

# 融合场景机制的机械作业操作识别算法

宋浪

上海理工大学机械工程学院, 上海

收稿日期: 2024年4月19日; 录用日期: 2024年5月14日; 发布日期: 2024年5月21日

## 摘要

机械作业场景下的操作识别可以大大提高安全监督效率, 针对传统的目标检测算法无法根据不同的机械作业场景识别操作行为的问题, 提出一种融合场景机制的机械作业操作识别算法, 融合两个在COCO数据集上预训练好的YOLOv5s模型, 先进行机械作业场景识别, 再进行关键目标检测, 同时利用交并比构建逻辑函数对作业人员的操作行为进行判定。以角磨机作业场景为例, 经实验验证, 本文的算法模型识别精确率为97.9%, 平均识别时间为0.114 s, 满足了精确性与实时性要求。

## 关键词

深度学习, 目标检测, 机械作业, YOLOv5s, 交并比

# Mechanical Operation Recognition Algorithm Incorporating Scene Mechanism

Lang Song

School of Mechanical Engineering, University of Shanghai for Science and Technology, Shanghai

Received: Apr. 19<sup>th</sup>, 2024; accepted: May. 14<sup>th</sup>, 2024; published: May. 21<sup>st</sup>, 2024

## Abstract

Operation recognition in mechanical operation scenarios can greatly improve the efficiency of safety supervision. Aiming at the problem that traditional target detection algorithms cannot identify operation behaviors according to different mechanical operation scenarios, a mechanical operation recognition algorithm integrating scene mechanism is proposed, which integrates two YOLOv5s models pre-trained on COCO data sets. First, the mechanical operation scene is identified, and then the key target is detected. At the same time, the operation behavior of the operator is judged by the logic function constructed by the intersection ratio. Taking the operation scene of Angle mill as an example, the experimental verification shows that the algorithm model recogni-

tion accuracy rate is 97.9%, and the average recognition time is 0.114s, which meets the accuracy and real-time requirements.

## Keywords

Deep Learning, Object Detection, Mechanical Operation, YOLOv5s, Intersection over Union

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

机械作业存在环境复杂、设备种类多以及作业场景多等情况，工人使用机械加工设备时面临高速、强光、高温等危险因素[1]。普通的人力监管效率低下，结合人力监督和高清视频监控[2]也始终存在疲劳与范围有限等因素，而结合高清视频监控与目标检测技术[3]对工人作业视频图像进行违规作业检测，可以做到高效监督，及时警告，大大提高巡查监督效率，同时节省了人力资源。

目标检测分为传统目标检测和基于深度学习的目标检测[4]-[9]。传统目标检测方法的检测流程主要分为三个步骤。首先是选取感兴趣的区域，通常是使用滑动窗口的方法来选取可能包含物体的区域；其次是使用人工设计的特征进行特征提取；然后使用预训练好的分类器对提取的特征进行检测分类。Yue 等[10]采用传统目标检测方法，利用定向梯度直方图(HOG)提取特征，然后采用随机二元测试方法构建分类器，并使用改进的 Real AdaBoost 算法构建 IBRF 进行安全帽检测。虽然该文章的算法基本满足检测要求，但检测目标过于单一，只对安全帽进行检测；且传统目标检测方法十分依赖人工设计的特征，这在实际应用中检测精度较低，只适用于特征明显，背景简单的场景；而且其运行速度较慢，可能产生多个正确结果，导致结果冗余和不一致。

除了传统的目标检测之外，基于深度学习的目标检测方法因其较高的精度和良好的实时性备受工业界和学术界的推崇。目前基于深度学习的目标检测主要分为两大类，一类是基于区域的两阶段方法，其主要代表为 R-CNN 系列算法，包括 Faster-RCNN 和 Mask-RCNN 等[11] [12]。另一类是基于端到端的单阶段方法，YOLO、SSD [13] [14]为此类方法的主要代表。Chen 等[15]针对 CNN 算法进行剪枝，进一步提高算法运行速度，并引入 AlexNet 和 Inception 对网络进行拓宽和深化，最后使用批量归一化方法处理输入批量样本实现对电力工人的穿戴检测。Ji 等[16]在 YOLOv4 的基础上，采用基于对象边界框构造位置特征的方法获得表征对象之间相对偏移的特征向量，并通过随机森林算法训练由位置特征组成的数据集，然后进行参数优化，得到分类器。最后通过推理机制分析分类器输出的信息来实现防护装备穿戴检测。Chen 等[17]在 YOLOv5 框架的基础上，采用自适应注意力嵌入模型(ASAE)来改进骨干网络，同时将原有的特征提取网络 PA-Net 替换为加权特征金字塔模块(WFPN)，用以检测工业场所中，工人不戴安全帽和吸烟的不安全行为。当前研究检测的穿戴防护用品类型数量较少，也无法判断工人的防护装备是否穿戴在正确的身体部位，也没有考虑到工人的设备操作行为，且算法模型不具备场景识别能力，无法根据不同的作业场景进行相应规范下的检测。

真实环境下，机械作业场景复杂多样，且不同的机械作业必须恪守相应的规范。因此，要求检测算法具备作业场景辨别能力；上述算法虽能进行目标检测，却无法根据不同的机械作业场景，判别工人操作是否违规。本文针对上述算法存在的问题提出了融合场景机制的机械作业操作识别算法。本文主要工

作如下：

1. 针对真实机械作业高质量图片数据不便获取问题，使用 COCO 数据集上预训练好的 YOLOv5s 模型进行训练，使用预训练权重代替随机初始化权重训练可以大大提高模型训练效率。
2. 采用模型融合方法，将两个预训练好的 YOLOv5s 模型进行融合；一个用于机械作业场景识别，另一个用于检测关键目标物。
3. 利用交并比(intersection over union, IoU)构建逻辑函数并对工人作业行为进行判定。
4. 以角磨机作业为例，构建数据集并进行实验验证，对实验结果进行分析。

## 2. 基于 YOLOv5 的目标检测与场景识别

YOLOv5 网络因其良好的检测精度和检测速度得到了广泛应用。YOLOv5 根据网络深度分为 YOLOv5s、YOLOv5n、YOLOv5m、YOLOv5l 和 YOLOv5x。YOLOv5n 满足实时性能要求，但识别精度无法满足要求。YOLOv5s 结合了 YOLOv5x 和 YOLOv5n 的优点。不仅满足实时性要求，而且具有良好的识别精度，因此，该文选择 YOLOv5s 进行目标检测与场景识别。

### 2.1. YOLOv5s 网络结构

该文采用的 YOLOv5s 6.0 版本网络模型由 Input、Backbone、Neck 和 Head 四部分组成，其网络结构如图 1 所示。Input 端包含了 Mosaic 数据增强、自适应锚框计算和自适应图像缩放，通过随机缩放、裁剪、排列 4 张图像，Mosaic 数据增强大大提高了数据集质量并增强了网络。BackBone 主干网络采用了 CSP-Darknet53 网络[18]，由 CBS 和 SPPF 等模块组成。CBS 模块由 Conv、BN 和 SiLU 组成，通过卷积操作提取局部空间信息，并通过 BN 层规范化特征值分布，最后通过激活函数 SiLU 引入非线性变换能力，从而实现对输入特征的转换和提取。SPPF 全称为快速空间金字塔池化层，其使用 3 个 kernel = 5 的卷积

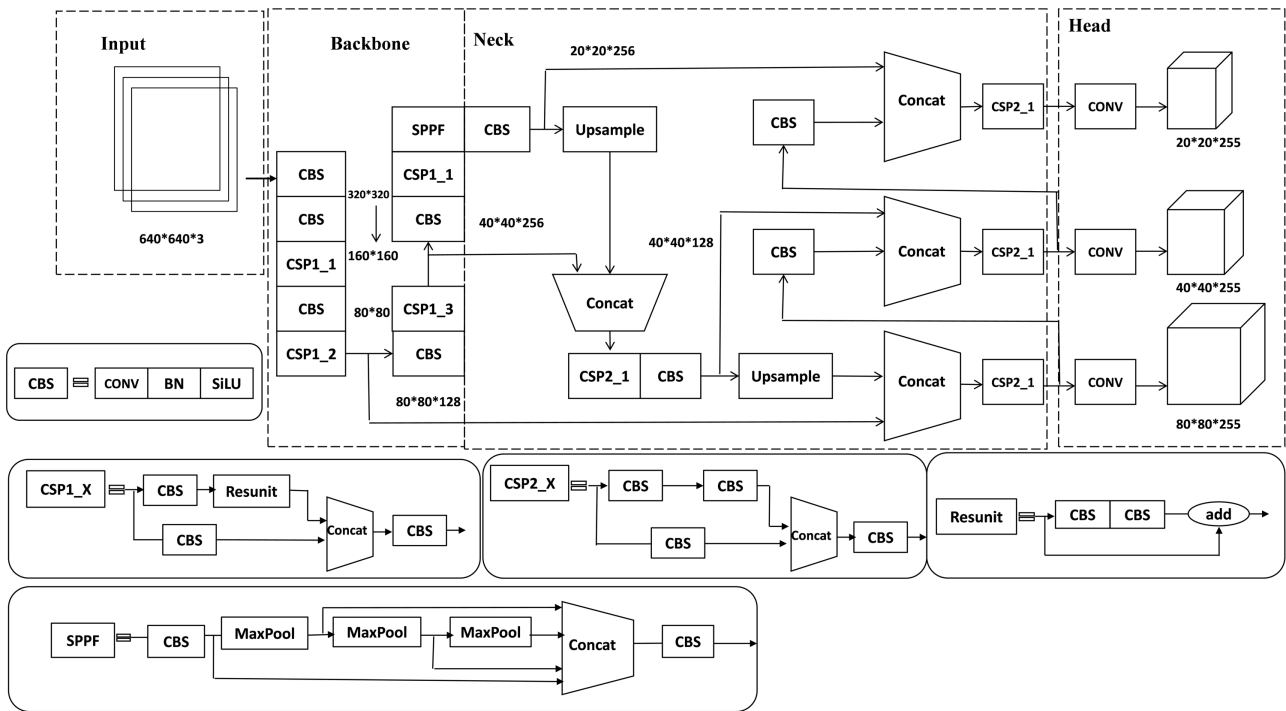


Figure 1. YOLOv5s network structure diagram

图 1. YOLOv5s 网络结构图

块进行最大池化，代替原来的  $\text{kernel} = (5, 9, 13)$  卷积块。SPPF 模块融合了丰富特征图信息，提高了网络运行速度。Neck 使用路径聚合网络(PANet)，可以将不同尺度特征图进行融合，在 PANet 中，底层特征使用一种新的特征金字塔网络 FPN 结构进行传播，PANet 可以利用精确的定位信号提高下层物体的定位精度。最后，Head 生成三个尺度的特征图，并使用三个  $\text{kernel} = 1$  的卷积块进行预测。

## 2.2. 机械作业场景识别

机械作业场景下的违规操作识别需先进行场景识别[19]，之后再根据该场景的相关作业规范，针对性地进行目标检测。常用的场景识别方法主要包括基于对象、基于上下文和基于区域。基于对象的场景识别是根据图像中的目标对象，如人、车、建筑等，来识别场景的类别，这种方法需要先进行目标检测，然后根据目标的特征或者与其他目标的关系，来判断场景的类别。上下文的场景识别主要考虑图像中的整体信息，如颜色、纹理、光照等，来识别场景的类别。这种方法不需要进行目标检测，而是利用全局特征或者局部特征，来判断场景的类别。区域的场景识别是根据图像中的不同区域，如天空、水面、草地等，来识别场景的类别。这种方法需要先进行图像分割，然后根据每个区域的特征或者与其他区域的关系，来判断场景的类别。该文选择基于对象的方法，以角磨机作业场景为例，以工人手握角磨机作为场景特征，同样利用 YOLOv5s 网络提取对象特征，识别角磨机场景。

## 3. 融合场景机制的机械作业操作识别算法

本文提出的融合场景机制的机械作业操作识别算法，融合两个在 COCO 数据集上预训练好的 YOLOv5s 模型，分别进行场景识别与目标检测，然后再根据检测到的目标，使用构建的逻辑函数进行操作判断。COCO 数据集是一个用于目标检测、分割的大型数据集，以场景理解为目标，并从复杂的日常生活中截取图像数据；其中包含了 80 个目标类别和 91 个 stuff 类别，共 328,000 影像数据和 2,500,000 个标签。

针对机械作业场景种类繁多，需要在识别作业场景的基础上再进行关键目标的检测识别，然后针对不同机械作业的作业规范，对工人的操作进行合法化判定。机械作业的操作识别主要分为穿戴物品识别和设备操作行为识别。本文接下来以角磨机作业为例进行研究。角磨机作业需要工人正确佩戴安全帽、护目镜、手套以及口罩，除此之外，使用角磨机作业时，工人还应正确握持角磨机握把，不能单手操作或是随意握持。

本文首先根据融合模型识别到的场景进行相应的关键目标检测，然后把检测到的目标对应的参数生成一个支持调用的参数列表，该列表中包含了目标的类别信息、识别框中心坐标和边框尺寸。当识别到角磨机场景时，调用图像对应的检测目标参数列表，使安全帽、护目镜、口罩、手套以及场景识别框同时存在，并抑制工人头部和手部的识别框的显示。此外，使用交并比构造判断函数对各个识别框的坐标信息进行位置判断。交并比在目标检测中的一个重要要素，表示预测框与真实框的重叠率，如式(1)。比值为 1 时即为完全重叠。交并比可以用来评价预测结果的准确性和确定正负样本。

式(1)中  $R(x_1)$  表示目标  $x_1$  的识别框，以下同理； $x_{he}$ 、 $x_{ha}$ 、 $x_h$ 、 $x_{go}$ 、 $x_m$ 、 $x_{gl}$ 、 $x_{gr}$  分别代表工人头部、工人手部、安全帽、护目镜、口罩、手套、握把。

$$IoU = \frac{R(x_1) \cap R(x_2)}{R(x_1) \cup R(x_2)} \quad (1)$$

式(2)表示工人头部与安全帽、护目镜、口罩交并比的乘积。若  $I_2 = 0$ ，则表示工人作业时没有正确佩戴安全帽、护目镜或者口罩。

$$I_2 = \frac{R(x_{he}) \cap R(x_h) * R(x_{he}) \cap R(x_g) * R(x_{he}) \cap R(x_m)}{R(x_{he}) \cup R(x_h) R(x_{he}) \cup R(x_g) R(x_{he}) \cup R(x_m)} \quad (2)$$

式(3)表示工人手部与手套的重叠程度, 若  $I_3 = 0$ , 则说明工人没有正确佩戴手套。

$$I_3 = \frac{R(x_{he}) \cap R(x_{gl})}{R(x_{he}) \cup R(x_{gl})} \quad (3)$$

式(4)为工人手部、手套以及角磨机握把的重叠程度, 若  $I_4 = 0$ , 则表示工人作业时未正确握持握把。

$$I_4 = \frac{R(x_{ha}) \cap R(x_{gr}) * R(x_{gl}) \cap R(x_{gr})}{R(x_{ha}) \cup R(x_{gr}) R(x_{gl}) \cup R(x_{gr})} \quad (4)$$

该文将式(2)~(4)的阈值统一设定为 0.5, 当阈值达到 0.5 时, 即判定为正常作业, 反之则判定为违规作业。

## 4. 实验验证

### 4.1. 数据集与实验配置

本文样本数据集为 1445 张现场拍摄的角磨机作业图片, 其中正常作业图片数量为 300 张, 其余各类违规操作图片数量为 1155 张。使用 labelImg 对图片数据进行标注, 工人头部标注为 Head、工人手部为 hand、安全帽为 helmet、护目镜为 goggles、口罩为 mask、手套为 gloves、握把为 grip。然后将数据集按照 7:3 的比例划分训练集和测试集。实验所用配置如下: 操作系统为 Windows10, CPU 为 Intel(R) Core(TM) i5-8300H CPU@2.30 GHz 2.30 GHz, GPU 为 GTX1050 Ti, 开发框架为 Pytorch1.8.1, CUDA10.2.89, CuDNN7.6.5, 编程语言为 Python3.8.5。

### 4.2. 评价指标

本文以精确率(Precision, P)、召回率(Recall, R)、平均精度(Mean Average Precision, mAP)等指标衡量算法模型预测的准确程度, 上述指标的值越高, 则代表模型的预测效果越好。精确率定义为:

$$P = \frac{TP}{TP + FP} \quad (5)$$

召回率定义为

$$R = \frac{TP}{TP + FN} \quad (6)$$

平均精度定义为

$$AP = \int_0^1 P(R) dR \quad (7)$$

$$mAP = \frac{\sum_{n=1}^N AP(n)}{N} \quad (8)$$

式(6)、(7)中的  $TP$  为正样本被正确预测数量,  $FP$  为负样本被误测为正样本数量,  $FN$  为负样本被误测为正样本数量。

YOLOv5s 损失函数包括边框损失( $loss_{box}$ )、分类损失( $loss_{cls}$ )与置信度损失( $loss_{obj}$ ), 总损失  $Loss$  表达式为:

$$Loss = \alpha * loss_{box} + \beta * loss_{cls} + \gamma * loss_{obj} \quad (9)$$

其中  $\alpha$ 、 $\beta$ 、 $\gamma$  为损失权重。本文使用 CIOU Loss 作为边框损失，分类损失与置信度损失选择二元交叉熵损失函数，并使用非极大值抑制 NMS 进行预测框筛选。

### 4.3. 实验结果分析

将融合模型的迭代次数 Epoch 设为 300，如果模型中途达到最佳训练效果，模型将会自动终止训练，并自动保存最后一次训练参数与最佳训练参数。该文融合模型在训练次数达到 269 时达到了最佳效果并终止训练。图 2 为本文模型的精度指标(P、R、mAP)曲线与损失值 Loss 变化曲线。

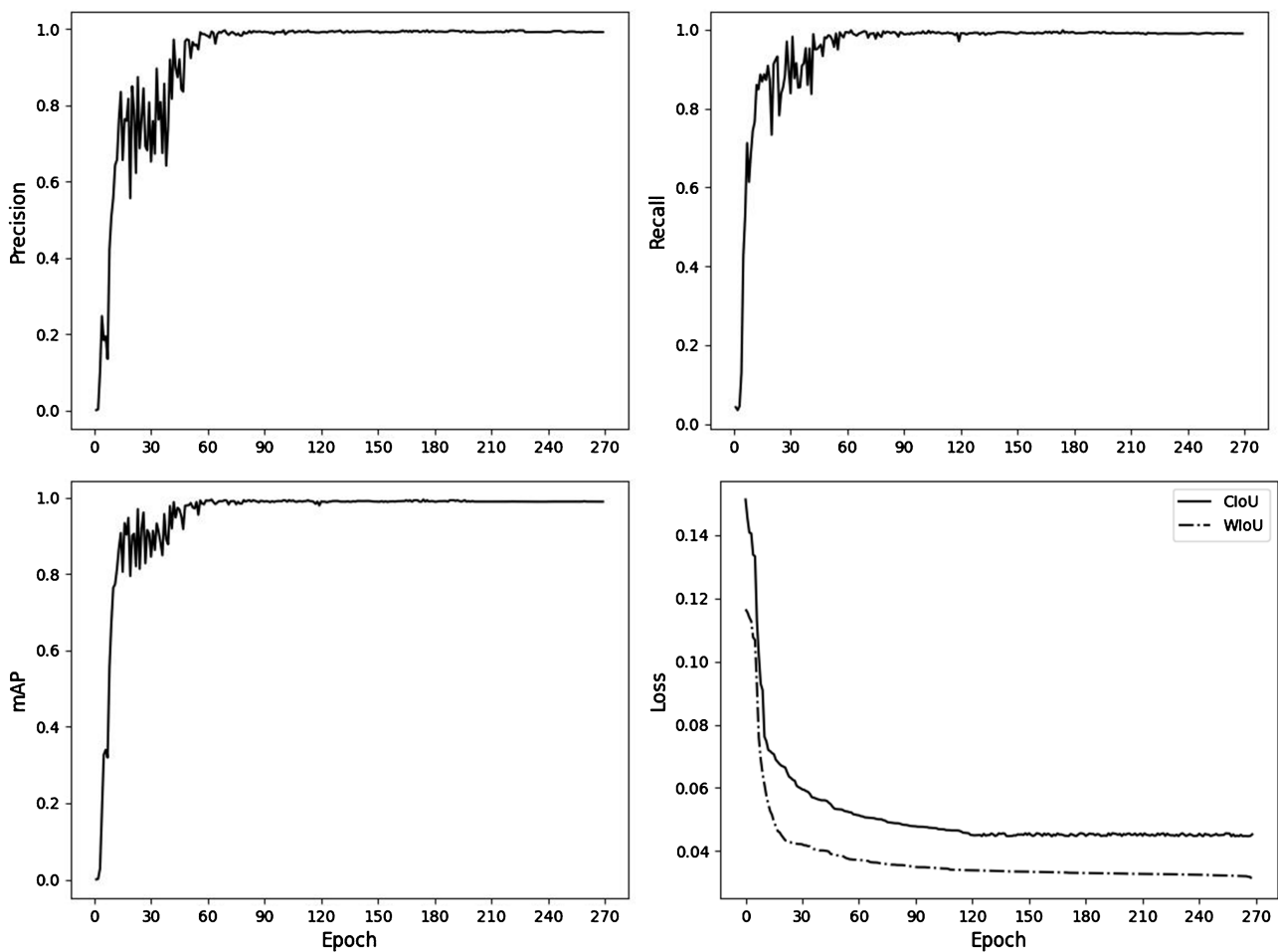


Figure 2. P, R, mAP and Loss variation curve

图 2. P、R、mAP、Loss 变化曲线

角磨机作业识别结果如图 3 所示，当工人头部、手部分别与安全帽、护目镜、口罩以及手套、握把等关键目标的识别框满足设定的 IoU 函数时，即判定为正常作业，并输出作业场景。当识别到非角磨机作业场景时，只需检测通用场景下的目标，例如安全帽。

## 5. 结论

本文针对当前工业场景下的安全检测算法无法判断工人防护设备是否穿戴正确，没有识别作业场景，





**Figure 3.** Inspection results (the picture on the left shows the operation of an angle Grinder, the picture on the right shows the operation of a non-angle grinder)

**图 3.** 检测结果(图左为角磨机作业场景；图右为非角磨机作业场景)

且未考虑工人的设备操作行为等问题，结合预训练模型与模型融合策略，同时利用交并比构建逻辑判断函数，提出了一种融合场景机制的机械作业操作识别算法。本文算法不仅实现了作业场景类型的识别，同时对工人的防护用品穿戴与设备操作行为进行了识别判断。本文以角磨机作业为例，构建样本数据集，经过实验验证，融合算法平均检测精度达到了 97.9%，平均检测时间为 0.114 s，满足实时性与准确性要求，可以有效进行机械作业场景下的操作识别，大大提高安全监督效率。

## 参考文献

- [1] 刘全龙, 彭雨蒙, 赵盼, 等. 基于 HFACS 模型的制造企业机械伤害事故致因分析[J]. 中国安全科学学报, 2023, 33(3): 51-59.
- [2] Bagga, P., Mitra, A., Das, A.K., *et al.* (2022) Secure Biometric-Based Access Control Scheme for Future IoT-Enabled Cloud-Assisted Video Surveillance System. *Computer Communications*, **195**, 27-39. <https://doi.org/10.1016/j.comcom.2022.08.003>
- [3] Zaidi, S.S.A., Ansari, M.S., Aslam, A., *et al.* (2022) A Survey of Modern Deep Learning Based Object Detection Models. *Digital Signal Processing*, **126**, Article ID: 103514. <https://doi.org/10.1016/j.dsp.2022.103514>
- [4] Zheng, A., Zhang, Y., Zhang, X., *et al.* (2022) Progressive End-To-End Object Detection in Crowded Scenes. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, New Orleans, 18-24 June 2022, 847-856. <https://doi.org/10.1109/CVPR52688.2022.00093>
- [5] Yang, J., Liu, S., Li, Z., *et al.* (2022) Real-Time Object Detection for Streaming Perception. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, New Orleans, 18-24 June 2022, 5375-5385. <https://doi.org/10.1109/CVPR52688.2022.00531>
- [6] Sowmyayani, S. and Rani, P.A.J. (2022) Salient Object Based Visual Sentiment Analysis by Combining Deep Features and Handcrafted Features. *Multimedia Tools and Applications*, **81**, 7941-7955. <https://doi.org/10.1007/s11042-022-11982-5>
- [7] Xu, M., Qin, L., Chen, W., *et al.* (2023) Multi-View Adversarial Discriminator: Mine the Non-Causal Factors for Object Detection in Unseen Domains. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Vancouver, 17-24 June 2023, 8103-8112. <https://doi.org/10.1109/CVPR52729.2023.00783>
- [8] Zhang, P., Li, X., Hu, X., *et al.* (2021) Vinvl: Revisiting Visual Representations in Vision-Language Models. *Pro-*

- 
- ceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, 20-25 June 2021, 5579-5588. <https://doi.org/10.1109/CVPR46437.2021.00553>
- [9] Xiao, Y., Tian, Z., Yu, J., *et al.* (2020) A Review of Object Detection Based on Deep Learning. *Multimedia Tools and Applications*, **79**, 23729-23791. <https://doi.org/10.1007/s11042-020-08976-6>
- [10] Yue, S., Zhang, Q., Shao, D., *et al.* (2022) Safety Helmet Wearing Status Detection Based on Improved Boosted Random Ferns. *Multimedia Tools and Applications*, **81**, 16783-16796. <https://doi.org/10.1007/s11042-022-12014-y>
- [11] He, Z., Zhang, L., Gao, X. and Zhang, D. (2023) Multi-Adversarial Faster-RCNN with Paradigm Teacher for Unrestricted Object Detection. *International Journal of Computer Vision*, **131**, 680-700. <https://doi.org/10.1007/s11263-022-01728-z>
- [12] Zhang, Z., Yin, X. and Yan, Z. (2022) Rapid Data Annotation for Sand-Like Granular Instance Segmentation Using Mask-RCNN. *Automation in Construction*, **133**, Article ID: 103994. <https://doi.org/10.1016/j.autcon.2021.103994>
- [13] Yan, J. and Wang, Z. (2022) YOLO V3+VGG16-Based Automatic Operations Monitoring and Analysis in a Manufacturing Workshop under Industry 4.0. *Journal of Manufacturing Systems*, **63**, 134-142. <https://doi.org/10.1016/j.jmsy.2022.02.009>
- [14] Wei, P., Yu, X., Di, Z., *et al.* (2022) Design of Robot Automatic Navigation under Computer Intelligent Algorithm and Machine Vision. *Journal of Industrial Information Integration*, **28**, Article ID: 100366. <https://doi.org/10.1016/j.jii.2022.100366>
- [15] Chen, Z., Ma, C., Ren, J., *et al.* (2023) Research on the Identification Method of Safety Wearing of Electric Power Workers Based on Deep Learning. *Frontiers in Energy Research*, **10**, Article 1091322. <https://doi.org/10.3389/fenrg.2022.1091322>
- [16] Ji, X., Gong, F., Yuan, X., *et al.* (2023) A High-Performance Framework for Personal Protective Equipment Detection on the Offshore Drilling Platform. *Complex & Intelligent Systems*, **9**, 5637-5652. <https://doi.org/10.1007/s40747-023-01028-0>
- [17] Chen, B., Wang, X., Bao, Q., *et al.* (2022) An Unsafe Behavior Detection Method Based on Improved YOLO Framework. *Electronics*, **11**, Article 1912. <https://doi.org/10.3390/electronics11121912>
- [18] 李永上, 马荣贵, 张美月. 改进 YOLOv5s + DeepSORT 的监控视频车流量统计[J]. 计算机工程与应用, 2022, 58(5): 271-279.
- [19] Wang, C., Peng, G. and De Baets, B. (2022) Class-Specific Discriminative Metric Learning for Scene Recognition. *Pattern Recognition: The Journal of the Pattern Recognition Society*, **126**, Article ID: 108589. <https://doi.org/10.1016/j.patcog.2022.108589>