

基于 Bi-SSD 的小目标检测算法^①



汪 能, 胡君红, 刘瑞康, 范良辰

(华中师范大学 物理科学与技术学院, 武汉 430079)

通讯作者: 胡君红, E-mail: 386393335@qq.com

摘 要: 针对目前目标检测技术中小目标检测困难问题, 提出了一种基于 SSD (Single Shot multibox Detector) 改进的小目标检测算法 Bi-SSD (Bi-directional Single Shot multibox Detector). 该算法为 SSD 的浅层特征设计了小目标特征提升模块, 在网络的分类和回归部分结合多尺度特征融合方法和 BiFPN (Bi-directional Feature Pyramid Network) 结构, 设计了 6 尺度 BiFPN 分类回归子网络. 实验结果表明, 在 PASCAL VOC 和 MS COCO 目标检测数据集上 Bi-SSD 相比原始的 SSD 算法有更好的检测性能. 其中 VOC2007+2012 上 Bi-SSD 算法的 mAP 指标达到了 78.47% 相较 SSD 算法提升了 1.34%, 在 COCO2017 上 Bi-SSD 算法的 mAP 达到 26.4% 提升了接近 2.4%.

关键词: 小目标检测; 特征融合; 多尺度检测; 深度学习; 卷积神经网络

引用格式: 汪能, 胡君红, 刘瑞康, 范良辰. 基于 Bi-SSD 的小目标检测算法. 计算机系统应用, 2020, 29(11): 139-144. <http://www.c-s-a.org.cn/1003-3254/7607.html>

Small Target Detection Algorithm Based on Bi-SSD

WANG Neng, HU Jun-Hong, LIU Rui-Kang, FAN Liang-Chen

(College of Physical Science and Technology, Central China Normal University, Wuhan 430079, China)

Abstract: Aiming at the difficulty of small target detection in current target detection technology, a small target detection algorithm named improved Bi-directional Single Shot multibox Detector (Bi-SSD) based on Single Shot multibox Detector (SSD) is proposed. This algorithm designed a small object feature improvement module for the shallow features of SSD. In the classification and regression parts of the network, a 6-scale Bi-directional Feature Pyramid Network (BiFPN) is designed as classification and regression sub-network according to multi-scale feature fusion method and BiFPN structure. Experimental results show that Bi-SSD has better detection performance than the original SSD on PASCAL VOC and MS COCO object detection datasets. On VOC2007+2012, Bi-SSD achieves 78.47% mAP, which is an increase of 1.34% compared to the original SSD algorithm. On COCO2017, Bi-SSD achieves 26.4% mAP, which was an increase of 2.4% compared to the original SSD algorithm.

Key words: small object detection; feature fusion; multi-scale detection; deep learning; Convolutional Neural Network (CNN)

1 引言

目标检测技术是计算机视觉领域的重要研究内容和应用领域^[1]. 在行人检测^[2]、车辆检测^[3]和遥感图像检测^[4]等诸多应用场景中, 目标检测技术扮演着非常

重要的作用. 目前目标检测算法对大目标的检测效果显著优于小目标, 然而实际场景目标尺寸差异很大, 针对小目标的目标检测技术的研究一直是计算机视觉领域的研究热点和难点^[1]. 近年基于深度神经网络的目标

① 收稿时间: 2020-02-28; 修改时间: 2020-03-17; 采用时间: 2020-03-27; csa 在线出版时间: 2020-10-29

检测算法的研究取得了较大的进展,但由于神经网络主要依赖深层特征进行目标检测,而对浅层特征利用不足,因此对小目标的检测有很大的困难.本文主要研究提升浅层特征的利用效率的方法,以此提升网络对小目标的检测性能.

深度学习目标检测框架大多是通用目标检测框架,并没有针对小目标进行优化设计,所以对小目标的检测效果并不理想.为解决小目标检测问题,近年有学者提出了不同的优化方案.Chen等^[5]在2016年提出了针对小目标的多区域语义信息目标检测器来优化对小目标的检测性能.Sommer等^[6]在2017年提出了一种由4个卷积层和3个全连接层构成的浅网络检测模型,对航空图像中的小目标进行检测.Li等^[7]在2017年提出将生成对抗网络应用于图像识别领域,通过挖掘不同尺度物体间的结构关联,提高小物体的特征表示,使之与大物体类似,帮助小目标检测接近大目标检测的水平.Lin等^[8]在2017年提出了特征金字塔的概念,通过网络的多尺度特征信息提升网络对小目标的检测性能.Singh和Bharat等^[9]在2018年提出了一种尺度归一化的目标检测图像金字塔,利用图像的尺度缩放来解决小目标的检测问题.

在以上优秀的研究成果的基础上,本文基于单阶段多目标检测器SSD(Single Shot multibox Detector)^[10]结构提出了针对小目标的改进模型,双向单阶段多目标检测器Bi-SSD(Bi-directional Single Shot multibox Detector).Bi-SSD模型在目标检测公共数据集PASCAL VOC2007 test和MS COCO2017 test上分别对原始的SSD算法取得了1.34%和2.4%的性能提升.本文的主要工作包括:为浅层特征输出设置了小目标检测专用模块,增强了网络对浅层输出的特征提取能力;参考双向特征金字塔网络BiFPN(Bi-directional Feature Pyramid Network)^[11]结构优化了SSD分类和回归部分的网络结构,设计了6尺度双向特征融合模块,使多尺度特征进行两次特征融合,分别增加了浅层特征的语义信息和深层网络的细节信息.

2 SSD算法分析

SSD网络是一种端到端的目标检测模型,以VGG16^[12]为基础网络,网络结构如图1所示.VGG16网络可分为7部分,前5部分Conv1~Conv5主要由卷积层构成,最后两个部分FC6~FC7为全连接层.

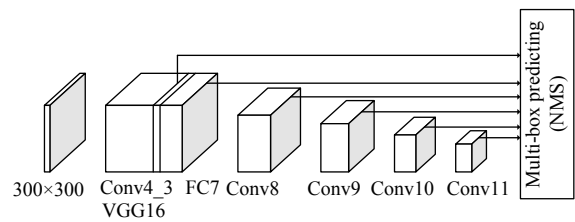


图1 SSD网络结构图

为实现全卷积结构,SSD模型将VGG16中的FC6和FC7替换为卷积层,并且移除了所有池化层和fc8层.算法设置了6个特征输出层,分别是Conv4_3, FC7, Conv8_2, Conv9_2, Conv10_2, Conv11_2,这些特征输出层各与一个输出卷积层相连.SSD模型根据每个特征输出层输出的特征图尺寸 N ,将图片划分为 $N \times N$ 个方格,以每个方格的中心点为中心,使每个方格与一定数量的大小固定的方框匹配,这些框称为候选框,其中包括两个正方形的候选框和两个矩形候选框,所以对于每个方格,SSD模型都会为其生成不同尺寸和宽高比的候选框,确保不同尺寸的目标均可以被候选框覆盖.

SSD算法为每个特征输出层设置了一个候选框尺寸缩放因子 s_k ,计算公式如式(1)所示,其中 s_{\min} 为0.2, s_{\max} 为0.9,第 k 层特征图的候选框尺寸与 s_k 成正比,所以浅层特征图对应的候选框尺寸较小,深层特征图对应的候选框尺寸较大.根据SSD候选框的匹配机制,较小的候选框通常与较小的真实框匹配,用于检测小目标,所以SSD模型中负责检测小目标的输出层主要是Conv4_3,特征输出层中其他深层特征图主要负责检测较大的目标.

$$s_k = s_{\min} + \frac{s_{\max} - s_{\min}}{m - 1}(k - 1) \quad (1)$$

作为小尺寸目标检测的主要来源,Conv4_3所处的卷积深度较浅,特征提取和表达能力有限,而较深的特征输出层又主要负责检测较大尺寸的目标,因而导致SSD网络对于小目标的检测性能较差.

3 SSD网络的改进

基于以上对SSD网络的分析,SSD算法的主要问题在于基础网络的特征输出层对于小目标的特征提取能力不足,导致算法对小目标的检测能力较差,本文将从以下几个方面对SSD算法进行优化:

在进行分类和回归网络前,针对浅层特征输出层,

设置了一个小目标特征优化模块,改善网络对小目标特征提取能力有限的问题。

改进分类和回归子网络,从现有的各种多尺度检测解决方案中探索多尺度检测在面对小目标时的更优解决方案,利用多尺度特征融合方案提升小目标检测性能。

修改后的网络结构图如图2所示。

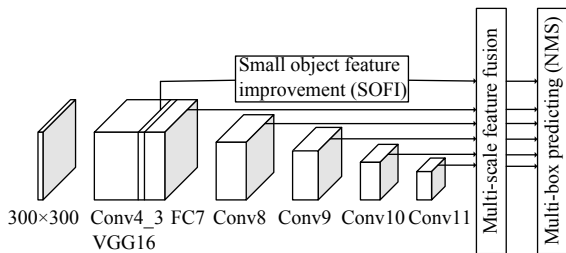


图2 改进后的 SSD 网络结构图

3.1 小目标特征优化模块

SOFI 模块结构如图3所示,由5个卷积层堆叠而成,卷积核尺寸分别为 1×1 和 3×3 ,其中 3×3 卷积核在模块中主要作用为特征提取^[12],确保 SOFI 模块能提高网络的特征提取能力, 1×1 卷积核的主要作用有两个:降维和跨通道信息交互^[13]。因为 1×1 卷积本身参数很少,通过降维可以降低模块的整体参数数量,而 1×1 通过通道间的线性组合增加了通道间的信息交互能力,使提取的特征能合理分配到各个通道中。

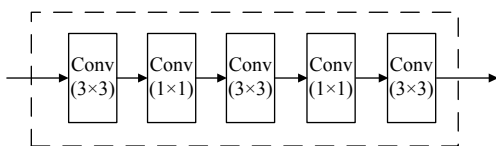


图3 小目标特征优化模块 SOFI

3.2 多尺度 BiFPN 结构分类回归网络

SSD 的多尺度特征直接与输出层连接,导致浅层特征提取能力不足,特征金字塔网络 FPN (Feature Pyramid Networks)^[8] 结构在输出层之前,增加了深层特征向浅层特征进行特征融合的结构,增加了浅层特征图中包含的语义信息,提高了浅层输出的检测能力。

本文利用 FPN 结构为 SSD 模型设计了特征融合模块,并基于 BiFPN 结构对特征融合模块进行了优化,如图4所示。BiFPN 对多尺度的输出要经过两次卷积操作,第一次卷积操作将深层特征向浅层特征整合,除

了最深层的特征直接向浅层整合外,其他的多尺度特征均经过卷积后向浅层整合,第二次卷积操作将浅层经过处理后的特征再向深层特征融合,实现了特征在多个尺度的输出间的传递,所有的特征融合均通过特征图间的加法运算实现。文献[11]比较了现有的不同特征融合方案,验证了这种结构在计算量和特征融合效果上取得了最优的平衡。

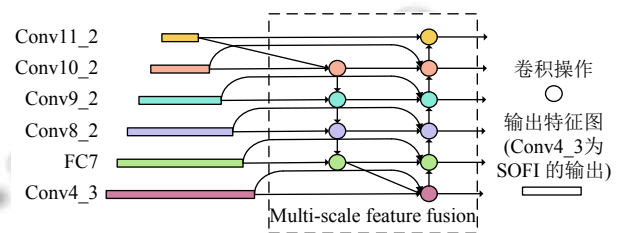


图4 Bi-SSD 分类回归网络结构图

4 实验验证与结果分析

4.1 实验数据集与训练参数设置

为了客观评估 Bi-SSD 算法的性能,本文分别在 PASCAL VOC 2007+2012 和 MS COCO 2017 数据集上对算法进行了训练和评估。VOC 2007+2012 数据集包含常见的 20 个分类,共计 21503 张图片,其中训练集为 16551 张,测试集为 4952 张;COCO 2017 数据集将样本分为 80 个类别,其中训练集的共包括 338 002 张图片,测试集共包括 14408 张图片,COCO 数据集中所有样本被粗略分为大,中,小 3 个尺寸,方便评价目标检测算法对不同尺寸目标的检测能力。

在模型训练过程中,本文将输入图片尺寸设置为 300, batch size 为 32,初始学习率为 10^{-3} ,优化算法选择随机梯度下降法 SGD (Stochastic Gradient Descent)。在 VOC 数据集上训练时,训练集为 VOC 2007 train 和 VOC 2012 train,测试集为 VOC 2007 test,迭代次数为 120 000 次,在前 50 000 次迭代过程中,学习率保持不变,从第 50 000 次迭代开始,每迭代 10 000 次,学习率减半。在 COCO 数据集上训练时,训练集为 COCO 2017 train,测试集为 COCO 2017 val,迭代次数为 400 000 次,在前 120 000 次迭代过程中,学习率保持不变,从第 120 000 次迭代开始,每迭代 40 000 次,学习率减半。

4.2 实验结果分析

SSD 模型和 Bi-SSD 模型在 VOC 数据集上的训

训练损失曲线如图5所示,图中包括SSD和Bi-SSD的损失曲线,3幅图从左到右分别为分类损失,回归损失和总体损失.从图中可知训练过程中,Bi-SSD模

型的分类损失和回归损失均小于SSD模型,以上现象说明了Bi-SSD模型相比SSD模型更少出现误检和漏检,有更好的识别精度和鲁棒性.

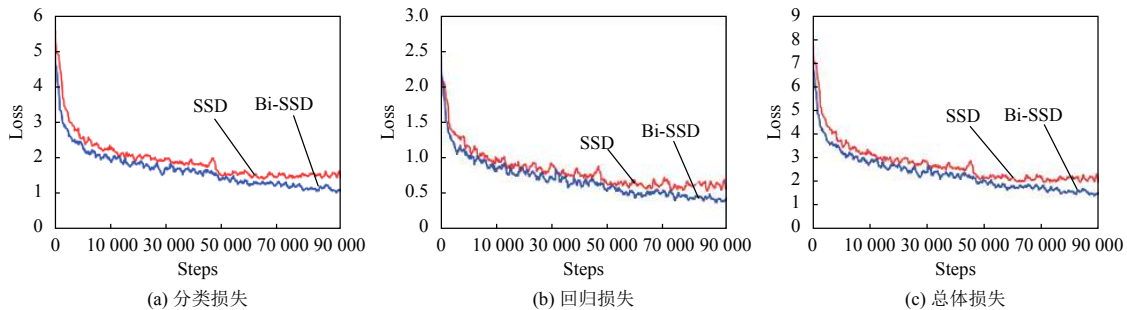


图5 VOC数据集训练损失曲线

SSD、SSD+SOFI、SSD+BiFPN和Bi-SSD模型在VOC数据集上的实验结果如表1所示.

表1 PASCAL VOC2007 test数据集上的检测结果

模型	mAP	aero-plane	bicycle	bird	boat	bottle	bus
SSD	0.7713	0.7991	0.8401	0.7613	0.7112	0.505	0.8426
SSD+SOFI	0.7760	0.8203	0.8429	0.7714	0.7096	0.5107	0.8438
SSD+BiFPN	0.7835	0.8150	0.8438	0.7571	0.7066	0.5441	0.8625
Bi-SSD	0.7847	0.8105	0.855	0.7725	0.7206	0.5397	0.8527
模型	car	cat	chair	cow	dining-table	dog	horse
SSD	0.8624	0.8775	0.5962	0.8274	0.7636	0.8563	0.8606
SSD+SOFI	0.8632	0.8758	0.6219	0.8479	0.7715	0.8538	0.8622
SSD+BiFPN	0.8630	0.8794	0.6207	0.8554	0.8037	0.8657	0.8810
Bi-SSD	0.8666	0.8851	0.6091	0.8604	0.7747	0.8684	0.8711
模型	motor-bike	person	potted-plant	sheep	sofa	train	tvmonitor
SSD	0.8366	0.7865	0.5069	0.7716	0.7881	0.8674	0.765
SSD+SOFI	0.8315	0.7870	0.4929	0.7805	0.8005	0.8633	0.7700
SSD+BiFPN	0.8459	0.7936	0.5087	0.7877	0.7983	0.8492	0.7878
Bi-SSD	0.855	0.7994	0.5198	0.8002	0.7949	0.8658	0.7724

表1中SSD+SOFI和SSD+BiFPN分别为仅添加SOFI模块和仅添加BiFPN模块的SSD模型.与SSD模型相比,SSD+SOFI、SSD+BiFPN、Bi-SSD模型在测试集VOC 2017 test的检测性能均有所提升,所有类别的均值平均精度mAP(mean Average Precision)提高了分别提高了0.47%、1.22%和1.34%,且在体积较小的目标类别如bird, bottle等的分类样本中,检测精度均有提升.

以上结果说明了SOFI模块和BiFPN模块对小目标检测的有效性.与SSD模型相比,同时具有SOFI模块和BiFPN模块的Bi-SSD模型对小体积目标的检测

能力有最大的提升.

图6是SSD模型和Bi-SSD模型在PASCAL VOC数据集上训练后的检测结果对比,为使检测结果更清晰,仅显示了对“person”类别目标的检测结果,3行图片从上到下分别是原图,SSD模型检测结果和Bi-SSD模型检测结果.测试图片中目标均为小目标,很明显Bi-SSD模型在小目标场景下,检测到的正样本更多,具有更高的查准率和查全率.实验结果证明Bi-SSD对小目标的检测性能确实优于SSD模型,由于选取的测试图片取复杂自现复杂实街景,也证明了Bi-SSD模型鲁棒性高.

SSD、SSD+SOFI、SSD+BiFPN和Bi-SSD模型在COCO数据集上的实验结果如表2所示,其中mAP@0.5和mAP@0.75分别为模型在正样本交并比(IoU)阈值分别为0.5和0.75时模型的均值平均精度,mAP@0.5:0.95为正样本IoU阈值分别为0.5至0.95,以0.05为间隔时所有均值平均精度的平均值.在mAP@0.5, mAP@0.75和mAP@0.5:0.95 3个指标上,Bi-SSD的得分均高于SSD.比较不同尺寸目标的检测结果,Bi-SSD在小、中、大3种尺寸的检测样本中检测平均精度AP分别为0.084, 0.288和0.43,相比SSD分别提高了2.4%, 3%和3.1%.对于另外一个指标平均召回率AR,Bi-SSD算法在小,中,大3种尺寸上的检测结果分别为0.129, 0.417, 0.569,与SSD相比分别提升了2.9%, 2.9%和2.1%.SSD在仅添加SOFI模块的情况下,对小目标的平均精度和平均召回率分别提升了0.9%和8%,在仅添加BiFPN模块的情况下,各尺度的平均精度和平均召回率均有所提升.



图6 SSD与Bi-SSD检测结果对比

表2 COCO 2017 test 数据集上的检测结果

模型	Avg.Precision, IoU			Avg.Precision, Area			Avg.Recall, Area		
	0.5:0.95	0.5	0.75	S	M	L	S	M	L
SSD	0.24	0.417	0.245	0.06	0.258	0.399	0.1	0.388	0.548
SSD+SOFI	0.24	0.419	0.244	0.069	0.254	0.396	0.108	0.379	0.537
SSD+BiFPN	0.263	0.450	0.270	0.083	0.288	0.434	0.128	0.417	0.576
Bi-SSD	0.264	0.451	0.274	0.092	0.288	0.43	0.132	0.419	0.574

从 $mAP@0.5$, $mAP@0.75$ 和 $mAP@0.5:0.95$ 3 个指标的检测结果中可以看出 Bi-SSD 模型和 SSD+BiFPN 模型的检测性能上全面优于 SSD 模型, SSD+SOFI 模型在小目标检测性能上优于 SSD 模型. 对不同尺寸检测样本的平均精度和平均召回率说明了 BiFPN 模块使 SSD 模型对不同尺寸目标的检测能力均有所提升, 其中对小目标的检测结果则说明 SOFI 模块使 SSD 模型更容易从背景中检测出小目标, 且对小目标的检测更为准确. 以上分析说明 BiFPN 模块和 SOFI 模块的有效性, 与 SSD 模型相比, 同时具备 SOFI 模块和 BiFPN 模块的 Bi-SSD 模型对小目标的检测能力更强.

综合以上 2 个实验的结果可知, Bi-SSD 算法相比 SSD 算法, 检测性能从整体上有所提升, 并且对于小目标, 不论是精确率还是召回率, Bi-SSD 算法均有较大的提升. 以上结果说明了, 本文以 SSD 为基础提出的 Bi-SSD 算法确实优化了模型对小目标检测能力.

5 结论与展望

本文以 SSD 模型为基础, 设计了一种新型的 One Stage 目标检测模型 Bi-SSD. 本文的关键思想是为提升小目标特征提取能力设计的 SOFI 模块和参考 BiFPN 结构设计的特征融合模块, 这两个模块分别从特征提取能力和多尺度特征融合两个角度优化了模型的小目标检测能力, 提升了模型在实际应用场景中使用价值. 在保留 SSD 算法的基础网络 VGG-16 的基础上通过添加上面两个模块, 在常用目标检测数据集 PASCAL VOC 和 MS COCO 上均取得了性能的提升, 在小目标上的提升尤为明显. 若使用更优秀的基础网络来进行特征提取, 模型的性能应该还有提升的余地, 而且 SOFI 模块和特征融合模块除了本文提出的结构外, 还可以根据不同的基础网络尝试更多的结构. 如何选择合适的基础网络和与之匹配的 SOFI 模块和特征融合模块将是以后的工作方向.

参考文献

- 1 Uijlings JRR, Van De Sande KEA, Gevers T, *et al.* Selective search for object recognition. *International Journal of Computer Vision*, 2013, 104(2): 154–171. [doi: [10.1007/s11263-013-0620-5](https://doi.org/10.1007/s11263-013-0620-5)]
- 2 曾接贤, 方琦, 符祥, 等. 融合多层特征的多尺度行人检测. *中国图象图形学报*, 2019, 24(10): 1683–1691. [doi: [10.11834/jig.190009](https://doi.org/10.11834/jig.190009)]
- 3 郑冬, 李向群, 许新征. 基于轻量化 SSD 的车辆及行人检测网络. *自然科学版*, 2019, 42(1): 73–81.
- 4 郭智, 宋萍, 张义, 等. 基于深度卷积神经网络的遥感图像飞机目标检测方法. *电子与信息学报*, 2018, 40(11): 2684–2690.
- 5 Chen CY, Liu MY, Tuzel O, *et al.* R-CNN for small object detection. *Proceedings of the 13th Asian Conference on Computer Vision*. Taipei, China. 2017. 214–230.
- 6 Sommer LW, Schuchert T, Beyerer J. Deep learning based multi-category object detection in aerial images. *Proceedings of SPIE Automatic Target Recognition XXVII*. Anaheim, CA, USA. 2017. 1020209.
- 7 Li JN, Liang XD, Wei YC, *et al.* Perceptual generative adversarial networks for small object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI, USA. 2017. 1951–1959.
- 8 Lin TY, Dollár P, Girshick R, *et al.* Feature pyramid networks for object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI, USA. 2017. 936–944.
- 9 Singh B, Davis LS. An analysis of scale invariance in object detection-SNIP. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT, USA. 2018. 3578–3587.
- 10 Liu W, Anguelov D, Erhan D, *et al.* SSD: Single shot MultiBox detector. *Proceedings of the 14th European Conference on Computer Vision*. Amsterdam, the Netherlands. 2016. 21–37.
- 11 Tan MX, Pang RM, Le QV. Efficientdet: Scalable and efficient object detection. *arXiv preprint arXiv: 1911.09070*, 2019.
- 12 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv: 1409.1556*, 2014.
- 13 Lin M, Chen Q, Yan SC. Network in network. *arXiv preprint arXiv: 1312.4400*, 2013.