

基于几何透视图像预处理和 CNN 的全景图像交通标志识别算法

曹峻凡 张向利* 闫坤 张红梅

(桂林电子科技大学广西无线宽带通信与信号处理重点实验室 广西 桂林 541004)

摘要 为解决深度学习方法在高清全景图像中检测交通标志遇到图形处理器资源不足、小目标容易漏检、检测速度过慢等问题,采用小目标过采样训练数据生成方法、图像分块和几何透视检测预处理方法以及改进的轻量神经网络 Improved-Tiny-YOLOv3,提出了一种基于深度学习的轻量级全景图像中交通标志检测方法。并在 Tsinghua-Tencent 100K 数据集上进行了实验,mAP 值达到 92.7%,在 Nvidia 1080Ti 显卡上检测速度可达到 20 FPS,实验结果验证了所提方法的有效性。

关键词 交通标志检测识别 Improved-Tiny-YOLOv3 几何透视法 随机裁剪 Clou 全景图像

中图分类号 TP3 **文献标志码** A **DOI**:10.3969/j.issn.1000-386x.2024.07.026

A PANORAMIC IMAGE TRAFFIC SIGN RECOGNITION ALGORITHM BASED ON GEOMETRIC PERSPECTIVE IMAGE PREPROCESSING AND CNN

Cao Junfan Zhang Xiangli* Yan Kun Zhang Hongmei

(Guangxi Key Laboratory of Wireless Wideband Communication and Signal Processing, Guilin University of Electronic Technology, Guilin 541004, Guangxi, China)

Abstract In order to solve the problems of insufficient graphics processor resources, small targets being easily missed, and slow detection speed when the deep learning method is used to detect traffic signs in high-definition panoramic images, this paper proposes a lightweight panoramic image traffic sign detection method based on deep learning. It used a small target oversampling training data generation method, image segmentation and geometric perspective detection preprocessing method, and an improved lightweight neural network Improved-Tiny-YOLOv3. Experiments were performed on the Tsinghua-Tencent 100K data set, and the mAP value reached 92.7%, and the detection speed on the Nvidia 1080Ti graphics card reached 20FPS. The experimental results show the effectiveness of the proposed method.

Keywords Traffic sign detection and recognition Improved-Tiny-YOLOv3 Geometric perspective Random crop Clou Panoramic image

0 引言

交通标志是道路行驶中非常重要的一环,其引导和规范了驾驶员的驾驶行为,保障了车辆行驶过程中的安全。在无人驾驶技术中,如何快速准确地检测出图像中交通标志一直都是研究的热点和难点,具有很高的现实意义与应用价值。

当前,卷积神经网络被广泛地应用于目标检测中,

基于卷积神经网络的目标检测方法主要可分为两种,基于候选区域的 Faster R-CNN^[1]、Cascade R-CNN^[2]和基于回归的 YOLOv3^[3]、YOLOv4^[4]与 SSD^[5]。这些网络模型通过一定改进能很好地应用于一般的道路交通数据集中,如德国交通标志数据集 GTSD^[6]。但在智能交通系统中通常捕获的是全景图像,交通标志的尺寸在图像中占比小且种类繁多,这会给检测带来很大难度。此外全景图像分辨率往往非常大,如本文选用的全景交通标志数据集(Tsinghua-Tencent 100K, TT100K)^[7],

每幅图像的分辨率是 $2\ 048 \times 2\ 048$, 应用深度学习方法处理高分辨率图像时, 图形处理器 (Graphics Processing Unit, GPU) 会遇到资源受限问题。

为解决上述问题, 近年来, 全景图像交通标志检测的研究主要是基于图像预处理和网络模型改进两方面来开展的。文献[8]在 Faster R-CNN 基础上引入注意力模块, 引导区域建议的生成, 在 TT100K 数据集进行实验, 将原图下采样至网络输入尺寸进行检测, 获得了 80.31% 的 mAP。文献[9-11]提出了端到端的卷积神经网络模型, 在检测端, 将全景图像进行分块处理, 依次对图像块进行检测, 解决了 GPU 资源不足的问题, 同时保留了交通标志细节信息, 取得了不错的准确率。文献[12]改进了 YOLOv3 的特征融合网络, 使之能更好地检测小目标, 并且使用了更符合交通标志比例的锚点框, 改进后的网络在 TT100K 上获得 82.73% 的 mAP。当前研究已取得很好的进展, 但在图像预处理阶段, 直接对输入图像下采样, 显然会丢失许多图像的细节信息, 使小交通标志更难以被检测。而纯粹的图像分块处理, 会大大地增加检测的冗余。同时, 针对交通标志识别的网络模型大多难以运行在性能较差的边缘设备上, 亟待轻量级交通标志检测模型的提出。针对上述问题, 本文提出了以下算法:

在检测阶段, 通过几何透视的先验信息提取图像交通标志感兴趣区域。只需少量的图像块就能保留绝大部分交通标志信息, 大大降低后续待检测图像块的冗余, 提高检测效率。

在训练阶段, 我们对训练集进行预处理, 提出基于交通标志的随机裁剪算法对包含小交通标志的区域过采样丰富训练集, 改善了对小目标检测的性能。

此外, 本文改进了 Tiny YOLOv3^[3] 算法。Tiny YOLOv3^[3] 是 YOLOv3 的简化版本, 其内存占用率低, 检测速度快。但在小目标密集的全景交通场景中 Tiny YOLOv3 漏检率较高。针对这一点, 我们提出 Improved-Tiny-YOLOv3 算法。充足的浅层特征对小目标检测有利, 我们通过 Passthrough^[13] 层将重组的浅层特征与深层特征相融合, 实现了更利于小目标检测的第三个尺度。并用更符合目标框回归机制的 CIoU^[14] (Complete Intersection over Union, CIoU) 改进了模型损失函数中的边界框位置误差项, 进一步提高网络性能。

1 方法设计

1.1 训练集图片预处理

在目标检测领域, 目标尺寸是根据其所占像素来

定义的, 可分为小 ($p < 32 \times 32$)、中 ($32 \times 32 < p < 96 \times 96$)、大 ($p > 96 \times 96$)。文献[15]提出对训练集小目标过采样能显著改善模型检测性能, 为改善小目标的检测性能, 本文提出基于交通标志的随机裁剪算法生成训练集。算法分为固定裁剪和随机裁剪两个部分。

固定裁剪确保原训练集中所有交通标志都至少被一个图像块完整包含, 在原训练集的高分辨率 (如 $2\ 048 \times 2\ 048$) 图像中裁剪以交通标志为中心的图像块, 图像块大小由网络模型输入 (如 416×416) 决定, 裁剪位置由式(1)、式(2)给出。

$$x = \begin{cases} x_c - T/2 & T/2 \leq x_c \leq W - T/2 \\ 0 & x_c < T/2 \\ W - T & x_c > W - T/2 \end{cases} \quad (1)$$

$$y = \begin{cases} y_c - T/2 & T/2 \leq y_c \leq H - T/2 \\ 0 & y_c < T/2 \\ H - T & y_c > H - T/2 \end{cases} \quad (2)$$

式中: x 和 y 表示图像块的在被裁图像中的左上角坐标; x_c 和 y_c 为交通标志中心的坐标; H 和 W 表示被裁图像的高和宽; T 表示图像块宽度。

随机裁剪在高分辨率图像和其下采样的低分辨率 (如 $1\ 040 \times 1\ 040$) 图像中进行。由交通标志的尺寸决定对其过采样次数。本文对小目标和中等目标分别过采样 2、1 个图像块。裁剪位置由式(3)求出。

$$x = x_c + T \times t_x, y = y_c + T \times t_y \quad (3)$$

式中: t_x 和 t_y 表示计算机产生的伪随机数, 服从以下分布:

$$t_x \sim U(t_{x_{\min}}, t_{x_{\max}}), t_y \sim U(t_{y_{\min}}, t_{y_{\max}}) \quad (4)$$

式中: $t_{x_{\min}}$ 、 $t_{x_{\max}}$ 和 $t_{y_{\min}}$ 、 $t_{y_{\max}}$ 分别代表横坐标和纵坐标方向产生随机值的范围。

$$t_{y_{\min}} = \max\left\{-1, \frac{-y_c}{T}\right\}, t_{y_{\max}} = \min\left\{0, \frac{H - T - y_c}{T}\right\} \quad (5)$$

$$t_{x_{\min}} = \max\left\{-1, \frac{-x_c}{T}\right\}, t_{x_{\max}} = \min\left\{0, \frac{W - T - x_c}{T}\right\} \quad (6)$$

通过随机函数产生随机值, 以被裁交通标志的中心坐标为基准, 产生随机抖动, 计算出图像块在被裁图像中的位置坐标, 使得图像块至少包含交通标志的 1/4 区域。最后将我们裁得的图像块制作成用于模型训练的训集。

1.2 基于几何透视的图像分块算法

在检测阶段, 本文提出基于几何透视的图像分块算法, 以图像中几何透视的先验信息为指导, 对全景交通图像进行分块。

1) 交通标志在全景图像中的分布规律。几何透视法告诉我们图片的成像存在一定特点, 即同一目标

由于与拍摄设备的相对距离不同,该目标的大小与位置在成像后的图片中呈现一定规律^[16]。在交通全景图像中,越接近地平线(地面与天空的分割线)的位置交通标志数量就越密集,尺寸越小,而距离地平线越远的位置,交通标志数量越少,尺寸相对越大。

为了探索交通标志在图像中的分布规律,对数据集 Tsinghua-Tencent 100K 训练集中 9 000 余幅图片进行统计分析。如图 1 所示,横坐标表示相对地平线的归一化范围区间,纵坐标为区间内交通标志数量。可以看出地平线下方 0.04 到地平线上方 0.16 范围内包含的完整交通标志数量最多,共有 13 145 个,占总体的 84.9%,该区域约占全图面积的 1/5。我们将该区域称为交通标志感兴趣区域。

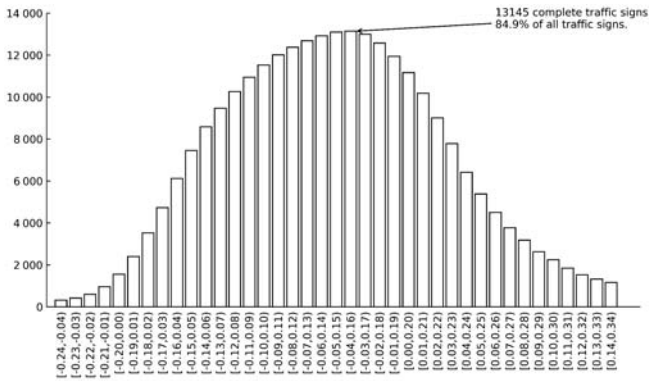


图 1 地平线附近交通标志数量

图 2 是交通标志感兴趣区域的交通标志数量与总体交通标志的数量对比图,交通标志感兴趣区域中尺寸为(0,32]的完整交通标志共有 6 227 个,占总体的 96.53%,而尺寸为(96,400]的交通标志共有 290 个,仅占总体 27.02%。从这一结果可以看出地平线附近的交通标志小而密集,而大交通标志分布较少。

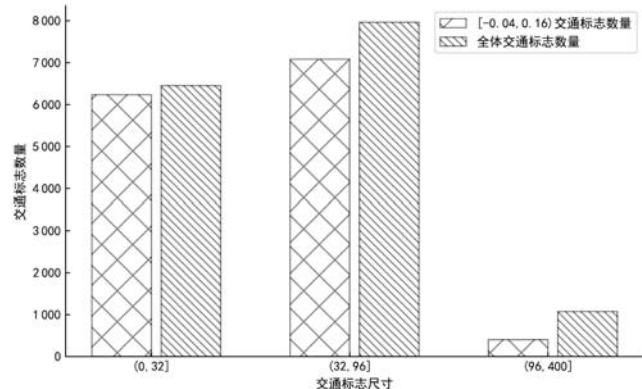


图 2 不同尺寸下交通标志数量与总体数量对比

2) 基于几何透视的图像分块算法。成功检测地平线是本文图像预处理方法的基础,当今对地平线检测的研究已经相对成熟,通过 CNN^[17]对 TT100K 数据集中图像的地平线检测有不错的准确率,且只需花费很短时间。

实际上,由于车载摄像头的安装位置固定,所以其拍摄图像的地平线位置也相对固定,为降低图像预处理时间复杂度,本文通过统计分析 TT100K 训练集图像交通标志在图像中的分布,得出类似结论。统计结果显示在分辨率为 2 048 × 2 048(图像左上角纵坐标 Y=0)的图像中。Y=620 到 Y=1 036 的区域包含完整交通标志数量最多,累计达到 13 112 个,同时在 Y>1 600 区域几乎不存在交通标志的任何一部分。

我们以该先验信息为指导,设计基于几何透视的滑动窗口算法。如图 3 所示,滑动窗口尺寸为模型输入尺寸(如 416 × 416),图像块由原始高分辨率(如 2 048 × 2 048)图像和其下采样一倍(如 1 040 × 1 040)的图像中产生。处理高分辨率图像时,将滑窗置于图像中交通标志感兴趣区域,由左到右进行滑动窗口算法,将其命名为待检测图像块一;处理下采样图像时,首先根据统计结果裁除图像中不存在交通标志的区域,紧接着着重对交通标志感兴趣区域之外的剩余区域进行滑动窗口处理。将其命名为待检测图像块二。

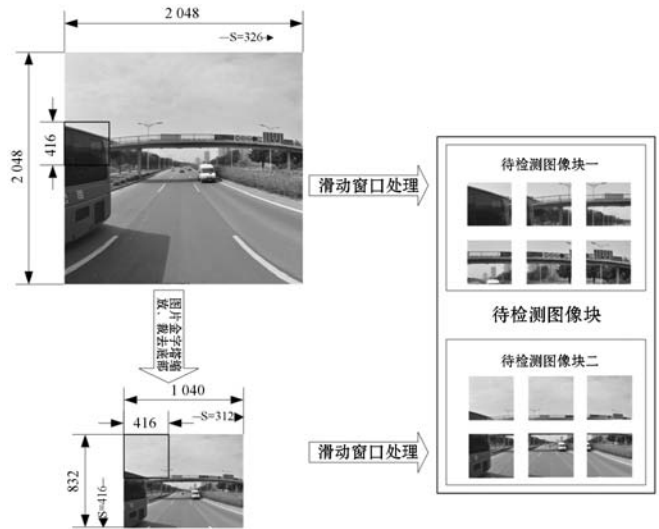


图 3 待测图像分块方法

待检测图像块一所对应区域为原图中交通标志小而密集的区域,所以该区域应保留原始分辨率,避免压缩引起小目标失真。待检测图像块二用于检测被图像块一分隔造成检测失败的交通标志以及图像块一区域外的交通标志。因为图像块一区域外目标数量少,且尺寸较大,所以可以通过图像压缩来降低检测的时间复杂度。此外,根据训练集统计信息,在全景交通图像的 Y>1 600 区域不包含交通标志的任何一部分,我们对该区域不予检测。

1.3 基于 Tiny-YOLOv3 算法的改进

Tiny-YOLOv3 算法是一个轻量级目标检测算法,相比 YOLOv3 算法有更快的检测速度和更低的显存占用率。Tiny-YOLOv3 网络结构如图 4 虚线外部分,其

主干网络由 7 个卷积层和 6 个最大池化层组成。由上文可知,全景交通场景中目标密集,但 Tiny-YOLOv3 对小目标检测精度低,漏检率高。为改善这些问题,本文在原有网络的基础上从网络结构和损失函数两个方面对其进行改进。在网络结构方面,我们在原有两个 YOLO 层(13×13,26×26)的基础上增加了 52×52 的 YOLO 层。卷积神经网络中浅层特征的感受野较小且包含更多图像的细节信息,对小目标检测有利。为了使得增加的 YOLO 层对小目标敏感,我们使用 Pass-through 对浅层的大分辨率特征进行特征重组,保留充足的浅层网络特征。如图 4 虚线中部分,首先我们将 104×104×64 的浅层特征通过 Passthrough 层特征重组为 52×52×256,然后将重组特征与 52×52×128 的深层特征进行融合,紧接着与第五层最大池化输出的 52×52×64 的特征融合成 52×52×448 的特征层,最后通过一个卷积操作输出为 52×52×150 的 YOLO 层。

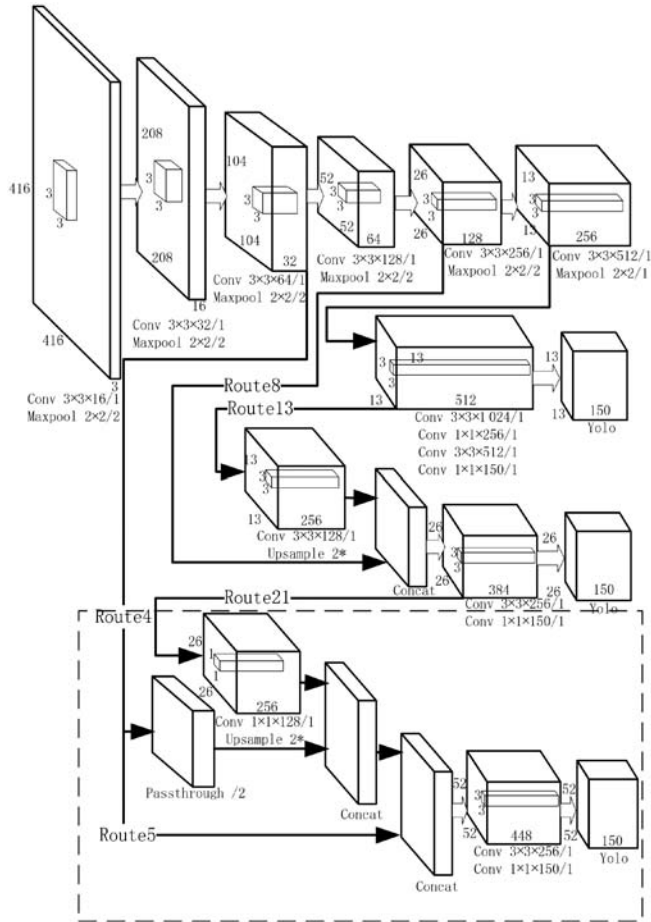


图 4 改进的 Tiny-YOLOv3 算法网络结构图

在损失函数方面,我们用 CIoU 改进了模型损失函数中的边界框位置误差项。Tiny-YOLOv3 中使用的交并比(Intersection over Union, IoU)定义为:

$$U_{IoU} = \frac{|B \cap B^{gt}|}{|B \cup B^{gt}|} \quad (7)$$

式中: B 表示预测框; B^{gt} 表示为真实框。

CIoU 损失在考虑交并比损失的基础上,同时考虑了边界框的中心点距离以及边界框的长宽比。CIoU 损失函数定义为:

$$L_{CIoU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \quad (8)$$

式中: $\frac{\rho^2(b, b^{gt})}{c^2}$ 为中心点距离损失项,该项中的 b, b^{gt} 分别表示 B, B^{gt} 的中心点; $\rho(\cdot)$ 表示欧氏距离; c 表示 B, B^{gt} 的最小外接矩形的对角线距离。这一项解决了 IoU 在预测框与真实框之间没交集时无法优化的问题,同时解决了真实框完全包裹预测框时,无法对其相对位置进行优化的问题。

αv 为长宽比损失项, α 是正权衡参数, v 是衡量长宽比一致性的参数,其中 v 定义为:

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right) \quad (9)$$

α 定义为:

$$\alpha = \frac{v}{(1 - U_{IoU}) + v} \quad (10)$$

综上所述,CIoU 相比 IoU 更符合目标框回归的机制,使得目标框回归变得更加稳定。

1.4 检测流程

本文算法的检测流程如图 5 所示。

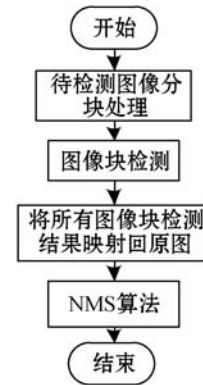


图 5 图像检测流程

首先对待测图像进行基于几何透视的图像分块处理,将滑窗产生的图像块串行输入到网络模型中进行交通标志检测与识别,然后将每个图像块的检测结果映射回原图中,最后通过非极大值抑制筛选出每个交通标志的最优检测结果。

2 模型训练

2.1 数据集介绍

本文选用的是 Tingshua-Tencent 100K 数据集,它是清华可视媒体研究中心制作的腾讯街景全景的交通标志数据集。本文选择其中标注的 9 000 余幅图片作为

训练集,3 000 余幅作为测试集,每幅图片的分辨率是 $2\ 048 \times 2\ 048$,这些图片覆盖了不同光线和天气状况,交通标志的分辨率在 $8 \times 8 \sim 400 \times 400$ 之间。为了保证每一类交通标志有充足的数据量,避免模型过拟合,本实验选取单类别交通标志数量大于 100 的交通标志作为检测和识别的对象,共 45 类。如图 6 所示。我们对 TT100K 训练集中的图像通过本文提出的训练集预处理算法处理,生成模型真正使用的训练集,将生成的训练集制作成 Pascal VOC 格式用于模型训练。



图 6 45 类交通标志图像及相应的类别名称

2.2 实验环境与参数设置

本文实验在一台 Linux 计算机上进行,该计算机处理器为 Core i7 7700k,显卡 GeForce GTX 1080Ti 11 GB 显存、32 GB 运行内存。网络的训练与图像预测在基于 C 语言的 darknet 深度学习框架上完成,训练的初始学习率为 0.001,最大迭代次数为 250 000 次, batch 为 64,在迭代到 200 000 以及 225 000 分别进行一次学习率衰减,衰减率为 0.9。通过对训练集进行聚类生成 9 个 anchor box,分别是(16,19)、(24,26)、(31,38)用于 52×52 特征图,(43,42)、(52,57)、(64,77)用于 26×26 特征图,(89,84)、(108,121)、(168,163)用于 13×13 特征图。

3 实验结果与分析

3.1 待测图像预处理方法对比

为了验证本文提出的待测图像预处理方法能在保证检测准确度的同时大大减少后续检测区域,我们对比了不同图像预处理方法所需检测的图像块的数量,以及在使用不同图像预处理方法时,网络模型 Improved-Tiny-YOLOv3 的 mAP 指标。对比结果如表 1 所示,由上文知图像块一区域仅占全图面积的 20%,却达到了 80% 多的 mAP,可看出图像块一有效提取了图像中交通标志感兴趣区域。由图像块二可看出,图像下采样后也可有效地检测出图像中的大部分交通标

志,达到在交通标志较少区域降低检测冗余的效果。文献[11]方法为了降低检测时间,仅在原始图像下采样一倍的图像中进行图像分块,图像压缩导致了小交通标志难以被检测。文献[10]也未考虑交通标志分布的先验信息,它在原始图像及其下采样图像上产生大量图像块,虽然其 mAP 稍高出本文算法,但检测冗余过大,难以应用在实际场景中。相比文献[9]和文献[10],本文方法采用分治的策略,在交通标志感兴趣区域采用原始分辨率图像,确保充足的交通标志信息,在交通标志少且较大的区域,将原图进行压缩,达到加速检测的目的,较好地权衡了图像块冗余度与交通标志信息保留程度之间的矛盾。

表 1 不同图像块裁剪方法对比

待测图像预处理方法	所需检测图像块数量	mAP/%
图像块 1	6	80.5
图像块 2	6	83.9
本文方法(图像块 1 + 图像块 2)	12	92.7
文献[10]方法	12	87.6
文献[9]方法	66	94.3

3.2 网络改进效果测试

为验证本文对网络结构和损失函数改进的有效性,进行如下消融实验。为确保实验对比的准确性,每个模型采用相同的训练集,测试集,相同的训练方式,以及测试集的预处理方法。以 Tiny-YOLOv3 为基准,对比了改进前后的平均检测精度(Mean Average Precision, mAP),IoU 阈值取 0.5,以及计算代价和模型尺寸。如表 2 所示,通过 Passthrough 层融合浅层信息将检测层提升到三个,大大提升了模型的 mAP,但同时增加了一定计算代价和模型权重。通过用 CIoU 改进损失函数,有效提升了模型的 mAP 且没有产生任何代价。此外改进网络模型的检测时间为 0.004 1 s/图像块。

表 2 网络改进效果对比

网络结构改进	损失函数改进	mAP /%	计算代价 /B	模型尺寸 /M
否	否	87.6	5.52	35.1
是	否	91.7	11.35	39.5
是	是	92.7	11.35	39.5

3.3 方法对比

为了说明本文方法在时间和性能上的优势,我们对比了不同算法间的 mAP 与平均准确度(Average Precision, AP),IoU 阈值取 0.5,以及检测识别的时间指

标。从表3中可以看出本文算法的 mAP(92.7%) 远远高于其他方法。虽然本文与文献[10]方法都是将原图像预处理为 12 个图像块进行检测,但不论从检测总时间上还是准确度上本文方法均优于文献[10]。考虑到边缘设备资源受限问题,YOLOv3 与 YOLOv4 输入为原图下采样至 416×416 的图像,虽然本文方法检测时间(0.049 s)慢于 YOLOv4(0.017 s)与 YOLOv3(0.023 s),但在 mAP 上分别高出 30.4 个百分点和 27.7 百分点。

表3 不同算法的性能对比

方法	检测总时间/s	mAP/%
Faster RCNN ^[10]	5 ~ 10	59.1
YOLOv4 ^[4]	0.017	62.3
YOLOv3 ^[3]	0.023	65.0
文献11方法	0.057	84.6
本文方法	0.049	92.7
TT100K Benchmark ^[10]	10	83.6

图7反映了45类交通标志的 AP 对比,可以看出本文方法的 AP 值几乎始终在图像最上方。实验结果充分说明了本文提出方法的有效性。

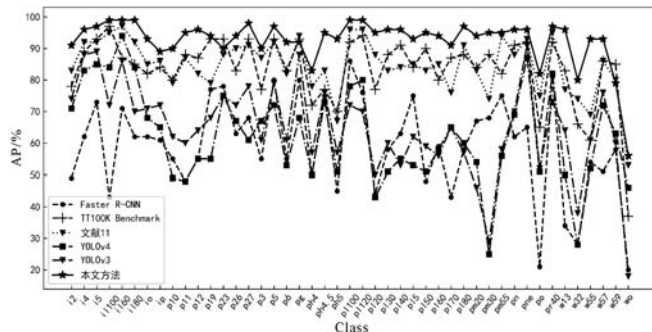


图7 不同算法的45类交通标志 AP 对比图

4 结语

本文考虑了几何透视的先验信息,对图像预处理,提取出了图像中的交通标志感兴趣区域,大大地减少了需要检测的图像块数量。通过用 CIoU 改进了 Tiny YOLOv3 的损失函数,使用 Passthrough 层充分融合了浅层特征信息,将二尺度的检测层提升为三尺度,使其能更好地检测小目标,提升了模型的 mAP。由于本文方法采用轻量级的网络模型,这对后续在性能较差的边缘设备上较好检测全景高清图像提供了可能性。

参考文献

[1] 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 Intelli-

gence, 2015, 39(6):1137-1149.

- [2] Cai Z, Vasconcelos N. Cascade R-CNN: Delving into high quality object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2018:6154-6162.
- [3] Redmon J, Farhadi A. YOLOv3: An incremental improvement[EB]. arXiv:1804.02767,2018.
- [4] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: Optimal speed and accuracy of object detection[EB]. arXiv:2004.10934,2020.
- [5] Liu W, Anguelov D, Erhan D, et al. SAD: Single shot multi-box detector[C]//Proceedings of the 14th European Conference on Computer Vision. Springer,2016:21-37.
- [6] Houben S, Stallkamp J, Salmen J, et al. Detection of traffic signs in real-world images: The German traffic sign detection benchmark[C]//International Joint Conference on Neural Networks. IEEE, 2013:1-8.
- [7] Zhu Z, Liang D, Zhang S, et al. Traffic-sign detection and classification in the wild[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016:2110-2118.
- [8] Yang T, Long X, Sangaiah A K, et al. Deep detection network for real-life traffic sign in vehicular networks[J]. Computer Networks, 2018, 136(8):95-104.
- [9] Meng Z, Fan X, Chen X, et al. Detecting small signs from large images[C]//2017 IEEE International Conference on Information Reuse and Integration (IRI). IEEE,2017:217-224.
- [10] Meng X, Zhang X, Yan K, et al. Real-time detection and recognition of live panoramic traffic signs based on deep learning[C]//2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS). IEEE, 2018:584-588.
- [11] 张淑芳,朱彤.基于残差单发多框检测器模型的交通标志检测与识别[J].浙江大学学报(工学版),2019,53(5):127-136.
- [12] 刘胜,马社祥,孟鑫,等.基于多尺度特征融合网络的交通标志检测[J].计算机应用与软件,2021,38(2):158-164.
- [13] Redmon J, Farhadi A. YOLO9000: Better, faster, stronger[C]//IEEE Conference on Computer Vision & Pattern Recognition,2017:6517-6525.
- [14] Rezatofighi H, Tsoi N, Gwak J Y, et al. Generalized intersection over union: A metric and a loss for bounding box regression[EB]. arXiv:1902.09630,2019.
- [15] Kisantal M, Wojna Z, Murawski J, et al. Augmentation for small object detection[EB]. arXiv:1902.07296,2019.
- [16] 何援军.透视和透视投影变换——论图形变换和投影的若干问题之三[J].计算机辅助设计与图形学报,2005(4):110-115.
- [17] Workman S, Zhai M, Jacobs N. Horizon lines in the wild[EB]. arXiv:1604.02129,2016.