

DOI: 10.13382/j.jemi.B2408070

卷积和自注意力机制双融合的医学图像分割网络*

范晓冬¹ 张楚浩¹ 周 婧² 侯利民¹ 韩 莹¹

(1. 辽宁工程技术大学电气与控制工程学院 葫芦岛 125105; 2. 渤海大学数学科学学院 锦州 121000)

摘要:针对医学图像分割中局部纹理细节与全局语义信息难以充分融合、易造成边界模糊与结构不完整的问题,设计了一种卷积与自注意力并行的双融合分支医学图像分割U形网络(dual fusion branch UNet, DFB-UNet)。编码器采用卷积神经网络(convolutional neural network, CNN)分支与Transformer分支并行提取局部与全局特征;在跨分支融合前端引入可变形卷积网络(deformable convolution network, DCN)以增强对不规则目标边界的建模能力;并构建由注意力融合模块(attention fusion module, AFM)与卷积融合模块(convolution fusion module, CFM)组成的双融合模块,其中CFM引入深度可分离卷积以在较低计算开销下实现有效特征交互与互补聚合。实验结果显示,在Kvasir-SEG数据集上mDice/mIoU达到0.89/0.83;在CVC-ClinicDB数据集上mDice/mIoU为0.810/0.723;跨数据集测试中,使用Kvasir-SEG训练并在CVC-ClinicDB测试时Dice/IoU为0.751/0.668;在ISIC2017数据集上IoU/Dice/Acc为0.816/0.889/0.965,在LUNA16数据集上Dice/IoU为0.966 0/0.934 2。结果表明,所提出的双融合并行编码结构能够在保证计算效率的同时提升分割精度与泛化性能,消融实验进一步验证了可变形卷积与双融合模块对网络性能提升的贡献。

关键词: 医学图像分割;卷积神经网络;Transformer;可变形卷积;深度可分离卷积

中图分类号: TN911.73;TP391.41 **文献标识码:** A **国家标准学科分类代码:** 510.4050

Medical image segmentation network based on dual fusion of convolutions and self-attention mechanisms

Fan Xiaodong¹ Zhang Chuhaohao¹ Zhou Jing² Hou Limin¹ Han Ying¹

(1. College of Electrical and Control Engineering, Liaoning Technical University, Huludao 125105, China;

2. College of Mathematical Sciences, Bohai University, Jinzhou 121000, China)

Abstract: To address the insufficient fusion of local texture details and global semantic information in medical image segmentation, which often leads to blurred boundaries and incomplete structures, a named dual fusion branch UNet (DFB-UNet) is proposed with parallel convolution and self-attention. The encoder employs a convolutional neural network (CNN) branch and a Transformer branch in parallel to learn local and global features. Before cross-branch fusion, a deformable convolution network (DCN) is introduced to enhance the modeling of irregular object boundaries. In addition, a dual-fusion module composed of an attention fusion module (AFM) and a convolution fusion module (CFM) is constructed; CFM incorporates depthwise separable convolutions to achieve effective feature interaction and complementary aggregation with lower computational cost. The experiment results are as follows: mDice/mIoU reach 0.89/0.83 on the dataset Kvasir-SEG; 0.810/0.723 on CVC-ClinicDB. In cross-dataset testing, training on Kvasir-SEG and testing on CVC-ClinicDB yield Dice/IoU of 0.751/0.668. On ISIC2017, IoU/Dice/Acc are 0.816/0.889/0.965, and on LUNA16, Dice/IoU are 0.966 0/0.934 2. The results indicate that the proposed dual-fusion parallel encoder improves segmentation accuracy and generalization while maintaining computational efficiency. The ablation studies further verify the contribution of DCN and the dual-fusion module.

Keywords: medical image segmentation; convolutional neural network; Transformer; deformable convolution; depthwise separable convolution

收稿日期: 2024-12-25 Received Date: 2024-12-25

* 基金项目: 国家自然科学基金(52177047, 62203197)项目资助

0 引言

医学图像通常由 X 射线、核磁共振、计算机断层扫描 (computed tomography, CT) 成像^[1], 是现代医学诊疗必不可少的诊疗技术。医学图像分割任务是由计算机从医学图像中智能地分割出人体组织以及病灶的区域, 有助于医生快速识别和检测病变的位置和形状, 从而提高医学诊断的效率^[2]。此类语义分割任务需要为输入图像的每个像素标注类别标签, 最终得到逐像素语义标注的掩码图^[3]。医学图像分割属于像素级语义标注任务, 通常依赖高质量掩码标注; 并且像素级标注获取代价高且一致性难保证, 这在一定程度上限制了模型的训练与推广。

自 20 世纪 70 年代起医学图像分割已有阈值分割、区域生长、活动轮廓、水平集与图割等经典方法, 并在部分特定模态与器官的规则场景中较为成熟; 但这类方法对手工特征与经验参数依赖较强, 在噪声干扰、对比度变化及跨设备分布差异等情况下鲁棒性与泛化能力易受影响, 因此近年来深度学习方法逐渐成为主流方向^[4]。

基于深度学习的医学图像分割技术在医学图像处理领域的广泛应用为医学诊断和治疗提供了更准确和高效的支持。Ronneberger 等^[5]提出了一种基于卷积神经网络 (convolutional neural network, CNN) 的 U 型编码器-解码器结构 U 形网络 (U-shaped network, U-Net), 对医学图像进行端到端分割, 使生物医学图像的语义分割技术取得很大的进步, 大大提升了语义分割的精确度。然而, U-Net 通过跳跃连接直接将编码器的特征发送给解码器, 将来自编码器的浅层信息和来自解码器的深层信息直接相连, 可能导致信息不匹配。在此基础上, Zunair 等^[6]改进了跳跃连接方式, 提出了 Sharp U-Net。该方法将来自编码器的特征通过锐化滤波器后再连接到解码器, 不再将编码器和解码器的特征直接相连, 这将更好的捕获输入特征的边界信息, 同时也有效缓解信息不匹配的问题。过去的五年中, 研究者基于 U-Net 框架在其跳跃连接和编码器结构上进行了有效的改进。DualA-Net^[7]应用多尺度跳跃连接以结合低级细节和高级语义信息, 从不同尺度中提取出更有利的信息。

随着跳跃连接的不断优化, U-Net 框架中的卷积单元也得到了进一步的改良。卷积固有的局部特性限制了网络性能的提升, 研究者通过堆叠卷积层来增加感受野, 提高网络获取全局信息的能力。Dual U-Net^[8]通过堆叠两个 U-Net 模型相互协作, 实现了对图像的高效编码和解码^[9]。CM-SegNet^[10]将多层感知机 (multilayer perceptron, MLP) 与卷积相结合, 实现了对补丁间相关性

的多尺度特征提取。UCR-Net^[11]增加了残差连接, 将不同功能的网络模块堆叠在一起, 捕捉图像的高级语义特征。然而, 网络模块的堆叠会大幅度增加模型的计算复杂度, 并且网络的加深在一定程度上会造成图像的边界信息丢失^[12]。为此, 研究者尝试使用大卷积核来增加网络的感受野。例如, RepLKNet^[13]使用 31×31 的超大卷积核提取输入图像特征, 以大幅度增加感受野区域。Tang 等^[11]提出的 CMUNet^[14]使用基于大卷积核的深度卷积来提取每个通道的全局信息。3D UX-Net^[15]使用大卷积核的深度卷积提取 3D 图像的特征信息。

U-Net 以其 U 型的编码器-解码器结构成为医学图像分割任务中的基本网络框架。在此架构之上, 卷积神经网络的性能不断提升, 如 BRAU-Net++^[16]、U-Net v2^[17] 和 FDR-TransUNet^[18]。然而, 卷积窗口逐行扫描的计算机制, 不利于提取特征间的全局相关性。因此, 仅依赖卷积神经网络往往难以实现建模远距离依赖关系, 在复杂结构或尺度变化明显的病灶场景下容易出现全局语义不一致的问题。此建模能力也是像素级的图像分割任务亟待解决的问题。

Transformer^[19]在自然语言处理中取得了长足的进步。主要是因为其中多头自注意力机制 (multi-head self-attention, MSA) 可以有效地在序列语义单元之间建立全局任意连接, 提高了模型的全局相关性建模能力。所以将获取全局信息能力较强的 Transformer 引入计算机视觉领域中就成为一种合理的尝试^[20]。Dosovitskiy 等^[21]提出了视觉 Transformer (Vision Transformer, ViT), 使用不重叠的图像补丁作为语义单元输入到 Transformer 网络。然而, 图像补丁之间并不具有自然语言中语义单元间的独立性。为了提升图像补丁间的信息交互能力, Liu 等^[22]提出的 Swin-Transformer, 设计了移动窗口来提取补丁之间的交互特征。在医学图像分割领域, TSCTA-Net^[23]以 MSA 为基本单元设计了分割网络的编码器和解码器, 网络性能有所提升。Transformer 架构固然能够捕捉全局信息^[24], 但对于细节信息的获取就有所欠缺, 特别是面对医学图像提取边界细节信息时。这使得 Transformer 结构在医学图像这种边界模糊、前景背景相似的场景中, 往往需要额外的细节补偿机制^[25]。

为了将卷积的局部特征提取能力与自注意力机制的远距离依赖性结合起来, 文献 [26] 试图将 CNN 和 Transformer 融合为一个网络, 优势互补。Zhang 等^[27]提出的 ETUNet 在基于 CNN 的编码器结构中引入 Transformer 单元, 以提升网络的全局特征提取能力。MobileUtr^[28]采用 U 型网络作为主干, Transformer 作为底部连接, 将 CNN 和 Transformer 结合起来, 实现编码器和解码器的通信。DEAU-Net^[29]采用了卷积和 Transformer 双编码器结构, 分别从宏观和微观角度实现特征提取, 残

差模块和注意力模块对特征进行补充。MS-TCNet^[30]利用多尺度 CNN 和 Transformer 构建网络编码器和解码器。UniFormer^[31]交替叠加 3D 卷积和自注意力机制构建了视频分割网络。

上述工作表明,将 CNN 与 Transformer 相结合在医学图像分割中具有很大的潜力。但是现有方法大多采用顺序式堆叠,顺序结构提取的局部与全局信息会存在二者特征学习不一致问题,且难以在同一层级实现两种信息的充分交互。为此本文在有效融合 CNN 和 Transformer,提出了双融合医学图像分割网络,应用深度可分离卷积单独设立为一个融合网络分支,设计了并行双分支编码器并引入融合机制,对编码器的输出特征进行局部和全局的双重融合,增强了两类特征信息的交互性。对比实验结果显示所设计网络在分割精度和计算效率上略有优势。消融实验验证了所设计网络模块的有效性。

1 本文方法

本文提出卷积和自注意力并行的医学图像分割双融合分支 U 形网络(dual fusion branch UNet,DFB-UNet),如图 1 所示。所设计网络采用卷积与自注意力机制并行的双分支编码器,以同时提取输入图像的局部细节与全局语义特征。针对医学病灶边界不规则、轮廓不清晰等特点,本文在两分支特征融合前引入可变形卷积网络(deformable convolution network,DCN),以增强对边界区域的表征能力。为进一步避免并行结构中可能出现的“有并行无交互”现象,设计由注意力融合分支与卷积融合分支组成的双融合模块,从不同角度实现跨分支特征的交互与互补融合。最终,经包含注意力门控机制的解码器逐级恢复空间分辨率,输出像素级分割结果。

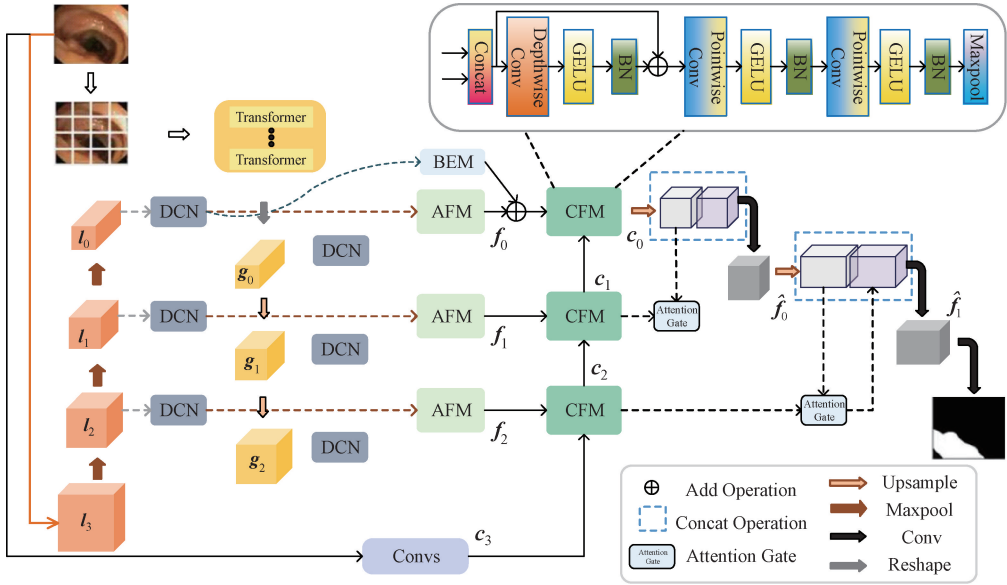


图 1 双融合 UNet 网络框架

Fig. 1 Framework of dual fusion UNet

1.1 融入可变形卷积的编码器

常规卷积的感受野均为规则矩形,可变形卷积^[32]引入采样点偏移量参数,可以根据图像学习不规则的感受野。同时,医学图像的病变区域多具有不规则的边界。可变形卷积更适合病灶位置不规则的医学图像分割任务。故本文在网络中使用了 6 个可变形卷积模块 DCN,并将其置于 CNN 分支和 Transformer 分支的输出特征融合前,以加强提取病灶的边界信息。

由图 1 可见,Transformer 分支 3 个阶段的输出

$$g_0 \in \frac{H}{16} \times \frac{W}{16} \times D_0, g_1 \in \frac{H}{8} \times \frac{W}{8} \times D_1 \text{ 和 } g_2 \in \frac{H}{4} \times \frac{W}{4} \times$$

D_2 以及 CNN 分支前 3 个阶段的输出 $I_0 \in \frac{H}{16} \times \frac{W}{16} \times 4C$ 、

$I_1 \in \frac{H}{8} \times \frac{W}{8} \times 2C$ 和 $I_2 \in \frac{H}{4} \times \frac{W}{4} \times C$ 通过卷积核大小

为 3×3 , 步长为 1, 填充为 1 的卷积获得可变形卷积的偏移量,以调整采样感受野提取病灶的不规则边界特征。这里之所以未将卷积分支的 I_3 特征进行可变形卷积,是因为卷积网络的浅层特征较为完整的保留了图像的边界特征,后续的消融实验也验证了该事实。

1.2 并行的卷积块注意模块(CBAM)

原始卷积块注意机制 CBAM^[33]由通道注意机制和

空间注意机制组合而成。在计算空间注意权重与输入特征逐元素相乘时,每个通道上的空间注意权重是一致的,导致空间注意力权重不能根据每个通道的具体特征进行动态调整,削弱网络的性能。本文旨在应用卷积模块改进空间注意机制,提高网络效率。改进的空间注意机制应用3层不同尺寸的条形卷积核和一个规则形状的卷积核并行提取输入特征,使用多尺度深度条形卷积核确保了有效的信息提取,同时降低了计算复杂度^[34]。并行卷积块注意模块(parallel convolutional block attention module, PCBAM)的流程如图2所示。首先,应用传统的通道注意机制获取通道注意权重。接下来,应用深度卷

积块模拟空间注意机制提取每个通道的空间位置信息,动态获得每个通道上的空间注意权重,然后再与输入特征逐元素相乘。同时,卷积块的特征提取能力受卷积核大小的影响,更大的卷积核具有更大的感受野,能够有效获取全局信息,从而增强其捕获语义信息的能力。相反,较小的卷积核在提取纹理信息方面是有效的^[35]。医学图像分割任务既需要掌握全局信息,又需要提取像素级细节信息。故本文分别选择5、7、11、21为4层卷积层的卷积核尺寸,兼顾大尺寸和小尺寸卷积核,能保证全面提取输入特征的信息。其中5×5卷积为常规卷积,其他均为条形卷积核。

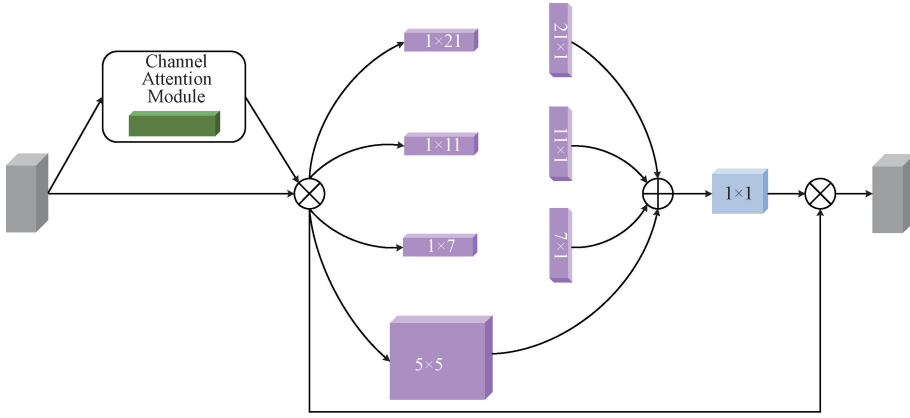


图2 并行的卷积块注意模块

Fig. 2 Parallel convolutional block attention module

PCBAM 的计算如式(1)和(2)所示。

$$F_c = CA(F) \otimes F \quad (1)$$

$$\hat{F} = Conv_{1 \times 1} \left[\sum_{i=1}^4 conv_i(F_c) \right] \otimes F_c \quad (2)$$

式中: $\hat{F}$ 为PCBAM的输出; $\otimes$ 代表逐元素相乘; $CA(\cdot)$ 表示通道注意机制计算过程; $Conv_{1 \times 1}$ 代表1×1卷积; $conv_i (i=1, 2, 3, 4)$ 表示不同大小卷积核的卷积过程, $conv_1$ 表示卷积核为5×5的卷积, $conv_2$ 表示卷积核为1×7,7×1堆叠的卷积,以此类推, $conv_3$ 、 $conv_4$ 表示图中上两层卷积。

1.3 注意融合模块(AFM)

为了更有效地将CNN分支和Transformer分支的特征融合起来,本文提出了AFM。由于医学图像在不同颜色通道上的信息分布不均匀,而且医学图像分割任务更多关注图像的病变区域,故AFM从空间和通道两个维度同时提取特征,融合来自CNN分支的局部特征和来自Transformer分支的全局特征。AFM由两条注意路径和一条加权路径组成。AFM的正向传播过程如图3所示。

首先,将来自Transformer分支的全局特征 g_i 与来自CNN分支的局部特征 l_i 分别进入相应的注意路径, g_i 和

l_i 对应元素相加后进入加权路径。注意路径的基本单元是CBAM,它是一个由通道注意机制(channel attention, CA)和空间注意机制(spatial attention, SA)串联而成的结构。CA为每个通道分配权重,解析特征的通道分布。SA在空间中学习权重,以突出对象并抑制背景。

加权路径为两个分支提供权重。输入依次经过全局平均池化,由两层全连接构成的MLP以及softmax函数,最终得到赋予权重的交互特征 b_i 如式(3)所示。

$$b_i = \text{softmax}(\text{MLP}(\text{Avgpool}(l_i + g_i))) \quad (3)$$

接下来,将两个注意路径的输出分别按对应元素相乘的方式与赋予权重的交互特征 b_i 结合,得到特征 $\hat{g}_i$ 和 $\hat{l}_i$ 。这两个特征按元素相加后经过残差结构得到该模块的最终输出如式(4)~(6)所示。

$$\hat{g}_i = \text{CBAM}(g_i) \otimes b_i \quad (4)$$

$$\hat{l}_i = \text{CBAM}(l_i) \otimes b_i \quad (5)$$

$$f_i = \text{Residual}(\hat{g}_i + \hat{l}_i) \quad (6)$$

式中: f_0 为AFM块的输出; $\text{CBAM}(\cdot)$ 为CBAM结构的计算过程; $\text{Residual}[\cdot]$ 为残差结构,在一定程度上防止了梯度消失、爆炸和网络退化的问题,有效地捕获了每个

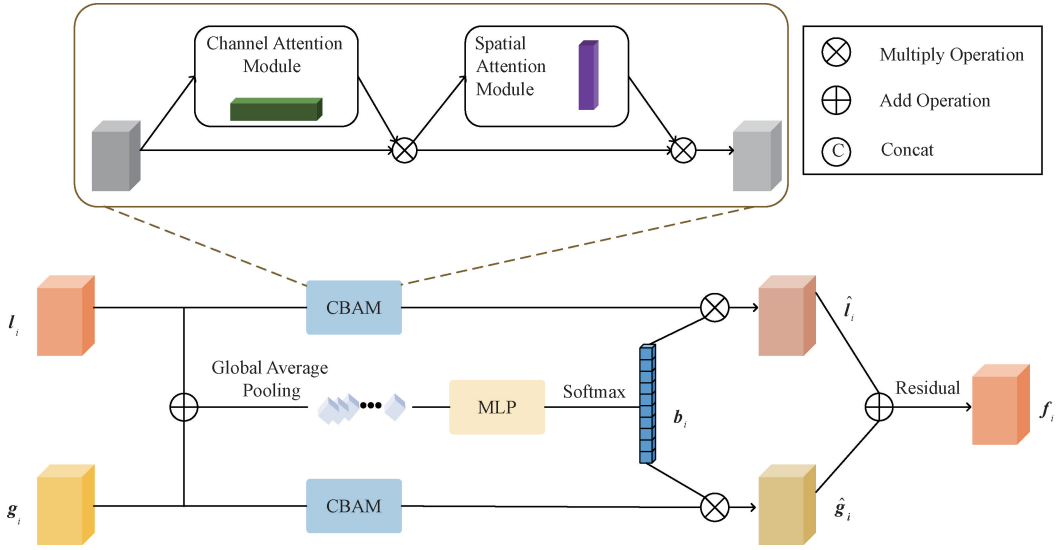


图 3 注意融合模块

Fig. 3 Attention fusion module

层次上的全局和局部特征信息。

1.4 边界提取模块 (BEM)

考虑到网络的浅层阶段主要关注高频信息,所以网络浅层会更多地用于捕获边界细节信息,故本文在第 1 个 AFM 前添加了一个 BEM。CNN 分支与 Transformer 分支第 1 个阶段的输出 g_0 与 l_0 在送入 AFM 的同时,也同时送入 BEM。所设计的 BEM 只由卷积操作构成,具体框架如图 4 所示。首先,将特征 g_0 与 l_0 分别经过 3×3 卷积,将其通道数分别变为 512 和 128。接下来的第 2 个 3×3 卷积层在提取特征的同时将两个分支的特征重塑为有相同大小的张量 $256 \times \frac{H}{16} \times \frac{W}{16}$ 。最后,将两种新的特征按对应元素相加,获得边界提取特征。

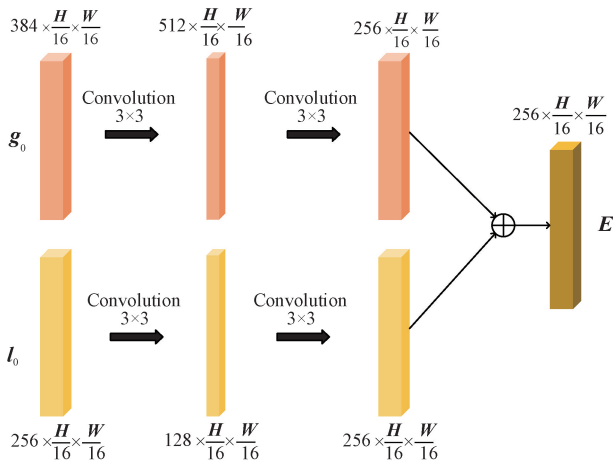


图 4 边界提取模块

Fig. 4 Boundary extraction module

1.5 双融合分支

双融合分支由注意融合分支和卷积融合分支合并而成。注意融合分支由 3 阶段构成,每个阶段的主要部件为并行的 PCBAM。这两个融合分支第 1 阶段的输出有所不同,该阶段注意融合分支的输出要考虑 AFM,即来自 CNN 分支和 Transformer 分支经过 DCN 的特征 l_0 与 g_0 分别经过 BEM 和 AFM,并且两个模块的输出进行逐元素相加后作为第 1 个 AFM 的输出如式 (7) 所示。

$$f_0 = f + E \quad (7)$$

式中: f 为第 1 阶段 AFM 的输出; E 为 BEM 的输出; f_0 为增加了 BEM 后注意融合分支第 1 阶段的输出。AFM 的具体计算过程如式 (8) ~ (11) 所示。

$$b_i = \text{softmax}(\text{MLP}(\text{GAP}(\tilde{l}_i + \tilde{g}_i))) \quad (8)$$

$$l_i = b_i \otimes \text{PCBAM}(\tilde{l}_i) \quad (9)$$

$$g_i = b_i \otimes \text{PCBAM}(\tilde{g}_i) \quad (10)$$

$$f_i = l_i + g_i \quad (11)$$

式中: $f_i (i = 1, 2)$ 为第 2 和第 3 阶段 AFM 的输出; $\tilde{l}_i$ 和 $\tilde{g}_i$ 分别为来自 CNN 分支和 Transformer 分支对应阶段的输出; b_i 为融合编码器两分支输出特征后的交互特征。

卷积融合分支由 3 个卷积融合模块 (convolution fusion module, CFM) 顺序堆叠而成,每个 CFM 主要应用深度可分离卷积融合来自注意融合分支的输出特征。由图 1 可见,深度可分离卷积由深度卷积和逐点卷积构成。不同于普通卷积,深度卷积首先将输入图像按通道分离,在每个通道进行卷积核不同的逐通道卷积。逐点卷积旨在改变特征的通道数,卷积核数输出输出特征数目。卷积融合分支具体计算过程如式 (12) ~ (15) 所示。

$$\mathbf{c}_i''' = \text{Concat}(\mathbf{f}_i, \mathbf{c}_{i+1}) \quad (12)$$

$$\mathbf{c}_i'' = \text{BN}(\sigma_1(\text{Depthwise}(\mathbf{c}_i''')) + \mathbf{c}_i''') \quad (13)$$

$$\mathbf{c}_i' = \text{BN}(\sigma_1(\text{Pointwise}(\mathbf{c}_i'')) + \mathbf{c}_i'') \quad (14)$$

$$\mathbf{c}_i = \text{BN}(\sigma_1(\text{Pointwise}(\mathbf{c}_i')) + \mathbf{c}_i') \quad (15)$$

式中: $\mathbf{c}_i (i = 0, 1, 2)$ 为注意融合分支不同阶段 CFM 的输出; $\sigma_i(\cdot)$ 表示高斯误差线性单元 (Gaussian error linear unit, GELU) 激活函数; $\text{BN}(\cdot)$ 代表批量归一化; $\text{Depthwise}(\cdot)$ 和 $\text{Pointwise}(\cdot)$ 分别表示深度卷积和逐点卷积的计算过程。

CFM 的输入为一对数据,故首先对输入对进行拼接操作,接下来再进行深度卷积。针对注意融合分支来说,随着网络的加深,输入特征的分辨率在逐层降低。考虑此情况,卷积融合分支的从底端介入。因为底端的特征分辨率大,很容易就能将输入图像下采样至 $\frac{H}{4} \times \frac{W}{4}$, 与对应阶段 AFM 的输出尺寸匹配。若从顶端进行,需要利用多层卷积将特征下采样至 $\frac{H}{16} \times \frac{W}{16}$ 才符合 AFM 的尺寸大小,这会增加图片的信息损失,降低网络性能。本文网络应用两层卷积将原始输入医学图像下采样至 $\frac{H}{4} \times \frac{W}{4}$, 记为 $\mathbf{c}_3$, 具体计算过程如式 (16) 和 (17) 所示。

$$\mathbf{c}_3' = \text{Maxpool}(\sigma_2(\text{conv}_1(\mathbf{X}))) \quad (16)$$

$$\mathbf{c}_3 = \text{Maxpool}(\sigma_2(\text{conv}_2(\mathbf{c}_3'))) \quad (17)$$

式中: σ_2 为修正线性单元 (rectified linear unit, ReLU) 激活函数; conv_1 、 conv_2 为参数不同的两个 3×3 卷积。

最终, $\mathbf{c}_3$ 与第 3 阶段 AFM 的输出 $\mathbf{f}_2$ 共同作为第三阶段 CFM 的输入对。

2 实验结果与分析

本文设计了 3 种类型实验,在 Kvasir-SEG 数据集、CVC-ClinicDB 数据集和肺分割数据集上对比本文网络和其他网络的性能。其中,Kvasir-SEG 含 1 000 张息肉图像及像素级标注,按常用划分使用 880 张训练、120 张测试;CVC-ClinicDB 含 612 张息肉图像及像素级标注,按常用划分使用 550 张训练、62 张测试,且由 29 个内镜视频序列抽帧构成,序列内样本可能存在较高相似性;ISIC 2017 含 2 000 张训练、150 张验证、600 张测试,用于皮肤病变分割对比实验;LUNA16 数据集用于肺部分割任务,包含来自 888 次 CT 扫描的 1 186 个肺结节注释,本文按 8 : 2 比例随机划分训练集和测试集,所有图像尺寸统一调整为 352×352 。

本文采用的评价指标为平均 Dice 系数 (mean Dice coefficient, mDice)、平均交并比 (mean intersection over

union, mIoU)、浮点运算数 (floating-point operations, FLOPs) 和准确率 (accuracy, Acc)。mDice、mIoU 和 Acc 的值域均为 $[0, 1]$, 数值越大表示分割性能越好。FLOPs 衡量模型的计算复杂度,数值越小表示模型效率越高。为了验证网络泛化能力,设计了交叉验证实验,在 Kvasir-SEG 数据集上训练,在 CVC-ClinicDB 上测试。同时,在肺分割数据集上对网络进行可视化实验,检测网络各模块的功能。在临床应用中,这种轻微的边缘偏差对肺结节检测等下游任务的影响有限,因为医生更关注肺实质的整体完整性而非精确的像素级边界;同时,DFB-UNet 生成的分割轮廓符合解剖结构,具有良好的可解释性,便于临床医生直观理解与快速复核。最终,针对所设计模块进行了相应的消融实验,以验证每个模块的有效性。

2.1 息肉分割任务性能比较

表 1 为本文的网络与 10 种典型网络在 Kvasir-SEG 数据集上的性能比较结果。可以看出,DFB-UNet 在 mDice 与 mIoU 指标上低于 SAM-MaGuP 网络,但都高于其他网络,同时在 352×352 输入下的浮点运算量较低,表明 DFB-UNet 网络在计算复杂度和分割精度之间取得了优越的平衡。在实际临床中,在有限的图形处理器 (graphics processing unit, GPU) 内存下可以处理更大的图片,为模型的远程医疗提供了可能。

表 1 在 Kvasir-SEG 数据集上的分割结果

Table 1 The segmentation results on the Kvasir-SEG dataset

方法	mDice	mIoU	GFLOPs
U-Net	0.60	0.47	25.965
Res-UNet	0.69	0.57	—
SFA	0.72	0.61	—
U-Net++	0.82	0.74	65.925
DCRNet	0.88	0.82	—
FANet	0.88	0.81	—
MSRF-Net	0.86	0.79	—
DeepLabV3+	0.88	0.82	—
SAM-MaGuP	0.95	0.89	—
DFB-UNet	0.89	0.83	10.213

在 CVC-ClinicDB 数据集上进行的对比实验进一步验证了网络的性能。如表 2 所示,在 mDice 和 mIoU 指标上低于 DUCK-Net^[36],但仍高于其他网络,表明 DFB-UNet 网络仍然具有一定的竞争性。

2.2 皮肤病变分割任务性能比较

DFB-UNet 在皮肤病变分割数据集 ISIC 2017 上的实验结果如表 3 所示。本文使用 IoU、Dice 和 Acc 作为评估指标来评价网络性能。与同样应用 Transformer 的方法

TransFuse 相比,本文的网络在 IoU 评分上提高了 2.1%。在 Dice 评分上比 MALUNet 高了 0.8%。与 MALUNet 相比,本文在 Dice 上提升 0.8%。此外,MUCM-Net^[37]在该数据集上取得更高指标,但 DFB-UNet 通过高效协同局部与全局特征,在精度与复杂度间实现了良好平衡,验证了该并行设计在皮肤病变分割中的稳健性。

表 2 在 CVC-ClinicDB 数据集上的定量结果
Table 2 Quantitative results on the CVC ClinicDB dataset

方法	mDice	mIoU
SFA	0.700	0.607
ResUNet-mod	0.779	0.456
U-Net	0.794	0.729
DUCK-Net	0.948	0.901
DFB-UNet	0.810	0.723

表 3 所提出框架在 ISIC 2017 数据集上的效果
Table 3 The segmentation results on the ISIC 2017 dataset

方法	Backbone	IoU	Dice	Acc
CDNN	-	0.765	0.849	0.934
DDN	ResNet-18	0.765	0.866	0.939
FrCN	VGG16	0.771	0.871	0.940
UTNetV2	-	0.774	0.872	0.958
UNet++	ResNet-34	0.775	0.858	0.938
DCL-PSI	ResNet-101	0.777	0.857	0.941
SLSDeep	ResNet-50	0.782	0.878	0.936
UNeXt-S	-	0.783	0.878	0.960
MALUNet	-	0.789	0.881	0.962
TransFuse	R34+Deit-S	0.795	0.872	0.944
MUCM-Net	-	0.849	0.919	0.970
DFB-UNet	R34+Deit-S	0.816	0.889	0.965

2.3 肺分割任务性能比较

针对肺分割任务的对比网络较少,仅选取在相同设置下的两个网络与本文方法进行对比,如表 4 所示。由表 4 可见,本文方法在少数样本的边界区域仍存在轻微的分割和欠分割。为进一步直观展示分割效果,在 LUNA16 上选取 3 张代表性切片进行可视化对比,原始图像、标注、本文预测如图 5(a)~(c)所示。可以看出,DFB-UNet 能较完整地分割双肺区域,并在边界灰度过渡较弱或形态不规则的位置保持较好的轮廓贴合。

表 4 所提出框架在 LUNA16 上的效果
Table 4 The segmentation results on the LUNA16 dataset

方法	Dice	IoU
U-Net	0.965 8	0.933 9
RU-Net	0.963 8	0.930 1
DFB-UNet	0.966 0	0.934 2

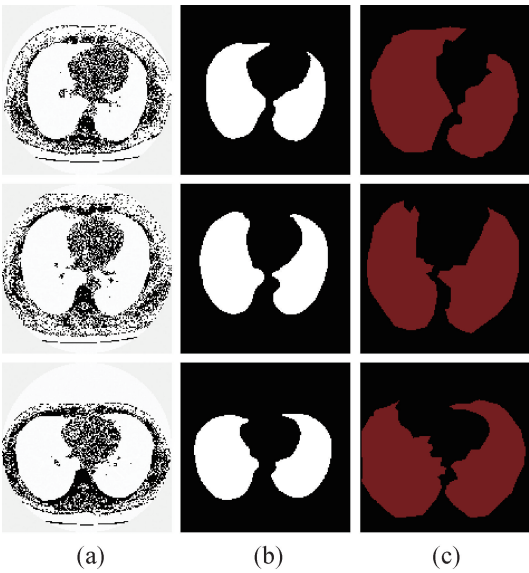


图 5 在 LUNA16 数据集上针对 DFB-UNet 的可视化结果
Fig. 5 Visualization for DFB UNet on the LUNA16 dataset

2.4 交叉数据集验证

交叉数据集验证以考察网络的泛化能力。将 DFB-UNet 在 Kvasir-SEG 数据集上训练,在 CVC-CLinicDB 上测试,验证结果如表 5 所示。本文的网络优于表中列方法。具体来说,本文网络相比于 HarDBet-MSEG 在两个指标上分别提高了 5.5%和 6.2%。而且,与 MSRF-Net 相比,DFB-UNet 在 mIoU 和 mDice 指标上的增量分别为 12.7%和 12.6%。交叉验证结果表明,DFB-UNet 在内镜息肉分割场景下具有一定的跨数据集泛化能力。

表 5 交叉数据集的定量结果
Table 5 Quantitative results of the cross dataset

方法	Dice	IoU
U-Net	0.634	0.543
ColonSegNet	0.613	0.509
ResUNet	0.336	0.243
MSRF-Net	0.624	0.542
DFB-UNet	0.751	0.668

2.5 消融实验

为了验证所设计的每个模块的有效性,本文在 ISIC 2017 数据集上设计了消融实验。主要针对 3 个方面,添加可变形卷积对网络性能的影响,注意融合模块的主要部件 CBAM 的改进是否提高了网络效率以及设计的双融合分支是否有效。评估指标采用 IoU、Dice 以及 Acc。针对可变形卷积实验结果如表 6 所示,可以看出,不加入可变形卷积的网络 D-1 性能不如添加了可变形卷积的网络。加入可变形卷积后,评估指标大约能提升 0.5%。本文猜测是由于可变形卷积的不规则感受野的

特殊性有利于同样目标边界也不规则的医学图像分割任务。

表 6 可变形卷积对网络的影响

方法	Add DCN	IoU	Dice	Acc
D-1	×	0. 811	0. 883	0. 963
DFB-UNet	√	0. 816	0. 889	0. 965

考虑改进的卷积块注意模块 PCBAM 对网络性能的影响,本文将原始的 CBAM 以及 PCBAM 分别作为 AFM 的主要部件,其中 D-2 表示应用原始 CBAM,对比结果如表 7 所示。从表 7 可以看出,本文网络的结果优于 D-2,在 Dice 指标上提升 0. 5%。表明 PCBAM 对于网络性能起到促进作用。说明给每个通道自适应加权的操作是有效的。

表 7 AFM 中 CBAM 改进的影响

方法	AFM Main Components	IoU	Dice	Acc
D-2	CBAM	0. 812	0. 884	0. 965
DFB-UNet	PCBAM	0. 816	0. 889	0. 965

针对双融合分支对网络的消融实验如表 8 所示。D-3 网络使用单融合分支,仅使用注意融合分支,不再经过卷积融合分支融合两分支特征。结果表明,结合了卷积和注意力机制的双融合分支有益于提升网络的性能,能多方面融合来自编码器两分支的输出特征。

表 8 双融合分支的影响

方法	Composition	Fusion	IoU	Dice	Acc
D-3	Parallel	AFM	0. 808	0. 881	0. 963
DFB-UNet	Parallel	AFM+CFM	0. 816	0. 889	0. 965

3 结 论

本文设计了一种新的医学图像分割网络,该网络采用双融合的卷积和自注意力机制并行方法。编码器结构是依赖 CNN 和 Transformer 并行的双分支结构,确保从局部和全局两个角度提取特征。同时,由于可变形卷积的卷积核具有不规则感受野,而且医学图像的目标病灶通常也是不规则的。将两分支的输出特征进行融合之前,经过可变形卷积以加强边界信息。融合分支由卷积融合分支以及注意融合分支并行构成,从不同角度融合网络特征。针对本文选取的内镜息肉、皮肤病变与 CT 肺部 3 类分割任务的数值实验结果显示本文的设计能够有效提

升网络性能。尤其在 ISIC2017 数据集上各项指标获得更好的性能。在 Kvasir-SEG 与 CVC-ClinicDB 之间的交叉数据集实验表明网络在内镜息肉分割场景下具有一定的跨数据集泛化能力。消融实验也证明了所提出各模块对网络性能提升具有贡献。

参考文献

[1] 钟翼龙, 张晓凤, 刘伟斌, 等. 填充材料 CT 图像中脱粘缺陷识别与深度测量[J]. 电子测量与仪器学报, 2024, 38(1): 178-186.
ZHONG Y L, ZHANG X F, LIU Y B, et al. Identification and depth measurement of debonding defects in CT images of packing materials[J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(1): 178-186.

[2] 薛钟毫, 姜金刚, 孙健鹏, 等. 基于多尺度卷积和选择性核双注意力机制的半监督全景 X 射线图像龋齿分割[J]. 仪器仪表学报, 2025, 46(6): 241-250.
XUE ZH H, JIANG J G, SUN J P, et al. Semi-supervised caries segmentation of panoramic X-ray images based on multi-scale convolution and selective kernel dual-attention [J]. Chinese Journal of Scientific Instrument, 2025, 46(6): 241-250.

[3] 龚瑾儒, 翟锦涛, 田峰, 等. 融合注意力和状态空间的儿科超声心动图分割[J]. 电子测量与仪器学报, 2025, 39(10): 32-40.
GONG J R, ZHAI J T, TIAN F, et al. Pediatric echocardiography segmentation combining attention and state space[J]. Journal of Electronic Measurement and Instrumentation, 2025, 39(10): 32-40.

[4] 张建, 宋志刚, 王书浩, 等. 基于胃部肿瘤病理数据特征提取的分型模型研究[J]. 仪器仪表学报, 2024, 45(7): 210-217.
ZHANG J, SONG ZH G, WANG SH H, et al. Research on feature classification model based on pathological data of gastric tumor [J]. Chinese Journal of Scientific Instrument, 2024, 45(7): 210-217.

[5] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[J]. International Conference on Medical Image Computing and Computer-assisted Intervention, 2015, 9351: 234-241.

[6] ZUNAIR H, HAMZA A B. Sharp U-Net: Depthwise convolutional network for biomedical image segmentation[J]. Computers in Biology and Medicine, 2021, 136: 104699.

[7] DOC Y Z, DOC S W, MRS Z H, et al. DualA-Net: A generalizable and adaptive network with dual-branch encoder for medical image segmentation[J]. Computer

- Methods and Programs in Biomedicine, 2024, 243: 107877.
- [8] JHA D, RIEGLER M A, JOHANSEN D, et al. Double U-Net: A deep convolutional neural network for medical image seg-mentation [C]. Computer-based Medical Systems (CBMS). IEEE, 2020: 558-564.
- [9] 胡冠华, 张永雷, 申立群. 基于改进 U-Net 的轻量级输电线分割算法[J]. 电子测量与仪器学报, 2024, 38(1): 211-218.
- HU G H, ZHANG Y L, SHEN L Q. Lightweight transmission line conductor segmentation algorithm with improved U-Net[J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(1): 211-218.
- [10] XING W, ZHU Z, HOU D, et al. CM-SegNet: A deep learning-based automatic segmentation approach for medical images by combining convolution and multilayer perceptron [J]. Computers in Biology and Medicine, 2022, 147: 105797.
- [11] SUN Q, DAI M Y, LAN Z Y, et al. UCR-Net: U-shaped context residual network for medical image segmentation. [J]. Computers in Biology and Medicine, 2022, 151: 106203.
- [12] 于天河, 徐博超, 侯善冲, 等. 基于图像的道路裂缝分割及量化方法研究 [J]. 仪器仪表学报, 2024, 45(9): 77-91.
- YU T H, XU B CH, HOU SH CH, et al. Research on image-based segmentation and quantification of road cracks [J]. Chinese Journal of Scientific Instrument, 2024, 45(9): 77-91.
- [13] DING X, ZHANG X, HAN J, et al. Scaling up your kernels to 31x31: Revisiting large kernel design in CNNs [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2022: 11963-11975.
- [14] TANG F, DING J, WANG L, et al. CMUNeXt: An efficient medical image segmentation network based on large kernel and skip fusion [J]. ArXiv preprint arXiv: 2308.01239, 2023.
- [15] LEE H H, BAO S, HUO Y, et al. 3D UX-Net: A large kernel volumetric convnet modernizing hierarchical transformer for medical image segmentation [J]. ArXiv preprint arXiv:2209.15076, 2022.
- [16] LAN L B, CAI P Z, JIANG L, et al. BRAU-Net++: U-Shaped hybrid cnn-transformer network for medical image segmentation [J]. ArXiv preprint arXiv: 2401.00722, 2024.
- [17] PENG Y, SONKA M, CHEN D Z. U-Net v2: Rethinking the skip connections of u-net for medical image segmentation [J]. ArXiv preprint arXiv: 2311.17791, 2023.
- [18] ZHANG C Y, SUN S B, HU W M, et al. FDR-TransUNet: A novel encoder-decoder architecture with vision transformer for improved medical image segmentation [J]. Computers in Biology and Medicine, 2024, 169: 107858.
- [19] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [J]. Advances in Neural Information Processing Systems, 2017, 30: 5998-6008.
- [20] 梁礼明, 王成斌, 康婷, 等. 多层次语义融合与特征耦合 Transformer 的视网膜血管分割 [J]. 电子测量与仪器学报, 2025, 39(12): 155-166.
- LIANG L M, WANG CH B, KANG T, et al. Multi-level semantic fusion and feature-coupled Transformer for retinal vessel segmentation [J]. Journal of Electronic Measurement and Instrumentation, 2025, 39 (12): 155-166.
- [21] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: Transformers for image recognition at scale [J]. ArXiv preprint arXiv: 2010.11929, 2020.
- [22] LIU Z, YAO Y, MA N, et al. Swin transformer: Hierarchical vision transformer using shifted windows [J]. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2021: 10012-10022.
- [23] FU Y H, LIU J F, SHI J. TSCA-Net: Transformer based spatial-channel attention segmentation network for medical images [J]. Computers in Biology and Medicine, 2024, 170: 107938.
- [24] 陈广秋, 任天蓉, 段锦, 等. 基于双分支多尺度特征融合的跨模态语义分割算法 [J]. 电子测量与仪器学报, 2025, 39(5): 144-154.
- CHEN G Q, REN T R, DUAN J, et al. Cross-modal semantic segmentation algorithm based on dual-branch multi-scale feature fusion [J]. Journal of Electronic Measurement and Instrumentation, 2025, 39(5): 144-154.
- [25] 颜文琴, 郑秀娟, 殷中平, 等. 融合面部外观与生理表征的作业人员疲劳判别 [J]. 电子测量与仪器学报, 2025, 39(10): 12-21,.
- YAN W Q, ZHENG X J, YIN ZH P, et al. Fatigue identification of workers by integrating facial appearance and physiological characteristics [J]. Journal of Electronic Measurement and Instrumentation, 2025, 39(10): 12-21.
- [26] FU B K, PENG Y S, HE J J, et al. HmsU-Net: A hybrid multi-scale u-net based on a cnn and transformer for medical image segmentation [J]. Computers in

Biology and Medicine, 2024, 170: 108013.

[27] ZHANG W, CHEN S, MA Y, et al. ETUNet: Exploring efficient transformer enhanced UNet for 3D brain tumor segmentation[J]. Computers in Biology and Medicine, 2024, 171: 108005.

[28] TANG F H, NIAN B K, DING J R, et al. MobileUtr: Revisiting the relationship between light-weight CNN and Transformer for efficient medical image segmentation[J]. ArXiv preprint arXiv:2312.01740, 2023.

[29] FU Z J, LI J J, HUA Z. DEAU-Net: Attention networks based on dual encoder for Medical Image Segmentation[J]. Computers in Biology and Medicine, 2022, 150: 106197.

[30] YU A, WEI L S, BAI J et al. MS-TCNet: An effective transformer-cnn combined network using multi-scale feature learning for 3d medical image segmentation[J]. Computers in Biology and Medicine, 2024, 170: 108057.

[31] LI K, WANG Y, GAO P, et al. UniFormer: Unified transformer for efficient spatiotemporal representation learning[J]. ArXiv preprint arXiv:2201.04676, 2022.

[32] DAI J, QI H, XIONG Y, et al. Deformable Convolutional Networks [C]. Proceedings of the IEEE International Conference on Computer Vision (ICCV). 2017: 764-773.

[33] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module[C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 3-19.

[34] HUANG Z, WANG X, HUANG L, et al. CCNet: Criss-cross attention for semantic segmentation [C]. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2019: 603-612.

[35] 靖永志, 倪胜, 贾兴科, 等. 基于列向语义分割的悬浮间隙视觉检测方法研究[J]. 仪器仪表学报, 2024, 45(9): 67-76.

JING Y Z, NI S, JIA X K, et al. Research on visual detection method of levitation gap based on column-oriented semantic segmentation [J]. Chinese Journal of Scientific Instrument, 2024, 45(9): 67-76.

[36] ETELEAZA D, CRACIUN C. Using DUCK-Net for polyp image segmentation [J]. Scientific Reports, 2023, 13(1): 9803.

[37] YUAN C, ZHAO D, AGALAN S S. MUCM-Net: A Mamba powered UCM-Net for skin lesion segmentation [J]. Exploration of Medicine, 2024, 5: 694-708.

作者简介



范晓冬(通信作者),2013 年于北京工业大学获得博士学位,现为辽宁工程技术大学副教授,主要研究方向为机器视觉。

E-mail: bhdxfxd@163.com

Fan Xiaodong (Corresponding author)

received his Ph. D. degree from Beijing University of Technology in 2013. He is currently an associate professor at Liaoning Technical University. His main research interest includes machine vision.



张楚浩,2023 年于辽宁工程技术大学获得学士学位,现为辽宁工程技术大学硕士研究生,主要研究方向为机器视觉。

E-mail: 135751599@qq.com

Zhang Chuha received his B. Sc.

degree from Liaoning Technical University in 2023. He is currently a M. Sc. candidate at Liaoning Technical University. His main research interest includes machine vision.



周婧,2024 年于渤海大学获得硕士学位,主要研究方向为深度学习、计算机视觉与图像处理。

E-mail:2567235749@qq.com

Zhou Jing received her M. Sc. degree

from Bohai University in 2024. Her main research interests include deep learning, computer vision, and image processing.



侯利民,2010 年于东北大学获得博士学位,现为辽宁工程技术大学教授,主要研究方向为滑模控制与永磁同步电机调速控制。

E-mail:764144847@qq.com

Hou Limin received his Ph. D. degree

from Northeastern University in 2010. He is currently a professor at Liaoning Technical University. His main research interests include sliding mode control and speed control of permanent magnet synchronous motors.



韩莹,2020 年于东北大学获得博士学位,现为辽宁工程技术大学副教授,主要研究方向为时间序列分析与应用、复杂系统预测等。

E-mail:hyfengyan@163.com

Han Ying received her Ph. D. degree

from Northeastern University in 2020. She is currently an associate professor at Liaoning Technical University. Her main research interests include time series analysis and application, complex system prediction, etc.