

基于自注意力机制和多尺度输入输出的医学图像分割算法

丁天舒, 陈媛媛

引用本文

丁天舒, 陈媛媛. [基于自注意力机制和多尺度输入输出的医学图像分割算法](#)[J]. 计算机科学, 2024, 51(2): 135-141.

DING Tianshu, CHEN Yuanyuan. [Medical Image Segmentation Algorithm Based on Self-attention and Multi-scale Input-Output](#) [J]. Computer Science, 2024, 51(2): 135-141.

相似文章推荐 (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)

[LNG-Transformer:基于多尺度信息交互的图像分类网络](#)

LNG-Transformer:An Image Classification Network Based on Multi-scale Information Interaction
计算机科学, 2024, 51(2): 189-195. <https://doi.org/10.11896/jsjcx.221100218>

[基于深度学习的图像分割综述](#)

Image Segmentation Based on Deep Learning:A Survey
计算机科学, 2024, 51(2): 107-116. <https://doi.org/10.11896/jsjcx.230900002>

[基于知识图谱与用户兴趣的推荐算法](#)

Knowledge Graph and User Interest Based Recommendation Algorithm
计算机科学, 2024, 51(2): 55-62. <https://doi.org/10.11896/jsjcx.221200169>

[基于特征拓扑融合的黑盒图对抗攻击](#)

Black-box Graph Adversarial Attacks Based on Topology and Feature Fusion
计算机科学, 2024, 51(1): 355-362. <https://doi.org/10.11896/jsjcx.230600127>

[基于异质图神经网络预训练的多标签文档分类研究](#)

Pre-training of Heterogeneous Graph Neural Networks for Multi-label Document Classification
计算机科学, 2024, 51(1): 143-149. <https://doi.org/10.11896/jsjcx.230600079>

基于自注意力机制和多尺度输入输出的医学图像分割算法

丁天舒 陈媛媛

四川大学计算机学院 成都 610065

(2020223045177@stu.scu.edu.cn)

摘要 更精细化的糖尿病性视网膜病变眼底图像分割结果,可以更好地辅助医生进行诊断。大规模高分辨率的分割数据集的出现,为更精细化的分割提供了有利条件。基于 U-Net 的主流分割网络,使用基于局部运算的卷积操作进行像素预测时无法充分挖掘全局信息,网络模型采用单输入单输出的结构,难以获取多尺度特征信息。为了最大程度地利用现有的大规模高分辨率的眼底图像病灶分割数据集,实现更精细化的分割,需要设计更好的分割方法。文中基于自注意力机制和多尺度输入输出结构对 U-Net 进行改造,提出了一种新的分割网络 SAM-Net,用自注意力模块代替传统卷积模块,增大网络获取全局信息的能力,引入多尺度输入和多尺度输出结构,使网络更容易获取多尺度特征信息。使用图片切片方法来缩小模型的输入尺寸,防止神经网络模型因为输入图片像素过大而导致训练难度增大。最终在 IDRiD 数据集和 FGADR 数据集上进行实验,结果表明, SAM-Net 可以达到比其他方法更优的性能。

关键词: U-Net; 自注意力机制; 糖尿病性视网膜病变; 分割; 神经网络

中图分类号 TP389.1

Medical Image Segmentation Algorithm Based on Self-attention and Multi-scale Input-Output

DING Tianshu and CHEN Yuanyuan

College of Computer Science, Sichuan University, Chengdu 610065, China

Abstract Refined fundus image segmentation results of diabetic retinopathy can better assist doctors in diagnosis. The appearance of large scale and high resolution segmentation data sets provides favorable conditions for more refined segmentation. The mainstream segmentation network based on U-Net, using convolution operation based on local operation, cannot fully excavate global information when making pixel prediction. The network model adopts single-input single-output structure, which makes it difficult to obtain multi-scale feature information. In order to maximize the use of existing large-scale high-resolution fundus image focal segmentation data sets and achieve more refined segmentation, better segmentation methods need to be designed. In this paper, U-Net is transformed based on the self-attention mechanism and multi-scale input/output structure, and a new segmentation network, SAM-Net, is proposed. The self-attention module is used to replace the traditional convolutional module, and the ability of the network to obtain global information is increased. The multi-scale input and multi-scale output structures are introduced to make it easier for the network to obtain multi-scale feature information. The image slicing method is used to reduce the input size of the model, so as to prevent the training difficulty of the neural network model from increasing due to the large pixel of the input picture. Finally, experimental results on IDRiD and FGADR data sets show that SAM-Net can achieve better performance than other methods.

Keywords U-Net, Self-attention, Diabetic retinopathy, Segmentation, Neural network

1 引言

糖尿病性视网膜病变(Diabetic Retinopathy, DR)是最常见的糖尿病微血管并发症之一,是慢性进行性糖尿病导致的视网膜微血管渗漏和阻塞引起的一系列眼底病变。根据国际方案^[1-2],糖尿病性视网膜病变的严重程度可以分为5个阶段(0-4),即无视网膜病变、轻度非增殖性病变、中度非增殖性病变、重度非增殖性病变和增殖性病变,分级通常由不同相关病变区域的数量和大小来进行判断。微动脉瘤(Micro Aneu-

rysm, MA)是糖尿病性视网膜病变最早的临床可见证据,通常表现为小红点状的局部毛细血管扩张。硬性渗出物(Hard Exudates, HE)是位于黄斑区域的明显的黄白色视网膜内沉积物,表现为小斑点或较大的斑块。软性渗出物(Soft Exudates, SE),有时也称为“棉絮斑”(Cotton Wool Spots, CWS),是神经纤维层变色或毛细血管前动脉闭塞的灰白色斑块,渗出通常出现在严重的 DR 阶段。此外,视网膜内微血管异常(Intraretinal Microvascular Abnormalities, IRMAs)是毛细血管扩张和视网膜内新血管形成的区域,一旦 IRMA 大量

出现,就可以被诊断为增殖性糖尿病性视网膜病变前期。4种病灶的数量和位置是对糖尿病性视网膜病变图片进行分级的重要依据。

近年来,神经网络方法在图像的分级分类^[3]、目标检测^[4]和语义分割^[5]方面取得了突出的成就,在糖尿病性视网膜病变眼底图像的智能诊断中发挥了很大的作用^[6]。但是,目前的大部分研究使用端到端的图片分级预测方法^[7-11],直接由眼底图片得出病变分级结果,不符合医生进行诊断时先看病灶,再根据病灶的种类和数量进行分级的习惯。因此,现在糖尿病性视网膜病变图像分级工作已经从端到端的直接分级变成先由语义分割网络或者目标检测网络进行特征提取,再将特征作为分级网络输入来进行分级,在这种趋势下,需要精细化的语义分割方法,来提取图像特征。

在图像分割领域,最经典的网络是 U-Net^[12],U-Net 以其创造性的编码器、解码器和跳跃融合结构,成为了最经典的分割算法。但是,由于传统 U-Net 结构全部采用卷积与池化操作来进行编解码,感受野有限,难以充分挖掘有效的上下文信息,单输入单输出,网络结构难以提取丰富的多尺度信息。

许多学者与医学专家合作,积累了大量的有标签数据集,尤其是带有病灶分割标签的数据集^[13-18],促进了糖尿病性视网膜病变神经网络诊断方法的突飞猛进。随着设备的进步,眼底图像的分辨率和清晰程度也在不断提高,积累的数据量不断增大,但是,由于卷积神经网络的局限性,目前已有的神经网络分割方法已经无法满足更精细化分割的要求。因此,我们需要设计更好的神经网络方法,来学习图中更丰富的信息,实现更精细的病灶分割。

本文提出了一种新型的多尺度输入输出自注意力语义分割神经网络模型 SAM-Net。SAM-Net 引入多尺度输入输出结构和自注意力机制,来提升模型对空间信息的感知能力,提升网络提取多尺度信息能力。对眼底图片进行切割后进入网络进行训练测试,避免高分辨率图像使模型运算的时空复杂度过高。在 IDRiD^[19]数据集和 FGADR^[20]数据集中的软渗出和硬渗出血灶上,验证了所提方法的性能。

2 相关工作

2.1 糖尿病性视网膜病变眼底图像

使用神经网络方法进行糖尿病性视网膜病变的研究,首先发展起来的是端到端的糖尿病性视网膜病变图片的分级工作,将眼底图片直接输入神经网络模型,得到输出结果。Gulshan 等^[1]采用 Inception-v3 架构训练 DR 分级模型,直接输入眼底图像得出分级结果。Jiang 等^[11]训练了多个深度学习模型,其中一些深度模型还与 AdaBoost 结合,以减少每个单独模型的偏差,多个模型集成后共同组成糖尿病性视网膜病变眼底图像分级系统。Zhang 等^[21]探讨了理想组件分类器数量与类标签数量之间的关系,以及不同组件分类器组合对最佳集成性能的影响,构建了最优模型。这些工作在直接进行端到端分类的神经网络模型中取得了很好的效果。

随着目标检测模型和语义分割模型的出现,糖尿病性视网膜病变的分级工作也从端到端的直接给出分级结果变为先使用目标检测和语义分割模型进行特征提取,再使用特征

作为分级网络的输入来进行分级工作。Yang 等^[22]提出将病灶检测和图片分级相结合,设计两阶段深度卷积神经网络。具体来说,首先训练一个网络,将斑块划分为不同的病变,然后第二个网络预测 DR 的严重分级。Wang 等^[23]提出了一个网络,用于标记异常区域,然后以全局和局部方式提供糖尿病性视网膜病变的分级。Lin 等^[24]引入病变检测模型,首先提取病变信息,然后将原始图像与病变特征融合,以识别糖尿病性视网膜病变。Zhou 等^[25]引入协同学习框架,端到端优化病变分割模型和疾病分级模型。为了提高分级的准确性,提出了病灶注意模块,对无标签数据提取的病灶图进行细化,以进行半监督分割。在 Wu 等的工作中^[26],分割和分类是并行进行的,将分割模型中的病变概率预测图和弱监督分类模型中的类激活图结合起来进行联合诊断。Zhou 等^[20]提出了一个全新的数据集,基于此数据集,首先构建分割网络,使用病灶分割图片来训练图像分割网络,然后将网络用于分类训练,这比直接端到端的分类效果更好。

这些学者的工作,在糖尿病性视网膜病变的分级、病灶检测、病灶分割方面,都取得了很好的成果,为眼底图像的智能诊断做出了突出的贡献。

高分辨率分割数据集的诞生,为更高精度的分割奠定了基础,但对分割网络的提取特征能力的要求也相应提高,我们需要针对糖尿病性视网膜病变的精细化分割工作,设计精度更高、学习能力更强的神经网络模型。

2.2 U-Net 分割算法

FCN 是对图像进行像素级分类的开山之作,提出了使用卷积代替全连接层,使用编解码的网络结构来实现语义级别的图像分割解决方案。U-Net^[12]则更进一步,在 FCN 的编解码结构上做出改进,在跨层连接时使用拼接来实现更厚的特征层,实现了更好的多尺度特征提取效果。之后,U-Net 成为了最经典的网络结构,目前主流的分割网络都是基于 U-Net 进行改进。

Res-U-Net^[27]和 Dense-U-Net^[28]分别受到分类网络中的残差连接模块和密集连接模块的启发,将 U-Net 的每一个子模块分别用残差连接模块和密集连接模块进行替换,分别在视网膜血管分割任务和 3D 粒子场全息成像上取得突破。

MultiResUNet^[29]对 Res-U-Net 中的残差连接模块进行扩展,使用残差路径,使编码器的特征在卷积操作后再与解码器中对应特征进行拼接,在多个数据集上进行了性能验证。

UNet++^[30]使用具有 3 个卷积层的密集卷积块来弥合编码器和解码器的特征图之间的语义鸿沟,大幅提升了医学图像的分割效果。

使用 U-Net 作为主干网络结构的诸多网络模型,在语义分割领域做出了突出的贡献。

2.3 引入自注意力机制

Transformer^[31]诞生于机器翻译任务中。在自然语言处理领域,基于 Transformer 的方法在各种机器翻译任务中取得了最优的性能^[32]。

在 Transformer 成功经验的驱动下,研究人员引入了一种开创性的视觉转换器(ViT)^[33],它在图像识别任务中实现了极佳的效果。与基于 CNN 的方法相比,ViT 的缺点是需要

使用大数据集进行预训练。Liu 等^[34]基于移动窗口机制和 Transformer 提出了一个称为 Swin Transformer 的模块,在图像分类、目标检测和语义分割等各种视觉任务中都取得了最先进的性能。

Cao 等^[35]将 Swin Transformer 块作为基本单元,构建一个具有跳跃式连接的 U 形编码器-解码器架构,用于医学图像分割,从而为 Transformer 在医学图像领域的发展提供了一个用于比较的基准。

Transformer 凭借其独特的提取全局特征的优势,在图像领域已经逐渐开始取代卷积,成为了最常用的特征提取基础模块。

2.4 多尺度输入输出

Li 等^[36]将多尺度图像分别送入多尺度输入输出网络,在输出层进行合并后输出最终预测图,在视觉显著性检测和边缘检测任务中取得了很好的效果。Fu 等^[37]提出使用降采样层,在每一层先将上层输入进行降采样后融合到网络中。

Fu 等^[37]提出了多尺度输出结构,每一层的输出进行上采样后得到一个预测图,结合所有预测图得到最终预测结果,在眼底图像的视杯视盘分割中取得了极好的效果。

多尺度输入输出方法通过不同尺度的多条通道,使模型可以提取到更丰富的尺度信息,并对网络进行拓宽,避免因为网络过深而遗忘特征。

3 眼底图像病灶分割方法

3.1 整体流程

本文中神经网络模型训练与测试的流程如图 1 所示。

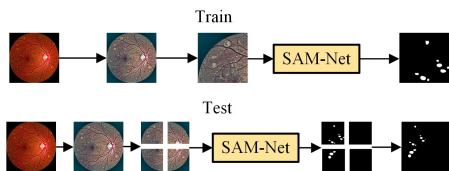


图 1 训练测试流程

Fig. 1 Train and test process

首先进行图片的预处理,将图片的 RGB 三通道颜色进行统一,避免图片偏向于某一种颜色,然后使用 CLAHA 算法

对图片进行处理,增强图片对比度,使图片中的信息更清晰。

由于图像的分辨率较高,分割模型所需要的时空复杂度过高,造成网络训练困难,因此在神经网络训练和测试时,不将整张图片放入网络中,而是首先将图片进行切割后,输入设计的 SAM-Net 模型进行模型训练,测试时将切片拼接后作为预测图像。

3.2 图像颜色均衡化和图像对比度增强

在 FGADR 中,存在多中心的数据,虽然相比其他数据,FGADR 数据集中的图片已经非常清晰,但是各中心的图片存在图片色差过大的情况,导致图片中的特征提取困难,因此使用图像颜色均衡化和图像对比度增强算法,来进行图片预处理。

颜色修正:

$$R^* = R - \text{mean}(R) \quad (1)$$

$$G^* = G - \text{mean}(G) \quad (2)$$

$$B^* = B - \text{mean}(B) \quad (3)$$

$$\text{image}^* = \text{image} / (\max(\text{image}) - \min(\text{image})) \quad (4)$$

进行颜色修正后,图片还存在部分病灶特征不清晰的情况,影响视觉效果,这会使得后期的网络训练中的模型难以准确提取特征,因此使用 CLAHA 算法对图片进行处理,使图片中的血管、病灶、视杯视盘等特征更加清晰。进行颜色修正和对对比度增强前后的眼底图片如图 2 所示。



图 2 预处理

Fig. 2 Preprocessing

3.3 SAM-Net 模型

本结构在传统 U-Net 的编码-解码的 U 形结构中引入多尺度输入输出,构成 M 形结构,并使用基于移动窗口机制的自注意力模块替换传统卷积模块,设计了一种新的 SAM-Net 模型,模型结构如图 3 所示。

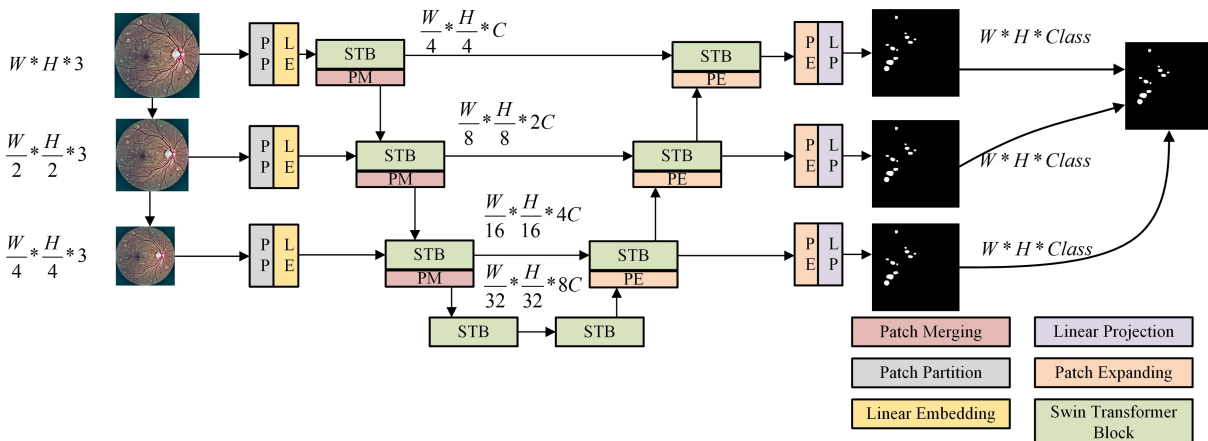


图 3 SAM-Net

Fig. 3 SAM-Net

模型由 3 部分组成,从输入到输出分别为多尺度输入结构、基于自注意力机制的编码解码结构和多尺度输出结构。

多尺度输入结构是 3 个并行的特征提取通道,使用 patch partition 和 linear embedding 方法提取 3 个不同尺度的特征。

基于自注意力机制的编码解码结构中,融合多尺度输入结构所提取的特征,使用 U-Net 的编码解码结构和基于自注意力机制的模块,来进行特征融合操作,形成 3 个特征图。其中的 self attention block 模块基于自注意力机制,可以比卷积模块更有效地提取特征;patch partition 用于降采样,将特征的分辨率减半,通道数加倍;patch expanding 用于上采样,将特征的分辨率加倍。

多尺度输出结构中,使用解码结构中的 3 个解码特征输出,构建 3 个并行的特征上采样通道。3 个通道分别输出与原图像素大小相同的特征图,融合后成为网络的输出特征图。

3.4 多尺度输入输出

传统 U-Net 结构采用独特的编码解码结构和跨层连接来进行分割工作,但是由于其单输入单输出的局限性,对不同尺度特征的提取能力不足,网络结构过深,容易造成特征丢失和训练困难。在编码解码路径上构建多尺度输入通道,可以拓宽网络结构,防止由网络过深导致的信息丢失,使神经网络更容易学习到不同尺度特征。

在多尺度输入结构中,首先采用平均池化层对图像进行下采样,并在编码器路径上构建并行的多尺度输入结构。每条路径的图像首先经过 patch merging 模块和 linear embedding 层进行特征提取,再将提取到的特征输入基于自注意力机制的编码解码结构中进行特征融合。

在多尺度输出结构中,使用编码解码结构中的解码后特征作为多条并行的侧向输出。每条通道经过 patch expanding 和 linear projection 模块后变为 $W * H * Class$ 像素,多个通道的输出进行平均后得到网络的预测特征图。

3.5 自注意力模块

本文使用自注意力机制构造的 swin Transformer 模块来代替传统的卷积模块^[34]。本模块由两个连续的自注意力模块组成,每个模块中主要包含对特征进行归一化的 layer norm 层、基于移动窗口的 MSA 模块以及进行特征提取的 MLP 模块。两个连续的自注意力模块分别使用 W-MSA 模块和 SW-MSA 模块。在模块内部,使用残差连接来进行跳跃特征融合,swin Transformer 模块的结构如图 4 所示。

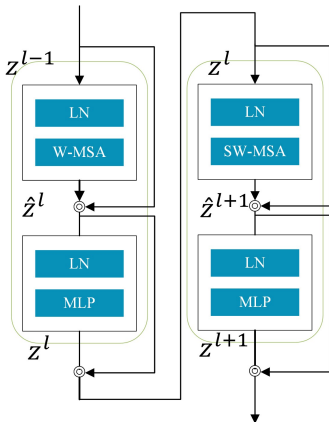


图 4 自注意力模块

Fig. 4 Self attention block

传统 U-Net 结构中使用的编码解码结构采用的卷积与反卷积等流行的运算操作是局部运算,没有充分挖掘有效的上下文信息对特征进行深入提取,使得网络在处理远层依存关系方面效率较低。相比传统卷积模块,本文中使用的自注意力模块没有感受野的局限性,为每一个预测像素关联更多的输入像素,可以更好地提取全局信息。缺点是由于缺少局部感受野的先验信息,需要更大规模的数据集进行训练。因此,我们采用多种数据增广方法,如切片、翻转、噪声等,来进行数据集扩充。

4 实验

4.1 糖尿病性视网膜病变相关数据集

本文采用 IDRiD^[19]和 FGADR^[26]两个数据集,进行神经网络模型的训练与测试。

IDRiD^[19]数据集提供带有典型糖尿病性视网膜病变和正常视网膜结构的专家注释的眼底图像。数据集包含 516 张图片,但其中只有 81 张有像素级的病灶标签来标记微动脉瘤、出血点、软渗出和硬渗出 4 种 DR 相关病变。

FGADR^[26]数据集包含 1 842 张带有像素级标注和图像级标注的糖尿病性视网膜病变图片,像素级标注包含 6 种病灶,图像级别标注符合 5 级分级标准,另外有 1 000 张图片带有像素级标注,该数据集中每一张图片像素规格均为 $1\ 280 * 1\ 280$ 。

4.2 图像切片

为降低模型训练时的计算复杂度,将图片进行切片后放入网络进行训练。

神经网络的时间复杂度,即模型的运算次数,由输出特征图面积 M_l^2 、卷积核面积 K_l^2 、输入通道数 C_{l-1} 、输出通道数 C_l 决定。神经网络的空间复杂度由两部分组成,即总参数量和各层输出特征图,总参数量只与卷积核面积 K_l^2 、输入通道数 C_{l-1} 、输出通道数 C_l 、层数相关,输出特征图的空间占用是输出特征图面积 M_l^2 和输出通道数 C_l 的连乘。

$$Time \sim O(\sum_{l=1}^D M_l^2 K_l^2 C_{l-1} C_l) \quad (5)$$

$$Space \sim O(\sum_{l=1}^D K_l^2 C_{l-1} C_l + \sum_{l=1}^D M_l^2 C_l) \quad (6)$$

将眼底图片进行切片后进行神经网络训练和测试,可以减小输出特征图的面积 M_l^2 ,大规模减少模型训练时所需要的时间和空间。

此外,切片是一种可以对数据集进行数据增广的方法。

4.3 实验设置

将 FGADR 数据集的 1 842 张图片按照 2:1 的比例进行随机划分,将 1 228 张作为训练图片,614 张作为测试图片。

IDRiD 数据集中包含 54 张训练图片和 27 张测试图片。

将 IDRiD 数据集中的图片像素缩放成 $1\ 280 * 1\ 280$,与 FGADR 数据集相同,方便模型进行训练与测试。

GPU 为 2 块 NVIDIA 3090 24G,内存 32GB,编码框架 PyTorch 0.4.1,CUDA10.0。采用 Adam 优化器,batchsize 为 8,学习率为 0.00001,迭代 epoch 为 100 次,终止次数为迭代次数,图片切割时切割大小为 640 像素大小,损失函数使用二值交叉熵和 dice 的联合损失函数。

4.4 图像分割评价方法

针对病灶分割使用 dice 进行评价时,由于并非每一张

图片都存在病灶,如图 5 所示,左图无渗出病灶,右图中有渗出病灶,因此不可以使用每张图片计算 dice 再进行平均的方法来进行总体计算。



图 5 渗出病灶

Fig. 5 Exudative lesions

在单张图片中, A 为图片的预测正标签像素, B 为图片的实际正标签像素, $dice$ 的计算式如下:

$$dice = \frac{2 * A \cap B}{A + B} \quad (7)$$

若图片中存在病灶,则 $dice$ 取值区间为 $[0, 1]$ 。若图片中不存在病灶,则 B 为 0, $dice$ 的值只有 0 和 1 两种情况, $dice$ 的取值区间为 $\{0, 1\}$,在进行每张图片的平均时会严重影响总体指标。

因此把所有图片中的每一个像素当作一个样本点,用于计算总体分数。当图片总数为 n 时, A_i 为第 i 张图片的预测正标签像素, B_i 为第 i 张图片的实际正标签像素,所有样本的评价方法的计算式为:

$$dice = \frac{2 * \sum_{i=0}^{n-1} A_i \cap B_i}{\sum_{i=0}^{n-1} (A_i + B_i)} \quad (8)$$

4.5 图像分割结果分析

4.5.1 使用自注意力机制和多尺度输入的 U-Net 模型优于传统的 U-Net 模型

将 FGADR 数据集和 IDRiD 数据集的图像分辨率压缩至 640 像素后进行训练与测试,在测试集上得到的结果如表 1 和表 2 所列。

表 1 FGADR

Table 1 FGADR

方法	软渗出	硬渗出
U-Net	0.344	0.590
Swin-Unet	0.378	0.601
U-Net++	0.369	0.587
Dense U-Net	0.382	0.595
SAM-Net	0.393	0.612

表 2 IDRiD

Table 2 IDRiD

方法	软渗出	硬渗出
U-Net	0.748	0.772
Swin-Unet	0.754	0.764
U-Net++	0.743	0.770
Dense U-Net	0.761	0.782
SAM-Net	0.757	0.788

根据表 1 和表 2 所列的实验结果可知,本文中提出的 SAM-Net 模型,在 FGADR 软渗出数据集、硬渗出数据集和 IDRiD 硬渗出数据集上达到了比其他方法更优的效果。

基于 U-Net 的主流分割网络,使用的卷积运算操作是局部运算,感受野有限,没有充分挖掘有效的上下文信息。网络模型采用单输入单输出结构,对多种尺度的信息提取能力有限,编码解码结构,使得神经网络模型较深而不够宽,在学习过程中容易造成特征遗忘。SAM-Net 引入自注意力机制,更加充分地挖掘图像中的上下文信息。引入多尺度输入输出结构,增加网络宽度,解决了网络过深引起的特征丢失问题,丰富模型提取多尺度信息的能力。

在 IDRiD 软渗出数据集上的表现略差,因为 IDRiD 数据集只有 54 张图片用作训练数据集,无法提供足够丰富的信息来进行训练,使用数据集扩增方法和大数据集预训练可以进行有效改善。

4.5.2 使用切片方法可以有效提升预测精度

无论 U-Net 还是 Swin-Unet,其最初设计针对的图片像素都是 200~400。而眼底图像的图片大小是传统语义分割任务的数倍,例如 FGADR 数据集,图片像素是 $1280 * 1280$,不进行图片的压缩,将高分辨率图像直接放入神经网络中,会增加神经网络计算时的时间复杂度与空间复杂度,导致模型训练难度极大。若将图像的分辨率压缩,会降低图像质量,严重影响分割精度。

使用切片方法,将图片切割为较小的图片后放入网络中,再将切片进行拼接后形成预测图像。该方法一方面可以降低神经网络计算时的时间复杂度与空间复杂度,另一方面也可以作为一种图像扩增方法。

将使用图像切片后的表 3 和表 4 所列的实验结果与未切片的表 1、表 2 中的实验数据进行对比可知,图像切片方法可以有效提升预测精度。

表 3 FGADR

Table 3 FGADR

方法	软渗出	硬渗出
U-Net(图像切片)	0.351	0.589
Swin-Unet(图像切片)	0.396	0.639
U-Net++(图像切片)	0.374	0.616
Dense U-Net(图像切片)	0.390	0.621
SAM-Net(图像切片)	0.403	0.652

表 4 IDRiD

Table 4 IDRiD

方法	软渗出	硬渗出
U-Net(图像切片)	0.751	0.780
Swin-Unet(图像切片)	0.765	0.803
U-Net++(图像切片)	0.759	0.785
Dense U-Net(图像切片)	0.767	0.793
SAM-Net(图像切片)	0.771	0.800

4.5.3 基于自注意力机制的神经网络在大数据预训练的前提下效果会明显提升

将表 5 中的实验结果与表 4 中的实验结果进行对比可知,基于 Transformer 的分割网络在经过大规模数据集预训练后,效果有明显的提升,原因是纯 Transformer 网络虽然弥补了传统 U-Net 感受野有限、难以获取全局信息的缺点,

但是由于缺少先验信息,需要更多的训练数据来学习图像中包含的信息。因此,基于 Transformer 的神经网络结构需要更大规模的训练数据,而 FGADR 数据集包含 1280 张含有像素级标准的眼底图像,可以满足大规模训练数据集的要求,经过 FGADR 数据集的预训练,SAM-Net 模型在 IDRiD 数据集上取得了最优的实验表现。

表 5 IDRiD 和 FGADR 的预训练

Table 5 pre-training of IDRiD and FGADR

方法	软渗出	硬渗出
U-Net(图像切片)	0.755	0.785
Swin-Unet(图像切片)	0.789	0.806
U-Net++(图像切片)	0.763	0.782
Dense U-Net(图像切片)	0.773	0.797
SAM-Net(图像切片)	0.803	0.812

结束语 针对经典 U-Net 全局信息提取不充分、易因局部化而出现分割误差、因网络过深而导致特征遗忘,无法充分提取多尺度特征等问题,本文提出了一种结合自注意力机制和多尺度输入输出结构的医学图像分割方法 SAM-Net,以获得尺度信息更丰富、特征更全面的病灶特征,提升了图像分割的精度。在 FGADR 和 IDRiD 中硬渗出和软渗出病灶数据集上进行实验,结果表明,本文方法相比其他方法具有更优的性能。后续我们会使用提取到的特征进行病变图像分级工作。

参考文献

- [1] GULSHAN V, PENG L, CORAM M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs [J]. *Jama*, 2016, 316(22):2402-2410.
- [2] HANEDA S, YAMASHITA H. International clinical diabetic retinopathy disease severity scale[J]. *Nippon Rinsho, Japanese Journal of Clinical Medicine*, 2010, 68(Suppl 9):228-235.
- [3] IRVIN J, RAJPURKAR P, KO M, et al. Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2019:590-597.
- [4] YAN K, WANG X, LU L, et al. Deeplesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning[J]. *Journal of Medical Imaging*, 2018, 5(3):036501.
- [5] FAN D, ZHOU T, JI G, et al. Inf-net: Automatic covid-19 lung infection segmentation from ct images[J]. *IEEE Transactions on Medical Imaging*, 2020, 39(8):2626-2637.
- [6] ASIRI N, HUSSAIN M, ADEL F, et al. Deep learning based computer-aided diagnosis systems for diabetic retinopathy: A survey[J]. *Artificial Intelligence in Medicine*, 2018, 8:41-57.
- [7] TU Z, GAO S, ZHOU K, et al. Sunet: A lesion regularized model for simultaneous diabetic retinopathy and diabetic macular edema grading[C]//2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). IEEE, 2020:1378-1382.
- [8] ARCADU F, BENMANSOUR F, MAUNZ A, et al. Deep learning algorithm predicts diabetic retinopathy progression in individual patients[J]. *NPJ Digital Medicine*, 2019, 2(1):1-9.
- [9] GARGEYA R, LENG T. Automated identification of diabetic retinopathy using deep learning [J]. *Ophthalmology*, 2017, 124(7):962-969.
- [10] SEOUD L, CHELBI J, CHERIET F. Automatic grading of diabetic retinopathy on a public database[C]//MICCAI. Springer, 2015.
- [11] JIANG H, YANG K, GAO M, et al. An interpretable ensemble deep learning model for diabetic retinopathy disease classification[C]//2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2019:2045-2048.
- [12] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C]//MICCAI. Springer, 2015:234-241.
- [13] Kaggle diabetic retinopathy detection competition[EB/OL]. <https://www.kaggle.com/c/diabetic-retinopathy-detection>.
- [14] DECENCIERE E, ZHANG X, CAZUGUEL G, et al. Feedback on a publicly distributed image database: the messidor database [J]. *Image Analysis & Stereology*, 2014, 33(3):231-234.
- [15] 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.
- [16] STAAL J, ABRAMOFF M, 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] International competition on ocular disease intelligent recognition[OL]. <https://odir2019.grand-challenge.org>.
- [18] Kaggle aptos 2019 blindness detection competition [OL]. <https://www.kaggle.com/c/aptos2019-blindness-detection>.
- [19] PORWAL P, PACHADE S, KOKARE M, et al. Idrid: Diabetic retinopathy-segmentation and grading challenge [J]. *Medical Image Analysis*, 2020, 59:101561.
- [20] ZHOU Y, WANG B, HUANG L, et al. A benchmark for studying diabetic retinopathy: Segmentation, grading, and transferability[J]. *IEEE Transactions on Medical Imaging*, 2020, 40(3):818-828.
- [21] ZHANG W, ZHONG J, YANG S, et al. Automated identification and grading system of diabetic retinopathy using deep neural networks[J]. *Knowl. Based Syst.*, 2019(175):12-25.
- [22] YANG Y, LI T, LI W, et al. Lesion detection and grading of diabetic retinopathy via two-stages deep convolutional neural networks[C]//ICCAI. Springer, 2017:533-540.
- [23] WANG Z, YIN Y, SHI J, et al. Zoom-in-net: Deep mining lesions for diabetic retinopathy detection[C]//MICCAI. 2017:267-275.
- [24] LIN Z, GUO R, WANG Y, et al. A framework for identifying diabetic retinopathy based on antinoise detection and attention-based fusion[C]//MICCAI. Springer, 2018:74-82.
- [25] ZHOU Y, HE X, HUANG L, et al. Collaborative learning of semi-supervised segmentation and classification for medical images[C]//CVPR. 2019.
- [26] WU Y, GAO S, MEI J, et al. Jcs: An explainable covid-19 diagnosis system by joint classification and segmentation[J]. arXiv:

- 2004.07054,2020.
- [27] CHAURASIA A,CULURCIELLO E. LinkNet; Exploiting encoder representations for efficient semantic segmentation[C]// 2017 IEEE Visual Communications and Image Processing (VCIP). 2017;1-4.
- [28] WU Y,WU J,JIN S,et al. Dense-U-net;Dense encoder-decoder network for holographic imaging of 3D particle fields[J]. Optics Communications,2021,493:126970.
- [29] IBTEHAZ N,RAHMAN M. MultiResUNet;Rethinking the U-Net architecture for multimodal biomedical image segmentation [J]. Neural Networks,2020,121:74-87.
- [30] ZHOU Z,SIDDIQUEE M,TAJBAKHS N,et al. UNet++ : Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation [J]. IEEE Trans Med Imaging, 2019, 39(6):1856-1867.
- [31] VASWANI A,SHAZEER N,PARMAR N,et al. Attention is all you need[J/OL]. Advances in Neural Information Processing Systems,2017,30. https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1e4a845aa-Paper.pdf.
- [32] DEVLIN J,CHANG M,LEE K,et al. BERT;Pre-training of deep bidirectional transformers for language understanding [C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1(Long and Short Papers). Minneapolis, Minnesota; Association for Computational Linguistics,2019;4171-4186.
- [33] DOSOVITSKIY A,BEYER L,KOLESNIKOV A,et al. An image is worth 16x16 words: Transformers for image recognition at scale[J]. International Conference on Learning Representations, 2021.
- [34] LIU Z,LIN Y,CAO Y,et al. Swin transformer: Hierarchical vision transformer using shifted windows[J]. arXiv:2103.14030, 2021.
- [35] CAO H,WANG Y,CHEN J,et al. Swin-Unet;Unet-like Pure Transformer for Medical Image Segmentation[J]. arXiv:2105.05537,2021.
- [36] LI G,YU Y. Visual saliency detection based on multiscale deep CNN features[J]. IEEE Trans. Image Process. ,2016,25(11): 5012-5024.
- [37] FU H,CHENG J,XU Y,et al. Joint Optic Disc and Cup Segmentation Based on Multi-Label Deep Network and Polar Transformation 2016[J]. IEEE Transactions on Medical Imaging,2018,37(7):1597-1605.



DING Tianshu, born in 1998, postgraduate. His main research interests include artificial intelligence and medical image analysis.



CHEN Yuanyuan, born in 1983, Ph.D, associate professor, master supervisor. Her main research interests include artificial intelligence, the theory and applications of neural networks and medical image analysis.

(责任编辑:喻黎)