

Semi-Supervised Medical Image Segmentation via Frequency Attention with DCT and Data Exchange: The FAS-Net Approach

Dai Lina^{12*}, Md Gapar Md Johar³, Mohammed Hazim Alkawaz⁴

¹School of Graduate Studies, Management and Science University, 40100 Shah Alam, Selangor, Malaysia

²School of Information Technology and Engineering, Guangzhou College of Commerce, Guangzhou, China

³Software Engineering and Digital Innovation Center, Management and Science University, 40100 Shah Alam, Selangor, Malaysia

⁴Department of Computer Science, College of Education for Pure Science, University of Mosul, Mosul, Nineveh, Iraq

dailn526@163.com, mdgapar@msu.edu.my, mohammed.ameen@uomosul.edu.iq

Abstract. In the domain of medical imaging, precise segmentation of organs or pathological regions is vital for disease diagnosis and various clinical practices. The inherent complexity of medical images poses challenges in procuring pixel-level exact annotations. Consequently, researchers have explored semi-supervised approaches to medical image segmentation, needing minimal pixel labeling for effective segmentation. Despite noteworthy advancements in this field, there remains potential for boosting model performance. This study proposes a new FAS-Net architecture for enhanced semi-supervised medical image segmentation. This model is based on the Mean Teacher structure, and a frequency attention guidance module is designed based on the discrete cosine transform. An input data augmentation strategy for data exchange is introduced to improve the efficiency of semi-supervised medical image segmentation. The frequency attention module empowers the segmentation network to mitigate non-relevant image regions while accentuating critical features. The data exchange strategy counteracts data distribution inconsistencies and empirical discrepancies by randomly selecting and interchanging segments between pairs of images. The efficacy of the FAS-Net model, when combined with the U-Net segmentation architecture, is demonstrated through validation on two datasets: ACDC for cardiac condition assessment and BraTS2019 for brain tumor assessment. Experimental outcomes demonstrate that FAS-Net surpasses existing state-of-the-art semi-supervised medical image segmentation techniques across these datasets.

Keywords: Attention, U-Net, Semi-Supervised, Medical Image Segmentation, Data Exchange

1. Introduction

Segmentation of medical images is an indispensable task within the domain of computer vision, especially when dealing with sophisticated medical imaging techniques such as Computed Tomography (CT) scans and Magnetic Resonance Imaging (MRI). These technologies hold significant value for healthcare practitioners in detecting numerous health issues. Doctors choose appropriate imaging modalities depending on specific clinical needs to procure detailed anatomical and pathological information about patients, thereby enabling precise diagnoses. Image segmentation in medicine stands as a core element and critical phase in the creation of computer-assisted diagnostic systems, significantly impacting disease detection, planning of therapeutic approaches, and fostering advancements in biomedical research (Devunooru, Alsadoon, Chandana, Beg, & Computing, 2021; Qureshi et al., 2023).

Progress in deep learning has sparked extensive adoption of diverse deep learning methodologies for segmenting medical images. Among these, Convolutional Neural Networks (CNNs) have exhibited unparalleled effectiveness in a myriad of medical image segmentation tasks, such as the precise outlining of neural structures, polyps, the liver, and the pancreas (Gul et al., 2022; M.-L. Huang, Wu, & Control, 2022; Yeung, Sala, Schönlieb, Rundo, & medicine, 2021). The development has notably improved the effectiveness and precision in tackling intricate medical image segmentation chores (Patil, Deore, & Computing, 2013). Despite the considerable strides made by fully supervised learning approaches in recent times (Han, Jian, & Wang, 2022; Jiang et al., 2023; X. Liu, Song, Liu, & Zhang, 2021), these techniques heavily depend on an abundance of meticulously labeled data, a resource that is both financially burdensome and time-intensive to obtain. While non-experts can annotate pixel-level labels for natural image semantic segmentation tasks, in medical image segmentation, collecting precise pixel-level labels requires considerable time commitment from medical experts, making the acquisition of medical image annotations particularly expensive and time-consuming. Additionally, segmentation results can vary due to the subjective nature of doctors' clinical experience, thus reducing the reliability of the segmentations. In recent times, the domain of semi-supervised medical image segmentation has garnered substantial interest, primarily due to its emphasis on lessening annotation expenses and enhancing the model's capacity to generalize from scarce pixel-wise annotated data.

This study aims to develop an enhanced semi-supervised medical image segmentation model by introducing a frequency attention module based on the Discrete Cosine Transform (DCT) and incorporating a data exchange strategy to mitigate the distribution mismatch between labeled and unlabeled data. The proposed approach aims to improve segmentation performance while minimizing the need for extensive pixel-level annotations.

Typically, semi-supervised image segmentation techniques can be broadly categorized into two primary groups. One of these is the pseudo-label semi-supervised learning approach, which initially predicts labels for unlabeled images using a segmentation model to generate pseudo-labels. These artificially labeled images are subsequently incorporated as additional instances for subsequent training. This self-improving learning approach enables the pseudo-label method to augment the training dataset, thereby serving as a viable technique for enhancing the efficacy of semi-supervised learning models (Basak & Yin, 2023; Xiaoyan et al., 2022). Nevertheless, the segmentation outcomes derived from the self-training pseudo-label method are frequently subpar due to the inconsistent nature of the predicted pseudo-labels. Another approach falls under the realm of semi-supervised learning with unsupervised regularization, which involves integrating unlabeled images with labeled data to develop segmentation models that encompass unsupervised regularization. This method enhances model efficacy by reducing the disparity between the outcomes of two forward passes employing distinct regularization techniques. The Mean Teacher technique is formulated on the principle of consistency regularization (A. Tarvainen & H. J. A. i. n. i. p. s. Valpola, 2017). Enhanced

learning efficiency and model performance are attained by concurrently managing a student model and a mentor model, thereby maximizing the exploitation of unlabeled data sources. Within the Mean Teacher structure, the architectural layout of both models remains identical, although they possess distinct parameter configurations. During the training phase, the student model's parameters are updated utilizing gradient descent methodologies. On the other hand, the teacher model enhances its parameters through the application of an Exponential Moving Average (EMA) approach, integrating the mean parameters of the student model (A. Tarvainen & H. J. A. i. n. i. p. s. Valpola, 2017). This design allows the teacher model to accumulate and smooth out the knowledge learned by the student model, thus providing stable and accurate guidance throughout the training process.

Several medical image segmentation models utilizing the Mean-Teacher framework have been suggested in the literature. Building upon the same structure, Wang et al. (2022) integrated additional objectives: the restoration of foreground and background elements to facilitate semantic data retrieval, along with the prediction of Signed Distance Fields (SDFs) to strengthen shape regulations, consequently enhancing segmentation performance (K. Wang et al., 2022), Mendel and colleagues (2023) introduced a novel semi-supervised learning approach known as the "Error-Correcting Mean-Teacher" to tackle the challenge of insufficient labeled data in the context of medical image segmentation. This method integrates a novel error-correction framework into the conventional Mean Teacher model, thereby improving its capacity to analyze unlabeled data (Mendel et al., 2023). Simultaneously, in that same year, Wang and his team proposed an innovative Mean Teacher framework with a regularization-centric approach, specifically designed for medical image segmentation. The Mean Teacher framework was enhanced with an integrated virtual adversarial training approach to elevate its segmentation proficiency (Q. Wang et al., 2022).

The optimization of semi-supervised medical image segmentation algorithms has consistently been a primary area of investigation. Toward this end, an improved model referred to as FAS-Net has been devised to augment the segmentation proficiency. FAS Net uses the Mean Teacher semi-supervised framework and adds a channel attention module based on discrete cosine transform to the segmentation networks of Teachers and Students. This module enables the segmentation network to reduce irrelevant image regions while emphasizing key features. The introduction of data exchange strategy in FAS-Net compensates for inconsistent data distribution and empirical differences by randomly selecting and exchanging segments between image pairs, thereby improving segmentation performance.

In summary, the main contributions of this paper are as follows:

- This study designed a frequency domain attention mechanism based on discrete cosine transform, which was integrated into the Mean-Teacher architecture to significantly enhance the model's ability to recognize key areas or features in images, thereby effectively improving segmentation performance. Additionally, combined with the U-Net architecture, the effectiveness and practicality of this frequency domain attention mechanism within the Mean-Teacher framework were validated.
- To further enhance the model's generalizability and reduce overfitting, this study introduced a cross-copy-paste strategy for optimizing the input data handling in the Student network. By combining cross-copy-paste with consistency regularization rules, this method effectively narrowed the empirical distribution gap between labeled and unlabeled data.
- Application results on the ACDC dataset for cardiac disease monitoring and the BraTs2019 dataset for brain tumor monitoring demonstrated that the model developed in this study surpasses several advanced semi-supervised medical image segmentation methods in terms of segmentation performance.

2. Related Work

2.1. Medical image segmentation

Medical image segmentation plays a crucial role in the field of medical image analysis, significantly aiding doctors in making rapid and accurate diagnoses by precisely identifying lesion boundaries. The primary goal of this technique is to correctly classify each pixel in medical images into its corresponding category. As an indispensable part of intelligent medical analysis, medical image segmentation is extensively used in various aspects such as lesion target identification, quantitative measurement of medical objects, and disease assessment (Qureshi et al., 2023). Over the past few decades, with the rapid advancement of deep learning technologies, multi-layer perceptual networks applied to medical image segmentation have significantly improved their ability to handle complex structures, successfully achieving efficient segmentation of various medical images including liver tumors, brain tumors, and pulmonary nodules (Y. Fu et al., 2021).

The domain of medical image segmentation has undergone a transformative shift with the advent of Neural Networks, particularly Convolutional Neural Networks (CNNs). Since LeCun and his colleagues first trained CNNs using backpropagation in 1989 (LeCun et al., 1989), the potential of deep learning has been gradually uncovered, particularly in image recognition and analysis. In medical image segmentation tasks, deep learning models are capable of automatically detecting and learning complex features from vast amounts of medical imaging data, surpassing the capabilities of traditional machine learning methods. Early segmentation algorithms, such as UNet, FCN, ResNet, and DeepLab (Cheng et al., 2022; C. Li et al., 2020; Sun, Peng, Guo, & Li, 2021; J. Wang, Liu, & Biomedicine, 2021), primarily relied on supervised learning to acquire stable features. Although these methods have improved performance, manually annotating lesion locations is not only time-consuming but also heavily dependent on the professional judgment of physicians. Obtaining reliable pixel-level labels from experienced experts is both costly and challenging, adding to the difficulties of medical image segmentation.

2.2. Semi supervised medical image segmentation

The fundamental distinction between semi-supervised and supervised learning stems from their approach to unlabeled data during training. In the domain of machine learning, semi-supervised techniques harness the potential of a limited set of labeled data in conjunction with an abundant amount of unlabeled data, unlike supervised learning that heavily depends on labeled instances. The Mean Teacher model, a notable illustration of such an approach, was proposed by Antti Tarvainen and Harri Valpola in their 2017 work (Tarvainen & Valpola, 2017). In scenarios where labeled data is scarce, the efficiency of semi-supervised learning is augmented by deploying a teacher model, which is created through the exponential moving average (EMA) of the student model's weights. This technique leverages unlabeled data through consistency loss to boost the model's generalization abilities. The core of the Mean Teacher method consists of two distinct models: a student model that continually evolves during training, and a teacher model, which serves as the EMA-weighted average of the student model (Tarvainen & Valpola, 2017). The teacher model acts as a consolidator and enhancer of the knowledge acquired by the student model, ensuring more stable and reliable guidance throughout the training process. A visual illustration of the Mean Teacher architecture is outlined in Fig. 1.

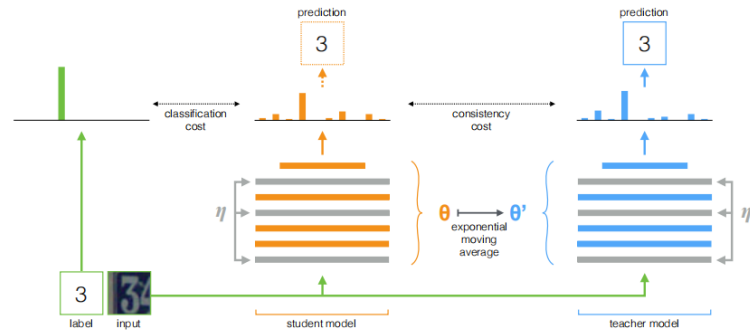


Fig. 1: The Mean Teacher method (Tarvainen & Valpola, 2017)

Recently, significant attention has been directed towards the application of the Mean-Teacher architecture for semi-supervised medical image segmentation. Wang et al. introduced supplementary objectives - the reconstruction of both foreground and background for semantic content retrieval, along with Signed Distance Field prediction to ensure shape regularity, thus enhancing their methodology in semi-supervised medical image segmentation as reported in (K. Wang et al., 2021). In response to biases present in the teacher model due to limited labeled data, a triple uncertainty-guided framework was devised to enhance the accuracy of pseudo-label generation by the teacher model (K. Wang et al., 2021). Furthermore, due to the potential issue of noisy predictions generated by teacher models in conventional Mean-Teacher frameworks for unlabeled data, which can impede the learning process of the student model, Wen and Ye introduced an enhanced semi-supervised approach for object detection in cell identification. This method is built upon an improved Mean-Teacher framework, aiming to notably improve the precision and effectiveness of cell detection within histopathological images when faced with constraints in labeling data (Wen & Ye, 2024).

2.3.U-Net in Medical Image Segmentation

In 2015, Ronneberger and colleagues developed the U-Net model, specifically designed for medical image segmentation (Ronneberger, Fischer, & Brox, 2015). This model, with its innovative network structure, significantly enhanced segmentation accuracy (Ronneberger et al., 2015). The U-Net model features a symmetrical contracting path and an expansive path, effectively integrating contextual and positional information in images, making it particularly suitable for handling small structures in medical images. Additionally, U-Net utilizes skip connections to preserve information in deeper layers of the network, greatly improving segmentation outcomes. Since its introduction, U-Net has solidified its position as a widely adopted method for medical image segmentation, playing a pivotal role in various medical image analysis tasks. Researchers have continually endeavored to improve upon the initial U-Net design, with one notable advancement being the creation of U-Net3+ by Huang et al., which incorporates sophisticated skip connection mechanisms (H. Huang et al., 2020), while Xiang et al. introduced the Bi-directional O-shaped Network (BiO-Net) which integrates skip connections along the decoder path and establishes an O-shaped cyclic pathway connecting the encoder and decoder (Xiang et al., 2020). This O-shaped reasoning path can be recursively iterated to enhance performance without introducing additional parameters, helping to prevent overfitting. Liu and colleagues developed a segmentation network named TransUNet+ (Y. Liu, Wang, Chen, Huangliang, & Zhang, 2022), featuring redesigned skip connections and a feature enhancement module. This module utilizes the column vectors of fraction matrices to establish relationships between features across images, thereby effectively enhancing feature representation. Moreover, Özgün Çiçek presented the 3D U-Net model, which expanded the capabilities of the U-Net architecture to process three-dimensional datasets, thereby enhancing its utility for intricate volumetric medical imaging tasks like analyzing CT and MRI scans (Çiçek, Abdulkadir, Lienkamp,

Brox, & Ronneberger, 2016). The 3D U-Net has made a substantial contribution to the field of medical image analysis due to its advanced contextual capture features and precise volumetric segmentation performance (Çiçek et al., 2016). Neural network architectures not only augment the automatization and accuracy in medical image partitioning, they also function as an indispensable asset for disease identification, contributing significantly to the realms of tailored medicine and precise therapeutic approaches.

2.4. Attention Mechanisms in Deep Learning

Attention mechanisms have emerged as a critical technology in augmenting the efficacy of medical image segmentation. By directing attention to important regions within the image, these mechanisms notably enhance the precision and effectiveness of analyzing intricate medical image datasets. In response to the constraints posed by fully convolutional networks, such as restricted perception fields and the incapacity to capture distant information, scholars have integrated attention mechanisms into semantic segmentation networks featuring multi-scale inputs to alleviate local neighborhood limitations (C. Li et al., 2020). In order to enhance the incorporation of contextual information, Fu and his team introduced spatial and channel attention modules. Designed to capture semantic connections in both spatial and channel aspects, these modules employ a self-attention scheme. This scheme facilitates the adaptive assimilation of regional characteristics while considering their holistic interdependencies (J. Fu et al., 2019). In the year 2023, Chen et al. presented a novel Self-Aware Attention (SAA) mechanism, consisting of Transformer-based Self-Attention (TSA) and Global Spatial Attention (GSA) elements. The primary objective of this module is to enhance the comprehension of long-range dependencies within the encoder's feature representations. Attention mechanisms, known for their ability to improve neural network focus on important regions, have demonstrated effectiveness in a range of medical image segmentation applications (Chen, Liu, Zhang, Lu, & Kong, 2023). Future research may continue to explore how to further integrate and optimize attention modules to achieve more precise segmentation outcomes in a broader range of medical applications.

2.5. Discrete Cosine Transform in computer vision

The Discrete Cosine Transform (DCT) is a widely used technique for converting signals or images from the time domain (or spatial domain) to the frequency domain (Ahmed, Natarajan, & Rao, 1974). Frequency analysis is a natural and effective tool for analyzing time series data. Incorporating frequency information into time series models is not only important but also intuitive. By applying the DCT, feature maps can be transformed into frequency domain images, which helps to reveal the fundamental frequency components of the images in the frequency domain. Using the frequency domain images obtained through the DCT, combined with attention mechanisms, allows for more effective capture and emphasis of critical features in the feature vectors, thus enhancing the outcomes of image processing and analysis.

The Frequency Domain Conversion technique known as the Discrete Cosine Transform (DCT) plays a substantial role in the realm of computer vision, particularly in transforming spatial domain components of RGB images into frequency domain components. Xu and others explored methods for object detection and segmentation in the DCT frequency domain after converting high-resolution RGB images (Xu et al., 2020). Shen and his team developed a DCT-based mask, DCT-Mask, that can seamlessly incorporate into the majority of pixel-focused segmentation algorithms without requiring preliminary processing or prior training, with negligible effects on computational speed (Shen et al., 2021). This method shows significant improvements, especially when dealing with high-quality annotations and complex architectures. Wang and his colleagues leveraged the properties of DCT to reduce energy loss during downsampling, creating new representations for ground truth masks (P. Wang & Zhao, 2022). Their model, through DCT transformation, can more efficiently process and compress mask data, reducing information loss while maintaining real-time segmentation speed. Cao's

group improved the resilience and visual integrity of watermarked images by integrating a channel attention mechanism within the DCT domain and incorporating a noise layer to steer the watermarking model, enabling the generated watermarked images to effectively resist various distortions (Cao et al., 2024).

2.6. Copy and Paste Data Augmentation Techniques

The method of copy-paste, serving as a valuable data augmentation strategy, showcases distinctive and robust functionalities within the realm of computer vision, specifically in the domain of image processing. This approach entails the identification and extraction of particular segments within an image, often objects or regions of significance, and relocating them to a fresh location within the same image or a separate one. Through this procedure, the dataset's variability is artificially expanded, consequently bolstering the efficacy and productivity of model training.

In recent times, the utilization of the copy-paste method has become increasingly prevalent in the realm of medical image segmentation. This technique has demonstrated significant value, particularly in semi-supervised learning and scenarios with limited data, owing to its distinctive capacity for data augmentation. This technique enhances the diversity of datasets by copying specific regions of interest or important parts within an image and pasting them into new locations within the same or different images, significantly improving the effectiveness of model training. For instance, in 2022, Fan and colleagues introduced a novel technique called Dynamic Cross-set Copy-Paste (DCSCP) (Fan, Gao, Jin, & Jiang, 2022), which not only expands the diversity of training samples but also effectively mitigates issues like mismatched data distributions and class imbalances. The DCSCP technique enhances the model's adaptability and generalization ability across different scenes by dynamically selecting key content in images and copying it into various backgrounds. Additionally, Tu and others in 2022 utilized the copy-paste technique to process both labeled and unlabeled data simultaneously, effectively solving the problem of missing critical information when handling these two types of data independently. Their approach ensures that the prior knowledge learned from labeled samples is effectively transferred to the learning process of unlabeled samples by sharing information between the two. Similarly, Wang and his team's CP2 method, which involves copying and pasting randomly cropped foreground areas of images into different backgrounds and pre-training a semantic segmentation model, not only learns to recognize features at the image level but also pays detailed attention to each pixel, providing richer and more accurate information for dense prediction tasks.

In recent years, the copy-paste technique has been widely adopted in the field of medical image segmentation due to its excellent data augmentation capabilities, particularly in semi-supervised learning environments and situations of data scarcity. For instance, the Dynamic Cross-set Copy-Paste (DCSCP) technology developed by Fan et al. (2022) not only significantly enriched the diversity of training samples but also effectively alleviated issues of data distribution mismatches and class imbalances (Fan et al., 2022). This technique enhances the adaptability and generalization of models across diverse scenarios by dynamically selecting and copying key content from images into various backgrounds. Furthermore, Tu et al. (2022) addressed the potential loss of information that can occur when labeled and unlabeled data are processed separately by using the copy-paste technique to handle both concurrently (Tu et al., 2022). Additionally, the CP2 method implemented by Wang and colleagues, which involves copying and pasting the foreground regions of images onto different backgrounds for pre-training, enables the model to not only recognize features at the image level but also pay close attention to each pixel, providing richer and more accurate data support for complex prediction tasks (F. Wang, Wang, Wei, Yuille, & Shen, 2022). In 2023, Bai and colleagues introduced an innovative bidirectional copy-paste method for semi-supervised medical image segmentation (Bai, Chen, Li, Shen, & Wang, 2023). This method effectively enhances the consistency and accuracy of the learning outcomes by reducing the differences in empirical distributions between labeled and

unlabeled data during the consistency learning process (Bai et al., 2023). The innovative application of the copy-paste technique in medical image segmentation not only addresses issues of data insufficiency and imbalance but also enhances data diversity and the robustness of model training, thereby advancing medical image analysis technology. The investigations emphasize the promising role of the copy-paste methodology in enhancing the efficiency of semi-supervised learning and the accuracy of medical image partitioning, revealing its extensive adaptability and potential for wider implementation.

While existing semi-supervised medical image segmentation methods have shown promising results, there is a need for approaches that can effectively leverage both spatial and frequency domain information to enhance feature extraction and segmentation performance. Furthermore, addressing the distribution mismatch between labeled and unlabeled data remains a challenge in semi-supervised learning. The research objective of this study is to improve the performance of segmentation and solve the problem of mismatch between labeled and unlabeled data in semi supervised scenarios. This study aims to improve the performance of semi supervised medical image segmentation by focusing on the attention mechanism based on discrete cosine transformation and the problem of data mismatch in semi supervised segmentation.

3. Method

In order to enhance the performance and precision of semi-supervised medical image segmentation models, this research has developed a frequency domain attention module utilizing Discrete Cosine Transform (DCT) and introduced a mixed data input approach for initial data augmentation. The incorporation of this frequency domain attention mechanism allows the model to concentrate on key regions within the image, leading to a notable enhancement in segmentation accuracy and complexity. This mechanism is particularly well-suited for analyzing intricate medical images, such as diverse tissue types and lesions, where precise detail recognition is essential for subsequent diagnostic and therapeutic purposes. The integration of diverse data inputs serves to mitigate disparities in distribution between annotated and unlabelled datasets. In order to substantiate the proposed approach's efficacy, the experiment employs Mean-Teacher as the fundamental network structure, upon which the U-Net network functions as the core component.

The training methodology employed in this research comprises three primary stages. Initially, the study utilizes existing labeled data to pre-train a supervised learning model, which then functions as a teacher network. The instructor network utilizes its accumulated wisdom to generate accurate proxy labels for untagged images. Subsequently, within each iterative phase, the disciple network's parameters undergo refinement through the Stochastic Gradient Descent method. Finally, the mentor network's parameters are enhanced employing an Exponential Moving Average approach, guaranteeing a dynamic realignment of the mentor's parameters to mirror the most up-to-date learning progress and the state of the disciple network. The architecture of this method is designed to promote consistent improvement and robustness throughout successive iterations.

Our dataset X consists of two parts: X^l and X^u . X^l is a labeled dataset, accompanied by its corresponding set of labels Y^l . X^u is an unlabeled dataset. X^l contains N^l labeled images, while X^u comprises N^u unlabeled images, $N^l \ll N^u$. Figure 2 displays the overall process framework of this study. From the dataset X , we arbitrarily choose two unlabeled images x_i^u, x_j^u , and two labeled images x_i^l, x_j^l . x_i^u, x_j^u are input into the Teacher network to generate pseudo-labels Y_i^u and Y_j^u , while data mixing operations using x_i^u, x_j^u, x_i^l and x_j^l produce mixed data x_i^m and x_j^m , which are then input into the Student network to produce pseudo-labels Y_i^p and Y_j^p . The mixed labels Y_i^m and Y_j^m generated using Y_i^u, Y_j^u and Group Truth Y_i^l and Y_j^l , Compare Y_i^m and Y_j^m with Y_i^p and Y_j^p , and train the model using a loss function. The identical backbone structure is employed in both Teacher and Student networks, a refined U-Net design that integrates a frequency domain attention module rooted in DCT conversion.

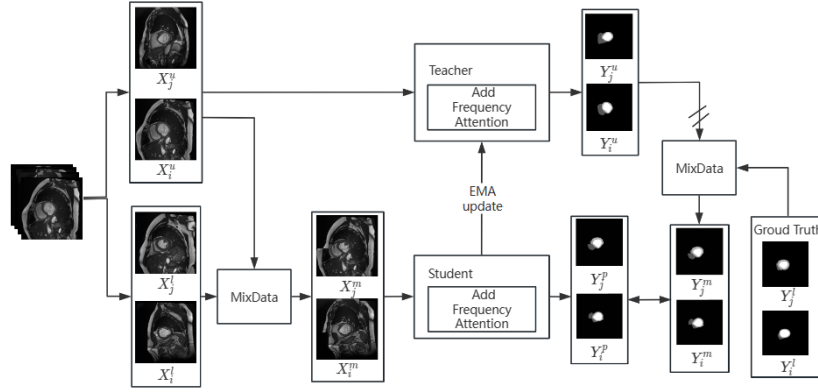


Fig. 2: The process of our proposed FAS-Net framework.

3.1. Frequency Attention Module Based on DCT

Attention mechanisms effectively reduce the model's response to features irrelevant to the task, focusing more on information in areas of interest, thereby extracting image features more precisely [9-10]. Discrete Cosine Transform (DCT) is a technique that converts signals or images from the time domain (or spatial domain) to the frequency domain (Ahmed et al., 1974). After DCT transformation, frequency-domain images can be more effectively integrated with attention mechanisms to capture essential features within the feature vectors.

This study combines DCT with channel attention mechanisms, applied within both the Teacher and Student network architectures, to enhance their feature extraction capabilities. The DCT-based attention mechanism's efficacy is ascertained by employing it within the Mean-Teacher framework, with both the Teacher and Student networks leveraging the U-Net architecture as their foundational segmentation models. Fig. 3 outlines the design of the DCT-based attention mechanism.

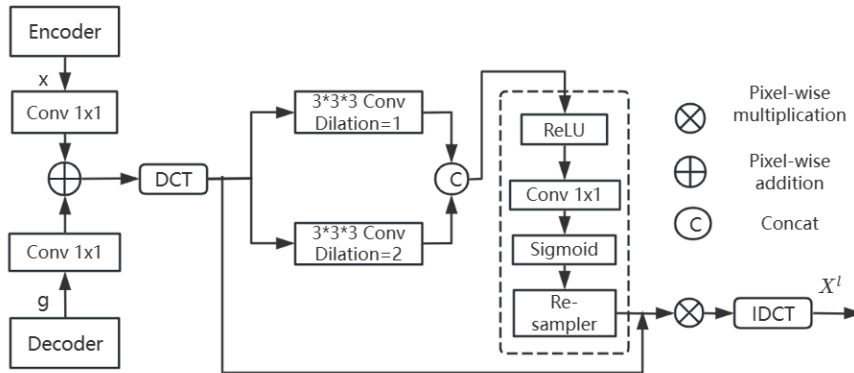


Fig. 3: The process of designing a frequency domain attention module based on DCT

In this study, X represents the input features for each layer of the network, while g is the next layer's input features generated during the decoder process. X and g are first processed through $1 \times 1 \times 1$ convolution operations to align the number of channels, and then they are added element-wise to obtain a merged feature map. This feature map is subsequently processed through a DCT-based discrete cosine transformation to obtain the transformed feature map f_{dct} . This f_{dct} is then processed through two parallel 3×3 convolution layers, each with different dilation rates, to capture multi-scale information. The outputs of these convolution layers are concatenated to form $f_{multi-scale}$, as shown in Equation 3-1. $f_{multi-scale}$ is then activated using the ReLU function and processed through a $1 \times 1 \times 1$ convolution, ultimately activated by the Sigmoid function to produce the attention coefficient α ,

as shown in Equation 3-2. This attention coefficient α multiplies element-wise with the DCT-transformed features at corresponding positions, and then the features are converted back to the spatial domain using the Inverse Discrete Cosine Transform (IDCT), resulting in the updated feature map X^l .

$$f_{\text{multi-scale}} = \text{Concat}(\text{Conv}_{3 \times 3}^{d=1}(f_{\text{dct}}), \text{Conv}_{3 \times 3}^{d=2}(f_{\text{dct}})) \quad (3-1)$$

$$\alpha = \text{Re-sampler}(\sigma(\text{Conv}_{1 \times 1}(\text{ReLU}(\text{Conv}_{1 \times 1}(f_{\text{multi-scale}})))))) \quad (3-2)$$

3.2.MixData

In order to rectify the discrepancy in data distribution between labeled and unlabeled datasets in the context of semi-supervised learning models, we adopted the technique of bidirectional copy-paste, as conceptualized by Bai et al. (2023), which facilitates integration of labeled and unlabeled data within the model architecture itself (Bai et al., 2023). Fig. 4 shows the data exchange framework introduced in our study.

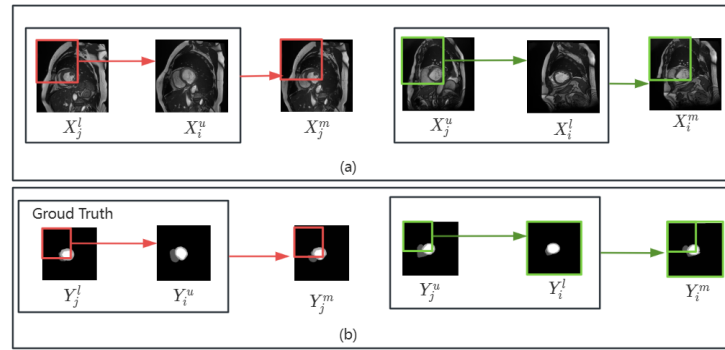


Fig. 4: Data exchange algorithm framework. (a) is the input data exchange process, and (b) is the corresponding label exchange process.

Data mixing process comprises two parts. The first part generates input data for the student network, x_i^m and x_j^m . We randomly select two labeled images x_i^l , x_j^l and two unlabeled images x_i^u , x_j^u , and then proceed with the data mixing process as illustrated in Figure 4(a). In this process, we select a specific region w from the image x_j^l as the foreground and copy it onto the background of the image x_i^u , creating a new image x_j^m . Similarly, we take the region w from the image x_j^u as the foreground and copy it onto the background of the image x_i^l , creating a new image x_i^m . These newly created images, x_i^m and x_j^m , are then fed into the student network for training. The selected location is randomly selected, and the size of the selected area is the specified size. The specific experimental process will be introduced later. Select and record the location and area size each time for the corresponding generation of mixed verification labels.

The second part generates mixed supervision labels using the same method applied for creating the mixed images, Generate mixed verification labels using the process shown in Figure 4(b). Based on the positions and sizes of the regions selected for creating x_i^m and x_j^m , similar copy operations are performed on the label images Y_j^l and Y_j^u to create new label images Y_j^m and Y_i^m . These mixed labels, Y_i^m and Y_j^m , serve as supervision to guide the predictions of the student network for x_i^m and x_j^m .

3.3.Loss Function

Within this study, the student network processes input imagery which integrates both tagged and untagged samples. It is commonly assumed that the accurate annotations given to the labeled images surpass the makeshift labels of their untagged counterparts in terms of reliability. To manage the influence of untagged pixel data on the loss calculation, a weighting element, referred to as γ , is introduced. Specifically, the loss functions for the mixed images x_i^m and x_j^m , $L_{i\text{loss}}$ and $L_{j\text{loss}}$, are defined as follows:

$$L_{iloss} = L_{con}^l (Y_i^p, Y_i^m) + \gamma L_{con}^u (Y_i^p, Y_i^m) \quad (3-3)$$

$$L_{jloss} = L_{con}^l (Y_j^p, Y_j^m) + \gamma L_{con}^u (Y_j^p, Y_j^m) \quad (3-4)$$

$$L_{all} = L_{iloss} + L_{jloss} \quad (3-5)$$

In this context, L_{con} denotes the integration of Dice loss with Cross-entropy loss. L_{iloss} and L_{jloss} represent the losses of x_i^m and x_j^m , respectively. The total loss function is the sum of L_{iloss} and L_{jloss} . During every iteration, the parameters within the student network are upgraded through the application of the loss function-driven Stochastic Gradient Descent algorithm. During the experiment, we set the values of γ to be 0.1, 0.5, 1.0, 1.5, and 2, respectively. This is to verify that the Dice and Jaccard obtained when γ is set to 0.1, 1.0, 1.5, and 2.0 are lower than those obtained when γ is set to 0.5. The main reason for this problem is that the loss function of unlabeled images is not particularly accurate compared to the loss function of labeled data, so the weight of the loss result of unlabeled labeled data is too high, which can affect the performance of the model.

4. Experiments:

4.1.Dataset

In order to assess the efficacy of our approach, we conducted thorough assessments using two publicly accessible datasets.

The ACDC dataset, an abbreviation for the Automated Cardiac Diagnosis Challenge, is specifically designed for the purpose of identifying heart diseases. It includes cardiac scan data from a total of 100 patients. The dataset features segmentation of four key components: the background, the right ventricle, the left ventricle, and the myocardium. For the smooth conduct of our research, we divided this data into three segments - training, validation, and testing. The breakdown consists of 70 patients in the training set, 10 patients in the validation set, and 20 patients in the testing set.

The BraTS 2019 dataset, referred to as the Brain Tumor Segmentation Challenge data collection from 2019, is designed for the purpose of segmenting brain tumors using Multimodal Magnetic Resonance Imaging (MRI) scans. It encompasses a total of 335 MRI scans, which consist of 259 instances of high-grade glioma (HGG) and 76 cases of low-grade glioma (LGG). The dataset includes four distinct imaging modalities, specifically T1, T1 with contrast enhancement (T1CE), T2, and Fluid Attenuated Inversion Recovery (Flair). Given our previous experience, we decided to focus on the Flair modality for our research endeavors.

4.2.Implementation Details

The experiments in this paper were conducted using an NVIDIA GeForce RTX 3090 graphics card and the PyTorch-GPU software environment. In our experimental setup, we performed 2D image segmentation on slices obtained from each 3D volume and implemented basic data augmentation techniques, including image rotation and flipping. The batch size for the experiments was set to 24, with a learning rate of 0.01. To enhance processing efficiency and accuracy, we only retained slices that contained lesions, excluding all pure background slices. Below are the detailed processing steps for the dataset:

ACDC dataset: In this study, the training dataset consisted of 136 labeled samples and 1,176 unlabeled samples, while the test set and validation set were composed of 20 and 40 samples, respectively. Throughout the training process, the input patch size for the 2D slices was established at 256×256, and the dimensions of the mixed data regions were defined as 170×170. The batch size, pre-training iterations, and self-training iterations were configured at 24,000, 10,000, and 30,000, respectively.

BraTS2019 dataset: In this study, the training dataset comprised 3,200 labeled samples and 28,976 unlabeled samples, while the test set consisted of 60 samples and the validation set comprised 25 samples. Throughout the training phase, the input patch size for the 2D slices was set at 256×256,

with the mixed data regions sized at 170×170. Parameters such as batch size, number of pre-training iterations, and number of self-training iterations were configured at 24,000, 10,000, and 30,000, respectively.

4.3. Evaluation Metrics

In this research, four assessment criteria were chosen, namely the Dice coefficient, Jaccard index, HD95, and ASD, to evaluate the effectiveness of the segmentation model.

The Dice coefficient is utilized to assess the similarity between two datasets, primarily in binary segmentation assignments. It is a metric that varies between 0 and 1, with a score of 1 denoting complete similarity and a score of 0 denoting no similarity.

The Jaccard index, also referred to as the Intersection over Union, serves as an alternative measure for assessing the likeness between the predicted and true segmentation. Analogous to the Dice coefficient, the Jaccard index spans from 0 to 1, where a value of 1 signifies complete alignment and 0 indicates no alignment.

The Hausdorff distance acts as a measuring tool that defines the maximum disparity between two sets of points. Within the domain of medical image segmentation, the 95th percentile Hausdorff Distance (HD95) is commonly used to estimate the greatest discrepancy between predicted and actual contour boundaries. To strengthen the reliability of assessment, it neglects the upper 5% of severe outliers. This criterion evaluates predictive accuracy under severe conditions, with lower values indicating enhanced segmentation precision.

The Average Surface Distance (ASD), a quantifiable evaluation tool, gauges segmentation precision by estimating the mean disparity between the estimated and authentic segmentation boundaries. It calculates the average shortest distance from every point on the predicted contour to its closest point on the actual contour. Reduced ASD values indicate an enhanced closeness between the predicted segmentation results and the factual reference.

4.4. Comparison With State-of-the-Art Methods

The performance averages of diverse models on the ACDC dataset with a labeling rate of 10% are depicted in Table 1. These models encompass UA-MT(Yu, Wang, Li, Fu, & Heng, 2019)、SASSNet(Li, Zhang, & He, 2020)、URPC(Luo et al., 2021)、MC-Net(Wu, Xu, Ge, Cai, & Zhang, 2021)、SS-Net(Wu, Wu, Wu, Ge, & Cai, 2022)、BCP(Bai et al., 2023), and our proprietary model. The table conclusively demonstrates that our algorithm surpasses the others in terms of effectiveness. The experimental results on the ACDC dataset demonstrate the effectiveness of the frequency-attention module proposed in this study in capturing relevant features, and the effectiveness of the data-exchange strategy in increasing the diversity of the data and improving the generalization ability of the model.

Table 2 displays the performance comparison of several algorithms on the FLAIR modality of the BraTS 2019 dataset. Methods such as MT(A. Tarvainen & H. J. a. p. a. Valpola, 2017)、UAMT(Yu et al., 2019)、URPC(Luo et al., 2022)、MCNet(Xiong et al., 2021)、SDMI-Net(Gai et al., 2024) use 3D segmentation techniques, whereas our algorithm employs 2D segmentation. Despite this, our algorithm surpasses all the aforementioned methods in terms of Dice and Jaccard indexes, demonstrating its superiority. However, the other two indicators HD95 and ASD are slightly insufficient on the other two indicators. Compared with cardiac images, the morphology, size and location of brain tumors are more variable, and the boundaries between brain tumors and normal brain tissue may be more blurred. In contrast, HD95 and ASD mainly measure the accuracy of the boundary and the fidelity of the shape, and are more sensitive to boundary errors and local inconsistencies. In subsequent studies, attention will be paid to how to further improve the performance of the model on different datasets by enhancing boundary feature learning and optimizing data enhancement strategies.

Table 1. Assessment of our approach against leading semi-supervised segmentation techniques on the ACDC dataset

Method	Scans used		Metrics			
	Labeled	Unlabeled	Dice	Jaccard	HD95	ASD
UA-MT	7(10%)	63(90%)	79.67	68.96	7.13	2.52
SASSNet	7(10%)	63(90%)	84.03	73.04	5.51	1.93
URPC	7(10%)	63(90%)	83.25	73.41	4.92	1.59
MC-Net	7(10%)	63(90%)	85.94	76.68	5.51	1.63
SS-Net	7(10%)	63(90%)	86.71	77.36	6.05	1.57
BCP	7(10%)	63(90%)	88.45	79.91	4.21	1.29
Ours	7(10%)	63(90%)	89.98	81.86	2.61	1.03

Table 2. Benchmarking with top-tier semi-supervised segmentation algorithms on the BraTS 2019 dataset

Method	Type	Metrics			
		Dice	Jaccard	HD95	ASD
MT	3D	78.07	67.42	10.56	2.73
UAMT	3D	79.09	68.05	11.97	3.5
URPC	3D	81.59	71.23	9.77	2.69
MCNet	3D	79.22	68.65	10.91	3.40
SDMI-Net	3D	81.87	71.47	9.64	2.41
Ours	2D	86.39	76.93	9.96	2.46

4.5. Ablation Studies

4.5.1 Ablation study of Main Network

In the experimental setup using the FLAIR modality of the BraTS 2019 dataset, this study conducted an ablation analysis by changing only the backbone network to the U-Net architecture and AttU-Net network structure while keeping other parts of the model constant. According to Table 3, there is a significant performance degradation when the backbone network is switched from Frequency Attention U-Net to standard U-Net. This result indicates that although the standard U-Net performs well under many circumstances, the Frequency Attention U-Net is more effective in highlighting crucial feature areas when dealing with medical images of the FLAIR modality, which exhibit complex heterogeneity. Thus, incorporating frequency attention mechanisms enhances overall segmentation accuracy.

5.5.2 Ablation study of Loss Function

In experiments conducted on the ACDC dataset, this study maintained all aspects of the model constant except for the loss function, which was changed to a weighted cross-entropy loss function. As shown in Table 4, replacing the loss function with weighted cross-entropy resulted in a decrease in performance. This suggests that while weighted cross-entropy is typically used to address class imbalance by assigning different weights to classes, it may not always lead to an improvement in performance for all types of tasks. This indicates the superiority of the selected loss function in this study.

Table 3. Ablation study of Main Network on BraTS 2019 dataset

Main Network	Scans used		Metrics			
	Labeled	Unlabeled	Dice	Jaccard	95HD	ASD
U-Net	7(10%)	63(90%)	84.63	74.78	16.19	4.53
AttU-Net			85.64	76.01	11.13	2.92
Ours			86.39	76.93	9.96	2.46

Table 4. Ablation study of Loss Function on ACDC dataset

Loss Function	Scans used		Metrics			
	Labeled	Unlabeled	Dice	Jaccard	95HD	ASD
Weighted cross entropy	7(10%)	63(90%)	88.55	80.08	6.27	1.64
Ours			89.98	81.86	2.61	1.03

4.6. Ablation Studies

4.6.1 Ablation study of Main Network

In the experimental setup using the FLAIR modality of the BraTS 2019 dataset, this study conducted an ablation analysis by changing only the backbone network to the U-Net architecture and AttU-Net network structure while keeping other parts of the model constant. According to Table 3, there is a significant performance degradation when the backbone network is switched from Frequency Attention U-Net to standard U-Net. This result indicates that although the standard U-Net performs well under many circumstances, the Frequency Attention U-Net is more effective in highlighting crucial feature areas when dealing with medical images of the FLAIR modality, which exhibit complex heterogeneity. Thus, incorporating frequency attention mechanisms enhances overall segmentation accuracy.

4.6.2 Ablation study of Loss Function

In experiments conducted on the ACDC dataset, this study maintained all aspects of the model constant except for the loss function, which was changed to a weighted cross-entropy loss function. As shown in Table 4, replacing the loss function with weighted cross-entropy resulted in a decrease in performance. This suggests that while weighted cross-entropy is typically used to address class imbalance by assigning different weights to classes, it may not always lead to an improvement in performance for all types of tasks. This indicates the superiority of the selected loss function in this study.

5. Conclusion:

This study proposes an enhanced semi-supervised medical image segmentation model FAS-Net. The FAS-Net model designs a frequency domain attention mechanism based on DCT and introduces a data mixing strategy for input data to improve the performance of semi supervised medical image segmentation on the basis of the Mean-Teacher semi supervised segmentation framework. Comparative analysis demonstrates superior performance of our algorithm across four evaluation criteria on the ACDC dataset. However, a weakness is revealed in the BraTS dataset where it lags in HD95 and ASD measurements.

While FAS-Net excels in overall segmentation accuracy, the results also highlight areas for further improvement, especially in capturing local segmentation details, as observed in the BraTS2019 dataset. Despite these limitations, the proposed FAS-Net model contributes to the advancement of an efficient and cost-effective medical image analysis pipeline by reducing the reliance on extensive pixel-level annotations. Future research directions could explore further

improving the model's ability to capture local segmentation details, such as integrating multi-scale or hierarchical attention mechanisms, or utilizing additional spatial or contextual information.

Acknowledgement

The authors are grateful to the Management and Science University in Malaysia, Guangzhou College of Commerce in China, and the University of Mosul in Iraq for their supports. This work is funded by Guangdong Provincial Education Science Foundation [grant number: 2023GXJK407], the Scientific Research Foundation of Guangzhou [grant number: 202201011667], the Philosophy and Social Science Foundation of Guangzhou [grant number: 2021GZGJ145], the Guangdong Province Undergraduate Teaching Quality and Teaching Reform Engineering Construction Project (2022SJXGG992), the school level scientific research project of Guangzhou Business School (2023XJYB33)

References

- Ahmed, N., Natarajan, T., & Rao, K. R. J. I. t. o. C. (1974). Discrete cosine transform. 100(1), 90-93.
- Bai, Y., Chen, D., Li, Q., Shen, W., & Wang, Y. (2023). Bidirectional copy-paste for semi-supervised medical image segmentation. Paper presented at the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.
- Basak, H., & Yin, Z. (2023). Pseudo-label guided contrastive learning for semi-supervised medical image segmentation. Paper presented at the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.
- Cao, F., Guo, D., Wang, T., Yao, H., Li, J., & Qin, C. J. E. S. w. A. (2024). Universal screen-shooting robust image watermarking with channel-attention in DCT domain. 238, 122062.
- Chen, B., Liu, Y., Zhang, Z., Lu, G., & Kong, A. W. K. J. I. T. o. E. T. i. C. I. (2023). Transattunet: Multi-level attention-guided u-net with transformer for medical image segmentation.
- Cheng, J., Tian, S., Yu, L., Gao, C., Kang, X., Ma, X., . . . Lu, H. J. M. I. A. (2022). ResGANet: Residual group attention network for medical image classification and segmentation. 76, 102313.
- Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., & Ronneberger, O. (2016). 3D U-Net: learning dense volumetric segmentation from sparse annotation. Paper presented at the Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19.
- Devunooru, S., Alsadoon, A., Chandana, P., Beg, A. J. J. o. A. I., & Computing, H. (2021). Deep learning neural networks for medical image segmentation of brain tumours for diagnosis: a recent review and taxonomy. 12, 455-483.
- Fan, J., Gao, B., Jin, H., & Jiang, L. (2022). Ucc: Uncertainty guided cross-head co-training for semi-supervised semantic segmentation. Paper presented at the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.
- Fu, J., Liu, J., Tian, H., Li, Y., Bao, Y., Fang, Z., & Lu, H. (2019). Dual attention network for scene segmentation. Paper presented at the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.
- Fu, Y., Lei, Y., Wang, T., Curran, W. J., Liu, T., & Yang, X. J. P. M. (2021). A review of deep learning based methods for medical image multi-organ segmentation. 85, 107-122.

- Gai, D., Huang, Z., Min, W., Geng, Y., Wu, H., Zhu, M., . . . Medicine. (2024). SDMI-Net: Spatially Dependent Mutual Information Network for semi-supervised medical image segmentation. 108374.
- Gul, S., Khan, M. S., Bibi, A., Khandakar, A., Ayari, M. A., Chowdhury, M. E. J. C. i. B., & Medicine. (2022). Deep learning techniques for liver and liver tumor segmentation: A review. 147, 105620.
- Han, Z., Jian, M., & Wang, G.-G. J. K.-B. S. (2022). ConvUNeXt: An efficient convolution neural network for medical image segmentation. 253, 109512.
- Huang, H., Lin, L., Tong, R., Hu, H., Zhang, Q., Iwamoto, Y., . . . Wu, J. (2020). Unet 3+: A full-scale connected unet for medical image segmentation. Paper presented at the ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP).
- Huang, M.-L., Wu, Y.-Z. J. B. S. P., & Control. (2022). Semantic segmentation of pancreatic medical images by using convolutional neural network. 73, 103458.
- Jiang, M., Roth, H. R., Li, W., Yang, D., Zhao, C., Nath, V., . . . Xu, Z. (2023). Fair Federated Medical Image Segmentation via Client Contribution Estimation. Paper presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.
- LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. J. N. c. (1989). Backpropagation applied to handwritten zip code recognition. 1(4), 541-551.
- Li, C., Tan, Y., Chen, W., Luo, X., He, Y., Gao, Y., . . . Graphics. (2020). ANU-Net: Attention-based nested U-Net to exploit full resolution features for medical image segmentation. 90, 11-20.
- Li, S., Zhang, C., & He, X. (2020). Shape-aware semi-supervised 3D semantic segmentation for medical images. Paper presented at the Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part I 23.
- Liu, X., Song, L., Liu, S., & Zhang, Y. J. S. (2021). A review of deep-learning-based medical image segmentation methods. 13(3), 1224.
- Liu, Y., Wang, H., Chen, Z., Huangliang, K., & Zhang, H. J. K.-B. S. (2022). TransUNet+: Redesigning the skip connection to enhance features in medical image segmentation. 256, 109859.
- Luo, X., Liao, W., Chen, J., Song, T., Chen, Y., Zhang, S., . . . Zhang, S. (2021). Efficient semi-supervised gross target volume of nasopharyngeal carcinoma segmentation via uncertainty rectified pyramid consistency. Paper presented at the Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part II 24.
- Luo, X., Wang, G., Liao, W., Chen, J., Song, T., Chen, Y., . . . Zhang, S. J. M. I. A. (2022). Semi-supervised medical image segmentation via uncertainty rectified pyramid consistency. 80, 102517.
- Mendel, R., Rauber, D., de Souza Jr, L. A., Papa, J. P., Palm, C. J. C. i. B., & Medicine. (2023). Error-Correcting Mean-Teacher: Corrections instead of consistency-targets applied to semi-supervised medical image segmentation. 154, 106585.
- Patil, D. D., Deore, S. G. J. I. J. o. C. S., & Computing, M. (2013). Medical image segmentation: a review. 2(1), 22-27.
- Qureshi, I., Yan, J., Abbas, Q., Shaheed, K., Riaz, A. B., Wahid, A., . . . Szczuko, P. J. I. F. (2023). Medical image segmentation using deep semantic-based methods: A review of techniques, applications and emerging trends. 90, 316-352.

- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. Paper presented at the Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18.
- Shen, X., Yang, J., Wei, C., Deng, B., Huang, J., Hua, X.-S., . . . Liang, K. (2021). Dct-mask: Discrete cosine transform mask representation for instance segmentation. Paper presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.
- Sun, J., Peng, Y., Guo, Y., & Li, D. J. N. (2021). Segmentation of the multimodal brain tumor image used the multi-pathway architecture method based on 3D FCN. 423, 34-45.
- Tarvainen, A., & Valpola, H. J. A. i. n. i. p. s. (2017). Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. 30.
- Tarvainen, A., & Valpola, H. J. a. p. a. (2017). Weight-averaged consistency targets improve semi-supervised deep learning results. CoRR. abs/1703.01780 (2017).
- Tu, P., Huang, Y., Zheng, F., He, Z., Cao, L., & Shao, L. (2022). Guidedmix-net: Semi-supervised semantic segmentation by using labeled images as reference. Paper presented at the Proceedings of the AAAI Conference on Artificial Intelligence.
- Wang, F., Wang, H., Wei, C., Yuille, A., & Shen, W. (2022). Cp 2: Copy-paste contrastive pretraining for semantic segmentation. Paper presented at the European Conference on Computer Vision.
- Wang, J., Liu, X. J. C. M., & Biomedicine, P. i. (2021). Medical image recognition and segmentation of pathological slices of gastric cancer based on Deeplab v3+ neural network. 207, 106210.
- Wang, K., Zhan, B., Zu, C., Wu, X., Zhou, J., Zhou, L., & Wang, Y. (2021). Tripled-uncertainty guided mean teacher model for semi-supervised medical image segmentation. Paper presented at the Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part II 24.
- Wang, K., Zhan, B., Zu, C., Wu, X., Zhou, J., Zhou, L., & Wang, Y. J. M. I. A. (2022). Semi-supervised medical image segmentation via a tripled-uncertainty guided mean teacher model with contrastive learning. 79, 102447.
- Wang, P., & Zhao, X. (2022). DCT-CenterMask: A Real-Time Instance Segmentation Network for Kidney Ultrasound Images. Paper presented at the Proceedings of the 5th International Conference on Big Data Technologies.
- Wang, Q., Li, X., Chen, M., Chen, L., Chen, J. J. P. i. M., & Biology. (2022). A regularization-driven Mean Teacher model based on semi-supervised learning for medical image segmentation. 67(17), 175010.
- Wen, Z., & Ye, C. (2024). A Robust Mean Teacher Framework for Semi-Supervised Cell Detection in Histopathology Images. Paper presented at the Medical Imaging with Deep Learning.
- Wu, Y., Wu, Z., Wu, Q., Ge, Z., & Cai, J. (2022). Exploring smoothness and class-separation for semi-supervised medical image segmentation. Paper presented at the International conference on medical image computing and computer-assisted intervention.
- Wu, Y., Xu, M., Ge, Z., Cai, J., & Zhang, L. (2021). Semi-supervised left atrium segmentation with mutual consistency training. Paper presented at the Medical image computing and computer assisted intervention–MICCAI 2021: 24th international conference, Strasbourg, France, September 27–October 1, 2021, proceedings, part II 24.

Xiang, T., Zhang, C., Liu, D., Song, Y., Huang, H., & Cai, W. (2020). BiO-Net: learning recurrent bi-directional connections for encoder-decoder architecture. Paper presented at the Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part I 23.

Xiaoyan, W., Yuan, Y., Guo, D., Huang, X., Cui, Y., Xia, M., . . . Chen, S. J. M. I. A. (2022). SSA-Net: Spatial self-attention network for COVID-19 pneumonia infection segmentation with semi-supervised few-shot learning. 79, 102459.

Xiong, Z., Xia, Q., Hu, Z., Huang, N., Bian, C., Zheng, Y., . . . Yang, X. J. M. i. a. (2021). A global benchmark of algorithms for segmenting the left atrium from late gadolinium-enhanced cardiac magnetic resonance imaging. 67, 101832.

Xu, K., Qin, M., Sun, F., Wang, Y., Chen, Y.-K., & Ren, F. (2020). Learning in the frequency domain. Paper presented at the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.

Yeung, M., Sala, E., Schönlieb, C.-B., Rundo, L. J. C. i. b., & medicine. (2021). Focus U-Net: A novel dual attention-gated CNN for polyp segmentation during colonoscopy. 137, 104815.

Yu, L., Wang, S., Li, X., Fu, C.-W., & Heng, P.-A. (2019). Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation. Paper presented at the Medical image computing and computer assisted intervention–MICCAI 2019: 22nd international conference, Shenzhen, China, October 13–17, 2019, proceedings, part II 22.