Journal of Physics: Conference Series

PAPER • **OPEN ACCESS**

Object recognition on patrol ship using image processing and convolutional neural network (CNN)

To cite this article: E Julianto *et al* 2020 *J. Phys.: Conf. Ser.* **1450** 012081

Object recognition on patrol ship using image processing and convolutional neural network (CNN)

E Julianto[1], A Khumaidi[2], P Priyonggo[1], M B Rahmat[2], S Sarena[2], R Adhitya[2], B Herijono[1], G Suharjito[1], I Munadhif[2]

[1] Department of Marine Engineering, Politeknik Perkapalan Negeri Surabaya, Jalan Teknik Kimia, Surabaya, Indonesia
[2] Department of Marine Electrical Engineering, Politeknik Perkapalan Negeri Surabaya, Jalan Teknik Kimia, Surabaya, Indonesia

E-mail: aguskhumaidi@ppns.ac.id

Abstract. Autonomous patrol vessels are state border patrol vessels equipped with cameras and image processing capabilities to detect objects around them. This prototype of ship can recognize a detected object; it used an image classification method called Convolutional Neural Network (CNN). So, it will minimize the occurrence of accidents on patrol boats. Input image in the form of RGB will experience feature extraction using a convolution layer. In the classification layer there is an artificial neural network with backpropagation to classify objects against a predetermined dataset. The detection value of the obtained vessel is operated by a predetermined actuator. In the final classification results the object recognition in the form of ships have a quite high accuracy. The average accuracy value is 96.59 percent with a sufficient light condition and RGB image input is taken in real-time.

1. Introduction

In 1982, the UN (United Nations) Law of the Sea Convention provided a legal basis for island states to determine the country's borders to the Economic Exclusive Zone and the Continental Shelf. With this basis a country has the authority to exploit natural resources in the zone. Various natural resources such as fisheries, natural gas, petroleum, and other mining materials can be utilized by the country concerned [1]. With this reinforced, every country has a security body to protect the country's borders. In this century the development of object detection was greatly increased. From one of the users of these technologies for tools that can help human work begin to be created. With this in the case of maritime border problems, patrol boats have a minimal number and there are still ship accidents. This paper is using an image processing algorithm to detect objects and learning algorithm using the convolutional neural network method to recognize the objects. For the recognition process, it is necessary to create a dataset according to the object that we want to recognize. The dataset is collected a lot; the more increasing number of dataset content, the level of accuracy will be higher.

Convolutional Neural Networks (CNN) that have been created will be combined with segmentation [2]. The results of the classification are bounding boxes or other terms and detected objects will be selected with a box. In addition, there are names of detection results and the value of the success rate of the detection. Finally, in one frame several objects can be detected [3-6]. The used architecture on CNN is feature extraction and classification stages. At the feature extraction stage, convolution occurs several times with different filters. the classification method in this study is using back propagation neural network. In this paper, an automated ship system for state border patrols was created to be able to detect objects in the form of ships. The detection process uses a camera with image processing and applied convolutional neural network method. The captured image will pass the extraction of the image feature to reduce the size of the input frame. Therefore the classification stage can be processed easily. Segmentation occurs after finding these results. The result will be a bounding-box frame that detects the object [7]. Finally, the value of (x, y) on the frame will be coordinated with the ship's actuators.

After the system above was trained to detect ships. The produced results of the ship detection experiments with RGB input are quite precise. The average accuracy value is 96.59% in sufficient light conditions. The proposed detection system can be implemented on the ship. To get the appropriate movement on the actuator a position trial will be carried out. To be clearer in this paper the structure is as follows. The structure of the system proposed in chapter 2 is methodology. And for further discussion on how the results of this study are discussed in chapter 3 is the result and data analysis. In the end the final results summarized in chapter 4 are the conclusion.

2. Methodology

The purpose of this study is to create an automated patrol system that can detect an object with image processing with the convolutional neural network method. The detection results can move the ship's actuator. So, the ship can move automatically to avoid the object. An explanation of how the system works can be seen in the form of a flowchart in Figure 1.

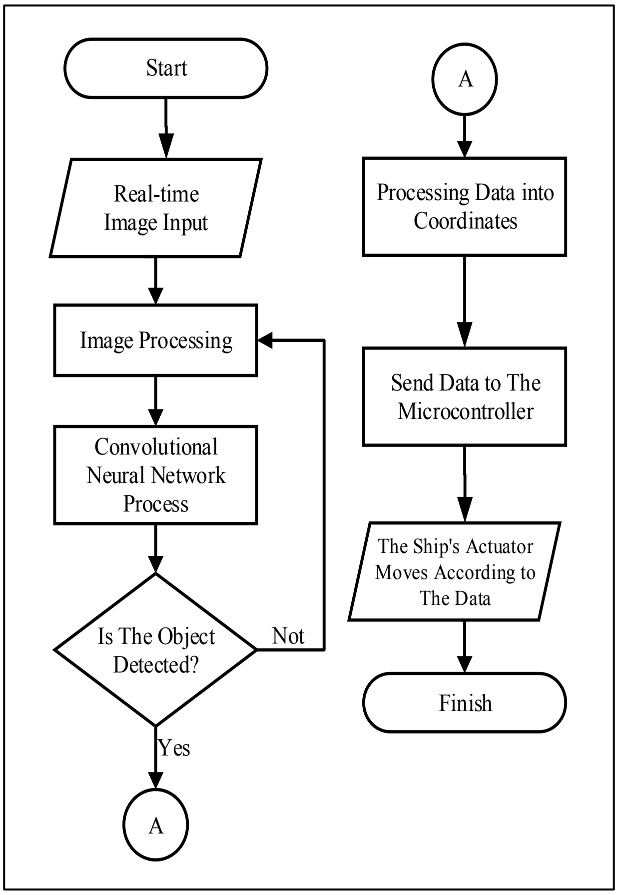

Figure 1. Flowchart of the system works.

From the flowchart structure in Figure 1, the detection system starts when the camera lights up, then input data in real-time starts to be made. This data is processed using image processing. From the results of image processing, it will continue to the CNN method algorithm. If there is an object in the input, therefore the value will be processed into coordinate data of pixel size. If there is no object, there will be no value to be processed into coordinate data. The coordinate data was sent to the microcontroller to be translated into an algorithm that can move the ship's actuator.

2.1. Convolutional neural network architecture
In Figure 2, CNN has 2 parts in general as follows: the first part is called feature extraction and the second part is called classification (Fully Connected). The feature extraction section contains processes for image convolution and max pooling. Which aims to reduce the size of the pixel image and sharpen the image. so that in the process of training content in the classification section, can run quickly and lightly.

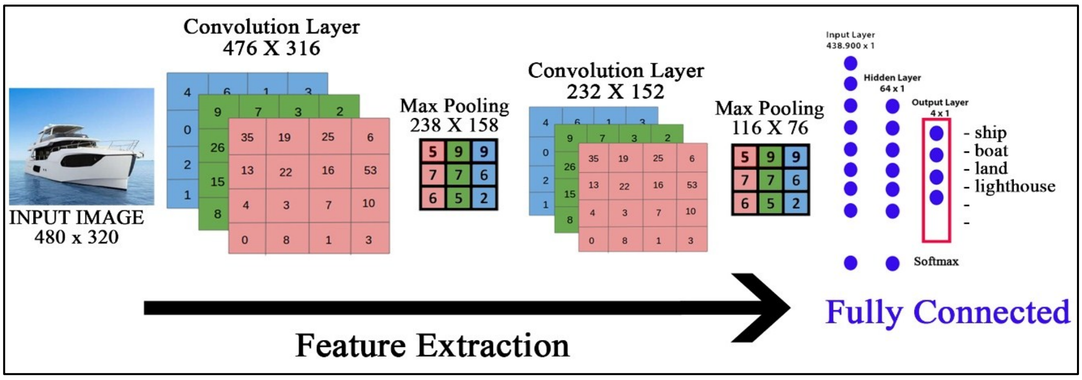

Figure 2. CNN architecture.

2.2. Convolution layer
The convolution layer in this study was conducted twice. With an input size of 480 × 320 after using the shrinking pixel feature to 116 × 76. This convolution layer uses the RELU activating function (rectified linear unit). In addition to set the activation function, this layer also has functions that affect it such as three parameters, depth, steps and zero padding settings [8]. The kernel in the convolution uses a size of 7 × 7.

2.3. Max pooling
This layer is the commonly used type for collection methods. At this layer there is a very large pixel reduction, up to half of the input for max pooling. But in this layer the number of output feature maps does not change [9]. In this study using a kernel sampling size of 3 × 3. Because in this study using 2 times convolution. Max pooling is applied in every convolution process. So, max pooling in this study also occurs 2 times.

2.4. Fully connected section (classification)
This section is an artificial neural network that that consists of neurons. This part of the system has a deep learning function. This system is usually called the Neural Network (NN). In NN there are various types of calculation methods. In this study is using the NN method which commonly used in many classification problems called back propagation. This method plays with the value of the system error; it can minimize the high error value. The basis of the backpropagation algorithm is gradient descent and the back propagation in this study uses an activation function called sigmoid [10-12]. The back propagation method is very effective for determining the weight of multilayer neural networks. The size of the output Mean Square Error (MSE) is obtained from the output (o) and target (t) difference, is described in equation below:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i}^{n}(o_i - t_i)^2 \quad (1)$$

t_i is a set of training weights for neural networks that consists of desired vector pairs. For training vector data, the values are consisting of 1 and -1.

2.5. Coordinates image

After successfully detecting objects with CNN. The ship should be able to avoid the object. Pixel information from the image can inform the coordinate of the object. The pixel values are in the coordinate format (x, y).

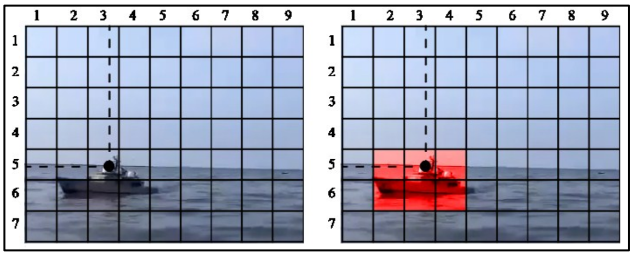

Figure 3. An object from extracted pixel.

With a pixel size of 9 × 7, as in Figure 3. The detected object is red. The object in pixel x is in the values of 5 and 6. For the value of y, the parameters are 2 to 4. Because the object is already known, the algorithm instructs the actuator to move towards the most pixels without objects. In Figure 3, the object on the left and right side has the most pixels with no object, and then moves it towards.

2.6. Hardware

To run the system in this study. requires hardware in the form of a ship prototype. The ship with the catamaran hull model was chosen because it has high stability [13]. With a system that resembles an original ship, this prototype also has a rudder and propeller as a driver. To move the actuator requires a microcontroller system that is connected to a PC. As in Figure 4, it shows how the prototype has been made. There is a camera on the front of the ship as the main sensor to provide image input.

Figure 4. Picture of a prototype of a border patrol boat.

3. Results and data analysis

In this chapter presented how the results after the system was made. The conducted trials were taking a place in the sea. There are also test results regarding the response of image processing parameters as shown in Figure 5.

Figure 5. The results of detecting ships with short distance.

The detection result has a good value if the object is not too far away and if the object is very far the value will be so small. It can be seen in Figure 5, the value of the object's similarity is quite high up to 90%. It is different with Figure 6, the distance of the ship is quite far and the similarity value of the object becomes small with a value of 58%.

Figure 6. The results of detecting ships with long distances.

The test is taken with a safe distance for detection. Which means the ship can still manoeuvre before approaching the object. Besides the ship there are still other objects that can be detected.

Figure 7. (a) The result of the image was detected by 2 boats, (b) The result of the image was detected by 3 boats.

In addition to one object that can be detected this research can also detect many objects on one image or frame. As in Figure 7 (a), there are 2 boats detected. In Figure 7 (b), there are 3 boats that can be detected. from this, patrol boats can avoid the detected boat.

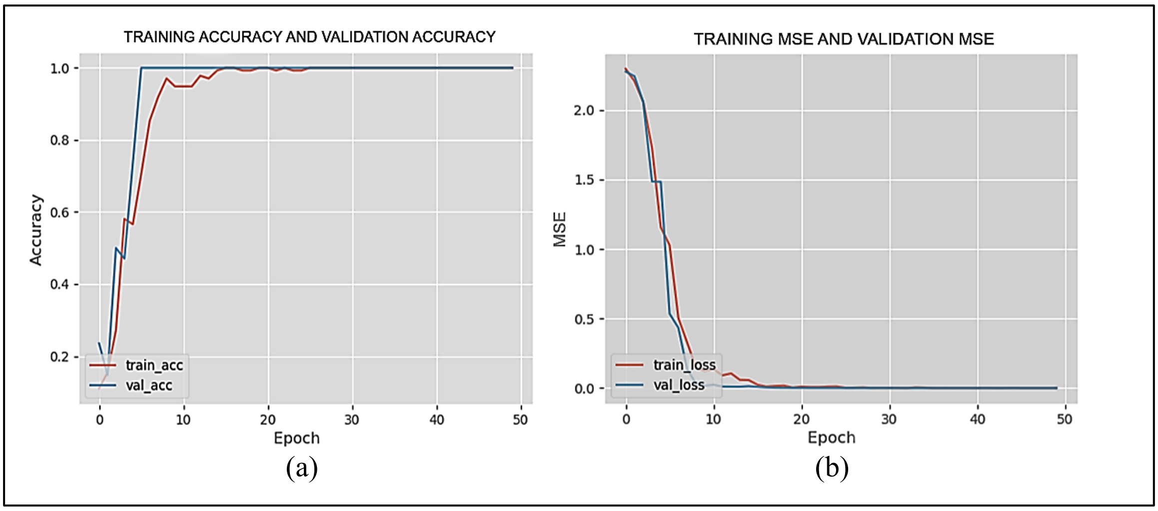

Figure 8. (a) Graphic images regarding performance accuracy, (b) Graphic image of MSE performance

In Figure 8 (a) is a graph of the comparison between the number of iterations with accuracy. The orange line graph is a training response and blue is a testing response. When the iteration was at 5 levels, the accuracy becomes optimal with a validation value of 100%. And at that level, the performance continues to stabilize. Figure 8 (b) is working performance for MSE between training and testing. The two images above have an influence on the depth of the layer. With these data after the average value is taken from different vulnerable times. The success rate for detecting object is equal to 96.59%.

4. Conclusions

In this paper the use of image processing and convolutional neural networks to detect objects on sea border patrol vessels has an average accuracy of 96.59%. These results have a small error or almost no error; the patrol boat can certainly detect the object correctly and it would not hit the object.

5. References

[1] Macnab R, Mukherjee P and Buxton R 2011 *Oceans '87*
[2] Kurniawan D A *et al.* 2018 *International Symposium on Electronics and Smart Devices (ISESD)*
[3] Khumaidi A, Yuniarno E M and Purnomo M H 2017 *International Seminar on Intelligent Technology and Its Applications (ISITIA)*
[4] Nazar A, Syai'in M, Nurwiyadi M N P and Khumaidi A 2019 *IOP Conference Series: Materials Science and Engineering* **462**
[5] Yang H, Zhou J T, Zhang Y, Gao B, Wu J and Cai J 2015 *Computer Vision and Pattern Recognition (CVPR)*
[6] Khumaidi A 2017 *Klasifikasi Image Sequence Hasil Pengelasan Menggunakan Metode Convolution Neural Network (Cnn) Untuk Non Distructive Test* (Surabaya: Institut Teknologi Sepuluh Nopember)
[7] Qian H, Zhou J and Xu J 2018 *Chinese Automation Congress (CAC)*
[8] Albawi S, Alzawi S and Mohammed T A 2017 *The International Conference on Engineering and Technology (ICET)*
[9] Yan K, Huang S, Song Y, Liu W and Fan N 2017 *Chinese Control Conference (CCC)*

[10] Mizutani H 2002 *Proceedings of IEEE International Symposium on Circuits and Systems - ISCAS '94*
[11] Putra R Y, Kautsar S, Adhitya R Y, Syai'in M, Rinanto N, Munadhif I, Sarena S T, Endrasmono J and Soeprijanto A 2017 *International Symposium on Electronics and Smart Devices (ISESD)* 153-157
[12] Mahardika W P *et al.* 2018 *International Symposium on Electronics and Smart Devices (ISESD)* 40-45
[13] Dzan W, Chang S and Hsu K 2013 *Second International Conference on Robot, Vision and Signal Processing*