

基于语义与形态特征融合的语义分割网络

魏博文¹, 全红艳²

(1. 华东师范大学软件工程学院, 上海 200062; 2. 华东师范大学计算机学院, 上海 200062)

摘要: 视网膜血管检测有助于医生诊断视网膜疾病, 而以往基于特征融合的算法难以解决视网膜血管检测中出现的漏分割问题, 且分割准确率较低. 本文对特征融合方式做出进一步探索, 并提出一种基于语义与形态特征融合的算法, 通过挖掘输入特征中蕴含的语义与形态信息, 建模特征间的相关关系. 随后, 使用特征融合模块实现多模态特征自适应地融合. 在公开数据集 DRIVE 以及 STARE 上的实验结果表明, 文章算法优于现有的语义分割模型, 尤其在敏感性上, 比传统 U-Net 网络提升了 8.20%.

关键词: 语义分割; 视网膜血管检测; 形态特征; 语义特征; 卷积模块; 特征融合

中图分类号: TP391

文献标识码: A

文章编号: 0372-2112(2022)11-2688-10

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20220074

Semantic Segmentation Network Based on Semantic and Morphological Feature Fusion

WEI Bo-wen¹, QUAN Hong-yan²

(1. Software Engineering Institute, East China Normal University, Shanghai 200062, China;

2. School of Computer Science, East China Normal University, Shanghai 200062, China)

Abstract: Retinal blood vessel detection is helpful for doctors to diagnose retinal diseases, but the previous algorithm based on feature fusion is difficult to solve the problem of missed segmentation in retinal blood vessel detection, and the segmentation accuracy is low. This paper further explores the feature fusion method and proposes an algorithm based on the fusion of semantic and morphological features. It models the correlation between features by mining the semantic and morphological information contained in the input features. Then, the feature fusion module realizes the adaptive fusion of multi-modal features. The experimental results on the public datasets DRIVE and STARE show that, the article algorithm is better than the existing semantic segmentation model, especially in sensitivity, which is 8.20% higher than the traditional U-Net.

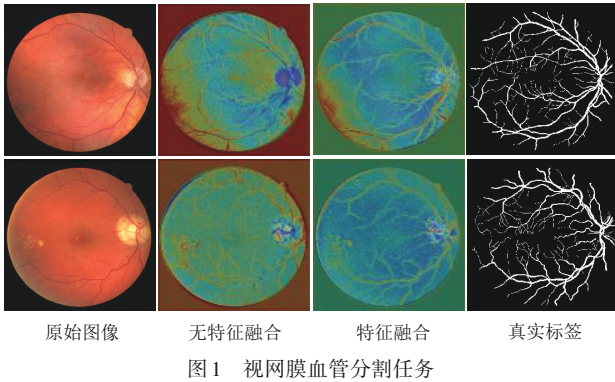
Key words: semantic segmentation; retinal vessel detection; morphological features; semantic features; convolution unit; feature fusion

1 引言

视网膜血管检测作为并发症检测手段, 有助于辅助诊断例如糖尿病、高血压, 以及精神退行性疾病等病症. 基于传统方法的视网膜血管检测^[1,2]. 难以适应不同的采样场景, 且精度较低. 随着 Ronneberger^[3]等人提出 U 型编解码卷积神经网络, 在医学影像分割上显示出巨大潜力. 此后, 许多研究人员基于此提出了改进的语义分割模型^[4,5]. 然而, 对于视网膜血管分割任务, 其前景血管像素相比背景非血管像素占整张图像比例极低, 这种前背景数据不平衡的问题使得视网膜血管分

割极易造成漏分问题. 近年来, 多尺度特征融合的方法在诸如目标检测等视觉任务上取得了良好成效^[6,7]. 这启发了研究者们对 U 型编解码网络与特征融合进行了更多探索. 多尺度特征融合能整合除同级外其余尺度的卷积特征, 这种融合方式使得特征图所含信息更为丰富. 如图 1 所示, 将多尺度特征融合引入视网膜血管分割任务, 可以显著提升网络对信息的整合能力, 获得更为清晰的特征图像. 因此, 寻求一种有效的特征融合方式是至关重要的. 文献[8]采用门控特征融合方法进行特征融合, 通过对编码器的特征进行提炼, 抑制无关特征, 使得网络更多地关注前景分割目标; 文献[9]提

出了空间金字塔融合模块,其将网络的整体框架划分为三个小型的子分割网络,分别用三种不同尺度的图像训练,通过空间金字塔模块实现多尺度特征的融合.



上述方法虽然在视网膜血管分割任务中取得了不错的效果,但是仍然存在不足,即特征间采用等权值连接(concat)或加和(adding)的融合方式.等权值连接增加了计算的开销;加和方式虽然不会带来额外的计算开销,但是由于各尺度特征之间存在差异,简单的等比例加和反而削弱了重要特征在特征图中的表达.此外,卷积神经网络的浅层提取形态特征,深层聚焦于语义特征.现有方法未能考虑到在浅层中弥补语义特征,在深层弥补形态特征.

针对上述存在的问题,本文受注意力机制启发^[10,11],提出了一种语义与形态特征自适应融合方法.其中,形态感知特征提取模块从空间维度强化重要的

前景特征,使得网络对前景区域更加敏感;语义感知特征提取模块除了提取语义信息外,还从通道维度对特征图信息进行整合,编码通道间的注意力关系,最大化地利用重要通道中蕴含的语义信息;语义与形态特征融合模块利用语义与形态特征提取模块中学习到的注意力系数,自适应地调整这两种模态特征在融合特征中的比例,并按照逐元素加和的方式进行特征融合.上述设计的多尺度卷积模块有效弥补了卷积神经网络浅层中缺失的语义信息和深层中缺失的形态信息,促进了形态特征与语义特征之间的交互融合,从而有效地提高了最终的分割精度.

2 本文方法

2.1 多尺度卷积模块

如图2所示,多尺度卷积模块作为编码器中的卷积单元,其内部级联了三个分支模块,分别是形态感知特征提取模块、语义感知特征提取模块、语义与形态特征融合模块.与图3以往分割模型广泛使用的基础卷积模块相比,本文设计的多尺度卷积模块通过构建语义与形态卷积深差,实现了语义与形态信息的协调互补. $X_{(i,1)}$ 与 $X_{(i,2)}$ 分别作为第*i*个多尺度卷积模块中形态与语义感知特征提取模块的输入. $X_{(i,2)}$ 的特征通道数为*C*,尺度为 $H \times W$. $X_{(i,1)}$ 的图像尺度是 $X_{(i,2)}$ 的两倍; $F_{(i,1)}$ 与 $F_{(i,2)}$ 分别表示第*i*个形态与语义感知特征提取模块的输出; F_i 表示自适应地融合语义与形态两种模态信息后的特征.

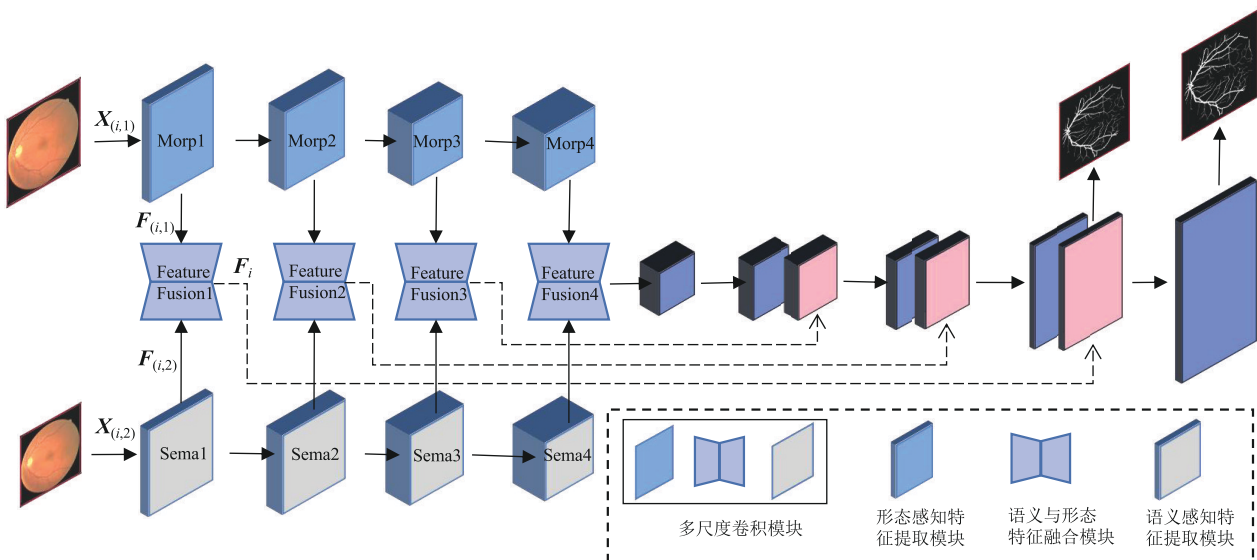


图2 基于多尺度卷积模块的语义分割网络模型整体结构

2.2 形态感知特征提取模块

Zhang^[12]等人在网络中引入边界监督,并证明了形态学特征可以提升分割的精度,但是这种依赖轮廓监

督的方法需要预先处理出分割图像的轮廓标签,但由于视网膜血管主干与分支仅有十几像素的差别,构造的轮廓标签和真实标签并不会有很大的区别,使得这

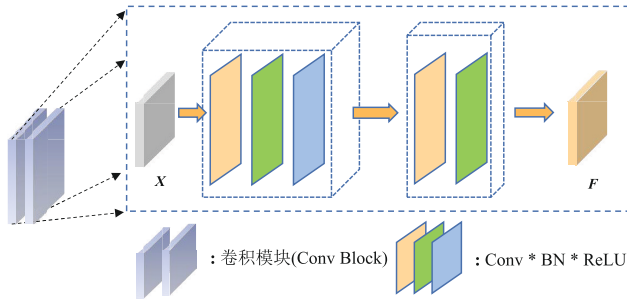


图3 基础卷积模块

种方法难以应用于视网膜血管分割任务,而本文形态感知特征提取模块可以引导网络提升对前景区域的关注度,从而适用于不同形态的医学目标,同时无需预先将真实标签处理成轮廓标签.形态感知特征提取模块如图4所示,Max和Mean分别表示通道最大池化与平均池化,输入特征经池化并在空间维度降维后得到特征尺度为 $1 \times 2H \times 2W$ 的掩膜 M_i^{Max} 与 M_i^{Mean} :

$$M_i^{\text{Max}} = \text{Conv}(\text{Max}(X_{(i,1)})) \quad (1)$$

$$M_i^{\text{Mean}} = \text{Conv}(\text{Mean}(X_{(i,1)})) \quad (2)$$

其中,Conv表示卷积核大小为 3×3 的卷积操作.最终得到的空间注意力 $S_{(i,1)}$ 可表示为

$$S_{(i,1)} = \delta(\text{Conv}(\text{concat}(M_i^{\text{Max}}, M_i^{\text{Mean}}))) \quad (3)$$

其中,concat表示将两种形态特征掩膜 M_i^{Max} 与 M_i^{Mean} 沿着通道维度等权值连接. $\delta(\cdot)$ 表示Sigmoid激活函数.经最后线性变换后得到最终尺度为 $C \times 2H \times 2W$ 的形态特征 $F_{(i,1)}$:

$$F_{(i,1)} = S_{(i,1)} \otimes X_{(i,1)} \quad (4)$$

2.3 语义感知特征提取模块

对于语义特征的挖掘,一些方法从特征尺度^[13,14]出发.例如,上下文融合方法^[15]使用空洞空间金字塔(atrous spatial pyramid pooling)构建比率分别为1、6、

12、18的空洞卷积核,在通道维度等权值连接不同尺度的语义特征.但这样做忽视了通道之间的关系.本文提出了语义感知特征提取模块,除了构建不同尺度的输入实现多尺度特征融合之外,还从通道维度思考了语义特征之间彼此的内在联系.本文利用全局平均池化(Global Average Pooling, GAP)和全局最大池化(Global Max Pooling, GMP)从空间维度将特征图压缩为1维向量,通过Sigmoid激活函数得到各个特征图的重要性系数,并在训练的过程中不断地更新和调整,最终寻找到通道间最优的线性组合关系,具体流程如图5所示.语义特征 $F_{(i,2)}$ 是通道注意力 $M_{(i,2)}$ 和多谱通道注意力 M_i^{DCT} 以及恒等残差连接(residual identity mapping)的输入特征 $X_{(i,2)}$ 三者的聚合.GMP与GAP将输入特征沿空间维度降维为两个特征尺度为 $C \times 1 \times 1$ 的注意力掩膜 M_i^{Gmp} 与 M_i^{Gap} :

$$M_i^{\text{Gmp}} = \text{Gmp}(\text{Conv}(X_{(i,2)})) \quad (5)$$

$$M_i^{\text{Gap}} = \text{Gap}(\text{Conv}(X_{(i,2)})) \quad (6)$$

随后通道注意力 $M_{(i,2)}$ 对这两种注意力进行综合:

$$M_{(i,2)} = \sigma(M_i^{\text{Gap}} \oplus M_i^{\text{Gmp}}) \quad (7)$$

然而,GAP实际上是频域中频率分量为0池化的特例,它并不适用于所有的图像.故除通道注意力外,还使用多谱通道注意力^[11],从不同频率角度提取特征图信息:

$$M_i^{\text{DCT}} = \delta(\text{Freq}) \quad (8)$$

$$\text{Freq} = \text{concat}([\text{Freq}^0, \dots, \text{Freq}^j, \dots, \text{Freq}^{n-1}]) \quad (9)$$

其中, Freq^j 表示计算出的第 j 个频率分量的结果,经等权值连接并Sigmoid激活函数归一化后得到 M_i^{DCT} ,其为0到1之间的通道注意力系数.最后多谱通道注意力 M_i^{DCT} 与通道注意力 $M_{(i,2)}$ 加和,并与输入特征 $X_{(i,2)}$ 恒等残差连接融合后,得到最终的语义特征 $F_{(i,2)}$:

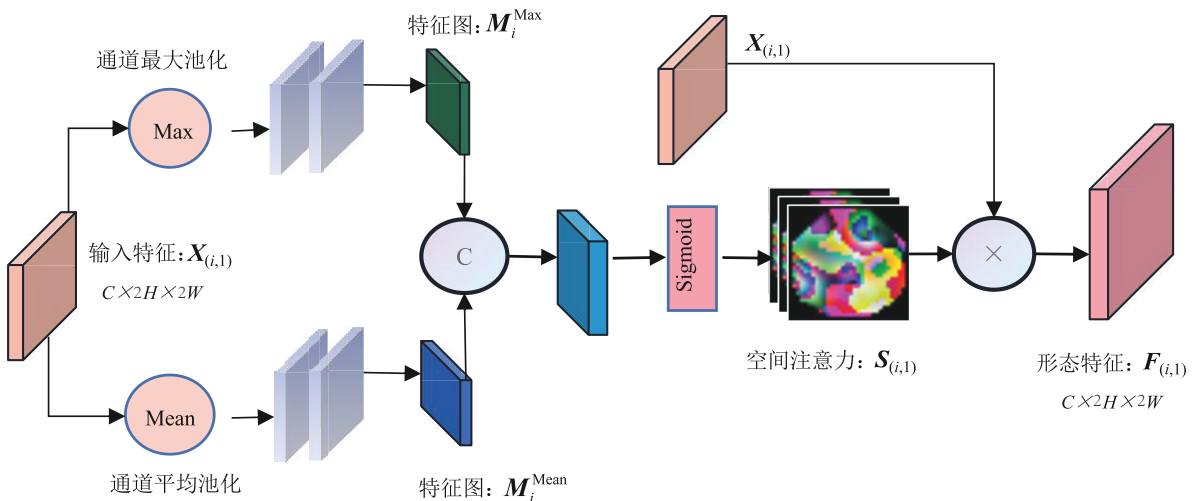


图4 形态感知特征提取模块

$$F_{(i,2)} = X_{(i,2)} \otimes (M_{(i,2)} \oplus M_i^{DCT}) \quad (10)$$

2.4 语义与形态特征融合模块

在之前的工作中,不同尺度特征间采用等权值连接,或加和方式实现特征融合,前一种方法在训练和推理过程中要求网络保存一份临时的特征图,加倍了计

算的开销.后者,简单地加和不同尺度的特征,存在特征不匹配的问题,为解决这个问题,本文设计的形态和语义融合模块利用以上两种特征提取模块中的注意力信息 $S_{(i,1)}$ 与 $M_{(i,2)}$,通过反向传播不断地更新形态和语义特征在融合特征中的比例,最终使得网络自动地学习到完美的融合比例.

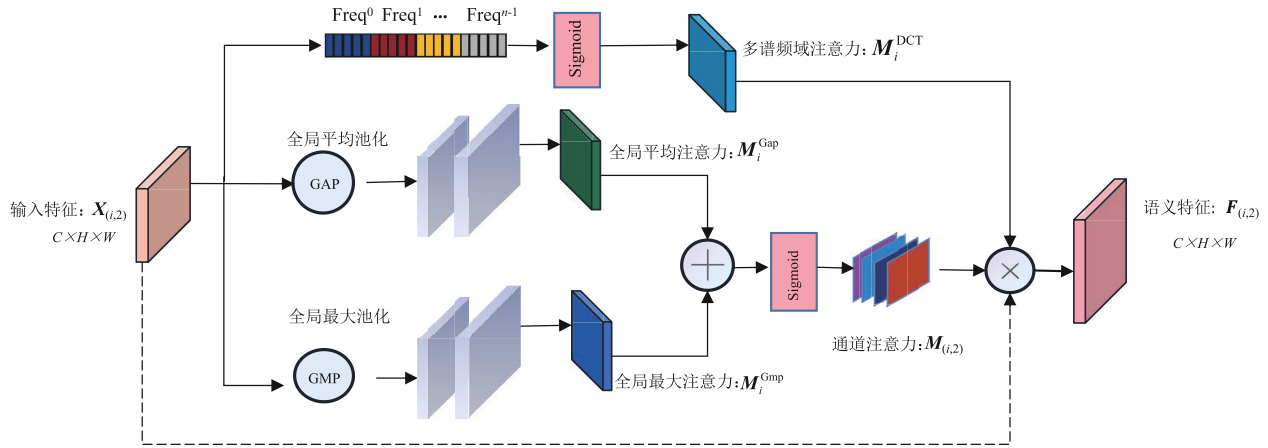


图5 语义感知特征提取模块

$$Prop_{(i,1)} = (1 - S_{(i,1)}) \otimes X_{(i,1)} \quad (11)$$

$$Prop_{(i,2)} = (1 - M_{(i,2)}) \otimes X_{(i,2)} \quad (12)$$

$$F_i = Prop_{(i,1)} \oplus Prop_{(i,2)} \oplus X_{(i,2)} \quad (13)$$

$Prop_{(i,1)}$ 与 $Prop_{(i,2)}$ 为按照相应注意力系数比例调整后的形态特征与语义特征. 最终融合的特征 F_i 为自适应形态特征 $Prop_{(i,1)}$ 、自适应语义特征 $Prop_{(i,2)}$ 以及输入特征 $X_{(i,2)}$ 三者的组合. 由于 $S_{(i,1)}$ 与 $M_{(i,2)}$ 可随着网络训练动态更新,这就保证了最终融合特征 F_i 的融合比例不是一成不变的,而是会根据输入图像中语义以及形态特征对分割结果的贡献,动态地融合多模态特征.

3 实验结果与分析

3.1 实验数据集与评估指标

3.1.1 数据集

DRIVE^[16]以及 STARE^[17]数据集是衡量视网膜血管分割算法模型使用最广的数据集. DRIVE 中含有 40 张照片,分辨率为 768×584 pixel,其中 33 张没有任何糖尿病视网膜病变的迹象,另外 7 张有轻度糖尿病视网膜病变的迹象. STARE 数据集图像大小为 605×700 pixel,含有 40 幅血管分割手工标注结果和两组共 80 幅血管检测手工标注结果.

3.1.2 评估标准

分析视网膜血管分割方法性能的指标为敏感性,特异性和准确性以及 AUC. 其中敏感性(sensitivity)又称真阳性率(true positive rate),即患者被诊断为阳性的概率,此值越大,说明诊断试验越灵敏,以 Sen 代表敏感

性,计算公式如下:

$$Sen = \frac{TP}{TP + FN} \quad (14)$$

其中 TP 表示真阳性,即被模型预测为正的正样本. FN 表示假阴性,即被模型预测为正的负样本. 特异性 (specificity) 又称真阴性率(true negative rate),即实际上未患病的人被诊断为阴性的概率,以 Spe 代表特异性,计算公式如下:

$$Spe = \frac{TN}{TN + FP} \quad (15)$$

此值越大,说明诊断结果越精确. 其中 TN 表示真阴性,即被模型预测为负的负样本. FP 表示假阳性,即被模型预测为负的正样本,又称为误报率. AUC 值作为衡量模型分类性能好坏的评价标准, AUC 数值更大的分类器效果更好. 准确率则是表示正确预测的像素点数 (TP+TN) 占整张图像中像素的比例. 以 Acc 代表准确率,计算公式如下:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (16)$$

3.2 实验细节

3.2.1 数据集预处理

本文实验中所使用的显卡为 NVIDIA GeForce RTX 2070, CPU 为 AMD Ryzen 5 2600x, 实验运行内存为 32 GB, 深度学习框架为 Pytorch 1.3. 针对 DRIVE 数据集中大小为 768×584 pixel 的图像,考虑到其四周不存在分割目标,故首先裁去四周的边界,随后在训练过程中将图像缩放为 448×448 pixel 作为模型的输入. 对于

STARE 数据集, 本文将分辨率大小为 605×700 pixel 的图像缩放为 448×448 pixel. 此外, 本文还使用了数据增强的方法, 例如: 调整色调饱和度和值、随机平移缩放旋转、随机水平竖直旋转等数据增强方法, 以及均值为 0、标准差为 3.2 的标准化操作.

3.2.2 训练细节以及超参数设置

本文实验采用 Adam 优化器, 初始学习率为 $2e-4$. 当损失函数连续 20 个 epoch 没有减少时, 将学习率降低到原来的一半, 直到学习率小于 $5e-7$ 停止训练. 本文损失函数为 Tversky^[18] 损失以及二元交叉熵损失. 其中, Tversky 损失如式 (17) 所示:

$$L(\varphi, \omega) = \frac{\sum_{i=1}^N p_{0i} g_{0i}}{\sum_{i=1}^N p_{0i} g_{0i} + \varphi \sum_{i=1}^N p_{0i} g_{1i} + \omega \sum_{i=1}^N p_{1i} g_{0i}} \quad (17)$$

其中 p_{0i} 表示第 i 个像素是血管的概率, p_{1i} 表示第 i 个像素不是血管的概率, g_{0i} 为真实标签, 表示第 i 个像素为血管, g_{1i} 与之相反. 通过调整超参数 φ 与 ω , 可以控制假阳性和假阴性之间的平衡. 在本文的实验中, 由于视网膜血管漏分割比过分割的情况严重, 经过多次实验测试, 最终设定 φ 为 0.3, ω 为 0.7. 二元交叉熵损失如式 (18) 所示:

$$L(x, y) = -y \log(\delta(x)) - (1-y) \log(1-\delta(x)) \quad (18)$$

其中 x 为预测值, y 为真实标签, $\delta(\cdot)$ 表示 Sigmoid 函数.

3.3 实验结果分析

3.3.1 与其他算法的对比结果

(1) 在 DRIVE 数据集上的对比结果

本小节首先与现有的算法在 DRIVE 数据集上进行了对比. 实验结果如表 1 所示, 本文算法在四种评估指标上均优于现有方法. 表 1 中 U-Net^[3]、R2U-Net^[4]、UNet++^[5]、DUNet^[19]、DDNet^[20] 方法同样使用 U 型编解码的网络框架, 但不同的是本文编码器中的卷积单元为多尺度卷积模块而非基础卷积模块. 从实验结果上可以看出多尺度卷积模块的性能全面优于基础卷积模块. 除却与其他语义分割网络的对比外, 表 1 中还存在另外一组对比实验: 与其他多监督方法的对比, Yan^[21] 等人以 U-Net 为主干, 在主监督外, 添加辅助的分类器一同对主干网络进行监督. 与 Yan 等人的方法相比, 本文的优势是通过 Tversky 损失函数, 灵活地调整假阳性以及假阴性之间的平衡, 从而解决了血管区域的漏分割问题以及把背景判定为前景的过分割问题.

(2) 在 STARE 数据集上的对比结果

如表 2 所示, 本文还在 STARE 数据集进一步与现有的方法进行了对比. 本文方法在特异性, 敏感性, 准确率以及 AUC 上同样领先视网膜血管检测任务上最新的 Xu^[9] 等人的方法. 值得一提的是, STARE 与 DRIVE

表 1 与其他算法在 DRIVE 数据集上的分割结果特异性、敏感性、准确率、AUC 的比较结果 单位: %

算法	年份	Spe	Sen	Acc	AUC
U-Net ^[3]	MICCAI'15	98.20	75.37	95.31	97.55
R2U ^[4]	NAECON'18	98.13	77.99	95.96	97.84
MS-Net ^[13]	MICCAI'18	98.19	78.44	95.67	98.07
Yan ^[21]	T-BME'18	98.18	76.53	95.42	97.52
UNet++ ^[5]	DLIA'18	98.36	79.59	95.47	97.79
DEU ^[14]	MICCAI'19	98.16	79.40	95.67	97.72
DUNet ^[19]	KBS'19	98.00	79.63	95.66	98.02
CE Net ^[15]	TMI'19	-	83.09	95.45	97.79
DDNet ^[20]	TMI'20	97.88	81.26	95.94	97.96
Xu ^[9]	ICASSP'21	98.07	82.13	96.15	98.15
Ours	2021	98.68	83.57	96.73	98.85

表 2 与其他算法在 STARE 数据集上的分割结果特异性、敏感性、准确率、AUC 的比较结果 单位: %

算法	年份	Spe	Sen	Acc	AUC
U-Net ^[3]	MICCAI'15	97.01	82.88	95.78	97.72
R2U ^[4]	NAECON'18	98.20	77.56	96.34	98.15
Yan ^[21]	TBE'18	98.46	75.81	96.12	98.01
DUNet ^[19]	KBS'19	98.78	75.95	96.41	98.32
CE Net ^[15]	TMI'19	98.43	80.72	96.57	98.29
DDNet ^[20]	TMI'20	97.69	83.91	96.85	98.58
Xu ^[9]	ICASSP'21	98.11	82.30	97.03	98.61
Ours	2021	98.34	84.95	97.43	98.88

数据集中的图像形态上存在着较大的区别, 前者 50% 的图像为患病视网膜采样图, 这导致了所成图像表现出参差不齐的色度变化. 而本文算法在两组视网膜血管数据集上均取得了领先的结果, 这证明了多尺度卷积模块具有不错的泛化性.

(3) 与其他卷积模块的对比

为了更清楚地了解多尺度卷积模块的性能, 本节进一步探讨了与其他基础卷积模块的区别. 表 3 中列举了三种卷积模块分别是 Res-block^[6]、Dense-block^[7] 和 SE-block^[22], 以上模型在本节的对比实验均使用了相同的 U 型编解码结构, 不同的是分别用 ResNet, DenseNet 以及 SeResNet 作为主干网络. 对于 ResNet 而言, 根据卷积模块类型将 ResNet 系列模型分为两类: 基于 Basic-block 的 ResNet18, 34 浅层网络, 以及基于 Bottleneck-block 的 ResNet50, 101, 152 深层网络. 对于 DenseNet 而言, Dense-block 的中心思想是对于每一层利用前面所有层的特征映射作为当前层输入, 且本身的特征映射为后续层提供输入, 形成全互连接. 而 SeRes-block 是

表 3 与其他卷积模块在 DRIVE 数据集上的分割结果特异性、敏感性、准确率、AUC 对比结果 单位: %

编码器	Block	Spe	Sen	Acc	AUC
ResNet18 ^[6]	Basic	97.03	77.43	94.35	97.05
ResNet34	Basic	97.21	77.56	94.53	97.10
ResNet50	Bottle	97.32	77.74	94.62	97.49
ResNet101	Bottle	97.53	78.03	94.72	97.58
DenseNet121 ^[7]	Dense	97.38	77.87	94.43	96.25
DenseNet169	Dense	97.46	78.09	94.76	96.34
SeResNet50 ^[22]	SERes	97.45	77.56	94.58	96.99
SeResNet101	SERes	97.51	78.21	94.79	97.14
Ours	MSCM	98.68	83.57	96.73	98.85

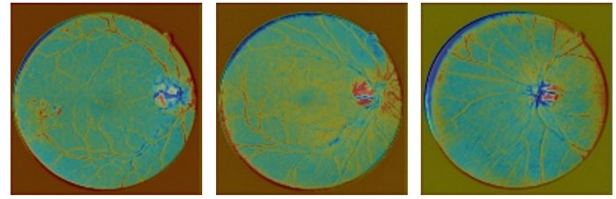
将SE模块应用到Res-block上,需要首先将特征维度降维,然后经过ReLu激活后再通过一个全连接(fully connection)层升回到原来的维度.这样做可以更好地拟合通道间复杂的相关性并且极大地减少了参数量和计算量.从实验结果可以看出本文方法在敏感性、特异性以及AUC上均领先于以上卷积模块网络,这是因为本文的多尺度卷积模块在编码阶段特意引入了语义特征以及形态特征,这两种特征彼此协同使得模型取得超越以往卷积模块算法的性能.

3.3.2 多尺度卷积模块的性能分析

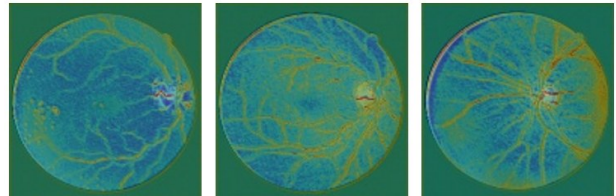
本节主要讨论本文中提出的语义感知特征提取模块与形态感知特征提取模块在网络模型中的作用.本实验以自定义的wide-UNet作为本文网络模型的骨架网络,在表4中用Backbone表示,先后替换骨架网络原编码器主分支为语义感知特征提取模块分支和形态感知特征提取模块分支,用以评估两种特征提取模块在主干网络上的性能.随后,替换骨架网络编码器为包含语义提取、形态提取、语义与形态融合模块的多尺度卷积模块,并分析了各模块的复杂度.

(1) 语义感知特征提取模块性能分析

图6为骨架网络与骨架网络添加语义模块前后特征图的可视化对比结果,可以看出使用了语义感知特征提取模块的网络解决了前景血管模糊不清的问题,这说明提取的语义特征,增强了网络对于前景血管的关注.



(a) 骨架网络编码器第三阶段特征图



(b) 添加语义感知特征提取模块第三阶段特征图

图6 骨架网络与添加语义感知特征提取模块前后特征图对比

(2) 形态感知特征提取模块性能分析

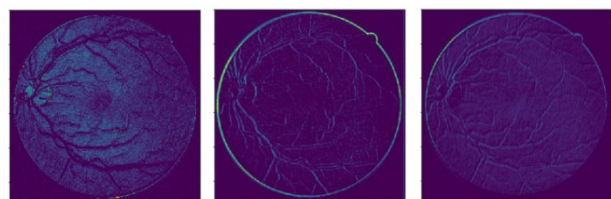
图7为骨架网络与骨架网络添加形态模块前后特征图可视化对比结果,可以看出使用了形态感知特征提取模块的网络对于视网膜血管中复杂的血管结构有着更强的识别能力,这是因为使用形态感知特征提取模块的网络,从空间维度提取了视网膜图像中的血管形态的特征表示,补充了网络中血管细节信息,从而得到了更为清晰的分割结果.

(3) 语义与形态特征融合模块性能分析

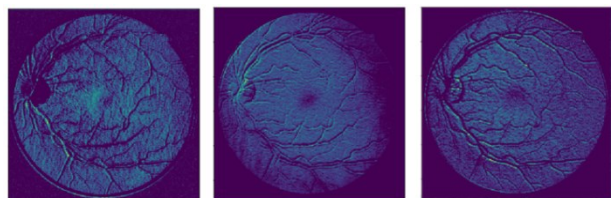
更进一步地,本节还对本文的语义与形态特征融合方法进行了分析.在以上形态提取与语义提取模块的基础上,网络添加语义与形态特征融合模块.实验结果如表4所示,对比仅使用形态特征提取模块或语义特征提取模块,各项指标均有大幅提升.其中,敏感性提升最多,而敏感性是用于衡量网络在处理视网膜血管漏分问题的性能.由于视网膜血管分割任务中存在的前背景数据不平衡,这使得血管漏分变得更难以解决.从表1、表2中各模型的敏感性指标数据可以看出,以往的方法均没有很好地解决漏分问题.但值得一提的是,引入语义与形态特征融合模块的网络与骨架网络相比敏感性指标提升了3.76,对比仅使用形态特征提取模块1.3的提升,以及仅使用语义特征提取模块1.68的提升,

表 4 各模块对网络性能的影响

算法	语义提取	形态提取	特征融合	Sen	Spe	Acc	AUC	Parameters	GFLOPs
Backbone	×	×	×	79.81	97.31	95.83	97.52	10.84M	11.70
With Sema	√	×	×	81.30	98.28	96.46	98.54	11.15M	12.47
With Morp	×	√	×	81.49	98.45	96.59	98.67	11.23M	12.90
Ours	√	√	√	83.57	98.68	96.73	98.85	11.45M	13.61



(a) 骨架网络编码器第一阶段特征图



(b) 添加形态感知特征提取模块第三阶段特征图

图7 骨架网络与添加形态感知特征提取模块前后特征图对比

这证明了语义与形态特征融合模块通过自适应地融合语义与形态特征,能有效地解决视网膜血管分割任务中的漏分问题。

(4) 模块复杂度分析

本节还对本文各模块的参数量以及浮点运算次数进行了分析。本文最终含有形态提取模块、语义提取模块,以及语义与形态融合模块的网络相比骨架网络在参数量上增加了 0.61 M,在浮点运算次数上增加了 2.91 GFLOPs。其中浮点运算次数计算是以推理一张 224×224 pixel 的视网膜血管图像为依据。本文算法以 0.61 M 的参数代价、2.91 GFLOPs 的计算代价,提升了骨架网络 1.37、3.76、0.90、1.33 的特异性、敏感性、准确率、AUC 性能。尤其在敏感性指标上提升较高,这将大大减少模型对视网膜血管结构漏分的概率,对于视网膜血管疾病的预筛具有重大

意义。

最后,如表 5 所示,本文还对比了其他特征融合方法,与使用求和融合方式的 Pyramid U-Net^[9]相比,本文的特征融合方法仅使用了语义与形态两个尺度的特征,但在特异性、敏感性、准确率以及上均优于 Pyramid U-Net,这也证明了本文语义与形态特征融合方法的有效性。上下文融合的 CE-Net^[15]以及引入 CBAM^[10]中通道注意力思想的 Att UNet^[23],这些方法从不同角度解决了多尺度特征融合问题,相比于原始的等权值连接方法,都有一定程度上的提升,但提升有限。本文的语义感知特征提取模块从语义角度出发,使得浅层卷积更快的识别与定位前景目标,形态感知特征提取模块从形态角度细化深层卷积语义特征的边界,并通过特征融合模块有效解决了视网膜血管检测中血管目标小,漏分割的问题,相比于现有的多模态特征融合方法有显著的提升。

表 5 与其他融合方式在 DRIVE 数据集上的指标特异性、敏感性、准确率、AUC 的对比结果 单位:%

融合方式	网络	Spe	Sen	Acc	AUC
等权值连接 ^[3]	U-Net	97.31	79.81	95.83	97.52
上下文融合 ^[15]	CE Net	-	83.09	95.45	97.79
求和 ^[9]	PUNet	98.07	82.13	96.15	98.15
通道注意力 ^[23]	Att UNet	98.49	81.35	96.58	98.52
门控融合 ^[8]	AG Net	98.32	83.40	96.43	98.59
语义与形态	Ours	98.68	83.57	96.73	98.85

3.4 结果可视化

图 8 展示了本文算法与其他卷积模块在 DRIVE 数据集上训练至收敛的 loss rate/epoch number 的曲线,在

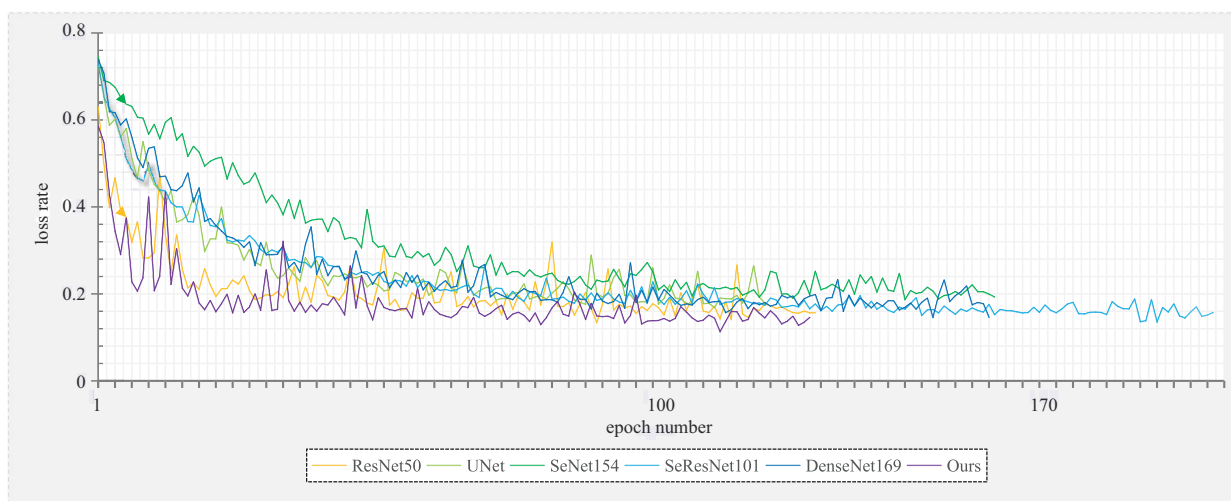


图8 与其他卷积模块算法训练至收敛曲线对比

本节实验中的所有网络均保持相同的训练设置,以及使用相同的交叉熵损失函数.紫色的曲线代表本文的方法,横轴表示训练轮数,纵轴为损失.可以看出,在相同的训练轮数下,本文基于多尺度卷积模块的模型收敛效果更好,且本文算法训练至收敛所需轮数,相比其他卷积模块算法花费更少.结合之前定量对比实验的结果,可以认为本文的方法在最短的训练时间内得到了收敛效果最好的模型,图9、图10为本文算法与其他算法在DRIVE数据集与STARE上分割结果的可视化效果对比.可以看出,采用等权值连接融合方法的算法只能识别较大的血管分支,并且预测的血管像素要比真实标签更宽,使用上下文融合算法模型整体分割效果相对较好,但在细小的血管位置仍然存在漏分的现象;相比于上下文融合算法,本文使用了多尺度卷积模块的网络在视网膜毛细血管密集区域分割结果更加精细.与求和方法使用三个尺度特征的Pyramid U-Net网络相比,本文仅使用语义与形态两个尺度的特征融合,取得了更好的分割结果.

可视化效果对比.可以看出,采用等权值连接融合方法的算法只能识别较大的血管分支,并且预测的血管像素要比真实标签更宽,使用上下文融合算法模型整体分割效果相对较好,但在细小的血管位置仍然存在漏分的现象;相比于上下文融合算法,本文使用了多尺度卷积模块的网络在视网膜毛细血管密集区域分割结果更加精细.与求和方法使用三个尺度特征的Pyramid U-Net网络相比,本文仅使用语义与形态两个尺度的特征融合,取得了更好的分割结果.

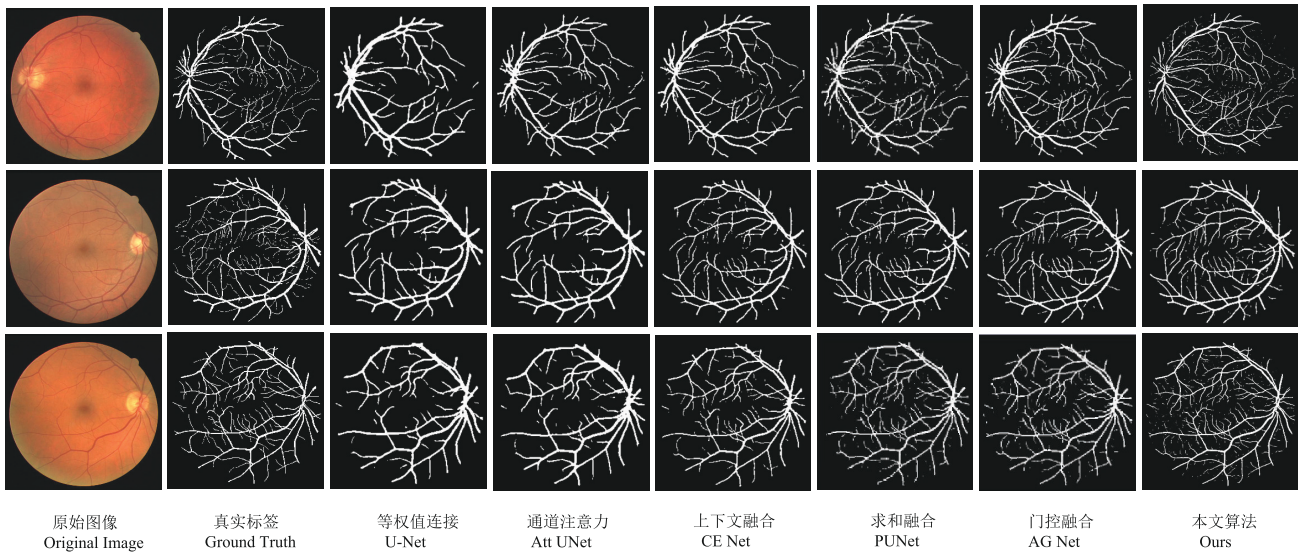


图9 与其他算法在DRIVE上分割可视化结果对比

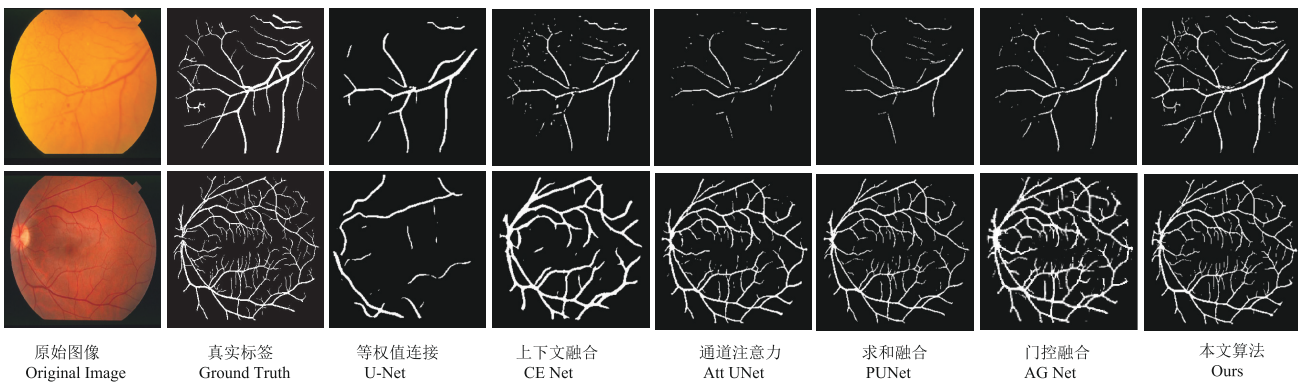


图10 与其他算法在STARE上分割可视化结果对比

4 结论

本文提出了一种基于语义与形态特征融合的语义分割算法,通过自适应融合多模态特征,结合深监督策略解决了以往视网膜血管检测算法的漏分割问题,这对于视网膜疾病的预筛查具有重要意义.其次,本文证明了应用语义与形态特征融合模块能够细化编码器各阶段的特征,有效地提升骨架网络的性能.最后,实验

结果表明本文方法能够有效应对不同的视网膜分割场景,在相同的训练时间下达到更好的分割精度,并在特异性、敏感性、准确率等指标上优于现有的算法.

参考文献

- [1] ZANA F, KLEIN J C. Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation [J]. IEEE Transactions on Image Processing, 2001, 10(7):

- 1010-1019.
- [2] SINGH N P, KUMAR R, SRIVASTAVA R. Local entropy thresholding based fast retinal vessels segmentation by modifying matched filter[C]//International Conference on Computing, Communication & Automation. Noida: IEEE, 2015: 1166-1170.
- [3] RONNEBERGER O, FISCHER P, BROX T. U-net: convolutional networks for biomedical image segmentation [C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich: Springer, Cham, 2015: 234-241.
- [4] ALOM M Z, YAKOPCIC C, TAHA T M, et al. Nuclei segmentation with recurrent residual convolutional neural networks based u-net(R2U-Net)[C]//NAECON 2018-IEEE National Aerospace and Electronics Conference. Dayton: IEEE, 2018: 228-233.
- [5] ZHOU Z, SIDDIQUEE M, TAJBAKHS N, et al. Unet++: redesigning skip connections to exploit multiscale features in image segmentation[J]. IEEE Transactions on Medical Imaging, 2020, 39(6): 1856-1867.
- [6] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 770-778.
- [7] HUANG G, LIU Z, MAATEN L VAN DER, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Hawaii: IEEE, 2017: 4700-4708.
- [8] ZHANG S, FU H, YAN Y, et al. Attention guided network for retinal image segmentation[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Shenzhen: Springer, Cham, 2019: 797-805.
- [9] ZHANG J, ZHANG Y, XU X. Pyramid u-net for retinal vessel segmentation[C]//ICASSP2021—2021 IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP). Toronto: IEEE, 2021: 1125-1129.
- [10] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision(ECCV). Munich: Springer, 2018: 3-19.
- [11] QIN Z, ZHANG P, WU F, et al. Fcanet: frequency channel attention networks[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021: 783-792.
- [12] ZHANG S, FU H, XU Y, et al. Retinal image segmentation with a structure-texture demixing network[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Lima: Springer, Cham, 2020: 765-774.
- [13] WU Y, YONG X, YANG S, et al. Multiscale network followed network model for retinal vessel segmentation [C]//International Conference on Medical Image Computing & Computer-assisted Intervention. Granada: Springer, Cham, 2018: 119-126.
- [14] WANG B, QIU S, HE H. Dual encoding u-net for retinal vessel segmentation[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Shenzhen: Springer, Cham, 2019: 84-92.
- [15] GU Z, CHENG J, FU H, et al. Ce-net: context encoder network for 2d medical image segmentation[J]. IEEE Transactions on Medical Imaging, 2019, 38(10): 2281-2292.
- [16] STAAL J, ABRAMOFF M D, NIEMEIJER M, et al. Ridge-based vessel segmentation in color images of the retina[J]. IEEE Transactions on Medical Imaging, 2004, 23(4): 501-509.
- [17] HOOVER A, KOUZNETSOVA V, GOLDBAUM M. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response[J]. IEEE Transactions on Medical Imaging, 2000, 19(3): 203-210.
- [18] SALEHI S S M, ERDOGMUS D, GHOLIPOUR A. Tversky loss function for image segmentation using 3D fully convolutional deep networks[C]//Machine Learning in Medical Imaging: 8th International Workshop. Quebec: Springer, 2017: 379-387.
- [19] JIN Q, MENG Z, PHAM T D, et al. Dunet: a deformable network for retinal vessel segmentation[J]. Knowledge-Based Systems, 2019, 178: 149-162.
- [20] MOU L, CHEN L, CHENG J, et al. Dense dilated network with probability regularized walk for vessel detection[J]. IEEE Transactions on Medical Imaging, 2019, 39(5): 1392-1403.
- [21] YAN Z, XIN Y, CHENG K T. Joint segment-level and pixel-wise losses for deep learning based retinal vessel segmentation[J]. IEEE Transactions on Biomedical Engineering, 2018, 65(9): 1912-1923.
- [22] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 7132-7141.

- [23] ISLAM M, VIBASHAN V S, JOSE V, et al. Brain tumor segmentation and survival prediction using 3D attention unet[C]//International MICCAI Brainlesion Workshop. Shenzhen: Springer, Cham, 2019: 262-272.

作者简介



魏博文 男,1996年6月出生,安徽六安人。现为华东师范大学软件工程学院硕士研究生。主要研究方向为计算机视觉、深度学习。
E-mail: 51194501070@stu.ecnu.edu.cn



全红艳(通讯作者) 女,1968年6月出生,黑龙江哈尔滨人,华东师范大学计算机科学与技术学院副教授,硕士生导师,研究方向为计算机视觉,人工智能。当前研究的兴趣领域为基于深度学习的医学影像分析、医学影像三维重建。
E-mail: hyquan@cs.ecnu.edu.cn