FFR-UNet: Feature Filter-Refinement UNet for Medical Image Segmentation

Weixin Xu[1]

[1]Beihang University

Abstract

Medical image segmentation poses a significant challenge in the field of computer vision. CNN-based methods often neglect long-range dependencies, and Transformer-based methods may overlook local context information. Moreover, in contrast to natural images, medical images present a distinct challenge wherein the foreground targets requiring segmentation are typically smaller, accompanied by a greater abundance of background information. To overcome these deficiencies, we propose a novel Feature Filter Module (FFM) designed to discern between informative and non-informative features. These features seamlessly transition into our proposed Feature Refinement Module (FRM), assigning them distinct roles to establish a robust connection between the two input features. Moreover, by integrating our proposed FFM and FRM into the encoder block of the UNet architecture, we introduce a novel framework named Feature Filter-Refinement UNet (FFR-UNet). Extensive experiments demonstrate the superiority of FFR-UNet, consistently achieving state-of-the-art (SOTA) performance.

Introduction

In this paper, to solve the above issues, we propose a novel Feature Filter Module (FFM) that filters out informative features and non-informative ones, to reduce the interference to the segmentation network. At the same time, to further refine the features and strengthen their interactions, we propose a Feature Refinement Module (FRM) to progressively suppress features in irrelevant background regions. Our main contributions are summarized as follows:

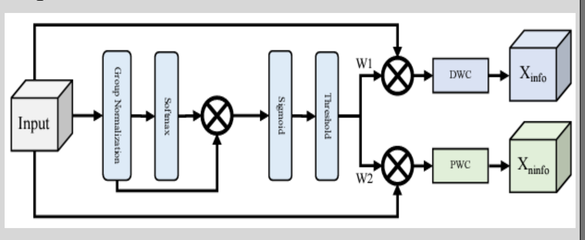
We propose a novel **Feature Filter Module (FFM),** aiming to enhance the network's ability to distinguish between informative and non-informative features to mitigate the impact of irrelevant features on the network and enhance the network's focus on crucial features.

Following the FFM, we introduce an innovative Feature Refinement Module. This module integrates the convolutional operation and cross-attention mechanism to not only refine features but also enhance interactions among features from the preceding layer. The distinctive combination of these two operations allows us to focus on long-range dependencies and local relations concurrently. This dual-focus approach sets our FRM apart, enriching the model's capacity to capture intricate dependencies across different spatial scales.

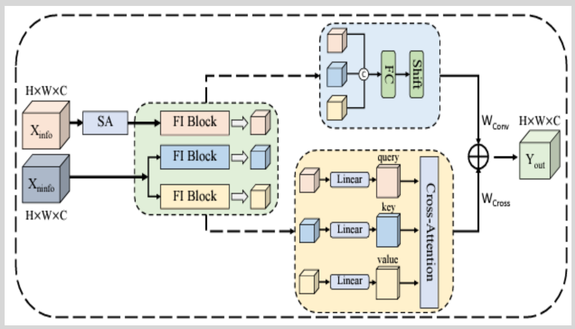
By incorporating the FFM and FRM into the encoder blocks of the UNet, we introduce a new framework dubbed Feature Filter-Refinement UNet (FFR-UNet). We evaluate our proposed FFR-UNet on three widely used public benchmarks for medical image segmentation tasks. Extensive experiments demonstrate the proposed FFR-UNet can achieve state-of-the-art (SOTA) performance.

Method

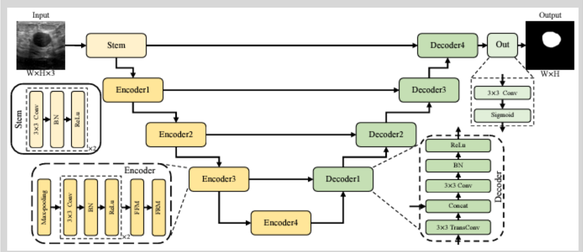
Feature Filter Module. As illustrated in Figure 1, The FFM is meticulously crafted to discriminate between informative and non-informative features in a multi-step process, adding a layer of sophistication to the segmentation framework. In detail, our proposed FFM encompasses the following steps: Firstly, a softmax operation, coupled with scaling factors in the Group Normalization (GN) layers. Subsequent to this evaluation, the weight values of feature maps, are normalized to the range (0, 1) through the sigmoid function, with a threshold of 0.5 applied to gate these values. This step yields informative weights, W1, by assigning weights above the threshold to 1. Simultaneously, non-informative weights, denoted as W2, are obtained by assigning weights below the threshold to 0. Following this weight discrimination, element-wise multiplication is performed on the two sets of weighted outputs, W1 and W2, with the input feature X. Finally, a DWC and a PWC are incorporated to get the final outputs.



Feature Refinement Module. Although non-informative features will introduce interference to the network, the rich structural information inherent is also crucial in medical images for segmentation tasks. Therefore, a direct and indiscriminate discarding of non-informative features is not advisable. This rationale underscores the introduction of the proposed Feature Refinement Module (FRM) following the feature filtering process, aiming to refine the features further and enhance their interactions. After analyzing convolution and cross-attention mechanisms, it's clear that each has distinct strengths and weaknesses. Convolution excels in local information processing but may overlook global context and long-range dependencies. To address this, we integrate the cross-attention mechanism into our Feature Refine Module (FRM), combining the strengths of both methods. This strategic fusion enables a comprehensive capture of intricate dependencies among extracted features, significantly enhancing our ability to refine embedded information by aggregating feature maps. The framework of our proposed FRM is shown in Figure 2.



Overall Architecture. The architectural framework of our Feature Filter-Refinement UNet (FFR-UNet) is illustrated in Figure 3. Within the confines of our architectural framework, we recognize that features derived from the convolutional layers encompass both informative and non-informative components. Being aware of the potential interference introduced by non-informative features in subsequent layers, addressing this concern is paramount to ensure the final segmentation results meet stringent quality standards. To this end, we introduce the Feature Filter Module (FFM) into the encoder layers, strategically positioned to discriminate between informative and non-informative features. Elevating this objective further, the Feature Refinement Module (FRM) is introduced. By synergistically integrating convolution and cross-attention mechanisms, the FRM is designed to refine features, fostering heightened interaction among distinct feature components. This nuanced approach aims to strike a delicate balance between preserving long-range dependencies and capturing local context information effectively, thereby enhancing the overall segmentation accuracy of the FFR-UNet.



Experiments

We evaluated our model on the BUSI, BUSIS, and TN3K datasets. The BUSI dataset comprises 780 breast ultrasound images, averaging 500×500 pixels, featuring normal, benign, and malignant cases of breast cancer with corresponding segmentations. Our focus on benign and malignant images (647 total) led to a balanced 7:1:2 split for training (453 images), validation (65 images), and testing (129 images). The BUSIS dataset includes 562 images from women aged 26 to 78 years, randomly divided into training (394 images), validation (56 images), and test sets (112 images) following a balanced 7:1:2 ratio. The TN3k dataset consists of 3493 thyroid ultrasound images with high-quality nodule masks, distributed for training and validation (2879 images) and testing (614 images) as per the official split.

To demonstrate the advantage of our proposed FFR-UNet, we compared it with seven widely used networks: UNet, ResUNet, AttUNet, TransUNet, UNeXt, CMUNet, and CMUNeXt. We re-implemented each of the compared approaches. We employ four commonly used metrics to perform a quantitative assessment of the performance of various segmentation models. These metrics include the Dice Similarity Coefficient (DSC), mean Intersection over Union (mIoU), Precision, and Recall.

The detailed results obtained on the BUSI and BUSIS datasets are presented in Table 1. Noteworthy advancements in performance metrics are observed when compared with previous state-of-the-art (SOTA) methods. Specifically, for the BUSI dataset, our model showcases improvements of 0.98% and 0.68% in mIoU and DSC, respectively. Similarly, on the BUSIS dataset, our method demonstrates enhancements of 0.75% and 0.48% in mIoU and DSC, respectively. The superior performance of our proposed method is further validated on the TN3K dataset, where we achieve significant improvements of 1.81% and 1.30% in terms of mIoU and DSC, as illustrated in Table 2.

Methods	BUSI dataset (%)				BUSIS dataset (%)			
	DSC	mIoU	Precision	Recall	DSC	mIoU	Precision	Recall
UNet [9]	76.18	67.62	79.09	79.71	91.18	84.89	93.02	90.88
ResUNet [17]	77.27	68.45	79.20	80.36	91.26	85.09	93.18	91.15
AttUNet [7]	76.62	68.09	79.72	78.44	91.04	84.65	93.04	90.66
TransUNet [2]	71.94	61.88	78.57	73.57	89.97	82.69	90.09	91.53
UNeXt [13]	72.27	61.64	77.06	75.39	89.97	82.41	90.69	90.57
CMUNet [11]	79.95	71.68	84.13	81.45	91.43	85.28	92.86	91.58
CMUNeXt [10]	74.52	65.32	77.53	76.46	90.43	83.41	90.89	91.29
Our Model	**81.17**	**72.92**	**82.41**	**82.72**	**92.11**	**86.26**	**93.73**	**91.80**

Table 1: Comparision results on BUSI and BUSIS datasets.

Methods	TN3K dataset (%)			
	DSC	mIoU	Precision	Recall
UNet [9]	77.69	67.38	74.91	87.08
ResUNet [17]	76.76	66.67	73.93	86.18
AttUNet [7]	77.80	67.86	74.94	87.04
TransUNet [2]	71.65	60.03	69.37	83.13
UNeXt [13]	74.35	63.15	72.07	85.19
CMUNet [11]	79.86	70.15	78.35	85.19
CMUNeXt [10]	75.83	65.73	73.39	85.34
Our Model	**81.16**	**71.96**	**80.28**	**87.49**

Table 2: Comparision results on TN3K datasets.

The ablation study results, detailed in Table 3 and 4, provide insightful analysis into the efficacy of our proposed Feature Filter Module (FFM) and Feature Refinement Module (FRM) on medical image segmentation tasks. The experimental outcomes reveal compelling improvements in performance metrics, emphasizing the significance of our proposed modules. The comprehensive ablation study affirms the efficacy and indispensability of our proposed FFM and FRM modules in the context of medical image segmentation tasks.

Methods	TN3K dataset (%)			
	DSC	mIoU	Precision	Recall
baseline	77.69	67.38	74.91	87.08
Ours w/o FRM	78.03	68.27	74.64	**87.61**
Our Model	**81.16**	**71.96**	**80.28**	**87.49**

Table 3: Ablation study results on TN3K datasets.

Methods	BUSI dataset (%)				BUSIS dataset (%)			
	DSC	mIoU	Precision	Recall	DSC	mIoU	Precision	Recall
baseline	76.18	67.62	79.09	79.71	91.18	84.89	93.02	90.88
Ours w/o FRM	79.08	69.86	81.30	81.96	91.46	85.15	93.25	90.98
Our Model	**81.17**	**72.92**	**82.41**	**82.72**	**92.11**	**86.26**	**93.73**	**91.80**

Table 4: Ablation study results on BUSI and BUSIS datasets.

Conclusion

In this paper, we unveil the FFR-UNet, an innovative framework meticulously crafted to not only elevate the precision of segmentation but also to augment the refinement of features in the domain of medical image segmentation. At the heart of our methodology lies the introduction of a groundbreaking FFM, strategically deployed to process features extracted from the preceding layer. Building upon the FFM, our contribution extends to the introduction of the FRM. The FRM, a carefully devised amalgamation of convolution operations and cross-attention mechanisms, emerges as a powerful tool for refining features. The culmination of the integration of the FFM and FRM into the encoder layers of the UNet architecture gives rise to the specialized FFR-UNet tailored explicitly for medical image segmentation.