

基于多尺度差分网络的息肉自动分割

Xiaoqi Zhao, Lihe Zhang, and Huchuan Lu¹

摘要

摘要：超过百分之九十以上的结直肠癌是由结直肠息肉逐渐转化而来的。在临床实践中，精确的息肉分割为大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。现有的方法大多是基于 U 形结构，在编码器中采用元素加法或级联的方法逐步融合不同层次的特征。但是，这两种操作都容易产生大量的冗余信息，削弱不同层次特征之间的互补性，导致息肉定位不准确，边缘模糊。为了解决这一难题，文章作者提出了一种多尺度差分网络 (MSNET) 来从结肠镜图像中分割息肉。具体来说，文章作者首先设计了一个差分单元 (SU) 来产生编码器中相邻电平之间的差异特征。然后，将不同层次的差分单元按金字塔形式配置不同的感受野，从而获得丰富的多尺度差异信息。此外，作者还构建了一个无需训练的网络 “Lossnet”，从底层到顶层全面监控息肉感知特征，驱动 MSNET 同步捕获细节线索和结构线索。在五个基准数据集上进行的大量实验表明，在不同的评估指标下，文章提出的 MSNET 相对于大多数最先进的方法都有良好的性能。但是基于上述的方法，容易对部分小目标以及边缘连接在一起的目标产生不好的分割效果，为了解决该问题，在复现文章的时候添加了边界主要共享模块以解决上述问题，同时在作者提供的 5 个数据集中，三个数据集的精度均有提升。

关键词：结直肠癌；自动息肉分割；差分

1 引言

根据 Globocan 2020 数据，结直肠癌是全球第三大常见癌症，也是第二大常见死亡原因。它通常开始于结肠内部形成的小的、非癌变的（良性的）细胞团，称为息肉。随着时间的推移，这些息肉中的一些会变成结肠癌。因此，预防结肠癌的最好方法是在息肉转化为癌之前识别并切除息肉。目前，结肠镜检查是最常用的检查手段，但这一过程涉及人工和昂贵的劳动，更别提其高误诊率。因此，自动准确地分割息肉具有重要的现实意义。选择本文做为前沿技术的复现文章的主要依据是该文章是目前对息肉自动分割最好的模型之一，其精度以及对前人算法的缺陷的补充更加完善，同时提出了差分单元进行息肉边缘的提取。成功且顺利的复现该文章的结果以及改进相关模块单元能够为我今后在医学图像分割的工作上积累更多的经验，由于医学图像分割大多数问题存在与边缘分割不清楚，界限不明确的部分，因此复现该文章对我的后续工作将会有极大的作用。

2 相关工作

息肉分割的主要方法主要分为两种，一种是基于传统人工设计特征的方法^[1]；另一种是基于深度学习的方法。基于深度学习的方法主要分为两类，一类是采用 FCN 类型网络进行息肉的分割；另一类是采用 U 型网络结构的进行息肉的分割。两种方法的区别在于前一种基于 FCN 的方法依赖于低分辨率特征生成最终预测，导致分割结果粗糙，边界模糊。而 U 型网络能够利用多层次信息重建高分辨率特征图，能够增强多边界的精准分割。

2.1 基于深度学习的方法

2.1.1 基于 FCN 的方法

Brandao 等人^[2]利用 FCN^[3]通过像素级预测来分割息肉。阿克巴里等人^[4]还采用基于 FCNN 的分割网络，结合斑块选择机制，提高了息肉分割的准确性。然而，基于 FCN 的方法依赖于低分辨率特征生成最终预测，导致分割结果粗糙，边界模糊。

2.1.2 基于 U 型网络的方法

在 UNET^[5]中，向上采样的特征映射与从编码器跳过的特征映射级联，并且在向上采样步骤之间添加卷积和非线性。UNET++^[6]使用嵌套和密集的跳过连接来减少编码器和解码器特征映射之间的语义鸿沟。最近的工作，SFA^[7]和 Pranet^[8]侧重于恢复息肉与其周围粘膜之间的尖锐边界。前者在一个共享编码器和两个相互约束的译码器下提出了一种选择性特征聚集结构和边界敏感损失函数。后者利用反向注意模块建立区域线索与边界线索之间的关系

3 本文方法

本文提出了一种新的多尺度差分网络 (MSNet) 用于息肉分割。文章首先设计了一个差分单元 (SU)，并将其应用于每一对相邻的层次特征。同时为了解决息肉边缘多样性问题，文章将多个 SUS 进行金字塔级联，捕捉大跨度的跨层次信息。然后，在解码器聚合特定的特征和多尺度特征的差分特征，生成最终的预测结果。同时，文章害提出了一个 Lossnet 网络来自动监控从网络从底层到顶层的特征映射，它可以通过简单的 L2 损失函数来优化从息肉分割细节到结构的分割。

3.1 本文方法概述

本文主要实现了 MSNet 网络，并对 MSNet 网络进行了改进，增加了共享边界注意解码器模块。改进前网络结构图，如图 1 所示。改进 MSNet 网络结构图，如图 2 所示此部分对本文将要复现的工作进行概述，图的插入如图 2所示：

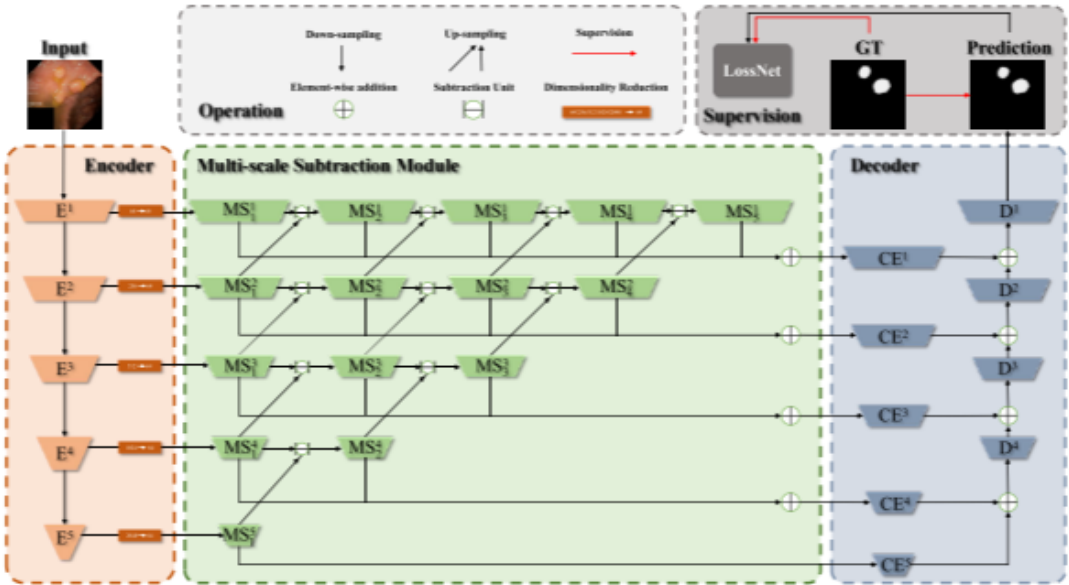


图 1: MSNet 改进前网络结构图

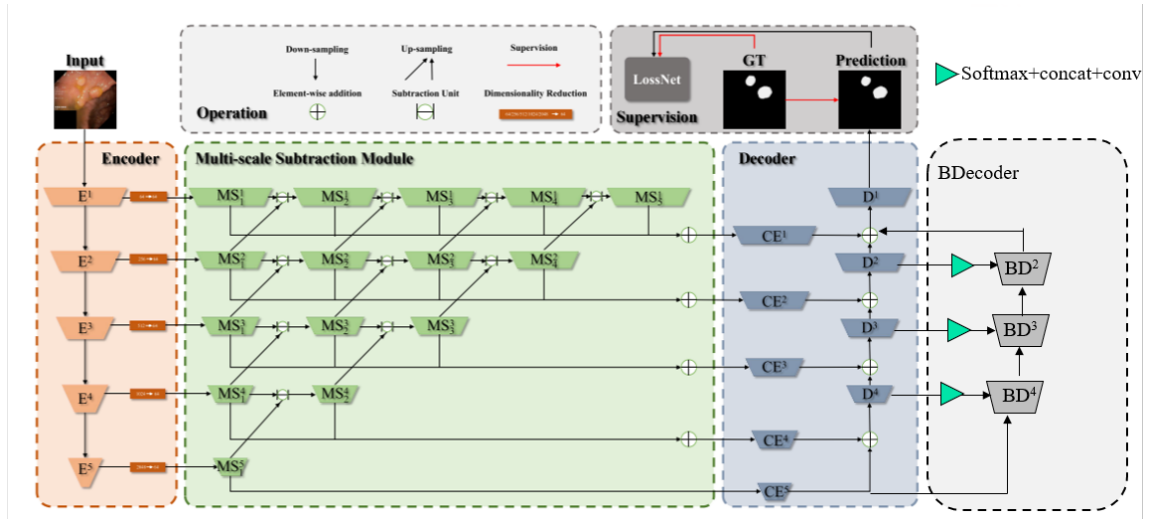


图 2: MSNet 改进后网络结构图

3.2 多尺度差分模块

作者使用 F_a 和 F_b 来表示相邻层的特征映射，它们都经过 Relu 函数并被激活。同时将基本减法单元 (SU) 定义为：

$$SU = \text{Conv}(|F_a \ominus F_b|) \quad (1)$$

其中 \ominus 是元素差分运算， $|\cdot|$ 计算绝对值， $\text{conv}(\cdot)$ 表示卷积层。SU 单元能够捕捉 F_a 和 F_b 的互补信息，突出它们的差异，从而为解码器提供更丰富的信息。同时为了获得跨多个特征层的高阶互补信息，我们通过将多个 SUS 水平和垂直串联起来，计算出一系列不同阶次和感受野的差分特征。我们将对应层与其他层之间的层特定特征 ($MS_{n=1}^i$) 和跨层差分特征 ($MS_{n \neq 1}^i$) 聚合起来，生成互补增强特征 (CE^i)。整体过程如下所示：

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

由于本文的网络在对于部分数据集的边缘提取的时候仍然存在不能够较好的将边缘分割出来的效果，因此将边界注意共享解码器模块加入本文的网络结构形成了新的网络结构，主要工作原理是在同一层的解码器模块增加将特征提取模块经过 softmax 函数生成边缘概率图后和同一层的特征进行 concat 操作后输入卷积神经网络进行特征提取，最后将所得到的边缘特征加入同一层的特征模块进行下一步操作，具体过程见图 2。整体过程可以定义为如下函数：

$$BD_i = \text{Conv}(\text{Concat}(\text{Softmax}(D_i) * D_i, D_i), i = 2, 3, 4 \quad (3)$$

其中 Conv 代表卷积操作， Concat 代表将两个矩阵合并， Softmax 代表 softmax 函数，将其与本层特征相乘得到边缘的概率图，最后和该层特征进行 concat 操作后输入卷积神经网络得到最后可以利用的特征。

3.3 损失函数定义

在所提出的 MSNet 模型中，损失函数可以定义为

$$L_{\text{total}} = L_{\text{IoU}}^w + L_{\text{BCE}}^w + L_f \quad (4)$$

其中 L_{IoU}^w 和 L_{BCE}^w 分别表示加权 IOU 损失和二进制交叉熵 (BCE) 损失，它们在分割任务中被广泛采用，因此我们将才采用该损失函数到我们的网络中。同时我们增加了额的 Lossnet 来进一步优化从细

节到结构的分割。具体来说，使用一个预先训练的 ImageNet 分类网络，如 VGG-16，分别提取预测和真实标签的多尺度特征。然后计算它们的特征差值作为损失，在复现文章中我们加入了不同的权重，公式如下：

$$\mathbf{L}_f = 0.2 * \mathbf{l}_f^1 + 0.2 * \mathbf{l}_f^2 + 0.3 * \mathbf{l}_f^3 + 0.3 * \mathbf{l}_f^4 \quad (5)$$

设 \mathbf{F}_P^i 和 \mathbf{F}_G^i 分别表示从预测和真是标签中提取的第 i 层特征图。计算 \mathbf{l}_f^i 作为它们的欧几里得距离 (L2-损失)，该距离在像素级进行监督，具体公式如下：

$$\mathbf{l}_f^i = \|F_P^i - F_G^i\|_2, i = 1, 2, 3, 4 \quad (6)$$

4 复现细节

4.1 与已有开源代码对比

本次复现工作主要对文章提供的源代码进行了改进工作以及代码调通工作。主要创新改进地方在于不同于作者仅仅使用单一的解码器，增加了边界注意力共享解码器，主要目的是为了解决作者在文章中对部分息肉的边界提取仍然不明显的问题。同时针对 lossnet 中的损失调整了权重比例，并非作者在文章中使用的 1: 1: 1: 1，在多次实验结果中证明使用 2: 2: 3: 3 的权重比能够达到较好的效果，在作者提供的 5 个测试集上均能够达到作者的精度，在其中三个测试集中精度均有提升。

4.2 实验环境搭建

本次前言技术代码复现实验环境基于 Pytorch 1.9 是基于 PyTorch 框架实现的，并在一个 TITAN X 的 GPU 上训练了 50 个周期，批量大小为 32。我们将输入调整为 352×352，Pranet^[8]。采用随机水平翻转和随机旋转数据增强来避免过拟合。对于优化器，我们采用随机梯度下降 (SGD)。动量衰变和重量衰变分别设为 0.9 和 0.0005。主干的最大学习速率为 0.005，其他部分的最大学习速率为 0.05。采用预热和线性衰减策略调整学习速率。

4.3 创新点

创新点：增加边界注意共享解码器，使网络能够增加额外的对与边界信息的利用，达到增强边界分割的能力。调整 Lossnet 中损失的权重比例。

5 实验结果分析

本次实验的复现结果在改进原有网络结构后，在原文的 5 个测试集中均能达到文章的效果，同时在 3 个测试集上均有精度的提升，我们分别在 M-dice, MIou 以及 MAE 等精度上评估了改进前后方法的对比，具体如下表 1。原文中的网络结构对于部分的息肉图不能够达到很好的分割效果，主要原因是针对部分多个区域的区域区分处不能够很好的分割处边界以及对于小目标区域不能够精准的分割，容易造成假阳性，具体如下图 3 所示。

表 1: 实验结果表

Dataset	改进前			改进后		
	Dice	MIou	MAE	Dice	MIou	MAE
CVC-T ^[9]	0.869	0.807	0.010	0.8956	0.8243	0.0065
ETIS ^[10]	0.719	0.664	0.020	0.7438	0.6721	0.029
Kvasir ^[11]	0.907	0.862	0.0278	0.9158	0.8650	0.0263

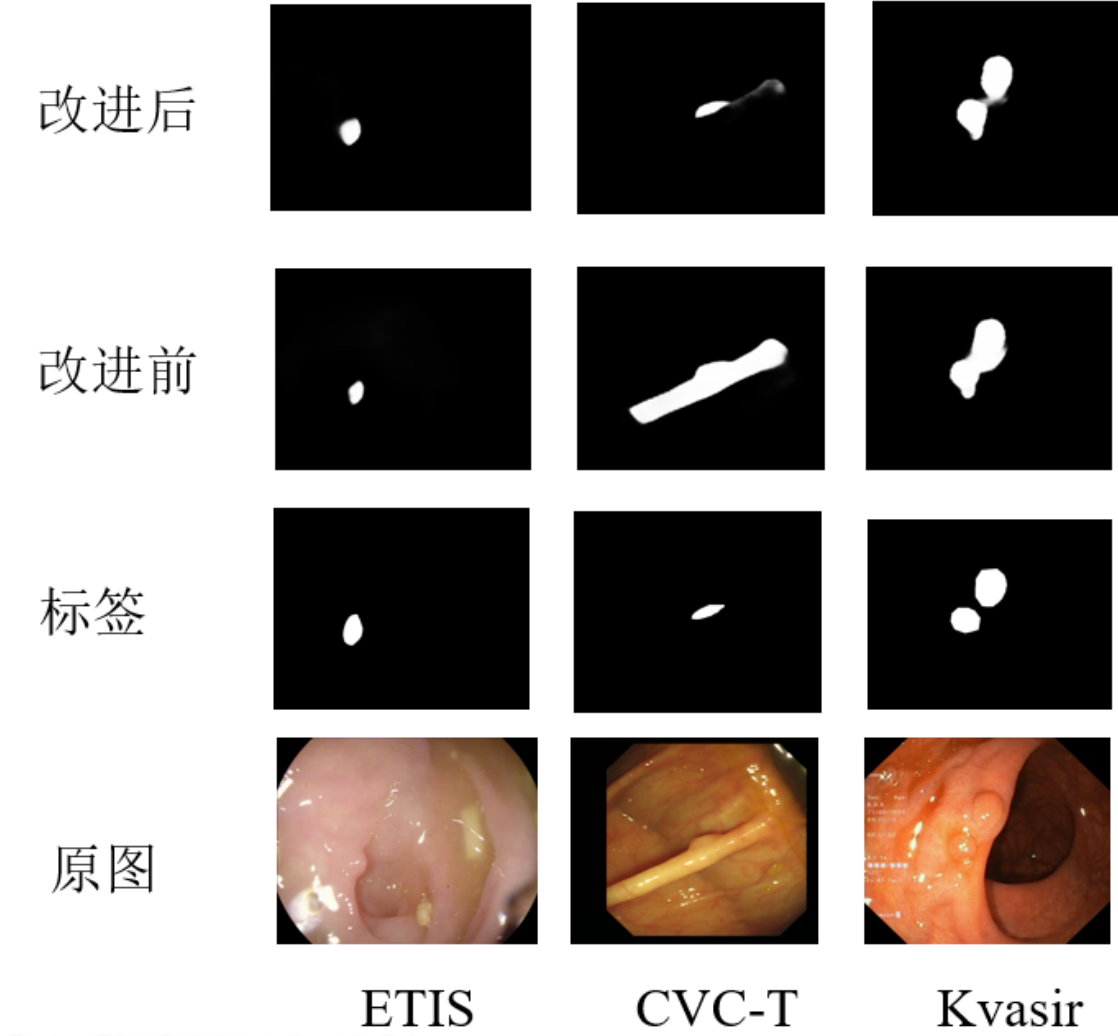


图 3: 实验结果示意

6 总结与展望

本次前沿技术的复现，对于原文的 MSNet 网络结构进行了深度的理解并对于其容易出现的错误分割的效果进行了阐述同时针对该情况进行了进一步的改进，改进效果体现在 Dice 的精度提升上，针对 $CVC-T$ 、 $Kvasir$ 、 $ETIS$ 三个测试数据集分别提升了 0.021、0.06 以及 0.028，同时在图 3 中也能够看出原文网络针对两个目标的连接处和细小目标的分割效果不佳，改进网络能够有效的改善该种情况。但是本次改进的网络仍然存在假阳性的情况，如果能够加入多尺度的注意力机制也许能够改善该种情况，将在以后继续改进网络结构以期待能够进一步改善该情况。

参考文献

- [1] TAJBAKHS N, GURUDU S R, LIANG J. Automatic polyp detection in colonoscopy videos using an ensemble of convolutional neural networks[J]. 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), 2015: 79-83.
- [2] BRANDAO P, MAZOMENOS E, CIUTI G, et al. Fully convolutional neural networks for polyp segmentation in colonoscopy[J]. Medical Imaging 2017: Computer-Aided Diagnosis, 2017, 10134: 101340F.
- [3] LONG J, SHELHAMER E, DARRELL T. Fully Convolutional Networks for Semantic Segmentation[J]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.
- [4] AKBARI M, MOHREKESH M, NASR-ESFAHANI E, et al. Polyp Segmentation in Colonoscopy Images Using Fully Convolutional Network[J]. 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2018: 69-72.
- [5] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional Networks for Biomedical Image Segmentation[J]. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, 2015: 234-241.
- [6] ZHOU Z, RAHMAN SIDDIQUEE M M, TAJBAKHS N, et al. Unet++: A nested u-net architecture for medical image segmentation[J]. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, 2018: 3-11.
- [7] FANG Y, CHEN C, YUAN Y, et al. Selective feature aggregation network with area-boundary constraints for polyp segmentation[J]. Medical Image Computing and Computer Assisted Intervention – MICCAI 2019, 2019: 302-310.
- [8] FAN D P, JI G P, ZHOU T, et al. Pranet: Parallel reverse attention network for polyp segmentation[J]. Medical Image Computing and Computer Assisted Intervention – MICCAI 2020, 2020: 263-273.
- [9] VÁZQUEZ D, BERNAL J, SÁNCHEZ F J, et al. A Benchmark for Endoluminal Scene Segmentation of Colonoscopy Images[J]. Medical Image Computing and Computer Assisted Intervention – MICCAI 2020, 2017: 2040-2295.
- [10] SILVA J, HISTACE A, ROMAIN O, et al. Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer[J]. International Journal of Computer Assisted Radiology and Surgery, 2014: 283-293.
- [11] JHA D, SMEDSRUD P H, RIEGLER M A, et al. Kvasir-SEG: A Segmented Polyp Dataset[J]. MultiMedia Modeling, 2020: 451-462.