

基于合成图像数据集的挖掘机关键点识别

姚宗伟¹, 陈 辰¹, 高振云², 靳鸿鹏¹, 荣 浩², 李学飞¹, 黄虹溥², 毕秋实¹

(1. 吉林大学 机械与航空航天工程学院, 长春 130022; 2. 柳州柳工挖掘机有限公司, 广西 柳州 545007)

摘 要: 本文提出了一种基于合成图像数据集的挖掘机关键点识别方法, 通过虚拟模型和场景的随机自动化以及关键点坐标和遮挡信息判定, 生成多样化的合成图像, 并利用基于平面视觉的深度神经网络完成关键点识别, 解决了传统大规模数据集采集困难的问题。试验结果显示: 该方法提高了关键点识别精度, 归一化误差和正确关键点百分比分别为 0.005 6 和 97.64%。因此, 本文方法能够满足监控挖掘机的作业安全和工作效率的实际应用需求, 同时避免了高质量工程数据集采集时安全风险高、时间/经济成本高、工况覆盖面窄且标签准确率低等问题, 有助于深度学习和大数据技术在挖掘机工作状态识别方面的应用部署。

关键词: 机械设计及理论; 挖掘机; 关键点识别; 深度学习; 合成图像

中图分类号: TP18 **文献标志码:** A **文章编号:** 1671-5497(2026)01-0076-10

DOI: 10.13229/j.cnki.jdxbgxb.20240677

Visual recognition of excavator keypoints based on synthetic image datasets

YAO Zong-wei¹, CHEN Chen¹, GAO Zhen-yun², JIN Hong-peng¹, RONG Hao²,
LI Xue-fei¹, HUANG Hong-pu², BI Qiu-shi¹

(1. School of Mechanical and Aerospace Engineering, Jilin University, Changchun 130022, China; 2. Liuzhou Liugong Excavator Co., Ltd., Liuzhou 545007, China)

Abstract: This paper proposes a method for excavator pose recognition using synthetic image datasets for model training. Initially, virtual models and scenarios are established. Programming is utilized to randomize the excavator pose, virtual camera position, and scene parameters. Subsequently, keypoint coordinates and occlusion information are computed to construct synthetic image datasets. Finally, excavator key points estimation is performed using a monocular camera. Experimental results demonstrate that training with synthetic image datasets improves model recognition accuracy, with a normalized error of 0.005 6 and a percentage of correct keypoint of 97.64%. Therefore, this method can meet the practical application needs of monitoring excavator operation safety and work efficiency. It also avoids issues such as high safety risks, high time/economic costs, narrow working condition coverage, and low label accuracy associated with high-quality engineering dataset collection. This contributes to the application and deployment of deep learning and big data technologies in excavator work state recognition.

收稿日期: 2024-06-17.

基金项目: 国家自然科学基金项目(52375246, 52105100, 52372428); 广西重点研发计划项目(2023AB09014).

作者简介: 姚宗伟(1985-), 男, 教授, 博士. 研究方向: 工程车辆智能化关键技术. E-mail: yzw@jlu.edu.cn

通信作者: 黄虹溥(1986-), 男, 高级工程师, 博士. 研究方向: 挖掘机智能化关键技术. E-mail: hhp@liugong.com

Key words: mechanical design and theory; excavator; keypoints estimation; deep learning; synthetic datasets

0 引言

土方作业是建筑行业中的重要工序,相关工程机械的作业安全和工作效率很大程度上决定着整个工程项目的进展^[1]。实时且深入地了解施工信息是有效管理工程项目的关键,其中就包括了对工程机械现场活动的监测^[2]。因此,监测整个施工现场以获取工程机械相关数据,对于避免安全事故的发生和生产分析十分重要。

土方工程机械被广泛应用于基建、农田水利、抢险救灾、采矿等领域^[3]。对于施工现场的挖掘机,传统监控往往采用目测方式,而这种方式费时费力,还存在人为疏忽导致无法及时预警的风险,也不能准确估计效率。在基于传感器的监测方法中,IMU是常用传感器,其优点是精度较高,且可在挖掘机不可见时测量位姿;然而,为了获取现场所有挖掘机的位姿,需在每台挖掘机上都安装惯性测量单元(Inertial measurement unit, IMU),费用高昂且费时费力,还需要定期停机维护以消除数据漂移^[4]。使用实时定位系统(Real-time locating system, RTLS)^[5]或全球定位系统(Global positioning system, GNSS)^[6]的方法同样存在上述问题。

随着计算机技术的发展,基于视觉的监控方法取得了长足的进步。有标记方法^[7]易出现标签损坏、脱落和遮挡等问题,因此,无标记方法受到越来越多的关注^[8,9]。传统计算机视觉的方法通常需要人工设计特征提取和决策标准才能获得期望输出^[10],在复杂的挖掘机作业现场应用效果不佳。因此,深度学习类方法成为当前的研究焦点^[11]。

为得到性能优良的深度学习模型,需要利用大规模数据集进行大量训练,而目前所使用的数据集,一般通过直接拍摄实际工作的挖掘机获得,存在如下缺陷:

(1)安全风险高。由于挖掘机工作场地内的大型机械较多,且环境嘈杂、盲区较多,对数据采集人员的安全具有一定威胁。

(2)时间/经济/人力成本高。现场采集图像不仅需要多方协调,还要求具有较强专业能力的

数据采集团队,导致综合成本很高。

(3)工况覆盖面窄。挖掘机的应用工况十分广泛,但由于前述原因,能够满足数据采集要求的工况较少,经其训练得到的模型在实际应用中存在较严重的“水土不服”。

(4)标签准确性差^[12]。实际拍摄获得的视频/图像,需要人工后期补加标签,而挖掘机工作时的影像数据无法像Cyber空间中的数据一样“干净”;同时,人眼识别难免存在误差,导致标注出的数据集中存在相当比例的“脏”数据。

为了绕过获取高质量工程数据集时的重重障碍,有研究利用合成图像对模型进行训练并达到真实图像训练模型近似的性能^[13];同时,由于合成图像时可以使视觉特征多样化,因此,较少的合成图像数据集就有可能达到较高的识别精度;此外,使用合成图像可以避免机密和版权问题^[14]。因此,本文通过自动化生成带标签信息的合成图像的方法建立数据集,进而开发了基于低成本视觉的挖掘机2D位姿识别方法。

1 研究现状

1.1 挖掘机关键点识别

随着人工智能的发展,尤其是卷积神经网络(Convolutional neural network, CNN)的提出,出现了诸多基于深度学习网络的人体位姿估计方法,并以此为基础改进得到了挖掘机状态估计方法,识别效果相较于早期使用计算机视觉的方法有了较大提升。Liang等^[15]对人体位姿估计的堆叠沙漏网络进行改进,建立了无须标记的挖掘机关键点识别方法,进而获得了2D姿态估计模型。Zhang等^[16]提出了一个挖掘机活动分析与安全监测系统:首先检测挖掘机并识别其10个关键点,然后估算2D位姿并结合基于规则的动作识别模块进行生产率分析,最后通过区域分割来划分工作区域并判别安全区域。Assadzadeh等^[8]定义了挖掘机的6个关键点,并利用高分辨率网络(HRNet)^[17]进行识别,进而估计挖掘机的2D位姿。

此外,有些研究识别了挖掘机关键点的3D坐标,并据此开发了3D位姿估计系统,分别为立体

视觉法和平面视觉法^[16]。立体视觉法一般利用具有特定位姿关系的多台相机,或者直接采用深度摄像机^[18],但各有局限:结构光法深度相机由于使用红外线测量距离,不适合在室外使用,且其体积大、功耗高;飞行时间(Time of flight, TOF)深度相机需要多次采样积分,测量时间较长,在测量运动物体时可能产生运动模糊,不能满足挖掘机对实时性和精度的要求;双目相机需要精确标定两台相机的位姿关系,而挖掘机应用场景中的恶劣条件很容易破坏相机之间的相对位置,使标定数据失效,因此,该类方法在实际应用时流程烦琐、可控性差。而平面视觉法训练数据集的获取往往需要到挖掘机工作现场拍摄影像数据并依靠传感器及人工添加标签信息,不仅同样存在上述传感器方法的缺陷,还存在前述获取高质量工程数据集困难的问题^[15]。此外,训练出的模型对与数据集中所包含的挖掘机有较大差别的目标识别效果较差,如需提高识别精度则需添加新的数据集。

1.2 合成图像数据集

如前所述,CNN需要经过大量标记数据的训练才能在目标检测和姿态估计等方面获得最佳性能。相比于人体姿态估计领域中大量开源的数据集,如 32.8 万张带标签图像的 COCO 数据集^[19]和包含 360 万个人体位姿标签的 Human 3.6M 数据集^[20],工程机械领域的公共数据集相对较少,如 MOCS 数据集^[14],但该数据集仅支持分类识别,无法进行关键点识别和位姿估计。此外,由于数据集中的标签信息难以获取,限制了研究中数据集的数量,如 Zhang 等^[16]的数据集包含 7 293 张带标记的图片和 479 个带标记的视频;Wen 等^[21]使用的数据集更小,只有 5 798 张带标签图像。

为了解决数据集难以获取的问题,有研究提出了自动生成^[22]或扩充数据集的方法。Luo 等^[23]提出使用数据增强的方式扩充数据集,如旋转、翻转、调整和缩放,以及颜色调整、饱和度和对比度的变化等,将 1 281 张带标签的图片扩充到 6 405 张。这些方法获取的数据集,只能称为“间接”数据集,其与实际挖掘机的特征关联性不高,所训练出的模型在实际应用时很难取得良好效果。

Torres 等^[24]提出了一种合成数据生成的方法以减少数据集的标注工作,将虚拟模型和视频帧中的挖掘机对齐,并自动输出对应的标签。Tian

等^[25]在 C4D 中自动获得挖掘机的合成图片及坐标,建立了用于位姿估计的数据集。与手动添加标签的方式相比,这种方式可以自动化生成样本量极大的数据集,能节约大量时间成本和人力成本,如 Assadzadeh^[8]基于域随机化的方法在 Unity 中建立虚拟模型生成了 1.5 万张合成图片训练集用于 2D 识别模型。不过,受技术手段限制,上述研究中生成的合成图片存在纹理与光照单一、背景与真实环境差距大等问题,导致所训练模型的识别精度不能满足实际使用要求^[14,24,25]。要想利用这种方式获得性能优良的挖掘机位姿识别模型,必须进行更加深入的研究。

2 本文方法

为了弥补合成图像数据集的本体和背景不拟真等问题所导致的识别模型精度低的缺陷,本文提出了一种基于平面视觉的挖掘机关键点识别框架,如图 1 所示。本文建立虚拟挖掘机模型并生成虚拟场景,随机化挖掘机位姿和其他条件,包括光照和相机参数等,并自动输出图像和对应的标签,生成合成图像数据集;此外,为了验证合成图像数据集模型的性能,制作了实际挖掘机数据集;最后,基于深度神经网络识别挖掘机关键点坐标。

2.1 关键点选择

由于不在挖掘机上安装可视化标记,因此,必须将关键点的坐标信息嵌入数据集中。为了避免某一部件发生部分遮挡时丢失该部件的整体位置信息,本文对关键点的选取提出了一定的要求。首先,为了得到工作装置的位姿,最少需要 4 个基础点,分别选择动臂铰点、斗杆铰点、铲斗铰点和铲斗尖点;但是,挖掘机作业时铲斗尖点常深入土壤中,难以拍摄到,且动臂铰点和铲斗铰点也经常被驾驶室和铲斗遮挡,如图 2 所示,因此,加入 3 个补充点:动臂油缸铰点、斗杆油缸铰点和铲斗油缸铰点。此时,挖掘机工作装置每个部件至少包含 2 个关键点,即使基础点发生遮挡而导致该点识别精度降低或识别失败,也可以利用补充点作为对应部件位置信息的补充以提高识别准确率。因此,标签共含 1 个挖掘机包围框和 7 个关键点(包含基础点和附加点),如图 2 所示,其中的 Point_1 到 Point_7 分别代表动臂铰点、斗杆铰点、铲斗铰点、铲尖尖点、动臂油缸铰点、斗杆油缸铰点和铲斗油缸铰点。另外,当关键点被遮挡时,

2D 图像上的特征会发生变化,应当加入判断遮挡信息以提高模型识别精度^[8]。因此,将关键点标

签设计为 $[x, y, v]$, x, y 为坐标, v 代表可见性,0 为超出视图、1 为遮挡、2 为可见。

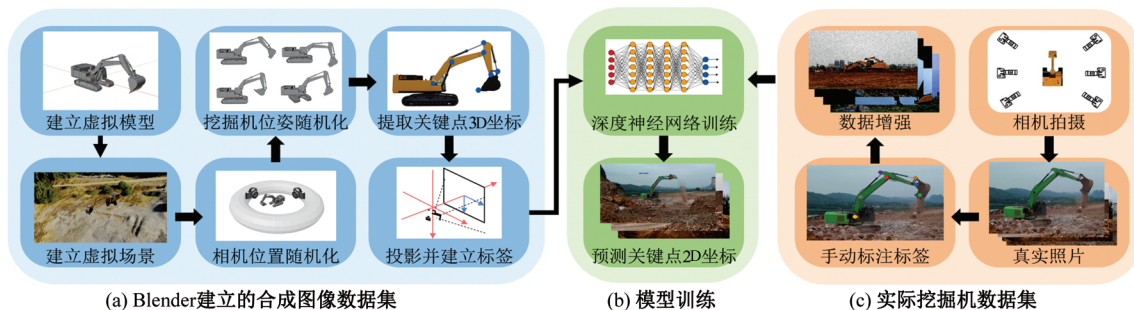


图 1 基于合成图像数据集的挖掘机关键点识别

Fig. 1 Keypoints recognition of excavator based on synthetic image dataset

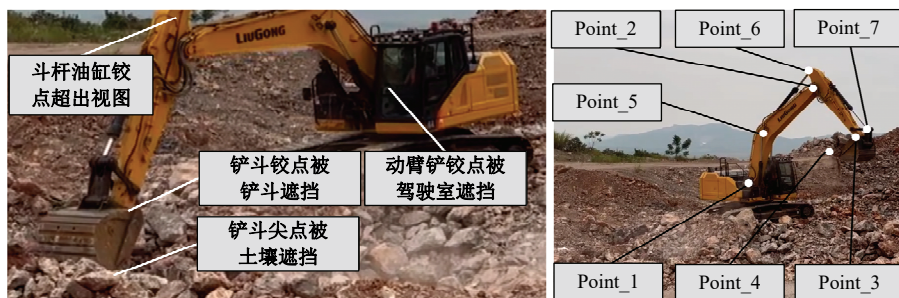


图 2 基础点遮挡以及关键点示意图

Fig. 2 Occlusion of base-points and keypoints

2.2 合成图像数据集

本文在 Blender 中建立合成图像数据集,其具有以下特点:①可自动对模型进行随机化,如更改背景参数、更改某一零件的形貌、添加遮挡物体等,能快速生成多样化数据集;②可通过编程自动读取模型任意点的 3D 坐标并计算投影后的 2D 坐标,进而生成数据集标签。

建立合成图像数据集的过程如图 1(a)所示:①建立挖掘机虚拟模型;②生成虚拟背景环境;③随机化参数并输出合成图片;④提取挖掘机关键点 3D 坐标,并投影为 2D 坐标,生成标签文件。

第①步,构造实际挖掘机的虚拟模型,包括行走履带、上部转台、动臂、斗杆和铲斗 5 个主要部分及油缸系统。可以通过骨骼功能和子父级关系功能设置挖掘机各部件之间的运动关系和各关节的活动范围,模拟实际挖掘机的作业姿态。

第②步,建立背景与环境虚拟模型,以提高训练后模型对真实挖掘机的识别精度^[12]。建立的场景包括建筑工地、城市道路、采石场和野外林间等,如图 3 所示。

第③步,参数随机化(包括虚拟相机的相关参数



图 3 合成图片的背景和环境

Fig. 3 Background and environment of synthetic image

以及挖掘机位姿和材质纹理等)并输出合成图像。

首先,在挖掘机周围放置虚拟摄像机,并随机化其焦距和传感器尺寸;然后,定义摄像机的放置范围以满足以下条件:该范围内任何位置的摄像机应该能够捕捉整个挖掘机,并且距离挖掘机有一定距离,因此,设定为挖掘机模型周围的水平环形区域,如图 1(a)所示。

将虚拟挖掘机的中心视为环形区域的形心,根据相机的视场角(Field of view, FOV)选择环形区域的最大半径 $R_{\max}$ 、最小半径 $R_{\min}$ 、最高点 $Z_{\max}$ 、最低点 $Z_{\min}$ 。首先使用摄像机到挖掘机的距离 d_{camera} 和摄像机绕挖掘机坐标系 Z 轴的角度 α_{camera} 确定虚拟相机在水平面内的坐标 $(X_{\text{camera}}, Y_{\text{camera}})$,然后,在范围内随机选取高度确定虚拟相机的垂向坐标 Z_{camera} 。

$$\begin{cases} d_{\text{Camera}} = \text{rand}(R_{\min}, R_{\max}) \\ \alpha_{\text{Camera}} = \text{rand}(0, 2\pi) \\ X_{\text{Camera}} = d_{\text{Camera}} \times \cos \alpha_{\text{Camera}} \\ Y_{\text{Camera}} = d_{\text{Camera}} \times \sin \alpha_{\text{Camera}} \\ Z_{\text{Camera}} = \text{rand}(Z_{\min}, Z_{\max}) \end{cases} \quad (1)$$

为了使相机的视角朝向虚拟模型,其视线应穿过挖掘机模型的中心,即世界坐标系的原点。分别定义 α_x 和 α_z 为虚拟相机绕自身坐标系 X 轴的角度与绕 Z 轴的角度,通过设置这两个角即可使相机朝向虚拟模型。

$$\begin{cases} \alpha_x = \frac{\pi}{2} - \arctan \frac{Z_{\text{Camera}}}{\sqrt{X_{\text{Camera}}^2 + Y_{\text{Camera}}^2}} \\ \alpha_z = \pi - \alpha \end{cases} \quad (2)$$

再对挖掘机姿态随机化。影响挖掘机姿态的变量主要有4个,分别是回转角 θ_0 、动臂关节转角 θ_1 、斗杆关节转角 θ_2 和铲斗关节转角 θ_3 。因此,通过随机化公式($\theta_i = \text{rand}(\theta_{i,\min}, \theta_{i,\max})$)给出以上角度值,即可模拟挖掘机的姿态,如图1(a)所示。

然后,将模型表面的纹理、材质、部分部件的颜色以及辉光、环境光遮蔽和空间反射等参数设置为在一定范围内的随机值,确保与现实中相似。

最后,调用渲染,输出并保存合成图片。

第④步,生成标签文件。在2.1节中选定了7个关键点作为数据标签,需要在Blender中生成其2D坐标并判断遮挡。

首先,通过虚拟相机和挖掘机的相对位置可以计算出虚拟相机的外参 E ,通过虚拟相机的焦距、FOV、虚拟传感器大小等可以计算出虚拟相机的内参 K 。

然后,建立关键点并通过子父级关系绑定到挖掘机模型的铰点上,提取其在水坐标系下的3D坐标,根据 E 和 K 变换为合成图像中的像素坐标系下的2D坐标,并进行归一化。

之后,通过添加辅助点的方式来确定关键点是否被遮挡,以关键点Point_3为例,在铲斗边缘位置建立两个辅助点,如图4所示,红点为挖掘机左侧辅助点,绿点为右侧辅助点(彩图见电子版,下同)。可以看出,在虚拟相机(即图片的视角)处于不同的位置时,Point_3是否被遮挡与其相对于辅助点的位置是相关的,即利用相机位置、关键点2D坐标、辅助点2D坐标等信息,可以判断关键点是否被遮挡。

最后,将归一化关键点2D坐标、遮挡判断信息一起输出为标签文件。

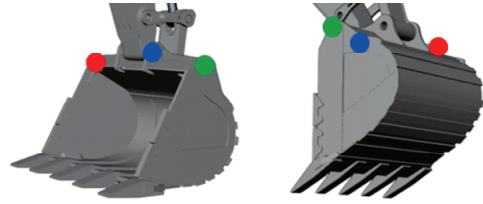


图4 利用辅助点判断关键点的遮挡

Fig. 4 Auxiliary points to determine occlusion of keypoints

在Blender中,利用Python脚本可以实现上述参数随机化和标签生成的自动化,并可以循环调用生成大量合成图片及标签。算法流程如图5所示。

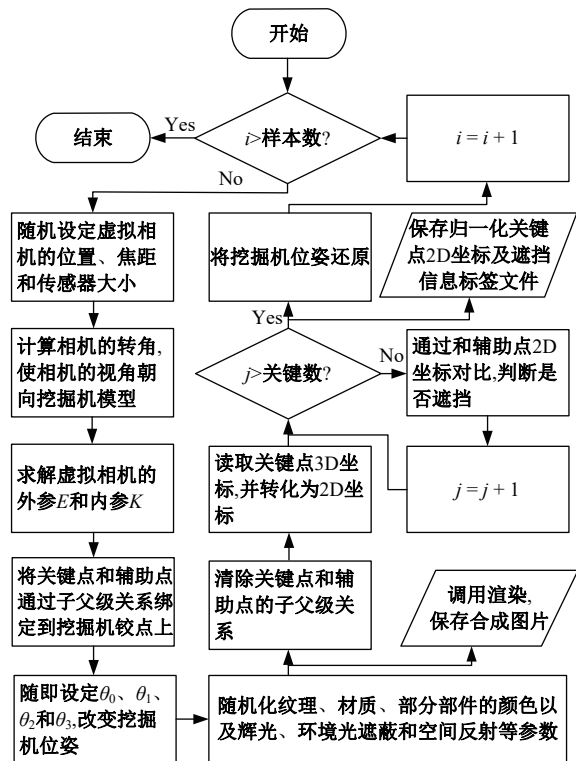


图5 虚拟数据集生成流程

Fig. 5 Virtual dataset generation process

2.3 深度神经网络模型

目前,能够进行位姿识别的神经网络模型有多种,其中性能较为突出的是YOLO系列。作为一种自底向上的图像识别算法,相较于其他自顶向下的深度学习图像识别算法,如CNN、堆叠沙漏网络等,YOLO系列网络能在保持精度的情况下达到更快的识别速度和更小的模型体积。YOLOv8是由YOLO前序版本改进而来,延续了YOLOv5中的CSP、特征融合方法和SPPF模块的思想,并做了部分改进^[26]:①为了满足不同项目的需求,基于类似于YOLOv5的尺度系数设计

了不同尺度的模型,包括 640 像素的 n、s、m、l、x 和 1 280 像素的 p6 目标检测网络;②在保留 YOLOv5 原始思想的前提下,参照 YOLOv7 中的 ELAN 结构,设计了 C2f 模块,如图 6 所示;③检测头部分采用了目前流行的方法;④分类损失使用了 BCE 损失,回归损失为 CIOU 损失+DFL,在损失函数中提出了非对称加权操作,正负样本非

对称加权、突出正样本为主样本。在 COCO 数据集上,YOLOv8 相较于其前序版本在精度和速度上都有所提高。

使用预训练模型初始化网络结构的权重是一种节省训练时间、提高精度并且降低过拟合概率的方法^[21],因此,本文使用 YOLOv8x-pose-p6 作为预训练模型进行训练。

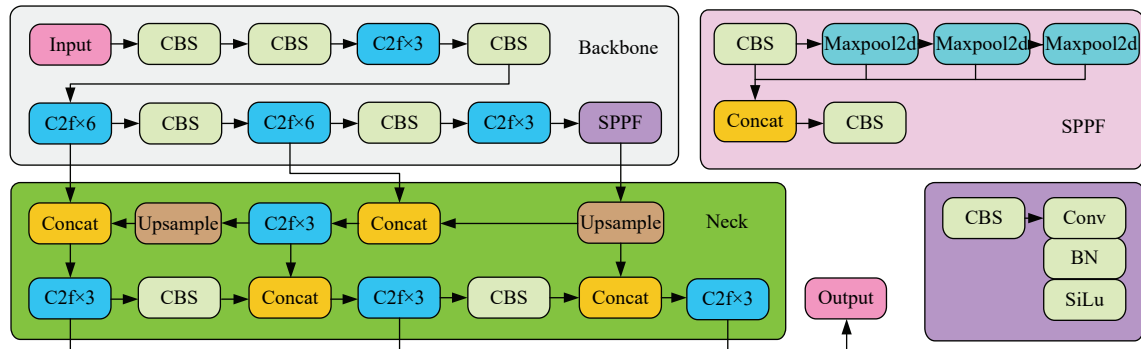


图 6 YOLOv8 网络结构

Fig. 6 Structure of YOLOv8

2.4 实际挖掘机数据集

为了探究本文方法的可靠性,需要利用实际挖掘机数据集进行验证,其获取过程如图 1(c)所示:①现场采集挖掘机图像;②手动标注标签;③图片数据增强。

拍摄的挖掘机作业工况应涵盖其常用工况,包括高地向下挖掘、平地挖掘、平地挖掘装车。为了保证数据集的多样性,应在挖掘机不同位姿、不同工况条件(见图 7(a))、不同相机角度(见图 7(b))、不同焦距、不同位置、不同光照和天气条件时进行拍摄。



图 7 实际挖掘机数据集获取

Fig. 7 Acquisition of actual excavator dataset

数据增强能够增加训练的数据量、丰富数据的分布,在一定程度上改善数据样本数量不平衡的情况,进而降低模型过拟合程度并提高泛化能力。因此,本文基于翻转、旋转、平移、缩放、对比度变换、添加高斯噪声、颜色变换等操作,对挖掘

机数据集进行增强,以模拟现实中出现的挖掘机斜坡作业、相机视野受限或可视性差等情况。

3 实验验证

3.1 数据收集

为了验证挖掘机关键点识别精度,需要准备深度神经网络的训练和测试数据;同时,为了验证使用真实照片数据集和合成图像数据集训练结果的差异,使用 2.2 和 2.4 节中提到的方法收集两种数据分别训练两个网络。训练过程中的参数设置均相同,优化器为 SGD, batch 为 8, 训练轮数为 100, 其他超参数根据经验选取。第一个模型使用的数据集 Dataset #1 为 37 368 张现场拍摄的实际挖掘机图片,其中 70% 作为训练集,30% 作为验证集。第二个模型使用的数据集 Dataset #2 为 Dataset #1 中随机选取 10 000 张图片,并加入 14 000 张由 Blender 随机化生成的合成照片,共计 24 000 张图片,其中 70% 作为训练集,30% 作为验证集。执行训练的硬件平台为:CPU Intel(R) Xeon (R) Silver 4210R @ 2.40 GHz、GPU Tesla V100S-32 GB。两模型的训练时间分别为 150 h 和 121 h。

测试挖掘机由专业驾驶员操作,工作模式为快速,连续挖掘两斗后停止,每次挖掘循环中挖掘机上部向左回转卸料,回转角度约为 120°,如图 8 所示。摄像机距离挖掘机约 35 m,摆放在机身左

前侧,且与挖掘机在同一水平高度。图像分辨率为 1 080 p,视频帧率为 60 fps,共拍摄了 39 s 的视频(2 个作业循环),提取出 2 340 张图片及附带的传感器数据。手动为测试数据图片添加每个关键点坐标标签,作为坐标的真实值 $P_{i,n}^{GT}$ 。

3.2 测试指标

将归一化误差(NE)和正确关键点百分比(PCK)作为评价 2D 坐标的估计值 $P_{i,n}^{es}$ 精度的指标。NE 为 $P_{i,n}^{es}$ 与 $P_{i,n}^{GT}$ 之间的距离差,并用图像对角线长度归一化,如式(3)所示。PCK 取决于关键点的 $P_{i,n}^{GT}$ 到 $P_{i,n}^{es}$ 的欧氏距离是否在阈值 α_{PCK} ,如式(4)所示:

$$NE_i = \frac{1}{f} \sum_{n=1}^f \frac{\|P_{i,n}^{GT} - P_{i,n}^{es}\|_2}{l_n} \quad (3)$$

$$PCK_i(\alpha) = \frac{1}{f} \sum_{n=1}^f \left(\|P_{i,n}^{GT} - P_{i,n}^{es}\|_2 < \alpha_{PCK} \times l_n \right) \quad (4)$$

式中: f 为测试集的总帧数; n 为帧序号; α_{PCK} 根据挖掘机关键点在图像中的实际大小,设为 0.02;

$\|P_{i,n}^{GT} - P_{i,n}^{es}\|_2$ 为第 n 帧中第 i 个关键点的真实 2D 坐标到估计 2D 坐标的欧氏距离; l_n 为第 n 帧中图像原始的对角线长度。

3.3 试验结果

首先使用 Dataset #1 训练的模型,验证本文方法对关键点的识别精度,然后将 Dataset #2 上训练的模型性能与真实照片数据集上训练的模型性能进行比较,验证合成图像数据集的可行性。识别到的关键点如图 9 所示。

使用 Dataset #1 训练后,模型在所有关键点上的平均 PCK 为 96.28%,平均 NE 为 0.005 9,如表 1 和表 2 所列,表明预测关键点坐标与真实值接近,模型具有良好的精度。使用 Dataset #2 训练后,模型在所有关键点上的平均 PCK 为 97.64%,平均 NE 为 0.005 6。可见,使用 Dataset #2 训练的模型精度略微高于使用 Dataset #1 训练的模型精度,说明使用合成图像数据集的方法不仅减少了传统建立数据集所需的大量时间和工作,而且经其训练的模型具有更出色的性能。



图 8 测试数据示例图像

Fig. 8 Samples of test data



(a) Dataset #1 模型识别结果



(b) Dataset #2 模型识别结果

图 9 关键点识别结果

Fig. 9 Keypoint recognition results

本文中识别的关键点与现有研究并不完全一致,本文识别了 7 个关键点(动臂铰点、斗杆铰点、铲斗铰点、铲斗尖点、动臂油缸铰点、斗杆油缸铰点和铲斗油缸铰点),而如 Tian 等^[25]的研究中识别

了上述前 4 个点和机体尾部点,因此,只对共同的关键点的精度进行比较。Tian 等^[25]使用从人体姿态估计问题转化来的非堆叠沙漏网络模型进行 2D 关键点的识别;Mahmood 等^[27]使用的是端到端的

CNN 模型,对比结果如表 1 所列。Assadzadeh 等^[8]使用的是 HRNet; Wen 等^[21]使用的是改进 R-CNN,对比结果见表 2。可以看到,本文方法得到

的关键点识别的平均 PCK 与平均 NE 均优于现有研究,而且本文方法不需要可视化标记,提高了算法的可靠性。

表 1 2D 关键点识别的 PCK 值

Table 1 PCKs of keypoints identification %

数据集	Point_1	Point_2	Point_3	Point_4	Point_5	Point_6	Point_7	平均
Dataset #1	99.95	99.05	98.57	82.38	99.99	94.24	99.76	96.28
Dataset #2	99.99	99.99	99.99	83.76	99.99	99.76	99.95	97.64
Tian 等 ^[1]	98.75	95.00	88.57	80.10	—	—	—	90.61
Mahmood 等 ^[2]	96.18	96.94	94.65	92.78	—	—	—	95.14

表 2 2D 关键点识别的 NE 值

Table 2 NEs of keypoints identification

数据集	Point_1	Point_2	Point_3	Point_4	Point_5	Point_6	Point_7	平均
Dataset #1	0.003 8	0.004 3	0.006 1	0.012 8	0.002 2	0.006 1	0.005 9	0.005 9
Dataset #2	0.003 4	0.004 2	0.006 0	0.013 1	0.003 1	0.004 8	0.004 9	0.005 6
Assadzadeh 等 ^[3]	0.028 7	0.020 8	0.031 1	0.099 2	—	—	—	0.045 0
Wen 等 ^[4]	0.039 9	0.028 2	0.021 9	0.029 0	—	—	—	0.029 8

4 结果讨论

与部分现有文献的结果相似^[8,24],本文实验结果中大部分关键点之间的 PCK 值和 NE 值接近,特例是铲斗尖点 Point_4,其精度明显低于其他 6 个点。为了探明可能的原因,本文额外评估了挖掘、回转、卸料各阶段中 Point_4 的 PCK 值,同时由于下降和提升阶段与挖掘机回转一般同时进行,因此,将上述两阶段包含在回转阶段中,如表 3 所示。可以看到,识别精度在挖掘阶段显著低于在回转和卸料阶段,而回转和卸料阶段的 PCK 值则与其他 6 个点保持在同一水平。由此可知,造成 Point_4 识别精度低的可能原因是挖掘实验过程铲斗尖点长时间被土壤遮挡,数量占测试集总数的 34.29%;同时,遮挡也使标注人员难以确定斗尖的真实位姿,只能依据经验推断 Point_4 所在的位置,导致难以得到准确的标注,而回转和卸料阶段因不受遮挡则没有影响。虽然对 Point_4 的识别精度不如其他点高,但是误差水平仍在可接受范围内,其 NE 值为 0.012 8,即平均距离差为 28.2 个像素,对于挖掘机作业时的现场监控和效率分析等后续流程的影响较小。

值得注意的是,测试数据为现场拍摄的实际挖掘机作业视频,Dataset #1 全部由实际挖掘机照片组成,Dataset #2 由 41.7% 的实际照片和 58.3% 的合成图像组成,Dataset #2 所训练的模型能够正确识别实际挖掘机的关键点,且相比于

表 3 各阶段 Point_4 的 PCK

Table 3 Point_4's PCK for each stage

数据集	PCK/%		
	挖掘	回转	卸料
Dataset #1	59.19	89.63	99.99
Dataset #2	67.22	89.05	95.30

Dataset #1 模型识别精度更高,这证明了本文提出的合成图像数据集方法的泛化性。同时,Dataset #1 的图像为 37 368 张,Dataset #2 的图像为 24 000 张,Dataset #2 以更少的样本数量达到了更高的识别精度,这与 Kim 等^[12]的研究结果吻合。其原因有以下两点:①虚拟化的挖掘机和场景可以模拟各种情况,如各种视角和位姿,丰富数据集中的样本种类,而实际数据集采集过程中由于现实条件限制往往难以实现;②本文合成图像生成方法中,虽然精细化建模了挖掘机和背景环境,使用了接近真实的光影和纹理等,但是合成图像的视觉特征仍与真实图片有本质不同。因此,将真实照片和合成图像混合作为数据集,既有利于减少数据集的工作量,又可以抵消实际照片和合成图像数据集的缺点,提高模型的识别精度和泛化性。

5 结束语

本文提出了一种基于平面视觉的挖掘机关键点识别方法,使用合成图像建立数据集,利用深度神经网络进行挖掘机关键点坐标估计。首先,设

定了7个待识别的关键点,其中4个基础点、3个补充点;然后,建立了挖掘机虚拟模型和虚拟作业环境及背景,并自动化生成合成图像和标签信息,建立了合成图像数据集;之后,在挖掘机作业现场拍摄获得了实际数据集,并使用两类数据集分别训练了识别模型;最后,进行了挖掘实验,对于识别的关键点坐标,合成图像数据集模型得到的平均PCK为97.64%、平均NE为0.0056,高于实际挖掘机数据集模型的结果。本文方法可以在施工现场服务于挖掘机的监控,估计挖掘机的位姿以进行安全区划分、碰撞预警以及生产率分析。此外,相比于传统的拍摄图片建立数据集,使用虚拟模型和场景自动化建立合成图片数据集是一种低成本且便捷的方案,还可以进一步提高识别精度。本文方法尚存在一些局限:首先,只识别了工作装置的关键点,没有识别挖掘机机身,在进行安全区划分、碰撞预警时可能会缺少机身的位姿信息;其次,虽然本文方法可以处理部分遮挡的情况,但当单目相机拍摄过程中距离过远或挖掘机被严重遮挡时,会无法正确识别出2D点或识别误差过大。

参考文献:

- [1] 于向军, 槐元辉, 姚宗伟, 等. 工程车辆无人驾驶关键技术[J]. 吉林大学学报:工学版, 2021, 51(4): 1153-1168.
Yu Xiang-jun, Huai Yuan-hui, Yao Zong-wei, et al. Key technologies in autonomous vehicle for engineering[J]. Journal of Jilin University(Engineering and Technology Edition), 2021, 51(4): 1153-1168.
- [2] 方成, 于盛鑫, 李永刚, 等. 基于深度学习的土木工程计算机视觉健康监测[J]. 同济大学学报:自然科学版, 2024, 52(2): 213-222.
Fang Cheng, Yu Sheng-xin, Li Yong-gang, et al. Deep learning-based computer vision for health monitoring in civil engineering[J]. Journal of Tongji University(Natural Science), 2024, 52(2): 213-222.
- [3] Yao Z W, Zhao S C, Tan X D, et al. Real-time task-oriented continuous digging trajectory planning for excavator arms[J]. Automation in Construction, 2023, 152: No. 104916.
- [4] Park J, Chen J D, Cho Y K. Self-corrective knowledge-based hybrid tracking system using bim and multimodal sensors[J]. Advanced Engineering Informatics, 2017, 32: 126-138.
- [5] Vahdatikhaki F, Hammad A, Siddiqui H. Optimization-based excavator pose estimation using real-time location systems[J]. Automation in Construction, 2015, 56: 76-92.
- [6] 王太海, 陈建宏, 金俊. 基于挖掘机GNSS精确定位的开采姿态监测系统[J]. 黄金科学技术, 2016, 24(4): 101-106.
Wang Tai-hai, Chen Jian-hong, Jin Jun. Mining attitude monitoring system based on GNSS precise positioning of excavator[J]. Gold Science and Technology, 2016, 24(4): 101-106.
- [7] Yang W J, Zhang X H, Ma H W, et al. Infrared leds-based pose estimation with underground camera model for boom-type roadheader in coal mining[J]. IEEE Access, 2019, 7: 33698-33712.
- [8] Assadzadeh A, Arashpour M, Brilakis I, et al. Vision-based excavator pose estimation using synthetically generated datasets with domain randomization[J]. Automation in Construction, 2022, 134: No. 104089.
- [9] 魏振忠, 冯广堃, 周丹雅, 等. 位姿视觉测量方法及应用综述[J]. 激光与光电子学进展, 2023, 60(3): 135-167.
Wei Zhen-zhong, Feng Guang-kun, Zhou Dan-ya, et al. A review of position and orientation visual measurement methods and applications[J]. Laser & Optoelectronics Progress, 2023, 60(3): 135-167.
- [10] 马伟, 宫乐, 冯浩, 等. 基于视觉的挖掘机工作装置位姿测量[J]. 机械设计与研究, 2018, 34(5): 173-176, 182.
Ma Wei, Gong Yue, Feng Hao, et al. Pose measurement of excavator device based on vision[J]. Machine Design and Research, 2018, 34(5): 173-176, 182.
- [11] 王连明, 吴鑫. 基于姿态估计的物体3D运动参数测量方法[J]. 吉林大学学报:工学版, 2023, 53(7): 2099-2108.
Wang Lian-ming, Wu Xin. Method for 3D motion parameter measurement based on pose estimation[J]. Journal of Jilin University(Engineering and Technology Edition), 2023, 53(7): 2099-2108.
- [12] 吴昊. 基于合成数据集的图像处理深度学习方法研究[D]. 兰州: 兰州大学信息科学与工程学院, 2021.
Wu Hao. Research on deep learning methods for image processing based on synthetic datasets[D]. Lanzhou: School of Information Science and Engineering, Lanzhou University, 2021.
- [13] Kim J, Kim D, Lee S, et al. Hybrid DNN training using both synthetic and real construction images to overcome training data shortage[J]. Automation in Construction, 2023, 149: No. 104771.

- [14] 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: No. 103482.
- [15] Liang C J, Kamat V R, Menassa C M. Real-time construction site layout and equipment monitoring[C]//Construction Research Congress, New Orleans, USA, 2018: 64-74.
- [16] Zhang S B, Zhang L J. Construction site safety monitoring and excavator activity analysis system[J]. *Construction Robotics*, 2022, 6: 151-161.
- [17] 郭晓新, 李佳慧, 张宝亮. 基于高分辨率网络的视杯和视盘的联合分割[J]. *吉林大学学报:工学版*, 2023, 53(8): 2350-2357.
- Guo Xiao-xin, Li Jia-hui, Zhang Bao-liang. Joint segmentation of optic cup and disc based on high resolution network[J]. *Journal of Jilin University(Engineering and Technology Edition)*, 2023, 53(8): 2350-2357.
- [18] Tang J Y, Zhang X, Wong P K Y, et al. Method on pose estimation of excavators based on onboard depth camera[J]. *IOP Conference Series: Earth and Environmental Science*, 2022, 1101: No. 072005.
- [19] Lin T Y, Maire M, Belongie S, et al. Microsoft coco: common objects in context[J]. *The 13th European Conference on Computer Vision*, 2014, 8693: 740-755.
- [20] Ionescu C, Papava D, Olaru V, et al. Human3.6M: large scale datasets and predictive methods for 3d human sensing in natural environments[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, 36(7): 1325-1339.
- [21] Wen L Y, Kim D, Liu M Y, et al. 3D excavator pose estimation using projection-based pose optimization for contact-driven hazard monitoring [J]. *Journal of Computing in Civil Engineering*, 2023, 37: 1-15.
- [22] 郑义桀, 罗健欣, 陈卫卫, 等. 基于 Unity3D 三维多视角虚拟数据集构建[J]. *计算机技术与发展*, 2023, 33(5): 173-179.
- Zheng Yi-jie, Luo Jian-xin, Chen Wei-wei, et al. 3D multi-view virtual dataset construction based on Unity3D[J]. *Computer Technology and Development*, 2023, 33(5): 173-179.
- [23] Luo H, Wang M Z, Wong P K Y, et al. Full body pose estimation of construction equipment using computer vision and deep learning techniques[J]. *Automation in Construction*, 2020, 110: No. 103016.
- [24] Torres C, Roberts D, Golparvar F M. Synthesizing pose sequences from 3D assets for vision-based activity analysis[J]. *Journal of Computing in Civil Engineering*, 2021, 35: No. 04020052.
- [25] Tian Z H, Yu Y, Xu F, et al. Dynamic hazardous proximity zone design for excavator based on 3D mechanical arm pose estimation via computer vision[J]. *Journal of Construction Engineering and Management*, 2023, 149: 1-17.
- [26] Lou H T, Duan X H, Guo J M, et al. DC-YOLOv8: small-size object detection algorithm based on camera sensor[J]. *Electronics*, 2023, 12: No. 2323.
- [27] Mahmood B, Han S, Seo J. Implementation experiments on convolutional neural network training using synthetic images for 3D pose estimation of an excavator on real images[J]. *Automation in Construction*, 2022, 133: No. 103996.