

一种应用于高分辨率遥感图像目标检测的 尺度自适应卷积神经网络

吴佳祥¹, 刘 辉², 贺光辉¹

(1 上海交通大学 电子信息与电气工程学院, 上海 200240; 2 上海卫星工程研究所, 上海 201109)

摘 要: 为了解决高分辨率遥感图像中的目标检测问题并提高检测准确率, 本文提出了一种基于卷积神经网络的检测方法. 该方法采用 Faster R-CNN 作为遥感图像目标检测的基础框架, 在此基础之上, 针对遥感图像中不同尺度物体的准确率差异较大这一问题, 提出了一种尺度自适应卷积设计网络. 经过在遥感图像数据集上的测试, 该方法与已有的方法相比, 大大提高了高分辨率遥感图像中目标检测的准确率.

关键词: 卷积神经网络; 多目标检测; 遥感图像

中图分类号: TP39

文献标识码: A

文章编号: 1000—7180(2018)08—0078—04

A Scale Adaptive Convolutional Neural Network for Object Detection in Remote Sensing Images

WU Jia-xiang¹, LIU Hui², HE Guang-hui¹

(1 School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China; 2 Shanghai Satellite Technology Institute, Shanghai 201109, China)

Abstract: A new approach based on convolutional neural network (CNN) is proposed to improve the accuracy of multi-object in remote sensing images. Faster region convolutional neural network has been employed as the basic framework. Moreover, a scale adaptive convolutional neural network (SA-CNN) is proposed to deal with the multi-scale target detection in remote sensing images. The comparative experimental results show that the proposed SA-CNN significantly improves the accuracy of multi-object detection.

Key words: CNN; multi-object detection; remote sensing images

1 引言

随着遥感技术的发展, 高分辨率遥感图像逐渐在很多领域中发挥出重要的作用, 如农作物种植区域调查, 森林火灾监测, 车辆监测以及导航系统等. 从这些遥感图像中获得有用信息已经逐渐成为遥感图像领域中的重要研究方向^[1].

现今, 卷积神经网络 (Convolutional Neural Network, CNN) 在计算机视觉任务中发挥着重要的作用, 如物体分类, 目标检测, 图像分割等. 通过大量的训练数据, 卷积神经网络可以学习到图像的深层

特征, 利用这些特征可以很好地对图像进行分类. 近年来, 相关论文提出了几种基于卷积神经网络的高分辨率遥感图像目标检测方法, 其中部分方法是针对特定目标比如船只, 车辆等, 还有部分方法是针对多目标进行检测. Z. Zou 利用奇异值分解网络实现了一种快速有效的卫星图像船只检测的方法^[2];

X. Chen 提出了一种混合卷积神经网络通过提取多种尺度特征来检测车辆, 大大提高检测准确率. 然而这些工作只是针对单一的目标类别进行检测, G. Chen 提出一种 Rotation Invariant CNN (RIC-NN) 用来检测遥感图像中多类别物体^[3], 然而, 由于

不同类别的物体尺度差异较大,该模型在有些类别物体,尤其是小物体的检测准确率并不是很高.针对这个问题,本文提出了一种有效的能够检测多类别的遥感图像目标检测方法.

2 基于 CNN 的目标检测方法

目标检测是与图像分类不同的任务,主要目的有两个,一个是识别一张图片中的各种物体,另一个是预测出各类物体的坐标(一般是能够覆盖物体区域的最小矩形). Region with Convolutional Neural Network(R-CNN)是第一个成功将 CNN 应用在目标检测任务中的方法^[4],针对 R-CNN 速度慢,效率低等问题,Fast R-CNN^[5]采用了 ROI Pooling 层, Fast R-CNN 可以处理多批次 ROI 而且避免了因为缩放带来的原图信息损失.然而, Fast R-CNN 的速度仍然不够快,没有形成一个完整的端到端的检测框架, Faster R-CNN 的出现解决了这个问题^[6],所以速度与效率相比 Fast R-CNN 提升很大.

除了上述的 R-CNN 系列之外,还有一些其他的目标检测方法,但与 Faster R-CNN 相比仍有差距.考虑到本文的目标是实现遥感图像目标的精确检测,准确度是最重要的指标,所以 Faster R-CNN 被采用为基础框架.

3 Faster R-CNN 检测方法

Faster R-CNN 是 S. Ren 于 2015 年提出的目标检测算法及框架,该算法在 PASCAL VOC 2007 测试数据集可以达到 78.8% 的检测准确率.如图 1 所示,该框架主要包括两个部分,选框产生网络 RPN 以及检测子网络.通常来说, Faster R-CNN 先通过一个通用的卷积神经网络结构如 ZF^[7], VGG16^[8],

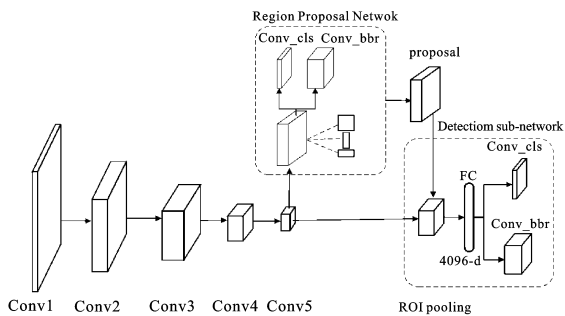


图 1 Faster R-CNN 结构图

这些通用的网络结构中的卷积层一般可以划分为五个 block(图 1 中的 Conv1 ~ Conv5), 每个

block 包括 1 层或多层卷积层, 同时每个 block 之后的池化层会减小特征图的尺寸, 一般是缩小两倍. 可以看到 RPN 采用 Conv5 部分的特征图来产生选框, 先使用一个 3×3 的卷积核在特征图上进行滑动, 对于每一个滑动窗口的中心采用 9 种不同比例, 不同尺度的 anchor, 可以逆向推导出 anchor 在原始输入图像中的区域, 这个区域就是选框. 假设输入图像的大小为 $1\,000 \times 600$, Conv5 的特征图大小为 60×40 , 于是最终可以产生的选框数目为 20 000 左右.

RPN 的训练过程如下, 设 anchor 映射选框区域集合 $\mathbf{P}$, 通常矩形区域可以用坐标 (x, y, w, h) 来表示, x, y 为矩形左上角的坐标, w 和 h 为矩形的宽度与高度, 原图像中物体的真实区域集合为 $\mathbf{G}$, RPN 生成的选框集合为 $\hat{\mathbf{G}}$. 图 1 RPN 网络中的 Conv_bbr 和 Conv_cls 都是卷积核大小为 1×1 的卷积层, Conv_cls 层输出的是选框为前景的概率, $(\hat{d}_x, \hat{d}_y, \hat{d}_w, \hat{d}_h)$ 为 Conv_bbr 层的输出, 与 $\hat{\mathbf{G}}$ 的关系为:

$$\hat{d}_x = (\hat{G}_x - P_x) / P_w \quad (1)$$

$$\hat{d}_y = (\hat{G}_y - P_y) / P_y \quad (2)$$

$$\hat{d}_w = (\hat{G}_w - P_w) / P_w \quad (3)$$

$$\hat{d}_h = (\hat{G}_h - P_h) / P_h \quad (4)$$

同样的 $\mathbf{G}$ 与 $\mathbf{P}$ 的关系也可以用另外一组偏移量 (d_x, d_y, d_w, d_h) 来表示, RPN 的作用就是令产生的选框集合 $\hat{\mathbf{G}}$ 更接近真实的物体区域集合 $\mathbf{G}$, RPN 网络的损失函数定义为:

$$L(p, y, d, \hat{d}) = L_{\text{cls}}(p, y) + \lambda [y \geq 1] L_{\text{loc}}(d, \hat{d}) \quad (5)$$

总体损失函数包括分类损失以及定位损失, λ 为平衡两者的超参, 一般设为 1; p 代表 Conv_cls 的输出, 即选框为各个物体类别的概率(物体类别包括背景类别, 背景类别的标签一般为 0); y 就代表实际选框的类别, $y \geq 1$ 代表了该选框为前景物体, 并非背景, 在式(5)中分类的损失函数为:

$$L_{\text{cls}}(p, y) = -\log(p_y) \quad (6)$$

定位的损失函数为:

$$L_{\text{loc}}(d, \hat{d}) = \sum_{i \in \{x, y, w, h\}} \text{smooth}_{\text{L1}}(d_i - \hat{d}_i) \quad (7)$$

其中

$$\text{smooth}_{\text{L1}}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise} \end{cases} \quad (8)$$

损失函数建立后, 在训练过程中, 通过梯度下降算法使得 $L(p, y, d, \hat{d})$ 最小. RPN 网络中的选框产生

后,通过筛选与真实选框接近的,进入后面的检测子网络,检测子网络主要包括 RoI Pooling 层,全连接层以及定位预测和分类的卷积层。RoI Pooling 层主要将不同大小的特征图固定到一定的大小,方便后面的全连接层得到特征向量。定位预测和分类的卷积层分别用于物体区域预测以及物体类别预测。

4 尺度自适应网络

高分辨率遥感图像目标检测与普通图像不同:

(1) 在尺寸比较大的遥感图像中(像素值 $1\,000 \times 1\,000$ 左右),有些比较小的物体如地面上的车辆,河里面的船只等,小目标由于区域像素较少,在 CNN 的前向中,特征也较少,较难识别;(2) 除了小目标以外,遥感图像中还有一些较大的目标,如操场,篮球场等,设计的检测网络需要同时兼顾不同尺度的目标。考虑到上述的问题与挑战,需要对 Faster R-CNN 框架进行改进,本文提出的尺度自适应网络,改进的方法主要由以下两点。

(1) 增加 Anchor 的种类

在 Faster R-CNN 中,Anchor 是区域选框产生的基准,尺度集合 S 为 $\{128^2, 256^2, 512^2\}$, 比例集合 R 为 $\{1:2, 2:1, 1:1\}$ 尺度和比例相互组合共同组成了 9 种不同的矩形区域。普通自然图像的物体尺度一般较大,所以采用集合 S 是合适,但是遥感图像中存在尺度较小的物体,将集合 S 扩增为 $\{32^2, 64^2, 128^2, 256^2, 512^2\}$, 实验证明,新尺度的增加可以提升目标检测的准确度。

(2) 多层特征融合

卷积的过程中,因为 Pooling 层的存在,特征图的尺寸会逐渐减少(一般 Pooling 层的 $\text{stride}=2$, 即尺寸缩小一半),所以小物体的特征在前向的过程中会逐渐消失,这对于识别和检测是相当不利的,解决这一问题的捷径是利用更低一层的特征输入到检测子网络中进行识别和检测。然而,低层的特征虽然对小物体的检测更有利,但是却不适合大物体,而更高层的特征虽然可能不包含小物体的特征,但感受也相对较大,适合大物体识别。为此,设计的网络需要能够既包含大物体信息也需要包含小物体信息,这里提出的尺度自适应网络采用反卷积来对高层的特征进行上采样,恢复到与低层特征大小相同,并与底层特征相加,完成特征融合。

如图 2 所示,提出的网络中,首先对 Conv5 的特征图进行反卷积,反卷积可以看做是卷积的逆操作,与卷积的局部映射相似,但反卷积可以在增大特

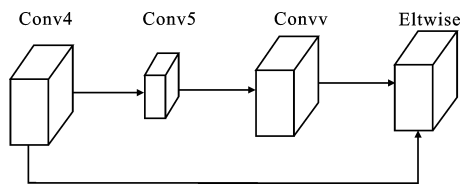


图 2 Conv4 与 Conv5 特征融合结构

征图大小的同时,保留原特征的相关信息。这里的融合方式采用的方式是相加,即每个通道相同大小的两个特征图点到点相加得到新的特征图。融合的特征既具有大物体特征,也具有能够识别出小物体的能力。

5 实验与结果

5.1 目标检测的评价标准

对于目标检测任务来说,Mean Average Precision (mAP) 是一个公认的评价识别精度的指标,多个类别物体检测中,每个类别都可以得到相应的 AP,即 precision-recall 绘制的曲线下的面积,mAP 定义为所有类别 AP 的平均

$$mAP = \frac{\sum_{q=1}^Q AP_q}{Q} \quad (9)$$

式中, Q 为类别数目,mAP 可以很好地衡量目标检测算法在测试数据集中的性能。

对于目标检测任务来说,衡量一个真实物体是否被检测到是由一个指标 Intersection of Union (IoU) 决定的,IoU 定义为

$$IoU = \frac{GT \cap DR}{GT \cup DR} \quad (10)$$

式中,GT 代表真实物体区域;DR 代表网络检测出的物体区域;IoU 为二者交集与并集的比例。当 $IoU > 0.5$ 时,我们认为已经检测到该物体。

5.2 实验结果

本文采用 NWPU VHR-10^[3] 数据集,该数据集主要由 10 个类别的高精度遥感图像组成,这十个类别主要是:飞机、船只、存储罐、棒球场、网球场、篮球场、操场、港口、桥梁以及车辆。为了验证提出的网络的识别精度,数据集被划分为训练数据集(70%)和测试集(30%)。

实验训练机器主要配置为单 CPU Intel(R) Xeon(R) CPU E5-1620 v4 @ 3.50 GHz,单 GPU NVIDIA GTX 1080Ti,主要基于深度学习平台 Caffe 实现。Faster R-CNN 和尺度自适应网络的基础结构都是 VGG16,在试验中,学习率初始设置为 0.001 并且每 25 000 迭代降低 10 倍,经过 50 000 迭

代后,模型训练终止并保存.最终迭代次数的模型在测试数据集中进行测试.整个训练过程持续约6.5小时,每次迭代大约为2.0秒.

表 1 不同算法在 NWPU VHR-10 数据集上的测试结果(%)

	COPD [86]	RICNN[4]	Faster R-CNN	SA-CNN
飞机	62.3	88.4	95.2	97.8
船只	68.9	77.3	78.0	87.6
存储罐	63.7	85.3	51.3	67.2
棒球场	83.3	88.1	94.2	94.8
网球场	32.1	40.8	93.5	99.5
篮球场	36.3	58.5	95.9	99.5
操场	85.3	86.7	99.7	95.9
港口	55.3	68.6	98.8	96.8
桥梁	21.5	61.5	66.9	68.0
车辆	0.45	71.1	60.3	85.1
mAP	54.6	72.6	83.4	89.2

表 1 给出了不同算法在 NWPU VHR-10 数据集上的测试结果,可以看到,传统的检测方法如 Collection of Part Detectors(COPD)检测效果一般,mAP 只有 54.6%,G. Chen 提出的 RICNN 大大提高了遥感图像中不同类别物体的检测准确率,但仍有部分物体尤其是小目标检测准确率不高,与这些方法相比,本文采用的 Faster R-CNN 以及改进的 SA-CNN 进一步地提升了检测准确率,尤其是 SA-CNN 拥有着最高的 mAP,并且在大部分目标中,mAP 已经提升到 95%以上.图 3 (a)-(f)为 SA-CNN 在一些实际测试图片中的效果,可以看到 SA-CNN 可以很好地检测遥感图像中的不同物体,尤其在小目标物体如飞机,车辆,船只具有着不错的检测准确率,该结果表明该网络拥有着很高的实际应用价值.

6 结束语

本文提出了一种应用在高精度遥感图像中的目标检测网络,在 Faster R-CNN 的基础上,针对遥感图像中不同目标尺度差异尤其是小目标这一难题,改进了选框产生网络并且提出了不同层次卷积特征图的融合.实验证明,SA-CNN 与以往的方法相比,拥有更高的检测精度(mAP 为 89.2%),可以很好地在检测不同民度的遥感目标.

参考文献:

[1] L. Zhang, L. Zhang, and B. Du. Deep learning for remote sensing data: A technical tutorial on the state of the art[J]. IEEE Geoscience and Remote Sensing Magazine, 2016, 4(2): 22-40.

[2] Z. Zou and Z. Shi. Ship detection in spaceborne opti-

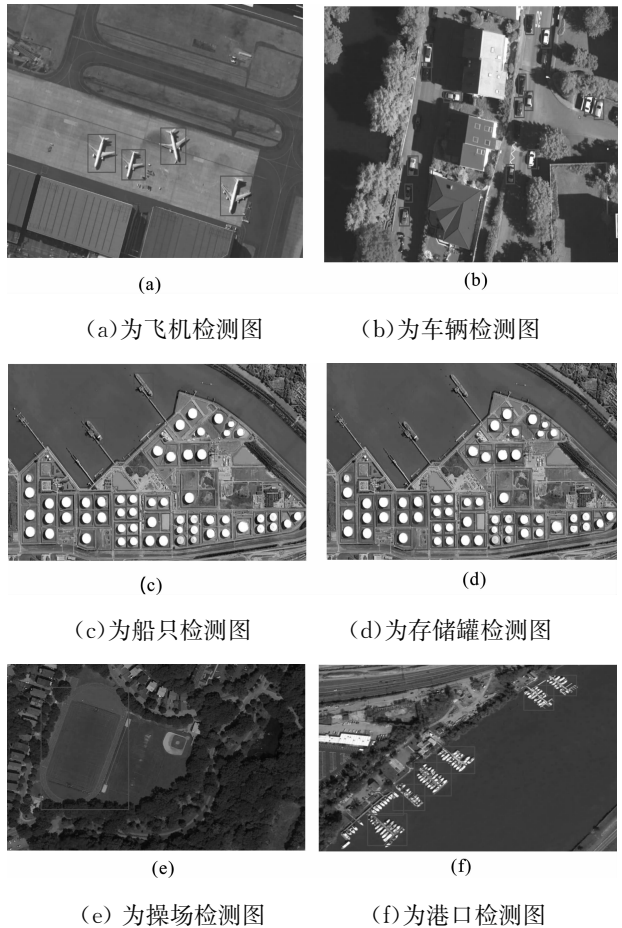


图 3 尺度自适应网络在高精度遥感图像中的检测效果

cal image with svd networks[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(10): 5832-5845.

[3] G. Cheng, P. Zhou, and J. Han. Learning rotation-invariant convolutional neural networks for object detection in vhr optical remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(12):7405-7415.

[4] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.

[5] R. Girshick. Fast R-CNN[C]// Proceedings of the IEEE International Conference on Computer Vision (ICCV). 2015: 1440-1448.

[6] S. Ren, K. He, R. Girshick, and J. Sun. Faster R-CNN: towards realtime object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.

分变量都量化为 16 bits,部分变量被量化为 18 bits. 脉动阵列结构的求倒结构完全分时复用,求倒结构的输入位宽为 13 bits,输出位宽为 18 bits,利用 SBTM 算法进行优化,取 $n_0=5, n_1=4, n_2=4, n_3=0$. 实现结果如下表所示:

表 1 FPGA 实现的资源消耗

资源种类	LUT	Slice	FF	DSP
消耗数量	6540	2106	6246	128

查找表大小为 13.5 kb. 由于该 FPGA 片上的 Block Memory 最小为 18 kb,而使用 SBTM 算法需要 2 个查找表,所以在 FPGA 上实现的时候需要消耗 2 个 18 kb 的 Block Memory.

本文提出的 LDLT 分解结构在系统时钟为 200 MHz 的情况下有 0.379 ns 的裕量,即系统时钟最高能够支持 216 MHz. 同时,该结构支持每个时钟周期输入一个矩阵列向量的吞吐率.

5 结束语

本文提出了一种适用于 MIMO 系统 MMSE 检测的 LDLT 分解算法的硬件实现. 该硬件实现充分利用了待分解矩阵的共轭对称性,仅需对其下三角矩阵进行操作,节省了硬件资源;同时针对不利于硬件实现的除法,提出了基于 SBTM 算法优化的求倒结构,并通过调整处理单元的 latency 使求倒结构能够完全分时复用,进一步减小了硬件资源的开销. 由于该求倒结构充分利用了待求倒数据的位宽,所以能在仅使用较小的查找表的情况下达到良好的数据精度,解决了近期文献中为了取得较好数据精度而需要在求倒结构中使用的查找表过大的问题.

参考文献:

[1] Burg A, Haene S, Perels D, et al. Algorithm and VLSI architecture for linear MMSE detection in MIMO-OFDM systems[C]. IEEE International Symposium on Circuits and Systems, 2006. ISCAS 2006. Proceedings. IEEE, 2006:4.

[2] Chockalingam A, Rajan B S. Large MIMO Systems [M]. IEEE, 2013.

[3] Studer C, Fatch S, Seethaler D. ASIC implementation of soft-input soft-output MIMO detection using MMSE parallel interference cancellation[J]. IEEE Journal of Solid-State Circuits, 2011, 46(7): 1754-1765.

[4] Gestner B, Anderson D V. Single Newton-Raphson iteration for integer-rounded divider for lattice reduction algorithms[C] // Circuits and Systems, 2008. MWCAS 2008. 51st Midwest Symposium on. IEEE, 2008: 966-969.

[5] Schulte M J, Stine J E, Wires K E. High-speed reciprocal approximations[C]// Signals, Systems & Computers, 1997. Conference Record of the Thirty-First Asilomar Conference on. IEEE, 1997, 2: 1183-1187.

作者简介:

梁子豪 男,(1993-),硕士研究生. 研究方向为 MIMO 系统的检测与硬件实现. E-mail:lzhhdzw@163.com.

刘 辉 男,(1984-),硕士,工程师. 研究方向为卫星通信系统.

贺光辉 男,(1980-),副研究员. 研究方向为通信系统 MIMO 技术、大规模集成电路等.

(上接第 81 页)

[7] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition[C]. In: International Conference on Learning Representations (ICLR), 2015.

[8] G. Cheng, J. Han, P. Zhou, and L. Guo. Multi-class geospatial object detection and geographic image classification based on collection of part detectors[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2014,98(12):119-132.

作者简介:

吴佳祥 男,(1995-),硕士研究生. 研究方向为计算机视觉. E-mail:wjxzju@foxmail.com.

刘 辉 男,(1984-),硕士,工程师. 研究方向为卫星通信系统.

贺光辉 男,(1980-),博士,副教授. 研究方向为无线通信和多媒体系统的信号处理算法优和 VLSI 架构设计.