



一种改进的BR-YOLOv3目标检测网络

宦海,陈逸飞,张琳,李鹏程,朱蓉蓉

(南京信息工程大学 电子与信息工程学院,南京 210044)

摘要:在目标检测任务中不同目标间尺寸差异较大,导致多尺寸目标难以被有效检测。基于YOLOv3提出BR-YOLOv3目标检测网络。利用空洞卷积提升网络层感受野尺寸的特性,使用不同数量、尺寸、膨胀率的卷积构建多层并行的空洞感受野模块。通过双向特征金字塔结构实现浅深层特征的双向融合,提升浅层预测分支分类、深层预测分支目标定位能力。使用 $LOSS_{GIOU}$ 定位损失函数实现目标回归过程整体化,从而降低目标漏检率。实验结果表明,BR-YOLOv3目标检测网络在Pascal VOC测试集上的测试平均精度均值达到79.24%,相比原网络提升3.52个百分点,且在检测精度上优于SSD、Faster RCNN等主流目标检测网络。

关键词:目标检测;目标尺寸差异;空洞感受野模块;双向特征金字塔;定位损失函数

开放科学(资源服务)标志码(OSID):



中文引用格式:宦海,陈逸飞,张琳,等.一种改进的BR-YOLOv3目标检测网络[J].计算机工程,2021,47(10):186-193.

英文引用格式:HUAN H, CHEN Y F, ZHANG L, et al. An improved BR-YOLOv3 object detection network[J]. Computer Engineering, 2021, 47(10): 186-193.

An Improved BR-YOLOv3 Object Detection Network

HUAN Hai, CHEN Yifei, ZHANG Lin, LI Pengcheng, ZHU Rongrong

(School of Electronic and Information Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China)

[Abstract] In the object detection task, there is a large size difference between different objects, which makes it difficult to effectively detect object with multiple size. Based on YOLOv3, the Bidirectional FPN Atrous Reception YOLOv3 (BR-YOLOv3) target detection network is proposed. Using atrous convolution can effectively improve the receptive field size of the network layer, using different numbers, convolution kernel size, and dilation rate convolution to build a multi-layer parallel Atrous Receptive Module (ARM), and by using Bidirectional Feature Pyramid Structure Network (BiFPN) realizes bidirectional fusion of shallow and deep features, improving the classified ability of shallow prediction branch, and enhancing the ability of deep prediction branch's target positioning. By using the $LOSS_{GIOU}$ positioning loss function, the target regression process is integrated, and the target miss rate is reduced. Experimental results show that the improved RB-Yolov3 on the Pascal VOC test set has a mean average precision of 79.24%, which is an increase of 4.65% on the basis of the original network. It is superior to mainstream target detection networks such as SSD and Faster RCNN in detection accuracy.

[Key words] object detection; object size difference; atrous receptive field module; Bidirectional Feature Pyramid Network (BiFPN); location loss function

DOI: 10.19678/j.issn.1000-3428.0059234

0 概述

目标检测作为计算机视觉领域的主要研究方向,长期以来受到广泛关注,相关技术已被应用于医学图像检测、火灾检测、汽车无人驾驶等领域。

基于深度学习的目标检测是近年来研究的热点,国内外研究人员针对目标检测网络中存在的不足进行改进。针对网络训练效率低的问题,HE等^[1]通过残差网络解决了深层网络难以收敛的问题。LOFFE等^[2]提出批归一化(Batch Normalization)解决网络训练过程中梯度消失的问题。REDMON等^[3]等提出目标预测与检测过程一体化,大幅缩短

基金项目:国家自然科学基金(41671345)。

作者简介:宦海(1978—),男,副教授,主研方向为通信与信息处理、图像处理;陈逸飞,硕士研究生;张琳,本科生;李鹏程、朱蓉蓉,硕士研究生。

收稿日期:2020-08-01 修回日期:2020-09-10 E-mail:641374881@qq.com

了网络检测时间。针对待检测目标尺寸差异大、难以有效检测的问题,LIU等^[4-5]通过多尺度特征信息的融合,提升了网络对多尺度目标的检测能力。SIEGEDY等^[6-7]使用Inception结构及改进金字塔结构,提升网络对尺度信息获取。DAI等^[8]通过空洞卷积思想构建可变形卷积,提升了网络对多尺寸目标的检测能力。针对网络检测时存在过多冗余信息的问题,MA等^[9]使用注意力机制,实现目标特征权重再分配,减少了冗余信息的干扰。ZHU等^[10]通过Anchor-free实现对目标位置及大小的预测,通过减少网络锚框产生的数量以减少冗余信息。针对网络目标定位能力弱的问题,XIAO等^[11]通过对多尺度显著图进行融合以提升对显著目标的检测完整性。ZHENG等^[12-13]通过引入目标区域惩罚项提升目标定位能力。QING等^[14]使用层级偏移,提升目标预测区域的定位以及检测精度。针对网络中浅层目标特征信息难以实现目标的分类问题。LI等^[15-17]通过使用空洞卷积提升网络感受野的大小,丰富网络中特征语义信息的同时提高了网络的目标识别能力。ZHU等^[18]采用扩张卷积的策略提升对浅层特征的提取。但鲜有文献针对目标检测网络难以检测多尺寸目标的问题进行研究。

本文针对经典网络YOLOv3在目标定位以及识别问题上的不足,使用 $\text{Loss}_{\text{GIOU}}$ 作为新的定位损失函数,实现目标预测区域回归过程的整体化以降低目标区域漏检的概率,将空洞感受野模块与双向特征

金字塔模块联结使用,增强各预测分支输出特征的语义强度,并通过浅深层特征的双向融合,提升整网络的对多尺寸目标的定位与分类能力。

1 YOLO v3 网络

1.1 网络结构

如图1(a)所示,YOLOv3^[12]共包含3个模块,分别为Darknet53特征提取模块、特征金字塔模(Feature Pyramid Network, FPN)及预测分支模块。图1(b)为特征提取模块Darknet53的基本网络结构,此模块由5个残差结构组成,每个残差结构将输入特征尺寸压缩至原尺寸的1/2。以输入图像大小为 $416 \times 416 \times 3$ 为例,残差模块1、2、3、4、5的输出特征图谱大小分别为 $208 \times 208 \times 64$, $104 \times 104 \times 128$, $52 \times 52 \times 256$, $26 \times 26 \times 512$, $13 \times 13 \times 1024$ 。其中残差模块3、4、5的输出特征图谱将作为下一模块的输入。图1(c)为FPN模块的结构图。图中DBL结构由卷积层、批归一化层、激活层串联组成。FPN模块包含3个预测分支结构,依次为大、中、小尺寸目标的检测提供特征信息。模块通过自顶向下的特征流,将来自高层预测分支中包含强语义特征信息融入到浅层特征中,为浅层预测分支提供更强语义特征信息。图1(d)为预测分支结构,以FPN结构生成的包含多尺度特征信息的融合特征1、2、3为特征,通过DBL结构与 1×1 大小的卷积层,产生模型的最终特征输出Output1、Output2、Output3。

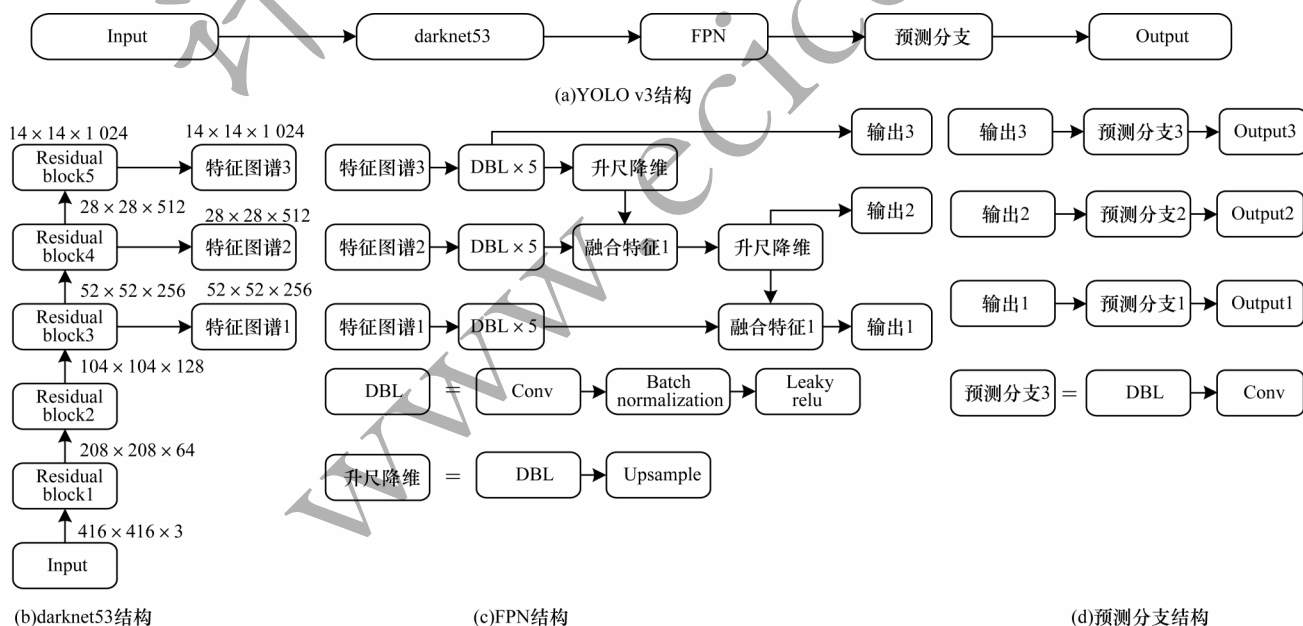


图1 YOLOv3网络结构

Fig.1 Structure of YOLOv3 network

1.2 损失函数

Yolov3网络损失函数由目标定位损失 $L_{\text{Loss}}^{\text{coor}}$,置信度损失 $L_{\text{Loss}}^{\text{conf}}$,以及目标分类损失 $L_{\text{Loss}}^{\text{cls}}$ 3个部分组成,

如式(1)所示:

$$L_{\text{Loss}} = L_{\text{Loss}}^{\text{coor}} + L_{\text{Loss}}^{\text{conf}} + L_{\text{Loss}}^{\text{cls}} \quad (1)$$

目标定位损失 $L_{\text{Loss}}^{\text{coor}}$ 以均方误差(MSE)作为损失

函数的目标函数。各预测区域与对应真实区域间的IoU值的计算公式如式(2)所示:

$$I_{IoU} = \frac{|\text{Area}(A) \cap \text{Area}(B)|}{|\text{Area}(A) \cup \text{Area}(B)|} \quad (2)$$

其中: $\text{Area}(A)$ 表示目标真实区域面积; $\text{Area}(B)$ 表示预测区域面积; \cap 表示交集; \cup 表示并集。通过真实区域与预测区域交集面积比上真实区域与预测区域并集面积, 获得2区域交并比(Intersection of Union, IoU)的值。通过预先设定好的IoU阈值对预测区域进行筛选, 筛选出IoU值大于阈值的区域。

对式(2)筛选出的区域, 计算其对应 $L_{\text{Loss}}^{\text{coor}}$ 的值, 函数表达式如式(3)所示:

$$L_{\text{Loss}}^{\text{coor}} = \sum_{i=1}^n \sum_{j=1}^3 \text{mask} \times [((x_p)_j - (x_t)_j)^2 + ((y_p)_j - (y_t)_j)^2 + ((w_p)_j - (w_t)_j)^2 + ((h_p)_j - (h_t)_j)^2] \quad (3)$$

其中: mask 表示预测框中包含目标的概率; x, y, w, h 依次为目标预测区域中心点横坐标、纵坐标、以及区域的宽和高的值; i 表示第 i 个预测框; n 表示真实框所对应预测框的总个数; j 表示对应预测分支号; 下标 p 表示该值为预测框的值; 下标 t 表示该值为目标预测框的值。

2 BR-YOLOv3 网络

针对YOLOv3在多尺寸目标检测精度较差的问题, 使用 $\text{Loss}_{\text{GloU}}$ 函数替换原有目标定位损失函数, 通过目标定位过程一体化以提升目标定位精度。通过在网络中添加空洞感受野模块, 提升网络层感受野大小, 从而增强特征语义强度。改进FPN为双向特征金字塔模块, 通过将高层特征与浅层特征实现双向融合, 最终构建BR-YOLOv3目标检测模型。

2.1 位置损失函数的改进

针对YOLOv3目标定位损失函数对目标预测区域中心点的横纵坐标, 以及预测区域的宽高进行独立的偏移量计算, 割裂了目标预测区域的整体性。使用 $\text{Loss}_{\text{GloU}}$ 作为目标定位损失函数, 通过将目标定位回归过程整体化以提升网络的目标定位能力。此处仍以式(2)筛选区域作为目标区域的产生形式。式(4)为惩罚项 σ_1 的计算公式:

$$\sigma_1 = \frac{|\text{Area}(C) / (\text{Area}(A) \cup \text{Area}(B))|}{|\text{Area}(C)|} \quad (4)$$

其中: $\text{Area}(C)$ 表示真实区域与预测区域外界矩形的面积; 符号 $/$ 表示从 $\text{Area}(C)$ 中排除 $\text{Area}(A) \cup \text{Area}(B)$ 的面积。

式5所示为GIoU的计算公式, 通过预测区域与真实区域间的IoU值减去惩罚项 σ_1 计算得出:

$$\text{GIoU} = \text{IoU} - \sigma_1 \quad (5)$$

$\text{Loss}_{\text{GloU}}$ 的计算公式如下:

$$\text{Loss}_{\text{GloU}} = \sigma_2 (1 - \text{GIoU}) \quad (6)$$

通过 $\sigma_2 (1 - \text{GIoU})$ 获得最终的预测区域定位损失值, 如式(7)所示:

$$\sigma_2 = 2 - \text{OB}_w \times \text{OB}_h / (\text{Image_size})^2 \quad (7)$$

其中: σ_2 为不同尺寸预测目标间的惩罚项; OB_w 为预测目标的宽度; OB_h 为预测目标的高度; Image_size 为输入图像尺寸。该惩罚项的使用有效改善了目标定位损失由较大预测区域所主导的弊端。

通过使用 $\text{Loss}_{\text{GloU}}$ 作为目标定位损失函数, 使预测区域的回归调整过程一体化, 从而满足目标区域整体性的要求。此外, 解决了预测区域回归调整过程中, 当出现调整后的预测区域与真实区域间的IoU为零时, 预测区域无法回归而导致的漏检问题。

如图2所示, 浅色框代表目标预测区域, 深色框代表目标真实区域, 图2中两个真实框为统一真实框。图2(a)中, 真实框与预测框的二范数 $\|\cdot\|_2$ 的值, 即定位损失值为10, 2框GIoU的计算值为17/56, $\text{Loss}_{\text{GloU}}$ 的值为39/56; 图2(b)中, 真实框与预测框的定位损失值同为10, 2框的GIoU值为7/18, $\text{Loss}_{\text{GloU}}$ 的值为11/18。图2中的例子对比显示, 通过使用 $\text{Loss}_{\text{GloU}}$ 作为损失函数能更好地刻画了目标预测区域与真实区域的位置关系。在目标回归计算的过程中, 若2区域的IoU值为0, 则无法进行回归计算, 导致该预测区域的直接丢失, 通过使用IoU作为目标位置关系的刻画函数, 在对预测区域进行回归调整时, 当出现2区域不相交时, IoU为一负值, 将会继续对目标区域进行回归计算, 而不会导致预测区域丢失, 即出现漏检情况。

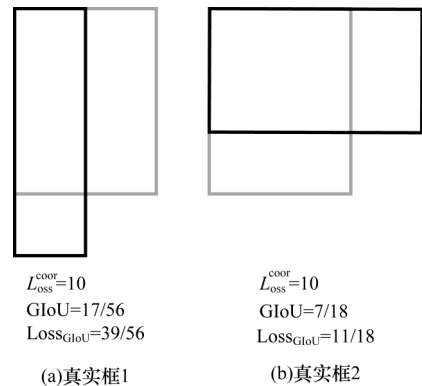


图2 同一目标区域定位损失函数值

Fig.2 Location loss function value for the same target area

2.2 添加空洞感受野模块

YOLOv3利用3个来自不同特征层级、具有不同分辨率的特征图谱, 实现对不同尺寸的目标进行差异化的检测。但浅层特征语义强度较弱, 不能充分

满足目标识别的任务需求。原检测网络中虽然使用了FPN将深层特征层中包含的强特征语义信息融入到浅层特征中,但由于深层特征层对输入图像进行了高倍数的特征压缩,因此其对于小目标的敏感性极低。如图1中的特征图谱4、5是在输入信息上进行16、32倍的特征压缩后的特征输出,其对于小目标的敏感性极低,对于尺寸小于 $16 \times 16, 32 \times 32$ 的目标,高层的特征信息无法为浅层特征带来高质量特征强度的提升,难以弥补网络在小目标上的检测缺陷。针对上述不足,构建空洞感受野模块(Atrous Receptive Module, ARM),并将该模块直接作用于特征提取网络输出的特征图谱上,直接增大相应网络层的感受野大小,提升输出特征的语义强度。

通过利用空洞卷积以包含更大的感受野,使其能够捕捉到多尺度信息。在感受野模块的不同分支使用不同大小、不同数量的传统卷积及不同膨胀率的空洞卷积,构建如图3所示的空洞感受野模块。其中Branch1、Branch2、Branch3、Branch4对应模块的4个分支结构。图3中 $1 \times 1 \text{conv}$ 、 $3 \times 3 \text{conv}$ 分别表示卷积核大小为 1×1 和 3×3 的卷积层; $3 \times 3 \text{conv rate}=2$,表示卷积核大小为 3×3 ,膨胀率为2的空洞卷积;当 $\text{rate}=3$,表示膨胀率为3的空洞卷积。DBL层是由卷积层、批归一化层和激活层组成的层级结构。concat表示在特征的最后一个维度进行堆叠运算,以产生融合特征。Add表示进行矩阵相加运算。其中 F_1 、 F_2 、 F_3 、 F_4 分别表示Branch1、Branch2、Branch3、Branch4的特征输出, F_5 为融合特征,O为模块最终输出。

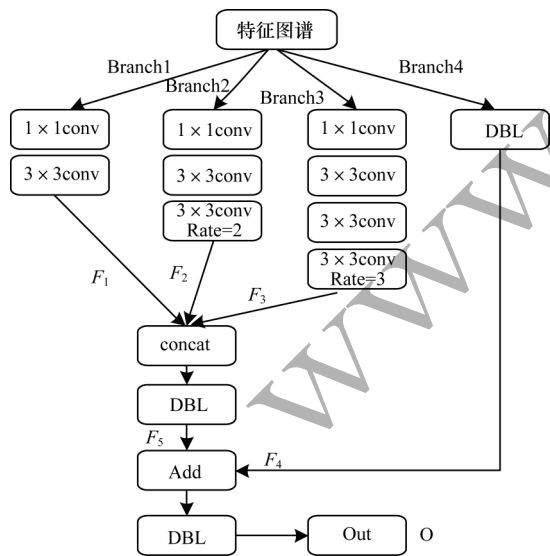


图3 空洞感受野模块

Fig.3 Atrous receptive module

感受野模块的输出特征计算过程如式(8)~式(12)所示。其中: F_0 表示感受野模块的输入; W 、

H 、 C 分别表示输入特征图谱的宽、高和特征通道数; A 表示空洞卷积; B 表示普通卷积; k 表示卷积核大小; r 表示空洞卷积的膨胀率; DBL 表示依次进行卷积、批归一化,以及激活运算; $*$ 表示卷积计算; \oplus 表示矩阵相加; \odot 表示进行特征通道的连接,即图3中的concat运算。在此模块中所有的卷积步长均为1。

对输入特征 F_0 进行2次卷积运算,输出特征图大小为 $W \times H \times (C/4)$,Branch1的感受野大小为 3×3 。 F_1 的计算公式如式(8)所示:

$$F_1 = F_0^{W \times H \times C} * B^{1 \times 1 \times (C/4)} * B^{3 \times 3 \times (C/4)} \quad (8)$$

输出特征 F_2 的大小为 $W \times H \times (C/4)$,Branch2感受野大小为 7×7 。 F_2 的计算公式如式(9)所示:

$$F_2 = F_0^{W \times H \times C} * B^{1 \times 1 \times (C/4)} * B^{3 \times 3 \times (C/4)} * (A^{3 \times 3 \times (C/4)})^2 \quad (9)$$

输出特征 F_3 的大小为 $W \times H \times (C/4)$,Branch3的感受野大小为 11×11 。 F_3 的计算公式如式(10)所示:

$$F_3 = F_0^{W \times H \times C} * B^{1 \times 1 \times (C/8)} * B^{3 \times 3 \times (3C/16)} * B^{3 \times 3 \times (C/4)} * (A^{3 \times 3 \times (C/4)})^3 \quad (10)$$

输出特征 F_4 的大小为 $W \times H \times C$,Branch4的感受野大小为 1×1 。 F_4 的计算公式如式(11)所示:

$$F_4 = F_0^{W \times H \times C} * DBL^{1 \times 1 \times C} \quad (11)$$

如图3所示,将Branch1、2、3输出 F_1 、 F_2 、 F_3 进行concat运算,并将输出结果进行一次DBL算法获得融合特征 F_5 。输出特征 F_5 的大小为 $W \times H \times C$ 。 F_5 的计算公式如式(12)所示:

$$F_5 = (F_1 \odot F_2 \odot F_3) * DBL^{1 \times 1 \times C} \quad (12)$$

如图3所示,将融合特征 F_5 特征与 F_4 特征进行跳跃连接,进行矩阵相加运算,并卷积大小为 1×1 的DBL层进行通道调整,获得最终输出特征 O 。特征 O 的大小为 $W \times H \times C$ 。 O 的计算公式如式(13)所示:

$$O = (F_5 \oplus F_4) * DBL^{1 \times 1 \times C} \quad (13)$$

2.3 双向金字塔模块

原网络中以FPN为特征融合方式,该结构仅包含单向的自顶向下的特征融合过程,仅为浅层特征提供了特征语义增强,忽略了高层特征对于上下文信息的缺失。高层特征由于经过高倍数的特征压缩导致特征图谱中丢失了大量的细节信息,难以满足精确目标定位任务的需求。本文通过使用双向特征金字塔(Bidirectional Feature Pyramid Network, BiFPN)替代原有FPN结构,实现浅、深层特征的双向融合,提升网络目标检测的能力。

如图4所示,BiFPN包含双向的特征传递信息流,即下行信息流和上行信息流。图中 F_i 表示各层级结构的特征输出, T_i 表示输入特征图谱编号。降维操作层表示对输入特征的维度降为原特征的 $1/2$;上采样降维层表示对传入网络特征,将特征维度降维

到原特征维度的 $1/2$, 对变换后的特征进行上采样, 将特征图谱的尺寸放大 1 倍。特征融合表示将 2 个传入特征在特征的最后一个维度上进行连接。图中的 $5 \times \text{DBL}$ 结构, 表示依次将特征传入 5 个 DBL 层进行运算, 每个 DBL 层由 1 个卷积层、1 个批归化层和 1 个激活层组成。该结构的使用可充分增加网络对于非线性特征的表达能力。下采样升维层表示对输入特征, 先将其特征图谱的维度数扩大 1 倍, 再将其特征图谱的尺寸下采样压缩为原尺寸的 $1/2$ 。

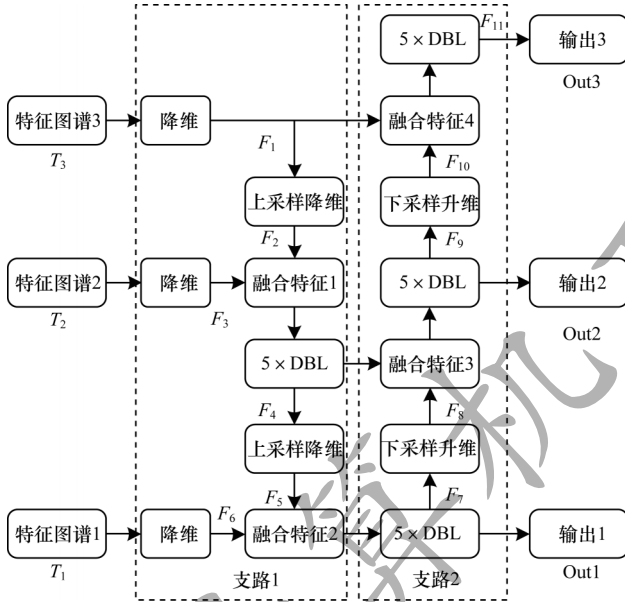


图4 双向特征金字塔结构

Fig.4 Bidirectional feature pyramid structure

改进模块中, 在每个特征图谱后添加了降维层, 目的在于降低输出特征维度数, 减少模块参数, 从而降低网络运算的开支。在改进模型中多次使用 $5 \times \text{DBL}$ 结构, 原因在于该结构可以提升模型对于非线性特征的表达能力。改进结构中通过自顶向下的特征融合信息流, 将深层特征中强语义特征信息融入到浅层特征中, 提升了网络浅层目标预测分支目标分类的能力。通过自底向上的特征融合特征流, 将浅层特征中丰富上下文细节特征信息融入到深层特征中, 提升了深层目标预测分支的目标定位能力。

2.4 实验步骤

为解决 YOLOv3 网络在多尺寸目标检测较差的问题, 首先通过使用改进 $\text{Loss}_{\text{GIOU}}$ 损失函数, 将预测目标回归区域整体化, 降低目标漏检率; 其次, 通过在网络不同位置添加空洞感受野模块, 测试相应模型目标检测精度, 确定空洞感受野模块的最佳添加位置; 然后, 通过对使用 FPN 以及 BiFPN 的网络检测其目标检测精度, 确定检测精度更高的模型结构; 最终将 2 个改进模块进行融合, 获得 BR-YOLOv3 目标检测网络, 并测试最终改进网络的检测精度。本文的实验流程如图 5 所示。

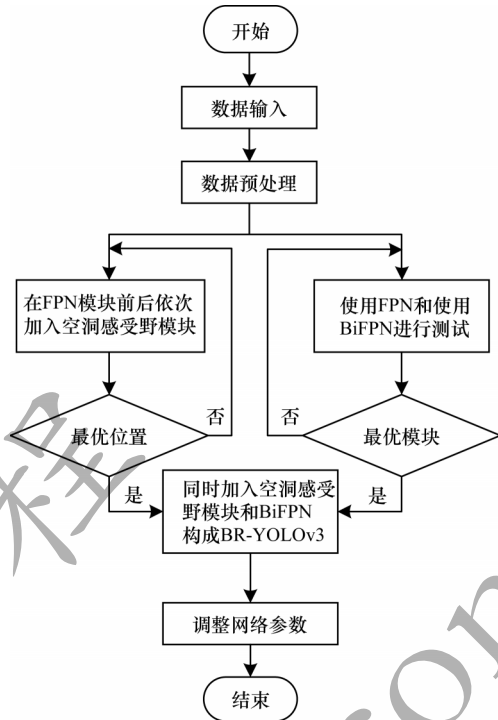


图5 实验流程

Fig.5 Experimental procedure

3 实验

3.1 实验数据

实验采用 Pascal VOC 数据集, 该数据集是图像识别与分割领域公开的标准化数据集。实验数据分为 2 部分: 训练数据集 (Trainval), 包含 VOC2012 数据集的全部以及 VOC2007 数据集的 Trainval 部分, 共包含 16 552 张完整标注的图片; 测试集 (Test) 采用 VOC2007 数据集的 test 部分, 共包含 4 953 张完整标注的图片。

3.2 实验结果评价标准

实验中采用 AP、Recall、mAP 作为实验结果评价标准。Precision 表示某一类别预测目标中预测正确占总真实标签个数的比例。Recall 表示预测目标正确的数量占目标预测总数的比例。AP 值由精度 (Precision) 和召回率 (Recall) 组成的 PR 曲线与 x, y 轴所围成面积计算得到。本文所采用的数据集共包含 20 个样本种类。mAP 为 20 个类别的 AP 值相加除以总类别数计算得到。

3.3 实验参数

实验中使用在 ImageNet 上训练好的 Darknet53 的参数为模型特征提取网络的初始化参数。在模型训练过的前 2 个批次采用预热学习率, 对特征提取网络以外的参数进行初步调整。在 2~50 个 epoch 间, 采用余弦退火学习方式, 学习率逐步由 0.000 1 下降到 0.000 001。在 2~20 个批次, 对特征提取网络部分模型参数进行固定, 对其余网络参数进行调整。在 20~50 个批次, 对整体网络的参数进行微调。训练过程中每次传入网络的图片数量 (Batch Size) 为 4。

3.4 实验环境

本文实验在 LINUX 系统下进行,使用显卡型号为 RTX2080Ti,CPU 型号为 i9-9900k,所使用的编程语言为 python3.6,使用深度框架为 Tensorflow 1.11.0。

3.5 实验结果

针对本文所使用的数据集,使用 kmeans 算法对数据集中包含的所有真实区域进行聚类计算,获得实验中所使用 Anchor 框的大小。通过 kmeans 获得的 Anchor 框的尺寸大小如下:(26,40),(48,98),(125,98),(73,199),(173,184),(123,297),(330,192),(220,329),(361,362)。

首先对定位损失函数 $\text{Loss}_{\text{GIOU}}$ 的改进模型与使用 MSE 做为定位损失函数的原模型的结果进行对比实验。实验结果如表 1 所示。

表 1 不同损失定位损失函数模型检测精度
Table 1 Different loss positioning loss function model detection accuracy

实验编号	损失函数	精度/%
A	MSE(Loss)	75.72
B	GIOU(Loss)	76.48

实验结果表明,通过使用 $\text{Loss}_{\text{GIOU}}$ 作为网络目标定位损失函数,实验检测精度达到 76.48%,在原网络的基础上提升了 0.76 个百分点。通过使用 $\text{Loss}_{\text{GIOU}}$ 损失函数有助于提升目标定位精度,并最终提升了目标检测精度。为进一步论证 $\text{Loss}_{\text{GIOU}}$ 定位损失函数的有效性及鲁棒性,通过分别计算测试集大、中、小尺寸的目标的检测精度。其中,目标尺寸大于 206×206 的目标为大尺寸目标,尺寸在 206×206 与 104×104 之间的目标为中尺寸目标,尺寸小于 104×104 的目标为小尺寸目标。多尺寸目标检测精度的实验结果如图 6 所示。使用定位损失函数为 $\text{Loss}_{\text{GIOU}}$ 的改进检测模型在大、中、小尺寸目标的检测精度分别为 88.45%、81.51%、70.53%;以 MSE 为目标定位损失函数的检测模型在大、中、小目标上的检测精度依次为 87.75%、78.51%、60.37%。实验结果表明,使用 $\text{Loss}_{\text{GIOU}}$ 作为定位损失函数的检测模型,在小尺寸目标的检测精度上较原模型检测精度获得较大提升,且改进目标定位损失函数具有良好的稳定性。

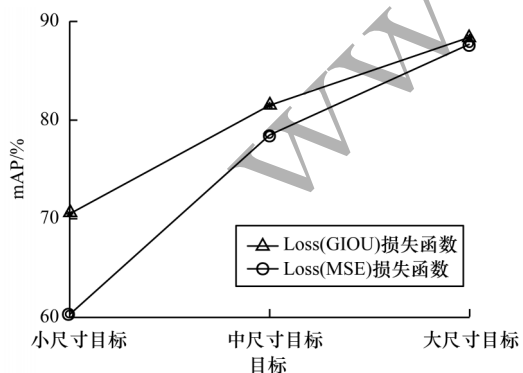


图 6 不同定位损失函数下不同尺寸目标检测精度

Fig.6 Object detection accuracy of different sizes under different positioning loss functions

实验第 2 阶段,将 ARM 模块加入网络的不同结构位置,分析在不同结构位置使用空洞感受野模块所带来的精度变化。如图 7(b)所示,将 ARM 连接到特征提取网络之后,并将该网络命名为 ARM(1)-YOLOv3;如图 7(c)所示,将 ARM 连接在 FPN 结构后,并将该网络命名为 ARM(2)-YOLOv3。图 7(a)为原网络结构。

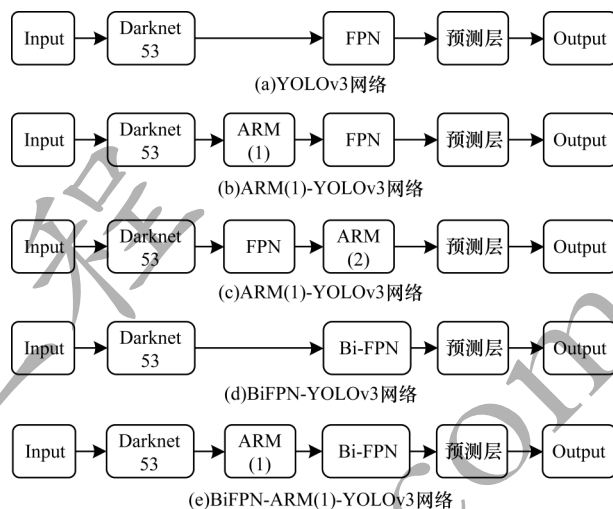


图 7 测试模型结构

Fig.7 Test model structure

实验对比结果如表 2 所示。由实验结果可知,在特征提取网路后添加 ARM 模块相较于在 FPN 结构后添加 ARM 模块对总体网络的检测精度的提升更大。实验 C 的检测精度达到了 76.93%,在原网络检测精度上提升 1.21 个百分点。在 E、F 这 2 组实验中,将 $\text{Loss}_{\text{GIOU}}$ 融入改进网络进行实验,实验结果表明,在使用空洞感受野模型的基础上使用 $\text{Loss}_{\text{GIOU}}$ 损失函数可以使目标检测精度得到进一步提升,达到 77.96%,在原网络检测精度的基础上提升了 2.96 个百分点。

表 2 不同位置添加 ARM 的模型检测精度

Table 2 Model detection accuracy of adding ARMs in different locations

编号	网络名	精度
A	YOLOv3	75.72
C	ARM(1)-YOLOv3	77.45
D	ARM(2)-YOLOv3	76.93
E	ARM(1)-YOLOv3(LossGIOU)	77.96
F	ARM(2)-YOLOv3(LossGIOU)	77.28

实验第 3 阶段,在 YOLOv3 网络的基础上使用双向特征金字塔结构代替原有的 FPN 网络结构,网络结构如图 7(d)所示。实验结果如表 3 所示,实验 G 的检测精度达到 77.20%,相较于原网络在检测精度上提升 1.48 个百分点。实验结果表明通过使用双向特征金字塔提升网络共享特征利用率,不仅提升了深层检测分支的目标定位能力,而且提升了整体网络的目标检测能力。实验 H 中,将双向特征金字塔结构融入到实验

E所用目标检测模型 ARM(1)-YOLOv3(Loss_{GIOU})中,得到模型 G,即 BiFPN-ARM(1)-YOLOv3。模型结构如图 7(e)所示。实验结果表明,该整体网络检测精度进一步提升到了 79.24%,在原 YOLOv3 的基础上提升了 3.52 个百分点。

表 3 采用 BiFPN 模块不同模型的检测精度

Table 3 Detection accuracy of different models using

BiFPN module		%
实验编号	网络名	精度
A	YOLOv3	75.72
G	BiFPN-YOLOv3	77.20
H	BiFPN-ARM(1)-YOLOv3(Loss _{GIOU})	79.24

以实验 H 中所获模型为最终改进目标检测网络的结构,改进网络结构如图 7(e)所示。

在 Darknet53 特征提取网络后添加 ARM-BiFPN 组合结构,代替原有的 FPN 网络结构,并替换原有目标定位损失函数为 Loss_{GIOU},获得最终改进检测网络 BR-YOLOv3。如表 4 所示,对主流经典目标检测网络 Faster RCNN(Faster Region Convolution Neural network)及 SSD 在同一实验环境下进行对比实验。实验结果表明,改进的 BR-YOLOv3 在 12 个目标类别的检测结果上均取得了最优的结果。改进网络在 Bird、Cat、Cow、Dog 等目标上,相较于 YOLOv3 网络的检测精度结果均获得了明显的提升。BR-YOLOv3 最终检测精度均优于当下的主流目标检测网络。

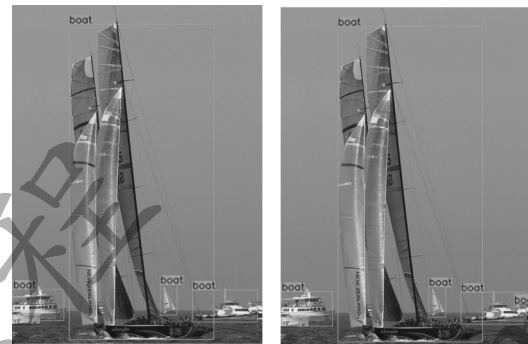
表 4 VOC2012 测试集检测精度

Table 4 VOC2012 test set detection accuracy %

目标类别	Faster RCNN	SSD	YOLOv3	BR-YOLOv3
Aero	83.65	86.14	82.99	88.61
Bicycle	78.85	80.78	84.76	87.31
Bird	74.59	68.28	72.62	81.07
Boat	54.38	60.48	64.66	67.13
Bottle	46.84	68.24	69.48	69.55
Bus	78.16	80.27	85.17	85.85
Car	77.97	77.95	88.72	88.95
Cat	85.52	89.22	80.38	87.63
Chair	47.12	60.37	59.41	59.28
Cow	78.35	77.06	77.00	84.42
Table	55.39	58.89	67.86	76.62
Dog	84.50	87.28	75.17	87.10
Horse	82.39	83.96	85.14	86.72
Mbike	81.91	81.93	83.19	87.71
Person	79.61	82.13	84.08	85.04
Plant	42.35	55.56	47.08	47.77
Sheep	71.65	79.82	78.74	79.74
Sofa	60.42	75.58	71.18	72.39
Train	82.24	86.23	78.62	84.93
Tv	60.56	73.15	78.22	77.03
mAP	70.32	75.66	75.72	79.24

3.6 多尺寸目标检测测试

图 8(a)和图 8(c)为 YOLO v3 的实例检测结果,图 8(b)和图 8(d)为改进网络 BR-YOLOv3 的实例检测结果。通过对比发现,图 8(a)和图 8(c)均未能检测到图像右下角的小尺寸目标,而图 8(b)和图 8(d)的改进模型成功检测到了位于图像右下角的 Boat 以及 Horse。



(a)YOLOv3检测效果

(b)BR-YOLOv3检测效果



(c)YOLOv3检测效果



(d)BR-YOLOv3检测效果

图 8 YOLOv3 与 BR-YOLOv3 多尺寸目标检测结果

Fig.8 YOLOv3 and BR-YOLOv3 multi-size target detection results

2 组模型在相同输入图像上产生差异化检测结果的原因在于:YOLOv3 采用了以 MSE 为目标函数的目标定位损失函数,在网络的训练过程中,通过对 1 个 batch 内所有预测框选区域进行回归调整,使小目标预测框所贡献的定位损失较小,而针对目标区域产生的损失值由尺寸较大的目标预测区域所主导,最终导致预测区域在调整过程中,出现由于调整尺度过大而造成小目标预测框偏离真实区域的现象,致使网络无法计算其回传梯度值,最终导致目标的丢失。这使得网络对于小目标的检测缺乏有效的学习,原网络在测试过程中出现边缘小目标漏检现象。改进网络通过引入 Loss_{GIOU} 定位损失函数,有效避免了当预测区域发生偏离无法

进行回归调整的问题。在 $\text{Loss}_{\text{GIoU}}$ 定位损失函数中添加对不同尺寸目标的惩罚项,通过衰减大目标预测区损失以提升小目标预测区域贡献的损失占比,为网络学习小目标检测提供更多机会。通过使用 ARM 模块提升浅层预测分支使用特征的语义强度,提升浅层目标预测分支对小目标的准确定位与有效分类的能力。通过以上分析可知,改进模型在多尺寸目标的检测上具有更优异的表现。

4 结束语

本文基于空洞感受野模块和双向特征金字塔结构提出 BR-YOLOv3 目标检测网络,将空洞感受野模块嵌入特征提取网络,提升各层预测分支输出特征的语义强度,使用双向特征金字塔结构提高共享特征的利用率。将丰富目标位置信息、边缘信息融入深层特征中,提升深层特征预测分支的目标定位能力。实验结果表明,改进模型提升了对多尺寸目标检测的性能,相较于 Faster RCNN、SSD 等主流目标检测算法具有更高的准确率,整体平均检测精度达到 79.24%。下一步将采用 Mixup、Cutmix 等图像增强方法及风格迁移对抗以提升对数据集中困难数据样本的识别率,并尝试使用随机遮挡目标的训练方法,提升网络对所识别目标各个部分的敏感性,以充分发掘网络对残缺目标的识别能力。

参考文献

- [1] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA: IEEE Press, 2016: 770-778.
- [2] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift [EB/OL]. [2020-07-01]. <https://arxiv.org/abs/1502.03167>
- [3] REDMON J, FARHADI A. YOLOv3: an incremental improvement[EB/OL]. [2020-07-01]. <https://arxiv.org/abs/1804.02767>.
- [4] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer, 2016: 21-37.
- [5] LIN T, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA: IEEE Press, 2017: 936-944.
- [6] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA: IEEE Press, 2016: 2818-2826.
- [7] 赵亚男, 吴黎明, 陈琦. 基于多尺度融合 SSD 的小目标检测算法[J]. 计算机工程, 2020, 46(1): 247-254.
- [8] ZHAO Y N, WU L M, CHEN Q, et al. Small object detection algorithm based on multiscale fusion SSD[J]. Computer Engineering, 2020, 46(1): 247-254. (in Chinese)
- [9] DAI J, QI H, XIONG Y, et al. Deformable convolutional networks[C]//Proceedings of IEEE International Conference on Computer Vision. Washington D. C. , USA: IEEE Press, 2017: 764-773.
- [10] 麻森权, 周克. 基于注意力机制和特征融合改进的小目标检测算法[J]. 计算机应用与软件, 2020, 37(5): 194-199.
- [11] MA S Q, ZHOU K. Improved small target detection algorithm based on attention mechanism and feature fusion[J]. Computer Applications and Software, 2020, 37(5): 194-199. (in Chinese)
- [12] ZHU C, HE Y, SAVVIDES M. Feature selective anchor-free module for single-shot object detection[C]//Proceedings of Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA: IEEE Press, 2019: 840-849.
- [13] 肖锋, 李茹娜. 语义信息引导下的显著目标检测算法[J]. 计算机工程, 2019, 45(4): 248-253.
- [14] XIAO F, LI R N. Salient object detection algorithm under guidance of semantic information[J]. Computer Engineering, 2019, 45(4): 248-253. (in Chinese)
- [15] REZATOFIGHI H, TSOI N, GWAK J Y, et al. Generalized intersection over union: a metric and a loss for bounding box regression[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA: IEEE Press, 2019: 658-666.
- [16] ZHENG Z, WANG P, LIU W, et al. Distance-IoU loss: faster and better learning for bounding box regression[C]//Proceedings of Conference on Artificial Intelligence. Washington D. C. , USA: IEEE Press, 2020: 435-451.
- [17] 秦升, 张晓林, 陈利利, 等. 基于人类视觉机制的层级偏移式目标检测[J]. 计算机工程, 2018, 44(6): 253-258.
- [18] QIN S, ZHANG X L, CHEN L L, et al. Hierarchical offset object detection based on human visual mechanism[J]. Computer Engineering, 2018, 44(6): 253-258. (in Chinese)
- [19] LI Z, PENG C, YU G, et al. Detnet: a backbone network for object detection[EB/OL]. [2020-07-01]. https://www.researchgate.net/publication/324584281_DetNet_A_Backbone_network_for_Object_Detection.
- [20] TIAN Z, SHEN C, CHEN H, et al. Fcos: fully convolutional one-stage object detection [C]//Proceedings of IEEE International Conference on Computer Vision. 2019: 9627-9636.
- [21] LIU W, LIAO S, REN W, et al. High-level semantic feature detection: a new perspective for pedestrian detection[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA: IEEE Press, 2019: 5187-5196.
- [22] 朱辉, 秦品乐. 基于多尺度特征结构的 U-Net 肺结节检测算法[J]. 计算机工程, 2019, 45(4): 254-261.
- [23] ZHU H, QIN P L. U-Net pulmonary nodule detection algorithm based on multi-scale feature structure [J]. Computer Engineering, 2019, 45(4): 254-261. (in Chinese)

编辑 赖玉玲