Enhancing Video Forensics: Deep Learning Approaches to Combat Advanced Video Manipulation Techniques

Suganthe Ravichandaran, Hema Moorthy, Iniya Kavitha Sadasivam, Keerthika Muthusamy

Department of Computer Science and Engineering, Kongu Engineering College, Erode-638 060, Tamil Nadu, India
suganthe_rc@kongu.ac.in; hemam.20cse@kongu.edu; iniyaks.20cse@kongu.edu; keerthikam.20cse@kongu.edu

Abstract: This research presents a method to efficiently detect facial tampering in videos, and particularly focuses on two recent techniques, used to generate hyper realistic forged videos: Deepfake and Face2Face. Traditional image forensics techniques are usually not well suited to videos due to the compression that strongly degrades the data. Thus, this work follows a deep learning approach and presents a network, with a smaller number of layers to focus on the mesoscopic properties of images. This work evaluates those fast networks on both an existing dataset and a dataset generated from online videos. Deepfake image detection is important because it helps everyone determine if the pictures seen online are real or fake. Due to advancements in computer vision techniques, people can create fake images that look extremely realistic. These could be used to spread lies or invade someone's privacy. The detection tools use smart technology to spot these fakes, ensuring that everyone can trust the pictures they come across and preventing the spread of misleading or harmful content on the internet. This work contributes to the growing body of research addressing the challenges posed by advanced video manipulation techniques, providing a valuable tool for applications in cybersecurity, media integrity, and the prevention of misinformation in an era dominated by sophisticated visual content manipulation.

Keywords: Video Forensics; Deepfake image; Meso4; face tampering; fake image detection; cyber security

1 Introduction

Deepfake videos have had a significant impact on many aspects of society in recent years, raising concerns about the spread of false information and the decline in public confidence in digital media. These extremely realistic modifications,

enabled by advanced AI algorithms, have the potential to mislead viewers by creating fake situations or changing the words and behaviors of people portrayed in the films. As such, there is a growing need to develop efficient methods for detecting and reducing the number of fake videos that are out there. If this urgent problem is not resolved, there could be serious consequences, such as the spread of false information widely, damage to people's and organizations reputations, and even social unrest.

An innovative approach that centered on model training with image datasets taken from video frames. The novel model uses deep learning techniques to analyze complex patterns and variations in visual content, allowing it to distinguish between real and fake videos. After extensive training on various sets of video frames, the model learns to identify common features and artifacts related to Deepfake manipulation. A proactive strategy like this for detecting Deepfake gives some hope for preserving the trustworthiness of digital media and preventing the spread of misleading content on different websites. Today, the danger of fake news is widely acknowledged and, in a context, where more than 100 million hours of video content are watched daily on social networks, the spread of falsified video raises more and more concerns. While significant improvements have been made for image forgery detection, digital video falsification detection still remains a difficult task. These initiatives are crucial in preventing the spread of false information and fostering confidence in online content, in addition to maintaining the reliability of digital media.

Deepfake detection in videos is more challenging than in images for several reasons, which include both the temporal and spatial complexities involved in video manipulation. Videos consist of a sequence of frames, meaning that deepfake detection must account for the temporal coherence between frames. Unlike images, which are static and independent, videos involve dynamic content with consistent motion across frames. Creating a convincing deepfake video requires modifying not only the facial features or expressions of individuals but also ensuring that the changes are consistent throughout the entire video sequence. Achieving such consistency over time increases the difficulty of detection, as subtle distortions might not be immediately visible in individual frames, but can become obvious when analyzed over a sequence of frames. In videos, both the visual and audio components must be analyzed to detect deepfakes. While deepfake images typically focus on altering the facial features of individuals, deepfake videos can also involve the manipulation of speech and lip-syncing. The audio may not match the facial movements, leading to subtle inconsistencies that are more challenging to spot.

2 Related Work

Rossler et al. [1] proposed an automated pipeline to detect fake faces from images. The authors selected four Deep Fake datasets, including Deep Fakes, Face2Face, Face Swap, and Neural Textures, along with a pristine dataset to evaluate precision. Dolhansky et al. [2] implemented light weight CNN model called TamperNet trained with Deep Fake Detection Challenge (DFDC) dataset which is used to detect only acute level manipulations on images such as cut-and-pasted objects. Nguyen et al. [3] adopted the idea of capsule architecture (CapsNets) and extended their work to detect different kinds of forgery in images and videos. This network generally requires more computational resources and training time due to their architecture. Chugh et al. [4] proposed an approach based on the modality dissonance score (MDS), which classifies forgery in Deep Fake video between audio and visual modalities through their dissimilarities. The contrastive loss was used to analyze the closeness features between audio and video.

Korshunov et al. [5] also evaluated baseline face-swap detection algorithms and found that the lip-sync-based approach failed to detect mismatches between lip movements and speech. Agarwal and Varshney [6], designed a statistical model based on hypothesis testing to detect face-swapped content or fraudulence in images. In this study, the authors focused on the theoretical boundaries and robustness of Deepfake detection systems, providing an in-depth statistical understanding of the limitations of these systems in challenging scenarios.

Kumar et al. [7] implemented several DL approaches and compared their results with the context of Deep Fake classification using metric learning. The authors used a Multitask Cascaded Convolutional Neural Network (MTCNN) to extract faces from images or videos. Rahul et al. [8] established a technique based on the common attributes of fabricated video clips that analyzed face interpretation and the manipulated videos are converted into frames and fed to the MTCNN to extract the facial features using the MobileNet model. Guera and Delp [9] proposed a two-stage analysis composed of a Convolutional Neural Network to extract the frame-level features. In the first stage, Inception-V3 with a fully connected layer at the top of the network was adopted. In the second stage, LSTM is applied to compute the intra-frame and temporal inconsistencies between frames. Li et al. [10] used Celeb-DF dataset and its overall performance is lowest across all datasets, with very less average AUC. Hence, the celeb-DF dataset still needs improvement. Yang et al. [11] proposed Avoid-df that focuses on addressing the challenge of detecting Deepfake videos by utilizing a novel approach that integrates audio and visual information.

Yan et al. [12] proposed a method called Uncovering Common Features (UCF), which focuses on identifying common visual patterns and features across different Deep Fake manipulation techniques, leading to more robust and generalizable detection models. Mcuba et al. [13] and Chen et al. [14] explored the impact of deep learning methods on the detection of Deep Fake audio, specifically in the

context of digital investigations. Li et al. [15] focused on the advancement of facial landmark detection algorithms with generative models such as GANs and VAEs, which has contributed to more convincing and realistic Deep Fakes.

Cozzolino et al. [16] demonstrated the application of Deep Fake in forensic investigation with machine-learning-based classification models. The temporal aggregation of convolutional representations and deep learning techniques were also mentioned as having demonstrated promising results in the detection of Deep Fakes. Li et al. [17] identified challenges in detecting high-quality Deep Fakes using synthesized videos and audio clips. They utilized deep neural networks for face detection and facial landmark extraction, emphasizing limitations in current Deep Fake generation methods that cannot produce a fine mapping of color shades for hair with respect to the human face. Ding, et al. [18] proposed a deep transfer learning model to detect swapped faces. Rashid et al. [19] illustrated the deep fake history and different deep fake technologies used to detect deep fake images.

Huang et al. [20] approached face swapping detection by focusing on face identity. Face swapping involves substituting the target face with another face to create a fake image that is indistinguishable from a real one.

Korshunov et al. [21] examined fundamental face-swap detection algorithms, revealing that lip-sync-based methods often overlook discrepancies between lip movements and speech. They further demonstrated that using image quality metrics alongside a Support Vector Machine (SVM) classifier effectively identifies high-quality Deep Fake videos, achieving an 8.97% equal error rate.

Korshunov et al. [21] focuses on face spoofing detection (presentation attacks like printed photos or replayed videos) using image quality assessment (IQA) techniques. It is about how to detect if someone is trying to fool a face recognition system with a fake face.

Agarwal and Varshney [22] designed a practical framework for hypothesis testing applied specifically to facial images and Deepfake detection. It focuses on real-world applications of detecting GAN-generated facial images using statistical tests to analyze facial features, and it includes empirical performance metrics to validate the proposed methods. In this study, the authors considered a mathematical bound value corresponding to the error probability based on the detection of genuine or GAN-generated images.

3 Proposed Methodology

This work aims to operate at a mesoscopic level of analysis, to detect forged facial videos. However, microscopic analyses based on image noise are unsuitable in compressed video contexts, where noise is significantly degraded. At a higher semantic level, distinguishing forged images—especially those depicting human

face—poses challenges for the human eye. Therefore, an intermediate approach utilizing a deep neural network with a limited number of layers is advocated. This approach leverages the Meso4 architecture, which has demonstrated superior classification scores in tests. Meso4 features a low level of representation and surprisingly few parameters. It is built upon a well-performing network for image classification, employing alternating layers of convolutions and pooling for feature extraction, along with a dense network for classification.

3.1 Dataset Description

The dataset consists of an array of images sourced from videos suspected to contain Deepfake manipulations, as well as authentic content obtained from various online sources. Extracting frames from these videos yield an extensive dataset of images, forming the basis for the analysis. This inclusion ensures a comprehensive representation of real-world scenarios, allowing for a thorough examination of various Deepfake methods. Deepfake refers to a technique that utilizes artificial intelligence (AI) to create synthetic media, typically involving the manipulation of visual or audio content to portray individuals saying or doing things they never did. Face2Face, on the other hand, is a specific Deepfake technique focused on facial manipulation. It involves mapping the facial expressions of a target individual onto another person's face in a video, creating a realistic but falsified portrayal of the target individual's actions. Both Deepfake and Face2Face techniques aim to deceive viewers by generating hyper realistic fabricated content, posing significant challenges for media authentication and trust worthiness. Both the real and fake datasets consist of a total of 961 images. Sample images from these datasets are shown in Figure 1 and 2, respectively.

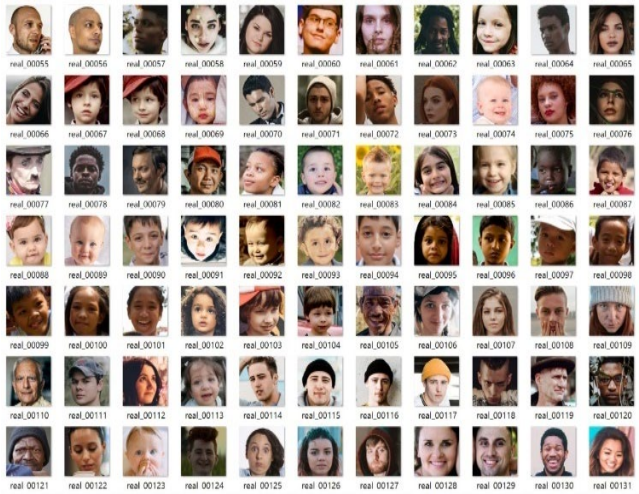

Figure 1
Real Images

Figure 2
Deep Fake Images

3.2 Design and Implementation

The Meso4 architecture is a deep learning model specifically designed for the task of detecting Deepfake manipulations within images or video frames. It comprises four layers of successive convolutions and pooling operations, followed by a dense network with one hidden layer. To improve generalization, the convolutional layers use ReLU activation functions that introduce non-linearities and Batch Normalization to regularize their output and prevent the vanishing gradient effect, and the fully connected layers use Dropout to regularize and improve their robustness. This structure enables the model to extract features from the input data, capturing both low-level details and higher-level deepfake artifacts. At the initial convolutional layers, the model performs spatial filtering operations to detect patterns and features within the input images. These operations are followed by pooling layers, which down sample the feature maps to retain important information while reducing computational complexity. The subsequent dense layer aggregates the extracted features and learns to classify the input as either authentic or manipulated based on learned patterns.

Despite its relatively simple architecture, Meso4 has demonstrated remarkable efficiency and effectiveness in detecting Deepfake content. By focusing on essential features and maintaining a low parameter count, it strikes a balance between computational resources and detection accuracy, making it a valuable tool in the fight against the spread of misleading and harmful Deepfake media. Techniques like learning rate scheduling or adaptive optimizers e.g., Adam,

RMSprop can be beneficial in dynamically adjusting the learning rate during training to improve performance. For deep fake detection with Meso4, Rectified Linear Unit (ReLU) is often a suitable choice due to its simplicity and effectiveness in promoting sparse activations, which can help the model learn discriminative features efficiently. As depicted in the architecture Figure 3, Meso4 comprises four layers of convolutional and pooling operations followed by a dense network.

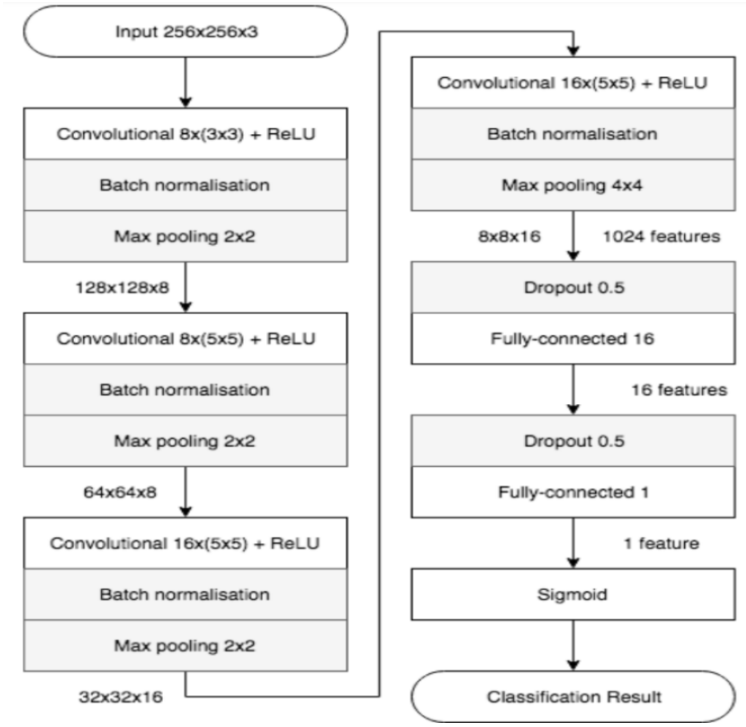

Figure 3
Meso4 Architecture

3.3 Deepfake Image Detection Process

Figure 4 represents the Deepfake image detection process, that begins with the collection of datasets containing images generated using the Face2Face and Deepfake techniques. These datasets are then loaded for feature extraction, focusing on key facial features such as eye shape, nose shape, and lip shape. Through temporal analysis, anomalies and inconsistencies in facial expressions introduced by Deepfake manipulation are identified. Subsequently, a Deepfake detection model is trained using the Meso4 algorithm, known for its effectiveness in analyzing micro-patterns and artifacts to detect manipulated facial images.

Once the model is trained, binary classification is performed to classify images as either real or fake based on the features learned during training. This binary classification approach enables the model to predict the authenticity of images, distinguishing between genuine and Deepfake images with the aim of improving the detection of manipulated content.

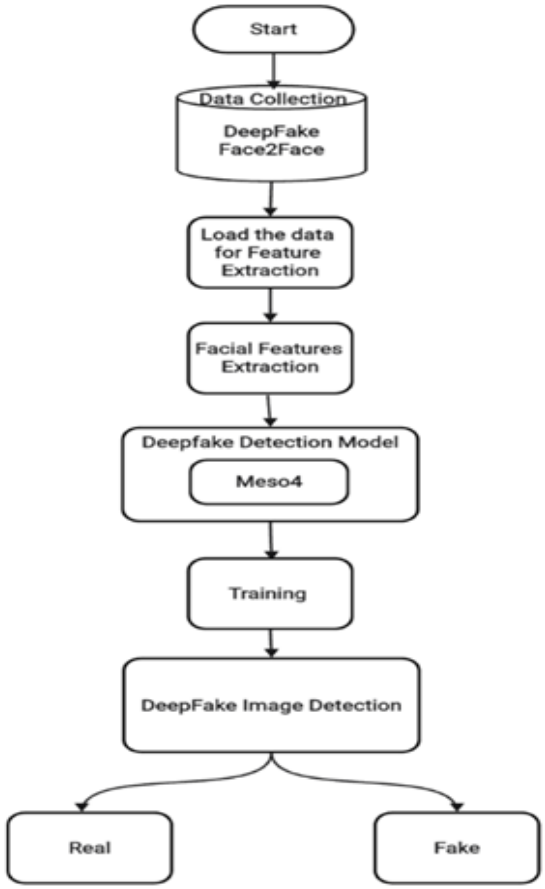

Figure 4
Fake Image Detection Process

Figure 5 describes a clear and straightforward process for preparing data to detect Deepfake videos. It starts with collecting input videos, which might be real or contain Deepfakes. Next, these videos undergo preprocessing, where they are split into individual frames. From these frames, facial regions are specifically identified and cropped out because these areas are most manipulated in Deepfakes. Finally, these cropped facial images are saved, creating a dataset ready for use in training a model to detect Deepfakes. This process ensures that the focus is on the most critical aspects of the videos for Deepfake detection.

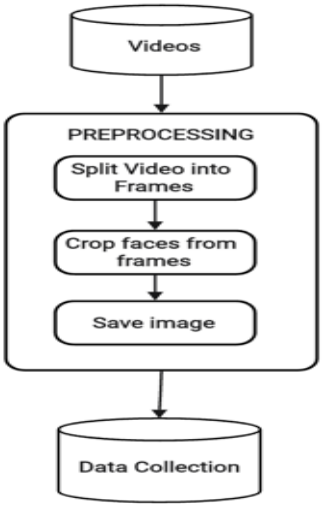

Figure 5
Data Collection Process

Forged videos were compressed, and faces have been extracted using the Viola-Jones detector. Approximately 50 frames were extracted per scene. The dataset has then been doubled with real face images, also extracted from various internet sources and with the same resolutions. Face Forensics set is to provide lossless compressed videos, which has enabled us to evaluate the robustness of our model with different compression levels. Both the real and fake datasets consist of a total of 961 images each shown in Figures 1 and 2.

The MESO4 architecture, renowned for its capability in capturing subtle inconsistencies characteristic of Deepfake images, forms the backbone of the model. Comprising convolutional layers, max-pooling operations, and fully connected layers, MESO4 excels in real-time Deepfake detection applications due to its compact design and efficient feature extraction capabilities. Once the dataset is prepared and the architecture selected, the model undergoes rigorous training, wherein parameters are optimized to minimize a predefined loss function, typically binary cross-entropy. Fine-tuning of hyper parameters such as learning rate, batch size, and number of epochs is crucial to achieve optimal performance. In the process of hyper parameter tuning using RMSprop optimizer with varying learning rate, the performance of the Meso4 model showed notable differences for Learning rate=0.1 shown in Table 1. Performance of MESO4 with Adam Optimizer and Learning Rate=0.001 is shown in table 2. This shows the performance of MESO4 model is high when it uses Adam optimizer and Learning rate as 0.001. The system performance is increased in terms of various parameters (such as accuracy, precision, recall, F1 score and loss) when number of epochs is increased, but the number of epochs used in this work is only 20 since it takes more than 3 hours for training.

Table 1
Performance of MESO4 with RMSPROP Optimizer and Learning Rate= 0.1

No. of Epochs	Training accuracy	Training Loss	Testing accuracy	Testing Loss
5	59.4	0.245	64.3	0.2242
10	65.8	0.225	67.8	0.2086
15	68.8	0.205	69.4	0.2026
20	73.5	0.187	71	0.2017

Table 2
Performance of MESO4 with Adam Optimizer and Learning Rate=0.001

No. of Epochs	Training Accuracy	Training Loss	Testing Accuracy	Testing Loss
5	72	0.12	71.2	0.12
10	74	0.10	73	0.11
15	77.4	0.09	75.6	0.10
20	79.2	0.08	77.39	0.10

In the context of real-time identification of Deepfake images, the MESO4 architecture emerges as a powerful tool, capable of discerning between authentic and manipulated content directly through camera input. By analyzing features such as inconsistencies in facial expressions, lighting, and texture, MESO4 effectively distinguishes between real and fake images with remarkable accuracy. This capability holds significant implications for various applications, including social media content moderation, news verification, and cyber security, where the rapid detection of Deepfakes is paramount. Through its integration with camera systems, MESO4 empowers users to mitigate the dissemination of misinformation in real-time, reinforcing the integrity of digital content and preserving trust in media platforms. Figure 6 demonstrates the real-time identification capabilities of spotting real and Deepfake images and manipulated content directly from camera inputs. This feature represents a significant advancement in digital security, offering immediate analysis of visual data to distinguish between authentic and altered images.

Figure 8 was generated using Deep Fake techniques characterized by blending facial features of Figure 7 to create a composite image. Despite the seamless merging of features to create a realistic image, the model correctly identifies this image as a Deepfake. This successful detection demonstrates the model's ability to spot even sophisticated manipulations, highlighting the importance of advanced detection systems in safeguarding against digital misinformation and protecting the authenticity of visual content. Figure 9 details the model's confidence levels in accurately predicting Deepfake images. This visualization provides insights into the model's certainty in correctly identifying manipulated content.

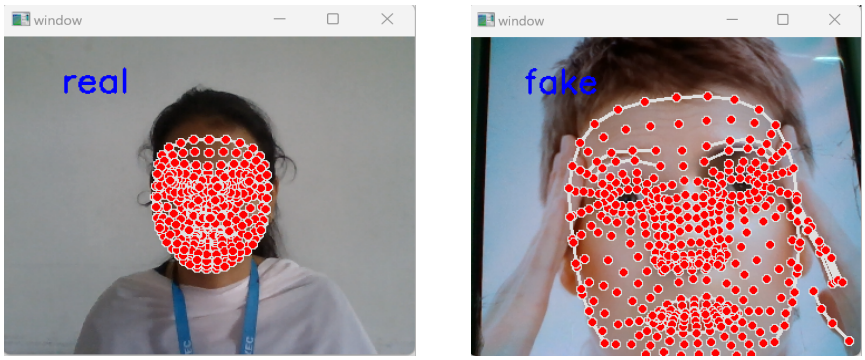

Figure 6
DeepFake Image Detection via Web Cam(Real and Fake images)

Figure 7
Real Image

Figure 8
Fake Image by Deepfake

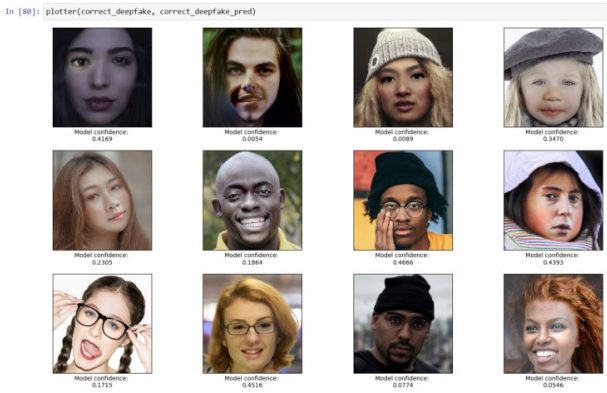

Figure 9
Model Confidence for correctly predicted Deep Fake images

4 Results and Discussion

4.1 Parameters for Evaluation

The metrics considered in this work used to evaluate the model are Precision, F1 score, accuracy, and recall, which are defined based on the values from the confusion matrix and the total number of samples within a specific class. The confusion matrix summarizes predictions against actual outcomes, highlighting the model's accuracy, errors, and misclassifications. The contents in confusion matrix are FP, FN, TP and TN represents False Positive, False Negative, True Positive and True Negative respectively.

Accuracy

Accuracy is determined using Equation (1), which involves comparing the total number of samples in a class to the total number of samples correctly identified as belonging to that class.

$$Accuracy = \frac{TP + TN}{FP + TN + TP + FN} \qquad (1)$$

Recall

Recall, also known as true positive or sensitivity, is the ratio of the total number of samples accurately classified as a specific class to the overall number of real samples within that class. This measure can be calculated using Equation (2).

$$Recall = \frac{TP}{FN + TP} \qquad (2)$$

Precision

Precision, also known as Positive Predictive Value, is a metric calculated using the given equation (3). It represents the fraction of samples correctly identified as a particular class out of all the samples classified as that same class.

$$Precision = \frac{TP}{FP + TP} \qquad (3)$$

F1-Score

The F1-Score, also known as the harmonic means of recall and precision, represents a balanced combination of these two factors. It is determined using Equation (4).

$$F1\ Score = \frac{2\ X\ Precision\ X\ Recall}{Precision + Recall} \qquad (4)$$

4.2 Performance Analysis

Table 3 outlines the training and validation performance metrics for four different neural network models: VGG-19, ResNet-50, Xception, and Meso4. Among these models, Meso4 exhibits the highest training accuracy at 79.20% with a corresponding training loss of 0.08 suggesting effective learning from the training data. In terms of validation accuracy, Meso4 also leads the group with a validation accuracy of 77.39% and validation loss of 0.3, indicating strong generalization performance on unseen data. Overall, Meso4 demonstrates promising performance both in terms of training accuracy and validation accuracy, highlighting its potential for various classification tasks, though further investigation into reducing validation loss may be warranted for improved robustness. Table 4 shows the performance analyzed during training and testing of different models.

Table 3
Performance of Various Models on Training and Validation Data

Models	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
VGG-19	74.92	1.46	73.28	2.49
ResNet-50	75.26	1.59	74.12	2.75
Xception	77.83	2.69	75.99	3.11
Meso4	79.20	0.08	77.39	0.3

Table 4
Performance Metrics of Various Models on Testing Data

Models	Precision	Recall	F1-Score	Accuracy
VGG-19	73	73	73	73
ResNet-50	74	73	73	74
Xception	76	74	74	76
Meso4	78	77	77	77

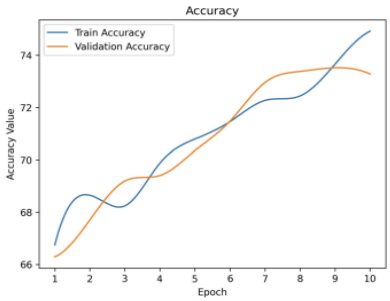

Figure 10
Meso4 - Epoch Vs Accuracy

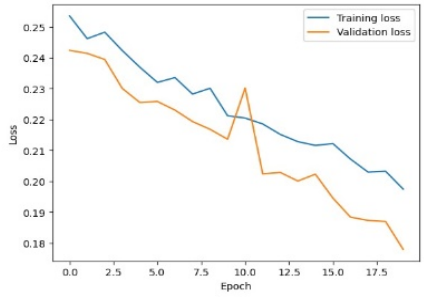

Figure 11
Meso4 -Epoch Vs Loss

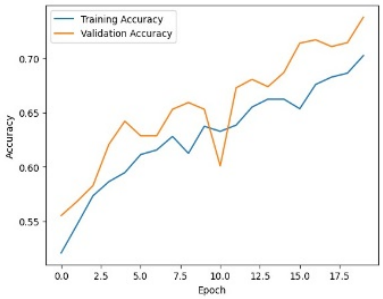

Figure 12
VGG-19-Epoch Vs Accuracy

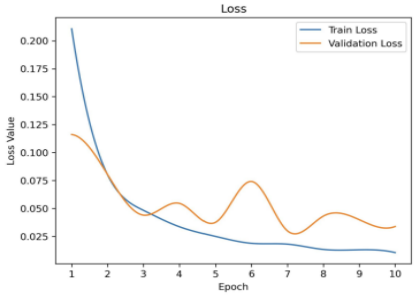

Figure13
VGG-19 - Epoch Vs Loss

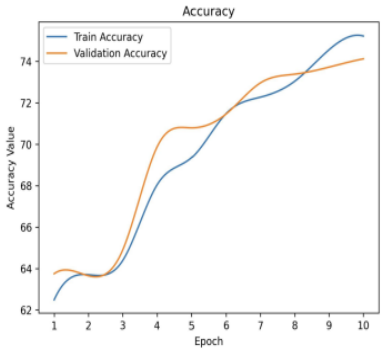

Figure 14
ResNet-50 -Epoch Accuracy

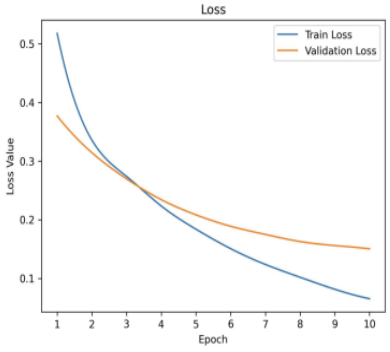

Figure 15
ResNet-50 - Epoch Vs Loss

Figure 10 illustrates that accuracy steadily increases with epochs and reaching approximately 79%, indicating effective learning and potential for further enhancement. Figure 11 shows that the loss initially decreases sharply, indicating rapid learning, but later stabilizes around 0.08, suggesting convergence to a stable point with minimal further reduction. Figure 12 represents the accuracy steadily increases with epochs, reaching around 75%, indicating effective learning but potential limitations in further improvement and Figure 13 initially decreasing the loss, but with fluctuations, suggesting convergence to a stable point with minimal further reduction. Figure 14 represents accuracy exhibits gradual improvement with epochs, reaching approximately 75%, indicating continuous learning but with slower progress compared to other models. Figure 15 depicts the loss curve displaying a declining trend, suggesting possible overfitting or convergence to a local minimum.

In the evaluation of network models, Figure 16 depicts VGG-19, ResNet-50, Xception, and Meso4—the precision, recall, F1-score and accuracy metrics provide insights into their performance across various classification tasks.

The VGG-19 and ResNet-50 demonstrate comparable precision, recall, F1-scores and accuracy where Xception exhibits slightly higher precision and accuracy values. However, Meso4 stands out with the highest precision, recall, F1-score and accuracy among the models, indicating its superior ability to accurately classify positive instances and effectively balance between precision and recall. These results suggest that Meso4 offers enhanced classification performance, making it a compelling choice for applications requiring high accuracy and reliability in identifying positive cases.

Regarding validation loss and model robustness, it is observed that the Meso4 model demonstrates the most favorable performance among all evaluated models. Specifically, it achieves the highest training accuracy of 79.20% with a low training loss of 0.08, indicating effective learning from the training data. More importantly, Meso4 also records the highest validation accuracy of 77.39% with a validation loss of 0.3, which is significantly lower than that of the other models—VGG-19 (2.49), ResNet-50 (2.75), and Xception (3.11). This comparatively lower validation loss suggests that Meso4 generalizes well to unseen data, indicating better robustness.

For 20 epochs, the model demonstrates consistent improvement in both training and testing accuracy, indicating effective learning and generalization capabilities. The model achieves its highest accuracy of 77.39% on the testing set after 20 epochs, highlighting its robust performance. The system's performance improves across multiple parameters with an increase in the number of epochs. However, due to training time constraints, this work utilizes only 20 epochs, as each training session exceeds 3 hours. The model can be trained for a greater number of epochs to achieve higher efficiency.

Conclusions

This research has made significant strides in addressing the challenge of face tampering in videos, a pressing issue in our digital age. Through the development of two innovative network architectures, which are efficient and provide low-cost solutions for detecting video forgeries, achieving high detection rates for both Deepfake and Face2Face manipulated videos under realistic internet conditions. Furthermore, the inner workings of deep learning models found the crucial roles those facial features, that the eye and mouth play a paramount role in the detection of faces forged with Deep fake. This insight not only enhances our current methodologies, but also sets the stage for future research.

The works success paves the way for exciting future enhancements and expansions. Some potential directions for future work include the development of real-time detection systems capable of operating within live video streams or social media platforms, providing instant alerts to potentially manipulated content. This capability would mark a significant leap in safeguarding information integrity in real-time communication channels.

Furthermore, exploring partnerships with technology companies and social media platforms could facilitate the implementation of these detection systems at a scale, creating a more secure and trustworthy digital ecosystem.

Lastly, there's a compelling need to address the ethical and privacy considerations associated with deploying Deepfake detection technologies, ensuring that efforts to protect digital authenticity do not inadvertently compromise individual rights or freedoms.

References

[1] Rossler, Andreas, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, and Matthias Nießner. "Faceforensics++: Learning to detect manipulated facial images." In Proceedings of the IEEE/CVF international conference on computer vision, pp. 1-11, 2019

[2] Dolhansky, Brian, Joanna Bitton, Ben Pflaum, Jikuo Lu, Russ Howes, Menglin Wang, and Cristian Canton Ferrer. "The Deepfake detection challenge (dfdc) dataset." arXiv preprint arXiv:2006.07397 (2020)

[3] Nguyen, Huy H., Junichi Yamagishi, and Isao Echizen. "Capsule-forensics: Using capsule networks to detect forged images and videos." In ICASSP 2019-2019 IEEEinternational conference on acoustics, speech and signal processing (ICASSP), pp. 2307-2311, IEEE, 2019

[4] Chugh, Komal, Parul Gupta, Abhinav Dhall, and Ramanathan Subramanian. "Not made for each other-audio-visual dissonance-based Deepfake detection and localization." In Proceedings of the 28th ACM international conference on multimedia, pp. 439-447, 2020

[5] Korshunov, Pavel, and Sébastien Marcel. "Deepfakes: a new threat to face recognition? assessment and detection." arXiv preprint arXiv:1812.08685 (2018)

[6] Agarwal, Sakshi, and Lav R. Varshney. "Limits of Deepfake detection: A robust estimation viewpoint." arXiv preprint arXiv:1905.03493 (2019)

[7] Kumar, K. Kranthi, Y. Kasiviswanadham, D. V. S. N. V. Indira, and Ch V. Bhargavi. "Criminal face identification system using deep learning algorithm multi-task cascade neural network (MTCNN)." Materials Today: Proceedings 80 (2023): 2406-2410

[8] Bharadwaj, Rakhi, Soham Ratnaparkhi, Rajendrasingh Rajpurohit, Kashish Rahate, Rahul Pandita, and Samarth Thosar. "Deepfake detection for preventing Audio and Video frauds using Advanced Deep Learning Techniques." In 2023 International Conference on Integrated Intelligence and Communication Systems (ICIICS), pp. 1-7, IEEE, 2023

[9] Güera, David, and Edward J. Delp. "Deepfake video detection using recurrent neural networks." In 2018 15th IEEE International Conference on advanced video and signal-based surveillance (AVSS) pp. 1-6, IEEE, 2018

[10] Li, Yuezun, Xin Yang, Pu Sun, Honggang Qi, and Siwei Lyu. "Celeb-df: A large-scale challenging dataset for Deepfake forensics." In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 3207-3216, 2020

[11] Zhou, Xiaoyu, Wenyuan Yang, Zhikai Chen, Bofei Guo, Zhongjie Ba, Zhihua Xia, Xiaochun Cao, and Kui Ren. "Avoid-df: Audio-visual joint learning for detecting Deepfake." IEEE Transactions on Information Forensics and Security 18 (2023): 2015-2029

[12] Yan, Zhiyuan, Yong Zhang, Yanbo Fan, and Baoyuan Wu. "Ucf: Uncovering common features for generalizable Deepfake detection." In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 22412-22423, 2023

[13] Mcuba, Mvelo, Avinash Singh, Richard Adeyemi Ikuesan, and Hein Venter. "The effect of deep learning methods on Deepfake audio detection for digital investigation." Procedia Computer Science 219 (2023): 211-219

[14] Chen, Tianxiang, Avrosh Kumar, Parav Nagarsheth, Ganesh Sivaraman, and Elie Khoury. "Generalization of Audio Deepfake Detection." In Odyssey, pp. 132-137, 2020

[15] Li, Yuezun, and Siwei Lyu. "Exposing Deepfake videos by detecting face warping artifacts." arXiv preprint arXiv:1811.00656 (2018)

[16] Cozzolino, Davide, Justus Thies, Andreas Rössler, Christian Riess, Matthias Nießner, and Luisa Verdoliva. "Forensictransfer: Weakly-supervised domain adaptation for forgery detection." arXiv preprint arXiv:1812.02510 (2018)

[17] Li, Yuezun, and Siwei Lyu. "Obstructing Deepfakes by disrupting face detection and facial landmarks extraction." Deep Learning-Based Face Analytics (2021): 247-267

[18] Ding, Xinyi, Zohreh Raziei, Eric C. Larson, Eli V. Olinick, Paul Krueger, and Michael Hahsler. "Swapped face detection using deep learning and subjective assessment." EURASIP Journal on Information Security 2020 (2020): 1-12

[19] Ahmed, Saadaldeen Rashid, Emrullah Sonuç, Mohammed Rashid Ahmed, and Adil Deniz Duru. "Analysis survey on Deepfake detection and recognition with convolutional neural networks." In 2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA) pp. 1-7, IEEE, 2022

[20] Huang, Baojin, Zhongyuan Wang, Jifan Yang, Jiaxin Ai, Qin Zou, Qian Wang, and Dengpan Ye. "Implicit identity driven Deepfake face swapping detection." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4490-4499, 2023

[21] Korshunov, Pavel, Sébastien Marcel, and Ivana Chingovska. "A study of the effectiveness of image quality assessment in the context of face spoofing detection." IEEE Transactions on Information Forensics and Security 13, No. 11 (2018): 2785-2797

[22] Agarwal, Nitin, and Lav R. Varshney. "Hypothesis testing for detecting GAN-generated images in facial analysis applications." IEEE Transactions on Information Forensics and Security 15, No. 11 (2020): 2965-2977