WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Deep Fake Manipulated Face Detection in Video Using Deep Learning

Ms. K. Lalithavani [1], S. Muthamizhan [2], R. Sankar [3], G. Sathyadevan [4], S. Venkatesan[5]

[1] Assistant Professor, Department of Information Technology, Dhanalakshmi Srinivasan Engineering College (Autonomous), Perambalur, Tamil Nadu.
[2,3,4,5] UG - Department of Information Technology, Dhanalakshmi Srinivasan Engineering College(Autonomous), Perambalur, Tamil Nadu.
E-Mail : lalithavani95@gmail.com, zhimat283@gmail.com, sankar3535sk@gmail.com, sathyadev28ae@gmail.com, svenkatesan312@gmail.com

ABSTRACT

Recent years have seen a significant increase in media manipulation due to the advancement of technology and simplicity of producing fake information. Video forgery is constantly rising in the digital world due to information security breaches, so need to creating picture and video content monitoring approach for forgery identification. Applications for video forgery detection include media science, forensic analysis, digital investigations, and video authenticity verification.

Proposed, an approach for Deepfake detection has been provided for forgery detection in video. DenseNet is a convolutional neural network (CNN) algorithm that employed as a method to identify Deepfake videos. DenseNet comprises dense blocks where each layer receives input from all preceding layers. Train the DenseNet CNN on the prepared dataset. During training, the model learns to distinguish between authentic and forged videos by extracting meaningful features from the input frames.

The proposed approach only uses the deep features extracted from the DenseNet CNN model then applies the conventional mathematical approach on these features to find the forgery in the video.

Keywords: Face Detection and Recognition, Inception DenseNet Algorithm.

1. INTRODUCTION

The vast amount of video content that is readily available to a wide audience online is made possible by the quick development of cross-platform, computationally inexpensive video editing software. Fake movies have become more common in recent years because to the availability of high-performance, user-friendly video editing tools, AI techniques, and a plethora of video data. Fake photos and videos are used in fraudulent activities to publish fake news, evade facial authentication, and for amusement. Because of information security breaches, video forgery is becoming more and more common in the digital world. As a result, there is a need to monitor image and video content in order to identify instances of forgery. The proliferation of phoney videos increases social unrest and security threats. Applications for video forgery detection include digital investigations, forensic analysis, media science, and video authenticity verification. Video forensic technology is used to extract attributes that allow real videos to be distinguished from phoney content frames. Deep learning has become so amazing because to the growing processing power that it was unimaginable only a few years ago. Like any extraordinary innovation, this has made new difficulties. Purported "DeepFake" created by deep generative adversarial models that can control video and brief snippets.

For identification of the DF (Deep Fake), it is very essential to recognize the manner Generative Adversarial Network (GAN) creates the DF.GAN takes as input a video and a picture of a particular individual (‘target’), and outputs some other video with the goal’s faces changed with the ones of some other individual (‘source’). The spine of DF is deep adversarial neural networks trained on face images and target films to robotically map the faces and facial expressions of the source to the target. With the right post processing, the ensuing films can gain an excessive degree of realism. The input picture is changed in each frame of the movie after the GAN splits it up into individual frames. In addition, it recreates the video. Auto encoders are typically used to conduct this interaction. We provide a novel deep learning-based method that is capable of effectively differentiating DF videos from real ones.

2. LITERATURE REVIEW

Wu, Rongliang, and Tao Chen, et al. [1] proposed Cascade Expression Focal GAN (Cascade EF-GAN), a novel network that performs progressive facial expression editing with local expression focuses . The introduction of the local focuses enables the Cascade EF-GAN to better preserve identity-related features and details around eyes, noses and mouths, which further helps reduce artifacts and blurs within the generated facial images.

Li, Yuezun, Xin Yang, and Siwei Lyu, et al. [2] presented a new large-scale and challenging DeepFake video dataset, Celeb-DF3 , for the development and evaluation of DeepFake detection algorithms. There are in total 5, 639 DeepFake videos, corresponding more than 2 million frames, in the Celeb-DF dataset. The real source videos are based on publicly available YouTube video clips of 59 celebrities of diverse genders, ages, and ethnic groups.

Shen, Yujun, Jinjin Gu, Xiaoou Tang, and Bolei Zhou, et al. [3] proposed a framework InterFaceGAN, short for Interpreting Face GANs, to identify the semantics encoded in the latent space of well-trained face synthesis models and then utilize them for semantic face editing. Beyond the vector arithmetic property, this framework provides both theoretical analysis and experimental results to verify that linear subspaces align with different true-or false semantics emerging in the latent space. In this framework, we conduct a detailed study on how different semantics are encoded in the latent space of GANs for face synthesis.

The proposed method is further applied to achieve real image manipulation when combined with GAN inversion methods or some encoder-involved models. Extensive results suggest that learning to synthesize faces spontaneously brings a disentangled and controllable facial attribute representation.

Nirkin, Yuval, Yosi Keller, and Tal Hassner, et al. [4] implemented a novel iterative deep learning– based approach for face reenactment which adjusts significant pose and expression variations that can be applied to a single image or a video sequence. For video sequences, we introduce a continuous interpolation of the face views based on reenactment, Delaunay Triangulation, and barycentric coordinates. For the bounding boxes, we apply the smoothing to the center and box dimensions separately, and for the face landmarks we apply the smoothing to each face part separately.

Nguyen, Thanh Thi, and Cuong M. Nguyen, et al. [5] deepfake detection is normally deemed a binary classification problem where classifiers are used to classify between authentic videos and tampered ones. This kind of methods requires a large database of real and fake videos to train classification models. The number of fake videos is increasingly available, but it is still limited in terms of setting a benchmark for validating various detection methods.

Videos from the publicly available VidTIMIT database [64] were used to generate low and high quality deepfake videos, which can effectively mimic the facial expressions, mouth movements, and eye blinking. These videos were then used to test various deepfake detection methods. Test results show that the popular face recognition systems based on VGG and Facenet are unable to detect deepfakes effectively.

Huang, Yihao, and Geguang Pu, et al. [6] proposed a novel approach, termed Fake Locator, to obtain high localization accuracy, at full resolution, on manipulated facial images. To the best of our knowledge, this is the very first attempt to solve the GAN-based fake localization problem with a gray-scale fakeness map that preserves more information of fake regions. The fake texture produced by upsampling methods is totally different from the real texture in the real image. Hence, the output fake face images of these GAN-based methods inevitably contain fake textures that are unlikely obtained from the real world scenario through a camera.

Wang, Zhi, Yiwen Guo, and Wangmeng Zuo, et al. [7] implemented an adversarial training model into the training process of DeepFake detection. We show that it improves the generalization ability and robustness of the models notably. A novel method of generating adversarial examples based on image blurring is proposed, and it is shown to be more suitable to the adversarial training framework of deepfake detection. DeepFake detection is normally cast into a binary classification task. Predictions can be made on the basis of one image (as model input) or a sequence of images in a single video. Given a training set D that includes a large number of images and their corresponding labels.

Rana, Md Shohel, and Andrew H. Sung, et al. [8] proposed a systematic literature review (SLR) on Deepfake detection. Though the technology has been mostly used in legitimate applications such as for entertainment and education, etc., malicious users have also exploited them for unlawful or nefarious purposes. For example, high-quality and realistic fake videos, images, or audios have been created to spread misinformation and propaganda, foment political discord and hate, or even harass and blackmail people.

To provide an updated overview of the research works in DeepFake detection, we conduct a systematic literature review (SLR) in this paper, summarizing 112 relevant articles from 2018 to 2020 that presented a variety of methodologies. Analyze them by grouping them into four different categories: deep learning-based techniques, classical machine learning-based methods, statistical techniques, and blockchain-based techniques.

Wubet, Worku Muluye, et al. [9] implemented an approach that aims to investigate deepfake challenges, and to detect deepfake videos by using eye blinking. DeepFake detections are methods to detect real or deepfake images and videos on social media. DeepFake detection techniques are needed original and fake images or video datasets to train the detection models.

In this study, first discussed deepfake technology and its challenges, then identified available video datasets. Also, the eye aspect ratio, used for eye blinking rate classification and the CNN and eye aspect ratio detect the eye blinking intervals. The detection models have been trained on UADFV publically available real and fake videos. The deepfake detection methods detect the deepfakes by eye blinking.

Tripathy, Soumya, Juho Kannala, and Esa Rahtu, et al. [10] presented a novel Facial Attribute Controllable rEenactment GAN (FACEGAN), which transfers the facial motion from the driving face via the Action Unit (AU) representation. Unlike facial landmarks, the AUs are independent of the facial structure preventing the identity leak. Moreover, AUs provide a human interpretable way to control the reenactment. FACEGAN processes background and face regions separately for optimized output quality.

Our model manipulates the source face-landmarks with driving facial attributes to generate a new set of landmarks representing the desired motion with the source identity and structure.

3. PURPOSE

The purpose of the project described above is to address the growing threat of video forgery, particularly in the form of deepfake videos, and to develop a reliable method for detecting such forgeries. The project serves several purposes:

Security Enhancement: By detecting and identifying deepfake videos, the project aims to enhance security in various domains such as media, law enforcement, and digital communications. Preventing the dissemination of false information can mitigate potential harm caused by misleading content.

Preservation of Media Integrity: Ensuring the authenticity of video content is crucial for preserving media integrity. Detecting and removing deepfake videos helps maintain trust in digital media platforms and prevents the spread of misinformation.

Forensic Analysis: The project can assist in forensic analysis by providing tools to verify the authenticity of video evidence. This can be valuable in legal proceedings, investigations, and other situations where the credibility of video footage is critical.

Technological Advancement: Developing and implementing deepfake detection techniques contributes to the advancement of technology in the field of multimedia analysis and artificial intelligence. It involves leveraging state-of-the-art neural network architectures like DenseNet and applying them to real-world problems.

Social Impact: By combatting the spread of deepfake videos, the project aims to mitigate the societal chaos and security risks associated with misinformation and fake news. It contributes to fostering a more informed and resilient society.

4. OBJECTIVES

The objective of this project is to develop an effective deepfake detection system capable of identifying forged videos amidst the vast sea of digital content. In response to the rising threat posed by video forgery, particularly with the proliferation of deepfake technology, the project seeks to harness the power of advanced neural network architectures like DenseNet. By curating a diverse dataset encompassing authentic and manipulated videos, the objective is to train the model to discern subtle cues and patterns indicative of forgery. Through rigorous experimentation and optimization, the goal is to achieve high accuracy and reliability in detecting deepfake videos across various scenarios and contexts. Furthermore, the project aims to not only contribute to the advancement of technology in multimedia analysis but also to address pressing societal concerns surrounding misinformation and media integrity. By deploying a robust deepfake detection system, the objective is to bolster security, foster trust in digital media platforms, and empower users with tools to navigate the increasingly complex landscape of online content with confidence and clarity.

5. EXISTING SYSTEM

Exiting system implemented a method based on meta-learning called the meta-deepfake detection (MDD) algorithm. With a meta-optimization objective, in order to learn efficient face representations on both synthetic source and target domains. The MDD shifts the source domain to the target domain. So as to increase model generalization, the gradients from the meta-train and the meta-test are combined using meta-optimization. The MDD can handle unseen domains without model updating for unseen domains. Here separated the source domains into the meta-train domain Ttrains and the meta-test domain Ttests during training to obtain domain generalization. In order to simulate the domain shift problem that existed when used in real-world situations, the model is driven to acquire generalizable information about how to generalize well on the new domains with different distributions. We also create meta-batches for training and testing by randomly splitting N source domains of TS. These data contain both real and fake face pairs and these patterns are not duplicated across domains. These pairs increase collation and comparison of information between real and fake images. Therefore, it also increases inter-class separability, which can be interpreted as a distinct dispersion of the feature distribution of samples, increasing differentiation during training as well as enhancing the model’s quality. More distinguishable characteristics may be learned by the network with less effort during optimization. The fact is that features learned by supervised learning have much less ability to generalize when subjected to unseen manipulation techniques. Therefore, the model is easier to generalize when the source domain is split into meta-train and meta-test. In addition, samples in the meta-train and meta-test are also shuffled and selected at random, which minimizes the problem of overfitting. Additionally, the data in the unseen domain is very diverse in reality, which the model has never seen or been trained in before. Thus, meta-splitting makes the model easier to train and also to generalize to unseen data.

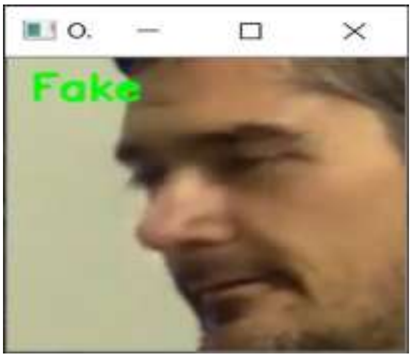
DISADVANTAGES

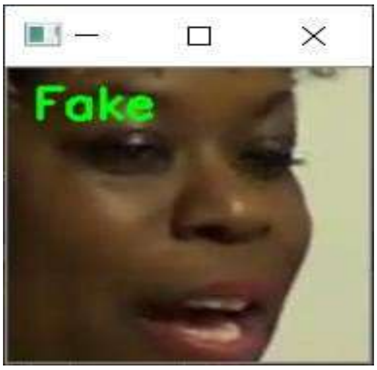
- Due to the rapid advancement of face forgery generation algorithms, some samples seem extremely similar to one another and only differ from one another by a few small features.
- It is getting harder to determine the difference between fake and real features in fake images.
- It is inefficient and perhaps impracticable to calculate the mean of the embedding of each class in each iteration, when taking into consideration the entire training set.

6. PROPOSED SYSTEM

Implement a brand-new deep learning-based approach that could successfully distinguish AI-generated fake videos (DF Videos) from actual videos. It’s incredibly essential to broaden technology that could spot fakes, so that the DF may be recognized and averted from spreading over the internet. Recently, Convolution Neural Network (CNN) has turn into a de-facto technique for classification of multi-dimensional data and it renders standard and also highly effectual network layer arrangements. But these architectures are limited by the speed due to massive number of calculations needed for training in addition to testing the network and also, it might render less accuracy. Leveraging the powerful DenseNet Convolutional Neural Network architecture, the system aims to distinguish between authentic and deepfake images or videos. Commencing with the assembly of a diverse dataset, encompassing both genuine and manipulated samples, the project involves meticulous preprocessing and augmentation to enhance model robustness. By implementing transfer learning, the model capitalizes on pre-trained DenseNet weights, fine-tuning its parameters for optimal performance in deepfake detection. Rigorous training and validation phases ensure the model's accuracy and generalization. Hyperparameter tuning and thorough testing against adversarial attacks contributes to the system's robustness.

Post-processing techniques refine results, and upon successful completion, the trained model will be deployed for real-time or batch processing, thus providing a reliable tool in the ongoing battle against the proliferation of deepfake content. Continuous monitoring and updates, along with adherence to metrics like accuracy and precision, guarantee the model's efficacy in real-world scenarios.

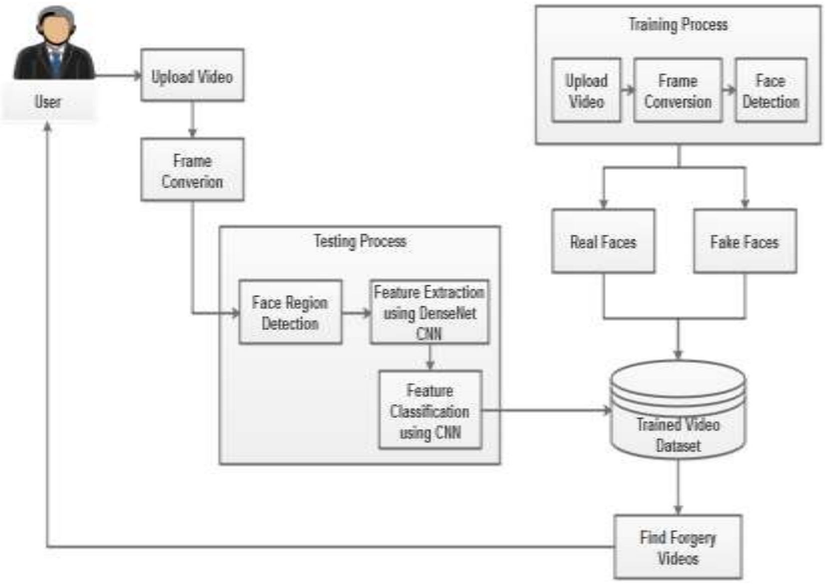




ADVANTAGES

- Proposed technique effectively classifies the forgery videos from the original video.
- Proposed DenseNet with CNN based feature classification achieves high performance on forgery detection.
- DenseNet helps the model learn discriminative features faster with less inference time.

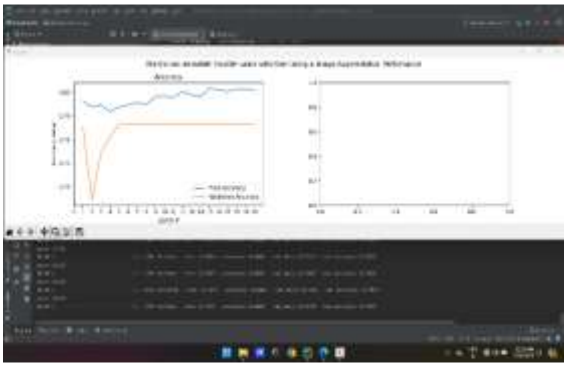
SYSTEM ARCHITECTURE



7. RESULT NALYSIS

Choosing a diverse dataset containing both authentic and deepfake manipulated videos is crucial. The dataset should cover various lighting conditions, angles, facial expressions, and backgrounds.

Analyze the model's performance in detecting deepfake manipulated faces. Pay attention to its ability to distinguish between authentic and manipulated faces while minimizing false positives and false negatives and obtained accuracy is 80%.



8. FUTURE ENHANCEMENT

DeepFakes are constantly evolving, and new convincing fakes. Need to develop other CNN models that are capable of detecting a wider range of deepfakes, including those created using new techniques. Real-time detection of deepfakes is critical for preventing the spread of misinformation on social media platforms. Future research work focus on developing other CNN models that can detect deepfakes in real-time, using limited computational resources.

9. CONCLUSION

In this work implemented DeepFake detection model to detect video forgery using DeepFake video dataset. DenseNet can be employed as a powerful convolutional neural network (CNN) architecture to effectively capture complex spatial features and patterns. The DenseNet architecture is characterized by densely connected blocks, enabling the seamless flow of information between layers. The detection network consisting of fully connected layers is employed to take the sequence descriptor as input and calculate probabilities of the frame sequence belonging to either authentic or deepfake class. Proposed model when trained with a large data set gave quiet impressive results compared to other deep learning models. Using the sequence descriptor as input, a detection network with fully connected layers is used to determine the likelihood that a frame sequence belongs to the legitimate or deep fake class. Proposed work can offer a defence in the detection of AI based fake videos and obtained accuracy is 80%.

REFERENCES

[1] Wu, Rongliang, Gongjie Zhang, Shijian Lu, and Tao Chen. "Cascade ef-gan: Progressive facial expression editing with local focuses." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5021-5030. 2020.

[2] Li, Yuezun, Xin Yang, Pu Sun, Honggang Qi, and Siwei Lyu. "Celeb-df: A large-scale challenging dataset for deepfake forensics." In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 3207-3216. 2020.

[3] Shen, Yujun, Jinjin Gu, Xiaoou Tang, and Bolei Zhou. "Interpreting the latent space of gans for semantic face editing." In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 9243-9252. 2020.

[4] Nirkin, Yuval, Yosi Keller, and Tal Hassner. "FSGANv2: Improved subject agnostic face swapping and reenactment." IEEE Transactions on Pattern Analysis and Machine Intelligence 45, no. 1 (2022): 560-575.

[5] Nguyen, Thanh Thi, Quoc Viet Hung Nguyen, Dung Tien Nguyen, Duc Thanh Nguyen, Thien Huynh-The, Saeid Nahavandi, Thanh Tam Nguyen, Quoc-Viet Pham, and Cuong M. Nguyen. "Deep learning for deepfakes creation and detection: A survey." Computer Vision and Image Understanding 223 (2022): 103525.

[6] Huang, Yihao, Felix Juefei-Xu, Qing Guo, Yang Liu, and Geguang Pu. "Fakelocator: Robust localization of GAN-based face manipulations." IEEE Transactions on Information Forensics and Security 17 (2022): 2657-2672.

[7] Wang, Zhi, Yiwen Guo, and Wangmeng Zuo. "Deepfake forensics via an adversarial game." IEEE Transactions on Image Processing 31 (2022): 3541-3552.

[8] Rana, Md Shohel, Mohammad Nur Nobi, Beddhu Murali, and Andrew H. Sung. "Deepfake detection: A systematic literature review." IEEE Access (2022).

[9] Wubet, Worku Muluye. "The deepfake challenges and deepfake video detection." Int. J. Innov. Technol. Explor. Eng 9 (2020).

[10] Tripathy, Soumya, Juho Kannala, and Esa Rahtu. "Facegan: Facial attribute controllable reenactment gan." In Proceedings of the IEEE/CVF winter conference on applications of computer vision, pp. 1329-1338. 2021