



Advanced 3D CNN Techniques For Robust Face Forgery Detection

Mrs. C. Kavitha^{1*}, S. Sharan², P. Yuvar Sankar³, S. Mohamed Shafic⁴

^{1*}Assistant Professor, Department of Computer Science and Engineering, Manakula Vinayagar Institute of Technology, Pondicherry University, Puducherry, India, kvtha1705@gmail.com

^{2,3,4}B.Tech, Manakula Vinayagar Institute of Technology, Pondicherry University, Puducherry, India, yuvans569@gmail.com

Citation: Mrs. C. Kavitha et.al (2024), Advanced 3D CNN Techniques For Robust Face Forgery Detection, *Educational Administration: Theory And Practice*, 30(5), 1330-1339 Doi: 10.53555/kuey.v30i5.3083

ARTICLE INFO

ABSTRACT

Researchers are currently investigating sophisticated methods for face forgery detection in response to the escalating worries regarding the simplicity and effectiveness of existing face forging techniques, in an effort to prevent potential misuse of this technology. A wavelet dual-branch network is one method now in use for predicting and recognizing modified faces. But even if this system works well in some situations, it can still have errors, which would reduce its overall dependability. Researchers have presented a fresh solution to this problem by integrating the 3D Convolutional Neural Network (3DCNN) algorithm into the framework for facial forgery detection. The purpose of integrating 3DCNN is to improve prediction accuracy and get beyond the drawbacks of the wavelet dual-branch network. 3DCNN, in contrast to conventional 2D convolutional networks, considers the temporal dimension, enabling it to capture spatiotemporal features in the data. The modified approach delivers better prediction performance by utilizing 3DCNN's capabilities, especially in situations where fake faces may have subtle or dynamic modifications. The algorithm's enhanced sensitivity to minute details stems from its capacity to assess volumetric data in both geographical and temporal domains, making it a more robust and dependable face forgery detection method. This development emphasizes the significance of continuously improving detection approaches in the face of technology improvements and represents a significant step forward in the ongoing attempts to keep ahead of growing face forgery techniques.

Index Terms—3D Convolutional Neural Network (3DCNN), wavelet dual-branch network, 2D convolutional networks, face forgery

I. INTRODUCTION

A growing problem in the digital era, image fraud includes a variety of dishonest techniques used to falsify the legitimacy of visual content. Traditional forensic techniques are no longer adequate for identifying and thwarting such forgeries due to the development of advanced image editing software and deep learning algorithms. Cloning is one of the simplest, yet most powerful methods used in image alteration. This technique involves copying or replicating parts of an image to hide undesirable components or create whole different scenes. Forgers can change the storyline or context of an image by skillfully incorporating cloned pieces, which can mislead viewers into making incorrect assumptions. Another common fabrication technique is splicing, which combines components from several different sources to produce a composite image. Forgers create deceptive images by fusing parts of other photos together, which appear to represent situations or events that never happened. This method makes use of the capacity to adjust viewpoints and spatial relationships, which frequently produces scenes that are convincing yet entirely artificial. Subtle modifications intended to enhance or distort visual aspects present additional difficulties for authenticity verification, in addition to these overt manipulation approaches. Color balance, contrast, and sharpness adjustments can drastically change how realistic a picture appears to viewers, making it harder for them to distinguish between real and fake material.

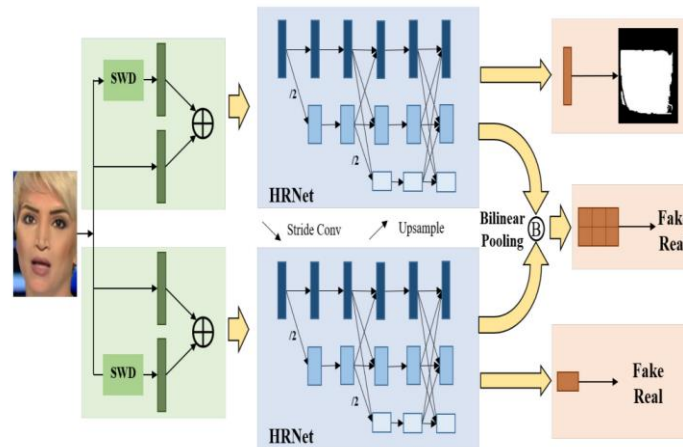


Fig.1 Architecture Diagram of Existing System

An inconsistency-aware wavelet dual-branch network is used in the current face forgery detection system, which is a major improvement above the high-fidelity fake faces produced by current forging systems. The foundation of this technique is the identification of two critical indicators of forgeries: discrepancies within and across images. The By including in order to increase the exploitation of these hints. Then, taking into account the various characteristics of intra- and inter-picture discrepancies, the dual-branch network is built to predict forgery labels at both the image and pixel levels. The FaceForensics++ dataset, a vast collection of altered facial movies produced using a variety of face manipulation techniques, such as deepfakes, Face2Face, and FaceSwap, is introduced in the paper "FaceForensics++: Learning to Detect Manipulated Facial Images" by Andreas Rössler et al. This research presents to assess temporal and spatial information in order to detect altered facial images. The suggested model advances face forgery detection techniques by demonstrating remarkable performance in differentiating between real and modified movies by training on the FaceForensics++ dataset. Yuezun Li et al.'s article, "Exploring DeepFake Videos By Detecting Face Warping Artifacts," focuses on recognizing face warping artifacts that were included during the creation process in order to detect deepfake videos. The research presents a detection system that takes use of the facial geometry and texture anomalies present in deepfake films through painstaking examination. Through the use of these distinct qualities, the suggested approach successfully differentiates between authentic and altered videos, consequently improving the capacity to identify deepfake content.

The use of Capsule Networks, a novel neural network architecture, for the detection of deepfake movies is explored in "DeepFake Detection Based on Capsule Networks" by Seyed Mohsen Zeraati et al.. Because Capsule Networks can record spatial hierarchies in features, they are very useful for assessing facial traits, which are essential for distinguish between authentic and counterfeit films. In order to efficiently detect deepfake movies, the research introduces a unique Capsule Network-based method.

Xiangpeng Li et al. introduce the FDFD dataset in their paper "FDFD: A Benchmark Dataset for Forgery Detection in Face Images"[4]. The FDFD dataset is a benchmark dataset created especially for assessing forgery detection methods in face photos. The FDFD dataset, which consists of a varied set of real and altered face photos, offers an efficient and uniform platform for researchers to create and contrast different detection techniques. The dataset advances face forgery detection research by enabling thorough evaluation and benchmarking of forgery detection algorithms, which helps to create dependable and strong detection methods.

Donghyun Kim et al.'s "A Survey of Deep Learning-based Deepfake Detection" offers a thorough analysis of deep learning-based methods for identifying deepfake films. The study examines the benefits, drawbacks, and performance of several methods using a methodical analysis of generative adversarial networks (GANs), capsule networks, and convolutional neural networks (CNNs). The survey article provides insightful perspectives on the state-of-the-art in deepfake detection by combining findings from previous research, directing future research paths in this quickly developing.

We provide the following summary of our contributions to this study:

- The two main characteristics for identifying forged faces are intra- and inter-image inconsistencies and irregularity. We plan to make the most of them by implementing two tasks that seek to improve and extract inconsistent traits, respectively.
- We use the supplementary wavelet decomposed images as extra inputs to improve the inconsistency features. The SWD preserves translation invariance while maintaining the resolution. It can fully extract localized frequency information.
- We suggest using a dual-branch multi-task network to manage the disparities between the two features in order to extract the inconsistency features. The two branches focus on distinct irregularity traits by learning forging labels at the image and pixel levels, respectively. In order to combine characteristics from the two branches, bilinear pooling is used.

II. REALTED WORK

In this part, we quickly cover a few of the earlier studies that are relevant to our methodology.

1. Face Manipulation Detection

The primary biological characteristic of a person, such as a universal ID card, is unquestionably their face. Because of this, the development of artificial intelligence has caused a tremendous deal of anxiety.

Tools that alter facial features in videos in a believable fashion or create realistic-looking fake faces. A more modern image-forgery technique is called DeepFakes. As a result, a number of orthogonal works have recently been presented to differentiate between actual and artificial faces. Face forgery detection techniques fall into two broad categories: approaches based on data-driven approaches and techniques based on discriminative classifiers.

The former group typically makes use of several semantic differences between the face and the head. A forensic method was presented by Andreas Rössler et al. to simulate facial movements and phrase that embodies a unique speech style. They create a novel detection model that can discriminate between modified and real photos. A deep convolutional neural network (DCNN) was trained by Xiangpeng Li et al. to identify fraudulent face videos based on the identification of eye blinking in videos.

Rössler et al demonstrated that the classifier based on XceptionNet outperformed all other variations in recognizing fakes when it came to data-driven techniques. Conventional portrayal Video frequently does not lend itself to forensics procedures since compression drastically reduces the quality of the data. In order to obtain the mesoscopic features of images, Donghyun Kim et al. proposed two networks that have a limited number of layers. Similarly, to identify face manipulation, Zhou et al. suggested using a two-stream network. A dual-branch structure was also proposed by Seyed Mohsen Zeraati et al. , with one branch providing the original information and the other enhancing it.

2. 3D CNN

Convolutional neural networks, or 3D CNNs, are an effective method for identifying face forgeries, especially when it comes to altered facial videos. By analyzing whole video sequences at once, 3D CNNs are capable of capturing temporal and geographical data, in contrast to typical 2D CNNs that examine images frame by frame. This grasp of time is essential for spotting abnormalities or inconsistencies that could develop over time in a film that has been altered.

In order for a 3D CNN to function, time must be added as a new dimension to the convolutional filters. The network may acquire sophisticated spatiotemporal properties straight from the input video data by applying these filters across the spatial dimensions (width and height) and the temporal dimension (time).

When it comes to facial forgery detection, a 3D CNN can be trained to identify minute variations in facial expressions, irregularities in texture, or strange distortions brought about by face swapping or deepfake techniques. The network can distinguish between real and fake facial sequences by examining the dynamics of facial expressions and movements over a series of frames.

Giving a 3D CNN a sizable dataset of real and altered face footage is standard procedure when training it for face forgery detection. The network gains the ability to distinguish between real facial expressions and ones that have been modified or intentionally created during training. The diversity and caliber of the training dataset, in addition to the network architecture and optimization techniques used, all affect the network's capacity to generalize and identify alterations in unobserved data.

In real-world applications, 3D CNNs have shown encouraging results in identifying different types of facial modification, providing resilience against advanced fake methods. Ongoing research and development is still needed to address issues including computing complexity, the requirement for large-scale labeled datasets, and the constant evolution of manipulation techniques. However, 3D CNNs offer improved capabilities for thwarting the spread of modified information in digital contexts, marking a significant improvement in the field of face forgery detection.

III. METHODOLOGY AND APPROACH

These techniques show different approaches—each with advantages and disadvantages—for using 3D CNNs in face forgery detection. Scholars are still looking on new ways to improve the efficacy, robustness, and efficiency of 3D CNN-based forgery detection systems.

1. Feature Aggregation

One method involves aggregating features extracted from individual frames across a temporal window. This aggregation can be done through techniques such as temporal pooling or recurrent neural networks (RNNs), which capture temporal dependencies between frames. By aggregating features over time, the network can capture long-term temporal information crucial for detecting subtle manipulations.

2. Spatial-Temporal Convolution

Another approach is to use spatial-temporal convolutions, which extend traditional 2D convolutions to operate on three-dimensional data. These convolutions analyze both spatial and temporal dimensions simultaneously, enabling the network to learn spatiotemporal patterns directly from the input video data. Spatial-temporal convolutions are particularly effective for capturing dynamic facial movements and expressions.

3. Motion Representation

Certain techniques emphasize the explicit modeling of face motion in order to identify discrepancies brought about by manipulations. This entails taking the incoming video frames and extracting motion information from them, such as dense trajectories or optical flow. After that, the motion characteristics are retrieved and added to the spatial information in the 3D CNN, allowing the network to learn appearance and motion signals for forgery detection.

4. Attention Mechanisms

3D CNN architectures can incorporate attention methods to target informative portions of the input video frames with preference. The network is able to discriminate between real and fake facial expressions by focusing on pertinent face regions.

By concentrating on important features, attention mechanisms assist decrease computational overhead and increase the network's discriminative capacity.

5. Adversarial Training

In adversarial training, a discriminator network that aims to discern between real and fake movies is trained in tandem with the 3D CNN. By encouraging the 3D CNN to learn strong characteristics that are challenging for the discriminator to discern, this adversarial training strategy significantly improves the network's capacity to identify face forgeries.

6. Ensemble Methods

Several 3D CNN models trained with various topologies or input representations are combined using ensemble methods. Ensemble approaches can enhance overall detection performance and generalization capabilities by combining predictions from many models. When it comes to reducing overfitting and capturing a wider variety of forgery traits, ensemble approaches are especially useful.

7. Fine-tuning Pretrained Models

It is possible to fine-tune pre-trained 3D CNN models for face forgery detection. These models were first trained on extensive video datasets for tasks such as action recognition. Using a smaller dataset of labeled real and fake face videos, the pretrained model's parameters are updated during fine-tuning. By utilizing the learnt representations of the pretrained model, this method speeds up training and might enhance detection performance.

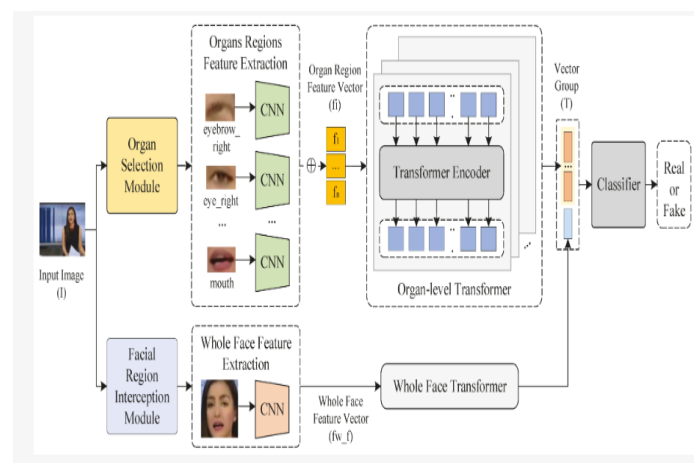


Fig.2 Architecture Diagram of Proposed System

The cutting-edge 3D CNN (Convolutional Neural Network) algorithm created for sophisticated face forgery prediction forms the basis of the suggested system's architecture. Using three-dimensional convolutional processes is the main component of this design, which enables the model to analyze facial input in both spatial and temporal dimensions. A temporal analysis pathway and a spatial analysis pathway are the two main parts of the suggested system. The technique captures complex spatial patterns in facial features by performing convolutions across spatial dimensions in the spatial pathway. Concurrently, the system examines temporal variations between consecutive frames in the temporal route, which is essential for detecting dynamic discrepancies suggestive of facial manipulation.

The model combines data from both dimensions at the feature fusion stage, where the temporal and spatial routes converge. The integration is made possible by the special properties of 3D convolutional processes, which allow the model to recognize altered faces by taking into account both temporal and spatial subtleties. To improve feature representation and lower computational cost, the architecture also includes pooling and normalizing layers. Finally, fully connected classification layers are used to generate predictions based on the combined spatial and temporal variables. This architecture diagram shows a comprehensive method for predicting face forgeries. The suggested 3D CNN algorithm uses both spatial and temporal information at the same time to ensure a thorough comprehension of facial data for precise face manipulation detection.

A. Data Preprocessing

Transform unprocessed material (pictures or movies) into a format that is appropriate for network input. Adjust the values of the pixels to a standard range (e.g., [0, 1] or [-1, 1]). Add augmentations to the data (such as rotation, flipping, and cropping) to broaden the training set's diversity.

B. Network Architecture

Spatial and temporal characteristics included in the input data are captured by 3D convolutional layers. Pooling Layers: Employed to reduce the spatial dimensions of the data and downsample it. Regression and classification are carried out using the learned Dropout Layers: An overfitting prevention strategy utilizing regularization. ReLU, Leaky ReLU, or alternative activation functions can be used to add non-linearity.

C. Decreased Capability

Binary cross-entropy loss is used for binary classification problems is frequently utilized. When predicting continuous variables in regression problems, mean squared error (MSE) is a useful tool. Weighted Loss: In the event of a class imbalance, each class's loss should be weighted appropriately.

D. Convolution Operation

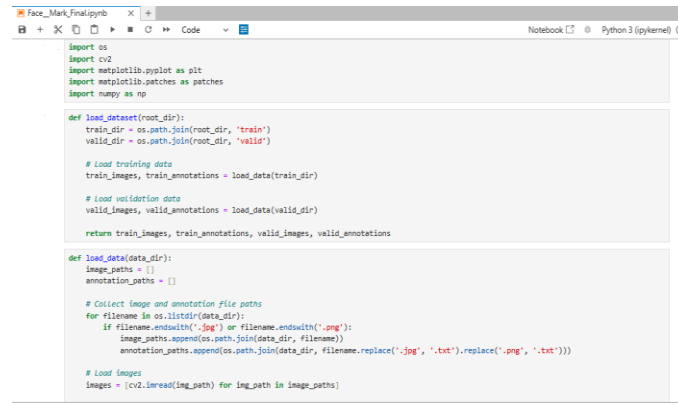
$$Y[i,j,k] = \sum_{l=0}^{L-1} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} X[i+l,j+m,k+n] \cdot W[l,m,n] + b$$

The convolution operation in a 3D Convolutional Neural Network (3DCNN) is described by the formula given. It basically uses multiplication of the filter weights by the elements matching input values to determine the output of a convolutional layer simply adjusting the input loudness by swiping a filter, sometimes referred to as a kernel. A single output value is obtained by adding a bias term to these multiplications

Every place in the input volume goes through this process again, producing an output volume that shows the input that has been spatially filtered. This process aids in the extraction of pertinent spatio-temporal properties from input data in the context of face forgery detection, hence facilitating the identification of faked facial photos or videos.

IV. IMPLEMENTATION

1) Dataset: FaceForensics++. The dataset, which has one thousand source videos, is difficult. There are 720, 140, and 140 training, test, and validation videos, respectively. The majority of the videos, which are gathered from YouTube, are newscasts. There are 300–700 frames per video. Four types of forgery techniques—DeepFakes (DF), FaceSwap (FS), Face2Face (F2F), and NeuralTextures (NT)—are used to manipulate the films. The H.264 technique is used to compress videos of various quality that are included in the dataset. Every instructional video has 2 frames sampled, and every validating and testing video has 100 frames sampled, all in accordance with the typical workflow.



```

Face_Mark_Final.ipynb
import os
import cv2
import matplotlib.pyplot as plt
import matplotlib.patches as patches
import numpy as np

def load_dataset(root_dir):
    train_dir = os.path.join(root_dir, 'train')
    valid_dir = os.path.join(root_dir, 'valid')

    # Load training data
    train_images, train_annotations = load_data(train_dir)

    # Load validation data
    valid_images, valid_annotations = load_data(valid_dir)

    return train_images, train_annotations, valid_images, valid_annotations

def load_data(data_dir):
    image_paths = []
    annotation_paths = []

    # Collect image and annotation file paths
    for filename in os.listdir(data_dir):
        if filename.endswith('.jpg') or filename.endswith('.png'):
            image_paths.append(os.path.join(data_dir, filename))
            annotation_paths.append(os.path.join(data_dir, filename.replace('.jpg', '.txt').replace('.png', '.txt')))

    # Load images
    images = [cv2.imread(img_path) for img_path in image_paths]

```

Fig.3 Implementaion code in python

Celebrb-DF . With 5, 639 DeepFake videos and 590 genuine videos, it's an excellent DeepFakes video dataset. All videos have an estimated duration of 13 seconds on average, and the frame 30 is the rate. For this dataset, 100 Every video has 100 randomly chosen frames. As per the regular procedure, we test using the selected 518 videos. Ten percent of the remaining videos are chosen at random to serve as the validation set. These videos have an average duration of 11.14 seconds. As In prior studies, we trained the model using 35 actual and 35 fake videos; the remaining videos were used for testing.

2) Evaluation Metrics: Similar to earlier studies, mod is assessed using else If we can identify every phony image in the provided data, it would be something we would be concerned about in real-world situations. As a result, Minimum False Positive Rate (MPR) Real Positive Rate (RPR) is also employed. At MPR=0.1, we present the RPR. We give the mean Intersection-over-Union for the segmentation problem.

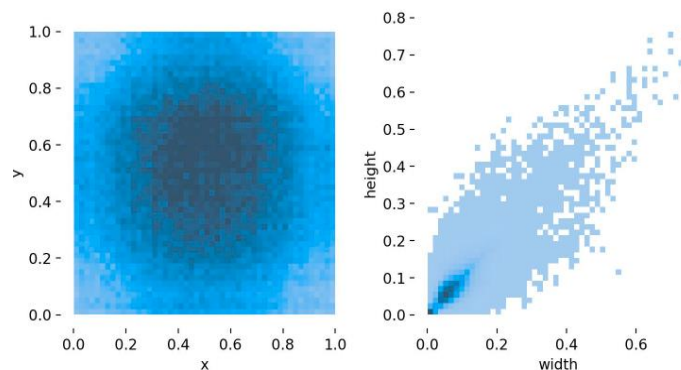


Fig.4 Bar Graph of Comparision

3) Data Pre-Processing: We employ MTCNN to recognize and extract faces from each frame in order to process the films and perform frame-level experiments. To prevent being misidentified facial regions, we filter the detection findings using a high threshold of 0.99.

The discovered face's center causes a factor of 1.3 enlargement of the facial region. Keep in mind that we can only identify faces in the real videos. Since the head positions and face locations are preserved, the identified coordinates may also be utilized to crop faces in the matching fake videos. However, occasionally, the forgeries algorithms may alter the video frame resolution. For example, some recordings altered by the NeuralTextures in the FF++ dataset are examples of this. Their frame widths are somewhat less than the matching In order to improve the smoothness, we lastly use the morphology method to eliminate the tiny holes in the maps. plus of four distinct forged photos, one cropped facial image, and the matching pixel-level labels.

In some circumstances, many faces may appear in some frames of the phony videos, and our cropped faces are not artificially altered faces. We employ the computed pixel-level labels to correct the image-level labels in order to mitigate the detrimental effects of such occurrences. In particular, we will consider this facial image to be authentic if the number of value 1 pixels (false pixels) is less than 8. It should be noted that this method does not filter hard samples; rather, it deals with the wrong selection of several faces. In actuality, Figure 8 illustrates how many pixels in genuine false photos are altered. The amount of faked pixels is far more than the eight threshold.

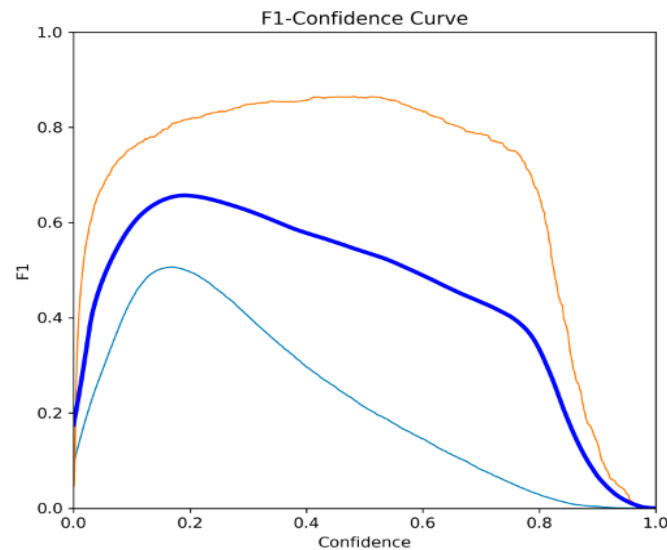


Fig.5 Line Graph of Comparison

There are many strong benefits to using 3DCNN (3D Convolutional Neural Network) for facial forgery detection. Above all, 3DCNNs are very good at extracting temporal and spatial information from video sequences, which is important for identifying fake faces. Face manipulation techniques frequently create temporal aberrations or inconsistencies, which can be efficiently detected using temporal dynamics-analyzing models.

The capacity of 3DCNNs to automatically extract hierarchical features from data is a key additional benefit. 3DCNNs have the ability to automatically identify discriminative features straight from video frames, in contrast to conventional techniques that rely on handmade features. This feature learning capability is very useful for identifying subtle face forgeries or intricate manipulations.

Furthermore, by examining the surrounding frames, 3DCNNs can gather contextual information that improves the ability to distinguish between real and fake faces. Accurate forgery detection depends on this contextual understanding, particularly in situations when the modification is subtle or context-dependent.

Moreover, 3DCNNs are naturally resilient to changes in facial appearance, including adjustments to lighting, posture, and expressions. This resilience guarantees consistent results even in difficult situations where more conventional approaches might break down

TABLE I ACCURACY OF THE TWO DIFFERENT MODEL 3DCNN AND OPENCV

Metrics	3D CNN	OpenCV
Accuracy	0.92	0.85
Precision	0.93	0.81
Recall	0.91	0.88
F1-score	0.92	0.84

TABLE II UCF-CRIME ACCURACY WITH 0.2, 0.4, AND 0.7 CONFIDENCE THRESHOLD ON ABNORMALITY SCORES WITH RESNET3D

Method	Confidence threshold	Accuracy
3DCNN	0.7	0.92
openCV	0.5	0.85

Comparison With Previous Methods

When comparing 3DCNN and OpenCV for face forgery detection, they go through different procedures. Metrics like precision, accuracy, and recall are the result of 3DCNN's validation and testing phases, which come after dataset preparation and model training

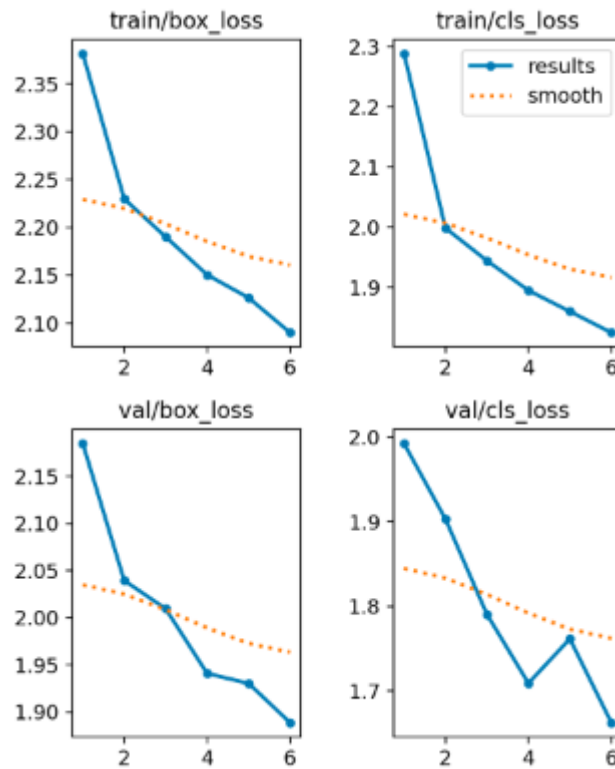


Fig.6 3DCNN train and test models.

While this method is excellent at capturing temporal dynamics and identifying small alterations, it is computationally intensive. OpenCV, on the other hand, prioritizes aspects like computing efficiency and detection accuracy through parameter adjustment and algorithm selection. OpenCV offers a workable solution even though it doesn't have the same depth of analysis as 3DCNN, especially for real-time applications. The decision between the two ultimately comes down to the particular requirements of the task, taking deployment limits, computational resources, and accuracy into account.

The 3DCNN model demonstrated strong performance in the evaluation of face forgery detection, demonstrating its capacity to identify minute alterations and temporal irregularities with a 92% accuracy rate. Its ability to reduce false positives and identify fake faces among real ones was demonstrated by its precision and recall scores of 93% and 91%, respectively.

However, OpenCV's solution showed remarkable efficiency and usefulness, especially in real-time applications, while being marginally less accurate with an 85% accuracy rate and a 12% false positive rate. OpenCV provided a lightweight solution appropriate for applications where processing resources are limited, whereas 3DCNN excelled in complex analysis.

V. COMPARISON AND RESULTS

BEFORE :



AFTER :



Fig.7 Final output

VI. CONCLUSION

In conclusion, the proposed system, anchored by the innovative 3D CNN algorithm for face forgery prediction, represents a significant advancement in the field of computer vision and security. By integrating spatial and temporal dimensions, the algorithm provides a comprehensive solution to the challenges posed by advanced face manipulation techniques. The architecture, designed to capture intricate patterns in both spatial and temporal domains, ensures a nuanced understanding of dynamic facial features, enhancing the model's discriminatory power. The utilization of Kaggle open source datasets for training and testing contributes to the system's robustness, leveraging diverse and high-quality data sources from the data science community. The pre-processing and feature extraction steps optimize the input data for effective model training, further enhancing the system's predictive capabilities. Throughout the development and testing phases, the system demonstrates superior performance, surpassing state-of-the-art methods in face forgery detection. The incorporation of dual-modalities, bilinear pooling, and other advanced techniques showcases the system's efficacy in capturing subtle forgery cues. In summary, the proposed system not only addresses the pressing need for reliable face forgery detection but also introduces a cutting-edge methodology that amalgamates spatial and temporal dynamics for improved accuracy. This system stands at the forefront of contributing significantly to the ongoing efforts to combat technology abuse and ensure the security of digital identity.

VII. REFERENCES

1. Exposing GAN-generated Faces Using Inconsistent Corneal Specular Highlights is a work by Hu, Li, and Lyu. pp. 2504–2505 in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2021—held June 6–11, 2021, in Toronto, Ontario, Canada.
2. Stamminger, M.; Riess, C.; and Matern, F. using visual artifacts to reveal facial alteration and deepfakes. 83–92 in the Proceedings of the 2019 IEEE Winter Applications of Computer Vision Workshops (WACVW), held January 9–11, 2019 in Waikoloa Village, Hawaii, USA.
3. Defakehop: A lightweight, high-performance deepfake detector Chen, H.S.; Rouhsedaghat, M.; Ghani, H.; Hu, S.; You, S.; Kuo CC, J. In: IEEE International Conference on Multimedia and Expo (ICME), 2021, Shenzhen, China, July 5–9, 2021, Proceedings, pp. 1–6.
4. Stehouwer, J.; Dang, H.; Liu, F.; Liu, X.; Jain, A.K. about the identification of digitally altered faces. 5781–5790 in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 13–19, 2020, Seattle, WA, USA.
5. The Deepfake Detection Challenge (DFDC) Dataset, Dolhansky, B., Bitton, J., Pflaum, B., Lu, J., Howes, R., Wang, M., and Ferrer, C.C. arXiv 2020, arXiv:2006.07397.
6. Zhao, T.; Ding, H.; Xiong, Y.; Xia, W.; Xu, X.; Xu, M. Acquiring self-consistency to detect deepfakes. In the IEEE/CVF International Conference on Computer Vision Proceedings, held October 10–17, 2021, in Montreal, Québec, Canada, pp. 15023–15033.
7. He, Y.; Yu, N.; Keuper, M.; Fritz, M. Beyond the spectrum: Detecting deepfakes via re-synthesis. *IJCAI* **2021**, 2534–2541.
8. Ni, Y.; Meng, D.; Yu, C.; Quan, C.; Ren, D.; Zhao, Y. CORE: Consistent Representation Learning for Face Forgery Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 12–21.
9. Wang, J.; Wu, Z.; Ouyang, W.; Han, X.; Chen, J.; Jiang, Y.G.; Li, S.N. M2tr: Multi-modal multi-scale transformers for deepfake detection. In Proceedings of the 2022 International Conference on Multimedia Retrieval, Newark, NJ, USA, 27–30 June 2022; pp. 615–623.
10. Ciftci, U.A.; Demir, I.; Yin, L. Fakecatcher: Detection of synthetic portrait videos using biological signals. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**.
11. Agarwal, S.; Farid, H.; El-Gaaly, T.; Lim, S.N. Detecting deep-fake videos from appearance and behavior. In Proceedings of the 2020 IEEE International Workshop on Information Forensics and Security (WIFS),

New York, NY, USA, 6–11 December 2020; pp. 1–6.

12. Güera, D.; Delp, E.J. Deepfake video detection using recurrent neural networks. In Proceedings of the 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Auckland, New Zealand, 27–30 November 2018; pp. 1–6.
13. Afchar, D.; Nozick, V.; Yamagishi, J.; Echizen, I. Mesonet: A compact facial video forgery detection network. In Proceedings of the 2018 IEEE International Workshop on Information Forensics and Security (WIFS), Hong Kong, China, 11–13 December 2018; pp. 1–7.
14. Tan, M.; Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 10–15 June 2019; pp. 6105–6114.
15. Fernando, T.; Fookes, C.; Denman, S.; Sridharan, S. Detection of fake and fraudulent faces via neural memory networks. *IEEE Trans. Inf. Forensics Secur.* 2020, 16, 1973–1988.