Analysis of Transfer Learning on Faster R-CNN for Vehicle Detection

Aldi Wiranata*‡, Suryo Adhi Wibowo*‡ and Raditiana Patmasari‡
*Center for Advanced Wireless Technologies (AdWiTech)
‡School of Electrical Engineering
Telkom University
Bandung, Indonesia, 40257
E-mail: aldiwiranata@student.telkomuniversity.ac.id,
{suryoadhiwibowo, raditiana}@telkomuniversity.ac.id

***Abstract*—Computer vision is one of the favorite research topics recently, especially for object detection task. Faster Region-based Convolutional Neural Network (R-CNN) is a state-of-the-art object detection algorithm. This method has an excellent performance influenced by several parameters such as the number of convolution layers, epoch, padding scheme, network initialization, etc. In this paper, we perform an analysis of the impact of transfer learning using pre-trained AlexNet on Faster R-CNN for vehicle detection. Transfer learning method enables us to use a small amount of training data and training time to achieve good performance. Based on the experimental results, the performance of transfer learning has significant improvement by 15.9% compared to the full-training model with mAP of 73.1% at 10th epoch.**

Keywords–Convolutional neural network; Transfer learning; Object detection; Vehicle detection.

I. Introduction

Artificial Intelligence is the most emerging research area lately, computer vision is one of this branches that has many developments. Computer vision algorithms teach a machine how to have a similar vision as human. We can implement object detection in several applications, such as video surveillance, face recognition and detection, also visual tracking [1], [2]. In this paper, we implement transfer learning with pre-trained AlexNet [3] using our dataset [4], then we use the results as our convolutional layer in Faster R-CNN [5] algorithm. We consider transfer learning because this method enables us to use a small amount of training data without suffering declining performance.

II. System Model

The main idea of transfer learning is to reuse the previous network knowledge, then we perform fine-tuning on the network using a small amount of data [6]. The key benefits from this method are we can use a smaller dataset to achieve good performance and also required little time to train the network. Transfer learning usually conducted in a pre-trained network such as [3] or [7], because these networks were trained in a large dataset and have broad knowledge in many classes or objects. The reasons why this transfer learning works because the existed knowledge from the network has some intersected class with the new dataset. The illustration of our system model for transfer learning is shown in Fig. 1.

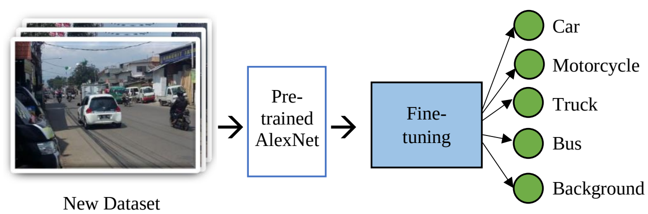

Figure 1. Illustration of transfer learning for vehicle detection.

III. Performance Analysis

In this paper, we use mean average precision (mAP) [8] as a performance parameter. In each class that is detected will be made a precision curve against the recall. A recall (r) is defined as the comparison of all positive samples from all samples in that class. While precision (p) is a comparison of the number of positive samples from positive classes:

$$r = \frac{tp}{tp + fn}, \quad p = \frac{tp}{tp + fp}, \tag{1}$$

where tp, fn, and fp denotes as true positive, false negative, and false positive.

The AP calculates curves from precision to recall, which are samples the curve at all unique recall values or for every true positive. AP can be defined as follows with M is the number of expected samples:

$$AP = \frac{1}{M} \sum_{r} p(r), \tag{2}$$

where r is the set of numbers by the formula

$$r = \left\{ \frac{k}{M} \middle| 0 \leq k \leq M, k \in \mathbb{N}_0 \right\}. \tag{3}$$

The mAP is the mean value of AP for each class. So that it can be defined as follows with N is the number of classes.

$$mAP = \frac{\sum_{i=1}^{N} AP(i)}{N}. \tag{4}$$

Detection is called successful if 50% of the bounding box (B_{p}) prediction area intersects with ground truth (B_{p}). This method is commonly referred to as Intersection over Union (IoU) which is formulated as follows:

$$IoU = \frac{area(B_{\mathrm{p}} \bigcap B_{\mathrm{gt}})}{area(B_{\mathrm{p}} \bigcup B_{\mathrm{gt}})}. \quad (5)$$

IV. Experimental Results

We are focused on vehicle detection with 4 different type of objects, there is a car, motorcycle, truck, and bus using [4] dataset. In Fig. 2 shows the comparison between transfer learning and full-training method. The transfer learning method provides outstanding performance in comparison with the full-training method with mAP gap of 15.9%, then the performance steadily produces small improvement with increasing epoch. Comparatively, the full-training method provides much lower performance, but then significantly improve and peaked at 50th epoch. The results proved that transfer learning results best when having a little amount of time or epoch for training.

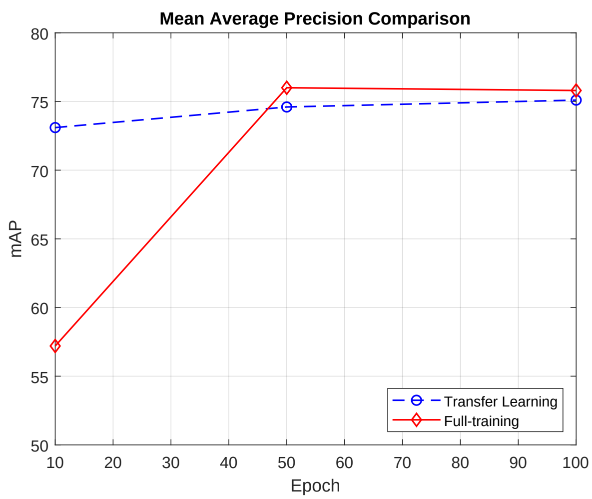

Figure 2. Detection mean average precision (percent) for transfer learning and full-training in every epoch sample.

V. Conclusion

We have concluded that transfer learning can achieve better performance with only using a small amount for training data and also with small epoch.

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, 2018, pp. 335–349.

[2] S. A. Wibowo, H. Lee, E. K. Kim, and S. Kim, "Convolutional shallow features for performance improvement of histogram of oriented gradients in visual object tracking," Mathematical Problems in Engineering, vol. 2017, 2017.

[3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in neural information processing systems, 2012, pp. 1097–1105.

[4] A. Wiranata, S. A. Wibowo, R. Patmasari, R. Rahmania, and R. Mayasari, "Investigation of padding schemes for faster R-CNN on vehicle detection," in 2018 International Conference on Control, Electronics, Renewable Energy and Communications (ICCEREC) (ICCEREC 2018), Bandung, Indonesia, Dec. 2018.

[5] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: towards real-time object detection with region proposal networks," IEEE Transactions on Pattern Analysis & Machine Intelligence, no. 6, 2017, pp. 1137–1149.

[6] S. J. Pan, Q. Yang et al., "A survey on transfer learning," IEEE Transactions on knowledge and data engineering, vol. 22, no. 10, 2010, pp. 1345–1359.

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

[8] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," International journal of computer vision, vol. 88, no. 2, 2010, pp. 303–338.