

RESEARCH ARTICLE

CNN-Based Copy-Move Forgery Detection Using Rotation-Invariant Wavelet Feature

SANG IN LEE, JUN YOUNG PARK, AND IL KYU EOM 

Department of Electronics Engineering, Pusan National University, Pusan 46241, South Korea

Corresponding author: Il Kyu Eom (ikeom@pusan.ac.kr)

This work was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science, and Technology, under Grant NRF-2018R1D1A1B07046213.

ABSTRACT This paper introduces a machine learning based copy-move forgery (CMF) localization method. The basic convolutional neural network cannot be applied to CMF detection because CMF frequently involves rotation transformation. Therefore, we propose a rotation-invariant feature based on the root-mean squared energy using high-frequency wavelet coefficients. Instead of using three color image channels, two-scale energy features and low-frequency subband image are fed into the conventional VGG16 network. A correlation module is used by employing small feature patches generated by the VGG16 network to obtain the possible copied and moved patch pairs. The all-to-all similarity score is computed using the correlation module. To generate the final binary localization map, a simplified mask decoder module is introduced, which is composed of two simple bilinear upsampling and two batch-normalized-inception-based mask deconvolution followed by bilinear upsampling. We perform experiments on four test datasets and compare the proposed method with state-of-the-art tampering localization methods. The results demonstrate that the proposed scheme outperforms the existing approaches.

INDEX TERMS Copy-move forgery, copy-move forgery localization, convolutional neural network, rotation-invariant, stationary wavelet transform, root-mean squared energy, simplified mask decoder module.

I. INTRODUCTION

Images are often used as important evidence to clarify events. However, the development of various image editing tools has allowed some persons to easily manipulate images. Furthermore, the use of manipulated images for malicious purposes can cause negative effects on human society. The authenticity of an image can become suspicious, thus, the determination of the authenticity of an image has emerged as an important issue. Because human eyes cannot easily detect forged image, we need to develop reliable image forgery detection methods. Extensive studies have been conducted on the detection of various image forgeries [1], [2], [3], [4].

A commonly used image tampering method is the copy-move forgery (CMF) in which part of an image is copied from one section of the image and is pasted elsewhere in the same image. An image can be forged to conceal or change its


The associate editor coordinating the review of this manuscript and approving it for publication was Vicente Alarcon-Aquino .



FIGURE 1. Typical example of a copy-move forgery.

meaning using the copy-move process. Therefore, verifying the reliability of the image and localizing the copied and moved areas are important. Fig. 1 shows a typical example of CMF where determining the CMF using naked eyes is difficult. Therefore, the development of a reliable CMF

detection (CMFD) method has become an important issue. In general, the copied part of the image is usually scaled or rotated before move process. Therefore, visual inspection alone cannot readily verify the authenticity of the image. The aim of CMFD is to accurately detect or localize the copy-moved area even if various operations, including scale and rotation, are applied before the image is manipulated.

The CMFD method is roughly classified into three categories: block-based, keypoint-based, and machine learning-based approaches. In the block-based approach, various block division and segmentation algorithms are used, followed by scale- and rotation-invariant feature extraction. Finally, matching is performed to find the copied and moved regions. Keypoint-based CMFD algorithms have attracted a much attention in recent years. Scale invariant feature transform (SIFT) is one of the most frequently used keypoint extraction methods for CMFD. Recently, a wide range of machine learning-based CMFD scheme has been introduced, which demonstrated promising detection results. In particular, the convolutional neural network (CNN) is drawing a lot of attention to CMFD. Because CNN shows good performance in the field of object detection, it is suitable for CMFD to find copied-moved objects. More detailed CMFD methods are reviewed and presented in the next section.

CMF frequently involves scale and rotation transformation. However, the basic CNN is not suitable for CMFD because it is not known to be invariant to rotation [5]. To overcome the shortcomings of the non-rotation-invariant property, most of the previous CNN-based CMFD studies focus only on structural changes to increase the accuracy of detection. The present study introduces a novel CMFD scheme using CNN and a wavelet domain energy feature. To provide a rotation-invariant property to the conventional CNN, the present work uses the energy feature of high-frequency wavelet coefficients. The proposed rotation-invariant feature can improve the detection performance for CMF images with rotational transformation and resizing.

The proposed CMFD network comprises four modules: rotation-invariant feature module based on the stationary wavelet transform (SWT), feature extraction module using CNN, correlation module to check similarity, and mask decoder module to generate a binary detection map for CMFD. The proposed network is robust to scale and rotation in CMFD using low-frequency and high-frequency channels in the wavelet domain instead of RGB channels as CNN input. The simulation results show that the proposed method generates superior copy-move localization results compared with the existing methods.

The remainder of this paper is organized as follows. Related works are briefly reviewed in Section II. The proposed CMFD network is presented in Section III. In Section IV, the performance of the proposed method is compared with that of existing methods using the experimental results. Section V provides the conclusion.

II. RELATED WORKS

A. BLOCK-BASED METHOD

The block-based CMFD methods mainly involve four steps: dividing a suspicious image into blocks, extracting features, matching features in the divided blocks, and localizing the forged regions. In the first step, various block division and segmentation methods can be used in preprocessing. An image can be divided into overlapping square blocks [6], [7], non-overlapping square blocks [8], or circular blocks [9], [10]. Image segmentation techniques [11], [12] are usually included in this step to separate the copied source region from the pasted target region.

Feature extraction is the main step in block-based CMFD algorithms. The extracted features must be invariant against scale and rotation, and must be robust against blurring, sharpening, background adjustment, compression, and noise addition when subjected to postprocessing steps. Transform-based methods are frequently used to remove information that is unnecessary for detection. A variety of transform methods, such as the polar cosine transform [13], Fourier-Mellin transform [14], and polar complex exponential transform [15] have been used for forgery detection together with the popular Fourier, discrete cosine [16], and various wavelet [17], [18] transforms. In addition to these transforms, histogram-based techniques [19] and statistical moment-based methods [20], [21] have also been introduced.

Feature matching determines a candidate pair of the original part and the corresponding copied-moved part using the extracted features. This step employs searching and similarity measurement techniques. The searching methods involve various sorting [18], [22] and hashing processes [12]. These algorithms usually involve the use of dimension reduction techniques. During the searching process, the matching methods search for possible matches to evaluate the similarity among the selected possible matches. The Euclidean distance is the most popular and simple similarity metric that is employed. Additionally, the Manhattan and Hamming distances between two features have also been employed to determine similarity.

The final step in the CMFD process is localization. The detection output can be visualized as a binary image that illustrates the detected copied-moved regions and their corresponding authentic image parts in the target image.

Most block-based methods suffer from high computational complexity because of a large number of features and lack robustness to geometric transformation attacks. Moreover, they are not invariant to various manipulations such as flipping, shift and blur.

B. KEYPOINT-BASED METHOD

Keypoint-based approaches for CMFD have been actively investigated. SIFT is the most frequently used keypoint extraction method. Because SIFT is robust against scaling, rotation, and occlusion, it is well suited for

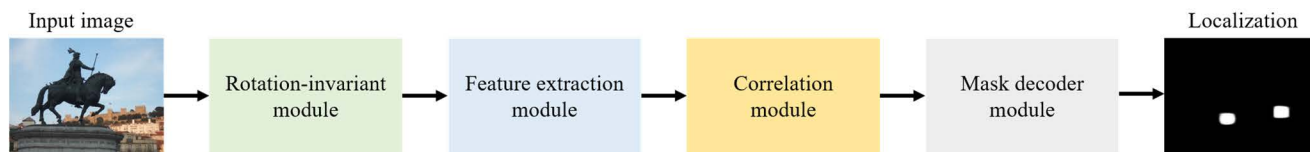


FIGURE 2. Architecture of the proposed CMFD network.

CMFD [23], [24], [25], [26]. First, SIFT generates both scale- and rotation-invariant keypoints, which are extracted at different scales using a scale-space representation by an image pyramid. Next, an orientation is assigned to each keypoint to achieve invariance in the image rotation. Finally, each keypoint is commonly represented by a 128-dimensional descriptor.

In the matching process, the keypoint descriptor of a pixel is compared with all other descriptors other than that pixel. However, false matching almost always occurs after completion of the matching process. Therefore, a wide range of methods have been proposed to eliminate false matches. Mismatched keypoint pairs are eliminated using various clustering algorithms, such as J-lineage [27], distance-based [28], and hierarchical [29] clustering. The random sample consensus algorithm [23] is the most frequently used algorithm to eliminate false matches alone or together with other false matching removal methods.

The performance of SIFT-based algorithms can be degraded when the CMF implementation involves small or smooth regions [25], [26]. Additionally, when a keypoint is present close to the boundary of the copied-moved portion and authentic region, or if an image is compressed after CMF, the keypoint descriptor of the copied portion is different from that of the moved portion because the conventional SIFT descriptor generation method only provides local information about a single keypoint. Therefore, the SIFT-based methods cannot obtain global information around the keypoint. Thus, coping with pixel changes, such as compression or differences in the background area due to the copy-move process can be difficult.

C. MACHINE LEARNING-BASED METHOD

Recently, machine learning has achieved breakthrough performance in image processing and computer vision tasks. Machine learning-based CMFD approaches have also been actively investigated.

In 2017, Bunk et al. [30] introduced a manipulated image detection and localization method using both CNN and long short-term memory-based network. This method used CNN as a patch classifier. Liu et al. [31] proposed the utilization of CNN to perform CMFD. A segmentation-based keypoint distribution strategy was proposed to generate homogeneous distributed keypoints and an adaptive oversegmentation method was adopted in their approach. This method slightly improves detection performance, however, it requires a high computational cost.

Wu et al. [32] proposed a dual-branch CNN scheme (BusterNet) that included Simi-Det and Mani-Det. The Simi-Det branch was designed for similarity detection, and the Mani-Det branch was employed for manipulation detection. After the two branches were implemented, both features were integrated to predict the pixel-level three-class results that contained the untampered, tampered, and untampered background regions. Each branch used VGG16 [33] in the feature extraction, and a mask decoder module based on BN-Inception [34]. However, when each BusterNet branch failed to accurately locate the regions, BusterNet could not distinguish between the source and target regions.

Zhong et al. [35] proposed a CMFD scheme using Dense-InceptionNet, which was an end-to-end, multidimensional dense-feature connection network. Dense-InceptionNet consisted of pyramid feature extractor, feature correlation matching, and hierarchical postprocessing modules. The detection accuracy of this network can be improved using multiple modules that perform similar roles. However, the number of training parameters may increase rapidly.

Zhu et al. [36] proposed an adaptive attention and residual refinement network for CMFD. They used position and channel attention features together with an adaptive attention mechanism to fully capture the context information, and adopted deep matching to compute the self-correlation among feature maps. The atrous spatial pyramid pooling was used to fuse the scaled correlation maps to generate a coarse mask. Finally, the coarse mask was optimized using the residual refinement module, which retained the structure of the object boundaries. They demonstrated that the proposed network exhibited strong robustness against noise, blur, and JPEG recompression during the postprocessing operations. However, the computational cost is high.

Chen et al. [37] have recently introduced a serial CMF localization scheme to serialize the similarity and manipulation detection branches. They used atrous convolution [38] in the final convolution layer of VGG16 and the double-level self-correlation in the correlation module for hierarchical feature comparisons. In this network, atrous spatial pyramid pooling and attention mechanism were proposed to capture the multiscale features. Finally, image-level network was employed to directly determine the regions obtained from the similarity detection branch whether tampered or untampered. The experimental results showed that this algorithm achieved superior performance over BusterNet. However, the performance of this algorithm remains to be improved.

Barni et al. [39] proposed a copy move source-target disambiguation method to identify the source and target regions of a copy-move forgery. They designed a multi-branch CNN architecture that solves the hypothesis testing problem by learning a set of features capable to reveal the presence of interpolation artefacts and boundary inconsistencies in the copy-moved area. The purpose of this network is to distinguish between target and source regions, not to find the copy-move regions through the original image and ground truth. Therefore, if only the input image is given without ground truth, the detection performance may be decreased.

Liu et al. [40] proposed a two-stage framework for copy-move forgery detection. The first stage is a backbone self-deephave matching network to enrich spatial information and leverage hierarchical features, and the second stage is Proposal SuperGlue to remove false-alarmed regions and remedy incomplete regions. However, post-processing has significant impact on forgery detection performance.

III. PROPOSED NETWORK

The proposed network for CMFD comprises four modules: rotation-invariant feature, feature extraction, correlation, and mask decoder modules. The architecture of the proposed network is shown in Fig. 2.

In the rotation-invariant module, the root-mean squared energy using high-frequency wavelet coefficients are used to provide rotation invariant features to the feature extraction module. In the proposed network, VGG16 [33] is used for the feature extraction module. VGG16 can extract image information of various scales by applying the 3×3 filter several times to provide different filter effects. Therefore, VGG16 is frequently applied to machine learning-based CMFD networks that must be robust to scale transformation. We exploit the correlation module used in Wu et al.' work [32] to check similarity of features generated by the feature extraction module. Finally, we propose a simplified mask decoder module to eliminate incorrectly detected regions. The details of each module are described in the next sections.

A. ROTATION-INVARIANT MODULE

Conventional CNN can obtain the scale-invariant characteristics by viewing images on various scales because the convolution operation is performed using filters with various sizes. However, rotation-invariance cannot be derived from the basic CNN structure. To address this problem, various attempts have been made to create CNN rotation-invariance [41], [42]. A copied part of an image is frequently scaled or rotated before the move operation in the CMF process. Therefore, rotation-invariance is essential in CMFD. In this paper, we employ the wavelet transform to provide rotation-invariance to CNN.

For given low-frequency subband image $\mathbf{W}_{i,l}$ at scale i , we perform SWT as follows.

$$\mathbf{W}_{i+1,o} = \text{SWT}(\mathbf{W}_{i,l}), \quad (1)$$

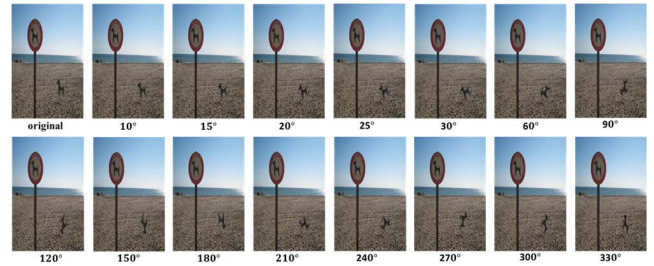


FIGURE 3. Sample copied-moved image and rotated images at various rotation angles.

where $\text{SWT}(\mathbf{z})$ is the undecimated wavelet transform of subband \mathbf{z} , $i (\in 0, 1, 2, \dots)$ is the scale of the wavelet subband, and $o (\in l, h, v, d)$ is the subband direction. h , v , and d present high-frequency subbands with horizontal, vertical, and diagonal directions, respectively. The suspicious input image is represented by $\mathbf{W}_{0,l}$. Scale i increases by one from zero each time the transform processes, which results in four lower resolution low-frequency and high-frequency subbands. In this paper, $\mathbf{W}_{1,l}$ is selected as a feature that represents the input of the feature extraction module. Because $\mathbf{W}_{1,l}$ is a low-pass version of the image, it can remove small noise in the image.

Even if a rotation operation is applied to an object in the image, we assume that the high-frequency energy does not change. According to this assumption, we propose a high-frequency energy feature to generate rotation-invariant feature. Let $W_{i,o}(x, y)$ be the wavelet coefficient at spatial location (x, y) . The root-mean squared energy is expressed as follows.

$$E_i(x, y) = \sqrt{\frac{W_{i,h}^2(x, y) + W_{i,v}^2(x, y) + W_{i,d}^2(x, y)}{3}}, \quad (2)$$

where $E_i(x, y)$ is the root-mean squared energy with scale i and location (x, y) . In this paper, a bold symbol is used for an image, and an italic symbol with spatial location is used for a single pixel value or a wavelet coefficient.

To verify the effectiveness of the proposed energy feature, we define average energy ratio $r_i(\theta)$ as

$$r_i(\theta) = \frac{\sum_{\text{all}(x,y)} E_{i,\theta}(x, y)}{\sum_{\text{all}(x,y)} E_i(x, y)}, \quad (3)$$

where $E_{i,\theta}(x, y)$ is the root-mean squared energy of the image with the objects rotated by θ . Fig. 3 shows a sample copy-move forged image and its rotated versions at various rotation angles. As shown in Fig. 3, an object is rotated at 15 different angles. Table 1 lists the $r_1(\theta)$ and $r_2(\theta)$ values, which indicate that all values of $r_1(\theta)$ and $r_2(\theta)$ are very close to one. This result indicates that the proposed root-mean squared energy value is almost maintained even when the object is rotated. In conclusion, $E_i(x, y)$ can be used as an input with rotation-invariant characteristics before being trained in the feature extraction module composed of CNNs.

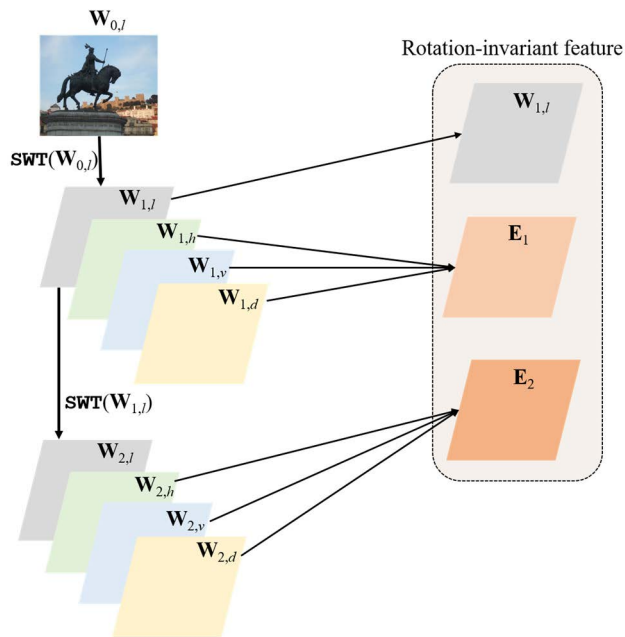


FIGURE 4. Overview of the proposed rotation-invariant module based on SWT.

Fig. 4 shows an overview of the proposed rotation-invariant module based on SWT. Before the rotation-invariant feature is extracted, all images are resized to 256×256 sized image. The filter kernel of SWT is the simplest Haar kernel, and the decomposition level of the wavelet transform is two. Instead of using RGB image channels, $W_{1,l}$, E_1 , and E_2 are used as input to the next CNN module in this paper.

TABLE 1. $r_1(\theta)$ and $r_2(\theta)$ values at various rotation angles of the sample images shown in Fig. 3.

Rotation angle θ (unit: degree)	$r_1(\theta)$	$r_2(\theta)$
10	0.9998	0.9997
15	0.9994	0.9989
20	0.9986	0.9979
25	0.9986	0.9978
30	0.9998	0.9978
60	0.9998	0.9981
90	1.0000	1.0006
120	1.0024	1.0002
150	0.9998	0.9992
180	1.0017	1.0011
210	0.9998	0.9983
240	1.0009	0.9992
270	1.0015	1.0006
300	1.0000	0.9998
330	1.0020	1.0009

B. FEATURE EXTRACTION MODULE BASED ON VGG16

VGG16 [33] has been successfully applied as a feature extractor in visual object classification, recognition and many other fields. BusterNet [32] uses VGG16 for CMF localization, and a serialized CMFD network [37] uses modified VGG16, which replaces the last convolution layer to atrous

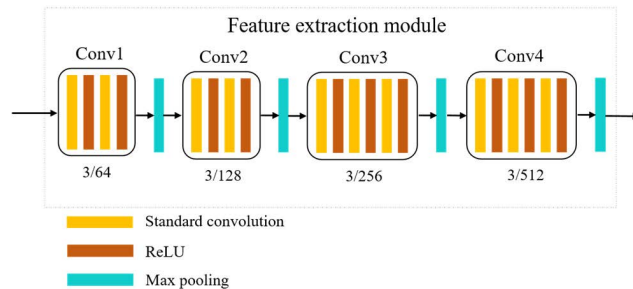


FIGURE 5. Basic VGG16 network as feature extraction module.

convolution [38]. Atrous convolution is used to enlarge the field-of views of filters. However, because these two networks do not consider the rotation-invariant feature, the localization performance is reduced for copied-moved images with rotations.

The proposed wavelet-based input contains rotation-invariance and a large field-of view. Hence, the basic VGG16 structure is sufficient to extract features for CMFD. VGG16 that is used in this paper contains four standard convolution groups as shown in Fig. 5. The kernel size of all standard convolution layers is 3×3 . The numbers of kernels in the four groups are 64, 128, 256, and 512. Each standard convolution possesses a ReLU activation function, and each group is followed by a max-pooling layer.

C. CORRELATION MODULE

The feature extraction module based on VGG16 generates 16×16 feature patches that amount to 512. To obtain the potential copied and moved patch pairs, the all-to-all feature similarity score is computed using the correlation module. The correlation module comprises self-correlation, percentile pooling, and batch normalization blocks. Fig. 6 presents the three blocks of the correlation module. This correlation module type is used in BusterNet [32]. The network of Chen et al. [37] modified the correlation module by adding a channel attention module.

For two given patches, the Pearson correlation coefficient, which quantifies the feature similarity, is used in the self-correlation block. The self-correlation block outputs 256 patches with a size of 16×16 because this block matches the two similar patches with the correlation coefficient. A larger correlation coefficient indicates a more similar pair of patches. The percentile pooling block sorts the similarity score in a descending order. The sorted similarity score contains sufficient information to determine what feature is matched in the next stage. Before moving on to the next module, batch normalization is performed in the batch normalization block. The self-correlation and percentile pooling blocks contain no trainable parameters.

D. MASK DECODER MODULE

The correlation module generates a 16×16 feature block whose resolution is lower than that of the input image.

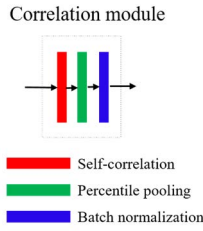


FIGURE 6. Architecture of correlation module.

Therefore, a decoding process that applies deconvolution to restore the original resolution is needed. In BusterNet, the BN-Inception-based mask deconvolution module, which is followed by bilinear upsampling [34], is used. Because the output resolution of the correlation module is 16×16 , four successive BN-Inception and bilinear upsampling blocks are required. To develop a pixel-level localization map, a single standard convolution layer followed by a sigmoid activation function is used.

Fig. 7(a) shows the mask decoder module used in BusterNet. The BN-Inception block consists of three convolution layers with $s_1, s_2,$ and s_3/n , where $s_1, s_2,$ and s_3 denote the kernel sizes, and n denotes the number of filters. In the BusterNet mask decoder module, the parametric BN-Inception block is used. In this paper, we use a simplified mask decoder module by removing the BN-Inception blocks in small-sized data. In the proposed network, two BN-Inception blocks of 16×16 and 32×32 patches are removed as shown in Fig. 7(b). Only two successive bilinear upsampling blocks are used to obtain a 64×64 patch.

To investigate the effect of the simplified mask decoder module, we compare the detection performance of the proposed network applying the mask decoder module used in BusterNet and the simplified module. Fig. 8 shows the comparison between the conventional and proposed simplified mask decoders. As shown in Figure 8, the small spots are eliminated with the use of the simplified mask decoder. However, the copied and moved regions are also reduced. The simplified mask decoder not only eliminates incorrectly detected regions, but also reduces the properly identified regions. The localization performance can be improved using the simplified mask decoder module.

E. SUMMARY OF PROPOSED NETWORK

A given color image is converted into a grayscale image. Next, two-level SWT is performed on this image. At each level, the root-mean squared energy is calculated using (3). In addition to the low-frequency image in the first level, two root-mean squared energy sources are fed into the feature extraction module, which is composed of four convolution groups. The output of each standard convolution layer passes through ReLU followed by a max-pooling layer. The output features of the VGG16 network are input to the correlation module. After self-correlation and percentile pooling,

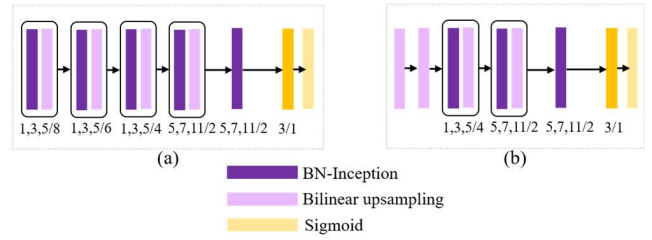


FIGURE 7. (a) Mask decoder module used in BusterNet, (b) simplified mask decoder module in the proposed network.

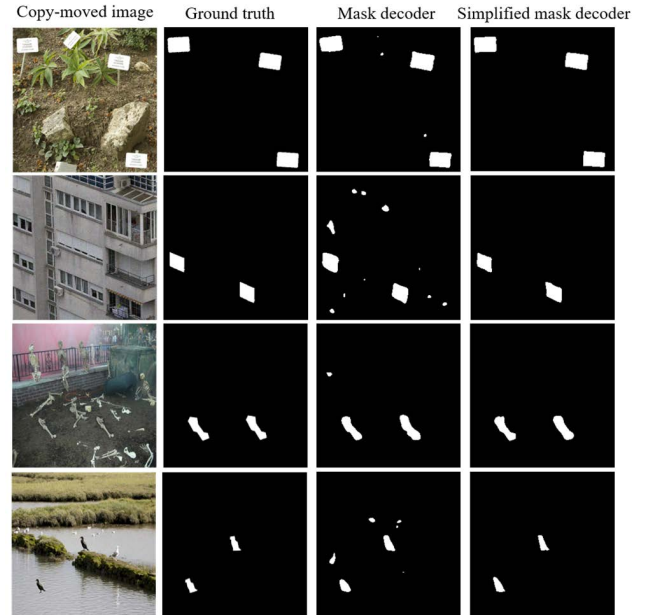


FIGURE 8. Comparison between mask decoder used in BusterNet and simplified mask decoder.

similar features are matched and normalized. In the mask decoder module, simple bilinear upsampling is performed twice, and the combination of upsampling and BN-Inception is followed by upsampling block. Finally, the upsampled 256×256 feature block passes through the BN-Inception net once more. The kernel sizes of the first BN-Inception net are $1 \times 1, 3 \times 3,$ and 5×5 , and the next kernel sizes of the second BN-Inception net are $5 \times 5, 7 \times 7,$ and 11×11 . The detection map is obtained by passing through the 1×1 convolution layer followed by the sigmoid activation function.

In our network, the binary cross-entropy loss [41] is used as follows.

$$L = -\frac{1}{m} \sum_{i=1}^m [y_i^t \log(y_i) + (1 - y_i^t) \log(1 - y_i)], \quad (4)$$

where L denotes the loss function, m is the number of training images, y_i^t is the true map of the i -th image, and y_i is the output of the network of the i -th image.

IV. EXPERIMENTAL RESULTS

A. DATASETS

In our experiment, we use four public datasets with ground truths, namely, CoMoFoD [44], MICC-F2000 [24], D [45], and COVERAGE [46]. The detailed information of these four datasets is listed in Table 2. Every tampered image has a corresponding ground truth. All images are divided into three categories, namely, training, validation, and test, at a ratio of 7:1.5:1.5, respectively. For reliable experimentation of the limited number of data, cross-validation is performed 10 times at random, and the final result is averaged and evaluated.

TABLE 2. Detailed information of four datasets.

Dataset	Description
CoMoPoD	5000 tampered images with resolution of 512×512
MICC-F2000	700 tampered images with resolution of 2048×1536
D	970 tampered images with resolution of 700×1000
COVERAGE	100 tampered images with resolution of 400×486

B. IMPLEMENTATION DETAILS

The proposed network is implemented using Keras in both training and testing. The Keras default function is used to initialize the parameters of all layers. We use the Adam optimizer with a learning rate of 0.01 and the binary cross-entropy loss function. The epoch and mini-batch sizes are set to 250 and 30, respectively. The detection algorithm is implemented on a 12-GB GeForce RTX 3080 Ti, Intel i7-11700K CPU @ 3.70 GHz with 64-GB RAM.

C. PERFORMANCE EVALUATION MEASURES

To evaluate the performance of the proposed method, we first define three measures at the pixel level. Let T_P , F_P , and F_N be the numbers of correctly detected, erroneously detected as forged, and falsely missed forged pixels, respectively. *Precision* denotes the probability that a detected forgery is truly forged. It is expressed as

$$Precision = \frac{T_P}{T_P + F_P}. \quad (5)$$

Recall denotes the probability that a forgery is detected, which is expressed as

$$Recall = \frac{T_P}{T_P + F_N}. \quad (6)$$

The overall detection performance, namely, *F*-measure is defined by the harmonic mean of *Precision* and *Recall* as follows.

$$F = 2 \frac{Precision \cdot Recall}{Precision + Recall}. \quad (7)$$

D. SIMULAITON RESULTS

To evaluate the performance of the proposed CMFD network, three SIFT-based algorithms are compared: SIFT and feature matching (SIFT + FM) [23], SIFT and adaptive segmentation

(SIFT + Seg) [47], and SIFT and reduced local binary pattern (SIFT + RLBP) [26]. Further, two machine learning-based networks are selected for comparison with the proposed network: BusterNet [32] and the serialized network (SeNet) [37]. The codes for the compared CMFD methods are downloaded from their respective project sites.

1) RUNNING TIME

To compare the computational costs, we measured the execution times for all CMFD methods and the number of parameters for CNN-based methods. The comparison methods were implemented in the same environment. The execution time was averaged over 20 operations.

Table 3 shows the execution times for test process and the number of parameters of CNN-based methods. As shown in Table 3, the proposed network achieves the fastest average testing time, and SeNet has the second-fastest testing times. Three SIFT-based methods have high computational costs. The number of parameters of the proposed network, BusterNet, and SeNet are 7,691,569, 9,709,277, and 15,526,813, respectively. In conclusion, it can be seen that the proposed CMFD network can be performed at high speed with a relatively small number of parameters.

TABLE 3. Comparison of the computational performance.

Method	Platform	Testing time	Number of parameters
SIFT+FM [23]	Matlab	10.97	-
SIFT+Seg [47]	Matlab	6.68	-
SIFT+RLBP [26]	Matlab	12.74	-
BusterNet [32]	Python	2.15	15,526,913
SeNet [37]	Python	1.83	9,709,277
Proposed	Python	1.70	7,691,569

2) DETECTION RESULTS

In this work, we divide the test images into two groups of forged images with and without a rotation attack to evaluate the effect of the proposed rotation-invariant wavelet energy feature. Evaluation of this classified image can provide a criterion for evaluating the performance of existing methods against a rotational attack.

TABLE 4. Comparison of the detection performance of the test image without rotation attack according to the pixel-level *Precision*, *Recall*, and *F* measures.

Method	<i>Precision</i>	<i>Recall</i>	<i>F</i>
SIFT+FM [23]	0.428	0.205	0.256
SIFT+Seg [47]	0.757	0.283	0.372
SIFT+RLBP [26]	0.526	0.662	0.556
BusterNet [32]	0.472	0.602	0.482
SeNet [37]	0.589	0.562	0.535
Proposed	0.955	0.872	0.905

Table 4 lists the three evaluation measures for forged images without a rotation attack. As shown in Table 4, the proposed method achieves the best performance. The SIFT-based algorithms, such as SIFT + FM and SIFT + Seg,

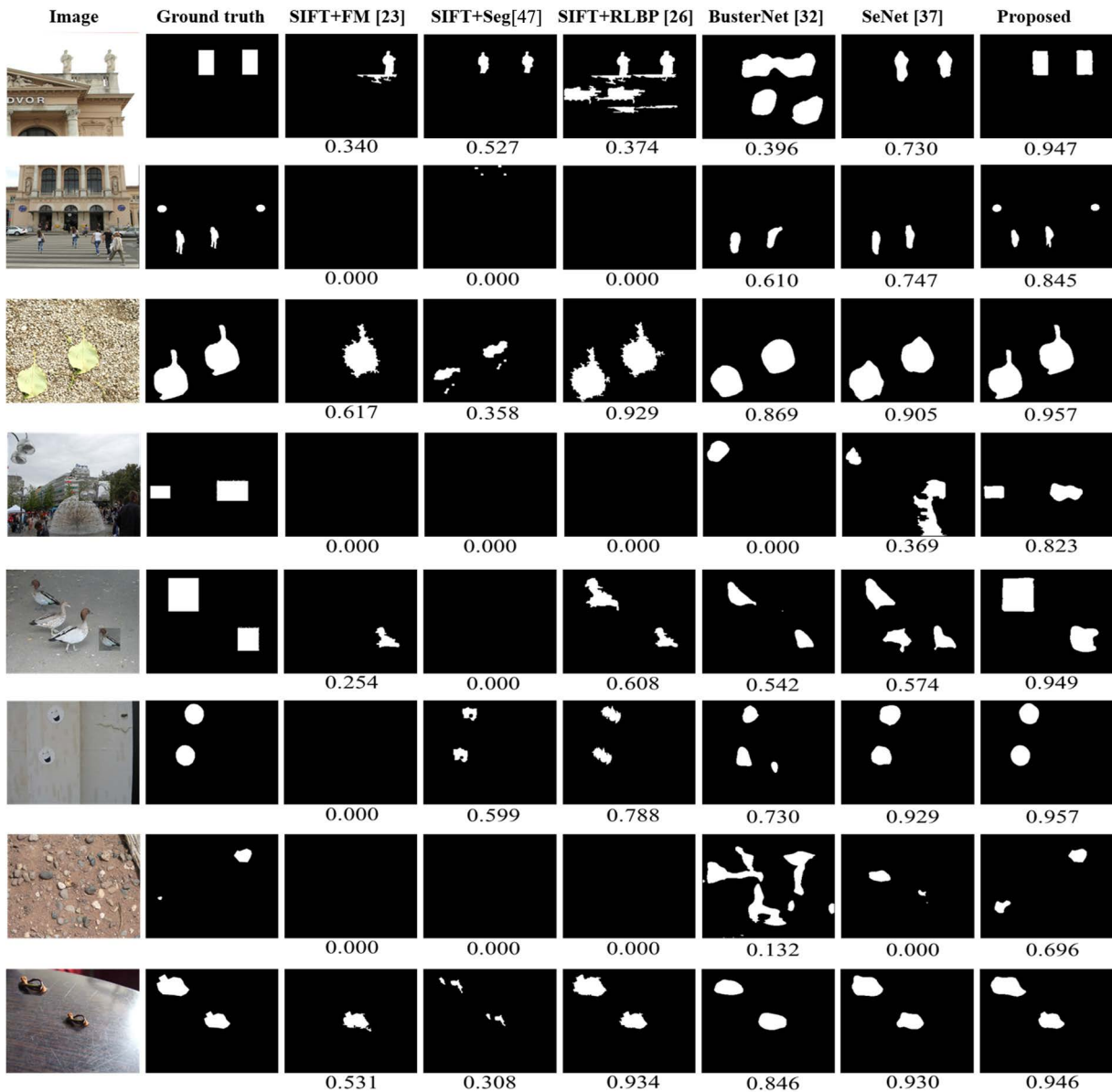


FIGURE 9. Comparison of the forgery detection results and their corresponding F -measure values for the test images without a rotation attack.

largely differ in terms of the *Precision* and *Recall* values. The value of the *Precision* is large, while that of *Recall* is very low because the forged region is found to be excessive in these two methods. On the other hand, the recently reported SIFT + RLBP method demonstrates a balance between the *Precision* and *Recall* values, and has a second highest F value. BusterNet and SeNet realize F values of 0.482 and 0.535, respectively, and obtain relatively balanced *Precision* and *Recall* values.

Fig. 9 presents the comparison of the forgery detection results of selected test images without a rotation attack. As shown in Fig. 9, the SIFT-based CMFD methods often fail to detect the forged regions. In contrast, the machine

learning-based methods, including the proposed method, identify even some of the copied or moved areas. The proposed network detects both copied and moved regions with few failures.

Table 5 lists the three evaluation measures for copied-moved images with a rotation attack. In general, the CMFD performance is degraded when a rotation attack is applied during the move process. As shown in Table 5, the F value of all methods is reduced compared to that when no rotation attack occurs, as listed in Table 4. BusterNet and SeNet reduce the F value from 0.482 and 0.535 to 0.367 and 0.375, respectively, for respective rates of 23.86% and 29.91%. The proposed method only reduces the F value by approximately

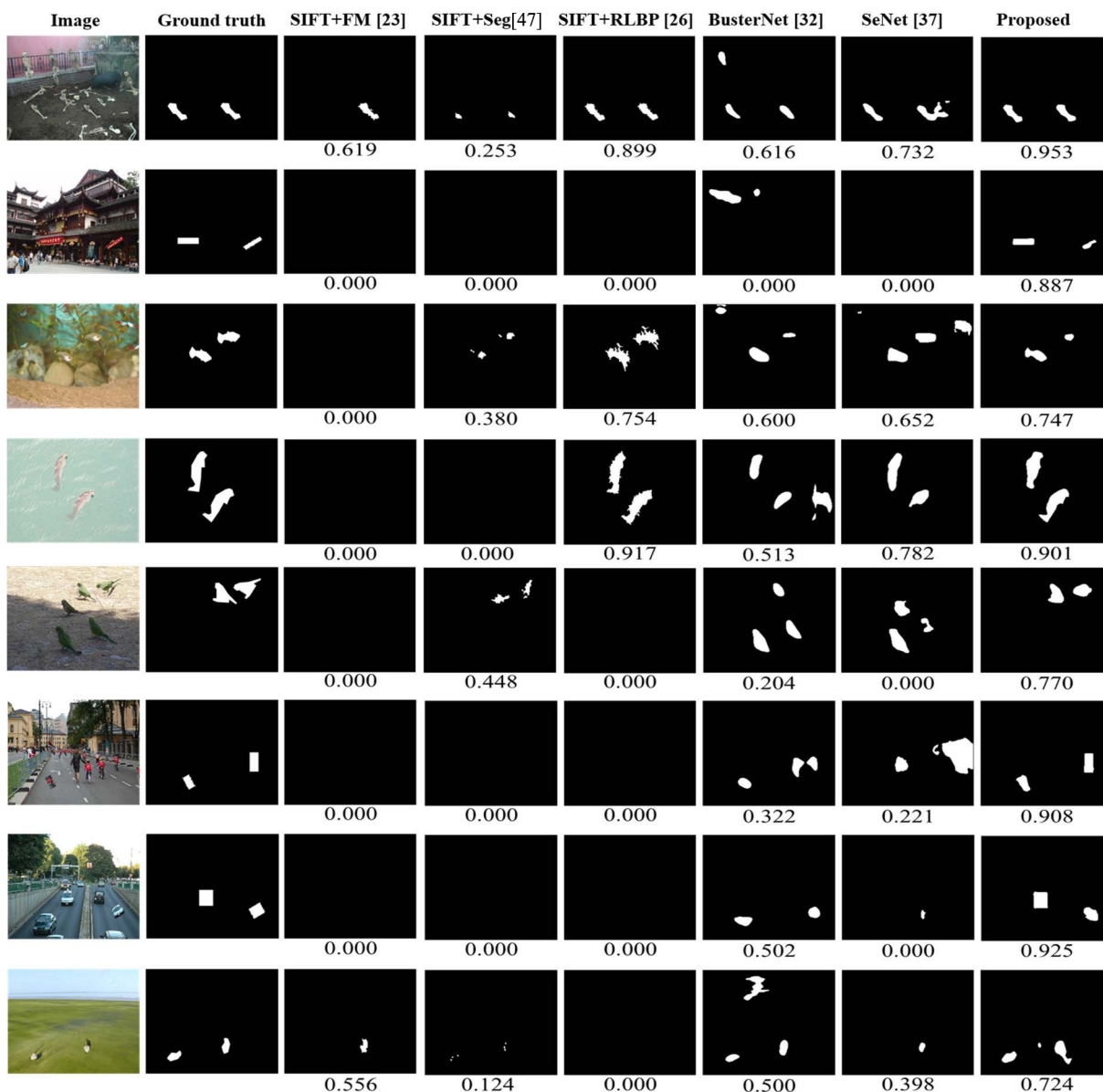


FIGURE 10. Comparison of the forgery detection results and their corresponding *F*-measure values of the test images with a rotation attack.

1% from 0.905 to 0.896. Because the proposed CMFD network uses a rotation-invariant wavelet energy feature, similar detection performance can be achieved with or without rotation attacks.

TABLE 5. Comparison of the detection performance on the test image with a rotation attack according to the pixel-level *Precision*, *Recall*, and *F* measures.

Method	<i>Precision</i>	<i>Recall</i>	<i>F</i>
SIFT+FM [23]	0.379	0.168	0.221
SIFT+Seg [47]	0.765	0.259	0.346
SIFT+RLBP [26]	0.473	0.556	0.489
BusterNet [32]	0.381	0.473	0.367
SeNet [37]	0.442	0.391	0.375
Proposed	0.937	0.870	0.896

Fig. 10 shows the comparison of the forgery detection results of selected test images with a rotation attack. As shown in Fig. 10, the proposed method adequately detects the forged regions despite the small errors, whereas other methods fail to localize the forged regions most of the time.

Table 6 shows the evaluation measures of all test images including scale, rotation, and various attacks. The proposed method exhibits the highest performance, followed by SIFT + RLBP, BusterNet, and SeNet on that order. Because the machine learning-based BusterNet and SeNet methods do not consider rotation-invariance, a large difference in the performance appears when no rotation attack occurs.

In our proposed network, a simplified mask decoder module is introduced to eliminate erroneous small spots.

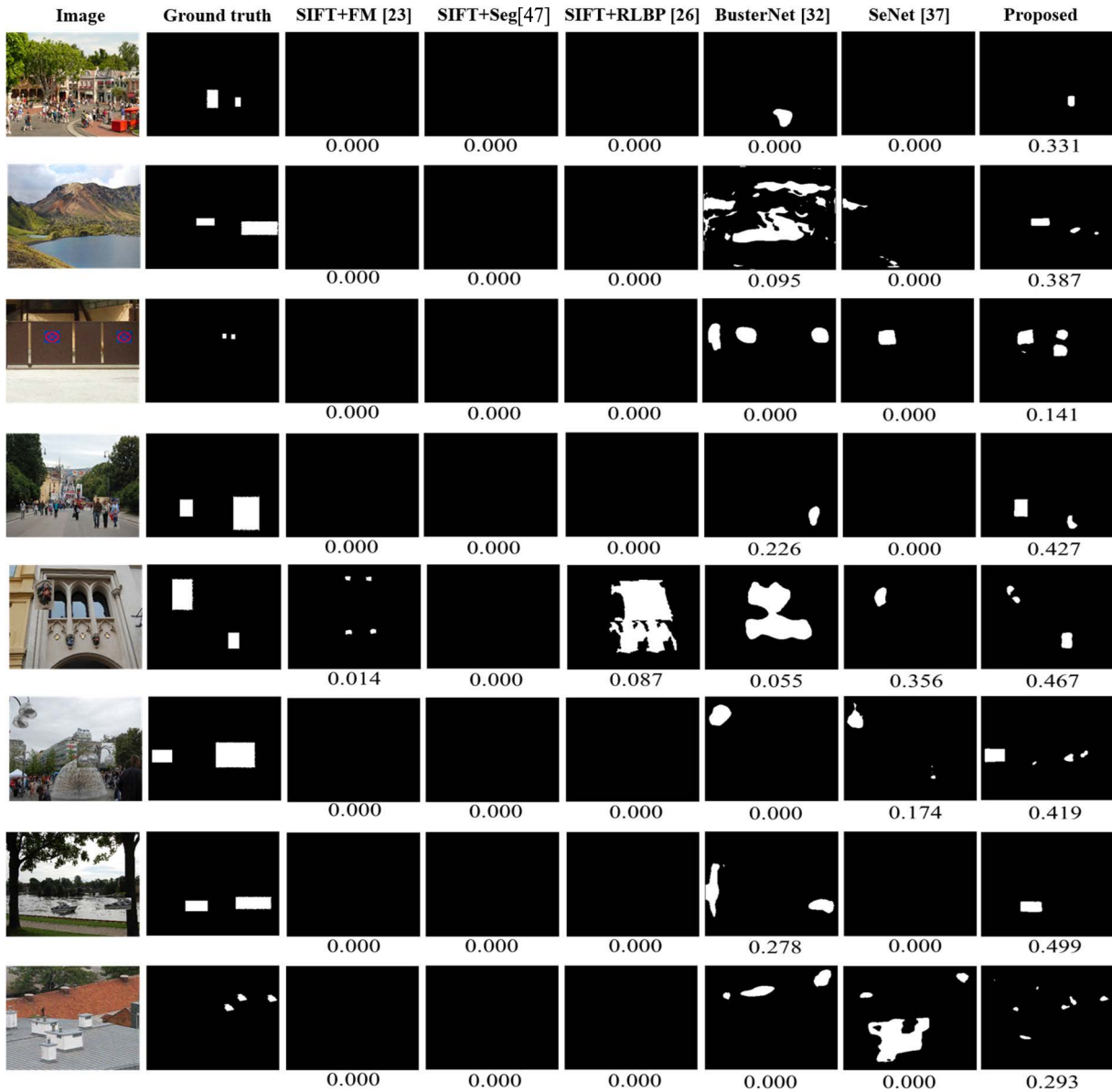


FIGURE 11. Failure cases and their corresponding F -measure values.

TABLE 6. Comparison of the detection performance on the total test image according to the pixel-level $Precision$, $Recall$, and F measures.

Method	$Precision$	$Recall$	F
SIFT+FM [23]	0.413	0.193	0.245
SIFT+Seg [47]	0.765	0.276	0.366
SIFT+RLBP [26]	0.521	0.629	0.542
BusterNet [32]	0.443	0.561	0.445
SeNet [37]	0.545	0.510	0.486
Proposed	0.949	0.871	0.902

Table 7 lists a summary of the effect of using the simplified mask decoder module. As previously explained, the simplified mask decoder not only eliminates the incorrectly detected regions (increases $Precision$), but also reduces the

properly identified regions (decreases $Recall$). However, the overall localization performance, which is represented by the F measure, can be improved using the simplified mask decoder module.

TABLE 7. Comparison of the detection performance between the conventional and simplified mask decoder modules.

Proposed method	$Precision$	$Recall$	F
Conventional mask decoder module	0.926	0.879	0.895
Simplified mask decoder module	0.949	0.871	0.902

Fig. 11 shows examples of detection failure, which is defined as a case where the F measure is less than 0.5. As shown in Fig. 11, a failure case can occur when the

manipulated region is very small or the scale difference between the copied and moved regions is large. Therefore, we believe that a future CMFD network should be designed to achieve high detection performance when very small areas are manipulated or even when a large difference in the size of the copied and moved regions exists.

3) ABLATION STUDY

The proposed network comprises rotation-invariant feature, feature extraction, correlation, and mask decoder modules. Three modules except the rotation-invariant feature module are simple and essential elements. Therefore, an ablation study was performed on the rotation-invariant feature module in this experiment.

TABLE 8. Detection performance according to various combinations of rotation-invariant features.

Rotation-invariant features	Precision	Recall	F
$\mathbf{W}_{i,1}$	0.693	0.937	0.774
$\mathbf{E}_1 + \mathbf{E}_2$	0.903	0.877	0.881
$\mathbf{W}_{i,1} + \mathbf{E}_2$	0.953	0.837	0.885
$\mathbf{W}_{i,1} + \mathbf{E}_1 + \mathbf{E}_2$	0.949	0.871	0.902

Table 8 shows the detection performance according to various combination of rotation-invariant features. As shown in Table 8, when only $\mathbf{W}_{i,1}$ that is the blurred version of the input image, the lowest F value is achieved. If only two levels energy features, \mathbf{E}_1 and \mathbf{E}_2 are used, the F value is expressed as 0.813. The best detection performance is achieved when using $\mathbf{W}_{i,1}$, \mathbf{E}_1 and \mathbf{E}_2 , together. From the results of the ablation study in Table 8, we can notice that the proposed rotation-invariant wavelet energy feature makes an important contribution to the CMFD performance.

4) DETECTION RESULTS FOR SYNTHETIC DATASET

The BusterNet and SeNet used synthetic 100,000 samples for training. However, two machine learning-based networks, only provide test codes in their respective project sites. Because machine learning model largely depends on training data, the performance of the proposed network using training data and test data in the same pool is superior compared to other networks. For a fairer comparison, we created 100,000 synthetic image pairs, similar to the methods of BusterNet and SeNet. All synthetic image pairs are divided into three categories, namely, training, validation, and test, at a ratio of 7:1.5:1.5, respectively.

Table 9 lists the evaluation measures for synthetic images. As shown in Table 9, the F values of SIFT-based methods decrease, while the F values of BusterNet and SeNet are slight increase. The F value of the proposed method decreases from 0.902 to 0.648. The results shown in Tables 6 and 9 show the data dependence of machine learning-based approaches. However, the performance of the proposed method is still the best for 100,000 synthetic samples.

TABLE 9. Comparison of the detection performance on the synthetic test images according to the pixel-level Precision, Recall, and F measures.

Method	Precision	Recall	F
SIFT+FM [23]	0.402	0.161	0.220
SIFT+Seg [47]	0.335	0.061	0.091
SIFT+RLBP [26]	0.531	0.464	0.472
BusterNet [32]	0.531	0.493	0.461
SeNet [37]	0.599	0.527	0.522
Proposed	0.657	0.697	0.648

5) LIMITATIONS

Machine learning-based CMFD approaches demonstrate promising detection results. However, because there are no standard training and test datasets for comparison, accurate detection performance comparison between various methods is difficult. In addition, the detection performance for a large synthetic dataset is still low. Therefore, machine learning-based CMFD is challenging work, and a more dedicated and well-structured network is required.

V. CONCLUSION

This paper proposed a novel copy-move detection network using CNN. We developed four blocks to localize the copy-move regions. A wavelet-based rotation-invariant module that used the root-mean squared energy in the high-frequency stationary wavelet subbands was proposed. This root-mean squared energy achieved a robust performance against rotation attacks. The conventional VGG16 structure was used as the main feature extraction module. The correlation module was adopted to generate potential copied and moved patch pairs, and the feature similarity score was computed using this module. Finally, we introduced the simplified mask decoder module to reduce the reinforcement of erroneous small spots using bilinear interpolation. The proposed method was compared with existing CMF localization algorithms. The simulation results demonstrated the proposed network had almost no difference in F values between the forged image with rotation and the tampered image with scaling, whereas the existing CNN-based methods showed a difference of more than 10%. For synthetic dataset, the proposed method outperformed state-of-the-art approaches by 12% in terms of the F measure.

REFERENCES

- [1] L. Zheng, Y. Zhang, and L. Vrizlynn, "A survey on image tampering and its detection in real-world photos," *J. Vis. Commun. Image Represent.*, vol. 58, pp. 380–399, Jan. 2019.
- [2] S. Teerakanok and T. Uehara, "Copy-move forgery detection: A state-of-the-art technical review and analysis," *IEEE Access*, vol. 7, pp. 40550–40568, 2019.
- [3] W. D. Ferreira, C. B. R. Ferreira, G. da Cruz Júnior, and F. Soares, "A review of digital image forensics," *Comput. Electr. Eng.*, vol. 85, Jul. 2020, Art. no. 106685.
- [4] R. Thakur and R. Rohilla, "Recent advances in digital image manipulation detection techniques: A brief review," *Forensic Sci. Int.*, vol. 312, Jul. 2020, Art. no. 110311.

- [5] Y. Gong, L. Wang, R. Guo, and S. Lazebnik, "Multi-scale orderless pooling of deep convolutional activation features," in *Proc. Eur. Conf. Comput. Vis.*, Zurich, Switzerland, Sep. 2014, pp. 392–407.
- [6] G. Lynch, F. Y. Shih, and H.-Y. Liao, "An efficient expanding block algorithm for image copy-move forgery detection," *Inf. Sci.*, vol. 239, pp. 253–265, Aug. 2013.
- [7] J. Zhao and J. Guo, "Passive forensics for copy-move image forgery using a method based on DCT and SVD," *Forensic Sci. Int.*, vol. 233, nos. 1–3, pp. 158–166, 2013.
- [8] Y. Sun, R. Ni, and Y. Zhao, "Nonoverlapping blocks based copy-move forgery detection," *Secur. Commun. Netw.*, vol. 2018, pp. 1–11, Jan. 2018.
- [9] Y. Cao, T. Gao, L. Fan, and Q. Yang, "A robust detection algorithm for copy-move forgery in digital images," *Forensic Sci. Int.*, vol. 214, nos. 1–3, pp. 33–43, 2012.
- [10] J. Zhong, Y. Gan, J. Young, L. Huang, and P. Lin, "A new block-based method for copy move forgery detection under image geometric transforms," *Multimedia Tools Appl.*, vol. 76, no. 13, pp. 14887–14903, Jul. 2017.
- [11] J. Li, X. Li, B. Yang, and X. Sun, "Segmentation-based image copy-move forgery detection scheme," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 3, pp. 507–518, Mar. 2015.
- [12] C.-M. Pun and J.-L. Chung, "A two-stage localization for copy-move forgery detection," *Inf. Sci.*, vols. 463–464, pp. 33–55, Oct. 2018.
- [13] Y. Li, "Image copy-move forgery detection based on polar cosine transform and approximate nearest neighbor searching," *Forensic Sci. Int.*, vol. 224, nos. 1–3, pp. 59–67, 2013.
- [14] R. Dixit and R. Naskar, "Copy-move forgery detection utilizing Fourier-Mellin transform log-polar features," *J. Electron. Imag.*, vol. 27, no. 2, p. 1, Mar. 2018.
- [15] K. M. Hosny, H. M. Hamza, and N. A. Lashin, "Copy-move forgery detection of duplicated objects using accurate PCET moments and morphological operators," *Imag. Sci. J.*, vol. 66, no. 6, pp. 330–345, Apr. 2018.
- [16] M. H. Alkawaz, G. Sulong, T. Saba, and A. Rehman, "Detection of copy-move image forgery based on discrete cosine transform," *Neural Comput. Appl.*, vol. 30, no. 1, pp. 183–192, Nov. 2016.
- [17] G. Muhammadiyah, M. Hussain, and G. Bebis, "Passive copy move image forgery detection using undecimated dyadic wavelet transform," *Digit. Investigat.*, vol. 9, no. 1, pp. 49–57, 2012.
- [18] T. Mahmood, A. Irtaza, Z. Mehmood, and M. T. Mahmood, "Copy-move forgery detection through stationary wavelets and local binary pattern variance for forensic analysis in digital images," *Forensic Sci. Int.*, vol. 279, pp. 8–21, Oct. 2017.
- [19] J.-C. Lee, C.-P. Chang, and W.-K. Chen, "Detection of copy-move image forgery using histogram of orientated gradients," *Inf. Sci.*, vol. 321, pp. 250–262, Nov. 2015.
- [20] S.-J. Ryu, M. Kirchner, M.-J. Lee, and H.-K. Lee, "Rotation invariant localization of duplicated image regions based on Zernike moments," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 8, pp. 1355–1370, Aug. 2013.
- [21] B. Mahdian and S. Saic, "Detection of copy-move forgery using a method based on blur moment invariants," *Forensic Sci. Int.*, vol. 171, no. 2, pp. 180–189, 2007.
- [22] M. Zandi, A. Mahmoudi-Aznavah, and A. Talebpour, "Iterative copy-move forgery detection based on a new interest point detector," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 11, pp. 2499–2512, Nov. 2016.
- [23] 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.
- [24] I. Amerini, L. Ballan, R. Caldelli, A. D. 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.
- [25] V. Christlein, C. Riess, J. Jordan, C. Riess, and E. Angelopoulou, "An evaluation of popular copy-move forgery detection approaches," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 6, pp. 1841–1854, Dec. 2012.
- [26] J. Y. Park, T. A. Kang, Y. H. Moon, and I. K. Eom, "Copy-move forgery detection using scale invariant feature and reduced local binary pattern histogram," *Symmetry*, vol. 12, no. 4, p. 492, Mar. 2020.
- [27] G. Jin and X. Wan, "An improved method for SIFT-based copy-move forgery detection using non-maximum value suppression and optimized J-Linkage," *Signal Process., Image Commun.*, vol. 57, pp. 113–125, Sep. 2017.
- [28] D. Uliyan, H. Jalab, A. A. Wahab, and S. Sadeghi, "Image region duplication forgery detection based on angular radial partitioning and Harris key-points," *Symmetry*, vol. 8, no. 7, p. 62, Jul. 2016.
- [29] N. B. A. Warif, A. W. A. Wahab, M. Y. I. Idris, R. Salleh, and F. Othman, "SIFT-symmetry: A robust detection method for copy-move forgery with reflection attack," *J. Vis. Commun. Image Represent.*, vol. 46, pp. 219–232, Jul. 2017.
- [30] J. Bunk, J. H. Bappy, T. M. Mohammed, L. Nataraj, A. Flenner, B. S. Manjunath, S. Chandrasekaran, A. K. Roy-Chowdhury, and L. Peterson, "Detection and localization of image forgeries using resampling features and deep learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Honolulu, HI, USA, Jul. 2017, pp. 1881–1889.
- [31] Y. Liu, Q. Guan, and X. Zhao, "Copy-move forgery detection based on convolutional kernel network," *Multimedia Tools Appl.*, vol. 77, no. 14, p. 18269–18293, 2018.
- [32] Y. Wu, W. Abd-Almageed, and P. Natarajan, "BusterNet: Detecting copy-move image forgery with source/target localization," in *Proc. Eur. Conf. Comput. Vis.*, Munich, Germany, Sep. 2018, pp. 170–186.
- [33] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*, 2015, pp. 1–14.
- [34] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Boston, MA, USA, Jun. 2015, pp. 1–9.
- [35] J.-L. Zhong and C.-M. Pun, "An end-to-end dense-InceptionNet for image copy-move forgery detection," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 2134–2146, 2020.
- [36] Y. Zhu, C. Chen, G. Yan, Y. Guo, and Y. Dong, "AR-Net: Adaptive attention and residual refinement network for copy-move forgery detection," *IEEE Trans. Ind. Informat.*, vol. 16, no. 10, pp. 6714–6723, Oct. 2020.
- [37] B. Chen, W. Tan, G. Coatrieux, Y. Zheng, and Y.-Q. Shi, "A serial image copy-move forgery localization scheme with source/target distinguishment," *IEEE Trans. Multimedia*, vol. 23, pp. 3506–3517, 2021.
- [38] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.
- [39] M. Barni, Q.-T. Phan, and B. Tondi, "Copy move source-target disambiguation through multi-branch CNNs," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 1825–1840, 2021.
- [40] Y. Liu, C. Xia, X. Zhu, and S. Xu, "Two-stage copy-move forgery detection with self deep matching and proposal SuperGlue," *IEEE Trans. Image Process.*, vol. 31, pp. 541–555, 2022.
- [41] Y. Chen, Z. X. Lyu, X. Kang, and Z. J. Wang, "A rotation-invariant convolutional neural network for image enhancement forensics," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Calgary, AB, Canada, Apr. 2018, pp. 2111–2115.
- [42] G. Cheng, J. Han, P. Zhou, and D. Xu, "Learning rotation-invariant and Fisher discriminative convolutional neural networks for object detection," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 265–278, Jan. 2019.
- [43] B. Bayar and M. C. Stamm, "Constrained convolutional neural networks: A new approach towards general purpose image manipulation detection," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 11, pp. 2691–2706, Nov. 2018.
- [44] D. Tralic, I. Zupancic, S. Grgic, and M. Grgic, "CoMoFoD—New database for copy-move forgery detection," in *Proc. ELMAR*, Zadar, Croatia, Sep. 2013, pp. 49–54.
- [45] E. Ardizzone, A. Bruno, and G. Mazzola, "Copy-move forgery detection by matching triangles of keypoints," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 10, pp. 2084–2094, Oct. 2015.

- [46] 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.
- [47] C. Pun, X. Yuan, and X. Bi, "Image forgery detection using adaptive oversegmentation and feature point matching," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 8, pp. 1705–1716, Aug. 2015.



JUN YOUNG PARK received the B.S. degree from the Department of Electronics Engineering, Pusan National University, South Korea, in 2017, where he is currently pursuing the Ph.D. degree in electronics engineering. His research interests include image enhancement, computer vision, copy-move forgery localization, and machine learning.



SANG IN LEE received the B.S. degree from the Department of Electronics Engineering, Pusan National University, South Korea, in 2021, where he is currently pursuing the M.S. degree in electronics engineering. His research interests include image processing, copy-move forgery localization, and machine learning.



IL KYU EOM received the B.S., M.S., and Ph.D. degrees from the Department of Electronics Engineering, Pusan National University, South Korea, in 1990, 1992, and 1998, respectively. From 1997 to 2005, he was a Faculty Member at Miryang National University, South Korea. He has been a Faculty Member with Pusan National University, since 2006. He is a Full Professor with the Department of Electronics Engineering. His research interests include image processing, computer vision, digital image forensic, and machine learning.

...