2025 First International Conference on Advances in Computer Science, Electrical, Electronics, and Communication Technologies (CE2CT) | 979-8-3315-1857-8/25/$31.00 ©2025 IEEE | DOI: 10.1109/CE2CT64011.2025.10939891

Detection of quality Deep fake images and Videos using Customised Convolutional Neural Networks

Alap Mahar
Department of Computer Science and Engineering
Bhagwant University
Ajmer-305004, Rajasthan, India
alapmahar@gmail.com

Pushpneel Verma
Department of Computer Science and Engineering
Bhagwant University
Ajmer-305004, Rajasthan, India
pushpuneelverma@gmail.com

Ajit Singh
Department of Computer Science and Engineering
VMSB University
Pithoragarh Campus, India
erajit@rediffmail.com

Abstract—**Due to advancements in artificial intelligence there are numerous deep fake images and video collections are available on the Internet and social media. The primary aim of the study is to analyse the deep fake images and videos since the quality is constantly improving a novel method was developed for accurate detection of quality deep fakes. The suggested approach uses a customised convolutional neural network (CNN) technique that uses facial landmark detection to extract structured data from photos and video frames before feeding it into the methods of CNN. The customised CNN model is augment-based CNN for generation of fake images and fake data. Involves about 260 films from the data set in which 202 images are made up while others were real images. About 300 videos are used in which 250 videos are fake and 50 were real the proposed model achieved the accuracy of 95.58% and 0.97 AUC score that outperforms the existing models like MLP-CNN and CNN. Additionally, the method succeeds with greater accuracy then the conventional models like DST-Net, VGG 16 Efficient Net. This research study's primary goal is to create a new CNN learning method for identifying high-quality deepfake photos and videos.**

Keywords*—deep fake images, video collections, customised CNN, facial landmark, Accurate, detection*

I. Introduction

These days, deep learning algorithms are being used to produce phoney videos. These phoney videos are frequently produced with malevolent intent, such as slandering a well-known political figure, fabricating pornography of well-known actors or actresses, altering forensic and court-related evidence, and launching frauds and schemes to manipulate identities. Since the phoney videos produced by deep learning systems closely resemble the real ones, they are employed for these purposes. People can't even tell the difference among real and fake images or videos. A Generative Adversarial Network (GAN) is the deep learning method used to create these kinds of films [1].

The availability of current developments has led to the proliferation of deepfake movies on social media [2]. An instance of a "deepfake" is a photograph or video where the subject's likeness is replaced with someone else's. One of the biggest issues confronting contemporary civilization is deepfake.

Celebrity's faces have been swiped over their photos in a number of pornographic videos using Deepfake. Deepfake was used in this capacity in addition to disseminating false information for politicians [3-5].

At the same time, attackers are generating and disseminating misinformation widely by abusing deepfakes. Humans are known to be significantly impacted by digital pictures [6]. Thus, the simplicity of creating convincing and manipulative deepfakes poses a serious danger to the reliability of information. There may be major political, social, financial, and legal repercussions as a result of these deepfakes directed against people and organisations being made extensively and easily accessible on social media sites [7].

The most prevalent kind of fake media appears to be deepfakes, which are composed of audio, video, and graphics. The first "deepfake" film, which featured a porn actor's face in place of a celebrity's, was released to the public in 2017. Deepfakes became popular after a Reddit user with the handle "Deepfake" demonstrated how to change a celebrity's look to make them the main character in a pornographic video clip [8].

Fake news on social media can be produced by using deep learning algorithms to create forged videos and photos that are used for face authentication. Facial forgeries on the targeted person's video [9, 10] or ID evidence [11, 12] can be easily simplified. There are three kinds of phony photos [13]: Face Trade, Lip Sync, and Lip Sync. Lip Sync is the term used to depict the source video that has been changed by keep sound with a consistent development in the mouth region. Vivified demeanors on the heads, faces, and eye developments of the individual being depicted are delivered utilizing the manikin ace strategy. To show the development of the manikin, this acting is additionally acted before the camera. In reality, Face Trading substitutes the objective face for the source face in source and objective recordings. The most well-known deepfake method is Face Trade [14, 15].

This study is driven by the necessity to create new techniques for detecting deepfakes since their quality is continuously increasing. The two primary classifier

979-8-3315-1857-8/25/$31.00 ©2025 IEEE

categories utilised in the identification of deep fakes are deep classifiers and shallow classifiers. By recognising the differences in their properties, shallow classifiers are able to distinguish between actual and fake photos and videos with accuracy. The goal of this examination is to make another CNN learning model that can dependably distinguish deepfake facial pictures. The strategy that has been proposed includes utilizing facial milestone recognition to remove organized information from video outlines. The CNN then, at that point, involves the gathered information as info. CNNs are fed video image frames directly for automated feature extraction.

II. Literature Review

We will examine some of the research that has been conducted on the production and detection of Deepfakes in this part. These and other deepfake movies are growing in popularity on social media platforms, which has made the whole community take the threat seriously and prompted scholars everywhere to develop advanced deepfake detection techniques. Numerous methods can be found in the most recent literature.

This study presents multilayer hybrid recurrent models of DL for deepfake video identification [16]. The structures proposed in this paper are developed using noise based worldly face convolutional highlights and fleeting learning of hybrid techniques for repetitive DL. Experiments have demonstrated how well these models perform in comparison to stacked recurrent DL models.

The VGG-Face [17] with ResNet50 architecture was used by the authors in [18] to create their detection model. To identify the phoney faces, they observed the deep face recognition neurone behaviours. AUC performance on the FF++, DFDC, and Celebrity-DF datasets is 98.5%, 68%, and 66.8%, respectively, according to evaluation.

Additional approaches have integrated CNN with additional learning models such Recurrent Neural Networks (RNN), LSTM, and Capsule Networks to further improve accuracy by detecting temporal disparities [19]. These techniques have shown encouraging results on datasets containing deepfake and Face Swap movies.

Despite several developments, it is still challenging to maintain a high level of accuracy while creating new methods for deepfake generation, and stronger algorithms are still needed to detect lower-quality deepfakes.

According to an analysis by [20], eye blinking is an unplanned, unconscious, and normal human behaviour. They claim that because eye blinking is natural and impossible to manufacture, we can determine whether a film is real or phoney by examining its pattern. This methodology does not provide good accuracy in a typical dataset and is a form of data leak. When applied to generalised data, this approach performs poorly; instead, it excels on a specific dataset that contains information pertaining to eye blinking.

They simply used a specific dataset to demonstrate their accuracy. If there were no blinking eyes in successive frames of a video, they labelled it as false.

For each deep neural network (DNN) utilised in deepfake generating methods, [21] offers model architectural charts. who concentrate on reenactment techniques for deepfake generation. The technological difficulties with generating and detection systems are not included in the survey.

Recently, a paper on different DeepFake detection techniques was proposed. In order to detect the phoney images, this uses a CNN and DNN that extracts eye blinks and combines the image's little noises with Long Short Term Memory (LSTM). [22] CNN and LSTM were used to extract the video's frame image elements. [23] used a CNN based on the Visual Geometry Group (VGG) 16 and Residual Network (ResNet) 50 to identify the distorted face. [24] used SVM to identify 68 landmarks that were derived from facial pictures.

Using residual noise, or the distinction between the first picture and its denoised partner, is the groundwork of the technique proposed in this work [25]. Research on leftover clamor has shown that it is successful in deepfake recognition because of its one of a kind and discriminative qualities, which CNNs with TL (Move Learning) can really record. To assess the viability of our methodology, we applied low-goal video cuts from FaceForensics++ and high-goal video cuts from Kaggle DFDC (Deepfake Identification challenge). The got results show a serious level of precision when contrasted with other contending approaches.

[26] focusses on distinguishing faked photographs and recordings or visual media uprightness check. DL-created deepfakes are talked about along with new methodologies of information driven measurable to counter them. They divide detection techniques into two categories: DL-based techniques and conventional techniques. The review also highlights the shortcomings of the forensic techniques used today as well as the prospects and difficulties that lie ahead.

III. Methodology

A. Description of data

A total of 300 videos were used, 250 of which were fake and 50 of which were actual. It utilizes 260 motion pictures from the DFDC dataset, 202 of which are phony while the leftover ones are genuine. Every video has fifteen second term.

It uses 260 movies from the DFDC dataset, 202 of which are fake while the remaining ones are authentic. Facebook established the DFDC in partnership with renowned academic experts and business leaders to accelerate the development of literature techniques for detecting deepfake videos.

For the challenge, Facebook created and shared a unique dataset with more than 110,000 video. To differentiate and assess their deepfake discovery calculations, explore novel systems, and trade particular data, the DFDC has advanced joint effort among experts from various regions of the planet. A fundamental dataset containing 7000 films showing two face change techniques and a piece of examination, as well as a complete dataset involving 130,000 recordings exhibiting eight facial change draws near and an exploration paper, include the DFDC dataset. The complete dataset was used by competitors in a Kaggle competition to create better approaches for detecting tampered medications. Facebook created the dataset by using paid actors who gave their permission for their appearances to be used and altered for the dataset.

B. Customised Convolutional Neural Network

The Customised CNN approach was developed with the goal of improving the quality of deep fake determination. A common DNN for identifying patterns in photos is the convolutional neural network. For consistency and speedier processing, every image is also scaled to fit inside the 220 px by 3 px restrictions. A part of the visual frames are given to CNN. The network is made fit for perceiving interesting features by utilizing input boundaries of 32, 64, and 128 channels of expanding sizes. The methods is developed utilizing a changed CNN design.

A comparison between the CNN and Multilayer Perceptron (MLP) models was carried out. One neural network model that is frequently used in computer vision tasks is the MLP On the other hand, CNN have outperformed it in this area. Due to its reliance on fully linked layers, in which each perceptron establishes a link with every other perceptron, the MLP is considered inappropriate for contemporary complex computer vision applications.

C. Proposed Framework

It is possible to retrieve individual picture frames from video after it has been initially delivered as input. The eyes, nose, and lips of a person can be found using a facial landmarks detector. This information can be used to infer face traits, such as eye blinks. This input needs to be preprocessed in some way before being fed into the model. By and by, the pictures are changed over into their mathematical portrayal through pre-handling. In the first, the information picture outlines are resized to 226 by 226 and the facial district of interest is trimmed. It is fundamental to guarantee that each picture is in the RGB channel. Utilizing this expert DL technique in light of the model utilized by CNN, the characterization step can decide if a specific video is a deepfake or not. The means of the proposed framework are displayed in Fig. 1.

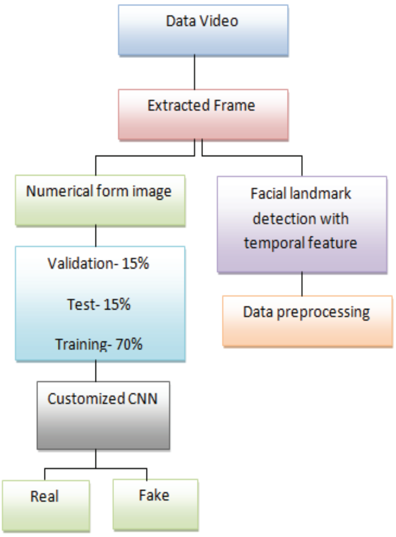

Fig 1. Proposed Framework

D. Temporal facial features are studied

It can be difficult to stop abnormal eye blinking from compromising deepfakes. In this stage, the subject's blinking pattern is recorded. Eye coordinates, which are derived from facial landmarks, are fed into the blink detector. One can tell if someone is blinking by looking at their eye's aspect ratio (EAR). Each eye can be independently represented by one of six different landmark positions.

E. Preparing the data

Preprocessing enables the removal of undesirable artefacts and enhances essential components required for the application under development. Some of these elements may change depending on the application. It establishes a baseline size for every image to take into consideration differences in photo sizes when obtaining images from cameras to be employed by the AI techniques.

F. Performance Measure

The correctness of the model influences both its training performance and its real-world behaviour. It will not, however, specify how it would be applied to the problem. Accuracy merely tells us how well the trained model works.

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

When assessing if there are more false positives than real positives, precision is essential..

$$Precision = \frac{TP}{TP + FP} \qquad (2)$$

Recall is useful when the number of false negatives is higher. The efficacy of the approach we use decreases when false negatives happen more frequently than.

$$Recall = \frac{TP}{TP + FN} \qquad (3)$$

The weighted mean of recall and precision is known as the F1-score.

$$F1 - score = \frac{2. precision. recall}{precision + recall} \qquad (4)$$

Erroneous data categorizations are corrected by using logarithmic loss, also known as log loss. It significantly improves the classification procedure for numerous categories [42]. The following formula is used to determine LL, where N is the number of examples that belong to M classes:

$$LL = \frac{-1}{N} \sum \sum X_{ij}. (\log P_{ij}) \qquad (5)$$

IV. Empirical Results

The results of this research can be used to identify deepfakes in videos. A voiceover, face swap, or both could be used in a deepfake. The labels "Fake" and "Real" in the training data's label column serve as indicators. The likelihood that the film is a fake has been determined here (Fig. 2). While the x-axis shows the video class, the y-axis shows the overall numbers. The numbers 0 and 1 represent the two categories of videos in this story: real and fraudulent. This graph shows that the numbers of counts for the two classifications are roughly identical. More specifically, there are 1965 REAL values and 1992 FALSE values.

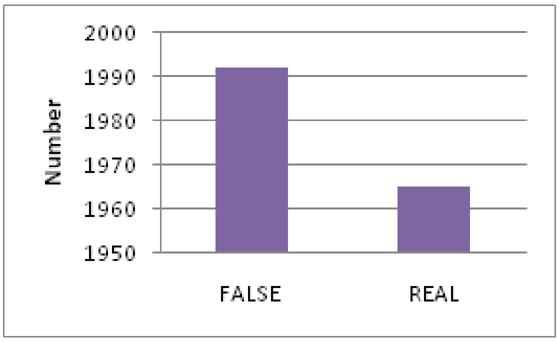

Fig 2.Data distribution for the collection of deepfake videos

Figure 3's comparison line graph illustrates the differences in training and validation loss values across the three methods.

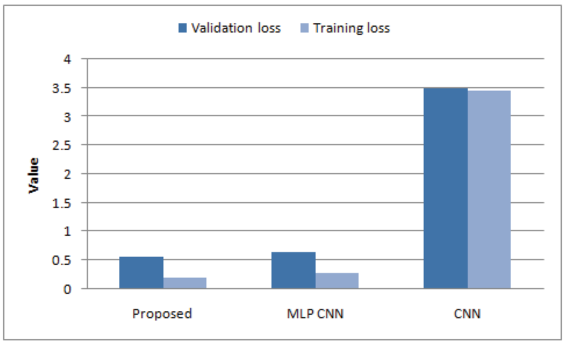

Fig 3. Comparsion of validation and training loss

It can infer from the data analysis in Figure 7 that distinct models are distinguished by varying Loss values. The MLP-CNN algorithm's loss values are CNN training 3.4544 and validation 3.492; suggested training 0.2114 and validation 0.564; and validation 0.6494 and training 0.2859.

Figure 4 displays a graph that contrasts the accuracy and AUC Score of the three methods.

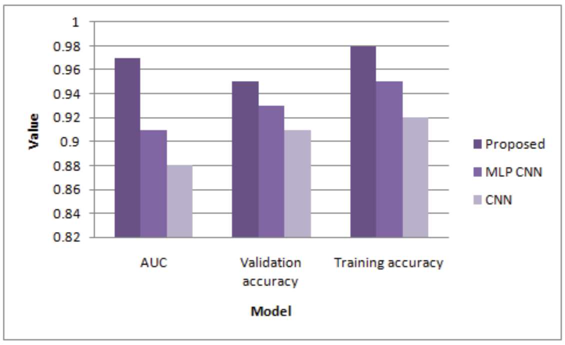

Fig 4. Comparison of AUC, validation accuracy and training accuracy

The comparison graph shows that there is a significant degree of resemblance between the CNN validation and training accuracy. In terms of accuracy on validation and training data, the proposed Customised CNN fared better than MLP-CNN, despite the latter producing good classification findings. The AUC score of 0.91 for the MLP-AUC CNN is higher than that of 0.88 for CNN alone, another popular method, according to the comparative line graph. As can be seen, MLP-CNN not only outperformed the suggested Customised CNN in terms of AUC score value, but also performed better than it, with an AUC score of 0.97.

To assess and gauge the efficacy of the suggested approach, in this paper it compared suggested method with the most cutting-edge techniques already in use. Table I displays the accuracy level of the current and suggested systems. This study carefully compared the suggested approach with the approaches described by [27] and [28].

Table I. Comparison Table

Method	Accuracy
Proposed	95.58
DST Net	90.94
VGG 16	81.03
Efficient Net	59.64

It is evident that models with accuracy ranges of 59.64 to 90.94% include Efficient Net, VGG-16, and DST-Net. The recommended approach outperforms the state-of-the-art models at the moment with the highest accuracy of 95.58%. The DST-Net method outperformed the Efficient Net method in the video-only modality, where total accuracy was the lowest.

V. Conclusion

In the modern world, deepfake detection is crucial and requires sophisticated detection methods because it will get harder to identify deepfakes in the future. Improvements in detecting techniques should continue since deepfakes can have serious social and political ramifications. This work, suggested a new technique for identifying AI-generated deepfake films. The suggested method demonstrated an AUC score of 0.97, a loss of 0.564, and a testing accuracy of 95.58%. In terms of performance, it outperformed CNN and MLP-CNN, two more techniques. Additionally, proposed approach outperformed more recent models like DST-Net, VGG-16, and Efficient Net in terms of accuracy.

Among the issues and restrictions that influence the viability and exactness of proposed models in distinguishing counterfeit news are the datasets utilized, worries with overfitting and underfitting, picture based credits, highlight vector encoding, techniques for AI, and information combination. The recommended strategy works well to stop the spread of fake films, particularly on social networking sites.

Concerning with overfitting and underfitting, picture based features, highlight vector encoding, strategies for machine learning, information combination, and the datasets utilized are a portion of the issues and impediments that influence the viability and exactness of proposed techniques in distinguishing fake news. Additionally, a reasonable blend should incorporate more temporal and spatial face data. Better models must also be tested on bigger, more balanced datasets utilising various DL techniques.

References

[1] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2020). Generative adversarial networks. Communications of the ACM, 63(11), 139-144.

[2] Almars, A. M. (2021). Deepfakes detection techniques using deep learning: a survey. Journal of Computer and Communications, 9(05), 20-35.

[3] Nataraj, L., Mohammed, T. M., Chandrasekaran, S., Flenner, A., Bappy, J. H., Roy-Chowdhury, A. K., & Manjunath, B. S. (2019). Detecting GAN generated fake images using co-occurrence matrices. arXiv preprint arXiv:1903.06836.

[4] Wang, S. Y., Wang, O., Zhang, R., Owens, A., & Efros, A. A. (2020). CNN-generated images are surprisingly easy to spot... for now. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 8695-8704).

[5] Hsu, C. C., Lee, C. Y., & Zhuang, Y. X. (2018, December). Learning to detect fake face images in the wild. In 2018 international symposium on computer, consumer and control (IS3C) (pp. 388-391). IEEE.

[6] Pollen, A. (2016). The rising tide of photographs: Not drowning but waving?. Captures, 1(1).

[7] Chesney, B., & Citron, D. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. Calif. L. Rev., 107, 1753.

[8] Güera, D., & Delp, E. J. (2018, November). 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.

[9] Thies, J., Zollhofer, M., Stamminger, M., Theobalt, C., & Nießner, M. (2016). Face2face: Real-time face capture and reenactment of rgb videos. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2387-2395).

[10] Korshunova, I., Shi, W., Dambre, J., & Theis, L. (2017). Fast face-swap using convolutional neural networks. In Proceedings of the IEEE international conference on computer vision (pp. 3677-3685).

[11] Tewari, A., Zollhoefer, M., Bernard, F., Garrido, P., Kim, H., Perez, P., & Theobalt, C. (2018). High-fidelity monocular face reconstruction based on an unsupervised model-based face autoencoder. IEEE transactions on pattern analysis and machine intelligence, 42(2), 357-370.

[12] Lin, J., Li, Y., & Yang, G. (2021). FPGAN: Face de-identification method with generative adversarial networks for social robots. Neural Networks, 133, 132-147.

[13] Chesney, R., & Citron, D. (2019). Deepfakes and the new disinformation war: The coming age of post-truth geopolitics. Foreign Aff., 98, 147.

[14] Lyu, S. (2020, July). Deepfake detection: Current challenges and next steps. In 2020 IEEE international conference on multimedia & expo workshops (ICMEW) (pp. 1-6). IEEE.

[15] Jafar, M. T., Ababneh, M., Al-Zoube, M., & Elhassan, A. (2020, April). Forensics and analysis of deepfake videos. In 2020 11th international conference on information and communication systems (ICICS) (pp. 053-058). IEEE.

[16] Jaiswal, G. (2021, November). Hybrid recurrent deep learning model for deepfake video detection. In 2021 IEEE 8th Uttar Pradesh section international conference on electrical, electronics and computer engineering (UPCON) (pp. 1-5). IEEE.

[17] Wanyonyi, D., & Celik, T. (2022). Open-source face recognition frameworks: A review of the landscape. IEEE Access, 10, 50601-50623.

[18] Wang, R., Juefei-Xu, F., Ma, L., Xie, X., Huang, Y., Wang, J., & Liu, Y. (2019). Fakespotter: A simple yet robust baseline for spotting ai-synthesized fake faces. arXiv preprint arXiv:1909.06122.

[19] Nguyen, H. H., Yamagishi, J., & Echizen, I. (2019). Use of a capsule network to detect fake images and videos. arXiv preprint arXiv:1910.12467.

[20] Jung, T., Kim, S., & Kim, K. (2020). Deepvision: Deepfakes detection using human eye blinking pattern. IEEE Access, 8, 83144-83154.

[21] Mirsky, Y., & Lee, W. (2021). The creation and detection of deepfakes: A survey. ACM computing surveys (CSUR), 54(1), 1-41.

[22] Bayar, B., & Stamm, M. C. (2016, June). A deep learning approach to universal image manipulation detection using a new convolutional layer. In Proceedings of the 4th ACM workshop on information hiding and multimedia security (pp. 5-10).

[23] Madaan, G. (2018). Various Approaches of Content Based Image Retrieval Process: A Review. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 3(1), 711-716.

[24] Tran, S. N., & Garcez, A. S. D. A. (2016). Deep logic networks: Inserting and extracting knowledge from deep belief networks. IEEE transactions on neural networks and learning systems, 29(2), 246-258.

[25] El Rai, M. C., Al Ahmad, H., Gouda, O., Jamal, D., Talib, M. A., & Nasir, Q. (2020, November). Fighting deepfake by residual noise using convolutional neural networks. In 2020 3rd International Conference on Signal Processing and Information Security (ICSPIS) (pp. 1-4). IEEE.

[26] Verdoliva, L. (2020). Media forensics and deepfakes: an overview. IEEE journal of selected topics in signal processing, 14(5), 910-932.

[27] Ilyas, H., Javed, A., & Malik, K. M. (2023). AVFakeNet: A unified end-to-end Dense Swin Transformer deep learning model for audio–visual deepfakes detection. Applied Soft Computing, 136, 110124.

[28] Khalid, H., Kim, M., Tariq, S., & Woo, S. S. (2021, October). Evaluation of an audio-video multimodal deepfake dataset using unimodal and multimodal detectors. In Proceedings of the 1st workshop on synthetic multimedia-audiovisual deepfake generation and detection (pp. 7-15).