AN ANALYSIS OF CONVNETS ON DEEPFAKE DETECTION

Thesis submitted

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

MASTER OF TECHNOLOGY
in
ARTIFICIAL INTELLIGENCE

Submitted by

Arpit Mahlawat

(Roll No. 2K22/AFI/03)

Under the supervision of

Ms. Anukriti Kaushal

Assistant Professor,

Department of Computer Science and Engineering

Department of Computer Science and Engineering
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi 110042

MAY, 2024

Department of Computer Science and Engineering
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

CANDIDATE'S DECLARATION

I, Arpit Mahlawat, Roll No - 2K22/AFI/03 student of M.Tech (Department of Computer Science and Engineering), hereby declare that the project Dissertation titled "An analysis of Convnets on Deepfake Detection" which is submitted by me to the Department of Computer Science and Engineering, Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of degree of Master of Technology, is original and not copied from any source without proper citation. This work has not previously formed the basis for the award of any Degree, Diploma Associateship, Fellowship or other similar title or recognition.

Place: Delhi Arpit Mahlawat

Date: 31.05.24

This is to certify that the student has incorporated all the corrections suggested by the examiners in the thesis and the statement made by the candidate is correct to the best of our knowledge.

Supervisor's Signature Signature of External examiner

Department of Computer Science and Engineering
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

CERTIFICATE

I hereby certify that the Project Dissertation titled "An analysis of Convnets on Deepfake Detection" which is submitted by Arpit Mahlawat, Roll No – 2K22/AFI/03, Department of Computer Science and Engineering, Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of the degree of Master of Technology, is a record of the project work carried out by the student under my supervision. To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Place: Delhi

Ms. Anukriti Kaushal

Date: 31.05.24

SUPERVISOR

Department of Computer Science and Engineering
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

ACKNOWLEDGEMENT

I wish to express my sincerest gratitude to Ms. Anukriti Kaushal for her continuous guidance and mentorship that she provided me during the project. She showed me the path to achieve our targets by explaining all the tasks to be done and explained to us the importance of this project as well as its industrial relevance. She was always ready to help me and clear our doubts regarding any hurdles in this project. Without her constant support and motivation, this project would not have been successful.

Place: Delhi Arpit Mahlawat

Date: 31.05.2024

Abstract

The exponential growth of technology has led to massive increase in the use of images and videos as the main medium of sharing information, flooding most of the internet. However, this growth has also brought about various set of challenges, which include an increase in crimes such as identity theft, privacy invasion, and the spread of fake news through manipulated media. Generative adversarial networks (GANs) and certain kinds of data augmentation techniques are utilized to construct a face dataset comprising both real and fake examples for training the classification model. The proposed approach is a highly versatile model with very low inference and training costs, leveraging transfer learning with MobileNet and EfficientNet architectures. Evaluation is performed on two popular datasets: 140k Real and Fake Faces, and Real and Fake Face Detection, both used as popular benchmarks. The model achieves accuracy exceeding 99.5% and 75% on them. This show the effectiveness of this CNN-based approach combined for detecting digitally altered images.

Contents

List of Tables

List of Figures

Chapter 1

INTRODUCTION

1.1 Deepfakes and effectiveness of current SOTA deepfake generation models

In recent years, the field of deepfake generation[3] has undergone remarkable advancements, pushing the boundaries of what is possible in the realm of synthetic media. Deepfakes, which are highly convincing, manipulated digital content, typically video or audio, have sparked both fascination and concern across various sectors, from entertainment to cybersecurity.

Deepfakes are named so due to their generation via deep learning methods. Deep learning [4]itslef is a subset of machine learning and it biggest breakthroughs in deepfake generation can be attributed to the development of deep neural networks, particularly generative adversarial networks (GANs) [5] and recurrent neural networks (RNNs) [6]. When GANs were introduced they created a major impact in creating deep fakes by letting two neural networks named generator and discriminator.In such a model the generator produces synthetic content, and the discriminator evaluates its authenticity, leading to a continuous improvement in the quality of deep fakes. This pitting of the two models means it makes it easier to beat existing detection models.

One of the most popular and perhaps important applications of deepfake technology is in the entertainment industry, where it has been used to superimpose the faces of actors onto other individuals or even manipulate historical footage to bring long-lost figures back to life all without the audience knowing the edit. This has enabled a new era of storytelling, where visual effects and character replication can be achieved with incredible realism. Additionally, deepfake technology has found its way into voice synthesis and audio manipulation, allowing for the creation of convincing voice forgeries, raising concerns about the potential for identity theft, misinformation, and privacy infringement.

However, this technology brings with itself some major challenges and ethical concerns, that need to be tackled immediately. The increasing accuracy and speed of deepfake methods has the potential to deceive, manipulate, and create confusion in the less technically educated masses, which has triggered public and regulatory concerns regarding their misuse. As a result, researchers, governments, and organizations are working tirelessly to develop countermeasures and detection systems to mitigate the adversely negative effects of deepfake technology.

Further we explore the current state-of-the-art methods in deepfake generation, highlighting details about them that include key advancements, ethical dilemmas, and the ongoing efforts to address the potential risks associated with these technologies. By providing a comprehensive overview of the current landscape of deepfake generation, we aim

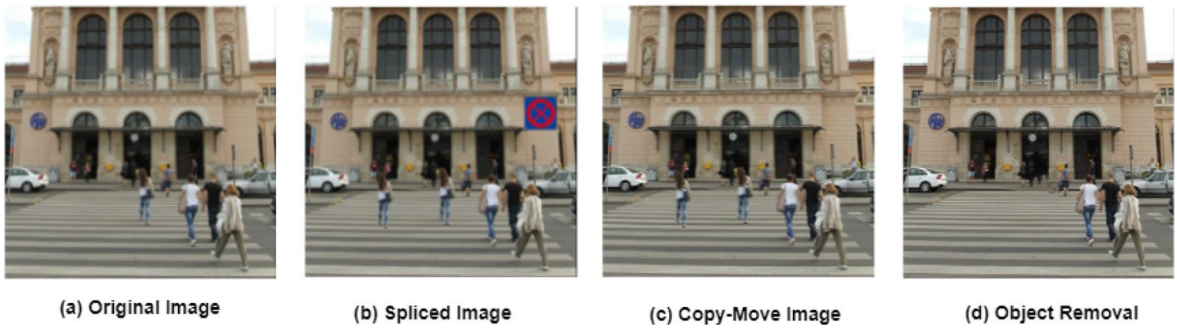

Figure 1.1: Left to Right a) Non-Tampered Image, b) Added a sign by splicing, c) Person duplicated using copy, d) Removal of person.

to equip people with a deeper understanding of the field's capabilities and challenges, and current detection solutions as well as the broader implications for society. The report also features our own solution based on previous research in order to provide deepfake detection using very few resources while maintaining a competitive advantage with latest methods.

1.2 Introduction to the Need for Deepfake Detection

In an age where digital manipulation and artificial intelligence have reached unprecedented heights, the emergence of deepfake technology has given rise to a compelling and urgent need for effective deepfake detection methods. Deepfakes, convincingly fabricated audio [7], video, or text content created using advanced machine learning techniques, pose a multifaceted challenge to our society. As we navigate the digital landscape, it has become increasingly essential to address the pressing need for deepfake detection to preserve trust, security, and authenticity in various domains.

1.2.1 Imperative for DFD

The need for deepfake detection arises from several factors:

- Deceptive Manipulation [8]: Deepfake technology enables the creation of highly realistic content that can convincingly impersonate individuals or fabricate events. This deception can have terrible consequences, from causing reputational harm to spreading disinformation and reducing our trust in the media.
- Information Integrity: With spread of deepfake content, the authenticity of information is at risk. The ability to distinguish genuine content from manipulated material is crucial for maintaining information integrity and reputation in areas such as journalism, politics, and public discourse.
- Identity Theft[8]: Deepfakes can be used for identity theft, allowing malicious actors to impersonate individuals for fraudulent purposes, such as financial scams or social engineering attacks. Detecting such impostors is essential for safeguarding personal security.

- Privacy Concerns: Deepfake technology has the potential to impact personal privacy. The creation of fake audio or video recordings can lead to the unauthorized dissemination of sensitive or compromising content. Detection methods are necessary to protect individuals from privacy violations.

- National Security [3]: In the context of national security, deepfakes can be employed to manipulate audiovisual content for political and geopolitical purposes. Accurate detection is crucial for identifying potential threats and preventing misinformation campaigns.

- Preservation of Trust: Trust in the authenticity of media is paramount in society. Deepfake detection is essential for preserving trust in various domains, including journalism, entertainment, and social interactions, where the lines between fact and fiction can blur.

- Emerging Threats[9]: As deepfake technology continues to evolve, so do the techniques employed by malicious actors. The need for proactive detection methods is essential to stay ahead of these emerging threats.

- Legislative and Regulatory Responses: Governments and regulatory bodies are recognizing the need for regulations regarding deepfake content. Detection mechanisms play a critical role in enforcing compliance with these regulations and holding wrongdoers accountable.

- Research and Innovation: The ongoing development of deepfake technology necessitates a concurrent focus on detection methods. Researchers, technologists, and organizations must continuously innovate to stay ahead of new and evolving deepfake techniques.

The need for deepfake detection is both an immediate and ongoing concern. The deceptive power of deepfakes, coupled with their potential for misuse and manipulation, underscores the critical role of detection in maintaining trust, privacy, and security in the digital age. This introduction highlights the various reasons why deepfake detection is an indispensable element of our technological landscape and a crucial tool for mitigating the adverse effects of synthetic media.

1.3 Objectives

Our objectives are to provide a comprehensive overview of the state of the art in deepfake detection techniques, highlighting the key challenges and trends in the field, and offer insights into potential future directions. At the same time we build our own approach inspired by current research in order to tackle this problem

Chapter 2

LITERATURE REVIEW

2.1 Sources and Time-line for chosen research papers

We accumulate a total of approximately 60-70 papers with a focus on 17 very solid, reputable and diverse studies from our determined sources within six years of the publication period from 2018-2023. These particular studies are based on solving pretty much most of the deepfake variations including but not limited to voice, video, picture etc. While the main focus on our report was on image deepfakes (partly due to resource constraints), studying various fields informs us about the variety of models, and how either they could be employed on completely different types of data or how combining them could possibly lead to a very effective model for detecting complex deepfakes.

2.1.1 Publication Period

The Deepfake related research primarily emerged in 2018 [10]. Therefore, the publication period which had our main focus mainly starts from the beginning of 2018 until 2023. Over the span of these 5-6 years the number of publications has increased at an exponential rate . For example during the first half of 2-18 only three publications with relevant research were published ,which doubled over the second half and continues to double every six months. This trend continues over the year, indicating the acknowledgment of research need on deepfake detection.

2.1.2 Source of Publications

Mainly we considered eight to ten different publication sources from recognized conferences, workshops, journals, and archives. According to our observation, most of the articles were published as archived papers, and only a few papers were issued in the considered journal. We didn't include the source in the table if the publication count is below two.

In the following text we have different researches according to the applied techniques and quick summaries respectively.

2.2 Quick overview of various methods and techniques

A comprehensive overview of various methods and techniques employed in the field of Deepfake detection spanning both image and video domains. Researchers have explored

a multitude of approaches, ranging from traditional computer vision techniques to advanced deep learning architectures. Summary of various ways deepfake detection has been enhanced over the years is given below

- **Deepfake Detection in Images** : Introduction of a GAN simulator replicating GAN-image artifacts for training a classifier. For Feature Extraction there are multiple kinds of proposals of networsk for extracting standard features from images. Deep Learning-Based Video Detection for example have used lates CNN models like MesoInception-4 [11]with mean squared error (MSE) loss for training.

- **Feature Extraction Techniques** : Implementation of extracted handcrafted features for input to networks. Spatiotemporal Features have also been used as features for detection. Common Textures in images along with Face landmarks frequently used for biometrics are also very useful for artifact detection

- **Enhancements and Innovations** : Data Augmentation, Super-Resolution Reconstruction [12] better feature extraction are commonly used for input dataset enhancement. Localization Strategies like pixel levels for frame analysis and Maximum Mean Discrepancy (MMD) Loss are used by some papers for discovering more general features.

- **Advanced Architectures and Techniques** : Attention Mechanism [13] was introduced to some models for better performance. Capsule-Network (CN) [14] is avery interesting architecture, which requires fewer parameters for training. Ensemble Learning techniques have of course frequently been used for increased performance.

- **Frame-by-Frame Analysis in Videos** : RNN-Based Networks [6]help extracting features at various levels. Another model uses Optical Flow-Based Technique [15] for analysis. Autoencoder-Based [16] architectures have been used with other architectures to address overfitting.

- **Frequency Domain Analysis**: Frequency Domain Analysis have been explored for analyzing image latent patterns.

- **Training types** : Some papers use Siamese networks [17], also commonly known as twin networks, are a type of neural network architecture commonly used for tasks that involve finding similarities or differences between two comparable data inputs. The term "Siamese" comes from the analogy of Siamese twins, where two identical sub-networks share the same weights and parameters. The main idea behind these networks and this kind of training is to learn a similarity metric from the data itself, rather than relying on predefined, hand-crafted similarity metrics. This is especially useful in domains where it is difficult to define an appropriate similarity metric making it a great fit for deepfake detection because when the similarity metric needs to be learned from the data itself.

This summary highlights the diversity of approaches in the field, including traditional methods, deep learning architectures, and novel techniques to improve the efficiency of current Deepfake detection systems. Researchers have explored various aspects, such as feature extraction, attention mechanisms, and frequency domain analysis, to improve the accuracy and reliability of detection systems. The following table is a quick well detailed summary of our literature review focused around the papers with most impact.

Table 2.1: Quick summary of the literature review conducted

Author	Type of operation Detected and Problem Addressed	Model Used / Detail of Model Used	Limitations
Nicolo Benneteni et al.[18] ICPR 2020	Forensics footprints are often minute and can be hard to detect . This is true for excessive compression, or strong downsampling	An ensemble different CNN models is proposed and model is obtained starting from a base network making use of : (i) attention layers (ii) siamese training.	Lack of embedding of temporal information.
Andreas Rössler et al.[19] ICCV 2019	Some additional domain related ideas increase detection to solid accuracy, even in the presence of a lot of compression.	XceptionNet, MesoInception and similar CNNs Added features based on Detection based on Steganalysis Features	Deepfakes have gotten much harder to detect by classical algorithms .
Shen Yu-Wang et al.[20] ICCV 2019	Detected Image Warping by Photoshop, where model was based around working without labeled data	Used a Dilated Residual Network variant (DRN-C-26) pretrained on the ImageNet[21] dataset. This type of architecture was designed originally for semantic segmentation	Limited to changes due to simple mathematical algorithms that work by classic editing tools
Xinru Chen et al.[22] IEEE 2022	For operations like Copy-move, Inpainting and Splicing there is a lack of generalization with current models.	By exploiting noise distribution and boundary artifact surrounding tampered regions,model learns semantic agnostic features. A new kind of network termed MVSS-Net is used.	Improved the specificity yet at the cost of a clear performance drop for pixel-level edit detection

Table 2.1: Quick summary of the literature review conducted

Author	Type of operation Detected and Problem Addressed	Model Used / Detail of Model Used	Limitations
Vishal Asnani et al.[23] 2022	Detect Removal, splicing, Copy-Move with Better robustness against Gaussian blur, Gaussian noise, JPEG compression and ISO noise	MSMG-Net Learn coarse-to-fine multiscale feature through resolution down (R-down) block and a stack of ResNet blocks.	Bad results due to artifacts in tampered regions incorporated from the subtleties with additional post-processing methods
Myung Joon-Kwong [24] IJCV 2022	Copy-Move and Splicing during Post Processing affect the DCT coefficients of processed image	A CAT-Net type proposal that uses DCT and RGB domain info for forgery localization. First study to use coefficients of DCT, straight into a segmentation network	Decreased performance when pre-trained
Shreya Ghosh et al.[25] 2023	Here, the goal is to learn some sort of task agnostic representations in self-supervised manner for face-edit tasks	MARLIN consists of (a) Representation Learning Module, (b) Facial region guided Tube Masking, (c) Downstream Adaptation.	As the model is trained on YouTube Face dataset, there could be potential bias in racial and cultural identities
David Coccomini [26] ISTI CNR 2021	Detects Image generated with VAEs and GANs as these have become accessible and accurate, resulting in videos that are difficult to detect.	Combined various types of Vision Transformers with another convolutional Network. Feature extraction done using EfficientNets.	Vulnerable against techniques such as neural textures.
Sahibzada Adil[27] APSIPA 2022	For changes such as Synthetic Lip movements , current methods are not sufficient to detect multi-modal manipulations both visual and acoustic in nature.	A Wav2lip model (pretrained) used to convert the audio in the video into a synthetic lip sequence, transformed into a vector sequence with a Resnet Model	Issues revolve around Lip occlusion, frontal face blocking. Adversarial attack may lead to poor performance of the detector.

Table 2.1: Quick summary of the literature review conducted

Author	Type of operation Detected and Problem Addressed	Model Used / Detail of Model Used	Limitations
Zhixi Cai et al.[7] DICTA 2022	A type of deepfake that contains only a small part of video/audio manipulation, which changes the meaning of the content i.e. sentimentally changed.	Used Ba-TFD known as Boundary aware Forgery Detection which combines both Audio and Video encoders.	The audio reenactment method used in the dataset does not always generate the goal style, the quality of dataset is verylimited.
Deressa Wodajo et al.[28] 2021	Face Synthesis, Face Swapping Even though each proposed mechanism have their strengths, current detection methods lack generalizability.	CNN, ViTs The First layers has structure of VGG which extracts learnable features. Then the ViT takes in the FL as input and converts these into a sequence of image pixels for the final detection process	Limited by diversity in the datasets.
Yongyi Zang et al.[29] 2023	Synthesized singing voices are typically mixed in songs with some strong background music that can easily hide those synthetic artifacts	AASIST model uses raw-waveform feature, using GNNs and incorporates spectrotemporal attention. LFCC is used for speech features that are fed into a ResNet	Unseen communication codecs and Interference from backing tracks along with Diverse musical genres
LingZhi Li et al.[30] CVPR 2020	Fake face images or deep fakes detected. When the model is used fakes generated by unseen manipulation models, most existing models see a huge performance drop	An HRNEt is used for face X-Ray twith concatenated representations from the four different resolutions to a CNN layer with one o/p channel. Here the final layers are mixed with fixed ImageNet weights	Method relies on blending step, which has low resilience against the detector

Table 2.1: Quick summary of the literature review conducted

Author	Type of operation Detected and Problem Addressed	Model Used / Detail of Model Used	Limitations
Sanjay Saha et al.[31] ICCV 2023	Deepfakes divided into frames called deepfake video temporal segmentation, but is not explored enough	Used a combo of ViTs based on scaling and shifting for spatial features and Timeseries Transformer for temporal features of the videos	Trading performance on some datasets for others.
Oscar de Lima et al.[32] ACM 2020	DeepFakes Current Deepfake detection methods only use individual video frames and therefore fail to learn from temporal information	DFT,RCN,R3D, ResNets, I3D are used with Added layers in the neural networks, that help with storing temporal information	Low ROC-AUC scores for most models used here
Ning Yu. et al. [33] IEEE 2021	Current techniques are unable to consistently prevent fakes misuse and as generative models evolve, they learn to better match the true distribution causing fewer artifacts	Uses steganography for embedding "artificial fingerprints" in training data.	Relies on proactive efforts from models that are generating deepfake
Shahroz Tariq et al. [34] 2020	Tries to address lack of generalizability of existing models. Such methods can be transferable to detect other deepfake methods	Uses CLRNet which is based on LSTMs, and takes a sequence of consecutive images as an input from a video for learning temporal information	Performance improved over previous methods but still lacking strong accuracy

Chapter 3

METHODOLOGY

3.1 Convolutional Neural Networks

Convolutional Neural Networks (ConvNets) [35]: ConvNets are a class of deep neural networks designed for processing structured linear data (usually represent in form of matrices), such as images. They typically consist of one or more layers with convolution operation performed between them, followed by pooling and fully connected layers.

- **Convolution Operation:** Convolution is a linear operation where a filter array (also known as a kernel) is slid over the data represented as matrices, performing element-wise multiplication between the filter values and the input values at each corresponding position. The result is then summed to produce the output. This operation is the major part of extracting features from the input data.
- **Conv2D Layer:** The Conv2D layer is a very common type of convolution operation used in ConvNets, especially for image processing tasks. In Conv2D, we take a 2D filter (kernel) and move it linearly over the 2D input data (e.g., an image in form of a matrix) or 3D data (the same image but now in different color channels), and perform element-wise multiplication. The output of this operation is called a "feature map" which represents the detected features by the model in the input data.
- **Filter (or Kernel):** The filter is a 2D array of weights that represent the impact of each cell while sliding over the input data and defines the pattern to be detected in the input data. It applies the same operation over the entire input matrix.
- **Depth and Channels:** Both input data and the filter need to have the same number of channels, i.e depth. For example, if the input image has the commonly used 3 channels representing the RGB channels, then the corresponding filter used for that image must have 3 channels too.
- **Feature Extraction:** Conv2D layers are used for feature extraction in ConvNets. These layers learn important features from the input data without human supervision, making them effective for the most obvious image processing tasks such as image classification, object recognition, segmentation, and in this case, image forgery detection.

Our proposed method relies on Conv2D layers, more specifically pretrained model layers for feature extraction in a ConvNet architecture. By learning important features

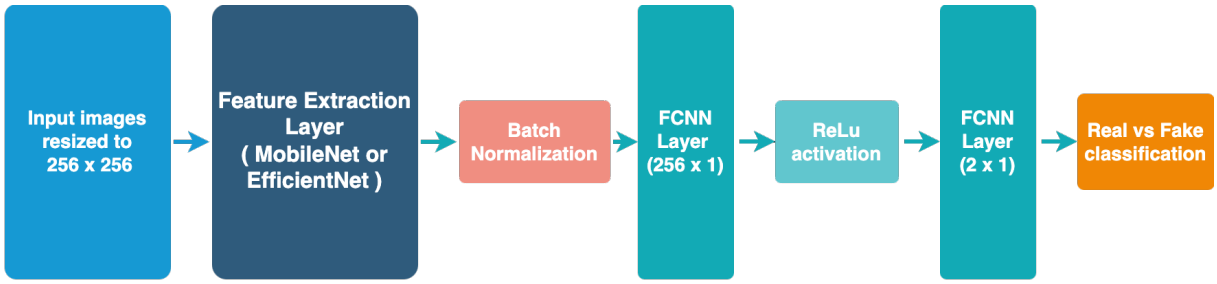

Figure 3.1: Proposed Model Architecture

from the input data, the proposed model can detect image forgeries regardless of image size and tampering type, highlighting the effectiveness of Conv2D layers in image processing tasks. We stuck to this particular model after realising the high effectiveness of the CNN architectures on binary classification of any king even though they were trained for multiclass classification. This makes sense given that with a deep model the kind of features that end up in the weights are a very diverse and effective set.

3.2 Proposed Model

3.2.1 Model Architecture

The proposed ConvNet architecture in Figure 2, is designed with simplicity in mind, aiming to efficiently detect image manipulation while minimizing model complexity by leveraging pre-trained models and modifying the outputs for good detection of manipulated images. Here's a breakdown of the key components:

1. **Feature Extractor :** We will use a pretrained CNN model without its final dense layer for feature extraction. The input data is fed to a pre-trained model of our choice with the parameters set for maximum efficiency in binary classification tasks, since our detection task is binary in nature (Fake and Real). The initial weights are imported from the model trained on ImageNet [21] dataset with the weights set to trainable. Batch normalization [36]is applied to the output from this model with parameters momentum set to 0.99 and epsilon (a small float variance to avoid division by zero during training) to 0.01.

2. **Dense layer 1:** The output of the feature extractor is thenfed to a dense layer of size 256 with L2 regularization for the kernels and L1 regularization for activity and bias. ReLu is used as the activation function for the output of this layer.

3. **Dense connected binary layer (Output Layer):** Dropout is applied to the output of previous layer with the parameter set to 0.4 (the fraction of values to be dropped during training). The final classification task is represented in 2 neurons in this layer using a softmax activation function

The model is compiled with Adamax as the optimizer and Categorical crossentropy is used for loss calculations.

Table 3.1: Pre-trained CNN models used

S.no	Model Type	Parameters
1	MobileNetV3-small	1.0 M
2	MobileNetV3-large	3.3 M
3	EfficientNetV2B0	6.3 M
4	EfficientNetV2B1	7.3 M
5	EfficientNetV2B2	9.1 M

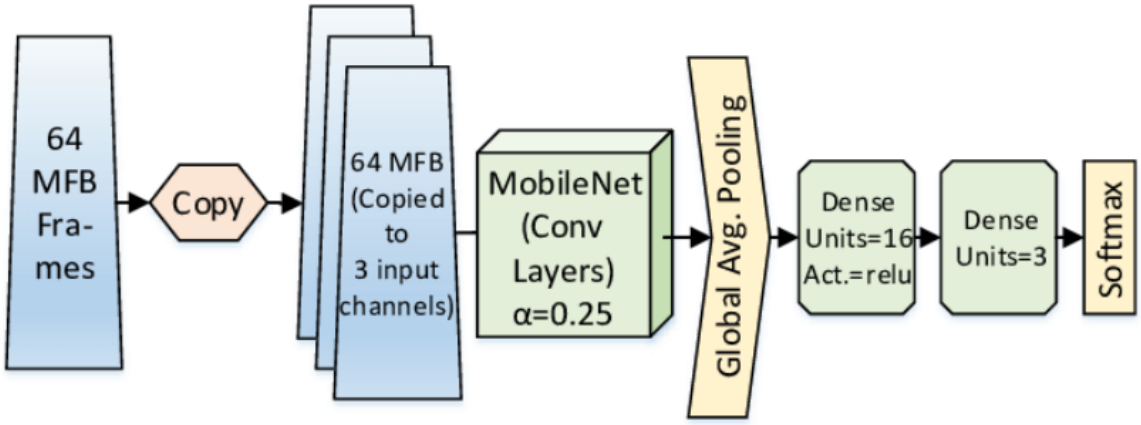

Figure 3.2: (The MobileNet architecture and [1].

3.2.2 Pre-Trained Models used

The initial goal of the model was to be highly effective even with the use of low number of parameters. For such a case we relied on relatively very small pre-trained ConvNet models called MobileNet [1]and EfficientNet [2]. Both these models have shown high performance metrics for general binary classification tasks.

MobileNet

MobileNet [1]is an efficient convolutional neural network architecture designed for mobile and embedded vision applications. It was developed by researchers at Google with the goal of building lightweight deep neural networks that can achieve near state-of-the-art accuracy while being computationally efficient and requiring fewer parameters.

MobileNet versions include MobileNet V1 which is the original MobileNet architecture introduced in 2017. MobileNet V2 [1] is an improved version released in 2018, which introduced linear bottleneck layers and shortcuts to improve performance. Finally, MobileNet V3 is the latest iteration released in 2019, with additional architectural improvements and two variants (MobileNetV3-Large and MobileNetV3-Small) for different resource constraints. For our proposed model we focus on the use of MobileNetV3-Large and Small models (3.3 and 1.0 million parameters respectively).

EfficientNet

EfficientNet [2] is a family of convolutional neural network models developed by researchers at Google AI. It was designed to achieve better accuracy and efficiency than

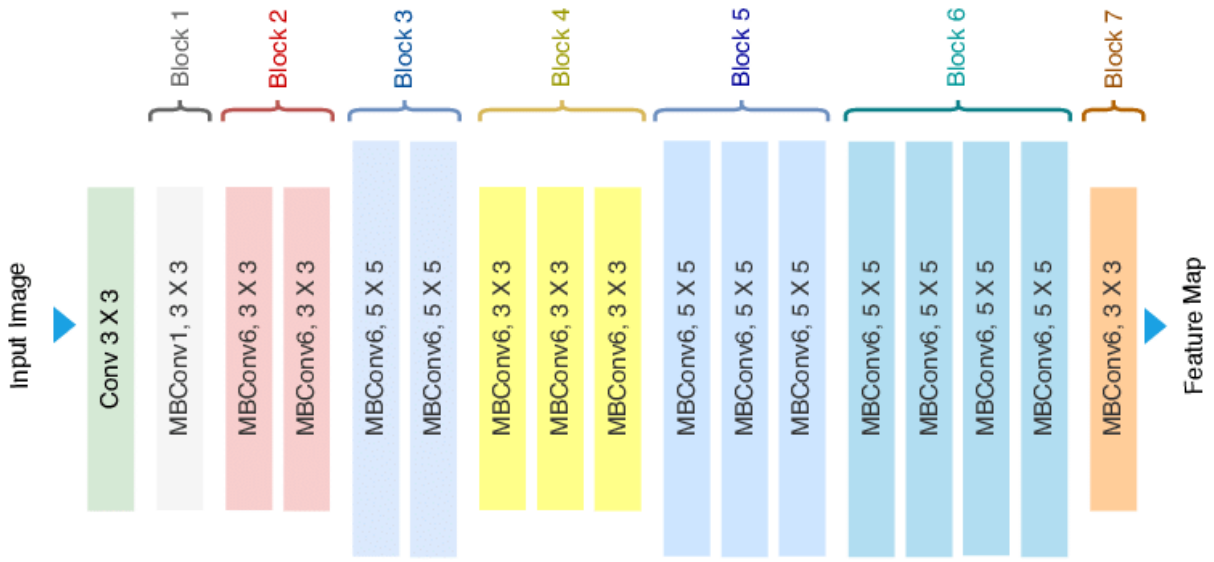

Figure 3.3: The EfficientNet architecture [2].

previous models like MobileNet, while also being scalable and capable of leveraging the benefits of model scaling.

There are several versions of EfficientNet models (EfficientNet-B0 to B7), with increasing model size and accuracy. The larger models achieve better accuracy but also have higher computational requirements. For our proposed model we focused on EfficientNet-B0 [37], B1 and B2 models with (6.3 , 7.3 , 9.1 million parameters respectively).

3.2.3 Optimizer

We have used Adamax (Adaptive Moment Estimation with Infinitely Huge Updates) as the optimizer which is is a variant of the Adam [38]optimization algorithm for stochastic gradient descent. It was proposed by researchers at Stanford University in 2015. Like Adam, Adamax is a combination of two other extensions of stochastic gradient descent: the momentum optimizer and the RMSProp optimizer [39]. The learning rate for each parameter is adjusted based on the average of recent gradients in the layers and the average of their recent squared gradients. In our testing Adamax had better performance metrics compared to the more conventional Adam optimizer.

The biased first moment estimate is updated as:

$$m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t \tag{3.1}$$

The biased second moment estimate is updated as:

$$u_t = \max(\beta_2 \cdot u_{t-1}, |g_t|) \tag{3.2}$$

The parameters (θ) are updated using these moment estimates:

$$\theta_t = \theta_{t-1} - \frac{\eta}{1 - \beta_1^t} \cdot \frac{m_t}{u_t + \epsilon} \tag{3.3}$$

3.2.4 Activation functions

We have used the Rectified Linear Unit (ReLU)[40] for every layer, except for the last classifier dense layer.Here's a summary of its key characteristics and advantages:

1. **Non-linearity:** ReLU introduces non-linearity to the network by "rectifying" aka modifying the values of inputs less than zero to exactly zero.

2. **Vanishing Gradient Problem:** ReLU helps to address the vanishing gradient problem plaguing larger neural net architectures, which can occur in deep networks during training phase. By allowing positive gradients for positive inputs and zero gradients for negative inputs, ReLU helps propagate gradients more effectively through the network, preventing them from vanishing or exploding.

3. **Faster Execution:**The biggest advantages of ReLU is its computational efficiency. Compared to other activation functions, ReLU involves only simple operations like comparison and selection. This results in faster execution and reduced computation time.

As mentioned, the mathematical representation of ReLU is

$$relu(\mathbf{z}) = max(0, \mathbf{z}). \tag{3.4}$$

Finally for the binary classification in the last layer we use sigmoid activation function .This activation function is a classic activation function used in neural networks which is non-linear in nature, and performs particularly well in the context of binary classification tasks. It reduces the input values to the range [0, 1], making it suitable for models like ours where the output needs to represent probabilities or binary decisions.

the mathematical representation of Sigmoid is

$$s(\mathbf{z}) = \frac{1}{1 + \exp(-\mathbf{z})}. \tag{3.5}$$

Overall, our proposed architecture is a minimal CNN model architecture suitable for detecting image manipulation efficiently. It employs convolutional layers for feature extraction, max-pooling layers for dimensionality reduction, dropout for regularization, and finally the two dense layers for a reduced set of features and classification. The architecture strikes a balance between simplicity and effectiveness in image forgery detection tasks.

3.3 Experiments

The model is evaluated on both the datasets and performance is analysed using standard metrics (Precision, Recall, F1-score).

3.3.1 Datasets

The first dataset, 140k RFF [41], is available on Kaggle as an open-source dataset. It consists of 140,000 images that are evenly divided between real and fake face images. 140k RFF is primarily used for image forgery detection tasks, specifically in distinguishing between real and fake faces. The second dataset Real and Fake face[42] detection was developed by the Computational Intelligence and Photography Lab in Yonsei University. The dataset contains photoshopped images divided into 960 fake and 1061 real images to train from.

Figure 3.4: Sample images from 140k RFF dataset

Figure 3.5: Sample images from real and fake faces datasett

3.3.2 Loss function and training settings

The Model is implemented using Python 3.11 with The help of multiple libraries such as Tensorflow[43], Keras [44], Pandas, Numpy etc. The Network is trained on 2 x T4 GPUs available on Kaggle. Batches of 32 images of size $128 \times 128 \times 3$ are used.

In the proposed work, ADAM emerges as the optimizer of choice for its efficiency and low memory requirements, making it well-suited for the task at hand. The binary classification problem, with its two distinct classes ('real' or 'fake'), lends itself naturally to the utilization of binary cross-entropy as the loss function. This loss function, also known as log loss, computes the negative average of the logarithm of corrected predicted probabilities, effectively capturing the model's performance in distinguishing between the two classes.

$$\mathcal{L}_{BCE}(y, \hat{y}) = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)] \tag{3.6}$$

The model is run for a certain number of epochs upto saturation of performance for every dataset, which is futher continued until we reach a saturation in terms of model accuracy. For the MobileNet models we run it for 15 and 20 epochs on 140k RFF and Readl Fake faces, while it is 10 and 15 for EfficientNet models for the respective datasets.

Chapter 4

RESULTS and DISCUSSION

Our model demonstrated exceptional performance on both 140k RFF and Real and fake faces dataset, achieving accuracies of nearly 99.5% and 77% respectively, either matching performance of other models on these datasets. This high level of accuracy underscores the robustness and reliability of our model across diverse datasets. These results validate the strength of this approach for practical applications. Furthermore, the solid performance across both datasets reinforces the generalizability of our model, showcasing ability to effectively learn and adapt to other data distributions.

As a final observation, strong CNN architectures can have very solid learning capabilities, even with a relatively small number of parameters for a complex task of deepfake detection.

Table 4.1: Performance metrics for 140k RFF(Macro Avg.)

Model Type	Precision	Recall	F1-score	Missclassification count
MNet-small	0.9819	0.9881	0.9850	790
MNet-large	0.9917	0.9988	0.9952	922
ENet-B1	0.9987	0.9987	0.9986	25
ENet-B2	0.9962	0.9950	0.9970	30

On the 140k RFF dataset, both the models perform extremely well quickly approaching saturation performance withing a few epochs. At the same time the dataset is pretty large as compared to the real fake face dataset. Therefore, the two architectures are ran for

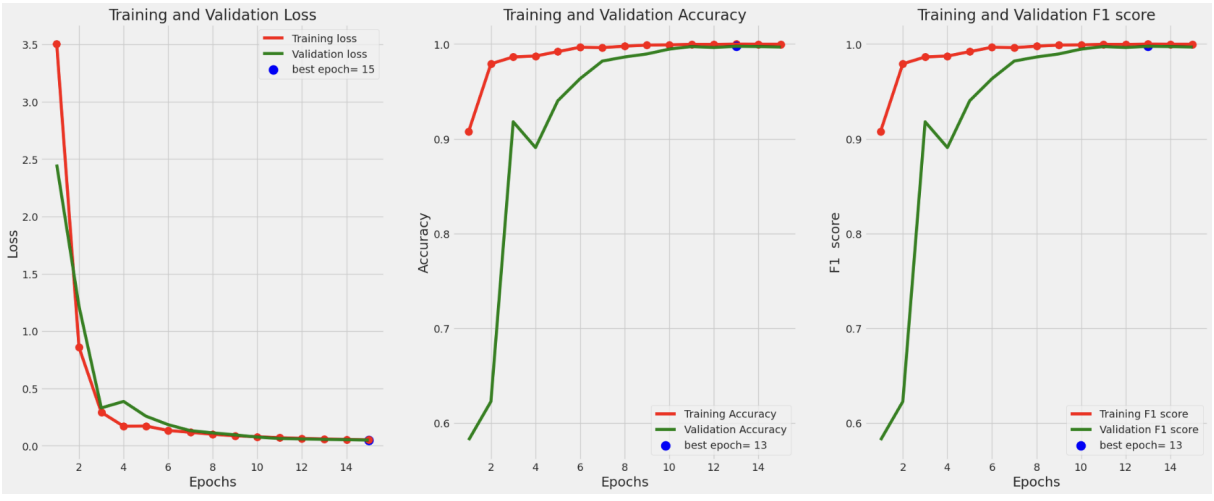

Figure 4.1: MobileNNetv3 performance on 140k RFF dataset

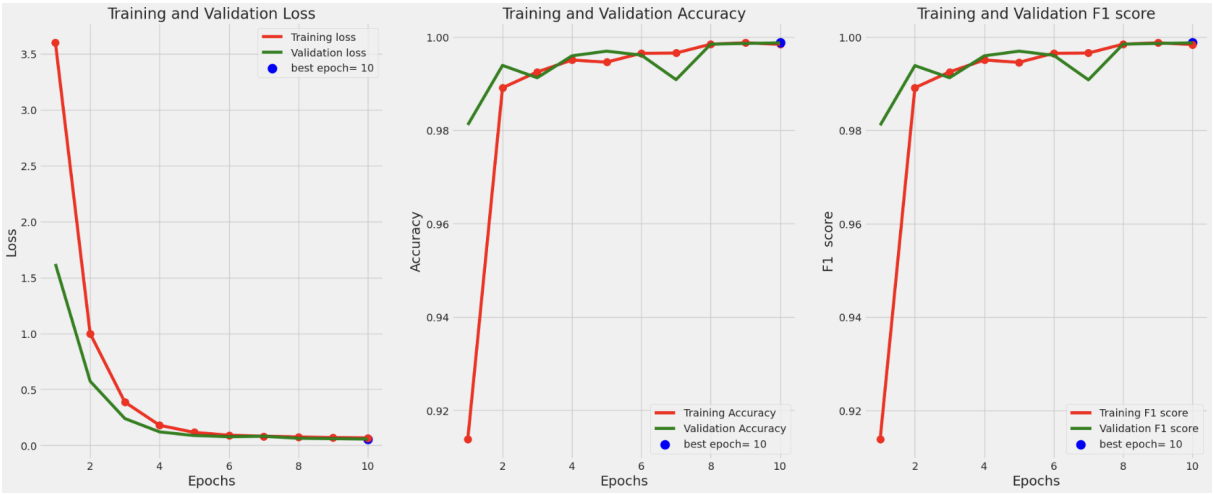

Figure 4.2: EfficientNet performance on 140k RFF dataset

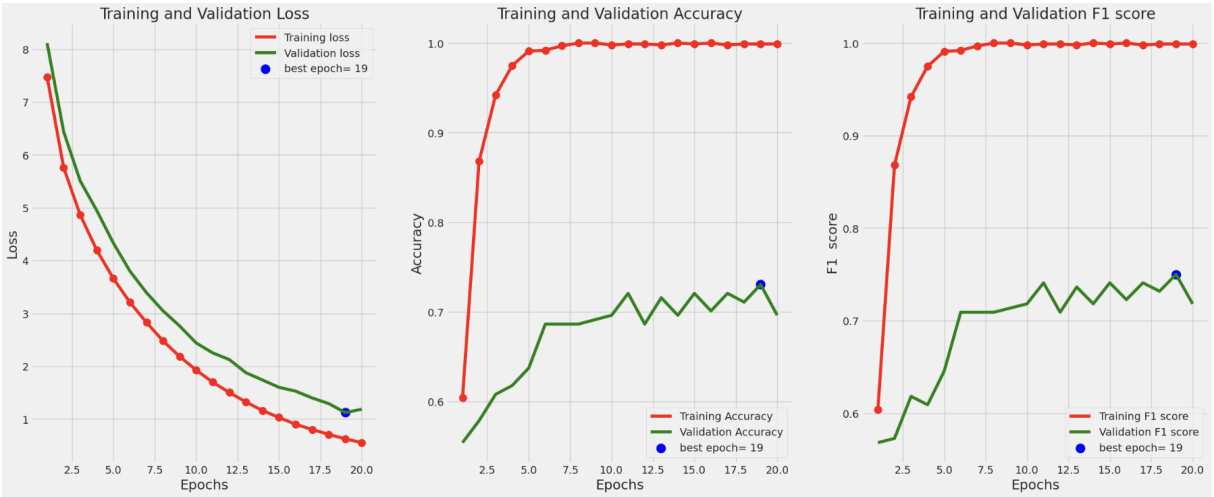

Figure 4.3: MobileNNetv3 performance on Real and Fake faces dataset

14 and 10 epochs respectively in order to fit in within our available runtime allocated by kaggle. The 10 epochs for EfficientNet compared to 14 for MobileNet is based on the fact that the previous is almost 8x bigger in parameters and hence more difficult to train.

On the other hand we are able to run both models for a few more epochs on the Real and Fake Faces dataset. Saturation is observed here too.

4.1 Performance Metrics

Performance metrics are crucial in evaluating the effectiveness of machine learning models and algorithms. They provide quantitative measures to assess how well a model performs on a given task. The performance metrics that we have relied for the evaluation of our model are precision, recall, support and accuracy.

- Precision is a measure of the accuracy of positive predictions made by the model. It calculates the proportion of true positive predictions out of all positive predictions made. A high precision value indicates that the model is making very few false positive predictions, which is desirable in applications where false positives are costly

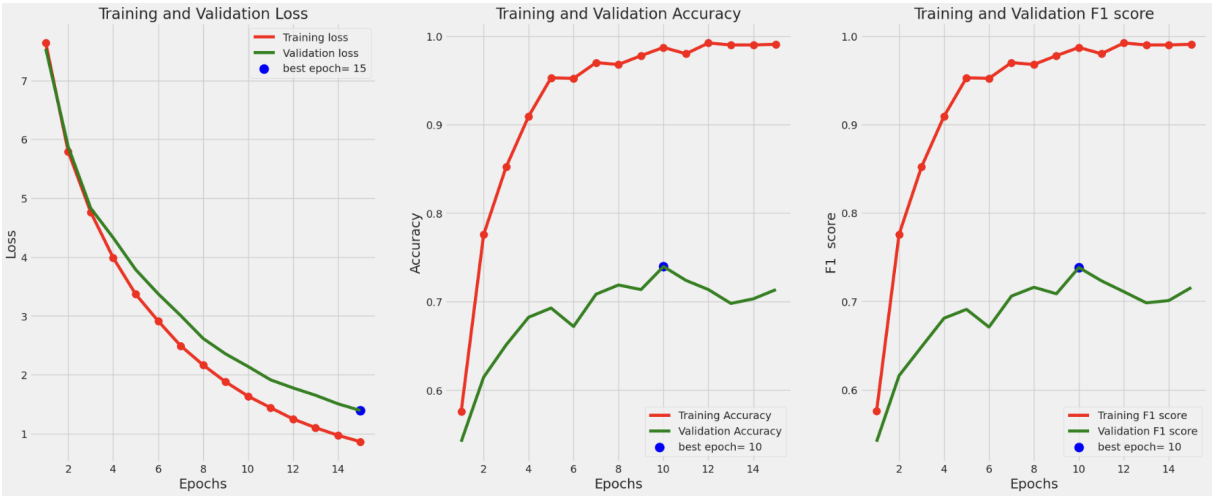

Figure 4.4: EfficientNet performance on Real and Fake Faces dataset

or unacceptable, such as fraud detection or medical diagnosis.

- Recall, on the other hand, measures the ability of the model to correctly identify positive instances. It calculates the proportion of true positive predictions out of all actual positive instances. A high recall value suggests that the model is effectively capturing most of the positive instances, which is important in applications where missing positive instances is undesirable, such as spam filtering or anomaly detection.

- Accuracy is a more general metric that measures the overall correctness of the model's predictions. It calculates the proportion of correct predictions (both true positives and true negatives) out of all predictions made. Accuracy is a useful metric when the classes are balanced, and the costs of false positives and false negatives are relatively equal.

- Additionally, the support metric simply provides us information about the number of instances of each class in the dataset. For a dataset, this can be helpful in understanding the class distribution and interpreting the other performance metrics in the especially when we consider the context of class imbalance or skewed data, a very common occurence in image based datasets.

For all the model and datasets the performance metric tables are given below

Table 4.2: Performance metrics for 140k RFF(Macro Avg.)

Model Type	Precision	Recall	F1-score	Missclassification count
MNet-small	0.9819	0.9881	0.9850	790
MNet-large	0.9917	0.9988	0.9952	922
ENet-B1	0.9987	0.9987	0.9986	25
ENet-B2	0.9962	0.9950	0.9970	30

Table 4.3: Performance metrics with MobileNet architure on 140k RFF

	precision	recall	f1-score	support
fake	0.9951	0.9989	0.9970	10000
real	0.9989	0.9951	0.9970	10000
accuracy			0.9970	20000
macro avg	0.9970	0.9970	0.9970	20000
weighted avg	0.9970	0.9970	0.9970	20000

Table 4.4: Performance metrics with MobileNet architure on Real and Fake Faces

	precision	recall	f1-score	support
fake	0.7000	0.8021	0.7476	96
real	0.8000	0.6972	0.7451	109
accuracy			0.7463	205
macro avg	0.7500	0.7497	0.7463	205
weighted avg	0.7532	0.7463	0.7463	205

Table 4.5: Performance metrics with EfficientNet architure on 140k RFF

	precision	recall	f1-score	support
fake	0.9983	0.9990	0.9987	10000
real	0.9990	0.9983	0.9986	10000
accuracy			0.9987	20000
macro avg	0.9987	0.9987	0.9986	20000
weighted avg	0.9987	0.9987	0.9986	20000

Table 4.6: Performance metrics with EfficientNet architure on Real and Fake Faces

	precision	recall	f1-score	support
fake	0.7500	0.8021	0.7426	93
real	0.7950	0.6972	0.7251	105
accuracy			0.7465	202
macro avg	0.7500	0.7497	0.7463	197
weighted avg	0.7532	0.7463	0.7463	197

Chapter 5

CONCLUSION AND FUTURE SCOPE

Deep learning models have emerged as powerful tools across various domains, demonstrating remarkable capabilities in tasks such as image recognition, natural language processing, and decision-making. However, as the complexity and specificity of these models increase, so does their computational cost. This chapter explores the existing research gap and problem statement surrounding the challenges posed by the expense and specificity of deep learning models.

5.1 Conclusion

Through our work we have highlighted the critical research gap in addressing the computational cost and specificity challenges associated with deep learning models. The problem statement emphasizes the need for innovative solutions that make deep learning more cost-effective, accessible, and adaptable to diverse tasks and datasets. As we found through our research and efforts in building a good deepfake detector deeper we have identified some key challenges in doing so.

The CNN model presented in this study demonstrates great performance with a significantly lower number of parameters compared to larger ones. This size and simplicity make it an attractive choice for resource-constrained environments, it is still obvious that there exists a performance gap between our model and much larger counterparts. Although this performance gap if bloated by the fact that models have gotten exponentially bigger.

Despite its limitations, our model's ability to achieve competitive results highlights the potential for further optimization and refinement. Future research could focus on enhancing the model's capacity to extract more intricate features to bridge the performance divide between compact models and their larger counterparts.

In summary, while our CNN model may lag behind bigger models in certain metrics, its efficiency and effectiveness in scenarios where computational resources are limited underscore its practical utility. Continued exploration and innovation in model design and optimization are essential for pushing the boundaries of performance while maintaining computational efficiency.

5.2 Research Gap

Despite the significant advancements in deep learning, a notable research gap exists in addressing the practical limitations associated with the high computational cost and speci-

ficity of these models. As deep learning architectures become increasingly complex and tailored to specific tasks, several critical challenges arise:

Computational Expense: Deep learning models, particularly large neural networks, demand large computational resources (both in terms of memory and energy) for training and inference. High computational costs hinder the accessibility and scalability of deep learning solutions, especially for researchers and organizations with limited resources.

Specificity and Overfitting: Specialized deep learning models tend to excel in specific tasks but may suffer from overfitting issues when applied to diverse datasets or broader domains. The lack of generalization in overly specific models limits their adaptability and robustness in real-world scenarios. The research problem addressed in this chapter revolves around finding solutions to mitigate the challenges associated with the cost and specificity of deep learning models.

- **Cost-Effective Deep Learning**: There is obviously a need to develop techniques to reduce the computational expenses associated with training . Deploying deep learning models without compromising performance is clearly a need especially for edge computing devices like mobile phones. There is also potential in finding methods to optimize existing architectures or devise novel algorithms that achieve comparable accuracy with reduced resource requirements.

- **Enhancing Generalization**: In the context of task-specific models,there is need enhance their ability to generalize across diverse datasets or tasks. Like the model presented here, we can find if there are transfer learning or domain adaptation strategies that can make specific deep learning models more versatile without losing out on performance, mostly due to exponentially increasing sizes of models. As presented in the literature review there almost no model which can generalize audio, video and semantic deepfakes together.

- **Scalability and Accessibility**: Deep learning needs to be made more accessible to a wider audience, including researchers and practitioners with limited computational resources. Models are quickly approaching billions of parameters in size, and as generators scale up so does a need for similarly performing detectors. Meanwhile we need to find approaches to scale deep learning solutions for deployment on edge devices or in resource-constrained environments.

- **Balancing Specificity and Generality**:Its a good goal to find methodologies help strike a solid balance between task-specificity and generalization, ensuring that deep learning models remain adaptable across various scenarios. We also need to develop frameworks that allow for dynamic adjustment of model specificity based on real-time demands or evolving datasets.

Bibliography

[1] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," 2017.

[2] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," 2020.

[3] M. Masood, M. Nawaz, K. M. Malik, A. Javed, and A. Irtaza, "Deepfakes generation and detection: State-of-the-art, open challenges, countermeasures, and way forward," 2021.

[4] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1–9.

[5] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," 2014.

[6] A. Sherstinsky, "Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network," *Physica D: Nonlinear Phenomena*, vol. 404, p. 132306, Mar. 2020. [Online]. Available: http://dx.doi.org/10.1016/j.physd.2019.132306

[7] Z. Cai, K. Stefanov, A. Dhall, and M. Hayat, "Do you really mean that? content driven audio-visual deepfake dataset and multimodal method for temporal forgery localization," 2023.

[8] J. Botha and H. Pieterse, "Fake news and deepfakes: A dangerous threat for 21st century information security," 03 2020.

[9] B. Mahmud and A. Sharmin, "Deep insights of deepfake technology : A review," 01 2020.

[10] M. S. Rana, M. N. Nobi, B. Murali, and A. H. Sung, "Deepfake detection: A systematic literature review," *IEEE Access*, vol. 10, pp. 25 494–25 513, 2022.

[11] Z. Xia, T. Qiao, M. Xu, X. Wu, L. Han, and Y. Chen, "Deepfake video detection based on mesonet with preprocessing module," *Symmetry*, vol. 14, no. 5, 2022. [Online]. Available: https://www.mdpi.com/2073-8994/14/5/939

[12] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: a technical overview," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp. 21–36, 2003.

[13] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," 2023.

[14] H. H. Nguyen, J. Yamagishi, and I. Echizen, "Use of a capsule network to detect fake images and videos," 2019.

[15] I. Amerini, L. Galteri, R. Caldelli, and A. Bimbo, "Deepfake video detection through optical flow based cnn," *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pp. 1205–1207, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:207900737

[16] D. Bank, N. Koenigstein, and R. Giryes, "Autoencoders," 2021.

[17] G. R. Koch, "Siamese neural networks for one-shot image recognition," 2015. [Online]. Available: https://api.semanticscholar.org/CorpusID:13874643

[18] N. Bonettini, E. D. Cannas, S. Mandelli, L. Bondi, P. Bestagini, and S. Tubaro, "Video face manipulation detection through ensemble of cnns," *CoRR*, vol. abs/2004.07676, 2020. [Online]. Available: https://arxiv.org/abs/2004.07676

[19] A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, "Faceforensics++: Learning to detect manipulated facial images," *CoRR*, vol. abs/1901.08971, 2019. [Online]. Available: http://arxiv.org/abs/1901.08971

[20] S. Wang, O. Wang, A. Owens, R. Zhang, and A. A. Efros, "Detecting photoshopped faces by scripting photoshop," *CoRR*, vol. abs/1906.05856, 2019. [Online]. Available: http://arxiv.org/abs/1906.05856

[21] K. Yang, K. Qinami, L. Fei-Fei, J. Deng, and O. Russakovsky, "Towards fairer datasets: filtering and balancing the distribution of the people subtree in the imagenet hierarchy," in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. ACM, Jan. 2020. [Online]. Available: http://dx.doi.org/10.1145/3351095.3375709

[22] X. Chen, C. Dong, J. Ji, J. Cao, and X. Li, "Image manipulation detection by multi-view multi-scale supervision," *CoRR*, vol. abs/2104.06832, 2021. [Online]. Available: https://arxiv.org/abs/2104.06832

[23] V. Asnani, X. Yin, T. Hassner, S. Liu, and X. Liu, "Proactive image manipulation detection," 2022.

[24] M.-J. Kwon, S.-H. Nam, I.-J. Yu, H.-K. Lee, and C. Kim, "Learning jpeg compression artifacts for image manipulation detection and localization," *International Journal of Computer Vision*, vol. 130, no. 8, p. 1875–1895, May 2022. [Online]. Available: http://dx.doi.org/10.1007/s11263-022-01617-5

[25] Z. Cai, S. Ghosh, K. Stefanov, A. Dhall, J. Cai, H. Rezatofighi, R. Haffari, and M. Hayat, "Marlin: Masked autoencoder for facial video representation learning," 2023.

[26] D. A. Coccomini, N. Messina, C. Gennaro, and F. Falchi, *Combining EfficientNet and Vision Transformers for Video Deepfake Detection*. Springer International Publishing, 2022, p. 219–229. [Online]. Available: http://dx.doi.org/10.1007/978-3-031-06433-319

[27] S. A. Shahzad, A. Hashmi, S. Khan, Y.-T. Peng, Y. Tsao, and H.-M. Wang, "Lip sync matters: A novel multimodal forgery detector," in *2022 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 2022, pp. 1885–1892.

[28] D. Wodajo and S. Atnafu, "Deepfake video detection using convolutional vision transformer," 2021.

[29] Y. Zang, Y. Zhang, M. Heydari, and Z. Duan, "Singfake: Singing voice deepfake detection," 2024.

[30] L. Li, J. Bao, T. Zhang, H. Yang, D. Chen, F. Wen, and B. Guo, "Face x-ray for more general face forgery detection," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 5000–5009.

[31] S. Saha, R. Perera, S. Seneviratne, T. Malepathirana, S. Rasnayaka, D. Geethika, T. Sim, and S. Halgamuge, "Undercover deepfakes: Detecting fake segments in videos," 2023.

[32] O. de Lima, S. Franklin, S. Basu, B. Karwoski, and A. George, "Deepfake detection using spatiotemporal convolutional networks," 2020.

[33] N. Yu, V. Skripniuk, S. Abdelnabi, and M. Fritz, "Artificial fingerprinting for generative models: Rooting deepfake attribution in training data," 2022.

[34] S. Tariq, S. Lee, and S. S. Woo, "A convolutional lstm based residual network for deepfake video detection," 2020.

[35] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," 2015.

[36] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015.

[37] M. Tan and Q. V. Le, "Efficientnetv2: Smaller models and faster training," 2021.

[38] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2017.

[39] H. Shaziya, "A study of the optimization algorithms in deep learning," 03 2020.

[40] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ser. ICML'10. Madison, WI, USA: Omnipress, 2010, p. 807–814.

[41] "140k rff," https://www.kaggle.com/datasets/xhlulu/140k-real-and-fake-faces, 2015.

[42] Y. U. CIP lab, "Real and fake faces," https://www.kaggle.com/datasets/ciplab/real-and-fake-face-detection, 2019.

[43] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens,

B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015, software available from tensorflow.org. [Online]. Available: https://www.tensorflow.org/

[44] F. Chollet, "keras," https://github.com/fchollet/keras, 2015.

Proof of Journal and Conferences attended:

1. Presented paper titled “Micro-Parameter MiniNet: A Low-Cost Approach for Image forgery and Deepfake Detection” in the 3rd International Conference on Recent Trends in Engineering, Technology and Business Management (ICRTETBM-2024) (Theme: Digitization Transformation and Business Operations) held during February 21-23, 2024 at the Amity International Business School, Amity University, Noida, Uttar Pradesh, India.

AMITY UNIVERSITY

Springer

3rd INTERNATIONAL CONFERENCE ON
RECENT TRENDS IN ENGINEERING, TECHNOLOGY AND BUSINESS MANAGEMENT (ICRTETBM 2024)
(Digitization Transformation and Business Operations)

Organized By Amity International Business School
Amity Centre for Interdisciplinary Research

24th INTERNATIONAL BUSINESS HORIZON
INBUSH ERA WORLD SUMMIT 2024

21st to 23rd February 2024

CERTIFICATE OF PARTICIPATION

This is to certify that Anukriti Kaushal, Arpit Mahlawat
from Department of CSE, Delhi Technological University, New Delhi, India has participated in 3rd International Conference on Recent Trends in Engineering, Technology and Business Management (ICRTETBM 2024) held from February 21st to 23rd February, 2024 at Amity University, Noida, India.

He/She also chaired a session / delivered a keynote / invited talk / presented a paper
Micro-Parameter MiniNet: A Low-Cost Approach for Image forgery and Deepfake Detection

Prof. (Dr.) Gurinder Singh
Group Vice Chancellor
General Chair, ICRTETBM 2024

Prof. (Dr.) P.K. Kapur
Director, ACIDR
Conference Chair, ICRTETBM 2024

Dr. Atul Chauhan
President, Ritnand Balved Education Foundation
& Chancellor, Amity University

2. Accepted paper titled “ An analysis of Augmented ConvNets on Image manipulation detection ” for presentation and publication in International Conference on Intelligent Computing and Communication Techniques (ICICCT-2024) at JNU New Delhi, India. (Paper Id 699).

Gmail Arpit Mahlawat <mahla.arpit@gmail.com>

Notification of acceptance of paper id 699

2 messages

Microsoft CMT <email@msr-cmt.org> Thu, May 16, 2024 at 4:29 PM
Reply-To: ICICCT 2024 <icicctcon@gmail.com>
To: Arpit Mahlawat <mahla.arpit@gmail.com>

Dear Dr./ Prof. Arpit Mahlawat,

Congratulations...

Your paper / article paper id 699: An analysis of Augmented ConvNets on Image manipulation detection has been accepted for publication in International Conference on Intelligent Computing and Communication Techniques at JNU New Delhi, India.

Kindly save your paper by given paper id only (eg. 346.docx, 346.pdf, 346_copyright.pdf)

Registration Link:

https://forms.gle/mSsHa8GMLtMkWuaq8

Please ensure the following before registration and uploading camera ready paper.
1. Paper must be in Taylor and Frances Format.
Template and copyright with author instruction are given in below link: https://icicct.in/author_inst.html

2. Minimum 12 references should be cited in the paper and all references must be cited in the body. Please follow the template.
3. The typographical and grammatical errors must be carefully looked at your end.
4. Complete the copyright form (available at template folder).
5. The regular fee (Available in registration section) will be charged up to 6 pages and after that additional Rs.1000 for Indian authors / 10 USD for foreign authors per additional page will be charged.
6. Reduce the Plagiarism below 10% excluding references and AI Plagiarism 0%. The Authors are solely responsible for any exclusion of publication if any.
7. Certificate will be issued by the name of registered author (Single author only).
8. Certificates may be issued to all other authors on the extra payment of 1000/- INR per author.
9. Last Date of registration and uploading copyright and camera-ready copy: 31/05/2024.
10. Make a single payment which includes registration fee + Extra certificates fee + Extra page fees.
11. Permissions: Kindly make sure the permissions for each copyrighted artwork file have been cleared ahead of the submission, with the details listed in the Permission Verification form (attached). All permission grants must be submitted along with your final manuscript.
12. Each Illustration must include a caption and an alternative text description to assist print impaired readers (‘Alt Text’).
(Alt Text is mandatory for each Illustrations)
Figures: Please make sure no figures are missing, and all figures are high resolution and alt text is included
Tables: Please ensure that there are no missing tables, and the tables in your manuscript are not pasted as figures.
Citation: Kindly ensure there are no missing citations in your manuscript

Registration Link: https://forms.gle/mSsHa8GMLtMkWuaq8

Registration Fee to be deposited in below account

Bank Account Details :

Indian Account Details:

Account Holder Name: EVEDANT Foundation

Account Number: 0674002190422900

IFSC Code: PUNB0067400
SWIFT Code: PUNBINBBGNM

Branch: Punjab National Bank, Navyug Market, Ghaziabad

Account Type: Current Account

International Conference on Intelligent Computing and Communication Techniques
28 & 29 June 2024.

Thanks and Regards
Convener
International Conference on Intelligent Computing and Communication Techniques
contact details : icicctcon@gmail.com, https://icicct.in/index.html

---

Arpit Mahlawat <mahla.arpit@gmail.com> Thu, May 16, 2024 at 4:32 PM
To: anukritikaushal@dtu.ac.in

[Quoted text hidden]

ANNEXURE-IV

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-42

PLAGIARISM VERIFICATION

Title of the Thesis: An analysis of ConvNets on Deepfake Detection

Total Pages: 26 Name of the Scholar: Arpit Mahlawat Supervisor (s)

(1) Ms. Anukriti Kaushal

(2)

(3)

Department: Computer Science and Engineering

This is to report that the above thesis was scanned for similarity detection. Process and outcome is given below:

Software used: Turnitin Similarity Index: 8 %, Total Word Count: 7705

Date: 30 / 05 / 24

Candidate's Signature **Signature of Supervisor(s)**

PAPER NAME

Delhi_Technological_University_Thesis_10-35.pdf

WORD COUNT	CHARACTER COUNT
7705 Words	**42268 Characters**
PAGE COUNT	FILE SIZE
26 Pages	**4.0MB**
SUBMISSION DATE	REPORT DATE
May 30, 2024 9:21 AM GMT+5:30	**May 30, 2024 9:22 AM GMT+5:30**

8% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 6% Internet database
- 5% Publications database
- Crossref database
- Crossref Posted Content database
- 5% Submitted Works database

Excluded from Similarity Report

- Bibliographic material
- Quoted material
- Cited material
- Small Matches (Less then 10 words)

8% Overall Similarity

Top sources found in the following databases:

- 6% Internet database
- Crossref database
- 5% Submitted Works database
- 5% Publications database
- Crossref Posted Content database

TOP SOURCES

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Jacobs University, Bremen on 2020-12-20 Submitted works	1%
2	arxiv.org Internet	1%
3	Shobhit Tyagi, Divakar Yadav. "MiniNet: a concise CNN for image forge... Crossref	<1%
4	Liverpool John Moores University on 2023-06-04 Submitted works	<1%
5	export.arxiv.org Internet	<1%
6	Kingston University on 2023-01-09 Submitted works	<1%
7	researchgate.net Internet	<1%
8	Higher Education Commission Pakistan on 2023-03-30 Submitted works	<1%

9	Mansi Rehaan, Nirmal Kaur, Staffy Kingra. "Face manipulated deepfake... Crossref	<1%
10	Md Shohel Rana, Mohammad Nur Nobi, Beddhu Murali, Andrew H. Sun... Crossref	<1%
11	codefinity.com Internet	<1%
12	section.iaesonline.com Internet	<1%
13	repository.northsouth.edu Internet	<1%
14	Singh, Shekhar. "Facial Expression Recognition Using Convolutional Ne... Publication	<1%
15	University of Warwick on 2023-07-10 Submitted works	<1%
16	Ahmed, Jishan. "Cost-Aware Machine Learning and Deep Learning for ... Publication	<1%
17	University of Greenwich on 2016-02-29 Submitted works	<1%
18	mdpi-res.com Internet	<1%
19	123dok.net Internet	<1%
20	Mansoura University on 2023-11-15 Submitted works	<1%

21	University of Nottingham on 2023-05-19 Submitted works	<1%
22	scholarworks.uaeu.ac.ae Internet	<1%
23	frontiersin.org Internet	<1%
24	mdpi.com Internet	<1%