Performance Analysis on Fine-tuned Region-based CNN for Object Recognition

Akhmad Yusuf Nasirudin*, Suryo Adhi Wibowo* and Rita Purnamasari**
*Center of Advenced Wireless Technologies, Telkom University,
Bandung, Indonesia,
E-mail:akhmadyusufnasirudin@gmail.com, suryoadhiwibowo@telkomuniversity.ac.id
**School Electrical Engineering, Telkom University
Bandung, Indonesia,
E-mail: ritapurnamasari@telkomuniversity.ac.id

***Abstract*—In this time, machine learning technology has developed rapidly, especially on deep learning architecture where many models have been created and already has a good result. However, to create a good system it took a long process such as designing a model architecture, creating a vast number of dataset and test the model many times to obtain the best performance. It is practically hard to create such a system. Therefore fine-tuning is a fascinating thing to discuss, by merely taking a model that has been known having a good result and configure the model to suit our needs, fine-tuning becomes more popular method than to create a model from scratch. In this experiment, we tried to fine-tune the R-CNN model where the pre-trained model used the Residual network architecture. Our best is at 90%.**

***Index Terms*—Deep learning, Neural network, Fine-tuning**

I. Introduction

The need for an artificial intelligence system has increased rapidly in the current era. Human jobs that have patterns begin to be replaced by artificial intelligence. This is one of the backgrounds for the creation of tools such as the autonomous car and robotic humanoid [1]. To support this development, a good performance of an artificial intelligence system is needed [2]. So, research on the artificial intelligence system began to bloom, starting from the architecture, methods, and datasets used to get the system model that has the best performance. Some models such as ResNet [3], VGG-16 [4], and MobileNet [5] have been developed and have quite good performance. But to get the model requires a lot of experiments, datasets, and very high hardware specifications. Therefore to overcome this, Fine-tuning was done.

Fine-tuning is a way in which we can utilize existing models and train them with far fewer new datasets. Besides being able to save time and be able to use hardware with lower specifications, fine-tuning become more popular because, with a much smaller dataset, the accuracy of the fine-tuning model approaches the original model.

II. System Model

We proposed [3] as an architecture that will be fine-tuned. This architecture allows to be done with deeper layers than others, because it has a shortcut that connects layers with the next layer directly, there will be no loss of depth data in deeper layers. Therefore the accuracy obtained will be even more significant when the depth of the layer increases [3]. The better the accuracy of the network that has been trained before, the better when the measurements are fine-tuned [6].

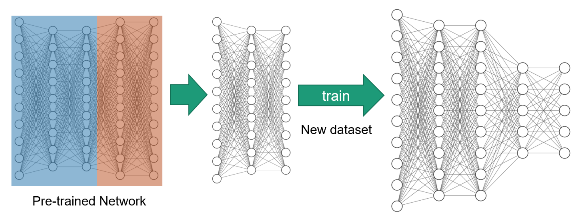

Fig. 1. Fine-tuning Process

III. Experiment

We make a model with [3] architecture that was trained using the Cifar-10 dataset with a total number of images of 60 thousand images divided into 10 classes [7], then fine-tuning with 75 training images divided into 5 classes.

Figure 1 explains that the process we performed in fine-tuning is that we freeze the weight of all layers except the last fully connected layer and the output layer in the pre-trained model, then we delete the two layers. Then from the model that has been obtained, we train the model with a new dataset and become a new model.

IV. Analysis

We conducted 6 experiments with 2 types of kernel sizes 3 and 5 where each kernel was tested on epoch 80, 100, and 120. From these six experiments we calculated the accuracy of each model both the pre-trained model and the fine-tuning model with the following formula:

$$accuracy = \frac{\sum P_t}{N} \quad (1)$$

Where Pt is the number of correct predictions and N is the total number of test data.

From figure 2, we know that by using fine-tuning, the fine-tuned model has an accuracy that is close to the accuracy of the pre-trained model, and even there are results where the fine-tuned model has higher accuracy than the pre-trained model.

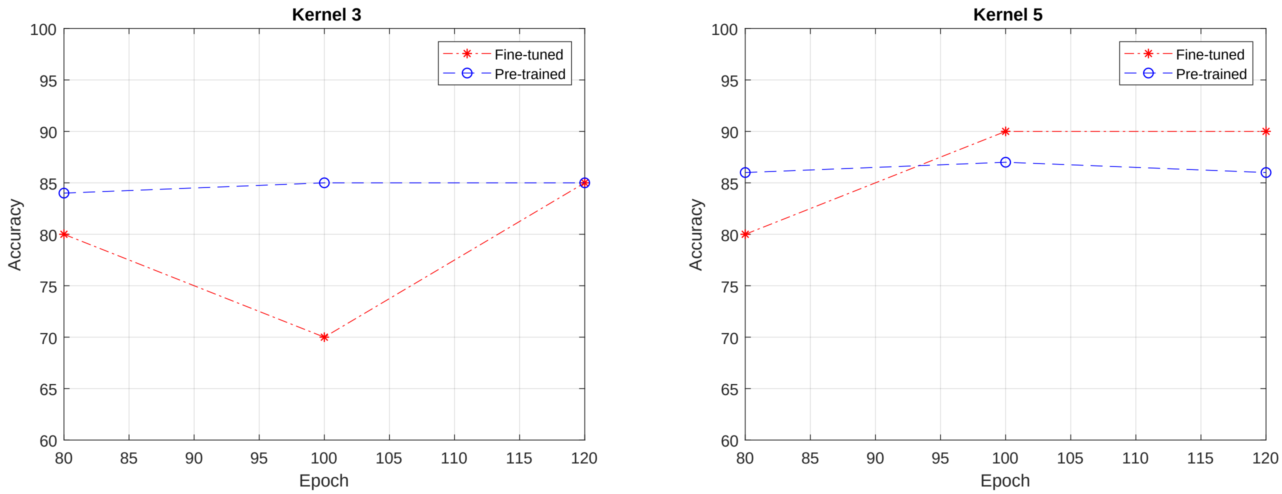

Fig. 2. Experimental Results

V. Conclusion

Fine-tuning is a good way when we want to make a model with good accuracy and only use a small dataset.

References

[1] S. A. Wibowo, H. Lee, E. K. Kim, and S. Kim, "Collaborative learning based on convolutional features and correlation filter for visual tracking," *International Journal of Control, Automation and Systems*, vol. 16, no. 1, pp. 335–349, 2018.

[2] ——, "Convolutional shallow features for performance improvement of histogram of oriented gradients in visual object tracking," *Mathematical Problems in Engineering*, vol. 2017, 2017.

[3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[4] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[5] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[6] Z. Zhou, J. Y. Shin, L. Zhang, S. R. Gurudu, M. B. Gotway, and J. Liang, "Fine-tuning convolutional neural networks for biomedical image analysis: Actively and incrementally." in *CVPR*, 2017, pp. 4761–4772.

[7] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," Citeseer, Tech. Rep., 2009.