

# 基于引导滤波和 Transformer 的双分支视网膜血管分割网络

闫本聪<sup>1</sup>, 王迎美<sup>1,2\*</sup>

(1. 山东理工大学数学与统计学院, 山东 淄博 255049; 2. 山东大学数学学院, 山东 济南 250100)

**摘要:**引入 Transformer 结构和 UNet++ 网络, 提出一个新的双分支视网膜血管分割网络, 该网络中的双分支编码器可以更好地关联图像中的全局信息, 使得整个网络在小数据集的训练中也有较好的效果。在此基础上, 为了进一步解决 UNet 网络中下采样操作导致的视网膜血管信息丢失问题, 在输出层和网络第二层的特征图引入引导滤波, 可以有效提高小血管分割精度。该网络使用 DRIVE 数据集(digital retinal images for vessel extraction)和 CHASEDB1 数据集(combined healthy and diabetic retinopathy database 1)进行实验, 在精确度、灵敏度等参数上有较大的提升, 并且血管分割图中正确分割出更多细小血管, 总体表现出较好的效果。

**关键词:** UNet++; Transformer; 引导滤波; 视网膜血管分割

**中图分类号:** O29 **文献标志码:** A

**引用格式:** 闫本聪, 王迎美. 基于引导滤波和 Transformer 的双分支视网膜血管分割网络[J]. 山东大学学报(理学版), 2025, 60(7): 22-31.

## A dual branch retinal vessel segmentation network based on guided filtering and Transformer

YAN Bencong<sup>1</sup>, WANG Yingmei<sup>1,2\*</sup>

(1. School of Mathematics and Statistics, Shandong University of Technology, Zibo 255049, Shandong, China; 2. School of Mathematics, Shandong University, Jinan 250100, Shandong, China)

**Abstract:** Based on the Transformer structure and the UNet++ network, a new dual branch retinal vessel segmentation network is proposed. The dual branch encoder in the network can better correlate global information in the image, so that the proposed network performs well on small datasets. Furthermore, in order to solve the problem of retinal vessel information loss caused by downsampling operations in UNet, guided filtering is introduced in the feature maps of the output layer and the second layer, which effectively improves the accuracy of small vessel segmentation. The effectiveness of the network is validated by experiments on the DRIVE (digital retinal images for vessel extraction) dataset and the CHASEDB1 (combined healthy and diabetic retinopathy database 1) dataset, which shows significant improvements in accuracy, sensitivity, and other parameters. In addition, more small blood vessels are visually segmented with better accuracy. All results demonstrate that the proposed method has better performance.

**Key words:** UNet++; Transformer; guided filtering; retinal vascular segmentation

## 0 引言

视网膜血管分割技术在眼科医学中具有重要的意义。首先, 它为眼科医生提供了高效而准确的眼底图像分析工具, 大大缩短了诊断时间, 提高了诊断的精确性; 其次, 该技术对于眼科疾病的早期诊断至关重要, 许多眼科疾病, 如糖尿病视网膜病变<sup>[1]</sup>和青光眼<sup>[2]</sup>, 早期症状不明显, 但通过视网膜血管分割技术, 医生可以在病情进展到不可逆的阶段前进行干预, 降低治疗难度; 此外, 该技术还为眼科研究提供了大量

收稿日期: 2024-03-01; 网络出版时间: 2025-03-11

基金项目: 山东省自然科学基金资助项目(ZR2022MA027); 中山大学广东省计算科学重点实验室开放基金项目(2021003); 山东省属普通本科高校教师访学研修项目

第一作者: 闫本聪(2000—), 男, 硕士研究生, 研究方向为医学图像处理. E-mail: 1720186576@qq.com

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

的图像数据,为深入研究眼科疾病的发病机制和治疗方法提供了宝贵的资源。随着计算机科学和医学技术的飞速发展,视网膜血管分割技术的研究进入了一个快速演进的时代。早期的研究主要集中在传统的图像处理方法上,包括滤波、边缘检测<sup>[3]</sup>和数学形态学<sup>[4]</sup>等。然而,这些方法在处理复杂的眼底图像和应对不同病变情况时存在一定的局限性。近年来,深度学习技术的兴起为视网膜血管分割<sup>[5]</sup>领域注入了新的活力。卷积神经网络(convolutional neural network, CNN)等深度学习模型通过对大量标注数据的学习,实现了对眼底图像中血管的高度自动化分割,这一创新性的方法为提高视网膜血管分割的准确性和鲁棒性<sup>[6]</sup>提供了新的可能性。

Ronneberger 等<sup>[7]</sup>提出的 UNet 网络被广泛应用于图像分割<sup>[8]</sup>任务中,UNet 网络结构以其特有的 U 形拓扑结构而著称,由编码器和解码器组成,通过跳跃连接有效地解决了传统分割模型中分辨率损失的问题。针对 UNet 网络中跳跃连接部分出现的语义差异问题,通过在跳跃连接中加入不同数量的融合多尺度信息的卷积块优化 UNet 网络,提出了 UNet++ 网络模型<sup>[9]</sup>。无论是 UNet 网络还是 UNet++ 网络都是通过 CNN 结构<sup>[10]</sup>提取图像特征,但是这种方式忽略了图像中像素之间的全局联系。Vaswani 等<sup>[11]</sup>提出用于自然语言处理任务的 Transformer 模型,该模型通过引入自注意力机制成功地捕获了输入序列中不同位置之间的依赖关系,从而在翻译等任务中取得了卓越的性能。随着对 Transformer 模型的理解深入,研究者们开始考虑将其应用于计算机视觉任务。通过将一张图片转化成图像块进而转化成序列的方式,Transformer 结构被应用于图像识别任务中<sup>[12]</sup>。TransUNet<sup>[13]</sup>网络将 Transformer 结构应用在 UNet 网络的编码器之中,建立了一个新的编码器结构使得网络可以关联图像的全局特征。在此基础上,TransResUNet 模型<sup>[14]</sup>将残差块引入了 TransUNet 网络模型。TransUNet 模型在解码器的特征提取中依然使用了 CNN 结构,SwiN-UNet 模型<sup>[15]</sup>使用 Transformer 结构完全代替了 CNN 结构进行特征提取,并且通过加入移动窗口的方法解决了 Transformer 结构在处理大型图片中计算量大的问题。尽管 Transformer 结构可以很好地关联图像的全局特征,但是该结构对大规模数据的需求较大,这在一些特定领域(如医学图像分割)可能会受到数据获取的限制。因此如何在小样本数据上训练出具有较好泛化性能的 Transformer 模型是一个亟待解决的问题。

针对上述模型出现的问题,本文结合 Transformer 结构和 UNet++ 模型提出了一种双分支编码器的 U 型网络结构,该结构不仅可以在视网膜分割中考虑像素之间的全局联系,并且在小型数据集上也表现出了较好的性能。同时,因为眼底图像中包含了大量的单像素级别的血管结构,这些血管结构在使用 UNet 网络模型进行视网膜血管分割时会因为下采样操作出现信息丢失的问题,导致这些单像素级别的信息难以恢复。因此本文在网络的浅层特征图中加入了引导滤波<sup>[16]</sup>,通过引导滤波恢复丢失的信息。本文研究的主要贡献如下:

- 1) 结合 Transformer 结构和 UNet++ 模型建立了双分支编码器结构,提高了网络的信息提取和关联全局信息的能力,并且使得 Transformer 结构在小数量训练集中也能表现出较好的性能。
- 2) 在 2 个分支的编码器与单一解码器的跳跃连接中加入了融合多尺度信息的 CNN 结构,使得网络可以更好地融合不同尺度编码器与解码器之间的特征信息。
- 3) 在网络的浅层特征图和输出层中使用输入图像的灰度图作为引导图像对特征图进行引导滤波,进而恢复在网络处理中使用下采样操作丢失的图像信息。

## 1 模型原理

### 1.1 UNet++

UNet++ 网络是对 UNet 网络的一种改进,UNet 网络的一个关键点是采用了一种独特的编码器-解码器结构,使其在特征提取和信息重建方面表现卓越。编码器负责捕获输入图像中的高级语义特征,而解码器则通过上采样将这些特征映射还原到原始输入图像的分辨率。这种结构的引入旨在克服传统方法中常见的分辨率损失问题,确保网络在分割任务中保持高分辨率的特征表达。编码器部分通常由卷积层和池化层交替构成,通过卷积操作提取图像的局部特征,然后通过池化层进行下采样,逐渐缩小特征图的尺寸,有助于网络逐渐理解图像的全局结构;解码器部分则负责将编码器提取的高级语义特征映射还原到原始输入图像的分辨率,通常由上采样层和反卷积层交替构成,通过这一过程,网络能够逐渐还原图像的细节信息。

U-Net 网络中的另一个关键设计元素是跳跃连接。这种连接方式将编码器中的某一层的输出直接与解码器对应层的输入相连接,形成一种直达通道,有助于维持网络中的局部信息,避免了在信息传递过程中的丢失。跳跃连接使得底层和高层的特征能够进行交流,从而更好地保留了图像的细节信息,有力地解决了传统分割网络中对细节捕捉能力不足的问题。UNet 网络将底层特征和高层特征直接进行融合,由于不同层次的特征图之间存在语义差距,这一操作会影响特征之间的融合。UNet++网络通过在跳跃连接中加入融合不同尺度信息的卷积块有效解决了这一问题,该网络通过底层特征与高层特征的差异大小加入了不同数量的卷积块,从而使得网络表现出更好的性能。

## 1.2 TransUNet

### 1.2.1 Transformer 结构

Transformer<sup>[11]</sup> 结构的核心创新之一是引入了自注意力机制,这是一种全新的关注机制,使得模型能够在处理输入序列的不同位置时,根据其他位置的信息进行加权聚合,能够有效地捕获长距离依赖关系,使得模型能够更好地理解序列中的全局信息。自注意力机制的计算过程可以分为如下步骤。

1) 计算注意力分数(attention scores)。对于输入序列中的每个位置,通过学习得到的权重,计算该位置与其他所有位置之间的关联度。

2) 权重归一化。将注意力分数进行 Softmax<sup>[17]</sup> 归一化,以确保每个位置的权重在 0 到 1 之间,并且所有位置的权重之和为 1。

3) 加权求和。使用归一化后的权重对所有位置的特征进行加权求和,得到该位置的最终表示。

为了进一步提升模型的表达能力,Transformer 引入了多头注意力机制,这种机制通过同时学习多组不同的自注意力权重,使得模型能够在不同的“注意”方面进行学习。每个头的注意力机制都能够捕获输入序列中的不同依赖关系,最终通过拼接或平均等操作,得到更丰富的表示。多头注意力机制不仅提高了模型的表达能力,而且有助于模型学习不同尺度的特征,使其更适用于图像分割任务中处理多尺度信息的需求,为 Transformer 模型在图像分割中的应用提供了更强大的能力。其次,Transformer 还包括了位置编码以及残差连接等操作。

### 1.2.2 编码器

在 TransUNet<sup>[12]</sup> 的编码器中,首先使用一个大小为 7 的卷积核进行处理,并进行正则激活和池化操作,使图片大小变为原本的一半,之后使用 ResNet50<sup>[18]</sup> 进行连续 3 次的特征提取,期间的每一层操作都会将图像的大小变为原本的一半,最终将得到一个只有原本十六分之一大小的特征图,然后通过 Transformer 结构对提取后的特征图进行处理,在该过程中网络会开始关联图像的全局特征。通过 Transformer 结构处理之后网络开始进行解码器的处理,与 UNet 的解码器使用相同的结构,在解码过程中,网络会直接拼接编码过程中的特征图,将图像的底层特征和高层特征直接进行拼接直到输出最后的图像。

## 1.3 引导滤波

引导滤波<sup>[16]</sup> 是一种基于局部的保边平滑算子,该滤波算子的实现方式是通过设置引导图像进行的,首先选取一幅图像作为引导图像,该图像可以是原图像对应的灰度图像,也可以是其他图像,然后利用该引导图像与被滤波图像之间的差异建立滤波器,通过这种方法建立的滤波器可以很好地考虑到引导图像的结构,从而起到保持边缘的平滑效果。

接下来定义引导滤波器,假设输出图像  $O$  是引导图像  $G$  在一个以像素  $p$  为中心的窗口  $W_p$  上的线性变换,图像  $O$  和  $G$  的关系可以表示为

$$O_i = a_p G_i + b_p, \quad \forall i \in W_p,$$

其中  $i$  为遍历引导图像  $G$  中  $W_p$  窗口下的像素。那么引导滤波算法的目标就是求解  $a_p$  和  $b_p$ , 构造目标函数

$$E = \sum_{i \in W_p} (O_i - I_i)^2 + \varepsilon a_p^2,$$

其中,  $\varepsilon > 0$  为一个正则化参数;  $I$  为输入图像。通过最小化上述目标函数得到

$$a_p = \frac{(1/|W_p|) \sum_{i \in W_p} G_i I_i - \mu_p \bar{I}_p}{\sigma_p^2 + \varepsilon}, \quad b_p = \bar{I}_p - a_p \mu_p,$$

其中,  $\mu_p$  和  $\sigma_p^2$  为引导图像  $G$  在窗口  $W_p$  下的均值和方差;  $\bar{I}_p$  为输入图像  $I$  在窗口  $W_p$  下的均值;  $|W_p|$  为窗口

$W_p$  中的像素个数。由于像素  $i$  可能存在于多个以不同的像素  $p$  为中心的窗口中,对应不同的窗口会计算出对应的  $a_p$  和  $b_p$ ,用  $|p|$  表示像素  $i$  所在的窗口个数,通过取平均得

$$\bar{a}_p = \frac{1}{|p|} \sum_{p:i \in W_p} a_i, \quad \bar{b}_p = \frac{1}{|p|} \sum_{p:i \in W_p} b_i,$$

因此得到

$$O_i = \bar{a}_p G_i + \bar{b}_p。$$

## 2 本文提出的网络结构

### 2.1 网络结构

本文提出了一种双分支的编码器网络结构,记为 DuGuTransUNet,第一个分支使用基本的卷积操作,如图 1 的绿色部分所示,该结构的每个 CNN 块都包括了 2 次卷积、GroupNorm<sup>[19]</sup> 以及 Relu 函数激活操作,然后进行最大池化,经过 4 次编码过程后该网络会将特征图直接传送给解码器。第二个分支如图 1 的蓝色部分所示,第一个 CNN 块包括 1 次卷积、GroupNorm 以及 Relu 函数激活和最大池化操作,其中卷积核的大小为 7,之后就是 ResNet 模块,通过 3 次该模块的操作最后将得到的特征图传送至 Transformer 模块,Transformer 模块处理后的特征图传送给解码器。

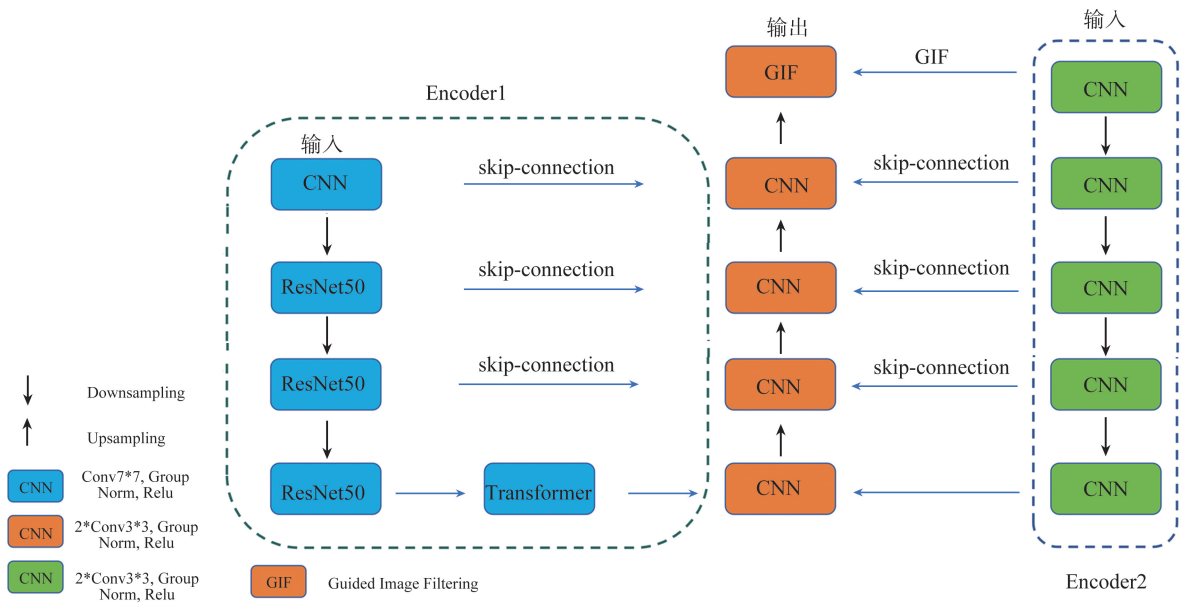


图 1 基于引导滤波和 Transformer 的双分支网络结构 DuGuTransUNet

Fig.1 A dual branch network structure DuGuTransUNet based on guided filtering and Transformer

图 1 中橙色部分为解码器,解码器的最底层将 2 个分支的特征图进行拼接然后进行卷积操作,和 UNet 的解码器一样通过多次卷积和 4 次上采样操作将图像恢复成原来的大小,最后输出图像。在解码器的处理过程中会拼接 2 个编码器不同阶段产生的特征图,从而使得网络的底层特征和高级特征可以直接建立联系,在图 1 的右分支中使用了 UNet++ 中的框架进行不同层级的信息拼接,而在左分支中考虑到不同层次特征图之间的差异性,在其中加入了融合不同尺度特征的卷积块。因为网络的下采样操作会导致视网膜血管信息的丢失,本文的网络结构还在网络的浅层特征和输出层中加入了引导滤波使得网络可以表现出更好的性能。该网络的双分支编码器结构会加强网络关联全局特征的能力,并且使得 Transformer 结构在小数据集上也可以表现出较好的效果。

### 2.2 引导滤波模块

视网膜血管含有大量的单像素的血管结构,但是在网络的下采样操作中会造成图像信息的丢失导致这些单像素的血管结构难以恢复。为了解决这个问题,本文的网络结构在网络的浅层特征和输出层加入了引导滤波模块(图 2 中的 GIF 模块)。使用网络输入图像的灰度图作为引导图像,对图 2 中第二层特征图反卷积上采样操作之后的特征图像进行引导滤波处理,然后把引导滤波后的特征图与图 2 所示的输入特征图进

行拼接,最后将这些拼接后的特征图进行卷积等操作,该结构在实验中表现出较好的效果。图1双分支结构的右分支中的跳跃连接加入引导滤波模块的结构如图2所示。

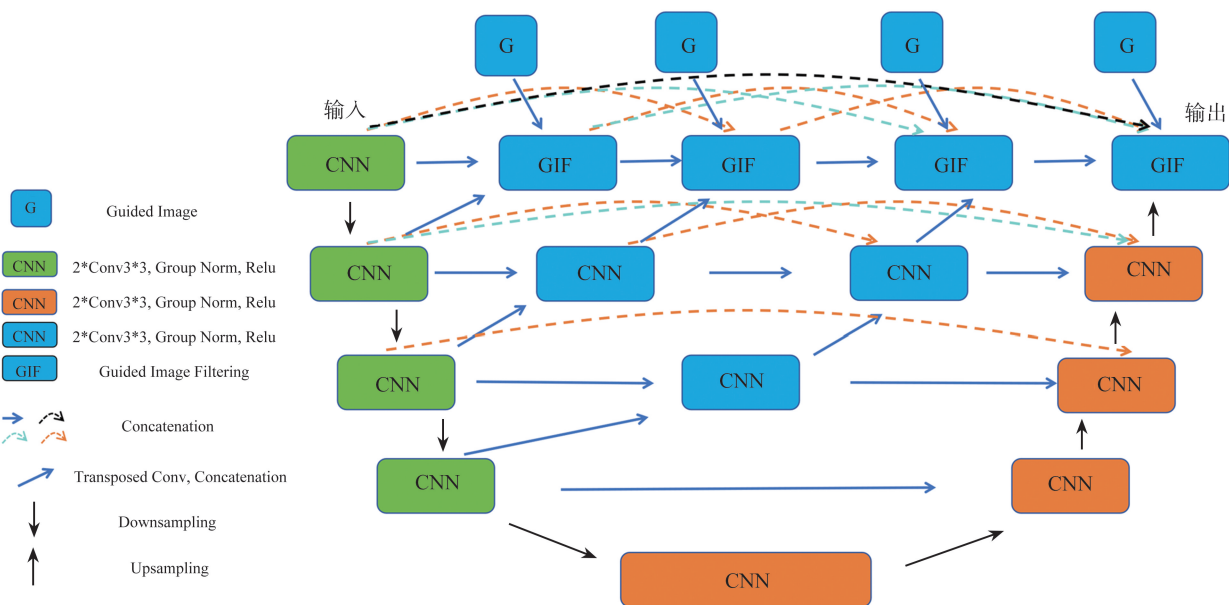


图2 加入引导滤波的跳跃连接的图1中的右分支  
Fig.2 The right branch in Fig.1 with skip-connections embedded with guided filtering

### 2.3 跳跃连接模块

UNet 网络直接将底层特征图与高层特征图进行拼接会由于不同层级之间的语义差异导致网络性能下降,UNet++模型通过加入融合不同尺度特征的卷积模块对 UNet 网络进行改进表现出较好的性能,基于这种思想本文在进行特征拼接时在每个编码器分支都加入了卷积块进行处理。编码器1采用了与 UNet++相同的跳跃连接结构,在跳跃链接中加入了卷积模块,具体结构如图3所示,编码器2除了引导滤波模块之外使用了与编码器1相同的跳跃连接结构。图3中蓝色线条表示将对应位置的特征图与后面的特征图进行拼接和卷积操作,当特征图的大小不同时先对特征图进行反卷积操作然后再进行拼接和卷积操作。

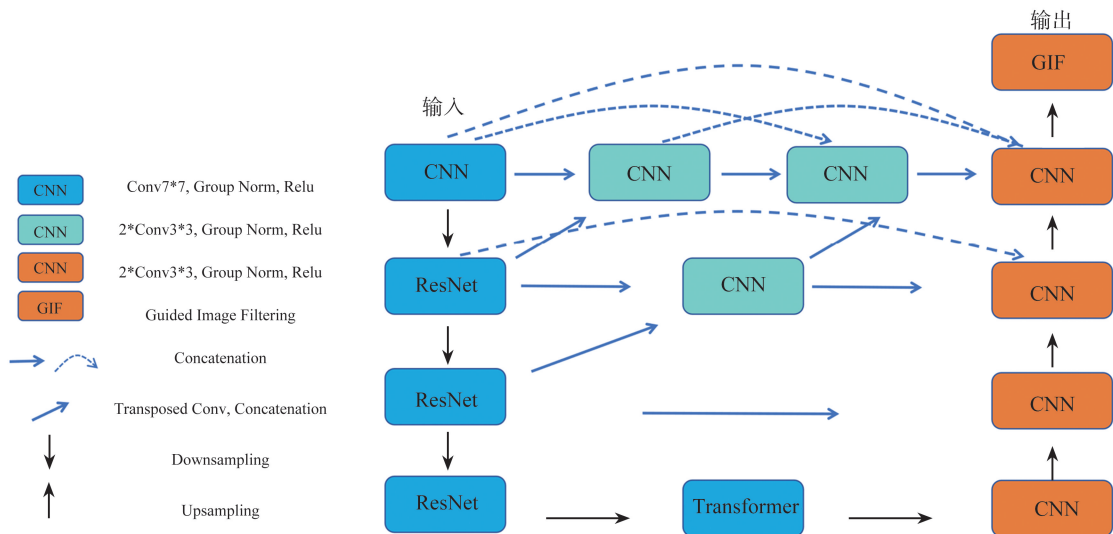


图3 加入详细跳跃连接信息的图1中的左分支  
Fig.3 The left branch in Fig.1 with skip-connections in detail

## 3 实验结果与分析

### 3.1 数据集与实验环境

为了验证网络的性能,本文使用 DRIVE 数据集<sup>[20]</sup> (digital retinal images for vessel extraction) 和

CHASEDB1 数据集<sup>[21]</sup>(combined healthy and diabetic retinopathy database 1)进行了实验。DRIVE 数据集包含 40 幅视网膜图像,其中 20 幅用于训练,20 幅用于测试。图像的分辨率为 584×565 像素,DRIVE 数据集是视网膜血管分割领域最早被广泛使用的数据集之一。CHASEDB1 数据集用于糖尿病视网膜病变研究,包含 28 幅视网膜图像,包含了不同疾病阶段的图像,是糖尿病视网膜病变分割算法评估的重要数据来源,该数据集使用 20 幅图片进行训练,8 幅图片进行测试。

本文模型在 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 Intel 酷睿 i5-1135G7 处理器,设备装有集成显卡。采用 Adam 优化器<sup>[22]</sup>对模型进行优化,Batch\_size = 1。总 epoch 数设为 150,初始学习率设置为 0.001,学习率随着迭代更新,每 50 次迭代缩小为原本的 1/10。损失函数使用交叉熵损失函数。

### 3.2 评价指标

为了客观地评估不同算法的性能,本文使用一系列的评价指标<sup>[23]</sup>,这些指标从不同角度衡量了算法的准确性、灵敏度、特异度等性能。灵敏度(sensitivity, Sen),也称为真正例率(true positive rate),表示算法成功检测到的血管像素占总血管像素的比例,数学表达式为

$$\text{Sen} = \frac{\text{TP}}{\text{TP} + \text{FN}},$$

其中,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)。特异度(specificity, Spe)表示算法成功排除非血管像素的能力,即非血管像素中被正确排除的比例,数学表达式为

$$\text{Spe} = \frac{\text{TN}}{\text{TN} + \text{FP}},$$

其中,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)。准确度(accuracy, Acc)表示算法正确分类的像素占总像素的比例,是一个综合考量,数学表达式为

$$\text{Acc} = \frac{\text{TN} + \text{TP}}{\text{TP} + \text{FN} + \text{TN} + \text{FP}}.$$

Dice 系数是一种集合相似度函数,通常被用来计算 2 个样本的相似度,Dice 的计算公式表示为

$$\text{Dice} = \frac{2|A \cap B|}{|A \cup B|},$$

其中,A 为预测标签的集合,B 为真实标签的集合。

### 3.3 对比实验分析

对比实验分别从 2 个数据集中进行,即 DRIVE 数据集和 CHASEBD1 数据集,在对比实验中分别比较了 UNet 模型<sup>[1]</sup>,UNet++ 模型<sup>[2]</sup>,TransUNet 模型<sup>[5]</sup>,DeepGuidanceNetwork 模型<sup>[24]</sup>以及 DATransUNet 模型<sup>[25]</sup>。分别从图像分割的准确度,特异度以及灵敏度和 Dice 来对比不同模型的性能,并且在对比中添加了各个模型分割一张图的平均测试时间。在模型训练中添加了 50% 概率的随机水平翻转和 [0°, 20°] 范围上的旋转,测试结果见表 1。

表 1 不同模型的 Acc、Sen、Spe、Dice 和平均测试时间  
Table 1 Acc, Sen, Spe, Dice, and average test time of different models

模型	DRIVE					CHASEBD1				
	Acc	Sen	Spe	Dice	时间/s	Acc	Sen	Spe	Dice	时间/s
UNet	0.958 0	0.724 5	0.986 5	0.882 1	1.00	0.957 8	0.828 8	0.970 1	0.864 5	1.500
UNet++	0.959 0	0.753 7	0.984 1	0.885 8	4.75	0.979 3	0.756 2	0.990 0	0.878 7	8.000
TransUNet	0.957 1	0.741 9	0.983 2	0.881 3	4.30	0.977 5	0.816 4	0.985 3	0.867 3	5.000
DeepGuidanceNet	0.960 9	0.760 9	0.985 2	0.879 1	3.00	0.979 9	0.837 8	0.986 8	0.890 2	4.870
DATransUNet	0.957 0	0.745 4	0.982 7	0.881 0	5.10	0.973 2	0.777 4	0.982 6	0.856 2	5.125
DuGuTransUNet	0.962 4	0.774 2	0.985 3	0.896 5	47.00	0.980 4	0.836 8	0.987 4	0.892 5	49.600

从表 1 中可以看出,无论是 DRIVE 数据集还是 CHASEBD1 数据集,本文的模型相对于实验中的其他模

型在准确度、特异度以及灵敏度上有着更好的表现,也就是说本文的模型有着更好的视网膜血管分割能力。相比较而言,由于 Transformer 结构在小数据集上的限制性,TransUNet 和 DATransUNet 在各个指标上表现较差。而本文的网络结构则有着更好的效果,这说明本文的模型在一定程度上缓解了小数据集的限制性问题。同时对比 UNet++ 和 DeepGuidanceNet,本文模型有着更好的结果,这表示加入双分支编码器的网络结构可以更好地关联图像的全局特征。在具备更好的分割效果的同时,由于本文模型结构更复杂,导致时间成本相对较高,在消融实验部分本文也进行了验证,但时间也在可接受范围内。

图 4 给出了数据集中具有代表性的 2 幅原始图像、标签图像和各种比较方法的处理结果。为了更好地展示分割效果,本文对应每幅整体图像(如图 4(a)–(f)和(m)–(r)所示),给出了对应的细节图像(如整体图像中的红色框部分)并进行了放大展示,如图 4(g)–(l)和(s)–(x)所示。从整体图像可以看出,相比于其他的网络模型本文的模型分割出了更多的血管细节,尤其是细小血管。从细节图像可以看出,本文的网络结构分割出的血管有着更好的连续性,而对比实验的其他网络出现了分割血管不连续的现象,更接近标签图像。这进一步验证了本文的网络结构有着更好的血管分割能力。

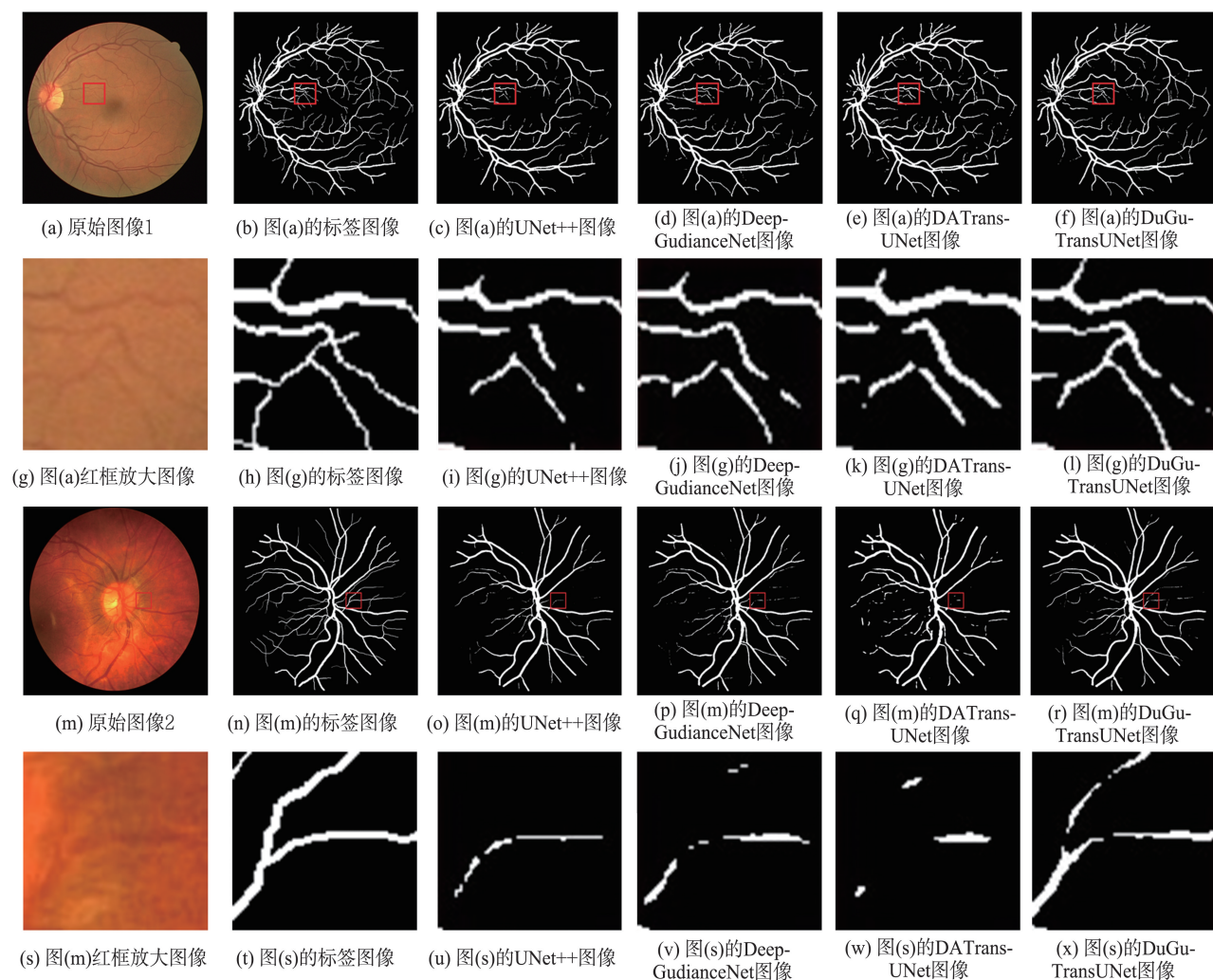


图 4 各种方法的实验结果  
Fig.4 Experimental results of different methods

### 3.4 消融实验及结果分析

为验证本文网络模型各模块的有效性,本节进行了消融实验,消融实验包括了本文模型中用 UNet++ 的跳跃连接替代引导滤波结构的模型(记为 DoubleUNet++);去掉左分支编码器 1 的模型(记为 DFUNet++);去掉右分支编码器 2 的模型(记为 TransUNet++);去掉左分支编码器 1 和用 Unet++ 跳跃连接替代引导滤波结构的模型(实际上就是 UNet++);以及本文提出的模型 DuGuTransUNet 模型,模型的设计思路和目标在表 2 中展示。这些模型在 DRIVE 和 CHASEBD1 数据集上进行实验,图 5 为不同模型

的分割结果。

表 2 模型描述与目标  
Table 2 Model description and objective

模型	模型描述	目标
DuGuTransUNet	本文所提基于引导滤波和 Transformer 的双分支网络	验证引导滤波跳跃链接和双分支的有效性
UNet++	DuDuTransUNet 模型去掉左分支和引导滤波	基础模型作为对照,验证本文设计模块的有效性
DFUNet++	DuDuTransUNet 模型去掉左分支	验证引导滤波模块的有效性
TransUNet++	DuDuTransUNet 模型去掉右分支	验证 Transformer 模块和跳跃连接的有效性
DoubleUNet++	DuDuTransUNet 模型去掉引导滤波模块	验证双分支编码器结构的有效性

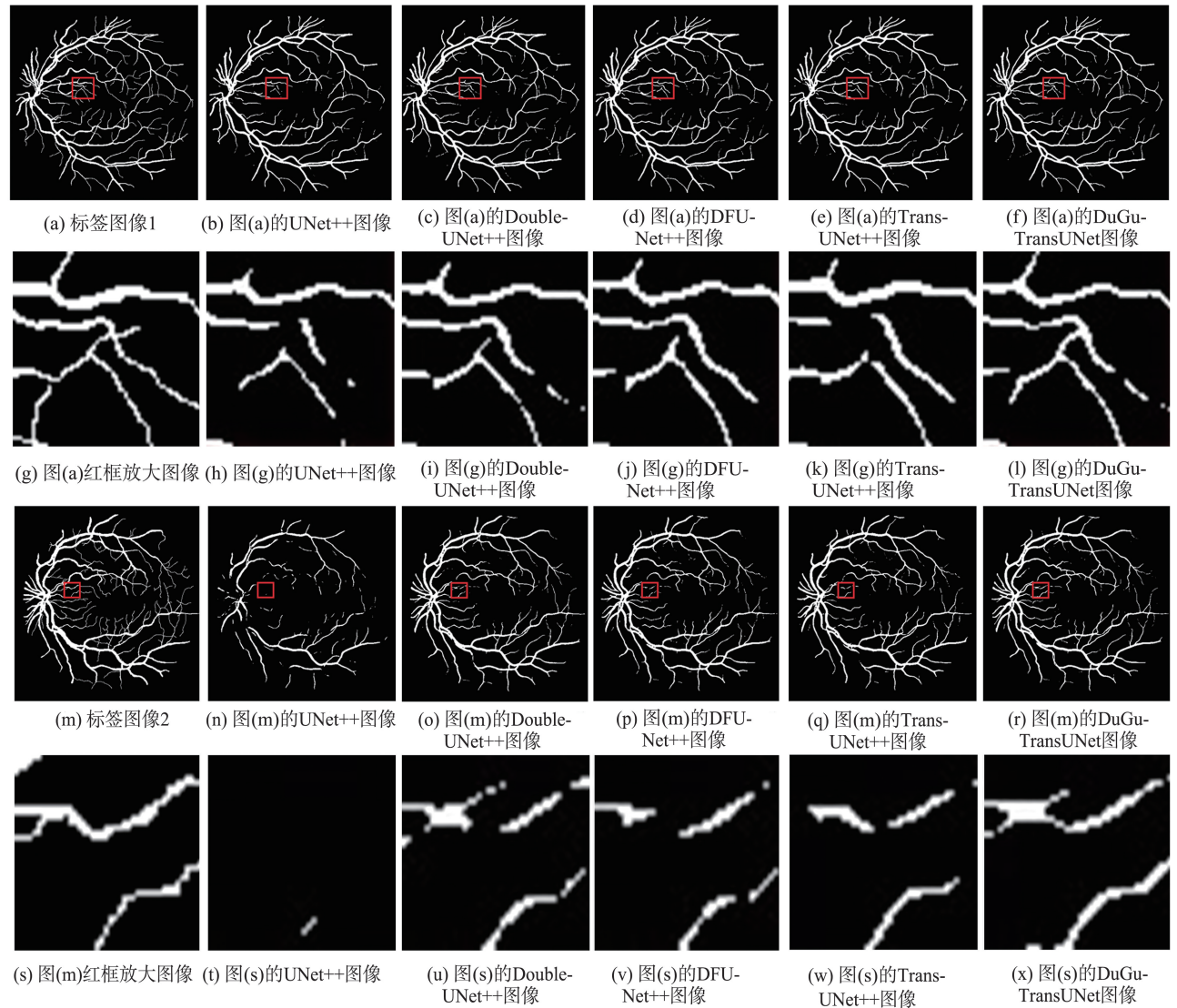


图 5 消融实验结果

Fig.5 Results of ablation experiments

从图 5(第 1 行和第 2 行是全局图,第 2 行和第 4 行是全局图中红框部分的放大图)以及表 3 可以看出,添加了引导滤波的 DFUNet++比原始的 UNet++模型在参数上表现更好,在分割图像中也有更多的细节并且血管的连续性更好,断连的地方较少,这说明引导滤波模块对增强血管细节发挥了重要作用。对比 TransUNet++与 UNet++模型的分割效果,可以看出 Transformer 模块相较于 CNN 在增强血管细节和连续性方面起到了作用。添加了双分支的 DoubleUNet++模型与 2 个只有单个分支的模型 DFUNet++和 TransUNet++相比有更好的 Acc、Sen 和 Spe 参数,并且在细节图中可以看出双分支结构的网络分割出的血管细节更多,这说明了本文模型的双分支结构具有增强网络性能的效果。本文所提方法比所有其他消融模型在客观参数和分割效果图上表现更好,且解决了 DoubleUNet++的 Spe 和 Dice 有所降低的问题,说明

加入引导滤波的跳跃连接、双分支和 Transformer 分割眼底图像效果更好。从表 3 可以看出,本文模型分割图片所用时间较长,这在很大程度上是由于使用双分支编码器导致计算量增加,目前硬件设备更新升级很快,相应的时间成本也会大大降低。总的来说,在分割效果上本文所提方法与先进方法相比具有更好的细小血管分割能力。

表 3 消融实验的 Acc、Sen、Spe、Dice 与平均测试时间  
Table 3 Acc, Sen, Spe, Dice and average test time of ablation experiments

模型	Acc	Sen	Spe	Dice	时间/s
UNet++	0.959 0	0.753 7	0.984 1	0.885 8	4.75
DFUNet++	0.961 9	0.770 9	0.985 2	0.895 2	16.15
TransUNet++	0.960 0	0.763 5	0.983 9	0.890 0	20.70
DoubleUNet++	0.962 1	0.775 3	0.984 9	0.822 2	40.60
DuGuTransUNet	0.962 4	0.774 2	0.985 3	0.896 5	47.00

## 4 结论

本文提出了加入引导滤波的双分支编码器 UNet++ 模型,该网络的第一个编码器由 ResNet 提取图像特征并使用 Transformer 结构进一步处理特征图,第二个编码器由普通的大小为 3 的卷积操作构成,通过卷积提取图像的信息。第一个编码器使得网络模型可以更好地关联图像的全局信息,而第二个编码器则保证了 Transformer 模型在小数据集上也可以发挥较好的作用。同时该网络在浅层特征和输出层中添加了引导滤波,通过引导图像对特征图进行增强,更好地恢复图像在网络处理中丢失的单像素血管结构信息,使得该网络有着更好的视网膜血管分割能力,具有一定的临床应用价值。但是由于网络结构复杂导致时间成本较高,可以通过优化网络结构和精细化每个模块进一步解决,这也正是接下来要做的研究工作。

### 参考文献:

- [1] ZHOU Jing, CHEN Bo. Retinal cell damage in diabetic retinopathy[J]. *Cells*, 2023, 12(9):1342.
- [2] JENA P K, KHUNTIA B, PALAI C, et al. A novel approach for diabetic retinopathy screening using asymmetric deep learning features[J]. *Big Data and Cognitive Computing*, 2023, 7(1):25.
- [3] YOU Ning, HAN Libo, ZHU Daming, et al. Research on image denoising in edge detection based on wavelet transform[J]. *Applied Sciences*, 2023, 13(3):1837.
- [4] XIAN Ronghao, LUGU Rikong, PENG Hong, et al. Edge detection method based on nonlinear spiking neural systems[J]. *International Journal of Neural Systems*, 2023, 33(1):2250060.
- [5] 李兰兰,张孝辉,牛得草,等. 深度学习在视网膜血管分割上的研究进展[J]. *计算机科学与探索*, 2021, 15(11):2063-2076. LI Lanlan, ZHANG Xiaohui, NIU Decao, et al. Research progress of deep learning in retinal vessel segmentation[J]. *Journal of Frontiers of Computer Science and Technology*, 2021, 15(11):2063-2076.
- [6] 纪守领,杜天宇,邓水光,等. 深度学习模型鲁棒性研究综述[J]. *计算机学报*, 2022, 45(1):190-206. JI Shouling, DU Tianyu, DENG Shuiguang, et al. Robustness certification research on deep learning models: a survey[J]. *Chinese Journal of Computers*, 2022, 45(1):190-206.
- [7] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation[C]// *Medical Image Computing and Computer-Assisted Intervention*. Cham: Springer, 2015:234-241.
- [8] SIDDIQUE N, PAHEDING S, ELKIN C P, et al. U-net and its variants for medical image segmentation: a review of theory and applications[J]. *IEEE Access*, 2021, 9:82031-82057.
- [9] ZHOU Z W, SIDDIQUEE M M R, TAJBAKHS 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.
- [10] LONG J, SHELHAMER E, Darrell T. Fully convolutional networks for semantic segmentation[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Boston: IEEE, 2015:3431-3440.
- [11] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]// *Proceedings of the 31st International Conference on Neural Information Processing Systems*. New York: ACM, 2017:6000-6010.
- [12] DOSOVISKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: transformers for image recognition at scale[EB/OL]. (2021-06-03)[2024-03-01]. <https://arxiv.org/pdf/2010.11929>.
- [13] CHEN Jieneng, LU Yongyi, YU Qihang, et al. Transunet: transformers make strong encoders for medical image

- segmentation[EB/OL]. (2021-02-08)[2024-03-01]. <https://arxiv.org/pdf/2102.04306>.
- [14] TOMAR N K, SHERGILL A, RIEDERS B, et al. TransResU-Net; transformer based ResU-Net for real-time colonoscopy polyp segmentation[EB/OL]. (2022-06-17)[2024-03-01]. <https://doi.org/10.48550/arXiv.2206.08985>.
- [15] CAO H, WANG Y Y, CHEN J, et al. Swin-unet; unet-like pure transformer for medical image segmentation[C]// European Conference on Computer Vision. Cham: Springer, 2022:205-218.
- [16] HE Kaiming, SUN Jian, TANG Xiaoou. Guided image filtering[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 35(6):1397-1409.
- [17] QIN Zhen, SUN Weixuan, DENG Hui, et al. Cosformer; rethinking softmax in attention[EB/OL]. (2022-02-17)[2024-03-01]. <https://doi.org/10.48550/arXiv.2206.08985>.
- [18] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, 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.
- [19] WU Yuxin, HE Kaiming. Group normalization[C]// Proceedings of the European Conference on Computer Vision (ECCV). Munich: Springer, 2018:3-19.
- [20] 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.
- [21] 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.
- [22] REYAD M, SARHAN A M, ARAFA M. A modified Adam algorithm for deep neural network optimization[J]. Neural Computing and Applications, 2023, 35(23):17095-17112.
- [23] QURESHI I, YAN J, ABBAS Q, et al. Medical image segmentation using deep semantic-based methods: a review of techniques, applications and emerging trends[J]. Information Fusion, 2023, 90:316-352.
- [24] YIN Pengshuai, YUAN Rui, CHENG Yiming, et al. Deep guidance network for biomedical image segmentation[J]. IEEE Access, 2020, 8:116106-116116.
- [25] SUN Guanqun, PAN Yizhi, KONG Weikun, et al. DA-TransUNet; integrating spatial and channel dual attention with transformer U-Net for medical image segmentation[J]. Frontiers in Bioengineering and Biotechnology, 2024, 12:1398237.

(编辑:李艺)

(上接第21页)

- [12] SHAFIQ M, GU Z. Deep residual learning for image recognition; a survey[J]. Applied Sciences, 2022, 12(18):8972.
- [13] XIE S, GIRSHICK R, DOLLÁR P, et al. Aggregated residual transformations for deep neural networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017:1492-1500.
- [14] DING Hu, XU Jinhui. Learning the truth vector in high dimensions[J]. Journal of Computer and System Sciences, 2020, 109:78-94.
- [15] KUMAR A, LIANG P, MA T. Verified uncertainty calibration[EB/OL]. (2019-09-23)[2023-12-29]. <http://arxiv.org/abs/1909.10155>.
- [16] GUO C, PLEISS G, SUN Y, et al. On calibration of modern neural networks[C]// Proceedings of the 34th International Conference on Machine Learning-Volume 70. Sydney: ACM, 2017:1321-1330.
- [17] 曹建军,常宸,翁年凤,等. 基于神经网络编码的真值发现[J]. 计算机工程与科学,2021,43(9):1546-1557.  
CAO Jianjun, CHANG Chen, WENG Nianfeng, et al. Truth discovery based on neural network encoding[J]. Computer Engineering & Science, 2021, 43(9):1546-1557.
- [18] XU Haitao, ZHANG Haiwang, LI Qianqian, et al. A data-semantic-conflict-based multi-truth discovery algorithm for a programming site[J]. Computers, Materials & Continua, 2021, 68(2):2681-2691.
- [19] CHOI Y S, BAE S, CHANG J H, et al. Fully automated hybrid approach to predict the IDH mutation status of gliomas via deep learning and radiomics[J]. Neuro-Oncology, 2021, 23(2):304-313.
- [20] TAHA B, LI T H, BOLEY D, et al. Detection of isocitrate dehydrogenase mutated glioblastomas through anomaly detection analytics[J]. Neurosurgery, 2021, 89(2):323-328.
- [21] KAWAGUCHI R K, TAKAHASHI M, MIYAKE M, et al. Assessing versatile machine learning models for glioma radiogenomic studies across hospitals[J]. Cancers, 2021, 13(14):3611.

(编辑:李艺)