



DOI:10.12404/j.issn.1671-1815.2406155

引用格式:于振华,闫本聪,王迎美.基于自注意力机制和侧输出损失函数的视网膜血管分割网络[J].科学技术与工程,2025,25(21):8993-9001.

Yu Zhenhua, Yan Bencong, Wang Yingmei. Retinal vessel segmentation network based on self-attention mechanism and lateral output loss function[J]. Science Technology and Engineering, 2025, 25(21): 8993-9001.

基于自注意力机制和侧输出损失函数的 视网膜血管分割网络

于振华, 闫本聪, 王迎美*

(山东理工大学数学与统计学院, 淄博 255000)

摘要 图像分割是医学图像分析中的一个基本问题,基于深度学习的典型 UNet 架构 (UNet architecture) 分割网络模型及其变式被广泛应用于视网膜血管分割之中。但是 UNet 网络通过局部卷积模块提取图像的特征信息,难以关联图像的全局信息,无法有效捕捉像素之间的长距离依赖关系。考虑到 UNet 网络模型存在的问题和视网膜血管图像的特点,在 UNet 跳跃连接中加入注意力模块,可以捕捉血管之间的长距离依赖关系。此外,为增强网络的分割能力,使用群归一化 (group normalization, GN) 代替 UNet 网络模型原始的批归一化 (batch normalization, BN),对不同的通道选择对应的分组。为更新参数和优化网络,利用侧输出层和最后的输出层设计了交叉熵损失函数。在 DRIVE 数据集和 CHASEDB1 数据集上进行了实验,实验结果表明所提出的模型有更好的图像分割效果。

关键词 UNet; Swin-UNet; 群归一化; 侧输出层; 视网膜血管分割

中图分类号 TP391 R774; 文献标志码 A

Retinal Vessel Segmentation Network Based on Self-attention Mechanism and Lateral Output Loss Function

YU Zhen-hua, YAN Ben-cong, WANG Ying-mei*

(School of Mathematics and Statistics, Shandong University of Technology, Zibo 255000, China)

[Abstract] Image segmentation is a fundamental problem in medical image analysis, the typical deep learning based UNet architecture (UNet) and its variants are widely used in retinal vessel segmentation. However, the UNet network extracts feature information from images through local convolution modules, which makes the global information of the images difficult to be correlated and the long-distance dependencies between pixels difficult to be effectively captured. Considering the problems with the UNet network model and the characteristics of retinal vascular images, an attention module was added to the skip connections of UNet to capture long-distance dependencies between blood vessels. In addition, to enhance the segmentation ability of the network, the group normalization (GN) was used instead of the original batch normalization (BN) of the UNet network model, and the corresponding groups were selected for different channels. To update parameters and optimize the network, the final cross entropy loss function was designed using the side output layer and the final output layer. Experiments are implemented on the DRIVE dataset and CHASEDB1 dataset, and the experimental results show that the proposed model has better image segmentation performance.

[Keywords] UNet; Swin-UNet; group normalization; lateral output layers; retinal blood vessel segmentation

眼科医学作为医学领域中的一个重要分支,在几十年来取得了令人瞩目的进展。随着人口老龄化和生活方式的变化,眼科疾病的发病率呈上升趋势,因此对于眼科疾病的早期诊断和治疗提出了更高的要求^[1]。在这一背景下,视网膜血管分割技术作为眼底图像处理的关键环节^[2],成为学者们的研究热点。

眼底图像是医生获取患者眼部信息的重要工具之一,一张正常的眼底图像主要包括血管、视盘和黄斑3个部分,并且正常的眼底呈现橘红色,眼底区域呈现出明亮的微黄色。通过观察血管结构、视盘大小及形状、黄斑和出血情况等,医生可以判断眼部疾病的类型并给出对应治疗方案。然而眼底图像^[3]中复

收稿日期:2024-08-17 修订日期:2025-04-10

基金项目:山东省自然科学基金面上项目(ZR2022MA027)

第一作者:于振华(2000—),男,汉族,山东滨州人,硕士研究生。研究方向:医学图像处理。E-mail:18766638125@163.com。

*通信作者:王迎美(1987—),女,汉族,山东新泰人,博士,副教授。研究方向:医学图像处理与重建。E-mail:yingmeiwang@sdu.edu.cn。

杂的结构和组织,尤其是视网膜血管网络的复杂性,使得人工分析和诊断变得烦琐且容易出现主观误判。

基于视网膜血管结构的复杂性,传统方法在视网膜血管分割中表现并不是很好。近年来随着深度学习技术的发展,使用网络来解决视网膜血管分割的方法越来越多。UNet 架构(UNet architecture)网络是图像分割领域的一个重要网络,该网络及其变式被广泛应用于各种图像分割任务之中。Ronneberger 等^[4]提出了由编码器和解码器构成的 UNet 网络,该网络使用局部卷积来提取图像的特征信息,并通过跳跃连接融合特征。近年来不断有学者对 UNet 网络进行改进^[5],包括在 UNet 网络中加入残差结构^[6]、注意力模块^[7]、Inception 模块^[8]等。田会娟等^[9]将残差结构加入 UNet 网络的编码器结构之中提出了 SRN-UNet (squeeze and excitation resnext UNet)网络,该网络可以激励图像的特征结构并抑制噪声,被用于低质量的虹膜图像分割之中。添加注意力模块的 UNet 网络结构包括在编码器结构中引入注意力机制,在解码器或跳跃连接中引入注意力机制等。郭宁等^[10]将注意力门控添加到了网络的解码器,使得网络增加了对特征区域和结构的关注,减少了网络在分割图像时受图像背景的影响,在肺部图像的分割中有着较好的效果。除了添加各种模块以外,还有部分学者将 UNet 网络与其他网络结合构造了一些复合网络。Vaswani 等^[11]提出了用于文字处理的 Transformer 结构,该结构是一种自注意力模块,可以使得网络更好地挖掘输入信息的全局特征,一开始被用于语言处理之中^[12]。Cao 等^[13]将 Transformer 结构与 UNet 模型相结合,提出了用于图像分割的 Swin-UNet (Swin UNet architecture)网络模型,同时通过添加移动窗口解决了 Transformer 结构在大型图像分割上计算量过大的问题,使得 Transformer 结构在图像分割领域得到了很好的应用。

除了在 UNet 网络中添加各种模块,还有一些学者对其连接方式进行改进。Huang 等^[14]为弥补 UNet 模型各尺度信息利用不充分的情况,设计了具有全尺度跳跃链接的 UNet3+ 模型。Xiang 等^[15]在编码器和解码器之间建立 O 形循环路径,这种路径可以多次递归提升网络的性能。

视网膜血管图像由主干血管和由主干血管延申出的众多分支血管构成,通过 UNet 网络模型进行视网膜血管分割时有以下问题亟待解决:一是视网膜血管分割的准确率较低,即在分割过程中网络不能很好的区分分割的主体和背景。二是对微小血管的分割效果较差,分割出的细小血管存在断裂、无法连续的现象。

分析上述亟待解决的问题,在视网膜血管的分割任务中分割主体和背景的像素数量存在巨大的差别,血管结构仅占眼底图像的 7% 左右,这容易引起网络对血管的关注不够,造成视网膜血管分割的准确率较低。一些细小血管的宽度仅由单像素构成,在网络分割的过程中很难识别该部分血管,并且血管的弯曲程度以及不同分支血管之间的重叠都会影响网络对血管的分割,造成细小血管分割较差。

针对上述问题,现添加可以关联全局信息的 Transformer 结构来捕捉血管之间的长距离依赖关系,通过局部卷积模块提取局部细节信息。为了增强视网膜血管图像的对比度,在数据预处理部分增加了原始图像的伽马矫正^[16]。为进一步提高网络血管分割的准确率,使用群归一化(group normalization, GN)代替批归一化(batch normalization, BN),使网络在训练中有更好的效果,从而保证网络的分割能力。此外,充分利用解码器的侧输出层,通过侧输出层构造损失函数。为了验证本文网络的分割能力,在 DRIVE 数据集和 CHASEDB1 数据集中进行实验。

1 模型原理

1.1 UNet

UNet 网络模型由 3 个部分构成,分别是编码器、解码器以及跳跃连接。编码器结构共有 4 层,每一层都是由两步相同的操作组成,即使用大小为 3 的卷积核进行卷积,BN 归一化以及 Relu 函数激活。然后进行下采样,同样进行这些操作,反复 4 次后网络就会得到一组只有原图 1/16 的特征图。

解码器首先将特征图进行上采样,每一次上采样将特征图大小扩大一倍,然后拼接编码器结构所提取的特征图,再进行下一步的卷积等操作。经过 4 次操作后,网络会输出与输入图像相同大小的分割图像。UNet 网络通过这种编码器解码器结构能够较好地挖掘图像的细节和结构信息,但是仅仅通过卷积模块进行信息的挖掘无法很好地挖掘图像的全局特征。在视网膜血管图像中,视网膜血管网络呈现复杂的树状结构,拥有丰富的分支,每一根细小的血管都依附于主干血管,同时细小血管与背景之间的对比度较弱。在视网膜血管分割的任务中,考虑图像的全局特征尤为重要,用于视网膜血管分割的网络应该具有图像的全局信息挖掘能力。

1.2 Transformer 结构

Transformer 结构首次在自然语言处理领域被提出,该结构主要由输入、多层编码器、多层解码器以及输出 4 部分构成。其中输入包括单词嵌入和位置编码^[17]处理,输出包括输出线性层以及 Softmax 激

活函数处理。该结构的核心在于它的编码器结构和解码器结构中的自注意力模块 (self-attention) 和多头注意力模块 (multi-head attention)。

自注意力模块通过查询向量 Q 、键向量 K 和值向量 V 来计算注意力,计算公式为

$$\begin{cases} Q = XW^Q \\ K = XW^K \\ V = XW^V \end{cases} \quad (1)$$

$$\begin{cases} X \in \mathbf{R}^{m \times D} \\ W^Q, W^K \in \mathbf{R}^{D \times d_k} \\ W^V \in \mathbf{R}^{D \times d_v} \end{cases} \quad (2)$$

式中: X 为输入的特征向量矩阵; D 为每个特征向量的维度; d_k 为 Q 和 K 中每个单词嵌入的维度; d_v 为 V 中每个单词嵌入的维度; m 为特征向量的数量。该结构通过将输入的 X 传递进一个线性层进行处理计算出 Q, K, V , 即式(1)的过程。然后通过得到的 Q, K, V 进行注意力的计算,公式为

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (3)$$

式(3)中:

$$\begin{cases} Q \in \mathbf{R}^{m \times d_k} \\ K \in \mathbf{R}^{m \times d_k} \\ V \in \mathbf{R}^{m \times d_v} \end{cases} \quad (4)$$

自注意力模块能够关联语句中单词之间的长距离关系。多头注意力模块是对自注意力模块的进一步拓展,该模块使用了多个式(1)中的权重矩阵 W^Q, W^K, W^V 来计算多个 W_i , 然后使用这些分段后的 W_i 计算 Q, K, V , 然后通过式(3)计算注意力,最后将这些计算后的注意力拼接起来,计算公式为

$$\text{MultiHeadAttention}(Q, K, V) = \text{Concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_n)W^p \quad (5)$$

式(5)中: $\text{head}_1, \text{head}_2, \dots, \text{head}_n$ 为使用不同的 W_i 计算出的 Attention, n 为使用权重矩阵的数量; Concat 为将多个 head_i 进行拼接; W^p 为一个矩阵,该矩阵的作用是将结果恢复到原本的维度。

1.3 Swin-Unet

Transformer 模型被广泛地应用在自然语言处理任务之中, Dosovitskiy 等^[11]将该模型应用于图像识别任务中。在该模型最初并不能广泛应用于图像处理任务中,由于在 Transformer 的位置编码中,将图像进行编码形成的向量过长,计算量过大。Swin-UNet 模型^[13]改进了 Transformer,通过移动窗口^[18]进行自注意力模块计算,大大减少了计算量,使得该模型可以在大型图片上进行运算,提高了 Transformer 结构的实用性。

Swin-UNet 网络采用了 UNet 的 U 形结构,由编码器、解码器及其之间的跳跃连接构成。该网络的基础模块是 Swin Transformer Block,利用该模块进行特征的提取,通过移动窗口进行 Transformer 处理,即在不同的窗口中计算注意力进一步提取图像的特征,该方法使得 Transformer 结构可以在较大的图片中使用,捕捉图片中的全局信息。

此外,该网络使用 Patch Merging 模块和 Patch Expanding 模块进行下采样、上采样和通道数增减操作。每进行一次 Patch Merging 操作,特征块的大小将变为原来的一半,通道数增加一倍, Patch Expanding 则将特征块的大小翻倍,通道数减半。Skip Connection 则是将编码器和解码器对应层次的特征进行拼接。Swin-Unet 网络完全使用 Transformer 结构提取图像的特征,然而该结构无法像卷积神经网络 (convolutional neural network, CNN) 结构一样捕捉图像局部细节特征,并且在较小的数据集上并没有很好的效果。

2 基于 Transformer 和 UNet 的网络结构

2.1 网络结构

提出一种基于 Transformer 和 UNet 的网络结构,命名为 AttenTransUnet,该网络结构既可以保证局部卷积对图像细节特征的提取能力,又可以兼顾 Transformer 模块联系全局的能力,网络结构如图 1 所示。首先对视网膜图像的绿色通道图像进行伽马矫正实现图像增强,然后将增强后的图像作为网络输入。该网络结构使用 UNet 网络的编码器、解码器及跳跃连接结构。编码器部分包括 4 层,每一层包括两步卷积核大小为 3 的卷积操作,GN 归一化和 Relu 激活。然后经过下采样后进入下一层,重复操作 4 次后进入解码器,解码器包括卷积核大小为 3 的卷积、GN 归一化、Relu 激活和上采样操作。

考虑到 UNet 网络在跳跃连接中将不同层次的特征图直接拼接时,可能会由于不同特征的语义差导致网络性能下降,同时单纯的 CNN 结构无法关联图像的全局信息,为减缓上述问题以及提高网络的全局信息关联能力,设计了简化 Swin-Unet 的 U 形结构作为编码器和解码器特征图之间的跳跃连接,命名为 Attention Block。即在特征图被卷积模块处理之后,会进入 AttentionBlock 模块,该操作会捕捉特征图中像素之间的全局联系,并且进行处理后的特征图再进行拼接会消除由直接拼接带来的影响,从而提高网络的分割能力。

所提网络结构充分利用侧输出层特征信息来构造损失函数,如图 1 右侧所示。通过上采样将解

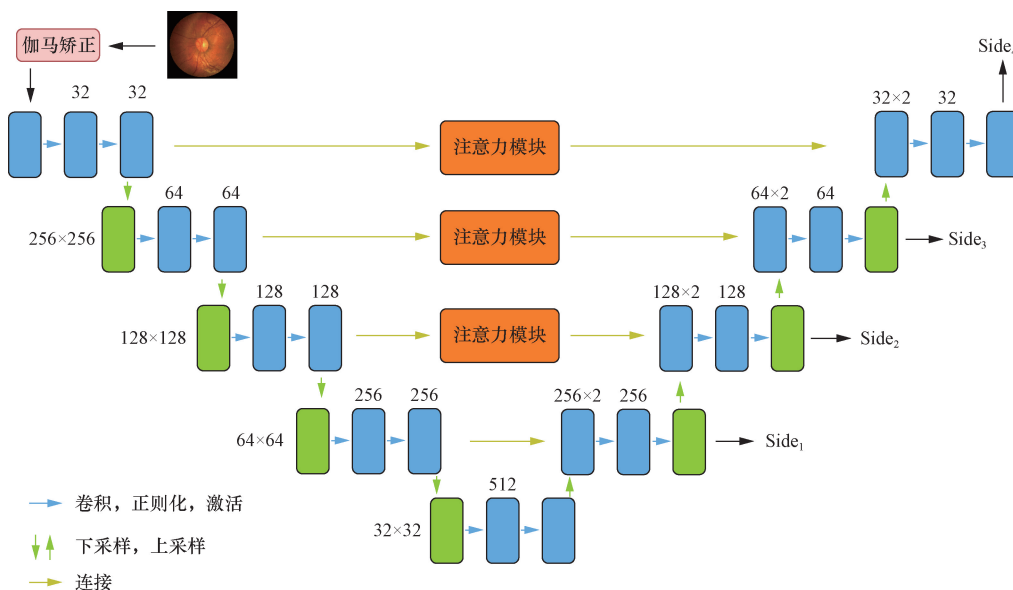


图1 AttenTransUnet 的网络结构
Fig. 1 The network structure of AttenTransUnet

码器中每一层的特征图与输入图像大小保持相同, 然后计算交叉熵损失函数。

2.2 伽马矫正

伽马矫正通过改变像素值的分布来实现对图像对比度增强, 公式表示为

$$\text{Output} = \left(\frac{\text{Input}}{\text{Max}} \right)^\gamma \text{Max} \quad (6)$$

式(6)中: Input 为输入图像像素的灰度值; Max 为输入图像像素灰度值的最大值; Output 为对应于输入图像位置的输出像素灰度值; γ 为伽马值, 通过调整 γ 大小可以调整图像亮暗区域的对比度。当 $\gamma > 1$ 时拉伸亮区的灰度范围起到增强亮区图像的对比度的作用; 当 $\gamma < 1$ 时拉伸暗区图像的灰度范围, 使图像暗区对比度增加。

针对视网膜血管图像, 实验数据预处理使用 $\gamma = 1.5$ 的伽马矫正进行了图像增强。通过实验证明,

对绿色通道使用伽马矫正相比于对其他通道或是三通道同时进行伽马矫正有着更好的效果, 因此使用绿色通道伽马矫正后的图像作为网络输入。

2.3 注意力模块

原始 UNet 网络直接将编码器特征图与解码器特征图进行特征融合, 由于编码器特征图与解码器的特征图包含着不同级别的信息, 这种特征图之间的信息差会导致网络分割能力下降。针对这种问题, Zhou 等^[19] 提出了 UNet++ 模型, 该模型通过在跳跃连接中添加卷积块的方式对 UNet 网络进行了改进, 减弱了直接跳跃连接带来的影响。

考虑到直接拼接的问题和局部卷积无法很好地提取图像全局信息的缺点, 设计了注意力跳跃连接模块, 结构如图 2 所示。该模块是一个简化的 Swin-UNet 的 U 形结构, 包括编码器和解码器两部分组成, 编

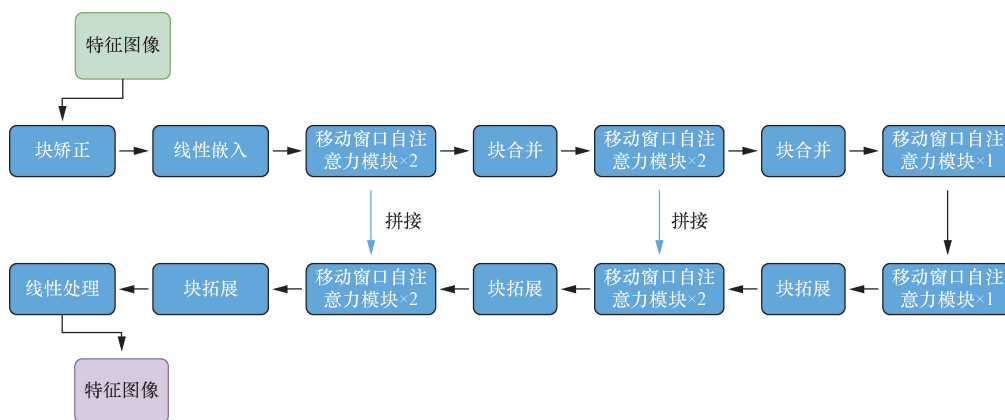


图2 Attention Block 跳跃连接模块的 U 形结构
Fig. 2 The U-shaped structure of the Attention Block skip connection module

码器由 3 层移动窗口自注意力模块 (Attention Block) 构成,该模块的原理是将整幅图片通过移动窗口分成不同的小的特征块,然后再对特征块进行 Transformer 处理,计算他们的注意力权重。每次进行完该处理后会通过块合并(patch merging)模块进行下采样以及提升特征层的数量。与编码器类似,解码器也由 3 层 Attention Block 构成,并通过块拓展(patch expending)进行上采样以及降低特征层的数量。提出的 AttenTransUNet 编码器通过 CNN 结构提取图像的细节特征,将特征图传递给 Attention Block 跳跃连接模块,该模块使用 Transformer 结构对特征图进行进一步处理,可以加强特征图中像素之间的全局联系。

通过实验对比了 Attention Block 模块深度的影响,最后将跳跃连接模块的 U 形结构确定为 3 层,并且对比了在不同位置该模块所带来的影响,发现仅在第一层、第二层和第三层加入该模块有较好的效果,同时可以减少计算量。

2.4 损失函数

为了参数优化和更好地网络训练,对原始 UNet 网络中的损失函数^[20]进行了改进。原始 UNet 模型利用最后输出层与标签图像进行交叉熵函数运算,以此作为最后的损失函数,计算公式如下。

$$\text{Lossfunction} = \text{CrossEntropy}(\text{side}_4, \text{label}) \quad (7)$$

式(7)中:side₄为网络最后输出图像;label为对应图像的标签。为了充分利用侧输出信息,将解码器中每层的输出层通过上采样处理,将图像大小恢复到与输入图像相同的尺寸,得到 side_i,*i* = 1,2,3,4。然后计算交叉熵,公式如下。

$$L_{s_1} = \text{CrossEntropy}(\text{side}_1, \text{label}) \quad (8)$$

$$L_{s_2} = \text{CrossEntropy}(\text{side}_2, \text{label}) \quad (9)$$

$$L_{s_3} = \text{CrossEntropy}(\text{side}_3, \text{label}) \quad (10)$$

$$L_{s_4} = \text{CrossEntropy}(\text{side}_4, \text{label}) \quad (11)$$

$$L_{s_5} = \text{CrossEntropy}[(\text{side}_1 + \text{side}_2 + \text{side}_3 + \text{side}_4)/4, \text{label}] \quad (12)$$

通过这些侧输出层构造了以上损失函数 L_{s_i} , $i = 1, 2, 3, \dots, 5$ 。通过消融实验,确定损失函数为

$$\text{Lossfunction} = L_{s_3} + L_{s_4} + L_{s_5} \quad (13)$$

3 实验结果与分析

3.1 数据集及其预处理与实验环境

为了验证网络分割的有效性,使用 DRIVE 数据集^[21]和 CHASEDB1 数据集^[22]进行实验。DRIVE 数据集(digital retinal images for vessel extraction)是研究视网膜病变的数据集,该数据集的图像来自荷

兰的一个糖尿病性视网膜病变^[21]的筛查项目,筛查人群是 400 名 25 ~ 90 岁的糖尿病受试者,并在图片中随机选取了 40 张图片。该数据集中的图片大小为 584 × 565,在实验中使用其中的 20 张作为训练图像,另外的 20 张作为测试图片。CHASEDB1 数据集(combined healthy and diabetic retinopathy database 1)是一个用于眼底图像分析的公开数据集,该数据集中的图像是由 14 名健康人和 14 名糖尿病性视网膜患者中采集而来,该数据集包含 28 幅视网膜图像,每张图像的大小为 999 × 960,在实验中使用 20 张图片进行训练,8 张图片进行测试。

使用伽马矫正进行了图像增强预处理,从而增强视网膜图像中血管的对比度,实验显示同时对 RGB 三色通道进行增强效果并不明显,而只对绿色通道进行图像增强会使得血管结构更加清晰,因此图 3 列出了在 DRIVE 数据集中不同通道使用 $\gamma = 1.5$ 进行矫正的结果图。除了伽马矫正,也进行了随机翻转预处理扩大数据量。

模型在 Windows 环境下实现,使用深度学习框架 PyTorch 实现所提出的网络架构。实验所用电脑硬件环境为 Windows 11 系统,搭载装有集成显卡的 Intel 酷睿 i5 1135G7 处理器。采用 Adam^[23]优化器对模型进行优化,Batch_size = 1,Transformer 模块依然使用层归一化^[24]进行归一化,其他的归一化使用 GN。总 epoch 数设为 100,初始学习率设置为 0.001,学习率随着迭代更新,每 30 次迭代缩小为原本的 1/10。

3.2 评价指标

采用以下评价参数进行各个网络视网膜血管分割能力的评估。

图像分割的准确率(accuracy, Acc)是评估图像分割算法性能的重要指标之一,它反映了算法对图像中目标物体分割的正确程度。准确率的计算通常基于像素级别的分类结果,即算法对图像中每个像素的类别判断与真实标签进行比较,计算正确分类的像素比例,公式为

$$\text{Acc} = \frac{\text{TN} + \text{TP}}{\text{TP} + \text{FN} + \text{TN} + \text{FP}} \quad (14)$$

式(14)中:TP为真正例(算法正确检测的血管像素数);FN为假负例(算法未能检测到的血管像素数);TN为真负例(算法正确排除的非血管像素数);FP为假正例(算法错误地将非血管像素标记为血管的像素数)。图像分割的灵敏度(sensitivity, Sen)计算的是分割图像中分割主体即血管像素被分割正确的比例。该参数可以体现网络模型对分割主体的识别能力,即分割网络对视网膜血管的识别能力,公式为

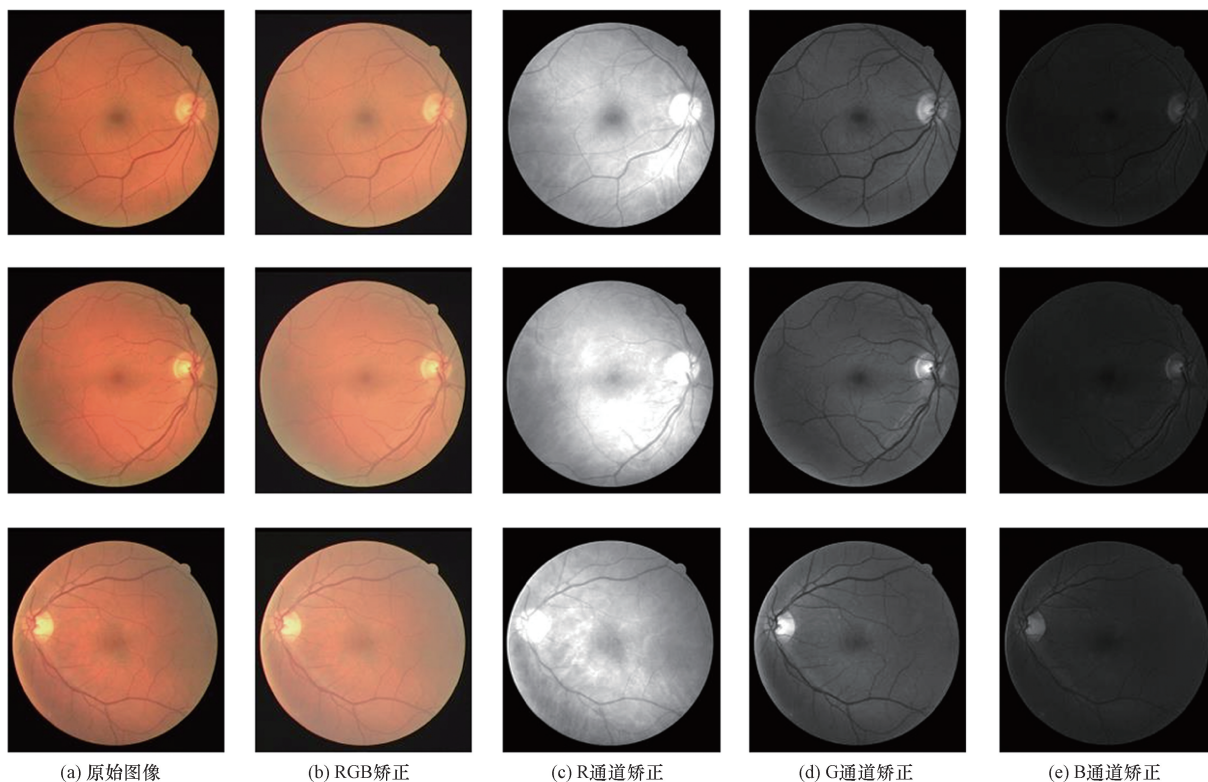


图3 不同通道的 Gamma 矫正图

Fig. 3 Gamma correction images for different channels

$$Sen = \frac{TP}{TP + FN} \quad (15)$$

图像分割中的特异度 (specificity, Spe) 是分割图像中分割背景被正确分割的比例。该参数体现的是分割网络对分割背景即非血管像素的识别能力, 计算公式为

$$Spe = \frac{TN}{TN + FP} \quad (16)$$

曲线下的面积 (area under curve, AUC) 是指受试者操作特征 (receiver operating characteristics, ROC) 曲线下面积的数值, ROC 曲线的横轴是假阳性率, 纵轴是真阳性率。AUC 的范围在 0 ~ 1, AUC 越大, 模型的分割性能就越好。当 AUC = 1 时, 表示模型可以完美地区分正例和负例, 越接近于 1, 模型的分类能力越强。当 AUC = 0.5 时, 表示模型的分类能力和随机猜测的结果相同。

3.3 对比试验分析

对比实验在 DRIVE 数据集和 CHASEBD1 数据集两个数据集中进行, 给出了 UNet 模型^[4]、TransUNet 模型^[25]、Swin-UNet 模型^[13] 以及所提出的 AttenTransUnet 模型的实验效果。通过精确率、灵敏度、特异度和 AUC 对分割效果进行了评估, 不同模型在两个数据集上的具体结果如表 1 所示。图 4 是不同网络结构的分割结果图, 选择了 DRIVE 数据集的第 1 幅图和第 7 幅图进行展示。

从图 3 可以看出, 对绿色通道进行伽马矫正的结果有着更好的效果, 即使是细小血管也有很好的对比度, 在分割结果中应该更能表现出更好的效果。在图 4 的细节图中不难看出, 在血管的连接性上, 本文网络有着更好的效果, 血管断联较少, 并且在血管分支上, 分割出了更多的细节信息。特别是

表 1 不同算法的 Acc、Sen、Spe、AUC

Table 1 Acc, Sen, Spe, AUC of different algorithms

模型	DRIVE				CHASEBD1			
	Acc	Sen	Spe	AUC	Acc	Sen	Spe	AUC
UNet	0.953 4	0.676 1	0.987 5	0.980 6	0.976 5	0.707 8	0.989 5	0.977 5
UNet + +	0.958 5	0.727 7	0.986 6	0.983 0	0.978 1	0.725 9	0.990 1	0.984 0
TransUNet	0.957 1	0.741 9	0.983 2	0.980 6	0.977 4	0.824 2	0.984 8	0.983 3
Swin-UNet	0.957 2	0.720 9	0.986 9	0.981 6	0.975 3	0.680 9	0.989 8	0.971 7
AttenTransUnet	0.960 3	0.757 7	0.984 9	0.985 0	0.979 4	0.822 1	0.987 0	0.987 3

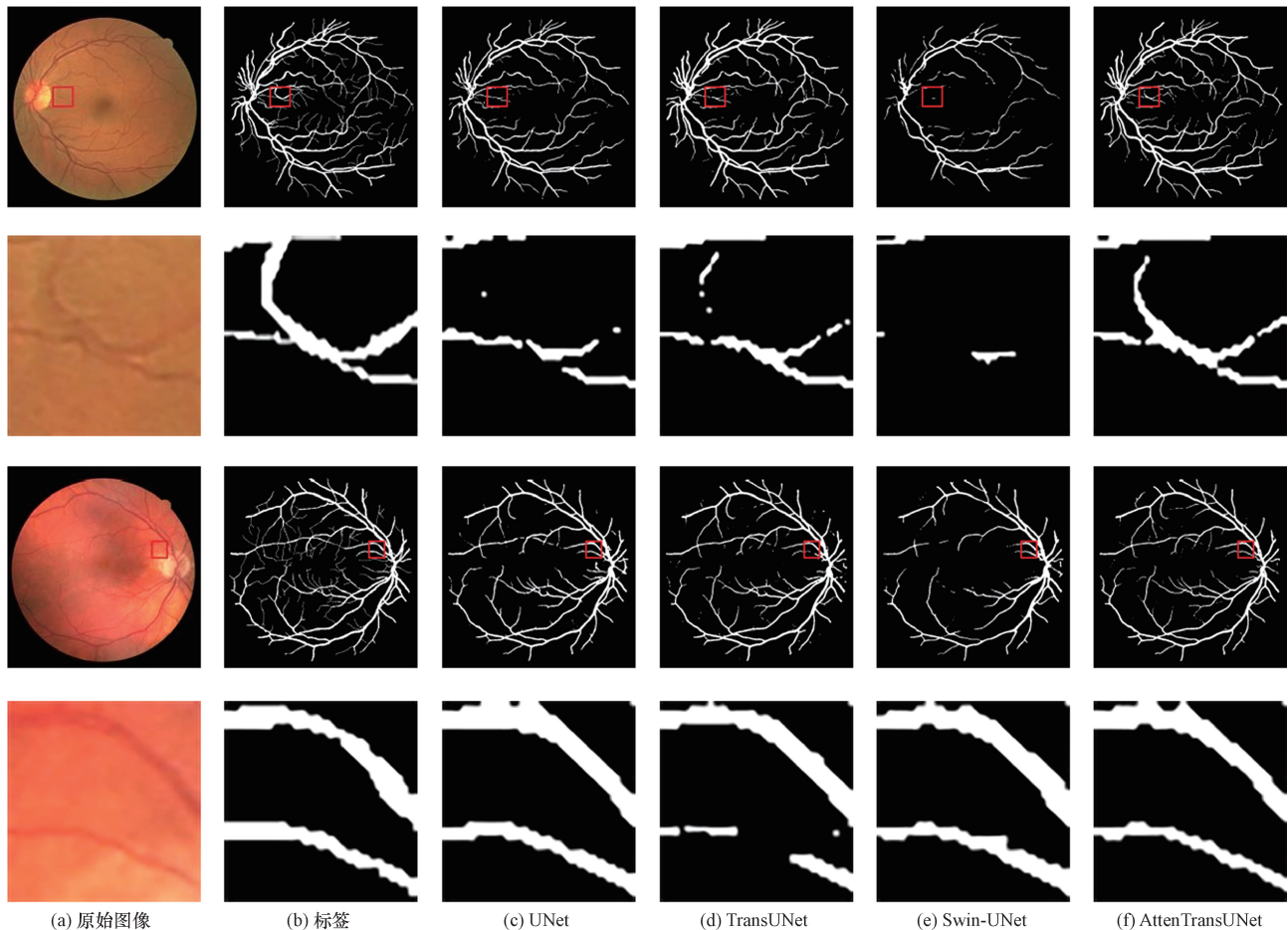


图4 各种比对算法的实验结果

Fig. 4 Experimental results of different comparison methods

从第三行和第四行图像中可以看出,本文网络模型分割出的血管与标签相近,误判的像素更少。这些结论可以通过表1中展示的各种算法的评价指标得到进一步验证,相比于同样使用注意力模块和卷积模块的TransUNet模型,本文网络模型在两个数据集上精确率分别提升了0.003 2和0.002。

3.4 消融试验及结果分析

为了验证本文网络模型设计的GN、伽马矫正、Attention Block跳跃连接的有效性,进行了消融实验,消融实验包括使用BN的UNet分割网络(模型1);使用GN的UNet分割网络(模型2);含有本文网络模型设置侧输出层的UNet网络(模型3);含有本文网络模型跳跃连接结构的UNet网络(模型4);使用伽马矫正的UNet网络(模型5)以及含有跳跃连接,GN和侧输出层的最终分割网络(模型6),在DRIVE数据集上进行训练和测试。图5是消融实验中不同模块的分割图,可以看出,进行伽马矫正后的分割效果比原始的UNet网络有着更好的分割效果,这说明该处理对网络效果的提升发挥了作用。同样,加入注意力模块的网络结构相比于之前

的网络细节信息更多。从表2可以看出本文网络模型的不同模块对整个网络结构都有着不同程度的提升。综上所述,本文网络结构在血管分割方面有着更好的效果。

为了显示本文网络模型设计的损失函数过程,对比了本文网络模型AttenTransUnet中加入不同层损失函数的参数表现,如表3所示,确定使得各项评价指标最高的损失函数为式(8)。损失函数的消融实验包括无侧输出损失函数(L_{s4} 模型),加入图2最后一层侧输出的损失函数($L_{s4} + L_{s5}$ 模型),加入图2的全部侧输出的损失函数($L_{s1} + L_{s2} + L_{s3} + L_{s4} + L_{s5}$ 模型),加入图2的后三层侧输出的损失函

表2 消融实验的Acc、Sen、Spe、AUC
Table 2 Acc, Sen, Spe, AUC of ablation experiments

模型	Acc	Sen	Spe	AUC
模型1	0.953 4	0.676 1	0.987 5	0.980 6
模型2	0.958 5	0.738 7	0.985 2	0.982 8
模型3	0.957 3	0.738 3	0.984 1	0.982 9
模型4	0.958 0	0.744 5	0.984 0	0.982 3
模型5	0.958 6	0.728 5	0.986 3	0.983 2
模型6	0.960 3	0.764 4	0.984 1	0.985 3

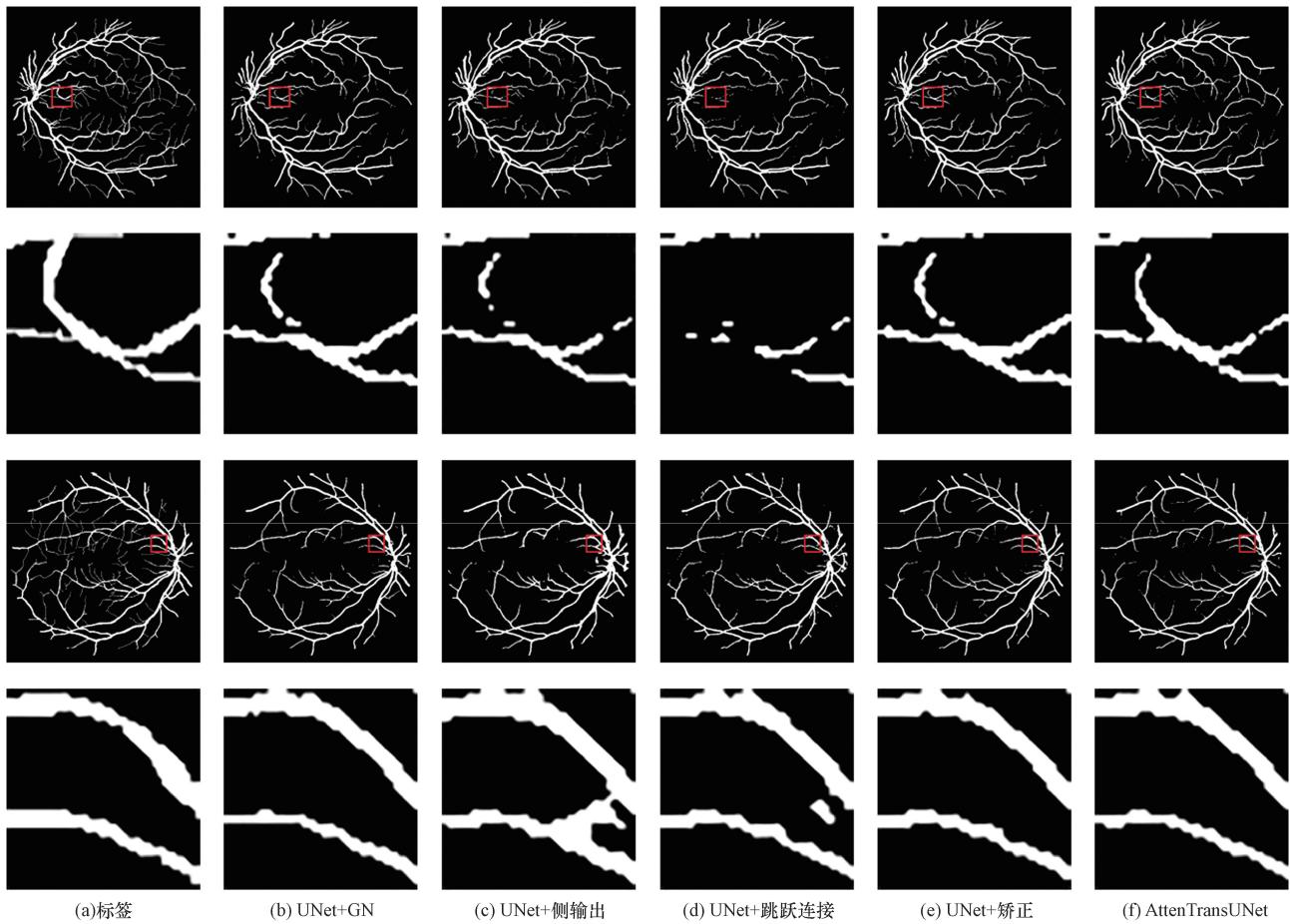


图5 消融实验结果

Fig. 5 Results of ablation experiments

表3 不同损失函数的 Acc、Sen、Spe、AUC

Table 3 Acc, Sen, Spe, AUC of different loss functions

模型	Acc	Sen	Spe	AUC
L_{s4}	0.953 4	0.676 1	0.987 5	0.980 6
$L_{s4} + L_{s5}$	0.956 8	0.727 6	0.984 8	0.980 9
$L_{s2} + L_{s3} + L_{s4} + L_{s5}$	0.957 1	0.711 6	0.987 0	0.981 8
$L_{s1} + L_{s2} + L_{s3} + L_{s4} + L_{s5}$	0.956 3	0.708 4	0.986 6	0.980 0
$L_{s3} + L_{s4} + L_{s5}$	0.957 3	0.738 3	0.984 1	0.982 9

数($L_{s2} + L_{s3} + L_{s4} + L_{s5}$ 模型),加入图2后两层的侧输出的损失函数($L_{s3} + L_{s4} + L_{s5}$ 模型)。

4 结论

提出了基于自注意力模块和侧输出损失函数的视网膜血管分割网络结构。自注意力模块可以加强血管之间的全局联系,增强血管的连接性。通过侧输出层构建的损失函数可以提高网络的分割能力。在进行分割之前,使用了绿色通道的伽马矫正,该过程增强了血管信息的对比度,使得网络可

以更好地区别血管结构,从而增强网络的分割效果。同时在训练过程中,在网络中使用 GN 使得网络在卷积之后有更好的归一化结果,从而提升了网络的分割效果。

参 考 文 献

- [1] Soleimani M, Cheraqpour K, Koganti R, et al. Cellular senescence and ophthalmic diseases; narrative review[J]. Graefe's Archive for Clinical and Experimental Ophthalmology, 2023, 261 (11): 3067-3082.
- [2] Yu L, Qin Z, Zhuang T, et al. A framework for hierarchical division of retinal vascular networks[J]. Neurocomputing, 2020, 392: 221-232.
- [3] Qin Q, Chen Y. A review of retinal vessel segmentation for fundus image analysis[J]. Engineering Applications of Artificial Intelligence, 2024, 128: 107454.
- [4] 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. Berlin: Springer International Publishing, 2015: 234-241.
- [5] 汪南洋, 沈疆海. 基于 MSHAM-UNet 的岩心孔洞图像分割方法[J]. 科学技术与工程, 2024, 24(24): 10362-10369. Wang Nanyang, Shen Jianghai. Image segmentation method of rock

- core hole based on MSHAM-UNet[J]. *Science Technology and Engineering*, 2024, 24(24): 10362-10369.
- [6] Zhang K, Sun M, Han T X, et al. Residual networks of residual networks: multilevel residual networks[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2017, 28(6): 1303-1314.
- [7] Baozhou Z, Hofstee P, Lee J, et al. An attention module for convolutional neural networks[J]. *ArXiv*, 2021: 2108.08205.
- [8] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision[C]//*IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas: IEEE, 2016: 2818-2826.
- [9] 田会娟, 翟佳豪, 柳建新, 等. 基于SRN-UNet的低质量虹膜分割算法[J]. *光子学报*, 2022, 51(2): 241-249.
Tian Huijuan, Zhai Jiabao, Liu Jianxin, et al. A low-quality iris image segmentation algorithm based on SRN-UNet[J]. *Acta Photonica Sinica*, 2022, 51(2): 241-249.
- [10] 郭宁, 柏正尧. 注意力机制下密集空洞卷积的肺部图像分割[J]. *中国图象图形学报*, 2021: 26(9): 2146-2155.
Guo Ning, Bai Zhengyao. The integration of attention mechanism and dense atrous convolution for lung image segmentation[J]. *Journal of Image and Graphics*, 2021, 26(9): 2146-2155.
- [11] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. *ArXiv Preprint ArXiv*, 2017: 1706.03762.
- [12] 汤翔中, 高丙朋. 融合注意力空洞卷积和Transformer的矿石图像分割[J]. *科学技术与工程*, 2023, 23(16): 6974-6982.
Tang Xiangzhong, Gao Bingpeng. Ore image segmentation based on attention hole convolution and Transformer[J]. *Science Technology and Engineering*, 2023, 23(16): 6974-6982.
- [13] Cao H, Wang Y, Chen J, et al. Swin-UNet: UNet-like pure Transformer for medical image segmentation[C]//*European Conference on Computer Vision*. Cham: Springer Nature Switzerland, 2022: 205-218.
- [14] Huang H, Lin L, Tong R, et al. UNet 3+: a full-scale connected UNet for medical image segmentation[C]//*ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Barcelona: IEEE, 2020: 1055-1059.
- [15] Xiang T, Zhang C, Liu D, et al. BiO-Net: learning recurrent bi-directional connections for encoder-decoder architecture[C]//*Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*. Berlin: Springer, 2020: 74-84.
- [16] Ibrahim H, A. Mohamed A E N, Ammar R, et al. Efficient color image enhancement using piecewise linear transformation and gamma correction[J]. *Journal of Optics*, 2024, 53(3): 2027-2037.
- [17] Wu K, Peng H, Chen M, et al. Rethinking and improving relative position encoding for vision transformer[C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*. Montreal: IEEE, 2021: 10033-10041.
- [18] Song Z, Yu J, Chen Y P P, et al. Transformer tracking with cyclic shifting window attention[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New Orleans: IEEE, 2022: 8791-8800.
- [19] Zhou Z, Siddiquee M M R, Tajbakhsh N, et al. UNet++: redesigning skip connections to exploit multiscale features in image segmentation[J]. *IEEE Transactions on Medical Imaging*, 2019, 39(6): 1856-1867.
- [20] Barron J T. A general and adaptive robust loss function[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach: IEEE, 2019: 4331-4339.
- [21] Staal J, Abràmoff 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.
- [22] Owen C G, Rudnicka A R, Mullen R, et al. Measuring retinal vessel tortuosity in 10-year-old children: validation of the computer-assisted image analysis of the retina (CAIAR) program[J]. *Investigative Ophthalmology & Visual Science*, 2009, 50(5): 2004-2010.
- [23] Tan T E, Wong T Y. Diabetic retinopathy: looking forward to 2030[J]. *Frontiers in Endocrinology*, 2023, 13: 1077669.
- [24] Ba J L, Kiros J R, Hinton G E. Layer normalization[J]. *ArXiv Preprint ArXiv*, 2016: 1607.06450.
- [25] Chen J, Lu Y, Yu Q, et al. TransUNet: Transformers make strong encoders for medical image segmentation[J]. *ArXiv Preprint ArXiv*, 2021: 2102.04306.