u_{w/s}), c). \quad (6)$$

Accordingly, the labels for $X^{in}_{w/s}$ and $X^{out}_{w/s}$ can be expressed as:

$$Y^{in}_{w/s} = M \odot Y^l_{w/s} + (1-M) \odot Y^u_{w/s}, \quad (7)$$

$$Y^{out}_{w/s} = M \odot Y^u_{w/s} + (1-M) \odot Y^l_{w/s}. \quad (8)$$

3.3. Non-linear Interpolation

We utilize a non-linear interpolation technique based on the exchange of low-frequency components to enhance the diversity of data samples, particularly for image augmentation. Our approach begins by decomposing an image I into its frequency components using the FFT:

$$I = F^{-1}(F(I)) \quad (9)$$

where F denotes the FFT and F^{-1} represents its inverse, the Inverse Fast Fourier Transform (iFFT). Subsequently,

we extract the low-frequency component $I^{\rightarrow \text{low}}$ of the image, defined as:

$$I^{\rightarrow \text{low}} = F^{-1}(F(I) \cdot H) \tag{10}$$

H is a low-pass filter employed to isolate low-frequency information.We conducted experimental investigations on the setting of H in the experimental section.

To enhance the diversity of the augmented images, we then perform non-linear interpolation by swapping the low-frequency components of the weakly augmented image I_w with those of the strongly augmented image I_s. This generates two new images $I_w^{->F}$ and $I_s^{->F}$:

$$I_{w/s}^{->F} = I_{w/s} - I_{w/s}^{->low} + I_{s/w}^{->low} \tag{11}$$

Next, to further increase the diversity of the augmented images, we perform non-linear interpolation by swapping the low-frequency components of the weakly augmented images X_w^{in} and X_s^{in} with those of the strongly augmented images X_w^{out} and X_s^{out}. This process generates four new images, defined as:

$$X_{w/s}^{\text{in/out}\rightarrow F} = X_{w/s}^{\text{in/out}} - X_{w/s}^{\text{in/out}\rightarrow \text{low}} + X_{s/w}^{\text{in/out}\rightarrow \text{low}} \tag{12}$$

We aim for different models to produce similar outputs across various samples to ensure they learn consistent feature representations. This consistency enables the models to maintain strong performance on unseen samples, thereby enhancing their generalization ability. To strengthen this effect, we concatenate the four images generated through non-linear interpolation into a new set X_{input}^F:

$$X_{\text{input}}^F = \text{Concat}\left[X_w^{\text{in}\rightarrow F}, X_s^{\text{in}\rightarrow F}, X_w^{\text{out}\rightarrow F}, X_s^{\text{out}\rightarrow F}\right] \tag{13}$$

3.4. Differentiated Training of Student Models

Student Model 1 is additionally trained using **labeled data**, whereas **Student Model 2** does not undergo this supplementary training. To enhance the robustness of Student Model 1, we incorporate two regularization techniques: linear interpolation consistency regularization and noise interpolation consistency regularization.

3.4.1. Linear Interpolation Consistency Regularization

Student Model 1 employs a pixel-wise data perturbation strategy along with consistency regularization that leverages unlabeled data. Given two unlabeled data points, X^{u_1} and X^{u_2}, we generate an interpolated data point $M_\beta(X^{u_1}, X^{u_2})$, defined as follows:

$$M_\beta(X^{u_1}, X^{u_2}) = \beta X^{u_1} + (1-\beta)X^{u_2} \tag{14}$$

In this equation, the hyperparameter β is sampled from a **Beta distribution**,following the setup in Mixup[46]. We apply the linear interpolation consistency regularization, which compares the output at the interpolated data point $f_{\theta_1}(M_\beta(X^{u_1}, X^{u_2}))$ with the outputs of the original data points:

$$M_\beta(f_{\theta_1}(X^{u_1}), f_{\theta_1}(X^{u_2})) \approx f_{\theta_1}(M_\beta(X^{u_1}, X^{u_2})) \tag{15}$$

3.4.2. Noise Interpolation Consistency Regularization

We also introduce a noise interpolation consistency constraint by reformulating equation 14 as follows:

$$\begin{aligned} M_\beta(X^{u_1}, X^{u_2}) &= \beta X^{u_1} + (1-\beta)X^{u_2} \\ &= X^{u_1} + (1-\beta)\cdot(X^{u_2} - X^{u_1}) \end{aligned} \tag{16}$$

In this formulation, we interpret $(1-\beta)\cdot(X^{u_2} - X^{u_1})$ as noise interference. The noise consistency loss can then be expressed as:

$$M_\beta(f_{\theta_1}(X^{u_1}), f_{\theta_1}(X^{u_2})) \approx f_{\theta_1}(M_\beta(X^{u_1})) \tag{17}$$

This noise consistency constraint ensures that Student Model 1 produces outputs that closely resemble those of the teacher model, even when the input data is slightly perturbed. By assessing the difference between the output at the interpolated point $M_\beta(X^{u_1}, X^{u_2})$ and the teacher model's output at X^{u_1}, this loss term encourages Student Model 1 to utilize information from more than just a single data point X^{u_1}. This approach effectively enhances the model's generalization ability.

In contrast, **Student Model 2** does not leverage labeled data for training. It employs cross pseudo-supervision and applies cross-consistency loss derived from the nonlinear interpolation process. Unlike Student Model 1, which utilizes both labeled and unlabeled data for consistency regularization, Student Model 2 focuses on leveraging shared learning signals between the two student models. This results in markedly different training trajectories for Student Model 1 and Student Model 2, promoting diversity in learning signals throughout the overall learning process.

Since Linear Interpolation Consistency Regularization and Noise Interpolation Consistency Regularization serve as two distinct additional training processes for labeled data, we refer to the training process with linear interpolation as **LICR**, and the one with noise interpolation as **NICR**.

4. Loss Functions

The overall loss function is comprised of three main components: **Cross Teaching Loss**, **Nonlinear Interpolation Loss**, and **Differentiation Loss**.The notation $\mathcal{L}_{\text{ce}}$ represents the Cross Entropy Loss, and $\mathcal{L}_{\text{dice}}$ represents the Dice Loss.

4.1. Cross Teaching Loss

Cross Teaching Loss leverages pseudo-labels from one model to supervise the other. It consists of a supervised loss and a cross pseudo-supervised loss. The supervised

loss ensures effective learning from ground truth labels by combining cross-entropy and Dice losses on both strongly and weakly augmented inputs:

$$\begin{aligned}\mathcal{L}_{sup}^{aug} = &\frac{1}{2}\left(\mathcal{L}_{ce,dice}(f_{\theta_1}(X_w^{in}), Y_w^{in}) + \mathcal{L}_{ce,dice}(f_{\theta_1}(X_w^{out}), Y_w^{out})\right) \\ &+\frac{1}{2}\left(\mathcal{L}_{ce,dice}(f_{\theta_2}(X_s^{in}), Y_s^{in}) + \mathcal{L}_{ce,dice}(f_{\theta_2}(X_s^{out}), Y_s^{out})\right)\end{aligned} \tag{18}$$

To enforce consistency, the cross pseudo-supervised loss aligns predictions with pseudo-labels from the other model:

$$\begin{aligned}\mathcal{L}_{cps}^{aug} = &\mathcal{L}_{dice}\left(f_{\theta_1}(X_w^{in}), \underset{c\in C}{argmax}(f_{\theta_2}(X_s^{in}), c)\right) \\ &+ \mathcal{L}_{dice}\left(f_{\theta_1}(X_w^{out}), \underset{c\in C}{argmax}(f_{\theta_2}(X_s^{out}), c)\right) \\ &+ \mathcal{L}_{dice}\left(f_{\theta_2}(X_s^{in}), \underset{c\in C}{argmax}(f_{\theta_1}(X_w^{in}), c)\right) \\ &+ \mathcal{L}_{dice}\left(f_{\theta_2}(X_s^{out}), \underset{c\in C}{argmax}(f_{\theta_1}(X_w^{out}), c)\right)\end{aligned} \tag{19}$$

4.2. Nonlinear Interpolation Consistency Loss

The Nonlinear Interpolation Consistency Loss ensures that the model outputs are consistent across interpolated data points. This is represented as follows:

$$\begin{aligned}\mathcal{L}_{consistency}^{Nonlinear} = &\mathcal{L}_{dice}\left(f_{\theta_1}(X_{input}^F), \underset{c\in C}{argmax}(f_{\theta_2}(X_{input}^F), c)\right) \\ &+ \mathcal{L}_{dice}\left(f_{\theta_2}(X_{input}^F), \underset{c\in C}{argmax}(f_{\theta_1}(X_{input}^F), c)\right)\end{aligned} \tag{20}$$

4.3. Differentiation Loss

Differentiated training primarily focuses on Student Model 1, incorporating supervised training with labeled data and a regularization term for unlabeled data. Consider labelled samples $(X_i^l, Y_i^l) \sim D^l$ from joint distribution $P(X, Y)$ and unlabelled samples $(X_i^u, X_j^u) \sim D^u$ from borderline distribution $P(X) = \frac{P(X,Y)}{P(X|Y)}$. Using SGD for every iteration t, the encoder-decoder parameter θ is updated minimising the objective function:

$$\mathcal{L}_{student1} = \mathcal{L}_{student1}^l + r(t) \cdot \mathcal{L}_{student1}^u \tag{21}$$

where $\mathcal{L}_{student1}^l$ is the cross entropy loss and dice loss applied over the labelled data D^l .Therefore, $\mathcal{L}_{student1}^l$ can be expressed as:

$$\mathcal{L}_{student1}^l = \frac{1}{2}\left(\mathcal{L}_{ce}(f_{\theta_1}(X^l), Y^l) + \mathcal{L}_{dice}(f_{\theta_1}(X^l), Y^l)\right) \tag{22}$$

$\mathcal{L}_{student1}^u$ is the interpolation consistency regularization loss applied over the unlabelled data D^u, $r(t)$ is the ramp function adjusting the weight of $\mathcal{L}_{student1}^u$ after every iteration. $\mathcal{L}_{student1}^u$ has two options: one is the LICR, and the other is the NICR.

LICR is calculated over (X_i^u, X_j^u) of sampled mini-batches and the pseudo labels $f_{\theta_T}(X_i^u)$ and $f_{\theta_T}(X_j^u)$. Next, interpolation $M_\beta(X_i^u, X_j^u)$ and model prediction $f_{\theta_1}(M_\beta(X_i^u, X_j^u))$ are computed updating θ to bring model prediction closer to the interpolation of the pseudo labels, $M_\beta(f_{\theta_T}(X_i^u), f_{\theta_T}(X_j^u)$. The deviation in model prediction and the interpolation of the pseudo labels is penalised using the mean squared loss. LICR can be expressed as:

$$\mathcal{L}_{LICR}^u = \mathbb{E}_{X_i^u, X_j^u}\left[\left\|f_{\theta_1}\left(M_\beta(X_i^u, X_j^u)\right) - M_\beta\left(f_{\theta_T}(X_i^u), f_{\theta_T}(X_j^u)\right)\right\|^2\right] \tag{23}$$

According to Equations 16 and 17, we can transform LICR to obtain a new loss, NICR, which is defined as:

$$\mathcal{L}_{NICR}^u = \mathbb{E}_{X_i^u, X_j^u}\left[\left\|f_{\theta_1}(M_\beta(X_i^u, X_i^u)) - f_{\theta_T}(X_i^u)\right\|^2\right] \tag{24}$$

5. Experiments

5.1. Dataset and Evaluation Metrics

ACDC Dataset: The ACDC dataset [5] consists of 200 annotated short-axis cardiac cine-MR images from 100 patients across four classes. 2D segmentation is more common than 3D [2]. Evaluation metrics include *Dice Similarity Coefficient* (DSC), *Jaccard*, *95% Hausdorff Distance* (95HD), and *Average Surface Distance* (ASD). Following the methods of BCP[3] and ABD[9], the input size was set to 256 × 256, with a batch size of 24 for training.

PROMISE12 Dataset: The PROMISE12 dataset [17] was introduced in the MICCAI 2012 prostate segmentation challenge, comprising MRI scans of 50 patients. All 3D scans are converted into 2D slices. DSC and ASD are used for evaluation. Following ABD[9], the input size was set to 224 × 224 with a batch size of 16 for training.

MS-CMRSeg 2019: The MS-CMRSeg 2019 dataset [35, 51] from the MICCAI 2019 challenge includes 45 multi-sequence cardiac MRI scans of cardiomyopathy patients. Evaluation metrics are DSC, Jaccard, 95HD, and ASD. Following the DiffRect[19] approach, the input size was 256 × 256. The training batch size was set to 8.

5.2. Comparison with SOTA Methods

Compared to SOTA methods on the ACDC test set, **β-FFT** demonstrates superior performance, particularly when using 10% labeled data, where the model achieves Dice and Jaccard scores of 90.50±0.04% and 83.12±0.12%, outperforming many recent methods, such as AD-MT and ABD.

On the PROMISE12 test set, β-FFT also outperforms existing methods with 20% labeled data, achieving a Dice score of 83.75±0.65% and an ASD of 1.20±0.07, surpassing AD-MT and ABD.

On the MS-CMRSEG 2019 dataset, β-FFT achieves a Dice score of 87.79±0.04% and a Jaccard index of 78.60±0.06% significantly outperforming popular semi-supervised approaches and approaching fully supervised performance with less labeled data.

Figure 3 presents a visual comparison of our method with other similar approaches.

Table 1. Comparisons with other methods on the ACDC test set.

Method	Scans used		Metrics			
	Labeled	Unlabeled	DSC↑	Jaccard↑	95HD↓	ASD↓
U-Net (MICCAI'2015) [29]	3(5%)	0	47.83	37.01	31.16	12.62
	7(10%)	0	79.41	68.11	9.35	2.70
	70(All)	0	91.44	84.59	4.30	0.99
DTC (AAAI'2021) [20]	3(5%)	67(95%)	56.90	45.67	23.36	7.39
URPC (MICCAI'2021) [21]			55.87	44.64	13.60	3.74
MC-Net (MICCAI'2021) [36]			62.85	52.29	7.62	2.33
SS-Net (MICCAI'2022) [38]			65.83	55.38	6.67	2.28
SCP-Net (MICCAI'2023) [47]			87.27	-	-	2.65
Cross Teaching (MIDL'2022) [22]			65.60	-	16.2	-
BCP (CVPR'2023) [3]			87.59	78.67	1.90	0.67
DiffRect (MICCAI'2024) [19]			82.46	71.76	7.18	1.94
ABD (CVPR'2024) [9]			88.96	80.70	1.57	0.52
AD-MT (ECCV'2024) [48]			88.75	80.41	**1.48**	**0.50**
Ours-β-FFT			**89.46**±0.12	**81.46**±0.22	1.78±0.32	0.55±0.10
DTC (AAAI'2021) [20]	7(10%)	63(90%)	84.29	73.92	12.81	4.01
URPC (MICCAI'2021) [21]			83.10	72.41	4.84	1.53
MC-Net (MICCAI'2021) [36]			86.44	77.04	5.50	1.84
SS-Net (MICCAI'2022) [38]			86.78	77.67	6.07	1.40
Cross Teaching (MIDL'2022) [22]			86.45	77.02	6.30	1.86
SCP-Net (MICCAI'2023) [47]			89.69	-	-	0.73
PLGCL (CVPR'2023) [4]			89.1	-	4.98	1.80
BCP (CVPR'2023) [3]			88.84	80.62	3.98	1.17
DiffRect (MICCAI'2024) [19]			89.27	81.13	3.85	1.00
ABD (CVPR'2024) [9]			89.81	81.95	**1.46**	0.49
AD-MT (ECCV'2024) [48]			89.46	81.47	1.51	**0.44**
Ours-β-FFT			**90.50**±0.04	**83.12**±0.12	2.38±0.87	0.62±0.13

Table 2. Comparisons with state-of-the-art semi-supervised segmentation methods on the PROMISE12 test set.

Method	Scans used		Metrics	
	Labeled	Unlabeled	DSC↑	ASD↓
U-Net [29]	7(20%)	0	60.88	13.87
	35(100%)	0	84.76	1.58
CCT [25]	7(20%)	28(80%)	71.43	16.61
URPC [21]			63.23	4.33
SS-Net [38]			62.31	4.36
SLC-Net [18]			68.31	4.69
SCP-Net [47]			77.06	3.52
ABD [9]			82.06	1.33
AD-MT [48]			79.82	1.77
Ours-β-FFT			**83.75**±0.65	**1.20**±0.07

Table 3. Segmentation results on MS-CMRSEG 2019 with 20% data labeled.

Method	Dice ↑	Jaccard↑	HD95↓	ASD↓
UAMT [45]	84.27	73.69	12.15	4.18
FixMatch [32]	84.31	73.57	17.79	4.81
CPS [8]	83.66	73.03	15.01	4.30
ICT [34]	83.66	73.06	17.24	4.85
MCNetV2 [37]	83.93	73.45	13.10	3.39
INCL [50]	84.33	73.92	9.95	2.61
DiffRect [19]	86.78	77.13	6.39	1.85
ABD [9]	87.25	77.77	11.74	4.25
AD-MT [48]	86.30	76.39	**3.56**	**1.21**
Ours-β-FFT	**87.79**±0.04	**78.60**±0.06	3.75±0.36	1.62±0.20
Supervised [29]	88.19	79.28	4.21	1.32

5.3. Ablation Study Analysis

The baseline method we use is an improved version of the BCP[3] from the ABD[9], combined with the Cross-Teaching framework[22]. Its structure consists of a mean-teacher framework with two student models and one teacher model.**All experiments are conducted on the ACDC dataset, with 10% of the data labeled**.

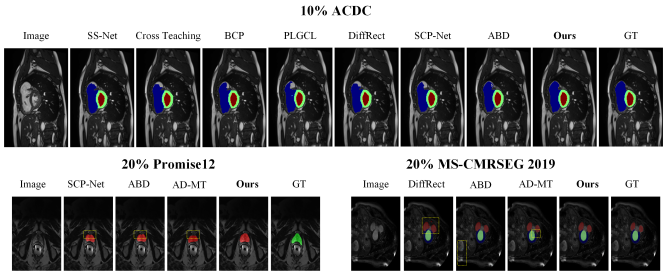

Figure 3. Visualization of segmentation results on ACDC dataset with 10% labeled data, PROMISE12 dataset with 20% labeled data,and MS-CMRSEG 2019 dataset with 20% labeled data .

5.3.1. Effect of Non-linear Interpolation Strategy

The experimental results demonstrate that the low-frequency component enhancement method, utilizing low-pass filters, effectively improves model performance on the ACDC dataset. Specifically, the 20x20 filter size achieved the best results on the validation set, while the 30x30 filter size excelled on the test set. This finding highlights the importance of selecting an appropriate filter size to balance detail and global information, ultimately optimizing the model's generalization ability.

Table 4. Effect of Low-Pass Filter Size on Model Performance.

H	ACDC Validation dataset		ACDC Test dataset	
	Dice↑	Jaccard↑	Dice↑	Jaccard↑
None	89.53	81.86	89.60	81.70
20x20	**90.23**	**82.94**	88.96	80.76
30x30	89.77	82.27	**89.96**	**82.29**
40x40	89.93	82.58	89.65	81.80
50x50	89.81	82.38	89.49	81.55

5.3.2. Impact of Differentiated Training Strategies on Model Performance

In this analysis, we use Student Model 1 as an example to validate the impact of differentiated training on model performance across the ACDC validation and test datasets.

The table 5 indicates that standardized training is conducted jointly for Student Model 1 and Student Model 2, while the differentiated training section represents additional training specifically for Student Model 1. Here, w denotes the use of weakly augmented data, and s denotes the use of strongly augmented data. The results show that by introducing the two differentiated training strategies, LICR and NICR, in the baseline model, the Dice value of the model on the test set significantly improves, with LICR(w) achieving a Dice value of 90.24 and NICR(w) reaching 90.23, highlighting the effectiveness of differentiated training in enhancing model performance.

Employing the Non-linear interpolation strategy, the

Table 5. Ablation study of Differentiated Training on ACDC Validation and Test Datasets.

Standardized Training		Differentiated Training				ACDC Validation		ACDC Test	
Baseline	Non-linear	LICR(w)	NICR(w)	LICR(s)	NICR(s)	Dice↑	Jaccard↑	Dice ↑	Jaccard ↑
✓						89.53	81.86	89.60	81.70
✓		✓				89.95	82.58	90.24	82.73
✓			✓			90.34	82.83	90.23	82.69
✓				✓		90.25	82.97	90.26	82.70
✓					✓	90.38	83.23	90.26	82.74
✓		✓		✓		90.09	82.79	90.13	82.47
✓		✓			✓	90.10	82.70	90.10	82.54
✓	✓					89.77	82.27	89.96	82.29
✓	✓	✓				90.53	83.54	90.27	82.76
✓	✓		✓			90.21	83.01	90.34	82.84
✓	✓			✓		90.30	83.14	90.38	82.95
✓	✓				✓	90.31	83.02	89.97	82.32
✓	✓	✓		✓		**90.77**	**83.81**	**90.54**	**83.23**

model's performance as measured by the Dice value on the test set reaches 90.54 when both LICR(w) and LICR(s) are applied simultaneously, confirming the effectiveness of our differentiated training approach. This outcome indicates that optimizing training strategies can significantly enhance model performance. Visualizing the training process, as shown in Figure 4, further illustrates that the performance of the two sub-models in β-FFT is significantly better than the Baseline, highlighting the benefits of our proposed method.

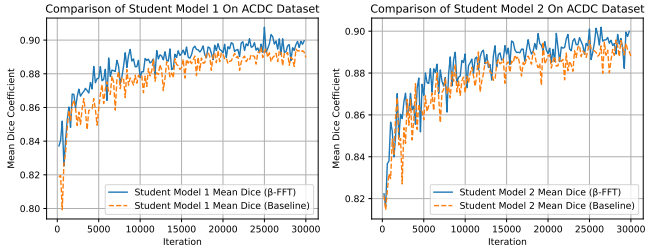

Figure 4. Comparison of the training process between β-FFT and Baseline methods.

5.3.3. Exploring the De-Homogenization Effects of Differentiated Training

To further investigate the impact of differentiated training on model homogenization, we first applied LICR(w) to Student Model 1 and LICR(s) to Student Model 2. This strategy aligns with the initial data augmentation strengths (strong and weak augmentation) assigned to the two student models, constituting a synchronized operation and enabling collaborative training. Experimental results demonstrate that this approach indeed enhances model performance. However, when we removed LICR(w) from Student Model 1 or LICR(s) from Student Model 2, the model performance further improved, indicating that applying LICR to a single sub-model is more effective than applying it to both simultaneously.

Further analysis reveals that when LICR(w) and LICR(s) are applied exclusively to Student Model 1, the model achieves its highest performance. As shown in **Table 6**, the Dice and Jaccard scores on the ACDC validation and test datasets reach **90.77 / 83.81** and **90.54 / 83.23** . In contrast,

Table 6. Ablation study results demonstrating the performance of two models after applying Non-Linear Interpolation, followed by differentiated training using LICR(w) and LICR(s) on the ACDC dataset.The gray row indicates that both Student Model 1 and Student Model 2 undergo simultaneous training, rather than differentiated training.

Differentiation Training		ACDC Validation dataset		ACDC Test dataset	
LICR(w)	LICR(s)	Dice↑	Jaccard↑	Dice ↑	Jaccard ↑
		89.77	82.27	89.96	82.29
Student 1		90.53	83.54	90.27	82.76
Student 1	**Student 2**	90.19	83.01	90.10	82.47
	Student 2	90.20	83.00	90.06	82.47
Student 1	**Student 1**	**90.77**	**83.81**	**90.54**	**83.23**

when both student models undergo simultaneous LICR(w) and LICR(s) training (as indicated by the gray row), performance slightly decreases. This further confirms that differentiated training effectively enhances model performance and mitigates homogenization issues.

5.3.4. Effect of Beta Distribution Sampling in Differentiated Training

We investigated the impact of Beta distribution parameter (a, a) on sample mixing and model performance. By adjusting a, we controlled sample diversity and feature complexity. Ablation results in Figure 5 show that smaller Beta parameters improve Dice and Jaccard scores, peaking at **Beta(0.1, 0.1)** with validation/test scores of 90.77 / 83.81 and 90.54 / 83.23, respectively. This indicates that lower Beta values enhance model generalization and robustness.

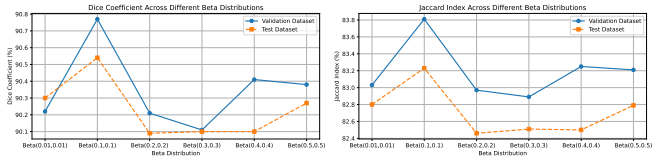

Figure 5. Differentiated training was conducted on Student 1 under the conditions of simultaneously applying nonlinear interpolation and both LICR(w) and LICR(s).

6. Conclusion

In this study, we address the issue of homogenization in co-training from both data and structural perspectives. We distinguish different sub-models using strong and weak augmentations and introduce a nonlinear interpolation method based on the Fast Fourier Transform (FFT) to generate more diverse training samples, thereby enhancing the model's generalization ability. Furthermore, we implement differentiated training by applying additional training to one of the models, effectively reducing homogenization. Extensive ablation experiments validate the effectiveness of our approach, with results demonstrating that β-FFT outperforms current state-of-the-art (SOTA) methods on three public medical image datasets.

Acknowledgment. This research was supported by: The Outstanding Award for Talent Project of the Chinese Academy of Sciences [Grant Number 29J20-052-III]; The Shaanxi Province Technological Innovation Guidance Special Project: Regional Science and Technology Innovation Center, Strategic Scientific and Technological Strength Category (No.2024QY-SZX-26); The Key Project for Teaching Research of the Medical Department of Wuhan University [Grant Number 2024ZD21]; The Key R&D Project of Hubei Province [Grant Number 2023BCB024].

References

[1] Eric Arazo, Diego Ortego, Paul Albert, Noel E O'Connor, and Kevin McGuinness. Pseudo-labeling and confirmation bias in deep semi-supervised learning. In *2020 International joint conference on neural networks (IJCNN)*, pages 1–8. IEEE, 2020. 1, 2

[2] Wenjia Bai, Ozan Oktay, Matthew Sinclair, Hideaki Suzuki, Martin Rajchl, Giacomo Tarroni, Ben Glocker, Andrew King, Paul M Matthews, and Daniel Rueckert. Semi-supervised learning for network-based cardiac mr image segmentation. In *Medical Image Computing and Computer-Assisted Intervention*, pages 253–260. Springer, 2017. 6

[3] Yunhao Bai, Duowen Chen, Qingli Li, Wei Shen, and Yan Wang. Bidirectional copy-paste for semi-supervised medical image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11514–11524, 2023. 2, 3, 6, 7

[4] Hritam Basak and Zhaozheng Yin. Pseudo-label guided contrastive learning for semi-supervised medical image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19786–19797, 2023. 7

[5] Olivier Bernard, Alain Lalande, Clement Zotti, Frederick Cervenansky, Xin Yang, Pheng-Ann Heng, Irem Cetin, Karim Lekadir, Oscar Camara, Miguel Angel Gonzalez Ballester, et al. Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: is the problem solved? *IEEE transactions on medical imaging*, 37 (11):2514–2525, 2018. 6

[6] Baixu Chen, Junguang Jiang, Ximei Wang, Pengfei Wan, Jianmin Wang, and Mingsheng Long. Debiased self-training for semi-supervised learning. *Advances in Neural Information Processing Systems*, 35:32424–32437, 2022. 2

[7] Xinlei Chen, Haoqi Fan, Ross Girshick, and Kaiming He. Improved baselines with momentum contrastive learning. *arXiv preprint arXiv:2003.04297*, 2020. 3

[8] Xiaokang Chen, Yuhui Yuan, Gang Zeng, and Jingdong Wang. Semi-supervised semantic segmentation with cross pseudo supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2613–2622, 2021. 2, 7

[9] Hanyang Chi, Jian Pang, Bingfeng Zhang, and Weifeng Liu. Adaptive bidirectional displacement for semi-supervised medical image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4070–4080, 2024. 3, 6, 7

[10] Kaiwen Cui, Jiaxing Huang, Zhipeng Luo, Gongjie Zhang, Fangneng Zhan, and Shijian Lu. Genco: Generative co-training for generative adversarial networks with limited data. In *Proceedings of the AAAI conference on artificial intelligence*, pages 499–507, 2022. 3

[11] Yves Grandvalet and Yoshua Bengio. Semi-supervised learning by entropy minimization. *Advances in neural information processing systems*, 17, 2004. 1

[12] Nastassya Horlava, Alisa Mironenko, Sebastian Niehaus, Sebastian Wagner, Ingo Roeder, and Nico Scherf. A comparative study of semi-and self-supervised semantic segmentation of biomedical microscopy data. *arXiv preprint arXiv:2011.08076*, 2020. 2

[13] Zhanghan Ke, Daoye Wang, Qiong Yan, Jimmy Ren, and Rynson WH Lau. Dual student: Breaking the limits of the teacher in semi-supervised learning. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6728–6736, 2019. 2

[14] Junnan Li, Caiming Xiong, and Steven CH Hoi. Comatch: Semi-supervised learning with contrastive graph regularization. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9475–9484, 2021. 3

[15] Xiaoxu Li, Yu Peng, and Min Xu. Patch-shuffle-based semi-supervised segmentation of bone computed tomography via consistent learning. *Biomedical Signal Processing and Control*, 80:104239, 2023. 3

[16] Yijiang Li, Xinjiang Wang, Lihe Yang, Litong Feng, Wayne Zhang, and Ying Gao. Diverse cotraining makes strong semi-supervised segmentor. *arXiv preprint arXiv:2308.09281*, 2023. 2, 3

[17] Geert Litjens, Robert Toth, Wendy Van De Ven, Caroline Hoeks, Sjoerd Kerkstra, Bram Van Ginneken, Graham Vincent, Gwenael Guillard, Neil Birbeck, Jindang Zhang, et al. Evaluation of prostate segmentation algorithms for mri: the promise12 challenge. *Medical image analysis*, 18(2):359–373, 2014. 6

[18] Jinhua Liu, Christian Desrosiers, and Yuanfeng Zhou. Semi-supervised medical image segmentation using cross-model pseudo-supervision with shape awareness and local context constraints. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 140–150. Springer, 2022. 7

[19] Xinyu Liu, Wuyang Li, and Yixuan Yuan. Diffrect: Latent diffusion label rectification for semi-supervised medical image segmentation. *arXiv preprint arXiv:2407.09918*, 2024. 6, 7

[20] Xiangde Luo, Jieneng Chen, Tao Song, and Guotai Wang. Semi-supervised medical image segmentation through dual-task consistency. In *Proceedings of the AAAI conference on artificial intelligence*, number 10, pages 8801–8809, 2021. 2, 7

[21] Xiangde Luo, Wenjun Liao, Jieneng Chen, Tao Song, Yinan Chen, Shichuan Zhang, Nianyong Chen, Guotai Wang, and Shaoting Zhang. Efficient semi-supervised gross target volume of nasopharyngeal carcinoma segmentation via uncertainty rectified pyramid consistency. In *Medical Image*

Computing and Computer Assisted Intervention, pages 318–329. Springer, 2021. 7

[22] Xiangde Luo, Minhao Hu, Tao Song, Guotai Wang, and Shaoting Zhang. Semi-supervised medical image segmentation via cross teaching between cnn and transformer. In *International Conference on Medical Imaging with Deep Learning*, pages 820–833. PMLR, 2022. 3, 7

[23] Fei Lyu, Mang Ye, Jonathan Frederik Carlsen, Kenny Erleben, Sune Darkner, and Pong C Yuen. Pseudo-label guided image synthesis for semi-supervised covid-19 pneumonia infection segmentation. *IEEE Transactions on Medical Imaging*, 42(3):797–809, 2022. 2

[24] Takeru Miyato, Shin-ichi Maeda, Masanori Koyama, and Shin Ishii. Virtual adversarial training: a regularization method for supervised and semi-supervised learning. *IEEE transactions on pattern analysis and machine intelligence*, 41(8):1979–1993, 2018. 1

[25] Yassine Ouali, Céline Hudelot, and Myriam Tami. Semi-supervised semantic segmentation with cross-consistency training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12674–12684, 2020. 2, 7

[26] Qianyao Qiang, Bin Zhang, Feiping Nie, and Fei Wang. Multi-view semi-supervised learning with adaptive graph fusion. *Neurocomputing*, 557:126685, 2023. 3

[27] Siyuan Qiao, Wei Shen, Zhishuai Zhang, Bo Wang, and Alan Yuille. Deep co-training for semi-supervised image recognition. In *Proceedings of the european conference on computer vision (eccv)*, pages 135–152, 2018. 3

[28] Ilija Radosavovic, Piotr Dollár, Ross Girshick, Georgia Gkioxari, and Kaiming He. Data distillation: Towards omni-supervised learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4119–4128, 2018. 2

[29] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer, 2015. 7

[30] Chuck Rosenberg, Martial Hebert, and Henry Schneiderman. Semi-supervised self-training of object detection models. In *IEEE Workshop on Applications of Computer Vision*, 2005. 1

[31] Mehdi Sajjadi, Mehran Javanmardi, and Tolga Tasdizen. Regularization with stochastic transformations and perturbations for deep semi-supervised learning. *Advances in neural information processing systems*, 29, 2016. 1

[32] Kihyuk Sohn, David Berthelot, Nicholas Carlini, Zizhao Zhang, Han Zhang, Colin A Raffel, Ekin Dogus Cubuk, Alexey Kurakin, and Chun-Liang Li. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *Advances in neural information processing systems*, 33:596–608, 2020. 2, 3, 7

[33] Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Advances in neural information processing systems*, 30, 2017. 2

[34] Vikas Verma, Kenji Kawaguchi, Alex Lamb, Juho Kannala, Arno Solin, Yoshua Bengio, and David Lopez-Paz. Interpolation consistency training for semi-supervised learning. *Neural Networks*, 145:90–106, 2022. 2, 7

[35] Fuping Wu and Xiahai Zhuang. Minimizing estimated risks on unlabeled data: A new formulation for semi-supervised medical image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(5):6021–6036, 2022. 6

[36] Yicheng Wu, Minfeng Xu, Zongyuan Ge, Jianfei Cai, and Lei Zhang. Semi-supervised left atrium segmentation with mutual consistency training. In *Medical Image Computing and Computer Assisted Intervention*, pages 297–306. Springer, 2021. 7

[37] Yicheng Wu, Zongyuan Ge, Donghao Zhang, Minfeng Xu, Lei Zhang, Yong Xia, and Jianfei Cai. Mutual consistency learning for semi-supervised medical image segmentation. *Medical Image Analysis*, 81:102530, 2022. 7

[38] Yicheng Wu, Zhonghua Wu, Qianyi Wu, Zongyuan Ge, and Jianfei Cai. Exploring smoothness and class-separation for semi-supervised medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 34–43. Springer, 2022. 7

[39] Yingda Xia, Dong Yang, Zhiding Yu, Fengze Liu, Jinzheng Cai, Lequan Yu, Zhuotun Zhu, Daguang Xu, Alan Yuille, and Holger Roth. Uncertainty-aware multi-view co-training for semi-supervised medical image segmentation and domain adaptation. *Medical image analysis*, 65:101766, 2020. 3

[40] Yang Xiu, Xinyi Zheng, Linlin Sun, and Zhuohao Fang. Fremix: Frequency-based mixup for data augmentation. *Wireless Communications and Mobile Computing*, 2022(1): 5323327, 2022. 3

[41] Qinwei Xu, Ruipeng Zhang, Ya Zhang, Yanfeng Wang, and Qi Tian. A fourier-based framework for domain generalization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14383–14392, 2021. 3

[42] Lihe Yang, Wei Zhuo, Lei Qi, Yinghuan Shi, and Yang Gao. St++: Make self-training work better for semi-supervised semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4268–4277, 2022. 3

[43] Yanchao Yang and Stefano Soatto. Fda: Fourier domain adaptation for semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4085–4095, 2020. 3

[44] Boon Peng Yap and Beng Koon Ng. Cut-paste consistency learning for semi-supervised lesion segmentation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 6160–6169, 2023. 3

[45] Lequan Yu, Shujun Wang, Xiaomeng Li, Chi-Wing Fu, and Pheng-Ann Heng. 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, pages 605–613. Springer, 2019. 2, 7

[46] Hongyi Zhang. mixup: Beyond empirical risk minimization. *arXiv preprint arXiv:1710.09412*, 2017. 5

[47] Zhenxi Zhang, Ran Ran, Chunna Tian, Heng Zhou, Xin Li, Fan Yang, and Zhicheng Jiao. Self-aware and cross-sample prototypical learning for semi-supervised medical image segmentation. *arXiv preprint arXiv:2305.16214*, 2023. 7

[48] Zhen Zhao, Zicheng Wang, Longyue Wang, Dian Yu, Yixuan Yuan, and Luping Zhou. Alternate diverse teaching for semi-supervised medical image segmentation. In *European Conference on Computer Vision*, pages 227–243. Springer, 2025. 3, 7

[49] Tianyi Zhou, Shengjie Wang, and Jeff Bilmes. Time-consistent self-supervision for semi-supervised learning. In *International conference on machine learning*, pages 11523–11533. PMLR, 2020. 2

[50] Ye Zhu, Jie Yang, Si-Qi Liu, and Ruimao Zhang. Inherent consistent learning for accurate semi-supervised medical image segmentation. *arXiv preprint arXiv:2303.14175*, 2023. 7

[51] Xiahai Zhuang. Multivariate mixture model for myocardial segmentation combining multi-source images. *IEEE transactions on pattern analysis and machine intelligence*, 41(12): 2933–2946, 2018. 6