

无线电工程

Radio Engineering

ISSN 1003-3106, CN 13-1097/TN

《无线电工程》网络首发论文

题目: 基于多尺度 YOLOv5 的交通标志检测
作者: 朱宁可, 张树地, 王翰文, 李红松, 余鹏飞
网络首发日期: 2023-09-04
引用格式: 朱宁可, 张树地, 王翰文, 李红松, 余鹏飞. 基于多尺度 YOLOv5 的交通标志检测[J/OL]. 无线电工程. <https://link.cnki.net/urlid/13.1097.tn.20230901.2300.028>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式 (包括网络呈现版式) 排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊 (光盘版)》电子杂志社有限公司签约, 在《中国学术期刊 (网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊 (网络版)》是国家新闻出版广电总局批准的网络连续型出版物 (ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

基于多尺度 Yolov5 的交通标志检测

朱宁可, 张树地, 王翰文, 李红松, 余鹏飞*
(云南大学 信息学院, 云南 昆明 650504)

摘 要: 针对小目标交通标志检测存在的检测精度低、漏检率高等问题, 提出了一种基于多尺度融合的 Yolov5 改进算法。在主干网络后输出四个有效特征层以便更好的融合多尺度信息, 另外在主干网络输出的三个特征层中添加改进的多尺度融合注意力机制 CBAM_U, 以提升网络的检测能力, 接着在 PANet (Path Aggregation Network) 下采样过程中添加 Fusion 模块, 促进不同感受野下特征的细融合, 最后在 Yolohand 前加入 ASFF (Adaptively Spatial Feature Fusion) 模块解决特征金字塔融合的不一致性, 进一步提升网络的表达能力。实验结果表明, 提出的方法相比于原始 Yolov5 网络在 CCTSDB 数据集中 $mAP@0.5$ 提升了 3.07%, 召回率提升了 3.83%, 查准率提升了 1.64%, $F1-score$ 提升了 2.66%, 相比于其他检测算法, 改进后的 Yolov5 算法在复杂场景中具有更好的鲁棒性。

关键词: 交通标志检测; 小目标; 多尺度融合; CBAM_U; 细融合

中国法分类号: TP391.41 文献标识码: A 标识码(OSID)



Traffic sign detection based on multi-scale Yolov5

ZHU Ningke, ZHANG Shudi, WANG Hanwen, LI Hongsong, YU Pengfei*

(College of Information, Yunnan University, Kunming 650504, China)

Abstract: A Yolov5 algorithm based on multiscale fusion is proposed to address the problems of low detection accuracy and high missed detection rate of small target traffic sign detection. Four effective feature layers are outputted after the backbone network for better fusion of multi-scale information, and an improved multi-scale fusion attention mechanism CBAM_U is added to the three feature layers outputted from the backbone network to improve the detection capability of the network, followed by the addition of the Fusion module in the process of PANet (Path Aggregation Network) downsampling to promote the fine fusion of features under different perceptual fields, and finally, the ASFF (Adaptively Spatial Feature Fusion) module is added before Yolohand to solve the inconsistency of feature pyramid fusion and further enhance the expression ability of the network. The experimental results show that the proposed method improves the $mAP@0.5$ by 3.07%, the recall rate by 3.83%, accuracy by 1.64%, and $F1-score$ by 2.66% over the original Yolov5 network on CCTSDB dataset, and the improved Yolov5 algorithm has better robustness in complex scenes compared with other detection algorithms.

Keywords: traffic sign detection; small target; multi-scale fusion; CBAM_U; fine fusion

基金项目: 国家自然科学基金 (62066046)

Foundation Item: National Natural Science Foundation of China (62066046)

0 引言

随着现代网络技术以及经济的发展, 汽车的使用方便大众的生活, 自动驾驶以及智能辅助驾驶系统逐渐进入商业领域。交通标志的检测对于行车安全至关重要, 但是交通标志常常会受到天气、光照、遮挡等因素的影响, 这给无人驾驶的应用带来了很大的安全挑战, 因此设计出一种即实时又准确的交通标志检测智能系统至关重要。

交通标志检测是智慧交通系统 (ITS) 以及无人驾驶系统的基础。在过去的几十年中, 目标检测的所采用的方法较为传统, 通过采用手工设计特征加浅层分类器来进行目标的获取以及分类。例如: 朱双东等人^[1]通过将 RGB 彩色交通标志图片转化到 HIS 颜色空间以降低光照对图片的影响, 但此过程仍具有一定的计算量; Gavrilin 等人^[2]通过方向梯度直方图 (Histogram of Oriented Gradients, HOG) 进行模板匹配的形狀检测; 另外汤凯等人^[3]提出一种

颜色特征、形状特征和尺度特征的多特征协同方法,采用支持向量机 (Support Vector Machine, SVM) 对融合的特征进行分类来获取预测结果;甘露等人^[4]使用 Gabor 滤波来提取交通标志的纹理特征以及形状特征,最后将纹理特征以及形状特征融合输入到 BP 神经网络进行分类,有效的提升了精度。时至今日,交通标志检测的主流方法是通过深度学习进行交通标志的分类以及预测位置。由于基于候选框的深度学习目标检测算法精度较高,被研究人员广泛关注,其中,两阶段目标检测算法代表性的是由 Girshick 等人^[5]提出的一种基于候选区域的目标检测算法 R-CNN (Regions with CNN features),该网络通过深度卷积神经网络 (Convolutional Neural Network, CNN) 对目标区域进行分类取得不错的效果,但是该网络过多的重复选取候选框造成大量存储资源以及计算资源的浪费。Girshick 等人提出的 Fast R-CNN^[6]以及 Ren Shaoqing 等人提出的 Faster R-CNN^[7]在 R-CNN 的基础上进一步将候选窗口的生成和候选窗口的分类与回归统一到一个网络中联合学习,大大提升了网络的检测速度。另外,一阶段代表性的目标检测算法主要包括 SSD^[8]算法以及 YOLO 系列算法,其中 SSD 算法采用不同特征层来预测不同大小的物体,YOLO 则是将图像分成 $N \times N$ 大小的子区域,并预测每个区域是否存在目标物体以及目标物体的类别、概率和位置。一阶段目标检测算法速度快,可以满足实时检测的要求,但是其精度不如两阶段目标检测算法。因此,本文选择 YOLO 系列进行改进,以提升交通标志检测的准确率。

YOLOv5 是目前比较主流的一种单级目标检测器,具有优秀的性能以及检测速度,同时还具有清晰的结构来进行分解和调整^[9]。虽然 YOLOv5 是一个强大的目标检测模型,但是其本身被设置为一种通用模型,对于交通标志这种小目标物体检测效果并不会过于显著^[10]。因此本文提出了一种多尺度的 YOLOv5 检测模型来融合更多尺度的信息,以此更好的提升检测效果。本文的主要贡献有:

(1) 将 YOLOv5 原本的三层特征融合改进为四层特征融合,以便于更好的保留底层信息 (包括位置信息、纹理信息等)。

(2) 为了使网络更好的关注小目标,本文对 CBAM 注意力机制进行改进,在空间注意力中引入 U 型结构以及残差结构以提升 CBAM 对小目标的敏感度。

(3) 将 Fusion 模块嵌入到 PANet 下采样过程中,促进底层信息的保留与融合

(4) 将 ASFF 模块嵌入到三个检测头前,以便于获取更加丰富的特征进行分类以及位置预测。

1 YOLOv5 模型结构

1.1 网络结构分析

YOLOv5 的结构是在 YOLOv4 的基础上进行改进的,其中改进的部分主要是在主干部分使用了 Focus 网络结构,使图片在下采样的过程中,在信息不丢失的情况下,将 W、H 的信息集中到通道上,再使用 3×3 的卷积对其进行特征提取,使得特征提取得更加充分。另外在主干部分以及特征加强网络中也加入了 CSPlayer,以便于更好的提取特征。中间部分是特征加强网络,所采用的结构是 PANet^[11]。最后是检测头部分,用于检测不同大小的目标物体。

1.2 Yolov5 不足之处

首先,Yolov5 是一种通用的一阶段目标检测模型,如果直接应用于自动驾驶领域存在着对小目标检测能力不足的问题,尤其是在目标存在模糊、密集以及遮挡的情况下。其次,Yolov5 中的特征增强模块 PANet 侧重于深度特征图的融合,而忽略了底层的细节特征,削弱了对小目标的检测能力,对于小目标的交通标志来说会产生大量的特征冗余,造成计算资源的浪费。最后,小目标物体在经过多次卷积之后会存在部分细节信息和位置信息消失的问题^[12]。因此本文对 Yolov5 模型进行改进,融合多尺度信息来加强网络对交通标志目标的检测能力。

2 模型改进

根据以上存在的不足之处,本文分别对特征加强网络、注意力机制进行改进。首先,从主干网络中多引出一层改进特征层,将引出的特征继续与加强特征网络融合,使得网络更多的融合底层的位置信息。其次,底层的特征中也存在着较多的噪声并且对于目标物体的聚焦程度不足,因此在主干网络输出到特征加强网络的特征中添加改进的注意力机制 CBAM_U,能够有效的融合底层的语义信息并过滤掉大部分噪声,使得网络对于目标物体更加敏感。然后在网络中添加细融合 Fusion 模块,促进自身特征之间的细融合。最后在特征加强网络后加入 ASFF 消除融合后特征之间的冲突。改进的 Yolov5 模型如图 1 所示。

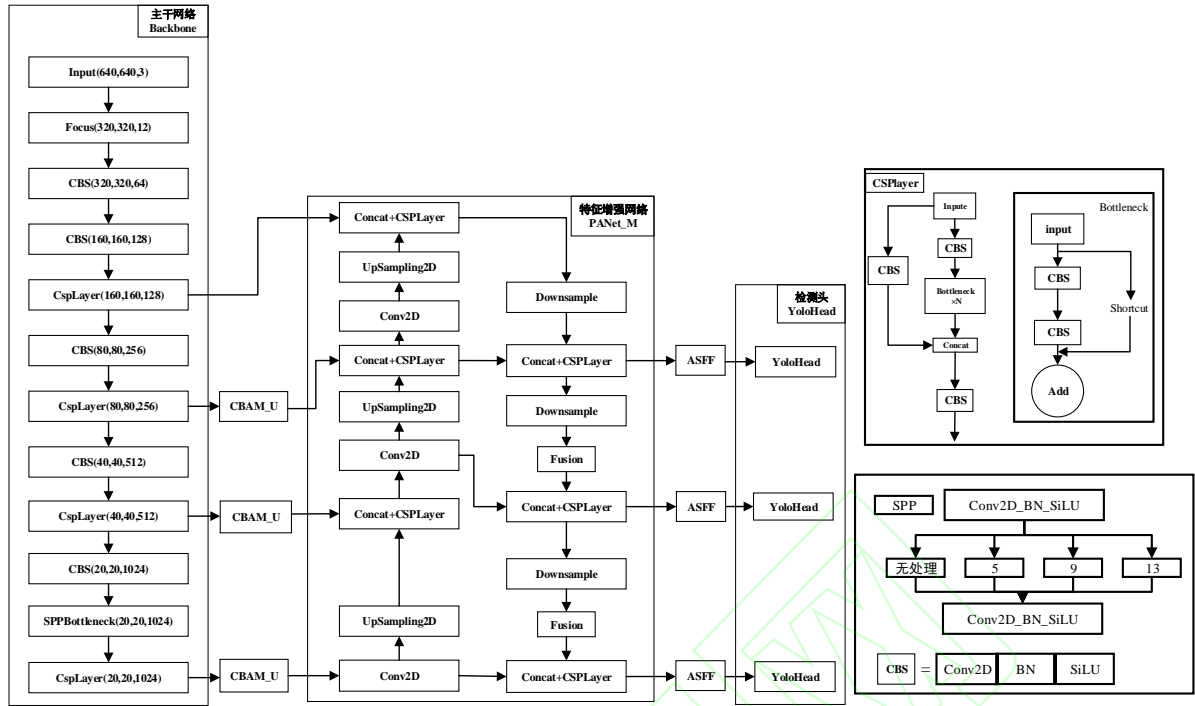


图1 改进 YOLOv5 结构

Fig. 1 The structure of Improved YOLOv5

2.1 特征加强网络（PANet）改进

在自动驾驶的真实场景图片中，交通标志的像素仅占很小一部分，另外不同的交通标志有着尺度不一的特点，想要更加精确地识别交通标志，低层次的语义信息显得格外重要。YOLOv5 与 YOLOv4 所使用的 PANet 结构相似，均是采用双向融合的方式，即由下向上融合以及由上向下融合。在 YOLOv4 的 PANet 结构中，每次进行 concat 拼接之后会经过五次卷积来调整通道数，加深网络，以此来增强特征的提取能力。而在 YOLOv5 中使用的是 C3 模块，对特征进行反复的提取，但在 C3 模块中，先将特征的通道数进行降低后再进行卷积，所以产生的参数量更少。在 YOLOv5 的 PANet 中，由主干网络传入三个有效特征层，层级之间经过尺度变换进行融合，使得卷积网络的高级特征包含对象的全局语义信息，这有助于区分不同的目标对象，具体结构如图 2 所示。

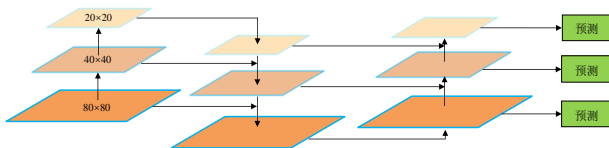


图2 原始 PANet 结构

Fig. 2 The structure of Primitive PANet

然而随着卷积层的加深，卷积网络对于目标物体位置的变化不再敏感，导致高层次特征在物体的位置定位方面显得不足^[13]。在低层次的特征中存在较多的位置空间信息，为了融合全局语义信息和空间位置信息，本文提出一种改进的 PANet，命名为

PANet_M，在 YOLOv5 的 PANet 基础上额外增加一层融合，将高层次语义信息与低层次特征相结合，以获取更加丰富的信息和表达准确的特征图。

在 YOLOv5 中，颈部特征加强网络共设定 3 个检测尺度，分别是 80×80 、 40×40 、 20×20 ，其中 80×80 的尺度用来检测小目标， 40×40 与 20×20 的尺度分别用来检测中目标以及大目标，具体结构如图 2 所示。本文从 YOLOv5 的主干网络中额外引出一层 160×160 的特征层与 80×80 的特征层进行融合，加强低层的语义信息，使每个特征层的语义信息更加丰富。具体改进结构如图 3 所示。

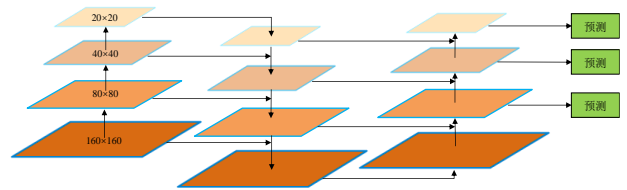


图3 PANet_M 结构

Fig. 3 The structure of PANet_M

2.2 CBAM 注意力机制改进

注意力机制能有效地将计算资源分配给更重要的任务，同时解决信息超载问题。CBAM 注意力是一种结合了空间（spatial）和通道（channel）的注意力机制模块^[14]。CBAM 中通道注意力机制聚焦于“什么”是有意义的输入图像，使用平均池化操作对输入特征图的空间维度进行压缩，来聚合特征图的空间信息并生成平均池化特征，与此同时使用最大池化聚焦于独特的物体特征并生成最大池化特征，所生成的两个特征被转发到一个共享的网络来

产生通道注意力映射，最后共享网络应用于每个描述符后使用Element-Wise求和的方式来合并并输出特征向量，最后通过sigmoid函数进行标准化得到通道特征权值。空间注意力模块聚焦在“哪里”是最具信息量的部分，首先将通道注意力模块产生的特征图经过最大池化与平均池化得到两个 $H \times W \times 1$ 的特征图，其次这两个特征图基于通道做concat拼接操作聚合成一个特征图的通道信息，再者将这些信息通过一个标准卷积产生2D空间注意力，最后通过sigmoid进行标准化得到最终注意力特征图，CBAM结构如图4和图5所示。

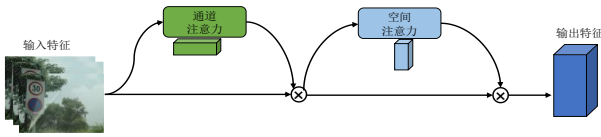


图4 CBAM 模块

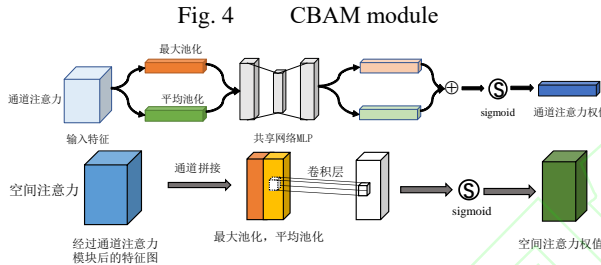


图5 通道注意力模块与空间注意力模块

Fig. 5 Channel attention module and spatial attention module

本文中使用CBAM注意力机制并对其进行改进，使其更适应小目标交通标志的特性。通过对交通真实场景进行分析，可以发现复杂道路背景的干扰信息是交通标志定位和识别过程的一大障碍^[15]。在目标检测的过程中，神经网络通常会滑动图片的所有区域进行特征提取，而不区分对象区域和非对象区域，因此没有特别的模块突出对象区域的重要性^[16]。通过以上分析，本文改进了CBAM中的空间注意力模块来解决上述问题，命名为CBAM_U，改进结构如图6、图7所示。

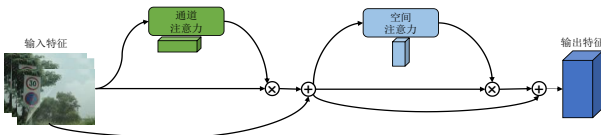


图6 CBAM_U 模块

Fig. 6 The module of CBAM_U

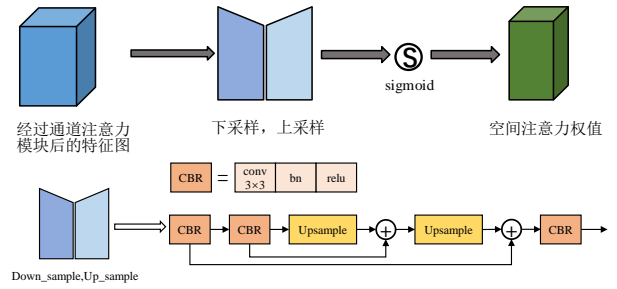


图7 改进空间注意力模块

Fig. 7 Improved spatial attention module

空间注意力机制主要模拟人眼的视觉神经，聚焦于特征图中重要的目标区域并过滤掉特征图中的干扰区域。受PANet以及U-Net^[17]的启发，U型结构能够有效的突出目标物体并且过滤掉部分干扰信息，因此本文在空间注意力模块中引入U型结构来探索每个特征图的空间信息，通过两次卷积快速获取高层次语义信息。高层次特征图含有高级语义信息和少量空间位置信息，通过U型机制，低层次的语义信息能够通过跳跃连接与高层次语义信息进行融合，将底层的位置信息发送给高层。同时为了在不破坏原始特征图的情况下使用与残差网络^[18]相似的结构，将高级特征与低级特征合并，并逐元素相加融合。最后使用sigmoid函数增加网络的非线性，输出特征权值。另外，在通道注意力机制中加入残差连接确保每次聚焦特征层的有效性。

2.3 Fusion 模块

在特征加强网络中，由于Yolov5在特征融合中使用了多个上采样进行多层次的融合，即不同层级之间实现粗融合，使得一些关键性的细节信息丢失。因此为了解决此问题在更细粒度的级别上合并多级特征，本文引入了一种Fusion模块来获取更精细的特征，具体结构如图8所示。本文提出的Fusion模块借鉴了RFB^[19]的Inception结构的思想，输入的特征通过两个不同的分支并行，其中一个分支通过一个 1×1 卷积获取原始特征，相当于保险丝的作用；另一个分支增加额外的 3×3 卷积来增加不同的感受野，不同的两个卷积核产生具有不同感受野的两个特征，相当于产生两个不同的尺度特征。简单的将这两个特征通过逐元素相加的方式实现精细的局部多尺度融合，使得自身某些细小的特征得到加强，即实现下采样特征的细融合。

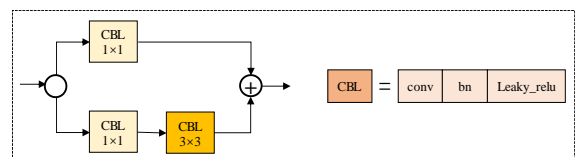


图8 Fusion 模块

Fig. 8 Fusion module

2.4 ASFF 模块

在Yolov5的特征金字塔中，多尺度融合是解决多尺度目标检测的有效方法，但是不同层级之间的特征存在一定的信息冲突，即不一致性，不同特征尺度之间的不一致性是基于特征金字塔的单阶段检测器的主要限制。为了抑制这种不一致性，Wang^[20]与Zhu^[21]等人将相邻级别的相应区域设置为忽略区域，但相邻级别的松弛往往会导致更差的预测为误报。本文引入ASFF模块^[22]来抑制不同层级之间的特征的信息冲突，提高PANet的融合效果，进而提高目标检测的能力，具体结构图如图9所示。

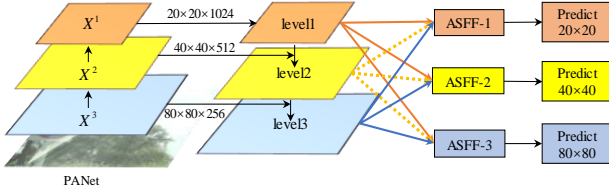


图9 ASFF 模块

Fig. 9 ASFF module

图9中左侧由PANet输出三个有效特征层分别为 20×20 、 40×40 以及 80×80 三种不同的尺度。以ASFF-1为例，特征图 X^2 通过一个 3×3 ， $\text{stride}=2$ 的卷积，特征图 X^3 则需要在 3×3 ， $\text{stride}=2$ 的卷积的基础上再添加一个 $\text{stride}=2$ 的Maxpooling，这样就能使得特征图 X^2 和 X^3 都获得与特征图 X^1 相同的维度，分别记为 X^{2-1} 和 X^{3-1} ；定义 $X_{i,j}^{n-1}$ 为特征图 X^{n-1} 位置 (i,j) 处的特征向量，则ASFF-1输出特征图 Y^1 位置 (i,j) 处计算公式如公式(1)所示：

$$Y_{i,j}^1 = \alpha_{i,j}^1 \times X_{i,j}^{1-1} + \beta_{i,j}^1 \times X_{i,j}^{2-1} + \gamma_{i,j}^1 \times X_{i,j}^{3-1} \quad (1)$$

公式(1)中 $\alpha_{i,j}^1$ 、 $\beta_{i,j}^1$ 、 $\gamma_{i,j}^1$ 分别表示三个不同层级之间到特征图 X^1 的重要性权重，定义 $\alpha_{i,j}^1$ 的计算公式如(2)所示：

$$\alpha_{i,j}^1 = \frac{e^{\lambda_{\alpha_{i,j}}^1}}{e^{\lambda_{\alpha_{i,j}}^1} + e^{\lambda_{\beta_{i,j}}^1} + e^{\lambda_{\gamma_{i,j}}^1}} \quad (2)$$

即通过不同层级之间自适应学习各尺度特征图融合的空间权重来解决不同层级之间存在的信息冲突。

3 实验与结果分析

3.1 数据处理

本文所使用的是长沙理工大学制作的2021版CCTSDB中国交通标志数据集^[23]，包含1.2万张图片以及对应的标注信息。本文的交通标志主要分为3类，分别警告类(warning)、禁止类(prohibitory)以

及指示类(mandatory)。为了提高模型的泛化能力以及鲁棒性，使用了Yolov5中的Mosaic^[24]数据增强。

Mosaic数据增强简单来说就是将4张图片通过随机缩放、随机裁剪、随机排布的方式进行拼接，该数据增强的方法能够丰富检测物体的背景以及小目标，极大的丰富了数据集，使网络更具有鲁棒性。

3.2 实验环境及评价指标

本文采用pytorch深度学习框架，Windows 10操作系统以及GeForce RTX 3090显卡进行实验。使用随机梯度下降法(SGD)来优化调整网络参数，初始学习率为0.01，权重衰减系数为0.0005，还使用了动态调整学习率进行训练，其中动态调整学习率的方式为cos，另外还使用了混合精度以节省显存资源。

本文所采用的评价指标为平均准确率($mAP@0.5$)、召回率(Recall)、查准率(Precision)以及F1-score。将测试结果划分为四大类，分别为真正样本(True Positive, TP)、假正样本(False Positive, FP)、真负样本(True Negative, TN)、假负样本(False Negative, FN)，具体方式如表1混淆矩阵。

表1 混淆矩阵

Tab. 1 Confusion matrix

混淆矩阵 (Confusion Matrix)		真实值	
		正样本	负样本
预测值	正样本	TP	FN
	负样本	FP	TN

查准率(Precision)是指所有正样本中检测结果为正样本所占的比例，能够衡量模型的错检程度，具体见公式(3)：

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

召回率(Recall)是指所有正样本中正确识别的概率，能够很好的衡量模型的漏检程度，具体见公式(4)：

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

平均准确率 $mAP@0.5$ 的值与AP有关，AP表示当混淆矩阵的IOU置信度取0.5时，针对目标检测中的某一类别具有x个正样本预测框，按照置信度进行降序排列，每一个预测框对应一个查准率Precision(P_i)，将所有的查准率求和取平均即可得到该类别AP的值，计算见公式(5)：

$$AP @ 0.5 = \frac{1}{x} \sum_{i=1}^x P_i \quad (5)$$

平均准确率 $mAP@0.5$ 定义为检测目标中所有类的 AP 值取平均, $mAP@0.5$ 衡量一个模型对于每一类的查准率随召回率的变化, 能更直观的表现出分类器的性能, $mAP@0.5$ 的计算见公式 (6):

$$mAP@0.5 = \frac{1}{n} \sum_{j=1}^n AP@0.5_j \quad (6)$$

$F1-score$ 是查准率与召回率的调和平均数, 能够很好的评价检测算法的优劣。由于交通标志识别关系到安全方面的问题, 需要较高的查准率与召回率, 本文将两者的权重设置为相同, 具体见公式 (7):

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (7)$$

3.3 实验分析

3.3.1 消融实验与分析

为了验证本文改进模块的效果进行消融实验, 消融实验所得出的结果如下表2所示。

由表2消融实验可得, 在CBAM通道注意力中加入残差连接相比于不加在 $mAP@0.5$ 上提升了0.52%, 确保了每次聚焦特征的有效性, 通过将不同尺度的特征层分别进行粗融合与细融合, 本文所提出的各个模块在原始模型中都有较大的提升。另外本文提出的改进Yolov5与原模型相比在 $mAP@0.5$ 上提升3.07%, 召回率提升3.83%, 做到交通标志检测的漏检程度更低, 通过引入改进的注意力机制使

得网络能更好的聚焦于目标物体, 使得查准率提升了1.64%, 准确率得到较大提升, 另外在 $F1-score$ 中, 提升了2.66%, 做到了召回率与准确率的平衡, 检测算法更加优秀, 具有良好的应用价值。

3.3.2 交通标志检测模型对比

从表3可以看出, 本文改进的Yolov5模型相比于SSD、FCOS、Yolov4、Centernet检测模型, $mAP@0.5$ 分别提升了4.92%、6.17%、2.59%、1.55%, 另外 $F1-score$ 分别提高了22.99%、9.33%、2.66%、0.53%。相比于较新的目标检测模型Yolov5-6.0、文献25、文献26, $mAP@0.5$ 分别高出0.7%、8.54%与0.42%, 召回率与查准率也都有部分提升, 由于模型参数数量的增加, 所以在检测速度方面存在少许下降, 但是仍能满足实时性的要求。可见通过将多尺度的不同信息进行融合能够使网络捕捉到更多的细节信息, 提升不同尺寸检测下的精度, 改进的Yolov5模型相比于这些模型更能满足在真实驾驶场景中高准确率的需求。

由表4可以看出, 在不同尺寸的检测中, 本文的算法相比于原始网络在 AP_s , AP_m 以及 AP_l 中分别提升19.5%、7.2%以及2.7%, 另外召回率分别提升18.1%、7.2%以及4.4%, 做到了精确度以及召回率的同时提升, 基本解决了小目标检测难、漏检率高的问题。

表2 消融实验

Tab. 2 Ablation experiments

模型	$mAP@0.5/\%$	$Recall/\%$	$Precision/\%$	$F1-score/\%$
Yolov5m	89.82	82.81	93.88	88.00
Yolov5m+cbam_nonres	91.81	84.97	95.62	89.67
Yolov5m+cbam_u	92.33	85.80	95.39	90.33
Yolov5m+ASFF	90.31	83.64	93.29	88.00
Yolov5m+PANet_M	91.35	82.33	95.29	88.67
Yolov5m+Fusion	92.17	86.29	94.05	90.00
Yolov5m+all	92.89	86.64	95.52	90.66

表3 交通标志检测模型对比

Tab. 3 Comparison of traffic sign detection models

模型	$mAP@0.5/\%$	$Recall/\%$	$Precision/\%$	$F1-score/\%$	FPS
Yolov5m	89.82	82.81	93.88	88.00	44.19
SSD	87.97	52.25	95.31	67.67	58.68
FCOS	86.72	70.59	96.33	81.33	38.13
Yolov4	90.30	83.45	92.69	88.0	36.82
Centernet	91.34	86.33	95.0	90.13	38.69
Yolov5m-6.0	92.19	88.55	95.52	90.0	46.67
文献25	84.35	84.71	85.92	/	/
文献26	92.47	/	/	/	20.0
Ours	92.89	86.64	95.52	90.66	35.09

表4 不同尺寸指标对比

Tab. 4 Comparison of different size indicators

模型	指标	Small	Medium	Large
Yolov5	<i>AP</i>	0.247	0.616	0.764
	<i>Recall</i>	0.321	0.676	0.807
Ours	<i>AP</i>	0.442	0.688	0.791
	<i>Recall</i>	0.502	0.748	0.851

3.3.3 图片检测结果与分析

在本节中随机抽取网络搜集的不同目标尺寸

的交通标志图片进行可视化比较,如图10所示,从左到右依次是Yolov5检测结果以及改进Yolov5检测结果,将目标物体分为小、中、大三种不同的尺寸进行检测。从图10可以看出,在小目标检测中,原始Yolov5检测算法存在漏检的情况,而本文中改进的算法对于小目标较为敏感。另外在中目标检测中,原始Yolov5算法存在误检的情况,将背景物体检测为交通标志,这对无人驾驶领域是不安全的。在大目标检测中,本文的检测算法精度更高。综上,改进的Yolov5算法在交通标志检测中具有更好的性能。



图10 CCTSDB 检测结果可视化

Fig. 10 Visualization of CCTSDB detection results

通过模拟不同场景下的检测来检验网络的合理性。本文在网络搜集的交通标志图片中模拟了雾天、雨天以及夜间的场景分别进行检测。通过图11雾天模拟场景可以看出,改进后的Yolov5能够检测出大部分的交通标志,而原始网络仅能检测处一个,并且存在误检的情况。雨天场景中两种网络均能检测出所有交通标志,但是改进的Yolov5算法的置信度更高,受噪声影响较小。最后在夜间检测的情况下本文的算法没有出现误检的情况,而原始网络将指示类标志(mandatory)检测为禁止类(prohibitory),这对于真实的应用场景来说是很不恰当的。

另外,本文通过热力图Grad-CAM[27]来解释改进的网络对于不同尺寸目标物体所带来的提升,通过图12能直观的观察到在小目标检测中,原始网络对于小目标的注意是不够的,在图12(a)的检测中存在多个目标并且小目标居多,原始网络在检测中存在漏检的问题,较小的警告类标志未检测出来。另外在较远距离的小目标检测图12(b)中对于目标物体的敏感程度是不够的,本文所提出的网络能更精确地定位到小目标交通标志,并且不存在漏检的情况。

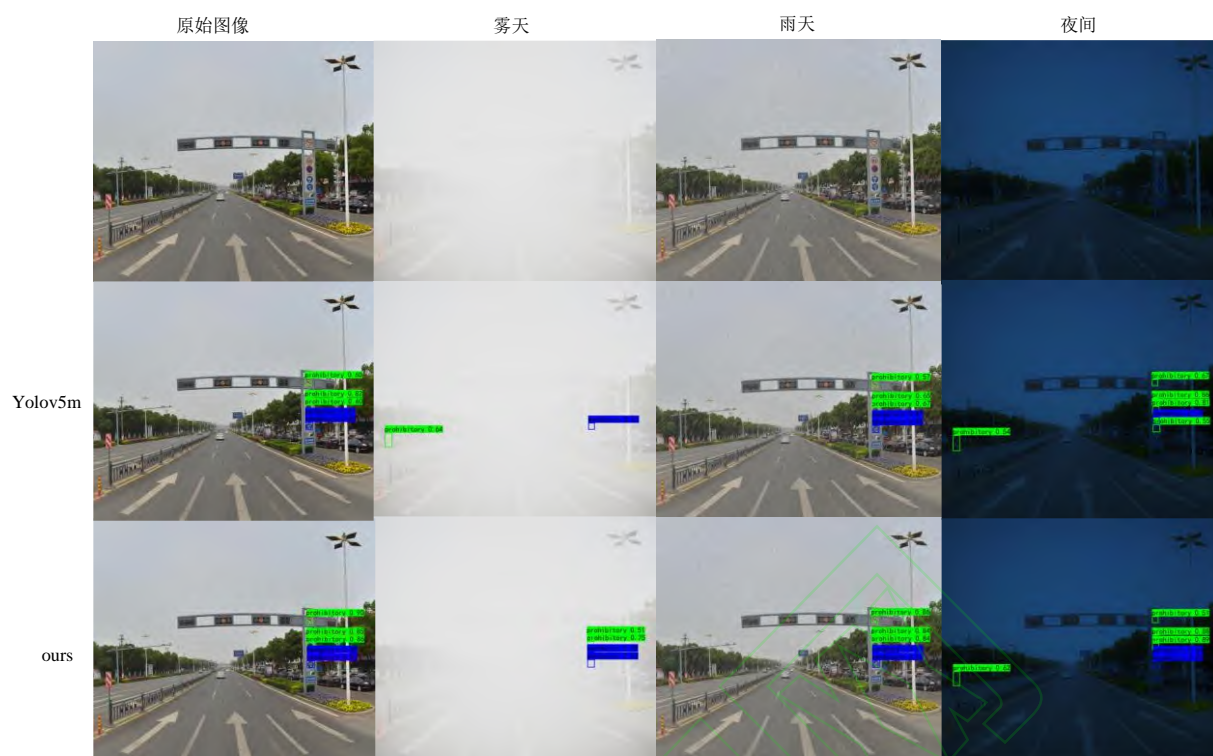


图11 不同场景检测结果可视化

Fig. 11 Visualization of detection results in different scenarios

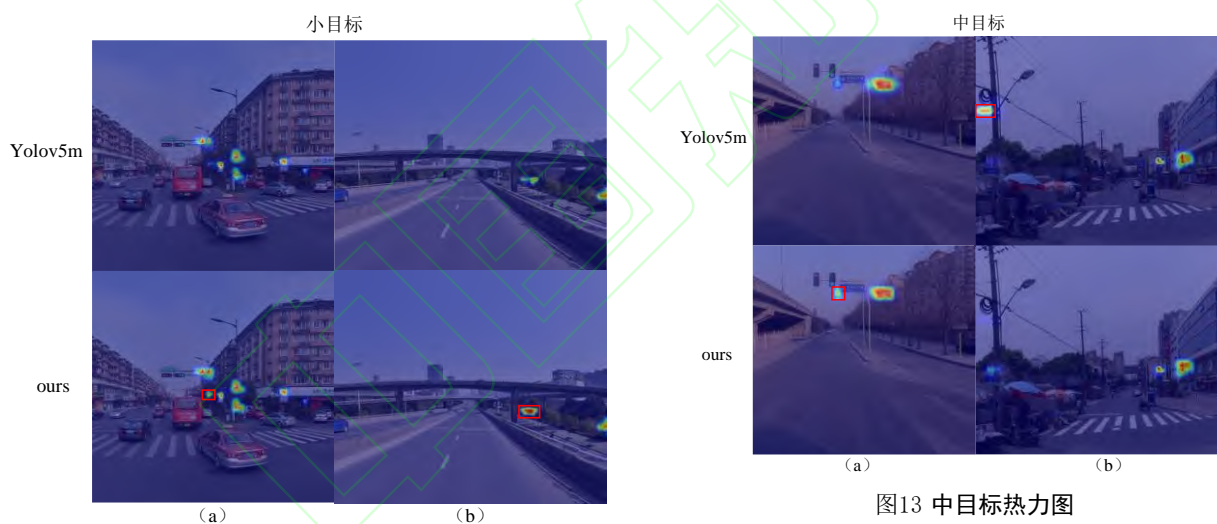


图12 小目标热力图

Fig. 12 Small target heatmap

对于中目标，通过图13的（a）能够看出，本文的网络能够很好的注意到目标区域，而原始网络受到过多的背景信息干扰使得指示类标志敏感度降低。在图13（b）的检测中，原始网络将背景信息的一卷线缆误识别为目标物体，这对交通标志检测领域来说是很不恰当的。

图13 中目标热力图

Fig. 13 Medium target heatmap

对于大目标物体来说，尺寸较大，相对较好检测，但是在背景信息比较复杂的情况下原始网络的目标聚焦程度可能会受到很大的影响。图14（a）中的背景较为复杂，原始网络很容易受到干扰，而本文的网络通过注意力以及多尺度融合使网络对于目标物体的聚焦程度增加，性能表现更优。图14（b）中的检测则是背景信息与目标信息较为相似，原始网络将目标物体误认为背景信息，而改进的Yolov5算法则表现更佳，能够注意到所有的目标物体。

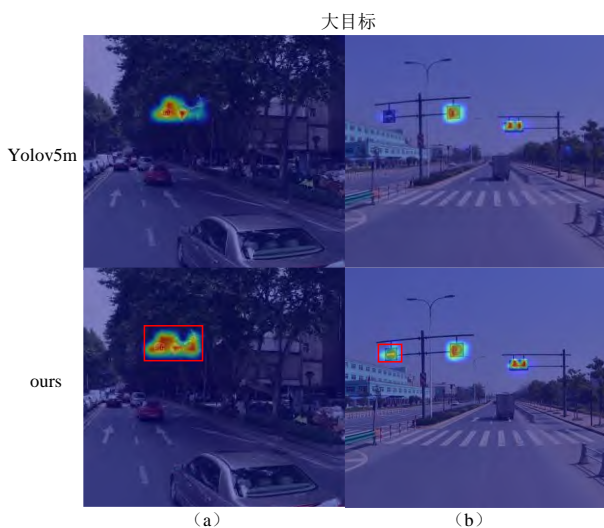


图14 大目标热力图

Fig. 14 Large target heatmap

4 结论

本文为了解决交通标志检测中检测难、精确度低、漏检率高等问题，提出了一种改进的Yolov5算法。通过融合多层的主干特征解决对目标物体位置信息不足的问题，再将输入的特征层通过改进的注意力机制滤除多余的噪声。另外在特征金字塔中加入Fusion结构促进特征之间的细融合，最后使用ASFF结构对输出的三个特征进行一致性融合，提升网络的检测能力。改进的Yolov5模型在小、中、大三种目标尺寸的检测中 $mAP@0.5$ 、查准率、召回率均有很大提升。所以本文提出的改进算法在自动驾驶辅助系统以及无人驾驶技术中有着较高的实用价值。同时本文所提出的模型也存在一定的局限性，由于CCTSDB仅存在3种类别，模型可检测的种类可以做进一步拓展。针对自动驾驶的应用还可以将改进的算法进一步优化，以方便移植到算力有限的移动设备中。这两点将会是接下来的研究目标。

参考文献

- [1] 朱双东, 刘兰兰, 陆晓峰. 一种用于道路交通标志识别的颜色-几何模型[J]. 仪器仪表学报, 2007, 28(5): 956-960.
- [2] Gavrila D M. Traffic sign recognition revisited[M]//Mustererkennung 1999. Springer, Berlin, Heidelberg, 1999: 86-93.
- [3] 汤凯, 李实英, 刘娟, 等. 基于多特征协同的交通标志检测[J]. 计算机工程, 2015, 41(3): 211-217.
- [4] 甘露, 田丽华, 李晨. 基于融合特征和 BP 网络的交通标志识别方法[J]. 计算机工程与设计, 2017, 38(10): 2783-2787.

- [5] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 580-587.
- [6] Girshick R. Fast r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.
- [7] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Trans Pattern Anal Mach Intell, 2017,39(6): 1137-1149.
- [8] Liu W, Anguelov D, Erhan D, et al. SSD: Single Shot Multibox Detector[C]//European conference on computer vision. Springer, Cham, 2016: 21-37.
- [9] Benjumea A, Teeti I, Cuzzolin F, et al. YOLO-Z: Improving small object detection in YOLOv5 for autonomous vehicles[J]. arXiv preprint arXiv:2112.11798, 2021.
- [10] Fan B B, Yang H. Multi-scale traffic sign detection model with attention[J]. Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, 2021, 235(2-3): 708-720.
- [11] Liu S, Qi L, Qin H, et al. Path aggregation network for instance segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 8759-8768.
- [12] Xiao J. exYOLO: A small object detector based on YOLOv3 Object Detector[J]. Procedia Computer Science, 2021, 188: 18-25.
- [13] Zhang J, Zhang L, Liu T, et al. YOLSO: You Only Look Small Object[J]. Journal of Visual Communication and Image Representation, 2021, 81: 103348.
- [14] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.
- [15] Wang J, Chen Y, Gao M, et al. Improved YOLOv5 network for real-time multi-scale traffic sign detection[J]. arXiv preprint arXiv:2112.08782, 2021.
- [16] Yao Y, Han L, Du C, et al. Traffic sign detection algorithm based on improved YOLOv4-Tiny[J]. Signal Processing: Image Communication, 2022, 107: 116783.
- [17] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015: 234-241.

- [18] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
- [19] Liu S, Huang D. Receptive field block net for accurate and fast object detection[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 385-400.
- [20] Wang J, Chen K, Yang S, et al. Region proposal by guided anchoring[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 2965-2974.
- [21] Zhu C, He Y, Savvides M. Feature selective anchor-free module for single-shot object detection[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 840-849.
- [22] Liu S, Huang D, Wang Y. Learning spatial fusion for single-shot object detection[J]. arXiv preprint arXiv:1911.09516, 2019.
- [23] Zhang J M, Zou X, Kuang L D, et al. CCTSDB 2021: A More Comprehensive Traffic Sign Detection Benchmark[J]. HUMAN-CENTRIC COMPUTING AND INFORMATION SCIENCES, 2022,12.
- [24] Yun S, Han D, Oh S J, et al. Cutmix: Regularization strategy to train strong classifiers with localizable features[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 6023-6032.
- [25] 吕禾丰, 陆华才. 基于YOLOv5算法的交通标志识别技术研究[J]. 电子测量与仪器学报, 2021,35(10):137-144.
- [26] 田智慧, 孙盐盐, 魏海涛. 基于SSD模型的交通标志检测算法[J]. 计算机应用与软件, 2021,38(12):201-206.
- [27] Selvaraju R R, Cogswell M, Das A, et al. Grad-cam: Visual explanations from deep networks via gradient-based localization[C]//Proceedings of the IEEE international conference on computer vision. 2017: 618-626.

作者简介:

朱宁可(2000-), 男, 河南驻马店人, 在读硕士研究生, 研究方向为深度学习、图像处理, 目标检测;

张树地(1997-), 男, 吉林白城人, 在读硕士研究生, 研究方向为语义分割、医学图像处理;

王翰文(1998-), 男, 云南曲靖人, 在读硕士研究生, 研究方向为深度学习、图像处理, 目标检测;

李红松(1974-), 男, 云南建水人, 博士, 副教授, 硕士研究生导师, 研究方向为深度学习、图像处理;

通信作者: 余鹏飞(1974-), 男, 云南贡山人, 博士, 副教授, 硕士研究生导师, 研究方向为深度学习、图像处理。