邵小强, 李 鑫, 杨 涛, 等. 改进 YOLOv5s 和 DeepSORT 的井下人员检测及跟踪算法[J]. 煤炭科学技术, 2023, 51(10): 291−301.
SHAO Xiaoqiang, LI Xin, YANG Tao, *et al*. Underground personnel detection and tracking based on improved YOLOv5s and DeepSORT[J]. Coal Science and Technology, 2023, 51(10): 291−301.

移动扫码阅读

改进 YOLOv5s 和 DeepSORT 的井下人员检测及跟踪算法

邵小强,李 鑫,杨 涛,杨永德,刘士博,原泽文
(西安科技大学 电气与控制工程学院, 陕西 西安 710054)

摘 要: 矿井移动目标的实时监测及跟踪系统是建设智慧矿山必不可少的内容，井下巡检机器人的出现可以实现对作业人员的实时监测，但是井下光照不均、煤尘干扰等因素的存在导致传统图像检测算法无法准确检测出作业人员。基于此提出一种可部署于井下巡检机器人的改进 YOLOv5s 和 DeepSORT 的井下人员检测及跟踪算法。首先利用监控摄像头与巡检机器人所录视频制作数据集，然后使用改进 YOLOv5s 网络对井下人员进行识别：考虑到井下人员检测及跟踪算法包含复杂的网络结构和庞大的参数体量，限制了检测模型的响应速度，使用改进轻量化网络 ShuffleNetV2 替代原 YOLOv5s 主干网络 CSP-Darknet53。同时，为减少图像中复杂背景的干扰，提升作业人员的关注度，将 Transformer 自注意力模块融入改进 ShuffleNetV2。其次，为了使多尺度特征能够有效融合且使得推理信息能够有效传输，将 Neck 中 FPN+PAN 结构替换为 BiFPN 结构。接着利用改进 DeepSORT 对人员进行编码追踪：考虑到井下环境黑暗，照度低，无纹理性，DeepSORT 难以有效提取到人员的外观信息，于是采用更深层卷积替换 DeepSORT 中小型残差网络来强化 DeepSORT 的外观信息提取能力。最后通过公开行人数据集及自建井下人员检测及跟踪数据集对本文改进算法进行验证，结果表明：改进的检测模型相比于原 YOLOv5s 模型平均检测精度提高了 5.2%，参数量减少了 41%，速度提升了 21%；改进 YOLOv5s-DeepSORT 的井下人员跟踪方法精度达到了 89.17%，速度达到了 67FPS，可以有效部署于井下巡检机器人实现作业人员的实时检测及跟踪。

关键词: 井下巡检机器人; YOLOv5s; 轻量化; DeepSORT; 实时检测及跟踪
中图分类号: TD76 **文献标志码:** A **文章编号:** 0253−2336(2023)10−0291−11

Underground personnel detection and tracking based on improved YOLOv5s and DeepSORT

SHAO Xiaoqiang, LI Xin, YANG Tao, YANG Yongde, LIU Shibo, YUAN Zewen
(*College of Electrical and Control Engineering, Xi'an University of Science and Technology, Xi'an* 710054, *China*)

Abstract: The real-time monitoring and tracking system of mine moving targets is an essential part of the construction of smart mines. The appearance of downhole inspection robots can realize the real-time monitoring of operators, but the existence of uneven lighting, coal dust interference and other factors lead to the traditional image detection algorithm can not accurately detect operators. Based on this, this paper proposes an improved YOLOv5s and DeepSORT algorithm for downhole personnel detection and tracking that can be deployed in downhole inspection robots. Firstly, the data set was made by using the video recorded by the surveillance camera and inspection robot, and then the improved YOLOv5s network was used to identify the underground personnel: Considering that the detection and tracking algorithm for downhole personnel contains complex network structure and huge parameter volume, which limits the response speed of the detection model, this paper uses an improved lightweight network ShuffleNetV2 to replace the original YOLOv5s backbone network CSP-

收稿日期: 2022−11−16 **责任编辑:** 周子博 **DOI:** 10.13199/j.cnki.cst.2022-1933
基金项目: 国家自然科学基金资助项目 (52174198)
作者简介: 邵小强(1976—), 男, 陕西商州人, 副教授, 博士。E-mail: shaoxq@xust.edu.cn
通讯作者: 李 鑫(1998—), 男, 山西太原人, 硕士研究生。E-mail: 1187751601@qq.com

Darknet53. Meanwhile, in order to reduce the interference of complex image background and improve the attention of operators, Transformer self-attention module is integrated into the ShuffleNetV2. Secondly, the FPN+PAN structure in Neck is replaced by BiFPN structure in order to effectively fuse multi-scale features and effectively transmit inference information. Then, improved DeepSORT was used to encode and track personnel: considering that the underground environment was dark, with low illumination and no texture, it was difficult for DeepSORT to effectively extract personnel's appearance information, so DeepSORT's small and medium residual network was replaced by deeper convolution to enhance DeepSORT's appearance information extraction ability. Finally, the improved algorithm is verified by open pedestrian data set and self-built underground personnel detection and tracking data set. The results show that compared with the original YOLOv5s model, the average detection accuracy of the improved detection model is increased by 5.2%, the number of parameters is reduced by 41%, and the speed is increased by 21%. The improved YOLOv5s-DeepSORT downhole personnel tracking method has a precision of 89.17% and a speed of 67FPS, which can be effectively deployed in downhole inspection robots to realize real-time detection and tracking of operators.

Key words: downhole inspection robot; YOLOv5s; Lightweight; DeepSORT; Real-time detection and tracking

0 引 言

为了扎实推进智慧矿山的建设,提升企业整体的信息化、数字化水平,对井下监控系统与巡检机器人的检测及跟踪能力进行全面升级是十分必要的。国家煤矿安监局最新出台的《煤矿井下单班作业人数限员规定》将矿井按生产能力分为 7 档,对于各档次矿井下单班作业人数及采掘工作面作业人数做出限制。于是对井下人员进行实时跟踪及统计是避免发生安全事故的有效手段。但井下工作环境存在着光照不均,煤尘干扰严重等问题,导致工作人员无法长时间有效对监控视频进行多场景监控[1],且定点监控覆盖面有限。因此,使用巡检机器人取代工作人员进行实时监控对于减轻职工工作强度,降低岗位安全风险,实现企业减人增效和建设智慧矿山有着积极的作用[2]。

当今目标检测算法分为 2 大类:传统机器学习与深度神经网络。传统机器学习算法分为三部分:滑动窗口、特征提取、分类器[3]。此类算法针对性低、时间复杂度高、存在窗口冗余[4];并且手工设计的特征鲁棒性差、泛化能力弱[5],这导致传统机器学习算法逐渐被深度学习算法所取代[6]。李若熙等[7] 通过 YOLOv4[8] 算法进行井下人员检测,在寻找目标中心点时引入聚类分析算法,提升了模型的特征提取能力。杨世超[9] 通过 Faster-RCNN[10] 算法进行井下人员检测,将井下监控采集的图像输入到检测模型中提取特征,利用区域建议网络和感兴趣区域池化得到目标的特征图,最后通过全连接层得到目标的精确位置。董昕宇等[11] 通过 SSD[12] 算法构建了一种井下人员检测模型,采用深度可分离卷积模块和倒置残差模块构建轻量化模型,提升了模型的检测速度。陈伟等[13] 提出一种基于注意力机制的无监督矿井人员跟踪算法,结合相关滤波和孪生网络在跟踪任务的优势,构建轻量化目标跟踪模型。以上文献都是利用深度学习算法实现井下人员检测与跟踪,但是当出现目标遮挡时,检测效果均不佳;同时缺少对井下人员编码统计的能力;而且模型参数量较大,检测速率也有待提高[14]。

针对上述问题,基于 YOLOv5s[15] 和 DeepSORT[16] 模型进行改进,使用改进轻量化网络ShuffleNetV2[17] 替代 YOLOv5s 主干网络 CSP-Darknet53[18],使得模型在保持精度的同时降低了计算量。同时在改进 ShuffleNetV2 中添加 Transformer[19] 自注意力模块来强化模型深浅特征的全局提取能力。接着使用 BiFPN[20] 结构替换原 Neck 结构,使多尺度特征能够有效融合。最后使用更深层卷积强化 DeepSORT 的外观信息提取能力,有效的提取图像的全局特征和深层信息,减少了目标编码切换的次数。实验结果表明,改进后的模型有效解决了人员遮挡时检测效果不佳及编码频繁切换的问题。

1 YOLOv5s 模型

YOLOv5 是当前深度学习主流的 One-Stage 结构目标检测网络,共有 4 个版本:YOLOv5s、YOLOv5m、YOLOv5l、YOLOv5x。考虑到井下巡检机器人的轻量化设计,本文采用深度最小,特征图宽度最小的网络 Yolov5s。其分为输入端 Input、主干网络 Backbone、颈部网络 Neck、输出端 Head 四部分。输入端通过 Mosaic 数据增强、自适应锚框计算、自适应图片缩放,使得模型适用于各种尺寸大小图像的输入的同时丰富了数据集,提升了网络的泛化能力。主干网络包含:焦点层 (Focus), Focus 结构在之前的 YOLO 系列[21-23,8] 中没有引入,它直接对输入的图像进行切片操作,使得图片下采样操作时,在不发生信息丢失的情况下,让特征提取更充分[24];跨

阶段局部网络层 (Cross Stage Partial Network, CSP)，CSP[25] 结构是为了解决推理过程中计算量过大的问题；空间金字塔池化 (Spatial Pyramid Pooling, SPP)，SPP[26] 结构能将任意大小的特征图转换成固定大小的特征向量。Neck 中采用的是 FPN+PAN 结构，负责对特征进行多尺度融合。Head 输出端负责最终的预测输出，使用 GIOU 损失函数作为位置回归损失函数，交叉熵损失函数作为类别损失函数，其作用是在不同尺度的特征图上预测不同大小的目标。

2　改进 YOLOv5s 井下人员检测算法

提出的井下人员检测框架如图 1 所示。首先将井下巡检机器人所采集的图像逐帧输入到改进 YOLOv5s 中进行训练，从而获取到网络的训练权重，最后利用测试集图像对本文改进的目标检测算法进行验证。

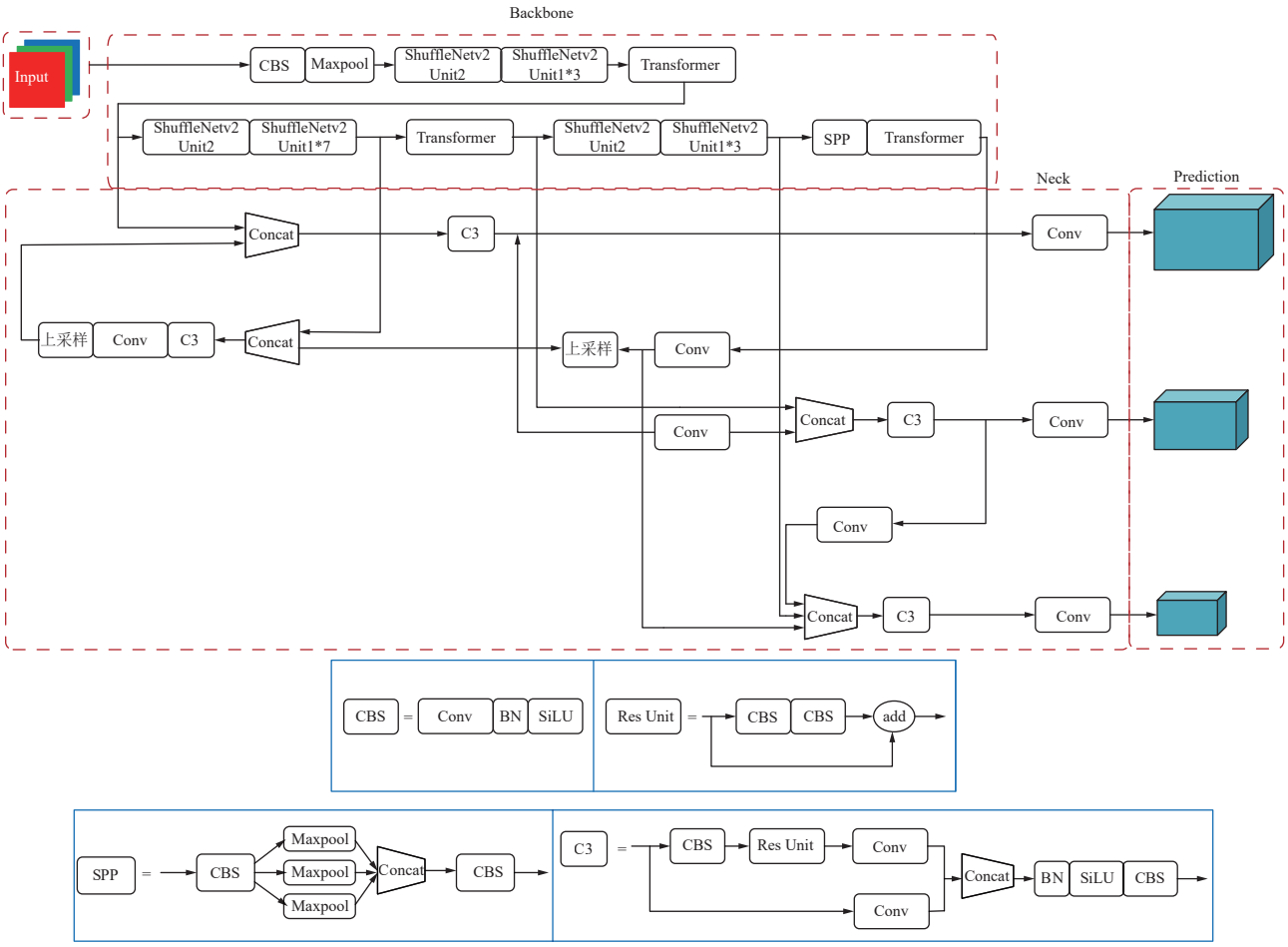


图 1　本文目标检测算法框架

Fig.1　Detection framework of the proposed algorithm

2.1　主干网络的替换

由于 YOLOv5s 具有较大的参数量，对于硬件成本要求较高，难以部署在小型的嵌入式设备或者移动端设备。因此使用轻量化网络 ShuffleNetV2 代替原主干网络 CSP-Darknet53，通过深度可分离卷积来代替传统卷积减小参数量的同时高效利用了特征通道与网络容量，使得网络仍保持较高的精度[27]。表 1 展示了改进 ShuffleNetV2 结构，本文将原结构中最大池化卷积层采用深度可分离卷积进行替换，实现了通道和区域的分离，增强了网络的特征提取能力同时也降低了参数量；使用全局池化层替换原结构中的全连接层进行特征融合，保留了前面卷积层提取到的空间信息，提升了网络的泛化能力。

2.2　Transformer 自注意力模块的融入

Transformer 整个网络结构由自注意力模块和前馈神经网络组成。Transformer 采用自注意力机制，将序列中的任意两个位置之间的距离缩小为一个定值，具有更好的并行性，符合现有的 GPU 框架[28]。本文在改进 ShuffleNetV2 中引入 Transformer 自注意力模块，与原始网络相比，添加 Transformer 模块

可以提取到更加丰富的图像全局信息与潜在的特征信息，提升了模型的泛化能力。

本文融入的 Transformer 块结构图如图 2 所示，其主要由以下 3 部分构成。

表 1　改进 ShuffleNetV2 结构

Table 1　Improve the structural ShuffleNetV2

层数	输出大小	核大小	步长	重复使用次数	通道数
Image	224×224	—	—	—	3
Conv1	112×112	3×3	2	1	24
DW conv	56×56	3×3	2	1	24
Stage2	28×28	—	2	1	116
Stage2	28×28	—	1	3	116
Stage3	14×14	—	2	1	232
Stage3	14×14	—	1	7	232
Stage4	7×7	—	2	1	464
Stage4	7×7	—	1	3	464
Conv5	7×7	1×1	1	1	1 024
Global pooling	1×1	7×7	—	—	—

高效自注意力层 (Efficient Self-Attention) 可以通过图像形状重塑，缩短远距离特征依赖间距，使网络更加全面地捕获图像特征信息[29]。自注意力公式如式 (1) 所示。

$$\mathrm{Attention}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=\mathrm{soft\,max}\left(\frac{QK^{\mathrm{T}}}{\sqrt{d_{\mathrm{k}}}}\right)V \tag{1}$$

式中，$(W_{\mathrm{q}},W_{\mathrm{k}},W_{\mathrm{v}})$为权重矩阵，负责将 $\boldsymbol{X}$ 映射为语义更深的特征向量 $\boldsymbol{Q}$, $\boldsymbol{K}$, $\boldsymbol{V}$，而 d_{k} 为特征向量长度。

高效自注意力层通过位置编码来确定图像的上下文信息，输出图像的分辨率是固定的，当测试集图像与训练集图像的分辨率不同时，会采用插值处理来保证图像尺度一致，但是这样会影响模型的准确率[30]。针对此问题，本文在高效自注意力层后连接混合前馈网络 (Mix Feedforward Network, Mix-FFN) 来弥补插值处理对泄露位置信息的影响。混合前馈网络计算公式如式 (2)、式 (3) 所示：

$$x_{\mathrm{out1}}=\mathrm{Conv}(MLP(\mathrm{Norm}(x_{\mathrm{in}}))) \tag{2}$$

$$x_{\mathrm{out}}=\mathrm{MLP}(\mathrm{GELU}(x_{\mathrm{out1}}))+x_{\mathrm{in}} \tag{3}$$

式中，x_{in}为上层输出；Norm 为归一化处理；MLP 为多层感知机；GELU 代表激活函数。

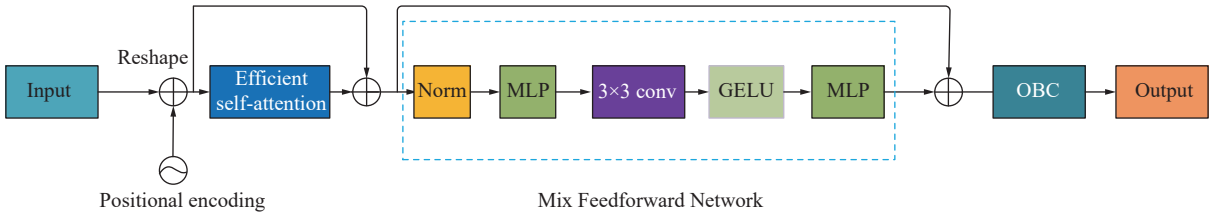


图 2　Transformer 块结构

Fig.2　Transformer block structure

重叠块压缩层 (Overlapping Block Compression, OBC) 用于压缩图像尺寸和改变图像通道数，保留尺度稳定的特征，简化模型复杂度和降低冗余信息。

2.3　多尺度特征融合网络

原始 YOLOv5s 的 Neck 部分采用的是 FPN+PAN 结构，FPN 是自顶向下，将高层的强语义特征向底层传递，增强了整个金字塔的语义信息，但是对定位信息没有传递。PAN 就是在 FPN 的后面添加一个自底向上的路径，对 FPN 进行补充，将底层的强定位信息传递上去。但是该结构的融合方式是将所有的结构图转换为相同大小后进行级联，没有将不同尺度之间的特征充分利用，使得最终的目标检测精度未达到最优。因此，本文采用一种更为高效的 Bi-FPN 特征融合结构进行替代。其结构如图 3 所示，相较于原始特征融合结构，BiFPN 能更有效的结合位于低层的定位信息与高层的语义信息，同时在通道叠加时将权重信息考虑进去，实现双向多尺度特征融合，通过不断调参确定不同分辨率的特征重要性，如式 (4) 所示。

$$\mathrm{Out}=\sum_i\frac{\omega_i}{\varepsilon+\sum_j\omega_j}\ln_i \tag{4}$$

式中，i 为第 i 个权重；j 为权重总个数；ln 为输入特征；Out 为输出特征；ω_{i}为权重。

将主干网络中 Transformer 模块提取出大小不同的特征图通过 BiFPN 进行融合，可以更加有效地融合全局深浅层的信息与关键的局部信息，将第一次下采样得到的特征图与后面的特征图进行跨层连接，使得定位信息能够获取充分，提升了模型小目标的检测性能；在特征融合时删除对模型贡献较低的

节点，在同尺度特征节点间增加跳跃连接，减少了计算量；最终在提高模型精度及泛化能力的同时降低了漏检率且几乎不增加运行成本。

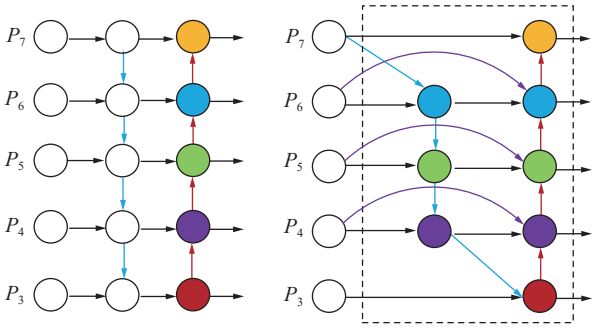


图 3　BiFPN 结构

Fig.3　BiFPN structure

3　DeepSORT 多目标跟踪算法及改进

使用本文提出的检测模型与改进 DeepSORT 跟踪算法搭配实现对井下人员的跟踪，首先将监测图像输入到改进 Yolov5s 目标检测网络，得到检测结果，然后通过改进 DeepSORT 算法逐帧对人员进行匹配，得到他们的轨迹信息，最后输出跟踪图像。

3.1　DeepSORT 算法

DeepSORT 是针对多目标的跟踪算法，其核心是利用卡尔曼滤波和匈牙利匹配算法，将跟踪结果和检测结果之间的 IOU (Intersection over Union, 交并比) 作为代价矩阵，实现对移动目标的跟踪。

为了跟踪检测模型找出的作业人员，DeepSORT 使用 8 维变量 x 来描述作业人员的外观信息和在图像中的运动信息，如式 (5) 所示。

$$x=(u,v,\gamma,q,\dot{u},\dot{v},\dot{\gamma},\dot{q}) \tag{5}$$

式中：(u,v)为井下人员的中心坐标；γ为人员检测框的宽高比；q为人员检测框的高；$(\dot{u},\dot{v},\dot{\gamma},\dot{q})$为$(u,v,\gamma,q)$相应的速度信息。

DeepSORT 结合井下人员的运动信息与外观信息，使用匈牙利算法对预测框和跟踪框进行匹配，对于人员的运动信息，采用马氏距离描述卡尔曼滤波的预测结果和改进 YOLOv5s 检测结果之间的关联程度，如式 (6) 所示。

$$d^{(1)}(i,j)=(d_j-\boldsymbol{y}_i)^{\mathrm{T}}\boldsymbol{S}_i^{-1}(d_j-\boldsymbol{y}_i) \tag{6}$$

式中：d_j为第 j 个检测框；$\boldsymbol{y}_j$为第 i 个检测框的状态向量；$\boldsymbol{S}_i$为 i 条轨迹之间的标准差矩阵。

当井下行人被障碍物长时间遮挡时，外观模型就会发挥作用，此时特征提取网络会对每个检测框计算出一个 128 维特征向量，限制条件为$\|r_j\|=1$，同时对检测到的每个人员构建一个确定轨迹的 100 帧外观特征向量。通过式 (7) 计算出这两者间的最小余弦距离。

$$d^{(2)}=(i,j)=\min\left\{1-\boldsymbol{r}_j^{\mathrm{T}}\boldsymbol{r}_k^{(i)}\,\middle|\,\boldsymbol{r}_k^{(i)}\in R_k\right\} \tag{7}$$

式中：$\boldsymbol{r}_j$为检测框对应的特征向量；$\boldsymbol{r}_k$为 100 帧已成功关联的特征向量。

马氏距离在短时预测时提供可靠的目标位置信息，使用外观特征的最小余弦距离可使得遮挡目标重新出现后恢复目标 ID，为了使两种度量的优势互补，最终将两种距离进行线性加权作为最终度量，公式如式 (8) 所示。

$$c_{i,j}=\lambda d^{(1)}(i,j)+(1-\lambda)d^{(2)}(i,j) \tag{8}$$

式中：λ为权重系数，若$c_{i,j}$落在指定阈值范围内，则认定实现正确关联。

3.2　DeepSORT 算法的改进

原始 DeepSORT 的外观特征提取采用一个小型的堆叠残差块完成，包含两个卷积层和六个残差网络。该模型在大规模路面行人检测数据集上训练后，可以取得很好的效果，但是井下环境光照不均匀，烟尘干扰严重，导致对井下人员跟踪的效果不理想，于是本文采用高效特征提取架构 OSA(one shot aggregation) 来替代原 DeepSORT 外观模型中的堆叠残差块以强化 DeepSORT 的外观特征提取能力，有效的提取图像中的全局特征和深层信息，达到减少人员编码切换次数的作用，OSA 结构如图 4 所示。

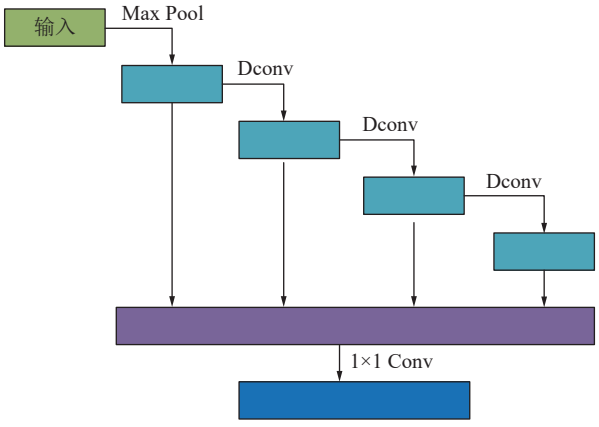


图 4　OSA 结构

Fig.4　OSA structure

在外观状态更新时，采用指数平均移动的方式替代特征集合对第 t 帧的第 i 个运动轨迹的外观状态进行更新。如式 (9) 所示。

$$e_i^t=\alpha e_i^{t-1}+(1-\alpha)f_i^t \tag{9}$$

式中：f_i^t为第 t 帧的第 i 个运动轨迹的外观嵌入；α为动量项。使用这种方式不仅减少了时间的消耗，同时提高了匹配的质量。

4　试验与分析

4.1　试验准备

本文采用 Caltech 行人数据集 (Caltech Pedestrian Detection Benchmark)、INRIA 行人数据集 (INRIA Person Dataset) 及自建井下人员检测及跟踪数据集对所提检测及跟踪算法井下进行验证。

1) Caltech 行人数据集:此数据集为目前规模较大的行人数据集，使用车载摄像头录制不同天气状况下 10 h 街景，拥有人员遮挡、目标尺度变化大、背景复杂等多种情形，标注超过 25 万帧，35 万个矩形框，2 300 个行人。同时注明了不同矩形框之间的时间关系及人员遮挡情况。

2) INRIA 行人数据集:此数据集为目前常见的静态人员检测数据集，数据集中人员身处不同光线条件及地点。训练集拥有正样本 1 000 张，负样本 1 500 张，包含 3 000 个行人；测试集包含正样本 350 张，负样本 500 张，包含 1 200 个行人，该数据集人员以站姿为主且高度均超 100 个像素，图片主要来源于谷歌，故清晰度较高。

3) 自建井下人员检测及跟踪数据集：采集井下巡检机器人与监控视频拍摄的 10 万帧图像，筛选其中 8 000 帧相似程度较低的图像构建数据集。首先使用 ffmpeg 工具将图像按帧切为图片，其中涵盖井下各种环境：光照不均 2 267 张、煤尘严重 1 568 张、目标遮挡 3 891 张、其余环境 1 200 张。其次采用 Python 编写的 Labelimg 对图片中人员进行标注，自动将人员位置及尺寸生成 xml 文件，最终转为适用于 yolo 系列的 txt 文件，包含每张图片中人员的中心位置 (x, y)、高 (h)、宽 (w) 三项信息。如图 5 所示，该数据集包含上万个人工标记的检测框。由于本文算法应用于井下人员的检测及追踪，故数据集中仅含“person”一个类。将图片数量按照 7∶2∶1 分为训练集、验证集和测试集。

试验使用平台参数如下：

配置	参数
操作系统	Windows 10
内存容量	32 GB
GPU	NVIDIA GeForce RTX 3070Ti
CPU	Intel 酷睿i7 12700H
模型框架	PyTorch1.7.1
编程语言	Python3.6

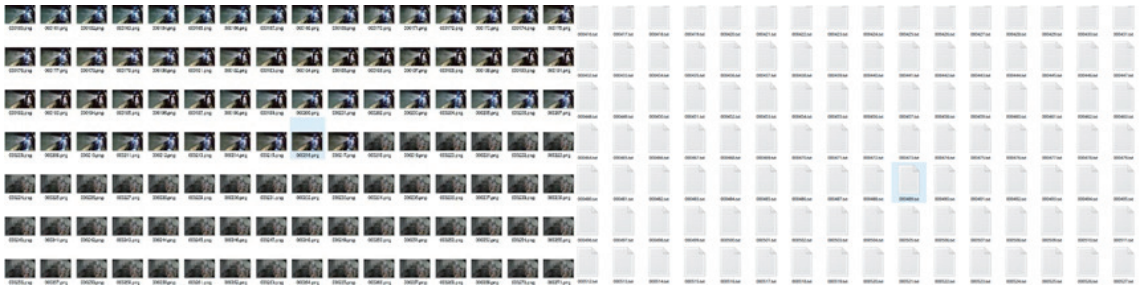


图 5　自建井下人员检测及跟踪数据集

Fig.5　Self-built downhole personnel detection and tracking data sets

检测算法评价指标：使用模型参数量、检测时间、召回率M_r、准确率M_p、漏检率M_m，误检率M_f及 mAP@0.5 作为检测算法的评价指标。

$$M_r = \frac{T_p}{T_p + F_N} \tag{10}$$

$$M_p = \frac{T_p}{T_p + F_p} \tag{11}$$

$$M_m = \frac{F_N}{F_N + T_p} \tag{12}$$

$$M_f = \frac{F_p}{F_p + T_N} \tag{13}$$

$$\mathrm{mAP} = \frac{T_p + T_n}{T_p + T_n + F_p} \tag{14}$$

式中：T_p为被正确检测出的井下人员；F_N为未被检测到的井下人员；F_P为被误检的井下人员；T_N为未被误检的井下人员；mAP 为不同召回率上正确率的平均值。

跟踪算法评价指标：

1) 编码变换次数 (ID switch, IDS)，跟踪过程中人员编号变换及丢失的次数，数值越小说明跟踪效果越好。

2) 多目标跟踪准确率 (Multiple Object Tracking Accuracy)，用于确定目标数及跟踪过程中误差累计

情况，如式 (15) 所示。

$$A_{\mathrm{MOT}}=1-\frac{\sum_{1}^{n}tM_{\mathrm{m}}+M_{\mathrm{f}}+\mathrm{IDS}}{\sum_{1}^{n}tGT_{t}} \tag{15}$$

式中:M_{m}为漏检率；M_{f}为误检率；IDS为编码转换次数；GT_{t}为目标数量；n 为图片数量；t 为第 t 张图片。

3) 多目标跟踪精度 (Multiple Object Tracking Precision, P_{MOT})，用于衡量目标位置的精确程度，如式 (16) 所示。

$$P_{\mathrm{MOT}}=\frac{\sum_{1}^{n}t,id_{t,i}}{\sum_{1}^{n}tc_{t}} \tag{16}$$

式中：$d_{t,i}$为目标 i 与标注框间的平均度量距离；c_t为 t 帧匹配成功的数目。

4) 每秒检测帧数 (Frames Per Second, FPS) 及模型参数量，体现模型运行的速率及成本。

4.2　目标检测试验结果与分析

将本文算法通过自建井下人员检测及跟踪数据集进行训练，输入图像大小为 608×608，迭代次数为 300，批次大小为 16，初始学习率设置为 0.01，后 150 轮的训练学习率降为 0.001。动量设置为 0.937，衰减系数为 0.005。训练损失变化如图 6 所示。可以看出模型三类损失函数收敛较快且都收敛于较低值，表明改进算法具有良好的收敛能力与鲁棒性。

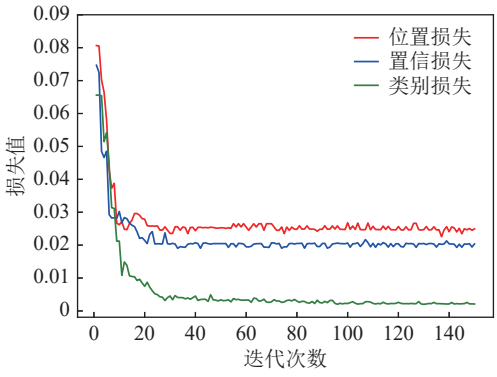


图 6　损失函数曲线

Fig.6　Loss function curve

为了验证本文改进检测算法的有效性以及轻量化主干网络选择的合理性，将本文算法与 YOLOv5s 模型和 YOLOv5s-ShuffleNetV2 通过自建井下人员检测及跟踪数据集进行对比。

从图 7 中可以看出，原始 YOLOv5s 算法迭代到 40 次时，准确率上升到 0.86 左右，最终收敛在 0.87 左右；YOLOv5s-ShuffleNetV2 在迭代到 40 次时，准确率上升到 0.84 左右，最终收敛在 0.85 左右；而本文所提算法在迭代 40 次时，准确率上升到 0.91 左右，最终收敛在 0.92 左右，较原始 YOLOv5s 模型提升了 5.1%。

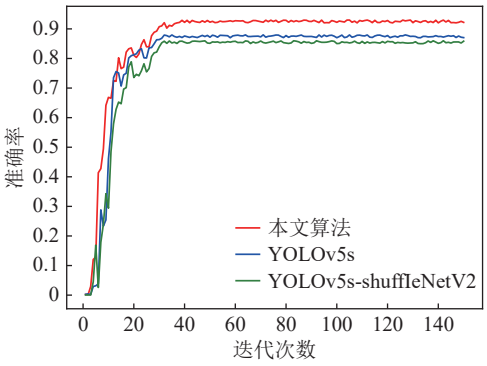


图 7　准确率曲线

Fig.7　Accuracy rate curve

从图 8 中可以看出，原始 YOLOv5s 算法在迭代到 40 次时，*mAP* 上升到 0.85 左右，最终收敛在 0.86 左右；YOLOv5s-ShuffleNetV2 在迭代到 40 次时，*mAP* 上升到 0.85 左右，最终收敛在 0.85 左右；而本文算法的迭代到 40 次时，*mAP* 上升到 0.89 左右，*mAP* 最终收敛在 0.90 左右，较原始 YOLOv5s 模型提升了 5.2%。综上所述，本文选取的轻量化网络 ShuffleNetV2 可以使得检测模型保持一定精度的同时降低计算量；轻量化主干的改进、注意力机制的引入以及多尺度的融合对于目标检测性能有着明显的提升，因此，本文检测算法对于井下复杂环境中的人员检测具有良好的精度。

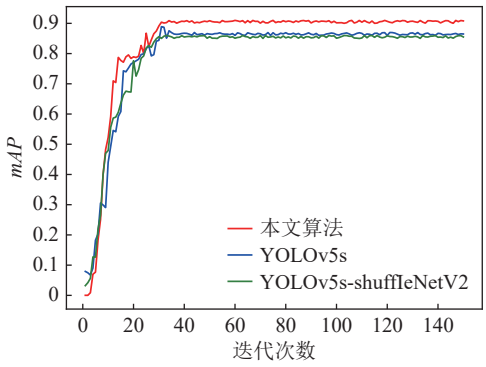


图 8　*mAP* 曲线

Fig.8　*mAP* curve

在 YOLOv5s 算法的基础上进行了改进轻量化主干网络的替换 ShuffleNetv2、Transformer 自注意力机制模块的融入、Neck 部分进行 BiFPN 的替换。为了检验本文对检测阶段各改进点的有效性，以 YOLOv5s 模型为基准，使用消融实验在相同环境下

进行进行验证，各模型参数设置保持一致，具体消融试验结果见表 2。

由表 2 可以看出，原始 YOLOv5s 的主干网络替换后，准确率下降了 1.4%，速率提升了 34%。在模型 2 中添加 Transformer 自注意力模块后，准确率提升了 2.8%。在模型 2 中使用 BiFPN 替代原来的特征融合结构后，准确率提升了 2.1%。在模型 2 中同时添加 Transformer 自注意力机制模块和 BiFPN 模块，准确率提升了 7.4%，平均漏检率下降了 40%，召回率提升了 8.4%，平均误检率下降了 51%。综上所述，单独添加 Transformer 自注意力模块和 BiFPN 模块，井下人员检测性能提升有限，而两种模块组合添加时，井下人员检测性能获得了很大的提升。相比于原始算法，准确率提升了 5.2%；参数量下降了 41%；检测速率提升了 21%，达到 0.014 8 s/帧；为部署于巡检机器人奠定了基础。

表 2　消融试验结果

Table 2　Ablation results

模型	ShuffleNetv2	Transformer	BiFPN	准确率	漏检率	召回率	误检率	时间/ms	参数量/MB
1	—	—	—	0.871	0.314	0.783	0.027	18.9	13.09
2	√	—	—	0.859	0.322	0.794	0.030	12.4	3.45
3	√	√	—	0.883	0.235	0.831	0.021	13.3	4.17
4	√	—	√	0.877	0.249	0.831	0.019	13.7	4.34
5	√	√	√	0.923	0.190	0.861	0.013	14.9	5.33

注：“√”表示对应部分已改进。

为了验证文中检测算法具有良好的泛化能力，在 2 个公开行人数据集 Caltech 行人数据集、INRIA 行人数据集上进行进一步验证，性能指标对比见表 3。通过比较 3 个不同数据集中的性能指标，可以看出文中算法不仅适用于井下人员检测，在目标尺度变化大、背景复杂、光照剧烈等多数场景中人员检测效果也均优于原始 YOLOv5s，因此，具有良好的泛化性与鲁棒性。

为了更加直观地体现文中检测算法的效果，选择 Faster-RCNN、YOLOv3、YOLOv4、YOLOv5s 4 种主流算法在自建数据集中选取光照不均、煤尘干扰、多目标移动、人员遮挡 4 种场景进行验证，检测结果如图 9 所示。

从第一组试验中，可以观察到光照不均严重，Faster-RCNN、YOLOv3、YOLOv4、YOLOv5s 均出现误检的情况，而本文算法使用了 BiFPN 结构使得多尺度特征能够有效融合，对于远处小目标检测能够起到了很好的识别作用。从第二组试验中，可以观察到粉尘干扰严重，除文中算法外，其余算法出现漏检、误检的情况，而文中算法由于融合了 Transformer 自注意力模块强化了模型深浅特征的全局提取能力，提升了目标在复杂环境中的对比度，有效抑制了粉尘的干扰。从第三、四组试验得出，本文算法

表 3　多数据集性能指标对比

Table 3　Comparison of performance indicators of multiple data sets

数据集	性能指标	YOLOv5s	本文算法
Caltech行人数据集	精确率	0.781	0.849
	召回率	0.691	0.733
	mAP	0.742	0.792
INRIA行人数据集	精确率	0.861	0.881
	召回率	0.788	0.791
	mAP	0.856	0.890
自建数据集	精确率	0.871	0.923
	召回率	0.783	0.861
	mAP	0.864	0.902

(a) 原图

(b) Faster-RCNN

(c) YOLOv3

(d) YOLOv4

(e) YOLOv5s

(f) 本文算法

图 9　主流目标检测结果对比

Fig.9　Comparison of detection results of mainstream targets

对于井下环境中多目标移动对象及遮挡人员的检测也具有良好的效果。综上所述，文中检测算法在井下各种复杂环境中检测效果良好，与主流目标检测算法相比更适用于井下人员的检测。

4.3　井下人员跟踪结果与精度分析

为了验证文中算法在井下人员多目标跟踪方面的表现，本文通过自建井下人员检测及跟踪数据集上进行验证，以 YOLOv5s-DeepSort 为基准，使用原算法的参数设置，对检测与跟踪阶段进行消融试验来验证文中两阶段改进各自的有效性，结果见表 4。

表 4　多目标跟踪结果对比

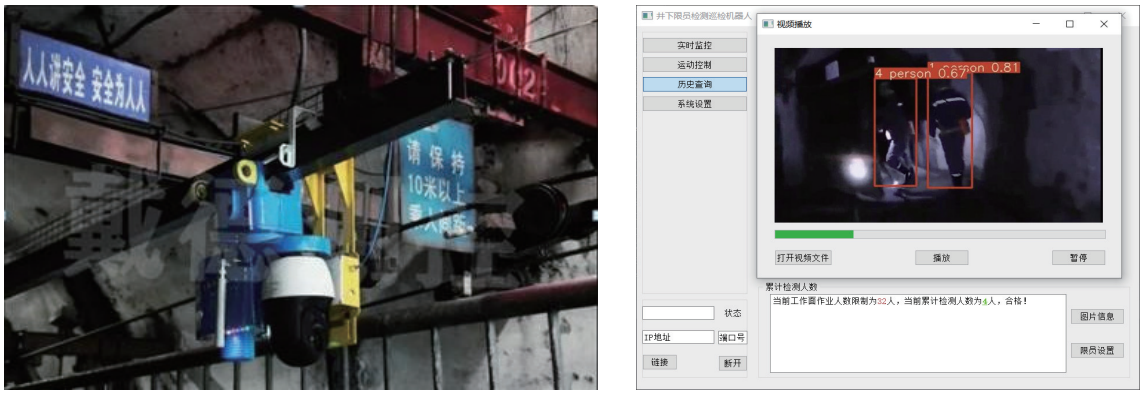
Table 4　Comparison of multi-target tracking results

算法	A_{MOT}/%	P_{MOT}/%	IDS	FPS	参数量/MB
YOLOv5s-DeepSORT	83.32	81.55	16	41	25.6
改YOLOv5s-DeepSORT	87.47	86.32	13	71	11.19
YOLOv5s-改DeepSORT	82.31	82.44	7	39	19.34
本文算法	89.17	87.91	4	67	5.91

由表 4 得出，文中目标检测阶段的改进在有效提升井下人员的检测精度的同时提升了检测速度，而跟踪阶段的改进有效减少了人员编号的转换，可以在出现人员遮挡的情况下有效提升检测的精度。文中检测及跟踪算法最终达到 89.17% 的精度；速率达到 67 帧；人员编码改变次数仅 4 次，目标编号改变次数降低了 66.7%；参数量缩减到原始跟踪算法的 23%。可以很好的满足井下人员实时检测及跟踪的需求。

为了更加直观展示文中跟踪算法的效果，文中选用戴德 KJXX12C 型防爆矿用巡检机器人进行验证，如图 10a 所示，该装置搭载本安型“双光谱”摄像仪，最小照度达彩色 0.002 lux，高粉尘环境下，可通过红外摄像仪辅助采集井下图像。采集与控制系统采用 STM32ZET6 芯片，上位机检测及跟踪主控系统采用 Windows 版工控机。图像信息会通过千兆无线通讯传输在远端上位机，将环境运行代码安装于上位机。图像信息经过本文算法处理，结果将存储并实时显示于主控界面，如图 10b 所示，主控界面采用 CS 架构，由 C#语言编写。监测人员通过主控界面实时及历史数据对工作面作业人数是否合格进行判断。

从图 10c，图 10d，图 10e 中可以观察到，在井下光照不足的环境中，井下 2 个作业人员相互遮挡并且持续行走一段距离后，巡检机器人能够进行稳定的检测跟踪并且其编号没有发生改变，实现有效计



(a) 巡检机器人井下安装效果　　(b) 主控界面监测

(c) 第 7 帧效果　　(d) 第 67 帧效果　　(e) 第 157 帧效果

图 10　巡检机器人多目标跟踪效果

Fig.10　Inspection robot multi-target tracking effect

数，也能够证明我们改进的算法在复杂环境中出现井下人员遮挡时，也会在后续帧中匹配到被遮挡人员，对于遮挡情况具有良好的鲁棒性。

5　结　　论

1) 提出了一种改进 YOLOv5s 和 DeepSORT 的井下人员检测及跟踪算法。在 YOLOv5s 模型的基础上，使用轻量化网络 ShuffleNetV2 替换了原主干网络 CSP-Darknet53，减少了模型的参数量。同时融入 Transformer 自注意力模块，可以提取到更多潜在的特征信息。使用多尺度特征融合网络 BiFPN 替换原 Neck 结构，能更好的融合全局深浅层信息与关键的局部信息。跟踪阶段使用更深层卷积强化了 DeepSORT 的外观信息提取能力。

2) 利用自建井下人员检测及跟踪数据集对本文算法进行验证。结果表明，本文井下人员检测算法的准确率达到了 92%，检测速率达到 0.0148 s/帧。多目标跟踪算法准确率提高到了 89.17%，目标编号改变次数降低了 66.7%，并且拥有良好的实时性。

3) 构建的改进 YOLOv5s 和 DeepSORT 的井下人员检测与跟踪算法能够实现在井下复杂环境中对人员的实时检测及跟踪，其参数量也缩减到原来的 23%，不仅可以部署于煤矿监控系统，也可以部署在井下巡检机器人等小型嵌入式设备上，可以为井下人员的安全生产提供良好的保障。对于国家矿山安全监察局出台的《煤矿井下单班作业人数限员规定》早日实现智能化监测具有重要意义。

参考文献(References):

[1] 龚　云，颉昕宇. 基于同态滤波方法的煤矿井下图像增强技术研究[J]. 煤炭科学技术，2023，51(3)：241−250.
GONG Yun, XIE Xinyu. Research on coal mine underground image recognition technology based on homomorphic filtering method[J]. Coal Science and Technology, 2023, 51(3): 241−250.

[2] 厉　丹. 视频目标检测与跟踪算法及其在煤矿中应用的研究[D]. 徐州：中国矿业大学，2011.
LI Dan. Research on object detection and tracking algorithm and its application in coal mine[D]. Xuzhou: China University of Mining and Technology, 2011.

[3] 刘　丽，赵凌君，郭承玉，等. 图像纹理分类方法研究进展和展望[J]. 自动化学报，2018，44(4)：584−607.
LIU Li, ZHAO Lingjun, GUO Chengyu, *et al*. Texture classification: state-of-the-art methods and prospects[J]. Acta Automatica Sinica, 2018, 44(4): 584−607.

[4] ZHU H G. An efficient lane line detection method based on computer vision[J]. Journal of Physics:Conference Series, 2021, 1802(3): 032006.

[5] ABBAS Q, LI Y. Cricket video events recognition using HOG, LBP and Multi-class SVM[J]. Journal of Physics:Conference Series, 2021, 1732(1): 012036.

[6] 阮顺领，李少博，顾清华，等. 基于双向特征融合的露天矿区道路障碍检测[J]. 煤炭学报，2023，48(3)：1425−1438.
RUAN Shunling, LI Shaobo, GU Qinghua, *et al*. Road obstacle detection in open-pit mines based on bidirectional feature fusion[J]. Journal of China Coal Society, 2023, 48(3): 1425−1438.

[7] 李若熙，吕　潇，张元生，等. 改进YOLOv4算法井下人员检测的

研究[J]. 矿业研究与开发, 2021, 41(11): 179−185.
LI Ruoxi, LYU Xiao, ZHANG Yuansheng, *et al*. Reserch on Underground Personnel Detection Based on Improved YOLOv4 Algorithm[J]. Mining Research and Development, 2021, 41(11): 179−185.

[8] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimalspeed and accuracy of object detection [J/OL]. arXiv: 2004.109342020[cs. CV]. (2020-04-23).https://arxiv.org/abs /2004.10934.

[9] 杨世超. 基于Faster-Rcnn的矿井人员识别检测[J]. 信息记录材料, 2020, 21(12): 236−238.
YANG Shichao. Mine personnel identification and detection based on Faster-RCNN[J]. Information Recording Materials, 2020, 21(12): 236−238.

[10] REN S, HE K, GIRSHICK R, *et al*. Faster RCNN: Towards real-time object detection with region proposal networks[C]//Advances in neural information processing systems, 2015: 91−99.

[11] 董昕宇, 师　杰, 张国英. 基于参数轻量化的井下人体实时检测算法[J]. 工矿自动化, 2021, 47(6): 71−78.
DONG Xinyu, SHI Jie, ZHANG Guoying. Real-time detection algorithm of underground human body based on lightweight parameters[J]. Industry and Mine Automation, 2021, 47(6): 71−78.

[12] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]//European conference on computer vision. Springer, Cham, 2016: 21 – 37.

[13] 陈　伟, 任　鹏, 田子建, 等. 基于注意力机制的无监督矿井人员跟踪[J]. 煤炭学报, 2021, 46(S1): 601−608.
CHEN Wei, REN Peng, TIAN Zijian, *et al*. Unsupervised mine personnel tracking based on attention mechanism[J]. Journal of China Coal Society, 2021, 46(S1): 601−608.

[14] 任志玲, 朱彦存. 改进CenterNet算法的煤矿皮带运输异物识别研究[J]. 控制工程, 2023(4): 703−711.
REN Zhiling, ZHU Yancun. Research on foreign body recognition in coal mine belt transportation based on improved centernet algorithm[J]. Control engineering, 2023(4): 703−711.

[15] 葛俏, 梁桥康, 邹坤霖, 等. 基于轻量化网络与嵌入式系统的喷码检测[J]. 控制工程, 2022(12): 2349−2356.
GE Qiao, LIANG Qiaokang, ZOU Kunlin, *et al*. Inkjet quality detection method based on lightweight network and embedded[J]. Control engineering, 2022(12): 2349−2356.

[16] BEWLEY A, GE Z, OTT L, et al. Simple online and realtime tracking[C] //2016 IEEE International Conference on Image Processing(ICIP), 2016: 3464−3468.

[17] MA NN, ZHANG XY, ZHENG HT, *et al*. ShuffleNetV2: Practical guidelines for efficient CNN architecture design[C]//Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich: Springer, 2018. : 122−138.

[18] WANG C Y, LIAO H Y M, WU Y H, et al. CSPNet: A new backbone that can enhance learning capability of CNN[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Seattle: IEEE, 2020: 1571−1580.

[19] XIA Z, PAN X, SONG S, *et al*. Vision transformer with deformable attention[J]. ArXiv preprint arXiv: 2201.00520, 2022.

[20] TAN M, PANG R, LE Q V. EfficientDet: Scalable and efficient object detection[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2020: 10778−10787.

[21] REDMON J, DIVVALA S, GIRSHICK R, *et al*. You only look once: unified, real-time object detection[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016: 779−788.

[22] REDMON J, FARHADI A. YOLO9000: Better, Faster, Stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 6517−6525.

[23] REDMON J, FARHADI A. YOLOv3: an incremental improvement [EB/OL]. (2018-09-30).https://arxiv.org/abs/1804. 02767.

[24] 张麒麟, 林清平, 肖　蕾. 改进YOLOv5的航拍图像识别算法[J]. 长江信息通信, 2021, 34(3): 73−76.
ZHANG Qilin, LIN Qingping, XIAO Lei. Improved algorithm for aerial image recognition based on yolov5[J]. Changjiang information communication, 2021, 34(3): 73−76.

[25] WANG C Y, LIAO H Y M, WU Y H, *et al*. CSPNet: A new backbone that can enhance learning capability of CNN[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. New York: IEEE Press, 2020: 390–391.

[26] HE K M, ZHANG X Y, REN S Q, *et al*. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904−1916.

[27] ZHANG X, ZHOU X, LIN M, et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Salt Lake City: IEEE Press, 2018: 6848-6856.

[28] ZAMIR S W, ARORA A, KHAN S, *et al*. Restormer: efficient transformer for high-resolution image restoration[J]. ArXiv preprint arXiv: 2111.09881, 2022

[29] PETIT O, Thome N, RAMBOU C, *et al*. U-net transformer: Self and cross attention for medical image segmentation[C]//International Workshop on Machine Learning in Medical Imaging. Springer, Cham, 2021: 267-276.

[30] CHU X, TIAN Z, ZHANG B, *et al*. Conditional positional encodings for vision transformers[J]. arXiv preprint arXiv: 2102.10882, 2021.