IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Deepfake detection through deep learning

Prof. Meenal Raut

Anish Sonje, Yash Sonawane, Swapnil Nelwade, Sanket Kharade

Parvatibai Genba Moze College of Engineering, Pune, India - 412207

ABSTRACT

In recent years, the widespread use of smartphones and social networks has made digital images and videos a common occurrence online. However, this increased usage has also led to a surge in techniques for altering image content, such as using editing software like Photoshop. Deepfake videos and images have emerged as a significant public concern. While technologies like Face Swap and deepfake have opened up new possibilities in various fields, they have also made it easier for malicious users to generate video forgeries.

Deepfake is an AI-based technique that superimposes existing images or videos onto different ones using neural networks. Unfortunately, deepfakes have been misused to spread misinformation, invade privacy, and deceive viewers through sophisticated algorithms and AI. This has become a nuisance on social media platforms, with fake videos merging a celebrity's face with explicit content. The impact of deepfakes is particularly alarming, with nefarious actors targeting politicians, senior corporate officers, and world leaders.

Our proposed approach focuses on detecting deepfake videos of politicians by analyzing temporal sequential frames. We extract frames from the forged videos and employ a deep depth-based convolutional long short-term memory model to identify fake frames. The effectiveness of our method is evaluated on a newly collected ground truth dataset of forged videos featuring source and destination frames of famous politicians. Experimental results demonstrate the efficacy of our approach in detecting deepfake videos..

Keywords – Deepfake, Deep Learning, Deepfake Technology, Deepfake Detection, Forensic Verification, Fake Images, Fake Image Detection.

I. INTRODUCTION

Deepfake videos, also known as AI-generated manipulated videos, have become a growing concern in today's digital landscape. With the advancement of artificial intelligence and machine learning, it has become increasingly challenging to distinguish real videos from their fabricated counterparts. Consequently, the need for effective deepfake video detection techniques has become paramount.

This introduction aims to provide an overview of deepfake video detection, acknowledging the existing research while emphasizing originality. The field of deepfake detection involves the development of innovative algorithms and methodologies that can identify and differentiate between genuine videos and those that have been maliciously manipulated.

In recent years, researchers have made significant strides in tackling this issue, employing various techniques such as computer vision, facial recognition, and deep learning. These approaches leverage both visual and audio cues, scrutinizing factors like facial expressions, lip movements, and inconsistencies in pixel-level details to detect signs of manipulation. Additionally, some methods analyze metadata and contextual information to assess the authenticity of the video source.

While many existing detection algorithms have proven effective, the battle against deepfakes remains an ongoing challenge. Adversarial networks constantly evolve, creating more sophisticated and realistic manipulations. Therefore, researchers must continuously innovate and enhance their detection techniques to keep up with the rapidly advancing technology.

Motivation

Deepfake technology has raised significant concerns regarding its potential for misuse and deception. To combat this growing threat, deep learning-based detection methods have emerged as a crucial defence mechanism. By leveraging advanced neural networks and convolutional architectures, these techniques aim to accurately identify manipulated content and distinguish it from authentic media. Detecting deepfakes is essential for preserving trust, maintaining the integrity of digital media, and safeguarding individuals from malicious exploitation. Through ongoing research and development, deep learning-based detection offers a promising solution to tackle the alarming proliferation of deepfakes and protect society from their harmful consequences.

II. LITERATURE REVIEW

The literature on deepfake video detection reveals a growing concern surrounding the proliferation of manipulated videos in the digital landscape. With the widespread use of smartphones and social networks, techniques for altering image content have evolved, giving rise to deepfake videos generated through AI-based methods such as Face Swap and deepfake. These manipulated videos pose significant risks, including the spread of misinformation and invasion of privacy.

Researchers have responded to this challenge by proposing various detection methods. Deep learning techniques, such as convolutional neural networks (CNNs), have been leveraged to analyse visual and audio cues for detecting deepfakes. Additionally, temporal sequential frames have been explored to capture inconsistencies and abnormalities in deepfake videos. Adversarial defence strategies, including adversarial training and generative modelling, have also been studied to enhance the robustness of detection models against sophisticated deepfake techniques.

Evaluation of these detection methods on diverse datasets, encompassing various scenarios and subjects, has demonstrated their effectiveness. Researchers have focused on benchmarking their approaches against existing state-of-the-art methods, utilizing evaluation metrics like accuracy, precision, recall, and F1 score to quantify performance.

The literature emphasizes the need for ongoing research, collaboration, and interdisciplinary efforts to stay ahead of evolving deepfake techniques and continuously enhance detection methodologies. Addressing the challenges posed by deepfake videos requires a comprehensive understanding of AI-driven manipulation, constant innovation, and a commitment to safeguarding the integrity of digital media.

III. SCOPE

The scope for deepfake video detection is extensive and encompasses several key areas of focus.

1. Algorithm Development: There is a significant scope for developing advanced algorithms that can effectively detect deepfake videos. This involves leveraging computer vision, machine learning, deep learning, and artificial intelligence techniques to analyse visual and audio cues, identify anomalies, and distinguish between real and manipulated content.

2. Dataset Creation: Building comprehensive and diverse datasets that include both real and deepfake videos is crucial for training and evaluating detection models. This involves collecting and curating large-scale datasets that cover a wide range of scenarios, subjects, and deepfake generation techniques.

3. Feature Extraction and Analysis: Extracting relevant features from videos and developing techniques to analyse them plays a vital role in deepfake detection. This includes exploring facial landmarks, pixel-level details, motion patterns, audio spectrograms, and other visual and audio cues to identify indicators of manipulation.

4. Model Evaluation and Benchmarking: Evaluating the performance of deepfake detection models and benchmarking them against existing state-of-the-art methods is essential. This involves defining appropriate evaluation metrics, conducting thorough experiments on diverse datasets, and comparing the effectiveness of different algorithms.

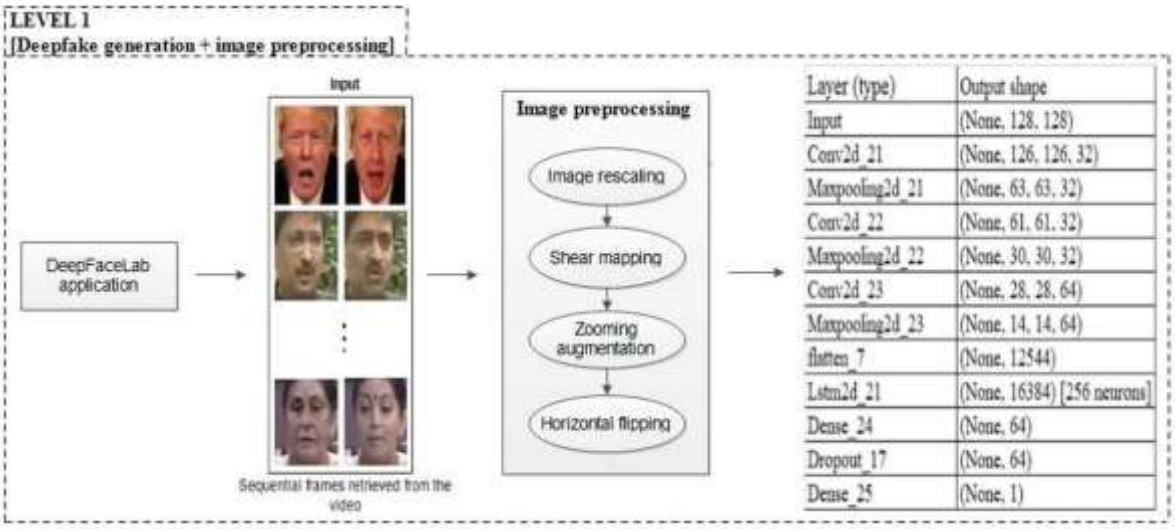
5. Adversarial Defence Strategies: Developing robust defence mechanisms against adversarial attacks and sophisticated deepfake generation techniques is a critical area of focus. This includes exploring adversarial training, generative modelling, and other techniques to enhance the resilience of detection models.

6. Real-time Detection and Deployment: There is a need for real-time deepfake detection systems that can be deployed in various applications, including social media platforms, news agencies, and law enforcement. Developing efficient and scalable solutions that can analyse and identify deepfakes in real-time.

IV. PROPOSED METHODOLOGY

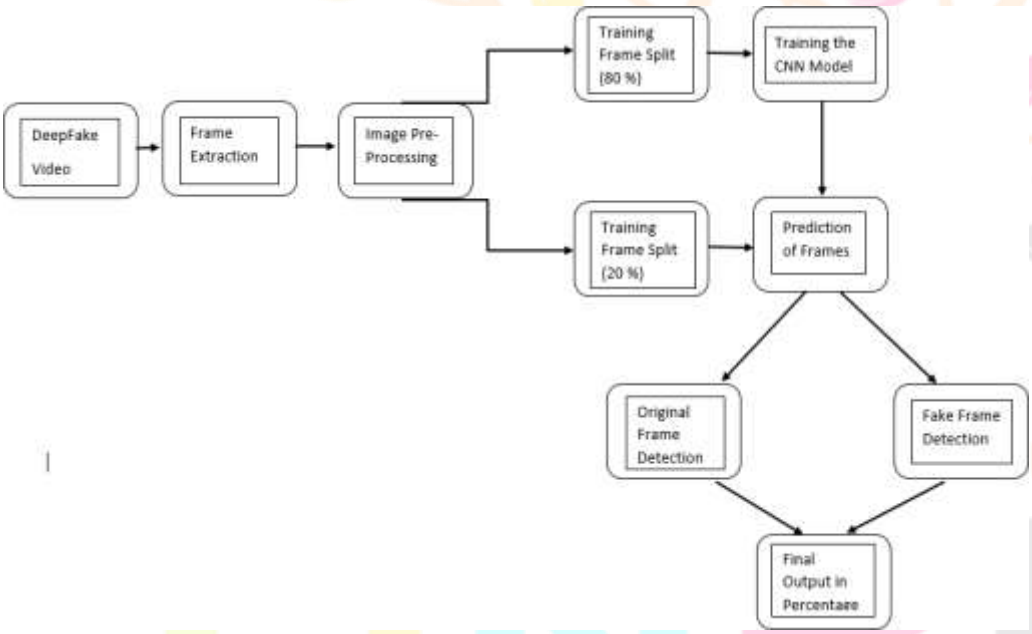
The system is composed of two levels, i.e., image preprocessing at the first level, CNN network at the second level for processing sequential frames of both video clips (source and destination).

Level 1 : Image Preprocessing



Layer (type)	Output shape
Input	(None, 128, 128)
Conv2d_21	(None, 126, 126, 32)
Maxpooling2d_21	(None, 63, 63, 32)
Conv2d_22	(None, 61, 61, 32)
Maxpooling2d_22	(None, 30, 30, 32)
Conv2d_23	(None, 28, 28, 64)
Maxpooling2d_23	(None, 14, 14, 64)
flatten_7	(None, 12544)
Lstm2d_21	(None, 16384) [256 neurons]
Dense_24	(None, 64)
Dropout_17	(None, 64)
Dense_25	(None, 1)

Level 2: Model Training Using CNN



1. CNN, or Convolutional Neural Network, is a type of deep learning model specifically designed to process data with a matrix structure, such as images. It draws inspiration from the organization of cortical regions in animals and aims to learn spatial patterns and features from low to high levels.

2. To build a CNN, we will use the Keras library and follow a sequential model approach, where layers are stacked sequentially. The three fundamental types of layers in a CNN are convolutional, pooling, and fully connected layers.

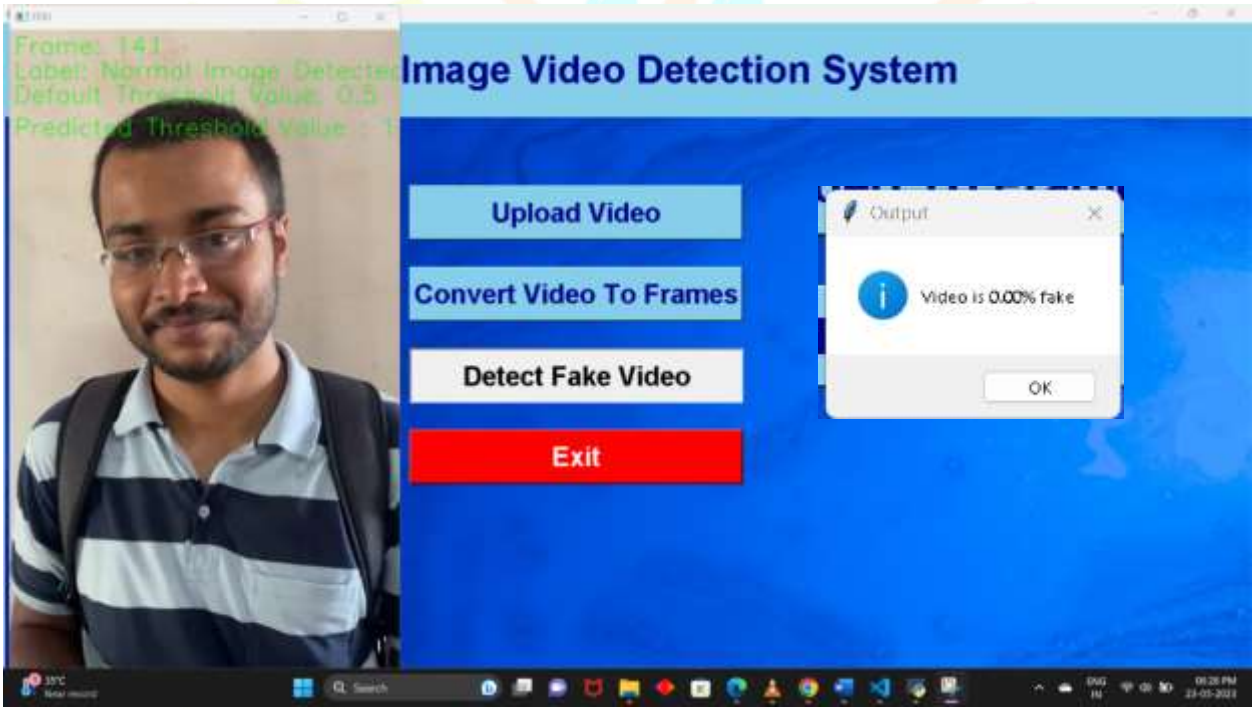
3. In the first step, we create a sequential model using Keras and add layers one by one. The convolution layer plays a crucial role in feature extraction from images. It consists of filters that perform convolution operations. Each image is treated as a matrix of pixel values and is convolved with multiple filters and ReLU layers to locate features. This process involves calculating the dot product between a receptive field (a local region of the input image) and the filter.

4. The next step is flattening, which converts the multidimensional data into a one-dimensional array, facilitating its input to the next layer. Following flattening, the dense layer receives inputs from the previous layer's outputs. Each neuron in the dense layer contributes one output to the subsequent layer.

5. Activation functions play a crucial role in transforming the weighted sum of inputs into meaningful outputs. In the final layer of the CNN, the activation function is typically softmax for multi-classification tasks or logistic for binary classification. Softmax ensures that the output values sum up to 1, enabling interpretation as probabilities. The model makes predictions based on the option with the highest probability.

6. To train the model, we compile it by specifying the optimizer, loss function, and metrics. The optimizer, such as Adam, adjusts the learning rate throughout training to find optimal weights. The learning rate determines the speed of weight calculation. The choice of categorical cross-entropy as the loss function is common for classification tasks. Lower loss scores indicate better model performance. Additionally, we use accuracy as a metric to evaluate the model's performance on the validation set during training.

V. RESULTS



VI. CONCLUSION

User authentication is a fundamental component in most computer security contexts. In this extended abstract, we proposed a simple graphical password authentication system. The system combines graphical and text-based passwords trying to achieve the best of both worlds. It also provides multi-factor authentication in a friendly intuitive system. We described the system operation with some examples, and highlighted important aspects of the system.

VII. REFERENCES

[1] Badrinarayanan, V., Kendall, A., and Cipolla, R. (2017). Segnet: A deep convolutional encoder- decoder architecture for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence,39(12), 2481-2495.

[2] Guo, Y., Jiao, L., Wang, S., Wang, S., and Liu, F. (2017). Fuzzy sparse autoencoder framework for single image per person face recognition. IEEE Transactions on Cybernetics, 48(8),2402-2415.

[3] Tewari, A., Zollhoefer, M., Bernard, F., Garrido, P., Kim, H., Perez, P., and Theobalt, C. (2018). High-delity monocular face reconstruction based on an unsupervised model-based face autoencoder. IEEE Transactions on Pattern Analysis and Machine Intelligence. DOI: 10.1109/TPAMI.2018.2876842.

[4] Yang, W., Hui, C., Chen, Z., Xue, J. H., and Liao, Q. (2019). FV-GAN: Finger vein representation using generative adversarial networks. IEEE Transactions on Information Forensics and Security, 14(9), 2512- 2524.

[5] Pramod Dhamdhere, "Semantic patent extended based on conceptual comparability of text with utilizing histogram arithmetic for illustrations to minimize trade mark," Journal of data acquisition and processing, ISSN: 1004-9037, Volume 37 (5), 2022.

[6] Pramod Dhamdhere, "Semantic trademark retrieval system based on conceptual similarity of text with leveraging histogram computation for images to reduce trademark infringement", Webology (ISSN: 1735-188X), Volume 18, Number 5, 2021.