Lecture Notes in Networks and Systems 1076

Cezary Biele · Janusz Kacprzyk ·
Wiesław Kopeć · Jakub Możaryn ·
Jan W. Owsiński ·
Andrzej Romanowski ·
Marcin Sikorski *Editors*

Digital Interaction and Machine Intelligence

Proceedings of MIDI'2023 – 11th Machine Intelligence and Digital Interaction – Conference, December 12–14, 2023, Warsaw, Poland (online)

Springer

Lecture Notes in Networks and Systems 1076

The series “Lecture Notes in Networks and Systems” publishes the latest developments in Networks and Systems—quickly, informally and with high quality. Original research reported in proceedings and post-proceedings represents the core of LNNS.

Volumes published in LNNS embrace all aspects and subfields of, as well as new challenges in, Networks and Systems.

The series contains proceedings and edited volumes in systems and networks, spanning the areas of Cyber-Physical Systems, Autonomous Systems, Sensor Networks, Control Systems, Energy Systems, Automotive Systems, Biological Systems, Vehicular Networking and Connected Vehicles, Aerospace Systems, Automation, Manufacturing, Smart Grids, Nonlinear Systems, Power Systems, Robotics, Social Systems, Economic Systems and other. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution and exposure which enable both a wide and rapid dissemination of research output.

The series covers the theory, applications, and perspectives on the state of the art and future developments relevant to systems and networks, decision making, control, complex processes and related areas, as embedded in the fields of interdisciplinary and applied sciences, engineering, computer science, physics, economics, social, and life sciences, as well as the paradigms and methodologies behind them.

Indexed by SCOPUS, EI Compendex, INSPEC, WTI Frankfurt eG, zbMATH, SCImago.

All books published in the series are submitted for consideration in Web of Science.

For proposals from Asia please contact Aninda Bose (aninda.bose@springer.com).

Cezary Biele · Janusz Kacprzyk ·
Wiesław Kopeć · Jakub Możaryn ·
Jan W. Owsiński · Andrzej Romanowski ·
Marcin Sikorski
Editors

Digital Interaction and Machine Intelligence

Proceedings of MIDI'2023 – 11th Machine Intelligence and Digital Interaction – Conference, December 12–14, 2023, Warsaw, Poland (online)

Springer

Editors
Cezary Biele
National Information Processing Institute
Warszaw, Poland

Janusz Kacprzyk
Polish Academy of Sciences
Systems Research Institute
Warsaw, Poland

Wiesław Kopeć
Polish-Japanese Academy of Information Technology
Warsaw, Poland

Jakub Możaryn
Institute of Automatic Control and Robotics, Faculty of Mechatronics
Warsaw University of Technology
Warsaw, Poland

Jan W. Owsiński
Polish Academy of Sciences
Systems Research Institute
Warsaw, Poland

Andrzej Romanowski
Institute of Applied Computer Science, Faculty of Electrical, Electronic, Computer and Control Engineering
Łódź University of Technology
LODZ, Poland

Marcin Sikorski
Department of Applied Informatics in Management, Faculty of Management and Economics
Gdańsk University of Technology
Gdańsk, Poland

ISSN 2367-3370 ISSN 2367-3389 (electronic)
Lecture Notes in Networks and Systems
ISBN 978-3-031-66593-6 ISBN 978-3-031-66594-3 (eBook)
https://doi.org/10.1007/978-3-031-66594-3

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Artificial intelligence (AI) is rapidly reshaping our lives due to significant research advancements and widespread adoption of interactive technologies, giving rise to novel social phenomena. Countries worldwide are actively seeking to comprehend and tackle these phenomena, necessitating interdisciplinary approaches and collaboration among researchers and practitioners.

The MIDI conference's central objective is to bridge the once-separate realms of AI and human-technology interaction. Commencing in 2020, the conference will encompass AI challenges alongside topics like interface design and user experience. While society grows increasingly aware of AI-related concerns, the development of AI technology outpaces the quest for solutions. Addressing these challenges effectively requires a blend of social research and AI expertise. The expanded conference format aims to foster rich exchanges among experts in artificial intelligence and human-technology interaction."

We anticipate that both enthusiasts of emerging trends and those engaged in developing end-user IT products and services will unearth a wellspring of inspiration and invaluable theoretical and practical insights within the pages of this book. The ultimate fate of any newly conceived product hinges on its ability to align with the evolving needs of individuals in a future where AI-based technological innovations are forged by human hands for the benefit of humanity. No matter how pioneering a technological solution may appear, unless it seamlessly integrates with the lifestyle and assorted factors influencing the social behavior of its intended users, it is bound to face rejection. Underestimating the significance of technology would be unwise, as evidenced by the annals of technological progress and the instances of groundbreaking solutions devised even by the world's most influential tycoons.

In this year's conference, two papers were honored with the Best Paper Award, dedicated to the memory of Professor Krzysztof Marasek. Professor Marasek played a pivotal role in initiating the inaugural MIDI conference, which delved into computer science from a user-centric perspective, and remained a driving force in subsequent editions. His profound expertise encompassed linguistics, voice user interfaces, and voice-based interactions, underscoring the critical importance of research on technological solutions approached through the user's lens. We aspire to see this year's and future iterations of the MIDI conference continue the legacy of Professor Krzysztof Marasek."

Cezary Biele
Janusz Kacprzyk
Wiesław Kopeć
Jakub Możaryn
Jan W. Owsiński
Andrzej Romanowski
Marcin Sikorski

Contents

Machine Intelligence

Maritime Vessel Detection and Classification in Harbor Environment Using Deep Learning

Khalid Waleed, Muhammad Umar Farooq(✉), and Rana Hammad Raza

Pakistan Navy Engineering College (PNEC), National University of Sciences and Technology (NUST), Karachi, Pakistan

{khalid.waleed2017,umar.farooq,hammad}@pnec.nust.edu.pk

Abstract. Automatic surveillance of docks, harbors, and seaports is considered crucial for ensuring smooth port operations. The identification and tracking of maritime vessels in these environments are generally facilitated by Radar and Automatic Identification System (AIS). However, radar faces challenges in detection of small non-metallic vessels, as well as there are other issues such as radar signal clutter, high electromagnetic radiation, and higher costs. Similarly, AIS also face challenges such as device malfunctioning or illegal manipulation etc. Consequently, cameras have been increasingly used in maritime traffic management systems. In recent years, Convolutional Neural Networks (CNNs) have shown significant progress in object detection and classification tasks. Though, obtaining large scale maritime datasets with focus on harbor environment still remains a challenge. In this study, experimentation with various pre-trained state-of-the-art CNN models were conducted on images captured from harbor environments in Karachi and Grand Canal in Venice. The dataset presented several challenges such as high-density traffic, occlusions, shadows, background infrastructure, wave motion, boat wakes and reflection on water surface. The pre-trained CNNs were finetuned using a combination of small datasets to obtain both detector and classifier which outperformed previously published results on the MarDCT dataset. Specifically, the proposed model achieved a DR (Detection Rate) of 0.81, FAR (False Alarm Rate) of 0.07 and an average accuracy of 98.78% on the MarDCT dataset.

Keywords: Maritime Traffic · Vessel Detection · Vessel Classification · Machine Learning · Deep Learning · Convolutional Neural Networks

1 Introduction

The increasing use of harbors for recreational activities at sea, passenger ferry services, and the significant increase in global maritime trade and economic activities highlight the need for automatic surveillance systems for docks, harbors, and ships [1]. These surveillance systems play an important role in addressing various challenges, including but not limited to identification of unknown vessels/ships and their potential collision paths, and controlling the maritime traffic in highly congested areas of ports/harbors. The International Chamber of Shipping also recognizes the importance of safe operations and recommends to implement advanced surveillance systems [2].

C. Biele et al. (Eds.): MIDI 2023, LNNS 1076, pp. 3–17, 2024.
https://doi.org/10.1007/978-3-031-66594-3_1

Traditionally, Vessel Traffic System (VTS) are employed to manage ship traffic, monitor vessel movement and collision avoidance. These systems rely on input from two main sources i.e., Radars and Automatic Identification Systems (AIS). Although, radars provide accurate position information, but they are costly and face limitations in detection of small vessels with non-metallic bodies. Radar signals are also prone to clutter, caused by reflections from various objects such as ground, water, buildings and other ships [3]. Similarly, the radar-based systems are not suitable for populated areas due to high electromagnetic radiation emissions, and therefore necessitate building of complex shadow zones to protect humans from exposure. On the other hand, AIS signals may also encounter issues due to device malfunctioning or illegal manipulation of the device.

Due to these challenges, there has been an increasing use of cameras for the maritime traffic systems. These cameras offer advantages such as affordability, flexibility, and compatibility with various platforms. Some applications rely solely based on camera inputs [4], while others integrate detection and tracking results from cameras with VTS information to provide a robust maritime traffic system.

However, visual detection and tracking using cameras in maritime environment has its own set of challenges. The conventional algorithms for vessel detection and tracking in standard video settings often fail to produce satisfactory results in maritime environments, due to the dynamic nature of background [5]. The constantly moving waves, the presence of shadows, reflections and boat wakes, as well as adverse weather conditions such as direct sun, rain and fog etc. can cause difficulties in capturing clear images with cameras. Similarly, tracking ships over large distances causes changes in viewpoint, which further complicate the scenario. Furthermore, cameras have very limited range of visibility.

Therefore, in order to address these complexities of the maritime environment, advanced algorithms are required. In recent years, Convolution Neural Networks (CNNs) have shown remarkable advancements for object detection in images/videos. CNNs excel at automatic feature extraction of the input data, which further improve the detection performances. However, there are also some associated challenges such as the need for extensive datasets, longer training times, and the requirement for specialized hardware such as Graphical Processing Units (GPUs), for high-speed trainings and inferences.

The hardware requirements can be overcome by utilizing online cloud-based resources such as Google Colaboratory [6] and Kaggle [7], providing affordable access to GPUs. Although, their free versions have some limitations, but users can always opt for inexpensive paid versions to access better resources. As for the datasets, large open datasets specific to maritime traffic in harbors are often unavailable due to security concerns. Therefore, researchers must rely on open datasets such as ImageNet [8], PASCAL [9] or MSCOCO [10] etc., which however, provide limited images/objects related to vessels. Transfer learning has emerged as a productive approach when dealing with scenarios of limited datasets and constrained training times. It simply freezes the feature extraction layers of pre-trained high performing models and then retrains the output layer to adjust the predictions as per the specific use case.

Existing research on detection and tracking of maritime vessels typically involves manual feature extraction (such as color and vessel length features etc.) or extraction using approaches such as Histogram of Oriented Gradients (HOG) and Scale Invariant and Feature Transform (SIFT) etc., combined with separate classifiers. Moreover, most of the research is focused on open sea scenarios, utilizing sensors onboard ships [11], Unmanned Aerial Vehicles (UAVs) [12], or satellite imagery [13] to detect, classify and track vessels. While, these conventional techniques demonstrate certain performance levels, exploring the CNNs which have shown promising performance in object detection scenarios, could significantly improve the performance. The existing research done in harbor contexts using CNNs has primarily relied on satellite images [14] or Synthetic Aperture Radar (SAR) images [15]. However, limited research has been conducted in this context, using image samples from fixed cameras installed in harbor environments (which generally exhibit varying maritime traffic densities and include backgrounds containing harbor infrastructures). The main contributions of this research are highlighted as follows:

- Transferred the convolution layers of CNN architectures for fast training and testing of vessel images, which achieved significant performance improvements compared to existing methods.
- Demonstrated better performance of the proposed method for maritime vessel detection and classification, specifically in local harbor environment, even with a limited data.

The rest of the paper is organized as follows. Section 2 provides an overview of the related research. Section 3 presents the working methodology of the proposed approach. Section 4 provides details for implementation of this research work. Section 5 discusses the results obtained from the experimentation. Finally, Sect. 6 concludes the paper and provides future research direction.

2 Related Works

Conventional Based Approaches: Several algorithms/techniques are discussed in the literature, which address the challenges of maritime traffic detection using cameras as the input source. Some of these algorithms have been tested in real systems at different seaports/canals, to monitor and manage maritime traffic. While, others have been implemented on diverse datasets to explore the efficacy of the proposed algorithms.

A two-step algorithm combining a detector based on HOG with Gaussian Mixture Model (GMM) and statistical based dynamic texture model was employed for detection maritime vessels [16]. It also incorporated latent Support Vector Machines (SVM) classifier. The effectiveness of the algorithm was evaluated on VOC2010 dataset, and a custom-made dataset captured using a fixed camera (consisting 600 vessel images). However, it was observed that without spatial and temporal correlations, the model suffered significant false detections. A WISARD (Wilkie, Stoneham and Aleksander's Recognition Device) weightless neural network-based detector along-with Kalman filter was used to track maritime vessels [4]. Tracking by detection method was employed, and the results were evaluated on custom-made dataset, consisting of 20 videos captured using fixed cameras. The videos covered both open sea and coastal areas. Moreover, the implementation was done with GPU acceleration. A detection approach using a 24-level cascade of boosted classifiers was proposed for detection of vessels [1]. The classifiers were based on edge, line and center-surround Haar features, incorporating a Canny edge detector having specific threshold values. An additional classifier level was also introduced to filter noise due to waves and boat wakes. For tracking purpose, PTracking algorithm based on distributed multi-clustered particle filtering was utilized. GMM estimated positions for all objects in a scene. Further, the noise due to reflections was filtered using Speed Up Robust Features (SURF) key-points. Multiple sensors including VTS tracks, Infra-Red (IR) cameras, and Pan-Tilt-Zoom (PTZ) cameras were utilized during the experimentation. The proposed approach was evaluated on MarDCT dataset having an open sea environment, and demonstrated high-quality results. A Boolean Map Saliency (BMS) with connected components as detector, and a Kalman filter as tracker was utilized for maritime object tracking [11]. The focus was to detect and track boats around a vessel in an open sea environment, utilizing the onboard fixed cameras. The performance of this approach was evaluated on PETS (Performance Evaluation of Tracking and Surveillance) 2016 dataset. SIFT key-points in combination with Bag of Visual Word (BoW) as a detector, and a modified Sparse Representation Classification (SRC) as a classifier was used for vessel detection and classification in satellite imagery [13]. The study utilized gray-scale satellite images to classify maritime traffic into barge, cargo, container, tanker, and bulk categories. However, the dataset used to conduct the research was not made publicly available. In another approach, background subtraction using unimodal gaussian distribution as a detector and active contours for tracking were employed for maritime vessels [2]. The model's performance was evaluated on a custom dataset captured using both fixed and PTZ cameras in an open sea environment. A robust real-time ship detection and tracking approach for visual surveillance of cage aquaculture was introduced [3]. The approach utilized a detector based on background subtraction (using median of RGB components of pixels) and adaptive template matching tracker. The model performance was evaluated on a custom dataset captured using a fixed camera in an ocean environment. HOG feature extraction was used to detect maritime vessels, and a hierarchical Kanade-Lucas-Tomasi (KLT) feature point tracker was used to track vessels at a port in Netherlands [17]. In order to reduce the computational complexity of HOG, it was only run on a small 3x3 pixel window predicted by the tracker in subsequent frames, after a successful detection. The model utilized a PTZ camera for vessel tracking. However, it faced challenges when tracked objects encountered occlusions, since there was no re-detection mechanism. Therefore, detections were further fused with VTS tracks. Background subtraction based on pixel color was used for detection, and statistics-based method based on color was used as a tracker for maritime traffic in a port at Greece [18]. The study utilized both fixed and PTZ camera as sensors. Model evaluation was carried out on a custom dataset from the port, achieving a real-time processing of 240x320 pixel videos. However, tracking was limited to a single object only. A pioneer video surveillance system for boat traffic monitoring i.e., ARGOS (Automatic Remote Grand Canal Observation System) was presented [19]. Blob formation was applied to the foreground image, and optical flow calculations were performed for each blob. To cope with under-segmentation and over-segmentation, clustering of sparse optical flow points was performed using Rek-means algorithm. Ellipse approximation was then employed to estimate the boats for detection. Kalman filter was used for tracking purposes. The algorithm was implemented in a grand canal of Venice, Italy, as part of the ARGOS, and aim to detect, track and manage maritime traffic. The generated dataset is known as MarDCT and resembles to a harbor environment, which includes high boat traffic density, surrounding infrastructure, docked/moored boats and shadows etc. An approach to classify various categories of boats within ARGOS system was presented [20]. The features based on color, number of edge pixels, boat length, and standard deviation in the horizontal and vertical directions of the edges were extracted. The classification was performed with weighted majority classification using K-Nearest Neighbor (KNN), Decision Tree (J48) and Random Forest (RF) approaches. An Independent Multimodal Background Subtraction (IMBS) was employed for vessel detection [21]. The results were evaluated on MarDCT and changedetection.net (CDNET) datasets. A novel method i.e., Multi-scale Consistence of Weighted Edge Radon Transform (MuSCoWERT) was presented [22]. The approach involved detecting the long linear features consistent over multiple scales using multiscale median filtering of the image, followed by Radon transform on a weighted edge map and computing the histogram of the detected linear features. Model evaluation showed superior performance against various contemporary methods on 84 challenging maritime videos.

Deep Learning Based Approaches: In recent years, deep learning techniques have also been utilized to address the maritime vessel detection challenge. The approaches discussed below show the ongoing efforts in this field, with traditional methods being complemented by deep learning approaches to enhance the overall performance:

A large-scale dataset comprising over 30,000 ship images from six different categories, collected using fixed cameras in a coastal environment was introduced [23]. Variants of Fast R-CNN, Faster R-CNN, SSD MobileNet, SSD VGG-16 and YOLOv2 models were used. Faster R-CNN with ResNet-101 backbone produced high Mean Average Precision (mAP) of 92% at 7 frames per second (FPS), whereas YOLOv2 produced mAP of 79% at 91 FPS. The satellite images from Google Maps were used to detect ships docked in harbors, achieving average precision of 85%, using a Mask R-CNN based method [24].

A VGG-19 model was used to classify various boats in the residential side of Tokyo Bay [25]. The dataset comprised of six classes and was collected using a camera. An F1-score of 0.7 was achieved. A transfer learning approach was utilized to retrain variants of ResNet and Inception model on ImageNet dataset [26]. These models were then applied to Maritime Vessels (MARVEL) dataset [27] and achieved 78% accuracy. An improved version of YOLOv3 model was used to detect ships in a sea environment [28]. The model was tested on a combined dataset comprising of MS COCO 2017, PASCAL VOC, and a custom dataset, achieving a mAP of 75% at 30 FPS.

3 Working Methodology

3.1 Selected Models

The specific details of the selected models are as follows. Different deep learning-based models were used in this study for comparative analysis:

Single-Stage Detector (SSD): SSD is a single-stage object detection model which can perform both object localization and classification in a single forward pass of the network. The research utilizes Inceptionv2 and MobileNetv2 based CNN architectures. Inceptionv2 is the second generation of Inception CNN architectures which notably uses batch normalization. While, MobileNetv2 is a CNN architecture which is highly optimized to run on low computation devices.

You Only Look Once (YOLO): YOLO [29] is a real-time object detection model. It uses single neural network to predict object locations in input image and classify them. Due to single network, it can predict at higher speeds compared to other models like variants of R-CNN [30][30–32]. The research utilizes YOLOv3 [33], as well as different architectures series including tiny (t), small (s), large (l) and extra-large (x) which vary on the basis of parameter sizes and provide various optimizations related to particular image sizes.

The key difference between YOLO and these models is their approach to object detection tasks. While YOLO can identify objects by using a single forward pass, SSD makes use of fixed-size anchor boxes and takes into consideration the IoU metrics.

3.2 Dataset

The datasets used to conduct this research work were taken from following three different sources. Overall distribution of the data is shown in Table 1.

Dataset 1 – MarDCT. The first set, MarDCT [20] is based on maritime traffic in Grand Canal, Italy. It constitutes 79.65% in terms of images (4,020) and 57.26% in terms of the total objects (4,029) with objects to image ratio of 1.00. Figure 1 shows sample of the dataset.

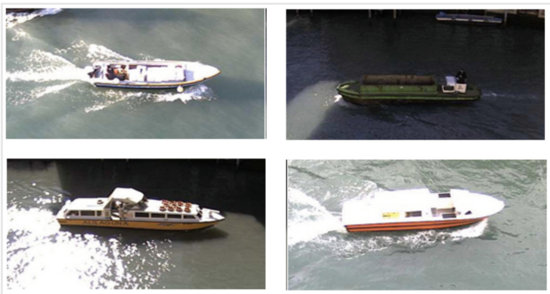


Fig. 1. Image samples from the MarDCT dataset.

Dataset 2 – VOC12. The second set, PASCAL VOC12 dataset is a general-purpose dataset. The dataset contains total 20 classes. However, there is a large variation in the number of objects per class throughout the image sets. The boats class is poorly represented related to other classes. Similarly, the objects of interest are occluded in some images. Such images may cause performance degradation of the models. Therefore, the dataset was refined to remove such images. Similarly, images which were not representative of harbor environment were also removed. The final dataset constitutes 5.29% in terms of images (267) and 5.81% in terms of the total objects (409) with objects to image ratio of 1.53. Figure 2 shows sample of the dataset.

Fig. 2. Image samples from the PASCAL VOC12 dataset.

Dataset 3 – Karachi Port Data. The third set, Karachi Port data was custom-made from the local harbor environment in Karachi Port area. The dataset was collected from 27 different locations along Karachi beach (as shown in Fig. 3) at morning, afternoon and evening times with different lighting conditions and zoom levels. All pictures were captured using high quality cellphone cameras. The dataset was passed through extensive cleaning which included orientation correction, zoom adjustment and cropping/removal of some irrelevant images etc. (as shown in Fig. 4). Therefore, the data was reduced to a total of 760 images. The data constitutes 15.06% in terms of images (760) and 36.92% in terms of the total objects (2598) with objects to image ratio of 3.42. Figure 4 shows sample of the dataset.

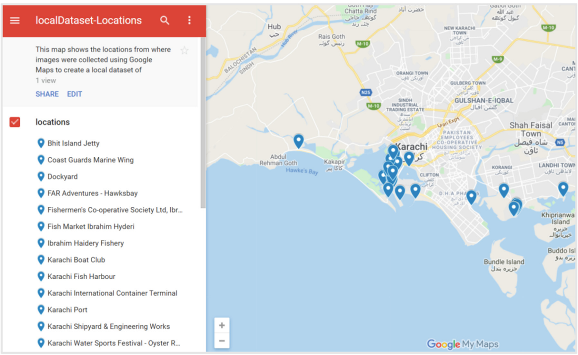
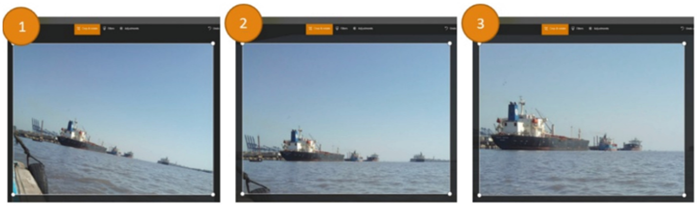


Fig. 3. Map locations for local dataset.

It can be seen that the MarDCT dataset has a higher percentage of images and objects, while the Karachi Port dataset has a higher no. of objects to images ratio.



(a) (b) (c)

Fig. 4. Clejaning of local dataset (a) Raw image (b) Rotation adjustment (c) Zoom adjustment

Table 1. Overall distribution of datasets

Dataset	Total Images	Total Objects	% of images	% of objects	Objects to images ratio
Dataset 1 (MarDCT)	4,020	4029	79.65%	54.48%	1
Dataset 2 (VOC12)	267	409	5.29%	5.53%	1.53
Dataset 3 (Karachi Port data)	760	2958	15.06%	39.99%	3.89
Total	5047	7396			

3.3 Evaluation Approach

All the experiments were conducted using Google Colaboratory Virtual Machine, having 12.69 GB RAM and NVIDIA Tesla T4 GPU. The performance of models was evaluated using Mean Average Precision (mAP). It is a commonly used metric to analyze performance of object detection models. It is calculated using Intersection over Union (IoU) thresholds, from 0.5 to 0.95 (with a step size of 0.05), and is represented as mAP@ [0.5:0.95]. Similarly, for comparison of model performances with that of other works, Detection Rate (DR) and False Alarm Rate (FAR) are used. DR and FAR are calculated using following equations.

$$DR = \frac{TP}{TP + FN}, FAR = \frac{FP}{TP + FP}$$

4 Implementation

The experiments were divided into following two different categories:

4.1 Detection Experiments

The detection experiments were conducted to investigate whether the use of general-purpose datasets could improve performance of the models and prevent overfitting in scenarios with limited available dataset. The capabilities of both pre-trained and custom-trained models was evaluated using different combinations of datasets from the local harbor environment.

Experiment 1 – MS COCO Dataset: The performance of pre-trained model trained on MS COCO dataset was tested on the local test dataset. This experiment served as a baseline comparison.

Experiment 2 – Local Dataset: The deep learning models were fine-tuned on local dataset, comprising of 660 and 100 images for train and test, respectively.

Experiment 3 – MarDCT Dataset: The models were fine-tuned on MarDCT dataset, comprising of 2,825 and 1,195 images for train and test, respectively.

Experiment 4 – Local + MarDCT Dataset: To diversify the training dataset, local dataset was combined with the MarDCT dataset. The overall dataset comprised of 3,485 and 1,295 images for train and test, respectively. The models were fine-tuned on the augmented dataset using transfer learning.

Experiment 5 – Local + MarDCT + VOC2012 Dataset: To diversify the training dataset even more, local dataset was combined with the MarDCT and VOC2012 dataset. The overall dataset comprised of 3,672 and 1,375 images for train and test, respectively. The models were fine-tuned on the augmented dataset using transfer learning.

4.2 Classification Experiments

Three deep learning models, namely Yolov3, Yolov5l and SSD MobileNet v2 were used to investigate the classification performance on MarDCT dataset. Thirteen different classes of boats were used to conduct these experiments. Transfer learning approach was utilized with weights trained on the MSCOCO dataset. Overall distribution of the data is shown in following Table 2.

5 Results and Discussion

The models obtained from various detection experiments were evaluated on a set of 100 images which were set aside from the local images' dataset for evaluation purposes. Table 3 below summarizes the results from each of the detection experiment discussed in previous Sect. 4.1.

During the training phase, the pre-trained models consistently resulted in low mAP score compared to other models. This performance was even more noticeable on the test dataset, indicating that these models are not suitable for direct use in harbor environment. Interestingly, the Yolov5s model trained on MarDCT dataset was able to achieve the highest score (i.e., mAP@0.5 = 99.73% and mAP@0.5:0.95 = 89.98%). However, the score dropped significantly on test dataset indicating that the model might have overfit. On the other hand, the Yolov5x model trained on Local + MarDCT + VOC2012 dataset achieved the highest score i.e., mAP@0.5 = 90.10% score. Moreover, the obtained result was quite close to its performance during training, indicating that the model is generalizing well to the unseen test data. Based on these observations, following are the conclusions from overall experimentation:

Table 2. Overall distribution of classes used in classification experiments

S. No	Boat Class	Total Images	Train Images	Test Images
1	Alilaguna	113	80	33
2	Ambulanza	85	60	25
3	Barchino	112	79	33
4	Lanciafino10mBianca	484	339	145
5	Lanciafino10mMarrone	355	249	106
6	Motobarca	215	151	64
7	Mototopo	878	615	263
8	Patanella	279	196	83
9	Polizia	71	50	21
10	Raccoltarifiuti	94	66	28
11	Topa	78	55	23
12	VaporettoACTV	949	665	284
13	Boat	154	108	46

- The pre-trained models are not a great fit for object detection in a maritime environment, and that they need to be re-trained on specific dataset for desired performance.
- Training / Finetuning the models by combining both general dataset and local dataset improves the performance of models due to dataset.

For the classification experiments, all the deep models performed well with average accuracy of above 95% with an overall best score of 98.78% achieved by Yolov5l model. Table 4 below summarizes the results obtained from classification experiments on the MarDCT dataset.

The results were also compared with online published results on MarDCT dataset to demonstrate the superior performance of deep learning models. Bloisi, Iocchi, and Pennisi (2013) used custom background subtraction method known as IMBS for vessel detection on MarDCT dataset [21]. Their approach resulted in DR = 0.54 and FAR = 0.14. Since, no information about ground truth or frames used for evaluation were provided, therefore, direct comparison with our model is not possible. However, to provide a comparison, a set of 16 frames were extracted from different videos having various boat sizes, pose, shadows, reflections, and boat wakes etc. The YOLOv5x model improved both the FAR and DR. Table 5 below shows the comparison of detection results with published work. Figure 5 shows the detection results obtained.

Similarly, for classification, Bloisi, Iocchi, and Pennisi (2015) used classical machine learning algorithms, KNN, J48 and RF to classify various boat categories in MarDCT dataset using a set of 11 features [20]. These features included color presence, number of edge pixels, length of boat, and standard deviation in the horizontal and vertical direction of the edges. Their approach achieved an accuracy of 73.14% by weighted majority

Table 3. Overall summary of detection results

S No	Model	Validation dataset		Test dataset	
		mAP@0.5	mAP@0.5:0.95	mAP@0.5	mAP@0.5:0.95
Experiment 1	Yolov3	55.40%	–	24.20%	07.76%
	Yolov3-tiny	33.00%	–	33.80%	07.59%
	Yolov5s	55.60%	–	51.00%	14.30%
	Yolov5x	68.70%	–	54.30%	17.70%
	SSD_inception_v2	–	24%	-1	-1
	SSD_mobilenet_v2	–	22%	-1	-1
Experiment 2	Yolov3	84.60%	–	84.60%	–
	Yolov3-tiny	65.20%	–	65.20%	–
	Yolov5s	86.20%	44.20%	86.20%	44.20%
	Yolov5x	88.80%	44.30%	88.80%	44.30%
	SSD_inception_v2	76.90%	34.20%	76.90%	34.20%
	SSD_mobilenet_v2	73.30%	28.81%	73.30%	28.81%
Experiment 3	Yolov3	99.68%	88.75%	24.20%	07.76%
	Yolov3-tiny	98.72%	81.45%	16.30%	04.05%
	Yolov5s	**99.73%**	**89.98%**	14.60%	04.59%
	Yolov5x	99.70%	89.26%	19.10%	06.38%
	SSD Inception v2	99.10%	79.40%	14.00%	05.20%
	SSD MobileNet v2	99.10%	77.90%	15.50%	06.20%
Experiment 4	Yolov3	98.44%	80.37%	87.00%	42.40%
	Yolov3-tiny	96.08%	71.29%	79.90%	32.60%
	Yolov5s	98.63%	82.35%	85.70%	42.90%
	Yolov5x	98.52%	82.78%	87.90%	**45.70%**
	SSD Inception v2	96.70%	74.30%	76.30%	34.50%
	SSD MobileNet v2	96.30%	73.20%	73.20%	31.40%
Experiment 5	Yolov3	97.02%	77.33%	86.80%	43.60%
	Yolov3-tiny	94.71%	68.77%	78.70%	32.20%
	Yolov5s	97.84%	79.44%	86.10%	43.20%
	Yolov5x	97.60%	80.04%	**90.10%**	44.20%
	SSD Inception v2	96.10%	72.40%	77.10%	35.20%
	SSD MobileNet v2	94.90%	71.00%	71.70%	32.20%

<u>*Note:*</u>– Metric unavailable due to the model and evaluation script incompatibility-1 Low metric score ~ 0

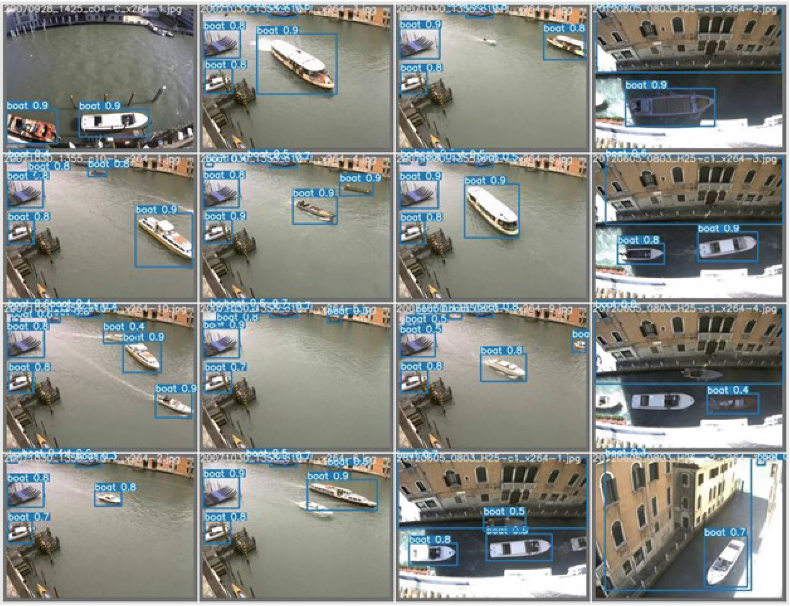


Fig. 5. Detection on MarDCT dataset using Yolov5x model.

Table 4. Overall summary of classification results

Model	Average Accuracy
Yolov3	97.48%
Yolov5l	**98.78%**
SSD MobileNet v2	96.36%

Table 5. Comparison of detection results with published results

Metrics	IMBS – Bloisi, Iocchi and Pennisi (2013)	Our Score
DR	0.54	**0.81**
FR	0.14	**0.07**

classification. As highlighted in Table 4, the YOLOv5l model achieved an average accuracy of 98.78%. This indicates superior capabilities of deep learning models to classify various boat classes. Table 6 below shows comparison of the proposed classification model with the published work.

Table 6. Comparison of classification results with published results

Metric	KNN + J48 + RF – Bloisi, Iocchi and Pennisi (2015)	Proposed approach using Yolov5l
Average accuracy	73.14%	**98.78%**

6 Conclusion and Future Work

This research used deep learning-based CNN models to detect and classify maritime vessels in a local harbor environment. The pre-trained YOLO and SSD models were used. The models were then re-trained with transfer learning using a combination of local, MarDCT and VOC12 datasets. The local data was manually collected at morning, afternoon and evening times with different lighting conditions and zoom levels, for a better representation of local harbor environment. The best performing object detection model, YOLOv5x achieved mAP of 90.10%. Similarly, the YOLOv5l model achieved the highest average accuracy score of 98.78% in classification experiments. The models outperformed previously published results on MarDCT dataset. The DR of 0.81, FAR of 0.07 and an average accuracy of 98.78% was obtained on the MarDCT dataset. The results suggest that the deep learning-based CNN models can be used to develop efficient maritime object detection and classification systems.

Further improvements can be made by collecting larger and more diverse datasets, which represent a wider range of maritime traffic and environmental conditions. Environment conditions such as fog, direct sunlight, reflections etc. introduce complexities in the detection stage. Therefore, a dataset collected at different times of the day and during different seasons of the year, would also improve the performance of deep learning models. Moreover, in harbor context, background often contains harbor area which may have buildings and trees etc. Training models to recognize these varying background elements can help to reduce false negatives. Similarly, potential methodologies for dataset expansion could also be explored for further improvements.

Acknowledgement. We acknowledge support from National Center of Big Data and Cloud Computing (NCBC) and Higher Education Commission (HEC) of Pakistan for conducting this research.

Data Availability. The custom dataset utilized in this study is available on request.

References

1. Bloisi, D.D., Previtali, F., Pennisi, A., Nardi, D., Fiorini, M.: Enhancing automatic maritime surveillance systems with visual information. IEEE Trans. Intell. Transp. Syst. **18**(4), 824–833 (2017). https://doi.org/10.1109/TITS.2016.2591321
2. Szpak, Z.L., Tapamo, J.R.: Maritime surveillance: tracking ships inside a dynamic background using a fast level-set. Expert Syst. Appl. **38**, 6669–6680 (2011)

3. Loomans, M.J.H., de With, P.H.N., Wijnhoven, R.G.J.: Robust automatic ship tracking in harbours using active cameras. In: 2013 IEEE International Conference on Image Processing (2013)
4. da Silva Moreira, R., Ebecken, N.F.F.: GWVT: a GPU maritime vessel tracker based on the wisard weightless neural network. Naval Eng. J. **129**, 109–116 (2017)
5. Moreira, R.D.S., Ebecken, N.F.F., Alves, A.S., Livernet, F., Campillo-Navetti, A.: A survey on video detection and tracking of maritime vessels. Int. J. Recent Res. Appl. Stud. **20** (2014)
6. Colab - Google Research. https://colab.research.google.com/
7. Kaggle. http://www.kaggle.com
8. Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L.: Imagenet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition (2009)
9. Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A.: The pascal visual object classes (voc) challenge. Int. J. Comput. Vision **88**, 303–338 (2010)
10. Lin, T.-Y., et al.: Microsoft coco: common objects in context. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) Computer Vision – ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V, pp. 740–755. Springer International Publishing, Cham (2014). https://doi.org/10.1007/978-3-319-10602-1_48
11. Cane, T., Ferryman, J.: Saliency-based detection for maritime object tracking. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (2016)
12. Lodeiro-Santiago, M., Caballero-Gil, P., Aguasca-Colomo, R., Caballero-Gil, C.: Secure UAV-based system to detect small boats using neural networks. Complexity **2019**, 1–11 (2019). https://doi.org/10.1155/2019/7206096
13. Rainey, K., Stastny, J.: Object recognition in ocean imagery using feature selection and compressive sensing. In: 2011 IEEE Applied Imagery Pattern Recognition Workshop (AIPR) (2011)
14. Joseph, S.I.T., Sasikala, J., Juliet, D.S.: Detection of ship from satellite images using deep convolutional neural networks with improved median filter. In: Jude Hemanth, D. (ed.) Artificial Intelligence Techniques for Satellite Image Analysis, pp. 69–82. Springer International Publishing, Cham (2020). https://doi.org/10.1007/978-3-030-24178-0_4
15. Dai, W., Mao, Y., Yuan, R., Liu, Y., Xuemei, Pu., Li, C.: A novel detector based on convolution neural networks for multiscale SAR ship detection in complex background. Sensors **20**(9), 2547 (2020). https://doi.org/10.3390/s20092547
16. Chan, M.T., Weed, C.: Vessel detection in video with dynamic maritime background. In: 2012 IEEE Applied Imagery Pattern Recognition Workshop (AIPR) (2012)
17. Hu, W.-C., Yang, C.-Y., Huang, D.-Y.: Robust real-time ship detection and tracking for visual surveillance of cage aquaculture. J. Vis. Commun. Image Represent. **22**, 543–556 (2011)
18. Kaimakis, P., Tsapatsoulis, N.: Background modeling methods for visual detection of maritime targets. In: Proceedings of the 4th ACM/IEEE International Workshop on Analysis and Retrieval of Tracked Events and Motion in Imagery Stream (2013)
19. Bloisi, D., Iocchi, L.: Argos—A video surveillance system for boat traffic monitoring in Venice. Int. J. Pattern Recognit Artif Intell. **23**, 1477–1502 (2009)
20. Bloisi, D.D., Iocchi, L., Pennisi, A., Tombolini, L.: ARGOS-venice boat classification. In: 2015 12th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS) (2015)
21. Bloisi, D.D., Pennisi, A., Iocchi, L.: Background modeling in the maritime domain. Mach. Vis. Appl. **25**, 1257–1269 (2014)
22. Prasad, D.K., Rajan, D., Rachmawati, L., Rajabally, E., Quek, C.: MuSCoWERT: multi-scale consistence of weighted edge radon transform for horizon detection in maritime images. JOSA A **33**, 2491–2500 (2016)

23. Shao, Z., Wu, W., Wang, Z., Du, W., Li, C.: Seaships: a large-scale precisely annotated dataset for ship detection. IEEE Trans. Multimedia **20**, 2593–2604 (2018)
24. Nie, S., Jiang, Z., Zhang, H., Cai, B., Yao, Y.: Inshore ship detection based on mask R-CNN. In: IGARSS 2018–2018 IEEE International Geoscience and Remote Sensing Symposium (2018)
25. Akiyama, T., Kobayashi, Y., Kishigami, J., Muto, K.: CNN-based boat detection model for alert system using surveillance video camera. In: 2018 IEEE 7th Global Conference on Consumer Electronics (GCCE) (2018)
26. Leclerc, M., Tharmarasa, R., Florea, M.C., Boury-Brisset, A.-C., Kirubarajan, T., Duclos-Hindié, N.: Ship classification using deep learning techniques for maritime target tracking. In: 2018 21st International Conference on Information Fusion (FUSION) (2018)
27. Gundogdu, E., Solmaz, B., Yücesoy, V., Koç, A.: Erhan Gundogdu, Berkan Solmaz, Veysel Yücesoy, Aykut Koç. In: Asian Conference on Computer Vision (2017)
28. Wang, Y., Ning, X., Leng, B., Fu, H.: Ship detection based on deep learning. In: 2019 IEEE International Conference on Mechatronics and Automation (ICMA) (2019)
29. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2016)
30. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2014)
31. Girshick, R.: Fast R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision (2015)
32. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. Adv. Neural. Inf. Process. Syst. **28**, 91–99 (2015)
33. Redmon, J., Farhadi, A.: Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767 (2018)

Fraud Detection Through Nature-Inspired Algorithms

Meenal Dugar(✉) and Aishwarya Asesh

University of Pennsylvania, Philadelphia, USA
u1278112@utah.edu, asesh@adobe.com

Abstract. The proliferation of credit card payments has led to increased convenience but also an uptick in transaction fraud. This necessitates data-mining approaches to detect and prevent fraudulent activities. Transaction fraud detection involves multiple steps, including data analysis, preprocessing, feature selection, and hyperparameter optimization. Given the prevalence of anonymized features in transaction datasets, we focused on enhancing model performance through feature selection using nature-inspired algorithms. Our experiments revealed that this approach, when combined with machine learning models, holds promise for fraud prevention. Among the tested nature-inspired algorithms, the Grey Wolf Optimizer stood out, improving the model's ROC AUC score by 1.4% while selecting only half of the features compared to recursive feature elimination.

Keywords: Fraud Detection · Nature-Inspired Algorithms · Feature Selection · Credit Card Transactions · Machine Learning · Data Mining

1 Introduction

Payments through credit cards have become increasingly prevalent, offering users a secure and convenient alternative to cash transactions. As this trend grows, the incidence of transaction frauds has proportionally escalated, highlighting an urgent need for robust automatic detection mechanisms. Credit card transactional data, inherently gathered by financial institutions for analytics and fraud investigation, is rich with features and attributes. This extensive data can be framed within the machine learning paradigm as a binary classification challenge, where transactions are labeled as either genuine or fraudulent.

However, one of the core challenges in dealing with such datasets is the presence of numerous, often anonymized, features which makes the feature selection process pivotal to the success of any predictive model. Inaccurate or redundant features can degrade model performance, making it imperative to employ advanced techniques for feature engineering. To this end, nature-inspired algorithms (NIAs), which are metaheuristic search methods inspired by natural processes, present a promising approach.

C. Biele et al. (Eds.): MIDI 2023, LNNS 1076, pp. 18–27, 2024.
https://doi.org/10.1007/978-3-031-66594-3_2

This paper aims to introduce a groundbreaking technique for transaction fraud detection by synergizing the capabilities of machine learning and NIAs. At its core, our method trains machine learning models on an optimized subset of features, meticulously handpicked by the NIAs. By conducting comparative analyses across various NIAs, we endeavor to determine the most effective algorithm tailored for this specific challenge. Through this confluence of data analytics, machine learning, and NIAs, we aspire to elevate the accuracy and efficiency of fraud detection systems in the credit card industry.

2 Related Works

With the surge in credit card transactions, the occurrence of transaction frauds has witnessed a parallel escalation. While data-mining techniques present themselves as a conventional defense against frauds, their efficacy is contingent upon the specificities of the data at hand, underscoring the absence of a universally effective solution. To bridge this gap, a detailed assessment of machine learning methodologies, predominantly focusing on neural network architectures, for addressing credit card fraud detection was undertaken [6]. Classical machine learning algorithms such as Support Vector Machines [10] and Random Forest [11] have been thoroughly evaluated in this context.

Interestingly, Genetic programming, a unique domain-specific approach, emerges as a potent tool, registering a noteworthy enhancement of almost 17

An intricate view of transactions reveals two analytical paradigms: treating each transaction as an isolated entity or analyzing a consolidated transactional history. For models to remain resilient to evolving transactional patterns, the incorporation of time series models, notably the Long-Short Term Memory (LSTM) recurrent neural networks, has been advocated [10].

Invariably, public datasets pertinent to transaction fraud detection are feature-rich, necessitating rigorous feature engineering and selection phases. Historically, NIAs have demonstrated superior prowess in addressing these feature selection challenges. Such problems, under the purview of metaheuristic approaches, particularly NIAs, align closely with combinatorial optimization paradigms [8].

Within the arena of nature-inspired algorithms for feature selection, the relevance of rough sets is undeniable, further elaborated in the context of data mining in [7]. Empirical studies have amalgamated rough sets with algorithms like the Firefly Algorithm [1], Particle Swarm Optimization [9], and the Bat Algorithm [2], drawing from diverse datasets to validate the efficacy of NIAs.

A series of comprehensive surveys have honed in on the application of NIAs in feature selection, presenting a plethora of algorithms including, but not limited to, Artificial Bees Colony, Cuckoo Search, Ant Colony Optimization, and Genetic Algorithms [5,8,15].

Generative AI models, by leveraging their capability to generate data that mirrors real-world patterns, can be instrumental in fraud detection by enriching datasets, especially when genuine anomalies are scarce. This enhancement aids in training more robust models, capable of discerning intricate fraudulent behaviors. However, challenges arise when these generative models inadvertently produce data that blurs the distinction between legitimate and fraudulent transactions, potentially leading to increased false positives. Additionally, the computational intensity of generative models and their inherent opacity can hinder their real-time applicability and interpretability, crucial for actionable fraud detection.

3 Problem Definition and Dataset Characteristics

Datasets tailored for transaction fraud detection in data science research commonly present specific challenges: they are often highly imbalanced, feature-rich, and predominated by anonymized attributes.

Our proposed approach leverages a dataset sourced from the Kaggle EEE-CIS Fraud Detection competition[1]. Characteristic of transaction fraud datasets, this compilation is abundant with diverse attributes, totaling 434 features, spanning demographics, credit card particulars, transaction details, and more. A significant portion of these features remain anonymized or ambiguously defined. Consequently, caution is imperative to prevent inadvertent label leakage and to ensure meaningful interpretation of data analyses or model outcomes tied to these anonymous attributes. With nearly 600k instances, the dataset is aptly suited for machine learning. Yet, the inherent class imbalance, as visualized in Fig. 1, necessitates meticulous handling during model training and evaluation.

Given the anonymity of the data, domain expertise can't guide feature selection. Manually curating from such an extensive feature set is not only arduous but nearly unfeasible. Implementing feature selection methodologies becomes paramount, not only for manageability but to enhance model simplicity and interpretability.

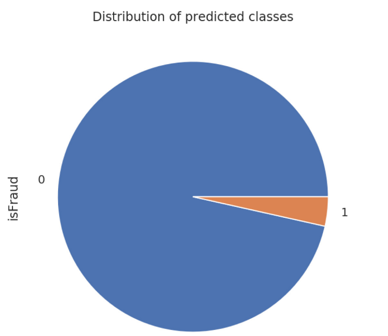


Fig. 1. Class Distribution Analysis: Evidence of High Data Imbalance.

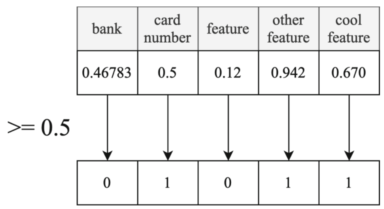


Fig. 2. Projection of float vector optimized by NIAs to a binary feature selection vector.

[1] https://www.kaggle.com/c/ieee-fraud-detection/data.