

基于卷积神经网络的多尺度 Logo 检测算法

江玉朝, 吉立新, 高超, 李邵梅

(国家数字交换系统工程技术研究中心, 河南 郑州 450002)

摘要: 针对自然场景图像中多尺度 Logo 的检测需求, 提出了一种基于卷积神经网络的多尺度 Logo 检测算法。该算法基于两阶段目标检测的实现思路, 通过构建特征金字塔并采取逐层预测的方式实现多尺度候选区域的生成, 通过融合卷积神经网络中的多层特征图以增强特征的表达能力。在 FlickrLogos-32 数据集上的实验结果显示, 相比基线方法, 所提算法能够提升生成候选区域的召回率, 并且在保证大中尺度 Logo 检测精度的前提下, 提升小尺度 Logo 的检测性能, 验证了所提算法的优越性。

关键词: Logo 检测; 卷积神经网络; 多尺度; 区域生成网络; 特征融合

中图分类号: TP391

文献标识码: A

doi: 10.11959/j.issn.2096-109x.2020026

Multi-scale Logo detection algorithm based on convolutional neural network

JIANG Yuchao, JI Lixin, GAO Chao, LI Shaomei

National Digital Switching System Engineering & Technological R&D Center, Zhengzhou 450002, China

Abstract: Aiming at the requirements for multi-scale Logo detection in natural scene images, a multi-scale Logo detection algorithm based on convolutional neural network was proposed. The algorithm was based on the realization of two-stage object detection. By constructing feature pyramids and adopting layer-by-layer prediction, multi-scale region proposals were generated. The multi-layer feature maps in convolutional neural networks were fused to enhance the feature representation. The experimental results on the FlickrLogos-32 dataset show that compared with the baseline, the proposed algorithm can improve the recall rate of region proposals, and can improve the performance of small Logo detection while ensuring the accuracy of large and middle Logo, proving the superiority of the proposed algorithm.

Key words: Logo detection, convolutional neural network, multi-scale, region proposal network, feature fusion

收稿日期: 2019-05-29; 修回日期: 2019-08-11

通信作者: 江玉朝, jingjiujiang@qq.com

基金项目: 国家自然科学基金资助项目 (No.61601513)

Foundation Item: The Nature Science Foundation of China (No.61601513)

论文引用格式: 江玉朝, 吉立新, 高超, 等. 基于卷积神经网络的多尺度 Logo 检测算法[J]. 网络与信息安全学报, 2020, 6(2): 116-124.

JIANG Y C, JI L X, GAO C, et al. Multi-scale Logo detection algorithm based on convolutional neural network[J]. Chinese Journal of Network and Information Security, 2020, 6(2): 116-124.

1 引言

在计算机视觉领域,多尺度目标检测一直是一项具有挑战性的基础课题。近年来,随着卷积神经网络(CNN, convolutional neural network)的发展,针对大中尺度的目标检测取得了重大突破,而对于小目标,由于受到小目标像素少、分辨率低、背景干扰等因素的影响,相应算法的检测性能仍十分受限。作为目标检测的一个特例,Logo 检测在品牌趋势预测、商标产权保护、车辆标志识别等领域有广泛的应用^[1-3]。

当前,基于 CNN 的多尺度目标检测主要有以下两类实现思路^[4]:一类是以 Faster R-CNN^[5]为代表的两阶段方法,此类方法首先生成一定数量的候选区域集合,然后使用分类算法对提取的候选区域进行类别判定和位置精修;另一类是以 SSD^[6]为代表的单阶段方法,此类方法直接在 CNN 的多尺度特征图上预测目标类别和边界框回归参数。与两阶段方法相比,单阶段方法在设计之初主要侧重于优化检测的速度,在算法精度方面始终与两阶段方法存在一定差距。因此,本文基于两阶段的实现思路研究自然场景图像中多尺度 Logo 目标的检测问题,即第一阶段利用候选区域生成算法提取出场景图像中可能包含 Logo 目标的子区域;第二阶段对该子区域进行类别判定并做进一步的位置精修。显然,对于这样的检测流程,第一阶段生成候选区域的优劣将直接影响检测算法的性能。

作为两阶段目标检测的代表方法, Faster R-CNN^[5]通过构建区域生成网络(RPN, region proposal network)生成候选区域。具体而言, RPN 在基础特征提取网络之后的单尺度特征图上通过预定义的多尺度锚点(anchor)进行密集采样得到初始候选区域集合,而后经前景分类筛选、边界框回归、非极大值抑制等步骤得到最终的候选区域。这其中存在两个问题:① 单尺度特征图的感受野难以适应多尺度目标,造成与感受野尺度不匹配目标的检测性能受限;② 预定义的锚点尺度未能适应小目标,导致小目标在 RPN 训练阶段生成的正负样本数量严重不平衡。

为解决上述问题并适应 Logo 检测应用场景,本文提出了一种基于 CNN 的多尺度 Logo 检测算法。算法在 Faster R-CNN 的基础上,通过构建特征金字塔并采取逐层预测的方式实现多尺度候选区域的生成,利用理论感受野指导特征金字塔层级的设计,结合有效感受野^[7-8]以及 Logo 对象的先验知识指导预定义锚点的设计,使特征金字塔的每层用于生成特定尺度的候选区域。此外,本文借鉴现今流行的多尺度特征融合思想,在基础网络和目标检测网络之间增加了多尺度特征融合网络,用于融合多层次特征以增强目标检测网络输入特征的表达能力。实验方面,本文在 Logo 检测的 benchmark 数据集 FlickrLogos-32^[9]上进行了详细的算法验证,相比于基线方法^[10],本文算法的实验结果取得了明显优势(mAP 85.7% VS 81.1%^[10]),验证了本文算法的优越性。

2 相关工作

2.1 多尺度目标检测

现阶段存在很大一部分工作致力于解决多尺度目标的检测问题。图像金字塔思想通过对原始图像进行简单缩放构造图像的多尺度副本,进而实现多尺度目标检测^[11-13],通常构造图像金字塔能够显著提升算法的检测精度,但与之而来的庞大内存和时间开销往往令许多实际应用难以承受。鉴于 CNN 的层次结构本身具有天然的多尺度金字塔形状,一些方法^[6,14]选择直接在特征金字塔各层上预测特定尺度的目标,此类单阶段方法一般在速度上具有明显优势,但浅层的高分辨率特征通常难以支撑细粒度的目标分类,导致针对小目标的检测效果一般。因此,文献[15-16]采用多尺度特征融合的方式弥补小目标检测性能的不足,此类算法通过融合语义信息丰富的高层特征和高分辨率的浅层特征,提升了多尺度目标检测的整体性能。

2.2 Logo 检测

随着 CNN 在计算机视觉领域取得的革命性突破,一些研究学者开始将 CNN 应用在 Logo 检测任务中。Iandola 等^[17]较早地开展了这项工作,通过 Fast R-CNN^[18]建立了使用 CNN 进行 Logo 检测的精度基准。Zhang 等^[3]提出多尺度平行卷

积网络用于车辆 Logo 的识别。为了促进深度学习方法在 Logo 检测问题中的研究, Hoi 等^[19]构建了大规模 Logo 检测与商标识别数据集 LOGO-Net。文献[10,20]则从大规模合成 Logo 图像的角度缓解训练样本缺乏的问题。Christian 等^[21]通过改进 Faster R-CNN 中候选区域的生成机制, 有针对性地提升了小尺度 Logo 的检测性能。

3 基于 CNN 的多尺度 Logo 检测算法

图 1 给出了本文基于 CNN 的多尺度 Logo 检测算法的整体框架, 该框架源于两阶段目标检测的实现思路, 算法主要分为基础特征提取网络、多尺度区域生成网络、多尺度特征融合网络、目标检测网络 4 个部分。其中, 图示基础特征提取网络为经典 VGG16 的全部特征提取层, 多尺度区域生成网络负责生成输入图像中可能包含 Logo 目标的候选区域, 多尺度特征融合网络用于整合多层次特征以增强特征的表达能力, 目标检测网络实现针对候选区域的类别预测和边界框回归。

3.1 多尺度区域生成网络

从结构上来说, 本文算法中多尺度区域生成网络为多级二分类回归的全卷积网络 (如图 1 所示), 在层次结构上与 SSD^[6]、S³FD^[14]等相似, 均在基础特征提取网络后添加了额外的

卷积层, 并选取不同尺度的特征层预测特定尺度的目标。不同的是, SSD 各预测层直接进行多类目标检测, 而 S³FD 针对的是人脸检测。对于 Logo 目标, 本文算法在设计多尺度区域生成网络时考虑了两个因素: ① 多层 CNN 的感受野与 Logo 目标尺度的适应关系, 这直接关系到网络层次结构的设计; ② 预定义锚点与 Logo 目标尺度的适应程度, 这直接关系到网络的精度和收敛速度。

根据 CNN 感受野理论, CNN 某层单位感受区域之外的任何位置都不会影响该层相应单元的输出结果。因此, 在设计多尺度区域生成网络时, 最后一层卷积特征图上预测单元所对应的理论感受野应当覆盖绝大多数尺度的目标, 此项约束实际上也决定了基于 CNN 的目标检测算法所能预测的目标尺度上限。表 1 中计算了本文多尺度区域生成网络各预测层经过 3×3 卷积操作之后的理论感受野, 显然经过一系列卷积和池化操作, 高层的理论感受野逐步增大。图 2 给出了 FlickrLogos-32 数据集对象大小的整体分布情况, 结合表 1 中相关参数可以看出, 预测层 Conv7_2 的理论感受野能够覆盖该数据集中绝大多数尺度的目标, 这在理论上保证了本文所设计的多尺度区域生成网络与数据集的目标尺度相适应。

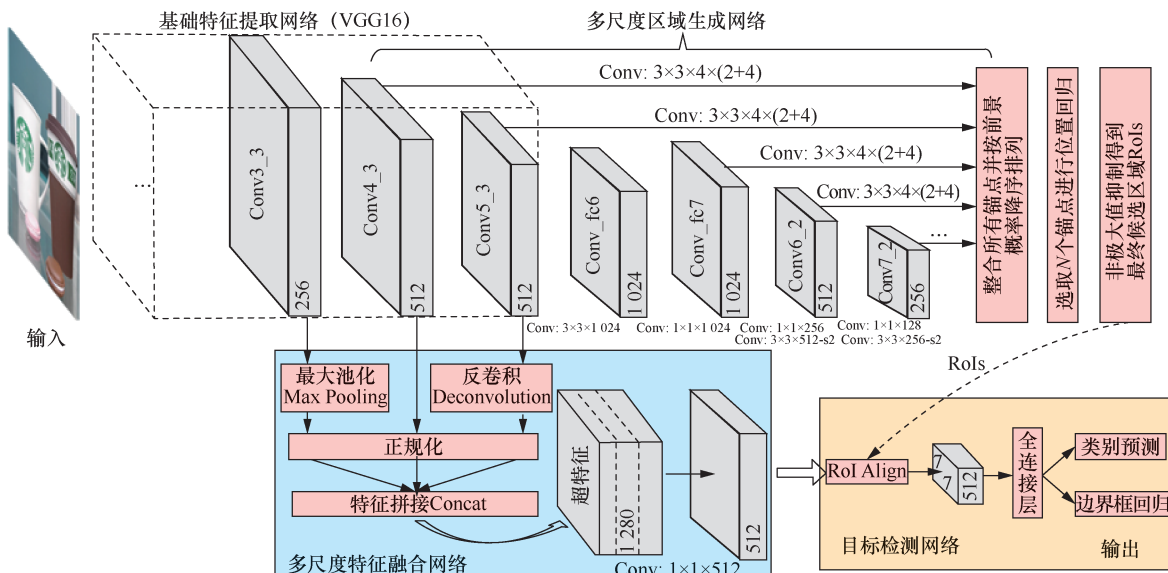


图 1 基于 CNN 的多尺度 Logo 检测算法
Figure 1 Multi-scale Logo detection algorithm based on CNN

表 1 本文多尺度区域生成网络各预测层的相关参数
Table 1 Parameters of prediction layer of the multi-scale region proposal network

预测层	步长	理论感受野	锚点尺度	锚点高宽比
Conv4_3	8	108	32	1/3
Conv5_3	16	228	64	
Conv_fc7	32	340	128	1/2
Conv6_2	64	564	256	1
Conv7_2	128	724	512	2

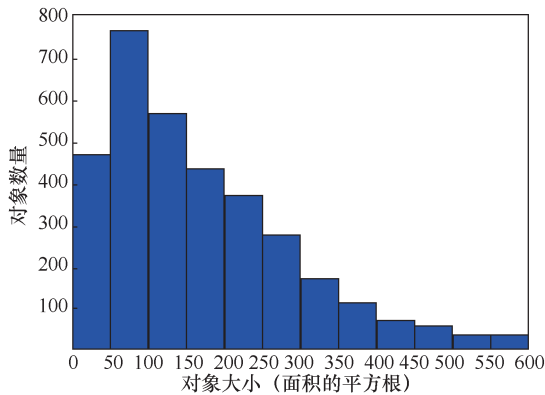


图 2 FlickrLogos-32 数据集对象大小分布
Figure 2 Distribution of object instance sizes in FlickrLogos-32

在确定网络层次结构的基础上，算法进一步考虑各预测层上预定义的锚点尺度和高宽比。根据 CNN 的有效感受野理论^[7-8]，CNN 的有效感受野满足高斯分布，仅占据理论感受野中心区域的较小部分（如图 3 所示，整体黑色区域为理论感受野，而白色点云区域为有效感受野，图片引自文献[7]）。本文为各预测层预设的锚点尺度如表 1 所示，所设置的锚点尺度均在一定程度上小于相应层的理论感受野，同时采取 S^3FD 中锚点设置的等比例间隔策略，即相应预测层的锚点尺度与该层的步长始终保持等比例，该策略能够保证不同预测层的锚点具备同样的空间密度分布。

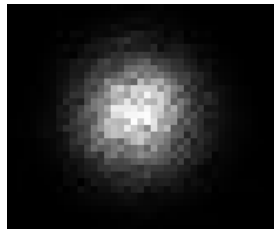


图 3 理论感受野与有效感受野关系
Figure 3 Relationship between theoretical receptive fields and effective receptive fields

在锚点高宽比的设置方面，图 4 展示了 FlickrLogos-32 数据集中 32 类 Logo 的官方示例，表 2 近似统计了该数据集中 32 类官方 Logo 以及所有 Logo 对象的高宽比，本文发现，仅从官方 Logo 的统计层面就能够对锚点的高宽比设置产生较强的指导意义。事实上，由于 Logo 对象平面性的设计原则，自然场景图像中出现的 Logo 通常具备较稳定的高宽比。进一步考虑实际场景中 Logo 目标可能存在旋转、扭曲等情况，本文预设的锚点高宽比如表 2 所示。



图 4 FlickrLogos-32 官方 Logo 示例
Figure 4 Exemplar images of Logo classes from the FlickrLogos-32

表 2 FlickrLogos-32 中官方 Logo 以及所有 Logo 对象高宽比的近似统计

目标数据集	高宽比				
	$\frac{1}{3}$	$\frac{1}{2}$	1	2	3
官方 Logo (32)	6	6	20	0	0
FlickrLogos-32 (3 404)	315	775	1 853	416	45

基于上述设计，本文多尺度区域生成网络通过大小为 $3 \times 3 \times 4 \times (2+4)$ 的滑动卷积核（如图 1 所示），针对各预测层中的每个位置分别预测 4 种不同高宽比的单尺度锚点的二分类概率（前景/背景）和 4 个位置回归参数。因而在网络训练阶段，各预测层应当配备与其预测尺度相对应的足量训练样本，以保证网络各层的充分训练，然而实际情况下，可用于训练的标记样本集通常难以满足该训练模式提出的较高要求。

因此，本文选择采取与 SSD 类似的粗放式正负样本生成策略，即对于正样本，算法允许任意标记样本在网络产生的全体锚点范围内进行匹配操作。具体而言，算法首先为每个真值框（ground truth）匹配与之 IoU（intersection over union）值

最大的锚点作为正样本，然后在剩余锚点中挑选与任一真值框 IoU 值高于 0.5 的也作为正样本。对于负样本，算法依据常用的难分负样本挖掘方法，首先筛选出与任意真值框 IoU 值都低于 0.3 的锚点子集作为初始负样本集合，接着按照预测背景概率由低到高的顺序对集合中的所有锚点进行排序，选取概率较低的锚点作为最终的负样本，并限制最终选定的正负样本数量之比不超过 1:3，以保证网络的快速收敛^[6,14]。

在训练目标方面，本文基本沿用原始 RPN^[5]中定义的多任务损失函数。由此，对于单个训练样本，本文多尺度区域生成网络的损失函数定义为

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*)$$

其中， i 为参与网络训练的正负样本集合中某锚点的索引， p_i 为锚点 i 的前景预测概率， $p_i^* \in \{0, 1\}$ 为锚点 i 对应的真实前景概率（若锚点 i 为正样本，则 p_i^* 为 1，否则为 0）， $t_i = (t_{ix}, t_{iy}, t_{iw}, t_{ih})$ 为预测边界框相对于锚点 i 的四维位置回归参数向量^[22]， t_i^* 为相应真值框相对于锚点 i 的真实位置回归参量（仅对正样本进行位置精修）。分类损失 $L_{cls}(p_i, p_i^*) = -[p_i^* \log p_i + (1 - p_i^*) \log(1 - p_i)]$ 为二分类 Softmax 交叉熵损失，回归损失 $L_{reg}(t_i, t_i^*) = \sum_{j \in \{x, y, w, h\}} smooth_{L1}(t_{ij} - t_{ij}^*)$ 为 Smooth L1 损失^[19]，两类损失分别由系数 N_{cls} 和 N_{reg} 进行归一化处理，权重 λ 用于控制两类损失的平衡，由于本文中正负样本 1:3 的设置， λ 取值 $\frac{1}{4}$ 。

3.2 多尺度特征融合网络

在原始 Faster R-CNN 框架中，RPN 生成的每个候选区域在特征层 Conv5_3 上完成位置映射后，由 RoI pooling 操作产生固定尺寸的特征图，而后送入目标检测网络得到最终的检测结果。为了进一步提升多尺度 Logo 检测算法的整体性能，本文借鉴现今流行的多尺度特征融合思想^[23-24]，在基础网络与目标检测网络之间增加了多尺度特征融合网络。如图 1 所示，网络选定 Conv4_3 作为

特征融合的基准层，因为该层相对于 Conv5_3 具有更小的步长，更有利于小目标的检测。同时，网络使用最大值池化（max pooling）实现针对 Conv3_3 特征层的下采样，使用文献[25]中的反卷积（deconvolution）实现针对 Conv5_3 特征层的上采样，并保证采样后的特征图均与基准层 Conv4_3 的大小相同。

由于特征层 Conv3_3、Conv4_3 和 Conv5_3 三者的特征尺度差异很大^[24]，直接融合将导致特征尺度较大的层对于后续的预测结果产生支配性的影响，而这些特征层本身并不能保证良好的检测性能，相反会增加后续参数学习的难度。因此，网络在对采样后的多层特征图进行融合前，先分别对 Conv3_3、Conv4_3 和 Conv5_3 层实施文献[24]中引入的 L2 正则化操作，并相应扩大各层特征的初始 L2 范数值为 10、8、5，以提升网络的训练速度^[14]。

在多层特征融合的具体实现方面，本文首先基于常用的 Concat 操作对正则化后的多层特征图进行拼接，形成高维的超特征，然后使用 1×1 的卷积实现针对超特征的降维和跨通道的信息整合。如图 1 所示，将针对超特征的降维工作放在目标检测网络之前，有利于避免后续各候选区域单独降维的重复计算，加快全流程的检测速度。

3.3 目标检测网络与联合训练

作为两阶段目标检测的第二阶段，目标检测网络针对候选区域做进一步的类别判定和位置精修。如图 1 所示，本文目标检测网络仍基于 Faster R-CNN 中实现，即利用全连接层产生最终的分类和回归结果。不同的是，为了缓解原始 RoI pooling 操作^[5]在候选区域映射过程中两次离散性的量化造成的区域不匹配问题^[26]，本文使用文献[26]中提出的 RoI Align 代替 RoI pooling 操作，这在一定程度上减轻了候选区域映射过程产生的位置偏差对于小目标检测的影响。

在多网络的联合训练方面，本文基于候选区域生成先于目标检测的顺序流程，参考 Faster R-CNN 中提出的四步交替训练方法，将本文多网络的联合训练过程分为以下 4 个阶段。① 单独训练区域生成模块，该模块包括基础网络和

多尺度区域生成网络, 其中基础网络、Conv_fc6 以及 Conv_fc7 基于 ImageNet 预训练模型初始化^[6], 其他新添加层使用标准差为 0.01 的零均值正态分布进行随机初始化。② 单独训练目标检测模块, 该模块包括基础网络、多尺度特征融合网络和目标检测网络, 其中基础网络以及目标检测网络同样基于 ImageNet 预训练模型初始化, 多尺度特征融合网络中的反卷积层基于双线性插值法初始化^[25], 降维卷积核参数使用标准差为 0.01 的零均值正态分布进行随机初始化, 训练所需的候选区域由区域生成模块提供。③ 再次训练区域生成模块, 其中基础网络由第二阶段训练所得的对应参数初始化, 并保持其参数在训练过程中固定, 只更新区域生成模块的独有部分。④ 再次训练目标检测模块, 固定第三阶段区域生成模块的所有参数, 只更新目标检测模块的独有部分。

4 实验与结果分析

4.1 数据集与评价指标

本文实验采用的数据集源自奥格斯堡大学多媒体计算和计算机视觉实验室维护并公开的 FlickrLogos-32^[9], 作为 Logo 检测的 benchmark 数据集, 其通常用于评估多类 Logo 检测/识别以及真实图像上的 Logo 检索方法。FlickrLogos-32 共有 8 420 张图像 (包含 6 000 张不含 Logo 的图像), 分为 32 类 Logo, 每类为 70 张且均具有较为平坦的表面。图 2 给出了该数据集对象大小的整体分布情况, 不难发现: ① 该数据集对象具有较大的尺度变化范围, 适用于评估检测算法的尺度不变性; ② 该数据集对象的平均尺度较小, 适用于检验算法在小目标检测方面的性能。

在实际训练过程中, 本文按照该数据集的官方划分标准, 为每类 Logo 划分 40 张图像用于训练, 剩余 30 张用于测试。由于本文算法中多尺度区域生成网络采取粗放式的正负样本生成策略, 该策略下网络的充分训练无疑对训练样本的数量和质量均提出了较高要求, 因此, 本文在将训练样本送入网络训练之前, 预先对每张训练图像随机采取如下数据增强操作: ① 直接使用原始图像; ② 对原始图像随机进行 2 倍或 4 倍的下采样,

原图范围内的剩余区域全 0 填充; ③ 对原始图像随机进行 1.5 倍或 2 倍的上采样, 并使用边长为 600px 的正方形区域对采样后的图像进行随机裁剪, 同时保证裁剪后子图像中包含的任一真值框区域与上采样后真值框的面积之比不低于 0.8。

目标检测任务通常需要一定的评价指标来评估检测算法的性能, Logo 检测领域中常用的评价指标是平均精度均值 (mAP, mean average precision)。mAP 综合表征了准确率 (precision) 和召回率 (recall), 其值越大, 表明算法性能越好, 本文中使用 mAP 值评估检测算法的整体性能。此外, 本文使用召回率均值 (mR, mean recall) 评估算法中多尺度区域生成网络的有效性, 该指标计算一组实验条件下多类目标召回率的算术平均值, 能够反映区域生成网络为多类目标生成高质量候选区域的平均性能。

4.2 对比实验与分析

本文算法基于深度学习框架 PyTorch 实现, 从节约内存和便于实现的角度出发, 每批次只训练一个样本, 使用随机梯度下降算法优化神经网络参数, 同时为了节约显存, 基础网络中的前四层卷积层的学习率设为 0。对于多网络的联合训练, 本文区域生成模块每阶段迭代 6×10^4 次, 目标检测模块每阶段迭代 4×10^4 次, 初始学习率均为 0.001, 学习率衰减步长为 4×10^4 , 衰减系数为 0.1。

表 3 列出了 4 种不同算法在 FlickrLogos-32 数据集上的实验结果, 可以看出, 与基线方法 Faster R-CNN^[10]相比, 本文算法取得了明显优势 (85.7% VS 81.1%)。在基线方法的复现过程中, 本文基于 VGG16 和 RoI Align^[26]的实验配置, 得到了与其相近的实验结果, 这表明本文算法相对于基线方法的性能提升直接来源于本文算法的改进部分, 即多尺度区域生成网络和多尺度特征融合网络。

在不同尺度目标的检测性能评估方面, 本文统计了 FlickrLogos-32 测试集中每类 Logo 对象的平均尺度, 然后将 Logo 类别按照平均尺度的降序排序, 顺序选取了前、中、后各 5 类 Logo 分别作为大、中、小尺度的代表。图 5 直观对比了本文算法与基线方法在所选类别上的平均精度, 可以发现, 对于大中尺度的目标, 本文算法与基线方法的整体检测性能基本相当; 而对于尺度较小的目标, 本文

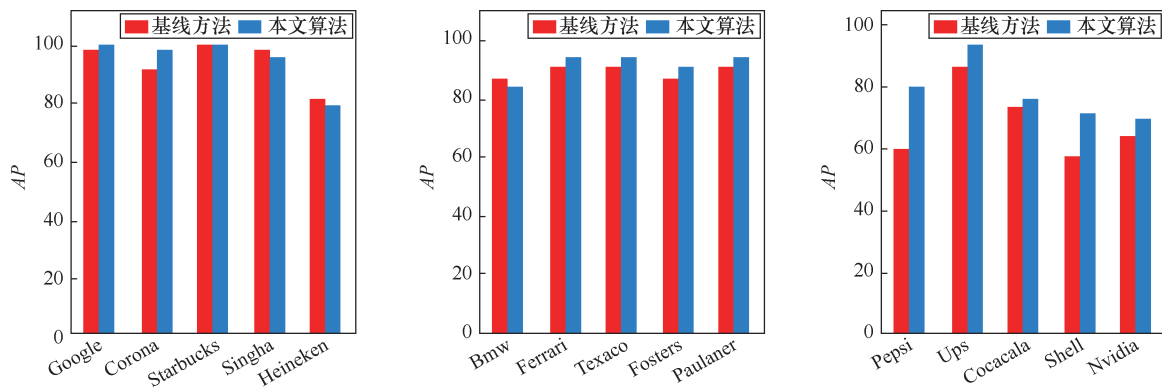


图 5 本文算法与基线方法在不同尺度类别上平均精度的对比
Figure 5 Comparison of the AP of different scale categories between our algorithm and the baseline

表 3 不同算法在 FlickrLogos-32 数据集上的实验结果
Table 3 Experimental results of different algorithms on FlickrLogos-32

算法	Adidas	Aldi	Apple	Becks	BMW	Carls	Chim	Coke	<i>mAP</i>
	Corona	DHL	Erdi	Esso	Fedex	Ferra	Ford	Fost	
	Google	Guin	Hein	HP	Milka	Nvid	Paul	Pepsi	
	Ritt	Shell	Sing	Starb	Stel	Texa	Tsin	Ups	
本文算法	78.3	88.7	80.7	86.1	84.2	77.4	83.1	76.0	85.7
	97.6	80.6	90.3	94.9	83.2	94.0	92.0	90.4	
	99.7	87.5	79.1	76.0	66.8	69.1	94.1	79.6	
	86.2	71.0	95.3	99.5	87.4	93.7	86.0	93.7	
基线方法 Faster R-CNN ^[10]	68.1	79.1	84.5	72.3	86.4	68.0	78.0	73.3	81.1
	90.9	77.4	90.9	88.6	71.1	91.0	98.3	86.2	
	98.0	90.7	81.3	67.0	54.5	64.0	90.9	59.6	
	81.0	57.3	97.9	99.5	86.7	90.4	87.5	85.8	
DeepLogo ^[17]	61.6	67.2	84.9	72.5	70.0	49.6	71.9	33.0	74.4
	92.9	53.5	80.1	88.8	61.3	90.0	84.2	79.7	
	85.2	89.4	57.8	58.4	34.6	50.3	98.6	34.2	
	63.0	57.4	94.2	95.9	82.2	87.4	84.3	81.5	
BD-FRCN-M ^[27]	—	—	—	—	—	—	—	—	73.5
	—	—	—	—	—	—	—	—	
	—	—	—	—	—	—	—	—	
	—	—	—	—	—	—	—	—	

算法的平均精度明显高于基线方法,这表明本文算法在小尺度目标检测方面更加稳健。

除此之外,为了验证本文算法中多尺度区域生成网络的有效性,本文分别计算了测试集上多尺度区域生成网络与原始 RPN^[5]在生成单位数量候选区域条件下的召回率均值,所有召回率的计算均基于 0.5 的 IoU 阈值,即仅当某真值框与该图像中任一候选区域的 IoU 值超过 0.5,该真值框被判定为召回。如图 6 所示,可观察到:① 多尺度

区域生成网络在任意数量候选区域条件下产生的召回率均值都明显优于原始 RPN,这表明利用特征金字塔逐层预测不同尺度候选区域的做法,能够有效提升生成候选区域的召回率;② 在 IoU 阈值 0.5 的设定下,两种区域生成算法的召回率均值都不高,尽管后续目标检测网络会对候选区域进行位置精修以保证最终的召回率,但目前区域生成算法的实际性能仍有进一步提升的空间,因此这方面成为下一步的研究重点。

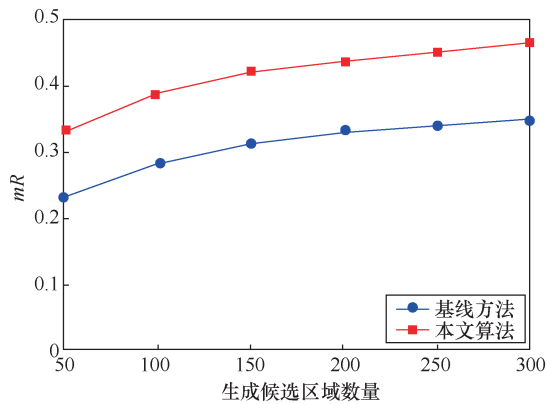


图6 本文算法与基线方法在区域生成网络性能方面的对比

Figure 6 Comparison of the PRN performance between our algorithm and the baseline

最后,本文实验探究了多尺度特征融合网络对于算法性能的影响,替换图1中目标检测网络的输入为单特征层,其他参数保持不变,得到的检测结果如表4所示。从表4中可以看出,基于权重可学习的特征融合算法能够在一定程度上提升检测算法的整体性能。图7给出了本文基于CNN的多尺度Logo检测算法与基线方法在FlickrLogos-32部分测试实例上检测结果的对比(右侧为本文算法的检测结果),可以发现本文算法对于目标的多尺度、小尺度等情形具备更强的稳健性。

表4 多层特征融合预测与单特征层预测在检测精度上的对比

Table 4 Comparison of detection accuracy between multi-layer feature fusion prediction and single feature layer prediction

目标检测网络的输入特征层	检测结果 mAP
Conv3_3+Conv4_3+Conv5_3	85.7
Conv3_3	83.2
Conv4_3	84.5
Conv5_3	84.1



图7 本文算法与基线方法在部分测试实例上检测结果的对比

Figure 7 Comparison of the detection results of our algorithm and the baseline on some test images

5 结束语

本文针对自然场景图像中多尺度Logo的检测需求,提出了一种基于CNN的多尺度Logo检测算法。算法在两阶段目标检测模型Faster R-CNN的基础上,通过构建特征金字塔并采取逐层预测的方式实现多尺度候选区域的生成,同时增加了多尺度特征融合网络,用于整合多层次特征以增强特征的表达能力。在FlickrLogos-32数据集上的实验结果表明:①采取特征金字塔逐层预测多尺度候选区域的方式,能够有效提升生成候选区域的召回率;②基于权重可学习的特征融合做法能够在一定程度上提升检测算法的整体性能。由于目前本文区域生成算法的性能仍与设计预期存在一定的差距,提升生成候选区域的召回率是下一步的研究方向。

参考文献:

- [1] SU H, GONG S, ZHU X. Scalable deep learning logo detection[C]//International Conference on Computer Vision Workshop. 2017: 270-279.
- [2] GAO Y, WANG F, LUAN H, et al. Brand data gathering from live social media streams[C]//International Conference on Multimedia Retrieval. 2014: 169.
- [3] PAN C, YAN Z, XU X, et al. Vehicle logo recognition based on deep learning architecture in video surveillance for intelligent traffic system[C]//IET International Conference on Smart and Sustainable City. 2013: 123-126.
- [4] ZHANG S, WEN L, BIAN X, et al. Single-shot refinement neural network for object detection[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2018.
- [5] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[C]//International Conference on Neural Information Processing Systems. 2015.
- [6] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multi-box detector[C]//European Conference on Computer Vision. 2016.
- [7] LUO W, LI Y, URTASUN R, et al. Understanding the effective receptive field in deep convolutional neural networks[C]//Conference and Workshop on Neural Information Processing Systems. 2016.
- [8] ZHOU B, KHOSLA A, LAPEDRIZA A, et al. Object detectors emerge in deep scene CNNs[J]. Computer Science, 2014.
- [9] ROMBERG S, PUEYO L G, LIENHART R, et al. Scalable logo recognition in real-world images[C]//ACM International Conference on Multimedia Retrieval. 2011: 25.
- [10] HANG S, ZHU X, GONG S. Deep learning logo detection with data expansion by synthesising context[C]//Applications of Com-

- puter Vision. 2017.
- [11] HU P, RAMANAN D. Finding tiny faces[C]//Computer Vision & Pattern Recognition. 2017.
- [12] EGGERT C, WINSCHER A, ZECHA D, et al. Saliency-guided selective magnification for company logo detection[C]//IEEE 23rd International Conference on Pattern Recognition (ICPR). 2016.
- [13] MENG Z, FAN X, XIN C, et al. Detecting small signs from large images[C]//IEEE International Conference on Information Reuse and Integration. 2017.
- [14] ZHANG S, ZHU X, LEI Z, et al. S3FD: single shot scale-invariant face detector[C]//International Conference on Computer Vision. 2017.
- [15] CAO G, XIE X, YANG W, et al. Feature-fused SSD: fast detection for small objects[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017.
- [16] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017.
- [17] IANDOLA F N, SHEN A, GAO P, et al. DeepLogo: hitting logo recognition with the deep neural network hammer[J]. arXiv preprint arXiv:1510.02131, 2015.
- [18] GIRSHICK R. Fast R-CNN[C]//International Conference on Computer Vision. 2015.
- [19] HOI S C H, WU X, LIU H, et al. LOGO-Net: large-scale deep logo detection and brand recognition with deep region-based convolutional networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 46(5): 2403-2412.
- [20] EGGERT C, WINSCHER A, LIENHART R. On the benefit of synthetic data for company logo detection[C]//The 23rd ACM International Conference on Multimedia. 2015.
- [21] EGGERT C, ZECHA D, BREHM S, et al. Improving small object proposals for company logo detection[J]. arXiv preprint arXiv:1704.08881, 2017.
- [22] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2014.
- [23] KONG T, YAO A, CHEN Y, et al. HyperNet: towards accurate region proposal generation and joint object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016.
- [24] LIU W, RABINOVICH A, BERG A C. ParseNet: looking wider to see better[J]. Computer Science, 2015.
- [25] LONG J, SHELHAMER E, DARRELL T. Fully convolutional

networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 39(4):640-651.

- [26] HE K, GKIOXARI G, DOLLAR P, et al. Mask R-CNN[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, (99): 1.
- [27] OLIVEIRA, GONÇALO, FRAZÃO, et al. Automatic graphic logo detection via fast region-based convolutional networks[C]//International Joint Conference on Neural Networks. 2016: 985-991.

[作者简介]



江玉朝（1994-），男，江苏盐城人，信息工程大学硕士生，主要研究方向为计算机视觉。



吉立新（1969-），男，江苏淮安人，博士，信息工程大学研究员、博士生导师，主要研究方向为通信与信息系统。



高超（1982-），男，河南郑州人，博士，信息工程大学助理研究员、博士生导师，主要研究方向为计算机视觉。



李邵梅（1982-），女，湖北钟祥人，博士，信息工程大学副研究员，主要研究方向为计算机视觉。

《网络与信息安全学报》

第二届编辑委员会

主 任 委 员：方滨兴

常务副主任委员：李风华

副 主 任 委 员：李 晖 俞能海 Kui Ren[USA] 刘建伟

邹德清 杜学绘 刘华鲁

委 员：（按姓氏笔画排序）

丁 勇 马建峰 王 标 王布宏 王连海

王良民 王国胤 王明生 王美琴 牛 犇

仇洪冰 田有亮 田俊峰 史国振 付章杰

冯 涛 朱建明 延志伟 仲 盛 向 涛

刘 哲 刘 渊 刘西蒙 刘哲理 闫 峥

孙东红 杨宏宇 肖 甫 肖 亮 吴 铤

吴晓平 邱卫东 何道敬 沈 超 张小松

张云勇 张红旗 张宏莉 张建标 张振江

陆月明 陈立全 陈兴蜀 陈性元 陈雷霆

苗春雨 林 鹏 易东山 赵 波 赵琉涛

咸鹤群 禹 勇 姚志强 秦 拯 贾 焰

殷丽华 翁 健 黄欣沂 曹春杰 彭国军

蒋文保 韩伟力 韩劲松 程 光 阚海斌

David Naccache [FRA] Gang Qu [USA] Jinjun Chen [AUS]

Willy Susilo [AUS] Yier Jin [USA]

收录声明

本刊对发表的文章,拥有出版电子版、网络版版权,并拥有和其他网站交换信息的权利。本刊支付的稿酬中已经包含上述费用。

Chinese Journal of Network and Information Security has the copyright to publish electronic edition, online edition of the published articles, and has the right to exchange information with other sites. The expenses have been included in the fee paid by editorial department.

道德声明

本刊发表的论文是作者独立取得的原创性研究成果,无一稿多投;论文内容不涉及国家机密;未曾以任何形式用任何文种在国内外公开发表过;论文内容不侵犯他人著作权和其他权利。若发生一稿多投、侵权、泄密等问题,论文作者将承担全部责任。

The authors of *Chinese Journal of Network and Information Security* guarantee that their submitted articles are original and contain nothing confidential. The said article is only submitted to *Chinese Journal of Network and Information Security*. The said article has not been published before and has not been submitted elsewhere for print or electronic publication consideration. The said article is no way whatever a violation or an infringement of any existing copyright or license from the third party. Otherwise, the authors of the said article shall take the blame for the violation or infringement of the related copyright and the leakage of secrets.



万众一心／同舟共济／安全自己／安全社会

新冠肺炎疫情防控工作
你我共担当



战“疫”必胜

COVID-19

网络 Chinese Journal of
Network and Information Security
信息安全学报



2020年4月15日出版 定价：60.00元

