

# 基于改进的特征提取网络的目标检测算法

乔婷, 苏寒松, 刘高华\*, 王萌

天津大学电气自动化与信息工程学院, 天津 300072

**摘要** 针对目标检测准确率低, 物体位置不精准的缺点, 设计了一种基于改进的特征提取网络的目标检测算法。首先将训练集进行数据增强; 其次设计了一种双通道网络, 用于目标检测算法 Faster R-CNN 的特征提取; 最后在算法的预测部分, 对非极大值抑制(NMS)机制进行了改进, 并采用加权平均方法获取存在多个相近的预测框的位置。在 VOC 2007 和 VOC 2012 数据库上进行实验, 表明本文算法比经典的目标检测算法效果要好, 准确率达到 79.1%, 提升了 3%~4%, 验证了本文算法的有效性。

**关键词** 图像处理; 深度学习; 目标检测; 特征提取; 卷积神经网络

中图分类号 TP391

文献标识码 A

doi: 10.3788/LOP56.231008

## Object Detection Algorithm Based on Improved Feature Extraction Network

Qiao Ting, Su Hansong, Liu Gaohua\*, Wang Meng

School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China

**Abstract** In this study, an object detection algorithm is designed based on an improved feature extraction network to solve the shortcomings of low object detection accuracy and inaccurate object position detection. Initially, the training set is enhanced; subsequently, a two-path network is designed for usage in feature extraction of the Faster R-CNN algorithm; finally, the non-maximum suppression mechanism is improved in the prediction part of the algorithm, and the weighted averaging method is adopted for obtaining the positions of multiple similar prediction boxes. The experiments conducted using the VOC 2007 and VOC 2012 databases denote that the proposed algorithm outperforms the classical object detection algorithm, with an accuracy rate of 79.1% and an improvement of 3%-4%. Thus, the effectiveness of the algorithm is verified.

**Key words** image processing; deep learning; object detection; feature extraction; convolution neural network

**OCIS codes** 100.4996; 330.1880; 100.3008; 200.4260

## 1 引言

随着人工智能时代的到来, 目标检测已经成为人工智能及计算机视觉领域中的一个研究重点和难点, 它的主要目标是识别出图像中感兴趣区域中的物体类别, 并预测出物体的具体位置。目标检测在各个领域都有着非常重要的应用, 如交通运输、人脸识别<sup>[1]</sup>、军事国防<sup>[2]</sup>、医学诊断、视频监控等领域。

基于深度学习的目标检测算法主要可以分为 2 大类: 一类是基于区域提名的算法, 通过区域建议的

方法生成一系列候选区域, 对候选区域进行特征提取, 然后进行分类和位置回归, 经典的算法有 R-CNN<sup>[3]</sup>、Fast R-CNN<sup>[4]</sup>、Faster R-CNN<sup>[5]</sup> 等; 另一类是端到端的算法, 通过将输入图像划分成一系列的网格, 对网格进行深度特征提取, 最后完成目标的分类识别, 该算法省去了生成候选框过程, 将目标检测问题转化为回归问题, 在保证一定精确度的前提下, 能够有效提高检测速度, 经典的算法有 YOLO<sup>[6]</sup>、SSD<sup>[7]</sup> 等。还有很多学者尝试着在此两类算法上进行创新, 如在单次多检测框检测(SSD)

收稿日期: 2019-04-26; 修回日期: 2019-05-24; 录用日期: 2019-06-03

基金项目: 国家自然科学基金(61471260)

\* E-mail: suppig@126.com

基础上进行创新的有 Faster-SSD 算法<sup>[8]</sup>,在改善感兴趣区域(ROI)分类效果上进行创新的有 R-FCN 算法<sup>[9]</sup>,除此之外还有基于自适应卷积特征的目标跟踪算法<sup>[10]</sup>等。

虽然现阶段优秀的目标检测算法有很多,但还是存在识别准确率不高的问题。光照、天气等环境干扰因素,对输入图像特征提取不充分,以及去除多余重复框后确定目标位置时准确率不高,都会影响目标检测效果。本文提出一种基于改进的卷积神经网络(CNN)特征提取网络的目标检测算法,旨在提高目标检测精度。首先将训练集进行数据增强,并对数据增强技术作定量分析;其次设计一种只利用 $1\times 1$ 、 $3\times 3$ 、 $1\times 1$ 卷积的双通道网络,用作目标检测算法 Faster R-CNN 的特征提取网络,对输入图像特征进行深层次利用;最后对非极大抑制(NMS)算法进行改进,使得预测阶段中含有目标的检测框能够因置信度的降低机制而更多地保留下来。

## 2 原理与方法

### 2.1 数据增强

利用数据增强来提高目标检测效果是一个很重要的环节,如果增强方法选择不当,则会降低准确性和稳健性。所以针对翻转、裁剪、旋转、添加高斯噪声 4 种方法,在 ImageNet 数据集的子集上,采用 AlexNet<sup>[11]</sup>对分类效果进行定量讨论。该子集共包含 15 类物体,每类物体有 500 张训练图片,输入图片大小为 $224\times 224\times 3$ 。实验硬件设备是 Ubuntu 16.04 操作系统,语言版本是 Python 2.7,使用 Caffe 平台。用这 4 种方法及其组合分别生成新的样本数量,数量是原始训练集的 1 倍和 3 倍,分别进行对比实验。

图 1 是 4 种方法各自生成 1 倍和 3 倍数据集的实验结果。从结果可知,1)翻转、裁剪、旋转 3 种方法,可以提高训练模型准确率,并且随着训练样本的成倍增加,效果越好;2)添加高斯噪声不能改善训练模型的效果;3)单独使用翻转方法效果较好,在只增加 1 倍数据量的情况下,准确率高出旋转方法 1.8%,高出裁剪方法 1.1%,在增加 3 倍数据量的情况下,准确率高出旋转方法 0.3%,高出裁剪方法 1.4%。

由于添加高斯噪声并不能改善结果,于是对剩下 3 种方法的组合作进一步分析。图 2 表明 2 种方法结合效果要优于只使用 1 种方法,3 种方法相结合时效果最好。在只增加 1 倍数据量和 3 倍数据量

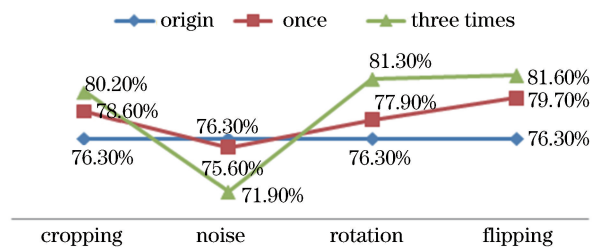


图 1 4 种数据增强技巧效果对比图

Fig. 1 Comparison of effects of four data augmentation techniques

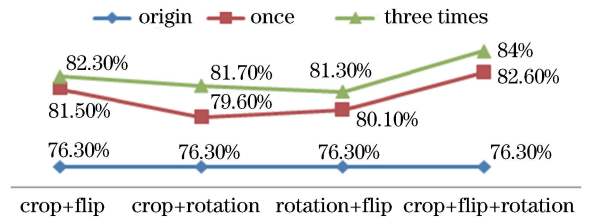


图 2 翻转、裁剪、旋转方法组合效果对比图

Fig. 2 Comparison of combination effects of flipping, cropping, and rotating methods

的情况下,准确率分别是 82.60% 和 84%。考虑到目标检测算法所用的 Pascal VOC 数据库容量并不大(2G),所以在实验结果与分析部分,对训练数据采取 3 种方法组合的形式,增加 3 倍数据集,以达到最好的检测效果。

### 2.2 改进的特征提取网络

采取最新的网络内部链接拓扑结构,结合 DensNet<sup>[12]</sup>和 ResNet<sup>[13]</sup>的优点,设计一种双通道特征提取网络,如图 3 所示。对于特征提取网络(表 1),有两条输入通道可以传输:一是经过 $1\times 1$ 、 $3\times 3$ 和 $1\times 1$ 卷积,与下方输入作相加操作,形成 ResNet 通道,利用跳跃连接技巧使得网络能够重复利用特征,减少冗余,提高传输效率;二是经过 $1\times 1$ 、 $3\times 3$ 和 $1\times 1$ 卷积,与上方输入作拼接操作,形成 DensNet 通道,利用密集连接技巧将网络的输入特征和输出特征拼接,从之前的网络中学习新的特征。

该特征提取网络能够更好地提取图片的信息,提高特征的复用和传播,同时,还能够通过共享中间三个卷积,提高网络的计算效率,能在减少参数量上取得较大的优势,最终提高了模型的性能,降低目标检测的错误率,相对于其他检测方法具有较高的精度。

为突出特征网络在参数量上的优势,这里选择较复杂的 100 层网络、VGG-16 网络、ResNet、DensNet 进行参数量对比。超参数  $K$  表征特征提取网络的增长率, $\alpha_1$ 、 $\alpha_2$ 、 $\alpha_3$ 、 $\alpha_4$  为网络层数的不同数量。从表 2 可以看出当  $K=48$  时,设计的 100 层特

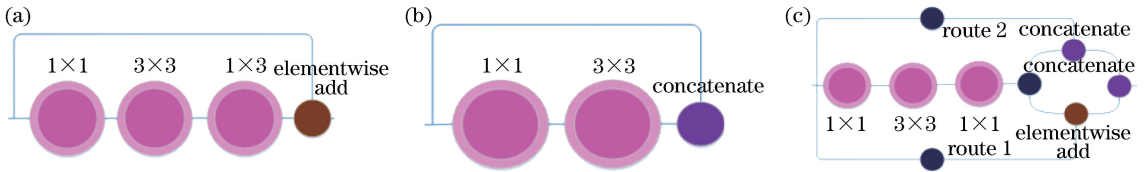


图3 ResNet、DensNet、双通道网络单元结构节点表示图。(a) ResNet 网络单元；(b) DensNet 网络单元；(c) 双通道网络单元  
 Fig. 3 Node representations of cell structures of ResNet, DensNet, and two-path networks. (a) ResNet network; (b) DensNet network; (c) two-path network

表1 特征提取网络结构

表2 不同特征提取网络的复杂性比较

Table 1 Structure of feature extraction network

Table 2 Comparison of complexity of different feature extraction networks

Layer	Output size	Detail
Conv1	112×112	7×7, 64, stride 2 3×3 max pool, stride 2
Conv2	56×56	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \\ 1 \times 1 \text{ conv} \end{bmatrix} \times \alpha_1$
Conv3	28×28	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \\ 1 \times 1 \text{ conv} \end{bmatrix} \times \alpha_2$
Conv4	14×14	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \\ 1 \times 1 \text{ conv} \end{bmatrix} \times \alpha_3$
Conv5	7×7	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \\ 1 \times 1 \text{ conv} \end{bmatrix} \times \alpha_4$

Feature extraction network	Depth	Parameter /10 <sup>6</sup>
VGG-16	16	168
DensNet(K=48)	161	111
ResNet	101	150
Ours( $\alpha_1 \alpha_2 \alpha_3 \alpha_4 = 6, 8, 16, 3; K=48$ )	100	134

征提取网络,相较于 VGG-16 特征提取网络,参数量减少了 20.24%,相较于 ResNet 特征提取网络,参数量减

少了 10.67%,因此在计算量上获得了更好的性能。

### 2.3 改进的非极大值抑制算法

Faster R-CNN 算法中建议窗口生成网络在计算过程中会得到 2 万多个带有置信值的预测框,这些预测框通过 NMS 的方法进行去冗余。但该算法仍然存在缺陷:如图 4(a)所示,图中虚线预测框 A 会因与实线预测框 B 重叠面积过大而被删除,从而导致 A 框漏检;同样如图 4(b)所示,鸟周围存在多个相近的预测框,传统的 NMS 无法保留最准确的那个框。

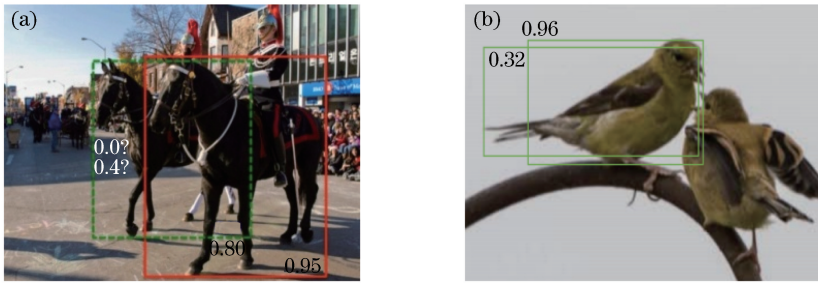


图4 传统的 NMS 问题举例图。(a)马；(b)鸟

Fig. 4 Examples of traditional NMS problems. (a) Horses; (b) birds

针对该问题作进一步改进,将所有预测框按置信度分数排序,分数最高的预测框选作对比框,剩下的预测框与对比框的重叠面积大于交并比(IoU,  $R_{IoU}$ )阈值<sup>[14]</sup>时,将该预测框的置信度分数与衰减函数相乘。为保证分数降低后连续,并且只对 IoU 重叠度大的有降低效果,选择高斯的衰减函数,  $e^{-\beta(R_{IoU})^2}$ ,其中  $\beta$  定义为  $\frac{1}{\sigma}$ 。为比较改进算法与传统算法的性能和稳健性,选择在一组公开提供的

MS-COCO 数据集上的 5000 张图片上进行实验。从表 3 得知,普通 NMS 算法随着 IoU 阈值的提升下降幅度较大,而改进后的 NMS 虽也有下降,但整体幅度较为稳定,并且在相同阈值下,平均准确率均高于普通 NMS 算法。随着  $\beta$  的减小,改进后的 NMS 平均准确率先提升后下降,在此数据集下 IoU 阈值为 0.5、 $\sigma$  为 0.6 时能达到最好的效果。为使对比公平准确,在后文实际训练中 IoU 阈值选择和原算法保持一致,设为 0.7,同时  $\sigma$  设为 0.6。

表3 IoU 阈值、 $\beta$  参数、加权平均对平均值 (AP) 的影响 (0.5、0.6、0.7 代表不同的 IoU 阈值, w 代表加权平均)Table 3 Influences of IoU threshold,  $\beta$  parameter, and weighted average on AP (0.5, 0.6, and 0.7 represent different IoU thresholds; w represents weighted average)

Different parameter	AP <sup>0.5</sup>	AP <sup>0.5w</sup>	AP <sup>0.6</sup>	AP <sup>0.6w</sup>	AP <sup>0.7</sup>	AP <sup>0.7w</sup>
Normal NMS	44.37	44.83	39.18	39.67	29.83	30.34
$\beta=2.5, \sigma=0.4$	46.42	46.92	42.83	43.40	34.68	35.24
$\beta=1.67, \sigma=0.6$	46.58	47.11	43.30	43.79	35.21	35.76
$\beta=1.25, \sigma=0.8$	45.93	46.45	41.68	42.21	33.01	33.53

除此之外, 针对一个物体周围存在多个位置相近的预测框的情况, 选择生成一个组合框, 即对这几个相近预测框位置作加权平均, 进一步改善预测框的位置。从表 3 中可以看出, 加权平均对 AP 值能够起到积极作用, 平均准确率能够提高的范围在 0.45~0.60 之间。对于预测框位置不太精确且距离范围特别大的情况, 在未来的研究工作中, 需要对加权平均方法作进一步改进, 以提高算法的普适性。

### 3 实现结果与分析

实验使用的环境配置为 CPU Intel (R) Core (TM) i7-6700, GPU 型号为 NVIDIA GTX 1080Ti, 操作系统是 Ubuntu16.04, 基于 Caffe 框架, python 2.7 版本, MATLAB 2014b 实现。本文采用端到端的联合训练法, 占用显存小, 并且训练速度快。

表4 数据增强和改进的 NMS 机制对准确率的影响

Table 4 Influences of data augmentation and improved NMS mechanism on accuracy

Detection framwork	Backbone	Training set	Testing set	mAP / %
Ours	Proposed	VOC2007+VOC2012	VOC2007	79.1
No augmentation	Proposed	VOC2007+VOC2012	VOC2007	76.6
No improved NMS	Proposed	VOC2007+VOC2012	VOC2007	78.0

#### 3.2 训练时热身 (warming up) 技巧的影响

本实验采取热身技巧设置了学习率, 在训练初始阶段, 先以小学习率训练一定次数, 然后再将学习率提升回来。对训练的前  $N$  个 epoch 采取 warming up, 并对  $N$  进行相应的讨论。由表 5 可知将前 5 个 epoch 初始学习率分别设置为 0.00001、0.0001、0.001、0.01、0.1 时效果最好, 能够提高 0.51%

表5 不同 epoch 对准确率的影响结果

Table 5 Influences of different epochs on accuracy

Nums of epoch	Learning rate setting	mAP / %
0	No warming up	78.20
2	0.01, 0.1	78.25
3	0.001, 0.01, 0.1	78.36
4	0.0001, 0.001, 0.01, 0.1	78.67
5	0.00001, 0.0001, 0.001, 0.01, 0.1	78.71

采用随机梯度下降法和反向传播进行训练, 动量设置为 0.9, 批次 (batch size) 设置为 16, 权重衰减 (weight decay) 设置为 0.0005。由于使用数据增强导致数据集增多, 因此训练次数也增加, 设置为 110 万次。起初学习率 (learning rate) 设置为 0.1, 迭代 65 万次后学习率降为 0.01, 继续迭代至 90 万次, 学习率降为 0.001, 直到训练结束, 得到最终模型。

#### 3.1 数据增强和改进的非极大值抑制的影响

表 4 表明使用数据增强和改进的 NMS 可以起到积极效果。将 VOC2007 数据集中的 5011 张训练集和 VOC2012 数据集中的 11540 张训练集进行数据增强, 准确率提高了 2.5%; 使用改进的 NMS 机制, 更多地保留了预测阶段中含有目标的检测框, 并且对存在多个相近的预测框进行加权平均, 使本文算法的准确度提高了 1.1%。

的准确率。该技巧让学习率有一个缓慢的提升过程, 有利于网络跨越随机梯度下降时的鞍点, 使网络更快地收敛, 同时提高了最终准确率。

#### 3.3 网络深度和网络增长率的影响

设计 52 层、100 层和 133 层 3 种不同深度的特征提取网络,  $\alpha_1, \alpha_2, \alpha_3, \alpha_4$  分别是 (3, 4, 6, 4)、(6, 8, 16, 3) 和 (6, 8, 26, 4), 图 5 中数据点从左至右依次代表 52 层、100 层和 133 层, 并且设置增长率为 12。由图 5 可知随网络层数的加深, 目标检测的 Top-1 错误率不断下降, 133 层的预测网络比 52 层预测网络错误率下降 1.5%, 比 100 层预测网络错误率下降 0.8%。由此可知双通道网络能够吸取 ResNet 和 DensNet 的优点, 当网络深度不断加深时, 能够很好地解决梯度爆炸或网络退化问题, 利用跳跃连接的技巧, 使得网络易于优化与泛化。



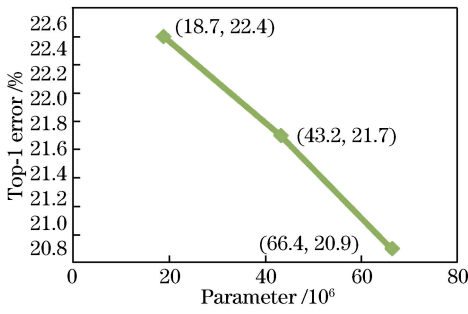


图5 52层、100层、133层深度的特征提取网络参数数量和 Top-1 错误率趋势图

Fig. 5 Trend of parameter quantity of feature extraction network with Top-1 error rate and 52, 100, and 133 layers

关于网络增长率  $K$  对比实验,为减少训练时间,选择 52 层网络,并设计了  $K$  为 12、18、24、48 四种情况(图 6 中数据点从左到右依次顺序排列)。由图 6 可得当特征提取网络层数固定时,随着  $K$  的增加,网络参数不断增加,同时 Top-1 错误率不断下降,这意味着准确率不断提高,模型性能不断提升。 $K$  为 48 时,Top-1 错误率比  $K$  为 18 时下降了 0.8%。证明该特征网络能够通过拼接前一层的特征,增加后续层的输入特征变化,更好地提取了图片的信息,提高了特征的复用和传播,从而能够学习到更加精准的模型。

### 3.4 检测精度和检测速度

表6为本文算法和其他算法的检测精度和检测

表6 不同算法在 VOC2007+VOC2012 训练集下的测试结果

Table 6 Testing results of different algorithms under VOC2007+VOC2012 training sets

Method	Backbone	Training set	Testing set	mAP/%	Frame rate / (frame · s <sup>-1</sup> )	
Fast R-CNN	VGG-16	VOC2007+VOC2012	VOC2007	70.0	0.50	
Faster R-CNN	VGG-16	VOC2007+VOC2012	VOC2007	73.2	7	
Two stage	Faster R-CNN	ResNet-101	VOC2007+VOC2012	VOC2007	76.4	2.40
	MR-CNN	ResNet-101	VOC2007+VOC2012	VOC2007	78.2	0.03
	ION	VGG-16	VOC2007+VOC2012	VOC2007	76.5	1.25
	Ours	Proposed	VOC2007+VOC2012	VOC2007	79.1	2.10
YOLO	GoogleNet	VOC2007+VOC2012	VOC2007	63.4	45	
YOLOv2	Darknet-19	VOC2007+VOC2012	VOC2007	78.6	40	
One stage	SSD321	ResNet-101	VOC2007+VOC2012	VOC2007	77.1	11.20
	SSD300 *	VGG-16	VOC2007+VOC2012	VOC2007	77.2	46
	DSOD300	DS/64-192-48-1	VOC2007+VOC2012	VOC2007	77.7	17.40
	DSSD513	ResNet-101	VOC2007+VOC2012	VOC2007	81.5	5.50

## 4 结 论

提出一种基于改进的特征提取网络的目标检测算法,将 Faster R-CNN 的特征提取网络替换成双通道网络;同时还对训练集进行了翻转、裁剪、旋转

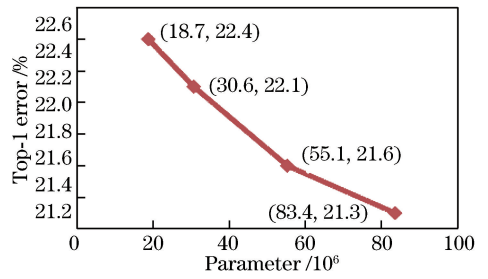


图6 网络增长率为 12,18,24,48 的特征提取网络的参数数量和 Top-1 错误率趋势图

Fig. 6 Trend of parameter quantity of feature extraction network with Top-1 error rate and network growth rates of 12, 18, 24, and 48

速度对比,并将目标检测主流算法做了分类。本文算法(网络层数为 133 层,网络增长率为 48)的检测精度为 79.1%,超过了 one stage 和 two stage 大部分算法,高于基于 VGG-16 的 Faster R-CNN 算法 5.9 个百分点,高于基于 ResNet-101 的 Faster R-CNN 算法 2.7 个百分点,充分说明本算法设计的特征提取网络起到了很好的效果,从而验证了本文算法的有效性。本文算法的检测速度是 2.10 frame/s,由于输入图像需要同时经过两路较大的网络,造成计算量比简单的特征提取网络要大,所以检测速度并没有明显的提升。但是本文算法的检测速度,优于 MR-CNN 算法 2.07 frame/s,优于 Fast R-CNN 算法 1.60 frame/s,在 two stage 系列算法中具有一定优势。

数据增强操作;在最后的预测部分改进了 NMS 算法,对所有 IoU 大于阈值的预测框的置信度分数进行降低,并对存在多个相近的预测框的位置进行加权平均以提高准确度。实验结果证明,本文算法的检测精度为 79.1%,检测速度为 2.1 frame/s,相比

于其他主流算法,有效提高了检测精度,同时在 two stage 系列算法中具有一定的检测速度优势。由于 two stage 系列的目标检测算法检测速度较低,难以满足不同情境下的实时性要求,所以在未来算法改进中,希望能够在保证检测精度的基础上,提高检测速度,以满足实时性的要求。

## 参 考 文 献

- [1] He Z C, Zhao L Z, Chen C. Convolution neural network with multi-resolution feature fusion for facial expression recognition[J]. *Laser & Optoelectronics Progress*, 2018, 55(7): 071503.  
何志超, 赵龙章, 陈闯. 用于人脸表情识别的多分辨率特征融合卷积神经网络[J]. *激光与光电子学进展*, 2018, 55(7): 071503.
- [2] Chen Y, Fan R S, Wang J X, *et al.* Cloud detection of ZY-3 satellite remote sensing images based on deep learning[J]. *Acta Optica Sinica*, 2018, 38(1): 0128005.  
陈洋, 范荣双, 王竞雪, 等. 基于深度学习的资源三号卫星遥感影像云检测方法[J]. *光学学报*, 2018, 38(1): 0128005.
- [3] 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, June 23-28, 2014, Columbus, OH, USA. New York: IEEE, 2014: 580-587.
- [4] Girshick R. Fast R-CNN [C] // 2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE, 2015: 1440-1448.
- [5] 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.
- [6] 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 (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 779-788.
- [7] Liu W, Anguelov D, Erhan D, *et al.* SSD: single shot MultiBox detector[M] // Leibe B, Matas J, Sebe N, *et al.* Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9905: 21-37.
- [8] Chen L L, Zhang Z D, Peng L. Real-time detection based on improved single shot MultiBox detector[J]. *Laser & Optoelectronics Progress*, 2019, 56(1): 011002.  
陈立里, 张正道, 彭力. 基于改进 SSD 的实时检测方法[J]. *激光与光电子学进展*, 2019, 56(1): 011002.
- [9] Dai J F, Li Y, He K M, *et al.* R-FCN: object detection via region-based fully convolutional networks[C] // Proceedings of the 30th International Conference on Neural Information Processing Systems, December 5-10, 2016, Barcelona, Spain. New York: ACM, 2016: 379-387.
- [10] Cai Y Z, Yang D D, Mao N, *et al.* Visual tracking algorithm based on adaptive convolutional features[J]. *Acta Optica Sinica*, 2017, 37(3): 0315002.  
蔡玉柱, 杨德东, 毛宁, 等. 基于自适应卷积特征的目标跟踪算法[J]. *光学学报*, 2017, 37(3): 0315002.
- [11] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C] // Proceedings of the 25th International Conference on Neural Information Processing Systems, December 3-6, 2012, Lake Tahoe, Nevada. New York: ACM, 2012: 1097-1105.
- [12] Huang G, Liu Z, van der Maaten L, *et al.* Densely connected convolutional networks[C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 2261-2269.
- [13] He K M, Zhang X Y, Ren S Q, *et al.* Deep residual learning for image recognition [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 770-778.
- [14] Neubeck A, van Gool L. Efficient non-maximum suppression[C] // 18th International Conference on Pattern Recognition (ICPR'06), August 20-24, 2006, Hong Kong, China. New York: IEEE, 2006: 9210072.