IJARIIT INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact Factor: 6.078
(Volume 9, Issue 6 - V9I6-1183)
Available online at: https://www.ijariit.com

Improvised approach to Deepfake detection

Lipika Chadha
lipika.chadha2020@vitstudent.ac.in
Vellore Institute of Technology, Vellore, Tamil Nadu

Hiya Kulasrestha
kulasresthahiya@gmail.com
Vellore Institute of Technology, Vellore, Tamil Nadu

Vishesh Bhargava
bhargavavishesh0811@gmail.com
Vellore Institute of Technology, Vellore, Tamil Nadu

Varun Jindal
Varunj4916@gmail.com
Vellore Institute of Technology, Vellore, Tamil Nadu

ABSTRACT

Deepfakes are realistic, human-synthesized videos that are incredibly simple to produce thanks to the advent of sophisticated algorithms fueled by advances in the field of deep learning. Deepfakes are being used to create fictitious news stories about terrorism, politics, and retaliation to incite societal unrest. The development of efficient techniques for identifying deepfakes is imperative, given the mounting concerns surrounding them. Our work in this field offers a brand-new deep learning-based method that effectively distinguishes between authentic videos and artificial intelligence-generated phony ones. Our technique can identify deepfakes that are both reenactments and replacements. To counter the threat posed by artificial intelligence (AI), we suggest a system that makes use of AI. To train an InceptionV3 model to categorize films as either real or manipulated, depending on whether they have undergone any kind of alteration, this method uses an MTCNN neural network to extract frame-level information. We assess our method on a large-scale balanced and mixed data set to mimic real-time scenarios **and improve the model's performance on real-time data. This dataset was painstakingly created by combining multiple available datasets**

Keywords: *MTCNN, CNN, InceptionV3, CLAHE, Deepfake Detection, Facial Embeddings*

I. INTRODUCTION

Within the field of artificial intelligence, deepfakes have become a powerful instrument that can produce lifelike audio or video clips of people acting or saying things they never would. These fakes are created by overlaying the voice or face of one individual over the body or voice of another. Applications for deepfakes could be found in satire, education, and entertainment, among other fields. They do, however, also provide serious risks, like spreading misleading information, creating pornography as retaliation, and damaging reputations. In this research, we successfully address these issues with a novel deepfake detection method. Our method consists of training a CNN classifier to differentiate between real and false photos, extracting face embeddings using MTCNN, improving the brightness of input images using the CLAHE technique, and training an InceptionV3 model to learn discriminative facial characteristics. The CLAHE algorithm is essential in bringing face image illumination back to normal, which lessens the effect of different lighting conditions on the execution of the next steps. Regardless of the surrounding lighting, our algorithm guarantees the accurate and consistent extraction of facial features. Using input photos, MTCNN, a multi-task convolutional neural network, accurately recognizes and extracts facial areas. MTCNN greatly minimizes computational overhead and removes the possibility of including extraneous background factors that can impede the detection process by concentrating only on the pertinent face features.

A vast dataset of actual and fake facial photos is used to train the InceptionV3 model, a deep neural network architecture. The model can acquire high-dimensional facial embeddings through this training procedure, which effectively captures the subtle variations between real and artificial faces. The model can decide whether an image is legitimate by using these embeddings, which capture the distinctive qualities of facial features. Ultimately, the retrieved facial embeddings are used to train a CNN classifier, which helps it distinguish between real and fraudulent photos. The CNN architecture makes good use of the discriminative information included in the facial embeddings. CNN is well known for its capacity to recognize patterns and extract complex features. Using the learned embeddings as training data, our classifier correctly classifies fresh images as either real or false.

II. LITERATURE REVIEW

A. Improving Facial Recognition of FaceNet in a small dataset using DeepFakes - A. A. Balde, A. Jain, and D. Patra - 2020 4th International Conference on Computer, Communication and Signal Processing (ICCCSP)

The paper delves into enhancing FaceNet's facial recognition on smaller datasets using DeepFakes. It highlights DeepFakes' capability to manipulate images and videos through AI, offering a means to augment small datasets for FaceNet training without additional data. Transfer learning emerges as a feasible technique for improving FaceNet's performance on limited datasets, leveraging pre-trained models on larger datasets. However, challenges arise due to facial variations caused by factors like age, pose, and lighting, impacting FaceNet's learning process. While computationally intensive, generating DeepFakes for augmentation may yield less realistic images, potentially affecting FaceNet's accuracy. The study acknowledges the simplicity of implementing transfer learning but stresses the need to address computational expenses and image realism to ensure its effectiveness in refining FaceNet's performance on smaller datasets.

B. Detecting Deepfakes with Metric Learning - A. Kumar, A. Bhavsar and R. Verma - 2020 8th International Workshop on Biometrics and Forensics (IWBF)

The paper investigates detecting Deepfakes using metric learning, employing triplet loss to establish a distance metric between real and fake videos. This approach proves effective even under high compression, showcasing robustness against variations in lighting and pose. Face alignment plays a crucial role in mitigating pose and lighting variations for accurate detection. Feature extraction from faces, bodies, or entire videos contributes to the classification process, which categorizes videos as real or fake using machine learning algorithms like support vector machines, random forests, or deep neural networks. While robust and straightforward to implement, this method demands substantial training data for optimal performance and may not be as effective with uncompressed Deepfakes. Overall, its efficacy in detecting compressed Deepfakes and resilience to lighting and pose variations mark its strengths, albeit with the requirement of significant training data.

C. Deepfakes Detection with Automatic Face Weighting -Daniel Mas Montserrat, Hanxiang Hao, S. K. Yarlagadda, Sri- ram Baireddy, Ruiting Shao Janos Horv´ath, Emily Bartusiak, Justin Yang, David G´uera, Fengqing Zhu, Edward J. Delp -2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)

The paper introduces an innovative approach, employing automatic facial weighting for improved deepfake detection. Recognizing the limitations of conventional methods in handling facial changes, the authors propose autonomously as-signing weights to facial components crucial for detection, such as structure, skin quality, and eye movements. These weights, determined by their significance in detecting deep fakes, enhance the system's ability to handle facial alterations effectively, resulting in higher accuracy in discerning real from fake videos. However, this method demands significant computational resources for training the machine learning algorithm on a substantial dataset of real and fake videos. Additionally, while the weighted features bolster detection accuracy, the system's vulnerability to adversarial attacks remains a concern, indicating the need for further security enhancements. Overall, the approach's strength lies in its capacity to adaptively assign weights to crucial facial features, leading to improved accuracy in identifying deepfake content, albeit at the cost of computational intensity and vulnerability to certain threats like adversarial attacks.

D. DeepFake Detection by Analyzing Convolutional Traces –
L. Guarnera, O. Giudice, S. Battiato – 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)

The paper introduces a novel approach for detecting deep- fakes through the analysis of convolutional traces in videos. By scrutinizing the activations of convolutional layers within deep neural networks, the method distinguishes between genuine and manipulated videos. Deepfake videos, created by superimposing one person's face onto another, produce distinct convolutional traces due to artifacts generated during the face-warping process. This technique extracts and utilizes these traces to train a classifier to discern between real and fake content. Notably, it excels in handling facial alterations like occlusions and pose changes, as these don't significantly affect convolutional traces. Its accuracy in differentiating between real and fake videos surpasses traditional methods. However, this method demands substantial computational resources to train deep neural networks on extensive datasets of genuine and fake videos. Moreover, while effective, its susceptibility to adversarial attacks highlights the need for further security measures. Overall, the approach's strength lies in its ability to leverage convolutional traces for accurate deepfake detection, despite the computational intensity and vulnerability to specific attack vectors.

E. A Novel Deep Learning Approach for Deepfake Image Detection - Ali Raza, Kashif Munir and Mubarak Almutairi
- Applied Sciences (2022)

The paper introduces a novel approach for deepfake image detection, fusing the VGG16 model with a convolutional neural network (CNN) architecture. Leveraging transfer learning, the hybrid model capitalizes on VGG16's pre-trained features from ImageNet, enhancing feature extraction for deepfake identification. This method demonstrates efficiency and optimization in processing data, utilizing a deepfake dataset from Yonsei University's Department of Computer Science, accessible via Kaggle. Hyperparameter tuning further refines model performance, boosting accuracy. However, VGG16's resource-intensive nature poses challenges for real-time deepfake de-detection due to its computational demands. Additionally, its extensive parameters might not suit smaller datasets, potentially leading to overfitting. Despite these limitations, the fusion of VGG16 and CNN offers significant advantages in leveraging pre-trained features for effective deepfake image detection, albeit with considerations for computational resources and dataset sizes.

F. Deepfakes Generation and Detection: A Short Survey –Zahid Akhtar - J. Imaging (2023)

The paper offers a comprehensive overview of deepfake generation and detection methods, delving into various ma- manipulation categories like identity swap and attribute alteration. It thoroughly reviews multiple approaches employed in both creating and identifying deepfakes, providing readers with a broad understanding of the subject's complexities. However, it falls short of deeply discussing specific methodologies, potentially leaving readers with only a surface-level grasp of certain concepts. Furthermore, it lacks a thorough comparative analysis of the effectiveness of different generation and detection techniques, limiting its applicability as prescriptive guidance. Despite these limitations, the paper sheds light on the ongoing battle between creators and detectors of deepfakes, emphasizing the intricacies of the field and advocating for further research and innovation in this evolving landscape.

G. Combining Deep Learning and Super-Resolution Algorithms for Deep Fake Detection - Nikita S. Ivanov, Anton V. Arzhskov, Vitaliy G. Ivanenko - 2020 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus)

The paper investigates deepfake detection methods by amal- gamating super-resolution algorithms with deep learning, preceding a comprehensive review and evaluation of existing techniques. This novel approach offers a holistic strategy to tackle deepfake identification, leveraging both established methodologies and fresh perspectives. However, the paper lacks specific implementation details, leaving gaps in understanding the practical execution, such as architectural specifics and training methodologies. Additionally, potential biases or limited diversity within the dataset might impact the suggested strategy's real-world applicability. Despite these limitations, the successful implementation of this approach could pave the way for a more reliable and effective means of identifying deepfake videos, crucial in combating misinformation and fake content proliferation.

H. Deep fake image detection based on pairwise learning - Hsu, Chih-Chung, Yi-Xiu Zhuang, and Chia-Yen Lee - Applied Sciences, MDPI (2020)

The paper introduces a deepfake image detection method leveraging pairwise learning, employing pairwise comparisons to train a model for precise identification. Utilizing deep neural networks for feature extraction enhances accuracy by extracting robust and intricate features from images. This approach applies a deep learning framework, showcasing theoretical innovation through pairwise comparisons, which enhances accuracy in deepfake detection. However, its implementation demands intricate setup and higher computational resources. Furthermore, the method is limited to detecting deepfake images, excluding other media types like videos or audio, posing a challenge in keeping pace with evolving deepfake techniques. Despite these limitations, the approach represents a promising step toward robust and accurate deepfake image detection, although it requires further development to encompass diverse forms of deepfake media and address computational complexities.

I. Improved Xception with Dual Attention Mechanism and Feature Fusion for Face Forgery Detection - H. Lin, W. Luo, K. Wei and M. Liu - 2022 4th International Conference on Data Intelligence and Security (ICDIS))

The paper introduces an enhanced approach for face forgery detection, combining the Xception model with a dual attention mechanism and feature fusion techniques. Leveraging the Xception model's effectiveness in extracting discriminative features from face images, the inclusion of a dual attention mechanism allows the model to focus on crucial features via learned attention maps. Feature fusion further enhances the model's discriminative power by combining features from various network layers, ensuring robustness to noise and scalability to large datasets. Notably, these methods offer explainability and real-time implementation for human-understandable decision-making and streaming media forgery detection. However, challenges persist due to the rising quality of face forgeries, abundant datasets enabling attackers to train their detection models, and the dynamic nature of forgeries hindering static detection methods. Despite these obstacles, the proposed enhancements showcase promise in addressing the complexities of detecting increasingly sophisticated face forgeries.

III. METHODOLOGY

The research uses a thorough pipeline to improve image quality and collect discriminative features for future classification in the context of deepfake detection. The Contrast Limited Adaptive Histogram Equalization (CLAHE) technique is used to efficiently enhance image contrast. Following contrast enhancement, Multi-Task Cascaded Convolutional Networks (MTCNN) are used to locate, and crop faces inside images, focusing on crucial regions for deepfake analysis. Following that, Inception V3, a strong convolutional neural network (CNN) architecture pre-trained on a large dataset, is used to extract embeddings from the facial images, which are high-dimensional feature representations. These embeddings serve as detailed descriptions of delicate facial features. The project then comprises training a modified CNN, with the architecture tailored for deepfake detection using the extracted embeddings. This customized CNN is intended to learn and detect patterns indicative of deepfake manipulations. The whole strategy combines image preprocessing, face detection, feature extraction, and customized CNN training to provide a holistic strategy for robust and reliable deepfake detection.

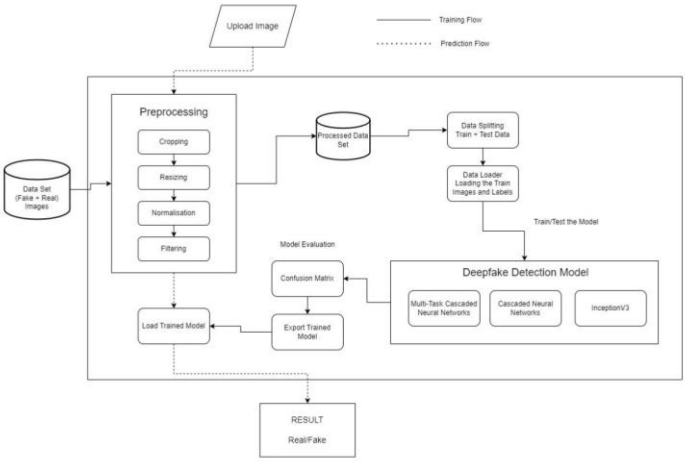

Fig. 1. Architecture Diagram

A. *Contrast enhancement using the CLAHE algorithm*

As a critical stage in the image preprocessing pipeline, the contrast-limited Adaptive Histogram Equalization (CLAHE) technique was implemented. The CLAHE settings, with a clip limit of 2.0 and a tile grid size of (8, 8), were designed to efficiently enhance visual contrast. A specific function, applied close to a folder, was created to systematically apply CLAHE to a picture folder and then save the processed images to an output folder. This tool aided in the implementation of CLAHE across many datasets by keeping the preprocessing processes organized and consistent. The application of CLAHE followed a defined method for each dataset, including Real Train, Fake Train, Real Test, and Fake Test. The input and output folders were initially constructed, with the latter created if it did not already exist. The function looped over the images in the input folder, reading each one with OpenCV. CLAHE was then applied to the image by converting it to the LAB color space, improving the luminance component, andmerging the processed components back together to create the final CLAHE-enhanced image in the BGR color space. Finally, the processed image was saved to the output folder, contributing to the creation of a preprocessed dataset with enhanced contrast,

an important step in the deepfake detection pipeline. This systematic use of CLAHE guaranteed homogeneity and improved the discriminative characteristics in the photos, ultimately boosting the performance of the deepfake detection model.

B. *Cropping images using MTCNN*

The exact localization and extraction of facial regions from dataset photos was a critical step in the preparation phase. This procedure was carried out independently for the datasets Real Train, FakeTrain, RealTest, and FakeTest. The implementation started by iterating through each image file within the relevant dataset directories, focusing on files with the extensions '.jpg' or '.png'. The OpenCV library was used to read the picture pixels for each image. Following that, a facial identification technique known as a detector was used to identify faces within the photos. The method found facial bounding boxes and provided the coordinates (x, y) and dimensions (width, height) of each found face. A loop was then started to run through the detected faces, allowing each facial region to be extracted by cropping the image based on the determined coordinates. These clipped facial portions were saved as individual image files in a designated output directory, with filenames that reflected the original image name as well as a suffix denoting the face index. In the case of the Real Train dataset, for example, the method entailed loading each image, finding faces, and then iteratively cropping and saving each discovered face to a predetermined output directory. This methodical procedure was repeated for the remaining datasets (Fake Train, Real Test, and Fake Test), assuring consistency and precision in face area extraction. This methodology functioned as an important preliminary step, separating facial traits for further analysis and feature extraction, ultimately contributing to the robustness of the deepfake detection model. Face detection and extraction techniques are used following best practices in preprocessing for facial recognition and deepfake detection applications.

C. *Embedding generation*

The use of the InceptionV3 architecture for extracting high-dimensional feature representations (embeddings) from pre-processed facial images was a critical stage. To collect discriminative facial features critical for further deepfake detection efforts, the InceptionV3 model was merged into a Sequential model in TensorFlow. The following major steps were included in the procedure: The InceptionV3 model, a complex convolutional neural network pre-trained on the ImageNet dataset, served as the foundation for the feature extraction procedure. The model was customized by inserting supplementary layers such as a global average pooling layer, a fully connected layer with 128 units using ReLU activation, and a dropout layer with a dropout rate of 0.5. The model was built with the Adam optimizer, with binary cross entropy as the loss function and accuracy as the evaluation metric. Image Embeddings in Action: Apply inceptionv3, a specialized function, was created to apply the customized InceptionV3 model to a collection of cropped facial photos. This function took a list of file paths corresponding to cropped photos, loaded the InceptionV3- based model, preprocessed the images to fit the model's input criteria, and then used the predict method to produce - beddings for the images. Dataset-specific Embeddings: Each dataset Real Train, Fake Train, Real Test, and Fake Test was implemented separately. The file paths of the cropped facial photos were collected for each dataset, and the InceptionV3-based model was used to construct embeddings. The resulting embeddings, which were saved as numpy arrays, enclosed high-dimensional representations of the unique facial features found in the individual datasets. These embeddings served as the basis for training the subsequent deepfake detection model. The use of the InceptionV3 architecture shows a transfer learning strategy, which leverages knowledge from a larger dataset to improve the model's performance in distinguishing complicated patterns and variations associated with authentic and altered facial photos. In conclusion, the incorporation of the

InceptionV3 architecture into the technique was important in extracting solid face embeddings, providing the framework for the following phases of the deepfake detection process.

D. Training CNN models on embeddings

We thoroughly investigated and refined various convolutional neural network (CNN) architectures for accurate deep-fake detection. The iterative procedure included experimenting with various model configurations and training strategies to improve the detection model's accuracy. Our models' evolution and the related variations in accuracy are detailed below: Initial CNN Model: We began our investigation with a simple CNN model consisting of a single convolutional layer with 32 filters, followed by a flattening layer and two tightly linked layers. The accuracy of this simple architecture was 73 percent. Callbacks in an Improved CNN Model: Recognizing the need for improvement, we added two convolutional layers and a dropout layer, as well as callbacks for early halting and learning rate scheduling. This change dramatically increased the accuracy to 82 percent. Scheduling Learning Rates and Increasing Patience: To improve the model, we used the ReduceLROnPlateau callback to construct a more advanced learning rate scheduler and increased the patience for early halting to allow for longer training. With this change, the accuracy increased to 87.9 percent. Modification of the Activation Function: We experimented with modifying the activation function to softmax to improve model performance. This change, however, resulted in a drop in accuracy, bringing it down to 81.9 percent. Because our work requires binary classification (differentiating between authentic and altered facial photos), the sigmoid activation function was more suitable. Sigmoid, unlike softmax, which is often used in multi-class classification settings, is capable of addressing binary classification problems by producing output values in the range [0, 1]. Deeper CNN Architecture: Recognizing the possibility of a more complicated design, we implemented a deeper CNN model with many convolutional layers, enhancing the model's capacity to collect intricate information. This change resulted in an 86 percent. accuracy.

E. Final Model

Our final and refined CNN architecture had three convolutional layers, each with a greater number of filters, resulting in a more sophisticated feature extraction process. Dropout regularization was also included in the model to boost generalization. The final model attained an accuracy of 89.9 percent thanks to the use of early halting and learning rate scheduling. The inclusion of ReduceLROnPlateau enabled the model to dynamically change its learning rate during training, which contributed to improved convergence and ultimate accuracy. The relevance of model design and training procedures in achieving optimal deepfake detection performance is shown by these iterative modifications. The final model, distinguished by a deeper architecture and attentive callback implementation, achieved the highest accuracy and represents the culmination of our efforts in optimizing the CNN architecture for the task at hand. Integration of Data: For both training and testing datasets, we used facial embeddings from actual and false photographs to ensure a complete representation of characteristics.

To distinguish between genuine (labeled as 1) and bogus (classified as 0) cases, binary labels were assigned. Model Architecture: The model design included a 1D CNN with three convolutional layers, each with increasing filters and a ReLU activation function. To facilitate deeper feature extraction, a flattening layer was used to move from convolutional layers to tightly linked layers. Two tightly connected layers culminated in a final layer with a sigmoid activation function for binary classification, incorporating a dropout layer for regularization. Optimization and training: The model was built with the Adam optimizer and binary cross-entropy loss as the chosen measure. To improve training efficiency, EarlyStopping, and ReduceLROnPlateau callbacks were used. EarlyStopping monitored validation loss and stopped training after 5 epochs if no progress was observed, whereas ReduceLROnPlateau dynamically changed the learning rate. Evaluation and training: The model was trained over 70 epochs with a batch size of 32 and a validation split of 20The stated callbacks were implemented into the training process, ensuring early halting for efficiency and dynamic learning rate adjustment. The model was evaluated on the test set after training, yielding crucial insights into its performance.

III. RESULTS

The implementation of two critical callbacks, EarlyStopping and ReduceLROnPlateau, significantly bolstered the efficacy of our model training process. EarlyStopping was configured to monitor validation loss, halting training if the loss stagnated for a specified number of epochs (patience) and restoring weights to the best-performing configuration. This tactic curbs overfitting, ensuring the selection of a well-generalized model. Conversely, ReduceLROnPlateau dynamically adjusts the learning rate based on validation loss, fine-tuning the rate for complex optimization landscapes.

To extend the exploration window, we increased both EarlyStopping and ReduceLROnPlateau patience from 3 to 5 epochs. This strategic adjustment prevented premature termination of training, granting the model more time to uncover potential accuracy improvements. Additionally, the adaptive learning rate from ReduceLROnPlateau facilitated a more refined search for optimal parameters, contributing to heightened accuracy compared to prior setups. The synergy between these callbacks was pivotal, culminating in our model achieving an impressive 89.9 percent accuracy in deepfake detection. The EarlyStopping callback's role in preventing overfitting by halting training upon validation loss plateauing ensures model generalization. Meanwhile, the ReduceLROnPlateau callback, with its adaptive learning rate, fine-tunes the training process, enhancing the model's ability to navigate complex optimization landscapes. Increasing the patience of both callbacks expanded the training window, enabling the model to explore potential accuracy gains and refine parameter optimization further. As a result, this orchestrated interplay between callbacks significantly contributed to our model's outstanding 89.9 percent accuracy in deepfake detection.

Modification Log		
SNo	Modification	Accuracy Achieved
1.	Learning Rate Scheduling	88 percent
2.	Used Early Stopping with a Patience of 5	88 percent
3.	Used Softmax Activation Function	82 percent
4.	Added More Layers to CNN Architecture	86 percent
5.	Added Class Weights	84 percent
6.	FINAL MODEL Combination of Points 1, 2, and 4	89.9 percent

TABLE I
TABULATED ACCURACIES AFTER MODIFICATIONS

Evaluation Metrics for Final Model		
SNo	Metric	Value
1.	Accuracy	89.9 percent
2.	Precision	87 percent
3.	Recall	80 percent
4.	F1 Score	83 percent

TABLE II
EVALUATION METRICS

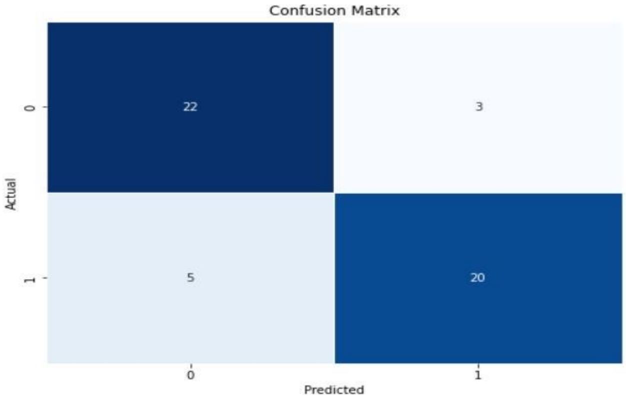

Fig. 2. Confusion Matrix

IV. REFERENCES

[1] [1]A. A. Balde, A. Jain, and D. Patra, “Improving facial recognition of facenet in a small dataset using deepfakes,” in *2020 4th International Conference on Computer, Communication and Signal Processing (IC- CCSP)*, 2020, pp. 1–6.

[2] [2]M. S. Rana, M. N. Nobi, B. Murali, and A. H. Sung, “Deepfake detection: A systematic literature review,” *IEEE Access*, vol. 10, pp. 25 494–25 513, 2022.

[3] [3]D. M. Montserrat, H. Hao, S. K. Yarlagadda, S. Baireddy, R. Shao,

[4] Horvath, E. Bartusiak, J. Yang, D. Guera, F. Zhu, and E. J. Delp, “Deepfakes detection with automatic face weighting,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020.

[5] [4]L. Guarnera, O. Giudice, and S. Battiato, “Deepfake detection by analyzing convolutional traces,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2020, pp. 666– 667.

[6] [5]A. Raza, K. Munir, and M. Almutairi, “A novel deep learning approach for deepfake image detection,” *Applied Sciences*, vol. 12, no. 19, p. 9820, 2022.

[7] [6]Z. Akhtar, “Deepfakes generation and detection: A short survey,” *Journal of Imaging*, vol. 9, no. 1, p. 18, 2023.

[8] [7]N. S. Ivanov, A. V. Arzhskov, and V. G. Ivanenko, “Combining deep learning and super-resolution algorithms for deep fake detection,” in *2020 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus)*. IEEE, 2020, pp. 326–328.

[9] [8]B. Malolan, A. Parekh, and F. Kazi, “Explainable deep-fake detection using visual interpretability methods,” in *2020 3rd International Conference on Information and Computer Technologies (ICICT)*.IEEE, 2020, pp. 289–293.

[10] [9]H. Lin, W. Luo, K. Wei, and M. Liu, “Improved xception with dual attention mechanism and feature fusion for face forgery detection,” in *2022 4th International Conference on Data Intelligence and Security (ICDIS)*. IEEE, 2022, pp. 208–212.

[11] [10]C.-C. Hsu, Y.-X. Zhuang, and C.-Y. Lee, “Deep fake image detection based on pairwise learning,” *Applied Sciences*, vol. 10, no. 1, p. 370, 2020.

[12] [11]S. Agarwal, H. Farid, T. El-Gaaly, and S.-N. Lim, “Detecting deep- -fake videos from appearance and behavior,” in *2020 IEEE International workshop on information forensics and security (WIFS)*. IEEE, 2020, pp. 1–6.

[13] [12]H. Zhao, W. Zhou, D. Chen, T. Wei, W. Zhang, and N. Yu, “Multi-attentional deepfake detection,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 2185–2194.

[14] [1] [2] [3] [4] [5] [6] [7] [8] [9] [10] [11] [12]