

结合注意力机制的深度学习图像目标检测

孙萍¹, 胡旭东², 张永军¹

1. 武汉大学 遥感信息工程学院, 武汉 430079

2. 中国科学院地理信息与文化科技产业基地, 中科天启, 江苏 苏州 215000

摘要:利用卷积神经网络进行目标检测时,提取的卷积特征具有很强的平移不变性,这将削弱模型的定位性能。事实上,目标对象通常具有不同的子区域特征和宽高比特性,但在目前流行的两阶段目标检测框架中,很少考虑这些具有平移尺度敏感性的特征成分。为了优化模型的特征表达,将在两阶段目标检测框架中引入与子区域特征和宽高比特性相关的注意力特征库,并生成注意力特征图对原始的ROI池化特征进行优化。另外,在注意力特征图的辅助下,模型特征维度可以有效地进行缩减。实验结果表明,引入注意力模块后,模型的检测精度和检测速度有明显提升。

关键词:目标检测;卷积神经网络(CNN);注意力机制;特征降维

文献标志码:A **中图分类号:**TP389.1 **doi:**10.3778/j.issn.1002-8331.1902-0155

孙萍,胡旭东,张永军.结合注意力机制的深度学习图像目标检测.计算机工程与应用,2019,55(17):180-184.

SUN Ping, HU Xudong, ZHANG Yongjun. Object detection based on deep learning and attention mechanism. Computer Engineering and Applications, 2019, 55(17): 180-184.

Object Detection Based on Deep Learning and Attention Mechanism

SUN Ping¹, HU Xudong², ZHANG Yongjun¹

1. School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China

2. Geo-Science and Technology Service Network, CAS, Image Sky, Suzhou, Jiangsu 215000, China

Abstract: In the Convolution Neural Network(CNN), convolutional layers are translation-invariant, which weaken the localization performance of object detector. Actually, objects usually have distinct sub-region spatial characteristics and aspect ratio characteristics, but in prevalent two-stage object detection methods, these translation-variant feature components are rarely considered. In order to optimize the feature representations, the sub-region attention bank and aspect ratio attention bank are introduced into the two-stage object detection framework and generate the corresponding attention maps to refine the original ROI features. In addition, with the aid of the attention maps, the feature dimension can be greatly reduced. The experimental results show that object detectors equipped with attention module improve the accuracy and inference speed significantly.

Key words: object detection; Convolution Neural Network(CNN); attention mechanism; dimension reduction

1 引言

近年来,深度学习在计算机视觉领域得到飞速发展,相关的理论和实践成果层出不穷,基于深度卷积神经网络的目标检测技术也取得了巨大突破。

卷积神经网络(Convolution Neural Network, CNN)^[1]有着强大的特征提取能力。R-CNN^[2]模型首次将卷积神经网络引入目标检测领域,用于提取候选区域(Region

Of Interest, ROI)的特征。之后,出现了一系列基于区域的两阶段目标检测模型,如Fast R-CNN^[3]、Faster R-CNN^[4]等,以及一些单阶段目标检测模型,如YOLO^[5-7]、SSD^[8-10]等。由于单阶段目标检测模型的检测精度有限,目前大部分目标检测器仍采用由候选区域生成器和区域分类器组成的两级框架。

在两阶段目标检测框架中,候选区域ROI发挥着重

作者简介:孙萍(1993—),女,硕士研究生,主要研究领域为数字图像处理、计算机视觉、深度学习、目标检测,E-mail:sp@whu.edu.cn;

胡旭东(1991—),男,研究员,主要研究领域为数字图像处理、计算机视觉;张永军(1975—),男,博士,教授,主要研究领域为数字图像处理、数字摄影测量与遥感、计算机视觉。

收稿日期:2019-02-25 **修回日期:**2019-04-03 **文章编号:**1002-8331(2019)17-0180-05

CNKI网络出版:2019-04-15, <http://kns.cnki.net/kcms/detail/11.2127.tp.20190409.1719.032.html>

要作用。如何生成功能强大、信息丰富的ROI特征成为目标检测领域的一项研究热点。

文献[11]提出的HyperNet通过累积多层特征图并压缩至归一化空间得到了多尺度超特征;文献[12]提出的特征金字塔FPN将顶层特征自上而下地逐层反馈至低层并进行融合,生成同时具备高语义信息和高分辨率准确度的特征;文献[13]提出的SDP根据ROI的尺寸在不同层级上进行级联池化;文献[14]提出的MR-CNN试图在ROI区域内部及周围构建更丰富的区域特征表示。这些方法一定程度上增强了模型的特征表达能力,但由于仍采用传统的ROI池化(ROI Pooling)操作,很少考虑模型的平移尺度协变性,使得模型定位性能不佳。针对这个问题,文献[15-16]提出了R-FCN和Deformable R-FCN模型,构造了一组位置敏感分数图,利用位置敏感池化操作对ROI子区域信息进行编码,有效地增强了模型的平移敏感性。不过,R-FCN模型只能保持子区域内的局部特征,不能很好地将各子区域的信息组合在一起形成全局结构信息。

值得关注的是,目标对象的不同子区域通常具有不同的空间特征,如边界部分更多的是边缘和轮廓特征,而中心部分更侧重于纹理特征,这些都是很好的定位识别依据。另外,不同类别或不同视角下的目标对象通常具有不同的宽高比特性。

因此,本文在R-FCN模型及视觉注意力机制^[17-18]的启发下,基于ROI子区域特征和宽高比特性的差异,在两阶段目标检测框架中引入特征注意力模块^[19],提取与ROI子区域特征和宽高比特性特定相关的注意力特征图,并与ROI池化特征进行组合优化,增强模型的特征表达能力,从而提高目标检测模型的定位识别性能。

2 引入注意力模块的两阶段目标检测模型

两阶段目标检测框架可细分为四部分:基础卷积特征提取、候选区域生成、ROI特征池化和ROI分类回归,如图1所示。模型首先利用卷积神经网络进行前向运算,生成整张图像的卷积特征图,维度为 $H \times W \times C$,然后利用候选区域生成网络RPN对anchor box^[4]进行分类和回归得到候选区域ROI。在得到了卷积特征图和候选区域ROI之后,重点需要研究ROI特征提取器的设计。

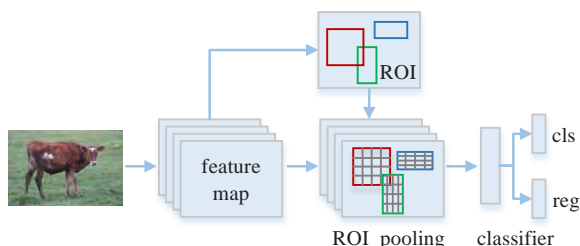


图1 两阶段目标检测框架主要流程

传统的ROI池化操作在卷积特征图的所有通道统一进行分块池化,很少考虑与ROI子区域特征和宽高比

属性相关的特征成分。为了增强模型的平移尺度敏感性,将在模型中引入与子区域特征和宽高比属性相关的注意力模块,模型的整体结构如图2所示。注意力模块作用于ROI特征提取过程中,新的ROI特征提取器主要包括ROI特征池化和注意力特征图生成两部分。首先,按照常规的ROI池化操作生成池化特征 f_i 。然后,设置子区域注意力特征库和宽高比注意力特征库,并根据各ROI的子区域特征和宽高比特性,从注意力特征库中选择性地生成子区域注意力特征图 M_{sr} 和宽高比注意力特征图 M_{ar} 。将注意力特征图进行合并,并与原始的ROI池化特征 f_i 进行加权融合,即可得到最终的ROI分类特征 rf_i ,过程描述如式(1)所示:

$$rf_i = f_i \cdot (M_{sr} + M_{ar}) \quad (1)$$

其中, $i=1,2,\dots,N$, N 表示ROI的个数。

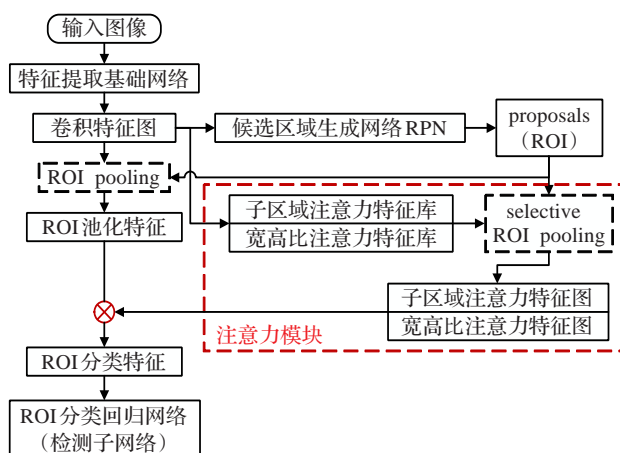


图2 引入注意力模块的两阶段目标检测框架

2.1 注意力特征库

生成候选区域ROI后,假设将每个ROI划分为 N_{sr} 个子区域,如划分为 3×3 的9个子区域。同时,根据宽高比(Asspect Ratio)属性将所有的ROI划分为 N_{ar} 类,如划分为 $Ratio < 0.75$ 、 $0.75 < Ratio < 1.3$ 和 $Ratio > 1.3$ 这三类。一旦给定一个ROI,就可以确定ROI内部各空间点所属的子区域及该ROI所属的宽高比类别。为了根据这些特性生成相应的注意力特征图,需要先设置注意力特征库。注意力特征库中包含的是众多ROI所有可能存在的注意力特征激活情况。

其中,子区域注意力特征库中存储的是与空间位置信息相关的特征。对于 $H \times W \times C$ 维的卷积特征,每个特征点可能隶属于各ROI的不同子区域,其注意力显著值与该点在ROI中的具体位置有关。编码 N_{sr} 个子区域空间特征的过程中,将划分为 N_{sr} 个子区域的ROI表示为 $\{G_{k_{sr}}\}_{1 \leq k_{sr} \leq N_{sr}}$,并将卷积特征图中 (i,j) 位置处的特征向量表示为 F_{ij} ,维度为 C 且 $1 \leq i \leq H$, $1 \leq j \leq W$ 。当 (i,j) 点位于ROI的第 k_{sr} 个子区域 $G_{k_{sr}}$ 时,对应的子区域注意力显著值为 $\omega_{sr}^{k_{sr}}(i,j)$,如式(2)所示:

$$\omega_{sr}^{k_{sr}}(i,j) = \Phi_{k_{sr}}(F_{ij}) \quad (2)$$

其中, $\Phi_{k_{sr}}$ 相当于一个注意力特征提取器, 其参数通过网络的反向传播进行学习。 (i, j) 点对应的子区域注意力特征库为 $\{\omega_{sr}^{k_{sr}}(i, j)\}_{1 \leq k_{sr} \leq N_{sr}}$ 的集合, 维度为 $N_{sr}C$ 。因此, 总体的子区域注意特征库 W_{sr} 的维度为 $H \times W \times N_{sr}C$ 。

在子区域注意力特征库中, 每个空间点都对应着一组注意力显著值, 反映的是该点隶属于ROI不同子区域时的空间属性特征。当点 (i, j) 位于第 k_{sr} 个子区域空间范围内时, 特征库中 $(k_{sr}-1)C+1$ 到 $k_{sr}C$ 这一组通道表示其注意力显著值。

在考虑ROI内部空间特征差异的同时, ROI的宽高比属性也值得关注。在实际应用过程中, 宽高比信息除了反映目标类别差异以外, 还可以反映目标对象的观测视角及姿势形态, 如行走的人和坐着的人在纵横尺度上有很大的不同。然而, 传统的ROI池化层为所有的ROI生成同一维度的特征表示, 如生成 7×7 的池化特征, 忽略了它们之间的纵横尺度差异。为了表征ROI的形态差异, 将在目标检测框架中继续引入宽高比注意力特征库。与子区域注意力特征库的设置类似, 对于卷积特征图中 (i, j) 位置处的特征点, 当 (i, j) 点所在的ROI属于第 k_{ar} 宽高比类别时, 其对应的宽高比注意力显著值为 $\omega_{ar}^{k_{ar}}(i, j)$ 。因此, (i, j) 点对应的宽高比注意力特征库 $W_{ar}(i, j)$ 为 $\{\omega_{ar}^{k_{ar}}(i, j)\}_{1 \leq k_{ar} \leq N_{ar}}$ 的集合, 而 W_{ar} 的总维度大小为 $H \times W \times N_{ar}C$ 。

2.2 注意力特征图

注意力特征库包含了所有可能的注意力特征激活情况。为了生成与某ROI特定相关的子区域注意力特征图和宽高比注意力特征图, 需要从注意力特征库有选择性地池化。

在选择性ROI池化(Selective ROI Pooling)过程中, 假设将ROI划分为 $h \times w$ 池化块。其中, 池化块 $(m, n)_{1 \leq m \leq h, 1 \leq n \leq w}$ 将分别对应子区域注意力特征库和宽高比注意力特征库中的索引 $k_{sr}(1 \leq k_{sr} \leq N_{sr})$ 和 $k_{ar}(1 \leq k_{ar} \leq N_{ar})$ 。索引 k_{sr} 表示 (m, n) 池化块的大部分区域位于ROI的 k_{sr} 子区域空间范围, 即 $bin(1, 1) \in G_1, \dots, bin(m, n) \in G_{k_{sr}}, \dots, bin(h, w) \in G_{N_{sr}}$ 。而索引 k_{ar} 表示的是该ROI属于第 k_{ar} 宽高比类别, 所有池化块共享相同的 k_{ar} 。确定了 k_{sr} 和 k_{ar} 指标后, 就可以选定注意力特征库中对应部分的通道列, 具体为子区域注意力特征库中 $(k_{sr}-1)C+1$ 到 $k_{sr}C$ 的通道列及宽高比注意力特征库中 $(k_{ar}-1)C+1$ 到 $k_{ar}C$ 的通道列。ROI各池化块分别在选定的通道上进行最大值池化, 即可得到与该ROI唯一相关的子区域注意力特征图 M_{sr} 和宽高比注意力特征图 M_{ar} , 如式(3)和式(4)所示:

$$M_{sr}(m, n, c) = \max_{(i, j) \in bin(m, n)} W_{sr}(i, j, c + (k_{sr}-1)C) \quad (3)$$

$$M_{ar}(m, n, c) = \max_{(i, j) \in bin(m, n)} W_{ar}(i, j, c + (k_{ar}-1)C) \quad (4)$$

其中, $1 \leq m \leq h, 1 \leq n \leq w, 1 \leq c \leq C$, 且 M_{sr} 和 M_{ar} 维

度大小均为 $h \times w \times C$ 。合并这两个注意力特征图, 并对初始的ROI池化特征进行加权激活, 然后将优化后的ROI特征输入到检测子网络中即可进行分类回归预测。

在模型训练过程中, 选择性ROI池化层遵循链式求导法则, 根据上述对应关系将注意力特征图的误差反向传播到注意力特征库中的对应位置, 并采用随机梯度下降算法进行参数更新。

2.3 特征降维

ROI池化特征通常是高维的, 如ResNet-50^[20]模型中为2048维。为了对高维特征进行建模, 通常需要高容量的检测子网络, 这将导致大规模的参数和运算。进行区域分类和回归时, ROI特征是否真的需要这么多通道呢? 考虑到生成的候选区域具有一定的准确性, 大部分ROI已经覆盖了相当一部分目标对象。并且引入注意力模块后, 可以针对ROI子区域特征和宽高比特性生成具有辨识性的特征分量。因此可考虑适当降低ROI池化特征的维度, 并以注意力特征图的形式保证ROI分类特征的有效性。当采用 1×1 卷积核将特征通道数从 C 维减少到 C_s 维时, 注意力特征库的维度需要分别设置为 $H \times W \times N_{sr}C_s$ 和 $H \times W \times N_{ar}C_s$ 。经过特征降维处理和注意力模块的特征再提取操作后, 生成的ROI分类特征更加简洁且信息更加丰富, 采用低容量的检测子网络即可完成分类回归预测。

3 实验结果与分析

为了验证注意力模块的有效性, 将在VOC0712数据集和安检X光图像数据集上进行实验, 实验数据示例图如图3所示。VOC0712数据集共包括21类别, 以VOC2007^[21]和VOC2012^[22]的训练数据集进行训练, 并在VOC2007的测试数据集上验证检测精度和检测速度; 安检X光图像数据集包括刀具、雨伞等10个类别, 共10889张图像, 由人工标注而成。

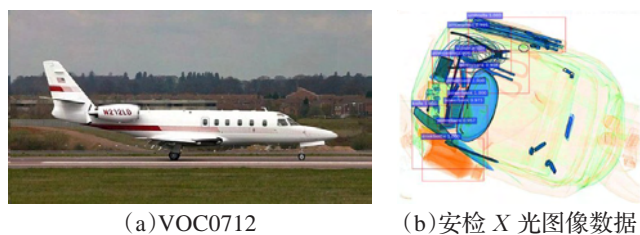


图3 实验数据示例图片

实验采用caffe^[23]深度学习框架, 以ResNet-50深度残差网络作为特征提取基础网络, 并以ImageNet^[24]数据集预训练好的模型进行参数初始化。采用随机梯度下降算法(Stochastic Gradient Descent, SGD^[25])对模型进行训练, 初始学习率设置为0.002, 冲量为0.9, 权重衰减为0.0005。

实验过程中, 将ROI划分为 3×3 的9个子区域和3个宽高比类别, 即 $Ratio < 0.75$ 、 $0.75 < Ratio < 1.3$ 和 $Ratio > 1.3$, 并设定降维后ROI特征维度为50。在res5c

卷积层后设置好子区域注意力特征库和宽高比注意力特征库后,利用选择性池化操作生成子区域注意力特征图和宽高比注意力特征图。为了便于分析,将设置三组对照实验,实验(a)只使用子区域注意力特征图,实验(b)只使用宽高比注意力特征图,实验(c)同时使用子区域注意力特征图和宽高比注意力特征图。除此之外,实验还复现了Faster R-CNN和R-FCN模型的检测结果,具体结果见表1和表2。

表1 VOC0712数据集检测结果

方法	子区域	宽高比	检测速度/s	mAP/%
Faster R-CNN			0.335	74.9
R-FCN			0.204	75.5
Ours(a)	√		0.209	77.9
Ours(b)		√	0.209	76.4
Ours(c)	√	√	0.209	78.2

表2 安检X光图像数据集检测结果

方法	子区域	宽高比	检测速度/s	mAP/%
Faster R-CNN			0.224	72.2
R-FCN			0.121	78.5
Ours(a)	√		0.169	79.3
Ours(b)		√	0.169	78.8
Ours(c)	√	√	0.169	79.7

目标检测模型常用的精度评价指标为平均精度均值(mean Average Precision, mAP),其主要与准确率Precision和召回率Recall有关。根据测试结果,每个类别都可得到一条PR(Precision-Recall)曲线,计算曲线下的面积可得到该类别的平均精度AP,mAP是各类别AP的平均值。

在VOC0712数据集上,Faster R-CNN模型的mAP为74.9%,而R-FCN模型的mAP为75.5%。相比之下,只使用子区域注意力特征图时模型(a)的mAP达到了77.9%,相对于R-FCN模型的精度提高了2.4%。而在宽高比信息的帮助下,模型(c)的mAP进一步达到了78.2%。在安检X光图像数据集上,也取得了类似的性能提升效果,比如同时使用子区域注意力特征图和宽高比注意力特征图的模型(c)精度达到了79.7%,相对于R-FCN模型提升了1.2%。这些结果充分验证了子区域注意力特征图和宽高比注意力特征图的有效性。

为了更直观地展示各模型的效果,将利用文献[26]中的可视化算法对Faster RCNN、R-FCN和引入注意力模块后模型的卷积特征图进行可视化,具体效果如图4所示。从中可直观地观察到,R-FCN模型利用位置敏感分数图对子区域信息进行编码后,一定程度上保持了子区域特征,但整体结构较为混乱。而引入注意力模块后,模型在保持局部结构特征的同时,也较好地维护了目标对象的空间结构关系。

引入注意力模块后,可以将ROI特征维度进行一定程度的压缩,这有助于减少模型容量并加快检测速度。

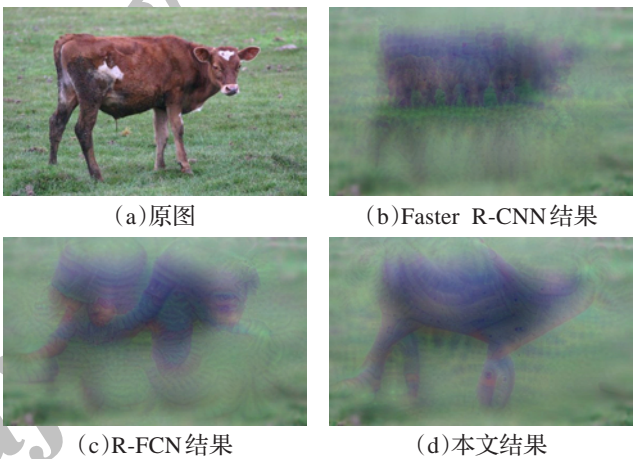


图4 特征可视化效果图(VOC0712数据图)

表3展示的是将ROI特征维度进行不同程度压缩后,在安检X光图像数据集上的检测结果。当卷积特征的通道数设置为1 000时,因为引入注意力模块后模型的容量过于庞大,在相同的训练迭代次数下,模型难以得到充分的训练,所以检测精度有限。若提供充分的训练数据和计算资源,模型应该能取得更好的检测效果。将特征通道数压缩至256维甚至50维时,模型仍取得了较好的检测结果,且特征维度的缩减对模型检测精度的影响不大。当维度压缩得过低时,检测精度有下降趋势,不过依然维持在较高的水平。这说明注意力模块可以有效地抽取ROI的主要特征,精炼的低维度特征即能满足检测子网络的识别需求,从而进一步验证了注意力特征模块的有效性。

表3 特征降维后模型的检测结果(X光图像数据集)

ROI特征维度	1 000	256	100	50	20
mAP/%	79.5	79.7	79.7	79.7	79.3

模型最终选择将特征通道数缩减至50维,并配备轻量级(如fc512+fc512+fc10)检测子网络进行分类回归预测。表1和表2中所示的检测速度是在GeForce GTX 1080 Ti显卡机器上的测试结果。从中可以看出,没有经过特征优化的Faster R-CNN模型只能使用重量级的检测子网络,检测速度较慢;R-FCN模型不采用检测子网络,而是使用位置敏感池化操作计算子区域的分类分数,并将其平均为最终的ROI分数,检测速度较快;而引入了注意力模块和降维操作的本文模型,在进行精简和有效的特征再提取后,检测速度与R-FCN模型相当,同时检测精度还得到了明显的提升。

4 结束语

本文在目前流行的两阶段目标检测框架中引入了注意力模块。首先基于ROI子区域特征和宽高比属性的差异,生成注意力特征库,然后以选择性池化的方式为每个ROI生成特定的子区域注意力特征图和宽高比注意力特征图。一方面,注意力特征图可以优化原始的

ROI池化特征,增强模型的定位识别能力;另一方面,生成的注意力特征图使得特征降维成为可能,有助于减少模型参数和计算量。实验结果表明,引入注意力模块后,模型的检测精度和检测速度都有明显的提升。

除了目标对象的子区域特征和宽高比属性值得关注以外,目标对象之间的相互关系以及场景信息等也是非常有意义的研究内容。另外,半监督、无监督目标检测算法及目标检测模型的域自适应能力^[27]等也都值得深入研究。后续将会围绕这些方面继续展开研究,期待目标检测领域有更大的突破和创新。

参考文献:

- [1] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]//Advances in Neural Information Processing Systems, 2012:1097-1105.
- [2] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//2013 IEEE Conference on Computer Vision and Pattern Recognition, 2013:580-587.
- [3] Girshick R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision, Santiago, 2015:1440-1448.
- [4] Ren S Q, He K, 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.
- [5] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.
- [6] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, 2017: 6517-6525.
- [7] Redmon J, Farhadi A. Yolov3: an incremental improvement[J]. arXiv: 1804.02767, 2018.
- [8] Liu W, Anguelov D, Erhan D, et al. SSD: single shot multibox detector[C]//European Conference on Computer Vision, 2016:21-37.
- [9] Li Z, Zhou F. FSSD: feature fusion single shot multi-box detector[J]. arXiv: 1712.00960, 2017.
- [10] Fu C, Liu W, Ranga A. DSSD: deconvolutional single shot detector[J]. arXiv: 1701.06659, 2017.
- [11] 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, 2016:845-853.
- [12] Lin T Y, Dollar P, Girshick R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, 2017:1-8.
- [13] Yang F, Choi W, Lin Y. Exploit all the layers: fast and accurate CNN object detector with scale dependent pooling and cascaded rejection classifiers[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2016:2129-2137.
- [14] Gidaris S, Komodakis N. Object detection via a multi-region and semantic segmentation-aware CNN model[C]//2015 IEEE International Conference on Computer Vision, 2015:1134-1142.
- [15] Dai J, Li Y, He K, et al. R-FCN: object detection via region-based fully convolutional networks[C]//2016 Advances in Neural Information Processing Systems, 2016:379-387.
- [16] Dai J, Qi H, Xiong Y, et al. Deformable convolutional networks[C]//IEEE International Conference on Computer Vision, 2017:764-773.
- [17] Yin W, Schutze H, Xiang B, et al. ABCNN: attention-based convolutional neural network for modeling sentence pairs[J]. Transactions of the Association for Computational Linguistics, 2016, 6(4): 259-272.
- [18] Zhou B, Khosla A, Lapedriza A, et al. Learning deep features for discriminative localization[C]//2016 of IEEE Conference on Computer Vision and Pattern Recognition, 2016:2921-2929.
- [19] Zhai Y, Fu J, Lu Y, et al. Feature selective networks for object detection[C]//2018 IEEE Conference on Computer Vision and Pattern Recognition, 2018:4139-4147.
- [20] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016:770-778.
- [21] Everingham M, Van Gool L, Williams C K I, et al. The PASCAL visual object classes (VOC) challenge[J]. International Journal of Computer Vision, 2010, 88(2):303-338.
- [22] Everingham M, Eslami M A, Gool L V, et al. The PASCAL visual object classes challenge: a retrospective[J]. International Journal of Computer Vision, 2015, 111(1):98-136.
- [23] Jia Y, Shelhamer E, Donahue J, et al. Caffe: convolutional architecture for fast feature embedding[J]. arXiv: 1408.5093, 2014.
- [24] Deng J, Dong W, Socher R, et al. Imagenet: a large-scale hierarchical image database[C]//2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009:248-255.
- [25] LeCun Y, Boser B, Denker J, et al. Back propagation applied to handwritten zip code recognition[J]. Neural Computation, 1989, 1(4):541-551.
- [26] Zeiler M, Fergus R. Visualizing and understanding convolutional networks[C]//European Conference on Computer Vision, 2014:818-833.
- [27] Chen Y, Li W, Sakaridis C, et al. Domain adaptive faster R-CNN for object detection in the wild[C]//2018 IEEE Conference on Computer Vision and Pattern Recognition, 2018:3339-3348.