

DOI:10.19651/j.cnki.emt.2211268

# 改进 DeepLabV3+ 网络的露天矿挡墙分割方法<sup>\*</sup>

马旭 杨立东 郭勇 赵艳锋

(内蒙古科技大学信息工程学院 包头 014000)

**摘要:** 为抑制露天矿场复杂环境下的背景干扰,对安全挡墙进行精确的语义分割,提出了一种基于改进 DeepLabV3+ 网络的露天矿安全挡墙分割方法。首先,主干网络采用轻量级的 MobileNetV2 网络,通过深度可分离卷积和倒残差结构有效降低了网络参数数量和计算量。然后,加入混合注意力模块进行通道及空间上的特征强化,可避免边缘信息丢失。最后,利用数据增强和迁移学习解决目标数据集较少的问题,提高模型的泛化能力。实验结果证明,该方法有较好的分割效果,MIOU 和 MPA 分别为 85.06%、92.94%,均优于原网络和其他经典网络模型。该方法能够精确地分割露天矿安全挡墙,有较好的实际应用价值。

**关键词:** 语义分割; MobileNetV2; 混合注意力

**中图分类号:** TP391 **文献标识码:** A **国家标准学科分类代码:** 510.4050

## Semantic segmentation method of safety retaining wall in open pit mine with improved DeepLabV3+ network

Ma Xu Yang Lidong Guo Yong Zhao Yanfeng

(School of Information Engineering, Inner Mongolia University of Science and Technology, Baotou 014000, China)

**Abstract:** In order to suppress the background interference in the complex environment of the open-pit mine and perform accurate semantic segmentation of the safety retaining wall, a segmentation method of the safety retaining wall of the open-pit mine based on the improved DeepLabV3+ network is proposed. First, the backbone network adopts the lightweight MobileNetV2 network, which effectively reduces the amount of network parameters and computation through the depthwise separable convolution and inverted residual structure. Then, a hybrid attention module is added for channel and spatial feature enhancement, which can avoid the loss of edge information. Finally, data augmentation and transfer learning are used to solve problems with fewer target datasets and improve the generalization ability of the model. The experimental results show that the method has a good segmentation effect, with MIOU and MPA of 85.06% and 92.94%, respectively, which are better than the original network and other classic network models.

**Keywords:** semantic segmentation; MobileNetV2; mixed attention

### 0 引言

露天矿为防止山体滑坡、填土失稳等灾害需要做安全挡墙,露天矿安全挡墙的安全状况不仅影响矿山的正常生产,也会影响到相关工作人员的生命安全,研发并实施安全挡墙自动识别报警平台不仅可以保障人身、生产安全,提高作业率,同时也能为企业带来可观的经济效益和社会效益<sup>[1]</sup>。由于安全挡墙在车辆排土过程中会受到损坏,可能导致车辆掉落到排土台阶下,若发现挡墙出现损坏情况,应及时通知相关人员进行修复,挡墙修复后,车辆才能继续进行排土作业。因此,进行露天矿安全挡墙的智能化监测是

基于当前发展情况必不可少的一项建设内容<sup>[2]</sup>,也是保障现场人员生命安全最有效的方式。

安全挡墙的智能化监测最关键的部分是图像的精确分割,图像分割是计算机视觉的重要组成部分,在场景应用中发挥着重要作用。常用的图像分割方法有区域生长法、边缘检测法、分水岭分割法和小波变换法等<sup>[3]</sup>。但是这些传统算法均存有特征提取差、分割准确度低、难以确定阈值、实时性较差等缺点。2012年以后,深度卷积神经网络正逐渐推动图像处理领域的发展,并在图像分类、目标检测、语义分割等方面取得了很大进步<sup>[4]</sup>。2015年,全卷积网络

收稿日期:2022-09-05

<sup>\*</sup> 基金项目:国家自然科学基金(62161040)、内蒙古自然科学基金(2021MS06030)、内蒙古科技计划项目(2021GG0023)、内蒙古自治区高等学校青年科技英才支持计划(NJYT22056)项目资助

FCN 改变了原来的网络结构,FCN 是最早提出的语义分割网络<sup>[5]</sup>,之后多数语义分割网络都继承 FCN 的基本框架。针对医学图像领域提出 U-Net 网络<sup>[6]</sup>,解决了医学影像数据量较少的问题。为解决外观相似物体的分割问题,提出 PSPNet 模型<sup>[7]</sup>,PSPNet 创新性的提出空间池化金字塔结构,提升了网络利用全局上下文信息的能力。随后,Google 提出的 DeepLab 系列网络使语义分割达到了新高峰。DeepLabV3+网络模型逐渐成为主流分割算法<sup>[8]</sup>,但因为它的结构复杂、参数过多,导致计算量过大,这在分割种类少的任务中不太适用。对此,Tian 等<sup>[9]</sup>提出基于 DeepLabV3+的 DUpsampling 算法,该算法把原网络解码器中的双线性插值法替换为一种数据相关型的上采样方法,可在恢复特征图时降低计算复杂度。Zhang 等<sup>[10]</sup>提出了改进 DeepLabV3+架构的金字塔注意力网络(PAN),该网络将特征金字塔注意力网络(FPA)替代编码器中的空洞空间卷积池化金字塔(ASPP)结构,进行特征提取,提升模型分割精度。以上研究虽然一定程度上优化了网络模型,

但对于复杂场景下的目标仍存在分割不连续以及边缘信息丢失的缺陷,因此本文基于 DeeplabV3+模型,提出一种基于改进 DeepLabV3+的露天矿挡墙分割算法。本文通过使用 MobileNetV2 替代 ResNet 网络<sup>[11]</sup>,使参数量减小,使主干网络轻量化。同时加入混合注意力机制,可以提高网络分割精确度,获得更准确的挡墙特征信息,在露天矿安全挡墙的分割任务中得到更好的效果。

## 1 网络结构

### 1.1 DeepLabV3+网络结构

DeepLabV3+是 Google 提出的深度卷积神经网络模型,此算法在 DeepLabV1-3 的基础上改进而来,在 Pascal Voc、CitySpace 等公共的数据集中都获得了很好的效果。DeepLabV3+网络由编码器和解码器组成,其结构如图 1 所示,编码器分为特征提取网络 ResNet 和空洞空间金字塔池化结构 ASPP<sup>[12]</sup>,解码器通过融合低层特征并进行特征图恢复。

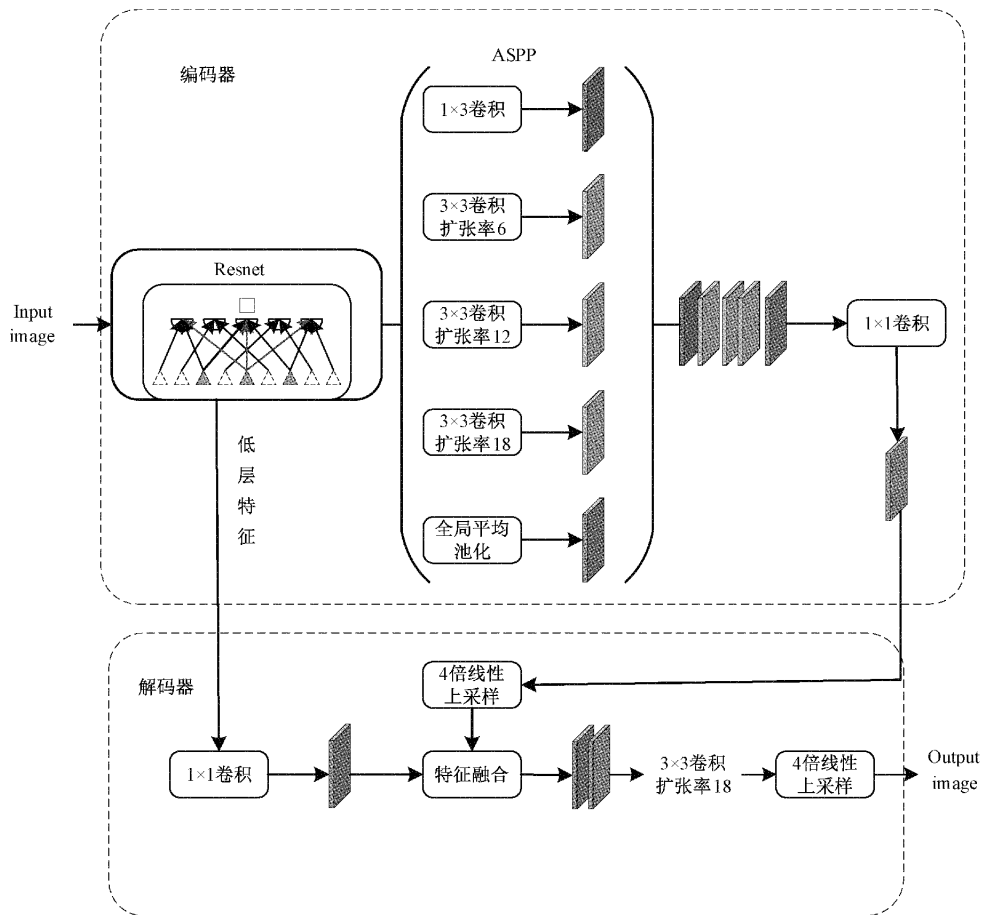


图 1 DeepLabV3+网络

首先,神经网络使用卷积层对图像进行特征提取,通过下采样操作减少计算量,但多次下采样会导致丢失部分目标边缘信息,这对语义分割任务不利。DeepLabV3+的

主干网络加入了扩张卷积,可以在增大感受野的同时减少下采样和参数量,使提取的特征不丢失相关边缘信息。其次,图像中种类不同则分割尺度也不相同,对多种类使用

同层特征进行分割不能确保分割效果。DeepLabV3+ 借鉴 PSPNet 网络并改进成空洞空间金字塔池化结构 (ASPP), 可以分割多尺度目标。ASPP 对特征图分别进行  $1 \times 1$  和  $3 \times 3$  的卷积及全局平均池化操作, 扩张率为 6、12、18, 然后进行特征融合并压缩通道, ASPP 能很好的完成多尺度的特征提取。最后, DeepLabV3+ 为获得目标的高层语义信息对输入图像进行下采样操作, 为弥补操作中丢失的边界信息采用编码-解码结构恢复目标边界信息。

### 1.2 MobileNetV2 主干网络结构

深度学习有大型网络和轻量级网络两个发展趋势<sup>[13]</sup>, 但是大型网络的计算量和参数量十分庞大, 一般在较高算力的 GPU 上运行, 所以神经网络开始进行轻量级研究。MobileNetV2 是轻量网络。MobileNetV2 通过倒残差结构和深度可分离卷积实现轻量化。深度可分离卷积和普通卷积不同, 分为深度卷积和逐点卷积。深度卷积在卷积时, 一个通道只由一个卷积核进行卷积, 卷积核数等于通道数。逐点卷积和普通卷积类似, 只是卷积核的尺寸设为  $1 \times 1$ 。MobileNetV2 在残差网络 ResNet 的基础上, 创新性的提出了倒残差结构<sup>[14]</sup>, 结构如图 2 所示。此结构首先通过逐点卷积升维, 再通过深度卷积提取通道特征, 最后通过逐点卷积降维。但只有步长为 1 时才有残差连接, 步长为 2 时使用串联结构。倒残差使通道维度先升后降, 这允许网络有更小的输入和输出, 降低了网络参数量和计算量, 同时通过残差连接能够使网络层更深, 提高梯度传播效率。

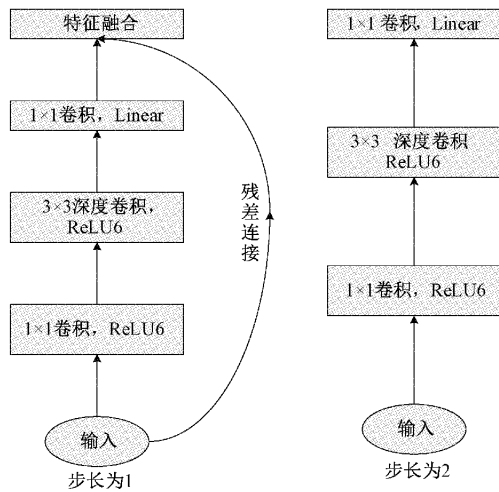


图 2 倒残差结构

### 1.3 混合注意力模块

注意力机制在深度学习的许多领域都有很好的效果。而混合注意力模块 CBAM 是一个通用且优秀的模块, 混合注意力模块可以关注图中重要特征并抑制无用特征<sup>[15]</sup>, 它将注意力过程分为通道注意力部分和空间注意力部分, 其结构如图 3 所示。通道注意力重点关注那些特征有意义, 使用最大池化和全局平均池化的方式来重视不同的特征,

核心思想就是增大有效通道权重, 减少无效通道权重。空间注意力模块关注哪里的特征是有意义的, 给出关键特征的位置, 使得到的特征图更显著。

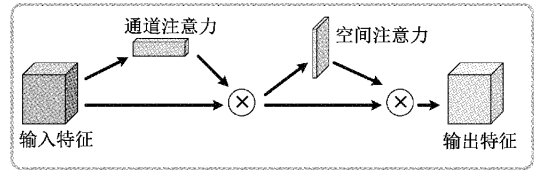


图 3 混合注意力模块

### 1.4 本文网络结构

排土场安全挡墙在图像上大多数是长条形状, 语义较简单, 但是有丰富边缘信息, 这需要网络有很高的细节提取能力<sup>[16]</sup>。DeepLabV3+ 网络在类别复杂的数据集上可以提取多样特征, 有较好效果。主干特征提取网络 ResNet 在提取多类别目标的任务有一定优势, 但 ResNet 结构复杂和庞大的计算量导致其不适合露天矿安全挡墙的语义分割。因此, 针对图像中的挡墙特征, 本文将 DeepLabV3+ 网络进行改进, 结构如图 4 所示, 模型首先用轻量级的 MobileNetV2 网络进行特征提取, 相比于 ResNet 网络, 具有计算量小和参数量较少、模型容易训练等优点; 再通过混合注意力进行特征图的通道和空间特征强化<sup>[17]</sup>, 然后通过空洞空间金字塔池化操作增大感受野并得到多尺度特征信息; 然后按照原网络方法进行特征图的解码恢复工作, 本方法整体上较原网络有更好的特征提取能力。

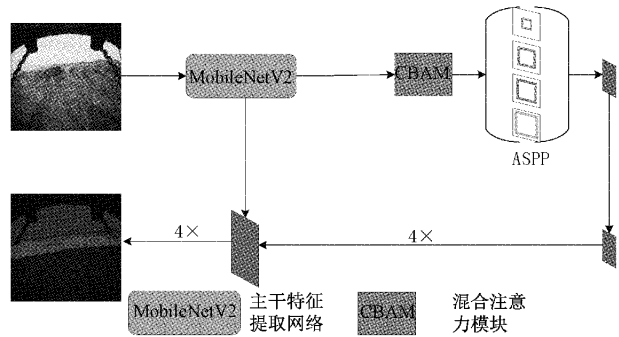


图 4 改进的 DeepLabV3+ 网络

## 2 实验数据集

实验所用安全挡墙图像来源于露天矿安全生产管理平台, 在平台上寻找含有排土场安全挡墙的视频回放, 通过软件处理成图像, 从中筛选出 600 张清晰图片, 裁剪图片大小为  $512 \times 512$ , 通过开源图像标注工具 Labelme 对图像进行人工标注, 将挡墙与背景进行分离, 得到 coco 数据集格式的 json 文件, 经过程序处理可以转成 png 格式。由于实验的数据量不足, 直接训练原始数据容易导致出现过拟合问题, 为了增加数据集的多样性, 提高模型的泛化能力, 对目标数据集进行数据增强处理增加样本容量<sup>[18]</sup>。数

据增强通过翻转、对比度增强、亮度增强和色彩增强等方式扩容到 3 000 张图像,数据增强如图 5 所示。最后通过 8:1:1 的比例随机划分目标数据集为训练集、测试集和验证集。

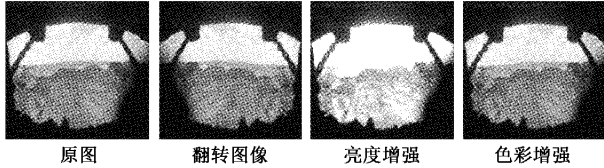


图 5 数据增强效果

### 3 实验与分析

#### 3.1 环境配置

实验所用的计算机硬件配置:CPU 为 intel Core i7-10700K,显卡设为 GTX1080Ti,内存 RAM 为 16 G。深度学习框架通过 Python3.6、Pytorch1.2 搭建,结合 Numpy、OpenCV、PIL 等库进行开发。网络超参数设置:Batch Size 设为 16,学习率设为  $7 \times 10^{-3}$ ,若在 5 轮训练中损失函数没有下降,学习率下降为原来的 0.1 倍;使用 Adam 优化器,可以同时结合一阶和二阶动量,和 Sgd、AdaGrad 等优化器相比,能够动态的调整每个参数的学习率,使网络收敛更快,动量 Momentum 为 0.9,Weight Decay 设为  $1 \times 10^{-4}$ 。Loss 采用语义分割常用的交叉熵损失函数,公式为:

$$CE = - \sum_{i=1}^k y_i \lg P_i \quad (1)$$

式中: $k$  为类别数量; $y$  为标签,若类别为  $i$ ,则  $y_i = 1$ ,否则等于 0; $P$  为神经网络输出,即类别为  $i$  的概率。预测结果与标签越接近,损失值越小。

#### 3.2 客观评价指标

评价指标分别是平均像素准确率(MPA)和平均交并比(MIOU)。MPA 表示正确识别的像素个数与图像的像素总量的比值,MIOU 是计算真实集合与预测集合的交集与并集的比值。计算公式如下:

$$MPA = \frac{1}{N} \left( \sum_{i=1}^N \frac{X_{ii}}{T_i} \right) \quad (2)$$

式中: $N$  为包括背景在内的类别总数; $X_{ii}$  代表识别为  $i$  类,实际为  $i$  类目标的像素; $T_i$  为  $i$  类目标像素的总数量。

$$MIOU = \frac{1}{N} \left( \sum_{i=1}^N \frac{X_{ii}}{\sum_{j=1}^N X_{ij} + \sum_{j=1}^N X_{ji} - X_{ii}} \right) \quad (3)$$

式中: $X_{ij}$  代表识别为  $j$  类,实际为  $i$  类目标的像素, $X_{ji}$  代表识别为  $i$  类,实际为  $j$  类目标的像素。

#### 3.3 迁移学习与模型训练

迁移学习能够解决目标数据集少,分割效果差的问题<sup>[19]</sup>。迁移学习分为两个阶段,第 1 阶段是网络模型在

Pascal Voc 数据集上初始化模型参数。此时,通过冻结空洞空间金字塔池化模块和解码器结构的批量归一层,进行特征迁移期间批量归一化层不更新,能够减少模型错误率并保证迁移效果,获得预训练权重。第 2 阶段是在目标数据集上训练,由于主干网络 MobileNetV2 提取的特征在整个模型中是通用的,因此本文将主干网络在训练初期冻结起来,以加快模型训练速度,防止权重在训练过程中被破坏。在训练后期再将主干网络解冻,使其参与整个模型的训练。将冻结训练的 Epoch 设置为 100,总的训练 Epoch 设置为 300;在此基础上,为防止模型过拟合,在训练过程中,设置了 EarlyStopping 监控验证集在训练过程中的 loss 值<sup>[20]</sup>,当验证集在训练了 20 个 Epoch 后其 loss 值仍没有下降,则提前停止整个模型的训练。

#### 3.4 实验结果分析

将所有网络模型在实验环境和相关参数一致的情况下进行对比试验,实验后的部分效果图如图 6 所示,分别为原图、标签图、本文方法测试结果、DeepLabV3+测试结果、U-Net 测试结果、HRNetV2 测试结果、PSPNet 测试结果。

由图 6 可以看出,PSPNet 的预测效果最差,大部分场景不能完成挡墙分割任务,而 HRNetV2 和 U-Net 因为受到车辆或远处矿山等背景影响会出现错误分割的问题。通过对比效果图,可以看出 DeepLabV3+原网络在场景复杂时会出现边缘分割不精确的问题,这是因为主干网络在特征提取过程中丢失边缘特征信息,从而导致分割效果不理想。而本方法使用混合注意力模块充分利用输入特征的所有特征信息,加强了上下文语义联系,所以效果图更接近标签图,在复杂场景下也能完成安全挡墙的精确分割。

由表 1 可知,改进后的 DeepLabV3+模型的实验结果明显优于原网络和其他网络模型,其 MIOU 对比 DeepLabV3+、U-Net、HRNetV2 和 PSPNet 模型,分别提升了 2.96%、3.93%、4.03%和 9.95%;且其平均像素精度(MPA)为 92.94%,比 DeepLabV3+、U-Net、HRNetV2 和 PSPNet 模型分别提升了 1.78%、4.08%、4.69%和 9.97%;显然,本实验方法分割的精确度更高,优于原网络和其他对比网络。

表 1 实验结果对比

网络模型	MIOU/%	MPA/%
本文方法	<b>85.06</b>	<b>92.94</b>
DeepLabV3+	82.10	91.16
U-Net	81.13	88.86
HRNetV2	81.03	88.25
PSPNet	75.11	82.97



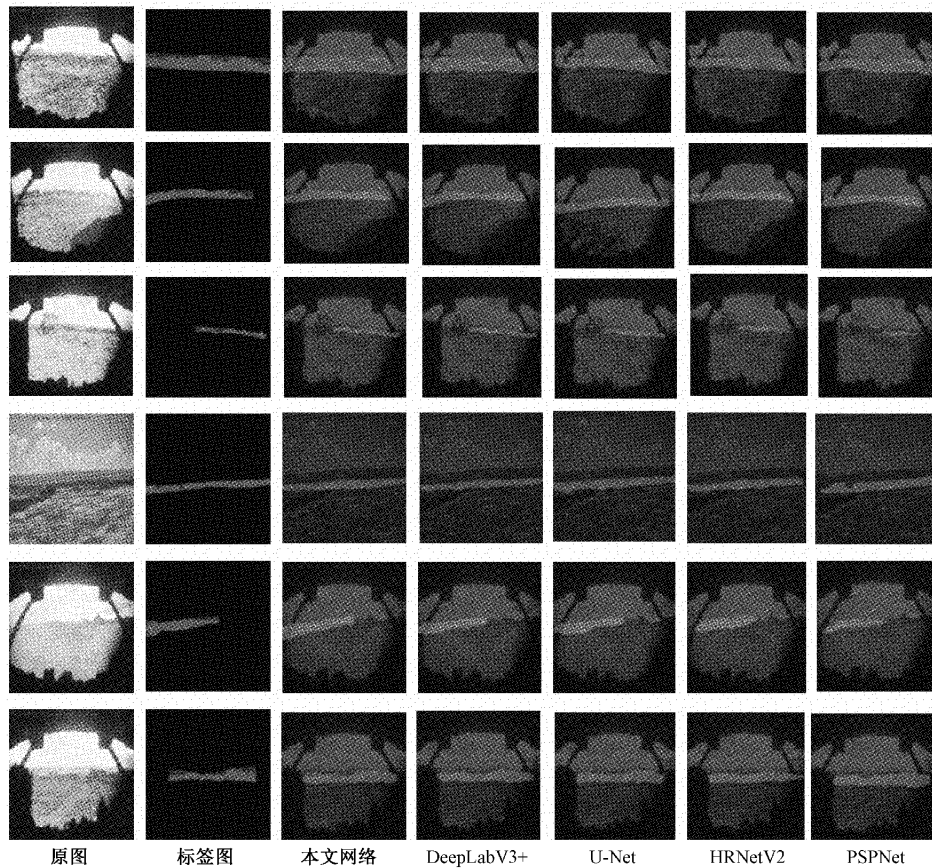


图 6 效果图对比

#### 4 结 论

传统方法不能实现对于露天矿安全挡墙的图像分割,因此本文提出了一种改进的 DeepLabV3+ 网络,该方法将骨干网络 ResNet 替换为轻量级的 MobileNetV2,同时加入混合注意力模块(CBAM),可以使网络参数量变少,细节提取增强。对于数据集较少的情况,采用数据增强和迁移学习的方法进行解决。由实验可知,本文方法能够在目标数据集上得到较好效果,改善了原网络边缘分割不精确的情况。后续会继续优化网络,在使网络更轻量化的同时做到预测更精准,实时性更强,从而更好的完成露天矿安全挡墙分割任务以及实现挡墙缺口自动识别系统。

#### 参考文献

- [1] 刘磊,郭二民,李忠华,等.加强“十四五”露天煤矿开采环境管理的建议[J].中国煤炭,2021,47(10):61-66.
- [2] 胡乃勋,陈涛,甄娜,等.基于卷积神经网络的面向对象露天采场提取[J].遥感技术与应用,2021,36(2):265-274.
- [3] 田萱,王亮,丁琪.基于深度学习的图像语义分割方法综述[J].软件学报,2019,30(2):440-468.
- [4] GUO Y, LIU Y, OERLEMANS A, et al. Deep learning for visual understanding: A review [J].

Neurocomputing, 2016, 187: 27-48.

- [5] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 3431-3440.
- [6] 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, Springer, Cham, 2015: 234-241.
- [7] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing network[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2881-2890.
- [8] CHEN L C, ZHU Y, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 801-818.
- [9] TIAN Z, HE T, SHEN C, et al. Decoders matter for semantic segmentation: Data-dependent decoding enables flexible feature aggregation[C]. Proceedings

- of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019; 3126-3135.
- [10] ZHANG Y, SUN X, DONG J, et al. GPNet: Gated pyramid network for semantic segmentation [J]. Pattern Recognition, 2021, 115: 107940.
- [11] SANDLER M, HOWARD A, ZHU M, et al. MobileNetV2: Inverted residuals and linear bottlenecks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018; 4510-4520.
- [12] 袁立,袁吉收,张德政.基于 DeepLab-v3+的遥感影像分类[J].激光与光电子学进展,2019,56(15):236-243.
- [13] 白欢欢.一种融合多注意力机制的服饰分割方法研究[J].国外电子测量技术,2021,40(7):52-58.
- [14] 袁建野,南新元,蔡鑫,等.基于轻量级残差网络的垃圾图片分类方法[J].环境工程,2021,39(2):110-115.
- [15] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module[C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018; 3-19.
- [16] 彭道刚,刘薇薇,威尔江,等.基于 CBAM-Res\_UNet 电厂高压蒸汽泄漏检测研究[J].电子测量与仪器学报,2021,35(12):206-214.
- [17] 刘艳菊,王秋霖,赵开峰,等.基于卷积神经网络的热轧钢条表面实时缺陷检测[J].仪器仪表学报,2021,42(12):211-219.
- [18] 石晨宇,周春,靳鸿,等.基于卷积神经网络的农作物病害识别研究[J].国外电子测量技术,2021,40(9):93-99.
- [19] 武凌,王浩,张晓春,等.基于深度迁移学习的垃圾分类系统设计与实现[J].沈阳大学学报(自然科学版),2020,32(6):496-502.
- [20] SINGH S, KRISHNAN S. Filter response normalization layer: Eliminating batch dependence in the training of deep neural networks[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020; 11237-11246.

### 作者简介

马旭,硕士研究生,主要研究方向为图像处理。

E-mail: mm27683486@163.com

杨立东,博士,教授,硕士生导师,主要研究方向为音频信号处理与模式识别。

E-mail: yld\_nkd@imust.edu.cn

郭勇,博士,硕士生导师,主要研究方向为信号与图像处理中的数学方法。

E-mail: guo\_yong@imust.edu.cn

赵艳锋,硕士,硕士生导师,主要研究方向为计算机网络。

E-mail: zyf@imust.edu.cn