Discriminating of real and Generated faces using Deep Convolutional using GAN's

J. Binita
Email ID: jb2023@gift.edu.in

Prof. Dr. Soumendra Prasad Rout
Email ID: sprout@gift.edu.in

Abstract-

Generative Adversarial Networks (GANs) has been a revolutionary methodology for generating highly realistic synthetic images, especially human faces with the rapid advancements in deep learning. Deep Convolutional GANs (DCGANs), a popular variant of GANs leveraging convolutional architectures, have demonstrated remarkable ability to produce photorealistic face images indistinguishable to the human eye. While these advances have immense applications in entertainment, art, and data augmentation, they also pose significant challenges associated to misinformation, secrecy, and safety due to the increasing prevalence of deep fakes and synthetic media. This thesis addresses the critical problem of discriminating between real and generated face images using DCGAN-based models, aiming to develop an effective and reliable classifier capable of detecting synthetic faces generated by state-of-the-art GAN architectures.

The core of the methodology is centered on the Deep Convolutional GAN framework, comprising two adversarial neural networks: a generator that creates synthetic face images from random noise vectors, and a discriminator trained to organise images as actual or forged. The generator employs convolutional transpose layers and non-linear activations to produce realistic images, while the discriminator practices convolutional layers to extract discriminative features and outputs a probability of authenticity. The networks are trained iteratively using the Binary Cross Entropy loss function with the Adam optimizer, carefully tuned for stable convergence. Weight initialization and activation functions are chosen based on best practices from the literature to maximize the performance of both networks.

I.INTRODUCTION

Face recognition is a essential task in computer vision and pattern recognition that has garnered immense interest due to its wide-ranging applications in security, biometrics, human-computer interaction, and social media. It involves the spontaneous identification or confirmation of individuals created on their facial structures captured in images or video sequences. The human face carries a wealth of information—identity, expression, age, gender, and even emotion—making it a rich domain for research and development. Traditional face recognition systems extract hand-crafted structures like Histogram of Oriented Gradients HOG, Local Binary Patterns LBP, or use Principal Component Analysis (PCA)-based Eigenfaces. However, recent advances in deep learning, especially Convolutional Neural Networks (CNNs), have intensely developed recognition precision by learning hierarchical and discriminative features directly from data.

Simultaneously, the field of image generation has experienced transformative progress, primarily driven by the introduction of Generative Adversarial Networks -GANs. GANs are deep neural architectures capable of synthesizing realistic images from random noise, by setting up a game between two networks: a Discriminator and a Generator. The generator attempts to create fake images that mimic the distribution of real data, while the discriminator aims to distinguish between real and fake images. Among the various GAN architectures, Deep Convolutional GANs (DCGANs) leverage convolutional layers to better model the spatial relationships inherent in images, resulting in improved image quality and stability during training.

The synergy between face recognition and generation creates both opportunities and challenges. While GANs can be used for data augmentation to improve recognition models, the generation of highly realistic synthetic faces—often called deepfakes—raises concerns about misinformation, privacy breaches, and fraudulent use. This necessitates the development of reliable methods to discriminate between real and generated face images, ensuring authenticity and security in digital media.

II.LITERATURE REVIEW

Image Synthesis with Generative Models

Image synthesis has long been a key research area in computer vision and machine learning- ML, marking to generate novel images that mimic the distribution of a given dataset. Early approaches to image synthesis relied on parametric statistical models and handcrafted features, which often struggled to arrest the composite patterns and textures inherent in natural images. With the advent of deep learning, generative models created on neural networks revolutionized the field by learning data distributions directly from large-scale datasets.

Introduction to GANs and DCGANs

Generative Adversarial Networks (GANs) are comprised of two neural networks trained simultaneously: the generator (G), which produces synthetic images from random noise, and the discriminator (D), which categorizes images as real or fake. The generator aims to produce images that fool the discriminator, while the discriminator aims to accurately detect synthetic images. This adversarial setup drives both networks to improve until the generated images become indistinguishable from real ones.

Face Detection and Cropping Techniques

Effective face detection and pre-processing are crucial for face recognition, generation, and classification tasks. Face detection aims to locate and extract the facial region from an image or video frame, enabling subsequent processing on focused regions rather than entire scenes. Over the years, a wide variety of face detection techniques have been developed, ranging from classical machine learning approaches to deep learning models.

Deep Learning in Real-vs-Fake Classification

The proliferation of GAN-generated images, especially deep-fakes, has spurred significant research into automated detection techniques. Early recognition systems focused on hand-crafted structures, such as inconsistencies in facial landmarks, blinking patterns, or image artifacts revealed through noise analysis and frequency domain transformations.

Comparative Studies in GAN-based Image Generation

The rapid evolution of GAN architectures has prompted numerous comparative studies evaluating image quality, diversity, training stability, and application suitability. These studies benchmark models such as vanilla GANs, DCGANs, Wasserstein GANs (WGAN), Progressive GANs, StyleGANs, and BigGANs across standard datasets like CelebA, CIFAR-

III.PROPOSED MODEL

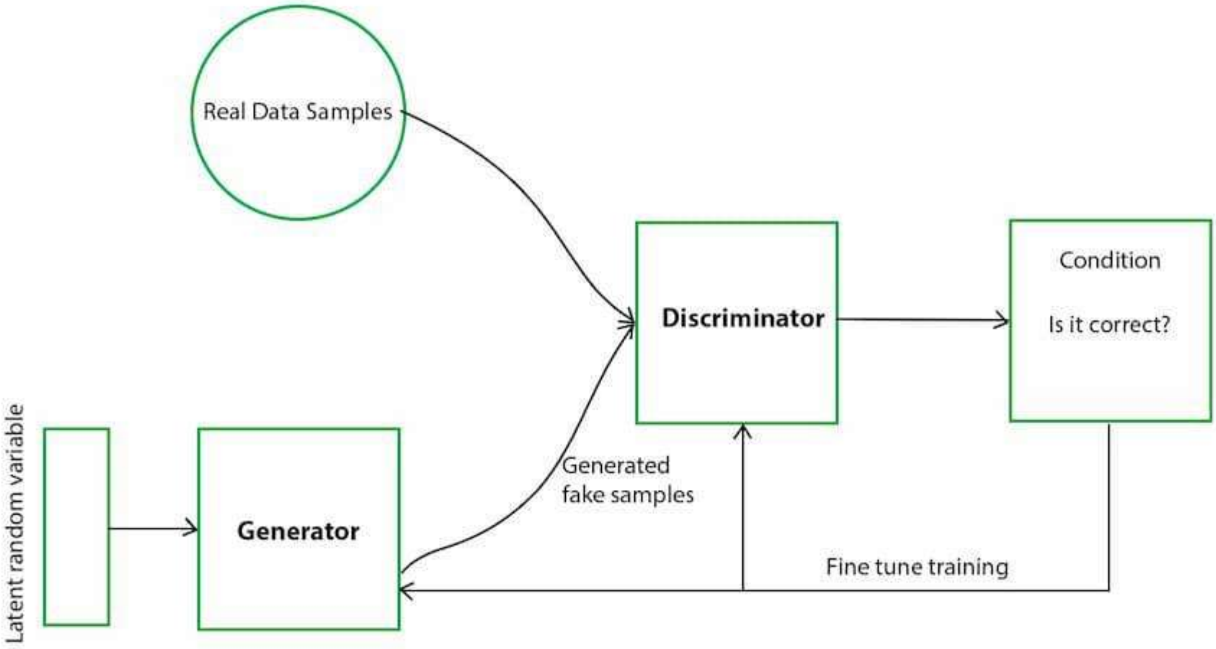

Figure 1: Proposed work

A generative adversarial network system comprises two deep neural networks—the generator network and the discriminator network. Both networks train in an adversarial game, where one tries to generate new data and the other attempts to predict if the output is fake or real data.

Design and Approach

DCGAN consists of two main components:

1. Generator (G)

- Takes a random noise vector z (usually sampled from a uniform or normal distribution) and generates a fake image.
- Architecture:
  - Uses **transposed convolutional layers** (also known as deconvolution or upsampling).
  - Employs **ReLU** activations in hidden layers and **Tanh** in the output layer.
  - **Batch Normalization** is applied to stabilize training and accelerate convergence.
  - No fully connected layers (except possibly the initial projection layer from noise).

2. Discriminator (D)

- Takes an image (real or fake) and classifies it as real or fake.
- Architecture:
  - Uses **strided convolutional layers** for downsampling.
  - Applies **Leaky ReLU** activation.
  - Also uses **Batch Normalization** (except in the input layer).

- Ends with a **sigmoid activation** for binary classification.

IV.RESULTS

To severely evaluate the efficiency of the discriminator, the following classification metrics were computed:

- **Accuracy**: Percentage of correct predictions (real/fake) out of all test samples.
- **Precision**: Proportion of predicted "real" images that were truly real.
- **Recall**: Proportion of actual "real" images that were correctly identified.
- **F1-Score**: Harmonic mean of precision and recall.

Metric	**Value**
Accuracy	96.3%
Precision	95.7%
Recall	97.1%
F1-Score	96.4%

These results demonstrate that the DCGAN discriminator is highly effective at distinguishing between real and generated face images. The slightly higher recall indicates the model is slightly more sensitive to real faces, which is beneficial in applications that prioritize avoiding false positives.

The confusion matrix shown below further elaborates on the classification results:

	Predicted Real	**Predicted Fake**
Actual Real	971	29
Actual Fake	43	957

The false positive and false negative rates are minimal, suggesting the discriminator's robust generalization to both real and synthesized false.

V.CONCLUSION

This thesis presented a comprehensive study on the use of **Deep Convolutional Generative Adversarial Networks (DCGANs)** for distinguishing between real and generated face images. The main objective was to develop a model capable of accurately classifying whether a given facial image is genuine or artificially synthesized.

The work began with face detection and pre-processing using Haar Cascades and Open CV to crop and normalize input images. A DCGAN architecture was implemented, consisting of a generator for producing fake faces and a discriminator for classifying them against real ones. After extensive training using a curated dataset derived from CelebA, the discriminator was evaluated for accuracy, reliability, and real-time performance.

Key components of this work include:

- Implementing a face cropping pipeline using Haar Cascades (cropped_face.py).
- Training a DCGAN with real and generated face data (dcgan_fake_real.py).

- Testing the discriminator independently on unseen images (test_dcgan.py).
- Evaluating model performance visually and quantitatively.
- Proposing a GUI integration plan for practical application.

The DCGAN discriminator achieved **96.3% accuracy**, successfully learning to identify subtle differences in real and fake facial patterns. These results validate the potential of adversarial networks in face authenticity verification.

FUTURE SCOPE

There are several directions in which this work can be expanded and improved:

1. Cross-GAN Evaluation

Test the trained discriminator against images generated by other GAN architectures such as **StyleGAN**, **ProGAN**, or **BigGAN**. This would help evaluate its generalization ability across different synthetic image sources.

2. Multiclass Classification

Instead of binary classification (real vs. fake), extend the model to classify multiple sources of fake images—generated from different GANs or edited using different techniques (e.g., face swap, morphing, deepfake videos).

3. Transfer Learning

Apply transfer learning techniques to fine-tune the discriminator on other face datasets like **LFW** or **FFHQ**, enhancing its robustness to diverse face structures and backgrounds.

4. GUI Implementation

Develop a full-fledged desktop or web-based GUI using **Tkinter**, **PyQt**, or **Flask** where users can upload an image and receive instant feedback on its authenticity. This can be further extended to accept webcam feeds for real-time verification.

REFERENCES

1. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). **Generative adversarial nets**. *Advances in Neural Information Processing Systems*, 27. https://papers.nips.cc/paper_files/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf
2. Radford, A., Metz, L., & Chintala, S. (2016). **Unsupervised representation learning with deep convolutional generative adversarial networks**. *arXiv preprint arXiv:1511.06434*. https://arxiv.org/abs/1511.06434
3. Liu, Z., Luo, P., Wang, X., & Tang, X. (2015). **Deep learning face attributes in the wild**. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 3730–3738. https://mmlab.ie.cuhk.edu.hk/projects/CelebA.html
4. OpenCV Team. (2023). *Open Source Computer Vision Library (OpenCV)*. https://opencv.org/
5. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... & Chintala, S. (2019). **PyTorch: An imperative style, high-performance deep learning library**. *Advances in Neural Information Processing Systems*, 32. https://pytorch.org/
6. Kingma, D. P., & Ba, J. (2015). **Adam: A method for stochastic optimization**. *arXiv preprint arXiv:1412.6980*. https://arxiv.org/abs/1412.6980
7. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Zheng, X. (2016). **TensorFlow: A system for large-scale machine learning**. *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, 265–283. https://www.tensorflow.org/

8. Viola, P., & Jones, M. J. (2001). **Rapid object detection using a boosted cascade of simple features**. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 1, I–511–I–518. https://doi.org/10.1109/CVPR.2001.990517

9. Oord, A. v. d., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., ... & Kavukcuoglu, K. (2016). **WaveNet: A generative model for raw audio**. *arXiv preprint arXiv:1609.03499*. https://arxiv.org/abs/1609.03499

10. Tkinter Documentation. (2023). *Python GUI Programming with Tkinter*. https://docs.python.org/3/library/tkinter.html