计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2018,27(11):155−160 [doi: 10.15888/j.cnki.csa.006615]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

卫星视频中目标的快速检测算法研究①

刘贵阳[1,2,3], 李盛阳[2,3], 邵雨阳[2,3]

[1](中国科学院大学, 北京 100049)
[2](中国科学院 太空应用重点实验室, 北京 100094)
[3](中国科学院 空间应用工程与技术中心, 北京 100094)
通讯作者: 李盛阳, E-mail: shyli@csu.ac.cn

摘　要: 随着视频卫星的不断发展, 如何在卫星视频数据中准确和快速地进行目标检测逐渐成为一个研究热点. 本文从两个方面改进了单阶段的目标检测网络. 针对卫星图像中目标尺寸小、分辨率低的特点, 利用反卷积操作丰富目标的上下文信息, 同时将对应尺度的卷积特征组合成超参特征, 丰富目标的细节特征; 并提出图像特征多级网格化, 将不同网格化的结果进行融合, 提高模型的检测准确率. 根据视频卫星对地凝视成像、场景移动缓慢的特点, 设计出内容一致性判别网络, 通过判别结果可以省略一些冗余的检测步骤, 提升整体的检测效率. 本实验使用“吉林一号”卫星视频数据, 通过具体的实验结果分析, 得出该系统对于对地凝视卫星视频中目标检测的准确率和速度都达到了较好的效果.

关键词: 视频卫星; 神经网络; 反卷积; 多级网格化; 内容一致性判别网络

引用格式: 刘贵阳,李盛阳,邵雨阳.卫星视频中目标的快速检测算法研究.计算机系统应用,2018,27(11):155–160. http://www.c-s-a.org.cn/1003-3254/6615.html

Fast Target Detection Algorithm in Satellite Video

LIU Gui-Yang[1,2,3], LI Sheng-Yang[2,3], SHAO Yu-Yang[2,3]

[1](University of Chinese Academy of Sciences, Beijing 100049, China)
[2](CAS Key Laboratory of Space Utilization, Chinese Academy of Sciences, Beijing 100094, China)
[3](Technology and Engineering Center for Space Utilization, Chinese Academy of Sciences, Beijing 100094, China)

Abstract: With the continuous development of video satellite technology, quick and accurate target detection in satellite video data has gradually become a research hotspot. This study improves the single-stage target detection framework from two aspects. In view of the features of small target size and low resolution in satellite images, the deconvolution operation is used to enrich the context information of the target, and the convolution features of the corresponding scales are combined into the super parameter features to enrich the details of the target. In addition, the image feature multilevel meshing is put forward, and the results of different meshes are fused to improve the detection accuracy of the model. According to the characteristics of satellite gaze imaging and the slow motion of the scene, we designed a content consistency discriminant network. Through the discriminant result, some redundant detection steps can be omitted to improve the overall detection efficiency. Through the concrete analysis of the experimental results of the “Jilin-1” Satellite, the accuracy and speed of target detection in satellite video were achieved by the detection system.

Key words: video satellite; neural networks; deconvolution; multilevel grid; content consistency discriminant network

① 基金项目: 遥感信息与图像分析技术国家级重点实验室基金 (Y8180711WN); 中国科学院国防科技创新基金 (Y6031511WY)
Foundation item: National Key Laboratory Foundation for Remote Sensing Information and Image Analysis Technology (Y8180711WN); National Defense Science and Technology Innovation Foundation of Chinese Academy of Sciences (Y6031511WY)
收稿时间: 2018-03-20; 修改时间: 2018-04-27; 采用时间: 2018-05-07; csa 在线出版时间: 2018-10-24

随着我国空间技术的不断发展, 在视频卫星技术领域取得了长足的进步. 在 2015 年 10 月, 我国第一套自主研发的商用视频卫星“吉林一号”成功发射, 其光学载荷对地成像的全色分辨率达到 0.72 米, 成为我国首颗米级高清动态视频卫星. 卫星获取的对地凝视数据可对工业、农业、交通等领域提供很多的基础应用[1]. 针对于卫星视频数据, 其中也存在着分辨率较低, 画面抖动, 图像周边畸变较为严重等问题[2]. 这些问题都对大场景中目标的快速检测带来较大的影响. 但同时针对视频相邻图像帧之间的信息的互补特征也给算法设计提供了很多新颖的思路.

随着大数据和云计算的不断发展, 深度学习技术得到了长足的发展, 视觉领域的机器学习模型不断挑战着人类在目标识别和目标检测领域的能力[3]. 在目标分类领域, 从 AlxeNet[4]、VGG[5]到 InceptionX[6]、ResNet[7], 模型的深度不断加深, 但是运算速度不断加快; 在目标检测领域, 从传统的滑动窗技术到二阶段检测框架 RCNN[8]、Fast RCNN[9]、Faster RCNN[10]、SPPNet[11]、RPN[12], 再到单阶段检测框架 YOLO[13]、SSD[14], 检测的准确率和检测效率也在不断的提高. 但是, 深度学习存在需要目标的样本量巨大, 且需要很大的计算量这两个问题, 而在视频卫星数据中缺少大量对于特定目标的标注数据, 要想利用深度学习技术解决视频卫星中目标检测问题问题, 目前必须另辟蹊径. Tianshu Yu 和 Ruisheng Wang 等在 2016 年使用图像数据和雷达点云数据, 同时辅助使用部分地理信息数据 (GIS) 通过改进图匹配技术较好的实现了街景数据中的场景解析[15], 虽然可以引入 GIS 信息进行遥感图像场景匹配之后在对感兴趣区域进行检测, 但融合 GIS 数据本身复杂度较高且场景中目标随机存在, 并不能较好的提高检测效率和准确率.

本文主要使用改进的 YOLO[13]模型和内容一致性检测模型, 利用视频数据中相邻帧之间内容的相似性降低精确检测的次数, 通过内容一致性判断确定当前帧是否需要再检测的思路, 在保证检测精度的前提下, 提高了视频数据的检测效率.

1 检测模型的改进

目前, 已有的目标检测模型分为两大类: 两阶段检测网络 Fast RCNN[9]、Faster RCNN[10]和单阶段检测网络 YOLO[13]、SSD[14]. 两阶段检测网络训练流程复杂, 且检测效率较低, 而单阶段网络检测网络虽然速度较快但是检测的准确率略低. 本论文主要改进 YOLO[13]模型, 在保证检测效率的前提下, 提高检测准确率.

原始的 YOLO[13]模型的设计主要是针对自然场景中的目标检测和识别任务, 自然场景中目标一般占据图像的主体, 且目标纹理较清晰, 因此模型在设计时卷积核一般选择成 7×7 和 3×3 大小, 并且经过不断的下采样将最后的卷积特征图固定为 7×7 这个尺寸, 将原始图像的尺寸 (448×448) 在宽和高的方向上同时各自缩小了 64 倍. 在纹理信息丰富的数据集中可以提高计算效率, 但是对于分辨率不高且经过压缩的卫星视频数据来说这会丢失大量目标的重要特征. 因此, 如何对图像做分割, 如何设计可以增强小目标特征信息的基础网络就变得至关重要.

1.1 多尺度网格分割策略

由于卫星图像具有低分辨率、视野广的特点, 图像中的目标会变的小且很模糊[1]. 若继续按照原始模型的网格划分, 会导致多个目标在相同的网格中, 导致模型检测任务失败.

对于现有的视频卫星数据, 先对图像中标注的目标的宽和高进行分别统计, 判断目标在尺度上的分布, 进而指导图像网格化策略: (1) 网格划分的粒度以不能同时包含任意两个或者多个目标为最佳; (2) 网格划分的粒度不宜过密, 否则会导致全联接层参数量爆炸, 模型难以学习; (3) 多尺度网格划分时, 不同尺度网格之间信息需要互补.

统计目标的宽和高的分布情况, 绘制分布散点图如下图所示. 根据图中可以得出, 目标宽和高的分布近似一致, 主体分布在 30～60 像素, 且目标飞机的宽和高的比例近似在 1:1.

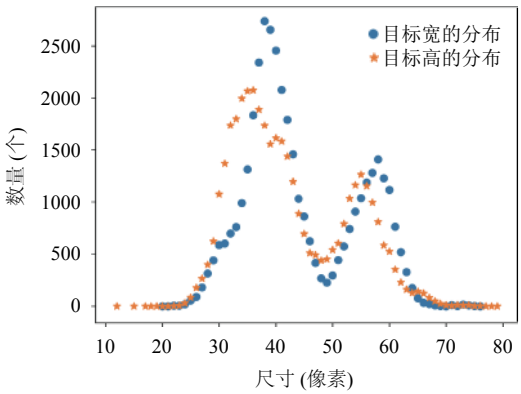
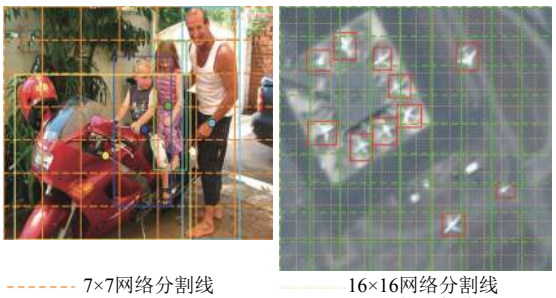


图 1 罗马·菲乌米奇诺机场图像飞机尺寸分布图

根据以上统计结果分析可以将图像 (512×512) 网格化为 13×13 大小, 但是图像中目标的分布是随机, 当

两个目标按照一定的角度相邻时, 13×13 的网格并不能很好的圈住某目标; 当目标恰好被某个网络恰好分割成 1/4 时由于 YOLO 模型的机制不能很顺利的检测到目标, 因此根据上述的指导规则 (2), 拟采用更小的尺度 16×16 对目标进行网格化. 对于 16×16 的网格化的特征图中每个单元的感受野较少, 图像的分辨率仅有 0.72 米, 因此, 再设计一层较大感受野的网格层, 使其可以较好的和 16×16 的细网格做检测的融合, 采用 9×9 网格去分割图像, 该设计也符合规则 (3), 将这两种网格化配合使用可以提高目标检测的准确率.



(a) 自然图像网络网络化示意图 (b) 小目标图像中多级网格化示意图

图 2 图像的网格化策略示意图

1.2 超参结构增强上下文信息

YOLO 目标检测模型反复使用卷积操作和最大池化操作将 448×448 的图像采样到 7×7 的特征图, 这种操作在卫星数据中并不能起到好作用, 批量的降采样导致目标的上下文信息损失的很厉害, 对于小目标的识别效果并不是很好[16], 且在视频数据中, 每一帧图像都是经过视频压缩编码处理之后, 单帧图像会变得很模糊, 更加不利于这种采样方式.

在 2015 年, 何凯明等人[7]提出残差网络模型, 不仅可以训练更深的网络模型, 还将底层卷积特征跨层传输到上层, 更加有利于目标的分类; 在 2016 年, 清华大学人工智能实验室提出 HyerNet 结构[12], 证明利用反卷积[17]构造出来的超参卷积层对小目标的检测有更好的效果; 在 2017 年, 康奈尔大学、清华大学、Facebook 人工智能研究院联合发表了 DenseNet 网络结构[18], 不仅有效的解决了梯度消失问题, 同时也强化了特征的传播和特征的重用.

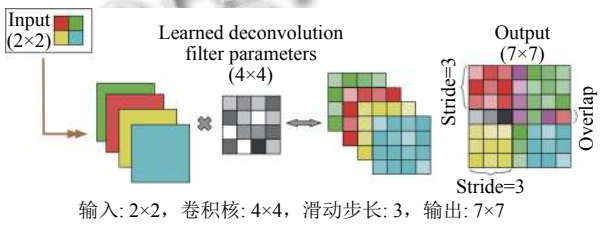


图 3 反卷积操作示意图

本文采用跨层连接和反卷积操作整合出不同尺度的超参特征立方体, 在不同尺度上强化小目标的细节特征, 具体的设计结构如图 4 所示. 其中网络的输入尺寸是 512×512, 经过卷积层和池化层之后将尺寸变换到 16×16, 在经过卷积操作和池化操作之后得到 4×4 的特征图, 之后对该层数据进行反卷积操作得到和 C5 相同尺寸 (9×9) 的特征图 DC5, 然后通过对这两层特征做连接后进行反卷积变化得到特征图 DC4, 同理得到特征图 DP3. 其中符号 P 的含义是该层特征数据来自池化操作, C 的含义是该层特征数据来自卷积操作, D 的含义是该层特征数据来自反卷积操作.

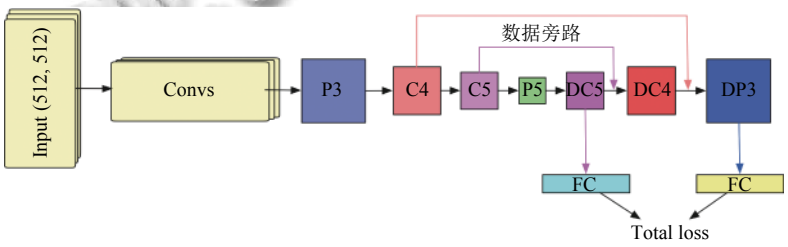


图 4 部分反卷积网络结构示意图

2 内容一致性判别网路

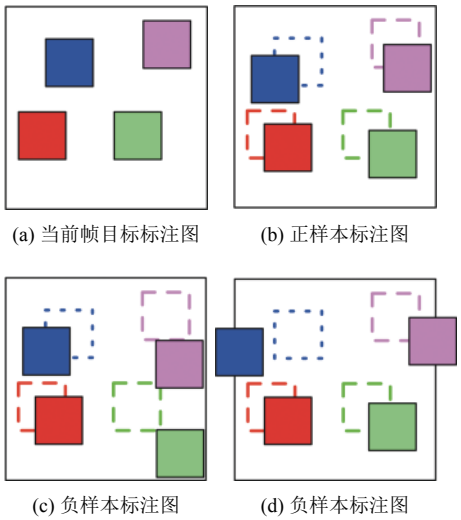
针对卫星视频数据而言, 视频场景很大, 虽然地物场景复杂, 但是整体内容的一致性较高, 场景中的目标运动相对在地面观察来说运动较慢[19]. 利用数据的这一特点, 本论文以孪生网络为主体, 设计出内容一致性检测网络, 跳过部分图像帧的检测, 提高检测的效率.

孪生网络 (Siamese Network) 是一种相似性度量网络, 核心思想是通过一组变换将输入数据映射到目标

空间, 在目标空间使用简单的距离度量进行相似度比较[20]. 算法主要关注连续图像序列中相邻的图像内容的差异, 如何定义内容差异程度, 如何对序列图像做标注等这些问题都会直接影响判别的结果. 主动拍摄视频数据大致分为固定镜头拍摄和运动镜头拍摄. 在固定镜头拍摄情况下, 往往更关注视野中内容的变化, 需要对相邻帧图像之间做判断; 在运动镜头拍摄情况下, 需要同时兼顾视野内场景的变化和场景内目标的变化, 这种变化往往来自场景的渐变和突变, 这时需要假定出关键帧图像, 不断的跟关键帧之间对比差异来判断内容的一致性. 本问题中卫星相对所拍摄的区域做凝视拍摄, 可以类比为摄像头相对目标场景静止, 但是镜头本身需要根据轨道卫星的状态做相对运动, 因此也会干扰到成像区域.

2.1 训练样本构造

实验对象是四段时间为 30 秒的“吉林一号”卫星视频数据[21], 视频中每帧图像的尺寸为 4096×3072, 将每一帧图像按照 8×6 的比例进行分割, 使得每个区域的尺寸为 512×512. 针对每段视频中相同的区域按照给定的规则进行标注: 当前帧图像中的每个目标和前一帧中对应的每个目标之间有位置的重合, 则判定这两帧图像的内容是一致的, 否则内容不一致. 图 5 中当前帧中有四个标注目标, 分别用四个色块进行表示, 图 (b), (c), (d) 分别表示下一帧图像内容, 实心色块表示目标在当前帧的位置, 不同颜色的虚线表示图 (a) 中相应目标的位置. 其中图 (a) 与 (b) 帧中内容标记为一致, 图 (a) 与 (c)、(a) 与 (d) 帧中内容均标记为不一致.



(a) 当前帧目标标注图 (b) 正样本标注图

(c) 负样本标注图 (d) 负样本标注图

图 5 内容一致判别标注示意图

2.2 网络结构设计

具体的网络结构设计如图 6, 其中卷积网络是共用参数, 将两帧图像输入到网络中, 得到相同维度的全连接层输出的特征, 根据标签信息去学习特征之间的相似性度量. 其中, 网络的主体部分的参数是共享的, 通过不同的输入图像, 得到在相同变换空间的特征, 再进行分类判别.

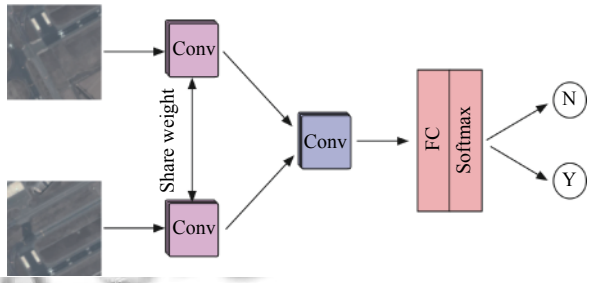


图 6 内容一致性判别网络结构示意图

3 系统检测流程

具备了对单帧图像的精细检测能力, 也具备了对于两幅图像内容一致性的判断能力, 通过这两个功能的配合使用, 本文设计出整体检测流程, 如图 7.

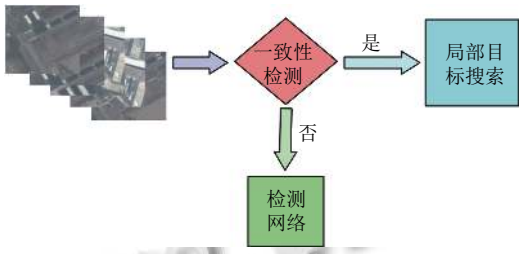


图 7 检测系统流程示意图

系统分为四大部分: 图像切割和图像增强、内容一致性判别网络、局部目标搜索模块、神经网络检测模块. 每帧图像按照 512×512 的尺寸进行切割, 得到 48 张小图像, 然后对这些图像进行增强变换, 镜像、裁减、形变、扩张这些操作按照一定的概率作用在每张图像上, 具体的流程如图 8 所示. 接着将当前帧图像和关键帧 (最近一次使用检测网络检测的图像) 进行内容一致性检测, 根据判断结果分别进行局部目标搜索和目标检测网络. 其中局部目标搜索, 主要以关键帧中目标的位置为核心, 按照检测出目标的尺度在周边进行查找.

4 实验结果分析

4.1 深度神经检测网络实验

本实验中共有四段标记数据, 分别是意大利·菲乌米奇诺机场、印度·英迪拉·甘地国际机场、美国·明尼

阿波利斯·圣保罗国际机场、突尼斯·迦太基国际机场. 我们使用后三段标注数据进行训练, 使用第一段视频作为测试, 在 Ubuntu 16.04 的 Linux 系统, 硬件信息为 Intel(R) Xeon(R) CPU E5-2620 v3 @ 2.40 GHz, NVIDIA GeForce GTX TITAN X GPU, 128 GB Memory 中进行 100 轮 (所有样本训练一次即为一轮) 训练. 对于飞机这一目标的检测的平均准确率是 0.912, 其中准确率和召回率的结果图如图 9 所示.

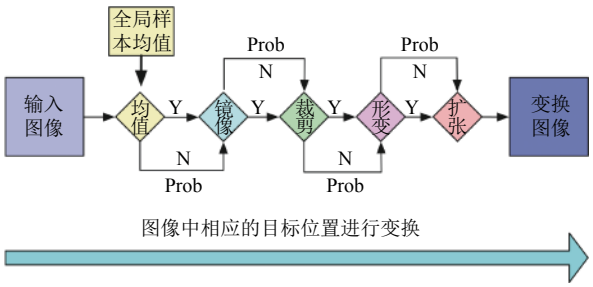
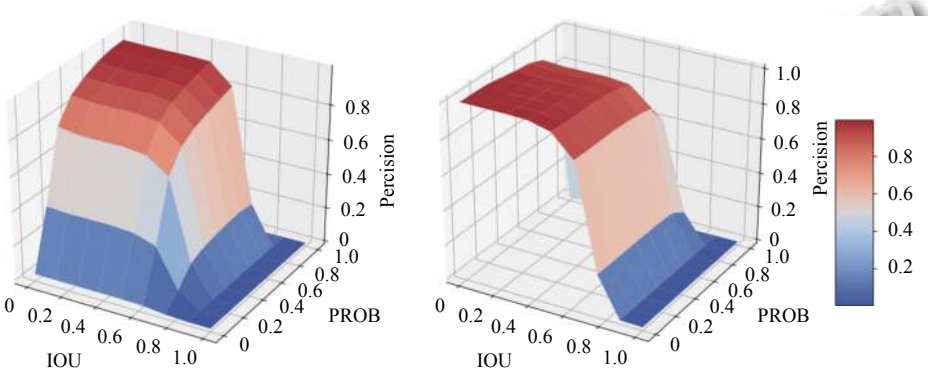


图 8 图像增强示意图



(a) 飞机检测的准确率 (b) 飞机检测的召回率

图 9 检测目标的准确率和召回率示意图

部分实验结果图如图 10 所示, 其中每张图像中的目标都进行了检测结果的标注, 每个框的左上方标记出该框的置信度大小. 红颜色的框的置信度至少在 0.8 以上, 绿颜色的框的置信度在 0.7～0.8, 蓝颜色的框在 0.6～0.7.

(a) 场景一的检测结果图 (b) 场景二的检测结果图

(c) 场景三的检测结果图 (d) 场景四的检测结果图

图 10 模型检测结果示意图

4.2 内容一致性判别实验

根据 2.1 中训练样本的构造规则, 预先生成相应的成对的训练样本, 设计了如下算法 1.

算法 1. 一致性判别网络训练样本构造算法

1) 先将视频中的单帧图像按照 512×512 的大小进行切割, 并记录切割出来的每个图像的相对原图像中的坐标和帧数;
2) 在相同的视频中, 随机抽取一帧作为关键帧, 找到前后对应的图像帧序列;
3) 记录出当前帧中某个切割区域的目标的位置, 同时按照目标的宽和高生成目标的合理潜在区域;
4) 将关键帧图像信息和其前后帧图像中对应的切割区域目标进行比较, 根据 2.1 节中的规则进行标记.

将上述算法生成的训练样本在相同配置的服务器中进行训练, 得到了较好的效果. 具体分析如下: 对地凝视的卫星视频数据中, 场景内容近似一致, 通过该模型, 可以较好的跳过相近内容的检测; 且场景中并非所有目标都在运动, 很多静止的目标可以很好的通过大图像切割和内容一致性检测策略直接确定相应的位置, 提高了整体的检测效率.

表中的数据的测试环境和训练环境一致 (在 TITAN X GPU, Tensorflow 1.4.1), 测试数据是将视频数据逐帧

提取 (每帧大小 4096×3072), 在内存中做 8×6 的切割后直接进行模型测试, 数据 IO 时间和大图像切分时间不在计算之内, 但是后处理时间计算在内. 通过表中对比可以看出内容判别网络将系统的整体检测效率提高了 50% 左右.

表 1 模型效率比较

模型	效率
改进版 YOLO	2.89 fps
改进版 YOLO+内容判别	4.14 fps

5 总结与展望

通过对检测模型的改进, 反卷积网络可以较好的对增强目标的上下文信息, 同时, 利用多个网格化的策略, 可以较好的弥补对于近邻目标检测的准确率. 结合内容一致性判别网络实验, 可以得出, 针对场景相对固定, 且场景内目标很多, 但运动目标一定的情况下, 通过该策略可以将局部静态信息过滤出去, 提高对视频整体的检测效率. 目前检测模型的设计仅考虑单帧图像, 对于视频而言, 如何考虑相邻帧信息, 将更多的信息整合到单阶段检测网络将成为下一步的研究热点.

参考文献

1 张过. 卫星视频处理与应用进展. 应用科学学报, 2016, 34(4): 361–370. [doi: 10.3969/j.issn.0255-8297.2016.04.001]

2 于渊博, 张涛, 郭立红, 等. 卫星视频运动目标检测算法. 液晶与显示, 2017, 32(2): 138–143.

3 周晓彦, 王珂, 李凌燕. 基于深度学习的目标检测算法综述. 电子测量技术, 2017, 40(11): 89–93. [doi: 10.3969/j.issn.1002-7300.2017.11.020]

4 Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Proceedings of the 25th International Conference on Neural Information Processing Systems. Lake Tahoe, NV, USA. 2012. 1097–1105.

5 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556, 2014.

6 Szegedy C, Liu W, Jia YQ, *et al*. Going deeper with convolutions. Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA. 2015. 1–9.

7 He KM, Zhang XY, Ren SW, *et al*. Deep residual learning for image recognition. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA. 2016. 770–778.

8 Chen CY, Liu MY, Tuzel O, *et al*. R-CNN for small object detection. In: Lai SH, Lepetit V, Nishino K, *et al*., eds. Computer Vision–ACCV 2016. Cham: Springer, 2016. 214–230.

9 Girshick R. Fast R-CNN. Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile. 2015. 1440–1448.

10 Ren SQ, He KM, Girshick R, *et al*. Faster R-CNN: Towards real-time object detection with region proposal networks. arXiv:1506.01497, 2015. 91–99.

11 He KM, Zhang XY, Ren SQ, *et al*. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904–1916. [doi: 10.1109/TPAMI.2015.2389824]

12 Kong T, Yao AB, Chen YR, *et al*. HyperNet: Towards accurate region proposal generation and joint object detection. arXiv:1604.00600, 2016. 845–853.

13 Redmon J, Divvala S, Girshick R, *et al*. You only look once: Unified, real-time object detection. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA. 2016. 779–788.

14 Liu W, Anguelov D, Erhan D, *et al*. SSD: Single shot multibox detector. In: Leibe B, Matas J, Sebe N, *et al*., eds. Computer Vision–ECCV 2016. Cham: Springer, 2016. 21–37.

15 Yu TS, Wang RS. Scene parsing using graph matching on street-view data. Computer Vision and Image Understanding, 2016, 145(C): 70–80.

16 Lei B, Wang B, Yue S. A fast detection method for small weak infrared target in complex background. Proceedings of Infrared, Millimeter-Wave, and Terahertz Technologies IV. Beijing, China. 2016, 30. 100301V.

17 Zeiler MD, Krishnan D, Taylor GW, *et al*. Deconvolutional networks. Proceedings of 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco, CA, USA. 2010. 2528–2535.

18 Huang G, Liu Z, Van Der Maaten L, *et al*. Densely connected convolutional networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA. 2017. 2261–2269.

19 满益云, 李海超. 低轨视频卫星成像特性分析. 航天器工程, 2015, 24(5): 52–57. [doi: 10.3969/j.issn.1673-8748.2015.05.008]

20 Bertinetto L, Valmadre J, Henriques JF, *et al*. Fully-convolutional siamese networks for object tracking. Hua G, Jégou H. Computer Vision–ECCV 2016 Workshops. Cham: Springer, 2016. 850–865.

21 徐伟, 金光, 王家骐. 吉林一号轻型高分辨率遥感卫星光学成像技术. 光学精密工程, 2017, 25(8): 1969–1978.