激光与光电子学进展 先进成像

基于注意力机制改进的轻量级目标检测算法

金梅，李义辉*，张立国，马子荐
燕山大学电气工程学院，河北 秦皇岛 066000

摘要 针对通用的目标检测算法在检测生活场景下的多类目标时检测精度低、速度较慢的问题，提出了一种基于注意力机制改进的轻量级目标检测算法YOLOv4s。该算法以CSPDarknet53-s作为主干特征提取网络提取图像特征，通过注意力模块进行特征选择，再利用特征金字塔网络对特征进行融合，最后通过检测头分别处理特征融合后的两个输出，进而提高对生活场景下多类目标检测的能力。实验结果表明：相比改进前的算法，YOLOv4s算法在PASCAL VOC数据集上的平均均值精度（mAP）及MS COCO数据集上的平均精度（AP）都有一定程度的提升；相较于轻量级算法Efficientdet，YOLOv4s算法在MS COCO数据集上的AP也有一定提高，并且实现了有效的显著目标检测。

关键词 机器视觉；目标检测；轻量级神经网络；注意力机制；特征金字塔；YOLOv4s
中图分类号 TP391.4 **文献标志码** A **DOI**：10.3788/LOP212947

Attention Mechanism-Based Improved Lightweight Target Detection Algorithm

Jin Mei, Li Yihui*, Zhang Liguo, Ma Zijian
School of Electrical Engineering, Yanshan University, Qinhuangdao 066000, *Hebei, China*

Abstract A lightweight YOLOv4s based on an attention mechanism is proposed to address the low accuracy and slow speed issue of the general target detection algorithm in multi-target life scenes. First, CSPDarknet53-s was used as the backbone network to extract image features, and these features were selected using the attention block. Subsequently, the feature pyramid network was adopted to fuse the features. Finally, the YOLOv4s head was used to process the two outputs after the feature fusion to improve the multi-target detection ability in living scenes. According to the experiment results, the YOLOv4s algorithm outperforms the prior algorithm in the PASCAL VOC and MS COCO datasets, exhibiting improvement in the mean average precision and average precision. Compared with the lightweight algorithm Efficientdet, the YOLOv4s algorithm also has a certain improvement in the AP on the MS COCO dataset, and achieves effective significant target detection.

Key words machine vision; object detection; lightweight neural network; attention mechanism; feature pyramid; YOLOv4s

1 引 言

目标检测的任务是找出图像中所有感兴趣的目标（物体），并确定它们的类别和位置，是计算机视觉领域的核心问题之一。由于各类物体有不同的外观、形状和姿态，加上成像时光照、遮挡等因素的干扰，目标检测一直是计算机视觉领域最具有挑战性的问题[1]。

在计算机视觉领域中，目标检测在各行各业应用广泛，在医疗卫生、公共交通、遥感检测及军事科技等方面有许多的研究和进展，但一些设备对模型的检测速度及大小仍然有一些要求，因此如何在提升检测精度的基础上提高速度是目标检测领域发展的方向[2]。目标检测领域已经逐渐从传统检测方法向基于深度学习的识别方法转变。目前，在目标检测领域比较流行的有两类算法：两阶段算法和一阶段算法。两阶段算法顾名思义就是由两个阶段组成的算法，首先在图像上输出大概包含的候选区域，之后对候选区域进行类别分类和位置回归。常见的两阶段算法有R-CNN[3]、Fast R-CNN[4]、Faster R-CNN[5]。两阶段算法精度高但速度慢，而一阶段算法由于不需要候选框，直接利用

收稿日期：2021-11-12；**修回日期**：2021-12-13；**录用日期**：2022-01-05；**网络首发日期**：2022-01-16
基金项目：河北省科学技术研究与发展计划科技支撑计划（20310302D）、河北省中央引导地方专项（199477141G）
通信作者：*liyihui819@163.com

神经网络得到分类及回归结果，节省大量的时间，并且占用的空间也较小，代表的一阶段算法有SSD[6]、RetinaNet[7]、YOLO系列[8-11]等。在基于深度学习的目标检测中，神经网络为其提供了最先进（SOTA）的方法，取得了非常优异的成绩。但是在这些成绩中，大部分都过分依赖于较高的计算量，在实际应用中没有办法通过廉价的设备去使用这些较为先进的方法。Wang等[12]从网络结构的角度出发，提出了跨阶段局部网络（CSPNet）。在网络优化过程中，存在着许多重复的梯度信息，CSPNet将网络特征图中的各个阶段整合起来，对于梯度的多样性比较敏感。通过ImageNet中的实验，与其他计算方法相比，CSPNet降低了20%的计算量，而准确率在保持不变的情况下甚至有所提升。在MS COCO数据集上，CSPNet的AP50对比其他的SOTA方法而言有很明显的提升，并且CSPNet相对来说比较通用且容易实现，可以与ResNet[13]、ResNeXt[14]、DenseNet[15]等融合，从而达到更好的效果。CSPNet在卷积神经网络（CNN）的学习能力方面进行了提升，能够在轻量化的同时继续维持较高的准确性，也降低了计算瓶颈和内存成本。虽然现在较为普遍的目标检测算法已经很大程度上提升了目标检测的精度及速率，但是由于在神经网络的深层网络中更易提取到图像的信息，网络结构往往过于庞大、参数量较多，将其应用到嵌入式的系统或平台上时显得过于笨重。现有的轻量级网络，例如MobileNetv3[16]、EfficientNet[17]，其精度及速率还达不到要求，所以较高的准确度及网络结构的轻量化在网络需要在嵌入式平台上进行工作时可起到至关重要的作用[18]。

本文在YOLOv4-tiny[19]算法的基础上进行改进，原网络由于过分追求轻量化，存在检测精度差的问题，基于此对主干网络中的部分结构、特征提取网络及激活函数进行了改进，使其保持高精度的同时，网络结构还相对较小。在主干网络中，将第2层的特征提取结构darknet改为resblock_body并将激活函数为修改leaky ReLU，加快网络进程。此外，在主干网络后增加两个注意力模块（ATTblock），并对特征金字塔网络（FPN）[20]进行改进，形成注意力特征金字塔结构（ATT-FPN），使主干网络提取到的特征先通过ATTblock进行通道加权，再让高层特征与底层特征互相融合，使得最终的结果既能满足大目标的检测也能检测较小目标。实验结果表明，所提算法不仅提高了检测精度和速度，而且满足轻量化的要求。

2　YOLOv4s目标检测算法

2.1　YOLOv4s整体结构

YOLOv4s主要由主干特征提取网络CSPdarknet53-s和基于注意力机制的加强特征提取网络ATT-FPN组成。CSPdarknet53-s由YOLOv4-tiny中的主干特征提取网络CSPdarknet53s改进而来，结构如图1所示，输入为416×416的图像，主干特征提取网络中残差块由3个增长为4个，即将原来第2层的特征提取结构

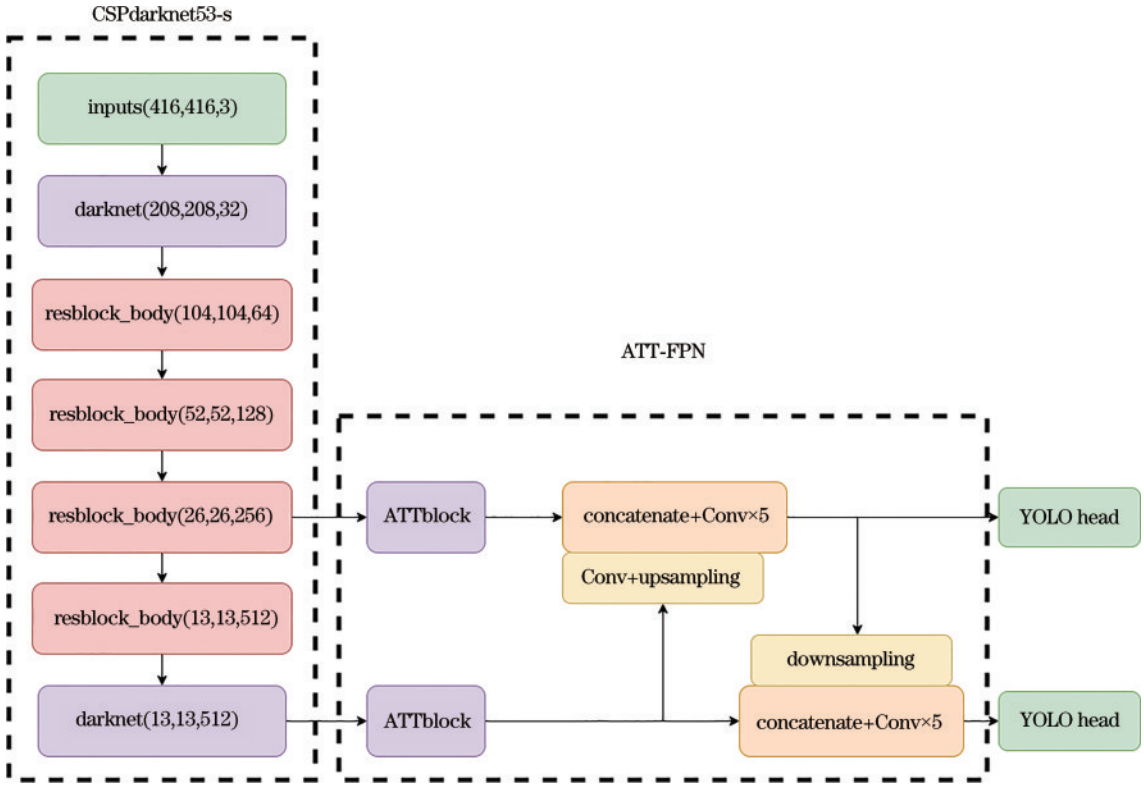

图1　YOLOv4s结构示意图

Fig. 1　YOLOv4s structure diagram

darknet改为resblock_body结构，在加强特征提取网络前增加了ATTblock注意力模块进行特征的通道加权，最终输出的两个特征层的尺寸为原输入尺寸的1/16和1/32，并得到浅层的目标位置信息和深层的语义特征，将其输入特征金字塔结构中，对两个输出特征进行concatenate拼接及上下采样，保留两个尺度特征的同时又进行了特征融合。由两个YOLO head结构推理出最终的分类预测结果，在预测过程中还利用非极大值抑制(NMS)[21]去除输出框中相对重合度较高的框，得到最终的预测框。

2.2　CSPDarknet53-s结构

基于CSPNet的目标检测器主要解决以下3个问题[12]：提升CNN的学习能力；去掉瓶颈结构的计算中算力较高的部分；降低内存占用。所采用的CSPdarknet53-s结构就由CSPNet改进而来，由2个darknet结构及4个resblock_body结构组成。在YOLOv4-tiny中的CSPdarknet53结构中，darknet结构由卷积、批归一化(BN)及Mish激活函数构成，而在YOLOv4s的darknet结构中，激活函数修改为leaky ReLU，这是由于此激活函数能够加快网络进程，达到提升网络检测速度的效果。而resblock_body结构块借鉴了CSPNet的结构，主干部分进行原来的残差块的堆叠，另一部分像一个残差边一样，经过一小部分处理直接连接到最后，相当于一次下采样和多次残差结构的堆叠。CSPdarknet53-s具体结构如图2所示，第3层由之前的darknet结构改为resblock_body结构块后，提高了对网络浅层特征的融合及空间细节信息的提取，此外CSPdarknet53-s的使用也减轻了现有方法对于高计算成本的依赖。

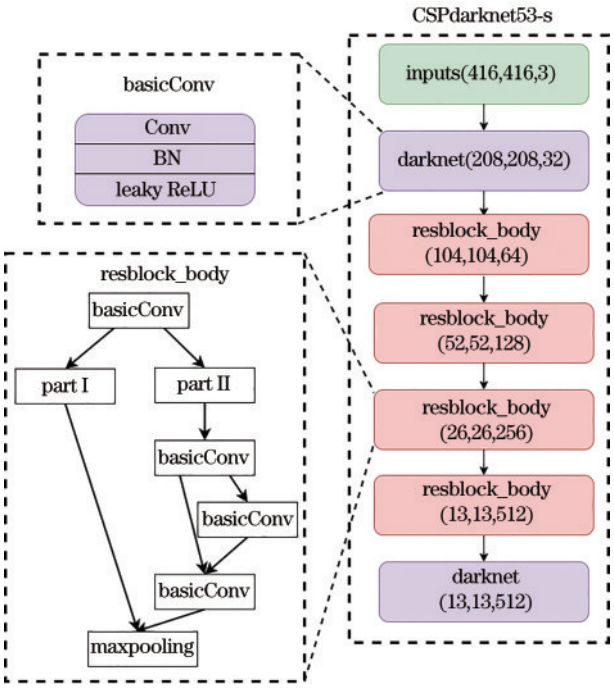

图2　CSPDarknet53-s结构示意图

Fig. 2　CSPDarknet53-s structure diagram

2.3　基于注意力的加强特征提取结构ATT-FPN

在神经网络中，感受野对于目标检测意义重大，适当的感受野可以提高检测物体的效果。因此，通过增大感受野得到被检测物体周围的有效上下文信息，能够提升对目标物体的检测效果。

FPN由Lin等[20]于2017年提出，FPN是一种多尺度自顶向下进行特征融合的目标检测方法，多尺度是指有多个特征预测层，例如在YOLOv4和YOLOv4-tiny中分别有3个和2个有效特征层。而在此算法提出以前，有些算法采用多尺度特征融合的方法来进行目标的预测，但是大部分算法都是将多个特征融合到一个尺度再进行预测，这就使得在预测过程中产生了一些偏差，融合后的结果对于最终的检测精度会产生一定的影响，而借助注意力机制可以使模型自动学习到不同channel的重要程度，故在主干网络与加强特征提取网络之间加入ATTblock注意力模块，将主干特征提取网络提取到的特征进行对应层数的通道加权，使得每个输出都是在对应大小的特征上进行提取的，更有利于后续特征的融合，从而提高对目标的检测精度。

由于ATT-FPN结构中有不同的预测层，ATT-FPN的模型训练方式与传统的Faster R-CNN的训练方式有些差异。与SSD方法有些类似，在ATT-FPN中位置相对靠前的拥有更高的分辨率，而在对较小目标进行预测时，高分辨率的特征起到了重要作用。在ATT-FPN结构中进行每个特征的融合时，使用K-means聚类去生成图像的候选框，然后对所有生成的候选框进行组合。一般来说，更小的目标应在更高分辨率的特征图上进行预测，所以需要分配这些候选框，分配的原理为

$$k = F\left[k_0 + \log_2\left(\sqrt{wh}/224\right)\right], \tag{1}$$

式中：k_0是一个常数，代表在分配的过程中分配到的预测层；w和h分别代表候选框的宽和高；$F[\cdot]$代表取数值的下界。例如，预测层有1、2、3、4等4个，分别代表了分辨率由大到小的4个特征融合层。k_0=4，将宽和高的乘积为112×112的候选框分配给k=3的预测层(k大于4时取4，k小于1时取1)，然后再进行后续检测。由式(1)可知，候选框的面积相对较大时，被分配到的预测层的数字也会变得越大，大的目标会被分配到分辨率较低的预测层进行预测；同理，较小目标会被分配到分辨率较高的预测层进行预测。从上一步骤得到的候选框进行预测层的分配，再送入相应的RoI pooling层，将RoI pooling层输出的结果进行融合，再经过两个全连接层(FC)，最后进行目标分类和位置回归。

ATT-FPN主要由ATTblock和FPN两部分组成，如图3所示。ATTblock主要由CBS、全局平均池化

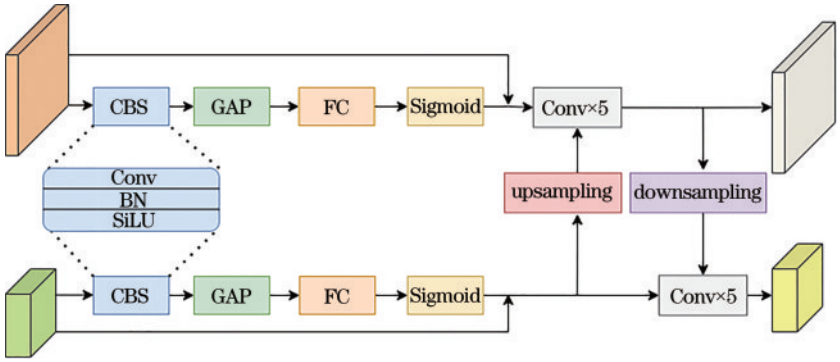

图 3　ATT-FPN 结构示意图

Fig. 3　ATT-FPN structure diagram

(GAP)、全连接层、Sigmoid 激活函数构成，其中 CBS 由卷积、批归一化、SiLU 激活函数组成。主干特征提取网络中的两个输出分为两路，一路通过 ATTblock 得到注意力通道加权后的特征，另一路像残差结构一样直接接到末尾与注意力结构的输出进行拼接，进而输入 FPN 中。SiLU 是 Sigmoid 和 ReLU 的改进版，具备无上界、有下界、平滑、非单调的特性，如图 4 所示。SiLU 在深层模型上的效果优于 ReLU，可以看作是平滑的 ReLU 激活函数。SiLU 激活函数表达式为

$$f(x)=x\cdot\mathrm{Sigmoid}(x)。\tag{2}$$

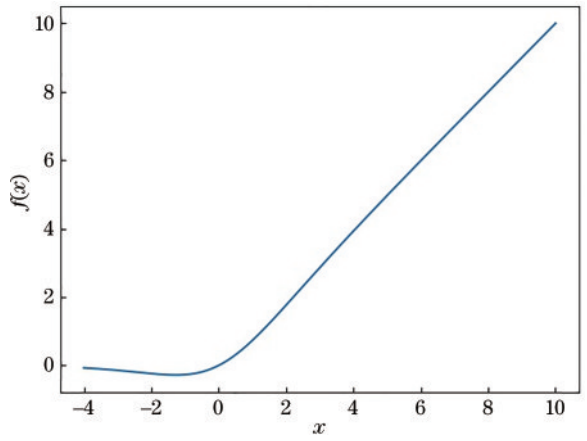

图 4　SiLU 激活函数

Fig. 4　SiLU activation function

所提 ATT-FPN 结构结合 FPN 的特点并融合注意力模块 ATTblock 进行通道加权，对最后两层特征层进行特征的反复提取，并且融合最后两层特征层，输入的两个特征层的尺寸为原输入尺寸的 1/16 和 1/32，这样就可以对大目标和较小目标的特征进行融合，使得多个不同尺度的特征融合得更加精确并分别进行预测，从而达到良好的检测效果。

2.4　YOLOV4s 损失函数

在 YOLOv4s 中，损失的计算主要分为 3 部分：位置损失、置信度损失和类别损失，即整个网络的损失函数为

$$L=L_{\mathrm{loc}}+L_{\mathrm{conf}}+L_{\mathrm{cls}}。\tag{3}$$

置信度损失和类别损失在 YOLOv4s 中均用二值交叉熵损失(BCE)计算。置信度损失分为两部分：第 1 部分是实际上存在目标的，将预测结果置信度的值与 1 进行比较；第 2 部分是实际上不存在目标的，在计算位置损失时将其最大的交并比(IOU)值与 0 进行比较，此部分是为了去除被忽略的不包含目标的框。类别损失表示实际上存在目标的框中预测类别与真实类别的差异。

位置损失的计算在 YOLOv4 中已经改进为 CIOU[22]，CIOU 是在 IOU 的基础上提出的。IOU 作为一种评估目标检测器的指标，反映预测检测框和真实检测框的检测效果，但是将其作为损失函数会出现极端情况。当两个检测框不相交时 IOU 为 0，此时不能反映两者的距离大小即它们的重合度，同时损失也为 0，没有进行梯度回传，就无法继续进行梯度训练。因此 GIOU[23]、DIOU[24]、CIOU 相继被提出，在坐标尺度的归一化、边界框的重叠面积和中心点距离方面都进行了考虑，在 CIOU 中引入一个惩罚项来改进预测框和真实框之间长宽比的一致性。CIOU 的表达式为

$$L_{\mathrm{CIOU}}=1-L_{\mathrm{IOU}}+\frac{\rho^2(b,b^{\mathrm{gt}})}{c^2}+\alpha\upsilon,\tag{4}$$

$$\upsilon=\frac{4}{\pi^2}\left(\arctan\frac{w^{\mathrm{gt}}}{h^{\mathrm{gt}}}-\arctan\frac{w}{h}\right)^2,\tag{5}$$

$$\alpha=\frac{\upsilon}{(1-L_{\mathrm{IOU}})+\upsilon},\tag{6}$$

式中：L_{CIOU} 为 CIOU 损失，如图 5 所示；$\rho^2(b,b^{\mathrm{gt}})$ 代表

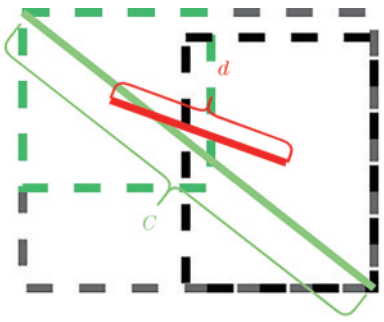

图 5　CIOU

Fig. 5　CIOU

预测框和真实框的中心点的欧氏距离；c代表的是能够同时包含预测框和真实框的最小闭包区域的对角线距离；w、h及w^{gt}、h^{gt}分别代表预测框和真实框的宽高；v用来度量宽高比的相似性；α为权重系数。

3　实验与分析

3.1　实验配置

实验是在 Windows 10 系统下进行的，采用 Pytorch 1.8.1 深度学习框架，硬件配置为 NVIDIA GeForce GTX 2060、6 GB 显存、Intel(R) Core(TM) i5-10400F CPU @ 2.90 GHz、16 GB RAM。训练过程中参数设置如下：输入大小为 416×416，采用 Adam 优化器，初始学习率为 0.001，权重衰减系数为 0.005，学习率采用余弦退火衰减，训练 200 个 epoch，batch_size 设置为 64。

3.2　实验数据集

主要在 PASCAL VOC 和 MS COCO 数据集上进行测试与训练，对于 VOC 数据集来说，训练集采用的是 VOC2012(train+val)，COCO 数据集采用的是 2017 年的版本，两种数据集的具体数量如表 1 所示。

表 1　实验数据集详情

Table 1　Experimental dataset details

Dataset	Class of number	Train set	Test set
PASCAL VOC	20	15412	1713
MS COCO	80	105539	11727

VOC2012 是 VOC2007 的升级版，拥有 20 个类别，有 11540 张图片可用于检测任务。MS COCO 数据集是具有 80 个类别的大规模数据集，包括训练集、验证集和测试集等 3 部分，每部分包含 118287、5000 和 40670 张图片，总大小约 25 GB，其中测试数据集没有标注信息，所以注释部分只有训练集和验证集。

将 COCO 数据集中训练和验证的数据集用来训练，并且将没有匹配到的图片剔除，最终剩下 117266 张图片。对于两种数据集，其中 90% 的图片用作训练，10% 的图片用作测试。

3.3　评价标准

本实验通过多个指标来评价不同算法的性能，包含精度均值(AP)、平均精度均值(mAP)、召回率(R)、F1 值、准确率(P)及每秒检测帧数(FPS)。

$$P_{\text{precision}}=\frac{N_{\text{TP}}}{N_{\text{TP}}+N_{\text{FP}}},\tag{7}$$

$$R_{\text{recall}}=\frac{N_{\text{TP}}}{N_{\text{TP}}+N_{\text{FN}}},\tag{8}$$

$$R_{\text{AP}_s}=\frac{1}{N_S}\sum_{R_{\text{recall}}\in R_{\text{recall},s}'}\rho_i(R_{\text{recall},s}'),\tag{9}$$

$$\rho_i(R_{\text{recall}})=\max_{R_{\text{recall},s}';\,R_{\text{recall},s}'\geqslant R_{\text{recall},s}}\rho(R_{\text{recall},s}'),\tag{10}$$

$$R_{\text{mAP}}=\frac{1}{N}\sum R_{\text{AP}_s},\tag{11}$$

式中：N_{TP}表示预测正确的正样本的图像数量；N_{FP}表示将负样本预测为正样本的数量；N_{FN}表示将正样本预测为负样本的数量；N_{TN}表示预测正确的负样本数量[25]；R_{AP_s}为 s 类时的精度均值；N_s为将 recall 分成的段数；$\rho_i(R_{\text{recall}})$代表将在每一段上的 precision 的最大值作为该段的代表。

3.4　实验结果对比

为了评估算法性能进行了一系列对比实验，输入图像的尺寸为 416×416，在计算召回率及精确率时，阈值设置为 0.3，IOU 设置为 0.5。将 YOLOv4s、YOLOv4-tiny、YOLOv4 在相同的训练集和测试集上进行相同批次的训练，得到对比实验的结果并进行评估。图 6 为通过注意力机制后得到的热力图与原图的对比，从图 6 可以看出，注意力机制对自行车和人这类显著特征进行了捕获，有利于网络在检测物体时，对物体进行定位与回归。

图 6　注意力机制热力图对比

Fig. 6　Comparison of thermogram of attention mechanism

对于主干网络层数的改进及注意力机制的加入，在同种实验设备下采用相同训练和测试方法进行消融实验。将 CSPdarknet53-s、ATT-block、ATT-FPN 这 3 种改进方法作用于 YOLOv4-tiny 网络上，在 PASCAL VOC 数据集上进行实验，结果如表 2 所示。

从表 2 可以看出：在 PASCAL VOC 数据集上，原始 YOLOv4-tiny 目标检测网络的 mAP 和 Recall 分别为 67% 和 75.6%，并没有表现出很好的检测效果；在

表 2　3种改进方法对比

Table 2　Comparison of three improved methods　unit:%

Method			PASCAL VOC	
CSPdarknet53-s	ATT-block	ATT-FPN	mAP	Recall
×	×	×	67	75.6
√	×	×	73.3	79.1
√	√	×	84.5	89.6
√	√	√	87.9	91.2

主干网络修改网络结构及在加强特征提取网络添加注意力机制后，网络的 mAP 及 Recall 提升到 87.9% 和 91.2%，分别提升了 20.9 个百分点和 15.6 个百分点。消融实验结果表明，在主干特征提取网络中改进网络层数及在加强特征提取网络中加入注意力机制并进行特征融合，能够有效提升网络的精确率和召回率，且漏检和误检的现象大大减少。图 7 为两种算法在两张相同的图片上的检测效果，其中图 7(a)为 YOLOv4s 算法的检测结果，图 7(b)为 YOLOv4-tiny 算法的检测结果。YOLOv4-tiny 出现了错检并且没有很好地将目标通过检测框标记出来，而 YOLOv4s 较全和较好地检测出了图中的物体，并选用了适当的框将物体进行了标记。

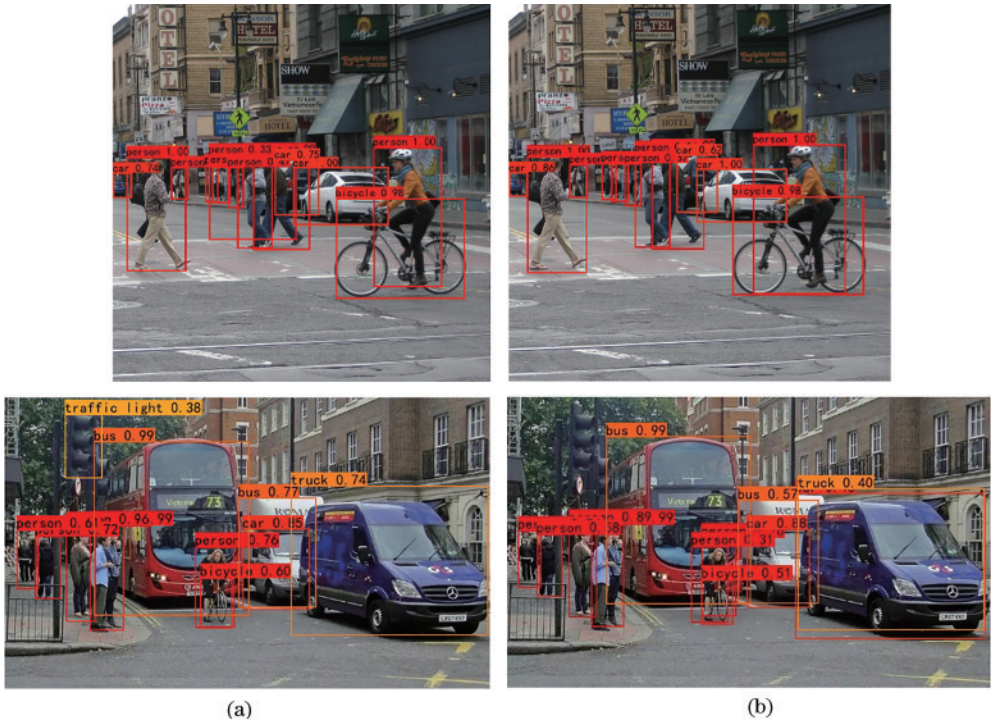

图 7　两种算法检测效果对比。(a) YOLOv4s；(b) YOLOv4-tiny

Fig. 7　Comparison of detection effects between two algorithms. (a) YOLOv4s; (b) YOLOv4-tiny

在 PASCAL VOC、MS COCO 数据集进行了对比实验。从表 3 可以看出：YOLOv4s 在 PASCAL VOC 数据集上比 Faster R-CNN、YOLOv4、YOLOv4-tiny 等 3 种检测算法的 mAP 分别高出 12.2 个百分点、3.9 个百分点、20.9 个百分点，并且召回率也分别提高了 12.9 个百分点、15.6 个百分点、1.7 个百分点；YOLOv4s 的检测速度比 Faster R-CNN、YOLOv4 分别提升了 10 倍和 3 倍，虽然比起 YOLOv4-tiny 速度略有下降，但是仍然能够满足实时检测的要求。表 4 为不同算法在 VOC2012 数据集上每个类别的 AP 对比，表中数据表明，改进后的算法在大多数物体上的检测效果要优于之前算法。

表 3　VOC2012 test 对比

Table 3　VOC2012 test comparison

Method	Backbone	mAP /%	Recall /%	FPS
Faster R-CNN	Resnet50	75.7	78.3	9
YOLOv4	CSPdarknet53	86	89.5	32
YOLOv4-tiny	CSPdarknet53s	67	75.6	106
YOLOv4s	CSPdarknet53-s	87.9	91.2	90

表 5 中，只修改主干网络结构的为 YOLOv4s-1，只修改加强特征提取网络的为 YOLOv4s-2。表 5 数据表明，YOLOv4s-2 效果更加优异，且 YOLOv4s-2 在小目标上的提升要比 YOLOv4s-1 好，说明注意力机制对于小目标的有效特征形成了明显的聚焦，更好抑制了目标周围的其他无关特征。基于 YOLOv4-tiny 改进的 YOLOv4s 检测算法在 MS COCO 数据集上比 YOLOv4-tiny 检测算法的 AP、AP_{50}、AP_{75}、AP_S、AP_M、AP_L 分别高出 6.9 个百分点、9.1 个百分点、11.8 个百分点、3 个百分点、8.6 个百分点、16.4 个百分点，而各项检测精度及 FPS 也要优于轻量型网络 Efficientdet，说明了 YOLOv4s 目标检测算法对于生活场景下的目标检测效果有明显的提升。

表 4 VOC2012 每个类别的 AP 对比

Table 4 Comparison of AP of each category in VOC2012 unit: %

Category	Faster R-CNN	YOLOv4	YOLOv4-tiny	YOLOv4s	Category	Faster R-CNN	YOLOv4	YOLOv4-tiny	YOLO v4s
Aeroplane	82.8	92.8	67.9	92.8	Table	80.4	79.5	74.5	88.6
Bicycle	71.0	80.6	67.5	85.4	Dog	90.4	97.6	85.6	94.3
Bird	90.6	94.7	78.3	93.9	Horse	60.8	51.9	64.9	87.8
Boat	71.5	82.7	54.7	78.5	Mbike	70.0	64.5	64.8	89.1
Bottle	50.0	76.1	31.6	77.4	Person	77.2	81.6	68.2	86.5
Bus	89.8	96.3	90.5	96.8	Plant	56.9	82.8	50.0	72.3
Car	63.3	81.1	53.1	82.5	Sheep	94.7	95.9	58.6	94.8
Cat	94.6	96.7	88.2	97.8	Sofa	79.4	89.1	74.7	92.0
Chair	51.1	67.6	46.4	77.8	Train	83.0	96.8	87.0	97.2
Cow	83.5	95.0	67.2	91.6	Tv	73.4	76.5	66.6	87.1

表 5 MS COCO test 对比

Table 5 MS COCO test comparison unit: %

Method	Backbone	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L	FPS
YOLOv4-tiny	CSPDarknet53s	18.1	37.0	15.4	5.6	20.6	29.3	106
Efficientdet-b0	Efficientnet	23.3	39.1	26.1	8.1	26.5	38.5	22
YOLOv4s-1	CSPDarknet53-s	19.3	38.9	20.2	6.2	22.3	33.5	108
YOLOv4s-2	CSPDarknet53s	22.5	43.9	26.5	8.3	25.6	40.8	89
YOLOv4s	CSPDarknet53-s	25.0	46.1	27.2	8.6	29.2	45.7	90

4 结 论

通过改进 YOLOv4-tiny 的加强特征提取网络得到 YOLOv4s，对主干网络最后两层的输出进行通道加权后再进行信息融合，有效地提升了生活场景下的目标检测的准确度并解决了效率问题，同时多层的信息融合也为目标的检测提供了更多的上下文信息。实验结果表明：在同一实验环境和实验设备上，YOLOv4s 在 PASCAL VOC 和 MS COCO 上的目标检测准确度较其他主流网络均有不同程度的提升；YOLOv4s 的模型参数量仅为 YOLOv4 的 1/6，且在检测速度上比 YOLOv4 快 3 倍，实现了实时检测，最终能在模型计算量与参数量较小的情况下保证高准确率的检测。

参 考 文 献

[1] 徐培，蔡小路，何文伟，等. 基于深度自编码网络的运动目标检测[J]. 计算机应用，2014，34(10)：2934-2937，2962.
Xu P，Cai X L，He W W，et al. Motion detection based on deep auto-encoder networks[J]. Journal of Computer Applications，2014，34(10)：2934-2937，2962.

[2] 李维刚，杨潮，蒋林，等. 基于改进 YOLOv4 算法的室内场景目标检测[J]. 激光与光电子学进展，2022，59(18)：1815003.
Li W G，Yang C，Jiang L，et al. Indoor scene target detection based on improved YOLOv4 algorithm[J]. Laser & Optoelectronics Progress，2022，59(18)：1815003.

[3] Girshick R，Donahue J，Darrell T，et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition，June 23-28，2014，Columbus，OH，USA. New York：IEEE Press，2014：580-587.

[4] Girshick R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision，December 7-13，2015，Santiago，Chile. New York：IEEE Press，2015：1440-1448.

[5] Ren S Q，He K M，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.

[6] Liu W，Anguelov D，Erhan D，et al. SSD：single shot MultiBox detector[M]//Leibe B，Matas J，Sebe N，et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham：Springer，2016，9905：21-37.

[7] Lin T Y，Goyal P，Girshick R，et al. Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2020，42(2)：318-327.

[8] Redmon J，Divvala S，Girshick R，et al. You only look once：unified，real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition，June 27-30，2016，Las Vegas，NV，USA. New York：IEEE Press，2016：779-788.

[9] Redmon J，Farhadi A. YOLO9000：better，faster，stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition，July 21-26，2017，Honolulu，HI，USA. New York：IEEE Press，2017：6517-6525.

[10] Redmon J, Farhadi A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08) [2021-06-08]. https://arxiv.org/abs/1804.02767.

[11] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. (2020-04-23)[2021-06-08]. https://arxiv.org/abs/2004.10934.

[12] Wang C Y, Liao H Y M, Wu Y H, et al. CSPNet: a new backbone that can enhance learning capability of CNN[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), June 14-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 1571-1580.

[13] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.

[14] Xie S N, Girshick R, Dollár P, et al. Aggregated residual transformations for deep neural networks[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 5987-5995.

[15] Huang G, Liu Z, van der Maaten L, et al. Densely connected convolutional networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 2261-2269.

[16] Howard A, Sandler M, Chu G, et al. Searching for MobileNetV3[EB/OL]. (2019-05-06)[2021-08-04]. https://arxiv.org/abs/1905.02244.

[17] Tan M X, Le Q V. EfficientNet: rethinking model scaling for convolutional neural networks[EB/OL]. (2019-05-28)[2021-06-05]. https://arxiv.org/abs/1905.11946.

[18] 孔维刚, 李文婧, 王秋艳, 等. 基于改进 YOLOv4 算法的轻量化网络设计与实现[J]. 计算机工程, 2022, 48(3): 181-188.
Kong W G, Li W J, Wang Q Y, et al. Design and implementation of lightweight network based on improved YOLOv4 algorithm[J]. Computer Engineering, 2022, 48(3): 181-188.

[19] 张欣, 张永强, 何斌, 等. 基于 YOLOv4-tiny 的遥感图像飞机目标检测技术研究[J]. 光学技术, 2021, 47(3): 344-351.
Zhang X, Zhang Y Q, He B, et al. Research on remote sensing image aircraft target detection technology based on YOLOv4-tiny[J]. Optical Technique, 2021, 47(3): 344-351.

[20] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 936-944.

[21] 王凤随, 王启胜, 陈金刚, 等. 基于注意力机制和 Soft-NMS 的改进 Faster R-CNN 目标检测算法[J]. 激光与光电子学进展, 2021, 58(24): 2420001.
Wang F S, Wang Q S, Chen J G, et al. Improved Faster R-CNN target detection algorithm based on attention mechanism and Soft-NMS[J]. Laser & Optoelectronics Progress, 2021, 58(24): 2420001.

[22] Zheng Z H, Wang P, Ren D W, et al. Enhancing geometric factors in model learning and inference for object detection and instance segmentation[EB/OL]. (2020-05-07)[2021-08-05]. https://arxiv.org/abs/2005.03572.

[23] Rezatofighi H, Tsoi N, Gwak J Y, et al. Generalized intersection over union: a metric and a loss for bounding box regression[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 658-666.

[24] Zheng Z H, Wang P, Liu W, et al. Distance-IoU loss: faster and better learning for bounding box regression[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12993-13000.

[25] 李成豪, 张静, 胡莉, 等. 基于多尺度感受野融合的小目标检测算法[J/OL]. 计算机工程与应用: 1-7[2021-05-24]. http://kns.cnki.net/kcms/detail/11.2127.TP.20210420.1358.074.html.
Li C H, Zhang J, Hu L, et al. Small target detection algorithm based on multiscale receptive field fusion [J/OL]. Computer engineering and application: 1-7[2021-05-24]. http://kns.cnki.net/kcms/detail/11.2127.TP.20210420.1358.074.html.