FIRAT UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
TÜRKİYE

COPY-MOVE FORGERY DETECTION IN DIGITAL IMAGES USING CONVOLUTIONAL NEURAL NETWORK

Khalid Jibril SANI

Master's Thesis

DEPARTMENT OF DIGITAL FORENSIC ENGINEERING

JANUARY 2022

FIRAT UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
TÜRKİYE

Department of Digital Forensic Engineering

Master's Thesis

COPY-MOVE FORGERY DETECTION IN DIGITAL IMAGES USING CONVOLUTIONAL NEURAL NETWORK

Author
Khalid Jibril SANI

Supervisor
Dr. Öğr. Üyesi Mustafa KAYA

JANUARY 2022
ELAZIG

FIRAT UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
TÜRKİYE

Department of Digital Forensic Engineering

Master's Thesis

Title:	Copy-Move Forgery Detection in Digital Images Using Convolutional Neural Network
Author:	Khalid Jibril SANI
Submission Date:	10 January 2022
Defense Date:	28 January 2022

THESIS APPROVAL

This thesis, which was prepared according to the thesis writing rules of the Graduate School of Natural and Applied Sciences, Fırat University, was evaluated by the committee members who have signed the following signatures and was unanimously approved after the defense exam made open to the academic audience.

		Signature
Supervisor:	Dr. Öğr. Üyesi Mustafa KAYA Firat University	Approved
Chair:	Doç.Dr. Şengül DOĞAN Firat University	Approved
Member:	Dr.Öğr.Üyesi Aytuğ BOYACI Milli Savunma Üniversitesi, Hava Harp Okulu	Approved

This thesis was approved by the Administrative Board of the Graduate School on

.... / .... / 20 ....

Signature

Prof.Dr. Kürşat Esat ALYAMAÇ
Director of the Graduate School

DECLARATION

I hereby declare that I wrote this Master's Thesis titled "Copy-Move Forgery Detection in Digital Images Using Convolutional Neural Network" in consistent with the thesis writing guide of the Graduate School of Natural and Applied Sciences, Firat University. I also declare that all information in it is correct, that I acted according to scientific ethics in producing and presenting the findings, cited all the references I used, express all institutions or organizations or persons who supported the thesis financially. I have never used the data and information I provide here in order to get a degree in any way.

28 January 2022

Khalid Jibril SANI

PREFACE

The advancement in technology has given rise to the development of software or tools that make image manipulation very easy to perform. The image forgeries commonly performed is the copy-move forgery. In this type of image forgery, a segment of the image is copied and pasted to a different segment of same image. This type of image forgery is done by attackers to share fake information. Therefore, the detection of forgeries in digital images is now becoming an interesting area of research.

This work proposed a deep learning method based on a pretrained ResNet50 network model to detect copy-move forgery in digital images. Deep learning methods have been proven to perform better in the detection of forgeries in digital images.

Khalid Jibril SANI
ELAZIG, 2022

TABLE OF CONTENTS

ABSTRACT

Copy-Move Forgery Detection in Digital Images Using Convolutional Neural Network

Khalid Jibril SANI

Master's Thesis

FIRAT UNIVERSITY
Graduate School of Natural and Applied Sciences

Department of Digital Forensic Engineering

January 2022, Page: ix + 044

The detection forgery in digital images has become a hot domain of research in the field of digital image forensics as a result of the prevalent use of image editing tools for manipulating an image to conceal or distort information in the image. One of the most common image forgeries performed is the copy-move forgery. This type of forgery involves copying a segment of the image which is then pasted to a different segment of the same image. The need for detecting whether an image is authentic becomes essential. The existing methods implemented for detecting image forgeries were based on traditional feature extraction algorithms such as block-based and key point-based algorithms. These traditional techniques employed produce a low-performance result. Deep learning techniques have proven to provide better performance in image processing tasks. In this research, a convolutional neural network that is based on a pre-trained ResNet50 network was proposed for the detection of copy-move forgeries in digital images. The proposed model uses the CoMoFoD image dataset in carrying out the experiment. The metric evaluation results achieved in the proposed model show that deep learning methods performance is more effective and efficient in digital image copy-move forgery detection.

Keywords: Digital image forensic, copy-move forgery, deep learning, convolutional neural network.

ÖZET

Evrişimsel Sinir Ağları Kullanılarak Sayısal Görüntülerde Kopyalama Sahteciliğinin Algılanması

Khalid Jibril SANI

Yüksek Lisans Tezi

FIRAT ÜNİVERSİTESİ
Fen Bilimleri Enstitüsü

Adli Bilişim Mühendisliği Anabilim Dalı

Ocak 2022, Sayfa: ix + 044

Görüntüdeki bilgileri gizlemek veya çarpıtmak için bir görüntüyü manipüle etmek için görüntü düzenleme araçlarının yaygın kullanımının bir sonucu olarak, görüntü sahteciliği tespiti, dijital görüntü adli tıp alanında sıcak bir araştırma konusu haline geldi. En sık yapılan görüntü sahtekarlıklarından biri kopyala-taşı sahtekarlığıdır. Kopyala-taşı sahtekarlığı, görüntünün bir bölümünün kopyalanmasını ve aynı görüntünün başka bir bölümüne yapıştırılmasını içerir. Bir görüntünün gerçekliğinin gerekli olup olmadığını tespit etme ihtiyacı. Görüntü sahteciliğini tespit etmek için uygulanan mevcut yöntemler, blok tabanlı ve anahtar nokta tabanlı algoritmalar gibi geleneksel özellik çıkarma algoritmalarına dayanıyordu. Kullanılan bu geleneksel teknikler, düşük performanslı bir sonuç üretir. Derin öğrenme tekniklerinin görüntü işleme görevlerinde daha iyi performans sağladığı kanıtlanmıştır. Bu çalışmada, dijital görüntülerde kopyala-taşı sahtekarlığını tespit etmek için önceden eğitilmiş bir ResNet50 ağına dayalı evrişimli bir sinir ağı önerilmiştir. Önerilen model, deneyi gerçekleştirirken CoMoFoD görüntü veri setini kullanır. Önerilen modelde elde edilen metrik değerlendirme sonuçları, dijital görüntülerde kopyala-taşı sahteciliğini tespit etmede derin öğrenme yöntemleri performansının daha etkili ve verimli olduğunu göstermektedir.

Anahtar Kelimeler: Dijital görüntü adli, kopyala-tası sahtekarlığı, derin öğrenme, evrişimli sinir ağı

LIST OF FIGURES

Page

LIST OF TABLES

1. INTRODUCTION

Presently, digital cameras have become commonly used by people. These cameras are mostly found in our mobile devices due to the advancement in computer technology. A recent study by [1] reported that more than one trillion of images were captured every year using mobile phones. This vast number of images are mostly manipulated after being captured. Due to rapid advancements in computer and internet technology, the use of software for picture editing has grown in recent years. The software editing tools for images are advanced and easy to use and have made altering digital images very easy. So, this has made it challenging to identify the effect of the alteration with the human eyes. In the work of [2], they reported that 10% of coloured images are digitally modified. These forged images are being created to spread rumors, make false claims, or earn money. For instance, during the G-20 summit in 2017, Markus Schreiber, an associate press photojournalist, snapped a picture [3]. The picture captured by the photojournalist was later modified and shared on social media platforms which generated much controversy [4] . The consequences of using forge images have made researchers in multimedia forensics analysis proposed methods of detecting manipulated images.

A digital image needs to be forensically analyzed to confirm the its authenticity. Before digital images can be used as evidence in legal proceedings of law, forensic analysis must be performed to validate their authenticity or uncover evidence of tampering. [5]. The importance of detecting forged images has increased in the number of researches in digital image forensics.

Digital image forensics is an emerging research area in the field of digital forensics that involves the authentication of digital images to know its source or know whether an image is genuine or modified [6]. In the past, digital images were broadly utilized with confidence and integrity in many fields. For example, pictures published in newspapers present some visual information that is considered authentic and accepted. However, today digital images used are less trustworthy as a result of the simplicity involved in generating images and the ease of use of software tools used for tempering the content of these images. One of the easiest ways to temper images involves copying a portion of an image and pasting it to another portion of the same image. This image forgery type is called the copy-move forgery. One of the reasons why copy-move forgery is performed is to hide some information or duplicate segments of the image [7]. Therefore, digital image forensics methods are required to ensure the integrity of digital images.

There have been approaches for detecting copy-move forgeries in images that have been employed in the previously. These existing methods are referred to as the traditional method, which is based on image geometrical, statistical, and physical properties [8]. The traditional methods are classified into two categories: block-based algorithm and a key-point algorithm.

The block-based algorithm used in detecting copy-move forgery divides the image into overlapping blocks and then features are extracted from each block of the image through the use of frequency transformations, image intensity, and texture [9]. The extracted features from each block are then compared with each other using a matching operation to discover the similarity of the blocks.

There are a number transformation techniques such as Principal Component Analysis (PCA), Discrete Cosine Transform (DCT), Singular Value Decomposition (SVD), and Dyadic Wavelet Transform (DWT) that are utilized for block-based algorithms. Local Binary Patterns (LBP) and DCT, for instance, have been used to detect copy-move forgery in digital images [10]. This method used by [10] involves the digital image been divided into overlapping blocks and then transforming each block into LBP form. A feature matrix is then obtained by applying DCT on each block LBP and sequentially sorted to detect similar blocks. Also, DCT and SVD were applied on digital images to detect region of forgery in the images [11]. Images are partitioned into blocks, and DCT extract features from the image blocks. After that, the dimensionality of the extracted features was reduced using SVD. The SVD also improves the model noise resistance.

The key-point-based algorithm used in image copy-move forgery detection deals with the extraction of specific features from an image which includes: blob, corners, and edges. Key-point descriptors are used in the key-point algorithm to extract the specific features and operate on the whole image. They commonly use key-point descriptors are the Scale Invariant Feature Transform (SIFT) and the Speeded Up Robust Features (SURF) [12]. When used to identify a copy-move forgery in digital images, the SIFT and SURF key-point descriptors have been shown to produce good results.

SIFT key-point descriptors and best-bin first nearest neighbor were applied in the detection copy-move forgery in images [13] which performs well for images with a large region of the forgery but performs poorly for images with a small region of forgery. In the work of [15], copy-move forged images that are being transformed geometrically via transition, rotation, and scaling were detected using the SIFT method. The method implemented by [14] uses the Euclidean distance threshold in finding similarity in the image feature extracted after carrying out SIFT feature extraction, key-point vector construction, key-point clustering, and estimating geometric transformation. The SIFT-based key-point feature extraction approach with scale comparison was used to identify rotational copy-move forgery, which was then corrected using the orientation adjustment technique [15].

[16] proposes a method that classifies forged and original images using Support Vector Machine (SVM), wavelet decomposition, and extraction of features in images with SURF key-point descriptors.

The block-based and key-point based algorithms have their strength and weakness in identifying copy-move forgeries in images. In block-based methods, its application is simple and copied, and pasted regions are easily detected but consume a lot of computational resources. When applying a key-point-based approach in detecting copy-move fraud in images is efficient and not computationally expensive. However, the key-point based method fails in detecting smooth copied portions in a copy-move image that are manipulated.

The research discussed above focused on using manual techniques in extracting features from images. So, this means the features extracted are limited and cannot provide much information to detect a forgery in an image. For instance, extracting features using LBP only focuses on the textual aspect of the image and cannot extract details about the color of the image. This type of problem is solved using deep learning techniques.

Recently, deep learning techniques have become widely used in detecting forgeries in images. The utilization of deep learning has proven to provide better performance in detecting forgeries in images [17]. A type of deep learning that is widely used in many image processing tasks is called a convolutional neural network (CNN). Also, applying CNN in sound processing has been successful [18]. CNN can automatically extract features from the image using the convolution operation at every network layer to generate feature maps [19].

Deep convolutional neural networks were used in some of the first efforts on image forgery detection which include: the work of [20] on detection of an anti-forensic technique called median filtering in image forgery, the work of [21], which focuses on identifying the model of camera that was used in capturing a digital image, the work of [22] detects multiple types of image manipulations techniques used in image forgery and the work of [23] on image steganalysis.

[24] employs a fully convolutional network to detect spliced sections of a picture produced using image splicing forgery techniques on the CASIA v2.0 dataset. The proposed method by [24] has an F1 score and MCC score performance metric of 0.8473 and 0.8527, respectively. It has also been used to identify copy-move forgery photos and is effective at localizing the forging region.

[25] implements a convolutional neural network for detection of copy-move manipulation using images from the copy-move forgery dataset (CMFD). The network architecture designed in [25] comprises three convolutional layers, two fully connected layers, one 3x3 max pool filter, and two 3x3 average pool filters within the convolutional layers, ReLu activation function within the intermediate layers, and a SoftMax activation function at the output layer for classification. The CNN model used to identify active copy-move forgeries photos extracts features from the images and detected active copy-move forgery images.

In some cases, to make detection of the forgery more difficult, copy-move forgery of an image is sometimes subjected to additional operations such as scaling, rotation, or blurring. [26] were able to propose a deep learning technique with a matching algorithm to detect copy-move tempered images. The deep learning model implemented uses the Visual Geometry Group-Net (VGGNet) to extract multi-scale features, which help in efficiently identifying a copied region in the fraud image even if it is rotated, scaled, or blurred.

[27] proposes a deep learning technique based on a pre-trained model to detect a copy-move forgery in images. The model was trained with a small dataset and performed well when applied on computer-generated forge images but poorly on real-scenario copy-move forge images. Furthermore, a deep learning model that extracts hierarchical features from an input image was developed and tested on multiple datasets for the classification of copy-move forge images and authentic images [28].

Similarly, [29] implemented a VGG16 deep learning architecture with semantic segmentation to detect an image copy-move forgery. This VGG16 transfer-learning approach was trained and tested on multiple datasets, providing excellent performance when evaluated. The accuracy achieved was 0.9869.

A deep learning technique based on the AlexNet model was experimented for the identification of image forgery [30]. The model could extract features from the input image, learn the features, and the generated feature maps are used to classify the tempered and original image.

[31] also proposes a technique for detecting copy-move forgery that uses an AlexNet model for feature engineering of the input image. Then SVM classifier was used for training the extracted features to differentiate between the forge and the original image. MICC-F220, an available to the public dataset, was used to train and evaluate the model. The accuracy obtained from the model was 93.94%.

[32] implemented an end-to-end deep neural network based on the VGG16 convolutional neural network to detect image forgery that are of the copy-move type. The VGG16 network implemented was used for feature extraction. Then a self-correlation matrix for feature pixels is computed for similarities, and finally, a deconvolutional operation was performed, which involves a point-wise feature extraction for determining matching pairs of pixels in locating the copy-move forgery in the image.

[25] presented an improved fully convolutional network for identification of forgery in images. The authors developed an upgraded convolutional network that is based on the VGG-16 architecture. The VGG-16 is a type of deep learning architecture for VGG-Net. A completely convolutional network contains five convolutional layers, three of which are fully linked. Convolutional layers were used to replace the last three fully-connected layers of the original VGG16 network. The CASIA v2.0 dataset was used to train and evaluate the model, and the F1-score and MCC-Score evaluation metrics revealed that the model performed 0.8473 and 0.8527, respectively.

[27] proposes a deep learning technique for detecting copy-move forgeries that includes a matching process. The model first segments an input image using the Simple Linear Iterative Clustering (SLIC) algorithm [27]. The segmented patches of the image are passed as input to the VGG-Net (Visual Geometry Group Net) used as a deep learning architecture in the proposed model where multi-scale features are extracted from image patches being extracted. When the features extraction is completed using VGG-NET, the depth of each and picture element (pixel) in the patches was reconstructed using an Adaptive Patch Matching (APM) approach for comparison with other parts of the picture. A comparison of the segmented patches was performed, and tempered areas in the image a found. Lastly, the matched key-point is combined with the segmented patches of the input image, and the region of forgery is revealed and displayed as output. The proposed model achieved an accuracy of 95% when tested on the MICC-F220 dataset.

[33] proposes an optimized pre-trained AlexNet model for image forgery detection, which performs well when trained and evaluated on multiple datasets. The AlexNet architecture implemented in the proposed model was modified, making it optimized. The following changes have been made to their proposed AlexNet model: In the seven network layers, Batch Normalization (BN) and the Max-Out activation function are utilized, and a SoftMax function utilized in the network's final layer instead of the Local Response Normalization (LRN) and Rectified Linear Unit (ReLu) activation function used in the original AlexNet model.

From the literature of deep learning methods applied in digital image forensics, it can be seen that there is a great improvement in the performance of the deep learning models proposed to solve the detection of forgery in image that are copy-move in nature.

The motivation of this study is inspired from the improvement deep learning techniques have proven to provide in image processing tasks and how it can be applied in digital image forensics. Image forgery has recorded an unexpected rise as a result development of image editing tools that can be used by both expert and non-expert.

The aim of the research is to offer an effective and efficient solution for detecting copy-move forgeries in digital photographs that will outperform existing methods. The proposed method in this study implements a convolutional neural network based on a pretrained ResNet50 network. The proposed method's efficacy was compared to the performance of current traditional and deep learning approaches utilized to solve the same problem.

2. DIGITAL IMAGE FORENSICS

In the world we live today, multimedia data such image and video are playing a key role in communication. The rapid development in computer technology has increase the number of multimedia data shared on the internet through social networks. Facebook, Instagram, Twitter and YouTube are examples of social media platforms that are mostly used by people to share their opinions and ideas. Digital images shared across different social media platforms can be easily manipulated because of the reachability of sophisticated image manipulation tool like Adobe Photoshop, Paint, Pixlr, etc. The alteration of digital images by manipulation software makes it impossible to tell whether a picture is real or fake. There has been an increase in the number of forge images as a result of affordability and availability of this modern image editing tools, hence trusting the originality of image become difficult [34], [35]. Recently, images are not just used as a means of communication but serve as digital evidence presented in the court of law. So, to forensically identify whether an image is authentic or fake, a new research domain in digital forensics discipline was introduced. Digital image forensic deals with the authentication of digital images to know its source or know whether an image is genuine or modified. Digital image forensic provides scientifically strong analysis and report about digital images when use as evidence in court of laws.

2.1. What is Digital Image?

A finite number of pixel intensities represented in a two-dimensional plane using the function f (x, y) is what describes a digital. The x and y in the function denotes the coordinate on the plane while f is the intensity value resulting from the amplitude of coordinates x and y paired as (x, y) [36]. The pixel values within a digital image are finite each having specific position. The two-dimensional plane of pixel intensities and their location is the representation of an object captured using sensors found in digital cameras. The sensor performs a sampling process called discretization to generate the pixel intensities of any pair of coordinates on the two-dimensional space [37]. Every pixel intensity value correlate to the light response of the real-world object received by the image sensor.

A digital image can a be grayscale or coloured which is defined by the how many color channels the image has. The grayscale image has a single-color channel that gives a black and white visual representation of an image. The intensity values in grayscale image are represented in a single two-dimensional plane based on the signal levels from the lowest (0) to the highest. The lowest signal level is for black and the highest signal level is for black. In coloured images, there are three color channels which implies that there are 3 intensity value represented at a particular pixel location. The most common representation of coloured images uses the Red, Green, Blue (RGB) color map. The RGB color channels in color images are combined together to produce the coloured visual representation of an object. Figure 2.1 and 2.2 below depict an example of gray scale and color image with normalize pixel intensities of certain portion in the image.

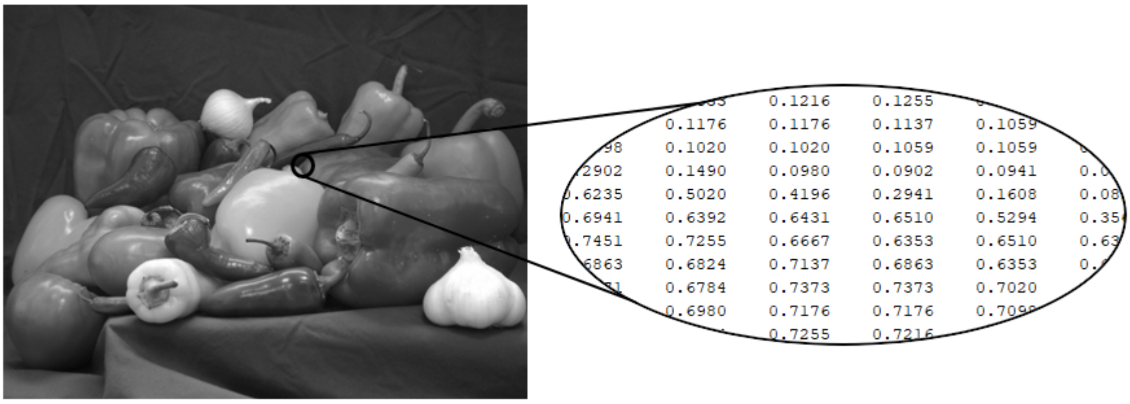

Figure 2.1. Gray Scale Image [38]

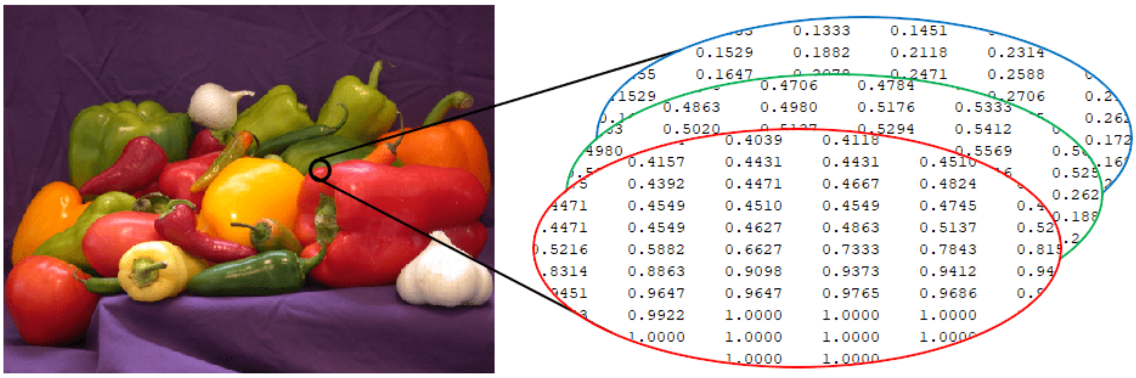

Figure 2.2. Color (RGB) Image [38]

2.1.1. Digital Image Formation

A digital image can be produced by following three major phases. The three phases involve in the formation of digital images as shown in Figure 2.3 below are: acquisition, storage and editing [39] – [41]. In the acquisition phase, a digital camera's lens captures light from a real-world object. The amount of light that passes through the lens which is controlled by the diaphragm falls on the image sensor. Complementary metal oxide semi-conductor (CMOS) or charge-coupled devices (CCD) [42] are generally used as image sensors in digital cameras. These image sensors are sensitive to light due to the presence of light sensitive diode. The light sensitive diodes in image sensors transform the light that falls on them into digital signals that corresponds to the pixel intensity of an image. Prior to receiving light on the image sensors, a thin film called color filter array (CFA) performs a filtering operation of light captured by the digital camera. The filtering done by CFA allows for the passage of only one-color channel for each pixel intensity while the other two-color channels are not permitted to pass through. The only allowed color-channel for each pixel intensity is recorded on the image sensor. Interpolation or demosaicing is a method that generates the complete color representation of every pixel in the image [43]. The CFA which has an arranged pattern of primary color (red, green and blue) as shown in Figure 2.3 below is responsible for producing a color image which constitutes different proportion of red, green, and blue color.

Now that the complete color representation of the image is formed at image sensor, this colored image is processed within the digital camera using variety of processing operations. The variety of process operations employed inside most digital cameras are: contrast enhancement, white balancing, noise reduction, image sharpening, gamma correction, and color processing.

In the storage phase of digital image formation, the processed image is compressed using an image compression algorithm and saved inside the storage location of the digital camera. Compression techniques use in digital cameras on digital image minimizes the amount of storage capacity needed to represent an image. The widely use compression algorithm in digital camera is a lossy compression algorithm called joint photography expert group (JPEG).

The final phase in digital image formation is editing. Unlike all previous phases of digital image formation, the editing phase takes place outside the camera. The compressed digital image from the storage phase can further be modified or transformed geometrically. The modification may be to enhance or reduce the visual appearance of the digital image, add or hide information to the digital image.

In all the phase of digital image formation, digital artifact called fingerprints are produced and left after performing the operations in each phase. These fingerprint traces in every phase of the digital image formation are embedded inside the fully-formed digital image. The traces in the fully-formed digital image may be examined to confirm the image's authenticity.

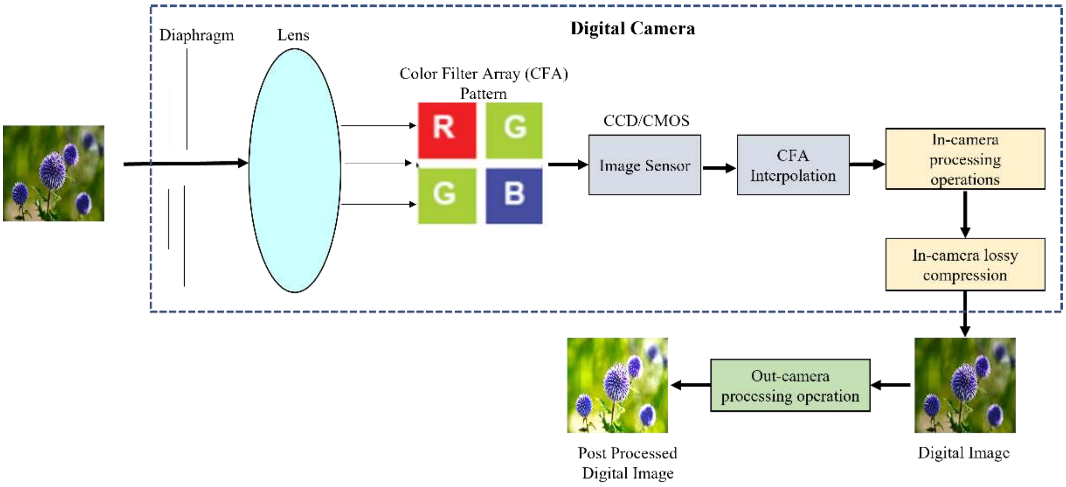

Figure 2.3. The Image Formation Process

Figure 2.3 above depict a visual description of how an image is formed using a digital camera.

2.1.2. Digital Image File Format

The file format of digital image describes how an image data is encoded and stored. The format of a digital image file defines the type of computer application that can open it. When a digital image is stored in a computer its format is known by the file extension on the filename. There are many formats for digital images that have been in existence in the past 30 years [37]. Some of widely used formats in digital images are explained below.

1. **JPEG or JPG:** This image format is an acronym for Joint Photographic Expert Group. This type of image format support different color modes of an image and can be displayed in most operating systems. Many images are saved as jpeg because of its lossy compression which makes it require less storage space. Image quality is slightly reduced when jpg format is used for an image. JPEG images are mostly used in printing, email, and power point.
2. **GIF:** This image format is an acronym for Graphic Interchange Format. This image format requires at most 256 colors for an image and low memory to save image with GIF format. GIF supports animations because it allows the combination of many image files in sequence. Digital images in GIF format are widely use on the internet due to their low file size and are supported by most applications.
3. **PNG:** This image format is an acronym for Portable Networks Graphics. PNG image format has support for millions of colors and also offer transparency. It was introduced to overcome the limitations provided by GIF. This image format is lossless and are mostly used on the web.

4. **TIFF:** This image format is an acronym for Tagged Image File Format. Images saved in this format are of high quality and require high storage space. TIFF uses a lossless compression. The TIFF images are most preferred in the production of graphics with high quality.
5. **BMP:** This image format is referred to as Bitmap Picture. It is considered as a simple image format because the pixels of the picture are represented as bits and encoded without compression. BMP images are most preferred in producing graphics or images that are realistic in nature.

2.2. Image Forgery Techniques

Image forgery detection techniques are basically classified into two major classes, which are: active and passive approach as shown in Figure 2.4 below. Active methods depend on prior information of the image, which is inserted at the point of capturing the image. The active methods include watermarking and digital signatures. In passive methods, preliminary information about the digital image is not required. The passive approach relies on the irregularities in the statistical properties of pixel intensities of the images resulting from the modification of the original image [27]. The passive approach is subdivided into three categories: copy-move, image splicing, and image retouching.

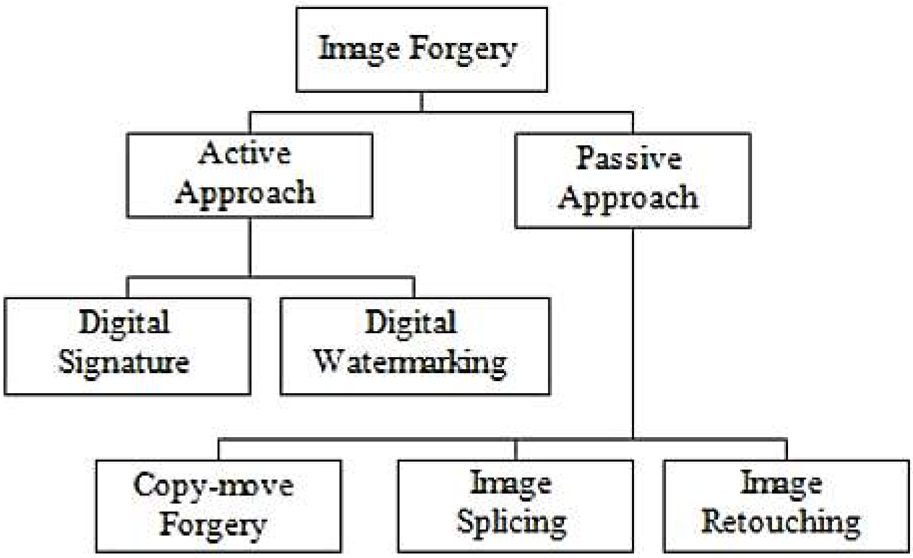

Figure 2.4 Classification of Image Forgery Techniques [44]

2.2.1. Copy-Move Forgery

As the name implies, this form of image forgery has to do with the copying a segment of an image and transferring it to another segment of the same image. This type of forgery is mainly done to hide information on the image. When a copy-move forgery operation is performed, traces of the modification are hidden by further editing the image using rescaling, filtering, and noise scattering operations [45]. The image below in Figure 2.5 a) and b) shows an example of what of copy-move forgery in image looks like when performed.

Figure 2.5. a) Original Image Before Copy-Move **b).** Copy-Move Forgery of Fig 2.5 a) [46]

2.2.2. Image Splicing Forgery

Image splicing forgery techniques involve combining at least two different images to create a forged image. This forgery technique uses the copy/cut and pastes operation. This means a part of one image is extracted and merged with another image. The resulting image after the merging operation creates a disturbance in the image pattern. Hence, forgery can be detected by analyzing the image pattern. An example of image slicing forgery is shown in Figure 2.6 below:

Figure 2.6. a) Original Image 1 Before Splicing

Figure 2.6. b) Original Image 2 Before Splicing

Figure 2.6. c). Image Splicing Forgery of Figure. 2.6 a) and Figure. 2.6 b) [27]

2.2.3. Image Retouching Forgery

Image retouching is considered a straightforward approach to modifying an image. This technique enhances or reduces specific features of the original images by adjusting their brightness, color contrast, sharpness, or size. In this method of forgery, the original image is not significantly changed. It is usually applied to make images look more attractive in magazine photo editors, but this modification may be considered unethical. An example of image retouching is shown in figure 2.7. a) and b) below.

Figure 2.7. a) Original Image Before Retouching

Figure 2.7. b) Image Retouching Forgery of Figure. 2.7 a)

Figure 2.7. a) on above shows a sample original image that is modified by applying an image retouching technique. The resulting modification is shown in Figure 2.7. b) above.

3. DEEP LEARNING

Deep learning is a subfield in machine learning based on artificial neural networks that imitate how the human brain works. One of the first artificial neural network (ANN) architectures introduced is known as the perceptron [47]. A perceptron is used to classify data that are linearly separable[48]. It has an input layer called neurons and the output layer for producing classification output results, as shown in Figure 3.1 below. In order to overcome the limitation of ANN, which is its application to non-linearly separable data, deep learning was introduced.

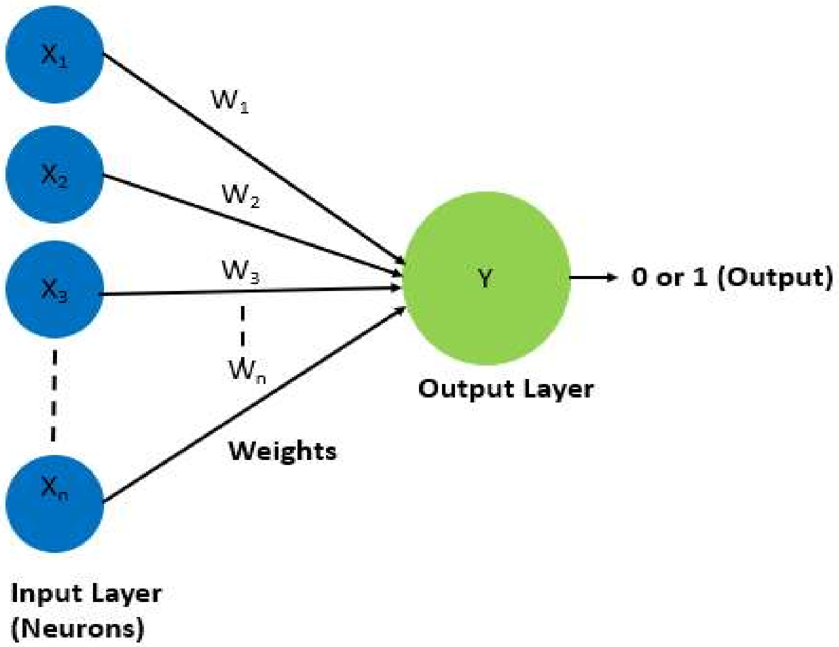

Figure 3.1. ANN Architecture: Perceptron

Figure 3.1 above shows depicts the architecture of an ANN.

Deep learning or deep neural network constitutes many layers of the network stack together [49] that requires a large amount of data to produce an effective result. It has to ability to learn from the data automatically and make predictions. Deep learning when it is implemented can solve a wide range of problems computationally such as visual recognition, natural language processing and sound recognition. Unlike ANN that only has input and output layer, a deep neural network has an arbitrary number of between the input and output layer called hidden layers. Figure 3.2 depict a deep neural network architecture with a number of neurons in its the input, hidden and output layers. When data is pass to a deep neural network it is been received as input by each neuron in the network. The input data received by each neuron is evaluated by multiplying it with a weight matrix, then summing it up with a bias. Equation 3.1 below shows how the input of each neuron is evaluated.

$$y = f\left(\sum_{i=1}^{n} w_i x_i + b\right) \tag{3.1}$$

From the equation above f denotes the activation function, w denotes the weight matrix, x denotes the input, n denotes the counts of neurons and b denotes the bias. The input of each neuron of the hidden layers are evaluated using the above equation and the result of prediction is produced at the output layer of the network. The number of neurons at the final layer is based on the type of problem the deep neural network is solving. For instance, in a binary classification problem two neurons are used at the output layer. The prediction result at the output layer is compared with the expected result to find the loss. Prediction result is then improved through backward propagation. Backward propagation aims to update the weight matrix using which will then be used to forward propagate the input values through the network. Therefore, the forward and backward propagation boost the deep neural network performance result.

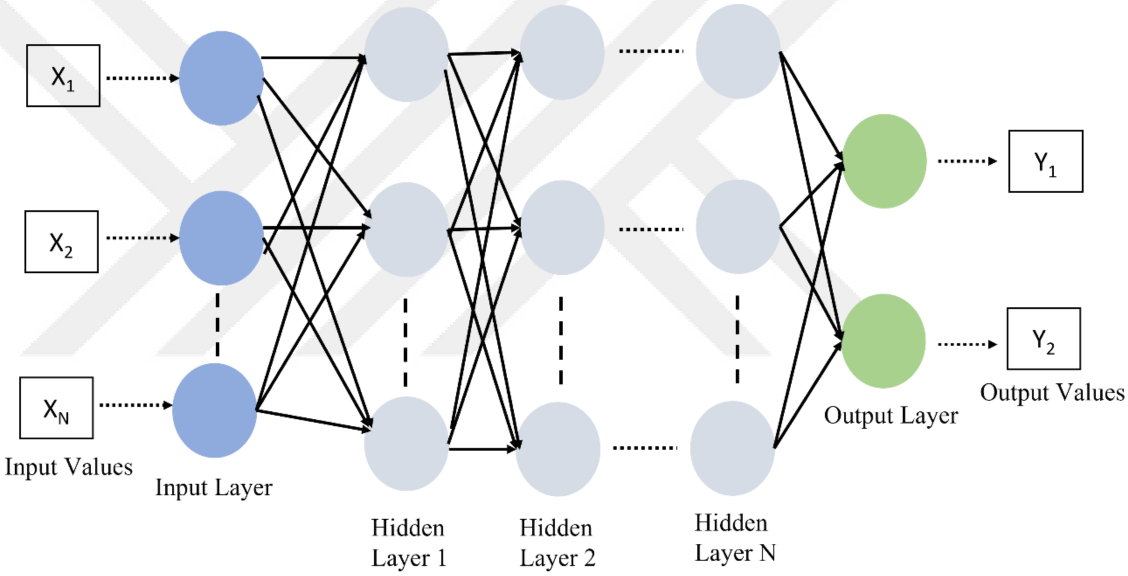

Figure 3.2. Architecture of Deep Neural Network

Recently, deep learning models have shown improved performance in solving image data problems using convolutional neural networks [50]. A detail description of the building blocks of CNN is given the following section.

3.1. Convolutional Neural Network

Convolutional neural networks are broadly employed in image processing tasks because they can learn features from images. A convolutional neural network comprises of an input and output layer, an arbitrary number of convolutional layers, and fully connected layers. CNN has the ability of extracting variety of features that are contained in an image using convolution. The features extracted are from the convolutional layers in the CNN where edges are detected from the raw pixels of the image at the first layer. Then shapes or blobs are detected in the second layer and higher-level features are detected in the deeper layers of the CNN. The ability of CNN to extract features in images has proven to be more robust than traditional feature extraction techniques because CNN are robust to translation, scaling, and rotation. The major component in building a CNN as shown in Figure 3.3 below are the convolutional layer, pooling layer and fully-connected layer.

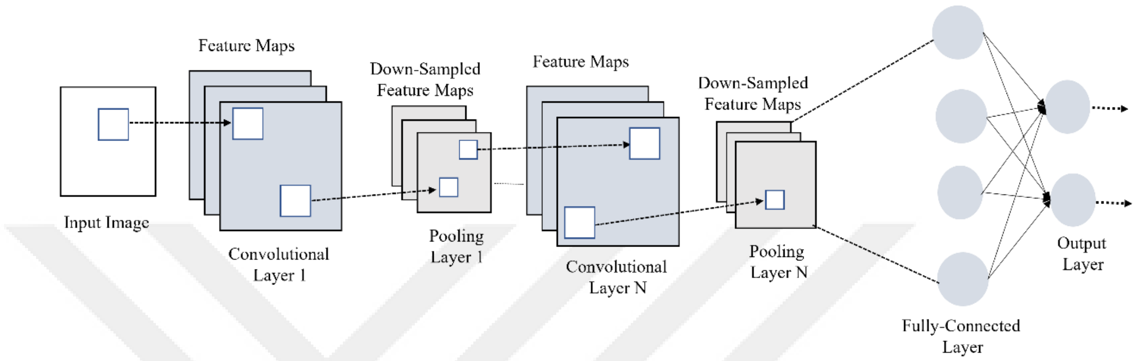

Figure 3.3. Basic Architecture of CNN [28]

3.1.1. Convolutional Layer

The convolutional layer is the layer that defines the CNN. It uses a number of kernels or filters to generate feature maps. The kernels or filters convolve over the inputs image to learn properties from the images. When the operation called convolution is performed on an input the spatial dimension is reduced. In a situation where the dimension of the input is to be preserved after convolution, padding is used. The type of padding commonly used is the zero padding where the zeros are added to the border of the input image. The convolution operation is illustrated in Figure 3.4 below showing the element-wise multiplication of pixel values with kernel values at a particular sliding window and then followed by a summation.

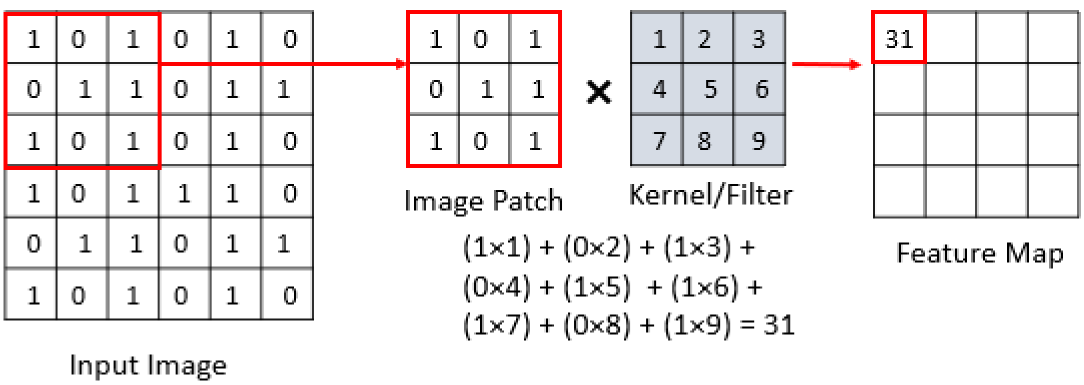

Figure 3.4. Illustration of Convolutional Operation [51]

3.1.2. Activation Function

Right after the convolutional operation is carried out at the convolutional layer, an activation function is applied. One of the advantages of activation function in CNN is that the learning performance of the model is during training is controlled with activation functions. In CNN a non-linear activation function called Rectified Linear Unit (ReLU) is commonly used. The ReLU activation function is expressed below in equation 3.2

$$f\ (x)\ =\ \max\ (0, x) \tag{3.2}$$

where x is the input to the function, and f (x) produces two possible values as output based on the input value. When the input value is less than 0 it outputs 0 and outputs the input value when input value is greater than or equals 0. Figure 3.5 a) and Figure 3.5 b) below shows the ReLU activation function graph and how it operates on a feature map.

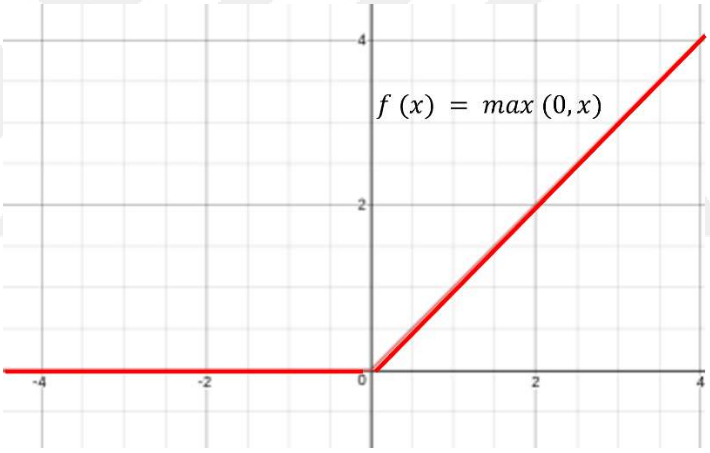

Figure 3.5 a). ReLU Activation Function Graph

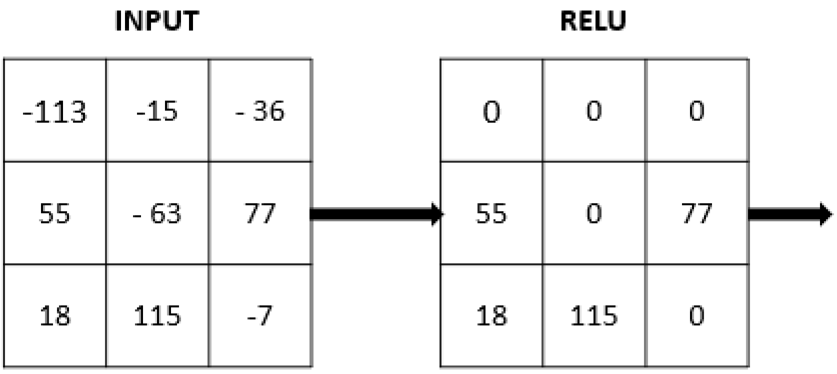

Figure 3.5 b). ReLU Operation on an Input [52]

3.1.3. Pooling Layer

The pooling layer function to reduce the spatial dimension of the input image. It is place right after the convolutional layer and this helps in controlling network overfitting and reducing computation. The pooling layer uses a pooling filter of small size that slides across the feature map to produce a down-sampled feature map. Some of the popular pooling techniques use in the CNN pooling layers are “max pooling” [53] and “average pooling” [54] [55]. The max pooling technique operate by choosing the largest pixel value of an image patch that corresponds to the dimension of the pooling filter used. The average pooling computes the average of the pixel values within an image patch that corresponds to the dimension of the pooling filter used. Figure 3.6 a) and b) below illustrates the max and average pooling operation.

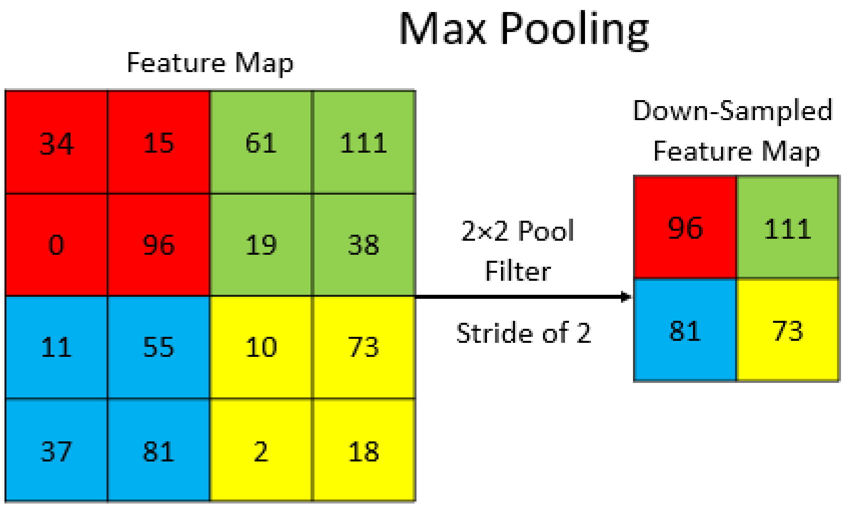

Figure 3.6 a). Max Pooling Illustration [56]

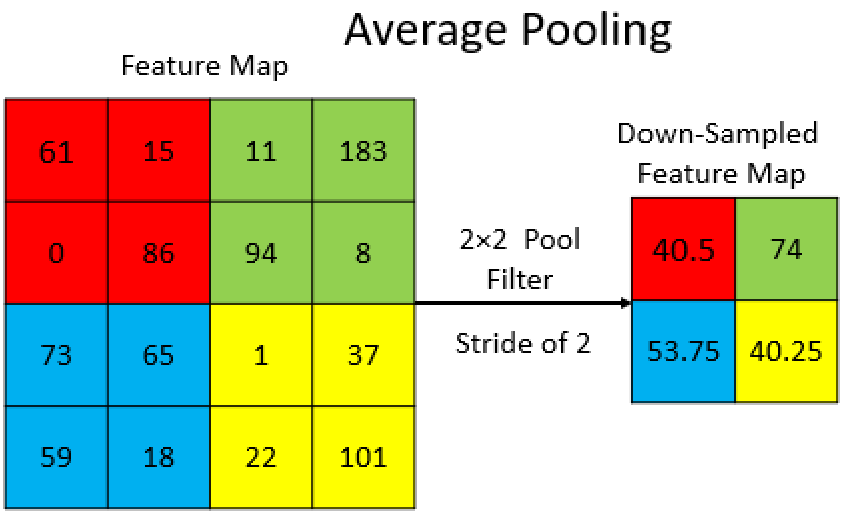

Figure 3.6 b). Average Pooling Illustration [56]

3.1.4. Fully-Connected Layer

The fully-connected layer also called the dense layer consist of neurons in each layer connected together. The convolutional layer's extracted features are flattened and given to the CNN's fully-connected layers to learn from. Typically, at least one fully connected used in CNN. In the fully-connected layer(s), the classification is at the last layer using a SoftMax classifier to calculate class probability.

4. METHODOLOGY

This chapter provides a detailed explanation of the proposed model used for detecting copy-move forgery in digital images. In coming up with the proposed model for copy-move forgery detection in digital images, we conducted 3 separate experiments that utilized the deep learning techniques. From the 3 implemented models, the model that performs best when evaluated was considered our proposed model. The deep learning techniques used in this research involve implementing a custom convolutional neural network and two fine-tuned pretrained deep neural networks. To carry out our experiment, the block diagram in Figure 4.1 shows the process flow of our proposed model. The remaining part of this chapter provides a detailed description of the various stages of the experiment carried out in the proposed model.

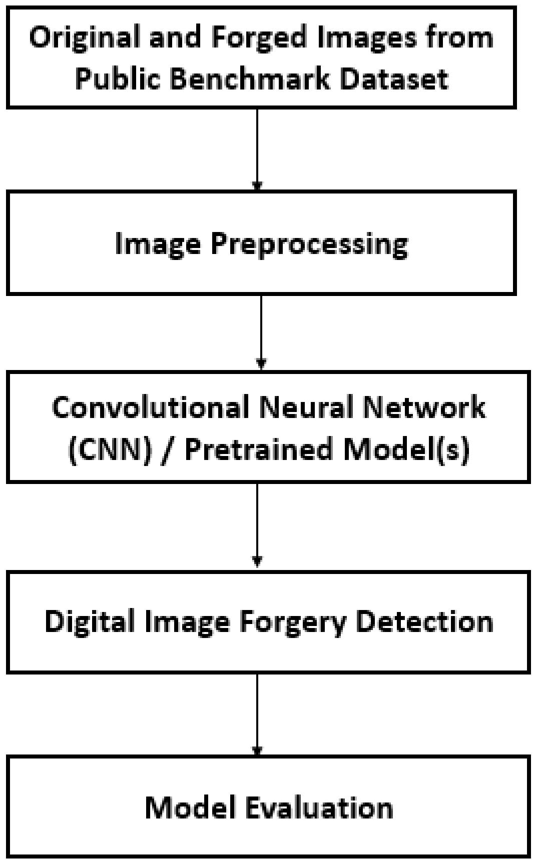

Figure 4.1. Block Diagram of the Proposed Model

Figure 4.1 above is a block diagram showing the processes involve in carrying out the research work.

4.1. Dataset Description

The dataset used in this research was retrieved from a database of images prepared for copy-move forgery detection in digital images. The copy-move forgery detection (CoMoFoD) database was prepared by [57]. The CoMoFoD database is a collection of 200 small-sized (512 × 512) and 60 large-sized (3000 × 2000) images grouped based on transformations applied to the images. The transformations applied to these images are translation, scaling, rotation, distortion, and combination.

1. Translation means the copied portion in the image is only pasted to a different portion in the image without transforming it.

2. Scaling means that the copied portion is scaled and pasted to a different portion in the image.

3. Rotation means that the copied portion of the image is rotated and pasted to a new portion in the image.

4. Distortion means that the copied portion of the image is distorted and pasted to a different location in the image.

5. Combination means that the coped region is applied two or more of the above transformation techniques and then pasted to a different location in the image.

In addition, various post-processing operations were applied to both the original and forge images in the CoMoFoD database. The post-processing operation applied are JPEG compression, noise addition, image blurring, color reduction, brightness change, and contrast adjustment. Table 4.1 and 4.2 below show some sample images from the CoMoFoD database and a description of the dataset used in the experiment, respectively.

Table 4.1. Sample Images from CoMoFoD Database

Transformation Techniques	Original Images	Forged Images
Translation		
Scaling		
Rotation		
Distortion		
Combination		

Table 4.1 above shows the sample images from the CoMoFoD database with their respective transformation techniques for both original and forge images.

Table 4.2. Description of Dataset Used in the Proposed Study

Name of Dataset	Dimension of Images	Number of Original Images	Number of Forge Images
CoMoFoD	512×512	5000	5000

Table 4.2 above describes the dimension of images in the CoMoFoD database and the number of original and forge images in the database.

4.1.1. Dataset Splitting and Preprocessing

The image dataset comprising 0f 10000 images from the CoMoFoD database used in this research was divided into three set: training, validation, and testing set.

Table 4.3. Description of Image Dataset Split Ratio

	Ratio	Number of Images
Training Set	8	8000
Validation Set	1	1000
Testing Set	1	1000

Table 4.3 above shows the splitting ratio and number of images in each set. The ratio of the image dataset set split was 8:1:1. This ratio means 80%, 10%, and 10% of the total number images were used for training, validation, and testing, respectively. The resulting number of original and forge images in the training, validation, and testing set are 8000, 1000, and 1000 respectively.

The image dataset used in the research was preprocessed before passing it to our CNN and pretrained model(s) for training. The preprocessing technique applied to the images in our dataset was resizing. The original dimension of images in our dataset was 512×512 and was down-sized to the dimension 224×224. Image resizing is an essential preprocessing step in image processing because limitations in system memory can make training a deep neural network model on high-resolution images challenging to perform. Also, high-resolution images may cause slow training and a poor model result [34] – [36]. Also, the pixel intensities of the images were normalized by dividing every pixel in the image by its largest pixel value. The pixel intensities of the images in our datasets fall within the range of 0 and 255. Thus, to normalize the images, every pixel was divided by 255. This normalization technique makes the image pixel fall within the range of 0 and 1, reducing the computational time during training [61].

4.2. CNN Architectures

In this research work, 3 experiments were carried out to detect a copy-move forgery in digital images. The CNN architectures implemented were: custom CNN model, VGG16, and ResNe50 pretrained network models in these experiments. The architectures of the CNN models used in our experiments are discussed as follows.

4.2.1. Custom CNN Model

The custom CNN model architecture implemented for detecting copy-move forgery in images comprises of 8 convolutional layers, 3 fully-connected layers. It uses the rectified linear unit (ReLU) activation function in the 8 convolutional layers and the first two fully-connected layers. In the last fully-connected layer, also called the output or classification layer, the SoftMax activation function was used. Batch normalization and dropout were added to the model layer as regularization technique to prevent overfitting during model training [62]. Table 4.4 shows the architecture of the custom CNN model used for the experiment.

The first and second convolutional layers used 32 filters of size 3×3 were used. In the third and fourth convolutional layers, 64 filters of size 3×3 were used. Following the fourth convolutional layer are the fifth and sixth convolutional layer which uses 128 filters of sizes 3×3. The last two convolutional layers, the seventh and eighth layers, use 256 filters of size 3×3. The activation function used in the convolutional layers was ReLU.

In the fully-connected layers, the first two layers comprises of 1024 neurons or nodes each. The last fully-connected layer has two nodes. The two nodes in the output layer indicate that the CNN model has two class labels to predict. Thus, in the case of our research, the two-class labels that the CNN model was predicting were either an image labeled as original or forged.

Table 4.4. Architecture of Custom CNN Model for Copy-Move Forgery Detection

Layer Type	Number Filters/Filter Size	Stride	Feature Map Shape
Input Image	-	-	224×224×3
Conv2D -1	32/3×3	1×1	224×224×32
Batch Normalization	-	-	224×224×32
Conv2D -2	32/3×3	1×1	224×224×32
Batch Normalization	-	-	224×224×32
Max Pooling	2×2	2×2	112×112×32
Dropout (0.25)	-	-	112×112×32
Conv2D -3	64/3×3	1×1	112×112×64
Batch Normalization	-	-	112×112×64
Conv2D -4	64/3×3	1×1	112×112×64
Batch Normalization	-	-	112×112×64
Max Pooling	2×2	2×2	56×56×64
Dropout (0.25)	-	-	56×56×64
Conv2D -5	128/3×3	1×1	56×56×128
Batch Normalization	-	-	56×56×128
Conv2D -6	128/3×3	1×1	56×56×128
Batch Normalization	-	-	56×56×128
Max Pooling	2×2	2×2	28×28×128
Dropout (0.25)	-	-	28×28×128
Conv2D -7	256/3×3	1×1	28×28×256
Batch Normalization	-	-	28×28×256
Conv2D - 8	256/3×3	1×1	28×28×256
Batch Normalization			28×28×256
Max Pooling	2×2	2×2	14×14×256
Dropout (0.25)	-	-	14×14×256
Flattened	-	-	50176
Fully-Connected	1×1	-	1024
Batch Normalization	-	-	1024
Fully-Connected	1×1	-	1024
Batch Normalization	-	-	1024
Dropout (0.25)	-	-	1024
Output Layer	-	-	2

4.2.2. VGG16 Pretrained Model

The VGG16 network is a Deep CNN proposed by a team [63] at the Visual Geometry Group of the University of Oxford during the ImageNet 2014 challenge. The ImageNet challenge evaluates models trained on about 14 million images from the ImageNet dataset to predict an image class from 1000 classes of the large-scale images used in training the network. The architecture of the VGG16 network architecture is depicted below.

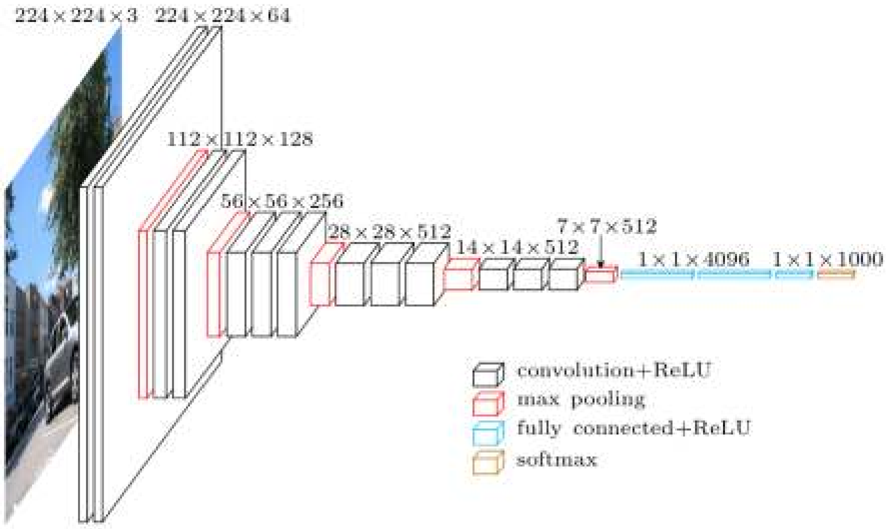

Figure 4.2. Visual Representation of VGG16 Network Architecture [64]

The VGG16 network as shown in figure 4.2 above consists of 13 convolutional layers with 3×3 convolution filters and 3 fully-connected layers. Its final layer uses a SoftMax activation function. Our research utilized the pre-trained VGG16 network model and then fine-tuned it to form a CNN model that can detect a copy-move forgery in digital images. Since the original VGG16 network has achieved a remarkable test accuracy of 92.7% when trained ImageNet dataset. We take advantage of the pre-trained VGG16 model by modifying its fully-connected layer.

The fine-tuned VGG16 model used to detect a copy-move forgery in images is configured as shown in Table 4.5 below. The original fully-connected layers in the VGG16 network model were replaced with new fully-connected layers. In the fined-tuned VGG16 network, 2 new fully-connected layers were created and connected to the 13 convolutional layers of the original VGG16 network model. The number of neurons used in fully-connected layers is 4096 and 2 for the first and second layers, respectively. The second layer is the output layer, which has 2 neurons representing the original and forge class label of images passed to the network. The output layer uses a SoftMax activation function.

Table 4.5. Architecture Fined-Tuned VGG16 Network Model for Copy-Move Forgery Detection

Layer Type	Number Filters/Filter Size	Stride	Feature Map Shape
Input Image	-	-	224×224×3
Conv2D -1	64/3×3	1×1	224×224×64
Conv2D -2	64/3×3	1×1	224×224×64
Max Pooling	2×2	2×2	112×112×64
Conv2D -3	128/3×3	1×1	112×112×128
Conv2D -4	128/3×3	1×1	112×112×128
Max Pooling	2×2	2×2	56×56×128
Conv2D - 5	256/3×3	1×1	56×56×256
Conv2D - 6	256/3×3	1×1	56×56×256
Conv2D -7	256/3×3	1×1	56×56×256
Max Pooling	2×2	2×2	28×28×256
Conv2D - 8	512/3×3	1×1	28×28×512
Conv2D - 9	512/3×3	1×1	28×28×512
Conv2D - 10	512/3×3	1×1	28×28×512
Max Pooling	2×2	2×2	14×14×512
Conv2D - 11	512/3×3	1×1	14×14×512
Conv2D - 12	512/3×3	1×1	14×14×512
Conv2D - 13	512/3×3	1×1	14×14×512
Max Pooling	2×2	2×2	7×7×512
Flattened	-	-	25088
Fully-Connected	1×1	-	4096
Output Layer	-	-	2

4.2.3. ResNet50 Pretrained Model

The ResNet50 model is one of the deep residual network models introduced by [65]. A residual network is a very deep neural network that has a residual block in which the layers within the block learn a residual function based on the input of the block. This residual function can be thought of as a function that fine-tunes the input feature map to produce a better-quality feature map. The residual block was introduced as a result of the performance degradation problem observed in deep neural networks [66], [67]. Although, there are very deep neural network models that have produced outstanding results ImageNet dataset [63], [68], [69]. However, increasing the depth of a neural network leads to a higher training error and rapid degradation of the network result. Figure 4.3 shows the connection of the residual block. In the residual block diagram, it can be seen that the skip connection was created that performs identity mappings.

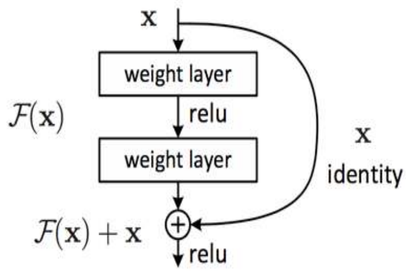

Figure 4.3. Residual Block

The architecture of the ResNet50 model, as shown in Figure 4.4 below, comprises of 3 convolutional layers within every residual block of the network. The first convolutional layer in the network has 64, 7×7 filters with a stride of 2. It is followed by a 3×3 max pooling operation with a stride of 2. The max-pool output is passed to the residual block with a series of 3 convolutions: 64, 1×1 filters, followed by 64, 3×3 filters, and lastly 256, 1×1 filters. This residual block is repeated 3 times. This sum-up to 9 convolutional layers in step. The subsequent residual block is repeated 4 times and consists of a series of 3 convolutions: 128, 1×1 filters, followed by 128, 3×3 filters, and 512, 1×1 filters. This sum-up to 12 convolutional layers in this stage. Next, is another residual block repeated 6 times and consists of 3 series of convolutions: 256, 1×1 filters, followed by 256, 3×3 filters, and lastly 1024, 1×1 filters. This gives us a total of 18 convolutional layers in this step. And then, the final residual block is repeated 3 times, containing 3 convolutions as follows: 512, 1×1 filters, followed by 512, 3×3 filters, and lastly, 2048, 1×1 filters. This gives us the sum of 9 convolutional layers in this step. Next, an average pooling operation is performed. Finally, the output of the average pooling is passed to a fully connected with 1000 neurons. This 1000 nodes fully-connected layer represent the output layer of the network. It uses the SoftMax activation function in the output layer.

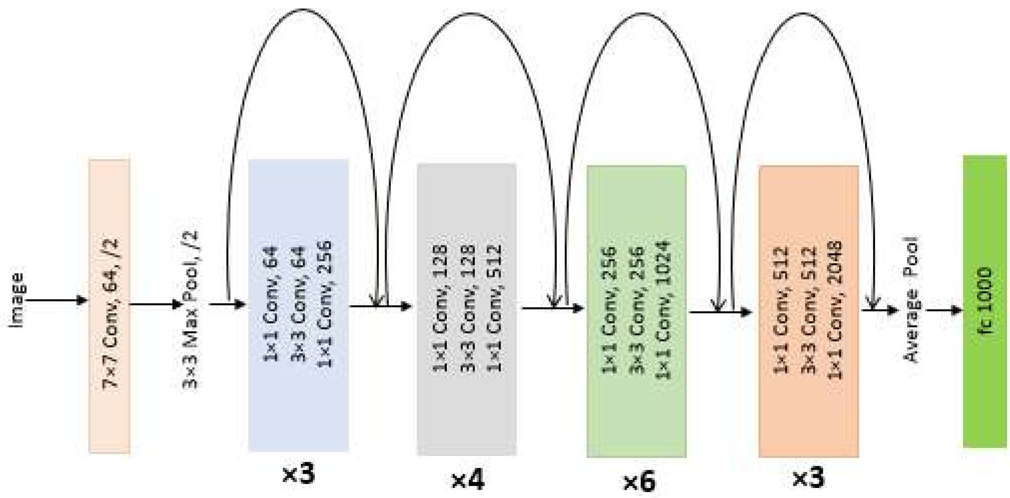

Figure 4.4. Architecture of ResNet50 Model

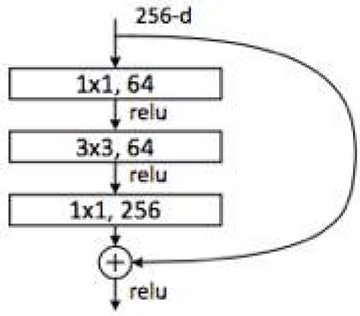

Figure 4.5. Architecture of ResNet50 Model Residual Block

Figure 4.5 above shows the residual block of used in the original ResNet50 model. It is the first residual block of the ResNet50 model which comprises of

The fined-tuned ResNet50 implemented for the detection of copy-move forgery in digital images replaces the fully-convolutional layers in the original ResNet50 network architecture. After the average pooling in the original ResNet50 model, the output was flattened and connected to 4 fully-connected layers. The first fully-connected layer has 1024 neurons followed by a dropout of 0.5. The next fully-connected layer has 512 neurons followed by a dropout of 0.5. The third fully-connected layer has 256 neurons, followed by a dropout of 0.5. The final fully-connected layer has 2 neurons that represent the output layer. The output layer uses the SoftMax activation function. The architecture of the fine-tuned ResNet50 model implemented for our problem is depicted in Figure 4.6 below:

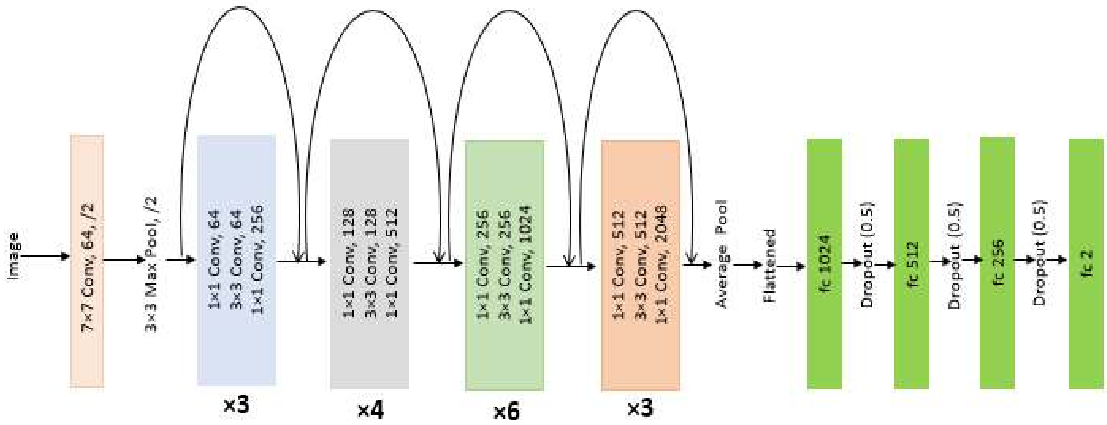

Figure 4.6. Fine-Tuned ResNet50 Architecture for Copy-Move Forgery Detection

4.3. Experiments

In this research, 3 experiments were conducted in proposing a model that detects copy-move forgery in digital images. These experiments were implemented using the python programming language. Python allows us to utilize the Keras library to build convolutional neural networks and load pre-trained models like VGG16 and ResNet50 network architecture. The system's hardware configuration for implementing our experiments has the following: Intel Xeon (R) CPU E4-2620 v4 @ 2.10GHz, 32.0GB system RAM, and a 24.0GB NVIDIA Quadro M4000 graphics processing unit (GPU).

In carrying out the experiments in this research, our dataset's training and validation set were passed to the CNN models implemented in this work for training. However, before passing the images in the datasets to our CNN models for the experiment, the original and forged labels of the images in our dataset were transformed into a one-hot encoded form. One-hot encoding is a coding technique for categorical variables. In transforming the original and forge labels to one-hot encoding, the labels were first converted to integer variables where 1 represents original, and 0 represents forge. Then the integer representation of the labels was transformed to their corresponding one-hot encoded form as [0, 1] and [1, 0] for original and forge label, respectively.

4.3.1. Experiment Analysis for Custom CNN Model

In training the custom CNN model implemented in this work, we used a categorical-cross entropy loss function for training the model. A learning rate of 0.0001 using the Adam optimization algorithm was initialized to control the step-size weight update of our model after every iteration [70]. The learning rate is a hyperparameter that determines how fast our model reaches the minimum of our loss function. Also, a learning rate decay was initialized in the model to reduce the learning rate of the model overtime during the training process. Using Keras as our deep learning framework, it allows us to set learning rate decay via the decay parameter in the Adam optimizer class. The setting is done by dividing our learning rate by the total number of epochs the model was expected to do during training.

The total number of epochs set for training this model was 100, and the batch size used was 32. Although the model was set to run for 100 epochs, an early-stopping method was initialized. Early stopping is a method used to monitor the improvement of certain performance metrics of a model over a consecutive number of epochs during the model's training process. When the conditions provided for the early stopping indicate no improvement, the training process stops automatically. In this model, the early stopping parameter used observes the validation loss of the model over 5 consecutive epochs with no improvement.

4.3.2. Experiment Analysis for Fine-Tuned VGG16 Model

The Fined-Tuned VGG16 model implemented for the detection of copy-move forgery in digital images was trained using a learning rate of 0.0001. However, the initialized learning rate of the model was set to decay over time during training. This model also uses the categorical cross entropy as its loss function and Adam as its optimization algorithm. The model was set to run for 60 epochs using a 32 as the batch size of the images passed to the network in each iteration. This model also utilizes the early stopping method, which monitors the validation loss of the model over 10 consecutive epochs.

The convolutional layers of the fined VGG16 model were freeze and were not trained during the training process. The model only trains on the new fully-connected layers implemented. Freezing the convolutional layers allows us to utilize and maintain highly rich image features extracted by the original VGG16 network to train our network.

4.3.3. Experiment Analysis for Fine-Tuned ResNet50 Model

In training the fined-tuned ResNet50 model for copy-move forgery in digital images, a learning rate of 0.0001 was initialized. This learning rate was also set to decay over time during the training process. The model was set to run for a total of 100 epochs while early-stopping was early stopping was monitored to control model overfitting. The validation loss was the metric monitored in the early stopping method for 10 consecutive epochs. The image batch size passed to this model in each iteration of the training process was 32.

The training of the fined-tuned ResNet50 model occurs from the last 3 residual blocks to the new fully-connected layers implemented in the network. This means the convolutional layers of the network before the last 3 residual blocks were all freeze during the training process. Training the model from the convolutional layers in the last 3 residual blocks allows us to extract highly discriminative features from the images passed to the network.

5. RESULTS AND DISCUSSION

This chapter gives a detailed description of the experimental results obtained during the training process, evaluations metrics employed to determine the performance of the trained model(s) in the experiment(s), and compares the results with findings of other studies.

To evaluate the performance of the proposed model of our experiments, an unseen dataset referred to as the test set data was used. The use of the images from the test dataset allows us to evaluate the trained models' performance in detecting copy-move forgery in an unseen image. Also, this ensures that overfitting has not happened. Overfitting occurs when a trained model performs very well on the training data but poorly on unseen data.

5.1. Experiments Performance Result

This section gives a detailed explanation of the performance results of the experiments conducted in this study. When the training process of the implemented models finishes, the plot for loss and accuracy curves of the models as shown in Figure 5.1 – Figure 5.6 below were generated. The generated loss (training and validation loss) and accuracy (training and validation accuracy) curves were for training and validation data used during model training. The loss curve shows how well the model performs in making the correct prediction after every epoch during training. The accuracy shows the fraction of correct predictions achieved during every epoch of the training process.

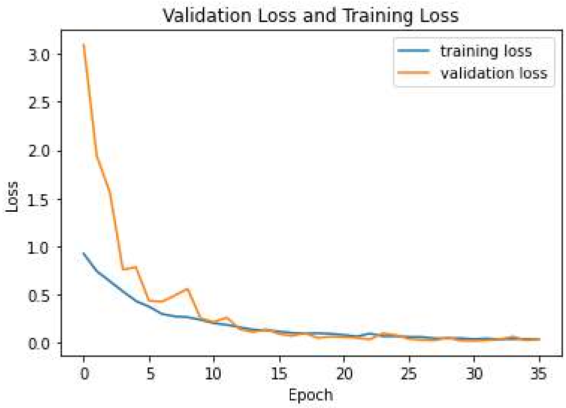

Figure 5.1. Custom CNN Loss Curve

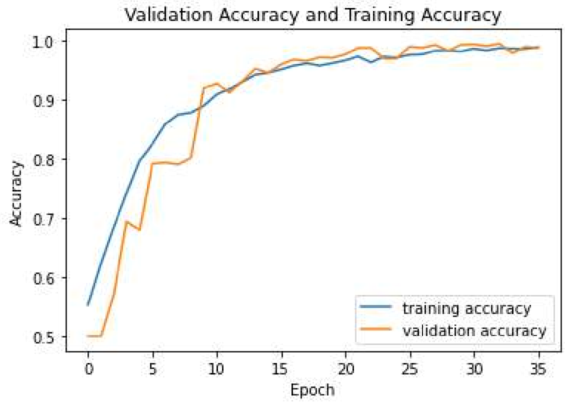

Figure 5.2. Custom CNN Accuracy Curve

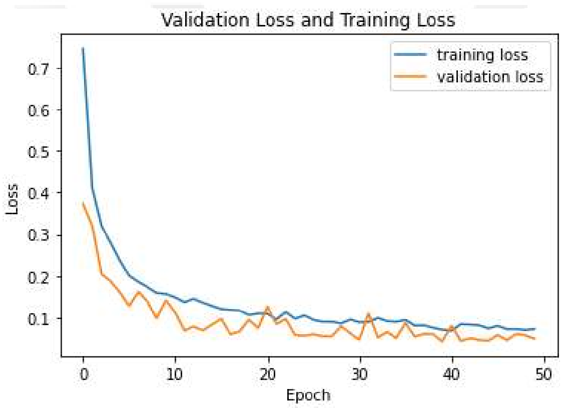

Figure 5.3. Fine-Tuned VGG16 Loss Curve

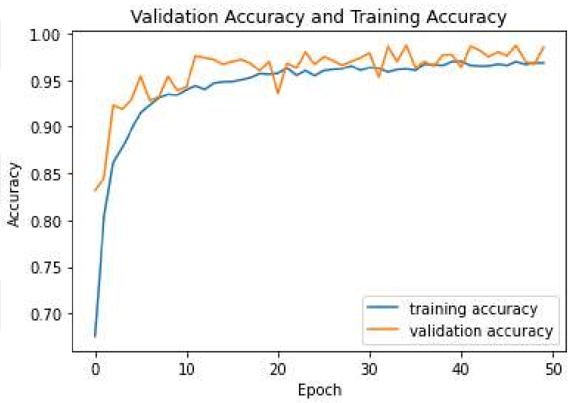

Figure 5.4. Fine-Tuned VGG16 Accuracy Curve

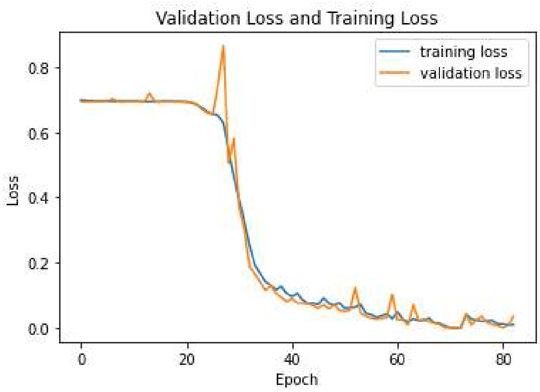

Figure 5.5. Fine-Tuned ResNet50 Loss Curve

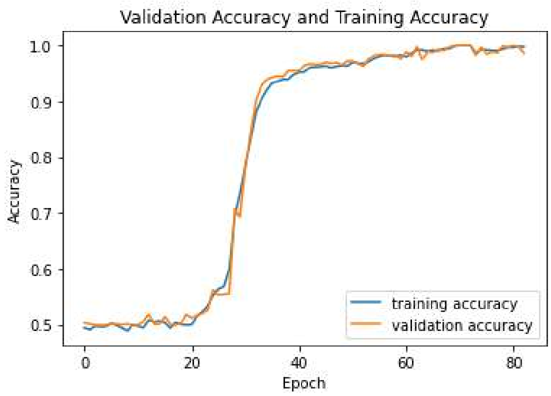

Figure 5.6. Fine-Tuned ResNet50 Accuracy Curve

Figures 5.1 and Figure 5.2 depict the loss and accuracy curve of the custom CNN model implemented. From the curves, it can be seen that model trains for 36 epochs. The training and validation loss/accuracy can be seen converging as the number of epochs increases.

Figures 5.3 and Figure 5.4 show the loss and accuracy curve of the fine-tubed VGG16 model. The loss and accuracy curve for the fine-tuned VGG16 model shows that both the training and validation loss/accuracy converged during the training process. The training process of this model stopped at the 50th epoch as a result of the early stopping method employed.

Figures 5.5 and Figure 5.6 depict the loss and accuracy curve of the fine-tuned ResNet50 model. From the curves of this model, it can be seen that the model training process start-off very slow as both the loss and accuracy were not improving until it reached the 22nd epoch. After the 22nd epoch, the loss and accuracy begins to improve until it reaches the 71st epoch. From there, the performance begins to saturate then the training process stops at the 83rd epoch due to the early stopping method employed.

5.1.1. Evaluation Metrics

The performance of the model in our experiments were evaluated using the following evaluation metrics.

1. **Accuracy:** Accuracy is a performance metric that provides the proportion of the accurate prediction of the forge and original image to all the images in the dataset. It is computed as shown in equation 4.1 below.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{5.1}$$

2. **Precision:** Precision is the proportion of accurately predicted forged class of images to the total predicted forged class of image. This is calculated using equation 4.2 below.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{5.2}$$

3. **Recall:** Recall is the proportion of accurately predicted forged class images to the total number of actual forged images. It is also referred to as Sensitivity or True Positive Rate. The mathematical expression for recall is shown in equation 4.3 below.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{5.3}$$

4. **F1-Score:** F1-Score is the harmonic average of precision and recall. It is mathematically expressed in equation 4.4 below.

$$\text{F1 Score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{5.4}$$

From the equations of the evaluation metrics defined above, TP is the true-positive rate that represent the number of forged images classified as forged images. TN is the true-negative rate. It denotes the number of original images classified as original images. FP is the false-positive rate that denotes the number of original images classified as forged images. And FN is the false-negative rate representing the number of forged images classified as original images. The true-positive, true-negative, false-positive, and false-negative rate are parameters of a confusion matrix as shown in Table 5.1 below.

Table 5.1. Confusion Matrix Table

	Predicted Forge	**Predicted Original**
Actual Forge	True Positive (TP)	False Negative (FN)
Actual Original	False Positive (FP)	True Negative (TN)

The performance result report of the custom CNN model, fine-tuned VGG16 model, and fine-tuned ResNet50 model are shown below in Table 5.1, Table 5.2, and Table 5.3. Also, the respective confusion matrices for the models are shown in Figure 5.7, 5.8 and 5.9 respectively.

Table 5.2. Custom CNN Model Performance Result

	Precision	**Recall**	**F1-Score**	**Number of Images**
Forge	1.00	0.98	0.99	490
Original	0.98	1.00	0.99	510
Accuracy			0.99	1000

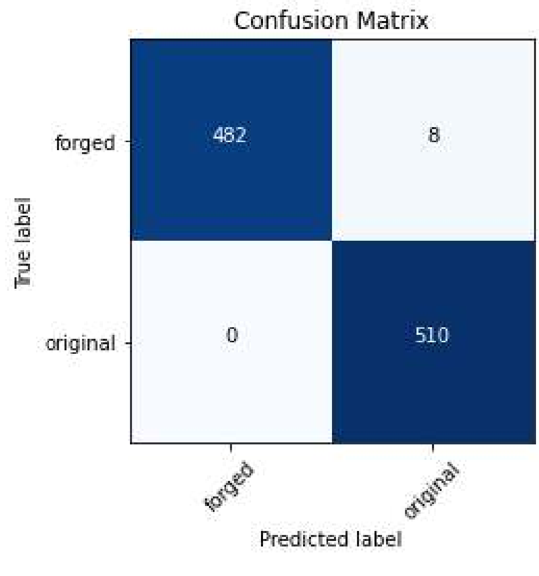

Figure 5.7. Confusion Matrix for Custom CNN Model

Table 5.3. Fine-Tuned VGG16 Model Performance Result

	Precision	Recall	F1-Score	Number of Images
Forge	0.95	0.99	0.97	490
Original	0.99	0.95	0.97	510
Accuracy			0.97	1000

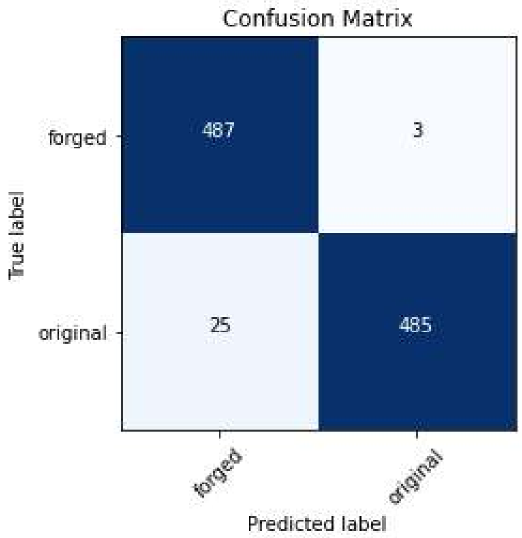

Figure 5.8. Confusion Matrix for Fine-Tuned VGG16 Model

Table 5.4. Fine-Tubed ResNet50 Model Performance Result

	Precision	Recall	F1-Score	Number of Images
Forge	1.00	1.00	1.00	490
Original	1.00	1.00	1.00	510
Accuracy			1.00	1000

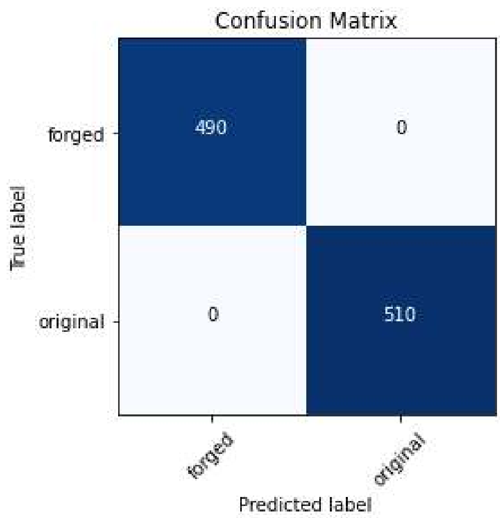

Figure 5.9. Confusion Matrix for Fine-Tuned ResNet50 Model

From the performance report of the models presented above, it can be seen that the model that has the best performance in detecting copy-move forgery in digital images was the fine-tuned ResNet50 model.

In Table 5.4 above, the accuracy, precision, recall, and f1-score achieved were all 1.00. This implies that 100% was obtained in all the evaluation metrics used to determine the performance of our model. The confusion matrix for the fine-tuned ResNet50 model as shown in Figure 5.9 above shows that model has no false-positives and false-negatives. This means all images in the test dataset were correctly classified as either forge or authentic.

Table 5.2 shows that the performance of the custom CNN model in detecting copy-move forgery in digital images was 0.99, 1.00, 0.98, and 0.99 for accuracy, precision, recall, and f1-score, respectively. And the performance for fine-tuned VGG16 network as shown in Table 5.3 was 0.97, 0.95, 0.99, and 0.97 for accuracy, precision, recall, and f1-score, respectively. The confusion matrix of the custom CNN model shown in Figure 5.7 shows that the model has 0 false-positive and 8 false-negatives. And in the fine-tuned VGG16 model its confusion matrix in Figure 5.8 depicts 25 false-positives and 3 false-negatives.

The performance results of the convolutional neural networks models show that they perform very well in extracting high-quality features for detecting forgery images. The best model found in our experiments was considered the proposed model in the research to detect copy-move forgery in digital images using the CoMoFoD database.

6. CONCLUSION

In this study, we propose a model for detecting copy-move forgery in digital images using CNN, which was based on a pre-trained model called ResNet50 network. A pretrained model allows us to utilize the small dataset to train our network and minimize the model training time. CoMoFoD image dataset contains original and forge images with post-processing attacks such as scaling, rotation, and noise addition was used to implement the proposed model to classify original and copy-move forge images. The performance of the proposed approach on the CoMoFoD dataset achieved an accuracy of 100.00%, precision of 100.00%, recall of 100.00%, and f1-score of 100.00% in the detection copy-move forgery in images. The proposed model performance shows that CNN is extremely effective and efficient in detecting forgery in digital images. Thus, it outperforms existing traditional methods.

RECOMMENDATIONS

Despite the achieved result from this proposed model, further research needs to be conducted to improve the model's ability to detect a forgery in digital images and localize the locations of forgery in the affected image in real-life scenarios. Image splicing and image retouching forgery are type of forgery that need explored deeply in order to proffer effective and efficient method for detecting them. Recently, generative adversarial network (GANs) is a deep learning technique for that use to generate a fake image from a real image. The fake image generated using GANs are realistic in nature and can be used deceive people viewing them. Therefore, proposing technique that can be able to differentiate fake GANs image from a real one will be interesting research to carry out.

REFERENCES

[1] N. Pantic, “How Many Photos Will Be Taken in 2021? - Mylio Blog.” https://blog.mylio.com/how-many-photos-will-be-taken-in-2021-stats/ (accessed Jan. 01, 2022).

[2] G. K. Birajdar and V. H. Mankar, “Digital image forgery detection using passive techniques: A survey,” Digital Investigation, vol. 10, no. 3. Elsevier Ltd, pp. 226–245, 2013. doi: 10.1016/j.diin.2013.04.007.

[3] M. Schreiber, “APTOPIX Trump Germany G20 | Buy Photos | AP Images | DetailView.” http://www.apimages.com/metadata/Index/APTOPIX-Trump-Germany-G20/84946d24e2d0479899aa37d7d9e22c6a (accessed Jan. 03, 2022).

[4] M. Novak, “That Viral Photo of Putin and Trump is Totally Fake.” https://gizmodo.com/that-viral-photo-of-putin-is-totally-fake-1796767457?IR=T (accessed Jan. 03, 2022).

[5] K. Asghar, Z. Habib, and M. Hussain, “Copy-move and splicing image forgery detection and localization techniques: a review,” Australian Journal of Forensic Sciences, vol. 49, no. 3. Taylor and Francis Ltd., pp. 281–307, May 04, 2017. doi: 10.1080/00450618.2016.1153711.

[6] A. Piva, “An Overview on Image Forensics,” ISRN Signal Processing, vol. 2013, pp. 1–22, 2013, doi: 10.1155/2013/496701.

[7] J. C. Lee, C. P. Chang, and W. K. Chen, “Detection of copy-move image forgery using histogram of orientated gradients,” Information Sciences, 2015, doi: 10.1016/j.ins.2015.03.009.

[8] V. Christlein, C. Riess, J. Jordan, C. Riess, and E. Angelopoulou, “An evaluation of popular copy-move forgery detection approaches,” IEEE Transactions on Information Forensics and Security, 2012, doi: 10.1109/TIFS.2012.2218597.

[9] N. B. A. Warif et al., “Copy-move forgery detection: Survey, challenges and future directions,” Journal of Network and Computer Applications. 2016. doi: 10.1016/j.jnca.2016.09.008.

[10] A. Boz and H. S. Bilge, “Copy-move image forgery detection based on LBP and DCT,” 2016. doi: 10.1109/siu.2016.7495802.

[11] X. Kang, L. Guangfeng, Y. Chen, E. Zhang, and G. Duan, “Detecting Tampered Regions in Digital Images Using Discrete Cosine Transform and Singular Value Decomposition,” International Journal of Digital Content Technology and its Applications, vol. 6, pp. 179–188, 2012, doi: 10.4156/jdcta.vol6.issue23.21.

[12] S. Sadeghi, S. Dadkhah, H. A. Jalab, G. Mazzola, and D. Uliyan, “State of the art in passive digital image forgery detection: copy-move image forgery,” Pattern Analysis and Applications, 2018, doi: 10.1007/s10044-017-0678-8.

[13] H. Hailing, G. Weiqiang, and Z. Yu, “Detection of copy-move forgery in digital images using sift algorithm,” 2008. doi: 10.1109/PACIIA.2008.240.

[14] 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, 2011, doi: 10.1109/TIFS.2011.2129512.

[15] C. C. Chen, W. Y. Lu, and C. H. Chou, “Rotational copy-move forgery detection using SIFT and region growing strategies,” Multimedia Tools and Applications, 2019, doi: 10.1007/s11042-019-7165-8.

[16] S. Dhivya, J. Sangeetha, and B. Sudhakar, “Copy-move forgery detection using SURF feature extraction and SVM supervised learning technique,” Soft Computing, 2020, doi: 10.1007/s00500-020-04795-x.

[17] A. Kuznetsov, “Digital image forgery detection using deep learning approach,” Journal of Physics: Conference Series, vol. 1368, p. 32028, Jan. 2019, doi: 10.1088/1742-6596/1368/3/032028.

[18] A. Bhandare, M. Bhide, P. Gokhale, and R. Chandavarkar, “Applications of Convolutional Neural Networks.” [Online]. Available: www.ijcsit.com

[19] N. Goel, S. Kaur, and R. Bala, “Dual branch convolutional neural network for copy move forgery detection,” no. December, pp. 1–10, 2020, doi: 10.1049/ipr2.12051.

[20] J. Chen, X. Kang, Y. Liu, and Z. J. Wang, “Median Filtering Forensics Based on Convolutional Neural Networks,” 2015.

[21] L. Bondi, L. Baroffio, D. Güera, P. Bestagini, E. J. Delp, and S. Tubaro, “First Steps Toward Camera Model Identification with Convolutional Neural Networks,” Mar. 2016, doi: 10.1109/LSP.2016.2641006.

[22] B. Bayar and M. C. Stamm, “A deep learning approach to universal image manipulation detection using a new convolutional layer,” in IH and MMSec 2016 - Proceedings of the 2016 ACM Information Hiding and Multimedia Security Workshop, 2016, pp. 5–10. doi: 10.1145/2909827.2930786.

[23] G. Xu, H. Z. Wu, and Y. Q. Shi, “Structural design of convolutional neural networks for steganalysis,” IEEE Signal Processing Letters, vol. 23, no. 5, pp. 708–712, May 2016, doi: 10.1109/LSP.2016.2548421.

[24] J. Zhang, Y. Li, S. Niu, Z. Cao, and X. Wang, “Improved fully convolutional network for digital image region forgery detection,” Computers, Materials and Continua, vol. 60, no. 1, pp. 287–303, 2019, doi: 10.32604/cmc.2019.05353.

[25] Y. Abdalla, M. Tariq Iqbal, and M. Shehata, “Copy-move forgery detection and localization using a generative adversarial network and convolutional neural-network,” Information (Switzerland), 2019, doi: 10.3390/info10090286.

[26] R. Agarwal and O. P. Verma, “An efficient copy move forgery detection using deep learning feature extraction and matching algorithm,” Multimedia Tools and Applications, vol. 79, no. 11–12, pp. 7355–7376, Mar. 2020, doi: 10.1007/s11042-019-08495-z.

[27] J. Ouyang, Y. Liu, and M. Liao, “Copy-move forgery detection based on deep learning,” Proceedings - 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics, CISP-BMEI 2017, vol. 2018-Janua, pp. 1–5, 2018, doi: 10.1109/CISP-BMEI.2017.8301940.

[28] M. A. Elaskily et al., “A novel deep learning framework for copy-moveforgery detection in images,” Multimedia Tools and Applications, vol. 79, no. 27–28, pp. 19167–19192, 2020, doi: 10.1007/s11042-020-08751-7.

[29] Abhishek and N. Jindal, “Copy move and splicing forgery detection using deep convolution neural network, and semantic segmentation,” Multimedia Tools and Applications, 2021, doi: 10.1007/s11042-020-09816-3.

[30] G. Muzaffer and G. Ulutas, “A new deep learning-based method to detection of copy-move forgery in digital images,” 2019. doi: 10.1109/EBBT.2019.8741657.

[31] A. Doegar, M. Dutta, and G. Kumar, “CNN based Image Forgery Detection using pre-trained AlexNet Model,” Proceedings of International Conference on Computational Intelligence & IoT (ICCIIoT) 2018, pp. 402–407, 2018.

[32] Y. Wu, W. Abd-Almageed, and P. Natarajan, “Image copy-move forgery detection via an end-to-end deep neural network,” Proceedings - 2018 IEEE Winter Conference on Applications of Computer Vision, WACV 2018, vol. 2018-Janua, pp. 1907–1915, 2018, doi: 10.1109/WACV.2018.00211.

[33] S. Samir, E. Emary, K. El-Sayed, and H. Onsi, “Optimization of a pre-trained AlexNet model for detecting and localizing image forgeries,” Information (Switzerland), vol. 11, no. 5, 2020, doi: 10.3390/INFO11050275.

[34] H. Farid, “Digital Doctoring: How to tell the real from the fake,” 2006.

[35] G. W. Meyer, H. E. Rushmeier, M. F. Cohen, D. P. Greenberg, and K. E. Torrance, “An Experimental Evaluation of Computer Graphics Imagery.”

[36] R. C. Gonzalez and R. E. (Richard E. Woods, Digital image processing. Prentice Hall, 2008.

[37] C. Solomon and T. Breckon, “Fundamentals of Digital Image Processing A Practical Approach with Examples in Matlab,” 2011.

[38] “Image Types in the Toolbox - MATLAB & Simulink.” https://www.mathworks.com/help/images/image-types-in-the-toolbox.html (accessed Jan. 06, 2022).

[39] W. Luo, Z. Qu, F. Pan, and J. Huang, “A survey of passive technology for digital image forensics,” Frontiers of Computer Science in China, vol. 1, no. 2. pp. 166–179, May 2007. doi: 10.1007/s11704-007-0017-0.

[40] V. Arul Karthick, “Forensic Technique for Detecting Tamper in Digital Image Compression,” 2013. [Online]. Available: www.ijarcce.com

[41] A. Swaminathan, M. Wu, and K. J. R. Liu, “Digital image forensics via intrinsic fingerprints,” IEEE Transactions on Information Forensics and Security, vol. 3, no. 1, pp. 101–117, Mar. 2008, doi: 10.1109/TIFS.2007.916010.

[42] M. Kirchner, “Efficient Estimation of CFA Pattern Configuration in Digital Camera Images.”

[43] D. Chauhan, D. Kasat, S. Jain, and V. Thakare, “Survey on Keypoint Based Copy-move Forgery Detection Methods on Image,” 2016. doi: 10.1016/j.procs.2016.05.213.

[44] E. Silva, T. Carvalho, A. Ferreira, and A. Rocha, “Going deeper into copy-move forgery detection: Exploring image telltales via multi-scale analysis and voting processes,” Journal of Visual Communication and Image Representation, 2015, doi: 10.1016/j.jvcir.2015.01.016.

[45] C. Kaur and N. Kanwal, “An analysis of image forgery detection techniques,” Statistics, Optimization and Information Computing, 2019, doi: 10.19139/soic.v7i2.542.

[46] P. Mantri and J. Thomas, “Design of Neural Networks,” IOP Conference Series: Materials Science and Engineering, vol. 1174, no. 1, p. 12023, Aug. 2021, doi: 10.1088/1757-899x/1174/1/012023.

[47] Z. Yanling, D. Bimin, and W. Zhanrong, “Analysis and study of perceptron to solve XOR problem,” in The 2nd International Workshop on Autonomous Decentralized System, 2002., 2002, pp. 168–173. doi: 10.1109/IWADS.2002.1194667.

[48] L. Deng and D. Yu, “Deep Learning: Methods and Applications,” Foundations and Trends® in Signal Processing, vol. 7, no. 3–4, pp. 197–387, 2014, doi: 10.1561/2000000039.

[49] R. Thakur and R. Rohilla, “Recent advances in digital image manipulation detection techniques: A brief review,” Forensic Science International. 2020. doi: 10.1016/j.forsciint.2020.110311.

[50] V. Lendave, “What Is A Convolutional Layer? -.” https://analyticsindiamag.com/what-is-a-convolutional-layer/ (accessed Jan. 08, 2022).

[51] A. Rosebrock, “Convolutional Neural Networks (CNNs) and Layer Types - PyImageSearch.” https://www.pyimagesearch.com/2021/05/14/convolutional-neural-networks-cnns-and-layer-types/ (accessed Jan. 08, 2022).

[52] M. aurelio Ranzato, Y. Boureau, and Y. Cun, “Sparse Feature Learning for Deep Belief Networks,” in Advances in Neural Information Processing Systems, 2008, vol. 20. [Online]. Available: https://proceedings.neurips.cc/paper/2007/file/c60d060b946d6dd6145dcbad5c4ccf6f-Paper.pdf

[53] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998, doi: 10.1109/5.726791.

[54] Y. LeCun et al., “Handwritten Digit Recognition with a Back-Propagation Network,” in Advances in Neural Information Processing Systems, 1990, vol. 2. [Online]. Available: https://proceedings.neurips.cc/paper/1989/file/53c3bce66e43be4f209556518c2fcb54-Paper.pdf

[55] M. Yani, Irawan S, and C. Setianingsih, “Application of Transfer Learning Using Convolutional Neural Network Method for Early Detection of Terry’s Nail,” Journal of Physics: Conference Series, vol. 1201, p. 12052, Jan. 2019, doi: 10.1088/1742-6596/1201/1/012052.

[56] D. Tralic, I. Zupancic, S. Grgic, and M. Grgic, “CoMoFoD - New database for copy-move forgery detection,” 2013.

[57] H. Touvron, A. Vedaldi, M. Douze, and H. Jegou, “Fixing the train-test resolution discrepancy,” in Advances in Neural Information Processing Systems, 2019, vol. 32. [Online]. Available: https://proceedings.neurips.cc/paper/2019/file/d03a857a23b5285736c4d55e0bb067c8-Paper.pdf

[58] C. F. Sabottke and B. M. Spieler, “The Effect of Image Resolution on Deep Learning in Radiography,” Radiology: Artificial Intelligence, vol. 2, no. 1, p. e190015, Jan. 2020, doi: 10.1148/ryai.2019190015.

[59] H. Talebi and P. Milanfar, “Learning to Resize Images for Computer Vision Tasks,” Mar. 2021, [Online]. Available: http://arxiv.org/abs/2103.09950

[60] P. Sane and R. Agrawal, “Pixel Normalization from Numeric Data as Input to Neural Networks,” Dec. 2017.

[61] C. Garbin, X. Zhu, and O. Marques, “Dropout vs. batch normalization: an empirical study of their impact to deep learning,” Multimedia Tools and Applications, vol. 79, pp. 1–39, Dec. 2020, doi: 10.1007/s11042-019-08453-9.

[62] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” 2015.

[63] T. Sugata and C. Yang, “Leaf App: Leaf recognition with deep convolutional neural networks,” IOP Conference Series: Materials Science and Engineering, vol. 273, p. 12004, Dec. 2017, doi: 10.1088/1757-899X/273/1/012004.

[64] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” Dec. 2015, [Online]. Available: http://arxiv.org/abs/1512.03385

[65] K. He and J. Sun, “Convolutional Neural Networks at Constrained Time Cost,” Dec. 2014, [Online]. Available: http://arxiv.org/abs/1412.1710

[66] R. K. Srivastava, K. Greff, and J. Schmidhuber, “Highway Networks,” CoRR, vol. abs/1505.00387, 2015, [Online]. Available: http://arxiv.org/abs/1505.00387

[67] K. He, X. Zhang, S. Ren, and J. Sun, “Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification,” CoRR, vol. abs/1502.01852, 2015, [Online]. Available: http://arxiv.org/abs/1502.01852

[68] S. Ioffe and C. Szegedy, “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,” CoRR, vol. abs/1502.03167, 2015, [Online]. Available: http://arxiv.org/abs/1502.03167

[69] D. P. Kingma and J. L. Ba, “Adam: A method for stochastic gradient descent,” ICLR: International Conference on Learning Representations, 2015.

CURRICULUM VITAE

Khalid Jibril SANI

[redacted]

[redacted] [redacted]
[redacted] [redacted]
[redacted] [redacted]
[redacted] [redacted]
[redacted] [redacted]
[redacted] [redacted]

[redacted]

[redacted] [redacted]
[redacted] [redacted]

[redacted]

[redacted] [redacted]
[redacted]
[redacted] [redacted]

[redacted]

- [redacted]
- [redacted]
- [redacted]

[redacted]

[redacted]

ACADEMIC ACTIVITIES