计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2023,32(2):83−93 [doi: 10.15888/j.cnki.csa.008971]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于 Mask R-CNN 卷积神经网络的虹膜分割①

敬红燕, 彭　静, 吴　锡, 李孝杰

(成都信息工程大学 计算机学院, 成都 610225)
通信作者: 彭　静, E-mail: pengj@cuit.edu.cn

摘　要: 针对虹膜图像中存在眼镜遮挡、模糊、角度偏差等不同噪声因素, 我们设计了一种基于 Mask R-CNN 的卷积神经网络 (convolutional neural network, CNN), 命名为 Mask-INet, 用于虹膜分割. 该网络在特征提取阶段为特征金字塔添加了一条自底向上的路径, 既提高了底层到顶层特征的定位信息, 增强语义信息融合, 又进一步加快了底层到顶层的传播效率, 有效提升对虹膜特征提取的准确性. 为了进一步挖掘特征图中的特征信息, 在掩模预测分支阶段, 我们引入上采样和 CBAM 网络 (convolutional block attention module), 利用上采样提高特征图的空间分辨率, 利用 CBAM 网络让特征图中的显著信息更加显著, 增强对特征的判别性. 该方法在 NIR-ISL 2021 比赛提供的虹膜数据集进行了验证. 在相同实验条件下与该赛事的冠军相比, 该方法的各项指标均优于其网络. 与基线 Mask R-CNN 相比, 该方法的 *Dice* 相似系数、平均交并比、召回率分别提升了 8.53%、11.97%、8.88%, 提升了虹膜分割效果.

关键词: 虹膜分割; 特征金字塔; Mask R-CNN; 残差网络; CBAM; 图像分割

引用格式: 敬红燕,彭静,吴锡,李孝杰.基于 Mask R-CNN 卷积神经网络的虹膜分割.计算机系统应用,2023,32(2):83–93. http://www.c-s-a.org.cn/1003-3254/8971.html

Mask R-CNN-embedded Convolutional Neural Network for Iris Segmentation

JING Hong-Yan, PENG Jing, WU Xi, LI Xiao-Jie

(School of Computer Science, Chengdu University of Information Technology, Chengdu 610225, China)

Abstract: In response to different noises in iris images, such as occlusion by glasses, blur, and angle deviation, this study designs a convolutional neural network (CNN) embedded with Mask R-CNN, named Mask-INet, for iris segmentation. The network adds a bottom-up path to the feature pyramid in the feature extraction stage, which not only improves the localization information of bottom-to-top features and enhances semantic information fusion but also further accelerates bottom-to-top propagation efficiency and effectively improves the accuracy of iris feature extraction. To further explore the feature information in the feature map, the study introduces upsampling and a convolutional block attention module (CBAM) network in the mask prediction branching stage. Upsampling is used to improve the spatial resolution of the feature map, and the CBAM network helps make the salient information in the feature map more significant so as to enhance the discrimination capacity for the features. The method is validated on the iris dataset provided by the NIR-ISL 2021 competition. The method outperforms the network of the champion of the event in terms of all indicators under the same experimental conditions. Compared with the baseline Mask R-CNN, the proposed method has the *Dice* similarity coefficient, mean intersection over union (*mIoU*), and recall improved by 8.53%, 11.97%, and 8.88%, respectively, which boosts iris segmentation performance.

Key words: iris segmentation; feature pyramid; Mask R-CNN; residual network (ResNet); convolutional block attention module (CBAM); image segmentation

① 基金项目: 国家重点研发计划 (2020YFA0608001); 国家自然科学基金面上项目 (42075142); 四川省科技厅科技计划 (2022YFG0026, 2021YFG0018, 2020JDTD0020, 2019ZDZX0007)
收稿时间: 2022-06-28; 修改时间: 2022-07-25, 2022-09-01; 采用时间: 2022-09-04; csa 在线出版时间: 2022-12-06
CNKI 网络首发时间: 2022-12-07

1 引言

虹膜是位于黑色瞳孔和白色巩膜之间的圆环状部分, 这个圆环部分里面包含了许多相互交错的细节特征, 而在生物特征中, 虹膜的这些细节特征并不会随着年龄的增长而发生任何改变, 因此虹膜可以作为识别生物身份的一种方法. 虹膜特征也因其独特性、稳定性以及不可更改性在国防和安全方面均发挥着十分重要的作用[1,2]. 虹膜识别因此也被认为是 21 世纪最有前途的生物识别技术[3], 被广泛应用于各种生物特征识别应用, 包括智能解锁[4]、边境控制[5]、取证等. 一个完整的虹膜识别系统流程通常包括以下 4 个步骤: 虹膜图像采集、虹膜预处理、特征提取和匹配[6]. 作为虹膜预处理的一部分, 虹膜分割定义了用于特征提取和匹配的图像区域, 因此直接影响虹膜整体识别性能[7]. 由此可见虹膜分割在虹膜识别中占有重要意义, 分割算法的鲁棒性和准确性直接决定了后续虹膜特征提取、验证和识别[8].

近年来, 基于深度学习的图像处理研究越来越深入, 取得的成就也非常可观. 相较于传统的虹膜分割方法, 基于深度学习的虹膜分割方法更加具有鲁棒性和识别性. 2015 年, Long 等人提出全卷积神经网络 (fully convolutional network, FCN)[9] 开创了语义级别的图像分割先河, 随后各种语义分割网络如雨后春笋, 纷纷涌现. 如 UNet[10]、SegNet[11]、PsPNet[12] 以及 DeepLab 系列. 其中 DeepLab 系列作为语义分割的经典模型, 取得了非常不错的分割成果. DeepLabv1[13] 针对池化降低分辨率问题, 提出了空洞卷积来扩展视野, 以便获取更多的上下文信息. DeepLabv2[14] 主要贡献在于提出了空洞空间金字塔池化 (atrous spatial pyramid pooling, ASPP), 其使用不同采样率的空洞卷积并行采样生成多尺度特征图, 用于处理尺度可变性问题. DeepLabv3[15] 基于图像层次将全局背景进行编码得到图像级特征, 增强了 ASPP, 进一步提升了分割精度. 2017 年 Transformer 横空出世, 随着 Transformer 在自然语言领域的应用和普及, 其也逐渐被应用到计算机视觉领域中. Valanarasu 等人[16] 提出的 MedT 基于 Transformer 的编码器架构来分割医学图像, 取得了很好的性能. 最早将深度学习方法应用到虹膜分割领域是 Jalilian 等人[17] 提出的全卷积编码解码网络 (FCEDNs), 实验表明 FCEDMNs 的分割结果优于传统的算法. Lian 等人[18] 提出了 ATT-UNet, 将 UNet 与注意力结合, 取得了不错的分割效果. Wang 等人[19] 基于转移学习提出一种新的训练方法, 该方法是以 ResNet34 为骨干的经典 UNet 架构, 采用两个基于 UNet 的独立模型来执行虹膜的分割和定位任务, 提高了模型的泛化能力. 最终获得了 NIR-ISL 2021 虹膜比赛的冠军. 由于语义分割只根据不同语义像素进行分割, 不能区分不同实体, 因此出现了实例分割, 即在检测到目标以后再对其进行分割. He 等人[20] 提出的 Mask R-CNN 在 Faster R-CNN[21] 的基础上增加了一个分支用于语义分割, 即对检测到的目标框进行目标分割. 从而实现实例分割并且通过大量实验证明该网络模型达到了较高的分割精度. 上述基于深度学习的分割算法网络被广泛应用于图像分割领域, 在虹膜分割任务中也取得了不错的成果. 但是对于存在各种噪声因素的困难样本分割效果不是很好, 存在分割边缘不够圆滑、漏分、错分等情况.

针对这一问题, 本文提出了基于 Mask R-CNN 神经网络的 Mask-INet 模型. 该模型以 Mask R-CNN 为基本框架, 结合了特征金字塔和混合注意力机制. Mask R-CNN 网络在语义分割的基础上对同类的物体能够进行更加精细的分割, 能进一步减少虹膜错分的情况. 我们利用特征金字塔高层特征进行上采样与底层特征自下而上的连接结构来提取虹膜图像特征. 特征金字塔这种连接结构融合了低分辨率语义信息较强的特征图和高分辨率语义信息较弱但空间信息丰富的特征图, 加深了对虹膜特征信息的挖掘. 特征金字塔在提取虹膜特征信息过程中虽然容易获取高层语义信息但是对于底层的定位信息却难以获取, 会导致检测精度不高. 为了提升虹膜最终的分割精度, 给特征金字塔添加了一条自下而上的路径, 这条路径不仅充分利用高层语义信息同时也充分利用了底层定位信息来挖掘虹膜特征, 极大提高了检测精度. 为了进一步挖掘虹膜特征, 我们在掩模预测分支引入两个上采样层将虹膜特征图的空间分辨率放大 4 倍. 再将注意力机制引入改进的掩模预测分支中, 注意机制能加强虹膜特征区域的权重信息, 抑制无关信息对模型的打扰, 能有效提升模型的分割精度和灵敏度.

2 准备工作

2.1 残差网络

卷积神经网络是目前计算机视觉领域中主要的特征提取技术[22]. 传统的卷积神经网络在信息传递

过程中总是存在着信息丢失、梯度消失或梯度爆炸的问题, 因此不能训练很深的网络. 理论上, 越深的网络输入表示能力越强的特征, 但随着网络的加深, 参数量和计算力也会增加, 从而影响网络的训练效果造成网络退化. He 等人[22] 提出的残差网络很好地解决了这个问题. 残差网络由一个个残差单元模块叠加组成, 一个残差单元的输入与输出可以用式 (1) 表示:

$$\begin{cases} y_l=h(x_l)+F(x_l,W_l) \\ x_{l+1}=f(y_l) \end{cases} \tag{1}$$

其中, x_l 和 x_{l+1} 分别代表当前这个残差单元的信号输入和输出, l 代表层, $h(x_l)$ 代表当前残差单元的恒等映射即 identity mapping, $F(\cdot)$ 是残差部分, 一般由 2 个或者 3 个卷积操作构成, W_l 表示为该部分卷积的权重, $f(\cdot)$ 为 ReLU 激活函数.

残差单元的残差式跳跃结构打破了传统 $n-1$ 层的输出只能传给 n 层的惯例, 使得随意某层的输出可以跳跃多层作为后层的输入, 这样做的好处在于虽然增加了网络结构的深度但网络并未退化且训练效果非常好, 为以后在模型叠加上提供了新的方向.

2.2 特征金字塔

识别不同尺寸的目标是计算机视觉的一个基本挑战[23]. 在目标检测任务中, 许多网络如 YOLO1[24], 利用卷积层提取特征, 经过多个池化层输出小尺度的特征图, 利用这个单个特征图进行后续的分类和边界框的回归, 但是对于目标大小不同的物体来说存在一定的缺陷. 因此文献 [23] 提出了特征金字塔网络.

传统的特征提取使用人工, 在图像金字塔上构建特征金字塔 (简称特征化图像金字塔)[25], 其任务是提取不同尺度图片的特征. 但是这样做增加计算量的同时还会消耗大量内存. 随后人们使用深度卷积网络 (ConvNets) 提取特征, 其做法是直接取高层语义特征进行预测, 但是由于感受野的问题, 对于小特征的物体可能存在检测不到的情况. 为了改善上述问题, SSD[26] 提出输出不同尺度的特征图预测. 但底层特征图的语义信息薄弱导致出现虽然框选出小物体但很容易将小物体错分类的情况. 针对上述 3 种问题, FPN 提出基于自底向上提取各层语义特征, 进行自顶向下的连接中融合自底向上的特征图, 再输出各个尺度特征图的预测.

FPN 的结构图如图 1 所示, 自底向上的过程中利用 ResNet 每级最后一个残差块的输出作为预测特征的输入, 其分别对应输入图片的下采样倍数为{4, 8, 16, 32}. 自顶向下的过程中通过上采样的方式将顶层的小特征图放大到同上一个 stage 的特征图尺度相同, 再同经过 1×1 卷积的残差块最后一层的特征图作逐元素相加操作.

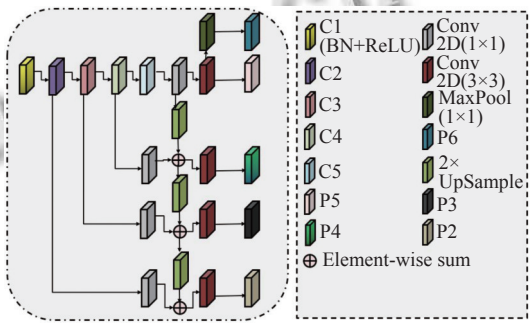


图 1 特征金字塔结构

2.3 CBAM

近年来, 为了提升网络的性能, 研究人员不仅将研究重心放在了如何利用网络的宽度、深度和基数这几个因素来提升网络的性能. 同时对注意力的研究也不断加深. Google mind 团队在文献 [27] 中将注意力机制引入 RNN 模型中进行图像分类, 注意力才正式进入计算机视觉领域.

在掩模预测分支进行特征提取时, 一些分布在特征图某些通道中的无效信息会被保留下来影响虹膜的分割精度. 因此, 我们引入 CBAM 强调所需分割的目标. 相较于 SE-Net[28] 只关注通道的注意力机制而言, CBAM 结合了空间注意力机制和通道注意力机制[29], 不仅考虑到不同通道像素的重要性, 还考虑到同一通道的不同位置的像素重要性. CBAM 是基于注意力机制的轻量级通用模型, 能融入到各种常规的卷积层中.

对上采样后的特征图, CBAM 从通道和空间两个维度计算该特征图的注意力图, 增强对目标物体的识别. 结构如图 2 所示, CBAM 由通道注意力机制和空间注意力机制串行组成, 通道子模块在共享网络中分别使用最大池化和平均池化聚合特征图的空间信息, 生成两个不同上下文的描述符, 对得到的两个特征做相

加和 Sigmoid 处理得到通道注意力图 M_c. 空间子模块利用沿通道轴汇聚的相似的两个输出, 并将它们转发到卷积层, 再做 Sigmoid 处理得到空间注意力图 M_s. CBAM 将得到的 M_c 同输入的 feature 按元组进行乘法操作再经过空间注意力机制得到 M_s 与通道注意力机制得到的特征同样按元素进行乘法操作. 可用如下公式表示:

$$F' = M_c(F) \otimes F \tag{2}$$

$$F'' = M_s(F') \otimes F' \tag{3}$$

其中, F 表示输入特征, $M_c(F)$ 表示经过通道注意力机制所获取的特征. $\otimes$表示按元素相乘. $M_s(F')$ 表示将通道注意力机制所得到的特征经过空间注意力机制得到的特征, F''表示 CBAM 最终获取的特征.

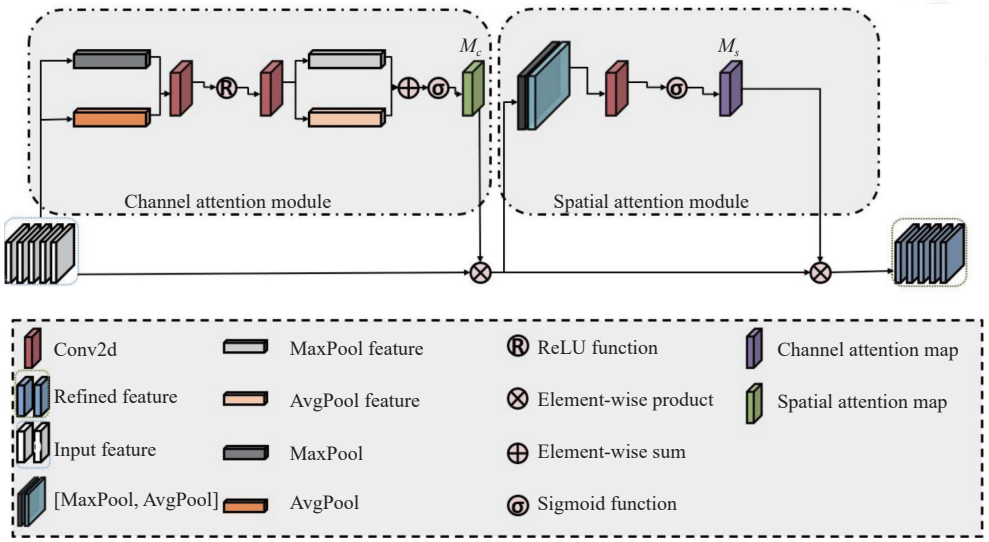


图 2 CBAM 网络结构

3 Mask-INet 算法

本文使用 ImageNet[30] 预训练的残差网络作为主干网络, 结合改进后的特征金字塔, 在掩模预测阶段利用上采样层来增大特征图的空间分辨率, 进一步挖掘虹膜特征的空间信息. 随后引入第 2.3 节介绍的 CBAM 机制激励重要特征信息, 抑制无用信息. 再同未经过 CBAM 网络的特征图进行逐元素相加, 从而提升虹膜的分割精度. Mask-INet 是一种端到端的网络, 以处理成 coco 格式的虹膜图像为输入, 经多次训练后输出虹膜的掩码和定位. 该网络总体结构如图 3 所示.

Mask-INet 模型使用残差结构有效避免了因网络加深而造成的模型退化问题. 同 FPN[[23] 相比, 改进后的 FPN 不仅保留了 FPN 模块融合后浅层高分辨率的特征如高层的语义信息和底层的边缘信息同时也充分利用了底层的边缘信息来获取定位信息, 使得每层特征既具有高层语义信息又具有底层的定位信息. 由于本文数据属于红外图像, 而在红外图像中, 背景和目标的对比度较低, 不同实例的辨别主要依靠轮廓特征信息, 因此在掩模预测分支中添加注意力机制来提升对有效特征提取的能力.

3.1 PA-FPN

在 Mask R-CNN 中引入特征金字塔网络结构能较好地检测不同尺度的物体, 但是在实验过程中仍存在检测精度较差的情况, 容易造成虹膜漏分. 这是由于 FPN 是自顶向下的模式, 只将相邻的特征进行直接融合, 而底层特征却无法影响高层特征, 也就是说高层特征难以获取底层特征的定位信息, 因此导致难以对目标进行精确定位[31].

为了充分利用底层定位信息, 提高对目标的检测精度, 为 FPN 添加了一条自底向上的路径, 如图 4, 该结构使得低层信息更容易传递到高层顶部, 有效利用底层的定位信息, 提高了检测精度. 之前底层特征只能通过特征金字塔, 现在能直接通过该结构传播到顶层, 进一步提高了传播效率. 为了方便后续使用, 将其简称为 PA-FPN.

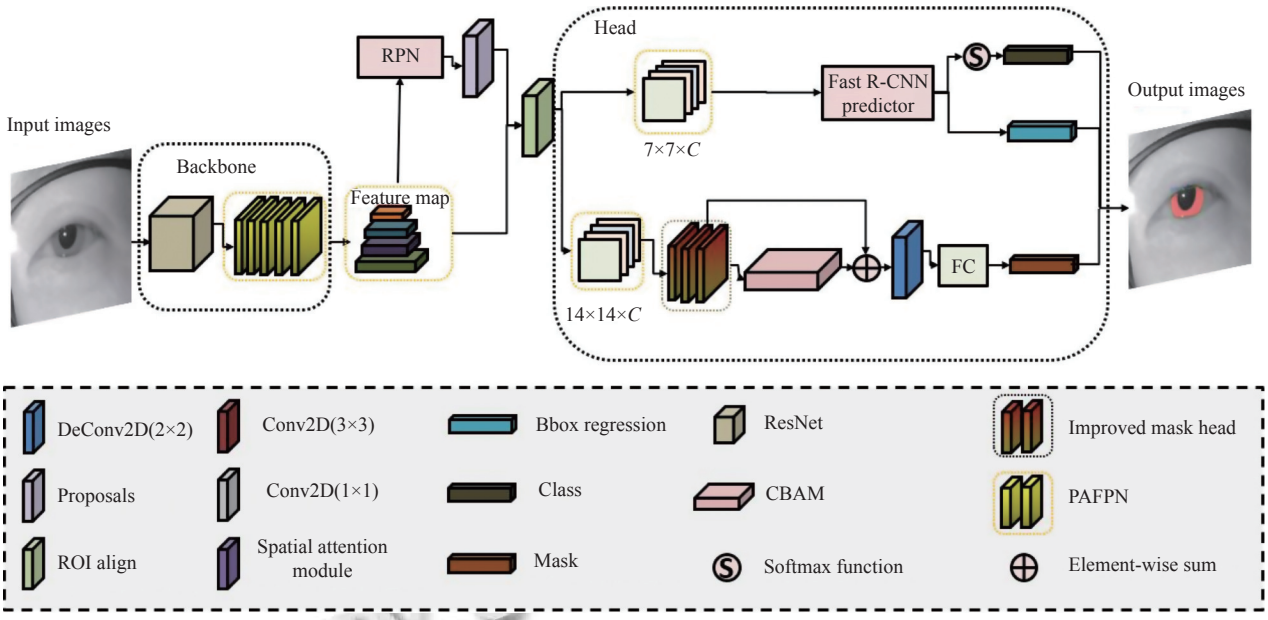


图 3 Mask-INet 模型的网络架构概述

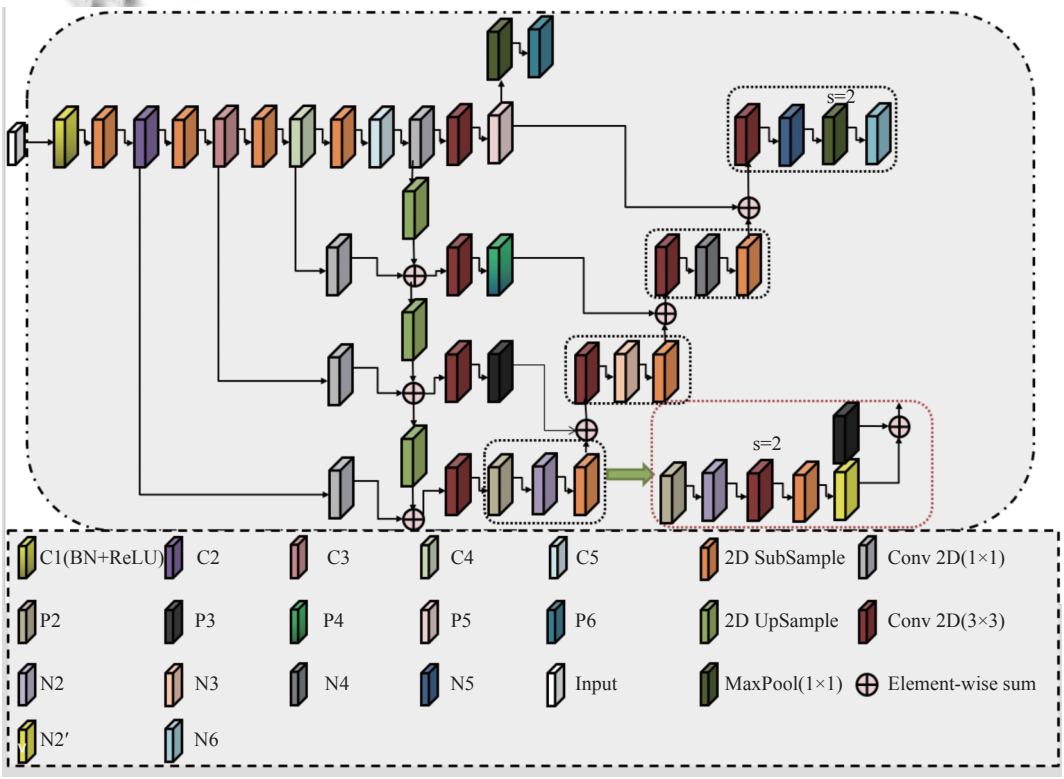


图 4 PA-FPN 结构图

图中输入为残差网络模块输出的权重向量. 具体如表 1 所示, 输入图片大小为 512×512×3, C2 到 C5 通道数列表为{64, 128, 256, 512}分别对应表 1 中 P2 到 P5 的输入尺寸. 为了使每个特征图在融合时保持一致, 需要对每个特征层得到的特征图进行一个 1×1 卷积操作, 再将高层特征图进行一次 2 倍上采样与同尺度特征图进行融合, 最后经过一个 3×3 的卷积得到 P2 到 P5, 通道数均为 256. 得到的特征金字塔特征图列表为

{P2, P3, P4, P5, P6}, 如图 3 红色虚线框部分, 令 P2=N2, 对 N2 进行 stride=2, kernel=3×3 的卷积操作, 然后再进行下采样, 得到为原来的一半的特征图记为 N2′, 将 N2′与 P3 进行逐元素融合, 为了消除上采样带来的混叠效果, 再通过一个 3×3 的卷积层, 最后生成特征图 N3, 在这过程中通道数均为 256. 重复此操作, 最后输出特征图记为{N2, N3, N4, N5, N6}, 其中 N6 是 N5 通过一个大小为 1×1, 步距为 2 的最大池化层实现下采样得到的. {N2, N3, N4, N5}的空间分辨率与横向连接传递的{P2, P3, P4, P5}互相对应. 我们将最终得到的金字塔特征图{N2, N3, N4, N5, N6}用于后续 RPN 网络的输入.

3.2 Improved-MaskHead (I-MH)

Mask R-CNN 中有 3 个分支, 一个用于预测分类、一个用于预测边界框回归、一个用于预测分割掩码, 这 3 个分支并行处理, 为每一个检测类别独立预测出掩码从而消除跨类别的竞争. 尽管 Mask R-CNN 相较其他算法有一定的优越性, 但是在掩模分支处理虹膜特征图的过程中发现空间分辨率较低, 导致信息损失较多, 所以本文在原始的掩码分支中加入了上采样层, 将原始特征图的分辨率增大了 4 倍. 预测虹膜图像掩模的本质问题是语义分割, 而语义分割对特征图的空间分辨率更为敏感, 分辨率更高的特征图, 更接近原的分辨率, 信息损失更少, 更有助于分割语义信息. 本文使用带有参数的反卷积层来进行上采样操作, 因为带有参数的反卷积层比不带参数的反卷积层更能适应不同的任务, 更具泛化能力.

Mask 分支模块的结构图如图 5 所示, 输入为经过 RoI Align 得到的特征, 通道大小为 256, 最终输出特征图大小为 56×56×2. 具体参数如表 2 所示, 其中 num_{cls} 代表类别数这里为 2.

表 1 PA-FPN 结构

Layer name	Input size	Output size	PA-FPN layer
P5	64×64×512	64×64×256	1×1, 3×3, stride 1
P4	64×64×512 128×128×256	128×128×256	1×1, 3×3, stride 2
P3	128×128×256 256×256×128	256×256×256	1×1, 3×3, stride 2
P2	256×256×128 512×512×64	512×512×256	1×1, 3×3, stride 2
N2	512×512×256	512×512×256	3×3, stride 1
N3	512×512×256 256×256×256	256×256×256	[3×3]×2, stride 2
N4	256×256×256 128×128×256	128×128×256	[3×3]×2, stride 2
N5	128×128×256 64×64×256	64×64×256	[3×3]×2, stride 2
N6	64×64×256	32×32×256	1×1, stride 2

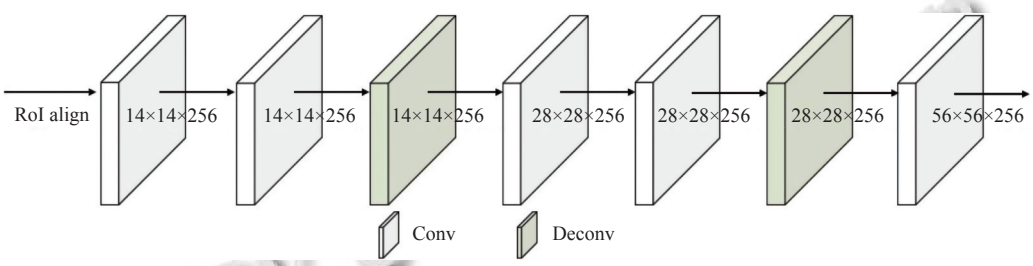


图 5 Mask 分支结构图

改进后的 mask 分支是由 4 个 3×3 的卷积层, 两个 2×2 的反卷积层和一个全连接层组成. 表 2 中 RoI Align 输入尺寸为 h×w×256 的特征图, 其中 h×w 是指输入任意的空间分辨率. 在这过程中通道数始终保持为 256. RoI Align 输入的特征图利用表中 mask_fcn1、mask_fcn2、mask_fcn3、mask_fcn4 等 4 个卷积层对特征进行空间信息挖掘, 同时利用 mask_deconv1、mask_deconv2 两个反卷积层将特征图的分辨率增大, 增强特征的空间信息, 便于生成质量更好的虹膜图像掩码. 最后通过一个 1×1 的卷积, 卷积核个数为分类个数, 得到预测后的每个类别的 mask 且大小均为 56×56.

表 2 添加了上采样层的掩模预测分支结构

Layer name	Input size	Output size	Mask branch layer
RoI Align	h×w×256	14×14×256	Max pool
mask_fcn1	14×14×256	14×14×256	3×3, 256, stride 1
mask_fcn2	14×14×256	14×14×256	3×3, 256, stride 1
mask_deconv1	14×14×256	28×28×256	2×2, 256, stride 2
mask_fcn3	28×28×256	28×28×256	3×3, 256, stride 1
mask_fcn4	28×28×256	28×28×256	3×3, 256, stride 1
mask_deconv2	28×28×256	56×56×256	2×2, 256, stride 2
mask_fcn5	56×56×256	56×56×num_{cls}	1×1, num_{cls}, stride 1

3.3 损失函数

在训练模型的过程中, 需要通过损失函数来定义该模型预测的好坏及优化的目标. 损失函数越小表明模型的鲁棒性越好.

总的损失函数为式 (4), L 是 3 个损失函数的总和.

$$L = L_{\text{cls}} + L_{\text{box}} + L_{\text{mask}} \tag{4}$$

其中, L_{cls} 为分类损失函数, 表示如式 (6):

$$S_i = \frac{e^{a_i}}{\sum_{k=1}^{T} e^{a_k}} \tag{5}$$

$$L_{\text{cls}} = -\sum_{i=1}^{T} y_i \log S_i \tag{6}$$

本文基于 Softmax 函数来计算 RPN 网络的交叉熵损失. 式 (5) 中符号 a_i 表示类别 i 经过网络前向传播后所得分, T 为分类的类别个数, S_i 表示类别 i 经 Softmax 函数计算得到的概率. 式 (6) 中 y_i 表示真实标签, S_i 表示所得概率.

式 (4) 中 L_{box} 为回归损失函数, 表示如式 (7):

$$L_{\text{box}} = \begin{cases} 0.5 \times x^2, \text{ if } |x| < 1 \\ |x| - 0.5, \text{ otherwise} \end{cases} \tag{7}$$

其中, $x = f(x_i) - y_i$ 为真实值与预测值之间的数值差值.

式 (4) 中 L_{mask} 为分割损失函数, 表示如式（8）:

$$L_{\text{mask}} = -\sum_{k=1}^{n} \widehat{y_k} \log y_k + (1 - \widehat{y_k}) \log (1 - \widehat{y_k}) \tag{8}$$

输入掩模预测分支的图像会经过一系列的卷积层、反卷积层之后输出总的类别的特征图, 这一分支的损失函数定义为平均二值交叉熵损失函数. 式 (8) 中 n 代表 n 种类别, $\widehat{y_k}$ 代表模型预测样本为正的概率, y_k 代表样本真实标签, 如果样本为正, 取值 1, 否则取值 0.

4 数据集

4.1 实验数据及数据增强

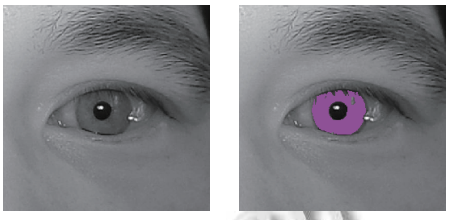
本文借助 2021 年举办的 NIR-ISL 2021 比赛—一项与 IJCB 2021 联合举办的基准测试挑战比赛中所提供的数据集, 包括 CASIA-Iris-Asia、CASIA-Iris-M1 和 CASIA-Iris-Africa[19].

CASIA-Iris-Asia 包含了亚洲人在非合作环境的各种近红外虹膜图像[19]. 该数据集是由 CASIA-Iris-Distance 和 CASIA-Iris-Complex 所组成. 这些数据集使用不同的移动设备从不同场景和环境下获得. 该比赛提供的 CASIA-Iris-Complex 数据集共 1 000 张, 其中包括遮挡虹膜图像 500 张和虹膜角度偏离图像 500 张. CASIA-Iris-Distance 数据集 400 张. 本文从这 3 部分数据集随机平等的抽取共 900 张虹膜图像作为训练数据, 剩余的图像作为测试数据.

CASIA-Iris-M1 是一个大型的近红外移动虹膜数据集, 包括 3 个子集: CASIA-Iris-M1-S1、CASIA-Iris-M1-S2 和 CASIA-Iris-M1-S3. 从 3 个子集中随机且平等的选择共 1 800 张图像作为训练数据, 以同样的方式选择不相交的 600 张图像作为测试数据.

CASIA -Iris-Africa 是非洲第一个大规模的黑人虹膜数据集. 从中随机平等的选取 400 张不同噪声类型的虹膜图像作为训练数据, 以同样的方式选择不相交的 250 张图像作为测试数据.

我们从 3 个数据集中抽取不同类型的虹膜图像, 组成我们的数据集. 比赛方提供的数据集为一般的分割数据集, 本文将该数据集处理成 coco 数据集格式以便使用. 所给原始数据如图 6(a) 所示, 虹膜的位置为标记区域, 如图 6(b) 所示.



(a) Eye (b) Iris

图 6 眼睛数据

一般而言, 成功的神经网络需要大量的参数, 而能使模型正常工作的大量参数是需要训练海量数据才能得到. 在实际情况中, 数据的获取非常困难, 不仅消耗大量人力财力还浪费时间. 所以对于数据量较小的数据集, 一般会采用数据增强. 本文虹膜数据集中的训练集图片数量较少, 很容易导致训练过程出现过拟合的情况, 因此使用数据增强来提高模型的泛化能力和鲁棒性. 常用的数据增强方法包括图像旋转、翻转、裁剪、缩放、移动等, 本文所用到的数据增强方法包括水平翻转、垂直翻转、裁剪和缩放. 最后将虹膜图像统一裁剪成 512×512 的固定大小.

5 实验结果与分析

5.1 实验环境及细节

本实验是在一台小型深度学习服务器上开展的,

该服务器的具体配置参数如下: 操作系统是 Ubuntu 18.04LTS, CPU 型号为 I7-7700K 内存大小为 128 GB, 显存为 11 GB 的 Nvidia GTX 2080Ti 显卡. Python 版本为 3.6, PyTorch 版本为 1.6.

该实验采用 Adamw 优化算法加快收敛速度, 设置 weight-decay 为 0.05, 训练 epoch 总数为 100, 每个 epoch 迭代次数为 1 000, Batch Size 设置为 8, 初始的学习率为 0.000 2, 每 20 个 epoch 学习率衰减为原来的 0.1. 绘制的训练集损失值变化如图 7 所示, 横坐标表示训练 epoch 次数, 纵坐标表示损失值. 当训练到 70 个 epoch 后模型损失值趋于收敛.

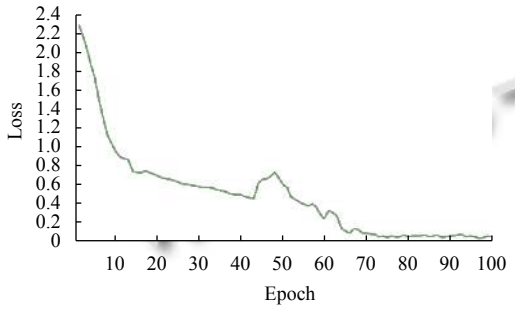


图 7 损失值变化

5.2 评价指标

为了验证所提出网络结构的有效性, 本文与不同方法作了对比实验. 实验结果评价指标包括常用的分割指标召回率 (*Recall*)、*Dice* 相似系数 (*Dice*)、平均交并比 (*mIoU*).

Recall 是衡量被标注为正的样本占总样本比例的重要指标. 如式 (9) 所示, 其中 *TP* 表示被预测为正样本, 真实值也为正样本; *FN* 表示被预测为负样本, 但真实值为正样本.

$$Recall = \frac{TP}{TP+FN} \tag{9}$$

Dice 是用于衡量真实掩码与分割结果相交集合相似性的度量指标, 如式 (10) 所示, 其中$|X \cap Y|$表示 X 和 Y 两个集合之间的交集. 分子系数设置为 2, 是因为分母重复计算 X 和 Y 之间相同的元素, 为了保证计算结果在 [0, 1] 之间. $|X|+|Y|$表示 X 和 Y 两个集合的元素总数量.

$$Dice = \frac{2|X \cap Y|}{|X|+|Y|} \tag{10}$$

mIoU 是衡量所有类别交集与并集之比的平均值的指标. 如式 (11) 所示, 其中 k 代表类别数, *FP* 表示被预测为正样本, 真实值为负样本.

$$mIoU = \frac{1}{k}\sum_{i=1}^{k}\frac{TP}{FN+FP+TP} \tag{11}$$

5.3 骨干网络选择实验

本文分别将 ResNet18、ResNet50 和 ResNet101 作为 Mask R-CNN 的 Backbone, 通过表 3 结果显示, 在不同场景下的虹膜分割任务中, ResNet50 作为 backbone 时, *Dice*、*mIoU* 和召回率均为最优. 因此本文采用 ResNet50 作为本文的 backbone.

表 3 Mask R-CNN 在 3 种不同 backbone 下的评估对比结果 (%)

Backbone	*Dice*	*mIoU*	*Recall*
ResNet18	85.11	77.53	85.17
ResNet50	87.09	80.81	87.48
ResNet101	86.34	78.76	86.62

5.4 FPN 实验

随着网络层数的加深, 检测所需的位置信息越差, 为了进一步增强特征图的语义信息以及目标物体的检测精度, 在 FPN 中引入一条自底而上的路径 (称为 PA-FPN).

由表 4 可以知道, 使用 PA-FPN 进行训练得到模型的 *Dice* 值达到了 92.43%, 相较原始的 FPN 模型 *Dice* 提升了 2.65%, *mIoU* 提升了 3.15%, 这说明 PAFPN 能有效提升图像的分割精度.

如图 8 所示是 PA-FPN 与原始 FPN 对比结果, 可以看出原始 FPN 对斜视图像的分割效果较差, 存在漏分的情况, 且虹膜边界分割比较粗糙. 相比之下 PA-FPN 分割出的虹膜内外边界更加平滑, 分割的最终效果更接近真实标签.

表 4 PA-FPN 同原模型性能评估对比结果 (%)

Method	*Dice*	*mIoU*	*Recall*
Mask R-CNN	87.09	80.81	87.48
FPN	89.78	85.86	90.71
PA-FPN	92.43	89.01	93.96

(a) Original image (b) Label (c) PA-FPN (d) FPN

图 8 FPN 与 PA-FPN 的模型预测结果对比

5.5 掩码预测分支改进实验

虹膜分割在虹膜识别中具有十分重要的作用, 本质上是对虹膜进行语义分割. 而在语义分割中, 空间分辨率的大小会影响对特征空间信息的挖掘. 本文利用上采样进一步提取空间信息, 如表 5 所示, 添加了上采

样的 Mask R-CNN 在 *Dice*、*mIoU* 指标上分别提升了 2.75%、2.91%.

表 5 I-MH 方法同原模型的对比结果 (%)

Method	*Dice*	*mIoU*	*Recall*
Mask R-CNN	87.09	80.81	87.48
I-MH	89.84	83.72	90.27

5.6 不同注意力机制模块的横向对比实验

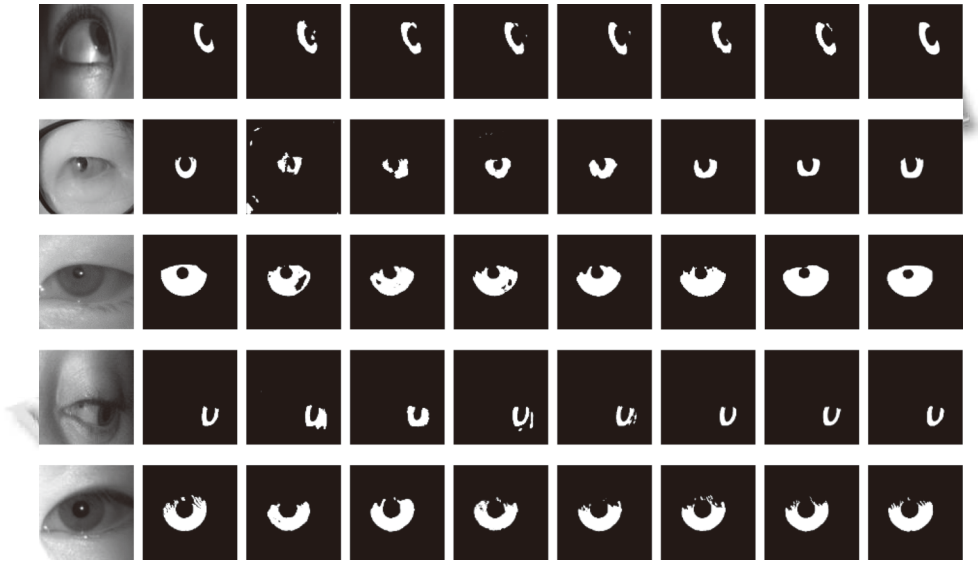
为了进一步挖掘虹膜的特征信息, 我们在第 3.2 节的 I-MH 中引入注意力机制, 并同未加注意力以前的特征图作逐元素相加操作, 进一步融合特征信息. 注意力机制能够让模型更加关注实例区域, 减少不相关的信息对目标检测性能的影响. CBAM 能同时兼顾空间和通道的特征信息, 能获取更好的效果. 本文同 SENet、ECA-Net[32] 等不同的注意力机制作了对比. 从表 6 可以看出, 添加注意力机制能有效提升虹膜的分割精度. 相比其他注意力机制而言, CBAM 的各项指标明显更好. 同未加注意力机制的模型对比, *Dice* 值、*mIoU* 分别提升了 2.02%、3.11%.

5.7 不同方法性能对比

为了说明 Mask-INet 模型的有效性和准确性, 我们在相同的数据集上, 使用具有相同实验条件的各种模型进行了对比实验, 实验结果采用相同的评价标准. 不同算法的分割结果如图 9 所示, 其中第 1 列为虹膜图像、第 2 列为虹膜的真实标签、第 3 列为 UNet[10] 的分割掩码、第 4 列为 SegNet[11] 的分割掩码、第 5 列为 PSPNet[12] 的分割掩码、第 6 列为 DeepLabv3[15] 的分割掩码、第 7 列为 T-UNet[19] 方法的分割掩码, 第列为文献 MedT[16] 的分割掩码结果, 最后一列是本文方法的分割掩码. 实验结果表明, UNet、SegNet、PSPNet、DeepLabv3、T-UNet、MedT 等方法对于包含斜视、模糊、眼镜遮挡、睫毛遮挡、瞳孔收缩等困难样本的分割效果较差, 存在错分、漏分的情况, 且虹膜边缘部分分割不够圆滑, 而本文提出的方法虹膜分割结果依然精确, 且边缘更加光滑, 分割结果更接近真实标签, 有效提升了分割精度.

表 6 不同注意力机制对本文测试数据集的评估对比结果 (%)

Mask R-CNN	I-MH	SENet	ECA-Net	CBAM	*Dice*	*mIoU*	*Recall*
√	√	—	—	—	89.84	83.72	90.27
√	√	√	—	—	90.04	85.11	90.30
√	√	—	√	—	90.97	85.95	90.73
√	√	—	—	√	91.86	86.83	91.04



(a) Original image (b) Label (c) UNet (d) SegNet (e) PSPNet (f) DeepLabv3 (g) T-UNet (h) MedT (i) Our model

图 9 不同方法的对比结果

表 7 展示了不同虹膜分割方法的定量指标结果, 可以看出本文方法在 *Dice* 相似系数、平均交并比和召回率 3 项评价指标上均优于其他几种方法. 在测试集上, *Dice* 相似系数上达到了 95.62%, 平均交并比达到了 92.78%, 召回率达到了 96.36%.

5.8 消融实验

为了验证各个模块对整体模型的有效性, 基于 Mask R-CNN 网络、以 ResNet50 为骨干网络, 加入

PA-FPN 结构、在掩码预测分支引入上采样和 CBAM 模块构成不同网络模型进行消融实验. 如表 8 所示, 相较其他方法, 本文方法在 *Dice* 相似系数、平均交并比和召回率的值上均取得了最高值. 分割精度在原模型的基础上 *Dice* 相似系数提升 8.53%, 平均交并比提升了 11.97%, 召回率提升了 8.88%, 充分证明了本文提出方法的有效性

表 7　不同方法性能评估对比结果 (%)

Method	*Dice*	*mIoU*	*Recall*
UNet[10]	90.56	83.81	90.30
SegNet[11]	90.72	84.15	91.48
PSPNet[12]	91.37	86.69	92.87
DeepLabv3[15]	91.53	87.66	92.71
T-UNet[19]	92.54	90.25	93.37
MedT[16]	93.78	91.16	95.23
本文方法	95.62	92.78	96.36

表 8　模型性能对比结果 (%)

Mask R-CNN	PA-FPN	I-MH (include CBAM)	*Dice*	*mIoU*	*Recall*
√	—	—	87.09	80.81	87.48
√	√	—	92.43	89.01	93.96
√	—	√	91.86	86.83	91.04
√	√	√	95.62	92.78	96.36

由图 10 可以看出, 本文改进的 Mask R-CNN 能够有效准确的分割出模糊样本中完整的虹膜, 分割结果接近真实标签.

(a) Original image (b) Label (c) Mask R-CNN (d) Ours

图 10　改进后模型对比结果

6　结论与展望

受残差网络、特征金字塔、注意力机制等多种网络的启发, 本文提出了基于 Mask R-CNN 网络的 Mask-INet 模型, 该模型对存在遮挡物、斜视、模糊等虹膜困难样本进行了有效、准确的分割. 本文的关键在于为特征金字塔添加了一条自底向上的路径, 缩短了信息路径, 增强了浅层的定位信息, 有效提升对虹膜特征的提取效果. 实验表明: 本文方法有效提高了虹膜图像的分割精度, 且有效改善了包含不同噪声因素的虹膜图像的分割. 相比于传统的 Mask R-CNN 网络, Mask-INet 网络的 *Dice* 值达到了 95.62%、*mIoU* 达到了 92.78%、*Recall* 达到了 96.36%、分别提升了 8.53%、11.97%、8.88%. 在下一步的工作中, 会探究如何捕获检测分支中不同大小的感受野, 提升生成 Mask 预测的质量, 从而进一步提高模型对虹膜困难样本分割的精确度.

参考文献

1 田会娟, 翟佳豪, 柳建新, 等. 基于 SRN-UNet 的低质量虹膜分割算法. 光子学报, 2022, 51(2): 0210006. [doi: 10.3788/gzxb20225102.0210006]

2 苑玮琦, 冯琪, 白晓光. 基于 2D-Gabor 滤波器的虹膜噪音检测方法. 光子学报, 2010, 39(2): 369–374. [doi: 10.3788/gzxb20103902.0369]

3 Zhou WB, Ma XT, Zhang Y. Research on image preprocessing algorithm and deep learning of iris recognition. Journal of Physics: Conference Series, 2020, 1621: 012008. [doi: 10.1088/1742-6596/1621/1/012008]

4 Cambier JL, Siedlarz JE. Portable authentication device and method using iris patterns: US, US6532298B1. 2003-03-11.

5 Sequeira AF, Chen LL, Ferryman J, *et al*. PROTECT Multimodal DB: Fusion evaluation on a novel multimodal biometrics dataset envisaging border control. Proceedings of the 2018 International Conference of the Biometrics Special Interest Group (BIOSIG). Darmstadt: IEEE, 2018. 1–5.

6 Wang CY, Muhammad J, Wang YL, *et al*. Towards complete and accurate iris segmentation using deep multi-task attention network for non-cooperative iris recognition. IEEE Transactions on Information Forensics and Security, 2020, 15: 2944–2959. [doi: 10.1109/TIFS.2020.2980791]

7 He ZF, Tan TN, Sun ZN, *et al*. Toward accurate and fast iris segmentation for iris biometrics. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(9): 1670–1684. [doi: 10.1109/TPAMI.2008.183]

8 Chen Y, Wang WY, Zeng Z, *et al*. An adaptive CNNs technology for robust iris segmentation. IEEE Access, 2019, 7: 64517–64532. [doi: 10.1109/ACCESS.2019.2917153]

9 Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015. 3431–3440.

10 Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. Proceedings of the 18th International Conference on Medical Image Computing and Computer-assisted Intervention. Munich: Springer, 2015. 234–241.

11 Badrinarayanan V, Kendall A, Cipolla R. SegNet: A deep convolutional encoder-decoder architecture for image

segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481–2495. [doi: 10.1109/TPAMI.2016.2644615]

12 Zhao HS, Shi JP, Qi XJ, *et al*. Pyramid scene parsing network. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 6230–6239. [doi: 10.1109/CVPR.2017.660]

13 Chen LC, Papandreou G, Kokkinos I, *et al*. Semantic image segmentation with deep convolutional nets and fully connected CRFs. Computer Science, 2014, (4): 357–361.

14 Chen LC, Papandreou G, Kokkinos I, *et al*. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 834–848. [doi: 10.1109/TPAMI.2017.2699184]

15 Chen LC, Papandreou G, Schroff F, *et al*. Rethinking atrous convolution for semantic image segmentation. arXiv:1706.05587, 2017.

16 Valanarasu JMJ, Oza P, Hacihaliloglu I, *et al*. Medical transformer: Gated axial-attention for medical image segmentation. Proceedings of the 24th International Conference on Medical Image Computing and Computer-assisted Intervention. Strasbourg: Springer, 2021. 36–46.

17 Jalilian E, Uhl A. Iris segmentation using fully convolutional encoder-decoder networks. Deep Learning for Biometrics. Cham: Springer, 2017. 133–155.

18 Lian S, Luo ZM, Zhong Z, *et al*. Attention guided U-Net for accurate iris segmentation. Journal of Visual Communication and Image Representation, 2018, 56: 296–304. [doi: 10.1016/j.jvcir.2018.10.001]

19 Wang CY, Wang YL, Zhang KB, *et al*. NIR iris challenge evaluation in non-cooperative environments: Segmentation and localization. 2021 IEEE International Joint Conference on Biometrics (IJCB). Shenzhen: IEEE, 2021. 1–10. [doi: 10.1109/IJCB52358.2021.9484336]

20 He KM, Gkioxari G, Dollár P, *et al*. Mask R-CNN. Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017. 2980–2988. [doi: 10.1109/ICCV.2017.322]

21 Ren SQ, He KM, Girshick R, *et al*. Faster R-CNN: Towards real-time object detection with region proposal networks. Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal: MIT Press, 2015. 91–99.

22 He KM, Zhang XY, Ren SQ, *et al*. Deep residual learning for image recognition. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 770–778. [doi: 10.1109/CVPR.2016.90]

23 Lin TY, Dollár P, Girshick R, *et al*. Feature pyramid networks for object detection. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 936–944. [doi: 10.1109/CVPR.2017.106]

24 Redmon J, Divvala S, Girshick R, *et al*. You only look once: Unified, real-time object detection. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 779–788. [doi: 10.1109/CVPR.2016.91]

25 王广学, 黄晓涛, 周智敏. SAR 图像尺度不变特征提取方法研究. 中国图象图形学报, 2011, 16(12): 2199–2205. [doi: 10.11834/jig.20111215]

26 Liu W, Anguelov D, Erhan D, *et al*. SSD: Single shot multibox detector. Proceedings of the 14th European Conference on Computer Vision. Amsterdam: Springer, 2016. 21–37.

27 Mnih V, Heess N, Graves A, *et al*. Recurrent models of visual attention. Proceedings of the 27th International Conference on Neural Information Processing Systems. Montreal: MIT Press, 2014. 2204–2212.

28 Hu J, Shen L, Albanie S, *et al*. Squeeze-and-excitation networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011–2023. [doi: 10.1109/TPAMI.2019.2913372]

29 Woo S, Park J, Lee JY, *et al*. CBAM: Convolutional block attention module. Proceedings of the 15th European Conference on Computer Vision. Munich: Springer, 2018. 3–19.

30 Deng J, Dong W, Socher R, *et al*. ImageNet: A large-scale hierarchical image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami: IEEE, 2009. 248–255. [doi: 10.1109/CVPR.2009.5206848]

31 Liu S, Qi L, Qin HF, *et al*. Path aggregation network for instance segmentation. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 8759–8768. [doi: 10.1109/CVPR.2018.00913]

32 Wang QL, Wu BG, Zhu PF, *et al*. ECA-Net: Efficient channel attention for deep convolutional neural networks. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2020. 11531–11539.

(校对责编: 牛欣悦)