

# 基于 MK-YOLOV4 的矿区人员无标注视频检索方法

赵云辉<sup>1</sup>, 程小舟<sup>2</sup>, 董铠文<sup>1\*</sup>, 云霄<sup>1</sup>, 孙彦景<sup>1</sup>, 韩英杰<sup>1</sup>

<sup>1</sup>中国矿业大学信息与控制工程学院, 江苏 徐州 221116;

<sup>2</sup>中钢集团马鞍山矿山研究院有限公司, 安徽 马鞍山 243000

**摘要** 对矿区人员入矿、出矿及重要生产活动行为进行精准定位与准确识别是实现矿区智能安全生产的重要基础。针对复杂的矿区生产环境,提出一种基于 MK-YOLOV4 的矿区人员无标注视频检索方法,实现对矿区重要关口监控无标注视频的多人员目标检测和各人员身份重识别。首先,提出 MK-YOLOV4 算法,实现无标注视频多人员检测,在 YOLOV4 上构建多尺度预测,结合 K-means++ 算法聚类生成符合样本特点的 anchor box,增强卷积神经网络对小目标的表征学习。其次,提出基于外观不变性的通道注意力特征提取网络,实现矿区人员身份的精确重识别,并针对矿区人员统一工作服的难点,提出基于权重约束的难样本采样损失函数,结合 Color jitter 和随机擦除两种数据增强策略,提高身份识别网络的精确性和鲁棒性。最后,针对现有训练数据集类别少、场景样本单一等特点,构建了更符合矿区场景特点的 Miner-Market 矿区人员重识别数据集,并在标准数据集和该数据集上对所提方法进行验证,充分证明了所提方法具有较高的检索性能和识别精度。

**关键词** 机器视觉; MK-YOLOV4; 人员检测; 人员重识别; Miner-Market 数据集

中图分类号 TP391.4

文献标志码 A

doi: 10.3788/LOP202259.0415003

## Unlabeled Video Retrieval Method of Mining Personnel Based on MK-YOLOV4

Zhao Yunhui<sup>1</sup>, Cheng Xiaozhou<sup>2</sup>, Dong Kaiwen<sup>1\*</sup>, Yun Xiao<sup>1</sup>, Sun Yanjing<sup>1</sup>, Han Yingjie<sup>1</sup>

<sup>1</sup>School of Information and Control Engineering, China University of Mining and Technology, Xuzhou, Jiangsu 221116, China;

<sup>2</sup>Sinosteel Maanshan Institute of Mining Research Co., Ltd., Maanshan, Anhui 243000, China

**Abstract** Precise positioning and accurate identification of personnel entering, exiting, and conducting important production activities in the mining area are important foundations for achieving intelligent and safe production in the mining area. This study proposes an unlabeled video retrieval method for personnel in the mining area using MK-YOLOV4 in the complex mining area production environment, which can realize multiperson target detection and reidentification of an individual's identity on unlabeled video of important gateway monitoring in the mining area. First, this study proposes the MK-YOLOV4 algorithm to achieve multiperson detection of unlabeled videos by building multiscale predictions on YOLOV4, and the K-means++ algorithm is combined to generate an anchor box that meets the characteristics of the samples, which can improve the representation learning of the convolutional

收稿日期: 2021-03-01; 修回日期: 2021-03-29; 录用日期: 2021-04-02

基金项目: 国家自然科学基金青年项目(61902404)、江苏省自然科学基金青年项目(BK20180640)、中国矿业大学大学生创新训练计划(20200142cx)

通信作者: \*dongkaiwen1996@outlook.com

neural network for small targets. Second, we propose a channel attention feature extraction network based on appearance invariance to achieve accurate reidentification of personnel in mining areas. Aiming at solving the problem of uniform work clothes for personnel in mining areas, this study proposes a weight-constrained difficult sample sampling loss function with two data enhancement strategies, where Color jitter and random erasure are combined to improve the accuracy and robustness of the identification network. Finally, according to the characteristics of the existing training dataset with few categories and single scene samples, a Miner-Market mining personnel reidentification dataset is constructed with the characteristics of the mining scenes, and the proposed method is verified on the standard dataset and Miner-Market dataset. The verification confirmed that the proposed method has high retrieval performance and recognition accuracy.

**Key words** machine vision; MK-YOLOV4; personnel detection; personnel reidentification; Miner-Market dataset

## 1 引言

矿区智能监控对实现我国安全生产具有重要意义。实际矿区环境复杂,视频监控摄像头距离目标较远,工作人员穿着统一工作服导致身份难以辨认,外界存在大面积遮挡、光照变化等因素,都会导致人员检索结果难以达到安全生产所要求的标准。Zheng等<sup>[1]</sup>从网络模型训练的角度出发,提出了端到端的人员检索网络,实现人员检测和识别的功能。Xiao等<sup>[2]</sup>从损失函数设计的角度出发,通过设计更为合理的 online instance matching(OIM)损失函数来提取更深层次的人员特征,从而实现人员检测和识别过程。Liu等<sup>[3]</sup>通过设计独特的 neural person search machine(NPSM)网络模型完成人员的检测和识别过程。以上工作都将人员检测和人员重识别结合在一起,从网络模型训练和损失函数改进为出发点进行研究,忽略了改进网络结构提取的更具判别力的特征,没有将二者高效结合。

人员检测作为人员检索方法的第一步,作为目标检测的子任务,目的是判断视频帧列中是否存在行人,并获取其位置和边界信息,然后再进行后续的人员重识别任务。例如,Zhao<sup>[4]</sup>提出基于模糊理论的检测算法,结合混合高斯建模和模糊函数实现矿区人员检测。Chai等<sup>[5]</sup>将背景建模与统计学习相结合,以实现矿区人员检测。崔鹏翔等<sup>[6]</sup>通过结合颜色空间和混合高斯阴影模型实现阴影环境下的人员检测。以上传统人员检测算法无法适应复杂的矿区场景变化,而以改进版 Faster R-CNN<sup>[7]</sup>和改进版 Mask R-CNN<sup>[8]</sup>为代表的 two-stage 算法,以改进版 SSD<sup>[9]</sup>和 YOLOV4<sup>[10]</sup>为代表的 one-stage 算法,是深度学习在目标检测上的应用,极大提高了目标检测的精度和模型的自适应性。但人员检测仅仅能够获取行人位置和边界信息,无法对其身份进行识别,对实现矿区安全监控具有一定的局限性。

人员重识别是指对于给定的目标行人图像,在不同的摄像头以及不同的场景下搜索视频中与该目标行人属于同一身份行人的技术。矿区人员重识别可以对检测到的人员身份进行辨认和识别,结合人员检测可实现矿区人员检索。传统的行人重识别方法分为人工特征设计和度量学习两种,然而无法适应复杂的外观变化。而基于深度学习的行人重识别方法不仅能学习到更丰富、更具识别力的特征,而且能实现特征提取和距离度量两个模块协同的端对端训练,主要包括分类模型<sup>[11]</sup>和验证模型<sup>[12]</sup>两种。以上行人重识别方法都是基于已裁剪好的行人图像实现匹配的,但在实际矿区场景中,需要先对行人区域进行预先标注才能进行下一步的人员重识别。

针对以上问题,本文提出一种基于 MK-YOLOV4的矿区人员无标注视频检索方法,先通过人员检测算法预先获得矿区人员区域图像,再对人员区域进行身份识别,实现矿区人员的无标注检索。所提方法包含人员检测和人员重识别两个模块,首先获取矿区环境下不同位置不同场景的无标注视频流,使用所提 MK-YOLOV4人员检测算法对每一帧人员图片进行检测,获取人员位置及边界信息;然后将获取的人员图片输入到所提基于外观不变性的通道注意力特征提取网络(CA-NET)中进行特征提取;最后依次对目标人员特征向量和候选人员特征向量进行相似度度量,找到目标人员在候选库中对应的人员图片,并返回其对应的身份ID信息。

## 2 人员检测模块

人员检测作为人员重识别的前提和基础,检测的精度对于识别的准确性至关重要。由于实际矿区场景下摄像头距离目标人员较远,增加了小目标识别的难度。为了解决该问题,提出基于小目标的 MK-YOLOV4人员检测算法。为了更加充分地利用浅层特征,增强卷积神经网络对小目标的表征学

习, MK-YOLOV4 人员检测算法在 YOLOV4 的基础上增加了多尺度预测数量, 并结合人员高度大于宽度这一固有属性, 使用 K-means++ 算法<sup>[13]</sup> 聚类生成符合样本特点的 anchor box。

### 2.1 多尺度预测

YOLOV4 检测算法作为目标检测领域的集大成者, 极大提高了目标检测精度和速度, 其网络结构主要包含 Backbone、Neck 和 Prediction 三部分, 如图 1 短虚线框所示。其中最右边的 CBM 模块和 CBL 模块分别表示卷积层、归一化层和激活函数层; res unit 模块表示深度残差网络, 主要由 CBM 模块组成; SPP 结构<sup>[14]</sup> 由  $1 \times 1$ 、 $5 \times 5$ 、 $9 \times 9$ 、 $13 \times 13$  四个不同的池化方式组成; csp-resn 表示结合 CSPNet<sup>[15]</sup> 设计思想的深度残差网络。Backbone 部分由 CBL 模块和 csp-resn 组成, 在深度残差网络基础上采用 CSPNet 设计思想, 对两个分支的特征向量进行跨阶段层次合并, 重复利用梯度信息, 在减

少模型计算量的同时保证准确率不变或略有提升。Neck 部分由 CBL 模块、SPP 结构、upsampling 和 concat 组成, 采用特征金字塔网络 (FPN) 结构和路径聚合网络<sup>[16]</sup> 设计思想, 通过由下至上和由上至下的特征融合过程, 获得不同尺度的目标信息, 一定程度上缓解了目标的尺度变化问题。YOLOV4 中的 Prediction 阶段输出  $y_1$ 、 $y_2$ 、 $y_3$  三个不同尺度的特征图, 由于浅层特征分辨率较高, 包含更多的位置和细节等信息, 高层特征分辨率较低, 具有更强的语义信息, YOLOV4 在特征融合阶段更多地使用了高层语义信息, 没有充分利用浅层特征, 使得小目标的位置和细节等信息丢失。为解决此问题, 在 YOLOV4 的基础上增加  $y_4$  和  $y_5$  两个不同的预测输出, 将多尺度预测数量扩展为 5 个, 在提取更加丰富浅层信息的同时, 获取更多的位置和细节信息, 提高真实场景下小目标的检测精度。模型的网络结构如图 1 实线框所示。

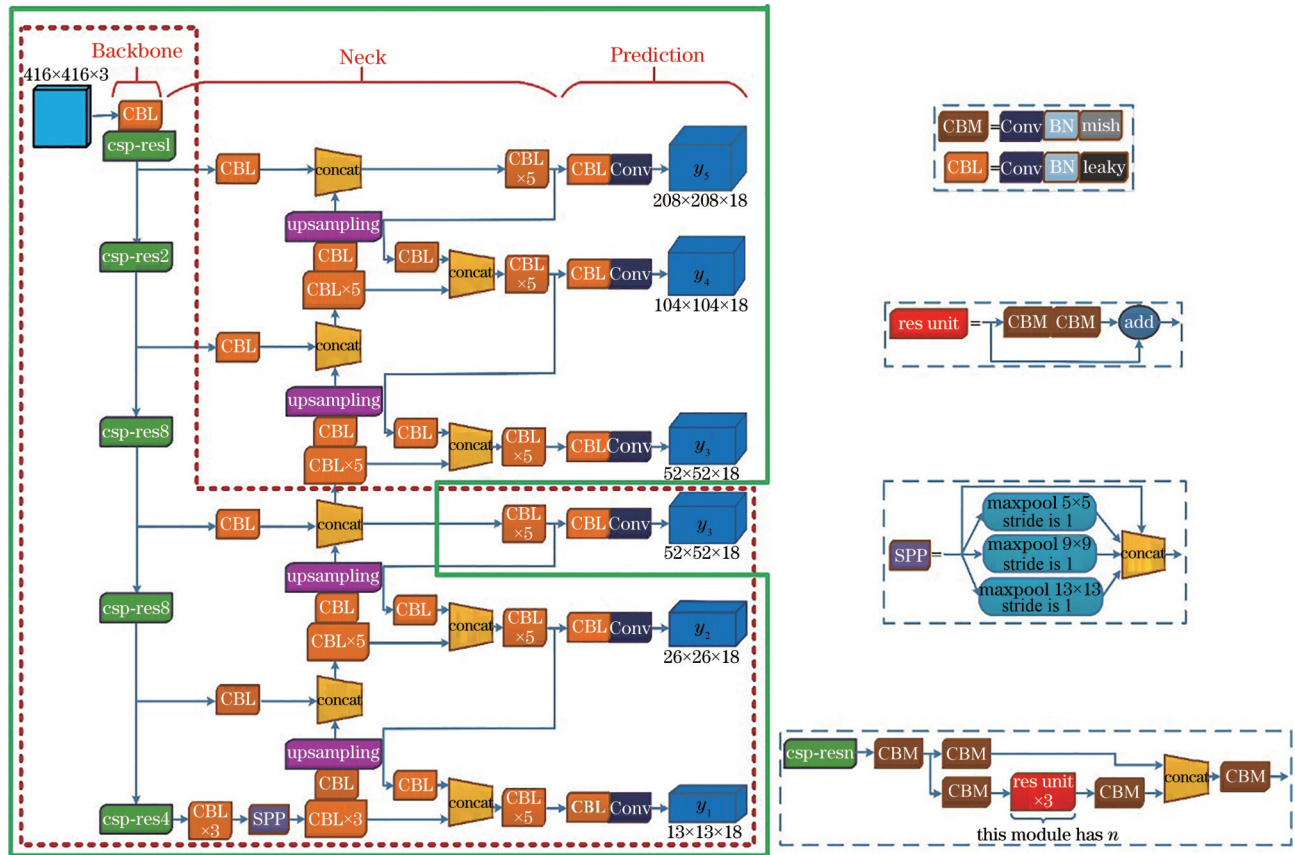


图 1 MK-YOLOV4 网络结构  
Fig. 1 MK-YOLOV4 network structure

### 2.2 anchor box 生成

常用的无监督聚类算法都是以欧氏距离作为相似度量, 这种距离度量方式会使大尺度框比小

尺度框产生更多误差, 而在目标检测领域, 聚类的目的是选择合适的 anchor box, 使得 anchor box 与 ground truth 有着更大的交互比 (IOU) 值。因此

Redmon 等<sup>[17]</sup>对这种度量方式加以改进,以 concat 作为评判标准,提出新的距离度量公式:

$$d(w, h, c) = 1 - \text{IOU}(w, h, c), \quad (1)$$

式中:  $w, h$  为样本点的宽,高;  $c$  为聚类中心。YOLOV4 的 anchor box 是在(1)式的基础上对 80 类目标进行聚类求得的,而对于所提人员检测模块,只检测人员这一类别且该类别具有高度大于宽度这一固有属性,若使用默认的 anchor box 会对训练结果产生较大影响。

为解决此问题,在(1)式基础上采用 K-means++ 算法,根据人员检测数据集聚类生成符合样本特点的 anchor box。K-means++ 算法随机选取一个聚类中心,使得初始选择的  $k$  个聚类中心之间的距离尽可能远,通过选择最优的聚类中心求得最佳的 anchor box,这种最佳的 anchor box 在一定程度上能提高定位的准确性,提高小目标的检测精度。anchor box 生成方法的流程如图 2 所示。

Input: the number and width and height of ground truth in the data set.

Output:  $k$  clustering centers with different sizes.

- 1) Obtain the quantity and width and height of ground truth. Assume that the quantity is  $n$  and the width and height are  $(w_i, h_i)$ , and satisfied  $i \in (1, 2, \dots, n)$ .
- 2) Randomly select 1 value from  $n$  as a clustering center, the clustering center is  $(w_j, h_j)$ , and satisfied  $j \in (1, 2, \dots, n)$ .
- 3) Calculate the distance between the remaining  $n-1$  values and the clustering center according to Eq. (1).
- 4) The sample with the largest distance is selected as the new clustering center by probability.
- 5) Repeat steps 2) and 3) several times until the optimal  $k$  clustering centers are selected.
- 6) Calculate the distance between  $k$  clustering centers and the remaining  $n-k$  values according to Eq. (1).
- 7) The  $k$  values with the smallest distance are re-used as the clustering center.
- 8) Repeat steps 6) and 7) until the clustering center no longer changes.

图 2 anchor box 生成方法流程

Fig. 2 Flowchart of anchor box generation method

### 3 人员重识别模块

矿区场景环境复杂,由于工作人员穿着、遮挡、光照等差异,极大降低了模型在真实场景下的泛化性和鲁棒性。为解决以上问题,针对矿区人员穿着问题,提出基于外观不变性的通道注意力特征提取网络(CA-NET),提高模型对人员外观变化的自适应性,实现矿区场景下人员身份的准确识别。针对矿区人员统一工作服的难点,提出基于权重约束的难样本采样损失函数(Trihardwc loss),保证模型获得更好的区分度和高性能。针对光照和遮挡问题,采用 Color jitter 和随机擦除(REA)<sup>[18]</sup>两种数据增强

策略,提高模型在实际场景下的泛化性和鲁棒性。最后根据矿区实际场景,构建更具有工程应用特点的 Miner-Market 人员重识别数据集。

#### 3.1 基于外观不变性的通道注意力特征提取网络

针对矿区场景下的人员重识别问题,提出基于外观不变性的通道注意力特征提取网络,模型的网络结构如图 3 所示。其中  $P \times K$  表示每一批次输入的训练数据量,Color jitter 和 random erasing 表示训练过程中采取的两种数据增强策略,Conv1 表示单独的卷积层,Conv2.x、Conv3.x、Conv4.x 和 Conv5.x 表示具有残差结构的残差单元,features 表示提取后的人员特征向量,FC layers 表示全连接层。CA-

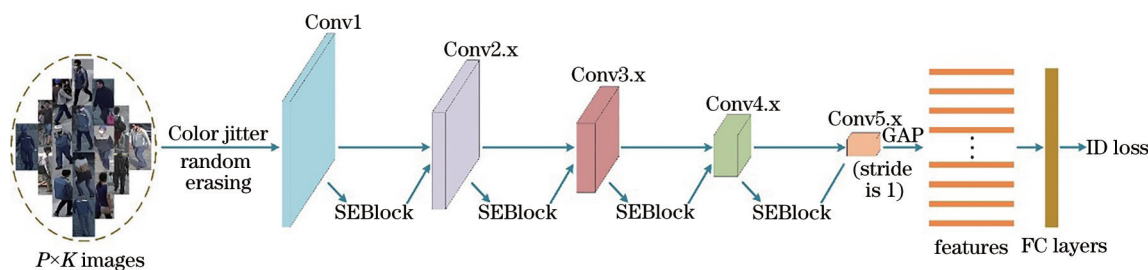


图 3 CA-NET 结构

Fig. 3 CA-NET structure

NET以ResNet50-IBN-a<sup>[19]</sup>为骨干网,同时在骨干网中加入通道注意力机制<sup>[20]</sup>,Hu等<sup>[20]</sup>在消融实验中证明,将注意力机制模块加入残差单元中能加强通道之间的相互依赖关系,取得更好的性能。因此在Conv2.x至Conv5.x四个残差单元中分别加入通道注意力模块,在图3中用SEBlock模块表示,通过显式建模通道来提高神经网络的表征力;其次,将Conv5.x中的stride改为1,通过增大特征图的尺寸来提取更加细粒度的特征,提升模型的性能,同时在Conv5.x后面增加全局平均池化层(GAP),将不同的输入转化为固定的输出,实现参数的自适应选择;最后,将全连接层的输出更改为训练集的人员身份数量,使用分类损失进行训练和优化。

### 3.2 基于权重约束的难样本采样损失函数

损失函数作为评价模型收敛好坏的函数,一定程度上决定了模型性能的高低,人员重识别中常用的损失函数除了分类损失外,还包括Triplet loss<sup>[12]</sup>和Trihard loss<sup>[21]</sup>。Triplet loss中,由于输入的三元组样本是随机挑选的,挑选出的很可能是简单样本,持续对简单样本进行学习训练会限制模型的泛化能力。因此Hermans等<sup>[21]</sup>在Triplet loss基础上引入难样本采样思想,提出的Trihard loss损失函数定义为

$$L_{\text{Thd}} = \frac{1}{P \times K} \sum_{a \in b_{\text{batch}}} \left( \max_{p \in A} d_{a,p} - \min_{n \in B} d_{a,n} + \alpha \right)_+, \quad (2)$$

式中:每一个batch都包含 $P \times K$ 张图片, $P$ 表示ID数量, $K$ 表示每一个ID下的图片数量;针对每一张图片 $a$ , $A$ 表示与 $a$ 的ID相同的图片集, $p$ 表示 $A$ 图片集中的某一张图片, $d_{a,p}$ 表示相同ID之间的距离度量, $B$ 表示与 $a$ 的ID不相同的图片集, $n$ 表示 $B$ 图片集中的某一张图片, $d_{a,n}$ 表示不同ID之间的距离

度量; $\alpha$ 通常设置为0.3。通过这种难样本采样方式,分别找出与 $a$ 特征距离最大的正样本对以及与 $a$ 特征距离最小的负样本对。Trihard loss通过训练最难的正样本对和最难的负样本对,改善了模型的性能,但是在训练过程中只考虑了正负样本对之间的相对距离,即正样本对特征距离的最大值和负样本对特征距离的最小值之间差值一定,导致损失结果相同,易忽略正样本对之间的绝对距离。

为解决此问题,提出的Trihardwc loss函数定义为

$$L_{\text{Thdwc}} = \frac{1}{P \times K} \sum_{a \in b_{\text{batch}}} \left( \max_{p \in A} d_{a,p} - \min_{n \in B} d_{a,n} + \alpha \right)_+ + \beta \times \frac{1}{P \times K} \sum_{a \in b_{\text{batch}}} \left( \frac{\max_{p \in A} d_{a,p}}{\min_{n \in B} d_{a,n}} \right), \quad (3)$$

式中: $\beta$ 表示权重系数。Trihardwc loss包含Trihard loss和权重约束项两部分,通过 $\beta$ 权重约束,在正样本对和负样本对距离度量之间差值一定时,正样本对特征距离的最大值越小,损失值就越小,正样本对的聚类性能就越好。

### 3.3 Color jitter和随机擦除

为解决实际矿区场景下光照变化的问题,CA-NET算法中引入Color jitter方法。Color jitter作为一种有效的数据扩充方式,能够随机改变输入图像的色彩,增加训练数据量以及训练数据特征的多样性,提高人员重识别模型在不同时段的泛化性。在CA-NET训练过程中,随机改变输入图像的亮度、对比度和饱和度,使得模型能够学习光照变化带来的差异,减小光照变化对模型性能的影响,如图4(a)所示。



图4 数据增强效果。(a) Color jitter;(b)随机擦除

Fig. 4 Data enhancement effect. (a) Color jitter; (b) random erasure

针对场景中的遮挡问题,采用随机擦除策略。随机擦除作为一种有效的数据增广方法,在图片中随机选择一个矩形区域,使用随机值擦除像素值,生成具有遮挡效果的训练图片,如图 4(b)所示。通过模拟人员遮挡来迫使网络利用局部未遮挡信息进行识别,降低了过拟合的风险,极大提高了模型的泛化性。

## 4 实验结果分析

### 4.1 实验环境和参数设置

本次实验使用 Intel(R) Xeon(R) Platinum 8163 4Core /2.5 GHz/33M 处理器,搭载 1 块 Tesla V100 显卡,基于 Ubuntu16.04 系统,使用 CUDA10.1、cudnn7 和 Pytorch 深度学习框架训练。人员检测训练阶段一般采用 SGD 优化器进行网络更新, batch size 设置为 64, epochs 设置为 454, 采用余弦学习率, 初始学习率设置为  $1 \times 10^{-2}$ , 输入尺寸设置为  $416 \times 416$ 。人员重识别训练阶段常采用

Adam 优化器进行网络更新, batch size 设置为 64, epochs 设置为 150, 采用线性学习率, 初始学习率设置为  $3 \times 10^{-4}$ 。

### 4.2 构建实验数据集

人员检测数据集包含 VOC2007、VOC2012 和 COCO2017, 首先将三者数据集中训练集中的人像类图片提取出来作为训练集, 其次将 VOC2007 和 COCO2017 数据集中测试集中的人像类图片提取出来作为测试集, 构建一个样本更大、更丰富的人员检测数据集, 增强矿区环境下的行人检测鲁棒性, 其中训练集图片数量为 70584, 测试集图片数量为 4790。

人员重识别数据集是在 Market1501 数据集<sup>[22]</sup>的基础上进行筛选和扩充的, 同时对矿区上的视频进行数据清洗和打标签分类, 构建更具有工程应用特点的 Miner-Market 人员重识别数据集, 部分示例如图 5 所示, 其中训练集包含 14447 张图片, 测试集包含 14359 张图片, 共有 1530 个不同的身份 ID。



图 5 Miner-Market 人员重识别数据集部分样例

Fig. 5 Some examples of Miner-Market personnel re-identification data set

### 4.3 人员检测实验结果分析

人员检测中常采用的评判标准有准确率 (Precision)、召回率 (Recall) 和平均精度 (AP)。为了综合衡量检测算法的性能, 以 Recall 为横坐标, Precision 为纵坐标的 P-R 曲线来反映准确率和召回率之间的关系, 同时 AP 值就是 P-R 曲线与坐标轴围成的面积, AP 值越大, 检测算法性能越好。针对人员检测数据集, 分别对 YOLOV4 算法与 MK-YOLOV4 算法进行性能比较, 对比结果如图 6 所示, 所提 MK-YOLOV4 检测算法的平均精度比 YOLOV4 高出 2.1%。

为了进一步验证 MK-YOLOV4 检测算法对小目标的检测性能, 对实际场景中的小目标进行检测, 检测对比结果如图 7 所示。从图 7 中可看出, MK-YOLOV4 对小目标检测的效果优于 YOLOV4 (黑白墙边的小目标人员没有被 YOLOV4 检测出)。

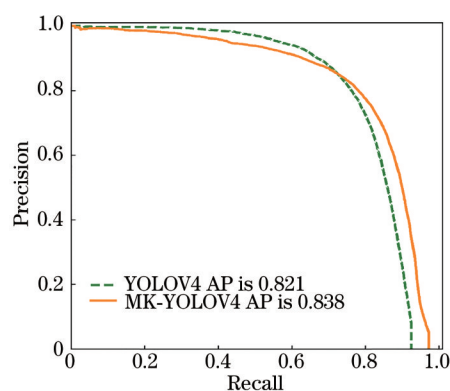


图 6 算法性能对比

Fig. 6 Algorithm performance comparison

### 4.4 人员重识别实验结果分析

#### 4.4.1 CA-NET 性能分析

人员重识别实验中常采用平均精度均值 (mAP) 和首次命中率 (Rank1) 作为评估标准。为了验证 CA-NET 的性能, 将 CA-NET 与主流的人员

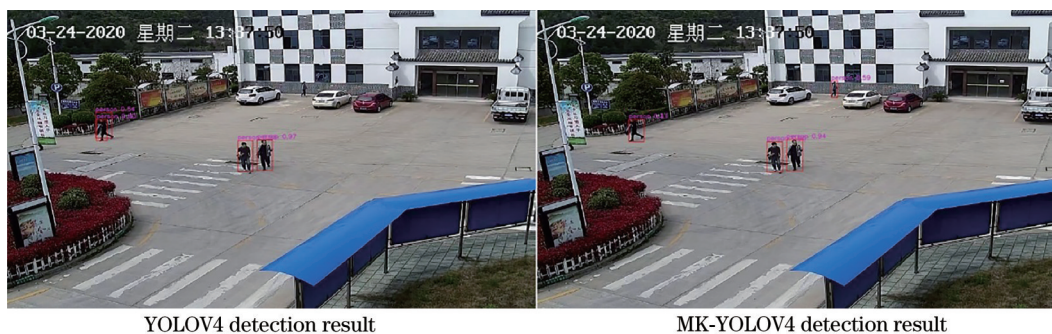


图 7 实际矿区场景检测结果

Fig. 7 Detection results in actual mining scene

重识别网络进行性能对比,对比结果如表 1 所示。由表 1 可以看出,无论是 mAP 还是 Rank1, CA-NET 均取得了最好的结果。

表 1 不同网络性能对比

Table 1 Comparison of different network performance

Backbone	mAP / %	Rank1 / %
ResNet50 <sup>[23]</sup>	65.6	82.7
SeResNet50 <sup>[20]</sup>	68.4	84.5
ResNet50-IBN-a <sup>[19]</sup>	67.0	83.9
CA-NET	69.0	85.6

#### 4.4.2 权重系数 $\beta$ 分析

在 4.4.1 节的基础上联合 Trihardwc loss 进行训练,同时对不同的权重系数进行实验,分析权重系数  $\beta$  的取值对算法性能的影响,对比结果如图 8 所示。

当  $\beta=0$  时, Trihardwc loss 相当于 Trihard loss, 当  $\beta$  取值为 0.1~1.0 时,表示具有不同权重系数的 Trihardwc loss。由图 8 可以看出,随着  $\beta$  的逐渐增大, mAP 和 Rank1 均呈现先上升后下降的趋势,且在  $\beta=0.1$  时, mAP 和 Rank1 均取得了最佳值。由此可以得出,当  $\beta=0.1$  时, Trihardwc loss 带来的性

能提升优于 Trihard loss。

#### 4.4.3 可视化结果分析

进一步探究 Color jitter 和随机擦除这两种数据增强策略对 Miner-Market 数据集的影响,实验结果如表 2 所示。由表 2 可以看出,所引入的 Color jitter 和随机擦除机制都一定程度上提高了模型的性能,增强了模型在真实场景下的鲁棒性和泛化性, Rank1 准确率也达到了 93.0%。为了进一步分析所提方法的优越性,结合 MK-YOLOV4 人员检测算法对视频中的矿区工作人员进行可视化检索展示,结果如图 9 所示。可以看出目标人员均识别准确,说明所提方法提高了矿区工作区域人员的识别精度,能够有效对目标人员进行准确识别和检索,对于矿区智能监控下的安全生产具有重要意义。

表 2 Miner-Market 数据集中不同模型性能分析

Table 2 Performance analysis of different models in Miner-Market data set

Model	mAP / %	Rank1 / %
CA-NET+Trihardwc loss+ Color jitter	78.6	91.2
CA-NET+Trihardwc loss+ Color jitter+REA	82.4	93.0

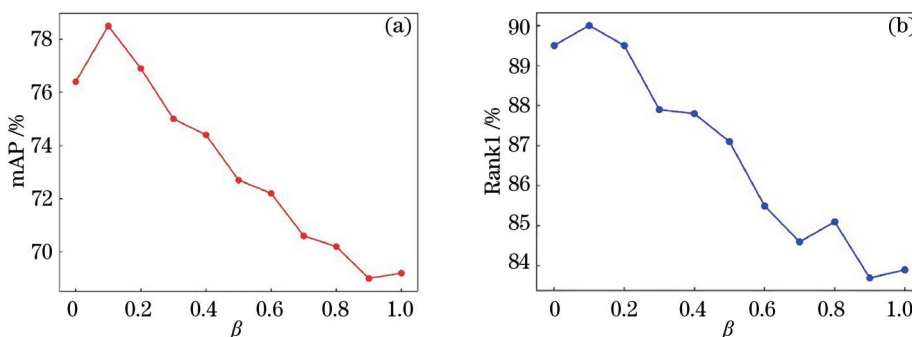


图 8 权重系数  $\beta$  分析。(a) mAP; (b) Rank1

Fig. 8 Weight coefficient  $\beta$  analysis. (a) mAP; (b) Rank1



图9 矿区工作人员重识别检索结果  
Fig. 9 Mining staff re-identification search results

## 5 结 论

针对矿区工作区域小目标人员检索问题,提出 MK-YOLOV4 无标注视频检索算法。MK-YOLOV4 算法在 YOLOV4 的基础上增加了多尺度预测数量,并结合人员高度大于宽度这一固有属性,使用 K-means++ 算法聚类生成符合样本特点的 anchor box,极大提高了对矿区工作区域小目标人员的检测精度。针对矿区工作区域人员穿着变化及统一工作服等问题,提出基于外观不变性的通道注意力特征提取网络和基于权重约束的难样本采样损失函数,并结合 Color jitter 和随机擦除两种数据增强策略,增强模型在实际场景下的泛化性和鲁棒性,实现矿区工作区域人员的准确识别。最后,构建更符合矿区场景特点的 Miner-Market 矿区人员重识别数据集,解决矿区场景下现有训练数据集类别少、场景样本单一等问题。

### 参 考 文 献

[1] Zheng L, Zhang H H, Sun S Y, et al. Person re-identification in the wild[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition

(CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 3346-3355.  
 [2] Xiao T, Li S, Wang B C, et al. Joint detection and identification feature learning for person search[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 3376-3385.  
 [3] Liu H, Feng J S, Jie Z Q, et al. Neural person search machines[C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice. New York: IEEE Press, 2017: 493-501.  
 [4] Zhao Y Q. Multi-level denoising and enhancement method based on wavelet transform for mine monitoring [J]. International Journal of Mining Science and Technology, 2013, 23(1): 163-166.  
 [5] Chai Y, Liu X L, Deng L J. An enhancement method for non-uniform illumination in coal mine [C]//2016 International Symposium on Computer, Consumer and Control (IS3C), July 4-6, 2016, Xi'an, China. New York: IEEE Press, 2016: 144-147.  
 [6] Cui P X, Yu F Q. Improved ViBe moving object



- detection algorithm to eliminate ghost and shadow[J]. *Laser & Optoelectronics Progress*, 2020, 57(16): 161006.
- 崔鹏翔, 于凤芹. 消除鬼影及阴影的改进 ViBe 运动目标检测算法[J]. *激光与光电子学进展*, 2020, 57(16): 161006.
- [7] Zhou B, Li R X, Shang Z H, et al. Object detection algorithm based on improved faster R-CNN[J]. *Laser & Optoelectronics Progress*, 2020, 57(10): 101009.
- 周兵, 李润鑫, 尚振宏, 等. 基于改进的 Faster R-CNN 目标检测算法[J]. *激光与光电子学进展*, 2020, 57(10): 101009.
- [8] Ren Z J, Lin S Z, Li D W, et al. Mask R-CNN object detection method based on improved feature pyramid[J]. *Laser & Optoelectronics Progress*, 2019, 56(4): 041502.
- 任之俊, 蔺素珍, 李大威, 等. 基于改进特征金字塔的 Mask R-CNN 目标检测方法[J]. *激光与光电子学进展*, 2019, 56(4): 041502.
- [9] Chen L L, Zhang Z D, Peng L. Real-time detection based on improved single shot MultiBox detector[J]. *Laser & Optoelectronics Progress*, 2019, 56(1): 011002.
- 陈立里, 张正道, 彭力. 基于改进 SSD 的实时检测方法[J]. *激光与光电子学进展*, 2019, 56(1): 011002.
- [10] Bochkovskiy A, Wang C Y, Liao H Y M. Yolov4: optimal speed and accuracy of object detection[EB/OL]. (2020-04-23)[2021-02-28]. <https://arxiv.org/abs/2004.10934>.
- [11] Chen Y B, Zhu X T, Gong S G. Person re-identification by deep learning multi-scale representations [C]//2017 IEEE International Conference on Computer Vision Workshops (ICCVW), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 2590-2600.
- [12] Schroff F, Kalenichenko D, Philbin J. FaceNet: a unified embedding for face recognition and clustering [C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 815-823.
- [13] Arthur D, Vassilvitskii S. K-means++ : the advantages of careful seeding[C]//Proceedings of the 2007 Annual ACM-SIAM Symposium on Discrete Algorithms, January 7-9, 2007, New Orleans, Louisiana, USA. New York: ACM, 2007: 1027-1035.
- [14] 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.
- [15] 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), June 14-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 390-391.
- [16] Liu S, Qi L, Qin H F, et al. Path aggregation network for instance segmentation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 8759-8768.
- [17] 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. New York: IEEE Press, 2017: 7263-7271.
- [18] Zhong Z, Zheng L, Kang G L, et al. Random erasing data augmentation[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, 34(7): 13001-13008.
- [19] Pan X G, Luo P, Shi J P, et al. Two at once: enhancing learning and generalization capacities via IBN-net[M]//Ferrari V, Hebert M, Sminchisescu C, et al. *Computer vision-ECCV 2018. Lecture notes in computer science*. Cham: Springer, 2018, 11208: 484-500.
- [20] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 7132-7141.
- [21] Hermans A, Beyer L, Leibe B. In defense of the triplet loss for person re-identification[EB/OL]. (2017-03-22)[2021-02-28]. <https://arxiv.org/abs/1703.07737>.
- [22] Zheng L, Shen L Y, Tian L, et al. Scalable person re-identification: a benchmark[C]//2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 1116-1124.
- [23] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.