DETECTING TAMPERED REGIONS IN JPEG IMAGES VIA CNN

A project report submitted to

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY ANANTAPUR, ANANTAPURAMU

in partial fulfillment of the requirements
for the award of degree of

BACHELOR OF TECHNOLOGY
in
ELECTRONICS AND COMMUNICATION ENGINEERING

Submitted by

NAGIREDDY CHANDRA MOULI REDDY-17121A04F0
SHAIK THOUHEED AHAMED -17121A04K4
TALAPALA GANESH -17121A04L3
N V SAI VIVEK -17121A04E9

Under the Guidance of

Dr. N. Padmaja, M.Tech, Ph.D.,
Professor, Department of ECE

Department of Electronics and Communication Engineering

SREE VIDYANIKETHAN ENGINEERING COLLEGE
(AUTONOMOUS)
Sree Sainath Nagar, A. Rangampet, Tirupathi - 517102.

(2017-2021)

SREE VIDYANIKETHAN ENGINEERING COLLEGE

(AUTONOMOUS)

Sree Sainath Nagar, A.Rangampet - 517 102

VISION

To be one of the Nation's premier Engineering Colleges by achieving the highest order of excellence in Teaching and Research.

MISSION

- To foster intellectual curiosity, pursuit and dissemination of knowledge.
- To explore students' potential through academic freedom and integrity.
- To promote technical mastery and nurture skilled professionals to face competition in ever increasing complex world.

QUALITY POLICY

Sree Vidyanikethan Engineering College strives to establish a system of Quality Assurance to continuously address, monitor and evaluate the quality of education offered to students, thus promoting effective teaching processes for the benefit of students and making the College a Centre of Excellence for Engineering and Technological studies.

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

Vision

To be a center of excellence in Electronics and Communication Engineering through teaching and research producing high quality engineering professionals with values and ethics to meet local and global demands.

Mission

- The Department of Electronics and Communication Engineering is established with the cause of creating competent professionals to work in multicultural and multidisciplinary environments.
- Imparting knowledge through contemporary curriculum and striving for development of students with diverse background.
- Inspiring students and faculty members for innovative research through constant interaction with research organizations and industry to meet societal needs.
- Developing skills for enhancing employability of students through comprehensive training process.
- Imbibing ethics and values in students for effective engineering practice.

B. Tech. (Electronics and Communication Engineering)

Program Educational Objectives

After few years of graduation, the graduates of B.Tech (ECE) will be:

PEO1. Enrolled or completed higher education in the core or allied areas of electronics and communication engineering or management.

PEO2. Successful entrepreneurial or technical career in the core or allied areas of electronics and communication engineering.

PEO3. Continued to learn and to adapt to the world of constantly evolving technologies in the core or allied areas of electronics and communication engineering.

Program Outcomes

On successful completion of the Program, the graduates of B.Tech. (ECE) Program will be able to:

PO1 **Engineering knowledge**: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.

PO2 **Problem analysis**: Identify, formulate, research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

PO3 **Design/development of solutions**: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.

PO4 **Conduct investigations of complex problems**: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

PO5 **Modern tool usage**: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.

PO6 **The engineer and society**: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.

PO7 **Environment and sustainability**: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

PO8 **Ethics**: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

PO9 **Individual and team work**: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.

PO10 **Communication**: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.

PO11 **Project management and finance**: Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.

PO12 **Lifelong learning**: Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

Program Specific Outcomes

On successful completion of the Program, the graduates of B. Tech. (ECE) will be able to

PSO1. Apply the knowledge of Electronics, Signal Processing, Communications, and VLSI & Embedded Systems to the solutions of real world problems.

PSO2. Analyze, Design and Develop solutions in real time in the domains of Electronics, Signal Processing, Communications, and VLSI & Embedded Systems.

PSO3. Conduct investigations and address complex engineering problems in the domains of Electronics, Signal Processing, Communications, and VLSI & Embedded Systems.

PSO4. Apply appropriate techniques, resources, and modern tools to complex engineering systems and processes in the domains of Electronics, Signal Processing, Communications, and VLSI & Embedded Systems.

Department of Electronics and Communication Engineering

SREE VIDYANIKETHAN ENGINEERING COLLEGE

(AUTONOMOUS)

Sree Sainath Nagar, A. Rangampet, Tirupathi - 517102.

Certificate

This is to certify that the project report entitled

"DETECTING TAMPERED REGIONS IN JPEG IMAGES VIA CNN"

is the bona fide work done and submitted by

NAGIREDDY CHANDRA MOULI REDDY -17121A04F0
SHAIK THOUHEED AHAMED -17121A04K4
TALAPALA GANESH -17121A04L3
N V SAI VIVEK -17121A04E9

in the Department of Electronics and Communication Engineering, ***Sree Vidyanikethan Engineering College,*** *A.Rangampet, affiliated to* **Jawaharlal Nehru Technological University Anantapur, Anantapuramu** *in partial fulfillment of the requirements for the award of the degree of* **Bachelor of Technology** *in* **Electronics and Communication Engineering** *during 2017-2021.*

GUIDE

Dr. N. Padmaja, M.Tech., Ph.D.,
Professor, Department of ECE.

HOD

Dr. N. Gireesh, M.Tech., Ph.D.
Professor & Head, Dept. of ECE.

INTERNAL EXAMINER

EXTERNAL EXAMINER

ACKNOWLEDGEMENTS

The successful completion of this project report is made possible with the help and guidance received from various quarters. We would like to avail this opportunity to express our sincere thanks and gratitude to all of them.

We are deeply indebted to our guide **Dr. N. Padmaja, M.Tech., Ph.D., Professor, Department of ECE**. We are really fortunate to associate ourselves with such an advising and helpingguide in every possible way at all stages for successful completion of this project work.

We extended our deep sense of gratitude to **Dr. N. Gireesh, M.Tech., Ph.D., Headof the Department of Electronics and Communication Engineering,** for his moral support and valuable advices during this project work and the course.

We thank our Principal **Dr. B. M. Satish, M.S., Ph.D.** for supporting us in completion of this project work successfully by providing the facilities. We are pleased to express our heartfelt thanks to our faculty in the Electronics and Communication Engineering Department for their moral support and good wishes.

Finally, we have notion to express our sincere thanks to our family and our friends and all those who guided, inspired and helped us in the completion of project work.

Team Members:

NAGIREDDY CHANDRA MOULI REDDY-17121A04F0
SHAIK THOUHEED AHAMED -17121A04K4
TALAPALA GANESH -17121A04L3
N V SAI VIVEK -17121A04E9

DECLARATION

We hereby declare that project report entitled "**DETECTING TAMPERED REGIONS IN JPEG IMAGES VIA CNN**" being submitted by us for award of degree of Bachelor of Technology in Electronics and Communication Engineering. Jawaharlal NehruTechnological University Anantapuramu is a bonafide record of the **SREE VIDYANIKETHAN ENGINEERING COLLEGE** and has not been submitted to any other courses or university of award of any degree.

Team Members:

NAGIREDDY CHANDRA MOULI REDDY-17121A04F0
SHAIK THOUHEED AHAMED -17121A04K4
TALAPALA GANESH -17121A04L3
N V SAI VIVEK -17121A04E9

ABSTRACT

In this work, we will detect the tampered regions in JPEG images using deep learning techniques (U Net architecture). Often, digital pictures are used as evidence in criminal investigations. Therefore, it is essential to check whether they have been tampered with or not. DCT coefficients play an important role in the detection of Tampered regions of images and these DCT coefficients are input to the CNN (Convolutional neural network). Convolutional neural network (U Net architecture) has been successfully used to achieve good performance in detecting tampered regions of images. U-net is convolutional network architecture for fast and precise segmentation of images mainly. Experiments will demonstrate that our model will provide best detection performance compared to the state-of-the-art methods.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

The hot topic known as “fake news” is becoming increasingly widespread across social media, which for many people has become their primary news source. Fake news is information that has been altered to represent a specific agenda. To support the production of fake news, images that have been tampered with are often presented with the associated reports. Fake news production has been enabled in recent years because of two main reasons: first, cost reductions of the required image-producing technology (e.g., cell phones and digital cameras); and second, the widespread accessibility of image-editing software from open-source tools and apps. Anyone with a cell phone or digital camera who has online access to the necessary software can now alter images easily and cheaply, for whatever purpose. At the same time, online access enables the images to be sent across a virtually limitless number of platforms, where they can be further altered through dedicated imagery software (e.g., Photoshop) using tools such as splicing, painting, or copy-move forgery.

Considering how easy it is to create fake images as part of a fake news report, there is a critical need for detection methods that can keep up with the latest technology in fraud production. The integrity of an image can be validated through one or more strategies, either alone or in combination, which test an image for authenticity. A popular strategy is copy-move image forgery, which involves copying or cloning an image patch into an identical image. The patches to be copied or clones can be either irregular or in regular form. Copy-move image forgery is increasing in popularity due in large part to its ease of use. Furthermore, because the copied or cloned patch has its source in the original image, photometric characteristics between the original and the forgery are essentially the same, making it that much more difficult for the fake to be detected.

Since its recent development at around the turn of the present century, the primary purpose for copy-move forgery detection (CMFD) has been determining if the imaging probe in question (otherwise known as the query) features any areas that are cloned, and whether this cloning has been performed with malicious intent. The three main types of copy-move forgeries are plain, affine, and complex. The earliest CMFD investigations dealt mostly with plain cloning. Interestingly, in a previous study, the researchers found that human judgment outsmarted machine learning regarding computer-generated forgeries. This could be caused by the current lack of photorealism found in most computer graphics tools.

Entering the post-mobile era, trillions of digital photos are produced every year. The explosion of image data promotes the development of digital image editing tools such as Photoshop and BeautyCam, enabling people to edit real-world photos from low-level pixels (e.g., adjust the lighting conditions in a selfie) to high-level 5 semantic contents (e.g., replace the sky in a wedding photo). The advancement of image manipulation techniques is a double-edged sword. On one hand, it facilitates the beautification of photos and thereby encourages people to express and share their ideas on visual arts of photo editing; on the other hand, it is much easier to forge the content of a given image without leaving any visible clues and thus helps forgers to deliver fake information. When image manipulation techniques are employed with unethical intention, the achieved compelling graphic effect of modified images would deceive the public. A well-known example is the Kerry Fonda 2004 election photo controversy.

In the US presidential election of 2004, the candidate John Kerry (the man in Fig. 1.1) was libeled by a photo showing that he was sitting next to Jane Fonda (the woman in Fig. 1.1) at an anti-war rally in 1970. This photo enraged many Vietnam veterans and caused negative sentiments towards John Kerry because of Jane Fonda's 15 opposition to the Vietnam War. However, this photo was later proven fake that it is actually a composite made from two different photos. Although the forger was finally sued in New York Federal District Court, the political impact of the published photo to the public had been irreversible. More recently, high-level image manipulation has been much automated by advanced computer vision technology.

Figure 1.1: The well-known Kerry-Fonda photo is actually a composite of two differentphotos.

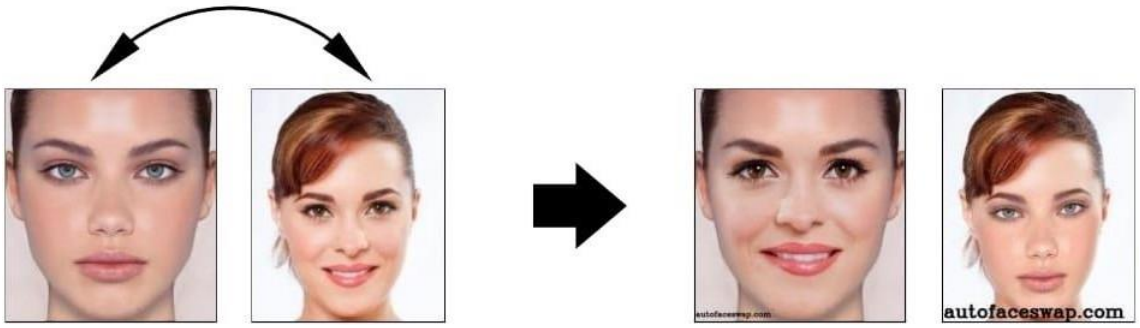


Figure 1.2: Swapping two faces using the auto face swap software.

For instance, localizing face regions in photos and swapping them with necessary distortion can be done in just a few 20 seconds by mobile or desktop applications. Figure 1.2 shows an example of swapped faces. Realistic visual effect can be achieved such that the fake photos are visually indistinguishable from the real ones. As mentioned above, the automation of such image tampering techniques not only makes it easier for forgers to produce realistic photos, but also encourages other people to abuse the software for fun. Consequently, the bad intention of the forgers becomes subtle and the increasingly large number of tampered images are hidden dangers on the internet. It is essential to 25 develop image tampering detection methods for efficiently searching and tracking tampered images over the internet, in order to prevent photo editing techniques from being used for any improper purposes.

1.1. Image Manipulation, Forgery and Tampering

Image editing or image manipulation refers to any operation that can be done to a digital image by software on a computer or other digital devices such as tablets and mobiles. Photo editing in a darkroom has been an outdated fashion since the film cameras were almost substituted by cheap but quality digital cameras in our life. Therefore, without special clarification, the term image manipulation used through this survey denotes digital image manipulation.

Common image manipulations not only include pixel-level operations such as resampling an image to reduce its size, but also include content-level operations such as removing an object from a digital photo. No matter which level operations are applied, an original image exists for each manipulation product. One can know changes by simply Common image manipulations not only include pixel-level operations such as resampling an image to reduce its size, but also include content-level operations such as removing an object from a digital photo. No matter which level operations are applied, an original image exists for each manipulation product. One

can know the changes by simply by slightly alter the image pixels at certain positions. The changes made to the original image are relatively subtle with respect to the magnitude of pixel values in the image, e.g., the least significant bit (LSB) embedding method alters the pixel values by only -1 or +1. Therefore, it is hard to see the manipulation clues by visual comparison between the product and the original. In fact, the objective of image steganography is to hide secret (sometimes malicious) information from human eyes but not to trick human eyes with fake image graph.

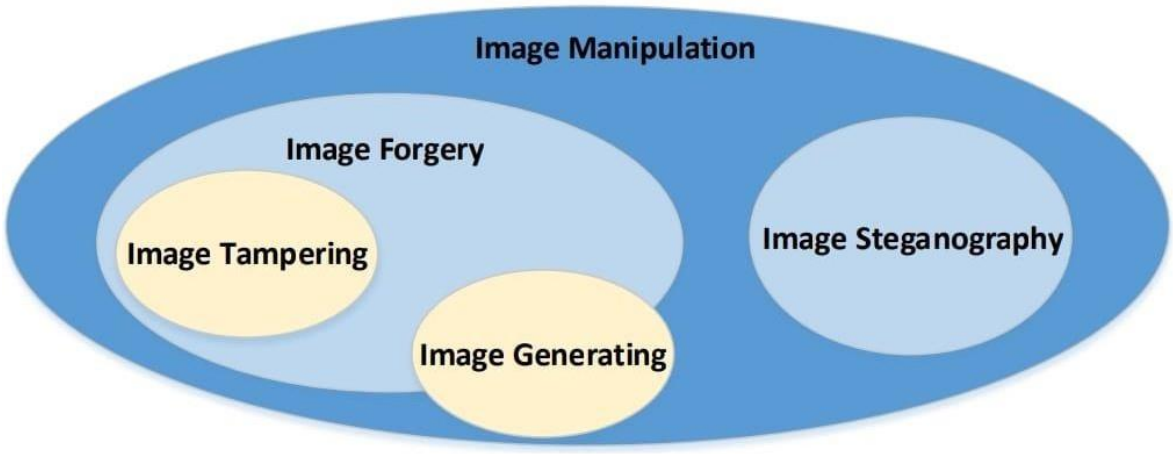


Figure 1.3: The hierarchical structure of image manipulation, forgery and tampering.

Image forgery emphasizes eye tricks. It denotes image manipulation that aims to deliver deceptive information through the image graphic content. In 2004, Farid defined and illustrated six different image forgery types, i.e., compositing, morphing, re-touching, enhancing, computer generating and computer painting. The first three techniques fundamentally alter the semantic content or appearance of an image; the enhancing technique can obscure or exaggerate the image details by altering the image color map or contrast. The last two directly create a virtual image artifact even without an original image. However, the graphic content of the virtual image may have violated some facts happened in the past, in order to deceive the viewers or improve storytelling. From this point of view, image forgery emphasizes the falsified part in an image that violates the happened facts, no matter they have been captured as an original image or not. For the same reason, some cases of image generating and image painting, that are not based on happened facts but only on the manipulator's art experience and creativity, do not belong to image forgery (see Fig. 1.3). Apart from the forgery examples in Fig. 1.1 and Fig. 1.2, more examples can be found in Farid's collection of "Photo Tampering Throughout History" .

Compositing may be the most common sub-category of image forgery that we can see in daily life. Indeed, making a dark photo brighter by enhancing its contrast is less likely to catch one's eye, but putting two irrelevant people in the same scene is worth some gossip for quite a long time. Wang et al. re-identified compositing as image tampering that alters portions of a given image to conceal an object in the scene or add a new object to the scene. There must be an original for a tampered image; the tampered image contains both tampered regions and untampered areas. It would be very easy for one to identify the tampered regions by visually comparing the product to the original. However, when one is given a tampered image without its original as the reference, it is hard to judge the manipulator's intention as good or evil. Blind detection of image tampering is thereby very challenging when the tampering clues in the product are visually imperceptible.

Table 1.1: Definitions of image manipulations, image forgery and image tampering.

Terminology	Definition
Image manipulation	Computing techniques that paint or edit a digital image.
Image forgery	Image manipulations that produce fake graphic content which falsify some facts happened in the past.
Image tampering	A special type of image forgery that alters a part or multiple parts of the graphic content of a given image.

From the above, the definitions of image manipulation, image forgery and image tampering can be given in Table 1.1. Image manipulation is the general name of all kinds of digital image editing and painting operations. Image forgery, as one can tell from its name, is a specific type of image manipulation for falsifying some happened facts. As one of the forgery categories, image tampering emphasizes that a region within a given image has been tampered. Besides, image steganography is not categorized to image forgery because it does not significantly change the graphic content of an image. Image generating and painting techniques may be used for forgery, but generating an image does not necessarily prove forgery in it. The relationship between the different categories is illustrated in Fig. 1.3. Here image generating is a particular type of image creating techniques that it utilizes computer software or algorithms to create an image mimicking real-world scenes.

Accordingly, the countermeasures to the three-level hierarchies are image manipulation detection, image forgery detection and image tampering detection, respectively. All of them can be further categorized to the general class of image forensics, which includes other digital investigation works like device forensics besides the work of image manipulation analysis. For

readers interested in a broader scope on digital forensics, we recommend a recent survey that has provided a critical review of current progresses. In contrast, the focus of this study will be image tampering and its detection in real-world photos.

1.2. Image Tampering in Real-world Photos

Tampering can be performed on either real-world photos or computer-generated images. Although we have defined in the above that image generating may not belong to image forgery, tampering on computer-generated images will indeed produce a tampered image. However, from the point of the view of a forger, it makes no sense that a forger first generates an image to picture some facts and then alters this image. Instead, he can directly generate an image telling the fake story. If so, the task becomes to distinguish computer-generated images from real-world photos. And once an image is confirmed as computer-generated, people would not trust it as a sound evidence to support the facts. Therefore, when we refer to tampered images, we imply that they are real-world photos.

Table 1.2: Common image tampering types, named and categorized according to the major manipulation operation used for tampering.

Category		Visible change in comparison	Single image source	Region duplication	Major use in tampering
Splicing	Copy-move	✓	✓	✓	Object removal
	Cut-paste	✓	×	✓	Object addition
Inpainting	Erase-fill	✓	✓*	×	Object removal

* Current image in painting methods use single image as tampering source, but it is possibleto extend to multiple images by someone.

According to its definition in the previous section, the procedure of image tampering is actually to replace the content within a region of the original by some new content. The source and composition of the new content determines the specific types of tampering. If the new content is an entire part within the original image itself, it is named as copy-move. If the new content is an entire part from another source image, it is cut-paste. If the new content is a composition of elementary patches within the original image or from another image, we call it as erase-fill.

In particular, current software suggest using neighbouring patches or pixels within the original image to replace the target region because using these patches is easy and more likely to achieve smooth filling effect than using patches from another arbitrary image. However, we can not exclude that someone may use another image rather than the original image as the patch source.

Other common terminologies: The term ”copy-move” has been widely used to denote the scenario that a region is duplicated within an image. In contrast, the term ”cut-paste” is rarely used and its substitute ”splicing” appears more often in published papers to denote region duplication between two or more images. In fact, some works used ”splicing” or ”composition” to cover the two types of tampering carried out by region duplication, no matterhow many images are used as sources. In order to avoid the confusion, we use ”cut-paste” to denote image splicing processing two or more images and ”splicing” is the collective name ofboth copy-move and cut-paste. The term ”erase-fill”, created by following the same naming convention of copy-move and cut-paste, actually has a more popular alias as ”image inpainting”, which originally refers to the application of restoring missing or corrupted parts inan image . Now we use ”erase-fill” to emphasize the process of filling the target region.

Table 1.2 summarizes the characteristics of the three tampering categories. First, all the three tampering types will cause significant change to the original image by replacing a region with new graphic content. This change can be pointed out by visually comparing the tampered image to the original. Second, copy-move and existing erase-fill applications operate with the original image as the only source. Third, copy-move and cut-paste manipulate relatively larger parts of the source image than erase-fill. Therefore, one can see obvious region duplication when all the source images are presented. Last, each type has its major use scenario. For example, due to its simplicity, copy-move is the most popular method for removing an undesired object by covering it with a part of the image background. Erasefill is an alternative object removal method that fills the region of the undesired object with neighbouring texture. In contrast, cut-paste is usually used for adding a new object to the original image. In general, both removal and addition are additive operations. Removing an object is adding a piece of image background (copy-move) or a composition of neighbouring texture (erase-fill); adding an

(a) Original image (b) Tampered image (c) Tampered region in white

Figure 1.4: The background in the original image is replaced in (b), the tampered region is marked in (c). Without given the original image (a), humans are difficult to identify (b) as a tampered image.

external object is removing a piece of image background. In practice, copy-move can be used to increase the number of existing objects; cut-paste can be applied for covering existing objects with a piece of background from another image; erase-fill can be adopted for restoringcorrupted parts of existing objects.

Post-processing: Some early works of image tampering detection did not consider image post-processing after one of the three tampering operations. For instance, splicing was defined as the simple cut-paste operation of image regions from one image onto the same or another image without performing post-processing. However, practical tampering often involves post-processing operations to smooth the boundaries of tampered regions, in order to make the final artifact less visually suspectable. There are two kinds of post-processing operations. One is active postprocessing for improving the tampering effect, e.g., image blurring, image resampling, brightness change and contrast adjustments. The other one is passive post-processing that may be unintentionally introduced to tampered images during data transmission, e.g., JPEG compression, noise adding and color reduction. Nowadays, when we refer to image tampering, we imply all the tampering processes with or without post-processing.

1.3. Humans are Easily Fooled by Tampered Images

It was found that when no original images are given for comparison, people have an extremely limited ability to detect and localize image tampering in real-world photos. For example, Figure 1.4 shows a tampering example from the CASIA image tampering detection evaluation dataset.

One can see that the backgrounds in the two images Fig. 1.4 (a) and (b) are different. Without prior knowledge, it is hard to know which of the two images is the original. In Fig. 1.4 (c), the tampered region is manually annotated in white by Zhang et al.in. It indicates that the image background in Fig. 1.4 (a) has been replaced. Recently, Schetinger et al. and Nightingale et al. independently conducted subjective evaluations to assess human's ability in image manipulation detection.

Schetinger et al. collected 17,208 subjective answers from 393 individuals, using 80 original images and 97 tampered images (i.e., 177 images in total) selected from three public image tampering datasets. In terms of tampering types, there are 20 erase-fill images, 35 copy-move images and 42 cut-paste images. All the 177 images were displayed with fixed resolution of 1024×768 to every subject, original resolution version were also available if the subject was keen to see. The subjects were first asked to answer if the shown image was authentic (i.e., the detection-only task), and then asked to point out the tampered region if they thought it was fake (i.e., the localization task). The major findings can be summarized as follows.

- The overall accuracy of tampering identification is 66.30%, indicating that the subjects' ability to detect tampered image was better than by chance, i.e., larger than 50%.
- The accuracy of identifying original images is 70.85% while the accuracy of identifying tampered images is 62.54% and only 46.50% of them can further correctly localize the tampered region. This indicates that the subjects had a bias towards saying an image is real when they were in doubt about the image's authenticity.
- The 9,417 answers on the tampered images are composed of 1,942 on the erase-fill images, 3,392 on the copymove images and 4,083 on the cut-paste images. The accuracies of correctly detecting and localizing image tampering are 38.5%, 46.9% and 59.4%, respectively. The authors adopted a bootstrap resampling approach to prove that erase-fill is statistically harder for people to identify than copy-move and cut-paste.

In contrast, Nightingale et al. studied combinations of two tampering types with fewer real-world photos but more subjects. The authors conducted two experiments, each using 10 photos of human faces in real-world scenes. Similar to the settings in, the subjects in the first experiment were asked to determine the shown image's authenticity, and then asked to select the tampered area if they thought it was fake. The overall accuracy on the detection-only task is 66%. Concretely, the subjects correctly identified 72% original images and 60% forged images. The accuracy of localizing the manipulated region in the forged images is 45%. One can observe that these four 160 reported results are coincidentally very close to that in the

previous evaluation by Schetinger et al. (i.e., 66.30%, 70.85%, 62.54% and 46.50%), proving again people's bias towards saying photos were authentic.

In the second experiment, subjects were asked to localize the manipulated region regardless of their response in the detection-only task. This condition seems to have changed people's bias on authentic photos, the detection accuracies on the original and tampered images are 58% and 65%, respectively. In term of the localization task, a mean accuracy of 56% was obtained, which is much higher than that in the first experiment (45%). As the detection accuracy of original images is lower than that of tampered images, we can know that the demand for the subjects to directly localize the tampered region has indeed encouraged their confidence in doubting the shown image as tampered. Furthermore, the subjects in the second experiment had indeed spent more time in searching the evidence of tampering. For example, when one is told that the image in Fig. 1.5 (a) is fake, he may carefully examine the image and find that the statue's head has been replaced.

From the above evaluations, a general conclusion can be drawn that when people has zero-knowledge of the original images, their ability of detecting and localizing image forgeries is

(a) Tampered image (b) Tampered region in white

Figure 1.5: When people are told that the given image is false they may be more confident to localize the manipulated region, because they generally would spend more time in carefully examining the image.

better than by chance but very limited. In contrast, people are better at searching online for original images and localizing tampering by comparison. One lesson we can learn from such subjective studies is that it is difficult to distinguish original and tampered images only through the visible channel of an image, i.e., the RGB patterns. When we develop automatic image manipulation detection methods, it may be better to analyze image transform domains such as

discrete cosine transform (DCT) and discrete wavelet transform (DWT), or analyze the image residuals.

Many studies were performed on the detection of image forgery. For example, Kaur proposed a method based on DCT(Discrete Cosine Transform) and SIFT(Scale-Invariant Feature Transform) to detect the copy–move forgery. Johnson described a technique for detecting image forgery by estimating the direction of a point light source. Popescu revealed the traces of digital tampering in color images interpolated using color-filter-array algorithms. Kobayashi used a method that was based on noise characteristics.

Different approaches were presented by He and Pevny; in these approaches, the target image was in the JPEG format. Notably, JPEG is the most frequently used image format, thereby making the approaches practical. The method proposed by He was a superior algorithm because it can automatically locate the tampered part irrespective of the shape and area of the tampered part. Similarly, we performed studies on the detection of tampered JPEG-format images and, consequently, reported that the overall performance of the MDBD(Multiple Detection using Block noise and Double JPEG) method is better than that of the method proposed by He.

In this study, we propose a novel technique to detect the tampered region in a JPEG image by using CNN. Compared with the MDBD method, in which feature values are selected via know-how, the proposed approach optimizes them using CNN and achieves a higher detection accuracy.

Chapter – 2

Literature Review

2.1 Detecting Doctored JPEG Images Via DCT Coefficient Analysis

J. He, Z. Lin, L. Wang, and X. Tang.

The steady improvement in image/video editing techniques has enabled people to synthesize realistic images/videos conveniently. Some legal issues may occur when a doctored image cannot be distinguished from a real one by visual examination. Realizing that it might be impossible to develop a method that is universal for all kinds of images and JPEG is the most frequently used image format, we propose an approach that can detect doctored JPEG images and further locate the doctored parts, by examining the double quantization effect hidden among the DCT coefficients. Up to date, this approach is the only one that can locate the doctored part automatically. And it has several other advantages: the ability to detect images doctored by different kinds of synthesizing methods (such as alpha matting and inpainting, besides simple image cut/paste), the ability to work without fully decompressing the JPEG images, and the fast speed. Experiments show that our method is effective for JPEG images, especially when the compression quality is high

Watermark [13] has been successful in digital right management (DRM). However, doctored image/video detection is a problem that is different from DRM. Moreover, plenty of images/videos are not protected by watermark. Therefore, watermark independent technologies for doctored image/video detection are necessary, as pointed out in [14, 19]. Farid et al. have done some pioneering work on this problem. They proposed testing some statistics of the images that may be changed after tempering [14] (but did not develop effective algorithms that use these statistics to detect doctored images), including the interpolation relationship among the nearby pixels if resampling happens when synthesis, the double quantization (DQ) effect of two JPEG compression steps with different qualities before and after the images are synthesized, the gamma consistency via blind gamma estimation using the bicoherence, the signal to noise ratio (SNR) consistency, and the Colour Filter Array (CFA) interpolation

relationship among the nearby pixels [15]. Ng [18] improved the bicoherence technique in [14] to detect spliced images. But temporarily they only presented their work on testing whether a given 128 × 128 patch, rather than a complete image, is a spliced one or not. Lin et al. [19] also

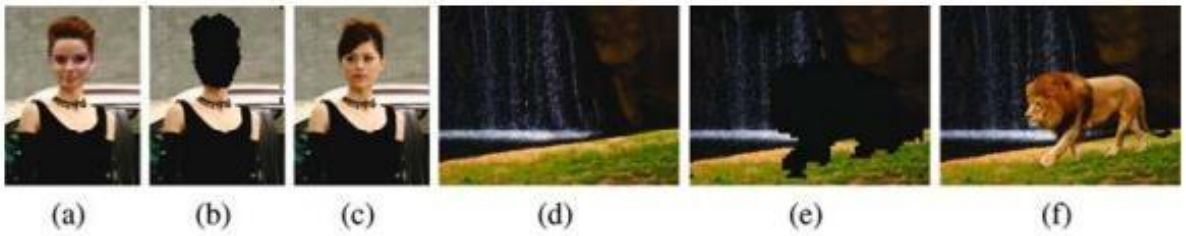


Fig. 2.1. Examples of image doctoring

proposed an algorithm that checks the normality and consistency of the camera response functions computed from different selections of patches along certain kinds of edges. These approaches may be effective in some aspects, but are by no means always reliable or provide a complete solution

Fig. 2.1. Examples of image doctoring and our detection results. (a) and (d) are two doctored JPEG images, where (a) is synthesized by replacing the face and (b) is by masking the lion and inpainting with structure propagation [9]. (b) and (e) are our detection results, where the doctored parts are shown as the black regions. For comparison, the original images are given in (c) and (f).

To proceed, we first give some definitions A "doctored" image means part of the content of a real image is altered. Note that this concept does not include those wholly synthesized images, e.g. an image completely rendered by computer graphics or by texture synthesis. But if part of the content of a real image is replaced by those synthesized or copied data, then it is viewed as "doctored". In other words, that an image is doctored implies that it must contain two parts: the undoctored part and the doctored part. A DCT block, or simply called a "block", is a group of pixels in an 8 × 8 window. It is the unit of DCT that is used in JPEG. A DCT grid is the horizontal lines and the vertical lines that partition an image into blocks when doing JPEG compression. A doctored block

Our method has several advantages. First, it is capable of locating the doctored part automatically. This is a feature that is rarely possessed by the existing methods. The duplicated region detection [16] may be the only exception. But copying a part of an image to another position of the image is not a common practice in image forging. Second, most of the existing methods aim at detecting doctored images synthesized by the cut/paste skill. In contrast, our method could deal with images whose doctored part is produced by different kinds of methods such as inpainting, alpha matting, texture synthesis and other editing skills besides image

cut/paste. Third, our algorithm directly analyses the DCT coefficients without fully decompressing the JPEG image. This saves the memory cost and the computation load. Finally, our method is much faster than the bi-coherence based approaches [14, 18], iterative methods [14], and the camera response function based algorithm [19]

Advantages : Ability to detect images doctored by different kinds of synthesizing methods, besides simple image cut/paste

Scope of Work : Procedure of DCT coefficients detection.

2.2 Exposing Digital Forgeries in Color Filter Array Interpolated Images

A. C. Popescu and H. Farid.

With the advent of low-cost and high-resolution digital cameras, and sophisticated photo editing software, digital images can be easily manipulated and altered. Although good forgeries may leave no visual clues of having been tampered with, they may, nevertheless, alter the underlying statistics of an image. Most digital cameras, for example, employ a single sensor in conjunction with a color filter array (CFA), and then interpolate the missing color samples to obtain a three channel color image. This interpolation introduces specific correlations which are likely to be destroyed when tampering with an image. We quantify the specific correlations introduced by CFA interpolation, and describe how these correlations, or lack thereof, can be automatically detected in any portion of an image. We show the efficacy of this approach in revealing traces of digital tampering in lossless and lossy compressed color images interpolated with several different CFA algorithms

2.2.1 COLOR FILTER ARRAY INTERPOLATION ALGORITHMS

A digital color image consists of three channels containing samples from different bands of the color spectrum, e.g., red, green, and blue. Most digital cameras, however, are equipped with a single charge-coupled device (CCD) or complementary metal-oxide semiconductor (CMOS) sensor and capture color images using a color filter array (CFA). The most frequently used CFA, known as the Bayer array [12], employs three color filters: red, green, and blue. The red

and blue pixels are sampled on rectilinear lattices, whereas the green pixels are sampled on a quincunx lattice;

Detecting Localized Tampering Since it is likely that tampering will destroy the periodicity of the CFA correlations, it may be possible to detect and localize tampering in any portion of an image. To illustrate this, consider the left-most image, taken with a Nikon D100 digital camera and saved in RAW format. Each color channel of this image (initially interpolated using the adaptive color plane technique) was blurred with a 3X3 binomial filter and down sampled by a factor of two in order to destroy the CFA periodic correlations. The 512X512 down sampled image was then resampled on a Bayer array and CFA interpolated. Next, composite images, which were 512X512 pixels in size, were created by splicing, in varying proportions (1/4, 1/2 and 3/4), the non-CFA interpolated image and the same image CFA interpolated with the bicubic algorithm; see Section II-A. The probability maps obtained from running EM on the red channel of the composite images. Notice that these probability maps clearly reveal the presence of two distinct regions.

2.2.2 Sensitivity and Robustness

From a digital forensics point of view, it is important to quantify the robustness and sensitivity of our detection technique. It is therefore necessary to devise a quantitative measure for the periodic correlations introduced by CFA interpolation. This is achieved by comparing the estimated probability maps of each color channel of a given image with synthetically generated probability maps

To simulate tampering, each image is blurred and down sampled in order to destroy the original CFA interpolation correlations. These images are then interpolated with each of the algorithms described in Section II. Given a three-channel color image, we first compute the similarity measures between the probability maps of each color channel and the corresponding synthetic maps. The image is labeled CFA interpolated if at least one of the three similarities is greater than its specified threshold or labeled tampered if all three similarities are less than the thresholds. Using this approach, we have obtained, with a 0% false positive rate, the following classification accuracies averaged over 100 images: bilinear: 100%, bicubic: 100%, smooth hue: 100%, median (3X3): 99% median (5X5): 97%, gradient-based: 100%, adaptive color plane: 97%, and variable number of gradients: 100%. To be useful in a forensic setting, it is important for our detection method to be robust to simple image distortions. We have tested

the sensitivity of our method to typical distortions that may conceal traces of tampering: 1) JPEG compression, 2) additive white Gaussian noise, and 3) nonlinear pointwise gamma correction. Fifty images taken with the Nikon Coolpix 950 camera were processed as described above in order to obtain CFA and non-CFA interpolated images. The detection accuracies (with 0% false positives) for different CFA interpolation algorithms as a function of the JPEG compression quality. Note that these detection accuracies are close to 100% for quality factors greater than 96 (out of 100) and that they decrease gracefully with decreasing quality factors. This decrease in accuracy is expected since the lossy JPEG compression introduces noise that destroys the correlations introduced by CFA interpolation. The decrease in accuracy depends on the CFA interpolation algorithm, e.g., the accuracy for smooth hue CFA interpolation is 56% at quality 70, whereas the accuracy for adaptive color plane CFA interpolation drops to 6%. The detection accuracies (with 0% false positives)

Advantages : Images can be classified even subjected to compression and nonlinearities

Scope of work : Studied about color-filterarray algorithms.

2.3 Detecting Video Forgeries Based on Noise Characteristics

M. Kobayash i, T. Okabe, and Y. Sato

The recent development of video editing techniques enables us to create realistic synthesized videos. Therefore using video data as evidence in places such as a court of law requires a method to detect forged videos. In this paper we propose an approach to detect suspicious regions in video recorded from a static scene by using noise characteristics. The image signal contains irradiance-dependent noise where the relation between irradiance and noise depends on some parameters; they include inherent parameters of a camera such as quantum efficiency and a response function, and recording parameters such as exposure and electric gain. Forged regions from another video camera taken under different conditions can be differentiated when the noise characteristics of the regions are inconsistent with the rest of the video

In the early days of the Internet, digital watermarking was the main countermeasure against illegal use of digital contents [6]. However, most images and videos do not have an embedded digital watermark. Once images or videos without watermarks are uploaded to the Internet, digital watermarks are ineffective even if they are embedded afterwards because the contents

may have already been tampered with by someone. Therefore digital watermarking is found to be limited in its ability to assure authenticity.

One of the most frequent digital evidence declared invalid in a court of law is a video recorded by a fixed surveillance camera. Tampering methods for a scene that contains a static background can be classified into two approaches. One is replacing regions or frames with duplicates from the same video sequence: forgers can hide unfavourable objects in a scene by overwriting these with the background. The other is clipping objects from other images or video segments and superimposing them on the desired regions in the video. This type of forgery aims to show objects that are advantageous for false evidence.

2.3.1 Forgery Detection Methods for Images and Video

The area of digital image forensics has progressed so markedly in the last few years that several approaches have been developed to detect forgeries in a digital image. Image tampering methods can be classified into two approaches. One is replacing regions with others in the same image and the other is superimposing regions clipped from other images. The first attempt of forgery detection was proposed by Fridrich et al [2]. This method targets the copy-move method of attack, which yields unnaturally high correlation between duplicated regions. The researchers introduced a detection method based on robust block matching, which was carried out by using Discrete Cosine Transform (DCT) coefficients in order to deal with lossy JPEG compression. Subsequent approaches target the superimposition-based forgeries, which verify the uniformity of characteristics in an image; therefore objects clipped from other images could be detected. Jonson and Farid proposed methods based on optical clues. They estimated the light source directions from some contours in an image and checked the consistency of estimated light source directions [4]. This technique showed so accurate estimation of light source directions for outdoor scenes that it could differentiate tampered objects in the image. JPEG is a compression technique for images; different manufacturers design different quantization tables used in a compression process. Ye et al. proposed a method to detect inconsistencies in an image based on the blocking artifact measure [17]. If blocks compressed with different quantization tables are combined in an image, the blocking artifact measure of forged blocks is much larger than that of an authentic block. They estimated the quantization table from the histogram of DCT coefficients and evaluated the blocking artifact measure of each block.

Wang and Farid proposed forgery detecting methods based on video duplication and a deinterlacing algorithm [15,16]. The first approach that detect duplication is similar to the correlation-based detection proposed by Fridrich et al., extended so that it could detect duplicated regions across frames. They combined spatial and temporal correlation for detecting duplicated frames as well. On the other hand, the deinterlacing algorithm is a technique of converting interlaced video into a non-interlaced form. Due to the half resolution of interlaced video, the deinterlacing algorithm makes full use of insertion, duplication, and interpolation of frames to create full-resolution video. Parameters in the interpolation and the posterior probability of forgery are estimated by using the Expectation Maximization (EM) algorithm. Wang and Farid referred to forgery detection for interlaced videos in the same paper. They suggested that the motion between fields of a frame is closely related to that across fields in interlaced videos. Evaluating the interference to this relation by tampering, they detect the forgeries in the given interlaced video.

The methods proposed by Wang and Farid are interesting attempts for digital video forensics. It should be pointed out, however, that these methods have limitations for forgery detection. The first forensic technique based on correlation assumes that forged regions are duplicated from the same video sequence. As a result, this method has the same limitation for forgery detection as the method proposed by Fridrich et al., that it cannot detect superimposed regions from other videos. The second method targeting deinterlaced and interlaced videos can detect superimposing from other video sequences, but it limits the form of the video to deinterlaced or interlaced form.

On the other hand, some researchers have recently introduced interesting attempts to make effective use of noise, rather than trying to remove it from images and videos. Matsushita and Lin exploited the distribution of noise intensity for each scene irradiance to estimate the camera response functions (CRFs) [11]. Noise distribution is by nature shown to be symmetric, but it is skewed by nonlinear CRFs. Conversely, the inverse CRF can be estimated by evaluating the degree of symmetry of back-projected irradiance distribution. Using the noise in an image, the detection ability of the method is not degraded by noise and thus the method can be used under conditions of high-level noise.

Noise information is available for camera identification and forgery detection as well. Due to the sensor imperfections developed in a manufacturing process, the CCD camera contains pixels with different sensitivity to light. This spatial variation of sensitivity is temporally fixed and known as fixed pattern noise. Since this non-uniformity is inherent in a camera, we can exploit it as a fingerprint. Luk´aˇs et al. determined the reference noise pattern of a camera by

averaging the noise extracted from several images [9]. They extracted fixed pattern noise from a given image using a smoothing filter and identified the camera that took the image. The authors also proposed a method for detecting forgeries in an image using the Same approach In this section, we propose a forgery detecting method using a noise characteristics model. In this paper, we will consider the inconsistencies of the characteristics of the noise mixed in the signal to be a clue to tampering. We first introduce a noise characteristic model in Section 3.1. As stated before, we focus in particular on photon shot noise for detecting forgeries in the given video. This is because the variance of observed intensity caused by photon shot noise is closely related to its mean. The relationship between the variance and mean of observed intensity is formulated as the noise level function (NLF), which is the clue to tampering. In Section 3.2, we propose a method to estimate NLF and detect forgeries by using the estimated NLF.

2.3.2 Detection of Forged Pixels

Based on the theoretical background described in the previous section, we analyze the noise characteristics and detect forgeries of the given video by the following process. First, the mean and the variance of the pixel value are calculated at each pixel. Next, the NLF is estimated by fitting a function to the noise characteristic points. Finally, each pixel is evaluated based on its distance from the estimated NLF.

Advantages : accurately evaluate the per-pixel authenticity of the given video

Scope of work : Performance of Noise characteristics.

CHAPTER-3

EXISTING METHOD

This study is worked to classify/distinguish Morphed images and normal images using svm classifier. The MDBD method, as depicted in Fig.1, is based on both the block noise analysis and double JPEG analysis. Upon pasting a region from a JPEG image onto the host image, assuming random placement of the forged region, the probability of the blocking artifacts not being aligned is 63/64. Therefore, high-frequency components derived from unaligned blocking artifacts exist in the tampered parts. The parameters that measure the effects of tampering have been expressed quantitatively by analyzing the extracted 19 high-frequency components from 64 DCT coefficients of each 8_8-pixel block and, subsequently, estimating the position of the blocking artifacts. In addition, the effects of double JPEG have been quantified. By inputting these parameters to a support vector machine, both the tampered image and original image can be identified.

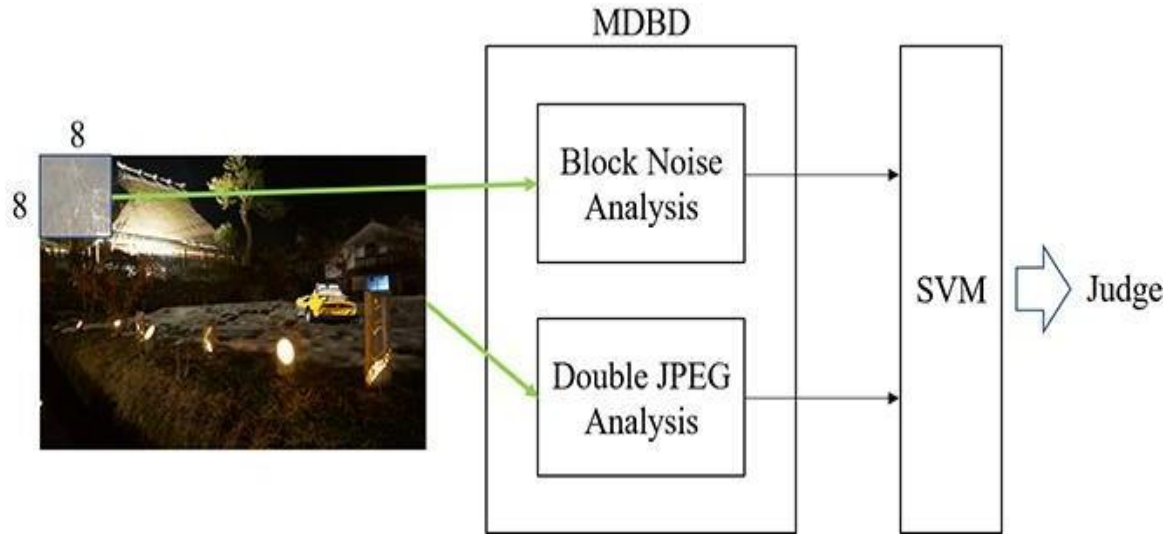


Figure 3.1: Overview of MDBD Method.

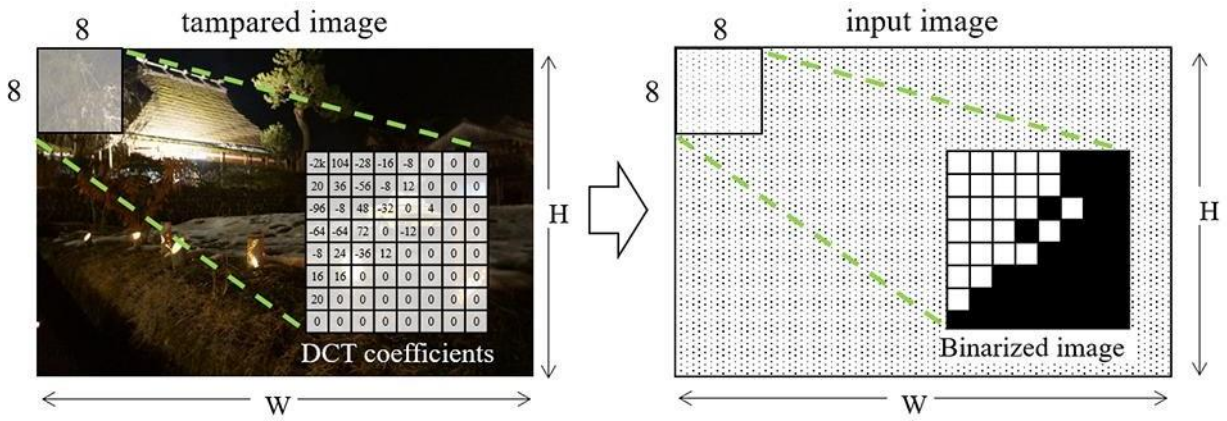


Figure 3.2. Input image with binarized DCT coefficients.

As a result of this experiment, the following two points have been revealed.

1. We must focus on high-frequency components of the DCT coefficients because the tampered part is affected by the blocking artifacts.

2. The spatial continuity of potentially tampered regions must be considered while detecting tampered parts. The MDBD method can discriminate between the tampered image and original image with a higher accuracy than that of the method by He; however, the accuracy of locating the tampered part must be improved.

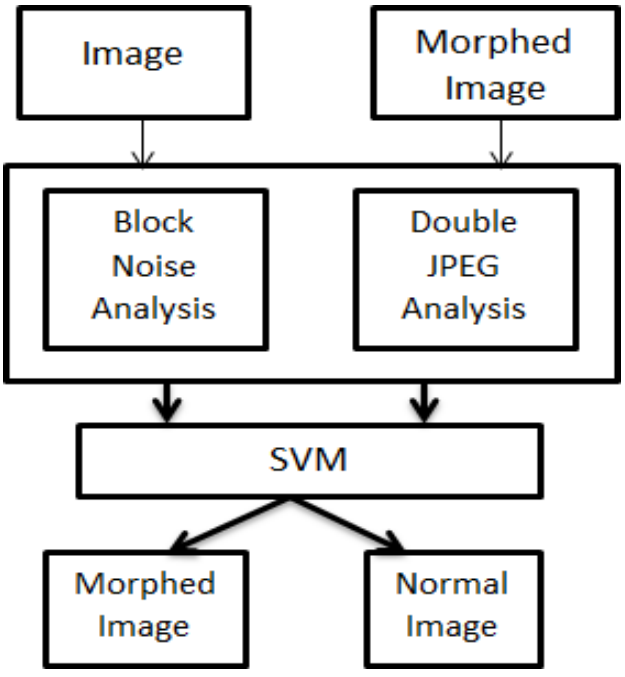


Figure 3.3: Block diagram of Existing Method.

3.1 Digital Morphing:

In the early 1990s computer techniques that often produced more convincing results began to be widely used. These involved distorting one image at the same time that it faded into another through marking corresponding points and vectors on the "before" and "after" images used in the morph. For example, one would morph one face into another by marking key points on the first face, such as the contour of the nose or location of an eye, and mark where these same points existed on the second face. The computer would then distort the first face to have the shape of the second face at the same time that it faded the two faces. To compute the transformation of image coordinates required for the distortion, the algorithm of beier and Neely can be used.

Present Use:

Morphing algorithms continue to advance and programs can automatically morph images that correspond closely enough with relatively little instruction from the user. This has led to the use of morphing techniques to create convincing slow-motion effects where none existed in the original film or video footage by morphing between each individual frame using optical flow technology. Morphing has also appeared as a transition technique between one scene and another in television shows, even if the contents of the two images are entirely unrelated. The algorithm in this case attempts to find corresponding points between the images and distort one into the other as they crossfade.

While perhaps less obvious than in the past, morphing is used heavily today. Whereas the effect was initially a novelty, today, morphing effects are most often designed to be seamless and invisible to the eye.

A particular use for morphing effects is modern digital font design. Using morphing technology, called interpolation or multiple master tech, a designer can create an intermediate between two styles, for example generating a semibold font by compromising between a bold and regular style, or extend a trend to create an ultra-light or ultra-bold. The technique is commonly used by font design studios.

3.2 Block noise analysis:

Noise is always presents in digital images during image acquisition, coding, transmission, and processing steps. It is very difficult to remove noise from the digital images without the prior knowledge of filtering techniques. In this article, a brief overview of various noise filtering techniques. These filters can be selected by analysis of the noise behaviour. In this way, a complete and quantitative analysis of noise and their best suited filters will be presented over here. Filtering image data is a standard process used in almost every image processing system. Filters are used for this purpose. They remove noise from images by preserving the details of the same. The choice of filter depends on the filter behaviour and type of data.

Filtering techniques:

We all know that, noise is abrupt change in pixel values in an image. So when it comes to filtering of images, the first intuition that comes is to replace the value of each pixel with average of pixel around it. This process smooths the image. For this we consider two assumptions.

Assumption:

1. The true value of pixels is similar to true value of pixels nearby
2. The noise is added to each pixel independently.

Let’s first consider 1-dimensional function before going into 2-dimensional image.

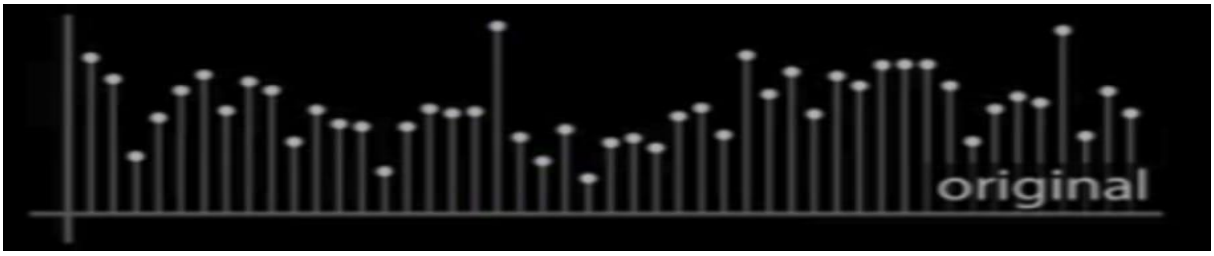


Figure 3.4: Original image signal in 1 Dimension with noise Source.

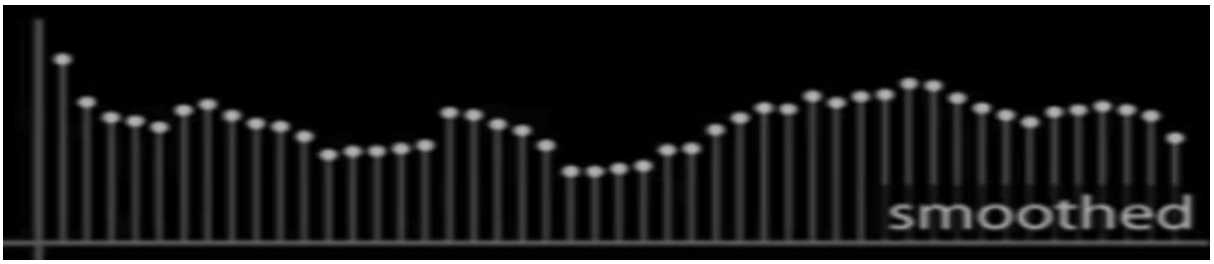


Figure 3.5: Image Signal in 1 Dimension after averaging Source.

In the above image of original function(fig-3.4), if we will consider each circle as pixel values, then the smoothed function(fig-3.5) is the result of averaging the side by pixel values of each pixel.

3.2.1 Filtering with weighted moving average uniform weight:

Instead of just thinking about averaging the local pixel, which is resulting in some loss of data, we consider a set of local pixel and assign them as uniform weights. Here we assume that noise is added to each pixel independently. According to this noise amount, we assign weights

noise is added to each pixel independently. According to this noise amount, we assign weights to different pixels.

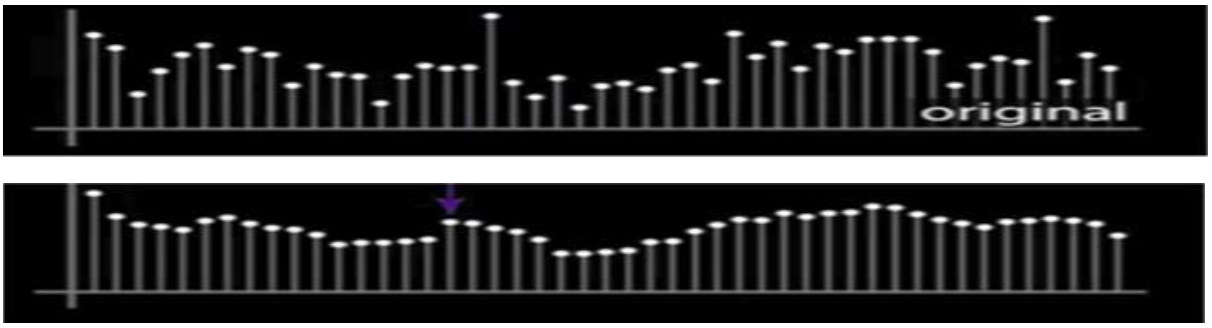


Figure 3.6 original image signal and image signal after uniform weighted moving averaging

3.2.2 Filtering with weighted moving average non-uniform weight:

Previously we took the assumption that the true value of pixels is similar to true value of pixels nearby. But it is not always true. So for higher accuracy we assign the nearby pixels with greater

weight then the pixels that are far away. This smooths the image and preserves the image information with less amount of data loss.

3.2.3 Weighted moving average in 2-dimensional image:

Thinking of image as a 2-dimensional matrix, we slide a small window (the red square in fig. 3.7)over the whole image to replace each pixel with the average of nearby pixels. This small windowis otherwise known as **mask or kernel.**

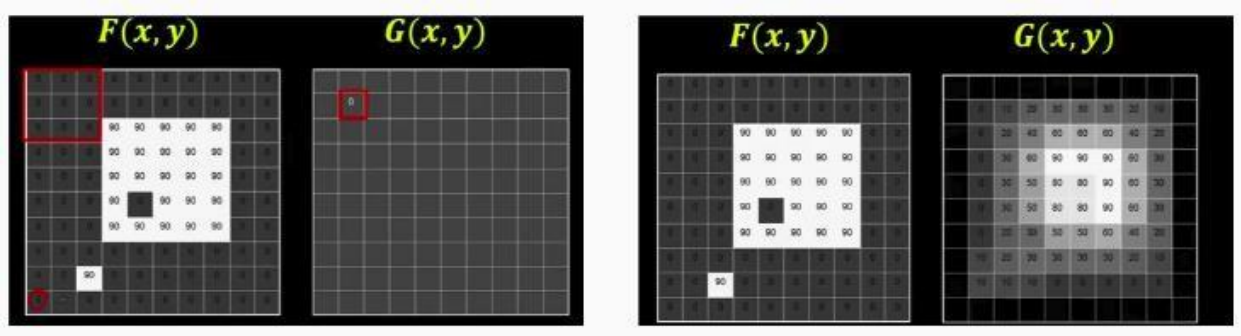


Figure 3.7 Calculating weighted moving averaging in a 2D image

The process used in filtering with uniform weights is also called correlation or **correlation filtering.**

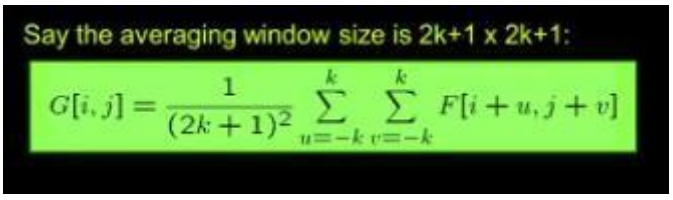


Figure 3.8 Correlation function for uniform weights

In correlation filtering with non-uniform weight, a function is used as non-uniform weights which is also called **mask or kernel** (function of the pixel values of the small sliding window)**.** The process used in it is called **cross-correlation.**

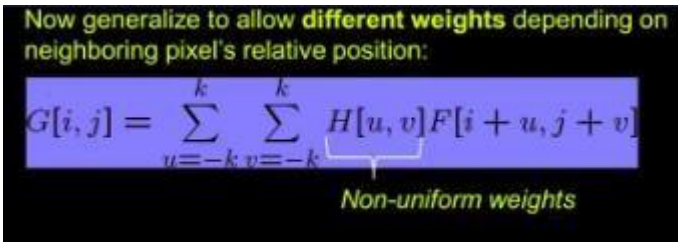


Figure 3.9 Correlation function for non-uniform weights

Types of Image noise filters:

There are different types of image noise filters. They can typically be divided into 2 types.

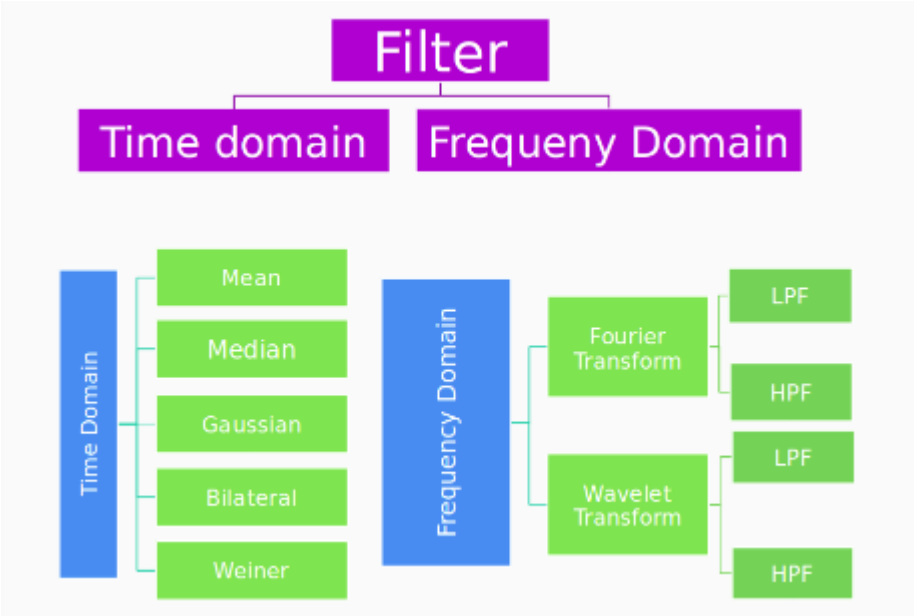


Figure 3.10 different types of noise filters

A low pass **filter** is applied on every random matrix frequently to obtained N random smooth pattern. System produce digital signature by applying signing process on digital **image**. Watermarking is also **used** for **image forgery** detection. In Checksum schema that it can add data into last most significant bit of pixels.

3.3 Double jpeg analysis

Since JPEG is the most widely used compression standard, detection of forgeries in JPEG images is necessary. In order to create a forged JPEG image, the image is usually loaded into a photo editing software, manipulated and then re-saved as JPEG. This yields to double JPEG compression artifacts, which can possibly reveal the forgery. Many techniques for the detection of double JPEG compressed images have been proposed. However, when the image is resized before the second compression step, the blocking artifacts of the first JPEG compression are destroyed. Therefore, most reported techniques for detecting double JPEG compression do not work for this case. In this paper, we propose a technique for detecting resized double JPEG compressed (called RDJPEG) images. We first identify features that can discriminate RD-JPEG images from JPEG images and then use Support Vector Machines (SVM) as a classification tool. Experiments with many RD-JPEG images with different quality and scaling factors indicate that our technique works well.

Due to the large number of available image processing tools, digital images can easily be altered without leaving visual evidence. Therefore, developing techniques for judging the authenticity of digital images became an urgent need. There are many types of image forgeries, which can be detected by different image forensic methods. Since JPEG is the most popular image type and it is supported by many applications, it is worthwhile to develop forensic techniques for JPEG images. Although there are many ways of making forgeries in a JPEG image, most share three main steps:

1) loading the JPEG image which is compressed by quality factor QF1 to a photo editing software, 2) manipulating this image and 3) re-compressing it as a JPEG file with quality factor QF2. Consequently, the forged image is doubly JPEG compressed (called D-JPEG). Detecting artifacts of double JPEG compression is an important step to judge whether a JPEG image is authentic. To this end, several techniques have been developed [2–8]. we found that when QF1 is different from QF2, periodic artifacts are present in the histograms of the DCT coefficients

of D-JPEG images. The periodicity can be recognized in the Fourier transform through peaks in the spectrum. expanded the global approach of by locating the tampered regions in D-JPEG images. Bianchi et al. proposed an enhanced version of, leading to an improvement of the accuracy of the algorithm. The authors in showed that the distribution of the most significant digit of the DCT coefficients in JPEG images follows the generalized Benford distribution. This distribution is very sensitive to double JPEG compression and this property can be applied to detect D-JPEG images. Chen et al. proposed a set of image features, which have subsequently been evaluated by a SVM based classifier. A limitation of these techniques is that they cannot detect D-JPEG images if the JPEG images are cropped before the second compression step is applied. The reason is that the corresponding blocking grids in the first compression and in the second compression are no longer aligned. In order to overcome this limitation, some other techniques have been proposed. In a blocking artifact characteristic matrix (BACM) is computed to measure the symmetric representation of the blocking artifacts introduced by JPEG compression. Since the symmetry of the BACM of a JPEG image is destroyed after the image is cropped, the BACM can be used as evidence for detecting cropped double JPEG compressed images. The authors of model the linear dependency of the "within-block" pixels (pixels that are not on the border of segmented 8×8 image blocks), compute the probability of the pixel being linearly correlated to its neighbors and form the map of the probabilities of all pixels in the image. The map is converted to Fourier domain and several statistical features from the different peak energy distribution are extracted in order to discriminate cropped D-JPEG images from non-cropped D-JPEG images. A simple yet reliable technique to detect the presence of cropped double JPEG compression has been introduced in this technique is based on the observation that the DCT coefficients exhibit an integer periodicity when they are computed according to the grids of the primary JPEG compression. Although work well for detecting cropped double JPEG compressed images, they are defeated if the images are resized before the second compression. The reason is that due to the effect of re-sampling, the blocking artifacts will be broken. The authors demonstrated the influence of resizing on the detection results of. To the best of our knowledge, there are only a few techniques for detecting resized double JPEG compressed (RD-JPEG) images. Kirchner and

Gloe apply a re-sampling detection technique (which was originally designed to work with uncompressed images) to JPEG images and analyze how the JPEG compression affects the detection output. A limitation is that the detection rates when applied to RD-JPEG images are very low if QF1 is much larger than QF2. Besides, if the JPEG images are down-sampled before the second compression, the technique is mostly defeated. The technique extracts neighboring joint density features and applies SVM to them. Although this technique works for both up-sampled images and down-sampled images by different interpolation methods, it is analyzed by the authors only for quality factors (both QF1 and QF2) of 75 and no informationon false positives is given. Bianchi and Piva proposed an algorithm, which can be summarizedby some steps: 1) estimate the candidate resizing factor; 2) for each candidate factor, undo theimage resizing operation and measure the NLDP (near lattice distribution property); 3) if the result is greater than a predefined threshold, label the images resized double JPEG compressed. Furthermore, the technique can estimate both the resize factor and the quality factor of the first JPEG compression of the analyzed image. The experimental results in show that it surpasses in the same test condition, but similar to, it seems more difficult to detect when QF1 is much larger than QF2. In this paper, we propose a new technique to detect RD-JPEG images. The technique first reveals specific features of JPEG images by using a re-sampling detector. These features are subsequently fed to SVM-based classifiers in order to discriminate RD-JPEG images from JPEG images. In comparison to, our technique does not require to distinguish in detail the peaks caused by JPEG compression from the peaks caused by re-sampling. In comparison to our approach does not need to extract complex image features for classification. The technique consists of some intricate steps, which mostly use for the purpose of reverse engineering of resized double JPEG compressed images. we briefly introduce state-of-the-art re-sampling detection techniques. Re-sampling detection is an important step of our technique and any of the mentioned re-sampling detectors can be used in our construction. The proposed detection algorithm for RD-JPEG images is explained in fig 3.11 and experimental results are shown.

To create a convincing forged image, the geometry of the image or some portions of it is often transformed. Once a geometric transformation (such as resizing or rotation) is applied to an

image, a re-sampling process is involved. Interpolation is the central step of re-sampling in order to estimate the value of the image signal at intermediate positions to the original samples. Based on specific artifacts created by interpolation, there are several techniques to detect traces of re-sampling in digital images. Gallagher realized that low-order interpolated signals introduce periodicity in the variance function of their second derivatives. Based on this observation, the author proposed a technique to detect whether an image has been re-sampled. A limitation of this technique is that it works only in the case of image resizing. Using the Radon transform, Mahdian and Saic improved so that their technique can detect not only image resizing, but also image rotation. Popescu noted that there are linear dependencies between neighboring pixels in re-sampled images. These correlations can be determined by using the Expectation/Maximization (EM) algorithm. The output of the algorithm is a matrix indicating the probability of every image sample being correlated to its neighbors (called p-map). The p-map of a resampled image usually contains periodic patterns, which are visible in the Fourier domain. A drawback of is that its computational complexity due to the use of the EM algorithm. Based on some improved techniques have been introduced. The author in showed that the p-map of a re-sampled image is periodic and this periodicity does not depend on the prediction weights that are used to compute the correlations of neighboring pixels. Therefore, he used a predefined set of prediction weights to compute the correlation probability of every image sample and designed a fast re-sampling detection technique which bypasses the EM algorithm. Although the values of prediction weights do not affect the periodicity of the pmap in theory, the authors in found that the selected set of predefined weights can strongly affect the obtained results: using one predefined set of weights for detection, peaks can be recognized in the transformed p-map, but using another set, peaks are not evident (though the periodicity exists in theory). Therefore, they use a predefined set, which is chosen through experimentation and apply the Radon transformation to the probability map of the analyzed image in order to enhance the frequency peaks and consequently the robustness of the overall technique. An example for detection of (uncompressed) re-sampled images by using EM algorithm and resampling techniques is presented in Fig. 3.11-3.14.

Figure 3.11 detection of(uncompressed) re-sampled images

Figure 3.12 resampled image

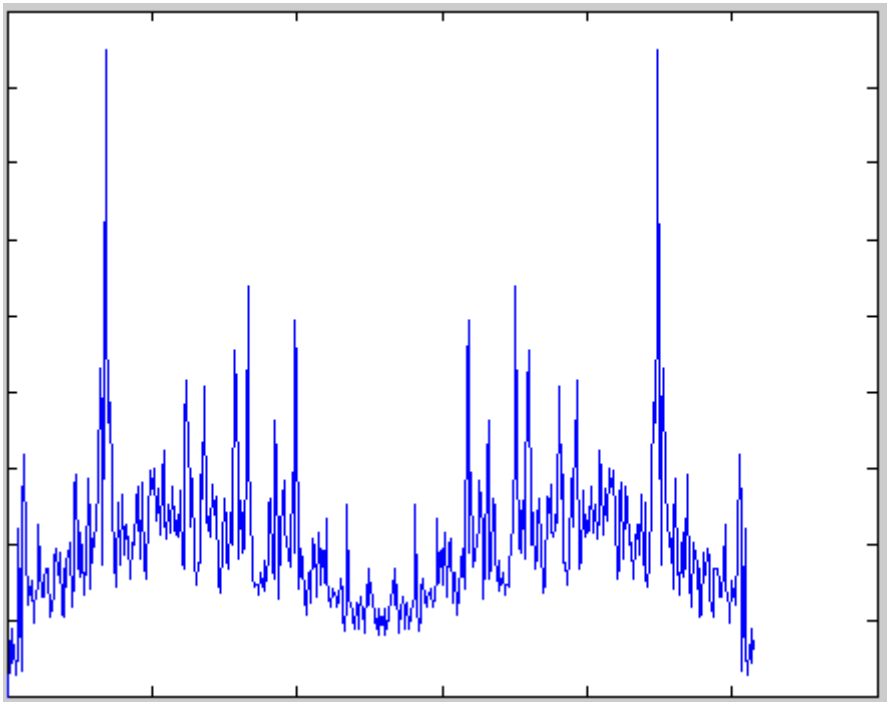


Figure 3.13 output without sampling

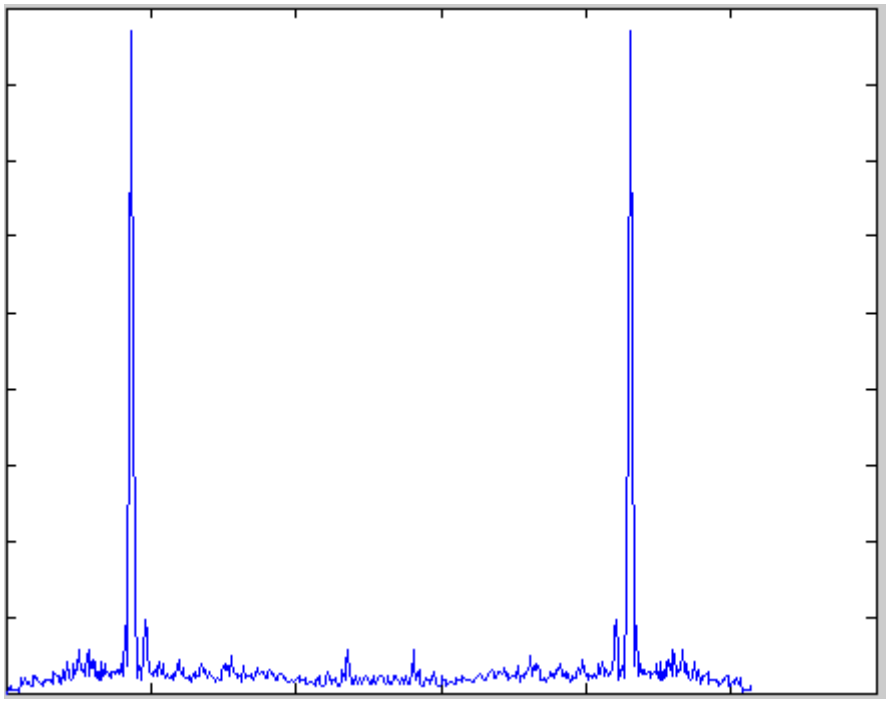


Figure 3.14 output with sampling

The techniques work well for detecting traces of re-sampling in uncompressed images. However, they fail when applied to JPEG images. The reason is that JPEG compression has an effect similar to nearest neighbor's interpolation and the resampling been re-sampled, yet the spectrum when applying the re-sampling detector contains strong peaks. In the next section, we propose a technique which uses one of these re-sampling detectors as the first step for detecting RD-JPEG images. Although any mentioned resampling detection technique in this section can be used, we choose this technique because of its efficiency as well as its speed.

When using data sets to detect re-sampling in both JPEG images and RD-JPEG images, we empirically found that the detection results of RD-JPEG images seem to have more peaks than those of JPEG images This is because the detection result of a RD-JPEG image contains not only the peaks introduced by JPEG compression, but also the peaks due to re-sampling. Nevertheless, the difference is not always easy to recognize by human eyes. Besides, it is necessary to automatically classify RD-JPEG images from JPEG images. To this end, we first apply the technique to JPEG images, and then extract the values of maximal peaks from the normalized Fourier spectrum. The extracted features are subsequently fed to SVM-based classifiers in order to discriminate RD-JPEG images from JPEG images. Since SVM is only a

binary classifier, we use two different approaches to design SVM classifiers for detecting RD-JPEG images. In the first approach, we design a single SVM classifier for directly distinguishing JPEG and RD-JPEG images, compressed by different quality factors. To this end, the features of a set of JPEG images and their re-sampled versions (the number of JPEG and re-sampled JPEG images are the same) are extracted for training a SVM classifier. This approach is simple and suitable for many situations in practice when we do not know the quality factors of the analyzed images. However, through experiments, reported in results, we find that this technique works well mostly when QF1 is lower than the QF2. The second approach is based on the idea that while QF1 of a double JPEG compressed image is usually not known to the analyst, QF2 can reliably be computed from the bitstream of the JPEG image. Thus, instead of using one single classifier for all quality factors, we design several different SVM classifiers, each of which distinguishes JPEG and RD-JPEG images for one specific value of QF2. Once the last quality factor of an analyzed JPEG image is known, the corresponding classifier will be applied to it. The method to design a classifier for a particular QF2 is similar to the first approach: we first use a set of JPEG images and another set of RDJPEG images (the numbers of images in both sets are the same and every image is compressed by QF2) and then extract image features for training. In other words, the last quality factor of a tested image is first identified, and then the image will be analyzed by the corresponding qualifier. In next section, we discuss experimental results for both approaches.

Experimental Results

First, we randomly choose 200 uncompressed images from the UCID image database. We create 5 datasets of JPEG images by compressing the uncompressed images with the quality factors of 40, 50, 60, 70, and 80. The JPEG images are subsequently resized by a scaling factor of 1.2 and recompressed by different factors of 40, 50, 60, 70, and 80. As a result, we obtained 5 datasets of 1000 RD-JPEG images corresponding to each dataset of JPEG images. To test the first approach, we create a single SVM classifier by using two groups of JPEG images and RD-JPEG images (with the scaling factor of 1.2) for training. After the training process (presented in Section 3) we apply the classifier to test RD-JPEG images. In training, we consider two cases of different quality factors: 1) 100 JPEG images compressed by a quality factor of 50 and100 RD-JPEG images re-compressed by a quality factor of 70 (QF1=50, QF2=70 and scaling factor =1.2) and 2) 100 JPEG images compressed by a quality factor of 70 and 100 RD-JPEGimages re-compressed by a quality factor of 80 (QF1=70, QF2=80 and scaling factor =1.2). Analyzing the detection results (see Table 1 and Table 2), we found that the technique works

well for detecting RD-JPEG images where QF1 is smaller than QF2. Otherwise, when QF1 is larger than QF2, the detection rate is reduced. In our experiments, the false positive rates (computed by testing the classifier on datasets of JPEG images which have been compressed by different quality factors of 40, 50, 60, 70, and 80) are lower than 10% in the first case and lower than 8% in the second case. In a more realistic scenario, we test the techniques on the RD-JPEG images, which have been resized with a different scaling factor than the factors are used in the training process. The datasets are created in the same way as above, except the scaling factor 1.1 is used instead of 1.2 (i.e. QF1=70, QF2=80 and scaling factor =1.1). Although the detection results are clearly worse compare with different data sets, we found that the degradation is not significant; therefore, the technique can potentially work in case the scaling factor is unknown. In the second approach, we consider 5 different cases corresponding to a QF2 of 40, 50, 60, 70, and 80. The case of QF2=40, we organize the training images into two groups: a group of 100 JPEG images (the quality factor of 40) and the other group of 100 RD-JPEG images (QF1=50, QF2=40 and scaling factor=1.2). The extracted features are used to train a SVM classifier that can be used to detect RD-JPEG images which compressed by the QF2 of 40. We repeat this process for the other cases when QF2 is 50, 60, 70, and 80. The detection results in testing RD-JPEG datasets are presented as we considered. We noticed that following the second approach, the technique works well even if QF1 is larger than QF2. The false positive rates are lower than 10% (9%, 8%, 5%, 6% and 3% when testing JPEG images compressed by the quality factors of 40, 50, 60, 70, and 80 respectively). Since JPEG compression with a lower factor produces stronger peaks in the Fourier spectrum, it obtains higher false positives.

3.4 SUPPORT VECTOR MACHINE

SVM is mainly used for classification purpose. In order to avoid computational complexity, it uses recognition tools by high dimension. There are many studies using SVM as a classifier in image forgery Detection. In this paper, a technique is developed to detect image forgery which includes removal, addition, and replacement of regions in an image. SVM classifier is used to find similar regions of an image by matching image blocks. Image, texture pixel value based features and edges are extracted to analyze the images for forged regions. After analyzing the image, hash values are calculated for feature extraction. This process consists of two phases, SVM train a set of identify the decision boundaries in the training phase and then the technique

will give the good generalization in high dimensional input images. Classification using SVM is mainly based on the concept of decision making and that defines the decision boundaries. A decision plane separates a set of objects having different class memberships and a set of objects having different class relationships. SVM determine a vectors called "support vectors"that easily identify the separators which gives the wide separation of classes and objects. SVM classifier supports both the binary and multiclass targets. Support Vector Machine models must have a similar functional form for block based network and radial basis functions, both are well-knowned data mining techniques. Since, neither of these algorithms has the very new theoretical approach to regularize the format, that forms the basis of SVM. The quality of generalization and ease of training in SVM is based on the capacities of those traditional methods. The SVM map the original data points from the input image to the high dimensional, feature block making classification problem simpler in feature space. This kind of mapping is done by a suitable choice of a kernel basis function.

Some researchers have pointed out that statistical and machine learning models are less different conceptually. Several new computational and machine learning methods are based on parameter estimation in statistics. Among these methods, support vector machines have attracted most interest in the last few years. Support vector machine (SVM) is a novel learning machine introduced first by Vapnik. It is based on the structural risk minimization (SRM) principle. In the last few years, SVM has yielded excellent generalization performance on a wide range of problems including bioinformatics, text categorization, fault diagnosis, image detection, power system, financial analysis, etc. In this article, we are interested in evaluating the performance of the SVM approach for supplier credit index prediction and selection in comparison with that of BPNN. A simple description of the SVM algorithm is provided here; for more details, the underlying algorithm of the class of supervised learning methods is to learn from observations. There are input space U ⊆ Rn, an output space Y, and a training set. S = ((u1, y1), (u2, y2),...,(ul, yl)), where l is the size of the training set. SVM belongs to the type of maximal margin classifier, in which the classification problem can be represented as an optimization problem, as shown in follows minw,bw, ws.t.yi(w, φ(ui+ b) 1 i = 1, 2,...,l (1) Vapnik illustrates how training a support vector machine for pattern recognition and regression leads to a quadratic optimization problem with bound constraints and one linear equality constraint[10]. Since the number of training examples determines the size of the problem, using standard quadratic problem solvers (Matlab toolbox) will easily make the computation impossible for large size training sets. Various methods have been proposed on solving the

quadratic programming problem in SVM, including gradient ascent methods, chunking and decomposition and Platt's sequential minimal optimization (SMO) algorithm (which extended the chunking approach to the extreme by updating two parameters at a time). max W(α) =li=1 αi − 1 2 l i,j=1 yiyjαiαjK(ui, uj) s.t. l i=1 yiαi = 0, αi> 0, i = 1, 2...,l (2) where, kernel function K(ui, uj) is the computation of inner product φ(ui), φ(uj) which maps the original data into a higher-dimension space and makes the input data set linearly separable in the transformed space. For noisy data, slack variables are introduced to relax the hard-margin constraints to allow for some classification errors[17], as shown in Eq.(3), where, noise level C >0 determines the tradeoff between the empirical error and the complexity term. minw,b,ξw, w Cn i=1 ξis.t.yi (w,φ(ui) + b) 1 − ξiξi 0, i = 1, 2,...,l (3) By solving the quadratic optimization, we obtain the support vector ui∈ SV s(αi> 0), where, SVs represents support vectors, therefore, the regressive function is f(u) = ui∈SV αiyiK(u, ui) (4) We experiment with different SVM parameter settings on the credit index data, including the noise level C and different kernel functions (including the linear, polynomial, radial basis, and sigmoid function), and experiment with different approaches for regression using SVM. The final SVM setting use a radial basis kernel function K(ui, uj) = exp(−γ|ui − uj | 2), γ = 0.2, and C=500. These parameter values achieve a relatively better performance on the supplier credit index data set.

Recently, there has been a decline in reliance on support vector machines (SVM), particularly kernel SVMs, as they require real-valued vectors. Users and researchers are instead turning to machine learning systems as end-to-end learning models. In general, deep learning architectures usually source the classifier function and feature representations solely from training examples, but we employed a linear SVM toward the end of every deep convolution neural network branch. We then trained them jointly by applying a backpropagation algorithm and stochastic gradient descent (input→→convNet→→SVM→→output). Note that the linear SVM has been given a linear activation function (similar to a regression function), with the hinge loss replacing the loss function, as follows: L(yˆi , yi) = max(0, 1 − yiyˆi) (1) where yi∈ [−1, 1] and yˆi = actual output. It is also possible to use a stochastic sub-gradient descent rather than an SGD, given that the hinge-loss cannot be differentiable. The training, thus, occurs end-to-end, with the hinge-loss error signal guiding the convNet and SVM weight learning. Furthermore, because hinge-loss causes the linear units to connect with learning maximum margin hyperplanes, linear SVMs are the outcome. As an example, a linear unit is connected by the hinge-loss, resulting in a single linear SVM + a trained convNet as feature detection after training. In this setup, the SVM is designed to learn the final splitting hyperplane and the

convNet is designed to learn the hierarchical features. Following the training procedure, a Heaviside step function can then be applied against the linear SVM output to obtain a binary output. The idea of using SVM in forgery detection is based on capturing the difference before knowing for certain that the patch is forged, which requires the use of a one-class SVM. This entity, which has been trained using feature vector h that is extracted from input images, quickly learns pristine feature distribution versus the forgery feature distribution. Next, it outputs a soft value that indicates the degree of possibility that the feature vector h is pristine or forged. The soft mask M0 is then defined as a matrix that has identical dimensions to the image, with every entry having a soft SVM output corresponding to image patches at identicalpositions. A final detection binary mask M can be obtained by employing the soft mask M0 forthresholding. More details will be provided in the following section. A summary of the copy- move forgery detection (CMFD) strategies employed in the present work is given in this section. Figure 3.3 illustrates the pipeline for the existing technique. As shown in the figure, our existing approach employs two distinct networks for forgery detection. One of the networksis GAN-based and detects any symptom of forgery, while the other locates any similarities existing in the image. As a joint network, the proposed method then locates any copy-move forgery found in the imagery target, demarcating the original source from the copy-moved region of the image. In CMFD-related tasks, the input data proceed through the two networks,GAN and CNN, and are then assigned to a linear classifier based on the proposed model selection. The CNN, in general, maintains feature extraction and ability to generate features versus strong discrimination skills, therefore the proposed model is used for data generation, feature extraction, data discrimination, and data classification.

With the increase in digital image forgery, the need for forgery detection algorithms has increased. In this paper, copy-move and splicing forgery detection are done at the same time. The suspicious image is taken as an input and then extract the features in it. Using SVM and PCA algorithm the original region of an image is eliminated. Then duplicate regions of an image is highlighted and marked as an output. In future work, we focus on accuracy of detecting duplicate regions.

3.5 SVM Disadvantages

- Choosing a “good” kernel function is not easy.
- Long training time for large datasets.

- Difficult to understand and interpret the final model, variable weights and individual impact.
- Since the final model is not so easy to see, we cannot do small calibrations to the model hence it's tough to incorporate our business logic.
- The SVM hyper parameters are Cost -C and gamma. It is not that easy to fine-tune these hyper-parameters. It is hard to visualize their impact

3.6 SVM Application

- Protein Structure Prediction
- Intrusion Detection
- Handwriting Recognition
- Detecting Steganography in digital images
- Breast Cancer Diagnosis
- Almost all the applications where ANN is used

CHAPTER 4

PROPOSED U-NET METHOD

4.1 U-Net and CNN

Artificial Intelligence has been witnessing a monumental growth in bridging the gap between the capabilities of humans and machines. Researchers and enthusiasts alike, work on numerous aspectsof the field to make amazing things happen. One of many such areas is the domain of Computer Vision.

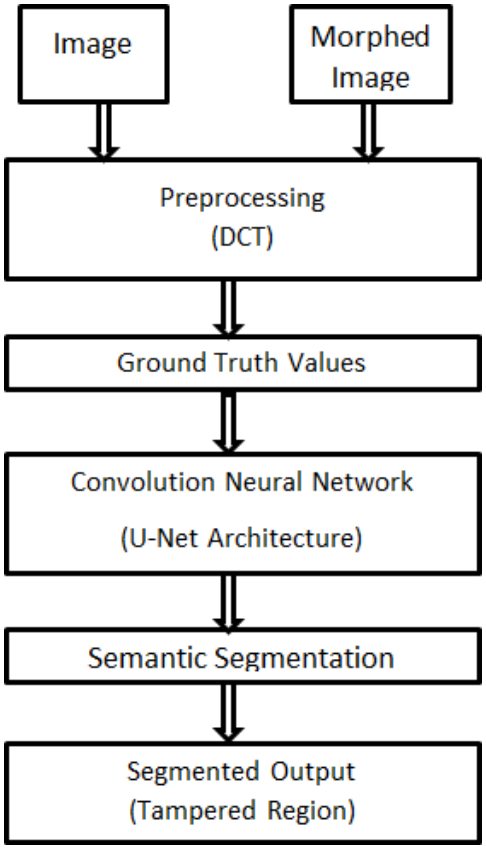


Figure 4.1 The block diagram of proposed method

4.1.1 Convolutional Neural Network

The agenda for this field is to enable machines to view the world as humans do, perceive it in a similar manner and even use the knowledge for a multitude of tasks such as Image & Video recognition, Image Analysis & Classification, Media Recreation, Recommendation Systems, Natural Language Processing, etc. The advancements in Computer Vision with Deep Learning hasbeen constructed and perfected with time, primarily over one particular algorithm — a Convolutional Neural Network.

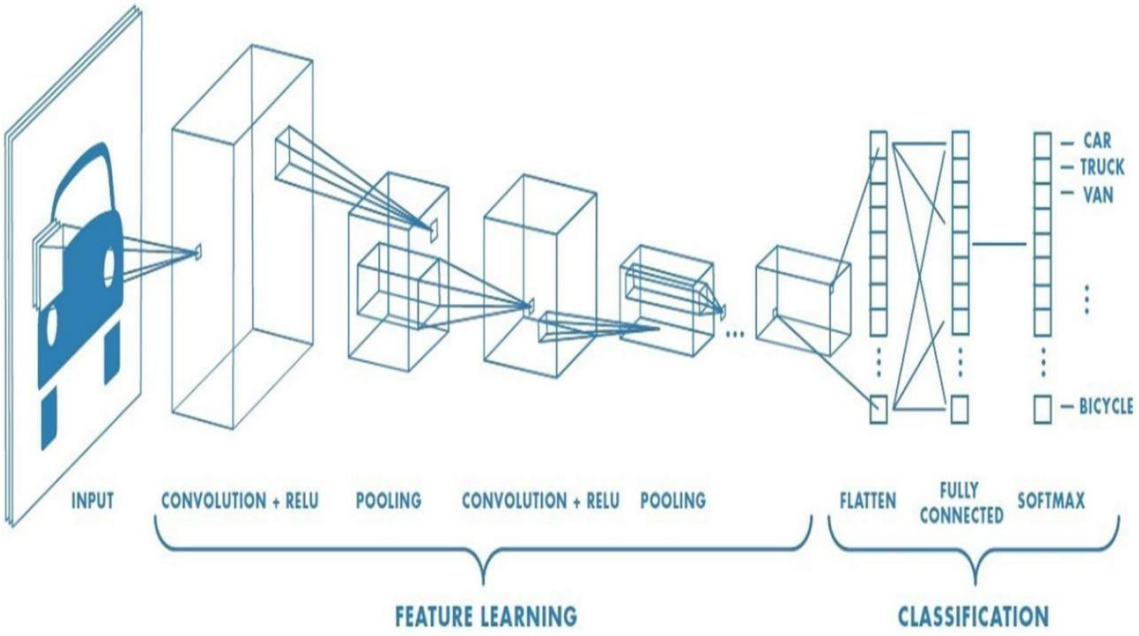


Figure 4.2 Convolutional neural network

A Convolutional Neural Network (Convent/CNN) is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other. The pre- processing required in a ConvNetis much lower as compared to other classification algorithms.While in primitive methods filters are hand-engineered, with enough training, Convent's havethe ability to learn these filters/characteristics.

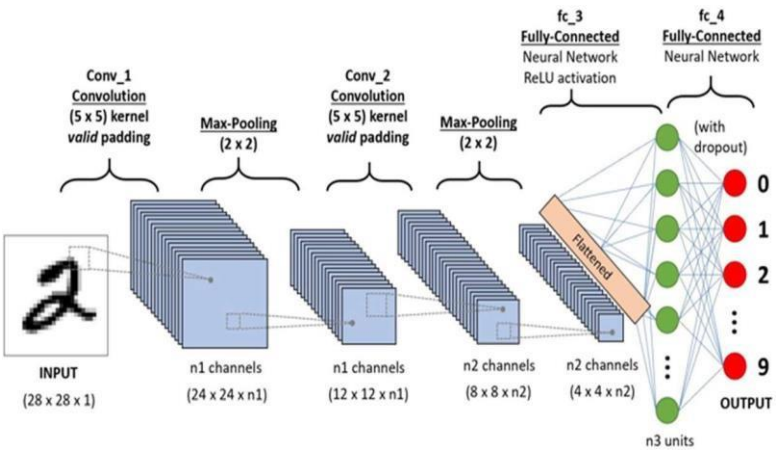


Figure 4.3 A CNN sequence to classify handwritten digits

The architecture of a ConvNet is analogous to that of the connectivity pattern of Neurons in the Human Brain and was inspired by the organization of the Visual Cortex. Individual neurons respond to stimuli only in a restricted region of the visual field known as the Receptive Field. A collection of such fields' overlaps to cover the entire visual area.

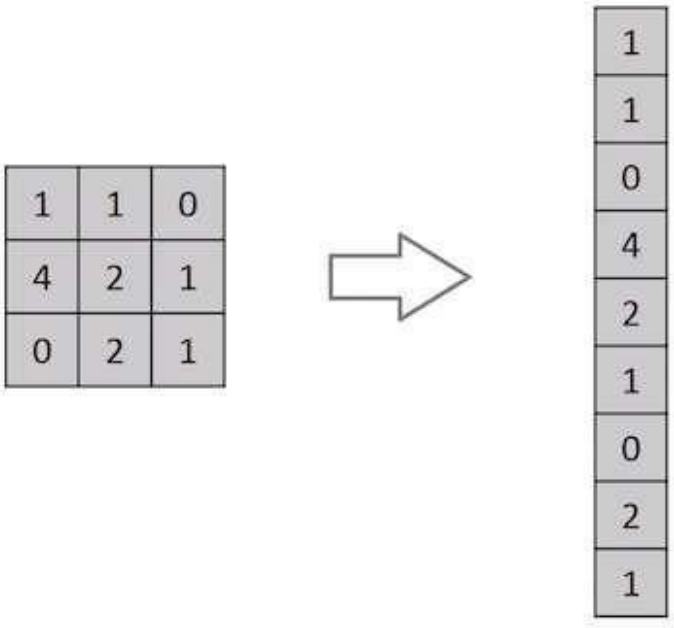


Figure 4.4 Flattening of a 3x3 image matrix into a 9x1 vector

An image is nothing but a matrix of pixel values, right? So why not just flatten the image (e.g.

3x3image matrix into a 9x1 vector) and feed it to a Multi-Level Perceptron for classification purposes?Uh. not really. In cases of extremely basic binary images, the method might show an average precision score while performing prediction of classes but would have little to no accuracy whenit comes to complex images having pixel dependencies throughout.

A ConvNet is able to successfully capture the Spatial and Temporal dependencies in an image through the application of relevant filters. The architecture performs a better fitting to the image dataset due to the reduction in the number of parameters involved and reusability of weights.In other words, the network can be trained to understand the sophistication of the image better.

4.1.2 Input Image

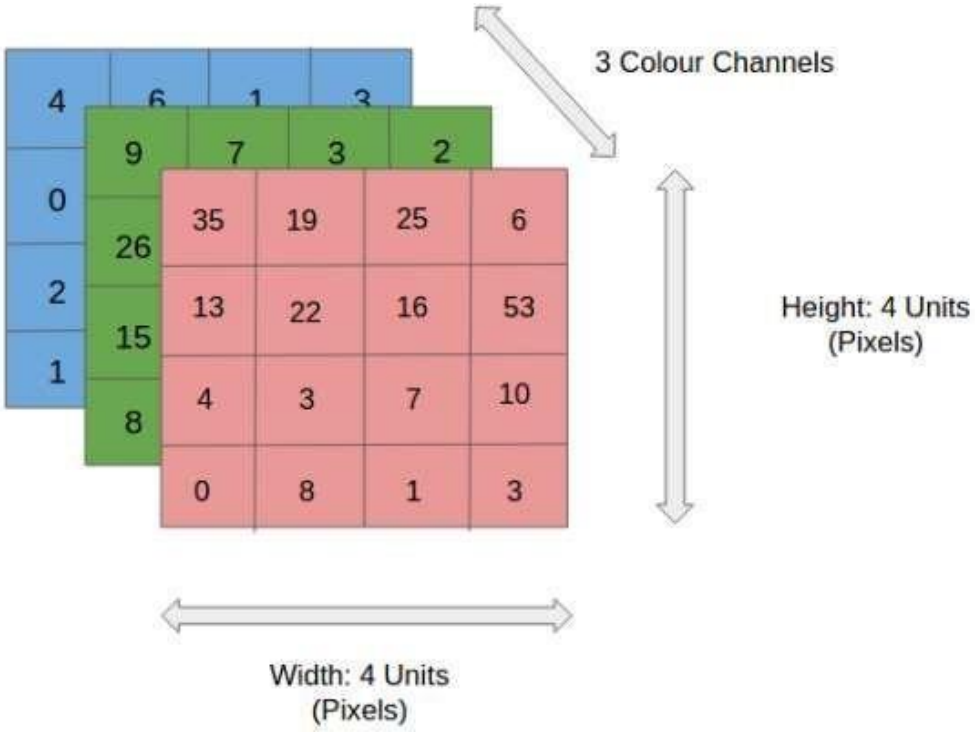


Figure 4.5 4x4x3 RGB Image

In the Figure 4.5, we have an RGB image which has been separated by its three-color planes — Red, Green, and Blue. There are a number of such color spaces in which images exist — Grayscale,RGB, HSV, CMYK, etc.

You can imagine how computationally intensive things would get once the images reach dimensions, say 8K (7680×4320). The role of the ConvNet is to reduce the images into a form which is easier to process, without losing features which are critical for getting a good prediction. This is important when we are to design an architecture which is not only good at learning featuresbut also is scalable to massive datasets.

4.1.3 Convolutional Kernel

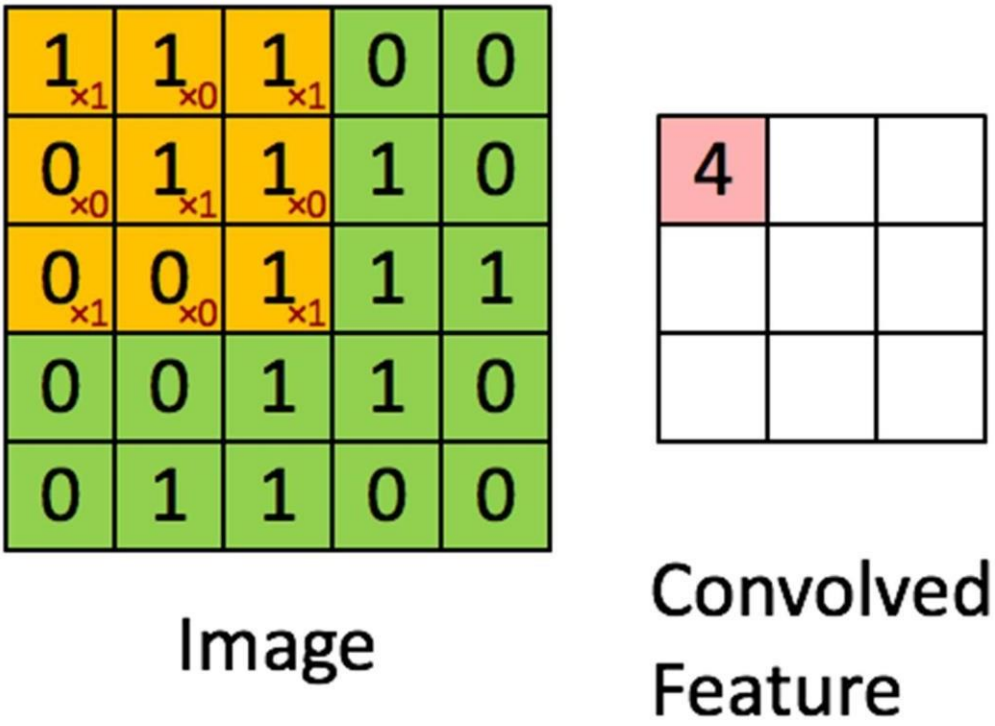


Figure 4.6 Convoluting a 5x5x1 image with a 3x3x1 kernel to get a 3x3x1 convolved feature

Image Dimensions = 5 (Height) x 5 (Breadth) x 1 (Number of channels, eg. RGB). In figure 4.6, the green section resembles our 5x5x1 input image, I. The element involved in carrying out the convolution operation in the first part of a Convolutional Layer is called the Kernel/Filter, K, represented in the color yellow. We have selected K as a 3x3x1 matrix.

The Kernel shifts 9 times because of Stride Length = 1 (Non-Stride), every time performinga matrix multiplication operation between K and the portion P of the image over which the kernel is hovering.

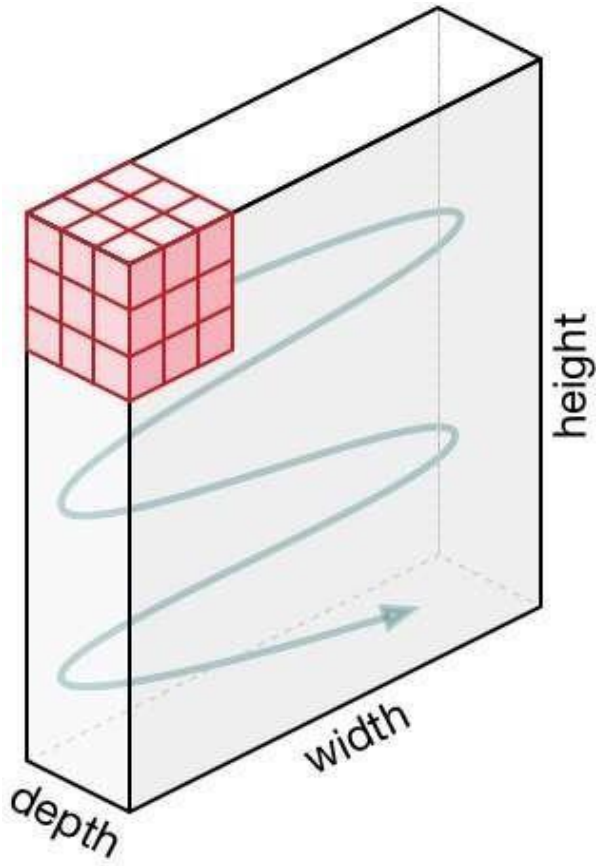


Figure 4.7 Movement of the Kernel

The filter moves to the right with a certain Stride Value till it parses the complete width. Moving on, it hops down to the beginning (left) of the image with the same Stride Value and repeats the process until the entire image is traversed.

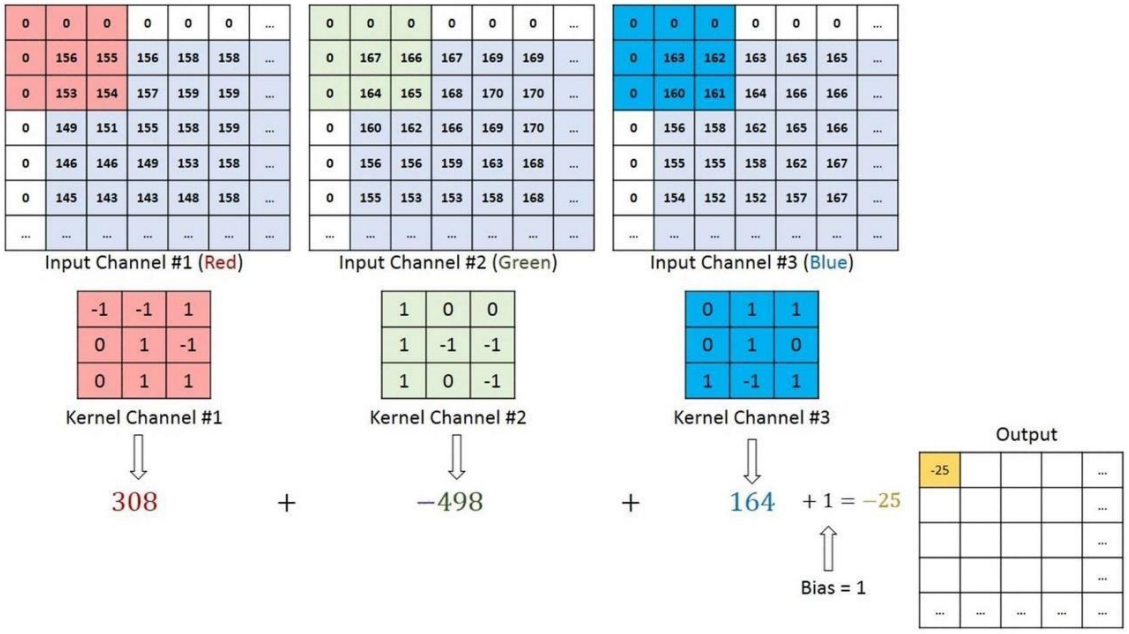


Figure 4.8 Convolution operation on a MxNx3 image matrix with a 3x3x3 Kernel

In the case of images with multiple channels (e.g. RGB), the Kernel has the same depth as that of the input image. Matrix Multiplication is performed between Kn and in stack ([K1, I1]; [K2, I2]; [K3, I3]) and all the results are summed with the bias to give us a squashed one-depth channel Convoluted Feature Output.

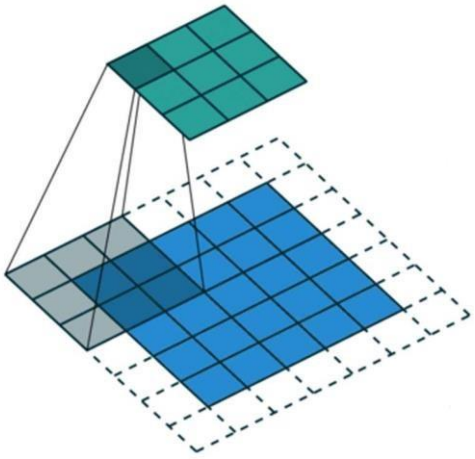


Figure 4.9 Convolution Operation with Stride Length = 2

The objective of the Convolution Operation is to extract the high-level features such as edges, from the input image. Convent's need not be limited to only one Convolutional Layer. Conventionally, the first ConvLayer is responsible for capturing the Low-Level features such as edges, color, gradient orientation, etc. With added layers, the architecture adapts to the High-Levelfeatures as well, giving us a network, which has the wholesome understanding of images in the dataset, similar to how we would.

There are two types of results to the operation — one in which the convolved feature is reduced in dimensionality as compared to the input, and the other in which the dimensionality is either increased or remains the same. This is done by applying Valid Padding in case of the former,or Same Padding in the case of the latter.

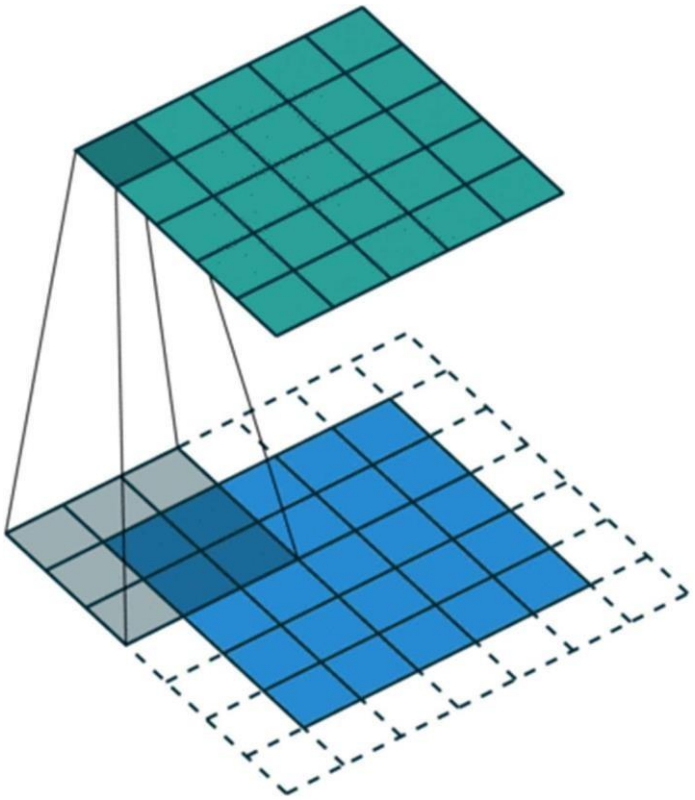


Figure 4.10 SAME padding: 5x5x1 image is padded with 0s to create a 6x6x1 image

When we augment the 5x5x1 image into a 6x6x1 image and then apply the 3x3x1 kernel over it, we find that the convolved matrix turns out to be of dimensions 5x5x1. Hence the name Same Padding. On the other hand, if we perform the same operation without padding, we are presented with a matrix which has dimensions of the Kernel (3x3x1) itself -Valid Padding.

4.1.4 Pooling Layer

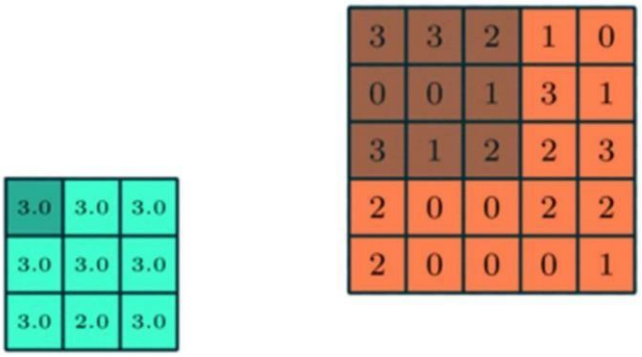


Figure 4.11 3x3 pooling over 5x5 convolved feature

Similar to the Convolutional Layer, the Pooling layer is responsible for reducing the spatial sizeof the Convolved Feature. This is to decrease the computational power required to process the datathrough dimensionality reduction. Furthermore, it is useful for extracting dominant features which are rotational and positional invariant, thus maintaining the process of effectively training of the model. There are two types of Pooling: Max Pooling and Average Pooling. Max Pooling returns the maximum value from the portion of the image covered by the Kernel. On the other hand, Average Pooling returns the average of all the values from the portion of the image covered by theKernel.

Max Pooling also performs as a Noise Suppressant. It discards the noisy activationsaltogether and also performs de-noising along with dimensionality reduction. On the other hand, Average Pooling simply performs dimensionality reduction as a noise suppressing mechanism. Hence, we can say that Max Pooling performs a lot better than Average Pooling.

The Convolutional Layer and the Pooling Layer, together form the i-th layer of a Convolutional Neural Network. Depending on the complexities in the images, the number of such layers may be increased for capturing low-levels details even further, but at the cost of more computational power.

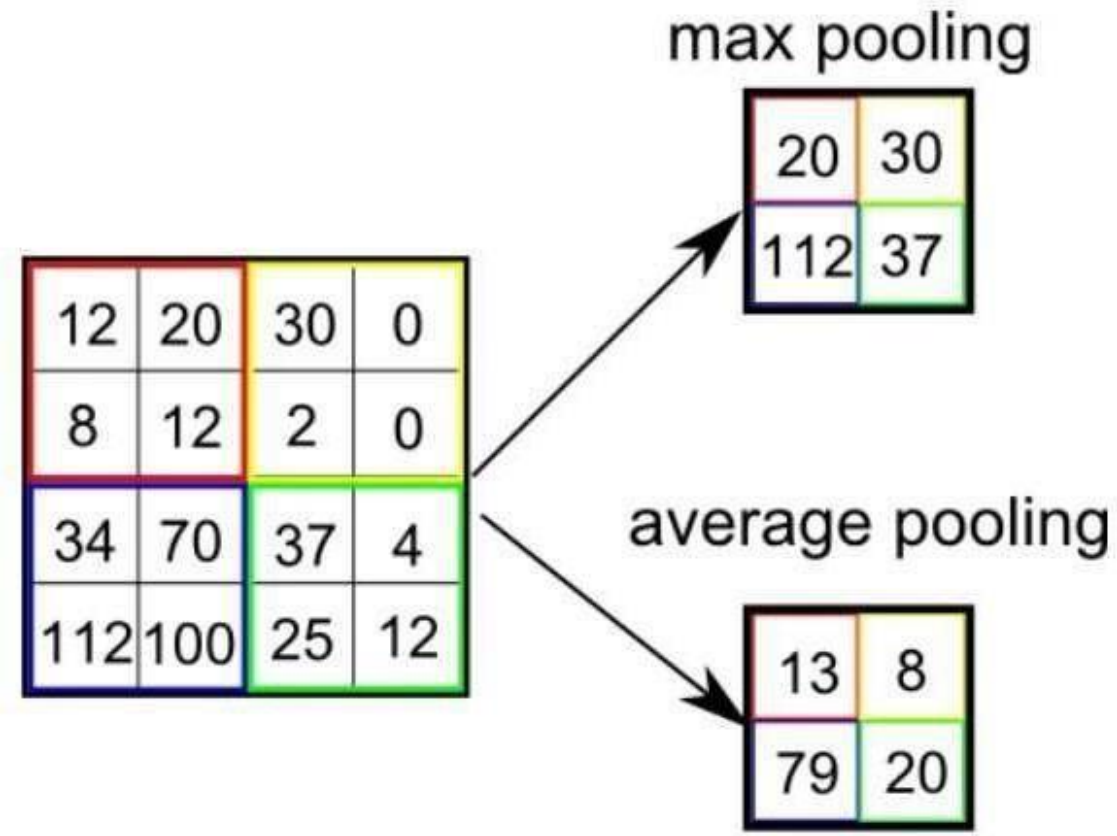


Figure 4.12 Types of Pooling

After going through the above process, we have successfully enabled the model to understand the features. Moving on, we are going to flatten the final output and feed it to a regular Neural Network for classification purposes.

4.1.5 Fully Connected Layer

Adding a Fully-Connected layer is a (usually) cheap way of learning non-linear combinations of the high-level features as represented by the output of the convolutional layer. The Fully-Connected layer is learning a possibly non-linear function in that space.

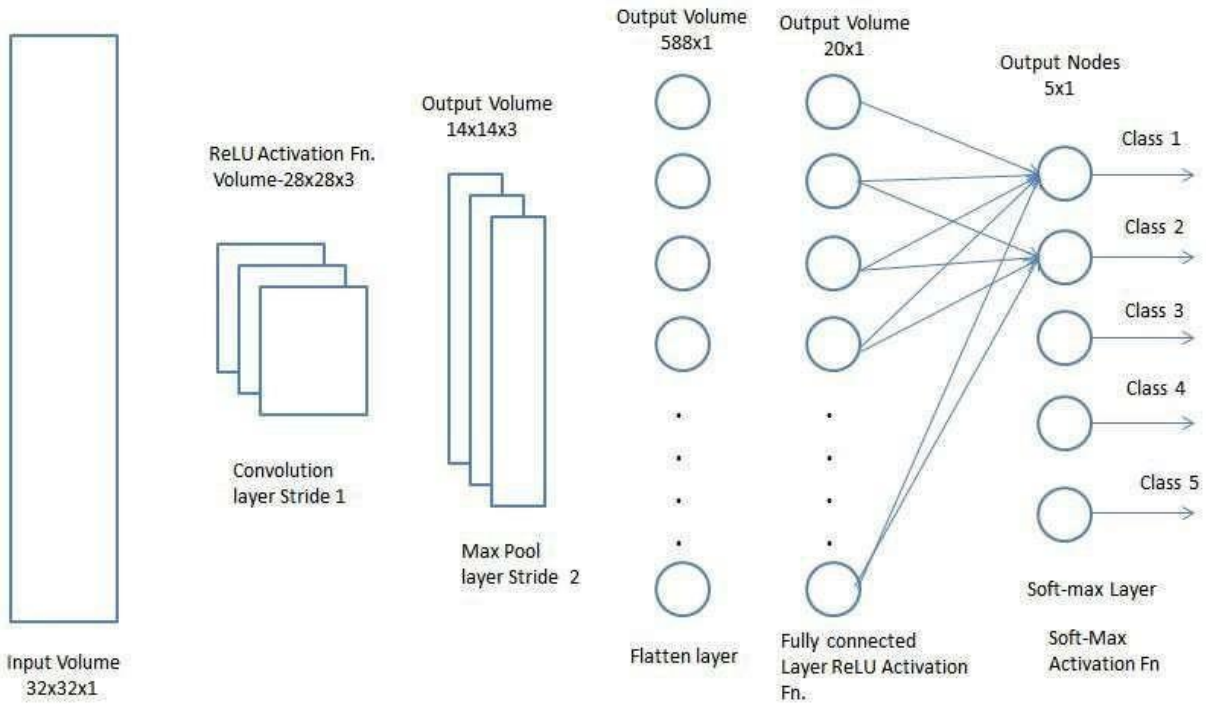


Figure 4.13 Fully connected layer

Now that we have converted our input image into a suitable form for our Multi-Level Perceptron, we shall flatten the image into a column vector. The flattened output is fed to a feed- forward neural network and backpropagation applied to every iteration of training. Over a seriesof epochs, the model is able to distinguish between dominating and certain low-level features in images and classify them using the SoftMax Classification technique.

U-Net, a kind of Convolutional Neural Networks (CNN) approach, was first proposed by Olaf Ranneberger, Phillip Fischer, and Thomas Brox in 2015 with the suggestion of better segmentation on biomedical images. Basically, segmentation is a process that partitions an image into regions. It is an image processing approach that allows us to separate objects and textures in images. Segmentation is especially preferred in applications such as remote sensing or tumor detection in biomedicine.

There are many traditional ways of doing this. For example; point, line, and edge detection methods, thresholding, region-based, pixel-based clustering, morphological approaches, etc. Various methods have been developed for segmentation with convolutional neural networks (a common deep learning architecture), which have become indispensable in tackling more advancedchallenges with image segmentation. In this post, we'll take a closer look at one such

architecture: u-net. In deep learning, it's known that we need large datasets for model training. But there are some problems we run into at this point! We often cannot afford the amount of data that needs to be collected for an image classification problem. In this context, affordability means time, money, and most importantly, hardware.

For example, it isn't possible to collect many biomedical images with the camera on your mobile phone. So, we need more systematic ways to collect data. There's also the data labeling process, for which a single developer/engineer will not suffice—this will require a lot of expertise and experience in classifying the relevant images. This is especially true with highly-specialized areas such as medical diagnostics. Another critical point is to provide education about the general image in classically convolutional neural networks through class labels. However, some problems require knowledge of localization/positioning with pixel-based approaches. In areas that require sensitive approaches, such as biomedical or defense, we need class information for each pixel.

U-Net is more successful than conventional models, in terms of architecture and in terms pixel-based image segmentation formed from convolutional neural network layers. It's even effective with limited dataset images. The presentation of this architecture was first realized through the analysis of biomedical images.

4.2 Differences That Make U-NET Special

As it's commonly known, the dimension reduction process in the height and width that we apply throughout the convolutional neural network—that is the pooling layer— is applied in the form of a dimension increase in the second half of the model.

The pooling layer reduces height and width information by keeping the number of channels of the input matrix constant. The calculation is a step used to reduce complexity (Each element of the image matrix is called a pixel). In summary, the pooling layer refers to a pixel that represents groups of pixels.

Note: Pooling layers can work with different approaches, including maximum, average, or medianlayers.

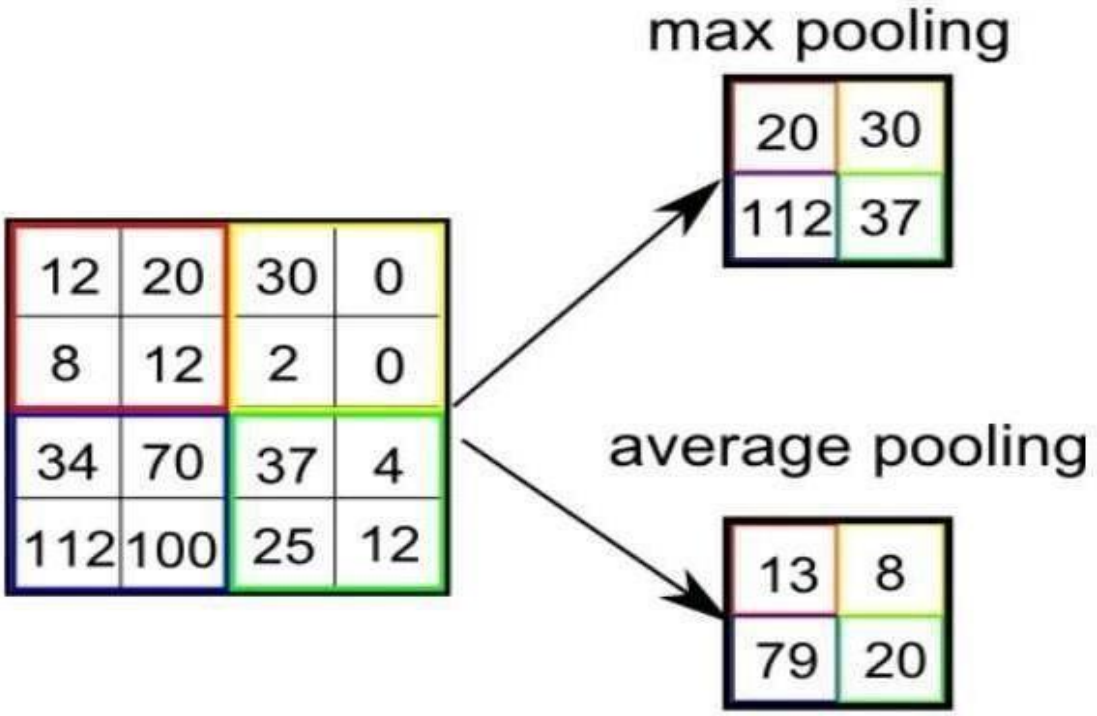


Figure 4.14 Representation: Max and Avg. Pooling

These layers are intended to increase the resolution of the output. For localization, the sampled output is combined with high-resolution features throughout the model. A sequential convolution layer then aims to produce a more precise output based on this information.

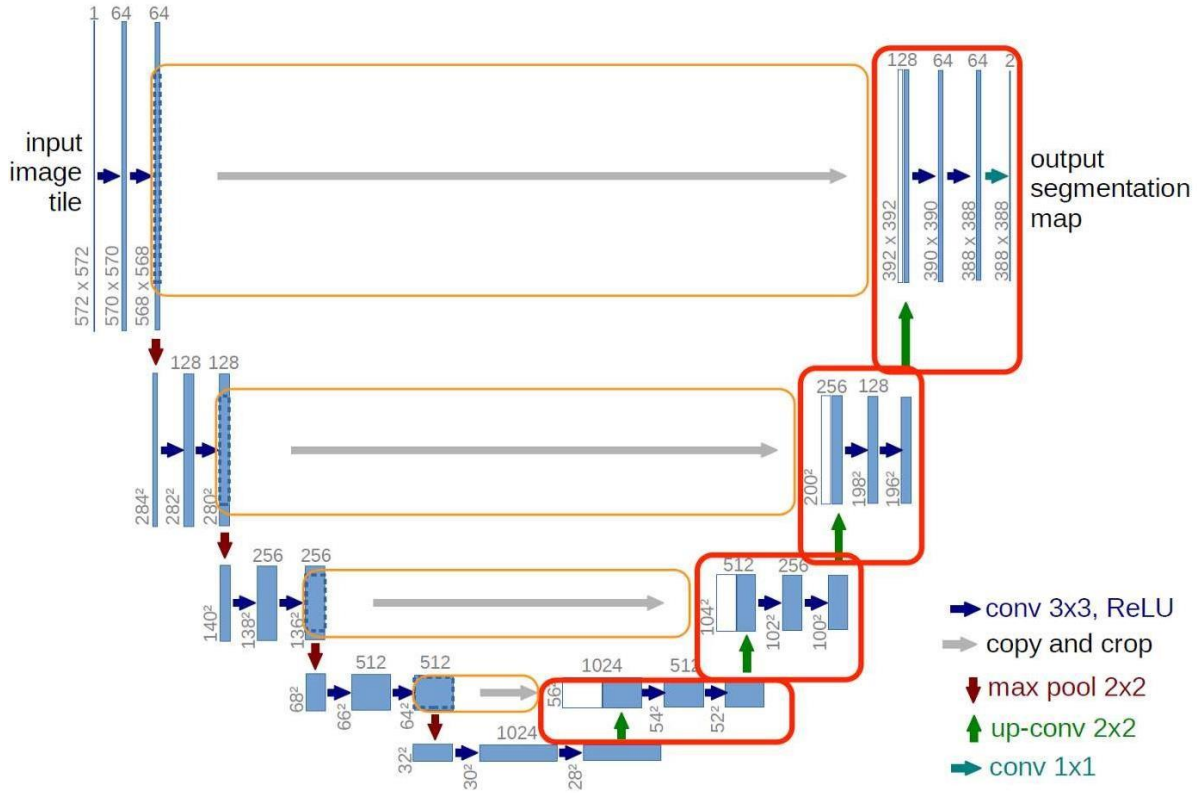


Figure 4.15 U-Net Model

U-Net takes its name from the architecture, which when visualized, appears similar to the letter U, as shown in the Figure 4.15. Input images are obtained as a segmented output map. The most special aspect of the architecture in the second half. The network does not have a fully-connected layer. Only the convolution layers are used. Each standard convolution process is activated by a ReLU activation function.

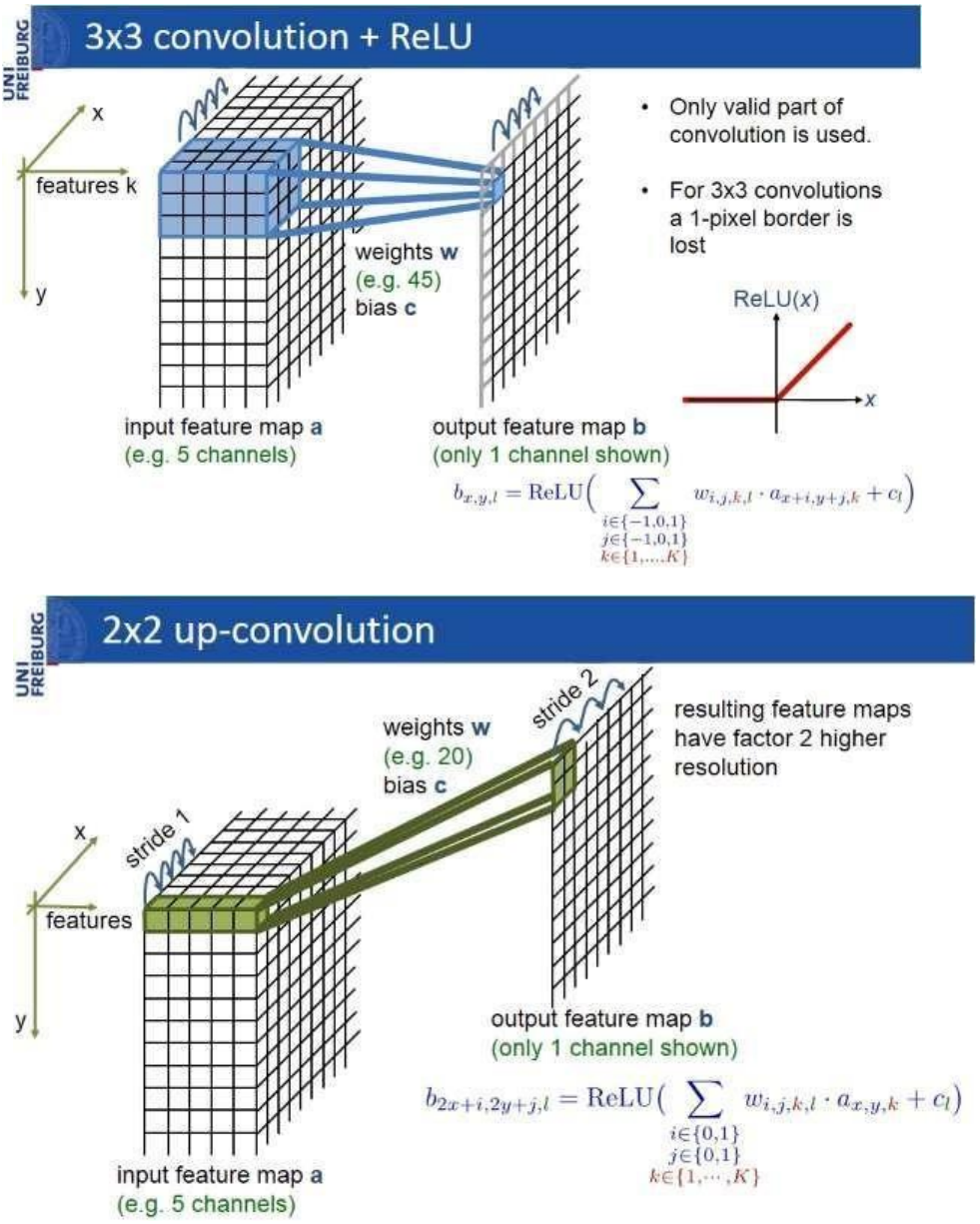


Figure 4.16 Representation of a convolution and deconvolution process in U-Net

The pixels in the border region are symmetrically added around the image so that images can be segmented continuously. With this strategy, the image is segmented completely. The padding (pixel adding) method is important for applying the U-Net model to large images; otherwise, the resolution will be limited by the capacity of the GPU memory.

CHAPTER 5

MATLAB

5.1 Introduction to MATLAB

What Is MATLAB?

MATLAB is an elite dialect for specialized registering. It incorporates calculation, representation, and programming in a simple to-utilize condition where issues and arrangements are communicated in natural numerical documentation. Run of the mill utilizes incorporate

- Math and calculation
- Algorithm advancement
- Data obtaining
- Modeling, reenactment, and prototyping
- Data examination, investigation, and representation
- Scientific and designing illustrations
- Application advancement, including graphical UI building

MATLAB is an intuitive framework whose important records thing is an exhibit that doesn't require dimensioning. This allows you to address several specialized processing issues, specifically people with framework and vector data, in a small quantity of the time it might take to compose a program in a scalar non intuitive dialect, as an instance, C or FORTRAN. The name MATLAB stays for grid studies facility. MATLAB modified into to start with composed to offer simple access to framework programming created thru the LINPACK and EISPACK ventures.

Today, MATLAB vehicles fuse the LAPACK and BLAS libraries, putting the cutting element in programming for network calculation.

MATLAB has advanced over a time of years with contribution from several customers. In college situations, it's miles the same old educational system for early on and propelled courses in arithmetic, designing, and technological know-how. In industry, MATLAB is the tool of selection for high-profitability research, development, and exam. MATLAB highlights a collection of extra utility-precise preparations called device booths. Important to most clients of MATLAB, tool kits allow you to examine and practice particular innovation. Tool compartments are exhaustiveaccumulations of MATLAB capacities (M-records) that extend out the MATLAB situation to take care of precise instructions of problems. Territories in which device stash are reachable incorporateflag dealing with, manage frameworks, neural systems, fluffy purpose, wavelets, recreation, and several others.

5.1.1 The MATLAB System

The MATLAB machine includes five predominant parts.

5.1.2 Development Environment

This is the set of tools and facilities that assist you use MATLAB functions and files. Many of those system are graphical consumer interfaces. It includes the MATLAB computing device and Command Window, a command history, an editor and debugger, and browsers for viewing help, the workspace, files, and the hunt direction.

5.1.3 The MATLAB Mathematical Function

This is a wonderful collection of computational algorithms starting from essential features like sum, sine, cosine, and complex mathematics, to extra sophisticated features like matrix inverse, matrix Eigen values, Bessel competencies, and fast Fourier transforms.

5.1.4 The MATLAB Language

This is an excessive-degree matrix/array language with manipulate go with the go with the flow statements, competencies, statistics systems, input/output, and object-oriented programming talents. It permits every programming in the small to hastily create quick and dirty throw-away

programs, and programming in the big to create entire massive and complicated software applications.

5.1.5 Graphics

MATLAB has huge facilities for showing vectors and matrices as graphs, further to annotating and printing those graphs. It consists of immoderate-degree capabilities for two-dimensional and three-dimensional facts visualization, image processing, animation, and presentation images. It additionally includes low-stage functions that allow you to completely customize the appearanceof image graphs further to assemble whole graphical consumer interfaces on your MATLAB packages.

5.1.6 The MATLAB Application Program Interface (API)

His is a library that lets in you to write C and FORTRAN packages that engage with MATLAB.It includes centers for calling workout routines from MATLAB (dynamic linking), calling MATLAB as a computational engine, and for analyzing and writing MAT-documents.

5.2 MATLAB Working Environment

5.2.1 MATLAB Desktop

MATLAB Desktop is the precept MATLAB utility window. The computing device consists offive sub windows, the summon window, the workspace software, the prevailing catalog window, the order records window, and at the least one Figure home windows, which are confirmed simply whilst the patron indicates a realistic.

The order window is the vicinity the consumer types MATLAB orders and expressions on the provoke (>>) and where the yield of these costs is shown. MATLAB characterizes the workspace because the association of factors that the purchaser makes in a work consultation. The workspace application demonstrates these elements and a few information approximately them. Double tapping on a variable in the workspace software dispatches the Array Editor, which can be applied to get statistics and wage instances regulate certain homes of the variable.

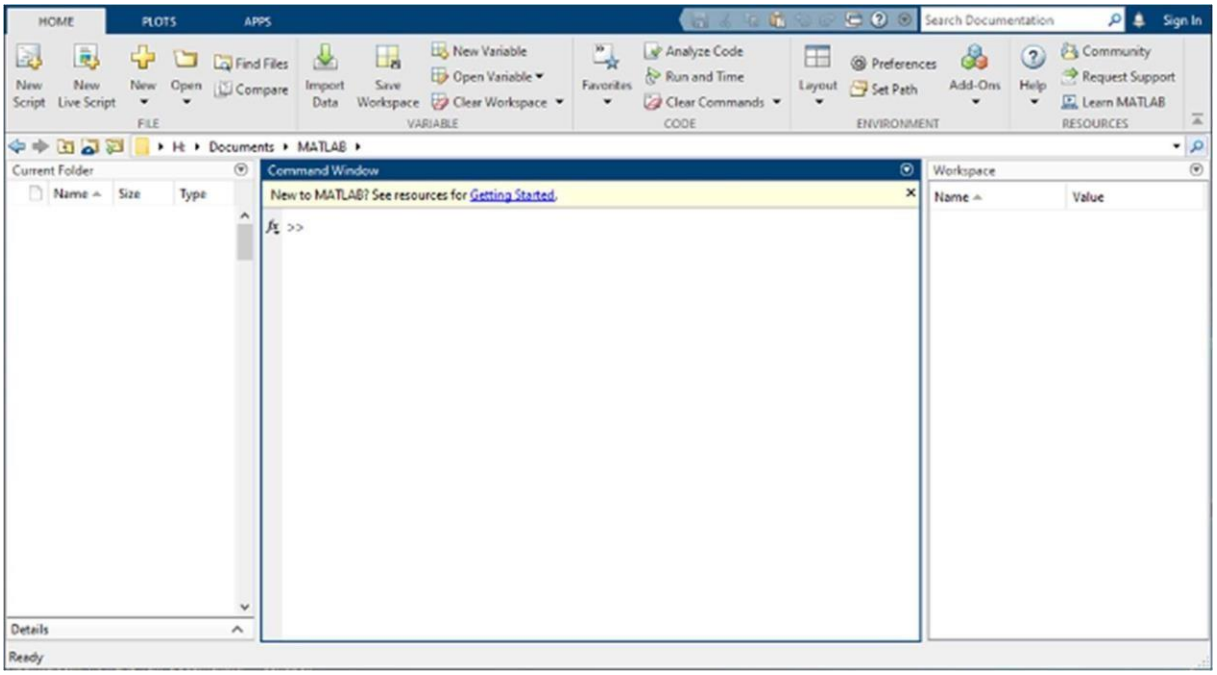


Figure 5.1: MATLAB Environment.

The gift Directory tab over the workspace tab demonstrates the substance of the prevailing registry, whose way is seemed within the gift index window. 1For case, in the home windows working framework the manner may be as consistent with the subsequent: C:MATLABWork, demonstrating that registry work is a subdirectory of the number one catalog MATLAB, that's added in force C. Tapping on the bolt inside the gift index window demonstrates a rundown of as of late utilized approaches. Tapping at the capture to at least one aspect of the window allows the customer to exchange the prevailing catalog.

MATLAB utilizes an inquiry way to discover M-information and other MATLAB associated documents, which might be kind out in catalogs within the PC report framework. Any report holds going for walks in MATLAB must reside in the ebb and go with the flow registry orin an index this is on are looking for manner. Of course, the facts provided with MATLAB and math works device kits are integrated into the inquiry manner. The least traumatic approach tolook which indexes are at the inquiry manner. The best method to peer which catalogs are quickly the search way, or to consist of or adjust an inquiry manner, is to pick out set way fromthe File menu the computing device, and after that utilization the set way change container. It is extremely good exercise to feature any normally applied catalogs to the pursuit manner to preserve a strategicdistance from again and again having the exchange the existing index.

The Command History Window incorporates a file of the orders a customer has entered in the rate window, together with each gift and beyond MATLAB classes. Already entered MATLAB orders may be selected and re-done from the fee history window by using right tapping on a sum on or association of orders. This interest dispatches a menu from which to choose exclusive picks however executing the orders. This is helpful to pick special choices notwithstanding executingthe summons. This is a precious element even as attempting different things with distinct ordersin a work consultation.

5.2.2 Using the MATLAB Editor to Create M-Files

The MATLAB manager is both a word processor specific for making M-information and a graphical MATLAB debugger. The proofreader can display up in a window without everybody else, or it could be a sub window inside the desktop. M-statistics are supposed by means of the growth .M, as in pixelup.m. The MATLAB editorial manager window has diverse draw down menus for errands, as an example, sparing, seeing, and troubleshooting files. Since it plays out some primary tests and furthermore makes use of shading to separate among exceptional components of code, this content material device is suggested because the apparatus of choice for composing and changing M-capacities. To open the proofreader, sort adjust at the incite opens the M-report filename. In a supervisor window, organized for changing. As referred to before, the file has to be within the momentum catalog, or in an index inside the pursuit way.

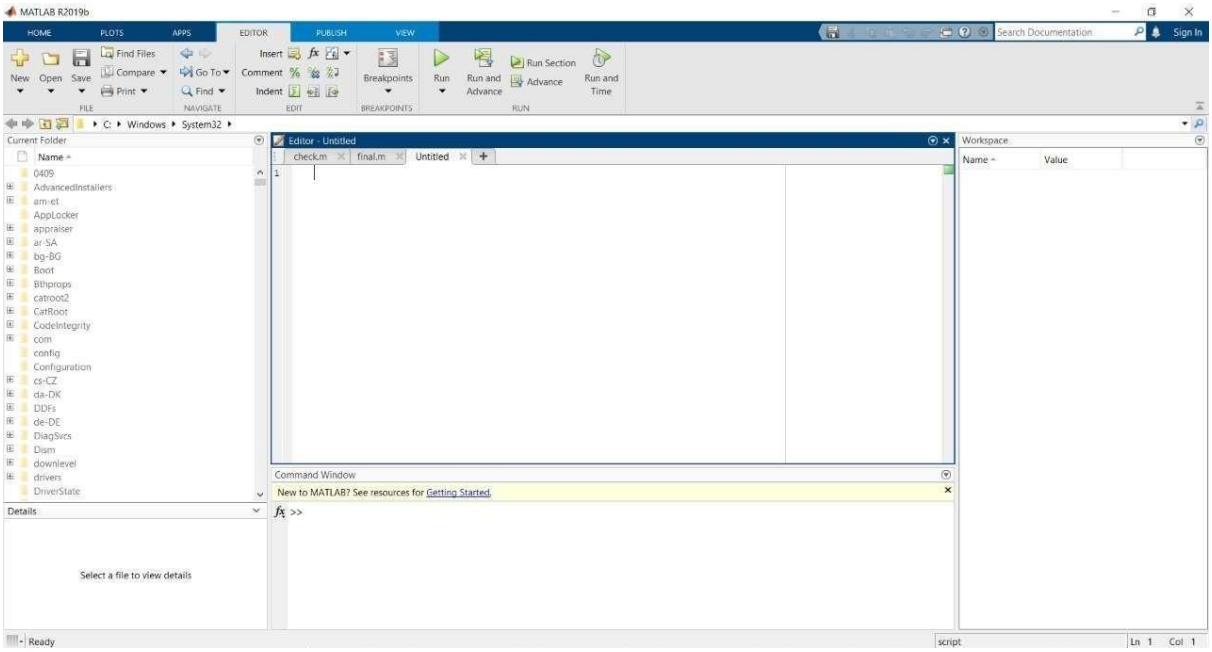


Figure 5.2 MATLAB Editor to create M-Files

5.2.3 Getting Help

The important approach to get help online is to utilize the MATLAB help program, opened as a different window either by tapping on the question mark image (?) on the desktop toolbar, or by writing help program at the provoke in the order window. The assistance Browser is a web programcoordinated into the MATLAB desktop that shows a Hypertext Markup Language (HTML) records. The Help Browser comprises of two sheets, the assistance pilot sheet, used to discover data, and the show sheet, used to see the data.

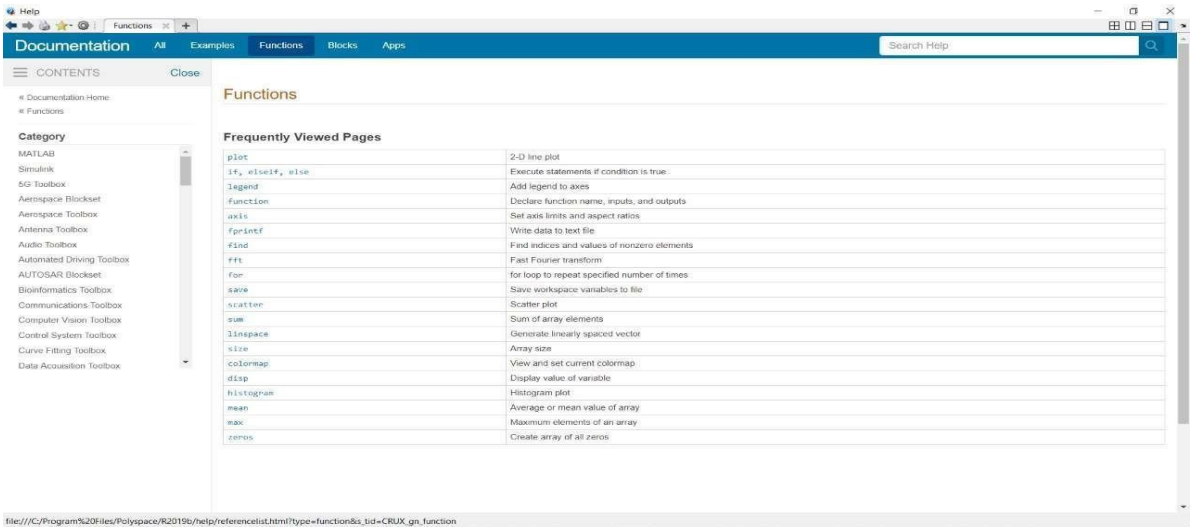


Figure 5.3 MATLAB Help Screen

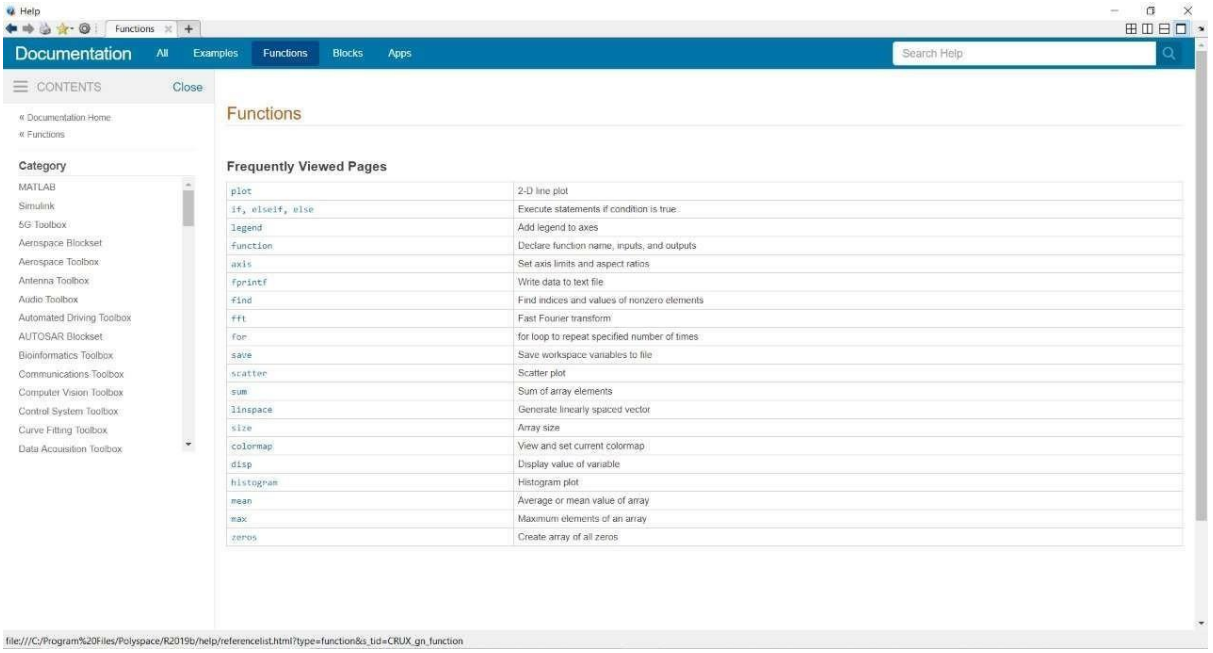


Figure 5.4 MATLAB Functions

CHAPTER 6

IMAGE PROCESSING

6.1 Introduction

Image segmentation objectives to split the desired foreground object from the historical past. Sincecoloration and texture in herbal pics are very complicated, automated segmentation of foreground objects from the complex heritage meets a vast impediment while foreground and history have similar functions. To cope with the problem, interactive image graph segmentation includes simpleuser interplay into image segmentation in a supervised or semi-supervised manner, and hasacquired a lot interest in current years. Interactive image segmentation extracts foreground objects from the complex historical past by using taking gain of the consumer's interactive enter. User interplay is employed to get prior information from customers which then play a critical role in guiding image segmentation. Thus, the project is to efficiently employ the constrained but treasured interactive inputs. In this painting, we attempt to maximize the guiding power of the given interactive statistics via propagating their characteristics through the whole image.

There have been increasing physical games within the exam institution to create smart self-loader image department structures. In the creators exhibited an intelligent image graph division system in light of diagram cut. Clients marked seed pixels which demonstrating clean foundation and frontal place have been applied as stable priors for portioning images into parent and floor. In the creators validated that diagram cut primarily based department calculations can be actualized brief. In the creators exhibited a division given incomplete amassing imperatives approach. Client inputs were applied as inclination to a characteristic collecting system, and the creators deliberate such one-sided collecting problem as a compelled advancement difficulty that proliferates scanty incomplete gathering facts to the unlabeled information by way of upholding gathering smoothness and reasonableness on the marked records focuses. They utilized the standardized reduce version and tackled the enhancement issue by using Eigen decay.

In the creators exhibited an intuitive image vanguard extraction approach that become computationally in light of diagram cut off however the creators provided a less complex customer collaboration gadget to decrease patron endeavors inside the connection procedure and an iterative model refreshing technique to enhance exactness. In an intuitive leading-edge basis division technique became provided on the subject of image tangling. The creators utilized Belief Propagation to iteratively engender customer named pixels to the unlabeled pixels.

A current work has built up an enhancement-based Figure ground division procedure, where a straightforwardness image was registered by improving a quadratic cost work with client provided direct limitations. The improvement issue has a one of a kind worldwide least and can be illuminated effectively by standard numerical strategies. In this paper, we acquaint measurable priors as imperatives with take care of the enhancement issue. For a few images, the factual priors can give sufficient requirements to consequently get palatable Figure ground division comes about. For more troublesome cases, client cooperation is important. In such cases, we utilize the division result in view of the measurable priors as a beginning stage for intelligent Figure ground division, and as a manual for help clients to put the requirements in the right areas to produce the coveted outcomes.

Along these lines, the factual priors control the client as well as enable diminishing clients' works in the connection to prepare. We have likewise built up another technique to make twofold (hard) division in view of the processed nonstop straightforwardness image. Another commitment of this paper is the augmentation of the streamlining based intelligent Figure ground division structure to intuitive multi-class division, where client can give multi-class seed pixels rather than simply frontal area foundation 2-class seeds, for dividing the given image into the coveted numberof areas by playing out a one shot advancement operation, which again has a one of a kind worldwide least and can be acquired by unraveling a vast arrangement of straight conditions.

6.2 Digital Image Processing

Computerized image preparing is a range portrayed by the requirement for broad test work to buildup the practicality of proposed answers for a given issue. A critical trademark hidden the

plan of image preparing frameworks is the huge level of testing and experimentation that typically is required before touching base at a satisfactory arrangement. This trademark infers that the capacity to plan approaches and rapidly model hopeful arrangements by and large assumes a noteworthy part in diminishing the cost and time required to land at a suitable framework execution.

6.2.1 What Is DIP

An image might be characterized as a -dimensional capacity f (x, y), where x, y are spatial instructions, and the adequacy of f at any integrate of commands (x, y) is known as the strength or dark diploma of the image via then. Whenever x, y and the abundance estimations off are all restricted discrete quantities, we name the image automated image. The area of DIP alludes to getting ready superior image by way of the use of strategies for automated PC. Advanced image graph is constituted of a constrained huge style of additives, every of which has a selected locationand esteem. The components are referred to as pixels.

Vision is the maximum progressive of our sensor, so it isn't brilliant that image play genuinely the most vital part in human statement. Be that as it could, numerous to humans, who are limited to the visible band of the EM range imaging machines cover almost the entire EM variety, going from gamma to radio waves. They can work likewise on image produced by means of resources that human beings are not acclimated to companion with image.

There isn't any massive information amongst creators almost about wherein image handling stops and one of a kind related territory, for instance, image exam and PC imaginativeand prescient start. Now and then a qualification is made via characterizing image dealing with asa teach wherein each the information and yield at a procedure are images. This is constraining and to a few diplomas artificial restrict. The range of image investigation is within the center of image making equipped and PC imaginative and prescient.

There aren't any obvious limits in the continuum from image graph coping with within the route of one component to finish imaginative and prescient at the opposite. In any case, one useful worldview is to don't forget 3 types of mechanized strategies on this continuum: low, mid and peculiar kingdom bureaucracy. Low-degree way consists of primitive operations, as an example, image on the brink of decrease commotion, differentiate upgrade and image honing. A low- diploma procedure is described with the aid of the manner that both its resources of

information and yields are image graphs.

Mid-stage procedure on images includes assignments, as an instance, department, depiction of that query diminishes them to a frame affordable for PC dealing with and characterization of person articles. A mid-degree process is portrayed by way of the way that its resources of info via and massive are image graphs however its yields are properties eliminated from those image graphs. At lengthy remaining greater multiplied quantity coping with includes Understanding an outfit of perceived gadgets, as in image graph exam and on the furthest stop of the continuum gambling out the highbrow capacities generally connected with human vision. Advanced image managing, as correctly characterized is utilized successfully in a wide scope of areas of great social and economic esteem.

6.2.2 What Is an Image

An image is spoken to as a two-dimensional capacity f (x, y) where x and y are spatial co-ordinatesand the adequacy of f at any match of directions (x, y) is known as the power of the image by then.

6.2.3 Gray Scale Image

A grayscale image is a capacity I (xylem) of the two spatial directions of the image plane. I (x, y)is the force of the image at the point (x, y) on the image plane. I (xylem) take non-negative esteemsexpect the image is limited by a rectangle [0, a] '[0, b] I: [0, a] [0, b] ® [0, data).

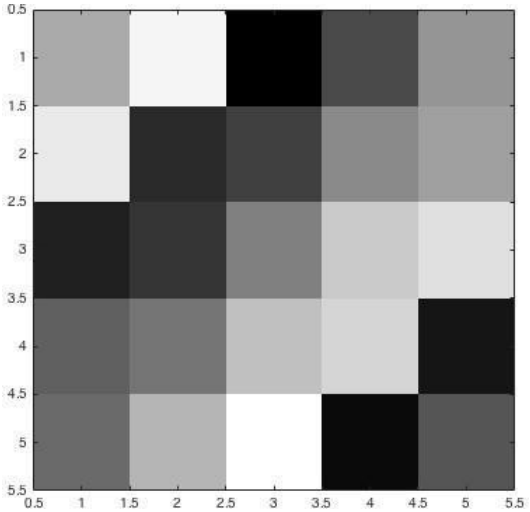


Figure 6.1: Grayscale image

6.2.4 Colour Image

It can be spoken to by way of three capacities, R (xylem) for purple, G (xylem) for green and B (xylem) for blue. An image is probably chronic as for the x and y allows and furthermore in adequacy. Changing over this type of image to advanced shape calls for that the directions and the adequacy to be digitized. Digitizing the facilitates esteems is called examining. Digitizing the adequacy esteems is known as quantization.

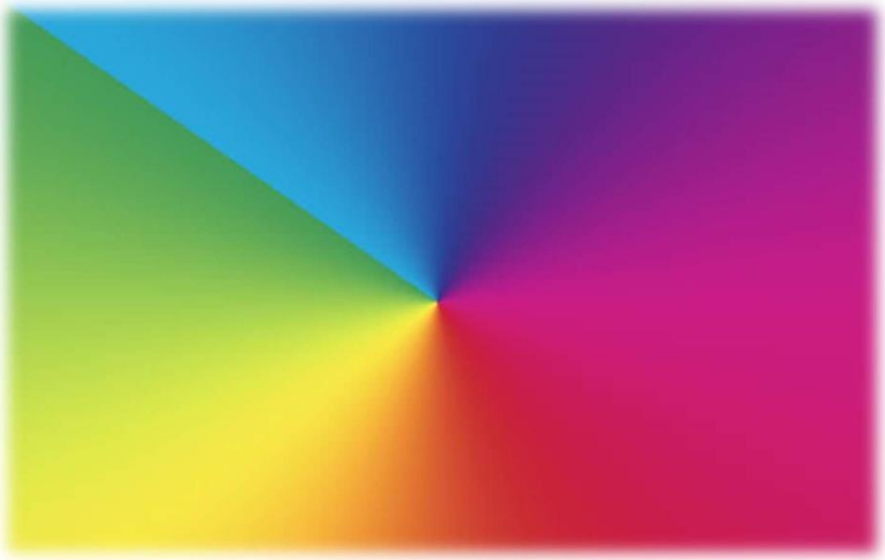


Figure 6.2 Colour Image

6.3 Coordinate Convention

The result of inspecting and quantization is a network of right numbers. We employ crucial methods to speak to advanced pix. Accept that an image graph f(x, y) is examined so that the subsequent image has M strains and N segments. We say that the image is of size M X N. The estimations of the pointers (xylem) are discrete quantities. For notational lucidity and lodging, we utilize complete variety esteems for these discrete pointers.

In many images dealing with books, the image start line is characterized to be at (xylem) = (zero, 0). The following direction esteems alongside the principle column of the image are (xylem) = (0, 1). It is vital to keep in mind that the documentation (0, 1) is utilized to suggest

the second instance alongside the predominant push. It doesn't suggest that these are the real estimations of physical instructions whilst the image graph changed into tested. Taking after parentdemonstrates the arrange way of life. Take notice of that x stages from 0 to M-1 and y from 0 to N-1 in whole number augmentations.

The arrange lifestyle applied as a part of the tool compartment to intend clusters is unique almost about the previous passage in minor methods. To start with, in choice to utilizing (xylem) the tool compartment utilizes the documentation (race) to reveal traces and segments. Note, anyhow, that the request of instructions is the same as the request tested within the beyond segment, as in the important difficulty of an organize topples, (alb), alludes to a line and the secondone to a segment. The other distinction is that the place to begin of the set-up framework is at (r,c) = (1, 1); because of this, r tiers from 1 to M and c from 1 to N in complete variety additions. IPT documentation alludes to the suggestions. Less frequently the device stash additionally uses another facilitate subculture referred to as spatial instructions which uses x to allude to sections and y to alludes to columns. This is the inverse of our utilization of things x and y.

6.4 Image as Matrices

The former exchange prompts the accompanying portrayal for a digitized image

work:f (0,0) f(0,1) … .. f (0,N-1)

f (1,0) f(1,1) … f(1,N-1)

f (xylem)= . . .

f (M-1,0) f(M-1,1) … f(M-1,N-1)

The correct element of this circumstance is a automated image through definition. Every detail of this show off is known as an image element, image element and pixel. The phrases image and pixel are implemented in the course of something is left of our exchanges to suggest a automatic image graph and its additives.

A computerized image can be spoken to normally as a MATLAB

grid:f (1,1) f(1,2) … . f(1,N)

f (2,1) f (2,2) … . f (2,N)

f = . . .

f (M,1) f(M,2) …. f (M, N)

Where f (1, 1) = f (0, 0) (take note of the utilization of a mono degree textual style to indicate MATLAB amounts). Obviously the two portrayals are indistinguishable, excluding the flow in beginning. The documentation f (p, q) shows the component located in line p and the phase q. For instance, f (6, 2) is the aspect in the sixth line and 2d section of the community f. Regularly we make use of the letters M and N in my view to mean the amount of traces and sections in a lattice. A 1xN grid is referred to as a line vector even as a Mx1 framework is known as a segment vector. A 1x1 grid is a scalar.

Lattices in MATLAB are positioned away in factors with names, for example, An, a, RGB,actual show off et cetera. Factors should start with a letter and contain simply letters, numerals andunderscores. As cited in the past section, all MATLAB quantities are composed utilizing mono- scope characters. We make use of ordinary Roman, italic documentation, for instance, f(x, y), for clinical expressions.

6.5 Reading Images

Images are perused into the MATLAB condition utilizing capacity imread whose punctuation is

Imread ('filename')

Arrange name Description perceived expansion

TIFF	Tagged Image File Format	.tif, .tiff
JPEG	Joint Image graph Experts Group	.jpg, .jpeg
GIF	Graphics Interchange Format	.gif
BMP	Windows Bitmap	.bmp
PNG	Portable Network Graphics	.png
XWD	X Window Dump	.xwd

Here filename is a spring containing the total of the image document (counting any appropriate augmentation). For instance, the charge line

>> f = imread ('8. jpg');

Peruses the JPEG (above table) image trunk beam into image cluster f. Take observe of theusage of single fees (') to delimit the string filename. The semicolon towards the end of a charge line is utilized by MATLAB for smothering yield, if a semicolon is excluded. MATLAB showsthe consequences of the operation(s) determined in that line. The incite image(>>) assigns the start of a fee line, as it suggests up within the MATLAB summon window.

6.6 Data Classes

In spite of the reality that we work with numbers organizes the estimations of pixels themselves aren't constrained to be entire numbers in MATLAB. Table above rundown exceptional information classes upheld by way of MATLAB and IPT are speak me to pixels esteems. The initial 8 sections inside the desk are alludes to as numeric facts instructions. The ninth segment is the singe magnificence and, as regarded, the remaining passage is alluded to as consistent data magnificence.

Every unmarried numeric calculation in MATLAB are carried out in twofold amounts, so that is likewise an everyday information magnificence experience in image making ready programs. Class unit eight moreover is skilled each once in a while, in particular while perusing information from stockpiles devices, as eight-bit pix are maximum normal portrayals found by andby means of. These two facts classes, classes shrewd, and, to a lesser diploma, magnificence unit sixteen represent the critical facts classes on which we center. Numerous Int capacities however bolster every one of the statistics classes recorded in table. Information magnificence twofold requires 8 bytes to talk to various uint8 and Int eight require one byte every, uint16 and int16 calls for 2 bytes and unit 32.

Int 32 and single required 4 bytes each. The burn information class holds characters in Unicode portrayal. A character string is simply a 1*n exhibit of characters legitimate cluster contains just the qualities 0 to 1, with every component being put away in memory utilizing capacity sensible or by utilizing social administrators.

6.7 Image Types

The toolbox supports four types of images.

A. Intensity images.

B. Binary images.

C. Indexed images.

D. R G B images.

Most monochrome image processing operations are carried out using binary or intensity images, so our initial focus is on these two image types. Indexed and RGB color images.

6.7.1 Intensity Images

An intensity image is a facts matrix whose values had been scaled to symbolize intentions. When the elements of an intensity photo are of class unit8, or class unit 16, they've integer values within the variety [0,255] and [0, 65535], respectively. If the image is of class double, the values are floating point numbers. Values of scaled, double depth images are inside the variety [0, 1] by way of using conference.

6.7.2 Binary Images

Binary images have an extraordinarily precise meaning in MATLAB.A binary photo is a logical array 0s and1s.therefore, an array of 0s and 1s whose values are of statistics kind, say unit8, and isn't always viewed as a binary image in MATLAB. A numeric array is transformed to binary using characteristic logical. Hence, if A is a numeric array including 0s and 1s, we createan array B utilizing the declaration.

B=logical (A)

If A contains elements other than 0s and 1s. Use of the logical function converts all nonzero quantities to logical 1s and all entries with value 0 to logical 0s.

Using relational and logical operators also creates logical arrays.

To test if an array is logical, we use the I logical function: illogical(c).

If c is a logical array, this function returns a 1. Otherwise returns a 0. Logical array can be converted to numeric arrays using the data class conversion functions.

6.7.3 Indexed Images

An indexed image has two components.A data matrix integer, x A color map matrix, map Matrix map is an m*3 arrays of sophistication double containing floating purpose values within the selection [0, 1]. The size m of the map is up to the quantity of colors it defines. Every row of map specifies the red, inexperienced and blue parts of one color. An indexed image uses direct mapping of component intensity values color map values. The color of every component relies upon creating use of the corresponding price the number matrix x as a pointer in to map. If x is of sophistication double, then all of its accessories with values not up to or up to one purpose to the primary row in map, all parts with price two purpose to the second row and lots of others. If x isof sophistication models or unit sixteen, then all components price zero point to the first row in map, all parts with price one point to the secondand plenty of others.

6.7.4 RGB Image

An RGB coloration picture is an M*N*three array of shade pixels the place every shade pixel is triplet just like the red, green and blue components of an RGB image, at a certain spatial scenario. An RGB image may also be considered as stack of 3 grey scale images that once fed in to the red, green and blue inputs of a color display.

Produce a color photo at the reveal. Convention the three images shots forming an RGB colour image are known as the red, green and blue element images. The records category of the components images determines their variety of values. If an RGB image is of class double the range of values is [0, 1].

In a similar fashion the kind of values is [0,255] or [0, 65535]. For RGB images of class units or unit 16 respectively. The quantity of bits uses to represents the pixel values of the factor images determines the bit depth of an RGB image. As an instance, if every aspect image is an 8bit image, the corresponding RGB image is stated to be 24 bits deep.

Commonly, the range of bits in all issue photographs is the identical. In this situation the number of feasible colors in an RGB image is (2^b) ^3, wherein b is some of bits in each aspect image. For the 8bit case the wide variety is 16,777,216 colors.

CHAPTER 7
RESULTS

7.1 Input Images

Figure 7.1 is the original image without any preprocessing techniques applied. Image enhancement is the process of adjusting original image so that the resultant image is more suitable to display CT Images.

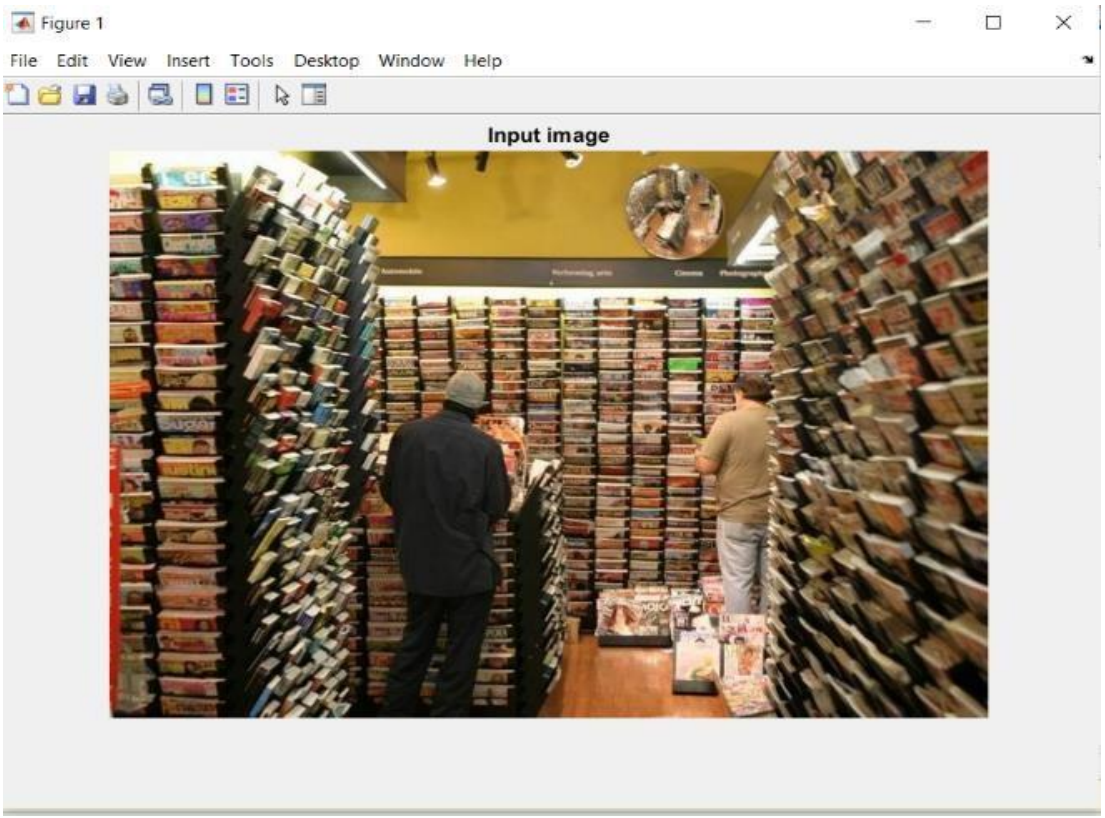


Figure 7.1 Input Image 1

A Grayscale (Or Gray level) image is simply One (1) in which the only colours are only shades of gray. The reason for differentiating such images from any other sort of colour

image is that less information needs to be provided for each pixel. It only contains the brightness information but not colour. A Grayscale image is one with all information removed. Here we areusing grayscale image as an input image because the colour increases complexity of the model. So, the inherent complexity of gray level images is lower than that of colour images.

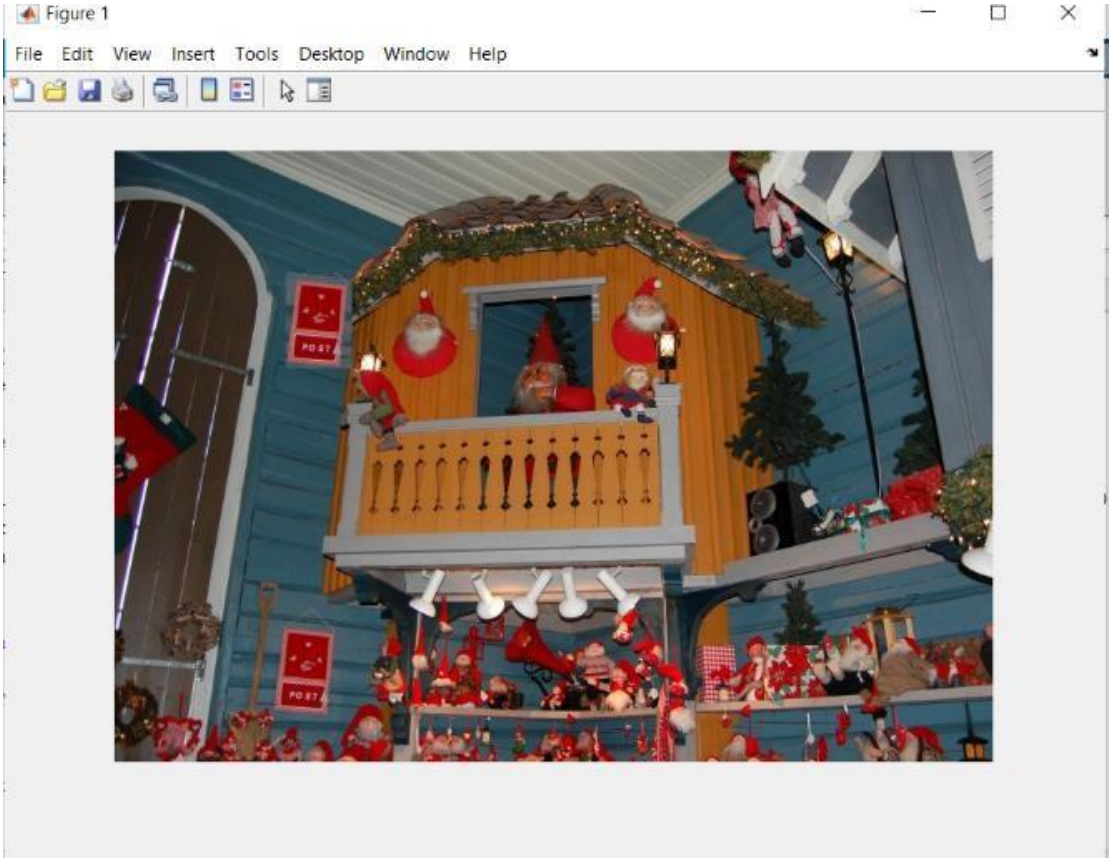


Figure 7.2 Input Image 2

7.2 Iterations

The Figure 7.3 shows the Iterations that are used to train the Input Images in CNN (Convolution neural networks). In this process 90 iterations are used to train the image. The above training progress the first plot shows the accuracy of the trained input image which is including in the both training and validation of that particular image.

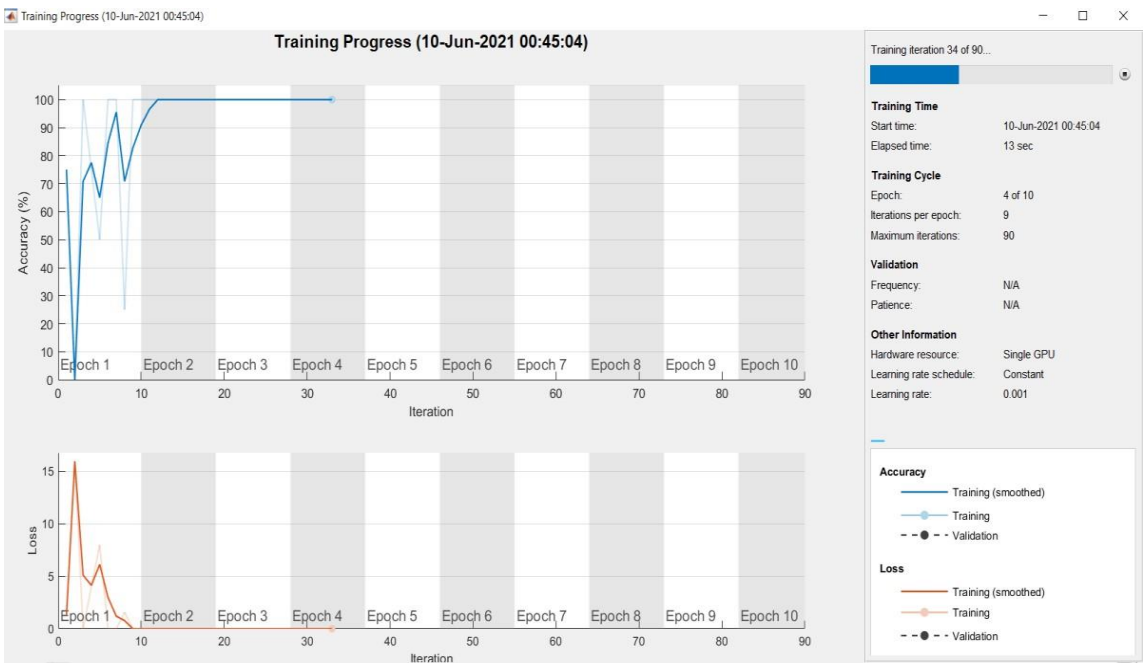


Figure 7.3 Iteration Screen in MATLAB

In the beginning of the plot the accuracy is less (for the first iteration) which can be further improved and enhanced for the last iteration.

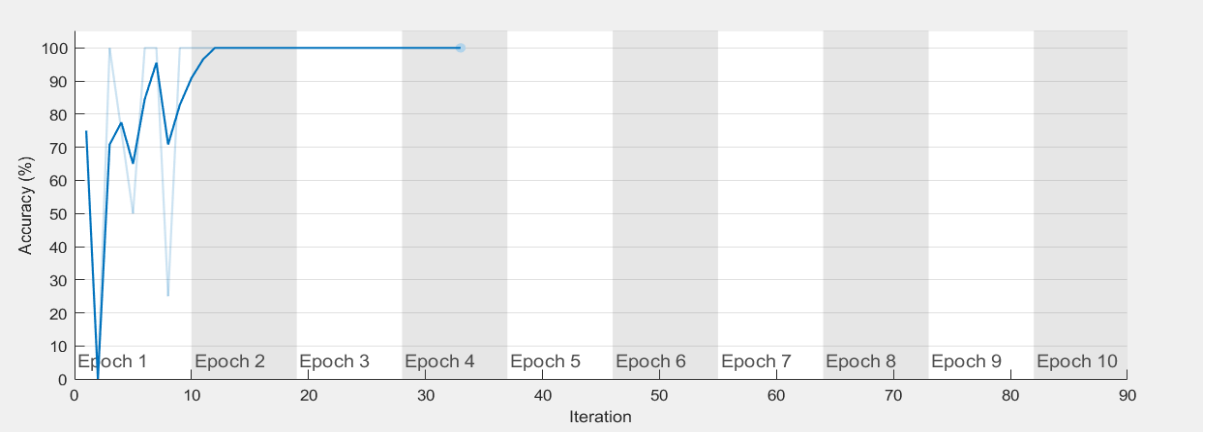


Figure 7.4 Graph of Accuracy of the Image

In the figure 7.5, the loss curve versus iterations. In the first iteration the last tends to be less as the iterations are increased. The loss of the image is Increased at the first iterations and gradually decreased in last iterations the loss tends nearly Zero.

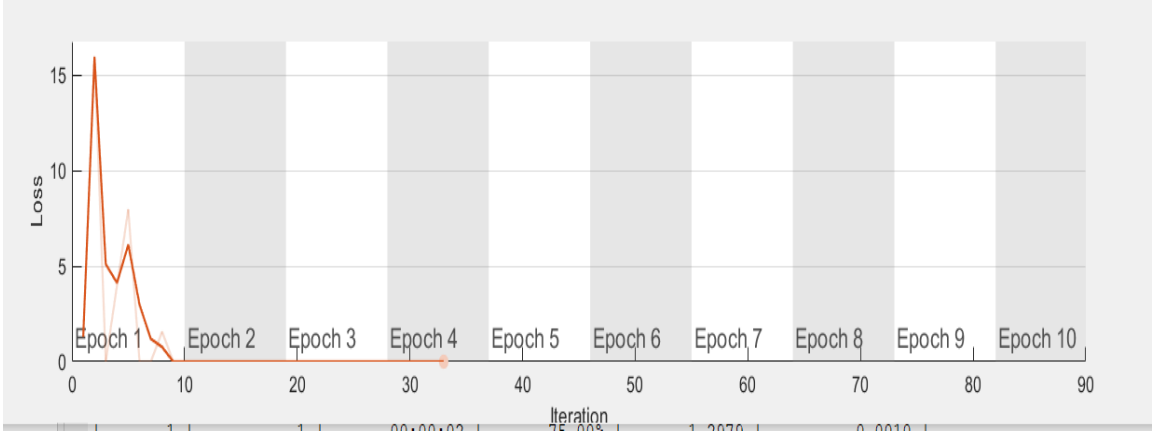


Figure 7.5 Graph of Loss of the Image

7.3 Classification of Tampered Image

The Figure 7.6 popups in the screen when the Input image is training is completed and displays Image is Tampered, if the input image is tampered. If the input image is not tampered, a message box popup and displays Image is not Tampered as shown in the Figure 7.7.

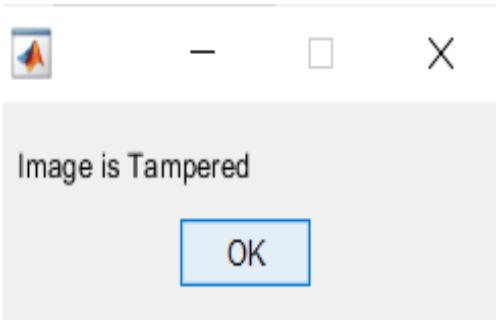


Figure 7.6 Image is Tampered Message Box

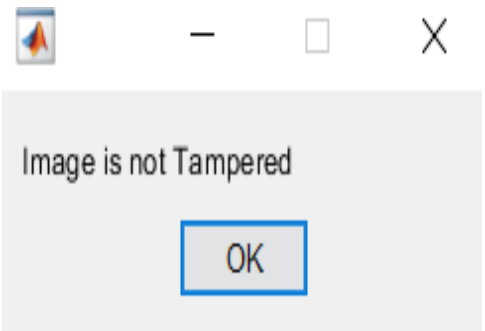


Figure 7.7 Image is not Tampered Message Box

7.4 Segmented Images

The input gray scale image is segmented using the U-Net architecture if the input image is tampered. If the input gray scale image is not tampered, the segmentation process is not required. The Figure 7.8 shows the final segmented image of an input image which is tampered. The final segmented image differentiates the Tampered part of the image with the rest of the image as shown in the Figure 7.9.

The segmentation process is not required for the image not being tampered. The segmentation process is with accuracy of about 84 percent which is very high as compared with the existing MDBD method using SVM.

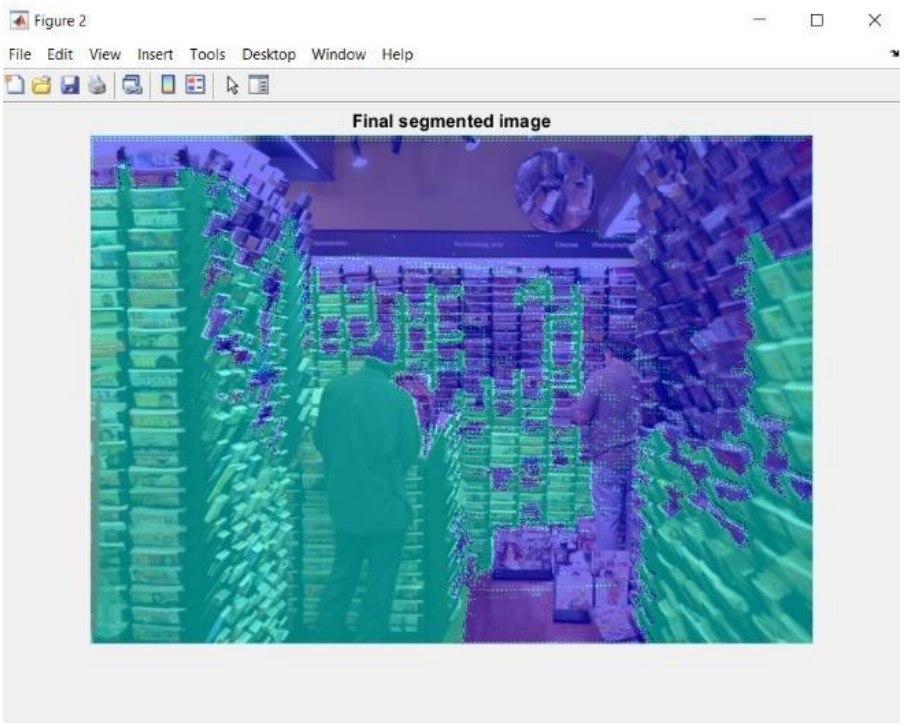


Figure 7.8 Final Segmented Image 1

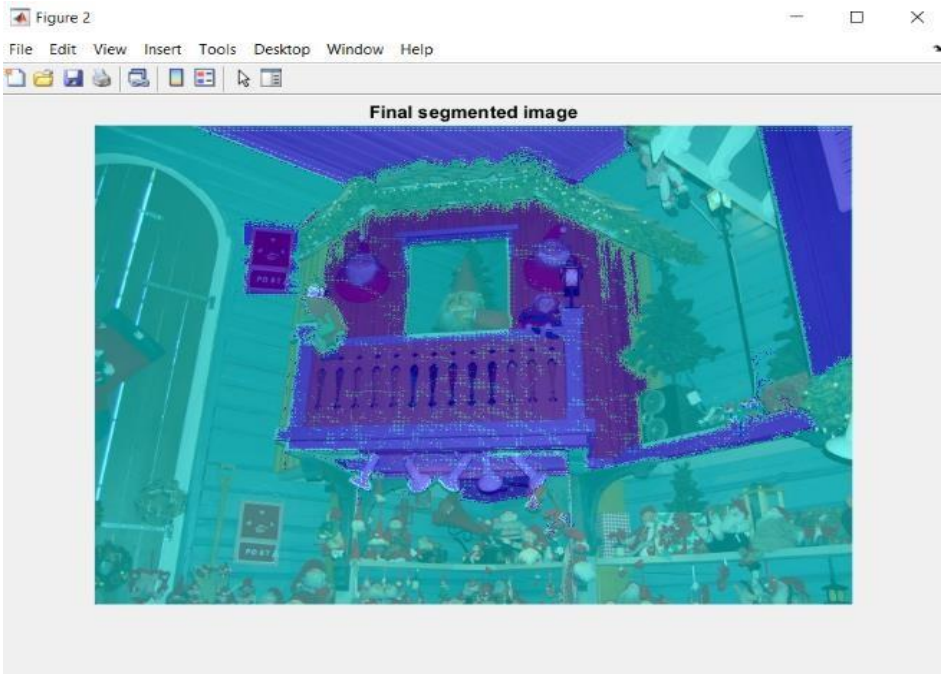


Figure 7.9 Final Segmented Image 2

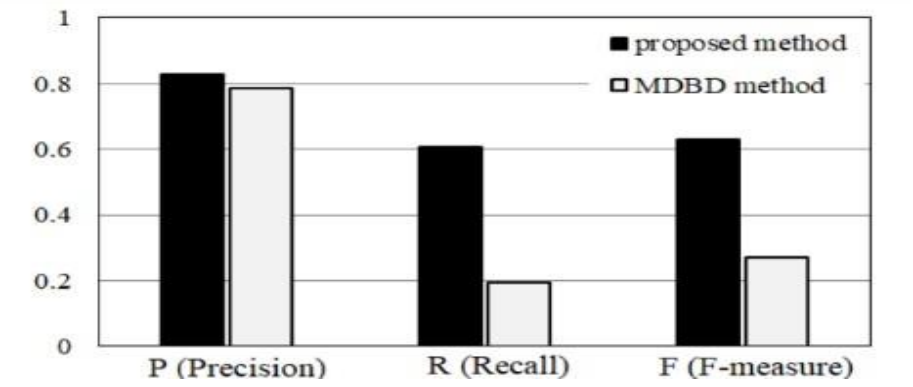


Figure 7.10 Detection accuracy comparison (average)

CHAPTER 8

PROJECT MANAGEMENT AND FINANCE FACTORS

8.1 Hardware

Since it is a simulation-based project a commercial PC with the following configuration is used.

Table 8.1 System Configuration and price

Unit type	Unit model	Price (Rs.)
CPU	Intel Core i5	55,000.00
Clock	3 GHz	
RAM	8GB	
SDD	512 GB	

8.2 Software

The software used for this project is MATLAB R 2019b with additional Toolboxes. The price ofeach of these Toolboxes along with the total cost is shown in Table 8.2.

Table 8.2 Software configuration and price

Module type	Price (Rs.)
MATLAB Student Version	3,283.00
Image Processing Toolbox	1,943.00
Total	5,226.00

8.3 Time Management

The total duration of the project work is 16 weeks with 36 hours spent each week. This entire duration can be divided in to two phases with each phase taking eighteen weeks. In phase 1, presentation 1, complete literature survey is done. The problem in existing method is identified. The software tools that are required to simulate the proposed algorithm i.e., MATLAB R2019b.

Table 8.3 Time Management

Description ↓	Total No. of weeks – 16, 36 hours per week						
weeks →	1-2	3-4	5-8	9-10	11-12	13-14	15-16
Problem Identification							
Literature Survey							
Identification of problem in existing techniques							
Proposed Solution and coding							
Documentation							

In phase 2, presentation 2, gone through some books and websites to acquaint myself to the MATLAB software and some important tool boxes that are available in the MATLAB. Developed the Mathematical model of proposed method. The code for the proposed algorithm is written and simulated the code using MATLAB R2019 b successfully.

8.4 Societal and Environmental Impact

The proposed method detection of tampered image using U-Net architecture is cost effective. The number of iterations using the U-Net architecture are more; due to this the latency in detecting and segmentation of the tampered region is increased. The detection and segmentation of tampered region is laborious and it finds application in various fields such as Bio-Medical image and Face Recognition applications of RNN, Legal, Banking, Insurance, Document digitization, Optical Character Recognition, Drug Discovery, etc.,

CONCLUSION

In the course of work, we reviewed research works on image tampering and its detection. Adjusting the color contrast in a real-world photo may not be with malicious intention but concealing an object or changing someone's face within the image should arouse suspicion. Image tampering is therefore defined as some content of a real-world photo has been replaced with new content from the same image or other images. Humans have difficulties in identifying if a photo has been altered, and it becomes worse when they try to localize where the tampering happens. For a tampered image, finding the corresponding original photos is the best means to prove the fraudulence. However, attackers would do their best to hide or destroy the original image. Therefore, inspecting the invisible artifacts left by possible image manipulation operations is naturally an important way for image tampering detection.

The proposed a method for detecting a tampered region in a JPEG image by using a CNN. In our proposed method, the detection accuracy of the tampered part was improved significantly while the erroneous detection of the non-tampered part as tampered with, was suppressed. In addition, the Fmeasure of our method was approximately 2.3 times that of the MDBD method. In future, we will further optimize the network structure and aim to improve the tampering-detection accuracy.

REFERENCES

[1] A. Kaur and R. Sharma, ”Copy-move forgery detection using DCT and SIFT”, International Journal of Computer Applications, vol.70-no.7, pp.30-34, May 2013.

[2] M. K. Johnson and H. Farid, ”Exposing digital forgeries by detecting inconsistencies in lighting”, MM&Sec 2005 Proceedings of the 7th workshop on Multimedia and security, pp.1-10, New York, USA, 2005.

[3] A. C. Popescu and H. Farid, ”Exposing digital forgeries in color filter array interpolated images”, IEEE Trans. Signal Processing, vol.53, pp.3948-3959, Oct. 2005, DOI:10.1109/TSP.2005.855406

[4] M. Kobayashi, T. Okabe, and Y. Sato, ”Detecting forgery from staticscene video based on inconsistency in noise level functions”, IEEE Trans. Information Forensics and Security, vol.5, pp.883-892, Sept. 2010, DOI:10.1109/TIFS.2010.2074194

[5] J. He, Z. Lin, L. Wang, and X. Tang, ”Detecting doctored JPEG images via DCT coefficient analysis”, Computer Vision-ECCV 2006, vol.3953 pp. 423-435, 2006

[6] T. Pevny and J. Fridrich, ”Detection of double-compression in JPEG images for application in steganography”, IEEE Trans. Information Forensics and Security, vol.3 pp. 247-258, June, 2008

[7] K. Taya, N. Takeda, T. Kobayashi, Y. Ozaki, and N. Kuroki: ”Detecting doctored JPEG image based on block noise analysis and double JPEG analysis”, IEEJ Trans.EIS, Vol.137 No.5, 2017 (in Japanese)

APPENDIX

Program Code to detect tampering of Image using U-Net architecture

```
clc;
clear;
close all;
close all hidden;

[file1,path1]=uigetfile('*.*');
rgb=imread([path1,file1]);
rgb=imresize(rgb,[512 734]);
figure,imshow(rgb);
title('Input image');



matlabroot = 'C:\Program Files (x86)\MATLAB2019\R2019b\bin';
digitDatasetPath = fullfile(matlabroot,'dataset');



imds =
imageDatastore(digitDatasetPath,'IncludeSubfolders',true,'LabelSource','foldernames');



layers = [
    imageInputLayer([512 734 3])
    convolution2dLayer(3,32,'Stride',1,'Padding','same','Name','conv_1')
    batchNormalizationLayer
    reluLayer
    maxPooling2dLayer(2,'Stride',2,'Name','maxpool_1')
```

```
    convolution2dLayer(3,64,'Stride',1,'Padding','same','Name','conv_2')
    batchNormalizationLayer
    reluLayer
    maxPooling2dLayer(2,'Stride',2,'Name','maxpool_2')

    convolution2dLayer(3,128,'Stride',1,'Padding','same','Name','conv_3')
    batchNormalizationLayer
    reluLayer
    maxPooling2dLayer(2,'Stride',2,'Name','maxpool_3')
    fullyConnectedLayer(3)
    softmaxLayer
    classificationLayer];


options = trainingOptions('sgdm','Plots','training-
progress','MaxEpochs',10,'MiniBatchSize',4,'initialLearnRate',0.001);


convnet = trainNetwork(imds,layers,options);
YPred = classify(convnet,rgb);

output=char(YPred);
if output=='1'
    msgbox('Image is not Tampered')
else
    msgbox('Image is Tampered')

I = rgb2gray(rgb);
gmag = imgradient(I);
title('Gradient Magnitude');


se = strel('disk',25);
Ie = imerode(I,se);
```

```
Iobr = imreconstruct(Ie,I);
Iobrd = imdilate(Iobr,se);
Iobrcbr = imreconstruct(imcomplement(Iobrd),imcomplement(Iobr));
Iobrcbr = imcomplement(Iobrcbr);
bw = imbinarize(Iobrcbr);

img4=zeros(512,734);
for i=1:512
   for j=1:734
      if bw(i,j)==1
         img4(i,j)=255;
      else
         img4(i,j)=0;
      end
   end
end


load('net.mat');
c=semanticseg(img4,net);
B = labeloverlay(rgb, c);
figure,imshow(B);
title('Final segmented image');


[file2,path2]=uigetfile('*.*');
groundtruth=imread([path2,file2]);
a=B;

u_bw_filename = im2bw(a);
b=groundtruth;
u_GT_filename = im2bw(b);
```

```
u_GT = [((u_GT_filename)) > 0];
u_bw = [((u_bw_filename)) > 0];

temp_obj_eval = objective_evaluation_core(u_bw, u_GT);

disp('Accuracy--');
disp(temp_obj_eval.Fmeasure);

disp('Specificity--');
disp(temp_obj_eval.Specificity*100);

disp('Sensitivity--');
disp(temp_obj_eval.Sensitivity*100);


end
```

SREE VIDYANIKETHAN ENGINEERING COLLEGE
(AUTONOMOUS)
SreeSainath Nagar, A.Rangampet - 517 102

Department of Electronics and Communication Engineering

PROJECT TITLE: Detecting tampered regions in JPEG images via CNN

ABSTRACT: In this work, we will detect the tampered regions in JPEG images using deep learning techniques (U Net architecture). Often, digital pictures are used as evidence in criminal investigations. Therefore, it is essential to check whether they have been tampered with or not. DCT coefficients play an important role in the detection of Tampered regions of images and these DCT coefficients are input to the CNN (Convolutional neural network). Convolutional neural network (U Net architecture) has been successfully used to achieve good performance in detecting tampered regions of images. U-net is convolutional network architecture for fast and precise segmentation of images mainly. Experiments will demonstrate that our model will provide best detection performance compared to the state-of-the-art methods.

PROJECT BATCH: 21C05

NAGIREDDY CHANDRA MOULI REDDY-17121A04F0
SHAIK THOUHEED AHAMED -17121A04K4
TALAPALA GANESH -17121A04L3
N V SAI VIVEK -17121A04E9

GUIDE: Dr. N. Padmaja, M.Tech., Ph.D.,
Professor, Department of ECE.

POs Attained:

	Program Outcomes												Program Specific Outcomes			
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3	PSO4
Detecting tampered regions in JPEG images via CNN	√	√	√	√	√			√	√	√	√	√	√	√	√	√

Signature of the Guide

BIODATA

NAME	**:**	NAGIREDDY CHANDRA MOULI REDDY
FATHER NAME	**:**	NAGIREDDY BALAKRISHNA REDDY
DATE OF BIRTH	**:**	05/08/2000
NATIONALITY	**:**	INDIAN

CONTACT DETAILS

Contact No.	**:**	8978178994
Email	**:**	cmouli056@gmail.com
Contact Address	**:**	Kadapa, Badvel, Siddavatam road, 516227

BIODATA

NAME	**:**	SHAIK THOUHEED AHAMED
FATHER NAME	**:**	SHAIK JAVEED
DATE OF BIRTH	**:**	31-10-1999
NATIONALITY	**:**	INDIAN

CONTACT DETAILS

Contact No.	**:**	7993735484
Email	**:**	shaikthouheednlr@gmail.com
Contact Address	**:**	Door No. 25-1-981, 6th street, Nethaji Nagar, Podalakur Road, Nellore, Andhra Pradesh-524005.

BIODATA

NAME	**:**	TALAPALA GANESH
FATHER NAME	**:**	TALAPALA SUDHAKAR
DATE OF BIRTH	**:**	11-4-1999
NATIONALITY	**:**	INDIAN

CONTACT DETAILS

Contact No.	**:**	7729828247
Email	**:**	talapalaganesh1999@gmail.com
Contact Address	**:**	3-36, Post Office street, Penubarthi, Nellore rural-524346

BIODATA

NAME	**:**	N V SAI VIVEK
FATHER NAME	**:**	G VENKATESULU
DATE OF BIRTH	**:**	06-07-2000
NATIONALITY	**:**	INDIAN

CONTACT DETAILS

Contact No.	**:**	**9963658702**
Email	**:**	saivivekroyal@gmail.com
Contact Address	**:**	**D/no-9e/268/12,parvati nagar,kalyandurg,anantapur,515761**

Detecting tampered regions in JPEG images via CNN

Kunihiko Taya
Forensic Science Lab.
Kyoto Prefectural Police H.Q.
Kyoto, Japan
162t802t@stu.kobe-u.ac.jp

Nobutaka Kuroki
Graduate School of Engineering
Kobe University
Kobe, Japan
kuroki@kobe-u.ac.jp

Naoto Takeda
FUSO PRECISION
Kyoto, Japan
takeda@fuso.co.jp

Tetsuya Hirose
Graduate School of Engineering
Osaka University
Osaka, Japan

Masahiro Numa
Graduate School of Engineering
Kobe University
Kobe, Japan

***Abstract*—Often, digital pictures are used as evidence in criminal investigations. Therefore, it is essential to check whether they have been tampered with or not. In this study, we propose a method for detecting the tampered region in a JPEG image by using a convolutional neural network (CNN). In the proposed method, DCT coefficients are input to the CNN. The output is a binary segmented image in which the tampered and non-tampered regions are represented using white and black pixels, respectively. In our experiment, 45 types of CNN models were created and compared with one another. The detection accuracy of the best model was 0.63 in terms of the F-measure, which is approximately 2.3 times that achieved using our preliminary method, which was based on a support vector machine.**

***Index Terms*—Tampered image, forgery, JPEG, convolutional neural networks (CNNs)**

I. Introduction

Because a digital camera records the subject as it is, a digital image acts as valuable evidence in criminal investigations. However, with the widespread use of digital cameras and image-editing tools in daily life, anyone can tamper with an image, even if he/she is not an expert in image processing. To judge whether the image conveys the truth or not, it is very important to check whether the image has been tampered with.

Many studies were performed on the detection of image forgery. For example, Kaur [1] proposed a method based on DCT(Discrete Cosine Transform) and SIFT(Scale-Invariant Feature Transform) to detect the copy–move forgery. Johnson [2] described a technique for detecting image forgery by estimating the direction of a point light source. Popescu [3] revealed the traces of digital tampering in color images interpolated using color-filter-array algorithms. Kobayashi [4] used a method that was based on noise characteristics.

Different approaches were presented by He [5] and Pevny [6]; in these approaches, the target image was in the JPEG format. Notably, JPEG is the most frequently used image format, thereby making the approaches practical. The method proposed by He was a superior algorithm because it can automatically locate the tampered part irrespective of the shape and area of the tampered part. Similarly, we performed studies on the detection of tampered JPEG-format images and, consequently, reported that the overall performance of the MDBD(Multiple Detection using Block noise and Double JPEG) method [7] is better than that of the method proposed by He.

In this study, we propose a novel technique to detect the tampered region in a JPEG image by using CNN. Compared with the MDBD method, in which feature values are selected via know-how, the proposed approach optimizes them using CNN and achieves a higher detection accuracy.

In Section 2, we briefly review the prior art and MDBD method. In Section 3, we describe the proposed method. In Section 4, we show the experimental results and compare them to those of the prior art. We conclude this study in Section 5.

II. Previous work - MDBD method

The MDBD method, as depicted in Fig.1, is based on both the block noise analysis and double JPEG analysis. Upon pasting a region from a JPEG image onto the host image, assuming random placement of the forged region, the probability of the blocking artifacts not being aligned is 63/64. Therefore, high-frequency components derived from unaligned blocking artifacts exist in the tampered parts. The parameters that measure the effects of tampering have been expressed quantitatively by analyzing the extracted 19 high-frequency components from 64 DCT coefficients of each 8×8 pixel block and, subsequently, estimating the position of the blocking artifacts. In addition, the effects of double JPEG have been quantified. By inputting these parameters to a support vector machine, both the tampered image and original image can be identified.

As a result of this experiment, the following two points have been revealed. 1. We must focus on high-frequency components of the DCT coefficients because the tampered part is affected by the blocking artifacts. 2. The spatial continuity of potentially tampered regions must be considered while

978-1-7281-7044-2/20/$31.00 ©2020 IEEE

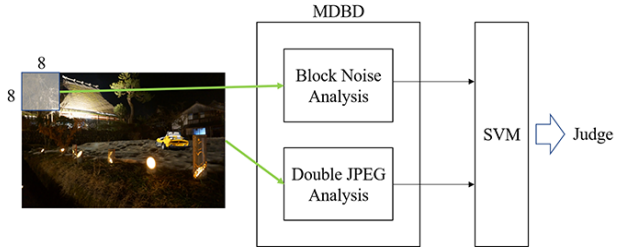


Fig. 1. Overview of MDBD

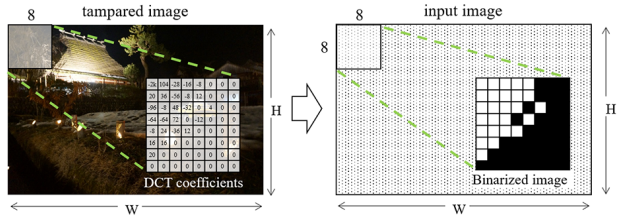


Fig. 2. Input image with binarized DCT coefficients

detecting tampered parts. The MDBD method can discriminate between the tampered image and original image with a higher accuracy than that of the method by He; however, the accuracy of locating the tampered part must be improved.

III. Proposed Method

Our proposed method, as depicted in Fig.2, extracts 64 DCT coefficients for each 8×8 pixel block in the tampered image and, subsequently, creates an 8×8 pixel image that is binarized to 0 or 1 depending on whether each component is 0 or not, respectively. This binarized image corresponds to the original 8×8 pixel block. Processing the entire image in a similar manner, a binary image with the same size as that of the inspection image is created. Fig.3 shows the CNN structure of the proposed method. The input image to the CNN is a two-dimensional representation of the DCT coefficients of the tampered image, and the final output image is a binarized image in which the tampered pixels are shown in white and non-tampered pixels in black.

Each feature map is expressed as follows:

$$F_1(Y) = max(0, Y * W_1 + B_1) \tag{1}$$

$$F_2(Y) = max(0, F_1 * W_2 + B_2) \tag{2}$$

$$F_n(Y) = F_{n-1} * W_n + B_n \tag{3}$$

$$F(Y) = \begin{cases} 1 & (F_n \geqq 0.5) \\ 0 & (F_n < 0.5) \end{cases} \tag{4}$$

where Y denotes the input image, W_n and B_n the filters and biases of the nth layer, respectively, and $F(Y)$ the output image.

The stride of the first layer is set to 8 and that of the other layers to 1. All the layers are set to zero padding, with no dropout layers.

Generally, the greater is the filter size, the wider is the input range reflected in the output of the CNN. In preliminary experiment 1, the reference range of the input is determined using an n-layer CNN (n:3 to 6). We measure the detection accuracy while varying the number of layers, i.e., n.

In preliminary experiment 2, we compare the detection accuracies of 45 types of network structures that satisfy the reference range. Finally, using the network determined in preliminary experiment 2, we compared the detection accuracy of the proposed method with that of the MDBD method.

IV. Experiments and Results

In the experiment, 50 JPEG-format images were taken using five digital cameras each from five different companies, namely, Casio, Sony, Canon, Olympus, and Nikon. The images contained various subjects, such as natural landscapes, car, people, and buildings. We performed the type of image forgery in which a part of the image was copied to another location in an image taken using the same camera, following which the tampered images were saved in the JPEG format. The image-editing tool was Photoshop CS, and the setting when saving the image was the highest image quality, i.e., lowest compression.

A. Input and Ground truth

We now describe a method of creating input images and ground truth. Because the size of the inspection image, which is taken using a digital camera, varies from 4M to 16M pixels, the input image is also significantly large. Owing to the specifications of the experimental apparatus, in preliminary experiment 1, we segmented each image into 256 × 256 pixels to be used as input segmented images. From the 50 tampered images, a total of 6,489 input images were obtained. The ground truth was taken as a binary image, wherein the pixel value of the tampered part was 1 (white) and that of the non-tampered part 0 (black). Among the 50 tampered images(i.e., 6,489 segmented images) created, 40 were training images(i.e., 5,635 segmented images), 5 were verification images(i.e., 427 segmented images), and 5 were test images(i.e., 427 segmented images). The training was performed in 100 epochs, and the testing was performed using the model with the highest validation accuracy.

As mentioned earlier, previous work has shown that the analysis of 8×8 pixel blocks is effective and that the considering the spatial connectivity along the x and y directions of potentially tampered region is important for detecting a tampered JPEG image. Therefore, setting the filter size of the first layer to 24(=8×3blocks) and stride to 8, the analysis is performed in the units of 8×8 pixel blocks while considering connectivity. To set the stride to 8 in the first layer, the ground truth must be of 32×32 pixels, which is 1/8 the size of the input image. The binary ground-truth image is divided into

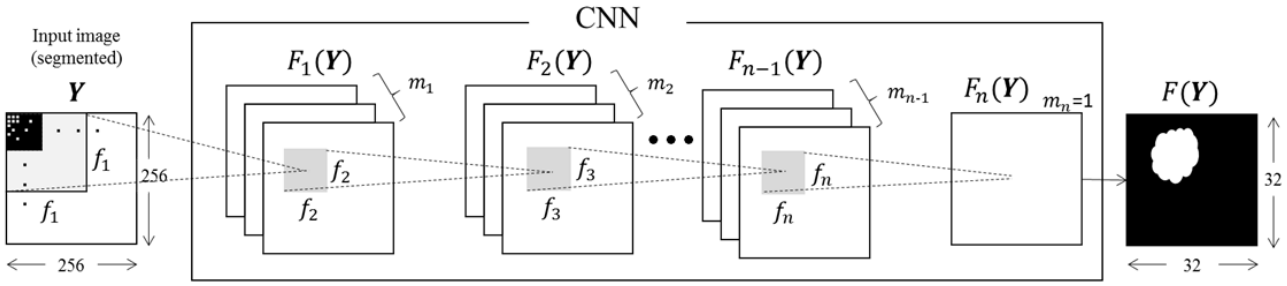


Fig. 3. CNN structure of the proposed method

TABLE I
PERFORMANCES FOR SEVERAL LAYER DEPTHS

n	$\{f_1,\cdots,f_n\}$	$\{m_1,\cdots,m_n\}$	P	R	F
3	{24,3,3}	{32,16,1}	0.77	0.58	0.66
4	{24,3,3,3}	{32,16,16,1}	0.74	0.72	0.73
5	{24,3,3,3,3}	{32,32,16,16,1}	0.75	0.79	0.76
6	{24,3,3,3,3,3}	{32,32,16,16,16,1}	0.76	0.80	0.77
7	{24,3,3,3,3,3,3}	{32,32,16,16,16,16,1}	0.76	0.82	0.79

256×256 pixels according to the input image, and the image reduced to 32×32 pixels via the bicubic method is used as the segmented ground truth. The following computing devices were used in the experiment: a CPU: Intel Xeon W3670, GPU: GeForce GTX1060 6GB, and framework: Theano (Keras).

B. Preliminary experiment 1 - Determining the reference range of input image

The average of the experimental results for 5 test images (i.e., 427 segmented images) with n layers is presented in TableI. In addition, P, i.e., precision, R, i.e., recall, and F, i.e., F-measure are defined as follows:

$$P = \frac{T_P}{T_P + F_P}, \quad R = \frac{T_P}{T_P + F_N}, \quad F = \frac{2 \cdot P \cdot R}{P + R} \tag{5}$$

where T_P denotes the number of detected tampered pixels, F_N the number of undetected tampered pixels, and F_P the number of non-tampered pixels erroneously detected as tampered.

From the results of preliminary experiment 1, it is evident that F reached almost the maximum value in 6 layers. Notably, we must calculate the reference range per output pixel from the network structure with six layers. In the first layer, the reference range is 24 pixels along the x and y directions. In the second and subsequent layers, because the filter size is $f_n = 3$ for an 1/8 size output map, the reference range is expanded by 8 pixels along the x and y directions. Therefore, the reference range throughout the entire network is $24 + 8 \times 5 \times 2 = 104$ pixels.

C. Preliminary experiment 2 - Determining the network structure

Following the condition that the reference range is 104×104 pixels, the number of filters and maps are changed, and the corresponding network structures are compared with one another. In preliminary experiment 1, zero padding was performed on all the layers, following which the boundaries of the output image became discontinuous. Therefore, zero padding is not performed in preliminary experiment 2. In this case, the size of the input segmented image is changed to 352×352 pixels because the reference range extends 48 pixels along both the x and y directions. The segmented ground truth measures 32×32 pixels as in preliminary experiment 1.

A total of 45 network structures, as listed in TableII, were created by combining the network layer n : $3to6$, filter size f_n : $24, 7, 5, 3$, and number of maps m_n : $128, 64, 32, 16, 1$, and, subsequently, the detection accuracy was measured. Because the network structure with the highest F is $(n, \{f_n\}, \{m_n\})$=(5,{24,5,3,3,3},{128,128,64,32,1}) and because P=0.95, R=0.81, and F=0.86, we will use this network structure and compare it to the MDBD method in the next section.

D. Comparison with the previous method

The evaluation has been performed using a 10-fold cross-validation method for the 50 tampered images To stabilize the learning, a dropout (0.5) was inserted between each layer. The model that maximized the validation accuracy was used for each group. The detection results of the proposed method and those of the MDBD method are depicted in Fig.4, and the detection accuracies are depicted in Fig.5. It is evident that the proposed method shows the tampered regions more effectively compared with the MDBD method. In addition, the F-measure of the proposed method achieves 0.63, which is approximately 2.3 times that of the MDBD method.

V. CONCLUSIONS

We proposed a method for detecting a tampered region in a JPEG image by using a CNN. In our proposed method, the detection accuracy of the tampered part was improved significantly while the erroneous detection of the non-tampered part as tampered with, was suppressed. In addition, the F-measure of our method was approximately 2.3 times that of the MDBD method. In future, we will further optimize the network structure and aim to improve the tampering-detection accuracy.

(1) tampared image:ex.1

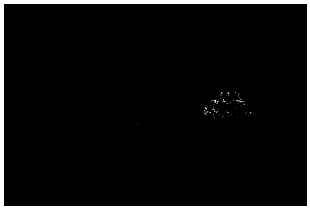
(2) tampared image:ex.2

(3) ground truth:ex.1

(4) ground truth:ex.2

(5) proposed method:ex.1

(6) proposed method:ex.2



(7) MDBD method:ex.1

(8) MDBD method:ex.2

Fig. 4. Detection results

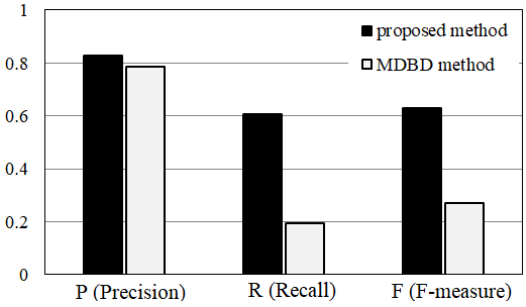


Fig. 5. Detection accuracy (average)

TABLE II
45 TYPES OF CNN MODELS AND THEIR PERFORMANCES

n	filter size $\{f_1,\cdots,f_n\}$	map $\{m_1,\cdots,m_n\}$	P	R	F
6	{24,3,3,3,3,3}	{16,16,16,16,16,1}	0.80	0.84	0.82
		{32,32,32,32,32,1}	0.76	0.85	0.80
		{64,64,64,64,64,1}	0.73	0.90	0.79
		{128,128,128,128,128,1}	-	0	-
		{32,32,32,16,16,1}	0.71	0.89	0.77
		{64,64,32,32,32,1}	0.59	0.91	0.70
		{128,128,128,64,64,1}	0.77	0.84	0.80
		{64,64,32,32,16,1}	0.83	0.81	0.82
		{128,128,64,64,32,1}	0.93	0.80	0.84
		{128,128,64,32,16,1}	0.93	0.80	0.85
5	{24,5,3,3,3}	{16,16,16,16,1}	0.79	0.82	0.80
		{32,32,32,32,1}	0.85	0.81	0.82
		{64,64,64,64,1}	0.91	0.73	0.81
		{128,128,128,128,1}	0.82	0.81	0.81
		{32,32,16,16,1}	0.82	0.78	0.80
		{64,64,32,32,1}	0.72	0.89	0.79
		{128,128,64,64,1}	0.91	0.34	0.43
		{64,64,32,16,1}	0.70	0.89	0.77
		{128,128,64,32,1}	**0.95**	**0.81**	**0.86**
		{128,64,32,16,1}	0.87	0.72	0.79
4	{24,5,5,3}	{16,16,16,1}	-	0	-
		{32,32,32,1}	0.87	0.81	0.83
		{64,64,64,1}	0.84	0.81	0.82
		{128,128,128,1}	-	0	-
		{32,32,16,1}	0.89	0.79	0.83
		{64,64,32,1}	0.84	0.79	0.80
		{128,128,64,1}	0.87	0.73	0.79
		{64,32,16,1}	0.83	0.84	0.83
		{128,64,32,1}	0.89	0.79	0.82
4	{24,7,3,3}	{16,16,16,1}	0.85	0.74	0.79
		{32,32,32,1}	0.80	0.84	0.81
		{64,64,64,1}	-	0	-
		{128,128,128,1}	-	0.31	-
		{32,32,16,1}	0.83	0.77	0.80
		{64,64,32,1}	0.82	0.85	0.83
		{128,128,64,1}	0.84	0.71	0.77
		{64,32,16,1}	0.85	0.85	0.84
		{128,64,32,1}	0.86	0.81	0.83
3	{24,7,5}	{16,16,1}	0.81	0.68	0.72
		{32,32,1}	0.84	0.75	0.79
		{64,64,1}	0.89	0.67	0.76
		{128,128,1}	0.66	0.78	0.70
		{32,16,1}	0.77	0.85	0.80
		{64,32,1}	0.59	0.71	0.63
		{128,64,1}	0.87	0.72	0.78

REFERENCES

[1] A. Kaur and R. Sharma, "Copy-move forgery detection using DCT and SIFT", International Journal of Computer Applications, vol.70-no.7, pp.30-34, May 2013.

[2] M. K. Johnson and H. Farid, "Exposing digital forgeries by detecting inconsistencies in lighting", MM&Sec 2005 Proceedings of the 7th workshop on Multimedia and security, pp.1-10, New York, USA, 2005.

[3] A. C. Popescu and H. Farid, "Exposing digital forgeries in color filter array interpolated images", IEEE Trans. Signal Processing, vol.53, pp.3948-3959, Oct. 2005, DOI:10.1109/TSP.2005.855406

[4] M. Kobayashi, T. Okabe, and Y. Sato, "Detecting forgery from static-scene video based on inconsistency in noise level functions", IEEE Trans. Information Forensics and Security, vol.5, pp.883-892, Sept. 2010, DOI:10.1109/TIFS.2010.2074194

[5] J. He, Z. Lin, L. Wang, and X. Tang, "Detecting doctored JPEG images via DCT coefficient analysis", Computer Vision-ECCV 2006, vol.3953 pp. 423-435, 2006

[6] T. Pevny and J. Fridrich, "Detection of double-compression in JPEG images for application in steganography", IEEE Trans. Information Forensics and Security, vol.3 pp. 247-258, June, 2008

[7] K. Taya, N. Takeda, T. Kobayashi, Y. Ozaki, and N. Kuroki: "Detecting doctored JPEG image based on block noise analysis and double JPEG analysis", IEEJ Trans.EIS, Vol.137 No.5, 2017 (in Japanese)