

引用格式:万昊,任永国.基于多尺度融合和特征对齐的实时高精度交通标志检测与识别[J].中国传媒大学学报(自然科学版),2024,31(01):65-73.

文章编号:1673-4793(2024)01-0065-09

基于多尺度融合和特征对齐的实时高精度交通标志检测与识别

万昊^{1*},任永国²

(1. 百度在线网络技术有限公司,北京 100085;2. 北京邮电大学,北京 100876)

摘要:高精度的实时交通标志检测和识别对安全自动驾驶和智能交通系统至关重要。本研究对基线网络YOLOV4进行了升级,增加了多尺度融合模块和注意力机制模块(AMM),丰富了不同尺度交通标志的特征表示。同时,颈部网络结合了特征选择模块和特征对齐模块,增强了高、低层特征图之间像素偏移的语义判别。具体地说,针对AMM,设计了一种转置的自注意力操作。其使用互协方差矩阵将令牌维度上的操作转换为通道维度,将时间复杂度从 $O(n^2)$ 降低到 $O(n)$ 。在TT100K交通标志数据集上的实验结果表明,与基线网络($mAP@0.5=76.4%$)相比,升级后的网络($mAP@0.5=83.4%$)取得了较好的改进,检测和识别速度可达39.45帧/秒,达到了目前最先进的水平。

关键词:交通标志识别;神经网络;自注意力;多尺度融合;特征对齐

中图分类号:TP183 文献标识码:A

Real-time high-accuracy traffic sign detection and recognition using multi-scale fusion and feature alignment

WAN Hao^{1*}, REN Yongguo²

(1. Baidu Online Network Technology (Beijing) Co., Ltd., Beijing 100085, China; 2. Beijing University of Posts and Telecommunication, Beijing 100876, China)

Abstract: Real-time traffic sign detection and recognition with high accuracy is crucial for safe autonomous driving and intelligent transportation systems. Many deep networks have been proposed, while there is still room for further improvement. In this study, the baseline network YOLOV4 was upgraded with a multi-scale fusion module and an attention mechanism module (AMM) for enriching the feature representation of traffic signs of different sizes. Meanwhile, a neck network combined a feature selection module and a feature alignment module for enhancing the semantic discriminant of pixel shift between high- and low-level feature maps. Specifically, a transposed self-attention operation was designed for AMM. It used cross-covariance matrix for transforming the operation on token dimension to channel dimension and reduced the time complexity from $O(n^2)$ to $O(n)$. Experimental results on the TT100K traffic sign dataset indicate that compared to the baseline network ($mAP@0.5=76.4%$), the upgraded network achieves good improvement ($mAP@0.5=83.4%$) with detection and recognition speed of 39.45 frames per second, and it also outperforms several state-of-the-art works.

Keywords: traffic sign recognition; neural network; self-attention; multi-scale fusion; feature alignment

1 引言

安全自动驾驶和智能交通需要对海量信息进行实时高精度处理,包括道路检测、车辆检测、行人检测、交通标志检测和识别等测量、分析和执行^[1]。

交通标志是智能交通系统中不可或缺的组成部分。作为一种重要的交通管理工具,它可以帮助人们明确道路规则和行车指导,减少事故发生的概率,提高道路交通安全性。随着智能交通系统的发展,交通标志检测成为了一项必不可少的研究内容。实时准确的交通标志检测和识别对安全的自动驾驶和智能交通至关重要^[2]。它主要涉及车辆碰撞的引导和警告、限速和转弯通知。另一方面,由于类内的可变性、类间的相似性和不良的标准化,它又很具有挑战性^[3]。此外,恶劣的天气、不断变化的照明、运动模糊、背景变化和物体遮挡都给交通标志的检测、识别和理解带来了困难。

为了解决这一具有挑战性的任务,研究者们已提出了许多算法。在以往的研究中,该任务一般分为交通标志定位、分割和分类几个连续的部分^[2,3]。大多数技术使用颜色和形状作为重要线索。然而,在恶劣条件下和背景变化的情况下,这些线索就不可靠了。具体来说,图像处理的整个过程对图像阈值分割、边缘检测、梯度分析和特征提取都很敏感,因此,容易造成高误报^[3]。

卷积神经网络(Convolutional Neural Networks, CNN)仍然是交通标志检测和识别的核心,因为它可以将检测和识别纳入到一个端到端的优化过程中^[1]。Yang等^[4]开发了一种具有双多尺度模块的全卷积网络(Fully Convolutional Network, FCN),可以抑制误报。Zhu等^[5]将FCN引导的交通标志提案与基于CNN的交通标志分类相结合。Li等^[6]将两个CNN整合成一个具有不对称核的新结构,并使用颜色和形状来细化定位和分类。Natarajan等^[7]设计了一种新的训练方法,对多个CNN进行加权以提高性能。Tabernik等^[8]使用在线硬例挖掘、样本加权和数据增强来适应Mask RCNN。Cao等^[9]使用颜色进行空间分割,使用形状特征进行符号定位,并使用Gabor核初始化显示CNN进行分类。Yuan等^[10]利用多尺度特征融合网络,并使用垂直空间序列关注模块获取上下文信息以获得更好的性能。Tian等^[11]将邻近层的相关上下文合并到检测架构中,并开发了一种基于循环注意力的新方法用于多尺度分析。Liu等^[12]设计了一个基于多尺度区域的CNN,它连接了深层和浅层特征层,并利用多尺度上下文区域来利用周围信息进行符号识别。

Sun等^[13]构建了一个密集连接相关的传输连接块来结合高、低层特征,并提出了一种锚点设计方法,为检测小型交通标志提供合适的锚点。Liu等^[14]开发了一种目标网格预测FCN,可以从粗定位到精定位检测灵活的潜在目标区域。基于交通标志的统计特征,Tang等^[15]引入了一种轻量级操作来提高推理速度,即集成用于解决样本不平衡问题的栅格运算,以及针对大小差异增强特征表示的特征聚合结构。Liu等^[16]学习尺度感知和上下文丰富的特征来识别小的和闭塞的交通标志,并使用用于多尺度自注意力学习的注意力驱动的双边特征金字塔网络和用于利用上下文丰富表示的自适应感受野融合块来更新网络。Lin等^[17]应用多尺度特征融合网络来提高识别精度和泛化能力。

深度网络YOLO^[18]在交通标志检测和识别中很受欢迎。文献[19]采用不同感兴趣区域的并行预处理来加快推理速度。文献[20]将网络与数据增强和多尺度空间金字塔池相结合,以提高性能。在文献[21]中,为了获得更好的结果,YOLO采用不同核大小的混合深度卷积和同层到跨层的注意力特征融合进行更新。文献[22]则使用自适应注意力和特征增强对网络进行升级,以减少信息损失以增强特征表示。

本文从特征提取和特征增强两个方面对骨干网络YOLOV4进行了升级,设计了一种新的注意力机制模块,降低了复杂度。主要贡献概括如下。

对特征提取,设计了一个多尺度融合模块,以在残差块内构建分层连接,以增加感受野,并修改了注意力机制模块,以纳入上下文信息,以丰富精细细节。

对特征增强,将特征选择模块和特征对齐模块相结合,以很好地处理冗余特征图以及补丁上采样和下采样之间的像素移位,以匹配高级别和低级别的语义特征图。

特征提取和特征增强有助于提高交通标志检测和识别性能,与TT100k数据集上的骨干网络YOLOV4相比,本文提出的网络取得了更好的效果。

2 网络架构

如图1所示,本文提出的网络架构是基于骨干网络YOLOV4的特征提取(FEx)和特征增强(FEh)进行升级的。前者集成了多尺度融合模块(Multi-Scale Fusion Module, MFM)和注意力机制模块(Attention Mechanism Module, AMM),后者集成了特征选择模块(Feature Selection Module, FSM)和特征对齐模块(Feature Alignment Module, FAM)。

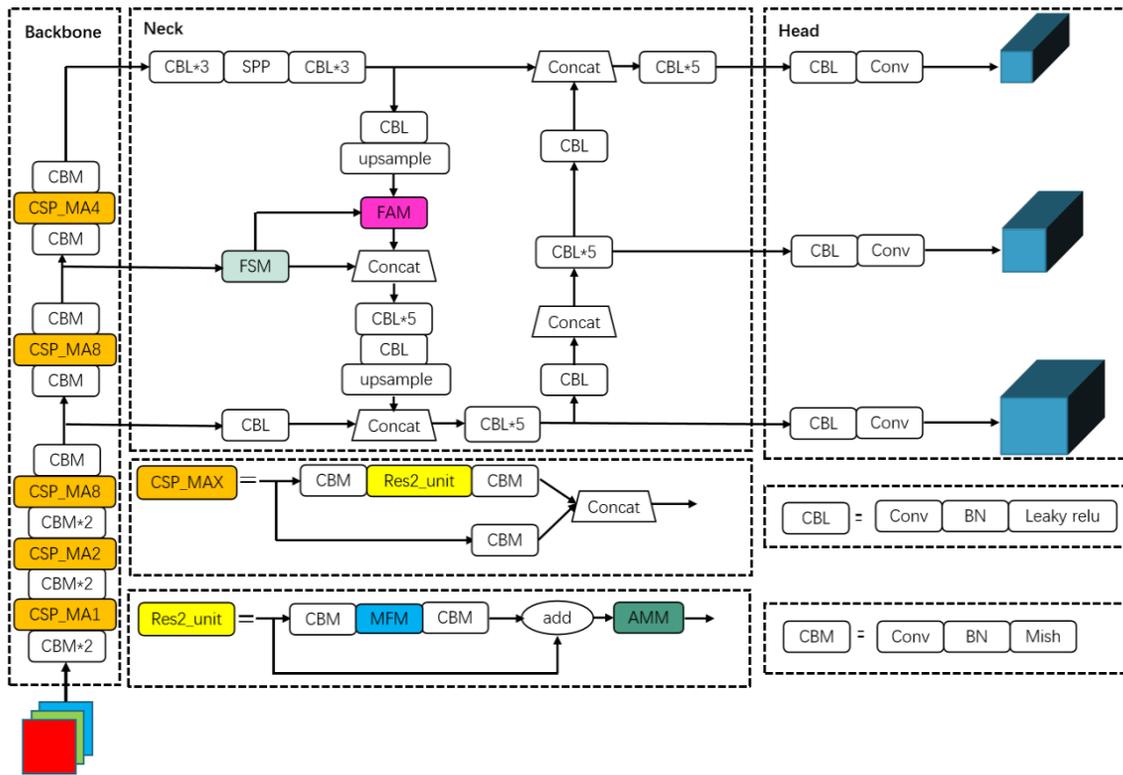


图1 升级后的YOLOV4架构

(注:颜色的方块代表升级后的模块。CSP MAX中的X是连续连接的此类单元的编号。MFM和AMM集成为Res2单元。)

2.1 多尺度融合模块(MFM)

现实环境中,需在不同的背景和灯光明下,在不同的范围内获取交通标志。多尺度特征融合对图像表示非常重要。由于实际场景中交通标志尺寸较小,在图像中整体像素占比不超过1%,若在特征提取阶段频繁使用卷积进行下采样,将会导致在深层特征图中小目标的特征难以被定位甚至导致消失。因此通常使用浅层特征层检测小目标,因为浅层的图像分辨率大、包含的细节信息更多。而深层特征

层通常被用来预测大、中尺寸目标,因为深层特征包含更丰富的语义信息。实际自动驾驶场景下的交通标志检测任务,一幅图像中通常包含各种尺寸不一的目标实例。现有网络更倾向于使用不同分辨率的数据输入来改进特征表达。在本研究中,相同的 3×3 卷积核被替换为单个残差块内的分层多尺度结构,因此,每个实例都可以在更细粒度的级别上表达,从而增加了每个网络层的感受野大小,提高了网络对不同尺寸目标的特征表达能力。MFM的结构如图2所示。

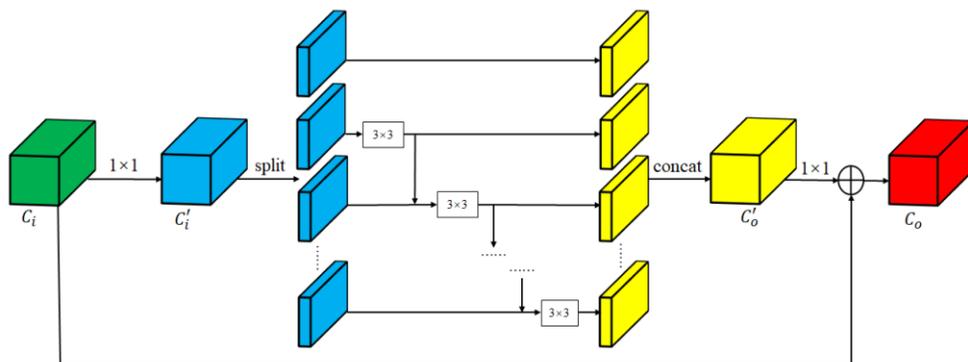


图2 多尺度融合模块(MFM)结构

具体来说, MFM对输入特征进行 1×1 卷积, 其输出被传送到 X_1, X_2, \dots, X_s 的不同通道。除 X_1 外, 其他通道将前一组特征映射的输出相加。过程可以用公式(1)表示, 其中 X_i 表示均衡化后的小块, s 为平均分组的个数($s = 4$), k_i 表示对特征图 i 进行卷积核为 3×3 的卷积运算, y 表示卷积后的输出。最后, s 个输出通道通过了用于特征图聚合的 1×1 卷积。

$$y_i = \begin{cases} X_i, & i = 1 \\ k_i(X_i), & i = 2 \\ k_i(X_i + y_{i-1}), & 2 < i \leq s \end{cases} \quad (1)$$

MFM在瓶颈残差块中采用了 3×3 卷积的块复用, 在不增加网络参数的情况下进行多尺度通道融合, 使得网络提取出来的特征表达能力更强。

2.2 自注意力机制模块(AMM)

由于交通标志检测任务对图像分辨率的要求较高, 若直接使用传统的自注意力机制处理长序列文本或高分辨率图像, 将会导致极大的计算开销。因

此对自注意力机制计算相似度矩阵的方式进行优化, 提出了一种轻量化的自注意力机制模块(AMM)。受非局部机制的启发^[23], 通过使用Key和Query的互协方差矩阵, 并将原始自注意力机制在空间维度上的点积操作转变为在通道维度上的点积操作。其保持了全局表示, 能够更加高效地处理高分辨率图像, 对小型交通标志的检测能力有所提高。同时, 计算复杂度从 $O(n^2)$ 大幅降低到 $O(n)$, 极大地降低了计算开销。

AMM结构如图3所示。首先, 该模块同样使用三个 1×1 卷积对输入特征图进行线性变换, 得到三个向量分支(Q、K和V)。然后, 通过乘法计算Q和K之间的相似矩阵, 并通过Softmax函数将值归一化。不同于自注意力机制直接在空间维度上进行点积操作得到相似度系数矩阵, 该模块是在通道维度上进行点积运算。接下来, 通过矩阵乘法将归一化的相似系数叠加到V上。最后, 通过卷积运算将叠加的特征维数恢复到与原始特征图相同的维数, 并通过与原始特征图X相加来补充上下文信息。

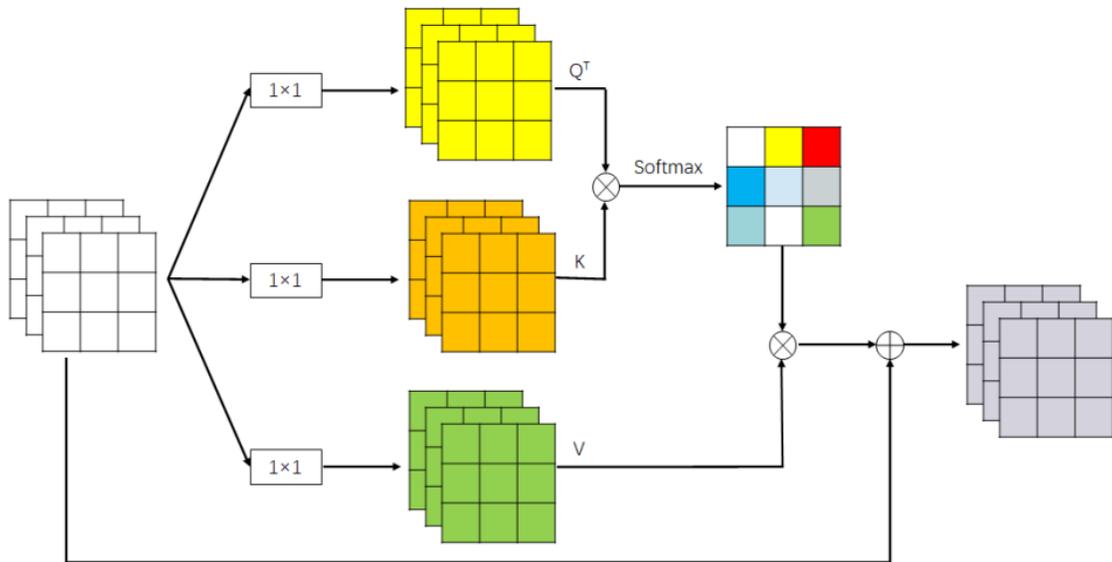


图3 注意机制模块(AMM)的结构

2.3 特征选择模块(FSM)

卷积神经网络的最新发展在交通标志检测领域取得了显著飞跃, 交通标志检测领域的研究者们经常会用到特征增强网络去融合不同语义的特征。而当大多数现有的方法为了设计简单而忽略了特征对齐问题, 上采样之后的特征和局部特征之间直接进行像素相加会导致未对齐的上下文信息, 进而导致预测

过程中的错误分类。

在对特征通道进行降维之前, 强调包含丰富空间细节信息的特征图是非常重要的, 这将便于资源的精准分配, 同时抑制冗余的特征。因此本节使用特征选择模块(FSM)来对特征图每个通道的重要性进行显式建模, 并相应地对它们进行重新校准。

FSM的目的是保持具有丰富空间细节的特征图, 同时降低特征维数。FSM的结构如图4所示。本节所

使用的特征选择模块首先会将每个输入特征图通过全局平均池化层以提取其全局信息,然后通过特征重要性建模函数学习使用全局信息对通道的重要性进行建模,再将重要性向量作用于原始的输入特征图上并将

其进行缩放,再将缩放之后的特征图与原始特征图进行相加,得到重缩放特征图。最后,在重新缩放的特征图上引入特征选择层,用于选择性地保留重要通道信息并丢弃掉无用的特征信息以进行通道降维。

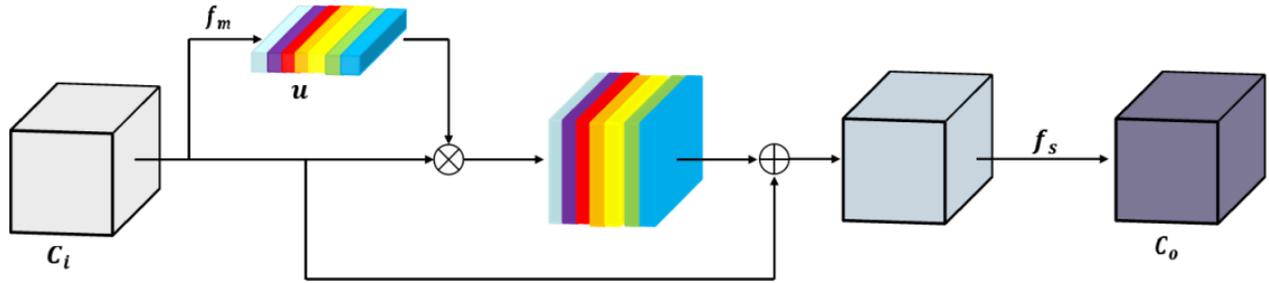


图4 特征选择模块(FSM)的结构

公式(2)定义了映射通道的特征选择过程。给定一个输入地图 C_i , 特征重要性函数 f_m 学习使用全局信息 z 对地图通道进行建模。同时, 形成重要性向量 μ 。最后, 引入特征选择层 s 对组合映射 $C_i + \mu * C_i$ 缩放, 以突出重要的通道信息并抑制相对冗余的特征。

$$\begin{cases} C_o = f_s(C_i + \mu * C_i) \\ \mu = f_m(z) \\ z = \frac{1}{H_i \times W_i} \sum_{h=1}^{H_i} \sum_{\omega=1}^{W_i} c(h, \omega) \end{cases} \quad (2)$$

本节所使用的特征选择模块的设计灵感是来自于通道注意力机制 SENet 的, 主要区别在于该模块在输入特征图和缩放特征图之间引入了额外的跳跃连接, 可以避免任何特定的通道相应被过度放大或者过度抑制。

2.4 特征对齐模块(FAM)

由于在特征提取过程中重复地使用下采样操作, 会使得上采样之后的特征图和其相应的局部特征之间存在可见的像素错位, 因此直接使用像素相加或者通道拼接的融合方式会损害图像边界的目标检测。所以在特征融合之前, 把上采样之后的高级语义特征和其对应的局部空间特征对齐是非常重要的, 因此本节使用一个特征对齐模块(FAM)来根据局部特征图所提供的空间位置关系来调整其所对应的上采样之后的特征图。

递归下采样和上采样导致可见的空间错位。直接添加或通道拼接会破坏物体边界的预测, 因此, 对齐相应地图的低级和高级语义特征成为必要步骤。根据相应的局部特征图提供的空间位置信息, 使用 FAM 对上采样后的特征图进行调整。FAM 的结构如图 5 所示。

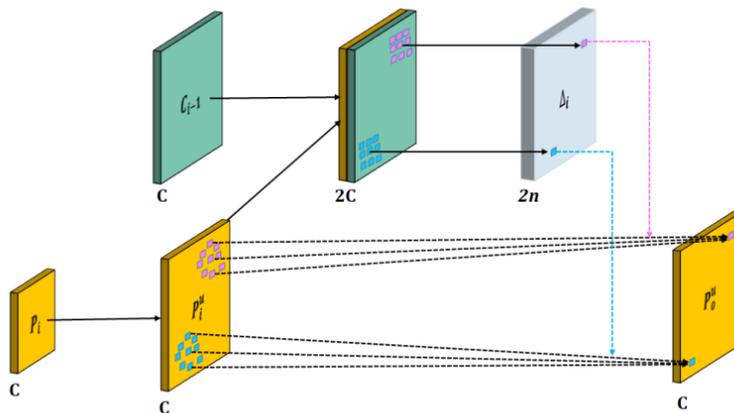


图5 特征对齐模块(FAM)的结构

如上图5所示,特征对齐模块首先会将上采样之后的特征图与其所对应的局部特征图进行拼接,然后通过一个普通的 3×3 卷积函数去学习这两层特征图之间的像素偏移量,可以得到一幅通道维度为 $2N$ 的偏移量图,前 N 维代表每个像素点在水平方向上的偏移,后 N 维代表每个像素点在垂直方向上的偏移,接着将学习到的偏移量用于上采样之后的特征图,最后在该特征图上使用一个 3×3 可变形卷积函数进行对齐。

公式(3)为特征对齐过程。其中, P_i^u 表示上采样后的特征图, C_{i-1} 表示其所对应的局部特征图, P_o^u 表示经过对齐之后的输出特征图。使用一个普通的 3×3 卷积函数 f_o 来学习这两层特征图之间的像素偏移,得到一个维数为 $2n$ 的偏移量图。前 n 维表示每个像素的水平偏移量,最后 n 维表示每个像素在垂直方向上的偏移量。最后,将学习到的偏移量 Δi 应用于上采样之后的特征图,与文献[24]一样,可变形卷积函数 f_o 用于像素级特征图对齐。

$$\begin{cases} P_o^u = f_a(P_i^u, \Delta i) \\ \Delta i = f_o([C_{i-1}, P_i^u]) \end{cases} \quad (3)$$

3 实验及结果分析

3.1 实验设计

3.1.1 数据集

大规模中国交通标志数据集 TT100K 数据集^[5]非常具有挑战性,其由清华大学和腾讯实验室联合收集和整理而成。该数据集包含近 100,000 张图像,其中包含 30,000 个交通标志实例,涵盖了光照和天气等各种条件变化,这些图像是由 6 个像素非常高的广角单反相机在中国的 300 多个城市拍摄的腾讯街景全景图。图像样本包含了远景拍摄、近景拍摄等场景,且存在复杂自动驾驶道路背景、以及各种遮挡等现象,该数据集相比于之前的中国交通标志数据集,在标注类别和图像清晰度等方面具有更大优势,也更适合国内交通标志检测的研究。TT100K 数据集在中国交

通标志标注方面非常全面,即使是很小的目标也能被精准标注,其中尺寸在 20×20 以下的目标占据了很大的比例。正是由于小尺寸交通标志数量众多,使得模型的检测过程变得更加具有挑战性,因此该数据集更能检测一个交通标志检测模型的性能和效果。其包含有详细注释的交通标志,涵盖了不同的现实环境。在数据集中,不同类别的交通标志的实例数量是不均匀的。

由于 TT100K 中国交通标志数据集中标记了大量的交通标志类别,但某些类别的实例数量却相对较少,这就导致了不平衡的数据分布,这种情况与实际自动驾驶场景的情况符合。在本研究中,选取数据量超过 100 张图像的 45 类交通标志进行后续实验。具体来说,训练集包含 6107 张图像,测试集包含 3073 张图像。

3.1.2 性能指标

本研究对比了 Faster RCNN^[25]、RetinaNet^[26]、YOLOv3、YOLOX^[27]、基线 YOLOV4^[18]以及本文提出的网络,利用每秒浮点运算数(FLOPs)、不同交点的平均精度(mAP)和每秒帧数(FPS)作为评价指标进行量化结果分析。

3.1.3 参数及环境设置

初始学习率为 0.01,采用余弦退火策略降低学习率。epoch 数和 batch size 分别固定为 300 和 64。其他参数设置为默认值。

神经网络由 Python(3.8 版本)和 Pytorch(1.7 版本)实现。代码部署在 Win10 工作站(64 位操作系统,8 核 CPU,16GB RAM,NVIDIA GeForce RTX3070,16GB 内存)上。

3.2 实验结果

表 1 给出了六种深度网络在 TT100K 数据集上交通标志检测与识别的对比结果。每个指标的最佳表现以黑体字表示。

可以看出,本文提出的网络在 mAP@0.5 和 mAP@0.5:0.95 指标上达到最高,其次是 YOLOX 和

表 1 TT100K 数据集上的性能比较

| | Backbone | FLOP | mAP@0.5 (%) | mAP@0.5:0.95(%) | FPS |
|-------------|-----------------|---------------|-------------|-----------------|--------------|
| Faster RCNN | ResNet50 | 283.14B | 73.5 | 60.8 | 3.03 |
| RetinaNet | ResNet101 | 315.32B | 71.9 | 56.4 | 17.61 |
| YOLOV3 | Darknet53 | 109.15B | 63.8 | 47.9 | 36.28 |
| YOLOX | CSPDarknet53 | 99.59B | 78.6 | 63.1 | 50.42 |
| YOLOV4 | CSPDarknet53 | 128.46B | 76.4 | 61.6 | 41.78 |
| Ours | CSPDarknet_MA53 | 146.79B | 83.4 | 65.5 | 39.45 |

YOLOV4。同时,6个网络中有4个满足实时性要求(≥ 20 FPS)。此外,发现基于CSPDarknet骨干网构建的网络比使用ResNet的网络具有更好的性能。最重要的是,与基线YOLOV4相比,提出的网络在mAP@0.5上高出7%,在mAP@0.5:0.95上高出4%。另一方面,由于集成了四个不同的模块,所提出的网络导致FLOPs增加,FPS略有下降。

测试结果表明,本文提出的网络提高了交通标志检测和识别性能,满足实时图像处理。

3.3 消融实验

进行消融实验以量化升级单元(FEx和FEh)对性能改进的有效性,表2显示了结果。

可以观察到,FEx和FEh都对改进有积极的贡献,FEx在精度、召回率、mAP@0.5和mAP@0.5:0.95上分别提高了3.3%,0.6%,2.9%和1.1%。值得注意的是,同时使用FEx和FEh结果明显提高(精度,7.9% \uparrow ;召回率,3.7% \uparrow ;mAP@0.5,7.0% \uparrow ;mAP@0.5:0.95,3.9% \uparrow),这表明升级基线网络有助于提高检测和识别性能。

与五个最先进研究的比较也显示在表2中。结果表明,本文提出的网络达到了最高的准确率和第三高的召回率和mAP@0.5,并且循环注意力^[11]、周围情境^[12]和迁移学习^[20]可能有利于检测和分类。另一方面,需要注意的是文献[11]将交通标志分为三类,文献[12]采用了不同的数据分割策略,文献[20]需要统一大小的输入图像。

表2 消融实验和其他最先进网络的比较

| | Precision(%) | Recall(%) | mAP@0.5(%) | mAP@0.5:0.95(%) |
|----------------------|--------------|-----------|------------|-----------------|
| YOLOV4 | 80.3 | 71.9 | 76.4 | 61.6 |
| YOLOV4+FEx | 84.1 | 72.5 | 79.3 | 62.7 |
| YOLOV4+FEh | 83.6 | 72.1 | 78.6 | 62.1 |
| YOLOV4+FEx+FEh(Ours) | 88.2 | 75.6 | 83.4 | 65.5 |
| 文献[11]算法 | | 92.3 | | |
| 文献[12]算法 | | 89.4 | 71.5 | |
| 文献[13]算法 | | | 83.25 | |
| 文献[20]算法 | | | 86.3 | |
| 文献[21]算法 | | | 83.73 | |

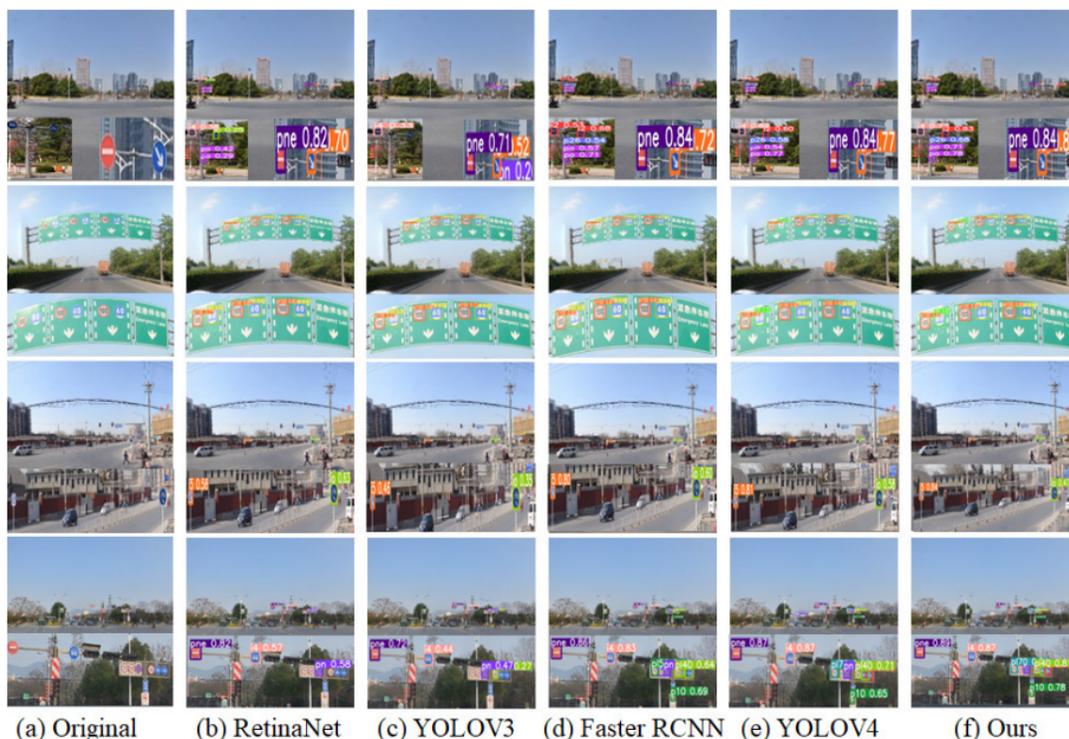


图6 交通标志检测和识别的对比示例

3.4 实例分析

图6展示了四个交通标志检测和识别结果的实例,在每个实例中用矩形定位,并将与置信度评分配对的交通标志进行了放大。

从图中可以看出, Faster RCNN、YOLOV4和所提出的网络对所有的交通标志进行了定位和识别,而RetinaNet和YOLOV3则错过了一些实例(第一行和第四行)。此外,所提出的网络产生了更高的置信度分数,因此可正确识别实例。

通过使用FEx和FEh对YOLOV4进行升级,可以显著提高交通标志检测和识别的性能,主要原因包括:(1)骨干网络YOLOV4相较于其他网络(如RetinaNet和Faster RCNN)具有更优越的性能,这可以从表1和图6中观察到;(2)FEx包含了交通标志的多尺度融合和自注意力机制以及丰富的层次表达,提高了检测和识别性能(见表2)。基于多尺度融合的自注意力学习也被设计用于拓宽接受域和增强上下文丰富表征^[16];(3)FEh可以校正像素偏移并对齐高低特征图,从而更好地进行语义识别(见表2)。

4 结论

实时高精度交通标志检测和识别在自动驾驶和智能交通领域具有挑战性。本研究通过多尺度融合和特征对齐对YOLOV4骨干进行升级,提高识别性能,并在注意力机制模块上设计了一种新的转置操作,大大降低了时间复杂度。在TT100K数据集上,升级后的网络达到了目前最先进的性能,满足了实时性要求。未来,将考虑进一步改进数据增强、循环注意力、迁移学习和高级架构。

参考文献(References):

- [1] Muhammad K, Ullah A, Lloret J, et al. Deep learning for safe autonomous driving: current challenges and future directions [J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 22, (7): 4316-4336.
- [2] Mogelmoose A, Trivedi M M, Moeslund T B. Vision-based traffic sign detection and analysis for intelligent driver assistance systems: perspectives and survey [J]. IEEE Transactions on Intelligent Transportation Systems, 2012, 13(4): 1484-1497.
- [3] Timofte R, Zimmermann K, Gool L V. Multi-view traffic sign detection, recognition, and 3D localisation [J]. Machine Vision and Applications, 2014, 25(3): 633-647.
- [4] Yang Y, Liu S, Ma W, et al. Efficient traffic-sign recognition with scale-aware cnn [DB/OL]. arXiv: 1805.12289, 2018.
- [5] Zhu Z, Liang D, Zhang S, et al. Traffic-sign detection and classification in the wild [C]//IEEE Conference on Computer Vision and Pattern Recognition, 2016: 2110-2118.
- [6] Li J, Wang Z. Real-time traffic sign recognition based on efficient CNNs in the wild [J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 20(3): 975-984.
- [7] Natarajan S, Annamraju A K, Baradkar C S. Traffic sign recognition using weighted multiconvolutional neural network [J]. IET Intelligent Transport Systems, 2018, 12(10): 1396-1405.
- [8] Tabernik D, Skočaj D. Deep learning for largescale traffic-sign detection and recognition [J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 21(4): 1427-1440.
- [9] Cao J, Song C, Peng S, et al. Improved traffic sign detection and recognition algorithm for intelligent vehicles [J]. Sensors, 2019, 19(18): 4021.
- [10] Yuan Y, Xiong Z, Wang Q. VSSA-net: vertical spatial sequence attention network for traffic sign detection [J]. IEEE Transactions on Image Processing, 2019, 28(7): 3423-3434.
- [11] Tian Y, Gelernter J, Wang X, et al. Traffic sign detection using a multi-scale recurrent attention network [J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 20(12): 4466-4475.
- [12] Liu Z, Du J, Tian F, et al. MR-CNN: a multi-scale region-based convolutional neural network for small traffic sign recognition [J]. IEEE Access, 2019, 7: 57120-57128.
- [13] Sun C, Ai Y, Wang S, et al. Dense-RefineDet for traffic sign detection and classification [J]. Sensors, 2020, 20(22): 6570.
- [14] Liu L, Wang Y, Li K, et al. Focus first: coarse-to-fine traffic sign detection with stepwise learning [J]. IEEE Access, 2020, 8: 171170-171183.
- [15] Tang Q, Cao G, Jo K. Integrated feature pyramid network with feature aggregation for traffic sign detection [J]. IEEE Access, 2021, 9: 117784-117794.
- [16] Liu Y, Peng J, Xue J, et al. TSingNet: scale-aware and context-rich feature learning for traffic sign detection and recognition in the wild [J]. Neurocomputing, 2021, 447: 10-22.
- [17] Lin S, Zhang Z, Tao J, et al. Traffic sign detection based on lightweight multiscale feature fusion network [J]. Sustainability, 2022, 14(21): 14019.
- [18] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection [C]//IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.
- [19] Avramovic A, Sluga D, Tabernik D, et al. Neural-network-based traffic sign detection and recognition in high-definition images using region focusing and parallelization [J]. IEEE Access, 2020, 8: 189855 - 189868.

- [20] Zhang H, Qin L, Li J, et al. Real-time detection method for small traffic signs based on yolov3[J]. IEEE Access, 2020, 8: 64145-64156.
- [21] Wan H, Gao L, Su M, et al. A novel neural network model for traffic sign detection and recognition under extreme conditions[J]. Journal of Sensors, 2021: 9984787.
- [22] Wang J, Chen Y, Dong Z, et al. Improved YOLOv5 network for real-time multi-scale traffic sign detection [J]. Neural Computing and Applications, 2023, 35: 7853-7865.
- [23] Wang X, Girshick R, Gupta A, et al. Non-local neural networks[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7794-7803.
- [24] Wang R, Shivanna R, Cheng D, et al. DCN V2: improved deep & cross network and practical lessons for web-scale learning to rank systems[C]. Proceedings of the Web Conference, 2021: 1785-1797.
- [25] Ren S, He K, Girshick R, 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.
- [26] Lin T, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]. IEEE International Conference on Computer Vision, 2017: 2980-2988.
- [27] Ge Z, Liu S, Wang F, et al. YOLOX: exceeding YOLO series in 2021[DB/OL]. arXiv: 2107.08430, 2021.

编辑:王谦