Transactions on Energy Systems and Engineering Applications

*Article*

# Technological prototype with artificial intelligence for military security in river environments

**Sergio Sánchez[1,*], Daniel Casillo[1], Andres Merlano[1], Julian Barrera[2], Alex Morales[1] and Elbert Contreras[1]**

[1] Facultad de Ingeniería, Corporación Universitaria Antonio José de Sucre -UAJS, Sucre, Colombia.
[2] Escuela de Formación Infanteria de Marina -ESFIM, Coveñas, Colombia.
[*] Correspondence: docente_investigador9@uajs.edu.co

*Received: 16 January 2024; Accepted: 21 October 2024; Published: 06 November 2024*

**Abstract:** Maritime and river security is one of the main concerns of military forces due to the large number of illicit activities that occur. Not to mention the extensive areas that must be monitored, and the weather conditions that can occur. Currently, technologies have become fundamental to leave aside manual surveillance for intelligent systems that allow remote sensing, traffic control, and object detection. Based on the aforementioned problems, the purpose of this research was to design a technological prototype with artificial vision based on an artificial intelligence model to detect water vessels and people in river environments as a support tool for military security. The prototype used at hardware level a Raspberry Pi 3 card and four pre-trained models based on R-CNN, YOLO, EfficientDet and SSD (single shot multibox). The best-performing model was the Mobilnet V2 SSD, with an mAP of 0.83 and an FPS of 5. Finally, this tool can contribute to strengthening the strategic, tactical, and operational capabilities of actors in the military intelligence sector, aimed at protecting sovereignty and territorial integrity to establish an environment of security in society.

© 2024 by the authors. Published by Universidad Tecnológica de Bolívar under the terms of the Creative Commons Attribution 4.0 License. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. https://doi.org/10.32397/tesea.vol5.n2.607

## 1. Introduction

River areas are very complex scenarios where different illicit activities take place; due to their large areas, multiple access points, sudden climate changes, lack of visibility and criminal groups, which represents a great challenge for the military forces. Not to mention that the electrical and telecommunications infrastructure is not very good [1].

On the other hand, river traffic has increased in recent decades with economic growth, also increasing the expansion of ports. These spaces are saturated with many incoming and outgoing vessels. Therefore, careful monitoring and proactive control of traffic is important to avoid illegal activities [2]. Not to mention,

**How to cite this article:** Sánchez, Sergio; Casillo, Daniel; Merlano, Andres; Barrera, Julian; Morales, Alex; Contreras, Elbert. Technological prototype with artificial intelligence for military security in river environments. *Transactions on Energy Systems and Engineering Applications*, 5(2): 607, 2024. DOI:10.32397/tesea.vol5.n2.607

that ship detection is a complex problem that depends on exogenous (weather and traffic density) and endogenous (ship type, on-board technologies, etc.) factors that influence navigation [3].

Monitoring and control of illegal vessel traffic is not simple, due to the following characteristics; increase of boats in ports, waterways are becoming increasingly congested, routes often merge, split or cross each other [4], the inertial motion of the vessel makes rapid changes of direction and speed difficult, the movement of the vessel is affected not only by the travel schedule, but also by the surrounding traffic and environment [5]. However, the fight against this type of illegal activities requires a large budget and deployment of personnel in the area, which is exposed to any type of confrontation with illegal armed groups.

Therefore, there is a need to use cutting-edge technologies of Industry 4.0, such as deep neural networks (CNN), for the detection of objects of interest in images and videos. Currently there are different pre-trained architectures with millions of images for multiple object recognition, which use the transfer learning strategy to take advantage of previously learned knowledge [6–8]. Where there are candidate region based object detection algorithms; (R-CNN), Fast R-CNN, Faster R-CNN and other models [9–11]. Y Regression-based object detection algorithms, also known as single-stage object detection algorithms, mainly include YOLO and SSD series algorithms [12, 13].

In the literature we can find some researches such as the one developed by the authors [14], who implemented an enhanced convolutional neural network (CNN) based on DarkNet53 architecture for ship detection in different weather conditions (rain, fog and low illumination). Experimental results obtained under different monitoring conditions showed that the proposed method significantly outperforms other strategies such as SSD, Faster R-CNN, YOLOv2 and YOLOv3. Using the metrics of accuracy, robustness and efficiency.

On the other hand, researchers [15] used a deep neural network based on RCNN architecture for maritime surveillance in coastal areas, due to the entry and exit of vessels from ports without permits. The model was trained in MATLAB software and compared with other architectures such as YOLO, where it presented better performance.

The authors [16] proposed an artificial intelligence model for ship detection in HR optical remote sensing images based on the Faster-R-CNN architecture, an active rotating filter and attention modules. The architecture was trained with the ocean remote sensing image dataset in PASCAL VOC format. The model was tested and evaluated on a computer with a 3.80 GHz Intel Core i7-10070 CPU and a GeForce GTX 2080 Ti GPU. The model was run on the open source Pytorch framework. The average accuracy of the proposed model showed an improvement of 5.49 % compared to the ResNet50 benchmark method.

For all of the above, this proposal is framed in the development of a support tool for river security using Industry 4.0 technologies, as a complementary prototype for military security.

## 2. Methods and Materials

### *2.1. Dataset description*

To train the models, images were taken from the KAGGLE platform, the Boat types dataset was used with 1000 images of vessels, of various sizes and different types (buoy, cruise ship, ferry, cargo ship, gondola, inflatable boat, kayak, paper boat and sailboat). We also worked with the people image dataset with 1200 photographs. All labels were made with the labelImg tool.

### *2.2. Implemented architectures*

The architectures used for the object detection task use feature extractors to transform visual information (pixels in an image) into meaningful and representative features. These features are learned patterns that help identify objects and their attributes in an image. These extractors are usually convolutional neural networks (CNNs) pre-trained on large datasets, which allows them to learn to detect common and generic features present in many images, such as edges, textures, shapes, among others [17–19]. The extractors used for this research can be seen below.

#### 2.2.1. R-CNN (Region-based Convolutional Neural Network)

It is based on the region of interest and is mainly used for object detection in images. It can detect multiple objects in an image, usually in the range of tens to hundreds of different classes [20, 21].

*Advantages:*

- High accuracy in object detection due to its region-of-interest approach.
- It can use region-specific features, which improves performance in detection tasks.

*Disadvantages:*

- It is computationally expensive and slow in execution time because it needs to extract regions of interest and perform classification and box fitting separately.
- No es adecuado para aplicaciones en tiempo real debido a su alta carga computacional.

#### 2.2.2. YOLO (You Only Look Once)

is a real-time object detection model that uses a grid approach to detect objects in images. It can detect multiple objects in a single pass through the image and can generally handle various classes of objects, such as people, cars, animals, etc [22, 23].

*Advantages:*

- Efficient and fast runtime, making it suitable for real-time applications.
- Can detect multiple objects in a single pass through the image, reducing computational complexity.
- Good performance on images with multiple small or overlapping objects.

*Disadvantages:*

- It may have slightly lower accuracy compared to other models, especially for small objects or in scenes with very close objects.

#### 2.2.3. EfficientDet

is an object detection architecture that is based on the concept of efficient neural networks. Like YOLO, it can detect multiple objects in a single pass and can generally handle multiple classes of objects [24, 25].

*Advantages:*

- It uses more efficient neural network architectures, allowing a balance between accuracy and computational efficiency.
- Can achieve high levels of accuracy with fewer parameters and computational resources.

*Disadvantages:*

- Requires more fine-tuning effort to obtain the best performance compared to other models.

#### 2.2.4. SSD (Single Shot Multibox Detector)

es otro modelo de detección de objetos en tiempo real que utiliza el concepto de predicciones a múltiples escalas para detectar objetos. Puede detectar múltiples objetos y generalmente puede manejar varias clases de objetos [26,27].

*Advantages:*

- Efficient and fast runtime, suitable for real-time applications.
- Good performance in small object detection

*Disadvantages:*

- May have slightly lower accuracy compared to other more complex models.
- Less efficient than YOLO in terms of detection speed.

A block diagram of the backbone used for this research is shown below.

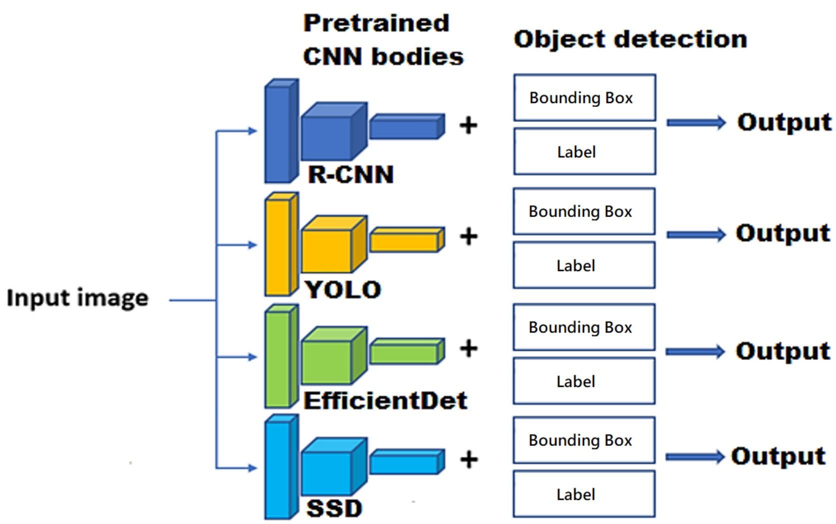


**Figure 1.** Architectures implemented in the proposed system.

In Figure 1 we can see the complete architecture for the object detection task, where four pre-trained models with different backbones were evaluated, two output layers were used, one for the detection of the boxes and the other for classes.

### *2.3. Experiment configuration*

For the development of this research, the pre-trained models SSD Mobilnet V2, R-CNN-Resnet 50, YOLOv2 and EfficientDet were used on the TensorFlow Lite framework and Darknet. On the other hand, the transfer learning strategy was implemented, with the objective of reusing the features learned from the frozen models, previously trained with databases such as MS COCO (Microsoft Common Objects in Context), ImageNet and PASCAL VOC. Then, a fine-tuning was performed for the object detection task to recognize the person class and aquatic vessels. For the training and testing of the model, the cross-validation technique was used, where 70% of the images were used for training and 30% for validation. In Figure 2, we can see the block diagram of the developed prototype.

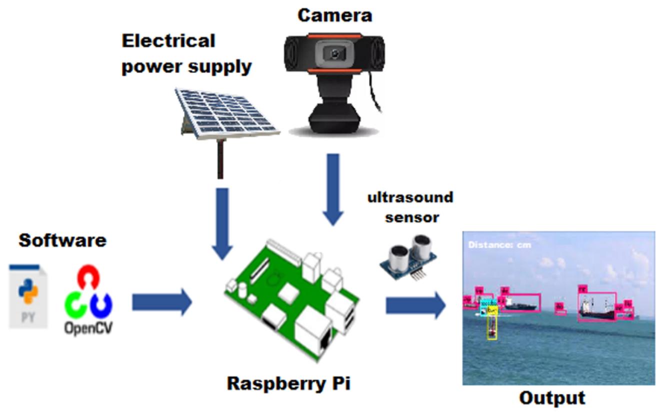


**Figure 2.** Block diagram of the proposed system.

From the figure above, it can be seen that for this research we used a low-cost arm card (Raspberry pi), a full HD webcam with CMOS sensor with a maximum video resolution of 720p, a camera image resolution of 2Mpx, a 20W photovoltaic system and a HC-SR05 ultrasound sensor with a maximum distance of 5 meters.

## 3. Results and Discussions

For the development of this research, four pre-trained nets were used to detect boats and people that may pose a risk in river areas at times and places that are not permitted. Four experiments were conducted in laboratory scenarios to validate the functionality of the architectures.

### 3.1. Experiment 1

We worked with the Mobilnet V2 SSD architecture, focusing on the detection of objects of interest in images and videos in real time. The qualitative results are shown below.

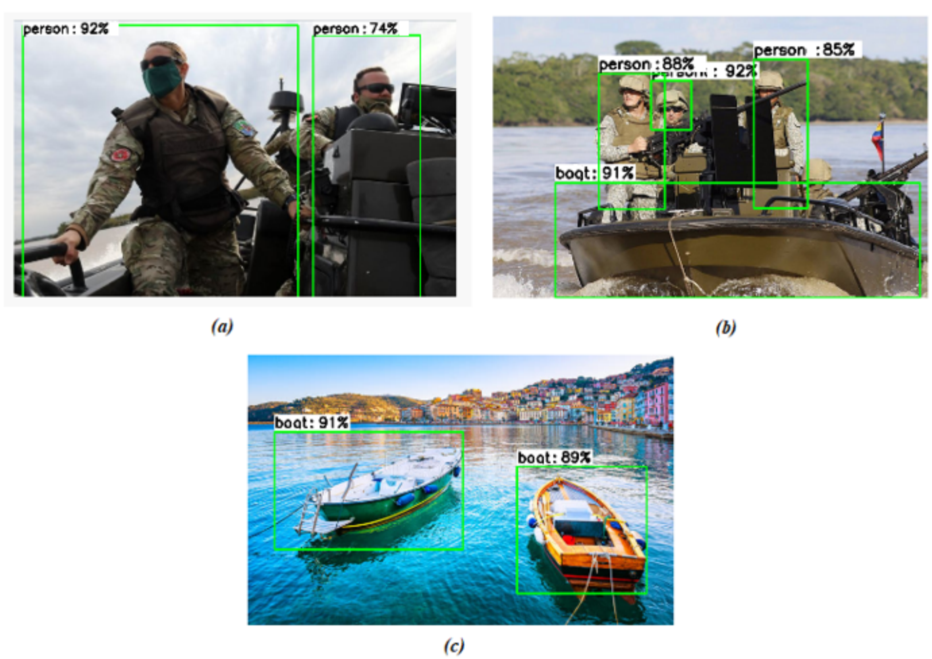


**Figure 3.** Model trained with the neural network SSD-MobileNet-v2.

From the previous results, it is possible to observe the percentages of success of the model in the objects of interest. Highlighting that this architecture is lightweight to be used in different electronic devices, detects objects at multiple scales and levels, can share feature learning and also has the ability to detect multiple classes of objects.

### 3.2. Experiment 2

In this test we used the R-CNN-ResNet architecture, which seeks to extract proposals or regions of interest to apply classification and regression techniques for object detection. From the above, we found the following qualitative results.

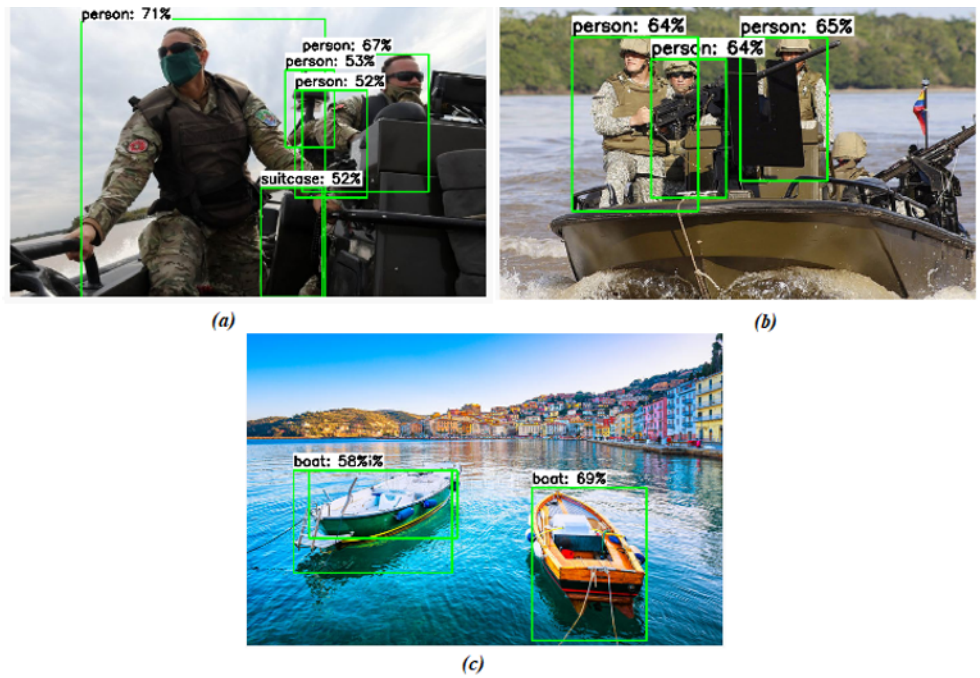


**Figure 4.** Model trained with the neural network R-CNN-ResNet 50.

From Figure 4, we can highlight a neural network model that is able to vary the depth, solve the vanishing gradient problem and the dimensionality. But it has drawbacks such as; high computational consumption in the inference stage, it cannot work in real time applications and it can be affected by over-fitting for training with few labeled images. On the other hand, it is observed to be less efficient and accurate than the SSD-MobileNet-v2 architecture.

### 3.3. Experiment 3

In this experiment we used the YOLO-V2 architecture (You Only Look Once, version 2) which is able to detect objects in real time and with good accuracy, finding the following qualitative results.


**Figure 5.** Model trained with the neural network YOLO-V2.

YOLO-V2 has high speed, efficiency and a good ability to detect multiple objects in a single pass. However, its performance is limited when compared to other detectors, especially on small objects and with little overlap with the background. Also, it suffers with low quality and low resolution input images. This can be contrasted with the results obtained previously, where it is observed that the model is not able to detect all objects accurately as SSD-MobileNet-v2 does and omits some objects of interest that the R-CNN-ResNet50 architecture recognizes.

### *3.4. Experiment 4*

For this last test, we worked with the EfficientDet (0) architecture, which uses the composite scaling technique to optimize the size and depth of the model, allowing it to achieve high levels of accuracy with fewer parameters compared to other architectures. The qualitative results can be seen below.

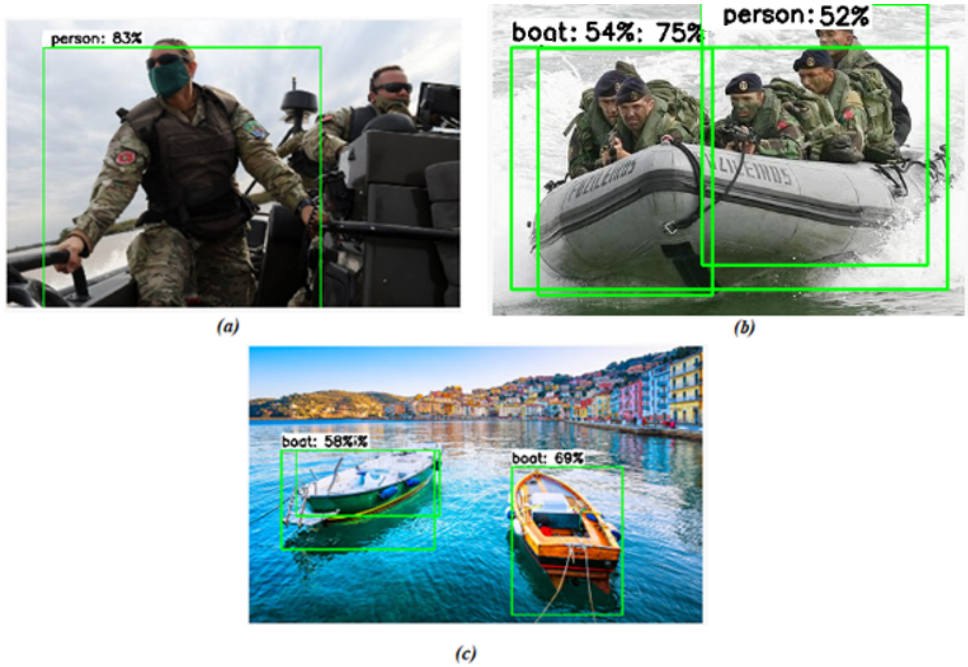


**Figure 6.** Model trained with the neural network EfficientDet (0).

The EfficientDet(0) architecture features good computational performance, a single-stage object detection approach, ability to detect objects at different scales and sizes, and good accuracy with fewer parameters. Its disadvantages mainly focus on the trade-off between efficiency and accuracy, the complexity of hyperparameter tuning, and limitations in detecting very small objects or objects with very particular shapes. From the above results, lower accuracy of the objects of interest is evident compared to the architectures of experiments 1, 2 and 3.

### *3.5. Experiment 5*

After selecting the best architecture, a joint test of the system was performed using the ultrasound sensor (HC-SR04) and the artificial intelligence model, for the estimation of the distance and location of the objects of interest. As detailed below.

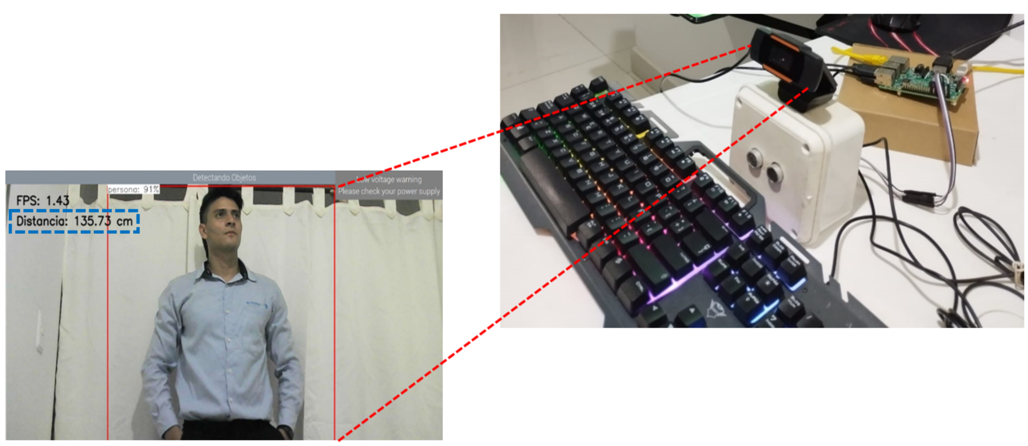


**Figure 7.** Complete technological prototype.

We can see in the Figure 7, that in addition to detecting the object, it estimates the distance at which it is located, which is given in the unit of centimeters (cm). An algorithm was also designed to store the objects of interest in real time, in a folder of the Raspberry Pi card operating system, where the name of the object, the date, the time at which it is detected and the distance of the object are recorded. In addition, screenshots of the most important objects (people and boats) are taken, organizing all this information in a folder.

Finally, all the results obtained from the previous experiments can be contrasted with the quantitative results.

**Table 1.** Quantitative results of the pre-trained models.

| Model | FPS | Resolution | mAP |
|---|---|---|---|
| **SSD Mobilnet V2** | 5 | 300x300 | 0.83 |
| **R-CNN-resnet 50** | 0.5 | 300x300 | 0.61 |
| **YOLOv2** | 0.9 | 416x416 | 0.51 |
| **EfficientDet (0)** | 1.8 | 512x512 | 0.69 |

In Table 1, we can validate that the model with the best performance is the Mobilnet V2 SSD, presenting a good mAP metric in the training stage, an acceptable FPS (frames per second) rate for a low-cost prototype. These results demonstrate the importance of using video surveillance systems with artificial intelligence to protect the military in areas where there is drug trafficking, weapons and criminal acts, as well as to protect local communities. Artificial intelligence technology allows you to quickly detect and react to suspicious activities, identify intrusions and monitor the environment in real time, providing constant and precise surveillance.

## 4. Conclusions

In the field of deep learning, significant advances in algorithm performance have been observed over time, however, challenges remain to be addressed. The use of parallel computing with high-powered graphics processing units (GPUs) has reduced the time required to train artificial neural networks. However, it has been shown that taking advantage of pre-trained models makes it possible to implement them without the need for full training, which in turn allows their storage on low-cost electronic cards, minimizing computational times without relying on a GPU. This situation provides researchers with an accessible starting point to enter this field, avoiding the tedious task of building and training an object detector from scratch.

Keras and TensorFlow, as deep learning platforms, offer a variety of pre-trained models available for this purpose. As we move forward, the gap between different detectors is narrowing. Single-shot detectors are adopting more complex designs to improve their accuracy, while region-based detectors are accelerating their operations for speed.

Pretrained convolutional neural networks have proven to be a very practical solution to problems related to the military field. Particularly the field of security using computer vision.

For future work, it is suggested to work with cards with higher capacity in order to run more accurate and real-time models.

**Funding:** This work was partially funded by Corporación Universitaria Antonio José de Sucre.

**Author contributions:** Conceptualization, S. Sanchez , D. Casillo and A.Merlano; Software, S. Sanchez , D. Casillo and A.Merlano; Investigation, S. Sanchez , D. Casillo and A.Merlano; Writing – Original Draft Preparation, S. Sanchez , D. Casillo, A.Merlano, J. Barrera, A. Morales and E. Contreras; Writing – Review

& Editing, S. Sanchez , D. Casillo, A.Merlano, J. Barrera, A. Morales and E. Contreras; Supervision, S. Sanchez , D. Casillo, A.Merlano, J. Barrera, A. Morales and E. Contreras.

**Disclosure statement:** The authors declare no conflict of interest.

## References


[1] Luis Orozco, Mikel Ibarra, Alejandro Chaparro Ortiz, Adriana Castañeda, Armando Triana, Carlos Forero, Carlos Moreno, Diego Peñuela, and Paula Ortíz. *Tecnologías emergentes para la seguridad y defensa nacional: los retos de los sistemas ciberfísicos para luchar contra el crimen organizado transnacional*. 06 2021.

[2] Mrunalini Nalamati, Nabin Sharma, Muhammad Saqib, and Michael Blumenstein. Automated monitoring in maritime video surveillance system. In *2020 35th International Conference on Image and Vision Computing New Zealand (IVCNZ)*, pages 1–6, 2020.

[3] Kwang-Il Kim, Jung Jeong, and Byung-Gil Lee. Study on the analysis of near-miss ship collisions using logistic regression. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 21:467–473, 05 2017.

[4] Kwang-Il Kim, Jung Jeong, and Byung-Gil Lee. Study on the analysis of near-miss ship collisions using logistic regression. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 21:467–473, 05 2017.

[5] Daewon Park. Syntactic-level integration and display of multiple domains' s-100-based data for e-navigation. *Cluster Computing*, 20, 03 2017.

[6] Fuzhen Zhuang, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Hui Xiong, and Qing He. A comprehensive survey on transfer learning, 2020.

[7] Fuzhen Zhuang, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Hui Xiong, and Qing He. A comprehensive survey on transfer learning. *Proceedings of the IEEE*, 109(1):43–76, 2021.

[8] Zhuangdi Zhu, Kaixiang Lin, and Jiayu Zhou. Transfer learning in deep reinforcement learning: A survey, 09 2020.

[9] Puja Bharati and Ankita Pramanik. *Deep Learning Techniques—R-CNN to Mask R-CNN: A Survey*, pages 657–668. 01 2020.

[10] Ross Girshick. Fast r-cnn. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 1440–1448, 2015.

[11] Madhusri Maity, Sriparna Banerjee, and Sheli Chaudhuri. Faster r-cnn and yolo based vehicle detection: A survey. 04 2021.

[12] Peiyuan Jiang, Daji Ergu, Fangyao Liu, Ying Cai, and Bo Ma. A review of yolo algorithm developments. *Procedia Computer Science*, 199:1066–1073, 2022. The 8th International Conference on Information Technology and Quantitative Management (ITQM 2020 2021): Developing Global Digital Economy after COVID-19.

[13] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg. *SSD: Single Shot MultiBox Detector*, page 21–37. Springer International Publishing, 2016.

[14] Wen Liu, Weiqiao Yuan, Xinqiang Chen, and Yuxu Lu. An enhanced cnn-enabled learning method for promoting ship detection in maritime surveillance system. *Ocean Engineering*, 235:109435, 07 2021.

[15] Mohana Mohana. Object detection and classification algorithms using deep learning for video surveillance applications. *International Journal of Innovative Technology and Exploring Engineering*, 8:386–395, 06 2019.

[16] Yuxin Dong, Fukun Chen, Shuang Han, and Hao Liu. Ship object detection of remote sensing image based on visual attention. *Remote Sensing*, 13(16), 2021.

[17] Yali Amit, Pedro Felzenszwalb, and Ross Girshick. *Object Detection*, pages 1–9. 01 2020.

[18] Zhengxia Zou, Keyan Chen, Zhenwei Shi, Yuhong Guo, and Jieping Ye. Object detection in 20 years: A survey, 2023.

[19] Xin Huang, Xinxin Wang, Wenyu Lv, Xiaying Bai, Xiang Long, Kaipeng Deng, Qingqing Dang, Shumin Han, Qiwen Liu, Xiaoguang Hu, Dianhai Yu, Yanjun Ma, and Osamu Yoshie. Pp-yolov2: A practical object detector, 2021.

[20] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Region-based convolutional networks for accurate object detection and segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(1):142–158, 2016.

[21] Xingxing Xie, Gong Cheng, Jiabao Wang, Xiwen Yao, and Junwei Han. Oriented r-cnn for object detection, 2021.

[22] Wei Fang, Lin Wang, and Peiming Ren. Tinier-yolo: A real-time object detection method for constrained environments. *IEEE Access*, 8:1935–1944, 2020.

[23] Chengji Liu, Yufan Tao, Jiawei Liang, Kai Li, and Yihang Chen. Object detection based on yolo network. pages 799–803, 12 2018.

[24] Mingxing Tan, Ruoming Pang, and Quoc V. Le. Efficientdet: Scalable and efficient object detection. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10778–10787, 2020.

[25] Delia Velasco-Montero, Jorge Fernández-Berni, Ricardo Carmona-Galan, and Ángel Rodríguez-Vázquez. Performance analysis of real-time dnn inference on raspberry pi. page 14, 05 2018.

[26] Sheping Zhai, Dingrong Shang, Shuhuan Wang, and Susu Dong. Df-ssd: An improved ssd object detection algorithm based on densenet and feature fusion. *IEEE Access*, 8:24344–24357, 2020.

[27] Wu Zheng, Weiliang Tang, Li Jiang, and Chi-Wing Fu. Se-ssd: Self-ensembling single-stage object detector from point cloud, 2021.