

甚高速区域卷积神经网络的船舶视频检测方法

杨 名¹, 阮雅端¹, 陈林凯², 张 鹏¹, 陈启美¹

(1. 南京大学 电子科学与工程学院, 南京 210093; 2. 江苏理工学院 计算机工程学院, 常州 213001)

摘要:为解决背景建模等传统视频目标识别算法在内河水运复杂环境误差过大的问题,提出了甚高速区域卷积神经网络的船舶识别检测方法. 分析了传统方法不足,阐述了卷积神经网络及后续的区域卷积神经网络的机制,给出了甚高速区域卷积神经网络特征模型,解析了损失函数的参数构建、参数设定,设定候选区域网络预测目标边界、计算匹配目标概率. 经实际内河运动船舶视频检测表明,该算法对船舶识别率优于90%,同时对不同清晰度、不同视角、不同船舶流量的场景具有很好的鲁棒性,比传统的背景建模算法提高25.75%.

关 键 词: 船舶视频检测; 深度学习; 背景建模; 卷积神经网络

中图分类号: TN929.53

文献标志码: A

New Video Recognition Algorithms for Inland River Ships Based on Faster R-CNN

YANG Ming¹, RUAN Ya-duan¹, CHEN Lin-kai², ZHANG Peng¹, CHEN Qi-mei¹

(1. School of Electronic Science and Engineering, Nanjing University, Nanjing 210093, China;

2. College of Computer Engineering, Jiangsu University of Technology, Jiangsu Changzhou 213001, China)

Abstract: To avoid the excessive error during the implementation of background modeling and other traditional video object recognition algorithms in complex inland waterway environment, a new ship identification detection method based on faster region convolutional neural networks (i. e. Faster R-CNN) was proposed. The shortcomings of these traditional methods were analyzed, elaborating the mechanism of convolutional neural network and the subsequent regional convolutional neural network was also described. A model of R-CNN was put forward, which worked out how to construct and set the parameter of Loss Function, set region proposal networks (RPN) to predict a target boundary, and calculated the probability of matching targets. The actual video detection for moving inland river ships indicated that the ship identification algorithm held a ship recognition rate of over 90%. Meanwhile, this new algorithm had good robustness, 25.75% higher than traditional background modeling algorithm in situations with different visual clarity, from different perspectives, and regardless of the number of ships.

Key words: ship video recognition; deep learning; background modeling; convolutional neural networks

内河航道日趋繁忙,借助基于计算机视觉的船舶目标识别已开始在监管服务中应用. 水运场景复杂,波光粼粼、光照变化、船尾同步拖纹、上下行行船交错、河边树木晃动等. 常用的3种目标识别方法

显得无能为力.

借助运动目标特征的方法对于检测到的目标具有持续性,更适合内河船舶视频目标检测. 卷积神经网络(CNN^[1], convolutional neural networks)对运

动目标的平移、比例缩放、倾斜等变形具有高度不变性。然而不能简单套用,需作相当研究,为此,提出了甚高速的区域卷积神经网络^[2]的船舶视频检测方法。

1 区域卷积神经网络

区域卷积神经网络(R-CNN^[3], region convolutional neural networks)是近些年较为流行的目标检测算法之一,其以区域选择算法和卷积神经网络为基础,用线性支持向量机(SVM, support vector machines)分类器实现分类。R-CNN 最大的缺点在于计算复杂度高,主要由以下三方面造成:1) 选择性搜索(Selective Search^[4]) 在计算过程中会提取 2 000 个候选区域,对于每一个候选区域都会输入 CNN 模型计算特征,会造成大量重复计算;2) CNN 模型特征 SVM 分类器中间有大量的硬盘 I/O 读写操作;3) 提取候选区域算法效率较低。

甚高速区域卷积神经网络(Faster R-CNN, faster region convolutional neural networks)针对这三方面不足作了优化。具体而言,Faster R-CNN 在卷积层输出的特征图上提取候选区域,在卷积层最后一层直接输出分类结果,不需要 SVM 分类器,避免了磁盘读写操作。

2 甚高速区域卷积神经网络模型架构

Faster R-CNN 主要由候选区域网络(RPN, region proposal networks)和快速区域卷积神经网络(Fast R-CNN^[5], fast region convolutional neural networks)两个模块组成。RPN 是一个能够预测某位置上的目标框边界和判断目标对象的全卷积网络,用

来产生高质量的区域检测。Faster R-CNN 将其与 Fast R-CNN 一并加入到一个网络里面,这样就不需要在之前单独进行区域检测,正因为这样的结构,可以完成端到端的监测任务。Faster R-CNN 同时将图片和感兴趣区域一起送入 Fast R-CNN 网络,每一个感兴趣区域被池化到一个固定大小的特征图,然后通过全连接到一个特征向量,两者通过交替微调实现共享网络权重,这样不仅解决了处理器实现选择搜索算法造成的速度慢问题,而且与深度网络相结合,共享前面的卷积计算,解决了 Fast R-CNN 中显著性目标检测对速度的瓶颈,从而提高了计算效率,达到了实时监测的效果。

3 网络与损失函数设定

3.1 卷积神经网络

以 CNN 开源项目 Caffe 为基础,构建了一个应用于船舶视频检测的卷积神经网络模型,如图 1 所示。该模型包括三个部分,输入层,隐藏层,输出层。输入层直接以图片作为输入,图片尺寸归一化为 $224 \times 224 \times 3$,长宽均为 224 像素,3 通道。隐藏层包括 5 个卷积层,3 个池化层和 2 个全连接层。5 个卷积层分别包含 96 个 55×55 、256 个 13×13 、384 个 13×13 、384 个 13×13 、256 个 6×6 的特征图。前 4 个卷积层的卷积核尺度依次设置为 $5 \times 5 \times 96$, $3 \times 3 \times 256$, $3 \times 3 \times 384$, $3 \times 3 \times 384$,卷积核数量依次为 256,384,384,256。第 5 五个卷积层经过池化后与 2 个全连接层相连,输出一个 4 096 维向量。池化方式为重叠最大池化,池化尺寸设置为 3×3 ,步长设置为 2。输出层以全连接层输出的 4 096 维向量作为输入,采用逻辑回归分类器,得到原始图像的分类结果。

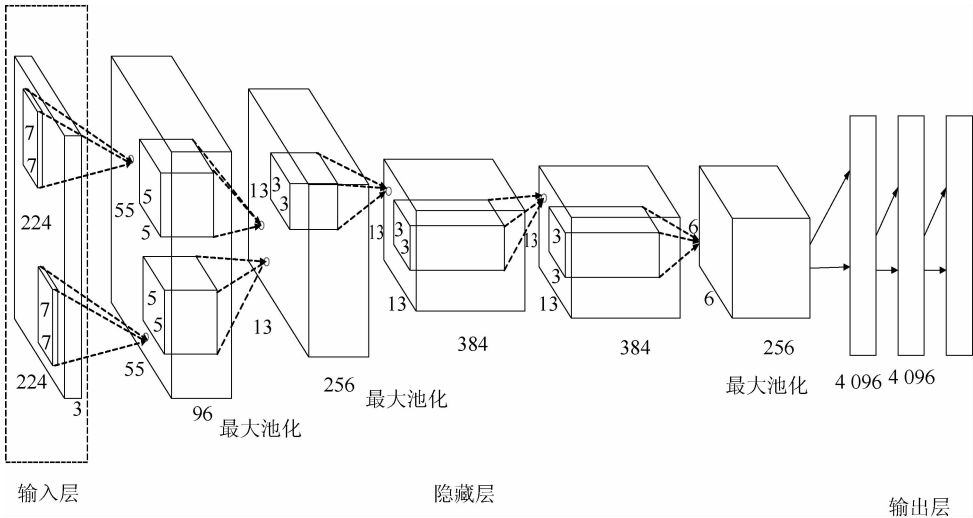


图1 CNN模型

3.2 RPN

RPN 作为全卷积网络,在共享卷积层后面,构造了1个卷积层和2个并列的全连接层的网络.如图2所示,RPN采用滑动窗口机制,定义滑动窗口尺寸为 $n \times n$,以每个滑动窗口的中心点为锚点(anchor),选取不同尺度、不同宽高比的矩形框,构造锚点金字塔,定义同一个滑动窗口位置有 k 个矩形框.滑动窗口映射到一个 N 维向量,可用包含 N 个卷积核的卷积层实现,作为2个全连接层的输入,2个全连接层分别实现矩形框的位置回归和目标分类.

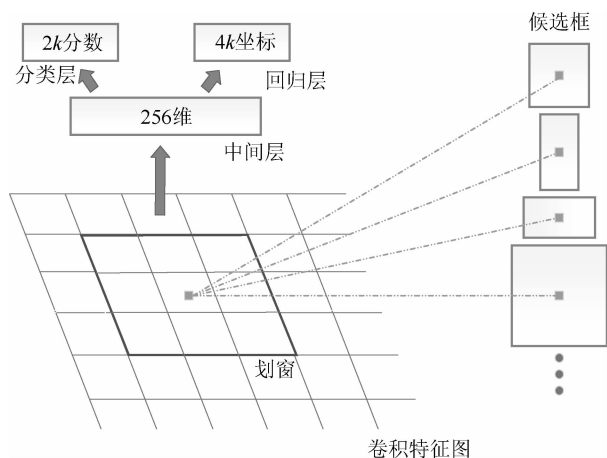


图2 RPN结构

文中滑动窗口尺寸为 3×3 ,用256的卷积核把矩形框映射为一个256维的向量,候选框宽高比分别为1:1、1:2、2:1;维度分别为128维、216维、512维,故 k 取9. RPN 2个全连接层的参数数量为 $256 \times (4 \times 9 + 2 \times 9) = 13824$. 在训练过程中,定义与人工标注矩形框交并比最大的矩形框为正样本,小于阈值的为负样本,实验中阈值设置为0.3. 其他样本不参与训练过程.

3.3 损失函数

将候选区域框分为两类. 其一与所有标记框交并比(IoU, Intersection over union)都小于0.3,另一与任意标记框IoU大于0.7或者与标记框具有较高IoU的. 因此对于每一个候选区域框,定义其损失函数

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{\text{cls}}} \sum_i L_{\text{cls}}(p_i, p_i^*) + \lambda \frac{1}{N_{\text{reg}}} \sum_i p_i^* L_{\text{reg}}(t_i, t_i^*) \quad (1)$$

对于 L_{cls} 分类损失函数和回归损失函数 L_{reg} ,分别定义为

$$L_{\text{cls}}(p_i, p_i^*) = -\ln e[p_i^* p_i + (1 - p_i^*)(1 - p_i)] \quad (2)$$

$$L_{\text{reg}}(t_i, t_i^*) = \sum_{i \in \{x, y, w, h\}} \text{smooth}_{L_1}(t_i, t_i^*) \quad (3)$$

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2 & |x| < 1 \\ |x| - 0.5 & |x| \geq 1 \end{cases} \quad (4)$$

其中: i 表示候选区域编号; p_i 表示候选区域 i 是目标的概率; p_i^* 表示样本正负,当矩形框为正样本时 p_i^* 为1,否则为0; t_i^* 表示与 t_i 关联的标记矩形框的坐标值; L_{cls} 是一个二分类(判断是否是目标)损失函数; $p_i^* L_{\text{reg}}$ 表示只针对正样本为1; N_{reg} 和 N_{cls} 为归一化参数,分别对应训练块(batch)尺寸和矩形框数量,实验中设置为256和2400; λ 为平衡因子,实验中设为10.

4 测试结果分析

测试采用了无锡新虹桥、葑溪大桥南、葑溪大桥北、蔡东桥、向化二号桥、山北大桥、锡宜高速的视频源. 实验采用系统为Ubuntu 15.10, GPU为Nvidia Quadro K620, 软件为Matlab 2014a. 考虑到实际情况下,船与非船图像特征差异较大,仅需要少量训练样本即可完成训练,故提取其中的山北大桥、新虹桥、葑溪大桥南图片作为训练集,其余作为测试集. 测试中提取160张训练图像,并通过镜像扩充得到320张图像的训练集,其中40张验证集,280张数据集,图像尺寸为 352×228 ,每批训练图片数量设置为10,迭代次数为2500.

将100张测试集分为两组,第一组80张为相同场景下Faster R-CNN算法和背景建模算法的对比检测,第二组20张为不同场景下Faster R-CNN的对比实验.

4.1 相同场景的比较分析

如图3和图4所示. 第一组对比了Faster R-CNN方法与背景建模方法检测效果,实验背景建模方法基于混合高斯建模算法. 实验测试集提取视频源80张图片,实验结果如表1所示,包括168个船舶目标, Faster R-CNN误检14个目标,漏检39个目标,准确率达90.21%,背景建模误检52个目标,漏检36个目标,准确率为71.74%. 对比图3(a)(b)(e)与图4(a)(b)(e)看出,背景建模算法得到的目标框图并不完整,而Faster R-CNN的框图较之相对完整. 对比图3(g)和图4(g),当船舶距离较近时,背景建模算法存在目标融合的问题, Faster R-CNN

算法对这种情况具有良好的检测效果. 对比图 3 (d)(f)与图 4(d)(f)发现,当水面波光或者船颜色与水面颜色接近时,背景建模出现了船舶分体和错检的情况,而 Faster R-CNN 算法并没有出现类似的情况. Faster R-CNN 算法基于卷积神经网络,对平移、比例缩放、倾斜具有高度不变性,对检测的目标具有持续性,而背景建模算法作为帧间像素解析法,依赖于背景图像效果,检测结果相对不稳定. 综上所述,Faster R-CNN 算法对于实际内河场景下波光问题具有较好的鲁棒性,框图完整,准确率比传统背景建模方法提高 25.75%.

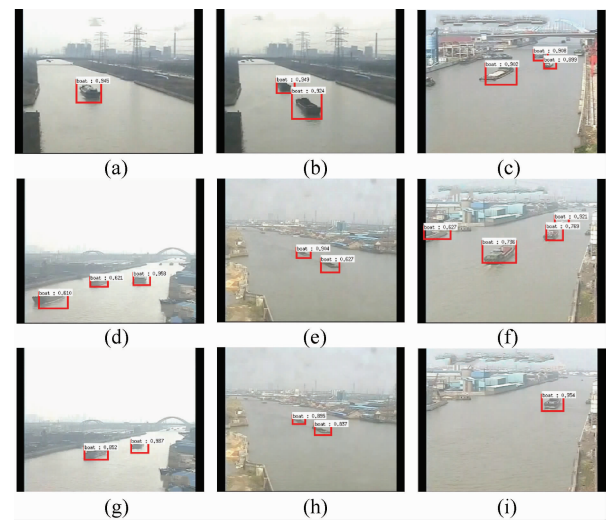


图3 Faster RCNN 第一组测试结果

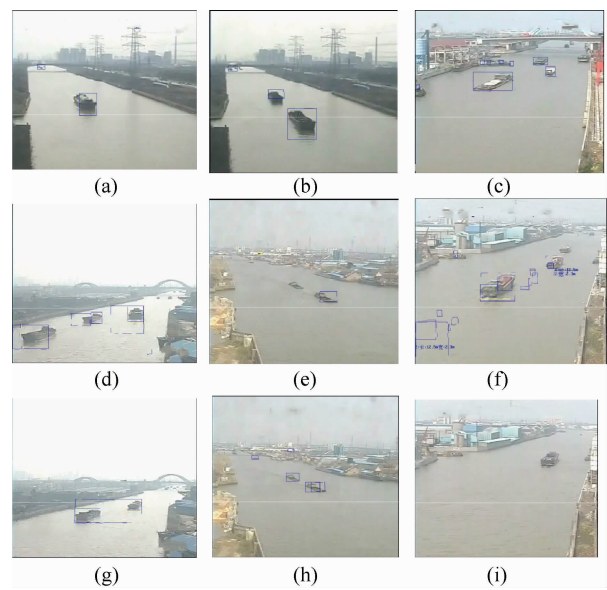


图4 背景建模第一组测试结果

表 1 第一组测试结果

算法	检测	误检	漏检	目标	准确率/%
背景建模	132	52	36	168	71.74
Faster R-CNN	129	14	39	168	90.21

4.2 不同场景的结果分析

第二组实验中提取了无锡市锡宜高速、蔡东桥、向化二号桥视频源中 20 张图片,包括 50 个船舶目标,实验结果如表 2 所示,误检 3 个,漏检 11 个,准确率达 92.86%. 测试结果如图 5 所示. 实验表明,Faster R-CNN 对于清晰视频源如图 5(c)(f)、模糊视频源如图 5(b)(e)、流量较大视频源如图 5(a)(d)均具有比较好的检测效果. 结果表明 Faster R-CNN 算法在多种实际场景下对船舶均具有较好的检测效果.

表 2 第二组测试结果

算法	检测	误检	漏检	目标	准确率/%
Faster R-CNN	39	3	11	50	92.86

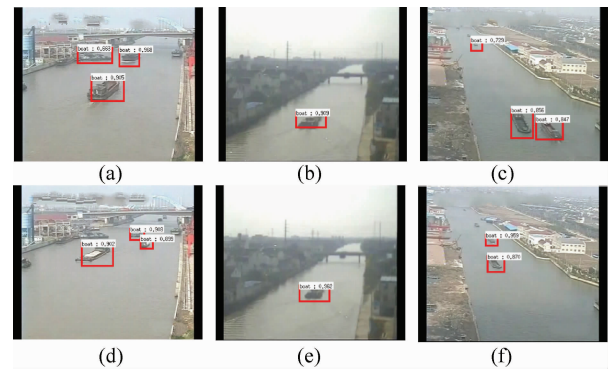


图5 Faster R-CNN 第二组测试结果

4.3 时间效率分析

统计实验中所有 99 张尺寸为 352×288 图片的测试时间,平均每幅图像耗时 379 ms,由于 Faster R-CNN 是一个端到端的网络,所以统计时间包括了卷积计算、分类等所有步骤. 在所有统计中,最高时间为 385 ms,最低时间为 371 ms,帧率为 2.6 Hz. 其中图像尺寸调整、卷积计算、区域提取占 39%,分类占 61%. 在实际应用中,尽管该算法时间复杂度较高,而实际上船舶行驶速度较慢. 该检测算法的准确率和实时性均满足实际要求.

5 实验小结

提出了基于 Faster R-CNN 的船舶识别检测算

法,其对不同清晰度、不同视角、不同船舶流量的场景具有很好的鲁棒性,明显优于传统的背景建模算法,已得以试用.

参考文献:

[1] Lécun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[2] Zhang Yu, Wei Xiushen, Wu Jianxin, et al. Weakly supervised fine-grained categorization with part-based image representation[J]. IEEE Transactions on Image Processing, 2016, 25(4): 1713-1725.

[3] Li Nanxi, Li Y, Porikli F. Deeptack: learning discriminative feature representations online for robust visual tracking[J]. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, 2015, 25(4): 1834-1848.

[4] Uijlings J R R, Sande K E A V D, Gevers T, et al. Selective search for object recognition[J]. International Journal of Computer Vision, 2013, 104(2): 154-171.

[5] Ren Shaoqiang, He Kaiming, Girshick R, et al. Object detection networks on convolutional feature maps[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 50(1): 815-830.