

# 利用多尺度减法网络实现自动息肉分割

梁宇

## 摘要

百分之九十以上的结直肠癌是由结直肠息肉逐渐转变而来的，因此若能早期发现结直肠息肉对预防结直肠癌有重要意义。在临床实践中，精确的息肉分割为结肠直肠癌的早期发现提供了重要的信息。因此，息肉自动分割技术对患者和医生都具有重要意义。现有的方法大多基于 u 型结构，利用元素级的加法或连接方式，在解码器中逐步融合不同层次的特征。然而，这两种操作都很容易产生大量的冗余信息，从而削弱了不同层次特征之间的互补性，导致息肉的定位不准确，边缘模糊。为了解决这一挑战，我们提出了一种多尺度减法网络（MSNet）从结肠镜图像中分割息肉。具体来说，我们首先设计了一个减法单元（SU）来产生编码器中相邻级别之间的不同特征。然后，我们在不同层次的多个减法单元中配备不同的感受域，从而获得丰富的多尺度差异信息。此外，我们建立了一个无训练的网络“LossNet”来全面监督从底层到顶层的息肉感知特征，从而驱动 MSNet 同时捕捉细节和结构线索。在五个基准数据集上进行的大量实验表明，我们的 MSNet 在不同的评估指标下优于大多数最先进的方法。此外，当处理一个 352×352 图像时，MSNet 以  $\square 70\text{fps}$  的实时速度运行。

**关键词：**结肠直肠癌；自动息肉分割；减法；LossNet

## 1 引言

请在此部分对选题背景，选题依据以及选题意义进行描述。根据 GLOBOCAN2020 年的数据，结肠直肠癌是全球第三大最常见的癌症，也是第二大最常见的死亡原因。它通常开始于小的、非癌性的（良性的）被称为息肉的细胞团块，形成于结肠内部。随着时间的推移，有些息肉会变成结肠癌症。因此，预防结肠癌的最好方法是在息肉变成癌症之前识别并切除它们。结肠镜检查是目前最常用的检查方法，但这一过程涉及到人工劳动和昂贵的劳动，更不用说其误诊率高了。因此，一款自动且精准的息肉分割技术具有重要的现实意义。

## 2 相关工作

息肉的自动分割是从传统的方法逐渐发展而来的，<sup>[1]</sup>基于手工设计的特征进行深度学习的方法<sup>[2]</sup>。虽然这些方法在临床上取得了进展，但它们受到 box-level 预测结果的限制，因此无法捕捉到息肉的形状和轮廓。为了解决这个问题<sup>[3]</sup>等人利用 FCN 通过像素级的预测来分割息肉。Akbari<sup>[4]</sup>等人也利用基于 FCN 的分割网络，结合斑块选择机制，提高了息肉分割的精度。然而，基于 FCN 的方法依靠低分辨率的特征来生成最终的预测，导致粗糙的分割结果和模糊的边界。近年来，U 型的网络结构<sup>[5]</sup>由于其利用多层信息重建高分辨率特征图的能力而受到了相当多的关注。许多息肉分割网络都采用 U 型架构。在 UNet<sup>[6]</sup>中，上采样特征映射与编码器跳过的特征映射连接，并且在上采样步骤之间添加卷积层和非线性层。UNet++<sup>[7]</sup>使用嵌套的和密集的跳过连接来减少编码器和解码器的特征映射之间的语义差距。后来 ResUNet++<sup>[8]</sup>结合了许多先进的技术，如残差计算，SEblock，空洞空间金字塔池，以及注意机制进一步提高性能。最近的研究中，SFA<sup>[9]</sup>和 PraNet<sup>[10]</sup>，重点是恢复息肉与其周围粘膜之间的

尖锐边界。前者提出了一种在共享编码器和两个相互约束的解码器下的选择性特征聚合结构和边界敏感损失函数。后者利用反向注意模块来建立区域和边界线索之间的关系。一般来说，编码器中不同级别的特征具有不同的特征。高级的有更多的语义信息，这有助于定位对象，而低级的有更详细的信息，可以捕捉对象的微妙边界。解码器利用特定级别和跨级别的特性来生成最终的高分辨率预测。然而，上述方法直接使用元素级的添加或连接来融合从编码器中的任何两个级别特征，并将它们传输到解码器。这些简单的操作并不更注意不同级别之间的差异信息。这一缺点不仅产生了冗余的信息，稀释了真正有用的特征，而且削弱了特定层次特征的特征，导致网络无法平衡精确的息肉定位和微妙的边界细化。在本文中，我们提出了一种新的多尺度减法网络（MSNet）的息肉分割任务。我们首先设计了一个减法单元（SU）并将其应用于每一对相邻的级别特性。为了解决息肉的规模多样性，我们金字塔地连接多个 SUs 以捕获大跨度跨级别信息。然后，我们聚合水平特定的特征和多路径跨水平差分特征，然后在解码器中生成最终的预测。此外，我们提出了一个 LossNet 来自动监督从底层到顶层提取的特征映射，它可以通过一个简单的  $\ell_2$ -损失函数来优化从细节到结构的分割。我们的主要贡献可以总结如下：1、本文提出了一种新的多尺度息肉自动分割减法网络。通过多级、多级级联减法操作，可以有效地获得不同层次之间从低阶到高阶的互补信息，从而全面增强了对息肉区域的感知。2、我们建立了一个通用的无训练损失网络，在特征层面实现细节-结构监督，这为基于预测本身的损失设计提供了重要的补充。3、所提出的 MSNet 可以准确地分割息肉，如图 1 所示。大量的实验表明，我们的 MSNet 在 5 个具有挑战性的数据集上以不同的评估指标下，大大推进了最先进的方法，实时推理速度约为 70fps。

### 3 本文方法

#### 3.1 本文方法概述

MSNet 架构如图 1 所示，其中有 5 个编码器块 ( $\mathbf{E}^i, i \in \{1, 2, 3, 4, 5\}$ )，一个多尺度减法模块和 4 个解码器块 ( $\mathbf{D}^i, i \in \{1, 2, 3, 4\}$ )，遵循 PraNet<sup>[10]</sup>，本文采用 Res2Net-50 作为骨干，提取五个层次的特征。首先，对每个编码器块的特征映射分别采用  $3 \times 3$  卷积，将通道数量减少到 64，这可以减少后续操作的参数数量。然后，将这些不同层次的特征输入多尺度减法模块，输出 5 个互补增强特征 ( $CE^i, i \in \{1, 2, 3, 4, 5\}$ )，最后，每个  $CE^i$  逐步参与解码器并生成最后的预测。通过对原文复现后发现，在训练阶段，预测和真实标签都被输入到 LossNet 中，以实现监督。

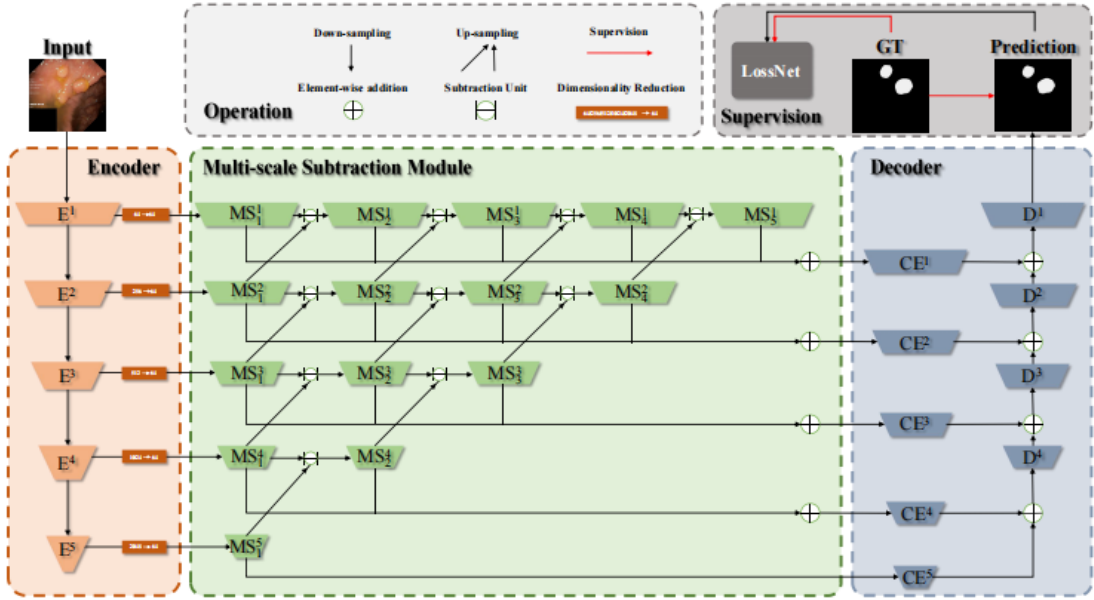


图 1: 原文方法示意图

### 3.2 多尺度减法模块

我们使用  $F_A$  和  $F_B$  来表示相邻层的特征图。它们都被  $\text{ReLU}$  操作激活。我们定义了一个基本的减法单位 (SU):

$$SU = \text{Conv}(|F_A \ominus F_B|),$$

其中  $\ominus$  为元素级减法运算,  $|\cdot|$  计算绝对值,  $\text{Conv}$  表示卷积层。SU 单元可以捕获  $F_A$  和  $F_B$  的互补信息, 并突出它们之间的差异, 从而为解码器提供更丰富的信息。为了获得多个特征层次的高阶互补信息, 我们水平和垂直连接多个 SUs, 计算一系列具有不同顺序和接受域的差异特征。多尺度减法模块的详细情况见图 1。我们在相应的级别和任何其他级别之间聚合了特定于级别的特性 ( $MS_1^i$ ) 和跨层次的差异特征 ( $MS_{n \neq 1}^i$ ) 从而生成互补性增强的特征 ( $CE^i$ )。此过程可表述如下:

$$CE^i = \text{Conv} \left( \sum_{n=1}^{6-i} MS_n^i \right) \quad i = 1, 2, 3, 4, 5$$

最后, 所有的 ( $CE^i$ ) 都参与解码, 然后对息肉区域进行分割。

### 3.3 LossNet

在所提出的模型中, 总的训练损失可以写为:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{IoU}}^w + \mathcal{L}_{\text{BCE}}^w + \mathcal{L}_f$$

其中,  $\mathcal{L}_{\text{IoU}}^w$  和  $\mathcal{L}_{\text{BCE}}^w$  表示加权 IoU 损失和二值交叉熵损失, 这在分割任务中被广泛采用。我们使用了与<sup>[10]</sup>中相同的定义, 其有效性已经在这些工作中得到了验证。与它们不同的是, 我们额外使用一个 LossNet 来进一步优化从细节到结构的分割。LossNet 的结构如图 2 所示。可以看出, 低级特征图包含了丰富的边界信息, 而高级特征图描述了位置信息。因此, LossNet 可以在特性级别上产生全面的监督。

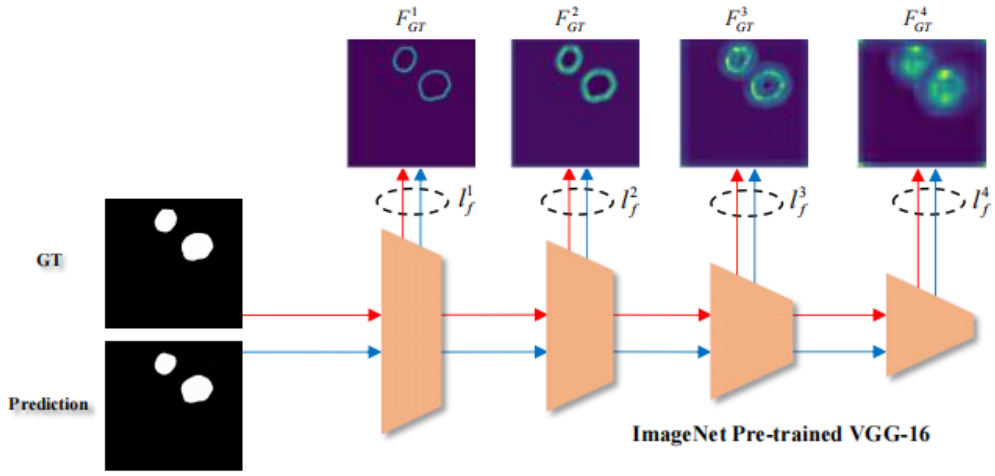


图 2: LossNet 示意图

## 4 复现细节

### 4.1 与已有开源代码对比

首先通过对作者在原文中的描述，尽量将原文网络架构复现，及利用 `dataset` 类封装输入的数据，对该方法进行实验。从复现结果中发现仍然存在在对目标区域分割的边界上仍存在模糊、不精确的问题，以此尝试做出一些改进。其一，我尝试在解码器中加入边界感知的模块，通过上采样和 `softmax` 层处理生成该尺度上的边界概率映射与输入相乘后再与原文得到的特征图拼接作为卷积层的输入，从而尝试解决息肉分割边界模糊的问题，进而优化分割性能，如图 3 所示。其二，实验过程中发现当目标区域较大时被分割的区域却较小的情况，通过加入注意力机制的方法，并引入了高斯调制函数<sup>[11]</sup>，通过突出次要激活，并一定程度上抑制最敏感和最不敏感的特征，试图引导模型扩大对目标区域的激活区域，从而解决目标区域分割不全的问题，如图 4 所示。

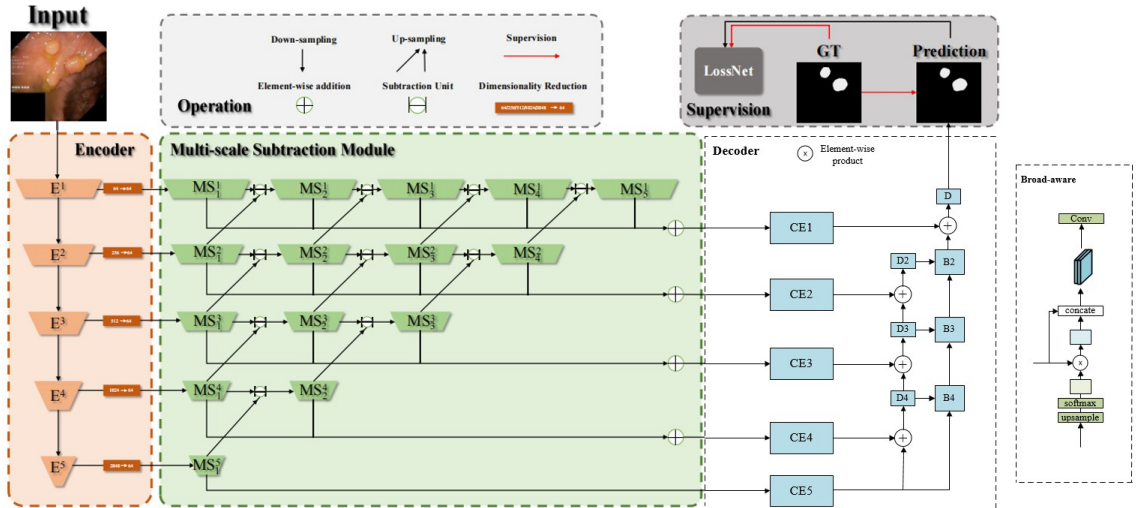


图 3: 改进后方法示意图

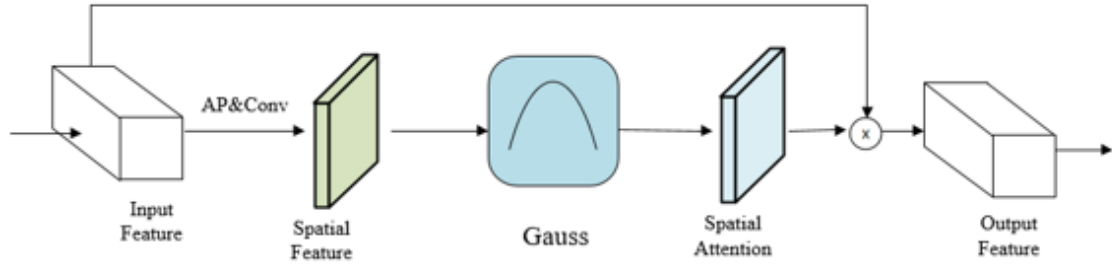


图 4: 高斯调制注意力示意图

## 4.2 实验设计

本次实验在五个基准数据集上评估了所提出的模型与原文模型对比，分别为: CVC-ColonDB [22], ETIS [21], Kvasir [11], CVC-T [24] 和 CVC-ClinicDB [2]。评价的指标采用的是广泛使用的平均 Dice, 平均 IoU 和 MAE。本次提出的模型是基于 PyTorch 框架实现的，并在一个 TITANXP GPU 上训练了 50 个轮次，小批量大小为 16。我们将输入的图片大小调整为  $352 \times 352$ ，采用随机水平翻转和随机旋转数据增强两种方法，以避免过拟合，对于优化器，我们采用随机梯度下降（SGD）。

## 5 实验结果分析

图 5 是通过复现后的结果与原文中的结果的对比，其中上三行为原文结果，下三行为复现后得到的结果，本次实验中将复现后的结果作为 baseline。图 6 是加入了改进一后与 baseline 的对比，可以发现除了在 Kvasir 数据集上效果和 baseline 结果接近外，其余的数据集都取得了比 baseline 要好的结果，改进后的分割结果如图 7 所示，其中第一行为真实标签，第二行为原文结果，最后一行为所提出改进后的结果，可以发现模型对边缘分割更加准确。

	ColonDB	ETIS	Kvasir	CVC-T	ClinicDB
mDice	0.755	0.719	0.907	0.869	0.921
mIoU	0.678	0.664	0.862	0.807	0.879
MAE	0.041	0.020	0.028	0.010	0.008
Baseline					
mDice	0.741	0.681	0.899	0.864	0.923
mIoU	0.665	0.639	0.852	0.802	0.880
MAE	0.041	0.0341	0.031	0.01	0.007

图 5: baseline 实验结果

		ColonDB	ETIS	Kvasir	CVC-T	ClinicDB
proposed	mDice	0.741	0.681	0.899	0.864	0.923
	mIoU	0.665	0.639	0.852	0.802	0.880
	MAE	0.041	0.034	0.031	0.010	0.007
	mDice	0.750	0.691	0.900	0.873	0.930
	mIoU	0.678	0.636	0.850	0.810	0.887
	MAE	0.038	0.025	0.033	0.008	0.007

图 6: 改进一实验结果

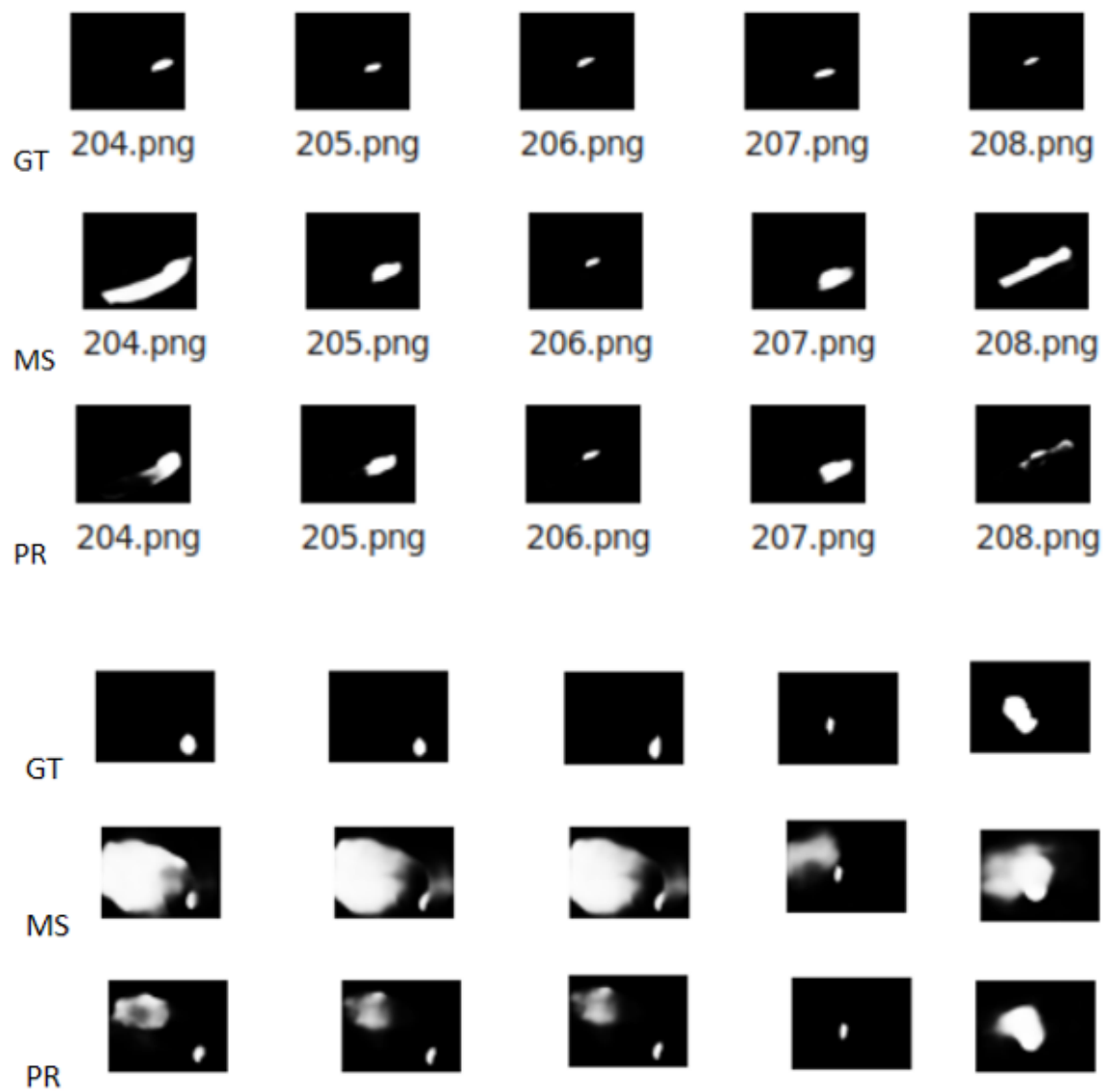


图 7: 改进一与原文方法实验结果对比

图 8 是加入改进二后的结果，效果表现得不是很好，图 9 是改进后的结果与原文的一些对比，其中第一行为真实标签，第二行为原文结果，最后一行为所提出改进后的结果，可以发现在某些实验的图片上可以扩大分割的区域从而更贴近真实的目标区域。猜测效果不好的原因主要是，其一加入的注意力模型没有训练好，而且相较于小目标来说未必合适；其二注意力模块插入的位置也会有一定的影响。

		ColonDB	ETIS	Kvasir	CVC-T	ClinicDB
w attention	mDice	0.741	0.681	0.899	0.864	0.923
	mIoU	0.665	0.639	0.852	0.802	0.880
	MAE	0.041	0.034	0.031	0.010	0.007
	mDice	0.735	0.676	0.880	0.894	0.900
	mIoU	0.664	0.614	0.826	0.830	0.850
	MAE	0.042	0.017	0.037	0.007	0.012

图 8: 改进二实验结果

er atte



图 9: 改进二与原文方法实验结果对比

## 6 总结与展望

本次实验中虽然通过加入边缘感知模块和注意力模块，使得模型的分割性能有所提升，但存在许多不足，加入边缘感知模块针对的小目标数据仍然无法较好的改善，同时可能受数据不平衡的影响较大，加入注意力模块后需要重新训练从而调整注意力模块的参数，需要消耗更多的时间且增加了参数量。往后还是通过从网络架构入手，进一步改善分割性能。

## 参考文献

- [1] TAJBAKHSH N, GURUDU S R, LIANG J. Automatic polyp detection in colonoscopy videos using an ensemble of convolutional neural networks[J]., 2015: 79-83.
- [2] BRANDAO P, MAZOMENOS E, CIUTI G, et al. Fully convolutional neural networks for polyp segmentation in colonoscopy[C]//Medical Imaging 2017: Computer-Aided Diagnosis: vol. 10134. 2017: 101-107.



- [3] LONG J, SELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 3431-3440.
- [4] AKBARI M, MOHREKESH M, NASR-ESFAHANI E, et al. Polyp segmentation in colonoscopy images using fully convolutional network[C] // 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2018: 69-72.
- [5] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2117-2125.
- [6] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C] // Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. 2015: 234-241.
- [7] ZHOU Z, RAHMAN SIDDIQUEE M M, TAJBAKHSN N, et al. Unet++: A nested u-net architecture for medical image segmentation[C] // Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4. 2018: 3-11.
- [8] JHA D, SMEDSRUD P H, RIEGLER M A, et al. Resunet++: An advanced architecture for medical image segmentation[C] // 2019 IEEE International Symposium on Multimedia (ISM). 2019: 225-2255.
- [9] FANG Y, CHEN C, YUAN Y, et al. Selective feature aggregation network with area-boundary constraints for polyp segmentation[C] // Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part I 22. 2019: 302-310.
- [10] FAN D P, JI G P, ZHOU T, et al. Pranet: Parallel reverse attention network for polyp segmentation[C] // Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part VI 23. 2020: 263-273.
- [11] QIN J, WU J, XIAO X, et al. Activation modulation and recalibration scheme for weakly supervised semantic segmentation[C] // Proceedings of the AAAI Conference on Artificial Intelligence: vol. 36: 2. 2022: 2117-2125.