

# 基于改进 YOLOv7-tiny 的轻量化道路目标检测算法\*

何泽江 蒋淑霞 柳霞

(中南林业科技大学机械与智能制造学院,长沙 410004)

**【摘要】**针对目标检测算法对算力和存储空间的高要求限制其在边缘设备中检测功能实时性的问题,提出了一种基于 YOLOv7-tiny 改进的轻量化道路目标检测算法。首先,通过 K-means++ 聚类算法生成适合道路目标检测的先验锚框;其次,改进 ELAN 结构轻量化主干网络,同时提出轻量型多尺度特征(LMS)模块优化颈部网络;最后,使用西格玛线性单元(SiLU)激活函数加速模型收敛,采用 MPDIoU 损失函数进一步提高检测精度。试验结果表明:改进后的模型参数量减少 18.3%,计算量降低 15.0%,且所有类别平均检测精度上升 1.1%。在 Jetson TX2 中,使用 TensorRT 加速后的检测速度达到 48 帧/s,基本满足道路目标检测的实时性要求。

**关键词:**自动驾驶 YOLOv7-tiny 道路目标检测 轻量化 Jetson TX2

中图分类号:TP391.4 文献标志码:A DOI: 10.19620/j.cnki.1000-3703.20231116

## An Improved Lightweight Road Target Detection Algorithm Based on YOLOv7-tiny

He Zejiang, Jiang Shuxia, Liu Xia

(College of Mechanical and Intelligent Manufacturing, Central South University of Forestry and Technology, Changsha 410004)

**【Abstract】**To address the high computational and storage demands of object detection algorithms that limit real-time performance of edge devices, this study proposes an improved lightweight road object detection algorithm based on YOLOv7-tiny. First, prior anchor boxes optimized for road object detection are generated using the K-means++ clustering algorithm. Secondly, the backbone network is streamlined by modifying the ELAN structure, while a Lightweight Multi-scale Feature (LMS) module is introduced to optimize the neck network. Finally, the Sigmoid Linear Unit (SiLU) activation function is adopted to accelerate model convergence, and the MPDIoU loss function is employed to further improve detection accuracy. Experimental results demonstrate that the improved model achieves an 18.3% reduction in parameters, a 15.0% decrease in computational complexity, and a 1.1% increase in mean average precision across all categories. When deployed on Jetson TX2 with TensorRT acceleration, the detection speed reaches 48 frames per second, essentially meeting real-time requirements for road object detection applications.

**Key words:** Autonomous driving, YOLOv7-tiny, Road target detection, Lightweight, Jetson TX2

**【引用格式】**何泽江,蒋淑霞,柳霞.基于改进 YOLOv7-tiny 的轻量化道路目标检测算法[J].汽车技术,2025(2):9-16.

HE Z J, JIANG S X, LIU X. An Improved Lightweight Road Target Detection Algorithm Based on YOLOv7-tiny[J]. Automobile Technology, 2025(2): 9-16.

## 1 前言

道路目标检测作为自动驾驶的核心技术,对车辆的环境感知和决策至关重要。然而,面对道路交通场景的复杂性和多变性,在保证检测精度的同时实现高效计算,仍是当前研究面临的重要挑战。

基于深度学习的单阶段目标检测技术,通过加入残

差网络<sup>[1]</sup>、检测层<sup>[2]</sup>和注意力机制<sup>[3-4]</sup>等方式,虽然能够增强网络的特征提取能力,但计算量和网络复杂度大幅上升,导致很难注入嵌入式设备。为此,可通过特征金字塔网络(Feature Pyramid Network, FPN)结构的轻量化,减少特征图的语义损失<sup>[5-6]</sup>,但主要用于缓解小目标像素稀疏。神经网络搜索可用于设计轻量化的网络结构<sup>[7]</sup>,但搜索成本巨大,且复杂网络结构的可解释性较

\*基金项目:教育部产学合作协同育人项目(230826311807159);湖南省自然科学基金项目(2022JJ50012);湖南省研究生科研创新项目(CX20220735)。

差。在网络中融入轻量化的高效网络,如MobilenetV3、ShufflenetV2和EfficientNetV2等<sup>[8-10]</sup>,虽然能够降低计算开销,但直接使用容易导致模型检测精度大幅下降,而且需要重新训练模型权重,时间成本较高。

为了能够在嵌入式平台或移动设备中部署,本文使用K-means++聚类算法生成适合道路目标检测的锚框;通过带瓶颈结构的高效远程注意力网络(Efficient Long-range Attention Network with Bottleneck, B-ELAN)和轻量化的多尺度特征(Lightweight Multi-Scale, LMS)模块优化网络结构;使用西格玛线性单元(Sigmoid Linear Unit, SiLU)激活函数来加速模型收敛,采用最小点距离交并比(Minimum Point Distance based Intersection over Union, MPDIoU)损失函数<sup>[11]</sup>增强预测锚框定位的准确性,进一步提高检测精度。同时,在Jetson TX2平台中验证本文模型的实时性。

## 2 改进的YOLOv7-tiny算法

按照网络的深度、宽度不同,YOLOv7共6个版本<sup>[12]</sup>,网络结构主要由输入端(Input)、主干网络(Backbone)、颈部网络(Neck)以及检测头(Head)组成。其中,输入端进行数据增强及图像的预处理,将图像缩放为640像素×640像素,送入主干网络;主干网络进行8倍、16倍及32倍下采样的特征提取,送入颈部网络进行特征融合;颈部网络采用路径聚合网络(Path Aggregation Network, PAN)和FPN的结构,使送入检测头前的特征同时拥有语义和定位信息;检测头对特征图进行多尺度目标检测和输出,从而预测物体的位置和类别。

YOLOv7-tiny作为同系列中最小版本,可保持较高的检测精度,但通过控制模型深度和宽度降低开销,对于小型边缘计算设备,计算成本依然很高,不适用于具有实时要求的道路目标检测任务。因此,需在提高检测精度的同时,对网络进行轻量化改进。

改进后网络结构如图1所示,其中,主干网络使用轻量化的B-ELAN;颈部网络使用结合跨阶段部分卷积的空间金字塔池化(Spatial Pyramid Pooling with Cross Stage Partial Convolutions, SPPCSPC)模块,利用轻量化多尺度模块替换颈部网络中的最后一个ELAN模块;CBS模块由卷积(Convolution)层、批归一化(Batch Normalisation)层和带泄漏的线性修正单元(Leaky Rectified Linear Unit, Leaky ReLU)激活函数组成,将Leaky ReLU激活函数更换为SiLU激活函数,生成CBS模块。

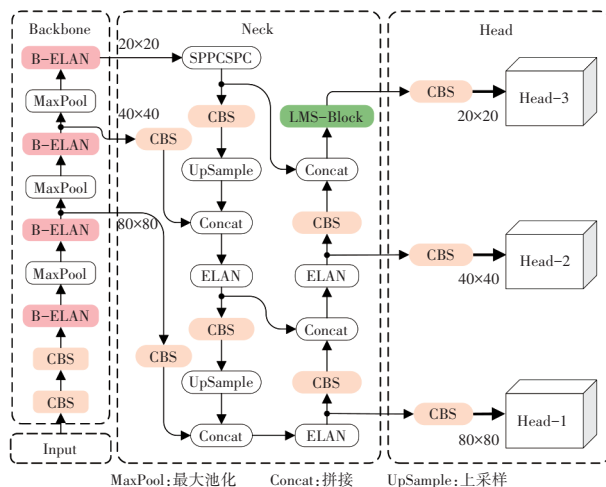


图1 改进的YOLOv7-tiny网络结构

### 2.1 先验锚框采集

YOLOv7-tiny使用自动锚定算法获得MS-COCO数据集的先验锚框,该算法由K-means聚类获得初始锚框和锚框进化部分。

K-means算法一次选择多个聚类中心,而K-means++算法只选择一个聚类中心,且克服了K-means随机选择的限制,从而提高了收敛性,更有可能实现全局最优解。为了获得适合道路目标检测的先验锚框,本文采用BDD100K数据集<sup>[13]</sup>,使用K-means++替代K-means聚类算法,锚框进化部分保持不变。K-means++算法获得初始锚框的主要流程为:

a. 采集BDD100K数据集中全部标注框的宽度、高度信息,作为聚类数据集 $X$ ,从 $X$ 中随机选择一个数据点作为初始质心 $c_1$ 。

b. 使用 $P(x_i) = \frac{D(x)^2}{\sum_{x \in XD(x)^2}$ 从样本中选择下一个质心 $c_i$ ,其中, $D(x)$ 为数据点 $x$ 与已选最近质心的最短距离。

c. 重复步骤b,直到选择了 $k$ 个质心,即不同尺寸锚框的个数。YOLOv7-tiny有3个不同尺寸特征的检测头,每个检测头对应3个锚框,所以 $k=9$ 。

d. 对于每个 $i \in \{1, 2, \dots, k\}$ ,形成聚类族 $C_i$ 。如果 $X$ 中某个数据点距离 $c_i$ 比 $c_j (j \neq i)$ 更近,则该数据点属于聚类族 $C$ 。

e. 对于每个 $i \in \{1, 2, \dots, k\}$ ,假设 $c_i$ 为聚类族 $C_i$ 的质心,则 $c_i = \frac{1}{|C_i|} \sum_{x \in C_i} x$ 。

f. 重复步骤d~步骤e,直到 $C \in \{C_1, C_2, \dots, C_k\}$ 不再变化。

通过K-means++算法生成的初始锚框分布更均匀,并与数据集中对象的特征对齐,有助于后续的锚框进化

得到更适合的先验锚框。

## 2.2 B-ELAN模块轻量化

经检测分析,YOLOv7-tiny主干网络中参数量和运算量主要集中在高效远程注意力网络(Efficient Long-range Attention Network, ELAN)模块,其中,2个 $3\times 3$ 标准卷积的占比超过60%。若使用深度可分离卷积替换标准卷积,虽然可以最大程度轻量化,但模型的检测性能会大幅降低。为此,轻量化ELAN模块结构设计如图2所示,改进部位使用2层瓶颈结构(Bottleneck), $1\times 1$ 卷积降低通道数后,使用 $3\times 3$ 卷积进行通道恢复。假设输入、输出特征通道相同,在不考虑偏置项的情况下,相较于标准卷积,该瓶颈结构能够减少4/9的参数量和计算量。

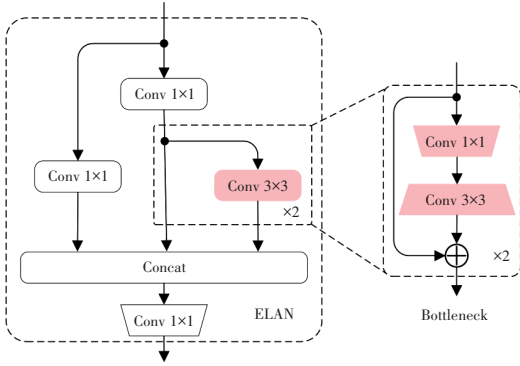


图2 B-ELAN结构

## 2.3 多尺度模块设计

网络中多尺度特征提取对于提升目标检测性能非常重要,模块结构如图3所示。Res2Net<sup>[14]</sup>(见图3a)在标准残差瓶颈结构基础上进行多尺度特征增强,首先对输入特征通过 $1\times 1$ 卷积进行通道数调整,然后将特征图分割为4份,即 $X_i, i\in\{1,2,3,4\}$ 。分别送入不同分支处理, $X_1$ 直接映射,其他组均经过标准卷积处理 $K_i$ ,Res2Net结构在提升多尺度特征的表达能力的同时,大幅降低了模块的参数量和计算量,不同分支 $Y_i$ 可表示为:

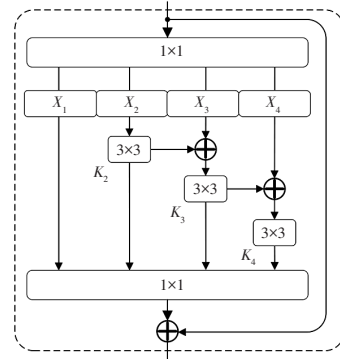
$$Y_i = \begin{cases} X_i, & i = 1 \\ K_i(X_i), & i = 2 \\ K_i(X_i + Y_{i-1}), & i = 3,4 \end{cases} \quad (1)$$

MS-Block<sup>[15]</sup>(见图3b)使用大核深度可分离卷积增大感受野,同时保持较快的推理速度。假设输入特征图 $X$ 通道数为 $C$ ,通过 $1\times 1$ 卷积将通道扩展至 $n\times C$ ,并将 $X$ 分割成4份,即 $X_i, i\in\{1,2,3,4\}$ 。 $X_1$ 直接映射,其他组均经过倒置的瓶颈结构 $IB_{k\times k}$ ,其中, $k$ 为卷积核尺寸。最后,将所有分支拼接,采用 $1\times 1$ 卷积学习编码不同尺度特征,不同分支 $Y_i$ 可表示为:

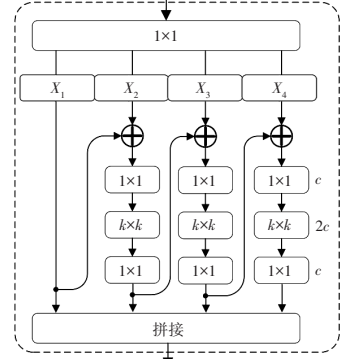
$$Y_i = \begin{cases} X_i, & i = 1 \\ IB_{k\times k}(X_i + Y_{i-1}), & 2 \leq i \leq 4 \end{cases} \quad (2)$$

综合以上2种多尺度特征模块的结构优点,提出一种轻量化LMS模块(见图3c)。输入特征 $X$ 通过 $1\times 1$ 卷积保持通道不变,随后将通道分割为4份,即 $X_i, i\in\{1,2,3,4\}$ 。其中 $X_1$ 直连,其他组在与前一组相加,再经过轻量化的瓶颈结构 $LB_i$ 。该结构由两层组成,使用 $1\times 1$ 卷积将通道扩展2倍,捕获特征的多样化信息,随后使用标准卷积将通道数减半,降低计算成本和模型复杂性。各分支 $Y_i$ 可表示为:

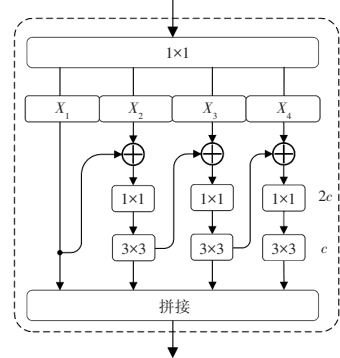
$$Y_i = \begin{cases} X_i, & i = 1 \\ LB_i(X_i + Y_{i-1}), & 2 \leq i \leq 4 \end{cases} \quad (3)$$



(a)Res2Net



(b)MS模块



(c)LMS模块

图3 多尺度模块结构

## 2.4 损失函数优化

损失函数常用来衡量模型预测值与真实值间的差异,损失函数的值越小,模型的鲁棒性越强。YOLOv7-

tiny的损失函数包括分类损失、定位损失和置信度损失。其中,定位损失指预测框与标注框间的误差,采用CIoU损失函数<sup>[16]</sup>:

$$\begin{cases} L_{\text{CIoU}} = 1 - n_{\text{CIoU}} \\ n_{\text{CIoU}} = n_{\text{IoU}} - \frac{\rho^2(\beta_{\text{gt}}, \beta_{\text{prd}})}{C^2} - \alpha V \\ n_{\text{IoU}} = \frac{\beta_{\text{gt}} \cap \beta_{\text{prd}}}{\beta_{\text{gt}} \cup \beta_{\text{prd}}}, \quad \alpha = \frac{V}{1 - \text{IoU} + V} \\ V = \frac{4}{\pi^2} \left( \arctan \frac{w^{\text{gt}}}{h^{\text{gt}}} - \arctan \frac{w^{\text{prd}}}{h^{\text{prd}}} \right)^2 \end{cases} \quad (4)$$

式中: $r$ 为两个中心点间的欧氏距离, $\beta_{\text{gt}}, \beta_{\text{prd}}$ 分别为标注框和预测框, $C$ 为标注框和预测框最小外接矩形的对角线长度, $w^{\text{gt}}, h^{\text{gt}}$ 分别为标注框的宽和高, $w^{\text{prd}}, h^{\text{prd}}$ 分别为预测框的宽和高。

CIoU解决了预测框中心点与标注框重合时产生的退化问题,但忽略了预测框和标注框具有相同的宽高比,但宽度、高度不同的情况,此时CIoU退化失效,将限制模型的收敛速度和精度。因此,考虑使用MPDIoU损失函数<sup>[11]</sup>:

$$\begin{cases} L_{\text{MPDIoU}} = 1 - n_{\text{MPDIoU}} \\ n_{\text{MPDIoU}} = n_{\text{IoU}} - \frac{d_1^2}{h^2 + w^2} - \frac{d_2^2}{h^2 + w^2} \\ d_1^2 = (x_1^{\text{prd}} - x_1^{\text{gt}})^2 + (y_1^{\text{prd}} - y_1^{\text{gt}})^2 \\ d_2^2 = (x_2^{\text{prd}} - x_2^{\text{gt}})^2 + (y_2^{\text{prd}} - y_2^{\text{gt}})^2 \end{cases} \quad (5)$$

式中: $(x_1^{\text{prd}}, y_1^{\text{prd}}), (x_2^{\text{prd}}, y_2^{\text{prd}})$ 分别为预测框左上点和右下点坐标, $(x_1^{\text{gt}}, y_1^{\text{gt}}), (x_2^{\text{gt}}, y_2^{\text{gt}})$ 分别为标注框左上点和右下点坐标, $h, w$ 分别为输入图片的高度和宽度。

MPDIoU损失基于最小化预测框左上点和标注框右下点的距离,当预测框和标注框具有相同的宽高比,但宽度和高度不同时,MPDIoU损失函数依然有效,保证了预测框回归的准确性,提高了收敛速度。

### 3 试验环境与评价指标

#### 3.1 数据集

本文使用BDD100K数据集,是目前目标检测领域规模最大、内容最具多样性的自动驾驶数据集之一,涵盖了常见类型的交通参与者。对于道路目标检测任务,共标注了10种常见道路目标,包括行人(Person)、骑行者(Rider)、乘用车(Car)、公交车(Bus)、卡车(Truck)、自行车(Bike)、摩托车(Motor)、火车(Train)、交通标志(Traffic Sign)和交通指示灯(Traffic Light)。该数据集共 $10 \times 10^4$ 张图片,按照7:1:2划分训练集、验证集和测试集。其中,训练集和验证集存在数据标注,而测试集无数据标注,故需要重新划分数据集。通过从训练集中随

机抽取20 000张图片作为新的训练集,在剩余的数据中随机抽取10 000张图片作为测试集。以上新形成的训练集、测试集,及原有的验证集,共同组成了试验的数据集。

#### 3.2 试验环境

本文试验使用英伟达GeForce RTX 3090图像处理器,显存容量为24 GB,处理器内存容量为32 GB。在深度学习PyTorch1.12.1框架中,使用Anaconda相关集成环境以及CUDA11.6加速。

在训练阶段,为了加快算法收敛,使用MS-COCO数据集预训练权重。加载权重时,通过调节权重的网络层,确保除更改网络层外的原始网络层权重全部被加载。同时,为了加速训练,将图片预先缓存到显存。

输入图像大小为640像素×640像素,使用默认比例的混合和马赛克技术进行数据增强,批大小为16。模型迭代训练为100次,采用梯度随机下降优化器,初始学习率为0.01,动量和权重衰减系数分别为0.937和0.0005。

#### 3.3 评价指标

将精确率(Precision) $P$ 、召回率(Recall) $R$ 、交并比为0.5时所有类别平均精度(mean Average Precision, mAP) mAP50作为模型性能的评价指标。其中, $P$ 为所有检测框中被正确检测的比例; $R$ 为所有检测框占全部标注框的比例。各指标计算公式为:

$$\begin{cases} P = \frac{TP}{TP + FP} \\ R = \frac{TP}{TP + FN} \\ A_{\text{AP}} = \int_0^1 P(R) dR \\ A_{\text{mAP}} = \frac{1}{n} \sum_{i=1}^n A_{\text{AP},i} \end{cases} \quad (6)$$

式中: $TP$ 为正确检测框的数量, $FP$ 为误检框的数量, $FN$ 为漏检框的数量, $A_{\text{AP}}$ 为平均精度, $A_{\text{mAP}}$ 为均值平均精度。

此外,使用模型的存储大小、参数量、浮点运算次数(Floating-Point Operations, FLOPs)以及帧率衡量模型的复杂度与计算消耗。

### 4 试验结果与分析

#### 4.1 锚框优化

分别使用K-means和K-means++聚类自动锚定算法生成3组先验锚框,同时分析对模型的影响,结果如表1所示。其中,小、中、大不同尺寸的锚框组分别对应 $80 \times 80, 40 \times 40$ 和 $20 \times 20$ 特征图,即分别检测图像中小型、中型及大型目标。

表 1 优化先验锚框结果

方法	数据集	特征层	特征图尺寸	锚框	$P$	$R$	mAP50
K-means	MS-COCO	P3	80×80	[10,13,16,30,33,23]	0.652	0.410	0.439
		P4	40×40	[30,61,62,45,59,119]			
		P5	20×20	[116,90,156,198,373,326]			
K-means	BDD100K	P3	80×80	[7,6,11,11,18,13]	0.667	0.410	0.448
		P4	40×40	[14,24,29,22,26,55]			
		P5	20×20	[48,34,89,65,170,153]			
K-means++	BDD100K	P3	80×80	[7,6,10,10,15,12]	0.684	0.405	0.451
		P4	40×40	[12,29,25,20,41,32]			
		P5	20×20	[70,49,124,86,193,149]			

由表 1 可知,由 K-means++ 自动锚定算法生成的锚框,与原始的先验锚框相比,测试集上所有类别平均检测精度 mAP50 提升 1.2%,与 K-means 的自动锚定算法在 BDD100K 生成的锚框相比,mAP50 提升 0.3%,表明在 YOLOv7-tiny 自动锚定算法中将 K-means++ 初始化聚类与锚框进化相结合,可以得到更适合道路目标检测的锚框,从而有效提高检测精度。

#### 4.2 主干网络选取

使用高效卷积替换 ELAN 结构的标准卷积,包括空间稀疏卷积<sup>[17]</sup>(SPConv)、沙漏卷积<sup>[18]</sup>(SandGlass)、部分卷积<sup>[19]</sup>(PConv)和分组空间卷积<sup>[20]</sup>(GSConv),分别形成的轻量化结构为 SP-ELAN、SG-ELAN、P-ELAN 以及 GS-ELAN。其中,SandGlass 为 MobileNeXt 网络专为移动轻量级网络设计的瓶颈结构,本文使用一半通道进行残差连接。

基于异构核的轻量化模块,通过使用大卷积核捕获更大的范围特征,同时实现轻量化。DC<sub>k</sub>-ELAN 基于异构卷积核的 ELAN,使用  $k \times k$  的深度可分离卷积代替 ELAN 模块中标准卷积。随着主干网络的加深,ELAN 的卷积核尺寸  $k$  依次为 3、3、5、7。同理,MS<sub>k</sub>-Block 使用 MS-Block 模块替换主干网络的 ELAN,同时随着主干网络的加深,MS-Block 的卷积核尺寸  $k$  依次为 3、3、5、7。

各轻量化网络均通过 K-means++ 聚类获得最优先验框,并使用 SiLU 激活函数,不同主干网络轻量化结果如表 2 所示。使用 B-ELAN 轻量化主干后,网络的参数量及计算量显著减低,mAP50 较高且运行速度仅次于 P-ELAN。P-ELAN 的模型帧率能够达到 112.4 帧/s,但检测中丢失了部分重要特征,准确度较基准网络 YOLOv7-tiny 下降 4.6%。

表 2 不同主干网络轻量化对比结果

模块	$P$	$R$	mAP50	参数量/ $\times 10^6$ 个	FLOPs/ $\times 10^9$ 次	帧率/帧 $\cdot s^{-1}$	模型大小/MB
B-ELAN	0.692	0.391	0.440	5.3	11.6	105.5	10.4
SP-ELAN	0.667	0.393	0.429	5.0	10.9	73.5	9.8
SG-ELAN	0.669	0.386	0.429	4.7	10.1	96.2	9.2
P-ELAN	0.638	0.398	0.421	4.6	9.7	112.4	8.9
GS-ELAN	0.624	0.400	0.421	4.5	9.8	103.1	9.0
DC <sub>k</sub> -ELAN	0.660	0.383	0.418	4.5	9.6	107.5	8.8
MS <sub>k</sub> -Block	0.623	0.359	0.388	4.6	10.3	94.3	9.2

#### 4.3 对比试验

为了验证本文模型的有效性,对比 YOLO 系列轻量化模型性能,包括:FPN 优化的 YOLOv7-tiny-EVC、YOLOv7-tiny-BiFPN;基于轻量化网络的 YOLOv7-tiny-VoVGSCSP、YOLOv7-tiny-MobileOne 和 YOLOv7-tiny-MobileNetV3 模型;基于视觉自注意力机制的 YOLOv7-

tiny-EfficientViT\_B1 和 YOLOv7-tiny-EfficientViT\_M3。所有模型迭代训练 100 次,并将各模型训练的最佳权重在测试集中进行测试,结果如表 3 所示。

由表 3 结果可知,在主流框架中,YOLOv8s 取得了最高的 mAP50,但其参数量、计算量和模型尺寸都远超其他模型。YOLOv5n 的参数量和计算量最小,运行速度最

快,但检测精度最低。YOLOv7-tiny的优化方法中,视觉自注意力机制对运行速度影响较大。本文模型在精确度和复杂度上获得较好的平衡,整体优于其他模型。

#### 4.4 消融试验

在BDD100K数据集上,通过消融试验检验本文模型中各模块的有效性,结果如表4所示。以原始YOLOv7-tiny为基线,SiLU激活函数使mAP50提升1.3%,模型的学习能力显著提升;使用K-means++聚类生成的先验锚

框,模型速度保持不变,mAP50提升0.6%,优化后的锚框更加适合道路目标检测;B-ELAN轻量化模型主干网络的计算量减少12.8%,能够在少量降低精度的前提下实现模型轻量化;MPDIoU不产生开销,能够提高检测精度的有效性。因此,相较于原始YOLOv7-tiny,本文模型的参数量减少18.3%,计算量减少15.0%,模型大小减少了17.9%,mAP50提高了1.1%,各模块能够轻量化网络结构,同时提升道路目标的检测精度。

表3 不同模型性能对比结果

模型	P	R	mAP50	参数量/ $\times 10^6$ 个	FLOPs/ $\times 10^9$ 次	帧率/帧 $\cdot s^{-1}$	模型大小/MB
YOLOv5n	0.645	0.348	0.375	1.8	4.3	133.3	4.0
YOLOv5s	0.700	0.404	0.447	7.0	16	113.0	14.0
YOLOv7-tiny	0.656	0.393	0.426	6.0	13.3	105.0	11.7
YOLOv8s	0.720	0.424	0.477	11.2	28.6	118.6	21.0
YOLOv7-tiny-VoVGSCSP <sup>[20]</sup>	0.715	0.383	0.433	5.7	12.4	97.3	11.2
YOLOv7-tiny-EVC <sup>[21]</sup>	0.659	0.402	0.433	7.1	16.6	83.0	13.8
YOLOv7-tiny-BiFPN <sup>[22]</sup>	0.665	0.398	0.432	6.0	13.2	90.4	11.7
YOLOv7-tiny-MobileOne <sup>[23]</sup>	0.563	0.341	0.350	4.5	10.9	114.0	14.7
YOLOv7-tiny-MobileNetV3 <sup>[24]</sup>	0.560	0.336	0.339	4.2	7.0	95.8	8.3
YOLOv7-tiny-EfficientViT_B1 <sup>[25]</sup>	0.663	0.417	0.444	8.1	14.1	78.0	15.6
YOLOv7-tiny-EfficientViT_M3 <sup>[26]</sup>	0.605	0.402	0.418	7.4	13.2	67.6	14.7
本文	0.685	0.393	0.437	4.9	11.3	120.2	9.6

表4 消融试验结果

SiLU	K-means++	B-ELAN	LMS模块	MPDIoU	mAP50	参数量/ $\times 10^6$ 个	FLOPs/ $\times 10^9$ 次	帧率/帧 $\cdot s^{-1}$	模型大小/MB
					0.426	6.0	13.3	105.0	11.7
√					0.439	6.0	13.3	103.0	11.7
	√				0.432	6.0	13.3	103.0	11.7
		√			0.417	5.3	11.6	105.5	10.4
			√		0.420	5.6	13.0	119.4	10.9
				√	0.428	6.0	13.3	105.0	11.7
√	√				0.451	6.0	13.3	103.0	11.7
√	√	√			0.440	5.3	11.6	105.5	10.4
√	√	√	√		0.434	4.9	11.3	120.2	9.6
√	√	√	√	√	0.437	4.9	11.3	120.2	9.6

以YOLOv7-tiny为基线,数据集中各常见类别目标检测结果如表5所示,同时,使用能见度有限的夜间行车图像对本文模型进行测试,结果如图4所示。本文模型能够检测出基线模型漏检的1个骑行者、1辆乘用车和2个交通标志,同时,检测目标的置信度也相应提升。改进后模型的交通标志检测精度提高了5个百分点,各道路目标的检测精度均有提高。

表5 不同道路目标的检测精度对比结果

类别	AP50	
	YOLOv7-tiny	本文
Car	0.734	0.753
Truck	0.555	0.565
Bus	0.550	0.569
Traffic Sign	0.534	0.584
Person	0.492	0.506



(a)YOLOv7-tiny



(b)本文

图4 夜间行车检测结果

#### 4.5 嵌入式试验

本文嵌入式平台使用英伟达 Jetson TX2,其配置环境为: Ubuntu 18.04 LTS 操作系统、JetPack 4.6.3、TensorRT 8.2.1、OpenCV 4.6.0、Torch 1.8.0、Torchvision 0.9.0、CUDA 10.2和cuDNN 8.2.1。

将各模型训练的最佳模型权重导入 Jetson TX2,并使用 TensorRT 加速推理行车视频。首先,从 PyTorch 中模型权重导出为 TensorRT 支持的 wts(Weights)格式文件;随后,在 TensorRT 中加载权重、定义网络、构建 TensorRT 引擎;最后,加载 TensorRT 引擎进行加速推理。

使用 TensorRT 加速的 YOLOv7-tiny 算法推理视频如图 5 所示,行车视频的采集地为湖南长沙万家丽路高架桥路段,得到实时推理速度为 48 帧/s,基本能够满足道路目标检测的实时要求。



图5 TensorRT加速推理行车视频

在 Jetson TX2 中,对比不同模型的运行速度,结果见图 6。输入图片尺寸为 600×600,相较于 YOLOv7-tiny,本文模型的帧率提升了 13 帧/s,相对提高 37.1%。同时,视频推理速度明显优于其他方法。

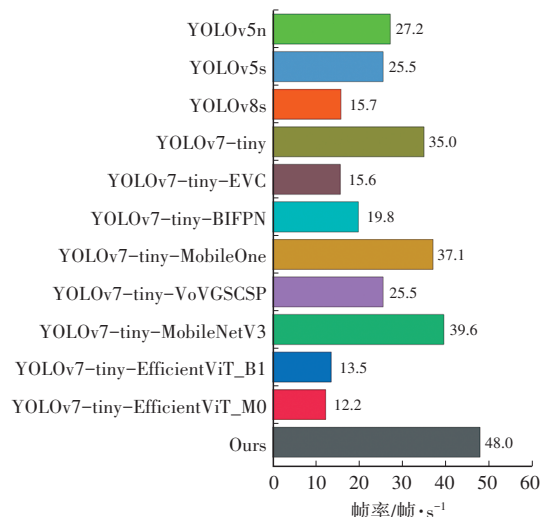


图6 不同模型的视频推理速度

## 5 结束语

本文基于 YOLOv7-tiny 模型进行改进,应用于道路目标检测时的轻量化问题。改进后的 YOLOv7-tiny 较原算法参数量、计算量和模型权重文件大小均大幅降低。通过在 BDD100K 数据子集上进行试验验证,证明了改进方法能够在轻量化网络的同时,提高了道路目标的检测精度。同时,在 Jetson TX2 上进行了 TensorRT 加速部署,验证了模型在嵌入式设备进行道路目标检测的实时性。

未来,将专注于改进后算法结合全景分割、可行驶区域检测、多目标跟踪、轨迹预测和深度估计等多重任务,进一步扩展其实际应用价值,并进行实车验证。

### 参考文献

- [1] 胡超超,刘军,张凯,等.基于深度学习的行人和骑行者目标检测及跟踪算法研究[J].汽车技术,2019(7):19-23.
- [2] HU C C, LIU J, ZHANG K, et al. Research on Target Detection and Tracking of Pedestrian and Cyclist Based on Deep Learning[J]. Automobile Technology, 2019(7): 19-23.
- [3] JIN Z Z, ZHENG Y F. Research on Application of Improved YOLO V3 Algorithm in Road Target Detection[C]// The 2020 International Conference on Mechatronics Technology and Intelligent Manufacturing (ICMTIM). Xi'an, China: Journal of Physics: Conference Series, 2020.
- [4] DONG X D, YAN S, DUAN C. A Lightweight Vehicles Detection Network Model Based on YOLOv5[J]. Engineering Applications of Artificial Intelligence, 2022, 113.
- [5] YAO J L, FAN X S, LI B, et al. Adverse Weather Target Detection Algorithm Based on Adaptive Color Levels and Improved YOLOv5[J]. Sensors, 2022, 22(21).
- [6] WANG J F, CHEN Y, DONG Z K, et al. Improved YOLOv5 Network for Real-Time Multi-Scale Traffic Sign

- Detection[J]. *Neural Computing and Applications*, 2023, 35(10): 7853–7865.
- [6] XUE Z Y, LIN H F, WANG F. A Small Target Forest Fire Detection Model Based on YOLOv5 Improvement[J]. *Forests*, 2022, 13(8).
- [7] LIU H, LI D, PENG J Z, et al. MTNAS: Search Multi-Task Networks for Autonomous Driving[C]// *Proceedings of the Asian Conference on Computer Vision*. Kyoto, Japan: Springer, 2020.
- [8] WANG P F, FU S, CAO X J. Improved Lightweight Target Detection Algorithm for Complex Roads with YOLOv5[C]// *2022 International Conference on Machine Learning and Intelligent Systems Engineering (MLISE)*. Guangzhou, China: IEEE, 2022: 275–283.
- [9] JIANG W Y, WEN J Y, XIE G M, et al. The Design of Lightweight Vehicle Detection Model Based on Improved YOLOv5[C]// *Third International Conference on Image Processing and Intelligent Control (IPIC)*. Kuala Lumpur, Malaysia: SPIE, 2023.
- [10] LIU C Y, HUANG D, WANG T, et al. Road Traffic Small Target Detection Algorithm Based on Lightweight YOLOv5[J]. *Frontiers in Computing and Intelligent Systems*, 2023, 3(1): 134–137.
- [11] MA S L, XU Y. MPDIoU: A Loss for Efficient and Accurate Bounding Box Regression[EB/OL]. (2023–07–14) [2023–12–27]. <https://arxiv.org/abs/2307.07662>.
- [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]// *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Vancouver, BC, Canada: IEEE, 2023: 7464–7475.
- [13] YU F, CHEN H F, WANG X, et al. BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning[C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, WA, USA: IEEE, 2020: 2636–2645.
- [14] GAO S H, CHENG M M, ZHAO K, et al. Res2net: A New Multi-Scale Backbone Architecture[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019, 43(2): 652–662.
- [15] CHEN Y M, YUAN X B, WU R Q, et al. YOLO-MS: Rethinking Multi-Scale Representation Learning for Real-Time Object Detection[EB/OL]. (2023–08–10) [2023–12–27]. <https://arxiv.org/abs/2308.05480>.
- [16] ZHENG Z H, 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*. Vancouver, Canada: PKP, 2020, 34(7): 12993–13000.
- [17] ZHANG Q L, JIANG Z Q, LU Q S, et al. Split to Be Slim: An Overlooked Redundancy in Vanilla Convolution[EB/OL]. (2020–06–22) [2023–12–27]. <https://arxiv.org/abs/2006.12085>.
- [18] ZHOU D Q, HOU Q B, CHEN Y P, et al. Rethinking Bottleneck Structure for Efficient Mobile Network Design [C]// *European Conference on Computer Vision (ECCV)*. Glasgow, UK: Springer International Publishing, 2020: 680–697.
- [19] CHEN J R, KAO S H, HE H, et al. Run, Don't Walk: Chasing Higher FLOPS for Faster Neural Networks[C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver, BC, Canada: IEEE, 2023.
- [20] LI H L, LI J, WEI H B, et al. Slim-Neck by Gsconv: A Better Design Paradigm of Detector Architectures for Autonomous Vehicles[EB/OL]. (2022–08–17) [2023–12–27]. <https://arxiv.org/abs/2206.02424>.
- [21] QUAN Y, ZHANG D, ZHANG L Y, et al. Centralized Feature Pyramid for Object Detection[J]. *IEEE Transactions on Image Processing*, 2023, 32: 4341–4354.
- [22] TAN M X, PANG R M, LE Q V. EfficientDet: Scalable and Efficient Object Detection[C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, WA, USA: IEEE, 2020: 10781–10790.
- [23] ZHANG Y L, FANG X P, GUO J, et al. CURI-YOLOv7: A Lightweight YOLOv7tiny Target Detector for Citrus Trees from UAV Remote Sensing Imagery Based on Embedded Device[J]. *Remote Sensing*, 2023, 15(19).
- [24] JIA L Q, WANG T, CHEN Y, et al. MobileNet-CA-YOLO: An Improved YOLOv7 Based on the MobileNetV3 and Attention Mechanism for Rice Pests and Diseases Detection[J]. *Agriculture*, 2023, 13(7).
- [25] CAI H, LI J Y, HU M Y, et al. EfficientViT: Lightweight Multi-Scale Attention for High-Resolution Dense Prediction[C]// *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. Paris, France: IEEE, 2023: 17302–17313.
- [26] LIU X Y, PENG H W, ZHENG N X, et al. EfficientViT: Memory Efficient Vision Transformer with Cascaded Group Attention[C]// *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Vancouver, BC, Canada: IEEE, 2023: 14420–14430.

(责任编辑 瑞 秋)

修改稿收到日期为2023年12月27日。