

基于改进 YOLOv3-SPP 算法的道路车辆检测

王涛^{1,2}, 冯浩^{1,2}, 秘蓉新³, 李林³, 何振学⁴, 傅奕茗^{1,2}, 吴姝^{1,2}

(1. 北京信息科技大学信息与通信工程学院, 北京 100101;

2. 北京信息科技大学光电测试技术及仪器教育部重点实验室, 北京 100192;

3. 国家计算机网络应急技术处理协调中心, 北京 100029;

4. 河北农业大学河北省农业大数据重点实验室, 河北 保定 071001)

摘要: 针对在城市道路场景下视觉检测车辆时, 车辆密集和远处车辆呈现小尺度, 导致出现检测精度低或者漏检的问题, 提出了一种基于改进的 YOLOv3-SPP 算法, 对激活函数进行优化, 以 DIOU-NMS Loss 作为边界框损失函数, 增强网络的表达能力。为提高所提算法对小目标和遮挡目标的特征提取能力, 引入空洞卷积模块, 增大目标的感受野。实验结果表明, 所提算法在检测车辆目标时 mAP 提高了 1.79%, 也有效减少了在检测紧密车辆目标时出现的漏检现象。

关键词: 车辆检测; YOLOv3-SPP 算法; 激活函数; 空洞卷积; 深度学习

中图分类号: TN92

文献标志码: A

DOI: 10.11959/j.issn.1000-436x.2024046

Road vehicle detection based on improved YOLOv3-SPP algorithm

WANG Tao^{1,2}, FENG Hao^{1,2}, MI Rongxin³, LI Lin³, HE Zhenxue⁴, FU Yiming^{1,2}, WU Shu^{1,2}

1. School of Information and Communication Engineering, Beijing Information Science & Technology University, Beijing 100101, China

2. Key Laboratory of Optoelectronic Testing Technology and Instrument, Ministry of Education,
Beijing Information Science & Technology University, Beijing 100192, China

3. National Computer Network Emergency Response Technical Team/Coordination Center of China, Beijing 100029, China

4. Hebei Key Laboratory of Agricultural Big Data, Hebei Agricultural University, Baoding 071001, China

Abstract: Aiming at the problem of low detection accuracy or missing detection caused by dense vehicles and small scale of distant vehicles in the visual detection of urban road scenes, an improved YOLOv3-SPP algorithm was proposed to optimize the activation function and take DIOU-NMS Loss as the boundary frame loss function to enhance the expression ability of the network. In order to improve the feature extraction ability of the proposed algorithm for small targets and occluding targets, the void convolution module was introduced to increase the receptive field of the target. Based on the experimental results, the proposed algorithm improves the mAP by 1.79% when detecting vehicle targets, and also effectively reduce the missing phenomenon when detecting tight vehicle targets.

Keywords: vehicle detection, YOLOv3-SPP algorithm, activation function, atrous convolution, deep learning

收稿日期: 2023-10-27; 修回日期: 2024-01-17

通信作者: 秘蓉新, mirongxin@cert.org.cn

基金项目: 国家自然科学基金资助项目 (No.62001034, No.62102130, No.6210031547); 北京市自然科学基金资助项目 (No.4232004); 北京市教育委员会科学研究计划基金资助项目 (No.KM202111232013); 河北省自然科学基金资助项目 (No.F2020204003); 河北省青年拔尖人才计划基金资助项目 (No.BJ2019008)

Foundation Items: The National Natural Science Foundation of China (No.62001034, No.62102130, No.6210031547), The Natural Science Foundation of Beijing (No.4232004), The Research and Development Program of Beijing Municipal Education Commission (No.KM202111232013), The Natural Science Foundation of Hebei Province (No.F2020204003), The Youth Top Talent Program Project of Hebei Province (No.BJ2019008)

0 引言

近年来,我国的汽车保有量不断增长。汽车的普及极大地提高了人们日常生活的效率,但这种发展也带来了一些如交通堵塞、交通事故频繁等交通问题^[1]。车辆检测算法研究发展快速,相关技术已经在智能交通系统中有了成熟的应用,并减少了事故的发生。强大的智能交通系统的一个重要作用是最大可能地保护个人免受交通事故的影响,从而车辆检测算法研究成为必要。传统的车辆检测算法是基于机器学习的算法。这种算法首先在图像中生成候选区域,其次在候选区域中采用 Haar-like^[2]、Log Gabor^[3]等算法提取特征,最后,通过分类器(支持向量机、AdaBoost 等)对提取到的特征进行分类识别。但此类算法具有一定的局限性,如受光照影响、依赖于人工提取特征和先验知识、鲁棒性较差、检测速度达不到实时性要求等。

基于深度学习的目标检测取得了多项重大突破,成为目标检测领域的主流算法。基于深度学习的目标检测算法按照处理图像中目标的任务可分为两大类。第一类是目标的识别和定位分开处理,即两阶段的目标检测算法,第一阶段对目标进行识别,通过选择性搜索或者区域生成网络生成候选框,该阶段专注于找出目标出现的位置,保证足够的召回率和准确率;第二阶段对目标进行定位,通过卷积神经网络(CNN, convolutional neural network)进行候选框的分类和位置的修正,代表算法有 RCNN (region CNN)^[4]、Fast RCNN^[5]、Faster RCNN^[6]等。第二类是目标的识别和定位同时处理,即单阶段的目标检测算法,从输入端输入原始数据到输出端即可得到目标的类别概率和位置坐标值,一个网络模型便可完成识别和定位的任务,代表算法有 SSD (single shot detector)^[7]、YOLO (you only look once)^[8-9]等。两阶段目标检测算法相较于单阶段目标检测算法有较低的识别错误率和漏检率,而单阶段目标检测算法相较于两阶段目标检测算法有更快的检测速度,更适用于实时检测场景中,因此提高单阶段目标检测算法的检测精度和检测速度更具实用价值。

目前,车辆检测算法的主要改进方向有不同层级的特征融合、改变网络结构和平衡数据集等。文献[10]基于改进 Faster RCNN,堆叠小卷积核来提取小目标特征,设计三通道的密集层来融合特征,在降低网络复杂度的同时保证车辆检测性能。文献[11]提出了一

种基于 Transformer 的新型目标检测框架 DETR (detection transformer),为 Transformer 在道路车辆检测任务中的应用奠定了重要的基础,但是 Transformer 基于自注意力机制建模图像数据的全局依赖,其计算复杂度的数量级是输入图像分辨率的二次方,因此,基于 Transformer 的算法在处理密集道路车辆图像或视频输入时会产生较大的计算量。文献[12]在分析传统的 Haar-like 车辆识别算法基础上,提出融合图像边缘特征的车辆识别算法 FE-CNN (fused edge features-CNN),融合后的图像特征不仅可以减少在提取图像边缘特征过程中产生的噪声,还可以有效补偿提取特征时丢失的部分特征信息,降低模型对训练图像高质量的要求,在真实车辆运行场景中表现较好。文献[13]将 ShuffleNetV2 作为主干网络,并去掉大部分的卷积层来构建轻量级的车辆检测网络,降低了复杂度的轻量级车辆目标检测网络更适用于性能有限的边缘计算设备。文献[14]提出了一种自适应多模态特征融合和跨模态指数模型,设计的多模态自适应融合模块在特征融合时选择高权重特征,压缩低权重冗余特征,利用多尺度特征金字塔进行自适应融合,另外建立的跨模态车辆指数在提取目标区域时,抑制了复杂背景信息,最大限度地减少了车辆检测中的误报。

在实际的道路车辆检测场景中,由于驾驶人员需要提前知道道路情况,以便及时做出反应,因此车载摄像头与车辆之间会存在距离较远的问题,距离较远的车辆目标在高分辨率图像中的占比很小,而且交通环境复杂,小目标车辆很容易与背景融合,导致提取特征较小,给检测任务增加了难度。此外,当道路车辆密集时,会存在严重的遮挡情况。车辆之间存在遮挡时,只能提取可见部分的特征信息,导致车辆信息不完整,检测出的目标回归框不准确,同时分类也会出现错误,导致误检或重检。另外,车辆行驶过程中难免会遇到雨天、雾天、雪天或冰雹等天气情况,这些天气情况都会使摄像头拍摄模糊,导致车辆检测任务出现漏检、误检等。在夜晚行驶时,摄像头捕捉到的画面不清晰,检测出现的漏检和误检的概率也很大。

YOLOv3-SPP 检测算法是在 YOLOv3 的基础上改进而来的。YOLOv3-SPP 分为两部分,第一部分为特征提取层,特征提取层以 DarkNet53 为主干网络,DarkNet53 网络结构使用了大量的 1×1 卷积、 3×3 卷积并配合残差模块对特征进行提取;第二部分为特征交互层,主要是利用空间金字塔特征图的

方式对大、中、小 3 种尺度的目标进行特征交互。YOLOv3-SPP 是在第五、六层卷积之间加入一个空间金字塔池化 (SPP, spatial pyramid pooling) 模块。该模块主要由 3 个最大池化层 (大小分别为 5×5 、 9×9 、 13×13) 和一个跳层连接组成。然而当 YOLOv3-SPP 网络提取小目标车辆特征时, 由于车辆尺度较小, 相应地包含的有效信息较小, 容易出现漏检; 当 YOLOv3-SPP 网络提取遮挡的车辆特征时, 由于图像中出现的是车辆的部分特征, 网络对遮挡目标的特征提取能力又较弱, 因此在这种情况下, 很容易出现漏检和误检的情况。

针对传统 YOLOv3-SPP 目标检测算法在道路车辆检测中出现的检测精度低、检测准确度低和漏检等问题, 本文提出了一种改进 YOLOv3-SPP 的道路车辆目标检测算法, 首先对激活函数和 YOLOv3-SPP 的特征交互层进行了优化, 然后对边界框损失函数进行了改进。本文的主要贡献如下。

1) 优化激活函数。改变激活函数在负半轴的表达式, 提高模型的非线性建模能力, 增加神经网络分层的非线性学习能力。

2) 优化 DarkNet53 网络结构。对 DarkNet53 网络中 128×128 的特征图进行膨胀率为 1 和 2 的空洞卷积操作后, 再进行下采样并与输出的 64×64 特征图进行特征融合, 提高小目标的感受野, 增强模型对小目标的检测能力。

3) 改进边界框损失函数。针对 CIoU (complete intersection over union) 损失函数的缺陷, 以 DIOU-NMS (distance intersection over union-non-maximum suppression) Loss 为边界框损失函数, 使预测框的损失值更加合理, DIOU-NMS Loss 对预测框的回归更加稳定, 另外 DIOU-NMS Loss 也不会像 CIoU 损失函数一样, 让损失的计算更加复杂, 导致训练过程的计算开销较大。

1 改进的 YOLOv3-SPP 算法

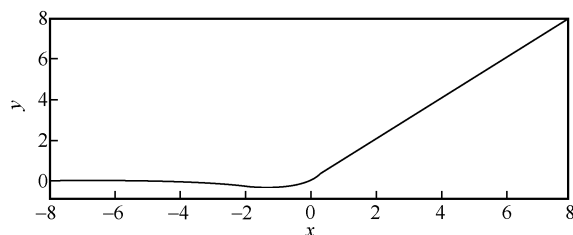
1.1 改进的激活函数 E-ReLU

在车辆检测数据集中, 如果车辆被遮挡或者受到光线的影响, YOLOv3-SPP 模型对车辆目标的提取性能将会受到影响。为了提高模型在复杂环境下车辆的检测性能, 本文使用 E-ReLU 激活函数来代替 YOLOv3-SPP 原本的 Leaky ReLU 激活函数^[15]。由于 Leaky ReLU 激活函数的导函数在 $(-\infty, 0)$ 上导函数恒为 0.01, 即表现为弱特征, 这会导致模型在

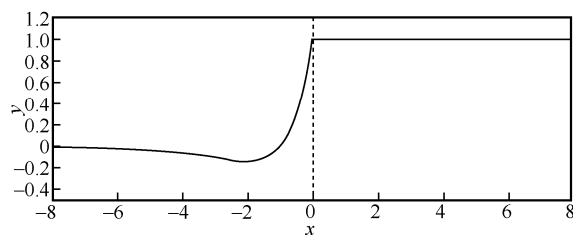
检测的过程中出现漏检情况, 因此就 Leaky ReLU 函数而言仍具有一定的局限性, 如果线性单元过大或者网络是循环结构, 就会导致梯度累积。由于在深度学习训练模型的过程中, 各层权值参数的更新与激活函数导数值的大小具有相关性, 在循环网络或者深层神经网络中参数会与激活函数的导数连乘相关, 因此对激活函数具有一定的要求。改进的 E-ReLU 激活函数的形式为

$$F(x) \begin{cases} x, & x > 0 \\ xe^x, & x \leq 0 \end{cases} \quad (1)$$

图 1 为 E-ReLU 的函数图像及其导函数图像。



(a) E-ReLU 的函数图像



(b) E-ReLU 的导函数图像

图 1 E-ReLU 的函数图像及其导函数图像

从图 1 可以看出, E-ReLU 激活函数满足以下条件: 函数为非线性函数; 函数具有一定的稀疏性, 减少了运算量; 函数输出具有正负值, 保证了参数更新的方向自由; 激活函数的导数没有过多为 0, 避免了“神经元坏死”现象的出现; 函数具有一定的弱饱和性, 即在模型训练的过程中, 刚开始训练时激活函数的导数应该较大, 能够加速模型收敛, 减少训练时间, 在模型训练的后半段激活函数的导数应逐渐趋于 0, 这样能平稳地使模型收敛至最优值, 不至于使模型跳过最优值而在一定的区间来回振荡; 导数的选取应当合适, 导数过小会出现参数更新时梯度消失的问题, 导数过大会出现参数更新时梯度爆炸的问题。

1.2 改进的 DarkNet53 网络结构

图片经过预处理后得到输入图像, 其尺寸为 $512 \times 512 \times 3$, 经过 YOLOv3-SPP 模型得到 3 个输出特征图, 分别为: $16 \times 16 \times 255$ 特征图用于大目标的

检测； $32 \times 32 \times 255$ 特征图用于中目标的检测； $64 \times 64 \times 255$ 特征图用于小目标的检测。YOLOv3-SPP 网络的思想是在进行目标检测时，对于图片中的小目标，当进行卷积的层数过多时，小目标在多次卷积后提取到的特征会很差，甚至提取不到特征，而对于较大的目标，进行的卷积操作越多，提取到大目标的特征信息就越细致，因此将 DarkNet53 网络最后生成的 16×16 大小的特征图作为大目标特征信息，然后对 16×16 的特征图进行上采样并与一定的卷积操作后得到新的 32×32 大小的特征图和 DarkNet53 网络中的 32×32 大小的特征图进行特征融合，将融合后的结果作为中目标的特征信息。这样结果作为中目标的特征相较于 DarkNet53 网络生成的中目标特征有更好的效果，对于小目标同样采用以上的操作思想。

YOLOv3-SPP 模型虽然利用空间金字塔池化操作并通过特征融合提升对小目标的检测性能，但是 YOLOv3-SPP 模型仍然存在一些问题，比如小目标检测不到或检测精度较低。由于 DarkNet53 中的 128×128 特征图含有更多目标信息，因此为了让模型获取到小目标更多的信息，本文将对 DarkNet53 中的 128×128 特征图进行特征提取，使小目标的信息更加丰富。为了扩大输出特征图中小目标的感受野，同时保证模型层数的增加不会导致小目标信息过多丢失，本文引入了空洞卷积（AC, atrous convolution）操作^[16]。空洞卷积是在标准卷积的基础上加入空洞，用来提升感受野并降低计算量。空洞卷积的卷积核中 2 个非零参数之间的水平或者竖直方向上的间隔为膨胀率，另外间隔中的值用 0 填充。图 2 为 3×3 标准卷积和空洞卷积。以 3×3 卷积核为例，图 2(a)卷积核中非零参数间隔为 1，因此若空洞卷积膨胀率为 1，则退化为标准卷积；图 2(b)卷积核中非零参数间隔为 2。

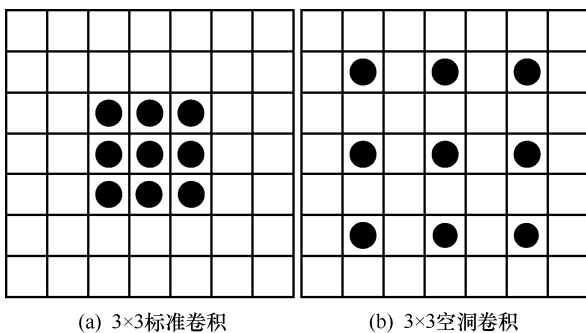


图 2 3×3 标准卷积和空洞卷积

图 3 为改进的 YOLOv3-SPP 网络结构，图 4 为改进的 YOLOv3-SPP 网络结构模块。

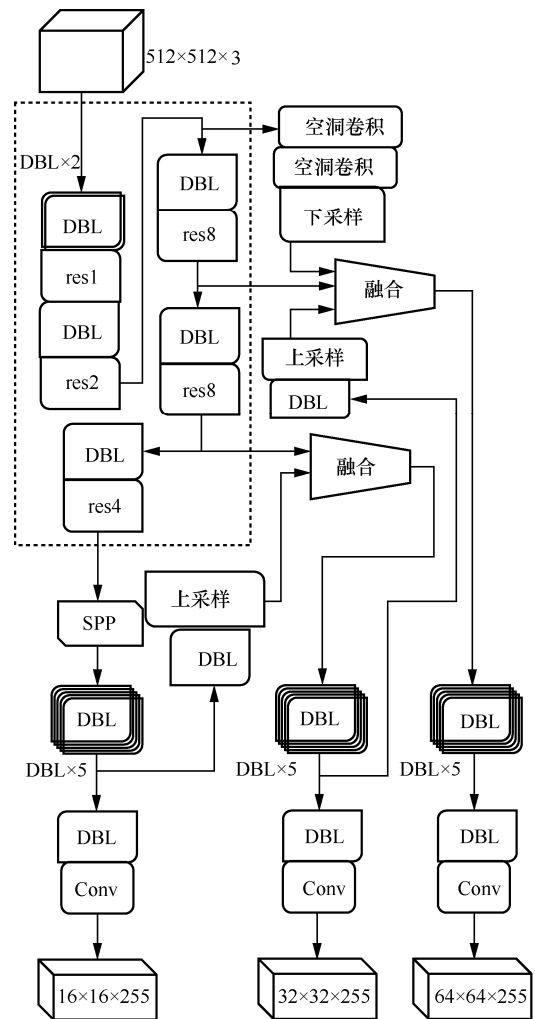


图 3 改进的 YOLOv3-SPP 网络结构

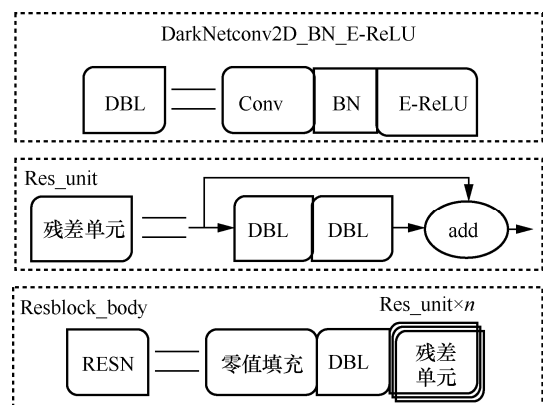


图 4 改进的 YOLOv3-SPP 网络结构模块

由于空洞卷积非零间隔中的值为 0，使用多个空洞卷积进行特征提取时，空洞卷积间隔与特征图对位置相乘结果为 0，可以理解为特征图中部分参数对目

标信息没有贡献,即使用单一空洞卷积会出现网格效应。本文采用混合空洞卷积模块,使用膨胀率为1的3×3标准卷积和膨胀率为2的3×3空洞卷积,将其串联成块,这种方法不仅增加了小目标的感受野,而且降低了计算量,同时也解决了在特征提取时由于网格效应所产生的信息流失严重的问题。将DarkNet53网络中128×128×128特征图通过混合空洞卷积模块和2倍下采样处理,得到64×64×128的特征图。将用于预测中目标的32×32×255特征图通过卷积模块和2倍上采样处理,得到64×64×512特征图。将上述64×64×128特征图、64×64×512特征图和DarkNet53中的64×64×256特征图这三者进行融合,得到新的64×64×256特征图,这种方法提升了检测小尺寸目标的准确率。

1.3 改进的边界框损失函数

对于位置信息的损失函数,YOLOv3-SPP采用在IOU^[17]、GIOW^[18]和DIOU^[19]的基础上改进而来的CIOU来表示,即

$$CIOU = IOU - \left(\frac{\rho^2(b, b^{gt})}{c^2} + \alpha \mathcal{G} \right) \quad (2)$$

$$\mathcal{G} = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2 \quad (3)$$

$$\alpha = \frac{\mathcal{G}}{(1 - IOU) + \mathcal{G}} \quad (4)$$

其中, b 、 h 和 w 表示预测框的中心点、高和宽; b^{gt} 、 h^{gt} 和 w^{gt} 表示真实框的中心点、高和宽; ρ 表示预测框和真实框两中心点之间的欧氏距离; c 表示能同时覆盖预测框和真实框的最小矩形的对角线长度。YOLOv3-SPP的位置信息损失为LCIOU=1-CIOU。

针对置信度损失函数,YOLOv3-SPP使用二元交叉熵^[20]来表示,如式(5)所示。

$$-\sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{obj} [C_i \log \widehat{C}_i + (1 - C_i) \log(1 - \widehat{C}_i)] - \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{noobj} [C_i \log \widehat{C}_i + (1 - C_i) \log(1 - \widehat{C}_i)] \quad (5)$$

其中, I_{ij}^{noobj} 表示第 i 个风格中的第 j 个预测框中是否存在目标,当第 i 个风格中的第 j 个预测框中不存在目标时, I_{ij}^{noobj} 为1,否则为0; λ_{noobj} 设置为0.5,作为权重因子; C_i 和 \widehat{C}_i 分别表示真实的置信度和预

测得到的置信度。

YOLOv3-SPP算法使用改进的Focal Loss来表示分类概率损失函数。Focal Loss主要针对样本不均衡问题,从损失角度提供另一种解决方法,其表达式为

$$L_{fl} = \begin{cases} -\alpha(1 - \widehat{p})^\gamma \log(\widehat{p}), & y=1 \\ -(1 - \alpha)\widehat{p}^\gamma \log(1 - \widehat{p}), & \text{其他} \end{cases} \quad (6)$$

令

$$p_t = \begin{cases} \widehat{p} & , y=1 \\ 1 - \widehat{p} & , \text{其他} \end{cases} \quad (7)$$

$$\alpha_t = \begin{cases} \alpha & , y=1 \\ 1 - \alpha & , \text{其他} \end{cases} \quad (8)$$

将Focal Loss表达式(6)统一为一个表达式,即

$$FL(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t) \quad (9)$$

其中, $\alpha \in [0, 1]$ 是一个超参数,用于平衡正负样本的权重, P_t 反映了预测类别与真实类别的接近程度, P_t 越大说明预测类别越接近真实类别,分类越准确,即 $(1 - P_t)^\gamma$ 能够降低易分类样本的损失贡献。

受文献[21]的启发,本文设计了DIOU-NMS Loss作为边框回归损失,对于定位损失,CIOU同时考虑了3种几何参数:重叠面积、中心点距离和长宽比。然而实验结果表明,采用CIOU的方式计算位置信息的损失并没有表现出比采用DIOU的方式计算位置信息的损失更好的效果,因为利用CIOU的方式计算边框损失,更倾向于往重叠区域增多的方向优化,而对于重叠区域较少的区域优化得并不好。本文采用DIOU-NMS的方式进行边框损失的计算。

YOLOv3-SPP中在模型预测阶段采用非极大值抑制(NMS, non-maxima suppression)算法^[22]来消除冗余的预测框,NMS算法的具体步骤如下。

1) 针对某一类别选取当前置信度最高的 S_i 的边框,记为 M 。

2) 计算该类别其他边框 B_i 和 M 的 $IOU(M, B_i)$,若 $IOU(M, B_i)$ 大于设定的阈值 ε ,就将该边框 B_i 的置信度置为0,并舍弃该边框;若 $IOU(M, B_i)$ 小于设定的阈值 ε ,则该边框 B_i 的置信度不变,并保留该边框。

3) 从除边框 M 外的剩余的边框中找出置信度

最高的一个边框, 重复步骤 2), 最终完成该类别的 NMS 的计算。

4) 循环执行步骤 1)~步骤 3), 完成所有类别的 NMS 计算。

NMS 对边框置信度的计算式为

$$S_i = \begin{cases} S_i, & \text{IOU}(M, B_i) < \varepsilon \\ 0, & \text{IOU}(M, B_i) \geq \varepsilon \end{cases} \quad (10)$$

$$S_i = \begin{cases} S_i, & \text{IOU} - R_{\text{DIOU}}(M, B_i) < \varepsilon \\ 0, & \text{IOU} - R_{\text{DIOU}}(M, B_i) \geq \varepsilon \end{cases} \quad (11)$$

$$R_{\text{DIOU}} = \frac{\rho^2(b, b^{\text{gt}})}{c^2} \quad (12)$$

由式(10)~式(12)可知, DIOU-NMS 考虑了 IOU 的值来确定边框是否被删除。改进后的 DIOU-NMS 能够充分判断重叠较高的 2 个观测边界框是否属于同一个目标, 从而进行边框抑制。本文采用 ε 为 0.42, 进一步降低了算法的漏检率。

2 实验结果和分析

2.1 数据集和评定指标

本文通过 PASCAL VOC2012 数据集和自行采集的实景数据对所提算法针对车辆的检测效果进行验证。PASCAL VOC2012 是标准数据集, 其包含小汽车、公交车、摩托车等多个类别, 本文仅对 PASCAL VOC2012 中小汽车、公交车、摩托车这三类车辆进行训练。在实验过程中, 本文将 PASCAL VOC2012 数据集的 5 717 张图片作为训练集, 并通过缩放、旋转和镜像等图像处理方式扩充训练集的数量, 进一步提高模型的性能。本文自行采集的测试数据集共包含 500 张图片, 用于训练模型的验证。

由于目标检测的目的是在图像中找到所有感兴趣的目标, 不同于分类和回归问题, 目标检测要确定目标在图像中的位置(定位)以及识别物体的类别(分类)。本文采用标准平均准确率(AP, average precision)和平均均值准确率(mAP, mean average precision)作为检测评价指标来进行模型检测性能的评定。AP 可以衡量模型在单个目标类别上的检测效果, mAP 则用来评价模型对所有检测目标类别的平均检测效果。目标检测的准确率(P, Precision)和召回率(R, Recall)的定义式分别为

$$P = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FP}}} \quad (13)$$

$$R = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FN}}} \quad (14)$$

其中, N_{TP} 表示将目标检测为正样本且检测正确的个数, N_{FP} 表示将目标检测为正样本但检测错误的个数, N_{FN} 表示将目标检测为负样本但检测错误的个数。

由于 PR 曲线是由 Precision 和 Recall 来确定的, 因此使用 PR 曲线可以表示模型对同类目标的 Recall 与 Precision 的权衡优化。AP 表示 PR 曲线的积分计算结果, 即 PR 曲线与坐标轴所围成的面积

$$\text{AP} = \int_0^1 P(r) dx, \quad r \in (0, 1) \quad (15)$$

mAP 是对所有类别的 AP 值进行求和后除以总类别数, 如式(16)所示。

$$\text{mAP} = \frac{1}{n} \sum_{i=1}^n \text{AP}_i \quad (16)$$

AP 和 mAP 可以用来衡量算法检测的准确程度, 而用来衡量模型检测的速度的是每秒检测图像的帧数(FPS, frame of per second)。FPS 的计算方法为: 首先将模型处理每张图片所用时间进行累加得到总的检测时间, 然后用模型检测图片的总张数除以总的检测时间, 定义以上计算得到的结果为模型的 FPS。FPS 的定义式为

$$\text{FPS} = \frac{N}{\sum_{i=1}^n t_i} \quad (17)$$

2.2 训练参数

硬件方面数据处理实验中央处理器采用 AMD 公司的 AMD Ryzen 7 4800H with Radeon Graphics 2.90 GHz, 显卡采用英伟达的 NVIDIA GTX1650ti; 软件方面采用 CUDA10.2 和 CUDNN8.70 加速库对显卡进行加速操作, 小组利用 Python 语言进行编程, 采用 Python3.9 版本中的 OpenCV 库进行图像处理, 深度学习的架构为 PyTorch1.7.1, 软件运行在 Windows 10 家庭中文版的操作系统上, 电脑运行内存为 DDR4 3200 MHz EEC Reg 32 GB, 本文实验所用硬件和软件的具体参数配置如表 1 所示。

表 1 参数配置

| 名称 | 相关配置 |
|---------|---|
| 中央处理器 | AMD Ryzen 7 4800H with Radeon Graphics 2.90 GHz |
| 内存 | DDR4 3200 MHz EEC Reg 32 GB |
| GPU | NVIDIA GTX1650ti |
| GPU 加速库 | CUDA10.2, CUDNN8.7.0 |
| 操作系统 | Windows 10 家庭中文版 |
| 数据处理 | Python3.9, OpenCV |
| 深度学习框架 | PyTorch1.7.1 |

模型训练过程, 初始输入图片尺寸为 512×512, 程序迭代次数为 1 000, 以 16 张图片为一批次, 归一化处理后进行输入, 即对输入的 16 张图片为一单元进行批量归一化处理。实验在迭代 1 000 次前, 通过调参, 对模型进行了多次迭代次数为 50 的训练, 发现当 IOU 的阈值为 0.7、置信度的阈值为 0.45 时, 模型损失值下降最快。本文对网络训练的部分参数如表 2 所示。

表 2 网络训练部分参数

| 参数名称 | 预设值 | 作用 |
|--------|---------|--------------------------|
| 图片大小 | 512×512 | 保持相同的图片预设处理方式, 发挥模型的推理效果 |
| 迭代次数 | 1 000 | 整个数据集迭代次数 |
| 批大小 | 16 | 对数据集划分, 批量处理数据 |
| IOU 阈值 | 0.7 | 目标检测可信度 |
| 置信度阈值 | 0.45 | 判断是否存在检测目标 |

2.3 与其他方法的性能对比

为了验证本文提出的改进的 YOLOv3-SPP 车辆检测算法的性能, 本节将其与未改进的 YOLOv3-SPP 算法进行对比。2 个算法均在 PASCAL VOC2012 数据集上进行训练, 分别在训练集上迭代

500 次, 完成训练后保存模型, 在自行收集的测试集上进行测试, 并进行小汽车、公交车和摩托车的 AP 计算和 mAP 计算, 另外, 计算模型在测试集上的检测速度 (FPS), 因为模型在训练过程中会自动产生小汽车、公交车和摩托车相应的 Precision 和 Recall, 根据三者的 Precision 和 Recall, 绘制出 PR 曲线, 以 PR 曲线与坐标轴所围成的面积为该模型识别 3 个类别的标准 AP, 三者 AP 值取平均为 mAP, 用 AP、mAP 和 FPS 来衡量模型检测性能的好坏, 接下来对单阶段目标检测算法 FE-CNN^[12]、NanoDet^[13]、SSD^[23]、Faster RCNN^[24]、RetinaNet^[25]、Proposed CNN^[26]、传统的 YOLOv3-SPP^[27]和改进的 YOLOv3-SPP 算法进行对比, 对比结果如表 3 所示。

从表 3 可以看出, 本文在自行收集的 500 张图片上对模型进行测试, 改进的 YOLOv3-SPP 比传统的 YOLOv3-SPP 虽然检测速度有所下降, 但 mAP 提升了 1.79%, 并且改进的 YOLOv3-SPP 算法训练的汽车、公交车和摩托车这 3 个类别各自的 AP 和 mAP 较 FE-CNN^[12]、NanoDet^[13]、SSD^[23]、Faster RCNN^[24]、RetinaNet^[25]、Proposed CNN^[26]和传统的 YOLOv3-SPP^[27]算法均有提升。这说明本文改进的 YOLOv3-SPP 算法在车辆检测任务上实现了精度的提升, 证明了改进的有效性。

图 5 为基于 PASCAL VOC2012 迭代 1 000 次训练的网络学习率和损失值变化曲线。从图 5 可以看出, 基于 PASCAL VOC2012 数据集训练的模型, 在本文参数设定模式下, 随着迭代次数的增加, 算法的学习率和损失值在不断减小, 当迭代次数超过 930 次后, 系统的学习率和损失值基本稳定, 另外, 实验结果表明, 当更新模型的权重参数为 930 次后训练得到的权重参数, 模型的检测效果也趋于稳定。

表 3 几种算法对比结果

| 指标 | mAP | AP | | | FPS/Hz |
|--------------------------------|--------|--------|--------|--------|--------|
| | | 小汽车 | 公交车 | 摩托车 | |
| FE-CNN | 80.31% | 82.60% | 79.64% | 78.69% | 25.69 |
| NanoDet | 87.59% | 88.72% | 85.33% | 88.72% | 17.61 |
| SSD ^[23] | 76.98% | 80.21% | 73.28% | 77.45% | 11.82 |
| Faster-RCNN ^[24] | 79.63% | 84.67% | 69.94% | 84.28% | 29.56 |
| RetinaNet ^[25] | 81.92% | 85.30% | 77.26% | 83.19% | 35.24 |
| Proposed CNN ^[26] | 84.37% | 86.51% | 81.92% | 84.68% | 47.63 |
| 传统的 YOLOv3-SPP ^[27] | 88.66% | 92.59% | 84.17% | 89.23% | 52.38 |
| 改进的 YOLOv3-SPP | 90.45% | 93.47% | 85.56% | 92.31% | 49.54 |

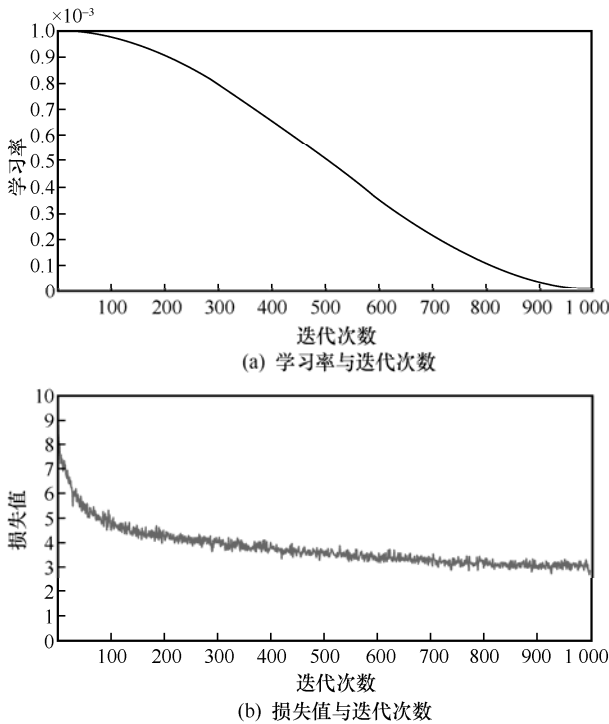


图 5 学习率、损失值与迭代次数的变化曲线

2.4 实验结果

为了对比本文改进的 YOLOv3-SPP 算法与其他主流算法在不同尺度上的检测性能，本节实验按照图像目标检测的划分方法，将目标尺度小于或等于 32×32 的物体设为小目标，尺度大于 32×32 而小于或等于 96×96 的物体设为中目标，将尺度大于 96×96 的物体设为大目标。由于道路上常见的车辆中小汽车、公交车和摩托车的占比较高，因此模型训练以这三类车辆为训练集和测试集，而这三类车辆在模型的测试集上能够直接产生 AP、mAP 和 FPS，因此基于产生的评价数据，本文在这三类车辆的测试集上以尺度大小来评价模型的好坏，并与 FE-CNN^[12]、NanoDet^[13]、SSD^[23] 算法和 Faster RCNN^[24] 算法，RetinaNet^[25] 和 Proposed CNN^[26] 进行比较。

本文中自行采集的测试数据集共包含 100 张交通实景图片，为了验证训练得到的模型性能，将这些图片又分别进行了旋转、缩放、扭曲和明暗处理得到了 400 张图片，最终整理出了 500 张图片作为模型性能的测试集。将这 500 张图片组成的数据集分别输入改进前后的 YOLOv3-SPP 模型进行测试，提取出小目标、中目标和大目标的准确率和召回率，得到 PR 曲线，需要注意的是，小、中、大三类目标中都含有小汽车、公交车和摩托车这三类车

辆，因此这里提取到的车辆目标的准确率和召回率仅与检测目标的尺寸有关，而与车辆目标的类别无关。图 6 为 FE-CNN^[12]、NanoDet^[13]、SSD^[23] 算法、Faster RCNN^[24] 算法、RetinaNet^[25]、Proposed CNN^[26]、传统的 YOLOv3-SPP^[27] 算法和改进的 YOLOv3-SPP 算法在测试集上不同目标尺度的 PR 曲线。

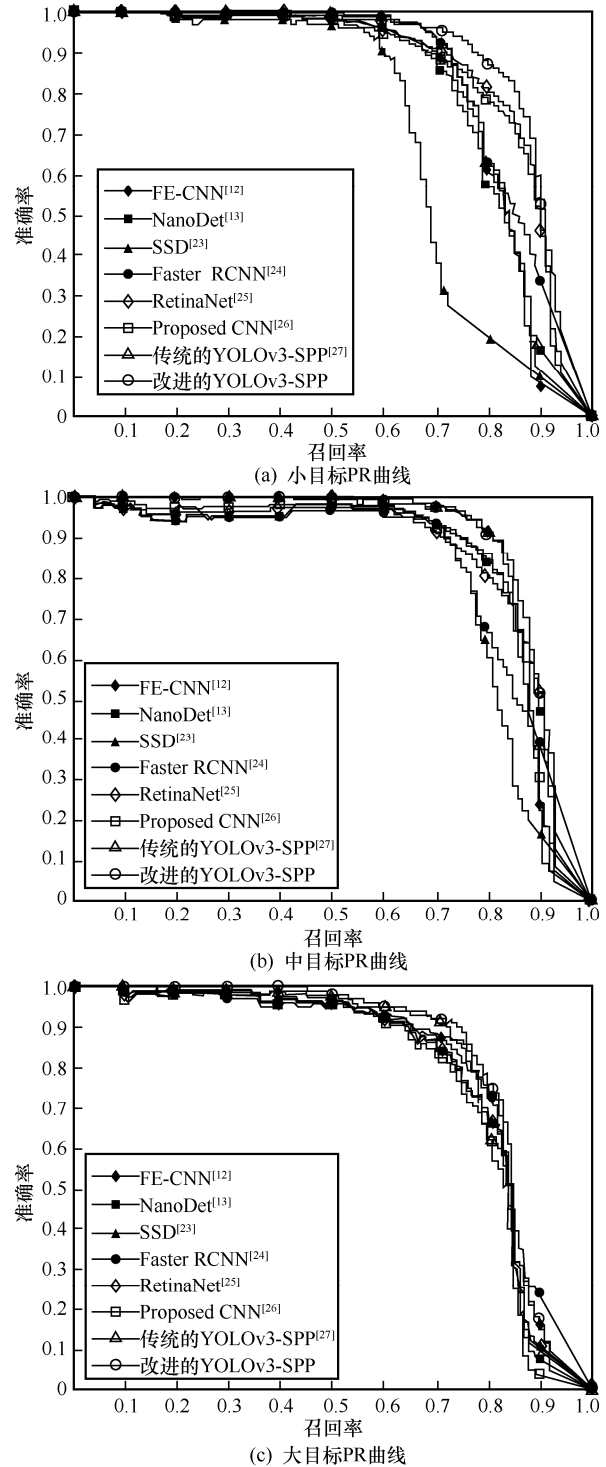


图 6 不同算法在测试集上不同目标尺度的 PR 曲线

图 6 直观地展示了 FE-CNN^[12]、NanoDet^[13]、SSD^[23]算法、Faster RCNN^[24]算法、RetinaNet^[25]、Proposed CNN^[26]、传统的 YOLOv3-SPP^[27]算法和改进的 YOLOv3-SPP 算法在本文采集的测试数据集上的检测效果。FE-CNN^[12]算法通过简单有效的边缘特征融合方法提高了识别精度和模型的收敛速度,但该算法在真实交通场景的车辆识别实验中,其响应速度和实时性有待提高。NanoDet^[13]算法最大限度地压缩网络卷积层数,提高了计算速度,但其就中小目标的特征提取并不理想。SSD^[23]算法主要分为 3 个模块,第一个模块是基础网络,用于提取低尺度的特征映射图;第二个模块是辅助卷积层,用于提取高尺度的特征映射图;第三个模块是预测卷积层,用于输出特征映射图的位置信息和分类信息,SSD 算法使用多尺度特征进行目标检测,极大地提高了检测精度,但 SSD 算法对小目标的识别性能仍比较差,达不到 Faster RCNN^[24]算法的水准。Faster RCNN^[24]算法主要分为 4 个模块,第一个模块是特征网络,用于提取特征;第二个模块为区域候选网络,用于生成候选框;第三个模块为兴趣域池化,用于提取第一个模块中特征图的坐标和生成第二个模块候选框的坐标;第四个模块进行目标分类和候选框的回归。Faster RCNN 在提高了结果准确度的同时,大大降低了参数数量。RetinaNet^[25]算法通过改进的损失函数 Focal Loss 缓解了前景和背景样本不均衡的问题,在一定程度上提升了模型的检测性能,但 RetinaNet^[25]的兼容性较差,其中锚框的比例调整比较复杂。文献[26]中通过对多个网络模型的训练和测试,表明 Proposed CNN^[26]算法在自主检测性能方面优于 PointPillars 和 YOLOv4,但 Proposed CNN^[26]算法存在数据遮挡造成的假阴性检测问题,因此并没有实现有效和高效的特征融合和特征提取。而本文改进的 YOLOv3-SPP 算法对中目标和大目标的检测性能 mAP 并没有较大提升,但对小目标的检测性能有所提升,另外实验检测结果表明,模型对大目标和中目标的检测仅提升了类别的置信度,而对小目标的检测降低了漏检率。

利用本文自行收集的数据集分别测试传统的 YOLOv3-SPP 和改进的 YOLOv3-SPP 的检测效果,可以发现,传统的 YOLOv3-SPP 测试结果中出现了一些重复检测和漏检的问题,改进的 YOLOv3-SPP 测试结果中检测较远处的车辆目标

依然存在检测不到的问题,但相较于传统的 YOLOv3-SPP 而言,改进的 YOLOv3-SPP 在复杂的交通场景下也能够保持很好的车辆检测效果,检测图像中近处密集的车辆和有干扰物遮挡的车辆都能够很好地被检测出来且没有误检,另外,检测出远处目标的数量和识别出车辆类别的置信度均有较大的提升。

3 结束语

基于 YOLOv3-SPP 算法,本文提出一种针对车辆检测任务改进的算法,实验结果表明,改进的 YOLOv3-SPP 算法能够在检测速度下降的情况下提升检测精度。本文对激活函数进行优化,改善了在复杂环境下检测不佳的问题;提出了一种新的损失函数 DIOU-NMS Loss 作为边界框损失函数,考虑边界框的重要点和宽高比使网络模型回归更加准确,进一步增强了网络的表示能力和学习能力;另外,改进 YOLOv3-SPP 的 DarkNet53 主干网络结构,引入空洞卷积模块,增加了小目标的感受野,降低了计算量,提高了模型对小目标和被遮挡车辆目标的特征提取能力和检测精度,解决了在检测时较远车辆呈现小尺度特征而出现检测精度低或者漏检的问题。但本文所提算法仍有许多不足,如在实验过程中由于参数调节的不合适容易出现梯度消失现象,后续研究将在保证检测精度的情况下提高检测速度,并消除梯度消失现象。另外,车辆检测在复杂多变的环境,比如受光照影响的夜间场景、隧道场景和雾霾天气或者车辆高速移动的场景,本文所提算法在实时性和检测精度上仍具有非常大的挑战,这也是后续的研究工作要解决的技术难点。

参考文献:

- [1] 刘耀辉. 基于深度学习的车辆检测研究[D]. 南京: 南京邮电大学, 2022.
LIU Y H. Research on vehicle detection based on deep learning[D]. Nanjing: Nanjing University of Posts and Telecommunications, 2022.
- [2] JABRI S, SAIDALLAH M, BELRHITI E A A, et al. Moving vehicle detection using Haar-like, LBP and a machine learning adaboost algorithm[C]//Proceedings of IEEE International Conference on Image Processing, Applications and Systems (IPAS). Piscataway: IEEE Press, 2018: 121-124.
- [3] DAVID H, ATHIRA T A. Improving the performance of vehicle detec-

- tion and verification by Log Gabor filter optimization[C]//Proceedings of Fourth International Conference on Advances in Computing and Communications. Piscataway: IEEE Press, 2014: 50-55.
- [4] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2014: 580-587.
- [5] GU R, BILLINGE S J L, DU Q. A fast two-stage algorithm for non-negative matrix factorization in smoothly varying data[J]. *Acta Crystallographica Section A, Foundations and Advances*, 2023, 79(2): 203-216.
- [6] ZHAO W Q, XU M F, CHENG X F, et al. An insulator in transmission lines recognition and fault detection model based on improved faster RCNN[J]. *IEEE Transactions on Instrumentation and Measurement*, 2021, 70: 1-8.
- [7] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//European Conference on Computer Vision. Berlin: Springer, 2016: 21-37.
- [8] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2016: 779-788.
- [9] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2017: 6517-6525.
- [10] SHEN J Q, LIU N Z, SUN H, et al. Vehicle detection in aerial images based on lightweight deep convolutional network and generative adversarial network[J]. *IEEE Access*, 2019, 7: 148119-148130.
- [11] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers[C]//European Conference on Computer Vision. Berlin: Springer, 2020: 213-229.
- [12] QIU L R, ZHANG D B, TIAN Y, et al. Deep learning-based algorithm for vehicle detection in intelligent transportation systems[J]. *The Journal of Supercomputing*, 2021, 77(10): 11083-11098.
- [13] SUN X C, XIA T R, WANG T Y, et al. A lightweight neural network for vehicle object detection[C]//Proceedings of IEEE 4th International Conference on Civil Aviation Safety and Information Technology (ICCASIT). Piscataway: IEEE Press, 2022: 1103-1106.
- [14] WU Y F, GUAN X R, ZHAO B Y, et al. Vehicle detection based on adaptive multimodal feature fusion and cross-modal vehicle index using RGB-T images[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2023, 16: 8166-8177.
- [15] XU J, LI Z S, DU B W, et al. ReLUpex made more practical: leaky ReLU[C]//Proceedings of the IEEE Symposium on Computers and Communications (ISCC). Piscataway: IEEE Press, 2020: 1-7.
- [16] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(4): 834-848.
- [17] 李琦林. 面向目标检测的边界框回归改进算法研究与实现[D]. 长春: 吉林大学, 2023.
- LI Q L. Research and implementation of improved bounding box regression algorithm for target detection[D]. Changchun: Jilin University, 2023.
- [18] JIANG K, ITOH H, ODA M, et al. Gaussian affinity and GIoU-based loss for perforation detection and localization from colonoscopy videos[J]. *International Journal of Computer Assisted Radiology and Surgery*, 2023, 18(5): 795-805.
- [19] ZHENG Z H, WANG P, LIU W, et al. Distance-IoU loss: faster and better learning for bounding box regression[J]. *arXiv Preprint, arXiv: 1911.08287*, 2019.
- [20] CAO J, SU Z, YU L Y, et al. Softmax cross entropy loss with unbiased decision boundary for image classification[C]//Proceedings of Chinese Automation Congress (CAC). Piscataway: IEEE Press, 2018: 2028-2032.
- [21] REZATOFIGHI H, TSOI N, GWAK J, et al. Generalized intersection over union: a metric and a loss for bounding box regression[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2019: 658-666.
- [22] 赵文清, 严海, 邵绪强. 改进的非极大值抑制算法的目标检测[J]. *中国图象图形学报*, 2018, 23(11): 1676-1685.
- ZHAO W Q, YAN H, SHAO X Q. Object detection based on improved non-maximum suppression algorithm[J]. *Journal of Image and Graphics*, 2018, 23(11): 1676-1685.
- [23] NI J J, SHEN K, CHEN Y, et al. An improved SSD-like deep network-based object detection method for indoor scenes[J]. *IEEE Transactions on Instrumentation and Measurement*, 2023, 72: 1-15.
- [24] YANG L, ZHONG J H, ZHANG Y, et al. An improving faster-RCNN with multi-attention ResNet for small target detection in intelligent autonomous transport with 6G[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2023, 24(7): 7717-7725.
- [25] GAO P, TIAN J W, TAI Y, et al. Vehicle detection with bottom enhanced RetinaNet in aerial images[C]//Proceedings of IEEE International Geoscience and Remote Sensing Symposium. Piscataway: IEEE Press, 2020: 1173-1176.
- [26] ZHOU S L, XU H, ZHANG G H, et al. Leveraging deep convolutional neural networks pre-trained on autonomous driving data for vehicle detection from roadside LiDAR data[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2022, 23(11): 22367-22377.

[27] YANG K L, JIAO Z Y, LIANG J X, et al. An application case of object detection model based on YOLOv3-SPP model pruning[C]//Proceedings of IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA). Piscataway: IEEE Press, 2022: 578-582.



李林（1986- ），男，安徽巢湖人，博士，国家计算机网络应急技术处理协调中心工程师，主要研究方向为网络信息安全等。

[作者简介]



王涛（1986- ），男，陕西渭南人，博士，北京信息科技大学副教授、硕士生导师，主要研究方向为车路协同感知、计算机视觉、雷达数据处理等。



何振学（1988- ），男，山东泰安人，博士，河北农业大学副教授、博士生导师，主要研究方向为智能装备、计算机视觉等。



冯浩（1996- ），男，河南驻马店人，北京信息科技大学硕士生，主要研究方向为车路协同感知、计算机视觉、雷达数据处理等。



傅奕茗（2000- ），男，江苏徐州人，北京信息科技大学硕士生，主要研究方向为智能感知、点云数据处理、激光 SLAM 等。



秘蓉新（1984- ），女，北京人，国家计算机网络应急技术处理协调中心助理研究员，主要研究方向为人工智能、区块链、网络信息安全等。



吴姝（1997- ），女，山西大同人，北京信息科技大学硕士生，主要研究方向为自动驾驶智能感知、激光雷达点云数据处理等。