

文章编号: 2095-2163(2023)04-0111-07

中图分类号: TH-39

文献标志码: A

深度学习行为识别的挖掘机生产效率监控系统

陈雪健¹, 秦水介², 白忠臣², 郭媛君³, 杨之乐³

(1 贵州大学 大数据与信息工程学院, 贵阳 550025; 2 贵州大学 贵州省光电子技术及应用重点实验室, 贵阳 550025;

3 中国科学院深圳先进技术研究院, 广东 深圳 518055)

摘要: 为了对大型施工设备的生产率进行准确的实时监测和分析, 提出了一种基于计算机视觉的设备识别和生产率监控系统。该方法利用 Faster-RNN 对视频序列中的挖掘机和卡车进行识别检测, 然后通过 ASFormer 对挖掘机动作进行动作分割。基于行为分割的结果对挖掘机的工作效率进行计算。比较了人工计算与行为分割网络的准确度的差距, 对 2 台不同的挖掘机进行动作的识别和分析, 结果显示平均准确度分别为 91.7% 和 94.8%。因此, 当多个场景并行处理时, 可以大大节省人力成本, 验证了本方法的有效性和实用性。本研究为慧工地的数字化管理提供了有效的技术基础, 开辟了计算机视觉应用的新场景, 提出了一种新的计算挖掘机生产效率的计算因子, 更适合于通过统计挖掘机动作来进行效率计算。

关键词: 深度学习; 机器视觉; 行为分割; 特征提取; 视频检测; 生产率检测; 监控系统

Productivity analysis system of earthmoving excavator based on deep learning action recognition

CHEN Xuejian¹, QIN Shuijie², BAI Zhongchen², GUO Yuanjun³, YANG Zhile³

(1 College of Big Data and Information Engineering, Guizhou University, Guiyang 550025, China;

2 Guizhou Key Laboratory for Photoelectron Technology and Application, Guizhou University, Guiyang 550025, China;

3 Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen Guangdong 518055, China)

[Abstract] In order to accurately monitor and analyze the productivity of large-scale construction equipment in real time, a method of equipment identification and productivity analysis based on computer vision is proposed. The method consists of two stages: firstly, the excavator and truck in the video sequence are identified and detected by Faster-RNN, and then the excavator action is segmented by ASFormer. The work efficiency of excavators is calculated based on the results of recognition and behavior segmentation. For two different excavators, the average accuracy is 91.7% and 94.8% respectively. When multiple scenarios are processed in parallel, the labor cost can be greatly saved, which verifies the effectiveness and practicability of this method. Therefore, this study provides an effective technical basis for the digital management of hui site, opens up a new scene of computer vision application, and puts forward a new fast calculation of excavator production efficiency, which can realize the real-time calculation of the work efficiency of construction equipment. It can play an important role in improving the production efficiency of construction site equipment and ensuring the safety of staff.

[Key words] deep learning; computer vision; action segmentation; feature extraction; video processing; productivity monitoring; monitoring system

0 引言

在现代施工现场, 已经基本实现了监控摄像头的全面覆盖。因此, 图像和视频数据不仅记录了非法入侵施工现场的信息, 还包括各种设备和建筑工

人的活动。但是这些视频数据的提取利用并不完善。近年来, 机器视觉领域取得巨大研究进展^[1-2], 相关机器视觉模型被广泛应用在各个方面, 例如能见度测定^[3]、基于手势识别的手部康复系统^[4]、辅助驾驶^[5]以及用图象识别的方法代替电阻等传统

基金项目: 国家自然科学基金-地区项目(61865002, 62065002)。

作者简介: 陈雪健(1997-), 男, 硕士研究生, 主要研究方向: 电路与系统; 秦水介(1963-), 女, 博士, 教授, 主要研究方向: 纳米量子点及其功能纳米结构的制备及在生物医学中的应用; 微型电子机械系统(MEMS)的微制造及性能; 白忠臣(1979-), 男, 博士, 教授, 主要研究方向: 纳米传感器; 郭媛君(1985-), 女, 博士, 副研究员, 主要研究方向: 电力大数据、故障检测、人工智能算法; 杨之乐(1987-), 男, 博士, 副研究员, 主要研究方向: 智能优化调度、电动汽车与新能源系统、智能制造与机器视觉。

通讯作者: 秦水介 Email: shuijie_qin@sina.com

收稿日期: 2022-08-04

火焰检测技术^[6]。近年来,Transformer 在自然语言处理、计算机视觉等人工智能领域取得了可观成果。并且因为其强大的全局信息捕捉能力被广泛应用在了计算机视觉研究中。土方挖掘机作为一种全地形施工设备,现已成为了土方作业中不可或缺的重要设备。并且土方作业运营成本大部分都来自于大型施工设备,因此对于施工设备进行监控并辅以相应调整是提高设备生产率的最佳手段^[7-11]。

本研究的主要贡献体现在以下几个方面。首先,将计算机视觉这个领域的先进技术(目标识别、动作分割)结合起来,应用于施工现场施工设备的活动分析,自动监控与分析施工现场大型设备的工作。与最近的研究相比,本文提出的模型基于长视频的动作分割模型,将未裁剪的长视频直接应用于训练和识别。此外,本文还提出了一种根据挖掘机

动作计算挖掘机工作效率的方法。该方法更适合通过挖掘机动作来计算挖掘机的工作效率,同时,该监控系统还可以帮助施工管理人员直观地了解土方挖掘机在一段时间内的详细活动信息,管理人员可以使用自动化监控手段更好地调度施工设备,促进项目的开展与实施。

1 研究框架

本文提出的深度学习方法主要包含2个阶段,框架如图1所示。由图1可知,首先使用Faster R-CNN^[12]模型对挖掘机视频进行处理,检测视频中是否存在卡车或挖土机。然后,采用三维卷积神经网络提取其时序特征,并将结果输入到ASFormer^[13]模型中,对挖掘机的动作序列进行训练和识别。最后,计算挖掘机土方作业的工作效率。

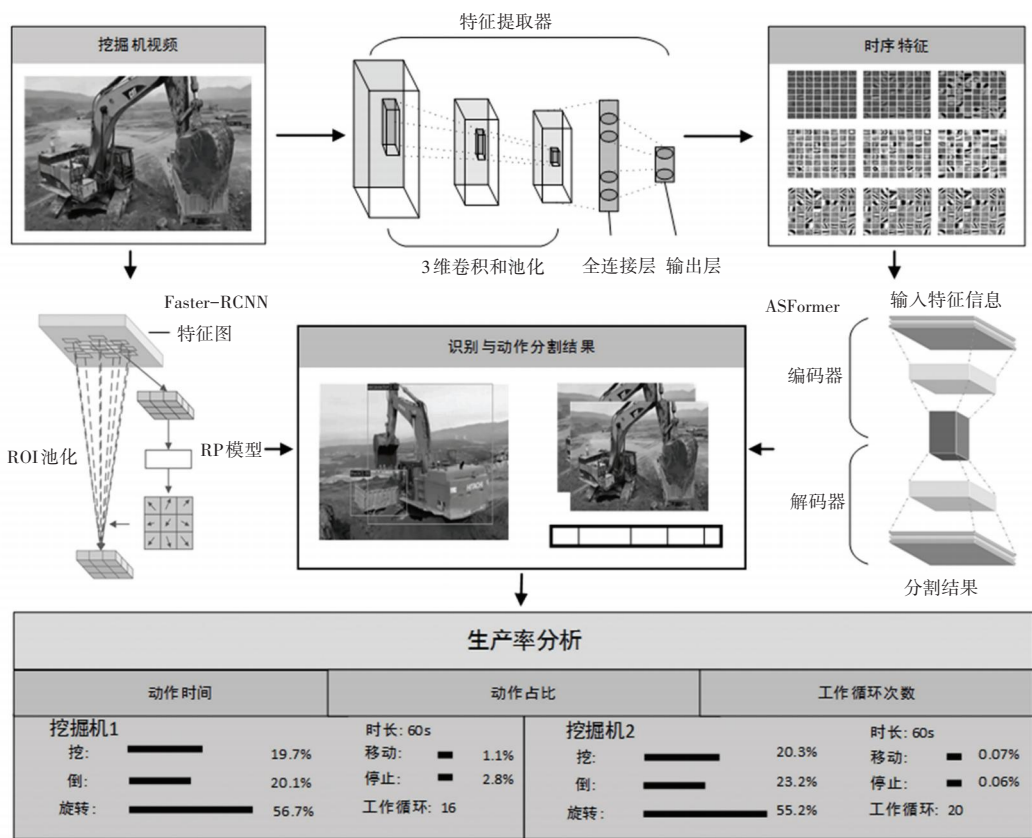


图1 框架流程图

Fig. 1 Workflow of the proposed framework

1.1 挖掘机与卡车检测

这个过程的目的确定挖掘机和卡车在视频的每一帧中的位置。为了识别视频中的挖掘机和卡车,本文采用了快速循环卷积神经网络(Faster R-CNN)该模型由Ren等学者^[12]开发,并已广泛应用于各个领域。在建筑领域,以往的研究已经证明了该模型在

恶劣施工条件下检测施工工人和设备的巨大潜力^[14]。

Faster R-CNN中使用了卷积神经网络(CNN)。该网络使用卷积层、激活层和池化层堆叠来提取图像的特征图,随后的Region Proposal Networks(RPN)和全连接层共享该部分提取的特征图。其

中, RPN 网络用于生成区域特征。随后, RoI 池化层的功能是通过整合信息提取出 proposal feature map, 发送给后续的全连接层, 并对目标类别做出判断。最后, 利用 proposal feature map 计算类别, 再次获得复选框的最终精确位置。

另外, 在实际的施工现场, 土方挖掘机在作业过程中会有各种动作, 因此在识别挖掘机的过程中, 要考虑到工作中挖掘机的各种动作识别。本文采用 3 种不同的长宽比(1:1、1:2、2:1)进行标记, 使所提出的模型能够检测到动作不断变化的土方挖掘机。同时, 为了提高模型的鲁棒性, 本文从现场不同角度采集了不同外观土方挖掘机的工作视频。除此之外, 还使用了常用的数据增强方法, 在标记图像时进行缩放、旋转和镜像, 以扩充数据集。

1.2 改进的动作分割网络

在挖掘机的动作分割识别过程中, 本研究应用了 ASFormer 模型。由于在施工现场获得的视频是未经编辑的长视频, 一个视频包含多个需要识别的动作。并且土方挖掘机的工作状态总是一系列动作序列, 所以挖掘机的动作识别与动作分割任务目标高度重合。考虑到 ASFormer 模型应用了 Transformer 结构, 其动作分割的性能得到了很大的提高。ASFormer 的模型如图 2 所示。由图 2 可知, ASFormer 使用带有编码器-解码器结构的 Transformer 模型。编码器首先根据预先提取到的视频的时序特征预测每帧的初始动作概率。然后由连续的解码器对初始预测结果进行优化。

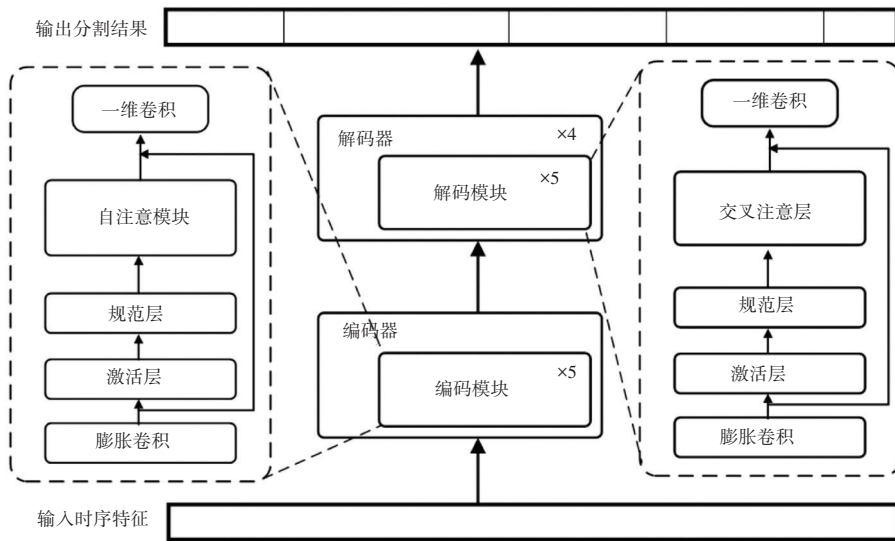


图 2 动作分割网络

Fig. 2 Action segmentation network

编码器的输入为预先提取的视频的时序特征, 大小为 $T \times D$, 其中 T 为视频帧数, D 为预先设定的特征维数。编码器在第一层使用全连接层, 输出用作后面一系列编码器块的输入。此后再将编码器块的输出送入全连接层以获得初始预测结果。每个编码器块由前馈层和带有残差连接的单头自注意层组成。

对于解码器, 首先其输入来自编码器的初始预测。解码器的第一层由一个调整输入大小的全连接层开始, 随后是一系列解码器块。解码器的结构类似于编码器。前馈层为时序卷积, 交叉注意层为分层结构。与编码器结构中的自注意层相比, 交叉注意层的主要区别在于查询 Q 和键值 K 是由编码器的

输出和上层的输出串联获得的。这样做的目的是为了了解码器中的每个位置都可以参与编码器的初始预测结果的细化。最后通过堆叠这些单独的解码器来细化结果。为了逐步减少外部信息对预测结果的影响, 避免误差积累, 在解码器接收每一个输入之前, 增加加权残差连接前馈层和交叉注意层:

$$out = feed_forward(x) \tag{1}$$

$$out = \alpha * cross_att(out) + out \tag{2}$$

其中, out 为前馈层输出; x 为前馈层输入; α 为交叉注意层系数。对第一个解码器设定 $\alpha = 1$, 而后对其余的解码器 α 则采用了指数增加策略。

1.3 挖掘机工作效率计算

在施工过程中, 挖掘机通常与大卡车等其他施

工设备一起工作。例如,在土方挖掘工作中,挖土机挖掘土壤,并将其装入卡车的铲斗中。当卡车满载时,挖掘机将土壤移至倾卸区,继而在装货区等待重新装货。如果单独研究土方挖掘机,其工作过程主要分为“挖掘”、“摆动”和“装载”。因此,将“挖掘-旋转-装载-旋转”定义为挖掘机的一个工作循环。

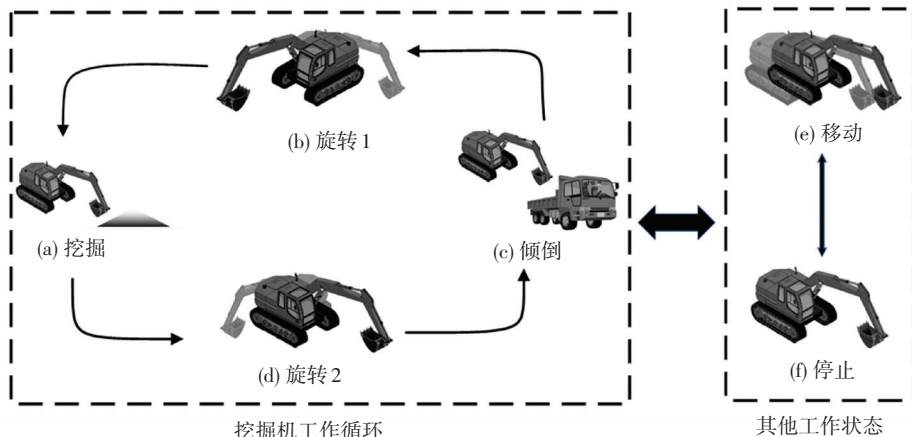


图3 挖掘机工作状态及流程

Fig. 3 Working state and process of the excavator

在土方挖掘过程中,当挖掘机铲斗载荷一定时,挖掘机的工作效率可以用单位时间内的工作循环次数来表示。在工作循环中,挖掘机的挖掘和倾卸对挖掘机的工作效率贡献最大。除此之外,挖掘机的工作效率还受到摆动动作时间占比例的影响。例如:当摆动时间占比较高时,其效率会受到影响而降低。经过以上分析,本文将挖掘机工作效率计算公式定义如下:

$$\theta = 1 - e^{1 - \frac{t_c}{t_s}} \quad (3)$$

$$EWP = \frac{c}{t_c} \times \theta \quad (4)$$

其中, c 表示工作循环数; t_c 、 t_s 分别表示工作循环总时间和旋转动作总时间; EWP 表示挖掘机的生产率,即单位时间的工作循环数。

2 实验过程和结果

2.1 数据集

为了训练深度学习模型,本研究使用普通光学相机和智能手机采集实际土方开挖现场的视频流数据,采集的视频数据可以记录挖土机的所有作业类型,所收集的视频数据示例如图4所示。为了避免光照和拍摄角度对结果的影响,实验人员在不同的时间和不同的角度进行视频采集工作。在视频采集过程中,使用了6个普通光学摄像头和2部智能手

另外,土方挖掘机还具有“移动”和“静止”两种作用状态。当挖掘机开始工作时,“移动”和“静止”对工作效率几乎没有贡献,因此本研究将2个动作列为其他动作,不在工作循环中,当计算挖掘机工作效率时只考虑工作循环中的动作类型。挖掘机的工作状态如图3所示。

机,总共收集了97 min的视频数据。经过选择,最终选取分辨率为380 * 640、帧率为30 fps、总图像帧数约80 000帧的44 min视频数据进行分析。对于识别任务的数据,本研究进一步将数据分为训练集数据(约60 000片)和测试集数据。在动作分割模型的训练过程中,将选出的44 min视频大致分成4部分,其中3/4作为训练集数据,其余作为测试集数据。

2.2 识别与动作分割结果

本研究训练的深度学习模型在挖掘机和卡车识别任务中表现良好,准确率为93.81%。识别结果如图5所示。对于挖掘机的活动识别部分,本研究训练的模型在测试集中达到了93.25%, F_1 值分别为0.87、0.856、0.75。测试集中某段视频的分割结果如图6所示。在图6中,从上到下展示了视频实例的真实分割、分割结果以及帧动作类别的可能性。另外,挖掘机和卡车识别模型可以在大多数环境中准确识别和检测这2种类型的施工设备。在分析挖掘机的活动时,动作分割模型能够准确识别其各种工作状态,包括:挖掘、倾卸、摆动、停止、移动。

本文提出的模型具有良好的性能,但仍存在一些不足。例如,在挖掘机、卡车识别过程中,若施工设备被大面积遮挡,识别就会产生较大的误差。在施工现场,这种遮挡通常发生在设备之间,模型不能对这种情景进行有效的判断,此时可以通过调整光学摄像机的位置来解决。



图 4 数据集示例

Fig. 4 Examples of the dataset



图 5 目标检测结果

Fig. 5 Excavator and truck identification results

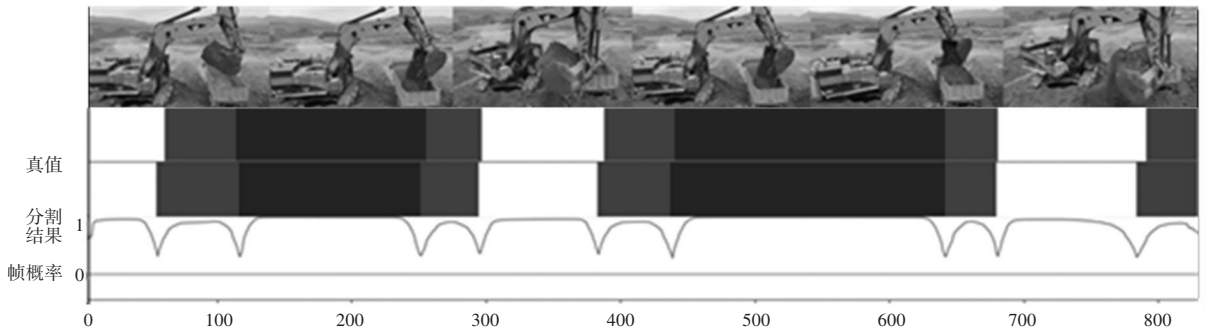


图 6 挖掘机动作分割结果

Fig. 6 Excavator action segmentation results

2.3 挖掘机生产率计算

挖掘机的生产率通过分析计算挖掘机动作分割的结果得来。挖掘机活动识别的结果可以直观地展示其活动的顺序。模型输出的识别结果对应于每个视频帧,因此可以通过视频的帧率来计算每个动作的持续时间,以获得各动作持续的准确时间以及次

数。将这些数据应用到 1.3 节提出的方法中,能够计算挖掘机的工作效率,从而可以进一步评价其对整个建设项目的贡献。2 台挖掘机的工作效率分析结果见表 1。由表 1 的计算结果可知,挖掘机 1 的工作效率为 1.18 个工作循环/min,挖掘机 2 的工作效率为 1.22 个工作循环/min。可以看出,挖掘机 2

的效率要高于挖掘机1的效率。表2所展示的是对挖掘机的动作进行人工识别结果,以及对挖掘机生产率的估算。通过将表1与表2内容对比,可以看出通过视频自动处理得到的结果与人工识别精度相

当,当需要监测的设备较多时,借助机器视觉的方法可以大大节省人力成本,同时也能保证分析的准确度。

表1 挖掘机生产率分析结果

Tab. 1 Analysis results of excavator productivity

设备 统计项目	挖掘机 1		挖掘机 2	
	动作持续时间/s	动作次数	动作持续时间/s	动作次数
挖掘	85	16	110	20
倾倒	87	16	126	20
旋转	243	34	300	42
移动	5	4	4	1
停止	12	5	3	1
工作循环次数	min(挖掘次数,倾倒次数) = 16		min(挖掘次数,倾倒次数) = 20	
生产率/(次·分钟 ⁻¹)	1.18		1.22	
准确率/%	91.7		94.8	

表2 挖掘机动作人工识别结果

Tab. 2 Manual segmentation results of excavator action

设备 统计项目	挖掘机 1		挖掘机 2	
	动作持续时间/s	动作次数	动作持续时间/s	动作次数
挖掘	87	16	106	20
倾倒	98	18	123	20
旋转	235	34	304	42
移动	0	0	5	1
停止	12	1	5	1
工作循环次数	min(挖掘次数,倾倒次数) = 16		min(挖掘次数,倾倒次数) = 20	
生产率/(次·分钟 ⁻¹)	1.25		1.19	

3 结束语

本文提出了一种基于计算机视觉的挖掘机作业和生产率自动化监测框架,该框架集成了基于长视频的挖掘机检测、动作识别和生产率计算模块。检测模块识别挖掘机在视频帧中的位置。动作分割模块对长视频中的挖掘机活动进行分割。最后,根据识别出的挖掘机活动信息计算挖掘机的生产率。本研究对施工监控视频进行识别分析,并将分析结果与人工分析结果进行了对比。对生产率分析的结果表明,应用生产率监测系统得到的挖掘机的工作效率的准确率与人工分析得到的准确率的误差很小,当施工工地监测设备较多时,机器视觉处理视频数据的速度比人工速度快,能够大大降低施工现场管理的人工成本,提高设备使用效率,降低安全风险,为智慧工地数字化提供了可靠的技术途径。

参考文献

- [1] 杜嘉豪,黄孝慈,邢孟阳. 基于机器视觉的停车场智能寻车系统[J]. 智能计算机与应用, 2022, 12(05): 102-106.
- [2] 李伟雄,闫加俊,杨丽丽. 利用图像识别技术的能见度观测研究[J]. 自动化仪表, 2021, 42(06): 63-67.
- [3] 宋婉娟,张剑. 融合回归网络和多尺度特征表示的实时行人检测[J]. 电子测量与仪器学报, 2018, 32(07): 15-20.
- [4] 徐胜,刘志诚,涂鑫涛,等. 基于手势识别的手部康复智能评估算法研究[J]. 自动化仪表, 2021, 42(12): 23-27, 32.
- [5] 王永宏. 基于机器视觉的汽车先进驾驶辅助系统中行人检测技术研究[D]. 杭州:浙江大学, 2019.
- [6] 戚凯旋,朱凌云. 基于图像处理的燃气锅炉燃烧状态检测研究[J]. 自动化仪表, 2021, 42(11): 34-38.
- [7] SLATON T, HERNANDEZ C, AKHAVIAN R. Construction activity recognition with convolutional recurrent networks [J]. Automation in Construction, 2020, 113: 103138.
- [8] NAGHSHBANDI S N, VARGA L, HU Y. Technologies for safe and resilient earthmoving operations; A systematic literature review [J]. Automation in Construction, 2021, 125: 103632.

(下转第 121 页)