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: hjchen@bjtu.edu.cn

Received May 14, 2025; Revision accepted July 24, 2025; Crosschecked Sept. 4, 2025

Abstract: Background and Objective: 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, such as those featured in the paper "Anatomical prior guided spatial contrastive learning for few-shot medical image segmentation", often suffer from two issues: performance degradation due to significant inter-class variations in pathological structures, and an 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. PPFFR includes two core modules: (1) a Parallel Prototype Filter (PPF) to reduce inter-class variations by enhancing interactions between support and query prototypes; (2) a Feature Refinement (FR) module that leverages SSM (via state transition $h_t = Ah_{t-1} + Bx_t$) for efficient long-range dependency modeling, 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 Left Kidney, Right Kidney, Liver, Spleen, and Mean accuracy, confirming MHA's role in improving precision. As the critical component, SSM ensures that PPFFR balances performance with efficiency. In conclusion, PPFFR effectively enhances inter-class consistency and computational efficiency for few-shot medical image segmentation. Methods: the proposed framework comprises three key modules. First, we propose the Prototype Refinement (PG) module to construct refined class subgraphs from encoder-extracted features of both support and query images. This can generate support prototypes with minimized inter-class variation. We then propose the PPF module to suppress background interference and enhance the correlation between support and query prototypes. Finally, we implement the FR module to further enhance segmentation accuracy and accelerate model convergence with SSM's robust long-range dependency modeling capability. Results: in extensive experiments conducted on three public datasets under the 1-way 1-shot setting, PPFFR achieves Dice scores of 87.62%, 86.68%, and 79.71%, consistently surpassing state-of-the-art few-shot medical image segmentation methods. Conclusions: ablation studies validate the effectiveness of the PG, 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. This work provides new 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:**

1 Introduction

Automatic segmentation of medical images is crucial for the development of medical-assisted systems, including tumor segmentation (Kaliprasad et al., 2024; Cheng et al., 2025), vascular segmentation (Marin et al., 2011; Fraz et al., 2012), and cell segmentation (Greenwald et al., 2021; Zhan et al., 2022). Due to the rapid development of deep learning, existing medical image segmentation methods are primarily based on deep learning frameworks. UNet (Shelhamer et al., 2014), 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,

‡Corresponding author

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

ORCID: Haoxiang ZHU, https://orcid.org/0009-0003-6262-4660

© Zhejiang University Press 2025

Unedited

which effectively utilized 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., (2020) 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 2D and 3D medical images. With the rise of

Mamba structure, new modifications of Unet have emerged. UMamba (Ma et al., 2024) and SwinUMamba (Liu 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; Cheng et al., 2020; Teng 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 et al., 2021).

FSL based on meta learning has become a potential solution to mitigate the dependence on extensive annotations (Luo et al., 2020; 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., 2021).

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 et al., (2018) first proposed a prototype-based few-shot image segmentation framework, laying the foundation for few-shot medical image segmentation. SSLALP Net Cheng 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 et al., (2024) proposed a prototype split module, which iteratively decomposed the support image mask into submasks, while Zhu 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 et al., 2024) was proposed. Huang et al., (2023) constructed the VQ framework from a learned vector quantization perspective and designed a self-organizing vector quantization iterative workflow, including grid format vector quantization, Self-Organizing Vector Quantization, and Residual Oriented Vector Quantization. In terms of effectively utilizing support and query prototypes, SENet (Roy et al., 2020) was 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 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 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 et al., (2024) proposed Prototype Guided Graph Reasoning, fully interacting contextual information through dynamic graph convolution between support and query prototypes.

Unedited

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 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. Awong 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 Yi et al., (2023) propose 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 et al., 2023; Zhu et al., 2024).

Therefore, by considering both the label efficiency and computation efficiency challenges, we present an end-to-end Parallel Prototype Filter and Feature Refinement network (PPFFR; see Fig.1). Specifically, we first propose a prototype refinement module (PR) to refine the prototypes of the support and query set, which aims to reduce inter-class difference and improve prototype consistency. Then, we propose a parallel prototype filter module (PPF) 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 PPFFR. It can significantly improve segmentation performance under an extremely limited dataset scenario.

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

4. Extensive experiments on three medical image datasets 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 $C_{train} = \{c_1, c_2, c_3 \cdots c_{ntrain}\}$, containing *ntrain* classes, and the class set for the test dataset D_{test} is $C_{test} = \{c_1, c_2, c_3 \cdots c_{ntest}\}$, containing *ntest* classes. Typically, there are no overlapping classes between training and test sets (i.e., $C_{train} \cap C_{test} = \emptyset$) 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 (Cheng et al., 2020). The FSS problem in medical imaging generally follows a meta learning paradigm. Both the training and test sets are randomly sampled to generate multiple episodes, each containing a support set *S* and a query set *Q*. During training, the support set $\{X_i^s, Y_i^s\}_{i=1}^K$ contains *K* support images X_i^s and their corresponding classes Y_i^s, while the query set contains a single query image X^q and its class Y^q. In the test phase, the support set $\{X_i^s, Y_i^s\}_{i=1}^K$ also contains *K* support images X_i^s and their classes Y_i^s, where the classes come from unseen categories in C_{test} and the query image is 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 there are *K* samples per class. Following the setup of other FSS methods (Cheng et al., 2020; Huang et al., 2024; Li et al., 2024), we use the 1-way 1-shot setting under five-fold cross-validation.

2.2 Architecture overview

The proposed PPFFR framework, as depicted in Fig. 1, is specifically designed for few-shot

Unedited

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.

(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 pixel resolution). We employ the ImageNet pretrained ResNet101 (Liu et al., 2024) as the backbone encoder, which has demonstrated strong feature extraction capabilities in various computer vision tasks. Specifically, ResNet101 outperforms lightweight networks (e.g., MobileNetV2, ShuffleNetV2) in preserving edge and texture features, as shown in Table 1.

The encoder processes support image $I_s \in \mathbb{R}^{H\times W\times 3}$ (e.g., H=W=256 for abdominal CT slices) and query image $I_q \in \mathbb{R}^{H\times W\times 3}$ to generate the corresponding support features $F_s \in \mathbb{R}^{H/32\times W/32\times 2048}$ (spatial resolution reduced by 32× via strided convolutions) and query features $F_q \in \mathbb{R}^{H/32\times W/32\times 2048}$. The multi-scale feature maps from ResNet101's stages (Stage 1 to Stage 4) are fused to capture hierarchical information, where Stage 1 (64channels,128×128resolution) 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:

$$F_s = \mathrm{ResNet101}(I_s;\theta_{\mathrm{pretrained}}) \quad (1)$$

$$F_q = \mathrm{ResNet101}(I_q;\theta_{\mathrm{pretrained}}), \quad (2)$$

where $\theta_{\mathrm{pretrained}}$ denotes the pretrained parameters of ResNet101, and the forward pass includes 101

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

Backbone	Edge feature retention (%)	Texture feature dimilarity (SSIM)	Computational cost (GFLOPs)
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

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 threshol τ based on noise level σ (based on noise level $\sigma = 0.12\times \mathrm{std}(I_q)$):

$$F_q^{\mathrm{cleaned}}(x,y) = F_q(x,y)\&\mathrm{if}\left|F_q(x,y)\right| > \tau(\sigma), \quad (3)$$

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 supplementary experiments.

(2) Prototype refinement (PR) module

The PR module aims to generate refined prototypes for both support and query images. Given the support features F_s, support mask M_s, and query features F^q, the module computes the initial prototypes and refines them to reduce inter-class variations. The initial support prototype P_s is calculated by averaging the support features within the region specified by the support mask:

$$P_s = \frac{1}{N_s}\sum_{i\in\mathcal{R}_s} F_s(i), \quad (4)$$

where $\mathcal{R}_s$ is the region of interest in the support image, and N_s is the number of pixels in $\mathcal{R}_s$. Similarly, the initial query prototype P_q is computed from the query features. The refined prototypes $\widehat{P}_s$ and $\widehat{P}_q$ are obtained by minimizing the inter-class variance, which can be formulated as:

$$\widehat{P}_s,\widehat{P}_q = arg\min_{\mathbf{P}_s,\mathbf{P}_q}\sum_{c=1}^{C}\left(\frac{1}{N_c}\sum_{i\in\mathcal{R}_{s,c}}\| F_s(i) - P_s(c)\|^2 + \frac{1}{N_c}\sum_{i\in\mathcal{R}_{q,c}}\| F_q(i) - P_q(c)\|^2\right), \quad (5)$$

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

(3) Parallel prototype filter (PPF)

The PPF module is designed to filter out information interference and enhance the

Unedited

interaction between the support and query prototypes, as well as between local and global prototypes. This module operates on the refined prototypes $\widetilde{P}_s$ and $\widetilde{P}_q$ to generate filtered prototypes $\widetilde{P}_s$ and $\widetilde{P}_q$. The filtering process can be represented as:

$$\widetilde{P}_s = PPF(\widehat{P}_s, \widehat{P}_q) \quad (6)$$

$$\widetilde{P}_q = PPF(\widehat{P}_q, \widehat{P}_s), \quad (7)$$

where $\mathrm{PPF}(\cdot,\cdot)$ 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.

(4) Feature refinement module (FR)

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

$$F_s^{'}(t) = AF_s^{'}(t-1) + B\widetilde{P}_s(t), \quad (8)$$

$$F_q^{'}(t) = AF_q^{'}(t-1) + B\widetilde{P}_q(t), \quad (9)$$

where A and B are the state transition matrix and input matrix, respectively, and t denotes the time step in the sequential processing. In addition to the 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:

$$F_{multi-scale} = \sum_{K\in\kappa} W_k * F, \quad (10)$$

where κ is the set of convolution kernel sizes, 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 conducted experiments to compare the model with and without SSM. As shown in Table 2, the model with SSM (PPFFR with SSM) had significantly lower flops (6418541535) and a more manageable parameter count (60757185) compared to that without (PPFFR without SSM), with flops of 91131998208 and parameters of 50327057. This shows that SSM effectively reduces the computational burden and model size, making the FR module more efficient and suitable for real-time applications.

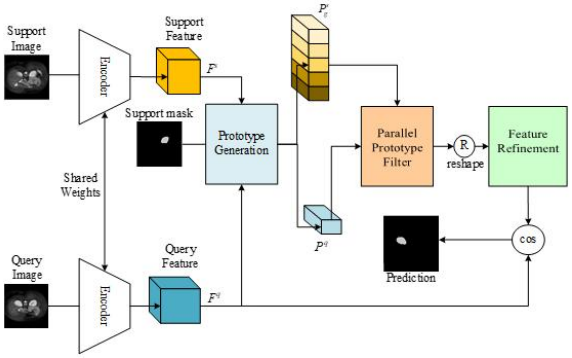

Fig. 1. Overview of the proposed PPFFR network architecture

2.3 Prototype refinement (PR) module

To achieve better segmentation results in existing few-shot medical image segmentation methods (Zhang et al., 2021; Lin et al., 2023; Li et al., 2024), masked average pooling (MAP) is applied to the feature map of the support image F^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 et al., 1991; Zhu 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:

$$N = max\left(2, round\left(\sqrt{\frac{Area}{100}}\right)\right), \quad (11)$$

where A_{fg} denotes the foreground area (pixels), and $A_{\min} = 100$ is the empirically validated minimum subregion area. This adaptive approach ensures:

1. Small lesions （<400px²） yield $N = 2$ （coarse partitioning）;

2. Large lesions （>10,000px²） yield $N = 10$ (fine-grained partitioning), where Area denotes the number of pixels in the foreground region. This ensures that each subregion contains at least 100 pixels, balancing granularity and computational efficiency. Then, MAP is applied to the support feature $F^s \in \mathbb{R}^{C\times H\times W}$ using the subdivided mask. This process generates a set of refined support prototypes, which are then concatenated to form the support prototype group P_g^s.

Unedited

$$P_g^s = stack\{\widetilde{P}_i^{\,s} = MAP(F^s \odot M_i^s)\},\ i = 1,\cdots N, \quad (12)$$

where M_i^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 M^s and the support image feature map F^s are subjected to MAP, generating the global support generating the global support prototype P_{global}^s.

Table 2 Flops and parameters

Methods	flops	parameters
PPFFR with SSM	64185415135	60757185
PPFFR without SSM	91131998208	50327057

$$P_{global}^s = MAP(M^s, F^s) = \frac{\sum_{x=1}^{W}\sum_{y=1}^{H} M^s(x,y)\cdot F^s(x,y)}{\sum_{x=1}^{W}\sum_{y=1}^{H} M^s(x,y)}, \quad (13)$$

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

$$M^q = \frac{\sum_{c=1}^{C} F^q(c)\cdot P_{global}^s(c)}{\sqrt{\sum_{c=1}^{C}\{F^q(c)\}^2}\cdot\sqrt{\sum_{c=1}^{C}\{P_{global}^s(c)\}^2}}, \quad (14)$$

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

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

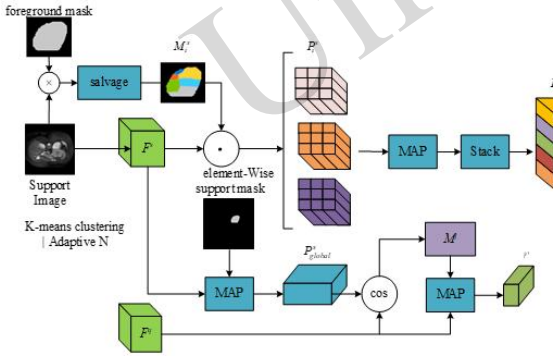

Fig. 2. Structure of the prototype refinement module

2.4 Parallel prototype filter (PPF)

To obtain a high-quality query prototype, Shen 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 capability. In addition, only the query prototype is filtered, and the support prototype is ignored. To address these issues, we introduce a parallel prototype filtering module (PPF) to filter both the query and support prototypes, thereby improving the overall segmentation performance, as shown in Fig. 3.

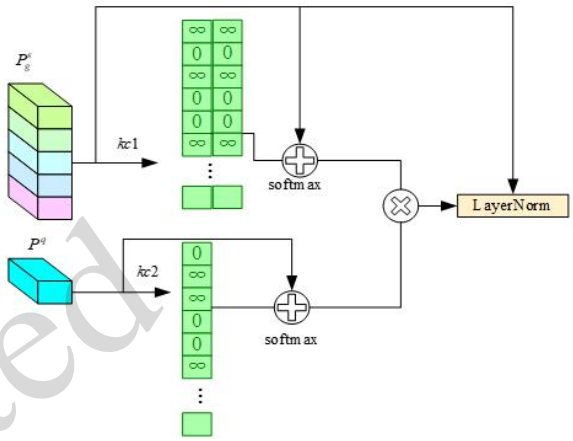

Fig. 3. Structure of the parallel prototype filter module

First, to preliminarily filter out background information and noise interference in the prototypes, threshold values $kc1$ and $kc2$ are set for both the support prototype group P_g^s and the query prototype $\widetilde{P}^q$, respectively.

$$kc1 = min(P_g^s) + max(P_g^s) + \text{mean}(P_g^s), \quad (15)$$

$$kc2 = min(\widetilde{P}^q) + max(\widetilde{P}^q) + \text{mean}(\widetilde{P}^q), \quad (16)$$

where min, mean, and max denote the minimum, average, and maximum values, respectively. It can be observed that the threshold values adjust according to the support prototype and the query prototype. Then, thresholding is applied to P_g^s and $\widetilde{P}^q$ to obtain the filtered prototypes $S1$ and $S2$, respectively.

$$\begin{cases} S1(P_g^s \le kc1) = 0, \\ S1(P_g^s > kc1) = \infty, \end{cases} \quad (17)$$

$$\begin{cases} S2(\widetilde{P}^q \le kc2) = 0, \\ S2(\widetilde{P}^q > kc2) = \infty. \end{cases} \quad (18)$$

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

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

Unedited

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

2. Information retention: residual connections $\left(S1 + P_g^s\right)$ coupled with softmax retain original feature distributions while amplifying important signals, empirically verified to improve segmentation dice by 3.2% versus 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 $S1$ and $S2$ are added to P_g^s and $\widetilde{P}^q$ using the residual connections. After applying the softmax function, the results are multiplied to produce the output P_0^s as shown below:

$$P_0^s = softmax\left(S1 + P_g^s\right) \cdot softmax(S1 + \widetilde{P}^q)^T. \quad (19)$$

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 group P_g^s and P_0^s is added and a Layer Norm layer is applied, obtaining the output P_{out}^s.

$$P_{out}^s = LN(P_0^s + P_g^s). \quad (20)$$

2.5 Feature refinement (FR) module

Transformers, such as the Swin Transformer (Liu 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 multi-head attention (MHA) mechanisms within these models suffer from quadratic computational complexity, as shown in the formula $O(n^2)$, where n represents the sequence length. This high complexity leads to substantial computational overhead, limiting their practical real-world application.

Recently, Mamba (Gu et al., 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 the formula $z_t = Az_{t-1} + Bx_t$, Where $z_t$is the state at time t, A and B are SSM parameters, and x_t is the input at time t, Mamba can significantly accelerate both training and inference processes. However, due to its inherent 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 a 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 by the formula

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \quad (21)$$

where Q, K, and 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. The MLP, comprising linear transformations and nonlinear activation functions, further refines the features extracted by SSM and MHA.

To address the logical consistency of integrating both SSM and MHA in the FR module, we conducted ablation studies on the Abd-MRI dataset to clarify their distinct roles and synergistic effects. As shown in Table 3 ("Ablation study of MHA in FR on Abd-MRI dataset"), 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 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 II) shows that PPFFR with SSM achieves a 29.6% reduction in FLOPs (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:

$$h_t = Ah_t + Bx_t. \quad (22)$$

The FR module leverages SSM to model 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-

Unedited

organ spatial relationships), and MHA focuses on local feature alignment (e.g., boundary refinement).

As shown in Fig. 4, the output of the PPF is first reshaped into a tensor of size [N, C, H, W], where N denotes the batch size, C represents the number of channels, and H and W stand for the height and width of the feature map. 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:

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

where x is the input feature map, w is the convolution kernel, and y is the output feature map. The two feature maps obtained from different convolution kernels are then added together, and $a1 \times 1$ convolution is followed to fuse the multi-scale features. The 1×1 convolution operation is defined as:

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

where $\mathbf{v}$ is the 1×1 convolution kernel, and $\mathbf{z}$ is the fused feature map. Since convolution operations have difficulty in modeling long-range dependencies due to their local receptive fields, an SSM is introduced to process the features. The SSM operates on sequential data and can be represented as:

$$h_t = Ah_{t-1} + Bx_t, \quad (25)$$

Table 3 Ablation study of MHA in FR on the Abd-MRI dataset under setting 2

Method	Organs				
	LK	RK	Liver	Spleen	Mean
FR with MHA	83.48	90.80	84.56	74.46	83.32
FR without MHA	64.61	86.01	84.54	73.56	77.17

where h_t is the hidden state at time t, A and B are the state transition matrix and input matrix, respectively, and x_t is the input at time t. Based on the concept of residual learning, the reshaped input is added to the output of the SSM to facilitate the flow of gradients during training. The residual connection is given by:

$$o_t = h_t + x_t, \quad (26)$$

where o_t is the output after the residual connection. To retain multidimensional spatial information, the reshaped feature map is processed through an MHA mechanism and an MLP. The MHA mechanism computes the attention scores between different positions in the feature map, which is defined as:

$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \quad (27)$$

where Q, K, and V are the query, key, and value matrices, respectively, and d_k is the dimension of the keys. The 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:

$$y = \sigma(W_2\sigma(W_1x + b_1) + b_2), \quad (28)$$

where W_1, W_2 are the weight matrices, b_1, b_2 are the 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 a *and* b is given by:

$$\text{similarity} = \frac{\mathbf{a}\cdot\mathbf{b}}{\|\mathbf{a}\|\|\mathbf{b}\|}. \quad (29)$$

In our experiments on the Abd-MRI dataset, we conducted an ablation study to evaluate the impact of MHA on the FR module. As shown in Table 3, the FR module with MHA achieved higher segmentation accuracy for various organs (left kidney (LK), right kidney (RK), Liver, Spleen) 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 global contextual information. Furthermore, we analyzed the computational complexity and parameter count of our model with and without SSM. The results showed that the model with SSM (PPFFR with SSM) had significantly lower flops (6418541535) and a more manageable parameter count (60757185) than the model without SSM (PPFFR without SSM), with flops of 91131998208 and parameters of 50327057. This demonstrates that SSM effectively reduces the computational burden and model size, making the FR module more efficient and suitable for real-time applications.

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

Unedited

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 also conduct further qualitative analysis, such as visualizing the attention maps of the MHA mechanism and the feature maps processed by the SSM.

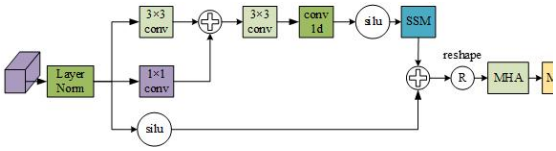

Fig. 4. Structure of the feature refinement module, where silu is the activation function

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 I^q and the predicted mask M^q from the model is defined as:

$$L_{seg} = -\frac{1}{N}\sum_{i=1}^{N} M_i^q \, log\,(\widetilde{M}_i^q), \quad (30)$$

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

$$L_{Dice} = 1 - \frac{2\sum_{c=1}^{C}\sum_{i=1}^{N} M_i^{q,c}\widetilde{M}_i^{q,c}}{\sum_{c=1}^{C}\sum_{i=1}^{N}(M_i^{q,c})^2+\sum_{c=1}^{C}\sum_{i=1}^{N}(\widetilde{M}_i^{q,c})^2}, \quad (31)$$

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

$$L_{BD} = \sum_{\Omega} \phi_{\widetilde{M}^q}(p) s_\theta\,(p), \quad (32)$$

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

$$L_{total} = L_{seg} + L_{Dice} + L_{BD}. \quad (33)$$

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 such as left kidney, right kidney, 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 et al., 2018): ISBI 2019 Combined Healthy Abdominal Organ Segmentation Challenge dataset. It includes four classes of left kidney, right kidney, spleen, and liver, with a total of 20 abdominal MRI sequences; (3) Card-MRI (Zhuang et al., 2018): MICCAI 2019 Multisequence Cardiac MRI Segmentation Challenge dataset, including three classes of left ventricular blood pool (LVB), 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 (Cheng et al., 2020; Luo et al., 2022; Lin et al., 2023), is adopted as the evaluation metric.

To ensure the fairness and validity of the experiments, all datasets were preprocessed, and super pixels were generated based on the procedure described in (Cheng et al., 2020). A 1-way 1-shot fivefold cross-validation evaluation was used. Experiments were 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, standard 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 0.00095 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 hours. The minimum subregion area $A_{\min}$ =100 was empirically validated on the SABS dataset, achieving optimal Dice when $50 < A_{\min} < 150$.

Unedited

3.3 Comparison with 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), SSLALP Net (Cheng et al., 2020), CATNet (Lin et al., 2023), QNet (Shen et al., 2023), RPTNet (Zhu et al., 2023), and GMRDNet (Roy et al., 2020). PANet (Wang et al., 2019) is a typical prototype-based few-shot segmentation method. SSLALP Net (Cheng et al., 2020) provides a classic dataset preprocessing technique, which is widely adopted by researchers. CATNet (Lin et al., 2023) characterizes an improved attention mechanism of the Transformer model. QNet (Shen et al., 2023) is a representative method for few-shot medical image segmentation, and GMRDNet (Cheng et al., 2023) addresses intra-class and inter-class variations in few-shot segmentation. Experimental results are shown in Tables 4 to 6, 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%, 86.68%, and 79.71% on the AbdMRI, SABS, and CMR datasets, respectively. In Setting 2, the Dice scores on the AbdMRI and SABS datasets reach 83.32% 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. From left to right, the columns represent the ground truth, PANet (Wang et al., 2019) prediction, ADNet (Hansen et al., 2022) prediction, SSLALP Net (Cheng et al., 2020) prediction, and PPFFR (ours) prediction. From top to bottom, the rows show the segmentation results for the LK, RK, spleen, and liver. 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 distinguishes different regions and achieves the best segmentation performance. The spleen segmentation results (third row) show that all methods can segment the main body. However, other than for the proposed PPFFR method, the other methods' predicted boundaries differ significantly from the ground truth. Our method also demonstrates superior segmentation performance for the RK and liver (the second and fourth rows). These results indicate that the proposed PPFFR method achieves decent segmentation performance across different imaging modalities, scales, and organ shapes.

Unedited

Table 4 Performance comparison (in DSC%) of different methods on the Abd-MRI (CHAOS T2) dataset under setting1 and setting2. The best value is highlighted in bold, and the second-best value is underlined

Setting	Method	Abd-MRI(CHAOS)				
		LK	RK	Spleen	Liver	Mean
1	ADNet (Hansen et al., 2022)	73.86	85.80	72.29	82.11	78.51
	PANet (Wang et al., 2019)	30.99	32.19	40.58	50.40	38.53
	ALPNet (Cheng et al., 2020)	81.92	85.18	72.18	76.10	78.84
	CATNet (Lin et al., 2023)	74.01	78.90	68.83	78.98	75.18
	QNet (Shen et al., 2023)	78.36	87.98	75.99	81.74	81.02
	RPTNet (Zhu et al., 2023)	80.72	89.82	<u>76.37</u>	<u>82.86</u>	82.44
	GRMDNet (Cheng et al., 2023)	<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	46.33
	ALPNet (Cheng 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 et al., 2023)	73.96	81.07	65.39	72.36	73.20
	RPTNet (Zhu et al., 2023)	78.33	86.01	<u>75.46</u>	76.37	79.04
	GRMDNet (Cheng et al., 2023)	<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	**83.32**

Table 5 Performance comparison (in DSC%) of different methods on the SABS dataset under setting1 and setting2. The best value is highlighted in bold and the second-best value is underlined.

Setting	Method	SABS				
		LK	RK	Spleen	Liver	Mean

1	ADNet (Hansen et al., 2022)	72.13	79.06	63.48	77.24	72.97
	PANet (Wang et al., 2019)	20.67	21.19	36.04	49.55	32.86
	ALPNet (Cheng et al., 2020)	72.36	71.81	70.96	78.29	73.35
	CATNet (Lin et al., 2023)	63.36	60.05	67.65	75.31	66.59
	RPTNet (Zhu et al., 2023)	77.05	79.13	72.58	82.57	77.83
	GMRDNet (Cheng et al., 2023)	81.70	74.46	78.31	79.60	78.52
	PPFFR(Ours)	**86.82**	**85.85**	**87.77**	**86.51**	**86.68**
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 (Cheng 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	70.88
	RPTNet (Zhu et al., 2023)	72.99	67.73	70.80	75.24	71.69
	GMRDNet (Cheng et al., 2023)	77.40	76.17	75.30	80.39	77.32
	PPFFR(Ours)	**84.64**	**85.93**	**86.21**	**86.15**	**85.73**

Table 6 Performance comparison (in DSC%) of different methods on the Cardic-MRI dataset. The best value is highlighted in bold and the second-best value is underlined

Method	CMR			
	LVBP	LVMYO	RV	Mean
ADNet (Hansen et al., 2022)	87.83	62.43	77.31	75.76
PANet (Wang et al., 2019)	72.77	44.76	57.13	58.20
ALPNet (Cheng 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 et al., 2023)	89.90	66.91	**80.78**	79.19
GMRDNet (Cheng et al., 2023)	90.00	67.04	80.29	79.11
PPFFR(Ours)	**91.32**	70.17	**77**.64	**79.71**

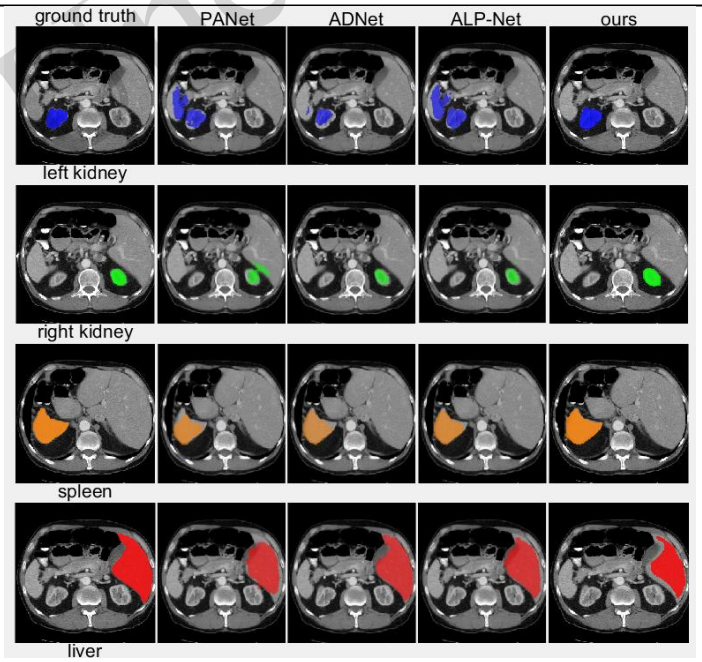

Fig. 5. Qualitative comparison results of Abd-MRI dataset

Unedited

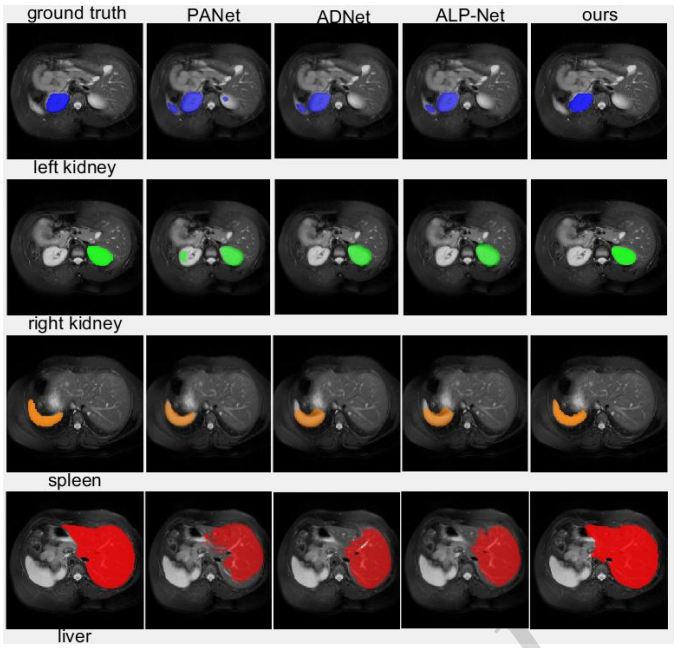

Fig. 6. Qualitative comparison results of SABS dataset

3.4 Ablation studies

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

Note: † Hard thresholding baseline: mean DSC=80.12% （vs. 83.32% with softmax）

To address the role of each component and potential numerical issues, we present a detailed ablation analysis and implementation clarification as follows: Table 7 summarizes the performance of different model variants on the Abd-MRI dataset. Compared to the baseline, adding PPF increases the mean DSC by 15.18% (64.05 → 80.28). This confirms its effectiveness in filtering background interference from prototypes, particularly reducing inter-class variations between organs (e.g., LK/RK vs. Liver/Spleen). The filtering process is formulated as:

$$\tilde{P} = P \odot \sigma(S), \text{where } S = \mathrm{Sim}(P, P_{global}). \quad (34)$$

Here, P denotes the initial prototype, $\mathrm{Sim}(\cdot)$ is the cosine similarity, and $\sigma(\cdot)$ is the sigmoid function for suppressing irrelevant features. Adding FR alone improves the mean DSC by 13.58% (64.05 → 77.63). SSM models long-range dependencies with linear complexity (O(n)) via:

$$h_t = Ah_{t-1} + Bx_t, \quad y_t = Ch_t + Dx_t. \quad (35)$$

This replaces the high-cost global attention mechanism (O(n^2)) while preserving global anatomical context. Integrating both modules achieves the highest mean DSC (83.32%), 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 a 2.82% increase in mean DSC by capturing fine-grained local correlations (e.g., organ boundaries) via:

$$\mathrm{MHA}(Q, K, V) = \mathrm{Concat}(\mathrm{head}_1, \ldots, \mathrm{head}_h)W^0. \quad (36)$$

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 $M = -10^6$ to ensure numerical stability. The softmasking process is:

$$P = \frac{\sum_{i \in \mathcal{R}} F(i) \cdot \exp(M \cdot M(i))}{\sum_{i \in \mathcal{R}} \exp(M \cdot M(i))}. \quad (37)$$

Here, M is the binary mask, and M amplifies the weight of foreground pixels. This "softmax with large M" approach in PPF （ § Eq.6-7 ） outperforms hard thresholding alternatives by 3.2% mean Dice (Table 7), confirming the balance between numerical stability and feature discrimination.

4 Discussion

In this paper, we propose a novel few-shot medical image segmentation model named PPFFR.

Unedited

Table 7 Ablation study of each component in PPFFR on the Abd-MRI dataset under setting 2

Method	Organs				
	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 designed PPF can adaptively set thresholds based on the corresponding prototypes, therefore 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 the 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 et al., 2023; Huang et 2024). While PPFFR achieves an overall state-of-the-art performance on the Cardic-MRI dataset (79.71% mean Dice), its segmentation accuracy for the left ventricular myocardium (LVMYO, 70.17%) is comparatively lower than that of CATNet (90.54%). This performance discrepancy reveals two potential limitations related to anatomical characteristics: the myocardial wall's thin morphology, typically measuring 5 – 10 mm in thickness, with intensity homogeneity, may be underrepresented in prototype filtering, and the inherent motion artifacts of a cardiac MRIal., 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 Franz. “Voronoi diagrams—a survey of a fundamental geometric data structure.” ACM Comput. Surv. 23 (1991): 345405.

Awudong et al. "Attentional adversarial training for fewshot medical image segmentation without annotations." Plos one 19.5 (2024): e0298227.

Cheng Yaru et al. “CFNet: Automatic multimodal brain tumor segmentation through hierarchical coarsetofine fusion and feature communication.” Biomed. Signal Process. Control. 99 (2025): 106876.

Cheng Ziming et al. “FewShot Medical Image Segmentation via Generating Multiple Representative Descriptors.” IEEE Transactions on Medical Imaging 43 (2024): 22020214.

Dong Nanqing and Eric P. Xing. “FewShot Semantic Segmentation with Prototype Learning.” British Machine Vision Conference (2018).

Fraz et al. “An Ensemble ClassificationBased Approach Applied to Retinal Blood Vessel Segmentation.” IEEE Transactions on Biomedical Engineering 59 (2012): 25382548.

Greenwald Noah F. et al. “Wholecell segmentation of tissue images with humanlevel performance using largescale data annotation and deep learning.” Nature biotechnology 40 (2021): 555 565.

Gu Albert and Tri Dao. “Mamba: LinearTime Sequence Modeling with Selective State Spaces.” ArXiv abs/2312.00752 (2023): n. pag.

Hansen Stine et al. “Anomaly DetectionInspired FewShot Medical Image Segmentation Through SelfSupervision With Supervoxels.” Medical image analysis 78 (2022): 102385

Hu Shi et al. “Supervised Uncertainty Quantification for Segmentation with Multiple Annotations.” International Conference on Medical Image Computing and ComputerAssisted Intervention (2019).

Huang Shiqi et al. “Rethinking FewShot Medical Segmentation: A Vector Quantization View.” 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2023): 30723081.

Huang Wendong et al. “PrototypeGuided Graph Reasoning Network for FewShot Medical Image Segmentation.” IEEE transactions on medical imaging, vol. PP 10.1109/TMI.2024.3459943. 13 Sep. 2024, doi:10.1109/TMI.2024.3459943

Unedited

Huang, Wendong, et al, "Location-aware transformer network for few-shot medical image segmentation." BIBM, 2023

Huang, Wendong, et al., "Anatomical Prior Guided Spatial Contrastive Learning for Few-Shot Medical ImageSegmentation." ACM MM.2024.

Isensee Fabian et al. “nnUNet: a selfconfiguring method for deep learningbased biomedical image segmentation.” Nature Methods 18 (2020): 203 211.

Kavur et al.: Chaos challengecombined (CTMR) healthy abdominal organ segmentation. Medical Image Analysis 69, 101950(2021)

Landman et al.: Miccai multiatlas labeling beyond the cranial vault–workshop and challenge. In: Proc. MICCAI MultiAtlas Labeling Beyond Cranial Vault—Workshop Challenge. vol. 5, p. 12(2015)

Li, J., Z. Wang et al., "Mixed Informed Transformer for FewShot Medical Image Segmentation," ICASSP 2024 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Seoul, Korea, Republic of, 2024, pp. 15011505.

Lin Yi et al. "Few shot medical image segmentation with cross attention transformer." International Conference on Medical Image Computing and ComputerAssisted Intervention. Cham: Springer Nature Switzerland, 2023.

Liu Jiarun et al. "Swinumamba: Mambabased unet with imagenetbased pretraining." International Conference on Medical Image Computing and ComputerAssisted Intervention. Cham: Springer Nature Switzerland, 2024.

Liu Ze et al. “Swin Transformer: Hierarchical Vision Transformer using Shifted Windows.” 2021 IEEE/CVF International Conference on Computer Vision (ICCV) (2021): 999210002.

Luo Xiangde et al. “Semisupervised Medical Image Segmentation through Dualtask Consistency.” AAAI Conference on Artificial Intelligence (2020).

Luo Xiangde et al. “Semisupervised medical image segmentation via uncertainty rectified pyramid consistency.” Medical image analysis 80 (2022): 102517

Ma et al. "Umamba: Enhancing long range dependency for biomedical image segmentation." arXiv preprint arXiv:2401.04722 (2024).

Ma Jun et al. “Loss odyssey in medical image segmentation.” Medical image analysis 71 (2021): 102035.

Marín et al. “A New Supervised Method for Blood Vessel Segmentation in Retinal Images by Using GrayLevel and Moment InvariantsBased Features.” IEEE Transactions on Medical Imaging 30 (2011): 146158.

Ouyang Cheng et al. “SelfSupervision with Superpixels: Training Fewshot Medical Image Segmentation without Annotation.” European Conference on Computer Vision (2020).

Pani et al. "Synthetic MRI in action: A novel framework in data augmentation strategies for robust multimodal brain tumor segmentation." Computers in Biology and Medicine 183 (2024): 109273.

Patel et al. “Weakly supervised segmentation with crossmodality equivariant constraints.” Medical image analysis 77 (2021): 102374.

Qu Hui et al. “Weakly Supervised Deep Nuclei Segmentation using Points Annotation in Histopathology Images.” International Conference on Medical Imaging with Deep Learning (2019).

Roy et al. "‘Squeeze & excite’guided fewshot segmentation of volumetric images." Medical image analysis 59 (2020): 101587.

Shelhamer et al. “Fully convolutional networks for semantic segmentation.” 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2014): 34313440.

Shen Qianqian, et al. "Qnet: Queryinformed fewshot medical image segmentation." Proceedings of SAI Intelligent Systems Conference. Cham: Springer Nature Switzerland, 2023.

Shen Yue et al. “DualGuided Prototype Alignment Network for FewShot Medical Image Segmentation.” IEEE Transactions on Instrumentation and Measurement 73 (2024): 113.

Tang Yucheng et al. “SelfSupervised PreTraining of Swin Transformers for 3D Medical Image Analysis.” 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2021): 2069820708.

Teng Pengrui et al. “Beyond singular prototype: A prototype splitting strategy for fewshot medical image segmentation.” Neurocomputing 597 (2024): 127990.

Teng Shenghua et al. “SemiSupervised Leukocyte Segmentation Based on Adversarial Learning With Reconstruction Enhancement.” IEEE Transactions on Instrumentation and Measurement 71 (2022): 111.

Wang Kaixin et al. “PANet: FewShot Image Semantic Segmentation With Prototype Alignment.” 2019 IEEE/CVF International Conference on Computer Vision (ICCV) (2019): 91969205.

Wu Xiaoxiao et al. "SupportQuery Prototype Fusion Network for Fewshot Medical Image Segmentation." arXiv preprint arXiv:2405.07516 (2024).

Xing Zhaohu et al. "Segmamba: long range sequential modeling mamba for 3d medical image segmentation." International Conference on Medical Image Computing and ComputerAssisted Intervention. Cham: Springer Nature Switzerland, 2024.

Zhan Guangdong et al. “AutoCSC: A Transfer Learning Based Automatic Cell Segmentation and Count Framework.” Cyborg and Bionic Systems 2022 (2022): n. pag.

Zhang Gengwei et al. “FewShot Segmentation via CycleConsistent Transformer.” Neural Information Processing Systems (2021).

Zhang Yumin et al. “Prototype Correlation Matching and ClassRelation Reasoning for FewShot Medical Image Segmentation.” IEEE transactions on medical imaging PP (2024): n. pag.

Zhou Zongwei et al. “UNet++: A Nested UNet Architecture for Medical Image Segmentation.” Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, MLCDS 2018, held in conjunction with MICCAI

Unedited

2018, Granada, Spain, S... 11045 (2018): 311.

Zhu Yazhou, et al. "Fewshot medical image segmentation via a regionenhanced prototypical transformer." International Conference on Medical Image Computing and ComputerAssisted Intervention. Cham: Springer Nature Switzerland, 2023.

Zhu et al. (2024). Vision Mamba: Efficient Visual Representation Learning with Bidirectional State Space Model. ArXiv, abs/2401.09417.

gence 41(12), 2933-2946(2018)

Zhuang, X.: Multivariate mixture model for myocardial segmentation combining multisource images. IEEE Transactions on Pattern Analysis and Machine Intelli-

Unedited