一种融合上下文信息及自适应感受野的多尺度目标检测算法

张 婷 兰时勇
(四川大学视觉合成图形图像技术国防重点学科实验室 四川 成都 610064)

摘 要 目标检测在实际应用各类复杂场景中面临着诸多的挑战,如目标遮挡、光照变化、目标尺度变化等。为了提高多尺度目标检测的性能,提出一种改进的特征金字塔(FPN)的目标检测算法。以特征金字塔网络框架为基础引入上下文信息融合模块,充分利用目标对象与其周围环境的关联属性,增强宽动态尺度范围的目标对象的特征表征,提高不同尺度目标的辨识能力。此外,构建一个跨通道注意机制,自适应调整不同尺度目标特征的通道灵敏度,学习到适应目标尺度的感受野范围。该算法在 Pascal VOC 数据集训练验证,其平均精确率(mAP)比基准方法提高了 3%。

关键词 目标检测 上下文信息融合 跨通道注意力机制

中图分类号 TP3 **文献标志码** A DOI:10.3969/j.issn.1000-386x.2024.10.046

AN MULTI-SCALE OBJECT DETECTION ALGORITHM COMBINING CONTEXT INFORMATION AND ADAPTIVE RECEPTIVE FIELD

Zhang Ting Lan Shiyong
(*National Key Laboratory of Fundamental Science on Synthetic Vision*, *Sichuan University*, *Chengdu 610064*, *Sichuan*, *China*)

Abstract Object detection faces many challenges in the practical application of various complex scenes, such as object occlusion, illumination changes, and object size changes in the practical application. In order to improve the performance of multi-scale target detection, this paper proposes an improved feature pyramid network (FPN) target detection algorithm. Based on the FPN framework, the context information fusion was introduced to utilize the relevance of an object to its surrounding environment and enhance feature representation of objects for wide dynamic range images and to improve the ability of detection ability for different scales. In addition, a cross-channel attention mechanism was constructed to adaptively adjust the channel sensitivity of target features at different scales. Experiments on the Pascal VOC dataset show that the proposed method improves the detection performance by 3% compared with the baseline method in terms of mean average precision (mAP).

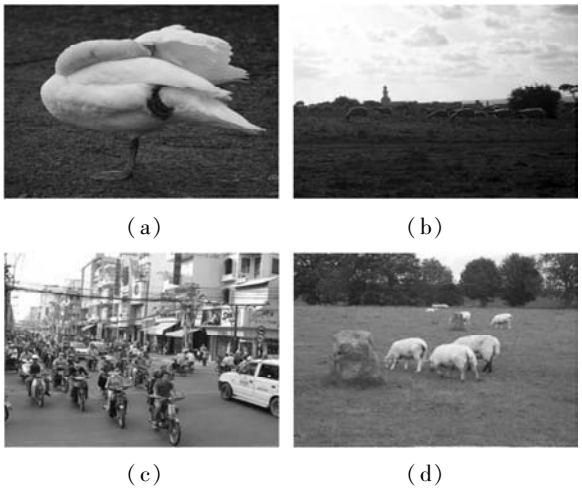
Keywords Object detection Context information extraction Cross-channel attention mechanism

0 引 言

目标检测是计算机视觉领域的基本任务之一,广泛应用在自动驾驶、医疗诊断、自动驾驶等领域,其目的是确定物体在图像中是否属于已知类别,如果属于已知类别,则使用矩形包围框来判定估计物体的位置。受限于现实物体成像,由于光、距离等诸多因素的干扰,造成被检测目标变形、模糊、重叠,导致目标漏检。如图 1 所示,图 1(a)中的丹顶鹤的脖子是卷曲的,由于目标自身的动作导致重要的特征检测缺失;在图 1(b)中,由于光照条件较差,动物的特征在图像中并不清晰。在实际应用中,物体遮挡经常发生,特别是当多个物体同时出现在一幅图像中,如图 1(c)中的人和摩托。另外,即使是同一类型的物体,由于成像视角的不同,它们的外观可能会有很大的不同,在距离较远的情况下,物体也很难检测到,如图 1(d)中的绵羊。在以上这些情况下的图像中,如何能够准确检测目标成为一大难题。

收稿日期:2021 - 05 - 05。四川省科技厅重点研发项目(2021YFG0300);视觉合成图形图像技术国防重点学科实验室开放研究项目。**张婷**,硕士生,主研领域:计算机视觉,深度学习。**兰时勇**,副研究员。

目前的目标检测方法中大多基于卷积神经网络,仅仅关注到目标本身的信息,忽略了图像中可能会出现干扰因素的影响,这些将会直接影响物体的外观和物体之间的差异。为了减少这些干扰因素造成的影响,专家们提出改进的基于上下文信息的方法,例如在 CoupleNet[1] 提出一个新的分支提取全局信息嵌入 R-FCN[2] 中,和 RoI(Region of Interest,感兴趣区域)池化、结合全局与局部上下信息提高检测精度;在 SIN[3] 中,将输入图像中的目标视为模型中的一个节点,将两个节点之间的连接视为它们之间的关联,从而考虑目标与全局场景信息之间的关联,这样将目标检测问题转化为结构推理问题。但是这些算法通常从某一层,特别是最后一层提取特征信息,单层网络构建单层特征,目标特征信息、位置信息有限。



(a)　(b)　(c)　(d)

图 1　在复杂场景中物体检测的具有挑战性的实例

针对以上的问题,本文旨在利用上下文信息来提升目标检测的准确性。由于在卷积网络中深层语义特征有助于目标识别,浅层特征有助于位置标定。因此采用了多尺度、多层次的网络模型。具体是利用 FPN (Feature Pyramid Network) [4] 中输出的特征层通过局部和全局上下文融合处理,获取与目标相关的周围环境中的信息。引入跨通道的注意力机制,保留更多的特征信息,抑制无用的信息,更多地关注特征通道中有效的特征信息。

1　相关工作

FPN 是用于检测多尺度的对象目标的最佳选择,具有丰富的语义信息。主要包括 3 个流程:

(1) 自下而上的通路是神经网络的正向传播过程,生成不同维度的特征,特征图在这一过程中通常会越变越小。

(2) 自上至下的过程是把语义更强的高层特征图进行上采样,然后将低层高分辨率的特征与语义更强的高层特征图进行结合,这样高层特征补充增强,保证了每一层都有合适的分辨率以及强的语义特征。

(3) 横向连接的过程是 CNN 网络层与最终的输出的各个维度的特征之间的关联表达。

由此,FPN 通过简单的网络连接解决了目标检测的多尺度问题,特征金字塔的网络结构如图 2 所示。但是此类方法仍然存在局限:单层网络构建特征对于在复杂的场景下的模糊、小的目标检测效果不佳。本文的改进算法旨在解决此类问题。

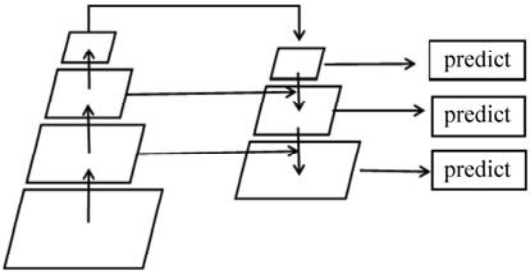


图 2　特征金字塔网络结构

2　本文改进算法

本文从两个方面出发,一方面是上下文特征信息的提取,另一个方面是通道注意力机制。在上下文特征信息提取模块中,使用一组不同的膨胀卷积核提取局部的文本信息,同时使用全局平均池化提取全局上下文特征信息,由此就会增加对象表示的特性。在注意力模块中,采用了基于通道的注意机制来获得 Feature Maps 通道之间的关系,根据不同的权重表示对相应通道的不同程度的注意。基于通道的注意力机制可以提高检测网络对通道特征的敏感性并增加与任务相关通道特征的权值[5]。最后,通过元素相加的方法融合这三个特征。网络结构如图 3 所示。

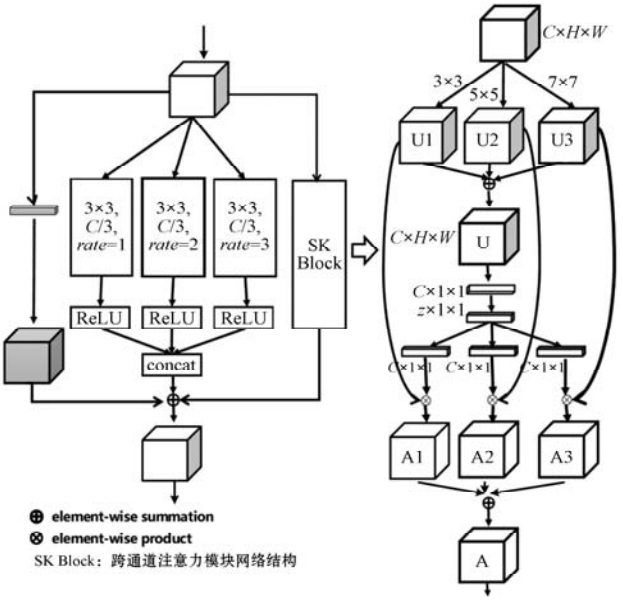
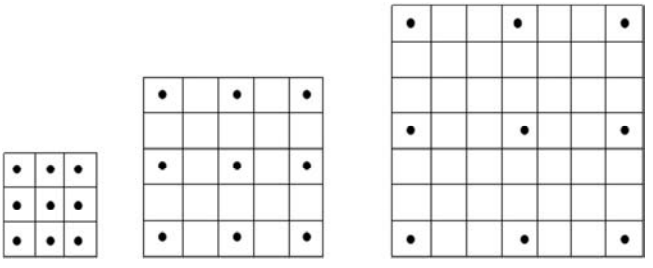


图 3　改进网络结构

2.1 局部和全局上下文信息

首先,通过对有效感受野的分析,对局部上下文信息的提取采用空洞卷积[5]的方法。输入层图像的梯度呈高斯分布,中心位置的梯度信号较强,离中心位置越远的梯度信号越弱。在前向传播中,输入的中心位置图层可以将信息扩散到输出的中心位置通过多个路径进行分层,具有更多的贡献点在远离输入层中心的位置也可将信息分散到输出的中心位置。通过较少的路径和较少的贡献进行分层,基于此有很多位置都可以忽略。因此本文使用3 个并列的空洞数分别为 1、2、3 的 3 ×3 的卷积核,这样做从一开始就完整地保留连续的 3 ×3 区域,同时保证感受野的连贯性,如图 3 所示。图 4(a)、(b)、(c)分别是相互独立的空洞数为 1、2、3 的卷积操作,黑色的点代表的是 3 ×3 的卷积核,白色的方格是卷积后的感受野。



(a) (b) (c)

图 4 不同空洞数的空洞卷积图

其次,为了提取全局上下文信息,我们使用全局平均池化[6]来提取目标对象的全局上下文信息,思想借鉴于 ParseNet[7]。全局平均池化可以减少有限的邻域而引起的估值方差,保留更多的图像背景信息。全局平均池化对空间信息进行求和,对输入的空间的鲁棒性更强,且不需要参数,避免在该层产生过拟合。

最后,融合局部特征和全局特征得到多尺度特征。

2.2 跨通道注意力机制

对于背景复杂的小目标、模糊目标的物体检测中,使用改进的基于局部和全局的上下文的 FPN 算法存在误检的情况。这是由于 FPN 的融合特征和上下文提取特征,存在特征提取复杂和冗余。因此引入注意力模块,聚焦“关心的”特征信息,抑制无用的信息,提高检测的精度。

本文设计了一个基于 SKNet[8]的跨通道注意力机制,根据通道间的依赖关系建模,提高网络对通道特征的敏感性,并自适应地调整感受野大小。

基于 SK 卷积[8],我们不使用全局平均池化来生成通道全局信息,而是计算每个通道特征的方差池化,因为信息的数量与信息的不确定性有关。

首先,我们对输入的图像的 Feature Map 分别进行卷积核分别是 3、5、7 的三个变换。为了提升效率,也可以将 5 ×5、7 ×7 的卷积替换为 3 ×3 的卷积,扩张率分别为 2、3。

然后,将三个分支的信息通过元素求和进行融合,接着使用方差值生成通道级统计信息嵌入全局信息。方差值在一定程度上反映了信息的不确定性,方差越大,不确定性越大。

进一步通过一个简单的全连接层实现降低维度来提高效率,得到一个较为紧凑的特征。

$$z = F_{fc}(s) = \delta(B(Ws)) \tag{1}$$

式中:s 为方差计算之后的值;δ 为 ReLU 函数;B 为 Batch Normalization;$\boldsymbol{W} \in \mathbf{R}^{d \times C}$;$d = \max(C/r, L)$ 其中 $L = 32$,C 为通道数,r 是 reduction ratio。

最后,在得到 z 的情况下,利用跨通道的软注意自适应地选择不同空间尺度的信息。式(2)应用 softmax 操作计算 channel-wise digits。

$$a_c = \frac{e^{A_c z}}{e^{A_c z} + e^{B_c z}} \tag{2}$$

$A \in \mathbf{R}^{C \times d}$,$A_c \in \mathbf{R}^{1 \times d}$是 A 的第 c 行,在三个分支上分别做计算。通过加权和累积得到特征的映射,通过注意力权重得到包含各种核的特征映射。

3 实 验

3.1 实验配置

本文算法在 UbunTu 16.04 操作系统上,采用深度学习框架 TensorFlow,实现多尺度目标检测算法。实验平台采用 GPU:GeForce RTX 2080 Ti,内存 16 GB。在 VOC07 数据集上进行实验,其中包括训练集 5 011 幅、测试集 4 952 幅,共包含 20 个类(加背景 21 个类)。骨干网络首先在 ImageNet 数据集上进行预训练,提高模型的泛化能力。在训练阶段,输入图像要调整大小,使较短的边有 800 个像素。设置 batch size 大小为 1,同时使用 SGD 优化器,具体的参数如表 1 所示。

表 1 实验中具体的参数设置

Iterations	Learning rate	Momentum	Decay
0 – 100k	0.001	0.9	0.000 05
100k – 140k	0.000 1	0.9	0.000 05
140k – 200k	0.000 01	0.9	0.000 05

3.2 实验分析

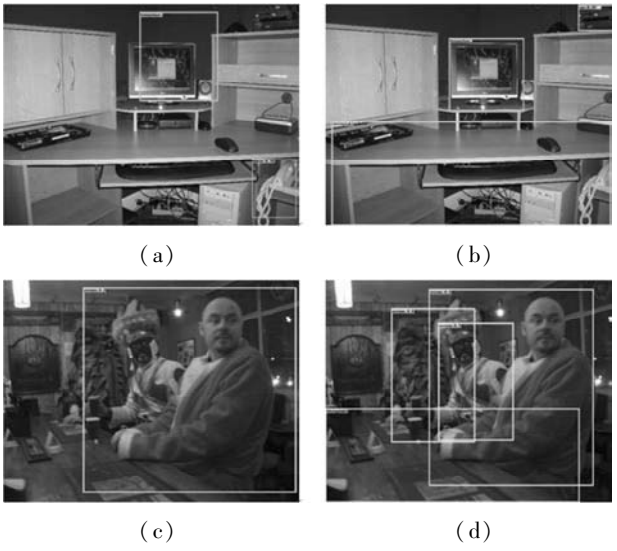
首先,通过消融实验对局部上下文和全局上下文提取以及跨通道域注意模块的有效性进行了验证。然后与其他目标检测算法进行比较,验证了本文提出的

基于 FPN 的 P2 – P6 层后上下文信息提取模块的有效性。在我们的实验中,首先在 FPN 的 P2 – P6 层添加了局部和全局上下文提取模块[9-10],发现 mAP 得到了改进,提高了 1.5%。然后,比较全局平均池化和方差计算的差异,当使用全局平均池化作为通道特征的全局信息时,检测效果下降,将跨通道的注意机制中的全局平均池化替换为方差计算,发现检测精度有所提高。由此利用通道特征的方差代替均值作为全局特征可以更有效地构建通道相关性。最终消融实验的对比结果如表 2 所示,验证了使用上下文提取模块和跨通道域注意机制模块的有效性,FPN 的 mAP 增加了 3% 以上。

表 2　消融实验结果对比

局部上下文	全局上下文	跨通道注意力模块(平均值)	跨通道注意力模块(方差)	mAP /%
				79.7
√		√		81.0
√			√	81.3
√	√			81.5
√	√	√		81.3
√	√		√	**82.5**

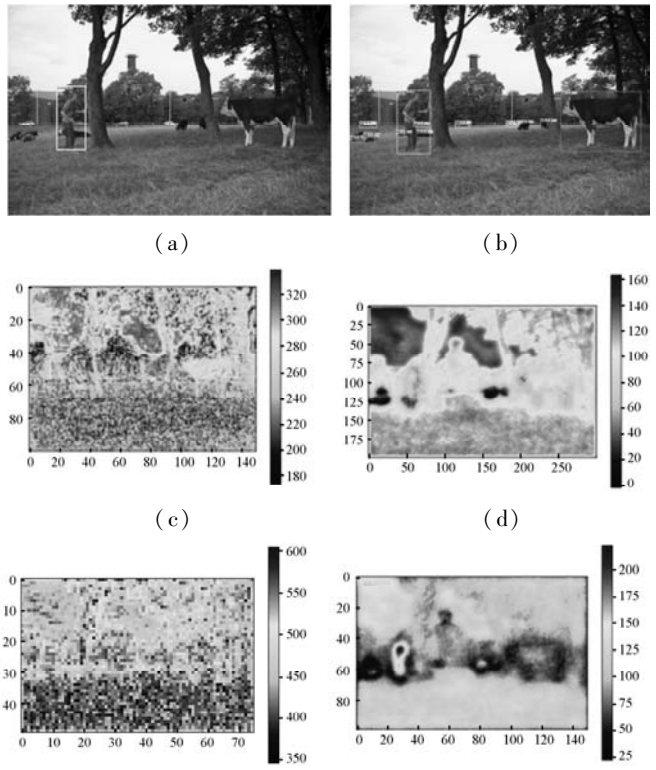
如图 5 所示,左列为 FPN 检测结果,右列为添加上下文提取模块和注意力模块后的检测结果。我们可以观察到,包含上下文提取模块可以提高对小物体(图 5(b)桌子上的模型汽车)和特征模糊目标(图 5(d)中人遮挡脸部特征信息)的检测。因此将语义信息的引入网络可以捕获对象与背景之间的相关性,通过目标周围环境信息,模糊特征的表达将会得到改善,这有助于检测被遮挡目标或小物体。



(a)　(b)　(c)　(d)

图 5　检测结果对比

接着,在图 6 中展示特征热力图的比较,其中左列和右列分别是 FPN 以及本文方法的 P3 层和 P4 层的特征图和热力图。可以清晰地看到在图 6(d)中,目标与环境的分界变得清晰,相比之下,图 6(c)中 FPN 的 P3 没有明显地将环境与目标区分出来。图 6(f)图像高度显示的人和牛(如图黑色较深的区域),这使他有别于他周围的对象。然而,FPN 的 P4 层和 P3 层之间的区别不明显。最后,通过对比图 6(a)和(b)的输出结果可以看出,本文使用的算法对较小目标的检测能力也有所提高。



(a)　(b)　(c)　(d)　(e)　(f)

图 6　热力图效果对比图

最后,将该方法与其他目标检测方法在使用相同的数据集的情况下进行了比较,如表 3 所示,ION[11] 和 HyperNet[12] 骨干网络使用 VGG16,其他网络模型的骨干网络使用 ResNet101。其中 R-FCN[2] 和 SSD[13] 是常用的目标检测算法,FPN、ION 和 HyperNet 都是基于 Faster RCNN[14] 的改进,而本文方法是基于 FPN。本文方法在 VOC07 的数据集训练验证,以 mAP 为度量,取得了较好的性能,同时也验证了方法的有效性。

表 3　实验结果对比

Method	Backbone	Train	Test	mAP
FPN	ResNet101	VOC07	VOC07	79.7
ION	VGG16	VOC07	VOC07	75.6
HyperNet	VGG16	VOC07	VOC07	76.3
R-FCN	ResNet101	VOC07	VOC07	80.5
SSD	ResNet101	VOC07	VOC07	80.6
Our method	ResNet101	VOC07	VOC07	**82.5**

4 结 语

本文提出了一种在 FPN 中结合上下文信息和跨通道域注意的算法。结果表明,空洞卷积可以有效地提取对象周围局部信息,而使用全局平均池化有助于提取对象的全局上下文信息。此外,跨通道域注意机制可以通过不同大小的感受野自适应调整,增强多尺度目标特征的表征能力。实验结果表明,该方法能够提高复杂场景下的多尺度目标检测性能。

参 考 文 献

[1] Zhu Y S, Zhao C Y, Wang J Q, et al. CoupleNet: Coupling global structure with local parts for object detection[EB]. arXiv:1708.02863v1,2017.

[2] Dai J F, Li Y, He K M, et al. R-FCN object detection via region-based fully convolutional networks[EB]. arXiv:1605.06409,2016.

[3] Liu Y, Wang R P, Shan S G, et al. Structure inference net: Object detection using scene-level context and instance-level relationships[C]//IEEE Conference on Computer Vision and Pattern Recognition,2018:6985 - 6995.

[4] Lin T Y, Dollar P, Grishick R, et al. Feature pyramid networks for object detection[C]//IEEE Conference on Computer Vision and Pattern Recognition,2017:936 - 944.

[5] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions[EB]. arXiv:1511.07122,2016.

[6] Lin M, Chen Q, Yan S C. Network in network[EB]. arXiv:1312.4400,2014.

[7] Liu W, Rabinovich A, Breg A C. ParseNet: Looking wider to see better[EB]. arXiv:1506.04579,2016.

[8] Li X, Wang W H, Hu X L, et al. Selective kernel networks [C]//IEEE Conference on Computer Vision and Pattern Recognition,2019:510 - 519.

[9] 张宽,腾国伟,范涛,等. 基于密集连接的FPN多尺度目标检测算法[J]. 计算机应用与软件,2020,37(1):166 - 212.

[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]. 计算机应用和软件,2020,37(5):195 - 199.

[11] Bell S, Zitnick C L, Bala K, et al. Inside-Outside net: Detecting objects in context with skip polling and recurrent neural networks[EB]. arXiv:1512.04143,2016.

[12] Kong T, Yao A, Chen Y, et al. Hypernet: Towards accurate region proposal generation and joint object detection [C]// IEEE Conference on Computer Vision and Pattern Recognition,2016:845 - 853.

[13] Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multibox detector[C]//European Conference on Computer Vision, 2016:21 - 37.

[14] Ren S Q, He K M, Grishick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transaction on Pattern Analysis & Machine Intelligence,2015,39(6):1137 - 1149.

(上接第 286 页)

[3] Marco A, Navigli R. Clustering and diversifying web search results with graph-based word sense induction[J]. Computational Linguistics,2013,39(3):709 - 754.

[4] Salton G. Automatic text processing: The transformation, analysis, and retrieval of information by computer[M]. New York: Addison-Wesley Longman Publishing,1989.

[5] Lin S H, Chen M C, Ho J M, et al. ACIRD: Intelligent internet document organization and retrival[J]. IEEE Transactions on Knowledge and Data Engineering,2012,14(3):599 - 614.

[6] 孙吉贵,刘杰,赵连宇. 聚类算法研究[J]. 软件学报,2008,19(1):48 - 61.

[7] Rodriguez A, Laio A. Clustering by fast search and find of density peaks[J]. Science,2014,344(6191):1492.

[8] Chang H C, Hsu C. Using topic keyword clusters for automatic document clustering[J]. Transactions on Information and Systems,2005,88(8):1852 - 1860.

[9] Ji W, Guo Q J, Zhong S, et al. Improved K-medoids clustering algorithm under semantic web[C]//2nd International Conference on Computer Science and Electronics Engineering,2013:527 - 531.

[10] 杨洁,季铎,蔡东风,等. 基于联合权重的多文档关键词抽取技术[J]. 中文信息学报,2008,22(6):75 - 79.

[11] 卜东波. 聚类/分类理论研究及其在文本挖掘中的应用[D]. 北京:中国科学院计算技术研究所.

[12] Thiesson B, Meck C, Chickering D M, et al. Learning mixtures of Bayesian networks[C]//Conference on Uncertainty in Artificial Intelligence,1997:321 - 323.

[13] Yan B, Zhang Y, Su H Y, et al. Cluster center initialization parallel algorithm for K-means algorithm[C]//International Conference on Computer and Information Technology,2014: 2169 - 2172.

[14] Dodge Y. Statistical data analysis based on the L1-norm and related methods[M]//Statistics for Industry and Technology. New York: Springer,1987.

[15] 仝鑫,王罗娜,王润正,等. 面向中文文本分类的词级对抗样本生成方法[J]. 信息网络安全,2020,20(9):12 - 16.