

引用本文:卢昱杰,刘 博,刘金杉,等.面向施工机械的深度学习图像数据集合成方法[J].建筑科学与工程学报,2022,39(4):100-107.  
LU Yu-jie, LIU Bo, LIU Jin-shan, et al. Image Dataset Synthetic Method for Construction Machinery Based on Deep Learning[J]. Journal of  
Architecture and Civil Engineering, 2022, 39(4): 100-107.

DOI:10.19815/j.jace.2021.07146

# 面向施工机械的深度学习图像数据集合成方法

卢昱杰<sup>1,2,3</sup>, 刘 博<sup>1</sup>, 刘金杉<sup>1</sup>, 赵宪忠<sup>1</sup>

(1. 同济大学 建筑工程系, 上海 200092; 2. 同济大学 工程结构性能演化与控制教育部重点实验室,  
上海 200092; 3. 同济大学 上海智能科学与技术研究院, 上海 200092)

**摘要:**工程现场环境复杂,获取包含丰富信息的图像难度大且标注成本高,造成基于计算机视觉的深度学习施工机械图像数据集构建困难。为满足快速、高质量构建建筑工程领域施工机械深度学习图像数据集,提出一种基于三维建模引擎的施工机械图像生成与自动标注方法,并以挖掘机为例构建了名为 SCED(Synthesized Construction Equipment Dataset)的挖掘机数据集。首先,采用三维建模引擎 UE4 对目标挖掘机设备进行模型构建,然后借助 UnrealCV 工具对原始模型进行多角度、多区域的图像采集,使用自编写模块实现自动语义分割与掩码图像生成,并完成图像的自动标注,最终生成包含 10 000 张图像的数据集。与现有公开机械数据集进行了目标尺寸、数量与构建工作量的对比,并比较了构建效率与成本,最后进行了图像数据集质量与效果验证。结果表明:该构建方法综合效率更高且成本更低,构建的 SCED 图像数据集丰富性和泛化能力更好,针对小目标物具有更好的检测效果;研究成果可为今后建筑施工领域深度学习图像数据集的构建提供参考依据。

**关键词:**计算机视觉;深度学习;建筑施工;施工机械;合成数据集

中图分类号:TU973.2

文献标志码:A

文章编号:1673-2049(2022)04-0100-08

## Image Dataset Synthetic Method for Construction Machinery Based on Deep Learning

LU Yu-jie<sup>1,2,3</sup>, LIU Bo<sup>1</sup>, LIU Jin-shan<sup>1</sup>, ZHAO Xian-zhong<sup>1</sup>

(1. Department of Building Engineering, Tongji University, Shanghai 200092, China; 2. Key Laboratory of Performance Evolution and Control for Engineering Structures of Ministry of Education, Tongji University, Shanghai 200092, China; 3. Shanghai Institute of Intelligent Science and Technology, Tongji University, Shanghai 200092, China)

**Abstract:** The complex environment of construction sites, the difficulty of obtaining images containing rich information and the high cost of image annotation make it difficult to build deep learning construction machinery image datasets based on computer vision. In order to build a fast and high-quality deep learning image dataset of construction machinery in the construction field, a method of generating and automatically annotating images of construction machinery based on a 3D modeling engine was proposed, and an excavator dataset named SCED (Synthesized

收稿日期:2021-07-03

基金项目:国家自然科学基金项目(52078374)

作者简介:卢昱杰(1985-),男,贵州贵阳人,教授,博士研究生导师,工学博士,E-mail:Lu6@tongji.edu.cn。

通信作者:刘金杉(1995-),男,河南洛阳人,工学硕士研究生,E-mail:supernova\_ks@outlook.com。

Construction Equipment Dataset) was constructed with an excavator as an illustration. Firstly, the target excavator equipment was modelled using the 3D modelling engine UE4, then the original model was captured from multiple angles and regions utilizing UnrealCV tool, automatic semantic segmentation and mask image generation were achieved using a self-written module, and the bounding boxes of excavator instance were automatically annotated, generating a dataset containing 10 000 images. The target size, number and construction effort were compared with existing publicly available mechanical datasets, and the construction efficiency and cost were also compared, and finally the quality and effectiveness of the image dataset were verified. The results show that the construction method is more efficient and less expensive, and the SCED image dataset is richer and better generalised, with better detection results for small targets. The research findings can provide a reference for the construction of deep learning image datasets in the field of building construction in the future.

**Key words:** computer vision; deep learning; building construction; construction machinery; synthetic dataset

## 0 引言

随着计算机硬件技术的快速发展以及算力的急速提升,基于计算机视觉原理和深度学习方法的目标检测技术已经被应用于各行各业,尤其在土木工程的建筑工程领域,深度学习与计算机视觉技术的结合已经为实际生产带来了诸多便利与革新,如基于目标检测方法的施工工人安全帽的检测<sup>[1]</sup>、基于图像尺寸识别的裂缝长度与宽度的计算<sup>[2]</sup>、基于语义分割的渗漏水识别与判别<sup>[3-5]</sup>以及基于多摄像设备的桥梁挠度计算与预警<sup>[6]</sup>等。

为更好运用目标识别与检测技术,专家与学者更多专注于提升检测效果的研究。提升基于深度学习的目标检测方法效果的要点包括深度学习模型算法的优化和图像数据集的快速、高质量构建。针对深度学习模型算法优化,国内外学者已经进行了大量研究<sup>[7-8]</sup>,VGG<sup>[9]</sup>、R-CNN<sup>[10]</sup>、ResNet<sup>[11]</sup>及 Det-Net<sup>[12]</sup>等从特征提取器角度进行了迭代与创新, IoU-Net 与 UnitBox 从 IoU 算法方面进行了优化改进, Focal Loss、Skrinkage Loss 及 Repulsion Loss 对 Loss 算法进行了改进提升,从基本原理上提升了目标检测的效率、精度及准确率等。但在快速、有效且高质量构建图像数据集方面,相关研究则刚刚起步。An 等<sup>[7]</sup>提出了基于图像合成的数据集生成概念,通过建模软件与实体扫描的方式获取三维模型,后根据场景进行图像的采集,该方法首次实现了建筑施工图像数据集的非实际场景生成,开创了合成图像数据集的先河,但复杂的建模过程和单一的图像背景限制了其使用范围。

在建筑施工场景下,由于环境条件所限,多采用人工拍摄、监控视频截取等传统方式对现场图像进行采集,难以保证构建图像数据集的高效、低成本以及图像多样性。在施工机械图像数据集构建方面,常出现覆盖机械种类不全、图像视角不均衡和光照等天气条件不充分等问题<sup>[13]</sup>。鉴于此,本文提出一种面向建筑施工机械的深度学习图像数据集构建方法,本质是通过建立的三维模型与多种场景环境进行组合,利用自动化多方位取景算法快速高效且低成本的收集多样化图像并完成掩码图的生成,实现施工机械图像数据集的合成与构建。

## 1 图像数据集构建方法

本文提出的基于三维建模的建筑施工机械图像数据集构建方法主要包含 2 个步骤:①目标机械三维模型的建立与多视角下机械图像的采集与掩码图像生成;②多环境下背景图像与目标机械的融合以及自动标注文件的生成。

完成图像数据集的构建后,基于常用目标检测模型与框架与现有公开的清华 MOCS<sup>[7]</sup>、阿尔伯塔 ACID<sup>[14]</sup>数据集进行数据集多维度性能指标和检测效果的对比。具体技术路线与流程如图 1 所示。

### 1.1 三维建模方法及建模依据

主流商用三维建模与动作实现软件 UE4 和 U3D 各具特点。本文选用具有良好插件兼容性与自编代码植入性的 UE4 引擎(图 2)与 UnrealCV<sup>[15]</sup>组件作为建模工具,通过自建模型与第三方模型导入相结合的方法实现多种建筑施工机械模型的建立。

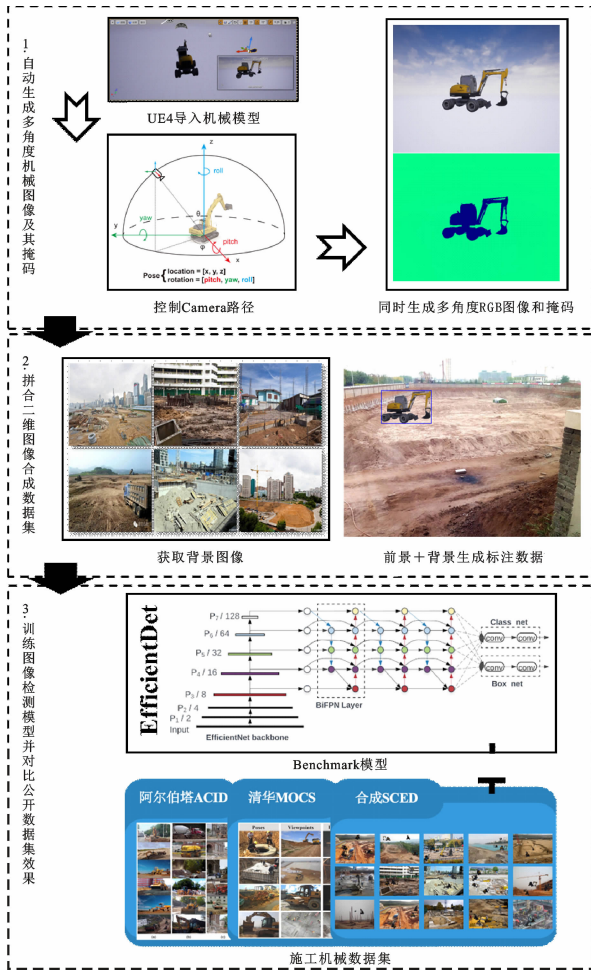


图 1 技术路线与流程

Fig. 1 Technical Route and Processes

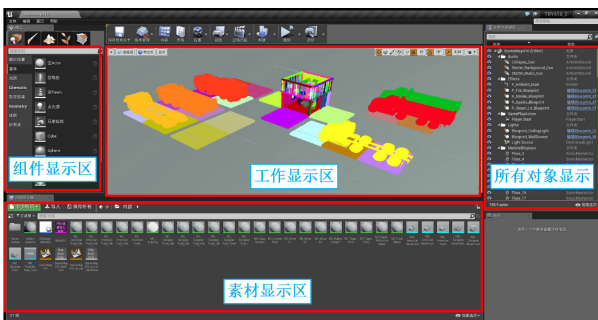


图 2 UE4 建模引擎界面与功能

Fig. 2 Interface and Features of UE4 Modeling Engine

## 1.2 模型建立与图像生成

采用 UE4 自行建立场景模型并使用 Twinmotion 商用施工机械模型进行多环境的三维模型建立。通过 UnrealCV 组件与自编代码实现针对目标机械模型的多角度环绕取景存图以及基于规划的场景线路进行多视角取景存图,原理如图 3 所示。其中多视角取图方式是通过 Camera 组件 360°环绕目标机械并每隔 10°取景 1 张实现,取景过程涉及到球

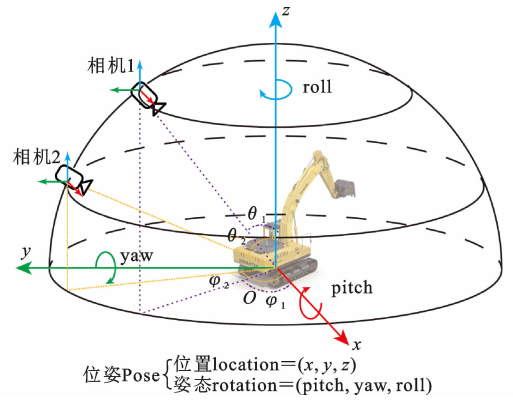


图 3 多视角取景原理

Fig. 3 Principle of Multi-view Framing

面坐标与笛卡儿坐标系的转换,方法为:

(1)假设三维机械模型中心位置坐标为  $(x_0, y_0, z_0)$ ,建立一个三维坐标系,预设机械模型姿态  $(pitch, yaw, roll)$  的初始值均为 0,  $\theta$  为取景组件 Camera 与机械模型中心连线与  $z$  轴正向夹角,  $\varphi$  为与  $x$  轴正向的夹角。

(2)令取景组件 Camera 的球面坐标的姿态为  $(x, y, z, pitch, yaw, roll)$ ,则坐标转换满足以下条件:

$$\begin{cases} x = r \sin(\theta) \cos(\varphi) + x_0 \\ y = r \sin(\theta) \sin(\varphi) + y_0 \\ z = r \cos(\theta) + z_0 \\ pitch = 0^\circ \\ yaw = \theta - 90^\circ (5^\circ < \theta < 50^\circ) \\ roll = 180^\circ + \varphi (0^\circ < \varphi < 360^\circ) \end{cases}$$

式中:  $\theta$  每隔 5°取一个值(共 10 组);  $\varphi$  每 5°取一个值(共 72 组),每一个独立的目标机械将从 720 个视角进行取景。

图像数据集主要考虑机械姿态和图像视角 2 个因素,部分图像效果展示见表 1。






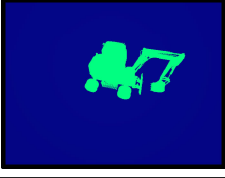

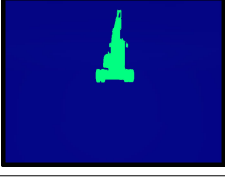




## 1.3 图像的自动标注

深度学习图像数据集标注的精确度是决定数据集质量的一个重要因素<sup>[16]</sup>。传统方法是通过人工点击或框选的方式对真实图像进行标注,这一过程费时费力且无法保证标注标准的统一性<sup>[17]</sup>。自动且规范的标注方法将极大提升工作效率与图像数据集的质量。

本文利用 UnrealCV 与自编代码配合 UE4 模型本身拥有的对象属性,实现了图像实例分割的掩码图生成。通过百度、谷歌、Bing 等网络搜索引擎,进行不同背景图像的检索与保存。为了不引入多余噪声与干扰,造成假阴性升高、检出率下降,搜集图

表 1 图像输出效果展示

Table 1 Image Output Effect Display

原始图像	掩码图像	姿态角
		$\theta=50^\circ$ $\varphi=310^\circ$
		$\theta=20^\circ$ $\varphi=340^\circ$
		$\theta=50^\circ$ $\varphi=15^\circ$
		$\theta=15^\circ$ $\varphi=85^\circ$
		$\theta=30^\circ$ $\varphi=115^\circ$
		$\theta=45^\circ$ $\varphi=175^\circ$

像均不包含挖掘机。利用生成的掩码图像分离出挖掘机作为前景,并随机选择 4 种数据增强方法(机械大小随机缩放 0.1 倍~1.2 倍、RGB 通道变换、随机水平翻转和随机亮度增强)中的 1 种或 2 种,与背景图像随机中心点且同 mask 包围框大小的图像进行随机组合,随机重复 1~3 次,得到一张具有包围框标注的合成挖掘机图像(图 4)。

为了保证图像数据集前景与背景亮度的一致性,消除合成数据的虚假感,本文在包围框中前景图像加入双轴模糊,具体做法为:先获取原始图像背景的平均亮度值,然后以包围框中心前景亮度值为峰值亮度,以包围框边缘中心亮度值为最低亮度值,在包围框纵横轴分别叠加高斯模糊,合成后的图像如图 5 所示。

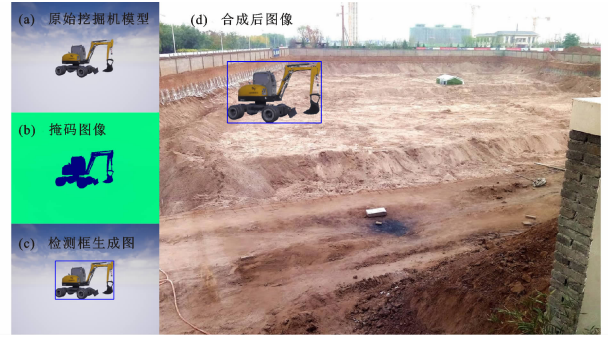


图 4 目标机械的掩码图像与图像合成

Fig. 4 Masked Image and Image Synthesis of Target Machinery



图 5 图像合成示例

Fig. 5 Examples of Image Synthesis

## 2 数据集的性能指标

为充分展现图像合成方法制作数据集的优越性与有效性,首先进行深度学习训练结果的对比。选用清华大学 MOCS 数据集的 EfficientDet-D5<sup>[18]</sup>模型的目标检测作为基准训练,模型输入尺寸为  $512 \times 512$ ,并利用水平翻转和标准化对数据进行增强处理,使用 COCO 预训练权重,训练参数 batch-size 为 16,学习率 0.001,分别训练 3 个数据集 60 epochs,使用 2 块 3090 显卡,共计训练 186 h,精度对比结果见表 2。EfficientDet 模型深度学习的训练损失曲线见图 6。

从表 2 可以看出,采用本文方法建立的数据集相比 ACID 与 MOCS 数据集在多个指标上均呈现出更好的精度,也表明 SCED 数据集质量优良。

从一个施工现场视频中分别截取第 0、21、189 帧图像(图 7),其中蓝色框为检测出的目标,绿色框为漏检目标,红色框为误检项目,可见 SCED 的泛化性能良好。

研究表明,图像中被标注物体的像素尺寸与被标注物体所处位置将影响基于深度学习的目标检测方法的泛化性能<sup>[19]</sup>。首先将清华 MOCS 数据集与阿尔伯塔 ACID 数据集进行统一清洗,处理为单一类别,将本文建立数据集 SCED 生成单一类别的标注数据,按照 EfficientDet 模型所需数据集的格式

表 2 各图像数据集精度对比

Table 2 Accuracy Comparison of Image Datasets

模型	数据集	AP0.5	AP0.5 : 0.95	AP0.5 : 0.95 (小目标尺寸)	AP0.5 : 0.95 (中目标尺寸)	AP0.5 : 0.95 (大目标尺寸)
EfficientDet-D5	ACID	0.918	0.767	0.368	0.579	0.791
	MOCS	0.783	0.538	0.069	0.361	0.611
	SCED	0.967	0.889	0.384	0.855	0.913

注:因 ACID 数据集笔者未提供 7 000 张图像训练集之外的验证集,本试验结果中 ACID 数据集存在过拟合倾向;AP0.5 为训练结果判别的重要指标,表示 IoU 阈值为 50% 时的检测精度;AP0.5 : 0.95 表示 IoU 阈值从 50% 至 95%,每 5% 取一个值的总平均值。

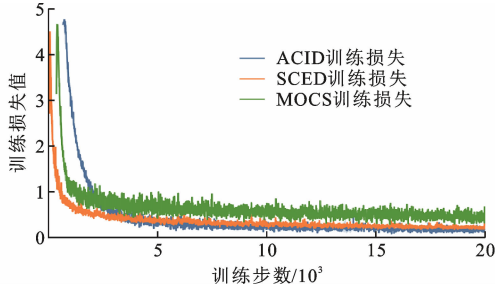


图 6 EfficientDet 模型训练损失曲线

Fig. 6 Training Loss Curves of EfficientDet Model

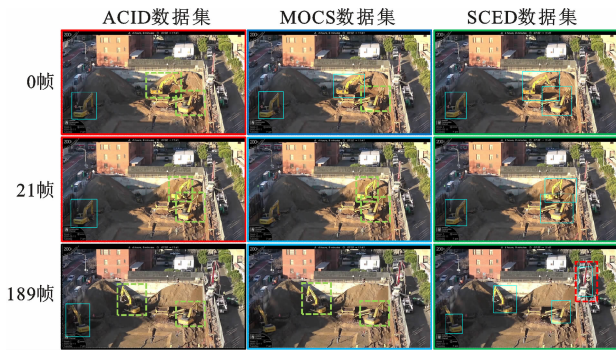


图 7 目标检测效果对比

Fig. 7 Target Detection Effect Comparison

(COCO 格式,训练集标注在一个 json 文件,标注框尺寸与左上角坐标均为绝对像素坐标)进行制作。

采用 COCO 的像素标准对 3 种数据集的目标尺寸进行统计,即小目标尺寸小于  $32 \times 32$ ,大目标尺寸大于  $96 \times 96$ ,中等目标介于两者之间,统计结果见图 8。

通常情况下深度学习图像数据集单张图像中会出现不止一个特征目标物,正确、全面地将每一张图像中每一个目标物进行标记对图像数据集的质量有正面效果,可以得到更优的训练结果。也正由于这个原因,SCED 图像数据集在创建过程就充分考虑了这一“多目标”特性,目标物数量远大于图像数量。挖掘机单类不同尺寸数量统计情况如表 3 所示。需要特别说明的是,ACID(7 000)指的是 ACID 数据集仅将 7 000 张图片对外公开。下文将在数据集英文名称后以括号形式表示此处使用某一数据集的图

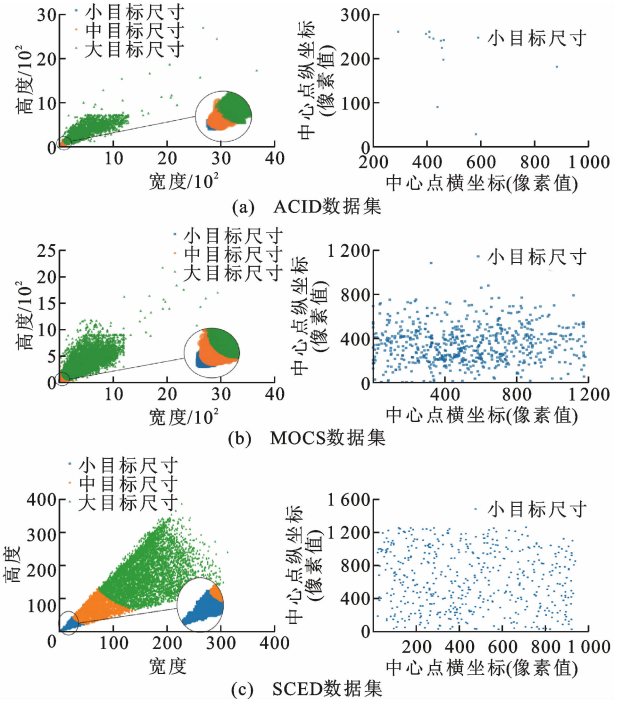


图 8 现有数据集的目标像素尺寸与位置分布

Fig. 8 Target Pixel Size and Location Distribution in Existing Datasets

表 3 挖掘机单类不同尺寸数量统计

Table 3 Quantity Statics of Excavators with Different Sizes

数据集	挖掘机实例数			总数
	包围框大小			
	小目标尺寸	中目标尺寸	大目标尺寸	
ACID(7 000)	12	180	1 750	1 942
MOCS	692	3 192	8 173	12 057
SCED	1 187	4 132	26 235	31 554

像数。

本文生成的图像数据集的目标尺寸分布均匀且多样(小目标比例较高),目标物在图像位置分布规律均匀,这一优良特性将大大提升图像数据集的质量,进而提升基于此图像数据集训练结果的精度。

### 3 合成图像数据集效果验证与对比

#### 3.1 数据准备与训练

本文制作的挖掘机数据集 SCED 共 10 000 张

图像,进行了标注文件的生成,并与清华大学 MOCS、阿尔伯塔 ACID 数据集制作时间与成本进行对比,结果如表 4 所示。可明显看出,由于可以进行多角

度图像的大批量生成与采集以及掩码文件的快速生成,在图像采集、筛选时间与标注综合成本方面,本文 SCED 数据集优势明显。

表 4 图像数据集工作量对比

Table 4 Comparison of Image Datasets Workload

数据集	图像量/张	数据集准备过程		标注过程		
		收集时间(段)	筛选时间/h	标注人力/人数	标注时间/h	标注花费/美元
ACID	10 000	—	162.0	50	225.0	3 650
MOCS	19 404	2005~2019 年	—	26	—	—
SCED	10 000	5 h	0.5	1	0.2*	0

注:“—”表示原文中未提及;“\*”为数据合成代码运行时间。

### 3.2 针对小目标的检测对比

在基于计算机视觉与深度学习方法的目标识别与检测领域,小目标的检测是一个重点主题,围绕如何提升小目标的检测效果,很多专家与学者进行了多种尝试<sup>[20]</sup>。本文 SCED 数据集在建立时便充分考虑了小目标检测问题,在小目标实例比例上相比 ACID 数据集更高。为验证 SCED 小目标图像的质量,将 SCED 中 1 000 张小目标图像随机替换掉 MOCS 与 ACID 数据集中的 1 000 张图像,同样依照上文 EfficientDet 模型的参数进行训练,效果展示见表 5,得到的精度变化对比情况见表 6、7。

由于 SCED 均为挖掘机图像,与 ACID 的对比中,替换后的 AP0.5 和 AP0.5 : 0.95 都有所提升,说明 SCED 作为挖掘机单一类别图像数据集的优质

特性。同时,由于替换的图像均为小目标属性,所以针对小目标检测精度的提升更明显,但由于替换可能造成大目标图像减少等因素,所以大目标尺寸下 AP0.5 : 0.95 的变化幅度甚微。而由于 MOCS 图像数据集数据量较大,所以每项数值虽都有提升,但幅度都不明显。此处也为小目标检测的优化提供了一种思路,即可通过提升有效小目标实例比例来提升检测效果。

## 4 讨论

随着计算机视觉技术与深度学习方法在土木工程领域应用的深入<sup>[21]</sup>,施工现场的检测技术也逐渐朝着智能化发展。为快速高质量地建立深度学习图像数据集,以便进一步开展相关检测与识别的工作,

表 5 MOCS 与 ACID 替换 1 000 张图像后的效果

Table 5 Effects of MOCS and ACID Replaced 1 000 Images

ACID 数据集		MOCS 数据集		数据集
				原始数据集
				替换后数据集

表 6 小目标替换后精度对比 (ACID)

Table 6 Accuracy Comparison After Small Target Replacement (ACID)

数据集	AP0.5	AP0.5 : 0.95	AP0.5 : 0.95 (小目标尺寸)	AP0.5 : 0.95 (中目标尺寸)	AP0.5 : 0.95 (大目标尺寸)
ACID(7 000)	0.918	0.767	0.368	0.579	0.791
ACID(6 000)+SCED(1 000)	0.927	0.774	0.443	0.592	0.795

表 7 小目标替换后精度对比 (MOCS)  
Table 7 Accuracy Comparison After Small  
Target Replacement (MOCS)

数据集	AP0.5	AP0.5 : 0.95
MOCS(19 404)	0.783	0.538
MOCS(18 404)+SCED(1 000)	0.795	0.549

本文依托图像快速合成技术提出了一种图像数据集构建方法。影响基于深度学习的目标检测算法效果的因素有很多,从图像数据集的质量分析,主要有:①目标物姿态的多样性;②背景(光照、环境物体)的多样性;③目标物尺寸的多样性;④目标物处于图像位置的均匀性等<sup>[23]</sup>。本文针对以上因素进行了一定的考虑与优化,除此以外,在基于图像合成方法的图像数据集构建方面,仍有其他因素值得进一步研究。

从上文可知,由于多姿态、多角度目标机械图像收集及背景图像替代的便捷性,以及掩码图像与标注文件生成的自动化,本文构建的 SCED 图像数据集从构建综合成本、图像丰富性以及图像质量方面都具有一定进步性,符合本文研究初衷。

同时,本文方法仍然具有进一步改进的空间。针对目标物清晰度方面,本文仅使用基于三维建模引擎建模的方法进行了目标物的构建。该模型与实际机械设备具有一定差异,未来可通过包括三维点云扫描在内的实体建模方法进行模型构建,进一步提升模型的质量,进而提升图像数据集质量。

针对图像融合方面,现有目标物与背景融合的方法还存在多种不一致性,如外观不一致与几何不一致等。这些不一致性将影响图像的真实性,从而降低图像数据集的总体质量。通过 GAN 网络以及其他算法消除亮度、位置和尺寸等不协调性将对提升图像数据集的质量具有重要意义。

## 5 结 语

(1)运用三维建模引擎与自编代码,融合预先获取的工地现场背景图像,实现了以挖掘机为目标的图像数据集构建与自动标注结果文件的生成,并构建了一个挖掘机图像数据集 SCED。

(2)本文方法生成的挖掘机图像数据集具有目标尺寸比例合理,目标位置分布均匀的特点,相比现有 ACID、MOCS 数据集具有更好的针对性、适用性与泛化性能,综合构建成本更低。

(3)基于 EfficientDet 目标检测算法,对本图像数据集进行了效果验证,结果表明 SCED 图像数据

集在小目标检测效果方面具有明显优势。

(4)本文提出的基于图像合成深度学习图像数据集快速构建方法可为建筑施工领域其他类型图像数据集的构建提供参考。

## 参考文献:

### References:

- [1] 方明,孙腾腾,邵楨.基于改进 YOLOv2 的快速安全帽佩戴情况检测[J].光学精密工程,2019,27(5):1196-1205.  
FANG Ming, SUN Teng-teng, SHAO Zhen. Fast Helmet-wearing-condition Detection Based on Improved YOLOv2[J]. Optics and Precision Engineering, 2019, 27(5): 1196-1205.
- [2] 孟诗乔,张啸天,乔甦阳,等.基于深度学习的网格优化裂缝检测模型研究[J].建筑结构学报,2020,41(增2):404-410.  
MENG Shi-qiao, ZHANG Xiao-tian, QIAO Su-yang, et al. Research on Grid Optimized Crack Detection Model Based on Deep Learning[J]. Journal of Building Structures, 2020, 41(S2): 404-410.
- [3] 李书进,赵源,孔凡,等.卷积神经网络在结构损伤诊断中的应用[J].建筑科学与工程学报,2020,37(6):29-37.  
LI Shu-jin, ZHAO Yuan, KONG Fan, et al. Application of Convolutional Neural Network in Structural Damage Identification[J]. Journal of Architecture and Civil Engineering, 2020, 37(6): 29-37.
- [4] 黄宏伟,李庆桐.基于深度学习的盾构隧道渗漏水病害图像识别[J].岩石力学与工程学报,2017,36(12):2861-2871.  
HUANG Hong-wei, LI Qing-tong. Image Recognition for Water Leakage in Shield Tunnel Based on Deep Learning[J]. Chinese Journal of Rock Mechanics and Engineering, 2017, 36(12): 2861-2871.
- [5] 薛亚东,李宜城.基于深度学习的盾构隧道衬砌病害识别方法[J].湖南大学学报(自然科学版),2018,45(3):100-109.  
XUE Ya-dong, LI Yi-cheng. A Method of Disease Recognition for Shield Tunnel Lining Based on Deep Learning[J]. Journal of Hunan University (Natural Sciences), 2018, 45(3): 100-109.
- [6] 叶肖伟,张小明,倪一清,等.基于机器视觉技术的桥梁挠度测试方法[J].浙江大学学报(工学版),2014,48(5):813-819.  
YE Xiao-wei, ZHANG Xiao-ming, NI Yi-qing, et al. Bridge Deflection Measurement Method Based on Machine Vision Technology[J]. Journal of Zhejiang Uni-

- versity (Engineering Science), 2014, 48(5): 813-819.
- [7] AN X H, ZHOU L, LIU Z G, et al. Dataset and Benchmark for Detecting Moving Objects in Construction Sites[J]. Automation in Construction, 2021, 122: 103482.
- [8] XIAO B, KANG S C. Development of an Image Data Set of Construction Machines for Deep Learning Object Detection[J]. Journal of Computing in Civil Engineering, 2021, 35(2): 05020005.
- [9] SIMONYAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-scale Image Recognition [J]. arXiv:1049. 1556.
- [10] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[C]//IEEE. Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas: IEEE, 2014: 580-587.
- [11] HE K M, ZHANG X Y, REN S Q, et al. Deep Residual Learning for Image Recognition[C]//IEEE. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 770-778.
- [12] LI Z, PENG C, YU G, et al. DetNet: Design Backbone for Object Detection [C]//FERRARI V, HEBERT M, SMINCHISESCU C, et al. Proceedings of the 15th European Conference, Munich: LNCS, 2018: 339-354.
- [13] 陈伟华, 南鹏飞, 闫孝姮, 等. 基于深度学习的采煤机截割轨迹预测及模型优化[J]. 煤炭学报, 2020, 45(12): 4209-4215.  
CHEN Wei-hua, NAN Peng-fei, YAN Xiao-heng, et al. Prediction and Model Optimization of Shearer Memory Cutting Trajectory Based on Deep Learning [J]. Journal of China Coal Society, 2020, 45(12): 4209-4215.
- [14] 张繁, 王通, 黄可蒙, 等. 基于 Unity3D 的改进实时红外仿真系统[J]. 计算机辅助设计与图形学学报, 2018, 30(7): 1177-1186.  
ZHANG Fan, WANG Tong, HUANG Ke-meng, et al. One Improved Real-time Infrared Simulation System Based on Unity3D [J]. Journal of Computer-aided Design & Computer Graphics, 2018, 30(7): 1177-1186.
- [15] QIU W, ZHONG F, YI Z, et al. UnrealCV: Virtual Worlds for Computer Vision [C]//LIU Q, LIENHART R, WANG H H. Proceedings of the 25th ACM International Conference on Multimedia, New York: Association for Computing Machinery, 2017: 1221-1224.
- [16] 张素兰, 郭平, 张继福, 等. 图像语义自动标注及其粒度分析方法[J]. 自动化学报, 2012, 38(5): 688-697.  
ZHANG Su-lan, GUO Ping, ZHANG Ji-fu, et al. Automatic Semantic Image Annotation with Granular Analysis Method [J]. Acta Automatica Sinica, 2012, 38(5): 688-697.
- [17] 高慧, 张继威, 来扬, 等. 深度学习的人体图像半自动标注系统[J]. 北京邮电大学学报, 2021, 44(1): 104-109.  
GAO Hui, ZHANG Ji-wei, LAI Yang, et al. Deep Learning Based Semi-automatic Labeling System for Human Images [J]. Journal of Beijing University of Posts and Telecommunications, 2021, 44(1): 104-109.
- [18] TAN M X, PANG R M, LE Q V. EfficientDet: Scalable and Efficient Object Detection [C]//IEEE. Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle: IEEE, 2020: 10778-10787.
- [19] 柯圣财, 赵永威, 李弼程, 等. 基于卷积神经网络和监督核哈希的图像检索方法[J]. 电子学报, 2017, 45(1): 157-163.  
KE Sheng-cai, ZHAO Yong-wei, LI Bi-cheng, et al. Image Retrieval Based on Convolutional Neural Network and Kernel-Based Supervised Hashing [J]. Acta Electronica Sinica, 2017, 45(1): 157-163.
- [20] 刘洋, 战荫伟. 基于深度学习的小目标检测算法综述[J]. 计算机工程与应用, 2021, 57(2): 37-48.  
LIU Yang, ZHAN Yin-wei. Survey of Small Object Detection Algorithms Based on Deep Learning [J]. Computer Engineering and Applications, 2021, 57(2): 37-48.
- [21] 刘金杉, 李元海, 卢昱杰, 等. 基于深度学习的隧道衬砌表观病害模拟检测系统[J/OL]. 哈尔滨工业大学学报, 1-10 [2022-02-14]. <http://kns.cnki.net/kcms/detail/23.1235.T.20220112.0853.002.html>.  
LIU Jin-shan, LI Yuan-hai, LU Yu-jie, et al. Tunnel Lining Surface Defect Simulation and Detection System Based on Deep Learning [J/OL]. Journal of Harbin Institute of Technology: 1-10 [2022-02-14]. <http://kns.cnki.net/kcms/detail/23.1235.T.20220112.0853.002.html>.
- [22] 卫星, 李佳, 孙晓, 等. 基于混合生成对抗网络的多视角图像生成算法[J]. 自动化学报, 2021, 47(11): 2623-2636.  
WEI Xing, LI Jia, SUN Xiao, et al. Cross-view Image Generation via Mixture Generative Adversarial Network [J]. Acta Automatica Sinica, 2021, 47(11): 2623-2636.