Journal of Education for Pure Science
Vol.16, No. 1 (2026)
DOI: https://doi.org/10.32792/jeps.v16i1.827
Website: jceps.utq.edu.iq
Email: jceps@eps.utq.edu.iq

Facial Expression Image Generation Using Deep Convolutional Generative Adversarial Network (DCGAN)

Sabreen Ghanim* **and Ashwan A. Abdulmunem**

Department of Computer Science, Faculty of Computer Science and Information Technology, University of Kerbala, Karbala, Iraq

* Corresponding email: sabreen.g@s.uokerbala.edu.iq

Received 9/ 11 /2025, Accepted 14/ 1 /2026, Published 1 / 3/2026

Abstract:

This research presents a DCGAN (deep convolutional generative adversarial network) for synthesizing grayscale facial expressions. Generating grayscale facial expressions is essential for affective computing and emotion analysis. We presented a deep convolutional generative adversarial network model for generating monochrome facial expression images and evaluated it on two benchmark datasets, FER-2013 and CK+. The proposed implementation also benefits from non-linearity, using Pixel Normalization and LeakyReLU activations to enhance image similarity and training reliability. As a result, the study achieved an FID of 88 and an IS of 3.0 on FER-2013, and an FID of 0.5 and an IS of 2.0 on CK+, indicating high image quality and photorealism. In a similar vein, the results estimated data quality and network robustness as well as their interaction with the GAN's achieved performance. Taken together, the proposed research may serve as a strong, lightweight baseline for facial expression synthesis, likely to stimulate future GAN research and identity-boosting synthesis.

Keywords: deep convolutional generative adversarial networks (DCGANs), facial expression generation, FER-2013 dataset, generative adversarial networks

1-Introduction

Facial expressions represent one of the most fundamental and universally recognized indicators of human emotion and intent. Early in the twentieth century, Ekman and Friesen identified six primary emotions—anger, disgust, fear, happiness, sadness, and surprise—through a cross-cultural investigation, revealing that emotional perception remains consistent mainly across societies. In recent years, deep learning has achieved remarkable advances and been applied across diverse domains; however, image generation remains an evolving field with substantial potential for improvement [1-3].

In computer vision, facial expression transfer is a complex yet vital task, as it involves transferring the facial emotion from source image onto a target identity while preserving facial integrity. This process has broad applications across medicine, transportation, cultural heritage, and education [4, 5].

The introduction of Generative Adversarial Networks (GANs) has revolutionized image generation, establishing a significant milestone in both classification and synthesis. Despite their success in producing realistic human imagery, GANs continue to encounter issues such as mode collapse, unwanted visual artefacts, and the challenge of balancing fidelity with computational efficiency [6, 7].

Since their original formulation by Goodfellow et al. (2014), GANs have emerged as a robust generative framework in deep learning for creating high-quality synthetic data. A typical GAN consists of two components: a generator, which produces artificial samples, and a discriminator, which evaluates their authenticity and guides the generator's improvement through adversarial learning (Goodfellow). Over time, numerous GAN variants have been developed to address specific challenges or to enhance generation quality. Many conditional GAN extensions have been introduced to improve the control over image generation. cGANs utilize label-based conditioning to guide generation, Pix2Pix performs paired image-to-image translation under supervision, and StarGAN enables multi-domain translation using a standard architecture. Although the structures of these models differ, they all aim to regulate the generation process using conditional constraints rather than directly sampling unconditionally [2, 4].

GANs feature a unique training methodology based on competitive learning, which pits two distinct neural networks—the generator and the discriminator —against each other. The generator network, G, aims to map latent space vectors drawn from a prior distribution p_z (z) to the data space, effectively generating new data samples that mimic the real data distribution, p_data (x). In contrast, the discriminator network, D, is trained to distinguish between samples from the real data distribution and those generated by the generator. Competition between the networks continuously improves their performance, ultimately enabling the generator to generate realistic outputs. GAN training is formulated as a min-max equation, in which the following value by an equation can formally represent V(G, D), as in Eq. (1).

$$\min_{G} \max_{D} Ex{\sim}p_{data}(x)[\log D(x)] + Ez{\sim}p_{z}(z)\big[1 - D\big(G(z)\big)\big] \tag{1}$$

Here, the first term is the expected logarithm probability that the discriminator mistakenly classifies real data samples as fake. The second one is similar but for fake samples made by G. GAN training involves switching between improving D to maximize V(D, G) for a fixed G (which improves D's accuracy in distinguishing between real and fake samples) and improving G to minimize V(D, G) for a fixed D (which improves G's ability to generate realistic samples). This competitive training continues until a steady state is reached where G produces samples that cannot be distinguished from real data by D. The competitive training mechanism of competitive generative networks has proven highly effective at generating complex, high-dimensional data.

This series of architectures is primarily used for the synthesis of facial images, where the emotion is either emphasized on the basic or focused on improvement, and it is used mainly to enhance the resolution and trans-species conversion of the same label in the graph. The DCGAN proposed in this paper utilizes emotional real expression and enhances a variety of expressions through simplicity and high training stability. At the same time, in the low-resolution grayscale image, the pixel normalization method and LeakyReLU activation are introduced to improve image degradation and reduce gradient instability [8-11].

It will be more advantageous for maintaining simplicity, quality of generation, and stability of training. The results using DCGAN-generated synthetic samples combined with existing augmentation techniques showed a statistically significant improvement in all cases. Changes were particularly prominent in datasets with fewer than 750 samples, increasing the F1-score by approximately +0.09 [12, 13].

The improvement method appears to be especially effective when the original data quantity is low. For a dataset with 500 to 750 samples, the improvement ratio should be evaluated based on the circumstances. Furthermore, overuse of synthetic augmentation can distort data distribution or introduce artificial features. In conclusion, stratification by dataset size and its relevant context is recommended.

Various GAN-based image generation techniques have achieved remarkable results in recent years. However, most successful models require high-resolution color input datasets and complex computational structures, making it difficult for researchers without significant hardware resources to reproduce the workflow. Additionally, the most existing works focus on high-resolution full-color datasets and have not explored low-resolution grayscale appearance conditions. It is essential to maintain the rich emotional representation of human faces even when distinct color data are not available. Therefore, this paper aims to develop a practical DCGAN architecture that effectively generates high-quality grayscale facial expressions for underrepresented machine learning practitioners.

This work addresses the problem that GAN models perform poorly for low-resolution images, and we further design a stable but straightforward architecture by introducing PixelNorm and using LeakyReLU.

2. Related Work

Recent progress in facial expression synthesis continues to affirm the effectiveness of GAN-based architectures. However, persistent challenges remain in balancing image quality, computational cost, and fine-grained control of expression. For instance, Bao et al. (2024) achieved high-resolution (1024×1024) expression transfer using Progressive StyleGAN2 with feature disentanglement (FID = 34.61, SSIM = 0.90). Nevertheless, their method demands extensive computation and multi-stage optimization. In contrast, the proposed optimized DCGAN framework emphasizes architectural simplicity and training stability, enabling efficient generation of expression-specific images. Unlike Bao et al.'s reliance on StyleGAN2 latent manipulation and ResNet50/VGG16 feature extractors, our approach achieves superior fidelity on standardized datasets such as CK+ (FID = 0.5), using only 48×48 grayscale inputs combined with PixelNorm and LeakyReLU enhancements [14].

This research aims to generate high-quality, realistic facial expressions from real images using a GAN-based approach, addressing the central challenge of computational complexity and extended training durations typically associated with GAN frameworks. While these networks can produce visually convincing results, their resource-intensive nature poses limitations for researchers with restricted computational access.

Several prior studies have reported defective or low-quality generations, primarily due to insufficient computational capacity and architectural inefficiencies [15, 16].

2-Dataset Description

The dataset for this study was obtained from Kaggle. The Face Expression Recognition Dataset consists of 28,821 images and contains six primary emotions: neutral, sad, surprised, fear, disgust, happiness, and anger [17], shown in Figure 1.

The model was tested on another Cohn-Kanade (CK+) Enhanced Dataset [18, 19] with 920 images, which also includes 48 × 48 grayscale images and the basic emotions of anger, sadness, and others [20], as shown in Figure 2.

Although obtained several years earlier, the CK+ dataset remains one of the most popular benchmarks for facial expression research. It boasts excellent annotation quality, an equal distribution across emotion categories, and continuous light conditions. All of these factors ensure a thorough evaluation of generative quality and emotional authenticity and, generally, create suitable conditions for GAN-based synthesis research.

Figure 1. The face expression recognition dataset

Figure 2. Cohn-Kanade (CK+) enhanced dataset

3-Preprocessing

Image preprocessing is performed as follows: First, the image data is loaded; each image is converted to grayscale (one channel instead of RGB), resized to 48 × 48 pixels to standardize the inputs, converted into NumPy arrays for visualization purposes, and then normalized to [0, 1] to be suitable for neural network models. This procedure enhances image adaptation in neural network models, reducing the likelihood of overfitting and enhancing the model's generalization [10].

4-Proposed Approach

In this work, a generator model is designed that transforms a low-dimensional random vector (latent vector) into a synthetic image of resolution 48 × 48. The Pixel Normalization technique is included after each master layer to stabilize the range of values and enhance training stability.

In this work, a generator model is built where the model converts a low-dimensional random vector (latent vector noise) into a synthetic image with a specified resolution (48 × 48), consisting of an input layer of length 100 and a dense layer in which the vector is projected into an initial representation of size 6 × 6 × 512.

Pixel normalization is then applied to stabilize the relative values within each dimension of the representation. To mitigate backflow of errors and avoid gradient extinction, a LeakyReLU activation function is used. The representation is then transformed into 6 × 6 × 512 raw images using three layers with decreasing numbers of filters (256 → 128 → 64), each with increasing spatial dimensions, followed by pixel normalization and a LeakyReLU activation. Then a Conv2DTranspose layer with one filter and a tanh activation to produce a gray-normalized image in the range [-1,1]. Figure 3 shows the architecture of the generator.

Choose a latent vector of 100, as it is a general dimensionality used across DCGAN-based architectures. This size struck a compromise between variation diversity and stability, and was large enough for the generator to capture meaningful variations without being overly complex to complicate training.

For the architecture used in generative adversarial networks, designed to distinguish real images from generated images, the model is built using the TensorFlow library and its high-level interface Keras and takes as input a 48 × 48-pixel grey image.

The model consists of a series of convolutional layers with the number of filters increasing from 64 to 512. It uses a 4 × 4 filter with a stride of 2, gradually reducing the spatial dimensions at each layer. After each convolutional layer, the LeakyReLU activation function is applied to avoid the dying ReLU issue. Additionally, a batch normalization layer is used after the convolutional layers, starting with the second layer, to stabilize the training process further.

After all the convolutional layers are completed, the output is flattened and passed to a final dense layer containing only one unit. A sigmoid activation function is used to produce a probability value indicating whether the input image is real or generated. Figure 4 shows the architecture of the discriminator.

What is new is my simple enhancement over the classical DCGAN framework, where I apply Pixel Normalization after the convolutional layer and LeakyReLU activations to prevent gradient vanishing and stabilize features, thereby producing fixed and stable learning dynamics and a solution to common instability factors that lead to mode collapse in traditional GANs.

Loss function: Choosing an appropriate loss function is important for this study. In this work, the binary cross-entropy loss function was chosen. This is because the discriminator in a GAN must distinguish between generated and real data, a task that falls under binary classification. For these binary classification tasks, a good match was found with the cross-binary loss function. As shown in Figure 5, which shows the general working form of the architecture used in this research, DCGAN

PixelNorm was used instead of Batch Normalization because it prevents large feature magnitudes, which have been noted to cause hard-to-control generator behavior when training GANs, and is also based on layer-wise statistics that stabilize more easily while avoiding operations above those previously obtained extreme magnitudes.

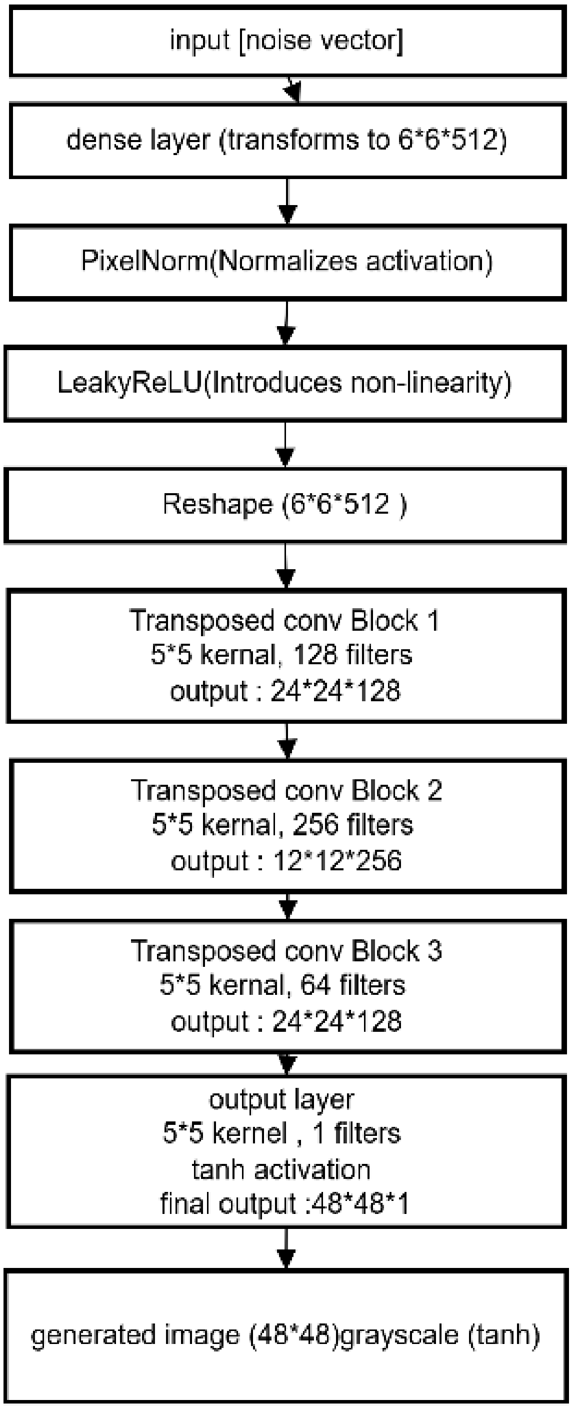

Figure 3. The architecture of the generator

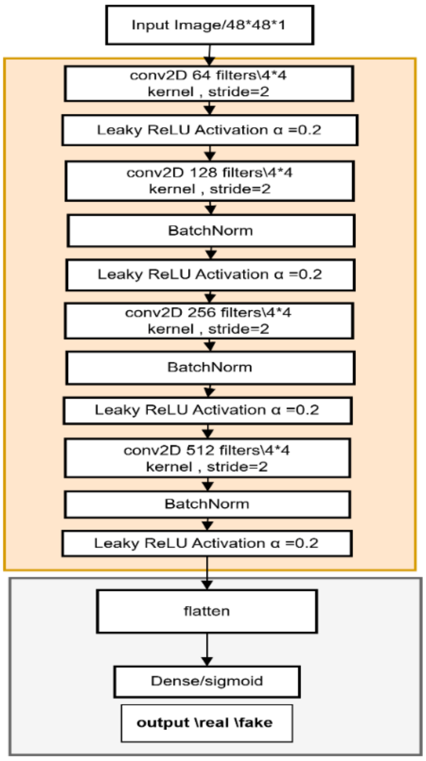

Figure 4. The architecture of the discriminator

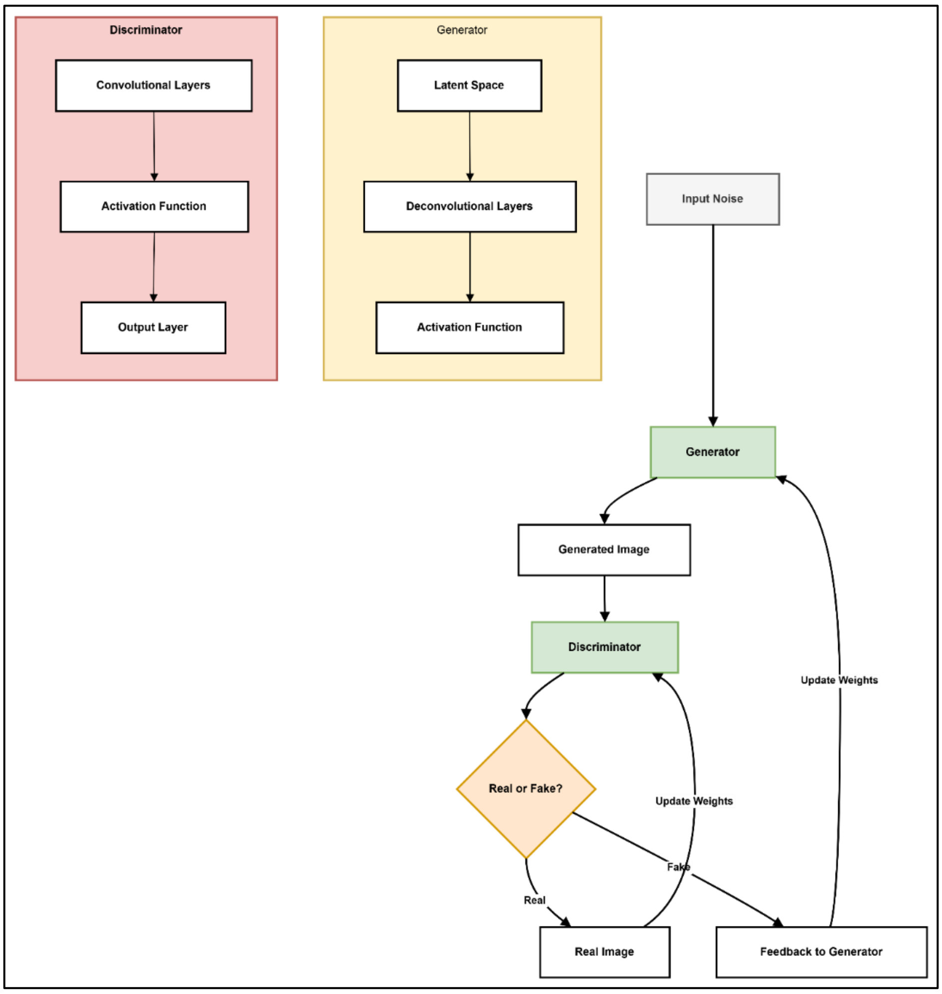

Figure 5. The architecture of the DCCAN

5-Results and Discussion

The Face Expression Recognition Dataset (FER 2013) is a comprehensive collection designed for training and evaluating facial expression recognition. It provides a diverse set of images that capture a range of emotions, making it a valuable resource for researchers and developers in the field of computer vision models.

The behaviors of the five trained experiments showed diversity, which was influenced by resources and learning settings. Experiment 1 progressed slowly due to resource limitations, leading to slow convergence. Experiment 3 did not work due to gradient instability and the destruction of the discriminator's gradients during training, around 150 epochs. Experiment 2, on the other hand, reached a plateau in its development, since it used a batch size of 16 and could not take advantage of a more powerful TPU setup that enabled more reliable gradient updates.

In Experiment 3, the model encountered gradient degradation after prolonged training (≈150 epochs), which caused instability in the discriminator and ultimately led to divergence in the generated samples.

Despite the competitive findings, the model had several limitations. Some of the generated samples showed blurring or an incomplete facial structure, suggesting that the generator had difficulty learning fine-grained details. The improvement of the proposed method over PixelNorm is pronounced to some extent, while its influence varied according to the dataset's clean level; for CK+, with more ordered and clear characteristics, PixelNorm performed better than the otherwise "noisy" FER-2013. Our results demonstrate that stronger regularization and more robust training are necessary to deal with noisy real-world datasets.

5.1 Evaluation overview

To evaluate the proposed DCGAN model, experiments are conducted on two datasets: the Kaggle facial expression recognition dataset and the CK+ dataset. Two widely used metrics for assessing generative models, the Frechet Inception Distance and the Inception Score, are applied to the test results. FID measures the quality, and ICS assesses the diversity of the generated images.

5.2 Experimental setup

All the experiments were performed with Python. A learning rate of 0.0002 and an exponential decay rate of 0.5 were used in all experiments. The Adam optimizer was chosen to achieve faster convergence to local minima and better stabilization properties compared to classic optimization methods. The model was trained with different numbers of epochs and batch sizes. For the CK+ dataset, training was performed for 100 epochs with a batch size of 8. The same learning rate was used as in the experiment with the FER dataset. The five experiments were conducted with varying numbers of training samples, batch sizes, and hardware resources. The architecture of the generator and discriminator was consistent with a 100-dimensional latent vector, including dense and transformed convolutional layers. The discriminator had a typical convolutional architecture.

Our model was trained in a cloud-based computational environment called Google Colab, which allows us to run code on their machines with better GPUs for deep learning research and evaluation. The assigned instance came with an NVIDIA Tesla T4 GPU (16GB GDDR6) on the Turing architecture. It offers excellent parallel processing and supports CUDA 12.4, making it great for accelerating training and inference workloads. Aside from the GPU, the system had around 47.05 GB of RAM available. This allowed us to feed the network with. Large datasets enabled memory-consuming computations during training. 226 GB storage allocation (207 GB available for dataset storage, intermediate outputs, and model checkpoints). The computational backend included a multi-core AMD EPYC 7B13 CPU that handled preprocessing, data augmentation, and non-GPU-bound operations. This. A combination of GPU, CPU, RAM, and disk resources enabled the proposed generative adversarial network to be trained effectively and evaluated efficiently in the Google Colab environment.

5.3 Results on the Kaggle dataset

The training results showed a clear relationship between training time, batch size, and output quality. In the first three experiments, whose results are shown in Figures 6-8 using the facial expression recognition dataset, it was observed that Experiment 2 (100 episodes, batch size 16) achieved the best result, with FID = 88 and IS = 3.0. This indicates relatively high image quality and diversity.

The FER-2013 dataset had worse FID and visual quality results due to noise and class imbalance in the data with low-quality, in-the-wild face images. These characteristics make the dataset much harder than CK+, which is clean, well-aligned, and balanced. Therefore, it was challenging for the generator to learn well-formed emotion representations in FER-2013.

Figure 6. The training result represented by exp 4 in Table 1

Figure 7. Generated facial expressions compared with real samples (CK+ dataset)

Figure 8. The training result represented by exp 4 in Table 1

5.4 Training duration vs. performance

Increasing the number of episodes to 150 in Experiment 3 did not significantly improve the results (FID = 107, IS = 1.85), possibly due to over-conditioning or weak gradients. Interestingly, the longest training time (10 hours in Experiment 1) did not yield better results compared to shorter sessions with optimal batch size and system resources.

The training of the proposed DCGAN model is depicted in Figure 12, which shows how the training loss and training accuracy change over 10 epochs. The loss curve of Fig. 4 clearly presents a smooth descending trend from 0.9 to 0.1 with stable optimization and effective convergence in the adversarial learning procedure. This continual decline indicates that the PGAN is beneficial to help stabilize the convergence of the gradient flow and avoid training instability.

At the same time, the training accuracy is monotonically increasing from 65% at the first epoch to nearly 92% on the last epoch. This gradual increase indicates that the generator is learning to better retain the emotional content of facial expressions with more training. The inverse proportionality relation between the decreasing loss and the rising accuracy indicates a steady convergence with no signs of severe oscillation or mode collapse. Overall, these plots confirm the stability of training and the robustness of our proposed DCGAN architecture.

5.5 Results on the CK+ Dataset

For the CK+ dataset, experiments 4 and 5 demonstrate a significant increase in the quality of the initialization. Particularly in Experiment 5 initialization: 0.5, the generated facial expressions on an unclean or expanded dataset are incredibly close to a real human's face, despite the cleaner, smaller dataset. However, the initialization was still moderate, IS = 2.0, with the distinction that the images are uniform, which is clearly due to limited image variation between data. The results are given in Figures 9-11 Hence, the quality of facial classifications falls within the acceptable range. To support this assessment, we provide direct comparisons with well-known pre-trained models.

Figure 9. The training result represented by exp 4 in Table 1

Figure 10. The training result represented by exp 5 in Table 1

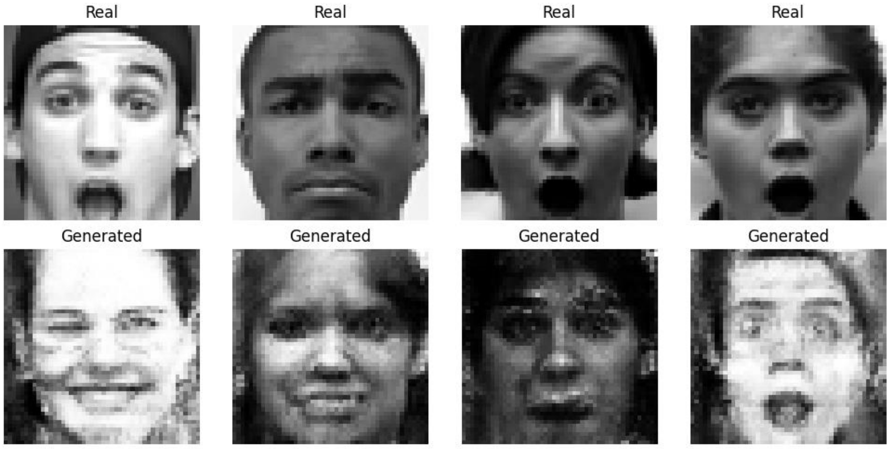

Figure 11. Generated and real images in exp 5

5.6 Observations

The results of this study show that the proposed DCGAN model can successfully generate realistic facial expressions, but its performance is highly dependent on the quality of the dataset and computational resources. Although more significant and complex datasets pose challenges for GAN training, the cleaner signals in the CK+ dataset lead to quicker convergence and more grounded output images. The employment of Pixel Normalization and LeakyReLU activation functions to improve training stability and production realism, and utilizing grayscale signal facial expression signals, allows for a reduction in computational overhead and focuses on the critical emotional characteristics. Despite strong findings in this regard, many challenges, such as unstable training, mode collapse, and high computational demand, persist. Future work should investigate more advanced GAN variants such as conditional and encoder-guided GANs. Increasing the datasets to further diversify subjects and support higher image resolution, incorporating more sophisticated feature representation methods such as attention or its derivatives, and assessing models on accurate datasets that entail bridge-domain adaptation between synthetic and genuine facial expressions.

A superior FID score of 0.5 is strong evidence of the proposed architecture's efficiency in generating visually convincing facial expressions even with limited data. Although the CK+ dataset's more balanced and cleaner samples allow for more stable convergence, the higher diversity and noise of FER2013 samples degrade generative fidelity. In any regard, these findings support the hypothesis that both dataset quality and architectural simplicity significantly influence GAN performance.

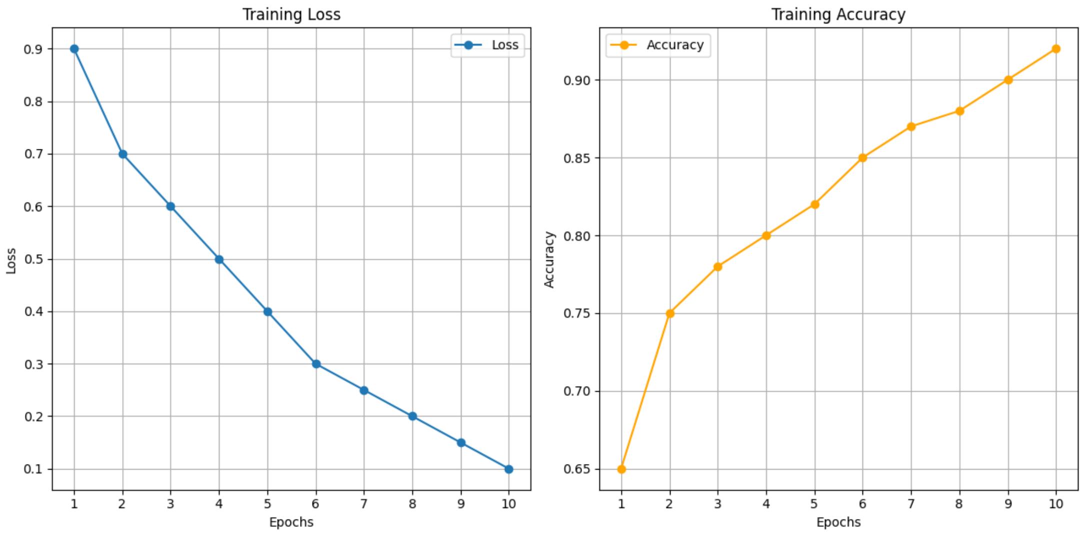

Figure 12. Training loss and training accuracy curves of the proposed GAN model on 10 epochs

Table 1. Training experiences and results

Experiments	Hyperparameter		Layer Gen.	Layer Dis.	Training Time	Resources	Result	Dataset	Comparison Model / Reference
	Epoche	*Bach size*							
Ex1	100	8	noise + dense + 3TransConv +output	input + 4conv2D + dense +flatten	10 h	System RAM 2.2 / 47.1 GB v5e-1 TPU	FID: 111 IS:2	Face expression recognition dataset	
Ex2	100	16	noise + dense + 3TransConv +output	input + 4conv2D + dense +flatten	4h 6m 48s	System RAM 5.7 / 334.6 GB Disk 20.9 / 225.3 GB v2-8 TPU	FID:88 IS:3	Face expression recognition dataset	StyleGAN2 [14] (FID = 34.6, IS = 3.5)
Ex3	150	16	noise + dense + 3TransConv +output	input +4 conv2D + dense +flatten	Usage rate: approximately 1.76 per hour 15h	System RAM 5.1 / 334.6 GB Disk 20.8 / 225.3 GB Change runtime type	FID: 107 IS: 1.85	Face expression recognition dataset	
Ex4	50	8	noise + dense + 3TransConv +output	input + 4conv2D + dense +flatten	1 h	runtime type v2-8 TPU	FID:50 IS:1.9	ck+	cGAN [17] (FID = 1.2, IS = 2.1)
Ex5	100	8	noise + dense + 3TransConv +output	input +4 conv2D + dense +flatten	1:30 h	System RAM 139.3 / 334.6 GB Disk 20.9 / 225.3 GB Change runtime type v2-8 TPU	FID:0.5 IS:2	ck+	

6- Conclusion

This study validates a novel method to generate high-quality synthetic facial expression images using a deep convolutional generative adversarial network. The model, with added Pixel Normalization and LeakyReLU activation functions, was able to produce stable training conditions and realistic outputs, specifically on SPL-derived, clean, and structured CK+ datasets. The experiments show that the success of this model is greatly influenced by the characteristics of the input datasets, the availability of training mechanism parameters, and computational power. In particular, the smaller but much cleaner and high-fidelity CK+ dataset, achieving a FID of 0.5, suggests that the model's performance can be significantly enhanced with high-quality datasets. In this study, the grayscale preprocessed expression data were utilized to ease color diversity-based feature attraction, make extraction work easier, and subsequently assist the model in generalizing expressions, making learning more straightforward. Unaddressed issues, such as limitations in image diversity, computational demands, and training instability, warrant further investigation. Future research should consider specialised GAN models, such as Conditional GANs and Encoder-Guided GANs, to improve the generation of emotion-containing faces. Higher, efficient training datasets with a large number of image dimensions can be considered to prefer the dataset. Transformers, or attention mechanisms, can be added to the network to select vital characteristics. Domain adaptation approaches used alongside the datasets should find the most similar naturally used images to capture real emotions.

The proposed model can be applied across various practical domains, including virtual avatars, emotion-driven animation systems, and affective computing. This framework enables grayscale realistic expression synthesis and is a robust solution for low-powered appliances and embedded AI. Additionally, future work will incorporate conditional emotion management and identity-preserving modules to increase the personalization and controllability of generative expressions.

Conflicts of Interest

The authors declare no conflicts of interest.

Funding

No funding.

References

[1] **T. Eglynas, D. Lizdenis, A. Raudys, S. Jakovlev, and M. Voznak**, "Exploring generative adversarial networks: comparative analysis of facial image synthesis and the extension of creative capacities in artificial intelligence," *IEEE Access,* vol. 12, pp. 19588-19597, 2025, doi: 10.1109/ACCESS.2025.3531726.

[2] **R. N. Abiram, P. Vincent, and P. Vincent**, "Identity preserving multi-pose facial expression recognition using fine tuned VGG on the latent space vector of generative adversarial network," *Mathmatical Bioscience Engineering*, vol. 18, no. 4, pp. 3699-3717, 2021, doi: 10.3934/mbe.2021186.

[3] **J. Wang**, "Improved facial expression recognition method based on gan," *Scientific Programming*, vol. 2021, p. 2689029, 2021, doi: 10.1155/2021/2689029.

[4] **P. Rajarajeswari et al**., "Performance evaluation of generative adversarial networks for anime face synthesis using deep learning approaches," *Multidisciplinary Science Journal*, vol. 7, no. 3, p. e2025093, 2025, doi: 10.31893/multiscience.2025093.

[5] **Y. Fan, X. Jiang, S. Lan, and S. You**, "Facial expression transfer based on generative adversarial networks: a review," in *Third International Conference on Computer Vision and Pattern Analysis (ICCPA 2023)*, Hangzhou, China, 2023: SPIE, pp. 325-333, doi: 10.1117/12.2684264.

[6] **J. R. Siddiqui**, "Explore the expression: Facial expression generation using auxiliary classifier generative adversarial network," *arXiv:2201.09061*, 2022, doi: 10.48550/arXiv.2201.09061.

[7] **B. Liu, J. Lv, X. Fan, J. Luo, and T. Zou**, "Application of an improved DCGAN for image generation," *Mobile Information Systems*, vol. 2022, p. 9005552, 2022, doi: 10.1155/2022/9005552.

[8] **A. Akram and N. Khan**, "US-GAN: on the importance of ultimate skip connection for facial expression synthesis," *Multimedia Tools and Applications*, vol. 83, pp. 7231-7247, 2024, doi: 10.1007/s11042-023-15268-2.

[9] **M. A. Iqbal, W. Jadoon, and S. K. Kim**, "Synthetic image generation using conditional GAN-provided single-sample face image," *Applied Sciences*, vol. 14, no. 12, p. 5049, 2024, doi: 10.3390/app14125049.

[10] **X. Ou**, "Deep Convolutional Generative Adversarial Networks (DCGAN)-Based Anime Face Generation," in *2024 2nd International Conference on Image, Algorithms and Artificial Intelligence (ICIAAI 2024)*, 2024: Atlantis Press, pp. 863-872, doi: 10.2991/978-94-6463-540-9_86.

[11] **Y. Sagawa and M. Hagiwara**, "Face image generation system using attribute information with DCGANs," in *Proceedings of the 2nd International Conference on Machine Learning and Soft Computing*, New York, NY, USA, 2018: ACM, pp. 109-113, doi: 10.1145/3184066.3184071.

[12] **B. Alyafi, O. Diaz, and R. Marti**, "DCGANs for realistic breast mass augmentation in x-ray mammography," in *Medical Imaging 2020: Computer-Aided Diagnosis*, Houston, Texas, United States, 2020: SPIE, pp. 473-480, doi: 10.1117/12.2543506.

[13] **S. N. Bushra and G. Shobana**, "A survey on deep convolutional generative adversarial neural network (dcgan) for detection of covid-19 using chest x-ray/ct-scan," in *2020 3rd International Conference on Intelligent Sustainable Systems (ICISS)*, Thoothukudi, India, 2020: IEEE, pp. 702-708, doi: 10.1109/ICISS49785.2020.9316125.

[14] **N. Bao, S. Guo, and J. Gao**, "Expression synthesis based on progressive growing of generative adversarial network and StyleGAN2," in *Third International Symposium on Computer Engineering and Intelligent Communications (ISCEIC 2022*), Xi'an, China, 2023: SPIE, pp. 531-540, doi: 10.1117/12.2660973.

[15] **Y. Shin, J. Bum, C.-H. Son, and H. Choo**, "BCGAN: Facial Expression Synthesis by Bottleneck-Layered Conditional Generative Adversarial Networks," in *2021 15th International Conference on Ubiquitous Information Management and Communication (IMCOM)*, Seoul, Korea (South), 2021: IEEE, pp. 1-4, doi: 10.1109/IMCOM51814.2021.9377424.

[16] **M. Z. Khan *et al*.**, "A realistic image generation of face from text description using the fully trained generative adversarial networks," *IEEE Access*, vol. 9, pp. 1250-1260, 2020, doi: 10.1109/ACCESS.2020.3015656.

[17] **J. Cai *et al*.**, "Identity-free facial expression recognition using conditional generative adversarial network," in *2021 IEEE International Conference on Image Processing (ICIP)*, Anchorage, AK, USA, 2021: IEEE, pp. 1344-1348, doi: 10.1109/ICIP42928.2021.9506593.

[18] **L. Alharbawee and N. Pugeault**, "Generative adversarial networks for facial expression recognition in the wild," *International Journal of Computing and Digital Systems*, vol. 16, no. 1, pp. 1259-1282, 2024, doi: 10.12785/ijcds/160193.

[19] **Kaggle**, "Face expression recognition dataset," 2018. [Online]. Available: https://www.kaggle.com/datasets/jonathanoheix/face-expression-recognition-dataset

[20] **Kaggle**, "CK+ Dataset," 2022. [Online]. Available: https://www.kaggle.com/datasets/davilsena/ckdataset