

中文引用格式:胡啸峰,滕腾,胡今鸣,等. 机械加工车间作业人员不安全行为识别模型[J]. 中国安全科学学报,2025,35(8):40-47.

英文引用格式:HU Xiaofeng, TENG Teng, HU Jinming, et al. Identification model of unsafe behaviors among operators in machining workshops[J]. China Safety Science Journal, 2025, 35(8): 40-47.

机械加工车间作业人员不安全行为识别模型^{*}

胡啸峰 副教授,滕腾,胡今鸣,温家骏

(1 中国人民公安大学 信息网络安全学院,北京 100038;

2 安全防范技术与风险评估公安部重点实验室,北京 100038)

中图分类号:X941

文献标志码:A

DOI: 10.16265/j.cnki.issn1003-3033.2025.08.0176

基金项目:国家自然科学基金资助(72174203)。

【摘要】 为提高机械加工车间作业人员的安全管理水平,构建基于 YOLOv11 的识别模型,融合元架构视觉模块(MetaFormer)、混合聚合网络(MANet)以及自适应特征网格卷积注意力机制(AFGC Attention),改进 YOLOv11 模型,并建立真实车间环境下的视频数据集,验证上述模型。结果表明:改进 YOLOv11 模型能够识别无人值守、未佩戴面罩作业、未穿着防护服作业 3 种行为, F_1 指数均达到 0.93 以上;改进的模型在小尺寸目标识别方面的性能具有显著提升,识别戴手套作业行为的 F_1 指数从 0.684 提升至 0.708,平均精度均值(mAP)@0.5 值从 0.604 提升到 0.651。

【关键词】 机械加工车间; 作业人员; 不安全行为; YOLOv11; 注意力机制

Identification model of unsafe behaviors among operators in machining workshops

HU Xiaofeng, TENG Teng, HU Jinming, WEN Jiajun

(1 School of Information and Network Security, People's Public Security University of China,

Beijing 100038, China; 2 Key Laboratory of Security Prevention and Risk Assessment,

Ministry of Public Security, Beijing 100038, China)

Abstract: To enhance the safety management of operators in machining workshops, an identification model based on YOLOv11 was constructed. The YOLOv11 model was improved by integrating the MetaFormer architecture, Mixed Aggregation Network (MANet) module, and Adaptive Feature Grid Convolution Attention (AFGC Attention) mechanism. A video dataset captured in a real workshop environment was established to validate the identification model. The results show that the improved YOLOv11 model can identify three types of behaviors, namely unattended operation, operating without a face shield, and operating without protective clothing, with F_1 scores exceeding 0.93 for all categories. The improved model demonstrates a significant enhancement in identifying small-sized targets, with the F_1 score for identifying glove-wearing behavior increasing from 0.684 to 0.708, and the mAP@0.5 value rising from 0.604 to 0.651. The research findings may provide technical support for the identification and early warning of unsafe behaviors among operators in machining workshops.

Keywords: machining workshops; operators; unsafe behaviors; YOLOv11; attention mechanism

0 引言

机械加工车间作业人员的不安全行为是引发安全事故的主要诱因。准确、及时识别与预警机械加工车间的人员不安全行为,并采取相应的预防和纠正措施,对保障车间的安全稳定运行具有重要意义。

当前,普遍采用视频图像数据与深度学习算法相结合的技术方案,智能识别建筑施工^[1]、矿山井下作业^[2]和公共场所^[3]等场景下的人员不安全行为。在深度学习算法中,单阶段目标检测类算法,如YOLO(You Only Look Once)^[4]系列算法,具有高识别准确率和轻量化特性,在生产安全领域得到广泛应用。如LI Lijuan等^[5]提出一种YOLOv5-SFE算法,实现了工厂工业场景中工人操作行为的实时检测,显著提升了检测与识别效率;FU Yuanyuan等^[6]基于YOLOv7框架设计了GD-YOLO网络,成功实现了化工园区内工人吸烟和使用手机行为的精准检测;XIN Fangfang等^[7]提出一种高精度实时检测方法DSY-YOLOv8n,适用于地下矿工行为的实时检测;罗国富等^[8]通过改进YOLOv5s显著提高了计算效率,实现了智能车间内工人不安全行为的实时检测;韩康等^[9]提出一种结合改进YOLOv7和ByteTrack的行为识别方法,实现了煤矿关键岗位人员的不安全行为识别;徐浩钧等^[10]针对化学实验室动火试验无人值守这一特定危险场景,设计了一种基于改进YOLOv7的行为识别方法,为实验室安全管理提供了有力保障。然而,目前针对机械加工车间,尤其是操作机床区域的作业人员不安全行为识别的研究还较为鲜见。

机械加工车间的目标识别环境通常较为复杂,且存在诸多需要识别的小尺寸目标,如手套等。在工程实践中,虽然理论上通过拉近摄像机的焦距可提高小尺寸目标的识别精度,但并非所有摄像机都具备智能轮巡、动态视角调整等功能。焦距的拉近可能会限制视野范围,导致周围重要区域被遗漏,进而影响全面监控的效果。此外,在机械加工车间这类复杂环境中,目标识别算法需要具备更强的抗干扰能力,特别是对低对比度和遮挡物等干扰因素的适应性,这也是解决小尺寸目标识别问题的关键所在。

鉴于此,笔者拟构建基于YOLOv11的机械加工车间作业人员不安全行为识别模型,并针对机械加工车间环境复杂、不安全行为的目标特征不明显等问题,改进YOLOv11模型,并建立机械加工车间的实际视频数据集验证该模型,以期对机械加工车间

的作业人员安全管理提供技术支撑。

1 构建机械加工车间不安全行为识别模型

1.1 基于YOLOv11的不安全行为识别模型

YOLO是一种实时目标检测算法,其中,YOLOv11是YOLO系列中的最新模型,能够实现更快的处理速度和更高的识别精度。YOLOv11模型架构主要由骨干网络、颈部网络和检测头组成,如图1所示。在骨干网络和颈部网络部分,YOLOv11模型引入了C3K2模块,与其前代模型YOLOv8中的C2F模块相比,C3K2模块是一个可配置模块,其中,包含一个名为c3k的参数,用于提升YOLOv11的处理效率。在骨干网络部分,YOLOv11通过引入结合空间注意力的跨阶段部分连接(Cross Stage Partial with Spatial Attention, C2PSA)模块,有效增强了模型的特征提取能力。检测头部分,YOLOv11采用2个深度卷积(Depthwise Convolution, DWConv)层,可有效减少计算复杂性和参数数量,使模型更为轻量化。

1.2 YOLOv11模型的改进

鉴于机械加工车间视觉环境的复杂性,以及不安全行为识别任务中小尺寸目标的存在,文中融合基于Transformer的元架构视觉模块(MetaFormer)、混合聚合网络(Mixed Aggregation Network, MANet)以及自适应特征网格卷积注意力机制(Adaptive Feature Grid Convolution Attention, AFGC Attention),改进YOLOv11x模型,旨在提高模型对小尺寸目标的识别精度,并增强其在复杂视觉环境下的抗干扰能力。具体改进方法是:①将主干网络部分的C3K2模块与MetaFormer架构下的卷积注意力Transformer模块(Convolutional Attention Former, CAFormer)和卷积Transformer模块(Convolutional Former, ConvFormer)融合;②将颈部网络中的C3K2模块替换为MANet模块;③在颈部网络中加入AFGC Attention模块。改进后的模型命名为YOLOv11x_MMA,其架构如图2所示。

1.2.1 骨干网络中C3K2模块的改进

在机械加工车间场景下,手套、防护面罩等小尺寸目标在图像中占据的像素区域较少,容易导致其特征信息在卷积操作中被忽略或丢失。在特征提取环节,YOLOv11模型的C3K2模块尽管提高了计算效率,但往往难以实现对小尺寸目标特征准确提取。因此,将主干网络中的C3K2模块与ConvFormer模

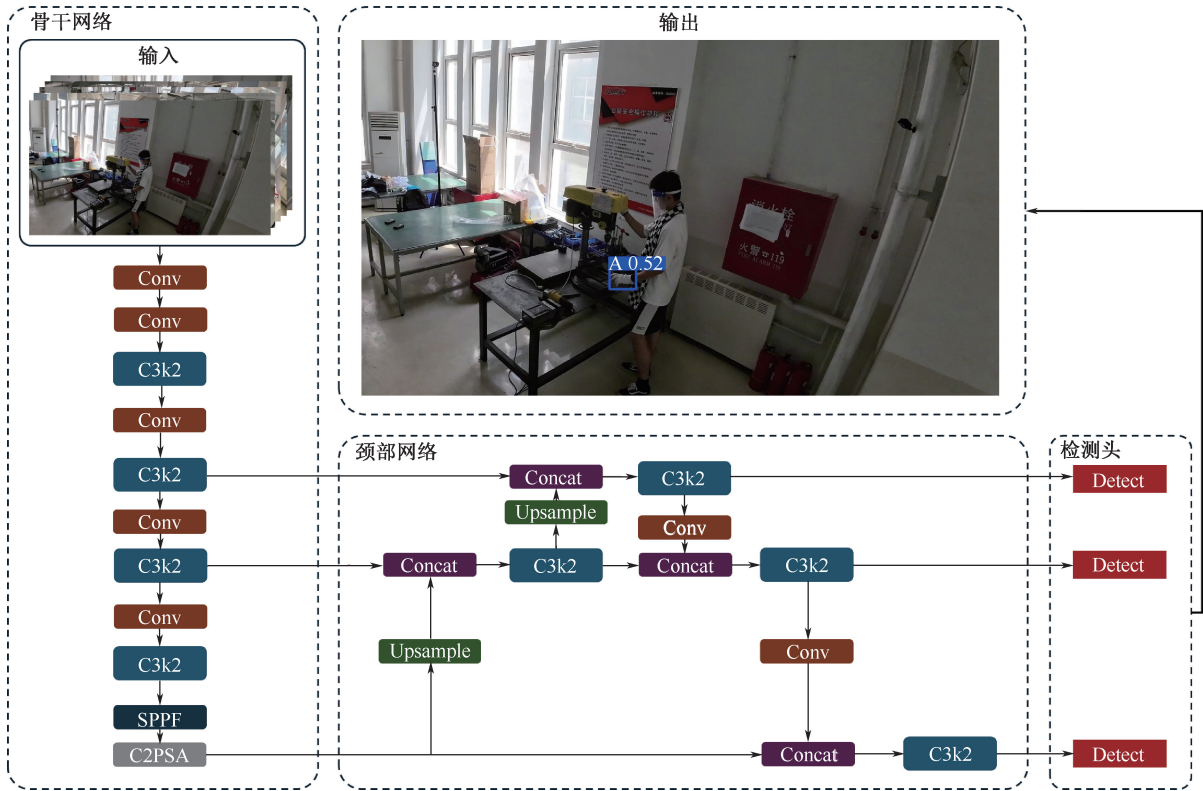


图1 YOLOv11 模型架构
Fig. 1 Architecture of YOLOv11 model

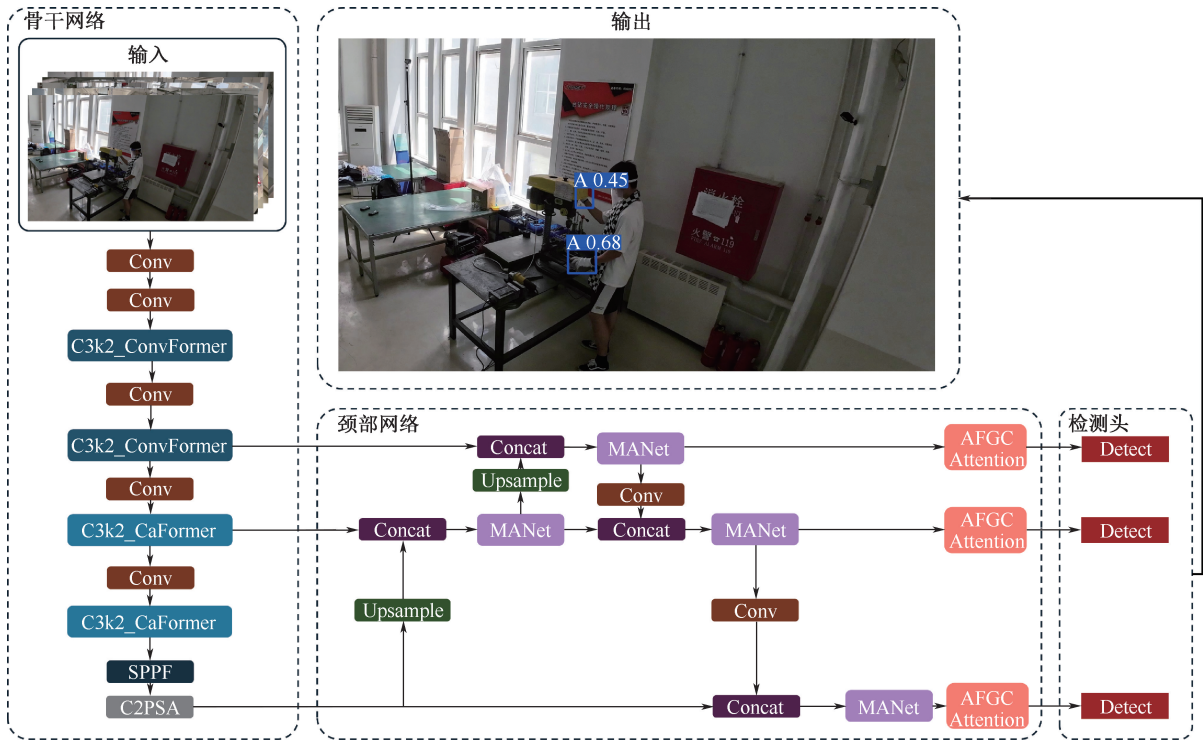


图2 YOLOv11x_MMA 模型架构
Fig. 2 Architecture of YOLOv11x_MMA model

块和 CAFormer 模块融合, 得到改进后的 C3K2_ConvFormer 模块和 C3K2_CaFormer 模块。通过强

化特征提取能力优化模型的基础特征表示, 提升模型性能。ConvFormer 模块和 CAFormer 模块是

MetaFormer 框架下的一类模块^[11]。ConvFormer 通过卷积操作强化局部特征提取,并结合变换器的全局特征识别能力,尤其在视觉任务中表现突出。CAFormer 与 ConvFormer 略有不同,CAFormer 在结构上增加了卷积与自注意力交替和协同工作的设计,优化了局部与全局特征的平衡,特别在处理长序列或复杂数据时表现出色。2 模块的卷积结构设计用 Convolutions($\mathbf{X}$)表示,定义如下:

$$\text{Convolutions}(\mathbf{X}) = \text{Conv}_{\text{pw}2}(\mathbf{Y}) \quad (1)$$

$$\mathbf{Y} = \text{Conv}_{\text{dw}}\{\text{StarReLU}[\text{Conv}_{\text{pw}1}(\mathbf{X})]\} \quad (2)$$

$$\text{StarReLU}(x) = \frac{[\text{ReLU}(x)]^2 - \mathbb{E}\{[\text{ReLU}(x)]^2\}}{\sqrt{\text{Var}\{[\text{ReLU}(x)]^2\}}} \quad (3)$$

式中: $\mathbf{X}$ 为输入的特征图; $\text{Conv}_{\text{pw}1}(\mathbf{X})$ 为对输入特征图做逐点卷积,用于调整通道的数量; $\text{StarReLU}(x)$ 为非线性激活函数; x 为非线性激活函数输入的神经元; $\text{StarReLU}[\text{Conv}_{\text{pw}1}(\mathbf{X})]$ 为对输入做非线性激活,通过引入非线性机制,使模型能够学习更为复杂的特征; $\text{Conv}_{\text{dw}}\{\text{StarReLU}[\text{Conv}_{\text{pw}1}(\mathbf{X})]\}$ 为对输入做深度卷积,用于提取局部空间信息; $\mathbf{Y}$ 为经深度卷积之后的中间特征图; $\text{Conv}_{\text{pw}2}(\mathbf{Y})$ 为对 $\mathbf{Y}$ 做第2个逐点卷积; $\mathbb{E}(\cdot)$ 代表求期望函数。

1.2.2 颈部网络中 C3K2 模块的改进

对于机械加工车间的小尺寸目标识别任务,颈部网络中 C3K2 模块的特征融合能力可能不足,易导致小目标的特征信息丢失。MANet 模块的混合聚合机制有助于保留小目标的特征信息,因此,将颈部网络中的 C3K2 模块替换为 MANet 模块,以期进一步提高小尺寸目标的识别精准度。MANet 是一种高效且灵活的神经网络架构,通过多路径设计、深度可分离卷积、模块化特征提取和混合特征聚合来增强特征表达能力^[12]。MANet 的结构表示如下:

$$\mathbf{Z}_m = \text{Conv}_1(\mathbf{Z}) \quad (4)$$

$$\mathbf{Z}_l = \begin{cases} \text{Conv}_2(\mathbf{Z}_m), l = 1 \\ \text{DSConv}[\text{Conv}_3(\mathbf{Z}_m)], l = 2 \\ \text{Split}(\mathbf{Z}_m), l = 3, 4 \\ \text{ConvNeck}_{l-4}(\mathbf{Z}_{l-1}) + \mathbf{Z}_{l-1}, l > 4 \end{cases} \quad (5)$$

$$\mathbf{Z}_o = \text{Conv}_o(\mathbf{Z}_1 \parallel \mathbf{Z}_2 \parallel \cdots \parallel \mathbf{Z}_l) \quad (6)$$

式中: $\mathbf{Z}$ 为颈部网络中 Concat 模块输出的特征图,其中,Concat 模块的输入为来自骨干网络的特征图或从上一个 MANet 模块输出的特征图; $\mathbf{Z}_m$ 为 $\mathbf{Z}$ 经 1×1 卷积后的中间特征图,其通道数为 $2c$; $\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_l$ 为通道数为 c 的特征图; $\text{Conv}_1(\mathbf{Z})$ 、 Conv_2

($\mathbf{Z}_m$)和 $\text{Conv}_3(\mathbf{Z}_m)$ 为对相应特征图进行 1×1 卷积操作; $\text{DSConv}[\text{Conv}_3(\mathbf{Z}_m)]$ 为对特征图进行深度可分离卷积; $\text{Split}(\mathbf{Z}_m)$ 为分割 $\mathbf{Z}_m$; $\text{ConvNeck}_{l-4}(\mathbf{Z}_{l-1}) + \mathbf{Z}_{l-1}$ 为第 $l-4$ 个卷积颈部操作,其处理输入特征图并加上原始输入,以保留信息并增强特征; $\mathbf{Z}_o$ 为生成 MANet 模块的输出特征图; $\mathbf{Z}_1 \parallel \mathbf{Z}_2 \parallel \cdots \parallel \mathbf{Z}_l$ 为拼接所有生成的特征图; $\text{Conv}_o(\mathbf{Z}_1 \parallel \mathbf{Z}_2 \parallel \cdots \parallel \mathbf{Z}_l)$ 为对拼接后的特征图进行 1×1 卷积操作。

1.2.3 颈部网络与注意力机制的融合

机械加工车间通常环境复杂,背景杂乱,光照条件多变,这些干扰因素会影响模型准确识别作业人员不安全行为。为提高模型性能,在颈部网络末端嵌入 AFGC Attention 模块,该模块利用相关矩阵捕捉全局与局部信息间的相关性,实现特征权重的更有效分配。这一改进可显著增强模型对复杂背景和光照变化的鲁棒性,有效减少因背景干扰和光照变化引起的漏报现象。AFGC Attention 作为一种精细化的通道注意力机制,通过融合全局与局部特征的交互建模,能够精准捕捉通道间的细粒度相关性,从而有效增强特征信息^[13]。其原理是先通过全局平均池化(Global Average Pooling, GAP)对特征图 $\mathbf{F} = \mathbf{Z}_o$ ($\mathbf{F} \in R^{C \times H \times W}$)提取特征向量,生成通道描述符 $\mathbf{U} \in R^C$,其中 $C = 2c$,代表通道数。在此之后,借助带状矩阵 $\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_k]$ 和对角矩阵 $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_h]$ 计算局部特征矩阵 $\mathbf{U}_{lc}$ 与全局特征矩阵 $\mathbf{U}_{gc}$,如下式:

$$\mathbf{U}_n = \text{GAP}(\mathbf{F}_n) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W \mathbf{F}_n(i, j) \quad (7)$$

$$\mathbf{U}_{lc} = \sum_{i=1}^k \mathbf{U} \cdot \mathbf{b}_i \quad (8)$$

$$\mathbf{U}_{gc} = \sum_{i=1}^h \mathbf{U} \cdot \mathbf{d}_i \quad (9)$$

式中: $\mathbf{U}_n$ 为 $\mathbf{U}$ 的第 n 个通道的元素; $\mathbf{F}_n$ 为特征图 $\mathbf{F}$ 中的第 n 个通道; $\text{GAP}(\mathbf{F}_n)$ 为对 $\mathbf{F}_n$ 进行全局平均池化,将特征图 $\mathbf{F}$ 的形状从 $C \times H \times W$ 转换为 $C \times 1 \times 1$; H 为高度; W 为宽度; $\mathbf{F}_n(i, j)$ 为特征图 $\mathbf{F}$ 第 n 个通道在位置 (i, j) 的值; $\mathbf{b}_i$ 为带状矩阵中的第 i 个向量; k 为相邻通道的数量; $\mathbf{d}_i$ 为对角 C 矩阵第 i 个向量; h 为全局通道的数量。

动态融合 $\mathbf{U}_{lc}$ 和 $\mathbf{U}_{gc}$,生成通道权重矩阵 $\mathbf{W}_e$,加权原始特征图,增强显著特征,得到增强后的特征图 $\mathbf{F}^*$,如下式:

$$\mathbf{W}_e = \sigma\{\sigma(\theta) \cdot \sigma(\mathbf{U}_{gc}^w) + [1 - \sigma(\theta)] \cdot \sigma(\mathbf{U}_{lc}^w)\} \quad (10)$$

$$F^* = W \otimes F \quad (11)$$

式中： U_{gc}^o 与 U_{lc}^o 分别为融合后的全局特征矩阵和局部特征矩阵； θ 为一个可学习的参数，通过 sigmoid 激活函数 $\sigma(\theta)$ 来动态调整全局和局部通道权重的融合； $\otimes$ 为 Kronecker 乘积算子^[14]。

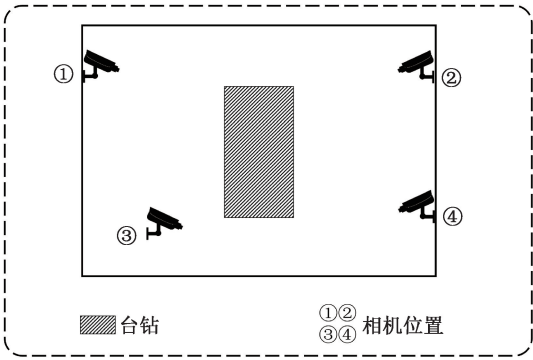
2 不安全行为识别模型算例验证

2.1 数据集与数据处理

鉴于目前缺乏用于专门研究机械加工车间不安全行为识别算法的公开视频图像数据集,以及在实际生产环境中实施不安全行为所存在的风险,采用行为模拟方法获取模型训练和测试所需的视频数据集。首先,选择 4 名经过安全培训的被试进入未进行机械加工生产的车间进行行为模拟,以台式钻床(简称台钻)生产工位所在场景为例,共模拟 6 种典型的不安全行为:①无人值守;②坐姿作业;③未佩戴防护面罩;④未穿着防护服;⑤戴手套作业;⑥戴围巾作业。这 6 种不安全行为分为违规操作和不安全着装 2 大类型。其中,无人值守和坐姿作业属于违规操作,而其他 4 种属于不安全着装。无人值守可能导致机械故障未能及时发现和处理,如传动机

构失灵、轴承磨损等,进而引发安全事故。坐姿操作台钻时,身体的灵活性和反应速度相对较差,一旦发生紧急情况,如工件飞出、设备故障等,操作人员无法迅速采取措施进行应对,增加了事故发生的可能性。在台钻加工过程中会产生大量金属碎屑和颗粒等微小物体,若操作人员未佩戴防护面罩和防护服,这些高速飞溅的微小物体极易接触到眼睛和皮肤,导致划伤、刺伤等伤害。台钻在运转时,主轴和钻头等部件会高速旋转,围巾很容易被卷入这些旋转部件中,导致围巾缠绕在设备上,不仅损坏设备,还可能将操作人员的身体部位一起卷入,造成严重的机械伤害。与围巾类似,手套也容易被卷入台钻的旋转部件,导致手套被缠绕在设备上,进而将手部甚至整个手臂卷入,造成严重的挤压伤、骨折等伤害。因此,操作台钻时戴手套作业和戴围巾作业都属于不安全行为,需要被识别。

在北京市某机械加工工厂的加工车间拍摄并采集该视频数据集。在加工车间的台钻四周,共架设 4 台摄像机,摄像机的参数设置为 2.7 K 分辨率、30 帧,以模拟车间实际监控摄像机的配置。机械加工车间台钻工位平面结构与实际环境如图 3 所示。



(a) 机械加工车间台钻工位平面结构



(b) 机械加工车间实际环境

图 3 机械加工车间台钻工位平面结构与实际环境

Fig. 3 Plan view and actual environment of drill press station in machining workshop

对于不安全着装类行为,模拟试验过程中,4 名被试依次进入试验场景,操作台钻并不断变化工作姿势,模拟机械加工过程。每次模拟试验由 1 名被试参与,进入试验场景的被试根据计划的着装更替方案依次更换着装,每套着装的模拟时长为 1 min,共 10 组着装方案。着装更替方案见表 1。对于坐姿作业行为,试验过程中,4 名被试佩戴护目镜并穿着防护服,依次进入试验场景。每次模拟试验由 1 名被试参与,进入试验场景的被试坐在台钻前的椅子上操作台钻,并不断变换操作姿势,每次模拟时

长为 3 min。对于无人值守行为,则在试验过程中,利用旋转、缩放、翻转和裁剪等方法,针对模拟试验间隙拍摄到的台钻无人操作的视频帧,进行数据增强,以此作为台钻无人值守数据集。无人值守行为识别模型设定为仅在工作时间段开启使用,如果出现台钻及周边区域无人的情况,则认为台钻无人值守,如果台钻及周边区域有人员出现,则认为台钻有人值守。因此,将无人出现的台钻工作台及周边区域整体作为识别目标,设置为无人值守标签,供模型训练使用。

表 1 着装更替方案

Table 1 Clothing replacement scheme				
方案编号	防护面罩	手套	防护服	围巾
1	—	—	—	—
2	√	—	—	—
3	—	√	—	—
4	—	—	√	—
5	—	—	—	√
6	—	√	—	√
7	√	√	—	—
8	√	√	—	√
9	√	—	√	√
10	√	√	√	√

对获取的数据集进行视频帧的抽取处理,并通过 Labelimg 软件进行标注。各识别目标对应的图像数量和标签数量见表 2。将标注后的数据集以 8 : 1 : 1 的比例随机划分为训练集、验证集和测试集。具体用途如下:训练集用于训练模型的权重和参数,通过大量的迭代学习,使模型能够掌握数据中的特征和模式,进而对目标进行检测和分类;验证集用来评估模型在训练过程中的性能,帮助选择最佳的模型参数,该集合提供一个独立的样本集,用于监测模型的过拟合情况,并据此在训练过程中调整模型;测试集用于最终评估模型的整体性能,以验证模型在实际应用中的效果,以及其对新、未见过的目标进行检测和分类能力。

表 2 各识别目标对应的图像数量与标签数量

Table 2 Number of images and labels corresponding to each identification target			
行为类型	识别目标	图像总数	标签总数
不安全着装	未佩戴防护面罩	2 085	1 843
	未穿着防护服	1 933	2 229
	戴手套作业	1 767	2 546
	戴围巾作业	1 789	1 836
违规操作	坐姿作业	4 161	4 295
	台钻无人值守	3 010	2 024

2.2 模型训练参数

采用以下硬件和软件配置进行模型训练与测试:处理器为 Intel Xeon Gold 5320 @ 2.2 GHz,图形处理器为英伟达 NVIDIA A100,深度学习框架采用 PyTorch(版本 2.5.0+cu121),编程语言为 Python 3.9。在训练过程中,采用 mosaic 数据增强方法,输入图像的分辨率为 640×640 像素。训练过程持续 300 个周期,初始学习率设定为 0.01,并逐渐降低至最终学习率 0.001。每次迭代处理的样本数量为

32,权重衰减参数为 0.000 5,优化器采用随机梯度下降。

2.3 模型性能评价指标

模型的性能评价指标采用精确率 P 、召回率 R 、平均精度均值(mean Average Precision, mAP)和 F_1 指数^[15]。

P 是指在所有被模型预测为正类的样本中,实际为正类样本的比例,计算公式为:

$$P = \frac{TP}{TP + FP}$$

(12)

式中:TP(True Positive)为正确识别为正类的样本数量;FP(False Positive)为错误识别为正类的样本数量。

R 是指在所有实际为正类的样本中,被模型正确识别为正类的比例,计算公式为:

$$R = \frac{TP}{TP + FN}$$

(13)

式中 FN(False Negative)为错误识别为负类(实际为正类)样本的数量。

mAP 是目标检测模型性能评估中的一个重要指标,衡量模型在 n 个类别上的平均精确率(Average Precision, AP)。mAP@0.5 是指在交并比(Intersection over Union, IoU)阈值为 0.5 时计算的 mAP。IoU 是一个衡量预测框和真实框重叠程度的指标,其值范围为 0~1,值越高表示重叠度越大。选择 mAP@0.5 作为平均精度指标。计算公式为:

$$AP = \int_0^1 P(y) dy$$

(14)

$$mAP = \frac{1}{n} \sum_{i=0}^n AP_i$$

(15)

F_1 是 P 和 R 的调和平均数,其取值范围为 0~1,值越高表示模型的综合性能越好。可见:当精确率和召回率都较高时, F_1 值也较高。计算公式为:

$$F_1 = 2 \cdot \frac{P \cdot R}{P + R}$$

(16)

3 不安全行为识别结果分析

表 3 为机械加工车间作业人员不安全行为识别模型的性能评价结果,其中,括号内为改进模型 YOLOv11x_MMA 的性能评价结果,括号外为未使用任何改进模块的 YOLOv11x 模型的结果。从表 3 可以看出,对于台式钻床无人值守、未佩戴面罩作业、未穿着防护服作业 3 种行为,YOLOv11x 模型与改进后的 YOLOv11x_MMA 模型的精确率、召回率、

mAP@0.5 以及 F_1 值均达到 0.9 以上。特别是对无人值守和未穿着防护服作业行为的识别,这些指标均接近 1.0。对于戴围巾作业和坐姿作业 2 种行为,2 个模型的精确率、召回率、mAP@0.5 以及 F_1 值均达到较高水平,均超过 0.8。其中,改进模型

YOLOv11x_MMA 对精确率的提升尤为显著。尽管改进模型的召回率有所下降,但 F_1 指数均有所提升。值得注意的是,由于手套这一识别目标尺寸小,且在加工过程中易被遮挡,YOLOv11x 模型对于戴手套作业行为的识别性能较为有限。

表 3 不安全行为识别模型的性能评价结果

行为类型	P	R	mAP@0.5	F_1
未佩戴口罩	0.944 (0.941)	0.926 (0.923)	0.954 (0.953)	0.935 (0.932)
未穿防护服	1.000 (0.999)	0.987 (0.983)	0.995 (0.995)	0.993 (0.991)
戴手套作业	0.706 (0.735)	0.663 (0.682)	0.604 (0.651)	0.684 (0.708)
戴围巾作业	0.842 (0.880)	0.839 (0.826)	0.830 (0.841)	0.840 (0.852)
坐姿作业	0.882 (0.911)	0.868 (0.845)	0.952 (0.944)	0.875 (0.877)
无人值守	0.991 (0.982)	0.989 (0.989)	0.994 (0.99)	0.990 (0.985)

表 4 为戴手套作业行为识别的消融试验结果。从表 4 可以看出,当采用 1.2 节所述的全部 3 种改进措施时,该模型对戴手套作业中的小尺寸目标(即手套)的识别性能最佳,精确率和召回率分别达

到 0.735 和 0.682,因此, F_1 指数也达到最高值 0.708,而 mAP @ 0.5 则仅次于仅采用 AFGC Attention 模块的结果。上述消融试验结果表明:改进措施具有显著的性能提升效果。

表 4 戴手套作业行为识别模型的消融试验结果

Table 4 Ablation experiment results for glove-wearing operation behavior identification						
MANet	MetaFormer	AFGC Attention	P	R	mAP@0.5	F_1
—	—	—	0.706	0.663	0.604	0.684
✓	—	—	0.687	0.663	0.606	0.675
—	✓	—	0.693	0.682	0.645	0.687
—	—	✓	0.734	0.654	0.661	0.692
✓	✓	—	0.731	0.668	0.629	0.698
✓	—	✓	0.718	0.605	0.621	0.657
—	✓	✓	0.657	0.67	0.596	0.663
✓	✓	✓	0.735	0.682	0.651	0.708

表 5 为 YOLOv11x_MMA 与其他模型(包括 Faster R-CNN、SSD 以及 YOLO 模型的其他版本:YOLOv8x、YOLOv9e、YOLOv10x、YOLOv11x)针对戴手套行为的对比试验结果。由表 5 可知:YOLOv11x_MMA 在精确率、召回率、 F_1 值以及 mAP@0.5 值上均优于其他所有模型。这表明:改进后的 YOLOv11x_MMA 模型对于机械加工车间戴手套行为的识别效果相较其他模型有显著的提升。

表 5 戴手套作业行为识别模型的对比试验结果

Table 5 Comparative experiment results for glove-wearing operation behavior identification				
模型	P	R	mAP@0.5	F_1
Faster R-CNN	0.278	0.356	0.202	0.312
SSD	0.628	0.427	0.405	0.508
YOLOv8x	0.710	0.644	0.633	0.675

续表 5

模型	P	R	mAP@0.5	F_1
YOLOv9e	0.693	0.632	0.644	0.661
YOLOv10x	0.660	0.621	0.621	0.640
YOLOv11x	0.706	0.663	0.604	0.684
YOLOv11x_MMA	0.735	0.682	0.651	0.708

4 结 论

1) 机械加工车间作业人员不安全行为识别模型的风险识别结果表明:对无人值守、未佩戴口罩作业、未穿着防护服作业 3 种行为识别的 F_1 指数均超过 0.93;对于戴围巾作业、坐姿作业 2 种行为识别的 F_1 指数均超过 0.84。

2) 该模型的小尺寸目标识别性能有所提升。然而,在机械加工车间场景中,工作人员的手部在加

工过程中极易被其他物体遮挡,这导致模型对手部目标的识别效果受到一定的影响。尽管如此,其对戴手套作业行为识别的 F_1 指数仍从 0.684 提高到 0.708, $mAP@0.5$ 值也从 0.604 提升到 0.651,可为风险预警提供依据。

3) 基于 YOLOv11 的机械加工车间作业人员不安全行为识别模型,在真实车间台式钻床区域环境中采集的视频数据集上显示出较高的识别性能,下一步可继续构建更丰富的视频数据集,持续改进和优化模型,不断提升其精准性与实用性。

参考文献

- [1] YANG Meng, WU Chengke, GUO Yuanjun, et al. Transformer-based deep learning model and video dataset for unsafe action identification in construction projects[J]. Automation in Construction, 2023, 146: DOI: 10.1016/j.autcon.2022.104703.
- [2] 佟瑞鹏,张艳伟.人工智能技术在矿工不安全行为识别中的融合应用[J].中国安全科学学报,2019,29(1):7-12. TONG Ruipeng, ZHANG Yanwei. Integration between artificial intelligence technologies for miners' unsafe behavior identification[J]. China Safety Science Journal, 2019, 29(1): 7-12.
- [3] 赵荣泳,韦炳宇,朱文杰,等.公共场所行人异常行为识别方法综述[J].中国安全科学学报,2024,34(2):83-93. ZHAO Rongyong, WEI Bingyu, ZHU Wenjie, et al. Overview of recognition methods of pedestrian abnormal behaviors in public places[J]. China Safety Science Journal, 2024, 34(2): 83-93.
- [4] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]. Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2016:779-788.
- [5] LI Lijuan, ZHANG Peng, YANG Shipin, et al. YOLOv5-SFE: an algorithm fusing spatio-temporal features for detecting and recognizing workers' operating behaviors[J]. Advanced Engineering Informatics, 2023, 56: DOI: 10.1016/j.aei.2023.101988.
- [6] FU Yuanyuan, RAN Teng, XIAO Wendong, et al. GD-YOLO: an improved convolutional neural network architecture for real-time detection of smoking and phone use behaviors[J]. Digital Signal Processing, 2024, 151: DOI: 10.1016/j.dsp.2024.104554.
- [7] XIN Fangfang, HE Xinyu, YAO Caoxiu, et al. A real-time detection for miner behavior via DYS-YOLOv8n model[J]. Journal of Real-Time Image Processing, 2024, 21: DOI: 10.1007/s11554-024-01466-0.
- [8] 罗国富,王源,李浩,等.基于改进 YOLOv5s 的智能车间工人不安全行为实时检测方法[J].计算机集成制造系统,2024,30(5):1 610-1 619. LUO Guofu, WANG Yuan, LI Hao, et al. Unsafe behavior real-time detection method of intelligent workshop workers based on improved YOLOv5s[J]. Computer Integrated Manufacturing Systems, 2024, 30(5): 1 610-1 619.
- [9] 韩康,李敬兆,陶荣颖.基于改进 YOLOv7 和 ByteTrack 的煤矿关键岗位人员不安全行为识别[J].工矿自动化,2024,50(3):82-91. HAN Kang, LI Jingzhao, TAO Rongying. Recognition of unsafe behaviors of key position personnel in coal mines based on improved YOLOv7 and ByteTrack[J]. Journal of Mine Automation, 2024, 50(3): 82-91.
- [10] 徐浩钧,胡啸峰,吴建松.化学实验室动火实验无人值守行为识别方法研究[J].中国安全生产科学技术,2023,19(12):135-141. XU Haojun, HU Xiaofeng, WU Jiansong. Research on identification method of unattended behavior in hot experiments of chemical laboratory[J]. Journal of Safety Science and Technology, 2023, 19(12): 135-141.
- [11] YU Weihao, SI Chenyang, ZHOU Pan, et al. MetaFormer baselines for vision[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(2): 896-912.
- [12] FENG Yifan, HUANG Jiangang, DU Shaoyi, et al. Hyper-YOLO: when visual object detection meets hypergraph computation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2025, 47(4): 2 388-2 401.
- [13] SUN Hang, WEN Yang, FENG Huijing, et al. Unsupervised bidirectional contrastive reconstruction and adaptive fine-grained channel attention networks for image dehazing[J]. Neural Networks, 2024, 176: DOI: 10.1016/j.neunet.2024.106314.
- [14] GRAHAM A. Kronecker products and matrix calculus with applications[M]. Horwood: Halsted Press, 1981: 21-35.
- [15] WITTEN I H, FRANK E, HALL M A, et al. Data mining: practical machine learning tools and techniques[M]. Burlington: Morgan Kaufmann, 2016: 181-186.



作者简介: 胡啸峰 (1986—),男,河北唐山人,工学博士,副教授,主要从事风险评估与预测预警技术研究。E-mail: huxiaofeng@ppsuc.edu.cn。