光电技术及应用

DOI：10.16818/j.issn1001－5868.2022110701

基于S3-YOLOv5s的矿井人员防护设备检测算法研究

代少升*，曾　奇，黄　炼，陈昌川，陈怡羽，卢正鑫

（重庆邮电大学 通信与信息工程学院，重庆 400065）

摘　要： 针对复杂矿井环境下光照度低、目标尺度变化大、目标间遮挡严重，现有的目标检测网络特征提取困难、检测效果差等问题，提出了改进的S3-YOLOv5s的矿井人员防护设备检测算法。在主干网络中加入无参注意力模块（SimAM），提升网络的特征提取能力；引入尺度均衡特征金字塔卷积，加强多尺度特征融合；最后采用SIoU作为边框回归损失函数并使用K-means＋＋算法进行先验锚框聚类，提高边框检测精度。实验表明，相比现有的YOLOv5s算法，所提算法在所有类别的平均检测精确度从89.64％提升到了92.86％，在复杂矿井环境条件下对人员防护设备有优良的检测能力，验证了所提方法的有效性。

关键词： 矿井环境；防护设备；YOLOv5s；注意力机制；尺度均衡

中图分类号： TP391　**文章编号：** 1001－5868(2023)01－0153－08

Research on Detection Algorithm of Mine Personnel Protection Equipment Based on S3-YOLOv5s

DAI Shaosheng，ZENG Qi，HUANG Lian，CHEN Changchuan，CHEN Yiyu，LU Zhengxin

（School of Communication and Information Engin.，Chongqing University of Posts and Telecommun.，Chongqing 400065，CHN）

Abstract： Aiming at the problems of low illumination，large change of target scale，serious occlusion between targets，difficult feature extraction of existing target detection network，poor detection effect，etc. in complex mine environment，an improved S3-YOLOv5s mine personnel protection equipment detection algorithm is proposed. A simple，parameter free attention module（SimAM）was added to the backbone network to improve the feature extraction capability of the network. Scale equalizing pyramid convolution（SEPC）was introduced to strengthen multi-scale feature fusion. Finally，SIoU was used as the frame regression loss function and K-means＋＋ algorithm was used for prior anchor frame clustering to improve the frame detection accuracy. The experimental results show that，compared with the existing YOLOv5s algorithm，the average detection accuracy of the proposed algorithm in all categories is improved from 89.64％ to 92.86％，and the algorithm has excellent detection capability for personnel protection equipment under complex mine environments，which verifies the effectiveness of the proposed method.

Key words： mine environment；PPE；YOLOv5s；attention mechanism；scale equilibrium

0　引言

在煤矿作业生产中，安全帽、防护口罩、矿灯、自救器、专业工作服等是生产工作者不可或缺的个人防护设备（Personal Protective Equipment，PPE）[1]。然而，目前大多数矿井人员安全防护设备着装检查采取的仍是常规的人工监管方式，人工成本高，监督不足，数据难以可视化等问题依然存在。随着基于深度学习的目标识别方式广泛应用于校园管理、身份识别、智慧城市等领域，国内外学者也将深度学习的识别应用于施工现场作业人员的安全检

收稿日期：2022－11－07.

基金项目：重庆邮电大学校企合作项目（E2021269SD，E2022026SD）.

* **通信作者：**代少升　**E-mail：**daiss@cqupt.edu.cn

测[2]，在矿井环境中个人防护设备的检测方面具有极大的实用性和可行性。

Li 等[3]提出基于 Faster R-CNN 网络的变电站环境下安全帽检测方法，并比较了不同主干网络下的预测速度，该文献表明使用 AlexNet 作为主干网络时速度可达 200MS/FPS，但检测效果不佳，使用 Resnet101 作为主干网络时效果有所提升但计算速率下降较大。Fu 等[4]提出基于 YOLOv3 的矿井环境安全帽检测方法，准确率可以达到 84.4%，Wang 等[5]亦将 YOLOv3 用于自然场景下的安全帽及工作服检测，mAP 最高达 95%，与 Faster R-CNN 这种 two-stage 网络模型不同，YOLO 只有一个 one-stage 过程，独特的结构使 YOLOv3 比其他网络更快[6]。Zhuo 等[7]提出基于 SSD 网络模型的煤矿井下人员识别方法，该模型能有效检测煤矿井下人员。Li 等[8]提出基于热红外图像的人员检测 YOLOv4 网络，YOLOv4 将主干网络从之前的 Darknet53 更改为 CSPDarknet53，同时采用了一些数据增强技巧，包括 CutOut，MixUp 和 CutMix，性能相比 YOLOv3 有一定的提升，且红外图像在一定程度上避免了复杂矿井环境造成的干扰。

基于上述分析，受矿井低光照影响，作业环境狭小导致各目标分布集中且目标间尺寸各异，传统目标检测网络在矿井环境人员防护设备检测方面效果不佳。本文基于 YOLOv5s 设计了一种尺度均衡的特征金字塔卷积目标检测网络，首先在主干网络中加入无参注意力模块（A Simple，Parameter Free Attention Module，SimAM）[9]无参注意力，同时结合多尺度特征融合的思想，将 YOLOv5s 中的 Head 部分重新设计为 3 层跨越金字塔特征尺度均衡的金字塔卷积（Scale-Equalizing Pyramid Convolution，SEPC）[10]。采用 SIoU Loss[11]替代 CIoU Loss 计算边框损失，最后根据数据集特性采用 K-means++聚类出最佳先验锚框参数。改进后的 S3-YOLOv5s 网络在 MINE WORKER 数据集中可进一步提高目标检测精度，降低漏检、错检率。

1 YOLOv5s 网络

YOLOv5s 的网络结构如图 1 所示，主要由 Backbone，Head 和 Direction 三部分组成，Backbone 用于提取 3 层不同尺度的特征金字塔，主要包含 CBL，C3 和 SPPF 模块。CBL 模块为复合卷积模块，即包含卷积、BN 和 SiLu 激活函数。C3 结构去掉了 BottleneckCSP[12]中的 1 个 Conv 卷积，剩下 3 层 Conv 卷积。SPPF 结构将原本 SPP 结构中并行的 3 路 MaxPool 改为串行的 MaxPool，在保证性能不下降的同时提升了前向及反向计算速率。Head 部分采用的是 FPN＋PAN 的特征融合结构，FPN 上采样传递高层特征，PAN 向下传递低层特征。

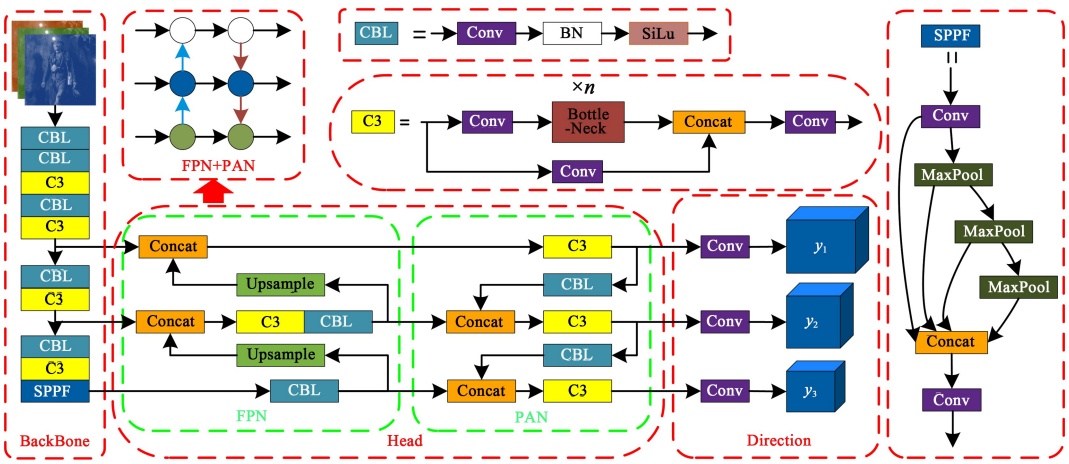


图 1 YOLOv5s 网络结构示意图

2 S3-YOLOv5s 网络

YOLOv5s 网络虽然在目标识别精度和识别速度上有较好的性能，但由于煤矿井下环境复杂（潮湿、煤尘大、光线不足且整体偏暗），导致所采集到的图像质量较差且目标的可分辨性不强。此外，工作人员工作服颜色较深，在光照不足、照度不均的情况下易造成目标与矿井背景的混淆[13]。各种个人防护设备尺度变化大，长宽比差异大，目标容易重叠，很难检测完全，造成误报。为此，本文在 YOLOv5s

算法的基础上，设计了一种基于注意力及尺度均衡特征金字塔卷积的特征融合 YOLOv5s 网络，加强特征表达和融合，有效提升了算法的识别精度。

2.1　SimAM 注意力

YOLOv5s 中采用金字塔特征提取作为骨干网络，考虑到煤矿井下的复杂环境，本文在主干网络中添加注意力机制以提高骨干网络的特征提取能力，增强特征表示能力。传统的注意力机制如 SE[14]，CBAM[15]，ECA[16]，CA[17] 等都是沿通道或空间位置来分配注意力权重且需要额外的子网络结构和模型参数，一方面基于通道或空间并不能生成真正的三维权重，另一方面，额外的子网络结构必然导致网络的复杂度增加。本文采用 SimAM 无参注意力机制，其注意力权值分配方式如图 2 所示。

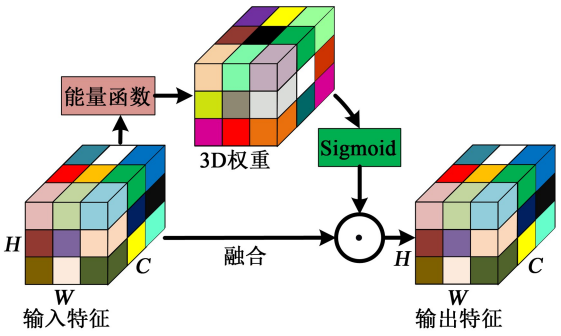


图 2　SimAM 注意力权重分配示意图

SimAM 通过设计能量函数来直接产生真正有效的三维权重，不需要增加额外的子网络结构，也没有额外的模型参数。能量函数的设计受神经科学理论启发，旨在度量神经元之间的线性可分性来寻找重要的神经元，对于输入的每一个神经元，具体的能量函数定义如下：

$$e_t(x_t)=\frac{4(\sigma^2+\lambda)}{(x_t-\mu)^2+2\sigma^2+2\lambda} \tag{1}$$

其中，x_t 表示当前通道中输入特征的目标神经元，$\mu=\frac{1}{W\times H}\sum_{i=1}^{W\times H}x_i$，$\sigma^2=\frac{1}{W\times H}\sum_{i=1}^{W\times H}(x_i-\mu)^2$ 为该通道中所有神经元的平均值和方差，以避免重复计算，降低计算成本，λ 为权重常数。综上，SimAM 无参注意力模块的输出公式为

$$Y=\mathrm{sigmoid}\left(\frac{1}{E(X)}\right)\odot X \tag{2}$$

其中，E 为跨通道及空间维度的能量函数矩阵，添加 Sigmoid 函数抑制注意力权重的异常值后与输入特征矩阵对应元素相乘即得到最终输出。本文经多次试验对比，将 SimAM 添加至 C3 模块的 BottleNeck 中，构建 C3SimAM 模块，如图 3 所示，选定 $\lambda=10^{-2}$，以达到最佳特征提取效果。

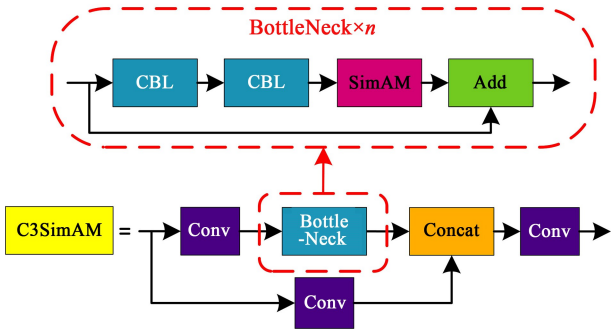


图 3　C3SimAM 模块示意图

2.2　特征融合网络

在特征金字塔提取时，网络从多个尺度对目标位置进行预测，以兼顾不同尺度的目标。但特征层每个层级所携带的信息并不对等，深层的特征图包含有较高的语义信息；浅层的特征图包含有较高的空间信息。这使得网络在预测目标时，无法直接从特征层获取语义及空间信息。因此，许多学者提出了不同的特征融合策略，包括 FPN[18]，PANet[19]，ASFF[20]，NAS-FPN[21]，BiFPN[22] 等。

YOLOv5s 特征融合网络采用了如图 4 所示 FPN＋PAN 结构。FPN 将高层特征进行上采样与前一层特征进行 Concat 操作，实现了将深层的语义信息带给浅层。PAN 将 FPN 中融合的特征进行反向融合，避免浅层特征中定位信息的丢失。

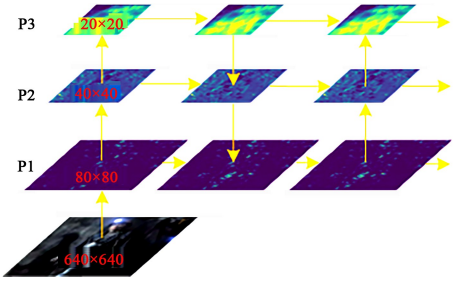


图 4　YOLOv5s 特征融合网络示意图

然而，这些特征融合的设计都是将特征图像上采样或下采样后，直接将其拼接，而没有研究特征金字塔的内在特性，也没有考虑来自不同路径的特征对网络的重要程度，以使所有特征映射都能做出最大化的贡献。本文采用金字塔卷积来融合不同层次的特征，其基本流程如图 5 所示。

尺度均衡金字塔卷积（SEPC）实际上是横跨特征层和特征分辨率的三维卷积。对于有 3 层特征层

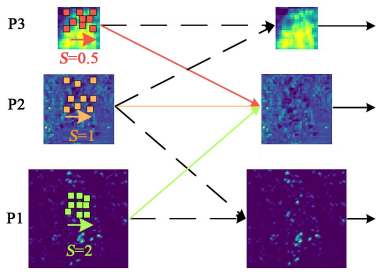


图 5 金字塔卷积示意图

输出的 YOLOv5s 网络，SEPC 可以表示为 3 个不同的二维可变形卷积[23]，以缓解特征层之间的非线性特征映射关系。然而，不同特征层之间存在分辨率不匹配。随着特征层级别的升高，分辨率将减小。为了兼容这种不匹配，在不同的特征层中进行卷积时，为不同的卷积核匹配不同的步长。SEPC 的输出为

$$y^l = w_1 s_{0.5} x^{(l+1)} + w_0 x^l + w_{(-1)} s_2 x^{(l-1)} \tag{3}$$

其中，l 表示特征层的层级，w_1，w_0 和 w_{-1} 是 3 个独立的二维卷积核，x 是输入特征层，s_2 表示步长为 2 的卷积。$s_{0.5}$ 表示步长为 1 的正常卷积与连续的双线性上采样层相结合。即：

$$y^l = \mathrm{Upsample}(w_1 x^{(l+1)}) + w_0 x^l + w_{(-1)} s_2 x^{(l-1)} \tag{4}$$

通过这种方式，将每个特征层中的特征通过变形偏移进行均衡，再进行层间融合，S3-YOLOv5s 网络结构如图 6 所示，图中 P1，P2，P3 分别为主干网络输出的 3 个特征层，通道数分别为 128，256，512。由公式(3)可以看出，SEPC 要求输入的 3 个特征层通道数一致，为此，在进行 SEPC 特征融合前，P1，P3 分别经过 1×1 的卷积，将通道数改为 256 与 P2 特征层保持一致。复合的 SEPC 模块包含 SEPC，iBN 和 SiLu 激活函数。iBN 为 3 层共享 BN，全部参数和统计共享，iBN 层的公式为

$$\begin{cases} \mu_{\mathrm{B}} = \dfrac{1}{3}\left(\dfrac{1}{m}\sum_{i=1}^{m} x_i^1 + \dfrac{1}{n}\sum_{i=1}^{n} x_i^2 + \dfrac{1}{k}\sum_{i=1}^{k} x_i^3\right) \\ \sigma_{\mathrm{B}}^2 = \dfrac{1}{3}\left(\dfrac{1}{m}\sum_{i=1}^{m}(x_i^1-\mu_{\mathrm{B}})^2 + \dfrac{1}{n}\sum_{i=1}^{n}(x_i^2-\mu_{\mathrm{B}})^2 + \dfrac{1}{k}\sum_{i=1}^{k}(x_i^3-\mu_{\mathrm{B}})^2\right) \\ y_i = \gamma\dfrac{x-\mu_{\mathrm{B}}}{\sqrt{\sigma_{\mathrm{B}}^2+\varepsilon}} + \beta \end{cases} \tag{5}$$

其中，γ 和 β 为可学习参数，μ_{B} 和 σ_{B}^2 为统计的 3 层特征层的共享均值及标准差，$x_i^l(l=1,2,3)$ 中，l 代表不同的特征层，ε 为极小非零正值，保证分母不为零。普通 BN 单独计算每个特征层的统计结果，iBN 统计所有的特征层，而不是单层统计，方差会变得更小更稳定。

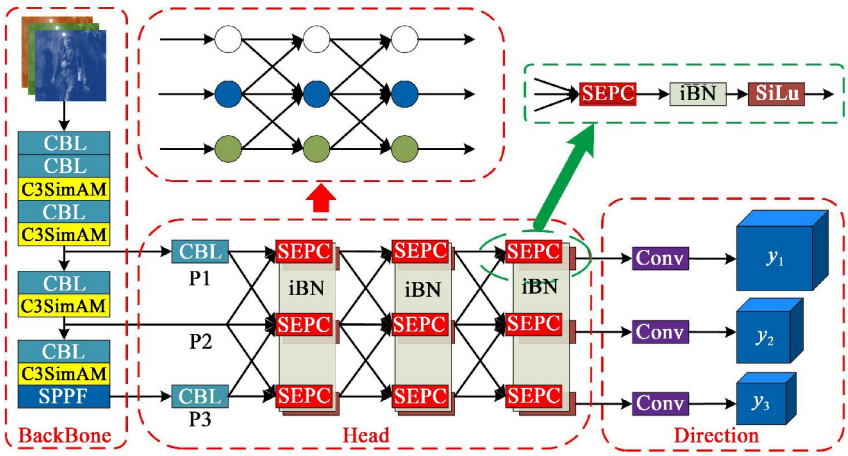


图 6 S3-YOLOv5s 网络结构示意图

2.3 SIoU 损失函数

YOLOv5s 的损失函数为边框回归损失、目标置信度损失和分类损失的加权和，其公式如下：

$$Loss = \lambda_1 L_{\mathrm{box}} + \lambda_2 L_{\mathrm{obj}} + \lambda_3 L_{\mathrm{cls}} = \sum_{i}^{N}\left(\lambda_1 \sum_{j}^{B_i} L_{\mathrm{GIoU}_j} + \lambda_2 \sum_{j}^{S_i \times S_i} L_{\mathrm{obj}_j} + \lambda_3 \sum_{j}^{B_i} L_{\mathrm{cls}_j}\right) \tag{6}$$

式中，N 为检测层个数，$S_i \times S_i$ 为相应检测层的分割网格数，B_i 为标签分配到的先验框目标个数，λ_1，λ_2 和 λ_3 为加权系数。L_{box} 和 L_{cls} 均采用二元交叉熵(Binary Cross Entropy，BCE)，其计算方式如下：

$$\begin{cases} y_i = \mathrm{Sigmoid}(x_i) = \dfrac{1}{1+e^{-x_i}} \\ L = -\sum_{i=1}^{N}(y_i^* \log(y_i) + (1-y_i^*)\log(1-y_i)) \end{cases} \tag{7}$$

式中，x_i 表示当前类别预测值，y_i 表示经过 Sigmoid 函数后将 x_i 映射到 0－1 的空间中反映当前类别概率，N 表示类别总数，y_i^* 表示当前类别的真实值(0 或 1)。L_{box} 使用 CIoU[24] 作为边界框回归损失计算，其公式如下：

$$\begin{cases} Loss_{CIoU}=1-IoU+\dfrac{\rho^2(p,p^{gt})}{|c^2|}+\dfrac{V^2}{(1-IoU)+V} \\ V=\dfrac{4}{\pi^2}\left(\arctan\dfrac{w^{gt}}{h^{gt}}-\arctan\dfrac{w}{h}\right)^2 \end{cases} \tag{8}$$

其中，P 和 P^{gt} 是预测框 B 和真实框 B^{gt} 的中心点，c 是对角线长度，ρ 表示欧氏距离，V 用来度量宽高比的一致性。

可以看到，CIoU 虽然解决了 GIoU 的退化问题，增加了宽高 Loss，避免了 DIoU 中心点重合时交并比一样无法区分的情况，但依旧没有考虑真实框与预测框之间不匹配的方向，导致模型训练收敛速度慢且预测框不够精确。本文使用 SIoU 作为边框回归损失计算，其公式如下：

$$\begin{cases} Loss_{SIoU}=1-IoU+\dfrac{\Delta+\Omega}{2} \\ \Delta=2-e^{-\gamma\frac{x^{gt}-x}{c_w}}-e^{-\gamma\frac{y^{gt}-y}{c_h}} \\ \Omega=(1-e^{-\omega_w})^2+(1-e^{-\omega_h})^2 \end{cases} \tag{9}$$

其中 $\omega_w=\dfrac{|w-w^{gt}|}{\max(w,w^{gt})}$，$\omega_h=\dfrac{|h-h^{gt}|}{\max(h,h^{gt})}$，$\gamma=1+2\sin^2\left(\arcsin\left(\dfrac{\max(y^{gt},y)-\max(x^{gt},x)}{\sqrt{(x^{gt}-x)^2+(y^{gt}-y)^2}}\right)-\dfrac{\pi}{4}\right)$。

SIoU 在考虑框中心之间的距离、重叠区域和纵横比的基础上，加入角度损失如图 7 所示，以有效减少损失的总自由度，让预测框很快地移动到最近的轴，加速训练收敛及边框回归效果。

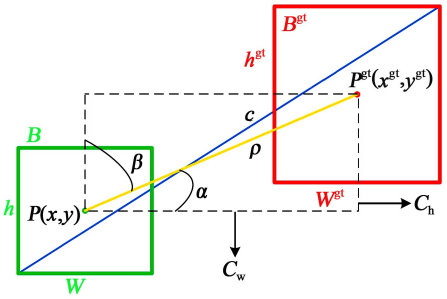


图 7　真实边框与预测边框

2.4　先验锚框聚类

YOLOv5s 网络将目标检测数据集 COCO 进行 K 均值维度聚类(K-Means)，得到初始先验锚框参数，但因 COCO 数据集的目标种类有 80 个类别，而本文中的个人防护设备检测种类为 8 个(person，face，hat，mask，lamp，rescuer，coat，pant)，且尺寸差异较大，默认的先验锚框参数不能满足矿井环境目标检测的实际需要，所以需要重新设计先验锚框大小。相较于只依赖人的先验知识而设计的锚框尺寸大小，本文采用 K-means＋＋聚类算法多次对 MINE WORKER 数据集中已标注的目标边界锚框聚类，产生不同数量、不同大小的先验框，尽可能增加先验框与实际目标框之间的匹配度，从而进一步提高检测准确率。K-means＋＋改进了 K-means 中随机初始质心的缺点，同时采用 IoU 来计算聚类中心的距离，公式如下：

$$S=\min\sum_{i=0}^{k}d(\text{box},\text{centroid})=\min\sum_{i=0}^{k}(1-IoU) \tag{10}$$

式中，S 表示聚类距离，box 为锚点框，centroid 为选定的聚类中心锚框。图 8 显示了 MINE WORKER 数据集标签锚框的分布及聚类效果，聚类后的最终锚框参数如表 1 所示。

表 1　先验锚框聚类结果

特征尺寸	先验锚框尺寸
[20×20]	[124×527]，[155×540]，[188×566]
[40×40]	[80×55]，[108×201]，[169×216]
[80×80]	[26×33]，[43×44]，[48×79]

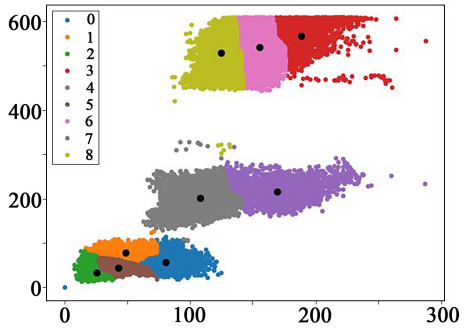


图 8　数据集标签锚框大小分布图

3　实验与分析

3.1　MINE WORKER 数据集

目前尚未有公开发布的矿井安全防护设备数据集，本文构建了一个新的 MINE WORKER 数据集，用于检测相关矿井防护设备。鉴于矿井作业现场要求携带相关安全防护设备，本文所采用的 MINE WORKER 数据集为专业作业人员所拍摄的画面共 5 131 张。但由于需要检测的目标种类较多，真实拍

摄的正负样本分布极度不均衡，如工作人员几乎都戴有安全帽及穿有防护服，导致训练时出现过拟合。本文在拍摄真实工作现场的基础上，针对真实场景数据集有限、数据集中各类别数量不均衡的情况，通过穿戴矿井人员专业设备，模拟采集不同角度、场景数据，制作模拟的数据集，增加数据集多样性。最终的 MINE WORKER 数据集包含 10 322 张图片，图 9 显示了相关类别的分布情况。

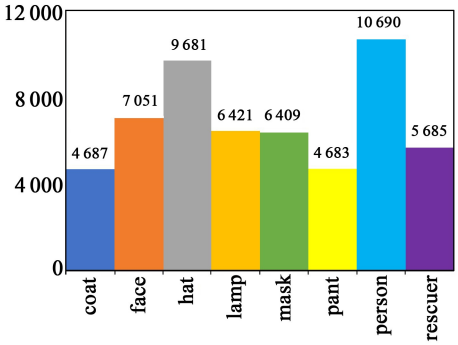


图 9 数据集类别分布情况

图 10 为部分真实场景数据集图片及模拟制作的数据集图片。本文参照 PASCAL VOC 目标检测数据集标准格式，采用 Labelimg 工具标注要检测的目标，标注的 8 类目标分别为 person，face，hat，mask，lamp，rescuer，coat，pant。训练集、测试集和验证集按 8∶1∶1 的比例随机分配。同时使用 Mosaic 数据增强、随机翻转、平移、颜色变换等进行处理，使数据集具有更好的泛化效果。



(a) 真实数据



(b) 模拟数据

图 10 部分数据集样本

3.2 实验环境及衡量指标

本文使用的实验环境参数如表 2 所示，训练的迭代次数为 300 次，迭代批量大小设置为 16，优化器选择 SGD。同时为验证改进后 S3-YOLOv5s 网络的性能，本文采用准确率(Precision，P)，召回率(Recall，R)，均值平均精度(mean Average Precision，mAP)，模型大小，前向推理速度对网络进行评估测试。相关计算公式如下：

$$\begin{cases} P = TP/(TP+FP) \\ R = TP/(TP+FN) \\ mAP = \sum AP/N \end{cases} \tag{11}$$

其中，TP 为正确分类的正样本数，FP 为被错误判定为正样本的负样本数。FN 为被错误判定为负样本的正样本数。AP 是以召回率为横轴，准确率为纵轴构成的 Precision-Recall 曲线下的面积，N 为类别总数。

表 2 实验环境参数

参数	配置
操作系统	Ubuntu 18.04
CPU	Intel Core i5-8500 (16 GB)
GPU	GeForce GTX 2070 (8 GB)
编程语言	Python 3.7.4
深度学习框架	Pytorch 1.9.0，torchvision 0.9.0
CUDA 版本	Cudatoolkit 10.2

3.3 对比试验

为验证本算法的先进性，本文使用 MINE WORKER 数据集，在目前主流的目标检测网络中分别进行训练及测试，形成对比实验，相关实验结果见表 3。

表 3 不同网络实验对比表

网络模型	分辨率	P/%	R/%	mAP@0.5(%)	mAP@0.5∶0.95(%)	模型大小/MB	速度/MS
Vgg16-Faster-RCNN	600×600	90.01	88.23	91.46	67.95	522	52
Resnet50-Faster-RCNN	600×600	94.12	91.75	93.36	70.85	108	94
YOLOv3	416×416	89.36	87.65	89.73	67.02	236	23
SSD	300×300	88.58	86.76	89.55	66.63	100	18
YOLOv4	416×416	91.63	90.11	92.00	69.20	244	27
S3-YOLOv5s(our)	640×640	92.24	90.34	92.86	70.05	19.8	13

分析表 3 可知，以 Resnet50 为主干网络的 Faster-RCNN 的 mAP 值最高，达到 93.36%，但前向预测速度达到 94MS，以 Vgg16 为主干网络的 Faster-RCNN 模型大小达到 522 MB；YOLOv4 的检测精度与本文方法相近，但模型大小依然有 244 MB，且前向预测速度仍然高达 27 MS，为 S3-

YOLOv5s 模型的两倍，以上两个网络均难以满足实时性及嵌入式设备的硬件需求。YOLOv3 及 SSD 在前向预测速度上有一定优势，但精度远低于 S3-YOLOv5s，综合对比不同网络，本文方法在检测精度上仅次于 Resnet50-Faster-RCNN，但前向预测速度及模型文件大小具有明显优势。

3.4　消融实验

为验证各项改进的有效性，本文以 YOLOv5s 模型为 baseline 进行不同改进来做对比实验，训练数据集为 MINE WORKER，表 4 给出了相关实验结果。

表 4　消融实验对比表

改进	*P*/%	*R*/%	mAP@0.5(%)	mAP@0.5∶0.95(%)	模型大小/MB	速度/MS
Baseline	88.31	86.75	89.64	65.95	13.8	7
Baseline+SIoU	89.05	87.35	90.36	66.31	13.8	7
Baseline+SimAM+SIoU	90.58	88.17	90.88	67.02	13.8	9
Baseline+SimAM+SEPC+SIoU	92.24	89.34	92.86	69.05	19.8	13

实验结果表明，原始 YOLOv5s 的精确率为 88.31%，召回率为 86.75%，IoU 标准取 0.5 的 mAP 为 89.64%，IoU 标准取 0.5～0.95 的 mAP 为 65.95%。以此为基础，即使只将损失函数改为 SIoU，各项指标也有一定的提升，且模型大小、FPS 几乎没有变化。在加入 SimAM 无参注意力模块后，mAP@0.5 和 mAP@0.5∶0.95 分别提升了 1.24 与 1.07。对于层间特征非线性变化问题，本文还比较了 SEPC 的实验结果。实验结果表明，引入可变形卷积作为尺度均衡的 SEPC 特征金字塔卷积模块可以帮助对齐不同层特征，加强特征融合，mAP@0.5 相对原始 YOLOv5s 提升了 3.22%，且 SEPC 只增加少量的卷积偏移参数，相比于 Baseline，模型大小增加了 6 M。但由于使用了共享权重的 iBN 层，在进行前向推理时，iBN 不需要计算现场统计数据，而且可以合并到前一个卷积层中。这在推理速度方面带来了显著的优势，网络总体的预测时间依然保持在 13 MS。

图 11 显示了部分预测结果，对比可以看出 S3-YOLOv5s 在目标误检及漏检方面均有所改进，许多漏检的标签在 S3-YOLOv5s 中有预测输出，同时 SIoU 的加入使得预测框精度也有所提升。



（a）标签　（b）YOLOv5s　（c）S3-YOLOv5s

图 11　部分预测结果

4　结论

本文提出了一种改进的 S3-YOLOv5s 方法，该方法首先在主干网络中加入 SimAM 无参注意力加强特征提取，采用 3 层尺度均衡的金字塔卷积进行特征融合，对比普通人为设计的特征融合模块，金字塔层间特征卷积为机器学习的自然结果，实验表明，改进后的网络提升了特征表达能力，进一步融合了

深、浅层特征信息，提高了目标的检测精度。同时，本文在尽量不增加模型复杂度的情况下，将损失函数改为 SIoU，也使模型有一定的性能提升。本文在自制的 MINE WORKER 数据集上进行对比实验，验证了相关改进的有效性。本文方法最终实现了矿井环境对工作人员安全防护设备的多目标实时检测，对矿井环境安全作业智能化发展有着重要意义。

参考文献：

[1] Nath N D, Behzadan A H, Paal S G. Deep learning for site safety: Real-time detection of personal protective equipment [J]. Automation in Construction, 2020, 112: 103085.

[2] Wang Z, Wu Y, Yang L, et al. Fast personal protective equipment detection for real construction sites using deep learning approaches[J]. Sensors, 2021, 21(10): 3478.

[3] Li T, Chen Q, Gan J, et al. Investigating into deep neural networks for applicable hard-hat wearing detection in substations[J]. J. of Physics: Conf. Series, 2021, 2002(1): 012062.

[4] Fu Chuan, Wang Rongxin. Research on safety helmet wearing YOLO-V3 detection technology improvement in mine environment[J]. J. of Physics: Conf. Series, 2019, 1345(4): 042045.

[5] Wang X, Niu D, Luo P, et al. A safety helmet and protective clothing detection method based on improved-Yolo V3[C]// 2020 Chinese Automation Congress (CAC), 2020: 5437-5441.

[6] Redmon J, Farhadi A. Yolov3: An incremental improvement [DB/OL]. (2018-4-8). http://arXiv.org/pdf/1804.02767.

[7] Zhuo N, Yun G. Research on application of target detection network based on SSD in underground coal mine[C]// 2021 5th Inter. Conf. on Computer Science and Artificial Intelligence, 2021: 21-26.

[8] Li X, Wang S, Liu B, et al. Improved YOLOv4 network using infrared images for personnel detection in coal mines[J]. J. of Electronic Imaging, 2022, 31(1): 013017.

[9] Yang L, Zhang R Y, Li L, et al. Simam: A simple, parameter-free attention module for convolutional neural networks[C]// Inter. Conf. on Machine Learning, 2021: 11863-11874.

[10] Wang X, Zhang S, Yu Z, et al. Scale-equalizing pyramid convolution for object detection[C]// Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition, 2020: 13359-13368.

[11] Gevorgyan Z. SIoU loss: More powerful learning for bounding box regression [DB/OL]. 2022-5-25. http://arXiv.org/ftp/arXiv/papers/2205/2205.12740.

[12] Wang C Y, Liao H Y M, Wu Y H, et al. CSPNet: A new backbone that can enhance learning capability of CNN[C]// Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition Workshops, 2020: 390-391.

[13] 周晨晖. 基于深度学习的煤矿复杂场景人员检测与统计分析方法研究[D]. 徐州：中国矿业大学，2019.
Zhou Chenhui. Research on personnel detection and statistical analysis in coal mine complex scenes based on deep learning [D]. Xuzhou: China University of Mining and Technology, 2019.

[14] Hu J, Shen L, Sun G. Squeeze-and-excitation networks [C]// Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, 2018: 7132-7141.

[15] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]// Proc. of the European Conf. on Computer Vision (ECCV), 2018: 3-19.

[16] Wang Q, Wu B, Zhu P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks[C]// Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, 2020: 11531-11539.

[17] Hou Q, Zhou D, Feng J. Coordinate attention for efficient mobile network design[C]// Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition, 2021: 13713-13722.

[18] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]// Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, 2017: 2117-2125.

[19] Liu S, Qi L, Qin H, et al. Path aggregation network for instance segmentation[C]// Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, 2018: 8759-8768.

[20] Liu S, Huang D, Wang Y. Learning spatial fusion for single-shot object detection[DB/OL]. 2019-10-21. http://arXiv.org/pdf/1911.09516.

[21] Ghiasi G, Lin T Y, Le Q V. Nas-fpn: Learning scalable feature pyramid architecture for object detection[C]// Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition, 2019: 7036-7045.

[22] Tan M, Pang R, Le Q V. Efficientdet: Scalable and efficient object detection[C]// Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition, 2020: 10781-10790.

[23] Dai J, Qi H, Xiong Y, et al. Deformable convolutional networks[C]//Proc. of the IEEE Inter. Conf. on Computer Vision, 2017: 764-773.

[24] Zheng Z, Wang P, Ren D, et al. Enhancing geometric factors in model learning and inference for object detection and instance segmentation[J]. IEEE Trans. on Cybernetics, 2021, 52(8): 8574-8586.

作者简介：

代少升(1974—)，男，河南省潢川县人，博士，教授，主要从事红外成像系统及 SOPC 嵌入式系统的设计与开发。