Hindawi
BioMed Research International
Volume 2021, Article ID 5546572, 12 pages
https://doi.org/10.1155/2021/5546572

Research Article

Cloud-Based Fusion of Residual Exploitation-Based Convolutional Neural Network Models for Image Tampering Detection in Bioinformatics

Amit Doegar,[1] **Srinidhi Hiriyannaiah**,[2] **G. M. Siddesh**,[3] **K. G. Srinivasa**,[1] **and Maitreyee Dutta**[1]

[1]*Department of Computer Science and Engineering, NITTTR, Chandigarh, India*
[2]*Department of Computer Science and Engineering, Ramaiah Institute of Technology, Bangalore, India*
[3]*Department of Information Science and Engineering, Ramaiah Institute of Technology, Bangalore, India*

Correspondence should be addressed to Amit Doegar; amit@nitttrchd.ac.in

Received 3 February 2021; Revised 9 March 2021; Accepted 31 March 2021; Published 12 April 2021

Academic Editor: B. D. Parameshachari

Cloud computing has evolved in various application areas such as medical imaging and bioinformatics. It raises the issues of privacy and tampering in the images especially related to the medical field and bioinformatics for various reasons. The digital images are quite vulnerable to be tampered by the interceptors. The credibility of individuals can transform through falsified information in the images. Image tampering detection is an approach to identifying and finding the tampered components in the image. For the efficient detection of image tampering, the sufficient number of features are required which can be achieved by a deep learning architecture-based models without manual feature extraction of functions. In this research work, we have presented and implemented a cloud-based residual exploitation-based deep learning architectures to detect whether or not an image is being tampered. The proposed approach is implemented on the publicly available benchmark MICC-F220 dataset with the k -fold cross-validation approach to avoid the overfitting problem and to evaluate the performance metrics.

1. Introduction

Computer vision and applications are the upcoming research areas in the field of computing and bioinformatics. The applications include object detection, video analytics, image segmentation, and image pixel analysis. Image forensics and forgery detection are the important applications that can be considered for computer vision applications. It has become easy for the online players to digitally tamper and alter the different dimensions of the images. The detection of forgery in the images is one of the key challenges in computer vision applications such as video surveillance. The different types of tampering include copy-move forgery and splicing. The recent research works on image tampering detection focus on the splicing [1] and copy-move methods [2–4]. The image tampering can be done in many ways and not restricted only copy-move and splicing methods. In some recent works, the residual methods are used for image tampering detection [5–7]. In this paper, the problem of image tampering detection with residual exploitation is dealt with the Convolutional Neural Network (CNN) models and fusion-based approach.

Recent advances in the machine learning (ML) and deep learning (DL) technologies are used for the tampering detection of the images. The underlying relationships are identified by the ML algorithms in analysing the data and making decisions. Speech and vision problems use the ML algorithms to understand and evaluate the different parameters for the speech and vision data [8]. However, these algorithms were limited to the capability of prediction of smaller amounts of data. DL techniques with CNN became more popular for solving the different challenging problems in speech and vision [9]. These techniques are used for image segmentation,

classification, detection, image context, and retrieval-related tasks [10, 11]. The optimal solutions to the computer vision problems using DL have paved the way for solving different types of problems in image tampering detection as well. In this paper, the residual nature of the CNN models is exploited for the image tampering detection.

CNN has given researchers to provide an insight into the image tampering detection using the feature maps provided at each layer. It was used to detect whether the image is tampered or not initially, but not to locate the tampered regions. However, there are some attempts to locate the tampered regions using the CNN but are not accurate [12]. A nonoverlapping image patch method was used for the image segmentation in [13]. However, when the image size is small, it fails to identify the tampering. The contextual information of the image is lost and leads to incorrect prediction because they use the image patch as a part of the input to the network. Once the CNN goes deeper, there will also be gradient degradation problem and weak discrimination of features as well. The weak discrimination of the features leads to the incorrect prediction. In this paper, the traditional extraction method of the image patch for image tampering detection is replaced with a fusion-based model based on the residual nets for optimal image tampering detection.

In this paper, image forgery detection is carried out through a novel decision method by residual exploitation-based deep learning models. The proposed approach consists of three phases on the pretrained and fine-tuned spatial exploitation-based CNN models, namely, ResNet-18, ResNet-50, and ResNet-101 [5]. In the first phase, a system to extract features using residual exploitation-based CNN models in the second-phase machine learning-based classifier is deployed on the extracted features, and in the final phase, the fusion of decision outcomes based on these residual exploitation-based CNN models is done to evaluate the accuracy of the model.

The main contributions of this paper are as follows:

(i) A decision fusion-based system is proposed using the CNN-based approach for image tampering detection. The residual exploitation-based CNN models used for the fusion decision are ResNet-18, ResNet-50, and ResNet-101

(ii) The fusion decision system is implemented in two phases. First, the pretrained weights for the residual exploitation-based CNN models are used to evaluate the tampering of the images. Second, the fine-tuned weights are used to compare the results of the tampering of the images with the pretrained model

(iii) The utilization of the residual exploitation-based CNN models leads to the reduction of the number of false matches, thereby reducing the false-positive rate and ultimately increasing the accuracy of the approach

The paper is further organized as follows. In Section 2, the related work is discussed on the image tampering detection methods and the CNN methods with spatial exploitation that are used for image tampering detection. In Section 3, the fusion model using the residual exploitation-based CNN models is proposed, and it follows the regularization applied on the fusion model in Section 4. The experiments and results are discussed in Section 5 followed by the conclusion.

2. Related Work

The different areas of research identified in the image tampering detection domain are resampling detection, JPEG artefacts, detection of copy-move operations, splicing, and object removal [14]. Digital content has evolved over a period of time with the advances in the computer graphics, internet, and digital contents. The advances are utilized for many applications in computer vision and image recognition applications [15]. However, the downside of the exploitation of these applications lies in the fact of analysis of creating fake images and videos. The current research focuses on identifying the forgeries in the images using different DL models and techniques. In this section, we discuss some of the related work based on some of image tampering detection methods and the residual exploitation of CNN models used for the image tampering detection methods.

In the method of copy-move forgeries, the image is divided into overlapping blocks, and correlation is determined for the cloned blocks. A patch detection-based algorithm [16] was used to approximate the neighbours for the forgery detection. Geometrical-based transformations with invariant features of the image were used for the copy-move forgery detection. Local binary pattern (LBP) and steerable pyramid transform (SPT) were used for image forgery detection [17, 18]. These methods are used for the traditional extraction of the tampered regions for the forgery detection. However, these methods fail for the images that are small in size and provide inaccurate tampered regions of the image.

DL methods are widely used for image forgery detection in recent works [19, 20]. The image manipulation tasks that are generally used are generic manipulations, resampling [21], and splicing [20]. One of the works in [22] used Gaussian-based CNN for Steganalysis. In [23], a stacked autoencoder was used for image tampering detection. Further, CNN combined with LSTM was used for image tampering detection using the various layers of CNN. Residual-based networks such as ResNet 50 were used in [15] for image tampering detection using the input of computer-generated images.

In 2015, a variant of CNN called U-Net was proposed in [24]. U-Net gained a huge success in neuronal structure segmentation, and because of its features which are propagated among layers, its framework is path breaking in the above field. The context information in U-Net is captured by successive layers; later, the output feature is up sampled and finally combined with the high-resolution features propagated by a symmetric expanding path. This enables precise location and also reduces the loss of detail information. This helped to propose some image segmentation methods [25, 26] based on U-Net. Most of the time in image splicing forgery detection, we need to segment out the tampered region in an image which is impossible to do with human eyes. Hence, image splicing forgery detection can be understood as a complicated image segmentation task which is

independent of the human visual system. Extraction of discriminative features plays a vital role in locating tampered regions of an image by providing the differences of image attributes. Even though U-Net can extract relatively shallow discriminative features, only two sides of the U-Net structure are interactive; this is not enough to locate the tampered regions. Besides, the gradient degradation problem [26] is observed when the network architecture becomes much more deeper.

VLAD [27] is a representation used in image recognition which is encoded by the residual vectors with respect to a dictionary. The formulated probabilistic version of VLAD [27] is used to form Fischer vector [28]. Both the representations are powerful for image retrieval and classification. Encoding residual vectors [28] is preferred over encoding original vectors for vector quantization. Multigrid method [29] is widely used in computer graphics and low-level vision to solve Partial Differential Equations (PDEs). This method develops subproblems at multiple scales, where each subproblem gives the residual solution between the finer and the coaster scale. Hierarchical basis preconditioning [30] is an alternative to Multigrid which relies on the variables that represent the residual vectors between the coarser and finer scales. As the standard solvers are unaware of the residual nature of the solutions, they converge slower compared to the Multigrid or Hierarchical basis preconditioning solvers [30]. These methods suggest that preconditioning or good reformulation can make the optimization easy.

Image classification using deep learning techniques involves the same three steps that are followed in machine learning algorithms for image classification. Those three steps are preprocessing, feature extraction, and classification. First, the input dataset is divided into two sets for training and testing. Both training and testing images are then preprocessed to resize the images according to the pretrained network size [31]. Further, these preprocessed images are sent through various layers of the network until the fully connected layer (FC-1000) extract features from the images sent. The classifier model is trained by passing the features extracted by FC-1000. The prediction of test images is done using the trained classifier model. Naïve Bayes, K-nearest neighbour, and multiclass model using SVM learner are the three classifiers used in this model.

In 2015, the ResNet architecture was proposed [32] which won the championship in the classification task of ImageNet match. For a few stacked layers in ResNet, residual mapping is defined as $y = F(x) + x(1)$ Equation (1), where x represents the input, the operation $F(x) + x$ is performed by a shortcut connection and element-wise addition, and y represents the output. The gradient degradation problem is a serious problem in image splicing forgery detection. This is generally seen in deeper networks; hence, the residual mapping technique is proposed to overcome this problem. The differences of image essence attributes are hard to discover through the multilayer structure as the discrimination of image essence attribute features will be weaker. To solve the above issue and to simultaneously strengthen the learning way of CNN, the residual mapping should be utilized more efficiently.

To make full use of features to detect tampering and to fuse features, adaptive attention mechanism and residual refinement network [33] are used which are robust to various postprocessing, such as blur, noise, and JPEG recompression. Residual-based [34] descriptors have proven extremely effective for a number of image forensic applications. Experimental results based on residual-based fully convolutional network [35] for image tampering detection for various datasets performed better than some existing methods in generalization ability, localization ability, and robustness against additional operations.

As CNN contains numerous parameters, weights, layers (spatial filters), biases, and so on, nowadays, they are widely used for detecting image forgery. The convolution operation in CNN considers the neighbourhood of the pixels in an image which results in different sizes of layers (spatial features). Various sizes of filters encapsulate the images with unique levels of granularity. The coarse-grained features of the image are extracted using the large-sized filters, while the fine-grained portions of the images are extracted using the small filters. Various researches on adjusting the size of the filters are conducted to optimize the performance of CNN to extract both coarse-grained and fine-grained features of an image.

The evaluation metrics play a vital role in estimating the tampering in images. There are two types of metrics used for the evaluation, pixel-based and image-based [36]. In the pixel-based method, the classification of the pixels is done as copy-move and authentic, whereas in the image-based method, the classification of image is done as either tampered or authentic. The measures used at image level are TP (true positive): tampered images are detected as tampered images, TN (true negative): nontampered images are detected as nontampered images, FP (false positive): nontampered images are detected as tampered images, and FN (false negatives): tampered images are detected as nontampered images. In this paper, the proposed method uses image-based methods to evaluate the accuracy. Among the existing methods discussed in this section, the CNN model is used to extract the spatial features of the image which includes the geometry, texture, wavelet, and transformations. The weights of the majority of the above-discussed models need to be altered each time for a new dataset of images as they use pretrained weights. In the proposed system, a fusion of decision-making is involved for image tampering detection based on the CNN models. The proposed fusion model is discussed in further sections.

3. Proposed System

The architecture of the proposed fusion system of residual exploitation-based CNN models is as shown in Figure 1. The residual exploitation-based CNN models chosen are ResNet-18, Resnet-50, and ResNet-101. It consists of three stages, namely, data preprocessing, fusion model, and the classification. In the data preprocessing stage, the input image is preprocessed based on the dimensions required by the fusion models. A support vector machine (SVM) is used for the classification of the image as tampered/forged or not.

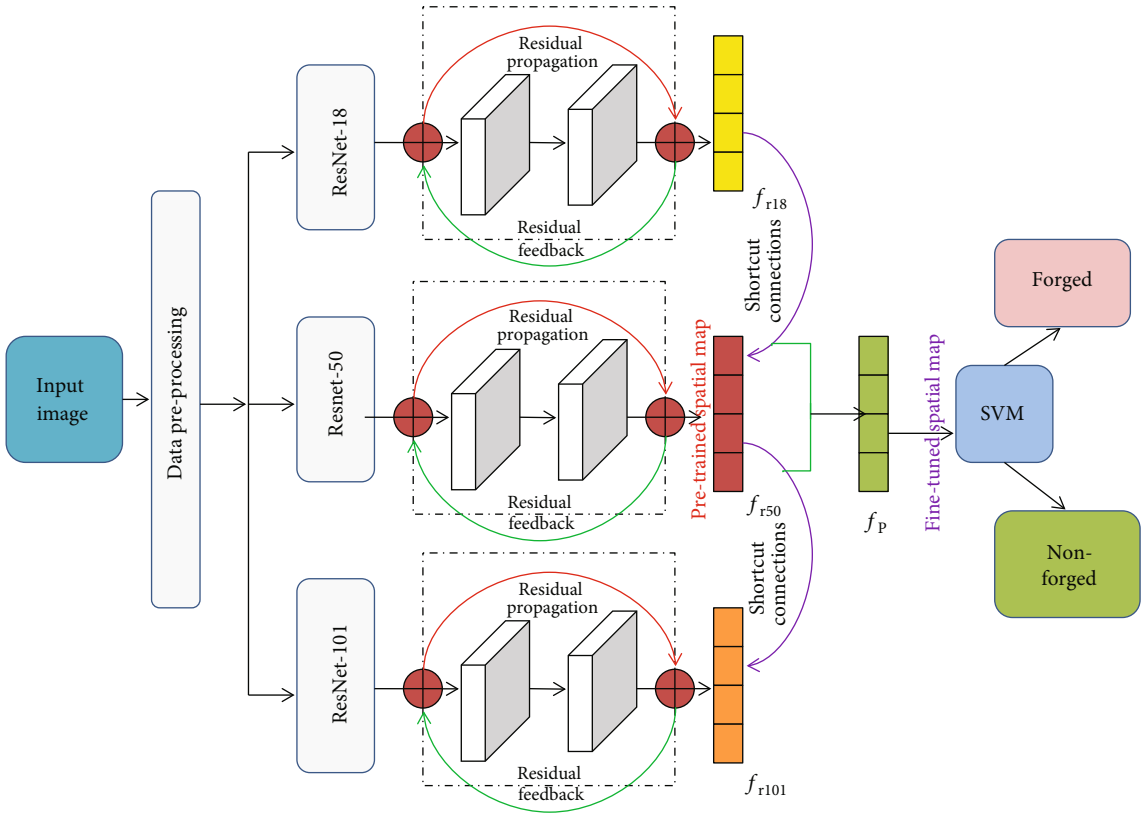


Figure 1: Architecture of fusion of residual exploitation-based CNN models.

The proposed system is implemented in two parts, i.e., pretrained and fine-tuned. In the pretrained implementation, regularization is not applied, and the pretrained weights are used for the classification. Regularization is a set of techniques that can prevent overfitting in neural networks and thus improve the accuracy of convolutional neural network-based models. Thus, to minimize the effect of overfitting regularization is applied in fine-tuned model implementation. In the fine-tuned implementation, regularization is applied for the classification. Initially, we discuss the residual exploitation-based CNN models, and then, the strategy used for the regularization is discussed in the further sections.

3.1. Data Preprocessing. In this stage, the input image that needs to be identified whether it is tampered or not is subjected to preprocessing. The dimensions of the input image required for ResNet-18 is 224×224. The dimensions of the input image required for ResNet-50 is 224×224. The dimensions of the input image required for ResNet-101 is 224×224 . The input image is preprocessed first based on the dimensions required for each of the residual exploitation-based CNN models. Each CNN model then takes the input image to produce the feature vector in the further stages.

3.2. Residual Exploitation-Based CNN Models for Image Classification. The different residual exploitation-based CNN models that are considered for fusion are ResNet-18, ResNet-50, and ResNet-101. These models are used for the image classification problems numerously. In this section, these models are discussed briefly. The residual deep learning models considered are summarized as shown in Table 1.

3.2.1. ResNet-18. It is a CNN trained on the ImageNet dataset with 18 layers deep and can classify the images upto 1000 categories. The network has learnt rich representations of the images with 11.7 million parameters. The network has an image input size of 224-by-224.

3.2.2. ResNet-50. It is a CNN trained on the ImageNet dataset with 50 layers deep and can classify the images upto 1000 categories. The network has learnt rich representations of the images with 25.6 million parameters. The network has an image input size of 224-by-224.

3.2.3. ResNet-101. It is a CNN trained on the ImageNet dataset with 101 layers deep and can classify the images upto 1000 categories. The network has learnt rich representations of the images with 44.6 million parameters. The network has an image input size of 224-by-224.

3.2.4. SVM. SVM is used as a classifier, as it is more suitable, popular, efficient, and widely used for binary classification problems as compared to other classifiers. Performance of the proposed approach is evaluated at image level by calculating the performance metrics as precision, false-positive rate (FPR), and recall, also known as true positive rate (TPR), F -score, and accuracy.

Table 1: Residual exploitation-based CNN models for image classification [36].

Residual exploitation-based CNN models	Depth	Number of parameters (million)	Input size
ResNet-18	18	11.7	224-by-224
ResNet-50	50	25.6	224-by-224
ResNet-101	101	44.6	224-by-224

3.3. Fusion Model and Regularization. The proposed system is first implemented as a CNN with pretrained weights for the image classification. Afterwards, the proposed system is implemented as a fusion of the residual exploitation-based CNN models as discussed in the previous section. Initially, the input image is passed to the residual exploitation-based CNN models to obtain their feature maps, respectively. The feature map from the ResNet-18 is denoted as f_{r18}, the feature map from the ResNet-50 is denoted as f_{r50}, and the feature map from the ResNet-101 is denoted as f_{r101}. For the fusion model, the pretrained CNN output feature mapping fp is used. This feature map f_p is a combination of the feature maps obtained from the residual exploitation-based CNN models as shown in Equation (1).

$$f_p = f_{r18} + f_{r50} + f_{r101}. \quad (1)$$

The fusion model uses feature map f_p as a local descriptor for input patch to extract the features of the image. The image for the fusion model is represented as a function $Y_{fusion} = f(x)$ where x is the patch in the input image. For a test image size $m \times n$, a sliding window of size $p \times p$ is used to compute the local descriptor Y_i is computed as shown in Equation (2). It is obtained as a concatenation of all the input patches X_i, and the new image representation is given by Equation (3) where s is the size of the stride used for transforming the input patch; this new image representation fusion is used as the feature map for the classification by the SVM as tampered or nontampered. In Equation (2), W_s represents the weights of the shortcut connections from the residual features of ResNet-18, ResNet-50, and ResNet-101 models.

$$Y_{fusion} = [Y_1 + Y_2 + \cdots + Y_T] + W_s, \quad (2)$$

$$f_{fusion} = \frac{m-w}{s} + 1 * \frac{n-w}{s} + 1. \quad (3)$$

For fine tuning of the parameters of the fusion model, the initialization of the weight kernels is used as shown in Equation (4). In this equation, W_f represents the weights of the fusion model, W_{r18} represents the weights of the ResNet-18 model, W_{r50} represents the weights of the ResNet-50 model, and W_{r101} represents the weights of the ResNet-101 model. The weight of the fusion model W_f is initialized as shown in Equation (5). The initialization of the weights acts as a regularization term and facilitates the fusion model to learn robust features of detecting the forgery rather than the complex image representations.

$$W_f = \left[W_{r18j} + W_{r50j} + W_{r101j}\right], \quad \text{where } j = 1, 2, 3, \quad (4)$$

$$W_f = \left[W_{r18}^{4k-2} + W_{r50}^{4k-2} + W_{r101}^{4k}\right] + W_s, \quad \text{where } k = ((j+1) \bmod 11) + 1. \quad (5)$$

4. Experiments and Results

In this section, the experiments and results of the proposed fusion model are discussed. The experiment is carried out in two stages. In the first stage, the residual exploitation-based CNN models are used with the pretrained weights, in the second stage, the fusion model with the strategy of weight initialization as discussed in the previous section. The configuration of the system used for the experiments is shown in Table 2.

4.1. Dataset. The dataset used for the experiment is benchmark publicly available MICC-F220 [37] of 110 nonforged images and 110 forged images with 3 channels, i.e., color images of size 722×480 to 800×600 pixels with 10 different combinations of geometrical and transformations attacks as shown in Figures 2 and 3. To avoid the problem of overfitting and to generalize the approach k-fold cross-validation with the value of k as 5 is used for training and testing sampling of images.

4.2. Baseline Models and Metrics. The baseline models that are used for the comparison of the fusion model are summarized as follows.

(i) SIFT: It uses the forensic method of the image tampering detection using a scale invariant features transform (SIFT) approach [37].

(ii) SURF: It uses a speeded up robust features (SURF) and hierarchical agglomerative clustering (HAC) for the image tampering detection [38].

(iii) DCT: It uses discrete cosine transform (DCT) features for each block and through lexicographical sorting of block-wise DCT coefficients for the image tampering detection [39].

(iv) PCA: It uses PCA on the image blocks to reduce the dimension space and perform lexicographical sorting for the image tampering detection [40].

(v) CSLBP: It uses center-symmetric local binary pattern (CSLBP) based on the combined features of Hessian points for the image tampering detection [41].

(vi) SYMMETRY: It uses the local symmetry value of an image to compute the key points for image tampering detection [42].

Table 2: Configuration of system.

Hardware	Intel(R) Xeon(R) Silver 4110 CPU with 2.10 GHZ, 128 GB RAM
GPU	Tesla P4
Software	Ubuntu 18.04 with Matlab release R2020a

Figure 2: Original image.

(vii) CLUSTERING strategy: It uses SIFT features with a clustering strategy to detect image tampering [43].

The basic metrics that are used for the evaluation of the fusion model are false-positive rate (FPR), recall (R), precision (P), f-score, and accuracy as shown in the equations (Equations (7) to (10)). The confusion matrix is used as the basis for the evaluation of the tampered and nontampered images as shown in Table 3, and the notations used are as follows:

(i) TP: Tampered image detected as tampered.

(ii) FN: Tampered image detected as nontampered.

(iii) FP: Nontampered image detected as tampered.

(iv) TN: Nontampered image detected as nontampered.

$$\text{FPR} = \frac{\text{FP}}{(\text{FP} + \text{TN})}, \tag{6}$$

$$\text{Recall} = \frac{\text{TP}}{(\text{TP} + \text{FN})}, \tag{7}$$

$$\text{Precision} = \frac{\text{TP}}{(\text{TP} + \text{FP})}, \tag{8}$$

$$F1\ \text{Score} = \frac{2 * (\text{Precision} * \text{Recall})}{(\text{Precision} + \text{Recall})}, \tag{9}$$

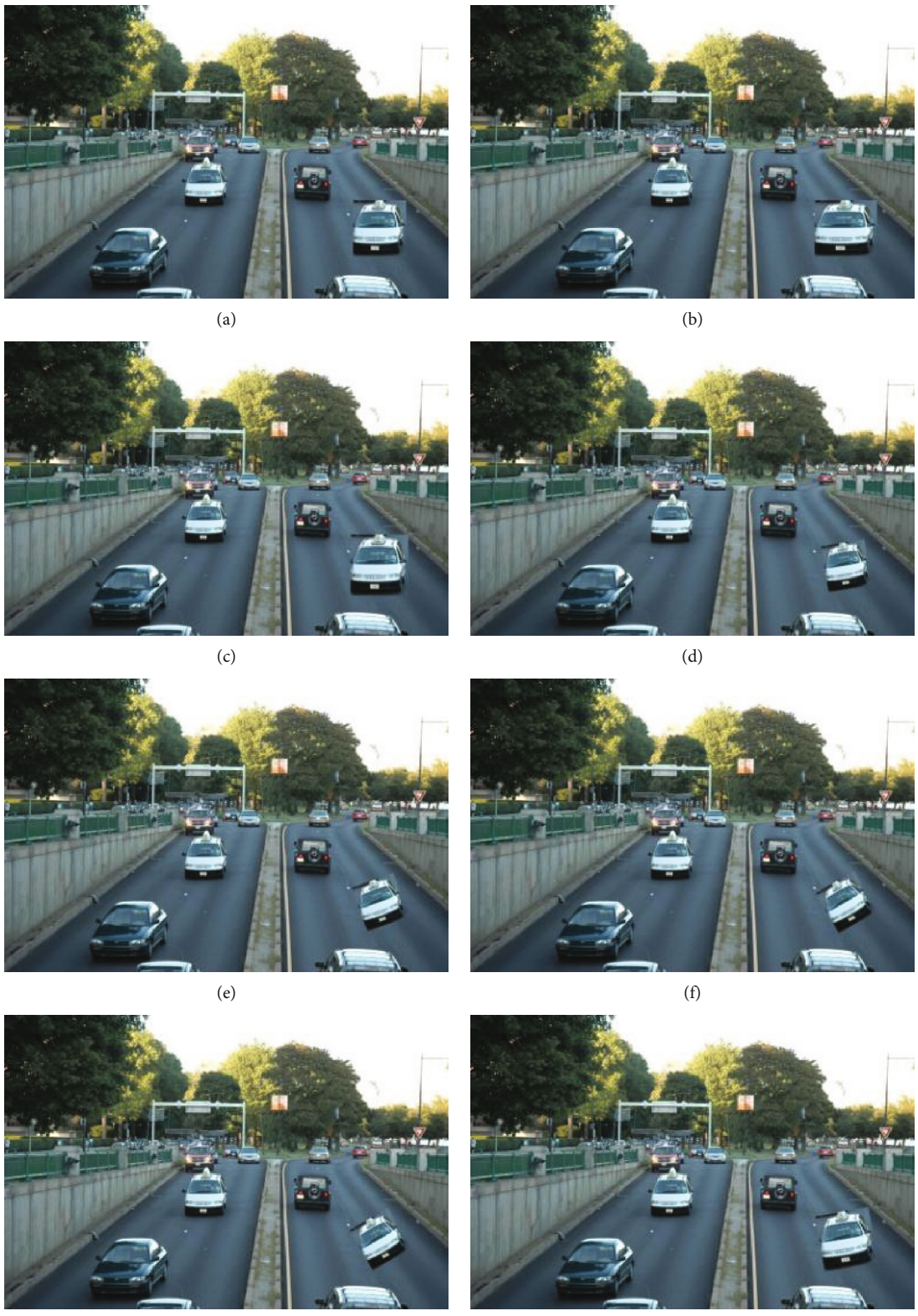
$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{TN} + \text{FP} + \text{FN}))}. \tag{10}$$

4.3. Pretrained Residual Exploitation-Based CNN Models. In this section, the results of the pretrained residual-based CNN models are discussed. The three models, namely, ResNet-18, ResNet-50, and ResNet-101 are used with the pretrained weights for the image tampering detection. Table 4 shows the confusion matrix for the ResNet-18 model. It can be observed that the accuracy of the ResNet-18 model is 92.27%, and the percentage of the prediction of the correct tampered images is 50% and correct nontampered images is 42.27%. However, the wrong tampered image prediction is 7.23%. Table 5 shows the confusion matrix for the ResNet-50 model. It can be observed that the accuracy of the ResNet-50 model is 92.27%, and the percentage of the prediction of the correct tampered images is 50% and correct nontampered images is 42.27%. However, the wrong tampered image prediction is 7.23%. Table 6 shows the confusion matrix for the ResNet-101 model. It can be observed that the accuracy of the ResNet-101 model is 91.81%, and the percentage of the prediction of the correct tampered images is 50% and correct nontampered images is 41.82%. However, the wrong nontampered prediction is 8.18%.

The ROC curve is used to estimate the AUC values for the pretrained residual exploitation-based convolutional neural networks as shown in Figure 4. Figure 4(a) represents the ROC curve for the pretrained ResNet-18 model with AUC of 97.57%. Figure 4(b) represents the ROC curve for the pretrained ResNet-50 model with AUC of 97.57%. Figure 4(c) represents the ROC curve for the pretrained ResNet-101 model with AUC of 96.52%.

4.4. Fine-Tuned Residual Exploitation-Based CNN Models. In this section, the results of the fine-tuned residual exploitation-based models are discussed. The three models, namely, ResNet-18, ResNet-50, and ResNet-101 are used with the fine-tuned weights for image tampering detection. Table 7 shows the confusion matrix for the fine-tuned ResNet-18 model. It can be observed that the accuracy of the fine-tuned ResNet-18 model is 95.0%, and the percentage of the prediction of the correct tampered images is 50% and correct nontampered images is 45.0%. However, the wrong tampered image prediction is 5.0%. Table 8 shows the confusion matrix for the fine-tuned ResNet-50 model. It can be observed that the accuracy of the fine-tuned ResNet-50 model is 90.90%, and the percentage of the prediction of the correct tampered images is 50% and correct nontampered images is 40.91%. However, the wrong tampered image prediction is 9.09%. Table 9 shows the confusion matrix for the fine-tuned ResNet-101 model. It can be observed that the accuracy of the fine-tuned ResNet-101 model is 87.27%, and the percentage of the prediction of the correct tampered images is 45.45% and correct nontampered images is 41.82%. However, the prediction of the wrong tampered images is 8.18%, and wrong nontampered image prediction is 4.55%.

The ROC curve is used to estimate the AUC values for the fine-tuned residual exploitation-based models as shown in



(a) (b) (c) (d) (e) (f) (g) (h)

Figure 3: Continued.

(i)

Figure 3: (a–i) Tampered images with different combinations of geometrical and transformations attacks.

Table 3: Confusion matrix.

Actual images	Predicted tampered	Predicted nontampered
Tampered	(True positive) TP	(False negative) FN
Nontampered	(False positive) FP	(True negative) TN

Table 4: Confusion matrix for the pretrained ResNet-18 model.

Image dataset	Tampered image prediction	Nontampered image prediction
Tampered images	50%	0%
Nontampered images	7.23%	42.27%

Table 5: Confusion matrix for the pretrained ResNet-50 model.

Image dataset	Tampered image prediction	Nontampered image prediction
Tampered images	50%	0%
Nontampered images	7.23%	42.27%

Table 6: Confusion matrix for the pretrained ResNet-101 model.

Image dataset	Tampered image prediction	Nontampered image prediction
Tampered images	50%	0%
Nontampered images	8.18%	41.82%

Figure 5. Figure 5(a) represents the ROC curve for the fine-tuned ResNet-18 model with AUC of 97.0. Figure 5(b) represents the ROC curve for the fine-tuned ResNet-50 model with AUC of 93.91%. Figure 5(c) represents the ROC curve for the ResNet-101 model with AUC of 92.0%.

4.5. Fusion Model. In this section, the results of the fusion models are discussed. Table 10 shows the confusion matrix for the pretrained fusion models. It can be observed that the accuracy of the pretrained fusion model is 90.91%, and the percentage of the prediction of the correct tampered images is 49.55% and correct nontampered images is 41.36%. However, the wrong tampered image prediction is 8.64%, and wrong nontampered image prediction is 0.45%.

Table 11 shows the confusion matrix for the fine-tuned fusion model. It can be observed that the accuracy of the fine-tuned fusion network is 93.18%, and the percentage of the prediction of the correct tampered images is 50% and the correct nontampered images is 43.18%. However, the wrong tampered image prediction is 6.82%. It can be clearly observed that the percentage of wrong tampered image prediction is less as compared to the pretrained residual exploitation-based models. The accuracy of the fine-tuned fusion model is higher than the pretrained fusion model.

4.6. Performance Comparison

4.6.1. Performance Comparison with Pretrained Residual Exploitation-Based Models. In this section, the performance comparison of the fusion model is carried out with pretrained residual exploitation-based CNN models. The metrics used for the comparison are precision, recall, f-score, and accuracy. The results of the performance comparisons are as shown in Table 12. The precision and recall metrics are important to determine the effectiveness of the CNN models. According to Equations (7) to (10), the values in Table 12 were obtained.

4.6.2. Performance Comparison with Fine-Tuned Residual Exploitation-Based Models. In this section, the performance comparison of the fusion model is carried out with fine-tuned residual exploitation-based CNN models. The metrics used for the comparison are precision, recall, f-score, and accuracy. The results of the performance comparison are shown in Table 13.

It can be observed from the values in Table 13 that the proposed fusion model achieves comparatively more precision, recall, and f-score than the fine-tuned residual exploitation-based CNN models. The results of the performance comparison of the fusion model with the baseline

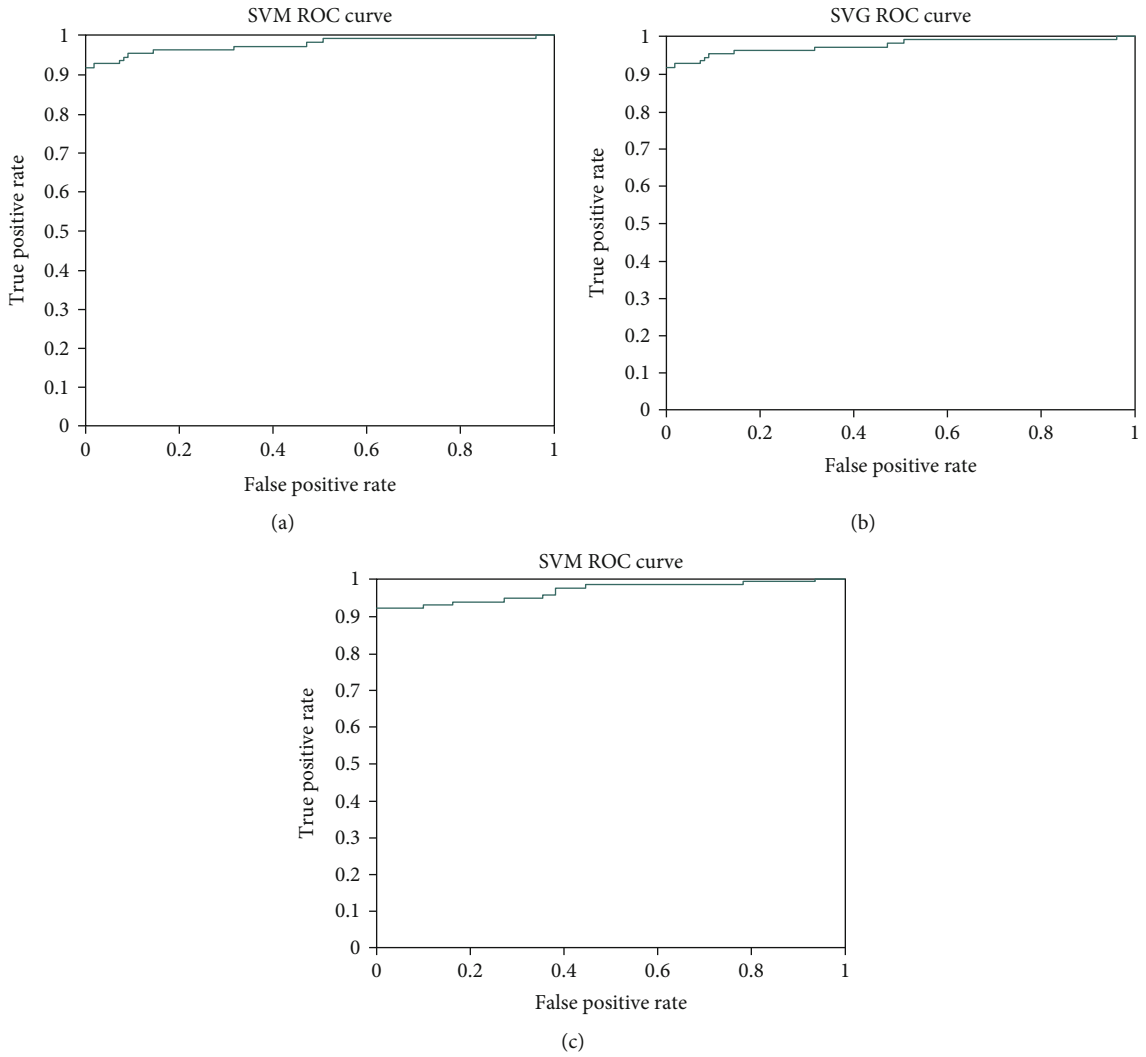


Figure 4: ROC curves for the pretrained residual-based models.

Table 7: Confusion matrix for the fine-tuned ResNet-18 model.

Image dataset	Tampered image prediction	Nontampered image prediction
Tampered images	50%	0%
Nontampered images	5.0%	45.0%

Table 8: Confusion matrix for the fine-tuned ResNet-50 model.

Image dataset	Tampered image prediction	Nontampered image prediction
Tampered images	50%	0%
Nontampered images	9.09%	40.91%

Table 9: Confusion matrix for the fine-tuned ResNet-101 model.

Image dataset	Tampered image prediction	Nontampered image prediction
Tampered images	45.45%	4.55%
Nontampered images	8.18%	41.82%

models are as shown in Table 14. The metrics used for the comparison are the FPR and TPR as they give the correctness of the model for the image tampering detection. The FPR for baseline 1 [37] is 8%, baseline 2 [38] is 3.64%, baseline 3 [39] is 84%, baseline 4 [40] is 86%, baseline 5 [41] is 2.89%, baseline 6 [42] is 5.45%, baseline 7 [43] is 7.63%, proposed pretrained fusion model is 17.27%, and proposed fine-tuned fusion model is 13.63%. The TPR for baseline 1 [37] is 100%, baseline 2 [38] is 73.64%, baseline 3 [39] is 89%, baseline 4 [40] is 87%, baseline 5 [41] is 96%, baseline 6 [42] is 83.64%, baseline 7 [43] is 97.87%, proposed pretrained fusion

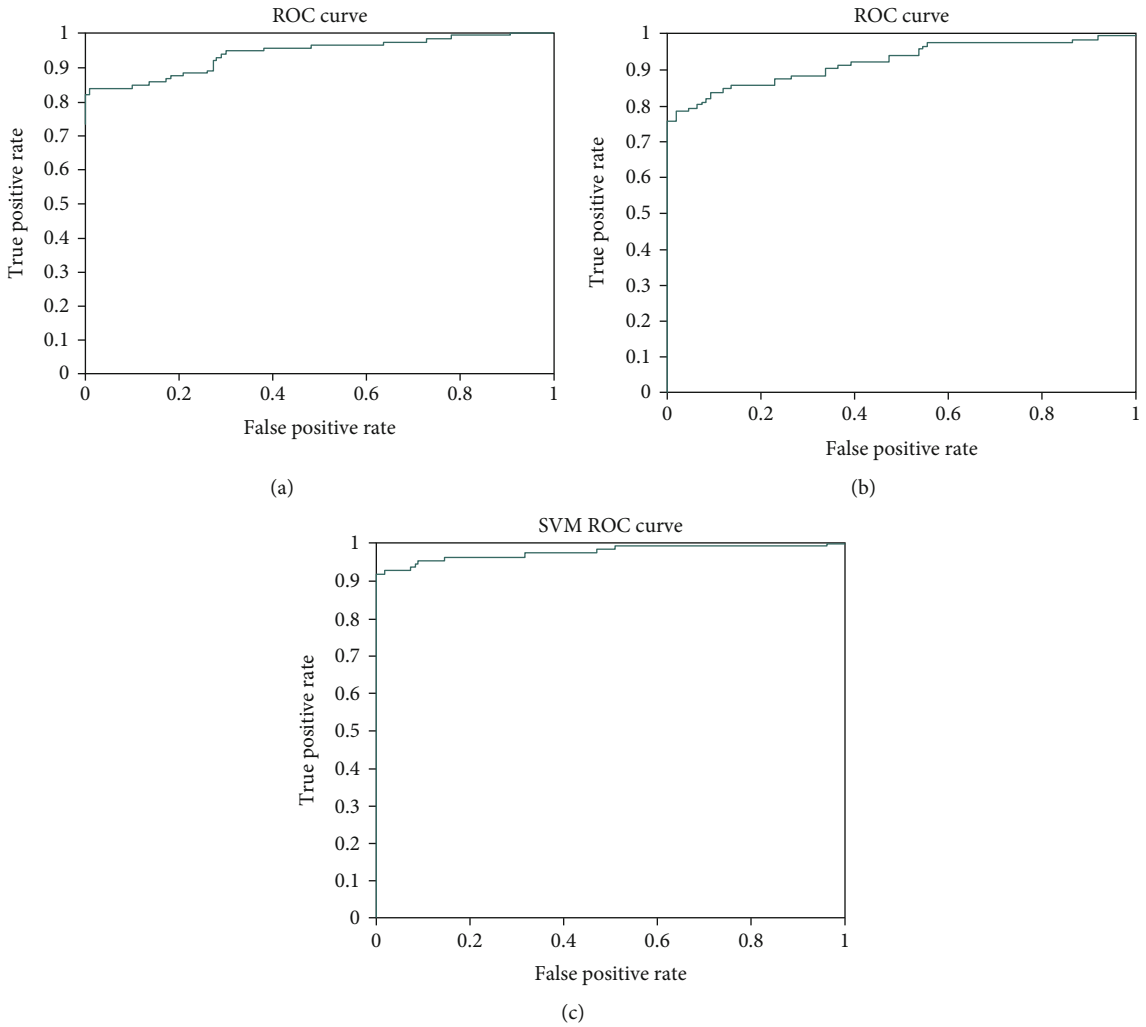


Figure 5: ROC curves for the fine-tuned residual-based models.

Table 10: Confusion matrix for the decision fusion for pretrained models.

Image dataset	Tampered image prediction	Nontampered image prediction
Tampered images	49.55%	0.45%
Nontampered images	8.64%	41.36%

Table 11: Confusion matrix for the decision fusion for fine-tuned models.

Image dataset	Tampered image prediction	Nontampered image prediction
Tampered images	50.0%	0%
Nontampered images	6.82%	43.18%

Table 12: Comparison of metrics for the pretrained residual exploitation-based models.

Model	Precision (%)	Recall (%)	*F*-score (%)	Accuracy (%)
ResNet-18	86.61	100	92.82	92.27
ResNet-50	86.61	100	92.82	92.27
ResNet-101	85.93	100	92.43	91.81
Fusion model	85.15	99.09	91.59	90.91

Table 13: Comparison of metrics for the fine-tuned residual exploitation-based models.

Model	Precision (%)	Recall (%)	*F*-score (%)	Accuracy (%)
ResNet-18	90.90	100	95.23	95.0
ResNet-50	84.61	100	91.66	90.90
ResNet-101	84.74	90.90	87.71	87.27
Fusion model	88.0	100	93.61	93.18

Table 14: Comparison of the proposed fusion-based models with baseline models.

Approach	FPR (%)	TPR (%)
SIFT [37]	8	100
SURF [38]	3.64	73.64
DCT [39]	84	89
PCA [40]	86	87
CSLBP [41]	2.89	96
SYMMETRY [42]	5.45	83.64
CLUSTERING strategy [43]	7.63	97.87
Pretrained fusion model (proposed)	17.27	99.09
Fine-tuned fusion model (proposed)	13.63	100

model is 99.09%, and proposed fine-tuned fusion model is 100%. Therefore, it can be observed that the fusion model has higher TPR as compared to the baseline models due to the weight initialization strategy used for the fusion model.

5. Conclusion

Image tampering detection helps to differentiate between the original and the manipulated or fake images. In this paper, a decision fusion of residual exploitation-based CNN models is implemented for image tampering detection. The idea was to use the residual exploitation-based CNN models, namely, ResNet-18, ResNet-50, and ResNet-101, and then combine all these models to obtain the decision to detect the tampering of the image. Regularization of the weights of the pretrained models is implemented to arrive at a decision of the image tampering. The experiments carried out indicate that the fusion-based approach gives more accuracy than the state-of-the-art approaches. In the future, the fusion decision can be improved with other weight initialization strategies for image tampering detection.

Data Availability

The data used for the research is already taken from public repository, and the link is provided in the paper.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] V. Manu and B. Mehtre, "Visual artifacts based image splicing detection in uncompressed images," in *2015 IEEE International Conference on Computer Graphics, Vision and Information Security (CGVIS)*, pp. 145–150, Bhubaneswar, India, 2015.

[2] Y. Cao, T. Gao, L. Fan, and Q. Yang, "A robust detection algorithm for copy-move forgery in digital images," *Forensic Science International*, vol. 214, no. 1–3, pp. 33–43, 2012.

[3] M. F. Hashmi, V. Anand, and A. G. Keskar, "Copy-move image forgery detection using an efficient and robust method combining un-decimated wavelet transform and scale invariant feature transform," *AASRI Procedia*, vol. 9, pp. 84–91, 2014.

[4] J. Li, X. Li, B. Yang, and X. Sun, "Segmentation-based image copy-move forgery detection scheme," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 3, pp. 507–518, 2015.

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, Las Vegas, NV, USA, 2016.

[6] X. Bi, Y. Wei, B. Xiao, and W. Li, "RRU-Net: the ringed residual U-Net for image splicing forgery detection," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1–10, Long Beach, CA, USA, 2019.

[7] A. K. Jaiswal and R. Srivastava, "Image splicing detection using deep residual network," in *Proceedings of 2nd International Conference on Advanced Computing and Software Engineering (ICACSE)*, UP, India, 2019.

[8] M. Heikkilä, M. Pietikäinen, and C. Schmid, "Description of interest regions with local binary patterns," *Pattern Recognition*, vol. 42, no. 3, pp. 425–436, 2009.

[9] Y. LeCun, B. Boser, J. S. Denker et al., "Backpropagation applied to handwritten zip code recognition," *Neural Computation*, vol. 1, no. 4, pp. 541–551, 1989.

[10] X. Liu, Z. Deng, and Y. Yang, "Recent progress in semantic image segmentation," *Artificial Intelligence Review*, vol. 52, no. 2, pp. 1089–1106, 2019.

[11] D. Ciresan, A. Giusti, L. Gambardella, and J. Schmidhuber, "Deep neural networks segment neuronal membranes in electron microscopy images," *Advances in Neural Information Processing Systems*, vol. 25, pp. 2843–2851, 2012.

[12] S. Ren, K. He, R. Girshick, X. Zhang, and J. Sun, "Object detection networks on convolutional feature maps," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 7, pp. 1476–1481, 2016.

[13] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431–3440, Boston, Massachusetts, 2015.

[14] J. H. Bappy, A. K. Roy-Chowdhury, J. Bunk, L. Nataraj, and B. Manjunath, "Exploiting spatial structure for localizing manipulated image regions," in *Proceedings of the IEEE international conference on computer vision*, pp. 4970–4979, California, USA, 2017.

[15] E. R. De Rezende, G. C. Ruppert, A. Theophilo, E. K. Tokuda, and T. Carvalho, "Exposing computer generated images by using deep convolutional neural networks," *Signal Processing: Image Communication*, vol. 66, pp. 113–126, 2018.

[16] D. Cozzolino, G. Poggi, and L. Verdoliva, "Efficient dense-field copy–move forgery detection," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 11, pp. 2284–2297, 2015.

[17] G. Muhammad, M. H. Al-Hammadi, M. Hussain, and G. Bebis, "Image forgery detection using steerable pyramid transform and local binary pattern," *Machine Vision and Applications*, vol. 25, no. 4, pp. 985–995, 2014.

[18] C. Guillemot and O. Le Meur, "Image inpainting: overview and recent advances," *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 127–144, 2013.

[19] B. Bayar and M. C. Stamm, "Design principles of convolutional neural networks for multimedia forensics," *Electronic Imaging*, vol. 2017, no. 7, pp. 77–86, 2017.

[20] Y. Rao and J. Ni, "A deep learning approach to detection of splicing and copy-move forgeries in images," in *2016 IEEE International Workshop on Information Forensics and Security (WIFS)*, pp. 1–6, Abu Dhabi, United Arab, 2016.

[21] B. Bayar and M. C. Stamm, "On the robustness of constrained convolutional neural networks to jpeg post-compression for image resampling detection," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2152–2156, New Orleans, LA, USA, 2017.

[22] Y. Qian, J. Dong, W. Wang, and T. Tan, "Deep learning for steganalysis via convolutional neural networks," in *Media Watermarking, Security, and Forensics, Vol. 9409*, International Society for Optics and Photonics, 2015.

[23] Y. Zhang, J. Goh, L. L. Win, and V. L. Thing, "Image region forgery detection: a deep learning approach," in *Proceedings of the Singapore Cyber-Security Conference (SG-CRC) 2016 - Cyber-Security by Design*, pp. 1–11, Singapore, 2016.

[24] O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Springer, Cham, 2015.

[25] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-net: learning dense volumetric segmentation from sparse annotation," in *International conference on medical image computing and computer-assisted intervention*, pp. 424–432, Springer, Cham, 2016.

[26] V. Iglovikov and A. Shvets, "Ternausnet: U-net with vgg11 encoder pre-trained on imagenet for image segmentation," 2018, http://arxiv.org/abs/11801.05746.

[27] H. Jégou, F. Perronnin, M. Douze, J. Sánchez, P. Pérez, and C. Schmid, "Aggregating local image descriptors into compact codes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 9, pp. 1704–1716, 2011.

[28] F. Perronnin and C. Dance, "Fisher kernels on visual vocabularies for image categorization," in *2007 IEEE conference on computer vision and pattern recognition*, pp. 1–8, Minneapolis, MN, USA, 2007.

[29] M. Brezina, A. J. Cleary, R. D. Falgout et al., "Algebraic multigrid based on element interpolation (AMGe)," *SIAM Journal on Scientific Computing*, vol. 22, no. 5, pp. 1570–1592, 2001.

[30] T. Vatanen, T. Raiko, H. Valpola, and Y. LeCun, "Pushing stochastic gradient towards second-order methods–backpropagation learning with transformations in nonlinearities," in *International Conference on Neural Information Processing*, pp. 442–449, Springer, Berlin, Heidelberg, 2013.

[31] P. Korus and J. Huang, "Multi-scale fusion for improved localization of malicious tampering in digital images," *IEEE Transactions on Image Processing*, vol. 25, no. 3, pp. 1312–1326, 2016.

[32] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, pp. 1440–1448, 2015.

[33] Y. Zhu, C. Chen, G. Yan, Y. Guo, and Y. Dong, "A R-net: adaptive attention and residual refinement network for copy-move forgery detection," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 10, pp. 6714–6723, 2020.

[34] D. Cozzolino, G. Poggi, and L. Verdoliva, "Recasting residual-based local descriptors as convolutional neural networks: an application to image forgery detection," in *Proceedings of the 5th ACM Workshop on Information Hiding and Multimedia Security*, pp. 159–164, New York, NY, USA, 2017.

[35] B. Chen, X. Qi, Y. Zhou, G. Yang, Y. Zheng, and B. Xiao, "Image splicing localization using residual image and residual-based fully convolutional network," *Journal of Visual Communication and Image Representation*, vol. 73, p. 102967, 2020.

[36] O. M. Al-Qershi and B. E. Khoo, "Evaluation of copy-move forgery detection: datasets and evaluation metrics," *Multimedia Tools and Applications*, vol. 77, no. 24, pp. 31807–31833, 2018.

[37] 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 Transactions on Information Forensics and Security*, vol. 6, no. 3, pp. 1099–1110, 2011.

[38] A. Kuznetsov, "Digital image forgery detection using deep learning approach," *Journal of Physics: Conference Series*, vol. 1368, no. 3, article 032028, 2019.

[39] A. J. Fridrich, B. D. Soukal, and A. J. Lukáš, "Detection of Copy-Move Forgery in Digital Images," *in Proceedings of Digital Forensic Research Workshop*, CiteSeerX, 2003.

[40] A. C. Popescu and H. Farid, *Exposing digital forgeries by detecting duplicated image regions*, Tech. Rep. TR2004-515, Dartmouth, 2004.

[41] D. M. Uliyan, H. A. Jalab, A. Abdul Wahab, S. Sadeghi, and S. Sadeghi, "Image region duplication forgery detection based on angular radial partitioning and Harris key-points," *Symmetry*, vol. 8, no. 7, p. 62, 2016.

[42] D. Vaishnavi and T. Subashini, "Application of local invariant symmetry features to detect and localize image copy move forgeries," *Journal of Information Security and Applications*, vol. 44, pp. 23–31, 2019.

[43] M. Abdel-Basset, G. Manogaran, A. E. Fakhry, and I. El-Henawy, "2-Levels of clustering strategy to detect and locate copy-move forgery in digital images," *Multimedia Tools and Applications*, vol. 79, pp. 5419–5437, 2020.