计算机工程
Computer Engineering
ISSN 1000-3428,CN 31-1289/TP

《计算机工程》网络首发论文

题目：结合轻量化与多尺度融合的交通标志检测算法
作者：兰红，王惠钊
DOI：10.19678/j.issn.1000-3428.006768269
网络首发日期：2024-03-06
引用格式：兰红，王惠钊．结合轻量化与多尺度融合的交通标志检测算法[J/OL]．计算机工程. https://doi.org/10.19678/j.issn.1000-3428.006768269

网络首发时间：2024-03-06 11:44:40
网络首发地址：https://link.cnki.net/urlid/31.1289.TP.20240304.1100.003

计 算 机 工 程
Computer Engineering

本文源代码链接：https://github.com/Wanguizhao/Multiple-YOLO.git

结合轻量化与多尺度融合的交通标志检测算法

兰红 王惠钊

江西理工大学信息工程学院 赣州 341000

摘要：交通标志检测在自动驾驶领域具有重要的应用价值，及时准确地检测交通目标对提高驾驶安全性和预防交通事故有重要意义。针对交通标志尺寸小，易受遮挡，在复杂环境下容易出现漏检、错检等问题，在 YOLOv8 的结构基础上提出一种结合轻量化与多尺度融合的交通标志检测网络架构 M-YOLO(Multiple-YOLO)，构建 M-YOLOs(Multiple-YOLO small)模型来应对高精度需求的检测任务，并调整网络深度得到更加轻量化的 M-YOLOn(Multiple-YOLO Nano)模型来解决不同环境下的检测需求。首先，针对交通标志目标尺寸小、图像特征流失的问题，通过增加小目标检测层，保留更多的特征信息，提高网络对于小目标的特征学习能力。提出高效多尺度特征金字塔融合网络 MPANet(Multiple Path Aggregation Network)，通过将浅层特征图进行降维，并进行跳跃连接，从而融合更多的图像特征信息。然后，提出融合稀疏注意力和空间注意力的 BRSA(Bi-Level Routing and Spatial Attention)注意力模块，能够有效的提取全局和局部的位置信息，减少复杂背景下对于关键信息的干扰。最后，设计两种轻量高效的 BBot 模块和 C2fGhost 模块，以提高模型运算速度并减少参数量。实验结果表明，M-YOLO 相较于 YOLOv8，参数量降低约 1/3。在 TT100K 数据集和 GTSDB 数据集上，M-YOLOs 检测精度分别提升了 9.7%和 2.1%，M-YOLOn 检测精度分别提升了 14.5%和 2.6%，轻量化的同时具备更高的检测效果。本文提出的 M-YOLO 架构解决了浅层特征图在特征提取过程中信息丢失的问题，并显著降低模型特征提取过程中冗余的计算开销，在实景采集的数据集上证实效果有效，表明在交通标志检测任务中具有应用价值。

关键词：卷积神经网络;模型轻量化;目标检测;注意力模块;多尺度融合

DOI：10.19678/j.issn.1000-3428.006768269

中图分类号：TP391

An Improved Algorithm Combining Lightweight and Multi scale Fusion for Traffic Sign Detection

LAN Hong WANG Huizhao

School of Information Engineering, Jiangxi University of Science and Technology, Ganzhou 341000

Abstract: It is very important to detect traffic signs in the field of autonomous driving. How to detect these signs in time and accurately has great significance for improving driving safety and preventing traffic accidents. Aim at the small size, obstructed of traffic signs, in complex environments some of them are missed and incorrect discretion. A traffic sign detection network architecture M-YOLO (Multiple YOLO) is proposed based on the structure of YOLOv8. M-YOLO combines lightweight and multi-scale fusion. To handle the most precise detection tasks, the M-YOLOs (Multiple YOLO small) model has been created, while the network depth has been altered to generate a lighter M-YOLOn (Multiple YOLO nano) model which can fulfil requirements for detection in differing environments. The first, in response to the problem of small traffic sign targets and loss of image features, improves the network's feature learning ability for small targets by adding a small target detection layer to retain more feature information. To propose an efficient multi-scale feature pyramid fusion network, MPANet (Multiple Path Aggregation Network). This will reduce the dimensionality of shallow feature maps and perform skip connections to fuse more image feature information. A BRSA (Bi Level

基金项目：江西省研究生创新专项资金资助项目(YC2023-S659);

E-mail：1304921383@qq.com

Routing and Spatial Attention) attention module is proposed which integrates sparse attention and spatial attention in order to extract global and local position information whilst reducing interference with key information in complex backgrounds. Finally, to reduce the number of parameters and improve the computational speed of the model, two lightweight and efficient BBot and C2fGhost modules were developed. The results of the experimental tests show that M-YOLO reduces the number of parameters by about 1/3 in respect to YOLOv8. For the TT100K and GTSDB datasets, the detection accuracy of M-YOLOs was improved by 9.7% and 2.1% respectively, and the detection accuracy of M-YOLOn was improved by 14.5% and 2.6% respectively. As a result of this lightweight approach, the detection performance is improved. The proposed M-YOLO architecture overcomes the information loss problem in feature extraction from flat feature maps and greatly reduces the unnecessary computation in model feature extraction. The effectiveness of the method has been demonstrated on the dataset collected in realistic scenarios, which shows its application value in traffic sign detection tasks.

Keywords: convolutional neural network; lightweight models; object detection; attention module; multi-scale fusion

0 概述

交通标志识别是计算机视觉和图像分类在自动驾驶中应用的重要研究内容，具有广泛应用[1]。随着科技的发展和人们对于高质量生活的追求，人们涌入城市生活导致城市道路压力不断增大，产生了道路拥挤，交通事故等一系列交通问题。如何准确地对交通标志进行精确检测是我们尚未解决的难题。影响交通标志检测精度的因素有很多，如小目标分辨率低[2]、相互遮挡、复杂背景等。此外，模型参数量过大也是影响检测速度的重要因素。因此交通标志相较于一般目标更难检测，是一项非常具有挑战性的任务。

近年来，目标检测算法主要分为传统目标检测算法和基于深度学习[3]的目标检测算法。目前深度学习目标检测算法主要有双阶段目标检测(Two-Stage)算法和单阶段目标检测(One-Stage)算法。双阶段目标检测算法如 R-CNN[4]、Fast RCNN[5]、Faster RCNN[6]、Cascade R-CNN[7]、D2Det[8]等，其主要任务是定位和分类，首先找出图像中感兴趣的区域，然后对所选区域的目标进行分类并标注位置信息，但有着模型太大，检测速度慢等问题。单阶段检测算法如 SSD(Single Shot Multi-Box Detector)[9]和 YOLO[10-14]系列算法，不需要找出感兴趣的区域，采用分类回归的思想，直接对图像处理一次就能得到目标分类和位置信息，有着检测速度快的优点，但检测精度有所不足。SSD 对主干网络提取不同尺度的特征图来检测目标，但是对小目标的浅层特征提取效果较差。YOLOv3[12]采用 FPN 进行多尺度特征融合，将浅层特征和深层特征联系起来，使得特征图具有较强的语义信息。YOLOv5、YOLOv6[13] 、YOLOv7[14]和 YOLOv8 采用 PAFPN 进行特征融合，在 FPN 的基础上添加一条自底向上的路径，将融合后的浅层信息再次与深层信息融合，减少浅层信息的流失。不同尺度的特征具有不同的语义信息，多尺度特征融合可以将浅层特征和深层特征关联，得到更加丰富的特征表达。多尺度特征融合需要对不同尺度的特征进行融合，因此需要大量的上下采样和通道升降维，导致需要增加计算量和存储空间，从而提高模型的复杂度。同时不同尺度的特征融合会导致一些关键信息的丢失，从而降低模型的性能。

目前目标检测网络通过减少网络宽度和深度实现轻量化的效果。这种方法牺牲模型特征提取的能力，轻量化的同时带来检测精度显著减低。设计轻量化模块是解决这一缺点的方法之一。VGGNet[15]通过 3×3 的卷积核代替大卷积核；ResNet[16]提出 BottleNeck 减少大卷积核的通道数；GhostNet[17]和 ShuffleNet[18]使用深度可分离卷积设计了轻量化高效特征提取模块，并构建了轻量化网络模型。因此，设计轻量化模块并维持原模型的检测性能是目标检测模型的首要目标。

为解决上述目标检测网络存在的问题，本文首先对多尺度特征融合造成的额外计算开销，设计了轻量高效的 MPANet(Multiple Path Aggregation Network)，提出将多尺度特征降维，以此减缓特征融合的计算开销，并进行特征跳跃连接，以此加强浅层特征的提取，减缓 PANet 的特征损失问题。针对模型提取特征信息存在大量参数和计算冗余的问题，构建并设计了两种轻量化模块，对模型存在的

冗余参数和计算进行了优化，设计了轻量高效的目标检测网络。本文基于 YOLOv8 网络设计了 M-YOLO(Multiple-YOLO)网络架构。在此基础上构建 M-YOLOs 与 M-YOLOn 两种模型以应对不同场景的需求。实验结果显示，本文提出的多尺度特征融合网络 MPANet 能显著降低参数和计算开销；提出的轻量化特征提取模块能够维持检测精度的同时显著降低计算开销并提高模型检测速度；在 TT100K 和 GTSDB 等交通标志数据集进行验证，与 YOLOv8 原网络相比，M-YOLO 在使用更少的参数和计算量的情况下，拥有更高的检测精度，明显优于目前主流算法。

1 YOLOv8 模型介绍

YOLOv8 算法是目前卷积神经网络最先进的算法之一，在交通标志目标检测任务中有重要的应用价值，能够解决交通标志种类多样，复杂环境导致的检测难题，具有检测速度快，检测准确率高的优点。其结构分为输入端、主干网络(Backone)、特征融合网络(Neck)和输出端组成。输入端使用 Mosaic 数据增强的方式，随机提取四张图片进行随机缩放、随机裁剪和随机排列，不但丰富了检测的数据集，还随机增加了很多小目标对象，增强了网络的鲁棒性。主干网络使用 Conv、C2f 和 SPPF 模块组成，其中 C2f 模块仿造 YOLOv7 算法的 ELAN 结构，通过多分支进行跨层次连接，获得更加丰富的梯度流信息，形成一个具有增强特征表示能力的特征提取模块。Neck 部分采用 PANet(Path Aggregation Network)，加强不同尺度特征的融合能力。对三种尺寸特征图进行预测，对不同尺寸特征图生成的预测框预测，最终输出网络的预测结果。

2 M-YOLO 网络结构

2.1 M-YOLO 模型简介

M-YOLO(Multiple-YOLO)以 YOLOv8 网络为基础，增加一层小目标检测层，提高对小目标图像特征的提取能力。将原有 PANet 替换为本文提出的 MPANet，将浅层特征图通过卷积操作提取特征，并跳跃连接来融合特征（图 1 虚线部分），提高浅层信息的权重，缓解图像特征提取过程中图像信息流失的问题。增加轻量化特征提取模块 BBtot（Bi-Level Routing and Spatial Attention Bottleneck）模块和 C2fGhost 模块，删除主干网络中最后一层的 C2f 模块并在主干网络末端添加 BBot 模块，将特征融合网络中的 C2f 模块替换为 C2fGohst 模块，以此加强图像特征提取能力并减少特征提取过程中计算花销。整体结构如图 1 所示。

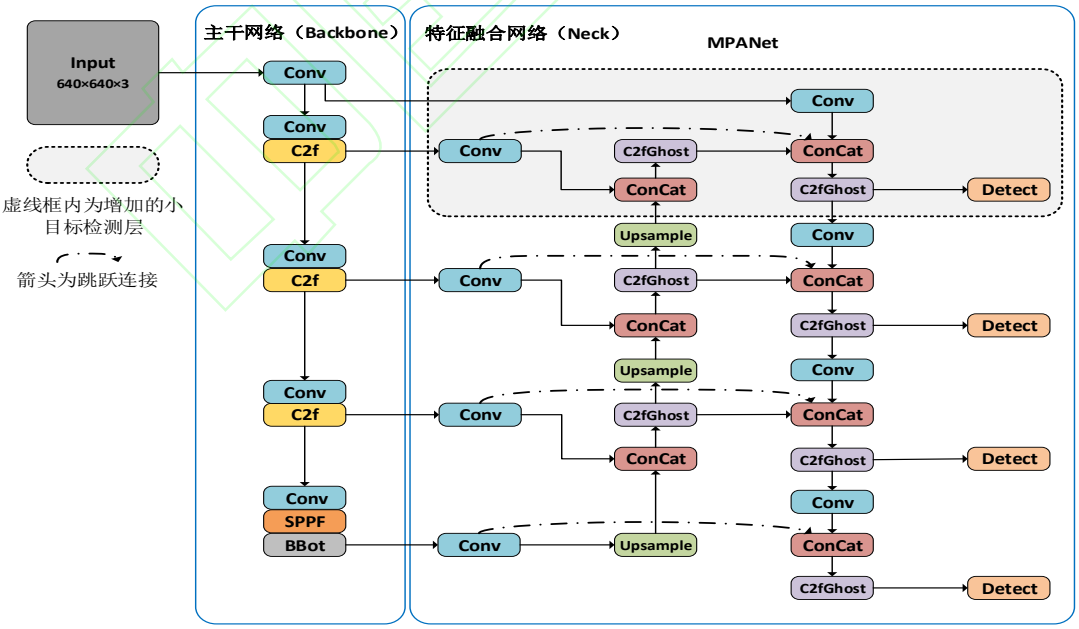


图 1 M-YOLO 网络结构图

Fig.1 M-YOLO structure diagram

2.2 小目标检测层

M-YOLO 网络通过增加小目标检测层来解决

YOLOv8 中存在的检测目标过小时，较深层特征图的小目标特征信息不充分从而出现漏检的问题，YOLOv8 网络 head 部分的主体是三个 Detect 检测器，对不同尺寸的特征图进行检测。当输入图像尺寸为 640*640 时，三个特征图的尺寸分别为 80*80，40*40，20*20。本文在原有的基础上新增 160*160 的检测层，相对于原有的三个检测层，保留了更多的特征信息，使得模型对于小目标检测的准确率更高。

2.3 多尺度融合网络 MPANet

为了增加特征提取，M-YOLO 模型采用 MPANet 替换原有的 PANet，改进 YOLOv8 中特征提取不充分的不足。PANet 直接对下采样得到的特征图进行特征融合，没有考虑到下采样操作带来的信息损失，导致对小目标的特征提取不充分，检测效果较差，同时产生冗余的计算开销。

为解决上述难题，本文提出轻量高效的多尺度融合网络 MPANet，其结构如图 2 所示。在多尺度特征融合网络中，从 backbone 到 Neck 同一级别的最短路径成为“干路”，其他路径称为“支路”。干路特征经过一系列的特征增强操作，原始特征被浅层特征和深层特征多次融合，其原始特征在特征图中的权重减弱，导致特征信息丢失。在 MPANet 中，为了缓解特征丢失的现象，将浅层特征通过卷积操作降维，并与支路特征层进行特征融合，同时与远端同一尺寸的特征层跳跃连接(图 2 虚线箭头)，以此加强原始特征在特征图中的权重。与原始 PANet 相比，MPANet 检测层可以直接获得浅层特征，解决 PANet 中的特征损失问题，并且经过多次特征融合，原始特征的权重得到增强，通过对原始特征进行卷积操作，可以显著降低模型的参数量和计算开销，在轻量化的同时拥有更高检测精度。

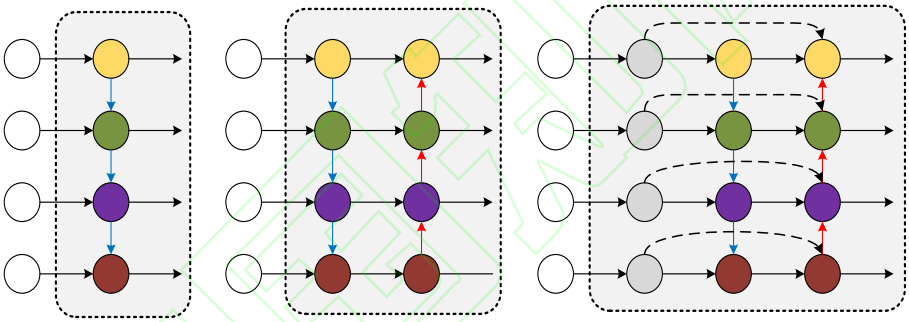


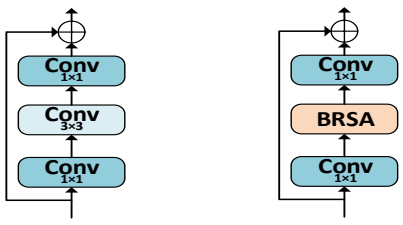
图 2 MPANet 原理图（由左到右分别为 FPN、PANet、MPANet）

Fig.2 MPANet schematic diagram (from left to right are FPN、PANet、MPANet

2.4 注意力与卷积融合模块 BBot

目前提取目标图像特征信息的方法主要有注意力机制和卷积操作。VASWANI 提出自注意力机制(Self-Attention)[19]，Dosovitskiy 等人提出的 Vision Transformer(ViT)计算机视觉上应用了自注意力机制[20]。自注意力机制可以减少外部信息的影响并且拥有捕获长距离依赖性的能力。卷积操作可以减少局部冗余的同时避免额外的计算并且有效学习大尺度图像中抽象和低分辨率的特征图。在此基础上，PAN 等人提出了一种兼顾了 Self-Attention 和 Convolution 的优点的混合模型 ACmix[21]。首先通过卷积对输入特征映射，得到了丰富的特征集，之后分别按照 Self-Attention 和 Convolution 的方式重用和聚合中间特征，充分结合了 Self-Attention 和 Convolution 的优点。

M-YOLO 网络中为了缓解图像特征提取过程中图像信息流失的问题，提出了轻量化特征提取模块 BBtot(BRSA Bottleneck)，该模块结合注意力机制和卷积操作的优点，通过将 BRSA 注意力(Bi-Level Routing and Spatial Attention)替换原 Bottleneck 结构的 3×3 的卷积，其中两个 1×1 的卷积分别用于降维和通道间的信息融合，充分地将 BRSA 注意力和卷积结合，在减少外部信息干扰并丰富了图像特征信息，同时减少了参数，降低了计算开销。其结构如图 3 所示。



(a)ResNet Bottleneck　(b)BRSA Bottleneck

图 3 ResNet Bottleneck 和 BRSA Bottleneck 结构对比图

Fig.3 ResNet Bottleneck and BRSA Bottleneck Structure Comparison Chart

BRSA注意力结合了稀疏注意力和空间注意力的优点，用于提取全局和局部的位置信息的，通过少量的开销提取特征图中重要的位置信息并减少无用信息的干扰。

BRSA 包含动态稀疏注意力模块 BRA 和空间注意力模块 SA，其结构如图 4 所示。主要思想是先筛选出大部分不相关的键值对，留下部分关键区域，随后通过空间注意力提升关键区域的特征表达。其工作流程如下：

1) 首先输入一张图片 X，$X\in R^{H\times W\times C}$，将其划分为 S×S 的不同的区域，其中每个区域包含 $\frac{HW}{S^2}$ 个特征向量，即 $X^r\in R^{S^2\times\frac{HW}{S^2}\times C}$，随后使用线性投影分别导出 Q、K、V 向量。其中 W^q、W^k、$W^v\in R^{C\times C}$ 分别是 Q、K、V 的投影权重。

$$Q=X^rW^q \tag{1}$$

$$K=X^rW^k \tag{2}$$

$$V=X^rW^v \tag{3}$$

2) 通过构造一个有向图来找到每个给定区域应该参与的区域。首先计算每个区域中 Q 和 K 的平均值，得到 Q^r、$K^r\in R^{S^2\times C}$，随后计算 Q^T 和 K^T 的区间相关性的邻接矩阵 A^r，逐行保存前 K 个连接的索引 $I^r\in N^{S^2\times K}$，I^r 是第 i 个区域的前 k 个最相关区域的索引。

$$A^r=Q^r(K^r)^T \tag{4}$$

$$I^r=topkIndex(A^r) \tag{5}$$

3) 随后聚集 K、V 的张量得到 K^g、V^g，其中 gather()函数的作用是将输入张量与索引组合得到新的张量，然后将 K^g、V^g 对使用注意力操作得到特征图 O。

$$K^g=gather(K,I^r) \tag{6}$$

$$V^g=gather(V,I^r) \tag{7}$$

$$O=attention(Q,K^g,V^g) \tag{8}$$

4) 最后将特征图 O 分别通过最大池化和平均池化操作，得到两个 $\frac{H}{S}\times\frac{W}{S}\times1$ 的特征图，经过 Concat 操作对两个特征图进行拼接，通过 7*7 卷积变为 1 通道的特征图，再通过 sigmoid 操作得到特征图,最后将特征图与特征图 O 相乘，得到最终生成的特征图 X。

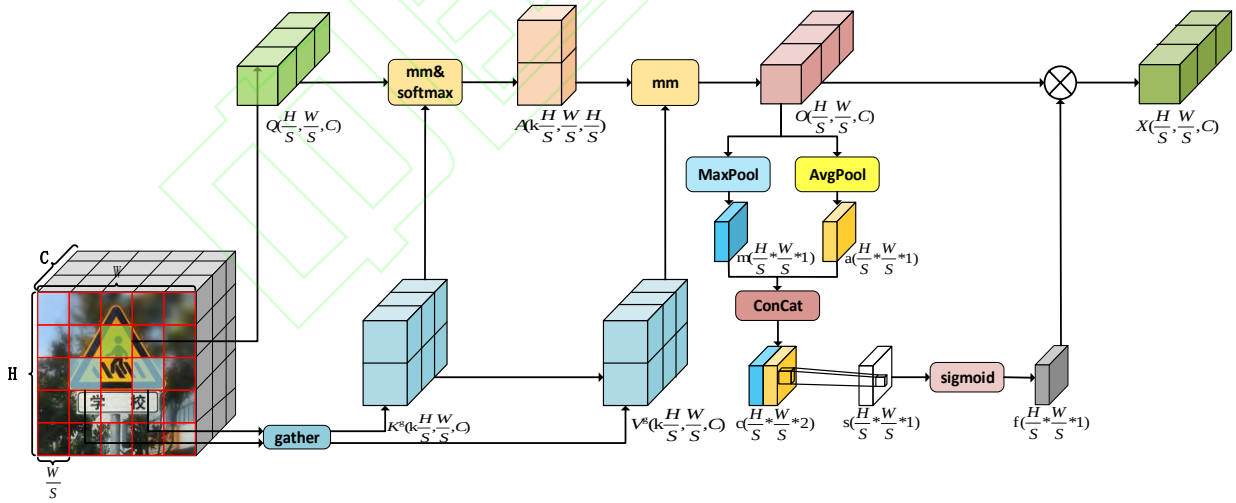


图 4 BRSA 注意力模块结构

Fig.4 Structure of attention module BRSA

2.5 轻量化特征提取模块 C2fGhost

在目标检测任务中，多尺度融合结构能促进不同尺寸的特征图之间的信息流通，但不同尺寸的特征图大量的升降维和上下采样操作导致产生大量的计算开销。针对这类问题，M-YOLO 中采用 GhostConv 模块设计了轻量高效的 C2fGhost 特征提取模块。GhostConv 模块提取图像特征过程可分为两部分，如图 5 所示。

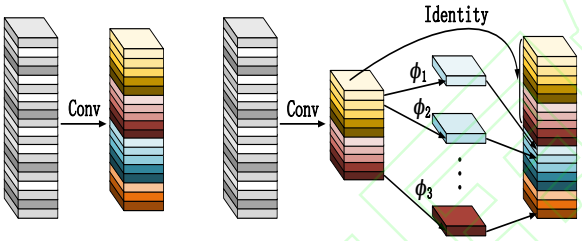
首先利用给定大小的卷积核进行普通卷积操作生成一部分特征图；另一部分利用深度可分离卷积分别对单通道进行卷积获得特征图，其中$\phi1$、$\phi2$、$\phi3$代表不同维度的特征图，Idendity 代表线性映射。Y'为卷积后的特征图，T 为卷积模块，f 为卷积前的特征图。

$$Y' = T * f' \tag{9}$$

原始特征经过卷积后生成 i 个特征图 Y_i'，i≤m，m 为卷积数量，然后经过线性映射后生成特征信息 y_{ij}，ψ_{ij}为线性映射，j≤n，n 为线性映射后特征图数量。最后拼接两部分特征图，得到最终特征图 Y。与原始卷积相比，GhostConv 只对其中一部分进行卷积操作，减少了卷积参数量，有效降低了模型的计算复杂度。

$$y_{ij} = \psi_{i,j}(Y_i'), \forall i = 1, \cdots, m; \forall j = 1, \cdots, n; \tag{10}$$

$$Y = Y' + y_{ij} \tag{11}$$



(a)Conv 模块　　(b)GhostConv 模块

图 5 Conv 和 GhostConv 结构图

Fig.5 Conv and GhostConv structural diagrams

GhostBottleneck 由下采样和深度可分离卷积实现，通过 GhostConv 进行特征提取，相比于 Bottleneck，使用更少的参数获得更高的特征提取。

C2fGhost 模块将 GhostBottleneck 层代替 Bottleneck 层，有效减少 Bottleneck 中普通卷积带来的冗余计算，在不削弱特征提取能力的情况下减少参数量。在保证轻量化的同时，对于不同尺度的模型采用不同的通道数，获得了更加丰富的梯度流信息。具体结构如图 6 所示。

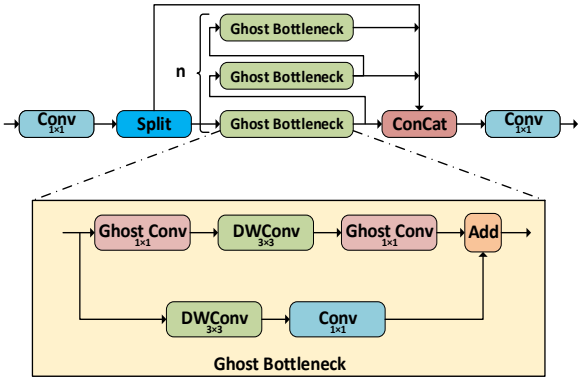


图 6 C2fGhost 结构图

Fig.6 C2fGhost structure diagram

3 实验分析

3.1 实验环境和参数配置

本文方法基于 PyTorch1.13.1 深度学习框架和 YOLOv8 框架实现，硬件规格为 NVIDIA GeForce RTX3060 12GB、Intel Core i5-12400F 和 16GB 内存。为确保实验公平性，不设置预训练权重，输入图片尺寸为 640*640，epoch 为300，动量参数设置为 0.937，初始学习率为 0.01，batchsize 设置为 16。

为了增强对比性，与 YOLOv8 网络类似，通过调整网络宽度和深度，本文使用 M-YOLO 网络架构构建了 M-YOLOs 与 M-YOLOn 两个模型，分别与 YOLOv8s 与 YOLOv8n 两个模型进行实验对比。其中 M-YOLOs 的深度为 0.33，宽度为 0.50，M-YOLOn 的深度为 0.33，宽度为 0.25。M-YOLOn 相较于 M-YOLOs 模型计算开销更低，适合应对不同环境下的检测需求。

3.2 数据集

1) TT100K 数据集[22]。TT100K 是我国的交通标志数据集，由清华大学联合腾讯公司整理并公布，提供 100000 张包含 30000 个交通标志，分辨率为 2048×2048，拍摄场景丰富，光照条件和天气变化有所不同，拍摄背景复杂，包含大量干扰信息，本文选取 9457 张图片，共 42 种类型的交通标志，其中 6598 张用于训练，1889 张用验证，970 张用测试。

2) GTSDB 数据集[23]。德国交通标志数据基准(GTSDB)采集不同天气条件下的真实道路场景，如雨天、雾天、晴天等，包含 900 张分辨率为 1369

×800 的道路交通图片，数据集内大部分为小目标，其被分为强制、禁止、警告和其他四种类型的交通标志，其中 600 张用于训练，300 张用验证。

3.3 评价指标

本文采用精确率(Precision)、召回率(Recall)、多类别平均精度(mAP)、模型权重和检测速度(FPS)作为评价指标。具体计算方法如下所示。

$$\mathrm{Pr}ecision=\frac{TP}{TP+FP} \tag{12}$$

$$\mathrm{Re}call=\frac{TP}{TP+FN} \tag{13}$$

$$\mathrm{AP}=\int_0^1 Precision(t)\,dt \tag{14}$$

$$\mathrm{mAP}=\frac{\sum_{n=1}^{N}AP_n}{N} \tag{15}$$

TP 表示样本为正且被预测为正的样本数量，FP 样本为负但被预测为正的样本数量，FN 表示实际为正，但被预测为负的样本的数量。

3.4 实验结果分析

为检验 M-YOLO 的检测性能，本文在 TT100K 数据集和 GTSDB 数据集进行训练与测试，如表 1 所示。与原架构相比，M-YOLO 架构参数量降低约 1/3，精确率和召回率显著提高。在 TT100K 数据集上，M-YOLOs 的 mAP@0.5 分别提升了 9.7%和 2.1%，在 GTSDB 数据集上，M-YOLOn 的 mAP@0.5 分别提升了 14.5%和 2.6%。

实验表明，M-YOLO 网络相较于原架构，在轻量化的同时显著提高交通标志目标检测的检测精度，在小目标检测的任务中有着更好的检测效果，具有良好的鲁棒性。

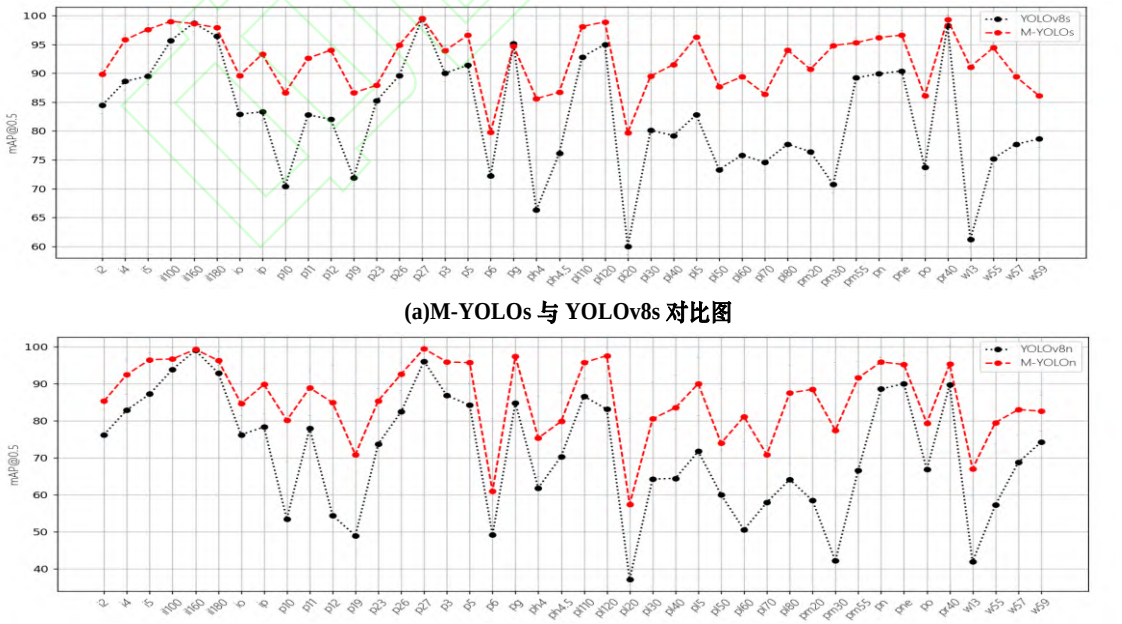
表 1 在 TT100K 和 GTSDB 数据集上的实验结果对比

Table 1 Comparison of experimental results on TT100K and GTSDB datasets

数据集	模型	P	R	mAP@0.5	参数量
TT100K	YOLOv8s	82.8	74.9	82.5	21.5
	M-YOLOs	**89.8**	**85.2**	**92.2**	14.2
	YOLOv8n	71.3	64.6	71.3	6.26
	M-YOLOn	84.1	76.9	85.8	**4.63**
GTSDB	YOLOv8s	95.4	89.0	94.5	21.5
	M-YOLOs	**95.9**	**95.3**	96.6	14.2
	YOLOv8n	91.4	92.6	94.6	6.24
	M-YOLOn	94.3	94.0	**97.2**	**4.54**

注：加粗字体为每列最优值。

为验证 M-YOLO 在 TT100K 数据集的各类别检测性能的提升，本文分别将 M-YOLOs、YOLOv8s 和 M-YOLOn、YOLOv8n 进行对比，由图 7 所示，M-YOLO 架构中各类别检测精度均大幅度提升，从整体来看，M-YOLO 相较于 YOLOv8 检测性能更加优秀。



(a)M-YOLOs 与 YOLOv8s 对比图

(a)M-YOLOn 与 YOLOv8n 对比图

图 7 TT100K 数据集各类别检测性能对比

Figure 7 Comparison of detection performance for various categories in the TT100K dataset

3.5 消融实验

为确保本文算法的有效性和合理性，在TT100K数据集上，对M-YOLO网络设计一系列实验，①-⑩是单独改进和各方法组合的模型的消融实验，如表2所示。

实验表明，本文提出的增加小目标检测层、MPANet和BBot模块能够在减低模型大小的情况下显著提高模型对小目标的检测精度，同对小目标的定位更加明确，使用C2fGhost模块能够少量降低模型参数量，减少卷积层的计算冗余的同时微量提升检测精度。

表2 消融实验表

Table 2 Ablation experiment table

模型	小目标层	MPANet	BBot	C2fGhost	P	R	mAP@0.5	参数量
①					82.8	74.9	82.5	21.5
②	√				84.4	80.6	87.0	21.1
③		√			85.1	74.8	83.8	15.0
④			√		86.4	73.0	83.2	19.7
⑤				√	82.5	74.4	82.7	18.1
⑥	√	√			86.6	84.3	90.7	16.9
⑦	√		√		86.8	79.8	88.1	19.4
⑧	√			√	86.9	79.4	87.3	17.6
⑨	√	√	√		88.9	85.9	91.9	15.1
⑩	√	√		√	88.8	84.6	91.0	16.0
All	√	√	√	√	**89.8**	85.2	**92.2**	**14.2**

注:√表示采用，加粗字体为每列最优值。

3.6 不同特征融合网络对比实验分析

为验证MPANet融合特征信息的优秀性能，本文选取BiFPN、AFPN、GFPN与MPANet在TT100K数据集上进行对比实验。为保证实验公平性，将主干网络Backbone固定，只比较不同特征融合网络Neck的性能。如表3所示，在TT100K数据集上精确率提升2.3%，mAP@0.5提升1.3%且参数量降低了30.2%。由此证明MPANet相比于其他特征融合网络在模型参数量上更具优势，并且能显著提升准确率。

表3 在TT100K数据集上不同特征融合网络对比实验

Table 3 Comparative Experiment on Different Feature Fusion Networks on TT100K Dataset

特征融合网络	P	R	mAP@0.5	参数量
PANet	82.8	74.9	82.5	21.5
BiFPN[24]	79.7	75.9	82.6	21.6
AFPN[25]	80.9	71.3	80.0	22.5
GFPN[26]	82.1	72.7	81.9	23.6
MPANet	**85.1**	74.8	**83.8**	**15.0**

注: 加粗字体为每列最优值。

3.7 注意力机制改进对比实验分析

为了验证本文提出的BBot模块中BRSA注意力对模型大小和检测精度的影响，本文在TT100K数据集上，选取MHSA、BRA、CBAM注意力对BBot模块的BRSA注意力进行替换，设计4种对比实验进行分析。实验结果如表4所示，BRSA在TT100K数据集上mAP@0.5提升0.7%，准确率提升3.6%，相比于其他注意力模块，能够更加充分地提取图像的有效信息，提高模型对交通标志检测的准确率。

表4 在TT100K数据集上不同注意力机制BBot对比实验

Table 4 Comparative Experiment on BBot of Different Attention Mechanisms on TT100K Dataset

注意力机制	P	R	mAP@0.5	参数量
--	82.8	74.9	82.5	21.5
MHSA[27]	80.7	73.7	82.0	19.6
BRA[28]	84.1	71.2	81.9	19.7
CBAM[29]	81.6	75.5	82.6	19.3
BRSA	**86.4**	73.0	**83.2**	19.7

注: 加粗字体为每列最优值。

3.8 TT100K数据集对比实验分析

本文在相同的环境配置下选取Faster-RCNN、Deformable DETR、TSP-RCNN、YOLOv3、YOLOv5、YOLOv6、YOLOv8s和当前优秀算法在

TT100K 数据集上进行对比试验。评价指标为精确率(Precision)、召回率 (Recall)，多类别平均精度(mAP@0.5)、模型权重和 FPS。实验结果由表 5 所示。M-YOLO 算法在 TT100K 数据集上的多类别平均精度和模型参数量有着显著优势。M-YOLOs 相较于双阶段目标检测算法 Faster-RCNN，M-YOLOs 在参数量为其 8.4%的情况下，mAP@0.5 和 FPS 分别提升了 22.7%和 37.2。通过对比可知，M-YOLOs 综合检测性能全面超过 Deformable DETR、TSP-RCNN、文献[30]和文献[31]等一般检测器。

在目标检测过程中，输入尺寸和 mAP@0.5 正相关，与 FPS 负相关，M-YOLOs 在输入尺寸为 640 的情况下，mAP@0.5 和 FPS 均高于文献[32]和文献[33]。在相同输入尺寸的条件下，mAP@0.5 远高于文献[34]、YOLOv3、YOLOv5s 和 YOLOv6，相较于 YOLOv8s 算法，参数量降低了 33.6%，精确率提高 7.0%，召回率提高 10.3%，mAP@0.5 提高 9.7%，达到了 92.2%，相比于其他模型在在精度和复杂度上具备更高的效率。同时参数量更小的 M-YOLOn 的 mAP@0.5 达到 85.8%，模型参数量仅为 4.63M，相比于 M-YOLOs，在牺牲了部分精度的情况下，获得更低的计算开销和更高的帧率，适用于缺乏计算资源的极端环境下的检测。

实验表明，本文提出的 M-YOLO 能够在显著降低模型参数量的同时准确提高对交通标志的检测精度，有效地解决小尺度交通标志的漏检问题，对小目标的精确定位和分类小尺度交通目标具有更加优秀的性能。

表 5 在 TT100K 数据集上和主流算法对比

Table 5 Comparison with mainstream algorithms on the TT100K dataset

模型	Neck	Size	P	R	mAP@0.5	参数量	FPS
Faster-RCNN	FPN	608	--	--	69.5	167.7	2.2
Deformable DETR	--	960	73.4	79.3	77.1	40.0	5.8
TSP-RCNN	--	960	82.5	84.3	81.2	--	4.7
YOLOv3	FPN	640	81.4	72.6	81.1	198.0	22.0
YOLOv5s	PAN	640	83.1	76.9	82.4	13.9	42.7
文献[30]	FPN	608	--	--	75.2	--	31.0
文献[31]	FPN	608	--	--	82	--	35.7
文献[32]	OverFeat	2048	--	--	81.6	--	5.8
文献[33]	RFB-F3+PAN	1024	--	--	86.6	--	34.5
文献[34]	PAN	640	71.7	66.7	71.9	6.1	--
YOLOv6	Rep-PAN	640	71.9	68.3	74.2	31.3	36.7
YOLOv8s	PAN	640	82.8	74.9	82.5	21.5	37.6
YOLOv8n	PAN	640	71.3	64.6	71.3	6.26	45.4
M-YOLOs	MPAN	640	**89.8**	**85.2**	**92.2**	14.2	39.4
M-YOLOn	MPAN	640	84.1	76.9	85.8	**4.63**	**47.2**

注: 加粗字体为每列最优值。

3.9 GTSDB 数据集对比实验分析

为了进一步验证 M-YOLO 架构对小交通标志的检测性能，将 M-YOLOs 和 M-YOLOn 与其他优秀交通标志检测算法在 GTSDB 数据集上进行实验，实验结果如表 6 所示。GTSDB 数据集中大部分交通标志目标为小目标，像素在 16×16~128×128 之间。禁止、建议和警告为单类别的平均精度。

表 6 在 GTSDB 数据集上和主流算法对比

Table 6 Comparison with mainstream algorithms on the GTSDB dataset

模型	mAP@0.5	禁止	建议	警告
LFFPN[35]	82.4	90.3	71.4	82.4
CAFFNet[36]	96.7	97.9	93.5	98.8
MTSDet[37]	92.9	90.9	88.6	99.1
YOLOv3	95.9	99.3	91.6	95.3
YOLOv6	94.2	99.2	92.7	94.9
YOLOv5s	93.2	97.8	87.9	95.5
YOLOv7	78.8	96.6	65.5	92.0
YOLOv8s	94.5	99.1	89.7	94.8
YOLOv8n	94.6	99.3	87.4	97.1
M-YOLOs	96.6	98.7	94.1	97.0

M-YOLOn	**97.2**	98.7	**94.4**	98.4

注: 加粗字体为每列最优值。

从表 6 可得出，本文提出的 M-YOLOs 和 M-YOLOn 在 GTSDB 数据集上 mAP@0.5 分别提升 2.1%和 2.6%,其中建议和警告的检测精度显著提升，mAP@0.5 明显优于其他算法。证明了本文提出的 M-YOLO 在小目标的检测上具有优秀的检测效果，具有良好的鲁棒性。和其他优秀算法相比，单类别检测平均精度相对平衡，更加适合应用于实际场景，具有一定的竞争力。

3.10 检测结果对比

为了更加直观地展示 M-YOLO 架构的检测性能，本文将 M-YOLOs 与 YOLOv8s 分别在 TT100K 和 GTSDB 数据集上选取部分图片进行检测对比。图 8 为 TT100K 数据集的检测结果，图 9 为 GTSDB 数据集的检测结果。考虑到不同地区路况差异性，本文在不同场景下进行实时采样进行检测，检测效果如图 10 所示。

由图 8 可知，YOLOv8s 检测前 4 张图片均出现漏检的情况，第 5 张图片将 pl30 标志错误检测为 pl60 标志，第 6 张图片将 i2 标志错误检测为 i4 标志，并且 p26 置信度为 0.66，M-YOLOs 检测置信度为 0.90 且其他交通标志均准确检测。图 9 观察可知，YOLOv8s 在昏暗环境下对小交通标志的检测效果较差，前四张图片均出现漏检，第 5 张图片将窗户错误检测为警告标志，第 6 张图片将车灯错误检测为警告标志，而 M-YOLOs 并没有出现错误检测的情况。图 10 观察可知，在实时采集的 6 张图片中对远处小交通标志均出现漏检，第 5 张图片的 pl40 标志被树木遮挡，导致 YOLOv8s 没有检测出来，而 M-YOLOs 能够很好地应对这类情况。由此证明 M-YOLOs 能够对各类别交通标志进行精准定位和识别，对远处小目标能实现良好的检测效果，并且对小尺度交通标志的定位更加明确，大大减少了错检、漏检的情况发生，因此 M-YOLO 架构具有优秀的检测性能和鲁棒性。

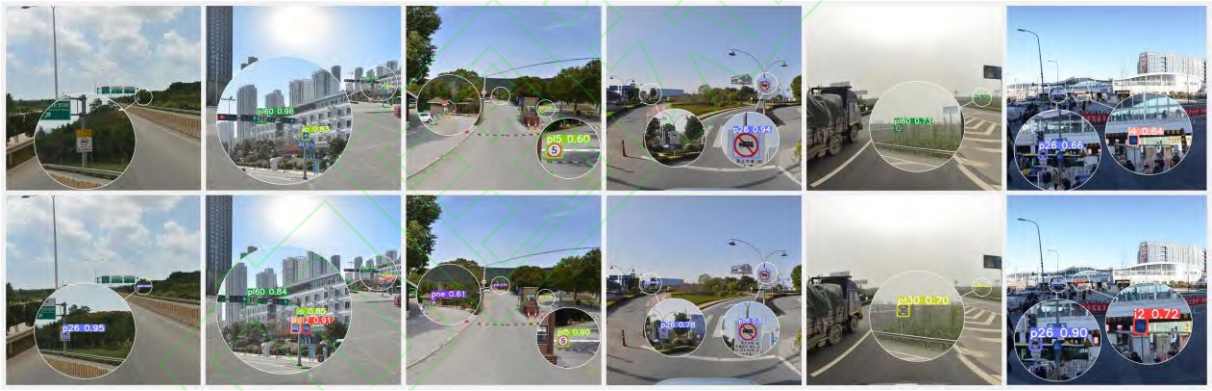


图 8 TT100K 数据集检测结果(上一行为 YOLOv8s，下一行为 M-YOLOs)

Fig.8 TT100K dataset detection results (Previous behavior YOLOv8s, next behavior M-YOLOs)

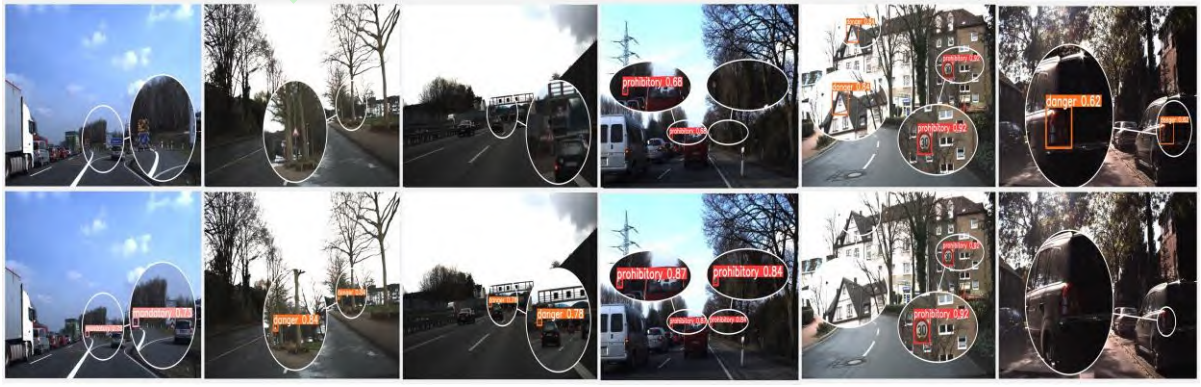


图 9 GTSDB 数据集检测结果(上一行为 YOLOv8s，下一行为 M-YOLOs)

Fig.9 GTSDB dataset detection results (Previous behavior YOLOv8s, next behavior M-YOLOs)

图 10 实时采样检测结果上一行为 YOLOv8s，下一行为 M-YOLOs)

Fig.10 Real time sampling detection results (Previous behavior YOLOv8s, next behavior M-YOLOs)

4 结论

为了平衡目标检测算法模型复杂度和检测精度，提高检测性能，通过对多尺度融合结构和特征提取模块轻量化进行优化，设计了轻量化架构 M-YOLO 并构建了 M-YOLOs 和 M-YOLOn 两种大小的模型。

针对交通标志目标背景复杂，目标之间存在遮挡和远处小目标等情况导致检测精度低，从而发生误检，漏检等问题。本文增加小目标检测层，提高对小目标特征提取能力，减少小目标特征信息的丢失。同时对多尺度特征融合结构进行改进，构建了 MPANet 结构，减少特征融合计算开销，加强浅层特征对检测的影响。在主干网络嵌入融合 BRSA 注意力的 BBot 模块，结合稀疏注意力和空间注意力的优点，减少了复杂环境下无用信息的干扰。最后在特征融合结构中使用 C2fGhsot 模块，在减少参数量的同时保证模型的精度。

本文在 TT100K 数据集和 GTSDB 数据集上进行大量实验，同时采集实际环境下的道路图进行验证，证明了 M-YOLO 平衡了模型复杂度和检测精度，以更少的参数获得更高的检测效果。相较于目前最新的目标检测算法，有着较大程度的性能提升，适用于不同条件下的交通标志目标检测。在后续研究中，将进一步研究本文特征融合结构和特征提取模块在其它 CNN 任务中的应用。

参考文献

[1] 董文轩,梁宏涛,刘国柱,胡强,于旭.深度卷积应用于目标检测算法综述[J].计算机科学与探索,2022,16(05):1025-1042.

[2] 刘洪江，王懋，刘丽华.基于深度学习的小目标检测综述[J].计算机工程与科学，2021,43(8):1429-1442.

[3] Hinton G E,Salakhutdinov R.Reducing the dimensionality of data with neural networks[J].Science,2006,313(5786):504-507.

[4] 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.

[5] GIRSHICK R. Fast R-CNN[C]// Proceeding of Conference on Computer Vision and Pattern Recognition. Boston, MA, US: IEEE, 2015: 1440-1448.

[6] REN S,HE K,GIRSHICK R,et al.Faster R-CNN:towards real-time object detection with region proposal networks[C]//Advances in Neural Information Processing Systems,2015.

[7] CAI Z W,VASCONCELOS N. Cascade R-CNN:delving into high quality object detection[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition.Washington D. C.,USA:IEEE Press,2018:6154-6162.

[8] CAO J L,CHOLAKKAL H,ANWER R M,et al. D2Det:towards high quality object detection and instance segmentation[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C.,USA:IEEE Press,2020:11482-11491.

[9] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single Shot MultiBox Detector[C]//European Conference on Computer Vision (ECCV). Amsterdam: Springer, 2016: 21-37.
[10] REDMON J,DIVVALA S,GIRSHICK R,et al.You only look once:unified,real-time object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2016:779-788.
[11] REDMON J,FARHADI A.YOLO9000:better,faster,stronger[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2017:7263-7271.
[12] REDMON J,FARHADI A.Yolov3:an incremental improvement[J].arXiv:1804.02767,2018.
[13]Li C, Li L, Jiang H, et al. YOLOv6: A single-stage object detection framework for industrial applications[J]. arXiv preprint arXiv:2209.02976, 2022.
[14]WANG C Y, BOCHKOVSKIY A , LIAO H. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[J]. arXiv preprints arXiv: 2207, 02696, 2022.
[15]Karen Simonyan, Andrew Zisserman. ”Very Deep Convolutional Networks for Large-Scale Image Recognition” Computer Vision and Pattern Recognition(cs.CV).[C]. 2014, arXiv:1409.1556.
[16]He K ,Zhang X, Ren S, et al. Deep ResidualLearning for Image Recognition[C]// IEEE Conference on Computer Vision & Pattern Recognition. IEEE Computer Society, 2016.
[17]Han K ,Wang, Y H, Tian Q, Ghostnet: more features from cheap operations[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press,2020:1577-1586
[18]ZHANG X,ZHOU X,LIN M, et al. Shufflenet: an extremely efficient convolutional neural network for mobile devices[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2018:6848-6856.
[19]VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: ACM,2017: 6000-6010.
[20] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[J]. arXiv preprint arXiv:2010.11929, 2020.
[21] PAN X, GE C, LU R, et al. On the integration of self-attention and convolution[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE Press 2022: 815-825.
[22]SHOJAEIFARD A，AMROUDI A N，MANSOORI A，et al. A Novel Genetically Optimized Convolutional Neural Network for Traffic Sign Recognition: A New Benchmark on Belgium and Chinese Traffic Sign Datasets[J].Neural processing letters,2019,50(3):3019-3043.
[23]HOUBEN S，STALLKAMP J，SALMEN J，et al. Detection of traffic signs in real-world images: The German Traffic Sign Detection Benchmark[C]//The 2013 international joint conference on neural networks.Washington D. C. ，USA:IEEE Press，2016: 761-769.
[24]Tan M, Pang R, Le Q V. Efficientdet: Scalable and efficient object detection[C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020: 10781-10790.
[25]Guoyu Yang, Jie Lei , Zhikuan Zhu. ”AFPN: Asymptotic Feature Pyramid Network for Object Detection” Computer Vision and Pattern Recognition (cs.CV).[C]. 2023, arXiv:2306.15988.
[26]Yiqi Jiang, Zhiyu Tan, Junyan Wang. ”GiraffeDet: A Heavy-Neck Paradigm for Object Detection” Computer Vision and Pattern Recognition (cs.CV).[C]. 2022, arXiv:2202.04256.
[27]SRINIVAS A, LIN T Y, PARMAR N, et al. Bottleneck transformers for visual recognition[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 16519-16529.
[28]Zhu L, Wang X, Ke Z, et al. BiFormer: Vision Transformer with Bi-Level Routing Attention[J]. arXiv preprint arXiv:2303.08810, 2023.
[29]Woo Sanghyun, Park Jongchan, Lee Joon-Young, Cbam: Convolutional block attention module[C]//European Conference on Computer Vision. 2018: 3-19.
[30]刘紫燕，袁磊，朱明成，等. 融合 SPP 和改进 FPN 的 YOLOv3 交通标志检测[J]. 计算机工程与应用, 2021,57(7): 164-170. LIU Z Y, YUAN L, ZHU M CH, et al. YOLOv3 Traffic sign Detection based on SPP and Improved FPN [J]. Computer Engineering and Applications, 2021, 57(7): 164-170.
[31]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. WANG B, HE Y.

Traffic sign detection based on improved YOLOv3[J]. Journal of Sichuan University (Natural Science Edition),2022,59(1):57-67.
[32]ZHU Z，LIANG D，ZHANG S，et al. Traffic-sign detectionand classification in the wild[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway，USA：IEEE Press，2016：2110-2118.
[33]陈梦涛，余粟. 基于改进 YOLOV4 模型的交通标志识别研究[J].微电子学与计算机，2022，39(1)：17-25.CHEN M T，YU S.Research on traffic sign recognitionbased on improved YOLOV4 model[J].Microelectronics＆Computer，2022，39(1)：17-25.
[34]熊恩杰，张荣芬，刘宇红，彭靖翔. 面向交通标志的 Ghost-YOLOv8 检测算法[J]. 计算机工程与应用, 2023, 59(20): 200-207.XIONG Enjie, ZHANG Rongfen, LIU Yuhong, PENG Jingxiang. Ghost-YOLOv8 Detection Algorithm for Traffic Signs[J]. Computer Engineering and Applications, 2023, 59(20): 200-207.
[35]REN K，HUANG L，FAN C，et al. Real-time traffic sign detection network using DS-DetNet and lite fusion FPN[J]. Journal of Real-Time Image Processing，2021，18(6): 2181-2191.
[36]LIU F，QIAN Y，LI H，et al. CAFFNet: channel attention and feature fusion network for multi-target traffic sign detection[J]. International Journal of Pattern Recognition and Artificial Intelligence，2021，35(07): 2152008.
[37]WEI H，ZHANG Q，QIAN Y，et al.MTSDet: multi-scale traffic sign detection with attention and path aggregation[J]. Applied Intelligence, 2023, 53(1):238-250.