云南民族大学学报(自然科学版),2025,34(2):197-205 CN 53-1192/N ISSN 1672-8513
doi:10.3969/j.issn.1672-8513.2025.02.010 http://ynmz.cbpt.cnki.net

基于改进YOLOv8的行人和车辆检测算法研究

刘佳琳,殷丽凤
(大连交通大学 软件学院,辽宁 大连 116028)

摘要:针对目前主流的目标检测算法在复杂的交通环境下对行人和车辆检测精度不高的问题,提出一种基于YOLOv8模型改进的目标检测算法.首先,在主干网络前增加一种可学习增强网络,该网络是通过拉普拉斯分解残差学习方式来构建的,不仅可以充分提取到目标的特征信息,还能增强算法的准确性和鲁棒性,从而减少不同强度的光照对图像目标检测的干扰;其次,在中等目标检测层之前添加提出的KSA注意力机制,检测时可以将注意力集中在图像中重要信息的区域,从而更准确地定位并识别出中等目标,同时也可以减少复杂的背景噪声对检测的干扰;最后,提出了基于像素点的Transformer结构,即Pixel Transformer结构.将该结构添加到主干网络中,从而进一步增强算法对图像全局特征的提取能力,使其能学习到更丰富、更全面的目标特征.使用KITTI数据集来进行消融实验以及各算法的对比实验,实验结果表明,设计的算法在相关指标上取得了一定程度的提升,其中mAP@0.5值提升了3.2个百分点,达到了96.7%,这充分体现了该算法的优越性.

关键词:目标检测;YOLOv8;注意力机制;全局特征
中图分类号:TP311 **文献标志码:**A **文章编号:**1672-8513(2025)02-0197-09

随着车流量的增长与道路环境的日益复杂,交通事故发生的概率也随之上升.准确地检测行人和车辆对于交通安全、智能交通系统、自动驾驶及城市规划[1-2]等方面拥有不可估量的价值.不太准确的检测可能会导致严重的安全事故[3],对生命和财产造成不可挽回的损害.行人和车辆是交通系统中最常见的参与者,因此本文的宗旨是提升算法在行人和车辆目标的检测效果.

目前对于行人和车辆的目标检测具有广泛的应用前景,而目标检测算法便是其核心技术之一[4].其中目标检测算法分为单阶段和两阶段算法,两阶段的目标检测算法[5]主要有R-CNN[6]和Fast R-CNN[7]等.单阶段的目标检测算法主要有YOLO[8]系列和SSD[9]等.近年来,针对如何提升复杂场景下行人车辆目标检测的精度问题,前人提出了一些基于YOLO系列算法的改进[10-19].

以上研究与改进在检测精度方面有所提升,但是在复杂的城市道路环境下实现准确的识别依然是一个重大挑战.针对以上问题,本文基于YOLOv8算法进行相关的改进,主要的贡献如下:

(1)引入一种可学习增强网络.该网络用于增强在低照度环境下目标的特征信息,从而降低光照背景对行人和车辆中目标检测的影响.

(2)加入中等目标增强注意力机制KSA.KSA注意力机制使得算法更准确地定位并识别出中等目标对象,同时可以减少噪声的干扰,从而提高检测的准确率.

(3)主干网络添加本文设计的Pixel Transformer结构.Pixel Transformer结构可以进一步增强算法对图像全局特征的提取能力,使其提取到更丰富、更全面的特征信息.

1 基础知识

本节对YOLOv8网络模型以及损失函数、注意力机制和Transformer模型的相关基础知识进行介绍.

收稿日期:2024-07-11.
基金项目:国家自然科学基金(61771087).
作者简介:刘佳琳(1998-),女,硕士研究生.主要从事计算机视觉研究.
通信作者:殷丽凤(1976-),女,博士,副教授,硕士生导师.主要从事大数据挖掘、机器学习算法、计算机视觉等研究.

YOLOv8作为最新最先进的模型,由提出YOLOv5[20]的Ultralytics公司于2023年1月发布[21]. YOLOv8根据模型的深度和宽度提供了5个不同版本:YOLOv8n、YOLOv8s、YOLOv8m、YOLOv8l、YOLOv8x. YOLOv8网络模型分为输入端、Backbone、Neck和Head 4个部分,如图1为其网络结构图.

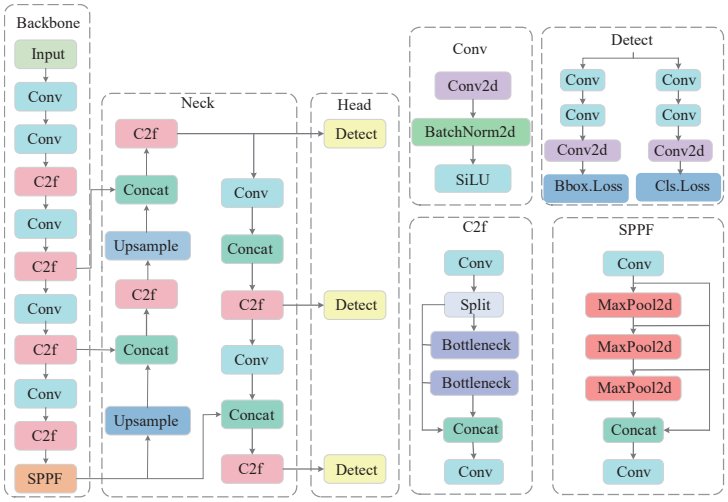


图1 YOLOv8网络模型

Backbone层主要用于提取输入图像的特征信息,主要包括卷积、C2f和SPPF等模块. 对YOLOv5中的C3模块用梯度流更丰富的C2f模块进行替换,而C2f模块依然采用残差连接的思想,同时也借鉴了YOLOv7中的ELAN结构,取消了分支中的卷积操作,并增加了额外的split操作和跳层连接,让提取到的特征信息更为丰富的同时,减少了计算量;SPPF则借鉴了SPP结构,将SPP的并行池化改为串行池化,进一步增大特征图的感受野,从而更好地检测不同尺度的物体.

Neck层主要将主干网络提取到的不同尺度的特征进行融合. 与YOLOv5相比,Neck层总体上依然采用FPN – PAN的颈部网络结构,但删除了FPN上采样阶段中的2个卷积模块,并将Neck层中的C3模块替换为C2f模块. 通过上采样和下采样的操作,将提取的不同尺度的特征信息融合到一起,进而捕获更为丰富的特征信息.

Head层主要用于目标的回归与预测,生成最终的目标检测结果. YOLOv8的Head层抛弃了以往YOLOv5的耦合头,换成解耦头结构,由Anchor – Based转为Anchor – Free. Anchor – Free将分类与检测头相分离,通过两条并行的分支结构,分别提取类别特征与位置特征,最后各用不同的卷积来完成分类与定位的任务,从而进一步提高网络的收敛速度与检测精度.

YOLOv8的边框回归损失由CIOU Loss + DFL Loss组成. CIOU Loss的计算如公式(1)所示,其中,IOU为交并比,b, b^{gt}分别为预测框的中心点坐标与真实框的中心点坐标,$p^2(b, b^{gt})$为b与b^{gt}两点之间的欧式距离,c为预测框和真实框最小外接矩形的对角线距离,w和h为预测框的宽和高,w^{gt}和h^{gt}为真实框的宽和高. v是惩罚项,α是权重函数见式(2 ~ 3).

$$\mathrm{CIOU} = 1 - \mathrm{IOU} + \frac{p^2(b, b^{gt})}{c^2} + \alpha v. \tag{1}$$

$$v = \frac{4}{\pi^2}(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h})^2. \tag{2}$$

$$\alpha = \frac{v}{(1 - \mathrm{IOU}) + v}. \tag{3}$$

DFL损失通过交叉熵的形式来优化与标签最接近的左右两侧位置的概率,从而更准确地识别并解析目标位置附近区域的分布情况,其值的计算如下公式(4)所示. 其中,y表示标签值,y_i和y_{i+1}为最接近y的2个数值,S_i和S_{i+1}表示全局最小解,S_i的取值如公式(5)所示,S_{i+1}的取值如公式(6)所示.

$$\mathrm{DFL}(S_i, S_{i+1}) = -((y_{i+1} - y)\log(S_i) + (y - y_i)\log(S_{i+1})). \tag{4}$$

$$S_i = y_{i+1} - y / y_{i+1} - y_i. \tag{5}$$

$$S_{i+1} = y - y_i / y_{i+1} - y_i. \tag{6}$$

为了进一步提高对中等尺寸目标的检测能力，提出了一种新的注意力机制，并将其添加到中等目标检测层．该注意力机制采用了通道注意力机制和核选择注意力机制中的部分结构．其中通道注意力机制（squeeze-and-excitation，记作 SE）[22]主要聚焦在通道维度，重点关注权重值大的通道，即当前有效的特征图通道．核选择注意力机制（selective kernel networks，记作 SK）[23]使算法能够通过自学习的方式来选择融合不同感受野的特征信息．

Transformer[24]结构是由 Google 团队于 2017 年提出的一种全新神经网络架构，主要由 Encoder 和 Decoder 两部分组成，其具备捕获全局特征信息的能力．Vision Transformer[25]是 Google 团队于 2020 年提出的，不再需要 CNN 结构，而是针对原输入图像进行分块处理，并将图像块 Patch 作为 Token 输入 Transformer．其准确率在 ImageNet1K 数据集上已经达到了 88．55%．

2　改进 YOLOv8 算法

为了优化 YOLOv8 算法在复杂场景下对行人和车辆的检测精度，提出一种基于改进 YOLOv8 的目标检测算法，其网络结构如图 2 所示．首先，针对暗光和强光等不同光照环境下对目标检测的影响，在 Backbone 层前增加一个可学习增强网络，该网络是利用拉普拉斯分解残差学习的方式进行构建的．其次，为了增加对一些重要的目标区域的关注，同时减少复杂背景的干扰，在 Head 层引入一种基于 SK 注意力机制与 SE 注意力机制部分结构组成的 KSA 注意力机制．最后，为了增强算法对全局特征信息的提取能力，在 Backbone 层的中间区域加入了本文提出的基于像素点的 Pixel Transformer 结构．

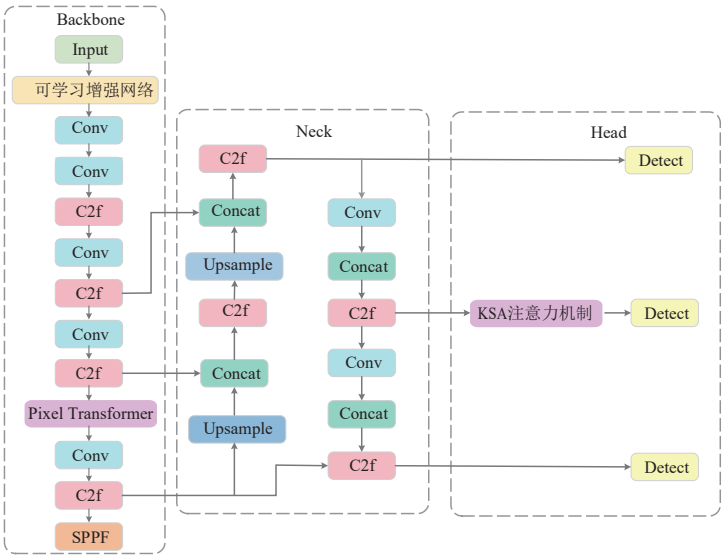


图 2　本文算法网络模型

2．1　可学习增强网络

在真实场景下，经常会存在诸如夜晚暗光和中午强光等背景环境，这些不利的光照条件会导致图像中的目标特征变得模糊不清，甚至可能被背景噪声所淹没，从而会对行人和车辆目标的检测存在一定的影响．借鉴 PE－YOLO[26]用于暗物体检测的金字塔增强网络的思想，设计了一种可学习增强网络，用于增强暗光以及强光背景下目标的特征信息，具体过程如图 3 所示．首先，输入原始图像，通过采用拉普拉斯分解得到高频图层，将原始图像与高频图层进行 Concat 拼接得到新的特征图，其可以保留原始图像的整体信息同时增强高频细节，再通过两次卷积操作，将生成的特征图与原始图像相加，最终得到增强后的图像．拉普拉斯分解的具体过程为：首先，对原始图像应用高斯滤波器进行模糊处理，从而去除图像中的高频噪声和细节，这一步的目的是提取图像的基本结构和轮廓信息；接着，将原始图像与经过高斯模糊处理后的图像进行相减并取绝对值，得到高频图层．高频图层主要包含图像的细节、边缘以及纹理等高频信息，这些信息在暗光条件下往往被噪声掩盖或减弱．

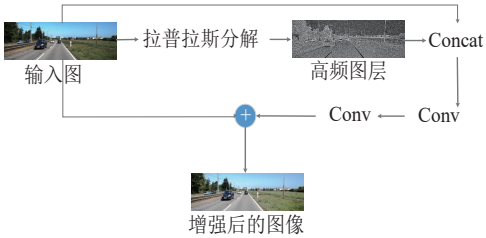


图 3　可学习增强网络

在输入端加入上文提出的可学习增强网络,将输入图像经过可学习增强网络处理,其不仅可以提取目标的高频特征信息,从而减少光照对行人和车辆目标检测的干扰,还能增强目标检测算法的鲁棒性. 总之,可学习增强网络为存在暗光以及强光的图像目标检测提供了一种有效的特征增强方法.

2.2　中等目标增强注意力机制 KSA

若图像背景中存在大量的建筑物、树木以及其他杂物,这些复杂的背景会干扰车辆与行人目标特征信息的提取. 在 YOLOv8 算法的检测层加入注意力机制,可以将检测时的注意力集中在图像重要信息的区域,其不但能增强对目标区域的关注,还能在一定程度上抑制对背景的关注,从而减少背景的干扰. 本文借助通道注意力机制(squeeze – and – excitation,记作 SE)与核选择注意力机制(selective kernel networks,记作 SK)两者的部分结构,提出一种注意力机制 KSA(简称为 KSA 注意力机制),如图 4 所示,其中,C 为通道数,r 为降维系数,一般取 16.

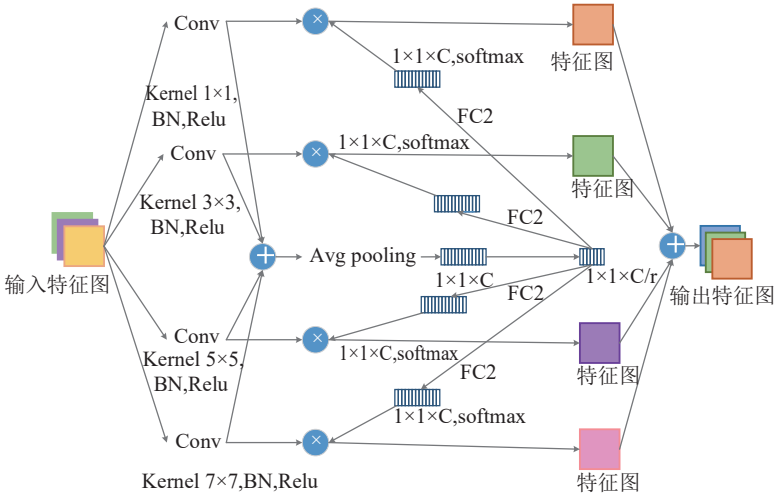


图 4　KSA 中等目标增强注意力机制

KSA 注意力机制的实现过程如下:首先,输入特征图分别经过 4 个分支处理来提取不同尺度的特征,每个分支分别由卷积、BN 和 Relu 激活函数等结构组成,卷积操作对应 4 个不同尺度的卷积核,分别为 1 × 1、3 × 3、5 × 5、7 × 7. 其次,将得到的 4 个特征图以相加的方式进行融合,再通过全局平均池化获得维度为 1 × 1 × C 的特征向量,并让特征向量进入全连接层降维到 1×1×C/r,再分别通过 4 个全连接层升维到原来的通道维度. 接着,利用 softmax 函数将 4 个特征值转化为注意力权重,再将 4 个特征图分别进行加权操作,从而得到处理后的特征图. 最后,将处理后的 4 个特征图进行融合.

KSA 注意力机制首先采用了 SK 注意力机制使用不同尺度的卷积核提取特征的设计思想,通过自学习的方式选择多个不同尺度的卷积核,能够更好地捕捉特征之间的关系,另外不同尺度的卷积处理也可以提升算法的表达能力:其次,KSA 注意力机制借鉴了 SE 注意力机制先降维再升维的操作思想,这种将通道维度先压缩再还原的操作能够自适应地学习每个通道的重要性,另一方面通过降维和升维,算法也能够以较少的参数学习通道间的复杂关系.

由于行人和车辆数据集中等目标的数量较多,因此,本文选择在中等目标检测层之前添加提出的 KSA 注意力机制,在目标检测任务中,能够更准确地定位并识别出中等目标对象,同时减少复杂的背景噪声干扰,从而提高检测的准确率.

2.3　Pixel Transformer

传统的 Vision Transformer 结构在处理图像时，通常会使用 Patch 的方式，先将图像分割成多个固定大小的 Patch，然后对每个 Patch 进行向量化得到 Token. 但 Patch 作为引入的超参数，一般需要手动设置 Patch 的大小，这样容易产生一些误差，导致对算法检测目标的精度存在影响. 为了使 Transformer 结构能够更具自适应性，从而提高算法的检测精度和鲁棒性. 由此本文提出了基于像素点的 Transformer 结构，即 Pixel Transformer.

在 Pixel Transformer 中，每个像素点可以直接作为 Token，避免了传统 Vision Transformer 中需要人为设置 Patch 大小的问题. 这使得算法能够更加自适应地处理不同尺度的图像目标特征，特别是在高分辨率图像中，能够保留更多的细节信息. 本文考虑到 YOLOv8 深层输出的特征图尺寸较小，通道数量较多，并且每个像素区域的感受野相对较大，每个像素已经包含了丰富的上下文信息. 因此在构建 Transformer 时，不再进一步划分 Patch，而是直接将每个像素视为一个 Token. 如图 5 所示，C、H、W 分别代表通道数、高度以及宽度. 输入特征图经过 Tokenize 向量化后，每个像素点作为一个 Token，再经过 3 层重复堆叠的 Transformer Encoder 处理，提取出更高层次的特征信息，并增加模型的非线性能力和泛化能力. 如图 5 和图 6 所示，最终输出具有全局信息以及丰富语义信息的特征图. 其中，Token 的总数为输入特征图宽高的乘积，每个 Token 的特征长度为通道维度. 此外，Transformer Encoder 主要由 Layer Norm、Multi-Head Attention、MLP Block 等结构组成. Layer Norm（层归一化）结构将每个 Token 进行归一化处理，用于提高算法的训练稳定性和性能. Multi – Head Attention（多头注意力机制）由多个 Self – Attention 组成，每个 Attention 头都能获得一个不同的表示空间. 这种机制通过在每个头中使用不同的查询（Query）、键（Key）和值（Value）权重矩阵，将输入序列投影到不同的子空间中，从而允许算法关注不同方面的特征信息. MLP Block（多层感知机块）通常由两个全连接层（也称为线性层）和一个非线性激活函数组成. 第一个线性层通常会将输入的特征维度扩展，第二个线性层则将扩展后的特征维度还原到原始维度. 这种结构有助于算法学习目标的复杂特征.

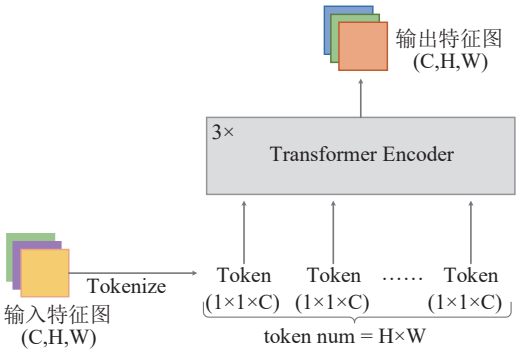


图 5　Pixel Transformer 结构

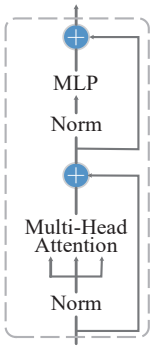


图 6　Transformer Encoder 结构

在 YOLOv8 主干网络中的 C2f 结构后添加 Pixel Transformer，可以进一步增强算法对目标全局特征的提取能力，提升多尺度特征融合的效果，优化梯度流信息的传递，并提高算法的泛化能力. 然而，另一方面值得注意的是，添加 Pixel Transformer 也会增加算法的计算复杂度和参数量. 由于行人和车辆数据集中等目标的数量较多，因此，考虑到算法的复杂度和参数量，本文只选择在主干网络中特征图尺寸为 40 × 40 的 C2f 结构后添加 Pixel Transformer. 加入 Pixel Transformer 后的算法可以捕获更多全局信息，使其能够学习到更丰富、更全面的目标特征.

3　实验结果与分析

3.1　实验环境

本实验的操作系统为 Linux，CPU 为 Intel（R）Xeon（R）Platinum 8280L，内存 32 G，显卡为 4 × NVIDIA GeForce RTX 3090，显存为 24 G × 4，深度学习框架为 Pytorch，cuda 的版本为 11.3，分布式训练方式为 DDP，学习率调整方式为 WarmUp，使用 YOLOv8n 作为基线模型. 超参数设置 batch size 为 16，epochs 为 1 000，初始

学习率为0.002 8.

3.2 数据集

采用KITTI－2D数据集作为实验数据集,该数据集由德国卡尔斯鲁厄理工学院和美国丰田技术研究所联合创建.它包含城市、高速以及校园等各种复杂交通场景下的真实交通图像,非常适用于对行人和车辆等目标的检测.数据集的类别分为Car、Van、Truck、Pedestrian、Person_sitting、Cyclist、Tram、Misc、Dontcare.由于交通环境中的目标大多数是行人与车辆,因此将原始数据集的类别标签重新组织并进一步划分类别.其中,将Car、Van、Truck、Tram标签合并到Car的类别中;将Pedestrian、Person_sitting标签合并到Pedestrian的类别中;去除Misc和Dontcare等类别.最后选择Car、Pedestrian和Cyclist作为最终的检测类别.数据集总共包含7 481张图像,按8∶1∶1的比例随机划分为训练集(5 985张)、验证集(753张)和测试集(753张).

3.3 评价指标

本实验选择YOLO系列算法中常用的性能评估指标:精度(precision,P)、召回率(recall,R)、所有类别的平均精度(meanaverage precision,mAP).其计算公式如式(7)~(10)所示.

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}. \tag{7}$$

$$R = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}. \tag{8}$$

$$\mathrm{AP} = \int_0^1 P(R)\,\mathrm{d}R. \tag{9}$$

$$\mathrm{mAP} = \frac{1}{n}\sum_{i=1}^{n}\int_0^1 P(R)\,\mathrm{d}R. \tag{10}$$

其中,TP为真正类,FP为假正类,FN为假负类.精度P表示模型预测的正类样本中真实正类所占比例;P的值越大证明某一类目标的识别效果越好.召回率R表示在真实正类样本中模型预测正类所占的比例;R值越大证明较多的正类被预测正确.AP表示单个目标类别的平均精度,其值是$P-R$曲线所围图像的面积.mAP是将数据集中所有类别的AP取平均值,用于全面评估算法在各个类别上的检测精度,其中mAP@0.5是指当IOU的值为0.5时,所有类别的平均精度.如果AP值与mAP@0.5值越大,则表明该算法的整体性能效果越好.

3.4 对比实验

为验证本文算法的有效性和可靠性,将其与当前几种主流算法进行对比.包括Fast－RCNN、SSD、YOLOv4、YOLOv5s、YOLOv8和本文基于YOLOv8改进的算法.并通过Car、Pedestrain、Cyclist各类别的AP值以及mAP@0.5值进行对比,不同算法的各项指标对比结果见表1所示.

表1 不同算法在KITTI数据集上的指标对比 %

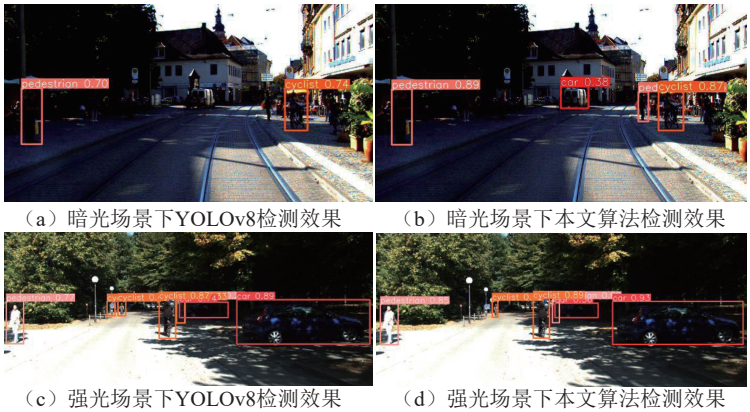
模型	Car	Pedestrain	Cyclist	mAP@0.5
Fast－RCNN	84.8	70.5	75.4	76.9
SSD	80.8	68.5	74.1	74.3
YOLOv4	95.0	76.1	86.0	85.85
YOLOv5s	97.2	83.6	90.1	90.3
YOLOv8n	98.1	83.3	90.8	93.5
本实验	99.4	92.2	97.7	96.7

实验表明,本文算法的mAP@0.5值均超越了常见的主流算法,检测精度方面效果最佳.相较于基准YOLOv8算法,本文算法的检测精度存在一定程度的提升.mAP@0.5达到了96.7%,提高了3.2个百分点.除此之外,各类别物体的精度均得到了提升,Car、Pedestrain、Cyclist等类别的AP值与YOLOv8基准模型相比,分别提升了1.3、8.9和6.9个百分点.实验结果显示改进后的算法对行人和车辆的检测具有一定的效果,这也充分证明了本文算法的可行性.

3.5 检测效果及分析

为了更直观地展示本文算法的检测效果,从KITTI数据集中选取复杂场景中具有代表性的两组图片进行对比.如图7所示.上图为原YOLOv8算法检测图,下图为本文算法检测图.

在图 7 的不同场景中，由第一组图片可见，此场景下 YOLOv8 的检测精度不高，且容易出现误检漏检．改进算法后，不仅能检测出暗光线下的行人和车辆，而且各物体类别的置信度也存在提升．相比 YOLOv8 至少提高了 10 个百分点．第二组图片中，YOLOv8 模型误将远处的交通标志牌识别为 Cyclist，但改进后的本文算法并未出现此种情况．由此可见，改进后的本文算法对于复杂交通环境下的目标识别能力有一定的提升．



（a）暗光场景下YOLOv8检测效果　（b）暗光场景下本文算法检测效果

（c）强光场景下YOLOv8检测效果　（d）强光场景下本文算法检测效果

图 7　不同场景下 2 种算法检测效果对比

3.6　消融实验

为了验证每个改进点相对于 YOLOv8 算法的有效性，在原 YOLOv8 算法的基础上，分别设计了可学习增强网络、KSA 注意力机制以及 Pixel Transformer 等模块的消融实验．实验结果如表 2 所示．

表 2　消融实验结果

组别	Base	可学习增强网络	KSA	Pixel Transformer	mAP@0.5/%
1	√				93.5
2	√	√			94.4
3	√		√		95.3
4	√			√	94.4
5	√	√	√	√	96.7

由表 2 可见，每个改进点相较于原 YOLOv8 算法的 mAP@0.5 值均有提升．第一组实验为原 YOLOv8 算法的基础实验，mAP@0.5 为 93.5%．第二组实验为可学习增强网络的检测结果，利用拉普拉斯分解残差学习方式进行构建可学习增强网络，用于增强极端光照下图像目标的特征信息，其 mAP@0.5 达到了 94.4%，相对于原算法提升了 0.9%．第三组实验为 KSA 注意力机制的检测结果，加入该注意力机制后，使得算法更加关注图像中集中存在目标的区域，其 mAP@0.5 达到了 95.3%，相对提升了 1.8%．第四组实验为 Pixel Transformer 的检测结果，添加该结构后可以捕获全局信息，使算法能够学习到更丰富、更全面的目标特征，其 mAP@0.5 达到了 94.4%，相对提升了 0.9%．第五组实验为整体改进后的检测结果，汲取了上述改进部分的优势，mAP@0.5 达到了最高，为 96.7%，相较于原 YOLOv8 算法提升了 3.2%．综合上述消融实验结果，证明了本文改进方法的有效性以及各方法之间具有较好的兼容性．

4　结语

为了提高 YOLOv8 算法在复杂交通场景中行人车辆的检测精度，使其更加准确地识别出目标，提出一种基于 YOLOv8 改进的行人和车辆目标检测算法．首先，设计一种可学习增强网络，其可以提取到目标丰富的特征信息，从而减少极端光照条件下对图像目标检测的干扰；其次，加入设计的 KSA 注意力机制，将检测注意力集中在图像中重要信息的区域，其不但能增强对目标区域的关注，还能减少背景的干扰；最后，添加提出的 Pixel Transformer 结构，可以进一步增强算法对目标全局特征的提取能力，促进多尺度特征的融合．为了证明各方法的有效性以及之间的兼容性，将各方法进行消融实验．最后与一些主流的目标检测算法进行

对比实验,证明了改进后的算法在行人和车辆等目标检测上取得了预期的效果.

参考文献:

[1] TERVEN J, CÓRDOVA – ESPARZA D M, ROMERO – GONZÁLEZ J A. A comprehensive review of YOLO architectures in computer vision: from YOLOv1 to YOLOv8 and YOLO – NAS[J]. Machine learning and knowledge extraction, 2023, 5(4), 1680 – 1716.

[2] SONG Y, HONG S, HU C, et al. MEB – YOLO: an efficient vehicle detection method in complex traffic road scenes[J]. Computers, Materials & Continua, 2023, 75(3): 5761 – 5784.

[3] 陈虹,郭露露,宫洵,等. 智能时代的汽车控制[J]. 自动化学报,2020,46(7):1313 – 1332.

[4] MOZAFFARI S, AL – JARRAH O Y, DIANATI M, et al. Deep learning-based vehicle behavior prediction for autonomous driving applications: a review[J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 23(1): 33 – 47.

[5] ZAIDI S SA, ANSARI M S, ASLAM A, et al. A survey of modern deep learning based object detection models[J]. Digital Signal Processing, 2022, 126: 103514.

[6] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurateobject detection and semantic segmentation[C]//Proceedingsof the IEEE conference oncomputer vision and pattern recognition. 2014: 580 – 587.

[7] GIRSHICK R. Fast r – cnn[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1440 – 1448.

[8] 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.

[9] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//Computer Vision – ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11 – 14, 2016, Proceedings, Part I 14. Springer International Publishing, 2016: 21 – 37.

[10] 胡淼,姜麟,陶友凤,等. 改进YOLOv7的自动驾驶目标检测算法[J]. 计算机工程与应用,2024,60(11):165 – 172.

[11] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. Yolov7: trainable bag – of – freebies sets new state – of – the – art for real – time object dete – ctors[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 7464 – 7475.

[12] 宋绍剑,夏海姐,李刚,等. YOLOv5的改进算法及其在自动驾驶多目标检测的应用研究[J]. 计算机工程与应用,2023,59(15):68 – 75.

[13] 刘辉,刘鑫满,刘大东,等. 面向复杂道路目标检测的YOLOv5算法优化研究[J]. 计算机工程与应用,2023,59(18):207 – 217.

[14] 魏陈浩,杨睿,刘振丙等. 具有双层路由注意力的YOLOv8道路场景目标检测方法[J]. 图学学报,2023,44(6):1104 – 1111.

[15] 田鹏,毛力. 改进YOLOv8的道路交通标志目标检测算法[J]. 计算机工程与应用,2024,60(8):202 – 212.

[16] 朱强军,胡斌,汪慧兰,等. 基于轻量化YOLOv8s交通标志的检测[J]. 图学学报,2024,45(3):422 – 432.

[17] 周飞,郭杜杜,王洋,等. 基于改进YOLOv8的交通监控车辆检测算法[J]. 计算机工程与应用,2024,60(6):110 – 120. .

[18] 熊恩杰,张荣芬,刘宇红,等. 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.

[19] 张利丰,田莹. 改进YOLOv8的多尺度轻量型车辆目标检测算法[J]. 计算机工程与应用,2024,60(3):129 – 137.

[20] ZHU X, LYU S, WANG X, et al. TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone – captured scenarios[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 2778 – 2788.

[21] SOHAN M, SAI RAM T, REDDY R, et al. A review on YOLOv8 and its advancements[C]//Prceedings of the International Conference on Data Intelligence and Cognitive Informatics. 2024: 529 – 545.

[22] HU J, SHEN L, SUN G. Squeeze – and – excitation networks[C]//Proceedings of the IEEE Conference onComputer Vision and Pattern Recognition. 2018: 7132 – 7141.

[23] LI X, WANG W, HU X, et al. Selective kernel networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 510 – 519.

[24] WASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the International Conference on Neural Information Processing Systems. 2017: 5998 – 6008.

[25] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[C]//Proceedings of the International Conference on Learning Representations. 2021: 10 – 20.

[26] YIN X, YU Z, FEI Z, et al. PE-YOLO: Pyramid enhancement network for dark object detection[C]//Proceedings of the International Conference on Artificial Neural Networks. 2023: 163 – 174.

Research on an improved pedestrian and vehicle detection algorithm based on YOLOv8

LIU Jia-lin，YIN Li-feng

（School of Software，Dalian Jiaotong University，Dalian 116028，China）

Abstract：In view of the problem that the mainstream target detection algorithms currently have low detection accuracy for pedestrians and vehicles in complex traffic environments，this paper proposes an improved target detection algorithm based on the YOLOv8 model. Firstly，a learnable enhancement network is added in front of the backbone network. This network is constructed through Laplacian decomposition residual learning method，which not only allows for full extraction of target feature information，but also enhances the accuracy and robustness of the algorithm，thus reducing the interference of different intensities of illumination on image target detection. Secondly，the proposed KSA attention mechanism is introduced before the medium target detection layer in this paper. After adding this attention mechanism，during detection，the attention can focus on the areas of important information in the image，so as to more accurately locate and identify medium targets，and at the same time，it can also reduce the interference of complex background noise on detection. Finally，this paper proposes a transformer structure based on pixel points，namely the Pixel Transformer structure. By adding this structure to the backbone network，the ability of the algorithm to extract global features of the image is further enhanced，enabling it to learn richer and more comprehensive target features. This paper uses the KITTI dataset to conduct ablation and comparison experiments of various algorithms. The experimental results show that the algorithm designed in this paper has achieved a certain degree of improvements in relevant indicators. Among them，the mAP@0. 5 value has increased by 3. 2 percentage points and reached 96. 7%，which fully reflects the superiority of the algorithm .

Key words：target detection；YOLOv8；attention mechanism；global features

（责任编辑　段　鹏）