DOI:10.11830/ISSN.1000-5013.202209025

图像抠图与 copy-paste 结合的数据增强方法

杨天成,杨建红,陈伟鑫

(华侨大学 机电及自动化学院,福建 厦门 361021)

摘要: 提出一种基于图像抠图与 copy-paste 结合的数据增强方法(matting-paste),采用图像抠图法获取单个垃圾实例的准确轮廓,并对单个实例进行旋转和亮度变换.根据物体轮廓信息,把实例粘贴到背景图上,无需额外的人工标注即可生成新的带有标注的数据,从而提高数据集的多样性和复杂性.结果表明:数据集扩充后的 mask 比数据集扩充前的识别精度提高了 0.039,matting-paste 能在已有数据集上有效地扩充数据,进一步提高模型的识别精度.

关键词: 数据增强;图像抠图;copy-paste;实例分割

中图分类号: TP 274;TP 183 **文献标志码:** A **文章编号:** 1000-5013(2023)02-0243-07

Data Enhancement Method Combining Image Matting and Copy-Paste

YANG Tiancheng, YANG Jianhong, CHEN Weixin

(College of Mechanical Engineering and Automation, Huaqiao University, Xiamen 361021, China)

Abstract: A data enhancement method (matting-paste) based on image matting and copy-paste is proposed. Using the image matting method to obtain the precise contour of a single waste instance, and rotation and brightness transformation are carried out for each instance. Instances are pasted onto the background image according to the object's contour information, and new annotated data can be generated without additional manual annotation, which improves the diversity and complexity of the dataset. The results show that the recognition precision of mask after dataset augmentation is improved 0.039 compared with before dataset augment. Matting-paste can effectively augment the data and further improve the the recognition precision of the model.

Keywords: data enhancement; image matting; copy-paste; instance segmentation

随着城市化的发展和城市人口的不断增加,城市生活垃圾(MSW)的数量急剧增加,类型也变得复杂.有效的废弃物管理可以回收 MSW 中的可回收物,减少环境污染和资源浪费[1].传统的回收工作需要大量的人工劳动力成本[2].深度学习技术可应用于垃圾的自动识别和分类,提高回收效率[3-4].实例分割可以很好地应用于固体废物的识别和分类[5],作为一种监督算法,检测效果依赖于标注数据集的数量[6].数据集通常是手动标注的,标注数据集是一件耗时的工作,如标注 1 000 个 COCO 实例需 22 h[7].

收稿日期: 2022-09-25

通信作者: 杨建红(1974-),男,教授,博士,主要从事多模态视觉检测方法及系统开发、基于多平台的机器深度学习算法、高效率智能分选机器人的研究.E-mail:yjhong@hqu.edu.cn.

基金项目: 福建省科技重大专项(2020YZ017022);福建省厦门市科技计划项目(2021FCX012501190024);深圳市科技计划项目(JSGG20201103100601004)

生活垃圾的形状是多变的,手动标注垃圾的精确轮廓需要大量的人工成本.数据增强可扩展可训练的数据集[8],传统的数据增强方法针对的是整个图像,只是简单地增加数据集的数量,并没有增加数据集的复杂性,不是专门为实例分割设计的.

Copy-paste是一种适用于实例分割的数据增强方法[9].它的核心思想是从原始图像中复制实例,根据实例的标注轮廓将其粘贴到另一张图像中.该方法可以有效提高数据集的多样性,扩展可训练的数据集.因此,对于copy-paste数据增强方法,每个实例轮廓的准确性会影响数据增强的效果.生活垃圾的形状复杂多变,人工标注很难得到准确的轮廓.使用3D相机可以获得物体的精确轮廓,但会额外增加设备的硬件成本.基于深度学习的图像分割可以有效分割复杂背景中物体的轮廓[10],是一种低成本且有效的方法,常见的实例分割网络如Mask R-CNN[11]和Mask Transfiner[12].自然图像抠图是从图像中准确估计出目标前景,抠图生成的前景比实例分割网络获取的轮廓更自然细腻.常见的图像抠图网络有MODNet[13]和HAttMatting[14].基于此,本文提出图像抠图与copy-paste结合的数据增强方法.

1 材料和方法

1.1 数据集

为采集可回收垃圾的高质量的RGB图像,搭建图像采集平台(图1).图像采集平台包括彩色相机和发光二极管(LED)光源,成本低,不需要昂贵的高精度的3D相机.当输送带将可回收垃圾运送至相机下方拍摄时,将可回收垃圾的RGB图像截取,缩放至分辨率为1 400 px×728 px,以去除亮度不均匀的区域.输送带的有效宽度为1 400 mm,因此,一张图像中可能有多个垃圾实例.为了防止图像失真,保证分割效果的准确性,当图像输入到实例分割网络时,图像的上、下边缘用0像素填充至分辨率为1 400 px×1 400 px.

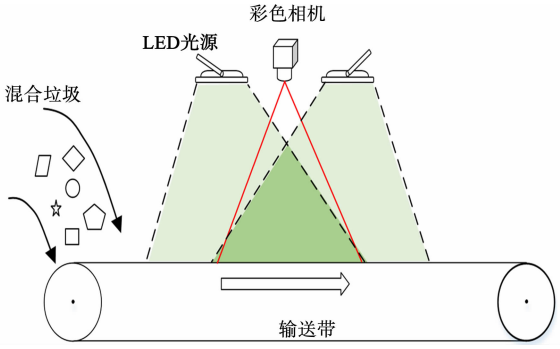

图1 图像采集平台

Fig. 1 Image acquisition platform

数据集分为利乐包、纸和纸杯3类.利乐包由纸、聚乙烯和铝[15]组成,而纸杯由纸和聚乙烯组成.由于数据集的成分和回收再生过程不同,需要进一步精细分类.数据集有2 274张图像,这些图像是在同一输送带上收集的.训练集由1 868张图像组成,物体稀疏放置,很少存在堆叠的情况.测试集有406张图像,物体密集放置,存在粘连堆叠的情况.数据集图片使用Labelme软件手工注释.

1.2 分割网络

1.2.1 Mask R-CNN Mask R-CNN是基于R-CNN的实例分割模型,在COCO实例分割任务中均优于以往的网络,它不仅可以检测出图像中实例种类的位置,还能为每个实例生成分割掩膜.Mask R-CNN结构,如图2所示.图2中:C2~C5为物体的低、高层特征;P2~P6为特征图;FPN为特征金字塔;RPN为区域提取网络;ROI层为目标层.

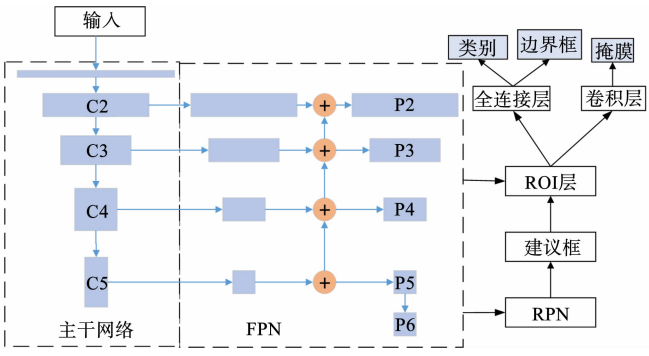

图2 Mask R-CNN结构

Fig. 2 Structure of Mask R-CNN

由图2可知:输入的图片首先在主干网络中经过多次卷积、池化操作后,图片分辨率逐渐减小,维度不断增加,从而提取到物体的深层特征;其次,使用FPN[16]进行特征融合;通过上采样和下采样,融合图

片的高层和低层特征，并得到特征图；RPN 生成一系列建议框，建议框代表着特征图上的一处矩形区域，矩形区域中可能包含有物体，也可能不包含物体；ROI 层根据建议框的位置，从特征图上截取相应的矩形区域并缩放到固定大小，然后传递给全连接层，对物体进行分类和边界框回归，得到物体的类别和位置；Mask R-CNN 在 ROI 层之后添加了卷积层，用于计算物体的二进制掩膜以分割物体的轮廓.

1.2.2　Mask Transfiner　这是一种优质高效的实例分割算法. 与现有方法不同，Mask Transfiner 不会统一处理整张图像，其识别容易出错并需要优化的像素区域（信息损失区域）. 这些像素区域点采用四叉树结构表示，并根据下采样物体掩膜的信息损失计算得到，主要分布在物体的边界或高频区域中，空间上不连续. 基于 Mask Transfiner[17]，四叉树结构只处理检测到的易出错的树节点，同时进行自校正. 由于信息损失区域的位置稀疏，仅占图像总像素的一小部分，这允许 Mask Transfiner 以较低的计算成本预测出高度准确的实例掩膜.

1.2.3　图像抠图　图像抠图是指图像和视频中准确的前景估计问题[18]. 抠图算法被应用于图像编辑和影片剪辑，可以精确地将图像或视频中的前景估计出来. 图像抠图的目的是从给定图像（I）中提取所需的前景（F）. 预测每个像素（i）具有精确前景概率 α 的 alpha 蒙版，即

$$I^i=\alpha^i F^i+(1-\alpha^i)B^i. \tag{1}$$

式(1)中：B 是 I 的背景.

图像抠图是具有挑战性的，因为式(1)右侧的所有变量都是未知的. 现有的抠图方法分为两类：一类是使用预定义的 trimap 图作为辅助输入；另一类是在不输入 trimap 图的情况下完成抠图. trimap 图有绝对前景（$\alpha=1.0$）、绝对背景（$\alpha=0$）和未知区域（$\alpha=0.5$）3 个区域的掩码. 由于创建 trimap 图会增加额外的工作量，因此，使用了 trimap-free 抠图方法. MODNet 是一种 trimap-free 的抠图网络，无需额外输入即可实现发丝级的人像抠图. MODNet 的抠图效果，如图 3 所示. 由图 3 可知：输入原始图像后，MODNet 会生成一张 alpha 图，alpha 图的白色部分代表前景，黑色部分代表背景.

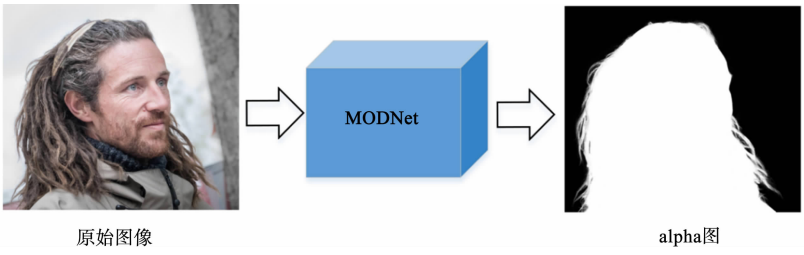

图 3　MODNet 的抠图效果

Fig. 3　Segmentation effect of MODNet

2　实验方法

2.1　实例轮廓

获取实例轮廓的流程，如图 4 所示. 图 4 中：从数据集图片中截取实例时，忽略有图片边缘的物体，

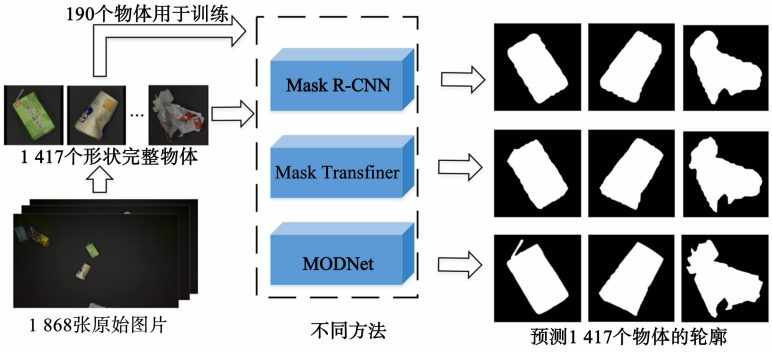

图 4　获取实例轮廓的流程

Fig. 4　Process of obtaining object's contour

以及存在粘连堆叠的物体,保证每个截取的物体都是独立且完整的;截取的图像使用 padding 方法进行填充,填充的形状为正方形,防止在输入分割模型时因缩放造成长宽比失真. 由于分割模型的初始权重不是专门表示分割垃圾图像的,因此,需要使用少量的垃圾图片样本进行迁移学习,使分割模型具有分割出垃圾轮廓的能力.

由图 4 可知:从 1 868 张原始图片中截取了 1 417 个形状完整物体,手工标注其中 190 个物体的准确轮廓,用于训练 MaskR-CNN,Mask Transfiner 和 MODNet;再使用训练好的分割模型分别预测截取的 1 417 个物体的轮廓(alpha 图);最后,使用 opencv 的轮廓算法,把物体轮廓转换成点集写入 json 文件,实现物体轮廓的自动标注.

不同方法的轮廓标注,如图 5 所示. 由图 5 可知:MODNet 获取的物体轮廓最准确,能得到利乐包的吸管部分和扭曲变形的纸准确的轮廓;Mask Transfiner 的分割效果比 Mask R-CNN 更好一些.

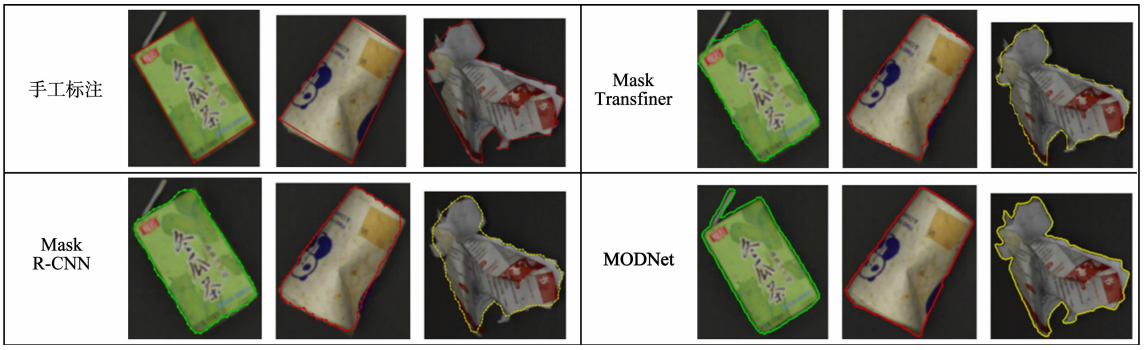

图 5 不同方法的轮廓标注

Fig. 5 Annotation contours with different methods

2.2 数据集的扩充

Matting-paste 可以根据已有标注的数据集,自动扩充带标注的数据集,从而提高数据集的多样性和复杂性. 生成图像数据的流程,如图 6 所示.

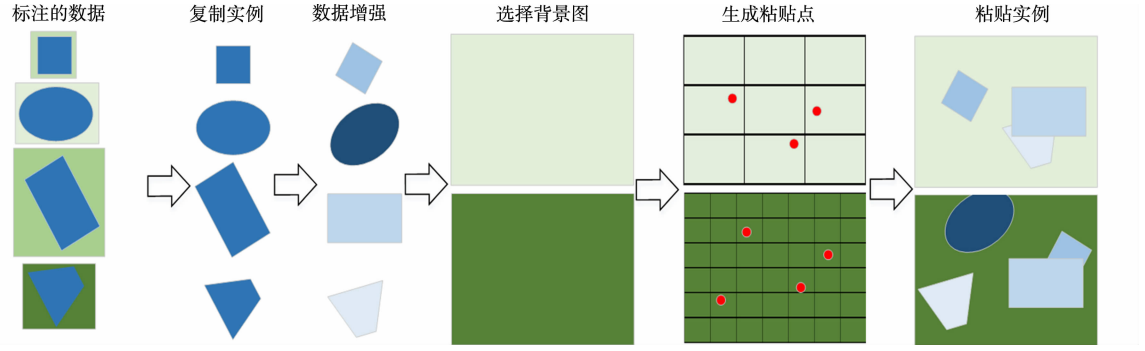

图 6 生成图像数据的流程

Fig. 6 Process of generating image data

1) 复制实例. 根据标注数据集的 json 文件中物体的轮廓信息,把轮廓内的像素抠下来. 如果标注的物体轮廓不准确,则轮廓内的像素可能会包含输送带的背景,或者轮廓没有完全包含物体.

2) 数据增强. 对每个实例进行旋转和亮度变换,以提高数据集的多样性. 因为相机的拍摄视野和垃圾的尺寸是固定的,因此,没有使用 copy-paste 中的大尺度抖动来改变实例的尺寸大小.

3) 选择背景图. 原始数据集虽然是在同一条输送带采集的,但是由于输送带不同区域受污染的程度不同,因此,在选择背景时从多张背景中随机选取一张,使生成的数据集更符合实际工况.

4) 生成粘贴点. 为保证每个实例粘贴的随机性,对每张背景图片随机划分为 $m\times n$ 的网格,其中,m 为行数,$3\leqslant m\leqslant 6$;n 为列数,$3\leqslant n\leqslant 8$. 随机生成 k 个粘贴点,当 $m\times n<15$ 时,$0\leqslant k\leqslant m\times n$,当 $m\times n\geqslant 15$ 时,$0\leqslant k\leqslant 15$. 粘贴点相对网格的中心位置发生 x 和 y 方向上的随机偏移,并限制每个网格最多只能有一个粘贴点.

5）粘贴实例. 根据生成的粘贴点，随机选择数据增强后的实例粘贴到背景图片上. 每个实例的标注轮廓根据 json 文件的点集进行 x 和 y 方向偏移，实例的类别标签使用 json 文件中的原始标签. 物体堆叠的情况，粘贴后实例会覆盖之前的实例轮廓，轮廓超过图片边界的部分会被截掉.

2.3　实验设置

为了验证数据扩充方法的有效性，对比不同数据扩充方法对识别精度的影响. 使用 MaskR-CNN 识别精度，以 ResNet 50[19] 作为主干网络，使用 FPN 融合多尺度特征，主干网络使用 ImageNet 数据集上预训练的权重进行迁移学习.

实验的深度学习框架为 pytorch 1.9；环境为 Python 3.7；设备的操作系统为 Windows 10 专业版；CPU 为 Intel i5-10400F；GPU 为 Nvidia RTX3090；内存为 16 GB. 每个模型训练 24 个 epoch，学习率为 0.001 25，batch size 为 2，前 1 000 步迭代执行学习率线性预热方法，增长率为 0.001.

采用均值平均精度（P_{mA}）来综合评估模型的性能，采用精度表示在所有被预测为正样本中实际为正样本的概率，采用召回率表示在所有实际为正样本中被预测为正样本的概率. 在一定的交并比（IOU）阈值下，利用不同精度和召回率的组合，可以得到一个特定类的平均精度（P_{A}），即

$$P=\frac{n_{\mathrm{TP}}}{n_{\mathrm{TP}}+n_{\mathrm{FP}}}, \tag{3}$$

$$r=\frac{n_{\mathrm{TP}}}{n_{\mathrm{TP}}+n_{\mathrm{FN}}}, \tag{4}$$

$$P_{\mathrm{A}}=\int_0^1 P(r)\mathrm{d}r, \tag{5}$$

$$P_{\mathrm{mA}}=\frac{1}{m}\sum_{i=1}^{m}P_{\mathrm{A},i}. \tag{6}$$

式(3)～(6)中：P 为识别精度；n_{TP} 为垃圾种类被正确识别的个数；n_{FN} 为垃圾被错误识别为背景的个数；n_{FP} 为背景被识别为垃圾的个数；r 为召回率；m 为类别数.

3　实验结果与分析

3.1　生成的图片效果

实际图片与生成图片的比较，如图 7 所示. 通过设置更多的粘贴点，可以使生成的图片中有更多的垃圾实例. Copy-paste 方法使用的物体轮廓是手工标注的，手工标注的轮廓可能包含传送带背景，物体存在堆叠时的效果不真实. 手工标注也可能丢失物体的部分轮廓，并且物体边角部分的轮廓不平滑，这导致 copy-paste 生成的图片不自然. 使用图像抠图方法获取的轮廓很准确，可以保留更多物体的原始轮廓信息，如利乐包的吸管部分. 因此，matting-paste 生成的图片更真实，物体堆叠时的效果和真实的图片很相似.

(a) 实际图片

(b) 手工标注轮廓生成图片

(c) 图像抠图轮廓生成图片

图 7　实际图片与生成图片的比较

Fig. 7　Comparison of actual image and generated image

3.2　识别精度结果

自动生成的数据集与原始的训练集混合，可以得到扩充后的训练集. 原始的训练集记为 $\mathrm{D_{OR}}$；原始训练集的物体轮廓是手工注释的，扩充后的训练集记为 $\mathrm{D_{CP}}$；使用 Mask R-CNN 的标注轮廓扩充后的训练集，记为 $\mathrm{D_{MR}}$；使用 Mask Transfiner 的标注轮廓扩充后的训练集，记为 $\mathrm{D_{MT}}$；使用 MODNet 的标注轮

廓扩充的训练集,记为 D_{MA}.

为了验证数据集扩充的有效性,确定最优的数据集扩充数量.先使用 matting-paste 的方法扩充训练集至 3 000 张,每次递增 1 000 张,并与原始 1 868 张训练集作为对比.不同训练集的识别精度,如表 1 所示.表 1 中:mask 为掩膜;box 为边界框,IOU 阈值为[0.50:0.95].由表 1 可知:扩充训练集之后,模型的精度明显提高;在数据集扩充至 6 000 张后,模型的识别精度达到最高,mask 的识别精度达到 0.692,比数据扩充前提高了 0.039,box 的识别精度达到 0.642,比数据扩充前提高了 0.028.因此,matting-paste 方法扩充数据集是有效的.

表 1 不同训练集的识别精度

Tab.1 Recognition precision of different trainsets

训练集数量	P(mask)	P(box)
1 868	0.653	0.614
3 000	0.681	0.628
4 000	0.686	0.637
5 000	0.683	0.638
6 000	0.692	0.642
7 000	0.688	0.627

为比较不同轮廓获取方法对数据集扩充后模型识别精度的影响,分别把训练集扩充到 6 000 张,对比不同数据集的识别精度(表 2).由表 2 可知:D_{CP} 相比于原始的训练集精度有所提高;使用分割网络获取的轮廓比手工标注的轮廓更准确,且 D_{MR},D_{MT},D_{MA} 的精度都比 D_{CP} 高,说明在数据扩充时,单个实例轮廓标注的准确性影响着模型的精度,轮廓的识别精度提高了 0.024,mask(0.50 以上)的识别精度提高了 0.014,box([0.50:0.95])的识别精度提高了 0.017,box(0.50 以上)提高了 0.011,数据增强效果优于 D_{CP}.

表 2 不同数据集的识别精度

Tab.2 Recognition precision of different datasets

数据集	P(mask)		P(box)	
	[0.50:0.95]	0.50 以上	[0.50:0.95]	0.50 以上
D_{OR}	0.653	0.841	0.614	0.843
D_{CP}	0.668	0.847	0.625	0.850
D_{MR}	0.677	0.849	0.628	0.850
D_{MT}	0.683	0.850	0.637	0.850
D_{MA}	0.692	0.861	0.642	0.861

Matting-paste 数据集扩充方法可以对数据集中的物体重新进行排列组合,自动生成带标注数据集图片以扩充训练集,从而提高模型的识别精度.当数据集扩充至 6 000 张时,模型的精度提升效果最优.相比于手工标注和 Mask R-CNN,Mask Transfiner 获取的轮廓,matting-paste 方法获取的物体轮廓最准确,扩充的数据集质量最好.

4 结束语

提出一种 matting-paste 的数据集扩充方法,首先,使用抠图方法获取单个垃圾实例的准确轮廓,并对单个实例进行旋转和亮度变换,以增加数据的多样性.其次,根据物体轮廓信息把实例粘贴到背景图上.此方法无需额外的人工标注,即可自动生成新的带有标注的数据集用于训练,从而提高训练集的多样性和复杂性.数据集扩充后的模型精度比扩充前模型的精度提高了 0.039.该方法可以应用于垃圾的目标检测和实例分割等分类任务中,在已有数据集上进一步扩充数据集,提高模型的识别精度.

参考文献:

[1] GUNDUPALLI S P, HAIT S, THAKUR A. A review on automated sorting of source-separated municipal solid waste for recycling[J]. Waste Management, 2017, 60: 56-74. DOI: 10.1016/j.wasman.2016.09.015.

[2] SEIKE T, ISOBE T, HARADA Y, *et al*. Analysis of the efficacy and feasibility of recycling PVC sashes in Japan[J]. Resources, Conservation and Recycling, 2018, 131: 41-53. DOI: 10.1016/j.resconrec.2017.12.003.

[3] ZHANG Qiang, YANG Qifan, ZHANG Xujuan, *et al*. A multi-label waste detection model based on transfer learning [J]. Resources, Conservation and Recycling, 2022, 181: 106235. DOI: 10.1016/j.resconrec.2022.106235.

[4] SOUSA J, REBELO A, CARDOSO J S. Automation of waste sorting with deep learning[C]// 2019 XV Workshop de Visão Computacional. Brazi: IEEE Press, 2019: 43-48. DOI: 10.1109/WVC.2019.8876924.

[5] LI Jiantao, FANG Huaiying, FAN Lulu, *et al*. RGB-D fusion models for construction and demolition waste detection

[J]. Waste Management,2022,139:96-104. DOI:10.1016/j.wasman.2021.12.021.

[6] HAFIZ A M,BHAT G M. A survey on instance segmentation: State of the art[J]. International Journal of multimedia Information Retrieval,2020,9(3):171-189. DOI:10.1007/s13735-020-00195-x.

[7] LIN T Y,MAIRE M,BELONGIE S,*et al*. Microsoft coco: Common objects in context[C]// European Conference on Computer vision. [S. l.]:Springer,2014:740-755. DOI:10.1007/978-3-319-10602-1_48.

[8] SHORTEN C,KHOSHGOFTAAR T M. A survey on image data augmentation for deep learning[J]. Journal of Big Data,2019,6(1):1-48. DOI:10.1186/s40537-019-0197-0.

[9] GHIASI G,CUI Yin,SRINIVAS A,*et al*. Simple copy-paste is a strong data augmentation method for instance segmentation[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S. l.]: IEEE Press,2021:2918-2928.

[10] MINAEE S,BOYKOV Y Y,PORIKLI F,*et al*. Image segmentation using deep learning: A survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2020,44(7):3523-3542. DOI:10.1109/TPAMI.2021.3059968.

[11] HE Kaiming,GKIOXARI G,DOLLÁR P,*et al*. Mask r-cnn[C]// Proceedings of the IEEE International Conference on Computer Vision. Piscataway:IEEE Press,2017:2961-2969. DOI:10.48550/arXiv.1703.06870.

[12] KE Lei,DANELLJAN M,LI Xia,*et al*. Mask transfiner for high-quality instance segmentation[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Orleans:IEEE Press,2022:4412-4421. DOI:10.48550/arXiv.2111.13673.

[13] KE Zhanghan,SUN Jiayu,LI Kaican,*et al*. Modnet: Real-time trimap-free portrait matting via objective decomposition[C]// Proceedings of the AAAI Conference on Artificial Intelligence. [S. l.]:AAAI Press,2022,36(1):1140-1147. DOI:10.1609/aaai.v36i1.19999.

[14] QIAO Yu, LIU Yuhao, YANG Xin, *et al*. Attention-guided hierarchical structure aggregation for image matting [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S. l.]: IEEE Press,2020:13676-13685.

[15] MA Yuhui. Changing tetra pak: From waste to resource[J]. Science Progress,2018,101(2):161-170. DOI:10.3184/003685018X15215434299329.

[16] LIN T Y,DOLLÁR P,GIRSHICK R,*et al*. Feature pyramid networks for object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu:IEEE Press,2017:2117-2125.

[17] VASWANI A,SHAZEER N,PARMAR N,*et al*. Attention is all you need[J]. Advances in Neural Information Processing Systems,2017,30:1-15.

[18] WANG Jue,COHEN M F. Image and video matting: A survey[J]. Foundations and Trends in Computer Graphics and Vision,2008,3(2):97-175. DOI:10.1561/0600000019.

[19] HE Kaiming,ZHANG Xiangyu,REN Shaoqing,*et al*. Deep residual learning for image recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas:IEEE Press,2016:770-778.

（责任编辑：陈志贤　　英文审校：吴逢铁）