

文章编号: 2095-2163(2022)01-0089-06

中图分类号: TP391.4

文献标志码: A

# 基于改进 Deeplabv3+的雾气图像分割

冯传盟<sup>1</sup>, 邢彦锋<sup>1</sup>, 李学星<sup>2</sup>, 蒋世谊<sup>1</sup>

(1 上海工程技术大学 机械与汽车工程学院, 上海 201620; 2 无锡职业技术学院 机械技术学院, 江苏 无锡 214121)

**摘要:** 针对传统雾气图像分割算法耗时长、分割结果存在凹陷等问题,提出一种基于 DeepLabv3+的雾气图像分割算法。改进算法将 DeepLabv3+原结构编码器的 Backbone 替换为更轻量的 Mobilenetv2 网络;将解码器的特征融合结构进行重新设计,同时加入注意力通道模块、边缘细化模块,通过消融实验得到分割效果最佳的雾气图像分割网络。实验结果表明,改进算法对雾气图像的分割精度 *MIOU* 可达 90.31%, 优于以 ResNet101 为基础的 DeepLabv3+算法,且分割速度提高了 57.26%,模型容量减少了 92.62%。

**关键词:** 雾气图像; DeepLabv3+; Mobilenetv2; 注意力机制; 边界细化

## Fog image segmentation based on improved DeepLabv3 Plus

FENG Chuanmeng<sup>1</sup>, XING Yanfeng<sup>1</sup>, LI Xuexing<sup>2</sup>, JIANG Shiyi<sup>1</sup>

(1 School of Mechanical and Automotive Engineering, Shanghai University of Engineering Science, Shanghai 201620, China;

2 School of Mechanical Technology, Wuxi institute of technology, Wuxi Jiangsu 214121, China)

**[Abstract]** In view of the time-consuming and denting of the traditional fog image segmentation algorithm, a fog image segmentation algorithm based on DeepLabv3 plus is proposed. The improved algorithm replaces the Backbone of DeepLabv3 plus original structure encoder with a lighter Mobilenetv2 network, redesigns the feature fusion structure of the decoder, introduces the attention channel module, edge refinement module, and obtains the fog image segmentation network with the best segmentation effect through ablation experiment. Experimental results show that the *MIOU* of the improved algorithm can be 90.31%, which is better than the DeepLabv3 plus algorithm based on ResNet101, and the segmentation speed is improved by 57.26 percent and the model capacity is reduced by 92.62 percent.

**[Key words]** fog image; Deeplabv3+; Mobilenetv2; attention mechanism; boundary refinement

## 0 引言

在现实生活中,有很多场景都需要使用高压喷雾,例如烟气调质<sup>[1]</sup>、汽车彩绘、焚烧炉降温等。其中,雾气几何形貌对其性能有着至关重要的影响,而精确将雾气与背景分割开来,则是准确测量雾气几何形貌的前提。传统分割算法分割速度较慢,分割结果存在缺陷。近年来,深度卷积神经网络由于其强大的特征提取能力,从而在分割任务上有着独特的表现。

语义分割是在像素级别理解图像,为图像中的每个像素分配一个类别<sup>[2]</sup>。目前,在语义分割上应用深度卷积神经网络主要面临分辨率下降及多尺度目标的挑战。

对于分辨率下降问题:在特征提取过程中,连续的池化或者步长为 2 的卷积操作,会导致输入特征张量的分辨率持续降低,从而损失原始图像信息,且

在上采样过程中难以完全恢复。Chen 等人<sup>[3]</sup>提出的 DeepLabv1 网络,将空洞卷积应用在语义分割任务中。空洞卷积的引入,能够在不增加额外学习参数的情况下,控制输出特征张量分辨率的同时,调整感受野的大小,从而有效解决这一挑战,这一技巧也被 PSPNet 网络<sup>[4]</sup>所采用。

对于多尺度目标的问题:在同一幅图像中,目标之间呈现出来的大小可能有差异,语义分割任务要求对不同大小的目标都要做到准确分割。Farabet 等人<sup>[5]</sup>通过将不同尺度的图像输入同一网络,得到不同大小的特征张量,将这些多尺度信息再做融合,解决多尺度目标。Jonathan Long 等人<sup>[6]</sup>提出的语义分割开山之作—FCN 网络,定义了一种跳跃连接结构,使得语义信息可以和表征信息相结合,用以精细分割结果。Ronneberger 等人<sup>[7]</sup>提出的 U-Net 网络,以 Concat 方式做了 4 次特征融合,以满足精细的医

**基金项目:** 上海市自然科学基金 (20ZR1422600)。

**作者简介:** 冯传盟(1994-),男,硕士研究生,主要研究方向:深度学习与图像处理;邢彦锋(1978-),男,博士,教授,主要研究方向:复杂薄板连接与增材制造;李学星(1986-),男,博士,讲师,主要研究方向:机电控制与光学测量;蒋世谊(1996-),男,硕士研究生,主要研究方向:机器视觉与光学测量。

**通讯作者:** 邢彦锋 Email: xyf2001721@163.com

**收稿日期:** 2021-07-19

学小目标分割任务。Vijay Badrinarayanan 等人<sup>[8]</sup>提出的 SegNet 网络,将编解码结构普适化,并在上采样过程中使用反池化操作,减少端到端的训练参数。Chen 等人<sup>[9]</sup>在 DeepLabv2 网络中提出 ASPP 结构来获得多尺度信息。在 DeepLabv3+ 网络<sup>[10]</sup>中,为了捕捉更清晰的目标边界,在 DeepLabv3<sup>[11]</sup>基础上增加了解码结构。

虽然 DeepLabv3+ 模型较好的解决了这上述两个挑战,但也存在一些不足。本文针对雾气研究对象,提出一种改进的 DeepLabv3+ 算法。研究的主要

贡献如下:

(1)对基础 DeepLabv3+ 模型融合层结构进行重新设计,并加入边缘细化模块、注意力通道模块。

(2)在雾气图像上具有良好的分割结果,实现了快速精细分割。

## 1 DeepLabv3+基础模型

DeepLabv3+ 基础模型网络结构如图 1 所示。该网络以 DeepLabv3 网络作为编码器,在此基础上增加解码模块进一步优化分割结果。

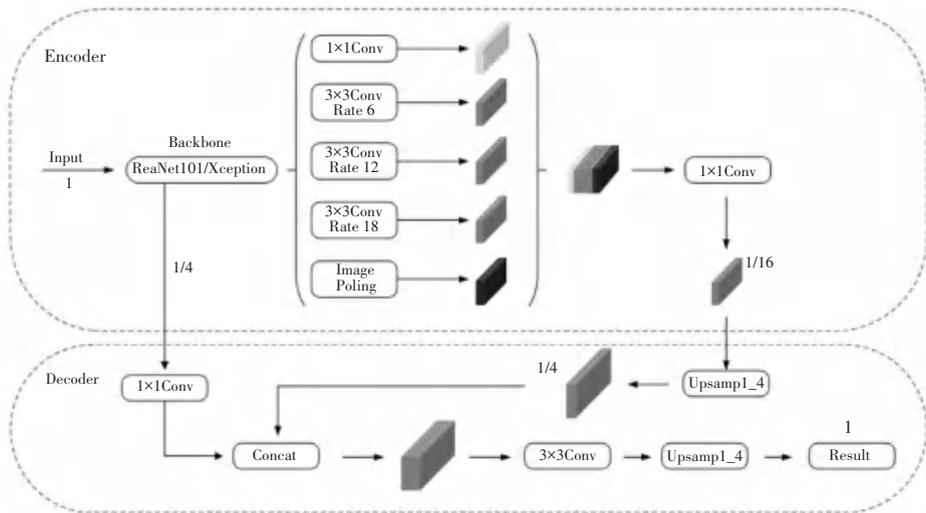


图 1 基础 Deeplabv3+ 模型网络结构

Fig. 1 Structure of the basic Deeplabv3 plus mode

在 Encoder 模块,初始输入图像在 Backbone 过程有选择地保留阶段特征,以存储局部信息做后续融合。通过空洞卷积控制,输出特征张量分辨率的同时调整感受野的大小。在 output-stride 等于 16 时,Backbone 最终输出特征张量分辨率为原始图像 1/16。不同于 DeepLabv3 中 ASPP 模块,DeepLabv3+ 中 ASPP 模块由 1 个  $1 \times 1$  的普通卷积层,分别以 6、12、18 为扩张率的并行空洞卷积层及 1 个全局池化特征层组成。ASPP 模块输出的多尺度信息以 Concat 方式拼接后,再通过  $1 \times 1$  卷积进行通道压缩。

在 Decoder 模块, $1 \times 1$  卷积降维后输出的特征张量,通过第一次双线性插值上采样 4 倍,与 Backbone 阶段储存的同分辨率特征张量做融合,融合后的特征张量再经过  $3 \times 3$  的卷积及第二次双线性插值上采样 4 倍,即可得到与原始输入图像相同分辨率的分割预测结果。

## 2 DeepLabv3+改进模型

### 2.1 融合层改进

DeepLabv3+ 基础模型以直接 4 倍上采样的方式

恢复图像分辨率,会造成特征张量中像素信息损失,破坏了图像中像素之间的连续性。且仅融合 Encoder 过程中储存的 1/4 大小的特征张量,忽略了更深层次特征张量包含的语义信息与浅层特征张量包含的空间细节信息,导致像素还原效果较差,最终雾气图像分割结果边缘信息丢失严重。基于此,本文提出的改进算法将 Encoder 阶段产生的低分辨率深层特征与高分辨率浅层特征进行逐级融合,以连续 2 倍上采样的方式,减少特征张量中像素信息损失,使分割边界更加精细。

### 2.2 边界细化模块

在 ResNet 网络残差模块的启发下,改进模型也使用到了类似的残差模块,用以进一步细化目标边界(Boundary Refinement Model)。模块结构示意图如图 2 所示。输入特征张量,经过一个  $1 \times 1$  的卷积进行通道压缩;在残差分支,通过 2 个  $3 \times 3$  的卷积进行特征细化;最后与输入特征张量做 Sum 融合。

### 2.3 通道注意力模块

以连续 2 倍上采样的方式,恢复图像分辨率包

含更多的特征融合操作。由于浅层特征与深层特征所处位置不同,经历的卷积层有所差异。浅层特征包含更多的空间细节信息,深层特征包含更多的语义信息,若简单的将浅层特征与深层特征直接融合效果可能较差。为了弥补浅层与深层特征之间的信息差距,改进模型将深层特征中的语义信息引入浅层特征,以进一步增强特征融合。

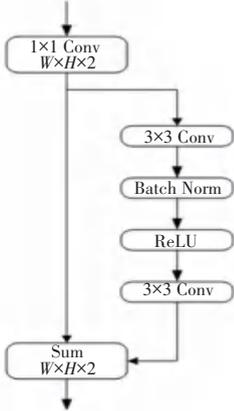


图 2 边界细化模块

Fig. 2 Boundary Refinement Model

在浅层与深层特征张量中,不同通道对目标预测能力的大小有差异。为了加强模型对重点通道特征张量的学习,快速提取重要特征并抑制对当前任务不重要的特征,改进模型使用到的注意力模块结构示意图如图 3 所示。浅层特征与深层特征 Concat 拼接后,输入全局池化层进行特征压缩。经第一个

$1 \times 1$  卷积进行通道压缩,  $ReLU$  激活函数保证输出为正, 第二个  $1 \times 1$  卷积进行特征映射后, 输入  $Sigmoid$  函数生成通道权值。输入特征张量各个通道与对应通道权值相乘即为重要特征通道突出后的特征张量。

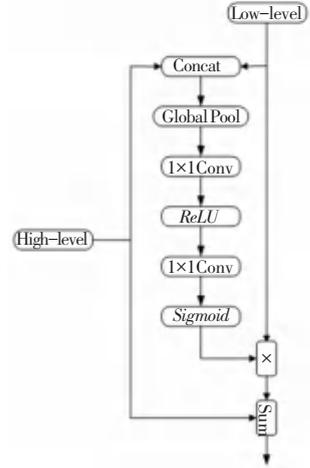


图 3 通道注意力模块

Fig. 3 Channel attention model

此外,改进模型将 ASPP 模块中的普通  $3 \times 3$  卷积替换为深度可分离卷积 (SeparableConv)。ASPP 模块输出特征张量以 Concat 方式拼接后,通道数变为 1 280,改进模型通过两个  $1 \times 1$  卷积减少通道压缩过程中的信息损失。DeepLabv3+改进模型整体结构示意图如图 4 所示。

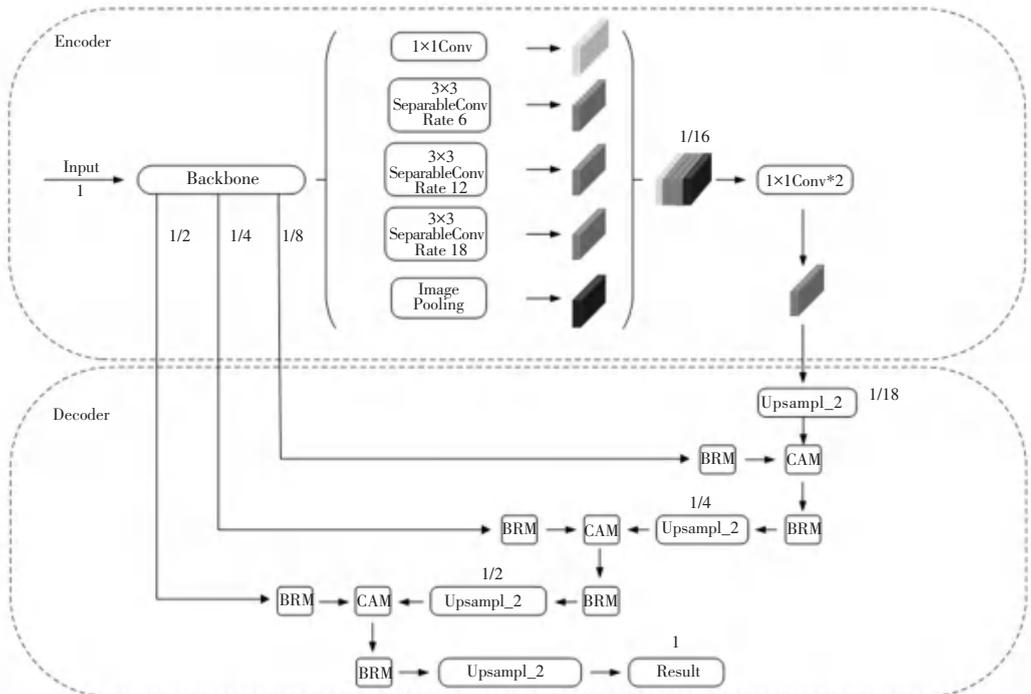


图 4 DeepLabv3+改进模型网络结构

Fig. 4 Structure of the improved Deeplabv3 plus mode

### 3 实验结果与分析

#### 3.1 实验环境

雾气图像分割模型实验环境基于 Ubuntu 系统、PyTorch 框架以及 Image 图像处理库等,具体的软硬件环境配置见表 1。

表 1 软硬件环境配置

Tab. 1 Hardware and software environment configuration

软硬件	版本
CPU	Intel Core i9-9920X
GPU	Nvidia GeForce RTX 2080Ti
OS	Ubuntu 18.04
CUDA	10.0
Cudnn	7.6
Python	3.7
Pytorch	1.0
Image	8.1.0

#### 3.2 数据集与预处理

由于目前尚未有公开的雾气图像分割数据集,所以改进模型实验数据集基于实拍制作而成。雾气图像采集装置见图 5 中 a 图,所用相机为大恒工业相机 MER-132-30 GM,采集数据集中包含训练图像 2 500 张,验证图像 800 张。

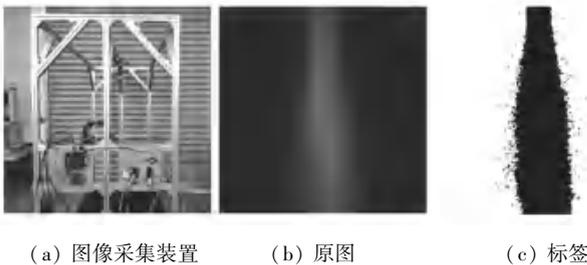


图 5 数据采集

Fig. 5 Data collection

通过工业相机拍摄得到的雾气原图如图 5 中 b 图所示,图 c 为通过传统阈值分割算法制作而成的标签图像。

为了使训练样本的形貌丰富多样化,避免模型过拟合并增强模型的泛化能力,实验使用到数据增强技术。其中包括随机缩放、随机裁剪、随机旋转、随机水平翻转、随机竖直翻转、及组合操作等。具体效果如图 6 所示。此外,通过随机改变图像的亮度、对比度和饱和度,以模拟现实世界中光照亮度等变化。

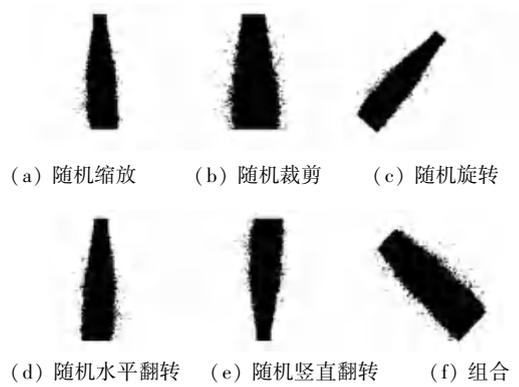


图 6 数据增强

Fig. 6 Data augmentation

#### 3.3 实验结果对比分析

实验过程中的超参数设置见表 2。

表 2 超参数设置

Tab. 2 Super parameter setting

超参数	设置
Batch	5
Input	512 * 512
Iterations	30 000
Learning rate	0.001
Momentum	0.9
Loss	Cross_Entropy
Output stride	16

实验总共迭代 30 000 次,训练过程中采用基于动量的随机梯度下降优化算法。其中, output - stride 等于 16,学习率采用 poly 策略, power 等于 0.9, ASPP 空洞率设置为 [6, 12, 18]。每隔 10 次打印一次 Loss,每隔 100 次进行一次验证,验证过程中保存当前最优 *MIOU* 值模型作为测试模型。为了动态观察模型收敛情况及分割结果,通过 Visdom 进行可视化展示。

常用的语义分割评价指标为:像素精度(*PA*)、均像素精度(*MPA*)、均交并比(*MIOU*)、频权交并比(*FWIOU*)。为了便于对比分析 DeepLabv3+改进模型与 DeepLabv3+基础模型及其它经典分割模型,评价时仅将 *MIOU* 作为分割结果评价指标。

针对 Deeplabv3+基础模型,本文对其进行不同骨干网络的替换与不同阶段特征融合对比分析实验,实验结果见表 3。其中, Backbone 表示骨干特征提取网络, Layer 表示进行特征融合时融合的特征张量分辨率大小。如 Layer 等于 4,代表只融合特征提取阶段储存的输入图像 1/4 大小的特征张量; Channel 表示输入图像经骨干特征提取网络后,输

出的特征张量通道数,也即 ASPP 模块输入通道数;  
*MIOU* 为分割评价指标;*Weigh* 表示模型所占内存大

小;*Parameters* 表示模型参数量;*Time* 表示在 CPU  
上运行单张图像所用时长。

表 3 不同骨干网络与特征融合的 Deeplabv3+基础模型分割性能对比

Tab. 3 Comparison of the segmentation performance of the Deeplabv3 plus basic model with different backbone networks and feature fusion

Group	Backbone	Loss	Layer	Channel	MIOU	Weigh/ M	Parameters/ Million	Time/ s
1	ResNet101	CE	4	2 048	90.11	183	42.649	2.616 8
2	Xception	CE	4	2 048	90.53	215	54.608	3.817 6
3	MobilenetV2	CE	4	1 280	90.06	19.4	6.986	1.528 9
4	MobilenetV2	CE	4	320	90.00	14.3	4.907	1.228 4
5	MobilenetV2	CE	4、8	320	90.06	14.4	4.916	1.395 6
6	MobilenetV2	CE	2、4、8	320	90.23	14.4	4.918	1.753 8

表 3 中,1、2 组实验为分别以 ResNet101<sup>[12]</sup>、  
Xception<sup>[13]</sup>网络为骨干网络,且仅融合 1/4 大小特  
征张量的 DeepLabv3+基础模型进行分割实验。由  
结果可知,基础模型分割速度较慢,且权重参数文件  
对内存需求较高,很难布置在移动设备以及嵌入式  
设备中。

表 3 中,3、4 组实验为分别以 1 280、320 通道  
的 MobileNetv2<sup>[14]</sup>网络为 backbone,且仅融合 1/4 大小  
特征张量的 DeepLabv3+模型进行对比分析实验。  
由分割结果可知,320 通道的 MobileNetv2 其分割精  
度相比 1 280 通道几乎没有变化,但模型参数量更

少,单张图像耗时更少。

表 3 中,4、5、6 组实验以 320 通道的 MobileNetv2  
网络为 backbone,分别融合不同阶段特征张量的模  
型进行对比分析实验。从分割结果可知,融合层为  
2、4、8 时,*MIOU* 值最高,分割边界最精细,从而验  
证了 2.1 节提出的改进思想。

通过消融实验,验证了在改进模型中加入的边  
界细化模块(BRM)、通道注意力模块(CAM)的有效  
性,实验结果见表 4。组别 1、2、3 皆以 320 通道  
的 MobileNetv2 网络为 backbone,融合层皆为 2、4、8  
层,实验超参数设置保持一致。

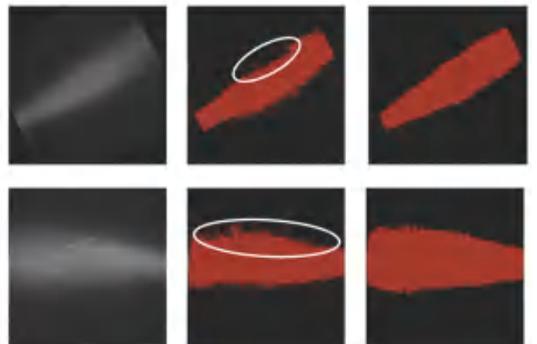
表 4 消融实验

Tab. 4 Ablation experiments

Group	CAM	BRM	MIOU	Weight/ M	Parameters/ Million	Time/ s
1	-	-	89.78	13.5	4.610 1	0.979 6
2	✓	-	90.07	13.5	4.610 3	0.986 6
3	✓	✓	90.31	13.5	4.610 6	1.118 4

由表 4 中 1、2 组实验对比结果可知,改进模型  
中通道注意力通道模块(CAM)的引入,加强了模型  
对重点通道特征张量的学习,提高了信息利用率;2、  
3 组实验对比结果表明,边缘细化模块(BRM)的引  
入进一步细化了目标边界。最终采取的改进模型即  
为组别 3。相比以 ResNet101 为 Backbone 的  
Deeplabv3+基础网络(表 3 中组别 1),最终改进的  
Deeplabv3+网络不仅带来了精度上的提升,分割速  
度亦提高了 57.26%,模型容量减少了 92.62%,更易  
布置在移动设备以及嵌入式设备中。

改进模型的雾气图像分割结果如图 7 所示。对  
于传统分割算法存在的分割凹陷问题(标签中椭圆  
部分),改进模型也能够较好的修补,且在较暗环境  
下也能得到良好的分割。



(a) 原始 (b) 标签 (c) 预测

图 7 改进模型分割结果可视化

Fig. 7 Improved model segmentation results visible

最后,将改进模型与当前一些比较流行的语义  
分割算法进行了比较分析,分析结果见表 5。由此

可见,本文提出的 DeepLabv3+改进模型,在分割精度及分割速度上都做到了最好,并且本文改进模型权重所占内存要远远少于其它模型,更易布置在移动设备以及嵌入式设备。

表5 不同分割模型性能对比

Tab. 5 Comparison of performance of different segmentation models

Model	MIOU	Weight/ M	Parameters/ Million	Time/ s
FCN8s	89.63	80.5	20.1	1.157 6
UNet	82.87	138.2	34.5	1.932 0
NestedUNet	90.31	36.7	9.2	2.766 9
Fusionnet	88.65	326.9	81.7	6.682 6
SegNet	90.23	117.9	29.4	1.777 5
Kernel_ Large_	90.08	233.8	58.3	3.280 2
Kernel_ Matters	90.09	344.9	86.0	3.325 5
ExFuse	90.24	351.4	87.7	3.589 1
DFN	90.21	262.9	65.6	2.559 3
PSPNet	90.21	262.9	65.6	2.559 3
DeepLabv3++	90.31	13.5	4.6	1.118 4

## 4 结束语

雾气图像的精确分割对工业领域具有重要的研究意义,本文针对基础 DeepLabv3+网络的不足之处,将原结构编码器的 Backbone 替换为更轻量的 Mobilenetv2 网络,将解码器的特征融合结构进行重新设计,同时加入注意力通道模块、边缘细化模块。实验结果表明,改进算法在分割精度带来提升的同时,分割速度亦大幅度提升,对于传统分割算法存在的分割凹陷问题也能够较好的修补,模型权重所占内存少使其更易布置在移动设备以及嵌入式设备,具有更大的工业应用价值。在后续研究中,将考虑利用深监督损失,进一步提升模型预测精度。

## 参考文献

[1] 王超. 选煤厂高压细水喷雾除尘系统研究[J]. 当代化工研究,

2021(5):48-49.

- [2] 卢旭,刘钊. 基于深度学习的图像语义分割技术综述[J]. 软件导刊,2021,20(1):242-244.
- [3] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Semantic image segmentation with deep convolutional nets and fully connected crfs[J]. Computer Science, 2014(4):357-361.
- [4] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing network[J]. IEEE Computer Society, 2016.
- [5] FARABET C, COUPRIE C, NAJMAN L, et al. Learning hierarchical features for scene labeling[J]. IEEE transactions on pattern analysis and machine intelligence, 2013, 35(8):1915-1929.
- [6] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 39(4):640-651.
- [7] 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 International Publishing, 2015, 39(4):640-651.
- [8] BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: A Deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017:1-1.
- [9] 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, 2018, 40(4):834-848.
- [10] CHEN L C, ZHU Y, PAPANDREOU G, et al. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation[J]. Springer, Cham, 2018.
- [11] CHEN L C, PAPANDREOU G, SCHROFF F, et al. Rethinking atrous convolution for semantic image segmentation[J]. 2017.
- [12] HE K, ZHANG X, REN S, et al. Deep Residual Learning for Image Recognition[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017.
- [13] CHOLLET F. Xception: deep learning with depthwise separable convolutions[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017.
- [14] 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.

(上接第88页)

- [14] JIE H, LI S, GANG S, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8):2011-2023.
- [15] WANG Q, WU B, ZHU P, et al. ECA-Net: efficient channel attention for deep convolutional neural networks[C]// 2020 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2020:11531-11539.
- [16] WANG X, GIRSHICK R, GUPTA A, et al. Non-Local Neural Networks[C]// 2018 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2018:7794-7803.

- [17] ZHANG S, CHI C, YAO Y, et al. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection[C]// 2020 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2020:9756-9765.
- [18] LI X, WANG W, WU L, et al. Generalized focal loss: learning qualified and distributed bounding boxes for dense object detection[J]. arXiv, 2020:1-14.
- [19] REZATOFIGHI H, TSOI N, GWAK J Y, et al. Generalized intersection over union: a metric and a loss for bounding box regression[C]// 2019 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2019:658-666.