

尺度敏感损失与特征融合的快速小目标检测方法

据长瑞,秦晓燕,袁广林,李 豪,朱 虹

(中国人民解放军陆军炮兵防空兵学院计算机教研室,安徽合肥 230031)

摘 要: 现有通用深度学习目标检测方法对中、大目标有着较好的检测精度,而对小目标检测精度较低,主要原因在于小目标训练数据少以及下采样后特征图分辨率过低. 针对上述问题,一方面,提出一种尺度敏感损失函数用于分类热图的训练,使小目标能够对模型更新产生更大的影响;另一方面,利用反卷积与可变形卷积提出一种自上而下的特征融合方法,获得高分辨率、强语义的特征图来检测目标. 在上述两个方面的基础上,设计一种尺度敏感与特征融合的小目标检测方法. 在PASCAL VOC数据集上,对提出方法进行了实验验证,实验结果表明:相比于现有目标检测方法,本文方法在保持较快检测速度的同时,提升了小目标检测的精度.

关键词: 深度学习; 小目标检测; 尺度敏感损失; 特征融合

中图分类号: TP391.4

文献标识码: A

文章编号: 0372-2112(2022)09-2119-08

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20210530

Fast Small Object Detection Method with Scale-Sensitivity Loss and Feature Fusion

JU Chang-rui, QIN Xiao-yan, YUAN Guang-lin, LI Hao, ZHU Hong

(Computer Teaching and Research Section, PLA Army Academy of Artillery and Air Defense, Hefei, Anhui 230031, China)

Abstract: The existing general deep learning target detection methods have good detection accuracy for medium and large targets, but the detection accuracy for small targets is low, mainly due to few data for small target training and low resolution of feature map after down sampling. To solve the above problems, on the one hand, a scale sensitive loss function is proposed for the training of classified heatmaps, so that small targets can have a greater impact on model updating; on the other hand, a top-down feature fusion method is proposed by using deconvolution and deformable convolution to obtain high-resolution and strong semantic feature map for target detection. On the basis of the above two aspects, a small target detection method based on scale sensitivity and feature fusion is designed. Experimental results on PASCAL VOC dataset show that compared with the existing target detection methods, the proposed method can maintain a faster detection speed and improve the accuracy of small object detection.

Key words: deep learning; small object detection; scale-sensitivity loss; feature fusion

1 引言

目标检测是计算机视觉领域的一个重要研究分支,其目的是预测图像中每个感兴趣的实例的边界框位置和类别标签. 目标检测在工业、农业、医学和军事等领域有着广泛的应用前景,具有较高的研究价值. 近年来,随着深度学习技术的发展,深度神经网络框架下的目标检测取得了重大进展^[1-5]. 虽然基于卷积神经网络的目标检测器在各种具有挑战性的数据库上取得了系列的成果,但是其仍然面临目标尺度小、数据不平衡和数据标注昂贵等诸多困难. 小目标只有十几到几十

个像素,在图像中的信息量有限,通用检测器对其检测难度较大. 另外,现有目标检测数据集中包含小目标的图片数量也相对较少. 因此,小目标检测一直是一个难点、热点问题. 为提升小目标的检测效果,目前主要有尺度建模以及数据增强两大类方法.

多尺度方法的基本思想是利用网络不同层特征检测不同尺度的目标. 基于不同尺度目标的语义特征所在特征层次的差异性,2016年Yang^[6]等人提出尺度相关池化并从不同尺度的特征层中检测目标,提升了小目标的检测精度. 基于相同的思路,刘伟^[7]等人提出的SSD(Single Shot MultiBox Detector)引入了多参考和多

分辨率检测技术,利用浅层特征检测小目标,利用深层特征检测大的目标,其小目标检测效果有了大的提升.针对小目标检测,2016年Cai^[8]等人提出在不同尺度特征图上铺设锚框(Anchor),同样实现了在不同尺度的特征图上检测不同大小的目标.由于目标检测任务同时需要物体的“语义和位置”特征,语义特征在深层特征图中,因此分层预测的网络结构难以兼顾小目标的语义和位置特征.针对该问题,2017年Lin^[9]等人提出了特征金字塔网络(Feature Pyramid Networks, FPN),用自顶向下的方法将深层特征图上采样后与浅层特征图融合,利用融合后的特征图做多尺度目标检测.为了提高小目标检测的精度,2018年,Liang^[10]等人提出一种两阶段的目标检测方法,该方法在区域建议阶段,采用横向连接的特征金字塔网络,使得小目标的语义特征更加突出,同时,为小目标设计了专门的Anchor,提升了小目标检测效果.2019年G. Ghiasi^[11]等人基于神经网络搜索技术,利用强化学习自动搜索对当前任务最优的FPN(Feature Pyramid Networks),进一步提升了小目标检测的精度.2020年Wu^[12]等人提出的多尺度正样本训练方法对小目标检测效果也有一定提升.针对小目标检测难题,在FPN的基础上,2020年Li^[13]等人提出一种提取并融合多尺度特征的目标检测网络,提高了对目标尺度变化的鲁棒性.上述多尺度检测方法虽然对小目标检测精度有提升,但也存在网络结构复杂和速度较慢等不足.解决小目标检测问题的另一类思路是数据增强.对基于锚框(anchor-based)的目标检测方法而言,通过增加检测小目标的anchor数量可以达到数据增强的目的.2017年Zhang^[14]等人提出的S3FD(Single Shot Scale-invariant Face Detector)通过调整候选框的生成步长产生更稠密的anchor增加小目标样本数量,使得小目标得到充分训练从而提升小目标的检测精度.2019年Kisantal^[15]等人对包含小目标的图像进行过采样和对小目标随机变换后复制到新位置的方式增加小目标数量.为了解决数据集中小物体数量不足的问题,2021年Chen^[16]等人使用图像拼接技术动态地生成拼接图像提升小目标数量,从而增强小目标的检测精度.

综上所述,现有针对小目标的检测方法存在网络结构复杂和速度较慢等不足.为了提高基于深度学习的目标检测方法对小目标的检测精度和速度,同时对中、大目标也有较高的检测性能,在CenterNet^[18]的启发下,本文提出一种基于尺度敏感损失与特征融合的单阶段无锚框小目标检测方法(Scale-Sensitivity Loss and Feature Fusion Detector, SSLFF-Det).该方法利用尺度敏感损失训练分类热图,提高小目标对网络模型参数学习的影响,利用反卷积和可变形卷积自上而下融合

特征,获得高分辨率、强语义的特征图来检测目标.在PASCAL VOC数据集上,对提出的方法进行了实验验证与分析,实验结果表明:本文方法达到了预期效果,即提升了小目标检测的精度和速度,同时对中、大目标也具有较优的性能.

2 尺度敏感损失

目前,常用的目标检测数据集有PASCAL VOC和MS COCO(Microsoft Common Objects in Context).文献[15]对MS COCO数据集进行统计分析发现:MS COCO数据集中小目标、中目标和大目标的占比分别是41.43%、34.32%和24.24%,但是只有51.82%的图像包含小目标,包含中、大目标的图像分别占70.07%和82.28%.虽然MS COCO数据集中小目标总数量较多,但包含小目标的图像占比较低.本文参考MS COCO数据集对小、中、大目标的划分标准,对PASCAL VOC数据集进行了数据统计,VOC 2007训练集中小目标、中目标和大目标的占比分别为11.20%、34.49%和54.31%,包含小目标的图像仅占12.67%,而包含中、大目标的图像分别占39.45%和90.52%.可以看出:在PASCAL VOC数据集中,小目标的数量占比以及包含小目标的图像数量占比均较低.PASCAL VOC中的部分图像样例如图1所示,其中多数图像只标注了中、大目标,而小目标未标注,这样会导致小目标训练样本较少.在现实应用中,小目标标注难度大等原因也会造成小目标训练样本较少的情况.

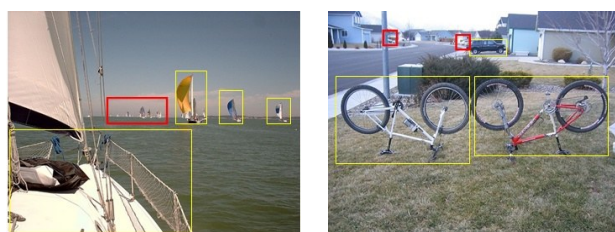


图1 PASCAL VOC数据图像(其中黄色框为标注目标,红色框内是未标注的小目标)

众所周知,目标检测网络的训练,是随机取一批图像计算其中标注样本的损失进行网络参数的训练.由于训练数据中包含小目标的图像较少,这会导致每一批图像中包含的小目标样本较少.这种尺度上的样本不平衡使得小目标对模型训练的影响相对中、大目标较小,导致模型对小目标的学习不充分.另外,目标在特征图上的信息量大小与目标尺度存在正相关关系,因此小目标在特征图上的信息较少,对模型训练的影响也较小.上述两个方面是导致小目标检测精度低的重要原因.为解决难易样本不平衡问题,何恺明等人提出Focal Loss^[19],在分类损失中加入一种与样本难易度

正相关的损失权重,有效提高了网络训练对难样本的关注度,提升了目标检测的性能.受此启发,为提高小目标对模型的影响,本文提出尺度敏感损失函数用于目标检测网络的训练,其定义如下:

$$SSL_{class} = k^{1-s} \frac{-1}{N} \sum_{cyc} \begin{cases} (1 - Y_{cyc}^{\wedge})^{\alpha} \log(Y_{cyc}^{\wedge}), & Y_{cyc} = 1 \\ (1 - Y_{cyc}^{\wedge})^{\beta} (Y_{cyc}^{\wedge})^{\alpha} \log(1 - Y_{cyc}^{\wedge}), & \text{其他} \end{cases} \quad (1)$$

其中, Y_{cyc} 和 \hat{Y}_{cyc} 分别表示类别热图中 (x, y) 处 c 类的标注值和预测值, N 是标注的样本个数, α, β 为超参数, 实验中分别设置为 2 和 4, s 是与目标尺度负相关的变量.

(1) 式中 k^{1-s} 定义为损失权重, 其中 s 为目标面积与图像面积的比值, 即目标的相对尺度. 对于 s 的取值, 如果一张图像中包含多个目标, 则取其中尺度最小的目标计算 s 的值, 以保证每张包含小目标的图像可以有较大的训练权重, 在训练中受到更多关注. k 为超参数, 用来调节尺度敏感权重的大小. 图 2 给出了损失权重 k^{1-s} 的函数图, 从图 2 可以看出, s 能保证 k^{1-s} 的取值在 1 到 k 之间, 这确保目标尺度对网络训练产生的影响在一定范围. 从函数性质上看, k^{1-s} 是指数函数, 它对自变量 s 的变化比较敏感, 这使得包含不同尺度目标的损失权重产生差别.

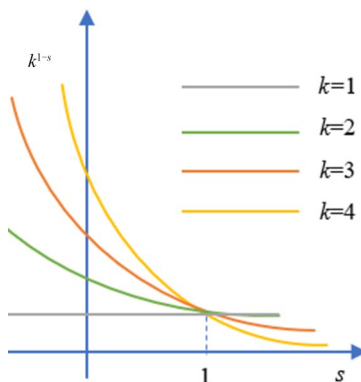


图 2 尺度敏感权重函数图像

基于上述分析,设计一批训练数据的损失计算方法,见算法 1.

算法 1 中, $w_{i,j}^o$ 和 $h_{i,j}^o$ 分别是 I_{batch} 中第 i 个训练图像中第 j 个目标的宽和高; W_i^l 和 H_i^l 分别是 I_{batch} 中第 i 个训练图像的宽和高; s_i^l 为第 i 个训练图像中最小目标的面积, s_i^2 为第 i 个训练图像的面积; N_i 是第 i 个训练图像中的目标数.

3 反卷积与可变形卷积特征融合网络

现有文献[6~13]表明:在特征提取网络中,深层特征感受野较大,包含丰富的语义信息,浅层特征感受野

算法 1 一批数据损失计算方法

输入:一批训练图像 $I_{batch} = \{I_1, I_2, \dots, I_M\}$, 超参数 k, α 和 β .

输出: $Loss_{batch}$

$Loss_{batch} = 0$;

FOR $i = 1$ to M

$$s_i^1 = \min(w_{i,1}^o \times h_{i,1}^o, \dots, w_{i,j}^o \times h_{i,j}^o, \dots, w_{i,N_i}^o \times h_{i,N_i}^o);$$

$$s_i^2 = W_i^l \times H_i^l;$$

$$s_i = s_i^1 / s_i^2;$$

$$SSL_{class}^i = k^{1-s_i} \frac{-1}{N_i} \sum_{cyc} \begin{cases} (1 - \hat{Y}_{cyc})^{\alpha} \log(\hat{Y}_{cyc}), & Y_{cyc} = 1 \\ (1 - Y_{cyc})^{\beta} (\hat{Y}_{cyc})^{\alpha} \log(1 - \hat{Y}_{cyc}), & \text{其他} \end{cases};$$

$$Loss_{batch} = Loss_{batch} + SSL_{class}^i;$$

END

较小,具有更多边缘和纹理等细节信息.目标检测的两个核心组件是分类和回归,语义信息的分类能力强,细节信息的定位能力大.小目标的语义信息在深层特征中能量小甚至消失,其细节信息主要在浅层特征中.中大目标的语义信息在深层特征中具有较高能量,其细节信息在浅层特征中更加丰富.为了提升小目标的检测性能,根据深层和浅层特征的上述特点,本文提出一种基于反卷积与可变形卷积的特征融合模块,其结构如图 3 所示.该模块首先对待融合特征图 f-map2 进行 3×3 的 2 倍上采样反卷积,得到与融合特征图 f-map1 相同大小的特征图;然后再通过一个可变形卷积,来缓解上采样带来的混叠效应和棋盘效应,并增强模型的拟合能力;最后将得到的特征图与 f-map1 相加得到本模块的融合特征图 ff-map.该模块可以根据网络深度及实际需求增加或减少,以达到最优效果.实验发现:在 ResNet-18 与 ResNet-50 上分别加入 2 个模块效果较好,而 ResNet-101 上使用 3 个效果更佳.

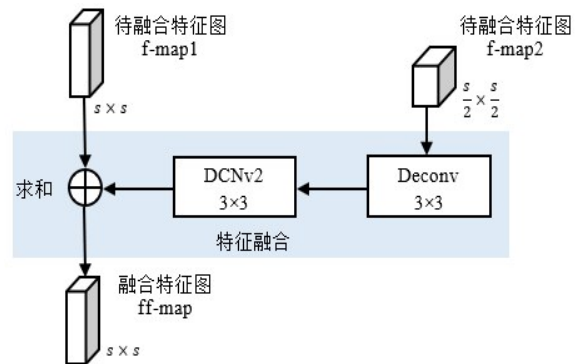


图 3 特征融合模块

利用提出的特征融合网络模块设计小目标检测网络,网络结构如图 4 所示,包括自底向上和自顶向下两个通路.自底向上通路是骨干网络用于特征提取,本文利用 ResNet-50 构建自底向上通路.为获得高分辨率的

特征图,删除了 ResNet-50 第一个池化层,利用其 block1~block4 得到尺度大小为原图的 1/2、1/4、1/8 和 1/16 的特征图.自顶向下通路用于特征融合,包括 2 个串联的特征融合模块,先将 block4 与 block3 的输出特征进行融合;然后再将得到的融合特征与 block2 的输出特征进行融合,从而得到高分辨率、强语义的融合特征图用于目标检测.网络的输出是目标类别、尺度和中心偏移量.

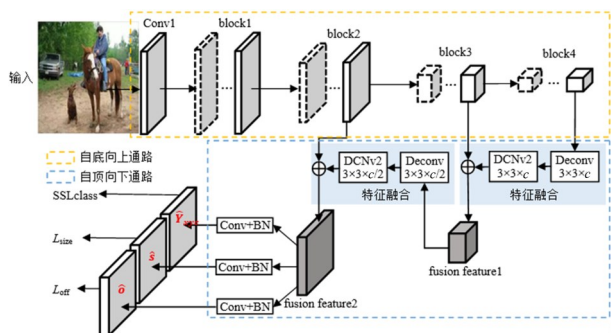


图 4 特征融合网络结构

图 5(a) 给出了 PASCAL VOC 数据集中一张图像,图 5(b)~(e) 显示了图 4(a) 在特征融合网络中 block1~block4 的输出特征图,图 5(f) 是其最终融合特征图.对比图 5(b)~(f) 各特征图可以看出:在最终融合特征图 5(f) 中,小鸟的融合特征既包含了深层特征的语义信息,又具有浅层的纹理边缘信息,大目标汽车的融合特征也同时包含深层语义信息和浅层纹理边缘信息.

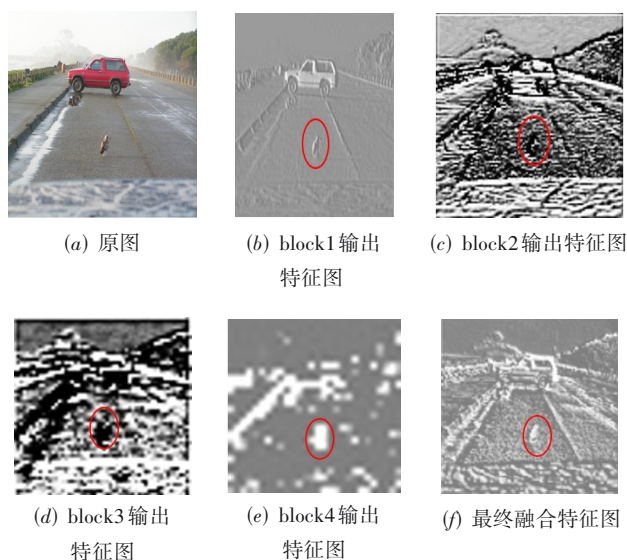


图 5 目标图像与特征图

4 网络训练与实现细节

本文用多任务损失函数训练网络,其定义如下:

$$L_{\text{total}} = \text{SSL}_{\text{class}} + \lambda_{\text{size}} L_{\text{size}} + \lambda_{\text{off}} L_{\text{off}} \quad (2)$$

其中, $\text{SSL}_{\text{class}}$ 是式(1)所示的尺度敏感分类损失; λ_{size} 和 λ_{off} 是超参数,本文 λ_{size} 取 0.1, λ_{off} 取 1.0; L_{size} 是尺度损失,其定义如式(3)所示; L_{off} 是中心偏移损失,其定义如式(4)所示.

$$L_{\text{size}} = \frac{1}{N} \sum_{k=1}^N |\hat{s}_{p_k} - s_k| \quad (3)$$

式(3)中, N 是样本个数, \hat{s}_{p_k} 和 s_k 分别是目标 k 的尺度预测值和标注值.

$$L_{\text{off}} = \frac{1}{N} \sum_{k=1}^N \left| \hat{\delta}_{p_k} - \left(\frac{p}{R} - \hat{p} \right) \right| \quad (4)$$

式(4)中, N 是样本个数, $\hat{\delta}_{p_k}$ 是目标 k 的中心偏移预测值, p 是目标 k 的位置标注值, R 是下采样倍数,本文 $R=4$.

网络训练时输入图像的大小归一化为 512×512 ,因此网络输出的目标类别热图、尺度热图和偏移量热图大小均是 128×128 . 目标类别热图的通道数与数据有关,对于 PASCAL VOC 数据集类别热图的通道数是 20,对于 MS COCO 数据集其类别通道数是 80,尺度热图的通道数和偏移量的热图通道数均是 2. 模型的初始化权重是由 ResNet-50 在 ImageNet 上预训练得到,并使用 4 张 GPU 对 ResNet-50 网络进行训练,训练批次大小 (batchsize) 设置为 32,共训练 100 个轮次 (epoch). 初始学习率设置为 $2.5e-4$,并在第 50 及 75 个轮次分别衰减至 $2.5e-5$ 及 $2.5e-6$.

5 实验结果与分析

5.1 实验环境与数据集

在 SYS-7048GR-TR 台式机 (CPU 型号为 Intel Xeon (R) ES-2630v4@2.20 GHz×20,内存为 64 GB, GPU 为 RTX2070S 8G) 上实现了本文方法,软件环境是: Ubuntu 18.04、Python 3.6、torch 1.4.0、cuda10.1 和 cudnn7.5. 为了验证本文方法的有效性,使用 PASCAL VOC 数据集对本文方法进行消融实验,并与现有优秀方法进行了对比分析.

5.2 消融实验

针对本文的两个改进方面进行消融实验,以验证尺度敏感损失和特征融合两方面对小目标检测性能的提升作用.

5.2.1 尺度敏感损失实验分析

首先,为验证尺度敏感损失和不同超参数对目标检测精度的影响,对取不同 k 值的本文方法进行实验验证与分析,实验结果如表 1 所示. 从表 1 可以看出:尺度敏感损失明显提升了小目标的检测精度,表明尺度敏感损失在训练时提高了网络对小目标的关注程度,有效提高了小目标对网络学习的影响力. 随着 k 值的增

加,小目标的检测精度逐渐提升.从表1也可以看出: k 值为4时,小目标的精度已经提升不大,而中、大目标的检测精度有一定的下降,导致总体精度下降.主要原因是当 k 值设置过大时,尺度敏感损失对网络训练的影响已经超出正常范围,对网络模型的学习有负面影响,因此尺度敏感损失对目标检测精度的提升存在上限.从表1中的实验结果可以看出: k 值设置为2时,尺度敏感损失的综合性能最佳.

表1 不同 k 值的检测结果

k 值	mAP	AP _{50:95} ^{large}	AP _{50:95} ^{medium}	AP _{50:95} ^{small}
$k=1$	77.1	80.4	75.6	25.5
$k=2$	75.5	77.5	73.5	28.6
$k=3$	73.4	75.3	70.6	29.1
$k=4$	70.2	73.3	69.1	29.3

5.2.2 特征融合实验分析

为验证特征融合对小目标检测的性能的影响,为使用不同深度 ResNet 的本文方法进行实验验证与分析,实验结果如表2所示.从表2可以看出:对不同深度的 ResNet 进行特征融合后,小目标的检测精度均有不同程度的提升,并且网络越深提升效果越明显.其中,轻量型网络 ResNet-18 提升最小,主要原因是其下采样倍数较低,它自身已能够提取丰富的细节信息,因此提升效果并不明显. ResNet-101 的小目标检测精度最低,特征融合后 ResNet-101 的小目标检测精度提升最大,但是检测速度有较大降低,这一结果表明:当网络层数超过某一阈值时,网络越深越不利于小目标的检测,同时检测速度也会明显下降.特征融合后 ResNet-50 的小目标精度提升了 3.4%,整体精度 mAP 有少许降低,但是其检测速度仍然较快.

表2 不同深度 ResNet 网络特征融合检测结果

骨干网络	mAP	AP _{50:95} ^{large}	AP _{50:95} ^{medium}	AP _{50:95} ^{small}	FPS
ResNet-18	73.2	75.0	72.5	26.9	142
ResNet-50	77.1	79.4	76.6	24.5	65
ResNet-101	80.5	82.4	78.5	23.4	45
FF-ResNet-18	65.1	67.9	60.2	27.6	122
FF-ResNet-50	76.8	77.1	76.4	27.9	50
FF-ResNet-101	78.5	79.0	77.5	28.8	29

5.2.3 综合影响实验分析

为验证尺度敏感损失与特征融合同时加入对小目标检测性能的综合影响,同时使用尺度敏感损失与特征融合网络进行训练和测试.其中尺度敏感损失 k 值设置为2,并采用不同深度的 ResNet 作为骨干网络,实验结果如表3所示.结合表1、表2与表3可以看出,当尺度敏感损失与特征融合网络同时加入时,相比两种策略分别加入,小目标检测精度得到进一步提升,该结果表明,本文提出的两种策略能够协同工作,共同提升小目标检测精度,相互之间没有抵消及干扰.另外,两种策略同时作用时,整体精度相比单独采用特征融合网络时有所下降,其主要原因是尺度敏感损失使得网络更加关注小目标,使得中、大目标的检测精度受到一定影响,小目标精度虽有所提升,但总体精度有少许下降,但是该影响依然处于可以接受的范围.

表3 综合影响实验结果

骨干网络	k	mAP	AP _{50:95} ^{large}	AP _{50:95} ^{medium}	AP _{50:95} ^{small}
FF-ResNet-18	2	64.9	67.2	60.0	28.3
FF-ResNet-50	2	76.5	76.9	74.5	30.6
FF-ResNet-101	2	77.4	78.3	74.6	31.4

5.3 对比实验

5.3.1 定量实验

该实验将本文目标检测方法 SSLFF-Det 与目前先进目标检测方法进行性能对比分析,在相同软硬件环境下,利用 VOC 2007 与 VOC 2012 合成数据集进行训练、测试对比分析.为了比较的公平性,目标检测方法均采用 ResNet-50 或者与之深度相近的网络作为骨干网络.实验结果如表4所示.从表4结果可以看出,本文目标检测方法 SSLFF-Det 对小目标的检测最优,在检测速度方面,SSLFF-Det 的检测速度处于较高水平,能够达到 50FPS.

5.3.2 定性实验

图6给出了本文方法、CenterNet^[18]、SSD513^[7]和 YOLOv4^[23]对 PASCAL VOC 数据集中4张图像的检测结果,其中图6(a)是 CenterNet 的检测结果,图6(b)为 SSD513 的检测结果,图6(c)为 YOLOv4 的检测结果,图6(d)为本文方法的检测结果.可以看出,本文方法对小目标检测具有更高召回率.

表 4 本文方法与各类方法对比

方法类型	方法	网络结构	mAP	AP _{small} ^{50:95}	FPS
两阶段方法	Faster-RCNN ^[20]	ResNet-50	74.5	24.3	15
	Cascade RCNN ^[21]	ResNet-50	75.6	26.2	17
	R-FCN ^[22]	ResNet-50	78.4	24.5	12
	FPN ^[9]	ResNet-50	77.5	29.4	9
	CenterNet2 ^[26]	ResNet-50	80.3	29.1	27
单阶段方法	SSD321 ^[7]	ResNet-50	77.2	27.5	23
	SSD513 ^[7]	VGG16	78.8	26.6	33
	DSSD ^[24]	ResNet-50	78.4	27.1	8
	YOLOv4 ^[23]	DarkNet-53	79.5	29.8	52
	YOLOv5	YOLOv5-l	78.3	23.9	121
	CornerNet ^[25]	ResNet-50	75.3	23.6	26
	CenterNet ^[18]	ResNet-50	77.1	25.5	65
	SSLFF-Det	ResNet-50	76.5	30.6	50



(a) CenterNet 检测结果



(b) SSD513 检测结果



(c) YOLOv4 检测结果



(d) 本文方法检测结果

图 6 不同检测方法检测结果

6 结论

针对现有基于深度学习的目标检测方法对小目标

检测的不足,本文提出一种基于尺度敏感损失与特征融合的通用快速小目标检测方法. 本文方法的主要创

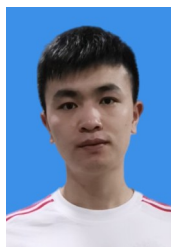
新点有两个方面:一是提出尺度敏感分类损失,使小目标在训练中更受关注,从而提升网络对小目标的学习能力;二是在 anchor-free 检测框架下加入了特征融合,从而得到融合深度语义和浅层细节的特征图进行目标检测. 对提出的目标检测方法进行了实验验证和分析,实验结果表明:本文提出的目标检测方法对小目标检测效果提升明显,且检测速度较快.

参考文献

- [1] WANG X J, ZHANG S L, YU Z R, et al. Scale-equalizing pyramid convolution for object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Piscataway: IEEE, 2020: 13356-13365.
- [2] GUO C X, FAN B, ZHANG Q, et al. AugFPN: improving multi-scale feature learning for object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Piscataway: IEEE, 2020: 12592-12601.
- [3] TAN M X, PANG R M, LE Q V. EfficientDet: scalable and efficient object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 10778-10787.
- [4] 刘颖, 刘红燕, 范九伦, 等. 基于深度学习的小目标检测研究与应用综述[J]. 电子学报, 2020, 48(3): 590-601.
LIU Y, LIU H Y, FAN J L, et al. A survey of research and application of small object detection based on deep learning[J]. Acta Electronica Sinica, 2020, 48(3): 590-601. (in Chinese)
- [5] TONG K, WU Y Q, ZHOU F. Recent advances in small object detection based on deep learning: A review[J]. Image and Vision Computing, 2020, 97: 103910.
- [6] YANG F, CHOI W, LIN Y Q. Exploit all the layers: Fast and accurate CNN object detector with scale dependent pooling and cascaded rejection classifiers[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 2129-2137.
- [7] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single Shot MultiBox Detector[C]//European Conference on Computer Vision. Cham: Springer, 2016: 21-37.
- [8] CAI Z W, FAN Q F, FERIS R S, et al. A unified multi-scale deep convolutional neural network for fast object detection[C]//European Conference on Computer Vision. Cham: Springer, 2016: 354-370.
- [9] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 936-944.
- [10] LIANG Z W, SHAO J, ZHANG D Y, et al. Small object detection using deep feature pyramid networks[C]//Pacific Rim Conference on Multimedia. Cham: Springer, 2018: 554-564.
- [11] GHIASI G, LIN T Y, LE Q V. NAS-FPN: Learning scalable feature pyramid architecture for object detection[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2019: 7029-7038.
- [12] WU J X, LIU S T, HUANG D, et al. Multi-Scale Positive Sample Refinement for Few-Shot Object Detection[C]//Computer Vision - ECCV 2020. Cham: Springer International Publishing, 2020: 456-472.
- [13] 李雅倩, 盖成远, 肖存军, 等. 基于细化多尺度深度特征的目标检测网络[J]. 电子学报, 2020, 48(12): 2360-2366.
LI Y Q, GAI C Y, XIAO C J, et al. Object detection networks based on refined multi-scale depth feature[J]. Acta Electronica Sinica, 2020, 48(12): 2360-2366. (in Chinese)
- [14] ZHANG S F, ZHU X Y, LEI Z, et al. S3FD: single shot scale-invariant face detector[C]//2017 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 192-201.
- [15] KISANTAL M, WOJNA Z, MURAWSKI J, et al. Augmentation for small object detection[EB/OL]. (2019-02-19). <https://arxiv.org/abs/1902.07296>.
- [16] CHEN Y K, ZHANG P Z, LI Z M, et al. Dynamic scale training for object detection[EB/OL]. (2020-04-26). <https://arxiv.org/abs/2004.12432>.
- [17] 李宝奇, 贺昱曜, 王伟, 等. 基于并行附加特征提取网络的SSD地面小目标检测模型[J]. 电子学报, 2020, 48(1): 84-91.
LI B Q, HE Y Y, QIANG W, et al. SSD with parallel additional feature extraction network for ground small target detection[J]. Acta Electronica Sinica, 2020, 48(1): 84-91. (in Chinese)
- [18] ZHOU X Y, WANG D Q, Krähenbühl Philipp, et al. Objects as Points [EB/OL]. (2019-04-16). <https://arxiv.org/abs/1904.07850>.
- [19] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//2017 IEEE International Conference on Computer Vision(ICCV). Piscataway: IEEE, 2017: 2999-3007.
- [20] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and

- Machine Intelligence, 2017,39(6): 1137-1149.
- [21] CAI Z W, VASCONCELOS N. Cascade R-CNN: Delving into high quality object detection[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 6154-6162.
- [22] DAI J, LI Y, HE K, et al. R-FCN: object detection via region-based fully convolutional networks[C]//Proceedings of the 30th International Conference on Neural Information Processing Systems. Red Hook: Curran Associates Inc, 2016: 379-387.
- [23] BOCHKOVSKIY A, WANG C Y, LIAO H, et al. YOLOv4: Optimal speed and accuracy of object detection [EB/OL]. (2020-04-23). <https://arxiv.org/abs/2004.10934>.
- [24] FU C Y, LIU W, RANGA A, et al. Dssd: Deconvolutional single shot detector[EB/OL]. (2017-01-23). <https://arxiv.org/abs/1701.06659>.
- [25] LAW H, DENG J. CornerNet: Detecting objects as paired keypoints[C]//Computer Vision - ECCV 2018. Cham: Springer International Publishing, 2018: 765-781.
- [26] ZHOU X Y, KOLTUN V, KRÄHENBÜHL P, et al. Probabilistic two-stage detection[EB/OL]. (2021-03-12). <https://arxiv.org/abs/2103.07461>.

作者简介



琚长瑞 男,1994年9月出生,安徽桐城人,陆军炮兵防空兵学院硕士研究生,主要研究方向为计算机视觉.

E-mail: juchangrui1994@163.com



秦晓燕 女,1980年2月出生,安徽淮北市人,副教授.主要研究方向为目标检测、机器学习及应用.

E-mail: 70853559@qq.com



袁广林(通讯作者) 男,1973年生,博士,副教授,主要从事计算机视觉、机器学习及其应用方面的研究.

E-mail: 1183212999@qq.com