

SCTD1.0:声呐常见目标检测数据集

周彦¹ 陈少昌¹ 吴可¹ 宁明强¹ 陈宏昆² 张鹏¹

¹ 海军工程大学电子工程学院 武汉 430033

² 中国人民解放军 92118 部队 浙江 舟山 316000

(986165364@qq.com)

摘要 近年来,卷积神经网络(CNN)在大规模自然图像数据集(如 ImageNet, COCO)中获得了广泛应用,但在声呐图像检测识别领域的应用研究较缺乏,其存在声呐图像目标检测和分类数据集缺乏且水下目标样本往往面临样本稀少、不平衡等问题。针对这一问题,在进行广泛收集声呐图像的基础上,构建了一个完全公开的、可以用于开展声呐图像检测和分类研究的声呐常见目标检测数据集 SCTD1.0,该数据集目前已包含水下沉船、失事飞机残骸、遇难者 3 类典型目标,共计 596 个样本。在 SCTD1.0 的基础上,文中采用迁移学习的方式测试了检测和分类的基准,具体来说:针对检测任务,使用特征金字塔网络对多尺度特征进行组合利用,比较了 YOLOv3, Faster R-CNN, Cascade R-CNN 这 3 种检测框架在本数据集上的性能表现;针对分类任务,对比了 VGGNet, ResNet50, DenseNet 3 种网络的分类性能,分类准确率达到了 90% 左右。

关键词: 声呐图像;数据集;卷积神经网络;迁移学习;检测与分类

中图法分类号 TP391;TB566

SCTD 1.0: Sonar Common Target Detection Dataset

ZHOU Yan¹, CHEN Shao-chang¹, WU Ke¹, NING Ming-qiang¹, CHEN Hong-kun² and ZHANG Peng¹

¹ School of Electronic Engineering, Naval University of Engineering, Wuhan 430033, China

² 92118 Troops of PLA, Zhoushan, Zhejiang 316000, China

Abstract In recent years, convolutional neural networks (CNN) have been widely used in large-scale natural image datasets (such as ImageNet, COCO). However, there is a lack of applied research in the field of sonar image detection and recognition, which suffers from a lack of sonar image target detection and classification datasets and often faces sparse and unbalanced samples of underwater targets. In response to this problem, based on the extensive collection of sonar images, this paper constructs a completely open sonar common target detection dataset SCTD1.0 that can be used for sonar image detection and classification research. The dataset currently contains three types of typical targets: underwater shipwreck, wreckage of crashed aircraft, and victims, with a total of 596 samples. On the basis of SCTD1.0, this paper uses transfer learning to test the benchmarks of detection and classification. Specifically, for the detection task, the feature pyramid network is used to combine and utilize multi-scale features, and the performance of the three detection frameworks YOLOv3, Faster R-CNN, and Cascade R-CNN on this dataset is compared. For classification tasks, the classification performance of the three networks of VGGNet, ResNet50, and DenseNet is compared, and the classification accuracy rate reaches about 90%.

Keywords Sonar image, Dataset, Convolutional neural network, Transfer learning, Detection and classification

1 引言

声呐图像是根据接收回声时间的分布产生的图像,先由发射基阵发射波束并在水中传播,当波束碰到水底或者障碍物等声学传播介质发生变化的界面时,声波会发生反射,接收基阵接收到反射波束后进行转换放大,接着采集系统对信号以图像的形式进行记录显示。声呐图像的检测识别研究在工业、环境以及军事等诸多领域都有广泛的应用价值,例如水底打捞^[1]、沉船目标识别^[2]、鱼类分布跟踪和识别计数^[3-5]、雷区探测^[6]和水雷目标探测^[7]等。声呐图像数据主要都是在水深探测以及底层探测获取数据时一并获取,从而直观地观测水

底浅层结构。侧扫声呐的发射基阵以一定的俯仰角和两个扇面向左右两侧的水中发射声波脉冲信号,而前视声呐的工作原理和侧扫声呐一样,只是声波发射基阵是以一个扇面向前或者向下发射脉冲信号。

近年来,声呐图像识别领域在不断发展,深度学习也广泛应用于图像分类^[8]、检测^[9-11]和图像分割^[12],深度学习概念的提出为声呐图像检测分类提供了一种新思路。Williams^[13]利用深度卷积神经网络进行声呐图像识别,取得了不错的识别效果。Kim 等^[14]设计了一个卷积神经网络分类器模型,并且将这种模型应用到水下航行器上,检测具有较高概率存在的目标,实验结果证实了该方法的有效性。Valdenegro-To-

基金项目:国家自然科学基金(61671461)

This work was supported by the National Natural Science Foundation of China(61671461).

通信作者:张鹏(pengzhang.ai@outlook.com)

ro^[15]针对前视声呐图像将使用卷积神经网络和使用模板匹配方法进行对比,实验结果表明卷积神经网络的性能更优,并且可以用较少的参数来提供更好的性能。Kim等^[16]发现,将卷积神经网络算法用于处理前视声呐图像非常有效,并且将其运用到水下机器人上,水下识别率得到了提高。Valdengro-Toro^[17]开发了一个卷积神经网络,对前视声呐图像的图像窗口进行了客观评分,并通过阈值处理生成检测建议,在它们的海洋垃圾对象数据集中效果很好,甚至该方法还可以检测不在训练集数据中的物体,因此这很有实际应用意义。

若想利用深度学习进行声呐图像的检测分类,则需要用相应的数据集进行训练测试,在其他计算机视觉领域有很多公开数据集作为支撑,如MNIST^[18],ImageNet^[19],PASCAL VOC^[20],MS COCO^[21]等,在合成孔径雷达领域也有MSTAR^[22],SSDD^[23],OpenSARShip^[24],AIR-SARShip-1.0^[25],但是声呐图像的采集需要耗费大量的资源,并且很多声呐图像都涉及到了保密问题,因此目前在声呐这个领域公开的数据集十分有限,这就导致深度学习在声呐图像检测分类的应用受到了极大限制。

为了促进声呐图像检测分类方面的研究,本文构建了声呐常见目标检测数据集(Sonar Common Target Detection Dataset,SCTD),并提取了衍生分类数据集(SCTD-C,Sonar Common Target Detection Dataset for Classification),它可以用于训练和测试检测分类识别算法,让研究人员在同一条件下对比算法的性能¹⁾。

2 数据集的准备和衍生分类数据集的提取

在选择数据对象类别时需要考虑通用性,这些类别还必须具有一定的代表性,并且与实际应用息息相关,出现的频率应足够高。作为水下声呐图像,最常用的还是打捞搜救,因此目标多为沉船、失事飞机和人,只需要判断出对象属于哪一大类即可满足需求,不再需要细分类别或区分个体,因此本文只要按照跨物种语义级别的图像检测分类来搜集构建数据集即可。

2.1 检测数据集

在声呐自动目标识别研究领域,由于水下样本稀缺,目前仍然缺乏公开的检测与分类数据集,近期,Jiang等^[26]提出了一种基于主动学习的声呐图像目标检测框架,并收集了216张声呐图像进行检测数据集标注,但是该标注方法适用于主动学习场景并且数据集没有公开。鉴于此,本文自行制作了声呐常见目标检测数据集SCTD。相比文献^[26]的工作,我们的优势在于:1)SCTD样本更为丰富,本文收集了497张分辨率较高的图像(包含596个目标,它的目标数量分布图如图1所示),并且同时包括侧扫声呐图像、合成孔径声呐图像、干涉合成孔径声呐图像、前视声呐图像;2)在数据标注方面,采用的是多组手工标注以减少人为偏差,并且提供了数据格式较为规范、应用广泛的PASCAL VOC和MS COCO格式,因为这两种数据格式比较规范,很多算法都是在这两种数据集上进行处理,这样对现有的算法稍加改动就可以直接在SCTD数据集上进行处理;3)我们将SCTD作为一个持久的

公开的项目以持续对其进行扩充,并提供开放下载。关于数据集的详细信息以及典型方法的基准可以参考项目主页。

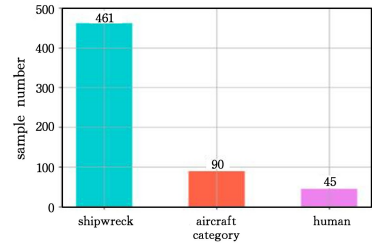


图1 各类目标样本数量的分布直方图

Fig. 1 Histogram of the distribution of various target sample numbers

雷达自动目标识别研究领域相对于声呐更活跃,并且有公开的数据集可以使用,例如合成孔径雷达舰船目标检测数据集(SSDD),SSDD属于单类目标检测数据集,且为灰度图像,不易探究检测器在特征提取、目标分类方面的优势。相比数据集SSDD而言,SCTD是一个更难的检测任务,主要体现在以下4个方面:

(1)SCTD目前收集到的有效的声呐图像数量是SSDD的一半左右,过拟合风险相比SSDD更大。

(2)SSDD只有单类舰船目标,SCTD目前收集了3类目标图像,并且类间样本不平衡,相对SSDD更为严重。

(3)由于成像声呐探测距离、成像方法等与合成孔径雷达图像的差异,SCTD图像分辨率相对SSDD更高,因此目标细节特征更加明显,如粉碎、断裂、变形等干扰检测的情况更加恶劣,如图2所示。

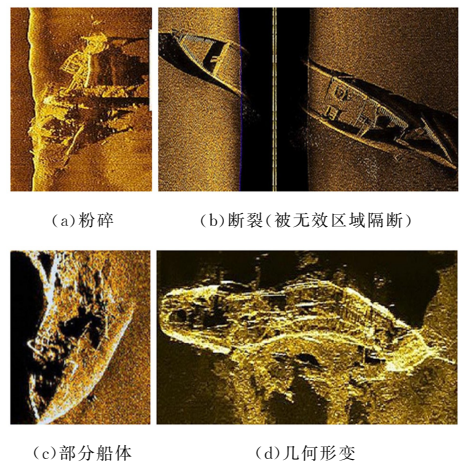


图2 典型声呐图像目标检测困难样例

Fig. 2 Hard examples of target detection in typical sonar images

(4)SCTD图像数据来源多样,图像渲染方式复杂,图像尺寸、分辨率以及长宽比差异较大,而SSDD图像统一采用灰度图显示且图像尺寸基本一致。

使用LabelImg标注软件,按照PASCAL VOC数据集的格式对图像进行标注,标注的框为水平矩形框,保存的文件格式为xml格式,xml文件中包含了图像的各种信息,包括文件名、文件路径、图像像素大小及通道数、标注目标的类别及位置等信息,标注的图像及xml文件的内容如图3(a)所示(由

¹⁾ <https://github.com/freepoet/SCTD>.

于页面关系只展示部分)。在搜集的图像中,出现了一张图像中包含多个目标的情况,如图 3(b)所示,图中有多个沉船目标,为了保持单标签图像分类的规范,需要对标注完的数据进行进一步的处理,使每张图像中只有一个目标。

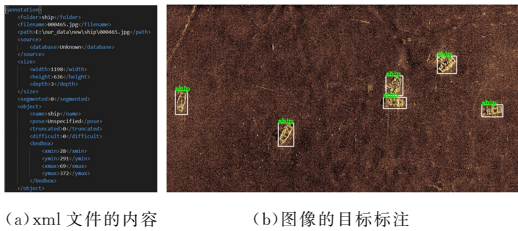


图 3 图像的目标标注及 xml 文件内容

Fig. 3 Target annotation of the image and the content of the xml file

2.2 衍生分类数据集的提取

从 SCTD 中根据标注提取衍生分类数据集,在数据集划分方面,本文进行数据集划分的原则为确保网络的训练仅使用检测数据集的训练集,另外从图 3(b)中还发现图中目标占整个图像的比例很小,图像中大部分的空间都被背景占据,因此把背景这个负样本单独作为一类,具体划分流程为:

(1)将检测数据集划分为训练、验证、测试数据集(7:2:1),用于检测器的整体训练;

(2)仅从检测数据集的训练集提取目标图像块(Target Patch)、背景图像块(Background Patch),构成衍生分类数据集的训练集和验证集,按照 8:2 进行划分。

在目标图像块和背景图像块的区域提取方面,目标区域的提取是简单的,但背景块的提取则需要对非矩形的不规则的背景区域进行特殊处理。区域提取程序设计的核心原则为保证目标区域与背景区域完全分离,并且尽可能地保留背景区域。因此,首先将原始框标注放大 1.1 倍进行目标区域提取,然后基于以下规则对背景区域进行划分:1)长宽比不超过 3:1;2)最短边不小于目标块最短边的一半;3)最长边不超过目标块的最长边的 2 倍。

按照上述要求,从 SCTD 的训练集中提取得到了衍生分类数据集 SCTD-C。在衍生数据集上,训练集(train-C)和验证集(val-C)中的样本数量如表 1 所列。

表 1 SCTD-C 样本数量

Table 1 SCTD-C sample size

Category	train	val
shipwreck	295	83
human	38	9
aircraft	72	18
background	7 129	1 790

以 SCTD-C 数据集为例,提取后的 3 类目标块和背景区域如图 4 所示。

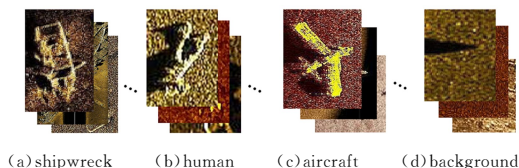


图 4 SCTD-C 提取的目标和背景图像块

Fig. 4 Target and background image blocks extracted by SCTD-C

从表 1 中发现,衍生分类数据集样本严重不平衡,background 样本数量达到 8919,而 human 样本数量却只有 47,这在分类模型中会严重影响分类模型的训练及预测结果。查看 background 类的样本发现,很多样本尺寸过小、图像长宽比不合理、存在无效区域、存在不明物体、图像上有文字等情况,如图 5 所示。

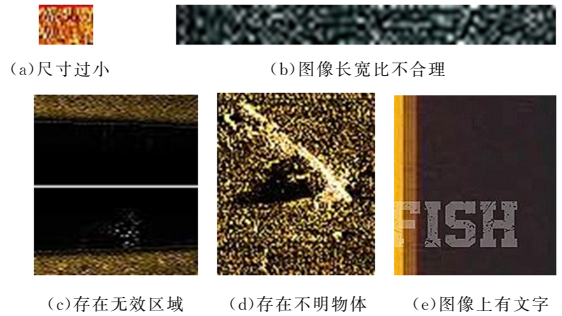


图 5 background 类中的不合理样本

Fig. 5 Unreasonable samples in background class

进一步对 background 类进行筛选,去掉上述情况中不合理的样本,保留不含任何不明物体的单一 background 类,最后剩余 background 类样本数量为 1000,按照 8:2 的比例划分训练集和验证集。

声呐图像与通用光学图像在成像机制、分辨率、包含信息等方面差异巨大,图像中目标的旋转和几何畸变严重地影响了检测效果。针对此问题,本文为声呐数据集设计了特定的数据增强方案。为了提高对旋转和形变的鲁棒性,在检测器训练时,对检测数据集采用随机裁剪(对输入图像随机切割掉一部分)、翻转变换(沿着垂直或者水平方向翻转图像)等,在衍生分类数据集上进行训练时,对目标图像块进行随机旋转(对输入图像随机旋转一定角度)和随机偏移(输入图像沿着水平或者垂直方向偏移某个比例)、随机裁剪、随机缩放(按照一定的比例放大或者缩小图像)等。

表 2 实验中使用的数据增强方法

Table 2 Data enhancement methods used in the experiment

检测训练数据增强方法	分类实验数据增强方法
随机裁剪	随机旋转
随机垂直翻转	随机水平偏移
随机水平翻转	随机垂直偏移
	随机裁剪
	随机缩放

3 SCTD 上的检测过程

深度学习目标检测器分为特征提取部分与预测部分。特征提取使用剪枝后的卷积网络作为骨干网络进行特征提取,在多尺度特征利用方面,考虑到声呐图像目标尺度变化大的问题,可以使用特征金字塔网络(Feature Pyramid Network, FPN)^[27]进行多尺度特征组合。

本节在自行构建的 SCTD 数据集上进行了检测器的目标检测实验,采用主流的检测器框架在相同实验环境下进行了对比,目前主流的检测算法有三大类:单阶段、双阶段以及多阶段。在每类算法中各选取一种具有代表性的算法作为检测器的框架,它们分别是 YOLOv3^[10],Faster RCNN^[9],Cascade RCNN^[28]。网络架构如图 6 所示,采用相同的数据增强策略和训练次数。在评价指标上,分别选用目标召回率(Recall)和

mAP50 指标^[29](下文中称为 mAP, mAP 指标在 $[0,1]$ 之间,该值越高说明预测出的边界框以及物体类别就越准确)。利用在 ImageNet 上预训练的 ResNet50^[30]骨干网络搭建 3 类检测器(选择 ResNet50 的原因在于其进行声呐目标识别时可以在准确度、模型复杂度与计算复杂度上做到很好的权衡)。进行实验时所有检测器的优化超参数设置相同:优化器均使用动量为 0.9 的随机梯度下降法,并且初始学习率(initial learning rate)均设置为 0.01,学习率进行预热的迭代次数(warm up iteration)为 500,权重衰减率(weight decay rate)为 0.0001,训练轮数为 36。

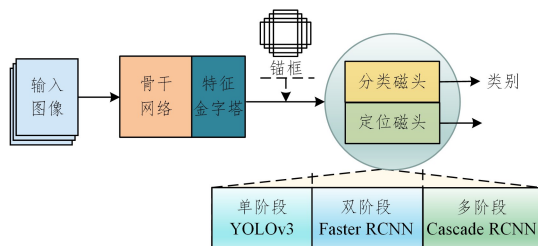


图 6 3 种基于深度学习的检测框架的典型架构

Fig. 6 Typical architectures of three detection frameworks based on deep learning

但是,在实验中发现采用统一的学习率无法取得较优的检测性,甚至导致检测器的训练失败。经过大量实验,将 YOLOv3 检测框架的学习率设置为 0.02, Faster RCNN 检测框架的学习率设置为 0.0001, Cascade RCNN 检测框架的学习率设置为 0.01,采用 R_1, R_2, R_3 分别表示检测器对 aircraft, human, shipwreck 3 类目标的召回率,使用 AP_1, AP_2, AP_3 表示对应的检测精度,可以得到这些检测器对各类目标的召回率、检测精度如表 3 所列,由于 SCTD 数据集中各类目标存在严重的样本量不平衡问题,表 3 中的 AR 和 mAP (AR 和 mAP 为平均召回率和平均检测精度)是加权平均得到的,即:

$$AR = P_1 \cdot R_1 + P_2 \cdot R_2 + P_3 \cdot R_3 \quad (1)$$

$$mAP = P_1 \cdot AP_1 + P_2 \cdot AP_2 + P_3 \cdot AP_3 \quad (2)$$

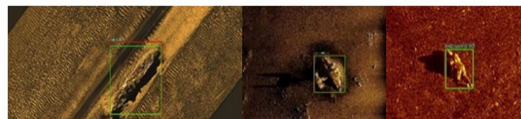
式(1)、式(2)中, P_1, P_2, P_3 分别为不同类别目标样本数在测试集中所占的比例, AR 和 mAP 的最终结果如表 3 所列。

表 3 检测器的召回率和检测精度

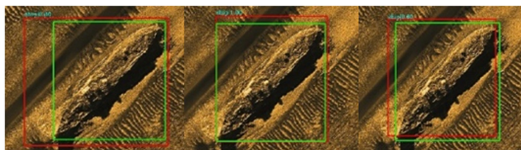
Table 3 Recall rate and detection accuracy of detectors

Method	R_1	AP_1	R_2	AP_2	R_3	AP_3	AR	mAP
YOLOv3	0.944	0.903	1.000	1.000	0.904	0.857	0.917	0.874
Faster RCNN	0.944	0.688	0.778	0.729	0.807	0.714	0.811	0.713
Cascade RCNN	0.889	0.345	0.444	0.210	0.819	0.610	0.766	0.529

从表 3 中可以发现,越是复杂的检测器在 SCTD 上的检测效果反而越差,3 类检测器的性能排序为: Cascade RCNN < Faster RCNN < YOLOv3。针对这一现象,首先尝试了多组不同的实验超参数,但仍然无法取得较好的结果,目前合理的解释可以是 SCTD 样本数较少且各类样本数极其不平衡。分别在测试集中选取不同类别的大、中、小型目标进行检测,边界框预测的效果如图 7 所示。



(a) 大、中、小 3 类目标检测效果



(b) 沉船图像的检测效果



(c) 失事飞机图像的检测效果



(d) 人图像的检测效果

注:绿色框为标注框,红色框为检测器的预测输出框,图 7(b)~图 7(d) 中的图像从左至右依次是 Cascade RCNN, Faster RCNN, YOLOv3 的预测输出

图 7 检测器的检测效果图(电子版为彩色)

Fig. 7 Detection effect of detector

4 SCTD-C 上的分类过程

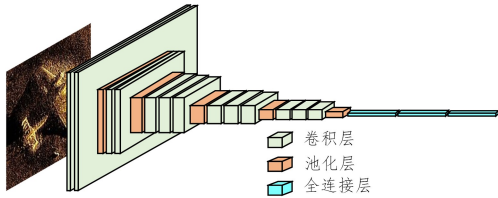
图像分类主要是研究图像的特征描述,深度学习利用设计好的网络结构,完全从训练数据中学习图像的层级结构性特征,其强大的学习能力可以提取数据的深层特征,能够学习更加接近图像高级语义的抽象特征,因此深度学习网络在图像分类任务上表现优异。

本节将在 SCTD 数据集的衍生分类数据集 SCTD-C 上进行目标分类训练,在相同的实验环境下采用经典的网络模型进行对比,经典网络模型中 VGGNet^[31], ResNet50^[30], DenseNet^[32]是比较有代表性的。VGGNet 结构简洁,通过反复堆叠卷积池化层构建网络,验证了可以通过加深网络深度来提升网络性能;ResNet50 通过设计残差网络结构,解决了随网络深度加深而准确率下降的问题;DenseNet 的基本思路与 ResNet50 一致,但是 DenseNet 是将前面所有层与后面层建立密集连接。它们的网络结构如图 8 所示,采用相同的数据增强策略,训练轮数都为 100,评价指标选用准确率(Accuracy),该值越高说明预测出物体所属的类别就越准确,优化器均使用 Adam 优化算法,并且初始学习率的设置为 0.001,每次参数更新后学习率的衰减值(decay)为 0.00001, batch_size 均为 32。

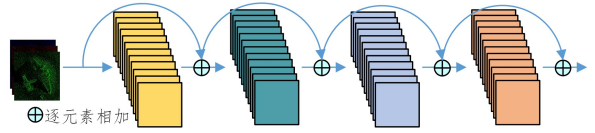
在实验过程中发现, VGGNet 网络在训练过程中训练数据集的准确率呈上升趋势,而验证数据集的准确率却始终维持在一个数值不变。查阅资料后通过分析可知,可能是网络出现了过拟合的现象,因此在原 VGGNet 网络结构的基础上进行了修改,增加了正则化项,具体的操作就是在每个卷积层

的激活函数后增加一个批归一化 (Batch Normalization)^[33] 层,它可以加快网络的训练和收敛速度,防止梯度爆炸、梯度

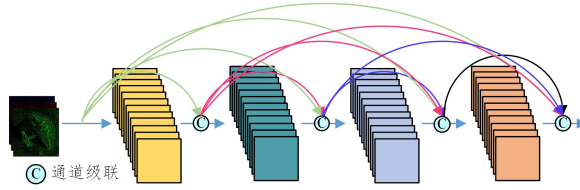
消失和过拟合。3个网络训练后的预测结果的准确率如表4所列。



(a) VGGNet 网络堆叠连接机制



(b) ResNet50 网络残差结构连接机制



(c) DenseNet 网络密集连接机制

图8 3种网络模型结构连接机制

Fig. 8 Three network model structure connection mechanisms

表4 分类网络的预测结果准确率

Table 4 Accuracy of prediction results of classification network

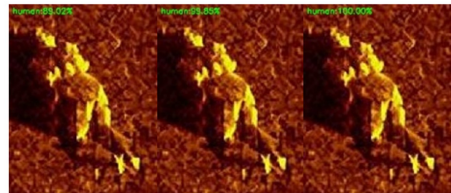
分类实验网络	预测结果准确率/%
VGGNet	89.24
ResNet50	92.01
DenseNet	93.40

从表4可以发现,3个分类网络的性能排序为:VGGNet<ResNet50<DenseNet,这与这3个网络在其他公开图像数据

集上的性能表现是相符的。一般随着网络层数的加深,获取的信息会越多,特征也越丰富,但实际实验表明准确率不会上升,效果反而会变差,ResNet50残差结构的提出使输入输出之间建立了一条直接关联的通道,以VGGNet网络结构为基础堆叠残差结构,增强了特征的表达能力,DenseNet更是将这种关联机制变成密集连接,特征的表达力进一步得到了增强。分别选取测试集中不同类别的目标进行预测分类,结果如图9所示。



(a) 失事飞机图像的预测分类效果



(b) 人图像的预测分类效果



(c) 沉船图像的预测分类效果



(d) 背景图像的预测分类效果

注:绿色文字为图像所属类别及预测为所属类别的概率,图9(a)~图9(d)中的图像从左至右依次是VGGNet、ResNet50、DenseNet的预测输出

图9 各分类网络的预测分类效果

Fig. 9 Prediction classification effect of each classification network

结束语 针对公开的声呐数据集较少的问题,本文构建了声呐常见目标检测数据集并提取了衍生分类数据集,通过搜集整理数据,并对数据进行相应的处理,得到一个用于检测和多分类的目标数据集,通过设计检测网络,比较了3种不同检测框架在本数据集上的性能,并在衍生分类数据集上比较了3种经典网络的分类性能,通过实验为检测和分类数据集分别建立了性能基准。接下来的工作要继续搜集数据,扩充样本数量,并解决数据集不平衡的问题,另外实验中在本数据集上运行的卷积神经网络的参数量较大,不利于在移动端上部署,因此接下来可以考虑对神经网络进行轻量化的改进。

参考文献

- [1] LUO X B, XU D M, HU J J, et al. Application Research of 3D Imaging Sonar System in Salvage Process[J]. Applied Mechanics and Materials, 2014, 643(8): 279-282.
- [2] DONG L Y, SHAN R, LIU H M, et al. Survey Recognition Method of Side Scan Sonar Image Based on Fractal Texture Features[J/OL]. Marine Geology and Quaternary Geology: 1-8. [2021-02-10]. <https://doi.org/10.16562/j.cnki.0256-1492.2020070301>.
- [3] EILER J H, GROTHUES T M, DOBARRO J A, et al. Comparing autonomous underwater vehicle (AUV) and vessel-based

- tracking performance for locating acoustically tagged fish[J]. *Marine Fisheries Review*, 2013, 75(4): 27-42.
- [4] SHEN J Y, LI L Y, DAI Y L, et al. Fish detection and monitoring system based on YOLO algorithm[J]. *Journal of Suzhou University of Science and Technology (Natural Science)*, 2020, 37(3): 68-73.
- [5] SHEN W, ZHU Z H, ZHANG J, et al. Fish target recognition and counting based on dual-frequency recognition sonar[J]. *Fishery Modernization*, 2020, 47(6): 83-89.
- [6] DOBECK G J, HYLAND J C, SMEDLEY L, et al. Automated detection and classification of sea mines in sonar imagery[J]. 1997, 3079: 90-110.
- [7] HU H B, MEI X H. A brief description of mine target detection technology based on sonar image[J]. *Digital Ocean and Underwater Warfare*, 2020, 3(4): 303-308.
- [8] KRIZHEVSKY A, SUTSKEVER I, HINTON G. ImageNet Classification with Deep Convolutional Neural Networks[J]. *Advances in Neural Information Processing Systems*, 2012, 25(2): 84-90.
- [9] 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.
- [10] REDMON J, FARHADI A. YOLOv3: An Incremental Improvement[J]. *Computer Vision and Pattern Recognition*, 2018, 4(2): 1-6.
- [11] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single Shot MultiBox Detector[C]// *European Conference Computer Vision*, 2016.
- [12] KAIMING H, GEORGIA G, PIOTR D, et al. Mask R-CNN[C]// *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2017.
- [13] WILLIAMS D P. Underwater target classification in synthetic aperture sonar imagery using deep convolutional neural networks[C]// *2016 23rd International Conference on Pattern Recognition (ICPR)*. IEEE, 2016.
- [14] KIM J, YU S C. Convolutional neural network-based real-time ROV detection using forward-looking sonar image[C]// *Autonomous Underwater Vehicles*. IEEE, 2016: 396-400.
- [15] VALDENEGRO-TOROM. Object recognition in forward-looking sonar images with Convolutional Neural Networks[C]// *Oceans*. IEEE, 2016.
- [16] KIM J, CHO H, PYO J, et al. The convolution neural network based agent vehicle detection using forward-looking sonar image[C]// *OCEANS 2016 MTS/IEEE Monterey*. 2016.
- [17] VALDENEGRO-TORO M. Objectness Scoring and Detection Proposals in Forward-Looking Sonar Images with Convolutional Neural Networks[C]// *IAPR Workshop on Artificial Neural Networks in Pattern Recognition*. Springer International Publishing, 2016, 21(3): 209-219.
- [18] 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.
- [19] DENG J, DONG W, SOCHER R, et al. ImageNet: A large-scale hierarchical image database[C]// *IEEE Conference on Computer Vision & Pattern Recognition*. IEEE, 2009.
- [20] EVERINGHAM M, GOOL L V, WILLIAMS C K I, et al. The Pascal Visual Object Classes (VOC) Challenge[J]. *International Journal of Computer Vision*, 2010, 88(2): 303-338.
- [21] LIN T Y, MAIRE M, BELONGIES, et al. Microsoft COCO: Common Objects in Context[J]. Springer International Publishing, 2014, 86(93): 740-755.
- [22] <https://www.sdms.af.mil/datasets/mstar/>.
- [23] LI J, QU C, SHAO J. Ship detection in SAR images based on an improved faster R-CNN[C]// *Sar in Big Data Era: Models, Methods & Applications*. IEEE, 2017.
- [24] HUANG L, LIU B, LI B, et al. OpenSARShip: A Dataset Dedicated to Sentinel-1 Ship Interpretation[J]. *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing*, 2018, 11(1): 195-208.
- [25] SUN X, WANG Z R, SUN Y R, et al. AIR-SARShip-1.0: High-resolution SAR ship detection dataset[J]. *Journal of Radars*, 2019, 8(6): 852-862.
- [26] JIANG L, CAI T, MAQ, et al. Active Object Detection in Sonar Images[J]. *IEEE Access*, 2020, 99(4): 1-14.
- [27] LIN T Y, DOLLARP, GIRSHICK R, et al. Feature Pyramid Networks for Object Detection[C]// *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, 2017.
- [28] CAI Z, VASCONCELOS N. Cascade R-CNN: Delving into High Quality Object Detection[C]// *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2018.
- [29] CHEN K, WANG J, PANG J, et al. MMDetection: Open MMLab Detection Toolbox and Benchmark[J]. *Computer Vision and Pattern Recognition*, 2019, 6(7): 1-11.
- [30] HE K, ZHANG X, REN S, et al. Deep Residual Learning for Image Recognition[C]// *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2016.
- [31] SIMONYAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. *Computer Science*, 2014, 23(7): 1-14.
- [32] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely Connected Convolutional Networks[C]// *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017.
- [33] IOFFE S, SZEGEDY C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift[J]. *Computer Science*, 2015, 12(27): 1-11.



ZHOU Yan, born in 1991, postgraduate. His main research interests include deep learning and computer vision.



ZHANG Peng, born in 1996, Ph.D. His main research interests include research on rotation pattern recognition of convolutional neural network and automated deep learning.