IMAGE SPLICING DETECTION WITH CONSTRAINED CONVOLUTIONAL NEURAL NETWORK

LEE YANG YANG

UNIVERSITI SAINS MALAYSIA

2019

IMAGE SPLICING DETECTION WITH CONSTRAINED CONVOLUTIONAL NEURAL NETWORK

by

LEE YANG YANG

**Thesis submitted in fulfilment of the
requirements for the degree of
Master of Science**

August 2019

ACKNOWLEDGEMENT

The thesis could not be completed without the encouragement, advice, and support of advisor, family, and friends. The author is grateful to have their blessing along the way. The research could never be able to be completed without them.

First of all, the author would like to thank Prof Madya Dr. Khoo Bee Ee. Thank you for being encouraging and supportive along the way. Even facing the new challenges and new knowledge, she keeps learning while guiding so that she can be helpful to her students. The research could have not been accomplished without her dedicated guidance.

To the Faculty of Electrical and Electronics Engineering, USM, the author would like to appreciate for providing the opportunity to be one of the students in this remarkable school. The author had a great time thanks to the hard work of lecturers and staffs.

Special thanks to the colleagues, especially Mr. Tan Chee Sheng, Mr. Anthony Uwaechia and Ms. Norsalina bt Hassan. They had been very supportive and keep helping one and another even they are of totally different field of studies and researches. Also thanks to other colleagues that helped the author along the way.

To the family members, the author could never say thank you enough for having him back during his most vulnerable. Special thanks to the elder brother for the technical advice and support. For the time, money and energy that have been given, the author could not ask for more but to be grateful to have their blessing.

To everyone that has been involved directly and indirectly in the completion of this study, the author would like to say thank you for your contribution and help that can never be forgotten.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
API	Application Programming Interface
BMP	Bitmap
CNN	Convolutional Neural Network
CPU	Central Processing Unit
CU	Compute Unit
CUDA	Compute Unified Device Architecture
DCT	Discrete Cosine Transform
DL	Deep Learning
DOF	Depth of Field
DWT	Discrete Wavelet Transform
ELU	Exponential Linear Unit
FC	Fully Connected
FCN	Fully Connected Neurons
FFT	Fast Fourier Transform
FN	False Negative
FOV	Field of View
FP	False Positive
GA	Genetic Algorithm
GB	GigaByte
GPU	Graphics Processing Unit
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
IoT	Internet of Things
JPEG	Joint Photographic Expert Group
k-NN	k-Nearest Neighbour
LBP	Local Binary Pattern
LSB	Least Significant Bit
MCC	Matthews Correlation Coefficient
ML	Machine Learning
MVS	Majority Vote System

NAS	Network Access Storage
NN	Neural Network
PCA	Principle Component Analysis
PCIe	Peripheral Component Interconnect Express
PNG	Portable Network Graphics
RAM	Random Access Memory
ReLU	Rectified Linear Unit
RGB	Red Green Blue
ROM	Read Only Memory
SGD	Stochastic Gradient Descent
SRDA	Spectral Regression Discriminant Analysis
SRM	Spatial Rich Model
SSD	Solid State Drive
SSIM	Structural Similarity Index
SVM	Support Vector Machine
TanH	Hyperbolic Tangent
TIFF	Tagged Image File Format
TN	True Negative
TP	True Positive
YCbCr	Luminance; Chroma: Blue; Chroma: Red
a.k.a	also known as
etc	et cetera (and so on)
i.e.	id est (that is)
sq.px.	square pixels

PENGESANAN IMEJ PENYAMBATAN DENGAN RANGKAIAN SARAF KONVOLUSI TERKEKANG

ABSTRAK

Satu cara pengesanan pemalsuan imej yang diperbaiki, khususnya pengesanan imej penyambatan dengan Rangkaian Saraf Konvolusi (CNN) Terkekang telah dicadangkan dalam kajian ini. Imej penyambatan adalah satu cara pemalsuan imej yang biasa dan sentiasa disalahgunakan untuk motif yang jahat seperti propaganda idea palsu. Pada masa ini ada banyak usaha berkaitan dengan pengesahan imej tersambat, tetapi kebanyakannya adalah sama ada algoritma yang rumit ataupun yang tertentu pada ciri-ciri imej. CNN terkekang pada asasnya adalah model CNN Pembelajaran Dalam dengan pemberat-pemberat di lapisan pertama dikekang supaya ia hanya menyari ciri-ciri manipulasi sambatan, bukannya ciri-ciri objek. Lapisan terkekang ini membolehkan model CNN untuk belajar ciri-ciri yang diperlukan secara langsung dari imej biasa dan terus melaksanakan pengkelasan. Dalam kajian ini, set data sumber terbuka, iaitu set data imej sambatan CASIA, CASIA 2, CUISDE, NIST dan Carvalho telah digunakan untuk latihan dan penandaarasan model CNN yang dicadangkan. Dengan data imej disiap sediakan, model CNN akan mempunyai siri uji kaji untuk menguji pelbagai parameter dan juga menyiasat faktor bukan parametrik seperti peratusan kawasan sambatan daripada data sendiri. Kemudian, hiperparameternya akan ditala untuk pengoptimuman. Dengan model CNN yang telah dilatih dan ditala, ujian klasifikasi secara pangkalan data silang akan dijalankan. Hasilnya menunjukkan bahawa CNN boleh mengklasifikasikan kumpulan gambar dalam 96.31% ketepatan dan 96.3% F1-Score, tetapi skor ini hanya sesuai untuk set data CASIA 2. CNN yang dioptimumkan terbukti berat sebelah kepada set data latihannya sendiri. Oleh itu, CNN ini telah dilatih semula dengan set data seimbang yang disatukan. Dengan sedikit pelarasan pada CNN yang dicadangkan, ia dapat menghasilkan keseluruhan prestasi dengan ketepatan tertinggi 94.3% dalam set data Carvalho dan minimum 75.56% dalam set data CASIA. Kemudian CNN yang dicadangkan telah diguna semula untuk operasi penyetempatan sambatan secara blok. CNN ini berfungsi dengan baik dalam penyetempatan sambatan pada ketepatan yang tinggi dalam set data Carvalho dengan markah MCC 0.3582. CNN ini boleh mendiskriminasi sempadan sahih dan palsu pada pelbagai imej dalam pangkalan data silang. Ia menunjukkan bahawa CNN dengan algoritma konvolusi terkekang boleh digunakan untuk pengesanan imej penyambatan secara umum.

IMAGE SPLICING DETECTION WITH CONSTRAINED CONVOLUTIONAL NEURAL NETWORK

ABSTRACT

An improved approach of image forgery detection, specifically image splicing detection with Constrained Convolutional Neural Network (CNN) is proposed in this research. Image splicing is a common method in image forgery and is often being misused for bad motives such as false idea propaganda. Nowadays there are many related efforts in detecting spliced images, but most of them are either feature-specific or complicated algorithms. Constrained CNN is basically a Deep Learning CNN model with its first layer weights being constrained so that it only extracts splicing manipulation features instead of object features. The constrained layer enables the CNN model to learn the required features directly from ubiquitous image input and then performs classification. In this research, the open source datasets, i.e. CASIA, CASIA 2, CUISDE, NIST, and Carvalho image splicing datasets were used for training and benchmarking the proposed CNN model. With the datasets prepared and assembled, the proposed CNN model will have a series of experiments to test for the various parameters as well as to investigate other non-parametric factors such as the data variation itself. Then its hyperparameters will be tuned for optimization. With the trained and tuned CNN model, a cross-database classification evaluation is carried out. The result shows that the CNN can classify image batch with 96.31% in accuracy and 96.3% in F1-Score, but the scores only apply to CASIA 2 dataset. An optimized CNN is shown to be biased to its own train dataset. Hence it is purposely retrained with a merged balanced dataset. With slight adjustment on the proposed CNN, it is found to be able to generalize the overall performance with the highest accuracy of 94.3% in

Carvalho dataset and minimum 75.56% in CASIA dataset. Then the proposed CNN is recasted for block-wise splicing localization operation. It performed well in splicing localization at high accuracy in Carvalho dataset with MCC mark of 0.3582. It is able to discriminate the authentic and splicing border in a wide range of images in the cross-database test. It is shown that CNN with constrained convolution algorithm can be used as a general image splicing detection task.

CHAPTER 1
INTRODUCTION

1.1 Overview

Nowadays a lot of information especially videos and pictures are stored in digital format. With higher accessibility to performance computation and digital storage by average consumers, digital information, however, can be digitally tampered with ease. Image splicing is one of the common image forgery techniques used with intention of covering or adding information in a picture by combining two or more image components into a single image (Zhang *et al.*, 2010) as shown in Figure 1.1. Simple cut-and-paste spliced images may be identifiable with visual inspection, but skilful users often use post-processing techniques such as blurring to cover up the splicing artefacts. Thus, the tampered image can sometimes be indistinguishable from the authentic image by human eyes.

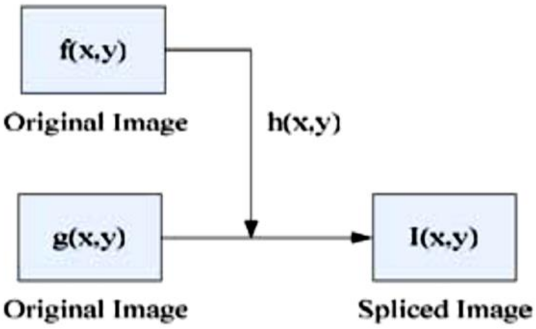

Figure 1.1 Process of image splicing (Zhang *et al.*, 2010)

Tampered images may seem as entertaining works in consumer space, but in the meantime, it also caused a serious threat in trustworthy and security of digital assets. For example, the originality of the non-watermarked digital sources such as uploaded videos and images are often questionable in police cases. To make matter worse, since the internet is more accessible than before with smartphones, new internet users especially the inexperienced elderly cannot differentiate the source and originality of information. Thus spliced images are very often embedded in fake news and mistaken as real information, leading to more uncontrollable social security. As such, there must be a way to countermeasure the possible threats caused by image forgery. Image Forgery Detection is a technique specifically to study and identify tampered images. Such technique is useful especially in verifying the authenticity of digital images in many applications such as crime investigation (Parameswaran and Sugitha, 2016).

There are many techniques currently being studied to discriminate spliced images. In this research, Convolutional Neural Network (CNN) is of great interest since it closely resembles the mechanism of the human brain by using a neural network model. Deep Learning (DL) is one of the specialized fields in many Machine Learning (ML) algorithms. While there are many new DL algorithms nowadays, CNN represents the best perception of machine sensing towards real world and real-life images. CNN is one of the most notable DL approaches for image classification and segmentation tasks. CNN can be trained with raw data and achieve robust performance (Guo *et al.*, 2016). However, its usefulness is still widely unexplored, especially in image forgery detection.

1.2 Motivation

It is undeniable that DL is gaining interest in both academic and industrial sectors. More applications nowadays are getting more dependent on computers such as auto-drive and data mining. The emergence of CNN allows users to process images without traditional hard-coded algorithms. With DL algorithm like CNN, it can “code” by itself and keep improving over time. With spliced image getting more complicated and people are more skillful in editing digital contents, hard-coded algorithms will be driven to obsolete. CNN, on the other hand, can keep learning from new data and improves over time, minimizing effort in developing specific programs such as splicing forgery detection.

1.3 Problem Statements

Simple direct cut-and-paste spliced images can be detected by looking for abrupt changes of object's edges or examining high-frequency components by using high-pass filters. Skilled forger often post-processed area of interest with various techniques such as median filtering, background blurring, and colour correction so that they looked natural by most people, making image splicing classifiaction more challenging, not to mention attempting spliced region localization.

Support Vector Machine (SVM) is good at feature classification method, but it requires more specific data input structure or variables such as preprocessed feature maps rather than direct image feeding, thus the process of feature extraction is needed for better classification performance (Cao and Chong, 2002). However, there also exist many options for feature extraction such as Principle Component Analysis (PCA) and domain transformation, resulting in more uncertain performance output and difficult to be tuned for general optimization. Another challenge of ML is that training on certain datasets only will render it database specific, making it underperformed in cross-database test (He *et al.*, 2012a).

The recent improvement in the computational economy had motivated researchers to shift focus from hand-crafted feature extraction approach to data-driven approach. Many researchers are looking into the DL model, more specifically CNN. Some researches did find that with the right conditions and customization, CNN can outperform the previous non-DL approaches. However, the DL model may get very complicated, considering that their hyperparameters are already hard to deal with. It is still unknown how much layer is considered adequate or optimum for such a specific task.

Although constrained convolution is proven to be superior in manipulation feature detection such as Gaussian Blurring and Median filtering (Bayar and Stamm, 2018), this novel pipeline is still not yet fully proven to be viable for generalizing image splicing detection. So far, there is no single research on whether or not a simple end-to-end CNN with slight modification with constrained convolution algorithm can do well in image splicing detection, i.e. splicing classification and splicing localization. Albeit there are DL algorithms developed specifically for splicing classification and splicing localization, none of them are being used for both classification and localization with the same DL models.

1.4 Objectives

With the problems of the research identified, the research objectives can be summarized as follows:

1). To develop a Constrained CNN for image splicing detection.

2). To classify spliced image and to localize spliced regions with the Constrained CNN developed from Objective 1.

3). To evaluate and generalize the performance of image splicing detection of the Constrained CNN.

1.5 Scope of Research

This research focuses solely on image splicing detection only. There are many types of image manipulation techniques, but only splicing techniques will be tackled to demonstrate the proof of concept. Images of other manipulation like added Gaussian noise or double compressed will not be in the scope. DL approach will be used as the area of focus in this research. More specifically, Constrained CNN will be investigated to perform image splicing detection. A higher level of classification like determining the authenticity of a whole full resolution image is not considered. The Python programming language, specifically Keras Application Programming Interface (API) will be utilized throughout the research.

1.6 Outline

The thesis is organized into six main chapters. This first chapter is about a brief overview of the image splicing forgery while pointing out the current dilemma of spliced image detection and seeking a way to improve it.

In the next Chapter 2, previous works from many related types of research will be reviewed and categorized according to the nature of the researches. The methods of performance measurement with critical comparisons of various works will then be presented. At the end of this chapter, the theory of CNN will be discussed in details prior to the experiment so that to understand how to better deal with the variables.

Chapter 3 is all about describing the methodology of the research. The methodology of the CNN model optimization and data preparation will be described in fine-grain details. The modified CNN algorithm will be proposed and the way it is being designed will be explained. The process of evaluation will be delineated for a better understanding of the method of performance measurement.

The execution of research and its output will be presented in both Chapter 4 and Chapter 5. Chapter 4 is of preliminary experiments whereby various parameters will be tested and verified in a series of experiments to verify the functionality of the

CNN model as well as optimizing the CNN in the process. Chapter 5 is the core experiment whereby the optimized CNN will be tested along with other databases in both classification and localization task. There will have minor optimization in the process of generalizing the overall performance. The performance will be compared across other databases and to other related works.

Lastly, Chapter 6 will conclude the outcomes and contributions of the research. Based on the finding of this research, a few suggestions are given so that to facilitate future works.

CHAPTER 2
LITERATURE REVIEW

2.1 Overview

In this chapter, current image splicing detection techniques will be discussed. Most of the works require the following techniques to identify the spliced image and localize a spliced region. Before diving into methods of splicing detection, it is necessary to review the history of image splicing and the current state of research in Section 2.2. The common properties of spliced images will be discussed in Section 2.3, as well as methods of feature extraction in Section 2.4. The methods of performance measurement metrics will be discussed in Section 2.5. Lastly in Section 2.6, the theory of CNN will be described in depth to understand the major parameters of CNN algorithm before proceeding to the next chapter.

2.2 Image Forgery and Image Splicing

The earliest forged image is believed to be dated back in 1840, where Hippolyte produced a fake image of himself committing suicide as frustration for not getting proper recognition as shown in Figure 2.1. That is when a new photographic process is discovered for manipulating images. The same process had also been introduced independently by Daguerre and Talbot in 1839 (Lester, 2015). Fast forward to February 2004, a photo showed that US Senator John Kerry and actress Jane Fonda sharing a stage at an anti-war rally during the Presidential primaries as Senator Kerry was campaigning for the Democratic nomination, is found to be spliced in Figure 2.2 (Qureshi and Deriche, 2015). Also in 2010, a Malaysian politician Jeffrey Wong Su En claimed to have been knighted by Queen Elizabeth II as recognition for his contribution to the international aid organization Médecins Sans Frontières. A picture of him being awarded by the Queen of England accompanied his statement, diffused in local media. However, the British High Commission in Kuala Lumpur made clear that his name was not included in the official knighthood recipients lists, and that the picture was not consistent with the usual protocol for knighthood ceremonies. The image was finally shown to be spliced between an original ceremony photo and Mr. Wong's face as shown in Figure 2.3 (Redi *et al.*, 2011).

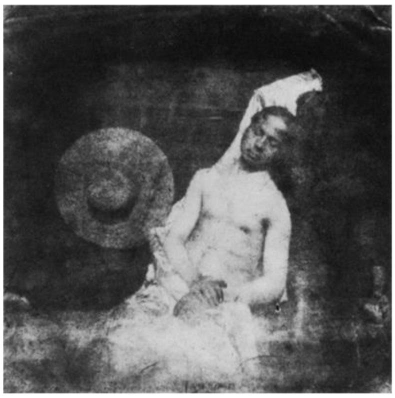

Figure 2.1 First fake photograph of Hippolyte Bayard committing suicide (Lester, 2015)

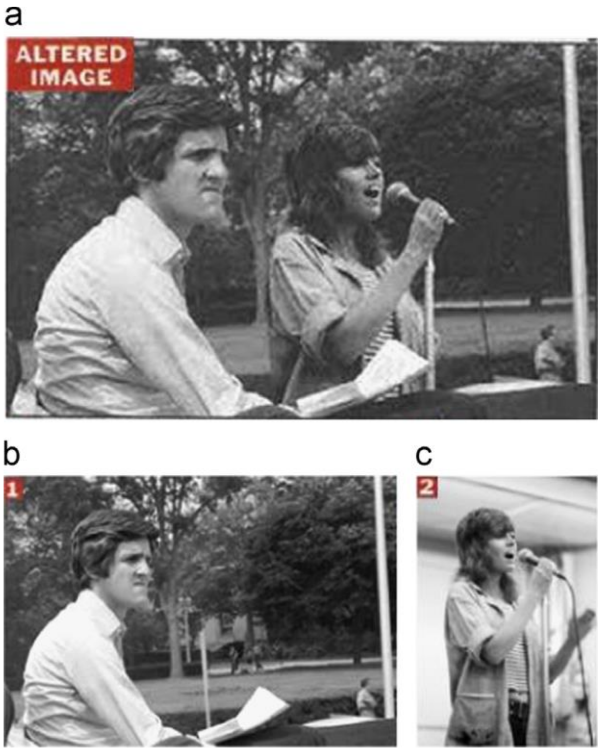

Figure 2.2 (a)Tampered image of Ex-U.S. presidential election candidate John Kerry and actress Jane Fonda. (b) Original image of John Kerry prepares to speak about war in Vietnam. (c) Original image of actress Jane Fonda speaks to a group of Vietnam veterans (Qureshi and Deriche, 2015)

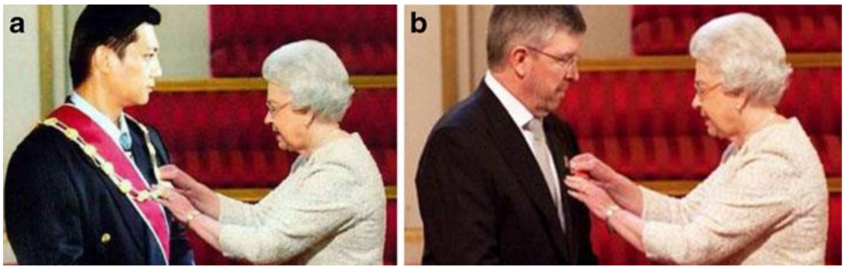

Figure 2.3 (a)The tampered image depicting Jeffrey Wong Su En while receiving the award from Queen Elizabeth II, published in Malaysian dailies. (b) The original picture of Ross Brawn receiving the Order of the British Empire from the Queen (Redi *et al.*, 2011)

Those are only a few examples of how the image splicing technique had been exploited with ulterior motives. The use of digital images as visual evidence is becoming more questionable in recent days. Professional image splicing techniques are becoming more accessible to users under a few computer mouse clicks. In order to confront the potential threat to trustworthy of digital assets, image forgery detection is necessary to determine the authenticity of digital media especially images.

Splicing is a subset of many forgery techniques such as copy-move and retouching. It is difficult to tell whether or not the image has been tampered with splicing or other tampering techniques without a ground reference image in real-world cases. Fortunately, there are a few approaches to assess the features that might be deterministic to the nature of spliced images as illustrated in Figure 2.4.

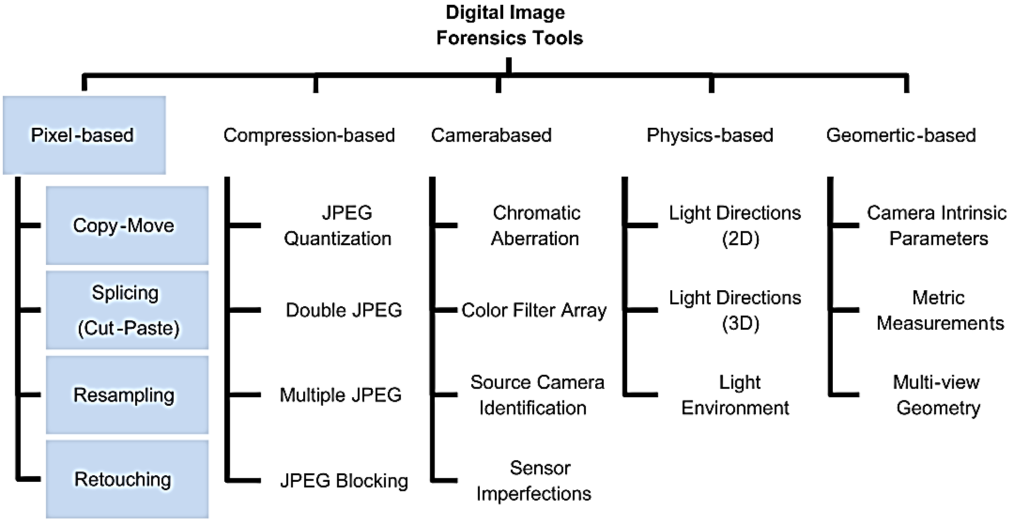

Figure 2.4 Image splicing is usually pixel-based. However, there are other types of features available to access the forged image (Qureshi and Deriche, 2015)

2.2.1 The Similarity with Image Steganalysis

Image steganography is a process where digital information is hidden in images. Frequently, images are tampered in low levels, such as binary least significant bit (LSB) shifting or model based tampering to hide information in images. Conversely, image steganalysis is an attempt to reveal the nature of the stego images by searching for anomalies in images such as by using Spatial Rich Model (SRM) (Fridrich and Kodovsky, 2012) and histogram features extraction (Luo *et al.*, 2011).

The important clue is that if a stego image can be viewed as tampered images since splicing is one of kind of image tampering, the algorithm for steganalysis might work as well in image splicing detection with some algorithmic modifications and adaptation(Qiu *et al.*, 2014) as illustrated in Figure 2.5.

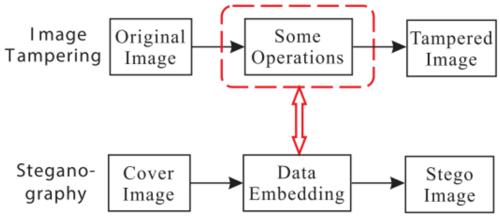

Figure 2.5 Illustration of the relationship between image tampering and image steganalysis (Qiu *et al.*, 2014)

Usually, the specific features of images are decoded or extracted before making further data processing and classification. The algorithm can be manually coded by targetting specific feature such as image noise, but there are many other possible usable features can be exploited such as edge detection and JPEG block artefact. As such, coding each algorithm for each feature extractor is time-consuming and the outcomes are not guaranteed. Hence a more effective self-learning feature or data-driven extraction technique is of huge interest. Deep Learning algorithm is one of the candidates and it can be modified to suit the specific needs such as image splicing detection.

2.2.2 The Rise of Deep Learning

ML is not new, it is an active research in computer science even before the existence of personal computers. Neural Network (NN) is the subset of ML other than Genetic Algorithm (GA), Support Vector Machine (SVM), etc. The NN is unique among other ideas as it resembles the mechanism of the biological neuron which is believed to contribute to the intelligence of the human brain. A simple NN only has a single layer design, but it can be designed to have many hidden layers that are capable of learning deeper or higher dimensional features and achieve better performances, thus it is named Deep Learning (DL). DL is a subset of ML which is just one part of the general AI field as illustrated in Figure 2.6. However, DL did not become a hot topic due to three major barriers.

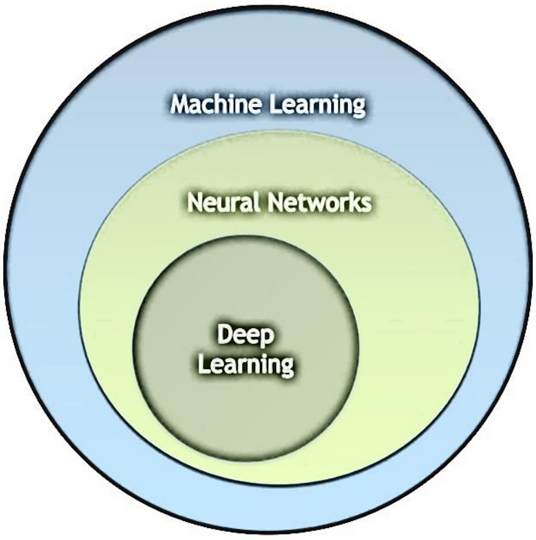

Figure 2.6 Relationship of Deep Learning to ML (Nardone, 2017)

2.2.2(a) Barrier 1: Right Algorithm

Firstly, in order to process multidimensional data such as videos and images, there needs the right algorithm to process those data. Simply connecting hundreds and thousands of neurons to every input pixel is not going to be practical considering the limited availability and accessibility of powerful computation. That makes the wall of the second barrier of computation which will be covered in the next section.

NN serves as the basis of all modern variation of Deep Learning algorithms. The evolution of NN and the first implementation of CNN started in 2012, where a group of researchers from the University of Toronto introduced CNN algorithm for ImageNet Large Scale Visual Recognition Challenge (ILSVRC) and the performance superseded all other algorithms, from classification error rate of over 20% down to 16% (Krizhevsky *et al.*, 2012). This is when the ANN, specifically CNN comes into play. The convolution algorithm is able to reduce the dimensions required for the computer to process and also extract higher dimensional features to improve accuracy. The tradeoff is the additional algorithm that convolutes image pixels demands more computation power and computes parallelism than before.

2.2.2(b) Barrier 2: Computational Viability

Central Processing Unit (CPU) is versatile in many types of digital operations, but not optimized for highly paralleled computation workload required by NN model optimization. The repetitive multiplication and addition of parameters simply cannot be parallelized well for CPU. Developing an ANN model with many neurons will only induce more computational cost and time. After the finding of CNN algorithm, people soon found that Graphics Processing Unit (GPU) can be exploited for general computation. GPU is a computer hardware initially designed to accelerate video game framerates, but newer games demand more realistic physical and visual effects. Thus, GPU has built-in small but highly paralleled Compute Unit (CU) (or Compute Unified Device Architecture (CUDA) core in Nvidia’s GPU) for computation workload. It is especially good in parallel processing which fits the need of convolution algorithm and neurons optimization, driving down the computational cost and time significantly.

It promotes more people to be actively involved in DL researches and also makes DL becomes practical in industries and in daily uses as it speeds up model inferencing. In addition to the development of more efficient inferencing algorithm like YOLO (Redmon *et al.*, 2015), real-time processing is made possible, opening up new industries such as autonomous vehicles.

2.2.2(c) Barrier 3: Availability of Big Data

Thirdly is the availability of data. Because CNN is data-driven machine intelligence in nature, it requires a lot of data for model optimization. It simply was not the right time for it to be practical decade ago since the Internet resources were not as ubiquitous as today. Nowadays many online services come to play such as social media, online shopping, exchange market, and cloud storage. Moreover, the Internet of Things (IoT) is strived to be the next Internet revolution (Kundhavai and Sridevi, 2016). Internet resources will only grow over time as more people use the internet and consume more digital data. Hence Big Data is readily available to be collected and will only become more accessible in the future. As a result, the year 2012 is also known as “Big Bang of Deep Learning” (Nardone, 2017).

In case of spliced image data, it is very fortunate that those data are also only readily available in recent years such as CASIA, CUISDE, and NIST Spliced Image datasets, thanks to the many other research efforts around the world.

2.3 Common Properties of Spliced Images:

Common properties of spliced images can be categorized into five major groups. Those properties serve as important footprint and distinctive features of spliced images relative to authentic images.

2.3.1 Camera-based Features

Digital images are not perfectly colour balanced, they have their own distinctive background noise that has been generated during the data acquisition stage from camera sensors. The distribution of image noise is unique across different cameras and sensors. If an image is cut and pasted into another image of different sources, their inherent noise distribution will be different from global noise, which makes it an important footprint for image splicing detection. Local noise variance estimation like one in Figure 2.7 is one of the techniques used to detect noise delta and localize the spliced region (Zhan and Yuesheng, 2015).

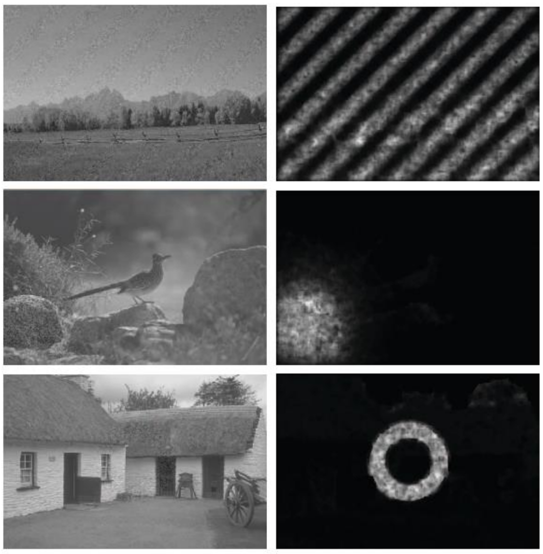

Figure 2.7 Local noise variance estimation (right) of corrupted images (left) (Zhan and Yuesheng, 2015)

Camera lenses are not perfect, so do the images taken by digital cameras. Very often, with a poor quality lens, chromatic aberration occurs at pixels that are further from the centre of images. Chromatic aberration is a phenomenon where blue light and red light are out of synchronization due to the nature of lens refraction. By identifying and magnifying the global vector of the aberration, the originality of an image can be determined (Johnson and Farid, 2006).

2.3.2 Compression-based Features

There are many digital formats to store images. Those formats can be categorized into two types, lossless and lossy compression format. Lossless format like Portable Network Graphics (PNG) and Bitmap (BMP) stores image as it is without digitally manipulating it, thus retaining the image quality. However, it comes at the cost of bigger file size which is not an option when it comes to online data transfer where data bandwidth is of great bottlenecking. If data bandwidth and digital storage space are the limiting factors, lossy formats like Joint Photographics Expert Group (JPEG) comes in handy to reduce the file size while retaining most of the visual details of images. JPEG format discards high-frequency elements so that to reduce the file size with virtually no perceptual loss of image quality. As a result, the JPEG format is the most widely used image compression format. However, it leaves traceable JPEG compression artefacts because JPEG uses the Discrete Cosine Transform (DCT) quantization and blocking to encode or decode an image. Hence an image of a spliced region can be traced by studying the global blockiness of JPEG image, especially double compressed JPEG where blockiness is more evident (Barni *et al.*, 2017). However such method is only limited to images of JPEG format. Although there are better compression format like JPEG 2000 format which uses wavelet transform instead of DCT, it too can be analysed from statistical difference to detect double compressed JPEG 2000 images (Zhang *et al.*, 2008).

2.3.3 Physics-based Features

Spliced image region also tends to have a different level of illumination than the global value. Hence by estimating the global illumination of an image, splice region can be revealed (Fan *et al.*, 2015). Speaking of illumination, the direction of global illumination can be an important footprint for spliced image detection as well since authentic images can have only illumination source of one directionality in most cases (Carvalho *et al.*, 2013).

Motion blur and out-of-focus blur are also common in images. Motion blur occurs when an image is taken on moving objects or the camera itself moves. Motion blur is mostly dependent on the camera shuttler's speed and the object's speed. Out-of-focus blur is due to optics where objects will be blurred when they are not in the depth-of-field (DOF) of the lenses. However, spliced components do not follow the global blurred types (Bahrami *et al.*, 2013) as one in Figure 2.8. A spliced region may have different types of blur stacking one another such as Gaussian blur in the out-of-focus blurred region. Therefore, spliced images can be discriminated by studying the consistency of blur types within an image (Binnar and Mane, 2016).

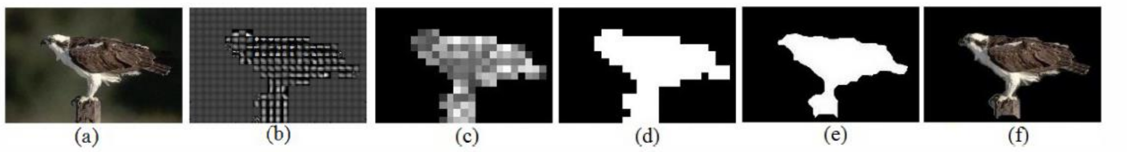

Figure 2.8 Image splicing detection: (a) Input image (b) Blur kernel map (c) blur map (d) binary map (e) Segmentation map (f) segmented output (Bahrami *et al.*, 2013)

2.3.4 Domain-based Features

Digital image is of at least two-dimensional data in grayscale, or three-dimensions in case of usual RGB images. Those data can be translated into other domain representation such as chrominance space and statistical domain.

In chrominance space, RGB (Red, Green, Blue) image can be converted to YCbCr (Luminance; Chroma: Blue; Chroma: Red) data format, which is still being used in television transmission nowadays. The colour space conversion is as simple as the following equation (Zhao *et al.*, 2011):

$$\begin{pmatrix} Y \\ B-Y \\ R-Y \end{pmatrix} = \begin{pmatrix} 0.299 & 0.587 & 0.117 \\ -0.299 & -0.587 & 0.886 \\ 0.701 & -0.587 & -0.114 \end{pmatrix} \cdot \begin{pmatrix} R \\ G \\ B \end{pmatrix} \tag{2.1}$$

One study found that simple image pre-processing methods such as converting the RGB colour space to YCbCr colour space can improve detection rate due to the difference in chrominance of different source of images (Zhao *et al.*, 2011).

Images can also be transformed into the frequency domain with Fast Fourier Transform (FFT). It converts the binary value into horizontal and vertical frequency component representation. FFT image is useful especially in editing any repetitive pattern or texture in images. Tampered images will have distinctive features in FFT, especially in high-frequency components. Hence FFT can be used as a preprocessing before feature extraction (Bunk *et al.*, 2017).

2.3.5 Geometry-based Features

Usually, camera lenses have a specific angular Field of View (FOV). Hence, even though an object is perfectly flat, it will seem curved from the centre toward the edge of images. This effect is called radial distortion and is getting more obvious in images taken with a wide-angle lens or fish-eye lens as illustrated in Figure 2.9. Spliced components usually do not take this space-warping effect into account, making radial distortion a good ground reference for splicing detection (Chennamma and Rangarajan, 2011).

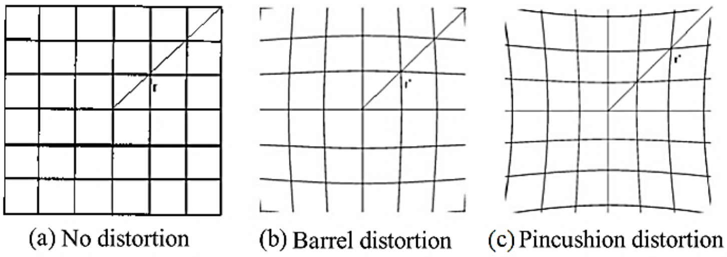

Figure 2.9 Phenomena of radial distortion, (a) no distortion, (b) Barrel distortion, (c) Pincushion distortion (Chennamma and Rangarajan, 2011)

2.4 Methods of Feature Extraction

The methods of feature extraction can be categorized into two approaches, targeted feature extraction and data-driven feature extraction despite the method of classification is more or less under the same subset of ML. The method that treats specific features as a target of input data is considered as targetted feature extraction. Targeted feature extraction uses specific algorithms to extract specific known features such as noise and edges. Statistical features such as histogram statistics (Stamm and Liu, 2010) are also considered as targetted features as the user needs to design the algorithm specifically to encode such feature.

Data-driven feature extraction, on the other hand, exploits big data to extract features without user's interference or guidance. The algorithm will learn to extract useful features in order to achieve a minimum cost function. The input image is usually in raw RGB format or in grayscale. Those algorithms may result in extracting features that are not understandable by user albeit the output can be of high accuracy.

2.4.1 Targeted Feature Extraction

Targetted features extraction techniques are widely used in the early image splicing detection approach. One of the early work exploits YCbCr colour space to extract distinct features with domain transformations such as Discrete Wavelet Transform (DWT) (Zhu and Zhen, 2012; Hakimi *et al.*, 2015) and DCT (Alahmadi *et al.*, 2013; Zhang *et al.*, 2016) were prevalent to extract the pattern of edges of objects in test images. Alternatively, image processing techniques such as Sobel edge detection (He *et al.*, 2011; Liu *et al.*, 2013) and Local Binary Pattern (LBP) (Alahmadi *et al.*, 2013; Hakimi *et al.*, 2015; Agarwal and Chand, 2016) were also being used like one in Figure 2.10. Other image properties such as YCbCr colour space conversion

(Hakimi *et al.*, 2015; Manu and Mehtre, 2015) as shown in Figure 2.11 and image histogram (Cozzolino *et al.*, 2015; Manu and Mehtre, 2015; Vaishnavi, 2016) are also found to be helpful and are being exploited to reveal hidden information of test images due to the nature of different chrominance and histogram entropy of spliced images.

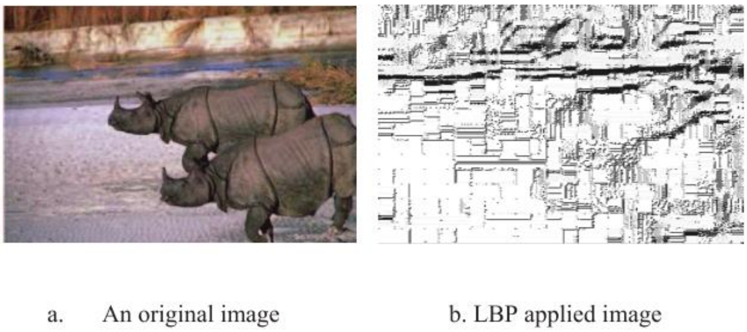
a. An original image b. LBP applied image

Figure 2.10 LBP applied image (Hakimi *et al.*, 2015).

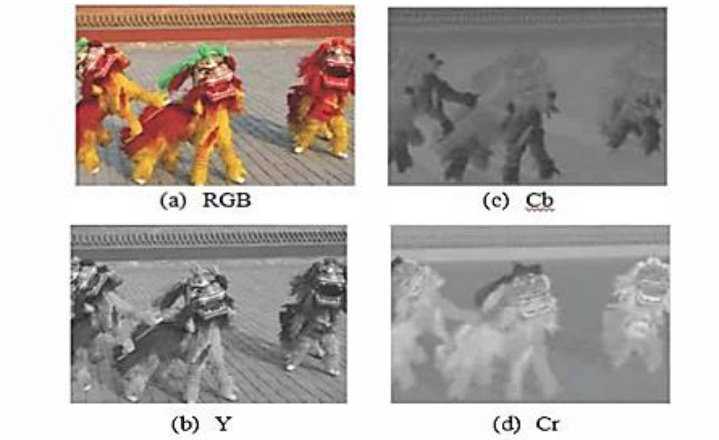

Figure 2.11 YCbCr colour space conversion from RGB colour space (Alahmadi *et al.*, 2013)

Other than features extraction, studying the quality of an image can also reveal the nature of spliced images. Non-reference image quality metrics such as blocking feature and zero-crossing were computed, then the outputs were combined to form image mapping to highlight the possible area of image tampering (Battisti *et al.*, 2012; Manu and Mehtre, 2015). Image splicing can also be detected by finding an abnormality in physical properties as mentioned in Section 2.3.3. Very often the blur degree of spliced region differs from the blur degree of its neighbouring, thus the blur degree and depth information of an image can be estimated and being used as references for image splicing (Bahrami *et al.*, 2013). Blur type inconsistency such as out-of-focus blur and motion blur are also being used to test subject images (Binnar and Mane, 2016). Properties of lighting are also a good source of evidence within a spliced image. Spliced images are most likely have inconsistent global illuminance, hence local illumination is estimated to reveal the spliced region (Fan *et al.*, 2015).

Intrinsic fingerprints of camera lens such as lens radial distortion (Chennamma and Rangarajan, 2011) and chromatic aberration (Johnson and Farid, 2006) are also some good source of references for image splicing detection. The true nature of photo images will be more evident by amplifying the red and blue pixels aberration vector amplitude like one in Figure 2.12. Nevertheless, most of classification processes still require traditional ML algorithms, mostly SVM (He *et al.*, 2012; Alahmadi *et al.*, 2013; Vaishnavi, 2016), k-nearest neighbour (k-NN) (Vaishnavi, 2016) and Spectral Regression Discriminant Analysis (SRDA) (Agarwal and Chand, 2016) classifiers to decide the nature of test images.

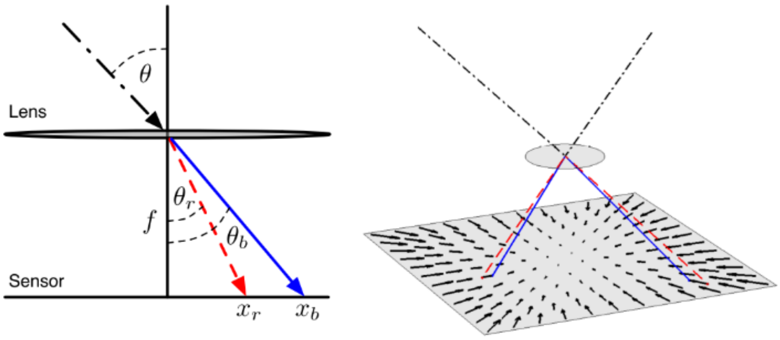

Figure 2.12 Chromatic aberration due to lens refraction (Chennamma and Rangarajan, 2011)

2.4.2 Data-driven Feature Extraction

Recent researches appeared to be focusing on Deep Learning feature extraction, more specifically CNN (Chen *et al.*, 2015) which can automatically learn and obtain features directly from two-dimensional data and beyond, such as images and videos. CNN is one of the Deep Learning algorithms that use multiple hidden layers of neural networks to achieve better performance in specific tasks but require more computational workhorse as well as larger datasets in the training stage. CNN, as shown in Figure 2.13, served as the basis of other popular variations such as AlexNet and Visual Geometry Group Net (VGGNet). CNN is superior in image classification task but it tends to extract features associated with image content only, hence CNN is ill-suited for directly extracting image manipulation features in many image forgeries such as image splicing (Bayar and Stamm, 2016). Many researchers had been trying to re-engineer the DL algorithms in the hope that it can be beneficial for image forgery detection.

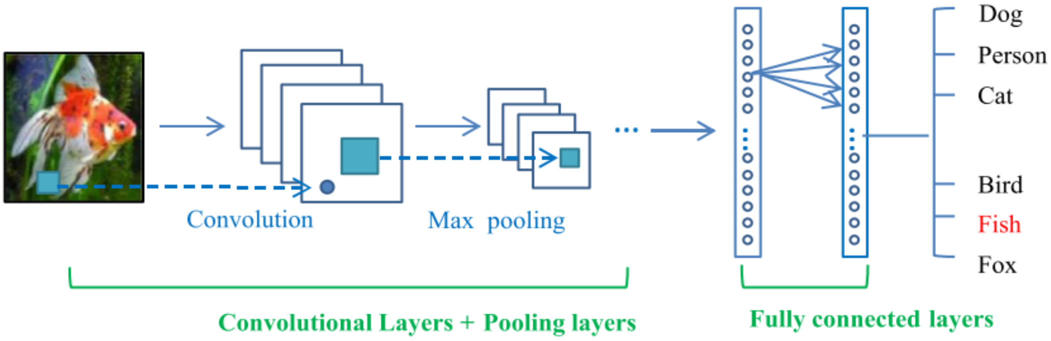

Figure 2.13 Pipeline of general CNN network *(Guo et al.*, 2016)

Image segmentation is gaining more attention recently. Unlike traditional CNN, it can retain the spatial information of the image features after convolution, i.e. the location and orientation of the object of interest in image space. In Salloum *et al.* (2018) works, they use VGGNet-like structure. Instead of ending with fully connected (FC) layer, they upsampled the feature maps and convolved them. The process was repeated until it matched the CNN input size so that to directly localize the spliced area as shown in Figure 2.14, in which they called the model Multitask Fully Convolutional Network (MFCN). Hence the localization can achieve pixel-wise localization instead of traditional block-wise detection. From Rao and Ni (2016), they made use of Spatial Rich Model (SRM) developed from steganalysis (Fridrich and Kodovsky, 2012) to strategically initialize their first layer weights so that to avoid object feature extraction at the initial stage of the training epoch.

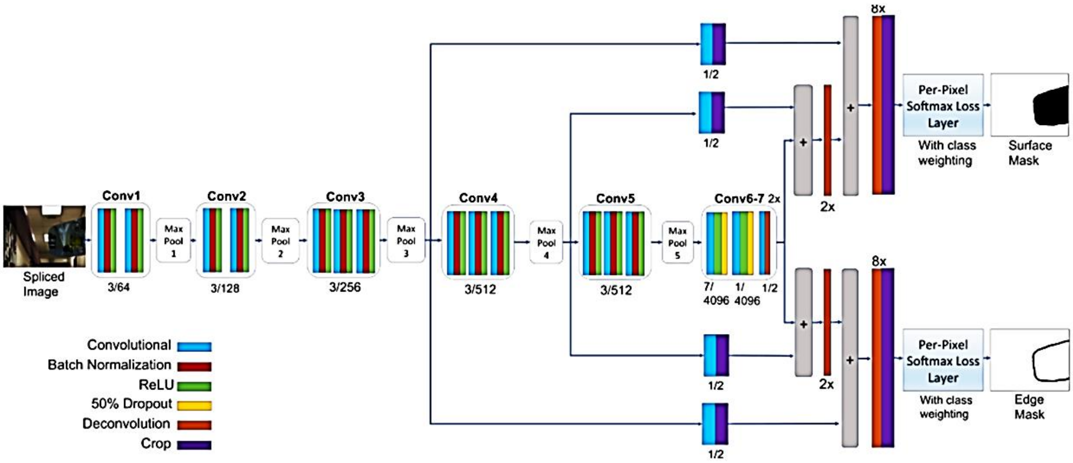

Figure 2.14 Multitask Fully Convolutional Network (MFCN) (Salloum *et al.*, 2018)

There are few researches where DL is used partially for feature extraction but the classification task is entrusted to more traditional SVM like what Zhang *et al.* (2016) did. They used a stacked autoencoder with Multi Layer Perceptron (MLP) in Figure 2.15 which is one of the DL algorithms to extract feature data. Then they used the extracted features to train SVM model, albeit the CNN nowadays can extract and classify directly in one process.

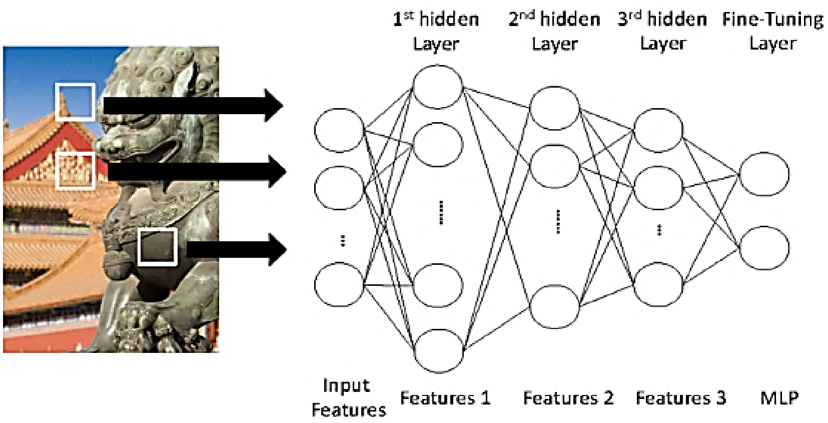

Figure 2.15 Stacked autoencoder algorithm (Zhang *et al.*, 2016)

Other than CNN algorithm, autoencoder splicing localization approach has been used by Cozzolino and Verdoliva (2017) to achieve pixel-wise localization but uses SRM for feature extraction instead. They use the binary mask as the label for autoencoder and trained it with the features extracted from their SRM algorithm as shown in Figure 2.16. Carvalho *et al.* (2013) on the other hand used colour illumination estimator to extract features then uses ML to learn and make decisions on human faces splicing detection.

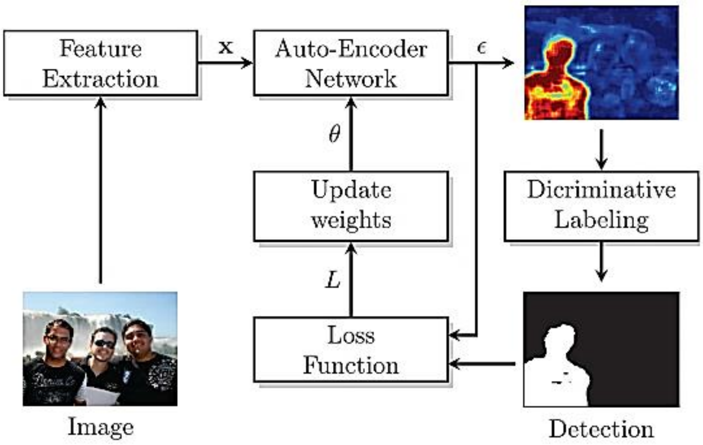

Figure 2.16 Fully autoencoder approach (Cozzolino and Verdoliva, 2017)

2.5 Methods of Performance Evaluation

There are two ways to evaluate the performance of the DL algorithm in spliced image detection, which are classification and spliced region localization. Classification is the operation that evaluates an image as a whole such that whether an image is authentic or spliced. Usually, the test image is resized or in cropped batch. Whilst localization is applied to localize the spliced region within a full-size image which is considerably more challenging than simple classification. Depending on algorithms or models, classification task can also be recasted into the localization task.

2.5.1 Available Datasets

Both tasks require data samples or datasets which are used by other researchers as the specimens of benchmarking. Table 2.1 shows the popular spliced image datasets available and of being cited frequently. These datasets are used in this research.

Table 2.1 Datasets available for image splicing detection works

Datasets	Description	No. of image

CASIA	The first version of CASIA 2, all of 128 pixel square images, pure authentic and spliced images. In grayscale only (Dong *et al.*, 2013).	Authentic = 800
		Spliced = 921

CASIA 2	Consists of copy-moved, removal, and spliced images. In RGB and of various sizes (Dong *et al.*, 2013).	Authentic = 7483 Spliced + others = 5122

CUISDE (a.k.a. DVMM)	Columbia Uncompressed Image Splicing Detection Evaluation, consists of high resolution authentic and spliced images with edge mask provided. In RGB (Hsu and Chang, 2006).	Authentic = 183
		Spliced = 180

NIST 2016	Nimble Challenge 2016, consists of various manipulation classes, including splicing image class (*NIST Media Forensics Challenge*, 2016).	Authentic = 874
		Spliced = 288

Carvalho	Novel dataset by Prof. Dr. Tiago J. Carvalho (Carvalho *et al.*, 2013).	Authentic = 100
		Spliced = 100

2.5.2 Splicing Classification

In image classification, an image will be determined as authentic or spliced. More specifically, since the goal of the CNN model is to detect image splicing, the authentic image will be framed as negative set (no splicing detected), and the spliced image will be the positive set (splicing detected). If the model can have the output classification as labelled, that will be a true response, either True Positive (TP) or True Negative (TN). If the output differs from the label, it will be a false response, either False Positive (FP) or False Negative (FN). Table 2.2 illustrates the discrimination between the measurement metric.