International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 21 (2017) pp. 11640-11646

Image Manipulation Detection using Convolutional Neural Network

Dong-Hyun Kim[1] and Hae-Yeoun Lee[2,*]

[1]Graduate Student, [2] PhD, Professor
[1,2] Department of Computer Software Engineering, Kumoh National Institute of Technology, 61 Dakhak-ro, Gumi, Gyeongbuk, 39177, Republic of Korea.

*(*Corresponding Author)*
[1,2]Orcid: 0000-0002-0693-431X, 0000-0002-6081-1492

Abstract

Using various methods, an image manipulation can be done not only by the image manipulation itself, but also by the criminals of counterfeiters for the purpose of counterfeiting. Digital forensic techniques are needed to detect the tampering and manipulation of images for such illegal purposes. In this paper, we present an image manipulation detection algorithm using deep learning technology, which has achieved remarkable results in recent researches. First, a convolutional neural network that is verified for image processing is applied. In addition, a high pass filter is used to acquire hidden features in the image rather than semantic information in the image. For the experiments, modified images are generated using median filtering, Gaussian blurring, additive white Gaussian noise addition, and image resizing for 256x256 images that were divided into 4 equal parts of Boss Base 1.01 images. Quantitative performance analysis is performed to test the performance of the proposed algorithm and image manipulation is detected with 95% accuracy.

Keywords: Multimedia forensic, Image manipulation, High pass filter, Convolutional Neural Network

INTRODUCTION

From the past several years, social media like KakaoTalk, Facebook, Instagram and SNS (Social Network Service) have been used by a large number of people and still the emerging of their use is increasing accordingly. They have become part of our lives. In particular, the development of smart devices such as smartphones has a remarkable role in uploading and downloading images to those social networks.

In the meantime, there has been a technique for manipulating an image using various methods with a specific purpose. Image tampering can be done by counterfeit criminals for the purpose of counterfeiting. Digital forensic techniques are needed to detect the tampering and manipulation of images for these illegal purposes and many researches have been studied on these forensic techniques. However, they use features designed by human intervention and their performance are totally dependent on the differentiation of these features among original, tampered, and modified images. Recently, the interest about deep learning has increased and many remarkable results are emerging. Hence, forensic researchers attempt to apply deep learning to detect the manipulation of images without human intervention.

In this paper, we propose an image manipulation detection algorithm using deep learning technology. The model based on a convolutional neural network (CNN) is designed. Especially, a high pass filter is used to acquire hidden features in the image rather than semantic information in the image. The convolutional layer is composed of 2 layers having maximum pooling, ReLU activation, and local response normalization. The fully connected layer is composed of 2 layers. For the experiments, modified images are generated using median filtering, Gaussian blurring, additive white Gaussian noise addition, and image resizing for 256x256 images. Quantitative performance analysis is performed to test the performance of the proposed algorithm.

The paper is organized as follows. In Section 2, we summary related works. The proposed algorithm for manipulation detection is described in Section 3. Experimental results are presented in Section 4 and Section 5 concludes.

RELATED WORK

Image Manipulation Detection

From the point view of digital forensics, image manipulation is a very important and raises serious criminal issues to take in consideration. That is why many researchers are conducting so many studies about it.

Rhee [1] is conducting research on forensic decision making using edge energy information of stochastic images. Using SA (steaking artifacts) and SPAM (subtractive pixel adjacency matrix), edge information are extracted from JPEG compressed images of an original image with various Q-Factor and a query image. This information is compared with TCJCR (threshold by combination of JPEG compression ratios) to detect image manipulation. TP (True Positive) and FN (False Negative) are 87.2% and 13.8%, respectively. Jeon et. al is studying the detection of copy-moving operation image using the mean value of the wavelet transform coefficient [2]. Also, in order to conceal the manipulation, they address a post-processing

method that can enhance the detection performance even in post-processing environments such as adding noise or compression. Bayram et. al is conducting research to detect the manipulation of images using binary similarity in images. The fact that the binary texture characteristics within the bit planes will differ between an original and a manipulated image is considered. They focus on image scaling up, rotational attack, brightness value manipulation, blurring attack, sharpening attack [3]. The similarity between binary images is measured using binary texture statistics.

Deep learning is the learning of existing artificial neural networks by stacking deeper layers. Deep learning models include Deep Neural Network (DNN), Convolution Neural Network (CNN) using convolution and pooling for image processing, and Recurrent Neural Network (RNN) [4]. In these days, to detect image manipulation, deep learning is studied to apply as follows.

Bayer and Stamm studies image manipulation detection by adding a new convolution layer [5]. In general, CNN learns the content semantics of images. Therefore, it is not suitable for detecting an operation that does not affect the image content. Therefore, it uses a convolutional layer to detect structural relationships between pixels regardless of the image content. Choi et. al studies CNN-based multi-operation detection to detect multiple attacks, not just one attack [6]. Their technique defines three types of attacks that have occurred frequently during image manipulation and detects when they are concurrently applied to images.

Recent research has shown that CNN performance is good for image processing. In this study, we have designed the image manipulation detection model using CNN.

Background Knowledge on Deep Learning

In this section, we give brief overview of Deep Learning and CNN. A lot of research has been done to solve the problem of artificial neural network mentioned above [7]. The computation of high complexity has been solved through the development of hardware performance. Furthermore, a variety of deep learning models have been proposed. One of them is a convolutional neural network model which is specialized in image processing. The convolutional neural network includes an input layer, a convolution operation layer, a fully connected layer, and an output layer as shown in Fig. 1.

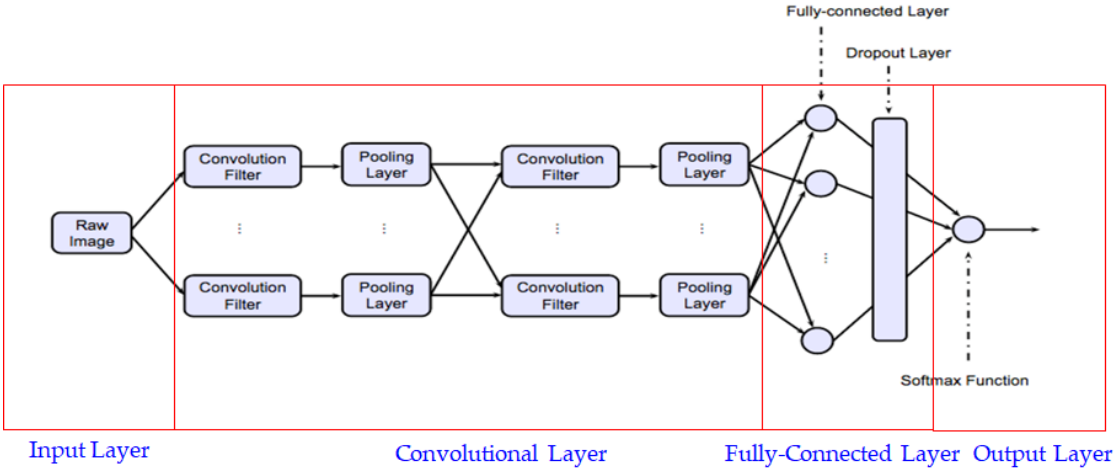
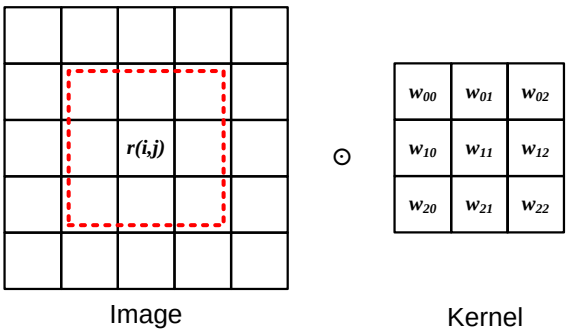


Figure 1: Overall structure of convolutional neural network



$$r_{(i,j)} = r_{(i-1,j-1)} \bullet w_{00} + r_{(i,j-1)} \bullet w_{01} + r_{(i+1,j-1)} \bullet w_{02} + \ldots$$
$$+ r_{(i-1,j+1)} \bullet w_{20} + r_{(i,j+1)} \bullet w_{21} + r_{(i+1,j+1)} \bullet w_{22}$$

Figure 2: Convolution operation

The input layer is the set of input units. It is a passage through which pixels of an image for learning are entered. Its size is related to the number of image pixels [7]. The convolutional layer consists of various convolution filters. Across the convolutional layer, the result values are passed to the next layer in a nonlinearity. In the pooling layer, the dimensionality of the data is reduced. Next, in the fully-connected layer, the classification is performed according to the learned results. Multiple fully connected layers can be stacked, Drop-out can be applied between each layer to prevent over-fitting or under-fitting. Finally, the output layer is learned to score each class and in general softmax function is used.

In particular, the convolutional layer consists of various combination of convolution, pooling, and activation operations. The computation of convolution in a neural network is a product of a two-dimensional matrix called an image and a kernel or mask. This is depicted in Fig. 2. Through this convolution, local features considering neighboring pixels can be extracted.

The pooling layer appearing after convolution is to select a pixel value having a certain characteristic among pixels in a specific region, such as maximum pooling and average pooling. Figure 3 shows an example of maximum pooling to select maximum values. Through this pooling, the size of input data can be minimized to improve the time performance. However, in aspect of detecting image manipulation, it is possible to lose important traces to determine the modifications.

The activation function is used to change the result of the hierarchy nonlinearly. Generally, ReLU, Sigmoid, tanh etc. are used.

PROPOSED DETECTION ALGORITHM

In this section, we describe the detection process of image manipulation using deep learning. The proposed algorithm is composed of two steps: learning and testing. <Fig. 4> shows a series of learning and testing processes. First, we manipulate the original image to produce a manipulated image. The generated image is classified into a learning set and a test set. Then, the learning set is fed into the proposed CNN model. Weights are then updated via back propagation. After learning, the test set is fed into the model that has been learned, and the accuracy is calculated by analyzing the result.

CNN-based deep learning model for detecting image manipulation is shown in Fig. 5. The model is composed of 1 high pass filter, 2 convolutional layers, 2 fully connected layers, and 1 output layer.

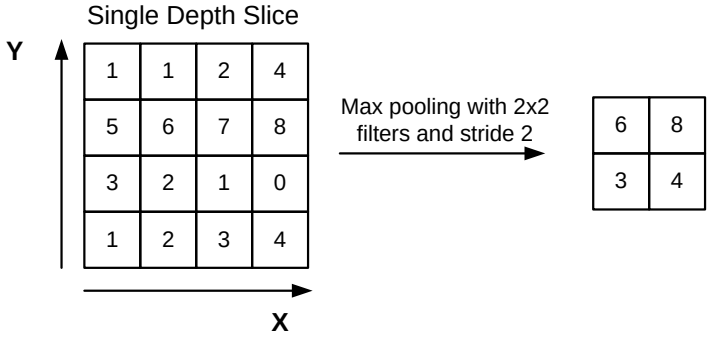


Figure 3: Maximum pooling operation

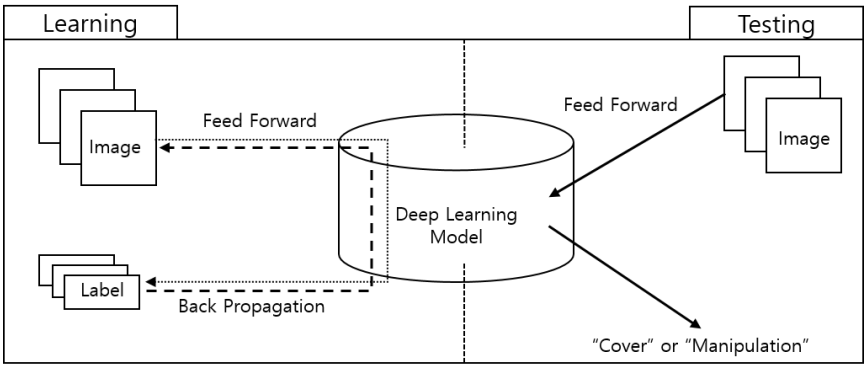


Figure 4: Image manipulation detection algorithm

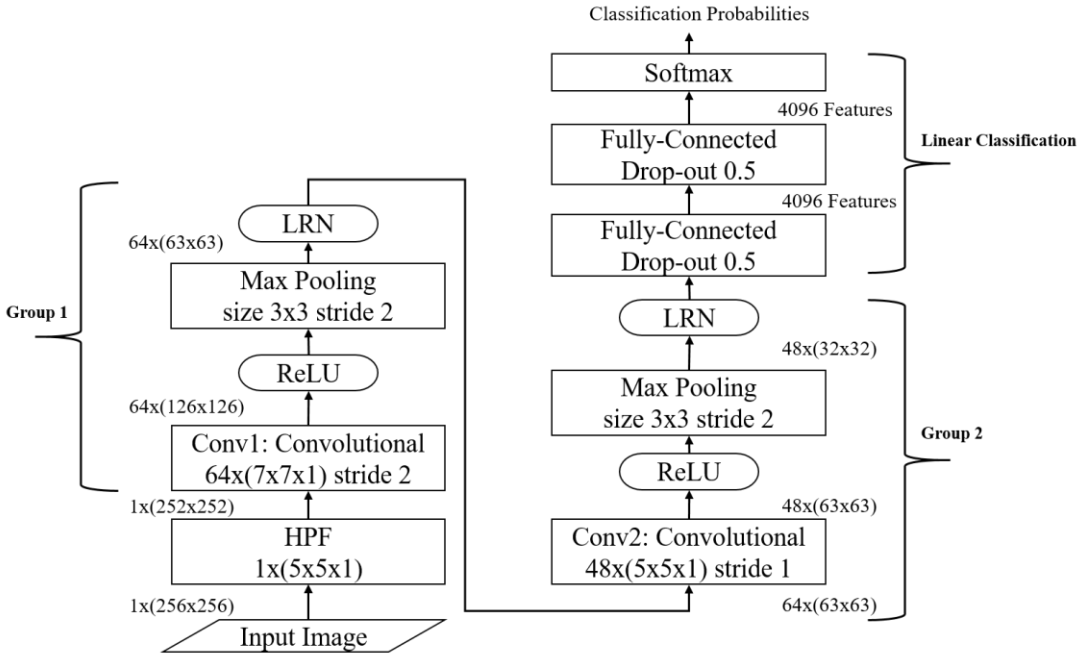


Figure 5: CNN model for image manipulation detection

High Pass Filter

In general, for image classification research, there is use of image pixel data. It behaves as if it was personally classified. However, our research targets those that cannot be distinguished by human eyes. So we apply HPF called by High Pass Filter to extract hidden features within the image as follows:

$$\mathrm{HPF} = \frac{1}{12}\begin{pmatrix} -1 & 2 & -2 & 2 & -1 \\ 2 & -6 & 8 & -6 & 2 \\ -2 & 8 & -12 & 8 & -2 \\ 2 & -6 & 8 & -6 & 2 \\ -1 & 2 & -2 & 2 & 1 \end{pmatrix} \quad (1)$$

Convolution Layer

In the process of passing through the above HPF high frequency filter, one channel image of 256x256 size is converted into 252x252. The neuron values of 252x252 pass through a 7x7 kernel of Conv1 layer with stride 2. Initialization of weight values in deep learning is important enough to influence overall performance. In this paper, we use Xavier initialization. Xavier initialization is the square of the input value and the input value of the random number value of the output value. Conv1 uses the Xavier initialization mentioned. Then, it goes through the pooling layer, which have been described above, and pass to the Conv2 layer. The Conv2 layer allows 126x126 neuron values to pass through a 5x5 kernel. The stride value of Conv2 is 2, and initialization uses Xavier initialization like in Conv1. At every convolution layer, we use an activation function called ReLU (Rectifier Linear Units). ReLU changes the output to the value of f(x) = max (0, x).

Pooling Layer

Convolutional neural networks generally have a very large number of neurons. It has been shown that this increases the complexity of learning problem. To solve this problem, we apply the pooling layer, such as maximum pooling or average pooling. Maximum pooling is used in this study. In the layer behind Conv1 and Conv2, the kernel size is 3x3 and the stride is 2x2. Also, after the Maximum Pooling layer, the brightness is standardized through a layer called LRN (Local Response Normalization). In this model, the depth radius is 4, the bias is 1.0, the alpha value is 0.001 / 9, and the beta value is 0.75.

Fully Connected Layer

In this study, two fully connected layers are used. Each has 4096 neurons, with a standard deviation of 0.4. This layer has the most basic DNN structure. We also have applied a drop-out of 0.5 to each layer to solve the over-fitting problem. Only 2048 of each of the 4096 neurons will be used for learning. We also use the softmax layer as the last layer for discrimination.

EXPERIMENTAL RESULTS

This section presents the results of the performance analysis of the proposed deep learning model. We first describe the experimental environment, describe the algorithms used to create the images which are used in the learning, and finally describe the results of the performance analysis.

Environmental Setup

In this study, we use Google Tensorflow as a deep learning framework. The GPU of NVidia GeForce 1080Ti is used and the image batch size is set to 100. The images are 10,000 images of 512x512 size of Boss Base 1.01. First, 40,000 images of 256x256 size are created by dividing the images into two horizontally and vertically divided images. The total of 200,000 images including the original images and images from 4 algorithms: median filtering, Gaussian blurring, AWGN (additive white Gaussian noise addition), and Re-Sampling are used. Of these, 80% (160,000 images) are used as learning data and 20% (40,000 images) are used as test images. Samples of test images are shown in Fig. 6.

As shown in Table 1, for each algorithm, a 5x5 kernel is used for median filtering, Gaussian blurring with a standard deviation of 1.1, and AWGN with a standard deviation of 2. In the case of Re-Sampling, bilinear interpolation is performed at a magnification of 1.5 times to enlarge and reduce to its original size.

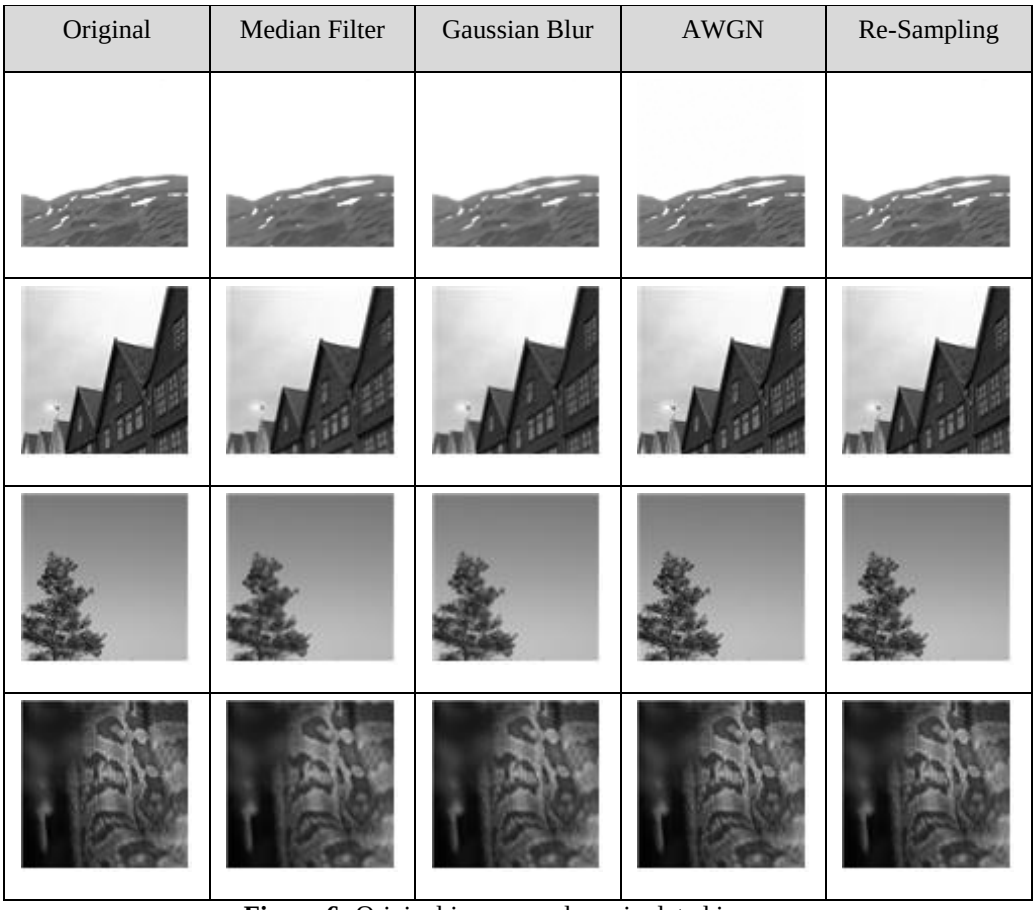


Figure 6: Original images and manipulated images

Table 1: Manipulation operation and parameter

Manipulation operation	Parameter
Median filtering	5x5 kernel
Gaussian blurring	5x5 kernel, Standard deviation=1.1
AWGN(Additive White Gaussian Noise)	Standard deviation = 2
Re-Sampling	Scaling = 1.5

Analysis Results

In the proposed algorithm, The HPF applied to the image is constructed equal to the kernel size 5x5. Generally, when working on a kernel of a certain size, a high-frequency filter of that size is used to extract the feature. The extracted feature is depicted in Fig. 7.

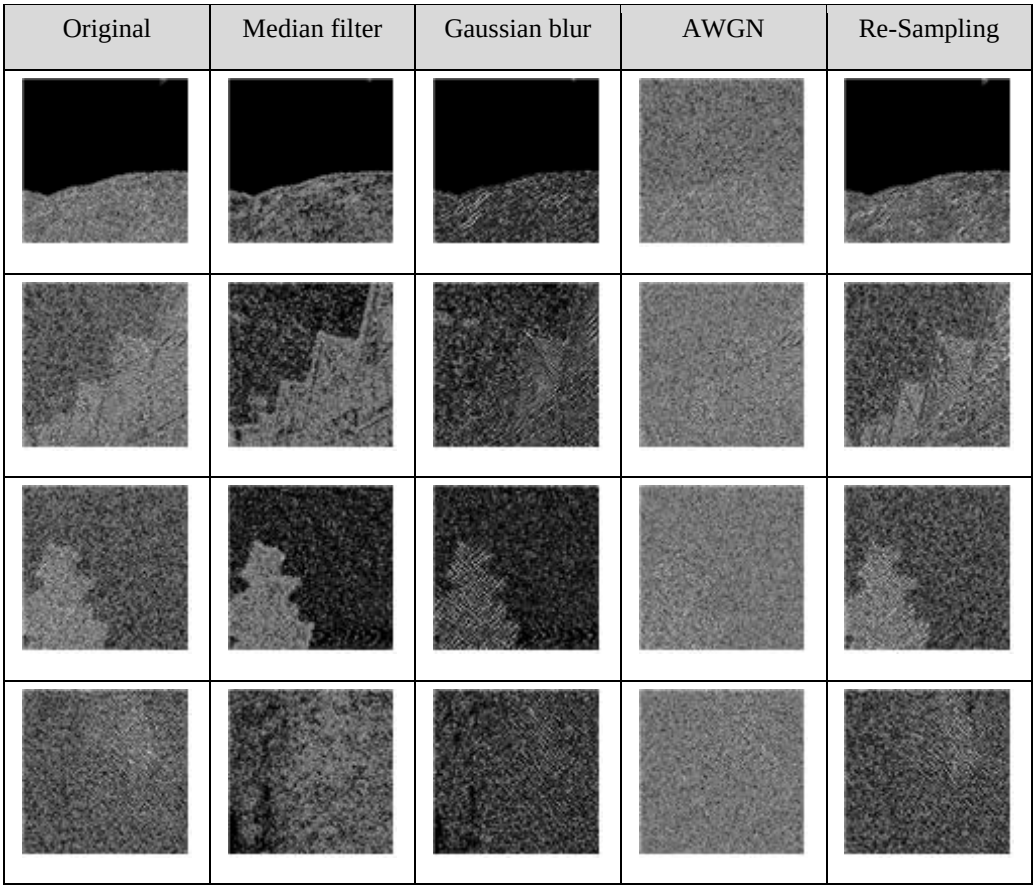


Figure 7: Extracted feature images with HPF

The CNN model is learnt by using the above images, and the accuracy is calculated in the test images process. The accuracy result is summarized in Fig. 8 where y axis represents the accuracy and x axis represents the number of learning. Table 2 shows the accuracy of each image modification detection in some epochs. Each number represents the image discrimination accuracy.

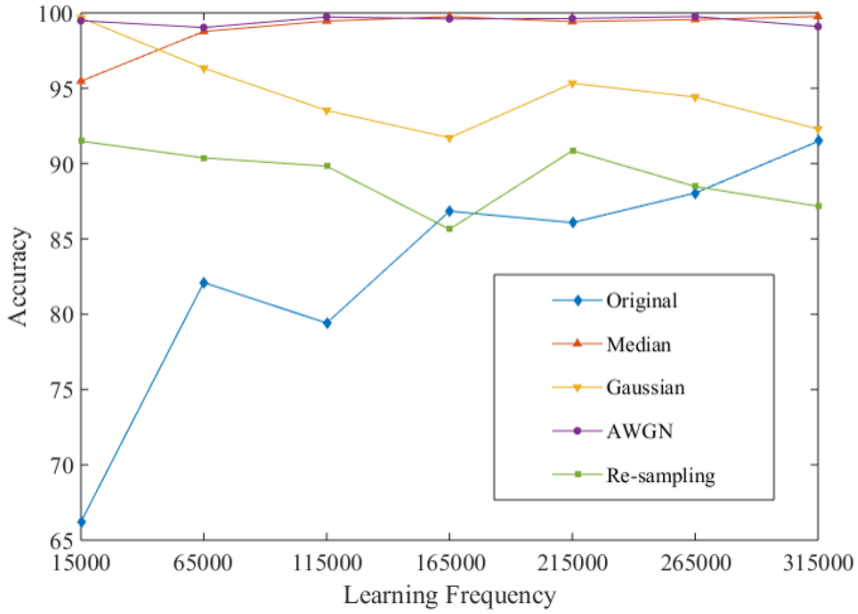


Figure 8: Accuracy by learning frequency

Table 2: Accuracy by learning frequency

Epoch	Original	Median filter	Gaussian blur	AWGN	Re-Sampling
15,000	66.21	95.47	99.67	99.48	91.49
120,000	92	99.66	94.16	99.15	89.67
240,000	94.99	99.95	90.84	99.23	85.04
315,000	90.92	99.45	97.5	99.48	95.98

As shown in Fig. 8, the initial accuracy of the original is low, but as the learning progress, the accuracy can approach 95%. It can also be seen that accuracy does not always increase as learning progress. We can confirm that some images are more accurate but other images are decreasing in accuracy. Therefore, proper learning may be better than over-learning. In general, CNN learns content semantics. However, using the HPF filter, we can learn hidden features inside the image.

CONCLUSION

As the internet advances rapidly in modern society, there are many social network services such as KakaoTalk, Facebook, Instagram and so on which have been used not only for good reasons but also some take the profit and use them for negative purposes. Under these circumstances, crimes against video are appearing for illegal purposes. Digital forensics are needed to detect these illegal purposes.

In this paper, we proposed image manipulation detection techniques using deep learning. After we briefly overviewed the related works, the proposed model was explained in detail. Through intensive experiments, the proposed model was analyzed and showed that at least 95 % accuracy was achieved.

The proposed model can be used to determine whether or not the image is manipulated, and can be applied for detection of more manipulation techniques if a better model is established in later studies. In addition, it will be possible to apply it to various multimedia as well as image in the further research. Under these circumstances, crimes against video are appearing for illegal purposes. Digital forensics will be needed to detect these illegal purposes.

ACKNOWLEDGMENT

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2017-2014-0-00639) supervised by the IITP(Institute for Information & communications Technology Promotion)

REFERENCES

[1] Rhee, K. H., 2014, “Image Forensic Decision Algorithm using Edge Energy Information of Forgery Image”, Journal of the Institute of Electronics and Information Engineers, 51 (3), pp. 75-81.

[2] Jeon, J.-J., Park, S.-H., Kim, Y.-I., and Eom, I.-K., 2014, “Copy-Move Forged Image Detection Using Average of Singular Values of Wavelet Coefficients”, Journal of Korean Institute of Information Technology, 12(11), pp. 119-126.

[3] Bayram, S., Avcibas, I., Sankur, B., and Memon, N, 2005, “Image manipulation detection with binary similarity measures”, Proceedings of the 2015 13th European Signal Processing Conference, pp. 1-4.

[4] Schmidhuber, J., 2015, “Deep learning in neural networks: An overview”, Neural networks, 61, pp. 85-117.

[5] Bayar, B., and Stamm, M. C., 2016, “A deep learning approach to universal image manipulation detection using a new convolutional layer”, Proceedings of the 4th ACM Workshop on Information Hiding and Multimedia Security, pp. 5-10.

[6] Choi, H. Y., Jang, H. U., Kim, D., Son, J., Mun, S. M., Choi, S., and Lee, H. K., 2017, “Detecting composite image manipulation based on deep neural networks”, Proceedings of the 2017 International Conference on Systems, Signals and Image Processing, pp. 1-5.

[7] Krizhevsky, A., Sutskever, I., and Hinton, G. E., 2012, “Imagenet classification with deep convolutional neural networks”, Proceedings of the 2015 25th International Conference on Neural Information Processing Systems, pp. 1097-1105.