IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 24 June 2023, accepted 11 August 2023, date of publication 22 August 2023, date of current version 30 August 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3307357

RESEARCH ARTICLE

Enhancing Digital Image Forgery Detection Using Transfer Learning

ASHGAN H. KHALIL[1], ATEF Z. GHALWASH[1], HALA ABDEL-GALIL ELSAYED[1], GOUDA I. SALAMA[2], AND HAITHAM A. GHALWASH[3]

[1]Computer Science Department, Faculty of Computers and Artificial Intelligence, Helwan University, Cairo 4271184, Egypt
[2]Department of Computer Engineering, MTC, Cairo 4393010, Egypt
[3]Ethical Hacking and Cyber Security, Coventry University–Egypt Branch, New Cairo, Egypt

Corresponding author: Ashgan H. Khalil (Ashganheissen_csp@fci.helwan.edu.eg)

ABSTRACT Nowadays, digital images are a main source of shared information in social media. Meanwhile, malicious software can forge such images for fake information. So, it's crucial to identify these forgeries. This problem was tackled in the literature by various digital image forgery detection techniques. But most of these techniques are tied to detecting only one type of forgery, such as image splicing or copy-move that is not applied in real life. This paper proposes an approach, to enhance digital image forgery detection using deep learning techniques via transfer learning to uncover two types of image forgery at the same time, The proposed technique relies on discovering the compressed quality of the forged area, which normally differs from the compressed quality of the rest of the image. A deep learning-based model is proposed to detect forgery in digital images, by calculating the difference between the original image and its compressed version, to produce a featured image as an input to the pre-trained model to train the model after removing its classifier and adding a new fine-tuned classifier. A comparison between eight different pre-trained models adapted for binary classification is done. The experimental results show that applying the technique using the adapted eight different pre-trained models outperforms the state-of-the-art methods after comparing it with the resulting evaluation metrics, charts, and graphs. Moreover, the results show that using the technique with the pre-trained model MobileNetV2 has the highest detection accuracy rate (around 95%) with fewer training parameters, leading to faster training time.

INDEX TERMS Deep neural network (DNN), image compression, image forgery detection (IFD), pretrained model, transfer learning.

I. INTRODUCTION

The tampering of a digital image is called digital image forgery, these forged images cannot be detected by the naked eye. Such images are the primary sources of spreading fake news and misleading information in the context of society with the aid of diverse social media platforms like Facebook, Twitter, etc. [1]. The editing software tools that can make these forgeries are available for free with some advanced features that are used for image tampering such as GNU, GIMP, and Adobe Photoshop [2]. Such forgeries can be detected using digital image forgery algorithms and techniques, these algorithms are used in image security especially when the original content is not available [3].

Digital image forgery means adding unusual patterns to the original images that create a heterogeneous variation in image properties and an unusual distribution of image features [3]. Figure 1 shows the classification of digital image forgery.

Active approaches require essential information about the image for the verification process. The inserted information within the picture is employed to observe the modification in that picture. The active approach consists of two types: digital signatures which insert some additional data obtained from an image by the end of the acquisition process, and digital watermarking which is inserted into images either during the acquisition phase or during the processing phase.

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

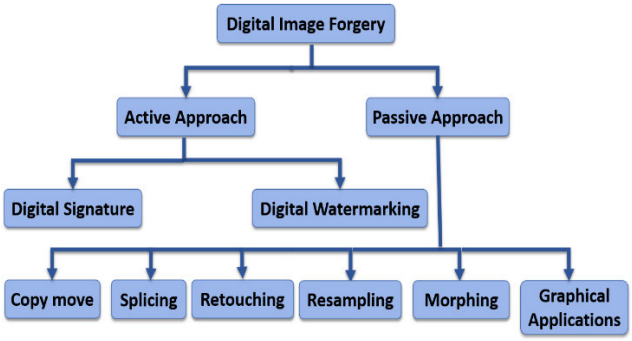


FIGURE 1. Digital image forgery classification.

The passive image forgery detection methods benefit from the features retained by the image allocation processes achieved in different stages of digital image acquisition and storage. Passive methodologies do not require past information about the image. These approaches exploit that the tampering actions modify the contents of information of the image that can facilitate tampering detection [4].

Copy move forgery involves duplicating a section or object within an image and pasting it again in a different location within the same image to replicate (or move) a specific scene in the image. Copy-move forgery is the most common technique used to manipulate images, it is also the most challenging type of forgery to detect due to the complexity of copying and replicating an object or section of the image with identical properties and feature distributions and pasting it within the same image [3]. some post-processing techniques can be added after CMF processes such as rotation, scaling, JPEG compression, etc. which makes the detection further difficult and complex [2].

Splicing forgery can be generated by adding or blending two images or set of images to produce an unprecedented image [3]. The source images used to generate a spliced image may include dissimilar color temperatures, illumination conditions, and noise levels based on various factors. Average filtering or some other related image processing operation can be applied as postprocessing like resizing, cropping, rotating, and retouching each of the source images to match the visual attributes, shape, and size of the target image so that the forged image can look realistic [5].

Retouching forgery involves modifying an image to hide or highlight particular features such as brightness, color, contrast, or other visual attributes and altering background coloring. It includes the visual quality enhancement of the image. **Resampling Forgery** is the act of altering the dimensionality of a particular object or section within an image to present a distorted or misleading view. **Morphing forgery** involves merging two scenes from different images to create an entirely new scene, this can be done through the use of graphic software to create a completely artificial image with no basis in reality [3]. The three major types of tampering are Copy Move, Image Splicing, and Image Retouching [4].

Digital Image Forgery Detection is a binary classification task, to classify the image as either forged or authentic. Recently, deep learning has become a promising tool for enhancing digital image forgery detection. In any Deep learning model, feature extraction is an important phase that affects the performance of the algorithm [6], where the database size is considered a significant factor. Transfer learning presents a viable alternative solution when dealing with limited sample size problems that supports taking the knowledge acquired from a previously trained model including features, weights, and other relevant information that was trained on a large dataset such as the ImageNet database, that contains 1.2 million images grouped into 1000 classes to solve the problem of small size dataset in the new target domain [7]. By utilizing a pre-trained model, significant amounts of time spent on training can be saved, and the model can be adapted to work with smaller datasets through retraining [8].

The motivation behind image forgery detection is to check the authenticity of digital images, especially when images are used as evidence in court and forensics, news, or historical data, or in the military, and medical diagnosis systems, it prevents the distribution of misinformation and fake news, particularly in social media and online platforms, these forged images can be used to destroy someone's reputation or mislead public opinion, or for distorting the truth in news reports, they can also be used to exaggerate the capabilities of the countries army.

Image forgery detection has several challenges due to the nature of image manipulation techniques. These challenges can be concluded as:

- Computational Complexity and the limitation of the CPU and memory is the main challenge, which takes a large training time and most of the time runs out of memory even with high memory specifications.
- Detecting more than one type of image forgery at the same time affects the accuracy rate, so there is a need to improve its accuracy rate.
- There is a need to solve the problem of the accuracy-speed trade-off.
- Most image forgery detection techniques that have high detection accuracy are very complex, there is a need for a simpler technique with high detection accuracy rate.
- Most image forgery detection techniques suffer from detecting images that lie under post-processing operations like image rotation, scaling, blurring, brightness adjustment, and adding noise.

This paper presents the following contribution:

- Detecting two types of passive image forgeries like image splicing and copy-move at the same time to be suitable for real-life scenarios.

- Achieving a high accuracy rate compared to the state-of-the-art results found in the literature. Moreover, using a pre-trained model and taking the power of transfer learning, with a small number of parameters, the developed lightweight model is well-suited for environments with memory and CPU limitations. This is an added value in favor of the proposed architecture.
- Evaluating the performance of eight different pre-trained models such as VGG16, VGG19, ResNet50, Reset101, ResNet152, MobileNetV2, Xception and DenseNet are considered.
- A comparative analysis of the eight forementioned pre-trained models and state of art is presented.
- Using the CASIAV2 dataset which is one of the best benchmark datasets that is considered as the main challenge itself, it contains two main types of image forgery (splicing and copy-move) with different sizes and contains many types of image formats (TIFF, JPEG, and BMP) and also the cropped parts in the forged images underwent some processing including distortion, rotation, and scaling, to create an image that seems to be real, involving blurring the spliced region's edge, which makes the detection process challenging.

The paper is organized as follows: A literature review is covered in section II. Section III discusses the proposed approach and presents the proposed architecture in detail. Section IV outlines the experimental results and discussion, along with the experimental setup and dataset structure. Section V has the conclusion and future work.

II. LITERATURE REVIEW

In image forgery detection field, various approaches were proposed. Traditional techniques mostly extract a set of hand-crafted based features, followed by a classifying technique like feature matching to differentiate between the authentic and forged images. In the machine learning approach, a set of classifiers can be used in the classifying process like Support Vector Machine and Naïve Bayes classifier. While more recent techniques employ convolutional neural networks (CNNs) and deep neural networks (DNN) methods, others employ the network with the help of pre-trained models and the power of transfer learning. CNN and deep learning-based techniques will be discussed moving over the use of different pre-trained models.

A. DEEP NEURAL NETWORK-BASED IMAGE FORGERY DETECTION TECHNIQUES

DNNs can autonomously learn an extensive number of features. Over the past few years, a variety of image forgery detection methods have been proposed, for detecting image forgery, where many of which relied on deep learning [5]. By constructing an appropriate neural network, deep learning networks can identify complex hidden patterns in data and effectively distinguish the forged parts from the original image [9]. Deep learning technique has proven to be effective in resolving many activities or issues that machine learning algorithms were previously unable to address [8].

When considering splicing detection, a scheme was proposed in [10] based on the local feature descriptor which is learned by a DNN. An improved initialization based on the (SRM) was proposed and developed a splicing localization scheme based on the proposed CNN model and fully connected conditional random field (CRF) with SVM which is robust against JPEG compression. In [11], a (CNN) model was developed using a relatively small number of parameters that can be used as an on-time detection model.

For splicing and copy-move separately, an end-to-end fully CNN that combines multi-resolution hybrid features, from RGB and noise streams was introduced in [12], where a tamper-guided dual self-attention (TDSA) module was designed to capture the difference between tampered and non-tampered areas and segments them from the image. A proposed hybrid features and semantic reinforcement network (HFSRNet) for IFD at the pixel level was proposed in [13], where the network employs an encoding and decoding approach and utilizes Long-Short Term Memory (LSTM) technology.

For copy move, [14] introduced a copy-move forgery detection and localization model based on super boundary-to-pixel direction (super-BPD) segmentation and deep CNN (DCNN). Starting with employing the segmentation technique that is used to enhance the connection among identical image blocks, thereby improving the accuracy of forgery detection, the DCNN is used to extract image features, ending by using image BPD information to optimize the edges of the rough detected image and obtain the final detected image. [15] developed a deep learning CNN model which used multi-scale input and multiple stages of convolutional layers, with two different parts, encoder, and decoder. In [16], a simple and lightweight convolutional neural network (CNN) has been proposed for the automatic detection of copy-move forgery detection, which has a high detection accuracy rate.

For copy-move and splicing together, [9] used a new image segmentation model U-Net by adding L2 regularization. Reference [17] introduced a system for IFD using double image compression, in which the difference between an original image and recompressed one was used in training the model, the method is capable of detecting both image splicing and copy-move together.

B. PRETRAINED NETWORK-BASED IMAGE FORGERY DETECTION TECHNIQUES

Different IFD techniques based on transfer learning will be discussed in this section. For splicing, [18] presented multiple image-splicing forgeries using Mask R-CNN and MobileNet-V1 backbone. A novel approach utilizing ResNet50v2 was

introduced in [19], that considered image batches as an input and used YOLO CNN weights with ResNet50v2 architecture.

For splicing and copy-move separately, [20] proposed a multi-task learning network called FBI-Net based on (DCT). The network employs a fully convolutional encoder-decoder architecture, and the Dilated Frequency Self-Attention Module (DFSAM) in the bridge layer adjusts fused features. Reference [21] introduced a lightweight model using mask R-CNN with MobileNet to detect copy-move and image-splicing forgeries.

For copy move, [22] used SmallerVGGNet and MobileNet-V2, time- and memory-saving deep learning models. In [23] an Optimal Deep Transfer Learning based Copy Move Forgery Detection (ODTLCMFD) technique was presented that derived a DL model for the classification of target images and then localized the copy moved regions. They used the MobileNet model with a political optimizer (PO) for feature extraction and the least square support vector machine (LS-SVM) model with an enhanced bird swarm algorithm (EBSA) for classification. They utilized the EBSA algorithm to modify the parameters in the Multiclass Support Vector Machine (MSVM) technique to enhance the classification performance. Reference [24] provided an automated deep learning-based fusion model for detecting and localizing copy-move forgeries (DLFM-CMDFC), that combined models of generative adversarial networks (GANs) and densely connected networks (DenseNet). The two outputs were merged in the DLFM-CMDFC technique to create a layer for encoding the input vectors with the first layer of an extreme learning machine (ELM) classifier. The ELM model's weight and bias values were modified using the artificial fish swarm algorithm (AFSA). The networks' outputs were supplied into the merger unit as input.

For splicing and copy-move together, a multimodal system was proposed in [25], which covers classification and localization, forgery detection through a deep neural network followed by part-based image retrieval classification. The localization of manipulated regions was accomplished using a deep neural network. InceptionV3 was employed for feature extraction. The Nearest Neighbor Algorithm was used to retrieve Potential donors and nearly duplicates. In [26] a novel approach to detect copy move and splicing image forgery using a Convolutional Neural Network (CNN), with three different models was presented, namely, ELA (Error Level Analysis), VGG16, and VGG19. The proposed method applied the pre-processing technique to obtain the images at a particular compression rate. These images were then utilized to train the model, where the images were classified as authentic or forged.

TABLE 1 summarizes the image forgery detection techniques based on deep learning, and TABLE 2 summarizes the image forgery detection techniques based on transfer learning. Both tables show that previous research reveals that some efforts were made for image splicing forgery detection with a high detection accuracy rate. Image splicing seems to be the easiest type to detect, Meanwhile, a lot of efforts were made to detect copy-move which seems to be difficult to detect. It is also worth to be noticed that there were a few research studies done for detecting both splicing and copy-move at the same time where less detection accuracy rate was recorded compared to the other techniques.

TABLE 1. Summary of deep learning-based image forgery detection techniques.

Forgery Type	Reference	Year	Features Extraction technique	Classification technique	Dataset	Evaluation
Splicing	[10]	2020	CNN-Based Local Descriptor Construction	SVM	CASIAv2 DVMM DSO-1	**Accuracy**: CASIAV2= 96.97%, DVMM= 97.04%, DSO-1= 97.5%
	[11]	2023	CNN	CNN	CASIAv1 CASIAv2 CUISDE	**Accuracy:** CASIAV1= 99.1%, CASIAV2= 99.3%, CUISDE= 100%
Splicing, Copy-Move Separately	[12]	2022	RGB stream + noise stream	End-to-end fully CNN + (TDSA)	NIST16 CASIA COLUMBIA	**Accuracy:** NIST16= 98.4%, COLUMBIA=97.7%
	[13]	2021	Hybrid Encoding+ Decoding CNN	Hybrid features and semantic reinforcement network HFSRNet	NIST16 COVERAGE CASIAv1	**Accuracy:** NIST16= 98.86%, COVERAGE= 92.76%, CASIAV1= 93.21%
Copy-Move	[14]	2022	DCNN	SD-Net: (super-BPD) +DCNN:	USCISI CoMoFoD CASIAv2	CoMoFoD **P**=59.11 **R**=57.69 **F**=50.77 CASIAv2: **P**=57.48 **R**=51.25 **F**=48.06
	[15]	2021	CNN (Encoder+ decoder)	CNN	CoMoFoD CMFD	**Accuracy**: CoMoFoD= 98.39%, CMFD= 98.78%
	[16]	2022	CNN	CNN	MICC-F2000	**Accuracy**= 97.52%
Splicing + Copy-Move Together	[9]	2021	Regularizing U-Net	Regularizing U-Net	CASIAv2	**F1-Score** = 0.9486
	[17]	2022	Difference Compression Quality +CNN	CNN	CASIAv2	**Accuracy**= 92.23%

Recently, image forgery detection techniques relied on deep learning only that needs the availability of large data sets for training which is not available in the real. This problem can be solved with the help of pre-trained models and the power of transfer learning. In addition, previous researches seems to be very complex to reach a high detection accuracy rate. Moreover, different evaluation matrices were considered in the studies which add difficulties when comparing such techniques. Also, not all studies are concerned with image pre-processing such as rotation, scaling, and blurring, which adds difficulties to the detection process. The above-mentioned considerations trigger the motivation to consider the transfer learning technique to build the proposed model.

In [17], the authors focused on the fact that CNNs can be utilized to detect image forgery, which is difficult for the human eye to detect, due to artifacts left by the forgeries. Also, the source of the forged region and the background images are different. This makes it easy to allocate the forged region by compression differences between both. This difference was utilized to train the CNN-based model to identify image forgery. The experimental results of [17] showed that the CNN model achieved a 92.3% detection accuracy rate which still needs improvement. In addition, the model has a large number of parameters that need to be reduced to save CPU and memory consumption. Also, the

TABLE 2. Summary of transfer learning-based image forgery detection techniques.

Forgery Type	Reference	Year	Features Extraction technique	Classification technique	Dataset	Evaluation
Splicing	[18]	2021	Mask R-CNN + MobileNetV1	Mask R-CNN + MobileNetV1	MISD CASIAv1 WildWeb Columbia Gray	**Average Precision**: MISD=82% CASIAv1=74% WildWeb= 81% Columbia Gray= 86% **F1-Score**= MISD=67% CASIAv1=64% WildWeb= 68% Columbia Gray= 61%
	[19]	2022	ResNet50v2+YOLO CNN weights	ResNet50v2+ YOLO CNN weights	CASIAv1 CASIAv2	**Accuracy**= 99.3%
Splicing, Copy-Move Separately	[20]	2022	DFSAM	ResNet-18	CASIAv1 CASIAv2 Carvalho Columbia Coverage IMD2020	**Average of IoU**= 70.99% and **F1-score**= 76.98%
	[21]	2022	Mask R-CNN with MobileNet V1	Mask R-CNN with MobileNet V1	COVERAGE CASIAv1 CASIAv2 MICCF220 MICCF600 MICCF2000 COLUMBIA	**F1-score:** MICC F600=70% for copy-move, CASIA1.0=64% for image splicing. **Average Precision**: MICC F2000, COVERAGE= 90% for copy-move, COLUMBIA= 90% for image splicing
Copy-Move	[22]	2021	SmallerVGGNet, MobileNetV2	SmallerVGGNet, MobileNetV2	CoMoFoD MICC-F2000 CASIAv2	**Accuracy**: SmallerVGGNet= 87% MobileNetV2= 85%
	[23]	2023	Political Optimizer (PO) +Mobile Networks (MobileNet)	EBSA + LS-SVM	MICC-F220 MICCF-2000 MICC-F600	**Accuracy**: MICC-F220, MICC-F2000 =98.6% MICC-F600= 98.3%
	[24]	2021	GANs + DenseNets	ELM classifier	MNIST, CIFAR10	MNIST: **Precision**= 95.42% **Recall**= 95.89%, **Accuracy**= 95.42%, **F-score**= 95.82%, CIFAR-10: **Precision**= 97.27% **Recall**= 96.46% **Accuracy**= 96.94% **F-score**= 96.06%
Splicing + Copy-Move Together	[25]	2022	DNN + InceptionV3	DNN + InceptionV3	CASIAv2 CoMoFoD NIST 2018	**Accuracy:** CASIAv2= 93.04% CoMoFoD+CASIAv2= 88.90% CoMoFoD + CASIAv2 +NIST 2018 =89.01%
	[26]	2022	ELA	CNN, VGG16, VGG19	CASIAv2 NC2016	**Accuracy**: CNN=70.6%, VGG16=71.6%, VGG19=72.9%

performance parameters e.g., F1 score, recall, precision, TPR, and TNR need to be improved by increasing their values. In addition, the model has large FPR and FNR which need to be decreased. All these evaluation matrices will be discussed in the following section.

III. PROPOSED APPROACH

The proposed approach considers the fact shown in [17], that copying a part of an image from one to another may impose some changes in the image properties due to the different sources of the images. Although these changes may not be detectable to the human eye, they can be detected by CNNs in manipulated images.

The proposed model aims to avoid all of the forementioned drawbacks, by adapting the idea of calculating the difference in compression qualities to produce the featured image as an input to a deep neural network with the assistance of a pre-trained model to benefit from the power of transfer learning. As a result, the evaluation matrix will be improved including the accuracy rate that will get better than that which was recorded when using CNN in [17]. This will be elaborated and discussed in the following section.

In a forged image, if the image is compressed, the forged section of the image will be compressed differently than the rest of the image. This is because the source of the original image differs from the source of the forged section. When analyzing the difference between the original image and its compressed version, the forgery component becomes more distinguished. Therefore, this aspect can be utilized to train a DNN-based model for detecting image forgery.

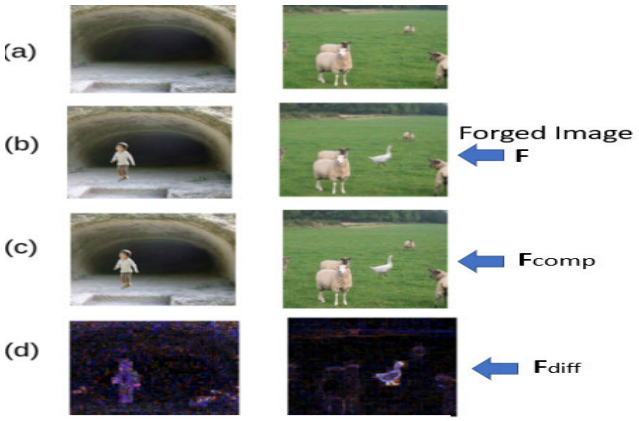


FIGURE 2. Set of images created in the proposed work.

The set of images created in the proposed work can be shown in Figure 2. The first image, (a) represents the original image without forgery, (b) represents the forged image that is denoted as F, (c) represents the compressed version of (b) that is denoted as Fcomp, (d) represents the mathematical difference between F and Fcomp denoted as Fdiff.

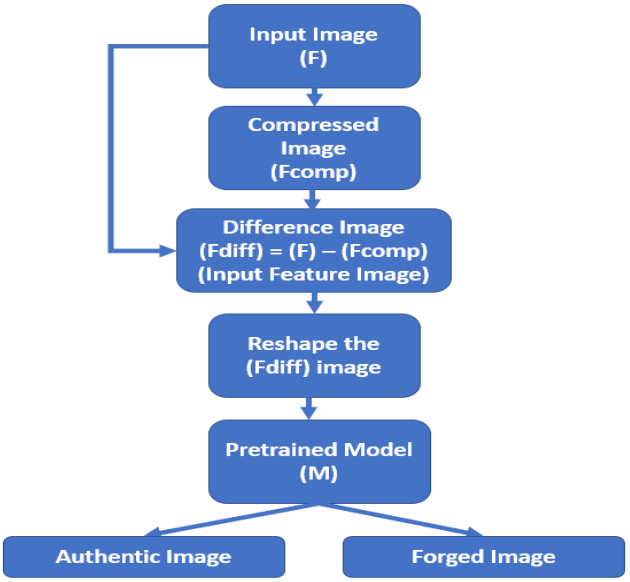


FIGURE 3. Flowchart of the proposed system (System Architecture).

In the proposed model, the preprocessing phase starts with the forged image (input image), denoted as F, as shown in Figure 3, which is compressed to get a compressed version of the input image, denoted as Fcomp. The difference between the forged image and its compressed version is then calculated by mathematical subtraction, denoted as (Fdiff), as shown in Equation (1).

$$\text{Fdiff} = \text{F} - -\text{Fcomp} \tag{1}$$

As a result, the forged part of the image appears in (Fdiff) due to the difference between the source of the forged and original parts. Fdiff is then reshaped to 160 × 160 pixels to fit as an input feature image for training a pre-trained model (M), which is then used to classify images as forged or authentic. Figure 3 shows the overall architecture of the proposed system.

In Figure 3, the pre-trained model, shown as block (M), is used to extract features from input images (Fdiff) and classify them as authentic images or forged images. In this block (pre-trained model), eight different pre-trained models are considered (one at a time) namely, VGG16 [27], VGG19 [28], ResNet [29], DenseNet [30], Xception [31], and MobileNet [32] for fine training with input images (Fdiff), to nominate the model with the best performance among them.

Each model of the forementioned eight pre-trained models has its own architecture which consists of a set of convolutional layers with activation functions and ends with a set of fully connected layers that can classify up to 1000 classes of images. So, each model architecture has to be modified to fit the binary classification problem with only two classes (authentic or forged images) as in the case of image forgery detection problems. Therefore, the native fully connected layers in each model are replaced with a new set of fully connected classification layers able to handle the binary classification problem at hand. The convolutional layers in every model should remain untouched since they contain all the trainable parameters used in transfer learning.

Figure 4 shows the detailed architecture of the proposed model classifier with the newly added layers. After removing the fully connected layers of the pre-trained model, a flatten layer is added to convert the input data, which is typically a multi-dimensional array, into a one-dimensional vector that can be fed to the next layers. The next two (new) layers are fully connected layers added with the ReLU activation function. The two layers have 1024 and 256 neurons, respectively. After each layer, a dropout 0.5 was added to prevent overfitting by randomly dropping out (setting to zero) about 50% of the output values of the previous layer will be randomly set to zero during the training phase. The last fully connected layer with a sigmoid activation function is added, which is the common activation function used in binary classification problems.

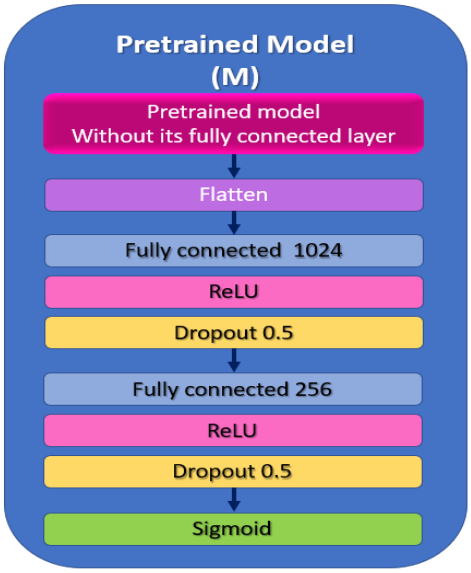


FIGURE 4. Detailed view of the proposed model classifier.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

This section will discuss the training and testing environmental setup used in the proposed approach and compare its performance with state-of-the-art techniques.

A. EXPERIMENTAL SETUP

1) ENVIRONMENT

For the experiment's environment, Google Colab Pro was used with premium GPUs and 100 compute units, with 25 GB system RAM and 16 GB GPU RAM.

2) DATASET

The CASIA 2.0 dataset is used in the experiments, the same dataset used in [17]. As described in [33], CASIA 2.0 has a total of 12614 images, with 7491 original images and 5123 forged images, including 3274 copy-move images and 1849 spliced images. The images are in JPEG and TIFF formats. The pixel dimensions of the images range from 320 × 240 to 900 × 600 [18]. TABLE 3 provides all information about the CASIA 2.0 database.

TABLE 3. CASIA.2.0 image forgery database specification.

	Authentic	Forged		Total	Size	Format
		Copy-move	Splicing			
Number of images	7491	3274	1849	12614	320x240 900x600	BMP, JPEG, TIFF
Total	7491	5123		12614		

Tampered images in CASIA v2 were created by combining two different authentic images or using the same authentic image. Cropped parts underwent some processing including distortion, rotation, and scaling, to create an image that seems to be real, involving blurring the spliced region's edge [19].

TABLE 4. Details division of CASIA-V2 dataset in the experiments.

	Authentic	Forged	Total
CASIA-V2	7491	5123	12614
Training Set 80%	5993	4098	10091
Testing Set 20%	1498	1025	2523

In the experiments, the dataset is divided into two sets, training, and testing sets with ratios of 80% and 20%, respectively. As TABLE 4 shows, the testing set is used as a testing and validation set as done in the paper [17]. The training set contains 10091 images, which are divided into 5993 authentic images and 4098 forged images, and the testing set contains 2523 images divided into 1498 authentic images and 1025 forged images.

B. EVALUATION METRICS

The performance of the proposed model is evaluated using the metrics specified in [16].

•**Accuracy:** The accuracy is determined by dividing the total number of correctly classified instances from both classes by the total number of instances in the dataset.

$$\text{Accuracy} = [(\text{TP} + \text{TN})/(\text{TP} + \text{FN} + \text{FP} + \text{TN})] \times 100 \quad (2)$$

•**Recall:** is the percentage of tampered images that were correctly classified out of the total number of images that were actually tampered.

$$\text{Recall} = \text{TP}/(\text{TP} + \text{FN}) \quad (3)$$

•**Precision:** is the proportion of images identified as forged and that are truly forged.

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP}) \quad (4)$$

•**F1 score:** is a measure of the accuracy of a test, which is defined as the harmonic mean of the precision and recall.

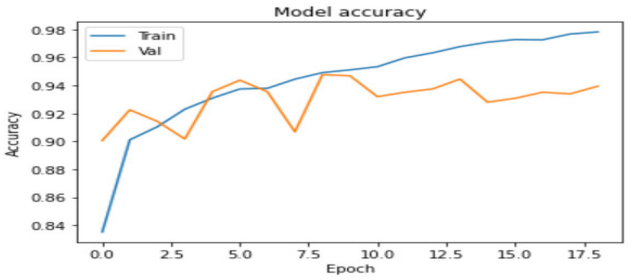
$$\text{F1 score} = [(2 \times \text{Recall} \times \text{Precision})/(\text{Recall} + \text{Precision})] \times 100 \quad (5)$$

C. MODEL TRAINING AND TESTING EVALUATION

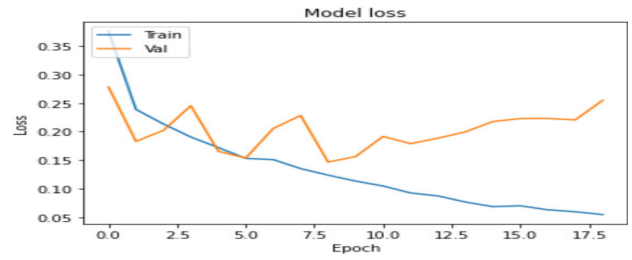
In order to fairly evaluate the training and testing phase for the eight different pre-trained models, a set of initial value parameters should be fixed all over the eight experiments. These parameters are as follows: The size of the input images is 160 × 160, with initial weight 'ImageNet', the number of epochs =100 with early stopping condition monitoring the minimum validation loss with patience = 10. The optimizer used is the "Adam" optimizer with learning rate = 1e-5 and loss 'binary cross entropy'.

In the experiments, the relation between the training and validation curve for the accuracy and the loss for each pre-trained model experiment is drawn, and three samples from them are displayed in Figures 5,6 and 7. In each figure, (a) displays the relationship between the training and validation accuracy, and (b) displays the relationship between the training and validation loss for each model.

These graphs are useful in many directions, the training accuracy curve shows how well the model is learning from the training data over time. As shown the curve generally increases as the model gets better at fitting the training data.

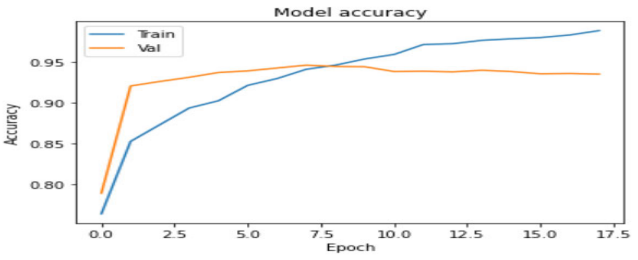


(a) VGG19 training and validation accuracy curve

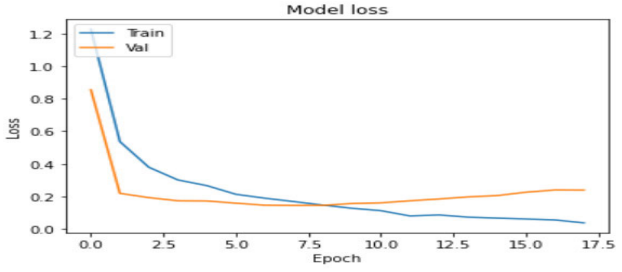


(b) VGG19 training and validation loss curve

FIGURE 5. VGG19 training and validation curves.

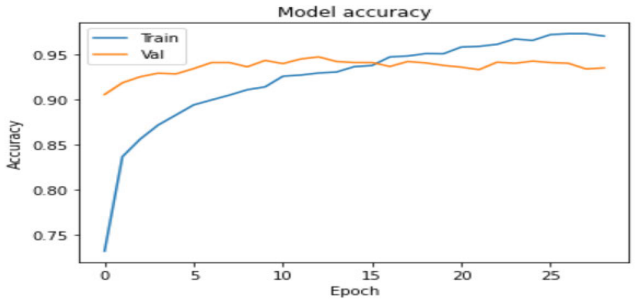


(a) ResNet50 training and validation accuracy curve

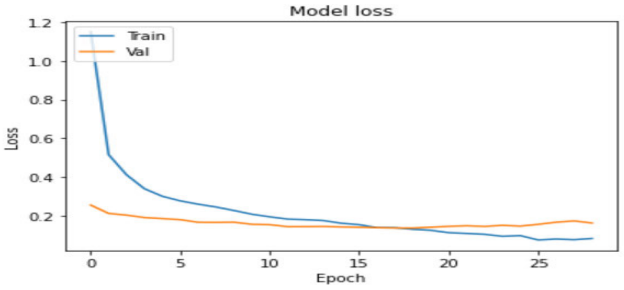


(b) ResNet50 training and validation loss curve

FIGURE 6. ResNet50 training and validation curves.



(a) MobileNetV2 training and validation accuracy curve.



(b) MobileNetV2 training and validation loss curve

FIGURE 7. MobileNetV2 training and validation curves.

On the other side, the validation accuracy curve shows how well the model is performing on a separate set of testing data that has not been seen during the training. The curve generally follows the training accuracy curve, but it may not increase as quickly or may plateau earlier. When the validation accuracy curve starts to decrease or diverge from the training accuracy curve, it indicates that the model is overfitting the training data, and is not generalizing well to new data.

The training loss curve shows how well the model is minimizing the training loss function over time. The curve generally decreases as the model gets better at fitting the training data. On the other hand, the validation loss curve shows how well the model is minimizing the loss function on a separate set of testing data that it has not seen during training. The curve generally decreases but does not decrease as quickly as the training loss curve or may plateau earlier. When the validation loss curve starts to increase or diverge from the training loss curve, it indicates that the model is overfitting the training data and is not generalizing well to new data, so the training process should stop immediately.

The accuracy and loss curves provide insights into how well the model is learning from the data, and whether it is generalizing well to new data. By monitoring these curves during the training process, the performance of the model can be improved

In the experiments, the dataset is divided into a training set and a validation set by a ratio of 80:20 respectively, the validation set serves as an independent dataset that allows for evaluating the network's performance, optimizing hyper-parameters, preventing overfitting, and making informed decisions during the training process. By monitoring the validation performance, the training process can stop at the point where the validation error is minimized, preventing overfitting and improving generalization. During the experiments' training processes, an early stopping condition was put in to monitor the training and validation loss, when the validation loss increases or diverges from the training loss, it indicates that the model is overfitting the training data and is not generalizing well to new data, so it will stop the training process immediately. This appeared in Figures 5(b), 6(b), and 7(b), which display the training and validation loss curves in the selected models.

The evaluation metrics used to compare the performance of the binary classification models for the proposed eight techniques with another state-of-the-art technique are shown in TABLE 5.

TABLE 5. Evaluation metrics of the proposed eight techniques with state-of-the-art technique.

	Total params	Accuracy	F1 Score	Precision	Recall	AUC
Reference [17]	28,577,474	92.23%	0.91	85.00%	91.00%	0.92
VGG16	15,502,914	93.83%	0.94	93.93%	93.71%	0.94
VGG19	20,812,610	94.77%	0.95	94.81%	94.73%	0.95
Xception	23,222,570	92.88%	0.93	92.92%	92.96%	0.93
DenseNet 121	8,350,018	94.14%	0.94	94.03%	94.10%	0.94
MobileNet	4,541,378	94.69%	0.94	94.21%	94.74%	0.95
ResNet50	25,948,802	94.61%	0.95	94.50%	94.69%	0.95
ResNet101	45,019,266	93.60%	0.94	93.36%	94.00%	0.94
ResNet152	60,692,738	93.43%	0.93	93.49%	93.12%	0.93

For all evaluation matrix, all deep pre-trained models achieve high values compared to the [17] which indicate that using deep pre-trained models improve the overall performance of the model.

Figure 8 shows the accuracy rate [17] of all eight techniques that relied on pre-trained models compared with the state-of-the-art accuracy in that employed CNN only, [25] that employed the pre-trained model InceptionV3, and [26] that employed pre-trained models VGG16 and VGG19. Generally, as shown in Figure 8, all the techniques that used pre-trained models recorded better detection accuracy rates than the accuracy given in [17] which used the CNN network only, regardless of [26] which has a lower accuracy rate. The proposed model with VGG19 recorded the best detection accuracy rate since it is classified as one of the best pre-trained models in image classification problems. It is worth to be noted that MobileNet and ResNet50 come next in order respectively.

It was shown in [28] and [29], that VGG19 and ResNet are known for their ability to learn rich hierarchical features that are robust to image transformations, such as scale and rotation, which can be useful in detecting different types of image manipulations. Moreover, MobileNet [32] is designed to be lightweight and efficient, making it suitable for mobile devices and embedded systems, while still achieving good accuracy on various computer vision tasks. In addition,

these models often include advanced techniques such as batch normalization, residual connections, and depth-wise convolutions, which can help to improve their performance on image classification tasks.

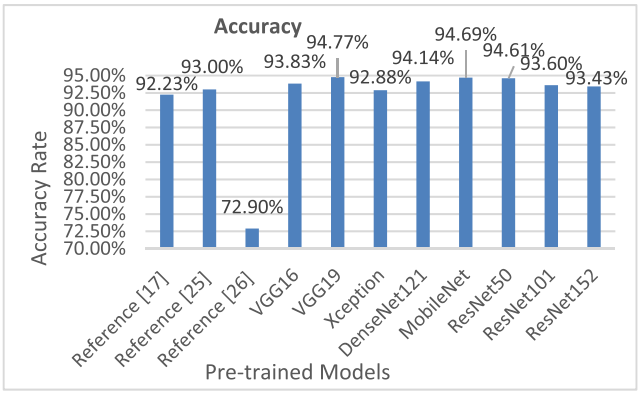


FIGURE 8. Accuracy chart for the eight experiments.

Overall, the combination of factors, such as deep architecture, large dataset training, robustness to image transformations, and advanced techniques, make MobileNet, VGG19, and ResNet well-suited for detecting image forgeries and achieving high detection accuracy rates.

The F1 score has the same meaning of precision and recall, commonly used to evaluate the performance of a classification process and how well the model can clearly classify the two classes. F1 score, precision, and recall should have a higher score. As shown in Figures 9,10 and 11, it is worth to be noticed that the models VGG19, reference [9], ResNet50, and MobileNetV2 achieved the best results in F1 score, and the models VGG19, ResNet50, and MobileNetV2 achieved the best results in precision and recall, respectively.

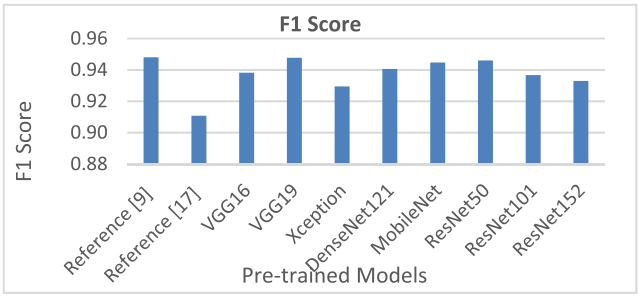


FIGURE 9. F1 Score Chart for the eight experiments.

Area Under Curve, (AUC) is another performance measure for a binary classification model. A higher AUC value means that the model can differentiate between positive and negative classes accurately with less error. In other words, the model can correctly identify true positives and true negatives, while minimizing false positives and false negatives. From the experimental results, shown in Figure 12, VGG19, MobileNet, and ResNet have the highest AUC value,

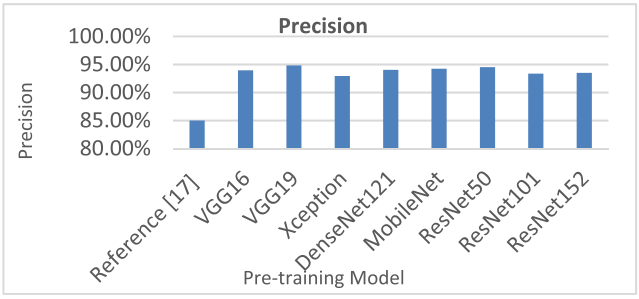


FIGURE 10. Precision chart for the eight experiments.

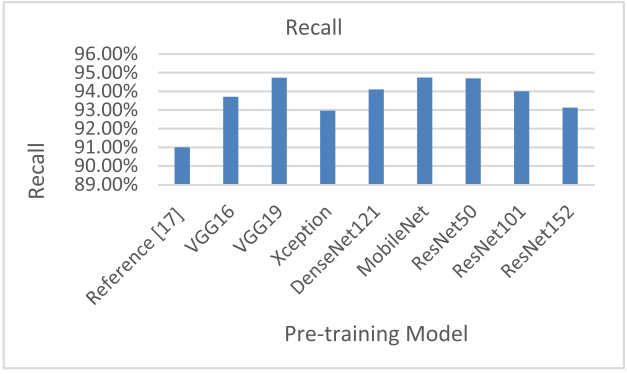


FIGURE 11. Recall chart for the eight experiments.

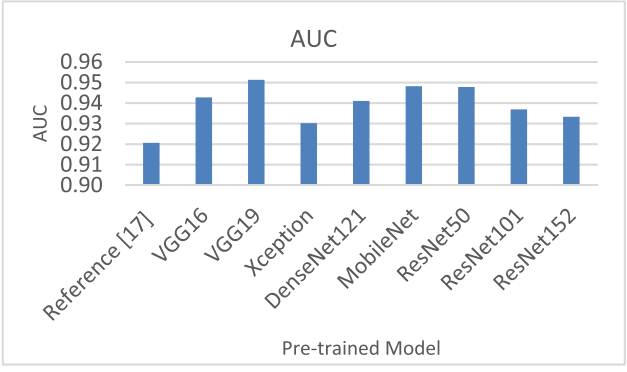


FIGURE 12. AUC chart for the eight experiments.

respectively, which means that these three models possess a better ability to correctly classify the images.

As known, a model with a smaller number of parameters can lead to faster training, lower computational costs, and reduced risk of overfitting the model to the training data. Complex models with many parameters relative to the available data can lead to overfitting, this is because of memorizing the training data instead of generalizing to new data.

From the experimental results and Figure 13, MobileNet, DenseNet121, and VGG16 have the lowest number of parameters that lead to faster training time, and lower computational costs. However, as mentioned earlier, a model with too many parameters can lead to overfitting. Therefore, it is important to achieve a balance between model complexity and generalization.

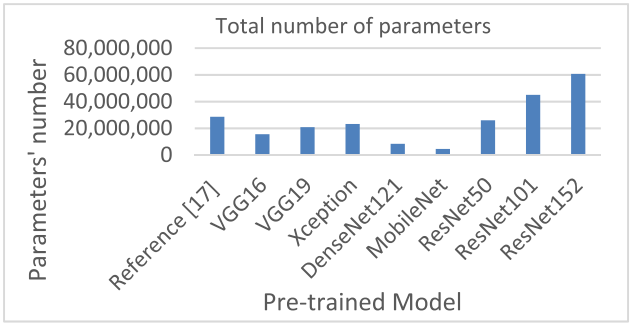


FIGURE 13. Total number of model's parameter Chart for the eight experiments.

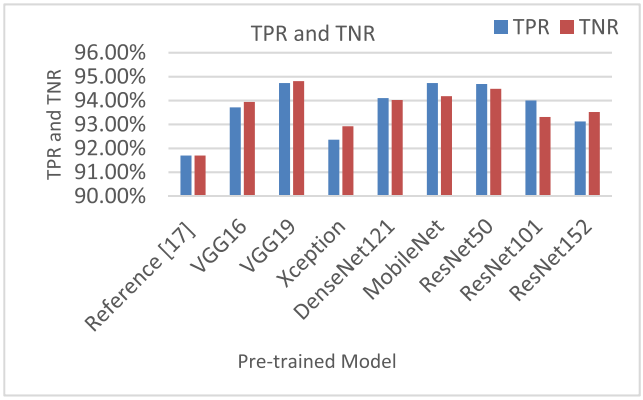


FIGURE 14. TPR and TNR Chart for the eight experiments.

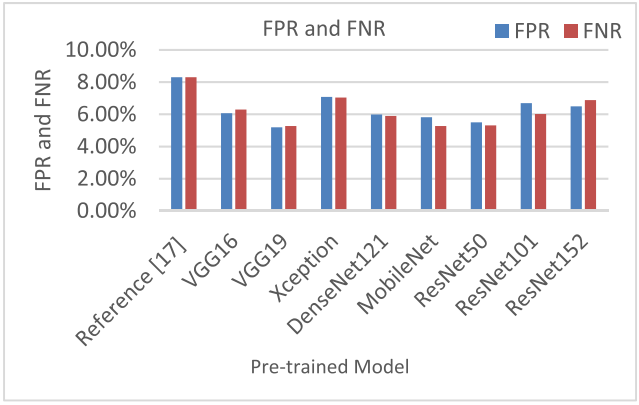


FIGURE 15. FPR and FNR Chart for the eight experiments.

As shown in Figures 14 & 15, MobileNet, ResNet50, and VGG19 have the highest TPR and TNR, indicating that these models can correctly detect the actual positive and negative cases that are identified as positive and negative, respectively.

As mentioned before, when considering the resulting evaluation metrics, charts, and graphs, the highest detection accuracy rates for the three models, VGG19, MobileNet, and ResNet50 are 94.77%, 94.69%, and 94.6%, respectively. Comparable results for the three models are also recorded in F1 score, recall, precision AUC, and TPR, which makes them the best choice in any image forgery detection system. In addition, weighing the number of network parameters as a measure of the system computational cost, MobileNet, DenesNet, and VGG16 have the minimum values. When compromising between the computational costs of the network and its accuracy, we found that the MobileNet pre-trained model is the best choice for an image forgery detection system that satisfies maximum detection accuracy rate and minimum computational costs and training time.

V. CONCLUSION AND FUTURE WORK

Image forgery detection techniques have become essential with the increased availability of image editing tools that can create forged. The paper presented an image forgery detection technique based on deep learning via a pre-trained model and transfer learning. The proposed technique considered the difference between an image and its compressed version to produce a featured image as an input to a pre-trained model that improved the detection accuracy rate. The technique with a given data set was applied to eight different pre-trained models adapted for binary classification. The recorded experimental results were compared with the state-of-the-art method. The results showed that using pre-trained models help achieve a higher detection accuracy rate than the state of the art which used CNN model.

Moreover, comparing the resulting evaluation metrics, charts, and graphs, for the eight pre-trained models, it was found that MobileNetV2 had the highest detection accuracy rate (around 95%) with a smaller number of training parameters which led to faster training, and lower computational costs, and lower system complexity and low memory consumption. So, it is highly recommended as a backbone with the image compression technique that effectively detects image splicing and copy-move at the same time with highly encouraging results.

Although the proposed model recorded high performance, compared to the performance of other studies, still there exist a set of limitations such as data generalization since the model performed well with training data and failed to generalize with unseen data. Forgery types generalization that the model was unable to detect all image forgery techniques including the novel forgery techniques. Localization is another issue since the model did not consider localizing the forged parts. Computational complexity and resource requirements for deep neural networks can be computationally intensive that require significant computational resources and time for training and inference.

In the future, an enhancement to the proposed technique can be added to increase the detection accuracy rate, keeping in mind the minimization of the training time and computational cost. Additionally, image forgery type detection, splicing, or copy move, can be extended with

localization. The combination of the proposed approach with other known image localization techniques will improve the accuracy, but it may increase the time complexity so it will need more improvement. The detection of forged videos that may be created by merging several videos is an incredibly challenging task.

REFERENCES

[1] K. D. Kadam, S. Ahirrao, and K. Kotecha, "Multiple image splicing dataset (MISD): A dataset for multiple splicing," *Data*, vol. 6, no. 10, p. 102, Sep. 2021.

[2] R. Agarwal, O. P. Verma, A. Saini, A. Shaw, and A. R. Patel, "The advent of deep learning-based," in *Innovative Data Communication Technologies and Application*. Singapore: Springer, 2021.

[3] M. A. Elaskily, M. H. Alkinani, A. Sedik, and M. M. Dessouky, "Deep learning based algorithm (ConvLSTM) for copy move forgery detection," *J. Intell. Fuzzy Syst.*, vol. 40, no. 3, pp. 4385–4405, Mar. 2021.

[4] A. Mohassin and K. Farida, "Digital image forgery detection approaches: A review," in *Applications of Artificial Intelligence in Engineering*. Singapore: Springer, 2021.

[5] K. B. Meena and V. Tyagi, *Image Splicing Forgery Detection Techniques: A Review*. Cham, Switzerland: Springer, 2021.

[6] S. Gupta, N. Mohan, and P. Kaushal, "Passive image forensics using universal techniques: A review," *Artif. Intell. Rev.*, vol. 55, no. 3, pp. 1629–1679, Jul. 2021.

[7] W. H. Khoh, Y. H. Pang, A. B. J. Teoh, and S. Y. Ooi, "In-air hand gesture signature using transfer learning and its forgery attack," *Appl. Soft Comput.*, vol. 113, Dec. 2021, Art. no. 108033.

[8] Abhishek and N. Jindal, "Copy move and splicing forgery detection using deep convolution neural network, and semantic segmentation," *Multimedia Tools Appl.*, vol. 80, no. 3, pp. 3571–3599, Jan. 2021.

[9] M. M. Qureshi and M. G. Qureshi, *Image Forgery Detection & Localization Using Regularized U-Net*. Singapore: Springer, 2021.

[10] Y. Rao, J. Ni, and H. Zhao, "Deep learning local descriptor for image splicing detection and localization," *IEEE Access*, vol. 8, pp. 25611–25625, 2020.

[11] K. M. Hosny, A. M. Mortda, N. A. Lashin, and M. M. Fouda, "A new method to detect splicing image forgery using convolutional neural network," *Appl. Sci.*, vol. 13, no. 3, p. 1272, Jan. 2023.

[12] F. Li, Z. Pei, W. Wei, J. Li, and C. Qin, "Image forgery detection using tamper-guided dual self-attention network with multiresolution hybrid feature," *Secur. Commun. Netw.*, vol. 2022, pp. 1–13, Oct. 2022.

[13] C. Haipeng, C. Chang, S. Zenan, and L. Yingda, "Hybrid features and semantic reinforcement network for image," *Multimedia Syst.*, vol. 28, no. 2, pp. 363–374, 2021.

[14] Q. Li, C. Wang, X. Zhou, and Z. Qin, "Image copy-move forgery detection and localization based on super-BPD segmentation and DCNN," *Sci. Rep.*, vol. 12, no. 1, Sep. 2022, Art. no. 14987.

[15] A. K. Jaiswal and R. Srivastava, "Detection of copy-move forgery in digital image using multi-scale, multi-stage deep learning model," *Neural Process. Lett.*, vol. 54, no. 1, pp. 75–100, Aug. 2021.

[16] S. Koul, M. Kumar, S. S. Khurana, F. Mushtaq, and K. Kumar, "An efficient approach for copy-move image forgery detection using convolution neural network," *Multimedia Tools Appl.*, vol. 81, no. 8, pp. 11259–11277, Mar. 2022.

[17] S. S. Ali, I. I. Ganapathi, N.-S. Vu, S. D. Ali, N. Saxena, and N. Werghi, "Image forgery detection using deep learning by recompressing images," *Electronics*, vol. 11, no. 3, p. 403, Jan. 2022.

[18] K. Kadam, S. Ahirrao, K. Kotecha, and S. Sahu, "Detection and localization of multiple image splicing using MobileNet v1," *IEEE Access*, vol. 9, pp. 162499–162519, 2021.

[19] E. U. H. Qazi, T. Zia, and A. Almorjan, "Deep learning-based digital image forgery detection system," *Appl. Sci.*, vol. 12, no. 6, p. 2851, Mar. 2022.

[20] A.-R. Gu, J.-H. Nam, and S.-C. Lee, "FBI-Net: Frequency-based image forgery localization via multitask learning with self-attention," *IEEE Access*, vol. 10, pp. 62751–62762, 2022.

[21] K. D. Kadam, S. Ahirrao, and K. Kotecha, "Efficient approach towards detection and identification of copy move and image splicing forgeries using mask R-CNN with MobileNet v1," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–21, Jan. 2022.

[22] M. N. Abbas, M. S. Ansari, M. N. Asghar, N. Kanwal, T. O'Neill, and B. Lee, "Lightweight deep learning model for detection of copy-move image forgery with post-processed attacks," in *Proc. IEEE 19th World Symp. Appl. Mach. Intell. Informat. (SAMI)*, Jan. 2021, pp. 125–130.

[23] C. D. P. Kumar and S. S. Sundaram, "Metaheuristics with optimal deep transfer learning based copy-move forgery detection technique," *Intell. Autom. Soft Comput.*, vol. 35, no. 1, pp. 881–899, 2023.

[24] N. Krishnaraj, B. Sivakumar, R. Kuppusamy, Y. Teekaraman, and A. R. Thelkar, "Design of automated deep learning-based fusion model for copy-move image forgery detection," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–13, Jan. 2022.

[25] S. Jabeen, U. G. Khan, R. Iqbal, M. Mukherjee, and J. Lloret, "A deep multimodal system for provenance filtering with universal forgery detection and localization," *Multimedia Tools Appl.*, vol. 80, no. 11, pp. 17025–17044, May 2021.

[26] D. Mallick, M. Shaikh, A. Gulhane, and T. Maktum, "Copy move and splicing image forgery detection using CNN," in *Proc. ITM Web Conf.*, vol. 44, 2022, Art. no. 03052.

[27] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[28] T.-H. Nguyen, T.-N. Nguyen, and B.-V. Ngo, "A VGG-19 model with transfer learning and image segmentation for classification of tomato leaf disease," *AgriEngineering*, vol. 4, no. 4, pp. 871–887, Oct. 2022.

[29] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[30] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.

[31] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1800–1807.

[32] R. Indraswaria, R. Rokhanab, and W. Herulambang, "Melanoma image classification based on MobileNetV2 network," in *Proc. 6th Inf. Syst. Int. Conf. (ISICO)*, 2022, pp. 198–207.

[33] 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.

ASHGAN H. KHALIL received the B.S. degree in computer science from the Modern Academy, Cairo, Egypt, in 2000, the Deplume Program in management and the M.S. degree in information systems from the Sadat Academy, Cairo, in 2004 and 2011, respectively, and the M.S. degree in computer science from the Arab Academy for Science, Technology and Maritime Transport, Cairo, in 2015. She is currently pursuing the Ph.D. degree in computer science with Helwan University, Cairo.

From 2000 to 2011, she was a Teaching Assistant with the Computer Science Department, Modern Academy. From 2012 to 2019, she was an Assistant Lecturer with the Computer Science Department, University of Wales, Validated Schemes, Faculty of Computer Science and Artificial Intelligence, and the Modern University for Technology and Information (MTI), Cairo. Her research interests include image processing and security techniques using deep learning.

ATEF Z. GHALWASH received the Ph.D. degree from the Faculty of Engineering, Computer Engineering Department, University of Maryland, USA, in 1988. He is currently a Professor with the Faculty of Computers and Artificial Intelligence, Helwan University, Egypt. He is involved in scientific research in the field of computer science and his specialty includes data and network security, artificial intelligence algorithms, machine learning, image processing, and software engineering. He has published more than 100 scientific papers in high quality journals and international conferences. He was the Head of the Scientific Promotion Committee, High Ministry of Education, in the computer and information systems field, from 2016 to 2022.

HALA ABDEL-GALIL ELSAYED is currently a Professor of artificial intelligence and the Head of the Computer Science Department, Faculty of Computers and Artificial Intelligence, Helwan University.

GOUDA I. SALAMA received the B.Eng. and M.Eng. degrees from MTC, Cairo, Egypt, in 1988 and 1994, respectively, and the Ph.D. degree in electrical and computer engineering from Virginia Tech University, USA, in 1999. He is currently a Faculty Member with the Department of Computer Engineering, MTC. His research interests include image and video processing, pattern recognition, and information security.

HAITHAM A. GHALWASH received the Ph.D. degree in computer science and engineering from the University of Connecticut, USA, in 2020. He was an Assistant Professor of residence with the University of Connecticut, in 2020, where he taught computer science courses until 2022. He is currently the Acting Course Director of the Ethical Hacking and Cybersecurity Program, Coventry University–Egypt Branch. His research interests include software-defined networks, applied cryptography, PKI systems, A.I, machine learning, and network security.