

DOI:10.19651/j.cnki.emt.2108743

基于注意力机制的交通标志识别*

李宇琼 周永军 蒋淑霞 梁 杨

(中南林业科技大学机电工程学院 长沙 410000)

摘要: 针对实际场景中的交通标志大多小而密集,导致小目标交通标志识别准确度较低的问题,提出一种改进YOLOv5算法。首先将CBAM同时嵌入YOLOv5网络的Backbone和Head部分,以提升网络特征提取能力。其次为解决GIoU Loss可能造成的模型收敛速度较慢问题,改用DIoU Loss作为网络回归损失函数。实验结果表明,改进后的算法对于交通标志图像的识别平均准确率达到96.40%,相较于原算法有了6.83%的提升。最后为验证模型的实时可行性,在TX2嵌入式系统中利用本文改进YOLOv5算法对实景视频中的交通标志进行识别,结果表明本文改进算法能在嵌入式系统中流畅运行。

关键词: 交通标志识别;YOLOv5;CBAM;DIoU Loss;嵌入式系统

中图分类号: TP389.1 **文献标识码:** A **国家标准学科分类代码:** 510.1050

Traffic sign recognition based on attention mechanism

Li Yuqiong Zhou Yongjun Jiang Shuxia Liang Yang

(School of Mechanical and Electrical Engineering, Central South University of Forestry and Technology, Changsha 410000, China)

Abstract: Aiming at the problem of low accuracy of small target detection in traffic sign recognition tasks, which caused by that most of traffic signs in actual scene are small and dense, this paper proposes an improved YOLOv5 algorithm. firstly, embedding the CBAM into the Backbone and Neck of YOLOv5 network to improve the network feature extraction ability. and in order to solve the problem of slow network converge caused by GIOU Loss, DIoU Loss was used as the regression Loss function of the network. Experimental results show that the improved algorithm reaches 96.40% mAP in traffic sign recognition task, which is 6.83% higher than the original YOLOv5 algorithm. Finally, sending the improved network into TX2 embedded system to recognize traffic signs in real video, the result shows that the improved algorithm can run smoothly in embedded system.

Keywords: traffic sign recognition; YOLOv5; CBAM; DIoU loss; embedded system

0 引言

交通标志的识别可分为检测与分类两个环节^[1]:1)判断图片或视频中是否含有交通标志,并标定其位置;2)对所含交通标志的进行分类。

基于深度学习在计算机视觉领域的优异表现,交通标志识别研究学者纷纷将研究重点转向深度学习。Ciresan等^[2]将HOG(histogram of oriented gradient)^[3]特征和支持向量机(support vector machine, SVM)^[4]引入到卷积神经网络(central neural network, CNN)中,在CCTSDB数据集^[5]上进行实验,获得良好的识别效果;Qian等^[6]利用Fast R-CNN与两个CNN网络级联后生成的网络对交通

标志进行识别,获得了90.2%的准确率。但该网络对计算机硬件要求较高,无法对交通标志进行实时识别;杨梦梦^[7]对Faster R-CNN算法进行了改进,用ResNet-101代替VGG16作为其主干网络,使该网络具有较好的鲁棒性;王卜等^[8]以YOLOv3网络为基础,采用空间金字塔池化模块对特征图进行分块池化,同时引入FI模块对特征图进行信息融合,网络mAP达到了82%;陈立潮等^[9]根据胶囊网络和多尺度思想,在主胶囊网络中采取双通道池化优化动态路由算法,在GTRSB数据集上进行实验,实验表明该网络识别准确率与识别速度均良好。

目前交通标志识别网络可分为一阶段式网络和两阶段式网络,两类网络各有不足。由于需要较强的计算能力,因

收稿日期:2021-12-31

* 基金项目:湖南省自然科学基金(2019JJ60076)、湖南省科技创新项目(2018NK2065)、湖南省教育厅重点研究项目(16A220)、湖南省自然科学基金面上项目(2017JJ2403)资助

而两阶段式网络难以对目标进行实时识别。而一阶段式网络虽然有着较快的识别速度,但识别准确率相对较低。因此本文以YOLOv5为基础,在其Backbone和Head部分嵌入CBAM(convolutional block attention module)^[10],以增强网络提取特征的能力。同时采用DIoU Loss^[11]作为回归损失函数,以解决GIoU Loss^[12]所造成的模型收敛速度较慢问题。

1 YOLOv5 网络模型简介

根据深度和宽度的不同,YOLOv5可分为YOLOv5s、YOLOv5m、YOLOv5l和YOLOv5x。出于对模型大小的

考虑,本文选用YOLOv5s作为识别网络。

图1为YOLOv5网络结构,网络分为Input、Backbone、Neck、Prediction四部分。Backbone部分,网络主要通过对输入图像进行多次卷积和Concat操作来实现3种不同尺度特征图的输出。Neck部分将路径聚合网络(perceptual adversarial network,PAN)^[13]添加到特征金字塔网络(feature pyramid network,FPN)^[14]后,以便更好地对网络不同主干层和不同检测层之间的信息进行融合。同时将CSP1_X和CSP2_X两种跨阶段网络(cross stage parital,CSPNet)^[15]分别应用于网络的Backbone和Head部分,极大地提升了网络特征融合能力。

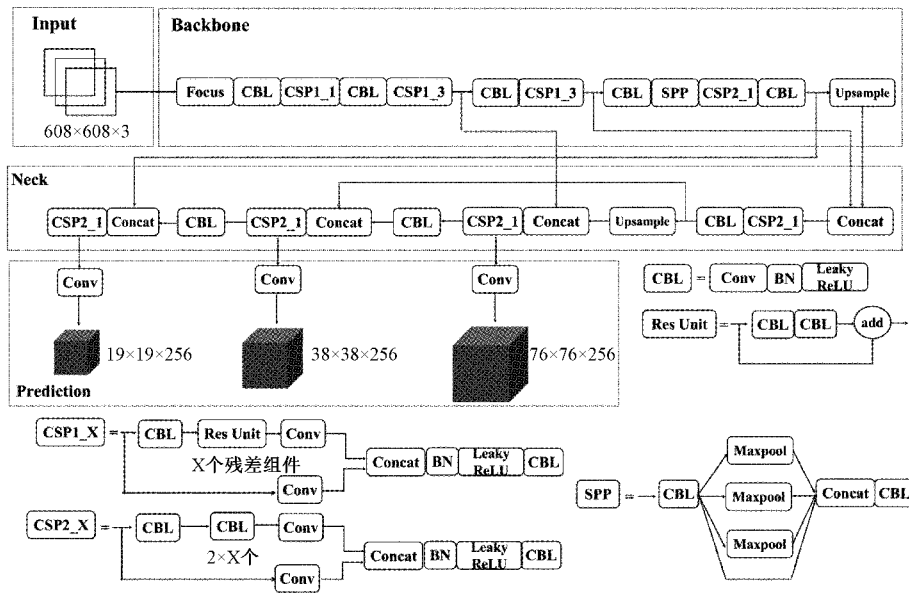


图1 YOLOv5网络结构

2 改进网络结构

2.1 基于CBAM注意力模块的YOLOv5算法

如图2所示,CBAM由通道注意力模块(channel attention module,CAM)和空间注意力模块(spatial attention module,SAM)组成,这两个模块通过对权重的重新分配使网络能更好地学习到原图中目标物体特征信息及其所在位置。

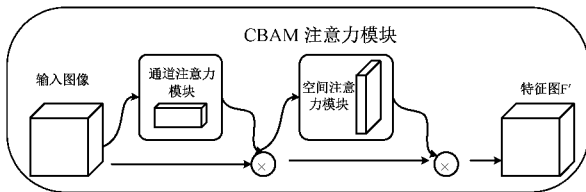


图2 CBAM结构图

图3和4分别为通道模块和空间模块结构图。这两个模块都先借助平均池化(average pooling)和最大池化(max pooling)以消除通维度和空间维度对于输入特征图的影响,

获得两个一维特征图。再对其进行卷积操作和非线性处理获得通道关注度 $M_c(F) \in R^{c \times 1 \times 1}$ 和空间关注度 $M_s(F) \in R^{H \times W}$ 。最后以所获得的关注度对通道权重进行再分配,得到最终特征图。 $M_c(F) \in R^{c \times 1 \times 1}$ 和 $M_s(F) \in R^{H \times W}$ 的计算公式如下:

$$M_c(F) = \sigma((W_1(W_0(F_{avg}^c))) + W_2(W_0(F_{max}^c))) \quad (1)$$

式中: σ 为sigmoid激活函数, $W_0 \in R^{c/r \times w}$ 、 $W_1 \in R^{c \times c/r}$ 为共享神经网络的权重。

$$M_s(F) = \sigma(f^{7 \times 7}([F_{avg}^s; F_{max}^s])) \quad (2)$$

式中: $f^{7 \times 7}$ 为卷积核大小为7x7的卷积核。

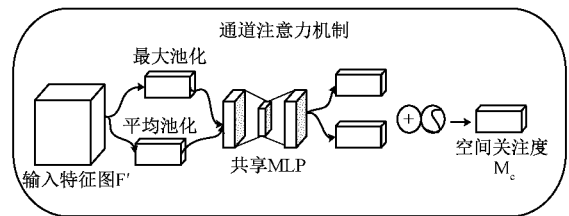


图3 CAM结构图

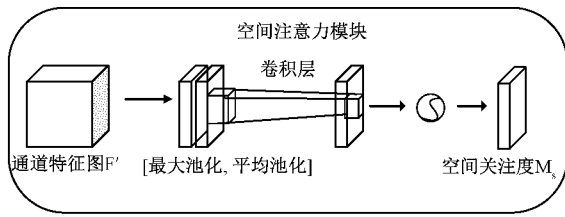


图 4 SAM 结构图

将嵌入 CBAM 的 YOLOv5 网络命名为 CBAM-YOLOv5,其结构简图如图 5 所示。

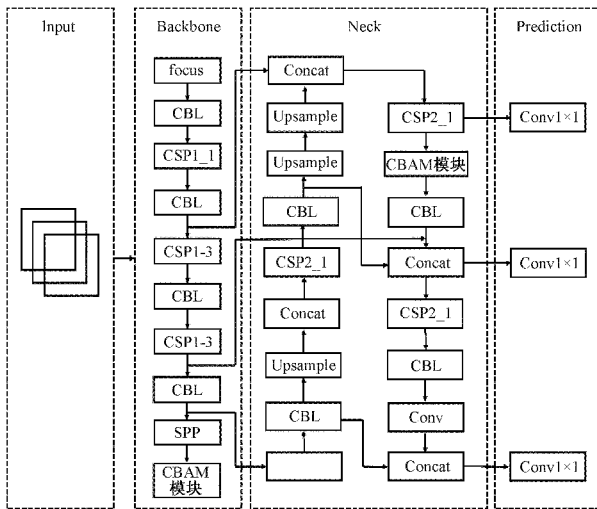


图 5 CBAM-YOLOv5 结构图

2.2 损失函数的改进

YOLOv5 网络采用 Giou Loss 作为 bounding box 的回归损失函数,其计算公式如下:

$$\begin{aligned} \ell_{GIoU} &= 1 - IoU + \frac{A^c - B \cup B^{gt}}{A^c} = \\ &= 1 - \frac{B \cap B^{gt}}{B \cup B^{gt}} + \frac{A^c - B \cup B^{gt}}{A^c} \end{aligned} \quad (3)$$

式中: B 、 B^{gt} 分别为预测框和真实框。

由图 6、7 可知,当真实框和目标框呈包含关系时, Giou 和 IoU 的损失值相等,此时两框的相对位置关系无法被反映出来。同时在训练过程中, Giou Loss 选择扩大预测框的大小来增大与真实框的交集,来最大化两框之间的重叠区域,因此网络收敛变慢。

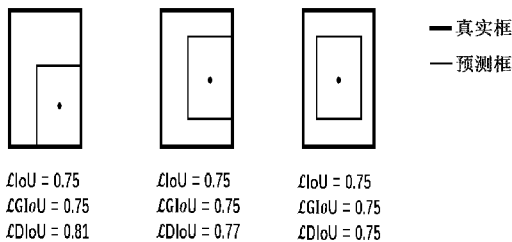


图 6 三种指标关系示意图

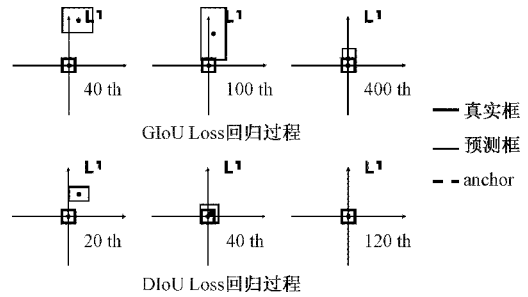


图 7 Giou Loss 和 DIoU Loss 回归过程图

针对上述问题,本文改用 DIoU Loss 作为网络回归损失函数,其公式如下:

$$\begin{aligned} \ell_{DIoU} &= 1 - IoU + \frac{\rho^2(b - b^{gt})}{c^2} = \\ &= 1 - \frac{B \cap B^{gt}}{B \cup B^{gt}} + \frac{\rho^2(b - b^{gt})}{c^2} \end{aligned} \quad (4)$$

式中: b 和 b^{gt} 分别为预测框和真实框的中心点, c 为最小外接矩 C 的对角线距离, ρ 表示 b 和 b^{gt} 这两个中心点的欧式距离。

3 实验结果及分析

3.1 数据集及评价指标

本文选取 TT100K 和 CCTSDB 两个数据集中部分图片,并加入由网络爬虫和实地拍摄所得的交通标志图片,最终获得包含禁止 (prohibitory)、警告 (warning)、指示 (mandatory) 三类交通标志图像共 12 360 张。

为更好地对模型进行评价,选用 FPS 和 mAP 来评估模型识别速度和准确度。FPS 为网络每秒识别的图像数,其数值越大,表明识别速度越快。mAP 为 P-R 曲线所围成面积的平均值,反映出网络对于目标物体的识别准确度。

3.2 实验环境

本文采用 Pytorch 深度学习框架,python 编程语言,Windows10 操作系统和 GeForce RTX 2080 显卡。具体实验配置如表 1 所示。

表 1 实验环境配置

参数	配置
GPU	GeForce RTX 2080
语言	Python3.7
系统环境	Windows10
加速环境	CUDA10.0

本文设置初始学习率为 0.01,在 30 epoch 和 45 epoch 时,学习率分别降至 0.001 和 0.000 1。batch size 设为 16,在 50 个 epoch 后停止训练。

3.3 实验结果分析

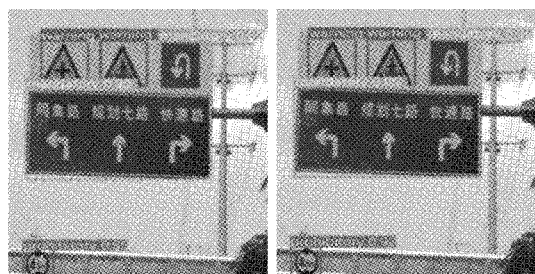
由表 2 可知,本文改进 YOLOv5 算法的平均识别准确

率达到了 96.40%，与 YOLOv5 算法相比提升了 6.83%，说明改进后的 YOLOv5 能更准确地对交通标志进行识别。

表 2 不同模型在测试集上的效能评估

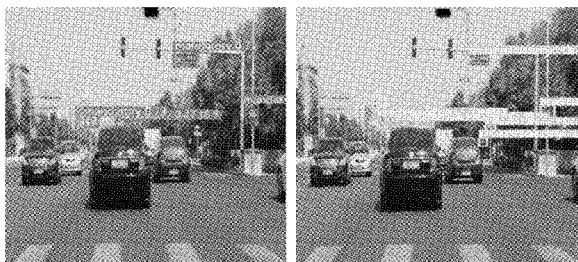
实验	CBAM 模块	DIoU Loss	P/ %	R/ %	mAP/ %
1			88.81	94.50	89.57
2	✓		92.40	86.30	95.16
3		✓	93.10	89.00	93.30
4	✓	✓	95.70	93.10	96.40

网络训练好之后需要利用测试集对其进行测试，图 8 和图 9(a)、(b)分别为简单和复杂场景下 YOLOv5 算法改进前后测试结果对比图，其中图(a)为改进前，图(b)为改进后。从中可以看出，无论是在交通标志较大的简单场景下，还是交通标志较小的复杂场景下，本文改进 YOLOv5 算法的识别准确率均优于原 YOLOv5 算法，说明改进后的 YOLOv5 算法能更好地获得交通标志的特征信息，获得更高的识别准确率。



(a) 原 YOLOv5 算法测试图 (b) 改进 YOLOv5 算法测试图

图 8 简单场景下改进前后测试结果对比图



(a) 原 YOLOv5 算法测试图 (b) 改进 YOLOv5 算法测试图

图 9 复杂场景下改进前后测试结果对比图

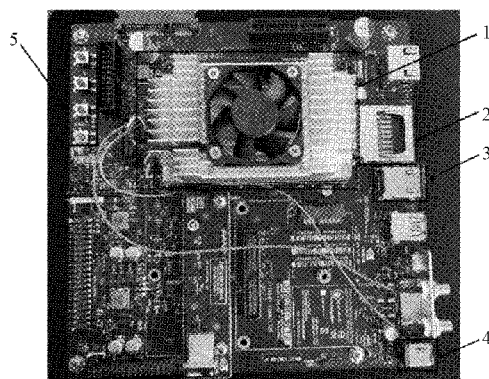
表 3 为相同环境下，本文改进 YOLOv5 算法与 SSD、YOLOv3、Improved YOLOv3、YOLOv5、SE-YOLOv5 五种经典模型的识别结果对比。从表中数据可知，由于 CBAM 的引入，网络的计算量有所增加，推理速度下降，因此本文改进 YOLOv5 算法识别速度与模型大小略低于嵌入 SE 注意力模块的 SE-YOLOv5 算法，但识别准确率为表中所列算法中最优，说明本文改进 YOLOv5 算法在保证模型大小和识别速度的前提下，有效地提高模型识别率。

表 3 不同算法目标检测结果对比

模型	mAP/%	FPS	模型大小
SSD ^[16]	75.68	15.65	100.23
YOLOv3 ^[17]	79.82	17.31	218.84
Improved-YOLOv3 ^[18]	83.36	21.81	198.57
YOLOv5	89.57	58.46	27.00
SE-YOLOv5	93.59	73.54	15.78
改进 YOLOv5 算法	96.40	66.92	20.34

4 嵌入式平台实验

为验证模型的实时有效性，将训练好的模型输送至 NVIDIA Jetson TX2 嵌入式系统中完成对实景视频中交通标志的识别，TX2 设备如图 10 所示。



1.内置CPU、GPU 2.HDMI屏幕接口
3.USB接口 4.电源开关 5.Reset按键等

图 10 TX2 设备

随后为使 TX2 能具备实验所需环境，利用 Ubuntu 系统的主机完成对 TX2 的刷机，刷机完毕的 TX2 实验环境配置为 Ubuntu18.04、Python3.7、CUDA10.0。最后在其中进行实验，识别结果如图 11 所示。

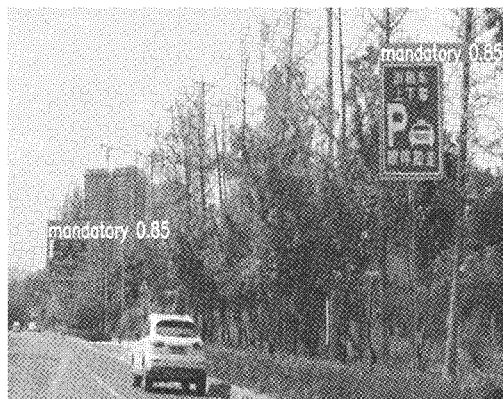


图 11 识别效果示意图

从图中可以看出，在 TX2 中的识别准确率与在主机中运行结果基本一致，FPS 达到 18 左右，说明本文改进算法基本能在嵌入式系统中完成对交通标志的识别。

5 结 论

本文提出一种改进 YOLOv5 算法以解决交通标志识别任务中小目标识别准确率较低的问题。引入 CBAM 使得网络特征提取能力得到增强。通过替换回归损失函数, 加快模型收敛速度。实验表明, 本文改进算法与原 YOLOv5 算法相比, 识别率有所提升。最后在 TX2 嵌入式系统应用本文改进算法完成对实景视频中交通标志的识别, 结果表明能流畅运行。

参考文献

- [1] 王文胜, 李继旺, 吴波, 等. 基于 YOLOv5 交通标志识别的智能车设计[J]. 国外电子测量技术, 2021, 40(10):158-164.
- [2] CIRESAN D, MEIER U, MASCI J, et al. Multi-column deep neural network for traffic sign classification[J]. Neural Networks, 2012, 32:338-338.
- [3] 包本刚. 融合多特征的目标检测与跟踪方法[J]. 电子测量与仪器学报, 2019, 33(9):93-99.
- [4] 常发亮, 黄翠, 刘成云, 等. 基于高斯颜色模型和 SVM 的交通标志检测[J]. 仪器仪表学报, 2014, 35(1):43-49.
- [5] JIANMING Z, MANTING H, XIAOKANG J, et al. A real-time chinese traffic sign detection algorithm based on modified YOLOv2 [J]. Algorithms, 2017, 10(4):127.
- [6] QIAN R, ZHANG B, YUE Y, et al. Robust Chinese traffic sign detection and recognition with deep convolutional neural network [C]. International Conference on Natural Computation, 2015:791-796.
- [7] 杨梦梦. 基于 Faster R-CNN 的交通标志检测方法[D]. 湖北:湖北工业大学, 2019.
- [8] 王卜, 何扬. 基于改进 YOLOv3 的交通标志检测[J]. 四川大学学报(自然科学版), 2022, 59(1):57-67.
- [9] 陈立潮, 张倩茹, 曹建芳, 等. 复杂场景下基于复合胶囊网络的交通标志识别[J]. 计算机工程与设计, 2021, 42(9):2627-2633.
- [10] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module[C]. Proceeding of the European Conference on Computer Vision(ECCV), 2018:3-19.
- [11] ZHENG Z, WANG P, LIU W, et al. Distance-IoU loss: Faster and better learning for bounding box regression[C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7):12993-13000.
- [12] REZATOFIGHI H, TSOI N, GWAK J Y, et al. Generalized intersection over union: A metric and a loss for bounding box regression[C]. CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2019:658-666.
- [13] LI H, XIONG P, AN J, et al. Pyramid attention network for semantic segmentation [J]. Computer Science, Computer Vision and Pattern Recognition(CVPR), IEEE, 2018, DOI:https://arXiv.org/abs/1805.10180.
- [14] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid network for objected detection[J]. Computer Science, Computer Vision and Pattern Recognition (CVPR), IEEE, 2017:936-944.
- [15] WANG C Y, LIAO H, WU Y H, et al. CSPNet: A new backbone that can enhance learning capability of CNN[C]. CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, 2020, 6:1571-1580.
- [16] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single Shot MultiBox Detector [J]. Computer Science, Computer Vision and Pattern Recognition (CVPR), IEEE, 2016:21-37.
- [17] REDMON J, FARHADI A. YOLOv3: An incremental improvement[J]. Computer Science, Computer Vision and Pattern Recognition(CVPR), IEEE, 2018:89-95.
- [18] 郭璠, 张泳祥, 唐璠, 等. YOLOv3- Λ : 基于注意力机制的交通标志检测网络[J]. 通信学报, 2021, 42(1):87-99.

作者简介

李宇琼, 硕士研究生, 主要研究方向为汽车自动驾驶技术、图像识别等。

E-mail: 1439245008@qq.com

周永军(通信作者), 硕士研究生导师, 工学硕士, 主要研究方向为汽车自动驾驶技术、图像识别等。

E-mail: zyj424@126.com