

深度学习技术及其在舰船目标识别领域的应用

马 啸,邵利民,金 鑫,徐冠雷

(海军大连舰艇学院 航海系,辽宁 大连 116018)

摘 要:近年来,中国海洋权益争端日益频繁,对海上舰船目标进行识别和监视显得尤为重要。图像技术的发展为舰船目标识别提供了新的感知来源,随着获取图像数据量的增加,传统通过人工判读识别舰船目标的方法资源消耗大且难以保证目标识别的精度和可靠性,迫切需要引入新的技术和方法以节省人力资源,提高舰船目标识别的精度和可靠性。大数据背景下,深度学习技术在语音识别、图像识别等领域的发展为舰船目标识别技术的突破提供了新思路。文中阐述了具有代表性的深度学习模型,介绍了主流的基于深度卷积神经网络的目标识别方法,将其中较典型的两种目标识别方法 Faster RCNN 和 YOLO 应用于舰船目标识别领域,通过客观分析比较两种方法在舰船目标识别中的优劣性。舰船目标识别结果表明,Faster RCNN 的准确率和召回率高于 YOLO,但其运行效率远低于 YOLO 的运行效率。由此,提出下一步的工作方向。

关键词:深度学习;模型;卷积神经网络;目标识别;舰船

中图分类号:TP391.9

文献标识码:A

文章编号:1673-629X(2019)12-0141-07

doi:10.3969/j.issn.1673-629X.2019.12.025

Deep Learning Technology and Its Application in Ship Target Recognition

MA Xiao, SHAO Li-min, JIN Xin, XU Guan-lei

(Department of Navigation, Dalian Naval Academy, Dalian 116018, China)

Abstract: In recent years, the conflicts of maritime rights and interests have become more and more frequent in China, therefore it is particularly important to recognize and monitor the ship target. The development of image technology provides a new source of perception for ship target recognition. With the increase of image data acquisition, the traditional ship target recognition method by manual interpretation consumes a lot of resources and is difficult to ensure the accuracy and reliability of target recognition, so it is an urgent to introduce new technologies and methods to save human resources and improve the accuracy and reliability of ship target recognition. Under the context of big data, the development of deep learning technology in speech recognition, image recognition and other fields provides a new idea for the breakthrough of ship target recognition. We describe some representative deep learning models and introduce the mainstream methods of target recognition based on deep convolution neural network. Two typical target recognition methods, Faster RCNN and YOLO, are chosen to be applied in ship target recognition. Through objective analysis to compare the advantages and disadvantages of these two methods in ship target recognition, the ship target recognition result shows that the accuracy and recall rate of the Faster RCNN are higher than YOLO, but its operation efficiency is much lower than YOLO. Therefore, the next step of work is proposed.

Key words: deep learning; model; convolutional neural networks; target recognition; ship

0 引言

对海上舰船目标进行有效识别和监控对维护海洋权益、保障海上航行安全具有重要的现实意义。随着信息探测技术的发展,获取到的舰船目标信息越来越多,传统通过人工判读识别舰船目标的方法资源消耗大且难以保证目标识别的精度和可靠性,迫切需要引入新的技术和方法以节省人力资源,提高舰船目标识

别的精度和可靠性。

深度学习作为机器学习领域的新分支,在2006年由 Geoffrey E. Hinton 等首次提出并迅速发展起来^[1]。它通过模仿人脑的学习机制来解释数据,如图像、文本、声音等,其本质上是一种特征学习方法,能够通过自主学习从大数据样本中剔除冗余信息,提取出更具表征能力和区分能力的目标特征,在文字识别、语音识

收稿日期:2019-01-09

修回日期:2019-05-21

网络出版时间:2019-09-24

基金项目:国家自然科学基金(61471412,61771020)

作者简介:马 啸(1993-),女,博士研究生,研究方向为图像处理、目标识别;邵利民,教授,研究方向为海洋气象。

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1450.TP.20190924.1534.016.html>

别、图像识别及视频处理等领域得到了广泛应用^[2]。将深度学习技术引入到舰船目标识别领域,为实现舰船目标识别技术的突破提供了一种新思路。

1 深度学习模型的发展

从 2006 年发展至今,深度学习的模型结构不断发展,与传统的机器学习方法,如支持向量机算法、聚类算法、Boosting 算法等相比,深度学习模型的中间隐层数更多,结构更深,有的模型甚至达到上千层,通过多隐层的分级变换,将底层特征转换到不同的特征空间,逐步抽象为高层特征,形成一个多层传递、逐步抽象、迭代细化的过程^[2-4]。目前深度学习模型中具有代表性的主要有基于受限玻尔兹曼机(restricted Boltzmann machines, RBM)的深度模型、基于循环神经网络(recurrent neural networks, RNN)的深度模型和基于卷积神经网络(convolutional neural networks, CNN)的深度模型。

基于 RBM 的深度模型是深度学习发展初期提出的一种模型结构,包括深度置信网络和深度玻尔兹曼机。RBM 作为这种模型结构的基本组成单元是一种典型的双层网络结构,其可视层单元和隐层单元相互连接,但层内没有连接^[5-7]。

循环神经网络是一类用于处理序列数据的神经网络,基于 RNN 的深度模型多用于机器翻译、语音识别等领域。

传统的神经网络模型中,从输入层到隐含层再到输出层,层与层之间是全连接的,层内节点之间无连接,这种结构无法对序列数据进行预测。循环神经网络 RNN 模型中,隐含层的输入不仅包括输入层的输出,还包括上一时刻隐含层的输出,且隐含层之间的节点是有连接的。这种结构使得 RNN 理论上能够处理任意长度的序列数据,实际应用中,一般假设当前状态只与前面几个状态相关以降低结构复杂性。

目前基于 RNN 的深度模型有 SRN、BRNN、DRNN、ESN、LSTM、GRU、CW-RNN 等。

简单循环神经网络(simple recurrent neural networks, SRN)是 RNN 结构的特例,它是一种三层网络,在隐含层增加了上下文单元,上下文每一个节点保存与之相连的隐含层节点上一时刻的输出,作用于当前时刻隐含层的状态,即隐含层的输入由输入层的输出和上一时刻自身状态决定,使得 SRN 能够实现序列数据的预测^[8]。

双向循环神经网络(bidirectional recurrent neural networks, BRNN)是由两个 RNN 上下叠加组成的,BRNN 模型假设当前输出与前后序列均相关,输出由两个 RNN 隐含层的状态决定^[9]。

深度循环神经网络(deep recurrent neural networks, DRNN)与 BRNN 相似,但模型包含层数更多,复杂性更高,具有更强大的表达与学习能力,需要的训练数据也更多^[10]。

回声状态网络(echo state networks, ESN)使用大规模随机连接的 RNN 取代经典神经网络的中间层,从而简化网络的训练过程^[11]。

长短时记忆模型(long short-term memory, LSTM)使用不同的函数计算 RNN 隐含层的状态,能够更好地对长短时依赖进行表达,在机器翻译、语言建模、多语言处理等方面应用广泛^[12-13]。

门循环控制单元(gated recurrent unit, GRU)是 LSTM 的变体,GRU 模型中将前面时刻状态对当前状态的影响进行了距离加权,距离越远权值越小。GRU 既保持了 LSTM 的效果,又简化了模型结构^[14]。

时钟循环神经网络(clockwork recurrent neural networks, CW-RNN)是一种使用时钟频率来驱动的 RNN。它将隐含层分为几组,每一组按照自己规定的时钟频率对输入进行处理。不同的隐含层模块在不同的时钟频率下工作,从而解决了长时间依赖问题^[15]。

CNN 是一种适用于处理网格化数据(如图像)的前馈式人工神经网络,具有自主分层学习自身特征的能力,最早由 Yann LeCun 等于 1989 年提出^[2]。基于 CNN 的深度模型多用于图像识别和视频处理等领域,在大样本集背景下,基于 CNN 的目标识别方法比传统目标识别方法具有更强的鲁棒性和泛化能力^[16]。经典的 CNN 模型由输入层、输出层、若干卷积层及池化层组成,如图 1 所示。卷积层与前一层采用局部连接和权值共享的方式相连,大大降低了模型的参数数量。池化层大幅降低特征的维度,减小分类难度,使网络具有更高的鲁棒性^[17]。

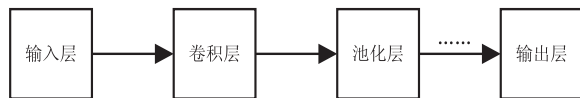


图 1 经典 CNN 模型的简易框图

2012 年以前,受计算机计算能力的限制,卷积神经网络在图像分类、目标识别等领域的发展一直没有取得质的突破,直到 Hinton 等提出了更深的卷积神经网络 AlexNet,并将该模型用于 2012 年的 ImageNet 图像识别竞赛中(ImageNet 是目前公开的最大的视觉数据库,包含 100 万幅图像,超过 1 000 个物体类别),得到了比当时最好的识别方法高出将近一半的正确识别率,引起了人们对深度卷积神经网络的高度重视。基于 CNN 的深度模型开始快速发展^[18],更深更广的深度卷积神经网络模型相继被提出,如 ZFNet^[19]、VGG16^[20]、GoogleNet^[21]以及 ResNet^[22]等。

在大规模视觉识别竞赛 ImageNet 中,基于深度学习的模型展现出了强大的学习能力:在图像分类竞赛单元中,2012 年 AlexNet 模型获得了比传统方法高出 10% 的正确率;2014 年 GoogleNet 模型的正确率为 93.344%,VGGNet 模型的正确率为 92.68%;2015 年 ResNet 模型的正确率为 96.43%;2016 年的 Inception V4 模型的错误率仅为 3.08%,而人类正常进行图像分类的错误率约为 5%,基于深度学习的图像分类能力已经明显高于人类的区分能力^[23],卷积神经网络逐渐被成功应用于目标检测、图像分割、物体识别等领域。

2 基于深度卷积神经网络的目标识别方法

随着深度卷积神经网络模型的不断变深,基于深度卷积神经网络的目标识别(或目标检测)方法也取得了重大进展,它可以通过自主学习从大量训练样本数据中学习到包含一定语义信息的目标特征,更有利于目标的识别和分类^[24]。目前有代表性的主要有 R-CNN、SPP-Net、Fast R-CNN、Faster R-CNN、YOLO、SSD 等。

R-CNN (region - based convolution neural networks) 是 Ross B. Girshick 在 2014 年提出的一种基于区域的深度卷积神经网络目标识别框架,它将候选区域与深度卷积神经网络组合代替传统目标识别算法中滑动窗口与手工设计特征的组合,是深度学习在目标识别领域的首次突破^[2,25-27]。

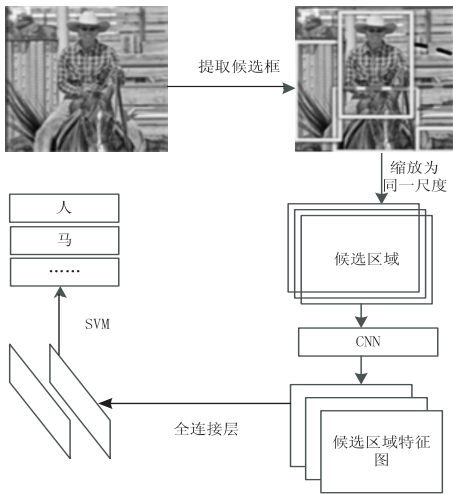


图 2 R-CNN 目标识别流程

R-CNN 的目标识别流程如图 2 所示。首先采用选择性搜索算法从输入图像中提取 2 000 个候选区域,然后将所有的候选区域缩放为同一尺度,通过预先训练好的卷积神经网络获取每一个候选区域的特征图,再经过两个全连接层把特征图转化为特征向量,最后利用支持向量机(SVM)对每个候选区域内的物体进行分类识别^[2,25-28]。相比之前的目标识别方法,该方

法的目标识别精度有了很大提升,这充分体现出深度学习的优势,但该方法存在提取的候选区域数量多、计算量大、中间数据需单独保存、算法运行速率慢、支持向量机分类模型需进一步优化等问题^[25-26]。

空间金字塔池化网络(spatial pyramid pooling in deep convolution networks, SPP-Net)是何凯明基于 R-CNN 改进的一种更加快速的目标识别方法。该方法在提取出图像中不同大小的候选区域后,通过预先训练好的卷积神经网络对整幅图像进行卷积得到整幅图像的特征图,将每一个候选区域投影到特征图上,找出候选区域对应的特征,再利用空间金字塔池化层对不同大小的候选区域进行池化,得到固定大小的特征向量并连接到全连接层,最后将全连接层的输出作为分类器的输入进行目标的分类识别^[2,26-29]。由于空间金字塔池化层的存在,使得 SPP-Net 可以处理任意大小的图像。SPP-Net 的目标识别流程如图 3 所示。

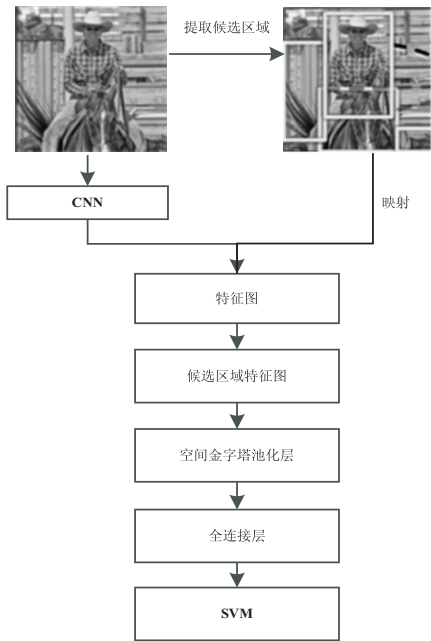


图 3 SPP-Net 目标识别流程

基于 SPP-Net 方法,2015 年 Ross Girshick 提出了一种改进的目标识别方法 Fast R-CNN。该方法将整幅图像通过训练好的卷积神经网络进行卷积,得到整幅图像的特征图,将提取的候选区域投影到特征图上,用简化的空间金字塔池化层(SSP)——感兴趣区域池化层(ROI pooling)提取候选区域的特征并连接到全连接层,得到特征向量,最后将边框回归直接加入到模型中进行训练,用多任务学习的方式同时进行分类和回归^[26,30]。该方法的中间数据不再需要单独存储,测试速度相较于 SPP-Net 方法提高了近 10 倍,在国际标准目标检测数据集 VOC2007 上的平均测试精度为 68%^[25-26]。

虽然 Fast R-CNN 的运行时间相较于 R-CNN 和

SPP-Net 大幅缩短,但候选区域的提取仍是速度提升的瓶颈。为进一步提高算法的运行时间,Girshick 又与何凯明等合作提出了一种 Faster R-CNN 目标识别方法,它由一个区域生成网络 (region proposal networks,RPN) 和 Fast R-CNN 组成。RPN 是一种直接用于提取候选区域的神经网络,与 Fast R-CNN 共享卷积特征,在对整幅图像卷积得到图像的特征图后,将特征图输入到 RPN 中提取出候选区域的特征,经过感兴趣区域池化层 (ROI pooling) 将候选区域特征输出到全连接层,最后使用分类器进行目标的分类识别。该方法通过 RPN 直接提取候选区域特征,进一步提高了目标识别的速度,在 VOC2007 上其平均测试精度达到了 73%^[16,26,31]。Fast R-CNN 和 Faster R-CNN 的目标识别流程分别如图 4(a) 和 4(b) 所示。

YOLO (you only look once) 是 Redmon 在 CVPR2016(计算机视觉领域国际顶会) 上提出的一个端到端的基于深度卷积神经网络的目标识别方法。该方法将目标检测作为回归问题求解,无需提取候选区域,通过将图像划分为固定大小的网格,在每个网格内预测边框内包含目标的置信度和边框区域属于不同类别的概率,直接在整幅图像上回归得到目标边框的位置和所属类别,运行速度有了极大提高。YOLO 的目标识别流程如图 5 所示^[32],与 R-CNN、Fast R-CNN、Faster R-CNN 等方法不同,YOLO 中训练和目标检测在一个单独网络中进行,大大提高了目标检测的速度且背景误检率低,但该目标定位误差较大,当多个目标的中点位于同一网格内时只能识别出一个目标,检测精度低^[26]。

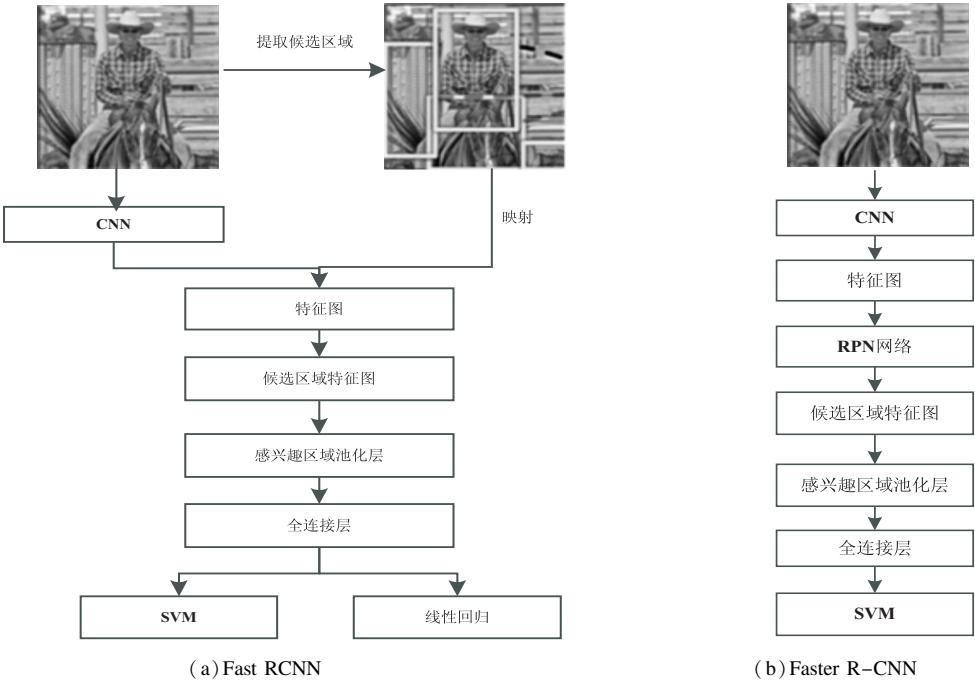


图 4 Fast RCNN 及 Faster R-CNN 目标识别流程

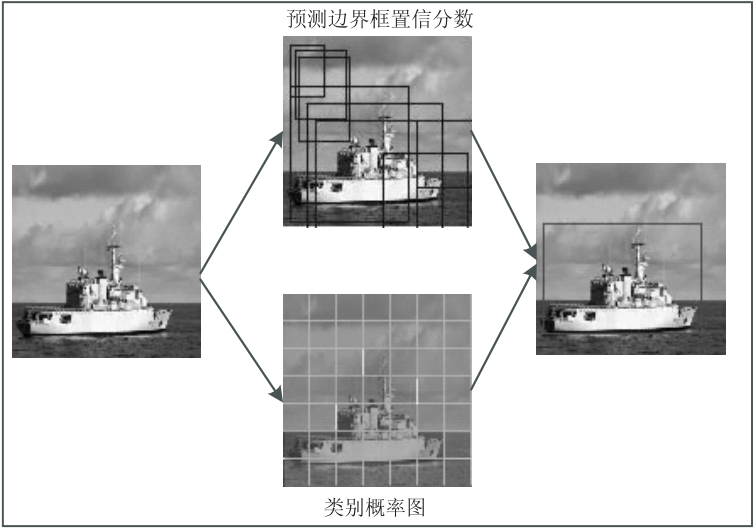


图 5 YOLO 目标识别流程

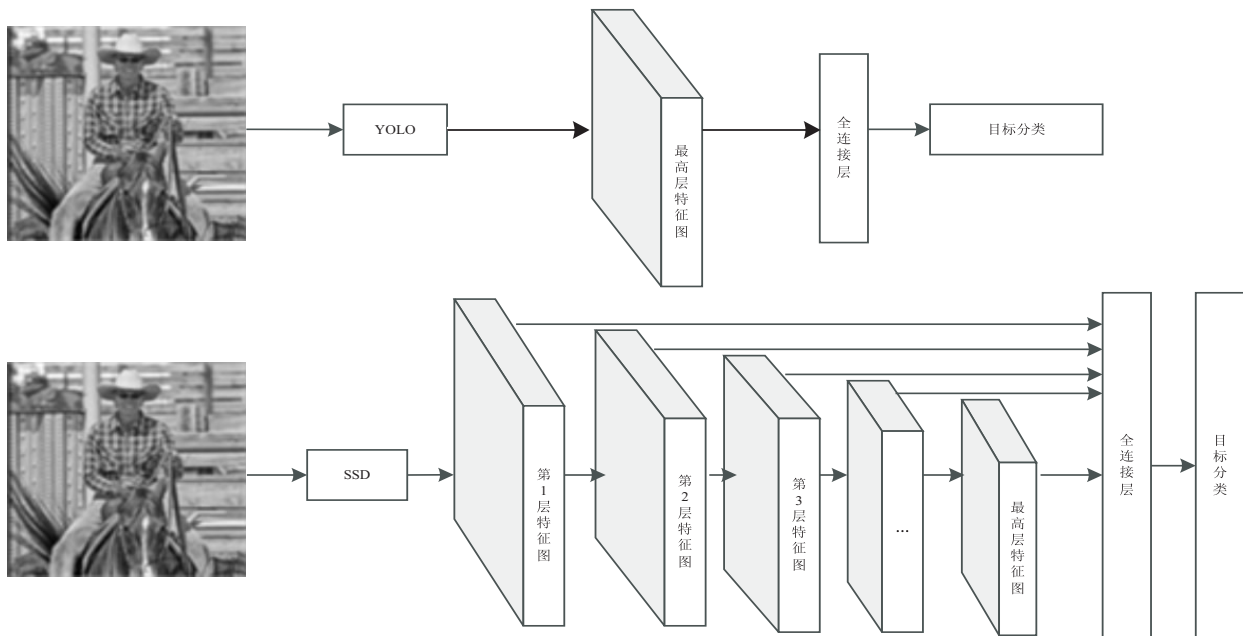


图 6 SSD 与 YOLO 网络结构对比

SSD (single shot multibox detector) 是 Wei Liu 提出的另一种端到端的目标识别方法。该方法借鉴了 Faster R-CNN 中的 RPN 网络和 YOLO 的回归思想, 基于前馈卷积神经网络, 将边界框的输出空间离散化为一组固定尺寸的默认边界框, 在每个默认边界框中预测目标及其归属类别的得分, 同时在不同层次的特征图上使用小的卷积核预测一系列回归边框的位置, 调整边界框从而更好地匹配目标形状, 实现目标的高精度检测^[33]。SSD 与 YOLO 的网络结构对比如图 6 所示。从图中可以看出, YOLO 的卷积层直接与全连接层相连, 在目标分类识别时只用了最高层的特征图, 而 SSD 采用了特征金字塔结构, 将不同层次的卷积层与全连接层相连, 使用低层特征图进行小目标识别, 使用高层特征图识别较大目标, 与 YOLO 相比, 该方法能够在保持检测速度的同时得到高精度的检测结果。

随着硬件性能的提升和计算机技术的进步, 基于深度学习的目标识别新方法不断被提出, 推动着计算机视觉和人工智能在理论和实践中的快速发展。2012 年, Google 公司在 Google Brain 项目中, 通过深度神经网络对 YouTube 上的视频进行无监督的训练和学习, 自动识别出视频中的猫^[17,34]。2013 年, 百度成立了深度学习研究院, 利用深度神经网络构建了百度大脑, 在人脸识别中达到了 99.7% 的正确识别率^[34]。专注于无人驾驶技术研究的 MobileEye、Tusimple 等公司也将深度学习技术引入到车辆目标识别系统中^[17,34]。目前, 深度学习在交通信号识别、行人检测、车辆检测、动物检测、人脸识别等目标识别领域已经得到了广泛的应用。

3 深度学习技术在舰船目标识别领域的应用比较

基于以上分析, 将具有代表性的基于候选区域的 Faster RCNN 目标识别方法和端到端的 YOLO 目标识别方法应用到舰船目标识别当中, 对两种方法得到的舰船目标识别结果进行客观分析, 从而比较不同方法在舰船目标识别领域的优劣性。

由于目前没有公开的用于深度学习模型训练的舰船样本数据集, 文中通过互联网搜集、拍摄船模以及出海采集舰船样本的方法构建了一个包含 20 000 张不同舰船样本的数据集用于模型的训练和学习。随机选择其中 15 000 张图像作为训练集, 剩下的 5 000 张图像作为测试集。训练集和测试集中部分舰船样本图像如图 7 所示。实验的硬件平台为: Intel(R) Core(TM) i5-6300HQ GPU960M@2.30 GHz, 8 GB 内存。

将训练集分别输入到 Faster RCNN 模型和 YOLO 模型中进行训练, 训练过程中, 设定最大循环次数为 2 000 次, 初始学习率设置为 0.01 以快速得到较优的参数值, 循环 500 次后将学习率调整为 0.001, 循环 1 000 次后将学习率调整为 0.000 1 防止模型陷入局部最优。训练完成后, 将测试集分别输入到训练好的 Faster RCNN 模型和 YOLO 模型中, 得到两种目标识别方法对测试集中舰船样本的识别结果, 部分识别结果分别如图 8 和 9 所示。从图中可以看出, 采用两种目标识别方法均能识别出不同种类的舰船目标。

为对两种方法的舰船目标识别结果进行客观评价, 采用目标检测中最常用的两个评价指标(准确率和召回率)对两种方法的识别结果进行比较。其计算

方法如式 1 所示：

$$\begin{cases} P = \frac{TP}{TP + FP} \\ R = \frac{TP}{TP + FN} \end{cases} \quad (1)$$

其中,TP 为正确识别出的舰船目标个数;TP + FP 为被判定为舰船目标的总数;TP + FN 为真实的舰船目标总数。经统计,舰船样本测试集中包含舰船目标总数为 5 256 个。



图 7 训练集及测试集中部分样本图像

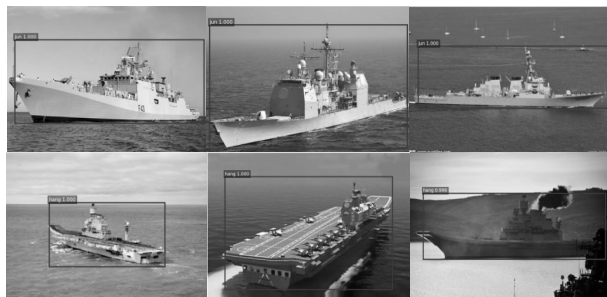


图 8 Faster RCNN 的舰船目标识别结果



图 9 YOLO 的舰船目标识别结果

对两种方法在 5 000 张舰船样本测试集中的识别结果进行统计。Faster RCNN 中,正确识别出的舰船目标个数为 4 585,被判定为舰船目标的总数为 4 931;YOLO 中,正确识别出的舰船目标个数为 4 406,被判定为舰船目标的总数为 4 852。由此可根据式 1 分别计算得出 Faster RCNN 及 YOLO 对舰船目标识别的准确率和召回率,如表 1 所示。

表 1 Faster RCNN 及 YOLO 识别舰船目标的
客观评价结果 %

目标识别方法	精确率 P	召回率 R
Faster RCNN	92. 98	87. 23
YOLO	90. 81	83. 83

此外,Faster RCNN 模型中,舰船样本训练集的训练过程共耗时 17 分 30 秒,测试过程中平均一张图像的检测时间约为 0.07 秒;YOLO 模型对舰船样本训练集的训练时间为 3 分 45 秒,测试过程中平均一张图像的检测时间约为 0.015 秒。

基于以上分析可以看出,采用 Faster RCNN 对舰船目标进行识别的准确率和召回率均比采用 YOLO 识别舰船目标的准确率和召回率高,但是 YOLO 由于直接采用端到端的方法实现舰船目标的识别,其训练时间和测试时间相比于 Faster RCNN 大幅缩短,其检测一张舰船目标图像的平均时间约为 Faster RCNN 检测一张舰船样本图像平均时间的 21.43%。

4 结束语

文中综述了近年来深度学习中具有代表性的深度模型,对基于深度卷积神经网络的几类主流目标识别方法进行了总结,将具有代表性的基于候选区域的 Faster RCNN 目标识别方法和端到端的 YOLO 目标识别方法应用到舰船目标识别领域,对两种方法得到的舰船目标识别结果进行客观分析,结果表明：

- (1)两种方法均能正确识别出舰船目标；
- (2)采用 Faster RCNN 方法得到的舰船目标识别结果,其准确率和召回率均比采用 YOLO 方法得到的舰船目标识别的准确率和召回率高；
- (3)采用 Faster RCNN 方法得到的舰船目标识别结果,其运行效率远低于 YOLO 方法的运行效率。

下一步将针对舰船目标识别领域,综合 Faster RCNN 方法检测目标的准确性和 YOLO 方法检测目标的快速性,提出更加适用于舰船目标识别的检测方法,在目标识别精度和算法复杂度之间取得平衡。

参考文献：

[1] HINTON G E,SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks [J]. Science, 2006,313 (5786) :504-507.

[2] 韩佳容. 基于区域卷积神经网络的目标检测方法研究 [D]. 长沙:国防科技大学,2016.

[3] 付若楠. 基于深度学习的目标检测研究 [D]. 北京:北京交通大学,2017.

[4] 段 萌. 基于卷积神经网络的图像识别方法研究 [D]. 郑州:郑州大学,2017.

[5] HINTON G E,OSINDERO S,TEH Y W. A fast learning algorithm for deep belief nets [J]. Neural Computation,2006, 18 (7) :1527-1554.

[6] SCHÖLKOPF B,PLATT J,HOFMANN T. Greedy layer-wise training of deep networks [C]//International conference on neural information processing systems. Cambridge:

- MIT Press,2006;153–160.
- [7] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278–2324.
 - [8] ELMAN J L. Finding structure in time[J]. Cognitive Science, 1990, 14(2): 179–211.
 - [9] SCHUSTER M, PALIWAL K K. Bidirectional recurrent neural networks[J]. IEEE Transactions on Signal Processing, 2002, 45(11): 2673–2681.
 - [10] GRAVES A, MOHAMED A, HINTON G E. Speech recognition with deep recurrent neural networks[C]//International conference on acoustics, speech and signal processing. Vancouver, BC, Canada: IEEE, 2013: 6645–6649.
 - [11] JAEGER H, HAAS H. Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication[J]. Science, 2004, 304(5667): 78–80.
 - [12] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735–1780.
 - [13] GERS F A, SCHMIDHUBER E. LSTM recurrent networks learn simple context-free and context-sensitive languages[J]. IEEE Transactions on Neural Networks, 2001, 12(6): 1333–1340.
 - [14] CHUNG J, GULCEHRE C, CHO K H, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling[C]//Proceedings of computer vision and pattern recognition. [s. l.]: [s. n.], 2014: 1–9.
 - [15] KOUTNÍK J, GREFF K, GOMEZ F, et al. A clockwork RNN[C]//Proceedings of computer vision and pattern recognition. [s. l.]: [s. n.], 2014: 1863–1871.
 - [16] BENGIO Y, COURVILLE A, VINCENT P. Representation learning: a review and new perspectives[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2013, 35(8): 1798–1828.
 - [17] 王 瑞. 基于卷积神经网络的图像识别[D]. 开封: 河南大学, 2015.
 - [18] 赵 锐. 基于深度卷积神经网络的智能车辆目标检测方法研究[D]. 长沙: 国防科技大学, 2015.
 - [19] ZEILER M D, FERGUS R. Visualizing and understanding convolutional networks[C]//European conference on computer vision. Zurich, Switzerland: Springer, 2014: 818–833.
 - [20] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[C]//Proceedings of computer vision and pattern recognition. [s. l.]: [s. n.], 2014: 1–14.
 - [21] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]//IEEE conference on computer vision and pattern recognition. Boston, MA, USA: IEEE, 2015: 1–9.
 - [22] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//IEEE conference on computer vision and pattern recognition. [s. l.]: IEEE, 2016: 770–778.
 - [23] 逢淑超. 深度学习在计算机视觉领域的若干关键技术研究[D]. 长春: 吉林大学, 2017.
 - [24] 郑 胤, 陈权崎, 章毓晋. 深度学习及其在目标和行为识别中的新进展[J]. 中国图象图形学报, 2014, 19(2): 175–184.
 - [25] 李传朋. 基于机器视觉和深度学习的目标识别与抓取定位研究[D]. 太原: 中北大学, 2017.
 - [26] 李美玲. 浅谈深度学习在目标检测中的发展[J]. 科技风, 2017(24): 237.
 - [27] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of computer vision and pattern recognition. [s. l.]: [s. n.], 2014: 580–587.
 - [28] UIJLINGS J R, SANDE K E, GEVERS T, et al. Selective search for object recognition[J]. International Journal of Computer Vision, 2013, 104(2): 154–171.
 - [29] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 37(9): 1904–1916.
 - [30] GIRSHICK R. Fast R-CNN[C]//IEEE international conference on computer vision. [s. l.]: IEEE, 2015: 1440–1448.
 - [31] REN Shaoqing, GIRSHICK R, 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.
 - [32] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//Proceedings of computer vision and pattern recognition. [s. l.]: [s. n.], 2016: 779–788.
 - [33] LIU Wei, ANGUELOV D, ERHAN D, et al. SSD: single shot multi-box detector[C]//European conference on computer vision. Amsterdam, The Netherlands: Springer, 2016: 21–37.
 - [34] 朱 煜, 赵江坤, 王逸宁, 等. 基于深度学习的人体行为识别算法综述[J]. 自动化学报, 2016, 42(6): 848–857.