ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 38,No. 1,Jan. 2023,pp. 162－173
DOI:10. 16337/j. 1004-9037. 2023. 01. 014

http://sjcj. nuaa. edu. cn
E-mail:sjcj@nuaa. edu. cn
Tel/Fax:＋86-025-84892742

融合深浅特征和动态选择机制的行人检测研究

沙梦洲[1,2]，沈　韬[1,2]，曾　凯[1,2]，马　倩[1,2]，曾文健[1]

(1.昆明理工大学信息工程与自动化学院,昆明 650500;2.昆明理工大学云南省计算机技术应用重点实验室,昆明 650500)

摘　要：针对无人驾驶场景下行人多尺度、小尺度造成漏检率升高,检测精度下降的问题,本文提出一种融合深浅层特征和级联动态选择机制的行人检测方法。首先,在YOLO v3-tiny的基础上基于密集连接的卷积神经网络改进特征提取部分,融合行人的深层特征和浅层特征加强网络对行人的识别能力;其次,在改进的主干网络上级联具有动态选择机制的注意力模块,使检测网络更加适应动态的行人尺度变化;最后,本文选择BDD 100K数据集和Caltech加州理工学院行人数据集进行实验,在保证实时性的前提下(25 ms/张),本文模型在BDD 100K数据集行人漏检率降低11.4%,平均检测精度提高11.7%,在Caltech行人漏检率降低10.1%,平均检测精度提高6.7%,适用于无人驾驶行人检测领域。

关键词：无人驾驶;小尺度;行人检测;密集连接;动态选择机制

中图分类号：TP391　　**文献标志码**:A

Pedestrian Detection Incorporating Deep and Shallow Features and Dynamic Selection Mechanisms

SHA Mengzhou[1,2], SHEN Tao[1,2], ZENG Kai[1,2], MA Qian[1,2], ZENG Wenjian[1]

(1. Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China; 2. Yunnan Key Laboratory of Computer Technologies Application, Kunming University of Science and Technology, Kunming 650500, China)

Abstract: Aiming at the problem that the multi-scale and small-scale of pedestrians in unmanned scenario causes the increase of missed detection rate and the decrease of detection accuracy, this paper proposes a pedestrian detection method that fuses deep and shallow layer features and cascade dynamic selection mechanism. Firstly, on the basis of YOLO v3-tiny, we improve the feature extraction part based on the densely connected convolutional neural network, and fuse the deep and shallow features of pedestrians to enhance the network's ability to recognize pedestrians. Secondly, we cascade the attention module with dynamic selection mechanism on the improved backbone network to make the detection network more adaptable to dynamic pedestrian scale changes. Finally, we choose the BDD 100K dataset and the Caltech pedestrian dataset to conduct experiments. Under the premise of real-time performance (25 ms/sheet), the missed detection rate of pedestrian is reduced by 11.4% and the average detection accuracy is improved by 11.7% in the BDD 100K dataset, and the missed detection rate of pedestrian is reduced by 10.1% and the average detection accuracy is improved by 6.7% in the Caltech dataset, which is suitable for unmanned

基金项目:国家自然科学基金(61971208);云南省中青年学术技术带头人后备人才基金(沈韬,2019HB005);云南省万人计划青年拔尖人才基金(沈韬, 朱艳,云南省人社厅2018 73);云南省重大科技专项基金(202002AB080001-8)。
收稿日期:2021-08-15;**修订日期**:2022-04-18

pedestrian detection.

Key words: driverless; small scale; pedestrian detection; dense connectivity; dynamic selection mechanisms

引　言

行人检测[1]作为机器感知重要的一部分,在机器视觉、无人驾驶[2]等领域都有较强的应用价值。在无人驾驶场景中,由于行人外观易受外部光照条件、姿态变化、目标遮挡等影响,在复杂的交通环境中进行行人检测容易出现漏检、检测精度不佳等问题,所以行人检测一直是一项富有挑战且极具研究价值的实际任务。因此,如何在提高行人检测效率的同时,提升面对复杂环境的鲁棒性具有深刻的意义。当前行人检测的方法主要分为两类方法:基于图像本身特征进行提取的传统手工特征和机器学习方法以及基于样本自动训练出行人特征分类器的深度学习方法。

(1)基于传统的特征提取方法:基于图像本身的手工特征进行提取

传统的目标检测方法主要是基于手工设计物体特征和选择分类器。主流的图像特征有Haar小波变换[3]、尺度不变特征变换(Scale invariant feature transform, SIFT)[4]、方向梯度直方图(Histogram of oriented gradient, HOG)[5]等,通过提取完手工特征再结合具体对象选择分类器,主流的物体分类器有支持向量机(Support vector machine, SVM)[6]、自适应增强(Adaptive boosting, Adaboost)[7]等。

2005年Xu等[8]提出一种行人检测跟踪方法,针对行人外观非刚性的特点采用SVM实现分类。该方法对于白天场景需要根据像素强度找出候选对象,通常需要对视频数据的每一帧进行详尽搜索,所以算法实时性较差。2010年Cerri等[9]提出基于AdaBoost级联元算法的全天候行人分类系统,基本思想是使用基于Haar特征的AdaBoost和基于AdaBoost特征的AdaBoost系统,以达到更好的行人分类,但是AdaBoost分类器对于异常值的鲁棒性较差。2012年Guo等[10]采用AdaBoost算法和级联方法从图像中分割行人候选,行人识别分类器使用SVM进行训练,在辅助驾驶系统中性能优于传统的单级分类器,但是两阶段的方式增加了算法的复杂度。

基于传统的特征提取方法需要根据不同的检测对象设计出不同的手工特征,然后挑选适合的分类器辨别行人,算法的实时性难以得到保证。

(2)基于深度学习的方法:设计不同的网络结构自动学习行人特征和实现分类

近年来,基于卷积神经网络的深度学习方法在无人驾驶领域实现了较多的成果,其中具有代表性的目标检测方法主要分为两阶段检测器、单阶段检测器。两阶段检测器主要代表为快速区域建议网络(Fast region proposal network, Fast R-CNN)[11]、掩码区域建议网络(Mask region proposal network, Mask R-CNN)[12]等,单阶段检测器主要代表为单个深层网络检测(Single shot multibox detector, SSD)[13]、一次检测网络(You only look once, YOLO)[14]等,单阶段检测器能实现更快的检测速度,但是检测精度一般低于两阶段检测器。

2016年Zhang等[15]发现Faster R-CNN中的区域建议网络(Region proposal network,RPN)处理小实例的特征图分辨率不足,提出首先使用RPN搜索行人区域,然后在高分辨率特征图级联增强森林对行人进行分类,但是两阶段的方法导致行人检测实时性较差。2018年Lan等[16]在实时性较强的原始YOLO网络中添加3个Passthrough层,可以很好地将网络的浅层行人细粒度特征传递到深层网络,使网络更好地学习浅层行人特征,但是没有考虑行人复杂的尺度变化问题,导致该方法对行人的检测精度较低。2019年Zhang等[17]对经典的LeNet-5卷积神经网络改进和优化,添加归一化层和动态自适应池化模型增强网络对尺度变化问题的鲁棒性,但是改进后的LeNet-5检测被遮挡行人、小尺度行人也会

出现误判现象。2019年Zhang等[18]考虑无人驾驶行人检测的实时性问题，借鉴Resnet的方法，在原有YOLO v3-tiny网络的基础上添加了3个卷积层，提高模型提取行人特征的能力，并在网络中增加1×1卷积层以减少计算复杂度，但是在考虑网络检测速度的同时没有考虑增强对小尺度行人特征的能力，该方法对小尺度行人存在漏检现象。

YOLO v3-tiny[19]是YOLO系列的简化版本之一，具有对硬件要求低、检测速度快的特点。但是面对行人检测小尺度容易漏检、多尺度复杂变化造成检测精度低的问题，如何在保证实时性的前提下，使YOLO v3-tiny网络更适用于无人驾驶行人检测领域，仍值得进一步研究。针对以上问题，本文提出使用深度学习方法对行人进行检测识别，主要贡献有：

(1)在YOLO v3-Tiny的基础上，提出基于密集连接[20]的思想改进YOLO v3-tiny的特征提取网络，该方法通过卷积、池化和密集连接块生成不同的特征层，增强深层特征和浅层特征之间交互性，提高网络模型对小尺度行人特征的提取能力。

(2)提出在主干网络中级联具有动态选择机制的注意力网络[21](Selective kernel network, SKNet)，改进后的检测方法能动态地选择不同尺度的卷积核，使网络更加适应交通场景行人复杂的尺度变化。

(3)实验结果表明，相比较YOLO v3-tiny网络，DK-YOLO v3-tiny在保证实时性的前提下，在BDD 100K数据集行人漏检率降低11.4%，平均检测精度提高11.7%；Caltech行人数据集[22]的漏检率降低10.1%、平均检测精度上升6.7%，适用于无人驾驶的应用场景。

1 网络模型

1.1 YOLO v3-tiny网络分析

对于无人驾驶计算能力较弱的轻型嵌入式设备，YOLO v3-tiny相较YOLO v3具备完全的轻量效应优势，被广泛运用在无人驾驶、无人机等实时性较高的应用场景。YOLO v3-tiny网络总共只有24层，相比较YOLO v3网络的107层精简很大，在检测端只有2个YOLO层，分别为YOLO13和YOLO26，其大小分别为13×13和26×26，每个网格产生3个anchor box的位置坐标和1个置信度，总共有6个anchors值。在交通场景下整个网络在前向传播过程中，被遮挡与受背景干扰的小尺度行人特征比较少，而YOLO v3-tiny主干网络比较浅，不能充分提取复杂多变的行人特征，从而导致这些目标的细节特征在整个模型的深层网络中消失。YOLO v3-tiny网络结构如图1所示。

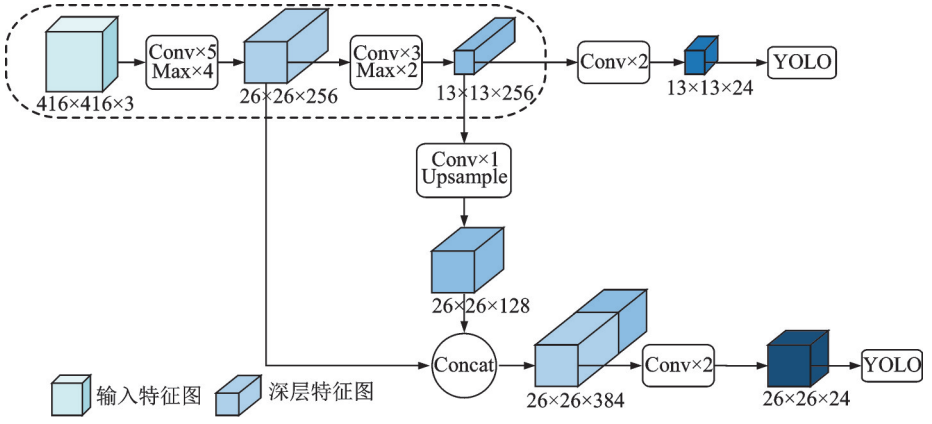


图1 YOLO v3-tiny网络结构图

Fig.1 YOLO v3-tiny network framework

1.2 DK-YOLO v3-tiny 网络结构

YOLO v3-tiny没有类似YOLO v3设计残差层加深网络,特征提取网络只使用不同的卷积、池化分开生成特征层,深层特征和浅层特征之间没有联系,导致特征层的利用率较低,不利于目标检测网络分类回归物体。本文基于密集连接思想在卷积层连接方面进行改进,利用Dense block模块每一层都融合其他层的特征输出,并在特征提取网络中结合具有自适应尺度的注意力网络SK,设计全新的DK-YOLO v3-tiny行人检测网络,如图2所示。

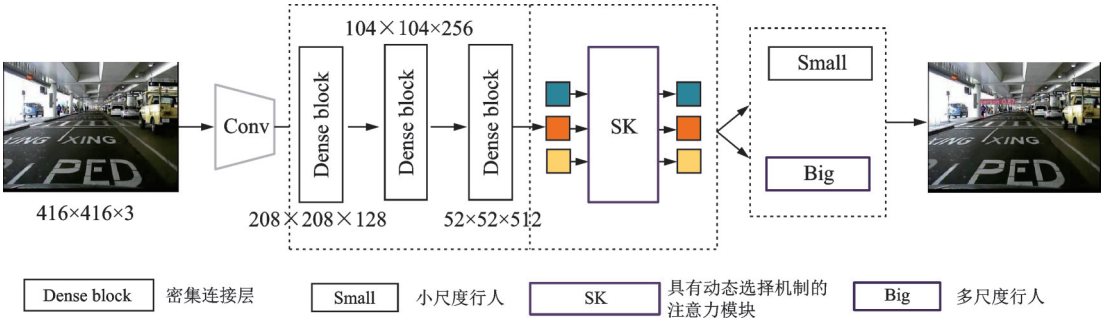


图2 DK-YOLO v3-tiny网络结构图

Fig.2 DK-YOLO v3-tiny network structure

2 研究方法

2.1 Dense 深浅层特征

密集连接网络(Dense network, DenseNet)在卷积层连接方面,主要思想在于建立不同层之间的连接关系,让每一层都融合其他层的特征输出,通过直接将浅层的卷积层和深层的卷积层建立密集连接,实现对特征的重复利用。在加深网络的同时,保存了低纬度的特征,使用不同层次的特征维度进行更平滑的决策。在充分利用特征层的同时解决梯度消失问题,网络训练效果较好。

DK-YOLO v3-tiny在原YOLOv3-tiny网络基础上,采用3层Dense block设计全新的特征提取网络,在加深卷积层的同时不会带来复杂的计算问题。行人的浅层特征与深层特征进行充分融合,在网络的前向传播中不用尺度行人的有效特征不易丢失,有利于回归预测中深层卷积层上采样与浅层卷积层的结合,从而加强DK-YOLO v3-tiny网络对小尺度行人特征的提取能力。在Dense block模块内,下一层的卷积层会联系前面所有卷积层的输出,以i层为例,第i层卷积层的输出为

$$X(i)=H([x_0,x_1,x_2,\cdots,x_{(i-1)}]) \tag{1}$$

式中:x_0为输入特征图;x_i为第i层的特征图;$H(i)$为层间变换函数,具有密集连接思想的Dense block结构如图3所示。

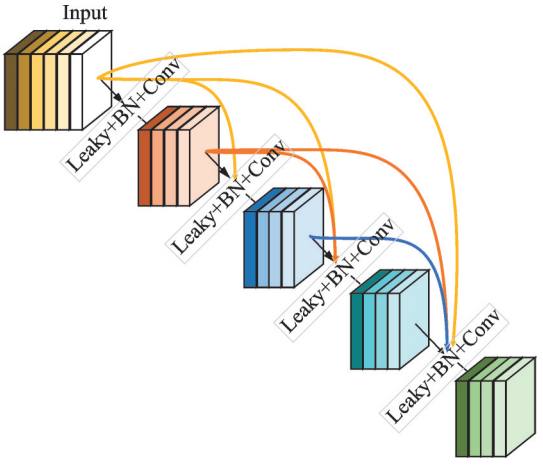


图3 Dense block结构图

Fig.3 Dense block structure

2.2 具有动态选择机制的注意力网络

交通环境复杂多变,无人驾驶汽车和待检测的行人都处于运动状态,行人具有多尺度、小尺度的特点。在YOLO v3-tiny网络中,同一CNN单元激活函数的

感受野大小相同,不能适应动态的尺度变化,然而不同大小的感受野(卷积核)对于不同尺度的行人会有不同的效果,所以检测网络需要更灵活的机制提高网络对局部和全局信息的提取能力。

DK-YOLO v3-tiny级联上具有动态选择机制的注意力网络SKNET,对输入的特征图可以自适应选择卷积核大小,扩大网络感受野捕获图像的局部和全局信息,适应不同尺度的行人目标。在多卷积核分支结构上,通过注意力机制赋予特征通道不同的权重,从而重新校准特征来提高网络的表示能力,动态选择机制结构如图4所示,图中$\boldsymbol{X}$表示特征映射,H、C、W表示三维坐标,$\boldsymbol{U}_1$为经过3×3卷积后的特征向量,$\boldsymbol{U}_2$为经过5×5卷积后的特征向量,$\boldsymbol{U}$为$\boldsymbol{U}_1$和$\boldsymbol{U}_2$相加后的特征向量,$\boldsymbol{S}_c$表示经过全局平均池化的特征,$\boldsymbol{Z}$表示经过全连接生成的特征,$\boldsymbol{V}$表示不同卷积核的注意力权重得到的特征图。

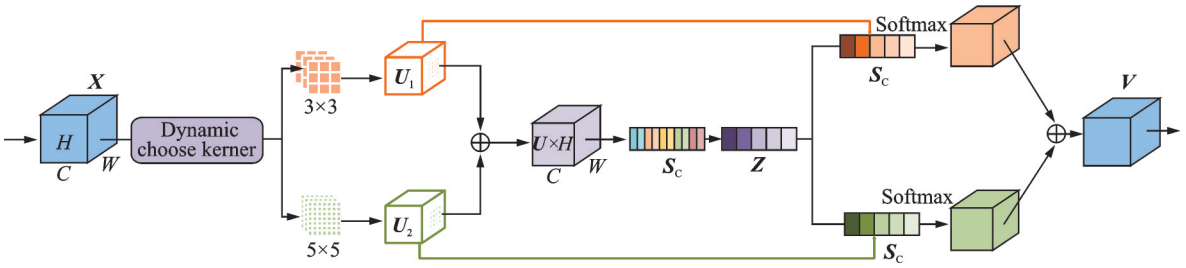


图4 动态选择机制注意力模块

Fig.4 Dynamic selection mechanism attention module

网络主要由分离,融合,选取3部分组成。

第一步:分离部分采用不同大小的卷积核并行处理特征图。对于输入的特征映射$\boldsymbol{X}\in\mathbf{R}^{H\times W\times C}$,并行采用不同尺度的卷积操作,卷积核尺寸大小为$2i+3$,其中$i\in N$。

$$F:\boldsymbol{X}\rightarrow\boldsymbol{U}_{2i+3} \tag{2}$$

第二步:融合部分是计算每个卷积核的权重大小。首先将每个特征维度信息进行聚合,然后$\boldsymbol{U}_c$通过全局平均池化(Global average pooling, GAP)统计不同通道的信息,表达式为

$$\boldsymbol{U}_c=\lim_{i\in[0,2]}U_{2i+3} \tag{3}$$

$$\boldsymbol{S}_c=F_{gp}(\boldsymbol{U}_c)=\frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}\boldsymbol{U}_c(i,j) \tag{4}$$

式中:$\boldsymbol{S}_c\in\mathbf{R}^{C\times 1}$;$F_{gp}$表示全局平均池化。

经过全连接生成紧凑的特征$\boldsymbol{Z}$,即

$$\boldsymbol{Z}=F_{fc}(s)=\delta(\beta(\boldsymbol{W}_s)) \tag{5}$$

式中:$\boldsymbol{Z}\in\mathbf{R}^{d\times 1}$;$\delta$为RELU激活函数;$\beta$表示批标准化(Batch noramlization);F_{fc}表示全连接;$\boldsymbol{Z}$的维度为卷积核的个数;$\boldsymbol{W}_s\in\mathbf{R}^{d\times c}$,其中$d$代表全连接后的特征维度,即

$$d=\max(C/r,L) \tag{6}$$

式中:$L=32$;r为压缩因子。

第三步:选取部分是通过不同权重的卷积核计算得出新的特征图。首先,通过使用柔性最大值传输函数(Softmax)对分支中的特征层(Feature map)设置权重,每个分支被赋予不同的通道权重。然后,网络可以根据行人尺度大小自主选择合适的感受野通道大小,实现动态选择自适应的感受野大小,选取部分示意图如图5所示,图中$\boldsymbol{A}$、$\boldsymbol{B}\in\mathbf{R}^{C\times d}$,$\boldsymbol{a}$、$\boldsymbol{b}$表示软注意力向量。

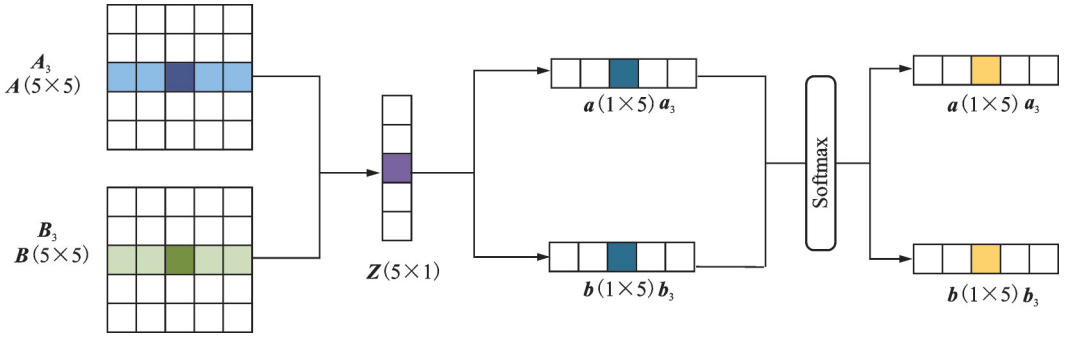


图5 选取部分示意图

Fig.5 Schematic diagram of selected part

2.3 多尺度训练方法

输入图片的尺寸对检测模型的性能影响相当明显。相比较单一尺度训练网络模型,在训练时设置不同尺度的输入图片,可使多尺度训练出来的模型对物体大小鲁棒性更强。该方法训练出来的模型在测试阶段不仅使小尺度的图片测试速度更快,而且大尺度的图片检测准确度更高。

多尺度训练方法对于卷积神经网检测分类任务可以提高模型的鲁棒性和提高检测精度。本文设计的DK-YOLO v3-tiny模型是比较典型的卷积神经网络,数据维度通过网络中不同卷积层和池化层进行传递,首先设置图片分辨率的输入范围,然后网络在每个Epoch的迭代训练前会随机选择范围内的一个图片输入分辨率,让网络模型在迭代训练的过程中可以适应不同尺度大小的图片。本文网络每批次训练的图片最小输入分辨率为320像素 × 320像素,最大输入分辨率为640像素 × 640像素,测试时选择尺度为640像素 × 640像素进行测试。

2.4 损失函数分析

在无人驾驶场景中,行人的物理尺寸较小且容易受到背景的干扰,DK-YOLO v3-tiny网络主要在于准确定位到小尺度行人的位置,减少背景干扰提高检测算法鲁棒性。因此,本文损失函数由3部分相加得到,分别为行人目标中心坐标和宽高坐标误差L_{pos}、置信度误差L_{obj}和分类误差L_{cls},具体计算为

$$L_{\text{pos}}=\lambda_{\text{coord}}\sum_{i=0}^{S^2}\sum_{j=0}^{B}l_{i,j}^{\text{obj}}\{[(x_i-\hat{x}_i^j)^2+(y_i-\hat{y}_i^j)^2]+[(\sqrt{\omega_i^j}-\sqrt{\hat{\omega}_i^j})^2+(\sqrt{h_i^j}-\sqrt{\hat{h}_i^j})^2]\} \tag{7}$$

$$L_{\text{obj}}=\lambda_{\text{noobj}}\sum_{i=0}^{S^2}\sum_{j=0}^{B}l_{i,j}^{\text{noobj}}(c_i-\hat{c}_i)^2+\lambda_{\text{obj}}\sum_{i=0}^{S^2}\sum_{j=0}^{B}l_{i,j}^{\text{obj}}(c_i-\hat{c}_i)^2 \tag{8}$$

$$L_{\text{cls}}=\sum_{i=0}^{S^2}l_{i,j}^{\text{obj}}\sum_{c\in\text{class}}([\hat{P}_i^j\lg(P_i^j)+(1+\hat{P}_i^j)\lg(1-P_i^j)]) \tag{9}$$

式中:$S\times S$为特征图的大小;B为先验框的数量;x_i、y_i为中心点的横坐标和纵坐标;ω_i、h_i为预测框的宽度和高度;$l_{i,j}^{\text{obj}}$表示第i个网格中的第j个锚帧,如果有行人目标则为1,否则为0;类似地,$l_{i,j}^{\text{noobj}}$表示在第i网格中的第j个锚框中不存在行人;c_i表示预测框包含目标对象的概率得分,$\hat{c}_i$代表真实值;P_i^j表示预测框属于类别c的概率,$\hat{P}_i^j$表示标签框所属类别的真实值;λ_{coord}为协调不同大小矩形框对误差函数贡献不一致的协调系数;λ_{noobj}表示当预测框没有预测到目标时其置信度误差在损失函数中所占权重;λ_{obj}表示当预测框预测到目标时其置信度误差在损失函数中所占权重。

3 实验结果与分析

3.1 实验数据与处理

3.1.1 BDD 100K数据集

本文研究无人驾驶领域行人检测问题,伯克利大学AI实验室发布的BDD 100K数据集是目前内容最具多样性的公开驾驶数据集,包含10万段高清视频,分辨率大小为1 280像素 × 720像素。采样每个视频的第10 s关键帧,共得到10万张图片,其中存在大量不同时间、天气等复杂环境的图片。本文只针对该数据集标签为行人(Person)的图片,并提取第5帧,共获得训练集4 420张,测试集3 220张,图6为BDD 100K数据集部分样例。

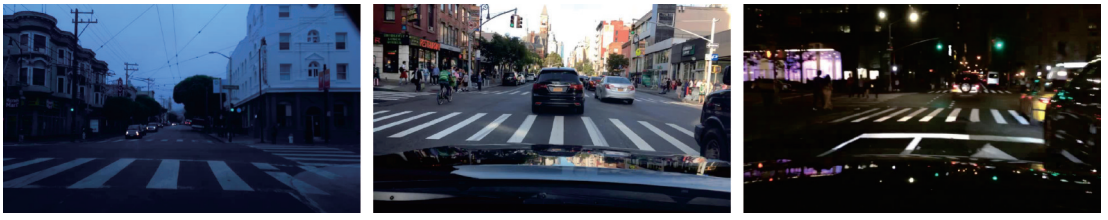


图6 BDD 100K数据集部分样例

Fig.6 Partial samples of the BDD100K dataset

3.1.2 加州理工学院Caltech行人数据集

加州理工学院Caltech行人数据集是无人驾驶领域规模最大的行人检测数据集,全部数据都是正常行驶的车辆采用车载摄像头在城市环境中采集得到,分辨率大小为640像素 × 480像素,其中包含了众多难检测的小尺度行人、被遮挡行人以及不同尺度的行人。

数据集视频序列包括set00~set10共11个序列。其中,加州理工学院将前6个序列划分为训练集,共有61 439张图片,其余为测试集,共有60 748张图片。因为数据集为视频流,相邻编号的图像具有较大的相似,所以本文采取每14张取1张图片的方式降低视频数据集前后帧的影响,共获得训练集4 389张,测试集4 340张。图7为Caltech数据集部分样例。

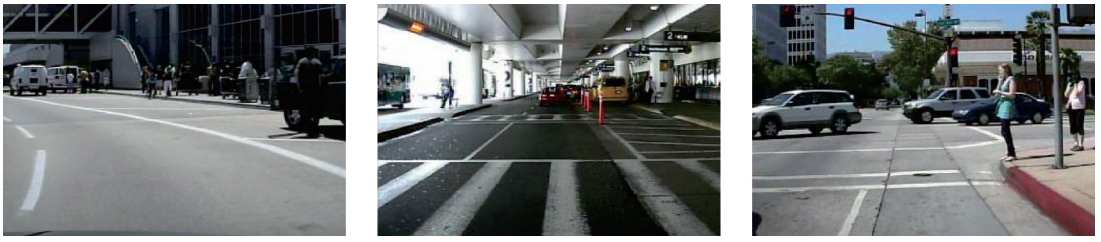


图7 Caltech行人数据集部分样例

Fig.7 Partial samples of the Caltech pedestrian dataset

3.2 实验设备

本文所有实验均使用Ubantu16.04作为主系统,工作站处理器型号为Intel(R)Core(TM)i5-9400F CPU @ 2.90 GHz,显卡型号为NVIDIA Geforce GTX2060,内存为16 GB。深度学习框架采用工程使用较多的PyTorch框架和OpenCV图像处理库。

本文训练阶段初始输入大小为416 × 416 × 3,批量大小(Batch size)设置为4,实验的迭代次数均设为500 epochs。其中,网络参数动量(Momentum)为0.9,学习率(Learning rate)为0.001,权值衰减

(Decay)为0.005,实验相关设施与参数配置如表1所示。

表1 实验相关设备与参数设置

Table 1 Experiment related facilities configuration

设备	相关配置
操作系统	Ubuntu 16.04
处理器	Intel(R)Core(TM)i5-9400F CPU @ 2.90 GHz
显卡	NVIDIA Geforce RTX2060
内存	16.0 GB
编程语言	Python 3.6
深度学习框架	Pytorch 1.6
图像处理库	OpenCV 3

3.3 实验评价指标

目标检测算法的评价指标主要有检测精度以及检测速度。准确率P代表模型能够正确识别物体类别,召回率R代表某一类别检测出来的识别框与真实框的比例,漏检率在数学关系表示为$1-R$,召回率越高越好,漏检率越低越好。平均精度(Average precision, AP)代表某一类别正确识别的数量占该类别所有被识别数的百分比,平均精度均值(Mean average precision, mAP)代表每个类别AP的和除以总类别数,通常将mAP作为模型检测精度的最终指标。上述指标表达式为

$$P=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \tag{10}$$

$$R=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{11}$$

$$\mathrm{AP}=\sum_{k=1}^{n}(r_{k+1}-r_k)\times p_k \tag{12}$$

$$\mathrm{mAP}=\frac{1}{N}\sum_{n=1}^{N}\mathrm{AP}(n) \tag{13}$$

式中:TP(True positive)代表正样本被正确识别为正样本;FP(False positive)代表负样本被错误识别为正样本;FN(False negative)代表正样本被错误识别出负样本,由于本文只针对无人驾驶场景的行人类别,即$N=1$;r_k、r_{k+1}表示PR曲线定积分横坐标上两点,p_k表示纵坐标;在检测速度方面,通常模型检测处理速度FPS(Frame per second)大于30,符合实时检测的标准。

3.4 消融实验

为探究本文不同创新点对检测性能的影响,本文在Caltech行人数据集进行了消融实验。通过表2可以看出,在Caltech数据集上本文的改进方式可以较大提升系统对行人的检测性能。基于密集连接思想可以最大化网络中所有层之间的信息流,增强提取行人特征时深层信息和浅层信息的交互能力,改进的YOLO v3-tiny虽然在检测精度上取得不错的效果,检测准确度上升11.4%,平均精度均值上升6%,但是在漏检率上仍有欠缺。特征提取网络通过级联可以动态适应行人尺度的卷积核后,采用非线性的方法融合来自不同尺寸卷积核的特征,实现动态调整模型感受野,捕获更多的行人特征,行人的漏检率进一步降低3.6%,平均检测精度上升0.7%,在保证实时性的前提下(25 ms/张),无人驾驶场景中的行人检测性能得到较大的提高。

表2 Caltech行人数据集消融实验

Table 2 Caltech pedestrian dataset ablation experiment

Baseline	Dense	SK	P	R	mAP	FPS
√			0.657	0.594	0.63	**120**
√	√		**0.771**	0.659	0.69	**50**
√		√	0.508	**0.712**	0.654	**90**
√	√	√	0.71	0.695	**0.697**	**40**

注:加粗字体为每列最优值。

本文模型初始输入图片大小为$416\times416\times3$,并对DK-YOLO v3-tiny的训练过程进行可视化,通过图8可以看出,该方法在训练至500 epochs时基本处于饱和状态。在图8中,平均损失在训练过程中

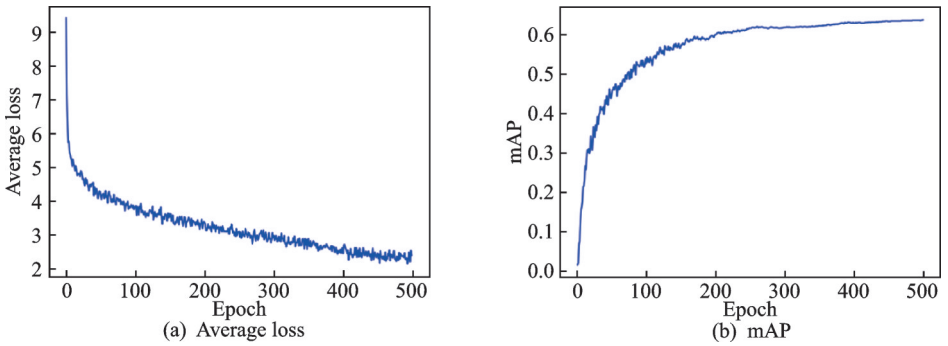


图8 平均损失和mAP趋势图

Fig.8 Training average loss and mAP trend graphs

平滑且稳定下降,到达500 epochs基于处于水平状态,说明本文模型的泛化性较强。DK-YOLO v3-tiny的mAP持续上升,最终处于0.6～0.7的区间保持相对稳定。

3.5 BDD 100K数据集对比实验

BDD 100K数据集包含黎明、白天和夜晚3个阶段,具有较高的挑战性。本文在BDD 100K采用5种方法对比,其中包含两阶段检测算法PANet,编解码网络ENet,经典目标检测算法Densenet201、YOLO v3-tiny、YOLO v4-tiny。Wang等[23]提出PANet在最先进的Mask R-CNN基础上提高信息流的传递效率,增强低层信息的利用率。ENet网络[24]由一个大的编码模块和一个小的解码模块组成,减少特征图分辨率引起的空间信息损失。陈立潮等[25]使用深层次的Dense结构Densenet201网络对车辆目标识别,该模型实现特征提取网络特征层的有效复用。实验结果如表3所示,与这3种检测算法和单阶段目标检测算法YOLO v3-tiny[19]和YOLO v4-tiny[20]相比,本文模型DK-YOLO v3-tiny检测准确率,召回率、平均检测精度可达到0.621、0.397、0.401,验证了本文模型在漏检率、检测精度方面的优越性。

表3 BDD100K常用检测算法性能对比表

Table 3 Comparison results of common detection algorithms on BDD 100K

检测模型	*P*	*R*	mAP
PANet[23]	0.529	0.384	0.373
ENet[24]	0.541	0.322	0.305
Densenet201[25]	0.562	0.29	0.272
YOLO v3-tiny[19]	0.576	0.283	0.284
YOLO v4-tiny[26]	0.330	0.377	0.306
本文模型	**0.621**	**0.397**	**0.401**

3.6 Caltech行人数据集对比实验

为了进一步验证DK-YOLO v3-tiny网络模型的有效性,本文将DK-YOLO v3-tiny和具有代表性的检测方法在Caltech行人数据集进行对比。其中包含两阶段检测器Faster R-CNN(Resnet)[27],单阶段检测器SSD(VGG)[27]、YOLO v3(Darknet53)、Densenet201(Densenet)以及YOLO轻量化模型系列YOLO v3-tiny、YOLO v4-tiny,这些方法包含不同的特征提取网络。实验对比结果如表4所示,可以看出DK-YOLO v3-tiny检测准确率、召回率、平均

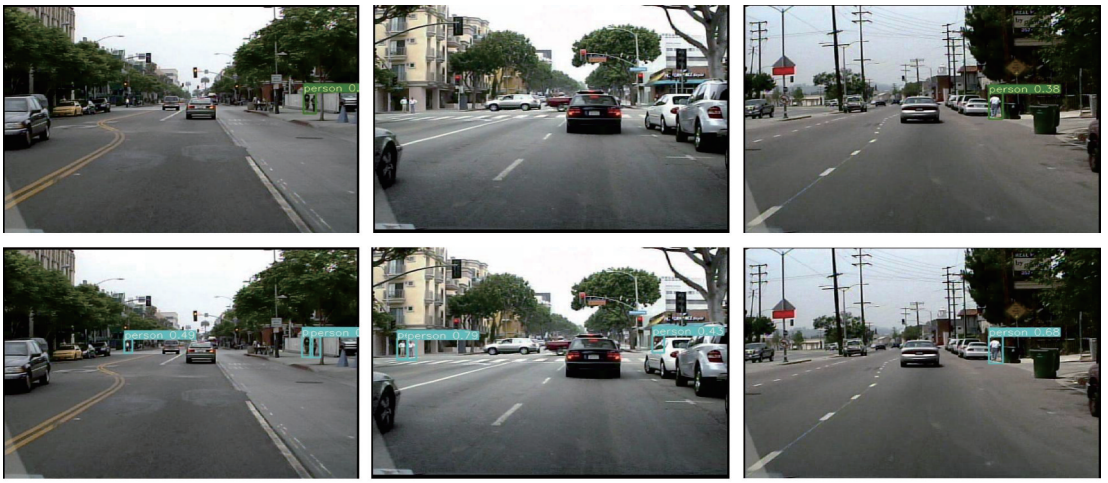
表4 Caltech常用检测算法对比表

Table 4 Comparison results of common detection algorithms on Caltech

检测模型	*P*	*R*	mAP
Faster R-CNN[27]	—	—	0.685
SSD[27]	—	—	0.596
YOLO v3[14]	0.596	0.564	0.574
YOLO v3-tiny[19]	0.657	0.594	0.630
YOLO v4-tiny[26]	0.445	**0.695**	0.644
Densenet201[25]	**0.787**	0.676	0.686
本文模型	0.710	**0.695**	**0.697**

检测精度可达到0.71、0.695、0.697,均达到最优性能。卷积神经网络不同卷积层具有层次性,浅层的语义特征和深层的语义特征往往代表不用的行人特征,选用不同的层进行判别会带来不同的结果,简单的特征提取网络(例如YOLO v3-tiny)不能有效地提取行人特征,所以造成检测效果较差。深层的特征提取网络(例如YOLO v3)在特征传递过程中,小尺度行人的浅层特征在信息传递过程中容易丢失,造成漏检率升高。本文基于密集思想融合深层和浅层特征,并级联动态选择机制的注意力提高模型局部感受野,模型检测时对多尺度变化、小尺度行人具有更强的鲁棒性,实验结果表明本文方法优于当前常用的两阶段、单阶段目标检测方法,在保证实时性的前提下,较大地提升了检测精度并降低了漏检率。

检测结果如图9所示,第1行是原始YOLO v3-tiny实现行人检测的结果图,第2行是本文方法的检测结果图。从图中可以看出YOLO v3-tiny对于小尺度行人漏检较多,多尺度的行人检测精度较低,本文方法在交通场景中面对多尺度、小尺度的行人目标具有更强的鲁棒性,甚至对于被车辆遮挡的行人也可以有效地检测出目标,检测性能良好。对于YOLO v3-tiny漏检、检测精度低的情况有较大的改进,本文方法更加适用于无人驾驶领域,符合实验的预期结果。



(a) Multi-scale pedestrian (b) Occluded pedestrian (c) Small-scale pedestrian

图9 两类模型检测效果对比图

Fig.9 Comparison of the detection effect of two models

4 结束语

本文使用深度学习的方法对交通场景行人进行检测和识别,在YOLO v3-tiny目标检测网络基础上提出一种融合深浅层特征和级联动态选择机制的方法,在保证算法实时性的前提下,解决了无人驾驶场景下行人复杂的尺度变化造成检测精度较低,小尺度行人容易漏检的问题。通过实验结果表明,本文方法在BDD 100K数据集行人漏检率降低11.4%,平均检测精度提高11.7%,在加州理工学院Caltech行人数据集上漏检率降低10.1%,平均检测精度上升6.7%,为无人驾驶行人检测相关技术提供了不同的思路。但是如果面对突变的恶劣天气,只通过摄像头传感器实现行人检测可能会面临巨大的挑战。因此未来的工作包括:研究多模态融合技术,在模型中引入毫米波雷达传感器和视觉进行融合,用传感器融合的方式减少特别恶劣环境对无人驾驶行人检测的影响。

参考文献：

[1] BRUNETTI A, BUONGIORNO D, TROTTA G F, et al. Computer vision and deep learning techniques for pedestrian detection and tracking: A survey[J]. Neurocomputing, 2018, 300: 17-33.

[2] BANSAL P, KOCKELMAN K M. Are we ready to embrace connected and self-driving vehicles? A case study of Texans[J]. Transportation, 2018, 45(2): 641-675.

[3] PRASANNA D, PRABHAKAR M. An efficient human tracking system using Haar-like and hog feature extraction[J]. Cluster Computing, 2019, 22(2): 2993-3000.

[4] HOSSEIN-NEJAD Z, NASRI M. An adaptive image registration method based on SIFT features and RANSAC transform[J]. Computers & Electrical Engineering, 2017, 62: 524-537.

[5] WEI Y, TIAN Q, GUO J, et al. Multi-vehicle detection algorithm through combining Harr and HOG features[J]. Mathematics and Computers in Simulation, 2019, 155: 130-145.

[6] LIN Y, LV F, ZHU S, et al. Large-scale image classification: Fast feature extraction and SVM training[C]//Proceedings of CVPR 2011. USA: IEEE, 2011: 1689-1696.

[7] BAIG M M, AWAIS M M, EL-ALFY E S M. AdaBoost-based artificial neural network learning[J]. Neurocomputing, 2017, 248: 120-126.

[8] XU FENGLIANG, XIA LIU, FUJIMURA K. Pedestrian detection and tracking with night vision[J]. IEEE Transactions on Intelligent Transportation Systems, 2005, 6(1): 63-71.

[9] CERRI P, GATTI L, MAZZEI L, et al. Day and night pedestrian detection using cascade AdaBoost system[C]// Proceedings of 13th International IEEE Conference on Intelligent Transportation Systems. USA: IEEE, 2010: 1843-1848.

[10] GUO Lie, GE Pingshu, ZHANG Mingheng, et al. Pedestrian detection for intelligent transportation systems combining AdaBoost algorithm and support vector machine[J]. Expert Systems with Applications, 2011, 39(4): 4274-4286.

[11] GIRSHICK R. Fast R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision. USA: IEEE, 2015: 1440-1448.

[12] HE K, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision. USA: IEEE, 2017: 2961-2969.

[13] WEI L, DRAGOMIR A, DUMITRU E, et al. SSD: Single shot multibox detector[C]// Proceedings of European Conference on Computer Vision. Cham: Springer, 2016: 21-37.

[14] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. USA: IEEE, 2016: 779-788.

[15] ZHANG L, LIN L, LIANG X, et al. Is faster R-CNN doing well for pedestrian detection?[C]// Proceedings of European Conference on Computer Vision. Cham: Springer, 2016: 443-457.

[16] LAN W, DANG J, WANG Y, et al. Pedestrian detection based on YOLO network model[C]// Proceedings of 2018 IEEE International Conference on Mechatronics and Automation (ICMA). USA: IEEE, 2018: 1547-1551.

[17] ZHANG C W, YANG M Y, ZENG H J, et al. Pedestrian detection based on improved LeNet-5 convolutional neural network [J]. Journal of Algorithms and Computational Technology, 2019. DOI:10.1177/1748302619873601.

[18] ZHANG Yi, SHEN Yongliang, ZHANG Jun. An improved tiny-yolov3 pedestrian detection algorithm[J]. Optik, 2019, 183: 17-23.

[19] ADARSH P, RATHI P, KUMAR M. YOLO v3-Tiny: Object detection and recognition using one stage improved model [C]// Proceedings of 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS). USA: IEEE, 2020: 687-694.

[20] HUANG G, LIU S, VAN DER MAATEN L, et al. Condensenet: An efficient densenet using learned group convolutions [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. USA: IEEE, 2018: 2752-2761.

[21] YANG J, WANG W, LI X, et al. Selective kernel networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. USA: IEEE, 2019: 510-519.

[22] WOJEK C, DOLLAR P, 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): 743-61.

[23] WANG K, LIEW J H, ZOU Y, et al. PaNet: Few-shot image semantic segmentation with prototype alignment[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. USA: IEEE, 2019: 9197-9206.

[24] NAZIR A, CHEEMA M N, SHENG B, et al. OFF-eNET: An optimally fused fully end-to-end network for automatic dense volumetric 3D intracranial blood vessels segmentation[J]. IEEE Transactions on Image Processing, 2020, 29: 7192-7202.

[25] 陈立潮,朝昕,潘理虎,等.基于部件关注DenseNet的细粒度车型识别[J].智能系统学报, 2022,17(2): 402-410.
CHEN Lichao, CHAO Xin, PAN Lihu, et al. Fine-grained vehicle-type identification based on partially-focused DenseNet[J]. CAAI Transactions on Intelligent Systems, 2022,17(2): 402-410.

[26] ALY G H, MAREY M, EL-SAYED A S, et al. YOLO v3 and YOLO v4 for masses detection in mammograms with ResNet and inception for masses classification[C]//Proceedings of International Conference on Advanced Machine Learning Technologies and Applications. Cham: Springer, 2021: 145-153.

[27] 王飞,王林,张儒良,等.基于融合FPN和Faster R-CNN的行人检测算法[J].数据采集与处理, 2019, 34(3): 530-537.
WANG Fei, WANG Lin, ZHANG Ruliang, et al. Pedestrian detection algorithm based on fusion FPN and faster R-CNN[J]. Journal of Data Acquisition and Processing, 2019, 34(3): 530-537.

作者简介:

沙梦洲(1996-),男,硕士研究生,研究方向:人工智能、计算机视觉,E-mail: 243360172@qq.com。

沈韬(1984-),**通信作者**,男,博士,教授,研究方向:太赫兹技术,智能感知与计算,E-mail:shentao@kmust.edu.cn。

曾凯(1985-),男,博士,副教授,研究方向:粒计算,分布式计算。

马倩(1997-),女,硕士研究生,研究方向:FPGA系统设计、模型轻量化。

曾文健(1994-),男,硕士研究生,研究方向:计算机视觉、深度学习。

(编辑:张黄群)