



## 基于高分辨率特征的舌象分割算法研究

马龙祥<sup>1,2</sup>, 杨 浩<sup>1,2</sup>, 宋婷婷<sup>1</sup>, 翟鹏博<sup>1,2</sup>, 余 亢<sup>1</sup>

(1. 中国科学院微电子研究所, 北京 100029; 2. 中国科学院大学, 北京 100049)

**摘 要:** 舌象的精准分割对舌诊中舌体识别与分类具有重要意义, 采用传统图像处理方法和深度学习方法分割舌象会丢失部分舌象边缘信息, 从而降低舌体识别精确度。针对该问题, 提出一种利用高分辨率网络的舌象分割算法。使用区域定位网络识别舌体并提取舌象原图特征生成建议框, 对其进行分类和回归处理以定位舌象所在区域, 同时构建高分辨率网络提取该区域高分辨率特征, 最终完成舌象分割。实验结果表明, 该算法可有效保留舌象边缘信息, 其分割结果平均交并比达到 98.2%, 较 SegNet、Mask-RCNN 算法分割舌象更精准。

**关键词:** 舌诊; 舌象; 深度学习; 高分辨率特征; 实例分割

开放科学(资源服务)标志码(OSID):



中文引用格式: 马龙祥, 杨浩, 宋婷婷, 等. 基于高分辨率特征的舌象分割算法研究[J]. 计算机工程, 2020, 46(10): 248-252.

英文引用格式: MA Longxiang, YANG Hao, SONG Tingting, et al. Research on tongue image segmentation algorithm based on high resolution feature[J]. Computer Engineering, 2020, 46(10): 248-252.

## Research on Tongue Image Segmentation Algorithm Based on High Resolution Feature

MA Longxiang<sup>1,2</sup>, YANG Hao<sup>1,2</sup>, SONG Tingting<sup>1</sup>, ZHAI Pengbo<sup>1,2</sup>, YU Kang<sup>1</sup>

(1. Institute of Microelectronics of Chinese Academy of Sciences, Beijing 100029, China;

2. University of Chinese Academy of Sciences, Beijing 100049, China)

**[Abstract]** The accurate segmentation of tongue images is of great significance for tongue recognition and classification in tongue diagnosis. The traditional image processing and deep learning methods will lose part of the edge information of tongue images, thus reducing the accuracy of tongue recognition. To solve the problem, this paper proposes a tongue image segmentation algorithm based on high-resolution network. The region location network is used to identify the tongue and extract the features of the original image of the tongue to generate suggestion boxes, which are classified and processed with regression to locate the tongue area. At the same time, a high-resolution network is constructed to extract the high-resolution features of the region, and finally complete the tongue image segmentation. Experimental results show that the proposed algorithm can effectively preserve the edge information of tongue images, and the mean Intersection over Union(mIoU) of segmentation results is 98.2%, which is more accurate than that of SegNet and Mask-RCNN algorithms.

**[Key words]** tongue diagnosis; tongue image; deep learning; high resolution feature; instance segmentation

DOI: 10.19678/j.issn.1000-3428.0056685

### 0 概述

舌诊作为中医望诊的重要方面, 在中医现代化进程中起着关键性作用。由于舌的形态、大小以及湿润程度可反映出各部位器官的状态, 是人体健康状况的客观表征, 因此舌诊的标准化研究对于整个中医现代化体系建设具有重要意义<sup>[1]</sup>。在当前舌诊

数字化流程中, 拍摄到图像所包含脸颊、下巴、嘴唇等其他部位信息会影响舌象特征的提取, 进而对舌诊造成干扰。如何从图像中精准提取舌体区域是当前亟待解决的问题。

舌诊数字化图像处理方法主要包括传统图像处理方法和深度学习图像处理方法。其中, 最大类间方差法<sup>[2]</sup>、基于 HSV 空间阈值的方法<sup>[3]</sup>以及基于 Snakes

基金项目: 国家科技重大专项(2018ZX01031201)。

作者简介: 马龙祥(1995—), 男, 硕士研究生, 主研方向为生物医学图像处理; 杨 浩, 副教授、博士; 宋婷婷, 助理研究员、硕士; 翟鹏博, 硕士研究生; 余 亢, 助理研究员、硕士。

收稿日期: 2019-11-12

修回日期: 2019-12-24

E-mail: malongxiang@ime.ac.cn

能量最小化原理的方法<sup>[4]</sup>等传统图像处理方法根据像素之间差异进行区域划分来完成舌象分割, 对像素之间差异较小的部分无法进行有效分割。深度学习方法因其具有全自动处理、鲁棒性高以及速度快等特点, 在图像分割特别是舌象分割领域具有一定优势<sup>[5-6]</sup>。文献[7]使用 VGGNet<sup>[8]</sup> 作为基础架构实现了 Fsnets 深度学习算法在舌象语义分割方面的应用。文献[9]对比了全卷积网络(Fully Convolutional Networks, FCN)<sup>[10]</sup>、SegNet<sup>[11]</sup>、Mask-RCNN<sup>[12]</sup>、Deeplab<sup>[13]</sup>等深度学习算法在舌象分割方面的应用。用于图像分割的深度学习方法主要分两种, 一种是由 FCN 演变得到的 SegNet 和 Unet<sup>[14]</sup> 语义分割方法, 另一种是改进 R-CNN<sup>[15]</sup> 所得 Mask-RCNN 实例分割方法。FCN 是一种 encoder-decoder 网络结构, 在进行舌象区域分割时, 其数次下采样和上采样的结构设计会使舌象区域丢失部分边缘信息。而基于 Mask-RCNN 的实例分割方法是将舌体区域单独作为实例进行识别, 在处理舌象数据时, 分割出的舌体区域较实际舌体区域偏小, 导致出现一定的欠分割问题。

针对上述舌象边缘信息丢失的问题, 本文提出一种结合区域定位和高分辨率网络<sup>[16]</sup> 的舌象分割算法。使用残差网络(Residual Network, ResNet)<sup>[17]</sup> 提取舌象原图特征生成建议框, 对其分类以判断不同区域所属类别, 同时进行回归处理以调整检测框位置使其接近真实舌象区域, 并采用高分辨率网络提取该区域特征完成舌象分割。

## 1 舌象分割方法及原理分析

舌象分割全过程分为区域定位和舌象分割 2 个阶段。图 1 为区域定位阶段, 先检测图像中舌体区域, 去除对舌象分割造成干扰的嘴唇、下巴等区域, 再采用卷积结构(卷积层、ReLU 激活层和池化层)提取图像特征得到特征图, 从中选取区域建议框并进行分类回归处理后, 获取与原图对应的候选区域, 从而完成区域定位得到舌体区域检测框。图 2 为舌象分割阶段, 针对区域定位阶段得到的舌体检测区域搭建高分辨率网络, 以最大程度保留舌象的边缘和形状等特征。分辨率网络结构不同于传统语义分割的 encoder-decoder 网络结构, 是在原图的 4 个变换尺寸上进行网络训练, 从而实现舌体区域的精准分割。

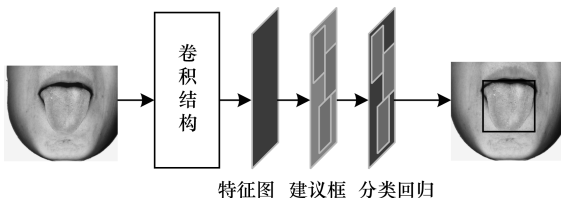


图 1 区域定位阶段

Fig. 1 Regional position stage

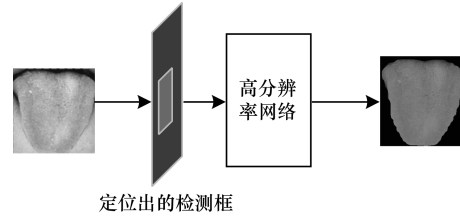


图 2 舌象分割阶段

Fig. 2 Tongue image segmentation stage

### 1.1 区域定位

由于采集的舌象数据通常较集中, 舌体区域在整个图像中的占比为 0.15 ~ 0.25, 除舌体区域以外其他部分均可视为背景区域, 因此如果直接分割原图会形成过多的背景区域, 对舌象特征提取造成干扰。采用区域定位技术可从图像中检测出舌体区域, 去除图像分割中的干扰, 从而实现图像精准分割。区域定位网络结构如图 3 所示。

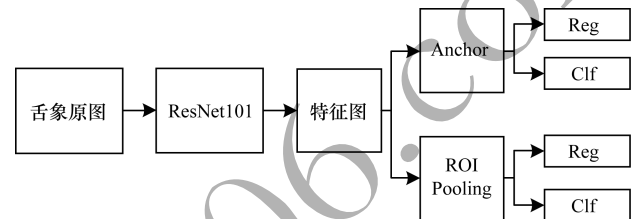


图 3 区域定位网络结构

Fig. 3 Regional position network structure

由于残差网络能提取丰富的图像特征, 其效果较 VGGNet 和 ZFNet<sup>[18]</sup> 更优, 因此本文选择 ResNet101 作为基础特征提取网络。具体实现步骤如下: 对于舌象原图, 先通过 ResNet101 提取图像特征, 在特征图中生成一定数量的建议区域, 从中选取含有正样本和负样本的区域建议框(该建议框和真实舌象区域重叠度大于 0.7 为正样本, 重叠度小于 0.3 为负样本, 正样本数量少于总数量的 1/2), 区域建议框图例如图 4 所示。



图 4 区域建议框图例

Fig. 4 Legend of regional suggestion box

然后利用锚点框 (Anchor) 确定候选框的位置, 再通过 ROI 池化 (ROI Pooling) 进行下采样, 最后对上述建议框进行分类 (Clf) 以判断不同区域所属类别并对其进行做回归 (Reg) 处理, 以调整检测框的位置使其更接近真实舌象区域。

分类和回归处理包括标准化处理、分类损失和回归损失计算。其中分类损失计算采用交叉熵损失<sup>[19]</sup>, 回归损失采用 Smooth\_L1 损失。在计算回归损失时, 只对正样本进行损失计算, 负样本不参与损失计算。分类损失和回归损失计算公式如下:

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{\text{cls}}} \sum_i L_{\text{cls}}(p_i, p_i^*) + \lambda \frac{1}{N_{\text{reg}}} \sum_i p_i L_{\text{reg}}(t_i, t_i^*) \quad (1)$$

其中,  $p_i$  表示候选框预测为目标的概率,  $N_{\text{cls}}$  表示每次选取图像的数量,  $N_{\text{reg}}$  表示归一化候选框的数量,  $t_i = \{x_i, y_i, w_i, h_i\}$  表示预测框左上角坐标以及预测框的宽和高,  $t_i^*$  表示真实舌体框坐标,  $L_{\text{cls}}$  表示分类损失,  $L_{\text{reg}}$  表示回归损失, GT 标签  $p_i^*$  的表达式为:

$$p_i^* = \begin{cases} 0, & \text{负样本} \\ 1, & \text{正样本} \end{cases} \quad (2)$$

对分类损失和回归损失分别进行优化, 再反向传播并更新网络参数, 最终完成网络训练。

## 1.2 舌象分割

针对区域定位出的舌象区域搭建高分辨率网络, 其结构如图 5 所示。使用该网络提取舌象区域高分辨率特征, 进而对舌象进行精准分割。

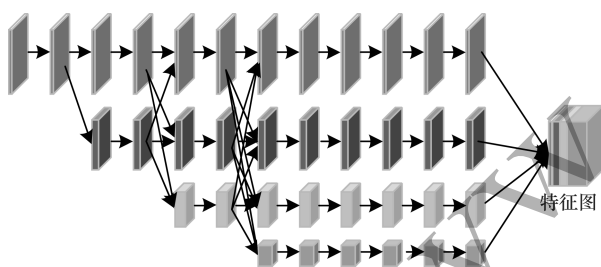


图 5 高分辨率网络结构

Fig. 5 High resolution network structure

高分辨率网络采用并行方式连接高、低分辨率卷积结果, 对不同分辨率的特征图通过多尺度融合<sup>[20]</sup>增强高分辨率表示, 从而保持舌象区域的高分辨率特征。此外, 舌象边缘信息对于病情诊断非常重要, 然而在使用 encoder-decoder 语义分割结构和 Mask-RCNN 实例分割结构的过程中, 均存在将特征图尺寸缩小再放大的步骤, 造成图像特征衰减, 进而导致舌象边缘信息丢失。本文设计的高分辨率特征网络基于原图特征而构建, 避免了上、下采样造成的

特征丢失问题<sup>[21]</sup>, 可实现舌象精准分割。高分辨率特征网络共有 4 个不同尺寸的通道, 由于最终输出的特征图需保持原图尺度, 因此分别对 3 个低分辨率通道进行双线性上采样, 再将所得特征图堆叠后得到最终的高分辨率特征图。

针对预测的特征图, 计算该特征图与真实标签之间的均方误差 (Mean Square Error, MSE), 优化均方误差后完成网络的权重更新。均方误差损失函数的表达式为:

$$\text{MSE} = \frac{\sum_{i=1}^n (y_i - z_i)^2}{n} \quad (3)$$

其中,  $z_i$  为预测值,  $y_i$  为真实值,  $n$  为样本数量。

## 2 舌象分割过程

舌象分割具体过程如下:

1) 搭建标准舌象采集环境, 以保证每张图像处于相同光强、亮度和拍摄距离的条件下, 从而最大程度避免外部条件对图像的干扰。

2) 构造数据集, 将数据分为 2 个部分, 分别作为训练集和测试集。使用 LabelMe 进行标注, 得到基于原始图像的真实标签 (uint8 单通道图像)。标签中每个像素点表示原始图像对应位置所属类别 (0 为背景, 1 为舌体区域)。

3) 使用标注好的训练集进行网络训练, 并使用测试集进行模型测试。

### 2.1 舌象标准采集环境搭建

在舌象采集过程中, 用户距离光源远近以及光线强弱均会造成所采集舌象的不同。舌体表面特征受光线影响较大 (如在光线较强情况下正常舌苔易发生表面反光而被识别为白苔), 且直射光照射舌面时会出现光线不均的现象。为解决该问题, 本文采集系统使用积分球光源。和传统光源相比, 积分球光源呈半封闭状, 其内部采用漫反射材料, 具有密闭空间和高显色指数, 可保证光照均匀, 同时球体直径限定了相机与用户之间的距离, 保证光照强度的一致性, 采集的图像舌面光线均匀, 无阴影和反光, 采集过程不受外界光源干扰。通过本文系统采集到的舌象图像避免了采集环境不稳定造成的干扰, 每张图像均在相同环境下获取, 保证了数据集的一致性, 有效减少数据集自身造成的系统误差。

### 2.2 舌象训练集和测试集划分与数据预处理

目前积分球光源系统使用工业相机进行舌象图像采集, 图像格式为 bmp, 像素为  $2\,300 \times 1\,944$ 。采集的舌象图像数据集共有 2 368 张舌象原图, 将

2 000 张和 368 张舌象图像分别作为训练集和测试集。使用 LabelMe 数据标注软件构建舌象分割标签,并采用数据增强方法<sup>[22]</sup>对图像进行预处理,以增加样本的数量和多样性,从而提升模型的鲁棒性。本文使用的数据增强方法主要包括:1) 翻转和旋转图像,以增加样本数量;2) 对图像进行颜色抖动,添加随机因子,并调整图像的饱和度、亮度以及对比度;3) 为图像添加高斯噪声,以增加样本多样性。通过上述数据增强方法,利用原始图像构造出更多不同场景下的舌象图像,为模型提供新的训练样本,并使模型学习到更具泛化性的特征,从而克服小数据集的局限性,提升分割算法的鲁棒性。

### 3 实验结果与分析

采用语义分割任务中常用的评估参数平均交并比(mean Intersection over Union, mIoU)作为本文训练模型分割效果的评价指标。mIoU 越大表明舌体预测区域与舌体真实区域越接近,模型分割效果越好。mIoU 的计算公式如下:

$$mIoU = \frac{1}{k+1} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij} + \sum_{j=0}^k (p_{ji} - p_{ii})} \times 100\% \quad (4)$$

其中, $i$ 表示真实值, $j$ 表示预测值, $p_{ij}$ 表示真实值为 $i$ 、预测值为 $j$ 的像素数量, $p_{ji}$ 表示真实值为 $j$ 、预测值为 $i$ 的像素数量, $p_{ii}$ 表示真实值和预测值都为 $i$ 的像素数量。

本文实验采用 encoder-decoder 网络结构的 SegNet 语义分割算法、Mask-RCNN 实例分割算法 2 种深度学习经典分割算法与本文基于区域定位和高分辨率网络的分割算法进行对比,以验证高分辨率网络对舌象分割性能的提升效果,不同算法在舌象图像测试集上的实验结果如表 1 所示。可以看出,和 SegNet 算法相比,Mask-RCNN 算法得到分割图像的 mIoU 更大,分割精度有一定幅度提升,但其所用预测时间较 SegNet 算法更长。本文算法所得分割图像的 mIoU 比 Mask-RCNN 算法更大,分割精度更高,但所用预测时间与 Mask-RCNN 算法相近。从分割精度与预测时间综合来看,本文算法的分割效果更好。

表 1 3 种算法在舌象测试集上的实验结果

Table 1 Experimental results of three algorithms on tongue image test set

算法	mIoU/%	预测时间/s
SegNet 算法	93.0	0.002
Mask-RCNN 算法	94.5	0.152
本文算法	98.2	0.164

图 6 为不同算法在舌象测试集上的分割效果。可以看出,SegNet 算法所得舌象边缘模糊且有毛刺,Mask-RCNN 算法预测的舌体区域偏小,上述算法均存在舌象边缘分割不精准的问题。而本文算法对舌象边缘分割精准,完整保留了舌体边缘信息。

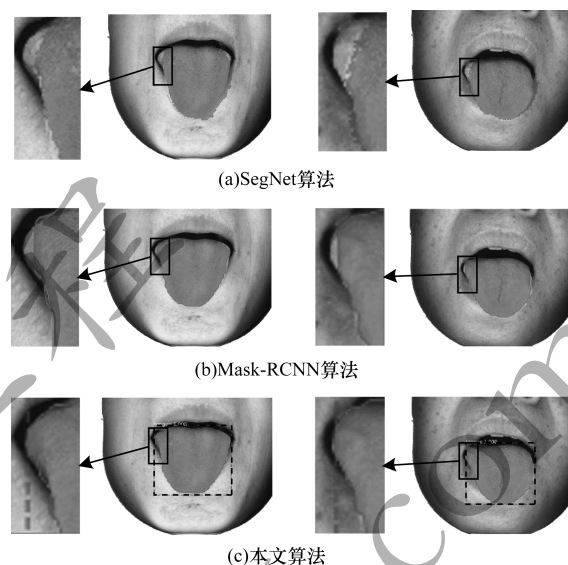


图 6 3 种算法在舌象测试集上的分割效果

Fig. 6 Segmentation effect of three algorithms on tongue image test set

### 4 结束语

针对舌象分割中舌体边缘信息丢失的问题,本文提出一种利用高分辨率特征的舌象分割算法。通过区域定位网络识别并获取舌体所在区域,构建高分辨率网络提取该区域舌体高分辨率特征,从而完成舌体区域分割。实验结果表明,该算法可有效保留舌象边缘信息,较 SegNet、Mask-RCNN 算法分割舌象更精准。后续考虑去除本文算法中冗余的模型参数以减少运算资源占用,同时对舌象识别算法进行研究,将齿痕舌、裂纹舌以及红点舌等不同舌体形态纳入该算法,以实现中医舌诊的智能化诊断。

### 参考文献

- [1] HUANG Shuqiong, ZHANG Yunlong, ZHOU Jing, et al. Research progress on the objectification, quantitation and standardization of tongue manifestation in traditional Chinese medicine[J]. China Journal of Traditional Chinese Medicine and Pharmacy, 2017, 32(4): 1625-1627. (in Chinese)  
黄淑琼,张云龙,周静,等. 浅谈中医舌象客观化、定量化、标准化研究[J]. 中华中医药杂志, 2017, 32(4): 1625-1627.
- [2] YI Sanli, ZHANG Guifang, HE Jianfeng, et al. Maximum entropy image segmentation based on maximum interclass variance[J]. Computer Engineering and Science, 2018, 40(10): 1874-1881. (in Chinese)

- 易三莉,张桂芳,贺建峰,等. 基于最大类间方差的最大熵图像分割[J]. 计算机工程与科学,2018,40(10): 1874-1881.
- [3] ZHENG Feng, HUANG Xiaoyang, WANG Boliang, et al. A method for tongue detection based on image segmentation[J]. Journal of Xiamen University(Natural Science Edition),2016, 55(6):895-900. (in Chinese)
- 郑丰,黄晓阳,王博亮,等. 基于图像分割的舌象检测方法[J]. 厦门大学学报(自然科学版),2016,55(6): 895-900.
- [4] HU Xuegang, QIU Xiulan. Novel image segmentation algorithm based on Snake model[J]. Journal of Computer Applications,2017,37(12):3523-3527. (in Chinese)
- 胡学刚,邱秀兰. 基于 Snake 模型的图像分割新算法[J]. 计算机应用,2017,37(12):3523-3527.
- [5] ESTEVA A, KUPREL B, NOVOA R A, et al. Dermatologist-level classification of skin cancer with deep neural networks [J]. Nature, 2017, 542 (7639): 115-118.
- [6] LIU Fei, ZHANG Junran, YANG Hao. Research progress of medical image recognition based on deep learning [J]. Chinese Journal of Biomedical Engineering, 2018, 37(1): 86-94. (in Chinese)
- 刘飞,张俊然,杨豪. 基于深度学习的医学图像识别研究进展[J]. 中国生物医学工程学报,2018,37(1):86-94.
- [7] WANG Liran, TANG Yiping, CHEN Peng, et al. Two-phase convolutional neural network design for tongue segmentation[J]. Journal of Image and Graphics,2018, 23(10):1571-1581. (in Chinese)
- 王丽冉,汤一平,陈朋,等. 面向舌体分割的两阶段卷积神经网络设计[J]. 中国图象图形学报,2018,23(10): 1571-1581.
- [8] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [EB/OL]. [2019-11-10]. <https://arxiv.org/abs/1409.1556>.
- [9] LU Yunxi, LI Xiaoguang, ZHANG Hui, et al. Review on tongue image segmentation technologies for traditional Chinese medicine: methodologies, performances and prospects [EB/OL]. [2019-11-10]. <https://doi.org/10.16383/j.aas.c180807>.
- [10] SHELHAMER E, LONG J, DARRELL T. Fully convolutional networks for semantic segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2017,39(4): 640-651.
- [11] 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.
- [12] HE K M, GKIOXARI G, DOLLAR P, et al. Mask R-CNN[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2020,42(2):386-397.
- [13] CHEN L C, PAPANIREOU 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,2018,40(4):834-848.
- [14] RONNEBERGER O, FISCHER P, BROX T. U-net: convolutional networks for biomedical image segmentation [EB/OL]. [2019-11-10]. <https://lmb.informatik.uni-freiburg.de/people/ronneber/u-net/>.
- [15] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA:IEEE Press,2014:580-587.
- [16] SUN Ke, XIAO Bin, LIU Dong, et al. Deep high-resolution representation learning for human pose estimation [C]//Proceedings of 2019 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA:IEEE Press,2019:5693-5703.
- [17] ZHANG Ming, LÜ Xiaoqi, WU Liang, et al. Multiplicative denoising method based on deep residual learning [J]. Laser and Optoelectronics Progress,2018, 55(3):197-203. (in Chinese)
- 张明,吕晓琪,吴凉,等. 基于深度残差学习的乘性噪声去噪方法[J]. 激光与光电子学进展,2018,55(3): 197-203.
- [18] ZEILER M D, FERGUS R. Visualizing and understanding convolutional networks [C]//Proceedings of ECCV '14. Berlin, Germany:Springer,2014:818-833.
- [19] DENG Zelu, GAO Jianbin, HUANG Tao, et al. Combining cross entropy loss with manually defined hard example for semantic image segmentation [C]// Proceedings of 2019 International Conference on Image and Graphics. Berlin, Germany:Springer,2019:32-43.
- [20] SUN Genyun, HUANG Hui, ZHANG Aizhu, et al. Fusion of multiscale convolutional neural networks for building extraction in very high-resolution images [J]. Remote Sensing,2019,11(3):227-228.
- [21] JAVED K, HUSSAIN G, SHAUKAT F, et al. A neural network approach to remove rain using reconstruction and feature losses [J]. Neural Computing and Applications, 2019,32(10):13129-13138.
- [22] LEE M B, KIM Y H, PARK K R. Conditional generative adversarial network-based data augmentation for enhancement of iris recognition accuracy [J]. IEEE Access,2019,7(8): 122134-122152.