International Journal of Latest Engineering Research and Applications (IJLERA) ISSN: 2455-7137

Volume – 09, Issue – 03, March 2024, PP – 32-36

A Real-Time Automated Port Entry and Exit Management Framework for Fishing Vessels Using YOLOv8

Jing-Yi Guo[1*], Yao-Chung Chen[2], Chen-Yu Hao[2], Tien-Yin Chou[2]

[1]*Ph.D. Program for Infrastructure Planning and Engineering Feng Chia University, Taichung, Taiwan*

[2]*GIS research center, Feng Chia University, Taichung, Taiwan*

Abstract: This study develops an automated management framework for the real-time recognition of boat entries and exits at the port. It utilizes the Real-Time Streaming Protocol (RTSP) for the acquisition of real-time image streams from the fishing port. The framework leverages computer vision techniques to analyze each frame and employs You Only Look Once version 8 (YOLOv8) for boat detection. Additionally, the BotSort algorithm (Bounding Box Object Sort and Tracking) is employed to track the detected boats. The framework also features an electronic fence designed in this research to determine whether boats are entering or departing from the port. It provides real-time cumulative counts, records relevant information in both text and images, and effectively automates the management of fishing boat entries and exits, thereby enhancing overall management efficiency.

Keywords: Real-Time Streaming Protocol, You Only Look Once, Bounding Box Object Sort and Tracking, electronic fence

I. Introduction

Certainly, the integration of Artificial Intelligence (AI) and computer vision technologies is becoming increasingly pivotal across various industries. In Taiwan, the strategic focus on Smart Governance is indicative of the nation's commitment to leveraging advanced technologies to enhance government operations and policymaking. The field of Image Recognition is experiencing rapid evolution, driven by technological advancements such as enhanced computer hardware, deep learning algorithms, the proliferation of the Internet of Things (IoT), and the advent of high-speed 5G communication. These developments have empowered real-time applications of image recognition, providing a multitude of opportunities for improving efficiency and decision-making.One notable application involves the utilization of computer vision and web cameras to capture real-time imagery of fishing vessels. This enables the automatic identification and logging of vessel entry and exit activities. The benefits of such a system are multifaceted, including increased operational efficiency for vessel management, substantial reductions in labor and time costs, minimization of data loss risks, and the establishment of real-time monitoring and response mechanisms. The automated data accumulation further supports in-depth analysis and informed decision-making, offering potential for substantial improvements in fisheries and port management. These enhancements encompass refined operations and more effective policy formulation. As technology continues to advance, the seamless integration of AI and computer vision will likely find even broader applications, revolutionizing various sectors and offering new ways to streamline processes, improve efficiency, and enhance decision-making capabilities.

II. Related Works

Deep learning (DL) is an integral component of machine learning (ML) and serves as a foundational theory in artificial intelligence (AI). Introduced in the early 2000s, DL initially did not garner significant attention due to its scalability requirements and the demand for extensive computational resources. It was not until after 2006, following the widespread adoption of the internet, the accumulation of diverse and abundant big data, and the advancement of information and communication technologies, such as high-end computing resources, fast data transmission, cloud computing, and edge computing, that DL gained rapid diffusion. As a result, it has progressively become a prominent choice for contemporary AI applications. These applications encompass various domains, including weather forecasting, stock market prediction, speech recognition, object detection, character recognition, intrusion detection, landslide detection, time series forecasting, text classification, gene expression analysis, bioinformatics, unstructured text data mining, and video processing.

Object detection and tracking based on image data are pivotal and highly valuable technologies within the realm of computer vision. These technologies have widespread applicability in smart transportation, smart agriculture, disaster management, smart manufacturing, and smart cities. Overfeat[1], introduced in 2013, was an important deep learning algorithm for object detection and image classification, marking a significant milestone in early deep learning's application to computer vision. Over the past decade, numerous object detection algorithms have been developed, typically categorized into two-stage (Two-Stage) and one-stage (Single-Stage) methods.

Two-stage object detectors, as their name suggests, divide the object detection process into two stages. The first stage, often referred to as "candidate region generation," is responsible for generating potential object locations. The second stage focuses on object classification and precise localization. These detectors typically deliver higher detection accuracy but come with increased model complexity and slower processing speed. Representative methods in this category include the R-CNN series (Regions with Convolutional Neural Networks). The pioneering R-CNN[2], introduced in 2014, marked an early breakthrough in two-stage object detection, introducing the concept of candidate region generation. Subsequent enhancements built upon this model, improving detection performance or introducing segmentation capabilities, as seen in Fast R-CNN[3], Faster R-CNN[4], Mask R-CNN[5], and the Feature Pyramid Network (FPN)[6].

One-stage object detectors usually employ a single neural network to process the entire image at once, including object detection, localization, and classification. These methods are typically designed for simplicity and speed, making them suitable for real-time applications. However, they may compromise object localization precision and bounding box accuracy due to single-pass processing. Examples of one-stage detectors include SSD（Single Shot MultiBox Detector）[7]、DSSD（Deconvolutional Single Shot Detector）[8]、RetinaNet[9]、RefineDet++[10]、M2Det[11]及 YOLO（You Only Look Once）[12-14] series. The YOLO series, introduced in 2015 with Version 1, has evolved through multiple versions, culminating in Version 8, and includes specialized models like X, R, PP, and NAS[13]. YOLO's ongoing evolution has established it as the mainstream algorithm for object detection. This project is built on Ultralytics' open-source software YOLOv8, which was released earlier this year.

III. Methods and results

This study utilizes computer vision technology to obtain real-time image streams from fishing ports through the RTSP. Subsequently, each frame of the image is parsed using OpenCV, and fishing boat detection is performed using YOLO. We then employ the BotSort algorithm for sorting and tracking of the detected fishing boats. Finally, the system determines whether a fishing boat is entering or leaving the port based on the intersection of tracking lines and electronic fence, as well as the spatial orientation distribution of the line starting and ending points. Relevant information is recorded in both text and images for future reference. The system is developed using Python, and its architecture is illustrated in Figure 1.

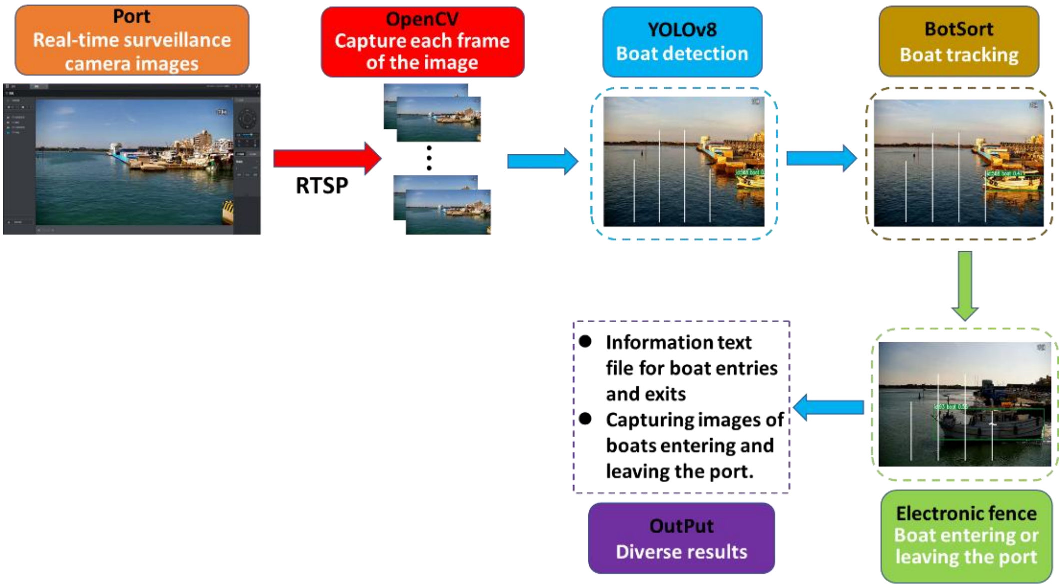


Figure 1. The proposed framework

3.1 capturing real-time images

To capture real-time images, the system takes advantage of the RTSP, which is specifically designed for real-time media streaming, enabling the transmission of audio and video over the network. This protocol is particularly well-suited for monitoring cameras and other real-time data sources. Additionally, OpenCV, a powerful computer vision library, is employed to facilitate a wide range of image processing and computer vision tasks. Therefore, this system utilizes RTSP and OpenCV for camera image streaming and frame capture. Subsequently, each obtained frame is sequentially passed into the steps of fishing boat detection and tracking in the order of the time series.

3.2 Fishing Boat Detection and Tracking

In this stage, the YOLOv8 architecture is used for fishing boat detection. Since there is currently insufficient image data of fishing boats in Penghu for model training, a pre-trained model (weights) from the COCO (Common Objects in Context) dataset is used for fishing boat detection. YOLO first detects whether there are fishing boat-related objects in each frame. If detected, it stores the fishing boat's rectangular bounding box, a class label, and the object's confidence score in a variable. The BotSort algorithm is then employed to track the detected fishing boats across consecutive frames, ensuring the continuity and distinction of objects between different frames. For each initially detected fishing boat in each frame, object tracking is established using the bounding box, assigning a unique identification code (ID) to each fishing boat. In subsequent frames, the motion trajectories of each fishing boat are tracked and their position information is continuously updated.

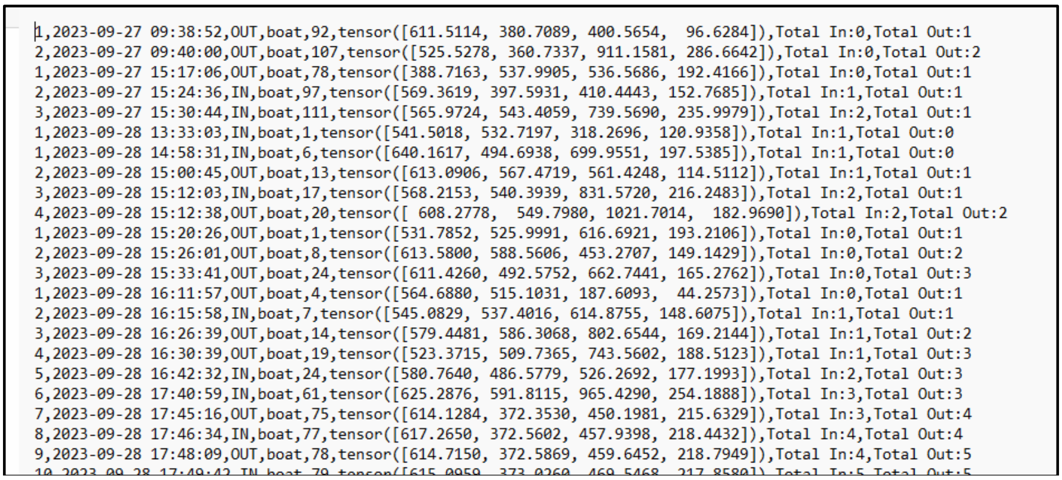
3.3 Boat Entry and Exit Determination

The determination of fishing boat entry and exit from the harbor involves setting electronic fence on the entry and exit routes (vertical white line in Figure 1) to ascertain the direction of fishing boat travel. If a fishing boat is traveling from left to right in the image, it is considered to be entering the harbor; if it is traveling from right to left, it is considered to be exiting the harbor. To enhance the accuracy of this determination and reduce the likelihood of missing cases, multiple electronic fence are set. When the line connecting the center point of the tracked bounding box intersects with any of the electronic fence, it is considered as crossing the fence. Subsequently, the spatial geometric relationship between the line connecting the center point of the bounding box and electronic fence is used to determine the direction of the fishing boat. Finally, the information of fishing boats entering or exiting at the moment, along with their images, is recorded (Figure 2).

Figure 2. Fishing boatentry and recognitionresults

3.4 Output of Results

To facilitate subsequent management and inspection by personnel, when the system identifies fishing boats entering or exiting the harbor, it immediately records the entry/exit serial number, time, direction (entry or exit), detection category serial number, tracking box center coordinates, length, width, total inbound, and total outbound quantities, among other information, which is accumulated and recorded in a TXT file. Simultaneously, the image of the identified fishing boat is captured, as shown in Figure 3 and Figure 4.



```
1,2023-09-27 09:38:52,OUT,boat,92,tensor([611.5114, 380.7089, 400.5654,  96.6284]),Total In:0,Total Out:1
2,2023-09-27 09:40:00,OUT,boat,107,tensor([525.5278, 360.7337, 911.1581, 286.6642]),Total In:0,Total Out:2
1,2023-09-27 15:17:06,OUT,boat,78,tensor([388.7163, 537.9905, 536.5686, 192.4166]),Total In:0,Total Out:1
2,2023-09-27 15:24:36,IN,boat,97,tensor([569.3619, 397.5931, 410.4443, 152.7685]),Total In:1,Total Out:1
3,2023-09-27 15:30:44,IN,boat,111,tensor([565.9724, 543.4059, 739.5690, 235.9979]),Total In:2,Total Out:1
1,2023-09-28 13:33:03,IN,boat,1,tensor([541.5018, 532.7197, 318.2696, 120.9358]),Total In:1,Total Out:0
1,2023-09-28 14:58:31,IN,boat,6,tensor([640.1617, 494.6938, 699.9551, 197.5385]),Total In:1,Total Out:0
2,2023-09-28 15:00:45,OUT,boat,13,tensor([613.0906, 567.4719, 561.4248, 114.5112]),Total In:1,Total Out:1
3,2023-09-28 15:12:03,IN,boat,17,tensor([568.2153, 540.3939, 831.5720, 216.2483]),Total In:2,Total Out:1
4,2023-09-28 15:12:38,OUT,boat,20,tensor([ 608.2778,  549.7980, 1021.7014,  182.9690]),Total In:2,Total Out:2
1,2023-09-28 15:20:26,OUT,boat,1,tensor([531.7852, 525.9991, 616.6921, 193.2106]),Total In:0,Total Out:1
2,2023-09-28 15:26:01,OUT,boat,8,tensor([613.5800, 588.5606, 453.2707, 149.1429]),Total In:0,Total Out:2
3,2023-09-28 15:33:41,OUT,boat,24,tensor([611.4260, 492.5752, 662.7441, 165.2762]),Total In:0,Total Out:3
1,2023-09-28 16:11:57,OUT,boat,4,tensor([564.6880, 515.1031, 187.6093,  44.2573]),Total In:0,Total Out:1
2,2023-09-28 16:15:58,IN,boat,7,tensor([545.0829, 537.4016, 614.8755, 148.6075]),Total In:1,Total Out:1
3,2023-09-28 16:26:39,OUT,boat,14,tensor([579.4481, 586.3068, 802.6544, 169.2144]),Total In:1,Total Out:2
4,2023-09-28 16:30:39,OUT,boat,19,tensor([523.3715, 509.7365, 743.5602, 188.5123]),Total In:1,Total Out:3
5,2023-09-28 16:42:32,IN,boat,24,tensor([580.7640, 486.5779, 526.2692, 177.1993]),Total In:2,Total Out:3
6,2023-09-28 17:40:59,IN,boat,61,tensor([625.2876, 591.8115, 965.4290, 254.1888]),Total In:3,Total Out:3
7,2023-09-28 17:45:16,OUT,boat,75,tensor([614.1284, 372.3530, 450.1981, 215.6329]),Total In:3,Total Out:4
8,2023-09-28 17:46:34,IN,boat,77,tensor([617.2650, 372.5602, 457.9398, 218.4432]),Total In:4,Total Out:4
9,2023-09-28 17:48:09,OUT,boat,78,tensor([614.7150, 372.5869, 459.6452, 218.7949]),Total In:4,Total Out:5
```

Figure 3 Output Text File Contents

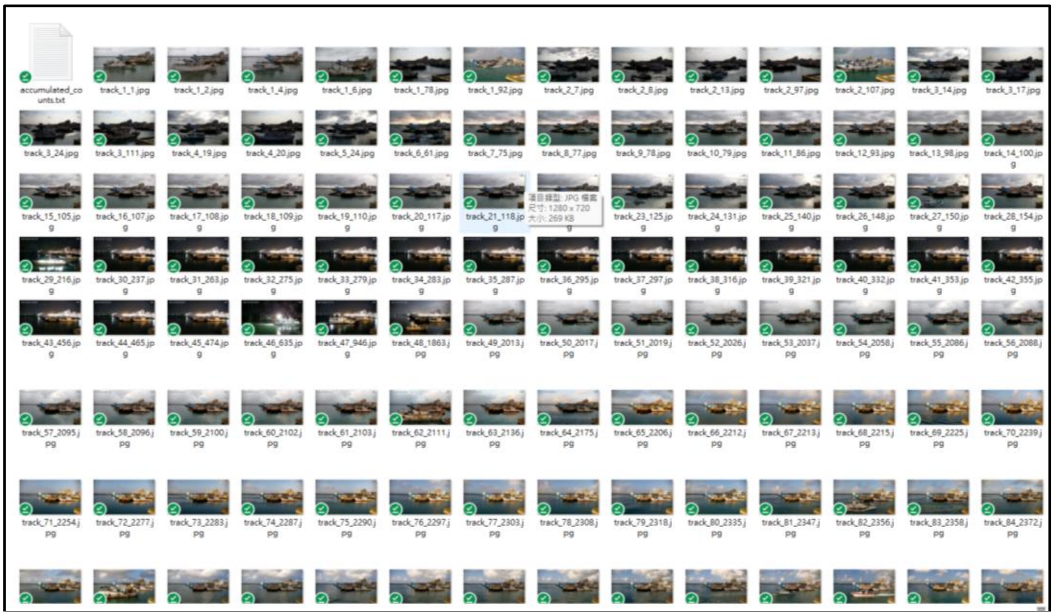


Figure 4 Output Text File and Screenshot Images

IV. Discussion and Conclusion

Automated image recognition technology for fishing boat entry and exit offers multiple benefits compared to traditional manual methods. Below, these benefits are elaborated to demonstrate the superiority of the technology:

1. Improved Management Efficiency: The system can continuously monitor 24/7, significantly enhancing management efficiency, compared to traditional methods reliant on manual patrols and monitoring. This technology can process a large volume of data in real-time, enabling the simultaneous identification of multiple fishing boats, which would be limited by the speed and efficiency of human operators in manual methods.
2. Enhanced Port Security: Traditional manual monitoring may be subject to subjective judgments, leading to the possibility of missing or failing to detect illegal entries. Automated image recognition technology can identify all fishing boats entering or exiting the harbor impartially and round the clock, enhancing port security. It can be designed to issue alarms automatically, assisting port management authorities in quickly responding to potential security issues.
3. Reduction in Human Errors or Interference: Manual methods are susceptible to errors influenced by the subjectivity and mistakes of operators. Automated image recognition eliminates these problems, improving the accuracy of identification. The technology can work continuously without being affected

by emotions, fatigue, or distractions and will not overlook critical information due to human errors or interference.

4. Time and Cost Savings: Automated identification technology saves a significant amount of time and labor costs by reducing the need for manual operations, thus lowering management expenses. Port management authorities can execute tasks more efficiently without the prolonged need for manual monitoring.
5. Real-Time Monitoring and Response: Real-time monitoring allows immediate identification of issues and alert generation. This helps take rapid actions before problems escalate, reducing potential risks and losses. Traditional methods may take longer to detect issues and respond.
6. Data Recording and Analysis: Automated image recognition technology generates detailed data records, including fishing boat identification information, activity logs, and other relevant data. This data is invaluable for subsequent analysis and decision-making. Port management authorities can optimize operations through data analysis and formulate more effective policies, thereby achieving better management and sustainability initiatives.

The automated image recognition technology for fishing boat entry and exit provides significant benefits compared to traditional manual methods in various aspects, including improved efficiency, enhanced security, support for environmental conservation, error reduction, time and cost savings, real-time monitoring, data recording, and sustainable management. This technology showcases the potential of smart technology in fisheries management, bringing smarter and more sustainable development to port and fisheries management.

The program development has been completed in a Python environment and has undergone initial testing. The image resolution is set at 1280x720, with a frame rate of 30 FPS. It is operable on a standard desktop computer (processor Intel(R) Core(TM) i5-10500 CPU @ 3.10GHz, 3096 Mhz, 6 cores, 12 logical processors; graphics card Intel(R) UHD Graphics 630; physical memory (RAM) 16.0 GB). Subsequent testing will be conducted on a higher-end computer to evaluate real-time detection, tracking, and counting objectives.

References

[1]. P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, “Overfeat: Integrated recognition, localization and detection using convolutional networks,” arXiv preprint arXiv: 1312.6229, 2013.

[2]. R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation." pp. 580-587.

[3]. R. Girshick, "Fast r-cnn." pp. 1440-1448.

[4]. S. Ren, K. He, R. Girshick, and J. Sun, “Faster r-cnn: Towards real-time object detection with region proposal networks,” Advances in neural information processing systems, vol. 28, 2015.

[5]. K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn." pp. 2961-2969.

[6]. T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection." pp. 2117-2125.

[7]. W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector." pp. 21-37.

[8]. C.-Y. Fu, W. Liu, A. Ranga, A. Tyagi, and A. C. Berg, “Dssd: Deconvolutional single shot detector,” arXiv preprint arXiv: 1701.06659, 2017.

[9]. T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection." pp. 2980-2988.

[10]. S. Zhang, L. Wen, Z. Lei, and S. Z. Li, “RefineDet++: Single-shot refinement neural network for object detection,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 31, no. 2, pp. 674-687, 2020.

[11]. Q. Zhao, T. Sheng, Y. Wang, Z. Tang, Y. Chen, L. Cai, and H. Ling, "M2det: A single-shot object detector based on multi-level feature pyramid network." pp. 9259-9266.

[12]. X. Cong, S. Li, F. Chen, C. Liu, and Y. Meng, “A Review of YOLO Object Detection Algorithms based on Deep Learning,” Frontiers in Computing and Intelligent Systems, vol. 4, no. 2, pp. 17-20, 2023.

[13]. J. Terven, and D. Cordova-Esparza, “A comprehensive review of YOLO: From YOLOv1 and beyond. arXiv 2023,” arXiv preprint arXiv:2304.00501.

[14]. M. Hussain, “YOLO-v1 to YOLO-v8, the Rise of YOLO and Its Complementary Nature toward Digital Manufacturing and Industrial Defect Detection,” Machines, vol. 11, no. 7, pp. 677, 2023.

