The use of mobilenet v1 for identifying various types of freshwater fish

E Suharto¹, Suhartono¹, A P Widodo¹ and E A Sarwoko¹

Department of Informatics, Faculty of Science and Mathematics, Diponegoro University
Jl. Prof.Dr.Soedarto, SH, Tembalang, Semarang 50275, Indonesia

Corresponding author: edys@lecturer.undip.ac.id

Abstract. In recent years, the business opportunities for freshwater fish potential utilization are very promising. Freshwater fish including pomfret, Nile tilapia, carp, goldfish, tilapia fish, snapper, and catfish have great economic value. They are usually exported in living conditions to several countries such as Singapore, Japan, Hong Kong, Taiwan, and Malaysia. The business of fish particularly freshwater fish is one of the national income sources. As the forms of fish vary, it is important to distinguish the types of freshwater fish utilizing object detection. This detection can be done by implementing deep learning. MobileNet V1 is a deep learning model that can be used for object detection or image classification. MobileNet V1 can work on smartphones or other embedded devices that still produce high-level accuracy. In this study, MobileNet V1 was trained with learning rate parameters 0.0004 and epoch 20,000. The use of these parameters obtained an accuracy rate of 90% in the detection of types of freshwater fish.

1. Introduction

Business opportunities by utilizing freshwater potential in recent years are increasingly promising. The business includes freshwater grass, sea cucumbers and various types of freshwater fish. Freshwater fish species include pomfret, tilapia, carp, goldfish, Nile-tilapia, snapper, catfish, and others. These fish have great economic value since these fish are usually exported in living conditions to several countries such as Singapore, Japan, Hong Kong, Taiwan, and Malaysia [1]. For example, the price of Nile tilapia, carp and snapper is indeed very promising, especially for business people. The price at the level of farmers/fishermen is IDR 60,000 - IDR 90,000 per kg for each live fish. As the business of freshwater fish is a source of national income for the country, that is so tempting to cultivate.

For this reason, the Directorate of Hatchery Management, the Directorate General of Fisheries at the Ministry of Agriculture, has paid great attention to the cultivation of several species of fish. Likewise, the State Ministry of Research and Technology seeks to create appropriate technology, so that the cultivation of tilapia, carp, and snapper truly produces products that have high business value [2]. Various activities related to the business of various types of fish include fisheries, trade, transportation and so on. For this reason, a model is needed to identify the type of fish.

The recognition of the pattern of any type of fish can be done through image processing. It has a more sensitive ability because it is equipped with an electro-optic sensor. This sensor is more precise and objective when compared to visual means that are subjective and strongly influenced by the psychic
Object recognition can be done by using the latest concept, namely deep learning. As a new branch of machine learning, deep learning is growing rapidly recently. Deep learning can produce a higher level of sensitivity compared to other machine learning methods. Research in deep learning has been carried out in various fields including health, safety, traffic, and various other fields. MobileNet is one of the deep learning models that can be used to detect objects. MobileNet v1 is a relatively new model and can be used on a variety of devices including embedded systems and smartphones [3]. MobileNet v1 can enable fish business managers to distinguish fish types through various devices.

2. Literature review

2.1. MobileNet v1
MobileNet v1 is a model developed by Andrew G. Howard and colleagues. MobileNet v1 can be used for object detection or classification [3]. MobileNet is developed using the depthwise separable convolution architecture. Depthwise separable convolution is divided into two, namely depthwise convolution and 1x1 pointwise convolution [4]. It is shown in figure 1 the depthwise separable convolution operation. Meanwhile, it is shown in figure 2, a 1x1 pointwise convolution operation. Each layer of MobileNet v1 is followed by the batch norm and ReLU. Batch normalization is a normalization process by reducing the average value and dividing it by standard deviation [5]. The MobileNet v1 architecture has 28 layers [4]. When MobileNet v1 architecture is used for object detection processes, the last layer section is given a Single Shot Detector (SSD) which functions as a detector object on the object [5].

2.2. Single shot detector
Single Shot Detector (SSD) is a feed-forward convolution-based object detector that generates a collection of bounding boxes along with values and presents classes on each bounding box. SSD requires image input and ground truth boxes, being a kind of square to mark objects to be identified, to be used in the training process. Figure 3 displays the SSD’s Architecture. Meanwhile, to produce high accuracy values, there are several features on SSD, namely multi-scale feature maps for detection, convolutional predictors, also default boxes and aspect ratios [6].
3. Research method

3.1. Data set

The dataset used was derived from Google Image and was divided into three classes, namely pomfret, carp and tilapia fish. The number of datasets was 700 images. The images were divided into three groups, 469 images for training images, 201 images for test images and 30 images for validation images. Training images, test images, and validation images were of different datasets and had no similarities. A total of 700 images were rotated images of 90°, 180°, and 270°.

3.2. Determination of ROI (Region of Interest) in the image.

The initial process before training was determining the ROI. It functioned so that the object to be detected was recognized. Determining ROI was done by using the labeling application. The application could be downloaded via https://github.com/tzutalin/labelImg. The results of determining ROI were xml files used in the training process. The xml file contained variables including:

- the filename that provided file name information.
- x1, x2, y1, y2 provided information in the form of coordinates of a square that functioned as markers of objects.
- class_name provided information with class names from the square.

The process of determining ROI in the image is shown in figure 4.

3.3. Training

The training was carried out using the TensorFlow library and the Python programming language. The training process used learning rate parameters that were worth 0.0004. These parameters were combined with epoch 20000. The training process used test images and training images that had been mentioned.
in section 3.1. After the process, the training finished producing a model file that could be used in the validation process. The latter process determined the ability of the model to detect validation images that were not found in the training process.

Various image sizes could be used as training images. The process of determining was carried out to determine the coordinates of the object detected. Then the training data used in the training process was carried out to change the sizing process to 300x300x3. The last number 3 denoted that the image input used was an RGB (Red Green Blue) image.

3.4. Testing
Testing was done by using the jupyter notebook application. It used the model file that had been formed in the training process. The testing process used a validation image that had been explained in section 3.1. The test aimed to determine the accuracy produced by a model file in detecting objects at images not used in the training process. Image input used in testing was the same as the input image in the training process.

4. Results and discussion
The MobileNet v1 training process was carried out to identify the types of freshwater fish. The training was conducted in about 8 hours. Each model file generated at each training produced different accuracy. Figure 5 displays accuracy information in the training process with learning rate parameters 0.0004 and epoch 20000. Training using learning rate parameters 0.0004 and epoch 20000 resulted in an accuracy of 90% and error detection occurred in the image numbers 9, 10 and 17. Figure 6, Figure 7 and Figure 8 show images of detection information on the image of carp fish, tilapia fish, and pomfret fish.

![Figure 5](image1.png)

**Figure 5.** Training with learning rate 0.0004 and epoch 20000 accuracies.

![Figure 6](image2.png)

**Figure 6.** Detection of Tilapia fish class.

![Figure 7](image3.png)

**Figure 7.** Detection of Carp Fish Class.
5. Conclusion
MobileNet v1 as a lightweight model developed with architectural depthwise separable convolution can be used on smartphones or other embedded devices. This model can be used for object detection or image classification. In this study, MobileNet v1 was trained with learning rate parameters of 0.0004 and epoch 20,000 to distinguish various types of freshwater fish. The use of these parameters trained the MobileNet v1 model to be able to distinguish three types of freshwater fish, namely Tilapia fish, Carp fish and Pomfret fish which obtained 90% accuracy.

References
[1] Indarjo A 2015 Manipulation on environment and feed for improving aquaculture export products (Indonesian title: Manipulasi Lingkungan dan Pakan Dalam Upaya Peningkatan Produk Ekspor Budidaya Perikanan), competitive grant
[2] Hartoto D I 2003 Reference series of modern public aquaculture management (Indonesian title: Seri Rujukan untuk Pengelolaan Perikanan Perairan Umum Modern) No:02/CFB/2003.
[3] Howard A G, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto Mand Adam H 2017 Mobilenets: Efficient convolutional neural networks for mobile vision applications. Retrieved from http://arxiv.org/abs/1704.04861
[4] Biswas D, Su H, Wang C, Stevanovic A, and Wang W 2018. An automatic traffic density estimation using Single Shot Detection (SSD) and MobileNet-SSD. Physics and Chemistry of the Earth. https://doi.org/10.1016/j.pce.2018.12.001
[5] Ioffe S and Szegedy C 2015 Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. Retrieved from http://arxiv.org/abs/1502.03167
[6] Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C and Berg A C 2016 SSD : Single Shot MultiBox Detector. DOI: 10.1007/978-3-319-46448-0_2 Retrieved from https://arxiv.org/abs/1512.02325