

改进的 Tiny YOLOv4 算法及其在行人检测中的应用

宣勇¹, 韩超^{1*}, 沙文瀚²

¹安徽工程大学高端装备先进感知与智能控制教育部重点实验室, 安徽 芜湖 241000;

²奇瑞新能源汽车股份有限公司, 安徽 芜湖 241000

摘要 针对 Tiny YOLOv4 目标检测算法在行人检测中精确度低和召回率不高的问题, 对特征提取网络及预测网络进行改进。在特征提取网络部分采用深度可分离卷积网络取代传统卷积网络, 这减少了参数并降低了计算量; 将注意力机制模块加入特征提取网络中以增强检测目标的感兴趣区域, 提高检测精确度; 在预测网络部分增加一个预测尺度, 对增加的尺度进行特征增强处理, 以提升目标检测的召回率。实验结果表明, 与原算法相比, 改进后的 Tiny YOLOv4 算法的检测精确度提升了 7.1%, 召回率提升了 6.6%。

关键词 机器视觉; 目标检测; 深度可分离卷积; 注意力机制; 尺度增强预测

中图分类号 TP391.4

文献标志码 A

DOI: 10.3788/LOP202259.1215002

Improved Tiny YOLOv4 Algorithm and Its Application in Pedestrian Detection

Xuan Yong¹, Han Chao^{1*}, Sha Wenhan²

¹Key Laboratory of Advanced Perception and Intelligent Control of High-End Equipment, Ministry of Education, Anhui Polytechnic University, Wuhu 241000, Anhui, China;

²Chery New Energy Automobile Co., Ltd., Wuhu 241000, Anhui, China

Abstract To solve the problem that the Tiny YOLOv4 target detection algorithm has low accuracy and low recall rate in pedestrian detection, the feature extraction network and prediction network are improved. In the part of feature extraction network, the traditional convolution network is replaced by a depthwise separable convolution network to reduce parameters and computation. The attention mechanism module is added in the feature extraction network to enhance the area of interest of detecting object and improve the detection accuracy. A prediction scale is added in the prediction network, and the added scale is enhanced by features to improve the recall of detection of objects. The experimental results show that compared with the original algorithm, the improved Tiny YOLOv4 algorithm improves the accuracy by 7.1%, and the recall rate also increases by 6.6%.

Key words machine vision; target detection; depthwise separable convolution; attention mechanism; multi-scale enhancement prediction

1 引言

随着科技不断发展, 目标检测广泛应用于各种

领域, 如生物识别考勤、无人驾驶和视频监控等。

目标检测是对图像进行内容描述, 从图像中获取感兴趣目标, 从而得到物体的类别信息和位置信息^[1]。

收稿日期: 2021-04-14; 修回日期: 2021-05-21; 录用日期: 2021-06-02

基金项目: 安徽工程大学检测技术与节能装置安徽省重点实验室开放基金(DTESD2020A06)

通信作者: hanchaozh@126.com

行人检测^[2]是通过在图片和视频中检测出行人、得到行人具体位置信息^[3]的目标检测,与一般的目标检测相比,行人检测较为复杂。静态检测的目标物体形态固定^[4],而行人检测的目标形态随时可能会发生变化,这增加了检测的复杂性^[5]。

传统的目标检测主要通过手工提取特征,较为经典的特征提取算法有尺度不变特征变换(SIFT)^[6]、梯度方向直方图(HOG)^[7]和可变形部件模型(DPM)^[8]。

随着深度学习^[9]理论与算法的快速发展,基于深度学习的特征提取算法已成为目标检测的主流算法。基于深度学习的特征提取算法主要分为两类。一类为两步算法模型,即先对目标对象生成候选区域,再进行分类,典型的两步算法有 R-CNN^[10]、Fast R-CNN^[11]、Faster R-CNN^[12]和 Mask R-CNN^[13]等。此类算法精确度高、召回率高,但速度慢,不能实现实时检测。另一类为一步算法模型,此类方法不需要候选区域,直接对目标生成类别概率与位置坐标信息,此方法的检测速度快,精确度相对于两步算法较低,经典的一步算法有 YOLO(You Only Look Once)^[14]、SSD(Single Shot Multibox Detector)^[15]、YOLO9000^[16]、YOLOv3^[17]、YOLOv4^[18]等。

Tiny YOLOv4^[19]是 YOLOv4 的简化版,其卷积层大大减少,网络结构十分简单,内存占用较小,检

测速度快,检测精确度有所降低。本文通过改进 Tiny YOLOv4 的网络结构,得到更好的特征提取网络,提高了检测精确度。

2 Tiny YOLOv4 网络

YOLOv4 网络结构由特征提取网络(Backbone)、加强特征网络(Neck)和预测网络(Prediction)组成。特征提取网络采用 CSPDarknet53,共有 5 个残差块;加强特征网络由 SPP(Spatial Pyramid Pooling)和 PANet(Path Aggregation Network)组成;预测网络由 13×13、26×26 和 52×52 三个预测尺度组成。Tiny YOLOv4 对 YOLOv4 网络结构进行了改进,形成了轻量级的检测算法,易移植到嵌入式平台,网络结构如图 1 所示,其 Backbone 部分采用 CSPDarknet53-tiny 结构,该结构由卷积-批规范化-激活函数(CBL)、跨阶段局部网络(CSP)和池化层(Pooling)三部分组成,减少了残差块网络,削减了网络训练层数,提升了检测速度;Neck 部分去掉了 SPP,采用 FPN^[20]结构代替 PANet 结构进行特征融合;Prediction 部分减少了一个预测尺度,利用 13×13、26×26 两个预测尺度。Tiny YOLOv4 相对于 YOLOv4 的检测精确度较低,但检测速度大大提升。

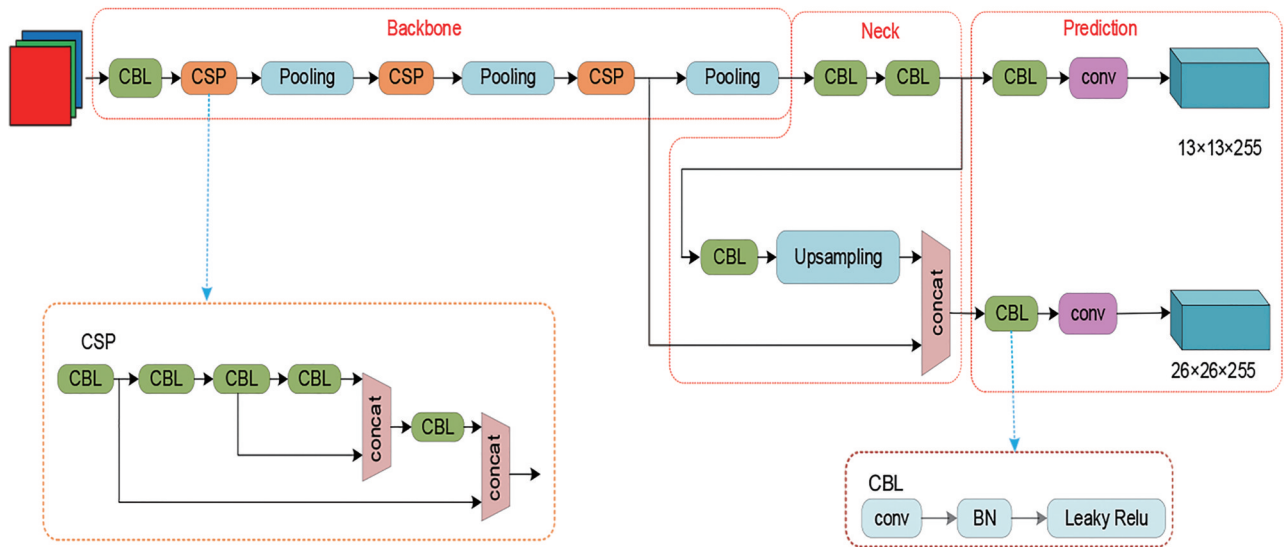


图 1 Tiny YOLOv4 网络结构
Fig. 1 Network structure of Tiny YOLOv4

就检测算法而言,检测精确度(Precision,可用 P 表示)和召回率(Recall,可用 R 表示)是衡量检测优劣的两个重要指标。检测精确度指正确检测出的目

标数(N_{TP})占已检测出目标的百分比,其计算式为

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}} \times 100\%, \quad (1)$$

式中: N_{FP} 为将错误目标当成正确目标检测出的数量。召回率指的是正确检测出的目标占检测总目标的百分比,其计算式为

$$R = \frac{N_{TP}}{N_{TP} + N_{FN}} \times 100\%, \quad (2)$$

式中: N_{FN} 表示目标存在但系统未将其检测出的目标数量。

3 Tiny YOLOv4 算法改进

在目标检测领域,神经网络算法已占主导地位,

尤其是在视频监控、无人驾驶等对检测速度和精确度要求较高的领域,Tiny YOLOv4 算法因其优异的检测速度,在行人等目标检测中被广泛应用,但其尚存在精确度低、召回率不高等问题。本文针对上述问题,对 Tiny YOLOv4 的网络结构进行改进,在特征提取网络部分采用深度可分离卷积(DSC)网络代替传统卷积网络,使网络结构更加简单,网络计算量更少,通过增加注意力机制模块增强行人的特征信息;通过增加一个预测尺度提高对小目标的检测精确度和召回率。改进后的网络结构如图 2 所示。

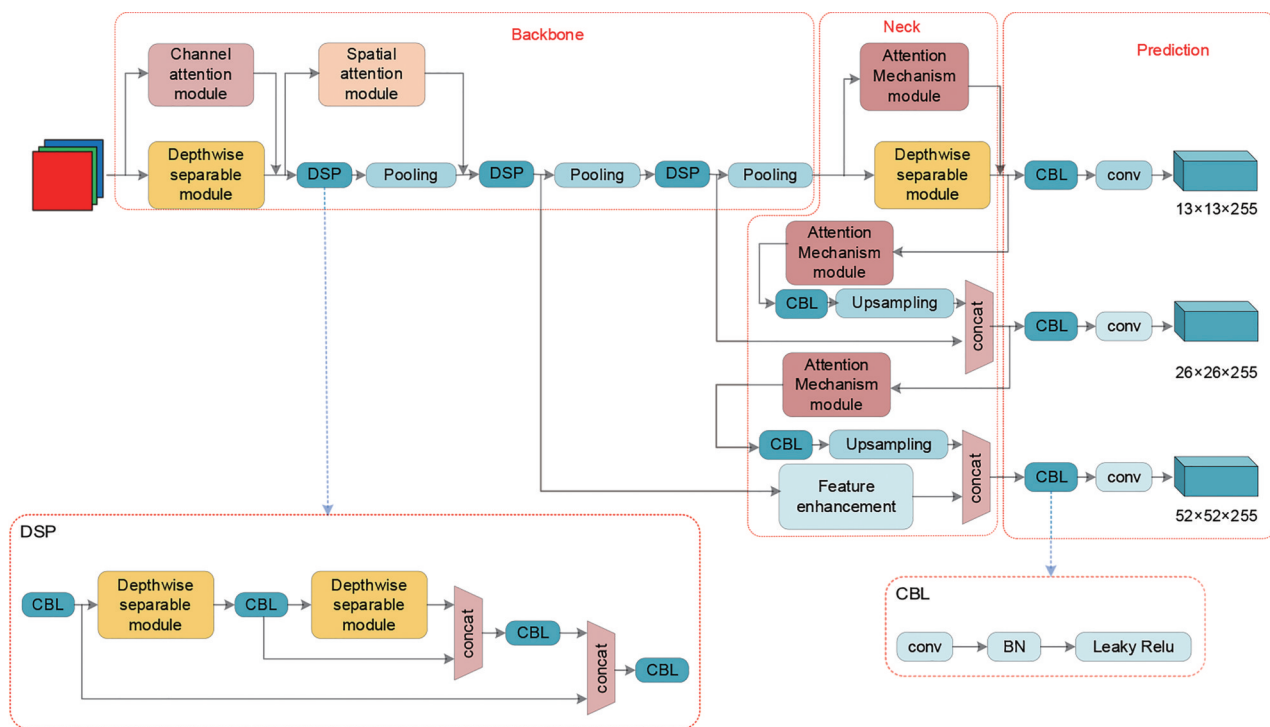


图 2 改进的网络结构

Fig. 2 Improved network structure

在改进的网络结构中,将特征提取网络(Backbone)中的卷积层改为深度可分离卷积层,把 CSP(Cross Stage Part)层改为带有深度可分离结构的 DSP(Depthwise Separable Part)层,利用深度可分离卷积来减少参数量。在输入图片与 DSP 层之间增加一个通道注意力机制层,在第一个 DSP 层和第二个 DSP 层之间增加一个空间注意力层。在加强特征网络增加注意力机制模块,利用注意力机制模块可增强目标物体的感兴趣区域,使感受野更关注目标特征。最后在预测网络增加一个预测尺度,形成多尺度预测,提高行人检测的精确度。

3.1 深度可分离卷积模块

目前神经网络结构广泛采用深度可分离卷

积^[21]作为特征提取的关键模块,使网络结构更加简单。深度可分离卷积的基本思想是用两个独立的层代替一个标准卷积层。第一层使用通道卷积(DC),它通过卷积核的计算获得每个通道中的空间信息,生成的特征通道与输入的特征通道完全相同。第二层采用 $1 \times 1 \times M$ 的点卷积(PC)得到跨通道信息, M 为前一层的通道数,此处的卷积运算将通道卷积的 feature map 加权组合成新的 feature map,最终得到的 feature map 值由卷积核的数量所得。传统卷积和深度可分离卷积的结构如图 3 所示。

假设从模型中输入的特征图大小为 $D_F \times D_F \times M$ (D_F 为宽/高, M 为输入通道数, F 为对输入特征图

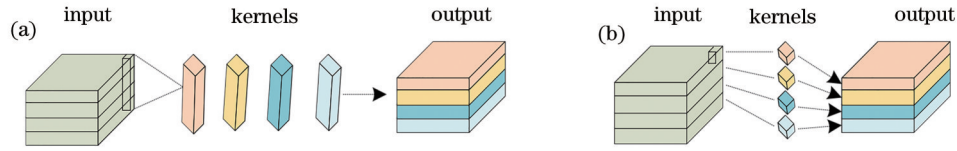


图 3 传统卷积和深度可分离卷积的网络结构。(a)传统卷积的网络结构;(b)深度可分离卷积的网络结构

Fig. 3 Network structures of traditional convolution and depthwise separable convolution. (a) Network structure of traditional convolution; (b) network structure of depthwise separable convolution

分的块数),输出的特征图大小为 $D_F \times D_F \times N$ (N 为输出通道数),卷积核大小为 $D_K \times D_K$ (K 为对卷积核分的块数),则传统卷积的计算量 (C_{on}) 为

$$C_{on} = D_K \times D_K \times M \times N \times D_F \times D_F \quad (3)$$

深度可分离卷积的计算量 (D_{sc}) 为

$$D_{sc} = D_K \times D_K \times M \times D_F \times D_F + M \times N \times D_F \times D_F \quad (4)$$

深度可分离卷积与传统卷积之间的计算量之比为

$$\frac{D_K D_K M D_F D_F + M N D_F D_F}{D_K D_K M N D_F D_F} = \frac{1}{N} + \frac{1}{D_K^2} \quad (5)$$

由式(5)可知,深度可分离卷积与传统卷积的参数计算量之比为 $\frac{1}{N} + \frac{1}{D_K^2}$,可得深度可分离卷积的计算量远小于传统卷积的计算量。为了减少计算量,在特征提取网络部分采用深度可分离层,从而减少卷积运算中的训练参数。表 1 为模型改进后的训练参数与 YOLOv4 和 Tiny YOLOv4 的训练参数对比。

表 1 不同网络模型的训练参数比较

Table 1 Comparison of training parameters of different network models

Detection algorithm	Number of training parameters
YOLOv4	64040001
Tiny YOLOv4	5939804
Proposed Tiny YOLOv4	4369606

3.2 注意力机制模块

注意力机制^[22]的本质是定位感兴趣区域的信息,抑制不需要的信息,将注意力集中在关注的特征上,对整段输入特征的某个部分更加关注,提取关键的特征,得到重要的信息。注意力机制由通道注意力和空间注意力构成。

通道注意力机制是对通道中感兴趣区域信息的增强,在特征提取网络中增加通道注意力机制可增强特征提取的通道信息,从而更好地提取行人特征信息。通道注意力结构如图 4 所示,图中 C 为通道数。

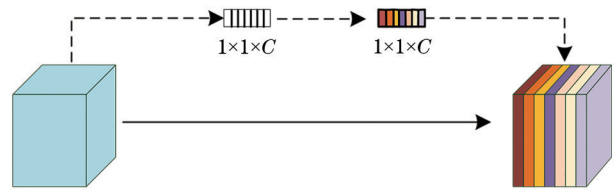


图 4 通道注意力结构

Fig. 4 Channel attention structure

空间注意力模块是在网络中找出需要提取的特征,然后对其进行处理。空间注意力模块的网络结构如图 5 所示。

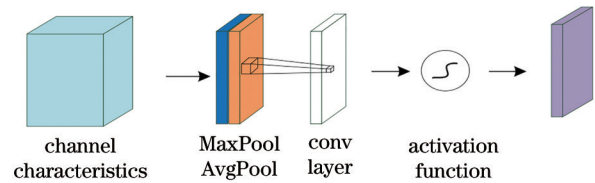


图 5 空间注意力模块的结构图

Fig. 5 Structure diagram of spatial attention module

空间注意力模块采用通道特征作为输入图,对通道特征分别进行最大池化和平均池化,将两次池化操作的结果进行一次卷积操作,将通道数改为 1 个通道,经过 sigmoid 激活函数生成空间注意力模块特征,将特征模块和该模块的输入特征相结合可使目标信息得到增强。因此,本文将通道注意力机制加在网络特征提取部分的 DSP 层,在第一个 DSP 层和第二个 DSP 层之间增加一个空间注意力机制,将通道注意力机制和空间注意力机制合成注意力机制模块并将其放在特征网络部分,最终使感兴趣目标区域的特征得到增强。

3.3 多尺度预测

Tiny YOLOv4 网络中的预测网络部分是预测目标在图片上的位置,由 13×13 和 26×26 两个预测尺度组成,因此,在预测网络部分增加一个 52×52 的预测尺度后,可得到更多的目标信息,大大提高检测目标在图片上的位置准确率。采用特征增强模块使行人的特征信息得到增强,检测精确度得

到进一步提升,特征增强结构如图 6 所示。

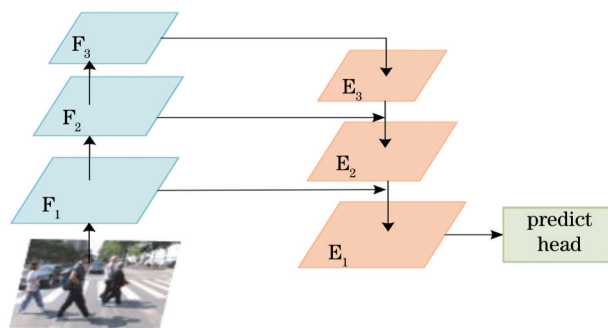


图 6 特征增强模块

Fig. 6 Feature enhancement module

特征增强模块采用特征金字塔^[23]结构,通过从下往上的路径和横向网络连接不断增强特征信息。对输入的图像特征进行三次上采样操作,对三次上采样得到的特征图像进行下采样并进行特征融合,行人目标信息得到增强,最终得到较好的检测效果。表 2 为改进的网络模型与 YOLOv4 和 Tiny YOLOv4 模型的对比结果。

表 2 不同网络模型的大小比较

Table 2 Size comparison of different network models

Detection algorithm	Model size /MB
YOLOv4	246
Tiny YOLOv4	22.7
Proposed Tiny YOLOv4	15.9

4 实 验

为了验证改进后 Tiny YOLOv4 算法的性能,将

表 3 不同算法对不同数据集的测试结果

Table 3 Test results of different algorithms on different datasets

Dataset	Algorithm	AP /%	FPS / (frame·s ⁻¹)
INRIA	YOLOv4	86.2	20.1
	Tiny YOLOv4	68.67	31.6
	Proposed Tiny YOLOv4	76.32	30.4
COCO	YOLOv4	90.76	20.8
	Tiny YOLOv4	69.76	32.4
	Proposed Tiny YOLOv4	78.2	30.2
VOC	YOLOv4	90.4	19.8
	Tiny YOLOv4	71.27	31.2
	Proposed Tiny YOLOv4	78.8	30.6
Mixed data	YOLOv4	92.5	19.5
	Tiny YOLOv4	73.42	32.2
	Proposed Tiny YOLOv4	80.52	31.4

其与经典目标检测算法在 INRIA、COCO、VOC 数据集和混合数据集上进行对比实验。用检测精确度、召回率和检测速率来评价检测算法的效果。

4.1 数据集制作

实验采用 64 位 windows 10 操作系统,实验程序语言是在 python 中的 Pytorch 框架实现的;利用 Cuda10.0 进行加速;硬件配置处理器:AMD Ryzen5 2600X Six-Core Processor;显卡:NVIDIA GeForce GTX 1070Ti;内存:8 GB。实验采用 Adam Optimizer 优化器,最初学习率值是 0.001,动量值为 0.9,权重衰减为 0.005,通过自适应调整学习速度,epoch 设为 100,先冻结前 50 个 epoch,当前 50 个 epoch 训练完,再解冻后 50 个 epoch,每经历 1 个 epoch 保存一次模型,最终选取训练中最低 loss 值的训练模型进行检测。

训练数据集选用 INRIA、COCO、VOC 数据集和它们的混合数据集进行行人检测实验,在 INRIA 中选择 902 张行人图片,在 COCO 中将行人选取出来,最终选择 5174 张图片,由于 VOC 数据集中含有 20 个类,通过数据筛选保留 person 类别的标签信息,并把 VOC 数据集中的行人这一类选取出来,最终得到混合的数据图片共 14642 张,其中 10% 的图片用于测试,20% 的图片用于验证,70% 的图片用于训练。

4.2 实验结果与分析

用 INRIA、COCO、VOC 和混合数据集分别对 YOLOv4 算法、Tiny YOLOv4 算法和本文改进算法进行训练与测试,计算检测精确度(AP)和检测速率(FPS)结果如表 3 所示。

由表 3 可知,用混合数据集训练后进行测试比用单个数据集训练后进行测试得到的检测精确度更高。因此,选用混合数据集在经典检测算法上进行对比实验,计算 AP、Recall、FPS,实验结果如表 4 所示。

表 4 检测算法结果对比

Table 4 Comparison of results of detection algorithm

Algorithm	AP / %	Recall / %	FPS / (frame·s ⁻¹)
Faster R-CNN	72.37	78.4	4.7
SSD	78.20	79.1	22.1
YOLOv3	85.34	77.6	17.3
Tiny YOLOv3	67.80	73.4	25.6
YOLOv4	92.50	81.5	19.5
Tiny YOLOv4	73.42	75.7	32.2
Proposed Tiny YOLOv4	80.52	82.3	31.4

由表 5 可知,在深度可分离的基础上增加注意力机制模块的精确度为 77.27%,与未修改的 Tiny YOLOv4 算法的 73.42% 精确度相比,提高了 3.85 个百分点;将多尺度预测与深度可分离和注意力机制模块相结合的精确度为 80.52%,与未修改的 Tiny YOLOv4 算法的 73.42% 精确度相比,提高了 7.1 个百分点。在检测速度方面,改进后的算法与 Faster R-CNN、SSD、YOLOv3、Tiny YOLOv3 和 YOLOv4 相比分别提高了 26.7、9.3、14.1、5.8、

11.9 frame·s⁻¹。改进后的算法不仅提高了检测精确度,也提高了召回率。

为了验证各个模块在改进算法上的检测效果,用混合数据集进行消融实验,实验结果如表 5 所示。

表 5 在混合数据集上的消融实验

Table 5 Ablation experiments on mixed datasets

Tiny YOLOv4 baseline	DSC	Attention mechanism	Scale enhancement	AP / %
✓				73.42
✓	✓			74.32
✓	✓	✓		77.27
✓	✓	✓	✓	80.52

图 7 是不同算法之间的损失值对比图。YOLOv4 的 Train_loss 值从最大值 13.4 左右开始回归,最终到 8.4 趋于稳定,Test_loss 从最大值 8.3 左右开始回归,最终到 5.6 趋于稳定;而 Tiny YOLOv4 的 Train_loss 值从 10.01 开始回归,最终在 7.4 达到稳定,Test_loss 从最大值 8.4 左右开始回归,最终到 6.6 趋于稳定;改进算法的 Train_loss 从最大值接近 10 开始回归,最终在 7.6 达到稳定,Test_loss 从最大值 8.1 左右开始回归,最终在 7.3 达到稳定。因此,YOLOv4、Tiny YOLOv4 及本文提出的改进算法的训练损失值和测试损失值最终均能趋于稳定,即达到收敛状态。训练损失值与测

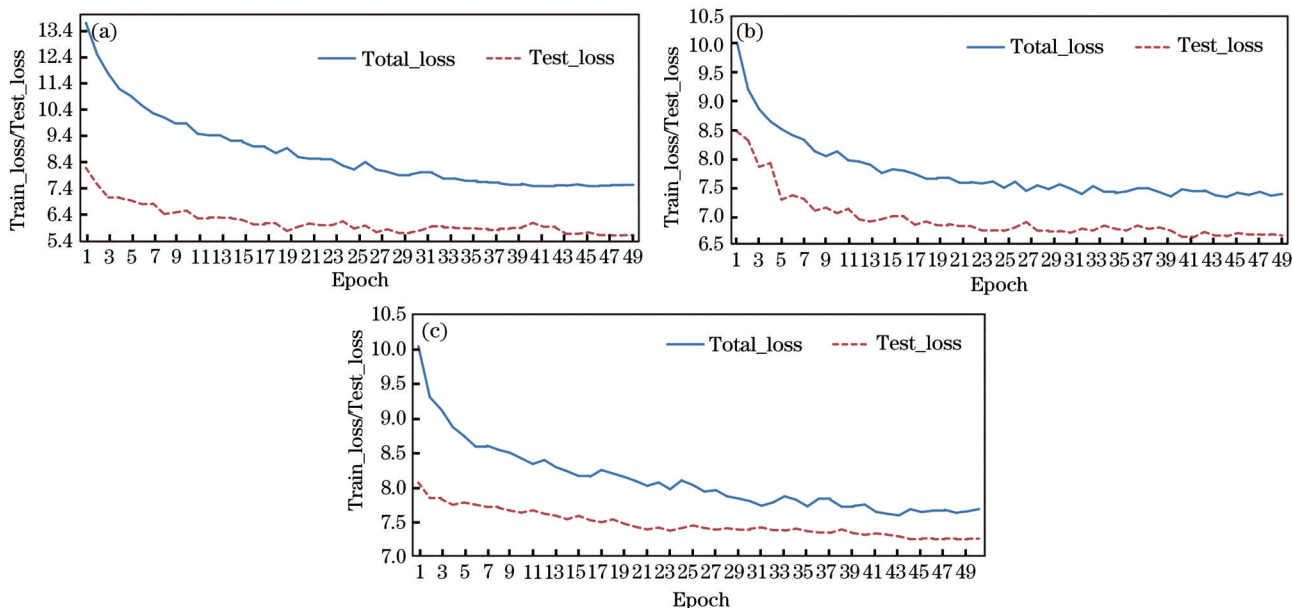


图 7 不同网络结构的损失值。(a) YOLOv4 的损失值;(b) Tiny YOLOv4 的损失值;(c)改进网络的损失值

Fig. 7 Loss values of different network structures. (a) Loss values of YOLOv4; (b) loss values of Tiny YOLOv4; (c) loss values of improved network

试损失值在达到稳定状态时,两者之间的差值分别为 1.88,0.73,0.43,改进算法的训练损失值和测试损失值的差值较小,测试结果表现较好。

图 8 是改进算法与 Tiny YOLOv4 算法的检测

结果的对比。其中图 8(a)~(d)为原图,图 8(e)~(h)是 Tiny YOLOv4 检测的结果,图 8(i)~(l)是改进算法的检测结果。由检测结果可见,改进后的算法使行人检测效果得到改善。



图 8 实验结果

Fig. 8 Experimental results

5 结 论

基于 Tiny YOLOv4 算法,将传统卷积网络改为深度可分离卷积网络,引入注意力机制模块,并在输出层增加一个尺度来改进 Tiny YOLOv4 算法,从而提高了检测算法的精确度。实验结果表明改进的 Tiny YOLOv4 算法的行人检测精确度提高了 7.1%,召回率提高了 6.6%,具有较好的目标检测效果。下一步将研究对于行人相互遮挡、物体对行人的遮挡和复杂背景下的行人检测问题,并在尽可能不增加算法计算量的情况下提升检测的精度和实时性。

参 考 文 献

[1] 段仲静,李少波,胡建军,等.深度学习目标检测方法及其主流框架综述[J].激光与光电子学进展,2020,57(12):120005.
Duan Z J, Li S B, Hu J J, et al. Review of deep

learning based object detection methods and their mainstream frameworks[J]. Laser & Optoelectronics Progress, 2020, 57(12): 120005.

[2] 曹帅,张晓伟,马健伟.基于跨尺度特征聚合网络的多尺度行人检测[J].北京航空航天大学学报,2020,46(9):1786-1796.

Cao S, Zhang X W, Ma J W. Trans-scale feature aggregation network for multiscale pedestrian detection [J]. Journal of Beijing University of Aeronautics and Astronautics, 2020, 46(9): 1786-1796.

[3] 鞠默然,罗海波,王仲博,等.改进的YOLO V3算法及其在小目标检测中的应用[J].光学学报,2019,39(7):0715004.

Ju M R, Luo H B, Wang Z B, et al. Improved YOLO V3 algorithm and its application in small target detection[J]. Acta Optica Sinica, 2019, 39(7): 0715004.

[4] 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.

Zhao B, Wang C P, Fu Q, et al. Multi-scale infrared

- pedestrian detection based on deep attention mechanism [J]. *Acta Optica Sinica*, 2020, 40(5): 0504001.
- [5] Lan W B, Dang J W, Wang Y P, et al. Pedestrian detection based on YOLO network model[C]//2018 IEEE International Conference on Mechatronics and Automation, August 5-8, 2018, Changchun, China. New York: IEEE Press, 2018: 1547-1551.
- [6] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. *International Journal of Computer Vision*, 2004, 60(2): 91-110.
- [7] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]//2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 20-25, 2005, San Diego, CA, USA. New York: IEEE Press, 2005: 886-893.
- [8] Felzenszwalb P, McAllester D, Ramanan D. A discriminatively trained, multiscale, deformable part model[C]//2008 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2008, Anchorage, AK, USA. New York: IEEE Press, 2008: 10139902.
- [9] 郑远攀, 李广阳, 李晔. 深度学习在图像识别中的应用研究综述[J]. *计算机工程与应用*, 2019, 55(12): 20-36.
Zheng Y P, Li G Y, Li Y. Survey of application of deep learning in image recognition[J]. *Computer Engineering and Applications*, 2019, 55(12): 20-36.
- [10] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, OH, USA. New York: IEEE Press, 2014: 580-587.
- [11] Girshick R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision, December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 1440-1448.
- [12] 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.
- [13] He K M, Gkioxari G, Dollár P, et al. Mask R-CNN [C]//2017 IEEE International Conference on Computer Vision, October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 2980-2988.
- [14] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 779-788.
- [15] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[M]//Leibe B, Matas J, Sebe N, et al. *Computer vision-ECCV 2016. Lecture notes in computer science*. Cham: Springer, 2016, 9905: 21-37.
- [16] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6517-6525.
- [17] Redmon J, Farhadi A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08) [2018-12-25]. <https://arxiv.org/abs/1804.02767>.
- [18] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. (2020-04-23)[2020-06-07]. <https://arxiv.org/abs/2004.10934>.
- [19] Jiang Z X, Zhao L Q, Li S Y, et al. Real-time object detection method based on improved YOLOv4-tiny [EB/OL]. (2020-11-09) [2020-12-02]. <https://arxiv.org/abs/2011.04244>.
- [20] Kirillov A, Girshick R, He K M, et al. Panoptic feature pyramid networks[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 6392-6401.
- [21] Chollet F. Xception: deep learning with depthwise separable convolutions[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 1800-1807.
- [22] Zhu X Z, Cheng D Z, Zhang Z, et al. An empirical study of spatial attention mechanisms in deep networks [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 6687-6696.
- [23] Adelson E H, Anderson C H, Bergen J R, et al. Pyramid methods in image processing[J]. *RCA Engineer*, 1984, 29(6): 33-41.