引文格式：刘海军,杨鸿海. 基于非对称卷积网络的遥感影像地物检测方法研究[J]. 地理空间信息, 2023, 21(9): 73-76. doi:10. 3969/j.issn. 1672-4623.2023.09.018

基于非对称卷积网络的遥感影像地物检测方法研究

刘海军[1,2]，杨鸿海[1,2]
（1. 青海省地理空间和自然资源大数据中心，青海 西宁 810001；
2. 青海省地理空间信息技术和应用重点实验室，青海 西宁 810001）

摘　要：针对高分遥感影像地物目标检测时存在的精度低、鲁棒性较差等问题，提出了一种单阶段目标检测模型。利用包含非对称卷积核的AC模块及其等效常规卷积核构建特征提取端分别进行模型训练和测试，使充分提取特征时不会产生大量计算参数；构建4层特征增强端，通过拼接不同倍数的特征图进一步丰富语义特征，利用K-means++聚类获取的锚点框来提高训练初期拟合速度。以RSOD数据集为基础，选取包含飞机、油罐、操场的影像作为数据集，对训练集进行影像质量增强与噪声样本扩充。结果表明，模型在3类目标上的精度表现均衡，综合检测精度达到91.03%；在测试环境下的检测速度可达27 $m \cdot s^{-1}$，达到实时检测的水平。

关键词：高分遥感影像；地物检测；全卷积网络；非对称卷积；多尺度特征增强
中图分类号：P237　　**文献标志码：**B　　**文章编号：**1672-4623（2023）09-0073-04

Research on Remote Sensing Image Ground Object Detection Method Based on Asymmetric Convolutional Network

LIU Haijun[1,2], YANG Honghai[1,2]
(1. Geographic Space and Natural Resources Big Data Center of Qinghai Province, Xining 810001, China;
2. Key Laboratory of Geospatial Information Technology and Application of Qinghai Province, Xining 810001, China)

Abstract: Aiming at the problems of low measurement accuracy and poor robustness in high-resolution remote sensing image ground object detection, we proposed a single-stage target detection model. The model uses the AC module including the asymmetric convolutional kernel and its equivalent conventional convolutional kernel to construct the feature extraction end to train and test the model respectively, to ensure that the model can fully extract features without generating a large number of computational parameters. This model constructs a four-layer feature enhancement end to further enrich semantic features by splicing feature maps of different multiples, and use the anchor boxes obtained by K-means++ clustering to improve the fitting speed in the early training stage. Based on the RSOD data set, we selected images including aircraft, oil tanks and playgrounds as the data set, and performed image quality enhancement and noise sample expansion on the training set. The experimental results show that the accuracy of model is balanced on the three types of targets, and the comprehensive detection accuracy reaches 91.03%. The detection speed can reach 27 $m \cdot s^{-1}$, reaching the level of real-time detection.

Key words: high-resolution remote sensing image, ground object detection, fully convolutional network, asymmetric convolution, multi-scale feature enhancement

高分遥感影像包含海量地物要素信息，对其进行快速精准的检测，能在智慧城市建设、智能交通、应急救灾、军事侦察等领域发挥重要价值，是长期以来的研究热点。许多学者曾以传统机器学习方法为出发点，通过梯度直方图[1]、颜色直方图[2]、SIFT[3]等人工归纳的特征实施检测，但其对目标的表达能力严重依赖设计者经验，且该类特征仅能表达影像中目标的浅层次信息，在实际应用中错检、漏检率较高。随着深度学习的兴起，基于卷积神经网络的目标检测模型因能挖掘目标深层次语义特征而得到广泛应用，如李妹燕[4]结合top-hat形态学滤波算法与形变卷积神经网络对红外遥感影像中的多种目标进行检测，取得了漏警率与虚警率均低于1.5%的结果；史超[5]基于两阶段的快速区域卷积神经网络对遥感影像中的船舶目标进行检测，在海洋卫星遥感数据集上取得了精度均值78.4%的结果；黄西尧[6]等通过高斯过程回归和模板匹

收稿日期：2021-12-28；修回日期：2022-09-02。
项目来源：青海省自然科学基金资助项目（2020-ZJ-927）。
第一作者简介：刘海军（1972—），工程师，主要从事省级基础测绘地理信息数据库建设、省级生态监测与建设、天地图省级节点建设、地理信息与地图编制、遥感等工作，E-mail：liuhaijun19720215@163.com。

配对多帧图像进行前景/背景分割与配准，并构建深度卷积神经网络对低精度CCD影像中的地球同步轨道目标进行检测，在测试集上检测精度最高可达98%。受卫星拍摄高度、拍摄角度、地面光照变换和地物阴影等因素影响，高分遥感影像内的目标往往背景复杂多样，且多以小尺寸目标的形式存在，因此常规方法难以精确检测遥感影像内的地物目标。

鉴于此，本文提出了一种单阶段的遥感影像地物检测模型，首先以非对称卷积核与小尺寸常规卷积核组成的AC模块构建特征提取端，然后使用4层跨层特征强化端对提取特征进行上采样，最终通过4个不同感受野的特征图实现多尺度目标检测。本文采用暗通道先验算法与颜色直方图均衡算法对训练集内的图像进行增强处理，同时利用几何变换、添加高斯噪声等方法制作包含噪声的样本数据，以提高训练后模型在不同场景下的鲁棒性。

1 遥感影像多目标检测模型构建

基于卷积神经网络的深度学习模型主要包括基于候选区域的二阶段法和基于目标回归的单阶段法[7]，前者先通过前端网络在整张图像中生成候选区，再使用后端检测网络输出目标的位置与类别检测结果，虽具有较高精度，但其结构复杂、计算参数量大、检测过程缓慢；后者将目标检测问题视为回归问题，通过整张自上而下的特征提取端从原始影像中提取不同感受野下的目标语义信息，再通过特征增强端对特征图进行融合增强，最后利用多个尺寸的锚点框来实现检测输出。本文采用单阶段目标检测模型结构。

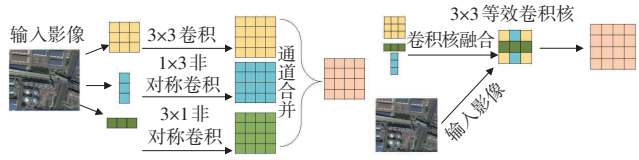
1.1 特征提取端

为确保模型能提取足够丰富的特征信息参与训练，且在测试阶段不会因模型结构过于复杂而导致检测速度过慢，本文借鉴了卷积核计算的可叠加性，在训练和测试阶段采用不同的等效卷积核来构建特征提取模块，表达式为：

$$M\times C^m+M\times C^n=M\times(C^m\oplus C^n) \tag{1}$$

式中，M 为输入的特征图；C^m、C^n 为卷积核；$C^m\oplus C^n$ 为二者的等效卷积核。

在训练阶段，特征提取层采用3×3常规卷积核与非对称卷积核的并联结构进行特征提取，再将提取的特征图进行通道合并以提高特征图内语义信息的丰富程度（图1a）。在检测阶段，为使模型采用参数一致的卷积核进行操作的同时尽可能减少计算量，对AC模块中的并联卷积核进行融合以获取其等效卷积核，并利用该卷积核进行特征提取。卷积核的融合通过核矩阵的对应位置求和操作来实现，检测阶段的计算过程见图1b。



a AC结构　　b 等效卷积核

图1 AC结构训练与测试过程示意图

为增强特征提取后非线性特征的表达输出能力，同时避免梯度消失问题[8]，本文采用无边界、非饱和、平滑的非线性函数Mish作为输出特征的激活函数，相较于Relu、Sigmoid等硬边界函数，Mish函数能够避免梯度饱和问题，且平滑的激活函数能让模型得到更好的准确性和泛化能力。其表达式为：

$$\mathrm{Mish}(x)=x\times\tanh(\log(1+e^x)) \tag{2}$$

式中，x 为输入特征值。

为加快模型的训练收敛，防止模型出现过拟合以及因上层输入特征分布不均而产生的协变量偏移[9]，在特征提取层后采用批量再规范化层对特征提取后的输入进行处理。为尽可能完整地保存原始图像内的语义信息，模型采用固定卷积核两倍下采样的方式来压缩图像尺寸，完整结构见图2。

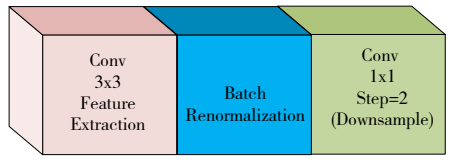


图2 特征提取—下采样层示意图

本文采用的损失函数由类别损失 L_{class}、置信度损失 L_{conf}、边界框损失 L_{bound} 组成，具体公式为：

$$\begin{cases} L_{\text{class}}=\lambda_{\text{class}}\sum_{i=0}^{K\times K}\sum_{c\in 3}I_{ij}^{\text{obj}}\left[\hat{p}_i\log p_i+(1-\hat{p}_i)\log(1-p_i)\right] \\ L_{\text{bound}}=\lambda_{\text{bound}}\sum_{i=0}^{K\times K}\sum_{j=0}^{M}I_{ij}^{\text{obj}}[(x_i-\hat{x}_i)^2+(y_i-\hat{y}_i)^2]+\lambda_{\text{bound}}\sum_{i=0}^{K\times K}\sum_{j=0}^{M}I_{ij}^{\text{obj}}(2-w_i\times h_i)[(w_i-\hat{w}_i)^2+(h_i-\hat{h}_i)^2] \\ L_{\text{conf}}=\lambda_{\text{conf}}\sum_{i=0}^{K\times K}\sum_{j=0}^{M}I_{ij}^{\text{obj}}\left[\hat{C}_i\log C_{ij}+\left(1-\hat{C}_{ij}\right)\log\left(1-C_{ij}\right)\right]+\lambda_{\text{conf}}\sum_{i=0}^{K\times K}\sum_{j=0}^{M}I_{ij}^{\text{nobj}}[\hat{C}_{ij}\log C_{ij}+\left(1-\hat{C}_{ij}\right)\log\left(1-C_{ij}\right)] \end{cases} \tag{3}$$

式中，λ_{class}、λ_{bound}、λ_{conf} 为权重系数；L_{class} 为多分类交叉熵型损失函数，类别个数 K 为图像网格数，$c=3$ 为目标类别个数，$I_{ij}^{\text{obj}}=0$ 或1用以判断第 i 个网格中第 j 个预测框是否负责检测该类物体的阈值，p_i、

$\hat{p}_i$ 分别为该物体存在的预测概率与真实概率；L_{bound} 为边界框定位损失函数，x_i、y_i、w_i、h_i 与 $\hat{x}_i$、$\hat{y}_i$、$\hat{w}_i$、$\hat{h}_i$ 分别为预测框与真实框的中心点坐标、边框长和宽；L_{conf} 为置信度损失函数，C_{ij}、$\hat{C}_{ij}$ 分别为第 i 个网格下第 j 个预测的预测置信度和真实置信度。

1.2　特征增强端

为丰富特征图的语义信息，实现精准的多尺度目标检测，将特征增强端连接在特征提取端末尾，通过上采样和特征图拼接的方式获取多个感受野的特征图，具体结构见图3，可以看出，特征增强端将提取端底层特征图为基础连续进行3次的上采样作为主干，为进一步增强上采样所得特征图内特征的丰富程度，不仅将提取端内同尺度特征图与其进行拼接，而且对增强端内特征图进行两倍上采样，对3种不同来源的特征图像进行拼接，将语义信息更丰富的特征图送入检测输出层。

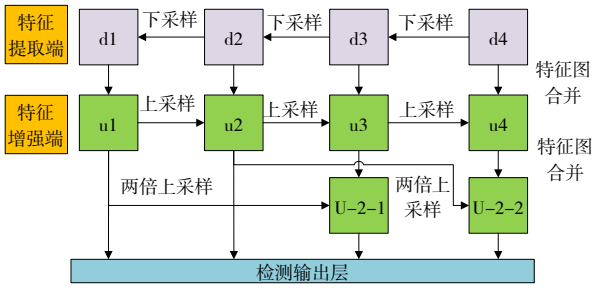


图3　特征增强端结构示意图

1.3　锚点框选取

锚点框是单阶段检测模型实现多尺度目标检测的必要参数，其宽高比与目标真实尺寸相近，能有效提高模型拟合速度与检测精度。本文以训练数据为基础，采用K-means++聚类获取锚点框。具体步骤为：①以全部训练目标的真实尺寸框为样本，随机选取一个点作为首个聚类中心；②获取其余样本与聚类中心之间的最小距离 $D=1-\text{IOU}(\text{box},\text{centroid})$，计算其余样本被选作聚类中心的概率值 $P=\dfrac{D(x)^2}{\sum_{x\in X}D_x}$，其中IOU为目标框box与真实框centroid的交并比；③利用轮盘算法迭代 m 次，直到选择出所有聚类中心（$m+1$）；④计算其余非中心样本与所有聚类中心的距离 D，将每个样本归纳到与其距离最近的中心类簇内；⑤为每个类簇重新选取一个聚类中心，重复直至所有中心固定。

结合特征增强端输出特征图个数与模型计算开销，本文通过12个聚类中心对目标框进行聚类。聚类结果为（19,16）、（22,22）、（27,29）、（28,20）、（34,37）、（40,41）、（46,46）、（51,61）、（55,51）、（73,48）、（78,77）、（121,78）。本文将12个锚点框平分给4个尺度的输出特征图，其中特征增强端输出的最下层特征图尺寸最小但感受野最大，适合检测大尺寸目标，因此将3个尺寸最大的锚点框（73,48）、（78,77）、（121,78）缩小4倍后用于该层；最上层特征图感受野最小，适合检测小尺寸目标，因此获得3个最小的锚点框（19,16）、（22,22）、（27,29），其余特征图均按该标准分配合适的锚点框。

2　模型训练与结果分析

2.1　数据集预处理

本文以RSOD卫星遥感影像集内包含飞机、油罐、操场目标的影像作为数据集，将其进一步划分为训练集、验证集和测试集。由于原始数据集内影像尺寸较大，无法直接放入模型进行训练，因此本文采用608×608窗口采集训练集内的影像，再通过人工筛选将包含样本的影像作为训练集。受卫星拍摄高度、角度、时间等因素影响，原始卫星影像质量参差不齐，为提高训练后模型检测精度，本文采用暗通道先验算法[10]和颜色直方图均衡算法对训练集进行图像增强；同时为提高训练后模型的鲁棒性，通过几何变换、高斯噪声等方法制作了噪声样本放入训练集内，部分增强图像和噪声样本见图4。处理后的训练集、验证集、测试集分别包含影像5 453张、545张和684张，按照Pascal VOC个数对训练集中的样本进行标注，标注后的飞机样本共4 587个，操场样本共873个，油罐样本共3 659个。



a 原始影像　　b 增强后影像　　c 包含高斯噪点的影像

图4　图像增强处理对比图

2.2　模型训练

本文采用的硬件配置：CPU为Intel i9 9900k，GPU为Nvidia GTX 3080，内存为16 G，硬盘为1T-SSD，运行环境为Ubuntu16.04、Python3.7，深度学习框架构建为TensorFlow，运算加速为Cuda10.0。训练过程中采用Adam优化器实现模型参数的动态更新，并采用动态学习率调整机制每隔1 500次迭代动态调整学习率一次，避免模型陷入局部极小值。初始学习率为0.000 35，每次调节的倍数为0.1倍，单次输入模型参与训练的图片数量为20张，训练总迭代次数为7 000次。模型训练迭代过程中的损失变化

见图5，可以看出，模型训练损失在训练初期快速收敛，而在后续训练中始终保持稳定降低，最终收敛到50以内，说明模型训练参数设置较好，训练拟合情况理想。

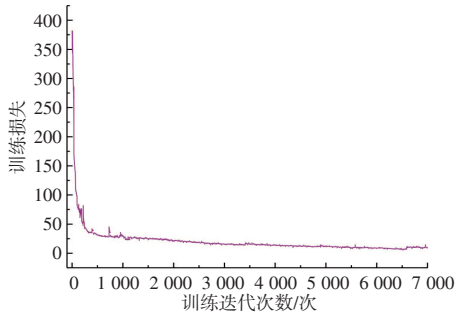


图5 训练拟合曲线示意图

2.3 模型测试与评价

为全面准确地评价本文方法，以每一类目标的平均准确率（*AP*）、全部类目标的平均准确率均值（mAP）和每秒传输帧率（FPS）为评价指标，以SSD、R-FCN、Yolov4深度学习算法为对照方法，对模型进行综合评价。测试检测速度时，测试集中的图像尺寸统一为608×608，将全部测试集送入模型检测后，按照处理单张图像所需时间来计算最终的FPS。评价结果见表1，可以看出，模型能稳定检出全部类别的目标，其中最高单类别检测精度可达96.53%，对于小尺寸的飞机目标也能达到87.24%的检测精度；模型的mAP达到了91.03%。由图6可知，模型能较好地检出不同角度、颜色背景下的多个尺度目标，说明本文方法有效提高了遥感影像中小目标物体的特征表达能力与丰富程度，使模型得到更充分的训练，具备了更好的泛化能力；进一步说明了AC模块能更有效地提取影像中的特征信息，利用4层特征增强结构也能更好地检测小尺度目标，因此模型在检测精度方面具备明显优势。模型的检测速度在测试硬件环境下可达27 m·s^{-1}，略低于SSD、Yolov4算法，其原因在于本文模型不仅在特征提取端的结构比SSD、Yolov4模型更复杂，而且为了提高小尺寸目标的检出精度，在特征增强端设置了4层多尺度检出输出窗口，增加了推理过程的计算量，牺牲了一部分检测速度，但整体上仍能达到实时检测水平。

表1 各类目标 *AP* 对比

指标	SSD	R-FCN	Yolov4	本文模型
飞机/%	78.56	79.54	82.16	87.24
操场/%	82.17	93.75	92.45	96.53
油罐/%	81.57	84.53	85.23	92.65
mAP/%	79.20	82.94	84.94	91.03
FPS/m·s^{-1}	47	15	29	27

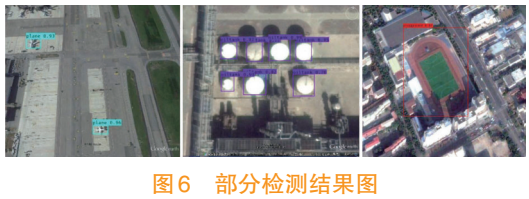


图6 部分检测结果图

3 结 语

针对高分遥感影像背景复杂以及现有方法对小目标检测精度较低、鲁棒性较差等问题，提出了一种单阶段遥感影像地物检测方法，以RSOD数据集为基础，对模型进行训练和测试，并采用 *AP* 、mAP和FPS等指标对模型进行评价。

1）模型对各类目标的检测精度表现均衡，且mAP达到91.03/%。实际检测结果证明模型能很好地检出不同背景复杂程度下多类别、多尺寸大小的目标，具备很好的稳定性和泛化能力。

2）模型的检测速度在测试硬件环境下可达27 m·s^{-1}，达到实时检测水平。

综上所述，模型能精准快速地检测高分遥感影像内不同尺度的地物目标，能在智慧城市建设、城市监督管理、军事侦察等多个领域发挥重要作用。在今后的研究中，将优化模型结构，扩充训练集，提高模型对密集、遮挡等更复杂目标的检测精度；同时采用轻量化手段对模型进行结构压缩，使其能部署在更多环境下开展流畅推理。

参考文献

[1] 刘方园,王水花,张煜东. 方向梯度直方图综述[J]. 计算机工程与应用,2017,53(19):1-7

[2] 张慧,王坤峰,王飞跃. 深度学习在目标视觉检测中的应用进展与展望[J]. 自动化学报,2017,43(8):1 289-1 305

[3] 穆莉莉,姚潘涛,郭枫,等. BP神经网络在SLAM特征匹配中的应用[J]. 测绘科学,2020,45(10):27-32

[4] 李妹燕. 基于深度学习网络的红外遥感图像多目标检测[J]. 激光杂志,2021,42(11):107-111

[5] 史超. 海面舰船遥感目标视觉检测技术研究[J]. 无线电工程,2021,51(11):1 189-1 194

[6] 黄西尧,何羿霆,杜华军,等. 地球同步轨道目标物深度学习检测方法[J]. 宇航学报,2021,42(10):1 283-1 292

[7] 罗会兰,陈鸿坤. 基于深度学习的目标检测研究综述[J]. 电子学报,2020,48(6):1 230-1 239

[8] 戴激光,杜阳,金光,等. 一种多尺度卷积神经网络道路提取方法[J]. 遥感信息,2020,35(1):28-37

[9] 卢海伟,袁晓彤. 基于层融合特征系数的动态网络结构化剪枝[J]. 模式识别与人工智能,2019,32(11):1 051-1 059

[10] 陈伟,贺元恺,李昭慧,等. 一种改进实时图像去雾的暗原色先验算法[J]. 现代电子技术,2020,43(6):163-167