An Efficient Deep Learning System for Improved Trust in Medical Imaging through Forgery Detection and Localization

Divya Mohan
Department of Computer Science & Engineering,
Albertian Institute Of Science & Technology,India
Email: divyamohan@aisat.ac.in

Aleena Carolin
Department of Computer Science & Engineering,
Albertian Institute Of Science & Technology,India
Email: aleenacarolin@aisat.ac.in

Anaya Babu
Department of Computer Science & Engineering,
Albertian Institute Of Science & Technology,India
Email: anayababu@aisat.ac.in

Diya P.S
Department of Computer Science & Engineering,
Albertian Institute Of Science & Technology,India
Email: diyashaimon@gmail.com

Natasha .K. George
Department of Computer Science & Engineering,
Albertian Institute Of Science & Technology,India
Email: natashageorge@aisat.ac.in

***Abstract*—In recent times, the rates of cybercrimes have been surging prodigiously. It has been proven incredibly easy to create fake documents with powerful photo editing softwares being as pervasive as ever. Documents can be scanned and forged within minutes with the help of these softwares that has tools readily available just to do that. While photo manipulation software is handy and ubiquitous, there are also means to deftly investigate these morphed documents. This project lays a foundation on investigation of digitally manipulated medical documents and provides a solution to distinguish original documents from a digitally morphed document. The purpose of the medical image forgery detection system is to verify that images related to healthcare are not changed or altered. This image forgery detection method finds the fraud medical images using the Convolutional neural network. This method is based on deep learning technique, which utilizes a convolutional neural network (CNN). Convolutional neural networks are implemented using Keras. The proposed network architecture takes image patches as input and obtains classification results for a patch: original or forgery. The result is expected to contain the percentage of forged image and area where the image is forged.**

Keywords—CNN, ELA, Forgery Detection

I. INTRODUCTION

Since the invention of photography, individuals and organizations have often sought ways to manipulate and modify images in order to deceive the viewer. Whilst originally a fairly difficult task requiring many hours of work by a professional technician, with the advent of digital photography it is now possible and fairly trivial for anyone to easily modify images, and even easier to achieve professional looking results. This has resulted in wide reaching social issues, ranging from the reliability of the images reported by the media to the doctrine of photographs of models in order to improve their looks or body image. With the sheer amount of methods available in which to manipulate an image, image forgery detection has become a growing area of research in both academia and the professional world alike. Many methods exist in order to detect forgery within digital images, however it is difficult to find which are the most efficient and practical to implement and run. Whilst one algorithm may have a good detection rate, it could also have a large rate of false positives. In addition, runtime is a major factor that contributes to the efficiency and overall usability of an algorithm, but tends to only be mentioned academically as opposed to in real world terms.

One of the most pressing issues is that there are many different ways of modifying an image, and due to a digital images' complex nature it's impossible to have an algorithm that detects every type of image forgery. Because of this, image forgery detection isn't widely used in the professional world. The underlying concept would be highly useful in the majority of professional fields that deal with images on a day to day basis, where the reliability and credibility of these images is crucial. In addition, with the large increase in the use of social media, individuals would also benefit greatly from being able to detect forgeries within images. Convincingly manipulated images are widely circulated on social media platforms [17], and are able to be spread rapidly within communities who believe them to be true. In order to detect these image forgeries, it is required that we understand some typical methods used in order to manipulate images. These include:

- Copy-paste Cloning - This is where existing areas within an image are cloned, allowing regions to be covered or objects to be duplicated. This is a commonly used method as the forgeries have the potential to look very convincing, due to the fact that they have come from the source image to begin with.

- Image Splicing - Whereby objects from another image are spliced together with the source image, adding objects that weren't present in the original image. Various blending techniques exist, such as blurring edges, reducing the contrast and utilizing cloning to help disguise the new object in with the surrounding area.
- Modification of existing regions - This is similar to copy-paste cloning,but instead of being an exact duplication, existing regions are modified inorder to suit the needs of the forgery. This can include simply resizing the object, mirroring or skewing it, or splicing two existing objects together.In all of these cases however, the duplicated region has been resampled, meaning that it has been modified enough not to be recognised by any clone detection algorithm.

II. Literature Survey

A. A Deep Learning Architecture for Classifying Medical Images of Anatomy Object

In this paper [1], a novel image forgery detection approach is proposed which utilizes a convolutional neural network (CNN) to automatically learn hierarchical representations from the input RGB color images. The proposed CNN is specifically designed for image splicing and copy-move detection applications. It consists of 8 convolutional layers, 2 pooling layers and a fully-connected layer with a 2-way softmax classifier. The input volumes of the CNN are patches of size 128×128×3. The first and second convolutional layers have 30 kernels with a receptive field of 5×5 while other layers all have 16 kernels of size 3×3. For activation function, Rectified Linear Units (ReLU) is applied to neurons to make them selectively respond to useful signals in the input. Both the second and forth convolutional layers are followed by a non-overlapping max-pooling with filter of size 2×2, which resizes the input spatially and discards 75% of the activations. To improve the generalization, local response normalization is applied to the feature maps before the pooling layer where the central value in each neighborhood is normalized by the surrounding pixel values. Finally, the extracted 400-D features (5×5×16) are passed to the fully-connected layer with 2-way softmax classifier through "dropout" which sets to zero the neurons in the fully-connected layer with probability of 0.5. The primary contributions are summarized as follows: (1) First a supervised CNN is trained to learn the hierarchical features of tampering operations (splicing and copy-move) with labeled patches (p×p) from the training images. The first convolutional layer of the CNN serves as the pre-processing module to efficiently suppress the effect of image contents. Instead of the random strategy, the kernel weights of the first layer are initialized with the 30 basic high-pass filters used in calculation of residual maps in spatial rich model (SRM), which helps to improve the generalization ability and accelerate the convergence of the network. (2) Then the features for an image are extracted with the pre-trained CNN on the basis of p×p patch by applying a patch-sized sliding-window to scan the whole image. The generated image representation is then condensed by a simple feature fusion technique, i.e. regional pooling, to obtain the final discriminative feature. (3) Finally, a SVM classifier is trained based on the resulting feature representation for binary classification (authentic/forged). The experimental results on several public datasets demonstrate that the proposed scheme can outperform some state-of-the-art methods

B. Discriminating Original Region from Duplicated One in Copy-Move Forgery

Copy-move forgery detection methods are mostly based on finding similar regions by providing a binary mask as their output in which each pixel is identified as either background or copy-move pixels [2]. Since the original and forged region are parts of the same image, detecting the duplicated snippet is a challenging task. In this paper, a method for discriminating the duplicated region from the original one is presented. This method employs texture information of the border regions of detected copy-move regions. The image texture describes the local arrangement of color and intensities. Local texture consistency might be damaged after any manipulation performed to mask the trace of forgeries. As a result, texture analysis can be exploited to discover local inconsistency. Local binary patterns (LBP) is a kind of visual descriptor and one of texture analysis methods which generate proper features for texture classification. Since LBP extracts statistical and structural features of the textures, they are considered as a powerful tool for texture analysis. In LBP, pixel brightness (intensity) is compared with the neighboring pixels brightness. Neighboring pixels can be selected in different radiuses and get a value zero or one based on differences with the central pixel. The values of neighboring pixels are converted from binary into decimal. In order to discriminate against the forged patches, LBP is applied to the grayscale image. Since the forged regions are usually modified by a low pass filter in order to disappear its borders with the background, it is expected that the LBP histogram of duplicated regions will be smoother. By calculating the standard deviation of the LBP histogram, it is possible to detect the copied patches.

The proposed method has been validated using the GRIP dataset. The presented method can detect the forged regions with accuracy of 67.5%.

C. Forgery detection in medical images with distinguished

recognition of original and tampered regions using density based clustering technique

This paper [3] proposes a passive keypoint-based approach for forgery detection in medical images. They applied boundary extraction followed by Laplacian blob detection to find the regions of the image having similar properties. For keypoint extraction from the image, they applied the Good Features To Track (GFTT) technique. To compute descriptors for extracted keypoints, BinBoost technique is used. For identification of similar descriptors, Hamming distance-based nearest neighbor search technique is utilized. Clustering over keypoints with similar descriptors is performed using Ant Colony Density-based Clustering (ACDC). Further, Fast Sample Consensus (FSC) technique is applied for selection of correct matches and removal of imprecise keypoints. Correlation map generation is utilized for localization of forged regions within an image. After detection of the forged region, the rectangular area is selected around the copy and moved regions with extended pixels in the surrounding area. The selected region is divided into blocks. Feature descriptor for each block is computed using GLCM and Haralick texture features. Further, Pearson product-moment correlation coefficient is computed for feature descriptors. The average value for correlation coefficient corresponding to the localized regions is computed. The region having high correlation value is distinguished as the original region of image while the other region is considered as a duplicated or cloned region. To analyze the performance of the proposed technique, images are collected from various medical image repositories such as NIH, TCIA, NAMIC, SICAS, etc. Medical images belonging to different modalities such as CT scan, Digital X-rays, Ultrasound, MRI, and PET are utilized for experimentation. This technique is able to detect forged medical images even when they are distorted using several post-processing and geometrical attacks. Proposed scheme has achieved improved forgery detection results as compared to state-of-the-art techniques. In addition, the proposed technique can also distinguish between original and tampered regions present within forged medical images using Haralick texture features and Pearson product-moment correlation coefficient computation. This technique has achieved better results while distinguishing between authentic and forged regions of image as compared to state-of-the-art methods.

D. *Image forgery detection using Deep Neural Network*

In this paper [4], a unique image forgery detection system based on neural networks and deep learning, emphasizing the CNN architecture approach is provided. To achieve satisfactory results, the suggested method uses a CNN architecture that incorporates variations in image compression. The difference between the original and recompressed images is used to train the model. The proposed technique can efficiently detect image splicing and copy-move types of image forgeries. It is a lightweight CNN-based architecture designed to detect image forgery efficiently.
The system will take the forged image and then recompress it. It is compressed using JPEG compression. When an image is recompressed, if it contains a forgery, the forged portion of the image compresses differently from the remainder of the image due to the difference between the source of the original image and the source of the forged portion. Now due to the difference in the source of the forged part and the original part of the image, the forged part gets highlighted. As a result, it is used to train the CNN based model to categorize an image as a forged image or a genuine one.

● CNN model consists of 3 convolutional layers after which there is a dense fully connected layer - The first and second convolutional layers consist of 32 filters of size 3-by-3, stride size one, and "relu" activation function.
● The third convolutional layer consists of 32 filters of size 7-by-7, stride size one, and "relu" activation function, followed by max-pooling of size 2- by-2.
● It is then followed by a dense layer that has 256 neurons with "relu" activation function, finally which is connected to two neurons (output neurons) with "sigmoid" activation.
The proposed technique explores numerous artifacts left behind in the image tampering process, and it takes advantage of differences in image sources through image recompression. While most existing algorithms are designed to detect only one type of forgery, this technique can detect both image splicing and copy-move forgeries and has achieved high accuracy in image forgery detection. Compared to existing techniques, the proposed technique is fast and can detect the presence of image forgery in significantly less time. Its accuracy and speed make it suitable for real-world application as it can function well even on slower devices. The experimental results are highly encouraging and they show that the overall validation accuracy is 92.23%, with a defined iteration limit.

III. Proposed Framework

In this paper, a medical image forgery detection method was introduced. The block diagram of the proposed system is shown in *Fig 1*.

The proposed framework consists of several components to detect whether the image is forged or not,to find the percentage of forgery & locate the forged area. The algorithm used in this work involves training a model using a Convolutional Neural Network (CNN) for forgery detection. In addition, the Error Level Analysis (ELA) technique was used for localization and calculating the percentage of forgery present in the images.In the preprocessing stage the image is resized and passed to the CNN model and it decides whether the image is forged or original.If the image is forged then it displays the percentage of forgery and locates the forged area.

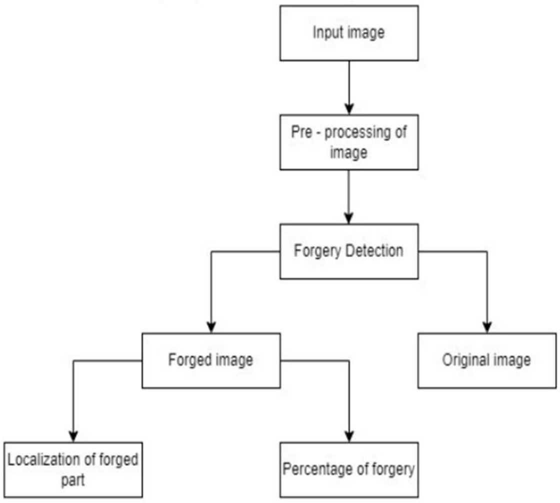


Fig.1. Block diagram of the proposed framework

IV. METHODOLOGY

There are three modules in this system, they are
(i) Forgery Detection
(ii)Percentage of forgery
(iii)Localization

A. *Forgery Detection*

Forgery detection is based on the CNN model, as shown in *Fig 2*. The model consists of 4 pairs of convolution layers followed by max pooling layer, 1 flatten layer & 3 layers of dense layers. Firstly the image is compressed and converted to ELA format and then the image is passed to model.

The convolution layers are feature mining, in which each convolution layer generates its feature maps using its own set of filters (i.e., ReLU). The feature maps produced from the first convolution layer are used in the next max-pooling layer to produce resized pooled feature maps, considered the inputs of the next convolution layer. The last feature maps merged with the final max-pooling are formatted as vectors and incorporated into Fully Connected.

The output from the last max pooling layer is then flattened into a one-dimensional vector of length 8x8x64 = 4096, which is passed through three fully connected (dense) layers with 64, 32, and 2 neurons respectively, where the final layer has a softmax activation function to produce a probability distribution over the two classes.

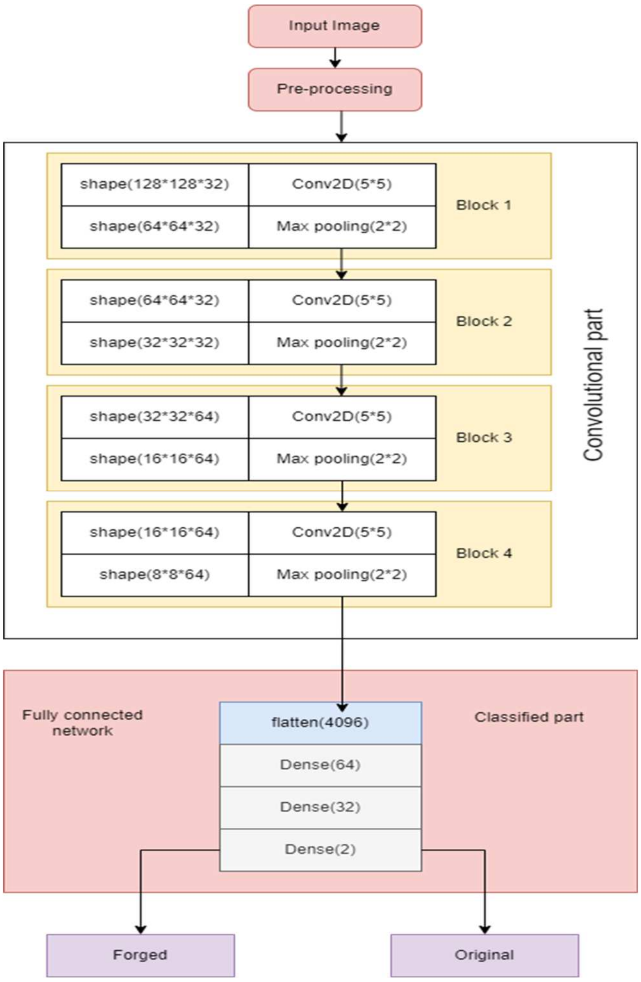


Fig. 2. Network architecture of the model

Final dense layer classifies the features extracted from the fully connected layer into two classes (original or tampered).

B. *Percentage of forgery*

Percentage of forgery is computed using Error Level Analysis (ELA) technique. Error level analysis (ELA) is the analysis of compression artifacts in digital data with lossy compression such as JPEG. To find the percentage of forgery of an input image using Error Level Analysis (ELA), you can follow these steps:

- Load the input image and apply JPEG compression to the image using a specified quality factor, and save the compressed image to a temporary file.
- Load the compressed image and the original image as numpy arrays using OpenCV's imread() function.
- Compute the ELA image by taking the absolute difference between the original and compressed images.
- Convert the ELA image to grayscale.
- Compute the mean pixel intensity of the grayscale ELA image and normalize it.

- Multiply the normalized mean pixel intensity by 100 to get the percentage of forgery.

C. Localization

Localization using error level analysis (ELA) is a method that can be used in a medical image forgery detection system to detect areas in an image that have been manipulated or edited. The basic idea behind ELA is to highlight the areas in an image where the compression level is inconsistent. ELA is a different method that compares the image with a copy of itself that has been saved at a different compression level. This process results in an image that highlights the areas of the original image that have been altered, as these areas will appear with a higher level of compression than the rest of the image. To locate the forgery in an input image using Error Level Analysis (ELA), you can follow these steps:

- The input image is opened and converted to RGB color mode and saved as a JPEG file with the specified quality level.
- The JPEG compressed input image is opened and its difference with the original input image is computed using the ImageChops module.
- The extrema (minimum and maximum pixel values) of the ELA image are obtained using the getextrema() method.
- The maximum difference between pixels in the ELA image is computed by taking the maximum of the second value in each tuple in extrema.
- A scaling factor is computed by dividing 255 (the maximum pixel value) by the maximum difference and is used to enhance the brightness of the ELA image
- The ELA image highlights the areas of the original image that have undergone compression, which can be used to localize the forged areas.

V. Result And Discussions

A. Dataset

Dataset used in the project contains a total of 600 images, in which 300 images are forged and 300 images are original. 80% of the images, that is, 480 images are used for training and 20% ,ie, 120 images are used for testing.

B. Evaluation metrics

The following evaluation metrics are used to estimate the accuracy of the proposed approach:

$$\text{Accuracy} = \frac{\text{TN+TP}}{\text{TN+FP+TP+FN}}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}}$$

$$\text{F1 score} = 2 * \frac{Precision * Recall}{Precision + Recall}$$

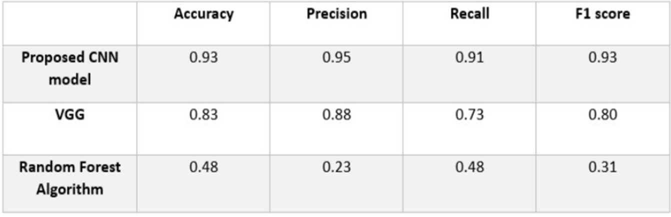
where,

- **True Positive (TP)**: number of forged images correctly identified as forged
- **False Positive (FP)**: number of authentic images incorrectly identified as forged
- **True Negative (TN)**: number of authentic images correctly identified as authentic
- **False Negative (FN)**: number of forged images incorrectly identified as authentic

C. Evaluation of performance

To evaluate the performance of the model, the proposed method is compared with other two previously published approaches, ie, VGG proposed in [26] and Random Forest algorithm proposed in [25]. Table 1 shows the comparison between the proposed method and other two methods. It shows that the proposed approach is superior to that of compared methods, with values 0.93, 0.95, 0.91 and 0.93 for accuracy, precision, recall and F1 score respectively.

TABLE 1: Proposed Method versus Previous Method



	Accuracy	Precision	Recall	F1 score
Proposed CNN model	0.93	0.95	0.91	0.93
VGG	0.83	0.88	0.73	0.80
Random Forest Algorithm	0.48	0.23	0.48	0.31

Fig.3 shows the accuracy, precision, recall and F1 score comparison of the proposed method with compared methods. In the figure, the red bar represents the proposed method. The graph indicates that the proposed model has higher value compared to other methods in terms of accuracy, precision, recall and F1 score.

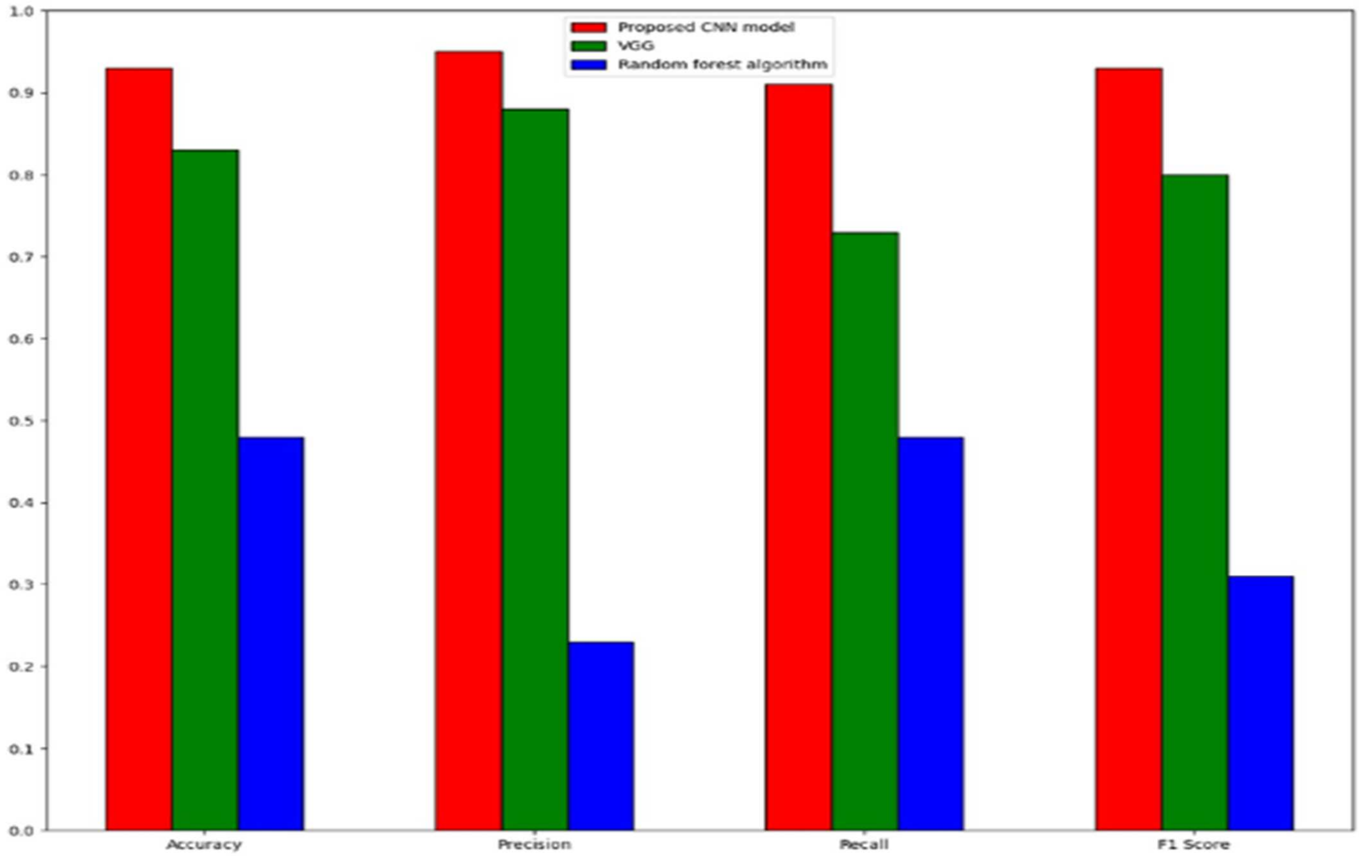


Fig.3. Accuracy, precision, recall and F1 score comparison of the proposed method with VGG and random Forest algorithm.

VI. Future Scope

In future, this system can be implemented in the hospital where the software can predict the disease and suggest the required treatment using the processed image. It can also be implemented as a mobile app.

Blockchain technology can be used to create a tamper-proof and transparent system for storing and sharing medical images, making it difficult for anyone to manipulate them without being detected.

VII. Conclusion

In conclusion, this study introduced a medical image forgery detection system based on deep neural learning. The proposed model can recognize the tampered images,classifying the input image into two types of classification: forged and original. It also locates the part of the image where it is forged and it will display the percentage of forgery. The numerical results after investigating and comparing with other approaches reveal superiority in favor of the proposed approach. The proposed method achieved 93% accuracy.

References

[1] Sameer Khan; Suet-Peng Yong, A Deep Learning Architecture for Classifying Medical Images of Anatomy Object, 2018

[2]Ahmed Ghoneim, Ghulam Muhammad, Syed Umar Amin, Brij Gupta, Medical Image Forgery Detection for Smart Healthcare,2018

[3] R. F. Olanrewaju, O. Khalifa, A. H. Hashim, A. Zeki, A. Aburas, Forgery detection in medical images using Complex Valued Neural Network (CVNN), 2018

[4] Guzin Ulutas, Arda Ustubioglu , Beste Ustubioglu , Vasif V Nabiyev , Mustafa Ulutas, Medical Image Tamper Detection Based on Passive Image Authentication, 2017

[5] Yuan Rao, Jiangqun Ni, A Deep Learning Approach to Detection of Splicing and Copy-Move Forgeries in Images, 2017

[6] S. McCloskey and M. Albright. Detecting gan-generated imagery using saturation cues. In 2019 IEEE International Conference on Image Processing (ICIP), pages 4584–4588, 2019.

[7] K. Zhang, Y. Liang, J. Zhang, Z. Wang, and X. Li. No one can escape: A general approach to detect tampered and generated images, IEEE Access, 7:129494–129503, 2019.

[8] Ning Yu, Larry S. Davis, and Mario Fritz. Attributing fake images to gans: Learning and analyzing gan fingerprints.In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pages 7556–7566, 2019.

[9] Lakshmanan Nataraj, Tajuddin Manhar Mohammed, BS Manjunath, Shivkumar Chandrasekaran, Arjuna Flenner,Jawadul H Bappy, and Amit K Roy-Chowdhury. Detecting gan generated fake images using co-occurrence matrices. Electronic Imaging, 2019(5):532–1, 2019.

[10] Joel Frank, Thorsten Eisenhofer, Lea Schönherr, Asja Fischer, Dorothea Kolossa, and Thorsten Holz. Leveraging frequency analysis for deep fake image recognition. In the International Conference on Machine Learning, pages3247–3258. PMLR, 2020.

[11] Ricard Durall, Margret Keuper, and Janis Keuper. Watch your up-convolution: Cnn based generative deep neural networks are failing to reproduce spectral distributions. In Proceedings of the IEEE/CVF Conference on ComputerVision and Pattern Recognition, pages 7890–7899, 2020.

[12] Francesco Marra, Cristiano Saltori, Giulia Boato, and Luisa Verdoliva. Incremental learning for the detection and classification of gan-generated images. In 2019 IEEE International Workshop on Information Forensics and Security (WIFS), pages 1–6. IEEE, 2019.

[13] Davide Cozzolino, Justus Thies, Andreas Rössler, Christian Riess, Matthias Nießner, and Luisa Verdoliva. Forensic transfer: Weakly-supervised domain adaptation for forgery detection. arXiv preprint arXiv:1812.02510, 2018.

[14] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In the International conference on machine learning, pages 214–223. PMLR, 2017.

[15] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation, 2017.

[16] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 4401–4410, 2019.

[17] Francisco Cruz, Nicolas Sidere, Mickael Coustaty, Vincent Poulain ¨ D'Andecy, and Jean-Marc Ogier. Local binary patterns for document forgery detection. In Document Analysis and Recognition (ICDAR), 2017 14th IAPR International Conference on, volume 1, pages 1223– 1228. IEEE, 2017.

[18] Anselmo Ferreira, Luca Bondi, Luca Baroffio, Paolo Bestagini, Jiwu Huang, Jefersson A dos Santos, Stefano Tubaro, and Anderson Rocha. Data-driven feature characterization techniques for laser printer attribution. IEEE Transactions on Information Forensics and Security, 12(8):1860–1873, 2017.

[19] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1– 9, 2015.

[20] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbig- niew Wojna. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 2818–2826, 2016.

[21] Tao Tsai, Yuadi and Yin. Source identification for printed documents. In 3rd IEEE International Conference on Collaboration and Internet Computing (CIC), pages 54–58, 2017

[22] A. Thakur, N. Jindal, Multimedia Tools and Application, Image Forensics Using Color Illumination, Block and Key Point Based Approach, (2018); 77: 26033.

[23] Anil Dada Warbhe, Rajiv V. Dharaskar, Vilas M. Thakare, ”Digital image forensics: An affine transform robust copy-paste tampering detection”, Intelligent Systems and Control (ISCO) 2016 10th International Conference on, pp. 1-5, 2016.

[24] Badal Soni, Pradip K. Das, Dalton Meitei Thounaojam. (2018) CMFD: a detailed review of block based and key feature based techniques in image copy-move forgery detection. IET Image Processing 12:2, pages 167- 178.

[25] Medical Image Forgery Detection: Rithin Krishna Dilipkumar,School of Computing, National College of Ireland, August 2022

[26] Medical Image Tampering Detection :a New Dataset And Baseline, Benjamin Reichman, Longlong Jing, Oguz Akin, Yingli Tian1, The City University of New York, New York, NY 1