

# 基于 YOLO 改进算法的安全帽和口罩佩戴自动同时检测

孙世丹<sup>1</sup>, 郑佳春<sup>1</sup>, 赵世佳<sup>1</sup>, 黄一琦<sup>2</sup>

(1. 集美大学海洋信息工程学院, 福建 厦门 361021; 2. 集美大学航海学院, 福建 厦门 361021)

**[摘要]** 针对工地、危险区域等场景需要实现同时佩戴安全帽与口罩的自动检测问题, 提出一种改进的 YOLOv3 算法以提高同时检测安全帽和口罩佩戴的准确率。首先, 对网络模型中的聚类算法进行优化, 使用加权核 K-means 聚类算法对训练数据集聚类分析, 选取更适合小目标检测的 Anchor Box, 以提高检测的平均精度和速度; 然后, 优化 YOLO 网络内部的 Darknet 特征网络层, 将 4 倍降采样提取的特征图进行 2 倍上采样, 再与 2 倍降采样进行卷积融合, 与 4 倍降采样、8 倍降采样以及 16 倍降采样一同输送到后续网络中, 来达到降低小目标的漏检概率。实验结果表明: 改进后的算法同时检测安全帽和口罩佩戴的平均准确率比原算法提高了 11.3%。

**[关键词]** 同时检测; YOLOv3 算法; K-means 聚类; 安全帽佩戴检测; 口罩佩戴检测

**[中图分类号]** TP 312

## Automatic Simultaneous Detection of Helmet and Mask Wearing Based on Improved YOLO Algorithm

SUN Shidan<sup>1</sup>, ZHENG Jiachun<sup>1</sup>, ZHAO Shijia<sup>1</sup>, HUANG Yiqi<sup>2</sup>

(1. School of Marine Information Engineering, Jimei University, Xiamen 361021, China;

2. Navigation Institute, Jimei University, Xiamen 361021, China)

**Abstract:** Aiming at the problem of automatic detection of wearing helmets and masks at the same time in the scene of construction site and dangerous area, an improved YOLOv3 algorithm is proposed to enhance the detection accuracy of helmets and masks. Firstly, the clustering algorithm in network is optimized. The weighted kernel K-means clustering algorithm is used to analyze the dataset, so as to select the anchor box more suitable for small targets detection and improve the average accuracy and speed of detection. Secondly, it optimizes the Darknet characteristic network layer in YOLO network. The extracted quadruple down sampling feature map is up sampling once. The double up sampling is fused with the previous double down sampling, and then it is transmitted to the subsequent network together with quadruple down sampling, eightfold down sampling and sixteen times down sampling to reduce the miss detection rate of small targets. Experimental results show that, the average detection accuracy of the improved algorithm is improved by 11.3% when the helmets and masks are worn at the same time.

**Keywords:** simultaneous determination; YOLO v3 algorithm; K-means clustering algorithm; helmet wearing test; mask wear detection

[收稿日期] 2020-11-22

[作者简介] 孙世丹(1995—), 女, 硕士生, 从事信号与信息处理研究。通信作者: 郑佳春(1965—), 男, 教授, 从事通信工程研究。E-mail: jchzheng@jmu.edu.cn

http://xuebaobangong.jmu.edu.cn/zkb

## 0 引言

新冠肺炎疫情,让人们认识到佩戴口罩的重要性;而在工地、危险区域进行作业活动,佩戴安全帽是安全规范要求。因此,为了确保安全,同时检测安全帽和口罩佩戴的问题应运而生。目前,针对口罩佩戴的检测主要采用人脸检测算法,例如,牛作东<sup>[1]</sup>等采用改进 retinaface 算法,通过人脸关键的三位分析,提高了口罩佩戴的检测速度。由于复杂场景中目标遮挡、人群密集、小尺度等问题,造成使用目标检测算法检测是否佩戴口罩的效果并不理想。对安全帽佩戴的检测研究主要采用基于深度学习的目标检测模型,通过训练和优化各种深度学习模型算法<sup>[2-4]</sup>来实现图像特征提取,获得安全帽的颜色、形状及佩戴是否正确等信息。

Girshick 等提出的区域卷积神经网络 (R-CNN)<sup>[5]</sup>、快速区域卷积神经网络 (Fast R-CNN)<sup>[6]</sup>和超快区域卷积神经网络 (Faster R-CNN)<sup>[7]</sup>大大提高了检测速度;Liu 等<sup>[8]</sup>提出的端到端多尺度检测算法 SSD (Single Shot MultiBox Detector),在检测精度和检测时间上取得了良好的效果;Redmon 等<sup>[9-11]</sup>提出了 YOLOv1、v2、v3 检测算法。YOLOv3 由于融合了不同的特征尺度,可进行分别检测,加上在检测效率、精确性的优势,在工程应用上更适合用于多种不同小目标的同时检测<sup>[12-13]</sup>。YOLOv3 是在 YOLOv2 算法基础上进行一系列改进而成的:YOLOv3 采用 Darknet-53 网络进行特征提取,利用一系列的卷积和上采样对提取的特征进行解析并标记输出;加入了 FPN 网络<sup>[14]</sup>中的多尺度网络架构,融合了不同尺度的特征图以提高预测效果;通过 K-means<sup>[15]</sup>对数据集标签进行聚类分析得到默认框的尺寸;将 ResNET<sup>[16]</sup>架构融合进 Darknet-19 中改变其网络的背景架构;采取平均交并比作为度量标签相似性的指标;每一个卷积层后都会有批量归一化操作<sup>[17]</sup>和去除 dropout 操作,防止出现过拟合现象;使用全卷积结构,在每个比例尺上,输出层后都会跟随着  $3 \times 3$  和  $1 \times 1$  的卷积层,以简化损失函数的计算,提高检测速度;采用上采样和融合做法将三个不同尺度的特征图融合然后对目标进行检测,提高算法对小目标检测的精确度。

综上所述,YOLOv3 是适合用于安全帽、口罩等尺寸中偏小物体的检测与识别的,但是,还需要进一步优化。因而,针对安全帽、口罩佩戴的同时检测问题以及实际情况,本文提出一种以 YOLOv3 算法为基础,以 YOLO 算法为 one-stage 网络架构的目标检测算法。由于其不需要先检测待测物体的位置而是直接进行特征检测,大大提高了检测速度。

## 1 YOLOv3 算法优化

为了提升安全帽、口罩佩戴的同时检测性能,本文对 YOLOv3 算法做了聚类算法优化和 Darknet 特征网络层优化。

### 1.1 聚类算法优化

YOLOv3 运用 K-means 聚类算法<sup>[15]</sup>对输入端的数据集进行处理,以达到优化网络内部初始锚框选取的目的。对数据集通过目标框的大小进行分类,得到 9 个先验框的大小。而先验框的大小与其尺度有关,尺度越大时得到的先验框越小,因而可以以此来获得更细致的目标边缘信息。聚类算法本质是计算每个样本点簇中心的距离。选取距离各点的均值最小的点作为新的点簇中心,之后依照新的点簇中心作为判断依据,将各点分配到距离最近的点簇中的一类,依次迭代循环直到点簇中心稳定收敛,收敛结果即是最后的聚类效果。核方法 (kernel methods)<sup>[18]</sup>则是将输入空间映射到高维的特征空间上来划分类别。将二维平面中的样本特征映射到高维的特征空间,使得在平面上无法分类或无法正确分类的特征变得线性可分,以此来区分不同类别的样本中心,等分类完成后映射回二维平面上。加权核 K-means 算法则是在核函数的基础上对不同占比的类别进行权重的分配,选取更加合适的锚框。

这种映射是非线性变换的,有利于对未知的目标框的检测。定义加权核 K-means 的目标函数为:

$$J(v) = \sum_{i=1}^k \sum_{a_i \in \pi_k} w(i) \|\varphi(a_i) - c_k\|^2。其中: i, a_i 表示输入空间中聚类的像素点; w(i) 表示对应像素点$$

的权重;  $\pi_k$  表示第  $k$  个聚类;  $c_k$  为每个子类的聚类中心;  $\varphi(a_i)$  为非线性的映射核距离函数。

非线性核函数的计算相对困难。从数学的角度出发, 存在一个低维空间中的函数  $K(x, x')$ , 当  $K(x, x') = \langle \varphi(x) \cdot \varphi(x') \rangle$  时, 它恰好等于在高维空间中这个内积。通过计算样本点投射到高维空间的值, 再进行内积运算, 得出非线性核函数的解。通过求出  $K(x, x')$  的内积函数, 得到高维空间中的样本点以及点簇中心的距离, 从而大大减少了计算量。经过  $K$  函数化简后的距离公式为:  $c_k = \sum_{a_i \in \pi_k} w(i) \varphi(a_i) / \sum_{a_i \in \pi_k} w(i)$ 。经过运算之后的聚类中心为:  $\|\varphi(a_i) - c_k\|^2 = \|\varphi(a_i) \cdot \varphi(a_i) - 2 \sum_{a_j \in \pi_k} \varphi(a_i) \cdot \varphi(a_j) / |\pi_k| + \sum_{a_j, a_l \in \pi_k} \varphi(a_j) \cdot \varphi(a_l) / |\pi_k|^2\| = K(x, x') - 2 \sum_{a_j \in \pi_k} \varphi(a_i) \cdot \varphi(a_j) / |\pi_k| + \sum_{a_j, a_l \in \pi_k} \varphi(a_j) \cdot \varphi(a_l) / |\pi_k|^2$ ,  $c_k = \sum_{a_i \in \pi_k} w(i) \varphi(a_i) / \sum_{a_i \in \pi_k} w(i)$ 。

YOLOv3 通常将公开数据集 coco 所训练得到的锚框的尺寸作为默认框, 其大小为 (10, 13), (16, 30), (33, 23), (30, 61), (62, 45), (59, 119), (116, 90), (156, 198), (373, 326)。在实际的安全帽、口罩佩戴的检测中, 由于安全帽的目标尺寸与 coco 数据集的差异较大, 使用 YOLOv3 默认框的检验效果并不好。而使用改进的聚类算法对自制安全帽数据集进行分析后, 得到了 9 组先验框的大小: (5, 7), (10, 13), (12, 15), (18, 23), (24, 30), (33, 50), (57, 72), (113, 127), (221, 201), 再按照先验框的尺度分配规律分别分配给 3 个尺度的特征图。经实验验证, 优化后的 K-means 算法的检测效果提升明显。

## 1.2 Darknet 特征网络层优化

YOLOv3 将 Darknet53 作为骨架网络对输入图片进行特征提取, 提取后的特征图经过一系列的解析操作得到最后的多尺度输出。其借鉴了 FPN 网络中多尺度预测的方法来提高检测结果。输入目标在 Darknet53 网络中经过 5 次降采样, 分别以 8 倍、16 倍、32 倍的降采样结果传输, 再经过上采样和张量拼接出的特征图最终被输出检测。这样的网络对于小目标的检测能力是不足的, 当检测目标小于  $8 \times 8$  像素时, 特征网络将会很难检测出来。为了提高网络对于头盔、口罩这类小目标物体的检测率, 将 4 倍降采样的特征信息进行一次上采样, 之后与上一层的 2 倍下采样信息进行特征融合, 得到融合后的 2 倍下采样融合图作为  $52 \times 52$  的尺度特征图检测, 融合图与之后的降采样信息进行卷积, 将融合之后的 4 倍、8 倍、16 倍降采样一同输送到后续网络中, 以提高小目标的检测特征, 进而增大小目标特征的占比, 最终增加小目标的检测率。

## 2 实验与分析

### 2.1 创建数据集

由于缺少公开的安全帽和口罩佩戴的数据集, 研究团队通过整合各种资源, 自建一份安全帽和口罩佩戴的数据集, 共 10 664 张图。其中: 戴有安全帽的图 4 604 张, 主要为施工工地工人佩戴环境; 戴有口罩的图 2 023 张, 主要为口罩佩戴展示及公共场所行人佩戴环境; 同时戴有安全帽和口罩的对比图 2 782 张, 主要为疫情期间工人佩戴; 未佩戴两者的对比图 1 255 张。采用平均准确率 (mean average precision, mAP) 作为评价指标, 与原网络进行比较实验。

### 2.2 实验环境与实验计划

本实验环境配置在 UBUNTU18.04 操作系统、TITAN RTX 2080GPU、CUDA10.1、CUDNN、128GB 内存平台上, 采用 pytorch1.2.0 框架进行训练。为了验证算法的有效性, 本研究进行了以下三种不同情况的实验。

1) 实验一 单检测安全帽佩戴的实验。以建筑工地为实验场景, 用安全帽佩戴数据集进行网络训练和检测实验。每批次随机选取 8 张图, 初始学习率为  $5 \times 10^{-4}$ , 且逐步递减, IOU 置为 0.5。采用反向传播对网络参数进行微调。将数据集划分为训练集和验证集分别进行评估。共 100 个 epoch 对网

络参数进行优化。

2) 实验二 单检测口罩佩戴的实验。以各大公共区域为实验场景, 每批次随机选取 8 张图, 初始学习率为  $5 \times 10^{-4}$ , 且逐步递减, IOU 置为 0.5。采用反向传播对网络参数进行微调。将数据集划分为训练集和验证集分别进行评估。共 100 个 epoch 对网络参数进行优化。

3) 实验三 同时检测安全帽和口罩佩戴的实验。在安全帽佩戴数据集和口罩佩戴数据集上添加工地上同时戴有安全帽和口罩的图 2 782 张, 进行网络训练, 同时检测安全帽和口罩的佩戴情况。训练批大小选取 64 张图, 验证批大小选取 2 张图, 初始学习率为  $1 \times 10^{-4}$ , 且逐步递减, 但不小于  $1 \times 10^{-6}$ , IOU 置为 0.5。采用反向传播对网络参数进行微调。将数据集划分为训练集和验证集分别进行评估。前 20 个 epoch 先对最后一层网络参数进行优化, 后 30 个 epoch 对整个网络的网络参数进行调整。

### 2.3 实验结果分析

本研究针对两大类目标进行检测, 为区分优化前后对比效果, 分别将单戴有口罩或安全帽的图进行了优化前后的检测对比, 以及同时戴有安全帽和口罩的图进行优化前后的检测对比。为评估改进后的算法模型的性能, 将各检测结果的准确率、平均准确率 (mAP) 作为模型性能的评估指标, 并与 YOLOv3 算法进行对比。

#### 2.3.1 实验一结果分析

图 1 展示了工地上建筑工人戴有安全帽的图优化前后的对比效果, 改进后安全帽佩戴的单检测效果明显高于原 YOLOv3 算法。算法优化前戴有安全帽的图的检测准确率为 87.4%, 未戴安全帽的图的检测准确率为 84.5%, 平均准确率为 85.9%; 优化后分别为 97.6%, 89.3%, 93.5%。改进后的 YOLOv3 对单检测安全帽佩戴的平均准确率提高了 7.6%。



a) 优化前安全帽检测效果  
Safety helmet detection effect before optimization  
b) 优化后安全帽检测效果  
Safety helmet detection effect after optimization

图 1 优化前后的单检测安全帽佩戴结果对比

Fig.1 Comparison of safety helmet detection results before and after optimization

#### 2.3.2 实验二结果分析

图 2 展示了车站场景下载有口罩的图优化前后的对比效果。算法优化前戴有口罩的图的检测准确率为 82.4%, 未佩戴口罩的图的检测准确率为 80.2%, 平均准确率为 81.3%; 优化后分别为 93.6%, 87.0%, 90.3%。改进后的 YOLOv3 对单检测口罩佩戴的平均准确率提高了 9%。

#### 2.3.3 实验三结果分析

图 3 显示出多人同时佩戴安全帽和口罩场景下的实验效果。为了方便查看, 这里省去了口罩检测的准确率, 只标识出了预测框。可以看出, 用本算法优化后同时检测安全帽和口罩佩戴的准确率均有所提高。

由表 1 可见, 同时检测安全帽和口罩佩戴的准确率, 改进后的算法相较于原算法均提升了 15.3%, 而平均准确率则提高了 11.3%。而相对于 SSD 算法, 改进后的 YOLO 算法对于同时检测安全帽和口罩佩戴的准确率更高, 速度也相比于 SSD 算法快了  $5 f \cdot s^{-1}$ 。



a) 优化前检测效果  
Detection effect before optimization



b) 优化后检测效果  
Detection effect after optimization

图2 优化前后的单检测口罩佩戴结果对比

Fig.2 Comparison of mask detection results before and after optimization



a) 优化前检测效果  
Detection effect before optimization



b) 优化后检测效果  
Detection effect after optimization

图3 优化前后的同时检测多人口罩和安全帽佩戴的结果对比

Fig.3 Comparison of multi person detection results before and after optimization

表1 YOLOv3 优化前后检测准确度

Tab.1 Detection accuracy before and after YOLOv3 optimization

算法 Algorithm	准确率 Accuracy/%			平均准确率 Mean average precision/%	帧速率 Frams per second /(f · s <sup>-1</sup> )
	安全帽检测 Safety helmet detection	口罩检测 Mask detection	未佩戴 Not wearing		
SSD	84.4	86.6	83.3	84.7	40
YOLOv3	80.4	74.3	72.0	75.6	48
改进的 YOLOv3 Improved YOLOv3	95.7	89.6	79.3	86.9	45

## 4 结束语

本文研究了基于 YOLO 同时检测安全帽和口罩佩戴的算法改进。该算法通过优化 K-means 聚类算法来选取更合适的锚框以适应数据集, 优化 Darknet 网络内部结构以及一些网络微调, 多次提取小目标的特征来增强其特征提取的效果。实验分析结果表明: 通过优化后该算法能够满足在实际场景中安全帽、口罩佩戴的同时检测需求, 可以应用于智慧工地、港口码头、危化矿区等安全监管, 具有一定的应用与市场价值。接下来, 会将算法移植到实际系统中实验应用并开展算法升级研究, 以进一步提升佩戴安全帽和口罩的检测性能。

### [ 参考文献 ]

[1] 牛作东, 覃涛, 李捍东, 等. 改进 RetinaFace 的自然场景口罩佩戴检测算法 [J]. 计算机工程与应用, 2020, 56(12): 1-7.

- [2] LONG X, CUI W, ZHENG Z. Safety helmet wearing detection based on deep learning [C] //3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC). Chengdu: IEEE, 2019: 2495-2499. DOI: 10.1109/ITNEC.2019.8729039.
- [3] DASGUPTA M, BANDYOPADHYAY O, CHATTERJI S. Automated helmet detection for multiple motorcycle riders using CNN [C] //2019 IEEE Conference on Information and Communication Technology. Allahabad: IEEE, 2019: 1-4. DOI:10.1109/CICT48419.2019.9066191.
- [4] CHEN S, TANG W, JI T, et al. Detection of safety helmet wearing based on improved faster R-CNN [C] //2020 International Joint Conference on Neural Networks (IJCNN). Glasgow: IEEE, 2020: 1-7. DOI: 10.1109/IJCNN48605.2020.9207574.
- [5] GIRSHICK R, DONAHUE J, DARRELL T, et al. Region-based convolutional networks for accurate object detection and segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 38(1): 142-158. DOI:10.1109/TPAMI.2015.2437384.
- [6] GIRSHICK R. Fast R-CNN [C] //2015 IEEE International Conference on Computer Vision (ICCV). Santiago: IEEE, 2015: 1440-1448. DOI:10.1109/ICCV.2015.169.
- [7] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 39(6): 1137-1149. DOI:10.1109/TPAMI.2016.2577031.
- [8] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector [C] //2016 European Conference on Computer Vision. Cham: Springer, 2016: 21-37. DOI:10.1007/978-3-319-46448-0\_2.
- [9] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C] //2016 Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 779-788. DOI:10.1109/CVPR.2016.91.
- [10] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C] //2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 6517-6525. DOI:10.1109/CVPR.2017.690.
- [11] REDMON J, FARHADI A. YOLOv3: an incremental improvement [J]. 2018 IEEE Conference on Computer Vision and Pattern Recognition, 2018: 89-95. DOI:10.1109/CVPR.2017.690.
- [12] WU F, JIN G, GAO M, et al. Helmet detection based on improved YOLO V3 deep model [C] //16th International Conference on Networking, Sensing and Control (ICNSC). Banff: IEEE, 2019: 363-368. DOI: 10.1109/ICNSC.2019.8743246.
- [13] LIN F, ZHENG X, WU Q. Small object detection in aerial view based on improved YoloV3 neural network [C] //2020 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA). Dalian: IEEE, 2020: 522-525. DOI:10.1109/AEECA49918.2020.9213538.
- [14] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C] //2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu: IEEE, 2017: 936-944. DOI: 10.1109/CVPR.2017.106.
- [15] SULAIMAN S N, ISA N A M. Adaptive fuzzy-k-means clustering algorithm for image segmentation [J]. IEEE Transactions on Consumer Electronics, 2010, 56 (4): 2661-2668. DOI:10.1109/TCE.2010.5681154.
- [16] HE K, ZHANG X Y, REN S, et al. Deep residual learning for image recognition [C] //2016 IEEE Conference on Computer Vision & Pattern Recognition. Las Vegas: IEEE, Computer Society, 2016:770-778. DOI:10.1109/CVPR.2016.90.
- [17] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift [J]. JMLR.org, 2015: 448-456.
- [18] GENG F Z, QIAN S P. An optimal reproducing kernel method for linear nonlocal boundary value problems [J]. Applied Mathematics Letters, 2018, 77: 49-56. DOI:10.1016/j.aml.2017.10.002.

(责任编辑 朱雪莲 英文审校 黄振坤)