

卫星视频舰船目标检测方法

张 驰¹, 王 朝², 盛 辉¹

(1. 中国石油大学(华东)海洋与空间信息学院, 山东 青岛 266580; 2. 青岛市勘察测绘研究院, 山东 青岛 266000)

摘要: 舰船目标检测是进行海上目标监管, 保障海上权益的重要手段。本文在 SSD(single shot multibox detector)算法的基础上, 利用残差网络(ResNet, residual network)作为骨干网络构建 SSD 模型, 将改进后的 SSD 算法应用于卫星视频舰船目标检测, 该算法采用残差连接替换原本的级联方式, 加强前后特征之间的联系, 减少模型参数, 在保证检测精度的同时提高检测速度。在本文构建的数据集上进行实验, 结果表明, 改进后的 SSD 算法在测试集上的均值平均精度(mAP, mean average precision)为 93%, 比原始 SSD 算法提高了 5.31%, 充分证明了该方法对于提升 SSD 模型性能的有效性。使用“吉林一号”视频 03 星图像进行验证, 结果表明, 该算法能够较准确地检测到舰船目标, 可为海上复杂环境条件下的舰船实时检测提供参考。

关键词: 卫星视频; 目标检测; 舰船; 深度学习

中图分类号: TP79 文献标识码: A 文章编号: 1000-3096(2021)05-0009-07

DOI: 10.11759/hyxx20201108006

海面是重要的经济区域, 在军事和民用领域具有显著的地位。舰船是海上交通运输以及装备的主要载体, 对其进行快速和精确的检测在沿海监管和防御等领域具有重要意义^[1]。视频卫星作为一种新型的对地观测卫星, 与传统遥感卫星相比, 其最大的特点是可以对某一区域进行“凝视”观测, 以“视频录像”的方式实时获取动态信息, 并实时传输至用户, 特别适于监视动态目标, 分析目标位置变化特性^[2]。

传统视频目标检测方法主要有帧间差分法、背景建模法、光流法等。Kopsiaftis 等^[3]针对 Skysat-1 卫星视频, 利用背景差分法将当前帧与背景模板进行计算得到待检测目标, 检测完整度为 80%, 精度为 80%, 但是难以适应海量数据。张过等^[4]利用“吉林一号”卫星视频数据, 在经典算法随机邻域和区域匹配法的基础上, 结合光流法的运动矢量辅助并加以精化处理取得较好的效果, 但是光流法计算量较大, 很难应用到遥感卫星视频目标的实时检测中。

随着目标检测在计算机视觉领域的发展, 各类硬件水平不断提高, 基于深度学习的目标检测方法逐渐成为主流, 主要算法有 Faster RCNN^[5], YOLO^[6] (you only look once)和 SSD^[7] (single shot multibox detector)等。近年来, 将深度学习类目标检测算法用于遥感视频卫星的研究也越来越多, 例如, 张作省等^[8]结合遥感影像特点对 YOLOv2 网络结构进行改进, 首次将深度卷积神经网络应用于凝视视频卫星

目标检测, 填补了光学视频卫星成像领域目标检测方案的技术空白; 刘贵阳等^[9]使用改进的 YOLO 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, 在保证检测精度的前提下, 提高了视频数据的检测效率。将深度学习用于视频卫星目标检测主要面临两个问题: 一是需要大量的待检测目标样本, 二是计算量大导致其检测速度可能无法满足实时性的要求。

本文结合视频卫星数据的特点, 将目前在传统图像目标检测中性能较为优异的 SSD 算法用于卫星视频舰船目标检测, 并进行改进, 在本文构建的舰船数据集上进行实验并分析实验结果, 最后利用“吉林一号”卫星视频数据进行验证。结果表明, 改进后的 SSD 算法在保证准确度的同时, 提高了舰船目标检测效率。

1 方法

1.1 SSD 算法原理

1.1.1 模型结构

SSD 算法模型结构如图 1 所示, 输入图像大小为 300×300 , 采用 VGG16^[10]作为基础网络, 保留了

收稿日期: 2020-11-08; 修回日期: 2021-01-20

基金项目: 国家重点研发计划项目(2017YFC1405600)

[Foundation: National Key R & D Plan, No. 2017YFC1405600]

作者简介: 张驰(1997—), 男, 山东济宁人, 硕士研究生, 主要从事目标检测研究, E-mail: 535474915@qq.com

VGG16 的卷积层部分用于特征提取, 并将 VGG16 的两个全连接层转换为了普通的卷积层(Conv6 和 Conv7), 之后又加了多个卷积层(Conv8_1, Conv8_2, Conv9_1, Conv9_2, Conv10_1, Conv10_2), 最后用一个全局池化层来变成 1×1 的输出(Conv11_2)。

1.1.2 多尺度特征图

SSD 算法采用多尺度特征图进行检测, 选取 Conv4_3, Conv_7, Conv8_2, Conv9_2, Conv10_2,

Conv11_2 这些大小不同的卷积块作为预测特征层, 进行多尺度目标检测。较浅的特征层由于感受野较小, 主要用于检测小目标; 而较深的特征层感受野较大, 主要用于检测大目标。如图 2 所示, 对于每一个预测特征层采用不同个数的 3×3 大小的卷积核计算类别得分和边界框回归参数, 分别对应图 2 中的 conf 和 loc, 其中, $\Delta(c_x, c_y, w, h)$ 表示边界框的中心坐标、宽和高的变换量, (c_1, c_2, \dots, c_p) 表示各类别得分。

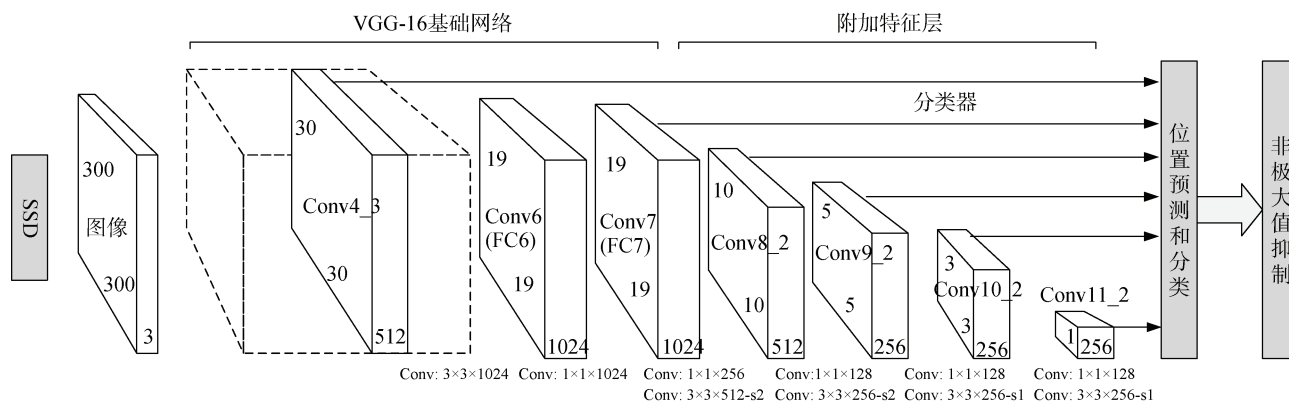


图 1 SSD 模型结构

Fig. 1 SSD model structure

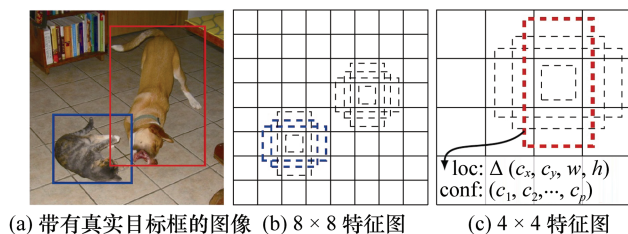


图 2 多尺度检测原理

Fig. 2 Principle of multiscale detection

1.1.3 先验框

SSD 算法参照 Faster RCNN 中的锚框机制, 在不同特征图的不同单元设置了尺度和长宽比不同的先验框, 预测的边界框是以这些先验框为基准的, 在一定程度上减少训练难度。一般情况下, 每个单元会设置多个先验框, 其尺度和长宽比存在差异, 如图 2 所示, 以每一个特征图的面而每一个单元作为中心, 使用了四种不同的先验框, 图片中猫和狗分别采用最适合它们形状的先验框来进行训练。假设对于特征图的每个位置预测 K 个边界框, 对于每一个边界框, 预测 C 个类别得分, 以及相对先验框中心坐标的 4 个偏移量, 则共需要 $(C+4) \times K$ 个分类器, 在 $m \times n$ 的特征图上面将会产生 $(C+4) \times K \times m \times n$ 个预测值。每

个特征层的尺度通过下面的公式计算得到:

$$S_k = S_{\min} + \frac{S_{\max} - S_{\min}}{m-1} (k-1), k \in [1, m], \quad (1)$$

其中, S_{\min} 表示最底层的尺度, 其值为 0.2; S_{\max} 表示最高层的尺度, 其值为 0.9; m 表示特征层数。

先验框的设置使用 5 种纵横比, 分别为 1, 2, 3, 1/2, 1/3, 另外, 当纵横比为 1 时, 先验框大小的计算公式为:

$$S'_k = \sqrt{S_k S_{k+1}}. \quad (2)$$

每个特征层先验框的尺度和比例设置如表 1 所示。利用不同尺度和比例的先验框, 可以基本覆盖各种形状大小的图像, 具有很好的适应性和选择性。

表 1 先验框大小设置

Tab. 1 Prior box size setting

特征层	尺度/ppi	比例	先验框数/个
Conv4_3	30	1, 2, 1/2	4
Conv7	60	1, 2, 3, 1/2, 1/3	6
Conv8_2	111	1, 2, 3, 1/2, 1/3	6
Conv9_2	162	1, 2, 3, 1/2, 1/3	6
Conv10_2	213	1, 2, 1/2	4
Conv11_2	264	1, 2, 1/2	4

1.1.4 损失函数

SSD 算法的损失函数包含两个部分，一个是类别损失，一个是定位损失。整个损失函数计算公式为：

$$L(x, c, l, g) = \frac{1}{N} [L_{\text{conf}}(x, c) + aL_{\text{loc}}(x, l, g)], \quad (3)$$

其中， N 是匹配到的正样本数量， c 是类别置信度预测值， l 是先验框对应的边界框预测值， g 是先验框的位置参数， x 代表网络的预测值。

类别损失计算公式为：

$$L_{\text{conf}}(x, c) = -\sum_{i \in \text{POS}} x_{ij}^p \log(\hat{c}_i^p) - \sum_{i \in \text{Neg}} \log(\hat{c}_i^0), \quad (4)$$

其中， \hat{c}_i^p 为预测的第 i 个默认框对应先验框的类别概率， x_{ij}^p 为第 i 个默认框匹配到的第 j 个先验框的概率。

定位损失计算公式为：

$$L_{\text{loc}}(x, l, g) = \sum_{i \in \text{POS}} \sum_{m \in \{cx, cy, w, h\}} x_{ij}^k \text{smooth}_{L1}(l_i^m - \hat{g}_j^m), \quad (5)$$

$$\hat{g}_j^{cx} = (g_j^{cx} - d_i^{cx}) / d_i^w, \quad (6)$$

$$\hat{g}_j^{cy} = (g_j^{cy} - d_i^{cy}) / d_i^h, \quad (7)$$

$$\hat{g}_j^w = \log\left(\frac{g_j^w}{d_i^w}\right), \quad (8)$$

$$\hat{g}_j^h = \log\left(\frac{g_j^h}{d_i^h}\right), \quad (9)$$

其中， l_i^m 为预测对应第 i 个正样本回归参数， \hat{g}_j^m 为正样本 i 匹配的第 j 个先验框的回归参数。

1.2 算法改进

SSD 算法是将 VGG16 作为骨干网络，并在其基础上修改并增加一系列卷积块构成的，VGG16 是一种深层的卷积神经网络，由 13 个卷积层和 3 个全连接层叠加而成，它使用较小的卷积核代替较大的卷积核来减少参数的同时通过增加网络的深度和宽度来提升性能。原始 SSD 模型的设计主要是针对自然场景中的目标检测任务，其在纹理信息丰富的数据集中可以提高计算效率，但是对于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。

ResNet^[11] 是一种残差网络，基本结构如图 3 所示，其采用残差连接替换原本的级联方式，增加前后特征层的联系，采用批归一化的方式并将激活函数设置为 ReLU(rectified linear unit)，在增加网络深度的同时减少模型的参数量，避免了因梯度爆炸或梯度消失带来的性能下降。

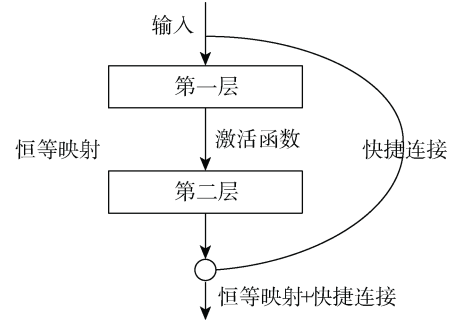


图 3 ResNet 基本结构

Fig. 3 ResNet basic structure

本文将原始 SSD 模型的骨干网络替换为 ResNet，构建 ResNet-SSD 模型，完整结构如图 4 所示，与原始 SSD 模型类似，ResNet-SSD 以 ResNet50(ResNet 中最具代表性的版本)作为特征提取网络，保留了 ResNet50 模型的前 4 个卷积层(图 4 中的 Conv1、Conv2_x、Conv3_x、Conv4_x)作为骨干网络，同时为了使输入图像经过 Conv4_x 后得到的预测特征层 1 尺度不变，我们将 Conv4_x 中的第一个残差结构中的步距全部修改为 1(图 4 中虚线指示部分)，增强对小目标特征信息的提取，并且去掉 ResNet50 卷积层后续的全连接层，增加额外的卷积层作为预测特征层，得到不同尺度的特征图。本文 ResNet-SSD 模型虽然增加了网络深度，但是由于使用了残差结构来减少参数量，模型精度提升的同时速度也有所提高。

2 实验

2.1 环境

实验在 Intel Xeon(R) CPU E5-2699、64 G 内存、Windows10 系统下搭建的 Pytorch 环境下进行，显卡为 Geforce GTX1080TI，实验环境配置参数如表 2 所示。

2.2 数据

目前，大部分用于目标检测的公开数据集主要是针对自然图像，遥感图像数据集较少，其中包含的舰船数量更少，因此，我们收集了来自不同卫星的高分辨率光学遥感图像，将其裁剪为 500×500 大小的图像，构建舰船数据集，其中包括 1963 张图像，共 2287 个舰船目标。为了提高模型训练精度，我们采用镜像、随机旋转、随机对比度变换等方法对数据集进行增强，将数据集扩充为 19630 张图像，使用 LabelImg 软件进行标注，得到完整的舰船检测数据集。在实验中，我们将数据集分为训练集、验证集和测试集，三者比例为 8 : 1 : 1。舰船样本如图 5 所示。

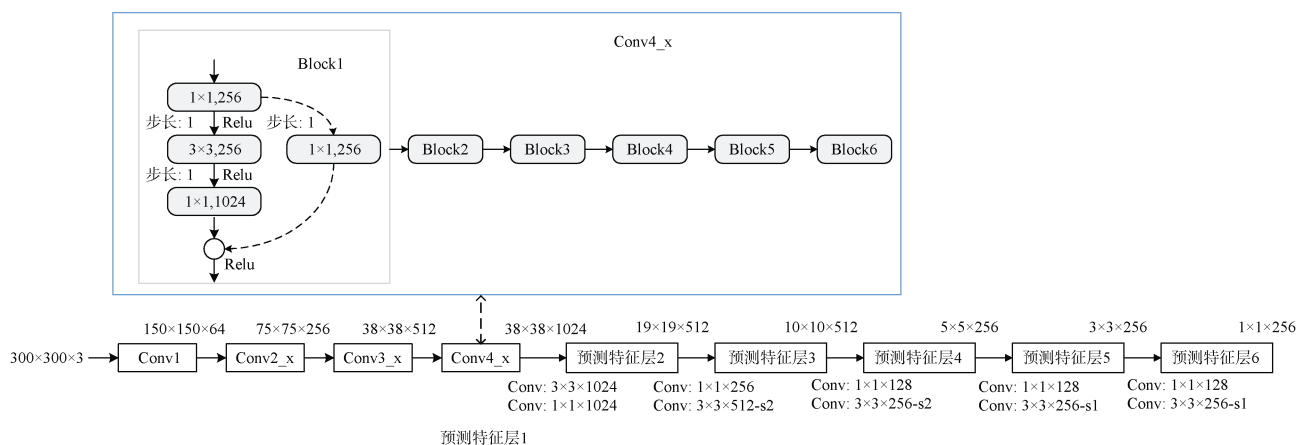


图 4 ResNet-SSD 模型结构
Fig. 4 ResNet-SSD model structure

表 2 实验环境配置参数

Tab. 2 The configuration parameters of experimental environment

配置名称	参数
操作系统	Windows10
CPU	Intel Xeon(R) CPU E5-2699
GPU	GTX1080TI/16 G
RAM	64 G
CUDA	10.2
cuDNN	7.6
Pytorch	1.5
Python	3.6



图 5 舰船样本
Fig. 5 Ship sample

本文卫星视频数据采用“吉林一号”卫星视频数据，由视频 03 星获取，分辨率为 0.92 m，时间为 30 s，共 300 帧，视频中每帧图像大小为 4 096 × 3 072，将每一帧图像分割为 500 × 500 大小。

2.3 结果

由于 SSD 模型中先验框参数的设置对检测精度会产生一定的影响，本文统计了所构建数据集中舰船

大小分布情况，如图 6，可以看出本文数据集中，舰船面积大多在 10² ppi 和 90² ppi 之间，因此原始 SSD 的参数大小并不适用于本文数据集，根据本文数据集特点，进行多组实验，最终取 S_{min} 为 0.02，S_{max} 为 0.4，计算得到各特征层先验框参数设置如表 3 所示。

表 3 先验框大小修改

Tab. 3 Prior box size modification

特征层	尺度/ppi	比例	先验框数/个
Conv4_3	3	1, 2, 1/2	4
Conv7	6	1, 2, 3, 1/2, 1/3	6
Conv8_2	34	1, 2, 3, 1/2, 1/3	6
Conv9_2	63	1, 2, 3, 1/2, 1/3	6
Conv10_2	91	1, 2, 1/2	4
Conv11_2	120	1, 2, 1/2	4

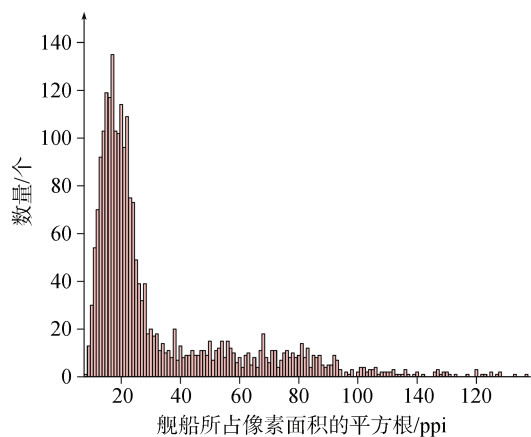


图 6 舰船大小分布直方图
Fig.6 Histogram of ship size distribution

为验证改进后 SSD 算法的有效性，本文在原始

SSD 模型的基础上, 利用 ResNet50 作为骨干网络构建 SSD 模型, 在本文构建的数据集上采用 SGD (stochastic gradient descent) 优化器对模型进行训练, 初始学习率设为 0.001, 采用动态学习率调整机制, 随着迭代次数增加, 学习率逐渐衰减直到模型收敛, 迭代最大次数设置为 100 000。利用训练得到的最终模型在测试集上进行测试, 并使用均值平均精度(mAP, mean average precision)评估模型性能, 得到结果如表 4 所示。

表 4 不同模型测试结果

Tab. 4 Test results of different models

检测算法	均值平均精度/%	时间/s
SSD	87.69	28
改进后的 SSD	93	21

2.4 分析

由表 4 可以看出, 改进后的 SSD 算法 mAP 值为 93%, 相比原始 SSD 算法提高了 5.31%, 并且运行时间加快了 7 s。这表明特征提取对于整体模型精度的影响较大, 本文使用的 ResNet50 特征提取网络由于使用跨层连接传递的方式将卷积处理后的特征与输入特征进行合并, 减少参数量的同时充分利用了各层特征信息, 将底层网络训练的更加充分, 使其网络深度增加, 精度也得到提高, 证明了残差网络的优异性能。

为测试改进后的 SSD 算法对卫星视频中舰船目标的检测能力, 利用“吉林一号”卫星视频数据进行实际舰船检测, 本文算法检测结果如图 7 所示, 图中绿色矩形框为检测出的舰船目标。为了综合评价算法的检测性能, 同时考虑虚警率和检测率, 本文利用舰船检测品质因数 F_{OM} 对检测结果进行定量分析:

$$F_{OM} = \frac{N_t}{N_f + N_g}, \quad (10)$$

其中, N_t 为舰船目标的正确检测数目, N_f 为虚警目标数目, N_g 为实际的舰船目标数目, 由上式知, F_{OM} 越大则检测性能越好。

由表 5 可知, 改进前后的 SSD 算法漏检目标数目分别为 15 和 10, 正确检测目标数目分别为 21 和 26, 说明对于大部分舰船目标, SSD 算法都能够准确地检测出来, 而漏检目标大部分为难分样本, 可以通过扩充数据集来减少漏检目标。改进前后的 SSD 算法虚警目标数目分别为 5 和 1, 说明残差网络能够很好的避免因梯度爆炸或梯度消失带来的性能下

降。综合来看, 改进后的 SSD 算法 F_{OM} 为 0.703, 比原始 SSD 算法提高了 0.191, 说明本文算法不仅能够很好地应用于卫星视频舰船目标检测, 并且检测效果比原始 SSD 算法更好。

表 5 不同模型检测结果

Tab. 5 Detection results of different models

检测算法	漏检目标	虚警目标	正确检测	F_{OM}
SSD	15	5	21	0.512
ResNet-SSD	10	1	26	0.703



图 7 本文算法舰船检测结果

Fig. 7 Satellite video detection results

3 结论

针对传统视频目标检测算法计算量大, 难以适应海量数据的问题, 本文将在传统图像目标检测中性能优异的 SSD 算法用于卫星视频舰船目标检测, 并在原始 SSD 模型基础上进行改进, 在本文构建的数据集上进行测试, 并利用“吉林一号”卫星视频数据进行验证, 主要结论如下。

1) 通过分析本文数据集中舰船目标大小分布, 进行多组实验确定最佳先验框的大小, 一定程度上提高了 SSD 算法对小目标的检测能力。

2) 为提高 SSD 算法对舰船目标的检测能力, 本文将残差网络与 SSD 算法相结合, 利用 ResNet 作为骨干网络构建 SSD 模型, 采用残差连接替代原本的级联方式, 在减少参数量的同时将深层特征与浅层特征结合, 相比原始的 SSD 算法, 在测试集上 mAP 提高了 5.31%, 速度上也有一定的提升。

3) 为验证改进后的 SSD 算法对卫星视频中舰

船目标的检测能力,利用“吉林一号”卫星视频数据进行舰船目标检测实验,分析检测过程中的漏检和误检目标,得到舰船检测品质因数为 0.703,总体检测效果较好,证明了改进后的 SSD 算法在卫星视频舰船目标检测中的有效性。但该算法对于训练过程要求较高,并且需要大量的数据集,后续还应进一步扩充数据集并对算法进行改进。

参考文献:

- [1] 唐沐恩,林挺强,文贡坚. 遥感图像中舰船检测方法综述[J]. 计算机应用研究, 2011, 28(1): 29-36.
TANG Muen, LIN Tingqiang, WEN Gongjian. Survey of ship detection methods in remote sensing images[J]. Research on Computer Application, 2011, 28(1): 29-36.
- [2] 孙志远,张刘,金光,等. 视频小卫星凝视姿态跟踪的仿真与实验[J]. 光学精密工程, 2011, 19(11): 2715-2723.
SUN Zhiyuan, ZHANG Liu, JIN Guang, et al. Simulation and experiment of gaze attitude tracking for video small satellite[J]. Optical Precision Engineering, 2011, 19(11): 2715-2723.
- [3] KOPSIAFTIS G, KARANTZALOS K. Vehicle detection and traffic density monitoring from very high resolution satellite video data[C]//Proceedings of the 2015 IEEE International Geoscience and Remote Sensing Symposium. Milan: IEEE, 2015, 1881-1884.
- [4] 张过. 卫星视频处理与应用进展[J]. 应用科学学报, 2016, 34(4): 361-370.
ZHANG Guo. Progress of satellite video processing and Application[J]. Journal of Applied Science, 2016, 34(4): 361-370.
- [5] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [6] 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 (CVPR 2015). Washington: IEEE Computer Society, 2016: 779-788.
- [7] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//Proceedings of the European Conference on Computer Vision. Berlin, Germany: Springer, 2016: 21-37.
- [8] 张作省,朱瑞飞. 凝视视频卫星目标检测算法[J]. 航天返回与遥感, 2018, 39(6): 102-111.
ZHANG Zuoxing, ZHU Ruifei. Satellite target detection algorithm based on staring video[J]. Space Return and Remote Sensing, 2018, 39(6): 102-111.
- [9] 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[J]. 计算机系统应用, 2018, 27(11): 155-160.
LIU Guiyang, LI Shengyang, SHAO Yuyang. Research on fast target detection algorithm in satellite video[J]. Application of Computer System, 2018, 27(11): 155-160.
- [10] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. Computer Science, 2014: 1409-1556.
- [11] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE Computer Society, 2016: 770-778.

Research on ship target detection based on satellite video

ZHANG Chi¹, WANG Zhao², SHENG Hui¹

(1. College of ocean and space information, China University of Petroleum (East China), Qingdao 266580, China; 2. Qingdao Geotechnical Investigation and Surveying Research Institute, Qingdao 266000, China)

Received: Nov. 8, 2020

Key words: satellite video; target detection; ship; deep learning

Abstract: Ship target detection is an important means to supervise and control maritime targets and protect maritime rights and interests. In this paper, based on SSD (single shot multibox detector) algorithm, using ResNet as the backbone network to build SSD model, the improved SSD algorithm is applied to satellite video ship target detection. The algorithm uses residual connection to replace the original cascade mode, strengthens the connection between the front and rear features, reduces model parameters, and improves the detection speed while ensuring the detection accuracy. The experimental results show that the map of the improved SSD algorithm on the test set is 93%, which is 5.31% higher than the original SSD algorithm, which fully proves the effectiveness of this method for improving the performance of SSD model. The results show that the algorithm can detect the ship target more accurately, which has a certain theoretical significance for real-time ship detection in complex marine environment.

(本文编辑: 丛培秀, 谭雪静)