

结合局部、全局感知与语义流对齐的脑白质高信号分割方法

张新峰, 郭依海, 刘晓民, 许忠贺, 李相生

引用本文

张新峰, 郭依海, 刘晓民, 许忠贺, 李相生. [结合局部、全局感知与语义流对齐的脑白质高信号分割方法](#)[J]. 计算机科学, 2026, 53(4): 291-298.

ZHANG Xinfeng, GUO Yihai, LIU Xiaomin, XU Zhonghe, LI Xiangsheng. [White Matter High Signal Segmentation Method Combining Local and Global Perception and Semantic Flow Alignment](#) [J]. Computer Science, 2026, 53(4): 291-298.

相似文献推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[融合ByteTrack的EAP-YOLOv8无人机Marker点检测与追踪](#)

Integrate ByteTrack's EAP-YOLOv8 UAV Marker Point Detection and Tracking
计算机科学, 2026, 53(3): 266-276. <https://doi.org/10.11896/jsjcx.241100115>

[基于文本-图像多模态融合的变电所布局图纸图符检测方法](#)

Method for Symbol Detection in Substation Layout Diagrams Based on Text-Image Multimodal Fusion
计算机科学, 2026, 53(1): 206-215. <https://doi.org/10.11896/jsjcx.250200090>

[基于深度学习的OCT/OCTA视网膜图像分析方法综述](#)

Review of Retinal Image Analysis Methods for OCT/OCTA Based on Deep Learning
计算机科学, 2026, 53(1): 128-140. <https://doi.org/10.11896/jsjcx.241100047>

[频域纹理先验与特征增强的医学图像分割模型](#)

Medical Image Segmentation Model Based on Frequency Texture Prior and Frequency Feature Enhancement Fusion
计算机科学, 2025, 52(11A): 241200125-8. <https://doi.org/10.11896/jsjcx.241200125>

[基于特征提取增强和金字塔结构的实时Transformer小目标检测模型](#)

Real-time Transformer Small Target Detection Model Based on Feature Extraction Enhancement and Pyramid Structure
计算机科学, 2025, 52(11A): 250100139-11. <https://doi.org/10.11896/jsjcx.250100139>

结合局部、全局感知与语义流对齐的脑白质高信号分割方法

张新峰¹ 郭依海¹ 刘晓民¹ 许忠贺¹ 李相生²

¹ 北京工业大学信息科学技术学院 北京 100124

² 中国人民解放军空军特色医学中心影像科 北京 100142

(zxf@bjut.edu.cn)

摘要 针对脑白质高信号目标小的特点,提出一种结合局部、全局感知与语义流对齐的脑白质信号分割方法 PGF-Net。首先,提出局部感知注意力模块(Patch Aware Attention, PAA),通过划分局部小图像块进行特征选择的方法,加强局部特征提取能力;然后,提出结合局部和全局感知的注意力模块(Patch Global Aware Attention, PGAA),利用 Transformer 全局感知的特点建立长程依赖;最后,提出门控语义流对齐模块(Gated Flow Alignment Module GFAM),在解码部分预测语义流偏移场,引导解码器中的高层特征扩张,实现与编码器对应低层特征的精准对齐融合。实验结果表明,PGF-Net 在自采数据集中,交并比(mIoU)达到 0.8769, Dice 系数为 0.8423,豪斯多夫距离(HD)降至 32.61,平均表面距离(ASD)仅为 1.7,达到了最优效果;在两种小目标公开数据集上也达到最优效果,验证了其泛化性和鲁棒性。此方法在辅助医生诊断方面具有一定的应用前景。

关键词: 图像分割;小目标;局部感知;全局感知;语义流对齐

中图分类号 TP391

White Matter High Signal Segmentation Method Combining Local and Global Perception and Semantic Flow Alignment

ZHANG Xinfeng¹, GUO Yihai¹, LIU Xiaomin¹, XU Zhonghe¹ and LI Xiangsheng²

¹ College of Information Science and Technology, Beijing University of Technology, Beijing 100124, China

² Department of Radiology, Air Force Medical Center, PLA, Beijing 100142, China

Abstract A white matter hyperintensity segmentation method called PGF-Net is proposed, which combines local and global perception with semantic flow alignment, to address the characteristic of small targets in high signal white matter. Firstly, it proposes the PAA (Patch Aware Attention) module, which enhances the ability to extract local features by dividing local small image blocks for feature selection. Secondly, it proposes to combine local and global aware attention modules (PGAA) and utilizes the characteristics of Transformer global perception to establish long-range dependencies. Lastly, it proposes a gated flow alignment module (GFAM) to predict the semantic flow offset field in the decoding section. Guide the expansion of high-level features in the decoder to achieve precise alignment and fusion with the corresponding low-level features in the encoder. Experimental results show that the PGF-Net achieves optimal performance in a self collected dataset, with a cross union ratio (mIoU) of 0.8769, a Dice coefficient of 0.8423, a Hausdorff distance (HD) of 32.61, and an average surface distance (ASD) of only 1.7. The model also achieves optimal performance on two small target public datasets, verifying its generalization and robustness. This method has certain application prospects in assisting doctors in diagnosis in the future.

Keywords Image segmentation, Small target, Local perception, Global perception, Semantic flow alignment

1 引言

脑白质高信号(White Matter Hyperintensity, WMH)是脑小血管疾病的重要影像学标志,常见于中老年人群的脑部磁共振成像(MRI)检查中,在 T2 加权像和液体衰减反转恢复 (FLAIR) 序列上呈现为高信号(亮区)。根据解剖位置,WMHs 可分为脑室旁白质高信号(Periventricular WMH, PWMH)和深部白质高信号(Deep WMH, DWMH),其形态多表现为点状或斑片状病灶。研究表明,脑白质高信号的进

展可能与认知功能下降、脑萎缩等神经系统退行性改变密切相关^[1]。分割 WMH 的体积、分布(如深部额叶和脑室周围)以及形态特征(如斑点状或融合性),有助于区分不同病因(如高血压性动脉硬化与脑淀粉样血管病),并为临床诊断提供客观依据^[2-3]。

脑白质高信号作为脑小血管疾病的重要影像学标志,在临床诊断中给影像科医生带来了巨大负担。医生需要逐层仔细阅片,在 T2-FLAIR 序列上寻找那些分布广泛、形态各异的点状或斑片状高信号病灶,这种重复性高、注意力高度集中的

工作极易导致视觉疲劳和漏诊。而基于深度学习的自动检测技术有望突破这一瓶颈,通过精准识别病灶并提供定量分析,能显著提升诊断效率,让医生从繁重的阅片工作中解放出来。

随着卷积神经网络(Convolutional Neural Network, CNN)的兴起,U-Net等编码器-解码器结构在医学图像分割任务中取得了显著进展。然而,CNN在建模长距离依赖关系方面存在局限,Transformer的引入很大程度上弥补了这一不足。在此背景下,Farkhani等^[4]提出了一种概率学辅助分支的改进TransUnet进行脑白质分割,通过概率辅助模块提高对医学图像中模糊目标的分类准确率,使用Transformer^[5]结构处理在卷积编码器提取的高层特征,显式建模卷积编码器不同层级特征间的长程依赖关系。Lee等^[6]在UNet中添加了SE模块,通过通道结果加权,抑制噪声通道(如背景干扰),增强关键特征通道(如纹理、形状),使Dice系数达到0.769,IoU达到0.633。Park等^[7]提出了一种用于WMH分割的具有多尺度(HF)的U-Net。该方法的Dice系数达到0.81,F1值到达0.79。Zhang等^[8]提出了结合多尺度与注意力机制的脑组织分割方法,同时建模全局上下文与局部细节,实现多尺度深层语义特征与浅层细节特征的高效融合,从而实现纹状体结构的精准分割。Ghafoorian等^[9]提出的将位置信息添加到卷积之后的全连接层的方法,实现了语义信息与位置信息的融合,使Dice系数达到了0.791。

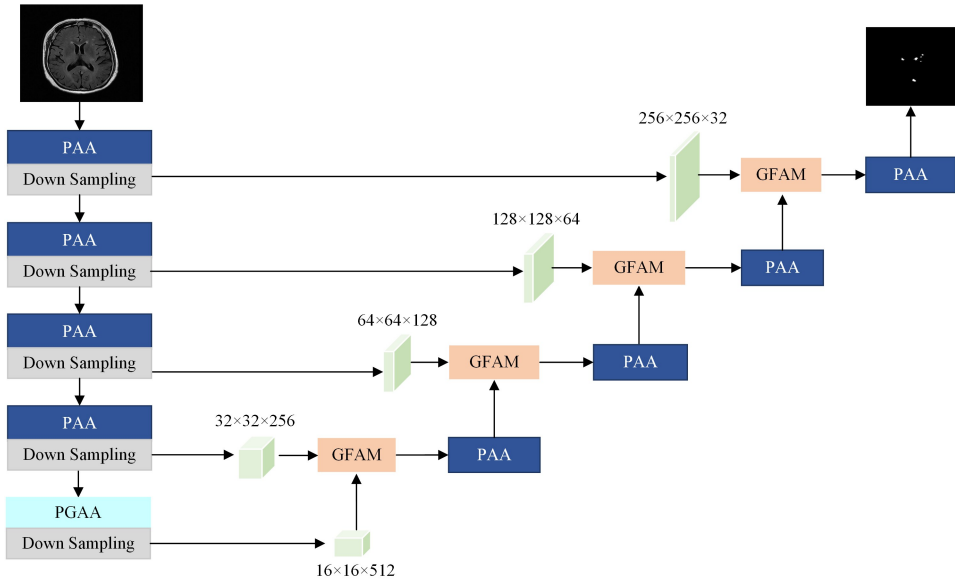


图1 模型的整体框架

Fig. 1 Overall framework of the proposed model

2.2 局部感知注意力模块 PAA

传统CNN因逐层下采样,导致特征图分辨率降低,高频细节丢失,在小目标检测场景中尤为显著:一方面,深层特征的感受野扩大,使得小目标的有效特征仅存个位数像素;另一方面,固定尺寸卷积核难以动态适应多尺度目标,造成小目标局部语义信息与空间上下文的关联被弱化。为解决上述问题,受到Xu等研究^[10]的启发,提出局部感知模块PAA。

如图2所示,PAA模块共分为两条支路,将两条支路的结果融合后,经过注意力机制得到输出。

具体步骤如下:

尽管这些方法在WMH分割任务上取得了显著效果,但由于脑白质高信号通常呈现体积小、分布稀疏的特点,现有方法仍面临小目标易丢失和边界模糊难区分的挑战。因此,增强模型对小目标的识别能力,并精细化边界细节,成为进一步提升分割精度的关键。针对这一问题,本文提出了PGF-UNet,在U型结构网络上,增加了局部感知注意力模块(PAA),以提高对小目标物体的特征提取能力;在瓶颈结构引入结合局部与全局感知的注意力模块(PGAA);最后在上采样阶段加入门控语义流对齐模块(GFAM),精细还原小目标边界细节。

2 针对小目标的分割模型

2.1 模型整体结构

模型整体结构如图1所示。编码器包括5个阶段 $Stage_i$ ($i=1,2,3,4,5$),第1-4每个阶段包括PAA模块和下采样,第5阶段为PGAA模块和下采样操作,输入图片经过4个编码阶段分别生成隐藏特征 $x_i \in \mathbb{R}^{\frac{D}{2^i} \times \frac{W}{2^i} \times 2^{i-1} C}$ ($D=W=512, C=16, i=1,2,3,4,5$)。解码器每个阶段包含GFAM模块和PAA模块。解码器对编码后得到的深层特征 $x_5 \in \mathbb{R}^{\frac{D}{32} \times \frac{W}{32} \times 16}$ 在每个阶段里通过GFAM模块与隐藏特征融合,并在PAA模块做进一步的特征提取,得到分割掩码。

1)第一条支路是经过PW卷积后,通过多次 3×3 卷积提取特征,并将每次的特征层通过跳跃连接。该条支路的目的是以更少的参数扩大感受野,捕获多尺度的空间特征。

2)第二条支路PA是重点提取局部特征:将特征图按照 2×2 大小进行图像块划分 $F \in \mathbb{R}^{\frac{HW}{4} \times 2 \times 2 \times C}$,然后将每个图像块通过通道平均降维为Token向量 $T \in \mathbb{R}^{\frac{HW}{4} \times 2 \times 2}$,最后使用FFN^[11]将局部特征映射到高维空间 $T \in \mathbb{R}^{\frac{HW}{4} \times C}$,并通过SoftMax进行注意力加权。

$$FFN(x) = LayerNorm(xW_1 + b_1)W_2 + b_2 \quad (1)$$

3) 在 Token 与通道层面,引入一种基于可学习参数的特征选择机制^[12]。具体来说,通过标准正态分布初始化一个可学习向量 ξ 和可学习矩阵 P 。对于所有 Token t_i ,计算其与向量 ξ 的余弦相似度 $\text{sim}(t_i, \xi)$,并将结果与 P 做矩阵乘法。输出结果如式(2)所示:

$$\hat{t}_i = P \cdot \text{sim}(t_i, \xi) \cdot t_i \quad (2)$$

其中, $P \in \mathbb{R}^{C \times C}$; $\xi \in \mathbb{R}^C$; $\text{sim}()$ 为余弦相似度函数,取值为 $0 \sim 1$ 。

在训练过程中,损失函数反向传播梯度到 ξ 向量和 P 矩阵, ξ 动态评估 Token 层面对任务的贡献度, P 矩阵在通道层面评估贡献度。Token t_i 首先与 ξ 计算余弦相似度,随后将该相似度值与 Token 相乘,实现 Token 选择;其次与 P 矩阵进行矩阵乘法,实现 Token 内部通道的选择。接下来通过 Reshape、上采样操作将 Token 维度转成与第一条支路一样的特征图维度 $F \in \mathbb{R}^{H \times W \times C}$,再通过卷积操作进一步融合特征。

4) 将以上两个支路的输出结果进行相加融合,再通过通道注意力机制^[13]和空间注意力机制^[14]得到最终的输出。具体的操作步骤如下。

(1)通道注意力:给定输入特征 $f_i \in \mathbb{R}^{H \times W \times C}$,将其沿空间维度通过全局平均池化压缩,生成通道统计向量 $M_C \in \mathbb{R}^{C \times 1 \times 1}$,再通过一维卷积和 Sigmoid 操作(见式(3)),得到通道注意力图。

$$M_c(F) = \sigma(\text{Conv1D}(\text{AvgPool}(F))) \quad (3)$$

其中, σ 是 Sigmoid 操作。

(2)空间注意力:通过平均池化操作和最大池化操作,得到两个二维图,将两个矩阵进行拼接、卷积和 Sigmoid 操作,生成空间注意力图。

$$M_s(F) = \sigma(f^{7 \times 7}([\text{AvgPool}(F); \text{MaxPool}(F)])) \quad (4)$$

其中, $f^{7 \times 7}$ 是卷积, $[\]$ 是拼接操作。

依次输入特征 F 与通道注意力图和空间注意力图,并依次执行矩阵点乘,得到的输出特征 F' 同时具备通道和空间维度的自适应增强能力,能够更聚焦于关键区域,提升模型的分割精度。

$$F' = M_c(F) \otimes F \quad (5)$$

$$F'' = M_s(F') \otimes F' \quad (6)$$

其中, \otimes 表矩阵点乘操作。

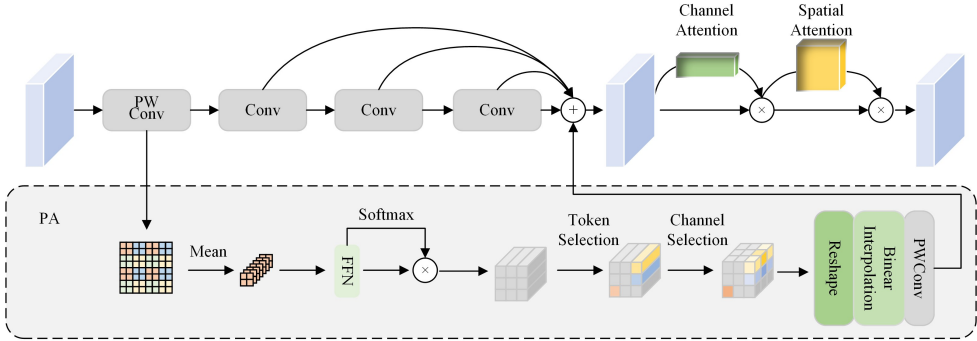


图2 PAA 模块结构

Fig. 2 Structure of PAA module

2.3 局部和全局感知注意力模块 PGAA

为解决小目标因缺乏上下文信息而被漏检的问题,本文提出了 PGAA 模块。如图 3 所示,该模块由 3 条支路并联组

成:PGA 支路、多卷积支路和 PA 支路。局部支路 PA 聚焦小目标细节,全局支路 PGA 提供语义上下文,两者通过并行相加实现多层次特征融合,提升检测的鲁棒性。

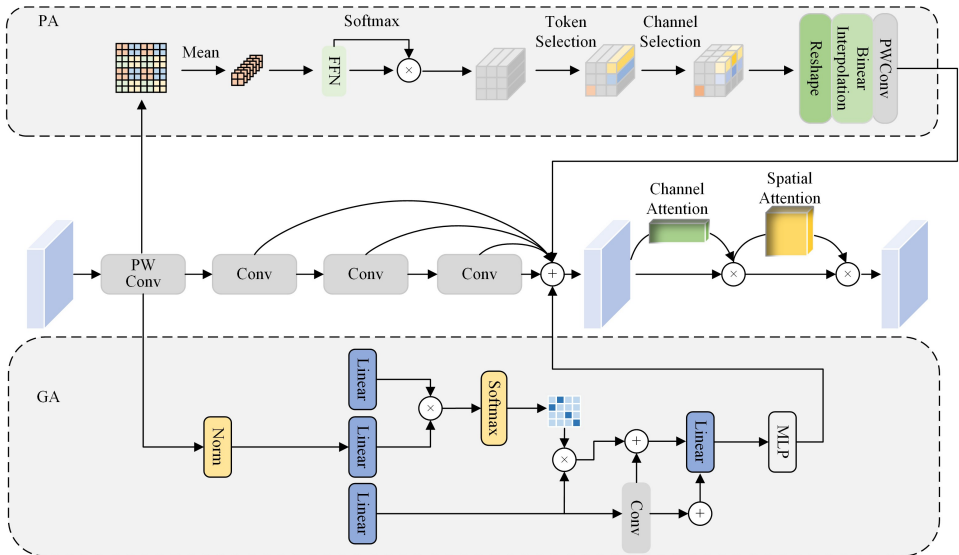


图3 PGAA 模块结构

Fig. 3 Structure of PGAA module

图 3 中,新增加的 GA 支路全局上下文建模的主要作用是捕获大范围上下文信息。具体来说,其利用 Transformer 的自注意力机制(Self-Attention)直接建模图像中任意像素或区域间的交互,突破了 CNN 局部感受野的限制;利用全局依赖补充小目标区域的上下文信息,增强特征表示,弥补了 PA 支路过度聚焦局部可能丢失的全局语义。

GA 先用可学习矩阵乘以输入特征图,得到 3 个矩阵 $\mathbf{Q}, \mathbf{K}, \mathbf{V}$, 将其划分窗口,通过循环位移进行移动窗口注意力^[15]计算。

$$\text{Attention}(\mathbf{Q}_i, \mathbf{K}_i, \mathbf{V}_i) = \text{Softmax}\left(\frac{\mathbf{Q}_i \mathbf{K}_i^T}{\sqrt{d}} + \mathbf{B}\right) \mathbf{V}_i \quad (7)$$

其中, \mathbf{B} 为相对位置偏置。

接下来,对矩阵 \mathbf{V}_i 进行卷积操作,以利用卷积高效提取局

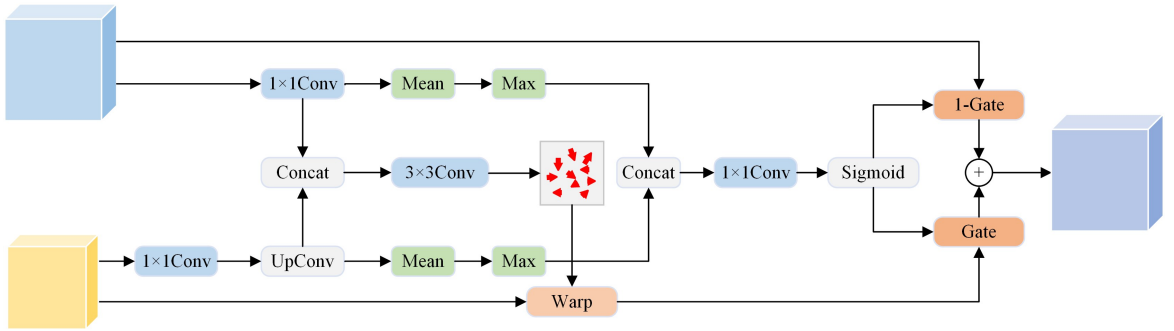


图 4 GFAM 模块结构

Fig. 4 Structure of GFAM module

具体来说,两个相邻特征图首先通过两个 1×1 卷积层压缩成相同的通道深度。设高层特征为 \mathbf{F}_l , 低层特征为 \mathbf{F}_{l-1} , 通过双线性插值将 \mathbf{F}_l 采样到与 \mathbf{F}_{l-1} 相同的大小,然后将二者拼接,并通过一个 3×3 的卷积层,得到语义流场的预测。

$$\Delta_{l-1} = \text{Conv}_l(\text{Concat}(\mathbf{F}_l, \mathbf{F}_{l-1})) \quad (8)$$

计算完 Δ_{l-1} 后,空间网格 Δ_{l-1} 上的每个位置上的偏移量通过加法运算映射到高层特征 \mathbf{F}_l 上的点 p 。最后,使用可微双线性采样^[17]得到尺寸扩展结果。

$$\mathbf{F}_l' = \sum_{p \in N(p_l)} \mathbf{w}_p \mathbf{F}_l(p) \quad (9)$$

其中, $N(p_l)$ 是变形后坐标 p_l 的 4 个最近邻点(左上、右上、左下、右下); \mathbf{w}_p 是根据 p_l 与邻域点的距离计算的双线性权重,满足 $\sum_p \mathbf{w}_p = 1$ 。

除此以外,使用门控机制来突出两个对齐特征上最重要的区域。在两个特征图上分别采用最大池化和平均池化,然后将全部 4 个图进行拼接,以生成可学习门控图。

$$\Delta_G = \text{conv}(\text{cat}(\text{Avepool}(\mathbf{F}_l, \mathbf{F}_{l-1}), \text{Maxpool}(\mathbf{F}_l, \mathbf{F}_{l-1}))) \quad (10)$$

最后,在融合相邻特征的过程中,采用 Δ_G 对高层特征加权,采用 $1 - \Delta_G$ 对低层特征加权,得到最终的输出结果。

3 实验及结果分析

3.1 数据集

本节使用的数据集包括自采数据集、IDRiD 数据集^[18] 和 WHM 分割挑战赛数据集^[19]。其中, IDRiD 数据集和

部特征的特性弥补细节信息。最后,将注意力输出和卷积结果融合,得到最终输出。

2.4 门控语义流对齐模块 GFAM

在传统的 UNet 网络解码器中,高层特征通过双线性插值上采样扩展特征图尺寸并与低层特征融合,这种方式无法处理下采样或残差连接导致的空间错位问题,导致语义信息的细节丢失,尤其是对于像素占比极低的小目标,可能导致严重的边界识别偏移,还会出现漏检现象^[16]。受到 Li 等研究^[16] 的启发,本文提出 GFAM 模块代替普通的上采样和跳跃连接部分。如图 4 所示,GFAM 模块可以通过预测偏移场,动态调整高层特征每个像素的位置,之后对语义流校准后的特征图进行双线性插值操作,使其与低层特征在空间上精确对齐。

WHMS 挑战赛数据集作为泛化性实验数据。

自采数据集来自某医院采集并由两位影像科医师标注的,部分示例如图 5 所示。共 740 张图片,其中 512 张作为训练集,128 张作为验证集,100 张作为测试集。



图 5 自采数据集示例图像

Fig. 5 Examples of self collected dataset

IDRiD 数据集专注于糖尿病视网膜病变,本文选择对膜病变中的硬渗病灶进行分割实验。其包含 81 张大小为 4288×2848 的 DR 图像及渗出物的标注,其中 54 张用于训练,27 张用于测试,部分示例图像如图 6 所示。

WHM 分割挑战赛数据集同样为脑白质分割数据集,部分示例图像如图 7 所示。本文选取其中 960 张图片,并将其

按照 7:2:1 的比例划分为训练集、验证集和测试集。

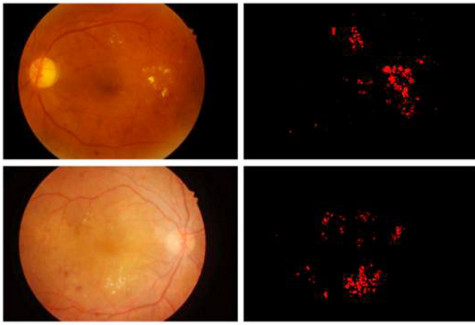


图 6 IDRiD 数据集示例图像

Fig. 6 Examples of IDRiD dataset

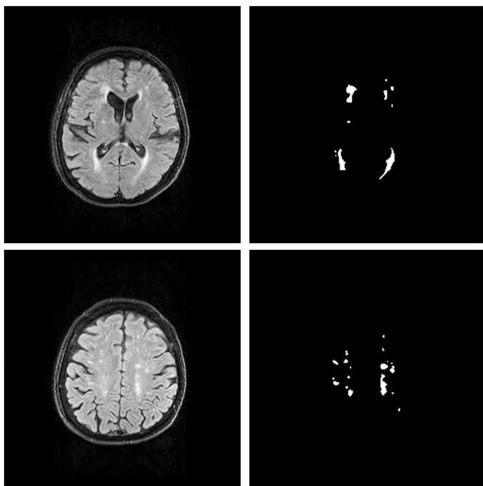


图 7 WHM 挑战赛数据集示例图像

Fig. 7 Examples of WHM segment challenge dataset

3.2 实验环境与参数

本文的实验环境如表 1 所列。

表 1 实验环境配置

Table 1 Experimental environment configuration

环境名称	环境参数
系统	Ubuntu 18.04 LTS
处理器	Intel(R) Core(TM) i9-11900
显卡	GeForce RTX 3080 Ti 12GB
编程语言	Python 3.9
深度学习框架	PyTorch 1.12.1

训练过程中,训练周期(Epoch)设为 2 400,批量大小(Batchsize)设置为 2,初始学习率设置为 0.01,使用 SGD 作为模型优化器,优化策略使用多项式衰减(Poly LR)策略。

3.3 评价指标

为了全面评估模型的分割性能,共使用 4 个评价指标,分别是交并比值(Mean Intersection over Union, MIoU)、Dice 系数(Dice Similarity Coefficient, DSC)、豪斯多夫距离^[20](Hausdorff Distance, HD)和平均表面距离(Average Surface Distance, ASD)。

MIoU 是语义分割任务中最常用的评估指标之一,通过计算预测区域与真实标注区域之间的交集与并集的比值,评估分割精度。

$$MIoU = \frac{1}{k+1} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij} + \sum_{j=0}^k p_{ji} - p_{ii}} \quad (11)$$

其中, $k+1$ 表示计算类别数, p_{ii} 表示第 i 类元素的实际数量, p_{ij} 表示将第 i 类元素预测为第 j 类元素的数量。MIoU 的特点是对类别不平衡数据敏感。

Dice 系数用于衡量预测分割结果与真实标注之间的相似性,定义为两倍交集面积与总像素数的比值。

$$Dice(A, B) = \frac{2|A \cap B|}{|A| + |B|} \quad (12)$$

Dice 系数的特点是对小目标分割敏感,常用于医学图像分析。

豪斯多夫距离用于量化两个点集之间的最大不匹配程度,反映分割边界的极端偏差。

$$HD(A, B) = \max(\max_{a \in A} \{\min_{b \in B} \|a - b\|\}, \max_{b \in B} \{\min_{a \in A} \|b - a\|\}) \quad (13)$$

其中, $\|a - b\|$ 为点 a 与 b 之间的欧氏距离。HD 的特点是捕捉最坏情况下的误差,适用于对边界精度要求严格的场景。

平均表面距离用于计算预测边界与真实边界间所有点的平均最小距离,提供整体边界偏差的平滑评估。

$$SD(A, B) = \frac{1}{|S(A)| + |S(B)|} (\sum_{a \in S(A), b \in S(B)} \min \|a - b\| + \sum_{b \in S(B), a \in S(A)} \min \|b - a\|) \quad (14)$$

其中, $S(A)$ 和 $S(B)$ 分别为预测和真实边界点集。ASD 的优点是综合反映全局边界误差,稳定性优于豪斯多夫距离。

3.4 损失函数

本文使用 Dice Loss^[21] 和 CE Loss^[22] 的加权组合作为损失函数。

$$L = \alpha L_{dice} + \beta L_{ce} \quad (15)$$

其中, α 和 β 均取值为 0.5。

Dice Loss 直接优化分割区域的重叠度(IoU),对小目标(如医学图像中的病灶)更敏感,针对性解决类别不平衡问题。

$$L_{dice} = 1 - \frac{2 \sum (p_i y_i) + \epsilon}{\sum p_i + \sum y_i + \epsilon} \quad (16)$$

其中, p_i 是预测概率; y_i 是真实标签; ϵ 是平滑项,防止除数为 0。

CE Loss 是分类任务的基础损失函数,通过概率匹配驱动模型学习。

$$L_{ce} = -\frac{1}{N} [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)] \quad (17)$$

3.5 实验与分析

3.5.1 编解码器模块选择

本文提出了两种编码器结构:基于局部感知支路的 PAA 模块,以及兼具局部与全局感知支路的 PGAA 模块。为探究 PGAA 模块在模型位置的适用性,进行了系统的对比实验。实验基于 WHM 分割挑战赛数据集,实验结果如表 2 所列。

表 2 编解码器模块选择

Table 2 Module selection of encoder and decoder

编/解码器	瓶颈结构	MIoU	DSC	HD(↓)	ASD(↓)
PGAA	PGAA	0.7825	0.6991	35.64	2.36
PAA	PGAA	0.7959	0.7257	31.67	2.16

实验结果表明,在网络全部层级采用 PGAA 模块的性能表现反而低于仅在瓶颈结构中部署 PGAA 模块的方案。经分析,出现这一现象的主要原因有以下两方面。

1)卷积操作更适合提取局部特征;在编码器阶段,网络主要从输入图像中提取局部特征(如边缘、纹理、细微结构)。卷积操作更适合此阶段,能够高效捕捉空间相关的局部模式。由于低层特征高度依赖局部信息,CNN 在该阶段优于 Transformer,避免了在早期阶段建模不必要的长程依赖。在瓶颈层,特征图被下采样至低空间分辨率,此时语义理解变得至关重要。该阶段需要全局上下文建模,以建立远距离区域间的关系。Transformer 的自注意力机制擅长捕捉此类长程依赖。

2)浅层编码器中 Transformer 存在过拟合风险;Transformer 是数据依赖性强的模型,依赖自注意力从头学习特征关系。在浅层网络中,低层特征间冗余交互(如边缘像素间的无关关联)。在数据量不足时,模型可能学习虚假相关性而非真实结构模式,导致过拟合。在瓶颈层,特征图经过空间压缩,同时保留了高层语义信息,使 Transformer 能够聚焦于关键的全局关系,避免了局部噪声的干扰。因此,瓶颈层是引入 Transformer 的理想位置。

3.5.2 消融实验

本文在 WHM 分割挑战赛数据集上通过消融实验,评估模型各部分模块的有效性,结果如表 3 所列。其中,基线表示不加任何模块的 Unet 网络,PAA 表示局部感知注意力模块,PGAA 表示结合了局部和全局感知的注意力模块,GFAM 表示门控语义流对齐模块。由实验结果可知,与基线相比,加入 PAA 模块使 MIoU 提高了 0.0315,说明该模块可以增强小目标识别能力,改善网络性能。将瓶颈结构换为 PGAA 模块,加入全局学习后,MIoU 再次提高,证明了 Transformer 模块的有效性。在此基础上,再次加入 GFAM 门控语义流对齐模块,在上采样过程中进行特征图对齐,分割精度得到进一步提升。此外,相较于 PAA,PGAA 和 GFAM 模块的两两组合方案,3 个模块共同作用的效果最优,MIoU 值达到 0.7959。

表 3 基于 WHMS 分割挑战赛数据集的不同模块消融实验

Table 3 Ablation experiments on different modules on the WHMS segment challenge dataset

基线	PAA	PGAA	GFAM	MIoU	DSC	HD(↓)	ASD(↓)
✓				0.7483	0.6773	47.51	3.61
✓	✓			0.7798	0.6985	36.40	2.59
✓		✓		0.7751	0.6974	38.25	2.67
✓			✓	0.7767	0.6977	37.32	2.52
✓	✓	✓		0.7843	0.7094	35.18	2.34
✓	✓		✓	0.7913	0.7106	34.77	2.31
✓		✓	✓	0.7825	0.6991	35.64	2.36
✓	✓	✓	✓	0.7959	0.7257	31.67	2.16

3.5.3 GFAM 模块可视化分析

为探究 GFAM 模块对特征提取过程的影响,对解码器第三阶段输出的特征图进行了可视化分析,结果如图 8 所示。

实验结果表明,引入 GFAM 模块后,模型在小尺度目标检测任务中展现出性能提升:1)细节特征提取能力增强,特征图中可见更丰富的纹理细节;2)边缘保持性能改善,目标轮廓的几何结构的完整性得到更好的保留。相比之下,无 GFAM

模块的基准模型虽能保持较高的特征图激活强度,但其空间分辨率存在明显退化,特别是在红色标注区域可见明显的边缘模糊现象。这表明了 GFAM 模块通过特征图对齐融合机制,提升了模型的特征表征,抑制了小目标细节信息的损失。

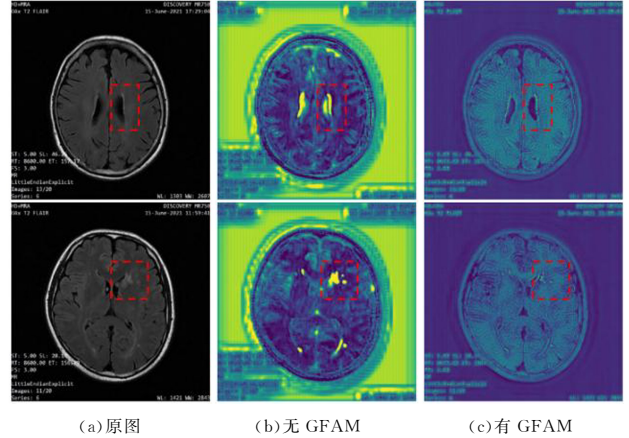


图 8 GFAM 模块激活可视化结果

Fig. 8 Class activation mapping based on GFAM block

3.5.4 与其他分割模型的比较

为系统评估 PGF-Net 模型在脑白质高信号数据集的分割任务中的性能表现,首先在自采数据集上选取了多种主流分割方法进行对比实验。这些方法包括:经典医学图像分割网络 nnUNet^[23]、基于 Transformer 架构的 Swin Transformer 网络,以及近年提出的先进分割模型 UNet3+^[24],nnSAM^[25]和 ConDSeg^[26]。

定量分析结果如表 4 所列。PGF-Net 模型在 4 个关键评价指标上均取得最优性能:mIoU 达到 0.8769,Dice 系数为 0.8423,HD 降至 32.61,ASD 为 1.7。这一系列实验数据表明了,PGF-Net 在脑白质高信号分割任务中的优越性,特别是在病灶边界精确划分和空间位置准确性方面展现出显著优势。

在模型效率方面,通过对比不同模型的前向传播所需的浮点运算次数(FLOPs),来分析模型的复杂度。从表 4 可以看出,PGF-Net 模型的浮点运算次数是 7.3040×10^7 ,虽然高于 nnUNet(3.1290×10^7),Swin Transformer(2.7020×10^7),UNet3+(2.6970×10^7)和 ConDSeg 模型(4.5550×10^7),但是对比 nnSAM 模型的 1.3348×10^8 次运算降低了 6.0440×10^7 次。这说明本文加入的 PPA,PGA 和 GFAM 模块为了提升分割准确率,引入了适度的计算开销,但能够将整体计算量控制在一定范围内。

表 4 各模型的性能对比(自采数据集)

Table 4 Performance comparison of models(self collected dataset)

模型	MIoU	Dice	HD	ASD	FLOPs
nnUNet	0.8452	0.8102	47.99	2.75	3.1290×10^7
Swin Transformer	0.7654	0.6612	48.61	2.82	2.7020×10^7
UNet3+	0.7613	0.6554	50.17	2.81	2.6970×10^7
nnSAM	0.8404	0.8097	40.15	2.56	1.3348×10^8
ConDSeg	0.7357	0.6315	33.86	2.47	4.5550×10^7
PGF-Net	0.8769	0.8423	32.61	1.70	7.3040×10^7

3.5.5 分割结果可视化

为了系统评估所提模型的分割性能,图 9 给出了各模型在

自采数据集中的分割可视化结果。结果表明,相比于其他模型,PGF-Net 在分割脑白质高信号的任务中生成的分割掩模与

专家标注的金标准具有更高的轮廓一致性,并且减少了漏分割现象,这进一步证明了 PGF-Net 模型具有更强的分割能力。

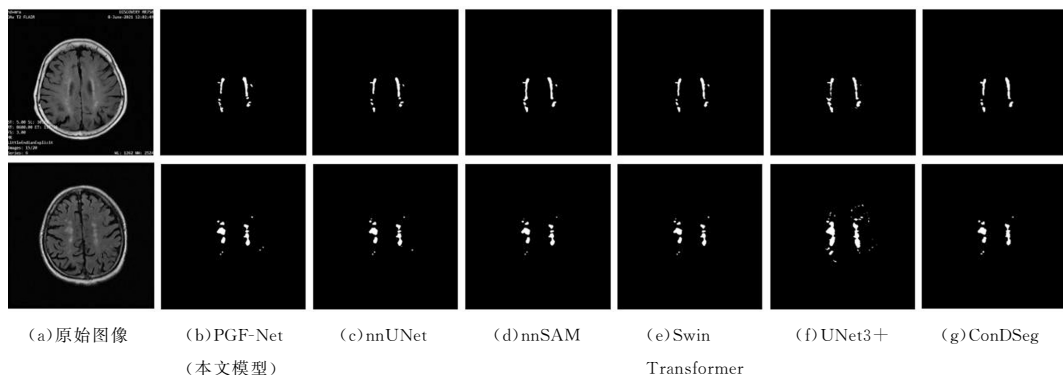


图9 各模型性能对比的可视化(自采数据集)

Fig.9 Visualization of performance comparison of various models(self collected dataset)

3.5.6 泛化性实验

为系统评估 PGF-Net 模型在分割任务中的泛化性能,选取了具有显著成像差异的 IDRiD 数据集和 WHM 分割挑战赛数据集进行实验。如表 5、表 6 所列,PGF-Net 在两种数据集上均展现出最优的分割性能:在 mIoU(0.738/0.7959)、Dice 系数(0.6351/0.7257)、HD(125.77/31.67)和 ASD(4.67/2.16)4 项指标上均优于对比方法。值得注意的是,尽管 IDRiD 数据集与 WHMS 分割挑战赛数据集在成像形态、颜色和分布等方面存在显著差异,但其分割目标均较小。这说明本文方法具备一定的泛化性和鲁棒性,为医学图像分析中的小目标分割提供了新的解决方案。

表5 各模型的性能对比(IDRiD 数据集)

Table 5 Performance comparison of models(IDRiD dataset)

模型	<i>MIoU</i>	<i>DSC</i>	<i>HD</i>	<i>ASD</i>	<i>FLOPs</i>
nnUNet	0.6845	0.6084	146.35	6.27	3.1200×10^7
Swin Transformer	0.6525	0.4490	155.05	8.13	2.7000×10^7
UNet3+	0.6241	0.3564	160.93	10.15	2.6970×10^7
nnSAM	0.7019	0.6210	144.81	5.94	1.3348×10^8
ConDSeg	0.5941	0.3164	177.93	12.24	4.5550×10^7
PGF-Net	0.7380	0.6531	125.77	4.67	7.3040×10^7

表6 各模型的性能对比(WHMS 分割挑战赛数据集)

Table 6 Performance comparison of models(WHMS segment challenge dataset)

模型	<i>MIoU</i>	<i>DSC</i>	<i>HD</i>	<i>ASD</i>	<i>FLOPs</i>
nnUNet	0.7483	0.6773	47.51	3.61	3.1200×10^7
Swin-Transformer	0.6881	0.6206	48.37	3.79	2.7000×10^7
UNet3+	0.6732	0.6160	53.26	3.75	2.6970×10^7
nnSAM	0.7524	0.6818	42.45	3.44	1.3348×10^8
ConDSeg	0.6647	0.6115	35.06	3.34	4.5550×10^7
PGF-Net	0.7959	0.7257	31.67	2.16	7.3040×10^7

结束语 本文提出一种基于 U 型架构的脑白质高信号分割模型 PGF-Net,通过三方面创新提升小病灶分割性能:在编码器各阶段引入局部感知注意力模块,增强对局部病灶信息的提取能力;其次,在瓶颈阶段引入局部与全局感知注意力模块,利用全局上下文增强小目标特征表示;利用全局依赖补充小目标区域的上下文信息,增强特征表示。最后,在解码器中利用门控语义流对齐模块,使上采样后的特征与编码器中

的低层特征精准对齐,并进行特征融合。

在自采数据集上的消融实验分析验证了各个模块的有效性,且 PGF-Net 在分割精度上显著优于现有方法。不同数据集下的结果也证实了所提模型在极具挑战性的小目标分割任务中优势明显,具有泛化性和鲁棒性。

PGF-Net 尽管在脑白质高信号分割任务中表现出色,但仍存在一些局限性,多级注意力机制与门控对齐模块的引入增加了计算开销,可能限制其在临床实时场景中的部署。未来可以探索知识蒸馏、模型剪枝或量化等方法,进一步优化计算效率,以提升模型的实用性。

参考文献

- [1] CAO J,ZHONG W,XIA Y,et al. The association between vascular white matter hyperintensities and cognitive function:a longitudinal study based on community populations [J]. Chinese Journal of Clinical Neuroscience,2025,33(2):200-209,220.
- [2] WANG C,XU J,FU Q,et al. The relationship between Fazekas grading of white matter hyperintensities and cognitive and neurological impairment in patients [J]. Clinical and Educational Journal of General Practice,2025,23(4):331-333,354.
- [3] ZHAN J,QIU W,LAN H,et al. Study on the severity of high signal intensity in white matter and the distribution of network disease syndrome types [J]. Modern Chinese Doctor,2025,63(2):1-4.
- [4] FARKHANI S,DEMnitz N,BORAXBEKK C J,et al. End-to-end volumetric segmentation of white matter hyperintensities using deep learning[J]. Computer Methods and Programs in Biomedicine,2024,245:108008.
- [5] DOSOVITSKIY A,BEYER L,KOLESNIKOV A,et al. An image is worth 16x16 words:Transformers for image recognition at scale[J]. arXiv:2010.11929,2020.
- [6] LEE A,WOO I,KANG D,et al. Fully automated segmentation on brain ischemic and white matter hyperintensities lesions using semantic segmentation networks with squeeze-and-excitation blocks in MRI[J]. Informatics in Medicine Unlocked,2020,21:100440-100440.
- [7] PARK G,HONG J,DUFFY B A,et al. White matter hyperin-

- tensities segmentation using the ensemble U-Net with multi-scale highlighting foregrounds [J]. *NeuroImage*, 2025, 237: 118140.
- [8] ZHANG X F, JIANG Y F, GUO S J, et al. Brain tissue segmentation method combining multi-scale and attention mechanisms [J]. *Journal of Jilin University (Engineering and Technology Edition)*, 2025, 55(10): 3352-3360.
- [9] GHAFORIAN M, KARSSEMEIJER N, HESKES T, et al. Location Sensitive Deep Convolutional Neural Networks for Segmentation of White Matter Hyperintensities [J]. *Scientific Reports*, 2017, 7(1): 5110.
- [10] XU S, ZHENG S, XU W, et al. HCF-Net: Hierarchical Context Fusion Network for Infrared Small Object Detection [C] // 2024 IEEE International Conference on Multimedia and Expo (ICME), 2024.
- [11] VASWANIA, SHAZEER N, PARMAR N, et al. Attention is all you need [C] // *Advances in Neural Information Processing Systems*, 2017.
- [12] SHI B, GAI S, DARRELL T, et al. Refocusing is key to transfer learning [J]. *arXiv: 2305. 15542*, 2023.
- [13] WANG Q, WU B, ZHU P, et al. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks [C] // 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020.
- [14] WOO S, PARK J, LEE J, et al. Cbam: Convolutional block attention module [C] // *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018: 3-19.
- [15] LIU Z, LIN Y, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows [C] // *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021: 10012-10022.
- [16] LI X, YOU A, ZHU Z, et al. Semantic flow for fast and accurate scene parsing [C] // *Computer Vision-ECCV*, 2020: 775-793.
- [17] JADERBERG M, SIMONYAN K, ZISSERMAN A. Spatial transformer networks [C] // *Advances in Neural Information Processing Systems*, 2015.
- [18] PORWAL P, PACHADE S, KAMBLE R, et al. Indian diabetic retinopathy image dataset (IDRiD): a database for diabetic retinopathy screening research [J]. *Data*, 2018, 3(3): 25.
- [19] KUIJF H, BIESBROEK J M, DE BRESSER J, et al. Data of the White Matter Hyperintensity (WMH) Segmentation Challenge [J]. *IEEE Transactions on Medical Imaging*, 2019, 38(11): 2556-2568.
- [20] HUTTRNLOCHE R, DANIEL P, GREGORY A, et al. Comparing images using the Hausdorff distance [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1993, 15(9): 850-863.
- [21] ZHAO R, QIAN B, ZHANG X, et al. Rethinking dice loss for medical image segmentation [C] // 2020 IEEE International Conference on Data Mining (ICDM). IEEE, 2020: 851-860.
- [22] WANG Y, MA X, CHEN Z, et al. Symmetric cross entropy for robust learning with noisy labels [C] // *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019: 322-330.
- [23] ISENSEE F, PETERSEN J, KLEIN A, et al. nnu-net: Self-adapting framework for u-net-based medical image segmentation [J]. *arXiv: 1809. 10486*, 2018.
- [24] HUANG H, LIN L, TONG R, et al. 3+: A full-scale connecticutnet for medical image segmentation [C] // *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020.
- [25] LIY, JING B, LI Z, et al. Plug-and-play segment anything model improves nnUNet performance [J]. *Medical Physics*, 2025, 52(2): 899-912.
- [26] LEI M, WU H, LYU X, et al. Condseg: A general medical image segmentation framework via contrast-driven feature enhancement [C] // *Proceedings of the AAAI Conference on Artificial Intelligence*, 2025: 4571-4579.



ZHANG Xinfeng, born in 1974, Ph.D, associate professor. His main research interests include signal and information processing and machine learning.



LI Xiangsheng, born in 1975, Ph.D, professor. His main research interests include functional MRI imaging research and early diagnosis of lung cancer.

(责任编辑:柯颖)