

基于激励压缩空洞卷积改进 U-Net 网络的物流托盘图像分割算法

doi:10.3969/j.issn.1674-7100.2021.05.005

魏占国 宋娅萍 李 亚

中南林业科技大学

物流与交通学院

湖南 长沙 410004

摘 要: 针对物流托盘种类多、形体规则度复杂,以及在工业生产环境下托盘存在被遮挡、光照条件变化等因素影响托盘图像分割效果的问题,提出一种新颖的基于激励压缩空洞卷积 (SEDC) 改进的 U-Net 网络,通过对特征通道间的相关性进行建模,强化重要特征,提升物流托盘图像分割性能。在 SEDC 模块中使用 $1 \times 1 \times 1$ 卷积进行数据降维与升维大幅降低计算量,利用正常卷积与膨胀率为 2 的空洞卷积探索不同视野下的图像特征,通过 SE 模块对不同层的重要程度进行自动学习。实验结果表明:相较于现有的一些经典图像分割算法,本文模型在尽可能保证图像分割性能的情况下大幅降低了模型的计算量,并提升了网络的鲁棒性,为物流托盘图像的智能分割提出了一种新的解决思路。

关键词: 物流托盘图像; 卷积神经网络; SEDC

中图分类号: TQ152

文献标志码: A

文章编号: 1674-7100(2021)05-0035-07

引文格式: 魏占国, 宋娅萍, 李 亚. 基于激励压缩空洞卷积改进 U-Net 网络的物流托盘图像分割算法 [J]. 包装学报, 2021, 13(5): 35-41.

0 引言

托盘在现代物流运输行业中起着关键的作用,它可以将包裹规格化和模块化,从而极大地提高运输效率。传统的人工叉车已不能满足现代物流运输搬运作业的高效率要求。因此,基于托盘智能分割识别的自动化叉车成为智慧物流行业的前沿热点。自动化叉车高效精准的作业关键在于可视化识别系统的物流托盘分割算法能准确识别托盘图像。

随着算法的提升和深度学习技术的飞速发展,卷积神经网络 (convolutional neural networks, CNN)

在图像语义分割^[1-3]任务上表现出了强大的性能。图像语义分割是一种将图像分割成具有精确语义类别属性的区域序列方法,即利用每个像素及其周围的图像块独立地对每个像素进行分类。加州大学伯克利分校的 J. Long 等提出的全卷积网络 (fully convolutional networks, FCN) 推广了传统的 CNN 结构,能在不带全连接层的情况下进行密集预测,但是 FCN 会造成图像位置信息的丢失^[4-7]。为使语义分割能够完全贴合图像,保留图像位置信息,研究者提出了两种不同形式的结构:

1) 编码器-解码器 (encoder-decoder) 结构。

收稿日期: 2021-07-27

基金项目: 国家林业和草原局科技成果推广计划基金资助项目 (2016-51); 湖南省研究生科研创新基金资助项目 (QL20210212)

作者简介: 魏占国 (1978-), 男, 安徽蒙城人, 中南林业科技大学副教授, 博士, 硕士生导师, 主要从事物流装备技术, 运输包装材料及单元载具检测技术研究, E-mail: jackwzg007@163.com

编码器利用池化层逐渐减小输入数据的空间维度,而解码器则利用像反卷积层这样的网络层来逐步修复目标对象的细节和空间维度。编码器和解码器通常有直接的信息连接,因而解码器可以更好地恢复目标对象细节。编码器-解码器结构的典型网络包括 U-Net^[8]、SegNet^[9]。

2) 空洞卷积 (dilated convolutions) 结构^[10-12]。空洞卷积结构可以在去除池化层的情况下扩大视野域,避免池化处理带来的位置信息丢失。该结构的典型网络包括 DeepLab V1、DeepLab V2^[10-11]。

受工业生产环境下物流托盘易被遮挡、光照条件变化以及托盘类型繁多等因素的影响,现有物流托盘分割算法难以进行精准的图像识别和有效分割,进而影响自动化叉车的精准装卸作业。基于此,本文以 U-Net 作为主干网络,提出一种新颖的激励压缩空洞卷积模块 (squeeze excitation dilated convolution,

SEDC) 来改进 U-Net 网络,通过对特征通道间的相关性进行建模,强化重要特征,提升物流托盘图像分割的准确率。

1 网络模型

1.1 改进的 U-Net 网络

为实现物流托盘的精准分割,本文构建了一个轻量级的物流托盘分割模型即基于 SEDC 模块改进的 U-Net 网络^[12-14],网络模型如图 1 所示。在 SEDC 模块(如图 2 所示)中,先用 $1 \times 1 \times 1$ 卷积进行数据降维,以降低计算量,再用传统卷积与膨胀率为 2 的空洞卷积获取多尺度感受视野下的图像特征,并通过 SE 模块 (squeeze-and-excitation) 提升 U-Net 网络对通道特征的敏感性,使其更加有效地对不同层的重要程度进行自动学习,最后以 $1 \times 1 \times 1$ 卷积进行数据升维。

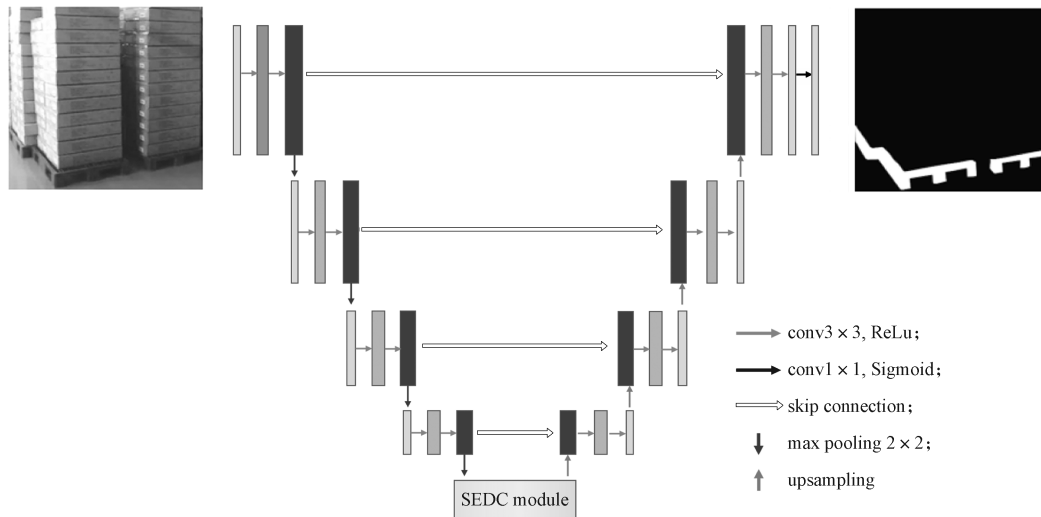


图 1 基于 SEDC 模块改进的 U-Net 网络

Fig. 1 Improved U-Net network model based on SEDC module

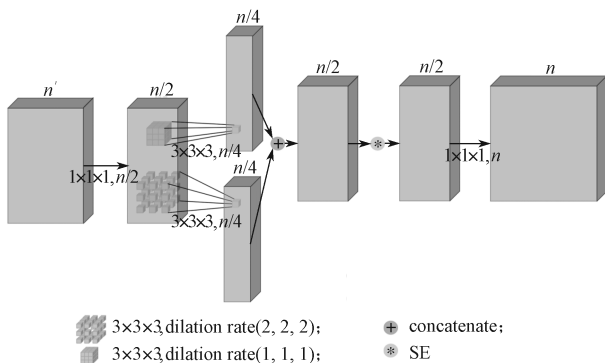


图 2 SEDC 模块示意图

Fig. 2 SEDC module

1.2 SEDC 模块

1.2.1 $1 \times 1 \times 1$ 卷积

采用 $1 \times 1 \times 1$ 卷积的动机如下。

1) 实现跨通道的交互和信息整合

$1 \times 1 \times 1$ 卷积对多个特征进行线性组合,能够保持数据的宽度、高度与深度不变,实现跨通道的特征整合,提升网络的表征能力。

2) 实现卷积核通道数的降维和升维

$3 \times 3 \times 3$ 卷积在几百个过滤器的卷积层上做卷积操作非常耗时,因此使用 $1 \times 1 \times 1$ 卷积在 $3 \times 3 \times 3$ 卷积层前后进行数据降维与升维操作,使参数数量进一

步减少, 缩短网络的训练时间。

1.2.2 空洞卷积

空洞卷积是在标准的卷积映射里注入空洞。相比传统卷积, 空洞卷积多了一个超参数, 即膨胀率 (dilation rate), 它代表卷积核 (kernel) 的间隔数量 (传统卷积的卷积核膨胀率为 1)。空洞卷积的优点是在不做池化操作、丢失特征信息的情况下, 加大感受视野, 让每个卷积输出都包含较大尺度的信息^[15-16]。

1.2.3 SE 模块

SE 模块^[17]的主要作用是关注通道之间的关系, 使模型可以自动学习到不同通道特征的重要程度。SE 模块包括压缩和激励两个关键操作, 如图 3 所示。图中, W' 、 H' 表示特征图宽和高, C' 表示通道数, 输入特征图大小为 $W' \times H' \times C'$ 。压缩操作是对卷积得到的特征图进行聚合, 将维度为 $W \times H$ 的特征图作为特征描述器, 以此获得全局感受视野信息。激励操作是一种自筛选机制, 使用样例特化激活函数对所有通道进行权值评比。

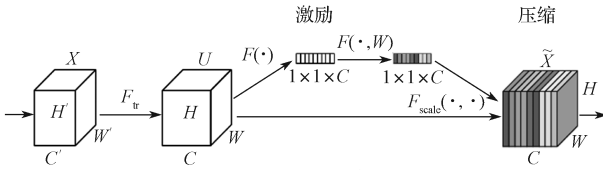


图 3 SE 模块

Fig. 3 SE module

SE 模块可以适用于任何映射 $F_{tr}: X \rightarrow U$, $X \in \mathbf{R}^{H' \times W' \times C'}$, $U \in \mathbf{R}^{H \times W \times C}$ 。以卷积过程为例, 卷积核为 $V=[v_1, v_2, \dots, v_C]$, 那么输出为 $U=[u_1, u_2, \dots, u_C]$ 。

$$u_c = v_c * X = \sum_{s=1}^{C'} (v_c^s * x^s),$$

式中: $*$ 为卷积操作; v_c 为第 c 个卷积核; v_c^s 为第 s 个通道的第 c 个卷积核; x^s 为第 s 个通道像素。输入一个通道上的空间特征, 卷积核会学习特征空间关系。由于对各个通道的卷积结果做了求和运算, 因而通道的特征关系与卷积核学习到的空间关系被混合在一起。而 SE 模块可抽离这种混杂, 使得模型能够直接学习到通道特征关系。

2 实验

2.1 评估指标

为了验证本模型对物流托盘图像分割任务的有效性, 本文采用自行采集的物流托盘图像

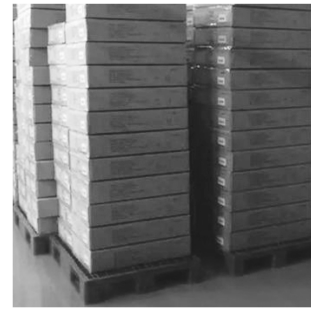
集 (如图 4 所示) 训练与测试网络模型, 用精度 (P)、召回率 (R) 和 $F1$ 分数 3 个评价指标评价算法的分割性能。3 个评价指标的公式如下:

$$P = \frac{TP}{TP + FP},$$

$$R = \frac{TP}{TP + FN},$$

$$F1 = \frac{2 \times P \times R}{P + R}.$$

式中: TP 为预测具有正类别标签数据的正类别数量 (即物流托盘被正确检测到的像素数); FP 为预测具有负类别标签数据的正类别数量; (FN 为预测具有正类别标签数据的负类别数量 (即物流托盘被正确检测但又被错误识别的像素数))。



a) 物流托盘图像



b) 分割图像

图 4 采集的物流托盘图像以及分割后的样例

Fig. 4 An example of the image of collected logistics pallet

2.1.1 指标函数

1) Dice 系数

Dice 系数, 即重合索引, 在信息检索领域也称为 $F1$ 分数, 其目标是在验证物流托盘图像分割效果时保证高查全率和查准率^[18]。较之自动分割结果与原标签数据的差异比较方法, Dice 系数可以更好

地刻画分割效果。Dice 系数公式为：

$$Dice = \frac{2|G \cap R'|}{|R'| + |G|} = \frac{2TP}{2TP + FP + FN},$$

式中： G 为原标签数据的真实分割结果； R' 为测试数据的自动分割结果。理想情况下，原标签数据的真实分割结果应与自动分割结果完全重合，即 $G=R'$ ，Dice 系数的值为 1。

2) 豪斯多夫距离

豪斯多夫距离表示一个集合到另一个集合中最近点的最大距离^[17]。从集合 A 到集合 B 的豪斯多夫距离是一个极大值函数，即：

$$h(A, B) = \max_{a \in A} \left\{ \min_{b \in B} \{d(a, b)\} \right\},$$

式中 $d(a, b)$ 为 a 、 b 两点之间的欧几里得距离。

豪斯多夫距离用来刻画轮廓的相似性，值越大代表越不相似，反之，值越小代表越相似。

3) 平均交并比

平均交并比 (mean intersection over union, MIoU) 是用于分割性能评价的常用指标，计算两个集合的交集和并集的比率^[19]。在本文中，平均交并比是原标签数据的真实分割结果和预测结果的比率，即：

$$MIoU = \frac{1}{k+1} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p'_{ij} + \sum_{j=0}^k p_{ij} - p_{ii}},$$

式中：MIoU 为平均交并比； p_{ii} 为真正样本； p'_{ij} 为假正样本； p_{ij} 为假负样本。

2.1.2 损失函数

用 Adam 作为算法优化器，因为它具有快速收敛的特性^[20]。将一阶矩估计的指数衰减率设置为 0.99；二阶矩估计的指数衰减率设置为 0.999；Epsilon 设置为 $1e-8$ ；Decay 学习率衰减设置为 $3e-8$ 。

2.2 实验分析

本实验采用的硬件平台为英特尔酷睿 i7-8700K CPU @ 3.70 GHz，GPU 为 GeForce GTX 1080，RAM 为 16.0 GB，操作系统为 Windows 10 专业版，编程语言为 Python 3.6.5，深度学习开发库为 Keras 2.1.5。

2.2.1 消融实验

为验证 SEDC 模块的有效性，本文对模型进行消融实验。5 个测试样本从测试集中随机选取，消融实验结果见图 5 和表 1。其中，U-Net with SE 表示使用 SE 模块改进的 U-Net 网络，U-Net with SE and $1 \times 1 \times 1$ conv 表示使用 SE 和 $1 \times 1 \times 1$ 卷积改进的 U-Net 网络。

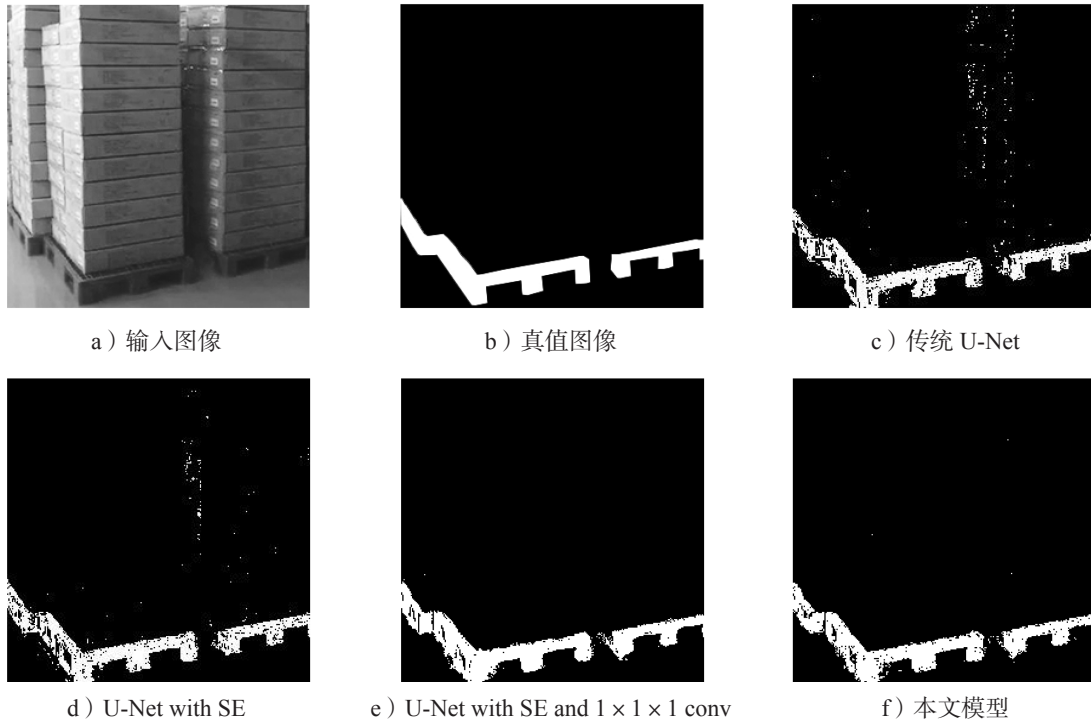


图 5 样本 1 对比实验的可视化结果

Fig. 5 Visualized results of the comparative experiment on the first sample

表 1 4 种网络模型对物流托盘图像的分割效果对比

Table 1 The results of the four network models compared on the logistics pallet image data set

样本序号	网络模型	P	R	$F1$	计算量/ 10^6
1	传统 U-Net	0.8825	0.9314	0.8897	19.07
	U-Net with SE	0.8901	0.9117	0.8997	19.15
	U-Net with SE and $1 \times 1 \times 1$ conv	0.8904	0.9387	0.8797	7.49
	本文模型	0.8903	0.9378	0.8945	8.68
2	传统 U-Net	0.8811	0.9211	0.8900	19.07
	U-Net with SE	0.8932	0.9200	0.8911	19.15
	U-Net with SE and $1 \times 1 \times 1$ conv	0.8914	0.9304	0.8836	7.49
	本文模型	0.8911	0.9299	0.8849	8.68
3	传统 U-Net	0.8722	0.9135	0.8959	19.07
	U-Net with SE	0.8643	0.9069	0.8912	19.15
	U-Net with SE with $1 \times 1 \times 1$ conv	0.8567	0.9125	0.8825	7.49
	本文模型	0.8699	0.9122	0.8887	8.68
4	传统 U-Net	0.9025	0.9415	0.8952	19.07
	U-Net with SE	0.8922	0.9398	0.9024	19.15
	U-Net with SE and $1 \times 1 \times 1$ conv	0.9012	0.9408	0.9152	7.49
	本文模型	0.9010	0.9416	0.9217	8.68
5	传统 U-Net	0.8625	0.9245	0.8925	19.07
	U-Net with SE	0.8704	0.9324	0.9025	19.15
	U-Net with SE and $1 \times 1 \times 1$ conv	0.8825	0.9311	0.8925	7.49
	本文模型	0.8911	0.9387	0.9025	8.68

由表 1 可知:

1) 与传统 U-Net 网络相比, U-Net with SE 网络的性能有小幅提升, 但参数量有小幅增长。

2) 与 U-Net with SE 网络相比, U-Net with SE and $1 \times 1 \times 1$ conv 网络的参数量仅为 U-Net with SE 网络的 39%, 但网络性能明显减弱。

3) 与传统 U-Net 网络相比, 本文模型的参数量下降了 54.5%, 网络结构更加轻量化。在大幅降低网络结构复杂度的情况下本文模型的性能没有大幅度变化。可见, 本文模型在尽可能保证图像分割性能的情况下大幅降低了模型的计算量, 并提升了网络的鲁棒性。

2.2.2 不同模型性能对比实验

为了进一步验证本文模型的优越性, 将本文模型与 Attention U-Net 模型^[17]、SegNet 模型和 PSPNet 模型^[21]进行对比实验。实验结果如表 2 所示。

表 2 物流托盘图像分割对比实验结果

Table 2 The results compared on the logistics pallet image data set

网络模型	P	R	$F1$
Attention U-Net	0.8188	0.8502	0.8722
SegNet	0.8417	0.9276	0.8511
PSPNet	0.8497	0.8915	0.8456
本文模型	0.8903	0.9378	0.8945

由表 2 可知, 本文模型的性能均优于其他 3 种模型, 其原因是本文模型有效地关注了通道之间的关系, 自动学习不同通道特征的重要程度, 从而得到更优的分割效果。

3 结语

本文提出了一种新颖的基于激励压缩空洞卷积改进的 U-Net 网络用于物流托盘图像分割识别任务。利用 $1 \times 1 \times 1$ 卷积降低参数量, 使用传统卷积与膨胀率为 2 的空洞卷积获取多尺度感受视野下的图像特征, 同时引入 SE 模块有效对不同层的重要程度进行自动学习, 提高模型鲁棒性。消融实验和对比实验的结果证明了本文模型的有效性。未来的研究工作将进一步研究编码与解码之间的关联性, 充分利用低级特征与语义信息对模型进行优化。

参考文献:

- [1] 刘桂红, 赵亮, 孙劲光, 等. 一种改进粒子群优化算法的 Otsu 图像阈值分割方法 [J]. 计算机科学, 2016, 43(3): 309-312.
LIU Guihong, ZHAO Liang, SUN Jinguang, et al. Otsu Image Threshold Segmentation Method Based on

- Improved Particle Swarm Optimization[J]. Computer Science, 2016, 43(3): 309–312.
- [2] 王 卓, 葛 斌, 涂明玉, 等. 基于改进大津法与人工鱼群优化的图像分割算法 [J]. 包装学报, 2019, 11(2): 81–86.
WANG Zhuo, GE Bin, TU Mingyu, et al. Image Segmentation Algorithm Based on Improved Otsu Algorithm and Artificial Fish Swarm Optimization[J]. Packaging Journal, 2019, 11(2): 81–86.
- [3] 胡 涛, 李卫华, 秦先祥. 图像语义分割方法综述 [J]. 测控技术, 2019, 38(7): 8–12.
HU Tao, LI Weihua, QIN Xianxiang. A Review on Image Semantic Segmentation[J]. Measurement & Control Technology, 2019, 38(7): 8–12.
- [4] LONG J, SHELHAMER E, DARRELL T. Fully Convolutional Networks for Semantic Segmentation[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston: IEEE, 2015: 3431–3440.
- [5] WANG W G, SHEN J B, SHAO L. Video Salient Object Detection via Fully Convolutional Networks[J]. IEEE Transactions on Image Processing, 2018, 27(1): 38–49.
- [6] LIU F, ZHOU Z Y, JANG H, et al. Deep Convolutional Neural Network and 3D Deformable Approach for Tissue Segmentation in Musculoskeletal Magnetic Resonance Imaging[J]. Magnetic Resonance in Medicine, 2018, 79(4): 2379–2391.
- [7] ZHAO X M, WU Y H, SONG G D, et al. A Deep Learning Model Integrating FCNNs and CRFS for Brain Tumor Segmentation[J]. Medical Image Analysis, 2018, 43: 98–111.
- [8] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional Networks for Biomedical Image Segmentation[C]//Medical Image Computing and Computer-Assisted Intervention(MICCAI 2015). Cham: Springer, 2015. DOI:10.1007/978-3-319-24574-4_28.
- [9] 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.
- [10] YU F, KOLTUN Y. Multi-Scale Context Aggregation by Dilated Convolutions[C/OL]//The 4th International Conference on Learning Representations(ICLR 2016). 2015: arXiv: 1804.03999[cs.CV]. <https://arxiv.org/pdf/1511.07122.pdf>.
- [11] GARCIA-GARCIA A, ORTS-ESCOLANO S, OPREA S, et al. A Review on Deep Learning Techniques Applied to Semantic Segmentation. [J/OL]. Transactions on Pattern Analysis and Machine Intelligence, 2017: arXiv: 1804.03999[cs.CV]. <https://arxiv.org/pdf/1704.06857v1.pdf>.
- [12] 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.
- [13] CHEN L C, ZHU Y K, PAPANDREOU G, et al. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation[J]//Computer Vision-ECCV 2018. Cham: Springer, 2018: 833–851.
- [14] ÇİÇEK Ö, ABDULKADIR A, LIENKAMP S S, et al. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation[C]//Medical Image Computing and Computer-Assisted Intervention (MICCAI 2016). Cham: Springer, 2016. DOI:10.1007/978-3-319-46723-8_49.
- [15] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-Excitation Networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011–2023.
- [16] YU F, KOLTUN V, FUNKHOUSER T. Dilated Residual Networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017: 636–644.
- [17] OKTAY O, SCHLEMPER J, FOLGOC L L, et al. Attention U-Net: Learning where to Look for the Pancreas[C/OL]//The International Conference on Medical Imaging with Deep Learning (MIDL 2018). 2018: arXiv: 1804.03999[cs.CV]. <https://arxiv.org/abs/1804.03999>.
- [18] TAHA A A, HANBURY A. Metrics for Evaluating 3D Medical Image Segmentation: Analysis, Selection, and Tool[J/OL]. BMC Medical Imaging, 2015, 15(1). <https://doi.org/10.1186/s12880-015-0068-x>.
- [19] 邓天民, 王 琳, 杨其芝, 等. 基于改进 SegNet 算法的车道线检测方法 [J]. 科学技术与工程, 2020, 20(36): 14988–14993.
DENG Tianmin, WANG Lin, YANG Qizhi, et al. Lane Line Detection Method Based on Improved SegNet Algorithm[J]. Science Technology and Engineering, 2020, 20(36): 14988–14993.

- [20] 唐贤伦, 杜一铭, 刘雨微, 等. 基于条件深度卷积生成对抗网络的图像识别方法 [J]. 自动化学报, 2018, 44(5): 855–864.
TANG Xianlun, DU Yiming, LIU Yuwei, et al. Image Recognition with Conditional Deep Convolutional Generative Adversarial Networks[J]. Acta Automatica Sinica, 2018, 44(5): 855–864.
- [21] ZHAO H S, SHI J P, QI X J, et al. Pyramid Scene Parsing Network[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017: 6230–6239.
- (责任编辑: 邓 彬)

Improved U-Net Network with Squeeze Excitation Dilated Convolution for Logistics Pallet Image Segmentation

WEI Zhanguo, SONG Yaping, LI Ya

(College of Mechanical & Electrical Engineering, Central South University of Forestry and Technology, Changsha 410004, China)

Abstract: Due to the various types of logistics pallets, complex shape rules, and the problems of pallets being blocked and changing light conditions in industrial production environment, a novel U-Net network based on squeeze excitation dilated convolution (SEDC) was proposed. By modeling the correlation between feature channels, the important features were strengthened and the segmentation performance of logistics pallet images was improved. Specifically, $1 \times 1 \times 1$ convolution in the SEDC module was used for data dimensionality reduction and dimensionality upgrade, which greatly reduced the amount of calculation, and image features were effectively explored under different fields of view through normal convolution and hole convolution with an expansion rate of 2, while automatically learning the importance of different layers through the SE module. Experimental results showed that compared with some existing classical image segmentation algorithms, the proposed model greatly reduced the computational burden and improved the robustness of the network while ensuring the performance of image segmentation as much as possible, and was expected to provide a new solution for intelligent segmentation of logistics pallet images.

Keywords: logistics pallet image; convolutional neural networks; SEDC