计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.191100206

基于改进 YOLO-V2 算法的遥感图像目标检测技术研究

张 曼 李 杰 丁荣莉 成昊天 沈 雳
上海航天电子技术研究所 上海 201109

摘 要 传统遥感图像目标检测方法的时间复杂度高且精准率低,如何快速准确地检测遥感图像中的特定目标成为当前的研究热点。为解决这一问题,文中在 YOLO-V2 目标检测算法的基础上进行改进,减少了卷积层数与维度,并结合特征金字塔思想,增加了检测尺度,达到了提高检测精度的目的。同时给出了一种基于深度学习的遥感图像目标检测算法的通用处理框架,解决了无法直接处理大幅遥感图像的问题。在 DOTA 数据集上进行对比实验,结果表明改进 YOLO-V2 算法在 15 个类别上的精准率和召回率均优于 YOLO-V2 算法,mAP 值提高了 0.12。在时间复杂度方面,所提方法略低于 YOLO-V2 算法;在大小为 416×416 的图像小块上,改进 YOLO-V2 算法相比 YOLO-V2 检测时间缩短了 0.1ms。

关键词:深度学习;遥感图像;目标检测;YOLO-V2;特征金字塔;多尺度
中图法分类号 TP751

Remote Sensing Image Object Detection Technology Based on Improved YOLO-V2 Algorithm

ZHANG Man,LI Jie,DING Rong-li,CHENG Hao-tian and SHEN Ji
Shanghai Aerospace Electronics Technology Research Institute,Shanghai 201109,China

Abstract Traditional method of remote sensing image object detection has the disadvantages of high time complexity and low precision. How to detect specific targets in remote sensing images quickly and accurately has become a hot research topic. In order to solve this problem,this paper improves the YOLO-V2 object detection algorithm,reduces the convolution layers and dimension,and combined with the ideal of feature pyramid to increase the detection features' scale,so as to achieve the purpose of improving detection accuracy. At the same time,a general processing framework of remote sensing image object detection algorithm based on deep learning is presented to solve the problem that large remote sensing images cannot be directly processed. Comparison experiments on the DOTA dataset show that the improved YOLO-V2 algorithm has better accuracy and recall rate in 15 categories than the YOLO-V2 algorithm,and the mAP value is increased by 0.12. In terms of time complexity,it is slightly lower than the YOLO-V2 algorithm. Specifically,on 416×416 image patches,the detection time of the improved YOLO-V2 algorithm is reduced by 0.1 ms compared to the YOLO-V2 algorithm.

Keywords Deep learning,Remote sensing image,Object detection,YOLO-V2,Feature pyramid,Multiscale

1 引言

高分辨率光学遥感图像中特定目标检测一直是遥感图像分析领域的研究热点。飞机、舰船、车辆、油罐等目标的识别定位,在军事与民事应用中有着重要的作用。传统遥感图像目标检测算法大多是基于手工特征[1]构建的。首先根据滑动窗口对候选区域进行分块处理,然后根据不同目标的不同几何或纹理特征进行定位。该类型算法主要有两个缺陷:1)基于滑动窗口的区域选择策略没有针对性,窗口冗余,时间复杂度高;2)基于背景知识设计的手工特征对多样性的环境变化没有很好的鲁棒性。近年来,深度学习技术发展迅速。在目标检测方面,2014 年 Girshick 等提出了 RCNN[2]算法,其利用选择性搜索法[3]进行图像分割,提取分割图像特征,最后进行分类与边框回归。由于分割图像具有重复区域,导致特征重复提取,影响算法运算效率。。针对这种情况,2015 年 Girshick 又提出了 Fast-RCNN[4]算法,该算法为 RCNN 的改进版,其将分类与边框回归通过多任务损失函数结合起来,提高了算法的运算速度。同年,Faster-RCNN[5]算法被提出,该算法通过 RPN 网络选取候选区域,将候选区域选取融合到网络中,提高了运算速度。基于区域建议框的 R-CNN 系列算法生成建议框重叠,网络结构复杂,导致算法运算速度慢,不易于工程的实现。2016 年,一种端对端的目标检测算法 YOLO[6]被提出,该算法将全图划分为格子,每个格子负责预测中心位于该格子的目标,解决建议框重叠的问题,实现了实时目标检测。2017 年 Redmod 等提出了 YOLO-V2[7]算法,在 YOLO 的基础上进行多方面改进,提高了定位准确率。

本文提出的遥感图像目标检测算法是基于 YOLO-V2 算法进行改进的,结合特征金字塔(FPN)[8]的思想,增加一种尺度进行检测,提高了目标定位精度。同时,本文还提供了一种基于深度学习的遥感图像目标检测算法的通用处理框架,解决了大幅遥感图像无法直接处理的问题。

2 YOLO-V2 算法

YOLO-V2 算法采用 Darknet-19 作为基础网络,图 1 为 YOLO-V2 算法的网络结构图,图中的虚线边框部分为基础网络部分。Darknet-19 网络结构类似 VGG-16[9],使用 3×3

通信作者:张曼(manzhang_sh@163.com)

卷积核大小，池化操作后，卷积层通道数翻倍，同时借鉴 NIN[10] 思想，将大小为 1×1 的卷积核置于大小为 3×3 的卷积核之间，压缩特征。每层卷积层后使用 BatchNormal[11] 算法加速训练进程，加快收敛。YOLO-V2 算法中 passthrough 层将 512 维 26×26 大小的特征转为 256 维 13×13 大小的特征，之后与 1024 维 13×13 大小的特征融合，低维特征传递给高维特征，类似 ResNet[12] 中残差的概念，相对 YOLO 算法目标定位的效果更好。

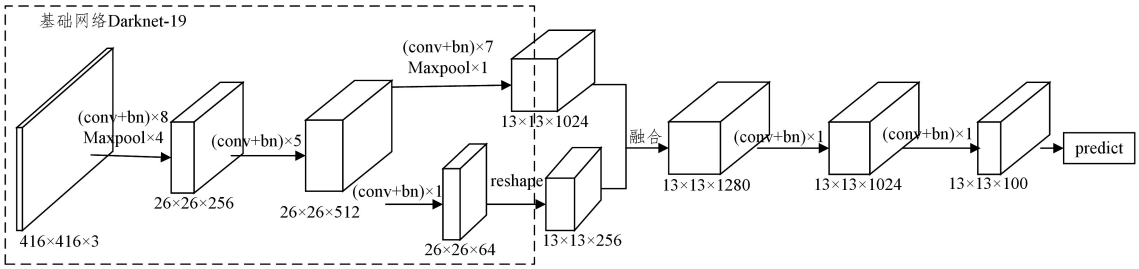


图 1　YOLO-V2 网络结构

Fig. 1　YOLO-V2 network structure

YOLO-V2 算法输入大小为 416×416 的图片，经过 5 次下采样，得到 13×13 大小的特征图，最后输出维度为：

$$N=S\times S\times B\times(5+C) \tag{1}$$

其中，S 表示最后特征图的大小，此处为 13。B 为一个栅格预测建议框的个数，此处根据 K-means[13] 聚类算法算得 5 个建议框时效果更佳。C 为预测类别个数，5 代表每个建议框预测的坐标信息和置信度包含 5 个信息。

YOLO-V2 算法在训练时采用多尺度训练方式，网络为全卷积结构，可动态调整输入图像的尺寸。在训练过程中使用 32 的倍数调整输入图像的尺寸，训练图像的尺寸在 320×320 到 608×608 之间随机变化。提高检测算法对图像大小的鲁棒性。

3　改进 YOLO-V2 遥感图像目标检测算法

3.1　算法处理框架

基于深度学习的目标检测算法无法直接处理大幅遥感图像，需要将图像裁剪后进行检测。图 2 给出了本文算法的处理框架。在训练阶段，首先将遥感图像裁剪为一定大小的图像小块，大幅遥感图像标签也随之变化；然后将裁剪后的图像小块与标签送入设计好的深度学习网络中进行训练；在训练时，将正确标签信息与预测标签信息的差值作为误差，迭代优化网络参数，最后保存训练好的网络模型。在检测阶段，同样先将测试遥感图像裁剪为小块图像；之后送入在训练阶段训练好的网络模型中，经过 NMS[14] 后处理算法消除重复定位窗口，得到小块图像的检测结果；最后将结果整合，得到原始大幅遥感图像的检测结果。

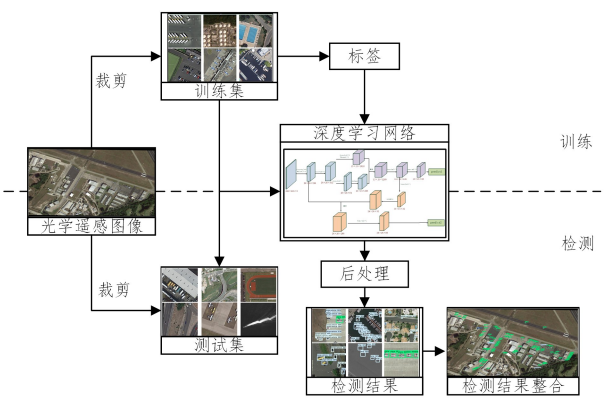


图 2　算法的处理框架

Fig. 2　Algorithm processing framework

本文算法处理框架的重点在于遥感图像和标签的裁剪与整合。遥感图像裁剪通过滑动窗口裁剪制定大小的图像小块，本文裁剪为 416×416 大小。图 3 为遥感图像裁剪示意图。上面的图为原始遥感图像，下面两张图为裁剪后的图像小块。一般相邻图像小块保持 15% 的重叠[15]，其目的是保证原图每个区域都能被完整检测到。重叠区域的重复检测可以通过 NMS 算法消除。这种有重合区域的裁剪方式可有效避免直接裁剪时边缘截断导致检测不出目标的情况。

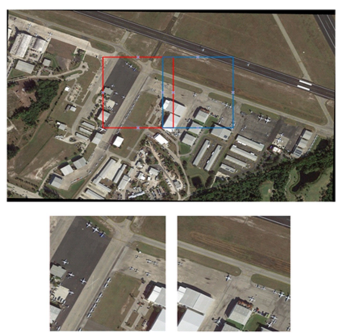


图 3　遥感图像裁剪示意图

Fig. 3　Cropping diagram of remote sensing image

在训练阶段，遥感图像的标签在裁剪过程中也会随之改变。原始遥感图像的标签在裁剪结束后，生成了对应图像小块数量的新标签。假设原始遥感图像的标签为 ImageName.txt，裁剪后小块遥感图像的标签名为：

ImageName_row_col_height_width.txt

其中，(col，row) 为图像小块左上角位于原始图像的坐标，height 和 width 为图像小块的高和宽。原始建议框坐标减去 (col，row) 得到新标签中的建议框坐标。

检测阶段的整合实现了图像小块检测标签的整合，将多张图像小块检测标签整合为原始遥感图像检测标签。整合时，首先将检测小块检测到的建议框坐标加上 (col，row)，得到该建议框原本位于原始图像的位置信息；之后，利用 NMS 算法消除重复定位建议框，保留置信度最高建议框；最后显示整合后的标签信息。

3.2　多尺度网络结构

在深度学习网络中，高层卷积层由于卷积池化操作，获得了高维特征，丢失了细节信息。低层卷积层相对保留了更详细的信息。特征金字塔网络结构通过将不同层次的特征融合，提高了目标检测的精准率。图 4 给出了特征金字塔网络

结构。图 4 中，左边为 FPN 自下而上的卷积池化过程，特征图经过下采样操作慢慢变小；右边为 FPN 自上而下的过程，特征图经过上采样变大。将左右相同大小的特征图融合以进行预测。FPN 网络这种将低分辨率、高语义信息的高层特征与高分辨率、低语义信息的低层特征自上而下进行侧边连接的方式，得到了所有尺度特征的丰富语义信息，可提高检测精度。

本文提出的基于 YOLO-V2 算法的改进多尺度网络结构，参考了 FPN 网络的思想。图 5 为本文提出的多尺度网络结构图。输入大小为 416×416 的三通道图像小块，经过 13 个 3×3 或 1×1 大小卷积层和 4 个最大池化层后变为 512 维 26×26 大小的特征图。之后经过类似 ResNet 网络单元结构的 passthrough 层，融合 13×13 大小的特征图和 26×26 大小的特征图得到 1 280 维 13×13 大小的特征图。相对 YOLO-V2 的 passthrough 层，支链减少了一个卷积层，提高了运算速度。接着经过 2 个卷积层后，进行 13×13 尺度的预测。另一方面，上阶段倒数第二层 13×13×128 大小的卷积层在经过上采样后，与上阶段 256 维 26×26 大小的特征图进行融合，接着通过 2 个卷积层提取信息，最后进行 26×26 尺度的目标预测。

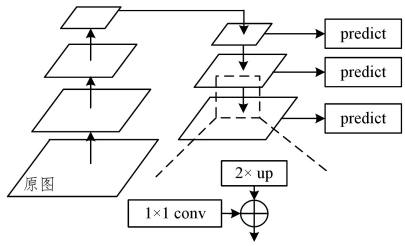


图 4 特征金字塔网络结构

Fig. 4 Diagram of feature pyramid network structure

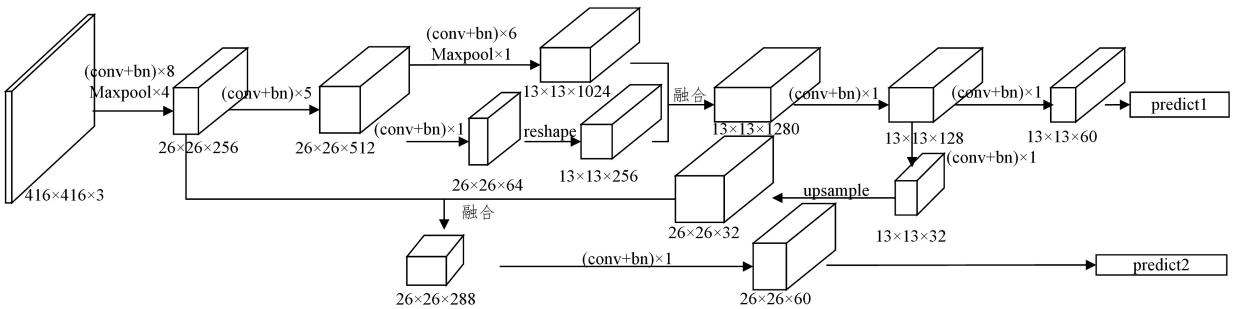


图 5 多尺度网络结构

Fig. 5 Diagram of multi-scale network structure

本文中多尺度的目标检测网络结构采用 13×13 和 26×26 两个尺度进行目标检测，每个尺度预测 3 个建议框，2 个尺度预测 6 个建议框。相对 YOLO-V2 算法的 5 个建议框，本文算法的预测效果更佳。

4 实验结果与分析

4.1 数据集

本文的实验图像来自网上开源航天遥感图像 DOTA[16] 数据集，共包含 1 869 张航空图像。大部分图像来自 GoogleEarch 和高分二号卫星。遥感图片的尺寸约为 4 000×4 000，包含飞机、舰船、卡车、小车、油罐等 15 个类别。裁剪为 416×416 大小后，得到 40 151 张图像小块，即数据集。图 6 给出了数据集示例。

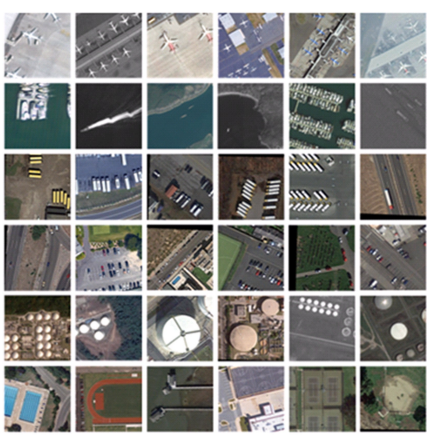


图 6 数据集示例

Fig. 6 Dataset examples

4.2 训练参数

训练时每批次处理 32 张图像，采用 Momentum[17] 梯度下降优化算法，初始学习率设为 0.001，迭代次数设置为 200 000。训练平台为 NVIDIA1080Ti×2。

4.3 检测结果定量评估

精准率(Presion)、召回率(Recall)和平均精准率(Average Precision，AP)为基于深度学习的图像目标检测算法的主要评价标准。假设 TF 表示预测为正向，实际预测正确的目标框数量；FP 表示预测为正向，实际预测错误的目标框数量；FN 表示预测为负向，实际预测错误的目标框数量；TN 表示预测为负向，实际预测正确的目标框数量。

精准率为正确检测为某类别数量与所有检测为该类别数量之比，其具体表达式如式(3)所示：

$$precision=\frac{TP}{TP+FP} \tag{2}$$

召回率为正确检测为某类别数量与所有标注为该类别数量之比，具体表达式如式(4)所示：

$$recall=\frac{TP}{TP+FN} \tag{3}$$

平均精准率是精确率 p 随召回率 r 变化的曲线在 r 从 0 到 1 上的积分，其表达式如式(5)所示：

$$AP=\int_0^1 p(r)\mathrm{d}r \tag{4}$$

表 1 列出了 YOLO-V2 算法与改进 YOLO-V2 算法的性能评估结果。可以发现，在 15 个类别上改进 YOLO-V2 算法相对 YOLO-V2 算法有更高的精准率、召回率和 AP 值。mAP 为 AP 的平均值，由 0.630 提高到 0.750，增长了 0.12，验证了改进 YOLO-V2 算法的有效性。

表 1　目标检测算法性能评估结果的对比

Table 1　Performance evaluation results comparison of object detection algorithm

类别	精准率		召回率		AP 值	
	YOLO-V2	改进 YOLO-V2	YOLO-V2	改进 YOLO-V2	YOLO-V2	改进 YOLO-V2
plane	0.928	0.947	0.775	0.995	0.689	0.844
baseball-diamond	0.853	0.900	0.7190.	0.989	0.620	0.819
bridge	0.879	0.899	0.833	0.959	0.719	0.817
ground-track-field	0.506	0.658	0.660	0.980	0.323	0.598
small-vehicle	0.886	0.886	0.911	0.983	0.804	0.805
large-vehicle	0.903	0.983	0.795	0.963	0.657	0.812
ship	0.916	0.915	0.873	0.974	0.750	0.832
tennis-court	0.872	0.914	0.776	0.993	0.665	0.792
basketball-court	0.855	0.870	0.703	0.985	0.622	0.791
storage-tank	0.870	0.897	0.908	0.988	0.791	0.816
soccer-ball-field	0.349	0.457	0.510	0.996	0.191	0.416
roundabout	0.795	0.820	0.827	0.973	0.651	0.745
harbor	0.861	0.880	0.736	0.989	0.639	0.782
swimming-pool	0.893	0.915	0.635	0.640	0.582	0.568
helicopter	0.889	0.902	0.871	0.994	0.738	0.809
mAP	—	—	—	—	0.630	0.750

表 2 列出了 YOLO-V2 算法与改进 YOLO-V2 算法在不同尺寸遥感图像上的检测时间结果。在 416×416 大小的图像小块与 4000×4000 大小的遥感图像上，改进 YOLO-V2 算法的检测时间略低于 YOLO-V2 算法。

表 2　目标检测算法的检测效率对比

Table 2　Detection efficiency comparison of object detection algorithm

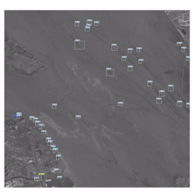
算法	遥感图像尺寸	
	416×416	4000×4000
YOLO-V2	10.17 ms	1.46 s
改进 YOLO-V2	10.26 ms	1.48 s

4.4　检测结果定性评估

图 7 给出了改进 YOLO-V2 算法在原始遥感图像上的检测结果。图 7(a)主要为飞机目标，图 7(b)主要为车辆目标，图 7(c)主要为舰船目标。可以发现，改进 YOLO-V2 算法，检测效果良好。

(a)飞机目标检测结果

(b)车辆目标检测结果



(c)舰船目标检测结果

图 7　原始遥感图像的检测结果

Fig. 7　Detection results of raw remote sensing images

图 8 给出了 YOLO-V2 算法和改进 YOLO-V2 算法在图像小块上的检测结果。可以发现，YOLO-V2 算法的漏检目标较多，改进 YOLO-V2 算法基本不漏检。在飞机、舰船、车辆等目标中，改进 YOLO-V2 算法的检测效果更好。

图 8　图像小块检测结果

Fig. 8　Detection results of image patches

结束语 本文给出了一种基于深度学习的遥感图像目标检测算法的通用处理框架，解决了大幅遥感图像无法直接处理的问题，并对遥感图像和标签的裁剪、整合方法进行了详细叙述。针对 YOLO-V2 算法检测精度低的问题，本文提供了一种改进 YOLO-V2 算法，结合特征金字塔思想，增加了 26×26 尺度进行检测，提高了目标定位精度。在开源遥感图像 DOTA 数据集上验证了该算法的有效性，在飞机、检测、车辆等目标图像中均有良好的检测效果，mAP 值达到了 0.75，相对 YOLO-V2 算法提高了 0.12。本文算法仍存在一些不足，由于不同类别训练数据不均衡，数量少的类别检测结果较差。在实时处理大幅遥感图像方面，检测效率的提升也是亟待解决的问题。近年来，深度学习技术发展迅速，可用于生成数据的生成对抗网络，目标检测模型的量化、剪枝等优化方法，都给相关研究带来了新的可能。

参 考 文 献

[1] CAO Q, ZHENG H, LI X S. A Cloud Detection Method for Satellite Remote Sensing Image Based on Texture Features[J]. Journal of Aeronautics, 2007, 28(3): 661-666.

[2] GIRSHICK R, DONAHUE 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. 2014: 580-587.

[3] UIJLINGS J R R, VAN DE SANDE K E A, GEVERS T, et al. Selective search for object recognition[J]. International Journal of Computer Vision, 2013, 104(2): 154-171.

[4] GIRSHICK R. Fast r-cnn[J]. arXiv: 1504.08083, 2015.

[5] 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. 2015: 91-99.

[6] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detect[C]// Las Vegas: Proceeding of the IEEE Conference on Computer Vision and Pattern Recogniton. IEEE, 2016.

[7] REDMOD J, FARHADI A. YOLO9000: Better, Faster, Stronger [C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu. IEEE, 2017: 6517-6525.

[8] SEFERBEKOV S S, LGLOVIKOV V I, et al. Feature Pyramid Network for Multi-Class land Segmentation[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. IEEE, 2018.

[9] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv: 1409.1556, 2014.

[10] LIN M, CHEN Q, YAN S. Network in network [J]. arXiv: 1312.4400, 2013.

[11] LOFE S, SZEGEDY C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift[C]// International Conference on Machine Learning. 2015.

[12] 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. 2016: 770-778.

[13] ZHANG S J, ZHAO H C. Algorithm research of optimal cluster number and initial cluster center[J]. Application Research of Computers, 2017, 34(6): 1617-1620.

[14] ROTHE R, GUILLAUMIN M, VAN GOOL L. Non-maximum suppression for object detection by passing messages between windows[C]// Asian Conference on Computer Vision. Springer, Cham, 2014: 290-306.

[15] VAN ETTEN A. You Only Look Twice: Rapid Multi-Scale Object Detection in Satellite Imagery[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2018.

[16] XIA G S, BAI X, et al. DOTA: A Large-scale Dataset for Object Detection in Aerial Images[C]// IEEE Conference on Computer Vision and Pattern Recogniton. 2018.

[17] QIAN N. On the momentum term in gradient descent learning algorithms[J]. Neural Networks, 1999, 12(1): 145-151.

ZHANG Man, born in 1993, master's degree. Her main research interests include image processing and remote sensing application.