基于 DeepLabv3+ 的轻量化铝带表面缺陷分割方法

郝铁文 [1]
HAO Tiewen

摘 要 目前，在利用机器视觉检测铝带表面缺陷技术领域，仍存在诸多问题亟待解决。其中，较为突出的是算法复杂度高以及模型参数繁多，致使该技术难以在实际生产环境中有效落地应用，尤其在检测速度与精度的平衡方面，更是难以达成预期目标，无法满足工业生产的实际需求。针对上述问题，文章基于 DeepLabv3+ 网络提出了一种轻量级语义分割方法。首先采用轻量级网络 ShuffleNetv2 作为特征提取网络，并在其中嵌入坐标注意力（coordinate attention，CA）机制以增强模型定位目标区域的能力；其次对空间金字塔池化（atrous spatial pyramid pooling，ASPP）模块进行轻量化改进，使其能在减小网络参数的同时高效地利用特征信息；最后采用加权特征融合的方法来使全局上下文信息有效地从低分辨率向高分辨率传播。实验表明，该方法的平均交并比（mean intersection over union，mIoU）优于原始模型，达到了 87.31%。且相较于原始模型，本方法的参数量减小了 87%，计算量降低了 95%，能够灵活地部署于嵌入式系统以便在生产实际中应用。

关键词 缺陷检测；语义分割；DeepLabv3+；特征融合；空间金字塔池化

doi：10.3969/j.issn.1672-9528.2024.12.018

0 引言

铝带材凭借高强度与轻量化等显著优点，在新能源汽车、高铁、航空航天以及国防工业等诸多关键领域得以广泛运用 [1]。而随着这些领域的迅猛发展，市场对于铝带材的表面质量也提出了愈发严苛的要求。而铝带材在生产过程中易受生产设备、原料和生产工艺等因素影响，表面会产生一些缺陷，如黑斑、凹坑、划痕等，对产品外观甚至性能造成严重影响。因此对铝板带进行精准的表面质量检测至关重要。目前，铝板带表面缺陷检测方法主要有：人工目视法、无损检测法 [2]、基于机器视觉的表面缺陷检测方法 [3]。

人工目视法，检测效率低、易受主观因素影响，无法适应生产力的要求；传统无损检测法，包括涡流检测法、红外检测法、漏磁检测法 [4-6] 等，识别缺陷种类少、准确度低，无法对产品进行综合的质量评估。随着工业现代化的发展，基于机器视觉的表面缺陷检测方法解决了上述问题，得到了广泛应用 [7]。其中，传统的基于机器学习的视觉检测方法依赖人为设计特征提取算子，对于特征简单的缺陷检测稳定可靠，但对于特征复杂的缺陷，该模型漏检率和误检率高，泛化性能差。基于深度学习 [8] 的表面缺陷检测方法通过卷积神经网络自动学习图像特征，适应性强、检测精度高，相比传统机器视觉方法有很大优势，已被广泛应用于实际生产。

在实际生产中，仅检测出铝带表面的缺陷种类已经无法满足高质量的生产要求，基于深度学习的图像语义分割方法可以检测出缺陷的具体位置、面积、密度等信息，这些信息可以辅助产品进行更全面的质量评估，以满足高质量的生产需求 [9]。目前以全卷积网络（fully convolutional network，FCN）[10] 为基础的语义分割网络在各种工业场景中广泛应用。例如，Dong 等人 [11] 提出了基于金字塔特征融合和全局上下文注意力网络，用于检测表面缺陷，实现了信息从低分辨率到高分辨率的有效传播；Liang 等人 [12] 提出了一种基于纹理去除的缺陷检测方法，通过不同膨胀率的膨胀卷积块聚合多尺度上下文信息，准确地去除复杂的纹理背景，提高缺陷检测精度。

DeepLabv3+[13] 分割网络融合了编 - 解码器和空间金字塔池化技术，能够更好地捕获图像中的边界信息和丰富的上下文语义信息，已经被广泛地应用在缺陷检测、自然图像分割等领域 [14-17]。本文基于 DeepLabv3+ 分割网络在自制的铝带表面缺陷数据集上进行缺陷检测，根据检测结果发现还存在多个问题：一是对于铝板带表面缺陷检测，

1. 一重集团大连工程技术有限公司 辽宁大连 116600
[基金项目] 大连市揭榜挂帅技术攻关项目（2022JB11GX004）；一重集团天津重工有限公司科研合作项目（wx7061240801S）

DeepLabv3+ 模型参数量大、推理速度慢，难以满足生产实际要求；二是由于缺陷存在类内差异，导致缺陷检测结果不准确；三是不同类型的缺陷在尺寸上存在较大差异，尺寸较小的缺陷可能在检测中丢失信息。对于类内差异，模型要适应缺陷特征的变化，可以通过金字塔特征融合网络提取不同阶段的上下文信息，来充分挖掘缺陷的特征信息，以实现对缺陷变化的感知。对于缺陷的尺度差异，可以通过融合多尺度特征图来适应缺陷尺度的变化。对此，本文对 DeepLabv3+ 算法进行了改进，提出了一种轻量级多尺度特征融合分割算法。

1 改进的 DeepLabv3+ 网络

改进后的模型如图 1 所示。首先采用轻量化网络 ShuffleNetv2[18] 作为主干，通过嵌入坐标注意力机制（coordinate attention，CA）[19]，使模型更关注有效信息，实现在减小网络参数的同时增强模型的定位能力；其次将 ASPP（atrous spatial pyramid pooling，ASPP）模块替换为经过轻量化改进的 ASPPF（atrous spatial pyramid pooling fast，ASPPF）模块，使网络能够有效地提取多尺度特征，更好地满足铝带表面缺陷检测任务的需求；最后选取主干网络经过下采样倍率为 4、16、32 的特征图，通过快速归一化融合 [20] 的方式进行特征融合，使特征信息有效地从低分辨率向高分辨率传播。

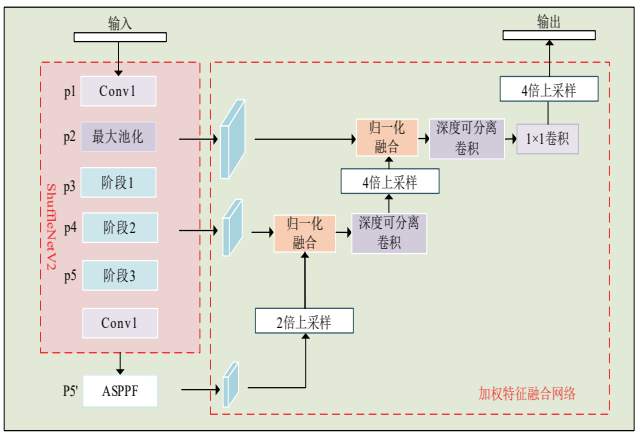


图 1 改进的网络结构

1.1 嵌入注意力机制的 ShuffleNetv2 轻量化主干

原始的 DeepLabv3+ 网络以 Xception[21] 为主干，网络参数量较大而且计算效率低，难以满足生产线上实时缺陷检测的要求。ShuffleNetv2 不同于先前的网络，直接基于运行速度设计网络结构，设计过程中除了考虑模型计算量，还考虑了内存访问成本和平台特点，因此该模型有着高效的计算性能，能够以极低的运算成本实现较高的识别准确率。

该网络包含五个阶段，共进行 5 次下采样。为了保证分割精度，本文在阶段 2、3、4 后添加了坐标注意力机制，以提高模型对缺陷特征的提取和定位能力。

1.1.1 坐标注意力机制（CA）

坐标注意力是一种即插即用的高效注意力机制，通过将横向和纵向的位置信息分别编码到通道信息中，帮助模型更好的定位缺陷信息，处理长距离依赖关系。操作过程包括坐标信息嵌入和坐标注意力生成两个步骤。其结构如图 2 所示。将坐标注意力模块嵌入到 ShuffleNetv2 主干网络中，可以在仅增加少量计算成本的前提下，提高网络对缺陷特征的提取和定位能力，将模型注意力聚集在重要特征的学习中，缓解背景噪音对特征挖掘造成的阻碍，从而显著提高模型的分割精度。

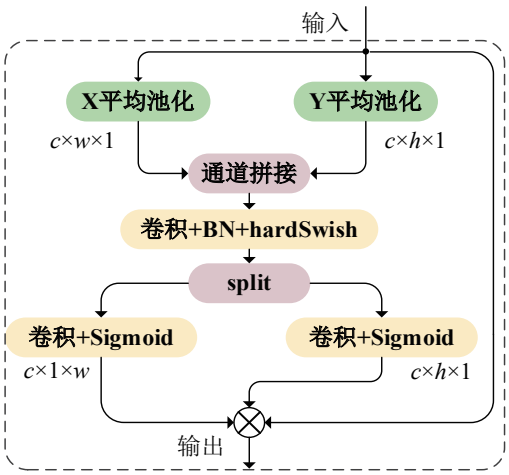


图 2 坐标注意力机制

1.2 轻量化的空间池化金字塔模块（ASPPF）

ASPP 模块模块由 4 个并行的空洞卷积和一个平均池化层组成，可以利用不同空洞率的空洞卷积提取多尺度特征，利用全局平均池化提取全局特征。然而，ASPP 模块还存在如下问题。

（1）随着空洞卷积空洞率的增加，有效滤波器的参数量会逐渐减小，导致模块无法提取有效的特征信息。

（2）由于空洞卷积稀疏采样，使远距离信息缺乏相关性。

（3）当输入通道数较大时，ASPP 模块内存占用率高。

针对上述问题，ASPPF 模块做出了改进，先采用 1×1 的卷积进行降维以降低模型参数。此外，重新设置了空洞率组合，以保证滤波器参数有效，避免丢失特征。在本文中，输入图像在主干网络中经过 5 次下采样，最终输出的特征图大小只有 16×16，因此空洞率设置不宜过大。

依照文献 [22] 提出的原则进行一组空洞卷积的设计。当连续使用多个空洞卷积时，应遵循以下原则来设置空洞率：（1）第一层卷积的空洞率应为 $r = 1$；（2）膨胀率的公约数

不能大于 1；（3）n 个卷积核为 $k\times k$ 的空洞卷积所对应的膨胀率为 $[r_1, r_2, \cdots, r_i, \cdots, r_n]$。由公式（1）计算得到的 M_i 应满足 $M_i \le k$，其中 r_i 表示第 i 层的膨胀系数，M_i 表示第 i 层空洞卷积的最大膨胀率。本文依照以上准则，重新设置了一组空洞卷积，其中空洞率分别为 1、3、7。

$$M_i = \max[M_{i+1} - 2r_i, M_{i+1} - 2(M_{i+1} - r_i), r_i] \quad (1)$$

ASPPF 结构如图 3 所示。输入特征先经过 1×1 卷积进行降维，再分别输入并行的空洞卷积。另一个分支与 ASPP 模块相同，进行全局平均池化操作，以提取全局特征。最后，将各层卷积的输出特征和全局特征进行拼接，得到铝带缺陷的多尺度特征。

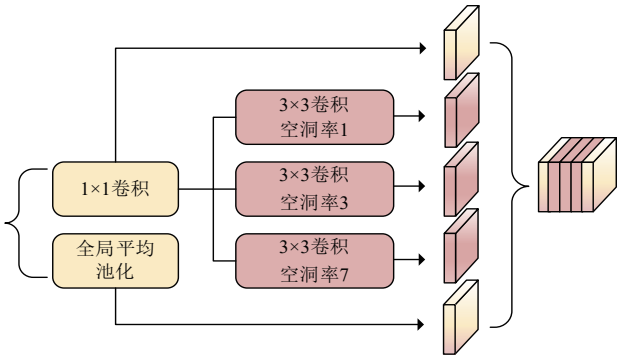


图 3 ASPPF 模块

1.3 加权特征融合网络

由于不同输入特征的分辨率不同，对输出特征的贡献各不相等，因此本文采用了加权融合的方法，为每个输入添加一个权重系数，让网络在迭代过程中学习每个输入特征的重要程度。快速归一化融合方法[23]是一种高效的加权融合方法，相比于其他方法如无界融合和基于 Softmax 的融合方法，它更加高效、简易。本文采用该方法进行特征融合，其公式表示为：

$$O = \sum_i \frac{w_i}{\varepsilon + \sum_j w_j} \cdot I_i \quad (2)$$

式中：w_i 是可学习的二维张量，通过 ReLU 激活函数来保证 $w_i \ge 0$；ε =0.000 1 用来保证数值稳定；I_i 为输入的特征。每个归一化权重的值介于 0 和 1 之间。

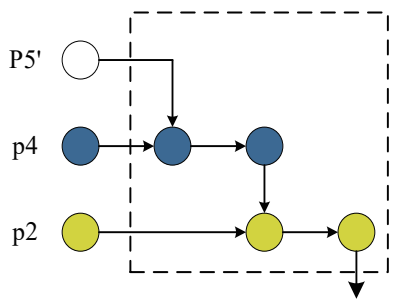


图 4 加权特征融合网络

加权特征融合网络如图 4 所示，先采用快速归一化方法融合不同分辨率特征，再通过深度可分离卷积（depthwise separable convolution，DSC）对输出特征进行进一步融合，可分别表示为：

$$P_4^{\text{out}} = \text{Conv}\left(\frac{w_1 \cdot P_4 + w_2 \cdot \text{Upsample}(P_{5'})}{w_1 + w_2 + \varepsilon}\right) \quad (3)$$

$$P_2^{\text{out}} = \text{Conv}\left(\frac{w_1' \cdot P_2 + w_2' \cdot \text{Upsample}(P_4^{\text{out}})}{w_1' + w_2' + \varepsilon}\right) \quad (4)$$

式中：Upsample 表示对特征进行空间上采样，Conv 代表深度可分离卷积。快速归一化融合使全局上下文信息可以有效地从低分辨率向高分辨率传播，增强了模型的分割性能。采用深度可分离卷积，可以提高效率、减少模型参数。

1.4 损失函数

本文将二元交叉熵损失函数（BCE loss）和 Dice 损失函数组合作为联合损失函数，将其表示为：

$$L_U(\hat{y}, y) = L_{\text{BCE}}(\hat{y}, y) + L_{\text{Dice}}(\hat{y}, y) \quad (5)$$

2 实验

2.1 实验设置

（1）参数设置：整个训练过程分为两个阶段，其中第一阶段将主干网络冻结，迭代次数为 50 次，批次大小为 16。第二阶段所有参数都参与训练，共迭代 150 次，批次大小为 8。训练开始使用了 ShuffleNetv2 预训练权重，采用随机梯度下降算法进行优化，初始学习率设置为 7e-3，衰减率为 1e-4，采用余弦衰减策略来更新学习率。

（2）计算平台：本文中所有实验都在 NVIDIA RTX A4000 16 GB 上进行。模型是基于 Python3.7.13、PyTorch1.7.1、CUDA11.0 构建的。

2.2 数据集

本文选取裂边、色差、黑斑、凹坑、擦痕、波浪和划痕这 7 种关键缺陷制作了铝板带表面缺陷数据集。使用 Labelme 对这些缺陷进行了人工分类和像素级标注。数据集共包含 5263 张图像，分为训练集、测试集和验证集，比例为 7:2:1。每张图像的分辨率为 512×512，每种缺陷的数量如表 1 所示。

表 1 数据集缺陷分布情况

单位：张

总计	裂边	色差	黑斑	凹坑	擦痕	波浪	划痕
5263	653	700	677	880	940	782	631

2.3 评价指标

为衡量不同模块和不同算法的性能，本文采用参数量和

浮点计算量（Flops）来衡量模型复杂度，用每秒处理图像的帧数（FPS）来衡量各个方法的实时性能，用平均交并比（mean intersection over union，mIoU）来评价分割结果。

2.4 消融实验

为了验证不同模块对分割结果的影响，在自制数据集上进行了一系列消融实验，实验结果如表 2 所示，将以 ShuffleNetv2 为主干的 DeepLabv3+ 作为基准网络，在主干中添加坐标注意力机制后 mIoU 提高了 0.64%，表明坐标注意力有助于实现对缺陷更准确的定位；将 DeepLabv3+ 中的 ASPP 模块替换成本文提出的 ASPPF 模块，精度提高 1.1%，表明该模块能够使输出特征包含更多可用信息；特征融合模块指将基准网络中的解码器替换成本文提出的加权特征融合模块，该模块使精度提升至 87.31%。

表 2 不同模块对缺陷分割结果的影响

DeepLabv3+（ShuffleNetv2）	CA	ASPPF	加权特征融合	mIoU/%
√				84.28
√	√			84.92
√	√	√		86.02
√	√	√	√	87.31

2.5 对比实验

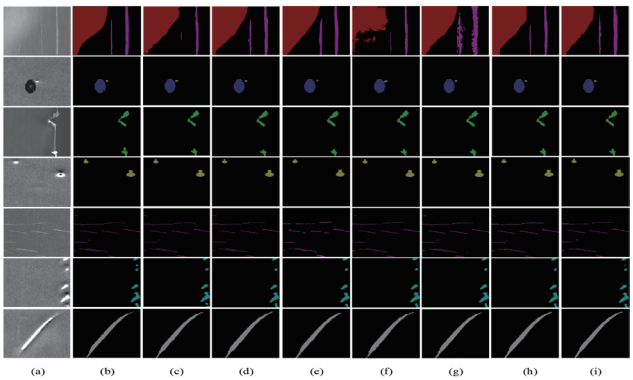
为了验证本文方法的有效性，在自制数据集上对比分析了本文方法和其他主流语义分割方法。为了保证实验公平，所有实验都采用相同的环境、图像大小和参数设置，实验结果如表 3 所示。

表 3 不同检测方法性能比较

方法	参数量 /(10^6)	Flops/(10^9)	mIoU/%	FPS
PSPNet[24]	46.710	59.110	81.69	30.394
SegNet[25]	29.448	161.470	82.34	25.603
FCN	32.950	277.751	86.19	23.804
UNet[26]	24.892	451.873	86.80	20.601
DeepLabv3+	54.710	166.891	87.02	22.876
HRNet[27]	65.851	93.700	88.07	8.399
Ours	7.316	8.428	87.31	29.885

与参数量最接近的方法 UNet 相比，本文模型精度提高了 0.51%；与计算量最接近的方法 PSPNet 相比，尽管本文模型的检测速率略低于 PSPNet，但是分割精度提高了 5.62%；HRNet 获得了最高的分割精度，但是该方法的检测速度较低。与基准模型 DeepLabv3+ 相比，本文的模型不仅降低了计算量和参数量，而且精度和检测速度还有所提升。与各个方法相比，本文的模型复杂度都是最低的。

为了更直观地对比不同方法的分割效果，对上述方法的分割结果进行可视化对比，结果如图 5 所示。图 5 第一行中色差缺陷与背景之间的对比度较低，UNet 和 SegNet 无法完整的检测出包含色差缺陷的全部像素，DeepLabv3+ 检测到的缺陷区域相对真实缺陷区域面积扩大；在第二行可以看到 SegNet 将凹坑错误地检测成了波浪；从第六行可以看出，部分模型如 PSPNet、DeepLabv3+ 和 FCN 漏检了尺寸较小的波浪缺陷。本文所提出的模型克服了上述模型在缺陷分割中存在的问题，能够相对准确地检测出不同尺寸的缺陷。



（a）原图；（b）标签；（c）UNet；（d）HRNet；（e）PSPNet；（f）SegNet；（g）DeepLabv3+；（h）FCN；（i）本文方法

图 5 分割结果对比图

4 结论

本文针对铝板表面缺陷检测问题，提出了一种基于 DeepLabv3+ 的轻量化语义分割方法，采用嵌入坐标注意力的 ShuffleNetv2 轻量级网络，使模型在降低参数的同时有效提取缺陷特征。通过轻量高效的 ASPPF 模块有效地提取多尺度语义信息。针对不同分辨率的特征映射，采用加权特征融合模块促进缺陷特征信息在模型中的高效流动。本文建立了铝板带表面缺陷图像数据集，并在该数据集上进行了对比实验，相比 DeepLabv3+，改进后的网络不仅参数量和计算量大幅降低，而且分割精度也有所提高，满足铝板带在生产线上进行缺陷检测的速度要求。

参考文献：

[1] 余欣木，蒋显全，谭小东，等．中国铝产业的发展现状及展望 [J]. 中国有色金属学报 ,2020,30(4):709-718.

[2] 张辉，宋雅男，王耀南，等．钢轨缺陷无损检测与评估技术综述 [J]. 仪器仪表学报 ,2019,40(2):11-25.

[3] 汤勃，孔建益，伍世虔．机器视觉表面缺陷检测综述 [J]. 中国图象图形学报 ,2017,22(12):1640-1663.

[4] 张荣华，叶松，马明，等．电涡流相位梯度及其在导电材料缺陷识别中的应用 [J]. 仪器仪表学报 ,2018,39(10):134-141.

[5] 沈功田 , 王尊祥 . 红外检测技术的研究与发展现状 [J]. 无损检测 ,2020,42(4):1-9+14.

[6] 沈功田 , 王宝轩 , 郭锴 . 漏磁检测技术的研究与发展现状 [J]. 中国特种设备安全 ,2017,33(9):43-52.

[7] 赵朗月 , 吴一全 . 基于机器视觉的表面缺陷检测方法研究进展 [J]. 仪器仪表学报 ,2022,43(1):198-219.

[8] YANN L ,YOSHUA B , GEOFFREY H. Deep learning[J]. Nature, 2015, 521(7553): 436-444.

[9] 陶显 , 侯伟 , 徐德 . 基于深度学习的表面缺陷检测方法综述 [J]. 自动化学报 ,2021,47(5):1017-1034.

[10] JONATHAN L, EVAN S,TREVOR D. Fully convolutional networks for semantic segmentation[C]// IEEE Transactions on Pattern Analysis and Machine Intelligence. Piscataway: IEEE, 2017,39(4): 640-651.

[11] DONG H W, SONG K C, HE Y, et al. PGA-Net: pyramid feature fusion and global context attention network for automated surface defect detection[J]. IEEE transactions on industrial informatics, 2019, 16(12): 7448-7458.

[12] LIANG Y, XU K, ZHOU P, et al. Automatic defect detection of texture surface with an efficient texture removal network[J]. Advanced engineering informatics, 2022, 53: 101672.

[13] 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). Berlin: Springer, 2018: 801-818.

[14] ZENG L, ZHANG S M, WANG P J, et al. Defect detection algorithm for magnetic particle inspection of aviation ferromagnetic parts based on improved DeepLabv3+[J]. Measurement science and technology, 2023, 34(6): 065401.

[15] ZHOU Q Q,LIU H L, TENGG S, et al. Automatic sewer defect detection and severity quantification based on pixel-level semantic segmentation[J]. Tunnelling and underground space technology, 2022, 123: 104403.

[16] XI D J,YI Q, WANG Z W. Attention Deeplabv3 model and its application into gear pitting measurement[J]. Journal of intelligent & fuzzy systems, 2022, 42(4): 3107-3120.

[17] JI M M, WU Z J. Automatic detection and severity analysis of grape black measles disease based on deep learning and fuzzy logic[J]. Computers and electronics in agriculture, 2022, 193: 106718.

[18] MA N N, ZHANG X Y, ZHNEG H T, et al. Shufflenet v2: practical guidelines for efficient CNN architecture design[C]// Proceedings of the European Conference on Computer Vision (ECCV). Berlin:Springer, 2018: 122-138.

[19] HOU Q B, ZHOU D Q, FENG J S. Coordinate attention for efficient mobile network design[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE, 2021: 13713-13722.

[20] SIFRE L, MALLAT S. Rigid-motion scattering for image classification[DB/OL]. (2014-03-07)[2024-06-22].https://doi.org/10.48550/arXiv.1403.1687.

[21]CHOLLET F. Xception: deep learning with depthwise separable convolutions[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017:1800-1807.

[22] WANG P Q, CHEN P F, YE Y, et al. Understanding convolution for semantic segmentation[C]//2018 IEEE Winter Conference on Applications of Computer Vision (WACV). Piscataway:IEEE, 2018: 1451-1460.

[23] TAN M X, PANG R M, LE Q V. Efficientdet: scalable and efficient object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE, 2020: 10781-10790.

[24] ZHAO H S, SHI J P, QI X J, et al. Pyramid scene parsing network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE, 2017: 2881-2890.

[25] 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.

[26] OLAF R, PHILIPP F, THOMAS B. U-Net: convolutional networks for biomedical image segmentation[C]//Medical image Computing and Computer-Assisted Intervention–MICCAI 2015. Berlin: Springer, 2015: 234-241.

[27] SUN K, XIAO B, LIU D, et al. Deep high-resolution representation learning for human pose estimation[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.Piscataway:IEEE, 2019: 5693-5703.

【作者简介】

郝铁文（1980—），男，辽宁大连人，博士研究生，一重集团天津重工有限公司室主任、首席工程师、高级工程师，研究方向：大型成形装备设计与制造、智能检测与智能制造技术。

（收稿日期：2024-08-28）