

基于 Yolopose 的挖掘机检测与工作状态识别^①



黄 健, 赵小飞, 王 虎, 胡其胜

(西安科技大学 通信与信息工程学院, 西安 710600)

通信作者: 赵小飞, E-mail: 760733186@qq.com

摘 要: 针对光缆、高压油气管道等地下基础设施周边容易受到挖掘机的野蛮入侵问题. 本文提出了一种结合 Yolopose 和多层感知机的挖掘机检测与工作状态判别方法. 首先, 设计了基于 Yolopose 的挖掘机 6 点姿势的提取网络 Yolopose-ex; 其次, 利用 Yolopose-ex 模型提取视频中挖掘机工作姿态的变化信息, 构建了挖掘机的工作状态特征向量 (MSV); 最后, 利用深度学习算法多层感知机 (multilayer perceptron, MLP) 分析了视频中的挖掘机的工作状态. 实验结果表明, 所提出的方法克服了复杂背景难以识别的问题, 对挖掘机工作状态识别准确率达到了 96.6%, 具有较高的推理速度和泛化能力.

关键词: 挖掘机; Yolopose; 姿势估计; 工作状态识别

引用格式: 黄健, 赵小飞, 王虎, 胡其胜. 基于 Yolopose 的挖掘机检测与工作状态识别. 计算机系统应用, 2024, 33(2): 299-307. <http://www.c-s-a.org.cn/1003-3254/9399.html>

Excavator Detection and Working State Discrimination Based on Yolopose

HUANG Jian, ZHAO Xiao-Fei, WANG Hu, HU Qi-Sheng

(College of Communication and Information Technology, Xi'an University of Science and Technology, Xi'an 710600, China)

Abstract: The surrounding areas of underground infrastructure such as optical cables and high-pressure oil and gas pipelines are vulnerable to brutal invasion by excavators. This study proposes an excavator detection and working state discrimination method combined with Yolopose and a multilayer perceptron. First, the Yolopose-ex extraction network based on Yolopose's six-point posture of the excavator is designed. Secondly, the Yolopose-ex model is utilized to extract the change information of the excavator's working posture in the video, and the working state feature vector (MSV) of the excavator is constructed. Finally, the multilayer perceptron (MLP) is adopted to analyze the working status of the excavator in the video. The experimental results show that the proposed method overcomes the problem of difficult discrimination of complex backgrounds, and the accuracy of the identification of the working state of the excavator reaches 96.6%, which has a high reasoning speed and generalization ability.

Key words: excavator; Yolopose; pose estimation; working state discrimination

近年来, 伴随我国经济的高质量发展, 地下高压油气管道、光缆等基础设施建设大幅度增加, 对地下基础设施入侵检测也提出了更高的要求. 地下国防光缆, 关系着国家命脉和人民的生命安全; 地下高压油气管道如果遭到野蛮入侵, 极易发生爆炸、火灾等重大安

全事故. 所以, 确保地下基础设施稳定运行至关重要.

随着现代化的进程加快, 挖掘机这种工程器械使用量也与日俱增, 使得地下基础设施受到挖掘机入侵事件频频发生. 传统的巡检方法通过人工对监控区域进行检测, 发现外力入侵目标, 派出工作人员到现场进

① 基金项目: 陕西省重点研发计划 (2023-YBGY-255)

收稿时间: 2023-07-12; 修改时间: 2023-08-11; 采用时间: 2023-09-26; csa 在线出版时间: 2024-01-02

CNKI 网络首发时间: 2024-01-03

行制止警告. 这种人工监督的方式效率低, 而且容易出现漏检现象. 随着维护视频监控规模持续扩大, 视频中心需要同时监控数百上千的监控点, 人工监控的方式很难应对. 因此, 迫切需要一种智能化的视频监控算法自动完成挖掘机的入侵检测与工作状态判别, 缓解人工监控的压力, 保障地下基础设施的安全性.

在国内, 目前利用计算机图像视觉识别技术进行挖掘机和判断挖掘机工作状态的相关研究相对较少, 李源等人^[1]利用挖掘机图像的颜色信息对挖掘机特征进行了提取, 由于该方法是基于 RGB 颜色特征来进行分割的, 在复杂的实际环境中效果比较差. 林焕凯等人^[2,3]基于混合可变形部件模型 (DPM) 的传统目标检测算法对挖掘机图像信息进行了研究, 该方法识别挖掘机的准确率达到 88.40%. 但该方法检测速度较慢, 精度低, 无法达到实际应用, 同时, 他们还提出了一种基于局部二值特征的物体形状回归算法来提取挖掘机的关键点, 构建了一种挖掘机运动状态特征描述子, 并利用机器学习算法支持向量机 (SVM) 对挖掘机运动状态进行分类, 达到了较高的准确率.

上述方法都是基于传统图像处理方法对挖掘机进行识别和运动状态判别, 处理速度慢、对环境的适应性不强, 容易出现漏检、错检. 如今, 人工智能算法与计算机视觉技术^[4,5]的研究如火如荼, 基于深度学习的智慧安防得到了广泛的应用. YOLO 算法^[6-8]因其出色的性能而备受关注, 其中最新的 YOLOv7 算法的性能最佳, 推理速度快, 鲁棒性强, 在实际工程项目取得了举足轻重的位置, 并且基于单阶段目标检测器的人体姿势模型 Yolopose^[9], 在人体姿势估计领域也取得了不错的效果. 目标检测领域的主要研究进展都无缝传递到姿态估计问题上, 消除了非标准的后处理^[10,11], 使得人体姿势估计可以达到端到端的训练, 极大地提高了姿势估计的推理速度.

本文以 Yolopose 为基础, 以挖掘机这种工程机械作为研究对象, 构造了 6 点挖掘机姿势图, 提出一种基于改进 Yolopose 算法的挖掘机检测与姿势估计算法 Yolopose-ex. 创新性地将 Yolopose 与注意力机制^[12,13]进行融合, 并且采用了 MobileNet^[14-16]所提出的深度可分离卷积替换原有的普通卷积. 应用以上方法可以显著提升模型的检测速度, 降低了模型的参数量、提高了模型对挖掘机的关注能力, 使得在复杂环境下模型具有很强的鲁棒性. 利用改进后 Yolopose-ex 模型预测

出挖掘机工作的姿势信息, 构建了挖掘机工作状态特征向量, 训练了基于神经网络搭建的多层感知机 (MLP) 对视频中的挖掘机状态进行判断.

1 挖掘机 6 点姿势图

本文的研究对象是工程建设当中经常使用的履带式挖掘机, 其工作行为主要由通过液压装置控制的铲斗、斗杆、动力臂所形成, 三者之间由铰轴相互连接. 所以如果想分析出挖掘机的工作状态, 必须先提取出挖掘机工作装置的关键点, 然后通过对关键点进行分析来判别挖掘机的工作状态. Yolopose 是一种人体关键点检测模型, 通过该算法可以提取人体关键点, 进行人体估计. 所以可以借鉴该算法来提取挖掘机关键点, 对挖掘机进行姿势进行估计. 核心思想为: 假设决定挖掘机运动有 n 个点为 $E = [x_1, y_1, \dots, x_n, y_n]^T$, 其中, x_n 和 y_n 对应挖掘机在图像中的横纵坐标. 通过训练大量的预标注的图片, 训练出能预测挖掘机关键点的模型. 如图 1 中的 6 个关键点 A、B、C、D、E、F 可以决定挖掘机的姿势, 本文定义其为挖掘机 6 点姿势图. 对 6 点姿势图进行分析来实现对挖掘机运动状态的判别.



图 1 挖掘机工作装置示意图

2 Yolopose 的姿势估计算法

Yolopose 算法是一种检测速度快、准确率都较高的姿势估计算法. 网络总共有 4 个部分, 分别为输入 (Input)、骨干网络 (Backbone)、特征融合网络 (Neck)、检测推理 (Head).

2.1 输入

在图像输入卷积神经网络前, 采用了马赛克增强 (Mosaic)、自适应锚框 (Autoanchor) 以及图片尺寸缩放等数据技术, 提高了模型的鲁棒性, 降低了模型对图像的敏感度, 增强了网络的泛化能力. 这些技术的结合有效地提升了模型的训练效果和性能表现.

2.2 骨干网络

Backbone 是 YOLOv7 目标检测算法中的主要特征提取网络,它由多个模块堆叠而成,包括 CBS、ELAN、MP、SPPCSPC 等模块. ELAN 及其变形结构是一种高效的多分支融合模块,其中,ELAN 模块主要应用于骨干网络,而 ELAN-W 应用在 Neck 中进行尺度融合. SPPCSPC 是加强特征提取模块,模块会对每个特征层进行空间金字塔池化操作,以获取不同的分辨率和语义信息,提供丰富的特征表示,用于预测关键点的位置和姿态角度.

2.3 特征融合模块

在 Neck 特征融合模块中,采用了 PANet 结构^[17]. 该模块接受骨干网络提取到的 4 种不同尺度的特征图,自上而下的路径将深层特征图中的语义信息传递到浅层特征图中,实现语义信息的融合. 而自下而上的路径,将浅层特征图中的细节信息传递到深层特征图中,实现细节信息的融合. 这种双向的信息传递和融合可以提高特征图的表达能力,使得模型能够同时捕捉到图像的语义信息和细节信息.

2.4 预测头

网络输出 4 个尺度的预测数据,4 个尺度分别拥有不同的感受野,预测大、较大、中、小物体的目标框、关键点坐标、置信度. 单阶段检测网络在 COCO 数据集上主要关注 80 个类别,每个 Anchor 需要预测 85 个值,其中包括 80 个类别的置信度、边界框坐标和物体准确度. 每个网格位置上都有 4 种不同形状的 Anchor 与之对应. 在挖掘机姿态估计任务中,本文将其视为一个单类物体检测问题,将负责预测 80 个类别置信度的信息转化为预测关键点的坐标和置信度. 每个挖掘机有 6 个关键点,每个关键点都有一对位置和置信度: $\{x, y, c\}$, 总共有 18 个关键点. 因此,对于每个 Anchor,需要预测 18 个元素来表示关键点,以及 6 个元素来表示边界框,总共需要预测 24 个元素. 这样的设计可以有效地关联关键点和边界框,实现对挖掘机姿态的准确预测.

3 改进后的 Yolopose 姿势估计算法

Yolopose 用于估计挖掘机 6 点姿势,参数量大,定位挖掘机关键点能力差,并且针对复杂背景的适应能力不强,容易产生误检,从而使检测受到干扰. 因此,本文首先引入 CBAM 注意力机制来抑制模型的无关信

息,凸显有用特征,降低冗余特征图,增加定位挖掘机 6 点姿势能力;其次,使用了深度可分离卷积替换骨干网络中的 ELAN 模块的普通卷积,极大地降低网络的参数量和计算量. 改进后的 Yolopose-ex 的网络结构模型如图 2 所示.

3.1 CBAM 注意力机制

由于本研究检测环境复杂多样,需要进行自适应特征细化,引入轻量级的注意力机制模块 CBAM 进行改进,可提高对复杂环境的适应性. CBAM^[12]由通道注意力模块(CAM)和空间注意力(SAM)两个模块组成. 给定 $C \times H \times W$ 的特征图, CAM 生成通道注意力图,来提取特征图的通道之间的关系,关注输入通道中哪个通道更有意义,它先通过平均池化和最大池化两个操作来聚合特征图,提取特征图的语义信息,然后将提取的特征向量输入一个共享的全连接层(MLP),最后经过激活函数生成不相同的通道注意力 $M_c(F)$,如图 3 所示,公式归纳为式(1)、式(2).

$$M_c(F) = \sigma(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F))) \quad (1)$$

$$M_c(F) = \sigma(W_1(W_0(F_{\text{avg}}^c)) + W_1(W_0(F_{\text{max}}^c))) \quad (2)$$

与通道注意力机制不同, SAM 的作用是关注哪块区域更有意义,信息更丰富,与通道注意力机制优势互补. 计算空间注意力,首先通过平均池化和最大池化两个操作,然后将二者拼接起来,生成一个高效的特征图描述符,该操作已被证明可以有效地表达信息丰富的区域,然后将拼接后的特征描述符通过卷积来生成一个空间注意力 $M_s(F)$. 如图 4 所示,空间注意力机制可以有效地使哪个区域凸显或抑制,公式归纳为式(3).

$$M_s(F) = \sigma(\text{Conv}([\text{AvgPool}(F), \text{MaxPool}(F)])) \quad (3)$$

对于输入的中间层特征图,顺序执行 channel attention 得到 M_c ,然后再通过 spatial attention 得到 M_s ,具体如图 5 所示,公式归纳为式(4)、式(5). 将其插入到网络当中,可以对挖掘机进行自适应特征细化.

$$F^1 = M_c(F) \otimes F \quad (4)$$

$$F^2 = M_s(F^1) \otimes F^1 \quad (5)$$

多分支融合模块 ELAN 进行了高效的特征融合,同时也产生了冗余特征,因此将 CBAM 注意力机制加入到 ELAN 模块之后,即 Backbone 和 PANet 二者的连接处,具体如图 1 所示,可提高模型的特征提取能力,同时增加了模型对环境的适应能力.

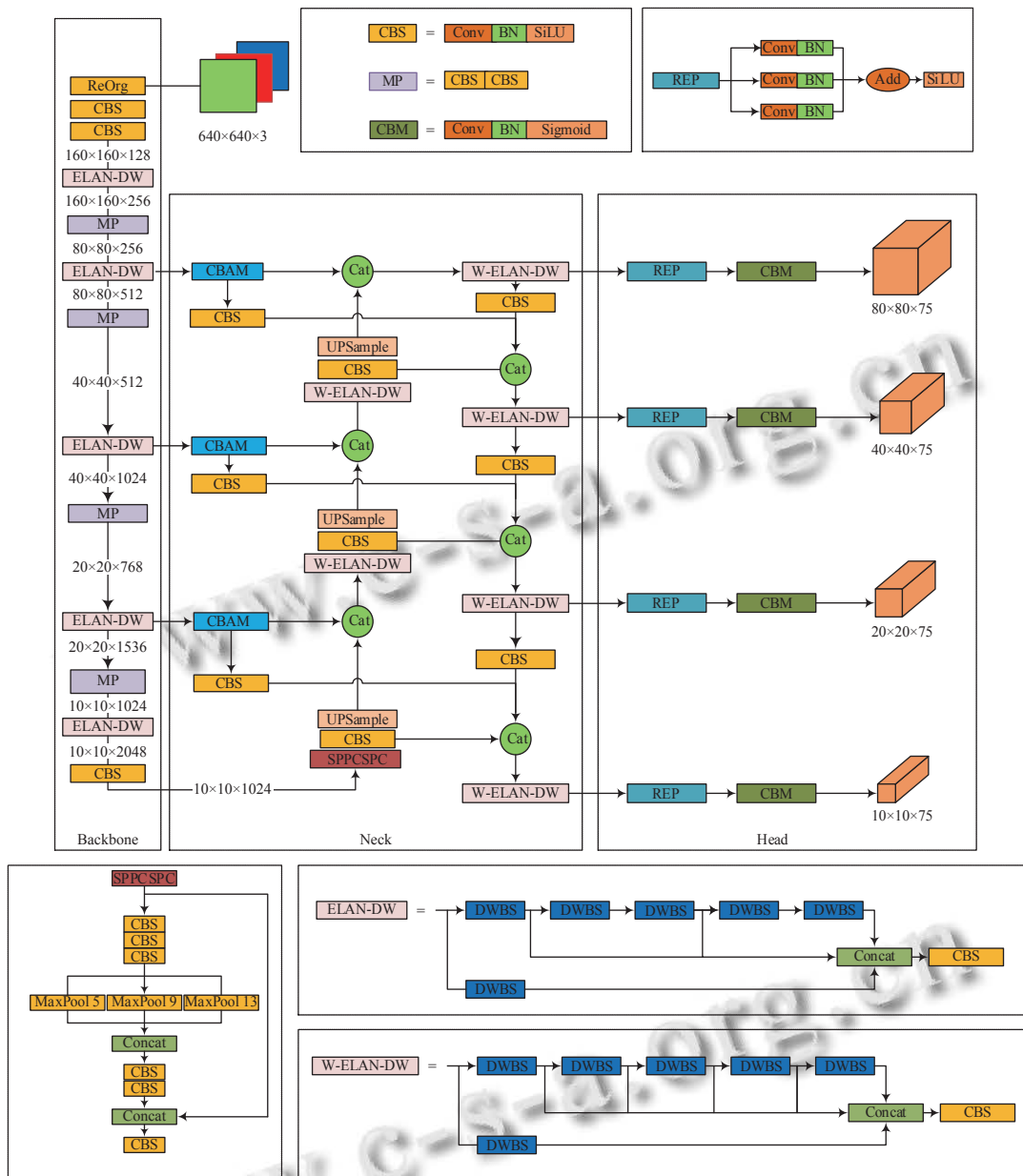


图2 YoloPose-ex 模型

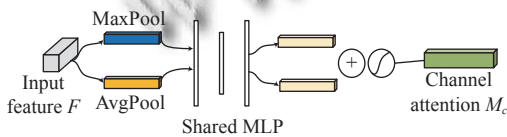


图3 通道注意机制

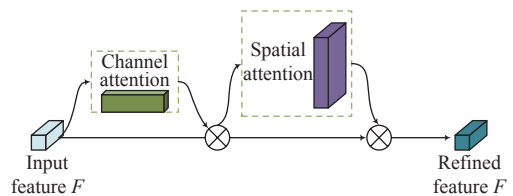


图5 CBAM 注意机制

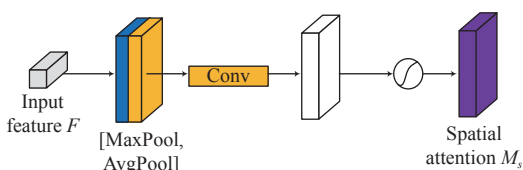


图4 空间注意机制

3.2 深度可分离卷积

深度可分离卷积 (depthwise separable convolution) 是 MobileNet 系列文章^[16-19]中提出的轻量化卷积神经网络的方法. 深度可分离卷积是一种将标准卷积操作

分解为深度卷积和逐点卷积两个步骤的方法. 深度卷积针对输入数据的每个通道进行独立的卷积操作, 而逐点卷积则用于将不同通道之间的信息进行融合. 相对于传统的卷积方法, 深度可分离卷积在模型性能和实时性方面提供了更好的表现. 通过拆分卷积操作并引入逐点卷积, 深度可分离卷积减少了计算量, 并且能够更有效地融合通道间的信息, 从而提高了模型的性能和实时性能.

Yolopose 中 ELAN 模块使用了大量的普通卷积进行多分支特征融合, 占用了大量的参数量和计算量, 采用深度可分离卷积代替普通卷积可以在不破坏 ELAN 高效结构的情况下大幅度降低 ELAN 模块的参数和计算量. 改进后的 ELAN-DW 模块如图 6 所示.

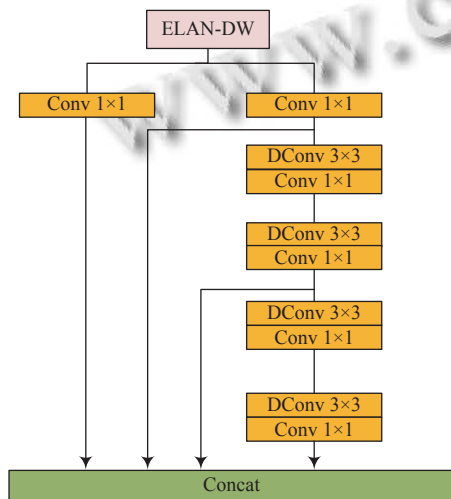


图 6 改进后的 ELAN-DW 模块

4 挖掘机的工作状态识别

4.1 算法实现

挖掘机工作状态识别可以分为两个阶段, 第 1 阶段为通过大量的挖掘机的 6 点姿势图基于 Yolopose-ex 训练出一个挖掘机检测与关键点提取模型, 第 2 阶段使用一阶段的 Yolopose-ex 模型得到挖掘机 6 点姿势图构建运动状态特征向量 MSV, 训练出多层感知机来判断挖掘机的工作状态, 算法流程图如图 7 所示.

4.2 挖掘机工作状态特征向量

通过改进的 Yolopose 挖掘机姿势估计算法, 可以获取挖掘机各个部分的关键点信息. 然而, 仅凭这些信息无法判断挖掘机是否在工作. 因此, 为实现挖掘机工作状态的判别, 本文根据 Yolopose-ex 挖掘机检测与姿

势模型所得到的挖掘机 6 点姿势, 计算出各个关键点连线之间的夹角, 然后根据各部分的角度变化, 来描述挖掘机的工作状态. 如图 1 所示, θ_1 、 θ_2 、 θ_3 、 θ_4 的变化 $\Delta\theta_1$ 、 $\Delta\theta_2$ 、 $\Delta\theta_3$ 、 $\Delta\theta_4$ 决定了挖掘机工作的姿态变化.

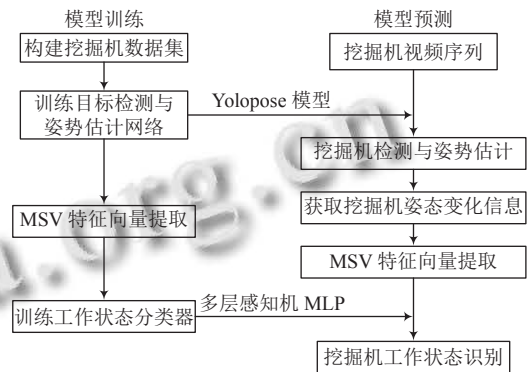


图 7 挖掘机工作状态识别算法

本文采用 $\Delta Angle$ 来表示挖掘机工作姿势变化的信息, 即 θ_1 、 θ_2 、 θ_3 、 θ_4 的变化 $\Delta\theta_1$ 、 $\Delta\theta_2$ 、 $\Delta\theta_3$ 、 $\Delta\theta_4$, 在结合视频中所检测出的挖掘机上一帧和下一帧的目标的长宽比 ΔWH , 形成挖掘机工作状态向量 $MSV^{[3]}$, 可表示为 $[\Delta WH, \Delta\theta_1, \Delta\theta_2, \Delta\theta_3, \Delta\theta_4]$. 以 MSV 作为多层感知机 MLP 的输入, 判别挖掘机的工作状态.

多层感知机是一种常见的人工神经网络模型, 它由多个神经元层组成, 每个神经元层与下一层相连. 通过在训练过程中不断调整网络的权重和偏置, 多层感知机能够学习到挖掘机不同工作状态的表示, 并用于分类任务. 利用这种方法, 可以实现对挖掘机工作状态的准确识别. 其中多层感知机的输入层为 5 维, 隐藏层为 8 维, 输出层为 2 维, 采用 ReLU 为激活函数, 损失函数为平方差损失函数, 反向传递采用梯度下降法, 多层感知机整体结构如图 8 所示.

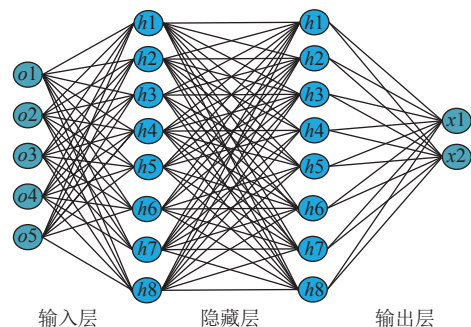


图 8 多层感知机

5 实验与结果分析

5.1 Yolopose 与 Yolopose-ex 对比实验

5.1.1 实验环境

实验环境采用 Ubuntu18.04 系统, GPU 为 NVIDIA GeForce RTX 3080 (24 GB), 处理器为 Intel(R) Xeon(R) CPU E5-2673v3 @2.40 GHz 12 核, 配置环境为 Python 3.8.0、PyTorch 1.8.1、cuda 11.6.

5.1.2 数据集处理

为了评估 Yolopose 和 Yolopose-ex 算法之间的差异以及相关算法的实时性和有效性, 本文采集了一个包含 10 000 张挖掘机工作照片的数据集. 该数据集涵盖了各种复杂的工作地点和背景情况, 图 9 展示了数据集中不同场景下的挖掘机工作姿势示例. 通过这个数据集, 可以全面地评估算法在挖掘机工作状态识别方面的性能, 并比较 Yolopose 和 Yolopose-ex 算法之间的差异.

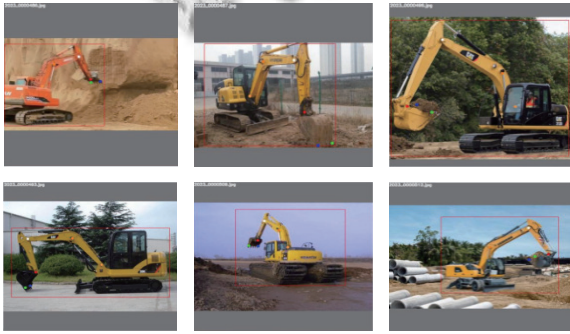


图 9 数据集部分图片

5.1.3 参数设定与训练

训练检测网络模型时, 输入的图像的尺寸大小调整为 640×640, 目标框的损失权重为 0.05, 关键点的损失权重为 0.2, 权重衰减系数设置为 0.000 5, 初始学习率设置为 0.01, 学习率动量设置为 0.937, 批量大小 (batch size) 设置为 16, 置信度为 0.3.

5.1.4 实验结果分析

为了评估改进的效果, 本文采用了多个目标检测评价指标进行验证, 包括平均精度均值 (mAP)、参数量 (Params)、计算量 (Gflops)、速度 (FPS) 和目标关键点相似度 (OKS). 其中 $Precision$ 、 $Recall$ 、 mAP 、 OKS 的计算公式如式 (6)–式 (9) 所示:

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

$$mAP = \frac{\sum_{i=1}^N AP_i}{N} \quad (8)$$

$$OKS = \frac{\sum_{i=1}^N \left[\exp\left(-\frac{d_i^2}{2s^2k_i^2}\right) d(v_i > 0) \right]}{\sum_{i=1}^N d(v_i > 0)} \quad (9)$$

在对比实验中, 使用了 Yolopose 算法、Yolopose-DW 算法、Yolopose-CBAM 算法和 Yolopose-ex 算法, 并采用了预训练权重为 YOLOv7-w6-pose.pt. 训练的结果如表 1 所示.

表 1 性能指标对比表

模型	$Precision$	$Recall$	$mAP@0.5$	参数量(万)	FPS
Yolopose	0.923	0.952	0.894	799	60
Yolopose-DW	0.902	0.933	0.906	512	74
Yolopose-CBAM	0.964	0.988	0.964	812	63
Yolopose-ex	0.955	0.986	0.953	574	71

根据表 1 的结果, Yolopose-ex 网络在几个性能指标上相较于 Yolopose 均有所提升, 具体而言, Yolopose 的准确率提升了 3.2%, $mAP@0.5$ 提高了 5.9%, FPS 提高了 11%, 这表明本文的改进方法不仅可以提高模型的检测精度, 还能显著加快模型的推理速度. Yolopose-DW 模型的使用说明了深度可分离卷积能够明显地提高模型的检测速度, 降低计算量和参数量, 但也会略微降低模型的准确率. 而 Yolopose-CBAM 的加入说明了 CBAM 注意力机制可以增强对目标的关注能力, 但会增加网络深度, 导致参数量增加, 从而使得模型的推理速度下降了 3%. 原始的 Yolopose 网络参数量为 79984616 个, 而经过深度可分离卷积和 CBAM 注意力机制模块的改进后, Yolopose-ex 的参数量减少为 57384311 个. 这种改进使得 Yolopose-ex 在推理速度、参数量和精度方面都有所提升, 从而降低了部署该模型所需的硬件成本, 同时仍能有效地检测挖掘机及其关键点.

图 10 展示了 Yolopose 和 Yolopose-ex 在训练结束后的模型测试曲线图. 其中蓝色表示 Yolopose, 黄色表示 Yolopose-ex. 横坐标表示训练轮数 (epochs), 纵坐标分别表示 $mAP@0.5$ 、 $mAP@0.5:0.95$ 、 $Precision$ 和 $Recall$. 从图 10 这 4 幅图中可以观察到, 当迭代次数接近 60 次时, 各个指标都趋于稳定, 并且接近于表 1 中的性能指标值. 图 11 为 Yolopose-ex 在一些图片的检测效果图, 挖掘机姿势可以得到精确的估计, 且检测的置信度均在 0.8 以上, 检测效果很好.

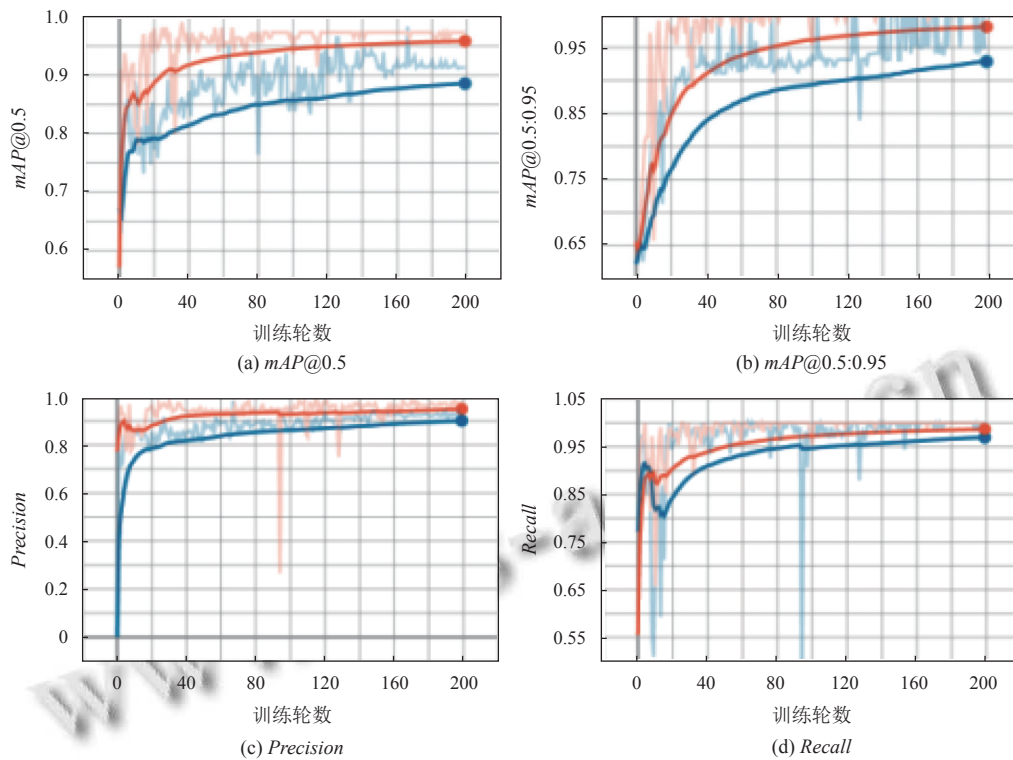


图 10 Yolopose 与 Yolopose-ex 在测试集的性能对比图

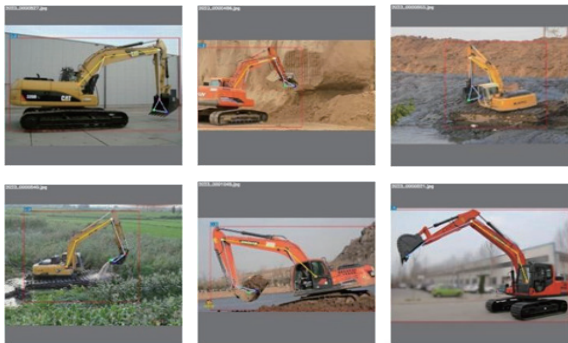


图 11 Yolopose-ex 检测效果图

5.2 挖掘机工作状态自动判别实验

5.2.1 数据集处理

挖掘机工作状态定义为正在执行沟渠挖掘、地基挖掘、深坑挖掘、土方挖掘、废旧拆迁等任务,而非工作状态则被定义为挖掘机静止或直线行走状态。

为了构建挖掘机工作状态和非工作状态的数据集,选取了 3000 个工作状态和非工作状态片段。每个片段包含 20 帧图像,并从中选择了第 1、5、10 和 15 帧图像(图 12, 图 13)。通过使用 Yolopose-ex 算法,得到了这 4 帧图像中挖掘机的 6 点姿势图数据。然后,计算这

4 帧之间的平均姿态变化信息,形成一个包含 3000 个挖掘机工作状态特征向量 (MSV) 的数据集。



图 12 挖掘机处于工作状态的关键帧



图 13 挖掘机非工作状态关键帧

5.2.2 实验结果分析

从数据集中随机选取 2400 个作为训练数据,剩下的 600 作为测试数据,训练多层感知机判别挖掘机工作状态。对不同的挖掘机工作状态特征向量进行了测试对比,结果如表 2 所示。

根据表 2 结果,使用 $\overline{\Delta WH}$ 作为工作状态的特征向量和多层感知机作为分类器时,判别准确率为 59.3%;使用 $\overline{\Delta Angle}$ 工作状态特征向量和多层感知机作为分类

器时, 判别准确率为 88.4%; 使用 \overline{MSV} 作为工作状态的特征向量时, 两种机器学习逻辑回归模型 (LR) 和支持向量机 (SVM) 的判别准确率分别为 87.6% 和 89.1%; 深度学习算法多层感知机工作状态的判别准确率达到 96.6%。通过对比, 说明选用挖掘机工作状态特征向量 \overline{MSV} 更能表达挖掘机工作状态, 准确率更高。

为了进一步分析基于挖掘机工作状态向量 \overline{MSV} 判断挖掘机工作状态的有效性, 对 \overline{MSV} 进行了细化分析。从表 3 可知, 当关键帧之间的挖掘机目标高宽比和挖掘机姿势变化较大时, 挖掘机被判别为工作状态; 当关键帧之间的挖掘机目标高宽比和挖掘机姿势变化较小时, 挖掘机被判别为非工作状态。从测试结果来看, 以 \overline{MSV} 为挖掘机工作状态的特征向量判别挖掘机运动状态具有较高准确性。

表 2 基于 MLP 的挖掘机工作状态识别结果

实验	算法	准确率 (%)
1	$\overline{\Delta WH}$ +MLP	59.3
2	$\overline{\Delta Angle}$ +MLP	88.4
3	\overline{MSV} +LR	87.6
4	\overline{MSV} +SVM	89.1
5	\overline{MSV} +MLP	96.6

表 3 挖掘机工作状态特征向量分析表

实验	$\overline{\Delta WH}$	$\overline{\Delta \theta_1}$	$\overline{\Delta \theta_2}$	$\overline{\Delta \theta_3}$	$\overline{\Delta \theta_4}$	实验状态	判别状态
1	0.241	33°	1°	23°	41°	工作状态	工作状态
2	0.112	52°	2°	15°	36°	工作状态	工作状态
3	0	0	1°	3°	1°	非工作状态	非工作状态
4	0	1°	0	5°	0	非工作状态	非工作状态

6 结论

本文提出了一种改进的挖掘机检测与姿势估计方法, 名为 Yolopose-ex。相较于传统的图像处理方法, 该模型能够有效地克服复杂背景对检测目标的干扰, 并具备较强的环境适应能力、鲁棒性和推理速度。Yolopose-ex 方法利用挖掘机工作所引起的挖掘机关键点在图像中的位置变化以及挖掘机目标在前后帧的长宽比构成了挖掘机特征向量 \overline{MSV} 。该特征向量被用作多层感知机 MLP 分类器的输入, 用于判断挖掘机的运动状态。实验结果显示, 该方法达到了 96.6% 的准确率。该算法的参数数量较少, 易于在监控中心进行部署, 为防护地下基础设施免受挖掘机的野蛮入侵提供了自动检测与运动状态判别的解决方案。

参考文献

- 李源, 何荣开, 王庆, 等. 基于颜色及投影特征的挖掘机图像分割算法. 小型微型计算机系统, 2013, 34(11): 2635-2638.
- 林焕凯. 复杂场景下挖掘机运动状态分析与识别 [硕士学位论文]. 广州: 华南农业大学, 2016.
- 毛亮, 薛月菊, 林焕凯, 等. 一种基于视频图像的挖掘机工作状态识别方法. 系统工程理论与实践, 2019, 39(3): 797-804. [doi: 10.12011/1000-6788-2017-1525-08]
- Liu Z, Lin YT, Cao Y, *et al.* Swin Transformer: Hierarchical vision Transformer using shifted windows. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE, 2021. 9992-10002. [doi: 10.1109/iccv48922.2021.00986]
- Mehta S, Rastegari M. MobileViT: Light-weight, general-purpose, and mobile-friendly vision transformer. Proceedings of the 10th International Conference on Learning Representations. ICLR, 2021.
- Redmon J, Divvala S, Girshick R, *et al.* You only look once: Unified, real-time object detection. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 779-788.
- Redmon J, Farhadi A. YOLOv3: An incremental improvement. arXiv:1804.02767, 2018.
- Wang CY, Bochkovskiy A, Liao HYM. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023. 7464-7475.
- Maji D, Nagori S, Mathew M, *et al.* YOLO-pose: Enhancing YOLO for multi person pose estimation using object keypoint similarity loss. Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022. 2636-2645.
- Fan XC, Zheng K, Lin YW, *et al.* Combining local appearance and holistic view: Dual-source deep neural networks for human pose estimation. Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston: IEEE, 2015. 1347-1355. [doi: 10.1109/cvpr.2015.7298740]
- Osokin D. Real-time 2D multi-person pose estimation on CPU: Lightweight OpenPose. Proceedings of the 8th International Conference on Pattern Recognition Applications and Methods. Prague: ICPRAM, 2019.
- Woo S, Park J, Lee J Y, *et al.* CBAM: Convolutional block attention module. Proceedings of the 15th European

- Conference on Computer Vision (ECCV). Munich: Springer, 2018. 3–19.
- 13 Hu J, Shen L, Sun G. Squeeze-and-excitation networks. Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 7132–7141.
- 14 Howard AG, Zhu ML, Chen B, *et al.* MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv:1704.04861, 2017.
- 15 Sandler M, Howard A, Zhu ML, *et al.* MobileNetV2: Inverted residuals and linear bottlenecks. Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 4510–4520.
- 16 Howard A, Sandler M, Chen B, *et al.* Searching for MobileNetV3. Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul: IEEE, 2019. 1314–1324. [doi: [10.1109/iccv.2019.00140](https://doi.org/10.1109/iccv.2019.00140)]
- 17 Liu S, Qi L, Qin HF, *et al.* Path aggregation network for instance segmentation. arXiv:1803.01534, 2018.
- 18 Ma NN, Zhang XY, Zheng HT, *et al.* ShuffleNet V2: Practical guidelines for efficient CNN architecture design. Proceedings of the 15th European Conference on Computer Vision. Munich: Springer, 2018. 122–138. [doi: [10.1007/978-3-030-01264-9_8](https://doi.org/10.1007/978-3-030-01264-9_8)]
- 19 Sun K, Xiao B, Liu D, *et al.* Deep high-resolution representation learning for human pose estimation. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 5686–5696.

(校对责编: 孙君艳)