DOI:10.19651/j.cnki.emt.2209283

基于可变形卷积改进 SSD 算法的目标检测方法*

蒋 晨 钱永明 姚兴田 李 壮
(南通大学机械工程学院 南通 226019)

摘 要: 为了提高传统 SSD 算法对小目标检测的准确率,提出一种改进的 SSD 目标检测算法:采用基于可变形卷积的 ResNet50 作为 SSD 算法的特征提取网络,提高对目标的处理能力;设计特征金字塔(FPN)来融合不同层的特征图,丰富浅层特征图的语义信息;在特征融合时引入通道注意机制,提取相应的通道权重,增加重要信息的比例,提高检测效果。最后采用 PASCAL-VOC2007 开源数据集进行仿真实验,并与传统 SSD 目标检测算法进行对比,准确率得到显著提高,验证了该算法对小目标检测的有效性。

关键词: 目标检测;可变形卷积;特征金字塔;注意机制
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

Target detection method based on deformable convolution improved SSD algorithm

Jiang Chen Qian Yongming Yao Xingtian Li Zhuang
(School of Mechanical Engineering, Nantong University, Nantong 226019, China)

Abstract: In order to improve the accuracy of the traditional SSD algorithm for small target detection, an improved SSD target detection algorithm is proposed: ResNet50 based on deformable convolution is used as the feature extraction network of the SSD algorithm to improve the processing ability of the target; the feature pyramid (FPN) to fuse feature maps of different layers and enrich the semantic information of shallow feature maps; introduce channel attention mechanism during feature fusion, extract corresponding channel weights, increase the proportion of important information, and improve the detection effect. Finally, the PASCAL-VOC2007 open source data set was used for simulation experiments, and compared with the traditional SSD target detection algorithm, the accuracy is significantly improved, which verifies the effectiveness of the algorithm for small target detection.

Keywords: object detection; deformable convolution; feature pyramid; attention mechanism

0 引 言

目标检测作为计算机视觉领域的一个重要分支有着广泛的应用。它的目的是在计算机输入的图像中找到目标物体并进行分类判断[1]。传统的目标检测方法主要为 3 个环节[2]:1)在输入图按照一定规则挑选相应的候选区域;2)使用特征提取器对候选区域进行特征提取;3)使用预训练过的分类器对其分类。传统的检测算法具有较高的时间复杂度,而且手工设计的特征对多样化对象不具有鲁棒性。2014 年,GRSHICKL[3] 等首次提出将卷积神经网络 R-CNN 应用于目标检测,极大提高了检测的准确率。将深度学习应用在目标检测上快速促进了该领域的发展,从此以后很快涌现了一些在精度与速度上更为优秀的算法。

基于深度学习的目标检测算法可以分为两类:两阶段目标检测算法和一阶段目标检测算法。两阶段目标检测算法是以 R-CNN 为核心展开的一系列算法[4],例如 Fast R-CNN 和 Faster R-CNN[5-7] 等算法,这类算法的主要检测步骤是:优先生成目标候选框;再通过卷积池化等操作提取特征层;最后进行候选框的位置回归以及样本分类[8]。而一阶段的目标检测算法如 YOLO[9-11] 以及 SSD[12-14] 等算法,它们没有预先生成目标候选框的步骤,而是直接对输入图像进行特征提取,同样最后进行目标分类和位置回归任务。由于一阶段的目标检测算法整个流程是在一个网络中进行的,比两阶段的目标检测算法速度更快,但特征提取过程丢失大量目标信息,对小目标检测效果较差。针对 SSD 目标检测算法的不足,Sun 等[15] 利用深层提取的语义信息

收稿日期:2022-03-15
* 基金项目:江苏省仿生功能材料重点实验室开放课题基金(BFM2101)项目资助

来增强目标检测特征,同时在不增加太多计算量的情况下提高了骨干网接受域的规模,从而大大提高了目标检测的性能。Zhang 等[16]提出一种光多次扩张卷积(LMDC)算子,LMDC 从特征映射中能够提取全局语义信息,从而使得特征信息更加完善和准确。Choi 等[17]通过合并注意流和特征映射级联流的输出,生成增强的特征映射,从而提高对小目标的检测。

传统 SSD 检测算法在 YOLO 和 Faster-RCNN 两者的基础上做了进一步的优化[18-19]。该算法有效融合了 RPN (region proposal network),其目标的分类任务和回归任务同步执行,较好地平衡了精度与速度,综合性能比较优秀,但是在对小目标进行检测时,提取的特征图不够完整,语义信息丢失严重,导致检测效果较差。针对上述问题,本文提出了一种基于变形卷积的改进型 SSD 目标检测方法。首先,利用基于变形卷积的 ResNet 网络作为特征提取网络,在不同尺度上准确定位目标的关键区域;其次,利用轻量级 FPN 有效的融合高级特征和低级特征,增加浅层特征的语义信息;最后,在局部特征融合之间引入通道注意机制,提高特征图语义信息的表达能力,最终实现不同尺度的目标检测。

1　本文方法

本文提出的网络模型结构如图 1 所示。首先针对检测目标的形状大小多样化,采用基于可变形卷积的 ResNet50 作为特征器适应性提取特征,(图 1 中可变形卷积用 dcn 表示);随后,针对浅层特征图缺乏深层语义信息,通过轻量型 FPN[20]+特征融合进行改善;最后,为了让模型更高效的检测目标,在 FPN 模块第一个特征融合后,采用通道注意力模块获取各通道权重,根据权重大小选择性的提高有效特征的表现,同时抑制无意义的特征。

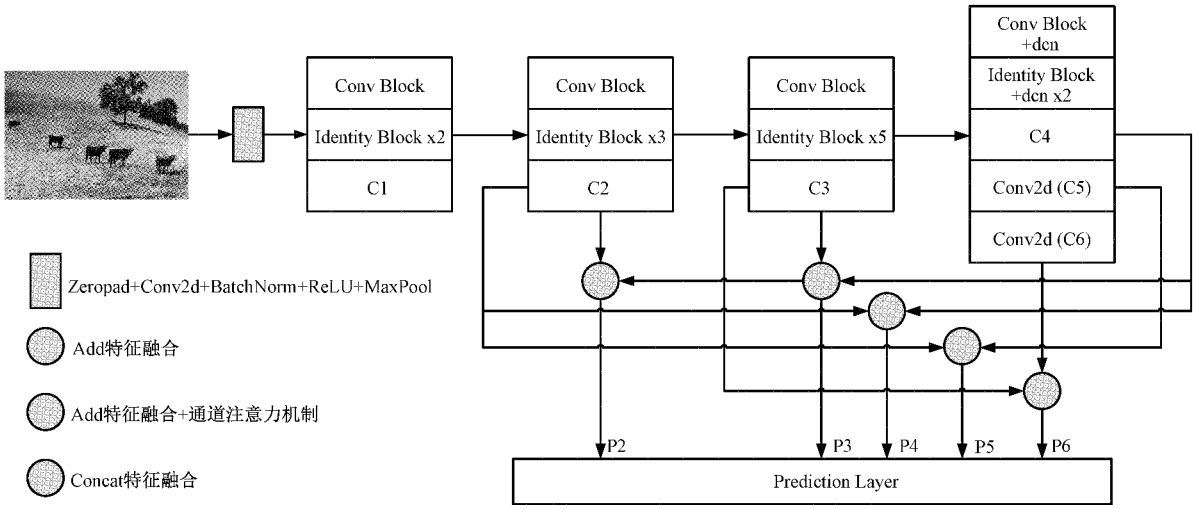


图 1　本文方法的网络模型结构

1.1　基于可变形卷积的特征提取网络

Gurita 等[21]使用可变形卷积模块来自适应性提高模型在小数据集上的分割性能。可变形卷积的改进之处是在传统卷积的基础上,调整卷积核的方向向量,使得卷积核能够根据目标的形状适应性采样。因此,为了适应各种形式的物体,本文引入了可变形卷积对采样位置进行自由采样,而不局限于方正的格点。

传统 CNN 定位采样方法难以适应物体的变形。该过程模型可表示为:

$$y(p_0)=\sum_{p_n\in R} w(p_n)\cdot x(p_0+p_n) \tag{1}$$

其中,x 表示输入特征图,卷积核按照方正的网格点对其进行采样。w 表示权重,对于输出 y 上的位置 p_0,输出特征映射等于 w 赋予的采样值之和。

对于每个 p_n,R 是 p_0 在的邻近点:

$$R=\{(-1,1),(-1,0),\cdots,(0,1)(1,1)\} \tag{2}$$

DCN(可变形卷积)的主要特点就是能够对特征自由采样,具有学习空间几何变形的能力。这非常适合于检测不同大小和形状的物体,而该方法只是在一定程度上增加了计算时间。

对于每个采样点具有额外学习目标偏移的可变形卷积:

$$y(p_0)=\sum_{p_n\in R} w(p_n)\cdot x(p_0+p_n+\Delta p_n) \tag{3}$$

对于每个附加 ΔP_0 后的 P_0,采样变得不规则,这使得新方法的变换建模能力优于传统 CNN。

如图 2 所示,一个 3×3 卷积单元在 9 个固定位置对输入图像进行采样,而 DCN 首先通过额外的 3×3 卷积层预测 9 个采样点的偏移位置,这些预定的偏移量与输入图像具有相同的空间分辨率。通道尺寸 18 对应于这 9 个采样点的 2D 偏移。因此,DCN 通过将这些预测偏移量加到固定的抽样位置来命名空间采样位置。

原始的 3×3 卷积层与 DCN 中 3×3 卷积层具有相同的超参数,但后者由于需要学习偏移量导致其输出通道为

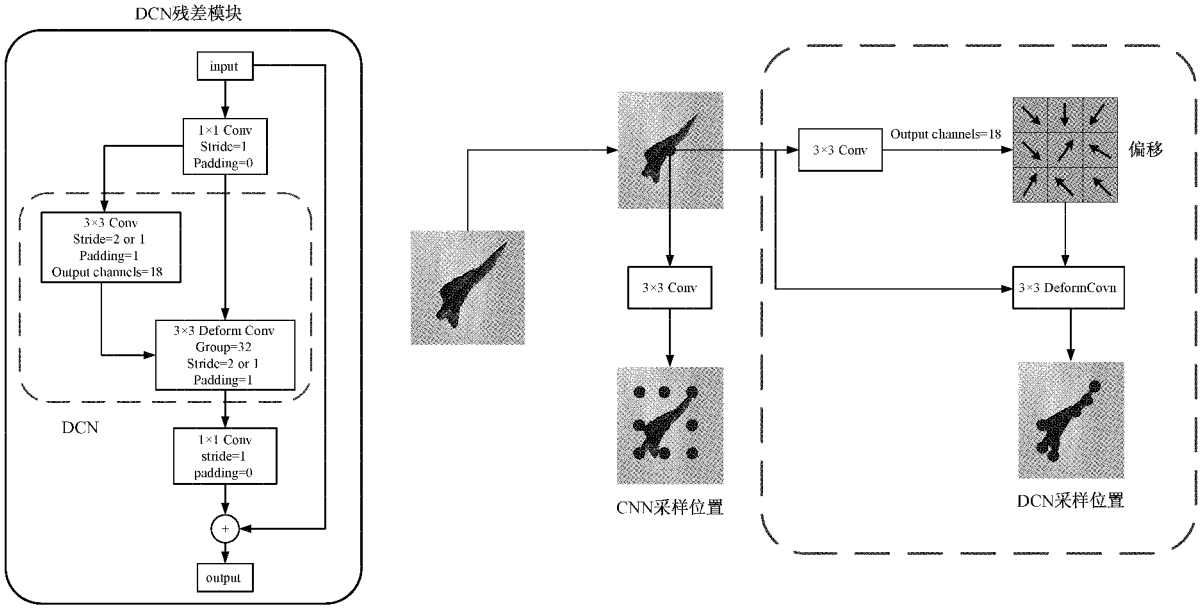


图 2　传统 CNN 与 DCN 对于相同区域的不同采样位置

18。在检测系统模型中使用 DCN 的最佳工作位置是主干部分即特征进行提取重要部分，它可以通过提高主干提取有用特征的能力。基于前人研究的启发，可以在主干中使用基于 DCN 的残差模块来替换传统 3×3 卷积。

本文的特征提取网络是基于 ResNet50，该网络中加入了跨层连接，能够在网络无效学习的时候保留原始信息，避免准确率下降的问题。本文设计的可变形 ResNet50 网络结构如图 3 所示，将 Conv4_x 层中的 3 个 3×3 传统卷积替换成可变形卷积，构成基于 DCN 的 ResNet50 特征提取网络。从以上对变形卷积的描述可以看出，基于可变形卷积 ResNet50 可以提取到更加准确的特征，提高网络检测变形目标的能力。

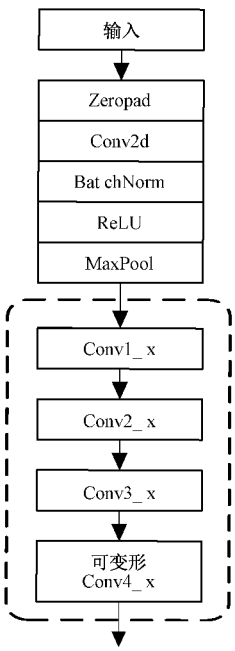


图 3　可变形 ResNet50 网络结构

1.2　轻量型 FPN 与特征融合

1)轻量型 FPN

本文采用基于可变形卷积的 ResNet50 作为 FPN 的基本网络。在该网络中，根据检测对象的形状大小特征选择语义信息较丰富的{C2，C3，C4}这 3 层特征图。在自下而上的网络中构建轻量型特征金字塔{P2，P3}，过程中使用了上采样操作与横向连接操作。FPN 构建原理如图 4 所示，左侧表示为将输入图像经过 3 个卷积模块进行特征提取的过程，步长设置为{16，32}。右侧表示为特征图的上采样以及融合过程，过程中使用大小为 1×1 且步长为 1 的卷积核进行降维操作以减少计算量。在每次融合后的特征图上可以进行研究一次 3×3 的卷积网络操作来消除上采样技术带来的混叠效应，最终得到两个主要特征图。

$$F_c = \sigma(SL(AvgPool(F)) + SL(MaxPool(F))) \tag{4}$$

$$F'_c = F + F_c \tag{5}$$

$$F''_c = F + F'_c \tag{6}$$

2)特征融合

特征融合方式主要分为两类：一种是特征值的对应叠加，通道数保持不变，即 Add 融合方式；另外一种是按照对应通道进行合并，每一特征下的信息量保持不变，即 Concat 融合方式。本文 FPN 过程中采用 Add 方式进行特征融合，对应的 3 个特征图{P4，P5，P6}的融合方式分别为{Add，Concat，Concat}。P4 采用 Add 方式融合增加了特征下的信息量，P5 与 P6 由于跨层融合时会带来噪声干扰，使用 Concat 方式能够减轻这类干扰。

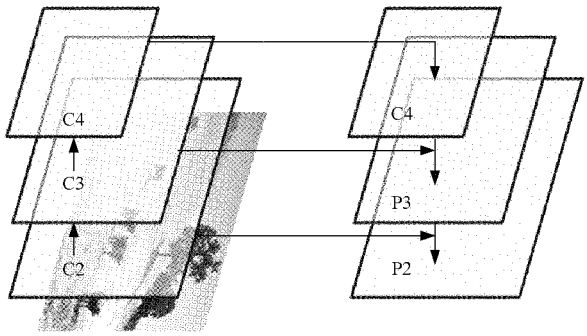


图 4　构建特征金字塔

1.3　通道注意力模块

通道注意力模块(coordinate attention, CA)通常作用于不同尺度特征的融合阶段。与直接加法或信道级联相比,CA 考虑了不同信道特征之间的相关性。CA 的基本理论思想是通过分析计算可以量化各通道的关联性,对信息的重要影响程度重新分配权重,最后进行特征图的加权操作。CA 的结构如图 5 所示,输入特征 F 通过 AvgPool(全局平均池化)以及 MaxPool(全局最大池化)得到两个均包含全局信息的 C×1×1 的特征图。使用两个不同的池操作可以避免单一池方法造成的信息损失。接下来,使用包含全连接操作和激活操作的参数共享层(parameter sharing layer, SL)来建模通道间的关系,通过添加全局池输出特征图的每个像素来聚合语义信息。接着使用 Sigmoid 激活函数求得相应的权重参数 F_c。最后将 F_c 与原始输入特征图按照逐元素相乘得到特征图 F'。

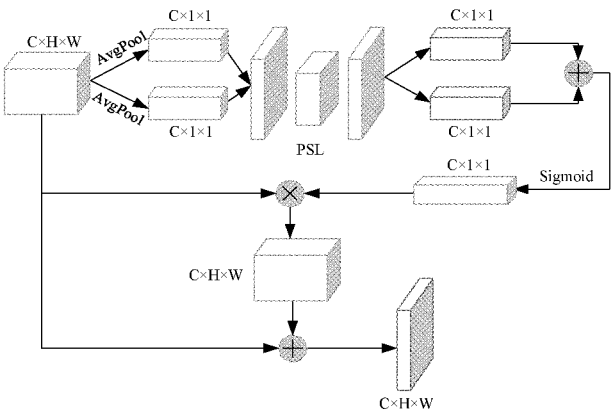


图 5　通道注意力模块

最后通过残差机制将 F′与 F 逐元素相加得到最终的输出特征图 F''_C。过程中通道注意力模块的计算公式如式(4)~(6)所示,其中 σ 表示为 Sigmoid 激活函数,SL 表示为参数共享层,F 表示为输入特征图。

2　实验结果及分析

2.1　数据集

开源的 PASCAL-VOC2007 数据集提供了 20 种类别的图片,表 1 为该数据集详细的物体类别。为了验证该算法对小目标识别的有效性,从上述数据集中选取 236 幅具有代表性的小目标图片作为实验数据集。选取的图片中包含物体的类别共有 9 种,分别为 aeroplane、bottle、bird、boat、cat、cow、person、cow 和 chair,对这 236 张图片进行相应处理后,共有 1 528 个标注物体的 ground truth。

表 1　PASCAL-VOC2007 数据集的物体类别

目录	类别
Vehicle	aeroplane、bicycle、boat、bus、car、motorbike、train
Animal	bird、cat、cow、dog、horse、sheep
Indoor	bottle、chair、diningtable、pottedplant、sofa、tvmonitor
Person	person

2.2　实验细节

本文实验的编程语言是 Python,使用的深度学习框架为 Pytorch。在 Ubuntu18.04 系统的服务器上,安装 CUDA11.1 版本的环境,采用 TC4 系列的 GPU 进行加速。本文使用基于 DCN 的 ResNet50 作为特征提取网络,对输入的图像随机水平翻转,并采用 BN(批规范化)进行正则化。模型训练的初始学习率设置为 5×10^{-4},且每 30 个周期降低为之前的 1/10,权值下降率为 5×10^{-4},Batch size 设置为 16。

2.3　实验结果

为了验证本文提出模型的有效性与优越性,将 VOC2007 数据集分别在 Faster R-CNN、传统 SSD、DSOD、LMDC-SSD、FSSD、DSSD 上进行训练测试。

检测结果如表 2 所示,本文算法的平均精确率(mean average precision, mAP)达到 80.0%,相较于检测效果优异的两阶段算法 Faster R-CNN 获得了 11.8%的提升,较传统 SSD 获得了 8.7%的提升,而且对于这 9 个对象的精确率(average precision, AP),本文算法除了对 aeroplane 检测的 AP 值比 DSOD 低 0.1%之外,对其它对象检测的 AP 值均为最高,这说明通过本文改进之后的方法能有效提升小目标的检测效果,整体检测效果优异。

从表 2 中数据可知,本文基于 SSD 改进算法的检测精确率比 DSOD、FSSD、DSSD 三种基于 SSD 改进的算法都高。其中 DSOD 算法的 mAP 值比本文算法低 7.1%,该算法是将 DenseNet 和 SSD 框架进行了融合,改进不多,仅在精度稍有提升,而本文算法不仅做了特征融合,提高了网络特征提取能力,而且通过可变形卷积提高了模型的自适应性。FSSD 算法的 mAP 值比本文算法低 7.7%,该算法是在 SSD 算法的颈部添加了 FPN,特征融合使得提取的特征信息更为精确,在这一方面本文算法除了根据数据集有效设计 FPN 之外增加了通道注意机制模块,提高了重要特征信息的占比,特征提取能力远大于 FSSD。DSSD 检测效果比 FSSD 与 DSOD 要好,但 mAP 值比本文算法低

表 2　PASCAL-VOC2007 数据集 9 类物体检测结果

算法	mAP	检测结果								
		aeroplane	bird	boat	bottle	cat	chair	person	cow	sofa
Faster R-CNN	68.2	80.0	78.7	62.8	45.9	86.9	34.8	78.3	75.3	70.9
SSD	71.3	83.0	84.6	65.9	44.3	88.6	43.1	81.0	74.9	76.5
DSOD	72.9	85.4	89.1	64.7	47.7	89.4	43.9	81.3	76.9	77.5
FSSD	72.3	84.8	86.0	65.7	45.6	90.2	43.7	81.4	74.9	78.0
DSSD	74.6	83.2	85.6	69.6	55.3	90.4	47.2	83.7	81.0	75.4
Ours	80.0	85.3	95.7	86.1	53.7	94.1	51.9	83.9	83.2	86.8

5.4%，该算法在 SSD 基础上做了特征融合以及通过聚类得到适应数据集对象的尺寸，但对于 VOC2007 数据集对象的形态大小多样性来说，通过本文算法的可变形卷积操作更为合适。总体来看，本文基于 SSD 改进的算法更为优异。

2.4　消融实验

为了验证本文提出改进方法的有效性，本文在上述 9 物体中选取 4 种(bird、boat、cat、cow)作为实验对象，分别设计不同模块的对比实验。

1)为了验证 DCN 的添加对模型检测精度的影响，分别在原始 ResNet50 不同位置加入 DCN 进行对比实验，实验结果如表 3 所示。实验结果表明，可变形卷积的添加可以提高模型的检测效果，在不同位置加入 DCN 后，模型的平均精度均得到提高，这是由于 DCN 在提取特征时能根据目标的形态大小自适应调整提取范围，更为精准提取到目标特征。同时发现 DCN 添加的位置越靠后对模型性能的提升就越大，即将 Conv4_x 层中的 3×3 标准卷积替换成可变形卷积的效果最好。因此本文在接下来的对比实验中均使用改进后的 ResNet50 网络作为特征提取网络。

表 3　DCN 对检测效果的影响

特征提取网络	bird	boat	cat	cow	mAP
ResNet50	86.34	67.21	90.72	77.18	80.36
DCN-Conv1	86.56	68.15	90.73	77.94	80.85
DCN-Conv2	86.56	68.30	91.02	78.63	81.13
DCN-Conv3	86.33	68.51	91.02	79.38	81.31
DCN-Conv4	87.29	68.58	91.87	79.82	81.89

2)为分析轻量 FPN 与特征融合的有效性，本文设计了以下 4 个对比实验并将结果记录在表 4 中。其中方案一表示无 FPN 无特征融合的情况，方案二表示有 FPN 无特征融合的情况，方案三表示无 FPN 有特征融合的情况，方案四表示有 FPN 和有特征融合的情况。实验结果表明，使用了 FPN 或者特征融合均能不同程度够提高模型的准确率，这两种方法有效增强了提取的特征信息，不仅能够增强浅层特征的语义信息，还能提高深层特征的定位信息。当两种均使用时效果最好，平均准确率达到了 84.57%。

表 4　FPN 与特征融合对检测效果的影响

方法	bird	boat	cat	cow	mAP
方案一	89.98	70.01	89.03	78.70	81.93
方案二	91.22	70.83	91.85	80.44	83.59
方案三	90.13	70.83	90.21	79.34	82.89
方案四	92.17	71.22	93.30	81.60	84.57

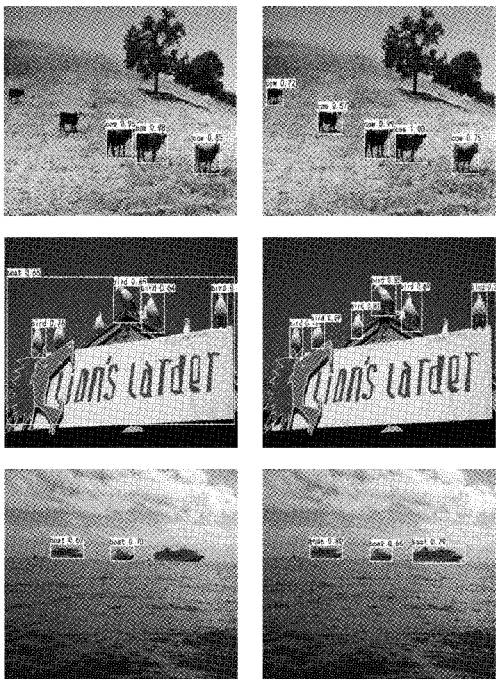
3)为验证通道注意力机制对本文模型的有效性，根据是否添加通道注意力机制设计对比实验，实验结果如表 5 所示(表中 att 表示注意力机制，表中 SSD 的特征提取器是基于 DCN 的 ResNet50)。由表 5 可以看出，添加注意机制的模型具有更好的检测效果，平均准确率为 83.28%，得益于注意机制能够提高重要信息的比重，体现了本文算法添加通道注意机制的有效性。

表 5　通道注意力机制对检测效果的影响

方法	bird	boat	cat	cow	mAP
SSD	87.29	68.58	91.87	79.82	81.89
SSD+att	89.53	70.08	92.05	81.44	83.28

2.5　定性结果分析

为了更直观体现本文方法的检测结果，图 6 可视化了当输入图像分辨率为 600×600 时，传统 SSD 目标检测算法与本文算法在 VOC2007 数据集上的测试结果。对比图 6 展示的效果可看出，传统 SSD 对分辨率相对较低及小目标进行检测效果较差，而本文改进的算法不但能够检测到更多的小尺寸的牛、鸟和船，降低小目标的漏检率，而且提高了对小目标检测的准确率，验证了本文算法对小目标检测的有效性。



(a) 传统SSD　　(b) 本文算法

图 6　传统 SSD 与本文算法检测结果对比

3　结　　论

针对小目标检测精度较低问题,提出一种轻量型 FPN 和通道注意机制 SSD 目标检测方法。该方法采用添加了可变形卷积的 ResNet50 作为特征提取器,有效提高了不同尺度目标的检测精度。在此基础上采用 FPN+通道注意力机制,丰富了浅层特征图的语义信息,提高了小目标检测的精确率,降低了漏检率。通过消融实验逐步验证了本文算法改进之处的有效性,并且与 DSOD、FSSD、DSSD 3 种基于 SSD 改进的算法相比,本文算法在特征提取与特征融合方面改进之后的检测精确率更高,相比传统的 SSD 算法,本文算法的 mAP 值提高了 8.7%。

参考文献

[1] XIAO X, WANG B, MIAO L, et al. Infrared and visible image object detection via focused feature enhancement and cascaded semantic extension [J]. Remote Sensing, 2021, 13(13): 2538, DOI: 10.3390/rs13132538.

[2] 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, 32(9): 970-978.

[3] GIRSHICK R, DONAHUE J, DARRELL T, et al. Region-based convolutional networks for accurate object detec-tion and segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 38(1): 142-158.

[4] MALINDA V, DEUKHEE L. Intervertebral disc instance segmentation using a multistage optimization mask-RCNN (MOM-RCNN) [J]. Journal of Computational Design and Engineering, 2021, 4(4): 1023-1036.

[5] WANG X, LU H. Application of faster R-CNN algorithm in weld position recognition [J]. International Core Journal of Engineering, 2021, 7(6): 408-416.

[6] 李厚杰, 王法胜, 贺建军, 等. 基于伪样本正则化 Faster R-CNN 的交通标志检测[J]. 吉林大学学报(工学版), 2021, 51(4): 1251-1260.

[7] 王标,周雅兰,王永红. 改进型 Faster R-CNN 网络在电子元件 LED 气泡缺陷检测中的应用[J]. 电子测量与仪器学报,2021,35(9):136-143.

[8] NI H X, WANG M Z, ZHAO M Z. An improved Faster R-CNN for defect recognition of key components of transmission line [J]. Mathematical Biosciences and Engineering: MBE, 2021, 18(4): 4679-4695.

[9] 彭继慎,孙礼鑫,王凯,等. 基于模型压缩的 ED-YOLO 电力巡检无人机避障目标检测算法[J]. 仪器仪表学报,2021,42(10):161-170.

[10] QING Y H, LIU W Y, FENG L Y, et al. Improved YOLO network for free-angle remote sensing target detection[J]. Remote Sensing, 2021, 13(11): 2171, DOI:10.3390/rs13112171.

[11] LIU M J, WANG X H, ZHOU A J, et al. UAV-YOLO: Small object detection on unmanned aerial vehicle perspective[J]. Sensors, 2020, 20(8): 2238, DOI:10.3390/s20082238.

[12] 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.

[13] DING F L, ZHUANG Z L, LIU Y, et al. Detecting Defects on Solid Wood Panels Based on an Improved SSD Algorithm [J]. Sensors (Basel, Switzerland), 2020, 20(18): 5315, DOI:10.3390/s20185315.

[14] YOU S, BI Q, JI Y M, et al. Traffic sign detection method based on improved SSD [J]. Information, 2020, 11(10): 475.

[15] SUN P, ZHAO Y, ZHU S. An approach to improve SSD through mask prediction of multi-scale feature maps [J]. Pattern Analysis and Applications, 2021, 24(3): 1357-1366.

[16] ZHANG X L, XIE H, ZHAO Y J, et al. A fast SSD model based on parameter reduction and dilated convolution [J]. Journal of Real-Time Image Processing, 2021, 18(6): 1-14.

[17] CHOI H T, LEE H J, KANG H, et al. SSD-EMB:

An Improved SSD Using Enhanced Feature Map Block for Object Detection[J]. Sensors, 2021, 21(8): 2842.

[18] WANG W J, HE M L, WANG X H, et al. Sewing gesture image detection method based on improved SSD model[J]. Electronics Letters, 2021, 57(8): 321-323.

[19] LIAO K Y, FAN B, ZHENG Y L, et al. Bow image retrieval method based on SSD target detection[J]. IET Image Processing, 2020, 14(17): 4441-4449.

[20] HE W M, WU Y, XIAO J, et al. MGFPN: Enhancing multi-scale feature for object detection[J]. Journal of Intelligent & Fuzzy Systems, 2021, 40(6): 11171-11181.

[21] GURITA A, MOCANU I G. Image Segmentation Using Encoder-Decoder with Deformable Convolutions [J]. Sensors, 2021, 21(5): 1570.

作者简介

蒋晨，硕士研究生，主要研究方向为计算机视觉与图像处理。

E-mail: 1261705105@qq.com

钱永明(通信作者)，教授，硕士研究生导师，主要研究方向为图像识别与数据处理。

E-mail: qian.ym@ntu.edu.cn

姚兴田，教授，硕士研究生导师，主要研究方向为目标检测。

E-mail: yao.xt@ntu.edu.cn

李壮，硕士研究生，主要研究方向为图像处理。

E-mail: 2390491438@qq.com