

doi:10.3969/j.issn.1673-9833.2021.02.007

基于改进 Tiny-yolov3 算法的安全帽佩戴检测

钟鑫豪, 龙永红, 何震凯, 李培云

(湖南工业大学 交通工程学院, 湖南 株洲 412007)

摘要: 针对深度学习方法在视觉上检测安全帽佩戴过程中存在对施工人员等小目标漏检率高和实际中需达到实时监测的要求, 提出一种改进的目标检测模型。首先, 在该算法的原网络上加入残差网络模块, 使得小目标的特征不会随着网络的加深而导致梯度消失的情况, 且能更好地改善对小目标的漏检率高的问题。然后, 对损失函数与筛选预测框进行了优化。理论分析与结果表明: 与原算法相比, 改进后算法的识别准确率提高了 4.6%, 召回率提高了 3.9%, 平均精确率均值提高了 4.1%, 帧率达 63 帧/s。可见提出的改进算法能更好地提取小目标特征, 同时也减少了边界框位置误差。

关键词: 目标检测; 目标特征; 损失函数; 残差网络

中图分类号: TP183; TP391.4 **文献标志码:** A **文章编号:** 1673-9833(2021)02-0046-05

引文格式: 钟鑫豪, 龙永红, 何震凯, 等. 基于改进 Tiny-yolov3 算法的安全帽佩戴检测 [J]. 湖南工业大学学报, 2021, 35(2): 46-50.

Hamlet Wearing Detection Algorithm Based on an Improved Tiny-yolov3

ZHONG Xinhao, LONG Yonghong, HE Zhenkai, LI Peiyun

(College of Traffic Engineering, Hunan University of Technology, Zhuzhou Hunan 412007, China)

Abstract: In view of such flaws as the high miss detection rate of small targets including constructors and the requirement of real-time monitoring in the process of visual safety helmet wearing detection by deep learning method, an improved target detection model has thus been proposed in this paper. First of all, a residual network module is added to the original network of the algorithm, so that the characteristics of small targets will not cause the gradient to disappear as the network deepens, thus better improving the problem of high missed detection rates for small targets. Moreover, an optimization can be achieved of the loss function and the screening prediction frame. Theoretical analysis and results show that, compared with the original algorithm, the recognition accuracy rate of the proposed method can be improved by 4.6%, meanwhile the recall rate is increased by 3.9%, and the average accuracy rate is increased by 4.1%, with the frame rate being 63 frames/s. The results show that the improved algorithm proposed in this paper is characterized with a better performance in extracting small target features, with the position error of the bounding box reduced accordingly as well.

Keywords: target detection; target feature; loss function; residual network

收稿日期: 2020-07-20

作者简介: 钟鑫豪 (1996-), 男, 江西萍乡人, 湖南工业大学硕士生, 主要研究方向为图像处理与机器视觉,
E-mail: 615508331@qq.com

通信作者: 龙永红 (1968-), 男, 湖南株洲人, 湖南工业大学教授, 博士, 主要从事机车, 汽车等智能检测与控制理论与应用, 机器视觉及其应用方面的教学与研究, E-mail: lyhcai@126.com

0 引言

在铁路施工中,安全是施工项目中最重要的一部分,每年在施工工程中由于未佩戴安全帽导致的安全事故时有发生。在国内,人工监测安全帽的佩戴占大多数,这不仅耗时耗力,如果监管不力还会导致严重的后果。因此,智能检测施工人员佩戴安全帽就十分有益,对此,本文通过图像处理的方式对施工人员佩戴安全帽的检测进行了研究。

国内外学者对此都做了大量的工作和研究,现有的检测佩戴安全帽的目标检测方法主要为两类。

第一类方法为传统的目标检测识别,蔡利梅等^[1]通过采集安全帽的外形特征信息来建立模型,用于矿井视频的安全帽识别。胡恬等^[2]采用小波变化和BP(back propagation)神经网络的应用,提出了针对安全帽的识别神经网络模型。刘晓慧等^[3]则利用施工人员的肤色与周围的环境颜色来定位人脸,再采用支持向量机(support vector machines, SVM)实现安全帽的识别。此类方法的优点在于不依赖于建筑工人的面部特征识别与特征描述准确,但严重依赖手工设计的特征。在铁路施工现场中,存在着光照反射,此类方法不适用于复杂的建筑施工环境并且检测速度太慢,满足不了实时监测的要求。

另一种方法则是基于深度学习的目标检测识别算法。该方法凭借着卷积神经网络的无需手动设计图像特征这一优势^[4-5],逐渐获得许多研究者的青睐。对于该类方法目前的目标检测方法主要有两类,第一类是“two-stage”深度学习算法,它基于区域提名,输入图片进去先生成候选框,然后对生成的候选框进行分类。例如R-CNN^[6-8]、R-FCN^[9]等算法。第二类则是“one-stage”的端对端深度学习算法,输入图片后直接对整个图片进行预测、定位和分类。例如yolo、SSD^[10]等算法,此类方法由于在检测步骤上比第一种方法简化了一些步骤,因此检测速度较第一类快,但是精度较低。虽然基于深度学习的目标检测方法相比于传统的目标检测方法有着鲁棒性较高、检测精度较好的特点,但是由于卷积神经网络的复杂性,对图片集的训练将消耗许多计算资源,并且要达到实时检测的目的对计算机的配置要求很高。因此,本文提出一种改进 Tiny-yolov3 的安全帽检测方法。

1 Tiny-yolov3 算法改进

1.1 Tiny-yolov3 算法简介

yolo 系列算法是 J. Redmon 等^[11-13]提出的基于深度学习的目标检测模型,由于它的提出是为实现

高精度的在线目标检测,因此 yolo 系列算法采用的是 one-stage 的目标检测算法,其中 yolov3 是该系列中最新提出的改进算法。并且该算法在 Coco 等数据集上有不俗表现,在工业和商用上均有广泛的应用。而本次采用的 Tiny-yolov3 算法是在 yolov3 算法的平台上的轻量级的实时检测算法,将 yolov3 上的特征网络 Darknet-53 的 53 层神经网络简化为 13 层,因此它的检测速度比 yolov3 快许多。其卷积结构仍然使用与 yolov3 相同的全卷积网络(convolutional networks, FNC)和批量标准化(batch normalization)等。

Tiny-yolov3 结构如图 1 所示,它主要由卷积层和池化层构成,分别有 7 层卷积层和 6 层池化层(max pooling),网络中每层卷积层或者池化层后的特征图尺寸分别表示分辨率宽、分辨率高、通道数。

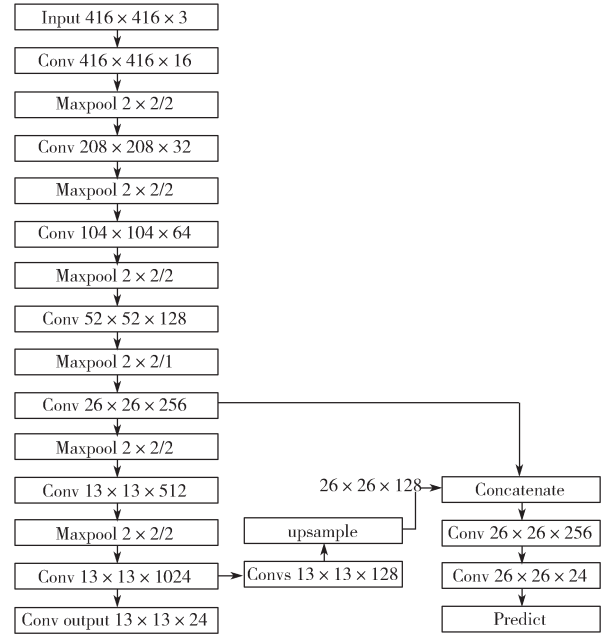


图 1 Tiny-yolov3 网络结构图

Fig. 1 Tiny-yolov3 network structure diagram

Tiny-yolov3 的工作流程如下:首先,输入一张图片,将图片划为 $S \times S$ 的网格,而每个网格内出现 B 个预测框;然后对预测框内的目标进行类型检测;最后输出检测目标的置信度和检测框。而置信度的计算式为

$$C = P_r(o)R_{iou} \quad (1)$$

式中: $P_r(o)$ 为目标 o 存在的概率,一般为 0 或者 1; R_{iou} 为预测框和真实包含目标框的交并比。

因此,置信度的数值是由每个网格中包含检测目标的概率和预测框的准确度共同决定的。

训练模型时, Tiny-yolov3 使用的损失函数与 yolov3 相同,主要由预测框的位置 (x, y) , 预测框的长宽 (w, h) , 预测类别(class), 以及预测置信

度 (confidence) 确定。损失函数公式如下:

$$L = \frac{1}{n} \sum_{i=1}^n L_{xy} + \frac{1}{n} \sum_{i=1}^n L_{wh} + \frac{1}{n} \sum_{i=1}^n L_{\text{class}} + \frac{1}{n} \sum_{i=1}^n L_{\text{confidence}} \quad (2)$$

式中: n 为预测框个数; L_{xy} 为预测框位置误差; L_{wh} 为预测框长宽误差; L_{class} 为类别预测误差; $L_{\text{confidence}}$ 为置信度误差。

1.2 添加残差网络

由于 Tiny-yolov3 的神经网络采用了 yolov3 的简化神经网络模型, 因此与 yolov3 相比, 其速度快上许多, 但是检测精度随之下降。提高深度学习模型的检测精度最常用的方法是增加网络深度, 即增加卷积层数。卷积神经网络层数越深, 收敛的可能性越小, 浅网络提取的小对象的特征随着网络的加深而被稀释。如果网络太深, 当在图层之间传递要素信息时, 也会导致要素信息丢失。因此, 课题组在原始网络的第 4 层和第 7 层之间添加了残差网络结构。残差网络使用 1×1 个卷积层和 3×3 个卷积层提取特征。将输入结构之前的特征图添加到残差结构之后生成的特征图中, 同时将浅层信息和深层信息传输到下一个卷积层以提取特征。这样, 可以减少在层之间通过时特征信息的这种丢失, 并且可以提高网络检测的准确性。残差模块如图 2 所示, 其中 n 指输入通道数, C 指通道数, t 指通道扩张或压缩的倍数。

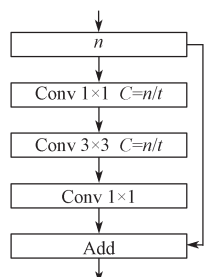


图 2 残差网络结构

Fig. 2 Residual network structure

1.3 损失函数与筛选预测框的改进

Tiny-yolov3 的评估方法与 yolov3 相同, 交并比 (intersection over union, IOU) 是预测框与真实框的交集, 是目标检测重要的评估方法。由交并比能得出损失函数, 目标检测的损失函数一般由分类损失函数 (classification loss) 和回归损失函数 (bounding box) 构成。而由于 Tiny-yolov3 的初始 R_{iou} 存在一些问题:

1) 当预测框和真实框不相交, 即 $R_{\text{iou}}=0$ 时, Loss 值无法进行评估;

2) 当存在两框重叠并且 R_{iou} 相同的情况, 但是位置预测框位置不相同, Loss 值无法区分相交情况下的不同。因此, 课题组提出用距离交并比

DIOU_Loss (Distance_IOU_Loss) 来代替 IOU_Loss 做损失函数的评估。距离交并比计算式如式 (3)。

$$R_{\text{diou}} = R_{\text{iou}} - \frac{D_{\text{euclidean}}^2}{D_{\text{center}}^2} = \frac{(A \cap B)}{(A \cup B)} - \frac{D_{\text{euclidean}}^2}{D_{\text{center}}^2} \quad (3)$$

式中: A 为真实框; B 为预测框; $A \cap B$ 为真实框和预测框的交集区域; $A \cup B$ 为真实框和预测框的并集区域; $D_{\text{euclidean}}$ 为最小外接框对角线距离; D_{center} 为两个中心点的欧氏距离。

本文将原算法中的筛选预测框的 NMS 算法 R_{diou} 改进为 L_{diou} 算法, 如式 (4) 所示:

$$L_{\text{diou}} = 1 - R_{\text{diou}} \quad (4)$$

当使用 DIOU 算法来检测目标时, 将会考虑预测框、真实框的重叠面积和两者的中心距离, 当存在预测框在真实框里面时, 直接计算两个框之间的距离, 从而达到快速收敛的目的, 并且解决了无法区分预测框在真实框内部时无法区分相对位置的问题。

2 实验

2.1 实验平台

本文的实验是在 windows10 环境下完成的算法搭建, 硬件环境如下: 处理器, AMD Ryzen5 3600 3.6 GHz; GPU 显卡, RTX2060SUPER, 显卡的内存为 8 GB; 计算机内存为 16 GB。软件上安装了 Visual Studio2015、python 3.6.5, 并且同时安装了 CUDA10.3 和 cudnn9.0.0, 以支持 NVIDIA GPU 的使用, 深度学习框架为 Caffe^[14]。

2.2 数据集制作

本文针对所研究的问题, 制作了一个规模适中的数据集。数据来源于监控视频截图、网络爬虫和图片库收集。为保证训练出的模型具备较高鲁棒性和多场景检测能力, 样本图片中的施工人员的拍摄角度各异, 光照的条件也有所不同。课题组按照 Pascal VOC 的数据集格式构建了自己的数据集, 数据集包括 5 323 张图片, 并采用 labellmg 工具对图像进行分类和标定后生成 XML 文件, 界面见图 3 所示。



图 3 labellmg 界面

Fig. 3 Labellmg interface

标记者的类型和坐标位置信息,通过程序将 XML 文件转换为 TXT 文件输入到训练集中。本实验按照 7:2:1 的比例设置训练集、验证集和测试训练后的模型集。

2.3 实验指标

本实验的结果评估标准主要为识别准确率 p (precision)、召回率 r (recall)、平均精确率均值 (mean average precision, mAP) 和检测速度 FPS。首先, p 表示在识别的正样本中,真实样本所占的比率,即

$$p = \frac{TP}{TP + FP} \quad (5)$$

式中: TP 为正确分类为正样本的数; FP 为将目标错误分为正样本的数。

r 表示识别正确的正样本在总样本数中所占的比例,即

$$r = \frac{TP}{TP + FN} \quad (6)$$

式中 FN 为被错误划分为负样本的正样本数。

如果只用 precision 或 recall 作为衡量一个模型检测精度的优劣显然不合适。因此,课题组还需要目标检测中最重要的指标之一的 mAP,是多个验证集的平均 AP 值, mAP 是由 Precision-recall 曲线与坐标轴包围区域的面积。 n 为计算的组数,则差值近似的公式为

$$mAP = \frac{1}{n} \sum_{i=1}^n \int_0^1 p(r) dr \quad (7)$$

2.4 实验结果分析

课题组使用官网上提供的 Tiny-yolov3 的权重参数作为权重训练的初始参数,并根据自制的训练集进行参数微调以达到最佳训练效果,部分试验参数调整如表 1 所示。

网络参数按照表 1 进行了 Tiny-yolov3 和本文改

进后的 Tiny-yolov3 算法训练和验证,并分别计算了识别准确率 p 、召回率 r 、平均精确率均值 mAP 和检测速度 FPS ,结果如表 2 所示。

表 1 网络参数说明表

Table 1 Network parameter description table

参 数 名	参数值
学习率 (learning rate)	0.001
动量 (momentum)	0.9
衰减 (decay)	0.000 5
批量大小 (batch size)	64

表 2 各算法的检测结果参数

Table 2 Parameters of detection results of each algorithm

算 法	$p/\%$	$r/\%$	$mAP/\%$	$FPS/(\text{帧} \cdot s^{-1})$
Tiny-yolov3	81.2	85.3	61.2	67
改进的 Tiny-yolov3	85.8	89.2	65.3	63

从表 2 所示结果来看,改进后的 Tiny-yolov3 在准确率和召回率上分别有着 4.6% 和 3.9% 的提升,在最关键的 mAP 上也有 4.1% 的提升。由于改进的网络上增加了残差结构,因此比原网络的 FPS 慢一点,但是影响不大,依然能满足实时性的要求。另外,为了更直观地表现两种算法的差异,选取了一些两种算法检测的效果图。图 4 为本文安全帽佩戴检测改进算法与原算法 Tiny-yolov3 的比较结果图。其中,图 a、d、g 为原图,图 b、e、h 为 Tiny-yolov3 检测结果,图 c、f、i 为本文改进的 Tiny-yolov3 算法的实验结果。对原图 1 的检测中,图 b 漏检了左前的一位佩戴了安全帽的人,图 c 没有漏检;对原图 2 的检测中,图 e 漏检了后面佩戴了安全帽的人和被遮挡的人,图 f 没有漏检佩戴安全帽人员,但是漏检了一些被遮挡的人;对原图 3 的检测中,图 h 中漏检了后面的 2 人与右侧的人,图 i 无人漏检。从实验结果比较来看,本文的改进算法能更好地检测出小目标。



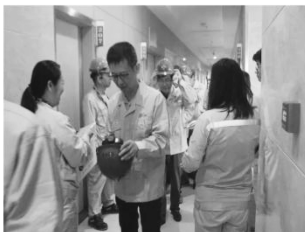
a) 原图 1



b) 原图 1 的 Tiny-yolov3 检测图



c) 原图 1 的改进算法检测图



d) 原图 2



e) 原图 2 的 Tiny-yolov3 检测图



f) 原图 2 的改进算法检测图



g) 原图 3



h) 原图 3 的 Tiny-yolov3 检测图



i) 原图 3 的改进算法检测图

图 4 实验结果

Fig. 4 Experimental results

3 结语

本文基于 Tiny-yolov3 的深度学习算法,通过在特征提取网络上加入残差网络模块,在不太影响检测速度的情况下,提高了小目标特征的获取。同时,在损失函数与筛选框的优化中,引入了 DIOU_Loss 的重合边界框的误差计算以提高目标的识别准确率。通过理论分析与实验结果表明:改进后的 Tiny-yolov3 与原算法相比,识别准确率提高了 4.6%,召回率提高了 3.9%,平均精确率均值提高了 4.1%,帧率达到 63 帧/s,满足实时监测的要求。但是,与 yolov3 等大型检测网络比起来,小目标检测的识别准确率有待加强。因此,接下来的工作是如何提高更小目标的检测准确率。

参考文献:

- [1] 蔡利梅. 基于视频的煤矿井下人员目标检测与跟踪研究 [D]. 徐州: 中国矿业大学, 2010.
CAI Limei. Research on Target Detection and Tracking of Underground Coal Mine Personnel Based on Video[D]. Xuzhou: China University of Mining and Technology, 2010.
- [2] 胡 恬, 王新刚. 基于小波变换和神经网络的安全帽识别系统分析与设计 [J]. 软件导刊, 2006(23): 37-38.
HU Tian, WANG Xingang. Analysis and Design of Helmet Recognition System Based on Wavelet Transform and Neural Network[J]. Software Guide, 2006(23): 37-38.
- [3] 刘晓慧, 叶西宁. 肤色检测和 Hu 矩在安全帽识别中的应用 [J]. 华东理工大学学报(自然科学版), 2014, 40(3): 365-370.
LIU Xiaohui, YE Xining. Skin Color Detection and Hu Moments in Helmet Recognition Research[J]. Journal of East China University of Science and Technology(Natural Science Edition), 2014, 40(3): 365-370.
- [4] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137-1149.
- [5] 姚群力, 胡 显, 雷 宏. 深度卷积神经网络在目标检测中的研究进展 [J]. 计算机工程与应用, 2018, 54(17): 1-9.
YAO Qunli, HU Xian, LEI Hong. Application of Deep Convolutional Neural Network in Object Detection[J]. Computer Engineering and Applications, 2018, 54(17): 1-9.
- [6] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[C]//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Washington DC: IEEE, 2014: 580-587.
- [7] GIRSHICK R. Fast R-CNN[C]//Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago: IEEE, 2015: 1440-1448.
- [8] YANG B, ZHANG Y, CAO J, et al. On Road Vehicle Detection Using an Improved Faster RCNN Framework with Small-Size Region Up-Scaling Strategy[C]// Pacific-Rim Symposium on Image and Video Technology. Wuhan: Springer, 2017: 241-253.
- [9] DAI J, LI Y, HE K, et al. R-FCN: Object Detection Via Region-Based Fully Convolutional Networks[C]// Proceedings of the Neural Information Processing Systems. New York: Curran Associates Inc, 2016: 379-387.
- [10] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single Shot Multibox Detector[C]//European Conference on Computer Vision. Amsterdam: Springer, 2016: 1-37.
- [11] REDMON J, DIVVALA S K, GIRSHICK R B, et al. You Only Look Once: Unified, Real-Time Object Detection[C]// Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 779-788.
- [12] REDMON J, FARHADI A. YOLO9000: Better, Faster, Stronger[C]//The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017: 7263-7271.
- [13] REDMON J, FARHADI A. YOLOv3: an Incremental Improvement[EB/OL]. [2020-04-11]. https://www.researchgate.net/publication/324387691_YOLOv3_An_Incremental_Improvement.

(责任编辑: 申 剑)