WWW.IJRPR.COM

International Journal of Research Publication and Reviews

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

Advancements in Deepfake Detection: A Deep Learning Perspective

Anirudh Kumar[1], Abhishek Jha[2], Ankit Kumar[3], Sarthak Dixit[4], Aditya Sam Koshi[5]

[1,2,3,4] Student, Department of Computer Science &Engineering, Bhagwan Parshuram Institute of Technology

[5] Assistant Professor, Department of Computer Science & Engineering, Bhagwan Parshuram Institute of Technology

ABSTRACT:

In the digital sphere, deepfake technology offers both promise and danger. Although it has the potential to be used lawfully, there are problems associated with it as well. First, it can be used to modify video footage, which raises serious social and security issues. The quickly developing technology used to construct deep fakes makes it difficult for traditional deep fake detection techniques, such as visual quality analysis or inconsistency detection, to stay up to date. Thus, the need for increasingly advanced detection methods is urgent.

This research presents an improved method employing graph neural networks (GNN) for deepfake video detection in order to close this gap. The detection process is split into two stages by the suggested method: a four-block CNN stream and a mini-batch graph convolution network stream. These streams are made up of necessary operations including activation functions, batch normalization, and convolution. The fattening operation at the end is what connects the convolutional layers to the dense layer. Three distinct fusion networks—FuNet-A (additive fusion), FuNet-M (element-wise multiplicative fusion), and FuNet-C (concatenation fusion)—assist in the fusion of these phases.

After 30 epochs, the suggested model shows an amazing 99.3% training and validation accuracy across a variety of datasets. This high level of accuracy highlights the need of improving detection techniques to counter the growing threat of deep fake technology. It also shows how effective the GNN-based approach is at detecting deep fake videos and highlights its potential to address the changing landscape of deceptive media content.

Keywords: Graph neural network, Convolutional neural network, Deepfake video detection, Multi-task cascaded convolutional neural network, Mini-GNN

Introduction:

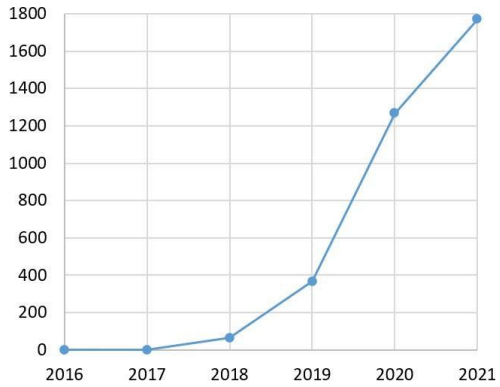
With the development of artificial intelligence, especially machine learning, it is now possible to create audio and video clips that are virtually identical to actual ones, making image and video forgeries a serious threat to civilization. Methods that use deep learning networks to edit images and videos, such as the deepfake technique, have become popular tools for content-changing video falsification. By seamlessly substituting the faces of people in photos or films with those of other people through the use of deep learning techniques, this technology streamlines and expedites the process of producing convincing false images and movies.

Artificial intelligence algorithms are used to create synthetic media called "deepfakes," which may convincingly portray people speaking or doing things that never happened. Deepfake producers may easily swap faces, change voices, and control actions in films with never-before-seen realism by utilizing deep learning techniques like generative adversarial networks (GANs) and autoencoders [2-8].

In order to train models to produce photo-realistic images and videos, deepfake techniques typically need a lot of image and video data. Deepfakes initially target public personalities, such as politicians and celebrities, because they often have a vast amount of photographs and videos available online. The faces of politicians or celebrities were replaced with bodies in pornographic photos and movies using deepfakes. In 2017, the first deepfake video featuring a porn actor's face in place of a celebrity's surfaced. When deepfake techniques are used to produce films of world leaders giving phony remarks for the aim of falsification, it poses a threat to international security. [9-12]

Therefore, deepfakes can be used to manipulate public opinion and influence election outcomes, incite conflict between nations over politics or religion, or destabilize financial markets by spreading false information. It can even be used to create fictitious satellite photos of the Earth that contain imaginary items in order to trick military analysts. For example, it can be used to create a bridge across a river that doesn't actually exist. This might lead a troop under guidance to cross the bridge during combat in the wrong direction.[13-15]

Since the turn of the millennium, another term which is linked to stalking has gained attention in the media and on the net, known as Group or Gang stalking. It generally involves a single stalker who may also recruit others into stalking by proxy, their involvement usually being ignorant or unsuspecting.



Importance of Deepfake Detection

Considering the situation where social and political are getting their image tarnished due to this technology, it is critical to develop efficient detection techniques. Deepfake detection is an essential safeguard against the fabricated content, allowing people, groups, and platforms to recognize and stop the spreading of false information and noxious propaganda.

Because of the rise in technology there are wide variety of manipulation techniques used to make deepfake video and image that detecting it is complex task. Furthermore, as deepfake technology is becoming more widely available there is greater chance that it will be abused, which make it precautionary steps to protect digital media integrity necessary.

Deepfake Creation

Deepfakes have gained popularity because of the high caliber of the manipulated movies and the ease with which their applications may be used by users of all computing skill levels, from experts to beginners. The majority of these applications were created using deep learning methods. The ability of deep learning to represent complicated and high-dimensional data is widely recognized. Deep autoencoders are one type of deep network with that capability; they are frequently used for image compression and dimensionality reduction. Using an autoencoder-decoder pairing structure, a Reddit user created the first deepfake creation, known as FakeApp. According to that technique, face images' latent features are extracted by the autoencoder and then rebuilt by the decoder [17-19].

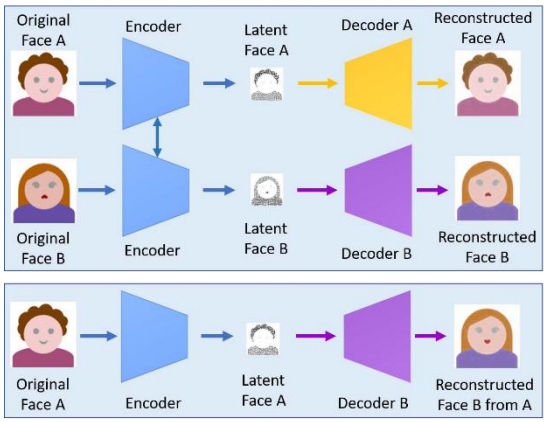


Fig. 3. This deepfake production approach uses two pairs of encoders and decoders.. For the training process (top), two networks utilize the same encoder but distinct decoders. To build a deepfake, a picture of face A is encoded using the common encoder and decoded using decoder B.

Through the incorporation of adversarial loss and perceptual loss, which are integrated into VGGFace, into the encoder-decoder architecture, faceswap-GAN—an enhanced variant of deepfakes—is created.

was suggested in. In order to smooth out segmentation mask artifacts and produce eye movements that are more realistic and consistent with input faces, VGGFace perceptual loss is used. This results in higher quality output movies. The generation of outputs with resolutions of 64x64, 128x128, and 256x256 is made easier by this approach.Furthermore, to improve face detection and face alignment, the multi-task convolutional neural network (CNN) from the FaceNet implementation is employed. In this model, the generative network is implemented using the CycleGAN.[20-21]

As shown in Fig. 4, a traditional GAN model consists of two neural networks: a discriminator and a generator. The goal of the generator G is to create images G(z) that are similar to genuine images x given a dataset of real images x with a distribution of pdata, while z are noise signals with a distribution of pz. The discriminator G's

Table 1 : Summary of notable deepfake tools

Tools	Links	Key Features
Faceswap	https://github.com/deepfakes/faceswap	- Using two encoder-decoder pairs. - Parameters of the encoder are shared.
Faceswap-GAN	https://github.com/shaoanlu/faceswap-GAN	Adversarial loss and perceptual loss (VGGface) are added to an auto-encoder architecture.
Few-Shot Face Translation	https://github.com/shaoanlu/fewshot-face-translation-GAN	- Use a pre-trained face recognition model to extract latent embeddings for GAN processing. - Incorporate semantic priors obtained by modules from FUNIT [41] and SPADE [42].
DeepFaceLab	https://github.com/iperov/DeepFaceLab	- Expand from the Faceswap method with new models, e.g. H64, H128, LIAEF128, SAE [43]. - Support multiple face extraction modes, e.g. S3FD, MTCNN, dlib, or manual [43].
DFaker	https://github.com/dfaker/df	- DSSIM loss function [44] is used to reconstruct face. - Implemented based on Keras library.
DeepFake_tf	https://github.com/StromWine/DeepFake_tf	Similar to DFaker but implemented based on tensorflow.
AvatarMe	https://github.com/lattas/AvatarMe	- Reconstruct 3D faces from arbitrary "in-the-wild" images. - Can reconstruct authentic 4K by 6K-resolution 3D faces from a single low-resolution image [45].
MarioNETte	https://hyperconnect.github.io/MarioNETte	- A few-shot face reenactment framework that preserves the target identity. - No additional fine-tuning phase is needed for identity adaptation [46].
DiscoFaceGAN	https://github.com/microsoft/DiscoFaceGAN	- Generate face images of virtual people with independent latent variables of identity, expression, pose, and illumination. - Embed 3D priors into adversarial learning [47].
StyleRig	https://gvv.mpi-inf.mpg.de/projects/StyleRig	- Create portrait images of faces with a rig-like control over a pretrained and fixed StyleGAN via 3D morphable face models. - Self-supervised without manual annotations [48].
FaceShifter	https://lingzhili.com/FaceShifterPage	- Face swapping in high-fidelity by exploiting and integrating the target attributes. - Can be applied to any new face pairs without requiring subject specific training [49].
FSGAN	https://github.com/YuvalNirkin/fsgan	- A face swapping and reenactment model that can be applied to pairs of faces without requiring training on those faces. - Adjust to both pose and expression variations [50].
StyleGAN	https://github.com/NVlabs/stylegan	- A new generator architecture for GANs is proposed based on style transfer literature. - The new architecture leads to automatic, unsupervised separation of high-level attributes and enables intuitive, scale-specific control of the synthesis of images [51].
Face2Face	https://justusthies.github.io/posts/face2face/	- Real-time facial reenactment of monocular target video sequence, e.g. Youtube video. - Animate the facial expressions of the target video by a source actor and re-render the manipulated output video in a photo-realistic fashion [52].
Neural Textures	https://github.com/SSRSGJYD/NeuralTexture	- Feature maps that are learned as part of the scene capture process and stored as maps on top of 3D mesh proxies. - Can coherently re-render or manipulate existing video content in both static and dynamic environments at real-time rates [53].
Transformable Bottleneck Networks	https://github.com/kyleolsz/TB-Networks	- A method for fine-grained 3D manipulation of image content. - Apply spatial transformations in CNN models using a transformable bottleneck framework [54].
"Do as I Do" Motion Transfer	github.com/carolineec/EverybodyDanceNow	- Automatically transfer the motion from a source to a target person by learning a video-to-video translation. - Can create a motion-synchronized dancing video with multiple subjects [55].
Neural Voice Puppetry	https://justusthies.github.io/posts/neural-voice-puppetry	- A method for audio-driven facial video synthesis. - Synthesize videos of a talking head from an audio sequence of another person using 3D face representation. [56].

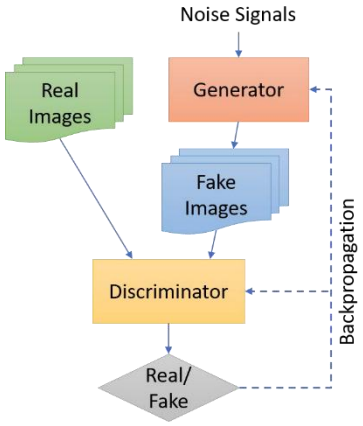


Fig 4 A neural network can be used to implement either of the two components of the GAN architecture, which are the generator and the discriminator.

$$\text{AdaIN}(x_i, y) = y_{s,i}\frac{x_i - \mu(x_i)}{\sigma(x_i)} + y_{b,i}$$

by G and actual photos x. To increase its capacity for classification, discriminator D is trained to maximize D(x), which denotes the likelihood that x is

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[\log(1 - D(G(z)))]$$

an actual image rather than a fictitious one created by G. Conversely, G is trained to minimize 1 − D(G(z)), or the likelihood that D would classify its outputs as synthetic images. This minimax game between players D and G can be explained by the value function that follows : min G max

Both networks get more proficient after receiving enough training; for example, generator G can create images that are strikingly comparable to genuine images, while discriminator D is highly capable of distinguish fake image from real ones .

A list of common deepfake tools and their characteristic properties is provided in Table 1. Among these is StyleGAN, a well-known face synthesis technique based on a GAN model that was first shown in. Style transfer drives StyleGAN, a unique generator network architecture capable of producing realistic face images. In a conventional GAN model, such as the progressive growing of GAN (PGGAN), the generator is represented by the feedforward network, whose input layer receives the signal noise (latent code).

A mapping network (f) and a synthesis network (g) are the two networks that are built and connected together in StyleGAN. A neural network, also known as the mapping network, made up of several completely connected layers is the characteristic of the non-linear function f: Z → W, which first converts the latent code z ∈ Z to w ∈ W (where W is an intermediate latent space). The intermediate representation w is specialized to styles $y = (y_s, y_b)$ using an affine transformation. These styles will be supplied to the adaptive instance normalization (AdaIN) operations, which are as follows:

where each feature map x_i undergoes independent normalization. Through the use of several sizes, the StyleGAN generator architecture enables control over the synthesis of images. Furthermore, this method generates a predetermined percentage of images utilizing two latent codes during training, as opposed to one random latent code. More specifically, the mapping network receives two latent codes, z_1 and z_2, which it uses to construct w_1 and w_2, respectively, which govern the styles by applying w_1 before and w_2 after the crossover point. Examples of graphics created by combining two latent codes at three different scales are shown in Fig. 5, where each group of styles regulates distinct high-level features that have meaning. Stated differently, StyleGAN's generator architecture may recognize the distinction between typically trained on human faces, such as position and identity), and it allows for intuitive, scale-specific control of the face synthesis [20-24].

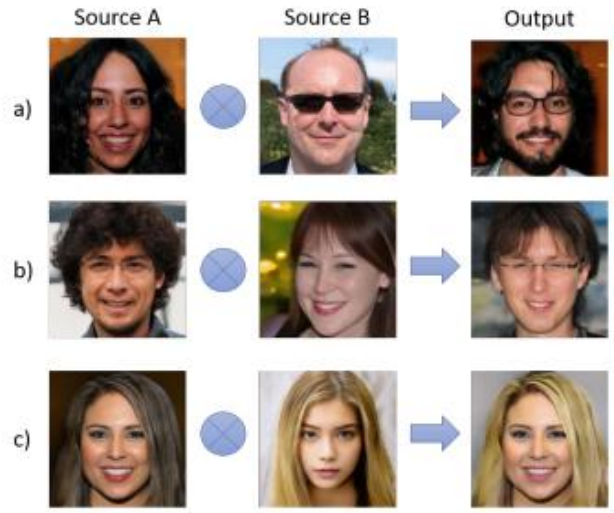


Fig 5 When using StyleGAN to blend styles, if one copies specific details from source B, the resulting images will retain source A's general characteristics and pose while acquiring B's color scheme and microstructure.

Deepfake Detection

Deepfake detection is commonly understood as a binary classification problem, in which altered and real videos are classified using classifiers. To train classification models, these kinds of approaches need a sizable collection of authentic and fraudulent movies. Although there are more and more false movies online, there are still not enough of them to provide a standard for validating different detection techniques. Korshunov and Marcel used the free source Faceswap-GAN code to create a significant deepfake dataset of 620 videos based on the GAN model in order to address this problem.movies from the VidTIMIT database, which is accessible to the public, were utilized to create deepfakes of varying quality. These movies can accurately replicate lip motions, eye blinking, and facial expressions [25-26].

Afterwards, different deepfake detection techniques were tested using these films. According to test results, popular face recognition systems built on VGG and Facenet are not able to identify deepfakes well. When used to detect deepfake videos from this recently created dataset, other techniques like lip-syncing approaches and image quality measurements with support vector machines (SVM) result in extremely high error rates. This raises questions regarding the urgent need to create more reliable techniques in the future that can distinguish between real and deepfakes [27-28].

This Section Present a survey of deepfake detection methods of fake video detection. It have smaller groups visual artifacts within single video frame based methods and temporal features. While most of the methods based on temporal features use deep learning.

Fake Video Detection

Recent years have seen a major advancement in video interpretation thanks to the development of improved models and bigger datasets . The emphasis on temporal modeling, which is thought to be the primary distinction between videos and images, is a recurring element in most techniques. The aforementioned encompasses several studies on low-level motion , temporal structure, long/short term interdependence , and modeling the action as a series of events/states.[29]

More precisely, state-of-the-art outcomes are attained by a wide range of deep learning architectures that seek to capture low-level motion through temporal convolutions. Motion-based action recognition has also been supported by hand-crafted features such as iDT. The real effect of simulating low-level motion is still unknown, though. One could contend that the scene and items in a frame are nearly adequate to infer the action, as illustrated in Fig. This theory is somewhat supported by the ability to recreate motion in a movie by matching deep features from a C3D model. We see that the network's pool-5 layer preserves all of the spatial information in the video but loses any discernible motion. Inspired by these observations, we carry out a thorough quantitative and qualitative investigation of motion's impact on video action identification.

We apply the popular 3D convolution model on the UCF101 and Kinetics video datasets as examples of our investigation. The most recent large-scale dataset created specifically for classification is called Kinetics, and UCF101 has long served as the industry standard benchmark for comparing and evaluating video models. Although 3D convolution has become the industry standard for comprehending videos, the suggested frameworks (generator and frame selector) are versatile and can be applied to any type of video model.

Temporal modeling for action recognition

Temporal modeling for action recognition: The main distinction between image and video models has been the focus on modeling the temporal information in a video.

Low-level motion, long- and short-term dependencies, temporal structure, representing the action as a series of events or states, and temporal pooling strategies are all included in this. It can be challenging to determine whether the models are indeed capturing motion information and whether motion is actually necessary for identifying action in existing video datasets because these methods are frequently assessed based on overall performance [30].

Model Analysis

Module 1 : Data Set Gathering

to improve the model's real-time prediction efficiency. Data from many publicly available data-sets, including Face Forensic++(FF)[1], Deepfake detection challenge (DFDC)[2], and Celeb-DF[3], have been collected by us. Moreover, we combined the previously gathered datasets to produce a brand-new dataset that allows for precise and quick detection of various types of films. We have taken into consideration 50% real and 50% fake videos in order to prevent the model's training bias.

The audio-alerted videos in the Deep Fake Detection Challenge (DFDC) dataset [3] are specific examples; audio deepfakes are not covered in this study. Using a Python script, we preprocessed the DFDC dataset and eliminated the audio-altered films from it.

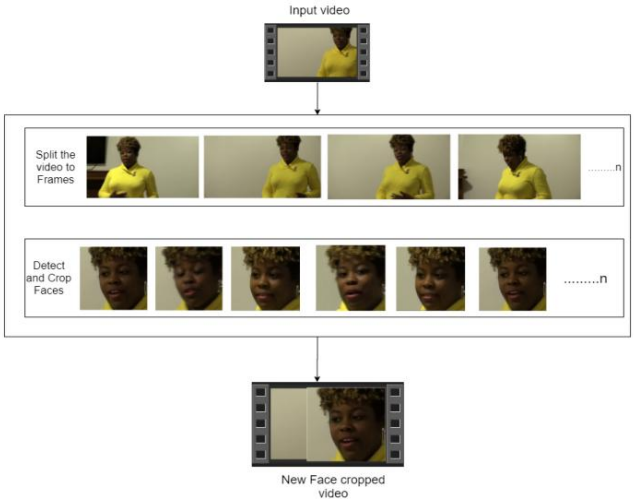


Fig 7 pre-processing of video

We extracted 1500 Real and 1500 Fake videos from the DFDC dataset after preprocessing it. The Face Forensic++(FF)[1] dataset has 1000 Real and 1000 Fake videos, while the CelebDF dataset contains 500 Real and 500 Fake videos. which means that there are 6000 total videos in our dataset—3000 real, 3000 fraudulent, and 3000 real. The distribution of the data sets is shown in Figure [31-32].

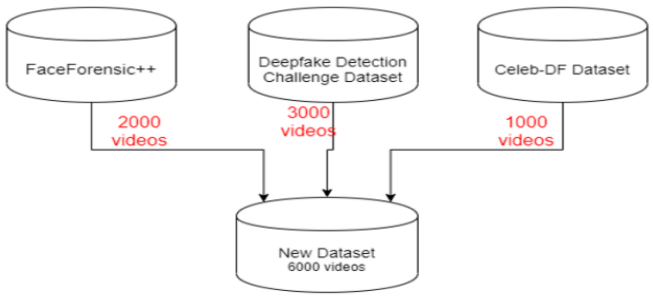


Fig 6 Dataset

Module 2 : Pre-processing

The videos undergo preprocessing, whereby all unnecessary elements and noise are eliminated. Just the necessary area of the video—the face—is identified and clipped. Dividing the video into frames is the first stage in the preparation process. The video is divided into frames, and each frame is cropped along the face once the face is identified in each frame. Afterwards, every frame of the video is combined to create a new video from the

cropped frame. Every video undergoes the same procedure, which results in the production of a processed dataset with just face videos. During preprocessing, the frame without the face is ignored. In order to keep the quantity of frames consistent, we have chosen threshold value based on the mean of total frames count of each video, and another reason of choosing it is to limit computational power

Module 3: Data set Split

The dataset is divided into train and test subsets, containing 4,200 train videos and 1,800 test videos, respectively. The ratio between the train and test is balanced; that is, 50% ofin each split, 50% bogus and 50% real videos.

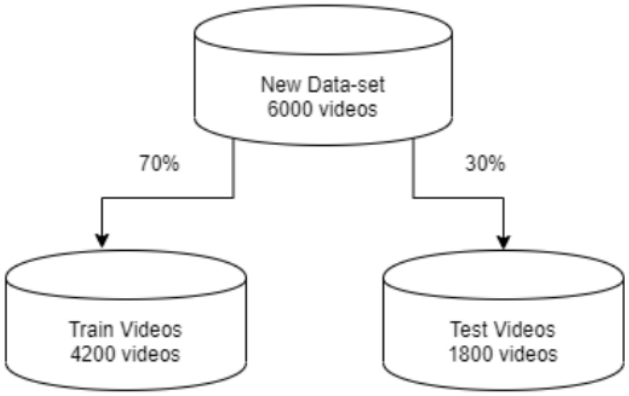


Fig 8 Train Test Split

Module 4 : Model Architecture

CNN and RNN are combined to create our model. After extracting the characteristics at the frame level using the Pre-trained ResNext CNN model, an LSTM network is trained to distinguish between pristine and deepfake videos. The labels of the training split of movies are loaded and fitted into the model for training using the Data Loader.

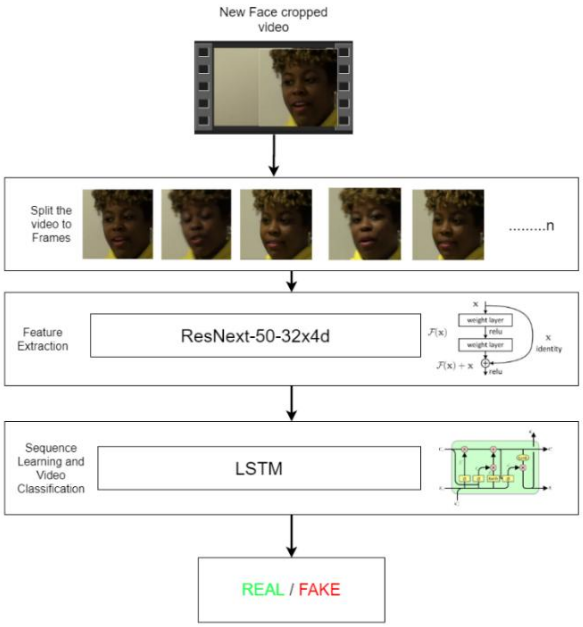
ResNext : For feature extraction, we utilized ResNext's pre-trained model rather than creating the code from scratch. The Residual CNN network, or ResNext, is designed to perform well on deeper neural networks. To conduct the experiment, we have employed the Resnext50_32x4d model. A ResNext with 50 layers and 32 x 4 dimensions has been utilized.The network will then be adjusted by adding more necessary layers and choosing the right learning rate to ensure that the model's gradient descent converges correctly. The sequential LSTM input is the 2048-dimensional feature vectors that follow the last pooling layers of ResNext.

LSTM for sequence Processing : The LSTM is fitted with 2048-dimensional feature vectors as input. One LSTM layer with 2048 latent dimensions, 2048 hidden layers, and 0.4 is what we are using. likelihood of dropout, which can help us accomplish our goal. The Department of Computer Engineering, GHRCEM-Wagholi, Pune, uses LSTM for the 2019–2020 28

By comparing the frame at "t" seconds with the frame of "t-n" seconds, Deepfake Video Detection sequentially processes the frames to enable temporal analysis of the video.where n is the number of frames that come before t.

The Leaky Relu activation function is another component of the model. To enable the model to learn the average rate of correlation between the input and output, a linear layer of 2048 input features and 2 output features is employed. The model makes use of an adaptive average polling layer with an output parameter of 1. It provides the intended output size of the H x W image. A sequential layer is used to process the frames in a consecutive manner. The batch training is carried out with a batch size of 4. The model's prediction confidence is obtained using a SoftMax layer [31-32].

Fig 9 Overview of the model



Model 5 : Hyper-Parameter tuning

It involves selecting the ideal hyper-parameters to attain the highest level of accuracy. following numerous iterations of the model. We select the hyper-parameters that work best for our dataset. To activate the Adamoptimizer with adaptive learning rate

using the model's parameters. The Department of Computer Engineering, GHRCEM-Wagholi, Pune, has set the learning rate to 1e-5 (0.00001) for the 2019–2020 academic year. Deepfake Video Detection produces a higher gradient descent global minimum. One weight decay, 1e-3, is employed [31-32].

Since this is a classification problem, the loss cross entropy method is applied to compute the result.Batch training is utilized in order to make the best use of the available processing capacity. Four is the batch size that is used. Four is the tested batch size that works best for training in our development environment.

Future Research Direction

Every system that is produced has room for improvement, especially when it is built with the newest, most popular technology and has a promising future.

Web Based Platform to Browser Plugin : Web based platform can be upscaled for browser plugin so it will be easy for user to access it

Full Body Deepfake Detection : currently the algorithm is used only for face, but the algorithm can enhanced in detecting full body deep fakes.

Privacy-Preserving Deepfake Detection : Introduce privacy preserving techniques that enable deepfake detection without compromising the privacy of individual

Real time Detection : Develop a real time deepfake detection that can identify manipulated it as soon as uploaded or stramed

Continuous Model Improvement : Implement mechanism for continuously updating and improving deepfake detection over time.

Conclusion

People's confidence in media content has started to decline as a result of deepfakes because believing in them no longer equates to seeing them. They might be upsetting and adverse consequences on the people who are the targets, increase hate speech and disinformation, and even ignite political unrest, public agitation, violence, or conflict. These days, this is especially important because deepfake technologies are becoming more accessible and social media sites can swiftly disseminate the fake content.

This review offers a current summary of deepfake creation and detection techniques along with a thorough discussion of the difficulties,

probable trends, and upcoming developments in this field. For the purpose of creating practical strategies to combat deepfakes, the artificial intelligence research community will find great value in this study.

By processing one second of video (10 frames per second), our project technique can accurately forecast the result. Using an LSTM for temporal sequence processing to identify changes between the t and t-1 frame and a pre-trained ResNext CNN model to extract frame-level features, we developed the model. The video in frame sequences of 10, 20, 30, 40, 60, 80, and 100 can be processed by our model.

REFERENCE :

1. Shruti Agarwal, Hany Farid, Yuming Gu, Mingming He, Koki Nagano, and Hao Li. Protecting world leaders against deep fakes. In *Computer Vision and Pattern Recognition Workshops*, volume 1, pages 38–45, 2019.
2. Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre- Antoine Manzagol. Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th Inter-national Conference on Machine learning*, pages 1096–1103, 2008.
3. Diederik P Kingma and Max Welling. Auto-encoding varia- tional Bayes. *arXiv preprint arXiv:1312.6114*, 2013.
4. Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in Neu-ral Information Processing Systems*, 27:2672–2680, 2014.
5. Alireza Makhzani, Jonathon Shlens, Navdeep Jaitly, Ian Good- fellow, and Brendan Frey. Adversarial autoencoders. *arXiv preprint arXiv:1511.05644*, 2015.
6. Ayush Tewari, Michael Zollhoefer, Florian Bernard, Pablo Garrido, Hyeongwoo Kim, Patrick Perez, and Christian Theobalt. High-fidelity monocular face reconstruction based on an unsupervised model-based face autoencoder. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(2):357–370, 2018.
7. Jiacheng Lin, Yang Li, and Guanci Yang. FPGAN: Face de- identification method with generative adversarial networks for social robots. *Neural Networks*, 133:132–147, 2021.
8. Ming-Yu Liu, Xun Huang, Jiahui Yu, Ting-Chun Wang, and Arun Mallya. Generative adversarial networks for image and video synthesis: Algorithms and applications. *Proceedings of the IEEE*, 109(5):839–862, 2021.
9. Siwei Lyu. Detecting 'deepfake' videos in the blink of an eye. http://theconversation.com/detecting–deepfake– videos- in- the-blink- of- an- eye- 101072, August2018.
10. Bloomberg. How faking videos became easy and why that'ssoscary. https://fortune.com/2018/09/11/deep– fakes-obama-video/, September 2018.
11. Robert Chesney and Danielle Citron. Deepfakes and the new disinformation war: The coming age of post-truth geopolitics. *Foreign Affairs*, 98:147, 2019.
12. Xinyi Zhou and Reza Zafarani. A survey of fake news: Fun- damental theories, detection methods, and opportunities. *ACM Computing Surveys (CSUR)*, 53(5):1–40, 2020.
13. Rohit Kumar Kaliyar, Anurag Goswami, and Pratik Narang. Deepfake: improving fake news detection using tensor decomposition-based deep neural network. *The Journal of Su-percomputing*, 77(2):1015–1037, 2021.
14. Bin Guo, Yasan Ding, Lina Yao, Yunji Liang, and Zhiwen Yu. The future of false information detection on social media: New perspectives and trends. *ACM Computing Surveys (CSUR)*, 53(4):1–36, 2020.
15. Patrick Tucker. The newest AI-enabled weapon: 'deep-faking'photosofthe earth. https://www.defenseone.com/technology/2019/03/next- phase- ai- deep- faking- whole-world-and-china-ahead/155944/, March 2019.
16. T Fish. Deep fakes: AI-manipulated media will be 'weaponised' to trick military. https://www.express. co . uk / news / science / 1109783 / deep – fakes –ai – artificial – intelligence – photos – video – weaponised-china, April 2019.
17. Abhijith Punnappurath and Michael S Brown. Learning raw image reconstruction-aware deep image compressors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(4):1013–1019, 2019.
18. Zhengxue Cheng, Heming Sun, Masaru Takeuchi, and Jiro Katto. Energy compaction-based image compression using convolutional autoencoder. *IEEE Transactions on Multimedia*,22(4):860–873, 2019.
19. Jan Chorowski, Ron J Weiss, Samy Bengio, and Aäron van den Oord. Unsupervised speech representation learning using WaveNet autoencoders. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 27(12):2041–2053, 2019.
20. Keras-VGGFace: VGGFace implementation with Keras framework. https:// github. com/ rcmalli/ keras- vggface.
21. Faceswap-GAN. https : / / github . com / shaoanlu /faceswap-GAN, .
22. Tero Karras, Samuli Laine, and Timo Aila. A style-based gen- erator architecture for generative adversarial networks. In *Pro- ceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4401–4410, 2019.
23. Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kil- ian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4700–4708, 2017.
24. Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of GANs for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*, 2017.
25. Pavel Korshunov and Sébastien Marcel. Vulnerability assess- ment and detection of deepfake videos. In *2019 International Conference on Biometrics (ICB)*, pages 1–6. IEEE, 2019.
26. VidTIMIT database. http://conradsanderson.id.au/vidtimit/.

27. Omkar M Parkhi, Andrea Vedaldi, and Andrew Zisserman. Deep face recognition. In *Proceedings of the British Machine Vision Conference (BMVC)*, pages 41.1–41.12, 2015.
28. Florian Schroff, Dmitry Kalenichenko, and James Philbin. FaceNet: A unified embedding for face recognition and clus- tering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 815–823, 2015.
29. Joon Son Chung, Andrew Senior, Oriol Vinyals, and An- drew Zisserman. Lip reading sentences in the wild. In *2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, pages 3444–3453. IEEE, 2017.
30. Darius Afchar, Vincent Nozick, Junichi Yamagishi, and Isao Echizen. MesoNet: a compact facial video forgery detection network. In *2018 IEEE International Workshop on InformationForensics and Security (WIFS)*, pages 1–7. IEEE, 2018.
31. What Makes a Video a Video: Analyzing Temporal Information in Video Understanding Models and Datasets. De-An Huang1 , Vignesh Ramanathan2 , Dhruv Mahajan2 , Lorenzo Torresani2,3, Manohar Paluri2 , Li Fei-Fei1 , and Juan Carlos Niebles1 1Stanford University, 2Facebook, 3Dartmouth Colleg IEEE xplore
32. Deepfake Video Detection using Neural Networks http://www.ijsrd.com/articles/IJSRDV8I10860.pdf
33. International Journal for Scientific Research and Development http://ijsrd.com/
34. Deepfake Detection: A Systematic Literature Review MD SHOHEL RANA 1,2, (Member, IEEE), MOHAMMAD NUR NOBI3 , (Member, IEEE), BEDDHU MURALI2 , AND ANDREW H. SUNG2 , (Member, IEEE)
35. Deepfake detection using deep learning methods: A systematic and comprehensive review Arash Heidari 1 | Nima Jafari Navimipour 2,3 | Hasan Dag4 | Mehmet Unal 5