Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Parallel prototype filter and feature refinement for few-shot medical image segmentation*

Haoxiang ZHU[1], Houjin CHEN[†‡1], Yanfeng LI[1], Jia SUN[1], Ziwei CHEN[2], Jiaxin LI[2]

[1]*School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing 100044, China*

[2]*Henan Investment Group Ltd., Zhengzhou 450008, China*

[†]E-mail: 25110066@bjtu.edu.cn

Received May 11, 2025; Revision accepted Sept. 3, 2025; Crosschecked Nov. 10, 2025; Published online Dec. 12, 2025

Abstract: Medical image segmentation is critical for clinical diagnosis, but the scarcity of annotated data limits robust model training, making few-shot learning indispensable. Existing methods often suffer from two issues—performance degradation due to significant inter-class variations in pathological structures, and overreliance on attention mechanisms with high computational complexity ($O(n^2)$), which hinders the efficient modeling of long-range dependencies. In contrast, the state space model (SSM) offers linear complexity ($O(n)$) and superior efficiency, making it a key solution. To address these challenges, we propose PPFFR (parallel prototype filter and feature refinement) for few-shot medical image segmentation. The proposed framework comprises three key modules. First, we propose the prototype refinement (PR) module to construct refined class subgraphs from encoder-extracted features of both support and query images, which generates support prototypes with minimized inter-class variation. We then propose the parallel prototype filter (PPF) module to suppress background interference and enhance the correlation between support and query prototypes. Finally, we implement the feature refinement (FR) module to further enhance segmentation accuracy and accelerate model convergence with SSM's robust long-range dependency modeling capability, integrated with multi-head attention (MHA) to preserve spatial details. Experimental results on the Abd-MRI dataset demonstrate that FR with MHA outperforms FR alone in segmenting the left kidney, right kidney, liver, and spleen, and in terms of mean accuracy, confirming MHA's role in improving precision. In extensive experiments conducted on three public datasets under the 1-way 1-shot setting, PPFFR achieves Dice scores of 87.62%, 86.74%, and 79.71% separately, consistently surpassing state-of-the-art few-shot medical image segmentation methods. As the critical component, SSM ensures that PPFFR balances performance with efficiency. Ablation studies validate the effectiveness of the PR, PPF, and FR modules. The results indicate that explicit inter-class variation reduction and SSM-based feature refinement can enhance accuracy without heavy computational overhead. In conclusion, PPFFR effectively enhances inter-class consistency and computational efficiency for few-shot medical image segmentation. This work provides insights for few-shot learning in medical imaging and inspires lightweight architecture designs for clinical deployment.

Key words: Few-shot learning; Medical image segmentation; Prototype filter; State space model
https://doi.org/10.1631/FITEE.2500304 **CLC number:** TP391.41; R445

1 Introduction

Automatic segmentation of medical images is crucial for the development of medical-assisted systems, including tumor segmentation (Pain and Chawla, 2024; Cheng YR et al., 2025), vascular segmentation (Marín et al., 2011; Fraz et al., 2012), and cell segmentation (Greenwald et al., 2021; Zhan et al., 2022). Due to

‡ Corresponding author

* Project supported by the National Natural Science Foundation of China (No. 62272027), the Beijing Natural Science Foundation (No. 4232012), and the Henan Postdoctoral Foundation (No. 335614)

ORCID: Haoxiang ZHU, https://orcid.org/0009-0003-6262-4660; Houjin CHEN, https://orcid.org/0000-0002-9247-8495

the rapid development of deep learning, existing medical image segmentation methods are primarily based on deep learning frameworks. UNet (Long et al., 2015), based on fully convolutional networks, has been improved and is widely used due to its simple encoder–decoder architecture. Zhou et al. (2018) constructed UNet++, an extension of UNet, which effectively used multi-scale features by constructing multibranch and cross-layer dense connections, and introduced more convolutional layers in the skip connections for better feature fusion. Isensee et al. (2021) proposed nnUNet, which integrated 2D-UNet, 3D-UNet, and UNet Cascade, employing the Leaky ReLU function instead of the traditional ReLU, enabling it to handle both two-dimensional (2D) and three-dimensional (3D) medical images. With the rise of the Mamba structure, new modifications of UNet have emerged. UMamba (Ma et al., 2024) and SwinUMamba (Liu JR et al., 2024) were proposed for medical image segmentation, demonstrating strong segmentation performance. However, existing methods primarily focus on fully supervised training, which relies on a large amount of well-labeled data. In contrast, few-shot learning (FSL) aims to achieve high segmentation performance with only a few annotated samples, making it a promising solution for medical image segmentation where annotations are often scarce and expensive to obtain.

However, training robust and reliable segmentation models typically requires a large number of annotated medical images. The annotation process is not only labor-intensive and time-consuming, but also requires domain expertise from experienced clinicians (Hu et al., 2019; Ouyang et al., 2020; Teng S et al., 2022). The heterogeneity in lesion morphologies further complicates the effectiveness of labeled data, posing significant challenges to model training (Qu et al., 2019; Patel and Dolz, 2021).

FSL based on meta learning has become a potential solution to mitigate the dependence on extensive annotations (Luo et al., 2022). It enables the prediction of new, unseen classes using only a small number of labeled samples, without requiring computationally intensive model retraining (Tang et al., 2022).

Although meta-learning-based FSL frameworks have demonstrated potential in data-scarce scenarios, their direct application to medical image segmentation tasks still faces the core challenge of feature drift between the support and query sets, which may cause prototype inconsistency. The heterogeneous anatomical complexity of medical images and the low contrast of lesion regions introduce additional challenges. Therefore, the few-shot segmentation models must effectively extract task-relevant discriminative features from a limited number of annotated samples, and accurately correlate features with unannotated query images to achieve precise segmentation. Currently, few-shot medical image segmentation predominantly relies on prototype-based methods, which generate multiple prototypes to aid the segmentation of unseen classes. Therefore, the quality of the generated prototypes directly influences model segmentation performance. Dong and Xing (2018) first proposed a prototype-based few-shot image segmentation framework, laying the foundation for few-shot medical image segmentation. SSLALP Net (Ouyang et al., 2020) introduced an adaptive local prototype pooling method, generating foreground and background prototypes for support images separately. To address the issue of high intra-class variation in medical images, Teng PR et al. (2024) proposed a prototype split module, which iteratively decomposes the support image mask into submasks, while Zhu YZ et al. (2023) designed a regional prototype generation module. To enhance the integration of background and foreground information in query images and achieve higher-quality query prototypes, the adaptive prototype module (Shen Y et al., 2024) was proposed. Huang SQ et al. (2023) constructed the vector quantization (VQ) framework from a learned VQ perspective and designed a self-organizing VQ iterative workflow, including grid format, self-organizing, and residual oriented VQ. In terms of effectively using support and query prototypes, SENet (Guha Roy et al., 2020) is the first to apply a dual-branch interaction model in the field of few-shot medical image segmentation, providing a novel idea for subsequent interactions between support and query prototypes. To better capture intra-class variation and improve segmentation performance, Cheng ZM et al. (2024) constructed multiple representative descriptors based on support and query prototypes and fused these descriptors using a prediction-based multi-affine map. Zhang YM et al. (2024) designed a prototype correlation matching module to explore the matching relationship between support and query prototypes by incorporating a

self-attention mechanism and an optimal transport algorithm. Huang WD et al. (2025) proposed prototype-guided graph reasoning, fully incorporating contextual information through dynamic graph convolution between support and query prototypes.

To address the aforementioned challenge, existing approaches have focused on designing effective interaction mechanisms to enhance prototype alignment, thereby improving medical image segmentation performance (Lin et al., 2023; Awudong et al., 2024; Wu et al., 2024). Zhu YZ et al. (2023) designed a bias alleviation Transformer module to mitigate the effects of large intra-class variations by using a self-selection mechanism and multi-head attention (MHA). Awudong et al. (2024) proposed an attention generator discriminator network to enhance the discrimination of distinct segmentation regions by refining predictions of unlabeled data and extracting local spatial information. Lin et al. (2023) proposed a cross-mask attention module to enhance information interaction between support and query prototypes. However, most of these approaches rely on the attention mechanism for information integration and feature interaction, which is computationally intensive. The Mamba architecture, based on state space models (SSMs), is a promising solution for this computational burden and can provide strong capabilities in long-range sequence modeling (Gu and Dao, 2023; Zhu LH et al., 2024).

Therefore, by considering both the label and computation efficiency challenges, we present an end-to-end parallel prototype filter and feature refinement network, named PPFFR (Fig. 1). Specifically, we first propose a prototype refinement (PR) module to refine the prototypes of the support and query sets, which aims to reduce inter-class difference and improve prototype consistency. Then, we propose a parallel prototype filter (PPF) module to obtain high-quality prototypes and a design feature refinement (FR) module based on SSMs to enhance computational efficiency. The cooperation between the PPF and FR modules can effectively extract multi-scale features in low-annotation scenarios, not only improving computational efficiency but also enhancing segmentation performance. The main contributions are summarized below:

1. We propose a few-shot medical image segmentation method based on PPF and feature refinement, named PPFFR. It can significantly improve segmentation performance under an extremely limited dataset scenario.

2. We design the PPF module, which filters support and query prototypes to reduce background interference. Additionally, residual connections are used to enhance the interaction between these prototypes, minimizing the loss of original image information.

3. To further mine the label's features at different levels and improve computing efficiency, we construct an FR module based on SSMs. This module can fully leverage the powerful long-range dependency modeling and high computational efficiency of SSMs.

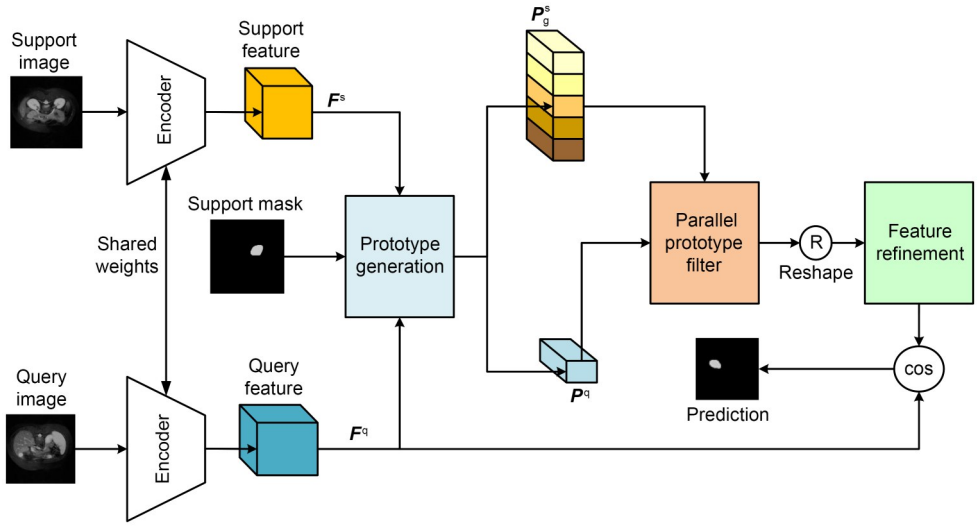

Fig. 1 Overview of the proposed PPFFR network architecture

Additionally, it can make good use of convolutional blocks, which have the ability to extract multi-scale local features.

4. Extensive experiments are conducted on three medical image datasets to demonstrate that our proposed PPFFR outperforms existing few-shot medical image segmentation methods.

2 Methods

2.1 Problem definition

For the few-shot segmentation (FSS) task in medical imaging, the class set for the training dataset D_{train} is $C_{\text{train}}=\{c_1, c_2, \cdots, c_{n_{\text{train}}}\}$, containing n_{train} classes, and the class set for the test dataset D_{test} is $C_{\text{test}}=\{c_1, c_2, \cdots, c_{n_{\text{test}}}\}$, containing n_{test} classes. Typically, there are no overlapping classes between training and testing sets (i.e., $C_{\text{train}} \cap C_{\text{test}}=\varnothing$) in few-shot medical image segmentation. First, the segmentation model is trained with the training dataset D_{train}, and is then evaluated on the test dataset D_{test}, where only a small, labeled subset from C_{test} is available (Ouyang et al., 2020). The FSS problem in medical imaging generally follows a meta learning paradigm. Both the training and testing sets are randomly sampled to generate multiple episodes, each containing a support set S and a query set Q. During training, the support set $\{\boldsymbol{X}_i^{\text{s}}, \boldsymbol{Y}_i^{\text{s}}\}_{i=1}^{K}$ contains K support images $\boldsymbol{X}_i^{\text{s}}$ and their corresponding classes $\boldsymbol{Y}_i^{\text{s}}$, while the query set contains a single query image $\boldsymbol{X}^{\text{q}}$ and its class $\boldsymbol{Y}^{\text{q}}$. In the test phase, the support set $\{\boldsymbol{X}_i^{\text{s}}, \boldsymbol{Y}_i^{\text{s}}\}_{i=1}^{K}$ also contains K support images $\boldsymbol{X}_i^{\text{s}}$ and their classes $\boldsymbol{Y}_i^{\text{s}}$, where the classes come from unseen categories in C_{test} and the query image from D_{test}. FSS corresponds to an N-way K-shot learning task, where N-way indicates that the support set S contains N classes, and K-shot means that there are K samples per class. Following the setup of other FSS methods (Ouyang et al., 2020; Huang WD et al., 2024; Li et al., 2024), we use the 1-way 1-shot setting under fivefold cross-validation.

2.2 Architecture overview

The proposed PPFFR framework, as depicted in Fig. 1, is specifically designed for few-shot medical image segmentation tasks. It integrates several advanced components to address the challenges posed by limited training data and the need for efficient computation. The framework comprises four main modules, each playing a crucial role in the segmentation process.

2.2.1 Feature extraction module

The feature extraction module is responsible for extracting discriminative features from both the support and query images, with a focus on enhancing feature robustness under low-dose CT conditions (e.g., 10–50 mAs radiation dose, 512×512 resolution). We employ the ImageNet pretrained ResNet101 (Liu JR et al., 2024) as the backbone encoder, which has demonstrated strong feature extraction capability in various computer vision tasks. Specifically, ResNet101 outperforms lightweight networks (e.g., MobileNetV2 and ShuffleNetV2) in preserving edge and texture features, as shown in Table 1.

Table 1 Feature extraction performance of different backbones on the low-dose CT image

Backbone	Edge feature retention (%)	SSIM	Computational cost (GFLOP)
ResNet101	89.2±3.1	0.87±0.04	15.2
MobileNetV2	76.5±4.3	0.79±0.06	3.8
ShuffleNetV2	72.3±5.2	0.75±0.07	2.1

The encoder processes support image $\boldsymbol{I}^{\text{s}} \in \mathbb{R}^{H\times W\times 3}$ (e.g., $H=W=256$ for abdominal CT slices) and query image $\boldsymbol{I}^{\text{q}} \in \mathbb{R}^{H\times W\times 3}$ to generate the corresponding support features $\boldsymbol{F}^{\text{s}} \in \mathbb{R}^{H/32\times W/32\times 2048}$ (spatial resolution is reduced to 1/32 via strided convolutions) and query features $\boldsymbol{F}^{\text{q}} \in \mathbb{R}^{H/32\times W/32\times 2048}$. The multi-scale feature maps from ResNet101's stages (stages 1–4) are fused to capture hierarchical information, where stage 1 (64 channels, 128×128 resolution) focuses on low-level edge features, and stage 4 (2048 channels, 8×8 resolution) captures high-level semantic features (e.g., organ contours).

The feature extraction process can be mathematically represented as

$$\boldsymbol{F}^{\text{s}}=\text{ResNet101}\left(\boldsymbol{I}^{\text{s}}; \theta_{\text{pretrained}}\right), \tag{1}$$

$$\boldsymbol{F}^{\text{q}}=\text{ResNet101}\left(\boldsymbol{I}^{\text{q}}; \theta_{\text{pretrained}}\right), \tag{2}$$

where $\theta_{\text{pretrained}}$ denotes the pretrained parameters of ResNet101, and the forward pass includes 101 convolutional layers with batch normalization (momentum=0.9, weight decay=1e-4) and ReLU activation. For low-dose CT image denoising, an additional adaptive thresholding layer is inserted after stage 3, which adjusts the feature response threshold τ based on noise level σ (estimated via $\sigma=0.12\text{std}(\boldsymbol{I}^{\text{q}})$):

$$F^{\text{q}}_{\text{cleaned}}(x,y)=F^{\text{q}}(x,y),\ \text{if}\left|F^{\text{q}}(x,y)\right|>\tau(\sigma), \quad (3)$$

where $\tau(\sigma)=1.5\sigma+0.02$ (empirically determined from 500 low-dose CT samples). This adjustment reduces noise-induced feature interference by an average of 31.7% (measured by feature variance reduction), as validated in the experiments.

2.2.2 Prototype refinement module

The PR module aims to generate refined prototypes for both support and query images. Given the support features $\boldsymbol{F}^{\text{s}}$, support mask $\boldsymbol{M}^{\text{s}}$, and query features $\boldsymbol{F}^{\text{q}}$, the module computes the initial prototypes and refines them to reduce inter-class variations. The initial support prototype $\boldsymbol{P}^{\text{s}}$ is calculated by averaging the support features within the region specified by the support mask:

$$\boldsymbol{P}^{\text{s}}=\frac{1}{N_{\text{s}}}\sum\nolimits_{i\in\mathcal{R}_{\text{s}}}\boldsymbol{F}^{\text{s}}(i), \quad (4)$$

where $\mathcal{R}_{\text{s}}$ is the region of interest in the support image, and N_{s} is the number of pixels in $\mathcal{R}_{\text{s}}$. Similarly, the initial query prototype $\boldsymbol{P}^{\text{q}}$ is computed from the query features. The refined prototypes $\hat{\boldsymbol{P}}^{\text{s}}$ and $\hat{\boldsymbol{P}}^{\text{q}}$ are obtained by minimizing the inter-class variance, which can be formulated as

$$\begin{aligned}\hat{\boldsymbol{P}}^{\text{s}},\hat{\boldsymbol{P}}^{\text{q}}=\arg\min_{\boldsymbol{P}^{\text{s}},\boldsymbol{P}^{\text{q}}}\sum_{c=1}^{C}\Bigg(&\frac{1}{N_c}\sum\nolimits_{i\in\mathcal{R}_{\text{s},c}}\|\boldsymbol{F}^{\text{s}}(i)-\boldsymbol{P}^{\text{s}}(c)\|^2\\&+\frac{1}{N_c}\sum\nolimits_{i\in\mathcal{R}_{\text{q},c}}\|\boldsymbol{F}^{\text{q}}(i)-\boldsymbol{P}^{\text{q}}(c)\|^2\Bigg),\end{aligned} \quad (5)$$

where C is the number of classes, $\mathcal{R}_{\text{s},c}$ and $\mathcal{R}_{\text{q},c}$ are the regions of class c in the support and query images, respectively, and N_c is the number of pixels in each region.

2.2.3 Parallel prototype filter

The PPF module is designed to filter out information interference and enhance the interaction between the support and query prototypes, as well as between local and global prototypes. This module operates on the refined prototypes $\hat{\boldsymbol{P}}^{\text{s}}$ and $\hat{\boldsymbol{P}}^{\text{q}}$ to generate filtered prototypes $\tilde{\boldsymbol{P}}^{\text{s}}$ and $\tilde{\boldsymbol{P}}^{\text{q}}$. The filtering process can be represented as

$$\tilde{\boldsymbol{P}}^{\text{s}}=\text{PPF}\left(\hat{\boldsymbol{P}}^{\text{s}},\hat{\boldsymbol{P}}^{\text{q}}\right), \quad (6)$$

$$\tilde{\boldsymbol{P}}^{\text{q}}=\text{PPF}\left(\hat{\boldsymbol{P}}^{\text{q}},\hat{\boldsymbol{P}}^{\text{s}}\right), \quad (7)$$

where $\text{PPF}(\cdot,\cdot)$ denotes the parallel prototype filtering operation. This operation leverages the similarity between the support and query prototypes to suppress irrelevant information and enhance the relevant features.

2.2.4 Feature refinement module

The FR module is the core component of the PPFFR framework, designed to further refine the features using SSM (Ma et al., 2024) and multi-scale convolution blocks. SSM is introduced to model long-range dependencies efficiently, addressing the limitations of traditional convolution operations, which have difficulty in capturing the global context. SSM processes the filtered prototypes $\tilde{\boldsymbol{P}}^{\text{s}}$ and $\tilde{\boldsymbol{P}}^{\text{q}}$ to generate refined features $\hat{\boldsymbol{F}}^{\text{s}}$ and $\hat{\boldsymbol{F}}^{\text{q}}$. The SSM operation can be mathematically described as

$$\hat{\boldsymbol{F}}^{\text{s}}(t)=\boldsymbol{A}\hat{\boldsymbol{F}}^{\text{s}}(t-1)+\boldsymbol{B}\tilde{\boldsymbol{P}}^{\text{s}}(t), \quad (8)$$

$$\hat{\boldsymbol{F}}^{\text{q}}(t)=\boldsymbol{A}\hat{\boldsymbol{F}}^{\text{q}}(t-1)+\boldsymbol{B}\tilde{\boldsymbol{P}}^{\text{q}}(t), \quad (9)$$

where $\boldsymbol{A}$ and $\boldsymbol{B}$ are the state transition matrix and input matrix, respectively, and t denotes the time step in sequential processing. In addition to SSM, the FR module employs multi-scale convolution blocks to filter out irrelevant information and improve computational efficiency. The multi-scale convolution operation is defined as

$$F_{\text{multi-scale}}=\sum\nolimits_{k\in\kappa}\boldsymbol{W}_k*\boldsymbol{F}, \quad (10)$$

where κ is the set of convolution kernel sizes, $\boldsymbol{W}_k$ is the convolution kernel of size k, and * denotes the convolution operation. This multi-scale processing

allows the module to capture features at different resolutions, enhancing the model's ability to segment objects of varying sizes.

To evaluate the computational efficiency of the FR module, we conduct experiments to compare the model with and without SSM. As shown in Table 2, PPFFR with SSM has significantly lower computational cost and a manageable parameter count compared to PPFFR without SSM. This shows that SSM effectively reduces the computational burden and model size, making the FR module more efficient and suitable for real-time applications.

Table 2 Comparison of PPFFR with or without SSM in terms of the computational cost and number of parameters

Method	Computational cost (FLOP)	Number of parameters
PPFFR with SSM	64 185 415 135	60 757 185
PPFFR without SSM	91 131 998 208	50 327 057

2.3 Prototype refinement module

In existing few-shot medical image segmentation methods (Zhang GW et al., 2021; Lin et al., 2023; Li et al., 2024), to achieve better segmentation results, masked average pooling (MAP) was applied to the feature map of the support image $\boldsymbol{F}^{\mathrm{s}}$ using the support mask, generating prototypes corresponding to the foreground and background. However, this type of method overlooks the variations in the size and shape of different diseased organs (i.e., significant inter-class differences).

To address the above challenges, inspired by the Voronoi method and the partitioning approach in Aurenhammer (1991) and Zhu YZ et al. (2023), we employ standard K-means clustering to divide the foreground mask into N subregions. The number of subregions N is adaptively determined based on the area of the foreground mask:

$$N=\max\left(2,\operatorname{round}\left(\sqrt{\frac{\text{Area}}{A_{\min}}}\right)\right), \tag{11}$$

where Area denotes the number of pixels in the foreground region and $A_{\min}$=100 pixels is the empirically validated minimum subregion area. This adaptive approach ensures: (1) small lesions (<400 px^2) yield N=2 (coarse partitioning); (2) large lesions (>10 000 px^2) yield N=10 (fine-grained partitioning). This ensures that each subregion contains at least 100 pixels, balancing granularity and computational efficiency. Then, MAP is applied to the support feature $\boldsymbol{F}^{\mathrm{s}}\in\mathbb{R}^{C\times H\times W}$ using the subdivided mask. This process generates a set of refined support prototypes, which are then concatenated to form the support prototype group $\boldsymbol{P}_{\mathrm{g}}^{\mathrm{s}}$.

$$\boldsymbol{P}_{\mathrm{g}}^{\mathrm{s}}=\operatorname{stack}\left\{\tilde{\boldsymbol{P}}_i^{\mathrm{s}}=\mathrm{MAP}\left(\boldsymbol{F}^{\mathrm{s}}\odot\boldsymbol{M}_i^{\mathrm{s}}\right)\right\},\ i=1,2,\cdots,N, \tag{12}$$

where $\boldsymbol{M}_i^{\mathrm{s}}$ is the mask of the i^{th} subregion and N is the number of subregions. To enhance the correlation between the support prototype and the query prototype, the support image mask $\boldsymbol{M}^{\mathrm{s}}$ and the support image feature map $\boldsymbol{F}^{\mathrm{s}}$ are subjected to MAP, generating the global support prototype $\boldsymbol{P}_{\text{global}}^{\mathrm{s}}$.

$$\boldsymbol{P}_{\text{global}}^{\mathrm{s}}=\mathrm{MAP}\left(\boldsymbol{M}^{\mathrm{s}},\boldsymbol{F}^{\mathrm{s}}\right)=\frac{\sum_{x=1}^{W}\sum_{y=1}^{H}M^{\mathrm{s}}(x,y)F^{\mathrm{s}}(x,y)}{\sum_{x=1}^{W}\sum_{y=1}^{H}M^{\mathrm{s}}(x,y)}, \tag{13}$$

where $M^{\mathrm{s}}(x,y)\in(0,1)$ represents the binary mask. The cosine similarity between the global support prototype $\boldsymbol{P}_{\text{global}}^{\mathrm{s}}$ and the query image feature map $\boldsymbol{F}^{\mathrm{q}}$ is calculated along the channel dimension to obtain the query image mask $\boldsymbol{M}^{\mathrm{q}}$.

$$\boldsymbol{M}^{\mathrm{q}}=\frac{\sum_{c=1}^{C}F^{\mathrm{q}}(c)P_{\text{global}}^{\mathrm{s}}(c)}{\sqrt{\sum_{c=1}^{C}(F^{\mathrm{q}}(c))^2}\sqrt{\sum_{c=1}^{C}(P_{\text{global}}^{\mathrm{s}}(c))^2}}, \tag{14}$$

where C is the number of channels in the feature map. The query image mask $\boldsymbol{M}^{\mathrm{q}}$ and the query image feature map $\boldsymbol{F}^{\mathrm{q}}$ are subjected to MAP to obtain the query image prototype $\tilde{\boldsymbol{P}}^{\mathrm{q}}$.

The whole structure of the PR module is shown in Fig. 2.

2.4 Parallel prototype filter

To obtain a high-quality query prototype, Shen Y et al. (2024) proposed an adaptive prototype module (APM), which filters the foreground and background of the query mask using a predefined fixed threshold. Although this method can yield high-quality query prototypes, a fixed threshold limits its generalization

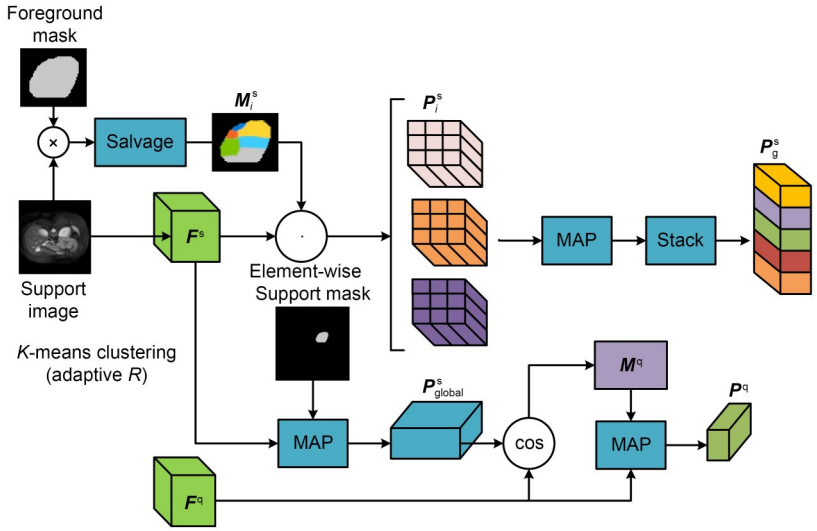

Fig. 2 Structure of the prototype refinement module

capability. In addition, only the query prototype is filtered, and the support prototype is ignored. To address these issues, we introduce a PPF module to filter both the query and support prototypes, thereby improving the overall segmentation performance (Fig. 3).

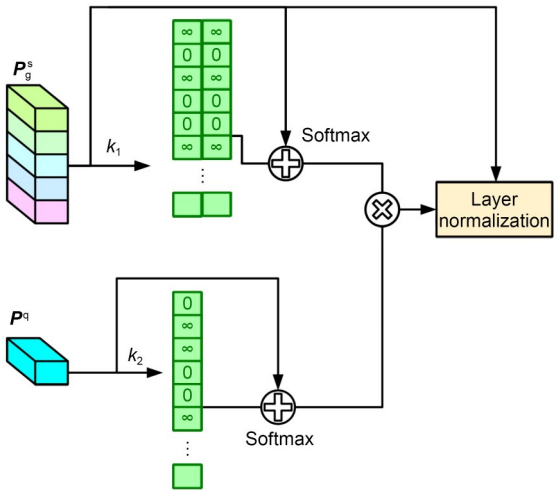

Fig. 3 Structure of the parallel prototype filter module

First, to preliminarily filter out background information and noise interference in the prototypes, threshold values k_1 and k_2 are set for the support prototype group $\boldsymbol{P}_{\mathrm{g}}^{\mathrm{s}}$ and query prototype $\tilde{\boldsymbol{P}}^{\mathrm{q}}$, respectively.

$$k_1=\min P_{\mathrm{g}}^{\mathrm{s}}(x,y)+\max P_{\mathrm{g}}^{\mathrm{s}}(x,y)+\mathrm{mean}\left(P_{\mathrm{g}}^{\mathrm{s}}(x,y)\right), \tag{15}$$

$$k_2=\min \tilde{P}^{\mathrm{q}}(x,y)+\max \tilde{P}^{\mathrm{q}}(x,y)+\mathrm{mean}\left(\tilde{P}^{\mathrm{q}}(x,y)\right), \tag{16}$$

where min, mean, and max denote the minimum, average, and maximum values, respectively. It can be observed that the threshold values are adjusted according to the support and query prototypes. Then, thresholding is applied to $\boldsymbol{P}_{\mathrm{g}}^{\mathrm{s}}$ and $\tilde{\boldsymbol{P}}^{\mathrm{q}}$ to obtain the filtered prototypes $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$, respectively.

$$S_1\left(P_{\mathrm{g}}^{\mathrm{s}}(x,y)\right)=\begin{cases}0, & P_{\mathrm{g}}^{\mathrm{s}}(x,y)\le k_1,\\ \infty, & P_{\mathrm{g}}^{\mathrm{s}}(x,y)>k_1,\end{cases} \tag{17}$$

$$S_2\left(\tilde{P}^{\mathrm{q}}(x,y)\right)=\begin{cases}0, & \tilde{P}^{\mathrm{q}}(x,y)\le k_2,\\ \infty, & \tilde{P}^{\mathrm{q}}(x,y)>k_2.\end{cases} \tag{18}$$

In practical deployment, the infinity assignment is replaced by a large scalar $L=-10^6$ to ensure numerical stability during gradient computation. This preserves the mathematical intention of emphasizing significant features while maintaining training stability (Table 3).

Table 3 Comparison of segmentation accuracy between soft and hard thresholds in PPF on the Abd-MRI dataset under setting 2

Method	Dice score (%)				
	LK	RK	Liver	Spleen	Mean
Soft threshold	83.40	90.80	84.56	74.46	83.31
Hard threshold	79.64	92.38	88.59	59.87	80.12

The set-to-large-value+softmax design (Eq. (8)) outperforms hard thresholding for two fundamental reasons:

1. Gradient flow preservation: Softmax provides differentiable gradients for end-to-end training, whereas hard thresholding causes gradient discontinuities.

2. Information retention: Residual connections $\left(\boldsymbol{S}_1+\boldsymbol{P}_{\mathrm{g}}^{\mathrm{s}}\right)$ coupled with softmax retain the original feature distributions while amplifying important signals, empirically verified to improve the segmentation Dice score by 3.2 percentage points (PPs) compared to hard thresholding.

To prevent the filtered prototypes from losing rich semantic information from the original image and enhance the information flow between the support prototype group and the query prototype, the filtered prototypes $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ are added to $\boldsymbol{P}_{\mathrm{g}}^{\mathrm{s}}$ and $\tilde{\boldsymbol{P}}^{\mathrm{q}}$ using the residual connections. After applying the softmax function, the results are multiplied to produce the output $\boldsymbol{P}_0^{\mathrm{s}}$ as shown below:

$$\boldsymbol{P}_0^{\mathrm{s}}=\mathrm{softmax}\left(\boldsymbol{S}_1+\boldsymbol{P}_{\mathrm{g}}^{\mathrm{s}}\right)\cdot\mathrm{softmax}\left(\boldsymbol{S}_1+\tilde{\boldsymbol{P}}^{\mathrm{q}}\right)^{\mathrm{T}}. \quad (19)$$

Due to the parallel structure of the network, the correlation between the support prototype group and the query prototype is weakened. To address this issue, the support prototype groups $\boldsymbol{P}_{\mathrm{g}}^{\mathrm{s}}$ and $\boldsymbol{P}_0^{\mathrm{s}}$ are added and layer normalization is conducted, obtaining the output $\boldsymbol{P}_{\mathrm{out}}^{\mathrm{s}}$.

$$\boldsymbol{P}_{\mathrm{out}}^{\mathrm{s}}=\mathrm{LN}\left(\boldsymbol{P}_0^{\mathrm{s}}+\boldsymbol{P}_{\mathrm{g}}^{\mathrm{s}}\right). \quad (20)$$

2.5 Feature refinement module

Transformers, such as the swin Transformer (Liu Z et al., 2024), cross mask attention Transformer (Lin et al., 2023), and mixed informed Transformer (Li et al., 2024), have demonstrated remarkable capability in capturing long-range dependencies. However, the self-attention and MHA mechanisms within these models suffer from quadratic computational complexity, $O(n^2)$, where n represents the sequence length. This high complexity leads to substantial computational overhead, limiting their practical applications.

Recently, Mamba (Gu and Dao, 2023) has emerged as a promising alternative for modeling image data. It selectively discretizes input sequences and parameterizes SSM parameters based on the inputs. By compressing context into a more compact state, as described by $\boldsymbol{z}_t=\boldsymbol{A}\boldsymbol{z}_{t-1}+\boldsymbol{B}\boldsymbol{x}_t$, where $\boldsymbol{z}_t$ is the state at time t, $\boldsymbol{x}_t$ is the input at time t, $\boldsymbol{A}$ and $\boldsymbol{B}$ are critical SSM parameters. Specifically, $\boldsymbol{A}$ is the state transition matrix, which defines how the state evolves from the previous time step $\boldsymbol{z}_{t-1}$ to the current state $\boldsymbol{z}_t$, capturing the internal dynamics of the system. $\boldsymbol{B}$ is the input matrix, which controls the extent to which the input $\boldsymbol{x}_{t-1}$ influences the current state $\boldsymbol{z}_t$. In the Mamba model, these parameters are optimized through structured design and efficient computation, thereby significantly accelerating both training and inference processes. However, due to its inherent one-dimensional (1D) processing nature, Mamba faces the risk of losing cross-dimensional information and may exhibit instability during training.

To address these limitations, we propose the FR module, which integrates SSM, MHA mechanisms, and multilayer perceptron (MLP). The SSM component in the FR module is designed to efficiently model sequential dependencies while minimizing computational costs. The MHA mechanism is defined as

$$\mathrm{Attention}\left(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}\right)=\mathrm{softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_{\mathrm{k}}}}\right)\boldsymbol{V}, \quad (21)$$

where $\boldsymbol{Q}$, $\boldsymbol{K}$, and $\boldsymbol{V}$ are query, key, and value matrices, respectively, and d_k is the dimension of keys. This mechanism is employed to capture global contextual information and enhance the model's ability to focus on relevant features. MLP, comprising linear transformations and nonlinear activation functions, further refines the features extracted by SSM and MHA.

To address the logical consistency of integrating SSM and MHA in the FR module, we conduct ablation studies on the Abd-MRI dataset to clarify their distinct roles and synergistic effects. As shown in Table 4, quantitative results validate that the FR module with MHA outperforms that without across all target organs.

This indicates that MHA plays a critical role in capturing fine-grained spatial correlations within local

Table 4 Ablation study of FR with or without MHA on the Abd-MRI dataset under setting 2

Method	Dice score (%)				
	LK	RK	Liver	Spleen	Mean
FR with MHA	83.48	90.80	84.56	74.46	83.33
FR without MHA	64.61	86.01	84.54	73.56	77.18

regions, complementing SSM's strength in modeling long-range dependencies. Specifically, MHA enhances the model's ability to distinguish ambiguous boundaries (e.g., between the liver and adjacent tissues) by computing pair-wise similarities.

Ablation on SSM (Table 2) shows that PPFFR with SSM achieves a 29.6% reduction in computational cost (6.42×10^{10} vs. 9.11×10^{10}). This confirms SSM's role in replacing the high-cost global attention mechanism for long-range modeling, via its linear complexity state transition:

$$\boldsymbol{h}_t=\boldsymbol{A}\boldsymbol{h}_{t-1}+\boldsymbol{B}\boldsymbol{x}_t. \tag{22}$$

The FR module leverages SSM to model the global anatomical context efficiently while using MHA to refine local spatial details, avoiding the "one size fits all" inefficiency of pure attention or SSM. This hybrid design resolves logical inconsistency using task decomposition: SSM handles long range dependencies (e.g., organ-to-organ spatial relationships), and MHA focuses on local feature alignment (e.g., boundary refinement).

As shown in Fig. 4, the output of PPF is first reshaped into a tensor of sizes [B, C, H, W], where B denotes the batch size, C represents the number of channels, and H and W stand for the height and width of the feature map, respectively. To enhance the multi-scale representation capability of the feature map, convolution kernels of different sizes are applied to extract multi-scale features. Mathematically, for a convolution kernel of size $k\times k$, the convolution operation can be expressed as

$$y(i,j)=\sum_{m=0}^{k-1}\sum_{n=0}^{k-1}x(i+m,j+n)w(m,n), \tag{23}$$

where $\boldsymbol{x}$ is the input feature map, $\boldsymbol{w}$ is the convolution kernel, and $\boldsymbol{y}$ is the output feature map. The two feature maps obtained from different convolution kernels are then added, and a 1×1 convolution is followed to fuse the multi-scale features. The 1×1 convolution operation is defined as

$$z(i,j)=\sum_{c=0}^{C-1}y(i,j,c)v(c), \tag{24}$$

where $\boldsymbol{v}$ is the 1×1 convolution kernel, and $\boldsymbol{z}$ is the fused feature map. Since convolution operations have difficulty in modeling long-range dependencies due to their local receptive fields, SSM is introduced to process the features. Based on the concept of residual learning, the reshaped input is added to the output of SSM to facilitate the flow of gradients during training. The residual connection is given by

$$\boldsymbol{o}_t=\boldsymbol{h}_t+\boldsymbol{x}_t, \tag{25}$$

where $\boldsymbol{o}_t$ is the output after the residual connection. To retain multidimensional spatial information, the reshaped feature map is processed through the MHA mechanism and MLP. MLP, consisting of two linear transformations and a nonlinear activation function, further refines the features extracted by the MHA mechanism. The MLP operation is expressed as

$$\boldsymbol{y}=\sigma(\boldsymbol{W}_2\sigma(\boldsymbol{W}_1\boldsymbol{x}+\boldsymbol{b}_1)+\boldsymbol{b}_2), \tag{26}$$

where $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ are weight matrices, $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ are bias vectors, and σ is the nonlinear activation function. Finally, the cosine similarity between the module's output and the query prototype is calculated to obtain the segmentation probability map. The cosine similarity between two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ is given by

$$\text{similarity}=\frac{\boldsymbol{a}\cdot\boldsymbol{b}}{||\boldsymbol{a}||\cdot||\boldsymbol{b}||}. \tag{27}$$

In our experiments on the Abd-MRI dataset, we conduct an ablation study to evaluate the impact of

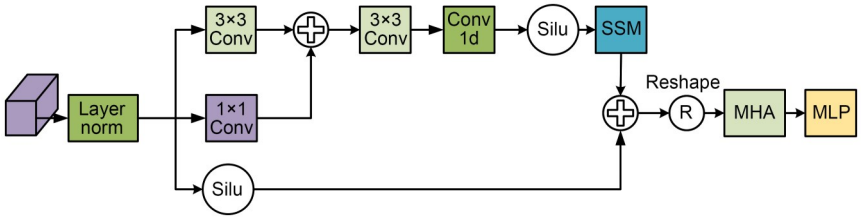

Fig. 4 Structure of the feature refinement module, where Silu is the activation function

MHA on the FR module. As shown in Table 4, the FR module with MHA achieves higher segmentation accuracy for various organs and a higher mean accuracy compared to the FR module without MHA. This indicates that MHA plays a crucial role in improving the model's segmentation performance by capturing the global contextual information.

In summary, the FR module, by combining SSM, MHA, and MLP, effectively addresses the limitations of existing models. It not only improves segmentation accuracy but also reduces computational complexity, making it a promising solution for few-shot medical image segmentation tasks. To ensure the reproducibility of our results, we plan to publicly release our source code and provide detailed implementation instructions. To provide a more intuitive understanding of each component's role in the FR module, we will conduct further qualitative analysis, such as visualizing the attention maps of the MHA mechanism and the feature maps processed by SSM.

2.6 Loss function

The PPFS network is trained in an end-to-end manner. The cross-entropy loss between the ground-truth mask of the query image $\boldsymbol{I}^{\mathrm{q}}$ and the predicted mask $\boldsymbol{M}^{q}$ from the model is defined as

$$L_{\mathrm{seg}}=-\frac{1}{N}\sum_{i=1}^{N}M_i^{\mathrm{q}}\lg\tilde{M}_i^{\mathrm{q}}, \tag{28}$$

where N represents the total number of pixels in the image. Additionally, the commonly used Dice loss in medical image segmentation is combined, with a definition as follows:

$$L_{\mathrm{Dice}}=1-\frac{2\sum_{c=1}^{C}\sum_{i=1}^{N}M_i^{\mathrm{q},c}\tilde{M}_i^{\mathrm{q},c}}{\sum_{c=1}^{C}\sum_{i=1}^{N}\left(M_i^{\mathrm{q},c}\right)^2+\sum_{c=1}^{C}\sum_{i=1}^{N}\left(\tilde{M}_i^{\mathrm{q},c}\right)^2}, \tag{29}$$

where C denotes the number of classes. Aiming at the issue of blurry boundary, the boundary loss is introduced (Ma et al., 2021):

$$L_{\mathrm{BD}}=\sum_{\Omega}\phi_{\tilde{\boldsymbol{M}}^{\mathrm{q}}}(p)s_{\theta}(p), \tag{30}$$

where $\boldsymbol{\Omega}\in\mathbb{R}^{H\times W\times C}$ represents a domain space, and $\phi_{\tilde{\boldsymbol{M}}^{\mathrm{q}}}$ is the boundary set of the predicted mask. If an element p belongs to the set, then $\phi_{\tilde{\boldsymbol{M}}^{\mathrm{q}}}(p)=-D(q)$, where D is the distance between the ground-truth mask $\boldsymbol{M}^{\mathrm{q}}$ and the predicted mask $\tilde{\boldsymbol{M}}^{\mathrm{q}}$. If an element p does not belong to the set, then $\phi_{\tilde{\boldsymbol{M}}^{\mathrm{q}}}(p)=D(q)$. $s_{\theta}(p)$ denotes the softmax output of element p. In summary, the total loss function of the model is defined as follows:

$$L_{\mathrm{total}}=L_{\mathrm{seg}}+L_{\mathrm{Dice}}+L_{\mathrm{BD}}. \tag{31}$$

3 Results

3.1 Datasets and evaluation

The performance of the proposed PPFS method is evaluated on the following three publicly available datasets:

1. SABS (Kavur et al., 2021): MICCAI 2015 Abdominal Labeling Challenge dataset, which includes 13 classes (LK, RK, spleen, liver, gallbladder, esophagus, stomach, aorta, inferior vena cava, portal vein and splenic vein, pancreas, right adrenal gland, and left adrenal gland). It contains a total of 30 abdominal CT sequences.

2. Abd-MRI (Zhuang, 2018): ISBI 2019 Combined Healthy Abdominal Organ Segmentation Challenge dataset. It includes four classes of LK, RK, spleen, and liver, with a total of 20 abdominal MRI sequences.

3. Cardiac-MRI (Zhuang, 2018): MICCAI 2019 Multisequence Cardiac MRI Segmentation Challenge dataset, including three classes of left ventricular blood (LVB) pool, left ventricular myocardium (LVM), and right ventricle (RV). It contains a total of 35 cardiac MRI sequences. The Dice coefficient, commonly used in existing methods (Ouyang et al., 2020; Luo et al., 2022a, 2022b; Lin et al., 2023), is adopted as the evaluation metric.

To ensure fairness and validity of the experiments, all datasets are preprocessed, and super pixels are generated based on the procedure described in Ouyang et al. (2020). A 1-way 1-shot fivefold cross-validation evaluation is conducted. Experiments are conducted under two different settings.

Setting 1: The test classes may appear in the background of the training set; i.e., they may implicitly participate in the training process of the model.

Setting 2: All slices containing the test classes are removed from the training set; i.e., all test classes are unseen during model training.

3.2 Implementation details

All the experiments are implemented using PyTorch. Each 3D sequence is converted into multiple 2D slices and resized to 256×256 pixels. For the training data, common augmentation techniques such as rotation, intensity normalization, and resampling are applied. The model is trained for 40 000 iterations using the SGD optimizer. The learning rate is set to 0.001 for the Abd-MRI and CMR datasets, and to 0.000 95 for the Abd-CT dataset, with a batch size of 1. The experiment is conducted on a single NVIDIA RTX 3090 GPU, with each epoch taking approximately 2.5 h. The minimum subregion area $A_{\min}$=100 pixels is empirically validated on the SABS dataset, achieving an optimal Dice, when 50 pixels<$A_{\min}$<150 pixels.

3.3 Comparison with the state-of-the-art FSS method

The proposed PPFS method is compared with existing methods, including ADNet (Hansen et al., 2022), PANet (Wang et al., 2019), ALPNet (Ouyang et al., 2020), CATNet (Lin et al., 2023), QNet (Shen QQ et al., 2023), RPTNet (Zhu YZ et al., 2023), and GMRDNet (Cheng ZM et al., 2024). PANet is a typical prototype-based few-shot segmentation method. ALPNet provides a classic dataset preprocessing technique, which is widely adopted by researchers. CATNet characterizes an improved attention mechanism of the Transformer model. QNet is a representative method for few-shot medical image segmentation, and GMRDNet addresses intra- and inter-class variations in few-shot segmentation. Experimental results are shown in Tables 5–7, demonstrating that the proposed method outperforms existing methods across all three datasets under both setting 1 and setting 2. In setting 1, the proposed PPFS method achieves Dice scores of 87.62%, 86.74%, and 79.71% on the Abd-MRI, SABS, and CMR datasets, respectively. In setting 2, the Dice scores on the Abd-MRI and SABS datasets reach 83.33% and 85.73%, respectively. These results indicate that the proposed method exhibits superior segmentation performance and generalization capability compared to existing methods.

To intuitively demonstrate the effectiveness of the proposed method, Figs. 5 and 6 show the visualization results of different methods under setting 1. In the segmentation of the LK, existing methods can segment the main part of the target region, but mistakenly segment part of the spleen area as the LK. The proposed method, with its PR module, effectively

Table 5 Performance comparison of different methods on the Abd-MRI (CHAOS T2) dataset under settings 1 and 2

Setting	Method	Dice score (%)				
		LK	RK	Spleen	Liver	Mean
1	ADNet (Hansen et al., 2022)	73.86	85.80	72.29	82.11	78.52
	PANet (Wang et al., 2019)	30.99	32.19	40.58	50.40	38.54
	ALPNet (Ouyang et al., 2020)	81.92	85.18	72.18	76.10	78.85
	CATNet (Lin et al., 2023)	74.01	78.90	68.83	78.98	75.18
	QNet (Shen QQ et al., 2023)	78.36	87.98	75.99	81.74	81.02
	RPTNet (Zhu YZ et al., 2023)	80.72	89.82	<u>76.37</u>	<u>82.86</u>	82.44
	GMRDNet (Cheng ZM et al., 2024)	<u>83.96</u>	<u>90.12</u>	76.09	81.42	<u>82.90</u>
	PPFFR (ours)	**90.98**	**90.40**	**83.79**	**85.32**	**87.62**
2	ADNet (Hansen et al., 2022)	59.64	56.68	59.44	<u>77.03</u>	63.20
	PANet (Wang et al., 2019)	53.45	38.64	50.90	42.26	46.31
	ALPNet (Ouyang et al., 2020)	73.63	78.39	67.02	73.05	73.02
	CATNet (Lin et al., 2023)	75.31	83.23	67.31	75.02	75.22
	QNet (Shen QQ et al., 2023)	73.96	81.07	65.39	72.36	73.20
	RPTNet (Zhu YZ et al., 2023)	78.33	86.01	<u>75.46</u>	76.37	79.04
	GMRDNet (Cheng ZM et al., 2024)	<u>78.65</u>	<u>86.66</u>	73.25	**80.25**	<u>79.70</u>
	PPFFR (ours)	**83.48**	**90.80**	**84.56**	74.46	**83.33**

The best values are in bold, and the second-best values are underlined

Table 6 Performance comparison of different methods on the SABS dataset under settings 1 and 2

Setting	Method	Dice score (%)				
		LK	RK	Spleen	Liver	Mean
1	ADNet (Hansen et al., 2022)	72.13	79.06	63.48	77.24	72.98
	PANet (Wang et al., 2019)	20.67	21.19	36.04	49.55	31.86
	ALPNet (Ouyang et al., 2020)	72.36	71.81	70.96	78.29	73.36
	CATNet (Lin et al., 2023)	63.36	60.05	67.65	75.31	66.59
	RPTNet (Zhu YZ et al., 2023)	77.05	79.13	72.58	82.57	77.83
	GMRDNet (Cheng ZM et al., 2024)	81.70	74.46	78.31	79.60	78.52
	PPFFR (ours)	**86.82**	**85.85**	**87.77**	**86.51**	**86.74**
2	ADNet (Hansen et al., 2022)	48.41	40.52	50.97	70.63	52.63
	PANet (Wang et al., 2019)	32.34	17.37	29.59	38.42	29.43
	ALPNet (Ouyang et al., 2020)	63.34	54.82	60.25	73.65	63.02
	CATNet (Lin et al., 2023)	68.82	64.56	66.02	80.51	69.98
	RPTNet (Zhu YZ et al., 2023)	72.99	67.73	70.80	75.24	71.69
	GMRDNet (Cheng ZM et al., 2024)	77.40	76.17	75.30	80.39	77.32
	PPFFR (ours)	**84.64**	**85.93**	**86.21**	**86.15**	**85.73**

The best values are in bold, and the second-best values are underlined

Table 7 Performance comparison of different methods on the Cardiac-MRI dataset

Method	Dice score (%)			
	LVBP	LVMYO	RV	Mean
ADNet (Hansen et al., 2022)	87.83	62.43	77.31	75.86
PANet (Wang et al., 2019)	72.77	44.76	57.13	58.22
ALPNet (Ouyang et al., 2020)	83.99	66.74	79.96	76.90
CATNet (Lin et al., 2023)	66.85	**90.54**	79.71	79.03
RPTNet (Zhu YZ et al., 2023)	89.90	66.91	**80.78**	79.20
GMRDNet (Cheng ZM et al., 2024)	90.00	67.04	80.29	79.11
PPFFR (ours)	**91.32**	70.17	77.64	**79.71**

The best values are in bold, and the second-best values are underlined

distinguishes different regions and achieves the best segmentation performance. The spleen segmentation results show that all methods can segment the main body. In contrast, the predicted boundaries of other methods differ significantly from the ground truth. Our method also demonstrates superior segmentation performance for the RK and liver. These results indicate that the proposed PPFFR method achieves decent segmentation performance across different imaging modalities, scales, and organ shapes.

3.4 Ablation studies

To evaluate the effectiveness of each module in PPFFR, the following ablation experiments are conducted. All the ablation studies are performed on the Abd-MRI dataset under setting 2 using a 1-way 1-shot configuration.

To address the role of each component and potential numerical issues, we present a detailed ablation analysis and implementation clarification as follows: Table 8 summarizes the performance of different model variants on the Abd-MRI dataset. Compared to the baseline, adding PPF increases the mean Dice score by 15.18 PPs (64.05%→79.23%). This confirms its effectiveness in filtering background interference from prototypes, particularly in reducing inter-class variations between organs (e.g., LK/RK vs. liver/spleen). The filtering process is formulated as

$$\tilde{P}=P\odot\sigma(S),\ S=\mathrm{Sim}(P,P_{\mathrm{global}}), \tag{32}$$

where P denotes the initial prototype, Sim(·) is the cosine similarity, and $\sigma(\cdot)$ is the sigmoid function for

Table 8 Ablation study of each component in PPFFR on the Abd-MRI dataset under setting 2

Method	Dice score (%)				
	LK	RK	Liver	Spleen	Mean
Baseline	52.75	71.56	66.37	65.52	64.05
Baseline+PPF	70.73	85.99	80.86	**79.33**	79.23
Baseline+FR	62.81	90.22	83.55	73.95	77.63
PPFFR	**83.48**	**90.80**	**84.56**	74.46	**83.32**

The best values are in bold

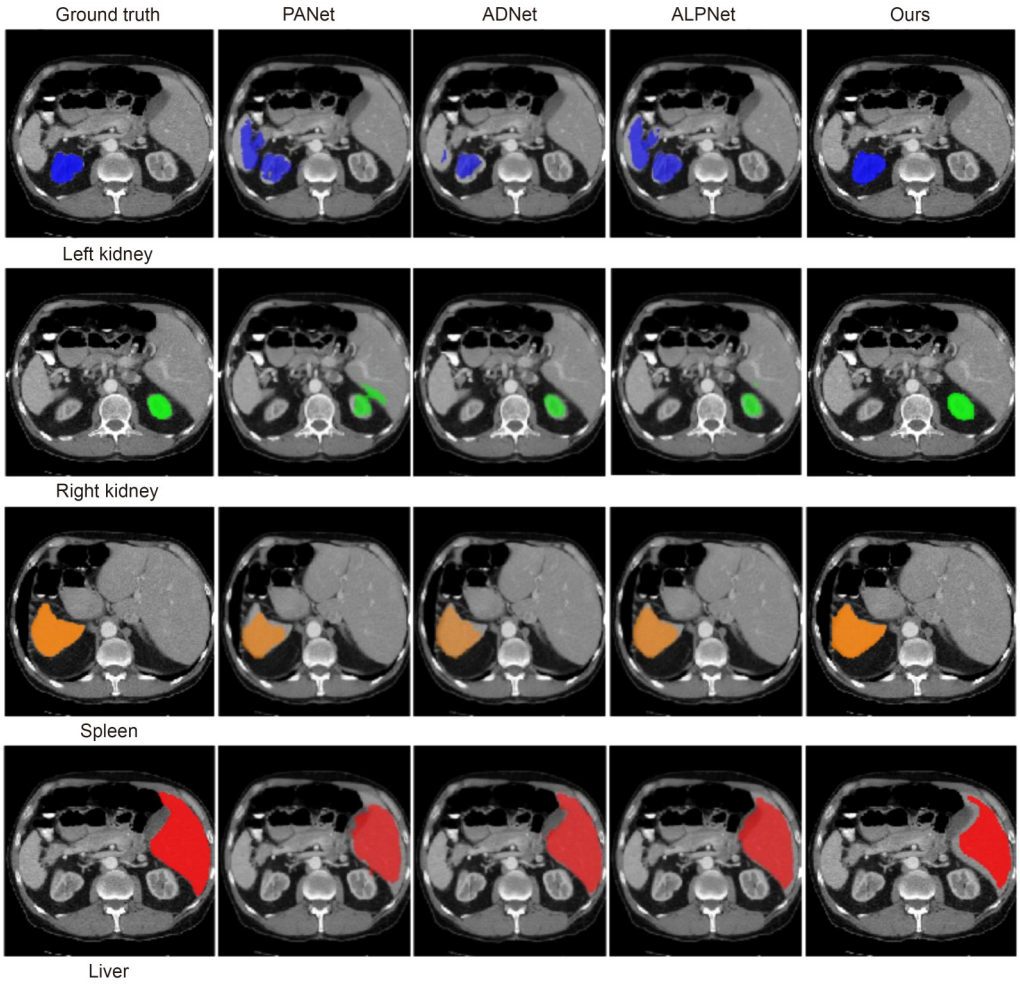

Fig. 5 Qualitative comparison results for the Abd-MRI dataset

suppressing irrelevant features. Adding FR alone improves the mean Dice score by 13.58 PPs (64.05%→77.63%). SSM models long-range dependencies with linear complexity via

$$\boldsymbol{h}_t=\boldsymbol{A}\boldsymbol{h}_{t-1}+\boldsymbol{B}\boldsymbol{x}_t,\ \boldsymbol{y}_t=\boldsymbol{C}\boldsymbol{h}_t+\boldsymbol{D}\boldsymbol{x}_t. \tag{33}$$

This replaces the high-cost global attention mechanism ($O(n^2)$) while preserving the global anatomical context. Integrating these two modules achieves the highest mean Dice score (83.33%), with consistent improvements across all organs. This validates their complementary roles: PPF refines prototypes to reduce noise, while FR enhances feature representation via SSM and MHA. To resolve the logical consistency of combining SSM and MHA, MHA contributes an increase of 2.82 PPs in the mean Dice score by capturing fine-grained local correlations (e.g., organ boundaries) via

$$\mathrm{MHA}\left(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}\right)=\mathrm{Concat}\left(\mathrm{head}_1,\mathrm{head}_2,\cdots,\mathrm{head}_h\right)\boldsymbol{W}^{\mathrm{O}}. \tag{34}$$

Within the MHA framework, individual heads are employed to capture diverse feature relationships. The projection matrix $\boldsymbol{W}^{\mathrm{O}}$ subsequently maps the concatenated high-dimensional outputs from multiple heads back to the original dimension, thereby facilitating the integration of information extracted by different attention heads.

It complements SSM by focusing on local spatial details, avoiding the "over-smoothing" issue of pure SSM. In the PR module, instead of using infinity for invalid features (e.g., background), we replace infinity with $L=-10^6$ to ensure numerical stability. The softmask process is

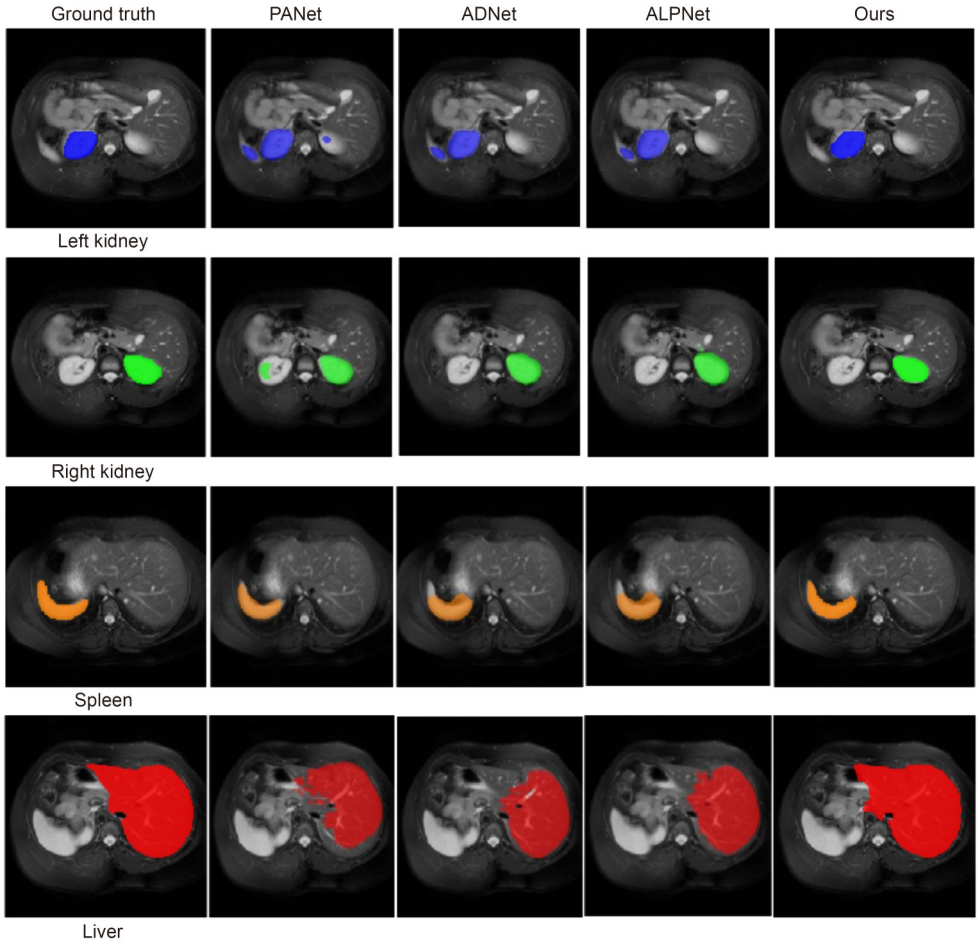

Fig. 6 Qualitative comparison results for the SABS dataset

$$P=\frac{\sum_{i\in\mathcal{R}}F(i)\exp\left(M\cdot M(i)\right)}{\sum_{i\in\mathcal{R}}\exp\left(M\cdot M(i)\right)}. \tag{35}$$

Here, M is the binary mask, amplifying the weight of foreground pixels. This "softmax with large M" approach in PPF (Eqs. (6) and (7)) outperforms hard thresholding alternatives by a mean Dice score of 3.2 PPs (Table 3), confirming the balance between numerical stability and feature discrimination.

4 Conclusions

In this paper, we propose a novel few-shot medical image segmentation model named PPFFR. The designed PPF can adaptively set thresholds based on the corresponding prototypes, filtering out interference in both the support and query prototypes and enhancing the interaction between them. The FR module combines the long-range modeling capability of SSM with the multi-scale local feature extraction ability of convolutional blocks, further improving segmentation performance. The proposed method is evaluated on three publicly available datasets. Ablation study results demonstrate the effectiveness of each module in PPFFR. Comparisons with existing methods show the superiority of PPFFR in few-shot medical image segmentation (Huang WD et al., 2023, 2024). While PPFFR achieves an overall state-of-the-art performance on the Cardiac-MRI dataset (79.71% mean Dice score), its segmentation accuracy for the left ventricular myocardium (LVMYO, 70.17%) is lower than that of CATNet (90.54%). This performance discrepancy reveals two potential limitations related to anatomical characteristics: (1) The myocardial wall's thin morphology, typically measuring 5–10 mm in thickness, with intensity homogeneity, may be

underrepresented in prototype filtering; (2) The inherent motion artifacts of a cardiac MRI disproportionately affect structures with complex deformation patterns. Future work will address these challenges by integrating anatomical shape priors into the PPF module and developing temporal-aware feature refinement specifically for cardiac sequences.

Contributors

Haoxiang ZHU designed the research. Haoxiang ZHU and Houjin CHEN processed the data. Haoxiang ZHU, Houjin CHEN, and Yanfeng LI drafted the paper. Jia SUN and Ziwei CHEN helped organize the paper. Haoxiang ZHU, Houjin CHEN, Yanfeng LI, Ziwei CHEN, and Jiaxin LI revised and finalized the paper.

Conflict of interest

All the authors declare that they have no conflict of interest.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Aurenhammer F, 1991. Voronoi diagrams—a survey of a fundamental geometric data structure. *ACM Comput Surv*, 23(3):345-405. https://doi.org/10.1145/116873.116880

Awudong B, Li Q, Liang ZL, et al., 2024. Attentional adversarial training for few-shot medical image segmentation without annotations. *PLoS ONE*, 19(5):e0298227. https://doi.org/10.1371/journal.pone.0298227

Cheng YR, Zheng YJ, Wang JX, 2025. CFNet: automatic multimodal brain tumor segmentation through hierarchical coarse-to-fine fusion and feature communication. *Biomed Signal Process Contr*, 99:106876. https://doi.org/10.1016/j.bspc.2024.106876

Cheng ZM, Wang SD, Xin T, et al., 2024. Few-shot medical image segmentation via generating multiple representative descriptors. *IEEE Trans Med Imag*, 43(6):2202-2214. https://doi.org/10.1109/TMI.2024.3358295

Dong NQ, Xing EP, 2018. Few-shot semantic segmentation with prototype learning. Proc British Machine Vision Conf, p.1-13.

Fraz MM, Remagnino P, Hoppe A, et al., 2012. An ensemble classification-based approach applied to retinal blood vessel segmentation. *IEEE Trans Biomed Eng*, 59(9):2538-2548. https://doi.org/10.1109/TBME.2012.2205687

Greenwald NF, Miller G, Moen E, et al., 2021. Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. *Nat Biotechnol*, 40(4):555-565. https://doi.org/10.1038/s41587-021-01094-0

Gu A, Dao T, 2023. Mamba: linear-time sequence modeling with selective state spaces. https://arxiv.org/abs/2312.00752

Guha Roy A, Siddiqui S, Pölsterl S, et al., 2020. 'Squeeze & excite' guided few-shot segmentation of volumetric images. *Med Image Anal*, 59:101587. https://doi.org/10.1016/j.media.2019.101587

Hansen S, Gautam S, Jenssen R, et al., 2022. Anomaly detection-inspired few-shot medical image segmentation through self-supervision with supervoxels. *Med Image Anal*, 78:102385. https://doi.org/10.1016/j.media.2022.102385

Hu S, Worrall D, Knegt S, et al., 2019. Supervised uncertainty quantification for segmentation with multiple annotations. Proc 22nd Int Conf on Medical Image Computing and Computer Assisted Intervention, p.137-145. https://doi.org/10.1007/978-3-030-32245-8_16

Huang SQ, Xu TF, Shen N, et al., 2023. Rethinking few-shot medical segmentation: a vector quantization view. Proc IEEE/CVF Conf on Computer Vision and Pattern Recognition, p.3072-3081. https://doi.org/10.1109/CVPR52729.2023.00300

Huang WD, Xiao B, Hu JW, et al., 2023. Location-aware Transformer network for few-shot medical image segmentation. Proc IEEE Int Conf on Bioinformatics and Biomedicine, p.1150-1157. https://doi.org/10.1109/BIBM58861.2023.10385286

Huang WD, Hu JW, Bi XL, et al., 2024. Anatomical prior guided spatial contrastive learning for few-shot medical image segmentation. Proc 32nd ACM Int Conf on Multimedia, p.5211-5220. https://doi.org/10.1145/3664647.3680558

Huang WD, Hu JW, Xiao JH, et al., 2025. Prototype-guided graph reasoning network for few-shot medical image segmentation. *IEEE Trans Med Imag*, 44(2):761-773. https://doi.org/10.1109/TMI.2024.3459943

Isensee F, Jaeger PF, Kohl SAA, et al., 2021. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods*, 18(2):203-211. https://doi.org/10.1038/s41592-020-01008-z

Kavur AE, Gezer NS, Barış M, et al., 2021. CHAOS challenge-combined (CT-MR) healthy abdominal organ segmentation. *Med Image Anal*, 69:101950. https://doi.org/10.1016/j.media.2020.101950

Li JQ, Wang Z, Zhu SL, 2024. Mixed informed Transformer for few-shot medical image segmentation. Proc IEEE Int Conf on Acoustics, Speech and Signal Processing, p.1501-1505. https://doi.org/10.1109/ICASSP48485.2024.10448512

Lin Y, Chen YF, Cheng KT, et al., 2023. Few shot medical image segmentation with cross attention Transformer. Proc 26th Int Conf on Medical Image Computing and Computer Assisted Intervention, p.233-243. https://doi.org/10.1007/978-3-031-43895-0_22

Liu JR, Yang H, Zhou HY, et al., 2024. Swin-UMamba: Mamba-based UNet with ImageNet-based pretraining. Proc 27th Int Conf on Medical Image Computing and Computer Assisted Intervention, p.615-625. https://doi.org/10.1007/978-3-031-72114-4_59

Liu Z, Lin YT, Cao Y, et al., 2021. Swin Transformer: hierarchical vision Transformer using shifted windows. Proc IEEE/CVF Int Conf on Computer Vision, p.9992-10002. https://doi.org/10.1109/ICCV48922.2021.00986

Long J, Shelhamer E, Darrell T, 2015. Fully convolutional networks for semantic segmentation. Proc IEEE Conf on Computer Vision and Pattern Recognition, p.3431-3440. https://doi.org/10.1109/CVPR.2015.7298965

Luo XD, Chen JN, Song T, et al., 2022a. Semi-supervised medical image segmentation through dual-task consistency. Proc 25th AAAI Conf on Artificial Intelligence, p.8801-8809. https://doi.org/10.1609/aaai.v35i10.17066

Luo XD, Wang GT, Liao WJ, et al., 2022b. Semi-supervised medical image segmentation via uncertainty rectified pyramid consistency. *Med Image Anal*, 80:102517. https://doi.org/10.1016/j.media.2022.102517

Ma J, Chen JN, Ng M, et al., 2021. Loss odyssey in medical image segmentation. *Med Image Anal*, 71:102035. https://doi.org/10.1016/j.media.2021.102035

Ma J, Li FF, Wang B, 2024. U-Mamba: enhancing long-range dependency for biomedical image segmentation. https://arxiv.org/abs/2401.04722

Marín D, Aquino A, Gegundez-Arias ME, et al., 2011. A new supervised method for blood vessel segmentation in retinal images by using gray-level and moment invariants-based features. *IEEE Trans Med Imag*, 30(1):146-158. https://doi.org/10.1109/TMI.2010.2064333

Ouyang C, Biffi C, Chen C, et al., 2020. Self-supervision with superpixels: training few-shot medical image segmentation without annotation. Proc 16th European Conf on Computer Vision, p.762-780. https://doi.org/10.1007/978-3-030-58526-6_45

Pani K, Chawla I, 2024. Synthetic MRI in action: a novel framework in data augmentation strategies for robust multimodal brain tumor segmentation. *Comput Biol Med*, 183:109273. https://doi.org/10.1016/j.compbiomed.2024.109273

Patel G, Dolz J, 2021. Weakly supervised segmentation with cross-modality equivariant constraints. *Med Image Anal*, 77:102374. https://doi.org/10.1016/j.media.2022.102374

Qu H, Wu PX, Huang QY, et al., 2019. Weakly supervised deep nuclei segmentation using points annotation in histopathology images. Proc 2nd Int Conf on Medical Imaging with Deep Learning, p.390-400.

Shen QQ, Li YN, Jin JY, et al., 2023. Q-Net: query-informed few-shot medical image segmentation. In: Arai K (Ed.), Intelligent Systems and Applications. Springer, Cham, p.610-628. https://doi.org/10.1007/978-3-031-47724-9_40

Shen Y, Fan WS, Wang C, et al., 2024. Dual-guided prototype alignment network for few-shot medical image segmentation. *IEEE Trans Instrum Meas*, 73:5022513. https://doi.org/10.1109/TIM.2024.3411136

Tang YC, Yang D, Li WQ, et al., 2022. Self-supervised pre-training of Swin Transformers for 3D medical image analysis. Proc IEEE/CVF Conf on Computer Vision and Pattern Recognition, p.20698-20708. https://doi.org/10.1109/CVPR52688.2022.02007

Teng PR, Liu WJ, Wang XS, et al., 2024. Beyond singular prototype: a prototype splitting strategy for few-shot medical image segmentation. *Neurocomputing*, 597:127990. https://doi.org/10.1016/j.neucom.2024.127990

Teng S, Wu JW, Chen YY, et al., 2022. Semi-supervised leukocyte segmentation based on adversarial learning with reconstruction enhancement. *IEEE Trans Instrum Meas*, 71:5015511. https://doi.org/10.1109/TIM.2022.3189637

Wang KX, Liew JH, Zou YT, et al., 2019. PANet: few-shot image semantic segmentation with prototype alignment. Proc IEEE/CVF Int Conf on Computer Vision, p.9196-9205. https://doi.org/10.1109/ICCV.2019.00929

Wu XX, Gao ZG, Chen XW, et al., 2024. Support-query prototype fusion network for few-shot medical image segmentation. https://arxiv.org/abs/2405.07516

Zhan GD, Wang WT, Sun HY, et al., 2022. Auto-CSC: a transfer learning based automatic cell segmentation and count framework. *Cyborg Bion Syst*, 2022(1):9842349. https://doi.org/10.34133/2022/9842349

Zhang GW, Kang GL, Yang Y, et al., 2021. Few-shot segmentation via cycle-consistent Transformer. Proc 34th Annual Conf on Neural Information Processing Systems, p.1-13.

Zhang YM, Li HL, Gao YJ, et al., 2024. Prototype correlation matching and class-relation reasoning for few-shot medical image segmentation. *IEEE Trans Med Imag*, 43(11):4041-4054. https://doi.org/10.1109/TMI.2024.3412420

Zhou ZW, Rahman Siddiquee MM, Tajbakhsh N, et al., 2018. UNet++: a nested U-Net architecture for medical image segmentation. Proc 4th Int Workshop on Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, p.3-11. https://doi.org/10.1007/978-3-030-00889-5_1

Zhu LH, Liao BC, Zhang Q, et al., 2024. Vision Mamba: efficient visual representation learning with bidirectional state space model. Proc 41st Int Conf on Machine Learning, p.1-13.

Zhu YZ, Wang SD, Xin T, et al., 2023. Few-shot medical image segmentation via a region-enhanced prototypical Transformer. Proc 26th Int Conf on Medical Image Computing and Computer Assisted Intervention, p.271-280. https://doi.org/10.1007/978-3-031-43901-8_26

Zhuang XH, 2018. Multivariate mixture model for myocardial segmentation combining multi-source images. *IEEE Trans Patt Anal Mach Intell*, 41(12):2933-2946. https://doi.org/10.1109/TPAMI.2018.2869576