

面向光学遥感目标的全局上下文检测模型设计

张瑞琰 姜秀杰 安军社 崔天舒

Design of global-contextual detection model for optical remote sensing targets

ZHANG Rui-yan, JIANG Xiu-jie, AN Jun-she, CUI Tian-shu

引用本文:

张瑞琰, 姜秀杰, 安军社, 崔天舒. 面向光学遥感目标的全局上下文检测模型设计[J]. *中国光学*, 2020, 13(6): 1302–1313. doi: 10.37188/CO.2020–0057

ZHANG Rui-yan, JIANG Xiu-jie, AN Jun-she, CUI Tian-shu. Design of global-contextual detection model for optical remote sensing targets[J]. *Chinese Optics*, 2020, 13(6): 1302–1313. doi: 10.37188/CO.2020–0057

在线阅读 View online: <https://doi.org/10.37188/CO.2020–0057>

您可能感兴趣的其他文章

Articles you may be interested in

基于卷积神经网络的候选区域优化算法

Region proposal optimization algorithm based on convolutional neural networks

中国光学. 2019, 12(6): 1348 <https://doi.org/10.3788/CO.20191206.1348>

高分辨率遥感图像SIFT和SURF算法匹配性能研究

Research on matching performance of SIFT and SURF algorithms for high resolution remote sensing image

中国光学. 2017, 10(3): 331 <https://doi.org/10.3788/CO.20171003.0331>

自适应上下文感知相关滤波跟踪

Adaptive context-aware correlation filter tracking

中国光学. 2019, 12(2): 265 <https://doi.org/10.3788/CO.20191202.0265>

复杂动背景下的“低小慢”目标检测技术

LSS-target detection in complex sky backgrounds

中国光学. 2019, 12(4): 853 <https://doi.org/10.3788/CO.20191204.0853>

基于改进的加速鲁棒特征的目标识别

Object detection based on improved speeded-up robust features

中国光学. 2017, 10(6): 719 <https://doi.org/10.3788/CO.20171006.0719>

自适应特征引流管故障智能识别方法

Intelligent identification algorithm of adaptive feature drainage tube fault

中国光学. 2017, 10(3): 340 <https://doi.org/10.3788/CO.20171003.0340>

文章编号 2095-1531(2020)06-1302-12

面向光学遥感目标的全局上下文检测模型设计

张瑞琰^{1,2}, 姜秀杰^{1*}, 安军社¹, 崔天舒^{1,2}

(1. 中国科学院国家空间科学中心 复杂航天系统电子信息技术重点实验室, 北京 100190;
2. 中国科学院大学, 北京 100049)

摘要:在复杂背景下的光学遥感图像目标检测中,为了提高检测精度,同时降低检测网络复杂度,提出了面向光学遥感目标的全局上下文检测模型。首先,采用结构简单的特征编码-特征解码网络进行特征提取。其次,为提高对多尺度目标的定位能力,采取全局上下文特征与目标中心点局部特征相结合的方式生成高分辨率热点图,并利用全局特征实现目标的预分类。最后,提出不同尺度的定位损失函数,用于增强模型的回归能力。实验结果表明:当使用主干网络 Root-Res-Net18 时,本文模型在公开遥感数据集 NWPU VHR-10 上的检测精度可达 97.6%AP⁵⁰ 和 83.4%AP⁷⁵,检测速度达 16 FPS,基本满足设计需求,实现了网络速度和精度的有效平衡,便于后续算法在移动设备端的移植和应用。

关键词:计算机视觉;目标检测;遥感图像;特征编码-特征解码;全局上下文特征

中图分类号:TP391.4 文献标志码:A doi:10.37188/CO.2020-0057

Design of global-contextual detection model for optical remote sensing targets

ZHANG Rui-yan^{1,2}, JIANG Xiu-jie^{1*}, AN Jun-she¹, CUI Tian-shu^{1,2}

(1. *Key Laboratory of Electronics and Information Technology for Space Systems, National Space Science Center, Chinese Academy of Sciences, Beijing 100190, China;*
2. *University of Chinese Academy of Sciences, Beijing 100049, China*)

* *Corresponding author, E-mail: jiangxj@nssc.ac.cn*

Abstract: To improve the detection accuracy and reduce the complexity of optical remote sensing of target images with a complex background, a global context detection model based on optical remote sensing of targets is proposed. First, a feature encoder-feature decoder network is used for feature extraction. Then, to improve the positioning ability of multi-scale targets, a method that combines global-contextual features and target center local features is used to generate high-resolution heat maps. The global features are used to achieve the pre-classification of targets. Finally, a positioning loss function at different scales is proposed to enhance

收稿日期:2020-04-07; 修订日期:2020-05-11

基金项目:中国科学院复杂航天系统电子信息技术重点实验室自主部署基金(No. Y42613A32S)

Supported by Laboratory Fund of Key Laboratory of Electronics and Information Technology for Space Systems, CAS (No. Y42613A32S)

the regression ability of the model. Experimental results show that the mean average precision of the proposed model reaches 97.6% AP⁵⁰ and 83.4% AP⁷⁵ on the NWPU VHR-10 public remote sensing data set, and the speed reaches 16 PFS. This design can achieve an effective balance between accuracy and speed. It facilitates subsequent porting and application of the algorithm on the mobile device side, which meets design requirements.

Key words: computer vision; object detection; remote sensing image; feature encoder-feature decoder; global-contextual feature

1 引言

随着科技的发展, 遥感卫星搭载的传感器分辨率逐渐提高, 遥感图像所包含的目标信息也日益丰富, 光学遥感图像检测在军事防备^[1]、民用生活^[2]等方面的利用价值越来越大。因此, 复杂背景下的光学遥感目标检测具有十分重要的意义。

光学遥感图像具有幅面大、多尺度等特点, 相比通用图像来说检测难度更大。以方向梯度直方图 (HOG)^[3], 尺度不变特征变换 (SIFT)^[4] 为代表的手工设计特征的方法表达能力不够, 检测精度不高。近年来, 卷积神经网络在遥感图像检测方面得到广泛应用。由于卷积神经网络的复杂度高及运算量大, 在实际部署网络时常采用以 SSD^[5]、YOLO^[6-7] 及 RetinaNet^[8] 为代表的高速单阶段检测方法。这些网络虽然能大幅度提高速度, 但是在光学遥感图像检测上仍存在若干问题, 例如生成大量候选框, 并采用非极大值抑制等限制速度的复杂操作, 并且在精度上低于两阶段检测网络。为了解决这些问题, 本文提出采用一种基于目标中心点的无预选框检测方式^[9-10], 并利用复杂度低的特征编码-特征解码网络来降低运算量^[11]。

在光学遥感目标检测中, 有效融合目标的局部特征和全局特征, 可以较大程度上提高网络的定位和回归能力。在双阶段网络中常采用融合局部上下文特征的方法, 如 Li K 等^[12] 提出 RICAOD 算法, 将局部上下文信息融入区域候选网络 (Region Proposal Network, RPN) 中, 以提高对候选框的筛选能力; Ma Wenping^[13] 等提出了一种基于上下文信息和多区域特征的多模型决策融合框架 MMDFN。在单阶段检测网络中常采用构造特征

金字塔的方式融合全局和局部的多尺度特征^[14], 如姚群力^[15] 等构造了一种深度特征金字塔 MSCNN, 以提取丰富的多尺度特征; 邓志鹏^[16] 等提出 MSDN 模型, 利用形变卷积和形变池化操作更准确地提取候选目标; 文献 [17] 基于 SSD 网络提出了融合注意力掩模的改进特征金字塔算法; 左俊皓^[18] 等则提出 MS-VANs 模型, 采用跳跃连接的编码-解码网络提取特征, 训练多尺度视觉注意力网络。

为了进一步提高特征编码-特征解码网络的定位和回归能力, 本文着重考虑全局上下文语义特征和局部特征在该网络中的融合。局部特征由中心点检测网络实现, 而全局特征则需添加相应辅助模块。全局特征可提供目标的背景信息, 常被用于分割、跟踪等多种视觉任务^[19-20]。如 Liu wei^[21] 等提出了 ParseNet, 添加全局特征后可大幅提高区域语义分割结果。Zhang Hang^[22] 等通过提取与特征图相关的上下文语义信息, 并结合扩张卷积和多尺度策略提升了语义分割性能。

综上, 本文提出了一种基于光学遥感目标的全局上下文检测模型 (Global-Contextual Detection Model, GCDN)。该模型采用特征编码-特征解码网络, 融合全局上下文特征辅助模块, 最后经分支检测模块输出预测热点图, 包含目标位置信息和目标框尺寸信息, 并采用定位热点损失、回归损失和偏移损失加权的方式训练网络。具体来说, 本文的创新点如下:

(1) 采用基于单阶段检测的特征编码-特征解码网络。在特征解码阶段引入全局上下文特征, 将其作为预分类, 并与经反卷积和变形卷积提取的局部特征相融合, 缓解目标检测中的误检漏检现象, 增强模型定位能力。

(2)针对多尺度检测问题,提出了基于不同尺度的定位损失函数,按目标框长宽比例确定目标各像素点对损失函数的贡献比例,增强模型回归能力。

(3)采用小型主干网络,并摒弃了推理阶段的非极大值抑制操作,有效减小了推理时间和部署难度,缓解了精度和速度的不平衡问题。

2 基于中心点的无预选框检测框架

2.1 特征编码-特征解码网络

本文基于特征编码-特征解码网络,预测包含目标中心点位置和长宽数值的热点图。图 1 为特征编码和特征解码网络基本结构。该结构较简单,无需中间层直通操作,特征编码模块为主干网络,采用较成熟的分类网络结构提取深层特征,特

征解码模块利用反卷积层输出高分辨率特征图。

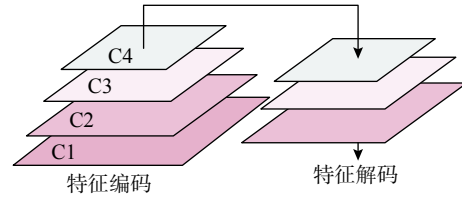


图 1 特征编码-特征解码网络架构

Fig. 1 Framework of the feature encoder-feature decoder network

2.2 检测模型总体架构

本文所提模型总体架构如图 2 所示,整体检测流程如下:首先,在检测网络中输入原始图像,依次经过特征编码、特征解码和分支检测网络,得到定位预测热点图、中心偏移预测热点图和回归预测热点图。三者尺度大小相同,输出结果与原始图像相比缩减了 4 倍。

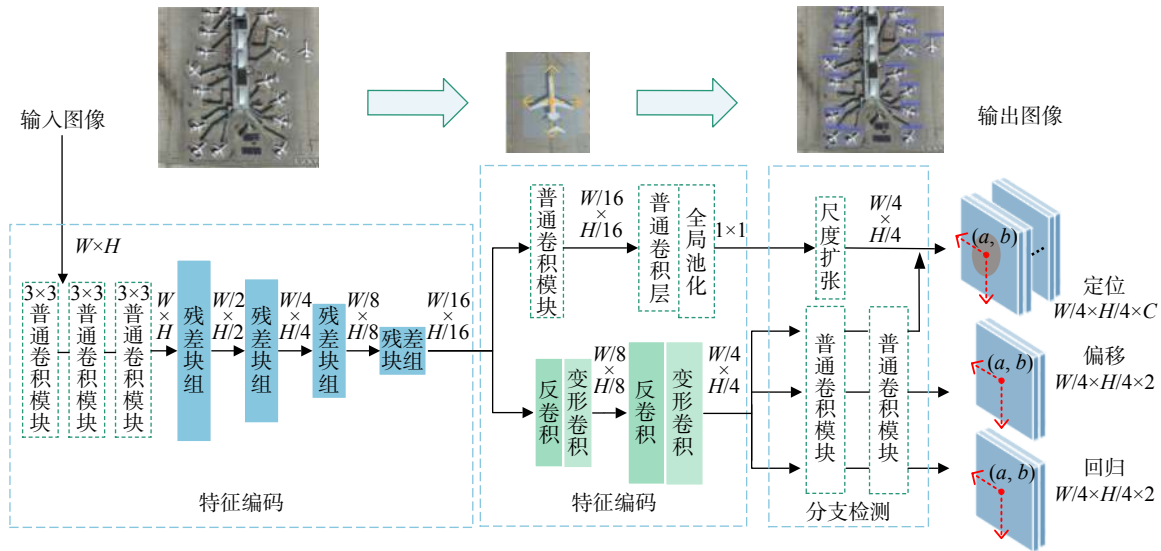


图 2 全局上下文检测模型总体架构

Fig. 2 Overall framework of the global-contextual detection model

之后,将定位热点图的各像素值排序,可得到前 K 个像素点的位置 (K 为最大检测目标数),由此确定检测目标的位置。具体操作如下:图 2 中,若检测到坐标为 (a,b) 的点为物体中心点时,根据该坐标可在另外两幅热点图的对应坐标 (a,b) 处得到像素中心偏移结果和回归结果。最终输出带预测框的输出图像。

其中特征解码阶段由两路并行处理,一路利用反卷积层和变形卷积层,增强深层特征在上采样时对感兴趣区域的建模能力,而后接入分支检测网络;另一路则采用普通卷积层和全局平均池化层提取各类全局语义信息。两路结果融合后便可得出预测值。

3 全局上下文检测算法

3.1 特征编码网络

本文采用含直连层的残差卷积网络 (Residual Network, ResNet)^[23] 来提取深层特征。考虑到实际移动端对检测速度的要求,以及主机显卡性能的影响,本文采用 18 层的残差网络。因层数较浅,为保证在残差网络中提取到足够有用的特征,需对 ResNet18 进行特征扩充。ZHU R 等人^[24] 提出一种可从头训练的 Root-ResNet18 网络。为挖掘更多的局部特征,该网络减少一次下采样,并利用 3 个相连的 3×3 卷积层代替原始 ResNet 的第一层 7×7 卷积层。其结构与原始 ResNet 网络对比如表 1 所示,原始图像大小设为 512×512。对检测网络而言,此举虽增加一定量的网络参数,但可以为小目标检测提供更丰富有效的特征。本文将该网络作为特征编码网络。

表 1 ResNet18 与 Root-ResNet18 结构

Tab. 1 Structures of ResNet18 and Root-ResNet18

阶段	输出尺寸	ResNet18	Root-ResNet18
			3×3, 64
			3×3, 64
C1	128×128	7×7, 64	3×3, 64
			3×3, MaxPool
			$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$
C2	64×64	$\begin{bmatrix} 3 \times 3, \text{MaxPool} \\ 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$
C3	32×32	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$
C4	16×16	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$
C5	8×8	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	—

3.2 全局上下文特征解码网络

本文在特征解码模块分两路对特征编码模块提取的信息进行融合,一路添加反卷积层和变形卷积层提取局部细粒度特征,另一路添加普通卷积层和全局平均池化层提取全局粗粒度特征。

第一路中的反卷积模块包含反卷积层、批量归一化层和 ReLu 激活函数层,反卷积层的步长为 2,采用填充操作扩大尺寸。同时考虑到所有卷积层为规则化的 3×3 卷积核操作,对任意形状的目标适应性较差,变形卷积通过给规则卷积的每个采样点添加偏移量和幅值信息,可极大增强对所提取特征的空间整合和建模能力。普通卷积和变形卷积的采样示意图如图 3(彩图见期刊电子版)所示,图中采样点颜色深浅代表该点在偏移和调制增量的占比不同。

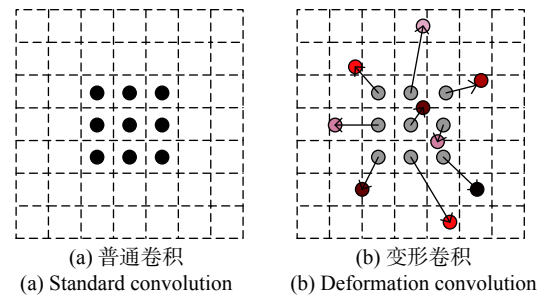


图 3 普通卷积采样和变形卷积采样示意图

Fig. 3 Sampling diagrams in standard convolution and deformation convolution

设 P_i 为该模块输出结果特征图 Y 上的一点, X 为输入特征图,其变形卷积操作定义如下:

$$Y(P_i) = \sum_k^K w_k \cdot X(P_i + P_k + \Delta P_k) \cdot \Delta m_k, \quad (1)$$

其中 K 为感受野区域, $k \in K$, w_k 为普通卷积中可学习的权重值, P_k 代表 K 中的元素, ΔP_k 和 Δm_k 代表变形卷积中可学习的偏移量和调制量。

第二路从全局的角度估计整张图片的场景内容,对目标的分类进行概率预判,该网络详细结构如图 4 所示。首先通过 n 个普通卷积模块进一步提取信息,卷积核为 3×3,将输入特征图维度降低至 64,这里网络较小, n 取 1 即可。其次,通过卷积核为 3×3 的卷积层和批量归一化层,将输出特征图的维度降低,使之等于目标类别数,这里不使用激活函数。然后,经过全局平均池化层^[25](Global Average Pooling, GAP) 和 Sigmoid 激活函数得出每类目标的预判概率向量,以便后续融合,该过程定义为:

$$q(k) = \text{Sigmoid} \left[\frac{1}{WH} \sum_i^w \sum_j^h X(P_{i,j,k}) \right], \quad (2)$$

其中, X 为输入特征图, $P \in X$ 。 i, j, k 分别代表输入特征图中的宽、高和维度。 $q(k)$ 为第 k 维度的输出值, 作为预判全局系数。 以强调输入图片中某类目标出现的概率, 且相对抑制其他类别。 两路采用式 (3) 进行融合, 得到分类热点图。

$$\hat{Y}_{\text{cls}}(P_{i,j,k}) = Y_{\text{cls}}(P_{i,j,k}) \cdot [1 + v \cdot q(k)], \quad (3)$$

其中, Y_{cls} 为第一路经分支检测网络后的定位预测热点图, 最终定位预测输出为 \hat{Y}_{cls} , v 为设定的超参数, 以调整预判系数对整体检测结果的影响。

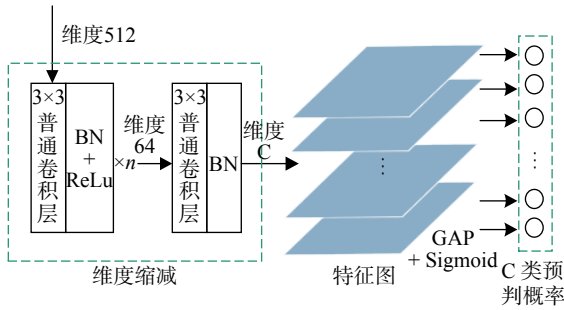


图 4 全局上下文特征提取流程

Fig. 4 Flow chart of global-contextual feature extraction

3.3 损失函数

本文所提模型将每个像素点视为一个目标, 精细化检测过程, 通过高分辨率预测图与软标签对比来得出最终结果, 因此, 更适用于中小型目标的检测。 该模型目标定位的关键点在于如何设置软标签, 即定位真实热点图、中心偏移真实热点图和回归真实热点图。

3.3.1 定位损失函数

设置定位真实热点图时, 将目标中心点设为正样本, 并在中心点处添加高斯掩模, 以指引网络收敛方向。 该掩模值设置为 0 至 1 的过渡值。 图 5(a) 为添加目标框的原图。

为更好地利用真实目标框信息, 本文将高斯椭圆半径值与目标框的宽高相关联, 区别于文献 [9,10] 中使用的二次方程求解半径的方法。 在某个类别上的高斯椭圆掩模定义为:

$$R_{i,j} = \frac{(i-i_0)^2}{2(\alpha/8)^2} + \frac{(j-j_0)^2}{2(\beta/8)^2}, \quad (4)$$

$$\alpha = \gamma w, \beta = \gamma h, \quad (5)$$

其中, (i_0, j_0) 为高斯椭圆的圆心, 也是正样本的位置所在, α, β 分别设置为目标框宽 w 和高的 γ 倍, $\gamma \in (0,1)$, γ 的具体取值见 4.4 节。 鉴于不同目标的正样本落在同一个像素点的概率较小, 相比中心点的重叠问题, 本文更关注于掩模的重叠现象, 这里取落在同一像素的掩模的最大值参与训练。

掩模对负样本的调节有较大作用。 如图 5(b) 所示, 颜色由暗至亮代表掩模值对损失函数中负样本的作用由重到轻, 最亮处代表目标的中心点。

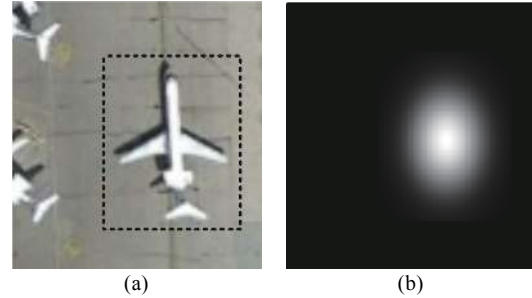


图 5 (a) 添加目标框的原图及 (b) 高斯椭圆掩模示意图

Fig. 5 (a) Original image with a target box and (b) schematic diagram of gaussian elliptical mask

本文的定位损失函数基于 Focal loss 做适应性调整, 如式 (6) 所示, 其中 ρ 沿用 Focal loss 中的值, 设为 2。

$$L_{\text{cls}} = -\frac{1}{N} \sum_{ijk} \begin{cases} [1 - \hat{Y}(P_{i,j,k})]^\rho \log(\hat{Y}(P_{i,j,k})) & \text{if } P_{i,j,k} \text{ 为正样本} \\ G_{i,j,k} \cdot [\hat{Y}(P_{i,j,k})]^\rho \log(1 - \hat{Y}(P_{i,j,k})) & \text{otherwise} \end{cases}, \quad (6)$$

其中, $P_{i,j,k}$ 为预测定位热点图上的一点, i, j 和 k 为三维坐标, $G_{i,j,k}$ 为所有类别上的定位损失掩模系数, 定义为:

$$G_{i,j,k} = \{1 - \exp[-R_{i,j,k}]\}^\omega, \quad (7)$$

其中, R 代表施加在目标中心点周围的掩模值。 参照文献 [8,10], 设置 ω 值设为 4。 当该点为目标框的中心点即正样本时, 采用式 (6) 上面的公式计算损失函数; 当该点不为中心点时, 采用式 (6) 下面的公式计算。 N 为图像中所有类别的正样本数之和。

3.3.2 中心偏移损失函数及回归损失函数

这里图像下采样比例为 4, 预测定位热点图的每一点均代表原图像中 4×4 范围的点, 因此, 需

要采用中心偏移预测来提高定位能力。偏移量真实值为目标框中心点坐标除4后的小数部分,而回归量真实值为目标框缩减4倍后的宽和长。二者真实热点图极其稀疏,只有正样本处有值,其余部分均为零。这里直接采用L1损失函数进行训练,定义如式(8)所示。其中输出预测热点图为 B ,真实热点图为 \bar{B} , N 为输入图像的各类别正样本个数。

$$L_{\text{off/box}} = \frac{1}{N} \sum_n |B_n - \bar{B}_n|. \quad (8)$$

3.3.3 混合损失函数

训练该网络时采用混合损失函数,对定位、回归和偏移损失函数做加权处理。该混合损失函数定义为:

$$L = L_{\text{cls}} + \lambda_{\text{off}} L_{\text{off}} + \lambda_{\text{box}} L_{\text{box}}, \quad (9)$$

其中, L_{cls} 代表定位分类损失, L_{off} 为中心点偏移损失, L_{box} 为预测框回归损失, λ 为调节系数。考虑到网络收敛需先大范围确定位置,再微调调节目标边框大小,因此,定位分类损失占较大比重,与文献[10]保持一致,取 λ_{off} 为1, λ_{box} 为0.1。

4 实验结果与分析

4.1 数据集简介及评价指标

为验证本文提出方法的有效性,采用复杂背景航天遥感十分类目标检测公开数据集NWPU VHR-10^[26]和大规模十五分类数据集DOTAv1.0^[27]进行验证。

NWPU VHR-10包含有标注信息的图像650张,背景信息图像150张。目标共有10类,分别为:飞机、舰船、油罐、棒球场、网球场、篮球场、田径场、港口、桥梁和车辆。表2对该数据集的各类目标框尺寸进行了统计,结果显示目标尺度大部分在300像素以下,最大不超过500像素。

DOTAv1.0数据集包含2806张标注信息的光学遥感图像。共15个类别,分别为:飞机、棒球场、桥、船只、田径场、小型车辆、大型车辆、船只、网球场、篮球场、储蓄罐、足球场、环形路线、海港、游泳池和直升飞机。其特点是尺度变化性更大,检测难度更高。

表2 数据集NWPU VHR-10目标尺寸统计表

Tab.2 Statistics of target sizes in the NWPU VHR-10 dataset

尺度(pixel)	0-10	10-40	40-100	100-300	300-500	500以上
宽	0	0.1327	0.6948	0.1599	0.0126	0
高	0	0.1448	0.7205	0.1242	0.0105	0

本文采用平均精确度(mean Average Precision, mAP)作为评估网络检测能力的标准。平均精确度代表所有类别模型检测精度(AP)的平均值,而AP由精确度和召回率构成的折线图的积分值确定。

精确度是指被预测为正样本中真实正样本所占比例,召回率是指原本正样本中依然被预测为正样本的比例。根据预测结果可得到 T_P (真阳性), F_P (假阳性), F_N (假阴性)和 T_N (真阴性)。从而得出检测精确度(p, precision)和召回率(r, recall)。

$$p = \frac{T_P}{T_P + F_P}, r = \frac{T_P}{T_P + F_N}. \quad (10)$$

本文通过判断预测框(P_R , Prediction Result)与真实框(G_T , Ground Truth)的交并比(I_{ou})来判断是否成功检测到物体,如式(11)所示。本文中主要评估交并比阈值为0.5和0.75两种情况,预测框与目标框的交并比在阈值之上即认为检测到有效目标。

$$I_{\text{ou}} = \frac{P_R \cap G_T}{P_R \cup G_T}. \quad (11)$$

为统计不同尺度对象的检测精度,与文献[13]保持一致,设置小目标(small)的尺寸为目标框面积小于 60^2 ,中目标(medium)为目标框面积大于小目标同时小于 120^2 ,大目标(large)则为目标框面积大于 120^2 。后文将对各尺度目标统计检测结果。

4.2 实验基本设置

由于遥感图像尺寸较大,需要对数据集进行裁剪处理。根据十分类数据集目标的宽高统计信息,长和宽最大不超过500像素,因此设定裁剪图片尺寸至少满足 500×500 ,故最终将图片裁剪为 640×640 ,并保证两张图片有140个像素的重叠区

域。可以保证完整覆盖最大尺寸目标。在裁剪时,检测物体中心点是否在所得图像内,若包含中心点,则保留目标框并调整标注的长宽;若不包含中心点,则抛弃该目标框。

对 NWPU VHR-10 数据集来说,本文采用数据集中含标注信息的 650 张图片进行训练和测试。处理后共获得 1743 张图片,取其中的 60% 作为训练集(1045 张),20% 作为验证集(349 张),剩余 20% 为测试集(349 张)。对 DOTA v1.0 数据集来说,对其进行同样操作,共得到 33892 张图片,随机取数据集的 1/2 为训练集(16946 张),1/6 为验证集(5649 张),1/3 为测试集(11297 张)。同时对图片进行数据增强操作,包括随机左右翻转,上下翻转及比例放缩等。

网络训练的优化器为 Adam,初始学习率设置为 1.25×10^{-4} ,在 NWPU VHR-10 数据集上训练次数为 280,并在 160 和 240 次后分别将学习率降低 10 倍,在 DOTA 数据集上训练次数为 140 次,并在 70 和 120 次后各将学习率降低 10 倍。批处理大小设置为 6 张图片。图像预处理阶段对输入图像进行统一缩放操作,进入特征编码网络的图像尺寸为 512×512 。本文训练及测试硬件均为 GPU GTX 1080 8G, CPU 为 i7-7700K 4.20GHz,算法平台采用 Pytorch 架构。

4.3 NWPU VHR-10 数据集检测结果

为凸显本模型的有效性,对比了 6 种基于深度学习的有代表性的目标检测模型。

(1)RICAOD 模型:基于两阶段检测方法,在

区域候选网络上融合了遥感图像的局部上下文特征,采用 ZF-Net^[28] 网络架构,其主干网络为 AlexNet。

(2)SSD 模型:作为对比的经典有效算法,采用 VGG16 网络提取特征,最后在不同尺度的特征图上得到多尺度融合预测结果。

(3)YOLOv3 模型:针对小目标检测,采用 Darknet53 网络结构,添加直连层和上采样操作实现不同层的特征图融合,在 3 个不同尺度的特征图上进行预测。

(4)MMDFN 模型:基于两阶段的检测方法,以 VGG16 分类网络为特征提取主网络,在 RPN 网络中构建多区域融合的子网络,也为基于局部上下文检测的方法。

(5)MSDN 模型:在特征提取网络 ResNet50 的最后一组残差块中采用可变形卷积,以获取多尺度形变卷积特征,在 3 个特征图上进行预测。

(6)MSCNN 模型:采用 RetinaNet 为基线架构,主干网络为 ResNet50,构建了一种新的特征金字塔方式以融合特征。

这里令 I_{ou} 阈值为 0.5,各总体实验结果如表 3 所示。其中 SSD 采用 mmdetection^[29] 检测框架训练,YOLO v3 采用 darknet 框架。RICAOD 模型和 MMDFN 模型均为两阶段检测网络中融合局部上下文特征,MSDN 模型和 MSCNN 模型均为融合多尺度特征提取方法的单阶段检测模型,而本文的 GCDN 则是基于单阶段检测中融合全局上下文特征检测和变形卷积的模型。

表 3 不同模型在数据集 NWPU VHR-10 上的平均精确度对比

Tab. 3 Comparison of mean average precisions of different models in the NWPU VHR-10 dataset

模型	主干网络	飞机	船舰	油罐	棒球场	网球场	篮球场	田径场	港口	桥梁	车辆	平均精确度(mAP)
RICAOD	AlexNet	0.9970	0.9080	0.9061	0.9291	0.9029	0.8031	0.9081	0.8029	0.6853	0.8714	0.8712
SSD	VGG16	0.9839	0.8993	0.8918	0.9851	0.8791	0.8481	0.9949	0.7730	0.7828	0.8739	0.8912
YOLOv3	Darknet53	0.9091	0.9091	0.9081	0.9913	0.9086	0.9091	0.9947	0.9005	0.9091	0.9035	0.9243
MMDFN	VGG16	0.9934	0.9227	0.9918	0.9668	0.9632	0.9756	1.0000	0.9740	0.8027	0.9136	0.9504
MSDN	ResNet50	0.9976	0.9721	0.8383	0.9909	0.9734	0.9991	0.9868	0.9719	0.9267	0.9010	0.9558
MSCNN	ResNet50	0.9940	0.9530	0.9180	0.9630	0.9540	0.9670	0.9930	0.9550	0.9720	0.9330	0.9600
GCDN	Root-ResNet18	0.9991	0.9983	0.9677	0.9916	0.9991	0.9759	0.9988	0.9412	0.9224	0.9636	0.9757

通过实验可以看出,MMDFN结构较为复杂,需要多个子系统的特征融合,在检测复杂度和计算资源利用率方面与单阶段检测相比不占优势。YOLOv3借鉴了SSD的多尺度特征图提取特征,采用多尺度预测和多宽高比预测框两种策略,可提高预测精度。但遥感目标尺寸差异较大且分布不均衡,针对单一像素点将生成大量无目标的预选框。本文的GCDN采用不生成预选框的方式,主干网络为Root-ResNet18,参数量较VGG16小且层数比ResNet50浅,结构更简单。

通过分析实验结果可以看出,上下文特征与变形卷积对多尺度检测效果的提高十分有效。MMDFN对于油罐、田径场和港口的检测精度有较大提高,体现了利用上下文语义特征的优势。MSDN模型的平均检测精度达95.6%,证明了变形卷积在特征提取网络的有效性,尤其是对海港的检测精度较高。本文所提GCDN模型融合了变形卷积和全局上下文特征,进一步提高了十分类遥感目标的检测精度,最终精度达到对比实验中的最高值97.57%,相比当前最好的方法MSCNN,提高了1.57%。

4.4 对比实验

4.4.1 不同高斯椭圆掩模半径对比

为了更好地确定高斯椭圆的掩模半径 γ 对于检测结果的影响,本文做了4组对比实验,分别取 γ 为0.25,0.45,0.75和0.95。实验结果如图6所示。通过实验可以看出 γ 的大小对结果有一定的影响, γ 较小时,对目标物体的覆盖率小,有利于网络的快速收敛,但会导致网络回归性能降低。而 γ 较大时,位于目标框内的负样本数目较多,网络计算得到的正样本的浮动范围较大,同时多个目标框聚集时,掩模的重合概率增大,不利于网络的学习和收敛。因此,本文最终取 γ 为0.45。

4.4.2 不同预判系数值对比

由式(3)可得预判系数 ν 代表全局上下文特征在定位预测结果中所占比例大小。因为预测主要贡献者为基于目标中心点的局部特征,因此 ν 的取值范围在(0,0.5)之间。为确定出比较合适的预测系数 ν ,本文设计了若干对照值以比较检

测效果。如图7所示,预测系数在[0.05,0.2]之间有较好的效果,过大会影响局部特征的表达能力,过小则不会对检测网络有提升作用。因此本文最终取 ν 为0.1。

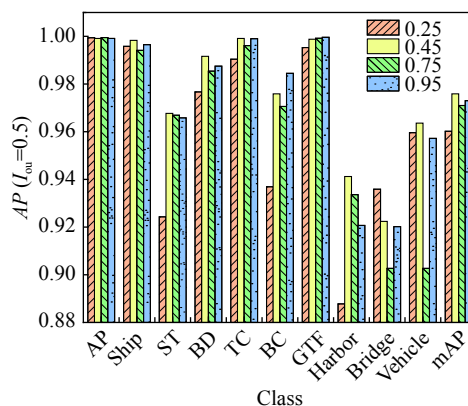


图6 不同 γ 值对结果的影响

Fig. 6 Effects of different γ values on results

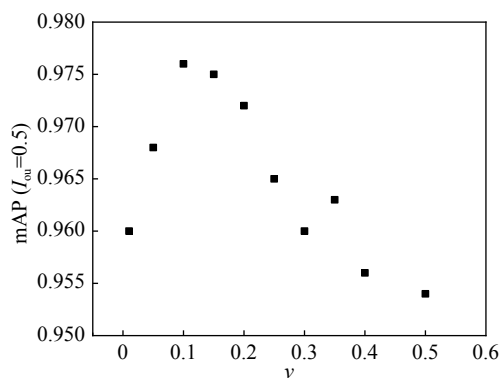


图7 不同 ν 值对结果的影响

Fig. 7 Effects of different ν values on results

4.4.3 不同交并比阈值对比

对于检测精度来说,AP⁵⁰可有效评估网络的分类定位能力,而AP⁷⁵则更能反映网络的回归效果,以反映预测框与目标框的差距。因此,本文将 I_{ou} 阈值提高至0.75,证明了本文所提方法可提高检测网络的回归能力。本文设置4个对比实验,分别为基线无预选框检测网络CenterNet,预选框检测网络MSCNN,不加全局上下文特征辅助模块的GCDN-woGC和添加全局上下文特征辅助模块的GCDN,实验结果如表4所示。

综合表4结果可以看出,CenterNet和GCDN-woGC的AP⁷⁵均低于MSCNN网络,说明中心点

检测方法在回归问题上略逊色于预选框检测网络。而本文所提模型可在一定程度上弥补该不

足, GCDN 相对 MSCNN 提高了 1.0% AP⁷⁵, 最终可达到与预选框检测相当甚至更优的回归能力。

表 4 检测阈值 0.75 下的不同模型平均精确度对比

Tab. 4 Comparison of mean-average precision of different models under the 0.75 detection threshold

模型	平均准确率 (AP)			平均准确率(AP)($I_{ou}=0.50:0.95$)			平均召回率(AR)($I_{ou}=0.50:0.95$)		
	$I_{ou}=0.50:0.95$	$I_{ou}=0.50$	$I_{ou}=0.75$	小目标	中目标	大目标	小目标	中目标	大目标
CenterNet	0.663	0.968	0.768	0.576	0.669	0.704	0.630	0.725	0.741
MSCNN	0.706	0.960	0.824	0.547	0.578	0.701	0.600	0.605	0.700
GCDN-woGC	0.679	0.973	0.775	0.572	0.690	0.750	0.639	0.744	0.786
GCDN	0.705	0.976	0.834	0.612	0.727	0.706	0.683	0.773	0.740

4.4.4 检测时间对比

为确定本文所提模型的实时性和有效性, 本文计算了测试集所有图片的平均检测时长, 该时长为图像预处理、网络检测和图像后处理时间之和, 并与 RICAOD 模型、MMDFN 模型、YOLOv3 模型和 MSDN 模型进行对比, 结果如表 5 所示。

表 5 不同模型的平均检测时间对比

Tab. 5 Comparison of the average detection times with different models

模型	输入尺寸	时间/s
RICAOD	400×400	2.89
MMDFN	400×400	0.75
YOLOv3	640×640	0.13
MSDN	600×800	0.21
GCDN	640×640	0.06

由表 5 可以看出, 本方法用时较短, 平均检测速度可达到 16FPS, 达到实际应用需求。较双阶段检测 RICAOD 与 MMDFN 来说, 检测速度有较大提升, 与单阶段检测 YOLOv3 和 MSDN 相比检测速度也较快, 更易于实现和操作, 反映出网络结构的紧凑性以及无预选框模型的优越性。

4.5 DOTA 数据集检测结果

为进一步验证本文所提模型的有效性, 将本文所提模型在大型数据集 DOTA v1.0 进行验证, I_{ou} 阈值取 0.5。采取的基准对比实验为模型 Faster-

RCNN、R-FCN、YOLOv2 和 SCFPN-scf^[30]。其中 Faster-RCNN 和 R-FCN 为提取候选区域的双阶段算法, 采用的主干网络均为 ResNet101。YOLOv2 的主干网络为 GoogleNet。SCFPN-scf 为只基于场景上下文特征进行的适应性优化算法, 其结构为融合了特征金字塔的双阶段检测网络。实验结果如表 6 所示, 可看出本文所提方法的平均精确度大幅超越经典通用检测网络, 同时在上下文特征的应用方面可媲美双阶段 SCFPN-scf 网络, 并有效减少了计算资源。

表 6 不同模型在数据集 DOTA 上的平均精确度对比

Tab. 6 Comparison of the mean-average precisions with different models in the DOTA dataset

模型	主干网络	平均检测精度(mAP)
YOLOv2	GoogleNet	39.20
R-FCN	ResNet101	52.58
Faster-RCNN	ResNet101	60.46
SCFPN-scf	ResNet101	75.22
GCDN	Root-ResNet18	75.95

4.6 可视化显示

图 8 为所提模型的部分检测效果, 涵盖 10 种光学遥感目标。从图中可以看出, 所提模型适合检测复杂背景下不同尺度的目标, 生成的预测框可完整覆盖目标物体, 进一步证明了所提模型的有效性。

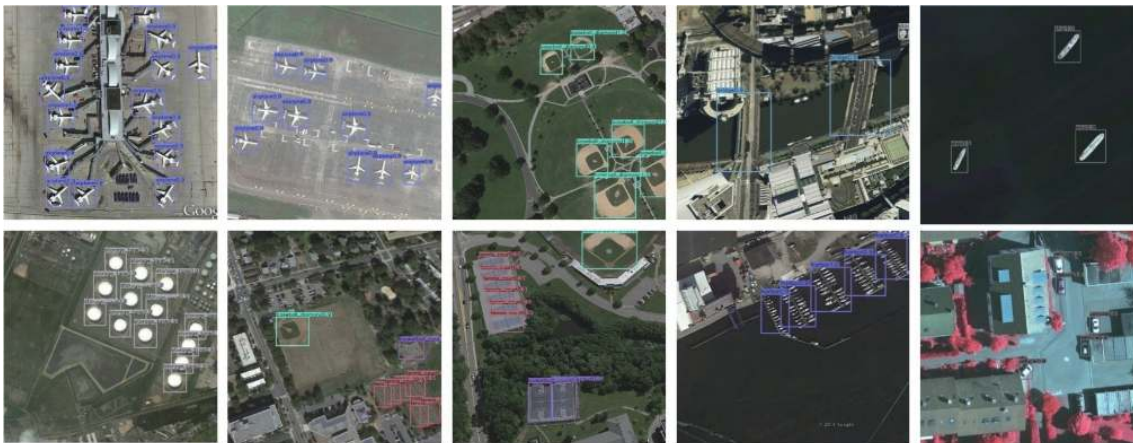


图8 GCDN的可视化检测效果图

Fig. 8 Visual detection results of the GCDN

5 结 论

在复杂背景的光学遥感目标检测中,针对网络复杂度和检测精度难以平衡的现状,提出了一种基于较浅层网络的全局上下文特征检测模型。该模型基于目标中心点的无预选框检测网络,结构简单且检测精度较高。结果显示,全局上下文

特征有益于增强网络的定位分类功能;融合目标框尺度的定位损失函数适用于多尺度目标的检测,可提升网络的回归能力。最终,本文所提方法在公开数据集 NWPU VHR-10 上达到 97.6%AP⁵⁰ 和 83.4% AP⁷⁵ 的检测精度。相较于其他单阶段检测网络,所提模型在网络层数更浅的情况下获得了更强的分类和回归能力。

参考文献:

- [1] 许凤晖,慕晓冬,柯冰,等. 基于遥感影像的军事阵地动态监测技术研究[J]. 遥感技术与应用, 2014, 29(3): 511-516. XU S H, MU X D, KE B, *et al.*. Dynamic monitoring of military position based on remote sensing image[J]. *Remote Sensing Technology and Application*, 2014, 29(3): 511-516. (in Chinese)
- [2] VALERO S, CHANUSSOT J, BENEDIKTSSON J A, *et al.*. Advanced directional mathematical morphology for the detection of the road network in very high resolution remote sensing images[J]. *Pattern Recognition Letters*, 2010, 31(10): 1120-1127.
- [3] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]. *Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE, 2005: 886-893.
- [4] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. *International Journal of Computer Vision*, 2004, 60(2): 91-110.
- [5] LIU W, ANGUELOV D, ERHAN D, *et al.*. SSD: single shot multibox detector[C]. *Proceedings of the 14th European Conference on Computer Vision*, Springer, 2016: 21-37.
- [6] REDMON J, FARHADI A. Yolov3: an incremental improvement[J]. arXiv: 1804.02767, 2018.
- [7] 马永杰, 宋晓凤. 基于YOLO和嵌入式系统的车流量检测[J]. 液晶与显示, 2019, 34(6): 613-618. MA Y J, SONG X F. Vehicle flow detection based on YOLO and embedded system[J]. *Chinese Journal of Liquid Crystals and Displays*, 2019, 34(6): 613-618. (in Chinese)
- [8] LIN T Y, GOYAL P, GIRSHICK R, *et al.*. Focal loss for dense object detection[C]. *Proceedings of 2017 IEEE International Conference on Computer Vision*, IEEE, 2017: 2999-3007.
- [9] LAW H, DENG J. Cornernet: detecting objects as paired keypoints[C]. *Proceedings of the 15th European Conference*

- on *Computer Vision*, Springer, 2018: 765-781.
- [10] ZHOU X Y, WANG D Q, KRÄHENBÜHL P. Objects as points[J]. arXiv: 1904.07850, 2019.
- [11] XIAO B, WU H P, WEI Y CH. Simple baselines for human pose estimation and tracking[C]. *Proceedings of the 15th European Conference on Computer Vision*, Springer, 2018: 472-487.
- [12] LI K, CHENG G, BU SH H, *et al.*. Rotation-insensitive and context-augmented object detection in remote sensing images[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2018, 56(4): 2337-2348.
- [13] MA W P, GUO Q Q, WU Y, *et al.*. A novel multi-model decision fusion network for object detection in remote sensing images[J]. *Remote Sensing*, 2019, 11(7): 737.
- [14] 梁华, 宋玉龙, 钱锋, 等. 基于深度学习的航空对地小目标检测[J]. *液晶与显示*, 2018, 33(9): 793-800.
LIANG H, SONG Y L, QIAN F, *et al.*. Detection of small target in aerial photography based on deep learning[J]. *Chinese Journal of Liquid Crystals and Displays*, 2018, 33(9): 793-800. (in Chinese)
- [15] 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): 1128002.
YAO Q L, HU X, LEI H. Object detection in remote sensing images using multiscale convolutional neural networks[J]. *Acta Optica Sinica*, 2019, 39(11): 1128002. (in Chinese)
- [16] 邓志鹏, 孙浩, 雷琳, 等. 基于多尺度形变特征卷积网络的高分辨率遥感影像目标检测[J]. *测绘学报*, 2018, 47(9): 1216-1227.
DENG ZH P, SUN H, LEI L, *et al.*. Object detection in remote sensing imagery with multi-scale deformable convolutional networks[J]. *Acta Geodaetica et Cartographica Sinica*, 2018, 47(9): 1216-1227. (in Chinese)
- [17] 董潇潇, 何小海, 吴晓红, 等. 基于注意力掩模融合的目标检测算法[J]. *液晶与显示*, 2019, 34(8): 825-833.
DONG X X, HE X H, WU X H, *et al.*. Object detection algorithm based on attention mask fusion[J]. *Chinese Journal of Liquid Crystals and Displays*, 2019, 34(8): 825-833. (in Chinese)
- [18] WANG CH, BAI X, WANG SH, *et al.*. Multiscale visual attention networks for object detection in VHR remote sensing images[J]. *IEEE Geoscience and Remote Sensing Letters*, 2019, 16(2): 310-314.
- [19] 左俊皓, 赵聪, 朱晓龙, 等. Faster-RCNN和Level-Set结合的高分遥感影像建筑物提取[J]. *液晶与显示*, 2019, 34(4): 439-447.
ZUO J H, ZHAO C, ZHU X L, *et al.*. High-resolution remote sensing image building extraction combined with Faster-RCNN and Level-Set[J]. *Chinese Journal of Liquid Crystals and Displays*, 2019, 34(4): 439-447. (in Chinese)
- [20] 于渊博, 张涛, 郭立红, 等. 卫星视频运动目标检测算法[J]. *液晶与显示*, 2017, 32(2): 138-143.
YU Y B, ZHANG T, GUO L H, *et al.*. Moving objects detection on satellite video[J]. *Chinese Journal of Liquid Crystals and Displays*, 2017, 32(2): 138-143. (in Chinese)
- [21] LIU W, RABINOVICH A, BERG A C. Parsenet: looking wider to see better[J]. arXiv: 1506.04579, 2015.
- [22] ZHANG H, DANA K, SHI J P, *et al.*. Context encoding for semantic segmentation[C]. *Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, 2018: 7151-7160.
- [23] HE K M, ZHANG X Y, REN SH Q, *et al.*. Deep residual learning for image recognition[C]. *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2016: 770-778.
- [24] ZHU R, ZHANG SH F, WANG X B, *et al.*. ScratchDet: training single-shot object detectors from scratch[C]. *Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, 2019: 2263-2272.
- [25] LIN M, CHEN Q, YAN SH CH. Network in network[J]. arXiv: 1312.4400, 2013.
- [26] CHENG G, ZHOU P CH, HAN J W. 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.
- [27] XIA G S, BAI X, DING J, *et al.*. DOTA: a large-scale dataset for object detection in aerial images[C]. *Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, 2018: 3974-3983.
- [28] ZEILER M D, FERGUS R. Visualizing and understanding convolutional networks[C]. *Proceedings of the 13th European Conference on Computer Vision*, Springer, 2014: 818-833.

- [29] CHEN K, WANG J Q, PANG J M, *et al.*. MMDetection: open MMLab detection toolbox and benchmark[J]. arXiv: 1906.07155v1, 2019.
- [30] CHEN CH Y, GONG W G, CHEN Y L, *et al.*. Object detection in remote sensing images based on a scene-contextual feature pyramid network[J]. *Remote Sensing*, 2019, 11(3): 339.

作者简介:



张瑞琰(1995—), 女, 河南商丘人, 博士研究生, 2016年于南开大学获得理学学士学位, 2018年于中国科学院大学国家空间科学中心转为硕博连读, 主要从事空间数据处理、遥感目标检测网络优化及压缩的相关研究。E-mail: zhangruiyan16@mailsucas.ac.cn



姜秀杰(1965—), 女, 黑龙江鹤岗人, 博士, 研究员, 博士生导师, 1988年于北京航空航天大学获得工学学士学位, 1991年于中国科学院大学获得工学硕士学位, 2007年于清华大学获得工学博士学位, 主要从事火箭探空技术研究、电场探测技术研究和空间综合电子技术研究。E-mail: jiangxj@nssc.ac.cn

《光学 精密工程》(月刊)

- 中国光学开拓者之一王大珩院士亲自创办的新中国历史最悠久的光学期刊
- 现任主编为国家级有突出贡献的青年科学家曹健林博士
- Benjamin J Eggleton, John Love 等国际著名光学专家为本刊国际编委

《光学 精密工程》主要栏目有现代应用光学(空间光学、纤维光学、信息光学、薄膜光学、光电技术及器件、光学工艺及设备、光电跟踪与测量、激光技术及设备);微纳技术与精密机械(纳米光学、精密机械);信息科学(图像处理、计算机应用与软件工程)等。

- * 美国工程索引 EI 核心期刊
- * 中国出版政府奖期刊提名奖
- * 中国精品科技期刊
- * 中文核心期刊
- * 百种中国杰出学术期刊
- * 中国最具国际影响力学术期刊

主管单位: 中国科学院

主办单位: 中国科学院长春光学精密机械与物理研究所

中国仪器仪表学会

地址: 长春市东南湖大路 3888 号

电话: 0431-86176855

电邮: gxjmgc@sina.com

定价: 100.00 元/册

邮编: 130033

传真: 0431-84613409

网址: <http://www.eope.net>