

基于改进YOLOv5的交通标志检测算法研究

刘晓倩, 李士心*, 李保胜, 董秀焕, 高雪苹, 陈东园, 张美会

天津职业技术师范大学电子工程学院, 天津

收稿日期: 2022年8月19日; 录用日期: 2022年9月16日; 发布日期: 2022年9月26日

摘要

针对交通标志检测网络复杂度高、计算量大、难以进行有效部署的问题, 提出一种基于改进YOLOv5的交通标志检测算法。该算法首先使用轻量级卷积神经网络MobileNetV3对原YOLOv5网络的主干特征提取网络进行替换; 使用深度可分离卷积替换部分标准卷积, 来缩减模型计算量; 为保证模型检测精度, 引入注意力机制CBAM模块。实验表明, 改进后的轻量化网络模型大小只为原YOLOv5模型大小的61%, 整理后TT100K数据集上的mAP达到了89.2%, FPS达到了25.1帧/毫秒。本文提出的算法在保证检测精度的前提下, 大幅度减少了模型参数数量和计算量, 并提高了检测速度。

关键词

目标检测, 交通标志, YOLOv5算法, 轻量化

Research on Traffic Sign Detection Algorithm Based on Improved YOLOv5

Xiaoqian Liu, Shixin Li*, Baosheng Li, Xiuhuan Dong, Xueping Gao, Dongyuan Chen, Meihui Zhang

College of Electronic Engineering, Tianjin University of Technology and Education, Tianjin

Received: Aug. 19th, 2022; accepted: Sep. 16th, 2022; published: Sep. 26th, 2022

Abstract

Aiming at the problems of high complexity, large amount of computation and difficulty in effective deployment of traffic sign detection network, a traffic sign detection algorithm based on improved YOLOv5 was proposed. Firstly, the lightweight convolutional neural network MobileNetV3 was used to replace the main feature extraction network of the original YOLOv5 network. And the standard

*通讯作者。

convolution was replaced by depth-separable convolution to reduce the computational cost of the model. To ensure the accuracy of model detection accuracy, the attention mechanism CBAM module is introduced. The experimental results show that the size of the improved lightweight network model is only 61% of the size of the original YOLOv5 model, the mAP on TT100K dataset reaches 89.2%, and the FPS reaches 25.1 frames per millisecond. The algorithm proposed in this paper can greatly reduce the number of model parameters and the calculation amount, and improve the detection speed under the premise of ensuring the detection accuracy.

Keywords

Object Detection, Traffic Sign, YOLOv5 Algorithm, Lightweight

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着城市化建设和道路网的不断发展完善, 车辆数量持续增加, 这也导致堵车和车祸事故的增多[1], 严重危害人民的生命安全。随着科技的发展, 自动驾驶技术也得到越来越广泛的关注, 尤其是道路交通标志识别系统引发了研究热潮。作为自动驾驶的关键环节, 交通标志识别系统可事先把道路标识信息迅速精确地提供给司机, 可以对整个交通系统起到指示和导航作用。

针对交通标志的检测方法主要分两大类: 基于传统特征提取的方法和基于深度学习的方法。基于传统特征提取的方法主要从交通标志的颜色、轮廓和形状等方面考虑, 传统方法提取的特征表达能力有限, 在复杂环境下识别准确率低、识别速度慢, 不能满足检测需求。自 2012 年[2]深度学习模型的检测应用在 Imagenet 大规模视觉识别挑战赛取得巨大的成功后, 基于深度学习的方法逐渐成为交通标志检测的主流方法。基于深度学习的目标检测算法有两类[3], 一类是两阶段算法, 分两步进行检测: 产生候选区域、进行分类和定位, 其中典型的算法有 R-CNN [4]、Fast RCNN [5]、Faster R-CNN [6]、Mask-RCNN [7]等; 另一类是单阶段算法, 该类算法模型可以直接进行类别的分类与定位, 省去了产生候选区域的阶段, 其以 YOLO [8]系列算法和 SSD [9]算法为代表。Zhang 等人[10]将卷积网络应用于交通标志检测中, 取得了不错的效果。P Cheng [11]等针对小目标检测提出了 Faster R-CNN 交通标志检测方法, 其在交通标志数据集上的检测较好。

单阶段检测模型在实时性与准确性上都有了较大提升, 但是依旧存在计算量大、检测时间长、车载移动端部署设备受限等缺陷, 为了减小计算量、提高检测效果, 本文提出了一种兼顾精度和速度的基于 YOLOv5 的改进交通标志检测算法, 首先, 使用 MobileNetV3 轻量级网络替换 YOLOv5 的主干网络, 其次, 在特征加强提取网络部分使用深度可分离卷积, 最后, 引入即插即用的注意力机制模块 CBAM, 在引入极少参数的情况下提升算法的检测精度, 并使用数据集验证了算法的有效性。

2. YOLOv5 算法

YOLOv5 算法[12]具有简洁、快速、易部署的优点, 被广泛应用于工业领域的目标检测、跟踪、分割等场景中。YOLOv5 网络可以分为输入端、Backbone、Neck 和 Head。图 1 为 YOLOv5 算法的网络框架示意图。

深度可分离卷积中深度卷积核大小为 $D_k \times D_k$ 时，深度可分离卷积计算量的下降倍数，即标准卷积与深度可分离卷积的运算计算量比值为：

$$J = \frac{D_k \cdot D_k \cdot M \cdot D_f \cdot D_f + M \cdot N \cdot D_f \cdot D_f}{D_k \cdot D_k \cdot M \cdot N \cdot D_f \cdot D_f} = \frac{1}{N} + \frac{1}{D_k^2} \approx \frac{1}{D_k^2} \quad (1)$$

由上式可以看出，当 $D_k = 3$ 时，相比于标准卷积，深度可分离卷积的计算量至少减少了 80%，极大的缩减了网络计算量。

MobileNetV3 保留了倒残差结构，引入了基于 SE 结构的轻量级注意力组件，弥补了轻量化模型精度过低的缺陷。此外网络改进了激活函数，利用 ReLU6 函数，提出了计算高效的 h-swish 函数。

3.2. 注意力机制

注意力机制来源于人类大脑对图像信息的处理。通过注意力在观察图像全局信息时，可以快速的锁定重点关注区域，自动屏蔽部分背景和冗余信息。延伸到神经网络中，注意力机制就是让网络自动关注到更需要关注区域的一个方式。注意力机制通常可以分为通道注意力机制、空间注意力机制，以及二者的结合。

CBAM [14]就是结合了空间和通道两个方面的注意力机制模块。通道注意力机制着重于全局信息，给定图像信息类别，空间注意力机制着重于局部信息，决定聚焦的位置。CBAM 结构如图 2 所示。

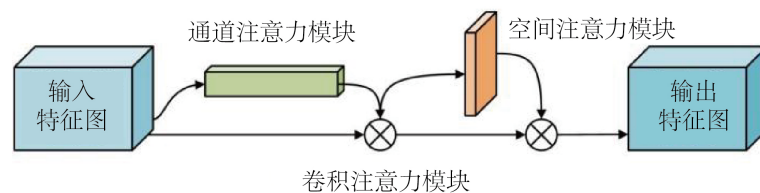


Figure 2. The structure of CBAM
图 2. CBAM 结构

3.3. 改进后 YOLOv5 网络

为了更快速高效地识别和定位交通标志，本文基于 YOLOv5 算法进行了三处改进：使用模型更小的轻量化模块 MobileNetV3 替换原网络的主干部分，来降低网络的体积和计算量；使用深度可分离卷积代替部分标准卷积，减少 Neck 部分的参数量；在 Neck 部分的检测层前使用注意力机制 CBAM 模块，来提升网络对某些特征和空间的注意力，以此加强网络对交通标志的特征提取能力。改进后的 YOLOv5 如图 3 所示。

4. 实验及结果分析

4.1. 交通标志数据集

为验证改进算法的效果，需要在数据集上进行验证。TT100K 是腾讯公司和清华大学联合开发的数据集[15]，几乎包括了真实场景下所有常见的高分辨率交通标志。本实验中，需对数据集进行适当处理，即将 TT100K 数据集格式转换为 xml 格式，按照交通标志的分类规则，将该数据集重新分为 mandatory、prohibitory 和 warning 三类，再将 xml 格式转换成 YOLOv5 对应标签格式。实验中共使用 9140 张数据集图片，其中包括训练集 5502 张，验证集 1834 张，测试集 1834 张。TT100K 数据集中小目标较多，可视化结果如图 4 所示。

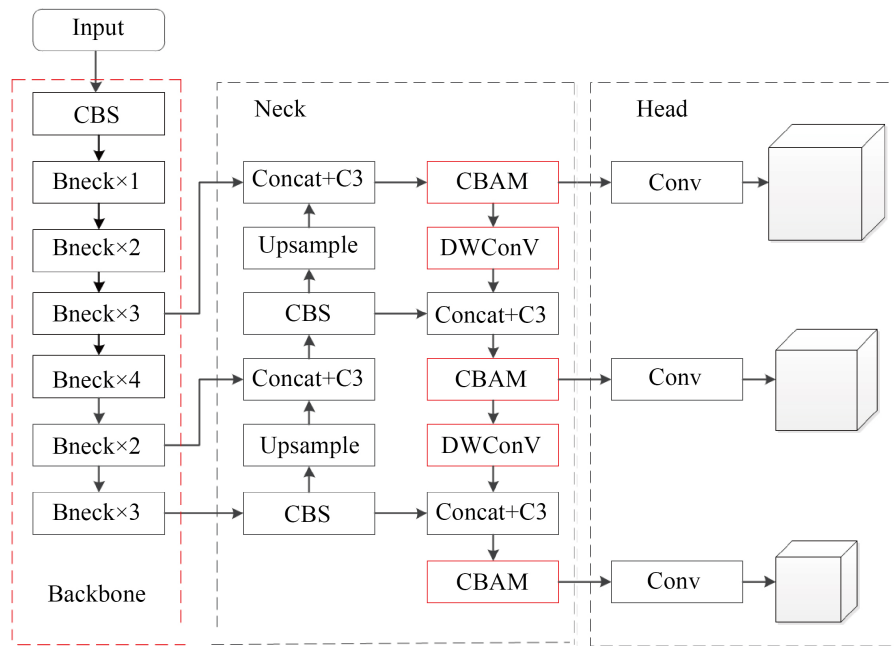


Figure 3. The improved YOLOv5 network
图 3. 改进后的 YOLOv5 网络

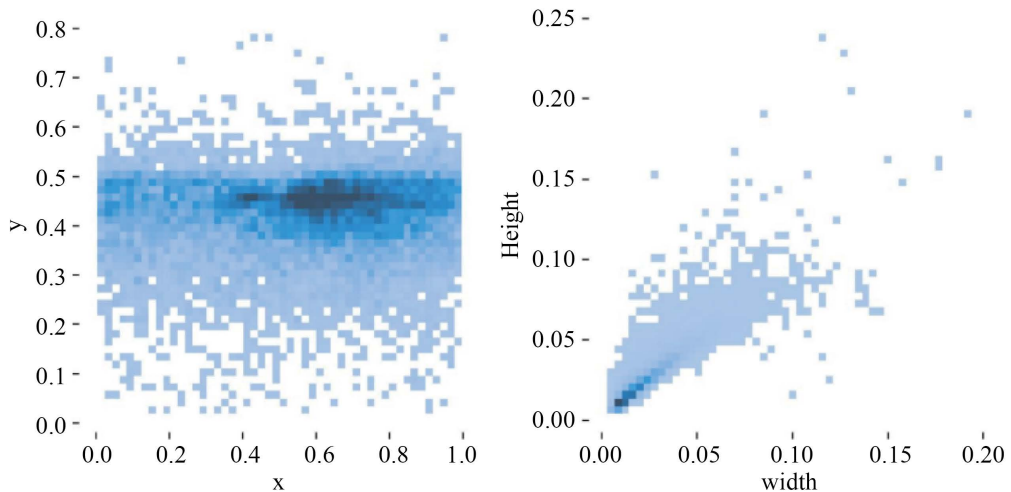


Figure 4. The result of the dataset target visualization
图 4. 数据集目标可视化结果

4.2. 评价指标

选用多种评估标准来衡量算法的有效性，包括 precision、recall、AP、mAP，其中 precision 代表模型的查准率，recall 代表模型的查全率，AP 表示平均精度，衡量的是模型对每个交通标志类别检测性能的优劣，其中 mAP 表示平均精度均值，衡量模型对所有交通标志类别检测性能的优劣。各评价指标的计算公式如下：

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3}$$

$$AP = \int_0^1 P(R) \tag{4}$$

$$mAP = \frac{\sum_{j=1}^c AP_j}{c} \tag{5}$$

式中 TP 为预测为正类样本且预测正确的个数，FP 为预测为正类样本且预测错误的个数，FN 表示预测为错误的正类样本的数量，c 表示类别总数，j 为当前类别的序号。

4.3. 模型训练与结果分析

本文采用的实验环境为，硬件配置：CPU 为(R) Core (TM) i9-10900X, 64GB RAM。GPU 采用了 NVIDIA GeForce RTX 3080。学习框架选用 PyTorch 1.11.1，编程语言为 Python。整个训练过程分为 300 个 epoch 进行，批次为 16，初始学习率为 0.001，动量大小为 0.735，权重衰减系数为 0.0005。

改进后的模型使用轻量化网络 MobileNetV3 替换主干网络使用深度可分离卷积替换部分标准卷积，并嵌入了 CBAM 模块。为验证模型改进后的效果，在训练集、验证集和测试集上进行实验，结果如表 1 所示。

Table 1. Experimental results

表 1. 实验结果

算法	MobileNetV3	DWConv	CBAM	mAP%	P	R	模型大小 (MB)	FPS (帧/ms)
YOLOv5	×	×	×	91.7	91.8	85.8	26.81	20.87
算法 1	√	×	×	88.9	90.6	83.6	18.74	24.10
算法 2	√	√	×	87.8	90.4	83.0	16.26	25.14
算法 3	√	√	√	89.2	91.1	83.2	16.42	25.06

从表 1 中可看出算法 1 相较原算法参数减少 30%，FPS 值提高 3.23，mAP 值略有下降。算法 2 在算法 1 基础上进一步减少了参数量，相较于原算法参数量减少 39%，FPS 值提高 4.27，但是 mAP 下降了 3.9%。算法 3，即本文最终改进后的 YOLOv5 算法，相较于算法 2，以极小的参数量换取了 mAP 的提高，算法 3 较原算法参数量减少 38%，模型大小仅有 16.42 MB，FPS 达到 25.06 帧/毫秒，虽然 mAP 值比原模型低 2.5%，但是模型复杂度大大降低，满足了交通标志的实时检测任务需求。

为了更直观地看到改进后算法的检测效果，在测试集上对算法进行测试。检测效果样例如图 5 和图 6 所示。

从图 5 可以看出，图中共有四个 prohibitory 类交通标志，改进后的 YOLOv5 算法相较原 YOLOv5 算法检测精度略有下降。从图 6 可以看出，图中共有三个 prohibitory 类交通标志，从图 6(a)可以看出，因为检测目标较小，原 YOLOv5 算法存在一个误检目标，由于注意力机制可以提高特征提取能力，使网络更关注局部位置信息，在图 6(b)中可以看出，改进后的 YOLOv5 算法并未检测出原算法的误检目标，检测效果较好。



Figure 5. Traffic sign detection renderings before and after improvement (1)
图 5. 改进前后交通标志检测效果图(1)



Figure 6. Traffic sign detection renderings before and after improvement (2)
图 6. 改进前后交通标志检测效果图(2)

5. 结论

本文针对交通标志检测网络计算量大、难以进行有效部署的问题，提出一种基于改进 YOLOv5 的交通标志检测算法。该方法在原 YOLOv5 模型的基础上，通过使用轻量化模块和注意力机制模块的方法来降低模型的复杂度。实验结果表明：该方法在保证检测精度的前提下，提高了检测速度，可满足交通标志实时检测的需求。但从实验结果看出，改进后的 YOLOv5 算法虽然大幅度降低了参数量，提高了检测速度，但检测精度却略有降低。后续将继续改进模型结构，在减小模型大小的同时，进一步提高模型检测精度。

项目基金

2021 年天津市研究生科研创新项目(2021YJSS224)。

参考文献

[1] 陈贺. 中国交通事故死亡状况及酒驾处罚对交通事故死亡率的影响[D]: [硕士学位论文]. 北京: 中国疾病预防控制中心

- 制中心, 2017.
- [2] Krizhevskya, Sutskever, I. and Hintong, E. (2017) ImageNet Classification with Deep Convolutional Neural Networks. *Communications of the ACM*, **60**, 84-90. <https://doi.org/10.1145/3065386>
 - [3] 曹前. 智能交通中的目标检测技术研究[D]: [硕士学位论文]. 广州: 华南理工大学, 2019. <https://doi.org/10.27151/d.cnki.ghnlu.2019.003535>
 - [4] Girshick, R., Donahue, J., Darrell, T., et al. (2014) Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 23-28 June 2014, 580-587. <https://doi.org/10.1109/CVPR.2014.81>
 - [5] Girshick, R. (2015) Fast R-CNN. *Proceedings of the IEEE International Conference on Computer Vision*, Santiago, 7-13 December 2015, 1440-1448. <https://doi.org/10.1109/ICCV.2015.169>
 - [6] Girshick, R., Donahue, J., Darrell, T., et al. (2014) Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. *IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 23-28 June 2014, 580-587. <https://doi.org/10.1109/CVPR.2014.81>
 - [7] He, K., Gkioxari, G., Dollár, P., et al. (2017) Mask R-CNN. *Proceedings of the IEEE International Conference on Computer Vision*, Venice, 22-29 October 2017, 2961-2969. <https://doi.org/10.1109/ICCV.2017.322>
 - [8] Bochkovskiy, A., Wang, C.Y. and Liao, H.Y.M. (2020) Yolov4: Optimal Speed and Accuracy of Object Detection.
 - [9] Liu, W., Anguelov, D., Erhan, D., et al. (2016) Ssd: Single Shot Multibox Detector. In: *European Conference on Computer Vision*, Springer, Cham, 21-37. https://doi.org/10.1007/978-3-319-46448-0_2
 - [10] Zhang, J.M., Xie, Z.P., Sun, J., et al. (2020) A Cascaded R-CNN with Multiscale Attention and Imbalanced Samples for Traffic Sign Detection. *IEEE Access*, **8**, 29742-29754. <https://doi.org/10.1109/ACCESS.2020.2972338>
 - [11] Cheng, P., Liu, W., Zhang, Y., et al. (2019) LOCO: Local Context Based Faster R-CNN for Small Traffic Sign Detection. In: *International Conference on Multimedia Modeling*, Springer, Cham, 329-341. https://doi.org/10.1007/978-3-319-73603-7_27
 - [12] Jocher, G., Nishimura, K., Mineeva, T., et al. (2020) YOLOv5. <https://github.com/ultralytics/yolov5>
 - [13] Sandler, M., Howard, A., Zhu, M., et al. (2018) MobileNetV2: Inverted Residuals and Linear Bottlenecks. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, 18-22 June 2018, 45610-4520. <https://doi.org/10.1109/CVPR.2018.00474>
 - [14] Woo, S., Park, J., Lee, J.Y., et al. (2018) CBAM: Convolutional Block Attention Module. *Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, 8-14 September 2018, 3-19. https://doi.org/10.1007/978-3-030-01234-2_1
 - [15] Zhu, Z., Liang, D., Zhang, S., et al. (2016) Traffic-Sign Detection and Classification in the Wild. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 2110-2118. <https://doi.org/10.1109/CVPR.2016.232>