

改进U-Net的注意力特征增强息肉分割网络

王志敏¹, 冯跃¹, 秦浩^{2,3}, 田霞³, 梁洁欣¹, 孙千帅¹, 曾军英¹

¹五邑大学电子与信息工程学院, 广东 江门

²武汉科技大学医学部医学院, 湖北 武汉

³武汉市第三医院(武汉大学附属同仁医院)消化内科, 湖北 武汉

收稿日期: 2024年12月13日; 录用日期: 2025年1月6日; 发布日期: 2025年1月15日

摘要

在肠道息肉分割任务中, 由于息肉病变区域与周围正常组织颜色对比度不高, 且边界模糊, 这导致分割过程中容易丢失关键信息, 并受到噪声干扰, 影响分割质量。为了解决这些问题, 提出一种改进的U-Net模型EEPSNet。首先, EEPSNet通过将普通卷积与空洞卷积相结合, 替换了原有的双卷积块。这种组合能捕捉更广泛区域的空间特征, 从而减少信息丢失。其次, 在解码阶段的特征融合过程中, 在U-Net的基础上引入了全局注意力机制特征融合, 这种融合方式不仅能够在空间和通道维度上关注显著特征, 还能增强模型对噪声干扰的抑制能力。EEPSNet在单卡NVIDIA Quadro RTX 5000 GPU上对四个公开的息肉分割数据集进行了实验, 包括Kvasir-SEG和CVC-ClinicDB, 用于评估模型的特征建模能力, 以及CVC-ColonDB和ETIS-LaribPolypDB, 用于评估模型的泛化能力。实验结果表明, EEPSNet模型在这些数据集上均取得了显著的性能提升, mdice分别提高至88.6%、91.1%、70.8%和66.0%, 同时也证明了EEPSNet在保持建模能力的同时, 也具有较好的泛化能力。

关键词

肠道息肉, 息肉图像分割, 深度学习, U型网络, 注意力机制, 医学图像分割, 注意力特征融合

Improved Attention Features of U-Net Enhance the Polyp Segmentation Network

Zhimin Wang¹, Yue Feng¹, Hao Qin^{2,3}, Xia Tian³, Jiexin Liang¹, Qianshuai Sun¹, Junying Zeng¹

¹School of Electronics and Information Engineering, Wuyi University, Jiangmen Guangdong

²School of Medicine, Department of Medicine, Wuhan University of Science and Technology, Wuhan Hubei

³Department of Gastroenterology, Wuhan Third Hospital (Tongren Hospital of Wuhan University), Wuhan Hubei

Received: Dec. 13th, 2024; accepted: Jan. 6th, 2025; published: Jan. 15th, 2025

文章引用: 王志敏, 冯跃, 秦浩, 田霞, 梁洁欣, 孙千帅, 曾军英. 改进 U-Net 的注意力特征增强息肉分割网络[J]. 图像与信号处理, 2025, 14(1): 74-84. DOI: 10.12677/jisp.2025.141007

Abstract

In the intestinal polyp segmentation task, due to the low color contrast of the polyp lesion area and the surrounding normal tissue and the blurred boundary, this leads to the loss of key information in the segmentation process and the interference of noise, which affects the segmentation quality. To solve these problems, a modified U-Net model EEPsNet is proposed. First, EEPsNet replaces the original double convolution block by combining ordinary convolution with void convolution. This combination captures the spatial features of a wider region, thereby reducing information loss. Secondly, in the feature fusion process in the decoding stage, global attention mechanism feature fusion is introduced on the basis of U-Net, and this fusion mode can not only focus on salient features in spatial and channel dimensions, but also enhance the ability of the model to suppress noise interference. EEPsNet experiments on four publicly available polyp segmentation datasets, including Kvasir-SEG and CVC-ClinicDB, to evaluate the feature modeling ability of the model, and CVC-ColonDB and ETIS-LaribPolypDB to evaluate the generalization ability of the model. The experimental results show that the EEPsNet model achieved significant performance improvement in these datasets, and the mdice increased to 88.6%, 91.1%, 70.8%, and 66.0%, respectively, which also proved that EEPsNet has good generalization ability while maintaining the modeling ability.

Keywords

Intestinal Polyp, Polyp Image Segmentation, Deep Learning, U-Shaped Network, Attention Mechanism, Medical Image Segmentation, Attention Feature Fusion

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着深度学习技术的快速发展,自从以卷积神经网络(CNN)为基础的全卷积网络(FCN) [1]被提出之后,医学图像分割技术有了更显著的进步。基于 FCN 的 UNet [2]网络已经成为医学图像分割领域最具代表性的骨干之一。UNet 网络及其改进的 U-Net++ [3]、ResUNet [4]、ResUNet++ [5]已经直接引入息肉分割领域并且取得了一些可接受的效果[6]。此外为了进一步提高息肉分割的网络性能,一些只针对于息肉分割任务的网络也相继被提出,并且取得了更好的效果。例如, SFA [7]提出了一种新颖的选择性特征聚合网络,具有用于息肉分割的区域和边界约束。DCRNet [8]使用外部和内部上下文关系模块来分别估计相同和不同图像中每个位置与所有其他位置之间的相似性。SCR-Net [9]通过两个关键模块有效地校准和优化语义信息在训练过程中学习语义-空间关系,将高级层的语义信息传递到低级层,增强特征对目标的区分能力。

尽管上述网络在息肉图像分割领域取得了显著的成就,但仍面临如下挑战:

- 1) 编码阶段的卷积层感受野尺寸不足导致不能感知远距离信息[10]。
- 2) 解码阶段,上采样进行特征融合时由于融合方式过于简单,可能会引入背景噪声,影响特征的鲁棒性,并可能导致语义信息的不匹配[11]。

为了解决上述问题,如图 1 所示,我们提出了 EEPsNet (Enhanced Feature Fusion and Enlarged Receptive Field Polyp Segmentation Net),一种改进的 U-Net 版本,它利用普通卷积和空洞卷积组合捕捉远距离的特

征信息并且通过注意力集中于融合特征来减少浅层特征和深层特征之间的语义不匹配程度。

首先，在编码阶段，EEPSNet 引入了感受野增强卷积块(Enlarged Receptive Field Block)，它结合了普通卷积和空洞卷积，以增强模型对特征的提取能力。这种卷积块能够捕捉到更丰富的纹理和形状信息，为后续的特征融合和分类提供了坚实的基础。

然后，在解码阶段，EEPSNet 采用了全局注意力特征融合模块(Global Attention Feature Fusion Module)。这个模块通过全局注意力机制来加强特征融合的过程，它能够识别并强化那些对于最终任务更为重要的特征，同时抑制那些可能由噪声引起的不相关特征。这种注意力机制的引入，不仅提高了模型对语义信息的匹配能力，还增强了模型对噪声的鲁棒性。

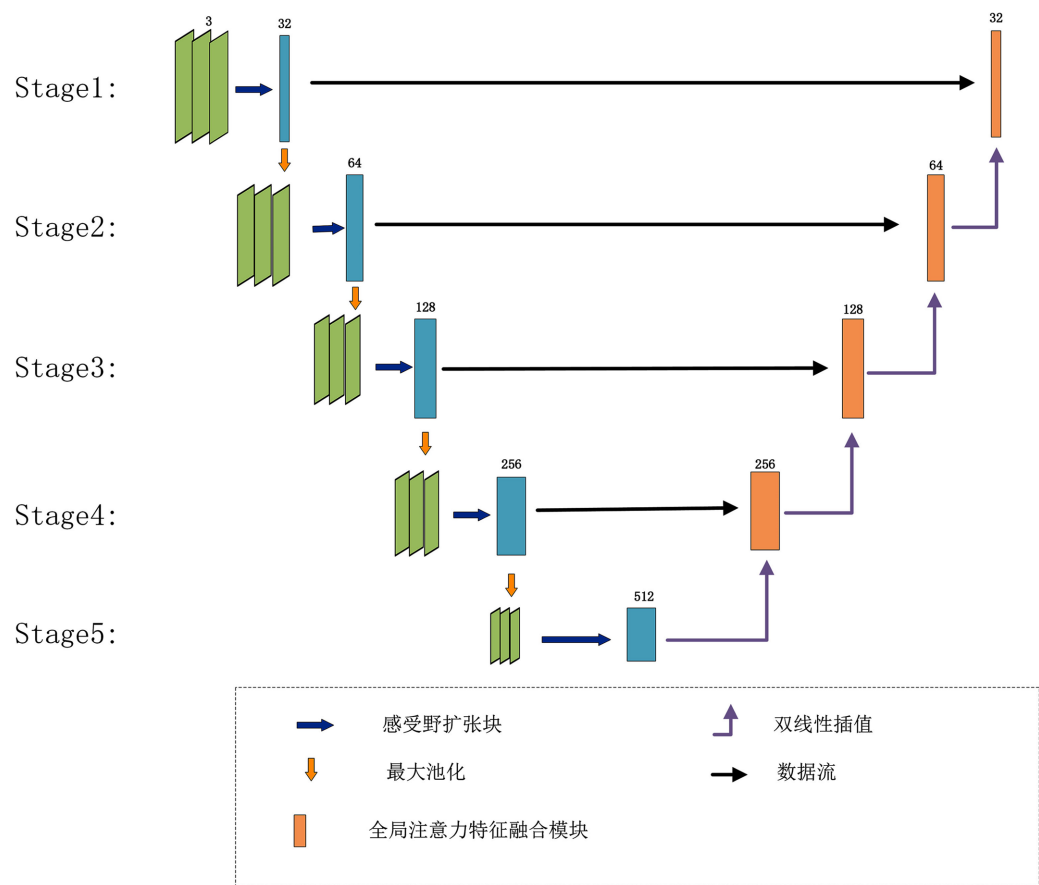


Figure 1. EEPSNet network architecture
图 1. EEPSNet 网络结构

2. EEPSNet 模型

2.1. 感受野增强卷积块(Enlarged Receptive Field Block)

普通卷积在捕捉局部特征方面表现优异，而空洞卷积则能够扩大感受野，捕获更广泛的上下文信息。因此，如图 2 所示，我们将这两种卷积操作结合起来，以期达到 $1 + 1 > 2$ 的效果。

具体来说，混合卷积块首先使用 3×3 的普通卷积对输入特征进行初步过滤，这一步骤有助于提取出图像的基本纹理和形状信息。紧接着，我们引入了空洞率为 2 的 3×3 卷积，这一操作能够在不增加计算复杂度的前提下，扩大感受野，捕获到更远距离的特征信息。这种设计使得模型能够同时关注到局部细

节和全局上下文，从而更全面地理解图像内容。在每次卷积操作之后，我们都会进行批量归一化(Batch Normalization)和 ReLU (Rectified Linear Unit)激活函数的处理。批量归一化有助于加速模型的收敛速度，同时减少内部协变量偏移的问题，而 ReLU 激活函数则能够引入非线性，增加模型的表达能力。

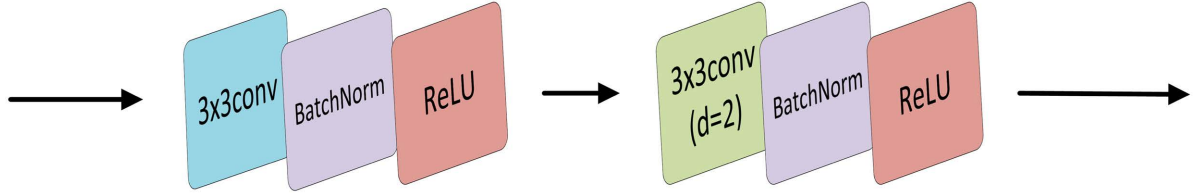


Figure 2. Enlarged receptive field block

图 2. 感受野增强卷积块

混合卷积块公式表达如下：

$$I_{out} = F_{dil=2}(F_{std}(I_{in})) \quad (1)$$

其中， I_{in} 表示输入特征图， I_{out} 表示经过混合卷积块后的输出特征图。 F_{std} 与 $F_{dil=2}$ 分别表示普通卷积与空洞卷积，公式(1)表示特征图先进行普通卷积再进行空洞卷积。此外对于 F_{std} 与 $F_{dil=2}$ 用公式可以分别表示为：

$$F_{std} = \text{ReLU}\left(\text{BN}\left(C_{std}(I; K_{3 \times 3}, S=1, P=1)\right)\right) \quad (2)$$

$$F_{dil} = \text{ReLU}\left(\text{BN}\left(C_{std}(I; K_{3 \times 3}, S=1, P=2, D=2)\right)\right) \quad (3)$$

BN 表示批量归一化，ReLU 则表示 ReLU 激活函数。对于 C_{std} ， I 代表输入特征图， $K_{3 \times 3}$ 表示卷积核的大小为 3×3 ， $S=1$ 代表卷积核的滑动步长为 1， $P=1$ 表示被卷积的特征图边界填充为 1， $D=2$ 则说明空洞率为 2，一般默认情况下为 1。

2.2. 全局注意力特征融合模块(Global Attention Feature Fusion Module)

如图 3 所示，全局注意力特征融合模块 GAFM (Global Attention Feature Fusion Module)由双线性插值上采样、全局注意机制 GAM (Global Attention Mechanism)以及一个混合卷积块组成。具体操作为，将低一级别的特征在空间维度上双线性插值尺寸扩大一倍，在通道维度上降低通道数量为原来的一半。目的是在通道维度和空间维度上与高一级别的特征图对齐进而进行特征融合操作。对齐尺寸后，再将低级特征进行全局注意力和一个普通卷积处理与高级的特征进行通道上的拼接。低级特征在与高级特征进行通道拼接前，对于高级特征则进行与经过尺寸对齐的低级特征的元素相乘的操作。在上述操作之后，将经过通道维度拼接的特征输入到混合卷积块中进行卷积操作。公式表达为：

$$I_{out} = F_{hc}\left(\left(I_1; F_{gam}\left(F_{bil}(I_2)\right)\right)\right) \quad (4)$$

其中， I_1 、 I_2 分别表示高一级别的特征和低一级别的特征， F_{hc} 代表混合卷积块(Hybrid Convolutional Block)具体操作参考公式(1)~(3)， F_{bil} 表示特征的双线性插值操作。

对于 F_{gam} 我们有如下的进一步描述，全局注意机制(GAM)是一种减少信息减少和放大大局维度的机制 - 交互特征，此机制采用 CBAM 的顺序通道 - 空间注意机制，并重新设计了子模块。整个过程如图 4、公式(5)、公式(6)所述。给定输入特征图 $F_1 \in R^{C \times H \times W}$ ，定义中间状态 F_2 和输出 F_3 为：

$$F_2 = M_c(F_1) \otimes F_1 \quad (5)$$

$$F_3 = M_s(F_2) \otimes F_2 \quad (6)$$

公式(5)和(6)中, M_c 和 M_s 分别为通道和空间注意图; $\otimes$ 表示逐元素的乘法。通道注意子模块使用 3D 排列来保留跨三维的信息。然后, 它用一个双层 MLP (多层感知器) 放大跨维通道空间依赖性。(MLP 是一个与 BAM 结构相同并且带有放缩比 r 的编码器 - 解码器结构)

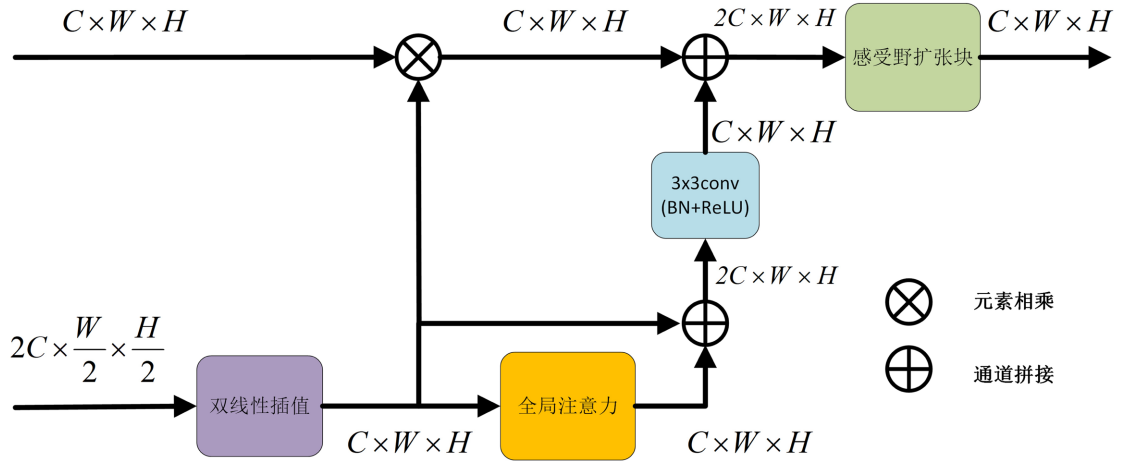


Figure 3. Global attention feature fusion module
图 3. 全局注意力特征融合模块

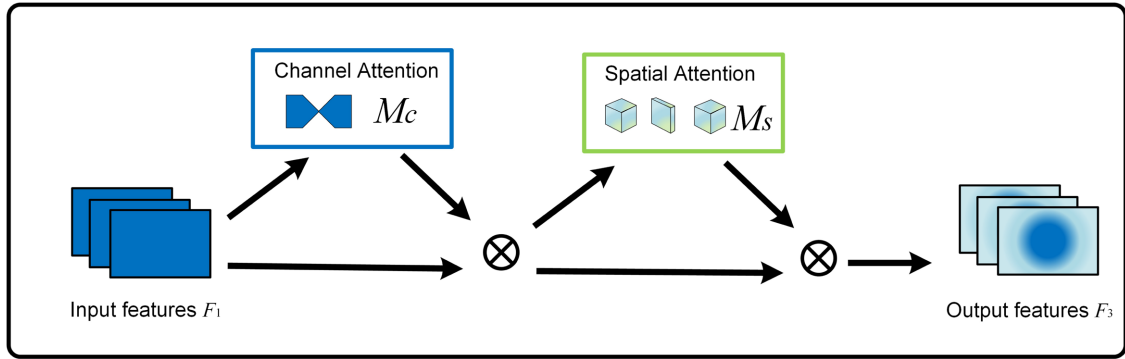


Figure 4. Global attention mechanism
图 4. 全局注意力机制

3. 实验

3.1. 数据集

为了全面评估我们提出的 EEPSNet 模型的性能, 我们遵循了 PraNet [12] 的实验设置, 在以下四个公开的结肠息肉图像数据集上进行了广泛的实验: Kvasir-SEG [13]、CVC-ClinicDB [14]、CVC-ColonDB [15] 和 ETIS [16]。这些数据集为我们提供了多样化的测试环境, 以全面评估 EEPSNet 在不同类型和质量结肠息肉图像上的表现。CVC-ClinicDB: 也称为 CVC-612, 包含 612 张从 25 个结肠镜检查视频中获取的结肠息肉图像, 图像分辨率为 384×384 像素。我们选取了 550 张图像用于训练, 剩下的 62 张用于测试。CVC-ColonDB: 一个小型数据库, 包含 380 张图像。Kvasir-SEG: 这是一个具有挑战性的数据集, 包含 1000 张息肉图像, 图像尺寸从 332×487 到 1920×1072 不等。通常, 我们使用数据集的 90% 进行训练, 剩余的 10% 用于测试。ETIS: 一个较早发布的数据集, 包含 196 张息肉图像, 图像大小为 1255×966 像素。

3.2. 实验设置

我们的实验是在 NVIDIA Quadro RTX 5000 GPU 上进行的, 该计算机配备了 10.0 版本的 CU-DA 和 2.2.2 版本的 PyTorch。使用 AdamW 优化器和余弦退火学习率调整策略, 学习率和权重衰减为 $1e-4$, 该模型总共训练了 150 个 epoch。通过多尺度训练、随机旋转、随机翻转来增强数据。输入图像尺寸调整为 352×352 。

3.3. 评价指标

我们使用与 PraNet [12]、Polyp-PVT [17] 相同的七个评估指标来评估所提出方法中的模型性能。我们使用平均骰子(mDice)、平均交并比(mIoU)、平均绝对误差(MAE)、加权 F 度量(F_β^ω)、S-measure (S_α)、平均和最大 E-measure (E_ξ) 作为评估指标来全面验证我们方法的性能。Dice 和 IoU 是息肉分割最常用的两个指标。Dice 系数是用于比较预测图和真实图之间的像素级分类结果的标准度量。IoU 关注分割对象的内部一致性。Dice 和 IoU 的取值范围在 $[0, 1]$ 之间, 值越接近 1, 分割效果越好。它们的定义如下:

$$\text{Dice}(A, B) = \frac{2 \times |A \cap B|}{|A| + |B|} = \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FP} + \text{FN}} \quad (7)$$

$$\text{IoU}(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \quad (8)$$

将 Dice 方程定义为公式(7), 将 IoU 方程定义为公式(8), A 变量为预测的像素, B 变量为掩码图的像素。其中, TP 表示真阳性, FP 表示假阳性, FN 表示假阴性。

$$\begin{cases} \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \\ \text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \\ F_\beta^\omega = \frac{(1 + \beta^2) \text{Precision}^\omega \cdot \text{Recall}^\omega}{\beta^2 \cdot \text{Precision}^\omega + \text{Recall}^\omega} \end{cases} \quad (9)$$

F_β^ω 被定义为公式(9)。其中 β 是权衡 Recall 和 Precision 的参数。Precision $^\omega$ 表示加权精度, 它是精确度的度量, Recall $^\omega$ 表示加权召回率, 它是完整性的度量。

$$\text{MAE} = \frac{1}{W \times H} \sum_{x=1}^W \sum_{y=1}^H |P(x, y) - Y(x, y)| \quad (10)$$

我们通过等式(10)计算 MAE 分数。P 和 Y 表示预测图和真实图。其中 W 和 H 分别是图像的宽度和高度。

$$S_\alpha = \gamma S_o + (1 - \gamma) S_r \quad (11)$$

我们通过等式(11)计算 S_α 。其中, S_o 和 S_r 表示区域感知和对象感知结构相似性, γ 默认设置为 0.5。

$$E_\xi = \frac{1}{W \times H} \sum_{x=1}^W \sum_{y=1}^H \phi FM(x, y) \quad (12)$$

E_ξ 方程定义为公式(12)。使用增强的对齐矩阵 ϕ 捕获二值图的两个属性(像素级匹配和图像级统计)。其中 W 和 H 分别是图像的宽度和高度。FM 代表前景图。我们使用这些指标来验证性能。

3.4. 损失函数

我们的损失函数可以表示为等式(13):

$$L = L_{IoU}^{\omega}(P, G) + L_{BCE}^{\omega}(P, G) \quad (13)$$

$L_{IoU}^{\omega}(\cdot)$ 和 $L_{BCE}^{\omega}(\cdot)$ 是加权并集交集(IoU)损失和加权二元交叉熵(BCE)损失, 它们在全局方面限制了预测图结构(对象级)和局部细节(像素级)视角。与平等对待所有像素的标准 BCE 损失函数不同, $L_{BCE}^{\omega}(\cdot)$ 考虑每个像素的重要性, 并为硬像素分配更高的权重。此外, 与标准 IoU 损失相比, $L_{IoU}^{\omega}(\cdot)$ 更关注硬像素。

4. 实验结果与分析

4.1. 定量分析

学习和泛化能力分析: 我们评估了模型在四个数据集上的性能, 其中 Kvasir-SEG 和 CVC-ClinicDB 用于评估模型的特征建模能力, 其他两个未见过的数据集 CVC-ColonDB 和 ETIS-LaribPolypDB 来评估泛化能力。我们在表 1~4 中列出了比较结果, 包括 U-Net [2]、U-Net++ [3]、ResUNet [4]、ResUNet++ [5]、Attention-UNet [18]、R2UNet [19]、SFANet [7]、SCR-Net [9]和 DCRNet [8]。

此外, 我们还展示了每个模型的可视化结果并分析了几个代表性案例, 以说明我们的模型具有更好的分割性能。

在特征建模方面, 我们的 EEPsNet 模型在 CVC-ClinicDB 数据集的七个评估指标中, 除了 $\max E_{\xi}$ 指标略低于 DCRNet 模型 0.2%且在 MAE 指标上与 DCRNet 持平外, 其他指标均优于其他模型。具体来说, 我们的模型在 mDice、mIoU、 F_{β}^{ω} 、 S_{α} 和 mE_{ξ} 五个指标上分别超越了次优模型 DCRNet 1.5%、1.2%、0.5%、0.1%和 0.5%。而在 Kvasir-SEG 数据集上, EEPsNet 的表现与 DCRNet 相当, 在 mIoU、 F_{β}^{ω} 、 $\max E_{\xi}$ 三个指标上分别高出 DCRNet 0.2%、0.1%和 0.5%, 而在 mDice 和 MAE 指标上与 DCRNet 持平。然而, 在另外两个指标 S_{α} 和 mE_{ξ} 上, 我们的模型分别落后于 DCRNet 0.4%和 1.1%。上述数据指标可以说明我们的模型在空间特征捕捉上有很大的提升。在 Kvasir-SEG 数据集上 S_{α} 和 mE_{ξ} 指标略低于 DCRNet 是因为 Kvasir-SEG 数据集中的图像数据尺寸方差较大, 模型训练结果存在一定的误差可忽略不计。

此外, 我们通过视觉对比分析了我们的模型与其他模型的分割效果, 如图 5 所示。第一行和第二行分别展示了 Kvasir-SEG 和 CVC-ClinicDB 数据集的分割结果。从这些图像中可以明显看出, 我们的模型在捕捉分割区域特征方面表现更佳, 不仅能够减少分割信息的丢失, 还能有效抑制非息肉区域的干扰, 这进一步验证了我们模型的有效性。

在泛化能力方面, 我们的模型在 CVC-ColonDB 和 ETIS-LaribPolypDB 这两个未见过的数据库上也展现出了显著的性能。在 ETIS-LaribPolypDB 数据集中, 与其它模型相比, 我们的模型在除了 MAE 指标略低于 SFA 方法 0.18%之外的所有六个指标上均取得了最佳成绩, 具体来说, 在 mDice、mIoU、 F_{β}^{ω} 、 S_{α} 、 mE_{ξ} 和 $\max E_{\xi}$ 指标上分别超越了次优模型 10.4%、8.9%、11.3%、7.2%、6.4%和 5.6%。在 CVC-ColonDB 数据集中, 我们的模型同样展现出了一定的优势, 除了 F_{β}^{ω} 与 S_{α} 两个指标分别低于 DCRNet 模型 0.4%和 0.1%, 并且在 mIoU 指标上与 DCRNet 持平外, 其他四个指标 mDice、 mE_{ξ} 、 $\max E_{\xi}$ 和 MAE 分别比 DCRNet 高出 0.4%、1.1%、2.3%和 1.3%。

此外, 我们在图 5 的第三行和第四行中, 对 CVC-ColonDB 和 ETIS-LaribPolypDB 两个数据集的分割效果进行了可视化对比。结果明显显示我们的模型在泛化能力上也优于其他模型。

Table 1. Experimental comparison of different network models on the CVC-ClinicDB dataset
表 1. 不同网络模型在 CVC-ClinicDB 数据集上的实验比较

CVC-ClinicDB							
Methods	mDice	mIoU	F_{β}^{ω}	S_{α}	mE_{ξ}	$\max E_{\xi}$	MAE
U-Net	0.823	0.755	0.811	0.889	0.913	0.954	0.019
U-Net++	0.794	0.729	0.785	0.873	0.891	0.931	0.022
ResUNet	0.732	0.621	0.690	0.814	0.868	0.878	0.036
ResUNet++	0.845	0.786	0.831	0.898	0.923	0.940	0.013
Attention-Unet	0.788	0.720	0.775	0.864	0.885	0.921	0.027
R2Unet	0.696	0.610	0.668	0.798	0.823	0.849	0.035
SFA	0.700	0.607	0.647	0.793	0.840	0.885	0.042
SCR-Net	0.791	0.711	0.772	0.855	0.886	0.906	0.028
DCR-Net	0.896	0.844	0.890	0.933	0.964	0.978	0.010
EEPS-Unet	0.911	0.856	0.895	0.934	0.969	0.976	0.010

Table 2. Experimental comparison of different network models on the Kvasir-SEG dataset
表 2. 不同网络模型在 Kvasir-SEG 数据集上的实验比较

Kvasir-SEG							
Methods	mDice	mIoU	F_{β}^{ω}	S_{α}	mE_{ξ}	$\max E_{\xi}$	MAE
U-Net	0.818	0.746	0.794	0.858	0.881	0.893	0.055
U-Net++	0.821	0.743	0.808	0.862	0.886	0.909	0.048
ResUNet	0.720	0.604	0.656	0.775	0.835	0.849	0.080
ResUNet++	0.826	0.750	0.796	0.861	0.888	0.893	0.051
Attention-Unet	0.843	0.771	0.823	0.875	0.903	0.909	0.045
R2Unet	0.785	0.696	0.765	0.831	0.864	0.872	0.059
SFA	0.723	0.611	0.670	0.782	0.834	0.849	0.075
SCRNet	0.798	0.711	0.755	0.836	0.871	0.888	0.060
DCRNet	0.886	0.825	0.868	0.911	0.941	0.933	0.035
EEPS-Unet	0.886	0.827	0.869	0.907	0.930	0.938	0.035

Table 3. Experimental comparison of different network models on the ETIS dataset
表 3. 不同网络模型在 ETIS 数据集上的实验比较

ETIS-LaribPolypDB							
Methods	mDice	mIoU	F_{β}^{ω}	S_{α}	mE_{ξ}	$\max E_{\xi}$	MAE
U-Net	0.398	0.335	0.366	0.684	0.643	0.740	0.036
U-Net++	0.401	0.344	0.390	0.683	0.629	0.776	0.035
ResUNet	0.411	0.315	0.364	0.656	0.678	0.685	0.044
ResUNet++	0.401	0.339	0.364	0.669	0.651	0.688	0.039

续表

Attenton-Unet	0.539	0.470	0.510	0.739	0.722	0.772	0.029
R2Unet	0.360	0.303	0.339	0.643	0.621	0.755	0.033
SFA	0.297	0.217	0.231	0.557	0.531	0.632	0.109
SCRNet	0.449	0.371	0.405	0.681	0.672	0.727	0.042
DCRNet	0.556	0.496	0.506	0.736	0.742	0.773	0.096
EEPS-Unet	0.660	0.585	0.619	0.808	0.806	0.829	0.037

Table 4. Experimental comparison of different network models on the CVC-ColonDB dataset

表 4. 不同网络模型在 CVC-ColonDB 数据集上的实验比较

CVC-ColonDB							
Methods	mDice	mIoU	F_{β}^w	S_{α}	mE_{ξ}	$\max E_{\xi}$	MAE
U-Net	0.512	0.444	0.498	0.712	0.696	0.776	0.061
U-Net++	0.483	0.410	0.467	0.691	0.680	0.760	0.064
ResUNet	0.528	0.416	0.489	0.690	0.755	0.761	0.060
ResUNet++	0.597	0.504	0.553	0.738	0.776	0.791	0.056
Attenton-Unet	0.635	0.549	0.606	0.765	0.797	0.832	0.055
R2Unet	0.629	0.537	0.597	0.767	0.793	0.823	0.046
SFA	0.469	0.347	0.379	0.634	0.675	0.764	0.094
SCRNet	0.616	0.527	0.579	0.752	0.788	0.832	0.049
DCRNet	0.704	0.631	0.684	0.821	0.840	0.848	0.052
EEPS-Unet	0.708	0.631	0.680	0.820	0.851	0.871	0.039

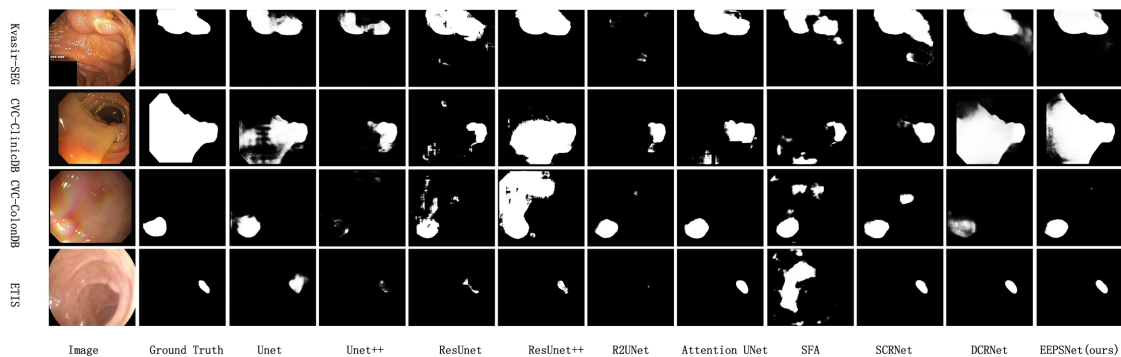


Figure 5. Comparison of the visual segmentation results for the different methods

图 5. 不同方法的视觉分割结果比较

4.2. 消融实验

为了验证我们模型中各个模块的有效性,我们在 Kvasir-SEG、CVC-ClinicDB、CVC-ColonDB 和 ETIS-LaribPolypDB 四个数据集上进行了消融实验,结果汇总在表 5 中。实验数据清晰地表明,无论是单独使用还是组合使用,我们的模块在这些数据集中都展现出了显著的效果。具体来说,混合卷积块(HCB)和全局注意力特征融合模块(GAFM)在单独加入网络时,就能显著提升性能,而当这两个模块联合使用时,性

能提升更为显著。这一结果证明了混合卷积在空间特征捕捉方面的能力优于传统的双卷积块。

此外，混合卷积块与感受野增强模块的组合在 ETIS-LaribPolypDB 和 CVC-ColonDB 数据集上的 mDice 指标分别提升了 26.2%和 19.6%，这一提升幅度大于在 Kvasir-SEG 上的 6.6%和 CVC-ClinicDB 上的 8.8%。这一差异表明我们的模型不仅在特定数据集上表现优异，还具有良好的泛化能力，能够在不同的数据环境中保持高效的表现。

Table 5. Comparison table of the ablation experiments
表 5. 消融实验对比表

	Methods	mDice	相比 baseline 提升
Kvasir-SEG	Baseline	0.818	0
	Baseline + HCB	0.880	6.2%
	Baseline + GAFM	0.882	6.4%
	Baseline + HCB + GAFM	0.886	6.6%
CVC-ClinicDB	Baseline	0.823	0
	Baseline + HCB	0.910	8.7%
	Baseline + GAFM	0.905	8.2%
	Baseline + HCB + GAFM	0.911	8.8%
ETIS-LaribPolypDB	Baseline	0.398	0
	Baseline + HCB	0.579	18.1%
	Baseline + GAFM	0.562	16.4%
	Baseline + HCB + GAFM	0.660	26.2%
CVC-ColonDB	Baseline	0.512	0
	Baseline + HCB	0.675	16.3%
	Baseline + GAFM	0.690	17.8%
	Baseline + HCB + GAFM	0.708	19.6%

5. 结语

本研究提出了一种基于 U-Net 架构的肠道息肉分割算法，该算法整合了混合卷积块(HCB)和全局注意力特征融合模块(GAFM)。我们设计的 EEPsNet 模型能够精确地识别并分割出肠道息肉区域，并且在四个公开的数据集上与其他现有模型进行了比较。实验结果证实，EEPSNet 在建模能力和泛化能力方面均有显著提升。

EEPSNet 模型通过增强空间特征学习能力，有效避免了信息丢失，同时在特征融合过程中，它能够高效地利用关键信息并抑制噪声干扰，从而增强了模型的鲁棒性。该模型的高分割性能使其适用于临床肠道息肉图像的检测，并为肠道息肉分割领域的研究提供了新的方向。

基金项目

广东省普通高校重点领域专项项目(2021ZDZX1032)；广东省国际及港澳台高端人才交流专项(2020A1313030021)；五邑大学科研项目(2018GR003)；江门市社会发展领域科技计划项目(2024)。

参考文献

- [1] Long, J., Shelhamer, E. and Darrell, T. (2015) Fully Convolutional Networks for Semantic Segmentation. 2015 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, 7-12 June 2015, 343-3440. <https://doi.org/10.1109/cvpr.2015.7298965>
- [2] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: *Lecture Notes in Computer Science*, Springer, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28
- [3] Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N. and Liang, J. (2018) U-Net++: A Nested U-Net Architecture for Medical Image Segmentation. In: *Lecture Notes in Computer Science*, Springer, 3-11. https://doi.org/10.1007/978-3-030-00889-5_1
- [4] Zhang, Z., Liu, Q. and Wang, Y. (2018) Road Extraction by Deep Residual U-Net. *IEEE Geoscience and Remote Sensing Letters*, **15**, 749-753. <https://doi.org/10.1109/lgrs.2018.2802944>
- [5] Jha, D., Smedsrud, P.H., Riegler, M.A., Johansen, D., Lange, T.D., Halvorsen, P., et al. (2019). ResUNet++: An Advanced Architecture for Medical Image Segmentation. 2019 *IEEE International Symposium on Multimedia (ISM)*, San Diego, 9-11 December 2019, 225-2255. <https://doi.org/10.1109/ism46123.2019.00049>
- [6] Fan, D., Ji, G., Zhou, T., Chen, G., Fu, H., Shen, J., et al. (2020) Prant: Parallel Reverse Attention Network for Polyp Segmentation. In: *Lecture Notes in Computer Science*, Springer, 263-273. https://doi.org/10.1007/978-3-030-59725-2_26
- [7] Fang, Y., Chen, C., Yuan, Y. and Tong, K. (2019) Selective Feature Aggregation Network with Area-Boundary Constraints for Polyp Segmentation. In: *Lecture Notes in Computer Science*, Springer, 302-310. https://doi.org/10.1007/978-3-030-32239-7_34
- [8] Yin, Z., Liang, K., Ma, Z. and Guo, J. (2022) Duplex Contextual Relation Network for Polyp Segmentation. 2022 *IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, Kolkata, 28-31 March 2022, 1-5. <https://doi.org/10.1109/isbi52829.2022.9761402>
- [9] Wu, H., Zhong, J., Wang, W., Wen, Z. and Qin, J. (2021) Precise Yet Efficient Semantic Calibration and Refinement in Convnets for Real-Time Polyp Segmentation from Colonoscopy Videos. *Proceedings of the AAAI Conference on Artificial Intelligence*, **35**, 2916-2924. <https://doi.org/10.1609/aaai.v35i4.16398>
- [10] Zhang, W., Lu, F., Su, H. and Hu, Y. (2024) Dual-Branch Multi-Information Aggregation Network with Transformer and Convolution for Polyp Segmentation. *Computers in Biology and Medicine*, **168**, Article 107760. <https://doi.org/10.1016/j.compbiomed.2023.107760>
- [11] Shu, X., Wang, J., Zhang, A., Shi, J. and Wu, X. (2024) CSCA U-Net: A Channel and Space Compound Attention CNN for Medical Image Segmentation. *Artificial Intelligence in Medicine*, **150**, Article 102800. <https://doi.org/10.1016/j.artmed.2024.102800>
- [12] Fan, D., Ji, G., Zhou, T., Chen, G., Fu, H., Shen, J., et al. (2020) Prant: Parallel Reverse Attention Network for Polyp Segmentation. In: *Lecture Notes in Computer Science*, Springer, 263-273. https://doi.org/10.1007/978-3-030-59725-2_26
- [13] Jha, D., Smedsrud, P.H., Riegler, M.A., Halvorsen, P., de Lange, T., Johansen, D., et al. (2019) Kvasir-Seg: A Segmented Polyp Dataset. In: *Lecture Notes in Computer Science*, Springer, 451-462. https://doi.org/10.1007/978-3-030-37734-2_37
- [14] Vázquez, D., Bernal, J., Sánchez, F.J., Fernández-Esparrach, G., López, A.M., Romero, A., et al. (2017) A Benchmark for Endoluminal Scene Segmentation of Colonoscopy Images. *Journal of Healthcare Engineering*, **2017**, 1-9. <https://doi.org/10.1155/2017/4037190>
- [15] Bernal, J., Sánchez, J. and Vilariño, F. (2012) Towards Automatic Polyp Detection with a Polyp Appearance Model. *Pattern Recognition*, **45**, 3166-3182. <https://doi.org/10.1016/j.patcog.2012.03.002>
- [16] Silva, J., Histace, A., Romain, O., Dray, X. and Granado, B. (2013) Toward Embedded Detection of Polyps in WCE Images for Early Diagnosis of Colorectal Cancer. *International Journal of Computer Assisted Radiology and Surgery*, **9**, 283-293. <https://doi.org/10.1007/s11548-013-0926-3>
- [17] Dong, B., Wang, W., Fan, D.P., et al. (2021) Polyp-Pvt: Polyp Segmentation with Pyramid Vision Transformers.
- [18] Oktay, O., Schlemper, J., Folgoc, L.L., et al. (2018) Attention U-Net: Learning Where to Look for the Pancreas.
- [19] Alom, M.Z., Yakopcic, C., Taha, T.M. and Asari, V.K. (2018) Nuclei Segmentation with Recurrent Residual Convolutional Neural Networks Based U-Net (R2U-Net) for Medical Image Segmentation.