

基于改进 YOLOv7 的交通路口目标识别算法

江 晟¹, 张仲义^{1,2}, 汪宗洋², 于 晴¹

(1. 长春理工大学 物理学院, 长春 130022; 2. 江苏集萃深度感知技术研究所, 江苏 无锡 214000)

摘要: 针对交通路口车辆目标检测算法存在精确度低、少检、漏检等问题, 提出一种基于改进 YOLOv7 的交通路口目标识别算法. 该算法首先利用前馈式卷积注意力机制 CBAM 从通道注意力和空间注意力两者提升网络对关键特征的注意力, 提高网络的运行速率, 优化网络的特征提取能力; 其次采取空间层到深度层连接全维动态卷积组成一个新的学习模块, 以此结构改进 YOLOv7 特征学习方式, 提升特征表达能力; 最后在实际采集的交通路口数据集上进行实验. 实验结果表明, 该方法在对应数据集上平均精度达到 96.1%, 训练耗时降低至 16.71 h, 因此针对交通路口小目标检测有明显的识别优势.

关键词: 深度学习; 目标检测; 卷积神经网络; 注意力机制; 全维动态卷积

中图分类号: TP301.6 **文献标志码:** A **文章编号:** 1671-5489(2024)03-0665-09

Target Recognition Algorithm of Traffic Intersection Based on Improved YOLOv7

JIANG Sheng¹, ZHANG Zhongyi^{1,2}, WANG Zongyang², YU Qing¹

(1. School of Physics, Changchun University of Science and Technology, Changchun 130022, China;

2. Institute of Deep Perception Technology, Wuxi 214000, Jiangsu Province, China)

Abstract: Aiming at the problems of low accuracy, under-detection, and missed detection in the vehicle target detection algorithm at traffic intersections, we proposed a target recognition algorithm of traffic intersection based on improved YOLOv7. Firstly, the algorithm used the feed-forward convolutional attention mechanism CBAM to enhance the network's attention to key features from both channel attention and spatial attention, improve the network's running speed, and optimize the network's feature extraction capabilities. Secondly, a new learning module was formed by connecting the spatial layer to depth layers to form a full-dimensional dynamic convolution, which improved the YOLOv7 feature learning method and enhanced the feature expression ability. Finally, the experiments were conducted on the actual collected traffic intersection dataset. The experimental results show that the proposed method achieves an average accuracy of 96.1% on the corresponding dataset, and the training time is reduced to 16.71 h. Therefore, it has obvious recognition advantages for small target detection at traffic intersections.

Keywords: deep learning; target detection; convolutional neural network; attention mechanism; full-dimensional dynamic convolution

收稿日期: 2023-06-12.

第一作者简介: 江 晟(1985—), 男, 汉族, 博士, 副教授, 从事多维智能感知与协同控制的研究, E-mail: js1985_cust@163.com.

通信作者简介: 张仲义(1999—), 男, 汉族, 硕士, 从事光电传感与探测技术的研究, E-mail: yuanqing288@icloud.com.

基金项目: 吉林省科技发展计划重点研发项目(批准号: 20210203214SF).

近年来,随着神经网络技术的不断深化^[1],基于深度卷积神经网络提出的深度学习算法因其精确度高、准确度高、使用方便,具备从静态或动态视频中快速检测出车辆目标及其相关特征信息^[2]等特点,实现了对路口交通流进行快速检测的良好功能,逐渐成为主流目标检测算法^[3].

但在目前的交通流量目标检测任务中,存在大量的远景小目标,其在图像中的分辨率和信息有限,使得对小目标的精确检测识别成为一项具有挑战性的任务.为更好地解决该问题,欧阳继红等^[4]提出了一种改进多尺度特征融合的目标检测方法,该方法将网络的学习参数作为特征间的平衡因子进行加权融合,从而提高模型多尺度特征融合能力,有效提高了模型的检测精度;逢晨曦等^[5]通过采用注意力改进的自适应空间特征融合策略生成金字塔形表示,在不影响小目标识别的情况下,提高了中、大目标的检测精度;姚庆安等^[6]用改进的带孔空间金字塔池化增加多维度目标的分割,用注意力细化模块捕获线索特征学习,并加入基于注意力机制的特征融合有队列地监督重要通道特征的学习,引导高、低阶特征融合,在面对多像素目标语义分割时效果较好;Singh 等^[7]提出了 SNIPER 模型,通过只处理目标周围的背景区域减少训练时间,提高 SNIP^[8]的训练效率,但这种多尺度特征训练的方法引用了大量的跨步卷积或池化层,导致细粒度信息的丢失和较低的特征学习能力.因此, Sunkara 等^[9]提出了 SPD-Conv 结构,通过引入由空间层到深度层结构连接非跨步卷积层的结构代替跨步卷积和池化,提升了对低分辨率场景的检测能力,降低了网络复杂度,提升了模型运行速度,但网络中的所有卷积均是静态卷积,静态卷积核共享权值^[10],导致模型特征学习能力下降.针对静态卷积,Chen 等^[11]在基于 CondConv^[12]的基础上,通过使用一组 K 个并行卷积核,对每个单独输入的 x ,根据关注度动态聚合多个并行卷积核,共享输出通道,相比于 CondConv,在注意力机制上采取 GAP+FC+ReLU+FC+Sofmax 的方法,提取注意力分配给不同的卷积核上,提升了特征表示能力,但其忽略了卷积核的其他维度. Li 等^[13]提出了全维动态卷积,通过采取多维注意力机制和并行策略,赋予了卷积核动态属性,有更强的特征表达能力,但引用过多的动态卷积后会提升模型的大小,增加网络复杂度.针对上述问题,本文采用一阶段算法 YOLO (you only look once) 系列最新的 YOLOv7^[14]网络,在原网络结构上进行改进,将改进后的网络与原有网络进行检测对比,对比结果表明,改进后的网络相比原网络的检测效果有明显提升,应用价值更广.

1 YOLOv7 网络模型

YOLOv7 是最新 YOLO 系列的检测器,相同体量下比 YOLOv5 精度更高,速度快 120%,其传输帧数达 161 帧/s;同时在 GPU V100 上进行测试,平均精度为 56.8%的模型可达 30 帧/s.无论是速度还是精度,都取得了较优异的检测效果. YOLOv7 的网络结构由输入 (Input)、骨干网络 (Backbone)、中间网络 (Neck) 和检测头 (Head) 四部分组成,其网络结构如图 1 所示.

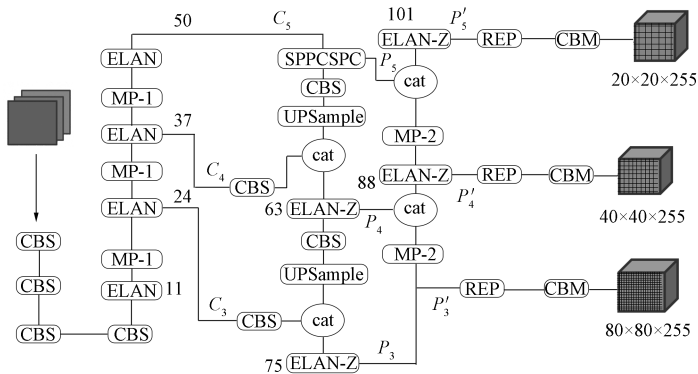


图 1 YOLOv7 网络结构

Fig. 1 Network structure of YOLOv7

1.1 骨干网络

YOLOv7 的骨干网络由 CBS 层、ELAN 层和 MP 层组成. 图像输入后首先经过 4 个 CBS 卷积层,

CBS 层由卷积(Conv)层、归一化(BN)层和损失函数(SiLu)层组成。

当特征图经过 4 个 CBS 后, 其特征尺寸由 $640 \times 640 \times 32$ 压缩至 $160 \times 160 \times 128$, 然后送至 ELAN 层^[8], ELAN 层是一个高效的网络结构, 由多个 CBS 构成, 第一条分支是经过一个 CBS 卷积改变通道数量, 第二条分支首先经过一个 CBS 卷积层改变通道数量, 其次经过 4 个 CBS 卷积层进行特征提取, 然后将结果叠加得到最后的特征提取结果. ELAN 层输入前后特征图尺寸大小保持不变, 通道数量增加至输入的一倍. 其通过控制最短和最长的梯度路径, 使网络能充分学习特征, 具有很强的鲁棒性. ELAN 的结构如图 2 所示。

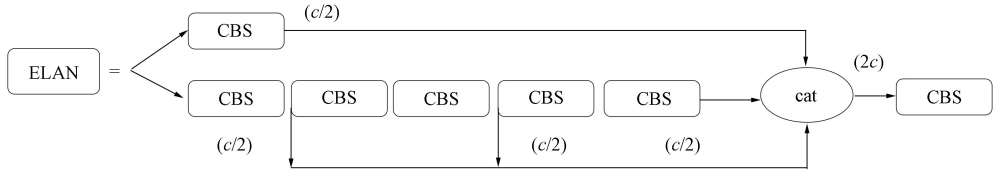


图 2 ELAN 的结构

Fig. 2 Structure of ELAN

经过 ELAN 层后特征图将会送入 MP 层, MP-1 层通道数保持不变, MP-2 层通道数增加至原来的一倍. MP 层主要由最大池化层(Maxpool)和 CBS 卷积层组成, 第一条分支是特征图先经过最大池化层, 进行下采样操作, 然后经过 CBS 卷积层, 改变通道数量; 第二条分支先经过一个卷积层改变通道数量, 然后进行下采样操作, 最后将结果叠加在一起。

经过 Backbone 网络后, 将会输出 3 个特征图 C_3, C_4, C_5 , 其大小分别为 $80 \times 80 \times 512, 40 \times 40 \times 1024, 20 \times 20 \times 1024$.

1.2 中间网络与检测头网络

YOLOv7 的 Neck 部分可视为一个 pafpn 结构, 将 Backbone 网络生成的 3 个特征图按其尺寸大小分别进行多尺度特征融合. 首先通过 SPPCSPC 模块, SPPCSPC 模块利用不同尺度的最大池化获得不同的感受野, 用于区分大目标和小目标, 其结构如图 3 所示。

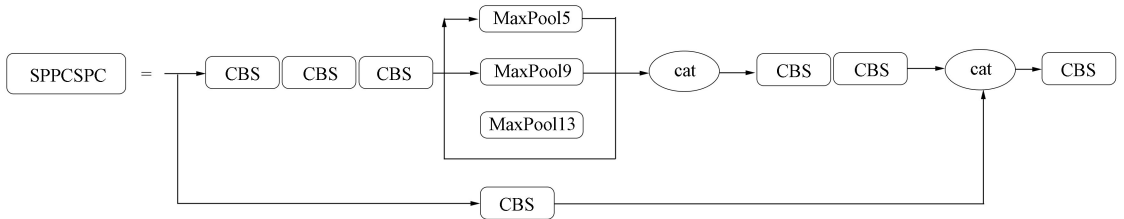


图 3 SPPCSPC 结构

Fig. 3 Structure of SPPCSPC

然后自顶向下经过 CBS 层、UPSample 层、ELAN-Z 层等依次与 C_4 和 C_3 进行融合, 得到 P_3, P_4, P_5 , 其中 UPSample 层是一个上采样模块, 它所使用的采样方式是最近邻插值 nearest 算法, 其结构如图 4 所示。

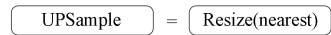


图 4 UPSample 结构

Fig. 4 Structure of UPSample

而 ELAN-Z 结构与 ELAN 结构类似, ELAN-Z 结构如图 5 所示. 由图 5 可见, 它的第二条分支所选取的输出数量不同, ELAN 模块选取 3 个输出进行最后相加, 而 ELAN-Z 选取 5 个输出进行相加, 从而提高特征提取的能力; 然后自底向上经由 MP-2 层、ELAN-Z 层与 P_4, P_5 进行融合, 最后 pafpn 将输出 P'_3, P'_4, P'_5 . 通过 RepConv 层调整通道数, 用卷积去预测 objectness, class 和 bbox 三部分。

RepConv 层有两种情况: 一种是训练(train), 另一种是推理(deploy). 在训练时, 其结构有 3 条分支: 分别是 3×3 卷积层+BN 层, 作用是特征提取; 1×1 卷积层+BN 层, 作用是特征平滑; 最后一条分支是 Identity, 由一个 BN 层构成. 最后将 3 条分支叠加在一起。

在推理时, 将训练时的第二条和第三条分支都转化为 3×3 的卷积, 然后进行矩阵融合, 权重相

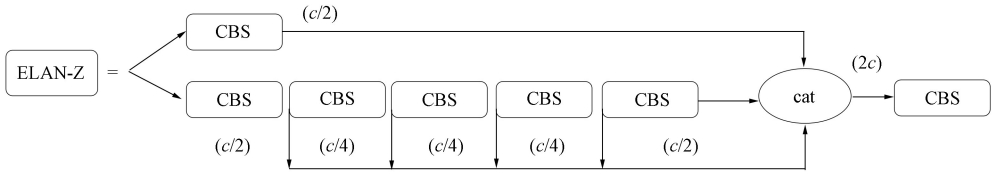


图 5 ELAN-Z 结构

Fig. 5 Structure of ELAN-Z

加, 3 条分支融合成 1 条分支, 其中只包含 1 个 3×3 卷积, 结构如图 6 所示.

在 YOLOv7 的 Backbone 和 Neck 中都不同程度地采用了跨步卷积层和 Maxpooling 层. 而在卷积神经网络(convolutional neural network, CNN)中, 使用跨步卷积或池化操作虽然会一定程度上增大感受野, 但在降维过程中会使网络忽略一些关键的特征信息, 易丢失特征细节信息, 常会导致小目标丢失, 出现漏检的现象, 从而无法满足检测的精度, 同时还会增加网络深度, 使整个网络计算量变大, 影响时空效率等问题; 同时, 对于传统的 CNN 网络, 传统卷积层是单个静态卷积核, 传统静态卷积中所有样本共享一个卷积核, 特征表达能力较差, 因此在一定程度上影响了 CNN 网络的准确性, 并且不能有效地进行推理.

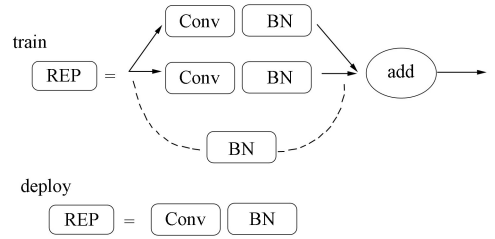


图 6 RepConv 结构

Fig. 6 Structure of RepConv

2 模型优化设计

在目标检测任务中, 目标检测模型不仅要在各种场景下准确地对各类目标进行检测, 而且还要降低网络参数和复杂度, 提高网络的运行速度, 进而提高目标检测的速度, 达到实时检测的目的. 基于此, 本文提出一种改进 YOLOv7 算法网络结构的方法.

2.1 空间-深度层设计

针对 YOLOv7 结构中的特征图, 对任何大小的 $S \times S \times C_1$ 中间特征映射 X , 将子特征映射序列切片为如下部分:

$$f_{0,0} = X[0 : S : \text{scale}, 0 : S : \text{scale}], \tag{1}$$

$$f_{1,0} = X[1 : S : \text{scale}, 0 : S : \text{scale}], \dots, \tag{2}$$

$$f_{\text{scale}-1,0} = X[\text{scale} - 1 : S : \text{scale}, 0 : S : \text{scale}], \tag{3}$$

$$f_{0,1} = X[0 : S : \text{scale}, 1 : S : \text{scale}], f_{1,1}, \dots, \tag{4}$$

$$f_{\text{scale}-1,1} = X[\text{scale} - 1 : S : \text{scale}, 1 : S : \text{scale}], \tag{5}$$

⋮

$$f_{0,\text{scale}-1} = X[0 : S : \text{scale}, \text{scale} - 1 : S : \text{scale}], f_{1,\text{scale}-1}, \dots, \tag{6}$$

$$f_{\text{scale}-1,\text{scale}-1} = X[\text{scale} - 1 : S : \text{scale}, \text{scale} - 1 : S : \text{scale}]. \tag{7}$$

例如 $\text{scale}=2$, 如图 7 所示, 可得到 4 个子图 $f_{0,0}, f_{0,1}, f_{1,0}, f_{1,1}$, 每个特征子图的尺寸为 $\left(\frac{S}{2}, \frac{S}{2}, C_1\right)$,

并下采样 2 倍. 然后按通道维度连接子特征映射, 得到一个新的特征映射 $X' \left(\frac{S}{\text{scale}}, \frac{S}{\text{scale}}, \text{scale}^2 C_1\right)$, 相比于特征映射 $f_{0,0}, f_{0,1}, f_{1,0}, f_{1,1}$, 空间维度减少一个 scale , 通道维度增加一个比例因子 scale^2 .

2.2 全维动态卷积优化

全维动态卷积是采取多维注意力机制和并行策略, 在任何卷积层沿内核空间的 4 个维度学习卷积内核的注意力, 运算公式如下:

$$y = (\alpha_{w_1} \odot \alpha_{f_1} \odot \alpha_{c_1} \odot \alpha_{s_1} \odot w_1 + \dots + \alpha_{w_n} \odot \alpha_{f_n} \odot \alpha_{c_n} \odot \alpha_{s_n} \odot w_n) * x, \tag{8}$$

其中: $\alpha_{w_i} \in \mathbb{R}$ 表示卷积核 w_i 的关注标量; $\alpha_{s_i} \in \mathbb{R}^{k \times k}$, $\alpha_{c_i} \in \mathbb{R}^{c_{in}}$ 和 $\alpha_{f_i} \in \mathbb{R}^{c_{out}}$ 表示 3 个新引入的关注

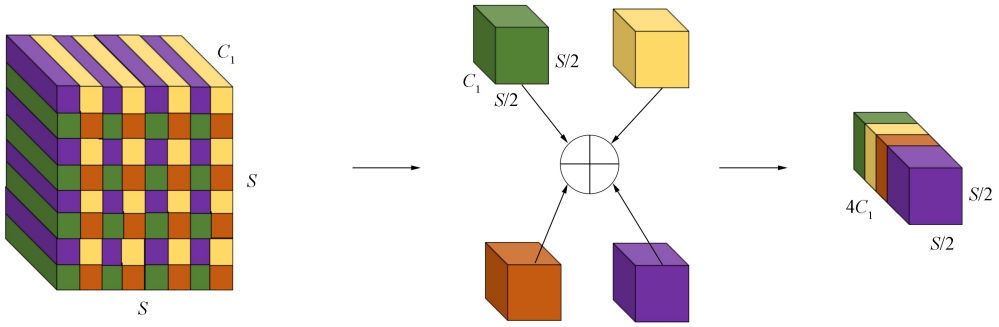


图 7 SPD 结构

Fig. 7 Structure of SPD

点, 分别沿卷积核 τ_i 核空间的空间维度、输入通道维度和输出通道维度计算; \odot 表示沿着核空间不同维度的乘法运算.

图 8 为全维动态卷积的多种注意力机制结构. 全维动态卷积通过对空间大小、输入通道数量和输出通道数量赋予了动态属性, 从而极大减少了额外的参数, 提升了卷积核自身的感受野, 有更强的特征表达能力.

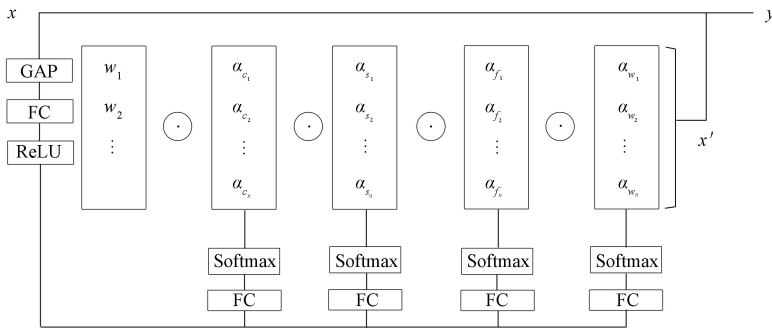


图 8 多种注意力机制结构

Fig. 8 Structure of multiple attention mechanism

2.3 卷积注意力机制设计

CBAM(convolutional block attention module)是一个轻量级的注意力模块, 由通道注意力模块 CAM(channel attention mechanism)和空间注意力模块 SAM(space attention mechanism)组成. CAM 可使网络提高对图像的前景和一些关键位置区域的注意力; SAM 使网络关注图片上下文中语义信息丰富的位置. 图 9 为 CBAM 的结构.

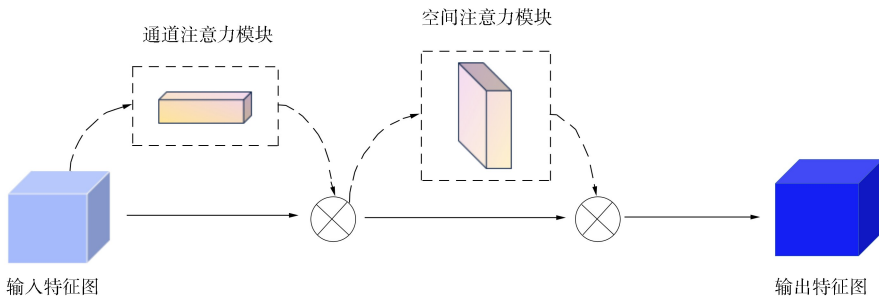


图 9 卷积注意力机制

Fig. 9 Convolutional block attention mechanism

对于特征图 $F \in \mathbb{R}^{C \times H \times W}$, $M_c \in \mathbb{R}^{C \times 1 \times 1}$ 表示通道注意力模块输出的通道权重数据为 $1 \times 1 \times C$, F' 表示通道注意力的输出; $M_s \in \mathbb{R}^{1 \times H \times W}$ 表示空间注意力模块输出的空间权重数据为 $2 \times H \times W$, 将通道注意力所输出的结果 F' 与空间注意力权重相乘后即得到 CBAM 输出结果 F'' , 公式如下:

$$F'' = M_c(F) \otimes F, \tag{9}$$

$$F'' = M_s(F') \otimes F'. \tag{10}$$

2.4 目标检测模型优化

YOLOv7 是一种基于卷积神经网络的一阶段算法, 本文在骨干网络与检测头之间的特征提取部分加入卷积注意力机制 CBAM, 如图 10 所示. 卷积注意力机制 CBAM 通过通道注意力机制模块 CAM 和空间注意力机制模块 SAM 分别沿通道和空间两个独立维度依次推断图片的注意力, 然后将注意力图乘以输入的特征图, 进行自适应特征细化, 优化了网络感兴趣特征的表现.

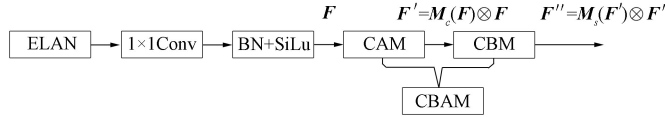


图 10 卷积注意力机制特征提取器

Fig. 10 Feature extractor of convolutional block attention mechanism

为改善网络训练速度, 优化网络的特征学习能力, 减少细节特征丢失, 本文采用带有 C_2 滤波器的非跨步全维动态卷积层连接空间-深度层的方式, 在检测头之前添加该结构, 以进一步对特征 X' 进行转换, 转换方式为

$$X'' \left(\frac{S}{scale}, \frac{S}{scale}, C_2 \right). \tag{11}$$

从而减少 YOLOv7 网络结构中跨步卷积和池化层的使用, 降低网络复杂度, 提升网络性能; 同时利用全维动态卷积代替传统网络中的静态卷积, 提升卷积核自身的感受野, 尽可能保留所有判别特征的信息, 获得更高的特征学习率. 改进的 YOLOv7 网络结构如图 11 所示.

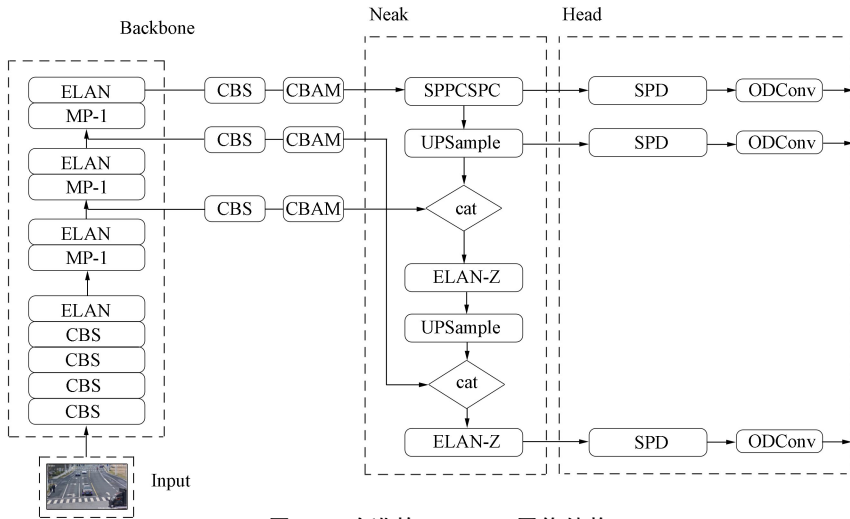


图 11 改进的 YOLOv7 网络结构

Fig. 11 Network structure of improved YOLOv7

3 实验验证

实验流程如图 12 所示. 通过自研的雷视一体机对交通路口进行拍摄, 收集一段时间内的交通目标数据; 将图片输入数据预处理模块对图片进行处理, 处理后的图片输入 YOLOv7 网络进行训练, 在反复前向、后向传播后, 训练完毕得到满足要求的权重模型. 进行检测时, 将图片输入到训练获得的模型中进行推理, 最后得到检测结果.

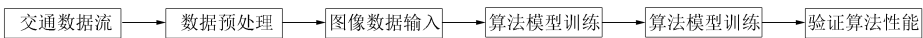


图 12 实验流程

Fig. 12 Experimental process

3.1 数据集的制作

本文数据集包括训练集、测试集和验证集, 数据集制作过程如下: 先利用由江苏省无锡市集萃深

度感知技术研究所提供的雷视一体机对路口进行监控并录制视频, 从录制视频中截取一定数量的图片, 图片中涵盖行人、非机动车、小型机动车、大型机动车 4 种交通目标; 然后利用 Vott 软件对数据集进行标注, 分别为 4 类识别目标, 其中“0”为行人, “1”为非机动车, “2”为小型机动车, “3”为大型机动车, 每一行是一个目标, 第一个元素是类别, 后面 4 个数据是经过归一化后的数据, 分别表示中心横坐标 X 、中心纵坐标 Y 、标注框宽度 W 、标注框高度 H 。

通过对所截取图片的筛选和检查, 从截取总数 6 000 张中选取 4 800 张图片作为训练集, 其中包括多个路口场景, 如图 13 所示的不同路口、白天、夜晚、雨天等场景, 以此增加数据集的多样性, 进而提升模型的鲁棒性. 然后按上述要求对图片进行标注. 此外, 剩余的 1 200 张图片不进行标注, 作为测试集对网络性能进行测试。

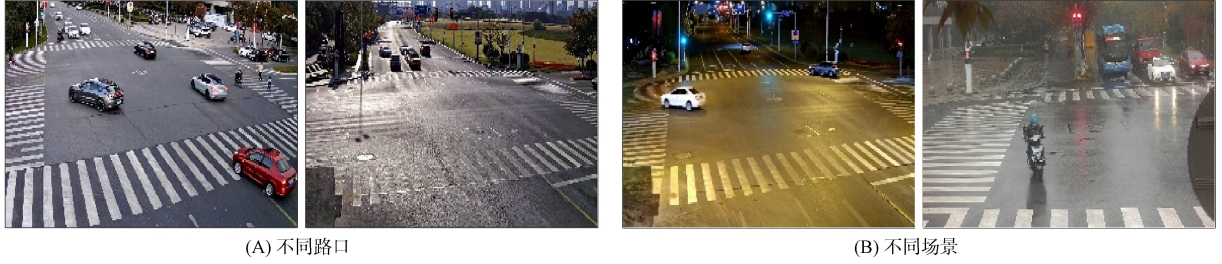


图 13 多样性数据集示例

Fig. 13 Examples of diversity datasets

3.2 实验设定与训练

实验环境配置如下: 操作系统为 Windows10/ubuntu18. 04, CPU 为 Intel(R) Core i5-10400F/ Intel(R) Core(TM) i9-9900, GPU 为 NVIDIA GeForce RTX3060/NVIDIA GeForce RTX2080Super, 编程语言为 Python3. 7, GPU 加速环境为 CUDA 11. 6/CUDA 11. 4, 显存为 12 GB/8 GB.

实验训练参数设置如下: epochs 设为 300, 表示对训练集的全部样本训练 300 次; batch-size 设为 8, 表示每次训练时在训练集中取 8 个训练样本进行训练; img-size 设为 640×640 , 表示输入图片像素大小为 640×640 。

3.3 结果分析

在 300 轮的训练后, 利用得到的最佳权重参数模型对待检测目标样本进行推理验证, 结果如图 14 所示. 由图 14(A)可见, 未检测出公交车的存在; 由图 14(C)可见, 未检测出周围处于图像暗部的车辆, 同时一些停止在车道线上的车辆也未能检测出; 由图 14(E)可见, 在夜晚灯光较暗的情况下, 在车道线上的小车对大车有一定的遮挡, 导致模型未对该大车进行有效识别; 由图 14(G)可见, 改进前模型将骑电瓶车的交通目标误识别为行人, 而且识别出的行人置信度较低. 而图 14(B), (D), (F), (H) 识别出了未能被检测出的目标, 并且相对提高了对目标的置信度, 有效改善了少检、漏检和误检的现象, 提高了目标检测的精度; 由图 14(I), (K), (M), (O)可见, 远处的目标很小, 可供学习的特征较少, 使改进前的网络模型对小目标检测能力较差; 由图 14(J), (L), (N), (P)可见, 改进后的模型能更深入地挖掘特征信息, 保证特征的完整性, 尤其是对一些关键特征的学习, 进而识别出了远处的车辆, 提升了小目标检测能力. 表 1 列出了不同网络模块的推理结果。

表 1 不同网络模块的推理结果

Table 1 Inference results of different network modules

网络模块	mAP/%	训练耗时/h	网络模块	mAP/%	训练耗时/h
YOLOv7	70.1	22.69	YOLOv7(CBAM)	86.9	17.41
YOLOv7(SPD-Conv)	89.4	21.35	本文	96.1	16.71
YOLOv7(ODConv)	87.2	20.01			

由表 1 可见, 改进后的网络模型平均精确度(mAP)比改进前有一定提高; 同时, 改进前的网络模型较复杂, 导致整个网络运行效率低, 训练时间较长, 且目标检测的精度也较低, 易出现目标漏检或误检等问题. 但经过改进后, 本文模型的训练时间大幅度下降, 运行效率提高, 改进前模型未被检测

出的目标均被成功检测出,且在交通数据集下平均精度相比于 YOLOv7(SPD-Conv)有一定提升。



图 14 网络结构改进前后对比结果

Fig. 14 Comparison results before and after network structure improvement

行人、非机动车和机动车的相对检出提升率列于表 2。由表 2 可见,行人在图中相对于其他目标较小,网络检测时易漏检,改进后的模型检测性能更强,尤其是对小目标的检测;在进行非机动车检测时,改进前的模型有时会误检,将非机动车目标识别为行人,而改进后的模型更精确,行人目标和非机动目标分属不同类,二者不同;对于机动车目标,由于目标较多会导致模型漏检和误检,改进后的模型具有更强的检测性能,能更好地识别交通目标。

表 2 改进 YOLOv7 网络的提升比率

Table 2 Improvement rate of improved YOLOv7 network

类别	YOLOv7	改进 YOLOv7 网络	提升比率/%
行人	17	28	64.7
非机动车	482	613	27.2
机动车	1 825	2 972	62.8

实验结果表明,改进后的算法增强了特征学习能力和特征表达能力,对在目标检测中出现的少检、漏检等问题有一定改善;同时在对小目标进行检测时,能有效检测出像素点较少的道路远处的小目标,提升了网络对小目标的检测能力,进一步提高了网络模型的检测精度和准确度。

综上所述,针对交通路口目标识别中存在的精确度低、少检、漏检等问题,本文提出了一种基于深度学习的交通目标识别方法。该方法通过在原有的 YOLOv7 网络基础上,采用前馈式卷积注意力机制 CBAM 提升对通道和空间的注意力,以提高对关键特征的注意力,降低网络复杂度,提高网络运行效率;同时采用 SPD 层与非跨步全维动态卷积 ODConv 相连接的方式,提升卷积神经网络对特征的学习能力和表达能力,提升了网络对小目标的检测精度,进而提升了整个网络的性能。实验结果表明,

改进后的目标检测网络在不同时间段、不同环境下的检测能力有一定提升, 对环境具有一定的鲁棒性, 实现了对交通路口目标识别精度的提升。

参 考 文 献

- [1] 侯学良, 单腾飞, 薛靖国. 深度学习的目标检测典型算法及其应用现状分析 [J]. 国外电子测量技术, 2022, 41(6): 165-174. (HOU X L, SHAN T F, XUE J G. Analysis of Typical Algorithms and Application Status of Deep Learning Object Detection [J]. Foreign Electronic Measurement Technology, 2022, 41(6): 165-174.)
- [2] 李明熹, 林正奎, 曲毅. 计算机视觉下的车辆目标检测算法综述 [J]. 计算机工程与应用, 2019, 55(24): 20-28. (LI M X, LIN Z K, QU Y. Overview of Vehicle Target Detection Algorithms under Computer Vision [J]. Computer Engineering and Applications, 2019, 55(24): 20-28.)
- [3] ALSHEMALI B, KALITA J. Improving the Reliability of Deep Neural Networks in NLP: A Review [J]. Knowledge-Based Systems, 2020, 191: 105210-1-105210-9.
- [4] 欧阳继红, 王梓明, 刘思光. 改进多尺度特征的 YOLO_v4 目标检测方法 [J]. 吉林大学学报(理学版), 2022, 60(6): 1349-1355. (OUYANG J H, WANG Z M, LIU S G. YOLO_v4 Object Detection Method with Improved Multi-scale Features [J]. Journal of Jilin University (Science Edition), 2022, 60(6): 1349-1355.)
- [5] 逢晨曦, 李文辉. 基于注意力改进的自适应空间特征融合目标检测算法 [J]. 吉林大学学报(理学版), 2023, 61(3): 557-566. (PANG C X, LI W H. Adaptive Spatial Feature Fusion Object Detection Algorithm Based on Attention Improvement [J]. Journal of Jilin University (Science Edition), 2023, 61(3): 557-566.)
- [6] 姚庆安, 张鑫, 刘力鸣, 等. 融合注意力机制和多尺度特征的图像语义分割 [J]. 吉林大学学报(理学版), 2022, 60(6): 1383-1390. (YAO Q A, ZHANG X, LIU L M, et al. Image Semantic Segmentation Based on Fusion of Attention Mechanism and Multi-scale Features [J]. Journal of Jilin University (Science Edition), 2022, 60(6): 1383-1390.)
- [7] SINGH B, NAJIBI M, DAVIS L S. Sniper: Efficient Multi-scale Training [C]//Proceedings of the 32nd International Conference on Natural Information Processing Systems. New York: ACM, 2018: 9333-9343.
- [8] SINGH B, DAVIS L S. An Analysis of Scale Invariance in Object Detection Snip [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2018: 3578-3587.
- [9] SUNKARA R, LUO T. No More Strided Convolutions or Pooling: A New CNN Building Block for Low-Resolution Images and Small Objects [C]//Machine Learning and Knowledge Discovery in Databases: European Conference. New York: ACM, 2022: 443-459.
- [10] 孙志军, 薛磊, 许阳明, 等. 深度学习研究综述 [J]. 计算机应用研究, 2012, 29(8): 2806-2810. (SUN Z J, XUE L, XU Y M, et al. A Review of Deep Learning Research [J]. Computer Application Research, 2012, 29(8): 2806-2810.)
- [11] CHEN Y P, DAI X Y, LIU M C, et al. Dynamic Convolution: Attention over Convolution Kernels [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2020: 11030-11039.
- [12] YANG B, BENDER G, LE Q V, et al. Condconv: Conditionally Parameterized Convolutions for Efficient Inference [C]//Proceedings of the 33rd International Conference on Neural Information Processing Systems. New York: ACM, 2019: 1307-1318.
- [13] LI C, ZHOU A J, YAO A B. Omni-dimensional Dynamic Convolution [EB/OL]. (2022-09-16)[2023-03-15]. <https://arxiv.org/abs/2209.07947>.
- [14] 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 [EB/OL]. (2022-07-06)[2023-04-01]. <https://arxiv.org/abs/2207.02696>.

(责任编辑: 韩 啸)