

DOI: 10.11991/ykj.201809014

网络出版地址: <http://kns.cnki.net/kcms/detail/23.1191.U.20181017.1655.002.html>

一种改进的卷积神经网络行人识别方法

陈聪, 杨忠, 宋佳蓉, 韩家明

南京航空航天大学 自动化学院, 江苏 南京 211106

摘 要:针对现有的行人检测算法存在的定位精度低、实时性差的问题, 借鉴目标检测的研究成果 You Only Look Once (YOLO) 算法, 提出一种实时的行人检测方法。以 Tiny-YOLO 为基础, 改变网络模型的输入尺寸, 获得更好的行人特征表达; 结合图像中行人尺寸特点, 使用聚类分析方法, 对数据集进行目标框聚类, 选取适合行人检测的候选框尺寸与数量; 通过增加一定数量卷积层的方法重新设计特征提取和目标检测网络; 在混合数据集上训练, 增强模型泛化性。实验结果表明, 在应对不同尺寸行人和部分遮挡情况时, 文中方法具有更低的漏检率、更好的定位精度与检测效果, 且检测速度可以满足实时性要求。

关键词:行人检测; 卷积神经网络; 深度学习; YOLO; 特征提取; 聚类分析; 多尺度特征; 行人数据集

中图分类号: TP751

文献标志码: A

文章编号: 1009-671X(2019)03-0051-07

An improved pedestrian detection method based on convolutional neural network

CHEN Cong, YANG Zhong, SONG Jiarong, HAN Jiaming

College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China

Abstract: To solve the problem of poor positioning accuracy and low real-time performance in current pedestrian detection algorithms, a real-time pedestrian detection method is proposed by applying the You Only Look Once (YOLO) algorithm of target detection. Based on the Tiny-YOLO network, by changing the size of input of the network model, it helps the network gain better expression of pedestrian characteristics. Combined with the pedestrian size characteristics in the image, the cluster analysis method is used to calculate the initial candidate boxes in dataset, and the size and number of anchor boxes suitable for pedestrian detection are selected. Convolution layers are added in order to redesign the feature extraction and target detection networks. Train the network on a mixed dataset to improve the generalization of the model. The experimental results show that the method has lower miss rate, better positioning accuracy and detection effect when it comes to the situation of different sized pedestrians and partial occlusion, and the detection speed can meet the real-time requirements.

Keywords: pedestrian detection; convolutional neural network(CNN); deep learning; YOLO; feature extraction; cluster analysis; multi-scale feature; pedestrian dataset

近年来, 随着与人工智能产业的兴起以及交通运输等行业的不断发展, 行人检测技术在智能监控、智能驾驶等领域得到了广泛的关注。传统行人检测方法中, 通过人工设计的特征提取器, 例如 Haar、方向梯度直方图 (histogram of oriented

gradient, HOG)、局部二值模式 (local binary pattern, LBP) 等从训练样本中提取行人特征, 再用提取的行人特征训练支持向量机 (support vector machine, SVM) 等分类器, 进行行人检测任务^[1]。例如可利用 HOG+LBP 特征处理行人遮挡问题, 提高行人检测的准确率^[2]。随着积分通道特征 (integral channel features, ICF)、聚集信道特征 (aggregated channel features, ACF)、梯度幅值特征, LUV 颜色特征等提出, 行人特征获得了更好地表达, 在行人检测任务中也取得了较好的成果^[3]。但通过手动设计的行人特征有着泛化性差、难以适应行人

收稿日期: 2018-09-14. 网络出版日期: 2018-10-17.

基金项目: 国家自然科学基金项目 (61473144); 江苏高校优势学科建设工程资助项目; 中国南方电网有限责任公司科技项目 (066600KK52170074)。

作者简介: 陈聪, 男, 硕士研究生;
杨忠, 男, 教授, 博士生导师。

通信作者: 杨忠, E-mail: YangZhong@nuaa.edu.cn.

形态变化等缺点,在实际应用中困难重重。

随着深度学习的迅速发展,硬件计算能力不断增强,相应的数据集不断构建,深度神经网络在不同的视觉任务中取得了巨大的成功,深度神经网络模型也在行人检测领域得到了广泛的运用^[3]。深度神经网络模型可以从图像像素中学习特征,提高行人检测器性能。在目标检测方面,基于区域建议 (regional proposal, RP) 的目标检测算法如基于区域的卷积神经网络^[4](region based convolutional neural networks, RCNN)、Faster-RCNN^[5],基于回归 (regression) 的目标检测算法如单镜头多盒检测器^[6](single shot multibox detector, SSD)、You Only Look Once^[7](YOLO) 系列算法,不仅大大加快了目标检测的速度,也维持了较高的准确率。

然而,现有的目标检测算法中,R-CNN、Faster R-CNN 由于实时性难以得到保证而较难实现应用;YOLO 系列算法在经过不断地改进后,在检测的实时性方面性能十分突出:YOLO V2^[8]、YOLO V3^[9]虽然具有较好的检测准确率,但是需要较大的 GPU 显存 (>4 GB),资源消耗量大,在部署方面存在一定的困难。因此,需要设计一种具有较好的实时性且对 GPU 等硬件配置要求不高的网络。我们选择 Tiny-YOLO^[10]网络,这是 YOLO 系列网络的简化版,拥有更少的卷积层、更快的检测速度以及较好的检测精度。对 Tiny-YOLO 网络进行改进,使其满足行人检测的实际需求。

1 Tiny-YOLO 原理

1.1 Tiny-YOLO 网络结构

Tiny-YOLO 主干网络拥有 7 个 3×3 的卷积 (convolution) 层,6 个池化 (maxpooling) 层,前 5 为个步长为 2 的池化层,最后一个为步长为 1 的池化层。模型输入为 416 pix×416 pix,在经过 5 个步长为 2 的池化层后,最终特征图大小为 13×13。网络结构如图 1 所示。

在神经网络中,输入的图片中目标大小不同,不同尺寸的目标在最终输出时获取的特征信息量不同。神经网络中较浅的卷积层可以很好地表征小尺寸目标,而较深的卷积层的特征能更好地描述大尺寸目标。因此,根据目标尺寸的不同,选择不同的卷积层特征获取更多的语义信息,对目标进行预测^[11],这样对不同尺寸的目标有更好的适应性。在 Tiny-YOLO 网络中使用了 2 种尺度对目标进行预测:

1) 在基础网络之后添加一些卷积层进行预测,此时特征图大小为 13 pix×13 pix。尺度 1) 对

大尺寸目标有较好的预测能力。

2) 从尺度 1) 中的倒数第 2 层的卷积层进行 2 倍上采样,上采样可以帮助网络学习细粒度特征,与最后一个尺寸为 26 pix×26 pix 的特征图相加,再次通过多个卷积后输出预测信息,特征图大小为 26 pix×26 pix。尺度 2) 使用了较大的特征图,对较小尺寸的目标有更好的预测能力。

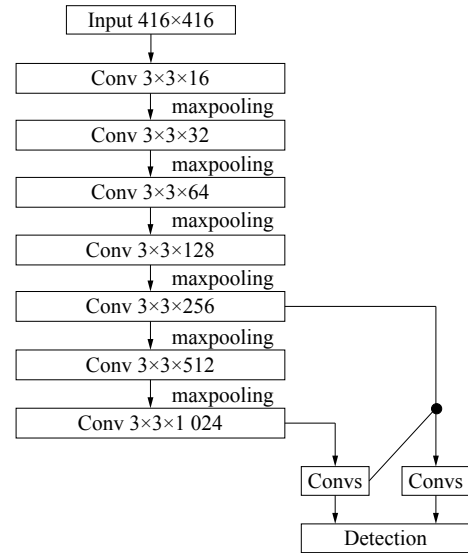


图1 Tiny-YOLO 网络结构

1.2 损失函数

在 YOLO 系列网络中,损失函数由坐标误差、交并比 (intersection over union, IOU) 误差和分类误差 3 部分组成。

$$E = \sum_{i=0}^{S^2} E_{\text{coord}} + E_{\text{IOU}} + E_{\text{class}}$$

在行人检测训练时,训练损失函数可表示为:

$$\begin{aligned} & \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B \prod \sum I_{ij}^{\text{person}} [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] + \\ & \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B \prod \sum I_{ij}^{\text{person}} [(\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + \\ & (\sqrt{h_i} - \sqrt{\hat{h}_i})^2] + \sum_{i=0}^{S^2} \sum_{j=0}^B \prod \sum I_{ij}^{\text{person}} (c_i - \hat{c}_i)^2 + \\ & \lambda_{\text{noperson}} \sum_{i=0}^{S^2} \sum_{j=0}^B \prod \sum I_{ij}^{\text{noperson}} (c_i - \hat{c}_i)^2 + \\ & \sum_{i=0}^{S^2} \prod \sum I_i^{\text{person}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2 \end{aligned}$$

式中: S 代表最终特征图的网格尺寸; B 代表每个网格的预测框个数; (x, y) 代表代表框的中心坐标, w, h 代表框的宽和高; c 代表框的置信度 (confidence); p 代表行人置信度; λ_{coord} 为候选框损

失权重, $\lambda_{\text{noperson}}$ 为分类损失权重; I_{ij}^{person} 表示第 i 个网格的第 j 个预测框存在行人的可能性, I_i^{person} 表示判定第 i 个网格存在行人的概率; $(\hat{x}, \hat{y}, \hat{w}, \hat{h}, \hat{c}, \hat{p})$ 为网络相应的预测值。

1.3 激活函数

在卷积神经网络中,需要引入非线性函数作为激活函数。引入激活函数可以增加神经网络模型的非线性,使神经网络模型具有更好的表达能力。在卷积神经网络中,常用的激活函数有 Sigmoid 函数、tanh 函数、ReLU 函数等。Tiny-YOLO 网络采用了 leaky ReLU 函数作为激活函数,这样可以有效地避免在训练阶段 Sigmoid、ReLU 等函数出现的梯度消失问题^[12]。leaky ReLU 激活函数公式为:

$$f(x) = \max(x, 0.1x)$$

1.4 边界框预测

Tiny-YOLO 使用由维度聚类得到 anchor box 作为先验框来预测边界框 (bounding box):

$$\begin{cases} b_x = \sigma(t_x) + c_x \\ b_y = \sigma(t_y) + c_y \\ b_w = p_w e^{t_w} \\ b_h = p_h e^{t_h} \end{cases}$$

式中: c_x 、 c_y 是网格的坐标偏移量; p_w 、 p_h 是预设的 anchor box 的边长,最终得到的 (b_x, b_y) 为边界框的中心坐标; b_w 、 b_h 为边界框宽度和高度; t_x 、 t_y 为边界框中心坐标的学习目标; t_w 、 t_h 为边界框宽度和高度的学习目标; σ 表示预测的边框的置信度。

2 改进的 Ped-YOLO 网络

Tiny-YOLO 网络作为针对多类别目标检测的一种简化网络,已经取得了较好的检测效果和实时性,但并不完全适合行人检测任务,因此,需要针对具体问题进行相应的改进。

在现有的行人检测模型如 HyperLearner^[13]中,一方面,利用额外的通道特征改善行人检测模型的性能;另一方面,通过采用更小的卷积核以及更多的采样获得更全面的行人特征表达。文中用于行人检测的 Ped-YOLO 模型相对于 Tiny-YOLO 模型改进主要体现在以下 3 个方面:

1) 改变训练时输入图像的分辨率;

2) 对行人检测数据集的目标框进行聚类,确定 anchor 参数;

3) 改进网络结构,使用更深的卷积神经网络对行人特征进行提取。

2.1 优化输入图像分辨率

行人识别研究表明,图像中行人身高和宽度为一定值,行人的横向特征表达比纵向特征表达少,增加横向特征表达有助于行人识别。文中采用改变网络模型输入的长宽比例,利用矩形输入网络提取更多的横向特征^[14],考虑到原网络输入图像尺寸为 416×416 ,为避免输入图像分辨率对网络的影响,选择了 608×288 分辨率图像作为网络输入。一方面可以有更好的横向特征提取效果;另一方面二者具有相近的像素数,不会影响检测实时性。

2.2 K-means 聚类分析

YOLO 系列算法引入了 Faster R-CNN 中的 anchor boxes 的思想^[15],在数据集上进行聚类确定 anchor boxes。YOLOv2 的初始候选框 anchor 参数针对 PASCAL VOC 数据集进行聚类确定,聚类结果为 5,因此其 anchor 数目为 5;YOLOv3 初始候选框 anchor 参数针对 COCO 数据集聚类确定,聚类结果为 9,因此其 anchor 的数目为 9,分属 3 个不同预测尺度;Tiny-YOLO 的 anchor 参数同样针对 COCO 数据集聚类确定,聚类结果为 6,因此其 anchor 的个数为 6,分属 2 个不同预测尺度。由于 PASCAL VOC 数据集与 COCO 数据集中类别较多,得到的 anchor 参数具有普适性,但是并不适用于行人检测。在行人检测中,行人的长宽比不会因行人的背景、姿态等发生大的改变,往往是一个相对固定的比值,因此需要对数据集进行聚类分析,重新选择最优的 anchor 个数和宽高维度。文中采用 K-means 聚类方法获得先验框的尺寸,得到 anchor 的个数和宽高维度。

通过设置先验框,可以提高预测框与 ground truth 之间的交并比。IOU 是产生的候选框 (candidate bound) 与原标记框 (ground truth bound) 的交叠率,即它们的交集与并集的比值。传统的 K-means 聚类方法对两点间距离进行计算常用的是曼哈顿距离、欧式距离等。如果使用欧式距离函数,这会导致较大的框比较小的框产生更多的错误,考虑到行人检测数据集中行人尺寸问题,采用 IOU 来衡量聚类结果。

聚类中的距离函数为:

$$d(\text{box}, \text{centroid}) = 1 - \text{IOU}(\text{box}, \text{centroid})$$

式中: centroid 代表簇的中心; box 代表样本聚类结果; $\text{IOU}(\text{box}, \text{centroid})$ 代表簇中心框和聚类框的交并比。

选择 608×288 作为模型输入大小,对行人检

测数据集进行聚类分析,使用递增的方法选择 k 值。通过计算不同 k 值下的 IOU 值,选择 k 的最优值。IOU 与 k 的关系如图 2 所示。

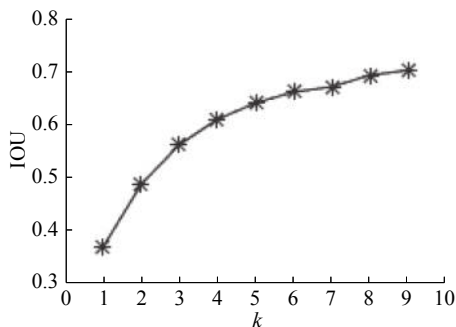


图2 IOU- k 折线图

根据图中的折线可知,在 $k=6$ 之后,IOU 值的增加逐渐变得平缓。考虑到网络的计算量,且改进的 Ped-YOLO 网络中使用原 Tiny-YOLO 网络在 2 种尺度上进行跨尺度预测的方法,最终采用 $k=6$ 的聚类结果。

2.3 网络结构改进

在卷积神经网络中,更多的卷积层与更深的网络结构往往对目标特征有更好的提取效果。而原 Tiny-YOLO 网络的层数较少,主干网络中只有 7 层卷积层,很难对行人目标特征有较好的提取效果,因此需要对网络结构进行改进,增加的卷积层如图 3 所示。一方面,考虑到行人区域相对较小,为了获得更高分辨率的信息,常用的 7×7 卷积层和 5×5 卷积层不适用于行人检测,因此选取 3×3 卷积核对原 Tiny-YOLO 网络进行加深。相对于原 Tiny-YOLO 网络, Ped-YOLO 在主干网络中增加了额外的 3 个 3×3 卷积层。

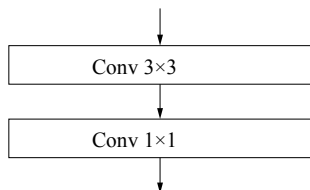


图3 Ped-YOLO 中增加的卷积层

另一方面,在网络加深的同时,额外的卷积层会造成网络模型参数过多,从而增加网络的运算量。通过借鉴 Resnet^[16]与 Densenet^[17]的思想,我们在网络中引入 1×1 卷积核。引入 1×1 卷积核的作用有两方面:一方面是卷积核通道数的降维,减少卷积核参数,简化神经网络模型;另一方面,引入 1×1 的卷积核,在前一卷积层的学习表示上添加了非线性激励,提升了网络的表达能力,同时也增加了模型深度,一定程度上提升了模型的特征能力。Ped-YOLO 最终网络结构如图 4 所示。

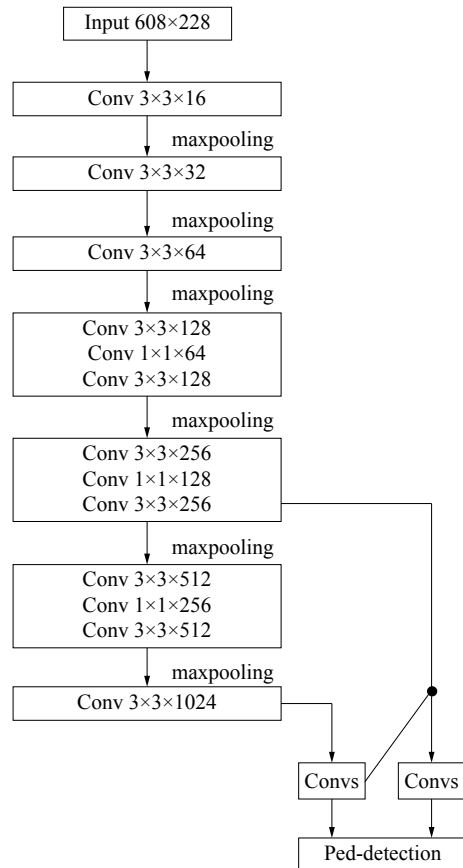


图4 Ped-YOLO 网络结构

3 实验结果及分析

3.1 训练及测试数据集

文中采用的行人检测数据库为 PASCAL VOC 与 INRIA 混合数据集。INRIA 数据集是目前使用最广泛的静态行人检测数据集, INRIA 数据集分为训练集和测试集,训练集中包含了正负样本,其中正样本 614 张,正样本中有 2 416 人,负样本 1 218 张;测试集中包含正样本 288 张,正样本中有 1 126 人,负样本 453 张。INRIA 数据集中图片的清晰度较高,背景也较为复杂,拍摄条件多样,存在光线变化、人体遮挡等情形,图片中人体大部分为站立姿势,高度大于 100 个像素^[18]。考虑到 INRIA 数据集训练样本较少,实际应用时泛化能力较差且容易出现过拟合现象,本文提取了 PASCAL VOC 数据集中行人图片 6 383 张,与 INRIA 数据集进行数据融合,提高 Ped-YOLO 网络的泛化能力。

3.2 实验平台

本文实验平台配置如表 1 所示。在该实验环境下,本文中的 Ped-YOLO 模型在 720 p 的 mp4 视频上检测速度可达到 66.8 f/s,满足实时性要求。

表1 实验平台软硬件配置

名称	相关配置
操作系统	Windows 10
中央处理器/GHz	Intel Core i7-7700, 3.6
内存/GB	32
GPU/GB	Nvidia GeForce Titan Xp, 12
GPU 加速库	CUDA 9.1, CUDNN 7.1
深度学习框架	Darknet

3.3 模型搭建与训练

在开源的深度学习框架 Darknet 上,搭建与训练 Ped-YOLO 模型。作为一种广泛应用的深度学习框架,Darknet 可适用于多种格式的图片输入,其中 INRIA 数据集的图片为 PNG 格式,PASCAL VOC 数据集的图片为 JPEG 格式。训练时学习率采用均匀分布策略 (polynomial decay),模型的初始学习率 (learningrate) 设为 0.001,在 25 000 和 30 000 次迭代后,学习率乘以 0.1,动量系数为 0.9,权值衰减系数为 0.001 0,训练中采用了图像随机调整曝光、饱和度、色调等方法对数据进行扩充。此外,在训练时,每迭代 10 次改变模型的输入尺寸,使模型对于不同尺寸的图像具有更好的检测效果。

3.4 实验结果与分析

3.4.1 混合数据集测试

分别在 INRIA 数据集和 PASCAL VOC 混合数据集上训练 Ped-YOLO 网络,并使用 INRIA 测试集进行测试。评价指标为召回率 (recall) 与平均精确值 (mean average precision, mAP),召回率指正确识别出来的行人个数与测试集中所有行人的个数的比值。

实验结果如表 2 所示。结果表明,使用 PASCAL VOC+INRIA 混合数据集训练与使用 INRIA 数据集训练相比,在混合数据集上,mAP 提高了 1.07%,召回率提高了 0.05。

表2 混合数据集测试结果

数据集	平均精确值/%	召回率
INRIA	89.54	0.87
INRIA+PASCAL VOC	90.61	0.92

3.4.2 平均交并比测试

为了验证文中设计的 Ped-YOLO 网络的定位准确性,利用平均交并比作为指标进行评测。本文在混合数据集上分别训练了 Tiny-YOLO 系列网络中的 YOLOv2-tiny 网络和 YOLOv3-tiny 网络作为 Ped-YOLO 网络的对照,并使用 INRIA 数据集中测试集测试平均交并比。

对比结果如表 3 所示。结果表明,相比较于 YOLOv2-tiny 网络,Ped-YOLO 网络平均交并比提高了 10.03%;相比较于 YOLOv3-tiny 网络,Ped-YOLO 网络平均交并比提高了 5.41%。这说明,在测试集上,Ped-YOLO 产生的预测框与原标记框的交叠率更高,对行人定位的精度更好。原因在于,通过对数据集的 K-means 聚类分析选择先验框以及网络结构的改进,可以更好地提升模型的定位精度。

表3 平均交并比测试结果 %

检测方法	平均交并比
YOLOv2-tiny	72.26
YOLOv3-tiny	76.88
Ped-YOLO	82.29

3.4.3 精度测试

使用 INRIA 数据集中测试数据集进行精度测试。本文选择 LAMR^[19] (log-average miss rate) 指标作为评判的标准。LAMR 指标表示的是平均每张图片误检数 (false positive per image, FPPI) 在 $[10^{-2}, 10^2]$ 上与漏检率之间的关系。在 FPPI=0.1 时,将本文的方法与传统的 HOG+SVM、Latent SVM+HOG 方法以及 YOLOv2-tiny、YOLOv3-tiny 的检测效果进行比较。

实验结果如表 4 所示。Ped-YOLO 相对于 SVM+HOG^[20] 方法漏检率减少 32.74%,相对于 Latent SVM+HOG^[21] 方法漏检率减少 6.72%,相对于 YOLOv2-tiny 网络漏检率减少 4.40%,相对于 YOLOv3-tiny 网络漏检率减少 4.07%。原因在于,采用人工设计的特征提取器 (如 HOG、LBP 等特征提取算法),行人特征表达不够充分;而 YOLOv2-tiny、YOLOv3-tiny 虽然通过卷积网络学习行人特征,但卷积网络层数相对较少;而 Ped-YOLO 网络解决了这个问题,因此对行人特征的表达能力优秀。

表4 精度测试结果 %

检测方法	漏检率
SVM+HOG	45.98
Latent SVM+HOG	19.96
YOLOv2-tiny	18.84
YOLOv3-tiny	17.31
Ped-YOLO	8.49

为了更加直观地检验 Ped-YOLO 网络的有效性,本文选取 INRIA 测试集中图像以及私有图像进行了检测效果测试,从中挑选了 3 张不同行人

尺寸以及存在遮挡情况的图像在 Tiny-YOLO 和 Ped-YOLO 这 2 个不同的模型上分别进行检测,

其结果如图 5、6 所示。



图5 Tiny-YOLO 检测结果



图6 Ped-YOLO 检测结果

可以看出,对于图像中较大尺度的无遮挡行人,2种网络具有相近的检测能力;对于较小尺寸的行人,改进的 Ped-YOLO 网络有着较好的检测结果,而 Tiny-YOLO 网络则会对小尺寸行人存在漏检的情况。此外,对于图像中存在的部分互相遮挡的目标行人,Tiny-YOLO 的定位精度较差,且存在漏检的情况;而 Ped-YOLO 网络在存在部分互相遮挡的行人情况下,表现出了非常好的性能,有着更好的定位精度,漏检率更低。这说明,改进了训练时的分辨率,并针对行人检测数据集重新聚类;加深了卷积神经网络结构,使用混合数据集,对行人的特征提取更加充分,对不同尺寸的行人以及部分遮挡的行人有更好的检测效果。

4 结论

本文以 Tiny-YOLO 网络为基础,通过改变网络输入尺寸、维度聚类分析、改进特征提取网络的方法,提出了一种改进的 Ped-YOLO 的行人检测网络。

1) 与传统算法和 Tiny-YOLO 相比,Ped-YOLO 网络具有更低的漏检率,Ped-YOLO 网络在保持了较高的检测速度的同时,对不同尺寸、存在互相遮挡的行人有更好的检测效果。

2) 本文提出的网络仍然存在一些问题,如训练受限于 PASCAL VOC 与 INRIA 数据集,行人训

练样本偏少,对小目标行人检测效果偏差,行人被较大范围遮挡时,模型的检测效果也不强。因此,可以通过融合多类行人检测数据集进行网络训练,改善网络对不同尺寸、不同环境下行人的检测效果,增强模型的泛化能力。

3) 在行人检测中,由于光照、遮挡等问题的存在,漏检率和误检率仍然达不到理想的水平。因此,需要提取更多的信息提升检测效果,如光流信息、语义信息等,提高特征表达能力,这也将成为行人检测的重要发展方向。

参考文献:

- [1] SABZMEYDANI P, MORI G. Detecting pedestrians by learning shapelet features[C]//Proceedings of 2007 IEEE Conference on Computer Vision and Pattern Recognition. Minneapolis, MN, USA: IEEE, 2007: 1-8.
- [2] WANG Xiaoyu, HAN T X, YAN Shuicheng. An HOG-LBP human detector with partial occlusion handling[C]//Proceedings of the 12th International Conference on Computer Vision. Kyoto: IEEE Press, 2009: 32-39.
- [3] 高宗,李少波,陈济楠,等.基于 YOLO 网络的行人检测方法[J].计算机工程,2018,44(5): 215-219.
- [4] 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. Columbus, OH, USA: IEEE, 2014.
- [5] REN Shaoqing, HE Kaiming, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[C]//Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal, Canada: ACM, 2015: 91–99.
- [6] LIU Wei, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//Proceedings of the 14th European Conference on Computer Vision. Amsterdam: Springer, 2016: 21–37.
- [7] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE, 2016: 779–788.
- [8] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE, 2017: 6517–6525.
- [9] REDMON J, FARHADI A. YOLOv3: an incremental improvement[J]. arXiv: 1804.02767, 2018.
- [10] REDMON J. Darknet: open source neural networks in C[DB/OL]. [2017-10-25]. <http://pjreddie.com/darknet/>.
- [11] 刘辉, 彭力, 闻继伟. 基于改进全卷积网络的多尺度感知行人检测算法[J]. 激光与光电子学进展, 2018, 55(9): 91504.
- [12] 冯帅, 张龙, 贺小慧. 基于 Jetson TK1 和深度卷积神经网络的行人检测[J]. 信息技术, 2017(10): 62–64.
- [13] MAO Jiayuan, XIAO Tete, JIANG Yuning, et al. What can help pedestrian detection?[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE, 2017: 6034–6043.
- [14] 郝旭政, 柴争义. 一种改进的深度残差网络行人检测方法[J/OL]. 计算机应用研究, 2019, 36(6). [2018–05–22]. (2018–03–16) <http://www.aocmag.com/article/02–2019–06–057.html>. DOI: 10.3969/j.issn.1001–3695.2017.12.0836.
- [15] DAI Jifeng, LI Yi, HE Kaiming, et al. R-FCN: object detection via region-based fully convolutional networks[C]//Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona, Spain: ACM, 2016: 379–387.
- [16] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE, 2016: 770–778.
- [17] HUANG Gao, LIU Zhuang, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE, 2017: 2261–2269.
- [18] 刘建国, 罗杰, 王帅帅, 等. 基于 YOLOv2 的行人检测方法研究[J]. 数字制造科学, 2018, 16(1): 50–54.
- [19] DOLLAR P, WOJEK C, SCHIELE B, et al. Pedestrian detection: an evaluation of the state of the art[J]. [IEEE transactions on pattern analysis and machine intelligence](#), 2012, 34(4): 743–761.
- [20] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]//Proceedings of 2015 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, CA, USA: IEEE Computer Society, 2005: 886–893.
- [21] FELZENSZWALB P F, GIRSHICK R B, MCALLESTER D, et al. Object detection with discriminatively trained part-based models[J]. [IEEE transactions on pattern analysis and machine intelligence](#), 2010, 32(9): 1627–1645.

本文引用格式:

陈聪, 杨忠, 宋佳蓉, 等. 一种改进的卷积神经网络行人识别方法[J]. 应用科技, 2019, 46(3): 51–57.

CHEN Cong, YANG Zhong, SONG Jiarong, et al. An improved pedestrian detection method based on convolutional neural network[J]. Applied science and technology, 2019, 46(3): 51–57.