

引用格式: ZHANG Yinhui, ZHANG Pengcheng, HE Zifen, et al. Lightweight Real-time Detection Model of Infrared Pedestrian Embedded in Fine-scale[J]. Acta Photonica Sinica, 2022, 51(9):0910001

张印辉,张朋程,何自芬,等. 红外行人目标精细尺度嵌入轻量化实时检测[J]. 光子学报, 2022, 51(9):0910001

红外行人目标精细尺度嵌入轻量化实时检测

张印辉,张朋程,何自芬,王森

(昆明理工大学 机电工程学院, 昆明 650500)

摘 要:针对红外图像颜色、纹理等信息不够丰富,导致检测精度相较于可见光图像低;夜间红外行人目标检测模型参数量大、依赖高性能 GPU 资源,导致检测速度慢等问题,提出一种融合行人目标精细尺度嵌入的多检测层、轻量化实时检测模型。首先为了获取更精确的红外行人位置特征,在原始 YOLOv4-tiny 结构上设计了 64×64 精细尺度检测层并增加残差结构加深主干网络,以融合红外行人的位置特征;其次针对红外行人目标宽高比相对固定的特点,应用 K-means++ 聚类分析出适用于红外行人检测的先验框预置参数;最后为了减少模型参数,通过通批量归一化层道剪枝实现模型轻量化,同时利用知识蒸馏算法完成 TIPRD 的微量调控。实验结果表明,轻量化红外行人实时检测模型检测速度达到了 88.7 帧/s,平均检测精度达到 89.2%,模型大小为 4 MB,相较于 YOLOv4-tiny 平均检测精度提升了 8.6%,模型大小缩小了 19.5 MB,相较于 YOLOv4 缩小了 264 MB。在 Jetson Nano 移动开发平台部署该模型验证了实际工程应用的有效性,对开发汽车辅助驾驶系统以降低夜间交通事故发生率具有重要意义。

关键词:红外检测;深度学习;多检测层;模型剪枝;知识蒸馏

中图分类号: TP391.4

文献标识码: A

doi: 10.3788/gzxb20225109.0910001

0 引言

世界经济的发展促使全球汽车保有量逐年攀升,因此由交通所带来的各种安全事故也变得更加频发。世界卫生组织最新调查表明,大多数国家的人口死亡人数中交通事故死亡占比率较大,其中夜间发生交通事故比例非常高,大约是白天的三倍^[1-3]。

为了减少夜间环境下交通安全事故的发生,各种增强驾驶安全性的辅助驾驶技术被广泛地应用,如超声波测距、毫米波雷达、摄像机辅助驾驶、自动驾驶场景构建测试^[4]等。综合各类技术对比,摄像机辅助驾驶技术是减少碰撞事故发生的重要手段之一^[5]。在夜间环境下,红外摄像机更是相较于可见光摄像机有着独特的优势,一方面普通相机的可见光成像容易被光照所影响,光照不足的夜间环境下获取的图像质量会严重下降、成像模糊,使得行人检测变得异常困难,而基于物体自身热辐射和反射成像的红外摄像机技术^[6],可以在不受光照条件的影响下实现无障碍夜视。另一方面,红外成像设备成本的逐年降低使其应用场景逐渐普遍,因此基于夜间红外摄像机的行人检测技术在车辆辅助驾驶中具有重要的研究价值和广阔的市场应用前景^[7]。

近年来,研究人员在夜间红外行人检测方面投入了大量的工作,如赵斌^[8]等利用注意力模块替代特征金字塔网络中传统的上采样模块,生成基于卷积特征的局部显著图,可以有效抑制不相关区域的特征响应,突出图像局部特性。蔡伟^[9]等在单阶段检测器(You Only Look Once, YOLO)检测框架的基础上,通过减少下采样次数,结合跨阶段局部模块、Focus 结构和空间金字塔池化结构设计了特征提取网络,在检测红外弱小

基金项目:国家自然科学基金(Nos.62061022, 62171206, 52065035, 61761024)

第一作者:张印辉(1977—),男,教授,博士,主要研究方向为图像处理、机器视觉及机器智能等。Email: yinhui_z@163.com

通讯作者:何自芬(1976—),女,副教授,博士,主要研究方向为图像处理、机器视觉及机器智能等。Email: zyhhzf1998@163.com

收稿日期:2022-03-04;录用日期:2022-04-21

<http://www.photon.ac.cn>

目标时具有较高的准确率和检测速度。

这些红外行人检测算法都根植于现有的行人检测算法,其中现有的行人检测算法又由两个分支组成。一个分支是传统的机器学习算法,但传统的机器学习算法存在着检测速度慢、精度低的问题;另一分支是深度学习,这种方法相较于传统的机器学习算法检测效果得到了显著的提升。目前基于深度学习的主流行人检测算法有区域卷积神经网络(Region Convolutional Neural Network, R-CNN)^[10]、Fast R-CNN^[11]、Faster R-CNN^[12]、单次多目标框检测器(Single Shot Multibox Detector, SDD)^[13]、YOLO系列^[14-17]等。通过添加注意力模块学习特征通道间关系和特征图空间信息,增强对行人目标可视区域的特征提取。众多检测算法中单阶段检测的YOLO系列凭借着较快的检测速度展现了极强大的研究潜力。邹梓吟^[18]通过添加注意力模块学习特征通道间关系和特征图空间信息,增强对行人目标可视区域的特征提取,提高了在遮挡行人检测方面的精度。白中浩等^[19]应用交叉熵损失函数替代Yolov2模型中的sigmoid激活函数,采用的多层特征图融合方法提升了复杂背景行人及小尺寸行人的检测精度。邓杰等^[20]在Yolov3的基础上融合了空间注意力和通道注意力机制,缓解了行人间遮挡的问题。何东等^[21]用深度可分离卷积对Yolov3网络进行压缩,有效减少参数量。针对计算资源紧张的移动设备,YOLO系列算法网络时空复杂度较高,参数量较大,因此严重依赖高性能计算机等问题;Yolo-tiny系列通过减少网络结构中的网络层数来减小模型大小,同时降低了参数量,这种轻量化的方式虽然提高了检测速度,但其检测精度也大幅下降。红外图像与传统可见光图像的差别主要在于缺乏色彩信息,并且空间和纹理细节信息不够丰富,因此提高红外图像行人的检测精度面临更大的挑战。

为了保证红外行人目标算法在检测精度和实时性上具有较好的平衡性,本文基于Yolov4-tiny算法进行改进并提出轻量化红外行人实时检测(Tiny Infrared Pedestrian Real-time Detection, TIPRD)模型,降低模型大小的同时又保证良好的检测精度。针对检测精度问题,通过循环残差结构(Cross Stage Partial Network, CSP)来强化网络红外行人特征提取能力,同时增加精细检测尺度层并通过不同倍数采样进行多尺度特征融合;其次,针对红外行人目标宽高相对固定的特点,通过K-means++聚类算法重新设计了先验框;最后,在轻量化方面,将提高精度后的模型进行批量归一化(Batch Normalization, BN)层剪枝处理,再通过知识蒸馏取代剪枝后精调处理,在保证精度的同时,大幅减少了模型参数,同时在Jetson Nano(2GB)移动端开发平台上验证了TIPRD模型的可行性。

1 TIPRD网络模型

1.1 精细尺度嵌入多检测层结构

红外图像相较于传统可见光图像,缺乏色彩信息,并且空间和纹理细节信息不够丰富,因此,如何丰富网络提取的特征空间及有效利用上下文相关语义信息显得尤为重要。针对上述问题,对不同网络层提取的特征图进行如图1(其中(a)为原图,(b)、(c)、(d)依次为网络由浅至深的特征图)所示的可视化,网络浅层输出特征主要以目标的边缘、形状等位置特征为主。随着网络层加深,浅层网络提取的特征累加,深层特征呈现为高尺度、抽象的语义信息。当仅采用深层网络提取的特征信息作为预测层输入特征,会缺失中距离目标丰富的位置信息,这会造成中距离目标行人检测率较低的现象,因此不利于汽车辅助驾驶系统的响应。

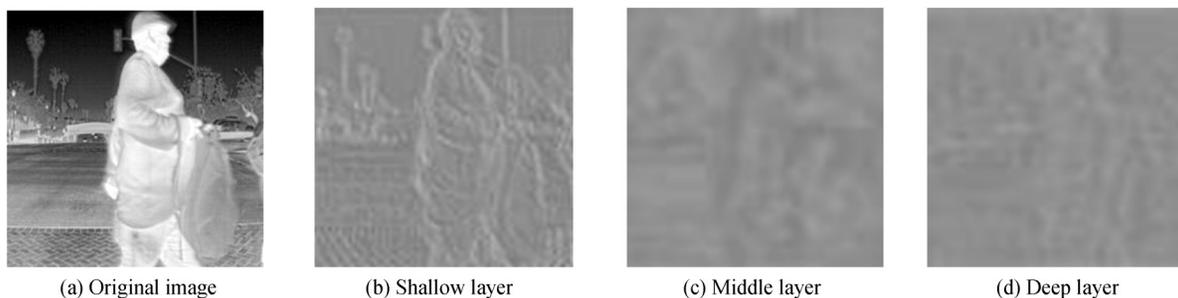


图1 特征图可视化

Fig.1 Visualization of the feature map

为强化网络特征提取能力,首先将CSP残差结构由三层增加至四层,一定程度上弥补红外图像颜色信息缺失的问题;其次,原始Yolov4-tiny网络只包含32倍下采样和16倍下采样所对应的 16×16 和 32×32 尺度的预测层。为了保证不同深度网络获得的位置信息和语义信息的充分结合,在该网络8倍降采样处添加由CBL(由Conv卷积层+Bn层+Leaky_relu激活函数三者组成)模块、 1×1 卷积层和Yolo检测头组成新的目标检测层。并且将8倍降采样特征图与经过上采样的16倍降采样特征图相融合,最终得到 16×16 、 32×32 和 64×64 三个预测尺度的检测层,极大丰富了特征图的信息,其结构如图2。

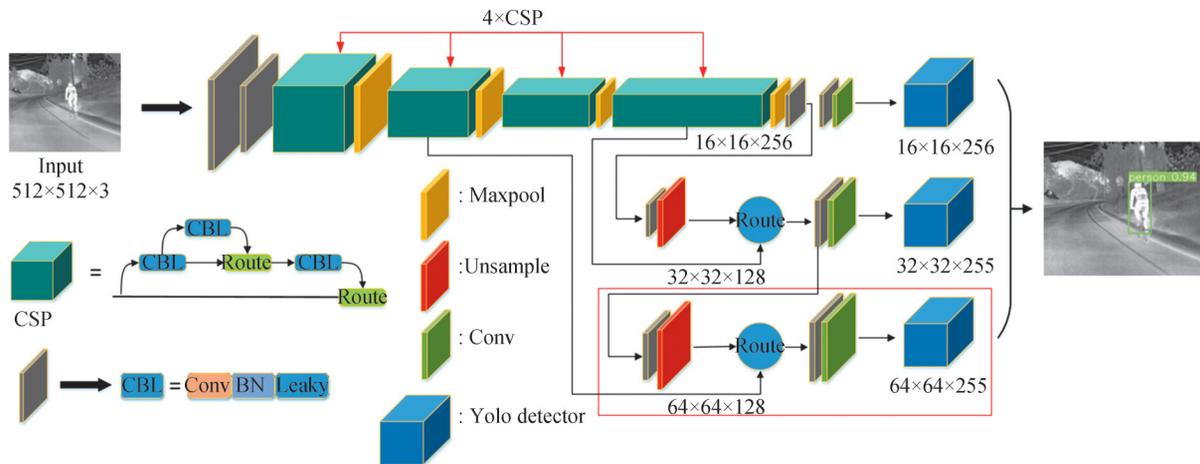


图2 TIPRD网络结构

Fig. 2 Structure of TIPRD network

1.2 行人先验框的聚类分析

Yolov4-tiny是Anchor-based思想的目标检测算法,先验框的形状与大小会对训练模型的效果产生较大的影响,先验框大小和检测对象尺寸不匹配,极可能会导致漏检或误检,从而降低定位精度,因此选择更为精准的先验框能够更好地反映目标特征,以提高模型的效果。

原Yolov4-tiny算法是在总计80个类别的COCO数据集上进行训练和测试,其数据集类别目标姿态、尺度差异较大,而本文使用的数据集中仅有红外行人一类,图像中红外行人目标的宽高比相对固定。

原先验框宽高维度参数设置分别为(10, 13)、(16, 30)、(33, 23)、(30, 61)、(62, 45)、(59, 119)、(116, 90)、(156, 198)、(373, 326),不适用于红外行人数据集。因此针对红外行人数据集,需要重新设计先验框大小。

本文利用K-means++聚类算法得到更适用于红外行人的先验框尺寸。普通K-means聚类算法初始化聚类中心的位置是完全随机的,会导致算法收敛很慢,而本文采用的K-means++聚类算法在聚类中心的初始化过程中保证每个聚类中心之间的相互距离尽可能远,避免了算法收敛慢的问题,其聚类结果和分配策略如表1所示。

表1 先验框的分配策略

Table 1 The allocation strategy of anchor

Detector size	Target size	Anchor size
16×16	Large target	(38, 132)
		(62, 205)
		(116, 321)
32×32	Medium target	(13, 49)
		(19, 60)
		(25, 89)
64×64	Small target	(5, 17)
		(8, 26)
		(9, 39)

如表1所示,本文采用 16×16 、 32×32 和 64×64 三种预测尺度的检测层,且每个预测尺度对应三个先验框,其中每个先验框所对应数值为先验框的宽高大小,显然聚类之后的先验框宽高比相对固定,与红外行人目标尺寸更匹配。

1.3 模型压缩

车辆辅助系统通常部署在移动开发平台,在此类平台部署精度与速度兼备的网络模型对它的计算能力和内存资源是一个巨大的挑战,因此需要将1.2节得到的网络模型进一步轻量化,实现模型在嵌入式平台部署中的性能优化。LIU Zhuang等^[22]提出一种方案实现通道稀疏化训练,其实质是对BN层的缩放因子增加L1正则化约束,降低模型复杂度。稀疏化训练时,尽量使缩放因子 γ 趋近于0,即可得到稀疏网络模型。稀疏模型缩放因子 γ 在L1正则化的约束下趋近于0,说明此通道在模型中的贡献程度比较低。本文将1.2节得到的模型进行了稀疏化训练,使用Tensorboard记录缩放因子 γ 权重值变化趋势,如图3所示。图3(a)为稀疏化训练前各BN层缩放因子 γ 的权重值分布情况, γ 值总体上分布在1附近,类似正态分布。图3(b)为稀疏化训练后各BN层缩放因子 γ 的权重值分布情况,稀疏化训练后 γ 的权重值在L1正则化的约束下趋近于0,满足稀疏化训练的目的。

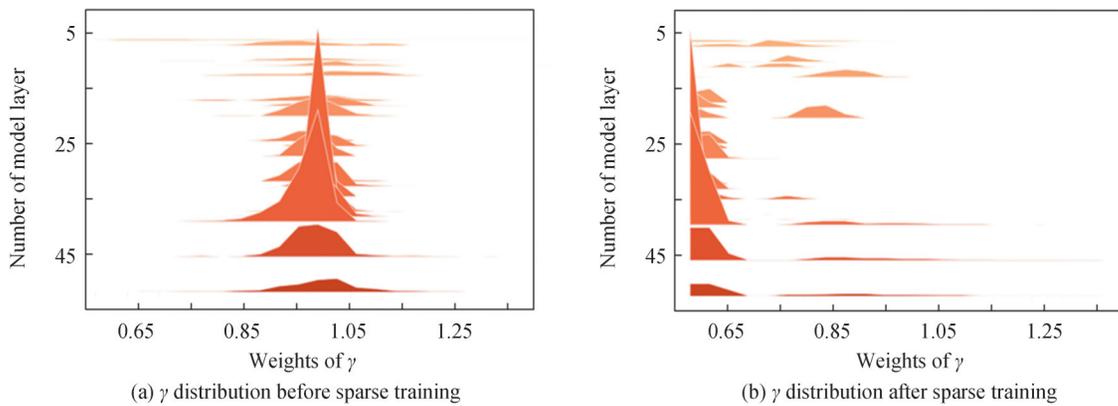


图3 稀疏化训练前后 γ 的分布情况

Fig.3 The distribution of γ before and after sparse training

由图3可知,模型稀疏化训练后大量缩放因子 γ 趋近于0,稀疏化效果比较明显,所以大部分通道贡献度较低,对于特征图的预测能力较弱,因此本文采用BN层通道剪枝策略来实现模型压缩。其实质是通过设定合适的剪枝率,将贡献程度低的通道进行删除,从而实现减少模型冗余参数,加快响应速度。通过实验,本文设置的最优剪枝率为0.8,模型剪枝前后的通道数变化情况如图4。

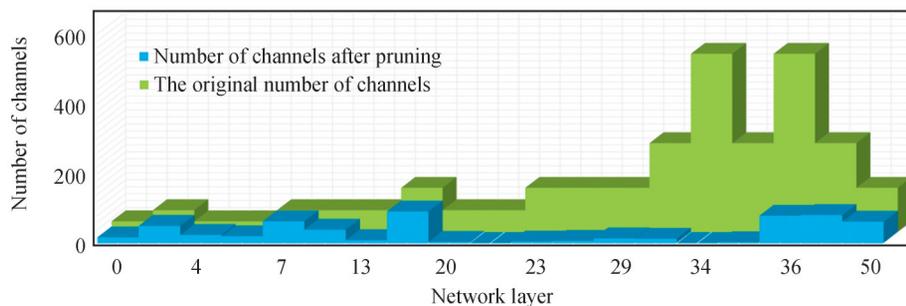


图4 剪枝前后通道数对比

Fig.4 Comparison of channel numbers before and after pruning

BN层通道剪枝后会导致模型检测精度存在不同程度的降低,其降低幅度取决于剪枝的力度,同时Yolov4-tiny相较于Yolov4等大型网络的参数量少,BN层通道剪枝后精度损失严重。

微调是传统剪枝后模型精度恢复的方法,其实质是对模型进行再训练;而知识蒸馏是利用参数量更大、

推理能力更好的 Teacher 模型辅助原模型再训练。在传统的训练过程 hard target(输入原始数据所对应的标签)中,所有负标签都被统一标记其概率为 0,但除了正标签之外,负标签也包含大量的信息,某些负标签对应的概率远远大于其他负标签,经过 Softmax 层输出的 Soft target(输入数据经过 Softmax 层输出的标签),可以提供更大的信息熵,所以知识蒸馏对于模型精度的微量调控更加有效。为了获得更多的信息熵,HINTON G 提出了 Softmax-T,可表示为^[23]

$$q_i = \frac{\exp(Z_i/T)}{\sum_j \exp(Z_j/T)} \quad (1)$$

式中, q_i 表示第 i 类的输出概率, Z_i, Z_j 表示 Softmax 层的输入, T 为温度系数,用来控制输出概率的 Soft 程度。如果 T 越大,则输出结果的分布越平缓,能保留更多的相似信息。

本文应用知识蒸馏^[23]方法进行模型精度的微量调控。在 0.8 剪枝率的 BN 层通道剪枝策略下,红外行人检测模型得到了压缩,但其检测精度跌至 0,因此本文将 BN 层通道剪枝前的红外行人检测模型权重文件 weights 和配置文件作为知识蒸馏的 Teacher 模型输入,经过 Softmax-T 得到 Soft target,从而辅助剪枝后的红外行人检测 Student 模型训练,完成其微量调控。模型压缩结构如图 5,其中 i 为卷积层序列号, C_i 表示第 i 卷积层的第 n 通道, j 则为 i 的下一卷积层, γ 为缩放因子, T 为温度系数, T -prediction 为 Teacher 模型的预测输出,prediction 为 Student 模型的预测输出。

本文模型压缩的实现流程如下:

- 1) 用红外行人数据集训练 1.2 节所得配置文件,得到基础模型;
- 2) 对基础模型进行稀疏化训练,使得对红外行人预测贡献程度较低通道的缩放因子 γ 趋于 0;
- 3) 确定剪枝比例,将低于剪枝比例设定阈值的缩放因子对应通道删除;
- 4) 将剪枝前的模型作为 Teacher 模型,剪枝后的模型作为 Student 模型,应用知识蒸馏实现模型的微量调控;
- 5) 生成新的红外行人检测模型文件。

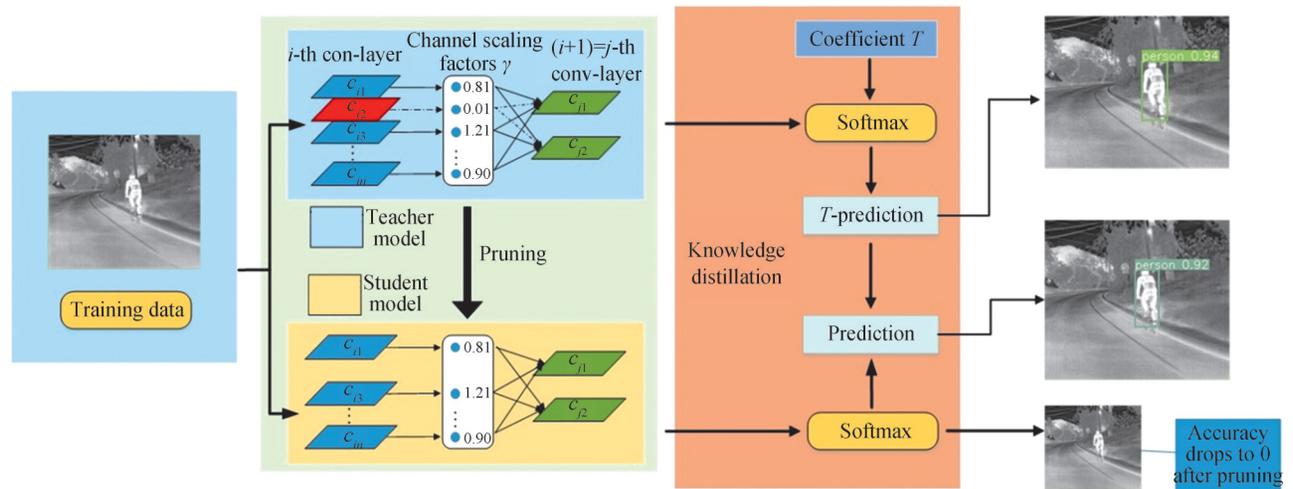


图 5 模型压缩框架

Fig.5 Model compression framework

2 实验及结果分析

2.1 数据集建立和实验环境

本文数据集选自 FLIR Thermal Starter 辅助驾驶红外数据集。该数据集中图像取帧间隔较短,相邻帧图像的相似度较大,针对原数据集进行稀疏化筛选,剔除相似度大的图像,只保留夜间条件下拍摄的红外行人图像作为数据集的图像,以降低过拟合造成检测精度低的问题。对数据集进行分析后发现,夜间红外图像行人姿态特征各异,不同行人姿态数量间存在不均衡的现象增大了算法识别难度,使得算法对模型的平

均检测精度与泛化能力均产生了不利影响。因此利用姿态扩充的方法,对不同数量的行人姿态特征进行均衡,进而为后续的神经网络提供更均衡的训练集,以获得更好的训练效果。

经过数据筛选与扩充后,得到数据集最终包含 1 815 幅图像。其中,1 364 幅图像作为训练集,451 幅图像作为测试集。数据集标注时只将行人、骑自行车与骑摩托车的人作为唯一的目标类,使用 YOLO-mark 样本标注软件对 1 815 张夜间红外行人图像、9 175 个行人目标真实值进行标注,数据集示例如图 6。



图 6 数据集示例
Fig.6 Example of data

实验环境是在 Ubuntu 系统下的 Pytorch 深度学习框架中进行训练,硬件配置和参数配置如表 2 和表 3 所示。

表 2 硬件配置

Table 2 Hardware configuration

Equipment	Model	Quantity
CPU	Intel(R) Core(TM) i5-10400F	1
GPU	NVIDIA GeForce RTX 2080Ti	1
RAM	8G DDR4 2666	2
Hard disk	512G SSD	1

表 3 参数配置

Table 3 Parameter configuration

Parameter	Value
Learn rate	0.001
Epoch	600
Momentum	0.9
Hue	0.1
Policy	Steps
Steps	3 200,3 600

2.2 评价指标

为了定量测试以上几种改进方法对红外行人检测模型性能所造成的影响程度,选取平均精度(Mean Average Precision, mAP)、每秒检测红外图像的帧数(Frame Per Second, FPS)、参数量以及模型大小作为模型性能的评价指标。

在辅助驾驶领域中的行人检测实际应用中,网络模型要部署在计算资源紧张的移动开发平台中,所以检测速度和模型大小也是至关重要的。一方面网络模型的检测速度可以反映出模型的响应快慢,车辆本身因具有较快的移动速度所以需要较快的检测速度;另一方面模型越大网络参数也就越多,其参数会影响模型的反应速度,从而不利于网络模型的移动端部署。因此评价指标 FPS 和模型大小也显得尤为重要。

2.3 实验结果分析

为验证针对精度提升所做工作的有效性,将 Yolov4-tiny 网络模型上分别添加多检测层、CSP 和先验框聚类的不同模型进行实验对比,实验结果如表 4。

表4 精度提升策略效果对比
Table 4 Comparison of the effect of precision improvement strategies

Model	Fine-scale detection layer	4×CSP	K-means++	Size/ MB	Speed/ (frames ⁻¹)	mAP/ %
Yolov4-tiny				23.5	90.7	80.6
Our	✓			24.9	84.0	86.5
Our	✓	✓		24.9	79.9	87.5
Yolo-pedestrian	✓	✓	✓	24.9	79.3	89.2

由表4可知,采用三种改进策略后的红外行人检测模型检测精度上升了8.6%,检测精度明显上升,但由于网络模型参数量的增加,其模型大小也增加了1.4 MB,FPS降低了11.4,因此需要对Yolo-pedestrian进行压缩。

2.3.1 模型压缩实验分析

为了使行人检测模型在精度提升的同时,又能保证在计算资源紧张的移动端车辆辅助系统中发挥性能,将存在较多的冗余参数和通道的Yolo-pedestrian进行BN层通道剪枝,实现模型的压缩和参数量的减少。通过精调和知识蒸馏两种策略实现剪枝后模型精度恢复可以弥补BN层通道剪枝后的模型精度损失。精调的本质就是进行再次训练,通过训练恢复损失的精度。本文采用的精调操作是将剪枝后的配置文件对剪枝得到的best.Weight权重进行再训练,训练迭代次数为500个epoch;知识蒸馏操作则是将剪枝前的模型作为Teacher模型,剪枝后的模型作为student模型。设置式(1)中 T 为3时,训练迭代次数为500个epoch。

本组实验的目的是验证不同剪枝率下以及剪枝后精调和知识蒸馏策略分别对Yolo-pedestrian红外行人检测模型精度恢复的影响。其结果如表5。

表5 不同微调策略结果对比
Table 5 Comparison of results of different fine-tuning strategies

Pruning rate	Fine tuning	Knowledge distillation	Number of channels	Size/MB	mAP/%
0.00			2 912	24.9	89.2
0.75	✓		728	4.6	89.1
0.75		✓	728	4.6	89.5
0.80	✓		583	4.0	88.6
0.80		✓	583	4.0	89.2
0.85	✓		374	3.5	87.7
0.85		✓	374	3.5	88.0
0.90	✓		184	3.0	83.9
0.90		✓	184	3.0	84.1

由表5可知,当网络模型进行BN层通道剪枝处理后,知识蒸馏比精调策略具有更好的精度恢复能力。当剪枝率为0.9时,剪枝后模型的精度通过精调和知识蒸馏两种策略都难以恢复,当剪枝率为0.8时,模型检测精度与Yolo-pedestrian保持一致,同时其模型大小压缩了20.9 MB。综上,Yolo-pedestrian模型在剪枝率为0.8时能保持一个均衡的检测性能,因此本文将此剪枝率下的红外行人检测模型命名为TIPRD。

2.3.2 对比实验

本组实验是为了对比TIPRD模型与Yolov3、Yolov3-tiny、Yolov4、Yolov4-tiny的检测性能,在本数据集上进行对比实验,实验结果如表6。

由表6可知,与经典的Yolov3、Yolov3-tiny、Yolov4和Yolov4-tiny模型相比,TIPRD模型在近似于Yolov4网络模型精度的前提下模型大小仅为Yolov4模型的1.5%,远远小于其他检测模型,比剪枝前Yolo-pedestrian模型FPS提高9.4帧/s。同时TIPRD模型还拥有着88.7帧/s的极快检测速度,满足实时性的要求,更加适用计算资源紧张的移动端夜间驾驶辅助系统的红外行人检测。模型最终检测结果如图7。

表6 不同算法对比
Table 6 Comparison of different algorithms

Model	Size/MB	Speed/(frame·s ⁻¹)	mAP/%
Yolov3	246.3	49.4	91.2
Yolov3-tiny	34.7	94.2	79.7
Yolov4	256	43.3	91.6
Yolov4-tiny	23.5	90.7	80.6
Yolo-pedestrian	24.9	79.3	89.2
TIPRD	4.0	88.7	89.2

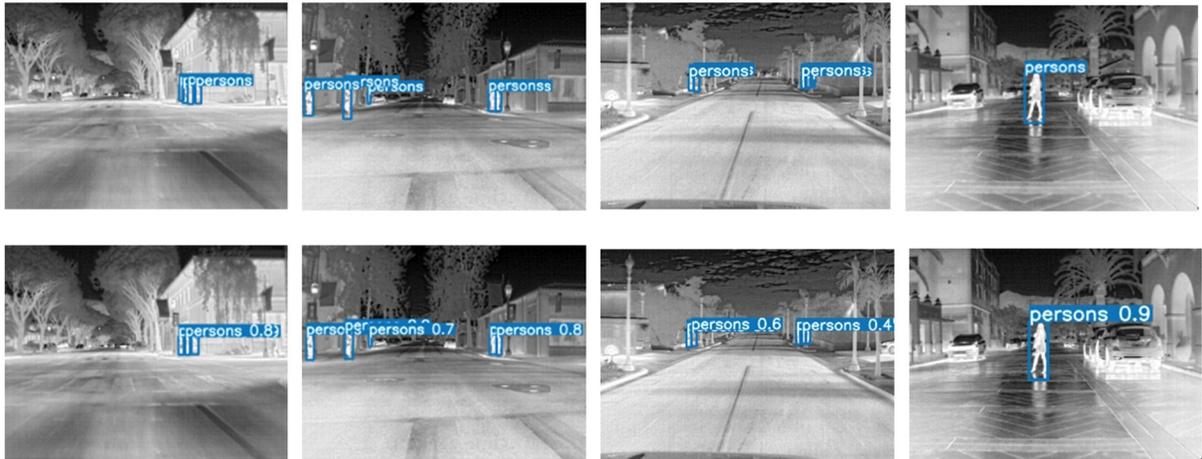


图7 检测效果
Fig.7 Test results

2.4 TIPRD 模型的移动端部署

为进一步验证 TIPRD 模型在移动端有效性及工程应用价值,采用 Jetson Nano 开发平台对模型进行部署,其中 TIPRD、Yolov3、Yolov3-tiny、Yolov4 以及 Yolov4-tiny 在 Jetson Nano 移动端开发平台的模型检测速度如表7。

表7 不同模型检测速度对比
Table 7 Comparison of detection speed of different models

Model	Speed/(frame·s ⁻¹)
Yolov4	/
Yolov3	0.95
Yolov4-tiny	5.2
Yolov3-tiny	2.6
TIPRD	6.9

由表7可知,TIPRD 夜间红外行人检测模型在计算能力较弱的 2GB 版 Jetson Nano 移动开发平台上测试的检测速度为 6.9 帧/s,相较于 Yolov4-tiny,TIPRD 夜间红外行人检测模型在计算能力有限的开发平台上 FPS 提升 1.7。本文提出的 TIPRD 模型不仅拥有轻量的模型,并且相对于其它模型检测速度有着明显的优势,更适合在移动端开发平台上部署。

3 结论

本文针对计算资源有限的辅助驾驶系统,提出了一种有较高精度的轻量级模型 TIPRD。首先基于 Yolov4-tiny 网络进行结构改进,在原网络结构上循环 CSP 残差结构来强化网络特征提取能力,并且增添 64×64 尺度的检测层,在新检测层与主干网络间添加一条特征融合线路,以融合红外行人的位置特征,丰富

特征图的语义信息。针对行人目标长宽相对固定的特点,使用K-means++聚类算法分析出适用于红外行人检测的先验框预置模型参数,提高先验框与行人目标的匹配度;模型精度提升8.6%,验证了针对Yolov4-tiny算法的改进有效性。其次在改进后行人检测模型基础上,对模型使用BN层通道剪枝策略实现压缩,同时利用知识蒸馏算法完成模型的微量调控,在保持精度前提下,实现了模型的深度压缩,使模型大小压缩至4 MB,同时测试速度到达了88.7 帧/s,比剪枝前提高了8.4 帧/s,满足实时性的需求。最后实现了TIPRD夜间红外行人检测模型在Jetson Nano(2GB)移动端开发平台上的部署,FPS提升了1.7 帧/s,进一步验证了本模型在移动端运行的有效性,展示了较好的工程应用价值。

本文所研究网络模型属于基于预设先验框的思想,目前已存在许多Anchor-free的目标检测算法,这类算法由于除去非极大值抑制这一步骤,所以能加快推理时间。本文网络结构可借鉴Anchor-free思想进一步优化。

参考文献

- [1] LI Jiankan, LI Zewei, HUA Wenwen, et al. Statistical analysis of urban road traffic accidents[J]. Science and Technology Innovation and Application, 2021, 11(21): 74-76.
黎健侃,李泽炜,华汶雯,等.城市道路交通事故统计分析[J].科技创新与应用,2021,11(21):74-76.
- [2] SOEHODHO S. Public transportation development and traffic accident prevention in Indonesia[J]. IATSS research, 2017, 40(2): 76-80.
- [3] JIANG Yongping. Research on the active steering headlight system of automobile[D]. Guangzhou: South China University of Technology, 2010.
蒋勇平.汽车主动转向灯系统的研究[D].广州:华南理工大学,2010.
- [4] XU Xiangyang, HU Wenhao, DONG Honglei, et al. Summary of key technologies for the construction of autonomous vehicle test scenarios[J]. Automotive Engineering, 2021, 43(4): 610-619.
徐向阳,胡文浩,董红磊,等.自动驾驶汽车测试场景构建关键技术综述[J].汽车工程,2021,43(4):610-619.
- [5] ZIEBINSKI A, CUPEK R, ERDOGAN H, et al. A survey of ADAS technologies for the future perspective of sensor fusion [C]. International Conference on Computational Collective Intelligence, 2016.
- [6] JI Biao. Principles of infrared temperature measurement and its influencing factors [C]. Proceedings of the 14th National Symposium on Infrared Heating and Infrared Medicine Development, 2013.
姬彪.红外测温原理及其影响因素[C].全国第十四届红外加热暨红外医学发展研讨会论文集,2013.
- [7] FAN Longfu. Research on classification technology of intelligent traffic pedestrian detection system [D]. Hefei: University of Science and Technology of China, 2010.
樊龙夫.智能交通行人检测系统的分类技术研究[D].合肥:中国科学技术大学,2010.
- [8] ZHAO Bin, WANG Chunping, FU Qiang, et al. Multi-scale infrared pedestrian detection based on deep attention mechanism[J]. Acta Optica Sinica. 2020, 40(5): 0504001.
赵斌,王春平,付强,等.基于深度注意力机制的多尺度红外行人检测[J].光学学报,2020,40(5):0504001.
- [9] CAI Wei, XU Peiwei, YANG Zhiyong, et al. Detection of small and dim targets in infrared images under complex background[J]. Applied Optics, 2021, 42(4): 643-650.
蔡伟,徐佩伟,杨志勇,等.复杂背景下红外图像弱小目标检测[J].应用光学,2021,42(4):643-650.
- [10] GIRSHICK R, DONAHUE J, DARRELL D, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[J]. 2014 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2014: 580-587.
- [11] GIRSHICK R. Fast R-CNN[C]. 2015 IEEE International Conference on Computer Vision (ICCV), IEEE, 2015.
- [12] REN S, HE K, 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.
- [13] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multi box detector[C]. Computer Vision - ECCV 2016, 2016.
- [14] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]. Computer Vision & Pattern Recognition, IEEE, 2016.
- [15] REDMON J, FARHAD A. YOLOv3: An incremental improvement[J]. Computer Vision and Pattern Recognition, 2018, 5(3): 12-22.
- [16] REDMON J, FARHAD A. YOLO9000: better, faster, stronger[C]. IEEE Conference on Computer Vision & Pattern Recognition, IEEE, 2017.
- [17] BOCHKOVSKIY A, WANG C Y, LIAO H. YOLOv4: Optimal speed and accuracy of object detection [J]. arXiv Preprint arXiv, 2020: 2004.10934.
- [18] ZHOU Z Y, GAI S Y, DA F P, et al. Occluded pedestrian detection algorithm based on attention mechanism [J]. Acta

- Optica Sinica, 2021, 41(15): 1515001.
 邹梓吟, 盖绍彦, 达飞鹏, 等. 基于注意力机制的遮挡行人检测算法[J]. 光学学报, 2021, 41(15): 1515001.
- [19] BAI Zhonghao, LI Zhiqiang, JIANG Binhui, et al. Real-time pedestrian detection in driver assistance system based on improved YOLOv2 model[J]. Automotive Engineering, 2019, 41(12): 1416-1423.
 白中浩, 李智强, 蒋彬辉, 等. 基于改进 YOLOv2 模型的驾驶辅助系统实时行人检测[J]. 汽车工程, 2019, 41(12): 1416-1423.
- [20] DENG Jie, WAN Wanggen. Dense pedestrian detection based on improved YOLOv3 [J]. Electronic Measurement Technology, 2021, 44(11): 90-95.
 邓杰, 万旺根. 基于改进 YOLOv3 的密集行人检测[J]. 电子测量技术, 2021, 44(11): 90-95.
- [21] HE Dong, CHEN Jinling, WANG Xi. Infrared pedestrian target detection based on improved YOLOv3 [J]. Chinese Science and Technology Paper, 2021, 16(7): 762-769.
 何东, 陈金令, 王熙. 基于改进 YOLOv3 的红外行人目标检测[J]. 中国科技论文, 2021, 16(7): 762-769.
- [22] LIU Zhuang, LI Jianguo, SHEN Zhiqiang, et al. Learning efficient convolutional networks through network slimming [C]. 2017 IEEE International Conference on Computer Vision (ICCV), IEEE, 2017.
- [23] HINTON G, VINYALS O, DEAN J. Distilling the knowledge in a neural network [J]. Computer Science, 2015, 14 (7): 38-39.

Lightweight Real-time Detection Model of Infrared Pedestrian Embedded in Fine-scale

ZHANG Yin-hui, ZHANG Peng-cheng, HE Zifen, WANG Sen

(Mechanical and Electrical Engineering, Kunming University of Science and Technology, Kunming 650500, China)

Abstract: The growing number of cars makes accidents more frequent. Due to the poor visibility conditions of drivers at night, the accident rate is higher than during the day. Therefore, various assisted driving technologies to enhance driving safety are widely used to reduce traffic accidents in the nighttime environment, among which infrared cameras have unique advantages at night. On the one hand, the visible light imaging of general cameras is easily affected by other interference light sources, and the low-quality images obtained in the nighttime environment with insufficient light will make pedestrian detection extremely difficult. The infrared camera technology based on the object's thermal radiation and reflection imaging can achieve barrier-free night vision without being affected by the interference light sources. On the other hand, the decreasing cost of infrared imaging equipment makes its application scenarios more and more common. Aiming at the night driving environment with a high accident rate, a night infrared image pedestrian detection model is proposed, which can detect pedestrians on the road at night in real-time. This research has important value and broad market application prospects in vehicle assisted driving, providing higher security for vehicles and pedestrians. Aiming at the problems of insufficient information such as color and texture of infrared images, low detection accuracy compared with visible light images, a large number of detection model parameters, and dependence on high-performance GPU resources, resulting in slow detection speed and other problems, a multi-scale embedding method fused with fine-scale pedestrian objects was proposed. Detection layer, lightweight real-time detection TIPRD model. First, to obtain more accurate infrared pedestrian location features, a 64×64 fine-scale detection layer is embedded on the original Yolov4-tiny structure to form a multi-detection layer structure, and a CSP module is added to deepen the backbone network to fuse the location features of infrared pedestrians; Secondly, in view of the relatively fixed aspect ratio of the infrared pedestrian target, K-means++ clustering is used to analyze the preset parameters of the a priori frame suitable for the infrared pedestrian target for improvement of the match between the a priori frame and the infrared pedestrian target. Finally, in order to reduce the model parameters, the model is processed through the BN layer channel pruning, and the model before pruning is used as the teacher model. At the same time, the model after pruning is used as the student model. The knowledge distillation algorithm is used instead of fine-tuning to complete the micro-control of TIPRD. While ensuring the detection accuracy, the model parameters are greatly reduced and the model is further lightened. Experiments based on the Yolov4-tiny network model show that using three strategies of fine-

scale multi-detection layer embedded in 64×64 , adding a CSP module and a priori box clustering can improve the detection accuracy of infrared pedestrian targets by 8.6%. But with the increase of model parameters, the model size increases by 1.4M, and the FPS decreases by 11.4 frames/s. Therefore, Yolo-pedestrian needs to be channel pruned to achieve model lightweight. After pruning the BN layer channel, the model detection accuracy will be reduced to varying degrees. Therefore, this paper uses knowledge distillation instead of fine-tuning to achieve the accuracy recovery of the model after pruning. When the pruning rate is 0.8, the model size is compressed by 20.9M, and the FPS is increased by 8.4 frames/s. The model can maintain the original accuracy after pruning through the knowledge distillation algorithm, achieving a lightweight model. Under the premise of approximating the accuracy of the Yolov4 network model, the size of the TIPRD model is only 1.5% of the Yolov4 model, which is far smaller than other detection models of the same type. Compared with the Yolo-pedestrian model before pruning, the FPS is improved by 9.4 frames/s. At the same time, the TIPRD model also has an extremely fast detection speed of 88.7 frames/s, which meets the requirements of real-time detection. For the assisted driving system with limited computing resources, a lightweight model TIPRD with high accuracy is proposed, which provides a certain reference value for the application of infrared pedestrian detection in the nighttime assisted driving system deployed on the mobile terminal. Firstly, the structure is improved based on the Yolov4-tiny network. The CSP structure is circulated on the original network structure to strengthen the network feature extraction ability, and a detection layer with a size of 64×64 is added. A feature fusion line is added between the new detection layer and the backbone network to fuse the location features of infrared pedestrians and enrich the semantic information of feature maps. And according to the relatively fixed length and width of pedestrian targets, the K-means++ clustering algorithm is used to analyze the preset model parameters of the apriori frame suitable for infrared pedestrian detection, which improves the match between the apriori frame and the pedestrian target; the model accuracy is improved by 8.6 percentage points, verifying the effectiveness of our improvements on the Yolov4-tiny algorithm. Secondly, based on the improved pedestrian detection model, the BN layer channel pruning strategy is used to achieve compression, and the knowledge distillation algorithm is applied to complete the micro-adjustment of the model. On the premise of maintaining accuracy, the deep compression of the model is realised, and the model's size is compressed. At the same time, the test speed reaches 88.7 frames/s, 8.4 frames/s higher than before pruning, which meets the requirements of real-time detections. Finally, the deployment of the TIPRD infrared pedestrian detection model at night on the Jetson Nano (2GB) mobile terminal development platform is realised, and the FPS is increased by 1.7 frames/s, by which the feasibility of running the model in the mobile terminal is further verified and good engineering application value is shown.

Key words: Infrared detection; Deep learning; Multiple detection layers; Model pruning; Knowledge distillation

OCIS Codes: 100.3008; 110.2970; 150.0155