基于上下文增强和特征融合的目标检测算法

杨海燕[1,2],王凤随[1,2],张兴旺[1,2]

1. 安徽工程大学 电气工程学院,安徽 芜湖 241000
2. 高端装备先进感知与智能控制教育部重点实验室,安徽 芜湖 241000

摘 要:目的 针对 CenterNet 无锚框目标检测算法表征能力不足的问题,提出一种基于上下文增强和特征融合的学习方法。**方法** 该方法采用多感受野和信息融合的思想,构建自适应上下文提取模块和特征融合策略。首先网络通过自适应上下文提取模块的多路径空洞卷积层获取目标的上下文特征,督促深层网络学习多尺度信息;然后,通过 ACON-C 激活函数在网络中加入非线性因素,对网络神经元自适应地激活,增强网络的数据拟合能力;最后,联合注意力特征融合策略对不同层次的特征信息进行合并,通过整合深层网络的语义信息和浅层网络的位置信息,来捕获对识别任务有用的特征信息,同时学习特征图在多个层次通道间的相关性,以加强网络对关键目标特征的专注度。**结果** 所提方法在 PASCAL VOC 公开数据集上 mAP 达到 83.82%,约比基线算法 CenterNet 增加了 3.72%。相较于经典算法 Faster R-CNN、SSD、YOLOv3 分别增加了 7.4%、9.5%、3.5%。**结论** 有效地提升了 CenterNet 算法的检测性能,并且改进的 CenterNet 相较于其他目标识别算法具有更高的识别准确度,在目标检测应用中具有良好的实用性,充分验证了所提方法的有效性。

关键词:目标检测;空洞卷积;上下文特征;特征融合;注意力机制;ACON 激活函数
中国分类号:TP181 **文献标识码:**A doi:10.16055/j.issn.1672-058X.2025.0003.013

Target Detection Algorithm Based on Context Enhancement and Feature Fusion

YANG Haiyan[1,2], WANG Fengsui[1,2], ZHANG Xingwang[1,2]

1. School of Electrical Engineering, Anhui Polytechnic University, Anhui Wuhu 241000, China
2. Key Laboratory of Advanced Perception and Intelligent Control of High-end Equipment, Ministry of Education, Anhui Wuhu 241000, China

Abstract: Objective Aiming at the issue of inadequate characterization capacity of CenterNet anchor-free object recognition algorithm, a method based on context enhancement and feature fusion was proposed. **Methods** This method adopted the concepts of multi-receptive field and information fusion to construct an adaptive context extraction module and feature fusion strategy. Firstly, the network obtained the contextual features of the target through the multipath dilated convolution of the adaptive context extraction module, prompting deep networks to learn multi-scale information. Then, nonlinear factors were added to the network through the ACON-C activation function, adaptively activating the neurons of

收稿日期:2023-10-10 **修回日期:**2023-12-11 **文章编号:**1672-058X(2025)03-0102-08
基金项目:安徽省自然科学基金项目(2108085MF197);安徽高校省级自然科学研究重点项目(KJ2019A0162);安徽工程大学国家自然科学基金预研项目(XJKY2022040).
作者简介:杨海燕(1999—),女,安徽六安人,硕士研究生,从事计算视觉研究.
通信作者:王凤随(1981—),男,安徽宿州人,教授,博士,硕士生导师,从事计算机视觉、图像与视频信息处理等研究. Email: fswang@ahpu.edu.cn.
引用格式:杨海燕,王凤随,张兴旺. 基于上下文增强和特征融合的目标检测算法[J]. 重庆工商大学学报(自然科学版),2025,42(3):102-109.
YANG Haiyan, WANG Fengsui, ZHANG Xingwang. Target detection algorithm based on context enhancement and feature fusion [J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2025, 42 (3): 102-109.

the network and enhancing the data-fitting ability of the network. Finally, a joint attention feature fusion strategy was used to merge feature information at different levels. By integrating semantic features from the high-level network and positional features from the low-level network, the feature information that is useful for the recognition task was captured. At the same time, the correlation between the feature maps in multiple levels of channels was learned to enhance the network' s focus on key target features. **Results** The proposed method achieved an mAP of 83.82% on the PASCAL VOC public dataset, an improvement of 3.72% compared with the CenterNet baseline algorithm. It also outperformed classic algorithms such as Faster R-CNN, SSD, and YOLOv3 by 7.4%, 9.5%, and 3.5%, respectively. **Conclusion** The proposed method effectively enhances the detection performance of the CenterNet algorithm, and the improved CenterNet has higher recognition accuracy compared with other target recognition algorithms. The proposed method proves to be practical in target detection applications, validating the effectiveness of the proposed approach.

Keywords: target detection; dilated convolution; context feature; feature fusion; attention mechanism; ACON activation function

1　引　言

目标检测的任务是识别出一张图像中多个物体的类别和位置。作为计算视觉领域的一个基本任务,广泛应用于面部分析[1]、人数统计[2-3]、行人检测[4-5]、目标跟踪[6-7]等领域。基于深度学习的目标检测算法按照是否预设锚框可分为有锚框的目标检测算法和无锚框的目标检测算法。无锚框网络通过预测关键点来生成候选框而不使用预设框,相比有锚框网络,网络结构更加简单,检测效率更高,更能满足实际应用场景中对实时检测的需求,但由于自然场景的复杂性以及目标物体的多变性,使得无锚框网络对目标识别能力不强。其主要原因有以下两点:(1)固定深度的网络结构缺乏有效的感受野,没有充分的上下文信息。(2)网络缺乏对不同层级共享特征的学习,造成深层网络表征能力不足。

对于前者,为捕捉感兴趣目标周围有效的上下文信息,现有的研究主要采用池化的方法扩充网络的感受野。如 ParseNet[8]提出一种基于上下文视角的网络架构,通过在深层网络嵌入全局平均池化层,以生成全局上下文信息指导局部信息判断,有效地缓解了局部特征区域上下文信息丢失的问题,同时拓展了深层网络的感知范围,为后续多尺度感受域的研究提供了新的思路。PSPNet[9]在 ParseNet 网络的基础上提出一种更为广泛的、基于不同特征区域的全局上下文集成结构,通过空间金字塔池化(Spatial Pyramid Pooling, SPP)结构来挖掘不同区域下的上下文信息,以加强网络的多尺度特征提取能力,进一步削减了多个次区域之间上下文特征的损失。上述方法分别通过全局平均池化和金字塔池化结构加强网络对上下文特征的关注度,虽缓解了部分无锚框网络感受野受限的情况,但其采用的池化操作会造成细腻特征的丢失,从而导致网络检测能力不强。对于后者,为提高深层特征图的表征能力,现有的研究主要采用信息交互模式使深层特征具有更多的特征信息描述符。如 U-Net[10]提出一种 U 型网络结构,在压缩网络学习框架下,采用跳级连接方式将全局特征融合到扩展层,以提高网络学习不同层级共享特征的能力,有力地提升了模型的目标检测性能,同时为后续多尺度特征融合网络的研究提供了知识储备。FPN[11]提出一种自顶向下的融合架构,利用横向连接的融合体系进行多级特征间的信息传递,以缓解不同深度特征之间的差异,使网络学习共享特征的能力进一步提升。PANet[12]在 FPN 网络的基础上,继续开发自底向上的融合路线,在利用低层级的准确定位信息指导高级信息判断的同时缩短信息路径,进一步减少了深层特征和浅层特征之间的距离。上述方法通过特征融合策略加强深层网络的表征能力,虽缓解了部分无锚框网络的深度特征提取能力不强的问题,但其采用的融合操作无法反映不同维度特征间的语义关联性,不能最大化利用和检测任务相关的特征信息。

针对以上问题,本文从充分利用感兴趣目标的上下文特征和最大化融合有用信息的角度出发,提出一种基于 CenterNet[13]的目标检测改进算法。首先,在骨干网络(ResNet-50)后添加自适应上下文提取模块(Adaptive Context Extraction Module, ACEM),通过使用多路径空洞卷积代替池化层对低分辨率特征进行不同大小感受野的特征提取,在保证扩充多尺度上下文信息的同时减少细腻特征的损失,以加强网络的上下文特征提取能力,并使用 ACON[14]激活函数,自适应地对网络神经元进行非线性和线性之间的参数切换,加强模型的数据拟合能力。其次,提出一种注意力特征融合模块(Adaptive Feature Fusion Module, AFFM),通过学习不同维度特征图通道间的相关性来加强网络对关键信息的专注度,从而最大化融合不同级别特征的有

用信息,提升模型的目标识别能力。最后在 PASCAL VOC[15] 数据集上验证模型的性能。

2 本文算法

2.1 网络总体框架

通过构建 ACEM 和 AFFM、引入 ACON 激活函数等三方面改进原始 CenterNet,以提升目标检测精度。改进后的网络简略结构如图 1 所示,由主干特征提取网络(ResNet-50),自适应上下文提取模块(ACEM),注意力特征融合模块(AFFM)和检测头(Head)四个部分组成。主干特征提取网络对输入图像进行特征学习,输出包含不同层次信息的特征图 P3、P2、P1 和 P0,自适应上下文提取模块(ACEM)对主干特征提取网络输出的高层语义特征图(P0)进行全局上下文特征学习,挖掘深层特征的上下文语义信息,然后通过注意力特征融合模块(AFFM)对不同层次特征图进行信息融合,实现深层网络的语义信息和浅层网络的空间位置信息的整合,最后,该融合特征图传入到检测头(Head)输出检测结果。

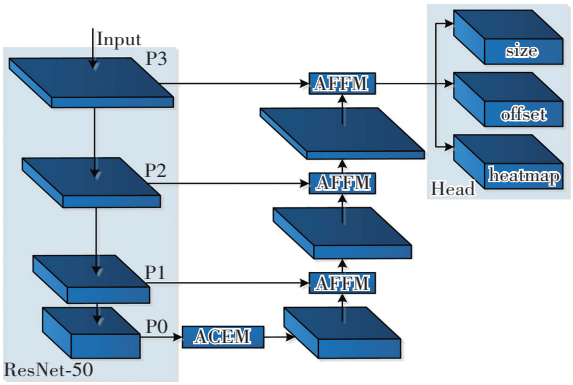

图 1 整体网络结构图

Fig. 1 Overall network structure

2.2 ACON 激活函数的原理

激活函数在神经网络中的作用是引入非线性特性,强化神经网络的学习能力。在很长一段时间 ReLU 激活函数都是最佳的神经网络激活函数,主要是由于其非饱和、稀疏性等优秀的特性,但是它会产生神经元坏死的严重后果。本文使用 ACON 激活函数,明确地学习函数的线性与非线性,给算法提供非线性建模能力。其原理如式(1)所示:

$$f_{\mathrm{ACON}}(\eta_a(x),\eta_b(x))=(\eta_a(x)-\eta_b(x))\times\sigma[\beta\times(\eta_a(x)-\eta_b(x))]+\eta_b(x) \tag{1}$$

其中,$\eta_a(x)$ 和 $\eta_b(x)$ 为输入样本,$\sigma(\cdot)$ 为 sigmoid 函数,β 为平滑因子,通过改变平滑因子 β 的大小可调节其渐进上界和下界的速度。$\eta_a(x)$ 和 $\eta_b(x)$ 在取不同的值时对应不同的激活函数,当 $\eta_a(x)=x,\eta_b(x)=0$ 时,对应 ACON-A 激活函数;当 $\eta_a(x)=x,\eta_b(x)=px$ 时,对应 ACON-B 激活函数;当 $\eta_a(x)=p_1x,\eta_b(x)=p_2x$ 时,对应 ACON-C 激活函数。详细运算过程如式(2)、式(3)、式(4)所示:

$$f_{\mathrm{ACON\text{-}A}}(x)=x\times\sigma(\beta\times x) \tag{2}$$

$$f_{\mathrm{ACON\text{-}B}}(x)=(1-p)x\times\sigma[\beta\times(1-p)x]+px \tag{3}$$

$$f_{\mathrm{ACON\text{-}C}}(x)=(p_1-p_2)\times\sigma[\beta\times(p_1-p_2)x]+p_2x \tag{4}$$

其中,p、p_1、p_2 为待训练参数,用于调节学习的上下界。ACON-C 激活函数通过改变 p_1 和 p_2 两个参数的大小可分别调节学习的上界和下界,并配合平滑因子 β 来调节其渐进上界和下界的速度,具有更好地数据拟合能力,效果优于 ACON-A 和 ACON-B 激活函数。因此,本文采用 ACON-C 激活函数,并应用在自适应上下文提取模块(ACEM)中。

2.3 自适应上下文提取模块 ACEM

具有较大感受野的模型可以获得更多的全局特征信息。利用全局上下文信息,可以提高模型对图像中遮挡物体的检测精度。科研人员在空洞卷积技术诞生前,为了扩大感受野,通常采取下采样,然而这样容易导致大量有用信息的流失,并会使特征图像的分辨率骤减。Fisher Yu 等[16] 在图像分割领域提出了空洞卷积模型,成功地解决了这一问题。扩张卷积在标准卷积中通过补零的方式以扩大卷积核尺寸,在保持特征图像分辨率的同时,有效地融合图像的上下文信息,从而扩大感受野,解决被遮挡物体的问题。然而,大量研究表明,单一尺度的空洞卷积可能导致格化效应[17],当空洞率过高时,由于卷积核内部的权值为 0 的像素过多,卷积变得过于稀疏,难以捕捉关键信息,对于遮挡目标的检测并不有利。因此,设计了一种自适应的上下文提取模块(ACEM),通过合并不同扩张率的卷积层产生的特征图,提升网络获得全局特征的能力,弥补了格化效应的不足。结构如图 2 所示。X 代表特征输入。

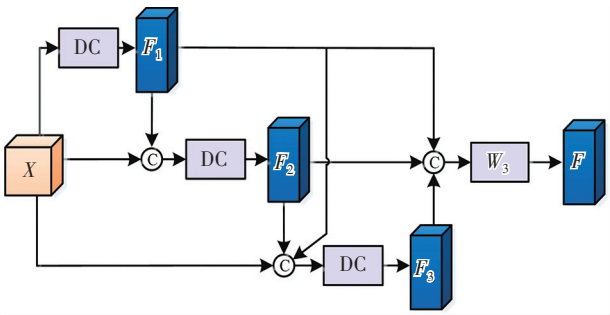

图 2 自适应上下文提取模块 ACEM

Fig. 2 The adaptive context extraction module(ACEM)

首先,利用三层并行的空洞卷积(Dilated Convolution, DC)模块对特征图进行多尺度特征采集,以此获取上下文信息用于遮挡目标检测,DC 模块是由 1×1 卷积和不同速率(r=3,6,12)的空洞卷积组合而成,并且三层并行的 DC 模块采用密集连接方式(即每一层输出特征 F_i 和输入 X 拼接后再传入下一层)进一步扩大感受野,来获得更丰富的多尺度上下文信息。多尺度特征 F_1、F_2、F_3 提取的过程如式(5)、式(6)、式(7)所示:

$$F_1 = D_{r=3}(X) \tag{5}$$

$$F_2 = D_{r=6}(\mathrm{cat}(X, F_1)) \tag{6}$$

$$F_3 = D_{r=12}(\mathrm{cat}(X, F_1, F_2)) \tag{7}$$

其中,$D_{r=i}(\cdot)$ 代表 DC 模块,r 代表的是扩张率,cat(·)表示沿通道维度拼接。接着,将多尺度特征 F_1、F_2、F_3 沿通道维度拼接得到多尺度特征图 cat(F_1,F_2,F_3),最后,通过 1×1 卷积模块融合粗粒度和细粒度特征,详细运算过程如式(8)所示:

$$Z = W_1(\mathrm{cat}(F_1, F_2, F_3)) \tag{8}$$

其中,cat(·)表示沿通道维度拼接,$W_1(\cdot)$ 代表 1×1 卷积运算。

2.4　注意力特征融合模块 AFFM

经 CenterNet 主干特征提取网络(ResNet-50)处理后可得到 4 种分辨率的特征。底层特征含有大量的位置以及其他细节信息,但其语义性较差,而顶层特征更具语义性,但对于目标细节的感知能力较低。如果仅使用深层维度的高级特征,则会丢失低级信息。为了整合不同维度的信息,需要对深层特征和浅层特征进行信息融合,而该方法存在不同层次信息间相互干扰的问题。为此,设计基于注意力机制的特征融合模块(AFFM),对不同层次的信息进行加权融合,从而对有用信息进行合并。

结构如图 3 所示,高阶语义特征图和低阶语义特征图作为输入,标记为 X 与 Y。首先,通过在通道方向上将特征 X 和 Y 连接得到特征图 cat(X,Y)。其次,特征图 cat(X,Y)传入 1×1 卷积和 3×3 卷积实施通道的压缩并整合跨通道信息获得特征图 $W_1(\mathrm{cat}(X,Y))$,为了避免梯度消失干扰检测效果,将结果与原始特征融合,输出一个初步的合成特征图 X_1。具体计算如式(9)、式(10)所示:

$$W_1(\mathrm{cat}(X,Y)) = W_{3\times3}(W_2(\mathrm{cat}(X,Y))) \tag{9}$$

$$X_1 = Y \oplus W_1(\mathrm{cat}(X,Y)) \tag{10}$$

其中,$W_2(\cdot)$ 代表 1×1 卷积运算,$W_{3\times3}(\cdot)$ 代表 3×3 卷积运算,cat(·)表示在通道维度上拼接,⊕表示按元素加合。

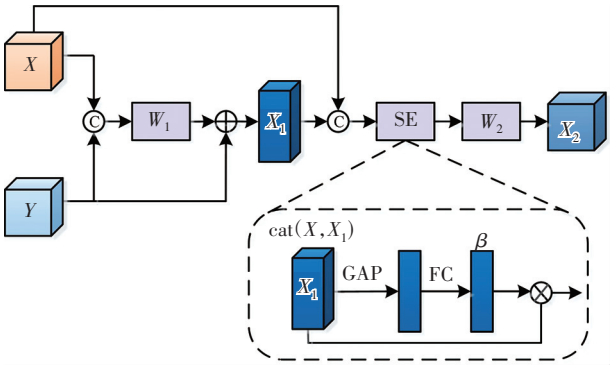

图 3　注意力特征融合模块 AFFM

Fig. 3　The attention feature fusion module(AFFM)

再次,沿通道层次上将特征 X_1 和 X 连接得到特征图 cat(X_1,X)。然后,特征图 cat(X_1,X)在压缩激励网络(Squeeze-and-Excitation networks, SE)中自动适应地融合有用的信息,SE 是通过全局平均池化技术来获得通道层次的信息,并利用全连接层(Fully Connected, FC)实现通道的数据交互,从而得到融合权重 β,再将融合权重 β 和特征图 cat(X_1,X)按元素相乘,输出加权融合的特性图 cat(X_1,X)⊙β。最后,特征图 cat(X_1,X)⊙β 输至 1×1 卷积,实现通道的压缩,从而输出特征图 X_2。具体计算如式(11)、式(12)所示:

$$\beta = \mathrm{FC}(\mathrm{GAP}(\mathrm{cat}(X_1, X))) \tag{11}$$

$$X_2 = W_2(\mathrm{cat}(X_1, X) \odot \beta) \tag{12}$$

其中,$W_2(\cdot)$ 代表 1×1 卷积运算,⊙表示按元素相乘,cat(·)表示沿通道维度拼接。

3　仿真实验与结果分析

3.1　实验设置

本文实验环境:CPU 为 4 核 Intel(R) Xeon(R) Silver 4110 CPU @ 2.10GHz,内存 16G,显卡 Nvidia GeForce RTX 2080Ti(11GB),操作系统为 Ubuntu 16.04,深度学习框架为 pytorch 1.2.0,CUDA 版本为 10.0。输入图像大小设置为 512×512。优化器采用 Adam,设定的起始学习率是 0.0005,权重衰减设置为 0,训练次数设置为 100 轮。

3.2　数据集和评价标准

本文在 PASCAL VOC 数据集上验证模型,PASCAL VOC 的数据集包含 20 类的物体,由 VOC2007 和 VOC2012 联合训练集构成,共 16 551 张图像和 40 058 个目标作为训练样本,4 952 张图像和 12 032 个目标作为测试样本,平均每个图像有 2.4 个目标。

在目标识别任务里,精确率(Precision, P)与召回率(Recall, R)是两个最常被应用的评估标准。

精确率(P):被分类器识别为正样本并且实际是正样本在被分类器识别为正样本中的占比,运算公式如式(13)所示:

$$P=\frac{N_{TP}}{N_{TP}+N_{FP}} \tag{13}$$

其中,N_{TP} 为正样本且预测结果是正类的数量,指的是被正确分类的正类。N_{FP} 为负样本但预测结果是正样本的数量,代表的是被错误分类的负类。

召回率(R):被分类器识别为正样本并且实际是正样本在所有实际是正样本中的占比,运算公式如式(14)所示。

$$R=\frac{N_{TP}}{N_{TP}+N_{FN}} \tag{14}$$

其中,N_{TP} 与式(13)中的含义相同,N_{FN} 为正样本但检测结果是负样本的数量,代表的是被错误分类的正类。平均精度(Average Precision, AP)为 PR 曲线面积。

目标检测模型通常会用精度(mean Average Precision, mAP)作为评价指标。mAP 是将所有种类的 AP 取均值后的结果。

3.3 消融实验

本文中,在相同的实验设定以及 VOC 数据集下进行消融实验,来评估所提方法的有效性。

如表 1 所示,ACEM 表示对主干特征提取网络(ResNet-50)输出特征进行上文信息提取 ACON-C 表示在 ACEM 中加入非线性因素 AFFM 表示对不同维度的特征进行信息融合 Experiment 1 表示在基线(CenterNet)上添加 ACEM,由实验结果可知,mAP 得到了 2.15%的提升,说明 ACEM 有加强网络特征提取能力的作用,能有效聚合上下文信息;Experiment 2 表示在 Experiment 1 的基础上对特征进行非线性处理,与 Experiment 1 方法相比,mAP 提升了 0.9%,说明 ACON-C 激活函数有提升网络数据拟合性能的作用;Experiment 3 表示在基线(CenterNet)上添加 AFFM,与基线算法相比,mAP 得到了 1.42%的提升;Experiment 4 表示在 Experiment 1 的基础上添加 AFFM,与 Experiment 1 方法相比,mAP 提升了 0.37%,说明 AFFM 有丰富和细化特征的作用,能提取对识别任务有用的关键特征;Experiment 5 表示在 Experiment 2 的基础上再对不同粒度的特征加权融合,与 Experiment 2 方法相比,mAP 增加了 0.67%,使目标检测性能得到进一步提升。

从表 2 中的每个类的检测精度数据去分析可以得到以下结论:大量密集的实例在空间定位上产生相互覆盖的状况会增加检测的复杂度。当在基线网络中添加 ACEM 后,并使用 ACON-C 激活函数对特征进行非线性处理,改进后的算法在复杂环境下对物体的检测效果要优于之前的算法。例如在 Bird(鸟)、Cow(牛)、Sheep(羊)这些易于成群出现的复杂场景中的目标 AP 分别有 3.71%、3.5%、5.46%的提升,在 Dining table(餐桌)、Potted plant(花盆)这些边缘模糊易出现遮挡的目标 AP 分别有了 5.72%、8.47%的提升。自适应上下文提取模块(ACEM)旨在借助被测物体周围的有效上下文信息,在不同的感受域中捕获上下文信息,挖掘标签中相互依赖的语义信息,从而提升检测性能;对于网络在进行下采样卷积操作时造成的目标位置信息的丢失,在 Experiment 2 的基础上加入注意力特征融合模块(AFFM)后,模型的位置定位能力得到了改善,且改进后的算法在检测多种类别的性能上都有所提升。例如在 Bird(鸟)、Bottle(杯子)这些类别 AP 分别又有了 1.80%、3.12%的提升,特征融合网络旨在将高层语义和低层位置信息相融合,弥补下采样操作过程中丢失的位置信息,从而提升模型的表征能力;在基线模型中同时引入自适应上下文提取模块(ACEM)、ACON-C 激活函数和注意力特征融合模块(AFFM),可以有效提升网络对多个类别的检测性能。例如,像 Dining table(餐桌)和 Potted plant(花盆)等物体的 AP 分别提升了 6.37%和 7.41%,使改良后的模型的 mAP 提高至 83.82%。

表 1 在 VOC 2007 测试集上的消融实验结果

Table 1 Ablation experiment results on VOC 2007 dataset

Method	ACEM	ACON-C	AFFM	mAP/%
CenterNet	×	×	×	80.10
Experiment 1	✓	×	×	82.25
Experiment 2	✓	✓	×	83.15
Experiment 3			✓	81.52
Experiment 4	✓		✓	82.62
Experiment 5	✓	✓	✓	**83.82**

表 2 不同类别的检测精度对比

Table 2 Comparison of detection accuracy of different categories

/%

Category	CenterNet	Experiment 1	Experiment 2	Experiment 3	Experiment 4	Experiment 5
Aeroplane	88.28	89.83	90.26	91.87	92.82	92.69
Bicycle	87.61	89.53	89.59	89.41	87.96	90.03
Bird	79.00	81.90	82.71	81.03	82.86	84.51
Boat	70.63	73.57	74.49	73.43	74.60	76.12

续表(表2)

Category	CenterNet	Experiment 1	Experiment 2	Experiment 3	Experiment 4	Experiment 5
Bottle	63.91	65.03	65.97	66.90	68.43	69.09
Bus	87.00	89.69	90.78	87.27	87.77	90.74
Car	90.03	91.03	91.13	90.93	91.66	91.41
Cat	90.20	91.57	91.36	89.08	91.30	91.94
Chair	64.18	64.50	65.92	64.79	66.43	67.58
Cow	86.43	90.31	89.93	85.69	89.10	88.29
Dining table	71.75	72.33	77.47	75.62	72.39	78.12
Dog	85.18	87.39	88.49	85.53	87.09	89.41
Horse	89.61	90.24	91.33	90.21	88.53	89.63
Motorbike	87.71	89.67	89.76	88.26	88.68	90.79
Person	85.15	85.49	85.83	86.23	85.96	86.77
Potted plant	48.33	54.88	56.80	52.92	56.61	55.74
Sheep	82.60	87.10	88.06	84.72	87.38	87.43
Sofa	77.09	79.81	80.09	78.32	80.59	81.07
Train	86.08	87.22	88.39	84.99	88.22	89.59
TV monitor	81.33	83.93	84.56	83.21	83.99	85.42
mAP	80.10	82.25	83.15	81.52	82.62	83.82

3.4　与其他方法进行比较

为了验证本文算法的有效性,对比实验基于公共数据集 PASCAL VOC,将所提算法与现有的目标检测先进算法进行比较,主要包括 Faster R-CNN[18]、CoupleNet[19]、SSD[20]、RefineDet[21]、RFBNet512[22]、CenterNet[13]、YOLOv3[23]、FCOS[24]、文献[25]、文献[26]、文献[27]等。

由表 3 可知,本文提出的基于上下文增强和特征融合的目标检测算法与基线(CenterNet)算法相比,在 VOC2007 测试集上 mAP 增加了 3.7%。由于 CenterNet 算法采用的网络结构表征能力不足,缺乏有效的上下文信息和目标空间位置信息,导致遮挡场景中目标识别准确率低,本文通过构建 ACEM 和 AFFM、引入 ACON 激活函数等三方面改进原始 CenterNet 网络,挖掘深层特征的上下文语义信息,并将深层特征的上下文语义信息和多维度的空间位置信息进行加权融合,加强网络对有用信息的关注度,从而提升目标检测性能。本文算法与 Faster R-CNN 和 CoupleNet 算法相比,mAP 分别增加了 7.4% 和 1.1%,Faster R-CNN 和 CoupleNet 算法利用单一尺度特征生成基于候选区的局部特征,缺乏本文对目标全局上下文信息的关注;与 SSD 算法相比,mAP 增加了 9.5%,SSD 算法使用不同分辨率的特征图去预测目标,缺乏本文对目标空间位置信息的关注;与 FCOS 算法相比,mAP 增加了 5.1%,FCOS 算法利用特征金字塔对不同层次特征进行融合,没有考虑到不同层次特征间的关联性,本文对不同特征图通道间的关联性进行建模,实现不同分辨率特征的充分融合;与在同一基线(CenterNet)算法上改进的文献[26]相比,mAP 增加了 1.5%,文献[26]构建轻量级的信息融合网络减少了网络参数,但丢失了目标的细腻特征,并且没有关注到不同层次特征间的相关性,本文算法使用注意力特征融合模块(AFFM)更能处理图像的细节,又通过学习不同层次特征间的语义关联性对多层特征加权融合,进一步加强了网络对细节信息的捕获能力。通过实验结果表明本文算法对目标多尺度上下文特征和空间位置特征提取效果更好,更能合并对目标识别有利的信息,更能应对复杂环境下的目标检测任务,进一步证明了所提算法的有效性。

表 3　在 VOC2007 测试集和其他方法对比实验结果

Table 3　Comparison of experimental results on the VOC 2007 dataset

Method	Backbone	mAP/%
Faster R-CNN[18]	ResNet-101	76.4
CoupleNet[19]	Reaidual-101	82.7
SSD[20]	VGG-16	74.3
RefineDet[21]	VGG-16	81.8
RFBNet512[22]	VGG-16	82.2
CenterNet[13]	ResNet-50	80.1
CenterNet[13]	ResNet-101	78.7
CenterNet[13]	DLA34	80.7

续表(表3)

Method	Backbone	mAP/%
YOLOv3[23]	DarNet53	80.3
FCOS[24]	ResNet-50	78.7
文献[25]	CBResNet101	83.2
文献[26]	Res101-FcaNet	82.3
文献[27]	ResNet-50	74.7
Ours	**ResNet-50**	**83.8**

3.5 主观评价结果

为了更直观地评价本文算法,将基线模型和改进后的模型分别在 VOC2007 测试集上测试。识别结果如图 4 所示。第一行图像为基线算法检测结果,第二行图像为改进算法的检测结果。从图中对比结果可以看出,在原始模型中,一些目标容易出现误检和漏检。但经过改进后,这些问题得到了缓解。从第一组羊群对比图和第四组花盆对比图可以看出,在密集场景下(图中被测物体之间相互遮挡),基线模型检测性能不佳,出现了漏检的现象,而本文算法采用了自适应上下文提取模块(ACEM)对基线模型进行改进后,借助被测物体的上下文信息,有效地解决了漏检的问题。从第二组狗的识别结果对比可以看出,基线模型将狗误判为背景,不能识别出被测物体,而本文算法通过采用 ACON-C 激活函数对特征进行非线性处理,具有良好的数据拟合能力,对于正样本表现为非线性拟合,重组后的特征加强了正样本的表征能力,而在负样本中变现为线性拟合,重组后的特征减弱了背景特征的干扰,解决了将物体误判为背景的问题。从第三组车辆的识别结果对比可以看出,原算法对车的定位缺乏精确性,本文提出的算法,通过整合注意力特征融合模块,将高级语义信息和底层位置信息混合在一起,大幅提升了车辆的定位精准度,同时也提升了置信度的评分。从第五组自行车和杯子的识别结果对比可以看出,原算法出现误检和定位不准的问题,而本文不仅减少了误检,对自行车和杯子的定位也更准确。

Baseline

Ours

图 4 图像识别结果比较

Fig. 4 Comparison of image recognition results

4 结 论

本文提出了一种基于上下文增强和特征融合的目标检测算法。针对 CenterNet 无锚框目标检测算法的特征提取能力不佳的问题,提出了自适应上下文提取模块和特征融合模块。前者通过提取深层网络多尺度特征,学习目标周围的上下文信息,增强了网络对遮挡目标特征的表达能力,并联合 ACON 激活函数,加强了网络对复杂数据的拟合性能。后者对深层网络的上下文语义信息和多维度的空间位置信息进行加权融合,提高了网络对关键信息的捕获能力,最终通过在公开数据集上的实验结果表明了所提方法的有效性。未来的工作将考虑适当做一些轻量化处理来提高检测速度。

参考文献(References):

[1] YANG S, LUO P, LOY C C, et al. WIDER FACE: A face detection benchmark[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 5525-55.

[2] SHAMI M B, MAQBOOL S, SAJID H, et al. People counting in dense crowd images using sparse head detections[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2019, 29(9): 2627-2636.

[3] STEWART R, ANDRILUKA M, NG A Y. End-to-end people detection in crowded scenes[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 2325-.

[4] WANG S, CHENG J, LIU H, et al. Pedestrian detection *via* body part semantic and contextual information with DNN[J]. IEEE Transactions on Multimedia, 2018, 20(11): 3148-3159.

[5] LI J, LIANG X, SHEN S, et al. Scale-aware fast R-CNN for pedestrian detection[J]. IEEE Transactions on Multimedia, 2018, 20(4): 985-996.

[6] CHEN K, TAO W. Learning linear regression *via* single-convolutional layer for visual object tracking[J]. IEEE Transactions on Multimedia, 2019, 21(1): 86-97.

[7] HU H, MA B, SHEN J, et al. Robust object tracking using manifold regularized convolutional neural networks[J]. IEEE Transactions on Multimedia, 2019, 21(2): 510-521.

[8] WEI L, ANDREW R, ALEXANDER C. et al. Parsenet: Looking wider to see better[C]//Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2015: 212-218.

[9] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing network

[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 6230-6239.

[10] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation[M]// Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015. Cham: Springer International Publishing, 2015: 234-241.

[11] 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. Piscataway: IEEE Press, 2017: 936-944.

[12] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 8759-8768.

[13] ZHOU X Y, WANG D Q, KRAHENBUHL P. Objects as points[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2019: 7263-7271.

[14] MA N, ZHANG X, LIU M, et al. Activate or not: Learning customized activation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2021: 8028-8038.

[15] 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.

[16] YU F, KOLTUN V, FUNKHOUSER T. Dilated residual networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 636-644.

[17] WANG P, CHEN P, YUAN Y, et al. Understanding convolution for semantic segmentation[C]//Proceedings of the IEEE Winter Conference on Applications of Computer Vision. Piscataway: IEEE Press, 2018: 1451-1460.

[18] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[C]// Proceedings of the IEEE Transactions on Pattern Analysis and Machine Intelligence. Piscataway: IEEE Press, 2017: 1137-1149.

[19] ZHU Y, ZHAO C, WANG J, et al. CoupleNet: coupling global structure with local parts for object detection[C]// Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2017: 4146-4154.

[20] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot MultiBox detector[M]//Computer Vision-ECCV 2016. Cham: Springer International Publishing, 2016: 21-37.

[21] ZHANG S, WEN L, LEI Z, et al. RefineDet: Single-shot refinement neural network for object detection[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31(2): 674-687.

[22] LIU S, HUANG D, WANG Y. Receptive field block net for accurate and fast object detection[M]//Computer Vision-ECCV 2018. Cham: Springer International Publishing, 2018: 404-419.

[23] REDMON J, FARHADI A. Yolov3: An incremental improvement [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2018: 3523-3541.

[24] 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. IEEE, 2019: 9627-9636.

[25] 戴坤, 许立波, 黄世旸, 等. 融合策略优选和双注意力的单阶段目标检测[J]. 中国图象图形学报, 20, 27(8): 2430-244.
DAI Kun, XU Libo, HUANG Shiyang, et al. Single stage object detection algorithm based on fusing strategy optimization selection and dual attention mechanism[J]. Journal of Image and Graphics, 20, 27(8): 2430-2440.

[26] 侯志强, 郭浩, 马素刚, 等. 基于双分支特征融合的无锚框目标检测算法[J]. 电子与信息学报, 20, 44(6): 2175-218.
HOU Zhiqiang, GUO Hao, MA Sugang, et al. Anchor-free object detection algorithm based on double branch feature fusion[J]. Journal of Electronics & Information Technology, 20, 44(6): 2175-2185.

[27] 王启胜, 王凤随, 陈金刚, 等. 融合自适应注意力机制的 Faster R-CNN 目标检测算法[J]. 激光与光电子学进展, 20, 59(12): 1215016.
WANG Qisheng, WANG Fengsui, CHEN Jingang, et al. Faster R-CNN target-detection algorithm fused with adaptive attention mechanism[J]. Laser & Optoelectronics Progress, 20, 59(12): 1215016.

责任编辑:陈　芳