

基于 YOLOv7 的交通目标检测算法研究

王沛雪, 张富春, 董晨乐

(延安大学 物理与电子信息学院, 陕西 延安 716000)

摘要: 针对交通场景中, 由于光照、遮挡、目标小以及背景复杂等因素导致目标检测精度低, 易出现漏检和误检问题的情况, 提出了一种基于 YOLOv7 的交通目标检测算法; 该算法在主干网络中融入多头注意力机制, 以增强网络特征学习能力, 从而更好地捕获数据和特征内部的相关性; 在 YOLOv7 颈部网络引入协调注意力模块 (CA), 将位置信息嵌入到注意力机制中, 忽略无关信息的干扰, 以增强网络的特征提取能力; 增加一个多尺度检测网络, 以增强模型对不同尺度目标的检测能力; 将 CIoU 损失函数更改为 SIoU 函数, 以减少模型收敛不稳定问题, 提高模型的鲁棒性; 实验结果表明, 改进的算法在 BDD100K 公开数据集上的检测精度和速度分别达到了 59.8% mAP 和 96.2 FPS, 相比原算法检测精度提高了 2.5%; 这表明改进的算法在满足实时性要求的同时, 具备良好的检测精度, 适用于复杂情况下的交通目标检测任务。

关键词: 交通目标检测; YOLOv7 网络; 注意力机制; 浅层网络检测层; SIoU 损失函数

Research on Traffic Object Detection Algorithm Based on YOLOv7

WANG Peixue, ZHANG Fuchun, DONG Chenle

(School of Physics and Electronic Information, Yan'an University, Yan'an 716000, China)

Abstract: Aiming at the factors of lighting, occlusion, small target and complex background in complex traffic scenes, it leads to low target detection accuracy and is prone to cause missed and false detections, a traffic target detection algorithm based on YOLOv7 is proposed. To better capture the internal correlation within the data and features, the algorithm incorporates the multi-head attention mechanism into the backbone network to enhance the network feature learning ability. The coordinated attention (CA) module is introduced into the YOLOv7 neck network, and the position information is embedded into the attention mechanism, which can ignore the interference of irrelevant information and enhance the feature extraction ability of the network. A multi-scale detection network is added to enhance the detection capability of the model for different scale targets. The CIoU loss function is modified to the SIoU function to reduce the convergence and instability of the model, and improve the robustness of the model. The experimental results show that the mean average precision (mAP) and detection speed of the improved algorithm on the BDD100K public dataset reach up to 59.8% and 96.2 frames per second (FPS), respectively, with an increase of 2.5% compared to that of the original algorithm. It shows that the improved algorithm has a good detection accuracy while meeting the real-time requirements, which is suitable for traffic target detections in complex situations.

Keywords: traffic object detection; YOLOv7 network; attention mechanisms; shallow network detection layer; SIoU loss function

0 引言

随着经济的发展和科技的进步, 自动驾驶技术的研究与发展受到人们的广泛关注^[1]。在自动驾驶领域, 良好的实时交通目标检测与识别对保障道路环境安全十分重要。道路交通目标检测通常分为 4 类: 车辆检测、行人检测、交通标志检测和其他障碍物检测。如何准确快速地识别前方道路上出现的物体, 为规范行车安全、保护公共安全并制定下一步如何避免障碍物的策略, 拥有重大的实际意义。

近年来, 深度学习的发展带动了基于卷积神经网络 (CNN, convolutional neural network)^[1] 的目标检测算法的迅猛发展, 其具有结构简单、检测效果好、速度快等优势, 已成为当前目标检测研究的主流方向^[2]。目前, 基于深度

学习的目标检测算法分为两种类型: 两阶段模型 (Two-stage) 和单阶段模型 (One-stage)^[3]。其中, 两阶段模型通过生成目标候选框区域来进行分类和目标定位任务, 代表算法以区域建议网络的 R-CNN 为主, 随后发展出 Fast R-CNN^[4]、Faster R-CNN^[5] 等算法, 检测精度大幅提升, 但参数量较大, 导致检测速度慢。单阶段模型将目标检测问题变为一个回归问题, 不生成指定的候选区域, 直接输出分类和定位信息, 以 YOLO 算法和 SSD^[6] 系列算法为代表。相比于两阶段目标检测算法, 单阶段目标检测算法在检测精度上虽然有所降低, 但在检测速度上得到了一定幅度的提升。尽管近些年来世界各地的专家学者们围绕交通场景中的目标检测算法的研究取得相当大的进步与发展, 但从

收稿日期: 2023-04-26; 修回日期: 2023-06-09。

基金项目: 国家自然科学基金 (62264015); 延安市科技创新项目 (2017CXTD-01)。

作者简介: 王沛雪 (1999-), 女, 硕士。

张富春 (1972-), 男, 博士, 教授。

引用格式: 王沛雪, 张富春, 董晨乐. 基于 YOLOv7 的交通目标检测算法研究 [J]. 计算机测量与控制, 2024, 32(4): 74-80.

长远的角度来看, 现有的目标检测算法并不能满足实际的应用需求, 仍有许多问题亟待解决: 如存在待检测目标密集且遮挡、由于设备或环境原因造成捕获图像存在分辨率低或模糊以及所捕获照片背景复杂等情况, 从而使检测精度和速度难以满足实际需求。

针对交通目标检测领域存在的诸多挑战, 文献 [7] 在基于 Tiny-YOLOv3 模型的基础上增加卷积层和引入 1×1 卷积核来增强模型特征提取能力, 减少模型计算力, 保证了检测实时性, 但不适用各种复杂交通场景下的检测任务; 文献 [8] 提出了一种新的多传感器和多级增强卷积网络结构模型 MME-YOLO, 在处理尺度、遮挡和照明等问题中该模型检测效果很好, 但模型复杂度较高; 文献 [9] 提出了一种 Attention-YOLOv4, 在主干网络中将注意力机制与残差块结合, 提高复杂背景下目标的检测精度, 但低分辨率图像中物体的检测能力不足; 文献 [10] 改进了 YOLOv4-Tiny 算法, 通过暗通道先验理论对图片进行去雾处理, 提高了在雨、雾等恶劣环境下小目标识别能力, 但模型结构复杂, 不易于移动端的部署; 文献 [11] 提出了 Faster R-CNN 和 YOLO 的混合模型来进行检测, 拥有较高的检测精度, 但检测速度慢, 训练时间长, 无法满足实时性要求。

本文提出了一种改进的基于 YOLOv7^[12] 交通目标检测算法, 旨在解决目前关于复杂交通场景下道路交通目标检测算法的问题。本文的主要贡献有以下几点: 在原有 YOLOv7 网络结构上新增加一个浅层网络检测头, 提高模型对小目标的检测能力; 提出一个新的模块 EL-MHSA, 将多头注意力机制 (MHSA, multi-head self attention) 融入主干网络中, 增强模型获取全局信息的能力; 在颈部网络加入协调注意力 (CA, coordinate attention) 模块, 使模型可以忽略无关信息的干扰, 提高模型检测能力; 将损失函数改为 SIoU, 减少损失函数收敛不稳定的问题, 提高了网络对于目标尺寸的鲁棒性。实验结果表明, 改进的 YOLOv7 算法在复杂交通目标检测任务中检测精度 $mAP@0.5$ 和 $mAP@0.5:0.95$ 分别提升了 2.5% 和 1.4% 的同时, 检测速度可以达到 96.2 FPS, 满足复杂交通场景下目标检测的准确性和实时性要求。

1 YOLOv7 算法概述

YOLOv7 是目前目标检测 YOLO 系列中比较先进的算法, 该网络进一步提高了目标检测的速度与精度。YOLOv7 算法相比于其他算法, 在 5~160 FPS 间的速度和精度方面表现得更优秀。

YOLOv7 算法采用了许多新的策略, 如高效聚合注意力网络 (ELAN, efficient layer aggregation network)、基于级联的模型 (concatenation-based models) 缩放、卷积重参数化等, 并在速度和精度之间取得了很好的平衡。YOLOv7 网络结构主要由输入端 (Input)、主干网络 (Backbone)、颈部网络 (Neck) 和头部 (Head) 组成, 如图 1 所示。Input 模块将输入的图像经数据增强、裁剪、缩放等一系列处理后, 将图像统一变为 640×640 像素, 以便能满足主干网络的输入要求。Backbone 模块主要由卷积 (BConv)、ELAN 层和 MP-

Conv 层这 3 个部分组成。其中, BConv 用于提取不同尺度的图像特征; ELAN 模块结构对不同尺度的连接路径进行控制, 在预防了由于输入信息和梯度信息在经过多层网络传播后出现信息丢失的同时还提升了网络精度。MPConv 模块通过融合当前特征层和卷积处理后的信息的方式, 从而提高网络的提取能力。Neck 模块中, 在颈部中引入路径聚合特征金字塔网络 (PAFPN, path aggregation feature pyramid network) 结构, 加强不同层次间的特征聚合。在 Head 模块中, YOLOv7 利用 3 种不同大小尺寸的 IDetect 检测头, 经过重参数化输出 3 个不同尺寸的预测结果。

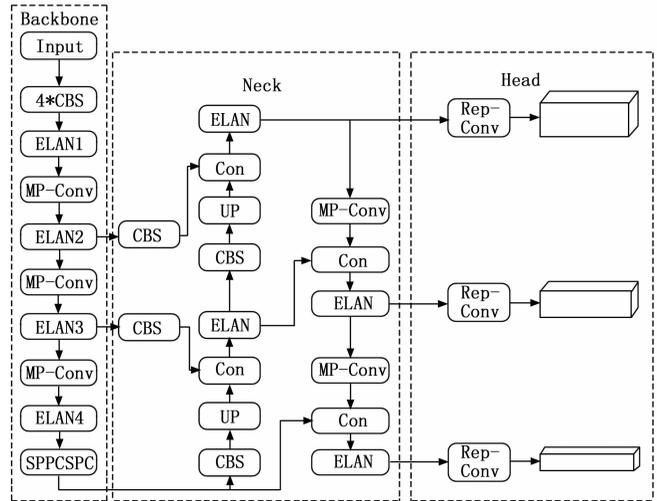


图 1 YOLOv7 网络结构图

2 改进的 YOLOv7 算法

2.1 总体网络结构

为了解决目前交通场景中目标检测由光照、遮挡、目标小以及背景复杂等原因^[13]而造成模型检测精度较低, 出现漏检和误检现象的问题, 本文使用 YOLOv7 网络进行改进, 改进后的网络模型如图 2 所示, 其中虚线框起来的部分即改进的模块。

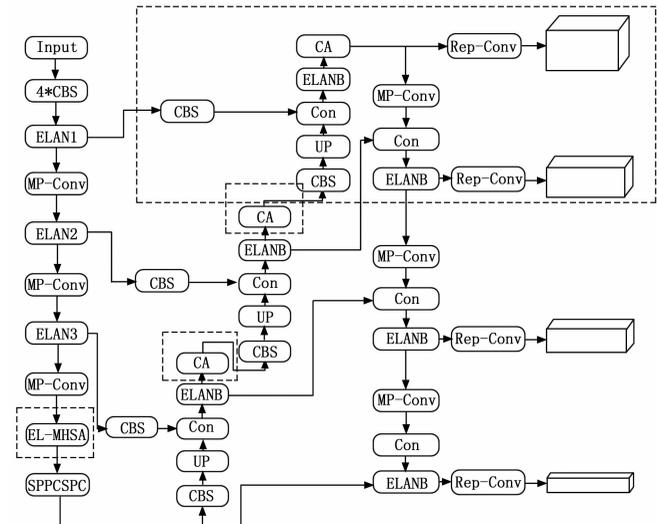


图 2 改进后的网络结构图

2.2 增加浅层网络检测层

原 YOLOv7 模型通过 8 倍、16 倍和 32 倍下采样的方式获取 3 种不同大小的特征图，其尺寸分别为 80×80 、 40×40 、 20×20 ，并将他们输入到特征融合网络中。在特征提取过程中，浅层网络更关注目标的细节信息，例如目标的纹理和边缘特征^[14]；而深层网络更侧重于提取目标的语义信息。由于特征图经过多次卷积和池化后，大部分远景小目标特征信息会丢失，通过原始网络的检测头很难检测到这些信息。

交通道路场景下存在大量小尺度目标和密集遮挡目标，而这些目标特征信息大多存在于网络更浅层，因此需要扩大网络检测尺度范围，检测出更小尺度目标，提高模型的检测性能。本文在 YOLOv7 模型上通过浅层网络 P2 增加了一个新的检测头，模型结构如图 3 所示。该检测头通过主干网络将输入图像经过 4 倍下采样操作而得到，具有 160×160 的分辨率，该特征含有待检测目标更丰富的细节特征信息，可以增强模型对小目标特征的提取能力。尽管添加小目标检测头增加了模型的计算量和内存开销，但是有助于增强微小目标的检测能力。

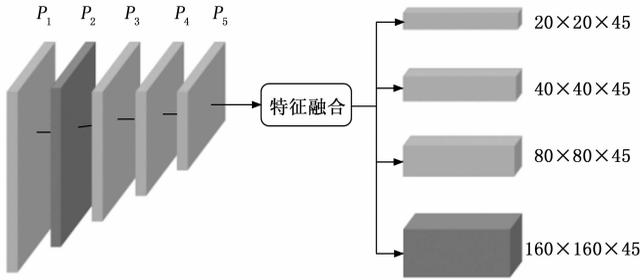


图 3 添加浅层网络检测层结构图

2.3 EL-MHSA 模块

由于交通目标检测的图像往往包含着许多复杂背景或者目标之间存在遮挡的问题，其图像经过多次卷积层提取特征后，存在许多非感兴趣区域信息，这些信息会干扰模型对目标的检测，从而影响了检测效果。为了更好地提取交通检测目标的关键信息并弱化无关背景的干扰^[15]，本文对 YOLOv7 网络中的 ELAN 模块进行改进，将多头自注意力机制融入 ELAN 模块中，如图 4 所示。具体来说是将 ELAN 模块中的要使用 Cat 进行融合的两个 3×3 卷积改为多头自注意力机制，可以在目标检测中建立长距离依赖关键，而且对于路测图像检测表现出更好的效果，解决了路测图像中因无效信息引起检测效果不佳的问题，能更高效地提取图像特征信息。在特征层间添加全局注意力虽然可以有效增加目标检测的精度，但是也会使模型带来额外的计算量和内存开销，因此本文只在主干网络 Backbone 中以最低分辨率的 ELAN 层（即 ELAN4 模块）进行多头注意力的融合。

多头注意力机^[16]是一种全局操作，在网络中融入多头注意力机制可以有效地获取特征图中目标周围的上下文信息，增强模型全局信息的关注能力。多头注意力机制的结

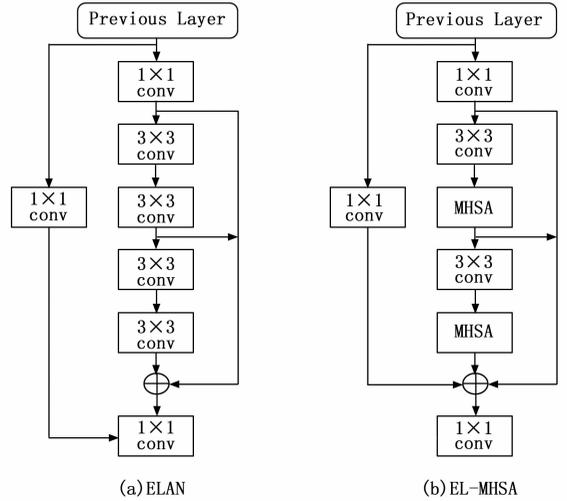


图 4 改进的 EL-MHSA 模块

构如图 5 所示。MHSA 由多个自注意力模块组成，每个 Self-Attention 模块在不同空间上捕捉全局特征信息。

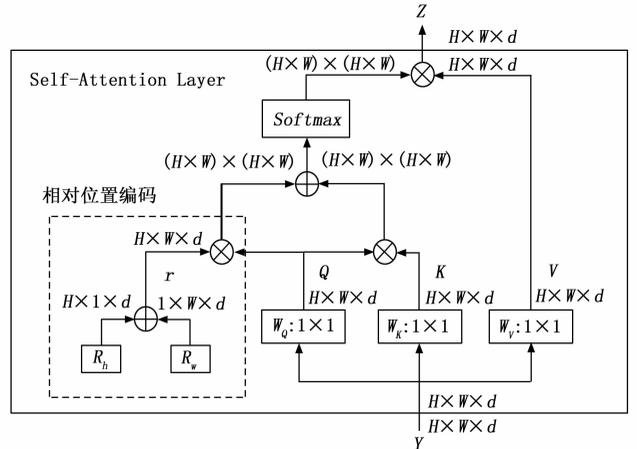


图 5 多头注意力机制结构图

Q 、 K 、 V 分别为 Query、Key、Value 矩阵， 1×1 表示逐点卷积， W_q 、 W_k 、 W_v 为注意力权重，整个过程中所学习并更新的就是注意力权重， R_w 和 R_k 分别为像素 a 到 b 之间宽、高的相对位置编码信息。使用多头注意力机制可以将特征图的全局信息与 CNN 卷积的局部信息相结合，从而形成参数量较少且能力更强的特征图^[17]。

2.4 协调注意力机制 (CA)

由于待检测的交通目标尺寸差异颇大且分布不均匀，使用原网络来检测时会出现漏检误检的情况，因此在网络中引入协调注意这种高效的注意力机制，让模型更加关注自己感兴趣的部分，增强网络对检测目标的提取能力。

CA 注意力机制相比较传统的 SE 注意力机制，它将位置信息嵌入到通道注意力机制中，避免使用 2 维全局池化造成的位置信息的丢失^[18]。CA 注意力机制还是一种轻量化的模型，它可以灵活地插入任何的网路，还不会产生大量的计算开销。

CA 的结构如图 6 所示，其中 $X_{Avg Pool}$ 和 $Y_{Avg Pool}$ 分别表示

沿着水平和垂直方向进行一维全局池化, 进行各个通道间特征的聚合, 使网络结构可以获得更多的感受野。

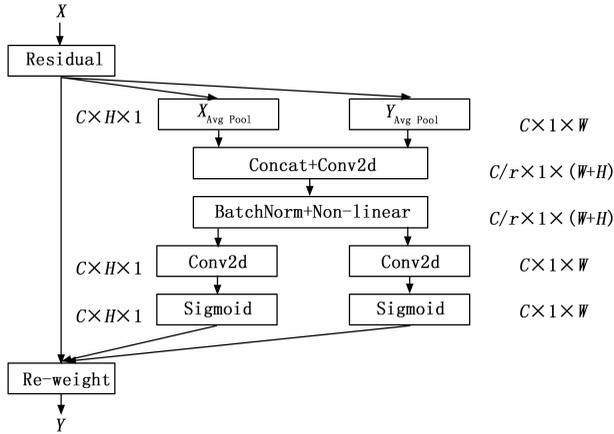


图 6 CA 注意力机制结构图

此过程可以表示为:

$$Z_c^w(\omega) = \frac{1}{H} \sum_{0 \leq j < H} x_c(j, \omega) \quad (1)$$

$$Z_c^h(\omega) = \frac{1}{W} \sum_{0 \leq i < W} x_c(h, i) \quad (2)$$

其中: H 、 W 为特征图的高和宽; $x_c(h, i)$ 为输入特征在第 c 个通道位置 (h, i) 处的值; $Z_c^h(\omega)$ 和 $Z_c^w(\omega)$ 指的是经过 W 、 H 方向池化后在第 c 个通道高为 h 时的值; 接着将得到的信道输出结果送入共享权值的 1×1 卷积进行归一化、激活操作, 生成一个中间特征图 f , 输出结果如公式 (3) 所示:

$$f = \delta\{F_1([\mathbf{z}^h, \mathbf{z}^w])\} \quad (3)$$

其中: $[\cdot, \cdot]$ 表示在某一空间维度进行拼接操作。然后将沿着空间维度拆分的两个张量 f^h 和 f^w , 利用 1×1 卷积、 F_h 、 F_w 将其变为具有相同特征的张量, 得到结果如下所示:

$$g^h = \delta[F_h(f^h)] \quad (4)$$

$$g^w = \delta[F_w(f^w)] \quad (5)$$

最后, 将得到的结果与输入特征图相乘, 得到 CA 输出结果如式 (6) 所示:

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j) \quad (6)$$

通过向 YOLOv7 网络中插入 CA 模块, 可以使模型更加关注自己感兴趣的区域, 削弱复杂背景下无关信息的干扰, 从而提高模型的检测效果。

2.5 损失函数的改进

YOLOv7 网络的损失函数同 YOLOv5 网络相同, 都是由置信度损失、分类损失和坐标损失这 3 部分组成。YOLOv7 网络的坐标损失采用 Ciou 来计算, 是在先前的损失函数的基础上增加了对宽高比之间的考虑, 但是并没有考虑预测框和真实框之间的角度和比例, 即真实框与预测框之间的方向不匹配的问题。为此, Gevorgyan 等人在 Ciou 的基础上提出了 SIoU 损失函数算法^[19], 它不仅聚合了以往损失函数的优点, 而且还考虑了真实框与预测框之间的矢量角, 并重新定义了损失函数。

简单来说, SIoU 损失函数更加全面细致地考虑了目标检测中各个因素的影响, 使得模型更加准确地学习目标特征和位置信息, 从而提高检测精度。因此, 本文采用性能更加高效的 SIoU 来代替模型的 Ciou 损失函数, 增加了网络损失函数的收敛性, 从而提高了模型的精度。SIoU 损失函数由 4 个 Cost 函数组成: 即角度损失、距离损失、形状损失和 IoU 损失。

角度损失是指在目标检测任务中, 用于度量预测框与真实框之间角度差异的一种损失函数。其计算方式通常采用如余弦距离等方法, 通过将角度误差映射为连续值来进行优化。角度损失的引入可以进一步提升目标检测模型对于目标位置和方向的识别能力, 提高模型的精度和鲁棒性。角度损失定义为:

$$\tau = 1 - 2 \times \sin^2 \left[\arcsin(x) - \frac{\pi}{4} \right] \quad (7)$$

为了进一步提高目标检测模型的精度和可靠性, 角度成本计算被引入到距离成本中。重新定义的距离成本不仅考虑了预测框与真实框之间的位置差异, 还融合了角度误差的测量, 通过同时优化位置和方向的损失函数来学习目标检测任务中的特征表示。这种重新定义的距离成本相对于传统的只考虑位置的距离计算方式, 能够更准确地对目标位置和方向进行建模, 并在目标尺度和旋转角度变化较大的情况下表现出更好的鲁棒性和泛化能力。角度成本计算被应用为距离成本, 为此重新定义距离成本如下:

$$\Delta = \sum_{t=x,y} [1 - e^{-(t-r)^2}] \quad (8)$$

距离损耗是目标检测中用于度量预测框和真实框之间距离差异的一种损失函数。它通常采用均方误差、IoU 等方式来计算。在距离损耗的优化过程中, 除了考虑位置信息, 还需要关注目标的尺度和旋转角度等其他关键特征。距离损耗定义为:

$$\gamma = \sum_{t=w,h} (1 - e^{-t^2})^q \quad (9)$$

由以上分析可知, 在目标检测任务中, 综合考虑位置、尺度和旋转角度等关键特征的 SIoU_{Loss} 损失函数公式可以表示为:

$$SIoU_{Loss} = 1 - IoU + \frac{\Delta + \gamma}{2} \quad (10)$$

3 实验结果及分析

3.1 实验环境

本文实验环境在 Linux 系统中运行, 利用两块内存为 32 G 的 v100 GPU 进行训练, Python 版本为 3.8, CUDA 版本为 10.2, 深度学习框架为 Pytorch1.8.0。

模型训练参数设置: 迭代次数 $epoch=100$, $batch_size=8$, 学习率选取 0.01, 权重衰减系数为 0.000 5, 动量为 0.937, 输入图片采用 640×640 的分辨率, 其余均采用默认设置。

3.2 数据集

本文选用由伯克利 AI 实验室发布的公开交通场景数据集 BDD 100 K 中的 10 K 来验证改进的 YOLOv7 模型在复杂交通情况下的有效性和优越性。该数据集是在 6 种不同

天气、场景和白天、黄昏、夜晚共 3 种不同时间段的所采集的图^[20]，且该数据集有 Person、Rider、Car、Bus、Truck、Bike、Motor、Traffic light、Traffic sign、Train 这 10 个感兴趣的对象类别，其各类别标签数量如图 7 所示，该数据集基本上涵盖了交通检测中所出现的所有基本目标。

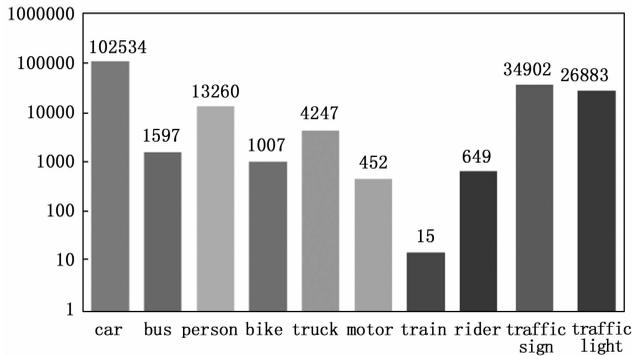


图 7 BDD100K 数据集各类标签分布图

在本次实验中，本文按照 8 : 1 : 1 的比例对数据集进行了划分。其中，训练集包含 8 000 张图像，验证集包含 1 000 张图像，测试集包含 1 000 张图像。这些数据集的使用，旨在评估和改进 YOLOv7 模型在交通场景下的表现，以期提高交通检测的准确度和效率。

3.3 评价指标

本文的性能评价主要采用以下 4 个指标：检测精度 (AP)、平均检测精度 (mAP)、准确率-召回率 (P-R) 曲线以及检测速度 (FPS)。相关指标的计算公式如下所示，旨在综合评估本文提出的基于 YOLOv7 的交通目标检测模型在不同方面的表现：

$$R = \frac{T_p}{T_p + F_p} \quad (11)$$

$$P = \frac{T_p}{T_p + F_n} \quad (12)$$

$$AP = \int_0^1 P(r) dr \quad (13)$$

$$mAP = \frac{\sum_{n=0}^N AP_n}{N} \quad (14)$$

式中， T_p 为模型预测正确的正样本数， F_p 为模型误检的正样本数， F_n 为模型漏检的正样本数。需要注意的是，由于检测结果是基于一定置信度阈值得出的，因此在计算 AP 时需要针对不同置信度下的结果进行平均化处理。具体而言，需要计算多个不同置信度阈值下的 AP 值，然后求其平均

值，得到最终的 mAP 指标。

3.4 消融实验

为了更好地验证本文所提出的改进模型与方法的有效性，使用 BDD00K 数据集进行消融实验。不同的改进策略使得模型的结构、检测精度、速度、参数量也进行变化，具体的实验结果如表 1 所示。

模型 1 为原始的 YOLOv7 算法的实验结果，将本次结果作为后面几组实验的对比基准，检测的 $mAP@0.5$ 为 57.3%， $mAP@0.5 : 0.95$ 为 31.9%，检测速度是最快的。模型 2 在原网络结构的基础上引入 EL-MHSA 模块，使 BDD100K 数据集 $mAP@0.5$ 和 $mAP@0.5 : 0.95$ 分别提升了 1.4% 和 0.9%，提升了网络的整体性能。模块 3 在颈部添加了 CA 协调注意力，挖掘各个通道间的相互依赖关系，增强对目标特征的提取能力，使 $mAP@0.5$ 和 $mAP@0.5 : 0.95$ 分别提升了 0.6% 和 0.2%。模块 4 通过增加一个小目标检测层保留了更多浅层细节信息，使模型的 $mAP@0.5$ 和 $mAP@0.5 : 0.95$ 分别提升了 2.3% 和 1.0%。模块 5 将损失函数改为 SIoU， $mAP@0.5$ 和 $mAP@0.5 : 0.95$ 分别提升了 1.6% 和 0.4%，虽然提升的效果较小，但是在模型收敛方面有着较好的结果。

通过上述实验，发现模块 4 添加小检测头与模块 6 整体改进的实验数据相差不大，这是因为将上述所有模块融入网络中后，模型参数量增加，网络结构相比单加模块的方式更为复杂，影响了检测精度。总的来说，每种改进的策略都可以让模型性能有不同程度的提升。将每个模块都融入后的算法与 YOLOv7 原算法相比， $mAP@0.5$ 和 $mAP@0.5 : 0.95$ 分别提升 2.5% 和 1.4%，表明该改进算法可以促进网络对交通目标特征的学习，更易避免误检与漏检情况的发生。该算法检测速度为 96.2 FPS，比最快的 YOLOv7 降低了 10.2，但是完全可以满足检测实时性的要求。

3.5 改进算法与 YOLOv7 算法在 BDD100K 中各个类别的比较

为了验证本文算法模型的优越性，让本文改进算法和 YOLOv7 算法各项参数指标相同的情况下，将两个算法在 BDD100K 数据集中训练的各个类别的 AP 值进行对比，结果如图 8 所示。从图中明显可以看到本文改进算法的每个类别的平均精度都高于原始算法。其中，car 的 AP 值从 80.5% 到 84.1%，提升了 3.6 个百分点；bus 的 AP 值从 66.4% 到 68.6%，提高了 2.2 个百分点；person 的 AP 值从 68.2% 到 71.1%，提高了 1.9 个百分点；bike 的 AP 值从 50.8% 到 52.9%，提高了 2.1 个百分点；truck 的 AP 值

表 1 算法在 BDD100K 数据集上的消融实验结果

Group	EL-MHSA	CA	Head	SIoU	P/%	R/%	$mAP@0.5$ /%	$mAP@0.5:0.95$ /%	Params/M	FPS
1					78.2	52.8	57.3	31.9	37.2	106.4
2	√				73.9	54.9	58.7	32.8	36.8	110
3		√			77.6	53.1	57.9	32.1	37.7	98
4			√		76.0	53.6	59.6	32.9	37.5	102
5				√	76.8	52.9	58.9	32.3	37.2	105
6	√	√	√	√	76.0	53.9	59.8	33.0	37.6	96.2

从 67.9% 到 69.1%，提高了 1.2 个百分点；motor 的 AP 值从 49.1% 到 53.5%，提高了 4.4 个百分点；rider 的 AP 值从 50.1% 到 52.4%，提高了 2.3 个百分点；traffic sign 的 AP 值从 73.6% 到 74.7%，提升了 0.9 个百分点；traffic light 的 AP 值从 68% 到 71.6%，提升了 3.6 个百分点；尤其是 train 目标，由于检测物体标注过于少，原算法没有检测出这个物体，而改进的算法检测出精度为 1.8%。从以上数据可以得出，本文改进的算法相比于 YOLOv7 原算法表现出更为优越的性能，对交通道路场景下检测任务具有更可观的检测效果。

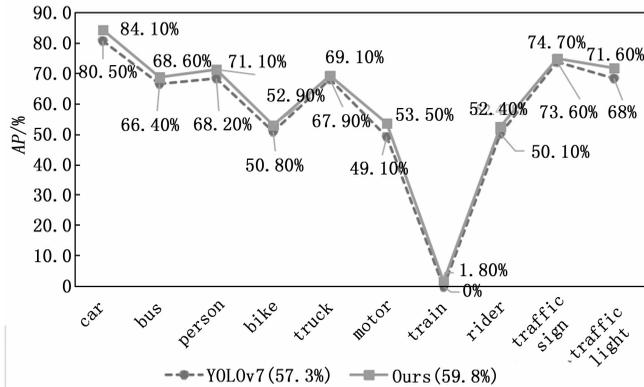


图 8 改进算法与 YOLOv7 算法在 BDD100K 中 10 个类别比较

为了更加直观地展示出本文改进算法在真实场景下的效果，在 BDD100K 数据集的测试集中随机选择了多张不同交通场景下的图片进行对比分析，如图 9 所示。其中，左侧为本文改进算法的结果图，右侧为 YOLOv7 算法的结果图。

图 9 的 (a) 组图片检测的是背景比较复杂的场景，右图将车子上面的梯子误检成交通灯，左图并没有出现误检的情况；(b) 组图片是在夜间图像像素模糊情况下进行检测的结果，右图将墙的边沿检测成了交通标志，而左图正确地检测出目标；(c) 组图片是在黄昏时光照不均匀的情况下的检测结果，明显可以看出，右图由于光线昏暗且树木遮掩，没有识别出车辆，而左图准确地检测出被遮掩的车辆；(d) 组图片在目标交叉密集时，左图明显比右图的精度检测高。总的来说，本文改进的算法相对于 YOLOv7 算法表现出了更好的性能，一定程度上改善了漏检和误检问题，可以更好地适应于交通场景下的检测任务。

3.6 对比试验

3.6.1 本文算法与其他同类改进算法对比试验

本实验将本文改进的算法与其它采用 BDD100K 公共数据集且应用于交通目标检测的改进算法的性能进行比较，如表 2 所示。本文改进的算法相比 AFFB_YOLOv5s^[20] 和 M-YOLO^[21] 算法的检测精度分别高了 8.3% 与 3.0%，检测速度分别高了 20.3 和 33.2。可见，本文改进的算法相较于其他改进的算法表现出更为显著的优势，可以更好地解决交通目标检测中出现漏检与误检问题的情况。

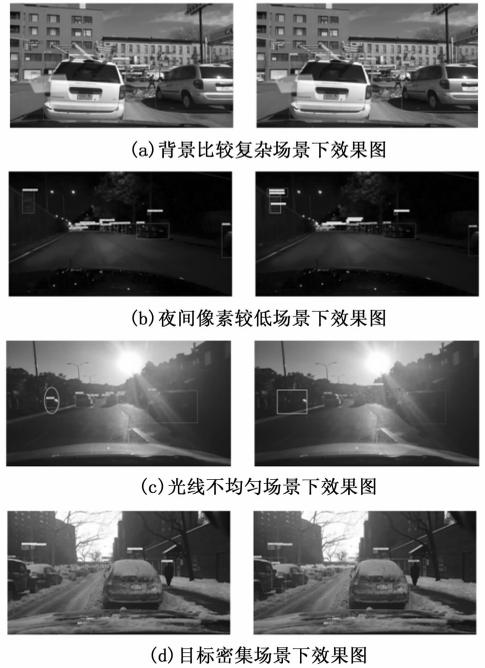


图 9 不同交通场景下模型检测对比结果

表 2 本文算法与同类改进算法对比结果

Models	P/%	R/%	mAP@0.5%	FPS
AFFB_YOLOv5s	33.0	58.3	51.5	63
M-YOLO	34.5	76.2	56.8	83.3
Ours	76.0	53.9	59.8	96.2

3.6.2 本文算法与其他主流算法对比实验

本实验将改进的算法与 Faster R-CNN、SSD、YOLOv3^[22]、YOLOv4^[24]、YOLOv5 和 YOLOv7 算法在同等条件下进行实验对比，实验结果如表 3 所示。

本文通过改进算法，针对现有目标检测算法在实时性和精度之间的权衡问题，提出了一种新的解决方案。实验结果表明，Faster R-CNN 算法的 mAP@0.5 值为 30.7%、FPS 值为 14.2，检测速度太低，无法满足实时性的需求；SSD 算法的 mAP@0.5 值为 28.1%、FPS 值为 39，相比 Faster R-CNN，在提高速度方面有提升，但是检测精度下降了 2.6 个百分点；而 YOLOv3、YOLOv4 和 YOLOv5 算法在检测速度上都有很大的提升，但其检测精度仍然存在局限。

表 3 不同算法在 BDD100K 数据集的性能对比结果

算法	输入尺寸	mAP@0.5/%	mAP@0.5 : 0.95/%	FPS
Faster R-CNN	640×640	30.7	—	14.2
YOLOv3	640×640	40.8	21.4	80
YOLOv4	640×640	45.3	22.7	84
SSD	640×640	28.1	—	39
YOLOv5	640×640	51.9	28.9	92
YOLOv7	640×640	57.3	31.9	106.4
Ours	640×640	59.8	33.0	96.2

相比之下, 本文改进算法在速度和精度方面达到了一个良好的平衡。虽然相较于原算法, 速度方面有所降低, 但 96.2 FPS 的速度仍然是除原算法之外最优的, 完全可以满足实时检测任务的需求。在精度方面, 本文算法在 $mAP@0.5$ 和 $mAP@0.5:0.95$ 等多个评价指标上排列第一, 明显优于其他主流算法。因此, 本文算法可以在实时性和精度之间取得一个更好的平衡, 为目标检测任务提供了更加高效和准确的解决方案。

4 结束语

本文在 YOLOv7 网络的基础上进行改进, 并将改进算法应用于交通场景的目标检测中。通过在 YOLOv7 原网络中融入全局注意力机制, 设计一个新的 EL-MHSA 模块, 获取全局结构化信息; 通过在颈部网络加入 CA 协调注意力机制, 使其更好地提取细节特征; 通过添加一个浅层网络检测头, 进一步提取小目标特征信息; 通过改变网络损失函数, 减轻了模型收敛不稳定问题。在 BDD100K 数据集上的实验结果表明, 该改进算法检测结果优于其它主流目标检测模型, 检测精度在达到 59.8% mAP 的同时, 检测速度为 96.2 FPS, 该算法可以在保证检测速度的同时还可以提高算法精度, 在一定程度上减轻了误检和漏检问题, 能够有效地完成交通场景下的目标检测任务。但是, 由于模型中添加检测头和注意力机制等方法使得模型复杂度上升, 参数量变大, 检测速度稍有下降, 后期可以考虑采用更加高级的特征提取算法和更加先进的目标检测网络结构, 以进一步提高模型的检测精度和速度。此外, 还可以探索基于深度学习的模型量化和剪枝技术, 以减小模型的计算和存储开销, 从而提高模型在边缘设备上的应用效果。

参考文献:

- [1] 来飞, 黄超群, 胡博. 智能汽车自动驾驶技术的发展与挑战 [J]. 西南大学学报 (自然科学版), 2019, 41 (8): 124-133.
- [2] 段仲静, 李少波, 胡建军, 等. 深度学习目标检测方法及其主流框架综述 [J]. 激光与光电子学进展, 2020, 57 (12): 59-74.
- [3] 张云川, 姜麟, 林莉. 基于单次双向特征金字塔网络的目标检测模型 [J]. 激光与光电子学进展, 2023, 60 (2): 273-283.
- [4] ROSS G. Fast R-CNN [C] //2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 2015: 1440-1448.
- [5] REN S Q, HE K M, ROSS G, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39 (6): 1137-1149.
- [6] LIU W, ANGELOV D, ERHAN D, et al. SSD: single shot multibox detector [C] //Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 2016, Proceedings, Part I 14. Springer International Publishing, 2016: 21-37.
- [7] ZHANG Y, SHEN Y L, ZHANG J. An improved tinyyolov3 pedestrian detection algorithm [J]. Optik, 2019, 183 (4): 17-23.
- [8] ZHU J X, LI X, JIN P, et al. MME-YOLO: multi-sensor multi-level enhanced YOLO for robust vehicle detection in traffic surveillance [J]. Sensors, 2020, 21 (1): 27-29.
- [9] LI Y, LI J G, MENG P. Attention-YOLOV4: a real-time and high-accurate traffic sign detection algorithm [J]. Multimedia Tools and Applications, 2022, 82 (5): 7567-7582.
- [10] ZHU D, XU G, ZHOU J, et al. Object detection in complex road scenarios: improved YOLOv4-tiny algorithm [C] //2021 2nd Information Communication Technologies Conference (ICTC). IEEE, 2021: 75-80.
- [11] USHA M, PRIYANKA C, RAJEEV T. EnsembleNet: a hybrid approach for vehicle detection and estimation of traffic density based on faster R-CNN and YOLO models [J]. Neural Computing and Applications, 2022, 35 (6): 4755-4774.
- [12] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 7464-7475.
- [13] 丁松涛, 曲仕茹. 基于深度学习的交通目标感兴趣区域检测 [J]. 中国公路学报, 2018, 31 (9): 167-174.
- [14] 张婷婷, 章坚武, 郭春生, 等. 基于深度学习的图像目标检测算法综述 [J]. 电信科学, 2020, 36 (7): 92-106.
- [15] 邬军, 邓月明, 何鑫, 等. 复杂背景下的移动目标检测算法 [J]. 现代电子技术, 2022, 45 (19): 59-65.
- [16] 齐向明, 董旭. 改进 Yolov7-tiny 的钢材表面缺陷检测算法 [J]. 计算机工程与应用, 2023, 59 (12): 176-183.
- [17] 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.
- [18] 朱高兴, 于璪. 基于 YOLOv5-CA 算法的野生动物目标检测研究 [J]. 信息技术与信息化, 2022 (6): 32-35.
- [19] GEVORGYAN Z. SiO loss: more powerful learning for bounding box regression [J]. Arxiv Preprint Arxiv: 2205.12740, 2022.
- [20] 张凯祥, 朱明. 基于 YOLOv5 的多任务自动驾驶环境感知算法 [J]. 计算机系统应用, 2022, 31 (9): 226-232.
- [21] LIAN J, YIN Y H, LI L H, et al. Small object detection in traffic scenes based on attention feature fusion [J]. Sensors, 2021, 21 (9): 3031-3031.
- [22] 胡晓兵, 罗耀俊. 基于 M-YOLO 的自动驾驶下目标识别算法 [J]. 计算机工程与设计, 2022, 43 (8): 2213-2220.
- [23] REDMON J, FARHADI A. Yolov3: An incremental improvement [EB/OL]. (2023-02-26) [2019-04-15]. <https://arxiv.org/abs/1084.02767>.
- [24] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: optimal speed and accuracy of object detection [EB/OL]. (2023-02-25) [2022-10-02]. <https://arxiv.org/abs/2004.10934>.