DOI：10.13382/j.jemi.B2205591

基于特征融合的自适应多尺度无锚框目标检测算法*

熊　磊[1,2,3]　王凤随[1,2,3]　钱亚萍[1,2,3]

（1.安徽工程大学电气工程学院　芜湖　241000；2.检测技术与节能装置安徽省重点实验室　芜湖　241000；
3.高端装备先进感知与智能控制教育部重点实验室　芜湖　241000）

摘　要：为了提高 CenterNet 无锚框目标检测网络的目标检测能力，提出一种基于注意力特征融合和多尺度特征提取网络的改进 CenterNet 目标检测网络。首先，为了提升网络对多尺度目标的表达能力，设计了自适应多尺度特征提取网络，利用空洞卷积对特征图进行重采样获取多尺度特征信息，并在空间维度上进行融合；其次，为了更好地融合语义和尺度不一致的特征，提出了一种基于通道局部注意力的特征融合模块，自适应地学习浅层特征和深层特征之间的融合权重，保留不同感受域的关键特征信息。最后，通过在 VOC 2007 测试集上对本文算法进行验证，实验结果表明，最终算法的检测精度达到 80.94%，相较于基线算法 CenterNet 提升了 3.82%，有效提升了无锚框目标检测算法的最终性能。

关键词：目标检测；无锚框；CenterNet；空洞卷积；特征融合；注意力机制

中图分类号：TP391.4　　**文献标识码**：A　　**国家标准学科分类代码**：520.20

Adaptive multi-scale anchor-free target detection algorithm based on feature fusion

Xiong Lei[1,2,3]　Wang Fengsui[1,2,3]　Qian Yaping[1,2,3]

(1. School of Electrical Engineering, Anhui Polytechnic University, Wuhu 241000, China; 2. Anhui Key Laboratory of Detection Technology and Energy Saving Devices, Wuhu 241000, China; 3. Key Laboratory of Advanced Perception and Intelligent Control of High-end Equipment, Ministry of Education, Wuhu 241000, China)

Abstract: In order to improve the target detection ability of CenterNet Ancor-free target detection network, an improved CenterNet target detection network based on attention feature fusion and multi-scale feature extraction network was proposed. Firstly, in order to improve the expression ability of the network for multi-scale targets, an adaptive multi-scale feature extraction network was designed. The feature map is resampled by cavity convolution to obtain multi-scale feature information, and the fusion was carried out on the spatial dimension. Secondly, in order to better integrate semantic and scale inconsistent features, a feature fusion module based on channel local attention was proposed. the fusion weight between shallow features and deep features was adaptively learned, and the key feature information of different perceptual domains was retained. Finally, the algorithm was verified on VOC 2007 test set. The experimental results showed that the detection accuracy of the final algorithm reaches 80.94%, which was 3.82% higher than the baseline algorithm CenterNet, and effectively improves the final performance of the Ancor-free target detection algorithm.

Keywords: target detection; anchor-free; CenterNet; dilated convolution; feature fusion; attention mechanism

0　引　言

目标检测算法要求计算机能够在含有多目标的数字图像中，定位感兴趣的目标并识别其类别，其在自动驾驶、智能监控等领域有着丰富的应用。

提升目标检测算法的性能一直是研究人员所关注的重点问题，当前，卷积神经网络通过更深、更宽、增加基数

收稿日期：2022-06-16　　Received Date: 2022-06-16

*基金项目：安徽省自然科学基金（2108085MF197，1708085MF154）、安徽高校省级自然科学研究重点项目（KJ2019A0162）、检测技术与节能装置安徽省重点实验室开放基金（DTESD2020B02）、安徽工程大学国家自然科学基金预研项目（Xjky2022040）、安徽高校研究生科学研究项目（YJS20210448，YJS20210449）资助

和动态改进特征来提高表示能力,除了这些策略之外,特征融合和多尺度特征提取对于现代网络架构来说可以有效提升网络的性能[1-4]。在无锚框目标检测算法中,ExtremeNet[5]利用关键点预测网络预测 4 个极值点(最顶部、最左侧、最底部、最右侧)和 1 个中心点共 5 个关键点,将 4 个极值点与中心点进行对齐构成一个物体检测框。FCOS[6]对特征图上的每一个像素点进行回归操作,利用特征金字塔网络进行多级预测,使用非极大值抑制后处理过的最终结果。相较于前两种无锚框目标检测算法,CenterNet[7]将目标检测框的中心点表示为目标并直接从中心点去回归目标的其他属性,网络结构更加简单且运行速度更快,但是该算法采用一种端到端的形式去生成目标检测框,未高效地使用不同层级的特征图,在不断下采样的过程中,造成部分目标的特征信息损失,同时缺乏对多尺度问题的解决能力。基于特征融合的方法来提高特征图利用率,基于特征融合的方法来提高特征图利用率,ResNet[8-9]残差网络及其后续改进网络中,身份映射特征和残差学习特征通过短跳跃连接融合为输入,从而可以训练非常深的网络。特征金字塔网络 FPN[10]和 U-Net[11]中,低级特征和高级特征采用长跳跃连接融合以获得高分辨率和语义强的特征。PANet[12]提出一种基于 FPN 的自下而上的路径增强特征层次,以增加深层的低级特征信息,并提出自适应特征池化来聚合所有级别的特征以进行更好的预测。DLA[13]通过引入迭代深度聚合和分层深度聚合结构,以更好地融合语义和空间信息。TridentNet[14]删除了特征金字塔的结构,并创建了具有不同感受域的多个尺度特定分支,以采用尺度感知训练和推理。InceptionNet[15-17]的多个版本中,相同的级别上具有多个大小不同卷积核的输出,被融合以处理对象尺度变化的问题。尽管它在现代网络架构中很适用,但大多数关于特征融合的工作都集中在构建复杂的路径以组合不同内核、组或层中的特征。特征融合的方法很少得到解决,通常通过简单的操作来实现,它们仅提供特征图的固定线性组合。SKNet[18]和 ResNeSt[19]提出基于全局通道注意机制,对同一层中多个内核或组的特征进行动态加权平均。尽管这种基于注意力的方法为特征融合提供了非线性方法,但仍然存在缺点:它们的融合权重是通过全局通道注意力机制生成的,仅仅在全局尺度上聚合上下文信息会削弱小对象的特征。针对多尺度问题的处理,PSPNet[20]提出利用金字塔池化的方法来得到多种感受野的目标特征信息,但是池化操作会使得目标的部分特征信息丢失。ASPP[21]在 SPP[22]的空间金字塔池化的方法上进行改进,使用多个具有不同采样率的并行空洞卷积层来替换池化层,并对每个采样提取的特征图进行拼接融合以生成最终结果,但是这样会忽略了不同尺度特征信息之间的不一致性。

针对目标检测在特征融合和多尺度特征提取中存在的上述问题,本文为了提升无锚框 CenterNet 目标检测网络的性能,对基于 ResNet-50 的无锚框 CenterNet 目标检测网络进行改进。首先,对不同层级的特征进行高效利用,提出一种利用注意力的优化特征融合模块,在对高层特征和低层特征的融合过程中同时考虑到全局和局部两个维度的权重,避免造成在特征融合过程中对小目标特征信息的削弱,提升无锚框网络的性能。其次,在 ResNet-50 下采样模块后嵌入了自适应多尺度特征提取网络,采用多分支空洞卷积层从高层语义特征中获取多尺度特征信息,利用空间重要性权重融合多尺度特征,提高目标检测精度。

1　本文算法

1.1　CentetrNet 算法

CenterNet 目标检测算法模型将目标检测分为中心定位和大小回归两部分,针对中心定位部分,CenterNet 采用高斯核来产生一个热力图,使网络能够在目标中心附近产生更高的激活;针对回归部分,CenterNet 将目标中心的像素定义为一个训练样本,并直接预测目标的高度和宽度;还预测中心点偏移来恢复输出步幅引起的离散误差。本文采用 ResNet50 作为 CenterNet 的主干特征提取网络提取特征图,CenterNet 模型如图 1 所示。

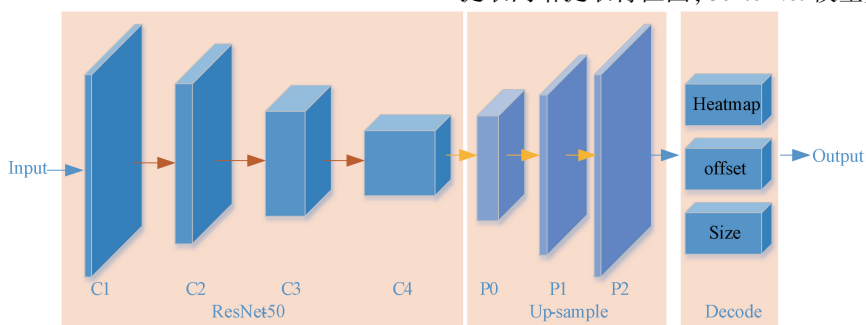

图 1　CentetNet 模型

Fig. 1　CenterNet model

CenterNet 使用 ResNet-50 进行下采样获取图像中目标的高层语义特征，然后将特征输入 3 层反卷积模块进行上采样，得到的特征图输入 Head 模块，获得目标的热力图、中心点偏移以及目标长宽。在上采样模块仅使用下采样模块中的最后一层特征图，未有效利用下采样模块中每一层的特征图，会导致图像中部分目标的特征信息因下采样而造成丢失。该网络同时忽略了对不同尺度目标的处理，造成了网络检测精度降低。

1.2　自适应多尺度特征提取网络

目标检测中，多尺度特征信息的获取可以有效改善目标检测网络的检测性能，本文提出的自适应多尺度特征提取网络（adaptive multi-scale feature extraction network，AMSF）通过学习不同尺度特征图之间的相关性保留有用的信息进行融合。首先，为了捕获多尺度对象和上下文的信息，本文采用多个并行的具有不同扩张率的空洞卷积层获取多尺度特征信息；其次，针对不同尺度的特征信息，自适应地从空间维度对其进行重采样学习空间融合权重；最后，根据获取的空间权重对多尺度特征信息进行融合。

自适应多尺度特征提取网络如图 2 所示。首先，针对 ResNet-50 输出的高层语义特征 $F \in R^{H\times W\times C}$，用 4 个并联的 DC block 对高层语义特征进行重采样获取多尺度特征信息 $F_1 \in R^{H\times W\times C/4}$，$F_2 \in R^{H\times W\times C/4}$，$F_3 \in R^{H\times W\times C/4}$，$F_4 \in R^{H\times W\times C/4}$，DC block 分别由一个 1×1 标准卷积块和一个 3×3 空洞卷积块组成。用式(1)表达：

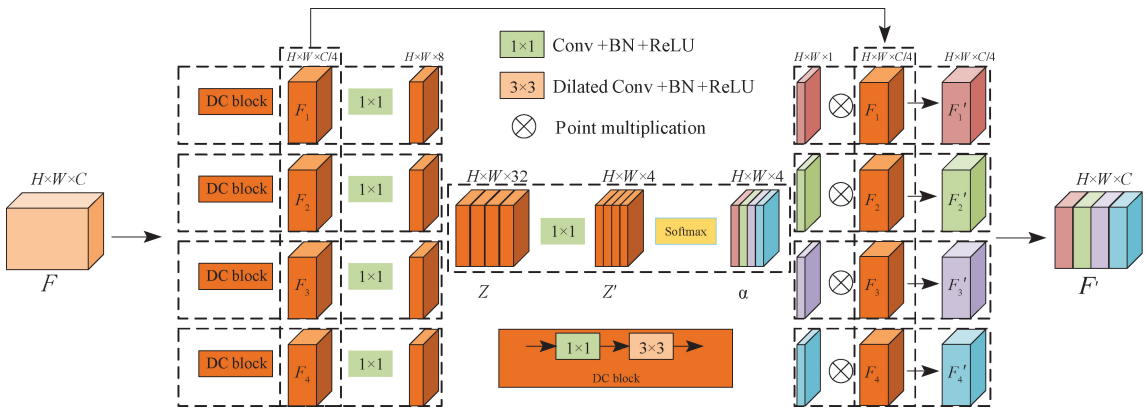

图 2　自适应多尺度特征提取网络

Fig. 2　Adaptive multi-scale feature extraction network (AMSF)

$$F_l = DC_i(F), l = 1,2,3,4 \tag{1}$$

式中：$DC_i(\cdot)$ 表示 DC block，i 表示空洞卷积的扩张率，经过多次实验，本文取 2、4、6、8。

为了自适应的融合多尺度特征图 F_1、F_2、F_3、F_4 并减少计算量，本文先将重采样后的 4 组多尺度特征图利用 1×1 标准卷积块分别对它们进行通道维度上压缩，将压缩后的特征在通道维度上拼接获得特征图 $Z \in R^{H\times W\times 32}$，之后再利用 1 × 1 标准卷积块将特征图 Z 继续压缩通道压缩获得特征图 $Z' \in R^{W\times H\times 4}$，采用分段压缩可以避免因一次压缩造成的特征信息丢失。然后将特征图 Z' 输入 SoftMax 激活函数获得权重 $\alpha \in R^{H\times W\times 4}$，将权重 α 与特征图 F_l 相乘获得新的多尺度特征图 $F'_l \in R^{H\times W\times C/4}$，最后将它们在通道维度上进行拼接获得特征融合之后的特征图 $F' \in R^{H\times W\times C}$。

令 F_{ij}^l 表示特征图 F_l 上位置 (i,j) 处的特征向量。自适应多尺度特征提取网络的公式表达为：

$$F' = cat(\alpha_{ij}^1 F_{ij}^1, \alpha_{ij}^2 F_{ij}^2, \alpha_{ij}^3 F_{ij}^3, \alpha_{ij}^4 F_{ij}^4) \tag{2}$$

其中，$\alpha_{ij}^1, \alpha_{ij}^2, \alpha_{ij}^3, \alpha_{ij}^4$ 表示 4 个不同尺度的特征图的空间重要性权重，它们是由网络自适应学习的。$\alpha_{ij}^1, \alpha_{ij}^2, \alpha_{ij}^3, \alpha_{ij}^4$ 可以是简单的标量变量，它们在针对多尺度特征图的所有通道中共享。受 ASFF[23] 的启发，令 $\alpha_{ij}^1 + \alpha_{ij}^2 + \alpha_{ij}^3 + \alpha_{ij}^4 = 1$ 和 $\alpha_{ij}^1, \alpha_{ij}^2, \alpha_{ij}^3, \alpha_{ij}^4 \in [0,1]$，并且定义为：

$$\alpha_{ij}^l = \frac{e^{\lambda_{\alpha_{ij}}^l}}{e^{\lambda_{\alpha_{ij}}^1} + e^{\lambda_{\alpha_{ij}}^2} + e^{\lambda_{\alpha_{ij}}^3} + e^{\lambda_{\alpha_{ij}}^4}} \tag{3}$$

本文中 α_{ij}^l 是通过使用 Softmax 激活函数来定义的，分别以 $\lambda_{a_{ij}}^1$、$\lambda_{a_{ij}}^2$、$\lambda_{a_{ij}}^3$、$\lambda_{a_{ij}}^4$ 作为控制参数。使用 1×1 卷积层分别从不同尺度特征图中计算权重标量图 α^1、α^2、α^3、α^4，因此可以通过标准反向传播来学习它们。

1.3　通道局部注意力特征融合模块

特征融合是指来自不同层次或分支的特征的重新组合，是现代网络架构中无所不在的组成部分。它通常通过简单的操作来实现，如求和或连接，如图 3 所示，这两种方法对于融合不一致语义和尺度的特征显然不是最佳方法，为了缓解因尺度变化和小物体特征信息引起的不平衡问题，本文提出了通道局部注意力特征融合模块（channel local attention feature fusion module，CLAFF），为

不同尺度的物体聚合来自不同感受域的上下文信息。

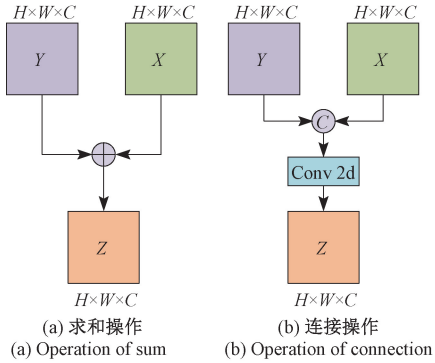

图 3　常规特征融合方法

Fig. 3　Conventional feature fusion method

CLAFF 的结构如图 4 所示，给定两个特征图 $X, Y \in R^{W*H*C}$，令 X 是高层语义特征图，Y 是低层特征图。首先，将 X、Y 选择按元素求和进行初步特征融合获得特征图 $X \oplus Y$；其次，将特征图 $X \oplus Y$ 输入通道局部注意力模块，通过全局平均池化（GAP）获取通道信息，利用一维卷积进行通道间信息交互并用 BN 层进行归一化得到全局上下文信息 $C(X \oplus Y) \in \mathbb{R}^{1\times1\times C}$，为了获取局部特征信息，选择二维卷积作为局部通道上下文聚合器并使用 BN 层进行归一化获取细粒度局部上下文信息 $L(X \oplus Y) \in \mathbb{R}^{H\times W\times C}$，最后将细粒度局部上下文信息整合到全局上下文信息中，利用 Sigmod 激活函数得到融合权重，通道局部注意力模块的公式表达为：

$$W(X \oplus Y) = \delta(L(X \oplus Y) \oplus C(X \oplus Y)) \tag{4}$$

式中：$W(X \oplus Y) \in \mathbb{R}^{H\times W\times C}$ 是全局注意力模块生成的融合权重，由 0 和 1 之间的实数组成；δ 表示 Sigmod 激活函数；$\oplus$ 表示按元素求和。

最后，利用融合权重对特征图 X、Y 分别进行处理，特征图 X 与融合权重 $W(X \oplus Y)$ 按元素相乘，特征图 Y 与融合权重 $1-W(X \oplus Y)$ 按元素相乘，将处理后的特征图 X、Y 按元素求和后进行最终特征融合，获得特征图 $Z \in R^{W\times H\times C}$，图 3 中的虚线表示 $1 - W(X \oplus Y)$。通道局部注意力特征融合模块的公式表示为：

$$Z = W(X \oplus Y) \otimes X \oplus Y \otimes (1 - W(X \oplus Y)) \tag{5}$$

式中：Z 是融合后特征图；$\otimes$ 表示按元素相乘。

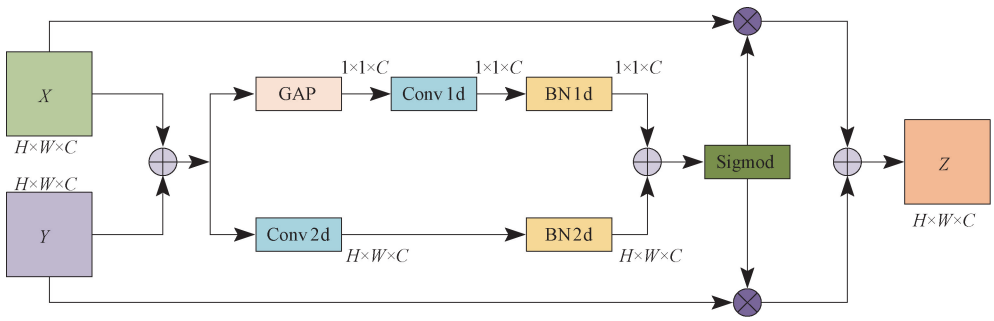

图 4　通道局部注意力特征融合模块

Fig. 4　Channel local attention feature fusion module (CLAFF)

1.4　网络结构设计

针对 CenterNet 无锚框目标检测模型本文主要对其进行如下改进，首先，将本文提出的自适应多尺度特征提取网络 AMSF 置于主干特征提取网络 ResNet-50 中的特征图 C4 之后；其次，将 AMSF 输出的特征图输入上采样模块，反卷积上采样得到特征图 P0 与 ResNet-50 中的特征图 C3 经过通道局部注意力特征融合模块 CLAFF 进行特征融合后进行上采样得到特征图 P1，特征图 P1 与 ResNet-50 中的特征图 C2 经过 CLAFF 进行特征融合后进行上采样得到特征图 P2，特征图 P2 与 ResNet-50 中的特征图 C1 经过 CLAFF 进行特征融合得到最终特征图；最后，将最终特征图输入 Head 模块进行解码获得中心定位和大小回归。改进后的 CenterNet 无锚框目标检测模型如图 5 所示。

2　实验结果与分析

2.1　数据集和评价标准

本文在 PASCAL VOC[24] 数据集上对网络验证，在 VOC 2007 和 VOC 2012 训练集上进行训练，它包含 16 551 张 20 个类别的训练图像和 40 058 个目标；在 VOC 2007 测试集上进行测试，它包含 4 952 张 20 个类别的测试图像和 12 032 个目标。如表 1 所示。

本文从客观评价和主观评价对改进后的 CenterNet 无锚框目标检测方法的性能进行了验证。测试集中每一类别的检测平均精度（average precision，AP）、误检率（log-average miss rate，MR）以及 mAP（mean average precision）作为客观评价的标准；利用改进前和改进后的网络对测试集中的图像进行检测，根据预测的目标检测

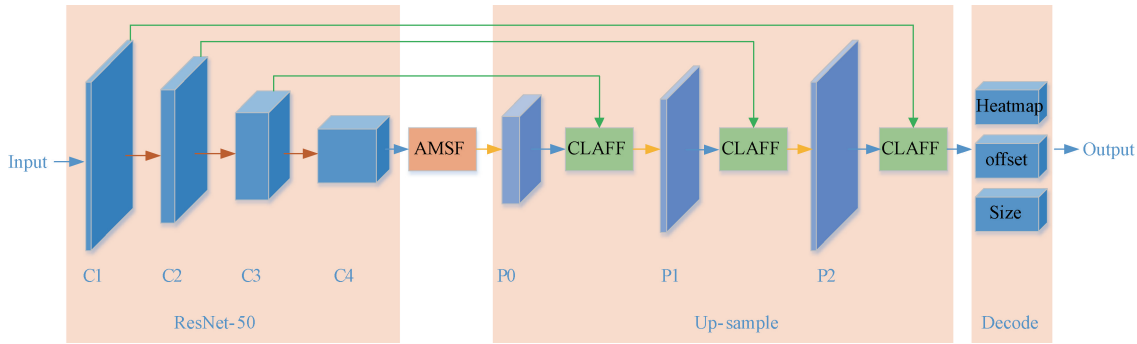

图5　改进后的CenterNet模型

Fig. 5　Improved CenterNet model

框质量作为主观评价的标准。

表1　VOC 2007和VOC 2012在目标检测任务中的训练和测试数据统计

Table 1　Training and test data statistics of VOC 2007 and VOC 2012 in target detection mission

Dataset	Trainval		Test	
	Images	Objects	Images	Objects
VOC2007	5 011	12 608	4 952	12 032
VOC2012	11 540	27 450	0	0
Total	16 551	40 058	4 952	12 032

平均精度(AP)是PR曲线面积,由精度(precision,P)和召回率(recall,R)组成,精度和召回率的计算公式如下所示:

$$P = \frac{TP}{TP + FP} \tag{6}$$

$$R = \frac{TP}{TP + FN} \tag{7}$$

2.2　实验参数设置

实验中所使用的操作系统为Linux,GPU为Nvidia GeForce RTX 3090(24 GB),处理器为英特尔Xeon Gold 6330@2.0 GHz,深度学习框架为Pytorch 1.8.1,训练时间耗费为14 h。本文算法在512×512的输入分辨率上进行训练,使用随机翻转、裁剪和颜色抖动作为数据增强,训练设置为100个epoch,网络采用等间隔调整学习率的策略,使用CenterNet的原始网络权重来对网络进行参数初始化,提高推理和网络收敛速度。前0~50个epoch对网络的主干特征提取网络的参数进行冻结训练,冻结网络训练阶段的Batch_size设置为8,初始学习率为0.001,每个epoch后学习率均调整0.94倍;后50~100个epoch解冻参数后整体训练,解冻网络训练阶段的Batch_size设置为4,初始学习率为0.000 1,每个epoch后学习率均调整0.94倍。实验参数设置如表2所示。

表2　实验参数设置

Table 2　Experimental parameter setting

迭代次数	批次	学习率	权重衰减率
0~50	8	0.001	0.94
50~100	4	0.000 1	0.94

2.3　PASCAL VOC上的定量评价

为了验证本文算法的有效性,与其他主流算法在VOC 2007数据集上的比较,其中分别列举了Faster-RCNN、SSD、YOLO、YOLOv2等经典算法。表3为不同算法在VOC 2007上得到的mAP结果。

表3　不同算法在VOC 2007上的mAP对比

Table 3　mAP comparison of different algorithms on VOC 2007

算法	Backbone	mAP/%	FPS
Faster R-CNN[25]	ResNet-101	76.4	2.4
SSD300[26]	VGG-16	74.3	46
CenterNet*	ResNet-50	77.1	111
CenterNet[7]	ResNet-101	78.7	30
CenterNet[7]	DLA34	80.7	33
YOLO[27]	Darknet	63.4	45
YOLOv2[28]	DarNet53	78.6	40
FCOS[6]	ResNet-50	78.7	–
文献[29]	ResNet-50	75.9	–
文献[30]	VGG-16	78.9	59
Ours	ResNet-50	80.9	88

从表3不同算法测试得到的mAP结果对比可以发现,本文提出的改进无锚框目标检测方法在检测平均精度上有较为明显的优势。在VOC 2007测试集上的检测精度,相较于双阶段目标检测算法,如以ResNet-101作为主干特征提取网络的Faster R-CNN,检测精度提升了4.5%,相较于单阶段的目标检测算法SSD300和YOLOv2,检测精度分别提升了6.6%和2.3%。相较于以ResNet-50、ResNet-101和DLA34作为主干特征提取网络的CenterNet,检测精度分别提升3.8%、2.2%和0.2%。

由于改进后网络相较于基线网络增加了 AMSF 模块和 CLAFF 模块，模型参数量有所增加，因此算法的复杂度稍有一定的增加，相应地，执行速度也有所下降，然而从表 3 不难看出，相较于其他经典的神经网络，提出算法在检测精度、速度等综合性能上具有明显的优势，并且能够满足实时性要求。

为了验证 AMSF 模块的有效性，本文在主干特征提取网络之后分别添加 PSP 模块和 ASPP 模块进行实验，表 4 中展示了不同模块的实验结果。

表 4　获取多尺度特征方法结果对比

Table 4　Comparison of results of multi-scale feature method

方法	PSP[15]	ASPP[16]	AMSF
mAP/%	79.06	79.14	79.56

从表 4 中的 mAP 结果对比可以发现，在多尺度特征信息获取过程中使用空洞卷积可以提高网络的检测性能，本文提出的 AMSF 模块避免因池化造成特征信息丢失，获取了不同感受的特征信息，又考虑到不同尺度特征信息之间的不一致性，AMSF 的性能相较于 PSP 和 ASPP 都有所提升。

从表 5 中的 mAP 结果对比可以发现，在网络中使用特征融合方法可以有效提升网络的表达能力，本文提出的 CLAFF 方法的检测平均精度最高，比 Add 方法高了 0.91%，比 Concate 方法高了 1.87%。

表 5　特征融合方法结果对比

Table 5　Comparison of Feature Fusion Methods

方法	Add	Concate	CLAFF
mAP/%	78.85	77.89	79.76

表 6　在 VOC 2007 测试集上的消融实验结果对比

Table 6　Comparison of ablation experiments on PASCAL VOC2007 dataset

方法	AMSF	CLAFF	mAP/%
CenterNet	×	×	77.12
Experiment 1	√	×	79.56
Experiment 2	×	√	79.84
Experiment 3	√	√	80.94

从表 6 中的 mAP 结果对比可以发现，本文提出的两种方法均能提高无锚框 CenterNet 目标检测方法的性能。实验 1 在网络中仅添加 AMSF 模块，使网络的 mAP 提高了 2.33%；实验 3 在网络中添加 3 次 CLAFF 模块，使网络的 mAP 提高了 2.72%；实验 3 在网络中同时添加 AMSF 模块和 CLAFF 模块，使网络的 mAP 提高了 3.82%。

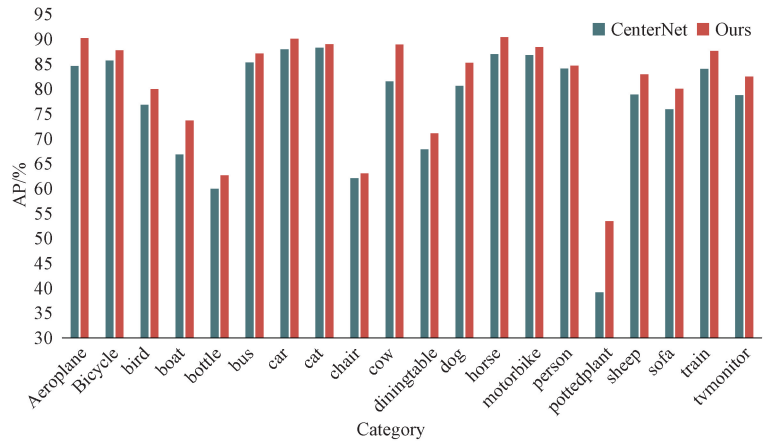

图 6　VOC 数据集上检测结果

Fig. 6　Test results on VOC datasets

从表 7 中的每个类的检测精度数据结果可以得出，当在基线网络中添加自适应多尺度特征提取网络 AMSF 时，实验结果表明网络处理多尺度问题的能力有明显提升，大部分类别的物体的检测精度相较于基线网络都有所提高，例如在 aeroplane（飞机）、bus（公共汽车）、boat（船）这些大物体类别上的 AP 分别有 4.92%、3.89% 和 5.16% 的提升；同时在 cat（猫）、dog（狗）和 pottedplant（花盆）这些小物体类别上的 AP 分别有 1.71%、3.44% 和 5.15% 的提升。本文提出的自适应多尺度特征提取网络，从深层特征中学习多感受野特征信息，通过空间权重自适应的融合不同感受野的特征，使得网络的检测精度有所提升。

当在基线网络中添加通道局部注意力特征融合模块 CLAFF 时，针对网络在下采样卷积操作中造成的特征信

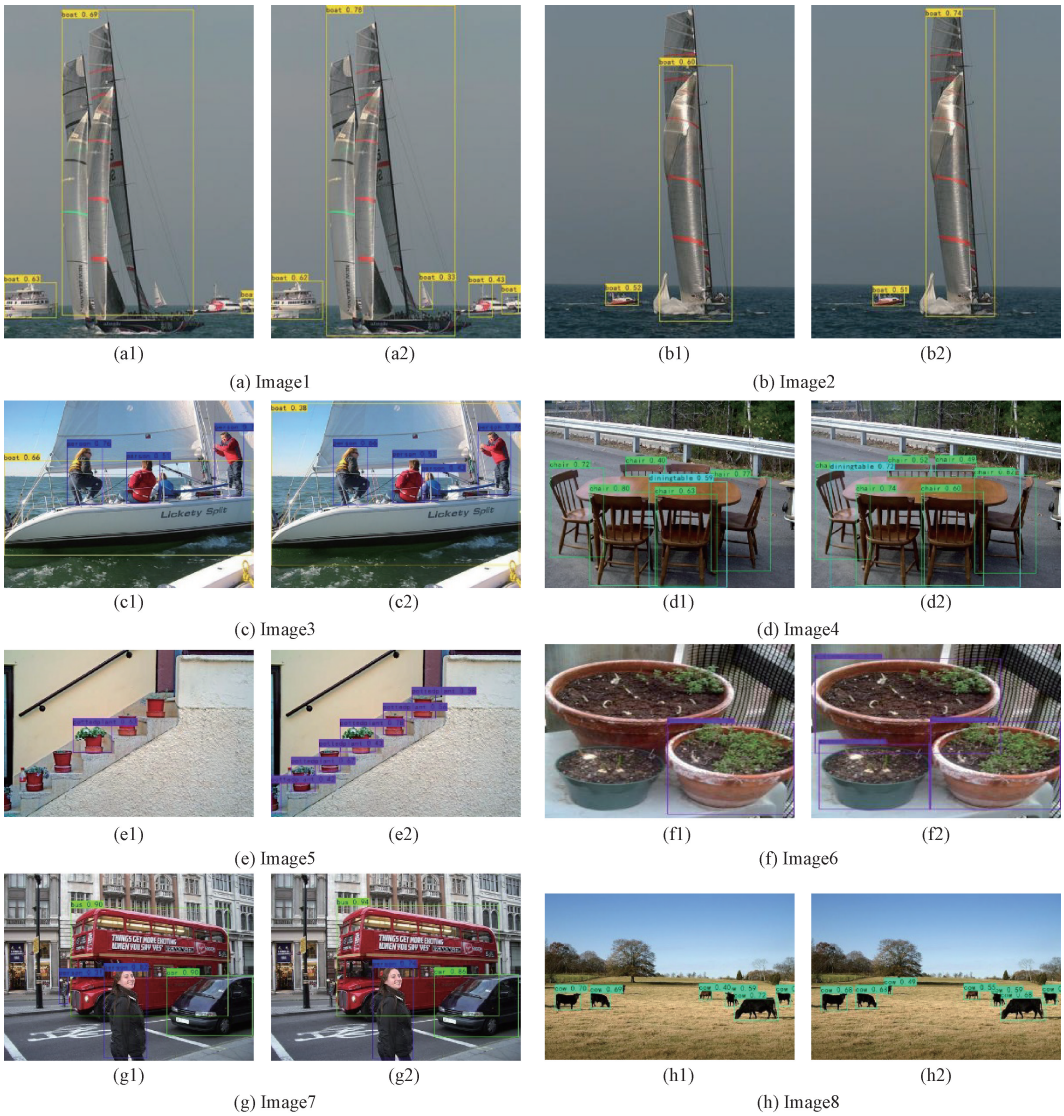

图7　图像检测比较

Fig. 7　Image detection comparison

息丢失,CLAFF利用注意力机制从浅层特征中获取特征信息与深层特征进行有效融合,实验结果表明CLAFF有效提升了网络的表达能力,相较于基线网络的mAP提升了2.72%。其中针对边缘模糊的类别,基线网络对其表达能力最低,例如pottedplant(花盆)的原始AP值只有39.14%,在引入CLAFF后,花盆的AP值相较于基线网络提升了9.34%。对于一些遮挡目标和小目标的AP值均有明显提升,例如cow(牛)、dog(狗)的AP值相较于基线网络分别提升了4.95%、3.61%。

当在基线网络中同时添加通道局部注意力特征融合模块和自适应多尺度特征提取网络时,改进后网络从特征融合和多尺度特征提取两个方面增强了网络的表达能力,有效地提升了不同物体的检测精度,例如pottedplant(花盆)、cow(牛)的AP值分别提升了14.34%、7.43%,最终网络的mAP达到了80.94%。VOC 2007测试集上可视化检测结果如图6所示。

2.4　主观评价结果

图7给出了原始算法和最终改进后算法在VOC 2007测试集不同图像的检测结果对比,图7中(a_1)~(h_1)为CenterNet网络检测结果,(a_2)~(h_2)为最终改进后CenterNet网络检测结果。通过Image1和Image2的检测结果对比可以发现,原始算法不能准确定位目标,最终改进后算法可以有效提高目标的定位精度;通过Image3和Image4的检测结果对比可以发现,当物体出现严重的遮挡现象,原始算法会造成物体的漏检,最终改进后算法不但可以检测出物体还能给出准确的检测框;通过Image5和Image6的检测结果对比可以发现,针对花盆这一边缘模糊的类别,原始算法会有严重的漏检现象,即使

表 7　20 个目标类检测平均精度对比

Table 7　Average precision comparison of 20 target classes

Category	CenterNet	Experiment 1	Experiment 2	Experiment 3
Aeroplane	84. 60	89. 52	89. 02	90. 22
Bicycle	85. 69	86. 91	88. 28	87. 78
bird	76. 83	77. 14	78. 79	79. 96
boat	66. 86	72. 02	71. 15	73. 64
bottle	59. 98	60. 87	62. 74	62. 66
bus	85. 35	89. 24	87. 39	87. 14
car	87. 99	89. 54	89. 50	90. 09
cat	88. 27	89. 98	89. 31	88. 96
chair	62. 08	61. 25	61. 97	63. 04
cow	81. 49	81. 56	86. 44	88. 92
diningtable	67. 90	71. 26	70. 65	71. 09
dog	80. 62	84. 06	84. 23	85. 24
horse	87. 01	90. 39	89. 60	90. 43
motorbike	86. 77	88. 51	86. 91	88. 41
person	84. 11	84. 74	84. 81	84. 66
pottedplant	39. 14	44. 29	48. 48	53. 48
sheep	78. 89	84. 70	84. 17	82. 96
sofa	75. 95	78. 20	77. 75	80. 06
train	84. 04	87. 06	84. 51	87. 65
tvmonitor	78. 78	80. 04	81. 00	82. 46
mAP/%	77. 12	79. 56	79. 84	80. 94

检测出花盆,检测框的定位也不够准确,最终改进后算法不但提高了花盆这一类别的检测精度而且边界框也更加准确;通过 Image7 和 Image8 的检测结果对比可以发现,当图像中出现小物体时,原始算法不仅会对物体误检而且还会漏检,最终改进后算法有效避免了图像中物体的漏检、误检现象。

综上所述,本文提出的最终改进算法在多方面提高了 CenterNet 算法的检测精度,利用特征融合方法有效提高了特征信息的传递,多尺度特征提取网络增强了网络对于多尺度问题的处理能力。

3　结　论

本文提出了一种基于注意力特征融合和多尺度特征提取网络的改进 CenterNet 目标检测算法。首先本文通过在主干特征提取网络后嵌入了提出的多尺度特征提取网络,使高层特征中拥有更丰富的多尺度特征信息,进一步提高了算法对多尺度问题的处理能力。其次对浅层特征和深层特征进行融合,同时在特征融合中采用通道局部注意力机制,利用通道注意力和局部注意力融合语义和尺度不一致的特征图,有效提高了特征的利用率,进而提升了网络的表达能力。最终在 VOC 2007 测试集上本文算法的检测精度有着明显的提升,充分证明了本文所提改进方法的有效性。在下一步的工作中,将主要从优化网络模型结构入手,进一步提高检测精度的同时进行网络轻量化设计。

参考文献

[1] 刘鸣瑄,刘惠义. 基于特征融合 SSD 的远距离车辆检测方法[J]. 国外电子测量技术,2020,39(2):28-32.
LIU M X, LIU H Y. Long-distance vehicle detection method based on feature fusion SSD [J]. Foreign Electronic Measurement Technology, 2020, 39(2): 28-32.

[2] 宋荣,周大可,杨欣. 基于特征融合的尺度感知行人检测[J]. 电子测量技术,2020,43(5):116-123.
SONG R, ZHOU D K, YANG X. Scale-aware pedestrian detection based on feature fusion [J]. Electronic Measurement Technology, 2020,43(5):116-123.

[3] 李晖晖,周康鹏,韩太初. 基于 CReLU 和 FPN 改进的 SSD 舰船目标检测[J]. 仪器仪表学报,2020,41(4):183-190.
LI H H, ZHOU K P, HAN T CH. Ship object detection based on SSD improved with CReLU and FPN [J]. Chinese Journal of Scientific Instrument, 2020,41(4): 183-190.

[4] 谢晓蔚,史健芳. 弱监督卷积神经网络的多目标图像检测研究[J]. 电子测量与仪器学报,2019,33(6):31-37.
XIE X W, SHI J F. Research of convolutional neural networks with weakly-supervised learning on multi-object image detection. [J]. Journal of Electronic Measurement and Instrumentation, 2019, 33(6):31-37.

[5] ZHOU X, ZHUO J, KRAHENBUHL P. Bottom-up object detection by grouping extreme and center points [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 850-859.

[6] TIAN Z, SHEN C, CHEN H, et al. Fcos: Fully convolutional one-stage object detection[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 9627-9636.

[7] ZHOU X Y, WANG D Q, KRAHENBUHL P. Objects as points[J]. arXiv preprint arXiv:1904. 07850, 2019.

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

[9] HE K, ZHANG X, REN S, et al. Identity mappings in deep residual networks [C]. European Conference on Computer Vision. Springer, Cham, 2016: 630-645.

[10] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117-2125.

[11] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C]. International Conference on Medical

Image Computing and Computer-Assisted Intervention. Springer, Cham, 2015: 234-241.
[12] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8759-8768.
[13] YU F, WANG D, SHELHAMER E, et al. Deep layer aggregation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 2403-2412.
[14] LI Y, CHEN Y, WANG N, et al. Scale-aware trident networks for object detection [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 6054-6063.
[15] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 1-9.
[16] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 2818-2826.
[17] SZEGEDY C, IOFFE S, VANHOUCKE V, et al. Inception-v4, inception-resnet and the impact of residual connections on learning [C]. Thirty-First AAAI Conference on Artificial Intelligence, 2017.
[18] LI X, WANG W, HU X, et al. Selective kernel networks[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 510-519.
[19] ZHANG H, WU C, ZHANG Z, et al. Resnest: Split-attention networks [J]. arXiv preprint arXiv: 2004.08955, 2020.
[20] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing network[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2881-2890.
[21] 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, 2017, 40(4): 834-848.
[22] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.
[23] LIU S, HUANG D, WANG Y. Learning spatial fusion for single-shot object detection[J]. arXiv preprint arXiv: 1911.09516, 2019.
[24] EVERINGHAM M, VAN GOOL L, WILLIAMS C K I, et al. The pascal visual object classes (voc) challenge[J]. International Journal of Computer Vision, 2010, 88(2): 303-338.
[25] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
[26] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector [C]. European Conference on Computer Vision. Springer, Cham, 2016: 21-37.
[27] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.
[28] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7263-7271.
[29] 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):405-416.
WANG F S, WANG Q SH, CHEN J G, et al. An improved target detection algorithm for Faster R-CNN based on attention mechanism and Soft-NMS[J]. Laser & Optoelectronics Progress, 2021,58(24):405-416.
[30] 许光宇,尹孟园. 基于空间-通道注意力的改进 SSD 目标检测算法[J]. 光电子·激光,2021,(9):970-978.
XU G Y, YIN M Y. Improved SSD object detection algorithm based on space-channel attention[J]. Journal of Optoelectronics· Laser, 2021, (9):970-978.

作者简介

熊磊,硕士研究生,主要研究方向为计算机视觉、目标检测。

Xiong Lei, M. Sc. candidate. His main research interests include computer vision and target detection.

王凤随(通信作者),博士,副教授,主要研究方向为图像与视频信息处理、视觉计算与智能分析。

Wang Fengsui (Corresponding author), Ph. D., associate professor. His main research interests include image and video information processing, visual computing and intelligent analysis.

钱亚萍,硕士研究生,主要研究方向为计算机视觉、行人重识别。

Qian Yaping, M. Sc. candidate. Her main research interests include computer vision and person re-identification.