Generation of Novelty Ground Truth Image Using Image Classification and Semantic Segmentation for Copy-Move Forgery Detection

KANG HYEON RHEE
School of Electronics Engineering, Chosun University, Gwangju 61452, South Korea
e-mail: khrhee@chosun.ac.kr

This work was supported by the National Research Foundation of Korea (NRF) under Grant 2021R1A2C1003406.

ABSTRACT Since the ground truth (GT) generated by CNN has pieces of patch information of the learned class, the accurate detection of Copy-Move is ambiguous. With various CNNs for image classification and semantic segmentation, the generated GT images are different yet similar to patch patterns for detecting forgery regions. It is difficult to determine which network model-generated GT image is suitable. Therefore, an optimal GT image is essential in image forensics. The proposed scheme in this paper generates a novelty GT image to solve this problem for the correct detection of Copy-Move forgery. The novelty GT image was configured using image classification and semantic segmentation. The variety of GT images is generated by adopting the state-of-the-art four image classifications and one semantic segmentation in the deep neural network. The proposed scheme implements mainly three tasks: 1) each network model generates the GT images (GT$_{net}$), 2) which are convergence synthesized into one (GT$_{conv}$), and 3) it decomposed again into GT images (GT$_{decomp}$) with a threshold value of the ‘Threshold Filter.’ Here, the GT$_{net}$ images involve two pieces of information about the image classification and semantic segmentation of the forgery image. The GT$_{conv}$ has two pieces of information as one GT image. The GT$_{decomp}$ is decomposed GT$_{conv}$ into various GT images by the threshold value, which is a permeated degree of the information about ‘Image classification’ and ‘Semantic segmentation.’ The proposed novelty GT image is accomplished with this operational flow for Copy-Move forgery detection. The results confirmed in the experiment for comparing the performance of the existing GT$_{net}$ image, and the GT$_{decomp}$ image of the proposed scheme showed that the Accuracy and F1 Score of the proposed scheme had the maximum improvement rate of 0.4% and 0.2%, respectively. Also, by estimating the proposed CMFD scheme, Area Under the Curve (AUC) is graded as ‘Excellent (A)’ with a value of 0.9 higher.

INDEX TERMS Copy-move image forensics, forgery detection, ground truth image, image classification, semantic segmentation, deep neural network.

I. INTRODUCTION

In modern social media, digital images contain a vast amount of information as an essential communication medium. While images can easily edit, transmit, and distribute information about our lives, trust in images is difficult to believe. It has emerged as an image forensics problem.

Using image editing tools (Premiere, Final Cut Pro, Vegas, Movavi, and After Effects, they registered trademarks.), a villain can easily manipulate or modify it. Therefore, Cut-Paste, Copy-Move manipulation, and distribution of malicious images cause severe personal infringement.

If there is image manipulation by such aggressive means, a defensive means to detect it is needed. Therefore, Cut-Paste [1], [2] and Copy-Move detection [3], [4] methods are being developed day by day as countermeasures.

Copy-Move operation selects one part of an image and copies it to another region of the same image. This manipulation method is used widely in the field of image forensics.

There are three main methods of detecting Copy-Move as follows:

1) ‘Block segmentation’: The forgery image to be tested is divided into several blocks, and features are extracted from each block. If blocks divided into $n \times n$ pixel square type have similar characteristics, these blocks...
are regarded as a Copy-Move area. There are some disadvantages: all blocks must be compared and computed, the computational cost is high, and other geometric transformations of moving patches cannot be handled. On the other hand, using the Simple Linear Iterative Clustering (SLIC) algorithm [5], the 2D superpixels of the image are calculated and divided into irregular shapes instead of the regular square division and groups the pixels into regions with similar values.

2) ‘Keypoint matching’: It is possible to find similar areas in the image by extracting and matching the key points of pixels without dividing the forgery image [6]. Scale-invariant feature transform (SIFT) [7], [8] is the most basic method for detecting Copy-Move. SIFT is an algorithm that extracts features that are invariant to the size and rotation of an image. The basic principle is to extract the SIFT features from two different images and find the corresponding parts by matching the most similar features in the two images. The disadvantage of SIFT is that the region of copy and moving are mismatched due to the bad key points. The expansion of the feature vector dimension for key point generation increases the computing cost and the decision time of the Copy-Move area, so it is not easy to manage the locale smoothly.

3) For pixel-based segmentation classification of images: BusterNet was proposed by Wu et al. [9] using the VGG16 net model structure. BusterNet has two branches: namely Simi-Det and Mani-Det. However, there are two drawbacks. One should ensure that both branches correctly locate regions, and the other, the Simi-Det branch, only extracts single-level and low-resolution features due to the four pooling layers in VGG16 [10].

Chen et al. proposed the CMSDNet (copy-move similarity detection network) and the STRDNet (source/target region distinction network) [11], which is an atrous convolution instead of the 4th polling layer of VGG16 was used to preserve ‘Field-of-views of filters’.

The above detection category is ‘Passive forensic detection’ methods that can detect the forgery area without needing a ground truth image of the Copy-Move area. On the other hand, the method [12] that uses CNN for Copy-Move detection is being developed. Still, since the class (copy area/move area), the ground truth image classification is trained together in the network model implementation. Hence, the author’s view seems to be regarded as ‘pre-Active forensic detection.’

The main words of Copy-Move forgery detection from the state-of-the-art [13] showed as a word cloud in Fig. 1.

Essentially, generating a good ground truth to detect the exact temper area of the forgery image is necessary for the image forensics field. To this end, this paper proposes a generation method of novelty ground truth for Copy-Move forgery detection. For this purpose, it adopts the state-of-the-art deep learning CNNs structure of existing ‘image classification’ [14]–[18] and ‘semantic segmentation’ [19]. The central concept of this paper is twofold:

1) The generated novelty GT image involves two characteristics of ‘image classification’ and ‘semantic segmentation’ of a forgery image.

2) Copy-Move detection is considered as ‘pre-Active forensic detection,’ referring to the ground truth image only when network training is performed, but while testing for forgery detection, the trained forgery detector becomes ‘Passive forensic ground truth’ in the actual field.

Also, the contributed operations of this paper for forensic detection are as follows:

1) Copy-Move forgery detection using the deep learning CNN structure for ‘image classification’ and ‘semantic segmentation.’ The generated various ground truth images converged to be one synthesized pattern.

2) The pattern image above in 1) is decomposed into a novelty ground truth image which can most accurately detect the Copy-Move patch by the ‘Crossing Over’ operation and the threshold value of the ‘Threshold Filter.’

3) The performance evaluation of general forgery detection calculates Accuracy and F1 Score from the detection result of Copy-Move after overlaying the ground truth and the forgery image. However, the many generated ground truth images already include Accuracy and F1 Score. Therefore, a novelty ground truth image with high Accuracy and F1 Score is ready to use for forgery detection; thus, CMFD is rapidly performed.

The above contribution in this paper is to reduce the generation cost of a ground truth image and improve the performance of an effective forensic detector.

The rest of this paper is organized as follows: Section 2 briefly introduces the state-of-the-art CNNs for ‘image classification’ and ‘semantic segmentation.’ In Section 3, the new scheme of the Copy-Move forgery
detection is proposed. Here, a novelty ground truth image is generated with the properties introduced in Section 2. The experimental results are discussed in Section 4, the comparison of performance evaluation with the generated existing and proposed ground truth images. Lastly, Section 5 draws the conclusion and the future research possibilities presented for the area of image forensics.

II. THEORETICAL BACKGROUND

A. ResNet

ResNet [14], [15] is based on the structure of VGG-19 [20] and is a CNN model that is basically used in YoLo model. It won the 2015 ILSVRC (ImageNet Large Scale Visual Recognition Challenge) and was developed by Microsoft. Before ResNet, net models tried to increase performance by stacking many deep layers, but the ResNet author found the Vanishing Gradient Problem (VGP) out of optimization, as shown in Fig. 2(a). This phenomenon in which the influence of gradient on backpropagation rapidly decreases as the model deepens.

Fig. 2(b) shows that a residual block called bottleneck architecture was used to solve this problem.

In this module, the input \( x \) is a shortcut structure added directly to the output ReLu \( F(x) + x \), and the VGP in Fig. 2(a) is solved when the number of layers increases, the input value is forgotten.

And in the version with more than 50 layers, the bottleneck skips connection structure, as shown in Fig. 2(c), is used.

B. MobileNetv2

Google developed MobileNetV2 [16]. Fig. 3(a) is the residual block structure of ResNet, and (b) is the inverted residual block of MobileNetV2.

In Fig 3,(a) is wide \( \rightarrow \) narrow \( \rightarrow \) wide type, and narrow makes a bottleneck, whereas (b) is narrow \( \rightarrow \) wide \( \rightarrow \) narrow type. The first feature is a linear bottleneck, and it is added to the last skip connection without going through ReLU. The necessary information is narrow and passed to deeper layers using the skip connection. Also, because skip connection is narrow, memory usage is reduced.

C. XCEPTION

“Extreme Inception,” [17] a powerful version of the Inception Architecture, is called ‘Xception’. If the size of the network is increased to improve the performance of the deep neural network, 1) overfitting occurs, and 2) computational resources increase. GoogLeNet [21] uses the Inception module as shown in Fig. 4 to solve this problem. Rhee [22] proposed a network model for Cut-Paste forgery detection as an application thereof.

D. INCEPTIONRESNETV2

As a model that surpassed VGG and won 1st place in ILSVRC 2014, a thesis was published under the name of Inception [18], and one of several versions of Inception is GoogLeNet [21]. Fig. 5(a) and (b) present the modules of the original and factorizing module types. Inception v2 has 2.5 times more training parameters than Google Net but is more efficient than VGG.
E. DEEPLAB V3PLUS STRUCTURE
When generating DeepLab v3plus networks [19] using Xception or Mobilenetv2 primary networks, separable depth convolutions are used in Atrous Spatial Pyramid Pooling (ASPP) and decoder subnetworks [23].

Convolutional layers are used for all other primary networks.

This DeepLab v3plus implementation does not include a global average pooling layer in ASPP but transforms it to a CNN for semantic image segmentation.

The network uses an encoder-decoder architecture, extended convolution. To use a CNN network for semantic segmentation, retrain the segment region by giving it a class.

FIGURE 6. Encoder-decoder with atrous convolution [23].

F. SEMANTIC SEGMENTATION
Semantic segmentation [24] segments objects in an image into meaningful units. More specifically, it predicts which class each pixel in an image belongs to. This task is sometimes called dense prediction because prediction is performed on all pixels in the image.

If semantic segmentation applies to an image, then know which class each pixel belongs to. That is, the input value of the semantic segmentation algorithm is a color image or a black and white image, and the output value is a segmentation map representing the predicted class of each pixel as in Fig. 7 [25], the semantic labels {1: person, 2: purse, 3: plants/grass, 4: sidewalk, and 5: building/structures}.

FIGURE 7. Semantic segmentation [24].

G. BUSTERNET
Wu et al. proposed a BusterNet [9] with a double-branched DNN structure based on the VGG16 [10] network as shown in Fig. 8, and it is composed of Mani-Det and Simi-Det.

The feature extraction modules of BusterNet use the VGG16 network, and the mask decoder module uses 4 pairs of BN-Inception [9] and performs bilinear upsampling. A correlation module is added to the Simi-Det branch to detect copy patches. This module consists of an autocorrelation layer, a percentile pooling layer, and a batch normalization layer.

FIGURE 8. BusterNet [9].

Mani-Det detects forged area (moved patch), and Simi-Det detects similar patch (copy patch). These two patches are defined into copy and move regions in the distinction map. BusterNet has two drawbacks [11]. The one should ensure that both branches correctly locate regions, and the other one, the Simi-Det branch, only extracts single-level and low-resolution features due to the four pooling layers in VGG16. Therefore, the detected forged area may not be correct.

H. SOURCE/TARGET DISTINGUISHMENT
Chen et al. proposed the Source/Target Distinguishment [11] as shown in Fig. 9 to compensate for the shortcomings of BusterNet. The parallel scheme BusterNet should ensure that each branch locates regions correctly. In contrast, this structure has the CMSDNet (copy-move similarity detection network) and the STRDNet (source/target region distinction network) connected in series. A detection map is inserted between them. That comprises three modules: the feature extraction module, correlation module, and mask decoder module. The peculiarity is that instead of the fourth polling layer of VGG16, an atrous convolution is used to preserve ‘Field-of-views of filters.

FIGURE 9. Source/target distinction [11].

III. PROPOSED COPY-MOVE FORGERY DETECTION SCHEME
A. CONFIGURE TO SEMANTIC SEGMENTATION OF NETWORK STRUCTURE
In this work, a novelty ground truth image to detect a Copy-Move region generated using the deep neural of the state-of-the-art CNN networks {Resnet50 [14], Resnet18 [15], Mobilenetv2 [16], Xception [17], Inceptionresnetv2 [18], and DeepLab v3plus structure [19]} that have excellent performance in ‘image classification’ and ‘semantic segmentation’ were developed, respectively. In Fig. 10, [14]–[18] used for ‘image classification’ are converted into that called
‘net_Model n’ (n = 1 ~ 5) in turn through the [19] used for ‘semantic segmentation.’ So, the structures of the net_Models are newly designed and have the characteristics of ‘image classification’ and ‘semantic segmentation,’ both.

In the manipulation area of the forgery image, copy and move areas are Class 1, and the other innocent area is Class 2. In the ground truth image, each class appears as ‘white’ and ‘black’ colors, respectively, shown in Fig. 11.

**C. PROPOSED SCHEME**

It is training to implement net_Models as a detector for Copy-Move detection. The training option used in the proposed scheme adopts SGDM (Stochastic Gradient Descent with Momentum) [35].

Subsequently, to improve the performance of CMFD, the workflow of the scheme proposed in this paper is shown in Fig. 12. The process from the GT generation of forgery image in net_Model (on the upper left) to the decision of the Copy-Move forgery (on the far right), described as follows:

1) The structure of the state-of-the-art five networks [14]–[19] transformed to net_Models (No. 1-5) which generate the ground truth image (GTnet) of the forgery image.
2) All GTnet images are convergence and synthesized into one GTconv image (thick purple line box).
3) Since GTconv depth ranges from 1 to 5, the decomposed images (GTdecomp) are generated by threshold values (thr = 1 to 5).
4) GTnet and original ground truth image (GTorg) become GTorg image by “Crossing Over” operation, which includes true positive (TP), true negative (TN), false positive (FP), and false negative (FN) information to detect the Copy-Move patch, is displayed in ‘white’, ‘black’, ‘red’, and ‘green’ colors, respectively. The procedure is shown in Fig. 13.
5) With GTorg image, Accuracy, Recall, and F1net Score are calculated from TP, TN, FP, and FN.
6) GTdecomp in 3) becomes GTorg image through the same procedure of 4) with GTorg.
7) With GTorg image, Accuracy, Recall, and F1decomp Score are calculated from TP, TN, FP, and FN through the same procedure of 5).
8) By higher value is selected from F1net Score in 5) and F1decomp Score in 7), CMFD is determined.

In Fig. 12, procedure 4) is depicted, also procedure 6) has a working same.

**D. PROPOSED FORGERY DETECTION METHOD**

In Fig. 14, the performance evaluation of Copy-Move forgery detection is the block described in green, and the green line indicates the Accuracy and F1 Score calculation processing.
The general and proposed methods are configured in (a) and (b), respectively.

In Fig. 14(a), the ground truth image generated by some method is overlaid with the forgery image, and then the evaluation is performed to examine the Copy-Move patch area.

If the result is not satisfactory, another method generates a ground truth image. Overlay with the forgery image again and repeat until the evaluation of the detection result is satisfactory.

In Fig. 12 (b), on the other hand, the performance evaluation of copy-move forgery detection in the proposed scheme is directly performed by overlaying the generated ground truth images and the forgery image.

**IV. EVALUATION OF EXPERIMENTAL RESULTS AND PERFORMANCE**

For the experiment of the proposed scheme, an environment is the MATLAB 2021a tool used as simulation software on a PC environment (64bit Win10 Pro, AMD Ryzen9 3950X® 16-Core CPU @3.5GHz, 128GB DDR4 memory, and NVIDIA 2080Ti 11GB Double graphic boards).

**A. NET_MODEL TRAINING EVALUATION**

The execution of each net_Model of the training process configured in Section III-B is shown, which includes {Training Loss, Training Accuracy, and Final Validation Accuracy} in Fig. 15 (a) ~ (e), respectively. These measured variable values are shown {Training time, the number of layers, Final Validation Accuracy} of net_Models, respectively in Table 2.
The Copy-Move forgery images are tested for test preparation in Section III-B on the trained net_Model of the proposed scheme.

TABLE 2. Training results.

| net_Model No. | Training time    | Number of Layers / Connections | Final Validation Accuracy |
|---------------|-----------------|--------------------------------|--------------------------|
| 1             | 16min.57sec.    | 226 / 227                      | 0.9440                   |
| 2             | 11min.22sec.    | 100 / 113                      | 0.9256                   |
| 3             | 43min.07sec.    | 186 / 201                      | 0.9280                   |
| 4             | 26min.13sec.    | 205 / 222                      | 0.9323                   |
| 5             | 145min.16sec.   | 853 / 956                      | 0.9281                   |

For the CMFD performance evaluation of 1) and 3) generated in the proposed scheme as shown in Fig. 12, the GT images generated in procedure 4) and 6) is a paired pattern of the “Crossing Over” operation (Fig. 13), with TP, TN, FP, and FN are displayed as the following color components.

- White area (TP): The same region of Class 1 of the original GT and the generated GT.
- Black area (TN): The same region of Class 2 of the original GT and the generated GT.
- Red area (FP): Copy-Move area in which the generated GT is erroneously detected.
- Green area (FN): The original GT is the Copy-Move area to be detected, but the generated GT is missing.

From procedures 5) and 7), the $F1$ Score is calculated by measurement equations (1) to (7). The meaning of the measurement is described in Table 3.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$
TABLE 3. Meaning: measurement items.

| Measurement | Meaning where Class 1: ‘white’ and Class 2: ‘black’ in GT image. |
|-------------|-------------------------------------------------------------------|
| Accuracy    | In overall predictions, the ratio of correct prediction (regardless, whether the prediction is ‘Class 1’ or ‘Class 2’). |
| Precision   | In the ratio of actual ‘Class 1,’ among the predicted as ‘Class 1’. |
| Recall      | The ratio of predictions of ‘Class 1,’ among actually ‘yes’. |
| Specificity | The ratio of predictions of ‘Class 2,’ among the actually ‘Class 2’. |
| Balanced Accuracy | Average of a Recall and a Specificity. |
| F1 Score    | The harmonize average of ‘Precision’ and ‘Recall’. (It has a larger value when both values are equally large than when only one of ‘Precision’ and ‘Recall’ is large) |
| IoU Score   | Jaccard similarity coefficient: A measure of statistical accuracy that penalizes FP. It is the ratio of the number of GT pixels of each class to the number of predicted pixels. |

\[
\text{Specificity} = \frac{TN}{FP + TN} + \text{Specificity} \tag{4}
\]

\[
\text{Balanced Accuracy} = \frac{2 \cdot \text{Recall} \cdot \text{Accuracy}}{\text{Recall} + \text{Accuracy}} \tag{5}
\]

\[
F1 \text{ Score} = \frac{2 \cdot \text{Accuracy} \cdot \text{Recall}}{\text{Accuracy} + \text{Recall}} \tag{6}
\]

\[
\text{IoU Score} = \frac{TP}{TP + FP + FN} \tag{7}
\]

Table 4 shows the measured items of net_Models according to the meaning of Table 3. net_Model No. (1 to 3) are above the average of each measurement item (blue bold: max.). The values in Table 4 are recorded with the final validation after the training.

TABLE 4. Measured items of net_Models on the training.

| net_Model No. | Measurement Items |
|---------------|-------------------|
|               | Global Accuracy   |
|               | Mean Accuracy     |
|               | Mean IoU          |
|               | Weighted IoU      |
|               | Mean BF Score     |
| 1.            | 0.940             |
| 2.            | 0.926             |
| 3.            | 0.928             |
| 4.            | 0.932             |
| 5.            | 0.928             |

Fig. 16 shows a confusion matrix to classify ‘Class 1’ (white: Copy-Move patch) and ‘Class 2’ (black: innocent region) for the generated GTnet images of the entire test forgery images by each net_Model (a) ∼ (e), see Fig. 11–1). The detection rate of the Copy-Move region is about 70% of ‘Class 1’, which is more than 50% of the class classification criterion, confirming the validity of semantic segmentation.

\[C. \text{ PERFORMANCE EVALUATION OF COPY-MOVE DETECTION ON TESTING}\]

Fig. 17 ∼ 35 shows the results of detecting Copy-Move by randomly selecting 19 images from the Copy-Move forgery image data set in Table 1. Each Figure is an example of a forgery image to which various manipulation methods and environments of Copy-Move are applied, and the description of the images in (a) ∼ (i) of Fig. 17 ∼ 35 is as follows:
Here, the brightness of the class overlay image (e) and (h) is differently displayed to distinguish the generated GT image of net_Model and the proposed scheme. Also, Table 5 shows how the GT generated from which model became the best GT\textit{decomp} image with how much threshold value.

Here, the brightness of the class overlay image (e) and (h) is differently displayed to distinguish the generated GT image of net_Model and the proposed scheme. Also, Table 5 shows how the GT generated from which model became the best GT\textit{decomp} image with how much threshold value.
FIGURE 22. Similar multiple objects.

FIGURE 23. Scale.

FIGURE 24. Blur.

FIGURE 25. Over glass.

FIGURE 26. Multiple moving patches.

FIGURE 27. Similar texture component.
D. COMPARISON OF GT IMAGES BETWEEN RETRANSFORMED NET_MODEL AND PROPOSED SCHEME FOR COPY-MOVE DETECTION

The experimental results were measured with six items \{Accuracy, Balanced Accuracy, Precision, Recall, F1 Score, and Specificity\} in Table 6 (next page). As a result of comparing the generated GT images of each net_Model and the proposed scheme, the performance was somewhat lower than that of net_Model 1; meanwhile, it was superior to the results of...
other net_Model 2 ~ 5 (red), and it was confirmed that the improvement rate was increased (blue bold).

The final validation accuracy in Table 4 is the accuracy of the verification process after net_Model training, and the accuracy in Table 6 is the results of test processing with the trained net_Model. The maximum Accuracy is a value of 0.9523 (red bold on highlight).

Fig. 36 presents the estimated proposed scheme with ROC curves: (a) the GTnet classification of the existing net_Models and (b) the GTcomp classification of the proposed scheme.

The aspect of the whole AUC values confirms that outstanding of the AUCs with a value of 0.9 higher. Thus, the evaluation [36] of the proposed CMFD scheme was graded as ‘Excellent’ (A)’.

In particular, it was also confirmed that all measurement items of net_Model 2 and 3 were ‘excellent’ (red) by increasing the improvement rate.

The measurement results in Table 6 present a graph for visual aid. In Fig. 37, GTnet (a), and GTcomp (b) are shown, respectively, and the improvement rates of each net_Models are shown in (c).

The performance evaluation of the proposed CMFD scheme was compared with the state of the arts [9], [11]
of the pixel-based segmentation classification described in Sections 1 and 2. Table 7 presents the executed results. The experimental environment is executed under the same conditions as in this paper.

It was confirmed that the deep learning network structure of the copy and move region patch classification for CMFD has a higher classification rate with the converged method: (image classification + semantic segmentation) against the structures of a parallel and serial configuration, respectively.

V. CONCLUSION

This paper proposed a new scheme for generating a novelty ground truth image detecting Copy-Moved patches in a forged image. The generated GT involves ‘Image classification’ and ‘Semantic segmentation,’ both in Copy-Move forgery images. According to the information of ‘Image classification’ and ‘Semantic segmentation,’ the best-fit GT image of the test forgery image is generated.

Variety kinds of forgery patches in Copy-Move {such as Rotate, Scale, and Blur, etc. } were included in the data set of the net_model training and were detected well in the test stage.

The proposed GT image of the dedicated CNN structures of excellent image classification and semantic segmentation improved the detection \( \text{Accuracy} \) and \( \text{F1 Score} \) for the Copy-Move matching. Consequently, the proposed scheme
in this paper could be applied more in-depth to the image forensics field (such as Cut-Paste, Inpainting, and Forgery feature extraction, etc.).

The detecting method of the Copy-Move patch region of the proposed scheme can be used quickly in an environment of online JIT (Just In Time). It is necessary to advance research on detecting the multiple classes of moving patches in the future.

ACKNOWLEDGMENT
The author would also like to thank the leading researchers in this study field, the hidden reviewers whose meaningful comments, and the official working of Associate Editor, IEEE Access. With the treatments, this paper could be written thoughtfully.

REFERENCES
[1] K. H. Rhee, “Detection of spliced image forensics using texture analysis of median filter residual,” IEEE Access, vol. 8, pp. 103374–103384, 2020, doi: 10.1109/ACCESS.2020.2999308.
[2] K. H. Rhee, “Forensic detection using bit-planes slicing of median filtering image,” IEEE Access, vol. 7, pp. 92586–92587, 2019, doi: 10.1109/ACCESS.2019.2927540.
[3] C. Wang, Z. Zhang, Q. Li, and X. Zhou, “An image copy-move forgery detection method based on SURF and PCET,” IEEE Access, vol. 7, pp. 170032–170047, 2019, doi: 10.1109/ACCESS.2019.2955308.
[4] H. Chen, X. Yang, and Y. Lyu, “Copy-move forgery detection based on keypoint clustering and similar neighborhood search algorithm,” IEEE Access, vol. 8, pp. 36863–36875, 2020, doi: 10.1109/ACCESS.2020.2974486.
[5] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, “SLIC superpixels compared to state-of-the-art superpixel methods,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 11, pp. 2274–2282, Nov. 2012, doi: 10.1109/TPAMI.2012.120.
[6] X. Pan and S. Lyu, “Region duplication detection using image feature matching,” IEEE Trans. Inf. Forensics Security, vol. 5, no. 4, pp. 857–867, Dec. 2010, doi: 10.1109/TIFS.2010.2078506.
[7] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” Int. J. Comput. Vis., vol. 60, no. 2, pp. 91–110, Nov. 2004.
[8] A. Costanzo, I. Amerini, R. Caldelli, and M. Barni, “Forensic analysis of SIFT keypoint removal and injection,” IEEE Trans. Inf. Forensics Security, vol. 9, no. 9, pp. 1456–1464, Sep. 2014, doi: 10.1109/TIFS.2014.2357654.
[9] Y. Wu, W. Abd-Almageed, and P. Natarajan, “BusterNet: Detecting copy-move image forgery with source/target localization,” in Proc. Eur. Conf. Comput. Vis., 2018, pp. 168–184. [Online]. Available: https://openaccess.thecvf.com/ECCV2018_search
[10] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” 2014, arXiv:1409.1556.
[11] B. Chen, W. Tian, G. Coatrieux, Y. Zheng, and Y.-Q. Shi, “A serial of-the-art technical review and analysis,” IEEE Trans. Inf. Forensics Security, vol. 16, no. 4, pp. 170032–170047, 2019, doi: 10.1109/TIFS.2019.2955308.
[12] K. H. Rhee, “Composition of visual feature vector pattern for deep learning in image forensics,” IEEE Access, vol. 8, pp. 188970–188980, 2020, doi: 10.1109/ACCESS.2020.3029087.
[13] L. Chen, G. Papandreou, F. Schroff, and H. Adam, “Rethinking atrous convolution for semantic image segmentation,” Dec. 2017, arXiv:1706.05587.
[14] Q. Zhao, G. Cao, A. Zou, X. Huang, and L. Yang, “Image tampering detection via semantic segmentation network,” in Proc. 15th IEEE Int. Conf. Signal Process. (ICSP), Dec. 2020, pp. 165–169, doi: 10.1109/ICSP4869.2020.9321086.
[15] J. Jordan. (Aug. 2021). An Overview of Semantic Image Segmentation. [Online]. Available: https://www.jeremyjordan.me/semantic-segmentation/
[16] COVeFoD—Image Database for Copy-Move Forgery Detection. Accessed: Aug. 2021. [Online]. Available: https://www.vcl.fher.unipa.it/comofd/comofd.html
[17] CVIP. Accessed: Aug. 2021. [Online]. Available: http://www.diid.unipa.it/cvip?_page_id=48
[18] GRIFF. Accessed: Aug. 2021. [Online]. Available: http://www.grip.unina.it/research/83-multimedia_forensics/90-copy-move-forgery.html
[19] CASIA V2. Accessed: Nov. 2021. [Online]. Available: https://paperswithcode.com/dataset/casia-v2
[20] COVERAGE. Accessed: Nov. 2021. [Online]. Available: https://paperswithcode.com/dataset/coverage
[21] B. Wen, Y. Zhu, R. Subramanian, T.-T. Ng, X. Shen, and S. Winkler, “COVERAGE—A novel database for copy-move forgery detection,” in Proc. IEEE Int. Conf. Image Process. (ICIP), Phoenix, AZ, USA, Sep. 2016, pp. 161–165, doi: 10.1109/ICIP.2016.7532339.
[22] I. Amerini, L. Ballan, R. Caldelli, A. Del Bimbo, and G. Serra, “A SIFT-based forensic method for copy-move attack detection and transformation recovery,” IEEE Trans. Inf. Forensics Security, vol. 6, no. 3, pp. 1099–1110, Sep. 2011, doi: 10.1109/TIFS.2011.2129512.
[23] CPH. Accessed: Nov. 2021. [Online]. Available: https://paperswithcode.com/dataset/cph
[24] M. Xu, T. Li, Z. Wang, X. Deng, R. Yang, and Z. Guan, “Reducing complexity of HEVC: A deep learning approach,” IEEE Trans. Image Process., vol. 27, no. 10, pp. 5044–5059, Oct. 2018, doi: 10.1109/TIP.2018.2847035.
[25] R. Pascuan, T. Mikolov, and Y. Bengio, “On the difficulty of training recurrent neural networks,” Proc. 30th Int. Conf. Mach. Learn., 2013, vol. 28, no. 3, pp. 1310–1318. [Online]. Available: http://proceedings.neurips.cc/paper/2013/pascuan13.pdf
[26] T. G. Tape. (2021). The Area Under an ROC Curve. [Online]. Available: http://gim.unmc.edu/dxtests/roc3.htm

KANG HYEON RHEE received the B.S. and M.S. degrees in electronics engineering from Chosun University, in 1977 and 1981, respectively, and the Ph.D. degree in electronics engineering from Ajou University, Suwon, South Korea, in 1991. He is currently a Professor Emeritus with the School of Electronics Engineering, Chosun University, Gwangju, South Korea. His current research interests include multimedia fingerprinting/forensics. He was a recipient of awards, such as the Haedong Prize from the Haedong Science and Culture Juridical Foundation, South Korea, which he received in 2002 and 2009, respectively.