

基于深度卷积神经网络的遥感影像目标检测

孙梓超¹, 谭喜成¹, 洪泽华², 董华萍^{1,3}, 沙宗尧¹, 周松涛¹, 杨宗亮¹

(1. 武汉大学遥感信息工程学院, 湖北 武汉 430079; 2. 上海机电工程研究所, 上海 201109;

3. 城市空间信息工程北京市重点实验室, 北京 100038)

摘要:随着遥感影像数据规模的快速扩张,如何高效准确地识别遥感影像中的典型目标成为当前的研究热点。为解决传统遥感影像目标检测方法准确率低的问题,用基于深度卷积神经网络进行遥感影像目标检测,在遥感影像数据集上用基于 Faster-RCNN 的神经网络模型对 VGG16 卷积网络进行训练,对输入的遥感影像通过区域推荐网络标注出待检目标的包围框和置信度,实现对遥感影像的目标检测。以飞机和油罐为例,在 TensorFlow 深度学习框架下实现了数据预处理、网络训练、目标检测等功能,并在当前测试数据集上取得了较高的检测准确率和置信度。该研究成果可应用于遥感影像解译和处理等相关领域。

关键词:深度卷积神经网络; 遥感影像目标检测; 区域卷积神经网络; 深度学习; TensorFlow 框架

中图分类号: TP753

文献标志码: A

DOI: 10.19328/j.cnki.1006-1630.2018.05.003

Remote Sensing Image Object Detection Based on Deep Convolution Neural Network

SUN Zichao¹, TAN Xicheng¹, HONG Zehua², DONG Huaping^{1,3}, SHA Zongyao¹,
ZHOU Songtao¹, YANG Zongliang¹

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

2. Shanghai Electro-Mechanical Engineering Institute, Shanghai 201109, China;

3. Beijing Key Laboratory of Urban Spatial Information Engineering, Beijing 100038, China)

Abstract: With the rapid expansion of remote sensing image data, how to identify typical object efficiently and accurately in remote sensing images has become a hot research area. In order to break through the bottleneck of the accuracy of traditional remote sensing image object detection methods, this paper proposes a deep convolution neural network method to detect remote sensing image objects, and uses the Faster-RCNN neural network model to train VGG16 convolution networks on remote sensing image datasets. The input remote sensing image is marked with the bounding box and confidence level of the object to be inspected through the region proposal network, thereby realizing the target detection of the remote sensing image. Taking aircraft and oil tank as examples, this paper realizes data preprocessing, network training, object detection and other processes under the TensorFlow deep learning framework, and achieves relative high detection accuracy and confidence on the current test dataset. The research results of this paper can apply to remote sensing image interpretation and remote sensing image processing.

Keywords: deep convolution neural network; remote sensing image object detection; region-based convolution neural network; deep learning; TensorFlow framework

收稿日期: 2018-08-01; 修回日期: 2018-09-08

基金项目:国家重点研发计划项目(2017YFB0504202); 上海航天科技创新基金项目(SAST2016006); 湖北省自然科学基金项目(2017CFB433); 智能地学信息处理湖北省重点实验室(中国地质大学(武汉))开放基金(KLIGIP-2017A09); 空间数据挖掘与信息共享教育部重点实验室(福州大学)开放基金(2016LSDMIS06); 城市空间信息工程北京市重点实验室经费资助项目(2017209)

作者简介:孙梓超(1997—),男,本科,主要研究方向为时空大数据分析、深度学习应用。

通信作者:谭喜成(1979—),男,副教授,主要研究方向为云计算、大数据分析、空间信息服务及地学高性能计算等。

0 引言

高分辨率遥感影像的目标检测是计算机视觉识别、数字影像处理等技术在遥感领域的重要应用^[1-2],被广泛应用于军事侦察、国土资源监测、违章建筑及违法用地识别等领域。传统遥感影像目标检测方法一般基于影像处理技术,如几何特征提取、纹理选择、影像分割,在此基础上用对比匹配的方法识别目标。该方法需要充分的特征工程建设,计算量大,工作流程复杂,检测准确率和效率较低,难以满足实际应用中检测质量和效率的要求^[2]。近年来,深度卷积神经网络在影像检测识别等领域表现出较大优势,该网络能根据标注,从海量影像中有效提取目标底层特征并进行训练。目标对象的各种特征几何关系被网络不断学习和优化,最终得到目标的高层深度特征^[1],能准确检测影像中的目标。本文基于深度卷积神经网络,设计了一种遥感影像目标检测方法,并以飞机、油罐目标为例,建立一个通用的遥感影像目标检测网络,在测试数据集上取得了较高的准确率和置信度。

1 研究现状

深度卷积神经网络是近年来深度学习应用到影像分类、识别等方面的一个重大成果。卷积神经网络最早可追溯到 1998 年提出的 LeNet-5, LeNet-5 是第一个多层结构神经网络模型,其训练过程采用反向传播算法^[3-5]。与传统的全连接神经网络相比,LeNet-5 增加了卷积层和池化层(降采样层)。卷积层对影像进行逐像素特征提取,和前一层局部连接且参数共享,既保证了对影像特征的充分提取,又合理减少了参数数量;池化层降低了输入到下一层的数据维度,从而降低了整个网络的复杂度。然而,在实际应用中,随着网络结构深度加深,参数数量逐渐增多。同时,受制于当时的硬件计算能力且缺乏大型训练数据集,训练中经常出现拟合现象,网络模型训练所得的参数只在训练数据集上表现良好,而在其他数据集上效果一般,可扩展性低。在此背景下,HINTON 等^[6]于 2006 年提出了对待训练参数进行预训练的初始化方法,该方法比随机初始化待训练参数的方法效果更好。随着计算机软硬件技术的突飞猛进,以及政府、企业等组织保

有数据量的不断增大,深度卷积神经网络迎来了发展高峰。2012 年,KRIZHEVSKY 等^[7]设计了 AlexNet 模型,用一种整流线性单元(ReLU)来解决梯度消失问题,以 Dropout^[8]正则化技术防止过拟合,同时使用 GPU 作为网络训练的硬件计算单元,大幅度提高了训练速度。该网络模型在 ImageNet 图片分类比赛中将错误率从 26% 降低到 15%。目前常用的深度卷积神经网络有 VGG16、ResNet、Inception、Xception 等。

深度卷积神经网络虽然在影像分类和识别领域取得了较大成就,但在目标检测中,不仅需要识别出影像中的目标,还需要检测出目标在影像中的位置,这对普通的深度卷积神经网络而言较为困难^[9]。与此同时,对于复杂的影像如遥感影像,一张影像可能存在不同数量的多种目标,如何识别出不同目标的种类并以包围框形式准确标定目标在影像中的位置,也成为深度卷积神经网络无法解决的难题。在此背景下,GIRSHICK^[9]提出了区域卷积神经网络(RCNN)。RCNN 相比传统目标检测方法,显著提高了检测准确率和效率,以 PASCAL VOC 数据集为例,检测精度从 35.1% 提高到了 53.7%。

但 RCNN 也存在如下问题:1) 候选区域影像预先提取,会消耗大量存储空间;2) 每个候选区域均输入卷积神经网络(CNN)计算,存在大量交叠区域,重复进行特征提取会浪费资源。针对此,GIRSHICK 团队在原有基础上提出了 Fast-RCNN^[9-10],其结构设计更精巧,利用对输入图片的大小归一化操作和均匀切割候选区域完成对应步骤,大大减少了每张图片的候选区域数量,显著提高了网络训练速度^[10]。

之后,GIRSHICK 团队又提出了 Faster-RCNN^[10-11],相比 RCNN 和 Fast-RCNN,其进步在于将目标检测的 4 个步骤统一在一个网络结构中。目标特征提取使用一个深度卷积网络,如 VGG16、Inception、ResNet,在候选框生成和类别判断中共享特征参数,大大减少了重复计算,利用区域生成网络完成候选区域选取,进一步加快了训练和测试速度,实现了与 GPU 的完全兼容,是目前最准确、快速的目标检测方法之一。

RCNN 的发展变化如图 1 所示。

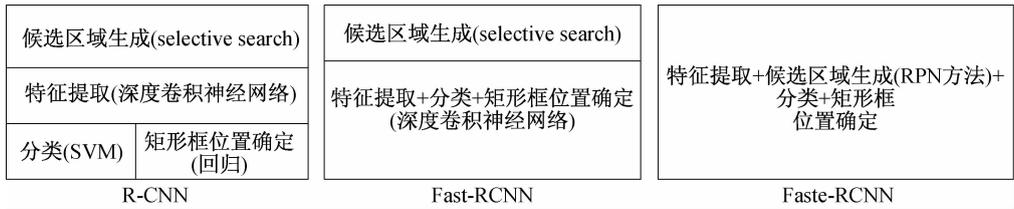


图 1 RCNN 的发展过程
Fig. 1 Development process of RCNN

2 基于 Faster-RCNN 的目标检测方法

2.1 Faster-RCNN 目标检测过程

Faster-RCNN 目标检测过程可分为如下 4 个步骤:

1) 利用 CNN 对图片进行特征提取,生成特征图(feature maps)。

2) 在步骤 1) 生成的特征图基础上用区域推荐网络(region proposal network, RPN)进行候选区域生成,该网络用 softmax 激活函数通过每个候选框中心锚点(anchor)判别该候选框属于背景还是目标,再通过矩形框回归对锚点修正获得精确的候选框(proposals)。

3) Roi Pooling 层获得步骤 1)、2) 中的产生的 feature maps 和 proposals,综合得出输入图片中可能为目标的部分及该部分在原图中的相对位置,称为 proposal feature maps,并输入全连接层进行目标所属类别判断。

4) 全连接层判断步骤 3) 中传入的每个 proposal feature maps 类别,并再次通过矩形框回归获得更精确的目标矩形框,完成目标检测过程。

Faster-RCNN 流程如图 2 所示。作为 Faster-RCNN 目标检测方法中的 2 个重要组成部分,CNN 和 RPN 分别解决了目标是什么和目标在哪里的问题。

2.2 CNN 结构

CNN 是深度学习方法应用到遥感影像目标检测领域的理论和技术基础,尽管 CNN 在 3 种 RCNN 中的位置和具体结构略有不同,但其作用均是进行物体特征提取和训练。CNN 结构中的主要组成部分包括卷积层、池化层、全连接层、激活函数。以本次实验使用的 VGG16 网络为例,其结构如图 3 所示。

卷积层用于图片特征的提取,其实现依赖于卷积核,如图 4 所示。卷积核在一幅 5×5 的影像上按从左到右从上往下进行逐像素卷积,将输入影像中的一个 3×3 子区的特征聚合为输出影像中的一个 1×1 子区。这样做可使产生的特征图中相邻神经元共享大部分输入,保留图像中目标的大部分特征。遥感影像中,目标与背景在纹理、范围等方面的差异在图片中表现为像素规律的差别,这种差别通过多

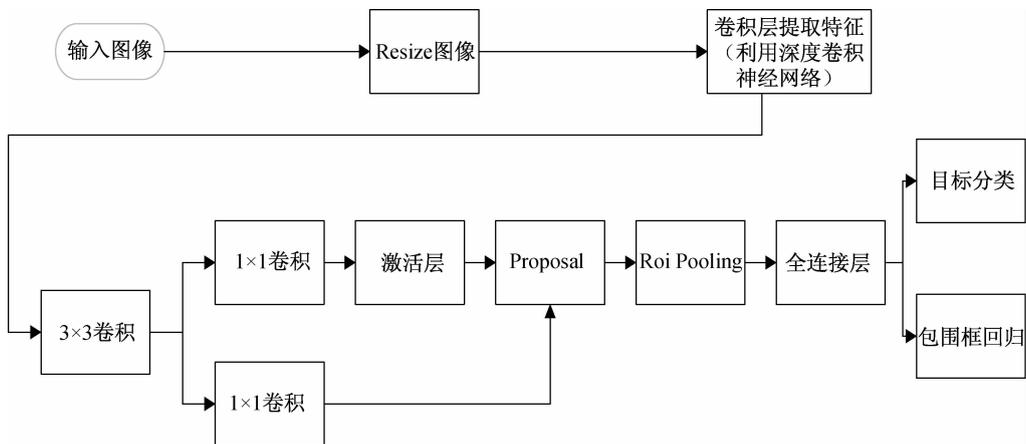


图 2 Faster-RCNN 流程
Fig. 2 Flowchart of Faster-RCNN

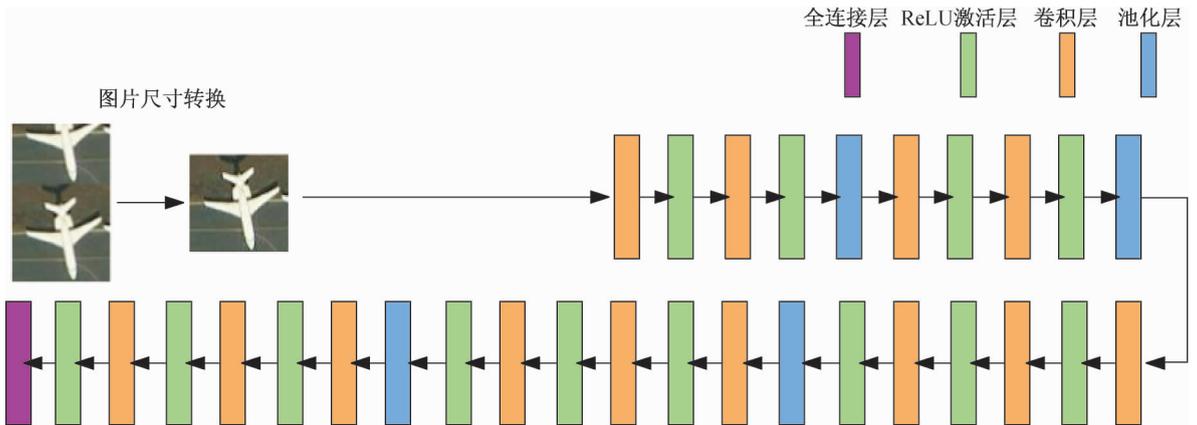


图 3 VGG16 网络结构

Fig. 3 Network structure of VGG16

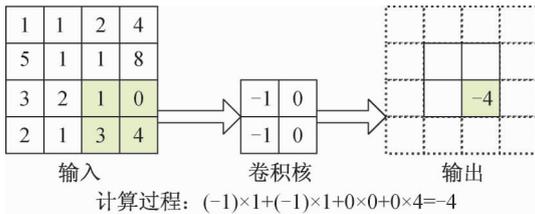


图 4 卷积层示意图

Fig. 4 Schematic diagram of convolution layer

层卷积核提取后转化为特征图上的黑色背景和非黑色的目标抽象线条,从而达到获得目标特征的目的。

池化层又称为降采样,在卷积后继续降低参数数量和模型复杂度,同时进行特征压缩,提取主要特征。常用的池化操作有最大池化法和平均池化法^[12]。最大池化示意图如图 5 所示,其中尺寸为 2×2 的过滤器以 2 为步长在输入特征图上进行滑动,输出由每个相邻 2×2 区域中最大值组成的特征图,从而达到降低参数数量和提取主要特征的目的。



图 5 最大池化示意图

Fig. 5 Schematic diagram of maximum pooling

遥感影像具有数据量大、图片蕴含信息复杂等特点,因此在 CNN 中合理使用池化层能大幅度降低计算参数,提高检测速度。遥感影像中目标和背景的像素值普遍存在差异,选择合理的池化层舍去

的信息绝大多数是无用信息,对网络的识别能力影响很小。

全连接层在整个 CNN 中起“分类器”作用,在实际使用中,全连接层可由卷积操作实现,相当于卷积核为 1×1 的卷积层^[8]。本实验中,图片经过 Roi Pooling 处理后,全连接层对其中包含的目标进行种类识别。

激活函数是非线性函数,用于在 CNN 中加入非线性元素。卷积层、全连接层计算完成后,一般会将结果通过某个激活函数转换为非线性形式,以此来增加网络的表达和泛化能力^[8]。目前,大多数卷积神经网络均采用 ReLU^[12] 作为激活函数。

2.3 RPN 结构

生成目标候选区域的质量直接影响到目标检测任务的准确率,是区域卷积神经网络中至关重要的一环,也是其与传统卷积神经网络的重要区别。在 RCNN 和 Fast-RCNN 中,候选区域生成采用选择搜索的方法,利用划窗对每幅影像提取约 2 000 个候选区域^[9-10]。

在 Faster-RCNN 中,用 RPN 改进候选区域的生成。RPN 是全卷积的网络,在提取特征的 CNN 后,实现特征参数共享,大大减少候选区域选取过程和特征提取过程中的重复计算,做到了端到端的目标检测,形成了统一的区域卷积神经网络结构^[11]。RPN 包括 1 个过滤器尺寸为 1×1 的卷积层和 2 个平行的全连接层。RPN 与选择搜索均用划动窗口策略生成候选区域,但 RPN 的划动窗口操作与选择搜索不同,在卷积层特征图上进行,经过卷积池化

操作,维度缩小了 256 倍。

在卷积层特征图上,基础划窗尺寸为 3×3 ,以每个划窗的中心点为锚点,通过锚点,按不同的缩放倍数和长宽比例得到不同大小的划窗尺寸,以适应待检测目标大小和比例不一的情况。常用的公共基准窗口有 3 个倍数(1,2,4)和 3 个比率(0.5,1,2),因此可得 9 个尺寸的划窗,如图 6 所示。利用上述方法,对每个锚点用不同尺寸的划窗划动生成不同尺寸的候选区域,这些候选区域经过 RPN 网络中的一层卷积后转换为 256 维的全连接特征,再分别输入 2 个全连接层用于判别目标是否存在,并进行后续的矩形框位置回归计算^[11],上述过程即 RPN 的工作原理,如图 7 所示。判别目标是否存在的具体方法为以候选标注框与人工标注框交并比为依据,对每个人工标注框,取与其交并比最大的候选框为正样本,小于某一阈值的候选框为负样本。

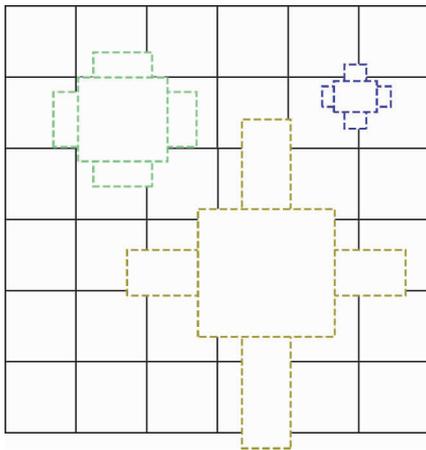


图 6 不同尺寸划窗的获取

Fig. 6 Acquisition of sliding windows with different dimensions

3 飞机目标检测实验

3.1 数据及实验过程

飞机、油罐的遥感影像数据集使用 RSOD-

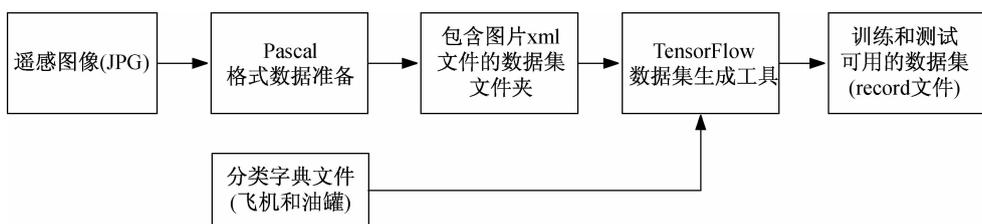


图 9 数据预处理流程

Fig. 9 Flowchart of data preprocessing

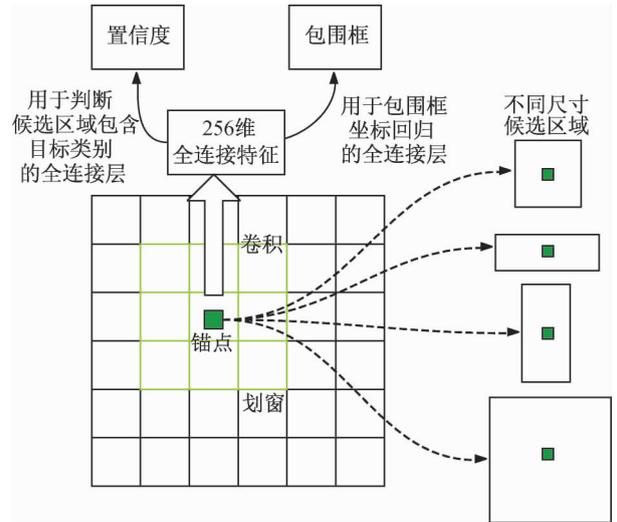


图 7 RPN 工作原理示意图

Fig. 7 Schematic diagram of operating principle of RPN

Dataset^[13-14],标注方法与 Pascal 数据集一致,飞机图像大小为 $1\ 044 \times 915$,油罐图像大小为 $1\ 075 \times 923$,如图 8 所示。本次实验所使用的遥感影像均为对地俯视图,未发生形变现象,为目标检测提供了便利。同时,待检测目标大小在 30×30 到 150×150 范围内,相对图像尺寸较小,且不同目标彼此大小差异较大,增加了准确检测的难度。本实验基于谷歌 TensorFlow 框架^[15],用数据集转换工具将遥感影像转换为 TensorFlow 可使用的数据集形式,上述数据预处理流程如图 9 所示。

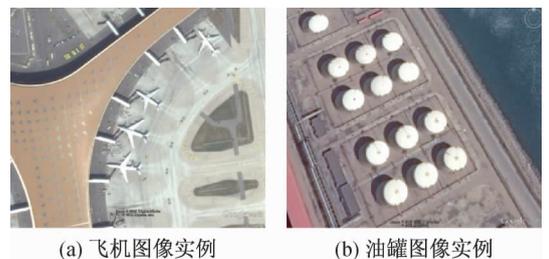


图 8 数据集图片实例

Fig. 8 Examples from image dataset

网络模型的训练使用 NVIDIA GEFORCE GTX1060 显卡,因此需在 Windows10 系统上搭建 cuda8.0+ cudnn6 的 GPU 运算环境。用基于 pascal 数据集预训练的 Faster-RCNN 模型进行定制训练,设待检测目标类别为 2,学习率为 0.002%,初步训练 100 000 轮,将训练参数存储到本地,并导出为静态网络参数模型以备后续测试^[16]。在测试数据集上进行网络测试,测试结果如图 10 所示。

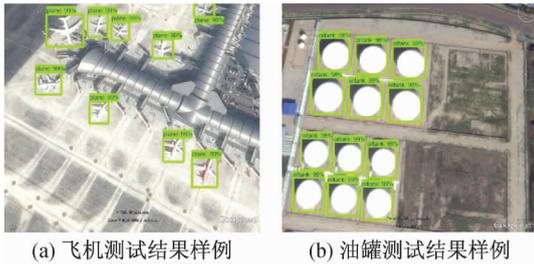


图 10 测试结果样例

Fig. 10 Samples of test results

3.2 讨论

针对每张测试图片,本文采用了一种直观的目标检测置信度计算方法,即检测目标置信度之和除以每张图中实际的目标数目,并定义未检测到的目标和检测错误的目标,置信度为 0,有

$$P = (\sum_1^j C_j) / j \quad (1)$$

式中: P 为每张图的检测置信度; j 为图中实际的目标数目; C_j 为每个目标的检测置信度。

用式(1)对测试结果进行统计,总平均置信度为 98.4%。同时,所有测试图片中每个目标均被框出,检测准确率达到 100%,较之传统遥感图像目标检测方法有较大提高。

本实验中,影响检测准确率的因素主要有:

1) 训练数据集规模较小。训练数据中遥感影像的光照、天气等过于接近,飞机种类、颜色、角度基本相同,网络泛化能力较弱。

2) 存在过拟合现象^[17-18]。这是因为借用了谷歌发布的预先训练的 Faster-RCNN 模型,模型内部已存在的参数对训练有一定影响。

3) 测试用例分辨率参差不齐。在不经影像处理的情况下,检测准确率受到较大影响。

后续研究可从上述 3 方面入手,扩大数据规模,覆盖多环境下的飞机遥感影像,并改进网络模型结构,以获得更好的检测效果。特征提取部分使用不同的深度卷积网络以获得更好的提取效果。用控制

变量的思想,对初始学习率进行微调,对比不同学习率下的检测准确率,探索模型初始参数对检测效果的影响^[19]。

4 结束语

本文详细介绍了深度卷积神经网络的一种变体—区域深度卷积神经网络,并以飞机、油罐的遥感影像为例,用 Faster-RCNN 区域卷积神经网络进行目标检测,提高了检测的准确率和置信度。本文使用的目标检测方法扩展方便,使用包含多种目标的数据集进行网络训练可使网络具备检测不同目标的能力,在遥感影像解译和快速信息获取等方面有着广阔的应用前景。本文使用的方法仍存在一些不足,如在不同光照、地形、天气等外部因素影响下,检测各种地物的遥感影像较为困难,在处理海量遥感数据时如何显著提升效率也仍待解决。随着遥感技术、芯片技术的发展,用于训练的数据集规模和训练所用的硬件设备计算力都会有较大提升,给相关研究带来了新的可能。

参考文献

- [1] 牛新, 窦勇, 张鹏, 等. 基于深度学习的光学遥感机场与飞行器目标识别技术[J]. 大数据, 2016, 2(5): 54-67.
- [2] CHENG G, HAN J. A survey on object detection in optical remote sensing images[J]. ISPRS Journal of Photogrammetry & Remote Sensing, 2016, 117: 11-28.
- [3] LECUN Y, BOSE B, DENKER J S, et al. Back-propagation applied to handwritten zip code recognition[J]. Neural Computation, 1989, 1(4): 541-551.
- [4] LIU Y J, LIU K Q. The fast convolution neural networks applied to license plate recognition[J]. Applied Mechanics & Materials, 2014, 562: 4802-4805.
- [5] MONIT F, BAROFFIO L, BONOI L, et al. Deep convolutional neural networks for pedestrian detection [J]. Signal Processing Image Communication, 2016, 47: 482-489.
- [6] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[C]// International Conference on Neural Information Processing Systems. Nevada: Curran Associates Inc, 2012: 1097-1105.
- [7] SMIRNOV E A, T IMOSHENKO D M, ANDRIANOV S N. Comparison of regularization methods for imagenet classification with deep convolution neural

- networks[J]. AASRI Procedia, 2014(6): 89-94.
- [8] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. Journal of Machine Learning Research, 2014, 15(1): 1929-1958.
- [9] GIRSHICK R, DONAHU 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. Columbia: IEEE, 2014: 580-587.
- [10] GIRSHICK R. Fast R-CNN[C]//Proceedings of the IEEE international conference on computer vision. Santiago:IEEE, 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]//Advances in Neural Information Processing Systems. Montreal:MIT Press,2015: 91-99.
- [12] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 37(9): 1904-1916.
- [13] LONG Y, GONG Y, XIAO Z, et al. Accurate object localization in remote sensing images based on convolutional neural networks[J]. IEEE Transactions on Geoscience & Remote Sensing, 2017, 55(5): 2486-2498.
- [14] XIAO Z, LIU Q, TANG G, et al. Elliptic Fourier transformation-based histograms of oriented gradients for rotationally invariant object detection in remote-sensing images[J]. International Journal of Remote Sensing, 2015, 36(2): 618-644.
- [15] 王一超, 韦建文. 基于高性能计算平台的 TensorFlow 应用探索与实践[J]. 实验室研究与探索, 2017, 36(12), 36: 125-128.
- [16] 宋倩, 黄昶, 余慧瑶. 基于 TensorFlow 的交通标志形状识别[J]. 信息通信, 2017(12): 286-288.
- [17] VIOLA P, JONES M. Robust real-time face detection[J]. International Journal of Computer Vision, 2004, 57(2): 137-154.
- [18] 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. Las Vegas:IEEE, 2016: 770-778.
- [19] 卢艺帆, 张松海. 基于卷积神经网络的光学遥感图像目标检测[J]. 中国科技论文, 2017, 12(14): 1583-1589, 1633.

(本文编辑:姚麒伟)