

李煜豫, 何宏, 陈家毓, 等. 融合注意力机制和空洞卷积的甲状腺眼病区域分割方法[J]. 智能计算机与应用, 2024, 14(12): 51-59. DOI:10.20169/j.issn.2095-2163.24080901

## 融合注意力机制和空洞卷积的甲状腺眼病区域分割方法

李煜豫<sup>1</sup>, 何宏<sup>1</sup>, 陈家毓<sup>1</sup>, 周慧芳<sup>2</sup>

(1 上海理工大学健康科学与工程学院, 上海 200093; 2 上海交通大学医学院附属第九人民医院眼科, 上海 200011)

**摘要:** 针对眼科门诊对甲状腺眼病分级诊断时, 存在诊断主观性较强、工作量大、判断不准确且医护资源消耗较大的问题, 为此提出一种融合注意力机制和空洞卷积的甲状腺眼病 (Thyroid Associated Ophthalmopathy, TAO) 患病区域多类别分割方法。该方法以 U 型网络 (U-Net) 作为基础, 提出了并行多尺度空洞卷积模块并将其引入下采样底部, 通过不同尺度感受野提取因下采样而丢失的有效信息, 以提升模型的特征提取能力; 此外, 针对眼瞳区域较小且 TAO 患者眼睑与巩膜病变后距离发生改变的问题, 提出了融合注意力机制模块, 提高模型在通道和空间上特征提取的有效性, 有效捕获眼部图像的高级与低级语义特征, 改善眼睑与巩膜边缘处分割效果。为验证本文方法的有效性, 将在数据增强后的 2 369 张巩膜和眼睑数据集上进行训练和预测, 实验结果表明所提方法的 *Dice* 相似系数、平均交并比 (*mIoU*) 分别达到了 93.33%、87.61%, 实现了对甲状腺眼病患病区域的准确分割, 可以有效地辅助甲状腺眼病的严重性分级诊断。

**关键词:** 甲状腺眼病; 多类别分割; 空洞卷积; 注意力机制; 损失函数

中图分类号: TP391.4

文献标志码: A

文章编号: 2095-2163(2024)12-0051-09

## A method for segmenting thyroid orbitopathy regions via integration of attention mechanism and dilated convolution

LI Yuyu<sup>1</sup>, HE Hong<sup>1</sup>, CHEN Jiayu<sup>1</sup>, ZHOU Huifang<sup>2</sup>

(1 School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China;

2 Department of Ophthalmology, Shanghai Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200011, China)

**Abstract:** In order to solve the problems of strong subjectivity, heavy workload, inaccurate judgment and large consumption of medical resources in the graded diagnosis of thyroid eye disease in ophthalmology clinics, a multi-category segmentation method of Thyroid Associated Ophthalmopathy (TAO) disease area integrating attention mechanism and cavitory convolution is proposed. Based on U-Net, a parallel multi-scale dilated convolution module is proposed and introduced to the bottom of downsampling, and the effective information lost due to downsampling is extracted through different scale receptive fields, so as to improve the feature extraction ability of the model. In addition, in order to solve the problem that the eye area is small and the distance between the eyelid and the scleral lesion changes in TAO patients, a fusion attention mechanism module is proposed in the paper to improve the effectiveness of the model in channel and spatial feature extraction, effectively capture the high-level and low-level semantic features of the eye image, and improve the segmentation effect between the eyelid and scleral edge. In order to verify the effectiveness of the proposed method, the similarity coefficient *Dice* and mean Intersection Union ratio (*mIoU*) of the proposed method reach 93.33% and 87.61%, respectively, which realizes the accurate segmentation of the affected area of thyroid eye disease, and could effectively assist in the graded diagnosis of the severity of thyroid eye disease.

**Key words:** thyroid eye disease; multi-class segmentation; dilated convolution; attention mechanisms; loss function

**基金项目:** 国家科技部项目 (G2021013008); 上海市科学技术委员会项目 (18070503000); 上海理工大学医工交叉重点项目 (1020308405, 1022308502)。

**作者简介:** 李煜豫 (2000—), 男, 硕士研究生, 主要研究方向: 医学图像处理; 陈家毓 (1997—), 男, 博士研究生, 主要研究方向: 医学图像处理; 周慧芳 (1977—), 女, 博士, 教授, 博士生导师, 主要研究方向: 甲状腺相关眼病。

**通信作者:** 何宏 (1973—), 女, 博士, 教授, 博士生导师, 主要研究方向: 医学图像处理。Email: hehong@usst.edu.cn。

收稿日期: 2024-08-09

## 0 引言

甲状腺相关眼病 (Thyroid Associated Ophthalmopathy, TAO), 又称格雷夫斯眼病 (Graves' ophthalmopathy, GO), 是与甲状腺功能异常密切相关的一种器官特异自身免疫性疾病, 其发病率居成人眼眶类疾病之首<sup>[1]</sup>。甲状腺相关眼病可导致患者单眼或双眼眼睑退缩、眼球突出、复视限制性斜视、暴露性角膜病变和 TAO 视神经病变 (Dysthyroid Optic Neuropathy, DON) 等<sup>[2]</sup>。临床对 TAO 进行严重性分级主要基于患者眼部症状与特征: 眼眶软组织充血水肿、眼部刺激症状、眼球突出、眼睑退缩、复视以及视神经病变。TAO 的严重性分级对于后续治疗方案具有重要影响, 目前世界对 TAO 严重性分级主要采用欧洲格雷夫斯眼病专家组 (European Group on Graves' Orbitopathy, EUGOGO) 制定的分级量表<sup>[3]</sup>。该量表通过眼睑退缩宽度、软组织受累、眼球突出、复视等症状进行评估, 将 TAO 严重性分为轻度、中重度和极重度三级。

临床 TAO 患者严重性分级的诊断中, 眼科医生往往通过临床经验为患者眼部症状与特征进行评估。其中, 软组织受累的判断包括眼睑是否水肿与巩膜是否充血, 由于其特征评估需要大量临床经验作为基础, 且受医生主观性影响较大, 在实际眼科门诊的评估诊断中, TAO 的严重性分级诊断需要消耗医生大量精力与时间。因此, 设计深度学习网络, 准确地对眼睑与巩膜同步进行多类别语义分割, 其分割结果能有效辅助眼科医生判断 TAO 患者软组织受累状况, 提高眼科门诊接诊效率; 同时, 也能为软组织受累自动识别模型提供精确的感兴趣区域, 提高自动识别模型的准确率。

近年来, 随着深度学习方法的发展, 图像分割技术水平得到显著提升, 医学领域越来越多地将图像分割技术应用于辅助临床诊断。Ronneberger 等学者<sup>[4]</sup>提出具有对称结构的 U 型网络模型—U-Net, 其采用端到端的编码器、解码器的结构, 较好地保留了图像的高低级语义特征, 医学图像领域分割已然成为研究热点。如今, U-Net 已经出现许多变换形式, 如嵌套 U-Net (U-Net++)<sup>[5]</sup>、注意力 U-Net (attention U-Net)<sup>[6]</sup>、U-Net3+ 等<sup>[7]</sup>, 这些变换体广泛应用于医学图像的分割中。Chen 等学者<sup>[8]</sup>基于卷积神经网络, 并利用多注意力密集连接, 提出用于虹膜分割的神经网络 MADNet, 实现了端到端的虹膜分割。Wang 等学者<sup>[9]</sup>提出在 U-Net 模型收缩路

径和扩展路径之间的中间瓶颈部分嵌入注意力模块, 以加强学习判别能力, 实现对巩膜的分割。Rot 等学者<sup>[10]</sup>提出了一种使用 SegNet 模型进行巩膜分割的方法, 在 SBVPI 数据集上取得较好的效果。Fuhl 等学者<sup>[11]</sup>提出了一个使用循环损失函数训练生成对抗网络 (GAN) 在眼睑区域进行分割, 优化后的 GAN 消除了图像中的噪点和遮挡。徐蓬泉等学者<sup>[12]</sup>构建多尺度语义模块等价代替较大的卷积核获得不同的感受野。褚张晴晴等学者<sup>[13]</sup>在 U 型网络跳跃连接中加入特征融合模块代替直接拼接操作, 有效融合不同层次的特征信息。胡帅等学者<sup>[14]</sup>在编码器结构加入提取低级特征信息的 MEE 模块, 获取医学图像丰富的边缘信息。梁礼明等学者<sup>[15]</sup>结合 Transformer 提出自校准注意力视网膜分割算法。陈家毓等学者<sup>[16]</sup>提出密集型挤压-激励通道注意力模块, 分别对眼睑、巩膜、泪阜进行分割, 但该模型无法实现眼睑、巩膜区域的多类别同时分割, 而且在实际硬件部署时存在冗余。综上所述, 以上算法依然存在不足: U-Net 系列在上下采样过程中可能会丢失空间信息, 从而导致分割精度下降。SegNet<sup>[17]</sup>模型包含大量的卷积和池化操作, 导致模型的计算量较大, 在面对精细图像分割时, 只利用了局部信息进行特征提取, 缺乏空间与通道特征的匹配, 因此提取的语义信息并不准确。

受上述工作启发, 本文提出一种基于并行多尺度空洞卷积和注意力机制的分割方法, 有效应用于 TAO 患病区域眼睑、巩膜的分割。该网络基于 U-Net 的编码-解码框架, 在底部融合并行多尺度空洞卷积 (Parallel Multiscale Atrous Convolution, PMAC) 通过不同尺度感受野捕获下采样过程中丢失的有效信息, 以提升模型的特征提取能力; 此外, 提出融合注意力机制模块 (Fusion Attention Module, FAM), 将通道注意力 (Channel Attention, CA) 与空间注意力 (Spatial Attention, SA) 融合并加入残差连接, 使网络能自适应地学习眼部图像的通道和空间特征信息, 提高病变眼睑与巩膜边缘区域的分割准确率。最后, 该网络在上海交通大学医学院附属第九人民医院眼科 TAO 患者真实面部彩色图像数据上进行训练与测试, 实现 TAO 患者眼睑、巩膜区域的同时准确分割, 结果表明本文方法具有较好的分割效果。

## 1 网络结构

### 1.1 整体框架

本文提出一种基于并行多尺度空洞卷积和注意

力机制的网络模型,主要由编码器、解码器、并行多尺度空洞卷积模块 (Parallel Multiscale Atrous Convolution, PMAC)

Convolution, PMAC)、融合注意力机制模块 (Fusion Attention Module, FAM) 组成,模型结构如图 1 所示。

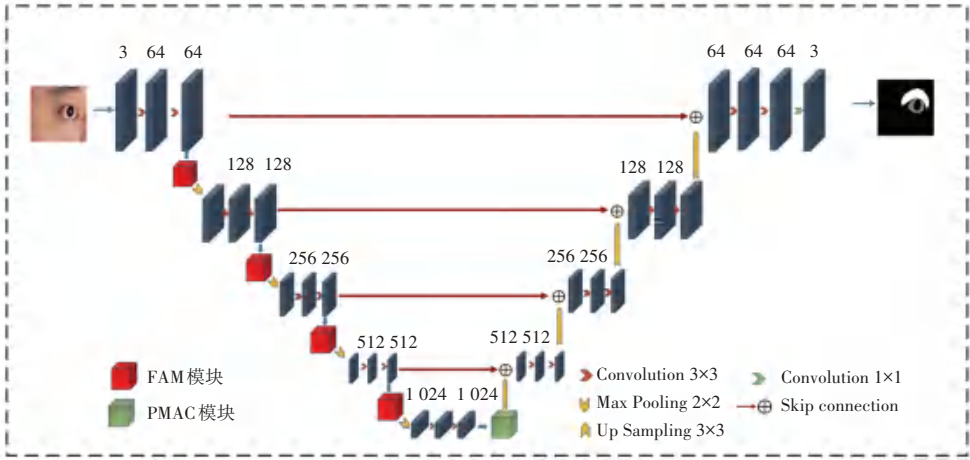


图 1 网络模型结构图

Fig. 1 Structure diagram of the network model

输入图像尺寸为  $256 \times 256 \times 3$  的彩色图像,经过编码器完成 4 次下采样,每次下采样经过 2 层  $3 \times 3$  卷积层,提取图像的低级特征,在每个卷积层后使用 ReLU 激活函数引入非线性和防止梯度爆炸,再经过  $2 \times 2$  的最大池化层完成一次下采样操作,这使得输出特征图的尺寸减半,同时保留了主要特征。经过 4 次下采样得到尺寸为  $16 \times 16 \times 512$  的特征图,接着将特征图输入并行多尺度空洞卷积模块 (PMAC),在不增加过多参数的同时,扩大感受野与获取不同尺度的目标特征。解码器同样为 4 层,每一层包含卷积核尺寸为  $3 \times 3$  的反卷积,将每一层输出特征图的尺寸扩大为输入特征图的 2 倍,每层编码器经过融合注意力机制模块 (FAM),自适应地调整特征图中通道和空间维度的权重,与解码器上采样输出特征图使用跳跃连接进行融合,提高网络的解码能力。最后,经过卷积核  $1 \times 1$  的卷积层压缩得到尺寸为  $256 \times 256 \times 3$  的眼睑、巩膜分割图。

## 1.2 并行多尺度空洞卷积模块

在下采样过程中特征图尺寸不断缩小,导致丢失了部分细节信息和全局上下文信息。因此,受空洞空间金字塔池化 (ASPP)<sup>[18]</sup> 和 CE-Net<sup>[19]</sup> 模块启发,在不引入过多参数的前提下,有效提取因下采样丢失的不同尺度的细节语义信息<sup>[20]</sup>,本文设计了并行多尺度空洞卷积模块 (PMAC)。该模块通过 2 组不同空洞率的空洞卷积<sup>[21]</sup> 连接,同时并联  $3 \times 3$  卷积和池化层,获得不同感受野下的特征图,与分别经过普通卷积和  $1 \times 1$  卷积的特征图进行拼接,从而获取分割目标的多尺度信息和细节信息,增强模型的

特征表达能力,在网络底部添加并行多尺度空洞卷积模块,提高了眼睑与巩膜的分割效果。

空洞卷积通过调整卷积核的膨胀率,可以在不增加参数数量的前提下实现增大感受野和保持分辨率的效果,空洞卷积与普通卷积的对比如图 2 所示。

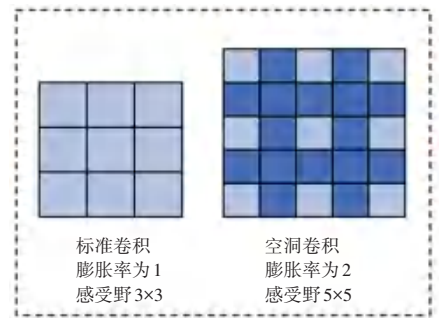


图 2 标准卷积与空洞卷积

Fig. 2 Standard convolution and dilated convolution

普通卷积可以看作是膨胀率为 1 的空洞卷积。空洞卷积核与标准卷积核的关系如下所示:

$$K = k + (k - 1) \times (r - 1) \quad (1)$$

其中,  $K$  表示空洞卷积核大小;  $k$  表示标准卷积核大小;  $r$  表示膨胀率。

并行多尺度空洞卷积模块 (PMAC) 结构如图 3 所示,由 4 层结构组成。第 1 层为  $3 \times 3$  标准卷积,输出特征图尺寸与输入特征图尺寸一致,假设输入特征图  $X_{input}$ , 第 1 层的计算如下:

$$y_1 = Conv(X_{input}) \quad (2)$$

其中,  $Conv$  表示  $3 \times 3$  卷积,  $y_1$  表示第一层的输出特征图。

第 2、3 层为空洞卷积层,每一层由 2 个空洞卷

积串联组成,第2层空洞卷积的膨胀率分别为1和3,第3层空洞卷积的膨胀率分别为3和5。第2、第3层的计算如下:

$$y_2 = \text{Conv}(\text{Dilated2}(\text{Conv}(\text{Dilated1}(X_{\text{input}})))) \quad (3)$$

$$y_3 = \text{Conv}(\text{Dilated3}(\text{Conv}(\text{Dilated2}(X_{\text{input}})))) \quad (4)$$

其中,  $\text{Conv}$  表示  $3 \times 3$  卷积;  $\text{Dilated}$  表示该卷积为膨胀卷积;  $y_2, y_3$  分别为第2、3层的输出特征图。

第4层将特征图进行全局平均池化获取全局信息,再利用双线性插值将特征图恢复至原输入特征图大小,第四层的计算如下:

$$y_4 = \text{Interpolate}(\text{Conv}(\text{Pool}_{\text{avg}}(X_{\text{input}}))) \quad (5)$$

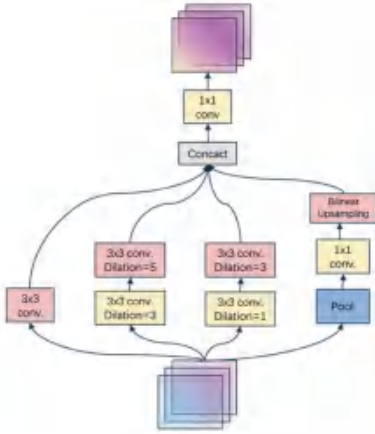


图3 并行多尺度空洞卷积模块

Fig. 3 Parallel multi-scale dilated convolution module

其中,  $\text{Pool}_{\text{avg}}$  表示全局平均池化;  $\text{Interpolate}$  在此采用双线性插值操作,避免引入过多参数量;  $y_4$  表示第4层的输出特征图。

最后,将4层网络获取的不同感受野的信息拼接起来从而获取多尺度信息,提高模块的整体特征提取能力,输出特征图  $Y_{\text{out}}$  如下:

$$y_4 = \text{Interpolate}(\text{Conv}(\text{Pool}_{\text{avg}}(X_{\text{input}}))) \quad (6)$$

$$Y_{\text{out}} = \text{Conv}(y_1 \oplus y_2 \oplus y_3 \oplus y_4) \quad (7)$$

其中,  $\text{Conv}$  表示  $1 \times 1$  卷积,用于调整输出通道数;“ $\oplus$ ”表示拼接操作;  $Y_{\text{out}}$  表示输出特征图。

### 1.3 融合注意力机制模块

眼睑与巩膜的特征辨别对于TAO严重性分级至关重要,由于TAO患者眼部病变后,巩膜与眼睑之间的距离发生改变,且因为眼部区域较小,在图像分割时往往难以准确地进行同时分割,且在模型下采样过程中有些无效信息会保留在输出特征图的一些通道和位置上。受SENet<sup>[22]</sup>和CBAM<sup>[23]</sup>启发,为了提高特征提取的有效性,本文设计融合注意力机制模块(FAM),将其添加到U型网络下采样与上采样的跳跃连接<sup>[24]</sup>之中。其中,空间注意力机制与通道注意力机制在训练过程中能更加精确地获取图像空间与通道上的高级与低级语义特征,并且去除特征通道上的无效信息,更准确地区分眼睑与巩膜的边缘部分。融合注意力机制模块(FAM)整体结构如图4所示。

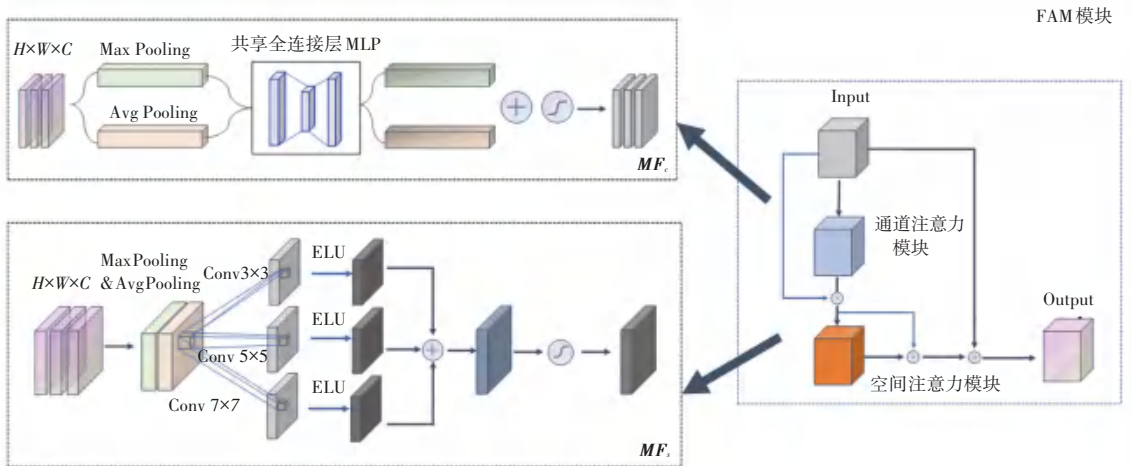


图4 融合注意力机制模块图

Fig. 4 Diagram of the fusion attention mechanism module

通道注意力机制的实现:首先将平均池化与最大池化同一时间作用于输入特征图  $F \in R^{H \times W \times C}$ , 分别得到2个大小为  $1 \times 1 \times C$  的特征向量  $F_1$  和  $F_2$ ; 接着,通过多层感知机对2个特征图进行降维,将  $F_1$

和  $F_2$  的通道数压缩至原来的  $1/r$ , 再将其扩张原通道数  $C$ ; 然后,将2个特征向量逐元素相加并通过 Sigmoid 激活函数得到具有通道位置权重的特征向量  $MF_c$ , 最后使用该特征向量乘以输入特征图求出

新特征图  $F_c$ 。计算过程如下所示:

$$\mathbf{MF}_c = \text{Sig}\{MLP(\text{pool}_{\text{avg}}(\mathbf{F})) \oplus MLP(\text{pool}_{\text{max}}(\mathbf{F}))\} \quad (8)$$

$$\mathbf{F}_c = \mathbf{MF}_c \otimes \mathbf{F} \quad (9)$$

其中,  $\text{Sig}$  表示 Sigmoid 激活函数;  $MLP$  表示多层感知机;  $\text{pool}_{\text{avg}}$  表示全局平均池化;  $\text{pool}_{\text{max}}$  表示最大池化; “ $\oplus$ ”表示拼接操作; “ $\otimes$ ”表示特征相乘。

空间注意力机制的实现: 首先, 将经过通道注意力机制的特征图  $F_c \in R^{H \times W \times C}$  作为输入, 分别经过平均池化和最大池化对输入特征图沿通道进行降维, 得到 2 个大小为  $H \times W \times 1$  的特征向量  $F_3$  和  $F_4$ , 然后将 2 个特征向量按通道进行拼接得到大小为  $H \times W \times 2$  的新特征向量; 接着利用卷积核尺寸分别为  $3 \times 3, 5 \times 5, 7 \times 7$  的卷积, 并行对特征进行提取, 再经过 ELU 进行非线性激活, 此后将不同尺寸的特征做加法运算实现特征融合。最后, 使用 Sigmoid 激活函数得到具有空间信息权重的特征向量  $\mathbf{MF}_s$ , 将  $\mathbf{MF}_s$  乘以特征图  $F_c$  得到空间注意力机制的输出特征图  $F_s$ 。计算过程如下所示:

$$\mathbf{MF}_{s1} = \text{ELU}\{Conv_{3 \times 3}(\text{pool}_{\text{avg}}(\mathbf{MF}_c) \oplus \text{pool}_{\text{avg}}(\mathbf{MF}_c))\} \quad (10)$$

$$\mathbf{MF}_{s2} = \text{ELU}\{Conv_{5 \times 5}(\text{pool}_{\text{avg}}(\mathbf{MF}_c) \oplus \text{pool}_{\text{avg}}(\mathbf{MF}_c))\} \quad (11)$$

$$\mathbf{MF}_{s3} = \text{ELU}\{Conv_{7 \times 7}(\text{pool}_{\text{avg}}(\mathbf{MF}_c) \oplus \text{pool}_{\text{avg}}(\mathbf{MF}_c))\} \quad (12)$$

$$\mathbf{MF}_s = \mathbf{MF}_{s1} \oplus \mathbf{MF}_{s2} \oplus \mathbf{MF}_{s3} \quad (13)$$

$$\mathbf{F}_s = \mathbf{MF}_s \otimes \mathbf{F}_c \quad (14)$$

其中,  $Conv_{3 \times 3}, Conv_{5 \times 5}, Conv_{7 \times 7}$  分别表示卷积核为  $3 \times 3, 5 \times 5, 7 \times 7$ ;  $\text{pool}_{\text{avg}}$  表示全局平均池化;  $\text{pool}_{\text{max}}$  表示最大池化; “ $\oplus$ ”表示拼接操作, “ $\otimes$ ”表示特征相乘。

最后, 在通道注意力与空间注意力中加入残差结构, 将输入特征图  $F$  与经过通道注意力和空间注意力的特征图  $\mathbf{MF}_s$  相加, 得到最终的特征图  $F_{\text{out}}$ 。在增加网络深度的同时减少额外的计算负担, 并且由于残差连接允许输入特征信息的直接传递, 避免了下采样过程中信息丢失, 提高模型泛化能力。计算过程如下所示:

$$\mathbf{F}_{\text{out}} = \mathbf{F} \oplus \mathbf{MF}_s \quad (15)$$

其中, “ $\oplus$ ”表示相加操作。

## 2 实验结果与分析

### 2.1 数据增强与参数设置

本实验数据来源于上海交通大学医学院附属第

九人民医院眼科, 共 658 张 TAO 患者的眼部彩色图像。由于在实际采集患者图像的场景中, 光照、对焦等因素对眼部图像会造成一定影响, 在模型训练过程中应当考虑上述情况, 因此, 本文对原始数据进行了数据增强。首先将训练集与验证集按 9:1 随机划分, 接着分别进行数据增强, 将原始图像分别经过随机亮度增强 (随机亮度因子为 0.5~1.0)、随机对比度增强 (随机对比度因子为 0.5~1.5)、随机饱和度增强 (随机饱和度因子为 0.5~1.5)。在增加训练量的同时, 降低模型过拟合现象发生的概率。

本文实验的硬件平台为 CPU: Intel (R) Xeon (R) W-2255 CPU @ 3.70 GHz, GPU: NVIDIA GeForce RTX 3090, 内存: 16 GB, 软件环境: PyTorch 2.0.1。优化器使用 RMSprop 算法, 初始学习率设置为  $1 \times 10^{-5}$ ,  $Batch\ size$  设置为 80,  $Epochs$  设置为 40。

### 2.2 损失函数

本文模型的损失函数采用交叉熵损失函数与 Dice 损失函数相加作为最终的损失函数进行模型优化, 分割结果越精确, 损失函数值越小。交叉熵损失函数的计算如下:

$$L_{CE} = \frac{1}{N} \sum_i L_i = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C y_{ic} \log(p_{ic}) \quad (16)$$

其中,  $N$  表示样本数量;  $C$  表示类别数;  $y_{ic}$  和  $p_{ic}$  分别表示第  $i$  个样本第  $C$  个类别的真实值和预测值。

Dice 损失函数计算如下:

$$L_{\text{Dice}} = 1 - \frac{2 \sum_{i=1}^N p_i y_i}{\sum_{i=1}^N p_i^2 + \sum_{i=1}^N y_i^2} \quad (17)$$

其中,  $N$  表示样本数量,  $y_i$  和  $p_i$  分别表示第  $i$  个样本的真实标签与预测结果。

### 2.3 评价指标

本文采用 Dice 系数、平均交并比 (MIoU)、召回率 (Recall)、准确率 (Acc) 四个评价指标对分割结果进行评估。各评估指标的计算公式如下:

$$\text{Dice} = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (18)$$

$$\text{MIoU} = \frac{1}{N} \sum_{i=1}^N \frac{TP_i}{TP_i + FP_i + FN_i} \quad (19)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (20)$$

$$\text{Acc} = \frac{TP + TN}{TP + TN + FP + FN} \quad (21)$$

其中,  $TP$  表示真正例;  $TN$  表示真负例;  $FP$  表示假正例;  $FN$  表示假负例。对于  $MIoU$  来说, 这里  $N$  为 3, 表示眼睑、巩膜、背景三个类别。理论上, 上述 4 个评价指标越接近于 1, 则证明分割效果准确。

## 2.4 对比实验结果

本文所提模型的训练损失曲线如图 5 所示。为验证本文所提模型的分割效果, 本文所提模型与 U-Net、UNet++、Attention U-Net、Deeplab v3+<sup>[25]</sup> 进行对比实验, 实验结果各项评估指标见表 1。

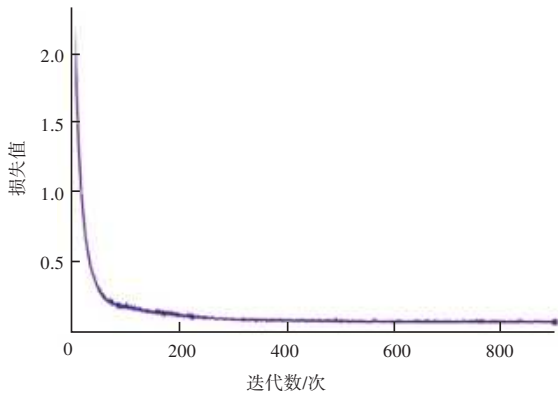


图 5 本文模型训练损失收敛曲线

Fig. 5 Convergence curve of the training loss of the model in this paper

表 1 本文模型与不同网络对比结果

Table 1 Comparison results of the model with different networks in this paper

Net module	Dice	MIoU	Acc	Recall
U-Net	91.87	86.78	97.34	93.31
UNet++	92.06	86.85	97.52	94.22
Att U-Net	92.70	86.50	97.30	94.37
Deeplab v3+	90.66	85.73	96.08	94.03
本文模型	<b>93.33</b>	<b>87.61</b>	<b>98.50</b>	<b>93.30</b>

从表 1 结果可以得出, 本文所提模型的  $MIoU$  为 87.61%、Dice 系数为 93.33%、准确率 (Acc) 为 98.50%、召回率 (Recall) 为 93.30%。其中,  $MIoU$ 、Dice 系数和准确率 (Acc) 较 U-Net 分别提高了 0.96%、1.59%、1.19%, 较 UNet++ 分别提高了 0.88%、1.38%、1.00%, 较 Attention U-Net 分别提高了 0.68%、1.29%、1.23%, 较 Deeplab v3+ 分别提高了 2.19%、2.94%、2.52%。这表明本文模型在 TAO 患者眼睑、巩膜区域能够充分提取有效信息, 较其他模型分割精度均有提升。

图 6 展示了本文所提模型与 U-Net、UNet++、Attention U-Net、Deeplab v3+ 深度学习模型在同一数据集上的分割结果对比。

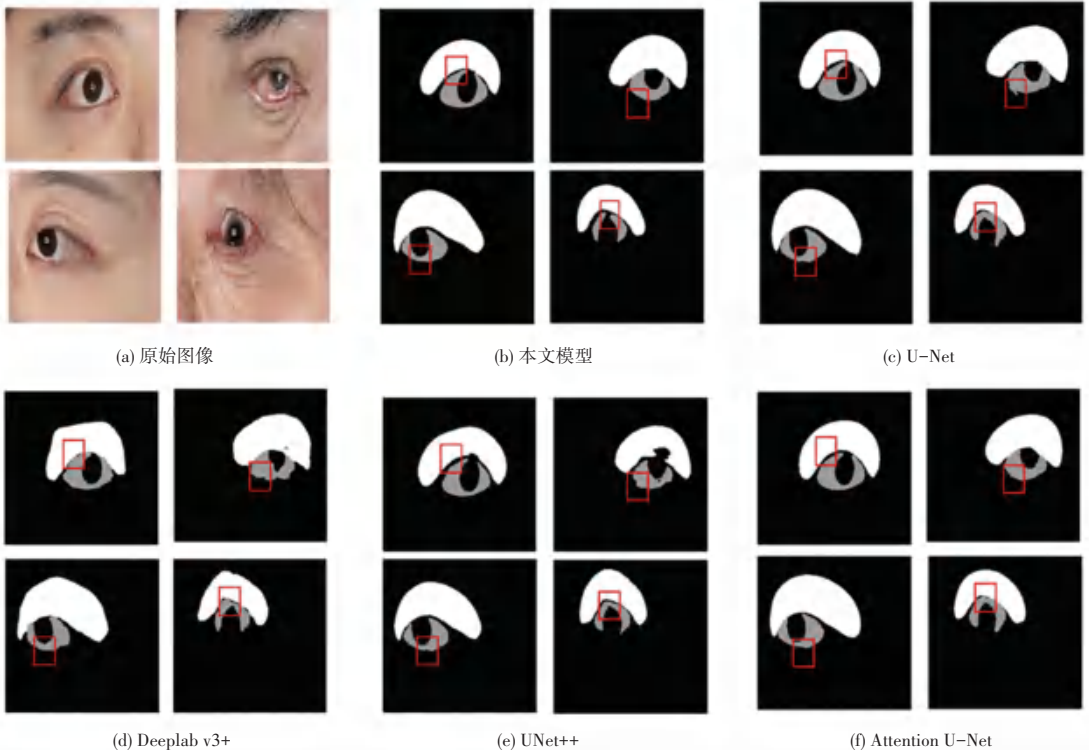


图 6 本文所提模型与不同网络分割结果对比图

Fig. 6 Comparison of the model proposed in this paper with the segmentation results of different networks

从第 1 个图可以看出本文模型对于巩膜上边缘与眼睑的分割效果较对比模型更加流畅, UNet++ 还存在分割缺失的情况, Deeplab v3+ 对于巩膜边缘分割较为粗糙。从第 2 个、第 3 个图可以看出, U-Net、UNet++、Attention U-Net、Deeplab v3+ 均出现了巩膜下边缘的细小分割错误, 且分割效果不清晰。从第 4 个图可以看出眼睑与巩膜边缘区域的细微分割处, 本文模型更加准确, U-Net、UNet++、Attention U-Net、Deeplab v3+ 在患者巩膜上边缘处分割出现错误。综上可知, 本文所提模型可以精准地同时分割眼睑、巩膜区域, 并且保留更完整语义信息, 对比模型存在分割边缘模糊、欠分割的情况, 表明本文所提模型对于眼睑、巩膜区域具有更优良的分割效果。

## 2.5 消融实验

为验证本文所提模型的有效性, 采用消融实验进行验证, 对本文所提出的 PMAC 模块、FAM 模块进行消融实验, 实验结果见表 2。相比于基础网络, *Dice* 系数和 *MIoU* 分别提升了 1.59%、0.96%。

表 2 本文模型不同模块消融实验结果

Table 2 Ablation experimental results of different modules of the model in this paper

Net module	<i>Dice</i>	<i>MIoU</i>	<i>Acc</i>	<i>Recall</i>
U-Net	91.87	86.78	97.34	93.31
Backbone+PMAC	92.72	86.54	98.36	92.94
Backbone+FAM	93.16	87.85	98.24	93.05
本文模型	<b>93.33</b>	87.61	98.50	<b>93.30</b>

图 7 展示消融实验的分割情况。实验结果表明, 引入 PMAC 模块和 FAM 模块均能使分割结果得到提升。从图 7(b) 可以看出, 未引入本文模块时巩膜上边缘与眼睑衔接处出现分割模糊现象。从图 7(c) 和图 7(d) 可以看出引入 PMAC 模块能有效优化巩膜下边缘的分割流畅性, 单独加入 FAM 模块对于巩膜下边缘的分割同样有效, 但是效果稍弱于 PMAC 模块。对比图 7(c) 和图 7(d) 可以看出 PMAC 模块和 FAM 模块对眼睑和巩膜下边缘的分割效果较未使用模块时更加光滑。从图 7(a) 中可以看出, 较未同时引入 PMAC 和 FAM 模块的模型, 同时引入 PMAC 和 FAM 模块时, 巩膜上边缘细微处的分割效果更加准确。

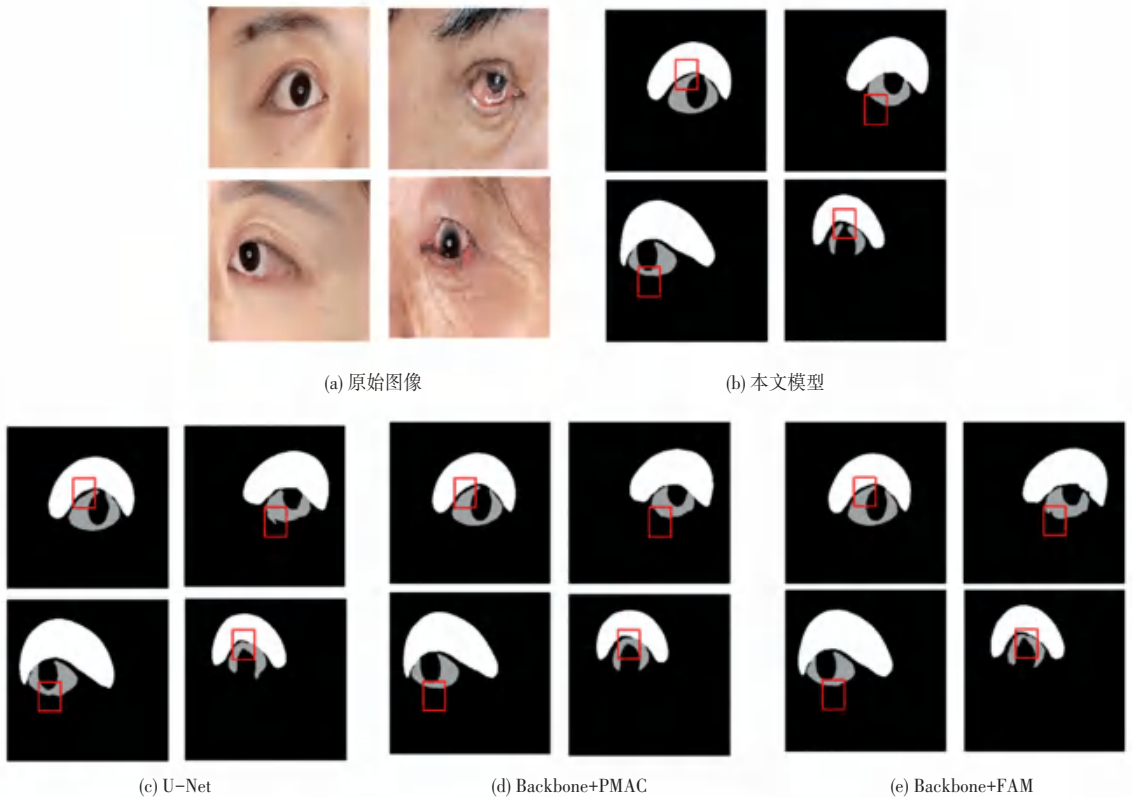


图 7 本文模型不同模块消融实验结果对比图

Fig. 7 Comparison of ablation experimental results of different modules of the model in this paper

消融实验结果表明加入并行多尺度空洞卷积模块后,模型能在一定程度上减少下采样过程的信息丢失,分割边缘区域更加光滑;加入融合注意力机制模块后,模型进一步加强对分割区域的位置与空间有效信息的提取,对眼睑与巩膜边缘的分割更加准确。综合来看,同时引入并行多尺度空洞卷积模块和融合注意力机制模块,在整体上和细节上都能保证分割区域的准确性和流畅性,实现较好的分割效果。

### 3 结束语

针对TAO患者严重性分级的诊断过程患病区域诊断不明确的问题,本文提出的基于并行多尺度空洞卷积和融合注意力机制的分割方法,有效应用于TAO患病区域眼睑、巩膜的多类别分割。该方法以U型网络为基础结构,在网络底部引入并行多尺度空洞卷积模块,在不引入过多参数数量的前提下,有效提取因下采样丢失的不同尺度的细节信息;同时,在下采样与上采样的跳跃连接中加入融合注意力机制模块,有效捕获眼部的高级与低级语义特征并融合通道和空间信息,提高了病变眼睑与巩膜衔接区域的分割准确率,针对数据集样本量少与实际采集条件影响,对医院提供数据集进行亮度、对比度、饱和度增强,实现样本量的扩增与数据增强。最终,本文所提模型在数据增强后的数据上获得了准确的分割效果, $MIoU$ 、 $Dice$ 系数、准确率( $Acc$ )、召回率( $Recall$ )分别达到了93.33%、87.61%、98.50%、93.30%。本文方法与其他参考文献方法相比,在 $Dice$ 系数、精度、召回率、准确率上均获得最佳性能。相比于基础网络, $Dice$ 系数和 $MIoU$ 分别提升了1.59%、0.96%。说明本文所提模型能够有效地对TAO患者眼睑、巩膜患病区域实现多类别同时分割,分割结果有望辅助眼科医生判断TAO患者软组织受累状况。

由于训练数据集的数量还有待补充,一定程度上影响了模型的学习效果和泛化能力,因此,后续将继续与医院展开深度合作,搜集更多TAO患者临床数据,进一步优化模型。

### 参考文献

[1] MALLIKA P S, TAN A K, AZIZ S, et al. Thyroid associated ophthalmopathy—a review[J]. Malaysian Family Physician: The Official Journal of the Academy of Family Physicians of Malaysia, 2009, 4(1): 8–14.  
[2] 中华医学会眼科学分会眼整形眼眶病学组,中华医学会内分泌

学会分会甲状腺学组. 中国甲状腺相关眼病诊断和治疗指南(2022年)[J]. 中华眼科杂志, 2022, 58(9): 646–668.  
[3] BARTALENA L, KAHALY G J, BALDESCHI L, et al. The 2021 European Group on Graves' orbitopathy (EUGOGO) clinical practice guidelines for the medical management of Graves' orbitopathy[J]. European Journal of Endocrinology, 2021, 185(4): G43–G67.  
[4] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C]//18<sup>th</sup> International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015). Cham: Springer, 2015: 234–241.  
[5] ZHOU Zongwei, RAHMAN S M M, TAJBAKHSH N, et al. UNet++: A nested u-net architecture for medical image segmentation[C]//Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4<sup>th</sup> International Workshop (DLMIA 2018) and 8<sup>th</sup> International Workshop (ML-CDS 2018). Cham: Springer, 2018: 3–11.  
[6] OKTAY O, SCHLEMPER J, FOLGOC L L, et al. Attention U-Net: Learning where to look for the pancreas[J]. arXiv preprint arXiv, 1804.03999, 2018.  
[7] HUANG Huimin, LIN Lanfen, TONG Ruofeng, et al. UNet 3+: A full-scale connected UNet for medical image segmentation[C]//2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2020). Piscataway, NJ: IEEE, 2020: 1055–1059.  
[8] CHEN Y, GAN H, CHEN H, et al. Accurate iris segmentation and recognition using an end-to-end unified framework based on MADNet and DSANet[J]. Neurocomputing, 2023, 517: 264–278.  
[9] WANG Caiyong, WANG Yunlong, LIU Yunfang, et al. ScleraSegNet: An attention assisted U-Net model for accurate sclera segmentation[J]. IEEE Transactions on Biometrics, Behavior, and Identity Science, 2019, 2(1): 40–54.  
[10] ROT P, VITEK M, GRM K, et al. Deep sclera segmentation and recognition[M]//UHL A, et al. Handbook of Vascular Biometrics, Advances in Computer Vision and Pattern Recognition. Cham: Springer, 2020: 395–432.  
[11] FUHL W, GEISLER D, ROSENSTIEL W, et al. The applicability of cycle gans for pupil and eyelid segmentation, data generation and image refinement[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. Piscataway, NJ: IEEE, 2019: 4406–4415.  
[12] 徐蓬泉, 梁宇翔, 李英. 融合多尺度语义和剩余瓶颈注意力的医学图像分割[J]. 计算机工程, 2023, 49(10): 162–170.  
[13] 褚张晴晴, 钟志强, 颜子夜, 等. 基于特征融合与注意力机制的脑肿瘤分割算法[J]. 计算机工程, 2023, 49(10): 154–161.  
[14] 胡帅, 李华玲, 郝德琛. 改进U-Net的多级边缘增强医学图像分割网络[J]. 计算机工程, 2024, 50(4): 286–293.  
[15] 梁礼明, 陈鑫, 周珑颂, 等. 基于U-Net多尺度自校准注意力视网膜分割算法[J]. 计算机应用研究, 2023, 40(3): 943–948.  
[16] 陈家毓, 何宏, 朱海鹏, 等. 基于DSE-Net的甲状腺相关眼病患病区域轻量型分割算法[J]. 数据采集与处理, 2023, 38(4): 915–925.  
[17] BADRINARAYANAN V, KENDALL A, CIPOLLA R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE Transactions on Pattern Analysis and



- Machine Intelligence, 2017, 39(12): 2481-2495.
- [18] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(4): 834-848.
- [19] GU Zaiwang, CHENG Jun, FU Huazhu, et al. CE-Net: Context encoder network for 2d medical image segmentation[J]. IEEE Transactions on Medical Imaging, 2019, 38(10): 2281-2292.
- [20] MEHTA S, RASTEGARI M, CASPI A, et al. Espnet: Efficient spatial pyramid of dilated convolutions for semantic segmentation [M]// FERRARI V, HEBERT M, SMINCHISESCU C, et al. Computer Vision-ECCV2018. Lecture Notes in Computer Science. Cham:Springer,2018,11211: 552-568.
- [21] YU F, KOLTUN V. Multi-scale context aggregation by dilated convolutions[J]. arXiv preprint arXiv,1511.07122, 2015.
- [22] HU Jie, SHEN Li, ALBANIE S, et al. Squeeze-and-excitation networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2018: 7132-7141.
- [23] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module [ M ]//FERRARI V, HEBERT M, SMINCHISESCU C, et al. Computer Vision-ECCV2018. Lecture Notes in Computer Science. Cham:Springer,2018,11211: 3-19.
- [24] 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. Piscataway,NJ:IEEE,2016: 770-778.
- [25] CHEN L C, ZHU Yukun, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation [ M ]// FERRARI V, HEBERT M, SMINCHISESCU C, et al. Computer Vision-ECCV2018. Lecture Notes in Computer Science. Cham:Springer,2018,11211: 801-818.