An Effective Deep-Learning Training Method using Data Augmentation

Changhyung Kim, Deayol Kim, Sooyoung Cho, Chanil Park, Kyounghak Lee

Abstract Background/Objectives: Deep learning is changing the research paradigm, showing dramatic performance improvements in many areas of computer vision.

Methods/Statistical analysis: It should be <70 words. Since Lecun's Lenet was released, Deep Learning has achieved significant performance improvements in object recognition and classification. However, it takes a huge data and takes a long time in order to learn, making it difficult to apply to real industrial environments. This method requires many manpower, high know-how and a lot of development time.

Therefore, we propose an effective deep-training training and performance enhancement method using data augmentation. After changing the original image to a YUV color space favorable to computer vision, the image is created by raising or lowering the luminance value in units of 5.Using the proposed data augmentation method can save time and cost.

Findings: In order to achieve satisfactory performance by applying deep learning to a real industrial environment, we must use our own method of producing a huge amount of data. In addition, the method of producing a direct dataset requires collecting a large amount of image data sets for a specific object and sorting the data with high quality.

In this paper, we propose efficient learning method of SSD (Single Shot MultiBox Detector)deepening learning image object recognition model based on MobileNet which is widely used in a mobile environment and embedded environment and data augmentation method to improve recognition performance (mAP).

Improvements/Applications: In SSDbased on MobileNet, the saturation of loss is faster than that of the original data set alone, and the mAP is improved by 0.7.

Keywords: Deep-Leaning, SSD, Mobilenet, Data Augmentation, Luminance Variation

I. INTRODUCTION

Deep learning is being applied in many areas of computer vision. Deep Learning uses a deep neural network to learn useful features directly from the dataset. The basis of this deep-learning technology is based on large-scale learning data called bigger and the hardware that can handle it.

Such a computer machine vision system starts with the recognition of the surrounding situation using the input image. This system cannot issue specific commands to the system if it cannot recognize nearby objects. For this reason, there are many challenges to improve object recognition efficiencies such as ILSVRC (Image Large Scale Visual Recognition Challenges) and PASCAL VOC(Visual Object Classes) challenge. As a result, it exceeded 94.9%, which is known as a human recognition rate[1,2].

Machine vision object recognition is a technique for determining the position and type of an object in an image. Since Lecun's release of LeNet, Deep Learning has achieved significant performance improvements in object recognition and classification [3]. This is superior to previous machine learning methods in the field of computer vision recognition.CNN is used for object classification and recognition in backbone network. In computer vision using deep learning, object classification and recognition work are used as a CNN network. The Convolutional Neural Network is a type of multi-layer neural network that is proposed to effectively recognize geometric relationships between data dimensions such as images. CNN network consists of feature extraction module and feature classification module [4].

Attempts have been made to apply deep-running techniques to real-world industrial environments based on high accuracy. However, the actual industrial environment is not open to data, so there is a limit to learning with only the data obtained directly. These problems are time and costly to collect and learn vast amounts of learning data. Also, the amount of data increases as the number of classes to be identified increases, and the more similar the characteristics of the class, the more data is required. Therefore, since the amount of data that can be obtained in a real environment is small, studies are being conducted to increase the amount of data for learning [5].

In this paper, we propose a method which is different from the method of vertically reversing and rotating the image which physically deforms the image. After changing the RGB domain to the YUV (YCbCr) color coordinate system, the amount of data was increased by adjusting the brightness of the image by adjusting the Y value. The proposed method is expected to improve the effective learning and learning method of depth learning using data augmentation in image recognition field.

Revised Manuscript Received on May 22, 2019.
ChanghyungKim, Department of Electronics and Communications Engineering, Kwangwoon University, Seoul, Korea
Deayol Kim, Department of Electronics and Communications Engineering, Kwangwoon University, Seoul, Korea
Sooyoung Cho, Department of Electronics and Communications Engineering, Kwangwoon University, Seoul, Korea
Chanil Park, Department of Electronics and Communications Engineering, Kwangwoon University, Seoul, Korea
Kyounghak Lee, IACF, Kwangwoon University, Seoul, Korea
*E-mail: goldbug@kw.ac.kr

II. MATERIALS AND METHODS

In this work, first, data augmentation by Luminance variation is performed, and the generated dataset is processed in order of performance verification using Mobilenet

2.1. luminance variation

Luminance variation method is shown in [Figure 1] as below. Firstly, the luminance variation method changes the original image to the YCbCr color coordinate system. Then, a new image is produced by changing the luminance value of the changed image by 5 units. Finally, the data set is composed of 20 data including the original image. The conversion formula from YCbCr to RGB is shown in Equation 1[6].

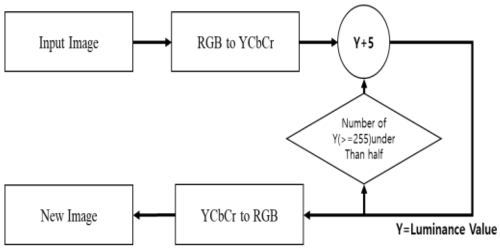

Figure 1.Generate dataset augmentation method

$$\begin{bmatrix} Y \\ C_b \\ C_r \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.169 & -0.331 & 0.500 \\ 0.500 & -0.419 & -0.081 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 0 \\ 128 \\ 128 \end{bmatrix} \begin{pmatrix} 0 \le Y \le 255 \\ 0 \le C_b \le 255 \\ 0 \le C_r \le 255 \end{pmatrix} \quad (1)$$

2.2.Proposed Object classification model

Use CNN-based computer vision with deep separable conventions. It is a model developed for the size and speed of the network. MobileNet using the Depthwise solution filters and Pointwise solution filters as shown in [Figure 2] below to solve the sequence structure as shown in [Figure 3]. CNN-based computer vision used to deep separable conventions. However, MobileNet is designed to reduce Convolution Layer structure because of its heavy weight. It is a model developed for the size and speed of the network. MobileNet used the Depthwise solution filters and Pointwise solution filters as shown in [Figure 2] below to solve the sequence structure as shown in [Figure 3]. MobileNet uses two sub-tasks of the Convolution layer, which first consists of a line layer that filters from the input terminals and a 1x1 Convolution layer that combines filtered values to create new functions[7].

Depth and point direction confluences consist of blocks of depthwise parallel confluence. It's like the traditional CNN network, but it's much faster. MoblieNet consists of a typical 3x3 console loop as the first layer in the overall architecture. And some of the depth orientation layers have structures to reduce data space dimensions. The point-wise layer doubles the number of output channels. And it's a batch of regularizations that use RELU6 for precision calculations. There is also a collective normalization. And Use RELU6 to make precision calculations. The shape of the function is similar to the sigmoid. And there is a hyper parameter called a depth multiplier Using this parameter allows you to change the number of channels for each layer. Through the structure of the convolution, MobilNetcan perform about 9 times less tasks than the existing neural networks of the same accuracy. Finally, we present a method of detecting objects in an image using one neural network. The approach is expressed by calculating the output space of the bounding box in sets with different aspect ratio and scale for different function map locations. MobilNet Network creates a score for each object category in each basic box and adjusts the box to match the appearance of each object. It also combines predictions of different functional maps with different resolutions to handle objects of different sizes. Adding objects also provides a high degree of accuracy and produces a faster.

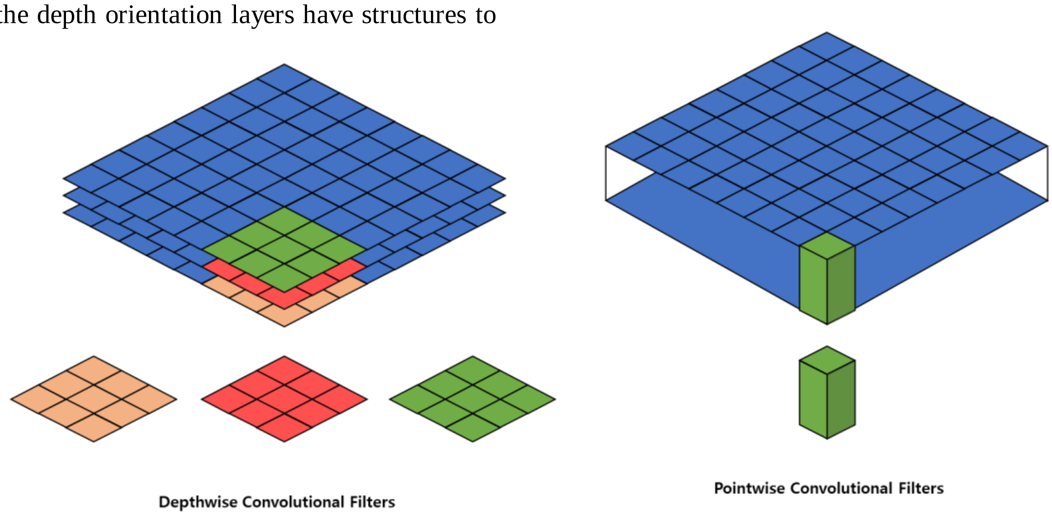

Figure 2.Depthwise and Pointwise Convolution Filters

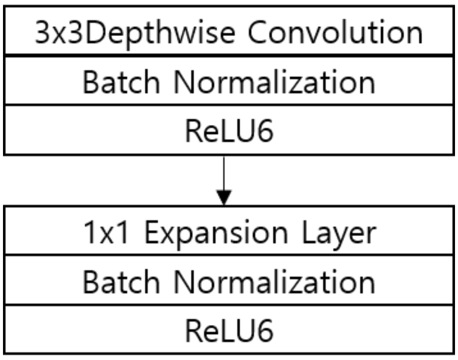

Figure 3Depthwise Separable Convolution block

III. RESULTS AND DISCUSSION

3.1. Dataset Augmentation

The data set used in the test was conducted using Kaggle's Food dataset and PASCAL VOC 2008-2011 dataset. The Food dataset provides 1000 images for each of 101 food categories, including apple pie and truce. The PASCAL VOC dataset used about 8000 images [8-9]. [Figure 4] shows Kaggle's Food dataset.

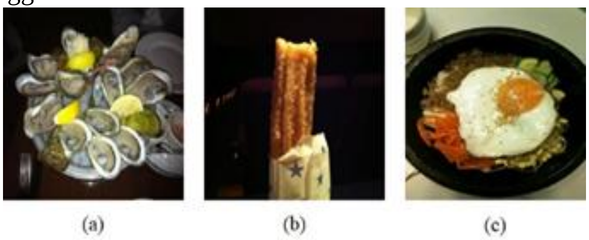

Figure 4.Kaggle Food dataset. (a) oyster (b) Churus (c) bibimbap

3.2. Results

In the experiment, 10 classes of food data set were selected. As a result of experiments, we could generate about 10 times more data than original data. Experiments were conducted using the original data and the generated data. The test environment used in the experiment is shown in [Table 1].

Table 1. Test environment

CPU	I7-8700K 4.30GHz
RAM	16GB
VGA	GeForce GTX 1080 Ti
Image size	256 x 256
Training Class	20
Training Model	SSD-MobileNet
Training Epoch	100
Training Batch Size	50

In the experiment, 10 classes of food data set were selected. As a result of experiments, we could generate about 10 times more data than original data. Experiments were conducted using the original data and the generated data.

Table 2.Simulation result

method	accuracy
original	72%
Proposed Method	90%

By using the generated data, the learning was performed and the result was confirmed to be about 18% higher than the original data.

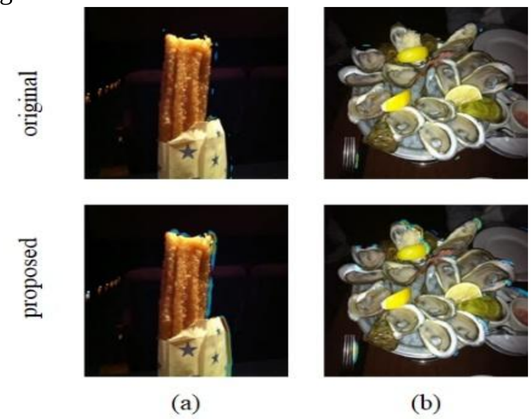

Figure 5.CAM results (a) oyster (b) Churus

[Figure 5] shows the CAM results of the model that learned only the original data and the generated data together [7]. The result of the model that we have learned the generated data together can confirm the characteristics of the image more accurately than it does not.

[Figure 6] shows the accuracy of image recognition.When the LV(Luminance Variation) was used, the mAP was 0.7 higher than that of the original training [10].

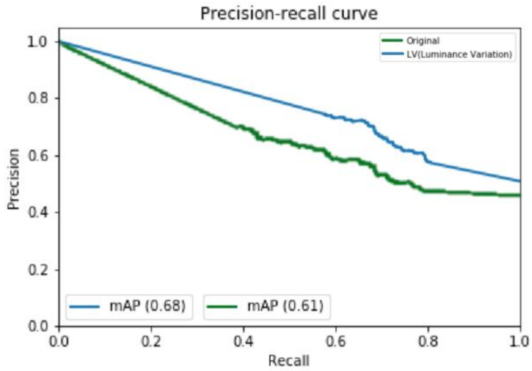

Figure 6.The result of comparing accuracy of original and LV method

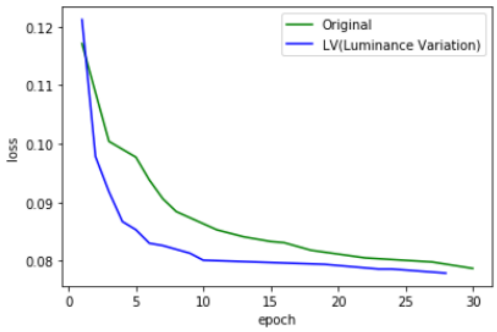

Figure 7.Loss saturation comparison result

As shown in [Figure 7], the LV method has a faster loss saturation than the original method. As the results show, the proposed method can reduce the training time.

IV. CONCLUSION

This paper proposes ways to generate data to increase the accuracy of the Deep Learning object recognition model. Deep learning requires a lot of data to be applied to the real environment. Building a satisfactory learning dataset requires the use of direct production of numerous data. Direct manufacturing involves collecting a lot of images about a particular object and classifying high-quality data. But this approach takes a lot of time and money. Therefore, this paper suggests using the method of increasing data can save time and money. Deep learning is the more data have higher the accuracy. Therefore, the amount of data to be applied to learning can be increased by using the above method. And applying it to the learning by using the above method in the specific environment, the advantage of the deep learning can be utilized to the maximum, and the step of data production process can be reduced at a moment. Using the above method, you can reduce the data generation process and reduce the increase of mAP and loss saturation more quickly.

ACKNOWLEDGMENT

The present Research has been conducted by the Research Grant of Kwangwoon University in 2018.

REFERENCES

1. Krizhevsky, A., Sutskever, I., & Hinton, G. E. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems.2012; 1097-1105.
2. Everingham, M., Van Gool, L., Williams, C. K., Winn, J., & Zisserman, A. The pascal visual object classes (voc) challenge. International journal of computer vision.2010;88(2):303-38.
3. LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. Gradient-based learning applied to document recognition. Proceedings of the IEEE. 1988; 86(11):2278-2324.
4. LeCun, Y., Bengio, Y., & Hinton, G. Deep learning. Nature. 521.7553.2015; 436.
5. Lakshmi, T. N., & Janaki, N. Deep learning based detection and recognition of objects using mobile nets and SSDs.
6. Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv.2017;1704.04861.
7. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. Learning deep features for discriminative localization. In Computer Vision and Pattern Recognition (CVPR), 2016 IEEE Conference on IEEE.2016 June;2921-29.
8. Kaggle DataSet.[updated 2018]. Available from: https://www.kaggle.com/(website)
9. PascalVOC.[updated2012]. Available from: http://host.robots.ox.ac.uk/pascal/VOC/index.html(website)
10. Li, K., Huang, Z., Cheng, Y. C., & Lee, C. H. A maximal figure-of-merit learning approach to maximizing mean average precision with deep neural networkbased classifiers. In Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on IEEE. 2014 May:4503-07