

# 基于改进 Faster R-CNN 的 Logo 目标检测方法<sup>①</sup>



黄明珠, 黄文清

(浙江理工大学 信息学院, 杭州 310018)

通讯作者: 黄明珠, E-mail: 851489278@qq.com

**摘要:** 社交网络的发展和经济全球化赋予了 Logo 巨大的商业价值, 让 Logo 检测有很大的应用前景. 而现实情况中 Logo 目标在图片中往往占据很小的一部分, Logo 的低分辨率导致检测性能难以进一步提升. 因此本文提出了一种基于改进 Faster R-CNN 的 Logo 检测方法. 该方法在 Faster R-CNN 框架中结合了生成对抗模型, 利用网络先将分辨率较低的 Logo 特征映射成高分辨率的表达能力更强的特征, 再送入完全连接层进行分类和回归, 从而提高检测的性能. 本文在公开的 Logo 数据集上进行了实验结果评估, 结果表明了本文提出的方法能有效地提高 Logo 物体检测的准确率的同时也没有因为结构的变化影响检测速度.

**关键词:** faster R-CNN; Logo 检测; 低分辨率; 生成对抗网络; 特征映射

引用格式: 黄明珠, 黄文清. 基于改进 Faster R-CNN 的 Logo 目标检测方法. 计算机系统应用, 2019, 28(2): 41-48. <http://www.c-s-a.org.cn/1003-3254/6766.html>

## Logo Object Detection Method Based on Improved Faster R-CNN

HUANG Ming-Zhu, HUANG Wen-Qing

(School of Information Science and Technology, Zhejiang Sci-Tech University, Hangzhou 310018, China)

**Abstract:** The development of social networks and economic globalization gives logo a great commercial value, which makes logo detection have a good application prospect. In fact, logo objects typically occupy a small portion of the image, and the low resolution of the logo makes it difficult to further improve detection performance. Therefore, this study proposes an improved detection method based on Faster R-CNN. This approach combines the generative adversarial networks and a Faster R-CNN framework, uses the network to map lower resolution features to highly expressed high resolution features, and then sends them to fully connected layers for classification and regression. The outcomes of the experiment are evaluated on a publicly available logo dataset. The results show that the method can effectively improve the accuracy of logo object detection without affecting the detection speed of the basic network.

**Key words:** faster R-CNN; Logo detection; low resolution; generative adversarial networks; feature mapping

随着计算机视觉技术的发展, 该领域中的目标检测技术越来越多地应用于实际生活中. Logo 作为一种识别和传达信息的视觉图形, 是人们在长期生活和实践中形成的视觉化表达方式. 不管是商品还是企业或是政府部门等, 都有专属的标志 Logo. Logo 的检测有巨大的应用前景, 是计算机视觉中的重要任务. 目前

对于 Logo 检测的研究已有不少<sup>[1]</sup>, 例如文献[2]中基于 SIFT 特征进行匹配的方法来检索带有特定 Logo 的图像. 文献[3]首次提出了基于全局不变特征的 Logo 检测与识别算法. 以及基于前面的成果, 文献[4]尝试利用一些约束条件, 如圆和点或是圆和线的关系等, 并且利用其中的文字信息进行更为精确的识别. 以及

① 收稿时间: 2018-08-10; 修改时间: 2018-09-06; 采用时间: 2018-09-11; csa 在线出版时间: 2019-01-28

文献[5]所做的主要针对文本图片中 Logo 目标的检测研究。

Logo 检测任务可看成是具一般性的目标检测任务。在卷积神经网络 (CNN) 广泛应用于物体检测以前, 以往的方法通常是通过人工选定特征描述符的传统检测手段, 这些方法有很多的局限性, 在实际中得不到很好的应用。并且这类传统的检测方法已达到了瓶颈。然而随着卷积网络在目标检测技术方面的发展, 各类研究成果被不断提出。如文献[6]中的 R-CNN 采用选择性搜索来查找包含高概率对象的预测框, 使用卷积层提取每个预测框的高维特征, 由支持向量机 (SVM) 确定目标类别。但 R-CNN 中因为每个预测框都有一些重叠, 存在大量重复计算的缺点。为了缓解这个问题, 文献[7]提出 Fast R-CNN 网络, 通过引入一个感兴趣区域池化层 (ROI Pooling) 来共享特征的计算。该网络使用整个图像作为输入来提取全局特征图, 然后感兴趣区域池化层从它们中提取每个预测框的固定长度特征。然而, Fast R-CNN 因采用选择性搜索导致过程分离, 限制了检测速度进一步提升。因此, 文献[8]提出了区域提名网络 (RPN) 以进一步提高检测速度, 相应的方法称为 Faster R-CNN。虽然 RPN 依赖于几百个预测框就能达到和 Fast R-CNN 上千个预测框相同的检测效果。但是在分辨率低的目标的检测效果上, Faster R-CNN 相较于 Fast R-CNN 是有所下降的。但是 Faster R-CNN 在检测精度和检测速度上取得了一个很好的平衡。此外, 在上述的检测方法中都应用了边框回归方法以提高检测精度<sup>[6-8]</sup>。

除了上面提到的方法之外, 还有一类目标检测方法不采用区域提名方法寻找预测框。如文献[9]中提到的 YOLO 网络, 它们的特点是直接从整个特征图上选取预测框和预测概率。并且它在检测时加入全局上下文信息, 很少在背景类的分类中出错, 并具有良好的泛化能力。但是, YOLO 对物体的比例非常敏感。它通过在每个单元格上滑动来预测目标框。如果单元格太小, 则无法提高速度; 如果单元格太大, 则难以检测到物体。为了解决这些弱点, 文献[10]中提到的 SSD 网络在每层卷积之后添加了池化操作, 逐层递减特征尺度以形成特征图金字塔。然后, 在每层特征图上的单元格滑窗提取预测框。这不仅改善了 YOLO 在预测框选择上的限制, 还保证了速度。

目前, 已提出的各类方法基本上都偏适用于中大

型目标的检测, 小目标的检测效果并不理想, 这里提到的小目标指的是分辨率低的目标。小目标在图像中往往占据很小的一部分, 通过卷积之后它的特征会变得更小, 特征的分辨率也就更低。这样的目标特征不明显而难以正确检测, 而相比之下中大型的目标卷积后的高分辨率的特征更加明显所以好检测。虽然不少学者针对小目标检测性能的改善要求, 已提出相应的解决办法<sup>[11-13]</sup>。但这些方法基本上都是通过获取多尺度特征来改善小物体的检测性能。这类方法不仅大大降低了检测速度, 而且不能保证提取的特征足以有效检测小物体。

为此, 我们提出一种设想: 如果能基于现有的对中大型目标已达到良好检测效果的网络框架, 通过改进, 使其能将小目标的特征映射成具有相似特征分布的高分辨率的大目标的特征, 那么就能提升 Logo 的检测性能。基于这一考虑, 本文针对现有的各类方法的分析结果和 Logo 目标的特点, 选取 Faster R-CNN 作为改进框架。Faster R-CNN 对中大型的目标具有良好的检测效果, 但对小目标的检测效果不佳<sup>[14,15]</sup>。我们将生成对抗模型应用到 Faster R-CNN 的网络架构中, 通过对抗训练方式, 使网络具备将低分辨率的特征图映射到高分辨率的特征图的能力, 从而提高整个网络的检测性能。

## 1 改进的 Faster R-CNN 网络结构

### 1.1 经典的 Faster R-CNN 的网络模型

Faster R-CNN 是经过 R-CNN 和 Fast R-CNN 的积淀, 由 Ross B. Girshick 在 2016 年提出的。它在 Fast R-CNN 的基础上引入区域提名网络 (RPN) 来提取预测框, 解决了采用选择性搜索导致的过程分离问题, 并通过交替训练使 RPN 和 Fast R-CNN 网络共享参数。因此在结构上, Faster R-CNN 已经将特征提取、预测框提取、边框回归 (bounding box regression)、分类都整合在了一个网络中, 使得综合性能有较大提高, 尤其在检测速度方面。Faster R-CNN 网络流程图如图 1 所示。

Faster R-CNN 的结构主要包含卷积/池化、RPN 网络、感兴趣区域池化层和分类器。首先由卷积/池化部分提取图片的特征图, 接着由 RPN 网络提取候选框, 然后将候选框和特征图输入感兴趣区域池化层提取各候选框的特征, 最后由分类器实现目标识别与定位。

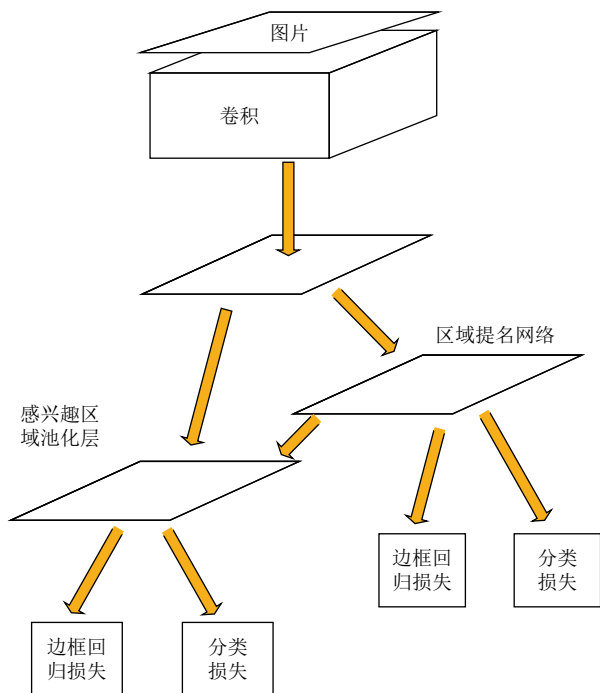


图1 经典的 Faster R-CNN 的检测流程图

RPN 网络是 Faster R-CNN 的核心部分, 类似于以往目标检测中的选择性搜索的作用, 通过卷积神经网络来选取候选框。由于图片中的目标的尺度和宽高比不一, 需要多种类型的窗口。因此, RPN 采用 anchor 机制, 即设定一个基准窗口大小, 按照 (8,16,32) 三种倍数和 (0.5,1,2) 三种比例得到 9 种尺度的窗口。通过在输入的特征图上滑动窗口, 即可得到关于这张图片的目标预测框。RPN 网络的输出是一个包含二类分类和边框回归的多任务模型。二类分类通过 Softmax 输出预测框属于前景类还是背景类, 边框回归则是用于计算预测框的偏移量, 以便获得更加准确的定位。最后, 综合预测框的概率和偏移量, 运用非极大值抑制 (NMS) 去除太小和超出边界的预测框, 然后将预测框送入完全连接层做进一步的分类和边框回归。

RPN 网络的提出虽然帮助 Faster R-CNN 加快了速度, 但是由于 RPN 只依赖几百个预测框选取目标位置, 在目标的召回率上必定有所下降。并且 RPN 网络的窗口类型有限难以满足多种尺度目标的检测需求。鉴于目前大部分数据集中的目标尺度适中, 小型目标偏少, 所以 Faster R-CNN 在窗口的设置上偏于适应中大型目标的检测, 对于小目标的检测效果并不理想。

## 1.2 对抗生成网络模型 (GANs)

生成对抗网络 (GANs)<sup>[16]</sup> 是用于学习生成模型的框

架, 由 Ian Goodfellow 于 2014 年首次提出。它并不一定要以卷积网络构成, 但已提出的各种用途的生成对抗网络基本上都是卷积网络。Mathieu 等人<sup>[17]</sup> 和 Dentonet 等人<sup>[18]</sup> 采用了 GANs 实现图像生成。在文献[19]和文献[20]中, GANs 是分别用于学习从一个流形到另一个流形的风格转移和修复的映射。除此之外, 使用 GANs 进行无监督表征学习在文献[21]中被提出。文献[22]中提出将 GANs 应用到超分辨率图像。

生成对抗网络是由竞争的两个神经网络模型组成。一个将噪声作为输入并生成样本, 所以称为生成器, 用  $G$  来表示。另一个模型称为判别器, 用  $D$  来表示。它的输入有两种, 一种是生成器生成的样本, 另一种是训练数据中的真实样本。生成器的目的是生成能欺骗判别器的样本, 判别器的目的是能够区分样本是生成的还是真实的样本。这两个网络的训练过程是對抗学习, 并且两个网络同时训练, 最后当生成器和判别器的损失函数达到一个平衡 (纳什平衡) 时, 即  $G$  可以生成足以“以假乱真”的样本  $G(z)$ 。对于  $D$  来说, 它难以判定  $G$  生成的图片究竟是不是真实的, 因此  $D(G(z))=0.5$ 。生成对抗网络模型的流程图如图 2 所示。

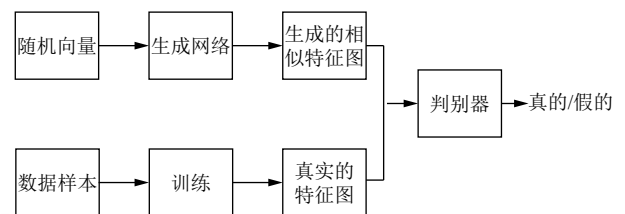


图2 生成对抗网络流程图

## 1.3 改进的 Faster R-CNN

本文基于 Faster R-CNN 网络对中大型目标已有良好的检测效果而对小目标的检测效果不佳的情况, 根据生成对抗网络可生成高分辨率图像的特性, 提出了一种新的检测器。网络的具体结构如图 3 所示, 整个网络可以分成两个子网络。上层子网络是由生成器和判别器组成的对抗网络, 用于为低分辨率目标生成适当的特征偏差。下层子网络是由 Faster R-CNN 框架构成的感知网络, 用于目标分类和回归。

对抗网络中, 生成器以 Faster R-CNN 的第一层卷积输出的特征作为输入, 然后进行  $3 \times 3$  的卷积和  $1 \times 1$  的卷积处理。这两次卷积处理的主要目的是为了与感知网络五层  $3 \times 3$  的卷积的最终输出的 256 个特征通

道保持一致. 五个残差块依次连接作为生成器的主干部分, 用于学习低分辨率目标特征和高分辨目标率特征的特征偏差, 再与下层的 Faster R-CNN 中的五层卷积得到的特征按像素相加得到高分辨率的特征. 每个残差块都是由 3×3 的卷积层, 批量规范化层以及 ReLU

激活函数层组成. 判别器的输入包括生成的高分辨率目标特征和真实的高分辨率目标特征两类, 生成的高分辨率目标特征作为负样本, 真实的高分辨率目标特征作为正样本, 通过为两种样本安排不同的标记帮助判别器学习区分真伪数据.

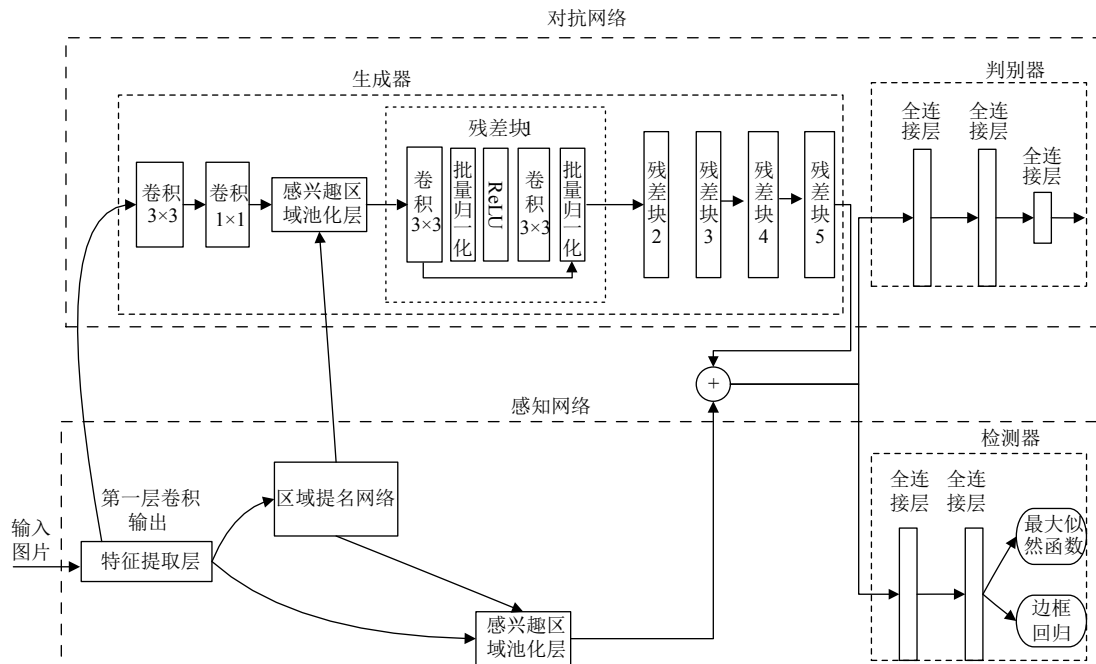


图3 改进的 Faster R-CNN 网络框架图

感知网络结构采用的是 Faster R-CNN 的检测框架, 以五层卷积进行特征提取. RPN 网络在特征图上选取预测框后会送入检测器做进一步的分类和回归.

## 2 改进的 Faster R-CNN 的训练过程

整个网络的训练过程是不断交替的迭代过程. 每次迭代的训练过程如图 4 所示, 可分为如下三步:

1) 训练 RPN 网络提取目标预测框.

2) 训练判别器和检测器: 用高分辨的图像训练感知网络和判别器, 判别器利用判别损失  $L_{dis}$  反向调参判别网络, 感知网络利用感知损失  $L_p$  反向调参感知网络, 相应的损失表达式将在下面具体介绍.

3) 判别器参数保持不变, 训练生成器和检测器: 用低分辨率的图像一起训练生成器和感知网络, 生成器会为低分辨率的目标生成合适的特征偏差, 与感知网络卷积后的目标特征按像素级相加得到生成的高分辨率的目标特征, 送入判别器和检测器. 此时因为生成器

和判别器的训练过程是相互对抗的, 为了提高生成器的生成能力, 判别网络不做反向调参. 我们利用生成损失  $L_g$  和感知损失  $L_p$  一起监督整个高分辨率特征生成过程, 相应的损失函数为  $L=w_1L_g+w_2L_c$ . 其中,  $w_1, w_2$  是两个权重. 在这里我们将它们设置为 1.

交替训练过程中, 生成器不断学习为低分辨率的目标特征生成合适的特征偏差, 判别器不断学习区分生成的高分辨率特征和真实的高分辨率特征. 通过迭代训练会不断加强生成器和判别器的能力, 从而提高低分辨率的小目标的检测准确率.

### 2.1 对抗损失

对于生成器, 我们的目标是训练一个生成函数  $G$  将  $f_s$  转换成  $f_l$  来欺骗判别器  $D$ . 这里  $f_l$  和  $f_s$  分别表示高分辨率的目标特征和相对应的低分辨率的目标特征. 然而, 生成器可能很难从  $f_s$  中包含的有限信息中学习直接生成高分辨率的特征. 因此, 从感知网络的低层卷积中引入特征  $f$  到生成网络中来学习高分辨率特

征和相对应的低分辨率特征之间的特征偏差,通过和 Faster R-CNN 卷积后的特征做像素级的叠加生成最终的高分辨率特征.在本文的设计中,我们的目标是改善小目标的检测性能,所以只需为低分辨率的目标特征做高分辨率特征映射.对于高分辨率目标我们需要生成器生成的特征偏差为零.因此,相应的生成损失函数定义如下:

$$L_g = -\log D(f_s + G(f_s|f)) \quad (1)$$

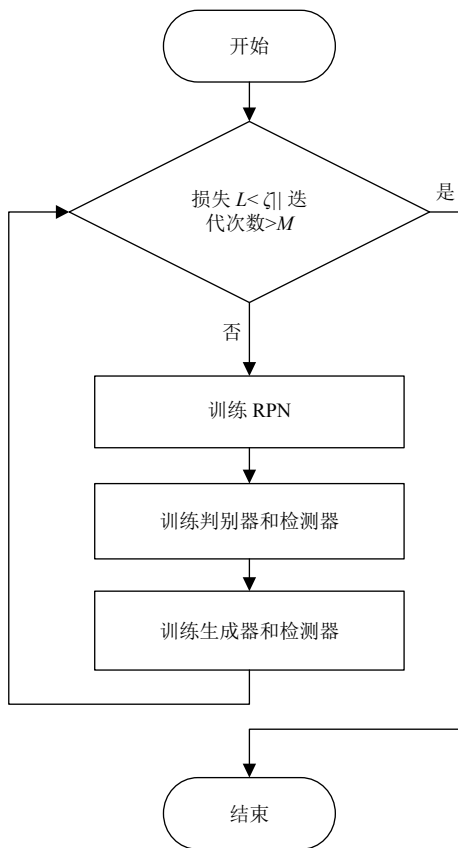


图4 改进的 Faster R-CNN 的训练流程图

对于判别器,对于输入的真实的高分辨率特征和生成的高分辨率特征,需要通过判别损失来反向调参,使判别器通过训练对于输入的真伪数据学习区分.显然当  $L_{dis}$  足够小时,判别器将能够区分所生成的特征与真实的特征之间的差异.最后,判别损失可以描述为:

$$L_{dis} = -\log D(f_i) - \log(1 - D(f_s|f)) \quad (2)$$

当对抗训练达到平衡时,即判别网络无法区分输入是真实特征图还是生成特征图时,此时的生成器已经可以为低分辨率的小目标生成以假乱真的高分辨率特

征,生成器的训练可以停止.

### 2.2 感知损失

在感知网络训练过程中,从特征图提取的每个预测框被送入检测器,得到两类输出.第一个输出层由 softmax 回归计算  $n+1$  个类中每个预测框的概率  $p=(p^0, p^1, \dots, p^n)$ .第二个输出层计算预测框坐标  $r^n=(r_x^n, r_y^n, r_w^n, r_h^n)$ .然后,使用公式(3)定义的多任务损失函数  $L_p$  对预测框的类型和坐标进行回归计算,同时计算得到预测框的坐标偏移.

$$L_p = L_{cls}(p, g) + \lambda[g \geq 1]L_{loc}(r^g, r) \quad (3)$$

其中,  $L_{cls}(p, g) = -\log p_g$  是真实类别  $g$  的对数损失.  $L_{loc}$  是边界回归损失函数,定义如下:

$$L_{loc}(r^g, r) = \sum_{i \in (x, y, w, h)} S_{L1}(r_i^g - r_i) \quad (4)$$

其中,  $S_{L1}$  是平滑损失函数,定义如下.

$$S_{L1} = \begin{cases} 0.5x^2, & |x| < 1 \\ |x| - 0.5, & \text{othersize} \end{cases} \quad (5)$$

训练结束时,每个训练预测框都被标记一个真值类别和真值框的目标位置.

## 3 实验结果分析

### 3.1 数据集

为了尽可能保证使用大数据量的原始图片进行训练,本文的原始图像来源于 Flickrlogos-32 和 BelgaLogos 数据集,它们都是互联网上公开的数据集资源.

在使用数据集训练网络前,为了顺利训练生成对抗网络,本文先对数据集进行了相应的预处理.预处理的步骤为:首先,选取两个数据集中目标尺寸大于  $64 \times 64$  的图片作为高分辨训练样本;接着,通过下采样使其尺寸小于  $32 \times 32$ ,作为低分辨率训练样本.测试样本从两个数据集中随机选取并不做预处理.

按照上述步骤,获得训练样本 7532 张,测试集 4000 张.然后,按照 Pascal VOC 2007 的数据格式对获取的训练样本数据集和测试集进行相应的转换工作.

因为 Logo 的种类众多,并且获取的每种类别拥有的图片量不一,有的数量太少不足以用来进行更进一步的细化分类训练.因此,所有本文的 Logo 目标检测只做两类分类和定位,并不对 Logo 再细致分类.

### 3.2 网络参数设置

对于 Logo 检测,先采用 ImageNet 大型分类数据

集对基础网络做预训练, 获得五层卷积的参数初始化. 除此之外, 网络中添加其它结构均采用“Xavier”<sup>[23]</sup>进行参数初始化.

整个网络使用随机梯度下降 (SGD) 训练, 动量为 0.9, 权重衰减量为 0.0005; 残差块的数量设置为 5. 在训练期间, 25% 的样本是前景, 剩下的是背景. 预测框与真值框的交并比 (IOU) 至少 0.5.

### 3.3 方法比较

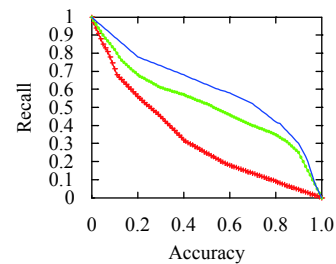
本文方法的测试是在 NVIDIA GeForce GTX 1070 GPU 和 Caffe 平台上实现的. 表 1 显示了本文的方法与其他方法在 Logo 检测上的平均准确率和召回率的比较. 从数据中, 我们可以明显看出, 改进的方法在检测性能方面优于 Faster R-CNN, 说明了我们的改进发挥了作用, 并且因为 Fast R-CNN 在候选框选取数量远远多于 Faster R-CNN, 所以在小物体的检测方面召回率要优于 Faster R-CNN. 然而, 在 Logo 的检测上, 我们的方法依赖于更少的预测框在小物体的检测效果上已经超过了 Fast R-CNN, 说明了本文提出的方法在小目标的检测上有优势.

表 1 本文方法与 Fast R-CNN 和 Faster R-CNN 在 Logo 测试集上的检测性能比较 (%)

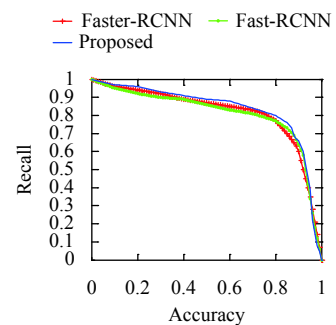
方法	Logo(像素<32×32)		Logo(All)	
	平均准确率	召回率	平均准确率	召回率
Fast R-CNN	57	76	68	86
Faster R-CNN	61	26	72	62
改进的方法	85	80	88	89

为了分析不同尺度下三种方法的性能, 我们将目标划分为两种类型的尺度: 像素小于  $32 \times 32$  小目标和像素大于  $32 \times 32$  的中大型目标. 通过 Accuracy-Recall 曲线比较它们在 Logo 数据集上的检测性能, 如图 5 所示. 总的来说, 我们的方法优于其他方法, 特别是在小物体的检测中, 证明了我们方法的有效性. 如图 5(a) 所示, Fast R-CNN 在小目标检测中优于 Faster R-CNN. 然而, 当将 GAN 引入 Faster R-CNN 时, 改进的 Faster R-CNN 比 Fast R-CNN 方法获得更好的性能. 它证明了生成器为小目标创造了适当的偏差, 使小目标的特征与大目标的特征相似, 从而获得了良好的性能. 对于大中型物体检测, 图 5 中所示的三种方法略有不同. 客观地说, Fast R-CNN 的性能比其他两种方法弱. 在测试

中型和大型物体时, Faster R-CNN 稍微弱于我们的改进方法. 为了更直观显示我们的检测效果, 我们随机选取了几张三种方法在 Logo 数据集上的检测效果图, 如图 6 所示. 图 6 中子图 (a)、(b)、(c) 分别代表的是 Fast R-CNN、Faster R-CNN 和改进的方法的检测效果图. 从图中可以看出, 相比于前两种方法, 改进的方法在定位和分类概率上都更加准确.



(a) 小目标检测



(b) 中大型目标检测

图 5 改进方法与 Fast R-CNN 和 Faster R-CNN 的 Accuracy-Recall 曲线图

## 4 结语

本文将 GAN 引入 Faster R-CNN 框架. 我们使用生成网络生成一个特征偏差, 使小目标的特征类似于大目标的特征, 从而欺骗了判别器. 通过这种策略, 与 Faster R-CNN 相比, 我们的方法有效地提高了小目标的检测性能. 尽管我们的方法使用 Logo 数据集进行训练和测试, 但是该方法的应用不只限于 Logo 的检测, 而且还可以用于其他情况下的小目标检测. 当然, 我们需要改进的地方还有很多. 两种模型的组合使网络结构更加复杂, 训练还需要分两个阶段完成, 训练时间和计算量的增加是不可避免的.



(a) Fast R-CNN (b) Faster R-CNN (c) 改进的方法

图6 改进的方法与 Fast R-CNN 和 Faster R-CNN 在测试集的检测效果图

参考文献

1 Li Z, Schulte-Austum M, Neschen M. Fast logo detection and recognition in document images. Proceedings of the 20th International Conference on Pattern Recognition. Istanbul, Turkey. 2010. 2716–2719.

2 Joly A, Buisson O. Logo retrieval with a contrario visual query expansion. Proceedings of the 17th ACM International Conference on Multimedia. Beijing, China. 2009. 581–584.  
 3 Doermann DS, Rivlin E, Weiss I. Logo recognition using geometric invariants. Proceedings of the 2nd International

- Conference on Document Analysis and Recognition. Tsukuba Science City, Japan. 1993. 894–897.
- 4 Doermann D, Rivlin E, Weiss I. Applying algebraic and differential invariants for logo recognition. *Machine Vision and Applications*, 1996, 9(2): 73–86. [doi: [10.1007/BF01214362](https://doi.org/10.1007/BF01214362)]
  - 5 Hassanzadeh S, Pourghassem H. A fast logo recognition algorithm in noisy document images. *Proceedings of 2011 International Conference on Intelligent Computation and Bio-Medical Instrumentation*. Wuhan, China. 2012. 64–67.
  - 6 Girshick R, Donahue J, Darrell T, *et al.* Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, OH, USA. 2014. 580–587.
  - 7 Girshick R. Fast R-CNN. *Proceedings of 2015 IEEE International Conference on Computer Vision*. Santiago, Chile. 2015. 1440–1448.
  - 8 Ren SQ, He KM, Girshick R, *et al.* Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137–1149. [doi: [10.1109/TPAMI.2016.2577031](https://doi.org/10.1109/TPAMI.2016.2577031)]
  - 9 Redmon J, Divvala S, Girshick R, *et al.* You only look once: Unified, real-time object detection. *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, NV, USA. 2016. 779–788.
  - 10 Liu W, Anguelov D, Erhan D, *et al.* SSD: Single shot MultiBox detector. *Proceedings of the 14th European Conference on Computer Vision-ECCV 2016*. Amsterdam, The Netherlands. 2016. 21–37.
  - 11 Bell S, Zitnick CL, Bala K, *et al.* Inside-outside net: Detecting objects in context with skip pooling and recurrent neural networks. *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, NV, USA. 2016. 2874–2883.
  - 12 Li HX, Lin Z, Shen XH, *et al.* A convolutional neural network cascade for face detection. *Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition*. Boston, MA, USA. 2015. 5325–5334.
  - 13 Yang F, Choi W, Lin YQ. Exploit all the layers: Fast and accurate CNN object detector with scale dependent pooling and cascaded rejection classifiers. *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, NV, USA. 2016. 2129–2137.
  - 14 Eggert C, Dan ZC, Brehm S, *et al.* Improving small object proposals for company logo detection. *Proceedings of 2017 ACM International Conference on Multimedia Retrieval*. Bucharest, Romania. 2017. 167–174.
  - 15 Li JN, Liang XD, Wei YC, *et al.* Perceptual generative adversarial networks for small object detection. *Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI, USA. 2017. 1951–1959.
  - 16 Goodfellow IJ, Pouget-Abadie J, Mirza M, *et al.* Generative adversarial nets. *Proceedings of the 27th International Conference on Neural Information Processing Systems*. Montreal, Canada. 2014. 2672–2680.
  - 17 Mathieu M, Couprie C, LeCun Y. Deep multi-scale video prediction beyond mean square error. *arXiv preprint arXiv:1511.05440*, 2015.
  - 18 Denton E, Chintala S, Szlam A, *et al.* Deep generative image models using a Laplacian pyramid of adversarial networks. *Proceedings of the 28th International Conference on Neural Information Processing Systems*. Montreal, Canada. 2015. 1486–1494.
  - 19 Li C, Wand M. Combining Markov random fields and convolutional neural networks for image synthesis. *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, NV, USA. 2016. 2479–2486.
  - 20 Yeh R, Chen C, Lim TY, *et al.* Semantic image inpainting with perceptual and contextual losses. *arXiv preprint arXiv:1607.07539*, 2016.
  - 21 Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.
  - 22 Ledig C, Theis L, Huszar F, *et al.* Photo-realistic single image super-resolution using a generative adversarial network. *Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI, USA. 2017. 105–114.
  - 23 Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics*. Sardinia, Italy. 2010. 249–256.