陈伟,赵广佳,路勇. 基于 YOLOv8 的油田井场人员安全服穿戴检测方法[J]. 智能计算机与应用,2025,15(4):61-68. DOI:10.20169/j.issn.2095-2163.24111503

基于 YOLOv8 的油田井场人员安全服穿戴检测方法

陈 伟[1], 赵广佳[2], 路 勇[2]
(1 西安石油大学 计算机学院, 西安 710065; 2 中国石油长庆油田分公司 第三采油厂, 西安 717600)

摘 要: 由于油田井场的监控视频视角范围广、拍摄距离远,人员图像较小且与周围设备混杂,目前使用的视频检测方法在检测井场人员安全服穿戴方面未能达到精度要求,本文提出了基于 YOLOv8 算法改进的检测模型。首先,添加 Shuffle Attention 注意力机制,加强网络提取正确着装和安全帽特征的能力,提高检测的精度;其次,在快速空间金字塔池化 SPPF 中添加可分离大核聚集模块 LSKA,通过提高长距离特征依赖增强全局特征信息融合;最后,将损失函数改为 *Wise-IoU*,使网络可以动态聚焦学习,更加专注于关键特征,提高了模型的泛化能力,减小低质量样本的负面影响。实验结果表明,改进后的模型在数据集上将精确率 *P* 提升了 7.4%,召回率、mAP_{50}/%以及 $mAP_{50:95}$/%都有所提升。说明本文改进后的模型对复杂场景中特征小目标的检测更加有效。

关键词: YOLOv8; 安全穿戴; 注意力机制; 损失函数; 油田井场
中图分类号: TP391.41 **文献标志码**: A **文章编号**: 2095-2163(2025)04-0061-08

Detection method of safety clothing for personnel in oil well site based on YOLOv8

CHEN Wei[1], ZHAO Guangjia[2], LU Yong[2]
(1 School of Computer Science, Xi'an Shiyou University, Xi'an 710065, China;
2 The Third Oil Production Plant of PetroChina Changqing Oilfield Branch, Xi'an 717600, China)

Abstract: Due to the wide viewing angle range and long shooting distance of the monitoring video of the oilfield well site, and the situation that the personnel images are small and mixed with the surrounding equipment, the current video detection methods can not meet the accuracy requirements in detecting the safety clothing of the personnel in the oil well site. This paper proposes an improved detection model based on the YOLOv8 algorithm. Firstly, the Shuffle Attention mechanism is added to strengthen the ability of the network to extract the features of correct dress and helmet, so as to improve the detection accuracy. Secondly, the Large Separable Kernel Attention (LSKA) is added after the Spatial Pyramid Pooling-Fast(SPPF) to enhance the global feature information fusion by improving the long-distance feature dependence. Finally, the loss function is changed to *Wise-IoU*, which allows the network to dynamically focus on learning and pay more attention to key features, thereby improving the generalization ability of the model and reducing the negative impact of low-quality samples. The experimental results show that the improved model improves the accuracy *P* by 7.4% in the data set, and the recall rate, mAP_{50}/% and $MAP_{50:95}$/% are all improved. It is verified that the improved model is more effective to detect small feature targets in complex scenes.

Key words: YOLOv8; safety clothing; attention mechanism; loss function; oilfield well site

0 引 言

由于油田井场中的监控视频大多视角范围广、拍摄距离远、图像中物体繁多,人物尺寸较小,目前使用的视频检测方法在检测井场人员安全服穿戴方面达到不到精度要求,因此,提升检测精度是急需解决的问题。针对油田作业现场监控视频中油田井场人员安全服穿戴和小目标检测效果较差的问题,田枫等学者[1]提出了改进 YOLOv5 的油田场景规范化着装检测方法 Cascade-YOLOv5。针对施工现场小目标工人检测困难的问题,李建华等学者[2]提出了一种融合改进的 YOLO 模型与帧差法的综合目标检测方法,方法实时性虽然能满足实际应用的需求,但对算力的要求较高。师盼武等学者[3]利用多尺度特征融合模块(MFFM)和 FSI-ECA 注意力模块,提出了一个安全帽佩戴检测网络 MFFMA-NET,

作者简介: 陈 伟(1999—),男,硕士研究生,主要研究方向:油田智能信息化。Email:674831348@qq.com; 赵广佳(1983—),女,采油工程师,主要研究方向:油气田开发; 路 勇(1981—),男,采油工程师,主要研究方向:油气田开发。

收稿日期: 2024-11-15

但是 MFFMA-NET 的检测速度并不很快,需要解决如何在保持检测精度的条件下提高检测速度的问题。张贝贝等学者[4]提出了采用 SSD 算法在 Caffe-SSD 框架下实现目标检测,但是小尺寸安全帽检测的精度仍有待进一步的提升。

本文为了解决油田井场人员安全服穿戴检测精度不高的问题,在 YOLOv8 的基础上进行改进,首先使用了通道混洗注意力机制 Shuffle Attention(SA)模块,将架构中的 C2f 替换成为 C2f-Shuffle Attention 结构,实现了空间注意力机制与通道注意力机制的有效结合,提高人物检测精度。其次,添加可分离大核卷积模块(LSKA),加强位置信息和全局信息关注,在提高特征提取能力同时也提高了特征表达能力且减少了计算成本。最后,将原有的 YOLOv8 的损失函数替换为 *Wise - IoU* 进行边界回归损失计算,降低低质量样本的影响,提高模型检测精度。

1 改进的安全服穿戴检测模型

1.1 YOLOv8 模型

YOLOv8 是 Ultralytics 在 2023 年提出的最新 YOLO 系列检测模型。YOLOv8 有 n,s,m,l,x 五种不同大小的模型,不同大小模型的运行速度及参数量见表 1。YOLOv8 模型的结构包括主干网络(Backbone)、特征融合网络(Neck)和预测网络(Head)三个部分。本文采用的是 YOLOv8n。改进后的模型结构如图 1 所示。

表 1 模型的运行速度及参数量

Table 1 Running speed and number of parameters of the model

Model	*Size* / pixels	Speed CPU ONNX/ms	Speed A100TensorRT/ms	*Params*/ M	*FLOPs*/ B
YOLOv8n	640	80.4	0.99	3.2	8.7
YOLOv8s	640	128.4	1.20	11.2	28.6
YOLOv8m	640	234.7	1.83	25.9	78.9
YOLOv8l	640	375.2	2.39	43.7	165.2
YOLOv8x	640	479.1	3.53	68.2	257.8

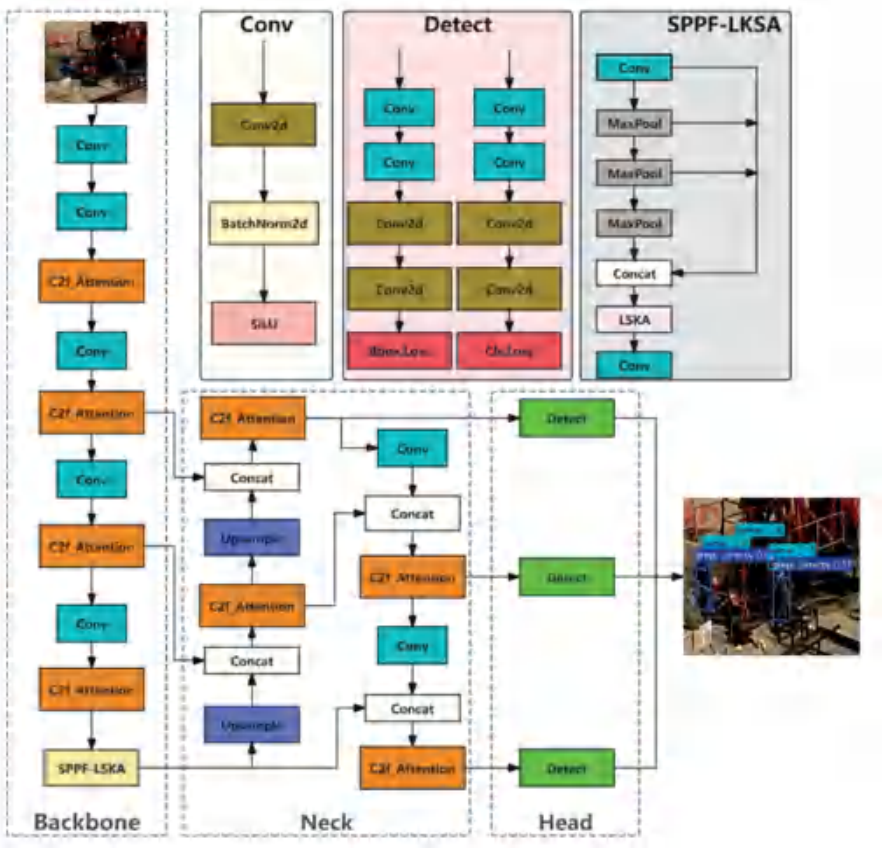

图 1 改进后模型结构

Fig. 1 Improved model structure

1.2　通道混洗注意力机制

通道混洗注意力机制[5]（Shuffle Attention，SA）是一种用于提升深度学习模型特征表示能力的轻量级注意力机制。用 SA 机制来替代原本模型中的 C2f 模块，可以解决训练速度过慢的问题。通过将通道混洗（Channel Shuffle）[6]操作与自注意力机制结合，旨在更有效地进行特征通道之间的信息交互。传统的注意力机制，如 SE 注意力机制[7]以及 CBAM 注意力机制[8]通常都伴随较大的计算开销。而本文采用的 SA 机制通过分组和通道混洗来打破这种限制，实现计算效率与性能的平衡。SA 机制的结构如图 2 所示，主要由以下 4 个部分组成。

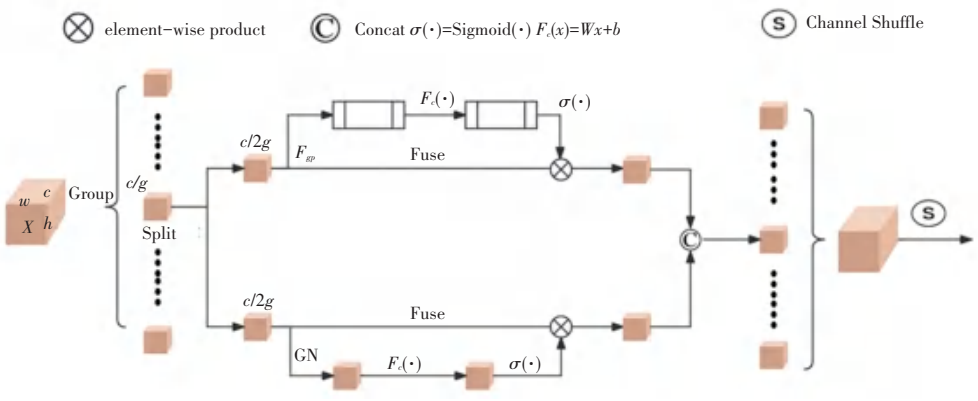

图 2　Shuffle Attention 结构图

Fig. 2　Shuffle Attention structure

1.2.1　特征分组

对于给定的特征图 $X \in R^{C\times W\times H}$，其中 C、W 和 H 分别表示特征图的通道、宽度和高度。首先，将特征 X 分解为 G 组 $X=[X_1,X_1,\cdots,X_1]$，$X_i \in R^{(C/G)\times W\times H}$；然后，将 X_i 沿输入通道维度分成 2 个分支 $X_{i1},X_{i2} \in R^{(C/G)/2\times W\times H}$，其中一个分支利用通道之间的相互关系来生成通道注意力图，而另一个分支则利用特征之间的空间关系来生成空间注意力图。

1.2.2　通道注意力

首先，利用全局平均池化函数获得通道统计信息；其次，利用融合线性函数增强特征表示；最后，将全局信息与 Sigmoid 激活函数激活后的原始特征值相乘，由此获得包含通道注意力权重 $X_{i1}^{'}$，从而增强特征。计算公式具体如下：

$$X_{i1}^{'}=\sigma(F_C(s_1))\times X_{i1}=\sigma(w_1s_1+b_1)\times X_{i1} \quad (1)$$

其中，线性函数 $F_C(x)=Wx+b(W,b \in R^{(C/2G)\times 1\times 1})$；$\sigma$ 表示 Sigmoid 激活函数；s_1 表示平均池化特征；w_1、b_1 可通过训练获得且 $w_1,b_1 \in R^{(C/G)/2\times 1\times 1}$。

根据以上分析，进一步推得：

$$s_1=F_{gp}(X_{i1})=\frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}X_{i1}(i,j) \quad (2)$$

其中，F_{gp} 表示全局平均池化函数。

1.2.3　空间注意力

与通道注意力不同，空间注意力[9]主要是对通道注意力的补充。首先，利用组归一化函数获得空间统计信息；然后，利用融合线性函数对特征表示增强；最后，通过与 Sigmoid 激活函数激活后的原始特征值相乘来嵌入全局信息，以获得包含空间注意力权重 $X_{i2}^{'}$ 的全局信息，从而增强特征对于特定区域的重要性。用到的公式为：

$$X_{i2}^{'}=\sigma(w_2\times GN(X_{i2})+b_2)\times X_{i2}=\sigma(w_2s_2+b_2)\times X_{i2} \quad (3)$$

其中，GN（Group Normalization）表示组归一化函数；s_2 表示归一化特征；w_2、b_2 可通过训练获得且 $w_2,b_2 \in R^{(C/G)/2\times 1\times 1}$。

1.2.4　聚合

融合通道注意力权重 $X_{i1}^{'}$ 和空间注意力权重 $X_{i2}^{'}$，返回分组维度 $CG^{-1}\times W\times H$。再次合并分组块，返回到原始维度 $C\times W\times H$。在完成注意力学习和重新校准特征后，拼接并聚合 2 个分支。当所有子特征聚合完毕后，执行信道分组操作。

1.3　可分离大核卷积模块

可分离大核卷积模块（LSKA）是一种可分离核注意力模块[10]，通过分解大核卷积操作来捕捉长程依赖性，将二维卷积核拆分为串联的一维卷积核，降低计算复杂度和内存需求。该模块能更好地建模全局上下文信息，提高空间关系和语义信息的捕捉能力，适应不同尺度的目标特征表示。LSKA 在大核聚焦模块的基础上将 2D 卷积核分解为级联的水平

和垂直 1D 核,进一步提高特征提取能力和特征表达能力,同时减少了计算成本。

LSKA 结构如图 3 左所示,SPPF-LSKA 模块结构如图 3 右所示。LSKA 输出的计算公式分别如下:

$$\bar{Z}^C = \left(\sum_{H,W} W^C_{1\times(2d-1)} * F^C\right) * \sum_{H,W} W^C_{(2d-1)\times 1} \tag{4}$$

$$Z^C = \left(\sum_{H,W} W^C_{1\times\lfloor k/d\rfloor} * \bar{Z}^C\right) * \sum_{H,W} W^C_{\lfloor k/d\rfloor\times 1} \tag{5}$$

$$A^C = Z^C * W_{1\times 1} \tag{6}$$

$$\bar{F}^C = F^C \otimes A^C \tag{7}$$

其中,F^C 表示输入特征图;C 表示输入通道数;$\bar{Z}^C$ 表示深度卷积后的输出;Z^C 表示膨胀卷积后的输出;H 和 W 分别表示特征图中的宽度和高度;“$\lfloor . \rfloor$”表示向下舍入操作;d 表示扩张率;k 表示内核大小;“ * ”表示卷积操作;A^C 表示注意力图;$\bar{F}^C$ 表示输出特征图,“$\otimes$”表示 Hadamard 积。

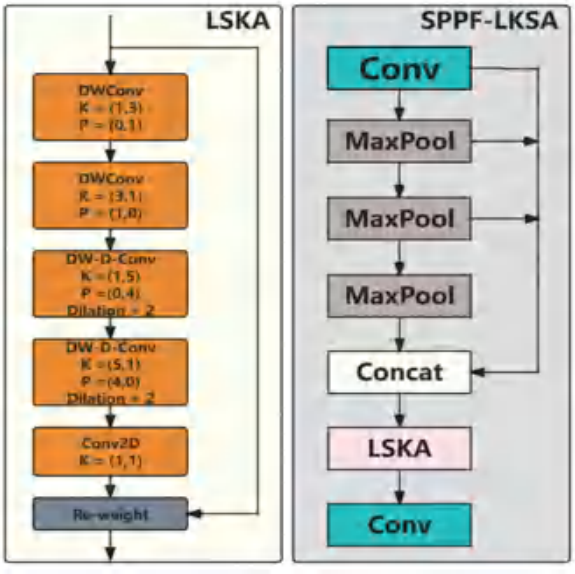

图 3 LSKA 结构图

Fig. 3 LSKA structure

1.4 *Wise-IoU* 损失函数

数据集制作过程中,标注框过大或过小会影响特征学习,错误标注会导致错误特征,数据集难以避免存在一些低质量示例,如果强化对低质量边界框的回归,会损害模型的检测性能。因此,本文采用具有动态聚焦机制的 *Wise - IoU*[11] 为回归损失,代替原本的 *CIoU*[12] 损失。*Wise - IoU* 的核心点是在传统 *IoU* 损失的基础上结合了注意力函数和非单调动态聚焦系数。计算方法见式(8) ~ 式(11):

$$R_{\mathrm{WIoU}} = e^{\frac{\rho^2(b,b^{\mathrm{gt}})}{W_g^2+H_g^2}} \tag{8}$$

$$\beta = \frac{L^*_{\mathrm{IoU}}}{\bar{L}_{\mathrm{IoU}}} \tag{9}$$

$$\gamma = \frac{\beta}{\delta\theta^{\beta-\delta}} \tag{10}$$

$$L_{\mathrm{WIoU}-v3} = \gamma \times R_{\mathrm{WIoU}} \times L_{\mathrm{IoU}} \tag{11}$$

其中,R_{WIoU} 表示注意力函数;W_g 和 H_g 分别表示真实框和预测框最小外接矩形的长和宽;β 表示离群度;L^*_{IoU} 表示 *IoU* 损失的梯度增益;$\bar{L}_{\mathrm{IoU}}$ 表示 *IoU* 损失的滑动平均;θ 和 δ 表示用于构造系数非单调性的超参数;γ 表示非单调动态聚焦系数。非单调动态聚焦系数用离群度衡量标注框的质量,质量较高的标注框示例具有较小的离群度,匹配的梯度增益较小,这使得 *Wise - IoU* 在每个时刻都能做出符合当前情况的最优梯度增益分配策略。

2 实验效果及评价

2.1 实验数据集

实验数据集通过对东部某油田井场的实时监控视频进行处理后获得,首先对监控视频进行分帧和剪裁处理,删除了没有人员的图片,然后采用 Labelimg 工具对图片进行标注,图片格式为 png,图片分辨率为 1 280×720。最终得到了 4 200 张图片,按照 8∶1∶1 的比例划分训练集、验证集和测试集,使用图片如图 4 所示。标注后图片如图 5 所示。

图 4 原始数据集样本

Fig. 4 Samples of original dataset

图 5　Labelimg 标签样本

Fig. 5　Labelimg label sample

2.2　实验环境

实验的操作系统版本为 Windows10，Torch 版本为 1.9.1，Python 版本为 3.8，GPU 为 NVIDIA GeForce GTX 1660S，CPU 为 12th Gen Intel Core i3-12100f。每次训练的轮次为 200，*batch* 为 16，工作线程为 8。

2.3　评价指标

在目标检测任务中，评价网络性能的指标主要有以下几种：精确率（P）、召回率（R）、平均精度（mAP）。精确率（P）的数学定义如下式所示：

$$P = \frac{TP}{TP + FP} \tag{12}$$

其中，TP 表示真正例，FP 表示是假正例。

召回率（R）的数学定义如下式所示：

$$R = \frac{TP}{TP + FN} \tag{13}$$

将每个类别的准确率（P）和召回率（R）绘制成 $P-R$ 曲线，对该线进行积分，得到检测精度（AP），数学定义如下式所示：

$$AP = \int_0^1 P(r)\,\mathrm{d}r \tag{14}$$

再对所有类别的检测精度 AP 取平均值，得到平均检测精度（mAP），数学定义如下式所示：

$$mAP = \frac{\sum_{i=1}^{c} AP_i}{C} \tag{15}$$

其中，C 表示所有的类别数，本文为 1。mAP 表示平均精度均值，对所有类别的平均精度求均值。mAP 常用 $mAP@_{50}/\%$ 和 $mAP@_{50:95}/\%$ 两个指标来衡量，前者表示在 *IoU* 阈值取 0.5 时的平均精度均值，后者则表示对 *IoU* 取值从 0.50 到 0.95、步长为 0.05 的 10 组 mAP 求均值的结果。

2.4　实验结果

2.4.1　Shuffle Attention 模块分析

为了验证 SA 模块在目标检测算法中的有效性和鲁棒性，实验以 YOLOV8n 为基准模型，在主干网络中分别引入了 EMA 注意力机制[13]、CBAM 注意力机制以及 TripletAttention 注意力机制[14]，并在自制数据集上进行实验来验证。对比检测结果见表 2。

表 2　不同模块对比实验

Table 2　Comparative experiments of different modules　%

Models	P	R	mAP_{50}	$mAP_{50:95}$
YOLOv8n	74.49	76.54	75.33	31.91
+EMA	75.62	75.98	75.72	29.74
+CBAM	75.90	73.47	77.33	32.66
+Triplet Attention	76.54	78.74	77.33	32.91
+Shuffle Attention	77.05	80.04	79.46	36.63

由表 2 可知，在自制数据集对比本文提出的用 SA 模块代替 C2f 模块表现出最好效果，通过分组和通道混洗实现精度和效率上的提高，在评价指标上较基准模型和其他几种注意力机制都有一定程度的提高。

2.4.2　SPPF 中加入不同注意力机制对比实验

为探究不同注意力机制对检测性能的影响，将常见的 CPCA 注意力机制[15]、以 MHSA 注意力机制[16]、DA 注意力机制[17]及本文使用的 LSKA 注意力机制分别添加到 YOLOv8n 中的 SPPF[18]结构中，对比检测结果见表 3。

表 3　不同机制对比实验

Table 3　Comparative experiments of different mechanisms　%

添加注意力机制	P	R	mAP_{50}	$mAP_{50:95}$
YOLOv8n	74.49	76.54	75.33	31.91
+MHSA	76.52	80.15	78.71	35.16
+DA	76.11	79.11	77.81	35.81
+CPCA	77.66	73.05	76.20	33.94
+LSKA	78.16	78.15	78.26	35.72

根据表 3 中的数据分析，引入注意力机制使网络聚焦于关键的特征部分，有助于提高目标的定位和分类精度。对比其他几种模型能得出加入 LSKA 后的检测效果最好。

2.4.3　损失函数改进效果实验

为了验证 *Wise - IoU* 损失函数在算法中的有效

性,实验以 YOLOv8 中默认的损失函数 *CIoU* 以及作为对比加入 *DIoU*[12]、*GIoU*[19] 与 *Wise - IoU*,实验结果见表 4。

表 4 不同损失函数对比实验

Table 4 Comparative experiment of different loss functions %

损失函数	*P*	*R*	mAP_{50}	$mAP_{50:95}$
CIoU	74.49	76.54	75.33	31.91
DIoU	75.15	75.82	74.61	31.56
GIoU	75.18	77.72	75.24	30.89
Wise - IoU	76.52	78.04	78.51	35.16

由表 4 可知,相较于 *CIoU* 损失函数,使用 *Wise - IoU* 损失函数后几种指标均有所提升,证实了结合了注意力函数和非单调动态聚焦系数的 *Wise - IoU* 在 YOLOv8 中有着显著优越性。

2.4.4 消融实验

为了验证本文改进算法的有效性,本文在自制数据集上使用相同训练参数进行了消融实验。在原始的 YOLOv8n 模型的基础上,本文依次进行了以下修改:使用 Shuffle Attention 模块代替原有模型中的 C2f 模块、在 SPPF 中加入 LSKA 模块、使用 *Wise - IoU* 代替原有模块中的 *CIoU*。对这些方法进行组合测试,最终得到的实验结果见表 5。

表 5 消融实验结果对比

Table 5 Comparison of ablation experiment results %

方法	Shuffle Attention 模块	SPPF 中加 LSKA	*Wise - IoU* 损失函数	*R*	mAP_{50}	$mAP_{50:95}$
YOLOv8n				76.54	75.33	31.91
优化 1	√			80.04	79.46	36.63
优化 2		√		78.15	78.26	35.72
优化 3			√	78.04	78.51	35.16
优化 4	√	√		84.06	82.63	38.44
优化 5	√		√	78.42	77.51	35.22
优化 6		√	√	83.33	81.94	37.74
优化 7	√	√	√	85.68	83.84	38.85

由表 5 可以得出,加入 Shuffle Attention 模块代替原有模型中的 C2f 模块、在 SPPF 中加入 LSKA 模块、使用 *Wise - IoU* 代替原有模块中的 *CIoU* 都能提升原始的检测效果。其中,在 SPPF 中加入 LSKA 模块对各项指标均有较大的提升,通过大核卷积和注意力机制,LSKA 模块能够更好地理解全局上下文,提高对目标的空间关系和语义信息的捕捉能力。使用 Shuffle Attention 模块代替原有模型中的 C2f 模块在回归率上有着显著的影响,相比原有的模型提升了 3.5%,这是由于 Shuffle Attention 模块通过分组和通道混洗来解决了传统的通道数多而开销大的这种缺陷,实现计算效率与性能的平衡。使用 *Wise - IoU* 代替原有模块中的 *CIoU* 则是相较于原模型将精确率 *P* 提升了 2.3%,回归率提升了 1.5%,mAP_{50}/% 提升了 3.18%以及 $mAP_{50:95}$/% 提升了 3.25%。在加入了所有的优化后,优化 7 则是相较于 YOLOv8n 原模型有了可观的提升,由于 Shuffle Attention 模块结合了自注意力机制,能够更有效地进行特征通道之间的信息交互,并且将 LSKA 与 SPPF 相结合,可以在保持多尺度特征融合的基础上,还使得特征表达能力得以大大增强。再者使用 *Wise - IoU* 代替原有模块中的 *CIoU* 后使得 *Wise - IoU* 在每个时刻都能做出符合当前情况的最优梯度增益分配策略,3 种优化相互补充,将精确率 *P* 提升了 7.4%,回归率提升了 9.41%,mAP_{50}/% 提升了8.51% 以及 $mAP_{50:95}$/% 提升了 6.94%。

2.4.5 改进前后效果图对比

为了验证改进模型的检测效果,选择了 2 组检测图片进行对比,如图 6 所示。分析可知,检测精度有显著提升。在第一组检测中,原模型只检测出了 1 个安全帽和 1 个正确着装,而改进后的模型检测出了图中 3 个安全帽和 2 个正确着装;在第 2 组检测中,原模型只检测出了 2 个正确着装,而改进后的模型则是将 3 个安全帽和 2 个正确着装全部检测到,验证了改进算法的有效性。第 3 组检测中,原模型只检测出了 1 个未戴安全帽和 1 个正确着装,而改进后的模型检测出了 3 个未戴安全帽和正确着装,并且在检测精度上获得显著提升。在实际场景中还将设置可视化报警系统,当检测到未戴安全帽或并未正确着装时可及时报警。

图 6　检测效果对比

Fig. 6　Detection effect comparison

3　结束语

本文针对油田井场人员安全服穿戴是否合规的问题，采用目标检测的方法，对监控图像中的安全服穿戴进行检测。为了提升目标的检测效果，本文在 YOLOv8n 的基础上使用 SA 模块改进了 C2f 模块，增强特征空间信息以及通道信息之间的关系，提升算法的特征学习能力，让算法更关注有用的信息；在 SPPF 中加入 LSKA 来加强全局信息特征提取，进一步提高特征提取能力，同时减少了计算成本；引入 *SIoU* 损失函数优化边框回归损失函数，使得网络在模型训练过程中稳定收敛，以提高训练速度和精度。本文在自制数据集上进行检测，为了提升目标的检测效果，用相同的训练参数，对比了主流算法和改进后算法的性能，验证了改进后的有效性，在精确率 P、回归率 R，mAP_{50}/% 以及 $mAP_{50:95}$/% 上均有提升。

参考文献

[1] 田枫，贾昊鹏，刘芳. 改进 YOLOv5 的油田作业现场安全着装小目标检测[J]. 计算机系统应用，2022，31(3)：159-168.

[2] 李建华，韩宇，石开铭，等. 施工现场小目标工人检测方法[J]. 图学学报，2024，45(5)：1040-1049.

[3] 师盼武，冯百明，丁洪文. 施工现场安全帽佩戴检测方法研究[J]. 微电子学与计算机，2024，41(10)：45-54.

[4] 张贝贝，程科，钱倩倩，等. 基于深度学习的施工现场安全帽佩戴检测的研究[J]. 计算机与数字工程，2023，51(7)：1657-1662.

[5] ZHANG Qinglong，YANG Yubin. SA-Net：Shuffle attention for deep convolutional neural networks [C] //Proceedings of IEEE International Conference on Acoustics，Speech and Signal Processing(ICASSP). Piscataway，NJ：IEEE，2021：2235-2239.

[6] ZHANG Xiangyu，ZHOU Xinyu，LIN Mengxiao，et al. ShuffleNet：An extremely efficient convolutional neural network for mobile devices [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway，NJ：IEEE，2018：6848-6856.

[7] HU Jie，SHEN Li，ALBANIE S，et al. Squeeze-and-excitation networks [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway，NJ：IEEE，2018：7132-7141.

[8] WOO S，PARK J，LEE J Y，et al. CBAM：Convolution block attention module[C]//Proceedings of the European Conference on Computer Vision(ECCV). Cham：Springer，2018：3-19.

[9] JADERBERG M，SIMONYAN K，ZISSERMAN A，et al. Spatial transformer networks [J]. arXiv preprint arXiv，1506. 02025，2015.

[10] LAU K W，PO L M，REHMAN Y A. Large separable kernel attention：Rethinking the large kernel attention design in CNN [J]. Expert Systems with Applications，2023，236：121352.

[11] TONG Zanjia，CHEN Yuhang，XU Zewei，et al. Wise-IoU：Bounding box regression loss with dynamic focusing mechanism [J]. arXiv preprint arXiv，2301. 10051，2023.

[12] ZHENG Zhaohui，WANG Ping，LIU Wei，et al. Distance-IoU loss：Faster and better learning for bounding box regression [J]. arXiv preprint arXiv，1911. 08287，2019.

[13] LI Xia, ZHONG Zhisheng, WU Jianlong, et al. Expectation-maximization attention networks for semantic segmentation[C] // Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Picataway, NJ: IEEE, 2019: 9166-9175.

[14] MISRA D, NALAMADA T, ARASANIPALAI U A, et al. Rotate to attend: Convolutional triplet attention module [C] // Proceedings of the 2021 IEEE Winter Conference on Applications of Computer Vision (WACV). Piscataway, NJ: IEEE, 2020: 3138-3147.

[15] HUANG Hejun, CHEN Zuguo, ZOU Ying, et al. Channel prior convolutional attention formedical image segmentation [J]. Computers in Biology and Medicine, 2023, 178: 108784.

[16] SRINIVAS A, LIN T Y, PARMAR N, et al. Bottleneck transformers for visual recognition [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2021: 16514-16524.

[17] FU Jun, LIU Jing, TIAN Haijie, et al. Dual attention network for scene segmentation [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2019: 3141-3149.

[18] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 37(9): 1904-1916.

[19] REZATOFIGHI S H, TSOI N, GWAK J Y, et al. Generalized intersection over union: A metric and a loss for bounding box regression [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2019: 658-666.