JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Disclosing Fake Faces Using Deep Neural Networks

[1]Nisha R, [2]Dr.K R Shylaja,

[1]Student, MTech, [2]Professor,
[1,2]Department of Computer Science,
[1,2]Dr. Ambedkar Institute of Technology, Bangalore, Karnataka, India

Abstract: Creating personalized photo-realistic talking head models or that are able to generate reliable video sequences that represent facial expressions and mimics of an individual. In this research, we provide an idea for creating talking head models with short time taken for training and only a couple of photos (so-called several shot learning). Deep learning techniques have become common place recently resulting in the generation of very realistic fake faces, raising about their possible abuse in different fields. A deep neural network architecture that combines content and trace feature extractors is used in this study to present a unique method for identifying and exposing fraudulent faces. Actually, our system can produce an acceptable result after learning from just one photo (one-shot learning) and adding a few more photos to improves the personalization's quality. According to this, our design talking heads are deep ConvNets that directly generate video frames through a series of convolutional neural networks rather than by warping them. A technology when dealing with few-shot capabilities is what we give us. It does extensive meta-learning on a huge dataset. The content extractor focuses on extracting high level semantic information and determining patterns that can tell real face characteristics apart from false ones. At the same time, the data extraction explores minute imperfections and discrepancies in facial patterns through the advantage of the trace left behind the generating process.

***Index Terms* - Deep Learning, Fake Faces, Convolutional Neural Network, Detection**

I. INTRODUCTION

Deep Learning is also called as the deep structural learning which is part of a border family of the machine learning methods. Learning can be of supervised learning, semi-supervised learning, and unsupervised learning. Deep learning Applications which are emerged as the Deepfake. Deepfake which allows for the automatic generation of creating the fake video content by using and Deep fake technology. Deepfake technology is a controversial technology which has many wide-reaching issues which has impacted the society for e.g., election biasing, cyber bullying. In this project we have proposed an integrated system with face forensic model which mashes up the conventional image forensic approach and the fake face image forensic approach. The system where we can detect and manipulated or altered the media with convolutional approaches. Convolutional neural networks are containing of two types that is a feature extractor and the trace features from a face images. The feature extractors are trained by transferring and fine-tuning models to the pretrained object recognition model. The extracted features are specialized to represent various contents in the face. The feature extractor are based on the local relationship between neighboring pixels by applying the multi-channel constrained convolution. This is an extended version of the single channel constrained convolution. The input images to be obtain the content excluded image and extract the features hierarchically. When the content is excluded, the color and the contrast of original image will get disappeared, leaving the outlines and some traces which verify the fake face detection performance of the model, we conducted experiments with facial image datasets manipulated by Deep fake and Face2Face algorithms. The model shows the higher detection accuracy and robustness of various video compression levels than the existing base line models. The technology that can create believable videos of speech expressions and mimic of an individual in the creation of customized photo-realistic talking head models. More particularly, we have thought about the challenging process of creating photo-realistic individualized head pictures from a collection of facial landmarks, which has been used to drive the model's motion. Such a capability has real-world uses in the telepresence, such as videoconferencing and multiplayer gaming, in addition to the special effects industry.

II. DEEP FAKE CREATION

Fake Face Creation is identified as the highly realistic human head images generated when training the Convolutional Neural Networks are known as fake faces. These tasks must be performed to train large dataset for one person in need to produce the customized talking head models. However, in many real-world circumstances.

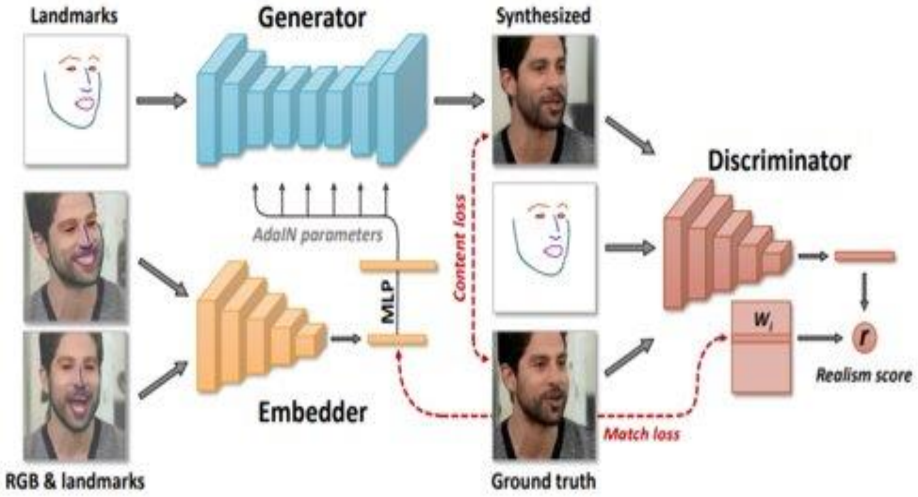


Figure 1. Block Diagram of Fake Creation

III. DEEP FAKE DETECTION

The basic architecture is to produce deepfake is encoder-decode architecture, where the encoder acquires target and the source face, and the decoder is to get encoding target face and then generate face video. Using high level processing, and the leftovers are removed but still few traces are left which are not visible by naked eye. These leftovers traces are the key features of our detection model. This model comprises of incetionResnetV2 for feature extraction. These features are used to train a recurrent neural network to analysis if the video has been put through manipulation or not. Only a small portion of video is manipulated which means the deepfakes are shorter in time, therefore, the video is spilt into small frames and these frames are given to input detection model.

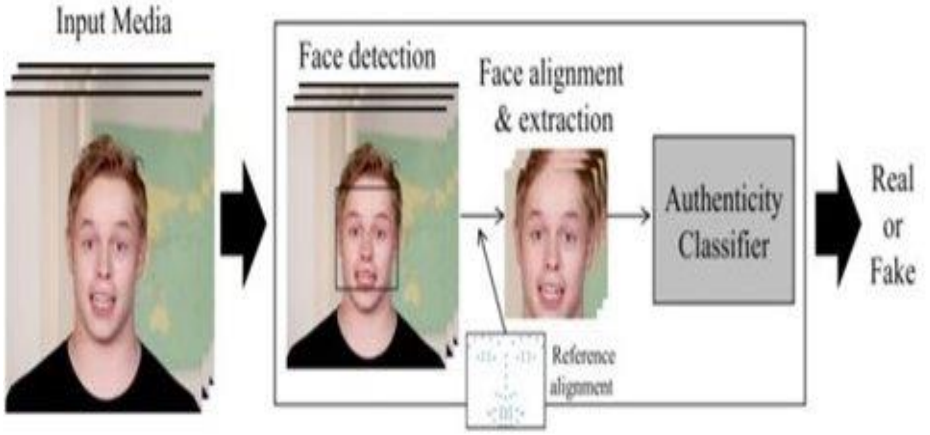


Figure 2. Block Diagram of Fake Detection

IV. LITERATURE SURVEY

Jee-Young Sun, Seung-Wook [1] Using comparatively basic handmade characteristics determined by two different order statistics, a paper that used Contrast Enhancement (CE) forensic methods may be done, although these techniques have trouble spotting contemporary counter forensics assaults. The Gray-level Co-Occurrences Matrix (GLCM), which provides traceable characteristics for CE forensics, is introduced into CNN in this study to assist in the process. The experimental findings shows that the suggested strategy, especially when dealing with counter-forensic attacks, outperforms traditional forensics methods in terms of forgery detection accuracy, the alternate and the more recent techniques can be used to generate the present generation of computer vision technology for fake faces.

A. Bromme, C. Busch, A.Dantcheva, C. Rathgeb, A.Uhl [2] Local Binary Patterns (LBP) and a set of the CNN based system that are to be considered. For this system we have used CNN and the CNN architecture are AlexNet, VGG19, ResNet50, Xception, and GoogleNet/Inceptionv3.The networks that perform better in detecting the CGI compared to the contents generated by FakeApp, or through the other techniques. These results that indicates event though the networks were not trained to detect CGI specifically, the are still somewhat effective for detecting of CGI videos.

Luisa Verdoliva, Christian Riess, Justus Thies, Matthias Niebner [3] With reference to the recent works, it is now possible to seamlessly generate manipulated images or videos using real-time technologies like image morphing, Snap-chat, and Face2Face. Considering the types of manipulation that is source-to-target, Self-reenactment.

Tong Che, Zoubin Grahahramani, Yoshua Bengio, Yangqiu Song [4] MetaGAN: It is defined as the adversarial Approach to few-shot of learning which presents MetaGAN as a general and flexible framework for few-shot learning also proposes MetaGAN and generic frameworks to boost the performance of few-shot learning models. The few learning problems are considered to be as the MetaGAN and help in processing the simple and the general frameworks.

Antreas Antoniou, Amos Strorkey, Harrison Edwards [5] DAGAN is identified as the (Data Augmentation Generative Adversarial Networks (DAGAN)) enables effective neural networks training even in low-data target domains. As the DAGAN is independent, it captures the cross-class transformations, moving data-points to other points of equivalent class. These three datasets used are The Omniglot Dataset, The EMNIST Dataset and the more complex VGG-Face Dataset.

DAGAN and then evaluate its performance on low-data target domains using. Standard stochastic-gradient neural networks training, and Specific one-shot meta-learning methods. DAGAN helps in the performance on low-data target domains using. Standard stochastic-gradient neural networks training and Specific one-shot meta-learning methods.

Nguyen Tran, Thanh-Toan Do, Ngai-Man Cheung [6] celeb-DF describes the dataset to deepfake detection methods. We have used the dataset as curated diverse which is manipulated videos and the deep neural network including CNNs, and with the representation of benchmarking and the evaluation of the deepfake methods.

V. METHODOLOGY

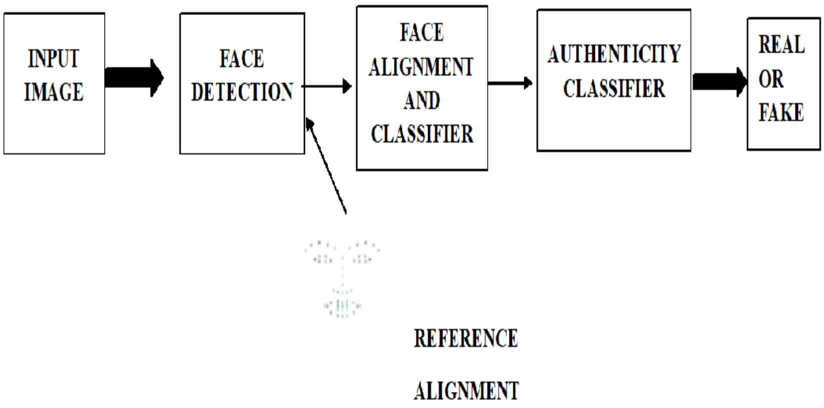


Figure 3. Fake Face Detection Framework

The Fake Face Detection framework consists of:

A. **Face Detection:** By taking an input image, the facial region is detected by a neural facial landmarks detection model that automatically localizes the facial components and facial contours such as eyes, mouth, and chin. Among those points, only 51 points are used excluding 17 points from chin because facial manipulation is performed inside the inner facial region.

B. **Face Alignment and Extraction**: The system aligns the face to appropriate the reference alignment because faces appearing in media are rarely frontal or unrotated. We apply the affine conversion on the image by finding the one-to-one mapping from the extracted landmark points to the reference alignment points. Through affine transformation, rotated or profile faces can be aligned according to the reference alignment, which helps to enhance the fake faces detection performance. Finally, the system crops the facial region from the images and feed it to facial authenticity classifier.

C. **Authenticity Classifier:** This proposed face authenicity classifier combines of content feature extractor (CFE) and the trace feature extractor (TFE). A convolution is depicted of containing its detail in the two feature extractors.

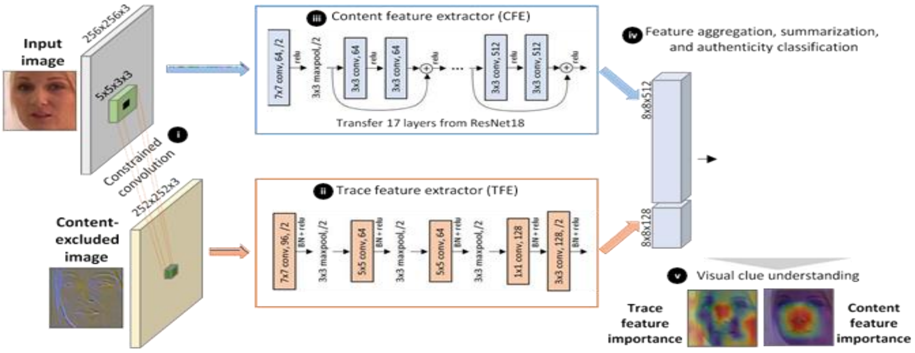


Figure 4. Face Authenticity Classifier Combines Content Feature Extractor

VI. IMPLEMENTATION

Photoshop that are affected and utilized each day by experts, yet that doesn’t imply that simply introducing both is everything necessary to make photorealistic pictures and recordings. In like this manner, making reasonable face traded recording is hard. Like any imaginative undertaking, the conclusive outcome is a blend of ability, duly and right tools. The primary endeavor of deepfake creation was FakeApp, created by a Reddit client utilizing autoencoder-decoder blending structure. In that strategy, the autoencoder extricates inert highlights of face pictures and the decoder is utilized to remark the face pictures. To trade faces between sources pictures what’s more, target pictures, there is a necessity of two encoder-decoder sets where each pair is utilized to prepare on a picture set and the two system sets are shared through the encoder’s parameters. The FakeApp software uses the AI Framework, TensorFlow of Google, which in other things was at that point for the Deep Dream. There are additionally open-source options in FakeApp program, as DeepFaceLab. Face Swap (right now facilitated on GitHub),and FakeApp (as of now facilitated on Bitbucket). Regardless of whichever the application we use to make a deepfake process involves in mainly three steps: Extraction, Training, Creation.

Extraction: The deepfake comes from deep learning and deep learning requires large data sets. Thousands on different pictures are required for deepfake video. The extraction process is of extracting all frames, identifying the face and aligning them. The alignment is a critical process, the neural network is performed to swap, and all the face should have the same size.

Training: Training is a specialized term acquired from machine learning. For this situation, it alludes to the procedure which permits a neural system to change over a face into another. Although it takes a few hours, the preparation stage should be done just a single time. When finished, it can change over a face from individual A to individual B.

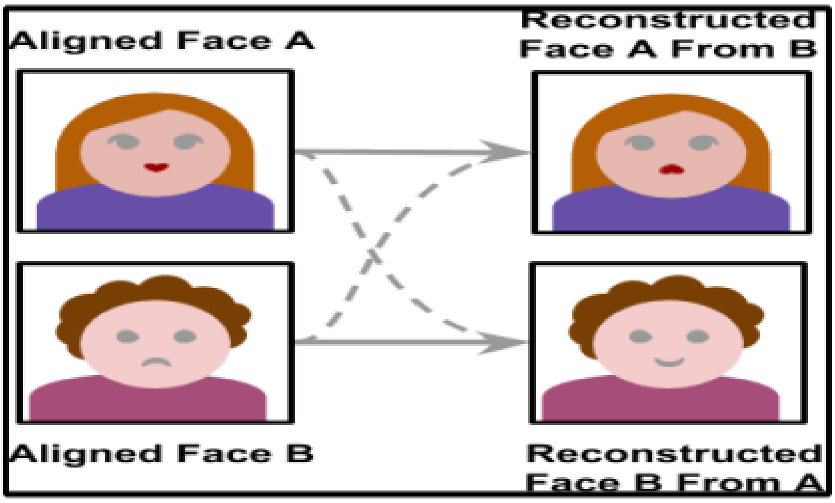


Figure 5. Shows the Training of an Image from Face A to Face B

Creation: When the training is finished, it is at last time to make a deepfake. Beginning from a video or an image, all casings are removed, and all appearances are adjusted. At that point, everyone is changed over-utilizing the prepared neural system. The last advance is to consolidate the change over the face once again into the first casing. While this seems like a simple errand, it is really where the most face-trade applications turn out badly. As already been told that autoencoders are helps to create a deepfake.

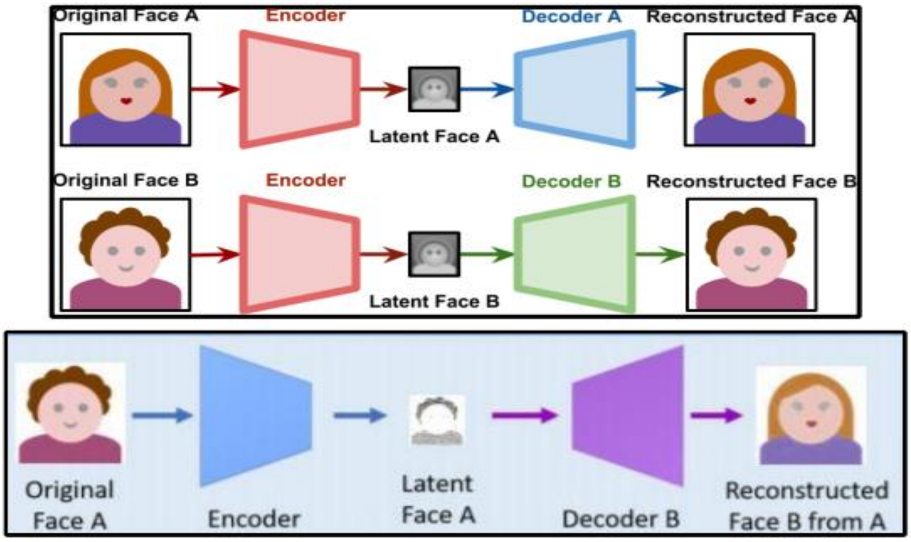


Figure 6. A deep fake creation model using two encoder-decoder pairs.

The module of the creation of fake images is a GAN. The block diagram which helps us to characterize typical GAN network. A GAN network is defined as a generator and a discriminator. During this training period, data set X is considered which includes many real images x under a distribution of pdata.

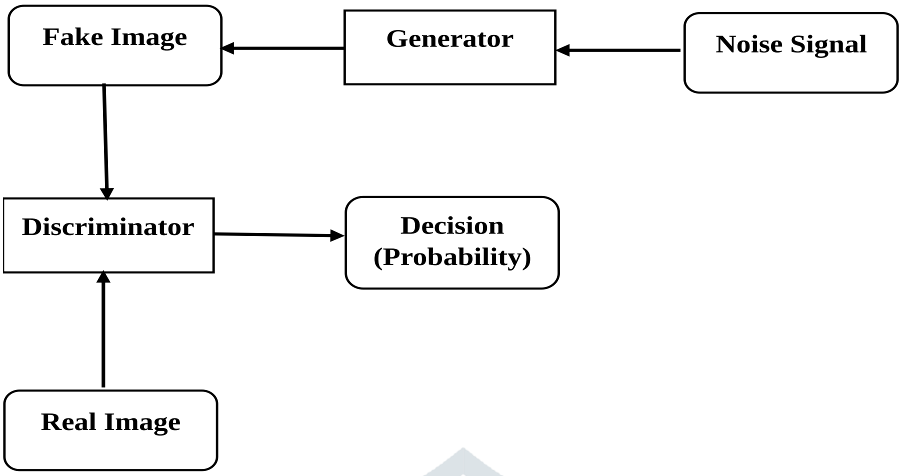


Figure 7. Generative Adversarial Networks (GAN)

6.1 Convolutional Neural Network

It is the most representative model of deep learning techniques. Each layer of CNN is known as a feature map. The feature map of the input layer is a 3D matrix of pixel intensities for the different color channels (e.g., RGB). The feature map of any internal layer is an induced multi-channel image, whose 'pixel' have a specific feature. Every neuron relates to the small portion of the adjacent neurons form the previous layer (receptive field). Different types of transformations can be conducted on features map such as filtering and the pooling. Filtering is defined as the operation which involves in the convolutes a filter matrix (learned weights) with the values of a receptive fields of neurons and takes a nonlinear function (such as sigmoid, ReLU) to obtain final responses. Next will be the Pooling operations, which consist of max pooling, average pooling, L2-pooling and local contrast normalization, summaries the response of a receptive fields into one value to produce more robust feature descriptions. Considering the interleave between the convolution and the pooling, an initial feature hierarchy is constructed, which are identified to be as the fine-tuned in a supervised manner by adding several fully connected (FC) layers in visual tasks. The final layer with different activations functions is added to get a specific conditional probability for each output neuron. The whole system are optimized on an objective function (e.g., mean squared error or cross-entropy loss). The typical VGG16 has totally 13 convolutional (conv) layers, fully connected layers, max- polling layers and a soft max classification layer. The convolutional future maps are produced by convoluting 3*3 filter windows, and future map resolution are reduced with 2 strid max -pooling layer. An arbitrary test image could be processed using a trained network. Rescaling or cropping operations are involved if the operation is needed if different size is provided.

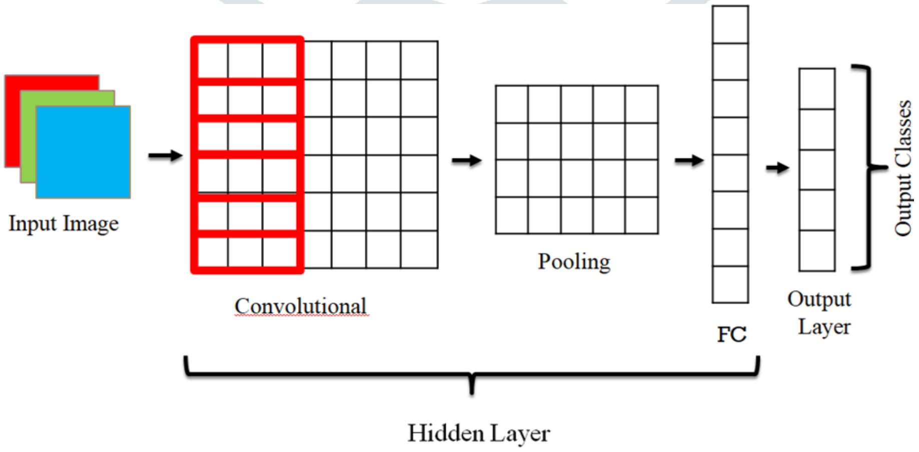


Figure 8. Typical CNN Architecture.

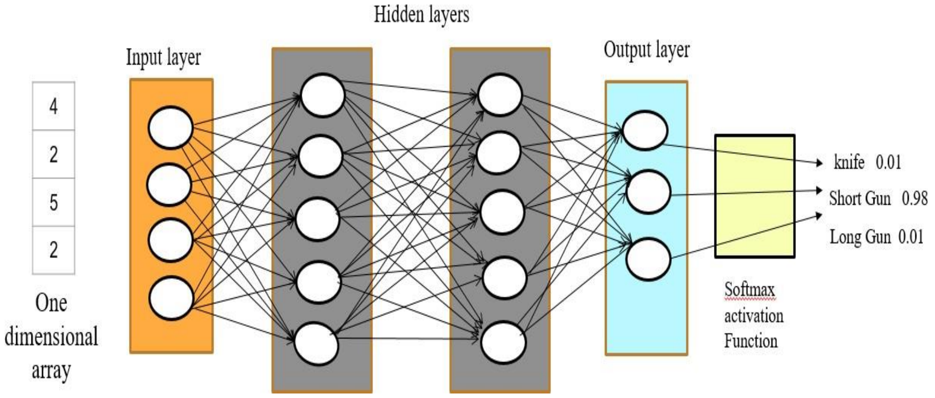
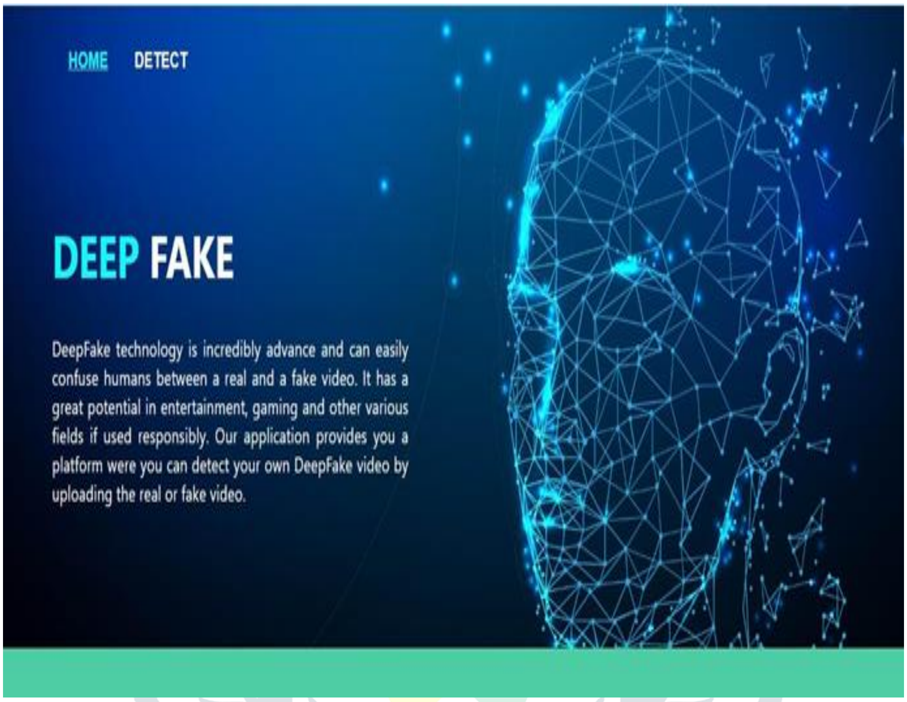


Figure 9. Fully Connected Layer and Output Layer.

VII. RESULTS AND DISCUSSION

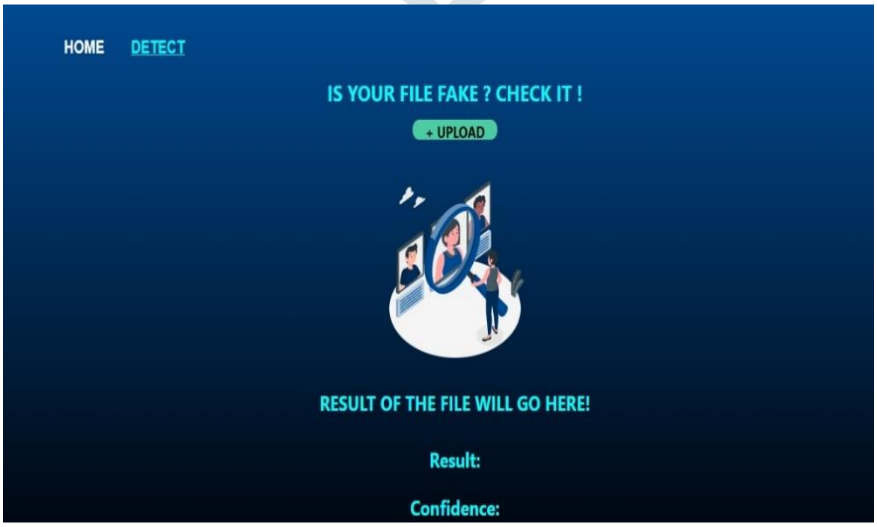


DEEP FAKE

DeepFake technology is incredibly advance and can easily confuse humans between a real and a fake video. It has a great potential in entertainment, gaming and other various fields if used responsibly. Our application provides you a platform were you can detect your own DeepFake video by uploading the real or fake video.

Figure 10. Home Page

Figure 10 shows that home page of this project. There are two buttons on this home page that is home and detect. To detect the fake faces, we need to click on the detect button we can identify whether the images are fake or real.



IS YOUR FILE FAKE ? CHECK IT !

RESULT OF THE FILE WILL GO HERE!

Result:

Confidence:

Figure 11. Detection Page

Figure 11 shows the detection page of this project. In this we have an upload button where we can upload the images and videos and it will identify fake or real and it will give how much the confidence is the image.

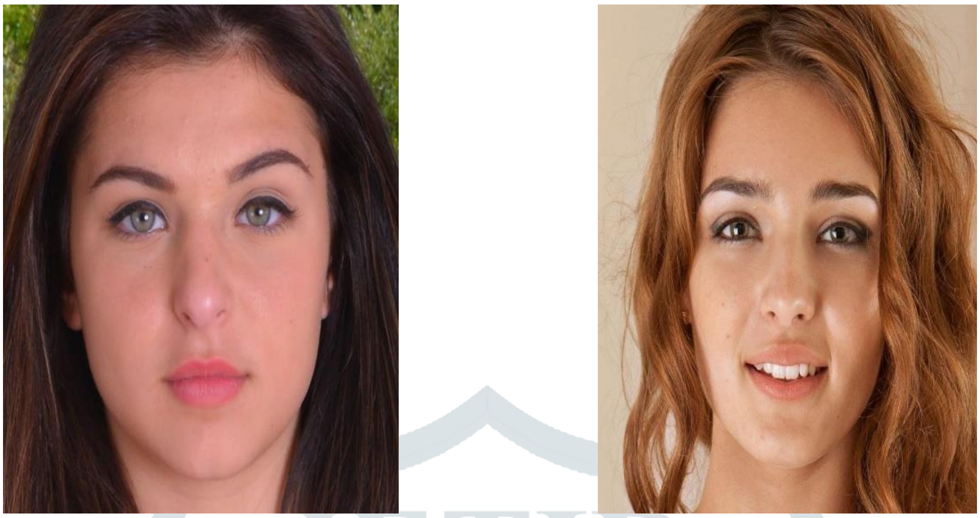


Figure 12. Input Images for Creation of Fake Image

Figure 12 shows that we have to merge the two images to create the fake image.

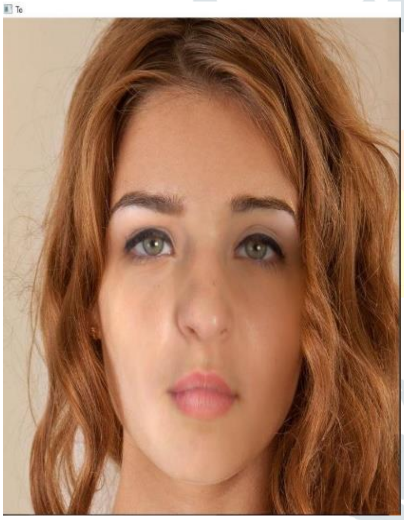
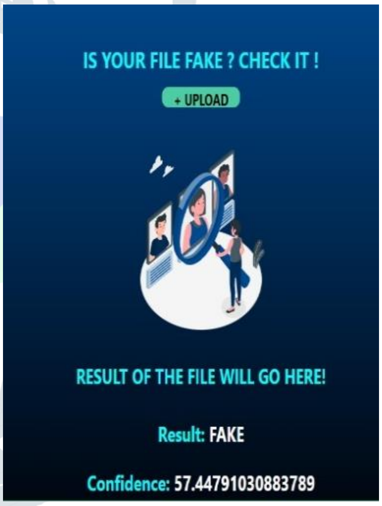


Figure 13. Creation of Fake Image



IS YOUR FILE FAKE ? CHECK IT !

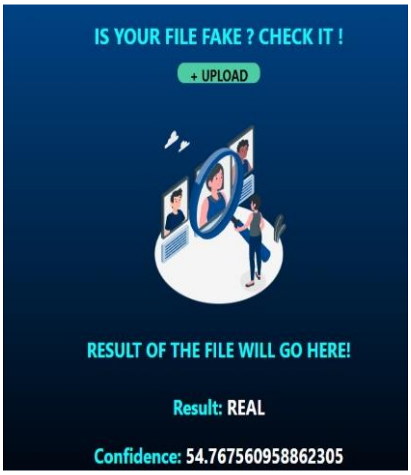
RESULT OF THE FILE WILL GO HERE!

Result: FAKE

Confidence: 57.44791030883789

Figure 14. Output of Fake Image

Figure 15. Input of Real Image



IS YOUR FILE FAKE ? CHECK IT !

RESULT OF THE FILE WILL GO HERE!

Result: REAL

Confidence: 54.767560958862305

Figure 16. Output of Real Image

VIII. CONCLUSION

Deepfakes are started to dissolve the trust of the individuals in the media substance as observing them is not at this point proportionate with putting in stock in them. They could make the pain and negative impacts those focused on, increase disinformation, and adore discourse, and even could animate political strain, excite general society, savagery, or war. This is particularly basic these days as the advancements for making the deepfakes are progressively agreeable furthermore, online networking stages can spread those phony substance rapidly. People who create deepfakes with a malicious purpose only need to deliver them to target audiences as part of their sabotage strategy without using social media. Recordings and photographs have been broadly utilized as confirmations in police examination and equity cases. We propose a well-trained system that can generate deepfakes using a one or more photographs which increases reliability of the personalization. The fake faces are mostly created for the fun, but abuse has been caused social unrest that maintains the individual privacy as well as social, political, and international security, it is imperative to develop the models that detect fake faces in the media. In the previous research can be divided into general-purpose image forensics and face image forensics. The former has been studied for the several decades and they focus on extracting images after the manipulation. which is inspired by object detection models to extract content features. The system is hybrid face forensics framework of convolutional neural network combining of the two forensics approaches that a general-purpose image forensics and face image forensics to manipulate detection performance.

IX. FUTURE ENHANCEMENT

Further advances in deep neural network-based fake face detection will probably integrate advanced information and track feature extractors. This improved algorithm will be created for not only identifying surface level visual signals but also to investigate more into the picture and the detailed patterns left behind by various image editing techniques. These networks will provide a multi-dimensional method for identifying modified faces by combining content evaluation, which looks at contextual anomalies, with trace evaluation, which reveals minor distortions.

REFERENCES

[1] J.Y. Sun, S.-W. Kim, S.-W. Lee, and S.-J. Ko, ‘‘A novel contrast enhancement forensics based on convolutional neural networks,’’ *Signal Process.,Image Commun.*, vol. 63, pp. 149–160, Apr. 2018.

[2] Bromme, C. Busch, A.Dantcheva, C. Rathgeb, A.Uhl, “Fake Face Detection Methods: can they be generalized”, 2018.

[3] Luisa Verdoliva, Christian Riess, Justus Thies, Matthias Niebner, “A large scale video dataset for forgery detecton in human faces”,2018.

[4] Tong Che, Zoubin Grahahramani, Yoshua Bengio, Yangqiu Song, “MetaGAN: An adversarial approach to few-shot learning”, 2018.

[5] Antreas Antoniou, Amos J. Storkey, and Harrison Edwards. Augmenting image classifiers using data augmentation generative adversarial networks. In Artificial Neural Networks and Machine Learning - ICANN, pages 594–603, 2018. 2

[6] Nguyen Tran, Thanh-Toan Do, Ngai-Man Cheung, “Celeb-DF: A Large-scale Challenging dataset for DeepFake Forensics”, 2020.

[7] Oleg Alexander, Mike Rogers, William Lambeth, Jen-Yuan Chiang, Wan-Chun Ma, Chuan-Chang Wang, and Paul De bevec. The Digital Emily project: Achieving a photorealistic digital actor. IEEE Computer Graphics and Applications, 30(4):20–31, 2010. 2

[8] SercanArik, Jitong Chen, KainanPeng, Wei Ping, and Yanqi Zhou. Neural voice cloning with a few samples. In Proc. NIPS, pages 10040–10050, 2018. 2

[9] HadarAverbuch-Elor, Daniel Cohen-Or, Johannes Kopf, and Michael F Cohen. Bringing portraits to life. ACM Transactions on Graphics (TOG), 36(6):196, 2017. 1, 14

[10] Facebook Wants to Stay ‘Neutral’ on Deepfakes. Congress Might Force it to Act. Accessed: Jun. 14, 2019. [Online]. Available: https://www.vox.com/future-perfect/2019/6/13/18677574/facebookzuckerbergdeepfakes-congress-house-hearing

[11] K. Jain, A. Ross, and S. Pankanti, ‘‘Biometrics: A tool for information security,’’ IEEE Trans. Inf. Forensics Security, vol. 1, no. 2, pp. 125–143, Jun. 2006.

[12] K. Jain, A. Ross, and S. Prabhakar, ‘‘An introduction to biometric recognition,’’ IEEE Trans. Circuits Syst. Video Technol., vol. 14, no. 1, pp. 4–20, Jan. 2004.

[13] D. Cozzolino, G. Poggi, and L. Verdoliva, ‘‘Recasting residual-based local descriptors as convolutional neural networks: An application to image forgery detection,’’ in *Proc. 5th ACM Workshop Inf. Hiding Multimedia Secur.*, Jun. 2017, pp. 159–164.

[14] M. Huh, A. Liu, A. Owens, and A. A. Efros, ‘‘Fighting fake news: Image splice detection via learned self-consistency,’’ in *Proc. Eur. Conf. Comput.Vis. (ECCV)*, 2018, pp. 101–117.

[15] Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan,V. Vanhoucke, and A. Rabinovich, ‘‘Going deeper with convolutions,’’in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015,

[16] R. Raghavendra, K. B. Raja, S. Venkatesh, and C. Busch, ‘‘Transferable deep-CNN features for detecting digital and print-scanned morphed face images,’’ in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 1822–1830.

[17] Afchar, V. Nozick, J. Yamagishi, and I. Echizen, “MesoNet: A com- pact facial video forgery detection network,” 2018, *arXiv:1809.00888*. [Online]. Available: http://arxiv.org/abs/1809.00888

[18] P. Zhou, X. Han, V. I. Morariu, and L. S. Davis, “Two-stream neural networks for tampered face detection,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 1831–1839.

[19] L. M. Dang, S. I. Hassan, S. Im, and H. Moon, “Face image manipulation detection based on a convolutional neural network,” *Expert Syst. Appl.*, vol. 129, pp. 156–168, Sep. 2019.

[20] Bulat and G. Tzimiropoulos, ‘‘How far are we from solving the 2D & 3D face alignment problem? (and a dataset of 230,000 3D facial Landmarks),’’in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017.

[21] P. Viola and M. Jones, ‘‘Rapid object detection using a boosted cascade of simple features,’’ in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Dec. 2001.

[22] Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, ‘‘Rethinking the inception architecture for computer vision,’’ in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.

[23] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, C. A. Berg, and L. Fei-Fei, ‘‘ImageNet large scale visual recognition challenge,’’ *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.

[24] P. Kingma and J. Ba, ‘‘Adam: A method for stochastic opti- mization,’’ 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv. org/abs/1412.6980.