DOI: 10.13382/j.jemi.B2205524

融合前景注意力的轻量级交通标志检测网络*

俞林森[1] 陈志国[1,2,3,4]

(1. 江南大学人工智能与计算机学院 无锡 214122;2. 江南大学先进技术研究院 无锡 214122;
3. 江苏省模式识别与计算智能工程实验室 无锡 214122;4. 江南大学模式识别与计算智能国际联合实验室 无锡 214122)

摘 要:针对目标检测算法模型在交通标志检测上容易出现错检和漏检等问题,提出一种融合前景注意力的轻量级交通标志检测网络 YOLOT。首先引入 SiLU 激活函数,提升模型检测的准确率;其次设计了一种基于鬼影模块的轻量级骨干网络,有效提取目标物特征;接着引入前景注意力感知模块,抑制背景噪声;然后改进路径聚合网络,加入残差结构,充分学习底层特征信息;最后使用 VariFocalLoss 和 GIoU,分别计算目标的分类损失和目标间的相似度,使目标的分类和定位更加准确。在多个数据集上进行了大量实验,结果表明,本文方法的精度优于目前最先进方法,在 CCTSDB 数据集上进行消融实验,最终精度达到 98.50%,与基线模型相比,准确率提升 1.32%,同时模型仅 4.7 MB,实时检测帧率达到 44 FPS。

关键词: 交通标志检测;激活函数;前景注意力;特征融合;VariFocalLoss;GIoU

中图分类号: TP391.4 **文献标识码**: A **国家标准学科分类代码**: 510.4050

Lightweight traffic sign detection network with fused foreground attention

Yu Linsen[1] Chen Zhiguo[1,2,3,4]

(1. School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi 214122, China; 2. Institute of Advanced Technology, Jiangnan University, Wuxi 214122, China; 3. Jiangsu Provincial Engineering Laboratory of Pattern Recognition and Computational Intelligence, Wuxi 214122, China; 4. International Joint Laboratory of Pattern Recognition and Computational Intelligence, Jiangnan University, Wuxi 214122, China)

Abstract: A lightweight traffic sign detection network incorporating foreground attention, YOLOT, is proposed to address the problem that object detection algorithm models are prone to error and miss detection on traffic sign detection. Firstly, the introduction of the SiLU activation function to improve the accuracy of model detection; secondly, a lightweight backbone network based on the ghost module is designed to effectively extract object features; thirdly, introduction of foreground attention perception module to suppress background noise; fourthly, we improve the path aggregation network by adding a residual structure to the feature fusion process; finally, we use VariFocalLoss and GIoU to calculate the classification loss of objects and the similarity between objects. Extensive experiments are conducted on several datasets, and the results show that the accuracy of the method in this paper is better than the current state-of-the-art methods. Ablation experiments are conducted on the CCTSDB dataset, and the final accuracy reaches 98.50%, with an accuracy improvement of 1.32% compared to the baseline model, while the model is only 4.7 MB, and the real-time detection frame rate reaches 44 frames per second.

Keywords: traffic sign detection; activation function; foreground attention; feature fusion; VariFocalLoss; GIoU

收稿日期:2022-05-24 Received Date: 2022-05-24

*基金项目:国家自然科学基金(62073155)项目资助

0 引 言

随着人工智能技术快速发展,自动驾驶[1]成为目前热门研究领域之一。交通标志的高效检测[2]是交通道路[3]安全的重要保障。通过对交通标志识别,交由车载智能系统对突发情况做出抉择,减少驾驶员发生意外的几率,保证驾驶员安全。由于交通标志检测存在物体成像尺度小和变化大等问题,在检测中会出现漏检,且交通标志在形状上存在相似性,容易导致误检。

目前,交通标志的检测方法主要分为两种,包括传统检测算法和基于深度学习的检测算法。传统检测算法首先通过滑动窗口分割出感兴趣区域,然后采用特征提取的方法获得交通标志的特征信息,最后通过机器学习识别交通标志类别。文献[4]通过对交通标志图像进行边缘检测,再根据图像的形状特征实现实时定位。由于滑动窗口产生的候选目标物冗余度较高,导致实时性和准确性都无法达到交通标志检测的要求。

基于深度学习的交通标志检测算法,通过搭建卷积神经网络(convolutional neural network, CNN)学习交通标志特征,再利用网络训练的方式实现交通标志的分类和定位。目前,基于深度学习的交通标志检测算法主要分为双阶段检测算法和单阶段检测算法。双阶段检测算法先使用选择性搜索算法,提取候选框,然后对候选框目标进行二次修正得到检测结果,代表算法有:R-CNN[5]、Fast R-CNN[6]和 Faster R-CNN[7]等。双阶段检测算法发展迅速,检测精度也在不断提升,但其自身结构限制了检测速度。而单阶段检测算法取消候选框推荐阶段,可以在一个阶段内直接确定目标类别和定位,代表算法有 YOLO[8-11]系列和 SSD[12-13]系列。以上算法均设置了锚框[14],在训练前通过聚类来确定一组最优的锚框,通过锚框实现对目标物的捕捉,但这也带来了一些问题:1)聚类得到的锚框只能用在特定数据集上;2)锚框增加了检测头复杂度。

因此,文献[15]设计了一种无锚框的检测算法,直接利用关键点的特征热力图进行检测,加快检测速度。文献[16]利用 YOLOv3 构建交通标志检测网络,提升检测速度,但是检测准确率还有待提升。文献[17]提出 MSA-YOLOv3,采用数据增强,并引入多尺度空间金字塔池化模块有效学习底层特征信息,提升了准确率,但是模型较大不易部署。基于深度学习的交通标志检测算法,在检测速度上有所提升,但检测的准确率依旧有待加强,且不易部署。

为有效提升交通标志目标检测的精度,同时便于模型在智能系统终端部署,本文提出融合前景注意力的轻量级交通标志检测方法。首先,引入 SiLU 激活函数,提升模型检测的准确率;其次,设计了一种基于鬼影模块的轻量级骨干网络,有效提取目标物特征;然后,采用前景注意力感知模块,减少背景噪声,聚焦于目标物体;借鉴残差结构,改进路径聚合网络,增强网络语义信息的同时有效学习底层特征信息;最后,为了进一步提高算法的分类和定位准确性,利用 VariFocalLoss[18]和 GIoU[19],分别做目标的分类损失和回归损失,加强对负样本的惩罚权重。

在 Pascal VOC,中国交通数据集(CSUST Chinese Traffic Sign Detection Benchmark, CCTSDB)和清华-腾讯 100K(Tsinghua-Tencent 100K 2021, TT100K_2021)上的实验证明,与不同的主流算法相比,本文算法在不同的指标上均有优势。在 CCTSDB 上的消融实验证明,本文算法的精度达到 98.50%,超过 YOLOX[20] 1.32%,权重仅 4.7 MB,检测速度达到 44 FPS。

1 算法整体结构

本文提出的融合前景注意力的轻量级交通标志检测方法,其结构如图 1 所示。

1.1 激活函数

YOLOX 延用 YOLOv3 的 LeakyReLU 激活函数,如图 2 所示。

其数学表达式为:

$$y = \max(0,x) + leak \cdot \min(0,x) \tag{1}$$

LeakyReLU 的优势在于保留了部分负轴的信息,但缺点是无法为正负输入值提供一致的关系预测。为避免上述问题,本文引入 SiLU 激活函数,如图 3 所示,其数学表达式为:

$$y = x \cdot sigmoid(x) \tag{2}$$

SiLU 既解决了负轴神经元坏死问题,又为正负输入值提供了区别性预测。

1.2 轻量级骨干网络

YOLOX 的骨干网络 CSPDarknet 是由基于跨阶段部分模块(cross stage partial, CSP)堆叠而成。本文在此基础上,提出一种基于鬼影模块的轻量级骨干网络 GCSPDatknet,由鬼影跨阶段部分模块(ghost cross stage partial, GCSP)堆叠而成。

CSP 模块结构如图 4(a)所示,CBA 由卷积操作(convolution)、批量标准化(batch normalization)和激活函数(activation function)搭建,Bottleneck 是参考 Resnet[21]中的瓶颈模块,K 表示 Bottleneck 堆叠的数量,CSP 模块的表达式为:

$$y = part1(x) \oplus part2(x) \tag{3}$$

式中:$part1(x)$ 表示残差,$part2(x)$ 表示卷积操作,$\oplus$

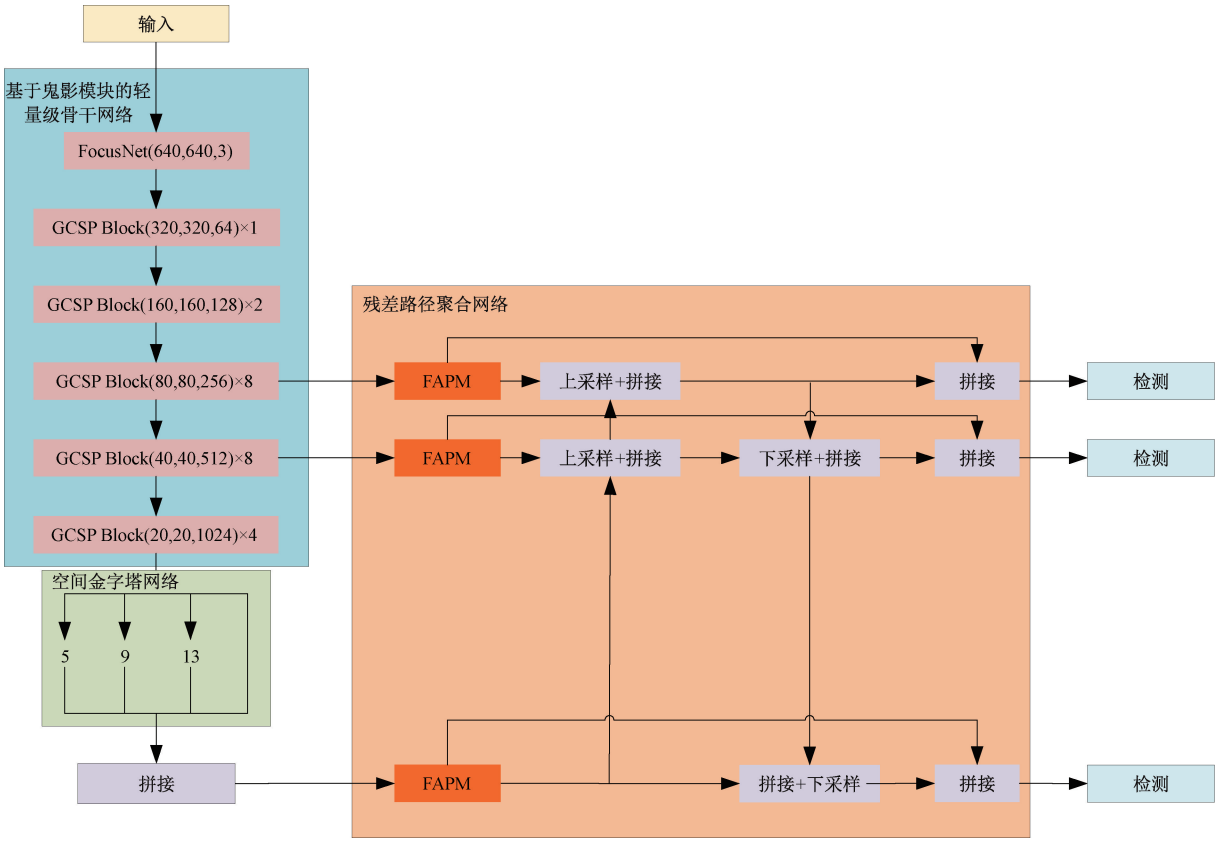


图 1　YOLOT 的结构示意图

Fig. 1　Structure diagram of YOLOT

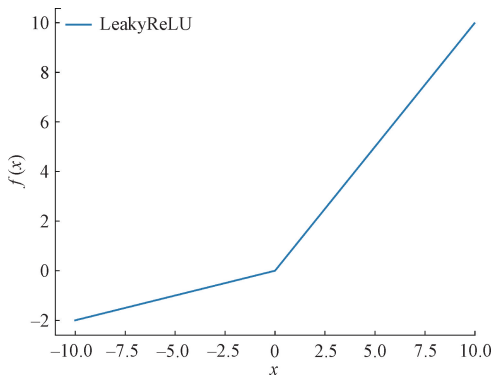


图 2　LeakyReLU 函数图

Fig. 2　LeakyReLU function diagram

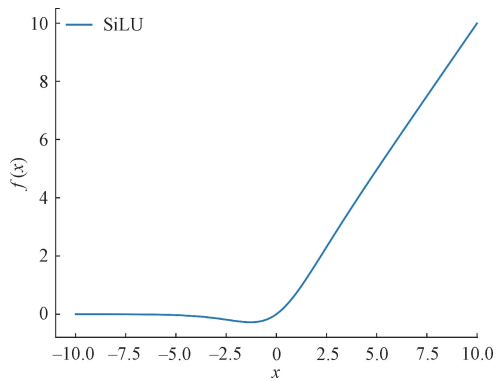


图 3　SiLU 函数图

Fig. 3　SiLU function diagram

表示按通道维度拼接操作。

利用 GhostNet[22] 的鬼影模块改进 CSP 模块，结构如图 4(b)所示，表达式为：

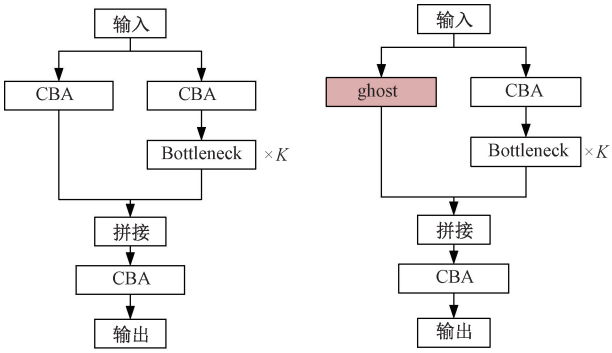
$$y = ghost(x) \oplus part2(x) \tag{4}$$

式中：*ghost* 表示鬼影模块，其结构如图 5 所示，利用鬼影模块代替 CSP 模块中的残差部分，实现高效特征提取。

1.3　前景注意力感知模块

在实时目标检测任务中，背景环境复杂，噪声会影响检测效果。因此，本文在网络中加入前景注意力感知模块(foreground attention perception module，FAPM)，对特征进行通道加权，抑制背景噪声。借鉴 CBAM[23] 注意力，本文的 FAPM 包括通道注意力和空间注意力，既考虑不同通道像素的重要性，又考虑到同一通道不同位置像素的重要性，其结构如图 6 所示。

通道注意力在空间维度进行全局最大池化和全局平均池化，得到 $1 \times 1 \times C$ 的特征描述，与原本的特征相乘，



(a) CSP模块 (b) GCSP模块
(a) CSP block (b) GCSP block

图 4 CSP 模块结构和 GCSP 模块结构

Fig. 4 Structure diagram of CSP block and GSP block

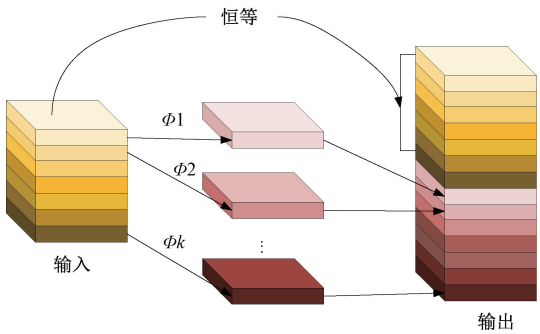


图 5 鬼影模块结构

Fig. 5 Structure diagram of ghost module

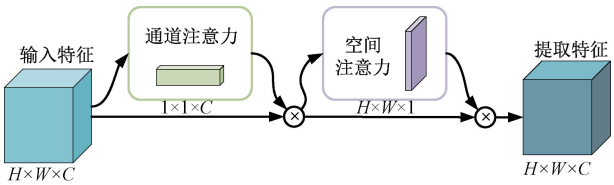


图 6 FAPM 结构

Fig. 6 Structure diagram of FAPM

进行特征加权,放大重要通道,其数学表达式为:

$$S1 = M_c(S) \otimes S \tag{5}$$

式中: S 表示输入特征图, M_c 表示获取通道注意力特征描述的操作,⊗表示元素乘积运算, $S1$ 表示通道注意力特征图。

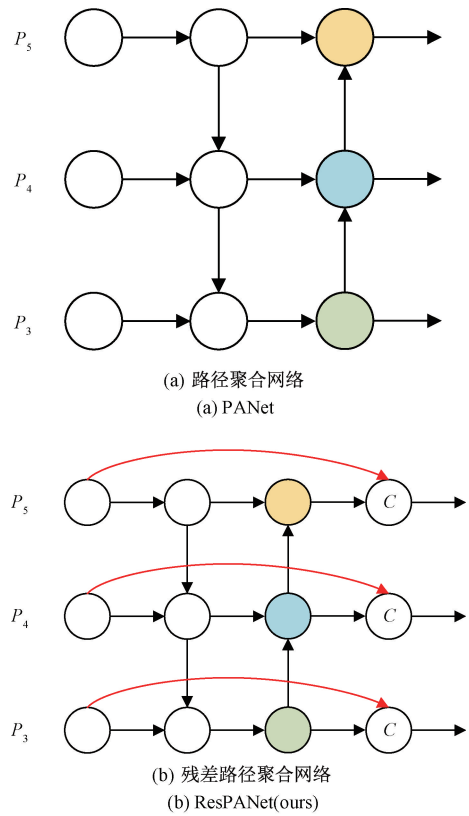
空间注意力机制将通道注意力特征图作为输入,在通道维度分别进行全局最大池化和全局平均池化,得到 $H \times W \times 1$ 特征描述,与输入的特征相乘,放大重要空间特征,得到最终特征,其数学表达式为:

$$S2 = M_s(S1) \otimes S1 \tag{6}$$

式中: $S1$ 表示输入特征图, M_s 表示获取空间注意力特征描述的操作,⊗表示元素乘积运算, $S2$ 表示最终的特征图。

1.4 残差路径聚合网络

神经网络中,深层特征图分辨率低,具有较强语义信息;而浅层特征图分辨率高,具有较强纹理信息。YOLOX 使用路径聚合网络 PANet[24] 融合多尺度特征,但经过逆向和正向的双向融合,增加网络深度,丢失了部分底层特征信息。因此,本文借鉴深度残差网络的思想,在 PANet 中加入残差结构,保留底部特征信息,生产残差路径聚合网络 ResPANet,其结构如图 7 所示。



(a) 路径聚合网络 (a) PANet

(b) 残差路径聚合网络 (b) ResPANet(ours)

图 7 PANet 结构以及 ResPANet 的结构

Fig. 7 Structure diagram of PANet and ResPANet

图 7(a)表示 PANet 结构, P_3 表示浅层特征, P_5 表示深层特征。首先,进行逆向特征融合,由 P_5 上采样,与 P_4 拼接,生成新特征 N_4,再由 P_4 上采样,与 P_3 拼接,生成新特征 N_3,其数学表达式为:

$$N_i = C(3, upSmapling(P_{i+1})) \oplus P_i \tag{7}$$

式中: N 表示新生成的特征, P 表示原特征, $C(3, -)$ 表示卷积核为 3 的卷积操作, $upSampling$ 表示上采样,⊕表示通道维度拼接。

然后进行正向特征融合,由 N_3 下采样,与 N_4 拼接,生成新特征 F_4,再由 N_4 下采样,与 P_5 拼接,生成新特征 F_5,数学表达式为:

$$F_{i+1} = C(3, subSampled(N_i)) \oplus N_{i+1} \tag{8}$$

式中：F 表示新生成的特征，N 表示原特征，$C(3,-)$ 表示卷积核为 3 的卷积操作，$subSampled$ 表示下采样，$\oplus$ 表示通道维度拼接。

图 7(b)表示 ResPANet 结构，在 PANet 中加入残差结构，将生成的新特征与底层特征拼接，得到最终特征，如以下公式所示：

$$F_5 = F_5 \oplus P_5 \tag{9}$$

$$F_4 = F_4 \oplus P_4 \tag{10}$$

$$F_3 = N_3 \oplus P_3 \tag{11}$$

将进行逆向融合和正向融合的特征与底部特征拼接，融合多尺度特征，又保留底层特征信息进行充分学习，促进后续网络进行检测。

1.5　损失函数

本文损失函数包括分类损失和回归损失。

1)分类损失

分类损失包括预测损失和置信度损失。预测损失使用二元交叉熵损失函数，即

$$loss = -(y\log(x) + (1-y)\log(1-x)) \tag{12}$$

式中：x 表示预测结果，y 表示真实标签。

置信度损失采用 VariFocalLoss，加强对负样本的训练，即：

$$loss = \begin{cases} -y(y\log(x) + (1-y)\log(1-x)), & y > 0 \\ -\alpha(sigmoid(x) - y)^{\gamma} \cdot \log(1-x), & y <= 0 \end{cases} \tag{13}$$

式中：x 表示预测结果，y 表示真实标签，α 表示负样本平衡因子，本文设置为 0.75，γ 表示调制因子，本文设置为 2.0。

2)回归损失

回归损失预测目标的坐标位置，YOLOX 用 IoU 计算真实框和预测框的损失，其不足是：

(1)当目标框和真实框不相交时，IoU 为 0，即损失为 0，网络不能更新。

(2)当目标框和真实框存在平移旋转时，IoU 相同，但是重合程度不同。

因此，本文采用 GIoU，如图 8 所示。

图 8 中 A 和 B 分别代表真实框和预测框，C 表示 A 和 B 的最小外包框，其数学表达式为：

$$GIoU(A,B) = IoU - \frac{C - A \cup B}{C} \tag{14}$$

式中：C 为真实框和预测框最小外包面积，$A \cup B$ 为预测框和真实框覆盖的并集面积，当 IoU 为 0，$GIoU$ 依旧有值，改进的回归损失函数定义如下：

$$loss = -\sum_{x,y \in R^P} 1 - GIoU(P_{bbox}^{x,y}, G_{bbox}^{x,y}) \tag{15}$$

式中：$P_{bbox}^{x,y}$ 表示点 (x,y) 所属的预测框，$G_{bbox}^{x,y}$ 表示点 (x,y) 所属的真实框。

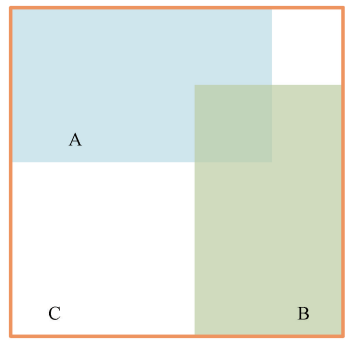


图 8　GIoU 示意图

Fig. 8　GIoU schematic diagram

2　实验与结果分析

2.1　数据集

本文在 Pascal VOC、CCTSDB 和 TT100K_2021 上验证算法的有效性。

1)Pascal VOC

Pascal VOC 包括自然界的 20 种常见物体，本文用 07trainval 和 12trainval 进行训练，共 16 125 张训练图片，在 07test 上进行测试，共 4 952 张测试图片。

2)CCTSDB

CCTSDB 包括 3 类交通标志：警告标志(warning)、指示标志(mandatory)和禁止标志(prohibitory)，如图 9 所示，共 14 229 张图片，其中 13 829 张用于训练，共 400 张用于测试。

(a) 警告标志

(a) Warning

(b) 指示标志

(b) Mandatory

(c) 禁止标志

(c) Prohibitory

图 9　CCTSDB 数据集的目标类别

Fig. 9　Target categories for the CCTSDB dataset

3)TT100K_2021

TT100K_2021 包括 3 类交通标志：警告标志(warning)、指示标志(instruction)和禁止标志

(prohibitory),与 CCTSDB 不同的是,TT100K_2021 在每个大类下分出具体的子类,如表 1 所示。最终,总共 4 552 张图片用于训练,2 314 张图片用于测试。

表 1 TT100K_2021 数据集分类

Table 1 TT100K_2021 dataset classification

类别名称	子类名称
warning	w13,w32,w55,w57,w59,wo
instruction	i2,i4,i5,il100,il60,il80,io,ip
prohibitory	p10,p11,p12,p19,p23,p26,p27,p3,p5,p6,pg,ph4,ph4.5,ph5,pl100,pl120,pl20,pl30,pl40,pl5,pl50,pl60,pl70,pl80,pm20,pm30,pm55,pn,pne,po,pr40

2.2 实验设置

1)实验环境设置

实验环境如表 2 所示,CPU 为 Intel(R)Core(TM)i7-10700,显卡为 NVIDIA GeForce RTX 3060,内存 32 GB,操作系统 Windows10,版本号 21H2,深度学习框架 Pytorch1.8,CUDA 版本 11.1,cuDNN 版本 8.1,Python 版本 3.8。

表 2 实验环境设置

Table 2 Experimental environment setup

类别	环境条件
CPU	Intel(R) Core(TM) i7-10700
显卡	NVIDIA GeForce RTX 3060
内存	32 G
操作系统	Windows10 21H2
深度学习框架	Pytorch1.8
CUDA 版本	CUDA11.1
cuDNN 版本	cuDNN8.1
脚本语言	Python3.8

2)实验参数设置

实验参数设置如表 3 所示。图片输入尺寸为 640×640,初始学习率(learning rate, lr)为 1×10^{-3},最小学习率 1×10^{-5},模型在训练集上迭代训练 100 次,迭代的批次大小设置为 2、16、32,根据模型大小而定,优化器采用随机梯度下降(stochastic gradient descent, SGD),学习率衰减为余弦退火(consine annealing, cos),动量设置 0.937,权重衰减设置 5×10^{-4},读取线程设置为 4。

表 3 实验参数设置

Table 3 Experimental parameter setting

参数名称	参数值
输入尺寸	-
初始学习率	1×10^{-3}
最小学习率	1×10^{-5}
迭代次数	100
批量大小	2/16/32
优化器	sgd
学习率衰减	cos
动量	0.937
权重衰减	5×10^{-4}
线程数量	4

与传统的学习率衰减方法不同,cos 随着 epoch 增加,学习率先模拟余弦函数快速下降,再线性上升,不断重复该过程,如图 10 所示。

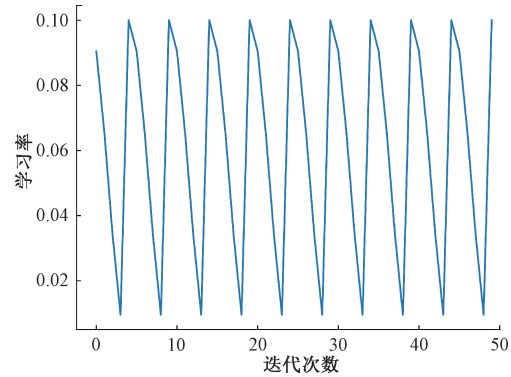


图 10 余弦退火优化算法下学习率的变化

Fig. 10 Variation of learning rate under cosine annealing optimization algorithm

2.3 模型评价指标

平均精度(mean average precision,mAP)是评价模型精度的重要指标。根据准确率和召回率绘制曲线,在 0~1 范围内曲线和坐标轴之间的面积即精度均值 AP,所有类别 AP 的均值评价模型检测准确度,即:

$$Precision = \frac{TP}{TP + FP} \tag{16}$$

$$Recall = \frac{TP}{TP + FN} \tag{17}$$

$$AP = \int_0^1 P(Recall)\,\mathrm{d}R \tag{18}$$

$$mAP = \frac{1}{c}\sum_{i=1}^{c} AP_i \tag{19}$$

式中:*Precision* 表示准确率,*Recall* 表示召回率,*TP* 表示正确检测的正样本个数,*FP* 表示错误检测的正样本个数,*FN* 表示漏检的正样本个数。*AP* 表示单个类别的平均精度,*mAP* 是所有类别 *AP* 均值。

2.4 YOLOT 性能分析与对比

1)与轻量级模型在 VOC 上的对比

为验证 YOLOT 作为轻量级网络的有效性,与轻量级网络 MobileNet-SSD[25]、Pelee[26]、YOLOv4-Tiny 和 YOLOv5-S 进行对比实验,如表 4 所示。

分析表 4 数据可知,YOLOT 在自然目标物数据集上获得了较高的准确度,超越了主流的轻量级检测模型,证明本文算法的有效性。

2)与主流模型在 CCTSDB 上的对比

在 CCTSDB 上验证 YOLOT 的有效性,将包括 Mobilenet-SSD、YOLOv4、YOLOv4-Tiny、EfficienetDet-b0[27]、Centernet、Retinanet 等主流算法与本文算法进行对比。

分析表 5 数据可知,基于锚框的模型,如 SSD、Mobilenetv2-SSD、YOLOv4-Tiny、EfficienetDet、YOLOv4 和 YOLOv5,因为锚框与交通标志的尺寸差异较大,所以在交通标志检测上表现较差。而无锚框的检测方法,如 Centernet 和 YOLOX,在 CCTSDB 验证集上达到 97%左右的准确率。本文模型 YOLOT 精度达到 98.50%,精度超越目前主流模型,达到最高的准确率。本文验证所有模型的速度,YOLOT 达到 45 FPS。同时 YOLOT 模型仅 4.7 MB,满足实时性要求,容易在嵌入式设备部署。

3)与主流模型在 TT100K 上的对比

为验证算法的泛化能力,在 TT100K_2021 上,与目前主流算法模型 SSD、MobileNet-SSD、YOLOv3、YOLOv4-Tiny、YOLOv5-S、CenterNet、YOLOX-Nano 进行对比,结果如表 6 所示。

分析表 6 数据可知,基于锚框的模型检测结果并不理想,而无锚框检测算法 YOLOX-Nano 有不错的检测效果,准确率达到 57.60%,而本文算法相较于 YOLOX-Nano,提升 7.87%。

表 4　本文算法模型与主流轻量级算法模型在 VOC 上的性能对比

Table 4　Performance of the algorithmic model in this paper versus the mainstream lightweight algorithmic model on VOC

模型	骨干网络	输入	速度/FPS	权重/MB	mAP/%
Mobilenet-SSD	Mobilenetv2	300×300	67	23.9	67.25
Pelee	Peleenet	304×304	13	19.3	71.26
YOLOv4-Tiny	CSPDarknet	416×416	**130**	22.6	75.25
YOLOv5-S	CSPDarknet	640×640	63	27.3	75.58
YOLOT(ours)	GCSPDarknet	640×640	43	**4.8**	**77.85**

表 5　主流算法模型与本文提出改进模型在 CCTSDB 数据集上的准确率对比

Table 5　Accuracy comparison between the mainstream algorithm model and the improved model proposed in this paper on CCTSDB dataset

模型	骨干网络	输入	速度/FPS	权重/MB	AP/%			mAP/%
					mandatory	prohibitory	warning	
SSD	VGG16	300×300	81	91.6	76.49	84.12	90.84	83.82
Mobilenet-SSD	Mobilenet	300×300	98	15.3	66.50	78.57	89.61	78.23
EfficientDet-b0	Efficientnet	512×512	24	15.0	70.83	79.98	89.52	80.11
YOLOv3	Darknet	416×416	48	235	92.70	95.60	94.66	94.32
YOLOv4	CSPDarknet	416×416	36	244	84.66	93.95	91.45	90.02
YOLOv4-Tiny	CSPDarknet	416×416	**137**	22.4	90.75	93.38	92.30	92.14
YOLOv5-S	CSPDarknet	640×640	71	27.1	96.60	94.97	97.36	96.31
Centernet	Resnet50	512×512	63	125	97.30	95.98	94.71	97.35
Retinanet	Resnet50	600×600	28	139	27.89	52.16	63.75	47.93
YOLOX-Nano	CSPDarknet	640×640	58	**3.7**	98.60	94.83	98.10	97.18
YOLOX-Tiny	CSPDarknet	640×640	63	19.4	98.25	97.12	98.54	97.97
YOLOT(ours)	GCSPDarknet	640×640	45	4.7	**98.81**	**97.64**	**99.06**	**98.50**

表 6　主流算法模型与本文提出改进模型在 TT100K_2021 数据集上的准确率对比

Table 6　Accuracy comparison between the mainstream algorithm model and the improved model proposed in this paper on TT100K_2021 dataset

模型	骨干网络	输入	速度/FPS	权重/MB	mAP/%
SSD	VGG16	300×300	52	113	15.53
Mobilenet-SSD	Mobilenet	300×300	64	36.6	9.51
YOLOv3	Darknet	416×416	46	235	32.35
YOLOv4-Tiny	CSPDarknet	640×640	**129**	22.8	23.73
YOLOv5-S	CSPDarknet	640×640	69	27.5	28.94
Centernet	Resnet50	512×512	57	124	25.97
YOLOX-Nano	CSPDarknet	640×640	57	**3.7**	57.60
YOLOT(ours)	GCSPDarknet	640×640	44	4.8	**65.47**

2.5 消融实验

为了验证 SiLU、GCSPDarknet、FAPM、ResPANet、GIoU 和 VariFocalLoss 对模型性能的影响。本文将 YOLOX 算法作为基线,在 CCTSDB 上对各部分进行消融实验,训练损失曲线如图 11 所示,横坐标代表迭代次数,纵坐标代表损失值,经过 100 次迭代后,网络的损失值趋于稳定。

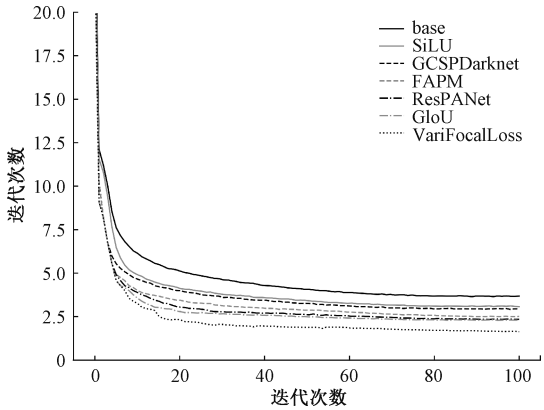


图 11 消融实验中各部分训练损失

Fig. 11 Loss of training in each component of the ablation experiment

分析表 7 模块性能影响可知,将 LeakyReLU 替换成 SiLU 后,mAP 提升 0.13%,使用 GCSPDarknet 网络后,mAP 提升 0.22%,说明鬼影模块有效提升骨干网络的特征提取能力,在引入 FAPM 后,mAP 提升 0.29%,抑制背景噪声,在加入 ResPANet 后,mAP 提升 0.22%,有效学习底部特征信息,使用 GIoU 训练后,mAP 提升 0.06%,使用 VariFocalLoss 进行训练,mAP 提升 0.22%,且各类目标的精度都达到了最优。

2.6 定性分析

从消融实验发现,GCSPDarknet 对检测结果有着较大的提升,为了更加直观的验证该部分的性能,将改进前后的模型在复杂环境下的检测结果进行定性对比分析,如图 12~14 所示。

如图 12(a)所示,雾天环境下图像的清晰度下降,检测的定位和分类难度提高,在改进前模型并没有检测出交通标志,在加入 GCSPDarknet 改进后,模型定位并检测出交通标志。

如图 13(a)是雨天环境下的图像,出现的噪声会影响模型检测结果,在改进前,模型出现漏检,在加入 GCSPDarknet 改进后,问题得到解决。

表 7 不同模块对模型性能的影响

Table 7 The impact of different modules on model performance

模型	mandatory			prohibitory			warning			mAP/%
	AP/%	Recall/%	Precision/%	AP/%	Recall/%	Precision/%	AP/%	Recall/%	Precision/%	
YOLOX-base	98.60	96.64	82.27	94.83	95.67	88.33	98.10	97.71	87.67	97.18
+SiLU	97.61	94.63	88.68	95.69	92.42	93.43	98.62	**98.47**	87.16	97.31(+0.13)
+GCSPDarknet	97.83	96.64	87.80	96.52	97.47	90.60	98.77	97.71	89.51	97.71(+0.40)
+FAPM	98.23	96.64	87.27	96.83	97.47	90.91	98.94	**98.47**	89.58	98.00(+0.29)
+ResPANet	98.53	96.64	86.75	97.17	**98.19**	**92.20**	98.97	**98.47**	89.58	98.22(+0.22)
+GIoU	98.25	97.32	88.41	97.59	97.11	91.81	99.00	97.71	90.78	98.28(+0.06)
+VariFocalLoss	**98.81**	**97.99**	**89.02**	**97.64**	**98.19**	**92.20**	**99.06**	97.71	**91.43**	**98.50(+0.22)**

(a) 原图 (a) Original image (b) 改进前 (b) Before improve (c) 改进后 (c) After improve

图 12 改进前后在雾天条件下检测结果对比

Fig. 12 Comparison of test results under foggy conditions before and after improve

(a) 原图 (a) Original image (b) 改进前 (b) Before improve (c) 改进后 (c) After improve

图 13 改进前后在雨天条件下检测结果对比

Fig. 13 Comparison of test results under rainy conditions before and after improve

如图 14(a)是经过几何变化的图像,在改进前,对该图像的检测出现 3 处漏检,而在加入 GCSPDarknet 改进后,漏检问题解决,图中所有交通标志检测出来。

2.7 检测效果可视化

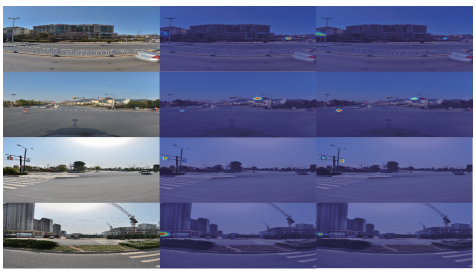
本文进行了特征图可视化,更加直观的证明 FAPM 的效果,如图 15 所示。

从图 15 可以看出,未加入 FAPM 时,网络在物体的

(a) 原图 (a) Original image　(b) 改进前 (b) Before improve　(c) 改进后 (c) After improve

图 14　改进前后在几何变化下检测结果对比

Fig. 14　Comparison of test results under geometric changes before and after improve



(a) 原图 (a) Original image　(b) 未加FAPM (b) Without FAPM　(c) 加入FAPM (c) Add FAPM

图 15　热度图可视化

Fig. 15　Heat map visualization

热度呈现出弥散或缺失,说明背景噪声对于目标物的聚焦是有影响的,加入 FAPM 后,明显抑制了物体周围噪声的影响,网络聚焦于目标,证明 FAPM 的有效性。

为对比本文算法在交通标志检测任务中的效果,这里测试了 YOLOv3、YOLOv4、YOLOv5、YOLOX 和 YOLOT 的实际检测效果,如图 16 所示。与 YOLOv3、YOLOv4、YOLOv5 和 YOLOX 相比,YOLOT 解决了存在的漏检和误检问题,同时对小目标和遮蔽目标的检测也有不错的表现。

3　结　论

本文提出一种融合前景注意力的轻量级交通标志检测方法。首先整体网络采用 SiLU 激活函数,提高模型检测的准确率;其次设计了一种基于鬼影模块的轻量化骨干网络 GCSPDarknet,有效提取目标特征;接着引入前景注意力感知模块,抑制背景噪声;然后改进路径聚合网络,在特征融合过程中加入残差结构,充分学习底部特征信息;最后引入 VariFocalLoss 和 GIoU,分别计算目标的分类预测和目标间的相似度,使目标的分类和定位更加准确。在多个数据集上与主流算法进行了对比实验,证明本文算法的有效性。改进算法在 CCTSDB 上的精度为

(a) YOLOv3　(b) YOLOv4　(c) YOLOv5　(d) YOLOX　(e) YOLOT(ours)

图 16　本文算法与其他主流算法检测效果对比

Fig. 16　The detection effect of this algorithm is compared with other mainstream algorithms

98.50%,与基线实验相比,提升 1.32%,同时模型仅 YOLOX.7MB,实时检测速度达到 44 FPS,易于在嵌入式设备上部署。改进算法有着较高的精度和较小的模型权重,但是检测速度并不突出,进一步提升检测速度是未来的研究方向。

参考文献

[1] 刘丹, 马同伟. 结合语义信息的行人检测方法[J]. 电子测量与仪器学报, 2019, 33(1): 54-60.
LIU D, MA T W. Pedestrian detection method based on semantic information [J]. Journal of Electronic Measurement and Instrumentation, 2019, 33(1): 54-60.

[2] 程换新, 郭占广, 刘文翰,等. 基于胶囊神经网络的交通标志识别研究[J]. 电子测量技术, 2020, 43(11): 112-116.
CHENG H X, GUO ZH G, LIU W H, et al. Research

on traffic sign recognition based on capsule neural network[J]. Electronic Measurement Technology, 2020, 43(11): 112-116.

[3] 郑少武, 李巍华, 胡坚耀. 基于激光点云与图像信息融合的交通环境车辆检测[J]. 仪器仪表学报, 2019, 40(12): 143-151.
ZHENG SH W, LI W H, HU J Y. Vehicle detection in the traffic environment based on the fusion of laser point cloud and image information[J]. Chinese Journal of Scientific Instrument, 2019, 40(12): 143-151.

[4] PICCIOLI G. Robust method for road sign detection and recognition[J]. Image and Vision Computing, 1996, 14(3): 209-223.

[5] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [J]. IEEE Conference on Computer Vision and Pattern Recognition, 2014, 81(1): 582-587.

[6] GIRSHICK R. Fast R-CNN[C]. Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440-1448.

[7] 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 & Machine Intelligence, 2017, 39(6): 1137-1149.

[8] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]. Computer Vision and Patterb Recognition, 2017: 6517-6525.

[9] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7263-7271.

[10] REDMON J, FARHADI A. YOLOv3: An incremental improvement[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018.

[11] BOCHKOVSKIY A, WANG C Y, LIAO H. YOLOv4: Optimal speed and accuracy of object detector[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2020.

[12] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector [C]. Computer Vision ECCV 2016. Cham: Springer, 2016: 21-37.

[13] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017 (99): 2999-3007.

[14] 张培培, 王昭, 王菲. 基于深度学习的图像目标检测算法研究[J]. 国外电子测量技术, 2020, 39(08): 34-39.
ZHANG P P, WAN ZH, WAN F. Research on image target detection algorithm based depth learning [J]. Foreign Electronic Measurement Technology, 2020, 39(8): 34-39.

[15] DUAN K, BAI S, XIE L, et al. Centernet: Keypoint triplets for object detection [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 6569-6578.

[16] RAJENDRAN S P, SHINE L, PRADEEP R, et al. Real-time traffic sign recognition using YOLOv3 based detector [C]. 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT). IEEE, 2019.

[17] 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.

[18] ZHANG H, WANG Y, DAYOUB F, et al. VarifocalNet: An IoU-aware dense object detector[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2021.

[19] REZATOFIGHI H, TSOI N, GWAK J Y, et al. Generalized intersection over union: A metric and a loss for bounding box regression[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2019.

[20] GE Z, LIU S, WANG F, et al. YOLOx: Exceeding YOLO series in 2021 [J]. arXiv preprint arXiv: 2107.08430, 2021.

[21] HE K, ZHANG X, REN S, et al. Deep residual learing for image recognition [C]. Proceeding of the IEEE Conference on Computer Vision and Pattern Recoginition, 2016: 770-778.

[22] HAN K, WANG Y, TIAN Q, et al. GhostNet: More features from cheap operations[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2020.

[23] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module [C]. European Conference on Computer Vision. Munich, Germany, 2018: 3-19.

[24] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8759-8768.

[25] SANDLER M, HOWARD A, ZHU M, et al. MobileNetV2: Inverted residuals and linear bottleneckss [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4510-4520.

[26] WANG R J, LI X, LING C X. Pelee: A real-time object detection system on mobile devices[C]. Advances in Neural Information Processing Systems, 2018: 1963-1972.

[27] TAN M, PANG R, LE Q V. EfficientDet: Scalable and efficient object detection[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2020.

作者简介

俞林森,2020 年于南通大学获得学士学位,现为江南大学硕士研究生,主要研究方向为目标检测、轻量化神经网络。
E-mail: 1812727794@qq.com

Yu Linsen received his B. Sc. degree from Nantong University in 2020. Now he is a M. Sc. candidate in Jiangnan University. His main research interests include object detection and lightweight neural network.

陈志国(通信作者),2001 年于江南大学获得学士学位,2006 年于江南大学获得硕士学位,2010 年于江南大学获得博士学位,现为江南大学副教授,主要研究方向为人工智能、机器学习。
E-mail: chenzg777@163.com

Chen Zhiguo (Corresponding author) received his B. Sc. degree from Jiangnan University in 2001, M. Sc. degree from Jiangnan University in 2006, and Ph. D. degree from Jiangnan University in 2010, respectively. Now he is an associate professor in Jiangnan University. His main research interests include artificial intelligence and machine learning.