Safe distance reminder system on ship against port for the standing process using image processing

A W B Santosa¹, A Hardianti², S Hasugian⁴, I Sutrisno⁴ and A Khumaidi⁵
¹Universitas Diponegoro, Indonesia
²⁴⁵Politeknik Perkapalan Negeri Surabaya, Indonesia
³Politeknik Pelayaran Surabaya, Indonesia

E-mail: Arikapal75@gmail.com

Abstract. Since ancient times, Indonesia has been known as a country that conducts trading activities. However, the number of accidents when the ship is about to dock is an indication of the need to improve the marine transportation system. The factors that cause accidents include human, technical and weather errors. Seeing these conditions and also related to the increasingly rapid development of technology, modern solutions for detecting objects using cameras can be developed. By using a PC as a video processor from a camera that works in real time and can classify several ship objects using the Convolutional Neural Network (CNN) method. As well as being able to estimate the distance of the object where the ship must turn off the engine before docked with the camera using Stereo Vision. And the results of the algorithm processing can be translated into a signal that will be sent to the harbormaster and also the ship that will dock as an output. Then the ship that will dock can find out the distance where the ship has to turn off the engine through the signal sent or siren. The average total accuracy of the system which is able to detect ships is 97.15%.

1. Introduction
Indonesia is the biggest archipelago country in the world. Since ancient times, Indonesia has been known as a country that has carried out many trade activities with water routes. Along with the times, now the world of world shipping is experiencing rapid development. Trade using sea transportation services is an option because the costs are cheaper or lower than other transportation services. The Port of Tanjung Perak is one of the gateways for sea transportation for eastern Indonesia, acting as a center for collectors and distributors of goods from Eastern Indonesia, particularly from East Java Province.

One of the facilities at the terminal is a pier, where the dock is a place to dock and unload ships. The jetty at the terminals at Tanjung Perak Port in Surabaya has different sizes. If the dock available at a terminal is longer than the other terminal, then the terminal can accommodate more ships than other terminals. The depth must not be less than 1.15 times the maximum draft of the largest vessel and must not be more than 100 m (Graillot, A., 1983). But the size of the ship also affects the number of ships that will moor at the dock. If ships with large sizes moor at the terminal, the terminal can only accommodate a small number of ships. So there are two important things that become the benchmark for ships that will moor, namely the length of the pier and the size of the ship. If the ship to be moored does not have a dock that matches the size of the ship, then the ship must wait. This will result in queues for ships afterwards and can also be the cause of ships colliding.
With the conditions and technological developments as above, the authors have the idea to create a tool to detect objects that will enter the dock using video processing which is supported using Stereo Vision. This prototype uses a camera to detect moving objects and monitor the boundaries of the area. Personal Computer (PC) is used to process images captured by cameras that can detect and classify a ship or moving object using the Convolutional Neural Network (CNN) method. So that the prototype can maximize the performance of the harbormaster in order to facilitate the work of the process of getting in and out of the ship at the dock can be optimal and minimize the occurrence of accidents.

2. Methodology
2.1. System Design
Convolutional Neural network is an architectural extension of feed forward artificial neural network with a larger number of layers. The CNN architecture is determined by the number of layers, the number of maps in each layer, the size of the kernel, the size of the subsampling windows, and how they are connected to each other. In Figure 2.7. You can see how the Convolutional Neural network architecture works in recognizing handwriting.

![Figure 1. Arsitektur LeNet5](image1)

2.1.1. System Design.
Dataset is an object that represents data and relations that have been created in storage. A dataset can be called a database, because it has the same structure. The dataset in this study is taken from several examples of several images that have been stored in the laptop storage space. And the dataset is used for object classification. the more input the dataset will make the higher the classification results and the heavier the laptop runs.

2.1.2. CNN Architecture.

This final project is focused on getting an object detection system with a camera. In addition, the main objective of this study is to obtain a classification result from the CNN process. In the CNN process in general, this process has 3 stages, namely pre-processing, processing, and classifying.

2.1.3. Euclidean Distance Method.

Euclidean Distance method is a method related to distance calculation. This concept was introduced by a mathematician named Euclid of Alexandria around 300 B.C.E (Wikipedia, n.d.-c). The principle of the Euclidean method is to find the proximity of the distance value between two variables. The following is the formula for measuring the distance of Euclidean Distance.

![Dataset](image)
\[ s = \sqrt{(Lat_1 - Lat_2)^2 + (Long_1 - Long_2)^2} \]

The formula above is a formula used to find distance values based on two coordinate points based on sources (Adiwilaga, 2014). Based on this formula, the data from the known ship distance values will then be compared to get the closest value to the ship.

3. Result and Discussion

3.1. GPS Testing Result

GPS testing, using ublox neo M8N GPS and Mobile GPS to find out the error value or difference obtained. Here are the points taken from the neo M8N GPS and Mobile GPS. The data used in this research is data obtained directly from the GPS device. The microcontroller will receive latitude and longitude data in the form of a decimal degree. From the latitude and longitude data will be processed into the length of the distance in kilometers. Where the authors here process the data using the Euclidean Distance Method.

3.1.1. Euclidean Method

\[ s = \sqrt{(x_2-x_1)^2 + (y_2-y_1)^2}) \times r \]

Where :
- \( s \) = distance (km)
- \( y_1 \) = longitude 1
- \( x_2 \) = latitude 2
- \( y_2 \) = longitude 2
- \( x_1 \) = latitude 1
- \( r \) = 1° earth = 111,319 km

Table 1. Neo M8N GPS test points

| Keterangan | Latitude  | Longitude  | Hasil Pengujian Jarak Metode Ecluidean |
|------------|-----------|------------|---------------------------------------|
| Titik 1    | -7.788983 | 112.009176 | 1.168,84 Cm                           |
| Titik 2    | -7.789044 | 112.009132 |                                       |

Table 2. Test points for mobile GPS

| Keterangan | Latitude  | Longitude  | Hasil Pengujian Jarak Metode Ecluidean |
|------------|-----------|------------|---------------------------------------|
| Titik 1    | -7.788950 | 112.009119 | 510,13 Cm                             |
| Titik 2    | -7.788964 | 112.009151 |                                       |

From the test table above, the error difference between GPS Neo M8N and GPS Hanphone is obtained, namely:

\[ = 1168.84 \text{ Cm} - 510.13 \text{ Cm} \]

\[ = 698.72 \text{ cm} \]
3.1.2. MobileNet Extracion Features

MobileNet SSD consists of 13 layer blocks which contain a depthwise separable convolution layer.

![Figure 4. Layer blocks which contain a depthwise separable convolution layer](image)

The image input becomes the first layer, then the depthwise convolution layer will process the image based on the RGB color channel (Red, Green, Blue). Figures below show the map feature of the activation function (RELU6) of the first three layers of the MobileNet architecture used for Ship detection. The map feature is a feature that has been successfully extracted from each layer. From this figure, there are 64 features extracted at each activation layer according to the output size on that layer, which is 64, which means 8x8.

![Figure 5. Layer 1 Extraction Feature Function](image)

![Figure 6. Layer 2 Extraction Feature Functions](image)
3.1.3. Evaluation Training

The training process is a process in which the processor learns the features of the input data so that the resulting system can recognize the specified object. In this study, the training process was carried out up to 40,000 steps and the network used was Mobilenet SSD. To be more optimal, the training process using a PC with an NVIDIA GeForce GTX 950 M GPU and the total time spent completing 40,000 steps is 8 hours.

Figure 7. Layer 3 Extraction Feature Functions

Figure 8. Chart of training loss against iteration
As previously explained, in this architecture, the author uses a multistep learning rate type with a step in the 20,000 and 40,000 iterations. In the 0-20,000 iteration, the initial learning rate used is 0.0005. Then in the iteration of 20,000 - 40,000 the learning rate value changes to 0.00025 because it is multiplied by the value $\gamma$.

![Train learning rate vs. Iters](image)

**Figure 11.** Changes in learning rate value against iteration

The initial learning rate value cannot be too small, because it will make the learning model not comprehensive in the initial phases. It is better if the learning rate value also decreases with the increase in iterations. Figure 11. shows the change in the value of the learning rate with respect to iterations. By using a configuration matrix, precision, recall, and accuracy values can be obtained.

![Precision x Recall curve](image)

**Figure 12.** Precision-Recall graph of model validation
In this project, the ship detection model is evaluated to obtain precision and recall values. In finding the precision and recall values, there are 4 parameters that must be fulfilled, namely true positive, true negative, false positive, and false negative. In this case, true positive is the prediction indicating Ship and groundtruth also indicating Ship. True negative does not predict Ship and groundtruth also does not indicate Ship. False positives are Ship's predictions but groundtruth doesn't show Ship. And false negative is not predicting the ship, but the groundtruth shows the ship.

The precision and recall values are calculated for all test data, which is about 530 images. From 530 images, it will produce mAP (Mean Average Precision), which is the average precision value of all test data. The mAP value on this ship detection model is 92.15%.

3.1.4. Vessel Detection Test Results
Ship detection testing is carried out to assess the system's ability to recognize objects in the form of ships. In this test, the confidence value will be obtained, which is the percentage of the model's prediction on the object of the ship.

| Input Gambar | Hasil Deteksi | Confidence |
|--------------|---------------|------------|
| ![Image](image1.png) | ![Image](image2.png) | 99.42% |
| ![Image](image3.png) | ![Image](image4.png) | 99.86% |
| ![Image](image5.png) | ![Image](image6.png) | 100% |

From the table above, it can be seen that the system's ability to detect ships is quite good, with an average confidence of above 90%.

3.1.5. Ships Detection Test Results from Various Distances
The second test is to vary the distance between the object and the camera when taking pictures. The variations in the distance used are 1 meter, 2 meters, 3 meters, 4 meters and 5 meters. This is done to
assess that the system can detect ships from a maximum distance of 5 meters. The test results of ship detection against distance can be seen in table 4 below.

**Table 4. Ships detection test results from various distances**

| Input | Distance | Confidence | Evidence |
|-------|----------|------------|----------|
| ![Image](image1.png) | 1 meter | 99.42% | Detected |
| ![Image](image2.png) | 2 meter | 100% | Detected |
| ![Image](image3.png) | 3 meter | 99.86% | Detected |
| ![Image](image4.png) | 4 meter | 51.29% | Detected |
| ![Image](image5.png) | 5 meter | 25.13% | Detected |

From the table we can see that the system’s optimal ability to detect ships is at a distance of 1 to 3 meters. This is evidenced by the confidence value that reaches more than 90%. At a distance of 2 meters, the system’s ability to detect ships began to decline as evidenced by the decrease in the system’s percentage of detecting ships. This can be due to the lack of variation in training data from various distances.

3.1.6. **Testing On Real Time Video**

After testing the system integration from the input in the form of an image. The next test is to calculate the fps (frames per second) from the input in the form of real time video. The fps test results from real time video can be seen in table 5.

From the table 5, the test results are obtained through real time video. From a total of 5 tests in the table, the system can recognize 3 ships with an accuracy above 90% and 2 ships with an accuracy be-
low 50%. The focus of this test is to prove the system can work in real time as shown by fps. The system can work in real time with a minimum fps value of 4.77 and the highest fps value is 11.90.

**Table 5. Testing Results on Real Time Video**

| Input | Confidence | FPS | Evidence |
|-------|------------|-----|----------|
| ![Image](image1.png) | 96.51% | 4.77 | Detected |
| ![Image](image2.png) | 100% | 6.53 | Detected |
| ![Image](image3.png) | 91.37% | 11.90 | Detected |
| ![Image](image4.png) | 49.5% | 11.47 | Detected |
| ![Image](image5.png) | 46.89% | 11.88 | Detected |

From all test data starting from ship detection, the overall accuracy value of the system is 78.9% and the system speed is 9.31 fps.

4. Conclusion
1. Ship recognition systems can be built with Mobilenet SSDs for detection on Ships. This method was chosen because it can recognize objects well, and is light and fast computational.
2. The results of the matrix evaluation for the ship detection model in the form of mAP (Mean Average Precision) of 92.15% were tested on 530 images.
3. The overall evaluation result of the ship recognition system is 78.9%, where the system is able to recognize ships with a level of truth of 78 per 100 characters.
4. The system can be implemented in real time and the computation rate in one frame is 9.31 frames per second.
5. The optimal distance between the camera and the object is under 5 meters.

5. References

[1] Khumaidi A 2017 Welding defect classification based on convolutional neural network (CNN) and Gaussian Kernel. International Seminar on Intelligent Technology and Its Application. pp 261-265
[2] Deshpande A 2016 A Beginner’s guide to understanding convolutional neural networks Part 2. Available at: https://adeshpande3.github.io/A-Beginner’s-Guide-To-Understanding-Convolutional-Neural-Networks-Part-2/
[3] Khumaidi A 2015 Implementasi pengolahan video dengan metode color threshold pada prototype kapal pendeteksi korban kecelakaan laut berbasis android. (Surabaya: Politeknik Perkapalan Negeri Surabaya)
[4] Lutfi M N 2018 Pendeteksii objek api dan boneka pada robot pemadam api berkaki menggunakan video processing berbasis mini PC. (Surabaya: Politeknik Perkapalan Negeri Surabaya)
[5] Marzuqi I, Arinata G P, Putra Z and Khumaidi A 2017 Segmentasi dan estimasi jarak bola dengan robot menggunakan stereo vision. 5th Indonesian Symposium on Robotic Systems and Control. pp 140-144
[6] Salsabila 2018 Penerapan deep learning menggunakan convolutional neural network untuk klasifikasi citra wayang Pukakawan. (Yogyakarta: Universitas Islam Indonesia)
[7] R Girshick J, Donahue T Darrell and J. Malik 2016 Region-based convolutional networks for accurate object detection and segmentation. IEEE Trans Pattern Anal. Mach. Intell
[8] Z Q Zhao, P Zheng, S T Xu and X Wu 2019 Object detection with deep learning: A review. IEEE Transactions on Neural Networks and Learning Systems
[9] Jami’in M A, Hu J, Marhaban M H, Sutrisno I and Mariun N B 2016 Quasi ARX neural network based adaptive predictive control for nonlinear systems. IEEJ Trans Elec Electron Eng. 11 pp 83-90
[10] Jami’in M A, Sutrisno I and Hu J 2012 Lyapunov learning algorithm for Quasi-ARX neural network to identification of nonlinear dynamical system. IEEE International Conference on Systems, Man, and Cybernetics (SMC). (Seoul). pp 3147-3152. doi: 10.1109/ICSMC.2012.6378275
[11] Jami’in M A, Sutrisno I and Hu J 2013 Deep searching for parameter estimation of the linear time invariant (LTI) system by using Quasi-ARX neural network. The 2013 International Joint Conference on Neural Networks (IJCNN). (Dallas, TX) pp 1-5. doi: 10.1109/IJCNN.2013.6707091
[12] Jami’in M A, Sutrisno I and Hu J 2015 Maximum power tracking control for a wind energy conversion system based on a quasi ARX neural network model. IEEJ Trans Elec Electron Eng. 10 pp 368-375
[13] Jami’in M A, Sutrisno I, Hu J, Mariun N B and Marhaban M H 2014 An adaptive predictive control based on a quasi-ARX neural network model. 13th International Conference on Control Automation Robotics & Vision (ICARCV). (Singapore). pp 253-258 doi: 10.1109/ICARCV.2014.7064314
[14] Sutrisno I, Jami’in M A and Hu J 2014 Modified fuzzy adaptive controller applied to nonlinear
systems modeled under quasi-ARX neural network. *Artif Life Robotics*. **19** pp 22–26 https://doi.org/10.1007/s10015-013-0137-6

[15] Sutrisno I, Jam'i'in A M and Hu J 2014 An improved elman neural network controller based on quasi ARX neural network for nonlinear systems. *IEEE Trans Elec Electron Eng*. **9** pp 494-501

[16] Sutrisno I, Che C and Hu J 2014 An improved adaptive switching control based on quasi-ARX neural network for nonlinear systems. *Artif Life Robotics*. **19** pp 347–353

[17] Sutrisno I, Jam'i'in M A and Hu J 2012 Neural predictive controller of nonlinear systems based on quasi-ARX neural network. *18th International Conference on Automation and Computing (ICAC).* (Loughborough) pp 1-6

[18] Rinanto N and Khumaidi *Aplikasi Android- Rasspberry Pi pada Kapal Tanpa Awak untuk Pencarian Korban Kecelakaan Laut*

[19] Widodo R B, Quita R M, Amrizal S, Gunawan R S, Wada C, Ardiansyah A, Sutrisno I, Rahmat M B, Arfianto A Z, Handoko C R, Santoso A D, Santosa A W B, Ardhana V Y P and Setiawan E Grasping and attached mode in human computer interaction in the study of mouse substitution. *Journal of Physics: Conference Series* **1196** 1 012034

[20] Sutrisno I, Jam'i'in M A, Hu J and Marhaban M H 2015 Self-organizing quasi-linear ARX RBFN modeling for identification and control of nonlinear systems. *54th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE).* IEEE

[21] Jam'i’in M A, Sutrisno I and Hu J 2014 Nonlinear adaptive control for wind energy conversion systems based on quasi-ARX neural networks model. *Proceedings of the International Multi Conference of Engineers and Computer Scientists.* March 12-14 2014 (Hong Kong)

[22] Sutrisno I 2009 *Pemrograman komputer dengan software matlab.* (Surabaya: ITS Press)