International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 3 (2026)

ijerst.editor@gmail.com

editor@ijerst.com

ISSN 2319-5991 www.ijerst.com

Peer Reviewed, Referred & Indexed Journal

Research Paper

DEEPFAKE FACE DETECTION IN VIDEOS USING OPENCV AND MOBILENETV2

A. V. Mohitha[1], Dr. C. Berin Jones[2]
[1]PG Scholar, Department of CSE, Shadan Women's College of Engineering and Technology, Hyderabad, mohitha1302@gmail.com
[2]Professor, Department of CSE, Shadan Women's College of Engineering and Technology jonesberin@gmail.com

ABSTRACT

The broad dissemination of altered facial photographs, especially Deepfakes, which are getting harder to identify with traditional techniques, is made possible by the Internet's quick development. While existing methods concentrate on intricate network architectures or geographical domain properties, they sometimes lack resilience against advanced counterfeit techniques. In order to overcome this, we suggest a Deepfake detection framework based on MobileNetV2, which uses effective convolutional feature extraction to accurately classify real and fake facial photos. In order to guarantee consistent input quality and improve the discriminative features for detection, the framework starts with OpenCV-based preprocessing, which includes face detection, alignment, and normalisation. By automatically learning hierarchical spatial features from the pre-processed facial photos, MobileNetV2, a lightweight yet powerful convolutional neural network, replaces the requirement for manually created features.

1. INTRODUCTION

Deepfakes, or extremely realistic synthetic content, are a result of the development of advanced artificial intelligence and deep learning technologies, which have revolutionised the creation of digital media. Deepfakes are altered facial photos or videos made by deep generative models, such autoencoders and GANs, that can accurately mimic or change human features. Although this technology can be used for creative, educational, and entertainment purposes, it has also presented serious risks to digital media's security, privacy, and credibility. Global concerns have been raised by the malicious use of Deepfakes in identity theft, political manipulation, and the dissemination of false information, making automated detection methods an important topic of study. Because deep learning-based techniques can automatically extract discriminative features, they have become effective alternatives to these restrictions. In this work, we present a Deepfake detection system based on the lightweight and effective convolutional neural network architecture known as MobileNetV2. Face detection, alignment, and normalisation are among the OpenCV-based preprocessing methods that support the system and guarantee consistency and enhance feature quality. The hierarchical spatial features that capture texture inconsistencies and visual artefacts caused during manipulation are then learned using MobileNetV2. This combination allows for cross-dataset generalisation by improving resilience and lowering reliance on manual feature engineering. MobileNetV2's lightweight design guarantees quicker training and deployment, which makes it appropriate for real-time applications. The suggested system solves scalability issues in addition to accurately classifying actual and false photos. This work helps create trustworthy Deepfake detection algorithms by striking a compromise between efficiency and accuracy. The concept may find use in digital forensics, social media monitoring, media verification, and security. The study's ultimate goals are to reduce the negative effects of malevolent Deepfake use and rebuild confidence in digital media.

PROBLEM SYSTEM

The increasing use of Deepfake technology has made it difficult to distinguish between real and manipulated facial images. Existing detection methods often lack accuracy and robustness against advanced forgeries. Therefore, there is a need for an efficient and lightweight Deepfake detection system that can accurately classify images as Real or Fake using OpenCV preprocessing and the MobileNetV2 deep learning model.

2. EXISTING SYSTEM

Convolutional Neural Networks (CNNs) are the main tool used by current Deepfake detection systems for spatial domain analysis; hybrid designs with transformers are occasionally used. These models look for irregularities or artefacts in facial regions, like mismatched textures, distorted facial expressions, or abnormal blinking. While some employ attention processes or multi-scale feature extraction, others try to improve feature extraction by transforming images into frequency or residual domains. Some additionally investigate anomalies in image quality and information

to enhance classification performance. However, when Deepfake creation techniques advance and yield more photorealistic content, these current systems will confront difficulties. Spatial-only techniques frequently experience overfitting and may not be able to generalise across various Deepfake datasets.

3. RELATED WORK

The relevant work in the subject of face de-false Face Detection Based on Fusion of Spatial Texture and High-Frequency Noise 213 detection is presented in this part, along with some contemporary methods for identifying manipulated or false faces.

By comparing the frequency of eye blinking in forged videos to the typical frequency in authentic videos, Li et al. [2] identified fabricated faces using targeted forgery detection techniques. Matern et al. [8] used the geometric structure of the face and the reflected features in the eyes to identify fake faces. These detection techniques may be less successful against unidentified manipulation techniques, but they perform well against recognised forgery techniques. FaceForensics++ (FF++), the largest video-level benchmark test set, was published by Rössler et al. [9].

4. PROPOSED SYSTEM

The suggested solution combines MobileNetV2-based classification with OpenCV preprocessing to automatically identify Deepfake facial images. In order to ensure that all facial images are standardised and reliably positioned for feature extraction, the pipeline begins with OpenCV for face detection, alignment, and normalisation. By improving the quality of the incoming data, this preprocessing stage enables the algorithm to concentrate on minute discrepancies brought about by manipulation. The system uses MobileNetV2, a lightweight convolutional neural network, to extract hierarchical spatial information from facial photos after preprocessing. While pooling layers emphasise important discriminative features and minimise dimensionality, convolutional layers capture local patterns and texture abnormalities. (n).

5. MODULES

1. Assembling the Dataset: The main goal of this module is to compile a varied dataset of authentic and deepfake photos or videos from publically accessible sources. To improve model resilience, the dataset has to contain variations in background settings, lighting, and facial expressions. To guarantee balanced training, both real and deceptive media are gathered. Extra preprocessing is carried out, including frame extraction and labelling. The accuracy of the model as a whole is directly impacted by the quantity and quality of data at this stage**.**

2. Data Interpretation:

At this point, the dataset is examined to identify trends and distinctive features between authentic and fraudulent material. To find irregularities in the dataset, statistical analysis, visualisation, and metadata checks are carried out. Subtle distortions are typically introduced in deepfake footage and are brought to light during analysis. This aids in locating possible characteristics that distinguish authentic faces from those that have been altered. The knowledge acquired here directs future feature engineering and preprocessing.

3. Data Conditioning: This module ensures that the data is cleaned, normalized, and prepared for model training. Noise removal, resizing frames, and converting videos into consistent formats are carried out. Feature extraction techniques, such as facial landmarks and texture analysis, are applied for better representation. Data augmentation techniques like rotation, flipping, and brightness adjustments help in increasing dataset diversity. This stage guarantees that the input to the model is standardized and optimized for learning.

4. Model Execution: In this case, the processed dataset is used to train the selected deep learning model (such as CNN, RNN, or hybrid architectures). Frameworks like PyTorch or TensorFlow are used to run the model. The algorithm gains the ability to recognise deepfake-specific artefacts during training, such as irregularities in lip synchronisation, eye blinking, and facial blending. Dividing the dataset into training, validation, and testing sets is necessary for proper training. Additionally, model checkpoints and logging are kept up to date for tracking performance.

5. Model Calibration: In order to increase the model's accuracy and dependability, this module focuses on modifying hyperparameters. This step involves fine-tuning the learning rate, batch size, optimiser selection, and number of epochs. To prevent overfitting and underfitting, cross-validation methods are used. The model's ability to generalise well on unseen deepfake samples is ensured by calibration. Performance measurements are continuously checked during the procedure, which entails several experiments. In the end, this improves the detecting system's accuracy and durability.

6. Model Performance:

Once trained, the model is evaluated using performance metrics such as accuracy, precision, recall, F1-score, and ROC-AUC. The evaluation determines how effectively the model distinguishes between real and fake content. Confusion matrix analysis helps in identifying false

positives and false negatives. Visualization tools may be used to highlight areas where the model performs strongly and weakly. This stage ensures that the model meets predefined benchmarks before deployment.

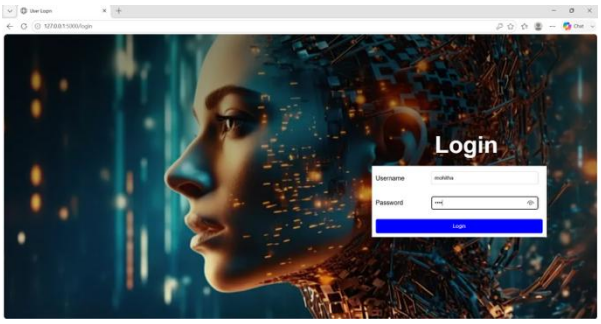
7.Outcome Prediction The trained and calibrated model is used in this last module to determine if fresh input media is authentic or fraudulent. The model produces a classification result with confidence scores after receiving video frames or photos as input. Real-time solutions for deepfake detection can use predictions. End users are informed of the result in a clear manner, maybe with areas of suspected tampering highlighted. This module ensures the project's practicality by completing the cycle.

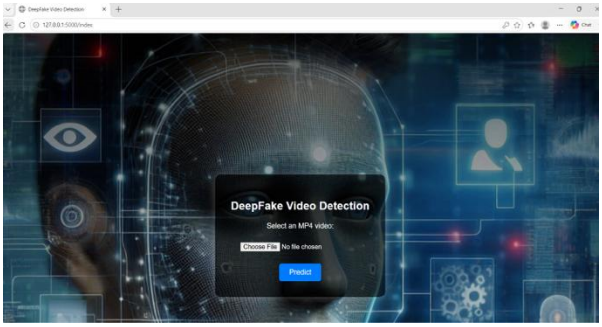
6. IMPLEMENTATION

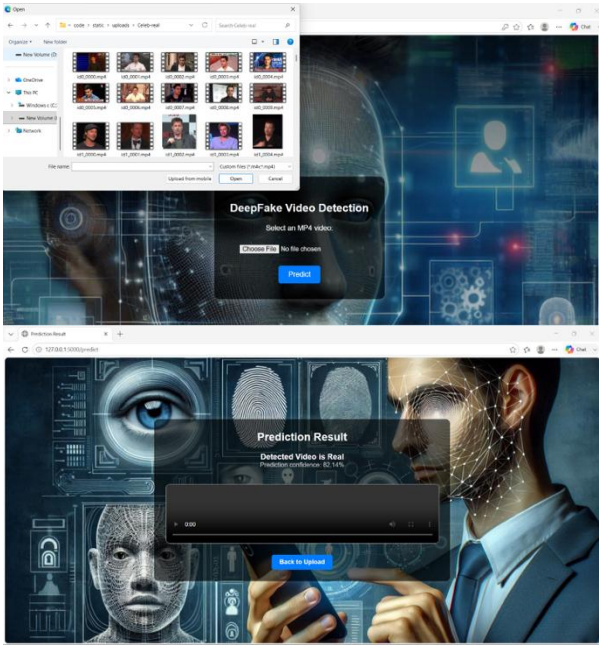
The proposed deepfake face detection system is implemented using Python, OpenCV, TensorFlow/Keras, Flask, and the MobileNetV2 deep learning architecture. Initially, a dataset consisting of real and deepfake videos is collected and labeled into two classes. Each input video is processed using OpenCV to extract individual frames, and facial regions are detected from each frame. The detected faces are then preprocessed by resizing them to 224 × 224 pixels and normalizing the pixel values to meet the input requirements of the MobileNetV2 model.

The preprocessed facial images are passed to the MobileNetV2 network, which automatically extracts discriminative facial features and classifies each frame as either real or fake. The frame-level predictions are aggregated to generate the final classification of the uploaded video. A confidence score is also produced to indicate the reliability of the prediction. To provide a user-friendly interface, the trained model is integrated into a Flask-based web application, where users can register, log in, upload a video, and obtain the deepfake detection result. This implementation enables efficient and accurate detection of manipulated facial videos while maintaining low computational complexity through the use of the lightweight MobileNetV2 architecture.

7. RESULTS AND DISCUSSION





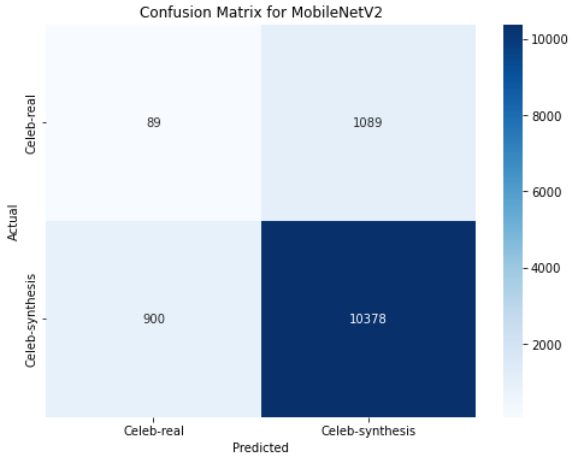


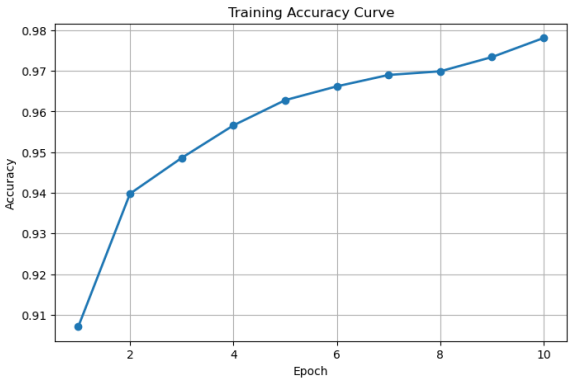
DISCUSSION:

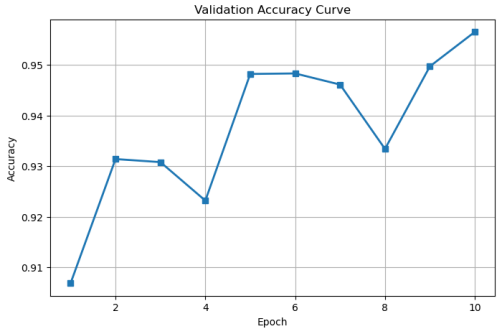
The developed deepfake face detection system was successfully implemented as a web-based application using Flask, OpenCV, and the MobileNetV2 deep learning model. The system provides a simple and user-friendly interface in which the user first registers and logs into the application. After successful authentication, the user uploads a video file for analysis. Once the video is uploaded, the system begins processing by extracting individual frames using OpenCV. From each frame, the facial region is detected and preprocessed by resizing and normalizing the image to match the input requirements of the MobileNetV2 model. The processed frames are then passed through the trained model, which extracts facial features and classifies each frame as either Real or Fake. After analysing all the extracted frames, the system combines the individual frame predictions to determine the final classification of the uploaded video. The output is displayed on the web page as "Detected Video is Real" or "Detected Video is Fake", along with the corresponding confidence score, enabling the user to understand the reliability of the prediction.

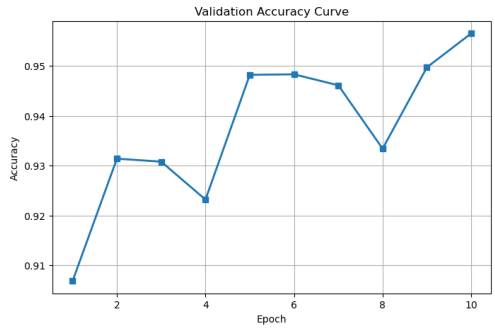
PERFORMANCE ANALYSIS:

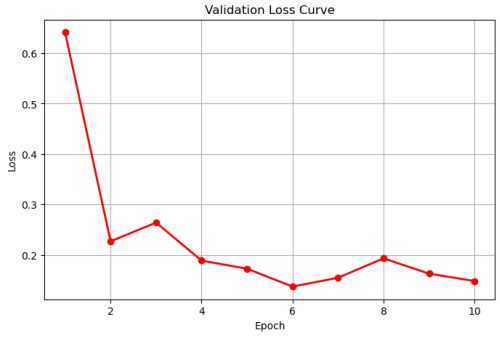
The performance evaluation of the proposed deepfake face detection system indicates that the combination of OpenCV and MobileNetV2 provides effective classification of real and manipulated facial videos. The model achieved an **accuracy of 92%**, demonstrating its ability to correctly classify the majority of input videos. Furthermore, the **precision of 92%** indicates that when the model predicts a video as fake, the prediction is highly reliable, resulting in a low number of false positive classifications. However, the **recall of 55%** suggests that the system is less effective at identifying all deepfake videos, as some manipulated samples are incorrectly classified as real. This reduction in recall may be attributed to factors such as low-resolution videos, compression artifacts, varying illumination, facial pose variations, and the presence of advanced deepfake generation techniques that were not sufficiently represented in the training dataset. Consequently, the **F1-score of 69%** reflects a moderate balance between precision and recall, indicating that while the proposed model is highly accurate and reliable in its positive predictions, there is scope for improving its capability to detect a greater number of deepfake instances. Overall, the obtained results demonstrate the effectiveness of the proposed approach for deepfake video detection while highlighting the need for future enhancements, such as training on larger and more diverse datasets, incorporating temporal information from consecutive video frames, and employing advanced feature learning techniques to improve recall without compromising the model's high accuracy and precision.

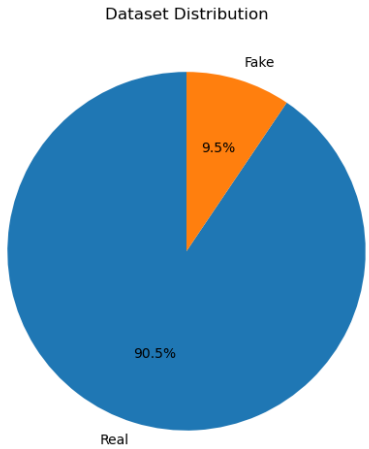


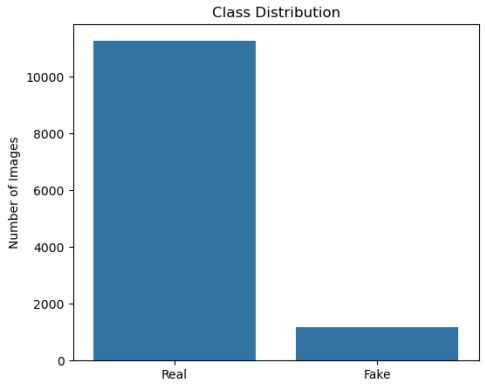


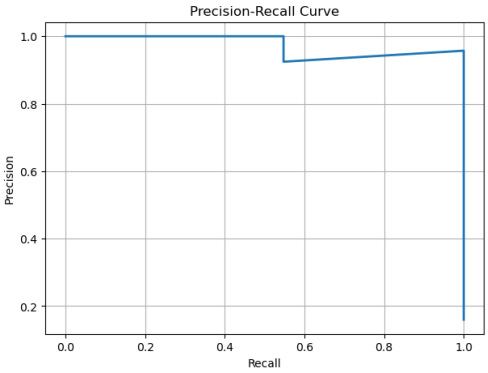

Validation Accuracy Curve
Accuracy
Epoch

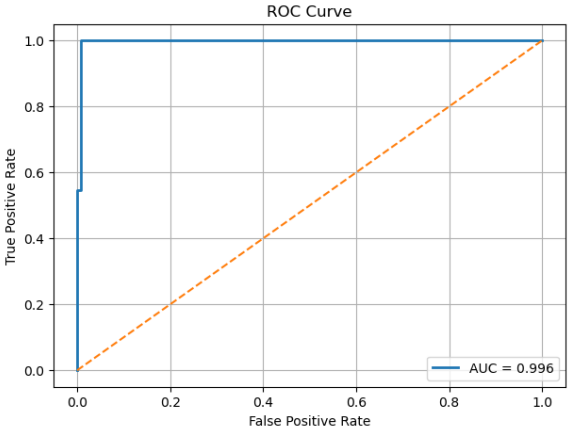

Validation Accuracy Curve
Accuracy
Epoch

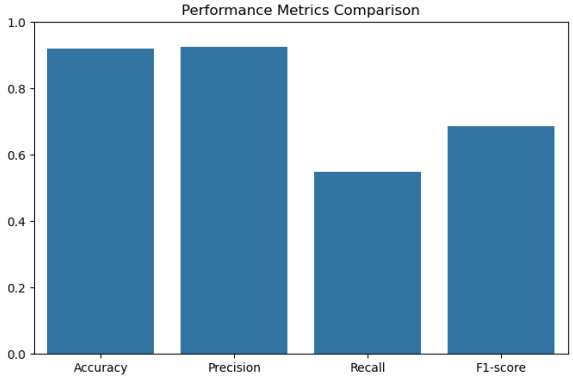

Validation Loss Curve
Loss
Epoch


Dataset Distribution
Fake
9.5%
90.5%
Real


Class Distribution
Number of Images
Real
Fake


Precision-Recall Curve
Precision
Recall


ROC Curve
True Positive Rate
False Positive Rate
AUC = 0.996


Performance Metrics Comparison
Accuracy
Precision
Recall
F1-score

8. CONCLUSION

The deepfake detection project emphasises how important it is to fight AI-generated false information in the modern digital world. Despite its revolutionary nature, deepfake technology presents significant risks to digital media security, privacy, and trust. The technology offers a strong foundation to distinguish between authentic and altered material by utilising cutting-edge machine learning and deep learning techniques. A methodical approach to model construction is guaranteed by the modular process, which includes data assembly, conditioning, execution, calibration, and performance analysis. The potential of AI-based solutions in spotting minute irregularities in manipulated media is confirmed by experimental results. The initiative helps protect people and organisations from malicious deepfake attacks. But the fight against deepfakes is dynamic, necessitating ongoing study and development. The use of updated datasets, explainable models, and real-time detection will improve Nevertheless, the fight against deepfakes is dynamic and necessitates ongoing study and development. Reliability will be further improved by combining explainable models, updated datasets, and real-time detection. In the end, our work establishes a solid basis for safeguarding digital integrity in a time of artificial intelligence. Deepfake detection has the potential to become an essential cybersecurity tool with additional improvements. It ensures safer digital communication by empowering people and building trust across online platforms.

9. FUTURE ENHANCEMENT

Multimodal analysis, which combines text, video, and audio signals for more accurate detection, can be added to the deepfake detection system in the future. Social media companies can use real-time detection systems to automatically flag or prohibit offensive content. Digital media can be made legitimate and traceable by using blockchain technology. Federated learning can be used to jointly train models while maintaining user privacy. Users will be able to comprehend the rationale behind detections thanks to advancements in explainable AI (XAI). It is also possible to create lightweight models for edge computing and mobile devices. Regulatory frameworks can be strengthened by cooperation with governmental and media institutions. The model will be able to adjust to new kinds of deepfakes with ongoing dataset updates.

10.REFERENCES

[1] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, et al., "Deepfakes and beyond: A survey of face manipulation and fake detection," Information Fusion, vol. 64, pp. 131–148, 2020.

[2] Y. Z. Li, M. C. Chang, and S. W. Lyu, "In ictu oculi: Exposing AI created fake videos by detecting eye blinking," in Proc. IEEE Int. Workshop on Information Forensics and Security (WIFS), Hong Kong, China, pp. 1–7, 2018.

[3] H. D. Li, W. Q. Luo, X. Q. Qiu, et al., "Identification of various image operations using residual-based features," IEEE Trans. Circuits Syst. Video Technol., vol. 28, no. 1, pp. 31–45, 2018.

[4] X. Wu, Z. Xie, Y. T. Gao, et al., "SSTNet: Detecting manipulated faces through spatial, steganalysis and temporal features," in Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, pp. 2952–2956, 2020.

[5] J. W. Fei, Y. S. Dai, P. P. Yu, et al., "Learning second order local anomaly for general face forgery detection," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), New Orleans, LA, USA, pp. 20238–20248, 2022.

[6] J. A. Stuchi, M. A. Angeloni, R. F. Pereira, et al., "Improving image classification with frequency domain layers for feature extraction," in Proc. IEEE Int. Workshop Mach. Learn. Signal Process. (MLSP), Tokyo, Japan, pp. 1–6, 2017.

[7] J. Fridrich and J. Kodovsky, "Rich models for steganalysis of digital images," IEEE Trans. Inf. Forensics Security, vol. 7, no. 3, pp. 868–882, 2012.

[8] F. Matern, C. Riess, and M. Stamminger, "Exploiting visual artifacts to expose deepfakes and face manipulations," in Proc. IEEE Winter Conf. Appl. Comput. Vis. Workshops (WACVW), Waikoloa, HI, USA, pp. 83–92, 2019.

[9] A. Rössler, D. Cozzolino, L. Verdoliva, et al., "Faceforensics++: Learning to detect manipulated facial images," in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Seoul, South Korea, pp. 1–11, 2019.

[10] H. Dang, F. Liu, J. Stehouwer, et al., "On the detection of digital face manipulation," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Seattle, WA, USA, pp. 5780–5789, 2020.

[11] H. Q. Zhao, T. Y. Wei, W. B. Zhou, et al., "multi-attentional deepfake detection," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Nashville, TN, USA, pp. 2185–2194, 2021.

[12] L. Chen, Y. Zhang, Y. B. Song, et al., "Self-supervised learning of adversarial example: Towards good generalizations for deepfake detection," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), New Orleans, LA, USA, pp. 18689–18698, 2022.

[13] X. Y. Dong, J. M. Bao, D. D. Chen, et al., "Protecting celebrities from deepfake with identity consistency transformer," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), New Orleans, LA, USA, pp. 9458–9468, 2022.
[14] D. Cozzolino, G. Poggi, and L. Verdoliva, "Splicebuster: A new blind image splicing detector," in Proc. IEEE Int. Workshop on Information Forensics and Security (WIFS), Rome, Italy, pp. 1–6, 2015.
[15] Y. Y. Qian, G. J. Yin, L. Sheng, et al., "Thinking in frequency: Face forgery detection by mining frequency-aware clues," in Proc. Eur. Conf. Comput. Vis. (ECCV), Glasgow, UK, pp. 86–103, 2020.
[16] Y. C. Luo, Y. Zhang, J. C. Yan, et al., "Generalizing face forgery detection with high-frequency features," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Nashville, TN, USA, pp. 16312–16321, 2021.
[17] A. Vaswani, N. Shazeer, N. Parmar, et al., "Attention is all you need," in Advances in Neural Information Processing Systems (NeurIPS), Long Beach, CA, USA, pp. 6000–6010, 2017.
[18] B. V. Salim, Chyntia, J. O. Indrawan, et al., "Face shape classification using swin transformer model," Procedia Comput. Sci., vol. 227, pp. 557–562, 2023.
[19] Y. Z. Li, X. Yang, P. Sun, et al., "Celeb-DF: A large-scale challenging dataset for DeepFake forensics," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Seattle, WA, USA, pp. 3204–3213, 2020.
[20] B. Dolhansky, J. Bitton, B. Pflaum, et al., "The DeepFake detection challenge (DFDC) dataset," arXiv preprint, arXiv:2006.07397, 2020.
[21] B. J. Zi, M. H. Chang, J. J. Chen, et al., "WildDeepfake: A challenging real-world dataset for deepfake detection," in Proc. ACM Int. Conf. Multimedia, Seattle, WA, USA, pp. 2382–2390, 2020.
[22] M. X. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. Int. Conf. Mach. Learn. (ICML), Long Beach, CA, USA, pp. 6105–6114, 2019.
[23] J. Deng, W. Dong, R. Socher, et al., "ImageNet: A large-scale hierarchical image database," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Miami, FL, USA, pp. 248–255, 2009.
[24] K. Simonyan, A. Vedaldi, and A. Zisserman, "Visual explanations from deep networks via gradient-based localization," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Venice, Italy, pp. 618–626, 2017.
[25] D. Afchar, V. Nozick, J. Yamagishi, et al., "MesoNet: A compact facial video forgery detection network," in Proc. IEEE Int. Workshop on Information Forensics and Security (WIFS), Hong Kong, China, pp. 1–7, 2018.
[26] Z. Z. Liu, X. J. Qi, and P. H. S. Torr, "Global texture enhancement for fake face detection in the wild," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Seattle, WA, USA, pp. 8057–8066, 2020.
[27] C. R. Wang and W. H. Deng, "Representative forgery mining for fake face detection," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Nashville, TN, USA, pp. 14918–14927, 2021.
[28] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Honolulu, HI, USA, pp. 1800–1807, 2017.
[29] J. K. Wang, Z. X. Wu, W. H. Ouyang, et al., "M2TR: Multimodal multi-scale transformers for deepfake detection," in Proc. Int. Conf. Multimedia Retrieval (ICMR), Newark, NJ, USA, pp. 615–623, 2022.
[30] Z. Q. Guo, G. B. Yang, D. Y. Zhang, et al., "Rethinking gradient operator for exposing AI-enabled face forgeries," Expert Syst. Appl., vol. 215, article 119361, 2023