

引用格式: LIUNan, MAO Zhaoyong, WANG Yichen, et al. Remote Sensing Images Target Detection Based on Adjustable Parameter and Receptive field[J]. Acta Photonica Sinica, 2021, 50(11):1128001

刘楠,毛昭勇,王亦晨,等. 基于参数量和感受野可调的遥感目标检测算法[J]. 光子学报, 2021, 50(11):1128001

基于参数量和感受野可调的遥感目标检测算法

刘楠¹, 毛昭勇², 王亦晨², 沈钧戈²

(1 西北工业大学 航海学院, 西安 710072)

(2 西北工业大学 无人系统技术研究院, 西安 710072)

摘 要: 针对光学遥感图像目标检测存在多尺度目标检测及实时性不佳的问题, 提出一种基于参数量和感受野可调的遥感目标检测算法, 在保证精度的基础上可同时兼顾实时性。该算法以更快的区域卷积神经网络为基础, 设计通道数可调模块和感受野可调模块, 分别提升速度和精度, 同时令全连接层维度根据目标类别数动态变化, 以减少参数冗余。在 DIOR 遥感数据集上的实验结果表明, 本文方法在达到最高精度时高于所有对比算法, 检测速度比 Faster R-CNN 略高; 在达到最高速度时, 可达到实时性且同时具有较高的精度。

关键词: 遥感图像; 目标检测; 深度学习; 感受野; 网络参数

中图分类号: TP751

文献标识码: A

doi:10.3788/gzxb20215011.1128001

Remote Sensing Images Target Detection Based on Adjustable Parameter and Receptive field

LIU Nan¹, MAO Zhaoyong², WANG Yichen², SHEN Junge²

(1 School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an 710072, China)

(2 Unmanned System Research Institute, Northwestern Polytechnical University, Xi'an 710072, China)

Abstract: In order to solve the problem of multi-scale and poor real-time performance in optical remote sensing image detection, a remote sensing target detection algorithm based on the adjustable parameter number and receptive field is proposed, which can not only reach high detection accuracy, but also achieve real-time performance. Based on the faster region-convolution neural network, a receptive field adjustable module and a channel number adjustable module are designed to improve the accuracy and speed respectively. At the same time, in order to reduce parameter redundancy, the dimension of the full connection layer changes dynamically according to the number of target categories. Experimental results on remote sensing datasets of DIOR, show that the proposed method is higher than all the comparisons with the highest accuracy, and the detection speed is higher than Faster R-CNN. When our algorithm achieved highest speed, it can achieve real-time requirement with proper precision.

Key words: Remote sensing imaging; Target detection; Deep learning; Receptive field; Network parameter

OCIS Codes: 280.4788; 100.3008; 100.4996; 110.2970

基金项目: 国家自然科学基金(No.61603233)

第一作者: 刘楠(1999—), 女, 硕士研究生, 主要研究方向为深度学习、目标检测。Email: 2020200646@mail.nwpu.edu.cn

导师(通讯作者): 沈钧戈(1987—), 女, 助理教授, 博士, 主要研究方向为人工智能与计算机视觉。Email: shenjunge@nwpu.edu.cn

收稿日期: 2021-05-07; 录用日期: 2021-07-06

<http://www.photon.ac.cn>

0 引言

近年来,随着光学遥感图像技术的发展,光学遥感图像体现出了信息量大、分辨率高、直观可读、便于获取的独特优点,使得越来越多的工作开始关注于光学遥感图像目标检测。

传统的遥感目标检测算法利用人工设计的特征来预测目标,但是其难以适应复杂多样的数据^[1-4],而基于深度学习的方法使用卷积神经网络,可以自动提取图像深层次的特征,不仅解放了人力,且更具鲁棒性。因此,近年来不少研究将深度学习引入遥感目标检测领域。

深度学习在普通光学图像目标检测领域已经有着广泛而成熟的应用,主要有一阶段算法和两阶段算法两大类。一阶段算法从图像直接预测目标,速度较快但精度较低,代表算法是YOLO(You Only Look Once)系列^[5-7]和单发多盒检测算法(Single Shot MultiBox Detector, SSD)^[8]等。两阶段算法先在图像上产生可能包含目标的候选框,再对候选框进行修正,该类算法精度高但速度较慢,以区域卷积神经网络(Region-Convolution Neural Network, RCNN)系列^[9-11]为代表。根据对速度和精度的不同需求,学者们将两类深度学习算法推广至遥感目标检测领域,但是由于普通光学图像和光学遥感图像之间具有较大的差异,后者目标实例的尺度变化范围极大,部分目标如车辆,舰船等在图像中占比不到0.1%,部分目标如机场,高尔夫球场等占比可能超过20%,这种多尺度问题使得遥感目标检测极具挑战。目前解决多尺度问题最主要的方法是使用更深的网络获取多尺度特征图进行特征融合,例如WANG Yingying等^[12]以深度残差网络(Deep Residual Network, ResNet)^[13]为主干,构造了一个改进多尺度网络将浅层特征图中的图像细节与深层特征图中的语义信息相融合;YANG Xue等^[14]将中间层与深层网络输出层进行融合丰富特征信息;姚群力等^[15]设计了一种多尺度特征金字塔网络,增强了对多尺度特征的表达能力。上述算法虽然实现了检测精度的上升,但网络参数量大,计算成本高,检测速度被严重拖慢。

基于此,本文以基于VGG16主干网络^[16]的Faster R-CNN作为框架,提出一种基于参数量和感受野可调的遥感目标检测算法,即本文算法可以对参数量与感受野进行调节得到不同模型,在解决光学遥感图像中的多尺度目标检测问题的同时兼顾检测精度和速度。

1 模型设计

1.1 网络结构

Faster R-CNN^[11]是经典的两阶段算法,网络结构如图1,该算法使用VGG16模型的五个卷积块作为特征提取网络,conv5_3输出特征图,下采样率为16。区域建议网络(Region Proposal Network, RPN)对特征

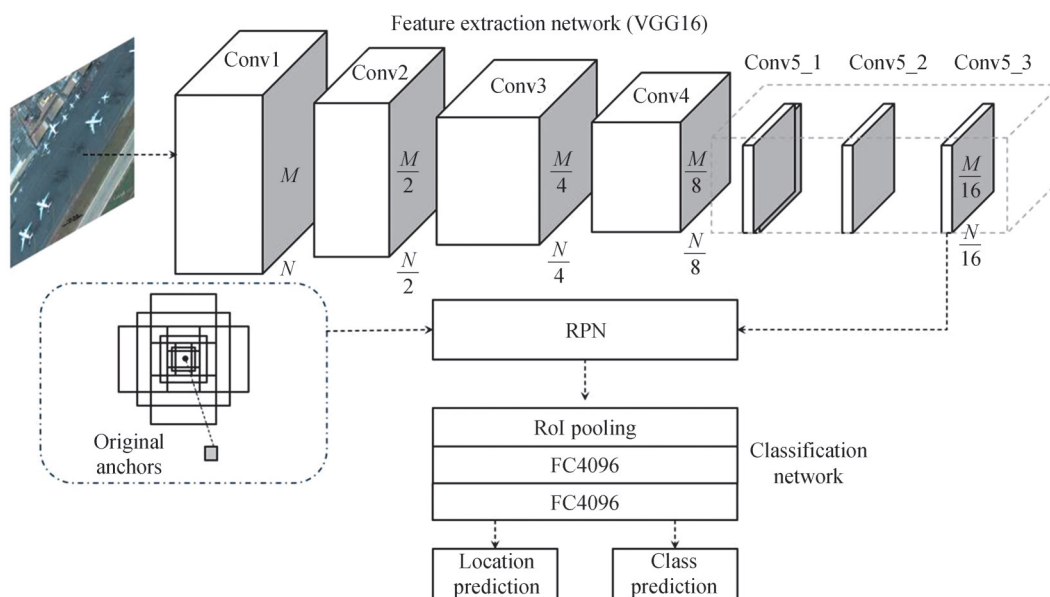


图1 Faster R-CNN 网络结构
Fig. 1 Faster R-CNN overall framework

图进行锚框分类和位置回归,得到感兴趣区域(Region of Interest, RoI),使用RoI Pooling层将RoI变换为同一尺寸,之后连接分类网络,其包括两个4 096全连接层,进行特征的线性组合和非线性变换,两个并行的全连接层,进行分类和位置回归。

Faster R-CNN算法虽然具有优秀的特征表达和分类能力,但是其特征图下采样率较大,导致小于 16×16 的小目标信息缺失,此外,其分类网络中的两个4 096全连接层最初是为1 000类别数的分类任务所设计,而遥感目标检测任务类别数一般在20类以下,这说明全连接层参数存在冗余,导致检测速度过慢。

为实现更好的检测效果,本文对Faster R-CNN进行改进,设计了一种基于参数数量和感受野可调的遥感目标检测算法(Adjustable Remote Sensing Target Detection Network, ASDN),网络结构如图2所示,主要改进点有三部分:1)基于空洞卷积设计了感受野可调模块(Receptor Field Adjustable Module, RFA),代替了特征提取网络的第五普通卷积块,同时去除了第四池化层,增大了感受野和特征图分辨率,提高了多尺度目标检测精度;2)在区域建议网络和分类网络之间增加了通道数可调模块(Channel Number Adjustable Module, CNA),可手动选择分类网络输入通道数大小,获得不同参数数量大小的模型,具有不同的速度和精度;3)调节全连接层(Adjust Full Connection layer, AFC),动态变化维度,以减少冗余参数。

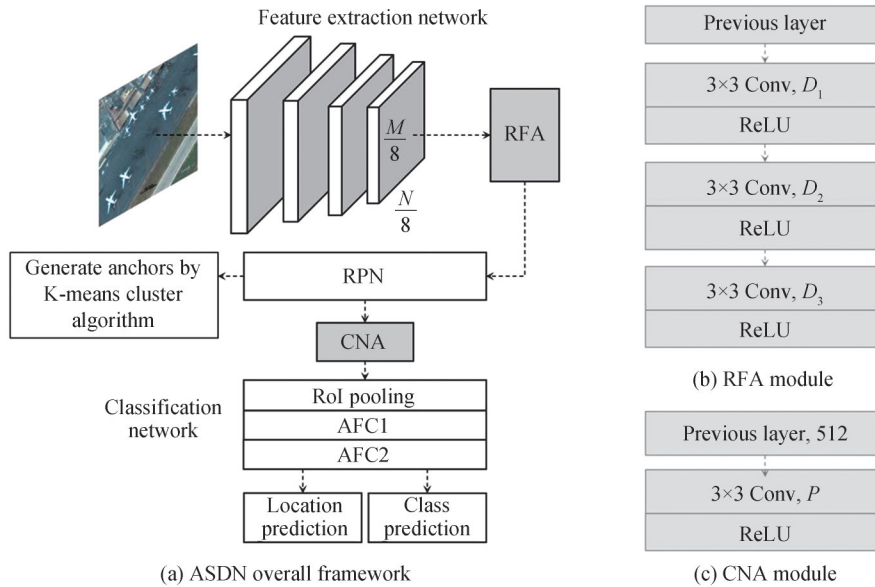


图2 ASDN网络结构
Fig.2 ASDN overall framework

1.2 感受野可调模块

ASDN算法去除了特征提取网络的第四池化层,将下采样率由16减少为8,保留了更多小目标特征信息,但这也导致特征图感受野由196减小为140,对大目标特征提取不利,在此引入空洞卷积思想。空洞卷积是在普通卷积核中加入空洞,这使得卷积操作略过了输入图上指定间隔的像素值,间隔大小称为空洞率。空洞卷积的优势在于扩大感受野的同时不影响输出层分辨率^[17]。

本文基于空洞卷积设计的感受野可调模块如图2(b),包括三个步长 S 为1的 3×3 卷积,通过设置不同的空洞率 D 组合,得到不同的感受野大小,感受野计算公式为

$$RF_i = (RF_{i-1} - 1) \times S + (D - 1) \times (K - 1) + K \quad (1)$$

式中, RF_i 表示第 i 层网络输出的特征层感受野的大小, $i=0$ 代表原始图片,此时 $RF=1$ 。 RF_{i-1} 表示第 i 层网络以 $i-1$ 层为输入时的感受野大小,一般以上一层为输入时感受野就等于卷积核大小。 S 表示第 i 层卷积的步长, D 表示空洞率,对于普通卷积,空洞率等于1, K 表示第 i 层卷积核的大小。

感受野并不是越大越好,其需要与锚框尺寸进行匹配,而ASDN算法使用了统计学中的K-means聚类算法^[18]获得更适合遥感数据集的锚框尺度,感受野可调模块根据聚类锚框的最大尺度来手动调节空洞率设置。调节原则有以下三点:1)为避免网格效应^[19],空洞率从正整数集合 $\{1, 2, 3, 4, \dots, n\}$ 中顺序选取,且彼此

不能有大于1的公约数;2)特征提取网络最终输出的特征图感受野应与聚类锚框最大尺度尽量接近;3)相邻卷积层空洞率不宜跨度过大,因此当不同设置感受野大小都接近聚类锚框最大尺度时,选择空洞率最大邻差最小的组合。

1.3 通道数可调模块

ASDN算法在区域建议网络和分类网络之间增加了通道数可调模块,如图2(c),使用一个 3×3 卷积,对特征图的通道数进行调节, 3×3 卷积输出通道数为通道数可调参数 P ,可自由设置,但为保证ASDN算法精度和速度不低于Faster R-CNN算法,通道数可调参数调节范围为 $[32, 512]$ 。通道数可调模块还使用了一个ReLU函数,用于增加模型非线性表达能力。

1.4 调节全连接层

Faster R-CNN算法在分类网络中使用了两个4 096全连接层,整个分类网络参数量达到了120M,几乎是模型其他部分参数量的10倍,严重拖慢了检测速度。ASDN算法重新设计了全连接层的维度,第一个全连接层其输入维度 ID_1 与通道数可调参数 P 对应,计算公式为

$$ID_1 = 7 \times 7 \times P \quad (2)$$

输出维度 OD_1 由目标类别数 N 和下采样率决定,计算公式为

$$OD_1 = \exp(\lceil \log_2[8 \times 8 \times (N + 1)] \rceil \times \lg 2) \quad (3)$$

第二个全连接层其输出维度 OD_2 一般由目标类别数 N 和通道数可调参数 P 共同决定,输出维度计算为

$$OD_2 = \text{Max} \left\{ \exp(\lceil \log_2[4 \times 4 \times (N + 1)] \rceil \times \lg 2), 4 \times P \right\} \quad (4)$$

2 实验环境

2.1 数据集

在公开的DIOR数据集^[20]上进行遥感目标检测实验。该数据集上共有20个遥感目标类别,包括:飞机(airplane)、机场(airport)、棒球场(baseball field)、篮球场(basketball court)、桥梁(bridge)、烟囱(chimney)、水坝(dam)、高速公路服务区(expressway service area)、高速公路收费站(expressway toll station)、高尔夫球场(golf field)、田径场(ground track field)、港口(harbor)、立交桥(overpass)、舰船(ship)、体育场(stadium)、储油罐(storage tank)、网球场(tennis court)、火车站(train station)、汽车(vehicle)和风车(wind mill)。DIOR数据集从Google Earth收集了23 463张图像,具有不同成像环境、天气、季节和清晰度,空间分辨率范围为0.5 m到30 m,并将所有图像都裁剪为800像素 \times 800像素,图像中一共含有192 472个实例,各实例数目如表1。该数据集图像中目标类别和实例数多,尺度变化范围大,具有类间相似性和类内多样性,整体数据集具有较大挑战性。本文实验将其分为训练集(11 725张图像)和测试集(11 738张图像)。

2.2 评价指标

目标检测中使用精确率(Precision)、召回率(Recall)、平均精度(AP)、全类平均精度(mAP)来判断检测

表1 DIOR数据集各类别实例数
Table 1 Number of instances of each category in DIOR

Category	Number	Category	Number
Airplane	10 104	Ground Track Field	3 038
Airport	1 327	Harbor	5 509
Baseball Field	5 817	Overpass	3 114
Basketball Court	3 225	Ship	62 400
Bridge	3 967	Stadium	1 268
Chimney	1 681	Storage Tank	26 414
Dam	1 049	Tennis Court	12 266
Expressway Service Area	2 165	Train Station	1 011
Expressway Toll Station	1 298	Vehicle	40 370
Golf Field	1 086	Wind Mill	5 363

结果。

由表2混淆矩阵可计算出各个评价指标。混淆矩阵中,TP表示将真正样本预测为正样本的数量,TN表示将真实负样本预测为负样本的数量,FP表示将真实负样本预测为正样本的数量,FN表示将真正样本预测为负样本的数量。

表2 混淆矩阵
Table 2 Confusion matrix

Actual	Predict	
	Positive	Negative
Positive	True Positive (TP)	False Negative (FN)
Negative	False Positive (FP)	True Negative (TN)

精确率表示预测正样本中真正样本的比例

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (5)$$

召回率表示真正样本被预测正确的比例

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (6)$$

每个预测样本经过分类后,都得到了一个置信度,表示该样本为正样本的概率,大于某一阈值的样本被预测为正样本,反之为负样本。置信度阈值不同,会使得混淆矩阵发生变化,得到不同的精确率和召回率,以精确率为纵轴,召回率为横轴,绘制出PR曲线辅助分析,平均精度AP即PR曲线与x轴围成的面积,AP越大,模型性能越好。而mAP是多个类别AP的平均值,用于衡量多分类检测器的性能。

此外,使用帧速度(Frames Per Second,FPS)用来评估目标检测的速度,即每秒内可以处理的图片数量。帧速度越大检测速度越快,当帧速度大于等于25时,可以认为算法满足实时性要求。

2.3 参数选择

使用python作为算法编程语言,使用Pytorch深度学习框架搭建网络,使用GeForce RTX 3 080平台进行实验。训练时批尺寸为4、初始学习率为 4×10^{-3} 、每经过8轮训练学习率衰减一个数量级,最大训练轮数为10,动量项Momentum设置为0.9,输入图像尺寸缩放为600像素 \times 600像素,并使用翻转进行图像增强。

本文算法对数据集进行K-means聚类,设置初始簇数目K值取值范围为[0, 25],相应的平均交并比结果如图3所示,可以看到当K值增加到一定数量后,曲线变换趋于平缓,平均交并比增加幅度小,考虑到锚框数量过大会增加计算成本,因此设置K值为9。

对于DIOR数据集聚类锚框尺寸为[9, 9]、[13, 19]、[18, 38]、[24, 11]、[33, 60]、[38, 22]、[67, 45]、[108, 135]、[306, 208]。

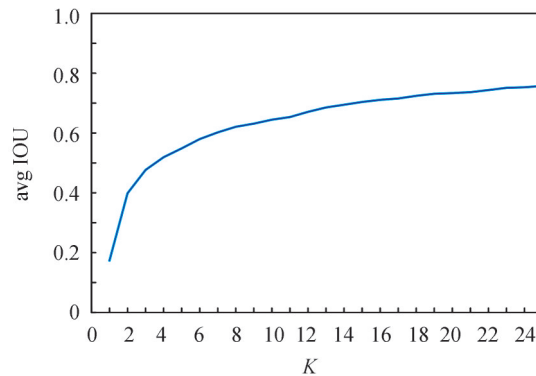


图3 平均交并比-聚类数曲线
Fig.3 Average IOU - cluster number curve

对于聚类锚框最大尺度[306, 208],由1.2节所述调节原则,空洞率应设置为(1, 3, 5),此时特征图感受野大小为236。为了保证准确,在对Faster R-CNN网络进行聚类和去池化后,使用不同空洞率组合进行对比训练,检测结果如表3,可以看到当空洞率设置为(1, 3, 5)时,算法获得最高精度,而感受野过大或过小都不利于精度提升,对于(1, 2, 5)组合,其感受野满足调节原则,但最大邻差大于(1, 3, 5)组合,精度也略低,证明调节原则可靠。

表3 不同空洞率组合的精度与感受野对比

Dilation ratios	(1, 2, 3)	(1, 2, 5)	(1, 3, 5)	(3, 5, 7)
mAP/%	67.5	67.8	68.1	67.4
Receptive field	188	220	236	332

3 结果与分析

3.1 主要结果

本文算法在DIOR数据集上进行训练,可得到通道数可调参数 P 变化时相应模型的全类平均精度(mAP)和帧速度(FPS),如图4所示,当 P 减少时,模型检测精度下降,检测速度提高,当 P 为512时算法具有最高精度0.681,且此时模型检测速度为20.4 FPS,要略高于原始Faster RCNN的检测速度(19.6 FPS),模型记为ASDN512。当 P 减小至256时,检测速度为25.6 FPS,已经满足实时要求,当 P 低于32时算法精度急剧下降,速度却提升微弱,因此 P 值最低取32,认为此时算法具有最高速度32.3 FPS,模型记为ASDN32。

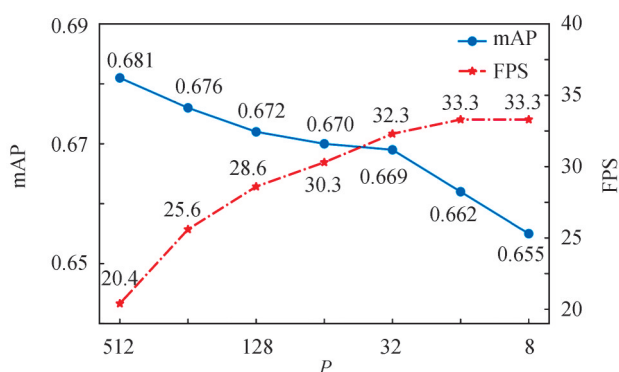
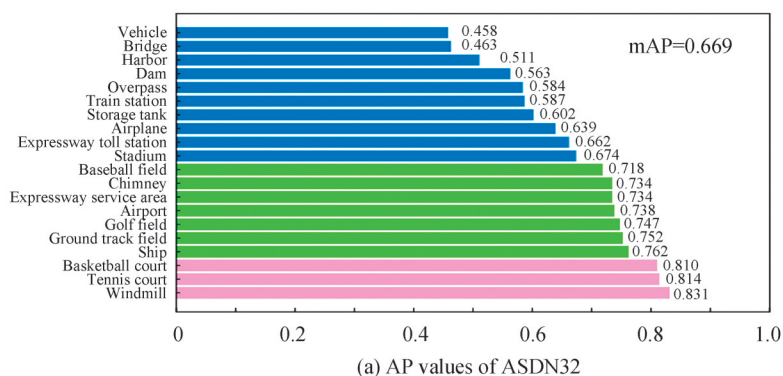


图4 模型检测精度和检测速度与 P 值的关系
Fig.4 Model mAP and FPS vs. P conditions

图5为模型ASDN32和模型ASDN512的各类的平均精度AP以及全类平均精度mAP,图6为各类检测结果可视化结果。可以看到模型ASDN32检测精度略低于ASDN512,这是由于前者参数较少,特征表达能力较弱。整体而言,模型ASDN32和模型ASDN512在DIOR数据集上均有十个类别AP超过0.7,包括目标实例较小的舰船,目标实例较大的高尔夫球场和高速公路服务区,其中AP超过0.8的三个类别为:篮球场、网球场以及风车,属于中等尺寸目标,说明二者均对多尺度遥感目标表现出较好的检测性能。



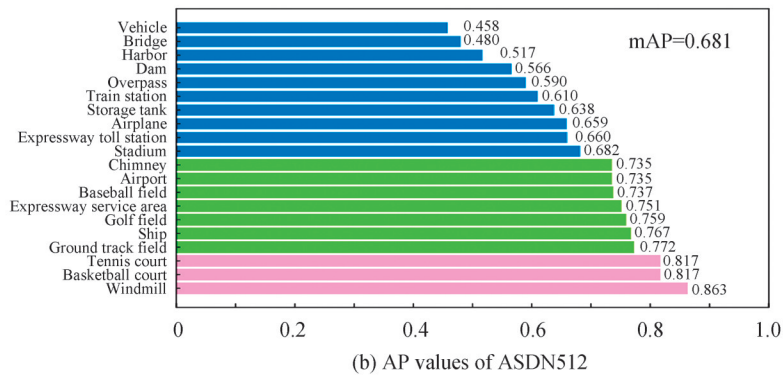
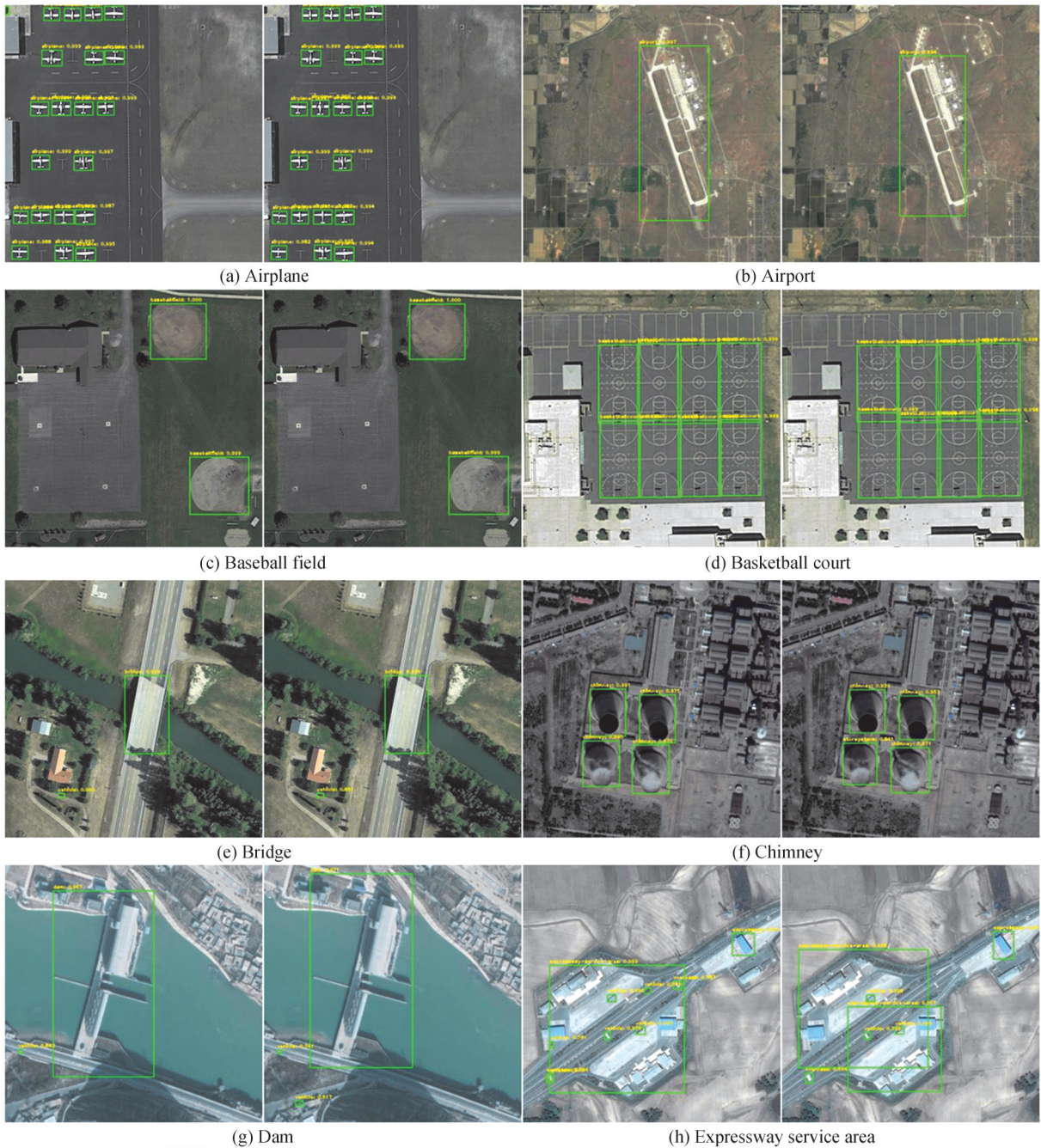
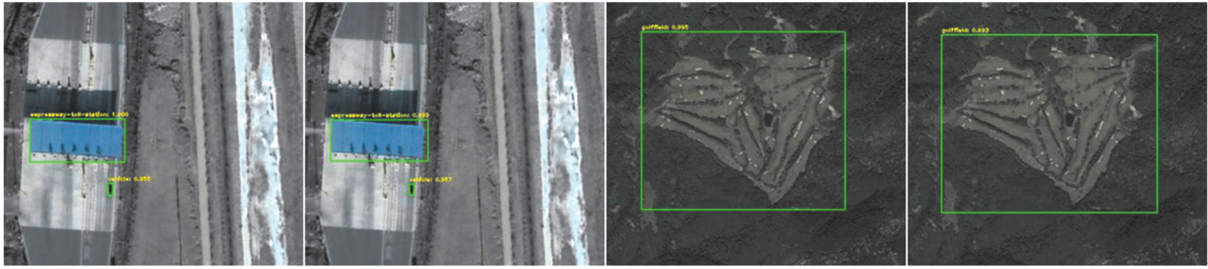


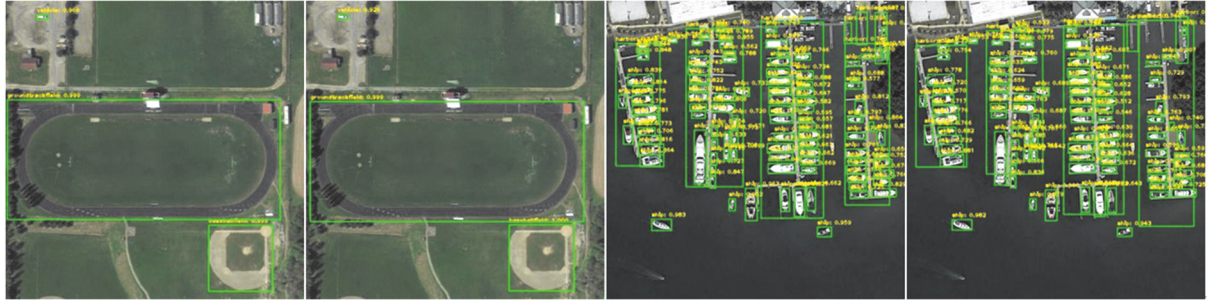
图5 在DIOR数据集上的测试精度
Fig.5 AP values for the DIOR dataset





(i) Expressway toll station

(j) Golf field



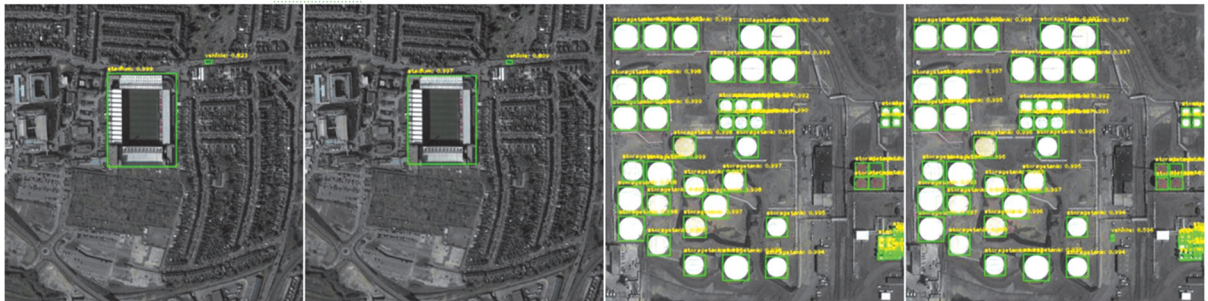
(k) Ground track field

(l) Harbor and small ship



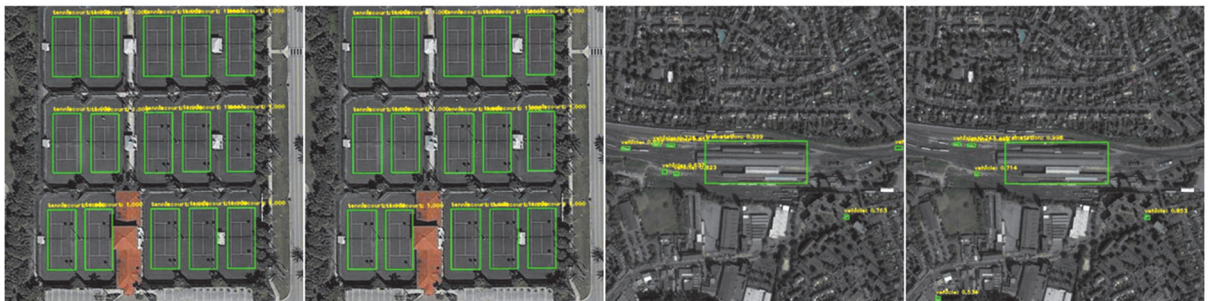
(m) Overpass

(n) Big ship



(o) Stadium

(p) Storage tank



(q) Tennis court

(r) Train station

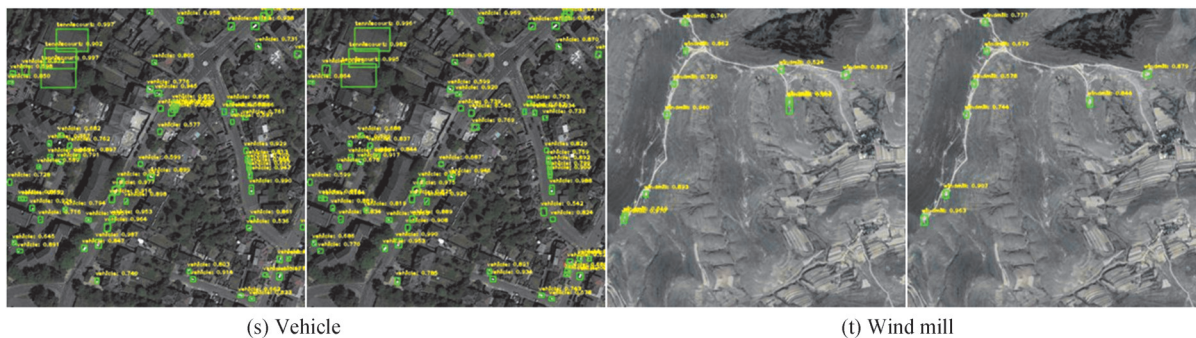


图6 基于ASDN的可视化结果,每一类别中左图为ASDN512,右图为ASDN32

Fig.6 Visual detection results of ASDN512. For each category, ASDN512 is shown on the left and ASDN32 on the right

ASDN算法在桥梁、港口以及车辆这三个类别上检测精度较低,第一是因为桥梁与立交桥较为相似,容易发生误检,第二是因为港口一般停泊有较多船只,背景复杂,检测较为困难,第三则是因为车辆有接近25%的实例尺寸小于 8×8 ,极其微小而导致漏检。图7为桥梁、港口以及车辆三个类别的PR曲线,可以看到三个类别检测召回率均过低,但其中车辆的精确率要高于另外二者,召回率低于另外二者,说明桥梁和港口主要问题在于特征更难表达,存在许多误检,而车辆的低精度主要源于实例过小所带来的漏检。

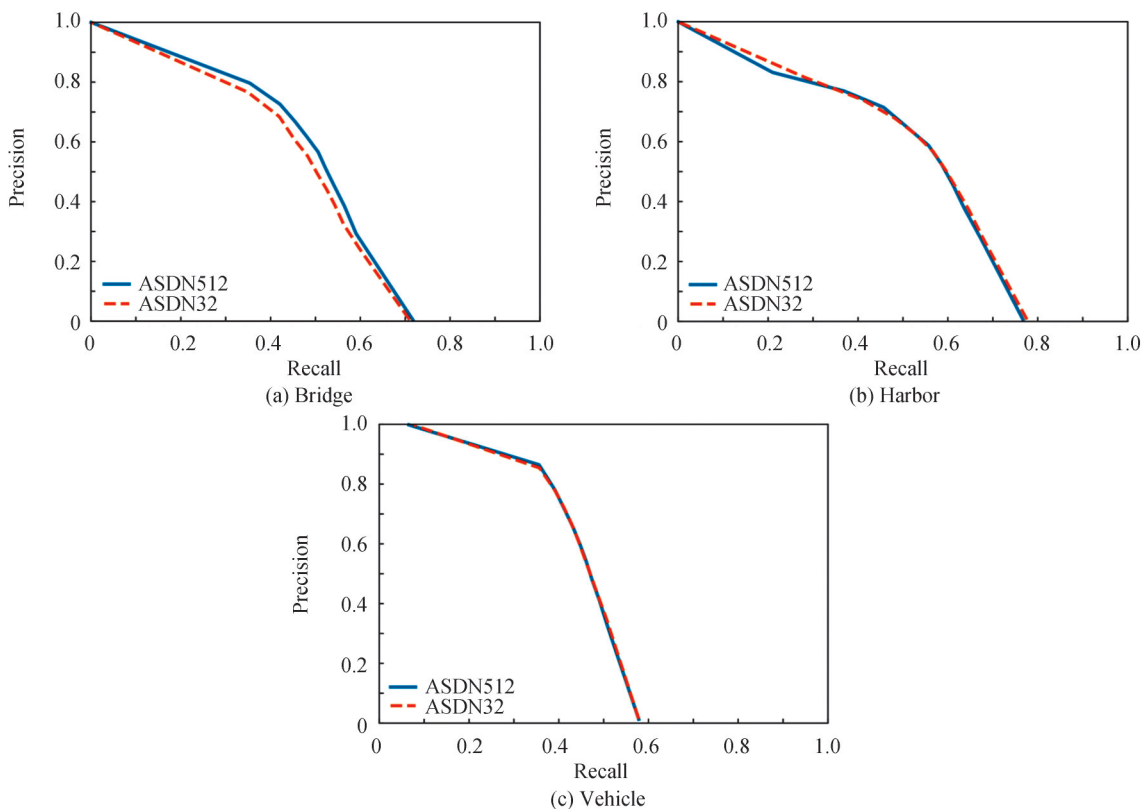


图7 PR曲线
Fig.7 PR curves

总的来说,ASDN算法对不同尺度的目标都具有良好的检测性能,通过调节 P 值可以在速度和精度之间做出权衡,满足不同检测任务的需求。

3.2 对比实验

本文算法在DIOR数据集上与当前一些主流遥感目标检测算法进行对比实验,结果如图5,红色表示精度排名第一,蓝色表示精度排名第二,绿色表示精度排名第三,类别与标签对应见表4,*号表示测试数据来源于文献[20]。从表5可以看出,ASDN512模型在所有对比算法中,具有最高精度,并且,在中小目标为主

的类别如飞机、舰船、汽车、风车上有较大提升。而 ASDN32 模型在高速一阶段算法中,具有最优性能。

表 4 DIOR 数据集目标类别
Table 4 Object classes in the DIOR dataset

C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Airplane	Airport	Baseball field	Basketball court	Bridge	Chimney	Dam	Expressway service area	Expressway toll station	Golf field
C11	C12	C13	C14	C15	C16	C17	C18	C19	C20
Ground track field	Harbor	Overpass	Ship	Stadium	Storage tank	Tennis court	Train station	Vehicle	Wind mill

表 5 不同算法在 DIOR 数据集上检测精度对比
Table 5 Comparison of detection accuracy of different algorithms on DIOR

Model	*RICNN ^[21]	Fater-RCNN ^[11]	*Fater-RCNN+FPN ^[22]	CBD-E ^[23]	ASDN512 (Ours)	*YOLO-V3 ^[7]	*SSD ^[8]	*CornerNet ^[24]	*Retinanet ^[25]	ASDN32 (Ours)
Backbone	VGG16	VGG16	ResNet-101	ResNet-101	VGG16	Darknet-53	VGG16	Hourglass-104	ResNet-101	VGG16
C1	39.1	61.5	54	54.2	65.9	72.2	59.5	58.8	60.2	63.9
C2	61	72.5	74.5	77	73.5	29.2	72.7	84.2	72	73.8
C3	60.1	68.3	63.3	71.5	73.7	74	72.4	72	70.6	71.8
C4	66.3	81.9	80.7	87.1	81.7	78.6	75.7	80.8	80.5	81
C5	25.3	45.1	44.8	44.6	48	31.2	29.7	46.4	43.6	46.3
C6	63.3	73.1	72.5	75.4	73.5	69.7	65.8	75.3	72.3	73.4
C7	41.1	57.6	60	63.5	56.6	26.9	56.6	64.3	61.4	56.3
C8	51.7	70.8	75.6	76.2	75.1	48.6	63.5	81.6	72.1	73.4
C9	36.6	63.9	62.3	65.3	66	54.4	53.1	76.3	66.7	66.2
C10	55.9	74.6	76	79.3	75.9	31.1	65.3	79.5	72	74.7
C11	58.9	79.8	76.8	79.5	77.2	61.1	68.6	79.5	73.4	75.2
C12	43.5	43.5	46.4	47.5	51.7	44.9	49.4	26.1	45.3	51.1
C13	39	59	57.2	59.3	59	49.7	48.1	60.6	56.9	58.4
C14	9.1	47.6	71.8	69.1	76.7	87.4	59.2	37.6	71.7	76.2
C15	61.1	67.9	68.3	69.7	68.2	70.6	61	70.7	70.4	67.4
C16	19.1	42.4	53.8	64.3	63.8	68.7	46.6	45.2	62	60.2
C17	63.5	81.1	81.1	84.5	81.7	87.3	76.3	84	80.9	81.4
C18	46.1	53.5	59.5	59.4	61	29.4	55.1	57.1	57	58.7
C19	11.4	34	43.1	44.7	45.8	48.3	27.4	43	47.2	45.8
C20	31.5	83.3	81.2	83.1	86.3	78.7	65.7	75.9	84.5	83.1
mAP/%	44.2	63.1	65.1	67.8	68.1	57.1	58.6	64.9	66.1	66.9

3.3 消融实验

本文对聚类(K)、去除池化层(-)、使用感受野可调模块RFA(+)、使用通道数可调模块CNA(通道参数为 P)和全连接层可调节AFC(C)的效果分别在DIOR数据集上进行消融实验,对比结果如表6。

从表6可以看出,Faster RCNN框架使用CNA模块后,减小参数 P 可令模型检测速度增加,精度降低,减少模型参数量。从表6第三行和第四行数据可知,AFC模块可以压缩冗余参数,进一步减少模型参数量,但对精度和检测速度影响较小。Faster RCNN框架使用聚类所得锚框替代默认锚框后,精度提升了5.4个百分点,这是因为聚类后锚框更加符合数据集目标尺度变化大的特点。从表6第六行和第七行数据可知,去除第四池化层和使用感受野可调模块均能给精度带来提升,且二者的结合对于多尺度目标能获得更好的性能。从表6最后两行可知,在ASDN512模型在检测速度比Faster RCNN算法略微提升的情况下,精度提升了8.6个百分点,模型参数量减小为后者的55%,ASDN32模型在达到实时检测速度32.3FPS时,精度仍比Faster RCNN高7.4个百分点,参数量小,仅为后者的16%,证明了各模块的有效性。

表6 消融实验
Table 6 Ablation experiments

Model	K	DP	RFA	CNA	AFC	mAP%	FPS	Parameters
Faster RCNN	×	×	×	×	×	59.5	19.6	136.85M
FRCNN32	×	×	×	✓	×	58.8	34.4	40.76M
FRCNN512	×	×	×	✓	×	59.4	19.2	139.31M
FRCNN512C	×	×	×	✓	✓	59.5	22.2	75.13M
FRCNN K	✓	×	×	×	×	64.9	19.6	136.85M
FRCNN K-	✓	✓	×	×	×	67.5	18.8	136.85M
FRCNNK -+	✓	✓	✓	×	×	68.1	18.1	136.85M
ASDN32	✓	✓	✓	✓	✓	66.9	32.3	21.44M
ASDN512	✓	✓	✓	✓	✓	68.1	20.4	75.13M

4 结论

本文以 Faster RCNN 算法为基础,针对遥感图像多尺度问题设计了 ASDN 算法,设计了感受野可调模块、通道数可调模并调节全连接层,以满足对模型速度和精度的不同需求。本文方法在 DIOR 遥感数据集上进行了实验,结果表明,ASDN512 模型具有最高精度,且高于所有对比算法,检测速度比 Faster R-CNN 略高,ASDN32 模型具有最高速度,满足实时性要求且精度超过一阶段算法,高于大部分两阶段算法。

参考文献

- [1] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005, 1: 886-893.
- [2] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.
- [3] RUBLEE E, RABAU D V, KONOLIGE K, et al. ORB: An efficient alternative to SIFT or SURF [C]. International Conference on Computer Vision, 2011: 2564-2571.
- [4] VIOLA P, JONES M J. Robust real-time face detection[J]. International Journal of Computer Vision, 2004, 57(2): 137-154.
- [5] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.
- [6] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7263-7271.
- [7] REDMON J, FARHADI A. Yolov3: An incremental improvement[J]. arXiv:1804.02767, 2018.
- [8] LIU W, ANGUÉLOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]. European Conference on Computer Vision. Springer, Cham, 2016: 21-37.
- [9] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.
- [10] GIRSHICK R. Fast R-CNN[C]. Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440-1448.
- [11] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [C]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [12] WANG Y, LI W, LI X, et al. Ship detection by modified RetinaNet [C]. 2018 10th IAPR Workshop on Pattern Recognition in Remote Sensing (PRRS), IEEE, 2018: 1-5.
- [13] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
- [14] YANG X, YANG J, YAN J, et al. Srdet: towards more robust detection for small, cluttered and rotated objects [C]. Proceedings of the IEEE International Conference on Computer Vision, 2019: 8232-8241.
- [15] YAO Qunli, HU Xian, LEI Hong. Object detection in remote sensing images using multiscale convolutional neural networks[J]. Acta Optica Sinica, 2019, 39(11): 346-353.
姚群力,胡显,雷宏.基于多尺度卷积神经网络的遥感目标检测研究[J].光学学报,2019,39(11):346-353.
- [16] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [J]. arXiv: 1409.1556, 2014.

- [17] YU F, KOLTUN V. Multi-scale context aggregation by dilated convolutions[J]. arXiv:1511.07122, 2015.
- [18] YANG Junchuang, ZHAO Chao. Survey on K- Means clustering algorithm[J]. Computer Engineering and Applications, 2019, 55(23): 7-14.
杨俊闯,赵超.K-Means聚类算法研究综述[J].计算机工程与应用,2019,55(23):7-14.
- [19] WANG P, CHEN P, YUAN Y, et al. Understanding convolution for semantic segmentation[C]. 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE, 2018: 1451-1460.
- [20] LI K, WAN G, CHENG G, et al. Object detection in optical remote sensing images: A survey and a new benchmark[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 159: 296-307.
- [21] CHENG G, ZHOU P, HAN J. Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(12): 7405-7415.
- [22] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117-2125.
- [23] ZHANG J, XIE C, XU X, et al. A contextual bidirectional enhancement method for remote sensing image object detection [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2020, 13: 4518-4531.
- [24] LAW H, DENG J. Cornernet: detecting objects as paired keypoints [C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 734-750.
- [25] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2980-2988.