Assessing Deep Learning Model Using AlexNet for Water Traffic Counting in Martapura River

Nahdi Saubari(✉) and Wang Kunfeng

College of Information Science and Technology, Program of Controlling Science and Engineering, Beijing University of Chemical Technology, Beijing, China
nahdi.vfp@gmail.com

Abstract. In recent years, the traffic of water transportation in Martapura river has been increased and creating many problems for the city and its environment. Hence, the traffic needs to be managed from time to time. Deep learning model might be used for traffic counting by detecting the ships. This study aims to assess AlexNet for traffic counting purposes in Martapura river. Data were collected two times a day for 3 months by using smartphone camera. Series of experiments were developed using Alexnet model to classify and detect ships or boats in Martapura River to draw a baseline for water traffic counting system. Result shows that Alexnet gives around 97% accurateness in detecting ships or other water vehicle as the main transportation. This certainly helps the traffic counting in Martapura river. Around 5 to 7 water vehicles were detected per hour. AlexNet also detect other floating objects like water plantation or plastic garbage. Other than object detection, AlexNet as Deep Learning technology can be used for water traffic counting globally.

Keywords: Traffic Counting, Martapura River, AlexNet

1 Introduction

In the era of computer vision, object detection has become one of the most important branches of knowledge. It can be implemented into various kind of fields, including face detection as the living object detection, aside from non-living object detection. However, during its early years, object detection mostly used to detect vehicle on the road, animals, etc. Vision based vehicle detection plays a significant role in monitoring the traffic and circulation situation, while at the same time can function as surveillance system [1]. For the past few years, deep learning technology has become a remarkable topic in computer vision. This has led to the benefits of deep learning in many areas of detection, processing and decision making[2]. There are three types of Deep learning models: the input layer (receives data), the hidden layer (extracts patterns), and the output layer (produces results). The output of one layer is used as an input into the next one. There are several architectures that can be applied in deep learning. One of the most widely used deep learning techniques is called Convolutional Neural Network (CNN)[3]–[5]. As a matter of fact, CNN as part of deep learning technology is one of the most well-known deep learning methods in the world since 2000s.

M. Setiyo et al. (eds.), *Proceedings of the 4th Borobudur International Symposium on Science and Technology 2022 (BIS-STE 2022)*, Advances in Engineering Research 225,
https://doi.org/10.2991/978-94-6463-284-2_41

Currently, deep learning technology is also applied to detect floating debris, and other floating objects like water transportation vehicle [6]–[8]. In contrast with non-water vehicle which run on land (cars, bus, etc), water vehicle detection faced many problems and the result might depend on several aspects such as the quality of the image input, condition of the weather, direction of the wind, reflections, etc. Many researches already tried to evaluate how deep learning-based vehicle detection can successfully give better result in terms of accurateness or time to detect floating objects. Most of the research took place in the sea, marine environment, lake, or urban water bodies. The method allows the usage of large data and has been successfully detect floating object in water environment.

Banjarmasin is located in South Kalimantan, Indonesia. The city has thousands of big and small rivers. The traffic of water transportation in Martapura river has been increased and creating many problems for the city and its environment. Climate change has also affected the water environment of Martapura river. At the moment, Banjarmasin is not only facing the problem of water transportation, but also has to deal with tidal flood which already rising since 2021. This situation needs to be managed with special approach, with the help of technology because in the future it might need real time detection.

Deep learning model might be used for traffic counting by detecting the ships per day. Deep learning techniques have emerged as a powerful strategy for learning feature representations directly from data, resulting in notable breakthroughs in the field of general object recognition, not to mention if it is used for real time data [9]. One of the most popular deep learning models is called AlexNet. AlexNet is an intensive CNN architecture and has proven to be very efficient in image classification tasks [10]. AlexNet is more suitable for identification and classification applications [10]–[12]. As water vehicle has more varied condition in terms of object detection, with the competence of AlexNet, it is reasonable to use this CNN architecture for detecting objects in Martapura river. The result of the training might be useful for the local government to control water traffic.

2 Method

Some previous researches to detect floating objects often use Faster R-CNN[13], [14], because it can give optimum performance in high accuracy due to its number of hidden layers. More hidden layers equal to higher accuracy, but it might be time consuming. It will also need better hardware specification to do the experiment. The other methods also can be implemented in water environment such as YOLO [15] and SSD which give better performance in speed detection. However, in this research, AlexNet was used because it only consists 8 hidden layers, and do not need high specification of hardware (core GPU).

This research was intended to assess deep learning model using AlexNet for detecting floating object in Martapura river. A conceptual understanding of AlexNet as a typical CNN architecture is firstly provided as a basic comprehension in image processing. AlexNet was preferred as the method in detecting water transportation due to its precision in object detection for so many years. Floating objects has unique characters because it is continuously moving, has reflection, and affecting by the climate and environment.

Data were collected two times a day for three months (June-August 2022) using smartphone camera in Martapura river. Series of experiments were developed using Alexnet to classify and detect boats or other floating objects in Martapura River.

3 Results

3.1 Deep Learning for Visual Recognition

Deep Learning model has broadly used for visual recognition. One of the most recognizable networks which commonly implemented for image processing is Convolutional Neural Network (CNN). This network is the type of deep neural network and gives so many advantages compared to conventional object detection. CNN can capture relevant features from an image unparallel to human brain and treat all elements of the vector input equivalently. No matter how many vector elements inputs, during the initial phase of data training, all vector elements as input will be treated equally [12]. For example, 10 inputs by 1 vector means all the 10 weights of each neuron are updated in the first layer. CNN is also a perfect feature extractor. In pre-training phase, CNN can be tuned in to extract useful attributes by feeding data on each level efficiently which spent less time and also saving memory. CNN allows to only train the classifiers at the end for labelling. Hence, in terms of identifying important features, CNN does not need any human intervention and can work by itself.

CNN also has some disadvantages. As part of deep learning model, it needs a high quality of GPU because it will allow the training process to perform faster. The more layers put into training; the more time will be needed. Compared to other neural networks, CNN might be a little bit slower due to its operation, but have higher accuracy.

CNN architecture consists of three to five main layers [12], [14], which are: convolutional layer, pooling layer, fully connected input layer, fully connected layer and fully connected output layer.

1. Convolutional layer: This layer is the spine of any CNN model, where the images will be scanned through its each pixel to create a feature map for classification.
2. Pooling layer: This layer brings the general pictures of the images by defining its dimension by limiting down necessary data from each convolutional layer. At this time, creating convolutional layers and pooling run continuously as a looping process. Therefore, it will take some times.
3. Fully connected input layer: This layer compressed the outputs into a single vector for the next upcoming layer.
4. Fully connected layer: This layer allocates weights to the input which will be used to propose a proper label after the feature analysis.
5. Fully connected Output layer: as the final layer, this layer shows the labels for classification which then appointed to the images.

Given its popularity in art neural network for maximum performance and efficiency, CNN as a strong deep learning model allows to be used in many fields like face recognition, object detection, analyzing textbook, etc. The problem with CNN is, it is also hard to be implemented for high resolution image.

3.2 AlexNet as a Typical CNN Architecture

AlexNet was first proposed by Alex Krizhevsky in 2012 and was built based on CNN architecture. Fig. 1 shows that AlexNet has 8 layers and can classify images into so many categories. In CNN, AlexNet architecture consists of five convolutional layers, and three fully connected layers. The difference with CNN is that AlexNet uses Rectified Linear Units (ReLU). ReLu is a linear function which allows the positive input to be processed as output directly. Due to its capability to train the data easier, ReLu has become the basic function for neural networks as it also often gives better performance compare to other [16].

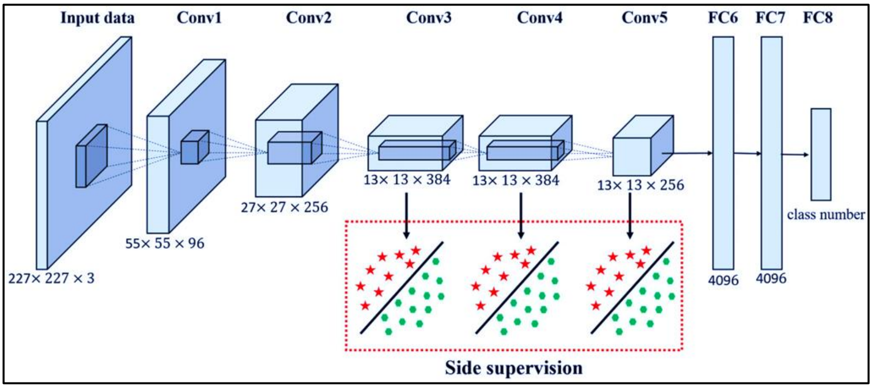

Fig. 1. Illustration of AlexNet Model

By learning the stable and unique features of the data, specific tasks and accurate classifications are performed [17], AlexNet has countless depth and durable feature capabilities. It can save a lot of model training time, and has faster prediction speed. Faster R-CNN can give more accuracy; however, this research was limited by time and quality of GPU. AlexNet which only consist eight hidden layer seems the most reasonable to apply in this research, without neglecting more recent detection model.

3.3 Experiments on Data Sets with AlexNet

In order to achieve better robustness using deep learning model, large data sets were needed. Therefore, this research took around three months to get those datasets. The images of the floating object used in the experiments were captured by videos taken with smartphone. Images were extracted from the videos, and each video duration is around 5 minutes or 300 seconds. The extracted images are in RGB then filtered to get obtained data. There were total of more than 1500 data sets were trained using AlexNet. These data were collected two times per day. First batch in the morning (9-10 a.m) and the second batch in the afternoon (2-3 p.m) local time.

Before collecting data, it was expected that there will be more water transportation passing the camera, such as small boats or bigger boats. After collaborating the whole data, eventually there are other floating objects in Martapura river. The non-water vehicle objects were dominated by water hyacinth, small debris, and garbage. As can be seen from Fig. 2 (a,b,c,d) AlexNet successfully detect small boats in Frame 3 of the 3rd second. Fig. 3 shows that boats and water hyacinth or plantation has also detected by AlexNet in Frame 5 of the 5th second. These detection shows that besides water vehicle, AlexNet detected other floating objects which might be considered as a threat for the water vehicle's track and the river. These whole overall detection process can be seen in Figure.4 including layers and output shapes which then used to define training accuracy and validation.

(a) Frame 12 of the 12nd second

(b) Frame 27 of the 27th second

(c) Frame 19 of the 19th second

(d) Frame 3 of the 3rd second

Fig. 2. Detecting Water Vehicle using AlexNet

Fig. 3. Frame 5 of the 5th second when AlexNet detect water hyacinth

```
Layer (type)                Output Shape              Param #
=================================================================
conv2d (Conv2D)             (None, 55, 55, 96)        34944

max_pooling2d (MaxPooling2D  (None, 27, 27, 96)       0
)

conv2d_1 (Conv2D)           (None, 23, 23, 256)       614656

max_pooling2d_1 (MaxPooling  (None, 11, 11, 256)      0
2D)

conv2d_2 (Conv2D)           (None, 9, 9, 384)         885120

conv2d_3 (Conv2D)           (None, 7, 7, 384)         1327488

conv2d_4 (Conv2D)           (None, 5, 5, 256)         884992

max_pooling2d_2 (MaxPooling  (None, 2, 2, 256)        0
2D)

flatten (Flatten)           (None, 1024)              0

dense (Dense)               (None, 4096)              4198400

dropout (Dropout)           (None, 4096)              0

dense_1 (Dense)             (None, 4096)              16781312

dropout_1 (Dropout)         (None, 4096)              0

dense_2 (Dense)             (None, 1000)              4097000

dropout_2 (Dropout)         (None, 1000)              0

dense_3 (Dense)             (None, 2)                 2002

=================================================================
Total params: 28,825,914
Trainable params: 28,825,914
Non-trainable params: 0
```

Fig. 4. Object Detection process using AlexNet

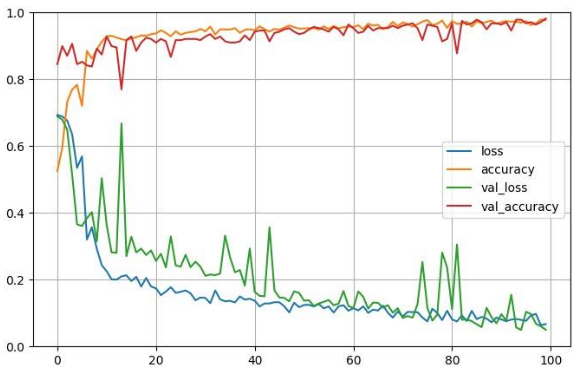

Fig. 5. Training Accuracy and validation

The training accuracy and validation for object detection using AlexNet shows in Figure.5 achieved 97% accurateness with 28.825.914 total parameters successfully trained and zero non-trainable parameters. This means that deep learning technology can detect floating objects in water environment effectively. Table 1 shows the results of data training taken in June-August 2022 which indicates that there were around 5 to 7 water vehicles were detected per hour, while there were 15 non water vehicle detected per hour. The average number of water vehicle per hour (6 objects per hour) actually is much lower than the average number of non-water vehicle per hour (15 objects per hour). Hence, classification of each data series needs to be prepared for next stage of this study.

Table 1. Results of Data Training

Date	Time	AlexNet Object Detection	
		Average Number of Water Vehicle per hour	Average Number of Non-water Vehicle per hour
1-5 June 2022	09.00 – 10.00	6	9
1-5 June 2022	14.00 – 15.00	5	5
6-12 June 2022	09.00 – 10.00	5	9
6-12 June 2022	14.00 – 15.00	7	17
13-19 June 2022	09.00 – 10.00	5	9
13-19 June 2022	14.00 – 15.00	7	8
20-26 June 2022	09.00 – 10.00	5	9
20-26 June 2022	14.00 – 15.00	7	12
27 June-3 July 2022	09.00 – 10.00	7	16
27 June-3 July 2022	14.00 – 15.00	7	18
4-10 July 2022	09.00 – 10.00	5	9
4-10 July 2022	14.00 – 15.00	7	18
11-17 July 2022	09.00 – 10.00	6	23
11-17 July 2022	14.00 – 15.00	7	6
18-24 July 2022	09.00 – 10.00	5	26
18-24 July 2022	14.00 – 15.00	7	21
25-31 July 2022	09.00 – 10.00	5	12
25-31 July 2022	14.00 – 15.00	6	13
1-7 Agt 2022	09.00 – 10.00	5	28
1-7 Agt 2022	14.00 – 15.00	6	13
8-14 Agt 2022	09.00 – 10.00	5	26
8-14 Agt 2022	14.00 – 15.00	7	10
15-21 Agt 2022	09.00 – 10.00	5	18
15-21 Agt 2022	14.00 – 15.00	7	19
22-31 Agt 2022	09.00 – 10.00	6	22
22-31 Agt 2022	14.00 – 15.00	6	15
AVERAGE		6	15,04

4 Discussion

Compared to other method, AlexNet provides much more advantages in terms of time and accurateness in detecting floating objects. Some preliminary research were done to detect floating objects using HLF methods [18], [19] and has shown the lack of competence in detecting floating objects. The method practically cannot be applied for detecting floating objects and more suitable for detecting living objects. Therefore, in this research deep learning technology has shown its efficiency and accurateness in detecting moving objects especially in water environment which has threads through its reflection, low-high tides and rapid movement.

Result shows that Alexnet performed around 97% accurateness (see Figure.5) in detecting boats or other water vehicle as the main water transportation. The data can be used to count number of water vehicle per hour. Furthermore, it is also possible to implement using real time data. However, these level of zero non-trainable parameter also might indicate that this research need more experiments with even larger set variation of data with different types of environments. The condition of the weather like heavy rain, tides, and streams has to be considered when detecting floating objects. In this term, this research needs to expand its experiments in wider and broader framework and method.

There were around 5 to 7 water vehicles were detected per hour, while there were 15 non water vehicle detected per hour. For the local government, this certainly helps the traffic counting in Martapura river and manage the cleanliness of the river environment. These number of results gave baseline for water traffic counting, where the Government can control the flow of water vehicle. However, AlexNet can work better with large amount of data. Therefore, the local government can consider to install camera along the riverside which can connect directly to the control panel of the city, and integrate the system as part of smart city application. In this way, the traffic counting can be monitored in real time.

5 Conclusion

Other than object detection, Alexnet as Deep Learning technology can be used for water traffic counting. During this research, Alexnet also detect other floating objects like water plantation or plastic garbage with small amount of accuracy. This experiment still have not classify each type of ship based on its dimension or typology. In the future, this research might be needing larger data in different weather condition to have a better and precise result.

References

1. H. Wang, Y. Cai, and L. Chen, “A Vehicle Detection Algorithm Based on Deep Belief Network,” *Sci. World J.*, vol. 2014, pp. 1–7, 2014, doi: 10.1155/2014/647380.

2. J. Wang, Y. Ma, L. Zhang, R. X. Gao, and D. Wu, “Deep learning for smart manufacturing: Methods and applications,” *J. Manuf. Syst.*, vol. 48, pp. 144–156, Jul. 2018, doi: 10.1016/j.jmsy.2018.01.003.
3. R. Khallaf and M. Khallaf, “Classification and analysis of deep learning applications in construction: A systematic literature review,” *Autom. Constr.*, vol. 129, p. 103760, Sep. 2021, doi: 10.1016/j.autcon.2021.103760.
4. N. Maharjan, H. Miyazaki, B. M. Pati, M. N. Dailey, S. Shrestha, and T. Nakamura, “Detection of River Plastic Using UAV Sensor Data and Deep Learning,” *Remote Sens.*, vol. 14, no. 13, p. 3049, Jun. 2022, doi: 10.3390/rs14133049.
5. D. Rong, L. Xie, and Y. Ying, “Computer vision detection of foreign objects in walnuts using deep learning,” *Comput. Electron. Agric.*, vol. 162, pp. 1001–1010, Jul. 2019, doi: 10.1016/j.compag.2019.05.019.
6. M. Ł. Kowalski *et al.*, “Detection of Inflatable Boats and People in Thermal Infrared with Deep Learning Methods,” *Sensors*, vol. 21, no. 16, p. 5330, Aug. 2021, doi: 10.3390/s21165330.
7. L. Steccanella, D. Bloisi, J. Blum, and A. Farinelli, “Deep Learning Waterline Detection for Low-Cost Autonomous Boats,” in *Intelligent Autonomous Systems 15*, M. Strand, R. Dillmann, E. Menegatti, and S. Ghidoni, Eds., in Advances in Intelligent Systems and Computing, vol. 867. Cham: Springer International Publishing, 2019, pp. 613–625. doi: 10.1007/978-3-030-01370-7_48.
8. A. M. Rekavandi, L. Xu, F. Boussaid, A.-K. Seghouane, S. Hoefs, and M. Bennamoun, “A Guide to Image and Video based Small Object Detection using Deep Learning : Case Study of Maritime Surveillance.” arXiv, Jul. 26, 2022. Accessed: Oct. 08, 2022. [Online]. Available: http://arxiv.org/abs/2207.12926
9. S. López-Tapia, R. Molina, and A. K. Katsaggelos, “Deep learning approaches to inverse problems in imaging: Past, present and future,” *Digit. Signal Process.*, vol. 119, p. 103285, Dec. 2021, doi: 10.1016/j.dsp.2021.103285.
10. A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.
11. B. Li, “Hearing loss classification via AlexNet and extreme learning machine,” *Int. J. Cogn. Comput. Eng.*, vol. 2, pp. 144–153, Jun. 2021, doi: 10.1016/j.ijcce.2021.09.002.
12. L. Alzubaidi *et al.*, “Review of deep learning: concepts, CNN architectures, challenges, applications, future directions,” *J. Big Data*, vol. 8, no. 1, p. 53, Dec. 2021, doi: 10.1186/s40537-021-00444-8.
13. L. Zhang, Y. Zhang, Z. Zhang, J. Shen, and H. Wang, “Real-Time Water Surface Object Detection Based on Improved Faster R-CNN,” *Sensors*, vol. 19, no. 16, p. 3523, Aug. 2019, doi: 10.3390/s19163523.
14. S. Kamada and T. Ichimura, “An Object Detection by using Adaptive Structural Learning of Deep Belief Network,” in *2019 International Joint Conference on Neural Networks (IJCNN)*, Budapest, Hungary: IEEE, Jul. 2019, pp. 1–8. doi: 10.1109/IJCNN.2019.8852145.
15. F. Lin, T. Hou, Q. Jin, and A. You, “Improved YOLO Based Detection Algorithm for Floating Debris in Waterway,” *Entropy*, vol. 23, no. 9, p. 1111, Aug. 2021, doi: 10.3390/e23091111.
16. A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.

17. M. Raveendra and K. Nagireddy, "Inter frame Tampering Detection based on DWT-DCT Markov Features and Fine tuned AlexNet Model," *Int. J. Comput. Sci. Netw. Secur.*, vol. 20, no. 12, pp. 1–12, Dec. 2020, doi: 10.22937/IJCSNS.2020.20.12.1.
18. N. Saubari, M. Gazali, and R. Ansari, "Metode HLF untuk Deteksi Objek Terapung pada Permukaan Sungai Martapura," *JISKA J. Inform. Sunan Kalijaga*, vol. 4, no. 2, p. 43, Dec. 2019, doi: 10.14421/jiska.2019.42-06.
19. N. Saubari, R. Ansari, and M. Gazali, "Deteksi Objek Terapung pada Sungai Martapura dengan Metode Haar Like Feature Menggunakan Kamera Smart Phone," *J. Sist. Inf. BISNIS*, vol. 9, no. 2, p. 141, Nov. 2019, doi: 10.21456/vol9iss2pp141-148.