INVESTIGATION AND DEVELOPMENT OF CONVOLUTIONAL NEURAL NETWORK BASED IMAGE SPLICING DETECTION

By

SITI MASTURA BINTI MD HASIM

A Dissertation submitted for partial fulfilment of the requirement for the degree of Master of Science (Electronic Systems Design Engineering)

August 2017

Acknowledgement

First and foremost, I would like to show my highest gratitude to Allah SWT for all the blessings that made this work possible to be completed and the chance of meeting amazing people along the way of this wonderful journey.

I am very thankful to my supervisor, Associate Professor Dr Khoo Bee Ee who have been a supportive and understanding supervisor who never stop sharing her knowledge and guidance. Thank you for teaching me everything I need to know on this topic and sharing all the wonderful study experience with me.

To my beloved family, thank you for your time and energy in assisting me with taking care of my little one during my study period. This dissertation is made possible because of the efforts from my incredible family.

To Faculty of Electrical and Electronics Engineering, Universiti Sains Malaysia, to all Electronics System Design Engineering classmates Sidang 2016/2017, it was a great experience to meet and know everyone. Last but not least, to everyone that have been involved directly and indirectly in the completion of this study, thank you for your contribution. May Allah SWT bless all of us.

Credits for the use of the Columbia Image Splicing Detection Evaluation Dataset are given to the DVMM Laboratory of Columbia University, CalPhotos Digital Library and the photographers listed in
http://www.ee.columbia.edu/ln/dvmm/downloads/AuthSplicedDataSet/photographers.htm.

Table of Contents

List of Figures

List of Tables

List of Abbreviation and Nomenclatures

CASIA	Chinese Academy of Sciences Institute of Automation
CISDE	Columbia Image Splicing Detection Evaluation
CNN	Convolutional Neural Network
CUISDE	Columbia Uncompressed Image Splicing Evaluation
JPEG	Joint Photographic Experts Group
MATLAB	Matrix Laboratory
SGDM	Stochastic Gradient Descent with Momentum
SRM	Spatial Rich Model

PENYIASATAN DAN PEMBANGUNAN PENGESAN PENYAMBATAN IMEJ BERDASARKAN KONVOLUSI RANGKAIAN NEURAL

Abstrak

Pengesan penyambatan imej telah menjadi satu bidang kajian yang sangat penting di seluruh dunia. Kepentingan untuk mengesan penyambatan imej tidak terhad kepada pihak yang berkuasa sahaja malah kepada semua pengguna biasa. Pengesan penyambatan imej memerlukan beberapa langkah untuk dipenuhi dan set data yang besar diperlukan. Kajian ini bermatlamat untuk menyiasat dan membina kaedah berdasarkan konvolusi rangkaian neural (CNN) untuk mengesan penyambatan imej. Tiga eksperimen awal telah dilakukan berdasarkan kajian sebelum ini untuk menyiasat bagaimana pra-pemprosesan membezakan prestasi CNN. Dari eksperimen awal yang dijalankan, satu rangka kerja dengan pengurangan nombor lapisan CNN telah dicadangkan tanpa sebarang pra-pemprosesan. Pengesahan silang dengan sepuluh lipatan telah digunakan untuk mendemonstrasi prestasi CNN. Eksperimen awal telah menunjukkan prestasi CNN sangat terjejas dengan saiz imej input. Oleh itu, reka bentuk yang dicadangkan telah diuji dengan tiga imej input yang berlainan saiz iaitu 28×28 piksel, 64×64 piksel dan 128×128 piksel. Dari pengesahan silang, 64×64 piksel imej input telah dikonklusikan sebagai saiz yang paling sesuai untuk pengesan penyambatan imej menggunakan CNN. Di akhir kajian ini, dapat dilihat bahawa dengan menggunakan reka bentuk yang dicadangkan, CNN dapat digunakan tanpa sebarang pra-pemprosesan.

INVESTIGATION AND DEVELOPMENT OF CONVOLUTIONAL NEURAL NETWORK BASED IMAGE SPLICING DETECTION

Abstract

Image splicing detection is an area of studies that have been studied widely all around the world recently. The importance to do image splicing detection is not only for the authorities but also for common user. Image splicing detection requires several steps to be completed and a huge dataset is needed to be used. This study is aimed to investigate and develop CNN based method for image splicing detection. Three preliminary experiments are done according to previous work to observe how pre-processing affects CNN performance. Based on the preliminary experiments, an architecture with reduced number of CNN layers are proposed without any pre-processing. Ten-fold cross validation is used to demonstrate CNN performance. Preliminary experiments shows that CNN performance are critically affected by input image size. Therefore, the proposed architecture are tested with different input image sizes. Three different input image sizes are tested which are 28×28 pixel, 64×64 pixel and 128×128 pixels. From cross validation is can be concluded that 64×64 pixels input image is the most suitable input image size for CNN image splicing detection. At the end of this study, it is observed that by using the proposed architecture, CNN can be used for image splicing detection without any pre-processing.

CHAPTER 1

INTRODUCTION

1.1 Image Forgery

Images are two-dimensional figures that are captured by optical devices such as cameras. Last two decades ago, images had been widely used as crime solving prime evidence. However, nowadays there are a lot of ways to forge images and even videos by adding elements that does not originally exist in the original image and video. There are several ways to tamper images such as image splicing and copy-move. Copy-move is a process of copying a region of an image and pasting it into the other region of the image to cover or eliminate some part of the image. Image splicing technique is when a composite image is made by using a combination of two or more images (Ng & Chang, 2004).

Images and videos have been widely used around the world as prime evidence upon proving suspects guilty in criminal investigation. All data captured by cameras nearby and at the crime scene around the time of incident will be gathered and reviewed to detect potential witnesses and suspects. All individuals located near the crime scene will then be brought into interrogation with law enforcer to listen to their side of stories to determine whether they could be the witness of the event or the potential suspects. Alibies of each suspect will be asked during the interrogation and been checked accordingly. Images and videos will become the prime evidence to determine whether their alibies are true or false.

However, with evolving technology, images and video can be forged to include things that are not originally exist at the scene during the moment those images and videos were captured. This kind of act has made images and videos lost their credibility to become solid evidence as the information obtained can be false at the same time has been disrupting in criminal investigation and might cause innocent people being charged on criminal offences they never did.

With current technologies, images can be easily edited in any free applications on smartphones and also there are several free software available on the internet. These applications and software can be easily accessed and used by any range of age. Increasing numbers of advanced technologies nowadays are one thing to be proud of but the danger lies in the question of how the technologies are being used.

Taking pictures and uploading into social media is a normal habit nowadays. However, before uploading pictures on the media social, users tend to edit out all the flaws on their face such as acne, wrinkles and even make their skin tone look fairer and their face size look slimmer. By the end of the process, the ‘online-look’ and the true appearances are totally different. To keep up with the imaginary look, the users tend to have low self-esteem and start to put on so much make up, consume unreliable supplements that can cause harm to their body just to look like the ‘online-look’. Taking pictures and uploading it is not as simple as it was last time that it can also cause death because of the obsession on getting their perfect look and availability of harmful supplements claiming that the products can give them that look in the market nowadays.

Media social has been a fantasy place for criminals to hide and socialize. They are able to create fake accounts using other people’s profiles and stalk on their victims. Images that have been uploaded by them seemed to be real therefore, people believe them without second guess. However, when being traced those pictures are actually edited to create a persona that fit to their fake profiles.

There are many types of image forgery that are available nowadays. Copy-move and image splicing are among the most commonly used techniques in image forgery. Copy-move technique is a famous technique that has been widely used to alter image information which involves in copying a part of host image and pasted onto the host image which makes this technique is a bit harder to be detected as compared to image splicing technique because the copied region share some similar features with the original host image. Copy-move images are usually aimed to cover some parts of the image such as acne or wrinkles of the object where clear skin of the object will be copied and pasted to cover the unwanted part of the image.

Image splicing technique is where a region of forged image is copied from another image and pasted on the host image. Danger of this technique is that the information of the image are changed such as an object can be added into the image or removed. In this case, this type of images are not eligible to be used as prime evidence in criminal investigations as it would be misleading and misconception will happen. Innocent person will be convicted for offense that he did not executed. In order to avoid using spliced images as prime evidence, studies have been done to detect spliced images apart from authentic images. Several way have been introduced to achieve this.

Image forgery detection has been studied in order to be able to detect forged images from authentic images. Image forgery detection can be classified into two groups namely passive and active. Figure 1.1 shows the classification of image forgery detection. Active image forgery detection are based on embedded information in digital images which are limited to types of camera used to capture the image such as digital watermark and digital signature. Passive approach able to detect image forgery without any additional information and classified into tampering and source device. Features are extracted to detect manipulation done based on tampering detection or source device identification. Tampering can be divided into dependent and independent tampering where dependent require copy-move of images to be pasted onto the tampered image such as by using copy-move or splicing while independent tampering are done by manipulating of digital information in the images (Warif *et al.*, 2016). This study is focusing on image splicing detection.

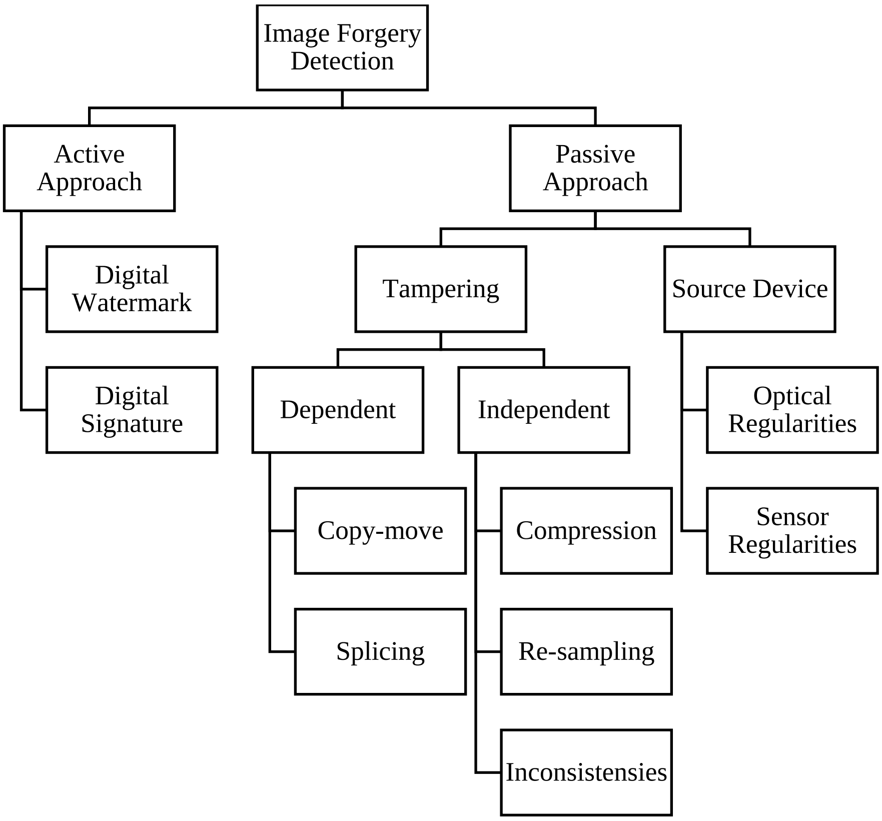


Figure 1.1: Illustration of Image Forgery Detection (Warif *et al.*, 2016)

1.1.1 Image Splicing Detection

Conventional image splicing detection requires image features to be predefined and extracted from the images, grouped into its classes and apply to the tested image to see which cluster it is in either in the original or fake image. These images are usually been cut in either overlapping or non-overlapping blocked and tested according to blocked and pasted back to its original position on whole image to detect which region had been altered (Mankar & Gurjar, 2015).

Previous work has shown that by doing image alteration, there are some other features that had been disturbed too such as image compression, pixel intensity difference as compared to its neighbouring pixel, colour intensity, edge abrupt discontinuity, contrast and noise (Park *et al.*, 2016). These are the features that had been used to detect original image apart from forged images. Features that are suitable in detecting image splicing technique is most of the features disturbed as there are high possibility that the pasted region is different from its host image in term of pixels intensity, colour and noise as they are originally from two different sources.

As a conclusion, image splicing is a rising research topic nowadays and studies have been done to detect spliced images apart from original images. Details on image splicing techniques that had been done in previous work will be discussed in Chapter 2 of this dissertation.

1.1.2 Convolutional Neural Network Technique

Machine learning that was aimed to mirror biological brain that can search for self-learn and differentiate from one thing to another. The concept of neurons connecting together to process information in human brains are adapted to this method (Ravi *et al.*, 2016). For example, given a red apple and a green apple and being asked what are the difference between the two of them, our brain will automatically think of the colour of the apples. Similar to when we are given a bus and a car to compare, our brain can list out several features that makes them different such as size, colour, number of seats, weight and so on. Deep learning is a field under machine learning, where multiple layers of learning where features can be self-learned and self-classified without the need of extra work on features extraction and classification.

Convolutional Neural Network (CNN) is a technique using deep learning method which had been an eye-opener to researches in current state as this might be the end to long image pre-processing in highlighting image features, features extraction and classification by using several methods combined together. The nature of self-learning features and classification of CNN has been a wonder when it comes to training massive numbers of dataset and large size of image. Researches involved in image forgery detection have not left out exploring other less-complex ways to achieve their goals. By using this technique, features extraction can be done automatically without changing much parameters as compared to conventional ways for image splicing detection.

CNN consist of three main layer namely convolution layer, pooling layer and fully connected layer (Guo *et al.*, 2016). Convolution layer act as features extraction, pooling layer follows convolution layer and is used to reduce parameters of feature maps and network and fully connected layer will then connect all the features and converted in a vector that will be used in classification. With increased number of CNN layers, more features can be learnt and used for classifications. However, too much layers can cause overfitting and reduce accuracy of the network. There are several parameters such as number of iterations, number of epochs and learning rate that play a role in CNN. Details on these CNN parameters will be shown in Chapter 2.

1.2 Problem Statement

Researches have been conducted to detect image splicing detection especially in obtaining great performance with minimal computational time and minimal human intervention (Bengio, 2009). Common framework of conventional image splicing detection requires features to be predefined at the beginning of the process. Features that are used to detect image splicing apart from authentic are limited to predefined features only. Human skills are required to predefine features. Mistakes done during the predefinition stage will cause wrong interpretation of spliced images during the training and testing stages (Bayar & Stamm, 2016).

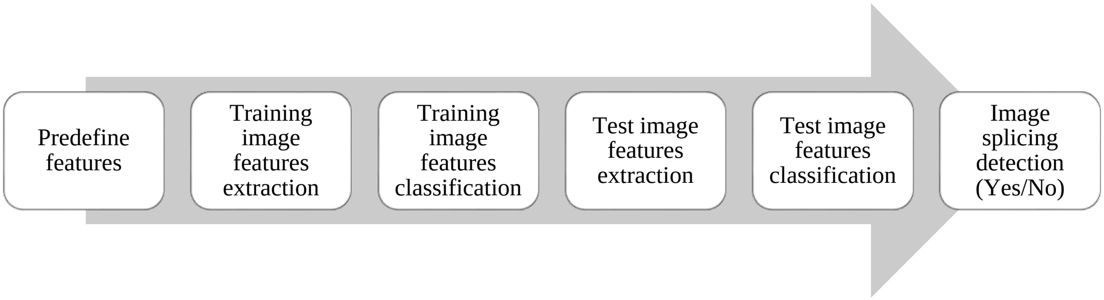

Predefine features
Training image features extraction
Training image features classification
Test image features extraction
Test image features classification
Image splicing detection (Yes/No)

Figure 1.2: Illustration of conventional image splicing detection

Figure 1.2 demonstrated steps of conventional image splicing detection where the process started with predefinition of features that are assumed to be able to detect spliced images apart from authentic images. The predefined features will be used for features extraction of training image and followed by classification of the predefined features. Kashyap *et al.* (2017) has shown that from most studies that they have reviewed, requires definition such as noise inconsistencies or JPEG compression to be defined in the program before features extraction can be done. Blind image splicing detection are not widely used yet. Features extraction and classification of testing images are also depending on predefined features. Features that are not defined will not be used during training and testing. Therefore, it is very crucial to correctly define features to be extracted and classified in the beginning of the framework.

Image splicing detection necessitates huge amount of image dataset in order to obtain good performance in terms of accuracy. Researchers who conducted works on image splicing detection usually used CASIA image tempering evaluation database and CISDE because of large number of images in the database. However, conventional image splicing detection have several steps that all the dataset images have to go through as shown in Figure 1.2.. To have a huge amount of dataset to be predefined correctly consumes a lot of time and expertise to ensure that features are correctly defined. As reducing the number of image dataset will affect detection performance results, a study is needed on reducing the number of steps and eliminating predefinition of features stage.

1.3 Objectives

This study is aimed to investigate the use of CNN for image splicing detection and develop convolutional neural network based of image splicing detection. To realise this aim, the following objectives are implemented:

1. To investigate and develop convolutional neural network based image splicing technique.
2. To measure the performance of the proposed image splicing technique.

1.4 Scope of Study

This work is focusing on grayscale and blocked images to reduce computational complexity and processing time to focus more on CNN training and testing architecture. Columbia Image Splicing Dataset (CISDE) has been used in this study as it complies with the necessity of this work which consist of grayscale and blocked images that can be localized into whole image according to their original position to detect spliced region in further study. Supervised training is done in this study as the dataset have already grouped spliced and authentic images. Supervised training means that both training and testing are done according to the groups of the input image and its accuracy are calculated according to true positive and true negative of the results.

To investigate CNN method to be applied to image splicing detection, three preliminary experiment is done according to three research paper that have used different CNN architecture. Preliminary experiment is aimed to observe CNN performance on CISDE dataset and how pre-processing such as filtering affect the performance. From this preliminary experiment, a CNN architecture is proposed in this study. Three different input image sizes are investigated to observe how input image sizes affect architecture performances.

Study has been done using MATLAB 2016a with Intel i5 6500. MATLAB toolboxes used in this study are Image Processing Toolbox, Computer Vision Toolbox, Neural Network Toolbox and Parallel Computing Toolbox. Memory used was 32GB with GPU Nvidia GTX1080Ti.

1.5 Dissertation Outline

This dissertation consists of five main chapters which explained the entire research study of this project.

Chapter 2 named as Literature Review shares several literature reviews on previous work related to this area of study such as image forgery techniques and image forgery detection techniques. The methods are compared to each other to see their limitations and advantages to choose the preferable method used in the preliminary experiment. Later in this chapter discusses on deep learning and CNN, the method that are used to simplify the architecture of conventional image splicing detection.

Chapter 3 describes on the methodology used in this study and each steps done in the process are discussed in details. The experimental setup of this project and validation is mentioned in this chapter for better overview on how the performance of the method is measured.

Chapter 4 shows results of this studies and comparisons between experiments have been done. The performance, effectiveness and stability of this current technique is discussed and validated k-fold validation.

The final Chapter 5 summaries all findings and concluded this study. Contributions and the significance of this project is mentioned and future recommendation in improving this research is suggested.

CHAPTER 2

LITERATURE REVIEW

2.1. Overview

In this chapter, existing image splicing detection will be discussed and the strength and weaknesses of the techniques will be discussed thoroughly. Later on, this chapter discusses on Convolutional Neural Network, a developing method that have been proven to have high performances in most of its network. This method is aimed to be used in this study for image splicing detection and to investigate how its parameters can affected the performance of this application. Steganalysis is also discussed in this chapter and its potential to be used in image splicing detection is highlighted considering how similar steganography and image splicing method are. This is a new findings that can be used later on in future studies. Three papers that are selected to be used in preliminary experiments will be shown in the later chapter to observe how these methods can be used to investigate CNN to be used with image splicing detection.

2.2. Image Splicing Detection Techniques

Recent years have shown high increment in digital images editing which is a great evolution in technologies. However, there are also irresponsible people using editing techniques to tamper and manipulate images to include false information (Park *et al.*, 2016). Copy-move is easy to be performed especially when both host image and pasted region are from the same image and they share similar properties such as illumination, intensity and noise making it is unnoticeable by human vision (Warif *et al.*, 2016)

Blocked-based approach is when whole image divided into smaller blocks either overlapping or non-overlapping blocks during pre-processing and feature extraction will done according to each blocks. There are several ways to detect image splicing such as frequency transform, texture and intensity similarities localization between spliced image and original image, moment's invariance and so on. These blocked images will be localized into the original image and spliced region can be obtained after the process (Warif *et al.*, 2016). Increasing of spliced area in a spliced image will increase the number of positive blocked images in the particular image. Authentic images will have only negative blocked images (Rao & Ni, 2016)

Introduction of noise to images are done by several researches and blind estimation of detected of background noise are done to check whether the image spliced or not. It is believed that spliced images will content higher noise compared to what have been inserted into the images. This kind of method is called noise inconsistency for image splicing detection. Blocked images that are used in this method can tell the user the exact location of spliced region once it is localized into the whole image. Dong *et al.* (2016) had introduced an image splicing detection by introducing artificial Gaussian noise and calculated blind noise estimation. They used Columbia uncompressed image splicing dataset and compared how block sizes of images affect the running time of splicing detection and compared detection result of spliced image to authentic images.

Park *et al.* (2016) has shown that image splicing detection can be done by transforming images into wavelet transform and difference in wavelet sub bands direction are calculated. Co-occurrence matrix are used to detect the offsets in the orientation and features are extracted by using characteristic function method that will reduce the dimension and classified by using support vector machine. This study has shown how reducing dimensionality can still remain the accuracy and Columbia dataset can achieve higher accuracy compared to colour dataset.

Another study had done on how multiple JPEG compression affected the consistency of quantization error. People usually use images available on the internet for splicing purposes as it is easily downloaded without registering any information of the downloader. Images on the internet are usually using JPEG format as it is compressed and easily uploaded. By using this method, quantization error of JPEG compression can be observed to be different according to how many times it is saved to the computer compared to the first timer saving (Han *et al.*, 2016). However, by using this method, images that have been edited using camera lenses cannot be detected as spliced and will be resulted as authentic.

2.2.1. Image Splicing Detection Dataset

There are several image splicing detection dataset available online such as CASIA Tampered Image Evaluation Database v1.0, CASIA Tampered Image Evaluation Database v2.0 (Dong, 2011), CISDE (Ng & Chang, 2004) and CUISDE (Hsu & Chang, 2006). These datasets have their own strength and weakness depending on the needs of the study.

CASIA v1.0 consists of both authentic and spliced images in RGB form without any post-processing after splicing is done to the image while CASIA v2.0 consists of images with post-processing spliced image. CASIA v2.0 is more unlikely to be detected by human vision compared to CASIA v1.0 that still can be distinguished by human vision. A total number of 1725 images with size of 374×256 pixels in CASIA v1.0 with 800 authentic images and 912 spliced images. In CASIA v2.0 dataset, there are 7491 authentic images and 5123 spliced colour images. The pixel size of the images are ranged from 240×160 to 900×600 pixels. Both CASIA dataset have categories such as scene, texture, animals, nature, plants and so one. However, in CASIA v2.0, some uncompressed image samples are included and also new indoor categories that will manipulate illumination of the images.

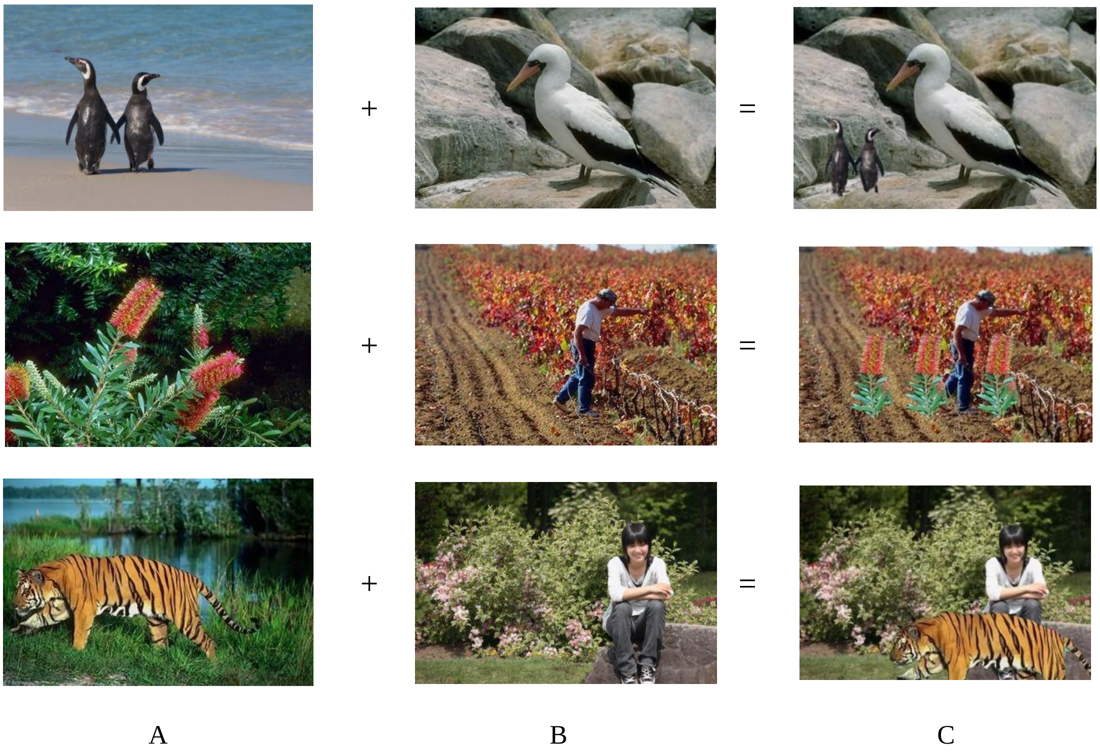


Figure 2.1: CASIA v1.0 Examples

Figure 2.1 shows example of CASIA v1.0 where the images in ‘Column A’ are cut and pasted into images in ‘Column B’ resulting with the image in ‘Column C’. Images in both Column A and B are classified as authentic image while images in Column C are categorised as spliced images.

CISDE is a grayscale authentic and spliced 128×128 pixels blocked-images. There are 933 authentic blocked images and 912 spliced blocked images. These images are grouped into several categories such as textured and homogeneous region, object boundary types and orientation of object boundaries either vertical or horizontal or others. CISDE have been a favourite dataset for most of researchers for its completed pre-processed that eased the job of the users and its small storage size that takes a short time to be downloaded and uploaded into the program. Figure 2.2 are some examples of Columbia Grayscale Blocked Image Splicing Dataset.

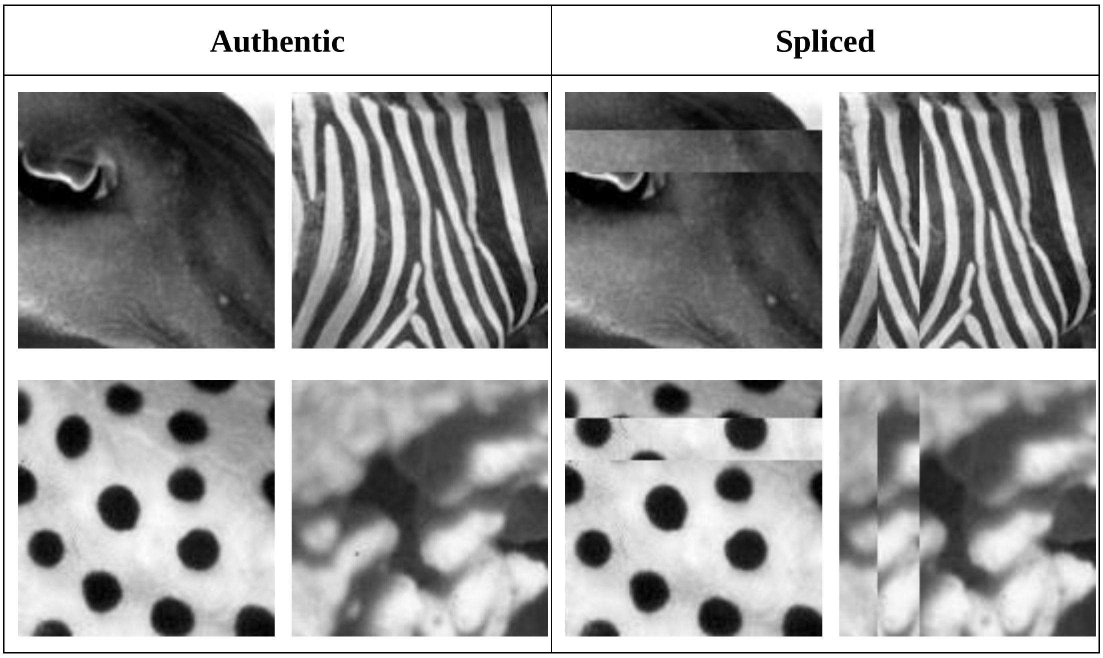


Figure 2.2: CISDE Examples

Most image splicing detection researches used either one or more from these three datasets. Table 2.1 datasets used by the researchers in image splicing detection.

Table 2.1: Datasets Used By Previous Researchers

	Dataset		
	CISDE	CASIA v1.0	CASIA v2.0
Shi *et al.* (2007)	√		
Chen et al (2007)	√		
Shi *et al.* (2008)	√		
Zhang *et al.* (2016)	√		
Park *et al.*(2016)	√	√	√
Rao & Ni (2016)	√	√	√

2.3. Convolutional Neural Network

Deep learning is a technique using machine learning principle and currently has increasing interest by researchers all over the world. Sudden interest increment is due to its availability to use GPU that is recently found in a research by Bayar & Stamm (2016) that increases chip processing abilities. Besides that, CNN has hardware computational cost that is lower as compared to other methods. CNN is considered as an advanced machine learning method that is able to learn features layer by layer and as deep as the user need to (Guo *et al.*, 2016).

CNN had been introduced in the era of early 1990s, however, there are no such clarity on what actually happen in the process and how does the network able to achieve such high performances. Therefore, Zeiler & Fergus (2014) had studied on how to visualize and understand every CNN layers by using deconvolution. Although in the studies done by Zeiler *et al.* (2011) had been used for features learning, the one used in Zeiler & Fergus (2014) are used as a platform of trained convolution network. It is shown that by the work of Zeiler & Fergus (2014) that CNN are able to learn simple features such as edges in the beginning of the layers and proceed with simple shapes in the next layers. More complex shapes with different orientation is learnt in the next layers followed by complex images that consists of several features such as different contrast and depth can be learnt later on.

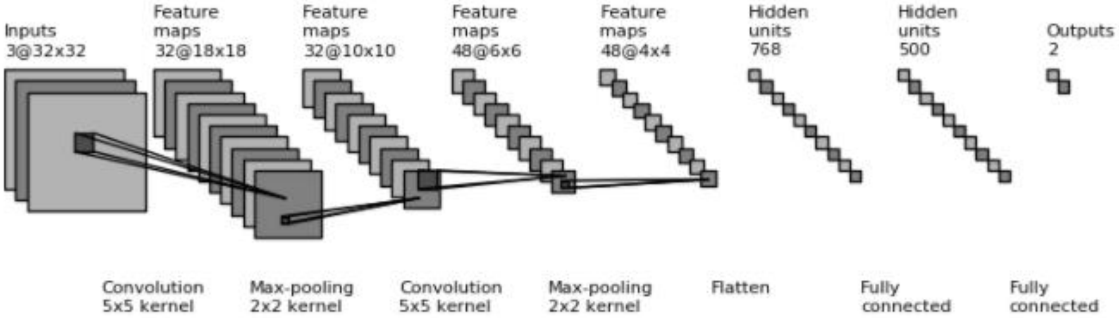


Figure 2.3: General CNN Architecture Illustration from Jefkine (2016)

Figure 2.3 by Jefkine (2016) shows a general CNN architecture with 32×32 pixels of input layer. First convolution was 5×5 kernel producing feature maps with size 18×18. Next, the architecture go through max pooling with 2×2 kernel resulting with feature maps with 10×10. Second convolution has the similar kernel size as the first convolution which is 5×5 making the output feature maps of 4×4. A total of 768 neurons are produced and went through two fully-connected layer and classified into two classes at the final layer.

Springenberg *et al.* (2015) had studied on simplifying convolutional neural network to only convolution layers without activation, pooling and fully-connected layers. They had come out with a framework consisting of only convolution layer pooling and fully-connected layer replaced with convolution layer. The aim of this study is to reduce the numbers of parameters for each network while maintaining good performance at the same time without data augmentation step.

Convolutional neural network have been widely used in previous work for object recognition and classification. A research done by Krizhevsky *et al.* (2012) were able to recognize up to 1000 different objects by using CNN. In their study, they have used over 15 million images as dataset which had awarded them as the largest CNN up to date of their research, 1000 object classes and have been referred and cited by hundreds of other studies done later than them.

2.3.1. Convolutional Neural Network Layers

In CNN, there are several main layers known as convolution layer, pooling layer and fully-connected layer. Convolutional layer act as features extraction layer where the output of previous layer will become input to the next layer. Pooling layer is used to reduce network size and prevent network overfitting issues. Output of pooling layer will be passed over to fully connected layer that will combine all features together and converted into a vector that will be used in classification.

2.3.1 (a) Input Image Layer

The first layer in CNN is input image layer. Input image layer is when input image to be trained or tested are read from the folder available in storage before undergoing the CNN framework. In this layer, input size of the images are justified and CNN features maps output from convolution layer depended on this input layer.

In this project, input image is CISDE image dataset that is labelled according to their folders either authentic or spliced. The original dataset image is resized to several pixel size to investigate how input image size affected the performance of CNN. In this study, images are resized to 28×28 pixels and 64×64 pixels and maintaining the 128×128 pixels. For each pixel sizes, the similar CNN architecture is applied but with different parameters according to the calculation of feature maps.

2.3.1 (b) Convolution Layer

In convolution layer, there are two parts; convolution and non-linearity. Output of convolution layer is feature maps which are known as feature representation of input in certain region. Input layer is convolved with convolution kernel added to variable output bias resulting in output layer which is also known as feature maps convolution layer included two main theory named as receptive fields and shared weights. Receptive field is where low level feature are convolved with input subset that will regulate total quantity of pixels connection and shared weights are useful to rise generalization result of the network and decreases number of free variables. Number of receptive field is number of local region of input volume where each neurons are connected to the convolution layer (Rao & Ni, 2016).

In convolution layer, there are several parameters that is required to be set such as number of filters, height and weight of the filters, padding size, stride size, number of channels, weight learning factor, bias learning factor, weight regularization factor, and bias regularization factor. In this study, only several important parameters have been used to simplify the architecture and calculations.

Table 2.2 are the definition of parameters used in this project.

Table 2.2: Parameters of Convolution Layer

Name of Parameters	**Definition**
Number of filters	• Integer number representing number of neurons connecting convolution layer to the input layer. • Turn out to be number of channels in output feature maps
Filter size	Height and width of filter • Scalar if filter have similar height and width • Vector if different height and weight
Padding size	Determine input borders horizontally and vertically
Stride size	Transverse the input vertically or horizontally according to the step size.
Number of channel	Feature maps of each filter. Depending on number of filters of the convolution layer. Number of channel for grayscale is 1 while number of channel for coloured images are 3

$$C^{ab} = \sum_{c=1}^{A^{b-1}} (W^{ab} * F^{c(b-1)}) \qquad (2.1)$$

Equation 2.1 shows convolution of weighing matrix and the previous feature map where C^{ab} = Results of layer b of the convolution with a-th kernel, W^{ab} = Weighing matrix, $*$ indicates usual convolution product, A^{b-1} = Number of kernel in the previous layer and $F^{c(b-1)}$ = final feature map produced by the previous (Couchot *et al.*, 2016).

For convolution layer, Figure 2.4 is the illusion by Deshpande (2016) for better understanding of how convolution layer and its parameters resulted with different outcomes.

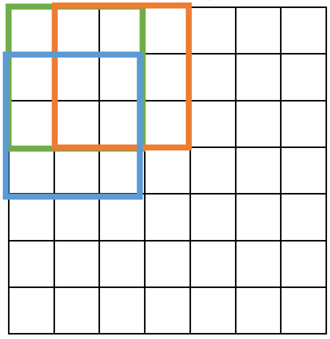
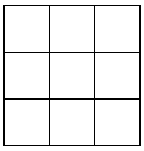
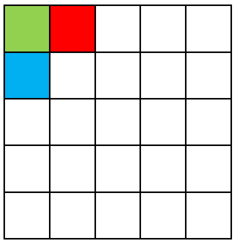
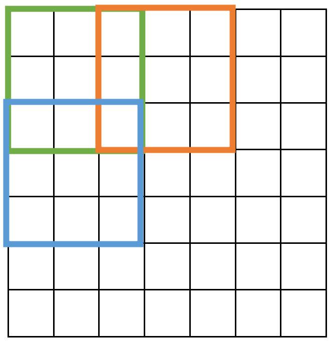
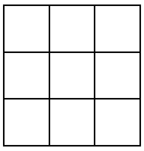
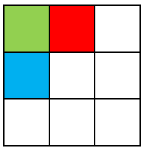
Input Size	**Filter size**	**Output Size**
7×7×1 with Stride 1 	3×3×1 	5×5×1 
7×7×1 with Stride 2 	3×3×1 	3×3×1 

Figure 2.4: Stride Illustration

From Figure 2.4, it can be seen that by using stride, the location of filter is moved according to its size and reduce the output feature map size. For large dataset, by using stride, there would be reduced numbers of overfitting as there are only certain feature maps are considered and not taking the whole feature maps into consideration. However, huge number of stride is not recommended as this would reduce the number of features studied and at the same time effect the accuracy of the network.

Stride and padding also controls the size of the output of convolution. The following formula shows the relation between the dimension of the feature maps, weight, padding and stride.

$$dim(\, C^{ab}) = \frac{dim(\, F^{c(b-1)}) - dim\ (W^{ab}) + 2 \times P}{S+1} \tag{2.2}$$

Equation 2.2 shows formula to calculate dimension of convolution where $\dim(\, C^{ab})$ = dimension of convolution between b with a-th kernel, $\dim(\, F^{c(b-1)})$ = dimension of previous feature map, $dim\ (W^{ab})$ = dimension of weighing matrix, P = Padding Size and S = Stride (Couchot *et al.*, 2016).

The second part in convolution layer is non-linearity activation function. In this project, Rectified Linear Units (ReLU) non-linearity are defined by $\mathrm{f(x)} = \max(0, \mathrm{x})$ are used as it has the ability for rapid convergence for huge networks trained with large dataset (Krizhevsky *et al.*, 2012; Chen *et al.*, 2015). Nonlinearity regularization are introduced in Krizhevsky *et al.* (2012) and had shown that by using ReLUs non-linearity makes CNN trains faster compared to saturating non-linearity

Feature maps that have been obtained from this convolution layer is considered features extracted from the input images. In this layer, features extraction is done and the output of this layer will become input to pooling layer.

2.3.1 (c) Pooling Layer

After the completion of convolution layer, features extracted from the input layer are ready for classification. However, as there are many features extracted from the previous layer, overfitting is prone to happen where the network might classify the images wrongly. Besides that, by having too many features representation to be learnt is computationally expensive. Therefore, pooling layer is introduced to reduce this mistake and at the same time increase the performance of the network (Bayar & Stamm, 2016).

There are two types of pooling either average pooling or max pooling. Max pooling will take highest value among the neighbourhood while average pooling will take the mean value of the overall neighbourhood. Max value is suitable to be used for layers that highly unlikely to meet high value while average value is more suitable for stable layer with almost the same neighbourhood value (Guo *et al.*, 2016).

In this study, max pooling have been used where only maximum number of feature representation is chosen to represent the selected region. The other non-selected pixels are converted into higher level features. Table 2.3 are the input arguments for max pooling layer: