



AIR UNIVERSITY

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

BACHELOR OF COMPUTER ENGINEERING

IDENTITY FRAUD DETECTION USING FACIAL ANTISPOOFING

Group Members:

Amna Yaseen 180384
Raza Ullah 180385
Rabia Khan 180360
Mahnoor Fatima 180395

Supervisor:

DR. M. HABIB MAHMOOD

Fall 2018

Approval of submission

It is to certify that the project titled
"Identity Fraud Detection Using Facial Anti-Spoofing"
has met the required standard of submission
in partial fulfillment of the requirement
for the award of degree of
Bachelor of Electrical Engineering
at Air University ,Islamabad

Project Supervisor
DR. M. HABIB MAHMOOD

Head Of Department
PROF. HAFIZ AASHIQ HUSSIAN
Chair Department

DEPARTMENT OF ELECTRICAL ENGINEERING

Submitted by:

AMNA YASEEN (180384)

RAZA ULLAH (180385)

RABIA KHAN (180360)

MAHNOOR FATIMA(180395)

Project Supervisor

Dr. M. Habib Mahmood

Assistant Professor

Head of Department

Prof. HAFIZ ASHIQ HUSSAIN

Acknowledgement

IN the name of Allah, Most Merciful, Most Gracious, Who has bestowed upon us abilities with persistence to achieve our academic goals. We are pleased and thankful to our families who believed in us and supported our journey. It is the result of their prayers that we are at the verge of completing our degrees. We are thankful to our respected supervisor, who assisted the project throughout. It is the result of his efforts that we are able to complete this project on time.

Abstract

IT has become easy for cyber criminals to impersonate an entity or a person to gain their trust. The cyberattack of spoofing any facial recognition systems is a big concern in the age where such systems are widely used. There is a need to develop such technology which can help to identify this kind of cyberattacks. Facial recognition biometric system is used in airport security systems, shopping malls, government buildings, banks, online-exams and in context of criminal investigation. Various biometric technologies have been widely used in real-world applications, including online pricing and e-commerce security, smartphone-based authentication, biometric passports, and border control assessments. Spoofing these facial recognition biometric systems can be done with false facial verification. This paper offers a short overview of face presentation attacks and anti-spoofing strategies. This paper additionally covers special methodologies for face spoofing detection, description of the experimentation databases available for face anti-spoofing detection and goals to offer new research route in this discipline.

Contents

1	Introduction	1
1.1	History	1
1.2	Introduction	1
1.3	Spoofing and Anti Spoofing	2
1.3.1	What is Spoofing?	2
1.3.2	What is anti spoofing?	2
1.3.3	Limitations in face anti spoofing?	2
1.3.4	Problem Statement	2
1.4	Objectives	3
1.5	Summary	3
1.6	Organization of Thesis	3
2	Literature Review	4
2.1	Introduction	4
2.2	Instruments of Presentation Attacks	4
2.2.1	Spoofing	5
2.3	Classes	5
2.4	Facial Spoofing	5
2.5	Presentation Attacks	5
2.6	Common Presentation Attacks	6
2.6.1	Photo attacks	6
2.6.2	Three-dimensional masking assaults	6
2.6.3	Video replay attack	6
2.7	Anti-Spoofing	6
2.7.1	Anti-Spoofing Solutions	6
2.7.2	Anti-Spoofing Techniques	7
2.8	Common Techniques	7
2.9	Eye Blink Detection	8
2.10	CNN in Deep Learning Feature	8
2.11	Challenge Response	9
2.12	3D Camera	9

Contents

2.13 Flash Active	9
2.14 HOW DOES IT FUNCTION?	10
2.15 Facial Anti Spoof Technique	10
2.16 Literature Review	10
2.17 Summary	17
3 Methodology	18
3.1 SIW	18
3.2 Algorithm	18
3.3 Image Classification	18
3.4 CNN	19
3.5 Pooling Layer	21
3.6 Fully connected layers	21
3.7 Models VGG16	22
3.8 ResNet	23
3.8.1 Designing of Network	24
3.9 Residual blocks and intuition	24
3.10 Viola Jones	24
3.11 Features of HAR	25
3.11.1 Software Aspect	26
3.11.2 Hardware Aspect	26
3.11.3 Code Description	27
3.12 Conclusion	28
3.13 Summary	29
4 Results and Analysis	30
4.1 Datasets	30
4.1.1 Labeled dataset	30
4.1.2 Unlabeled dataset	31
4.2 Output Analysis	31
4.3 Results	33
5 Environment and Sustainability	36
5.1 Background	36
5.2 Machine in Sustaining Environment	37
5.3 How our project help in Sustaining Environment	38
5.4 Energy use and carbon emission	39
5.5 Cost of deep learning	40
6 Conclusion and Future Work	41
6.1 ACHEIEMENTS	41
6.2 Deep Learning Classification	41
6.3 CONCLUSION	42
6.4 FUTURE WORK	42

List of Figures

2.1 Common presentation attacks: Photo attack, three dimensional masking assaults and video replay attack comparison using their demonstration .	6
2.2 Showing methoda of detection: eye blink detection, CNN, Active flash, Challenge response, 3d camera amd their specifications	7
2.3 CNN detection to distinguish which photograph are genuine and which are fake. CNN would see and understand collection of frames	8
2.4 Common presentation attacks: Photo attack, three dimensional masking assaults and video replay attack comparison using their function examples	10
2.5 Detection of real and fake face using the first publically available dataset i.e NUAA	11
2.6 Detection of real and fake face using the dataset produced by the user .	11
2.7 Block Schematic of the proposed system that converts the image to YCbCr CIELUV color space, pass it through CNN and obtain feature vector from it.	12
2.8 VGG Face Architecture that show its 16 layers consisting of 2D convolution, max pooling and average pooling	12
2.9 Receiver Operating Characteristics Curve that tells the likelihood weather the images were correctly classified or not	13
2.10 Spoof detection of a single face using VGG16 architecture which consists of 2D convolution, max pooling and average pooling	13
2.11 Spoof detection of a multiple face using VGG16 architecture which consists of 2D convolution, max pooling and average pooling	13
2.12 Tge flowchart for the proposed MC FBC system for 3 dimensional face spoofing is as follows	14
2.13 Face spoofing database for RGB and YCbCr color space for real faces and faces made from wax	14
2.14 The database used for MC FBC approach comprises of these fake images	15
2.15 The database used for MC FBC approach comprises of these real images	15
2.16 CNN architecture that consists of con2D convolution, max pooling, Relu and fully connected layers	16

List of Figures

2.17 Epoch Loss	17
2.18 Epoch Accuracy	17
3.1 CNN.	19
3.2 CNN architecture that consists of con2D convolution, max pooling, Relu and fully connected layers	20
3.3 VGG Face Architecture that show its 16 layers consisting of 2D convolution, max pooling and average pooling	22
3.4 CNN architecture that consists of con2D convolution, max pooling, Relu and fully connected layers	22
3.5 Features if Haar that includes edge features, line features, four rectangle features and some other important features	25
4.1 Real image.	32
4.2 Fake image.	32
4.3 Fake image.	32
4.4 No face image.	33
4.5 Custom Vgg implemented on celeb A Spoof dataset.	33
4.6 Vgg 16 implemented on celeb A Spoof dataset.	34
4.7 Confusion matrix obtained from dataset real and fake images.	34
4.8 Vgg 16 on dataset real and fake images.	35

List of Tables

2.1	NUAA total photos and custom dataset	11
2.2	Total images used	12

CHAPTER 1

Introduction

1.1 History

In the previous twenty years, the progression of innovation in gadgets and software engineering has given admittance to high even out innovation gadgets at moderate costs to a significant extent of the world populace. Different biometric frameworks have been generally conveyed, in actuality, applications, for example, on-line installment and web based business security, cell phone based validation, gotten admittance control, biometric visas, and line checks are all included. Face recognition has been one of the most studied biometric advances since the 1990s [1], owing to its several advantages over other biometrics. For sure, faces are profoundly unmistakable among people and face acknowledgment can be carried out even in non-meddling obtaining situations, or from a good ways. As of late, profound learning has significantly improved the cutting edge execution of numerous PC vision errands, for example, picture grouping and item acknowledgment [2, 3, 4]. With these huge advances, face acknowledgment has additionally made extraordinary forward leaps, for example, the accomplishment of DeepFace [5], DeepIDs [6], VGG Face [7], FaceNet [8], SphereFace [9], and ArcFace [10] are some of the most popular facial recognition systems.

1.2 Introduction

Over the last two decades, advances in electronics and computer science have enabled a large section of the world's population to obtain top-of-the-line technology at cheap prices. Biometric systems have been widely used in real-world applications such online payment and e-commerce security, smartphone-based authentication, protected access control, and biometric passport and border inspections. Since the 1990s, facial recognition has been one of the most researched biometric technologies, owing to its various

advantages over other biometrics. Face recognition can be used even in nonintrusive acquisition circumstances or from a distance because faces are extremely different across persons. Due to the ever-increasing flow of face photos available on the internet and the availability of low-cost but high-resolution digital equipment, photo and video replay attacks are the most common attacks. Imposters can simply collect and reuse authentic user face samples. Photo assaults are carried out by submitting a photo of a legitimate person to the face authentication system. Imposters typically employ a variety of tactics, including printed photo attacks, which involve displaying a photograph printed on paper. Photo display attacks, on the other hand, display the image on the screen of a digital device such as a smartphone, tablet, or laptop before presenting it to the system.

1.3 Spoofing and Anti Spoofing

1.3.1 What is Spoofing?

Spoofing is the act of disguising a communication or identity so that it appears to be associated with a trusted, authorized source. Spoofing attacks can take many forms, from the common email spoofing attacks that are deployed in phishing campaigns to caller ID spoofing attacks that are often used to commit fraud.

1.3.2 What is anti spoofing?

Antispoofing is a technique for identifying and dropping packets that have a false source address. In a spoofing attack, the source address of an incoming packet is changed to make it appear as if it is coming from a known, trusted source.

1.3.3 Limitations in face anti spoofing?

The human face is the most reliable biometric modality owing to its unique characteristic. It has been used in large-scale applications for identifying millions of authentic users. Recently, the increasing demand for biometric technology based on the human face has led to research work in the field of face spoofing which poses a challenge in face recognition. In the last few years, the research community gives more attention to detecting fake and genuine face samples. However, the proposed liveness and replay attack detection algorithms work on familiar spoof materials. Therefore, generalized liveness algorithms need to be implemented for unseen and unpredictable spoofing attacks. The vulnerability of features against spoofing attacks must be taken into account so that additional features can be added to make the system more secure and computationally efficient for unseen and unpredictable spoof attacks.

1.3.4 Problem Statement

The implementation of facial anti spoofing in order to detect fraud of identity caused by imposters.

1.4 Objectives

Following are the objectives of implementation of this project:

- To deploy the fraud detection system where identity of employers need to be secured.
- To train the datasets on different neural networks and observe their working.
- To choose such a neural network whose accuracy is maximum evaluated by evaluation metric.
- To learn how images are manipulated and can serve as security threat.

1.5 Summary

Face recognition systems have been extensively used in government as well as commercial applications such as mobile, banking and surveillance systems etc. In the last 10 years, the whole biometric community such as researchers, developers, and retailers work on this challenging task and develop a more accurate protection methods against spoofing threat. This challenge affects high-security field in the companies, government sectors, rising small and medium sized endeavors. Although several face antispoofing or liveness detection methods have been proposed, the issue is still unsolved due to difficulty in finding the features and methods for spoof attacks. Recently it has been shown that the traditional face biometric techniques are more vulnerable to spoofing attacks, so entire research community required more concentrate to solve this challenging task. The goal of this paper is to provide a detailed study of antispoofing methodologies and evaluation databases. However, there is a need to provide more generalized algorithms for detection of unpredictable spoofing attacks in order to make the system more secure, computationally efficient and reliable.

1.6 Organization of Thesis

In chapter 2 basically we described about the Literature review of Spoofing and Anti Spoofing. In which we specified the different types of attacks and Deep learning techniques as well. Then moving forward we have the other chapter the third one Methodology. In methodology we briefly explained the models and its architecture, all the algorithms as well the hardware and software aspects. After this we explained our results and analysis in which the datasets are described, moreover the graphs and matrices are used to show the obtained results with respect to the accuracy and loss our models obtained. Chapter 5 is how our project has the impact on environment and its sustainability. How it helps in sustaining the environment. By explaining all these aspects then we conclude our project by summarizing it and explained the achievements we achieved till now and the future work we expect to perform with the time being.

CHAPTER 2

Literature Review

2.1 Introduction

Now we'll look into facial recognition. We'll concentrate on facial liveness detection in this article. There are numerous parallels between the two modes. Face liveness



detection systems use texture and motion analysis, as well as artificial intelligence. These elements are combined in the most advanced approaches. They also rely largely on the precision of face detection algorithms to detect changes in stance and emotion, as well as reduce noise (brightness, background).

2.2 Instruments of Presentation Attacks

Again, we provide a (non-exhaustive) list of artefacts (PAIs) that represent human traits and may be used to perpetrate face spoofing attacks utilising 2D, 3D models on a screen (2D), or 3D objects.

- In 2D static assaults, high-definition facial pictures on flat paper are employed, as well as simple flat paper masks with holes. The high-definition display is used to mislead low-resolution cameras. A video sequence comprising visuals can be used to reply to simple challenge and answer procedures. The perforations, in particular, facilitate blinking.

These two-dimensional attacks have been thoroughly reported.

- In 3D static assaults, impersonators use 3D printouts, wax heads, or sculptures.
- In 3D dynamic assaults, fraudsters can use resin, latex, or silicone masks with holes for the eyes and other specialised regions like the mouth, lips, and eye brows.

2.2.1 Spoofing

A cyberattack that happens when a con artist is veiled as a confided in source to gain access to important data or information. Sites, texts, calls, messages, IP addresses, and workers may all be used to parody. In this study, we focus on impersonation (spoofing) attacks, in which the imposter may either take biometric data from a legal user or develop a Presentation Attack, the fundamental objective of caricaturing is to get to individual data, take cash, sidestep network access controls or spread malware through contaminated connections or connections. For it to be effective, the ridiculing assault needs to consolidate a specific degree of social designing. This implies that the strategies that tricksters use can viably fool their casualties into giving out their own data. Tricksters utilize social designing to play on weak human attributes, like ravenousness, dread, and gullibility. An illustration of this kind of friendly designing is the place where the trickster depends on the casualty's sensations of dread trying to acquire data or cash.

2.3 Classes

Email spoofing, caller ID spoofing, website spoofing, text messaging spoofing, gps spoofing, ip address spoofing, DNS spoofing, and ARP spoofing are all examples of spoofing.

2.4 Facial Spoofing

Face recognition has been one of the most discussed biometric developments since the 1990s, owing to its several advantages over other biometrics. Face authentication technologies, on the other hand, became key targets of Presentation Attacks due to their prevalence (PAs). Facial spoofing is another sort of caricaturing that depends on facial recognition programming to open gadgets or access a protected structure. This sort of ridiculing is moderately uncommon yet with progresses in facial acknowledgment innovation and more organizations utilizing facial acknowledgment as a feature of their security framework, the dangers with facial parodying will develop. A cybercriminal can utilize pictures found via online media to construct a similarity of an individual and afterward utilize this to open any security framework that utilizes facial acknowledgment.

2.5 Presentation Attacks

PAs are carried out at the sensor level, avoiding the need to enter the framework. PAs are recognised based on biometric flaws alone. Gatecrashers seek to get access to the

biometric framework by using a relic, which is frequently phoney (e.g., a face photograph, a veil, an engineering finger impression, or a printed iris image), or by imitating the role of certified customers (e.g., walk, signature).

2.6 Common Presentation Attacks

2.6.1 Photo attacks

Photo assaults are carried out by submitting a photo of a legitimate user to the face authentication system.

2.6.2 Three-dimensional masking assaults

Attacks using 3D masks rebuild 3D facial artefacts.

2.6.3 Video replay attack

The image is first shown on a digital device's screen, such as a smartphone, tablet, or laptop, and then provided to the system.

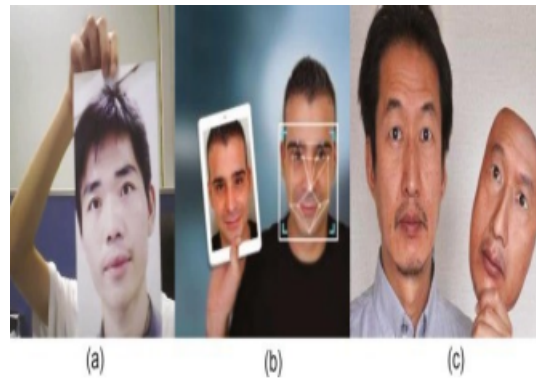


Figure 2.1: *Common presentation attacks: Photo attack, three dimensional masking assaults and video replay attack comparison using their demonstration*

2.7 Anti-Spoofing

The task of avoiding fraudulent facial verification by utilising a photo, video, mask, or other substitute for an authorised person's face is known as facial anti-spoofing. Anti-spoofing strategies should be a major focus from the start when designing a secure face recognition system in a real-world setting.

2.7.1 Anti-Spoofing Solutions

Methods for detecting liveness can be combined. One example is challenge-response and LB. The first method uses movement to detect whether a face is alive. The second prevents these movements from being displayed on a flat screen.



	Static PAI	Dynamic PAI	Image input	User involvement	Generalized	Environment- independent	Cost
LBP							
Eye blink detection							
CNN							
Active flash							
Challenge- response							
3D camera							

Figure 2.2: Showing methods of detection: eye blink detection, CNN, Active flash, Challenge response, 3d camera and their specifications

2.7.2 Anti-Spoofing Techniques

In our increasingly digital society, it's no wonder that cybercrime is on the rise. Many businesses are now looking into machine learning engineers' biometric face recognition as a possible security solution. This cutting-edge technology has a lot of promise and has the potential to transform how we access sensitive data. Face recognition technology will be one of the aspects that will shape the AI environment in the future. Face recognition, as exciting as it seems, is not without problems. User photographs are easily accessed on social media and can be used to deceive facial recognition software. Consider the use of paper photos, screenshots, or 3D facial reconstruction. That is why, in order to protect sensitive data, decrease theft, and mitigate fraud, businesses must implement face anti-spoofing systems. What's to stop someone from impersonating someone else to obtain access to sensitive information? This is where anti-spoofing solutions are required. To verify an individual's identification, we use liveness detection. These checks can determine whether a person is genuinely present or is attempting to fool the system with a photo.

2.8 Common Techniques

Presentation assaults are the most common type of face spoofing attack. To trick facial recognition software, these assaults use 2D and 3D (static or dynamic) techniques. Static 2D presentation assaults utilise images, flat paper, or masks, and dynamic variations use screen video replays or a series of pictures. 3D printing, sculptures, and masks are examples of static 3D presentation attacks, whereas dynamic attacks use highly complicated robots to mimic face emotions, replete with makeup. Of course, these examples do not represent the absolute truth. Presentation attacks emerge in tandem with technological advancements. We believe it's critical to focus on strategies that:

- Prevent static and dynamic 2D spoofs when designing solutions for this challenge.

- No user participation was required
- Images were used instead of videos a trustworthy solution must accomplish maximum accuracy, take minimal time, and put the user's experience first. Most significantly, it had to work with current facial recognition software.

2.9 Eye Blink Detection

One of the most accurate liveness detection tests is eye blink detection. Natural blinking is a simple technique to tell if a face is alive. Humans blink 15â30 times each minute on average. During a blink, the eyes are closed for roughly 250 milliseconds. Modern cameras record videos with much shorter frame intervals (50 milliseconds at 30 frames per second). We can detect frames with closed eyelids in videos and count them to acquire the anticipated numbers. Eye blink detection can be implemented by analysing face landmarks and calculating the surface area of the eyes.

For this objective, we can also use deep learning.

2.10 CNN in Deep Learning Feature

Additional anti-spoofing technologies include deep learning and convolutional neural networks (CNN). When we were researching technologies, we began to think of anti-spoofing as a binary classification problem. We could teach CNN to distinguish which photographs are genuine and which are fake. It will also work. However, there is one issue. The convolutional network would "see" and "understand" no consistent (stable) collection of features.



Figure 2.3: CNN detection to distinguish which photograph are genuine and which are fake. CNN would see and understand collection of frames

Both of the images above are parodies. Yes, our CNN is capable of detecting both of them as spoofing. Even in the left one, there are several non-eye evident aberrations. However, it only works with specific datasets under specific parameters, such as camera quality, surroundings, and light. The neural network will not produce correct results if any of them are modified. As a result, this strategy is only useful in specific situations.

2.11 Challenge Response

Another possible anti-spoofing technology is challenges and responses. This method employs a unique action known as a challenge. The algorithm checks to see if the challenge took place during a video sequence. A challenge-response system verifies an individual's identity through a series of challenges.

- Smiles are one of these problems.
- Sad or happy expressions on the face
- Head movements

While successful, this strategy necessitates additional input and has the potential to negatively damage the user experience.

2.12 3D Camera

The most reliable anti-spoofing cameras are those that employ 3D technology. Because we can tell the difference between a face and a flat object, accurate pixel depth information can protect us from presentation assaults with high precision. Despite the difficulty of 3D assaults, cameras remain one of the most reliable face anti-spoofing techniques. Despite the fact that cameras are widely accessible, not all computer users have them installed. That is why we believe it is critical to work with standard RGB photos.

2.13 Flash Active

Active flash is a fascinating method with a lot of potential, in our opinion. We decided to put it to the test for our particular assignment. It also avoids the "black box dilemma," unlike some of the other alternatives. We were able to detect spoofing utilising light reflections on a face with this technique. The concept is to use the additional light from a device's screen to create a shifting light environment. On the face, the white light casts an adequate reflection.



This technique can be used to distinguish real from artificial faces.

2.14 HOW DOES IT FUNCTION?

We use the data from the frames taken before and after the flash to train our network. Active flash aids in the classification and separation of facial characteristics. It is feasible to construct an independent model based on the face angle (with reasonable limits). When we calculate the pixel difference, however, face alignment is required. However, it is apparent that this strategy is effective. Based on the specific use cases that need to be solved, the technology could be made more complicated.

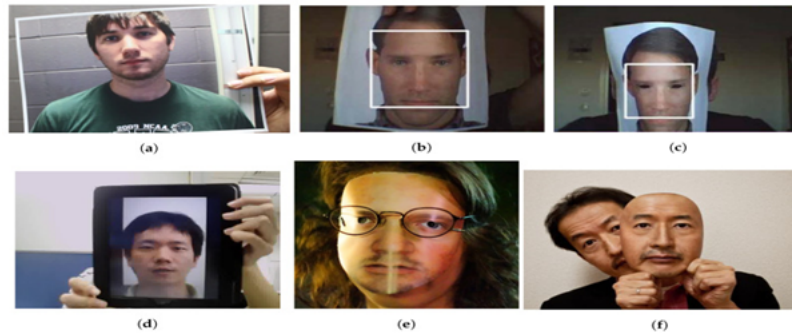


Figure 2.4: Common presentation attacks: Photo attack, three dimensional masking assaults and video replay attack comparison using their function examples

2.15 Facial Anti Spoof Technique

Using anti-spoofing techniques, we may create a presentation attack detection system (PAD) and combine it with the facial recognition system.

- Jourabloo et al. [70] suggested estimating the noise of a given spoof facial picture to identify photo/video replay assaults in 2018. (the authors also claimed that the proposed method could be applied to detect makeup attacks).
- George et al. [71] suggested Deep Pixel wise Binary Supervision (DeepPixBiS) for face PAD in 2019, based on DenseNet [132]. A real face input is tagged as 1 in the feature map, whereas a fake face input is annotated as 0. Only the mean value of pixels in the feature map is utilised as the score for face PAD during the evaluation / test phase.
- In 2017, Pan et colleagues proposed that for face PAD, they combine eye-blinking detection with texture-based scene context matching

2.16 Literature Review

For spoof detection they utilized two of the datasets. First are the NUAA dataset and the custom data set. There number of images are shown below:

Following are the samples of NUAA and custom created dataset:

For the purpose of face detection, the neural network used was MTCNN stands for Multi Task Cascaded Convolutional Neural Network. This neural network is made up of three networks that are connected and used one after the other in three phases. Face

Table 2.1: *NUAA total photos and custom dataset*

Class	NUAZ DATASET	CUSTOM DATASET
REAL	3781	919
SPOOF	2398	2371

**Figure 2.5:** *Detection of real and fake face using the first publically available dataset i.e NUAA***Figure 2.6:** *Detection of real and fake face using the dataset produced by the user*

locations and bounding boxes are predicted from the input image in the first phase, and the result contains false positives. The output is then subjected to regression analysis, in this way the regression of bounding boxes of output is made to remove the false positives. The bounding boxes are then refined, the faces that are detected in the output are denoised. Denoising is done through non-local algorithms. They work in the manner where they replace the value of target pixel with the mean value of all the pixels present in the image. The later step is the conversion of face into various color space like YCbCr and CIELUV

The following is a block schematic of the proposed system:

As name of this paper suggests the neural network used for the purpose of feature extraction from the output image is VGG-Face Convolutional Neural Network which is an update form of VGG-16. It specifically targets the features of the face. It is a 19 layer deep neural network, its architecture is given below:

The feature that are extracted by this network is then passed through classifier for the classification purpose of real and spoof faces. The features of the faces are detected after they passed through various color spaces, then they are passed through VGG-Face NN to extract the features and are sent as an input for classifier. The accuracy, specificity, sensitivity, and receiver operating sensitivity curve of the approach are all examined. Accuracy indicates whether or not the forecasts were right, whereas specificity and sensitivity indicate whether or not the predictions were correct.

Receiver operating curve shows the likelihood that whether the images were cor-

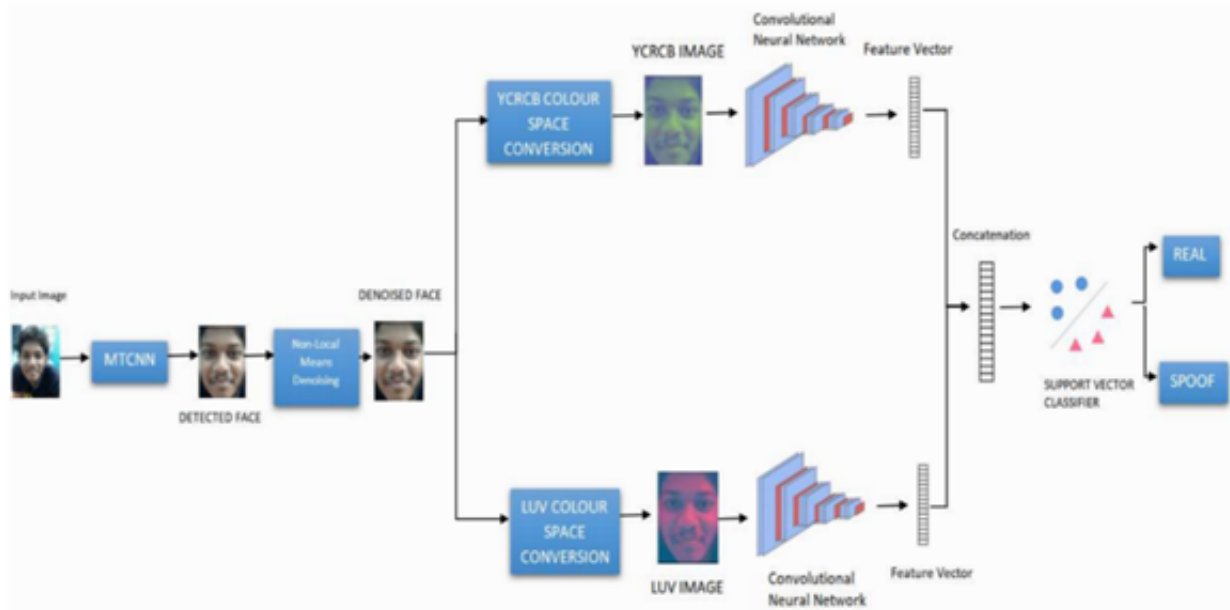


Figure 2.7: Block Schematic of the proposed system that converts the image to YCbCr CIELUV color space, pass it through CNN and obtain feature vector from it.



Figure 2.8: VGG Face Architecture that show its 16 layers consisting of 2D convolution, max pooling and average pooling

Table 2.2: Total images used

Class	TRAIN	TEST
REAL	2777	1924
SPOOF	2904	1864
TOTAL	5681	3788

rectly classified or not, it lies between 0 and 1.

Following images shows the output of spoof detection for single and multiple faces:

The accuracy of prediction is 99.6%.

This paper[3] focuses on detection of 3D face spoofing attacks on recognition systems that are a greater threat than 2D presentation attacks. The database used for this purpose consisted of three types.

- 3D face spoofing database.
3DMAD is the database that produces 17 masks of the users and records video for the purpose of presentation attacks. 3DFS database produces 26 print models

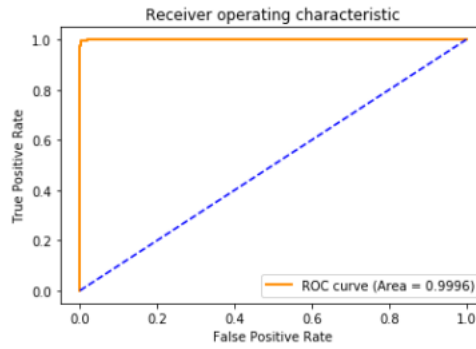


Figure 2.9: Receiver Operating Characteristics Curve that tells the likelihood whether the images were correctly classified or not

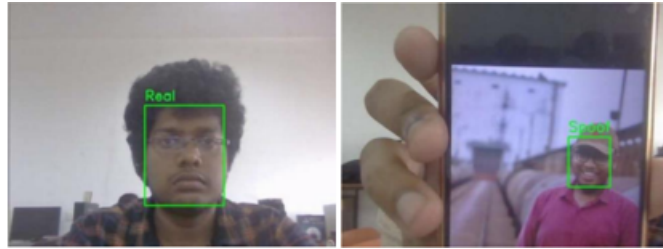


Figure 2.10: Spoof detection of a single face using VGG16 architecture which consists of 2D convolution, max pooling and average pooling

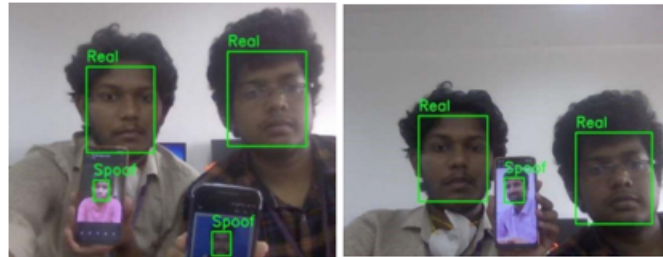


Figure 2.11: Spoof detection of a multiple face using VGG16 architecture which consists of 2D convolution, max pooling and average pooling

using printers. For efficient detection various kinds of 3D spoofing databases were used under different light conditions.

- Still picture database of wax figures. All the images that were of low quality and difficult to recognize the faces were removed from this database. There are 2300 photos in all, with 745 subjects (both real and wax faces).
- Wax figure face of database of videos. Video wax faces were collected from the internet as short videos. Then the data is cleaned by keeping only those videos that has frontal pose. A total of 285 video were included in this database out of which 145 were wax figure face videos and rest were real face videos.

Distinctive features of human face were generated using a fine grained manner in order to detect human face. To carry out detection of face spoofing they combined the skin



color with factorized bilinear pooling.

The flowchart for the proposed MC FBC system for 3D face spoofing is as follows:

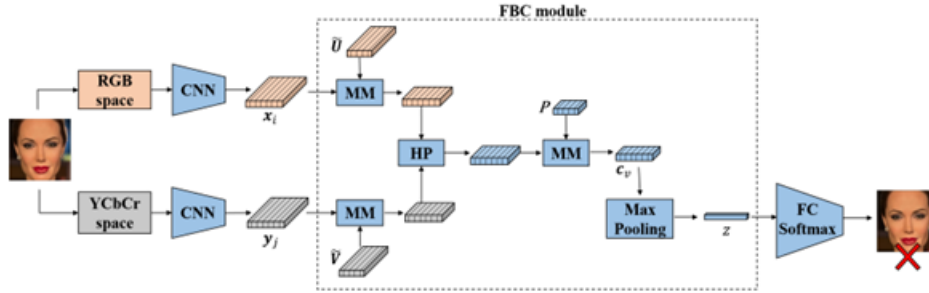


Figure 2.12: The flowchart for the proposed MC FBC system for 3 dimensional face spoofing is as follows

MM stands for matrix multiplication, HP for Hadamard product, and x_i and y_j for features obtained via convolutional neural networks. They evaluated the influence of colour spaces on detection performance using the super realistic 3D face faking database. In addition to the MC FBC approach, there are a variety of feature fusion techniques. Finally, the method's generalizability is demonstrated by inter-database testing utilising a database and various 2D/3D face spoofing databases.

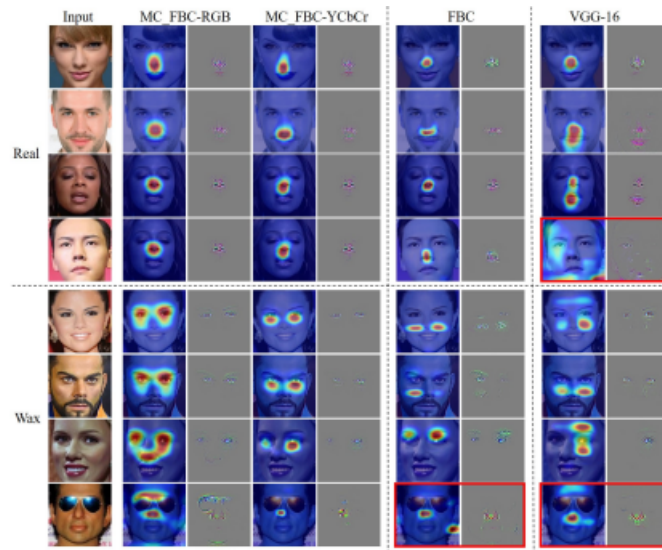


Figure 2.13: Face spoofing database for RGB and YCbCr color space for real faces and faces made from wax

The suggested MC FBC method combines complementary features derived using CNN models employing factorised bilinear coding from two colour spaces (VGG-16 and ResNet-50) (RGB vs. YCbCr). We established a new database (WFFD) with wax figure faces comprising both images and videos with significant diversity and large topic size as a super realistic face presentation attack. Extensive experimental results have shown that the suggested method outperforms various existing PAD algorithms, as well as human-based spoofing detection, in distinguishing actual faces from wax figure faces. Our method beat earlier 3D mask and 2D face spoofing databases in both intra-database and inter-database testing.

It's worth noting that the suggested scheme's best performance during interdatabase testing still has a 10% error rate. The performance merits further examination. Face spoofing attacks that are extremely realistic are difficult to differentiate from other sorts of assaults. Even for humans, there are actual ones learning-centered when paired with liveness cues, approaches are a viable combination in the future to enable effective and universal spoofing detection. However, as AI technology improves, when things move quickly, more difficult spoofing is likely. Deepfakes and other such attacks will become increasingly potent.

This paper[2] discusses about the method used to prevent spoofing of faces using deep Convolutional Neural Network.

The dataset used for this approach is comprised of real and fake images, where 1634 images are used for fitting in the model and for the purpose of the dataset, 80 percent of the total photos are used for training. Twenty percent of the photographs are utilised for testing.



Figure 2.14: *The database used for MC FBC approach comprises of these fake images*



Figure 2.15: *The database used for MC FBC approach comprises of these real images*

In this model CNN is used to recognize image patterns and classify them. images were passed through various layers of CNN. Following image shows the complete architecture of CNN:

In the beginning convolutional layer extract the input image, using different kernels various functions are performed on the input image these includes edge detection, sharpening, box blur and gaussian blur. When there is a difficulty is fitting the image padding was used. The next step is usage of ReLU for non-linear function.

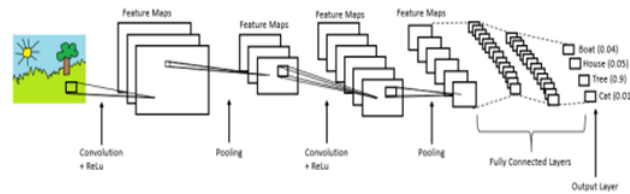


Figure 2.16: CNN architecture that consists of conv2D convolution, max pooling, Relu and fully connected layers

The model of CNN for implementation here is given:

Model: "sequential"		
Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 48, 48, 256)	2560
activation (Activation)	(None, 48, 48, 256)	0
max_pooling2d (MaxPooling2D)	(None, 24, 24, 256)	0
conv2d_1 (Conv2D)	(None, 22, 22, 256)	590080
activation_1 (Activation)	(None, 22, 22, 256)	0
max_pooling2d_1 (MaxPooling2D)	(None, 11, 11, 256)	0
conv2d_2 (Conv2D)	(None, 9, 9, 256)	590080
activation_2 (Activation)	(None, 9, 9, 256)	0
max_pooling2d_2 (MaxPooling2D)	(None, 4, 4, 256)	0
flatten (Flatten)	(None, 4096)	0
dense (Dense)	(None, 1)	4097
activation_3 (Activation)	(None, 1)	0
dense_1 (Dense)	(None, 1)	2
activation_4 (Activation)	(None, 1)	0

The images that were convolved using more layers of convolution gave best results. So in this model there were 3 activation and 3 pooling layers which provided 98.9% accuracy of images in detection.

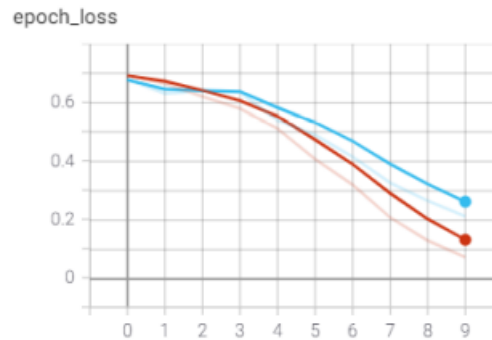


Figure 2.17: *Epoch Loss*

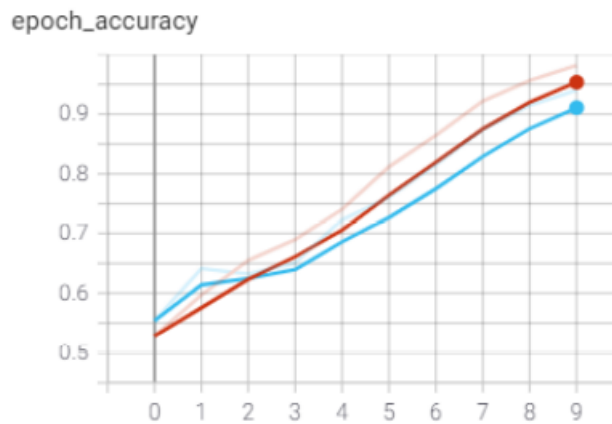


Figure 2.18: *Epoch Accuracy*

The loss decreased whereas the accuracy of epoch increased. Testing accuracy shown in blue color indicates that it is a bit less than that of training accuracy and it results in overfitting. The paper concludes that using more number of layers in convolutional neural network makes it crucial to analyze the performance of the network with accuracy.

2.17 Summary

In this chapter we discussed about presentation attacks and their types. There are four types of presentation attacks that are photo attacks, video replay attacks and three-dimensional masking attacks. Later we discussed about anti spoofing solutions that are eye blink detection, CNN, active flash, challenge response and 3D camera. Their functional working is described. We studied about first publically available dataset and other created by the user themselves.

CHAPTER 3

Methodology

3.1 SIW

The SiW Database [33] is the first to integrate face spoofing assaults with a variety of tagged postures and emotions. This collection contains 3300 assault films and 1320 actual access videos from 165 people. It covers subjects from a wider range of ethnicities than the databases described above. This dataset includes two types of print (photo) assaults and four types of video replay attacks. Four spoof media (PAIs) were used to construct video replay attacks: two cellphones, a tablet, and a laptop. Four separate sessions were recorded, each with distinct head poses/camera distances, facial emotions, and lighting circumstances.

3.2 Algorithm

Various methodologies are there in order to deal with facial spoofing or to detect the presentation attack. One of the main method is PAD presentation attack detection. Presentation attacks can be of 3 types the image based, video relay based and the mask based.

3.3 Image Classification

Image classification means the process in which the image is grouped or segmented into different parts depending upon there features. Features of every image differs from the other some of which can be recognized by the normal human eye but some cannot. These characteristics might include an image's borders, pixel intensity, or pixel value changes. The most difficult aspect of working with photographs is the ambiguity of the attributes that the images rely on. A picture is made up of little indivisible parts called

DATASETS	INTELLECTUAL PROPERTY	VEDIOS AND ATTACKS
MSSPOOF	Idiap Research institute	21 clients,35 images in both VIS and NIR
REPLAY MOBILE	Idiap Research institute	1190 video clips of photo/video attack attempts to 40 clients
REPLAY ATTACK	Idiap Research institute	1300 video clips of photo/video attack attempts to 50 clients
3D MAD	Idiap Research institute	76500 frames of 17 persons/3 different sessions each session have 5 videos of 300 frames
SIW	MSU (Michigan state university) COMPUTER VISION LAB	4,478 videos
CASIA SURF	Paper With code	21,000 videos
QULU-NPU	University of Oulu	4950 real access and attack videos
CASIA MFSD	Paper With code	600 video recordings/ 240 videos of 20 subjects(training), 360 videos of 30 subjects(testing)

Figure 3.1: CNN.

pixels, each of which has a different strength known as pixel intensity. The images are either colored or grayscale we usually work with gray scale images because they are simple to work with. In colored images or RGB image there are three color channels Red, Green, Blue and with All colour palettes are conceivable when these three colours are combined. We operate with three channels whenever we work with a colour image. Because the image is made up of several pixels, each pixel in a colourful image has three channels. When working with photos, we must use more complex approaches to identify the image's edges or underlying patterns of the many characteristics in a face, which may then be used to classify or label the images.

3.4 CNN

Artificial intelligence has made significant progress in narrowing the gap between human and computer capacities. To reach fantastic outcomes, researchers and enthusiasts concentrate on a range of aspects of the area. One example is the field of computer vision. The goal of this field is to enable machines to see and perceive the world in the same way that humans do, and to use that knowledge to perform tasks such as image and video recognition, image analysis and classification, media recreation, recommendation systems, natural language processing, and so on.

A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning system that can take an input picture and assign importance (learnable weights and biases) to various aspects/objects in the image, as well as differentiate between them. ConvNet re-

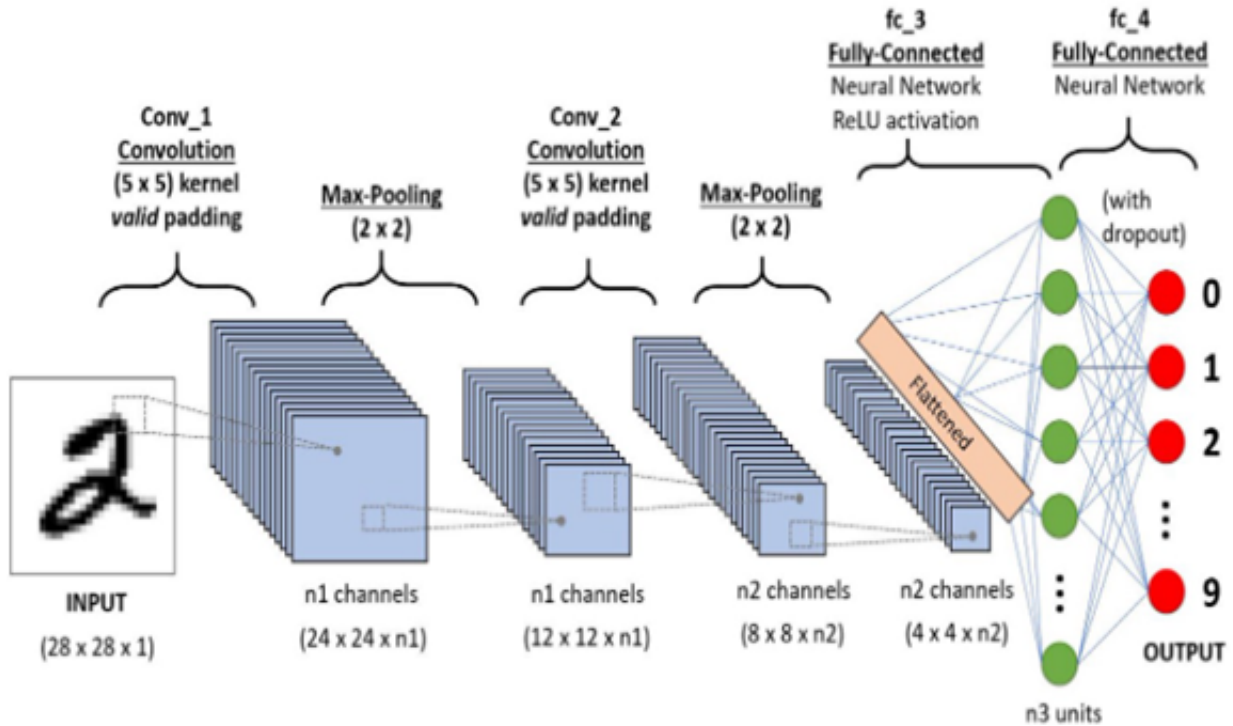


Figure 3.2: CNN architecture that consists of con2D convolution, max pooling, Relu and fully connected layers

quires far less preprocessing than other classification techniques. FilterConvNets can learn these filters//characteristics with enough training, whereas simple techniques need handengineering. A ConvNet's design is inspired by the Visual Cortex's organisation and is similar to the connection pattern of Neurons in the Human Brain. Individual neurons can only respond to stimuli in the Receptive Field, a tiny portion of the visual field. Several of these fields come together to fill the complete visual field. An image is nothing more than a collection of pixel data. Why not simply flatten the picture (e.g., a 3×3 image matrix into a 9×1 vector) and use a Multi-Level Perceptron to classify it? No way, no how.

The approach may provide an average precision score when doing class prediction on relatively basic binary pictures, but it will have little to no accuracy when dealing with sophisticated images with pixel dependencies throughout. A ConvNet may capture the spatial and temporal relationships in an image by using appropriate filters. The architecture achieves better fitting to the picture dataset because to the reduced number of parameters and reusability of weights. To put it another way, the network may be trained to better recognise the image's sophistication.

The purpose of the Convolution Operation is to extract high-level properties such as edges from the input picture. There is no requirement for ConvNets to have only one Convolutional Layer. Traditionally, the first ConvLayer is in charge of recording

LowLevel data such as edges, colour, gradient direction, and so on. The architecture responds to HighLevel properties as more layers are added, resulting in a network that understands all of the photographs in the dataset in the same manner that people do. The procedure yields two sorts of results: one in which the convolved feature's dimensionality is reduced when compared to the input, and another in which the dimensionality is raised or unaltered. This is performed if Valid Padding is utilised.

We discover that the convolved matrix is of dimensions $5 \times 5 \times 1$ when we augment the $5 \times 5 \times 1$ picture into a $6 \times 6 \times 1$ image and then apply the $3 \times 3 \times 1$ kernel over it. Same padding, thus the name. When we execute the same operation without padding, however, we get Valid Padding, which is a matrix with the same dimensions as the Kernel ($3 \times 3 \times 1$).

3.5 Pooling Layer

The Pooling layer, like the Convolutional Layer, is responsible for reducing the spatial size of the Convolved Feature. The computer power required to process the data is lowered as a result of dimensionality reduction. It also helps maintain the model's training process going smoothly by extracting rotational and positional invariant dominating features. There are two forms of pooling: maximal pooling and average pooling. Max Pooling returns the maximum value from the part of the image covered by the kernel. On the other hand, Average Pooling returns the average of all the values from the image's Kernel section. Noise can also be reduced by using Max Pooling. It eliminates all noise reduction and dimensionality reduction. Average Pooling, on the other hand, is essentially a noise suppression approach that reduces dimensionality. As a consequence, we may say that Max Pooling performs better than Average Pooling.

The Convolutional Layer and the Pooling Layer make up the i -th layer of a Convolutional Neural Network. The number of such layers may be increased depending on the picture complexity to capture even more low-level characteristics, but at the cost of increased processing power. After using the aforementioned strategy, we were able to get the model to grasp the characteristics. The final result will then be flattened and sent to a traditional Neural Network for classification.

3.6 Fully connected layers

Adding a Fully-Connected layer is a (usually) low-cost method of learning non-linear combinations of high-level information represented by the output of the convolutional layer. The Fully-Connected layer is learning a possibly non-linear function in this area. Now that we've converted the image to a format suitable for our MultiLevel Perceptron, we'll flatten it into a column vector. Backpropagation is used in each round of training to transmit the flattened output to a feedforward neural network. Over a number of epochs, the model can discriminate between dominant and low-level characteristics in pictures and categorise them using the Softmax Classification technique.

Several CNN designs have proven effective in designing algorithms that enable AI today and will continue to do so in the future. A few examples are as follows:

- LeNet.
- Alex Net

- VGGNet
- Google Net
- ResNet.
- ZFNet.

3.7 Models VGG16

K. Simonyan and A. Zisserman of the University of Oxford proposed the VGG16 convolutional neural network model in their paper "Very Deep Convolutional Networks for Large-Scale Image Recognition." The model achieves 92.7 percent top-5 test accuracy in Image Net, a dataset of over 14 million pictures belonging to 1000 classes. The model that was presented to the ILSVRC-2014 was a well-known one. It outperforms Alex Net by substituting large kernel-size filters with many 3x3 kernel-size filters in the first and second convolutional layers, respectively. VGG16 had been training on NVIDIA Titan Black GPUs for weeks.

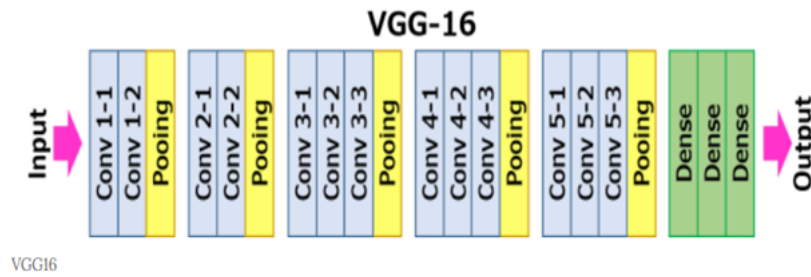


Figure 3.3: VGG Face Architecture that show its 16 layers consisting of 2D convolution, max pooling and average pooling

With the assistance of an image, the architecture is better defined: A $224 \times 224 \times 3$ RGB

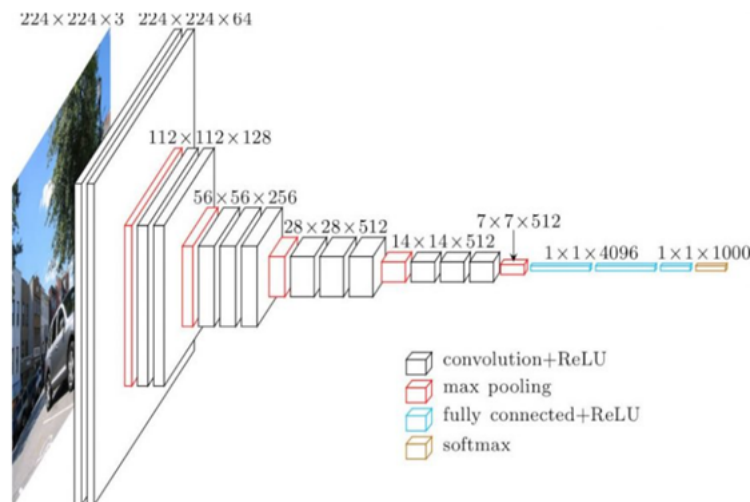


Figure 3.4: CNN architecture that consists of con2D convolution, max pooling, Relu and fully connected layers

image with a defined size is sent into the cov1 layer. The picture is processed using a stack of convolutional (conv.) layers with a very small receptive field: 3x3 (the smallest size that can capture the concepts of left/right, up/down, and centre). In one of the settings, it additionally includes 11 convolution filters, which may be regarded of as a linear change of the input channels (followed by nonlinearity). For 3x3 convolution layers, the convolution stride is set to 1 pixel, and the spatial padding of convolution layer input is set to 1 pixel to maintain spatial resolution after convolution. To conduct spatial pooling, five maxpooling layers follow sections of the conv (not all the conv. layers are followed by maxpooling). Stride 2 is used to maxpool over a 2x2-pixel frame. Three FullyConnected (FC) layers are added after a stack of convolutional layers (of varied depth in different architectures): the first two have 4096 channels each, while the third does 1000-way ILSVRC classification and so has 1000 channels (one for each class). The final layer is the softmax one. The fully linked tiers in all networks are set up in the same way. All hidden layers have the rectification (ReLU) nonlinearity. It's also worth noting that, with the exception of one, all of the networks use Local Response Normalization (LRN), which does not enhance performance on the ILSVRC dataset but does increase memory use and computation time. USE CASES AND IMPLEMENTATION:

Unfortunately, VGGNet has two major flaws:

- The network architecture weights (in terms of disk/bandwidth) are rather large.
- Training is extremely slow.

Due to its depth and number of fully linked nodes, VGG16 is over 533MB in size. This makes VGG deployment harder. Although VGG16 is used in many deep learning image classification problems, smaller network topologies are typically favoured (such as Squeeze Net, Google Net, etc.). However, because it is simple to apply, it is an excellent learning tool.

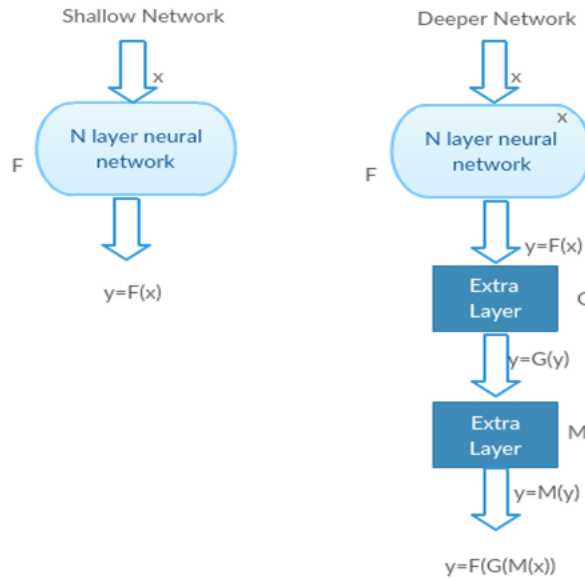
- Pytorch
- Tensor flow
- Keras

3.8 ResNet

When deeper networks start to converge, a degradation problem arises: as network depth grows, accuracy gets saturated and rapidly degrades. Shallow networks can be used in the early levels of a deeper model, whereas the remaining layers can simply act as identity functions (Input equal to output). The deeper network's extra layers approximate the mapping better than the shallower equivalent and reduce inaccuracy by a significant margin.

In the worst-case scenario, the network's shallow and deeper variations should produce the same accuracy. In the rewarded case, the deeper model should yield higher accuracy than its shallower equivalent. Deeper models, on the other hand, score badly in testing with our present solvers. As a result, the model's performance suffers when deeper networks are used. To address this problem, this study employs the Deep Residual Learning paradigm. Instead of learning a straight $x \rightarrow y$ mapping, a function $H(x)$ is

used (A few stacked non-linear layers). $H(x) = F(x)+x$, where $F(x)$ and x respectively signify the stacked non-linear layers and the identity function (input=output).



3.8.1 Designing of Network

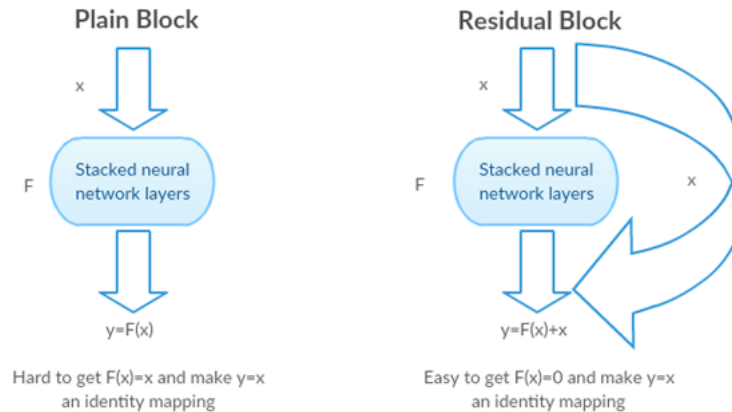
- Most 3*3 filters are used in network design
- CNN layers with stride 2 for down sampling
- A 1000-way fully-connected layer with Soft-max at the end, as well as a global average pooling layer.

3.9 Residual blocks and intuition

Instead of employing a stack of non-linear layers to fit an identity mapping (x , input=output), we may easily push the residuals to zero ($F(x) = 0$) if the identity mapping is optimal. It's much easier to come up with a solution like $F(x) = 0$ rather than $F(x) = x$ using a stack of non-linear cnn layers as a function (Think about it). $F(x)$ is the Residual function, according to the authors.

3.10 Viola Jones

The Viola-Jones algorithmic rule, developed by Paul Viola and Michael Jones in 2001, is an associate degree object-recognition framework that permits the identification of visual characteristics in real time. Viola-Jones, although being an outdated framework, is rather strong, and its use has attempted to be particularly significant in real-time face recognition. In viola jones training and testing, there are two levels. Viola-Jones draws a box (as seen on the right) and looks for a face at regular intervals within it. It's all



about determining these Haar-like characteristics, which will be detailed later. When questioning about each tile in the picture, the box slides a step to the right. A huge box is employed in this example, as well as large stairs for presentation. With fewer stages, a variety of boxes observe face-like attributes, and the data from all of these boxes together aids the rule in determining where the face is.

3.11 Features of HAR

The Haar-like choices are called after King Haar, a Hungarian physicist who discovered the Haar wavelet concept in the nineteenth century (kind of just like the relative of Haar-like features).

The alternatives below depict a box with a light-weight and a dark aspect, but the machine selects which feature is active.

One side is usually lighter than the other, and the central piece is usually shinier than the surrounding boxes, which may be mistaken for a nose.

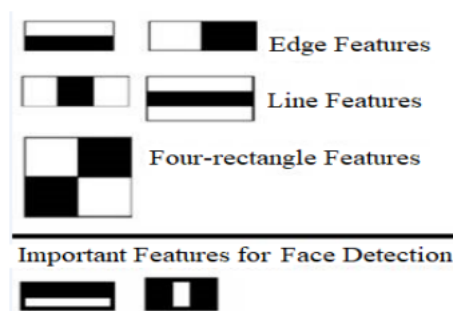


Figure 3.5: Features of Haar that includes edge features, line features, four rectangle features and some other important features

Viola and Jones detected three sorts of Haar-like characteristics in their research: Edge characteristics Line-features Four-sided characteristics These characteristics help the machine understand what the image is. Consider how a table's sting might seem in black and white. One facet will be brighter than the other, creating a bw characteristic as seen in the image above. The horizontal and hence the vertical options in the two

mandatory Face Detection settings explain how the system perceives the eyebrows and thus the nose separately. Design Procedure: Coming towards the procedure for our project, it has following two domains of work:

- Software.
- Hardware.

3.11.1 Software Aspect

Project includes two main methods for software operation:

- Image processing.
- AI based machine learning.

Image Processing

Digital image processing is the process of processing digital pictures using a computer and an algorithm. We will process digital photographs using various image processing methods. We'll analyse the image for machine learning after applying different smoothing, sharpening, and other filters.

MACHINE LEARNING

Machine learning is a branch of artificial intelligence that deals with data self-learning. AI (ML) is the study of computer computations that improve over time as a result of experience and data collection. It is regarded as a component of artificial intelligence. AI computations build a model based on example data, known as "preparing data," to make projections or decisions without being explicitly programmed to do so.

3.11.2 Hardware Aspect

We will decide while working that on which hardware we are going to perform or implement in our fyp according to the suitability and advancement.

- GPU.
- RASPBERRY PI.

Graphic Processing Unit

Graphics processing technology has progressed to provide different advantages in the computing world. New graphics processing units (GPUs) provide up new possibilities in vice, content production, machine learning, and other areas. What Does a Graphics Processing Unit Do? The graphics processing unit, or GPU, has become one of the most important types of computing technology for both personal and corporate use. The GPU, which is designed for multiprocessing, is used in a broad range of applications, including graphics and video rendering. GPUs have become more popular in creative production and computer science, although being primarily known for their skills in vice (AI). They got more adaptable and programmable throughout time, expanding their capabilities. Advanced lighting and shadowing techniques allowed graphics

programmers to create more attention-grabbing visual effects and realistic sceneries. Other developers began to use GPUs to drastically speed new tasks in high-performance computing (HPC), deep learning, and other areas. GPU and CPU: Although CPUs have continued to improve performance through beaux arts breakthroughs, higher clock rates, and therefore the addition of cores, GPUs are specifically designed to speed special effects operations. When purchasing a system, it's helpful to understand the function of the CPU and GPU so you can design the best of both. What's the Difference Between a GPU and a Graphics Card? While the phrases graphics card (or video card) are sometimes used interchangeably, there is a subtle difference between the two. A graphics card is an add-in board that contains the GPU, similar to how a motherboard contains the CPU. This board also comprises the many components required for the GPU to function and connect to other devices the remaining components of the system. There are two types of GPUs: integrated and separate. An integrated GPU does not appear on its own separate card in any way, but is instead incorporated into the CPU. A separate GPU is a chip that is installed on its own printed circuit and is often coupled to a PCI slot.

Raspberry PI

The Raspberry Pi Foundation, in collaboration with Broadcom, developed a series of small single-board computers (SBCs) in the United Kingdom. The Raspberry Pi project began with the goal of demonstrating fundamental software engineering in schools and developing countries.

3.11.3 Code Description

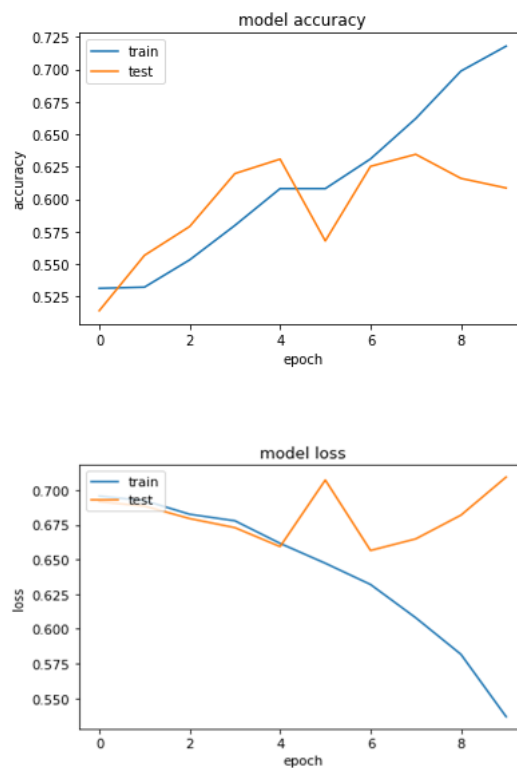
- A CNN network has been used using Keras library
- Keras acting as interface for tensor flow library
- Once a person captures his/her photo in video, the model detects the face and labels it accordingly. The dataset is big and stored on a hard disc. So, using the code below, we can directly link Google Drive to Colab. 20 epochs are used which are iterations basically. Loss function is binary cross entropy. For binary classification, it is employed, whereas categorical cross entropy is used for multiclass classification.
- In below code, function is used to preprocess images of dataset.
- It is the first step after connecting with the dataset. This function basically reads all images and converts them in 64 by 64 size.
- Then, it normalizes all pixels to reduce sharp intensity in images.
- Meanwhile it gets the name of all images and then after preprocessing, an array is generated which contains normalized images and their names.
- This code contains a machine learning model used in it. It is a sequential model which contains 2D convolution block.
- In this block, 32 filters are applied with kernel size 3 by 3.

Chapter 3. Methodology

- Input shape is 64 by 64 with 3 color channels.
- RELU Activation function used.
- Pooling size is of 2 by 2.
- The dropout rate is used to minimise the number of activated neurons in each layer.
- In end, all layers are flatten and the whole model is compiled.

A confusion matrix is a table that shows how well a categorization method performs.

Graph of training and test accuracy on dataset using this model.



To test code in realtime, camera permission is necessary

Machine learning model has an output layer which uses 0,1,2 â. As labels for the number of classes of dataset. So, 0 indicate fake faces, 1 for real and 2 for no face in video

3.12 Conclusion

The main objective of this project is to detect real and fake faces in real time video. For this purpose, a machine learning model is developed using a convolutional neural network which comes through Keras library. It provides a Python interface for artificial neural networks. Keras acts as an interface for the Tensor Flow library. Once a person captures his/her photo in video, the model detects the face and labels it accordingly. After analyzing our output we can see our model can detect between real fake faces

with an accuracy of 75 percent. The VGG 16 model is trained by using real and fake face dataset through 20 epochs then by default learning rate is set to 0.2 the value of cost function is optimized by adam optimizer. The classifier is used for classification of real and fake users. Loss function is binary cross entropy which is used for multiclass classification. Preprocessing images of dataset is done followed by normalization of pictures after preprocessing an array of normalized images is generated and 32 filters of size 3x3 are applied .input of size 64x64 of three color channel is used . The model is trained using google colab to get faster output during the training process .VGG 16 model achieved 75 percent accuracy on real and fake face dataset to detect.

3.13 Summary

In this chapter we discussed about datasets and their types, labeled and unlabeled dataset. Later we discussed about SIW dataset that is the first to integrate face spoofing assaults with a variety of tagged postures and emotions. Then we discussed about working of convolution neural network's each layer in detail. First is the pooling layer, second is fully connected layer and its working. We shifted towards VGG16 model and its architecture description. VGG-16 is a convolutional neural network that is 16 layers deep. You can load a pretrained version of the network trained on more than a million images from the ImageNet database [1]. The pretrained network can classify images into 1000 object categories, such as keyboard, mouse, pencil, and many animals. Then we discussed about resnet. ResNet is a powerful backbone model that is used very frequently in many computer vision tasks. ResNet uses skip connection to add the output from an earlier layer to a later layer. This helps it mitigate the vanishing gradient problem. Lastly we talked about viola jones algorithm and its steps.

CHAPTER 4

Results and Analysis

4.1 Datasets

Models for machine learning are only as good as the data on which they are trained. Even the most efficient machine learning algorithms will fail to execute without high-quality training data. Early in the training process, quality, accurate, comprehensive, and relevant data is required. Only with good training data can the algorithm quickly pick up the features and find the associations it needs to forecast the future. Training datasets, learning sets, and training sets are all terms for training data. It's an important part of any machine learning model that helps it generate correct predictions or complete a task. Simply, the machine learning model is built using training data. It explains how the desired result should seem. The model analyses the dataset frequently in order to fully comprehend its characteristics and to improve its performance. Training data can be divided into two types: labelled data and unlabeled data.

4.1.1 Labeled dataset

A collection of data samples containing one or more relevant labels is referred to as labelled data. Labels describe certain qualities, properties, classifications, or enclosed items, and it's also known as annotated data. Fruit photos, for example, can be labelled as apples, bananas, or grapes. In supervised learning, labelled training data is used. It allows machine learning models to learn the traits associated with specific labels, which can then be used to classify subsequent data points. In the example above, this means that a model may learn the characteristics of distinct fruits from labelled image data and apply that knowledge to group fresh photographs.

4.1.2 Unlabeled dataset

Unlabeled data is the polar opposite of labelled data, as one might imagine. It's unlabeled data, or data that hasn't been labelled with any classifications, features, or attributes. It's utilised in unsupervised machine learning, where ML models must detect patterns or similarities in data in order to draw conclusions. In the prior example of apples, bananas, and grapes, the images of such fruits will not be classified in unlabeled training data. Each image will be evaluated by the model based on its features, such as colour and shape.

The dataset we are using have total 960 training fake images training real as 1081

Training fake has been further divided into 80% training fake data 20% testing fake data

Training real has been further divided into 80% training real data 20% testing real data
As a whole data for training is 1633 and testing is 408.

The CASIA-FASD Database [19] is the first freely accessible face PAD dataset with printed picture and video replay assaults. The CASIA-FASD database contains three forms of spoofing attacks: distorted printed images (which imitate paper mask assaults), printed photos with cut eyes, and video attacks (motion cue such as eye blinking is also included). Each actual face video and spoofing assault video is categorised into three quality levels: low, medium, and high. The high-resolution video is 1280 720 pixels, while the low/normal-resolution video is 640 480 pixels. However, rather than rigorous quantitative criteria, the poor and normal quality is defined experimentally by the perceptual experience. A training set (with 20 participants) and a testing set are created from the whole database (containing 30 subjects).

4.2 Output Analysis

This project's main goal is to identify authentic and phoney faces in real-time video. A convolutional neural network from the Keras toolkit is used to create a machine learning model for this purpose. It gives artificial neural networks a Python interface. The Tensor Flow library is accessed using Keras. Once a person captures his/her photo in video, the model detects the face and labels it accordingly. After analyzing our output we can see our model can detect between real fake faces with an accuracy of 75 percent.



Figure 4.1: *Real image.*



Figure 4.2: *Fake image.*



Figure 4.3: *Fake image.*

Our model will also detect a no face image in below image.

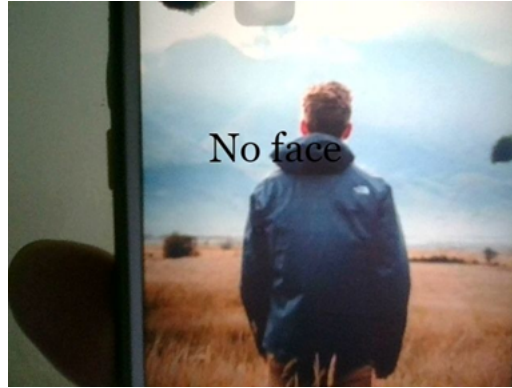


Figure 4.4: *No face image.*

4.3 Results

In our final year project we were required to implement the facial Anti Spoofing. We implemented it using the model of VGG 16 a custom layered model and the other one on two datasets.

We run the custom vgg 16 model on Celeb A Spoof dataset by taking round about 4k images and the results we obtained are given in the below diagram.

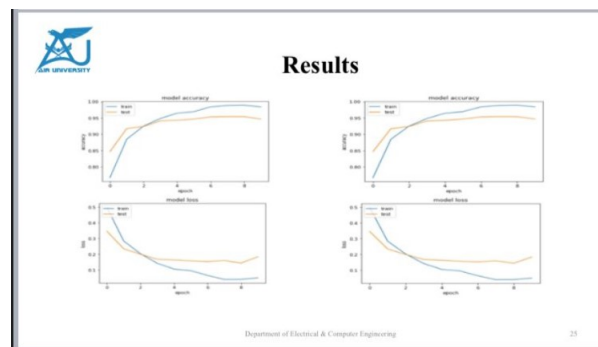


Figure 4.5: *Custom Vgg implemented on celeb A Spoof dataset.*

Chapter 4. Results and Analysis

Then we implemented the vgg 16 model on the same dataset and obtained the results. The results and the confusion matrix showing the test set accuracy is given in figure 4.6.

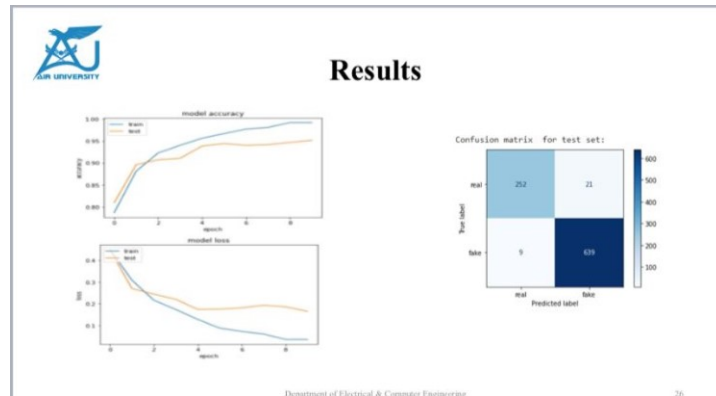


Figure 4.6: Vgg 16 implemented on celeb A Spoof dataset.

The result of other dataset (real and fake images containing about 2000 images) in a form of confusion matrix when vgg 16 is implemented on it is given below:

	Confusion Matrix	
real	144	71
fake	89	102
	real	fake

Department of Electrical & Computer Engineering

Figure 4.7: Confusion matrix obtained from dataset real and fake images.

4.3. Results

Following are the graphs representing the model accuracy and loss when the vgg 16 model is implemented on real and fake images dataset.



Figure 4.8: Vgg 16 on dataset real and fake images.

CHAPTER 5

Environment and Sustainability

5.1 Background

Artificial intelligence has made such rapid development over the past 10 years that it is now seen as the instrument of choice to address environmental problems, particularly greenhouse gas emissions (GHG). The deep learning community started to discover that training models with a growing number of parameters uses a lot of energy, which leads to increased greenhouse gas emissions. To our knowledge, there has never been a direct discussion of the whole net environmental implications of AI solutions for the environment (AI for Green), beyond GHG. We suggest examining the potential drawbacks of AI for the environment in this post. We examine the various AI implications first, and then offer the various assessment approaches.

The process of assessing the effects of an AI service is basically similar to that used for other digital services. However, AI has several unique characteristics that must be considered since they magnify its environmental implications. The first is that deep learning techniques, in particular, require a lot of data. These information must be obtained, handed over, put in storage, and processed. All of these actions have an influence on the environment and demand resources and energy. When using a surveillance satellite, the information will likely be many, however the quantity of acquisition devices may be constrained; in the case of an infrastructure for smart buildings, even though there may not be as much data, numerous devices will be necessary.

Deep neural models require a lot of computation time and resources to train, in part because the model itself learns a thorough representation that allows it to perform better data analysis. In contrast, a human will often propose a handcrafted answer when using other models, which just provide a portion of this knowledge. If the model uses continuous learning, the computation cost could even be greater.

At the same time, AI is becoming more and more popular, and AI for Green ini-

tiatives frequently promote AI as a solution to environmental issues [12, 26, 32]. No quantification of all the environmental costs associated with AI is proposed to close the loop between AI for Green and Green AI, despite the negative environmental effects being briefly alluded to, particularly rebound effects [26?] where unitary efficiency gains can result in an increase in global GHG. Because of this, it is even more crucial to be able to evaluate the actual affects while taking both good and negative effects into account. Even though AI has a considerably wider reach and at least two significant historical trends, those works frequently use the term AI to refer to deep learning techniques [9]. In this essay, we'll also pay attention to deep learning techniques, which raise particular environmental concerns and, as we've seen arXiv:2110.11822v2 [cs.AI] 21 Apr 2022, are frequently touted as potential answers to environmental challenges.

With a predicted market size of \$116 billion by 2028 and a projected CAGR of 40 percent, artificial intelligence (AI) and deep learning (DL) have been accelerating the digital transformation of technologies, industries, and society at an unheard-of rate since the 2010s. This is because DL opens up RD prospects in the automotive, health-care, defense, and other industries. More activity may be seen in academia, businesses, and individuals in an effort to address these issues in light of climate change, greenhouse gas emissions, and the shortage of fossil fuels that hang over civilizations. AI/DL is frequently positioned as the solution.

5.2 Machine in Sustaining Environment

Deep neural networks (DNNs), which are virtual constructs responsible for the explosion in AI developments, are the foundation of deep learning (DL). DNNs are able to process large amounts of big data (images, text, and numbers) to extract increasingly abstract features and learn an approximation of those features. The more abstract the features, the more computing is needed to use them for DL inference. At their most basic level, DNNs are simply a series of instructions, similar to any other software, and are thus bound by identical assumptions about performance or costs. For instance, it takes more computing to identify items in a digitized painting than it does to identify letters in handwritten text.

Functionally, DNNs imitate how the brain works by employing weights, which are N-dimensional structures of floating-point integers stored in RAM, to represent connections between artificial neurons. Artificial neurons calculate their firing by combining their inputs, weights, and an activation function. They then transmit the result on to successive network neurons, who in turn output the answer.

DL workloads depend on transistors, which represent logical 1s and 0s by switching electrical current from high to low, whether they are running on server-grade CPUs, accelerators, or mobile devices. The basic blocks of current hardware, adders, multipliers, and control units, are made up of as many as 54b transistors and exhibit simplistic behaviour when used singly.

Energy is dissipated (consumed) during transistor switching, with the amount depending on the number of active transistors and the switching speed set by the clock. Either more engaged transistors or more processing time are needed to handle more challenging DL workloads that represent numerous abstract data aspects, both of which raise the circuit's overall energy consumption proportionately. DL is more than just

training; models have a complex lifecycle that starts with experts choosing the model's goal, training data, network structure, and hyper parameters indicating how training should proceed. Once trained, DL models are used in a variety of applications, run on numerous platforms, and extract knowledge from data points that are related but not exactly the same as the ones they trained on. As a result, DL models can be seen as exact copies of the human brain that exist in numerous contexts and are subjected to various types of information that reach their sensory inputs. Each of these situations calls for calculation, which uses energy. While the environmental impact of training the instance-zero can be compared to operating five mid-sized cars for a year (transformer model), the in-field impact of its deployment for inference could have millions of replicas operating all over the world for an unpredictably long time, consuming enormous amounts of energy. Although efforts to build a more effective computational infrastructure for DL are ongoing, there are still several sociological and economic systemic obstacles that must be overcome before DL can be sustainable and have a good impact on other businesses. These obstacles include both growth and demand.

Although artificial intelligence may hold the solution to preserving our environment, climate change is a pressing issue for people everywhere. AI uses machine learning to uncover patterns in data that can be used to identify trends. Other capabilities of AI, including as object and image recognition, conversational assistants, and autonomous systems, are also assisting in the fight against climate change. AI uses a lot of energy at the same time. The massive amounts of data required to fuel AI systems must be stored in data centers, but doing so requires a lot of energy. Furthermore, complex artificial intelligence systems, such as deep learning models, can require powerful GPUs to run continuously for days at a time during training. Artificial intelligence has long had the potential to support our environment. 74 percent of survey participants agreed that AI will help to address long-term environmental concerns, according to a 2018 Intel study. Additionally on board, Intel has pledged to restore 100% of its global water use by 2025. In order to combat environmental damage, 200 research grants totaling \$50 million will be given out as part of Microsoft's AI for Earth program, which was introduced in 2017. Researchers can easily exchange knowledge and data thanks to Microsoft's service. The objective is to combine the efforts of specialists to fight climate change and avert a sixth mass extinction, using environmental research collected from all over the world and processed by an AI system. Key development for next study. For complicated weather patterns like tropical cyclones, weather fronts, and atmospheric rivers, storm researchers have AI algorithms that can predict them with up to 99 percent accuracy. Without the aid of AI, it is incredibly difficult for humans to monitor or anticipate atmospheric rivers.

5.3 How our project help in Sustaining Environment

Organizations can employ a variety of strategies to decrease their negative effects on the environment. Farmers in India employ AI-enabled technologies to increase groundnut crop yields by 30%, according to a report from the Earth Institute at Columbia University. The quantity of energy required by modern AI models is enormous, and this demand is increasing at an astounding rate. In the deep learning era, the computational resources required to create a best-in-class AI model have increased by 300,000 times

between 2012 and 2018, doubling on average every 3.4 months. The most recent example of this exponential trajectory is GPT-3. The main truth is that AI already has a sizable carbon footprint, and if current market trends continue, it will soon get much worse. In the coming years, the field of artificial intelligence may turn against us in the fight against climate change unless we are willing to reevaluate and revise the current AI research agenda.

The AI-enabled technology can assist the farmers in creating a map of the field to identify the best times to apply fertiliser to particular farmland areas and the best times to plant. Artificial intelligence advancements are essentially made possible by sheer size in the deep learning-centric research paradigm of today: greater datasets, larger models, and more computing. This phenomena is well-illustrated by GPT-3. A staggering 175 billion parameters make up the model. To put this number into context, the GPT-2, which was unveiled last year and was thought to be state-of-the-art, had only 1.5 billion parameters. The GPT-3 required many thousand petaflop-days of training time, compared to the few dozen petaflop-days needed for the GPT-2 of previous year. The issue with depending on larger and larger models to propel AI advancement is that doing so requires a significant amount of energy consumption, which results in carbon emissions. To be sure, this estimate is for a particularly energy-intensive model. Training an average-sized machine learning model today generates far less than 626,155 pounds of carbon output. At the same time, it is worth keeping in mind that when this analysis was conducted, GPT-2 was the largest model available for study and was treated by the researchers as an upper bound on model size. Just a year later, GPT-2 looks tiny—hundred times smaller, in fact—compared to its successor. For each piece of data input to them during training, neural networks perform a lengthy series of mathematical operations (both forward propagation and back propagation), adjusting their parameters in intricate ways. Therefore, larger datasets result in increased computational and energy needs. The significant experimentation and adjustment necessary to create a model is another aspect contributing to AI's enormous energy consumption. Today's machine learning still relies heavily on trial and error. During training, practitioners frequently create hundreds of variations of a particular model, experimenting with various neural architectures and hyper parameters before settling on the best layout.

5.4 Energy use and carbon emission

A compelling case study is presented in the aforementioned 2019 paper. Much smaller than attention-grabbing behemoths like GPT-3, the researchers chose an average-sized model and looked at more than simply the energy needed to. The connection between AI's energy use and carbon emissions is a fundamental premise of this discussion. What is the most effective method to approach this relationship? The EPA estimates that one kilowatt-hour of energy use results in an average of 0.954 pounds of CO₂ emissions in the United States. This average takes into account the various carbon footprints and proportional distributions of various electrical sources across the American energy grid (e.g., renewables, nuclear, natural gas, coal). This U.S. national average is used in Strubell's analysis, as was previously discussed, to determine the carbon emissions of various AI models depending on their energy requirements. It makes sense to assume that. For example, the power source mix for Amazon Web Services nearly reflects that

of the United States as a whole, and most AI.

Longer training improves accuracy while more complicated DL models permit inference of more abstract data features. Consequently, the standard approach to creating a "better" model is to invest computing resources in it, climbing the DL leader board hierarchy. Such a mindset causes DL models to outperform the requirements of their intended application areas, and development expenses are rarely published because they are challenging to measure. Globally, DL expansion may appear unstoppable and unconcerned with costs; nevertheless, we propose that this is not the case because of the "lavishness" or competition of specialists, but rather because there aren't any established ideas about the intellectual worth of DL. Being stronger or faster is the easiest approach to compete on a fundamental level, assuming no other presumptions about the competition. We think the global economy is growing.

5.5 Cost of deep learning

Based on their chip architecture and operational assumptions, two very dissimilar devices (cloud servers and mobile phones) running the same DL model could have quite distinct energetic footprints (e.g., battery-powered vs constant access to electricity). This is made worse by the fact that even small changes to the model's initial design can have a significant impact on training costs and, more importantly, in-field operation costs. For example, altering the model's data layout after training can have an impact on operating costs by changing cache access patterns and (co)processor scheduling decisions.

Due to the black box nature of DNNs and knowledge gaps in the general public, as well as the fact that management is prevented from connecting organisational-level issues like environmental costs to the practicalities of product operation, examining these relationships is challenging from the perspective of both DL designers and engineers. The primary forces that enable more sustainable AI should be the ability to explain DL models and sufficient (nongratuious) intellectual capacity.

CHAPTER 6

Conclusion and Future Work

6.1 ACHEIEMENTS

Talking about the milestones that were achieved in this project, we were successful in achieving data pre-processing, implementing deep learning models . All there achievements are listed below:

Data Preprocessing We applied a set of data pre-processing techniques to enhance our data before using it for training. This includes:

- Reshaping was used to make sure that each sample in the data set is of the same shape.
- Rescaling was used to scale down the pixels of the images to a certain range (0 - 1 in this case).
- Rotation was used to rotate some instances in the training set by a certain angle. This was done to introduce some sort of variability in our data.
- Shifting was used to shift the pixels of some images by different values (within the given range). This was also done to introduce variability in the data.
- Filling was also done to introduce variability in the data. We used nearest filling.

6.2 Deep Learning Classification

- Custom VGG-16(Real and Fake images) On our Custom model, that was built on the architecture of VGG-16, we were able to achieve the maximum accuracy of 71 Percent of the accuracy and the dataset on which the model was made comprised of almost 2000 images including the fake and real one.

Chapter 6. Conclusion and Future Work

- Custom VGG-16(Celeb A Spoof) On our Custom model, that was built on the architecture of VGG-16, we were able to achieve the maximum accuracy of 91 Percent of the accuracy before regularization and the almost 95 percent after regularization. The dataset on which the model was made comprised of almost 4000 images including the fake and real one.
- VGG-16(Real and Fake images): The accuracy we obtained was about 76.
- VGG-16(Celeb A Spoof) The accuracy we obtained by implemented this model on Celeb A Spoof was almost 94.

6.3 CONCLUSION

The main objective of this project is to detect real and fake faces in real time video. For this purpose, a machine learning model is developed using a convolutional neural network which comes through Keras library. It provides a Python interface for artificial neural networks. Keras acts as an interface for the Tensor Flow library. Once a person captures his/her photo in video, the model detects the face and labels it accordingly. After analyzing our output we can see our model can detect between real fake faces with an accuracy of 75 percent. The VGG 16 model is trained by using real and fake face dataset through 20 epochs then by default learning rate is set to 0.2 the value of cost function is optimized by adam optimizer. The classifier is used for classification of real and fake users. Loss function is binary cross entropy which is used for multiclass classification. Preprocessing images of dataset is done followed by normalization of pictures after preprocessing an array of normalized images is generated and 32 filters of size 3x3 are applied .input of size 64x64 of three color channel is used . The model is trained using google colab to get faster output during the training process .VGG 16 model achieved 75 percent accuracy on real and fake face dataset to detect.

6.4 FUTURE WORK

This project holds a lot of potential to improve and get deployed in future.

- Autoencoders: We can work on embedding auto encoders to increase the accuracies of the model further. Auto encoders can help extract lower dimensional feature map. It's a data processing approach that extracts features from raw data so that it can be utilised to train a machine learning model.
- Training on other models: To improve the accuracies we can test and train datasets on any deeper model than mobile net i.e. Resnet or other models. Also, by using Resnets we can solve the vanishing gradient problem. As we know that resnet does not allow vanishing gradient problem to occur. It has skip connections that function as gradient superhighways, which allow the gradient to flow unhindered. It makes it possible for gradients to propagate to deep layers before they can be attenuated to small or zero values. .
- Deployment: The other main objective or our future goal is deployment of our implemented project by making its web application and then further its deployment.

Bibliography

- [1] Hoogsteden, C., and P. CROSS. "Public access to GPS: government duty, economic rationality or international philanthropy?." *CISM Journal*, vol. 46, no. 1, pp. 41-53, 1992.
- [2] Schuckers, Stephanie AC. "Spoofing and anti-spoofing measures." *Information Security Technical Report*, vol. 7, no. 4, pp. 56-62, 2002.
- [3] Kollreider, Klaus, Hartwig Fronthaler, and Josef Bigun. "Evaluating liveness by face images and the structure tensor." In *Fourth IEEE Workshop on Automatic Identification Advanced Technologies (AutoID'05)*, pp. 75-80, 2005.
- [4] Galbally, Javier, S bastien Marcel, and Julian Fierrez. "Biometric antispooofing methods: A survey in face recognition." *IEEE Access*, vol. 2, pp. 1530-1552, 2014.
- [5] Bagga, Manpreet, and Baljit Singh. "Spoofing detection in face recognition: A review." In *3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, pp. 2037-2042, 2016.
- [6] Patel, Keyurkumar, Hu Han, and Anil K. Jain. "Secure face unlock spoof detection on smartphones." *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 10, pp. 2268-2283, 2016.
- [7] Galbally, Javier, and Riccardo Satta. "Three-dimensional and two-and-a-half-dimensional face recognition spoofing using three-dimensional printed models." *IET Biometrics*, vol. 5, no. 2, pp. 83-91, 2015.
- [8] Menotti, David, Giovanni Chiachia, Allan Pinto, William Robson Schwartz, Helio Pedrini, Alexandre Xavier Falcao, and Anderson Rocha. "Deep representations for iris, face, and fingerprint spoofing detection." *IEEE Transactions on Information Forensics and Security* vol. 10, no. 4, pp. 864-879, 2015.
- [9] Pinto, Allan, William Robson Schwartz, Helio Pedrini, and Anderson de Rezende Rocha. "Using visual rhythms for detecting video-based facial spoof attacks." *IEEE Transactions on Information Forensics and Security*, vol 10, no. 5, pp. 1025-1038, 2015.

Bibliography

- [10] Yang, Jianwei, Zhen Lei, Dong Yi, and Stan Z. Li. "Person-specific face antispoofing with subject domain adaptation." *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 4, pp. 797-809, 2015.
- [11] Chingovska, Ivana, and Andr   Rabello dos Anjos. "On the use of client identity information for face antispoofing." *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 4, pp. 787-796, 2015.
- [12] Gragnaniello, Diego, Giovanni Poggi, Carlo Sansone, and Luisa Verdoliva. "An investigation of local descriptors for biometric spoofing detection." *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 4, pp. 849-863, 2015.
- [13] Anjos, Andr  , Murali Mohan Chakka, and S  bastien Marcel. "Motion-based counter-measures to photo attacks in face recognition." *IET Biometrics*, vol. 3, no. 3, pp. 147-158, 2014.
- [14] Chingovska, Ivana, Andre Rabello Dos Anjos, and Sebastien Marcel. "Biometrics evaluation under spoofing attacks." *IEEE transactions on Information Forensics and Security*, vol. 9, no. 12, pp. 2264-2276, 2014.
- [15] Biggio, Battista, Zahid Akhtar, Giorgio Fumera, Gian Luca Marcialis, and Fabio Roli. "Security evaluation of biometric authentication systems under real spoofing attacks." *IET Biometrics*, vol. 1, no. 1 pp. 11-24, 2012.
- [16] Poh, Norman, Chi Ho Chan, Josef Kittler, S  bastien Marcel, Christopher Mc Cool, Enrique Argones R  a, Jos   Luis Alba Castro et al. "An evaluation of video-to-video face verification." *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 4, pp. 781-801, 2010.
- [17] Evans, Nicholas, Stan Z. Li, Sebastien Marcel, and Arun Ross. "Guest editorial: Special issue on biometric spoofing and countermeasures." *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 4, pp. 699-702, 2015.
- [18] Wen, Di, Hu Han, and Anil K. Jain. "Face spoof detection with image distortion analysis." *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 4, pp. 746-761, 2015.
- [19] Hadid, Abdenour, Nicholas Evans, S  bastien Marcel, and Julian Fierrez. "Biometrics systems under spoofing attack: an evaluation methodology and lessons learned." *IEEE Signal Processing Magazine*, vol. 32, no. 5 pp. 20-30, 2015.
- [20] Ratha, Nalini K., Jonathan H. Connell, and Ruud M. Bolle. "Enhancing security and privacy in biometrics-based authentication systems." *IBM Systems Journal*, vol. 40, no. 3, pp. 614-634, 2001.
- [21] Erdogmus, Nesli, and Sebastien Marcel. "Spoofing face recognition with 3D masks." *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 7, pp. 1084-1097, 2014.
- [22] Agarwal, Akshay, Richa Singh, and Mayank Vatsa. "Face anti-spoofing using Haralick features." In *8th IEEE International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pp. 1-6, 2016.

- [23] Boulkenafet, Zinelabidine, Jukka Komulainen, and Abdenour Hadid. "Face Spoofing Detection Using Colour Texture Analysis." *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 8, pp. 1818-1830, 2016.
- [24] Pinto, Allan, Helio Pedrini, William Robson Schwartz, and Anderson Rocha. "Face spoofing detection through visual codebooks of spectral- temporal cubes." *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 4726-4740, 2015.
- [25] Gavrilescu, Mihai. "Study on using individual differences in facial expressions for a face recognition system immune to spoofing attacks." *IET Biometrics*, vol. 5, no. 3, pp. 236-242, 2016.
- [26] Arashloo, Shervin Rahimzadeh, Josef Kittler, and William Christmas. "Face spoofing detection based on multiple descriptor fusion using multiscale dynamic binarized statistical image features." *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 11, pp. 2396-2407, 2015.