International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 2 Volume No.6 March – 2022 DOI:10.46647//ijetms.2022.v06i02.001
ISSN: 2581-4621

Deep Fake Face Detection Using Machine Learning and Various GAN Models

Kaku Riya Dharmendra[1], Upasana Tiwari[2]

[1]*Department Of Computer Engineering, Marwadi University, Rajkot, Gujarat, India*
[2]*Department Of Computer Engineering, Marwadi University, Rajkot, Gujarat, India*
[1]rdk96rjt@gmail.com
[2]upasana.tiwari@marwadieducation.edu.in

***Abstract*— Remarkable advancements in Deep learning, incredibly genuine AI generated pretended videos have been created. To appropriately construct the Deep fake phenomenon, new Deep fake detection algorithms must be designed for the exploitation of this powerful A.I. technology has a major effect on person lives. However, traces left by Generative Adversarial Network (GAN) engines during the construction of Deep fakes can be discovered by studying ad-hoc frequencies. We'll utilize the Deep Convolutional Generative Adversarial Network (DCGAN) and Style GAN, which have both shown to be quite good at creating images. We covered the theoretical aspects of GAN as well as our methods for developing a DCGAN Model using MNIST and CelebA data sets.**

***Keywords*— Generative Adversarial Network (GAN), CNN, DCGAN.**

I. INTRODUCTION

Astonishing enhancement within the area of deep learning leads the way to thickening of deep fake video tape. The help of expert system architecture like Generative Adversarial network (GAN) and auto encoders and a substantial quality of motion picture of selected topics anybody can produce such a pervasive pretend video tape. One of the reporter genus's Rana Ayyub suffers from threatening a deep fake plan. An incident video tape appears in the internet community. For that reason researchers are enlarging a structure which can find deep fake video tape. Deep fake is a technique that uses To superimpose face photos of a target person onto a video of a source person to create a video of the target person doing or saying things that the source person does, Deep Learning models such as auto encoders and generative adversarial networks are used. It's been used to solve a variety of difficulties in computer vision. This model has also looked at the person's facial expression and motions in the photograph or video clip.To build a convincing image and video tape, the deep fake algorithm requires a vast amount of data sets.

II. LITERATURE REVIEW

Mingzhu Nilotic et al. developed a system that recognises faces within frames. The difficulty with the CNN algorithmic programme is that the image is shrunk and used as input, which reduces the system's precision despite the loss of image data. As a result, the issue of knowledge loss no longer exists. CelebA, AFW, and FDDB are the datasets used in the proposed system. The planned system's unique accuracy is 95%, whereas the continuous system's is 74%.The technology also gave an answer that might be used to keep the CNN part's accuracy high.

David Guerra et al. demonstrated how to construct deep fake images and how to detect them using CNN and LSTM.For high grade GANs are used to create deep fake videos. For this creation approach, the first real image's encoder is used to swap faces with the target image's decoder. They used a variety of strategies on deep fake videos to come up with the best way to detect deep fakes: partitioning the video using a mixture of CNN and LSTM, the video was converted into 80 frames per second. The maximum accuracy level achieved was 97.1%. This accuracy was attained on high-resolution photographs.

Dig Vijay Yadav, et al. have described how deep fake techniques function and how they can accurately swap looks. They employed a variety of methods to determine whether or not the video was created via deep fake. CNN is most well-known for its visual content. acknowledgement & when in combination with LSTM, it slightly modifies the frame and this information can be utilized to detect deep fakes, according to one of the examples mentioned in this study .On the Face2Face dataset, the article claims that the deep fake video can be detected with 95% to 98 % accuracy using the Meso-4 and Misconception-4 architectures.

Peng Chen, et al. developed face swapping spotting film by behavioral and geospatial hints and named it FSSPOTTER, which may be a unified framework for exploring abstraction and temporal data within videos at the same time. The abstraction Extractor (SFE) separates

Website: ijetms.in Issue: 2 Volume No.6 March – 2022 DOI:10.46647//ijetms.2022.v06i02.001
ISSN: 2581-4621

the video into a number of segments, each of which comprises a specific range of frames. It accepts clipping and frame-level generation choices. The extraction of data from the backbone network encourages additional alternatives through making use of the super pixel- unit of binary categorization that is wise. A two-way LSTM is used by the Temporal Feature someone to search for temporal irregularities within the framework. After that, a completely linked layer and a soft max layer are used to calculate likelihood of a footage being genuine or phony. For the analysis, Face Forensics ++, Deep Fake TIMIT, UADFV, and Celeb-DF were utilized, with an accuracy of 77.6%.

During this paper, Andreas Rossler, et al. have said that finding Deep Fake, whether mechanically or by humans, is difficult. To discover all of these Deep Fakes, the article employs the CNN model. They disguised the forgery detection as a flaw in the modified videos' per-frame binary classification. They utilized a total of seven methods to uncover the deep fake of movies of varying quality. The handmade feature, as well as the SVM classifier, is used in the Steganalysis process. The results proved beyond a shadow of a doubt that the best accuracy for the caliber photos was eighty one for the Xception Net algorithm.

XTao, et al. has developed a technique which stresses the idea that perfect frame alignment and motion compensation are required to achieve greater results. In a CNN framework, the author has implemented a sub-pixel motion compensation layer. They also took 975 sequences from the camera, which contained 1080p HD video clips of excellent quality that were freely accessible over the internet, as well as down sample first frame of 540 X 960 pixels.In their suggested technique,The Peak Signal to Noise Ratio is a 36.71 % ratio of peak signal to noise.

With the use of an optical flow test, Irene Amerini et al. a system was proposed to leverage dissimilarities between frames as a possibility. Two successive frames for an authentic video and a deep fake video are pictured and displayed simultaneously to determine the optical flow field. For example, the chin movement in the original sequence is more powerful than the smoother movement in the altered film. The Face Forensic++ dataset was used, with 120 videos used for validation and testing and 720 videos used for training. In Face2Face video, they used two neural networks: VGG16 (81.6%) and ResNet50 (75.46%). This paper's uniqueness comes from its analysis of inter-frame dissimilarity, whereas other systems rely on inter-frame anomalies and use optical flow to resolve them on the Convolution neural network approach.

III. Objectives

During the early stages of the research, the majority of the methodologies employed to detect the faces were machine learning-based.They all normally go the same route.They either collect the dataset or download it beforehand.Following that, the images are preprocessed with several sorts of processing.The research then extracts these properties from the dataset.

The classification accuracy is determined by the methodologies used in picture analysis and extracted functions.Finding deep fake images is a time-consuming and difficult task.We focused on photos in this study and defined deep fake images as those created by machine learning algorithms, particularly GANs. GANs models are capable of producing high-quality photographs. In fact, humans have a hard time distinguishing between deep fake and real pictures.

IV. Methodology

The most successful rule for Deep Fake is the GAN, it combines two neural networks that may generate realistic visuals. This machine learning approach can learn from a large number of sources of pictures & afterwards combining these pictures to create a picture that appears authentic to human eyes. The architecture that is Fig-1 shows the various steps to ensure deep fake pictures and gives accurate results. The procedure is divided into four stages, as shown below.:

Step-1: CelebA & MNIST dataset : Downloads the data set from the kaggle repository.
Step-2: It uploads the following Dataset for process it includes all the pictures for restoration and rescalling.
Step-3: After that,The dataset is then divided into two sections: training and testing.
Step-4: GAN Model Created: The training dataset is applied, as a DCGAN intake and the weight optimizer is tweaked to determine if the photos are real or phony.

A. Data Collection

The datasets we used were MNIST and CelebA Dataset. The MNIST dataset (Modified National Institute of Standards and Technology) is a set of data from the National Institute of Standards and Technology. It's made up of 60000 tiny square gray scale images of handwritten single numerals ranging from 0 to 9 in a 28 X 28 grid. CelebA stands for CelebFaces Attributes Collection,containing approximately 200K photos, a large-scale facial characteristics dataset

B. Process Model

Machine learning technologies are used to suggest the system. To begin, we use the CelebA Dataset, which contains all photographs with resolutions of 4x4, 8x8, and 16x16 pixels to

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 2 Volume No.6 March – 2022 DOI:10.46647//ijetms.2022.v06i02.001
ISSN: 2581-4621

speed up the training process and produce a realistic model test result.

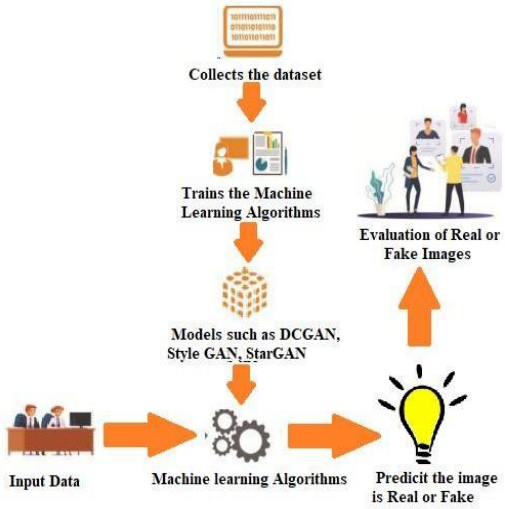

Fig 1: Architecture Diagram

C. How GAN's Work

The most successful rule for Deep Fake is the GAN, it combines two neural networks that may generate realistic visuals. This machine learning approach can learn from a large number of sources of pictures & afterwards combining these pictures to create a picture that appears authentic to human eyes. As an example, we can use GAN to assist us generate realistic pictures of animals, clothing styles, and anything else that GAN has been trained for. The GAN combines two neural networks, one of which is referred to as a generator, and another as the human. The human is a different neural network that quantifies the visuals created by the generator for credibility, and it helps in producing a lot of realistic pictures that look as real as human eyes. The Generator is a neural network that tries to come up with pretend pictures from the pictures information set, that is it attempts to make the best of what has been given to it and come up with real pictures. Min max approach is being taught to both the generator and the human. The zero indicates fictitious production, while the one represents genuine output. Humans want to get closer to one another in order to construct a convincing Deep Fake .If we train the generator and human for a long time, we will be able to generate a large number of realistic and authentic Deep fake photos, which will then be used to swap person 's face in comparison to person B's movie.

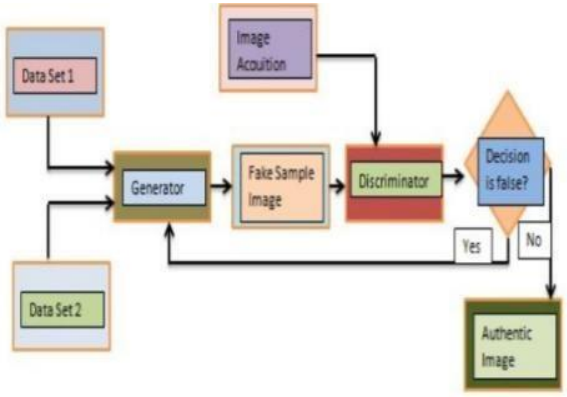

Fig 2: Working of GAN

D. Build Model

1. DCGAN: GANs are used to teach a deep learning model how to create new data from the same training data distribution. DCGAN is a deep learning model that is taught to generate new data from the same distribution of training data.

2. Style GAN: Finally, the Style Generative Adversarial Network architecture, also known as StyleGAN, is a recent and powerful method for entire-face synthesis, in which the framework regulates the style output at each phase of the generation process by mapping points in latent space to an intermediary latent space. StyleGAN is thus capable of producing not just photorealistic and high-quality images of faces, but also control settings for the overall style of the image at varying levels of detail.

3. Star GAN: Choi et al. proposed StarGAN as a method for doing image-to-image translations across multiple domains (e.g., modifying hair color, gender, and so on) with a single model. CELEBA, which was trained on two types of face datasets and has 40 labels linked to facial factors including hair color, gender, and age, can do image-to-image translation with excellent visual results

E. Loss Function of Generation And Discriminator

Discriminator Loss: While the discriminator is being taught, it classifies both genuine and fraudulent data from the generator. It avoids itself from misidentifying a real example as fake or a fake example (generated by the generator) as real by optimizing the below code. For discriminator learning to train , the following equation is ideal:

$$\nabla_{\theta}\frac{1}{n}\sum_{i=1}^{m}\left[\log D\times(x^{i})+\log\left(1-D\left(G\left(z^{i}\right)\right)\right)\right]$$

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 2 Volume No.6 March – 2022 DOI:10.46647//ijetms.2022.v06i02.001
ISSN: 2581-4621

Maximizing log(1-D(G(z)) will aid in accurately recognising the generator's fake image, whereas log(D(x)) refers to the likelihood of the generator successfully classifying the genuine image.

Generator Loss: The generator takes image noise samples and generates the desirable result. The input is then transmitted to the discriminator, who determines whether it is "Real" or "Fake" depending on the discriminator's ability to tell the difference.Using the discriminator's loss, the generator loss is determined. If it properly deceives the discriminator, it is rewarded; otherwise, it is penalized.

$$\nabla_o \frac{1}{m} \sum_{i=1}^{n} \log\left(1 - D(G(z^i))\right)$$

F. Training & Testing

At the time of training the data we first trained the discriminator and generator and got the G_Loss and D_loss.

Part-1: Training of discriminator:

Remember that the purpose of discriminator training is to increase the likelihood of accurately identifying a given input as authentic or phony. We want to "upgrade the discriminator by rising its stochastic gradient," as Good fellow puts it. We aim to maximize log(D(x))+log(D(G(z)) in practise .Following will be calculated in two steps: To obtain the values, we will first generate a set of real pictures from the training dataset, just pass through D and calculate the loss function of (log(D(x)), and finally reverse the pass. Second, we will construct a batch of false pictures using the current generator, then forward pass them via D, compute their loss (log(D(G(z)), & reverse transmit a value.

Part-2: Training of Generator

We want to put the Generator through its paces by minimizing log(D(G(z)), as stated in the original study. As noted previously, this method does not give appropriate gradients, particularly early on in the educational process, as demonstrated by Goodfellow. As a workaround, we'd want to increase log(D(G(z)).

We do this in the programme by first utilizing the Discriminator to categorize the Generator output from Part 1, then estimating G's loss with real marking as GT, a backward pass to compute G's gradients, and an optimizer step to alter G's parameters.The use of actual labels as GT labels for the loss function may appear weird, but it allows us to use the log(x) component of the BCELoss (rather than the log(1x) component, which is more efficient.

Finally, we'll present some data and perform our constant At the conclusion of each epoch, run a noise batch through the generator to see how G's training is developing.
The training statistics are as follows:

- Loss_D: The total amount of money lost on all batches, real and false is determined as the discriminator loss (log(D(x)+log(D(G(z)).
- G stands for Generator loss and thus it is computed as log(D(G(z)).
- D(x) is the discriminator's median result for the entire real batch .This should start near 1 and theoretically converge to 0.5 as G improves.
- The first number indicates when D was last updated, whereas the second indicates when D was last updated.

Here is the graph showing the Loss_G Loss_D, D(x) and D(G(z)).

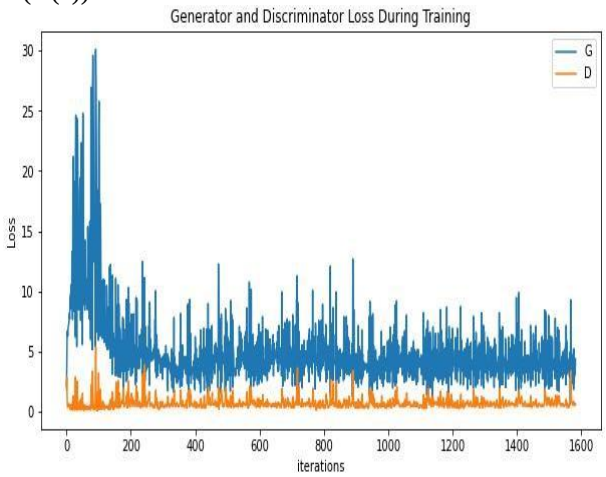

Fig 3: Shows the result of Generator and discriminator loss.

G. Data Visualization

In these we have shown the sample images of the CelebA, MNIST dataset .

FIG: 4 SHOWS THE MNIST DATASET

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 2 Volume No.6 March – 2022 DOI:10.46647//ijetms.2022.v06i02.001
ISSN: 2581-4621

FIG: 5 SHOWS THE DATA SET OF CELEBA.

V RESULTS

A. Result of Style GAN :

TABLE I

Model	Loss_D	Loss_G
Resolution 4 X 4	3.1302	1.0665
Resolution 8 x 8	-0.4877	3.7509
Resolution 16 x 16	-1.2683	11.4094

Shows the Stable resolution of Loss_G and Loss_D

TABLE II

Model	Loss_D	Loss_G
Resolution 8 x 8	3.9530	1.8761
Resolution 16 x 16	5.5708	7.0519

Shows the Transition resolution of Loss_G & Loss_D

Style GAN output for the Images after it has been trained and tested .

Fig: 6 Shows the result when it was run during the interface using the pre-trained 64 x 64 pixel.

B. Result of DCGAN using MNIST data set

Fig: 7 Shows the image of the MNIST dataset after the training of the dataset.

C. Result of DCGAN using CelebA dataset

Fig: 8 Shows the Real Images which we have used while giving the input to our model.

Fig: 9 Shows the Fake Images which are created by our model which is the DCGAN model.

Website: ijetms.in Issue: 2 Volume No.6 March – 2022 DOI:10.46647//ijetms.2022.v06i02.001
ISSN: 2581-4621

D. Result of Star GAN using CelebA dataset:

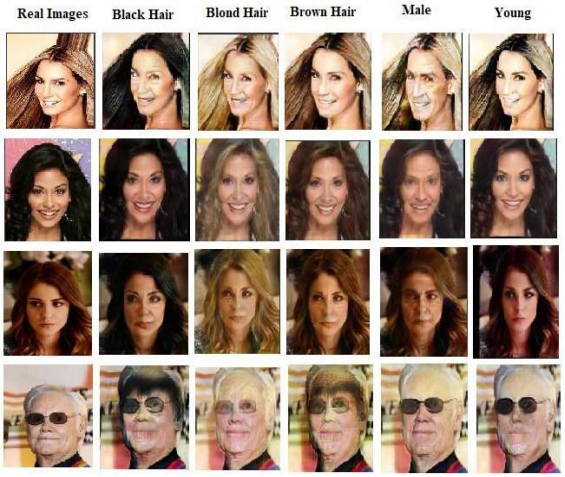

Fig: 10 Shows the original image and the generated image with various features such as black hair, brown hair, Blond hair Male & young.

VI CONCLUSION

Given that deep fake generated photos are becoming increasingly popular on social media platforms, it is becoming increasingly difficult to distinguish between the original and the faked versions.Deep learning methodologies and computer facilities are used.We uncovered some of the images that were generated using GAN using models like DCGAN, StarGAN, and StyleGAN.Brown hair, Black hair, Blond Hair, and other features in StarGAN were used to create the photos.The Loss Of Generator and Discriminator Graph has been displayed. This paper indicates that 95% of the photos are genuine, while the other 5% are fraudulent. We plan to raise the image resolution in the future, as well as add more models and datasets, to make detecting deep fake images more precise.

References

[1] Yadav, D., & Salmani, S. (2019, May). Deepfake: A survey on facial forgery technique using generative adversarial network. In 2019 International Conference on Intelligent Computing and Control Systems (ICCS) (pp. 852-857). IEEE.

[2] Amerini, I., Galteri, L., Caldelli, R., & Del Bimbo, A. (2019). Deepfake video detection through optical flow based cnn. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (pp. 0-0).

[3] Chen, P., Liu, J., Liang, T., Zhou, G., Gao, H., Dai, J., & Han, J. (2020, July). Fsspotter: Spotting face-swapped video by spatial and temporal clues. In 2020 IEEE international conference on multimedia and expo (ICME) (pp. 1-6). IEEE.

[4] Aneja, S., & Nießner, M. (2020). Generalized Zero and Few-Shot Transfer for Facial Forgery Detection. arXiv preprint arXiv:2006.11863.

[5] Tao, X., Gao, H., Liao, R., Wang, J., & Jia, J. (2017). Detail-revealing deep video super-resolution. In Proceedings of the IEEE International Conference on Computer Vision (pp. 4472-4480).

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

[7] Ranjan, P., Patil, S., & Kazi, F. (2020, March). Improved Generalizability of Deep-Fakes Detection using Transfer Learning Based CNN Framework. In 2020 3rd International Conference on Information and Computer Technologies (ICICT) (pp. 86-90). IEEE.

[8] Luo, M., Xiao, Y., & Zhou, Y. (2018). Multi-scale face detection based on convolutional neural network. In 2018 Chinese Automation Congress (CAC) (pp. 1752-1757). IEEE.

[9] Rossler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Nießner, M. (2019). Faceforensics++: Learning to detect manipulated facial images. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 1-11).

[10] Matern, F., Riess, C., & Stamminger, M. (2019, January). Exploiting visual artifacts to expose deepfakes and face manipulations. In 2019 IEEE Winter Applications of Computer Vision Workshops (WACVW) (pp. 83-92). IEEE.

[11] McCloskey, S., & Albright, M. (2019, September). Detecting GAN-generated imagery using saturation cues. In 2019 IEEE International Conference on Image Processing (ICIP) (pp. 4584-4588). IEEE.

[12] https://www.indiatoday.in/india-today-insight/story/india-s-deepfake-problem-videos-are-neither-deep-n or-fake-1643883-2020-02-06

[13] Xin Yang, Yeuzen Li and Siwei Lyu, "EXPOSING DEEP FAKES USING INCONSISTENT HEAD POSES", ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).