基于双主干网络的雾天交通目标检测方法研究

李习习[1],强 俊[1],刘无纪[1],杜云龙[1],刘 进[1,2]
1. 安徽工程大学 计算机与信息学院,安徽 芜湖 241000
2. 东南大学 计算机网络和信息集成教育部重点实验室, 南京 210096

摘 要:车辆和行人安全监测是城市交通监测的一项重要任务。针对雾霾等复杂恶劣天气条件下,监测采集的图像视觉效果差、噪声高、目标检测困难等问题,提出了一种双主干网络(MobileNets VGG-DCBM Network, MVNet)用于雾天交通目标检测,结构受 PCCN 和 CBNet 网络结构的启发,由改进的深度可分离卷积神经网络 MobileNets 和基于 VGGNet 构建的 VGG-DCBM 网络组成;采用并行方式构建双主干目标检测网络结构,以改进的 MobileNets 为主主干网络,VGG-DCBM 为辅助主干网络,共同提取特征信息,实现不同网络间特征层信息的融合;MVNet 网络结构采用并行方式获取两个不同网络提取的不同特征层信息,通过采用通道拼接的方法实现不同网络特征信息之间的融合,以获得更丰富的细节特征;在 RTTS 和 HazePerson 数据集上,平均精度均值(mean Average Precision, mAP)分别达到 71.50%和 89.84%;实验结果表明:在雾霾等复杂恶劣天气条件下具有较强的鲁棒性且能够准确的检测到车辆和行人,在目标检测性能上优于对比方法。

关键词:雾天交通目标;双主干网络;并行方式;特征融合;通道拼接
中图分类号:TP391.41 **文献标识码**:A doi:10.16055/j.issn.1672-058X.2023.0004.004

Research on Traffic Object Detection Method in Fog Based on Dual Backbone Network

LI Xixi[1], QIANG Jun[1], LIU Wuji[1], DU Yunlong[1], LIU Jin[1,2]
1. School of Computer and Information, Anhui Polytechnic University, Anhui Wuhu 241000, China
2. Key Laboratory of Computer Network and Information Integration Ministry of Education, Southeast University, Nanjing 210096, China

Abstract: Vehicle and pedestrian safety monitoring is an important task of urban traffic monitoring. Aiming at the problems of poor visual effect, high noise, and difficult target detection of the images collected under complex and bad weather conditions such as fog and haze, a dual backbone network (MobileNets VGG-DCBM Network, MVNet) was proposed for traffic object detection in fog. Inspired by the network structure of PCCN and CBNet, this structure was composed of improved depthwise separable convolutional neural network MobileNets and VGG-DCBM network based on VGGNet. The dual backbone object detection network structure was constructed by using the parallel method. The improved MobileNets was the main backbone network and VGG-DCBM was the assistant backbone network to extract the feature information and realize the fusion of feature information between different networks. MVNet network structure adopted the parallel method to obtain the information of different feature layers extracted by the two networks, and realized the fusion of different network feature information by using the method of channel splicing, so as to obtain richer detailed

收稿日期:2022-04-14 **修回日期**:2022-05-12 **文章编号**:1672-058X(2023)04-0025-10
基金项目:国家自然科学基金项目(61801003);安徽省高校优秀拔尖人才培育资助项目(GXYQZD2021123).
作者简介:李习习(1996—),男,安徽淮南人,硕士研究生,从事计算机视觉和模式识别研究.
通讯作者:强俊(1981—),女,安徽芜湖人,副教授,硕士,从事计算机视觉和模式识别研究. Email: chiang_j@ahpu.edu.cn.
引用格式:李习习,强俊,刘无纪,等. 基于双主干网络的雾天交通目标检测方法研究[J]. 重庆工商大学学报(自然科学版),2023,40(4):25—34.
LI Xixi, QIANG Jun, LIU Wuji, et al. Research on traffic object detection method in fog based on dual backbone network[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2023, 40(4): 25—34.

features. On RTTS and HazePerson datasets, the mean average precisions (mAP) reached 71.50% and 89.84%, respectively. The experimental results show that this method has strong robustness in complex bad weather conditions such as fog and haze, and can accurately detect vehicles and pedestrians. It is better than the comparison method in object detection performance.

Keywords: tratfic object in foggy weather; dual backbone network; parallel model; feature fusion; channel splicing

1 引 言

城市车辆和行人的安全监测一直是亟待解决的问题,其中复杂多变的车流和人流,对交通路段的视频监控造成了很大的困扰。尤其是在雾霾和其他恶劣天气环境中的交通事故数量[1-2]明显高于简单环境中的交通事故数量。为了有效提高车辆和行人识别的效率,减少交通事故的发生,车辆和行人的准确识别成为近年来的研究热点。

近年来,基于深度学习[3]的目标检测研究进展迅速,特别是在基于深度卷积神经网络[4](Deep Convolution Neural Networks, DCNNs)的图像目标检测研究中。目前,在目标检测和识别方面取得突破性进展的优秀检测算法主要分为两大类:(i) 基于"two-stage"目标检测方法,如以 R-CNN(Region Convolution Neural Network)系列[5-7]等为代表的神经网络模型。Ren 等[7]提出了一种 Faster R-CNN 网络模型,模型引入了基于 Fast R-CNN[6]的区域建议网络(Region Proposal Network, RPN),网络主要用于生成区域建议,并将区域建议和目标检测识分别引入网络进行训练。此类方法,首先要多次检测生成候选框,然后再对每一个候选框进行分类,因此检测速度较慢,但是目标检测精度有所提高。(ii) 基于"one-stage"目标检测方法,如以 YOLO(You Only Look Once)系列[8-11]、SSD[12](Single Shot MultiBox Detector)等为代表的神经网络模型。特点是"一步到位",仅仅需要将图片送入网络一次就可以预测出对应的目标边界框,检测速度相对较快,但是目标检测精度略有不足。

计算机视觉等相关领域的学者提出了许多目标检测和识别技术。然而,针对一些特殊的应用场景,如:雾、霾等天气条件下的目标检测方法大多都是结合一些去噪的方法。方法主要分两步实现,即先去噪后检测。具体步骤是:去除在复杂恶劣天气条件下拍摄的图像所对应的噪声;利用现有的目标检测方法对去除噪声的图像进行检测。

Li 等[13]选择 Faster R-CNN 和 AOD-Net 进行联合优化训练,提出了先对图像进行去雾处理,然后利用现有的目标检测方法对去雾图像进行检测的方法,以及先对图像进行去雾处理,然后利用自适应目标检测方法对去雾图像进行检测的方法,从而解决了真实雾场景中的目标检测问题。陈琼红等[14]提出将 AOD-Net 除雾算法与 SSD 目标检测算法相结合。方法通过选取浓雾、中雾和薄雾 3 种不同级别的雾气浓度图像,对车辆和行人的检测精度进行分析,证明方法能够有效地检测出雾天环境下的车辆和行人。解宇虹等[15]提出了一种基于图像去雾和目标检测(Joint Learning in Dehazing and Object Detection, DONet)的联合学习框架。DONet 将去雾模型和目标检测模型级联进行联合学习,使网络能够完成图像去雾任务和目标检测任务。方法将去雾与检测相结合,虽然有效的提高了目标检测精度,但是增加目标检测过程中额外的去雾工作量,实时性相对较差。

针对上述问题,提出了一种双主干网络(MobileNets VGG-DCBM Network, MVNet)用于雾天目标检测,结构受 PCCN(Parallel Cross Convolutional Network)[16]和 CBNet(Composite Backbone Network)[17]网络结构的启发,由改进的 MobileNets[18]和基于 VGGNet[19]构建的 VGG-DCBM 网络组成。双主干网络之间采用并行的方式,获取两个网络的不同特征层信息,辅助主干网络是为了提取不同于主主干网络的细节特征。MVNet 网络结构采用并行方式获取两个网络提取的不同特征层的信息,并在不同层实现两个网络之间的通道拼接,从而解决复杂恶劣天气(雾、霾等)下的目标检测问题。在两个有雾自然图像雾天数据集上,平均精度均值分别达到 71.50% 和 89.84%,同时能够快速准确地检测到车辆和行人。

2 方 法

2.1 MVNet 框架

提出了一种双主干网络框架 MVNet。MVNet 的整体框架如图 1 所示,加粗箭头表示通道拼接。使用通道拼接的方法来增强 MobileNets 和 VGG-DCBM 对细节特征提取。主干特征提取网络采用改进的 MobileNets 网络结构,辅助特征提取网络采用 VGG-DCBM 网络结构。

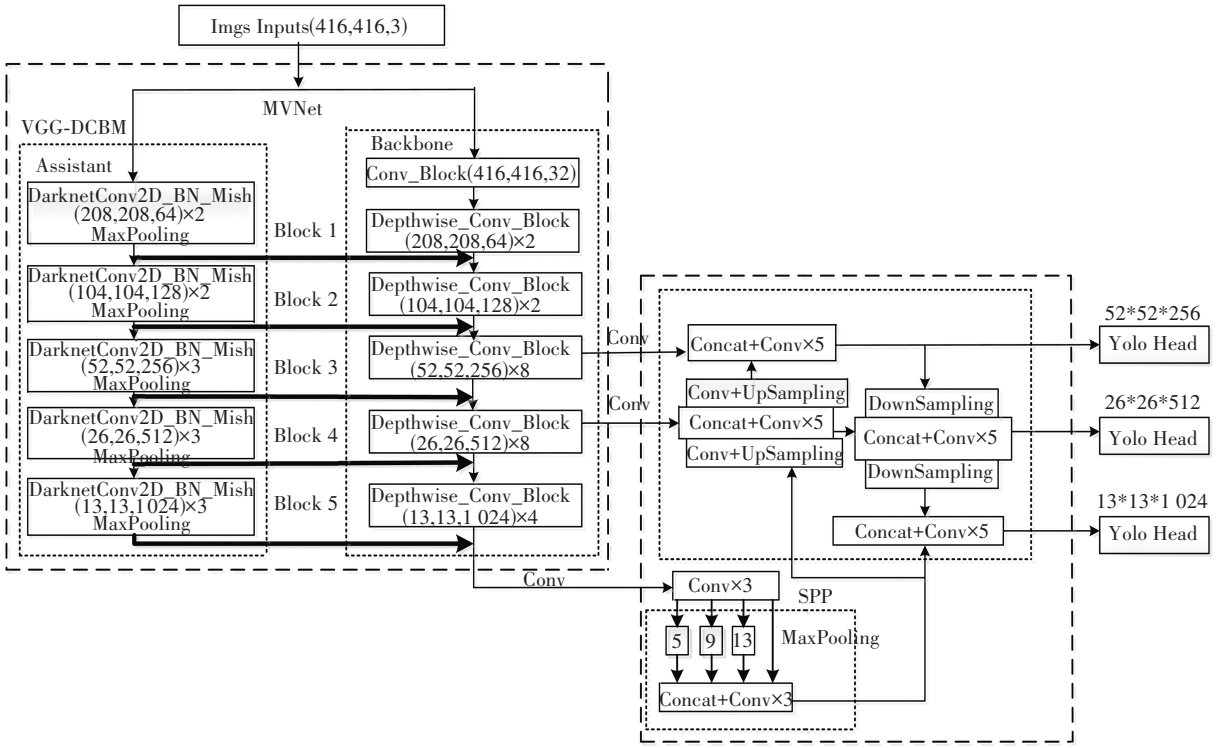


图 1　MVNet 的整体框架

Fig. 1　Architecture of MVNet

2.1.1　主主干网络

MobileNets 的网络结构是基于深度可分离卷积[20]的堆叠设计,深度可分卷积与传统卷积有很大不同。其基本思想是完全分离通道之间的相关性和空间相关性,同时大大减少了计算量和参数量,这样可以减少过多的冗余特征。因此,选择 MobileNets 作为 MVNet 的主主干网络。深度可分离卷积可分解为深度卷积(Depthwise Convolution, DW)和逐点卷积(Pointwise Convolution, PW)。DW 针对每个输入通道使用不同的卷积核,而 PW 使用普通的卷积核,其卷积核大小为 1×1。图 2 显示了深度可分离卷积块,图 3 显示了深度可分离卷积机制。

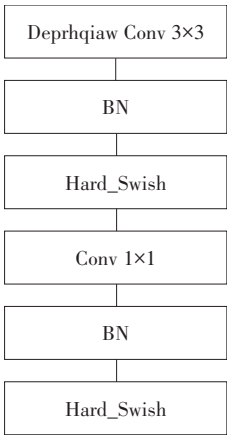


图 2　深度可分离卷积块

Fig. 2　Depthwise separable convolution blocks

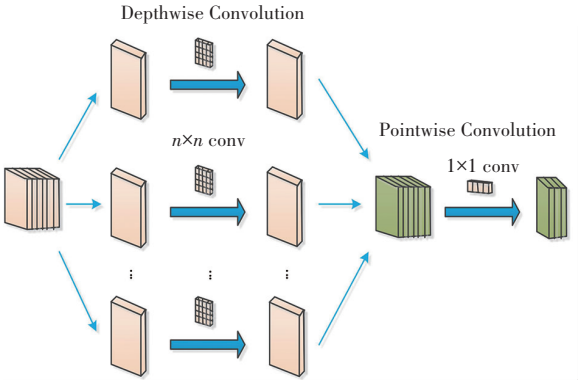


图 3　深度可分离卷积机制

Fig. 3　Depthwise separable convolution mechanism

考虑 MobileNets 网络太浅,在原有的基础上扩大了网络层数,增加了网络的深度。此外,将深度可分离卷积块中的激活函数 ReLU6(x)替换为 *hard_swish*(x)[21]。*hard_swish*(x)激活函数很好地解决了网络中一些神经元永远不会被激活而导致相应的参数无法更新的问题。*hard_swish*(x)激活函数如式(1)所示:

$$h\text{-}swish[x]=x*\frac{\mathrm{ReLU6}(x+3)}{6} \tag{1}$$

2.1.2　辅助主干网络

借鉴 PCCN[16]和 CBNet[17]的网络结构引入辅助主干网络,帮助主主干网络提取更多有用的特征信息。VGG-DCBM 是由 VGG16 改进而来的,将 VGG16 网络的末端所有全连接层全部舍弃,只保留了卷积层和池化层。

此外,还将在 VGG16 中用 DarknetConv2D_BN_Mish(DCBM)卷积块替换传统的卷积块,并在卷积块之后添加池操作。其目的是通过计算小区域内的值来减少冗余特征,然后只输出一个值作为特征从而有效地避免了网络学习过多的冗余特征。图 4 显示了 VGG-DCBM 网络结构,在这个网络结构中,卷积块是由 DarknetConv2D+批标准化+激活函数(Mish)+ MaxPooling 组成。

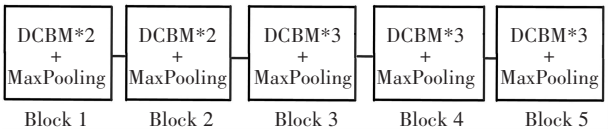


图 4 VGG-DCBM 网络结构

Fig. 4 VGG-DCBM network structure

卷积块中的激活函数采用 Mish[22],如式(2)所示:

$$y=x+\tanh(\ln(1+\exp(x))) \tag{2}$$

2.2 特征融合

在计算机视觉领域,人们提出了许多网络来解决小目标或模糊目标的分类检测性能问题。这些网络主要是从特征层面进行改进,使网络学习更具表现力。受 PCCN[16] 结构的启发,采用并行逻辑将两个不同的网络结合起来,形成一种双网络结构。同时,使用通道拼接的方法来连接不同的网络层,以获得更多的特征信息。通道拼接操作不会改变特征图的大小,只会增加通道的数量。在辅助特征提取网络中,每个阶段的结果都可以看作是一个更高层次的特征;每个功能级别的输出是主干特征提取网络输入的一部分,并流向后续主干的并行阶段。这种并行操作可以融合多个高级和低级特征,以生成更丰富的特征表示。图 5 显示了特征融合总体框架。这个过程如式(3)和式(4)所示:

$$f_o=f_a\oplus f_b \tag{3}$$

$$f_O=\varphi(f_o) \tag{4}$$

其中,$\oplus$是特征通道拼接融合的过程,f_a 表示当前阶段辅助主干网络的输出特征,f_b 表示当前阶段主主干网络的输出特征,f_o 表示特征拼接融合的结果,f_O 作为主主干网络下一层的输入值。从 f_o 到 f_O 的过程通过通道调整进行。如式(4)所示,φ 是 1×1 的卷积运算。

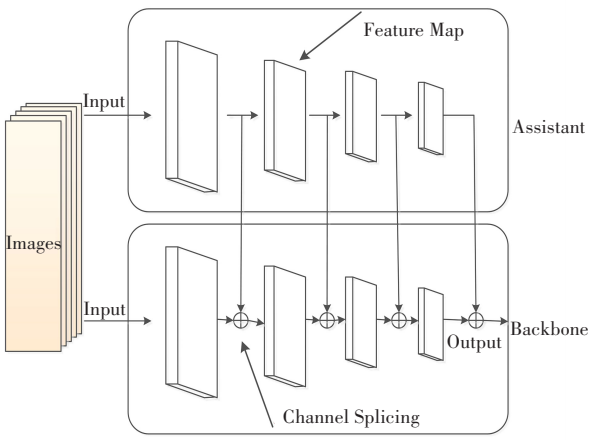


图 5 特征融合总体框架

Fig. 5 Overall framework of feature fusion

2.3 损失函数设计

目标检测任务的损失函数由位置损失、置信度损失和分类损失之和组成。位置损失:是描述检测框中被检测对象的具体位置。选取 CIOU Loss[23] 作为位置损失的损失函数,这使得检测框的回归速度和准确性更高。IOU 的定义如式(5)所示:

$$IOU=\frac{|B\cap B^{gt}|}{|B\cup B^{gt}|} \tag{5}$$

其中,B 为检测框的面积,B^{gt} 为真实框的面积。因此,CIOU Loss 的定义如式(6)所示:

$$L_{\mathrm{CIOU}}=1-IOU+\frac{\rho(b,b^{gt})}{c^2}+\alpha v \tag{6}$$

其中,α 是平衡正数的参数,v 是确保长宽比具有一致性的参数,b 是中心点,$\rho(b,b^{gt})$ 分别代表了预测框和真实框的中心点的欧式距离,c 代表的是能够同时包含预测框和真实框的最小闭包区域的对角线距离。α 和 v 的定义如式(7)和式(8)所示:

$$\alpha=\frac{v}{1-IOU+v} \tag{7}$$

$$v=\frac{4}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}}-\arctan\frac{w}{h}\right)^2 \tag{8}$$

其中,w 和 h 为预测框的宽高,w^{gt} 和 h^{gt} 为真实框的宽和高。

置信度损失:是判断检测框中被检测对象是否存在或局部存在(分为有意义和无意义)。选取 binary cross-entropy[24] 作为置信度损失的损失函数。置信度定义为 $P_{(obj)}*IOU_{pre}^{gt}$,$P_{(obj)}\in\{0,1\}$。若预测框中存在被检测对象,则 $P_{(obj)}=1$,此时置信度等于 IOU。若预测框中无被检测对象,则 $P_{(obj)}=0$,此时置信度等于 0。置信度损失的定义如式(9)所示:

$$L_{\mathrm{Con}}=\frac{-\sum_{i=1}^{N}[\beta^{obj}y_i\log(p(y_i))+\beta^{nobj}(1-y_i)\log(1-p(y_i))]}{N} \tag{9}$$

其中,N 表示样本总数,β^{obj} 和 β^{nobj} 是惩罚项,表示当检测框中没有检测到的对象时,对整个损失函数的贡献,y_i 为模型估计的标签($y_i\in\{0,1\}$),$p(y_i)$ 是所有 N 个样本中 $y_i=1$ 的预测值($p(y_i)\in[0,1]$)。

分类损失:是针对被检测对象的类别而言的,选取 Focal Loss[25-26] 作为分类损失的损失函数。分类损失的定义如式(10)所示:

$$L_{\mathrm{Cla}}=-\lambda_c(1-p_c)^{\gamma}\log(p_c) \tag{10}$$

其中,λ_c 为平衡变量($\lambda_c=0.25$),$\lambda_c\in[0,1]$,γ 为可调节聚焦参数($\gamma=2$),$\gamma\geqslant 0$。

p_c 定义如式(11)所示:

$$p_c=\begin{cases}p & \text{if } z=1\\ 1-p & \text{otherwise}\end{cases} \tag{11}$$

其中,z 为真实框类($y_i\in\{-1,+1\}$),p 是由模型估计的标签为 $z=1$ 类的概率($p\in[0,1]$)。

3　实验验证

3.1　实验设计

实验环境采用 Tensorflow2.2 机器学习框架,在操作系统为 Windows10\64 位,CPU 为 Intel(R) Core(TM) i9-10940X,主频为 3.31GHz\14 核,运行内存为 64 GB,GPU 为 NVIDIA GeForce RTX3080 的实验环境背景下进行的。

3.2　实验参数

为了减少网络模型的收敛时间以及提升生成坐标框的准确性,采用 K-means 聚类算法[27]针对所需的实验数据生成 archors 坐标框为[(12,16),(19,36),(40,28),(36,75),(76,55),(72,146),(142,110),(192,243),(459,401)]。训练阶段将数据集中不同尺寸图片统一转换为 416×416×3,每次训练设置 100 个 Epoch。采用分段方式,前 0~50 个 Epoch 的批量(batch size)设置为 16,初始学习率(learning rate)设置为 1×10^{-3},设置损失函数 3 次不下降时,以原学习率为基准的 50%自动调节学习率,最低降为 1×10^{-6};后 50~100 个 Epoch 的批量(batch size)设置为 8,初始学习率设置为 1×10^{-4},设置损失函数 5 次不下降时,以原学习率为基准的 50%自动调节学习率,最低降为 1×10^{-6}。此外,采用早停(early stopping)机制,在损失函数多次不下降时自动结束训练。在测试阶段,设置 *IoU* 阈值为 0.5 的非极大抑制(NMS)[28],对训练结果进行测试。

3.3　实验数据集

使用的是公开的有雾自然图像数据集(Real-world Task-driven Testing Set, RTTS[29])以及雾天行人图像数据集 HazePerson[30]。

(1) RTTS 数据集雾天图片共有 4 322 张,五分类,分别是 bus, bicycle, car, motorbike and person,取 3 889 张作为训练集,433 张作为测试集。由于雾天环境复杂多样,所面临的检测难度也有所不同;因此,在测试集中挑选出薄雾、中雾和浓雾 3 种不同等级雾气浓度图片进行检测结果分析。

(2) HazePerson 数据集共有图片 1 195 张,属于单一类别,取 1 075 张作为训练集,120 张作为测试集。由于数据集图像数量不够,目标对象单一,采用水平翻转(fliplr), 对比度(contrast)两种图像增强技术来扩充数据集;因此,在测试集中挑选出水平翻转图像、对比度图像和原始图像进行检测结果分析。

3.4　评估方法

选择了目标检测中常用的评估方法:精确度(Precision, P_r),召回率(Recall, R_e),*AP*(Average Precision),*mAP*(mean Average Precision)和 *FPS*(Frames Per Second)来分析网络模型的检测效率和性能。P_r 和 R_e 的定义如式(12)和式(13)所示:

$$P_r=\frac{TP}{TP+FP} \tag{12}$$

$$R_e=\frac{TP}{TP+FN} \tag{13}$$

其中,*TP* 表示正确预测的对象边界框的数量。*FP* 表示无法预测的对象边界框的数量。*FN* 表示遗漏的预测对象真实框的数量。

对于多目标检测算法中,*mAP* 是衡量多类目标检测结果的平均精度。*mAP* 是所有类别的平均 *AP* 值,计算方法如式(14)和式(15)所示:

$$AP=\int_0^1 P_r(R_e)\,d_{R_e} \tag{14}$$

$$mAP=\frac{1}{M}\sum_{k=1}^{M}AP_k \tag{15}$$

其中,*M* 是目标类别的数量。

3.5　实验结果分析

为了研究模型在雾霾等复杂恶劣天气条件下对车辆和行人的检测效率,比较了该模型与其他常用神经网络模型的性能。网络模型设计的初衷是采用端到端的设计理念,即直接检测有雾图像,进而可以省去前期对雾霾图片进行去雾霾操作,这样可以降低工作量,减少模型对图片的推理时间,增加检测的连贯性和实时性。

首先,将 MVNet 与“two-stage”目标检测方法中表现较好的 Faster RCNN-resnet101 进行了比较。通过分析表 1 和表 2 中的实验结果可知,MVNet 的 *mAP* 值与 Faster RCNN-resnet101 相当,但是,在图像推理速度(*FPS*)方面,MVNet 在不同的数据集上分别达到 19.61 fps 和 20.60 fps,大约是 Faster RCNN-resnet101 的 3.5 倍。因此,提出的 MVNet 方法在目标检测的实时性方面具有明显的优势。

表 1　不同的检测方法在 RTTS 数据集上的检测结果

Table 1　Detection results of different detection methods on RTTS dataset

Method	*mAP*/%	R_e/%	P_r/%	*FPS*
SSD-vgg16	67.39	51.00	85.53	55.13
YOLOv4-CSPdarknet53_tiny	65.21	47.53	85.13	36.66
YOLOv3-Darknet53	68.36	56.63	82.92	33.55
Faster RCNN-resnet101	72.10	61.78	85.69	5.60
YOLOv4-CSPdarknet53	69.10	47.44	86.76	25.20
MVNet	71.50	72.56	95.31	19.61

表 2 不同的检测方法在 HazePerson 数据集上的检测结果

Table 2 Detection results of different detection methods on HazePerson dataset

Method	*mAP*/%	R_e/%	P_r%	*FPS*
SSD-vgg16	83.12	69.33	92.34	56.42
YOLOv4-CSPdarknet53_tiny	83.27	74.12	92.80	40.42
YOLOv3-Darknet53	88.42	82.43	93.14	34.73
Faster RCNN-resnet101	89.34	83.07	93.31	6.02
YOLOv4-CSPdarknet53	88.47	82.43	92.47	25.97
MVNet	89.84	84.66	95.19	20.60

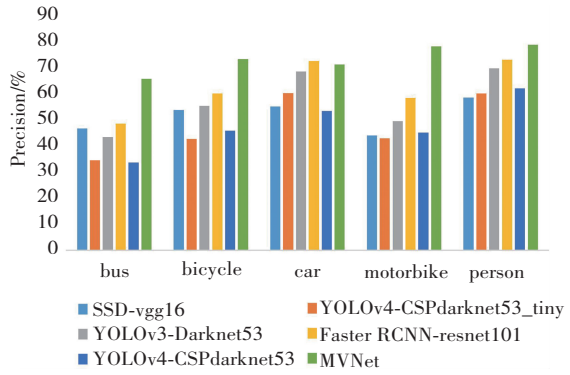
其次,将 MVNet 与“one-stage”目标检测方法进行了比较。由图 6(a)和图 6(b)以及表 1 和表 2 的实验结果可知,方法与“one-stage”目标检测方法 SSD-vgg16、YOLOv4-CSPdarknet53 _ tiny、YOLOv3-Darknet53 和 YOLOv4-CSPdarknet53 相比,在精确度、召回率和 *mAP* 性能指标上有明显的提升。

最后,将分析网络结构,并验证方法的先进性。

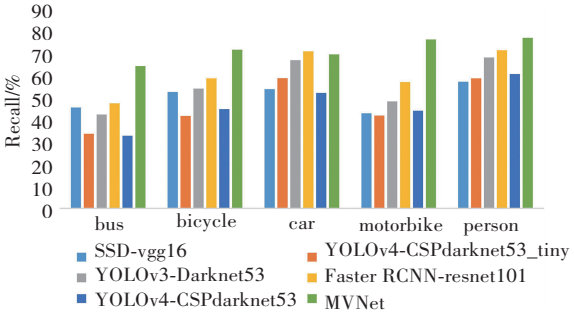
(1) 主主干网络的改进与未改进之间的性能比较。由图 7(a)和图 7(b)中的实验结果可知,改进的 MobileNets 比未改进的 MobileNet 在精确度上有明显提高,而且在召回率方面前者也优于后者。同时,由表 4 可知,在单分类数据集 HazePerson 上改进的 MobileNets 依然具有较强的鲁棒性。

(2) 单一网络结构与双网络结构之间的性能比较。由表 3 和表 4 的实验结果分析可知,YOLOv4-CSPDarknet53 _ vgg-dcbm 的 *mAP* 高于 YOLOv4-CSPdarknet53。此外,通过分析不同数据集上的实验结果,在主干 CSPdarknet53 网络结构的基础上引入辅助主干网络可以提升目标检测方法的性能。

(3) 双网络结构之间的性能比较。通过分析表 3 至表 4 的实验结果可知,与 YOLOv4-CSPDarknet53_vgg-dcbm 方法相比,提出的方法在目标检测的结果上有很大的提高。其中从图 7(a)和图 7(b)的实验结果分析上,可以发现 MVNet 在精确度和召回率方面有较大的提升。尤其是在召回率方面,在某些目标类别(如 bus and bicycle),几乎是 YOLOv4-CSPDarknet53_vgg-dcbm 的 2 倍。此外,在不同数据集上,可以发现提出的方法在图像推理速度上比 YOLOv4-CSPDarknet53_vgg-dcbm 大约高出 5 fps。



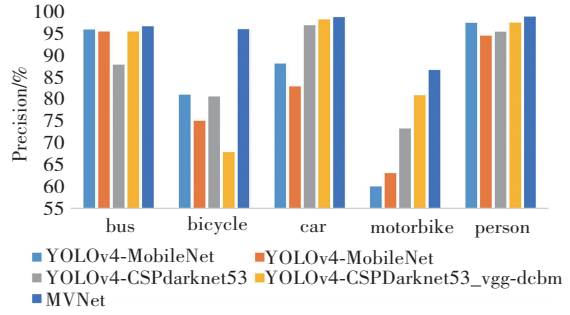
(a) 精确度



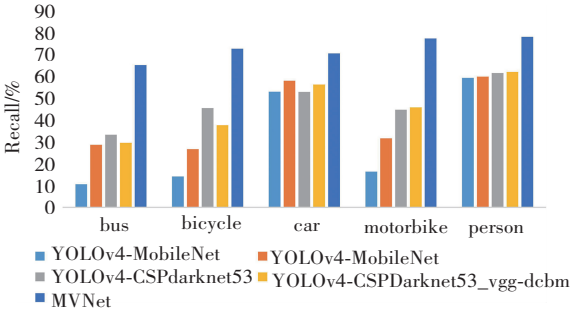
(b) 召回率

图 6 不同的检测方法在 RTTS 数据集上的表现

Fig. 6 Performance of different detection methods on RTTS dataset



(a) 精确度



(b) 召回率

图 7 改进的不同检测方法在 RTTS 数据集上的表现

Fig. 7 Performance of different improved detection methods on RTTS dataset

表 3　改进的不同的检测方法在 RTTS 数据集上的检测结果

Table 3　Detection results of different improved detection methods on RTTS dataset

Method	*mAP*/%	R_e/%	P_r/%	*FPS*
YOLOv4-MobileNet	58.63	33.09	84.44	33.35
YOLOv4-MobileNets	60.32	38.34	82.14	26.38
YOLOv4-CSPdarknet53	69.10	47.44	86.76	25.20
YOLOv4-CSPDarknet53_vgg-dcbm	69.79	46.13	87.93	12.39
MVNet	71.50	72.56	95.31	19.61

表 4　改进的不同检测方法在 HazePerson 数据集上的检测结果

Table 4　Detection results of different improved detection methods on HazePerson dataset

Method	*mAP*/%	R_e/%	P_r/%	*FPS*
YOLOv4-MobileNet	73.57	23.32	61.93	38.06
YOLOv4-MobileNets	79.64	63.58	93.43	30.00
YOLOv4-CSPdarknet53	88.47	82.43	92.47	25.97
YOLOv4-CSPDarknet53_vgg-dcbm	88.74	73.16	95.03	16.23
MVNet	89.84	84.66	95.19	20.60

4　实验仿真测试

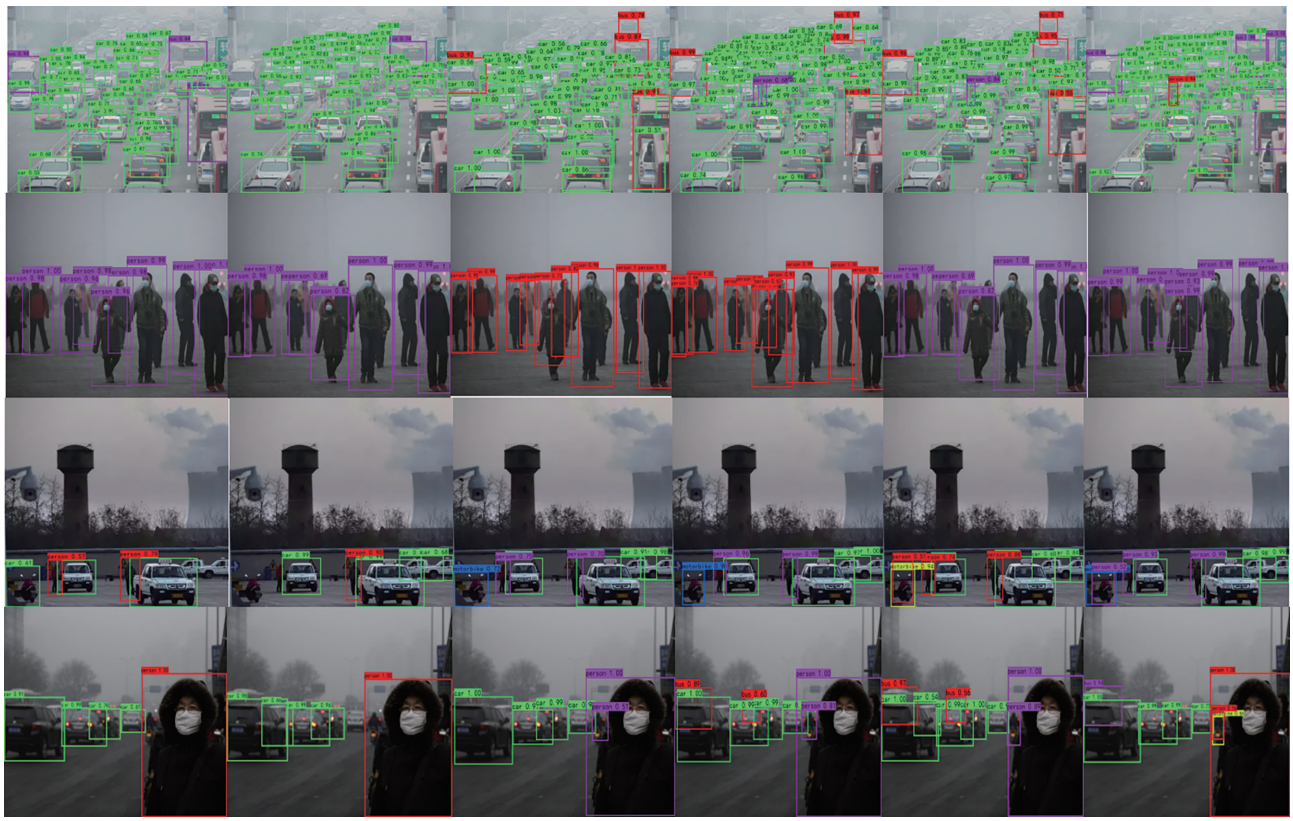
图 8、图 9 和图 10 显示了 MVNet 和几个现有目标检测方法在 RTTS 测试集中选择的 3 种不同雾浓度水平下的检测结果比较。

图 8 显示了在薄雾环境下的检测结果。尽管在不同场景下而且光线照射不均匀,但是所提出的 MVNet 仍有较好的检测性能。

图 9 显示了在中雾环境下的检测结果。在道路场景中,提出的方法在目标比较密集的情况下都有较好的检测效果。在行人与车辆繁多的场景中识别效果尤为突出。此外,在密集小目标的远距离场景中,例如远距离拍摄的车辆,MVNet 都能较好的识别出车辆的类别。

图 10 显示了在浓雾环境下的检测结果。在浓雾环境下由于光线很弱导致检测到的目标较少。虽然在密集小目标检与识别到的目标数量略有减少,但是 MVNet 仍然能够在繁杂的交通道路上有较好的检测效果。

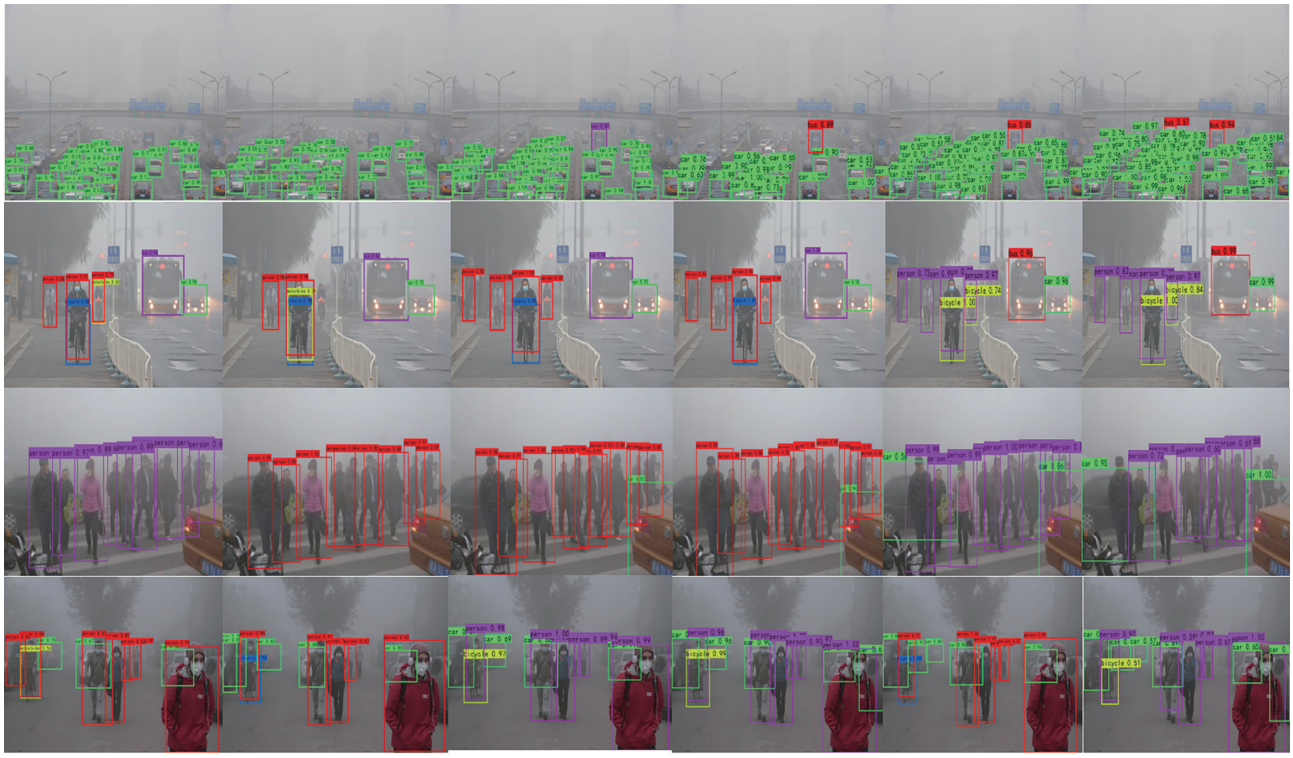
图 11 显示了在 HazePerson 测试集中选择的 3 种图像类型下,MVNet 和几个现有网络的检测结果比较。在不同的角度和对比度下,该网络框架也具有较高的检测效果。



(a) SSD-vgg16　(b) YOLOv4-CS-darknet 53_tiny　(c) YOLOv3-Dark-net 53　(d) YOLOv4-CSP-darknet 53　(e) Faster RCNN-resnet 101　(f) MVNet

图 8　薄雾环境下的检测结果

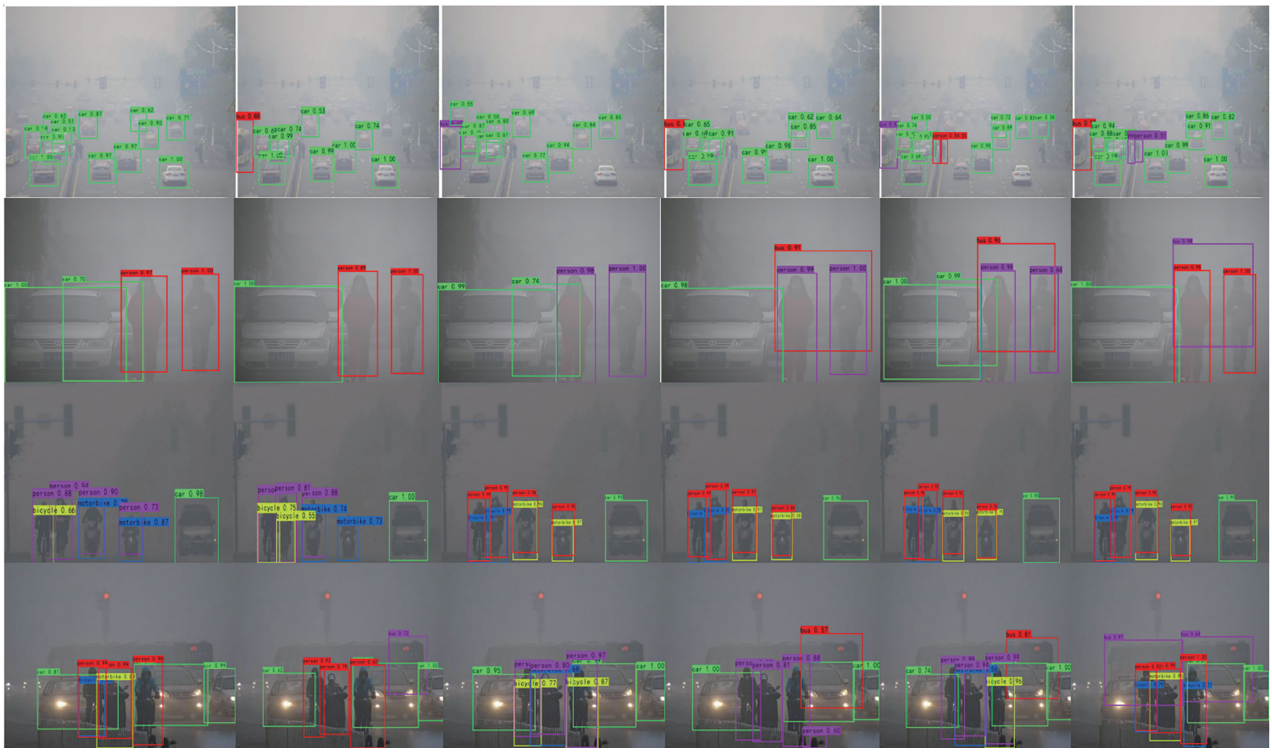
Fig. 8　Detection results in mist environment



(a) SSD-vgg16 (b) YOLOv4-CS -darknet 53_tiny (c) YOLOv3-Dark -net 53 (d) YOLOv4-CSP -darknet 53 (e) Faster RCNN- resnet 101 (f) MVNet

图 9 中雾环境下的检测结果

Fig. 9 Detection results in medium fog environment



(a) SSD-vgg16 (b) YOLOv4-CS -darknet 53_tiny (c) YOLOv3-Dark -net 53 (d) YOLOv4-CSP -darknet 53 (e) Faster RCNN- resnet 101 (f) MVNet

图 10 浓雾环境下的检测结果

Fig. 10 Detection results in dense fog environment

(a) SSD-vgg16　(b) YOLOv4-CS -darknet 53_tiny　(c) YOLOv3-Dark -net 53　(d) YOLOv4-CSP -darknet 53　(e) Faster RCNN-resnet 101　(f) MVNet

图 11　不同角度和不同对比度的检测结果

Fig. 11　Detection results at different angles and different contrasts

综上所述,在浓雾环境中,稠密小目标检测和识别的目标数量略有减少;但在薄雾和中雾环境中,被检测和识别到的目标数量显著增加,检测和识别的准确率较高,漏检率较低。

5　结束语

针对复杂恶劣天气(雾、霾等)下的车辆和行人检测问题,提出了一种双主干网络 MVNet 用于雾天交通目标检测,以提高复杂恶劣天气(雾、霾等)下的目标检测性能。MVNet 网络结构是由主主干网络 MobileNets 和辅助主干网络 VGG-DCBM 组成。辅助主干网络提取图像中的一系列特征,然后通过并行方式获得两个网络的特征层信息,并实现不同网络特征层信息之间的通道拼接融合,以获得更详细的特征,从而获得更好的检测效果。在未来工作中,为提高密集小目标检测和识别的准确率,考虑进行去雾处理和图像增强,以进一步提高复杂恶劣天气(雾、霾等)下目标识别的效率和识别速度。

参考文献(References):

[1] KHORSHIDI A, AINY E, SOORI H. Iranian road traffic injury project: assessment of road traffic injuries in Iran in 2012[J]. Injury Prevention, 2016, 22(2): 172—179.

[2] DAVIDSON P, DHARMARATNE S. Disastrous but preventable: road traffic accidents[J]. Health Care for Women International, 2016, 37(7): 706—706.

[3] WANG J, ZHANG T J, CHENG Y, et al. Deep learning for object detection: a survey[J]. Computer Systems: Science & Engineering, 2021, 38(2): 165—182.

[4] ASIFULLAH K, ANABIA S, UMME Z, et al. A survey of the recent architectures of deep convolutional neural networks [J]. Artificial Intelligence Review, 2020, 53(8): 5455—5516.

[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, 2016, 38(1): 142—158.

[6] GIRSHICK R. Fast R-CNN [C]//Proceedings of the IEEE International Conference on Computer Vision. Santiago, Chile, 2016: 1440—1448.

[7] 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 and Machine Intelligence, 2017, 39(6): 1137—1149.

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

[9] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Washington, USA, 2017: 6517—6525.

[10] LI G F, LIU X, TAO B, et al. Research on ceramic tile defect detection based on YOLOv3[J]. International Journal of Wireless and Mobile Computing, 2021, 21(2): 128—133.

[11] YAO G, SUN Y J, WONG M P, et al. A real-time detection method for concrete surface cracks based on improved YOLOv4 [J]. Symmetry, 2021, 13(9): 1716—1723.

[12] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//Proceedings of the European Conference on Computer Vision. Springer, Cham, 2016: 21—37.

[13] LI B Y, PENG X L, WANG Z Y, et al. AOD-Net: all-in-one dehazing network for dehazing and beyond[C]//Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy, 2017: 4780—4788.

[14] 陈琼红, 冀杰, 种一帆, 等. 基于AOD-Net和SSD的雾天车辆和行人检测[J]. 重庆理工大学学报(自然科学), 2021, 35(5): 108—117.
CHEN Qiong-hong, JI Jie, CHONG Yi-fan, et al. Vehicle and pedestrian detection based on AOD-Net and SSD algorithm in hazy environment[J]. Journal of Chongqing University of Technology(Natural Science Edition), 2021, 35(5): 152—156.

[15] 解宇虹, 谢源, 陈亮, 等. 真实有雾场景下的目标检测[J]. 计算机辅助设计与图形学学报, 2021, 33(5): 733—745.
XIE Yu-hong, XIE Yuan, CHEN Liang, et al. Object detection in real-world hazy scene[J]. Journal of Computer-Aided Design & Computer Graphics, 2021, 33(5): 733—745.

[16] ZHANG Y, CHEN X, GUO D, et al. PCCN: parallel cross convolutional neural network for abnormal network traffic flows detection in multi-class imbalanced network traffic flows[J]. IEEE Access, 2019, 23(7): 119904—119916.

[17] LIU Y D, WANG Y T, WANG S W, et al. CBNet: a novel composite backbone network architecture for object detection [C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 11653—11660.

[18] CHONGKE B, WANG J M, DUAN Y L, et al. Mobilenet based apple leaf diseases identification[J]. Mobile Networks and Applications, 2022, 27(1): 172—180

[19] TAKISAWA N, YAZAKI S, ISHIHATA H. Distributed deep learning of ResNet 50 and VGG16 with pipeline parallelism [C]//Proceedings of the 2020 Eighth International Symposium on Computing and Networking Workshops, Naha, Japan, 2020: 24—27.

[20] QIN Y Y, CAO J T, JI X F. Fire detection method based on depthwise separable convolution and YOLOv3 [J]. International Journal of Automation and Computing, 2021, 18 (2): 300—310.

[21] HOWARD A, PANG R, ADAM H, et al. Searching for mobilenetv3 [C]//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 2019: 1314—1324.

[22] MATHAYO, PETER B, KANG D K. Beta and alpha regularizers of mish activation functions for machine learning applications in deep neural networks[J]. The International Journal of Internet, Broadcasting and Communication, 2022, 14(1): 136—141.

[23] WANG X F, SONG J Y. ICIoU: improved loss based on complete intersection over union for bounding box regression[J]. IEEE Access, 2021(9): 105686—105695.

[24] ZADEH S G, SCHMID M. Bias in cross-entropy-based training of deep survival networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(9): 3126—3137.

[25] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [C]//Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 2017: 2999—3007.

[26] HUANG S C, LE T H, JAW D W. DSNet: joint semantic learning for object detection in inclement weather conditions[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(8): 2623—2633.

[27] FATEMEH M, HAMID S. An improved k-means algorithm for big data[J]. IET Software, 2022, 16(1): 48—59.

[28] LUO Z K, FANG Z, ZHENG S X, et al. NMS-Loss: learning with non-maximum suppression for crowded pedestrian detection[C]//Proceedings of the 2021 International Conference on Multimedia Retrieval, 2021: 481—485.

[29] LI B Y, REN W Q, FU D P, et al. Benchmarking single-image dehazing and beyond[J]. IEEE Transactions on Image Processing, 2019, 28(1): 492—505.

[30] LI G F, YANG Y F, QU X D. Deep learning approaches on pedestrian detection in hazy weather[J]. IEEE Transactions on Industrial Electronics, 2020, 67(10): 8889—8899.

责任编辑:田 静