Tampering Detection and Segmentation Model for Multimedia Forensic

Manjunatha S[1], Malini M Patil[2], Swetha M D[3], Prabhu Vijay S S[4]
Dept. of Information Science & Engineering, Global Academy of Technology (Affiliated to Visvesvaraya Technological University, Belagavi-590018), Bengaluru, Karnataka, India[1]
Dept. of Computer Science & Engineering, RVITM (Affiliated to Visvesvaraya Technological University, Belagavi-590018), Bengaluru, Karnataka, India[2]
Dept. of Computer Science & Engineering, BNMIT (Affiliated to Visvesvaraya Technological University, Belagavi-590018), Bengaluru, Karnataka, India[3]
Dept. of Information Science & Engineering, BMSCE, Bengaluru, Karnataka, India[3]
Senior Software Engineer and Data Analyst, Navshyatechnologies, Bengaluru, Karnataka, India[4]

***Abstract*—When an image undergoes hybrid post-processing transformation, detecting tamper region, localizing it and segmentation becomes very difficult tasks. In particular, when a copy-move attack with hybrid transformation has similar contrast and illumination parameters with an authenticated image it makes tamper detection difficult. Alongside, under small-smooth attack existing tamper identification model provides a very poor segmentation outcome and sometimes fails to identify an image as tampered. This article focused on addressing the difficulty through the adoption of the Deep Learning model. The proposed technique is efficient in detecting tampering with good segmentation outcomes. However, existing models fail to distinguish adjacent pixels' relationships affecting segmentation outcomes. In this paper, an Improved Convolution Neural Network (ICNN) assuring correlation awareness-based Tamper Detection and Segmentation (TDS) model for image forensics is presented. This model brings good correlation among adjacent pixels through the introduction of an additional layer namely the correlation layer alongside vertical and horizontal layers. The TDS-ICNN is very effective in localizing and segmenting tamper regions even under small-smooth post-processing tampering attacks by using a feature descriptor built using aggregated three-layer ICNN architecture. An experiment is done to study TDS-ICNN with other tamper identification models using various datasets such as MICC, Coverage, and CoMoFoD. The TDS-ICNN is very efficient under different post-processing hybrid attacks when compared with existing models.**

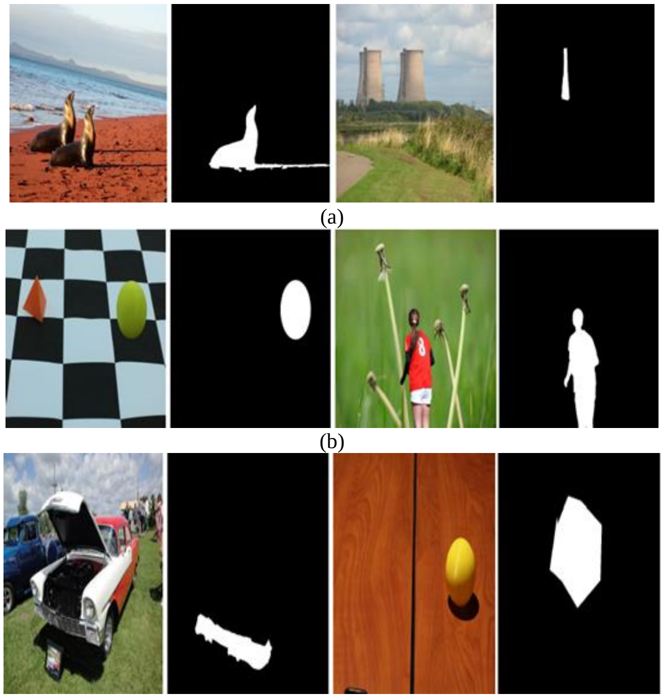
Keywords*—Convolution neural networks; digital image forensic; hybrid image transformation; resampling feature; segmentation

I. INTRODUCTION

Image authentication methods are characterized in the following two classes: (1) Active and (2) Blind or Passive. Digital Watermarking has been proposed as an active method using which an image can be authenticated [1]. The main aim of watermarking is to ensure the protection of copyright, authentication of content, ownership recognition, and data integrity. Watermarking ensures content from modification only and also provides data integrity and content authentication. Watermarks generally are indivisible from the digitized picture element they are embedded in. Further, the watermarks undergo a similar transformation in the picture. The major drawback of using watermarking is that it prerequisites watermark to be embedded during capturing of the image. This also binds/restricts its applicability to real-time environment usage. Thus, they are used only in controlled surroundings such as in armed forces and surveillance environments. Furthermore, some watermarks may break down the image quality.

Passive or blind forgery detection considers images without any digital signature, digital watermark, or any other prior information and checks the authenticity and origin of the image. Image forgery may not leave any visual clues of tampering being done. But there are high chance that it most probably perturbs the underlying statistical characteristics of an original image or modifies the scene of an image. These inconsistencies are utilized for tampering identification. Since this method doesn't require any prior knowledge of the picture. Passive forgery authentication techniques are further divided into forgery-dependent techniques and forgery-independent techniques as shown in Fig. 1. Forgery-dependent methodologies are delineated to identify a particular class of tampering for example splicing, copy-clone, etc. which relies on the forgery class type used on a picture. Whereas latter, the independent methods identify tampering using artifact traces left in the procedure of carrying out light inconsistencies and re-sampling. Existing forgery detection techniques recognize various traces of forged segments and identify them and the forged segment is localized [2].

Over many years, several attempts have been made for the classification of whether given images are authenticated or forged. Nonetheless, just a couple of works [3] endeavor to localize tampering at the pixel level. Recent methodologies [4] have aimed at addressing the localization issue by characterizing patches as tampered. Establishing the location of the tampered region ring is an exceptionally difficult job and also well-crafted tampering of pictures doesn't leave any visual hints. A sample repetition of well-crafted image tampering is shown in Fig. 1, where image one and three defines the tampered image, and image two and four defines the ground truth of the respective image that has been forged through transformation attacks. In Fig. 1(a), a copy-clone attack is presented where a set of objects is copied and pasted into the different regions within the image. Here one image is the source and the other is the copy-moved object. Fig. 1(b), defines the spliced attack, here an object within an image is spliced and copied into a different image. Fig. 1(c), shows an object removal attack, where an image is blended on top of some-other object.


(a)

(b)

(c)

Fig. 1. Different types of tampering attacks [7].

The majority of the cutting-edge image forgery detection approaches uses statistical properties through frequency domain feature. In [5], the artifact was introduced by applying a different level of JPEG compression for the identification of tampered images. In [6], additional noisy information was added to images that were compressed through JPEG to work on the presentation of resampling identification. Recently, the DL method has provided some very good results in computer vision applications, for example, object detection, hyperspectral crop classification, image registration, and segmentation, etc. Recently, DL models like auto-encoders [7] and Convolutional Neural Networks (CNN) [8], [9] have been employed for image tampering detection with good results. Existing tampering detection models are predominantly designed to detect only one kind of attack [10], [11]. In this way, one methodology probably won't excel in different sorts of tampering attacks. Additionally, it appears to be not realistic to expect this sort of attack well in advance.

Segmenting the tampering region is more challenging as compared to object segmentation because here only the region that is tampered only must be segmented. Recently, CNN has been emphasized with good effect for object segmentation strategies [12], [13]. In [12], a fully connected CNN has been used for studying object features and shape features through the extraction of features in a hierarchical manner. The CNN-based method provides good performance in the field of segmentation and object classification. In image tampering only the tampered region must be segmented and well-crafted tampered image differentiating between genuine and tampered is very difficult because they look very comparable. Although CNN produces spatial guides for various districts of sections, it can't sum up some different statistical noise made by various tampering methods. Consequently, the tampering region localization using a standard CNN-based design may not provide the ideal performance requirement of a realistic attack. In [13], studied different image forgery segmentation models were studied [14]. The study shows that the existing model performs badly in detecting copy-clone and object-removal. Using the resampling feature [4] the artifacts were created (i.e., resampling, compression) using tampered images can be learned [15]. The resampling attack generally occasionally allows correlation among pixels because of interpolation. The CNN-based [16], [17], image forgery identification model learns resampling features [18] very well using spatial maps produced through translation invariance of various regions of images [19], [20]. Thus, this research work aims to build an efficient resampling feature detection through CNN to detect hybrid attacks and achieve better-tampered region segmentation outcomes [21], [22]. The significance of the research work is as follows:

- This paper presented an improved CNN for tampering detection and segmentation in the image by adding to additional layer to retain the correlation between horizontal and vertical streams for exploiting good-quality resampling features.
- The TDS-ICNN model can work well considering different attacks such as scaling, compression, and rotation attacks.
- The TDS-ICNN is efficient in detecting multiple tampered regions within the same image.
- The TDS-ICNN can even detect image tampering attacks under noisy and small-smooth regions. An improved tampering area segmentation outcome using TDS-ICNN for tampering dataset with hybrid transformation attack is achieved. On the other side, the existing model works well i.e., good segmentation for some datasets, and for other datasets, very poor result is achieved.
- This shows the robustness of the TDS-ICNN model. An improved ROC performance is achieved using the TDS-ICNN model for carrying out classification tasks such as whether a given image is authenticated or tampered with considering diverse tampering datasets such as CoMoFoD, Coverage, and MICC.

The manuscript is arranged as follows: Section II discusses various current methodologies to detect tampering in multimedia content. Section III presents the material and method used for performing tampering detection methods. Section IV presents with working structure of the proposed tampering detection and segmentation model. Section V presents the experiment analysis of the proposed method with various other tampering detection methodologies. Section VI concludes the research significance with future research direction.

II. Related Work

The section studies various recent methodologies for detecting tampering in multimedia content. In [8] developed a robust image tampering detection method using CNN, where an image undergoes double compression tampering attacks; the model attains an accuracy of 92% using the CASIA v2 dataset. Similarly, [9] used ResNet50v2 for constructing batchwise CNN to detect image tampering. Experiment outcomes show 99.3 accuracy on the Casia v1 dataset and 81% accuracy on the CASIA v2 dataset. In [11] designed a tampering detection by training CNN with both unseen noise and predictable noise for online social network platforms. The model works well for social platforms; however, considering other domains the model fails to accurately detect tampering in images. In [18] designed pulse-CNN model to extract the contour features of potential tampering that had undergone complex tampering attacks like noise, scaling, and rotation attacks. The experiment outcome shows the model achieved a precision of 95.27% and 95.3% on the CASIA and CoMoFoD datasets, respectively.

In [23] introduced an end-to-end deep neural network namely BusterNet with two layers to capture the tamper feature followed by a fusion layer to merge the feature for segmentation of copy-move tamper region. Experiments are done on CASIA and CoMoFoD and segmentation output is given at pixel-level. Similarly, in [26] designed adaptive-attention residual refined network (AR-Net) to extract tampered object features, and feature maps correlation is done after which the fusion of features is performed using pyramid pooling. The experiment is done using CASIA II, Coverage, and CoMoFoD. In [27] developed a copy-move tampering detection mechanism using source-target region distinguishment network (STRDNet) by extending BusterNet. The model additionally introduces a filter at the pooling layer with a double self-correlation layer for establishing feature matching hierarchically. The experiment is done using CASIA, CoMoFoD, and Coverage datasets and the segmentation outcome is given at the pixel level. In [29] introduced an effective block-level feature optimization trained with deep CNN. The deep CNN uses a feature pyramid for robust detection accuracy against scaling attacks. The experiment is done using CASIA II with 57.48% and the CoMoFoD dataset with a precision of 50.11% and the boundary pixel direction aids in the detection of segmentation edges and can tolerate noise, compression, blurring, and color addition.

In [24] designed a key-point-based clustering method to detect tampering attacks under small-smooth regions. Experiments are done on MICC, GRIP, FAU, and Coverage with good true positive rates of 97.5%, 100.0%, 100%, and 80.22%, respectively. In [25] designed a new SIFT key points extraction through effective clustering for identifying tampered regions utilizing similarities. The clustering process to identify similarities is done considering color with different scales and smaller cluster size is considered to reduce computational overhead. In obtaining more quality outcomes pixel level similarity is done iteratively. The experiment is done using D0 datasets and pixel-level analysis segmentation accuracy is measured. In [30] combined both accelerated KAZE (A-KAZE) and speeded up robust features (SURF) for extraction of features by keeping the contrast level reasonably low. Then, to eliminate the mismatch density-based spatial clustering (DBSCAN) is used. Then, the affine matrix is applied to improve the tampering localization accuracy. The experiment is done with Ardizzone (D0) with 92.75% precision and the CoMoFoD dataset [31] with 95.23%. The overall survey shows key-point-based tampering detection is predominantly studied its performance using the MICC and DO dataset and the CNN-based model is predominantly studied using CASIA, Coverage, and CoMoFoD dataset.

The result attained using existing tampering detection methods have obtained satisfactory results; however, there is still wide scope to improve the results, as the existing model failed to provide good segmentation result under small-smooth robust tampering attacks which undergoes diverse post-processing attacks. Further, the model must be tested under different kinds of datasets; and most of the existing methods failed to provide pixel-level segmentation analysis. The current methods failed to extract feature correlation between horizontal and vertical layers; as a result, higher false positive is experienced with poor segmentation outcomes. In overcoming the research issues in the next section, the proposed methodology is presented.

III. Materials and Methods Used

A. Dataset Used

The dataset used in this work is listed below:

1) MICC: The dataset is composed of 600 images out of which, 160 images are forged, and the remaining 440 images are authenticated. The dataset is composed of different attacks like scaling and rotation made of plants, artifacts, animals, etc.

2) CoMoFoD: The dataset is made of a total of 260 images where several post—processing attacks are done. The resolution images are 512×512 for 200 images where different post-processing attacks have been done to obtain a total of 10,400 images. The other 60 images have a resolution of 3000×2000 and different post-processing attacks like compression, blurring, scaling, rotation, and noise addition have been created to obtain a total of 3120 images.

3) Coverage: The dataset is composed of different copy-clone attacks with a total of 100 images of tampered and as well as authenticated ones. The image size is 400×486 with complex attacks like rotation, scaling, and illumination attacks.

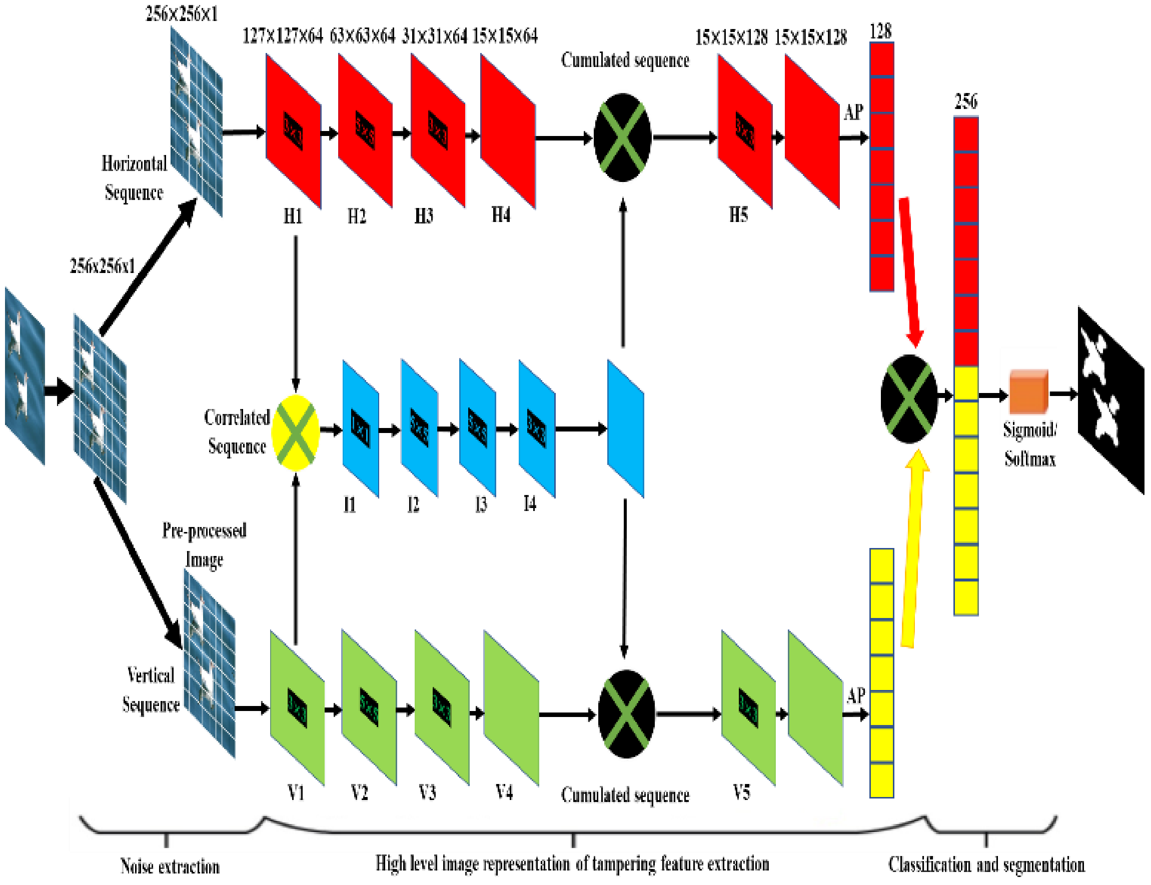


Fig. 2. Architecture of tampering detection and segmentation using improved convolution neural network.

B. Preprocessing

The work has used a total of three datasets; in this first, the image is resized to 512 × 512 into the non-overlapping region of 64 $(\text{i.e.}, 8 * 8)$ similar to the work presented in [17]; thus, induces certain artifacts. In [17] used space-filling curve for extracting correlation among both horizontal and vertical streams; the model achieved good, tampered region detection accuracies; however, with poor segmentation accuracies; especially under small-smooth hybrid attacks. In addressing the segmentation problem, this paper introduces an improved CNN model that with an additional layer to obtain a good correlated features-map for achieving improved tampered region detection and segmentation outcome.

IV. Proposed Methodology

The methodology to localize and segment tamper regions considering hybrid attacks using TDS-ICNN is presented in this section. The feature extraction process using TDS-ICNN. Lastly, extracted features are highly correlated and training is done to create a good descriptor in classifying whether a given image is authenticated or tampered with.

A. System Model and Architecture

The preprocessed image are passed into proposed improved CNN model for extraction of resampling features and identify the tampered region and segment it. In this work 50% of images from three different dataset is taken as input during training process of ICNN and tampering detection and segmentation model is constructed. The architecture of TDS-ICNN is given in Fig. 2. The working of tampering detection using ICNN architecture is given in Algorithm 1.

Algorithm 1: the ICNN-based tampering detection and segmentation.

Step 1. Start

Step 2. Load the images.

Step 3. Preprocess image into 512×512 into the non-overlapping region of 64 (i.e.,8×8)

Step 4. Pass the image into a three-layer ICNN.

Step 5. The first layer extracts the multi-dimensional RSF with the presence of noise. The RSF is captured by considering the difference between adjacent pixels across vertically and horizontally directions.

Step 6. The middle layer extracts the high-level feature across vertical and horizontal directions. The features that are correlated across both horizontal and vertical directions are aggregated.

Step 7. The, using last layer i.e., SoftMax and sigmoid function takes aggregated features as input for learning diverse features and optimizing binary tampering detection problems in multimedia forensics, respectively.

Step 8. Store the result and segmentation outcome.

Step 9. Stop

A detailed explanation of the different layers is given below.

B. Extraction of Noisy Features

In multimedia forensic extraction of resampling features is difficult as it is dependent on the information presented in a respective image. Nonetheless, some existing methodologies showed RSF extraction is not dependent on an image by extracting RSF through spatial domain using redundant feature properties. In this work, the noise is modeled by interpolating the current pixel with neighboring pixels and the difference in estimates is computed considering the image size of 256×256. In extracting the initial resampling feature with minimal training overhead two high-pass filters are used CNN kernel namely horizontal 3×1 and vertical 1×3 filters. Then the image is convoluted with padding and stride set to 1 using these filters, after that the difference (i.e., correlation) between neighboring pixels in vertical and as well horizontal direction are extracted to obtain a residual map of $256 \times 256 \times 1$.

C. High-level Feature Extraction

This layer takes input from the previous layer for extraction of high-level features. The standard tampered region detection and segmentation model extracts features and correlates through each direction individually; as a result, exhibits very poor performance. However, in this paper, the RSF features are extracted and weighted in both directions individually, where it is composed of five similar groups. The group encompasses 4-layer such as convolutional layer, batch normalize layer, activation layer, and pooling layer. The fifth group has correlated features collected from the middle layer of TDS-ICNN. Finally, the features from different layers are aggregated to obtain the final RSF feature to perform tapering detection classification.

The middle layer in TDS-ICNN fuses the correlated features from both directions. The middle layer is composed of 4 groups such as convolutional layer, batch normalize layer, activation layer, and pooling layer. The feature extracted from group 1 from horizontal and vertical streams is fused considering 1×1 convolutional kernel with stride set to 1. The remaining three groups are utilized for the extraction of high-level RSF illustrations of aggregated tampering information. Finally, by interpolating in both directions backward the feature map is established.

D. Classification

The ICNN introduces fully-connected CNN employing SoftMax/Sigmoid operation. The model takes input features from the middle layer and performs classification based on probability estimates that belong to the tampered or non-tampered group using the following equation.

$$P(z = 1|y) = \frac{1}{1+f^{-a}} \tag{1}$$

$$P(z = k|y) = \frac{f^{a_k}}{\sum_{l=0}^{L} f^{a_k}} \tag{2}$$

where Eq. (1) defines the sigmoid operation of a fully connected layer for performing classification of establishing whether an image is tampered with or not as output. The parameter $P(z = 1|y)$ defines the probability of whether y is classified into the respective group. Eq. (2) is used for detecting multiple tampered regions using SoftMax operation, where a_k is the fully connected layer output of the k^{th} neuron. The parameter $P(z = k|y)$ defines the probability of whether y belongs to $the\ k^{th}$ group.

E. Convolution Layer

The feature extraction done using the convolutional layer is as follows

$$G_k^{(o)} = \sum_{l=0}^{L} G_l^{(o-1)} * \alpha_{lk}^{(o)} + c_k^{(o)} \tag{3}$$

where $G_k^{(o)}$ defines the k^{th} feature-map established inside $the\ o^{th}$ layer, $G_l^{(o-1)}$ represents the j^{th} feature-map established inside $(the\ o-1)^{th}$ layer, $\alpha_{lk}^{(o)}$ defines $the\ l$ channel of k^{th} convolutional kernel inside $the\ o^{th}$ layer, and $c_k^{(o)}$ represents k^{th} bias parameter of o^{th} layers, and $*$ represent two-dimension convolution operation. The convolutional layer is set to 3 filters with sizes of $(1 * 1, 3 * 3,$ and $5 * 5)$ and a stride of 1.

F. Batch Normalization

The feature map extracted in the previous layer is normalized according to feature variance according to its distribution in the middle layer. The batch normalizer operates between activation and convolutional layers. The average between total information inside the batch is described as follows

$$\beta = \frac{1}{n} \sum_{j=0}^{n} y_j \tag{4}$$

where β defines the average, n defines the overall size of the feature used, and y_j represents the j^{th} information used. In a similar, manner the difference between the total features inside the batch is estimated as follows

$$\gamma^2 = \frac{1}{n} \sum_{j=0}^{n} \left(y_j - \beta\right)^2 \tag{5}$$

where γ^2 defines the difference. In this work, normalization is done on each feature to obtain new feature sets $\hat{y}_j$ with average initialized to 0 and difference initialized to 1 and is obtained using the following equation

$$\hat{y}_j = \frac{y_j - \beta}{\sqrt{\gamma^2 + \delta}} \tag{6}$$

where δ defines a trivial floating-point parameter higher than 0 that is used for avoiding dividing by zero error. The final batch-normalized feature is expressed as follows

$$z_j = \varphi \hat{y}_j + \omega \tag{7}$$

where, φ and ω are the CNN extracted features, and z_j defines batch normalization j^{th} output. In this work to obtain better features an activation function is used that is non-linear. The adoption of such a layer will not cause significant changes due to smaller fluctuations in prediction error.

G. Activation

In this work, the TDS is represented in the form of different spaces for achieving better-tampered region detection in multimedia forensics. The work uses TanH as an activation function instead of ReLu and Sigmoid because it works well for features with higher differences.

H. Pooling Layer

The element size is reduced by down-sampling the feature maps and establishing the hierarchical structure by observing continuous features' convolutional filter. The max pooling kernel size is set to 3×3 and stride of 2 and is applied to all pooling layers except the 5th layer of both streams for providing maximum parameter in each input feature-maps by capturing patterns on neighboring pixels. The average pooling is used in the last pooling layer of both streams for down-sampling the feature maps to 1 to minimize the model parameter of fully connected CNN. The adoption of such a mechanism significantly aided in achieving improved tampered region identification and segmentation using the proposed methodology.

V. Experimental Study

In this section experiment is done to validate the performance of TDS-ICNN over existing tampering detection methodologies like copy-move forgery detection using binary descriptor feature (CMFD-BDF) [22], BusterNet [23], fast and efficient CMFD (FE-CMFD) [24], AR-Net [26], and STRDNet [27].

A. Setup and Metrics

The TDS-ICNN model is modeled utilizing Python, C++, and Matlab libraries. The Intel I-7 processor with 16 GB RAM running with Windows 10 platform is used for conducting the experiments. Performance is evaluated using MICC-600, Coverage, and CoMoFoD dataset. The MICC-F600 dataset undergoes scaling and rotation post-processing tamper attacks. The CoMoFoD dataset undergoes compression, scaling, and rotation post-processing tamper attacks. The coverage dataset undergoes compression, scaling, and post-processing tamper attacks. The ROC metrics used are recall/ true positive rate (TPR), false positive rate (FPR), and F1-score for validating different tamper identification models.

$$False\ positive\ rate\ (FPR) = \frac{FP}{(FP+TN)} \qquad (8)$$

$$True\ positive\ rate\ (TPR) = \frac{TP}{(TP+FN)} \qquad (9)$$

$$F1 - Score = \frac{2TP}{((2TP+FN+FP))} \qquad (10)$$

$$Accuracy = \frac{((TN+TP))}{((TP+FP+TN+FN))} \qquad (11)$$

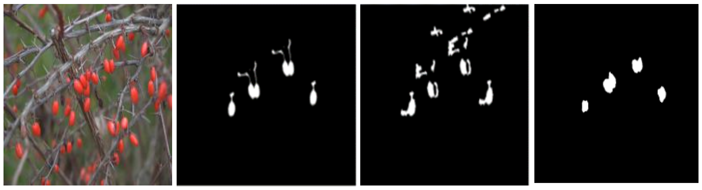

(a) original image
(b) Ground Truth
(c) Segmentation outcome utilizing existing method [28]
(d) Segmentation outcome utilizing TDS-ICNN

Fig. 3. Segmentation outcome of different methodologies.

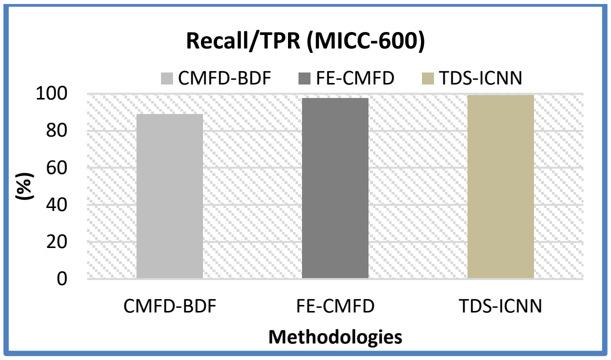


Fig. 4. Recall performance for MICC-600 dataset.

B. MICC Dataset

The experiment is conducted using the MICC-F600 dataset. The tampering segmentation result utilizing TDS-ICNN and other recent tamper identification models is graphically represented in Fig. 3. From Fig. 3 it is seen that TDS-ICNN provides improved tampering region segmentation outcomes when compared with existing models. Fig. 4 shows recall performance achieved utilizing TDS-ICNN and other existing tampering detection methodologies. Fig. 5 shows false positive rate performance achieved utilizing TDS-ICNN and other existing tampering detection methodologies. Fig. 6 shows the F1-score at image level performance achieved utilizing TDS-ICNN and other existing tampering detection methodologies. Fig. 7 shows that the F1-score at pixel-level performance was achieved utilizing TDS-ICNN and other existing tampering detection methodologies. The outcome obtained from Table I shows that TDS-ICNN improves detection accuracy and reduces false positives; thus, it can be adapted to provide a reliable tamper identification model.

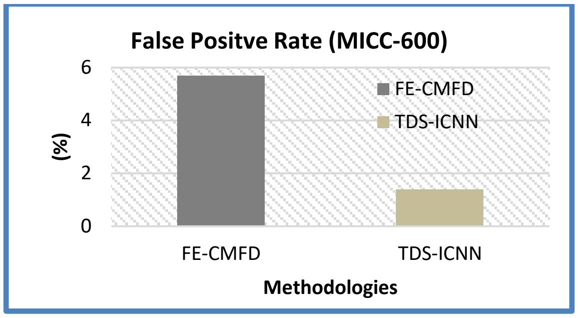


Fig. 5. False positive rate for MICC-600 dataset.

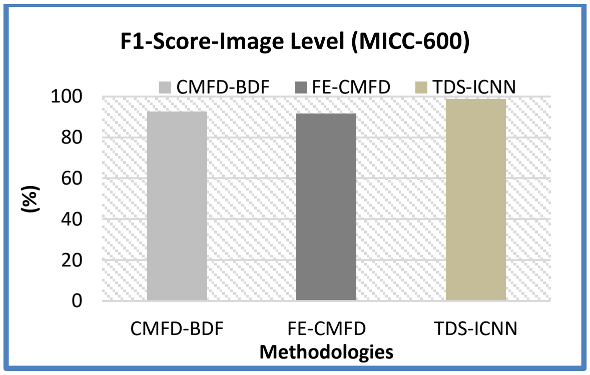


Fig. 6. F1-Score at image level performance for MICC-600 dataset.

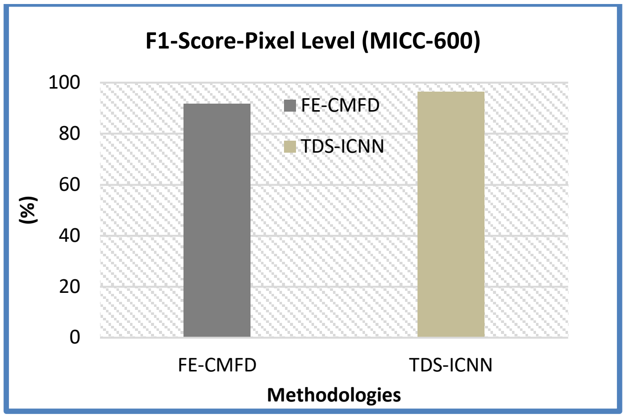
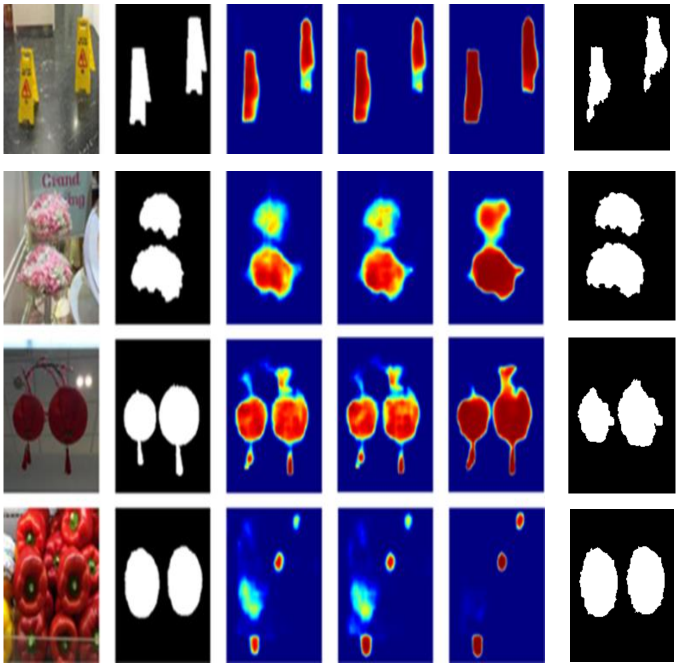


Fig. 7. F1-Score at pixel level performance for MICC-600 dataset.

TABLE I. COMPARATIVE STUDY FOR MICC DATASET

Methodology used	Performance metrics			
	TPR	***FPR***	***F1-Score image***	***F1-Score pixel***
CMFD-BDF [22]	89.14		92.6	
FE-CMFD [24]		5.68	91.5	91.8
TDS-ICNN [Proposed]	99.1	1.4	98.6	96.5

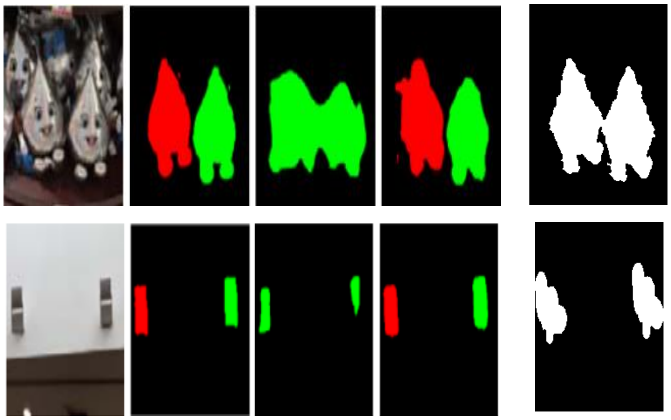


(a) Tampered image (b) Tamper region ground truth (c) Base (d) Base-Ada-Aten (e) Ar-Net [26] (f) TDS-ICNN

Fig. 8. Tampering region segmentation outcome using coverage dataset of proposed tampering and existing AR-Net tampering detection method.

C. Coverage Dataset

Here experiment is carried out using a coverage dataset. In the dataset is very difficult to classify which is authenticated and which is tampered one. The tampering segmentation results utilizing TDS-ICNN and other recent tamper identification models are graphically represented in Fig. 8 and Fig. 9. The result proves improved tamper region segmentation outcomes utilizing TDS-ICNN concerning recent tamper identification models. Fig. 10 shows the accuracy of performance achieved utilizing TDS-ICNN and other existing tampering detection methodologies. Fig. 11 shows the F1-score utilizing TDS-ICNN and other recent tamper identification methodologies. The outcome obtained from Table II shows that TDS-ICNN improves detection accuracy and reduce false positive and hence, it can be adopted to provide a reliable tamper identification model.



(a) Tampered image (b) Tampered region ground truth (c) BusterNet [23] (d) STRDNet [27] (d) TDS-ICNN

Fig. 9. Tampering region segmentation outcome using Coverage dataset of proposed tampering and existing STRDNet tampering detection method.

TABLE II. COMPARATIVE STUDY FOR COVERAGE DATASET

Methodology used	Performance metrics	
	Accuracy	***F1-Score***
Base [26]	0.8581	
Base-Ada-Atten [26]	0.8542	
AR-Net [26]	0.8488	
BusterNet [27]		0.464
STRDNet [27]		0.677
TDS-ICNN [Proposed]	0.8563	0.7456

D. CoMoFoD Dataset

The CoMoFoD dataset is utilized for studying the performance of TDS-ICNN with other recent tamper identification models. The dataset has diverse post-processing attacks being accrued out; thus, making it extremely challenging to detect tamper regions and localize them. The tampering segmentation result utilizing TDS-ICNN and other recent tamper identification models is graphically represented in Fig. 12. From Fig. 12 it can be stated that TDS-ICNN improves tamper region segmentation outcomes when compared with existing models. Fig. 13 shows recall performance achieved utilizing TDS-ICNN and other existing tampering detection methodologies. Fig. 14 shows the precision performance achieved utilizing TDS-ICNN and other existing tampering detection methodologies. Fig. 15 shows the F1-score result utilizing TDS-ICNN and other recent tamper identification methodologies. The outcome obtained in Table III shows that TDS-ICNN improves detection accuracy and reduces false positives; thus, can be adapted to provide a reliable tamper identification model.

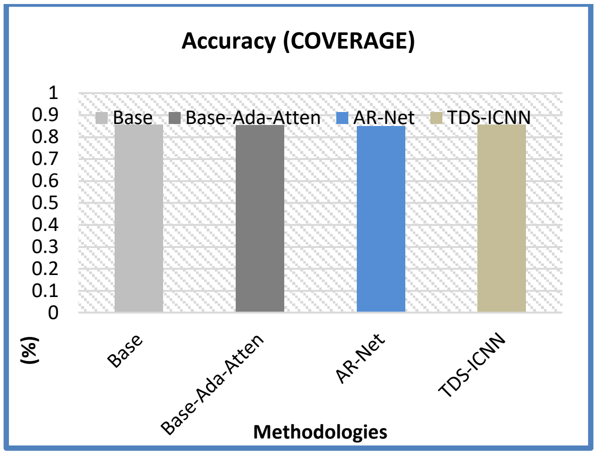


Fig. 10. Accuracy performance for coverage dataset.

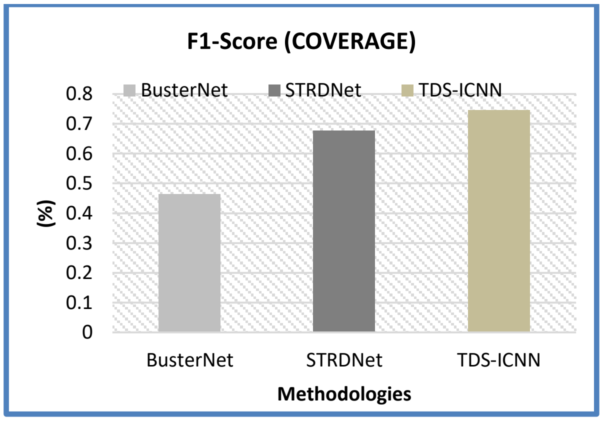
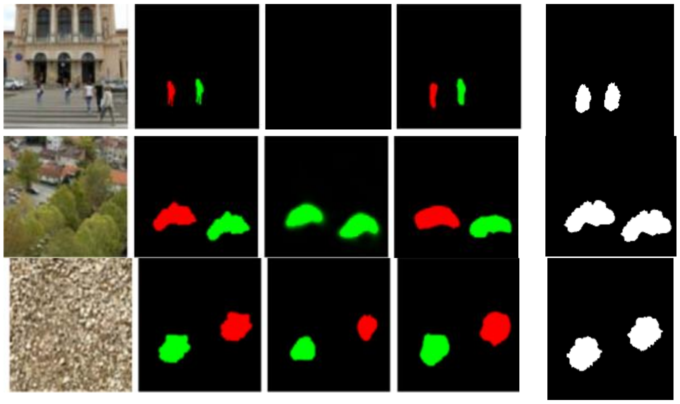


Fig. 11. F1-Score performance for coverage dataset.



(a) original image
(b) ground truth image
(c) BusterNet [23]
(d) STRDNet [27]
(d) TDS-ICNN

Fig. 12. Tampering region segmentation outcome using CoMoFoD dataset.

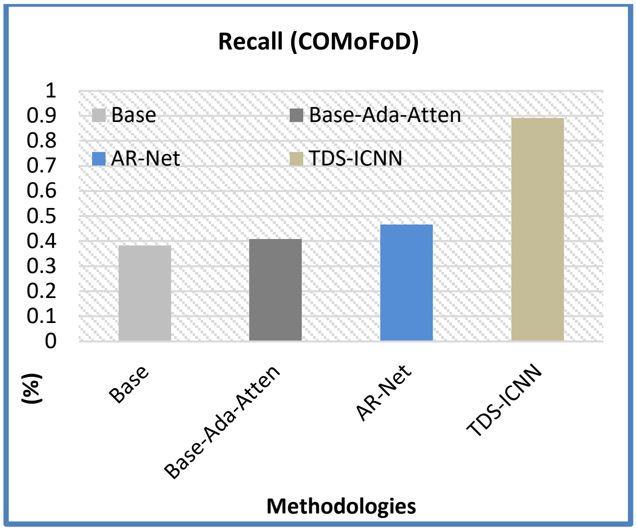


Fig. 13. Recall performance for the CoMoFoD dataset.

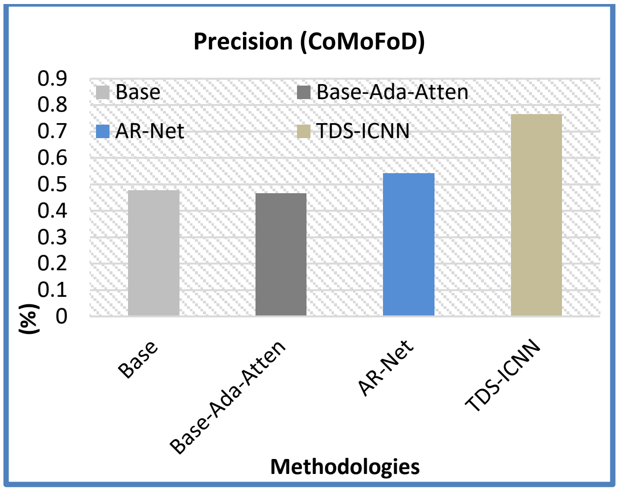


Fig. 14. Precision performance for CoMoFoD dataset.

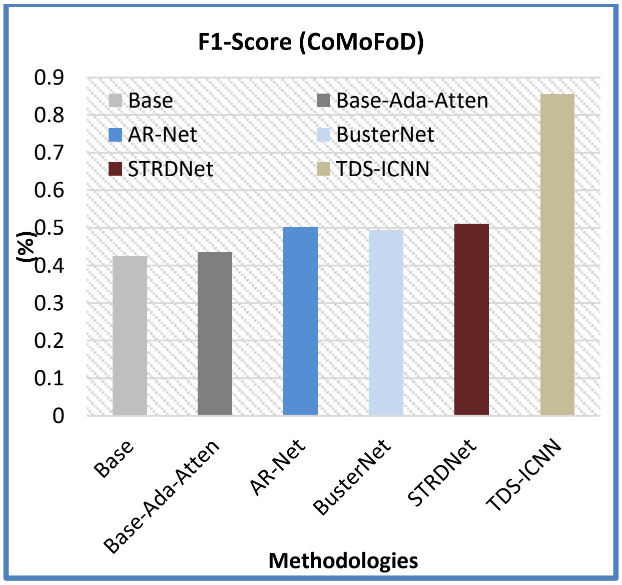


Fig. 15. F1-Score performance for CoMoFoD dataset.

TABLE III. COMPARATIVE STUDY FOR COMOFOD DATASET

Methodology used	**Performance metrics**		
	Recall	***Precision***	***F1-Score***
Base [26]	0.3811	0.4768	0.4236
Base-Ada-Atten [26]	0.4075	0.4661	0.4349
AR-Net [26]	0.4655	0.5421	0.5009
BusterNet [27]	×	×	0.493
STRDNet [27]	×	×	0.511
TDS-ICNN [Proposed]	0.89	0.7654	0.856

VI. CONCLUSION

The research work has presented a technique namely TDS-ICNN to identify whether an image is authenticated or tampered with. The preprocessing technique and feature extraction technique adopted in TDS-ICNN can retain spatial features concerning different patches. Alongside this, a good correlation exists among both horizontal and vertical curves through the introduction of a correlation layer. To eliminate spatial dependencies, the features extracted are aggregated and a descriptor is constructed to perform classification. The experiment is conducted using three datasets, such as MICC-600, Coverage, and CoMoFoD. For the MICC dataset the existing methods namely CMFD-BDF attains a TPR and F1-Score of 89.14% and 92.6%, respectively; however, the proposed TDS-ICNN attains a TPR and F1-score of 99.1% and 98.6%, respectively. For the Coverage dataset the existing methods namely AR-Net attain an accuracy of 84.88% and the proposed TDS-ICNN attains an accuracy of 85.63%, respectively. Similarly, the STRDNet attains an F1-score of 67.7%, and the proposed TDS-ICNN attains an F1-Score of 74.56%. For the CoMoFoD dataset the existing methods namely AR-Net attains a recall, precision, and F1-Score of 46.55%, 54.21%, and 50.09%, respectively; however, the proposed TDS-ICNN attains a recall, precision, and F1-Score of 89.0%, 76.54%, and 85.6%, respectively. The result attained shows that superior performance is achieved using TDS-ICNN in comparison with other standard tamper detection methods. A good ROC performance such as TPR, FPR, F1-Score, and accuracy in comparison with other existing tamper detection methodologies is achieved. The significant result provides a satisfactory benchmark for using it for real-time tampering image circulation in social media platforms and WhatsApp messenger; thereby can prevent misleading information circulation.

Future work would be focused on studying the model performance on other standard datasets like CASIA, and DO. The work would further investigate how the proposed model can be used to detect tampering in video. Further, would focus on developing an ensemble learning model to improve tampering detection accuracy with fewer false positives.

REFERENCES

[1] Dadkhah, S., Mazzola, G., Uliyan, D., Sadeghi, S., Jalab, H.A.: State of the art in passive digital image forgery detection: copy-move image forgery. Pattern Anal. Appl. 21, 291–306, 2017.

[2] H. Li, W. Luo, X. Qiu and J. Huang, "Image Forgery Localization via Integrating Tampering Possibility Maps," in IEEE Transactions on Information Forensics and Security, vol. 12, no. 5, pp. 1240-1252, 2017. DOI: 10.1109/TIFS.2015.2423261.

[3] J. H. Bappy, A. K. Roy-Chowdhury, J. Bunk, L. Nataraj, and B. Manjunath. Exploiting spatial structure for localizing manipulated image regions. In ICCV, 2017.

[4] J. Bunk, J. H. Bappy, T. M. Mohammed, L. Nataraj, A. Flenner, B. Manjunath, S. Chandrasekaran, A. K. Roy-Chowdhury, and L. Peterson. Detection and localization of image forgeries using resampling features and Deep Learning. In Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on, pages 1881–1889, 2017.

[5] W. Wang, J. Dong, and T. Tan. Exploring DCT coefficient quantization effects for local tampering detection. IEEE Transactions on Information Forensics and Security, 9(10):1653–1666, 2014. DOI: 10.1109/TIFS.2014.2345479.

[6] Yang, Hong-Ying & Qi, Shu-Ren & Niu, Ying & Niu, Pan-Pan & Wang, xiang yang. (2019). Copy-move forgery detection based on adaptive keypoints extraction and matching. Multimedia Tools and Applications. 78. 10.1007/s11042-019-08169-w.

[7] J. H. Bappy, C. Simons, L. Nataraj, B. S. Manjunath and A. K. Roy-Chowdhury, "Hybrid LSTM and Encoder–Decoder Architecture for Detection of Image Forgeries," in IEEE Transactions on Image Processing, vol. 28, no. 7, pp. 3286-3300, July 2019, doi: 10.1109/TIP.2019.2895466.

[8] Ali, S.S.; Ganapathi, I.I.; Vu, N.-S.; Ali, S.D.; Saxena, N.; Werghi, N. Image Forgery Detection Using Deep Learning by Recompressing Images. Electronics 2022, 11, 403. https://doi.org/10.3390/electronics11030403.

[9] Qazi, E.U.H.; Zia, T.; Almorjan, A. Deep Learning-Based Digital Image Forgery Detection System. Appl. Sci. 2022, 12, 2851. https://doi.org/10.3390/app12062851.

[10] Shivanandappa, Manjunath & Patil, Malini. Extraction of image resampling using correlation aware convolution neural networks for image tampering detection. International Journal of Electrical and Computer Engineering. 12. 3033. 2022, https://doi.org/10.11591/ijece.v12i3.pp3033-3043.

[11] H. Wu, J. Zhou, J. Tian, J. Liu and Y. Qiao, "Robust Image Forgery Detection Against Transmission Over Online Social Networks," in IEEE Transactions on Information Forensics and Security, vol. 17, pp. 443-456, 2022, doi: 10.1109/TIFS.2022.3144878.

[12] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In IEEE Conference on Computer Vision and Pattern Recognition, 2015.

[13] S. Manjunatha. and M. M. Patil, "Deep learning-based Technique for Image Tamper Detection," 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), 2021, pp. 1278-1285, doi: 10.1109/ICICV50876.2021.9388471.

[14] Wang, Chengyou & Zhang, Zhi & Zhou, Xiao. (2018). An image copy-move forgery detection scheme based on A-KAZE and SURF features. Symmetry. 10. 706. 10.3390/sym10120706.

[15] Liang Y, Fang Y, Luo S and Chen B. Image Resampling Detection Based on Convolutional Neural Network. 2019 15th International Conference on Computational Intelligence and Security (CIS), Macao, China, 2019. pp. 257-261. DOI: 10.1109/CIS.2019.00061

[16] Shivanandappa, Manjunath & Patil, Malini. Efficient resampling features and convolution neural network model for image forgery detection. Indonesian Journal of Electrical Engineering and Computer Science. 25. 183, 2022. https://doi.org/10.11591/ijeecs.v25.i1.pp183-190.

[17] Shivanandappa, Manjunath & Patil, Malini. Tampering Detection using Resampling Features and Convolution Neural Networks. Turkish Journal of Computer and Mathematics Education; Trabzon Vol. 12, Iss. 11, pp. 2791-2800, 2021.

[18] Zhou, G., Tian, X. & Zhou, A. Image copy-move forgery passive detection based on improved PCNN and self-selected sub-images. Front.

Comput. Sci. 16, 164705 (2022). https://doi.org/10.1007/s11704-021-0450-5.

[19] Flenner, Arjuna & Peterson, Lawrence & Bunk, Jason & Mohammed, Tajuddin Manhar & Nataraj, Lakshmanan & Manjunath, B. Resampling Forgery Detection Using Deep Learning and A-Contrario Analysis. Electronic Imaging. 2018. 10.2352/ISSN.2470-1173.2018.07.MWSF-212, 2018.

[20] Qazi, Tanzeela & Ali, Mushtaq & Hayat, Khizar & Baptiste, Magnier. (2022). Seamless Copy–Move Replication in Digital Images. Journal of Imaging. 8. 69. 10.3390/jimaging8030069.

[21] Huang, H., Ciou, A. Copy-move forgery detection for image forensics using the superpixel segmentation and the Helmert transformation. J Image Video Proc. 2019, 68 (2019). https://doi.org/10.1186/s13640-019-0469-9, 2019.

[22] Raju, P.M., Nair, M.S.: Copy-move forgery detection using binary discriminant features. J. King Saud Univ. - Comput. Inf. Sci. 2018.

[23] Yue Wu, Wael Abd-Almageed, Prem Natarajan. BusterNet: Detecting Copy-Move Image Forgery with Source/Target Localization. Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 168-184.

[24] Y. Li and J. Zhou, "Fast and Effective Image Copy-Move Forgery Detection via Hierarchical Feature Point Matching," in IEEE Transactions on Information Forensics and Security, vol. 14, no. 5, pp. 1307-1322, May 2019, doi: 10.1109/TIFS.2018.2876837.

[25] H. Chen, X. Yang and Y. Lyu, "Copy-Move Forgery Detection Based on Keypoint Clustering and Similar Neighborhood Search Algorithm," in IEEE Access, vol. 8, pp. 36863-36875, 2020, doi: 10.1109/ACCESS.2020.2974804.

[26] Y. Zhu, C. Chen, G. Yan, Y. Guo and Y. Dong, "AR-Net: Adaptive Attention and Residual Refinement Network for Copy-Move Forgery Detection," in IEEE Transactions on Industrial Informatics, vol. 16, no. 10, pp. 6714-6723, Oct. 2020, doi: 10.1109/TII.2020.2982705.

[27] B. Chen, W. Tan, G. Coatrieux, Y. Zheng and Y. Q. Shi, "A serial image copy-move forgery localization scheme with source/target distinguishment," in IEEE Transactions on Multimedia, 2020. doi: 10.1109/TMM.2020.3026868.

[28] J. Li, X. Li, B. Yang, and X. Sun. Segmentation-based image copy-move forgery detection scheme. IEEE Transactions on Information Forensics and Security, 10(3):507–518, 2015.

[29] Li Q, Wang C, Zhou X, Qin Z. Image copy-move forgery detection and localization based on super-BPD segmentation and DCNN. Sci Rep. 2022 Sep 2;12(1):14987. doi: 10.1038/s41598-022-19325-y.

[30] Fu, G.; Zhang, Y.; Wang, Y. Image Copy-Move Forgery Detection Based on Fused Features and Density Clustering. Appl. Sci. 2023, 13, 7528. https://doi.org/10.3390/app13137528.

[31] Manaf Mohammed Ali Alhaidery, Amir Hossein Taherinia, Haider Ismael Shahadi, A robust detection and localization technique for copy-move forgery in digital images, Journal of King Saud University - Computer and Information Sciences, Volume 35, Issue 1, 2023, Pages 449-461, https://doi.org/10.1016/j.jksuci.2022.12.014.