

基于改进 U-Net 网络的吹氩图像分割方法<sup>①</sup>

黄禹康<sup>②\*</sup> 熊 凌<sup>③\*</sup> 刘 洋<sup>\*\* \*\*\*\*</sup> 邓 攀<sup>\*\*\*</sup> 但斌斌<sup>\*\*</sup>

( \* 武汉科技大学冶金自动化与检测技术教育部工程研究中心 武汉 430081)

( \*\* 武汉科技大学冶金装备及其控制教育部重点实验室 武汉 430081)

( \*\*\* 宝钢股份中央研究院(武钢有限技术中心) 武汉 430081)

**摘 要** 针对传统 U-Net 网络模型参数量大、图片处理时间长、无法满足工业生产实时性要求的问题,提出了一种改进 U-Net 网络的吹氩图像分割方法。该方法以 U-Net 框架为主体,使用传统 U-Net 网络的特征融合模块高效利用图像的特征信息,利用 MoblieNet 网络中的深度可分离卷积方法替代传统卷积,降低网络的参数量和计算量,缩短了分割所需的时间。实验结果表明,改进的 U-Net 网络在保持精度的同时,具有良好的实时性。与传统 U-Net 网络相比,其参数量缩小 15 倍,在 GPU 上运行平均耗时降低 6 倍。改进的 U-Net 网络处理一张分辨率为 224 × 224 像素的图片的平均耗时为 30 ms,可以满足工业生产对图像处理实时性的要求。

**关键词** 深度可分离卷积; U-Net; MobileNet; 吹氩图像; 图像分割

0 引 言

钢包底吹氩是炉外精炼的一种重要手段<sup>[1]</sup>,目前,往往采用人工观察的方法对吹氩流量进行手动调节,难以实现吹氩流量的精准控制<sup>[2]</sup>。为了实现氩气流量智能控制,需要通过采集钢包底吹氩图像并对图像进行分割,检测出钢渣、钢水区域,并计算裸露钢水区域面积,根据面积与吹氩流量之间的关系调节吹氩流量。吹氩流量智能控制的效果直接影响钢水的纯净度、质量和生产成本,因此,有必要对钢包底吹氩图像分割技术开展研究。

近年来,随着人工智能技术的发展,深度学习技术被广泛应用于图像分割领域。语义分割是一种像素级的分类任务,其通过将目标图像分为不同的语义类别,对不同类的像素点进行标记<sup>[3]</sup>。早期的语义分割方法主要使用传统聚类算法,在全卷积语义

分割网络出现后,深度学习技术在图像语义分割领域得到了快速的推广。Long 等人<sup>[4]</sup>在 2015 年提出全卷积神经网络(fully convolutional networks, FCN),使用卷积层替换网络中原有的线性层,使网络可以适应任意尺寸图像的输入,实现像素级的图像语义分割。全卷积神经网络在网络层之间使用了跳级结构,将低层特征与高层特征结合起来,使分割的效果更好。Ronneberger 等人<sup>[5]</sup>提出了 U-Net 网络,该网络具有对称的 U 型结构,通过对图像特征的编码与解码,融合了网络高层与低层的语义特征,提高了分割的准确性。与 FCN 网络相比,U-Net 网络对图像特征的利用率更高,可以在较少样本量的情况下完成模型训练并实现图像语义分割。Noh 等人<sup>[6]</sup>提出了 DeconvNet 网络,对 FCN 网络感受野固定、上采样粗糙等问题进行了改进。Yue 等人<sup>[7]</sup>在 U-Net 网络及 DeconvNet 网络基础上,提出了 SegNet 网络,在解码器中使用反池化对特征图进行上采样,并在分

① 国家自然科学基金(62073250)、湖北省重点研发计划(2020BAB021)和湖北省中央引导地方科技发展资金(2020ZYD022)资助项目。  
② 男,1997 年生,硕士;研究方向:图像处理,深度学习;E-mail: 793201261@qq.com。  
③ 通信作者,E-mail: 981217284@qq.com。  
(收稿日期:2021-2-3)

割中保持高频细节的完整性,在保证网络分割精度的前提下,减少了网络的参数量。其他还有如 PSP-Net、E-Net、Link-Net、Mask R-CNN 等网络基础上发展而来的网络模型在语义分割研究方向取得了很大的进展<sup>[8-11]</sup>。这些方法虽然在分割精度上不断提升,但模型对硬件设备和计算量的需求也在不断增加,导致图像处理速度较慢,不适合应用于工业实践。在实践生产过程中,关键在于如何使用较少的计算资源,既好又快地完成指定任务。目前对于轻量化模型的研究并不多见,主要的轻量化模型结构有 Xception、SqueezeNet、ShuffleNet、MobileNet 等<sup>[12-14]</sup>,虽然它们能够有效地降低模型的参数和计算量,但是仍存在训练过程复杂、模型精度不高的缺点。

本文针对吹氩生产中对图像处理实时性的要求改进了 U-Net 网络,该网络使用传统 U-Net 网络的特征融合方法,使网络具有较好的图像分割效果,参考了 MobileNet 网络对深度可分离卷积方法的运用,利用深度可分离卷积参数量和计算量较小的特点,设计了轻量化的语义分割网络。改进的 U-Net 网络同时具有较好的实时性和图像分割精度,可以满足吹氩生产中对图像处理快速性和准确性的需求。

## 1 改进的 U-Net 网络模型

为了实现吹氩流量自动控制,需要得到裸露钢水面积和吹氩流量之间的关系,将裸露钢水区域从钢液液面中分割出来是其中的关键步骤。U-Net 网络作为语义分割的经典模型,适用于吹氩图像分割。将 U-Net 网络应用于工业生产时,U-Net 网络存在着模型参数量大和计算量大等问题,不能满足工业生产中的高实时性要求,也对便携式硬件设备有较大的局限性。因此需要对 U-Net 网络进行轻量化设计,减少网络参数,提高图像处理速度。

### 1.1 深度可分离卷积

为了对 U-Net 网络进行轻量化设计,降低网络的参数量和计算量,采用 MobileNet 网络中的深度可分离卷积方法来替代传统卷积。深度可分离卷积与传统卷积方法的不同点在于,输入的特征张量的每

个通道都对应了一个单独的卷积核来进行卷积操作,卷积后输出特征张量。传统卷积和深度可分离卷积对比如图 1 所示。

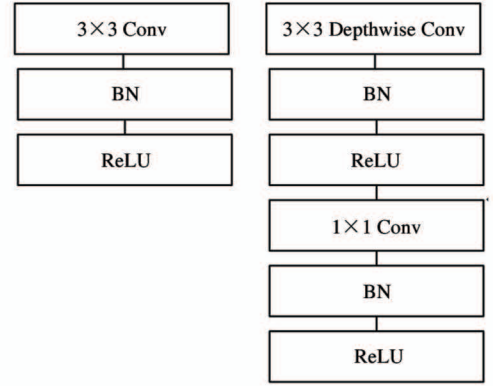


图 1 传统卷积和深度可分离卷积对比图

由图 1 可知,传统的卷积运算由标准卷积层、批量归一化(batch normalization, BN)和 ReLU 激活函数组成。深度可分离卷积运算由深度可分离卷积、逐点卷积、批量归一化和 ReLU 激活函数操作组成。

假设输入张量尺寸为  $D_F \times D_F \times M$ , 输出张量尺寸为  $D_k \times D_k \times N$ , 卷积核的尺寸是  $K \times K$ , 那么普通卷积的参数量和计算量分别为

$$P_c = K \times K \times M \times N \quad (1)$$

$$C_c = K \times K \times M \times N \times D_F \times D_F \quad (2)$$

深度可分离卷积的参数量和计算量分别为

$$P_d = K \times K \times M + M \times N \quad (3)$$

$$C_d = K \times K \times M \times D_F \times D_F + M \times N \times D_F \times D_F \quad (4)$$

则网络参数量和计算量之比为

$$\frac{P_d}{P_c} = \frac{C_d}{C_c} = \frac{1}{N} + \frac{1}{K^2} \quad (5)$$

由式(5)可以看出,深度可分离卷积操作的参数量和计算量明显低于普通卷积操作,因此利用深度可分离卷积这一特点设计轻量化的快速语义分割网络。

### 1.2 改进的 U-Net 网络结构

本文以 U-Net 框架为主体,网络结构如图 2 所示。

改进的 U-Net 网络仍然使用编码与解码结构,网络将低层特征与高层特征融合在一起,加强网络



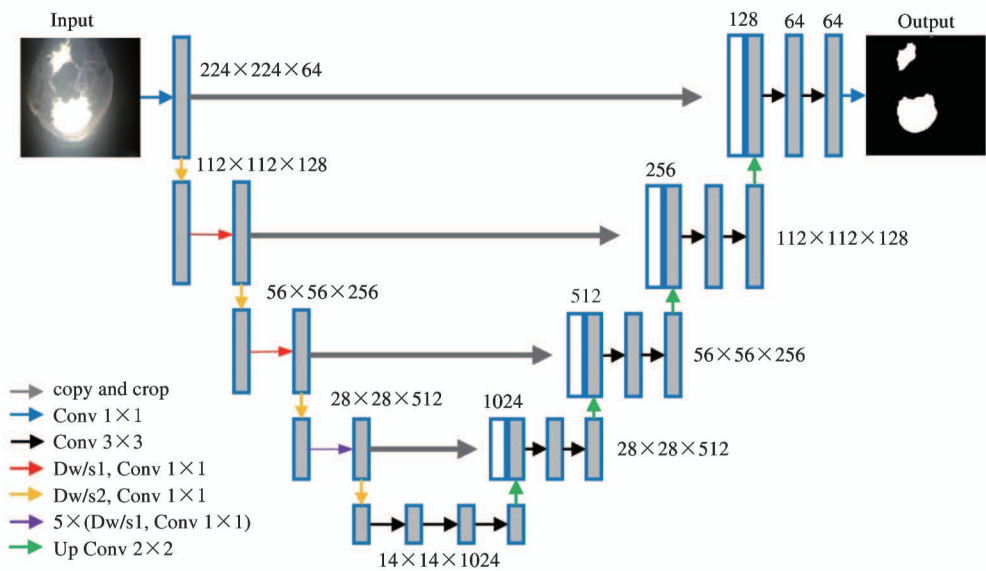


图 2 改进的 U-Net 网络结构

对图像细节的处理能力。相较于传统 U-Net 网络，本网络在以下方面进行了改进。

(1)使用深度可分离卷积替代传统卷积,为了弥补快速语义分割网络损失的精度,在提取特征的过程中连续使用 5 次步长为 1 的深度可分离卷积,使用多层深度可分离卷积增强网络对抽象特征的提取能力,可以提高特征的利用率,增加网络分割精度,使网络模型在准确率和实时性上达到较好的平衡。

(2)由于吹氩图像中钢渣区域与钢水区域在图像中所占比例不平衡,选择结合 Dice 相似系数(dice similarity coefficient, DSC)损失函数与二值交叉熵(binary cross entropy, BCE)损失函数的混合损失函数作为网络的损失函数,减少目标区域与背景区域在图像中占比不平衡对网络训练效果的影响,提高网络的分割精度。

(3)网络中的激活函数均采用 Leaky-ReLU 激活函数,避免过滤掉激活函数的负值信息导致卷积神经元的失活从而影响网络的训练。

1.3 损失函数

在只区分目标和背景的语义分割任务中,常使用二值交叉熵公式作为网络的损失函数,计算公式为

$$L_B = \frac{1}{N} \sum_{i=1}^N ((1 - G_i) \cdot \ln(1 - P_i) + G_i \cdot \ln(P_i)) \quad (6)$$

式中,  $G_i$  为像素点  $i$  的真实类别,  $P_i$  为语义分割网络对像素点  $i$  的预测结果。

在吹氩图像中,裸露钢水区域占图像的比例通常比钢渣区域要小。由于二值交叉熵损失函数在优化网络的过程中对图像上的每一个类别都平等地进行评估,再将各类对应的梯度进行回传,当两类区域占图像的比例不平衡时,占图像比例更大的那一类对网络的优化方向影响更大,从而使分割结果产生误差。

Dice 相似系数损失函数也是一种常用于语义分割任务的损失函数,在优化网络权重的过程中,使预测结果逐渐接近真实结果。但是当有像素预测错误时,会导致梯度变化十分剧烈,影响网络的训练过程。其公式如下:

$$L_D = 1 - \frac{2 \sum_{i=1}^N G_i P_i}{\sum_{i=1}^N G_i + \sum_{i=1}^N P_i} \quad (7)$$

本文综合二值交叉熵损失函数和 Dice 系数损失函数的特点,针对吹氩图像的特性,提出了一种混合损失函数,公式为

$$L = L_D + L_B \quad (8)$$

2 实验与分析

本文的模型在 TensorFlow 和 Keras 框架下实

现。硬件环境为 CPU 为 Intel Core i7-9700,GPU 为 NVIDIA RTX 2070。软件环境为 Window10、TensorFlow、Keras、Python3.7。

2.1 精度指标

为了验证网络的有效性,选取了精准率 (precision,P)、召回率 (recall,R)、F 分数和平均交并比 (mean intersection over union,MIOU) 指标对网络的分割性能进行了测试。为了量化每种网络的分割效果,通过对原始图像进行手工标注,得到真实标签 (ground truth,GT) 图像,如图 3 所示。



图 3 人工标注的真实标签图像

通常,基于 GT 图像的精度指标是从真阳性 (true positive, TP)、假阳性 (false positive, FP)、假阴性 (false negative, FN) 和真阴性 (true negative, TN) 计算出来的。将分割结果与 GT 图像进行比较,如果都将一个像素划分为背景类,则将其表示为 TP;如果仅有分割结果将其归类为背景类,则将其视为 FP;如果仅有 GT 图像将其分类为背景类,则将其视为 FN;如果此像素既没有在 GT 图像中出现,也没有在分割结果中出现,则被认为是 TN。精准率和召回率的公式为

$$P = \frac{|TP|}{|TP| + |FP|} \tag{9}$$

$$R = \frac{|TP|}{|TP| + |FN|} \tag{10}$$

其中 $|\cdot|$ 运算符表示相关区域中的像素数。当使用 P 和 R 指标比较不同的分割方法时,需要一个单独的性能指标进行度量。F 分数是 P 和 R 指标的结合,表示算法在 P-R 空间中的性能,其公式如式 (11)所示。

$$F_s = \frac{2PR}{P + R} = \frac{|TP|}{|TP| + (1/2)(|FN| + |FP|)} \tag{11}$$

其中 $|\cdot|$ 运算符表示相关区域中的像素数。

MIOU 指标的公式如下所示:

$$MIOU = \text{mean}(\frac{A \cap B}{A \cup B}) \tag{12}$$

式中,A 是人工标注的图像,B 是网络预测图像, $\frac{A \cap B}{A \cup B}$  是两类图像交集与并集之比。

2.2 数据集

本文采用吹氦工业生产现场采集的数据,数据集由 110 张图像组成,其中 70 张是训练集,30 张是验证集,10 张是测试集。图像分辨率为  $224 \times 224$  像素,每张图像都有对应的手工标注标签。为了增加训练数据,本文采用数据扩充的方式增加可用的训练样本数量。本文采用了 5 种数据扩充的方式,包括水平方向平移、水平方向翻转、垂直方向平移、垂直方向翻转和随机角度图像旋转。部分吹氦图像数据集扩充结果如图 4 所示。

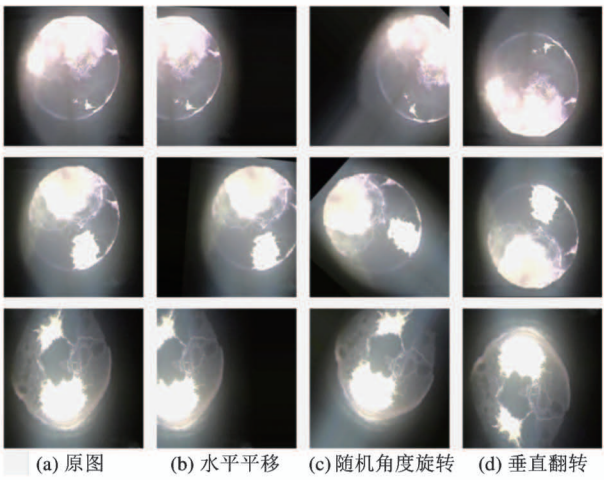


图 4 部分数据集扩充结果

在扩充了数据集之后,训练集、验证集和测试集在数据前后包含的图像数量如表 1 所示。

表 1 训练集、验证集和测试集的大小

| 扩充前后 | 合计   | 训练集 | 验证集 | 测试集 |
|------|------|-----|-----|-----|
| 扩充前  | 110  | 70  | 30  | 10  |
| 扩充后  | 1450 | 945 | 405 | 100 |



2.3 模型训练

训练时设置批尺寸为 8,使用 Adam 优化器对网络参数进行优化。改进 U-Net 网络模型训练集和验证集的准确率变化曲线如图 5 所示,损失函数随训练次数的变化曲线如图 6 所示。

可以看出,训练 50 轮时,损失函数已经收敛,此时训练集的准确率达到 99% 以上。

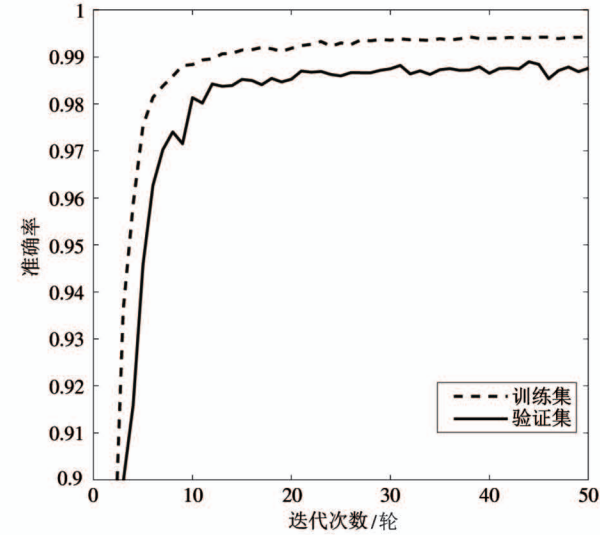


图 5 准确率变化情况

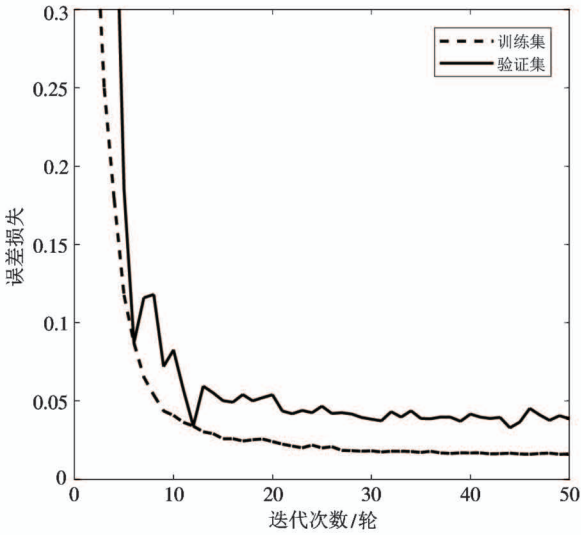


图 6 损失函数变化情况

2.4 实验结果

实验在吹氩数据集上进行,实验中将原始图像缩放至  $224 \times 224$  像素分辨率,再进行数据扩充操作。各网络模型实验结果如图 7 所示。

从图 7 可以看出,对比人工标注的真实标签图像,本文提出的网络分割图像能够取得与传统 U-Net 网络接近的效果,且在分割细节上优于 FCN 网络,

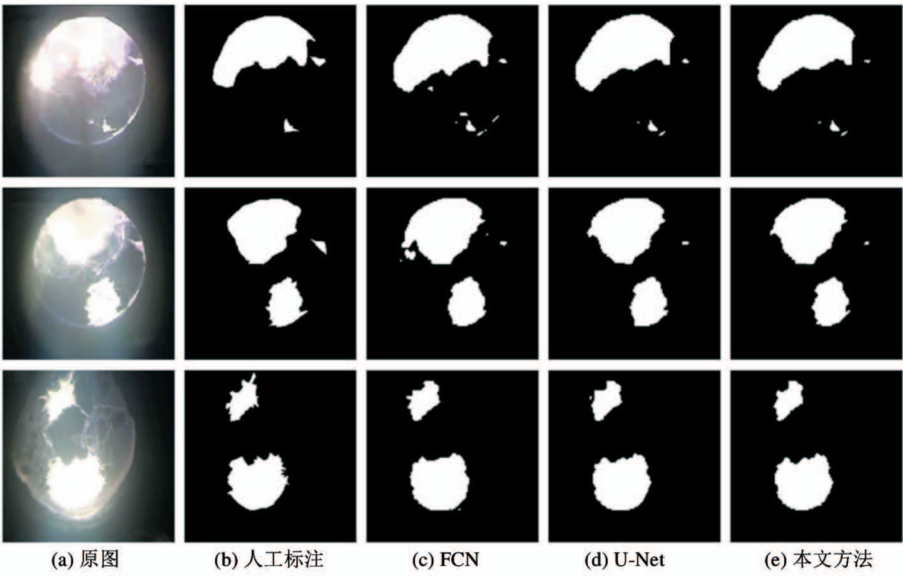


图 7 吹氩图像分割结果对比图

FCN 网络将部分钢渣区域误分为了裸露钢水区域。本文方法整体的计算量更少,计算速度表现得更快。在 CPU 上运行时,传统 U-Net 网络分割一张图片需

要 1.95 s,不能达到实时处理的要求,改进的 U-Net 网络需要 0.43 s,将速度提升了近 5 倍。在 GPU 运行条件下 U-Net 网络分割一张图片需要 0.18 s,而

改进的 U-Net 网络只需要 0.03 s。

实验的各项性能指标与参数对比如表 2 所示。

从表 2 中可以看出,在吹氩数据集中,本文方法的各项精度指标略低于传统 U-Net 网络,优于 FCN 网络,在保持分割精度的同时,大幅缩小了网络的参数量和计算量。参数量为 24.3 MB,对比传统 U-Net 网络缩小了 15 倍;计算量为 11.2 GFlops,对比传统 U-Net 网络缩小了 21 倍,通过减少网络的参数量和计算量,降低了设备的计算负担,加快了运行速度。本文方法在 2 种计算条件下的平均耗时均为最低,通过对比实验证明本文改进的 U-Net 网络在保证网络具有较好分割精度的同时,可以完成快速语义分割的任务,同时网络具有轻量化的结构,适合应用于吹氩生产实践场景中。

表 2 吹氩数据集不同网络分割结果

| 网络结构           | FCN   | U-Net | 本文方法 |
|----------------|-------|-------|------|
| MIOU           | 0.83  | 0.87  | 0.86 |
| P              | 0.89  | 0.93  | 0.92 |
| R              | 0.98  | 0.99  | 0.98 |
| F <sub>s</sub> | 0.94  | 0.96  | 0.95 |
| 参数量/MB         | 195.6 | 383.3 | 24.3 |
| 计算量/GFlops     | 90.6  | 242.8 | 11.2 |
| CPU 平均耗时/s     | 1.87  | 1.95  | 0.43 |
| GPU 平均耗时/s     | 0.26  | 0.18  | 0.03 |

3 结 论

工业生产应用对图像语义分割的实时性和准确性都有较高的要求,本文在传统 U-Net 网络框架的基础上,进行轻量化网络设计,实现了快速语义分割。相较于传统 U-Net 网络,改进的 U-Net 网络使用深度可分离卷积方法降低网络的参数量和计算量,在保持较好的精确度的同时,提高了网络运行速度。实验结果表明,改进的 U-Net 网络大幅减少了设备的计算负担,加快了语义分割的速度,在硬件设备方面更适合便携式设备使用,同时能保证工业生产的实时性要求,能够应用于工业生产实践。后续研究将引入更先进的算法和网络结构,提高轻量化网络的分割精度。

参考文献

[ 1 ] 周业连, 朱苗勇, 刘建斌, 等. 吹氩精炼钢包内非金属夹杂物去除机理[J]. 钢铁, 2016, 51(6): 39-46

[ 2 ] CHU K Y, LAI P H, CHEN H H, et al. The effects of bottom blowing gas flow rate distribution during the steel-making converter process on mixing efficiency[J]. *Metalurgical and Materials Transactions B: Process Metallurgy and Materials Processing Science*, 2016, 47(2): 948-962

[ 3 ] 郎辉宇, 吴俊君. 基于深度学习的图像语义分割技术研究综述[J]. 计算机工程与应用, 2019, 55(19): 12-21

[ 4 ] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Boston, USA, 2015: 3431-3440

[ 5 ] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation [C]//The 18th International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 2015: 234-241

[ 6 ] NOH H, HONG S, HAN B. Learning deconvolution network for semantic segmentation[C]//2015 IEEE International Conference on Computer Vision, Santiago, Chile, 2015: 1520-1528

[ 7 ] YUE Y J, LI X S, ZHAO H, et al. Image segmentation method of crop diseases based on improved segnet neural network [C] // 2020 IEEE International Conference on Mechatronics and Automation, Beijing, China, 2020: 1986-1991

[ 8 ] FONDEVIK S K, STAHL A, TRANSETH A A, et al. Image segmentation of corrosion damages in industrial inspections[C] // International Conference on Tools with Artificial Intelligence, Baltimore, USA, 2020: 787-792

[ 9 ] MOHAN S K, KRISHNA R, MANISH K R, et al. Semantic segmentation using ENet architecture [C] // The 2nd International Conference on Energy Systems, Drives and Automations, Kolkata, India, 2020: 631-637

[ 10 ] TRAN A, ZONOOZI A, VARADARAJAN J, et al. PP-LinkNet: improving semantic segmentation of high resolution satellite imagery with multi-stage training[C] // The 2nd ACM Workshop on Structuring and Understanding of Multimedia heritAge Contents, Seattle, USA, 2020: 57-64

[11] HE K M, GKIOXARI G, DOLLAR P, et al. Mask R-CNN[C] //International Conference on Computer Vision, Venice, Italy, 2017: 2980-2988

[12] BALADO J, OLABARRIA C, MARTINEZ-SANCHEZ J, et al. Semantic segmentation of major macroalgae in coastal environments using high-resolution ground imagery and deep learning[J]. *International Journal of Remote Sensing*, 2021, 42(5):1785-1800

[13] ZHANG X Y, ZHOU X Y, LIN M X, et al. ShuffleNet: an extremely efficient convolutional neural network for mobile devices[C] //IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 6848-6856

[14] DEBJYOTI S, MOHAMED E S. Thin MobileNet: an enhanced MobileNet architecture[C] //IEEE Annual Ubiquitous Computing, Electronics and Mobile Communication Conference, Columbia, USA, 2019: 280-285

## Argon blowing image segmentation method based on improved U-Net model

HUANG Yukang<sup>\*</sup>, XIONG Ling<sup>\*</sup>, LIU Yang<sup>\*\* \*\*\*</sup>, DENG Pan<sup>\*\*\*</sup>, DAN Binbin<sup>\*\*</sup>

(<sup>\*</sup> Engineering Research Center for Metallurgical Automation and Measurement Technology of Ministry of Education, Wuhan University of Science and Technology, Wuhan 430081)

(<sup>\*\*</sup> Key Laboratory of Metallurgical Equipment and Control of Ministry of Education, Wuhan University of Science and Technology, Wuhan 430081)

(<sup>\*\*\*</sup> Baosteel Central Research Institute (WISCO Limited Technology Center), Wuhan 430081)

### Abstract

Aiming at the problem that the traditional U-Net network model has large parameters and long processing time, which cannot ensure the real-time requirements of industrial production, an improved U-Net network method for argon blowing image segmentation is proposed. This method takes the U-Net framework as the main body and uses the feature fusion module of the traditional U-Net network to make efficient use of the image feature information. In the meantime, the depthwise separable convolution method in the MoblieNet network is used to replace the traditional convolution, which reduces the number of network parameters and calculations, and shortens the time required for segmentation. Experimental results show that the improved U-Net network has excellent real-time performance while maintaining accuracy. Compared with the traditional U-Net network, the number of parameters is reduced by 15 times, and the average running time on GPU is reduced by 6 times. The average time for the improved U-Net network to process a picture with a resolution of  $224 \times 224$  pixels is 30 ms, which can satisfy the real-time requirements of industrial production.

**Key words:** depthwise separable convolution, U-Net, MobileNet, argon blowing image, image segmentation