IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 17 February 2024, accepted 10 March 2024, date of publication 21 March 2024, date of current version 28 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3380460

RESEARCH ARTICLE

CNN-Keypoint Based Two-Stage Hybrid Approach for Copy-Move Forgery Detection

ANJALI DIWAN[1], (Senior Member, IEEE), AND ANIL K. ROY[2], (Senior Member, IEEE)
[1]Department of CE-AI, Marwadi University, Rajkot, Gujarat 360003, India
[2]DA-IICT, Gandhinagar, Gujarat 382007, India

Corresponding authors: Anil K. Roy (anil_roy@daiict.ac.in) and Anjali Diwan (anjali.diwan@ieee.org)

ABSTRACT Authenticating digital images poses a significant challenge due to the widespread use of image forgery techniques, including copy-move forgery. Copy-move forgery involves copying and pasting portions of an image within the same image while applying geometric transformations to make the forged image appear genuine. Furthermore, additional processing techniques such as additive noise scaling, JPEG compression, and rotation can be employed to further conceal evidence of forgery. These factors contribute to the complexity of detecting and verifying the authenticity of digital images. The proposed work uses a combination of the CenSurE keypoint detection and a CNN architecture to detect and localize copy-move forgery in digital images. The use of CNN architecture allows the algorithm to update its learning via training data repeatedly, making it a data-driven approach. By combining keypoints with CNN features, the proposed approach can enhance the detection of copy-move forgery even in the presence of attacks such as geometrical transformations, scale, and rotation. Additionally, the proposed approach can effectively handle post-processing operations such as JPEG compression, additive noise, image blur, colour reduction, brightness change, and contrast adjustment. One important aspect of the proposed approach is its ability to handle images with different textures, including smooth and self-similar structural images with dense textures. The proposed approach can produce stable results in images with various attacks, making it a functional and reliable tool for detecting copy-move forgery in a diverse range of forged images. The proposed approach represents an important contribution to the field of multimedia forensics, providing an effective and reliable means of detecting and localizing copy-move forgery in digital images.

INDEX TERMS CenSurE detector, copy-move, CNN forgery, deep learning, digital image forgery, hybrid approach, image forgery detection, image duplication, image keypoint, multimedia forensics.

I. INTRODUCTION

In recent years, the significance of digital data, especially in the form of images, has grown immensely in information sharing. The accessibility of budget-friendly digital cameras has simplified image capture. The impact of social networking platforms is substantial, with individuals frequently sharing and forwarding images. This turns every user into a potential source of digital content.

Furthermore, the prevalence of user-friendly image editing software empowers both beginners and experts to craft plausible modifications. These alterations might range from harmless modifications for amusement to more insidious image manipulations aimed at deception [1]. This reality introduces doubt into the credibility of digital images. The situation becomes particularly critical when these digital images are employed for specific purposes such as news dissemination, research publications, and legal proceedings.

Recent years have witnessed the role of digital data in the form of images for providing information. The availability of low-cost digital camera-enabled devices made it easy to capture images. Statistics show that the use of social networking sites has influenced people very seriously. Now people go online and upload their pictures while they forward pictures uploaded by others. This makes every user a probable source of digital information.

The associate editor coordinating the review of this manuscript and approving it for publication was Qichun Zhang.

The accessibility of simplistic image-modifying software is advantageous for both novices and experts, enabling them to create authentic modifications. These tweaks range from minor alterations made for entertainment to substantial image manipulation carried out maliciously [2]. This fact poses uncertainty in the authenticity of the digital image. When these digital images are used for specific purposes like news broadcasts, research journals, and the court of law, this problem becomes more critical.

One prevalent form of image forgery involves inserting a portion of the image within the same image, a technique termed copy-move forgery. This manipulation entails copying a segment of the original image and pasting it elsewhere within the image [3]. Illustrative examples of images exhibiting copy-move forgery from the Coverage, MICC-F600, and CASIA II datasets are depicted in Figure (1).

Copy-move forgery detection involves two primary approaches: keypoint-based and block-based. The keypoint-based approach emphasizes the identification of regions with high entropy in an image, which encompasses elements like edges, blobs, and corners, commonly referred to as keypoints [4]. These keypoints are selected based on their uniqueness and robustness, enabling effective detection of copy-move forgery even in the presence of various image processes. Conversely, the block-based approach divides the image into various shapes of blocks, such as squares, circles, or rectangles [5]. Features are derived from each block, and a comparison is performed block-by-block to identify similarities indicating copy-move forgery. Additionally, deep learning models have also been employed for copy-move forgery detection, offering new possibilities and improved performance in this field.

Deep Neural Networks (DNN) have demonstrated their effectiveness in learning hierarchical representations from input data, making them highly versatile in various applications, including image classification and speech recognition [6]. Researchers have leveraged the automatic feature learning capability of DNN for image forgery detection. By utilizing pre-trained neural networks, real-time comprehensive forgery detection approaches can be developed [7]. Deep learning approaches excel at detecting copy-move forgeries by autonomously acquiring distinct features from images, allowing for the precise recognition of even minor changes with enhanced accuracy, robustness, and speed when compared to traditional techniques, thus enhancing security and precision in diverse applications [2].

A. MOTIVATION

Despite numerous advancements, we observed that current copy-move forgery detection methods in image forensics still possess certain limitations.

Our analysis revealed that distinct approaches have diverse capabilities in identifying copy-move forgery. Block-based approaches have excellent detection performance in straightforward copy-move and copy-move with post-processing scenarios such as JPEG compression. However, they encounter difficulties when faced with complex post-processing techniques like geometric transformation [5], [8], [9]. Conversely, keypoint-based approaches exhibit better resistance to geometric alterations but are susceptible to the influence of image texture. Smoother images yield fewer keypoints, while denser textures result in more keypoints [10], [11]. However, achieving the appropriate equilibrium is crucial; an insufficient number of keypoints renders the detection of manipulation difficult, whereas an excessive number of keypoints hampers the identification of copy-move areas.

Complex image attacks, such as extensive rotation and combined attacks involving scale, rotation, and JPEG compression, are frequently employed to conceal the traces of forgery [12]. However, the detection of these sophisticated forgery techniques has been largely neglected by researchers, with only a limited number of studies addressing this specific direction. This highlights the need for more comprehensive research and methodologies to effectively detect and mitigate the impact of these complex image attacks in forgery detection.

The accuracy of forgery detection is significantly influenced by the texture of an image. It has been observed that image features can be significantly influenced by the texture present in the image [13]. However, only a limited number of researchers have explored and addressed the diverse nature of image textures, indicating a gap in understanding and addressing this important aspect of forgery detection.

B. CONTRIBUTION

To overcome the limitations of the existing approaches as mentioned in Section (I-A), we propose a novel approach that combines the CNN architecture and CenSurE keypoint detector for copy-move forgery detection. The CNN architecture allows us to learn and extract meaningful features from the image, enabling efficient detection of copy-move forgery. Simultaneously, the CenSurE keypoint detector is employed to capture structural information within the image. By merging the strengths of both approaches, our proposed method seeks to enhance the accuracy of copy-move forgery detection. Several key advantages of the proposed approach include:

- Our proposed approach introduces a two-channel model that combines a CNN architecture with a CenSurE keypoint detector for improved copy-move forgery detection.
- For boundary prediction of the forged region, we use the residual refinement stage. The decision is further optimized with the help of the softmax layer.
- In addition to widely recognized attacks like noise addition, JPEG compression, rotation, and scaling, our proposed approach possesses the capability to detect a forgery in images subjected to novel attacks such as image blur, color reduction, brightness change, and contrast adjustment. These less-explored attacks are

FIGURE 1. Some example images depicting copy-move forgery here the first row consists of an authentic image and the second row consists forged image.

frequently employed to conceal the location of forgery or camouflage the forged regions. By addressing these novel attacks, our proposed method becomes crucial in achieving comprehensive and effective forgery detection in digital images.

- The proposed approach is capable of detecting and localizing complex image attacks, including extensive rotation and combined attacks involving scale, rotation, and JPEG compression. These sophisticated attack techniques are commonly used to obscure the evidence of forgery.
- By leveraging the advantages of CNN, including its ability to learn statistical features and adaptive feature learning, we aim to address forgery in images with diverse textures, such as smooth, coarse and structural textures.

The structure of this paper is as follows: In Section II, we discussed the Related work. In Section III, we discuss the basics of image feature extraction. In Section IV, we discuss the proposed CNN-based fusion approach. In Section V, we discuss the proposed algorithm for the detection of forgery. In Section VI, we discuss the dataset and evaluation metric. In Section VII, we discussed experiments and results. In Section VIII we draw a conclusion.

II. RELATED WORK

In the field of image forensics, forgery detection methods can be broadly categorized as passive or active. Active methods involve embedding digital information such as watermarks or signatures into images to protect their integrity, but they require prior embedding using specialized equipment. Conversely, passive methods do not depend on prior knowledge and instead analyze the inherent characteristics of the images to ascertain their authenticity. Passive methods are practical ways to handle forgery and have received significant attention from researchers.

Among passive methods, detection approaches of copy-move image forgery can be categorized into keypoint-based, block-based, and deep learning-based approaches. Block-based approaches divide the image into blocks for comparison, while keypoint-based approaches focus on identifying unique points of interest. Deep learning-based methods utilize deep neural networks to learn and extract features for copy-move forgery detection. These different approaches within the passive category offer diverse strategies to detect a copy-move forgery in digital images.

A. BLOCK-BASED

Several researchers have explored a block-based approach for detecting copy-move forgery. Some of these studies include; Babu and Rao [14] have employed a combination of different versions of LBP local ternary pattern (LTeP), local phase quantization (LPQ), and local Gabor binary pattern. The features of LBP are used to train the classifier model and SVM is used for the classification of copy-move forgery verification. Diwan et.al [12] introduced an LPP-based approach for copy-move forgery detection. Their block-based method demonstrates effectiveness for both post-processed and original images. Hosny et al. [15] presented a sub-sampled image method using QPCETMs, incorporating the Sobel operator to identify edges and eliminate small regions, yet its effectiveness might be limited for images with smooth or dense textures. Gani and Qadir [16] utilized Cellular Automata on individual DCT block features within the image, but the method's time complexity is notably high.

The algorithm for block-based copy-move forgery detection compares blocks of the image to identify similar regions that may indicate copy-move forgery. This approach fails to detect forgery when the copied region is geometrically transformed, such as rotated, scaled, or flipped. This is because the transformed region may not match exactly with any of the blocks in the image [17]. Therefore, a different approach is needed to detect geometrically transformed copy-move forgery.

B. KEYPOINT-BASED

The keypoint-based approach can be used as an alternative when a block-based approach fails to detect copy-move

forgeries [18]. The keypoint-based approach depends on identifying and matching distinctive keypoints, which are regions of an image with unique visual features, instead of using fixed-size blocks.

Researchers have employed a variety of keypoint-based forgery detection techniques in the past. Kumar and Meenpal [19] utilized a silent keypoint section approach with SIFT features along with KAZE image keypoint features for the detection of copy-move forgery. Lee et al. [20] Proposed a copy-move detection methodology that utilizes a rotation-invariant characteristic and high-frequency wavelet coefficients. This approach uses a correlation module and a reduced mask decoder module. Venugopalan and Gopakumar [21] employed SIFT keypoint with DBSCAN for the clustering of the extracted keypoints. Additionally, they used Hu invariant moment for getting for identification of similar regions in copy-move images. Wang et al. [22] utilize simple linear iterative clustering (SLIC) and the K-multiple-means (KMM) for feature extraction along with Fast Quaternion Generic Polar Complex Exponential Transform (FQGPCET) and the texture features based on the Gray-level co-occurrence matrix (GLCM) to enhance the robust feature extraction for copy-move forgery detection.

Keypoints are usually detected by finding areas in the image that have high-contrast changes in texture or color. The count of keypoints detected in an image is influenced by its texture. In the case of smooth images, there may be fewer keypoints available for detection compared to images with more textured regions. This can lead to a lower detection rate for keypoint-based approaches for smooth images [10].

C. DEEP LEARNING-BASED

Both keypoint-based and block-based approaches have their strengths and weaknesses, and the choice of approach depends on the specific requirements of the application and the complexity of the forgery. However, it is important to note that simply extracting features of blocks and keypoints may not be enough to provide a foolproof forgery detection result. In addition to the feature extraction process, an effective feature selection process and classification techniques are also necessary for accurate forgery detection.

In recent years, deep learning-based approaches have demonstrated significant potential due to their capacity to autonomously learn intricate features and deliver precise predictions [23]. These approaches have been used for copy-move forgery detection and have shown improved performance compared to traditional methods.

Deep learning-based approaches have been used for copy-move forgery detection by researchers. Some of them are; Xiong et al. [24] have used a multiscale fusion network model for copy-move forgery detection and localization. Zhang et al. [25] have used CNN and transformer-based generative adversarial networks for feature extraction and copy-move forgery detection. Kaur et al. [26] have used contrast-limited adaptive histogram equalization with CNN to make a deep neural network that effectively detects copy-move forgery. Zhu et al. [27] have proposed end-to-end AR-Net along with deep matching to capture context information fully. Kuznetsov et al. [28] have used an optimized tailored CNN base classifier for copy-move forgery detection.

III. BASICS OF IMAGE FEATURE EXTRACTION

As discussed in section II image feature extraction process is the most decisive step in copy-move forgery detection. The result of forgery is highly dependent on the image features we are extracting and processing. We are focusing on two types of image features first is local and global keypoint features and second image statistical features extracted by neural networks. In this section, we will discuss briefly the keypoint the CenSurE detector, FREAK descriptor, and Convolutional Neural Network (CNN).

A. KEYPOINT DETECTOR AND DESCRIPTOR

Keypoints are often described by a set of descriptors that capture the local image information around the keypoint. There are many methods for detecting keypoints in images, such as Speeded-Up Robust Features (SURF), Scale-Invariant Feature Transform (SIFT), the Harris corner detector, and CenSurE (Center Surround Extrema). Each method has its advantages and disadvantages, depending on the application and the characteristics of the images.

To achieve efficient copy-move forgery detection, the keypoints should possess a sparsity, repeatability, and distinctiveness to optimize matching accuracy [29]. The repeatability of a detector, quantifying the consistency of detected regions across images, is crucial in this context. It's a measure based solely on feature geometry, defined as the ratio between simultaneously detected points in a pair of images [10]. We understand that we need a combination of the detector (of the keypoints in the image) and descriptor (of similarity) to successfully detect and localize various types of copy-move forgery. Diwan et al. [10] have done a detailed experiment on the various combinations of detectors and descriptors to determine the best response for repeatability, geometric transformation, and other post-processing attacks. It has been found that the CenSurE detector and FREAK descriptor give the best responses for the different images.

CenSurE Keypoint Detector: CenSurE (Center Surround Extremas) is a keypoint detector that was introduced by Agrawal et al. [30]. The detector is based on the idea that keypoints in an image are regions where the intensity variation is significant with respect to the surrounding pixels.

CenSurE operates by convolving an image with a filter bank that consists of two filters: a center filter and a surround filter. The center filter is designed to respond to edges and the surrounding filter is designed to suppress the responses of the center filter in homogeneous regions. The output of the convolution is then thresholded to obtain the keypoints.

The CenSurE detector has three main steps, which include:

- Detecting edges involves filtering out weaker responses and computing the response to a bilevel LoG.

- Detecting local extrema using Harris measure, which has a strong corner response.
- Calculating features at various scales and then identifying significant points across all locations and scales to identify large-scale features.

CenSurE is using a box filter where the size of the inner box is $(2n+1)\times(2n+1)$ and the size of the outer box is $(4n+1)\times(4n+1)$. Haar wavelet of seven scales is applied on the box filters with different block size e.g. $n = [1, 2, 3, 4, 5, 6, 7]$. Seven filter responses are calculated for every pixel in the image. After that, non-maximal suppression is computed for each scale. A response is suppressed when it is higher (in the case of maxima) or lower (in the case of minima) than its neighboring responses in a local area across different scales. Pixels pointed as either maxima or minima within this area are considered feature point locations.

Let I be the input image, and let G_c and G_s be the center and surround filters, respectively. The response of CenSurE at a pixel (x,y) is defined as:

$$C(x, y) = Max[(I * G_c)(x, y) * R((I * G_s)(x, y))]$$

Here $*$ denotes the convolution operation and $R()$ thresholding function and $max()$ returns the maximum value of its arguments.

The center filter G_c can be defined as a derivative of the Gaussian filter, which responds to edges in the image. The surround filter G_s can be defined as a box filter, which suppresses the responses of the center filter in homogeneous regions. The threshold function $R()$ is:

$$R(x) = 1\ if\ |x| > T,\ \ 0\ otherwise \tag{1}$$

where T is a threshold value that is determined empirically. The thresholding function is used to suppress responses that are not significant with respect to the surrounding pixels.

CenSurE can be effectively calculated using integral images, which allows for robust and quick detection of keypoints. Hence, image features detected by CenSurE give higher accuracy in copy-move forgery detection.

One advantage of CenSurE is its computational efficiency. The convolution operation can be efficiently implemented using integral images, which makes it much faster than other keypoint detectors like SIFT and SURF. Moreover, CenSurE is relatively robust to changes in image scale and rotation. CenSurE can be employed in the field of image forensics to identify instances of copy-move forgery within digital images. It is done by identifying regions of the image that have been duplicated.

FREAK Descriptors: Image keypoint descriptors are compact representations of the local image information around specific keypoints. FREAK (Fast Retina Keypoint) [31] is a binary feature descriptor that is used for efficient and robust keypoint matching in computer vision applications. FREAK is based on the Retina keypoint detector, which detects keypoints in images by finding the local extrema of the Laplacian of the Gaussian function. FREAK then computes a binary descriptor for each keypoint by analyzing the intensity patterns around the keypoint. The binary descriptor is generated using a combination of binary tests that compare pairs of pixels around the keypoint.

FREAK employs a sample pattern based on the characteristics of the human retina, the FREAK descriptor is distinctive. In concentric rings with various radii around a keypoint, it chooses a collection of point pairs (sampling points) at random. The intensity difference between these point pairs is then calculated. The binary string that makes up the descriptor is created using these intensity differences. The process is repeated for multiple radii and orientations to capture information at different scales and orientations. The binary string F is constructed through a sequence of one-bit Differences of Gaussian (DoG), as demonstrated in Equation (2).

$$\sum_{0\leqslant n\leqslant M} 2^n T(P_n) = F. \tag{2}$$

The receptive field pair is represented as P_n, while M signifies the descriptor's size, which can be configured as needed.

$$T(P_n) = \begin{Bmatrix} 1{=}\text{if}\ \ I(P_n^{r1}) - I(P_n^{r2}) > 0 \\ 0{=}\text{Otherwise} \end{Bmatrix} \tag{3}$$

Every pair of the receptive fields undergoes smoothing by a Gaussian filter as represented Eq. (3). $I(P_n^{r1})$ is first field of pair P_n.

FREAK is constructed through intensity comparisons of 512 sampling pairs. This involves randomly selecting two circles and then identifying pairs that yield more informative results. For each pair, if the intensity of the first point exceeds that of the second, a '1' is recorded; otherwise, a '0' is recorded in the corresponding descriptor bit. Performance can be tuned by modifying Gaussian kernel sizes or allowing receptive fields to overlap. Overlaps capture more information, enhancing performance. Moreover, pixels are averaged and concentrated around the keypoint, resulting in a more precise description of the keypoint.

B. CNN ARCHITECTURE AND IMAGE FEATURE EXTRACTION

CNNs, belonging to the Deep Neural Networks category, excel at recognizing, classifying, and analyzing visual features in images. In the context of CNN, the term 'Convolution' signifies a mathematical operation where two functions undergo multiplication to generate a third function, portraying how one function's structure is altered by the other. In more straightforward words, this operation involves matrix multiplication of image representations, producing an output that aids in feature extraction from the image.

Two prevailing parts of CNN architecture are:

- A convolution tool extracts and isolates unique image features during a process known as feature extraction, enabling their distinct identification and subsequent analysis.

- A fully connected layer relies on the features that were extracted in earlier stages to forecast the class of the image using the results from the convolution process.

Convolution Neural Network: CNN architecture consists of multiple cascaded convolution layers which end with a layer that is fully connected as represented in Figure (4). CNN mainly has input, output, and hidden layers. The hidden layers consist of ReLU, pooling, and fully connected layers. Here every layer performs a definite function of transforming its input into a distinct representation..

The very first layer is the input layer which is the convolution layer that applies convolution operation to the input. The main goal of this layer is to use filters or kernels to extract features from the input image by applying them to the image by sliding pixel-by-pixel. Element-wise matrix multiplication is done between the image and the kernel and summation follows thereafter. This sum contributes to the creation of a feature map. Ultimately, an array known as a feature map or activation map of the image is generated. This accumulation leads to the formation of a feature map, culminating in the generation of an array termed as a feature map or activation map for the image.

Next is the pooling layer that combines the outputs of the convolution layers. It primarily reduces the spatial dimension of the previous convolution layer output before passing it to the next convolution layer. Pooling primarily aids in extracting distinct and refined features. Additionally, it serves to diminish variance and computational complexity. Max-pooling is effective in capturing basic features such as edges and points. We need detailed low-level image features hence we are using max pooing for proposed copy-move forgery detection.

Following the convolution and pooling stages, fully connected layers comprise the final phase of the CNN architecture. The fully connected layer encompasses neurons, weights, and biases, facilitating connections between distinct layers. Typically positioned before the output layer, these layers constitute the concluding sections of a CNN architecture. After undergoing prior layer processing, the input image is transformed into a flattened format and directed toward the fully connected layer. This flattened vector is then subjected to additional fully connected layers, where mathematical functions are frequently applied. This stage marks the commencement of the classification process.

The purpose of using the CNN network is that it learns image features automatically and accurately in training the model through adaptive learning. VGG16 is good for finding image copy-move parts because it can understand underlying details in the texture of the image. The deep convolutional layers of VGG16 excel at extracting hierarchical features, a crucial aspect for tasks like image tampering detection that necessitate a profound understanding of patterns and textures. We used VGG16 [32] for feature extraction of the test image. In VGG16 network architecture multiple 3×3 sized filters are applied in the convolution layer and 2×2 sized filter is applied in the pooling layer. It has its apparent advantage as the multiple small-sized layers increase the depth of the CNN, eventually resulting in more detailed learning and less computational cost [33]. This ensures the extraction of representative features even in the case of images with low textures and very smooth structures.

IV. PROPOSED CNN-BASED FUSION APPROACH

Accurate and efficient feature extraction is a prime requisite in the copy-move forgery detection problem. CNN is especially known for learning image features in an adaptive manner when working on a large dataset. Multiple layers in CNN can learn detailed image features and automatically detect similar features present in them. In this novel approach we took advantage of two different types of feature extraction and final detection is suggested based on the output of the fusion of both CenSurE keypoint and CNN architecture. This can be achieved by a proposed two-channel network: The CenSurE branch (Channel-1) and CNN branch (Channel-2). The schematic of this approach is shown in Figure (2).

A. CHANNEL-1

The **Channel-1** of Figure (2) is doing the keypoint-based copy-move forgery detection in this model. In this channel, we are using the CenSurE detector, which gives rotation and scale-invariant keypoints, and the FREAK descriptor, which has geometrical invariant features. Keypoint-based copy-move forgery detection is done on this channel of the model.

In the proposed keypoint-based channel, attention was paid to getting image features that are invariant to geometrical transformations like scale and rotation, as well as getting stable keypoints. Stable keypoints can be extracted by creating an image pyramid, but this method suffers from poor localization of scaled features at higher levels of the pyramid. To address this, the CenSurE detector was employed, generating features that encompass all pixels of the image across various scales.

Once the keypoints are detected using the chosen detector, the repeatability rate is calculated to determine how many of the detected regions in one image are also present in another image. This is important for detecting the same features despite variations in angle or other transformations. To eliminate outliers and improve detection and localization, a threshold is set for different textured images with different ranges of keypoints. Once the keypoints are detected, the nature of the geometrical transformation between the copied and moved regions is determined. Subsequently, the coordinate data of both copied and moved regions are used for an affine homographic matrix preparation. Homography, rotation, and scaling calculations are done using homographic matrix decomposition, while translation is determined by identifying the centroid of the cluster.

Figure (3) shows the framework of Channel-1. Considering the targeted image (I) we are detecting a set of keypoints ($X = x_1, x_2, \ldots x_i$) and then finding descriptor $f_1, f_2, .. f_i$ for each keypoint. The similarity between the

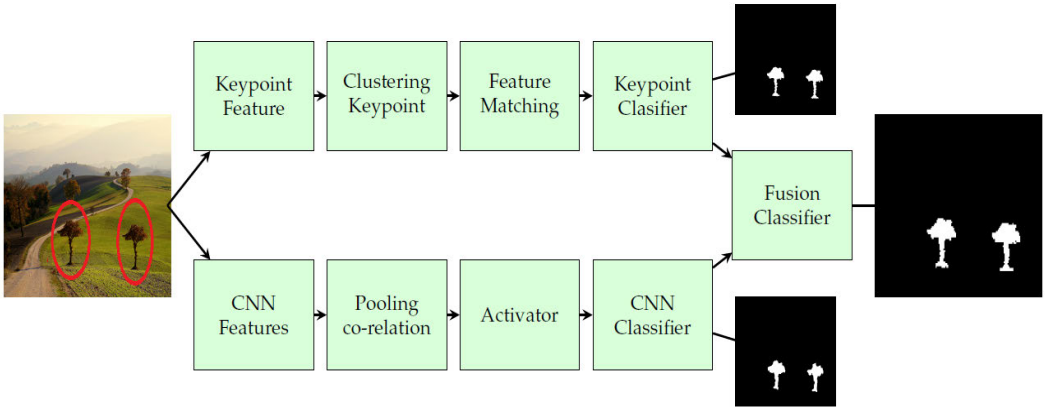

FIGURE 2. Proposed CNN-CenSurE based copy-move forgery detection framework.

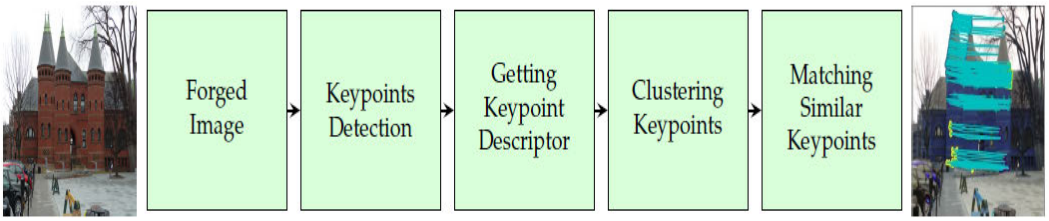

FIGURE 3. Framework of Channel-1.

two keypoints is determined by matching the descriptor of the corresponding keypoints. The matching process involves the nearest neighborhood search applied to the keypoints. For this, the Euclidean distance between two descriptors is utilized. A global threshold is set to assess the distance between descriptors. Given the high-dimensional feature space, certain descriptors need to be notably more discriminative. Thus, we calculate the ratio between the two closest neighbors and compare it with the threshold. A similarity vector $S = d_1, d_2, \ldots d_n - 1$ of the keypoint is defined. This similarity vector represents the descriptor with the shortest distance concerning other descriptors. The 2NN (2)-nearest neighbour) test must satisfy for keypoints matching, and for this ratio between the distance 1 (d1) and distance 2 (d2) must be higher than the threshold (T).

$$\frac{d_1}{d_2} > T \text{ , where } T \in 1, 0 \tag{4}$$

Outliers can affect the estimation of the forged region, so it is essential to remove them. The Random Sample Consensus algorithm RANSAC is used for this purpose. In RANSAC, a set of matched points is chosen randomly to estimate the homography. Then, the remaining keypoints are transformed and their spatial distance is evaluated with respect to their matched points. An outlier is identified when the distance of points surpasses a predefined threshold.

B. CHANNEL-2

Figure (4) shows the framework of channel-2. In the CNN channel, the input image is passed through several convolution layers to capture the spatial and temporal dependencies in the images. We are using VGG16 and at first, we are giving input matrix to the network. It starts with the application of two convolution layers having 64 channels of 3×3 sized filters and max-pooling of stride (2,2). This combination is repeated twice. Then three convolution layers of 256 channels of 3×3 sized filters and the same max-pooling layers are

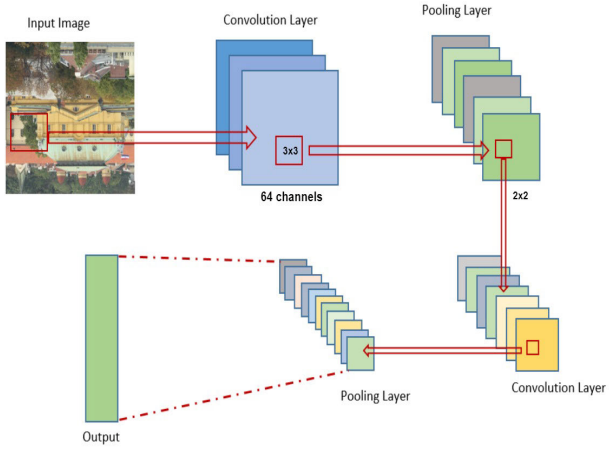

FIGURE 4. Framework of Channel-2.

applied. ReLU activation is summoned after each layer to drop all the negative values before it enters the next layer. The output of a convolution is passed to the set of two dance layers and then to the softmax layer of two units.

The self-correlation module is used for computing feature similarity. Pooling is also used for the collection of useful statistics about the feature. We are extracting the image feature (f_b^I) of size (16 × 16x512) from the CNN channel. We are applying deconvolution so that we can restore the original size of the image before applying further steps. Following a sequence of layers of type hidden and pooling, the data proceeds through several fully connected layers where the classification task is conducted. A binary classifier is employed to predict the mask image. This mask is generated using Inception and Bilinearuppool2D one after another same as done in [34] and produces tensor d_b^I with the dimension (256 × 256 × 6). Ultimately, within the output layer, the error is computed and subsequently backpropagated through the network to adjust the filter weights. This iterative process of feed-forward and backpropagation aims to minimize the error and train the network.

$$f_b^I = f_b^I|i_r, i_c|\text{where}, i_r, j_c \in [0 \ldots .15] \tag{5}$$

To get similarities between the copy-move regions, the similarity score is calculated by getting self-correlation between regions. Matched regions are located by the Max Pooling stage by collating meaningful statistics about regions.

C. FUSION CLASSIFIER

The two inputs to our fusion network are defined as follows: input A and input B. These inputs to the fusion network are the output of the two channels (256 × 256 × 6). The output filter dimension is six because of the last inception network, which concatenates three convolution 2D responses with two output filters (i.e., 2 × 3 = 6). BN-Inception net is used for combining the output of the two channels. Two more fully connected layers follow the above step. The first layer consists of two nodes, followed by a ReLU activation, but just one node with a linear activation is present in the second layer. The multi-input model builds the final step, and it defines the objectives of the model which are:

1) To Accept two inputs and
2) To define the output.

V. PROPOSED ALGORITHM FOR DETECTION OF FORGERY

The proposed approach for copy-move forgery detection utilizes a CNN architecture and CenSurE keypoint detection to extract features from the input image. The FREAK binary descriptor and Agglomerative Hierarchical Clustering algorithm are used to identify and match keypoints. Copy-move forged regions are identified as similar patches using a feature classifier. The training process requires a large number of images for diverse sample selection, and a single-class classifier is used to classify whether an image is forged or authentic. The functional steps of this approach are described in Algorithm (1).

Algorithm 1 CNN-CenSurE Based Copy-Move Forgery Detection

1: Input image (copy-move forgery image).
2: Input Image X to Key-point feature extraction (Channel-1)
3: Clustering of the Key-point
4: Feature matching and classification
5: Generate Binary classifier mask (output of channel-1)
6: Input image X to CNN feature extractor (Channel-2)
7: Compute self-co-relation for feature similarity
8: Percentile pooling is done to get statistics for matched patches
9: The four convolution layers consist of 16,32,64, and 128 kernels respectively. Each of these four convolution layers is followed by a pooling layer which reduces by a factor.
10: A dropout layers are added in between which switched off the neurons to find the path
11: Percentile pooling Binary classifier Mask is generated (output of channel-2)
12: Binary mask features from both Channel are taken to Fusion Copy-move perdition
13: Fuse feature using the BN-Inception
14: Predict the mask using a Conv2D with one filter followed by the softmax activation

1) Input Image: The first step in the algorithm is to provide an input image to both the CNN channel and the CenSurE detector channel.
2) CenSurE Detector Channel: The CenSurE detector channel identifies the keypoints in the input image. CenSurE stands for Center Surround Extrema. This algorithm identifies keypoints based on the scale-space extrema detection approach. The output of the CenSurE detector channel is a set of keypoints.
3) Feature Descriptor: The FREAK descriptor is used to create a binary descriptor for each keypoint identified in the previous step. This descriptor is used to match the keypoints in the next step.
4) Keypoint Matching: Agglomerative Hierarchical Clustering (AHC) is used to match the keypoints identified in step 3. The AHC algorithm undergoes iterations across keypoints, generating a collection of matched points. Singular points are excluded, and the remaining points are earmarked for subsequent processing.
5) Homographic Matrix Decomposition: Homographic matrix decomposition is applied to address high-level scale and rotation in the matched keypoints.
6) CNN Channel: The CNN channel extracts image features from the input image. The architecture of the CNN is designed to extract features that are useful for detecting copy-move forgeries. The output of the CNN channel is a feature map.

7) Copy-Move Forgery Classification: Copy-move forged regions are classified as similar patches using a feature classifier. The feature classifier is trained on diverse image samples, and it classifies each image as either authentic or forged.

The proposed CenSurE-FREAK keypoint-channel and CNN feature extraction-channel approach for copy-move forgery detection involves two main steps. The initial stage involves subjecting the input image to the CenSurE keypoint detector, aiming to detect keypoints and their respective descriptors. Subsequently, a matching operation is executed among the descriptor vectors of keypoints to detect analogous patches within the test image. In the second step, the input image is processed by the CNN architecture to extract image features. The convolution layer extracts edges, and the pooling layer reduces the dimensionality and extracts the dominant features of the image. Finally, the outputs of both channels are combined in the Fusion Classifier stage to include the benefits of both approaches, make the final decision of copy-move forgery, and localize the copy-move forged region. The fusion classifier outputs a superset of all the features of the image, which is helpful in detecting copy-move forgery in a variety of images.

The procedure of training VGG16 for copy-move forgery detection entails a methodical approach to endow the model with the capability to differentiate between genuine and altered areas within images. The initial pivotal stage involves creating a dataset that consists of photos containing authentic content as well as examples of copy-move forgeries. The dataset must be meticulously annotated to ensure precise and reliable ground truth information regarding the locations and characteristics of the forgeries. We have images with rotation, flipping, scaling, and modifications in brightness and contrast that improve the model's capacity to generalize across a wide range of forgery variants.

The training procedure involves exposing the model to the curated dataset, wherein it acquires the ability to differentiate between genuine and altered regions. The performance of the model is assessed on a validation set to guarantee that it can effectively apply to new and unexplored data. Refinement can be carried out by utilising the outcomes of the validation process to make adjustments to hyperparameters or modify the architecture. Subsequent evaluation on an independent test set is conducted to examine the accuracy, precision, recall, and F1-score of the model for detecting copy-move forgery

VI. DATASET AND EVALUATION METRIC

To assess the strength and efficacy of our forgery detection method across different types of copy-move forgery, we performed extensive experiments using a diverse range of images. These experiments involved the utilization of seven well-known open-source datasets, allowing us to thoroughly evaluate the performance and robustness of our approach.

A. DATASET USED

We have used seven open source datasets CMFD [35], GRIP [36], CoMoFoD [37], MICC-F600 [38], MICC-F220 [38], COVERAGE [39] and CASIA-II [40]. Table (1) provides an overview of the dataset details, while Table 2) presents specific information on the different levels of attacks found in CMFD and CoMoFoD datasets. Additionally, Table (3) lists the details of the datasets used for both training and testing purposes. Further elaboration on all seven datasets is provided below.

1) **CMFD:** The CMFD dataset consists of 48 authentic images with different textures. These images are affected by copy-move forgery with translation, rotation, scaling, and their combined effect. The translation is a simple copy-move without any angle rotation or scale change. All the forged and original images are processed with different types of post-processing operations like JPEG compression and additive Gaussian noise. As discussed in Table (2), this dataset has nine levels of JPEG compression from 100 to 20 and five levels of additive noise. The combination attack setup consisted of a 2° rotation, 1% scaling, and JPEG compression at a level of 80. The subsequent setups involved increasing the rotation and scaling while decreasing the JPEG quality, resulting in combinations of (4°, 3%, 75), (6°, 5%, 70), and (8°, 7%, 65).
2) **GRIP:** The GRIP dataset consists of 80 original and 80 forged images with corresponding ground truth images. It has a wide range of textures in images like smooth, coarse, and self-similar structural images of the monuments. Diversity in the textural property of the image makes this dataset challenging. Note that we have used primary images of GRIP dataset that has only simple copy-move images without any post-processing and geometrical transformation attacks.
3) **CoMoFoD:** The CoMoFoD dataset consists of 200 base images, with forged images in each of the following categories: translation, rotation, scaling, combination, and distortion. Furthermore, all forged images undergo post-processing operations including JPEG compression, additive noise, brightness change, color reduction, contrast enhancement, and image blur.
4) **MICC-F220:** The MICC-F220 dataset consists of 110 original and 110 forged images. The manipulation of copy-move is done by translation, rotation, scaling, or any combination of these three processes. Post-processing like JPEG compression or additive Gaussian noise is applied to images to hide traces of forgery. This dataset has single and multiple copy-move forged images.
5) **MICC-F600:** The MICC-F600 dataset comprises a total of 600 images, with 440 being original and 160 being forged images. The forged images are generated using copy-move forgery techniques, involving manipulations such as translation, rotation, scaling,

TABLE 1. Details of the dataset used to collect images for experimental work.

Dataset	Total Image	Image Size	Image Content	Image Format
CMFD	1632	388x2592 800x533	Outdoor places, Animals, Building, Indoor scene	PNG
CASIA II	12323	800x600 320x240	Indoor scenery, Plant Animal, Architecture	BMF, TIFF JPEG
GRIP	240	1024X768	Animal, Bird, Flowers, Building, Desert, Human, Sky.	PNG
MICC-F600	600	2048x1536	Flower, Animals, Human House, Natural scene	JPEG
MICC-F220	220	737x492	Beach, Humans, tree bookshelf, Road, mountain	JPEG
CoMoFoD	2160	3000X2000 512X512	Human, City view, Vehicles Grass, Wall, Roof, Building	JPEG PNG
COVERAGE	200	400X486	Indoor view, Rooms, Stores Public places, objects	TIFF

TABLE 2. Details of the range of different attacks applied on copy-move forged images.

Attacks	Criterion	CMFD	MICC-F600
Rotation	Degree	2 to 10, 180, 60	2 to 10
Scaling	Ratio	0.91 to 1.09	0.91 to 1.09
JPEG	Compression	100 to 20	100 to 20
AWGN	Noise Level	0.02 to 0.1	20 to 100

or a combination of these processes. Furthermore, post-processing techniques like JPEG compression and AWGN are applied to the forged images.

6) **Coverage:** The Coverage dataset contains 100 original, forged, and ground truth images with a similar object. This dataset consists of images from indoors and outdoors. Six different types of manipulations are applied to these images. These operations are translation, rotation, scaling, illumination change, free form, and any combination of these five. Also, it has 20 images with a combination of different copy-move forgeries.
7) **CASIA-II:** The CASIA-II dataset has a total of 7491 images and 5123 forged images. Out of all images, 3274 images are copy-move forged images, with various manipulation and post-processing. This dataset has images with translation, rotation, and scaling manipulations. Post-processing like JPEG compression and edge blurring is applied to some of these images.

B. EVALUATION METRIC

All experiments are conducted on a pixel level to ensure a comprehensive evaluation. True Positive (TP) signifies the total pixels correctly identified as forged that genuinely belong to the forged region. False Positive (FP) signifies the total pixels erroneously identified as forged, despite not being part of the actual forged region. Similarly, False Negative (FN) indicates total pixels incorrectly identified as not forged, even though they belong to the forged region. From these metrics, Precision (P), Recall (R), and F1-Score are calculated. F1-Score serves as the primary performance metric for evaluating the efficacy of our proposed approaches and also for comparison with other reported methods Its value ranges from 1 (best) to 0 (worst). We have shown it in terms of percentage (after multiplying it by 100) in this thesis. P, R and F1-Score are related to TP, FP and FN as under:

$$P = \frac{TP}{TP + FP}, \tag{6}$$

$$R = \frac{TP}{TP + FN}, \tag{7}$$

$$F1 = 2\frac{P \times R}{P + R}. \tag{8}$$

VII. EXPERIMENTS AND RESULTS

The performance evaluation of our proposed approach for forgery detection was conducted using the OpenCV-Python public library on a computer system comprising an Intel Core (TM) i7 processor, 8 GB RAM, and a 2.80 GHz CPU. In this section, we present the results of our evaluation, encompassing both quantitative and qualitative outcomes, for various types of forgeries. Additionally, we compare the performance of our approach with state-of-the-art techniques in the field of copy-move forgery detection.

A. PERFORMANCE EVALUATION

In our experiment, we aimed to assess the performance of detectors and descriptors on forged images that underwent various types of forgery and post-processing. We focused on nine distinct types of copy-move forgery attacks, which can be considered a comprehensive list of forgeries followed by post-processing attacks. Some of these attacks have been previously studied, allowing for comparisons with existing works. The types of copy-move forgery attacks considered in our study are as follows:

TABLE 3. Details of the image used for the training and testing phase of experimental work.

Dataset	CMFD	CASIA II	CoMoFoD	MICC F600	MICC F220	COVERAG	GRIP
Training Images	864	1200	1220	340	110	100	120
Testing Images	768	1000	940	260	110	100	120

1) Simple copy-move: This scenario involves directly pasting a copied region without any additional pr-processing, post-processing, or transformations. It is also referred to as translation.
2) Multiple copy-move: In this scenario, there are two possibilities. Firstly, the copied region is pasted multiple times in different regions of the image. Secondly, multiple regions are copied and pasted into the image.
3) Rotation: The copied region is rotated by a certain angle before pasting onto the image. We have taken images with small angles like 10°, 20°, 30° etc. and large angles like 60°, 120°, 180° also.
4) Flip: The copied region is rotated by 180° and pasted onto the image.
5) Scaling: Scaling can be performed in two ways: by adjusting the height or the width of the copied region. In our experiments, we used datasets that included images with varying percentages of up or down scaling applied to the copied region.
6) Combined transformation: In the case of combined transformation, multiple attacks, like rotation and scaling, are done to the targeted copied part of the image prior to pasting it. This technique is used to create more complex and realistic copy-move forgeries.
7) JPEG compression: In the JPEG compression attack, the copy-move region is subjected to lossy compression using the JPEG algorithm. The experiment involved using datasets with varying levels of compression applied to the forged images, allowing for evaluation of the method's performance under different compression settings.
8) Additive noise: Additive noise is a commonly employed attack to conceal evidence of forgery in copy-move forged images. The experimental datasets utilized in the study include forged images with different levels of AWGN added, creating a more challenging scenario for the detection of forgery.
9) Additional post-processing: In addition to the basic copy-move attack, post-processing techniques such as image blur, brightness change, color reduction, and contrast adjustment are commonly employed in combination with the copy-move attack to conceal evidence of forgery. The CoMoFoD dataset includes forged images with varying levels of these post-processing techniques, applied after the copy-move operation, to remove forgery footprints.

B. DISCUSSION OF EXPERIMENTAL RESULTS

In this section, we present the evaluation results of our proposed copy-move forgery detection technique on different types of attacks. The objective of copy-move forgery detection is to identify regions in the image that exhibit similarity, indicating the presence of forgery. Hence, we assess the performance of our technique on various types of attacks, including simple copy-move, rotation, flip, scaling, combined attacks, JPEG compression, AWGN, and additional post-processing attacks. We analyze the effect of these attacks on the accuracy of detection and localization of forged regions.

1) SIMPLE COPY-MOVE

Simple copy-move forgery detection refers to the identification of forgery where no post-processing or geometric transformation is applied. It involves the duplication of a specific region within an image and pasting it onto another region of the same image without any additional modifications. Our approach effectively detects and handles both single and multiple instances of copy-move forgery. In Figure (5), we have included some of the detected forgeries for single and multiple copy-move cases.

We conducted a comparative analysis of our proposed algorithm with several recently reported copy-move forgery detection approaches, and the results are presented in Table (4). Our approach consistently performed well across different datasets, as indicated by the high F1-Score values obtained. The proposed approach has shown improved results in the CMFD, GRIP, CoMoFoD, MICC-F600, and CASIA-II datasets compared to the keypoint-based approaches. However, in the MICC-F220 and COVERAGE datasets, the performance of the CNN-CenSurE based approach is slightly lower.

TABLE 4. Comparison of our results (F1-Score) with recently published works for simple copy-move forgery in images.

Dataset	Bi [41]	Li [42]	Diwan [10]	Proposed
CMFD	92.87	98.91	97.61	97.99
GRIP	92.98	100	95.12	98.15
MICC-F600	–	91.50	97.14	97.64
MICC-F220	–	99.10	98.43	97.64
CoMoFoD	–	–	98.43	98.56
COVERAGE	–	72.28	97.50	96.85
CASIA-II	–	–	95.36	97.38

Our approach outperforms recently published keypoint-based works [10], [41] in terms of detection results on the GRIP dataset. This improvement can be attributed to the proposed approach's capability to effectively detect forgery in images with diverse textures. However, it is worth noting that Li's approach [42] demonstrates exceptionally high performance on the GRIP dataset, surpassing the proposed approach. We compared our work with neural network-based work proposed by Zhu et al. [27] besides others. From the results of [27] we can infer that the use of only neural networks is not helping detection and localization in diverse images. In a comparison of time complexities, our proposed approach surpasses other methods in image processing. It processes images in just 15.17 seconds, showcasing superior efficiency compared to Li's approach (86.6 seconds), Bi's approach (74.17 seconds), and Diwan's approach (15.75 seconds). This rapid processing significantly enhances the effectiveness of forgery detection, especially for real-time applications.

The proposed CNN and CenSurE fusion approach includes the benefits of image keypoints and image features extracted by the CNN model, which makes detection and localization even better. Figure (5) illustrates examples of images for copy-move forgery detection.

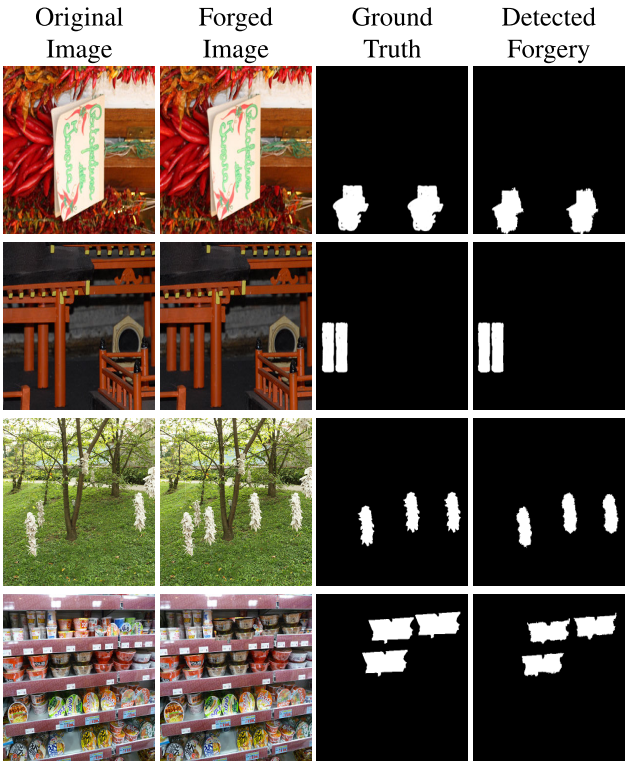

FIGURE 5. Detection results for single and multiple copy-move forgeries from CMFD, CoMoFoD and MICC-F600 datasets.

Image Texture: As we know if we use only keypoint information of the image, extraction of useful keypoints for extreme image texture is difficult. Like in a smooth image number of keypoints is so low that they are not sufficient to detect and localize forgery. Whereas in highly textured images with self-similar structures, there is a large number of similar keypoints that create false localization of the tempered region. In our proposed approach, since we used a fusion of keypoint features of images with the neural network, it is evident from Figure (6), that it was quite effective in forgery detection and localization for diverse images.

Multiple Copy-Move: As mentioned in Section (VII-A), we address multiple copy-move forgery explicitly, some example images for multiple copy-move forgery detection are shown in Figure (7) and Figure (5) in the following section.

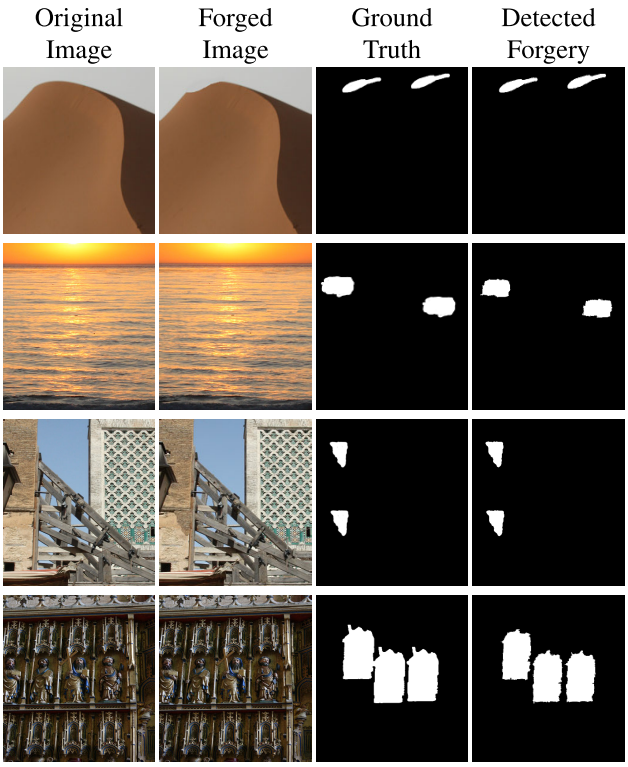

FIGURE 6. Detection results for images with different textures from GRIP, CMFD and MICC-F600 datasets.

2) DETECTION OF POST-PROCESSED COPY-MOVE

Skilled forgery of digital images often involves applying various post-processing attacks to hide the traces of manipulation. In our study, we investigated several commonly used post-processing attacks, such as JPEG compression, noise addition, brightness change, color reduction, contrast adjustment, and image blurring. These operations are extensively used for hiding copy-move forgery in a digital image. We analyzed the impact of these attacks on forged images to understand their effects and develop effective detection methods.

CNNs are powerful deep-learning models capable of learning complex patterns and features from images. They can automatically learn and extract relevant features from images, including those affected by post-processing attacks. By training a CNN on a dataset that includes post-processed copy-move forged images, the network can learn to identify the subtle changes and artifacts introduced by different

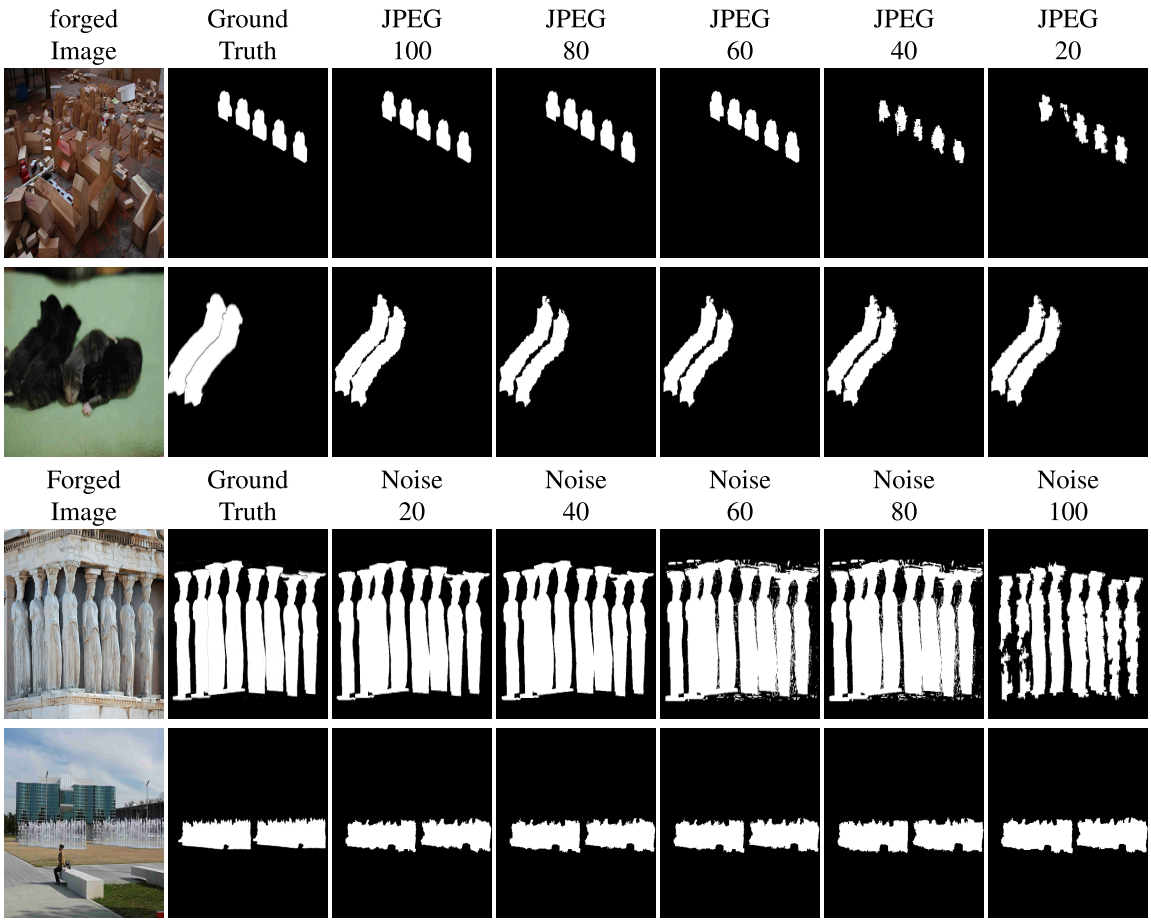

FIGURE 7. Detection results for Copy-move images with low to high JPEG compression and low to high additive noise for CMFD and CoMoFoD datasets.

post-processing techniques. This enables the CNN to detect and classify forged regions accurately. CenSurE keypoint detector, on the other hand, is designed to extract distinctive local features from an image. These keypoints represent areas with significant variations in intensity. Indeed, the robustness of the combined CNN and CenSurE approach plays a crucial role in detecting copy-move forgery, even in the presence of post-processing techniques like JPEG compression and noise addition.

JPEG Compression: We are taking post-processed images from CMFD and CoMoFoD datasets for the experiment. For robust experimentation of our approach, we have used six levels of JPEG compression, e.g., JPEG 100, 90, 80, 70, 60, and 40, the lesser the number more the compression. As shown in Table (5) for both datasets results gradually deteriorated when the compression level in images went on increasing. As expected, as we move from JPEG 100 to JPEG 40, the F1-Score deteriorated, but still, we got good results in all cases for both datasets. We know that as compression increases high-frequency information like edges, corners, and gradient change of the image gradually gets smooth out. This high-frequency information in the image is the key feature of the image that was used by the image feature detectors.

TABLE 5. The average result of copy-move forgery detection for images with various JPEG and Noise levels for CMFD and CoMoFoD datasets and its comparison with keypoint-based approach [10].

Attack	Value	F1-Score		F1-Score	
		Proposed CMFD	Diwan [10] CMFD	Proposed CoMoFoD	Diwan [10] CoMoFoD
JPEG Quality	100	98.11	94.87	96.98	94.51
	90	97.99	94.87	96.61	93.92
	80	97.89	93.96	96.01	93.75
	70	97.36	93.61	95.79	92.80
	60	96.27	92.88	95.40	91.77
	40	94.81	90.18	93.57	90.53
Noise Level	20	96.34	91.90	96.28	90.89
	40	96.01	91.90	–	–
	60	95.21	91.90	95.41	90.36
	80	93.96	91.01	–	–
	100	92.99	90.88	94.71	89.91

In JPEG compressed images, when the compression level is below 60, image visual degradation and blocking artifact were noticeably visible. This factor is affecting forgery detection in images with higher compression. However, CNN architecture curtails the effect of higher compression by detecting low-level image features. The CNN can learn to recognize the statistical patterns and artifacts introduced

by JPEG compression. As shown in the Table (5) the keypoint-based approach proposed by Diwan et al. [10], yielded significantly lower results, particularly for higher levels of compression, while our proposed approach effectively detected forgery even under challenging compression conditions. The first two rows of the Figure (7) showcase images with JPEG compression with different levels of compression.

Additive Noise: Additive noise is a commonly used post-processing attack aimed at concealing image forgeries by introducing random pixel values. The introduction of noise in an image can generate undesired edges or corners, which can disrupt the background of the image. The next experiment in post-processed copy-move forgery detection was to see the impact of additive noise in images after the simple copy-move operation. The miscellaneous edges and corners created during noise addition become severe when the noise level increases, it creates a blur effect. These miscellaneous edges affected the detection of keypoints but multi-level CNN architecture extracted detailed image features, which in turn helped in better detection and localization of forgery after the fusion classifier stage of the proposed approach.

Table (5) illustrates the results of forgery detection for copy-move images with added noise. The findings demonstrate that our proposed approach surpasses the keypoint-based method introduced by Diwan et al. [10] in terms of accuracy, particularly for lower levels of noise. This highlights the effectiveness of our approach in detecting forgery in the presence of noise. However, it is worth noting that the performance of our approach diminishes for higher levels of noise. The last two rows of the Figure (7) showcase images with additive noise with different levels of compression.

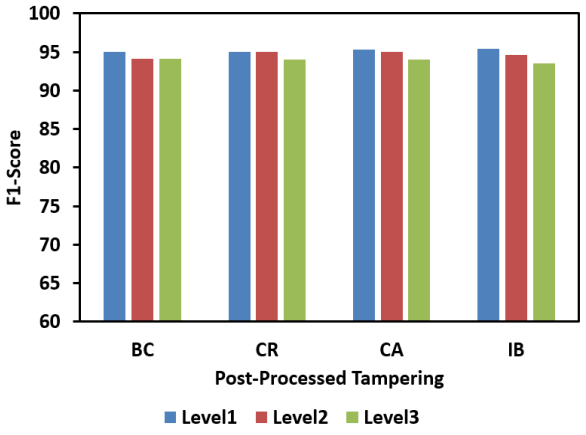

FIGURE 8. Copy-move forgery detection results for images with additional post-processing, e.g., Brightness change (BC), Colour reduction (CR), Contrast adjustment (CA), and Image blur (IB) for the CoMoFoD dataset.

3) ADDITIONAL POST-PROCESSING

The proposed approach for copy-move forgery detection exhibits robustness against additional post-processing techniques, such as brightness change, color reduction, contrast enhancement, and image blur. These operations make changes in the image at the pixel level and camouflage traces of forgery and are commonly employed to conceal traces of copy-move forgery in images. By modifying the pixel values, these operations make it challenging to detect forgery. For example, increasing the brightness of a forged image can decrease contrast, leading to a higher number of false negatives and a decrease in the recall, ultimately reducing the overall detection accuracy or F1-Score. Similarly, in the case of color reduction, reducing intensity levels can impact edge detection and detection accuracy.

In our experiments, we evaluated the performance of the proposed CNN-CenSurE-based approach on additional post-processing techniques using the CoMoFoD dataset. The results indicate that our approach effectively detects and localizes forgery across different levels of these post-processing operations. Figure (8) illustrates the consistent results obtained even when the processing levels range from soft to harsh. The CoMoFoD dataset includes forged images with three levels of additional post-processing: level1, level2, and level3. The label3 corresponds to the harshest processing, which is more severe compared to level2 and level1. Similarly, level1 represents the lighter or softer processing in comparison to level2 and level3. This pattern holds true for contrast adjustment, color reduction, and brightness change as well.

By clustering keypoints and incorporating CNN features, our approach successfully preserves similarity in keypoints and registers local image features, enhancing the overall effectiveness of forgery detection and localization. Our proposed approach shows effective forgery detection capabilities even with varying levels of post-processing, from soft to harsh. However, as depicted in Figure (8), the level of post-processing becomes more severe, leading to a lower overall F1-Score. This is mainly due to the fact that harsher post-processing attacks, like high levels of image Blur or contrast adjustment, can obscure keypoint information and make it more difficult to accurately detect copy-move forgery.

4) DETECTION OF GEOMETRICALLY TRANSFORMED COPY-MOVE

Our research now aims to detect a forgery in images where the copy-move regions have undergone rotation or scaling. Detection of forgery becomes further complicated when the copied region is geometrically transformed before moving it. These transformations are commonly employed to make forgery appear realistic. However, not all image features are invariant to such transformations, posing a challenge for detection. We specifically focus on scenarios where the copied region is rotated by large angles, such as 40°, 60°, and 180°, as well as cases involving significant scaling combined with rotation, known as a combined attack.

Copy-Move With Angle Rotation: Detection of forgery becomes increasingly challenging when copy-move regions

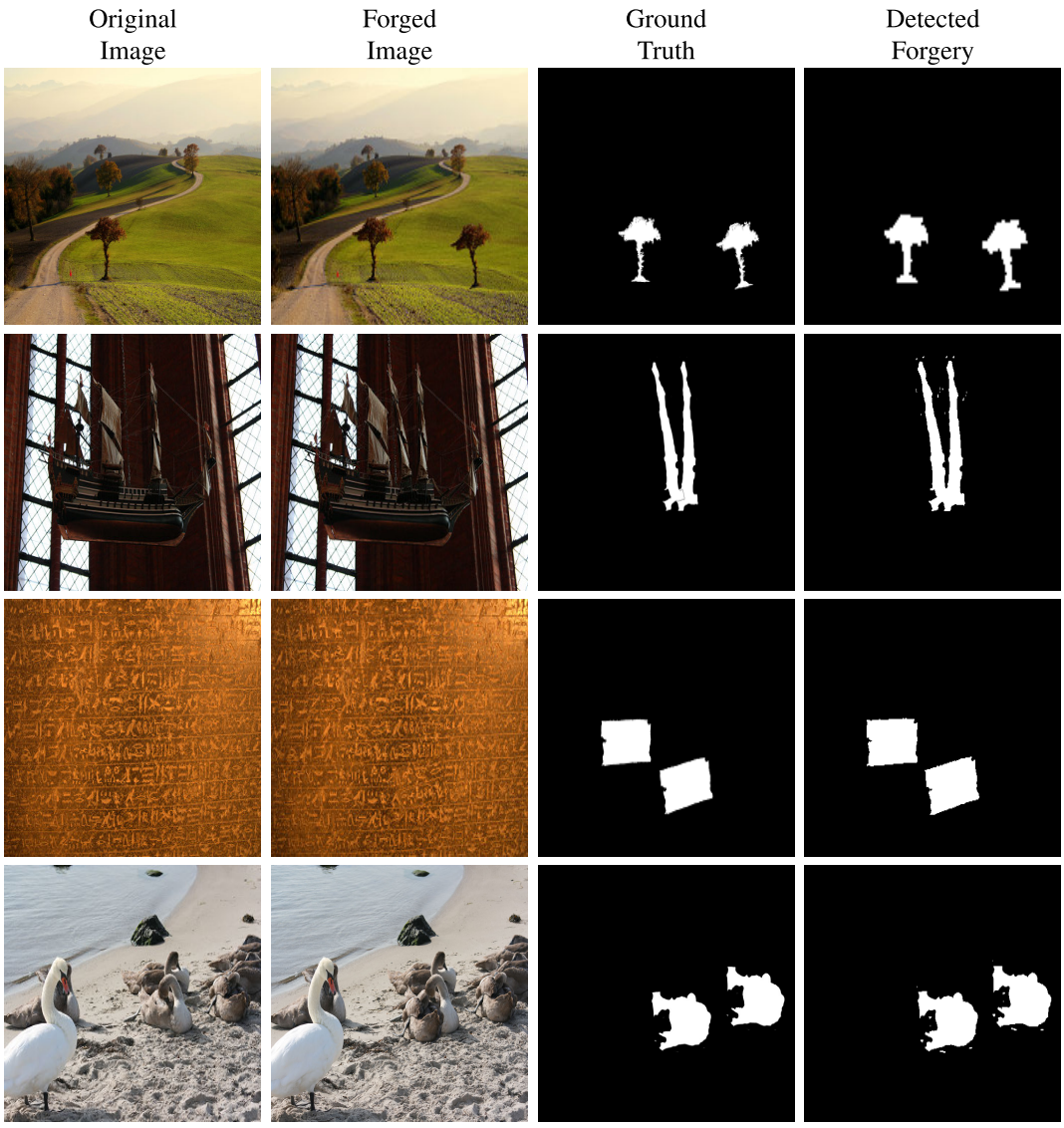

FIGURE 9. Detection results for copy-move images with rotation as the geometrical transformation from CMFD, CoMoFoD and MICC-F600 datasets.

are affected by rotation, particularly at higher degrees of rotation. To address this challenge, our approach leverages the features extracted from a CNN architecture and utilizes the CenSurE keypoint detector. By combining these two components, our method benefits from both the discriminative power of CNN features and the robustness of CenSurE keypoints in capturing important information from forged regions. The statistical information extracted from these features boots the effectiveness of forgery detection for the images which involves angle rotation.

Our approach demonstrates successful forgery detection in images with small degrees of rotation, such as 2°, 4°, 6°, and 10°. Notably, our approach excels in detecting and localizing forgery even at higher degrees of rotation, such as 20°, 40°, 60° and 180° rotation, surpassing the performance of the method proposed by Diwan et al. [10]. This highlights the effectiveness of our approach in addressing the complexity of detecting forgery with higher degrees of rotation, an aspect that has not been extensively studied by previous researchers. The superior results obtained for higher degrees of rotation validate the robustness and capability of our approach in handling such challenging forgery scenarios.

However, when examining Table (6), it is evident that as the rotation angle increases, the F1-Score decreases. This decline is attributed to the loss of correspondence among the extracted image feature, making it more challenging to detect and localize the forgery accurately. Despite this decrease in performance, the proposed algorithm maintains stable computational complexity across various attack conditions, including large geometrical transformations like 180° rotation (flip). The stable computational complexity contributes to faster computation, allowing for efficient forgery detection in real-time applications. Some images with rotation are shown in Figure (9).

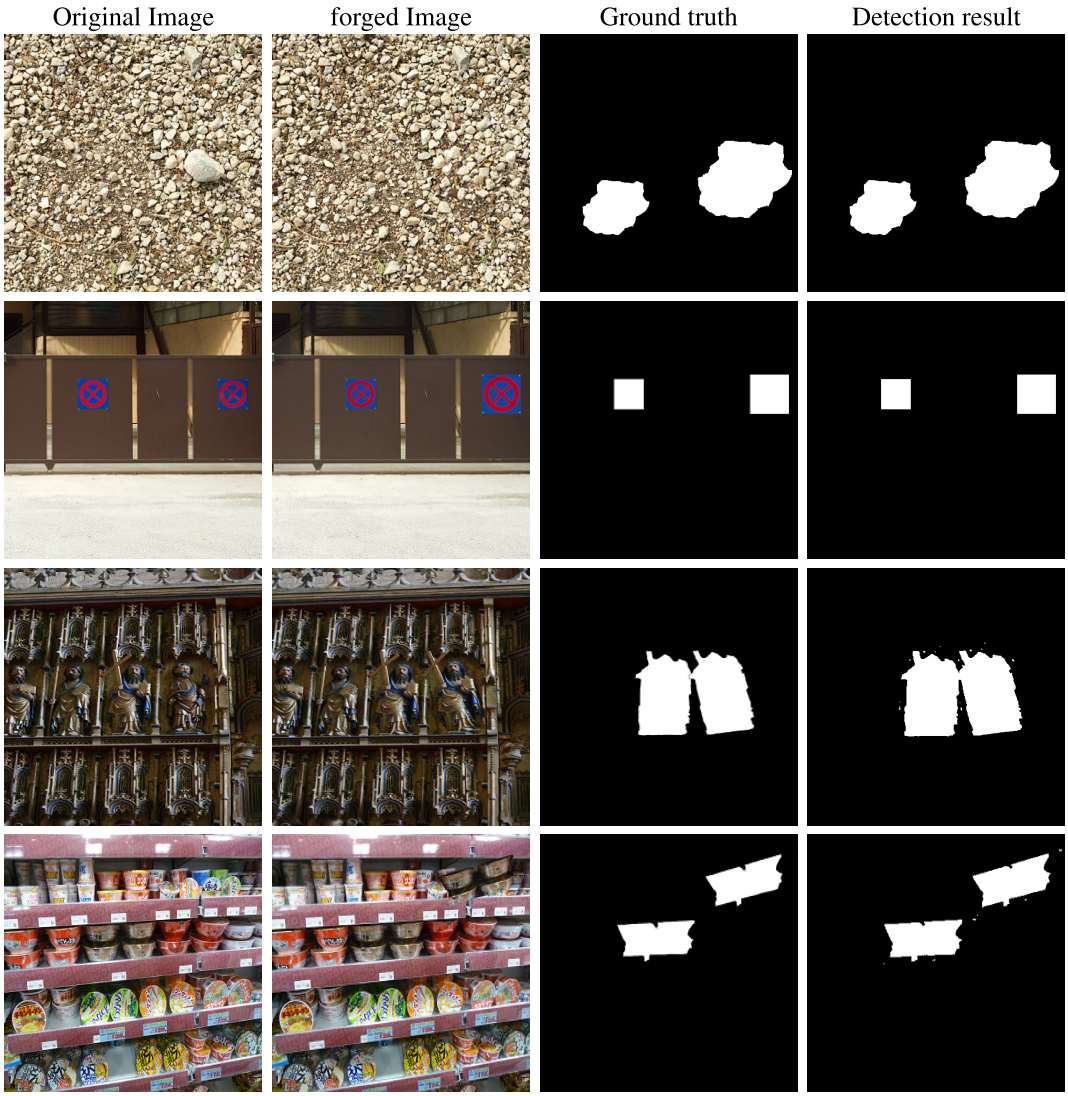

FIGURE 10. Detection results for copy-move images with scale and combined attack as the geometrical transformation from CMFD, CoMoFoD and MICC-F600 datasets.

TABLE 6. The average result of copy-move forgery detection for images with various levels of angle rotation for CMFD dataset and its comparison with the keypoint-based approach [10].

Angle	F1-Score		Angle	F1-Score	
	CMFD			CoMoFoD	
	Proposed	Diwan [10]		Proposed	Diwan [10]
2	98.19	94.91	20	94.67	89.88
4	97.99	94.05	40	93.99	87.65
6	96.98	94.99	60	93.81	87.65
8	96.55	93.90	180	93.77	84.88
10	94.99	93.88	–	–	–

Copy-Move With Scale Change: The accuracy of copy-move forgery detection is significantly affected when there are changes in the scale of the manipulated region. This challenge becomes even more prominent when scaling is combined with rotation, resulting in a combined attack. As mentioned in section (VI-A), even small-scale changes and rotations in combined attacks may not be visibly apparent in the copy and move regions. However, they disrupt the correspondence between these regions, posing a significant challenge to forgery detection. Moreover, the dataset includes a small amount of JPEG compression, further deteriorating the correspondence between the copied and moved regions and exacerbating the difficulty of detection. Consequently, the detection accuracy is considerably compromised in such scenarios, highlighting the importance of addressing this critical issue in forgery detection algorithms.

Our proposed approach combines the strengths of CNN and CenSurE keypoints to boost the detection and localization of forgery, particularly in cases involving angle rotation and scale change. By utilizing the pooling layers of CNN, the dominant features of the images, which are rotation and scale-invariant, are extracted. This combination of CNN and CenSurE keypoints strengthens the detection and

localization capabilities, especially for scenarios where large scale changes are involved. The results presented in Table (7) demonstrate the effectiveness of our approach in detecting forgery.

Table (7) provides insights into the impact of scale change and combined attacks on the performance of our approach. It is worth noting that our proposed approach consistently outperforms the approach presented by Diwan et al. [10] in terms of detecting forgery with scale changes, especially for scale variations of 2%, 4%, 6%, 8% and 10%. These results indicate that our approach excels in handling scale variations and demonstrates superior performance in detecting copy-move forgery under such conditions.

The impact of scale changes on detection accuracy is evident from the results obtained in our experiments. However, our proposed approach exhibits strong capabilities in detecting forgery even in the presence of complex combined attacks. This is demonstrated by the experiments run on the CMFD dataset, and the accuracy of our forgery detection is visually represented in Figure (10). The figure showcases the precise identification of forged regions, with the first two rows depicting results for scaled copy-move forgery and the last two rows showcasing the detection performance for combined scale and rotation attacks. These results serve as evidence of the robustness and effectiveness of our approach in handling complex scenarios of copy-move forgery.

TABLE 7. The average result of copy-move forgery detection for images with various levels of scale factor and its comparison with the keypoint-based approach [10].

Sacle Factor	F1-Score Proposed	F1-Score Diwan [10]
2	97.99	93.49
4	97.94	93.95
6	95.83	92.46
8	94.92	92.01
10	94.28	92.88

VIII. CONCLUSION

The proposed end-to-end trainable copy-move fusion approach incorporated the advantages of CenSurE keypoint and CNN architecture to detect and localize the copy-move forgery in digital images. It employs a data-driven approach that allows the algorithm to update its learning through training data continually. Creating a fusion of keypoints with CNN features enhances forgery detection in the presence of different copy-move forgery attacks.

The CNN-CenSurE approach works pretty well on all varieties of copy-move forgery including simple copy-move (single and multiple), post-processed copy-move (JPEG and Noise), geometrically transformed copy-move (angle rotation and scale change), and some additional processing like brightness change, colour reduction, contrast adjustment, and image blur. We can say that we could meet our objective of making a robust approach that can detect a forgery in all types of copy-move forged images. We specifically addressed images with various textures like smooth and coarse images with dense textures. We are getting stable results in images with different attacks, which makes our detection approach useful for a diverse range of forged images. Extensive experiments on seven datasets available in the public domain have been carried out, demonstrating the superior performance of the proposed CNN-based fusion approach over other image forgery detection approaches. This algorithm takes very little time to process, so it can be used for faster detection.

REFERENCES

[1] H. Farid, "Digital forensics in a post-truth age," *Forensic Sci. Int.*, vol. 289, pp. 268–269, Aug. 2018.

[2] H. Ding, L. Chen, Q. Tao, Z. Fu, L. Dong, and X. Cui, "DCU-net: A dual-channel U-shaped network for image splicing forgery detection," *Neural Comput. Appl.*, vol. 35, no. 7, pp. 5015–5031, Mar. 2023.

[3] 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.

[4] S. Weng, T. Zhu, T. Zhang, and C. Zhang, "UCM-net: A U-net-like tampered-region-related framework for copy-move forgery detection," *IEEE Trans. Multimedia*, vol. 26, pp. 750–763, 2024.

[5] A. Diwan, A. K. Roy, and S. K. Mitra, "Locality preserving projection based multiple copy-paste forgery detection," in *Proc. IEEE Appl. Signal Process. Conf. (ASPCON)*, Oct. 2020, pp. 158–162.

[6] R. Agarwal, D. Khudaniya, A. Gupta, and K. Grover, "Image forgery detection and deep learning techniques: A review," in *Proc. 4th Int. Conf. Intell. Comput. Control Syst. (ICICCS)*, May 2020, pp. 1096–1100.

[7] H.-T. Wang and P.-C. Su, "Deep-learning-based block similarity evaluation for image forensics," in *Proc. IEEE Int. Conf. Consum. Electron. (ICCE-Taiwan)*, Sep. 2020, pp. 1–2.

[8] A. Diwan, P. A. Koringa, A. K. Roy, and S. K. Mitra, "Neighbourhood projection embedding based image tampering detection and localization," in *Computer Vision, Pattern Recognition, Image Processing, and Graphics*, R. V. Babu, M. Prasanna, and V. P. Namboodiri, Eds. Singapore: Springer, 2020, pp. 387–396.

[9] A. Kashyap, K. D. Tyagi, and V. B. Tyagi, "Robust and optimized algorithm for detection of copy-rotate-move tempering," *IEEE Access*, vol. 11, pp. 66626–66640, 2023.

[10] A. Diwan, R. Sharma, A. K. Roy, and S. K. Mitra, "Keypoint based comprehensive copy-move forgery detection," *IET Image Process.*, vol. 15, no. 6, pp. 1298–1309, 2021. [Online]. Available: https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/ipr2.12105

[11] K. H. Rhee, "Generation of novelty ground truth image using image classification and semantic segmentation for copy-move forgery detection," *IEEE Access*, vol. 10, pp. 2783–2796, 2022.

[12] A. Diwan, V. Mall, A. Roy, and S. Mitra, "Detection and localization of copy-move tampering using features of locality preserving projection," in *Proc. 5th Int. Conf. Image Inf. Process. (ICIIP)*, Nov. 2019, pp. 397–402.

[13] A. Diwan, D. Kumar, R. Mahadeva, H. C. S. Perera, and J. Alawatugoda, "Unveiling copy-move forgeries: Enhancing detection with SuperPoint keypoint architecture," *IEEE Access*, vol. 11, pp. 86132–86148, 2023.

[14] S. B. G. T. Babu and C. S. Rao, "Copy-move forgery verification in images using local feature extractors and optimized classifiers," *Big Data Mining Analytics*, vol. 6, no. 3, pp. 347–360, Sep. 2023.

[15] K. M. Hosny, H. M. Hamza, and N. A. Lashin, "Copy-for-duplication forgery detection in colour images using QPCETMs and sub-image approach," *IET Image Process.*, vol. 13, no. 9, pp. 1437–1446, Jul. 2019.

[16] G. Gani and F. Qadir, "A robust copy-move forgery detection technique based on discrete cosine transform and cellular automata," *J. Inf. Secur. Appl.*, vol. 54, Oct. 2020, Art. no. 102510.

[17] X.-Y. Wang, C. Wang, L. Wang, L.-X. Jiao, H.-Y. Yang, and P.-P. Niu, "A fast and high accurate image copy-move forgery detection approach," *Multidimensional Syst. Signal Process.*, vol. 31, no. 3, pp. 857–883, Jul. 2020.

[18] A. Diwan and U. Sonkar, "Visualizing the truth: A survey of multimedia forensic analysis," *Multimedia Tools Appl.*, pp. 1–28, Oct. 2023.

[19] N. Kumar and T. Meenpal, "Salient keypoint-based copy–move image forgery detection," *Austral. J. Forensic Sci.*, vol. 55, no. 3, pp. 331–354, May 2023.
[20] S. I. Lee, J. Y. Park, and I. K. Eom, "CNN-based copy-move forgery detection using rotation-invariant wavelet feature," *IEEE Access*, vol. 10, pp. 106217–106229, 2022.
[21] A. K. Venugopalan and G. Gopakumar, "Keypoint-based detection and region growing-based localization of copy-move forgery in digital images," in *Computer Vision and Machine Intelligence: Proceedings of CVMI 2022*. Cham, Switzerland: Springer, 2023, pp. 513–524.
[22] X.-Y. Wang, X.-Q. Wang, P.-P. Niu, and H.-Y. Yang, "Accurate and robust image copy-move forgery detection using adaptive keypoints and FQGPCET-GLCM feature," *Multimedia Tools Appl.*, vol. 83, no. 1, pp. 2203–2235, Jan. 2024.
[23] K. H. Rhee, "Composition of visual feature vector pattern for deep learning in image forensics," *IEEE Access*, vol. 8, pp. 188970–188980, 2020.
[24] L. Xiong, J. Xu, C.-N. Yang, and X. Zhang, "CMCF-net: An end-to-end context multiscale cross-fusion network for robust copy-move forgery detection," *IEEE Trans. Multimedia*, pp. 1–13, 2023.
[25] Y. Zhang, G. Zhu, X. Wang, X. Luo, Y. Zhou, H. Zhang, and L. Wu, "CNN-transformer based generative adversarial network for copy-move source/target distinguishment," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 5, pp. 2019–2032, May 2023.
[26] N. Kaur, N. Jindal, and K. Singh, "A deep learning framework for copy-move forgery detection in digital images," *Multimedia Tools Appl.*, vol. 82, no. 12, pp. 17741–17768, May 2023.
[27] 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.
[28] O. Kuznetsov, E. Frontoni, L. Romeo, and R. Rosati, "Enhancing copy-move forgery detection through a novel CNN architecture and comprehensive dataset analysis," *Multimedia Tools Appl.*, pp. 1–35, Jan. 2024.
[29] P. Mer, R. Sunil, and A. Diwan, "From traditional to deep: A survey of image forgery detection techniques," in *Proc. IEEE 3rd Appl. Signal Process. Conf. (ASPCON)*, Nov. 2023, pp. 103–108.
[30] M. Agrawal, K. Konolige, and M. R. Blas, "CenSurE: Center surround extremas for realtime feature detection and matching," in *Proc. Comput. Vis. ECCV*, D. Forsyth, P. Torr, and A. Zisserman, Eds. Berlin, Germany: Springer, 2008, pp. 102–115.
[31] A. Alahi, R. Ortiz, and P. Vandergheynst, "FREAK: Fast retina keypoint," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 510–517.
[32] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
[33] M. A. Elaskily, M. M. Dessouky, O. S. Faragallah, and A. Sedik, "A survey on copy move forgery detection (CMFD) technique," in *Proc. Int. Conf. Intell. Innov. Technol. Comput., Electr. Electron. (IITCEE)*, Jan. 2023, pp. 439–443.
[34] K. Zhao, X. Yuan, T. Liu, Y. Xiang, Z. Xie, G. Huang, and L. Feng, "CAMU-net: Copy-move forgery detection utilizing coordinate attention and multi-scale feature fusion-based up-sampling," *Expert Syst. Appl.*, vol. 238, Mar. 2024, Art. no. 121918.
[35] 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.
[36] D. Cozzolino, G. Poggi, and L. Verdoliva, "Copy-move forgery detection based on PatchMatch," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 5312–5316.
[37] D. Tralic, I. Zupancic, S. Grgic, and M. Grgic, "CoMoFoD—New database for copy-move forgery detection," in *Proc. ELMAR-*, Sep. 2013, pp. 49–54.
[38] 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.
[39] 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)*, Sep. 2016, pp. 161–165.
[40] J. Dong, W. Wang, and T. Tan, "CASIA image tampering detection evaluation database," in *Proc. IEEE China Summit Int. Conf. Signal Inf. Process.*, Jul. 2013, pp. 422–426.
[41] X. Bi and C.-M. Pun, "Fast copy-move forgery detection using local bidirectional coherency error refinement," *Pattern Recognit.*, vol. 81, pp. 161–175, Sep. 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0031320318301183
[42] Y. Li and J. Zhou, "Fast and effective image copy-move forgery detection via hierarchical feature point matching," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 5, pp. 1307–1322, May 2019.

ANJALI DIWAN (Senior Member, IEEE) received the Ph.D. degree from the Dhirubhai Ambani Institute of Information and Communication Technology (DA-IICT), Gandhinagar, Gujarat. She is currently a Faculty Member with the CE-AI/Big Data Department, Marwadi University, Rajkot, Gujarat, India. She is a highly experienced academic and software industry professional with over 19 years of expertise. Her areas of academic and research interests include machine learning, image processing, artificial intelligence, deep learning, data security, multimedia forensics, and the application of technologies to address humanitarian challenges. She also serves as a member for the SAC Team of IEEE R10 (2023–2024) and the Section Chair for the IEEE Young Professionals Affinity Group of Gujarat Section (2022–2024).

ANIL K. ROY (Senior Member, IEEE) was with the Centre for Theoretical Studies, Indian Institute of Science, Bangalore, India, from 1993 to 1994, followed by the Optical Fiber Group, IIT Delhi, India, till 1996. He is currently a Faculty Member with DA-IICT, Gandhinagar, India, is an alumnus of IIT Delhi and IIT Roorkee. His research interests include sensors, the Internet of Things, image processing, fiber optics, semiconductor physics, and applications of technologies for humanitarian challenges. Some of the notable awards which he won are: the 2019 IEEE India Council Section Chair Lifetime Achievement Award, the 2017 IEEE Sensors Council Meritorious Service Award, the 2012 IEEE MGA Leadership Award, and the 2010 IEEE Region 10 Outstanding Volunteer Award. He is also the VP of the Technical Operations of IEEE Sensors Council (2022–2023). He was the Chair of the IEEE Conference Publications Committee (2019–2020). He has been involved in various efforts of spreading awareness about cyber security among students, parents, and professionals as well. He is the General Co-Chair of many international conferences, such as INDICON 2009, IEEE SENSORS 2018, IEEE APSCON 2023, and IEEE R10 HTC 2023, and the Technical Program Committee Co-Chair of INDICON 2019, IEEE International Symposium on Smart Electronic Systems (iSES 2023), and IEEE APSCON 2024.