计算机科学
COMPUTER SCIENCE

改进型FCOS目标检测算法

陈金令，程茂凯，徐紫涵

引用本文

陈金令，程茂凯，徐紫涵.改进型FCOS目标检测算法[J]．计算机科学, 2022, 49(11A): 210900220-6.

CHEN Jin-ling, CHENG Mao-kai, XU Zi-han. Improved FCOS Target Detection Algorithm[J]. Computer Science, 2022, 49(11A): 210900220-6.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.210900220

改进型 FCOS 目标检测算法

陈金令　程茂凯　徐紫涵
西南石油大学电气信息学院　成都 610500

摘　要　针对经典无锚框目标检测算法 FCOS(Fully Constitutional One-Stage Object Detection)难以充分提取目标特征,位置与内容信息结合能力不足,正负样本区分不充分导致性能减弱等问题,提出了一种改进型 FCOS 目标检测算法。该方法首先在 ResNet50 特征提取网络中加入可变形卷积模块与全局注意力模块,提高特征信息捕获能力;然后,将 FPN 特征金字塔与深层链路层相结合,构成多尺度特征融合模块,提升特征提取效果。最后,加入自适应划分正负样本模块,增强检验框的准确性以达到提高回归精度的效果,从而提升检测结果。为了测试算法的检测效果,分别使用了 COCO 数据集与 VOC 数据集进行实验。与原 FCOS 算法相比,所提算法在两个数据集上的平均精度分别提高了 2.3%和 1.8%,其中,对 COCO 数据集中的小目标检测的效果有明显提升。

关键词:目标检测;可变形卷积;全局注意力;多尺度特征;特征金字塔;正负样本
中图法分类号　TP391.41

Improved FCOS Target Detection Algorithm

CHEN Jin-ling,CHENG Mao-kai and XU Zi-han
School of Electrical Information,Southwest Petroleum University,Chengdu 610500,China

Abstract　An enhanced FCOS object detection algorithm is proposed to address the problems that the classical anchorless frame object detection algorithm FCOS(fully constitutional one-stage object detection) has difficulty in extracting target information,insufficient ability to combine location and content information,and weak performance due to insufficient differentiation between positive and negative sample. The method first adds a deformable convolution module and a global attention module to the ResNet50 feature extraction network to improve the feature information capture capability. Then,the FPN feature pyramid is combined with the deep link layer to form a multi-scale feature fusion module to improve the feature extraction effect. Finally,the adaptive division of positive and negative samples module is added to enhance the accuracy of the test frame to achieve the effect of improving the regression accuracy. In order to test the detection effect of the algorithm,the COCO dataset and VOC dataset are used for experiments. Compared with the original FCOS algorithm,the average accuracy of the proposed algorithm on the two datasets improves by 2.3% and 1.8%,respectively. Among them,there is a significant improvement for the detection of small targets in the COCO dataset.

Keywords　Target detection,Deformable convolution,Global attention,Multi-scale features,Feature pyramid,Positive and negative samples

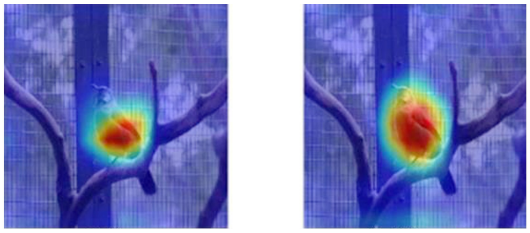
1　引言

在机器视觉中,目标检测是一项基本且富有挑战性的任务,它要求算法为图像中每一个“感兴趣”的实例预测一个带有类别标签的边界框。当前主流的算法可以分为两大类:One-stage 检测算法和 Two-stage 检测算法。One-stage 检测算法是将可能的边界框(称为锚)的复杂排列滑动到图像上,并直接对其进行分类,无须指定框的内容。最经典的算法有 YOLOv2[1],YOLOv3[2]等。Two-stage 检测算法是重新计算每个潜在框中的图像特征,再进行特征分类,最后进行后处理。这种后处理的缺点是很难区分前景和背景,同时训练过程耗时过长。经典的算法如 SSD[3],RetinaNet[4],Faster R-cnn[5]等。这些使用 Anchor 机制的算法统称为 Anchor-Base 算法。但是 Anchor-Base 算法存在着泛化能力不足等问题,例如形状变化较大的候选对象(尤其是小对象)会存在一定的困难。后来,可以与 Anchor-Base 算法相媲美的 Anchor-Free 算法应运而生。这类算法去除了 Anchor-Base 算法中复杂的先验框,使用关键点检测的方法替代,获得了较高的召回率。代表性的算法有 FCOS[6],CenterNet[7],CornerNet[8],RepPoints[9]等。

现有的 Anchor-Free 算法中,FCOS 是一种基于中心点的经典网络,其继承了 FCN[10]算法中的全卷积和特征金字塔(Feature Pyramid Networks,FPN[11])。然而,FCOS 虽然使用全卷积获取了更多的空间位置信息,但是忽略了整个网络的上下文信息交流不充分对特征特取和网络优化的重要性。FCOS 中的特征金字塔(Feature Pyramid Networks,FPN[11])

基金项目:四川省重点研发计划(重大科技专项)(2022YFS0020);南充市市校科技战略合作专项(22SXQT0292)
This work was supported by the Sichuan Provincial Key R&D Plan(Major Science and Technology Project)(2022YFS0020)and Nanchong City-School Science and Technology Strategic Cooperation Project(22SXQT0292).
通信作者:陈金令(chengjl2002@163.com)

虽然试图通过横向连接进行特征集成，但是 FPN 中的顺序方式使得集成特征更多地关注相邻特征，而较少关注其他层特征。FCOS 的 head 部分使用的是传统的 IoU 的策略，对前景和背景区分不充分，从而导致检测精度下降。因此本文改变传统策略，采取一种自适应样本选择的方法，筛选出更准确的正样本点用于真正的训练中，有助于提高训练效率以及准确度。



(a)SENet (b)BoTNet

图 1 SENet 与 BoTNet 热力图对比

Fig. 1 Comparison of SENet and BoTNet heat map

为解决特征提取网络存在的缺乏全局视野以及特征提取不充分的问题，本文提出了一种改进型的 FCOS 目标检测算法。首先，在特征提取网络部分使用 BoTNet[12] 替换原有的 ResNet[13] 网络，其中 BoTNet 使用 MHSA(Multi-Head Self-Attention)[14] 层替换掉 Resnet-50 最后一层网络中的 3×3 卷积，再在 c3 和 c4 特征提取层加入可变形卷积模块。该方法可以使整个网络拥有全局视野，从而能够更好地提取不同阶段特征的丰富信息，其效果优于 SENet[15] 中所使用的注意机制的方法，如图 1 所示。接着，在特征提取模块后面添加了多尺度特征融合模块。该模块是一个 U 型结构，通过自上到下与自底向上相聚合的模块构成。针对回归不够高效、精度有待提高的问题，从确定其正负样本方面入手，构建一个自适应筛选器，用来筛选出真正需要回归部分的 Anchor，从而使得收敛速度更快，回归更加准确。

本文的主要工作可以总结为以下 3 点：1)在主干网络中引入 MHSA 层与可变形卷积层，提高全局性视野能力；2)特征提取中使用 U 型结构的多尺度特征提取模块，进一步增强了不同阶段特征交互能力；3)在 head 部分增加一个自适应筛选器，提高回归精度，加快收敛速度。通过以上 3 点改进对 FCOS 算法中存在的不足进行完善。同时在 MS COCO 数据集[16] 和 PASCAL VOC 数据集[17] 上的实验验证了该改进算法的检测性能，表明了本文方法的有效性。

2 改进型 FCOS 目标检测算法

改进型 FCOS 网络的整体架构如图 2 所示。

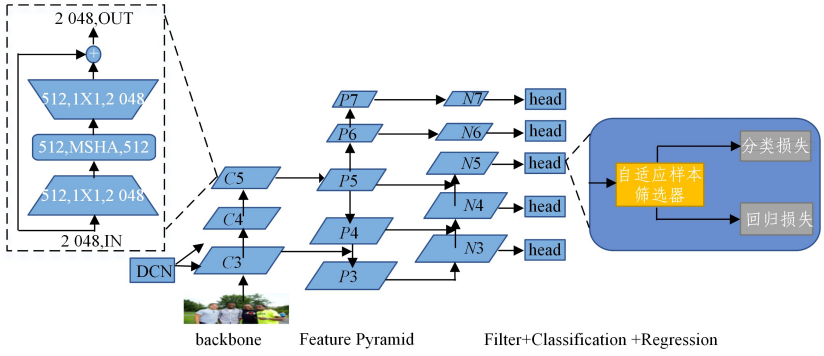


图 2 改进型 FCOS 算法整体架构

Fig. 2 Overall architecture of improved FCOS algorithm

首先输入一张 RGB 彩色图片，然后，在加入的 MHSA 层和可变形卷积的 BoTNet 主干网络中提取图像特征。在特征提取的多尺度预测部分使用 U 型结构的特征提取模块，接着输出 5 层多尺度融合特征图，并在每层特征上连接一个 head 部分，head 部分有两条分支，一条是用于分类，另一条接自适应筛选器用于计算中心度和边界框回归。

2.1 特征提取模块

图 2 左侧部分是基于自注意力机制的 ResNet-50 特征提取模块。ResNet-50 是原始 FCOS 算法中的主干网络，使用 ResNet 每个阶段中最后一个残差块作为输出特征图，即 $C3$，$C4$ 和 $C5$，这 3 层特征构成自下而上的路径。而这 3 层特征分别通过横向连结和自顶部向底部的纵向连接部分得到。$P6$，$P7$ 是经过步长为 2 的 3×3 卷积得到，最后 $P7$，$P6$，$P5$，$P4$ 和 $P3$ 这 5 层特征便构成 FPN 则定上下的路径。由于原有的 ResNet+FPN 的主干网络依然存在不能充分提取图片特征信息的问题，因此，本文算法加入 MHSA 自注意力机制来进一步抓取特征信息。

ResNet 通常有 4 个阶段(残差块)，统称为{$c2$，$c3$，$c4$，$c5$}，分别对应的输入图像的步长为{4，8，16，32}。堆栈[$c2$，$c3$，$c4$，$c5$]由多个具有剩余连接的瓶颈块组成(例如，R50 有[3，4，6，3]个瓶颈块)。本方法是将原始 ResNet 中 $c5$ 层中的 3 个 3×3 空间卷积替换为 MHSA 层，再将 DCN 模块加入到 $c3$，$c4$ 层中。其对比模型结构如表 1 所列，下面将详细介绍替换的 MHSA 自注意力层。

表 1 改进 ResNet50 结构

Table 1 Improve ResNet50 structure

子层名	输出图像大小	ResNet-50	New ResNet-50
Conv_1	512×512	7×7，64，步长为 2 3×3max pool，步长为 2	7×7，64，步长为 2 3×3max pool，步长为 2
Conv_2	256×256	$\begin{bmatrix}1\times1 & 64\\3\times3 & 64\\1\times1 & 256\end{bmatrix}\times3$	$\begin{bmatrix}1\times1 & 64\\3\times3 & 64\\1\times1 & 256\end{bmatrix}\times3$
Conv_3	128×128	$\begin{bmatrix}1\times1 & 128\\3\times3 & 128\\1\times1 & 512\end{bmatrix}\times4$	$\begin{bmatrix}1\times1 & 128\\DCN & 128\\1\times1 & 512\end{bmatrix}\times4$
Conv_4	64×64	$\begin{bmatrix}1\times1 & 256\\3\times3 & 256\\1\times1 & 1024\end{bmatrix}\times6$	$\begin{bmatrix}1\times1 & 256\\DCN & 256\\1\times1 & 1024\end{bmatrix}\times6$
Conv_5	32×32	$\begin{bmatrix}1\times1 & 512\\3\times3 & 512\\1\times1 & 2048\end{bmatrix}\times3$	$\begin{bmatrix}1\times1 & 512\\MHSA & 512\\1\times1 & 2048\end{bmatrix}\times3$
参数量		25.5×10^6	21.8×10^6

2.1.1　MHSA 自注意力层

自我注意是一个计算原语，它基于内容寻址机制实现成对实体交互，从而在长序列中学习丰富的关联特征层次。此方法已经成为 NLP（Natural Language Processing[18]）中 Transformer 块形式的标准工具，突出的例子有 BERT[19] 模型。

在机器视觉中利用自我注意的一种简单方法是使用 Transformer 中提出的多头自我注意力层（MHSA）代替原有的空间卷积。另一方面，视觉转换器（ViT）[20] 则提出在非重叠面的线性投影上叠加 Transformer 块。这些方法展现出两种不同的模型结构。但是在计算机视觉中使用自我注意依然存在几个挑战：1）其图像比分类任务中使用的图像大得多；2）自我注意的记忆的计算量和占用资源比空间维度更大。因此 Srinivas 等[12] 提出了一种很好的解决方法 BoT 块。这种混合的方法不仅可以满足全局性的自我关注，还可以通过卷积进行空间下采样使注意力集中在较小的分辨率上，有效地提高了对高分辨率图像的处理能力。具体结构如图 3 所示。

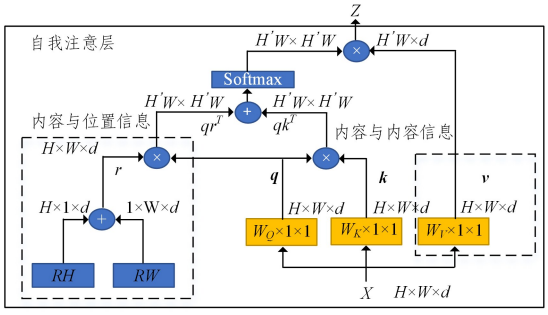


图 3　MHSA 的层在 Bot 块中的详细结构

Fig. 3　Detailed structure of MHSA layer in Bot block

这种全局视野是在 2D 的特征图上执行。这种二维特征图的高度和宽度分别采用相对位置编码 R_H 和 R_W。通常的自注意力可定义为：

$$Attention(\boldsymbol{q},\boldsymbol{k},\boldsymbol{v})=\mathrm{softmax}\left(\frac{\boldsymbol{q}\boldsymbol{k}^{\mathrm{T}}}{\sqrt{d_k}}\right)v \tag{1}$$

其中，$\boldsymbol{q},\boldsymbol{k},\boldsymbol{v}$ 分别表示查询信息、关键词和数值，为矩阵形式，d_k 则表示为均值。

由图 3 可知，Bot 块中所得自注意力可定义为：

$$Attention(\boldsymbol{q},\boldsymbol{k},\boldsymbol{v},r)=\mathrm{softmax}(\boldsymbol{q}\boldsymbol{k}^{\mathrm{T}}+\boldsymbol{q}r^{\mathrm{T}})\boldsymbol{v} \tag{2}$$

其中，$\boldsymbol{q},\boldsymbol{k},\boldsymbol{r},v$ 分别表示查询、键、位置编码（这里使用是相对位置编码）和数值。加法和乘法分别表示为元素求和与矩阵乘法，1×1 则表示逐点卷积。

为了使注意力拥有感知位置信息的能力，在 Transformer 的基础上使用更适合视觉任务的相对位置编码。这是因为注意力不仅仅考虑到了图像的内容信息，同时还考虑到了图像中不同位置特征之间的相对距离，所以可以有效地将跨越对象的信息与位置信息关联起来。

Transformer 中的 MHSA 和 BoTNet 中的 MHSA 有以下区别：1）Transformer 中使用的是 Layer Normalization[21]，而 BoTNet 使用的是 Batch Normalization[22]。2）Transformer 仅仅使用了一个非线性激活，而 BoTNet 使用了 3 个非线性激活。3）Transformer 的 MHSA 包含一个输出投影而 BoTNet 没有。

由于 BoTNet 中的 MHSA 层可以将空间位置信息和上下文语意信息相结合，高效地获取了全局信息从而可以提取到更多的特征信息。因此该方法有效缓解了 FCOS 中存在的难以充分提取全局特征的问题。

2.1.2　可变形卷积

传统的 CNNs 卷积神经网络[23] 是通过增加网络的深度与宽度来提升检测精度。DCN（Deformable Convolutional Networks）[24] 认为卷积核不应该只是一个矩形，而是在不同阶段、不同特征图和不同的像素点都有最优的卷积核结构存在，因此，在原有卷积核上的每个点增加一个可学习的偏移量。可变形卷积流程如图 4 所示。

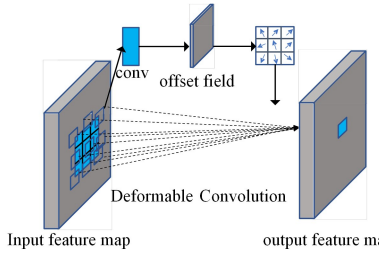


图 4　可变形卷积详细过程

Fig. 4　Detailed process of deformable convolution

可变形卷积是在原二维卷积中得到启发，通常二维卷积可以定义为：

$$y(P_0)=\sum_{P_n\in R} w(P_n)\cdot x(P_0+P_n) \tag{3}$$

其中，$y(P_0)$ 表示每个位置在输出特征图上对应的结果，w 表示加权系数，x 表示输入特征映射图。R 代表一个膨胀成度为 1 的 3×3 卷积核，使用这个卷积核在输入特征图中进行抽样，其大小与膨胀程度代表着感受野的大小，如式(4)所示：

$$R=\{(-1,-1),(-1,0),\cdots,(0,1),(1,1)\}$$

其中，R 代表着一个膨胀程度为 1 的 3×3 卷积核，w 表示加权系数，x 表示输入特征映射图。

可变形卷积不是改变卷积核的大小，而是通过对采样结果的改变间接达到改变卷积核的大小的效果。引入一个偏移量 ΔP_n，即：

$$y(P_0)=\sum_{P_n\in R} w(P_n)\cdot x(P_0+P_n+\Delta P_n) \tag{5}$$

其中，$\{\Delta P_n\mid n=1,\cdots,N\}$ 且 $N=|R|$。由于偏移量往往是小数，因此通过双线性插值的方法实施。

由于可变形卷积可以通过增加偏移量的方式来增加网络的感受野，可以得到更多的特征信息，因此有助于缓解 FCOS 中难以充分提取全局特征信息的问题。

2.2　多尺度多级预测模块

原始模型采用 5 层特征金字塔作为连接部分，再进行不同尺度的目标回归。这 5 层尺度回归的目标大小分别是[0，64]，[64，128]，[128，256]，[256，512]和[512，∞]，分别对应 FPN 中的 $P3,P4,P5,P6,P7$。但针对检测小物体与多目标重叠问题，原算法还是有一定的不足。对于 FCOS 模型，每层每个像素点都会回归固定尺度大小范围内的目标。相对地，假设目标集中在某个尺度范围内，将会使得检测层的工作量非常大，导致检测质量不佳。此问题同样是影响模型性能的因素之一，在多目标检测场景中会导致 FCOS 模型的监测功能稍有降低，同时也阐明，当检测任务复杂、检测目标数量较大时，本文提出的多尺度多级检测会使 FCOS 检测性能提高。

如图 2 所示，在 BoTNet 特征提取模块后面连接具有

自底向上路径的路径聚合模块。$P7$,$P6$ 层特征图不经过任何操作,直接作为 $N7$,$N6$。以 $P4$ 层特征图为例,$P4$ 与经过 2 倍下采样的 $N3$ 按元素相加,相加得到的特征图再经过 3×3 卷积,即生产特征图 $N4$。以此类推,生产特征图 $N5$,最后生成 5 层通道数都为 256 的特征。$N3$,$N4$,$N5$,$N6$ 和 $N7$ 就构成了具有自底向上路径的聚合模块。该模块通过扩充自下而上的路径,在较低特征层上用准确的定位信号增强整个特征金字塔,有效加强了信息流。

2.3 自适应样本筛选模块

在以前的样本选择策略中存在一些敏感的超参数,如基于锚框的检测器中的 IoU 阈值和无锚检测器中的标度范围。设置这些超参数后,所有真实地面框必须按照固定规则来抉择其正样本,这适用于大多数对象,但会遗漏一些外部对象。

这些超参数设置不同的值将产生差异很大的结果。因此本文选择加入 ATSS[25] 自适应正负样本筛选模块,几乎不需要任何超参数。ATSS 的整体结构如图 5 所示。

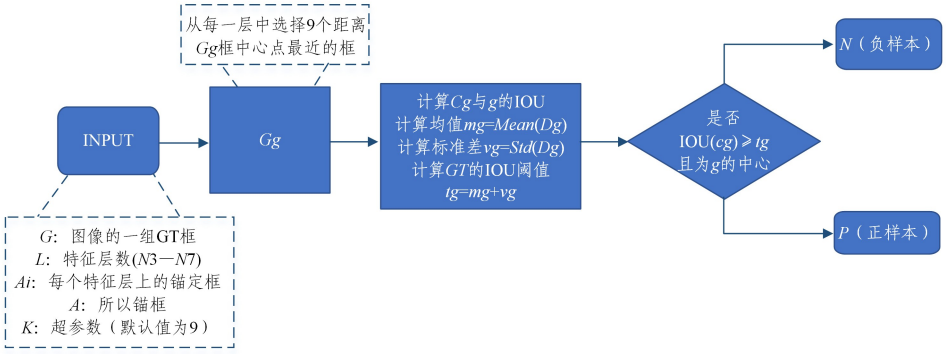


图 5 ATSS 整体结构

Fig. 5 Overall structure of ATSS

由原 FCOS 所得论断:离目标中心越近的定位点将产生更高质量的检测锚框。从而得出结论,更好的候选锚点是那些与对象中心点距离接近的点。

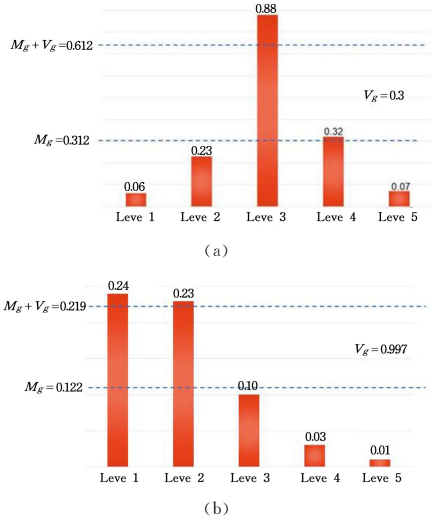


图 6 每个级别都有其候选者及对应的 IOU

Fig. 6 Each level has its candidates and corresponding IOUs

为了得到更好的候选锚点,本文使用标准差与均值之和作为 IOU 的阈值,原因是目标对象的 IoU 平均值是该对象预设锚框的适用性度量。如图 6 所示,高 M_g 代表着其拥有高质量的候选者,同时 IOU 阈值相对较高。相反,M_g 较低代表着其大多数候选对象质量较低,IOU 阈值应较低。对象的 IoU 标准差则代表应该在哪些层适宜检测该对象的指标。如图 5 所示,高 V_g 意味着存在一个专门适用于该对象的金字塔层,因此仅从该层选择正样本。如若 V_g 值较低,意味着同时存在二个或两个以上的适合该对象的金字塔层,将 M_g 与 V_g 相加后能够得到一个低阈值 T_g,再根据统计特性,从适当的金字塔层中为每个对象自适应地抉择足够的前景框。

3 实验结果与分析

3.1 数据集与评价指标

该改进算法在 MS COCO 和 PASCAL VOC 两大公共数据集上进行实验,其中 MS COCO 数据集包含 80 个类别,100 000 张用于训练的图片,5 000 张用于验证的图片,20 000 张用于测试的图片。在 test-dev 2017 上进行本文算法实验并与最新的目标检测算法相比较,然后使用 val 2017 进行消融实验,实验结果都遵循 MS COCO 标准的平均精度(Average Precision,AP)指标。其中,AP 表示 IoU 从 0.5 开始,每隔 0.05 作为阈值,直到取到 0.95 得到的平均精度再平均的结果,AP_{50} 表示 IoU 阈值为 0.5 时的平均精度,同理,AP_{75} 表示 IoU 阈值为 0.75 时的平均精度。APs,AP_M 和 AP_L 分别代表小、中、大目标的平均精度。PASCAL VOC 数据集包含 20 个类别,其中训练图片 20 000 张(trainval2007+2012),测试图片 4 000 张(test2012),实验结果遵循 VOC 数据集的指标,其中平均精度表示该类别的 IoU 阈值为 0.5 时的平均精度。

3.2 训练参数设置

本文实验基于 PyTorch1.4 深度学习框架实现,操作系统为 Ubuntu,使用了 2 块 NVIDAIA GeForce RTX GPU 训练,显存为 12GB。由于 FCOS 模型要求较高,存在内存不够的问题,本实验通过线性策略调整了 batch_size 大小和 IMS_PER_BATCH 的数量。本文的基线模型是 FCOS,其他超参数设置都沿用 FCOS 原有的参数。

3.3 实验结果对比与分析

本节在 COCO 测试集与 PASCAL VOC 测试集上评估了本文提出的改进型 FCOS 目标检测算法的性能。为了更好地体现本文改进算法的优势,本文选择了以下几种在单阶段、两阶段、有锚框和无锚框的经典算法进行对比实验。表 2 列出了比较算法与本文算法的 AP 值。从表中可以看出改进的 FCOS 算法的 AP 可以达到 41.6%,性能得到显著提升。与原 FCOS 算法相比,AP 值提高了 3%,

特别是对于小目标检测的精度提高了近两个百分点，大大提高了检测物体的准确性。

表 2　改进型 FCOS 算法与其他最新目标检测算法对比

Table 2　Comparison of improved FCOS algorithm and other latest target detection algorithms

（单位：%）

方法	主干网络	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
YOLOv3[2]	DarkNet-53	33.0	57.9	34.4	18.3	25.4	41.9
RetinaNet[4]	ResNet-50	36.9	56.3	39.3	20.6	39.9	46.5
Faster R-CNN[5]	ResNet-101	36.2	59.1	39.0	18.2	39.0	48.2
Mask R-CNN[26]	ResNet-101	38.2	60.3	41.7	20.1	41.1	50.2
CornerNet[8]	Hourglass-104	40.5	56.5	43.1	19.4	42.7	53.9
CenterNet[7]	DLA-34	39.2	57.1	42.8	19.9	43.0	51.4
FCOS[6]	ResNet-50	38.6	57.4	41.4	22.3	42.5	49.8
本文方法	BotNet-50+DCN	41.0	60.1	45.4	24.1	43.2	52.6

在 VOC 测试集上，实验主要是将改进型 FCOS 目标检测算法的各个类别与原 FCOS 算法作比较，其中实线表示原 FCOS 算法，虚线表示改进型 FCOS 算法，具体结构如图 7 所示。从图 7 不难发现，改进型 FCOS 算法的平均精度普遍高于原 FCOS 算法，其中有个别类的平均精度远高于原算法，特别是对于小目标和重叠目标的图片。

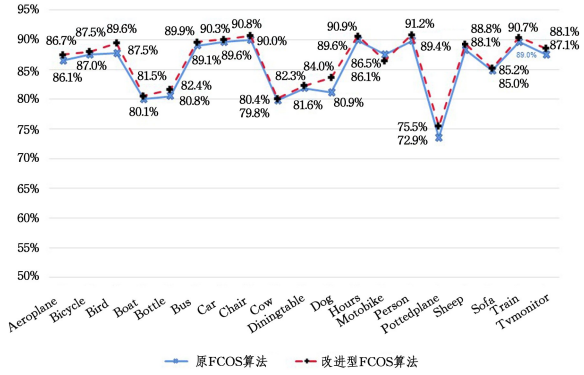
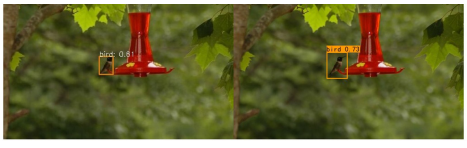


图 7　改进型 FCOS 算法与原 FCOS 算法在 VOC test2007 中 20 类的目标比较

Fig. 7　Comparison of 20 types of targets in VOC test2007 between improved FCOS algorithm and original FCOS algorithm

下面是从 MS COCO 数据集中随机选取的一些图片进行可视化，本文选取两组具有代表性的检测结果进行对比。图 8(a)给出了 FCOS 算法的可视化结果，图 8(b)给出了改进型 FCOS 算法的可视化结果。可以看出，改进型 FCOS 目标检测算法的检测结果能够更好地识别物体，检测边框也更精确。



(a)

(b)

图 8　COCO 数据集上可视化结果对比

Fig. 8　Comparison of visualization results on COCO data set

3.4　消融实验

消融实验如表 3 所列，其中 Bot 代表 BoTNet 主干网络，Ut 代表多尺度路径融合模块，ATSS 代表自适应正负样本筛选器。

(1)采用 BoTNet 作为 Backbone 的优势

在 FCOS 上采用 BoTNet 作为 Backbone 起到了有效的提升作用，相比原 ResNet，降低了网络中的参数量，以便能使用尽可能少的计算资源来训练。本文分别计算了 ResNet50 与 BoTNet50 的参数量，如表 3 所列。

表 3　ResNet50 与 BoTNet50 的参数量对比

Table 3　Comparison of parameters between ResNet50 and BoTNet50

Backbone	*Epoch*	*Params*
ResNet50	100	25.5 MB
BoTNet50	100	20.8 MB

(2)ATSS 模块中超参数 k 对精度的影响

通过对 VOC 数据进行针对超参数 k 的不同值对最终精度的影响的实验，如表 1 所列，证明 k 值的大小对精度的影响可以忽略不计。

表 4　超参数 k 值的不同大小对精度的影响

Table 4　Impact of different values of hyperparameter k on accuracy

k	AP/%
3	38.0
5	38.8
7	39.1
9	39.3
11	39.1

(3)对比 ATSS 在不同网络中的使用情况

本文还比较不同的网络模型使用 ATSS 模块的情况，如表 5 所列。本文算法使用 ATSS 的效果最佳。

表 5　不同算法使用 ATSS 的使用情况对比

Table 5　Comparison of different algorithms using ATSS

algorithm	AP/%
RetinaNet+ATSS	36,9(+1.2)
FasterRCNN+ATSS	36.2(+0.9)
本文算法+ATSS	40.3(+1.4)

(4)3 个改进部分不同方式组合的比较

如表 6 所列，通过 3 个模块的不同组合情况进行两两结合比较分析可以发现，在原 FCOS 基线上单独加入一个模块，或者加入任意两个模块都不能达到最佳性能。原因在于 3 个部分都很重要，3 个模块分别对应着特征提取、多级预测、边框回归，因此在改进目标检测算法时，不能着眼于其中的某一个部分，而要从全局的视角来分析所存在的问题，不同问题要有针对性的解决方法。

因此，这 3 个改进部分相结合不仅填充了缺失全局视野的空缺并解决了难以充分提取目标信息的问题，还优化了正负样本区分不够精确导致的边界框回归不够精确的准确问题，验证了该改进算法的有效性。如表 6 所列，在 COCO val 数据集上的检测精度比 FCOS 算法高出了 2.8%，性能显著提高。

表 6 MHSA+DCN、改进的 neck 部分、ATSS 自适应筛选模块这 3 部分的对比分析实验

Table 6 Comparative analysis experiment of MHSA+DCN, improved neck part and ATSS adaptive screening module

（单位：%）

MHSA+DCN	U_t	ATSS	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
			38.6	57.4	41.4	22.3	42.5	49.8
		√	39.2	57.3	42.4	22.7	43.1	51.5
√			40.1	58.0	42.6	23.0	43.6	51.1
√	√		40.3	58.4	43.2	23.2	44.5	51.3
√	√	√	41.4	60.1	45.4	24.1	45.0	52.6

结束语 本文针对 FCOS 算法中存在的缺乏全局视野和难以充分提取目标信息以及正负样本区分不够充分的问题，提出了一种改进型 FCOS 目标检测算法。具体来说，本文在特征提取网络中分别融入了多头注意力模块与可变性卷积，然后与 U 型路径特征提取模块相结合构成多尺度特征提取网络。最后，在边界框回归中添加自适应正负样本筛选模块，在提取出“真正的”正样本的同时，提高了边界框回归不够精确的问题。通过以上改进，有效地解决了传统 FCOS 算法中存在的问题，其检测性能也在 COCO 和 VOC 两大公告数据集上取得显著提升。但在实验中也发现，本文提出的方法在实际应用中，检测帧率与参数量方面还需要改善。下一步针对如何提高实际检测场景中的检测帧率进行后续研究。

参 考 文 献

[1] REDMON J, FARHADI A. YOLO9000: Better, Faster, Stronger [C]// IEEE Conference on Computer Vision & Pattern Recognition. IEEE, 2017: 6517-6525.

[2] REDMON J, FARHADI A. YOLOv3: An Incremental Improvement[J]. arXiv: 1804.02767, 2018.

[3] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single Shot MultiBox Detector[C]// European Conference on Computer Vision. Cham: Springer, 2016: 21-37.

[4] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal Loss for Dense Object Detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2020, 42(2): 318-327.

[5] REN S Q, HE K M, GIRSHICK, 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.

[6] TIAN Z, SHEN C, CHEN H, et al. FCOS: A simple and strong anchor-free object detector[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 44(4): 1922-1933.

[7] ZHOU X, WANG D, KRHENBÜHL P. Objects as Points[J]. arXiv: 1904.07850, 2019.

[8] LAW H, DENG J. CornerNet: Detecting Objects as Paired Keypoints[J]. International Journal of Computer Vision, 2020, 128(3): 642-656.

[9] YANG Z, LIU S, HU H, et al. RepPoints: Point Set Representation for Object Detection[C]// 2019 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, 2019.

[10] LONG J, SHELHAMER E, DARRELL T. Fully Convolutional Networks for Semantic Segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 39(4): 640-651.

[11] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature Pyramid Networks for Object Detection[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, 2017.

[12] SRINIVAS A, LIN T Y, PARMAR N, et al. Bottleneck Transformers for Visual Recognition[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 16519-16529.

[13] 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.

[14] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]// Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). 2017: 6000-6010.

[15] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 7132-7141.

[16] LIN T Y, MAIRE M, BELONGIES, et al. Microsoft COCO: common objects in context [C]// European Conference on Computer Vision (ECCV). Cham: Springer, 2014: 745-755.

[17] EVERINGHAM M, VAN G 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.

[18] KANTOR P B. Foundations of Statistical Natural Language Processing[J]. Information Retrieval, 2001, 4(1): 80-81.

[19] CARION N, MASSA F, SYNNAEVE G, et al. End-to-End Object Detection with Transformers[C]// European Conference on Computer Vision. Cham: Springer, 2020: 213-229.

[20] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale[J]. arXiv: 2010.11929, 2020.

[21] BA J L, KIROS J R, HINTON G E. Layer Normalization[J]. arXiv: 1607.06450, 2016.

[22] LOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]// International Conference on Machine Learning. PMLR, 2015: 448-456.

[23] LECUN Y, BENGIO Y. Convolutional networks for images, speech, and time series[M]// The Handbook of Brain Theory and Neural Networks. MIT press, 1998: 255-258.

[24] DAI J, QI H, XIONG Y, et al. Deformable Convolutional Networks[C]// Proceedings of the IEEE International Conference on Computer Vision. 2017: 764-773.

[25] ZHANG S, CHI C, YAO Y, et al. Bridging the Gap Between Anchor-Based and Anchor-Free Detection via Adaptive Training Sample Selection[C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020.

[26] FANG L P, HE H J, ZHOU G M. A review of target detection algorithm research[J]. Computer Engineering and Applications, 2018, 54(13): 11-18, 33.

CHEN Jin-ling, born in 1975, Ph.D, professorate senior engineer. His main research interests include deep learning and image processing.