

基于 SSD 的桥梁主动防船撞目标检测方法与应用

夏焯,陈李沐,王君杰,孙利民[†]
(同济大学 土木工程学院,上海 200092)

摘要:针对现有桥梁主动防撞视频监测系统在精度、鲁棒性和效率等方面的局限性,提出基于 SSD(Single Shot MultiBox Detector)目标检测的桥梁主动防撞视频监测系统框架,建立并使用海量船舶图像数据集对 SSD 模型进行神经网络训练,生成适用于复杂环境下的船舶目标检测模型,通过样本选取、数据增强、广域划分优化等技术手段,实现航区船舶目标的稳定、智能检测.与传统方法相比,提出的船舶目标检测方法在稳定性、准确性、高效性和智能化方面优势明显.以上海松浦大桥为工程背景进行了实桥试验,验证了该方法在桥区航道长期稳定监测的可行性,为进一步的航迹跟踪预测、防撞预警和航道监控提供信息支持.

关键词:主动防船撞;视频监控;神经网络;目标检测;松浦大桥;危险源辨识

中图分类号:U447

文献标志码:A

Single Shot MultiBox Detector Based Vessel Detection Method and Application for Active Anti-collision Monitoring

XIA Ye, CHEN Limu, WANG Junjie, SUN Limin[†]
(College of Civil Engineering, Tongji University, Shanghai 200092, China)

Abstract: Single Shot Multibox detector based video surveillance system for active anti-collision between vessel and bridge is proposed aiming at the limitations of the existing systems in accuracy, robustness and efficiency. A vessel-exclusive dataset with tons of image samples is established for neural network training, and an SSD (Single Shot MultiBox Detector) based object detection model with both universality and pertinence is generated with tactics of sample filtering, data augmentation and large-scale optimization, which can realize the stable and intelligent vessel detection. Comparison results with conventional methods indicate that the proposed navigational object detection method shows remarkable advantage in robustness, accuracy, efficiency and intelligence. In-situ test is carried out at Songpu Bridge in Shanghai, and the results illustrate that the method is qualified for long-term monitoring and provides information support for further analysis and decision making.

Key words: active anti-collision, video surveillance, neural networks, object detection, Songpu Bridge, hazard source identification

* 收稿日期:2019-01-13

基金项目:国家重点研发计划资助项目(2018YFC0809606), National Key R&D Program of China(2018YFC0809606);上海市科学技术委员会资助项目(19DZ1203004), Science and Technology Commission of Shanghai Municipality(19DZ1203004)

作者简介:夏焯(1983—),男,湖南永州人,同济大学副研究员,博士

[†] 通讯联系人, E-mail: lmsun@tongji.edu.cn

船撞桥事故在桥梁事故中占比极高,对桥梁公路安全运输、人民的生命财产乃至社会经济的发展均造成重大影响。随着经济持续增长和对交通设施的需求日益加大,我国桥梁的建设数目不断增加,大型桥梁的建设可以缓解交通压力、促进陆路交通。但是对于水上船舶而言,桥梁却是人工障碍物,船舶在桥下航行通行时,存在着碰撞桥墩或桥跨结构的危险,从而对桥梁、船舶的安全性构成威胁,同时对航道和陆路交通的正常运行也构成威胁。

传统的桥梁主动防撞一般通过船舶交通管理系统(VTS)、船舶自动识别系统(AIS)建立桥梁船撞预警来引导船舶的航行,通过视频监控系统(CCTV)监控桥区现场情况。VTS、AIS需要安装大量昂贵的设备,造价较高,而传统CCTV又无法实现主动预警的功能。近年来,基于视频目标检测的桥梁防撞主动预警技术受到部分学者的关注和研究,该技术通过将视频目标检测结合CCTV,实现对桥区水域主动预警的功能。然而,现有的桥梁防撞视频监测系统大多采用基于运动的目标检测技术^[1,2],既依赖背景环境的稳定性,又无法对伪目标进行智能甄别,面对复杂多变的航道环境,其适应性差、智能化程度低的缺点凸显。

本文提出基于视频传感的桥梁主动防撞监测预警系统框架,阐述航区目标检测的具体流程和技术方法。相较于传统基于运动目标检测方法的监测系统,深度学习目标检测方法有潜力改善现有技术,克服现有方法在复杂环境下的适应性问题 and 目标的智能识别问题,使得该系统能实现复杂且多变环境下稳定的船舶目标智能检测,具备可靠的实用性,推动主动防船撞系统的技术发展。

1 技术框架和研究方法

本文提出基于视频传感桥梁主动防撞监测预警系统,由视频获取、目标检测、航迹跟踪、航迹预测、风险评估、预警决策六个模块组成,技术框架图如图1所示。

视频获取模块通过桥载监控摄像头获取视频图像信息,并进行图像增强等预处理,传入目标检测模块;目标检测模块使用预先训练好的卷积神经网络模型检测画面中的船舶目标,根据空间投影关系将目标的像素坐标转换为世界坐标,传入航迹跟踪模块;航迹跟踪模块在时间维度上进行目标关联形成航行轨迹,并自动检查和去除伪航迹;航迹预测模块

将目标当前航迹输入神经网络,输出对目标将来航迹、抵桥时间和通过位置概率分布的预测,并通过对在线学习的机制基于累积的航迹数据不断更新神经网络模型;风险评估模块参考航迹跟踪和预测的结果,对船桥撞击风险进行综合评估;预警决策模块根据风险评估结果做出相应的决策,如在检测到危险的情况下对船舶进行警示和引导,在无可避免的撞击发生前通过中止上桥方向交通等方式避免进一步的人员和财产损失,在撞击发生后对桥梁状态进行撞后评估,从而对是否通行或维修作出决策,并且对撞击过程进行录像取证。

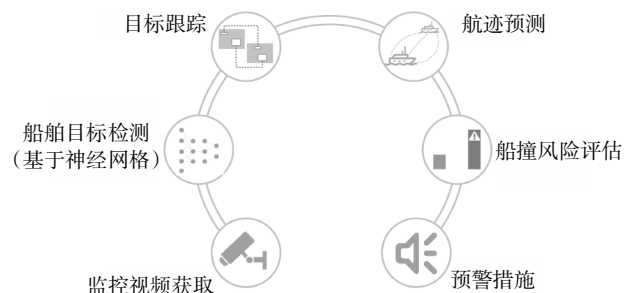


图1 基于视频传感的桥梁主动防撞监测预警系统框架

Fig.1 System framework for anti-collision monitoring

目标检测是整个防撞监测系统的核心和基础,识别准确性及稳定性影响整个系统的成败。本文将针对航区目标检测的特点和需求,通过模型选取、数据集建立、模型训练和针对性优化,研究该模块的技术方法,为后续跟踪预警模块提供稳定有效的信息支持。

2 船舶目标检测方法

2.1 神经网络模型

2.1.1 方法比较

目标检测的核心任务在于找到图像中感兴趣的目標,并确定目標在图像中的位置和大小,航区目标的检测具有以下特点:其一是航区的背景环境通常较为复杂,波浪、光照变化、水汽、阴影倒影等均可能对目标检测形成干扰;其二是船舶的种类繁多,外观、大小不一而足;其三是出于预警的需求,监测范围较大,画面中的远处目标较小;其四是目标交会、重叠的情况时有发生,对图像识别造成困扰。上述特点使得航区目标的检测面临其它环境下未有的挑战。现有桥梁主动防撞视频监测系统大多采用基于运动的目标检测方法,基本原理是提取视频图像序列中目标的运动特性,将运动目标与静止背景进行分割,

从而实现目标检测.根据分割原理不同,常用的基于运动的目标检测方法可分为三类:光流法^[3]、帧间差分法和背景减除法^[4].然而,基于运动的目标检测算法的缺点显而易见:依赖于“静止”的背景,无法适应快速变化的背景,而航道受光照、风等的影响,其水面往往伴随剧烈的变化;检测出的目标仅仅是“运动”的物体,而无法甄别该物体是否为航区船舶,并且无法识别静止的船舶;运算速度较慢,无法充分利用高分辨率监控图像的像素信息.上述的缺点致使现有桥梁防撞视频监测系统缺乏应对复杂环境的稳定性,智能化程度低,无法真正适应长期监测的需求.

基于特征的目标检测方法是另一种主流目标检测方法,近年来随着深度学习技术的飞速发展,逐渐开始渗透到各个应用领域,其在桥梁防撞监测领域亦有开发前景.基于特征的目标检测算法是对目标的颜色、轮廓等特征进行提取与判断,从而实现对检测目标的识别与定位.基于特征的目标检测算法经历了依靠人工设计特征的早期模式到使用卷积神经网络进行特征提取的发展阶段.传统的机器学习算法基于人工设计的特征,如 Haar 特征、LBP 特征和 HOG 特征^[5],这种依赖先验知识的方法面对目标多样性物体外形往往泛化能力不足,无法满足外观多样的船舶目标的检测需求;而基于深度学习的目标检测方法^[6-8]通过训练从庞大的数据集提取和学习特征,对物体外观多样性、角度、形变以及环境和光照的变化均有强大的适应能力,并且在检测的同时能够实现目标类型的智能识别,能够较好地适应航区目标检测的特殊性.

SSD 是目前较为成熟的一种深度学习目标检测算法,除上述优势外,其将特征提取、边框回归和分类在一个无分支的卷积网络中完成,在保证精度的同时达到极高的检测速度,能够有效满足航区监测范围大、目标数量多、检测精度和信息处理效率要求高的需求,因此,选择 SSD 作为航区目标检测模块的模型基础.

2.1.2 SSD 模型

SSD 模型即 Single Shot MultiBox Detector, 由 Wei Liu 等人于 2016 年首次提出^[9].相比同类型网络^[10,11],其最大特点在于使用多尺度特征图检测,适应不同大小物体的检测,能够满足航区目标尺度多样性的需求.

SSD 模型的架构如图 2 所示^[9],网络的主体由基础网络和附加特征网络组成.图像缩放成 300×300

标准尺寸后前向传播通过基础网络;基础网络由高质量图像分类网络的标准架构(如 VGG、GoogleNet、AlexNet 等)在分类层之前截断形成,其作用是进行特征提取形成特征图,本文以 VGG-16 网络作为基础网络,如该图所示;在经过基础网络的特征提取后,进入辅助网络结构,辅助特征网络是一种卷积金字塔结构,各层的特征图尺寸逐渐减小,包括基础网络的末层在内,特征图大小从 38×38 到 1×1 不等,上述诸层通过卷积滤波器,生成目标物体框和类别的预测;最后通过非极大值抑制过滤过度重合的结果,得到最终的输出.

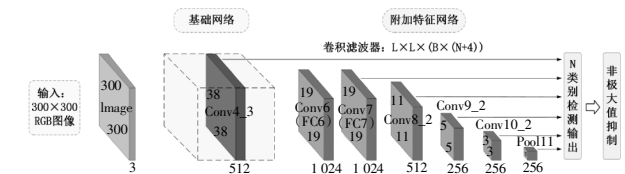


图 2 SSD 网络架构
Fig.2 SSD framework

卷积滤波器的工作原理如图 3 所示^[9],特征图的每个像素都有不同大小和宽高比的默认框,使用卷积滤波器对特征图每个像素默认框的位置(location)和各类别的置信度(confidence)进行预测.位置输出有 4 个:(l^x, l^y, l^w, l^h),分别为物体框中心和宽高相对于默认框的偏移;置信度输出有 N 个:(c^0, c^1, \dots, c^{N-1}),分别为物体属于各类别的分数,其中 c^0 为背景(负样本)分数.因此在每个特征层的输出维度为 $L \times L \times B \times (N+4)$,其中 L 为特征图尺寸, B 为每个像素默认框数量,在 conv4_3 层为 3,其余层为 6.

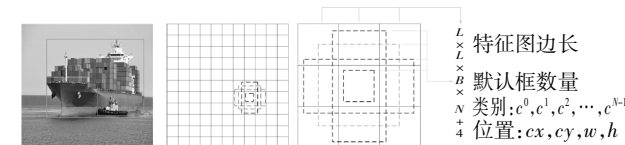


图 3 多尺度特征图检测

Fig.3 Multi-scale feature maps for detection

金字塔式的特征层结构使网络能够适应不同大小物体的检测,如在训练过程中,图中小艇和大船所在的物体框分别与 11×11 和 3×3 特征图中较小和较大的默认框相匹配,则这两个框标记为正样本,其余默认框为负样本.因此,小的物体在较大的特征图中检测,而大的物体在较小的特征图中检测.多尺度特征图检测是 SSD 模型的特点与先进之处.如此,可以提高检测结果的精度,同时还得到了不同长宽比例的检测结果.

SSD 模型训练过程中的损失定义如下^[8]:

$$L(x, c, l, g) = \frac{1}{N} (L_{\text{conf}}(x, c) + \alpha L_{\text{loc}}(x, l, g)) \quad (1)$$

式中: N 为匹配目标的默认框数量, 特殊地当无目标即 $N=0$ 时, 取损失 $L=0$, 以利于数据统计; α 为权重系数, 默认为 1; L_{conf} 为置信度损失, 定义如下:

$$L_{\text{conf}}(x, c) = - \sum_{i \in \text{Pos}} x_{ij}^p \log(\hat{c}_i^p) - \sum_{i \in \text{Neg}} \log(\hat{c}_i^0) \quad (2)$$

$$\hat{c}_i^p = \frac{\exp(c_i^p)}{\sum_p \exp(c_i^p)} \quad (3)$$

式中: x_{ij}^p 为当第 i 个默认框与类别 p 的第 j 个真实框匹配时取 1, 否则取 0; c_i^p 为第 i 个默认框属于类别 p 的预测置信度分数; \hat{c}_i^p 为第 i 个默认框属于类别 p 的预测置信度概率, 满足 $\sum_p \hat{c}_i^p = 1$.

L_{loc} 为位置损失, 定义如下^[8]:

$$L_{\text{loc}}(x, l, g) = \sum_{i \in \text{Pos}} \sum_{m \in \{x, cy, w, h\}} x_{ij}^m \text{smooth}_{L1}(l_i^m - \hat{g}_m^j) \quad (4)$$

$$\hat{g}_j^{cx} = (g_j^{cx} - d_i^{cx}) / d_i^{cx} \quad (5)$$

$$\hat{g}_j^{cy} = (g_j^{cy} - d_i^{cy}) / d_i^{cy} \quad (6)$$

$$\hat{g}_j^w = \log\left(\frac{g_j^w}{d_i^w}\right) \quad (7)$$

$$\hat{g}_j^h = \log\left(\frac{g_j^h}{d_i^h}\right) \quad (8)$$

式中: l_i^m 为第 i 个预测框 m 参数的相对偏移量; \hat{g}_m^j 为第 j 个真实框 m 参数相对于默认框的偏移量; g_j^m, d_i^m 分别对应真实框和默认框的 m 参数; smooth_{L1} 为绝对损失函数, 常用于回归。

SSD 模型的训练过程, 就是将真实的标签框与事先定义好的一系列固定大小的默认框配对, 计算并通过迭代最小化损失的过程。

2.2 船撞专用数据集

建立合适的数据集对 SSD 模型进行训练, 才能在桥梁主动防船撞系统中以达到预期的目标检测效果. 充分考虑到检测模型在不同船型、视角、环境下的适应性以及特定工作场景下的针对性, 训练所用到的原始图像数据既来自网络爬取, 也有现场采集. 目标检测任务训练所用的目标检测模型通常可分为两类: 一类是广谱数据集, 即涵盖物体类别较多、数据量较大、适应多种任务的数据集, 常用的包括 PASCAL VOC、MS-COCO 等; 另一类是针对特定目标检测任务, 包含一个或多个物体类别的专用数据集. 上述广谱数据集尽管数据量大、涵盖类别多, 但在特定目标检测且对精度要求较高的任务上, 由于物体本

身外观的多样性, 往往难以达到理想的效果. 因此, 本文建立了用于船舶目标检测的专用数据, 即船撞专用数据集 (VC 数据集, Vessel Collision Dataset) 进行训练.

目标检测模型应具有普适性, 这意味着模型必须适用于多船型、多角度、复杂环境下的检测任务. 因此, VC 数据集的建立充分考虑了样本涵盖面的丰富性. 为此, 从网络上搜集了近 4 万张船舶图片作为样本, 样本涵盖了包括集装箱船、散货船、滚装船、帆船、小艇等在内的多种常见船型, 包含海洋、港口、内河在内的多种地理环境, 多种拍摄角度和不同的气候光照条件. 样本的气候光照条件包含晴朗的白天、雾天、阴天、傍晚和黑夜, 并通过算法筛选提高了其中暗环境条件样本的比例. 这是由于在自然条件下暗环境样本的分布比例相对较少, 人为提高比例可以提高该环境下的检测性能. 经过筛选, 最终共标注了 3 万余张图片, 作为数据集的基础.

为了验证 VC 数据集训练的模型效果, 图 4 比较了其与广谱数据集 VOC 0712 训练 SSD 模型的性能对比图, 共包含四幅图. 第 1 幅图为白天能见度较好情况下的检测结果, 可见上图准确地检测出船体的边界, 而下图的检测存在误差; 第 2 幅图是在傍晚逆光条件下的检测结果, 上图准确地地区分了船体与阴影, 而下图的结果包含了阴影; 第 3 幅图是在夜间有外部光照下的检测结果, 可见上图准确检测出两个目标, 而下图仅检测出其中一个; 第 4 张图是在黑夜无外部光源下的结果, 可见下图未能检测出目标, 而上图在此条件下依然保持较好的检测效果. 可以看出, 基于 VC 数据集的 SSD 模型对船舶目标的检测无论在准确性还是各环境条件下的鲁棒性上均远远优于基于广谱数据集的模型检测结果, 能够充分保证模型的普适性.

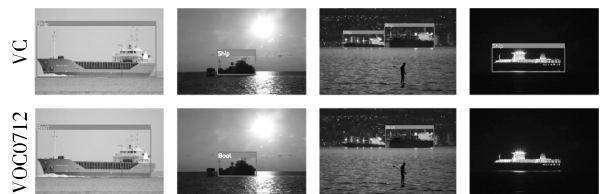


图 4 VC 数据集与 VOC 0712 数据集检测效果对比

Fig.4 Performance comparison between VC and VOC 0712

目标检测模型应具有针对性. 利用海量样本数据建立的数据集具备适应不同条件检测任务的普适性, 但其在实际应用场景下并不一定能达到最佳. 通过有针对性地加入数据量远小于已有样本的现场图像数据到数据集, 就能以较小代价使模型在实际场

景下的工作性能得到显著的提升,使模型兼具普适性和针对性.为此,采集试验桥梁通航区的监测视频,拍摄时间从白天持续到夜晚,从视频中按一定时间间隔截取帧图像,并对打乱顺序后的图像样本进行标注,对已有数据集进行扩充.为进一步提升检测性能,将实拍样本划分为简单样本和困难样本,如图 5 所示,简单样本通常为单艘船舶,边界鲜明,困难样本包括船舶间遮挡、桥墩遮挡以及相似背景干扰等情况.在正常分布下,简单样本占比较高,并且大多相似,而困难样本较少,根据损失函数的定义,此类样本在损失计算中所占权重相对较低,这将导致其特征不能得到充分学习,如此易导致模型在上述困难情况下表现不佳.因此,通过对困难样本减小采样间隔、裁剪等数据增加手段,使实拍样本集中的简单样本和困难样本的比例接近 0.5:0.5,从而提升模型对困难样本的识别能力.试验结果显示,在基础数据集中增加仅 1/20~1/10 的针对性场景扩充数据,便能提升模型的目标检测能力.试验结果显示,在基础数据集中增加仅 1/20~1/10 的针对性场景扩充数据,便能提升模型的目标检测能力,避免了在船舶间遮挡、桥墩遮挡以及相似背景干扰等情况下的误检和漏检.

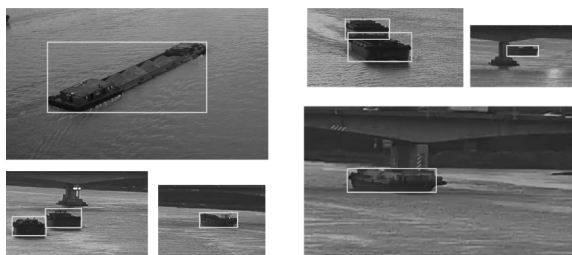


图 5 样本划分
Fig.5 Sample division

数据集中样本的涵盖面和比例之外,标注的质量同样影响模型检测的准确性.目标检测任务的数据标注包括了目标的边框和类别.对于船舶目标,影响标注质量的因素主要有 2 个,如图 6 所示.



图 6 标注质量影响因素
Fig.6 Impacts of annotation quality

其一是部分船舶上部有细长高耸的桅杆或天

线,甚至高度远超船体部分,若将其标注在内,框上部绝大部分区域为背景,而实际目标的特征极少,如此将导致模型训练过程中引入了过多背景特征形成干扰,在实际工作中表现为误警,因此在标注过程中应当舍弃过细过长的部分.其二是目标受到部分遮挡时,若遮挡面积较大,同样会引入过多干扰因素,使得实际工作中将障碍物误识别为目标,因此当遮挡面积超过 40%时不标注.

在 SSD 模型的训练过程中,对训练数据的图片按照 0.1 至 1 的尺寸比例随机裁剪,并以 50% 的概率随机翻转,使模型适应不同尺寸和外形的目标.

数据集建立后,可采用数据增强策略来提升模型对目标尺寸的适应性.将 VC 数据集按照 8:1:1 的比例分为训练集、验证集和测试集,训练中每进行 10 000 次迭代使用测试集评估损失.图 7 为模型经过 80 000 次迭代的损失变化曲线,可见模型在 50 000 次迭代后训练集损失逐渐收敛,测试集损失在 10 000 次迭代后便已平稳,经过 80 000 次训练的模型在测试集上的损失收敛到 0.757.

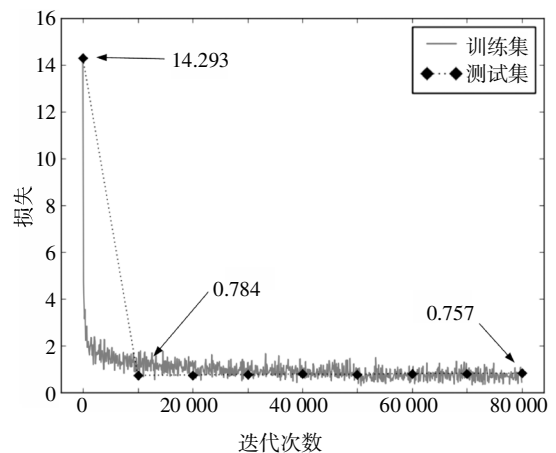


图 7 损失-迭代次数曲线

Fig.7 Loss-iteration curve

2.3 航区广域划分方法

航区桥梁主动防撞目标检测的一大特点和难点就是监测范围大,远近目标在监控画面中大小悬殊,导致小目标难以识别与定位.为解决这一问题,提出航区广域划分方法.定义目标大小的衡量指标如下:

$$\hat{g}_j^h = \min\left(\frac{h}{H}, \frac{w}{W}\right) \quad (9)$$

实测结果表明 SSD 模型能够检测 λ 在 0.1~1 间的目标,且随着 λ 的降低准确率有所下降. SSD 的架构决定了该模型对小目标的检测有局限性,原因有

三:其一,小物体通常在大的特征图中检测(如图3所示),但由图2可见,较之于小的特征层,大的特征层在网络中所处层次较浅,而小物体的检测通常需要更深层的特征,这一矛盾制约了小物体的检测能力;其二是检测前对图片的缩放使得小物体原本少的特征信息被进一步压缩,尽管SSD512相较SSD300增加了输入图像的尺寸,但小物体的检测能力提升不大;其三是训练过程中正负样本的划分,如图3所示,小物体检测所在的特征图中负样本数远大于正样本,使得正样本的特征未能被充分地学习。上述原因制约了SSD模型在小目标检测上的表现,这是检测速度与精度妥协的结果,尽管RetinaNet^[11]等提出通过改进损失函数减小正负样本比例失衡的影响,但 λ 远小于0.1时目标特征信息不足依旧是最大的制约,充分利用监控视频的高清图像信息方是解决此问题的关键。

因此,针对桥梁主动防撞的大范围监测,本文提出分区域检测方法,利用算法自动将监测区域划分为 $(m \times n + 1)$ 个区域,其中靠近摄像头且无河岸部分设为1个矩形区域,画面中两岸围成的梯形区域分为 $m \times n$ 个区域,其中 m 为横向分割数, n 为纵向分割数,横向宽度按下底两岸宽度均分,纵向高度满足:

$$h_i = 1.618h_{i+1} \quad (10)$$

$$\sum_{i=1}^n h_i = H_T \quad (11)$$

式中: h_i 为由下向上第 i 组矩形区域高度; H_T 为梯形高度。

划分后的区域在画面中为 $(m \times n + 1)$ 个矩形,将矩形放大1.1~1.2倍使之相互交错,如图8将画面划分为 $(3 \times 2 + 1)$ 个区域,然后将各区域图像分别输入模型进行检测,将检测结果重组,对交错区域的重合结果进行非极大值抑制,得到监测区域的总检测结果。各区域的相互交错可避免区域交界处目标割裂导致的漏检。通过上述方法,能够使各区域船舶目标的 λ 保持在检测准确率较高的区间,充分利用高分辨率图像信息的同时将天空、河岸的无关区域舍弃,从而得到覆盖广大监测范围的、稳定的检测结果。如图9所示,近处较大的目标和远处极小的目标能够在同一相机视野中检测。

航区广域划分方法避免了单帧输入导致的像素丢失,使高清相机的图像信息得到充分利用,解决了大范围监测下小目标的识别定位问题。

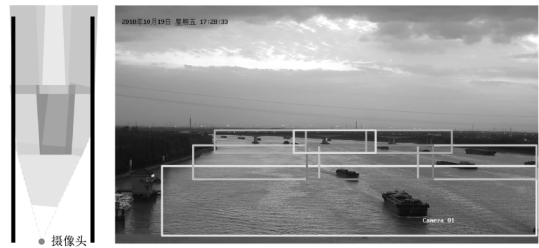


图8 区域划分方法示意

Fig.8 Region division



图9 检测效果展示

Fig.9 Detection effect

3 目标检测算法综合评价

本文提出的船舶目标检测算法在准确性、稳定性、效率和智能化方面均远优于传统防撞系统所采用的船舶目标检测算法。现有桥梁主动防撞视频监测系统大多采用基于运动的目标检测方法,包括帧间差分法和背景减除法,其中后者是发展最为成熟、在防撞领域使用最广泛的方法。背景减除法的基本原理是通过统计学理论建立检测场景的背景图像,再将每一帧待检测目标图像与背景图像进行差分,实现运动前景的检测,并通过背景图像的动态更新来适应动态场景的情况。以下将本文所采用的船舶目标检测方法与该方法做了比较。

根据背景模型的不同,基于背景减除原理的目标检测方法很多,如单高斯背景建模法、混合高斯背景建模法等^[12],本文选取了ViBe算法(Visual Background Extractor,视觉背景提取)为比较对象,该方法由Barnich等人于2009年提出^[13],基于邻域随机取样的背景建模方法,是目前最为先进的运动目标检测方法之一。如图10所示,使用SSD和ViBe对同一段航道视频进行处理,将船舶目标检测效果进行比较。

由第1s的检测结果可以看出,ViBe能够在第2帧就开始检测,这是其较之于同类算法的优势(背景建模通常需要前若干帧),但由于水面波浪、光照等动态因素的存在,ViBe所得到的前景中存在大面积的噪声,这直接导致了大面积的误检。此外,由于初始

背景模型将全图视为背景,还会导致目标移动后在其初始位置残留鬼影^[4].本文所提方法不存在初始化和帧间关联的问题,从第 1 帧开始便可稳定检测.

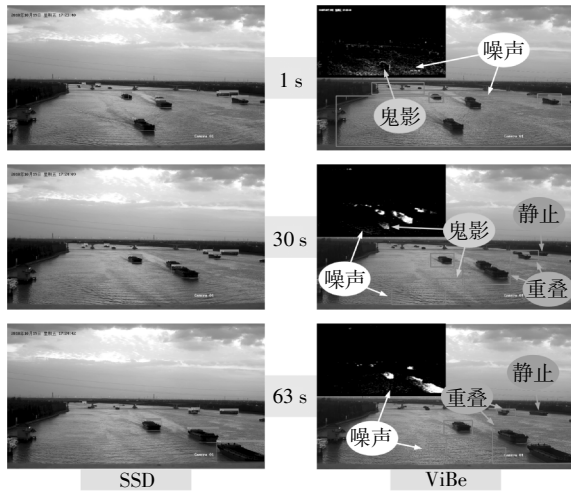


图 10 SSD 与传统背景减除法性能比较

Fig.10 Comparison between SSD and conventional method

由第 30 s 和第 63 s 的检测结果可见,尽管通过背景的更新,噪声面积减小,但其影响依旧存在.另外,从 30 s 的结果可以发现,鬼影的影响仍然存在.噪声和鬼影将导致背景减除法的检测结果中存在大量误警.值得一提的是,画面的抖动也将导致大量噪声的出现.此外,可以发现,ViBe 并不能将静止的目标检测出来,而对于相邻或重叠的目标,该方法无法将其有效区分.从检测框的精度层面,ViBe 的检测框中不仅包含目标本身,还有目标的阴影、倒影以及噪声,这直接影响到目标定位的精度.另外,此类方法需要通过高斯平滑来减小噪声的影响,这也导致了有效图像信息的丢失,对小物体的检测能力有限.

综上所述,传统的背景减除算法的检测需要经过初始化,才能进入稳定检测阶段,并且在抗噪声、应对物体重叠、检测精度上均有不少局限性.此外,传统方法仅仅是机械式地区分运动物体和静止背景,对静止目标的检测和运动非目标的区分无能为力.而本文所采用的基于神经网络的目标检测方法,从第 1 帧开始便能够稳定检测,且应对上述问题均有良好的表现,更为精确、稳定、智能,尤其在应对光照和水面波浪变化剧烈、桥梁振动等复杂航道监测环境上,其优势显而易见.

本文还将两种方法的运行速度进行了比较,如图 11 所示,该对比试验采用 AMD Ryzen 7 2700X+NVIDIA GTX 1080Ti 的硬件环境.神经网络运算支持 GPU 并行,因此 SSD 使用 GPU+CPU 运算,而 ViBe

仅使用 CPU.由于 SSD300 模型的输入大小为 300×300,因此将测试视频的图像尺寸统一缩放到 300×300 进行处理.在该分辨率下,ViBe 平均每秒所处理的帧数仅为 1.76 帧,而 SSD300 的速度达到了 30.27 fps.由于本文采用了分区域检测的方法来提高图像信息的利用率,事实上将单帧图像划分为多个图像进行处理,在按图 8 划分为 7 个区域进行检测的情况下,其每秒处理的帧数也达到了 8.72 帧.将 ViBe 的输入尺寸提高到 512×512 来提高像素利用率,其运行速度降低到仅 0.66 fps,而检测效果并无明显提升.因此,在相同输入尺度下,本方法的运行速度比传统方法快 395%以上,且准确度更高.

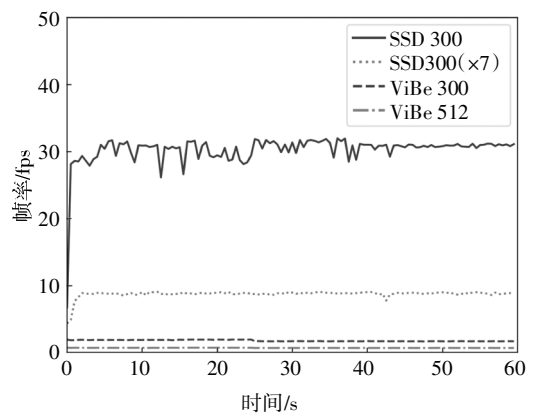


图 11 SSD 与 ViBe 算法检测速度比较

Fig.11 Speed comparison between SSD and ViBe

表 1 对本文提出的目标检测方法传统的船舶目标检测方法进行了综合评价.显然,基于深度学习的船舶目标检测方法在各方面均优于传统方法,更加稳定、准确、高效和智能.

表 1 本文方法与传统方法综合评价

Tab.1 Comprehensive evaluation

综合性能	评价科目	本文方法	传统方法
高效性	并行运算	支持	不支持
	检测速度	快	较慢
	信息利用率	高	低
准确性	阴影区分	区分	不区分
	重叠区分	区分	不区分
	小物体	检测	不检测
稳定性	抗噪声	强	弱
	动态背景	不影响	影响
	起始帧数	第 1 帧	第 n 帧
智能性	类型识别	支持	不支持
	静止物体	检测	不检测

4 松浦大桥航区试验

以上海松浦大桥为背景进行了实桥航区试验, 试验目的在于: 其一, 获取实际航道的船舶图像数据, 用于模型训练; 其二, 优化视频获取模块的集成与布设方案; 其三, 量化本文提出的船舶目标检测方法的定位精度和有效范围; 其四, 为进一步的航迹预测模型的建立提供数据基础. 上海市松浦大桥为预应力连续钢桁架结构, 正桥长 419.6 m, 四孔, 中间二孔跨径各 112 m, 两边二孔跨径各 96 m. 上层公路桥面宽 12 m, 其中车行道宽 9 m, 两端引桥共长 1 440 m, 南北两岸各 22 孔, 跨径 32.7 m. 下层铁路桥两端引桥共长 2 628 m, 桥面宽 12 m, 南岸 38 孔, 北岸 42 孔, 跨径为 32.7 m, 桥位处河道宽度约 400 m.

试验步骤包括标定点的布设与测定、相机架设与调节、相机内参标定、视频采集灯, 其中标定点的布设测定用于后期摄像机像素坐标-世界坐标的换算. 如图 12 所示, 在 3 个桥墩顶部桥面处各架设 1 台摄像机, 拍摄航道视频. 其中 #1 和 #3 摄像机为普通网络摄像机, 分别用于拍摄近距离和中距离监控区域的视频图像; #2 为云台摄像机, 使用长焦镜头, 用于拍摄远距离监控区域的视频图像. 试验所拍摄的视频图像用于船舶图像数据集的扩充和模型训练, 以提升神经网络模型在实桥场景下的工作性能.

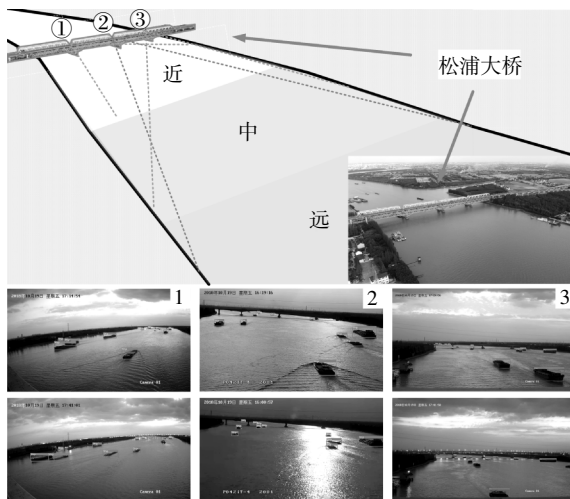


图 12 摄像采集点布设与模型检测效果

Fig.12 Camera arrangement and detection effect

利用采集到的实桥监控视频验证目标检测模型在不同视角、不同光线环境下的工作性能, 如图 12 所示. 首先, 由结果可见该模型在不同视角下均表现出了良好的检测性能, 能够很好地应对船舶目标姿

态多样性的挑战. 其次, 该视频拍摄于秋季下午 16:00~18:00 间, 监测区域为桥梁西偏南方向, 该时段的场景变化十分剧烈, 船舶目标检测面临逆光、水波和日落等因素的挑战. 从不同时段光线条件下的检测结果可以看出, 该模型能够实现正常光照环境下的船舶目标稳定检测, 在应对强烈逆光、夜间环境光均有良好的表现.

如图 12 所示, 摄像头 #1 上下两图距桥 500 m 以内各有 5 艘和 6 艘船舶, 均被成功检测, 且各有 2 艘发生不同程度交叉的船舶, 亦得到精准的区分; 摄像头 #2 上下两图中各有 10 艘和 11 艘船舶, 均被成功检测, 尤其下图中存在强烈的反射光和镜头抖动干扰; 摄像头 #3 上下两图中分别有 13 艘和 10 艘船舶, 同样得到精确检测, 且上图中最近和最远的船舶目标距离超过 2 000 m, 大小相差超过 10 倍, 均能被该机位覆盖.

试验还对本系统的有效监测范围和定位精度进行了验证. 图 13 是航迹跟踪模块基于本模型目标检测结果所得到的船舶目标航迹 (桥墩处圆角框为处于视野盲区的目标). 有赖于使用区域划分进行大范围监测优化, 使用中等焦距镜头的 #3 相机的有效监测范围达到了 2.5 km, 使得检测方法对小物体的检测能力不再是有效监测范围的制约因素, 取而代之的是河道走向、视野障碍等外在因素. 此外, 检测精度的提升也为目标的精准定位提供了良好的数据基础.

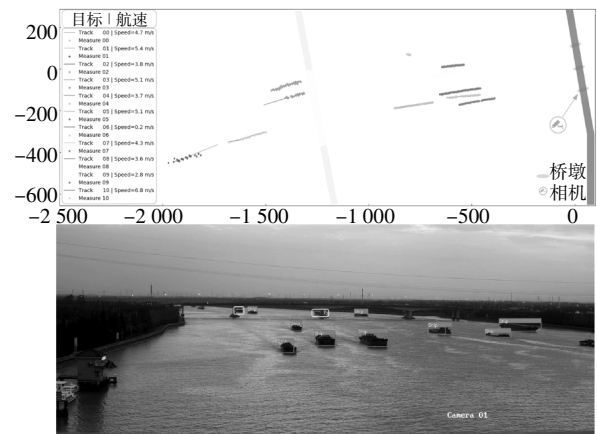


图 13 #3 摄像头检测目标航迹和有效监测范围

Fig.13 Tracking results and monitoring range under camera #3

综上, 实桥试验验证了本文提出的主动防船撞系统框架中的基于深度学习的目标检测方法, 其能够充分满足针对桥梁主动防撞航区目标检测各方面的需求, 应对包括监测时间长、环境复杂性高、监测范围大、目标外观多变、检测精度要求高等问题, 从

而为系统进一步的航迹跟踪预测、防撞预警决策和航道监控管理提供信息支持。

5 结 论

本文针对桥梁主动防撞视频监控研究了基于深度学习的航区目标检测方法,主要进行了以下工作:选择SSD作为航区目标检测的模型基础,建立神经网络模型;建立船撞专用数据集,包含35 155张船舶图像,通过样本选取和数据增强兼顾了模型普适性和实际工作场景下的针对性;通过航区广域划分方法解决小物体检测问题,中等焦距下单目标监测范围达到100~2 500 m。

综合评价结果表明,本文提出的航区目标检测方法在各方面均较传统方法有较大提升,包括:有效应对噪声和动态背景,且无需初始化过程,全程稳定工作;不受阴影、倒影和目标重叠影响,定位精准,且对 λ 小于0.02的物体亦保持较高检测率;速度快395%以上,同时像素利用率更高;区分目标类型,减少误检和漏检,智能化程度更高。

实桥试验验证了本文提出的基于深度学习的航区目标检测方法能满足航区全天候、多视角、大范围、多目标的目标识别定位需求,为进一步的航迹跟踪预测、防撞预警和航道监控提供信息支持。

参 考 文 献

- [1] 陈亮,吴善刚,肖英杰,等. 基于视频目标检测的桥梁防船撞主动预警技术研究[J]. 水运工程,2012,(6):150—154.
CHEN L, WU S G, XIAO Y J, *et al.* Active forewarning technology for anti-collision between ship and bridge based on video object detection [J]. Port & Waterway Engineering, 2012, (6): 150—154. (In Chinese)
- [2] 任慧,王莹,肖刚,等. 桥梁防船撞主动预警系统设计及实验研究[J]. 激光与红外,2013,43(1):66—70.
REN H, WANG Y, XIAO G, *et al.* Active early-warning system for bridge piers against ship collision and its experiment [J]. Laser & Infrared, 2013, 43(1): 66—70. (In Chinese).
- [3] LUCAS B, KANADE T. An iterative image registration technique with an application to stereo vision[C]// Proceedings of the 7th international joint conference on Artificial intelligence. Vancouver: IJ-CAI, 1981: 674—679
- [4] GLOYER B, AGHAJAN H K, SIU K Y, *et al.* Video-based freeway-monitoring system using recursive vehicle tracking [J]. Proc Spie, 1995, 2421: 173—180.
- [5] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection [C]// 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). San Diego: IEEE, 2005: 886—893.
- [6] 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.
- [7] HE K, ZHANG X, REN S, *et al.* Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904—1916.
- [8] GIRSHICK R. Fast R-CNN [C]// 2015 IEEE International Conference on Computer Vision (ICCV'15). Santiago: IEEE, 2015: 1440—1448.
- [9] LIU W, ANGUELOV D, ERHAN D, *et al.* SSD: single shot multibox detector [C]// Proceedings of the 14th European Conference on Computer Vision. Netherlands: ECCV, 2016: 21—37.
- [10] REDMON J, DIVVALA S, GRISHICK R, *et al.* You only look once: unified, real-time object detection [C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'16). Las Vegas: IEEE, 2016: 779—788.
- [11] LIN T, GOYAL P, GRISHICK R, *et al.* Focal loss for dense object detection [C]// 2017 IEEE International Conference on Computer Vision (ICCV'17). Venice: IEEE Computer Society, 2017: 2999—3007.
- [12] DAN D, GE L, YAN X. Identification of moving loads based on the information fusion of weigh-in-motion system and multiple camera machine vision [J]. Measurement, 2019. doi: 10.1016/j.measurement.2019.05.042.
- [13] BARNICH O, DROOGENBROECK M. ViBe: A powerful random technique to estimate the background in video sequences [C]// Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing. Washington D C: IEEE Computer Society, 2009: 945—948.
- [14] 李鹏, 王延江. 一种运动背景下视觉注意辅助的目标检测方法[J]. 湖南大学学报(自然科学版), 2018, 45(8): 138—149.
LI P, WANG Y J. A target detection method aided with visual saliency under moving background [J]. Journal of Hunan University (Natural Sciences), 2018, 45(8): 138—149. (In Chinese).