计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2019,28(9):174−179 [doi: 10.15888/j.cnki.csa.007065]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于 YOLO 的安全帽检测方法①

林　俊[1], 党伟超[1], 潘理虎[1,2], 白尚旺[1], 张　睿[1]

[1](太原科技大学 计算机科学与技术学院, 太原 030024)
[2](中国科学院 地理科学与资源研究所, 北京 100101)
通讯作者: 党伟超, E-mail: dangweichao@tyust.edu.cn

摘　要: 安全帽作为作业工人最基本的个体防护装备, 对作业人员的生命安全具有重要意义. 但是部分作业人员安全意识缺乏, 不佩戴安全帽行为时常发生. 本文聚焦于复杂场景下对工作人员是否佩戴安全帽的实时检测. YOLO (You Only Look Once) 是当前最为先进的实时目标检测算法, 在检测精度和速度上都取得了良好的效果, 将 YOLO 应用于安全帽检测. 针对未佩戴安全帽单类检测问题, 修改分类器, 将输出修改为 18 维度的张量. 基于 YOLOv3 在 ImageNet 上的预训练模型, 对实际场景下采集到的 2010 张样本进行标注并训练, 根据损失函数和 IOU 曲线对模型进行优化调参, 最终得到最优的安全帽检测模型. 实验结果表明, 在 2000 张图片测试集上取得了 98.7% 的准确率, 在无 GPU 环境下平均检测速度达到了 35 fps, 满足实时性的检测要求, 验证了基于 YOLOv3 安全帽检测方法的有效性.

关键词: 安全帽检测; 卷积神经网络; 目标检测; YOLO; 实时检测

引用格式: 林俊,党伟超,潘理虎,白尚旺,张睿.基于 YOLO 的安全帽检测方法.计算机系统应用,2019,28(9):174–179. http://www.c-s-a.org.cn/1003-3254/7065.html

Safety Helmet Detection Based on YOLO

LIN Jun[1], DANG Wei-Chao[1], PAN Li-Hu[1,2], BAI Shang-Wang[1], ZHANG Rui[1]

[1](School of Computer Science and Technology, Taiyuan University of Science and Technology, Taiyuan 030024, China)
[2](Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China)

Abstract: As the most basic personal protective equipment, helmets are of great significance to the safety for workers. However, some workers lack safety awareness and often do not wear safety helmets. This study focuses on the detection of safety helmet in complex background. You Only Look Once (YOLO) is a state-of-the-art, real-time object detection algorithm, we propose to apply the YOLO detector for safety helmets detection, which achieves high accuracy. For the single-type detection problem without wearing a helmet, the classifier is modified and the output is modified to a tensor of 18 dimensions. We train YOLOv3 for safety helmets detection on the 2010 datasets based on the pre-training model in ImageNet. Then we optimize the model according to the loss function and IOU curve. The experimental results show that the safety helmet detector gets 98.7% accuracy and 35 fps on the 2000 detection test sets without GPU, which meets the real-time detection requirements. The effectiveness of the YOLOv3 safety helmet detection method is verified.

Key words: safety helmet detection; convolution neural network; object detection; YOLO; real-time detection

① 基金项目: 山西省中科院科技合作项目 (20141101001); 山西省重点研发计划 (一般) 工业项目 (201703D121042-1); 山西省社会发展科技项目 (20140313020-1); 山西省应用基础研究项目 (201801D221179)
Foundation item: Science and Technology Collaborative Program Between Shanxi Province and Chinese Academy of Sciences (20141101001); Key Research and Development Program for General Industry of Shanxi Province (201703D121042-1); Science and Technology Program for Social Development of Shanxi Province (20140313020-1); Applied Basic Research Projectof Shanxi Province (201801D221179)
收稿时间: 2019-03-01; 修改时间: 2019-03-29; 采用时间: 2019-04-01; csa 在线出版时间: 2019-09-05

图像视频中的场景目标物体检测已经成为当前人工智能、计算机视觉领域的一个研究热点[1, 2]. 而生产安全问题一直是一个社会关注度极高的问题, 每年近百万起安全事故给社会和家庭带来巨大的压力. 根据相关报告显示, 95% 的安全事故是由于工人的违规违章造成的. 安全帽作为作业工人最基本的个体防护装备, 对工作人员的生命安全具有重要意义. 但是, 部分操作人员安全意识缺乏, 不佩戴安全帽行为时有发生. 安全帽检测已经成为构建生产安全视频监控的一项重要技术, 在煤矿、变电站、建筑工地等实际场景中需求广泛.

目标检测是指找出输入图像中的目标物体, 包含物体分类和物体定位两个子任务, 判断物体的类别和位置. 传统的目标检测方法如帧差法[3]、光流法[4]、背景差分法[5]、viola-Jones 检测器[6]、HOG 检测器[7]、可变性部件模型 (Deformable Part based Model, DMP)[8]等. 这些方法在检测时主要分为三个步骤: 第一步获取前景目标信息或者采用滑动窗口对图像中的每一个尺度和像素进行遍历, 第二步进行特征提取, 第三步利用提取到的特征建立数学模型或者利用分类器 (如 SVM[9]、AdaBoost[10]) 进行分类得到目标检测结果. 传统的检测方法在特定的场景下可以取得良好效果, 但在开放环境下, 如角度变换、光照不足、天气变化等, 其准确性难以得到保证, 且泛化能力差. 除此之外, 基于传统的手工特征设计依赖大量的先验知识, 主观性强, 分三步走的检测过程繁琐、计算开销大, 在一些要求实时检测的场景, 往往具有挑战性.

近年来人工智能快速发展, 计算机视觉作为人工智能的一个重要研究方向, 也迎来了第三次热潮[11]. 目标检测作为计算机视觉领域的一个研究热点, 大量的基于卷积神经网络的优秀目标检测算法取得了巨大的成功[12], 激励着越来越多的学者开始致力于深度学习目标检测算法的研究. YOLO[13](You Only Look Once) 是由 Joseph Redmon 等人最早在 2016 年 CVPR 上提出的基于卷积神经网络、快速、高效、开放的目标检测算法, 截止 2018 年, 已有 3 个改进的版本: YOLO, YOLO9000[14], 和 YOLOv3[15]. YOLO9000 在 YOLO 的基础上进行优化和改进, 加入了批标准化层 (Batch Normalization, BN) 和类 Anchor 机制, 在保证实时性的前提下, 准确率有了较大的提升, 可以检测 9000 类目标. YOLOv3 在 YOLO 和 YOLO9000 的基础上进行改进, 采用特征融合和多尺度预测, 在检测速度和检测精度上都达到了最先进的水平.

本文首先根据是否佩戴安全帽单类检测, 修改分类器, 将输出修改为 18 维度的张量. 之后采用 YOLOv3 在 ImageNet 上的预训练模型, 在此基础对实际场景下采集到的 2010 张数据样本进行标注并训练, 根据损失函数曲线和 IOU 曲线对模型进行优化, 得到最优的安全帽检测模型. 基于 YOLOv3 的安全帽检测方法能够自主学习目标特征, 减少手工特征设计人为因素干扰, 具有较高的准确率, 对复杂场景下的不同颜色、不同角度的安全帽检测展现出很好的泛化能力和健壮性.

1 安全帽检测

基于深度学习的目标检测算法主要分为两类: 一类是以 RCNN[16–18]系列算法为代表的、“两步走”的基于候选区域的目标检测算法, 一类是以 YOLO、SSD[19]为代表的、“一步走”的基于回归的目标检测算法. 基于候选区域的目标检测算法从理论上来讲比基于回归的目标检测算法精准度更高, 以 Faster-RCNN 为代表, 基于候选区域的目标检测算法由卷积层 (convolution layers)、区域候选网络 (Region Proposal Networks, RPN)、感兴趣区域池化层 (ROI Pooling)、分类层 (Classification) 四部分组成. 卷积层由一组基础的卷积层、激活层和池化层组成, 用来提取特征, 产生后续所需要的特征图; 区域候选网络主要用于生成区域候选框; 感兴趣区域池化层负责收集特征图和区域候选框, 将信息综合起来进行后续类别的判断; 最后一层分类层, 根据区域候选网络综合的信息进行目标类别的判定, 同时修正候选框的位置. 总的来说, Faster-RCNN 首先采用 RPN 网络产生候选框, 之后再对候选框进行位置的修定和目标的分类. 由于其复杂的网络构成, 检测速度相对来说比较慢一点. 基于回归的目标检测算法真正意义上实现了端到端的训练, 以 YOLO为代表, 基于回归的目标检测算法一次性完成目标的分类与定位, 整个网络结构只由卷积层组成, 输入的图像只经过一次网络, 所以基于回归的目标检测算法更快. 改进版的 YOLOv3, 不论在速度上还是在精度上都到达了最先进的水平.

由于 YOLOv3 在目标检测上取得优异成绩, 将 YOLOv3 算法应用于安全帽检测. 基于 ImageNet 上的预训练模型, 修改分类器, 用采集到的 2010 张样本数

据训练安全帽检测器 (Helmet Detector). 利用训练得到的安全帽检测器对包含 2000 张图片的测试集进行测试, 图 1 展示了安全帽检测器结构.

1.1 YOLOv3 网络结构

YOLOv3 以 darknet-53 作为基础网络, 采用多尺度预测 (类 FPN[20]) 的方法, 分别在大小为 13×13、26×26、52×52 的特征图上进行预测. 多尺度预测和特征融合提高了小目标的识别能力, 从而提升整个网络的性能. 图 2 显示了 YOLOv3 的网络结构.

YOLOv3 整个网络只由一些卷积层 (convolution layers)、激活层 (leaky relu)、批标准化层 (Batch Normalization, BN) 构成. 对于一张指定的输入图像, 首先通过 darknet-53 基础网络进行特征的提取和张量的相加, 之后在得到的不同尺度的特征图上继续进行卷积操作, 通过上采样层与前一层得到的特征图进行张量的拼接, 再经过一系列卷积操作之后, 在不同特征图上进行目标检测和位置回归, 最后通过 YOLO 检测层 (YOLO Detection) 进行坐标和类别结果的输出.

图 1 安全帽检测器结构

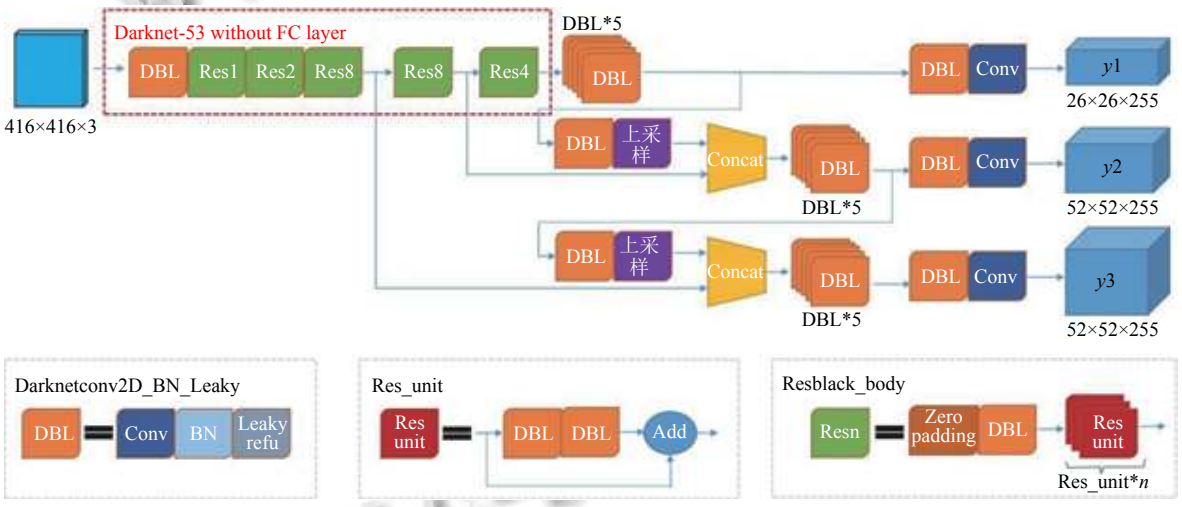

图 2 YOLOv3 网络结构

1.2 分类器设置

YOLOv3 算法在 COCO 数据集上检测 80 种物体类别. 本应用场景中, 只需要检测没有佩戴安全帽一类, 可以将安全帽检测转化为一个单分类问题, 减少网络计算开销.

YOLOv3 输出 3 个不同尺度的特征图: $y1$、$y2$、$y3$, 如图 2 所示. 不同的特征图对应不同的尺度, 分别为 13×13、26×26、52×52, 深度均为 255. 在 YOLOv3 中, 采用类 Anchor 机制, 通过维度聚类的方法确定模版框 (anchor box prior), 模版框的个数 k=9, k 为超参数, 在实验的基础上得出. 9 个模版框由 3 个输出张量平分, 每个输出张量中的每个网格会输出 3 个预测框, 所以针对有 80 种类别的 COCO 数据集来说, 输出张量的维度为 3×(5+80)=255, 其中 3 代表每个网格预测的 3 个模版框, 5 代表每个预测框的坐标信息 (x, y, w, h), 以及置信度 (*confidence*, c).

根据实际场景, 修改分类器, 改变网络最后一层的输出维度. 只检测不戴安全帽一类, 输出维度变为 3×(5+1)=18. 这样可以在不影响实际需求的前提下, 减少网络运算量, 提高检测精度和速度.

2 实验

2.1 数据集制作

实验数据集来源于建筑工地 3 号通道口视频监控. 为了使训练数据集具有较高的质量、模型具有多场景检测能力, 采集到的监控中工作人员佩戴安全帽样本来自后方、前方、左侧方、右侧方等不同的检测角度, 不仅仅局限为某一特定方向. 而且这些数据来源于一天中的不同时间阶段, 具有不同的光照条件. 在这样的数据集上进行模型训练更具有代表性. 制作数据集时, 首先将获取到的不同时间段的视频监控按 1 帧/秒进行截图, 获取训练数据样本后再进行筛选, 过滤掉没有目标的样本, 同时兼顾不同角度的样本的数量, 使各个角度的样本数量基本达到均衡. 根据 Pascal VOC 和 COCO 数据集的图像标注标准, 将获取到的样本使用 yolo-mark 进行标注, 产生训练所需要的 xml 文件. 训练样本示例和标注后的训练样本示例如图 3 和图 4 所示. 本实验训练集包括 2010 张戴安全帽和不戴安全帽样本, 1500 张验证数据集, 以及 2000 张测试集. 训练样本、验证样本和测试样本都不重复. 如表 1 所示.

图 3 训练样本示例

图 4 标注后的训练样本示例

表 1 数据集

	训练集	验证集	测试集	总计
数量	2010	1500	2000	5510

2.2 模型训练细节

制作好训练集, 采用 darknet53.conv.74 预训练权重和 yolov3-cov.cfg 配置文件, 在此基础上利用标注好的训练集进行 YOLOv3 模型的训练. 在训练过程中, 保存日志文件和权重文件. 从日志文件中提取 *loss* 值和 IOU 值做图, 根据损失函数和 IOU 的变化曲线图进行优化调参并确定最优权重. 在测试时, 采用 *loss* 值最小的迭代次数产生的权重作为检测的最终权重文件.

3 实验结果分析

3.1 *Loss* 和迭代次数实验结果分析

Loss 值是整个网络结构的损失函数部分, 它的值越小越好, 期望值为 0. 本实验中将网络结构参数进行微调, 在学习率 (learning rate) 为 0.001, steps=8000, 12 000, scales=0.1, 0.1 下迭代 20 000 次. 由图 5 可以看出: 在前 200 次迭代中, 损失函数值较大, 在迭代到大约 600 次的时候损失函数值骤然下降, 从 700 到 8000 轮迭代过程中, 损失函数值继续较快速下降. 在进行到 8000 次的时候, 学习率降低为之前的 0.1, 损失值缓慢下降. 在 12 000 次的时又将学习率降低为上一次的 0.1, 学习速度变慢, 损失函数小幅度减小, 16 000 次以后, 损失函数值几乎趋于平稳, 不再减少.

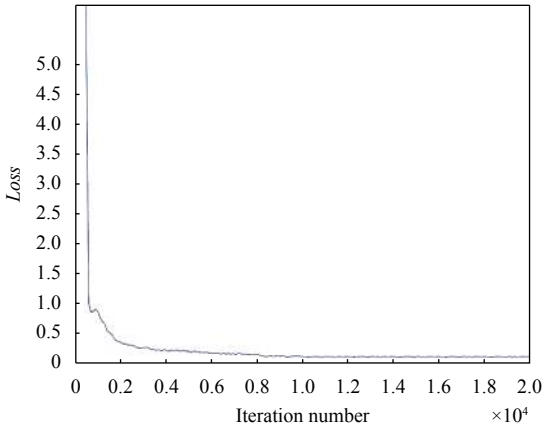

图 5 平均 *loss*

3.2 Avg IOU 实验结果分析

Avg IOU (Intersection Over Union) 指的是在当前迭代次数中, 产生的候选框与原标记框之间交集与并集的比值, 该值越大越好, 期望值为 1. 本实验从训练日志文件中提取 IOU 值信息, 采用滑动平均算法对 80 000 条数据进行平均, 使得曲线更加平滑, 观察变化趋势. 图 6 可以看出, 从第 1 轮到 50 000 轮, 随着迭代次数的增加, 平均 IOU 值总体呈上升趋势, 从 50 000 轮以后, 波动逐渐趋于平缓.

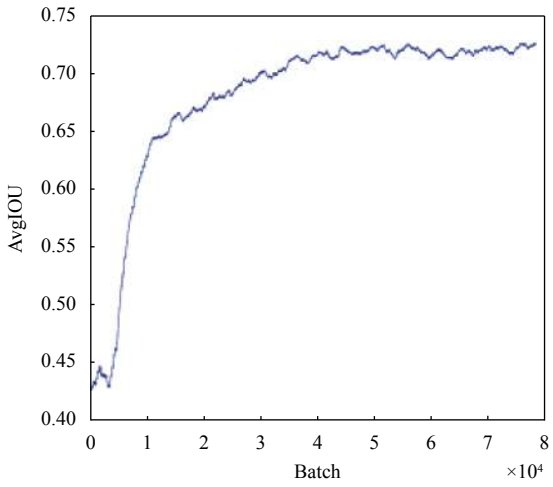

图 6 平均 IOU

3.3 目标检测准确度分析

对 *loss* 曲线分析, 模型采用迭代 17 000 次时的权重文件作为检测模型的最终权重, 将 2000 张测试集用训练好的模型进行测试验证. 分析实验结果, 得出检测准确率. 如表 2 所示.

表 2 安全帽检测器检测结果分析

	验证集	测试集
样本总数	1500	2000
漏检数	7	17
错检数	2	9
总错误数	9	26
准确率	0.994	0.987

实验在无 GPU 的环境下, 平均检测速度达到了 35 fps, 满足实时性的要求. 同时, 本实验在较少训练样本下到达了 98.7% 的准确率, 显示了安全帽检测方法的优越性. 图 7 展示了安全帽检测模型结果示例.

4 结论与展望

详细阐述了基于 YOLO 的安全帽检测方法, 包括分类器设置、训练以及模型优化. 实验结果表明, 基于 YOLO 的安全帽检测方法不论在测试精度上还是在检测速度上都取得了良好的效果. 在 2000 张测试集上进行评估, 达到了 98.7% 的准确率; 在无 GPU 的环境下, 平均检测速度达到了 35 fps. 但是基于 YOLO 的安全帽检测模型在重叠目标上会出现漏检现象, 下一步可针对重叠场景、密集目标进行网络结构改进加子网络进行重叠目标的判断, 也可以增加训练样本的多样性、提高训练样本质量, 在未来需要深入探究.

图 7 安全帽检测器结果示例

参考文献

1 Li Y, Tang S, Lin M, *et al*. Implicit negative sub-categorization and sink diversion for object detection. IEEE Transactions on Image Processing, 2018, 27(4): 1561–1574. [doi: 10.1109/TIP.2017.2779270]

2 Li YK, Yu TS, Li BX. Simultaneous event localization and recognition in surveillance video. Proceedings of the 201815th IEEE International Conference on Advanced Video and Signal Based Surveillance. Auckland, New Zealand. 2018. 1–6. [doi: 10.1109/AVSS.2018.8639169]

3 Cheng YH, Wang J. A motion image detection method based on the inter-frame difference method. Applied Mechanics and Materials, 2014, 490–491: 1283–1286. [doi: 10.4028/www.scientific.net/AMM.490-491]

4 Horn BKP, Schunck BG. Determining optical flow. Proceedings of SPIE 0281, Techniques and Applications of Image Understanding. Washington, WA, USA. 1981. 185–203.

5 Lee DS. Effective gaussian mixture learning for video background subtraction. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(5): 827–832. [doi: 10.1109/TPAMI.2005.102]

6 Viola P, Jones M. Rapid object detection using a boosted cascade of simple features. Proceedings of 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Kauai, HI, USA. 2001. I–I.

7 Dalal N, Triggs B. Histograms of oriented gradients for human detection. Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, CA, USA. 2005. 886–893.

8 Felzenszwalb P, McAllester D, Ramanan D. A discriminatively trained, multiscale, deformable part model. Proceedings of 2008 IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, AK, USA. 2008. 1–8.

9 Burges CJC. A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery, 1998, 2(2): 121–167. [doi: 10.1023/A:1009715923555]

10 Zhu J, Zou H, Rosset S, *et al*. Multi-class AdaBoost. Statistics and its Interface, 2009, 2(3): 349–360. [doi: 10.4310/SII.2009.v2.n3.a8]

11 李开复, 王咏刚. 人工智能. 北京: 文化发展出版社, 2017. 40–49.

12 张玉宏. 深度学习之美 AI 时代的数据处理与最佳实践. 北京: 电子工业出版社, 2018. 414–434.

13 Redmon J, Divvala S, Girshick R, *et al*. You only look once: Unified, real-time object detection. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA. 2016. 779–788.

14 Redmon J, Farhadi A. YOLO9000: Better, faster, stronger. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA. 2017. 6517–6525.

15 Redmon J, Farhadi A. YOLOv3: An incremental improvement. arXiv: 1804.02767, 2018.

16 Girshick R, Donahue J, Darrell T, *et al*. Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA. 2014. 580–587.

17 Girshick R. Fast R-CNN. Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile. 2015. 1440–448.

18 Ren SQ, He KM, Girshick R, *et al*. Faster R-CNN: Towards real-time object detection with region proposal networks. Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal, Canada. 2015. 91–99.

19 Liu W, Anguelov D, Erhan D, *et al*. SSD: Single shot MultiBox detector. Proceedings of the 14th European Conference on Computer Vision. Amsterdam, The Netherlands. 2016. 21–37.

20 Lin TY, Dollár P, Girshick R, *et al*. Feature pyramid networks for object detection. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA. 2017. 936–944.