Artifact Based Deepfake Detection Methods

Preeti
Department of Computer Science and Engineering
Maharishi Markandeshwar Engineering College, (Deemed to be University)
Mullana, Ambala, Haryana, India, 133207
ranapreeti106@gmail.com

Sandhya Bansal
Department of Computer Science and Engineering
Maharishi Markandeshwar Engineering College, (Deemed to be University)
Mullana, Ambala, Haryana, India, 133207
sandhya12bansal@gmail.com

2023 Second International Conference on Informatics (ICI) | 979-8-3503-4383-0/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICI60088.2023.10421629

***Abstract*— One of the biggest innovations of Artificial Intelligence (AI) is the ability to generate manipulated or synthesized media (images, videos). When these generated media is created in to look like real and original people then it is called a Deepfake. Deepfakes have gained attention due to their potential to create convincing and misleading content that can be difficult to distinguish from authentic media. While they have garnered positive value in entertainment sector but have seen as a biggest threat regarding its ethical and societal implications particularly in the context of misinformation, identity theft, privacy invasion, and defamation. Rigorous research has been done since its beginning to prevent and detect media forgery and many detection techniques have been developed and discussed in literature. There is no clear winner in identification of DeepFakes however the artefact-based approaches seem to be much superior in the literature. In order to find the lean and identify best detection technique, an evaluation of existing techniques is needed which is the main purpose of our paper. This compares the video forgery detection techniques in terms of convergence, accuracy and equal error rate (EER).**

Keywords*— Artificial Intelligence, Machine Learning, Generative AI, DeepFakes, Artifacts.

I. Introduction

Due to the innovative digital technology, discern real and fake media are now more and more difficult. With the emergence of Deepfakes [1], [2], which are hyper-realistic videos that use artificial intelligence (AI) to show somebody saying and doing things that actually never happened, and is today counted among one of the most recent innovations causing the issue. [3]. AI programmes generate "deepfakes," videos or images that impersonate real ones by replacing, combining, and superimposing images as well as videos to seem as real one [4]. Deepfake has a wide range of beneficial uses in a variety of industries, including film, social media, games, educational media, digital communications, healthcare, entertainment, material science, and many economic sectors like e-commerce and fashion [3]. Additionally, deepfakes have the potential to be dangerous. For instance, early versions of deepfakes featured politicians, actors, comedians, and celebrities whose faces were incorporated into pornographic videos. [5], Future deepfakes are supposed to be utilised for things like extortion, market manipulation, political sabotage, bullying, revenge porn, fake news, and could even present a false video evidence in court [4].

Deepfakes have garnered popularity as a result of the high quality of their altered videos and the easy accessibility of their applications to users with varying degrees of computer expertise, from novices to experts[6]. Deep synthetic videos are created using a variety of techniques, models, and software applications. Some well-known and frequently employed methods [7] are: Faceswap [8], is a method that employs two encoders and decoders with shared parameters. AvatarMe [9], which regenerate 3D features through images, also can reform authentic 3D images with 4K by 6K pixel through a low-resolution image. Based on style transfer literature, StyleGan [10]is a novel architecture for GANs [11]. This architecture does automatic, unsupervised separation of high-level attributes and allows intuitive, scale-specific control image synthesis.

There are various ways to entail deepfakes and most frequently used [12] such as Head puppetry [13] entails synthesizing a video of a target persons whole head and upper-shoulder using a video of a source persons head, so the synthesized target appears to behave the same way as the source. Face swapping [8] involves generating a video of the target with the faces replaced by synthesized faces of the source while keeping the same facial expressions. Lip syncing is to create a falsified video by only manipulating the lip region so that the target appears to speak something that s/he does not speak in reality [14] . Deepfake technology has now become very sophisticated and is evolving exponentially. So, it is important for the detection methos as well as the laws regarding deepfake grow equally. There are various methods present to battle deepfakes [3]: legislation and regulation, corporate policies and voluntary action, education and training, and deepfake detection methods.

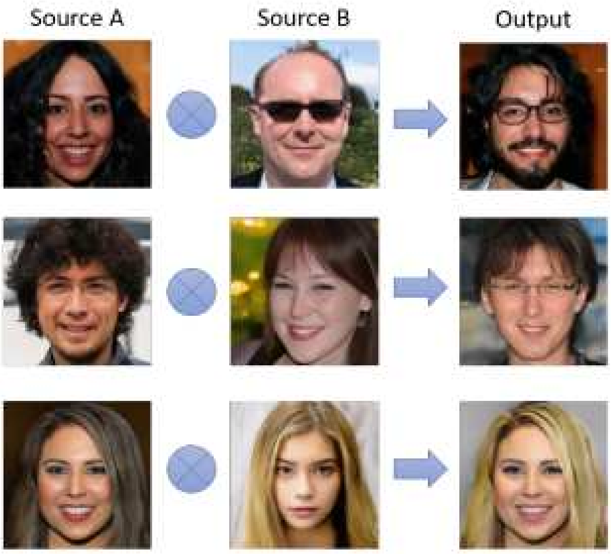

Fig. 1. Examples of DeepFakes frame Generation using StyeGAN [7]

Various methods have been proposed to detect manipulated videos. Early attempts mainly focused on inconsistent features caused by the face synthesis process while current detection methods mostly target at fundamental features. Some of the frequently used detection methods[15] are Network based methods, Artifacts based methods, Bio-signal Based methods, Statistical Based etc. General network-based approaches [16], [17] treat DeepFake detection as a

979-8-3503-4383-0/23/$31.00 ©2023 IEEE

frame-level classification problem. The Temporal consistency-based deepfakes methods are used to identify inconsistencies between adjacent frames due to algorithmic flaws. Consequently, RNN [16] is used to detect these inconsistencies. Visual artefacts-based methods use the blending operation in the generation process would result in intrinsic image differences at the blending edges. These artefacts are distinguished using CNN-based techniques.

As all devices leave unique traces in captured images due to their individual generation processes. In Forensic Analysis Faces and background images are simultaneously recognised as originating from distinct devices. Consequently, the detection operation can be accomplished using these traces. It is difficult for GAN to comprehend the concealed biological signals of features, making it challenging to synthesise human expressions with plausible behaviour. In Biological signals-based methods [18] biological signals are extracted from this observation to detect deepfake videos. Whereas, in the Statistical Analysis methods which analyse various statistical features of the video frames, such as colour distribution, noise patterns, or compression artifacts. Deepfakes often exhibit different statistical properties compared to authentic videos, and these discrepancies can be detected using machine learning algorithms.

Other methods such as blockchain and cryptographic Techniques use blockchain technology and cryptographic signatures to verify the authenticity of videos. By securely timestamping and storing video data on a blockchain, it becomes difficult to manipulate the content without leaving traces. Also, in Audio-Visual Correlation based deepfake detection [19] can be enhanced by analysing the correlation between audio and visual elements in a video. Inconsistencies between the lip movements and the corresponding audio can be indicative of deepfake manipulation. In certain cases, human experts trained in identifying deepfakes can manually review videos for signs of manipulation. While this approach is time-consuming and not scalable for large-scale detection, human expertise can be valuable for analysing suspicious cases [20]. Out of these methods it was observed that the Artifacts based methods performed significantly better than the others in terms of overall accuracy. However, as the picture is not that clear due to advent of deep learning methods, an evaluation is need to validate the performance of existing State-of-the-art Artifacts based methods. Which is the main scope of this paper.

The main contributions of this work lie in its systematic evaluation of artifact-based deepfake detection methods, the comparison of specific techniques, and the use of well-established datasets and evaluation metrics. This research aims to advance the understanding of the effectiveness of these methods in identifying deepfakes, which is a crucial area of study given the increasing concerns about media forgery. Which can be summarised as:

1. Development of an evaluation framework for assessing artifact-based deepfake detection methods.
2. Comparative study of four specific methods (DCL, GFF, Face X-ray, F3-Net) using two common datasets.
3. Evaluation based on metrics including convergence, Estimated Error Rate (EER), and accuracy to identify the most effective technique for face forgery detection.

The paper is structured into five distinct sections to systematically present the research findings. In the "Introduction" (Section I), the groundwork is laid by introducing the overarching problem of deepfake detection and the need to explore artifact-based methods. Section II, "Artifact-Based Methods," delves into the theoretical background of these detection techniques, providing a foundational understanding. Following this, Section III, "Evaluation of Artifact-Based Methods," introduces the evaluation framework and methodology used to compare the effectiveness of four specific techniques. The core experimental data and results are detailed in Section IV, "Experiment Setup and Evaluation Results." Finally, Section V, "Conclusion and Future Scope," concludes the paper by summarizing the key findings and discussing potential avenues for future research, thereby offering a comprehensive overview of the study's contributions and implications.

II. Artifact Based Methods

Artifact-based detection methods focus on analysing the artifacts, anomalies or inconsistencies that are typically present in deepfake videos due to the manipulation process. These artifacts can arise from the synthesis techniques used to create deepfakes leading to the abnormal behaviour of the generated image, like inconsistent brightness and boundary anomaly and can be detected using various techniques. In Literature there are many artifact-based detection methods commonly used such as Face warping artefacts [21] refers to a specific type of visual distortion that can occur in deepfake videos or image manipulations. It refers to the abnormal warping or stretching of facial features in a manipulated face. Head pose inconsistency refers to a visual anomaly that can be observed in deepfake videos or image manipulations, specifically related to inconsistencies in the pose or orientation of the subject's head. Bleeding Boundary can occur in deepfake videos or image manipulations, particularly at the boundaries of the manipulated region. It refers to the phenomenon where the colour or texture from the manipulated area appears to bleed or spill over into the surrounding background or adjacent objects. Compression Artifacts: Deepfake videos often exhibit different compression artifacts compared to authentic videos. By analysing the compression patterns, such as blockiness or blurring, it is possible to identify potential deepfake manipulations.

TABLE I. COMPARISON OF ARTIFACTS USED FOR DETECTION

Artifact	Description	Common Causes	Detection Techniques/Features	Pros	Cons
Face warping artifacts	Distortions in the facial structure or features	Image manipulation, face morphing	Analysis of geometric inconsistencies, facial landmarks, texture anomalies	Can detect facial morphing techniques	Less effective for subtle warping
Head pose inconsistency	Inconsistent head orientation or alignment	Pose estimation errors, improper alignment	Head pose estimation, facial landmark alignment	Effective for identifying manipulated head poses	Limited to head-related inconsistencies
Bleeding boundary	Unnatural blending or bleeding of object boundaries	Poor segmentation, edge artifacts	Edge detection, boundary analysis	Can identify tampered or merged objects	Less effective for complex or subtle boundaries
Compression artifacts	Distortions caused by lossy compression algorithms	High compression ratio, blockiness, blurring	Compression artifact analysis, DCT coefficient analysis	Effective for identifying lossy compression artifacts	May generate false positives in heavily compressed but

					authentic videos
Splicing artifacts	Visual inconsistencies where two images are merged	Image tampering, object insertion	Forensic analysis, image similarity comparison	Can identify image splicing or tampering	Less effective for advanced blending techniques
Eyeblink artifacts	Artifacts related to irregular or incomplete blinks	Partially closed eyes, frame misalignment	Blink detection, frame analysis	Can detect irregular or unrealistic eye movements	Limited to eye-related artifacts
Mouth artifacts	Issues with mouth region, such as unrealistic shapes	Mouth manipulation, lip syncing	Lip movement analysis, facial expression detection	Effective for identifying manipulated mouth movements	Less effective for subtle mouth artifacts
Reflection and lighting	Inconsistent lighting conditions or reflections	Varying light sources, reflective surfaces	Lighting analysis, reflection detection	Can detect inconsistencies in lighting and reflections	Limited to lighting-related artifacts
Inconsistencies in Pupil Dilation	Inconsistencies in the size or shape of pupils	Artificial dilation, pupil tracking errors	Pupil tracking, pupil shape analysis	Effective for detecting manipulated pupil dilation	Limited to pupil-related artifacts
Sensor and metadata	Analysis of sensor data and metadata for discrepancies	Metadata manipulation, device-specific artifacts	Source device analysis, metadata verification	Can detect anomalies in sensor data and metadata	Limited to the available sensor and metadata information

Splicing Artifacts When creating deepfakes, different facial or body parts are often spliced together from different sources. This can result in visible seams or irregularities at the boundaries between the manipulated regions. Detection algorithms can look for these splicing artifacts as evidence of deepfake manipulation. Eyeblink Artifacts Deepfake algorithms sometimes have difficulty generating natural-looking eye blinks. As a result, deepfake videos may exhibit unusual or abnormal eye blink patterns. By analysing the eye movements and blink patterns, it is possible to identify potential deepfakes. Mouth Artifacts Manipulating the mouth region to synchronize it with altered speech can introduce artifacts. Deepfake videos may exhibit unnatural lip movements, inconsistent lip-sync, or mismatched mouth shapes. These irregularities can be indicative of deepfake manipulation.

Reflection and Lighting Inconsistencies Deepfake algorithms often struggle with accurately replicating lighting conditions and reflections. Therefore, analysing inconsistencies in lighting and reflections across the video frames can help identify deepfake manipulations. Inconsistencies in Pupil Dilation Pupil dilation is an involuntary response that occurs in real humans in response to changes in lighting conditions or emotional states. Deepfake algorithms may fail to accurately replicate these changes, leading to inconsistencies in pupil dilation. Detecting such irregularities can be useful in identifying deepfakes. Sensor and Metadata Analysis Deepfake videos may lack sensor noise or metadata that are typically present in authentic videos. Analysing the absence or inconsistencies in sensor noise patterns, EXIF data, or other metadata can provide indications of potential deepfake manipulations. Furthermore, the artifact-based Methods can be divided into the approaches based on the type of classifier used, which are deep and shallow classifiers.

A. *Shallow classifiers*

A shallow classifier typically refers to a simpler model with a smaller number of layers or parameters. It may use basic machine learning algorithms such as logistic regression, support vector machines (SVM), or decision trees. Shallow classifiers often rely on handcrafted features or statistical properties extracted from the input data to make predictions These features can include various visual, statistical, or temporal characteristics that are believed to be indicative of deepfake manipulation. Some common features used in shallow classifiers for deepfake detection include texture patterns, colour distributions, noise statistics, motion characteristics, and facial landmarks. One advantage of using shallow classifiers for deepfake detection is that they often require less computational resources compared to deep learning models. They can be trained and deployed more efficiently, making them suitable for real-time or resource-constrained applications. However, shallow classifiers may have limitations in capturing complex and high-level representations that deep learning models excel at. They heavily rely on handcrafted features, which may not fully capture the intricacies of deepfake manipulations. As a result, their detection performance might be limited, especially against advanced and sophisticated deepfake techniques.

B. *Deep classifiers*

A Deep Classifier or Deep Learning [1], [14] refers to a more complex model that consists of multiple layers of interconnected nodes, such as deep neural networks. Deep classifiers can automatically learn complex features and representations from raw input data without relying on explicit feature engineering. These models can capture intricate patterns and hierarchies present in the data. This method produces artifacts that CNN models like VGG16 [22], ResNet50, ResNet101, and ResNet152 [23] can identify due to the resolution mismatch between the warping face area and surrounding context. One of the challenges in using deep classifiers for deepfake detection is the requirement for significant computational resources, especially during training. Deep learning models often require substantial amounts of data and computational power to effectively learn and generalize from complex patterns. Additionally, deep classifiers may be more susceptible to overfitting if the training data is not diverse or representative enough. To overcome limitations of deep and shallow classifiers, a combination of shallow classifiers and deep learning models can be used in ensemble-based approaches. This approach leverages the strengths of both techniques, utilizing the simplicity and speed of shallow classifiers while benefiting from the powerful representation learning capabilities of deep learning models.

III. Evaluation of Artifact based Methods

This evaluation work focused on assessing the performance of artifact-based methods for detecting face forgery in digital images. In this research work we have considered four recent research papers [24] [25] [26] [27] from literature that deal with the face forgery detection to compare and finding out the best technique among four for face forgery detection. The Four specific methods which were evaluated are named Dual Contrastive Learning (DCL), Generalizing Face Forgery Detection (GFF), Face X-ray, and Frequency in Face Forgery Network (F3-Net). These Methods are

A. *Dual Contrastive Learning (DCL)*

For face forgery detection authors in [26] have proposed Dual Contrastive Learning (DCL) framework, which simultaneously compare and contrast characteristics between instances and within instances. This technique trains the model through contrastive learning framework in a supervised way. First step in DCL is Data Views Generation (DVG) unit to produce various views of inputs via special designed data augmentation, next step is features got fetched through this specially designed supervised contrastive training framework. Further, to manage the feature distribution and improve the inconsistency of forgery faces, the Inter-Instance Contrastive Learning module and Intra-instance Contrastive Leaning module are used, respectively. But the overall loss (cross-entropy loss based supervised learning and losses based contrastive learning) is determined by:

$$L_{all} = \phi(L_{inter} + L_{intra}) + (1 - \phi)L_{ce} \qquad \text{Eq (1)}$$

Where in Eq (1) hyper-parameters ϕ is used to balance the cross-entropy loss and contrastive loss.

B. *Generalizing Face Forgery (GFF)*

Authors [27] proposed a model for face forgery detection. To make use of image noises, authors design three novel modules. First module is *Multiscale high-frequency* feature extraction. To extract high-frequency noises from images, they have used high-pass filters. So as to fetch more rich and informative features they apply these filters to low-level features at multiple scales. A two-stream network is constructed to process the two modalities by utilizing both high-frequency sounds and low-frequency textures, respectively. To attach more importance to forgery traces, second module is the *residual guided spatial attention* is used which is applied at the entrance part to guide the RGB modality. Third module is a *dual cross-modality attention* which is used as a connection of interaction between the two modalities. In this manner, the two modalities provide complementary information based on their correlation and facilitate representation learning mutually. They have also used loss function

$$\mathcal{L}_{AMS}\& = -\frac{1}{n}\sum_{i=1}^{n} \log \frac{e^{s\cdot\left(W_{y_i}^T f_i - m\right)}}{e^{s\cdot\left(W_{y_i}^T f_i - m\right)} + \sum_{j=1, j\neq y_i}^{c} e^{sW_j^T f_i}}$$

Eq (2)

C. *Generalizing Face Forgery (GFF)*

Authors [28] follows the approach that When an image is created by integrating two images, inherent image differences exist across the blending boundary. Their approach to detecting forgeries employs a technique known as face X-ray. They explain that enhanced generalization skills result from two factors: First, detecting the face X-ray as opposed to focusing on manipulation-specific anomalies.; Second, built a huge training dataset automatically and readily fused from real images so that the model is trained to focus more on the face X-rays. They applied the cross-entropy loss to estimate the model's accuracy.

$$L_b = -\sum_{\{I,B\}\in\mathcal{D}} \frac{1}{N}\sum_{i,j} \left(B_{i,j}\log \hat{B}_{i,j} + \left(1 - B_{i,j}\right)\log\left(1 - \hat{B}_{i,j}\right)\right)$$

Eq (3)

where N is the count of pixels in the feature map$\hat{B}$ In Eq (3), The loss for classification is

$$L_c = -\sum_{\{I,c\}\in\mathcal{D}} (c\log(\hat{c}) + (1 - c)\log(1 - \hat{c}))$$

Eq (4)

Hence, in the Eq (4),loss function is $L = \lambda L_b + L_c$ where λ is the loss weight balancing L_b and L_c .

D. *Frequency in Face Forgery Network (F3 -Net)*

Authors [29] propose a new framework named F 3 -Net (Frequency in Face Forgery Network), that take advantage of on the frequency-aware forgery evidences. This framework is made up with two frequency-aware modules, one module goal is to learn subtle forgery patterns via Frequency-aware Image Decomposition (FAD), and the other to fetch high-level semantics from Local Frequency Statistics (LFS) to check the frequency-aware statistical anomaly between forged as well as real images. Both modules FAD and LFS, are then gradually merged by a cross-attention module known as MixBlock, which facilitates interactions between FAD and LFS. The whole face forgery detection model is learned end-to-end through cross-entropy loss.

IV. Experiment Setup and Evaluation Results

In our research paper we have compared four different image and video forgery detection techniques based on their experiments performed on two common datasets. One is the FaceForensics++ [30] which a very huge forgery face dataset composing of 720 videos for training and 280 videos for validation or testing. For performing generalization experiments the dataset has two kinds of video quality: high-quality and low-quality. Another most common dataset used is CelebDF [31] which is consisting of 590 real videos collected from YouTube with groups of different ages, ethnic groups and genders, and 5639 5639 deep fake videos. Forgery videos are created using face swap for each pair of the 59 groups. The evaluation work compared the performance of artifact-based methods, namely Dual Contrastive Learning (DCL), Generalizing Face Forgery Detection (GFF), Face X-ray, and Frequency in Face Forgery Network (F3-Net), using several evaluation metrics including convergence, Estimated Error Rate (EER), accuracy. The Results of the Evaluation are presented in the Table 2 in terms of Convergence, EER, and ACC which are further explained.

TABLE II. Evaluation results in terms of Convergence, EER, and ACC

Work	Convergence	EER	ACC
DCL	99.3	3.26	87.7
GFF	98.36	3.85	95.6
Face Xray	87.4	7.62	85.2
F3-NET	98.1	3.58	96.8

A. *Convergence*

Convergence refers to the training process's stability and the model's ability to reach an optimal state. The evaluation assessed the convergence behaviour of each method, examining factors such as training time, convergence speed, and the presence of oscillations or plateaus in the learning curve. The higher convergence of F3-Net implies that the training process took longer to stabilize, possibly due to the complexity of the frequency-based analysis and the deep neural network architecture employed by F3-Net. The method's focus on analysing variations in frequency components to detect face forgery may have introduced additional computational complexity, leading to a longer training time. On the other End DCL's minimum convergence suggests that it was able to learn discriminative representations and capture the desired artifacts efficiently during the training process. The method's design, such as the utilization of dual contrastive learning framework, likely contributed to its faster convergence.

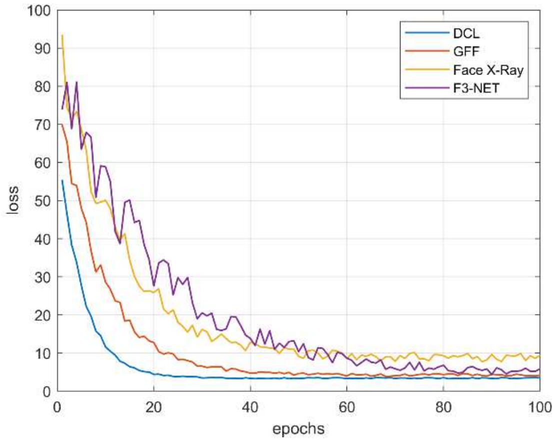

Fig. 2. Convergence curves of the DCL, GFF, Face X-ray and F3-Net.

By leveraging contrastive learning techniques, DCL effectively learned to differentiate between manipulated and authentic face images, leading to quicker convergence. Although GFF performed almost equaly well than the DCL method but it did take GFF more than 50 epochs to converge whereas DCL took around 30. As illustrated in Figure 3, it is clear that DCL has minimum convergence indicating DCL with take the minimum time of all four to get trained followed by GFF, Face Xray, and F3 net has the heist convergence indicating F3 net will take the maximum time to be trained. The Figure also explains the loss experienced during training and it is clear that DCL has the minimum loss followed by GFF, Face Xray and F3 net has the maximum loss among all four.

B. EER

EER is another important metric used in biometric systems evaluation. It represents the point where the false acceptance rate (FAR) and false rejection rate (FRR) are equal. The lower the EER value, the better the performance of the method. The evaluation work analysed the EER values to determine the accuracy of each method in correctly classifying manipulated and authentic face images.

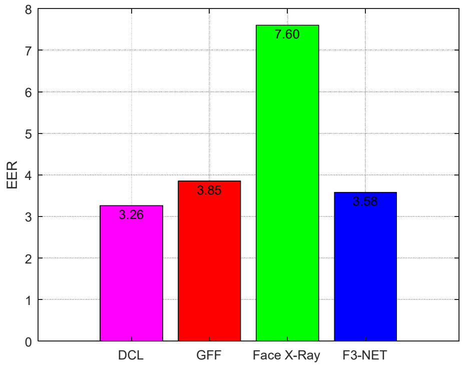

Fig. 3. EER of the DCL, GFF, Face X-ray and F3-Net methods.

During the study it was observed (Fig 3) that Face Xray has the maximum EER of 7.46 while others, GFF 3.85, F3 net has 3.58 and DCL has minimum EER of all with the value 3.26.

C. Accuracy

Accuracy is a widely used performance metric that measures the overall correctness of a model's predictions. The evaluation work calculated the accuracy values for each method to evaluate their effectiveness in accurately detecting face forgery.

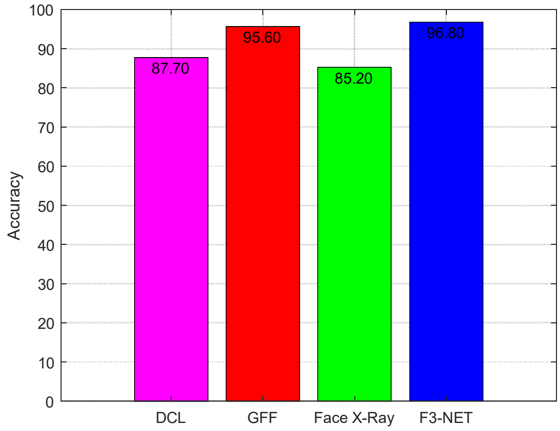

Fig. 4. Accuracy of the DCL, GFF, Face X-ray and F3-Net methods.

DCL and F3-NET demonstrate the highest accuracies among the evaluated methods, with DCL achieving an accuracy of 95.7% and F3-NET achieving an accuracy of 96.8%. These methods exhibit strong performance in accurately detecting face forgery and classifying manipulated and authentic face images. GFF closely follows with an accuracy of 95.6%, showcasing its effectiveness in detecting face forgery. It performs on par with DCL in terms of accuracy. Face X-ray has the lowest accuracy among the evaluated methods, with an accuracy of 85.2%. This indicates that Face X-ray may have challenges in accurately detecting face forgery or may not capture all relevant artifacts effectively, resulting in lower accuracy.

When comparing the methods solely based on accuracy, we can conclude that F3-NET achieves the highest accuracy, closely followed by DCL and GFF. Face X-ray lags behind in terms of accuracy. However, it's important to consider that accuracy alone may not provide a comprehensive assessment of a method's performance. Other metrics such as precision, recall, and F1 score should also be considered to evaluate the methods' effectiveness in detecting face forgery accurately. For Example, DCL has the Fastest Convergence among all even having slightly lower accuracy.

V. Conclusion and Future Scope

Deepfakes and their quality are on rise exponentially which can lead to various threats. So, the detection techniques need more sophistication and improvement to cope this situation. In search of best forgery detection technique, we have compared four resent face forgery detection techniques from literature. The evaluation work compared the performance of artifact-based methods, namely Dual Contrastive Learning (DCL), Generalizing Face Forgery Detection (GFF), Face X-ray, and Frequency in Face Forgery Network (F3-Net), These research techniques are evaluated in terms, EER and Convergence. Based on the study it is observed that DCL is the best among four techniques under consideration with minimum EER, convergence and accuracy and ERR. In future we would like to proceed with development of a new artifact-based face forgery detection method using deep learning approaches.

.

REFERENCES

[1] T. Zhang, "Deepfake generation and detection, a survey," Multimed Tools Appl, vol. 81, no. 5, pp. 6259–6276, 2022.

[2] M. S. Rana, M. N. Nobi, B. Murali, and A. H. Sung, "Deepfake detection: A systematic literature review," IEEE Access, 2022.

[3] M. Westerlund, "The Emergence of Deepfake Technology: A Review."

[4] M.-H. Maras and A. Alexandrou, "Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake videos," The International Journal of Evidence & Proof, vol. 23, no. 3, pp. 255–262, 2019.

[5] H. R. Hasan and K. Salah, "Combating deepfake videos using blockchain and smart contracts," Ieee Access, vol. 7, pp. 41596–41606, 2019.

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

[7] T. T. Nguyen et al., "Deep learning for deepfakes creation and detection: A survey," Computer Vision and Image Understanding, vol. 223, Oct. 2022, doi: 10.1016/j.cviu.2022.103525.

[8] B. Huang et al., "Implicit Identity Driven Deepfake Face Swapping Detection," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 4490–4499.

[9] A. Lattas et al., "AvatarMe: Realistically Renderable 3D Facial Reconstruction" in-the-wild"," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 760–769.

[10] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and improving the image quality of stylegan," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 8110–8119.

[11] S. Singh, R. Sharma, and A. F. Smeaton, "Using GANs to synthesise minimum training data for deepfake generation," arXiv preprint arXiv:2011.05421, 2020.

[12] S. Lyu, "DeepFake Detection: Current Challenges and Next Steps," Mar. 2020, [Online]. Available: http://arxiv.org/abs/2003.09234

[13] S. J. Pipin, R. Purba, and M. F. Pasha, "Deepfake Video Detection Using Spatiotemporal Convolutional Network and Photo Response Non Uniformity," in 2022 IEEE International Conference of Computer Science and Information Technology (ICOSNIKOM), 2022, pp. 1–6.

[14] S. Salman and J. A. Shamsi, "Comparison of Deepfakes Detection Techniques," in 2023 3rd International Conference on Artificial Intelligence (ICAI), 2023, pp. 227–232.

[15] P. Yu, Z. Xia, J. Fei, and Y. Lu, "A Survey on Deepfake Video Detection," IET Biometrics, vol. 10, no. 6. John Wiley and Sons Inc, pp. 607–624, Nov. 01, 2021. doi: 10.1049/bme2.12031.

[16] D. Güera and E. J. Delp, "Deepfake video detection using recurrent neural networks," in 2018 15th IEEE international conference on advanced video and signal based surveillance (AVSS), 2018, pp. 1–6.

[17] S. R. Ahmed, E. Sonuç, M. R. Ahmed, and A. D. Duru, "Analysis survey on deepfake detection and recognition with convolutional neural networks," in 2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), 2022, pp. 1–7.

[18] X. Jin, D. Ye, and C. Chen, "Countering spoof: towards detecting deepfake with multidimensional biological signals," Security and Communication Networks, vol. 2021, pp. 1–8, 2021.

[19] Y. Zhou and S.-N. Lim, "Joint audio-visual deepfake detection," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 14800–14809.

[20] P. Korshunov and S. Marcel, "Deepfake detection: humans vs. machines," arXiv preprint arXiv:2009.03155, 2020.

[21] Y. Li and S. Lyu, "Exposing deepfake videos by detecting face warping artifacts," arXiv preprint arXiv:1811.00656, 2018.

[22] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.

[23] L. Zhang, H. Li, R. Zhu, and P. Du, "An infrared and visible image fusion algorithm based on ResNet-152," Multimed Tools Appl, vol. 81, no. 7, pp. 9277–9287, 2022.

[24] L. Li et al., "Face x-ray for more general face forgery detection," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 5001–5010.

[25] Y. Qian, G. Yin, L. Sheng, Z. Chen, and J. Shao, "Thinking in frequency: Face forgery detection by mining frequency-aware clues," in European conference on computer vision, 2020, pp. 86–103.

[26] K. Sun, T. Yao, S. Chen, S. Ding, J. Li, and R. Ji, "Dual contrastive learning for general face forgery detection," in Proceedings of the AAAI Conference on Artificial Intelligence, 2022, pp. 2316–2324.

[27] Y. Luo, Y. Zhang, J. Yan, and W. Liu, "Generalizing face forgery detection with high-frequency features," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021, pp. 16317–16326.

[28] L. Li et al., "Face X-ray for More General Face Forgery Detection." [Online]. Available: https://29a.ch/photo-forensics/#noise-analysis

[29] Y. Qian, G. Yin, L. Sheng, Z. Chen, and J. Shao, "Thinking in Frequency: Face Forgery Detection by Mining Frequency-aware Clues," Jul. 2020, [Online]. Available: http://arxiv.org/abs/2007.09355

[30] A. Rossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, "Faceforensics++: Learning to detect manipulated facial images," in Proceedings of the IEEE/CVF international conference on computer vision, 2019, pp. 1–11.

[31] Y. Li, X. Yang, P. Sun, H. Qi, and S. Lyu, "Celeb-df: A large-scale challenging dataset for deepfake forensics," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 3207–3216.