International Journal of Electrical and Computer Engineering (IJECE)
Vol. 15, No. 3, June 2025, pp. 3418~3428
ISSN: 2088-8708, DOI: 10.11591/ijece.v15i3.pp3418-3428

A novel hybrid generation technique of facial expressions using fine-tuning and auxiliary condition generative adversarial network

Mai Elbadwey[1], Hani Attar[2,3], Ayman Amer[2], Khaled H. Ibrahim[4], Somaya Kayed[5], Tamer Barakat[4]
[1]Department of Electronics and Communication Engineering, Pyramids Higher Institute for Engineering and Technology, Giza, Egypt
[2]Faculty of Engineering, Zarqa University, Zarqa, Jordan
[3]Faculty of Engineering, University of Business and Technology, Jeddah, Saudi Arabia
[4]Department of Electrical Engineering, Faculty of Engineering, Fayoum University, Fayoum, Egypt
[5]Department of Electronics and Computer Department, Nahda University, Beni Suef, Egypt

Article Info

Article history:

Received Sep 12, 2024
Revised Mar 10, 2025
Accepted Mar 20, 2025

Keywords:

Condition generative adversarial networks
Convolution neural network
Deep learning
Discriminator
FERG dataset
Generator
Hyperparameter tuning model

ABSTRACT

The facial expression generation continues to be interesting for researchers and scientists worldwide. Facial expressions are an excellent way of transmitting one's emotions or intentions to others, and they have been extensively studied in areas such as driver safety, human-computer interaction (HCI), deception detection, health care, and monitoring. The facial expression generation starts with a single neutral image and generates sequences of facial expression images, which are combined to create a video. Previous methods generated facial expression images of the same person. However, they still suffer from low accuracy and image quality. This article overcomes this problem using a novel hybrid model for facial expression video generation using fine-tuning and condition-generative model architectures to optimize the model's parameters. Results indicate that the proposed novel approach significantly improves the expression generation of the same person. The proposed method can reliably and accurately generate facial expressions, with a testing accuracy of 98.7% and a training accuracy of 99.9%.

Corresponding Author:

Mai Elbadwey
Department of Electronics and Communication Engineering, Pyramids Higher Institute for Engineering and Technology
Giza, Egypt
Email: me1523@fayoum.edu.eg

1. INTRODUCTION

The importance of facial information has been reflected in several research papers devoted to its recognition during the last century. Recent advancements in deep learning technology have enabled machines to create facial expressions effectively [1]. While facial expression recognition has progressed, it remains a subject of ongoing research.

Facial expression generation is an essential topic in emotion detection. The goal is not just to assign an image to a set of predefined labels but also to build a realistic image that accurately reflects a given facial emotion. Facial expression generation is especially significant since it requires a deeper understanding of emotions, moving beyond simple classification to capture the complexities of human expression. This capacity enables more advanced applications, such as virtual avatars or human-computer interaction systems, that require realistic and dynamic expressions to communicate naturally [2].

Generative adversarial networks (GANs) can generate new facial expressions based on the desired expression label from an input face picture. Although GANs have shown extraordinary effectiveness in facial expression generation, current approaches have several limitations. Psychological studies show that the mouth and eyes are crucial for facial expression [3].

Previous studies in diagnostic research have yielded consistent findings relating to basic emotions. The eyes are more expressive of intense emotions like anger, fear, and sadness than other facial parts, while the mouth is better at showing feelings of happiness and disgust. However, earlier GAN-based algorithms have often focused on the face, ignoring these smaller facial regions in facial expression synthesis, resulting in overlapping and blurring of the obtained results. In light of this discovery, researchers investigated and applied local facial region features to synthesize facial expressions [4], [5].

Ian Goodfellow proposed the GAN architecture for convolution neural networks in 2014 [6]. A basic GAN design consists of an adversarial generator and discriminator network that work together to achieve the same goal. The GAN architecture is designed to solve the problem of synthesizing facial expressions. However, most current studies on creating facial expression pictures rarely focus on improving the emotional impact of facial expression image creation [6], [7].

To address the challenges of low quality, noise, and identity preservation skills that frequently plague one-step approaches, researchers suggest two feature extractors: the identity feature extractor and the spatial feature extractor, which encode the input image into multiple codes in the latent space. Additionally, experts recommend leveraging the excellent performance of modern face recognition systems by using the identity encoder as a previously trained feature extractor for facial recognition [8].

Many research efforts have focused on facial emotion transmission from a single image using traditional graphics-based methods. Specific techniques employ 2D or 3D warps to manipulate the input face to approximate the goal expression directly. Generative modeling techniques based on deep learning are data-driven, whereas computer graphics techniques are considered theory-driven approaches. Many strategies have been proposed for translating facial expressions from one identity to another using the GAN [9], [10].

Several studies aimed to create a face with controlled position and expression using a neural network [11], [12]. While these methods are often difficult to achieve with high accuracy, in [12], a deep neural network based on VGG-face analysis attained an accuracy of 76%. Wiles *et al.* [13] used a similar approach but a different strategy. In a previous study, some researchers used an automatic single-hot vector to encode the desired expression in a GAN, resulting in faces representing discrete emotion states [14], [15]. Despite its simplicity, this technique only produces specific facial expressions, leading to a limited range of samples. Ding *et al.* [16] proposed an expressive GAN (ExprGAN) to change facial expressions containing an emotion controller system. This system stores advanced information about the desired expression, including intensity fluctuations, which are a vector of true values conditioned by the label.

Other studies use facial geometry to encode a desired expression via facial landmark placement [17], [18]. This strategy is more adaptable, allowing for continuous facial emotions and the transmission of facial expressions. The outcomes for expression synthesis achieved 66.6% accuracy. Most researchers proposed a technique for fitting the overall direction of various face emotions in the feature space by encoding the supplied image and then applying a generalized linear model (GLM) [19]. This method can change the emotion associated with an input image and its related affect condition. Its design is built on a single generator and three discriminators.

Using paired datasets, researchers used U-Net-based generators and PatchGAN discriminators to learn the mapping between two domains. Sanghyuck *et al.* [20] developed CycleGAN, a network consisting of two generators and two discriminators that learn cyclical cross-domain mappings [21]–[23]. They employed Johnson *et al.* [21] design in their generator. Pix2pix and CycleGAN can be used to learn facial emotion mappings, however the results are not realistic. Arbish *et al.* [24] presented US-GAN as a more compact and efficient model for facial expression synthesis. US-GAN outperforms previous cutting-edge methods by 25% in facial realism, 43% in emotional quality, and 58% in identity preservation.

Facial expression generation using GANs has significantly advanced computer vision. However, various challenges hinder their practical deployment and effectiveness. One major issue is inefficient training, as GAN-based models for facial expression generation often require large amounts of high-quality annotated data and significant computational resources. This makes them challenging to implement real-time applications, especially on resource-constrained devices like mobile phones or embedded systems.

Furthermore, these models often struggle to control expression variability precisely, resulting in realistic facial expressions with consistent intensity across diverse demographics. Another issue is poor generalization across datasets, meaning that GANs trained on one dataset may fail to generate high-quality expressions when applied to new or unseen datasets with different lighting conditions, ethnicities, or facial characteristics.

Finally, many existing models face a trade-off between realism and diversity, excelling at producing realistic expressions for a single emotion but failing to represent the full range of human emotions with equal accuracy. These issues highlight the need for a more robust, efficient, and adaptable facial expression generation system that balances realism and variety while remaining computationally efficient, capable of generalizing across different datasets, and suitable for real-time deployment on resource-constrained devices.

Motivated by the discussion above, this article presents a novel hybrid generation technique using a tuning model with a conditional GAN. To address the issues of limited training samples and low classification accuracy in facial expression recognition, this article employs auxiliary classifier GAN (AC-GAN) to produce sample data, enhancing the dataset and reducing overfitting. To summarize, the key contributions are:

- An efficient generation technique for facial expressions.
- Utilization of cutting-edge facial recognition technology, featuring a novel hybrid model for facial expression video production using fine-tuning and conditional generative model architecture to improve the model's parameters.
- The model is evaluated directly and qualitatively against public facial expression benchmarks, including the FERG dataset.
- The hybrid model can reliably and accurately generate facial expressions, with a testing accuracy of 98.7% and a training accuracy of 99.9%.

2. PROPOSED MODEL

The proposed model consists of the following procedures, as illustrated in Figure 1. This article aims to generate more accurate dynamic facial expressions for the same individual. The architecture of the proposed model includes an auxiliary classifier GAN and hyperparameter tuning.

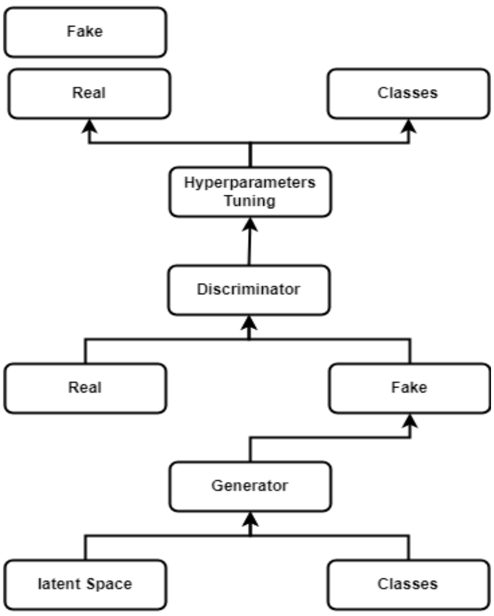

Figure 1. The hybrid generation method of facial expression

The auxiliary classifier GAN comprises two phases: the generator and the discriminator. The generator produces facial expression images of the same person, while the discriminator classifies the generated and real images to enhance the model's accuracy. The final phase involves hyperparameter tuning to determine the best settings for optimizing the model. Each phase of the model will be explored further in the subsequent sections.

The AC-GAN, is a modified version of the conditional GAN that predicts an image's class label using the discriminator rather of receiving it [25], [26]. The training process stabilizes and allows for the generation of large-sized, high-quality images while learning a representation in the latent space that is independent of the class label. The architecture is designed so that the discriminator and auxiliary classifier can be considered independent models with shared weights. The discriminator and the auxiliary classifier can be combined to generate a single neural network with two outputs.

The initial output is a single probability generated by the sigmoid activation function, which demonstrates the "realness" of the input image and can be improved using binary cross-entropy, similar to a regular GAN discriminator model. The second output is the chance that the image belongs to each class, determined using the softmax activation function, which, like any other multi-class classification neural network model, has been augmented with categorical cross-entropy. Figure 2 demonstrates how the conditional attribute vector combines with the visual representation in convolutional layers.

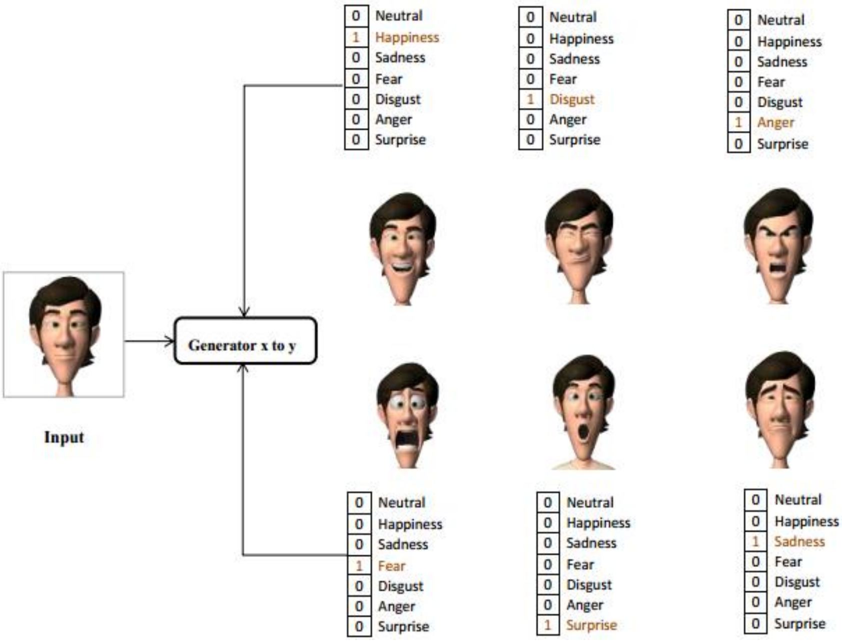

Figure 2. The framework of the proposed technique. Image input can be converted to a new, changed expression face image, led by the facial expression attribute vector

2.1. Auxiliary classifier generative adversarial networks

AC-GAN, an enhanced version of GAN, employs two deep neural networks: a generator and a discriminator. These networks engage in a competitive, adversarial training process to generate synthetic data indistinguishable from real data. Using a random Gaussian noise vector as input, the generator aims to generate realistic data that fools the discriminator, whose goal is to consistently distinguish between "real" data from training data and "fake" data produced by the generator.

AC-GANs differ from the conditional GAN design in that they allow the discriminator to guess both the class label and its information source (i.e., "real" or "fake" status). The generator aims to diminish the discriminator's ability to distinguish between real and false data while improving its ability to predict the class label of manufactured data. This frequently results in a more stable training process, allowing the GAN to obtain a latent space description.

2.1.1. Generator model

The generator architecture includes an encoder and a decoder network [27]. By learning its distribution, the encoder network converts an input image to a latent space. The encoder network contains several blocks of down sampling layers. Downsampling blocks have three levels: convolution, batch normalization, and activation function layers.

Downsampling blocks repeatedly occur until the image is one-dimensional. The earlier layers have more information than the subsequent layers, which have a set number of features. The output of the decoder's last layer is flattened and joined to a fully connected set of n-dimensional layers (μ, σ). These layers define the parameters of the latent space distribution. Figure 3 illustrates the design of an encoding and decoding network. The decoder network receives the effect and latent vectors from the encoder output. Another set of thick layers is added, which are combined and modified. There are six upsampling blocks, each with two-dimensional transpose convolutions, batch normalization, and activation layers.

The generator receives two parameters instead of one. It aims to create a picture for a specific class using a class label and random points from the latent space as the source. The name refers to how the process of producing and classifying images depends on the class label when it is added as an input. The discriminator network takes the generating network's output and validates it by discriminating between actual and fraudulent outputs.

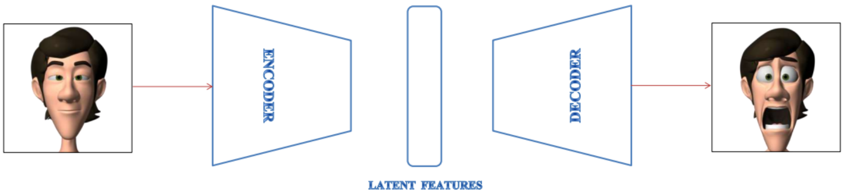

Figure 3. Generator architecture

Each attribute is distinguished from the others using a vector that represents the conditional attribute. Only the element in the attribute vector that matches the label is set to 1, with the remaining elements being set to 0. Next, in the generator G ($G_{x \to y}$) a fully connected layer concatenated the vector with the image embedding vector. By switching the two matching expressions in the generator $G_{x \to y}$, the expression vector can be altered. The conditional label can guide the transition from one expression to another. GAN autonomously learns this correspondence during training. Because of this, Expression condition GAN (ECGAN) can add the conditional vector to modify a face.

2.1.2. Discriminator model

The discriminator's role is to know the difference between untrue and true data [28]. The discriminator model is used to update the generator model rather than directly. The architecture of the discriminator tends to be very similar to that of traditional convolutional neural networks (CNNs). The discriminator network consists of three main layer blocks, each with three layers: a batch-normalization layer, a rectified linear unit (ReLU) activation function, and a 2D convolution layer with strides. A final dense classification network is created by gradually increasing and flattening the features learned at each block. Its last layers comprise a softmax layer which predicts facial expression in an input image while the sigmoid layer decides whether a decision is real or fake. By feeding it with real and fake images, the discriminator can determine if an image is real or fake and predict its class label.

The feature information obtained from the damaged samples will be returned to the generator. Similarly, the simulation results of real samples are also used as feedback. In this backpropagation strategy, all feedback methods pass the appropriate feature values, and the generator corrects newly produced expression characteristics based on the feedback characteristic values.

Although this is an arbitrary number, the model is intended to run for 50 training epochs. The batch size is set to 64 samples. During training, the discriminator model is adjusted with half a batch of true and half a batch of false samples, resulting in one set of weight changes. The generator is then changed using the merged GAN model. When utilizing the train_on_batch() function, both the discriminator and the hybrid model give three loss values. The first number represents the sum of the loss values, which can be ignored. The second value represents the loss in the real/fake output layer. The third value is the loss for label classification.

2.1.2. Loss functions

GANs use loss functions to train both the discriminator and the generator. Throughout training, the loss function aids in adjusting the weights of these models to enhance their performance. The mini-max function involves the generator's encoder and decoder network blocks, incorporating the reconstruction and GAN loss [29]. As shown in (1) illustrates the adversarial loss in the generator. The aggregate loss, encompassing the adversarial, reconstruction, and KL-divergence losses, signifies the overall loss of the generator.

$$l_{Adv}^{G}\left(I_g, A_t, y_d, y_{dm}\right) = \emptyset_b(1, y_d) + \emptyset_m(A_t, y_{dm}) \tag{1}$$

where, I_g generated image, A_t the target effect, $y_d, y_{dm}, \emptyset_b, \emptyset_m$. The output of the discriminator D.

Moreover, not only does the generator network generate characters presented on input images, but it also learns how to change facial expressions given as a target effect. This network learns to generate realistic facial expressions by considering the specific input image and the desired expression. As can be seen from (2) and (3), discriminator loss is separately computed for true images and untrue ones.

$$l_{real}(A_s, y_d, y_{dm}) = \emptyset_b(1, y_d) + \emptyset_m(A_s, y_{dm}) \tag{2}$$

$$l_{fake}(A_t, y_d, y_{dm}) = \emptyset_b(0, y_d) + \emptyset_m(A_t, y_{dm}) \tag{3}$$

Following this, the discriminator loss is expressed in (4) as the sum of the losses on the real and fake images. The total generator loss is shown in (5), which depends on Alpha, Beta, and Gamma values. The model's performance and specific requirements will determine the values to choose for Alpha, Beta, and Gamma. To achieve the best results for your particular transfer and dataset, consider using Alpha=0.8, Beta=0.2, and Gamma=5.

$$l_{Adv}^{D} = l_{real} + l_{fake} \quad (4)$$

$$gen\_loss = Alpha * gan\_loss + Beta * kl\_loss + Gamma * l1\_loss \quad (5)$$

2.2. Hyperparameter tuning model

Begin the search for the optimal hyperparameter setup [30]. All arguments supplied to the search function are forwarded to the model.fit() method called in each run. Remember to pass validation_data when evaluating the model. The model-building function is invoked with varying hyperparameter values in each trial during the search. To construct the model, the tuner generates another set of hyperparameter values for each trial. Next, the model is fitted and evaluated, with metrics being recorded. After thoroughly exploring the parameter space, the tuner eventually identifies a good set of hyperparameter values. Once the search is complete, you can obtain the top model or models. The model is saved at the epoch where it performed best on the validation_data. The tuning model has several layers that work together to fine-tune the parameters and get the best results. These layers include convolution, pooling, flattening, and dense layers.

The convolutional layers set the number of filters and their settings for a 2D convolution layer. Convolutional Layers specify the number of filters and their hyperparameters for a 2D convolution layer. Steps sixteen are used to determine The number of filters for both initial convolution layers, ranging from 32 to 128 for the first layer and 32 to 64 for the second layer, using the ReLU activation function. Pooling layers use maximum pooling to reduce the size of feature maps. The flattening layer connects to the dense (connected) layers, turning 2D matrices into a 1D vector. The dense layer is a linked layer with settings that control the number of units (from 32 to 128 in step 16). The final dense layer gives out probabilities for seven classes using softmax activation.

2.3. Comparison between previous studies and the proposed model

Previous models noticed a reduction in the quality of their outputs when they included more classes using the same model. However, the AC-GAN model separates huge datasets into subsets based on classes. This enables the generator and discriminator models to be trained independently for each subset. A fine-tuning model is an excellent method for adjusting significant models to specific parameters. It is an important step in enhancing the quality and efficiency of the model.

3. SIMULATION, RESULTS AND DISCUSSION

3.1. Case study

As described in the introduction, the dataset used for this study is FERG-DB. The dataset included in this study consists of synthetic images of characters produced in a previous study [26] to study facial expressions. 555,767 images of six characters with seven distinct emotions make up the dataset. There are 10,000 images for every effect, each featuring a range of facial expressions to convey the desired effect. This offers a wealth of knowledge that is difficult to obtain when utilizing pictures of actual people.

The dataset is preprocessed using a set of techniques. Specifically, the image is scaled, normalized, and then subjected to random noise via cropping and jitter. The normalization is usual, producing pictures ranging from -1 to 1. After preprocessing all photographs, the dataset is split into training and validation. This study's training set features 44.6 K images, whereas the validation set has 11.1 K.

3.2. Simulation of the proposed model

The generator and discriminator models are trained separately using AC-GAN. Weights are modified throughout every training phase depending on the generator and discriminator's loss computations. The target photos for each batch are randomly selected based on the intended effect and character in the source image. This study conducts training in 1875-batches size for each epoch, with the process repeated 50 epochs. After applying the encoder to a set of neutral pictures of each letter that serve as reference pictures, the developers get a collection of latent vectors. These latent vectors are then sent into the decoder network, along with a list of intended face expressions (one for each effect). As shown in Figure 4, the proposed strategy was used to convert every expression translation from the same person into a video. After applying hyperparameter tuning to the model, the best value for the model's parameters was obtained.

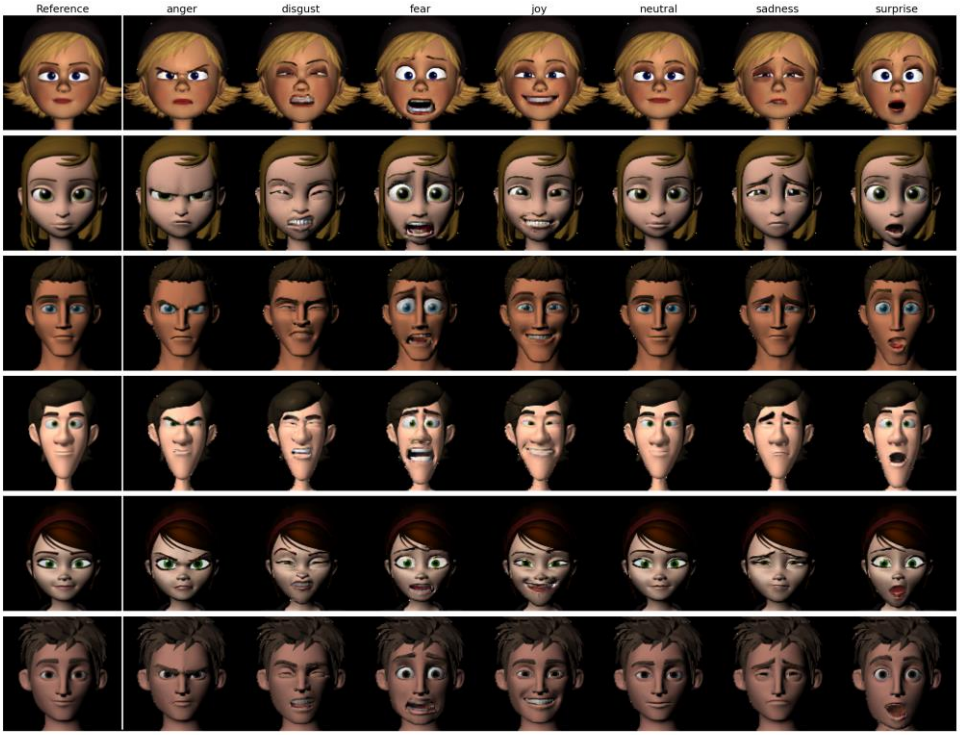

Figure 4. Results of different emotional expressions for each identity

The results of the hyperparameter optimization indicate that AC-GAN, using specific parameter values, outperforms other combinations. The optimal parameters for the model are listed in Table 1. The proposed model evaluates over ten epochs with a batch size of 1875. The relationship between loss value, training accuracy, and validation accuracy over the ten epochs is depicted in Figure 5.

Table 1. The best parameters for the hyperparameter tuning model

Best Value	Hyperparameter
Conv_1_Filter	80
Conv_2_Filter	64
Dense_1_units	32
Learning_rate	0.01

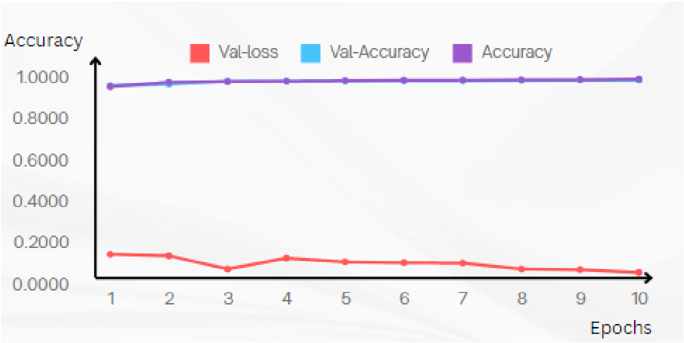

Figure 5. Loss and accuracy values for ten epochs

3.3. Discussion

To evaluate the effectiveness of the proposed method, a comparison is made between state-of-the-art GANs, such as StarGAN and GANImation. Like CycleGAN, StarGAN is an image-to-image conversion but takes a unified method in which a generator is trained to map input images to a user-selected multi-target domain. This strategy significantly reduces the number of parameters by sharing generator weights across domains. The proposed method achieved improved levels of accuracy. Table 2 compares the validation accuracy of various approaches.

Table 2. Studies in literature and accuracy performances of the proposed approach

Method	Dataset	Accuracy %
VGG-FACE using Landmark Localization [12]	4DFAB	0.7687
IF-GAN [31]	RAF-DB	0.8833
StarGAN [32]	RAF-DB	0.8851
	CFEED	0.8187
Cascade EFGAN [33]	RAF-DB	0.9365
FEX-GAN [26]	FERG	0.9138
The proposed method	FERG	0. 9876

4. CONCLUSION

This article enhances image-to-image generation and image-to-video transformation. The proposed model can create high-quality images of a person's facial emotions from one neutral image. This approach's effectiveness is validated by tests comparing discriminative indexes and image-generating effects. The novel hybrid technique effectively solves the problem of overfitting and low image quality for generating facial expressions. A future study aims to increase the accuracy and reliability of facial expression classification by generating higher-quality facial expression images. As confirmed by test results, this method greatly outperforms current baselines for dynamic facial expression generation. The proposed model solved these problems to produce higher-quality dynamic facial emotions with a validated accuracy of 98.7%. This study has significant limitations. The model is applied to only one dataset, FERG. In future studies, researchers aim to apply this method to other datasets. They hope to enhance the proposed technique by including micro-expressions and simulated behavior.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Mai Elbadwey	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	
Hani Attar		✓			✓	✓		✓	✓	✓	✓	✓		✓
Ayman Amer	✓				✓	✓			✓	✓	✓		✓	✓
Khalid H. Ibrahim	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	
Somaya Ismail					✓	✓	✓			✓	✓	✓	✓	
Tamer Barakat		✓			✓	✓	✓		✓	✓	✓	✓	✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

INFORMED CONSENT

We have obtained informed consent from all individuals included in this study.

DATA AVAILABILITY

The data that support the findings of this study are available from the University of Washington. Restrictions apply to the availability of these data, which were used under license for this study. Data are available from "D. Aneja, A. Colburn, G. Faigin, L. Shapiro, and B. Mones" with the permission of the University of Washington [26].

REFERENCES

[1] L. Zhang *et al.*, "ARFACE: Attention-aware and regularization for face recognition with reinforcement learning," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 4, no. 1, pp. 30–42, Jan. 2022, doi: 10.1109/TBIOM.2021.3104014.

[2] L. Liang, C. Lang, Y. Li, S. Feng, and J. Zhao, "Fine-grained facial expression recognition in the wild," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 482–494, 2021, doi: 10.1109/TIFS.2020.3007327.

[3] E. A. Martin *et al.*, "Behavioral measures of psychotic disorders: Using automatic facial coding to detect nonverbal expressions in video," *Journal of Psychiatric Research*, vol. 176, pp. 9–17, Aug. 2024, doi: 10.1016/j.jpsychires.2024.05.056.

[4] S. Jia, S. Wang, C. Hu, P. J. Webster, and X. Li, "Detection of genuine and posed facial expressions of emotion: Databases and methods," *Frontiers in Psychology*, vol. 11, Jan. 2021, doi: 10.3389/fpsyg.2020.580287.

[5] S. Li and W. Deng, "Deep facial expression recognition: a survey," *IEEE Transactions on Affective Computing*, vol. 13, no. 3, pp. 1195–1215, Jul. 2022, doi: 10.1109/TAFFC.2020.2981446.

[6] M. Castelle, "The social lives of generative adversarial networks," in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, Jan. 2020, p. 413. doi: 10.1145/3351095.3373156.

[7] M. Alghanim, H. Attar, K. Rezaee, M. Khosravi, A. Solyman, and M. A. Kanan, "A hybrid deep neural network approach to recognize driving fatigue based on EEG signals," *International Journal of Intelligent Systems*, vol. 2024, no. 1, Jan. 2024, doi: 10.1155/2024/9898333.

[8] H. Bouzid and L. Ballihi, "Facial expression video generation based-on spatio-temporal convolutional GAN: FEV-GAN," *Intelligent Systems with Applications*, vol. 16, Nov. 2022, doi: 10.1016/j.iswa.2022.200139.

[9] S. Ramis, J. M. Buades, F. J. Perales, and C. Manresa-Yee, "A novel approach to cross dataset studies in facial expression recognition," *Multimedia Tools and Applications*, vol. 81, no. 27, pp. 39507–39544, Nov. 2022, doi: 10.1007/s11042-022-13117-2.

[10] N. Otberdout, C. Ferrari, M. Daoudi, S. Berretti, and A. Del Bimbo, "Sparse to dense, dynamic 3D facial expression generation," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2022, pp. 20353–20362. doi: 10.1109/CVPR52688.2022.01974.

[11] Y. Ren, G. Li, Y. Chen, T. H. Li, and S. Liu, "PIRenderer: Controllable portrait image generation via semantic neural rendering," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct. 2021, pp. 13739–13748. doi: 10.1109/ICCV48922.2021.01350.

[12] D. Kollias, S. Cheng, E. Ververas, I. Kotsia, and S. Zafeiriou, "Deep neural network augmentation: Generating faces for affect analysis," *International Journal of Computer Vision*, vol. 128, no. 5, pp. 1455–1484, May 2020, doi: 10.1007/s11263-020-01304-3.

[13] O. Wiles, A. S. Koepke, and A. Zisserman, "X2Face: A network for controlling face generation using images, audio, and pose codes," in *Computer Vision – ECCV 2018*, 2018, pp. 690–706. doi: 10.1007/978-3-030-01261-8_41.

[14] Y.-H. Lai and S.-H. Lai, "Emotion-preserving representation learning via generative adversarial network for multi-view facial expression recognition," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, May 2018, pp. 263–270. doi: 10.1109/FG.2018.00046.

[15] N. Otberdout, M. Daoudi, A. Kacem, L. Ballihi, and S. Berretti, "Dynamic facial expression generation on Hilbert hypersphere with conditional Wasserstein generative adversarial nets," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 2, pp. 848–863, Feb. 2022, doi: 10.1109/TPAMI.2020.3002500.

[16] H. Ding, K. Sricharan, and R. Chellappa, "ExprGAN: Facial expression editing with controllable expression intensity," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, Apr. 2018, doi: 10.1609/aaai.v32i1.12277.

[17] L. Song, Z. Lu, R. He, Z. Sun, and T. Tan, "Geometry guided adversarial facial expression synthesis," in *Proceedings of the 26th ACM international conference on Multimedia*, Oct. 2018, pp. 627–635. doi: 10.1145/3240508.3240612.

[18] C. Liu, K. Hirota, J. Ma, Z. Jia, and Y. Dai, "Facial expression recognition using hybrid features of pixel and geometry," *IEEE Access*, vol. 9, pp. 18876–18889, 2021, doi: 10.1109/ACCESS.2021.3054332.

[19] M. Chen, C. Li, K. Li, H. Zhang, and X. He, "Double encoder conditional GAN for facial expression synthesis," in *2018 37th Chinese Control Conference (CCC)*, Jul. 2018, pp. 9286–9291. doi: 10.23919/ChiCC.2018.8483579.

[20] K. Rezaee *et al.*, "Diagnostic tools for detecting autism spectrum disorder: A review," in *2022 International Engineering Conference on Electrical, Energy, and Artificial Intelligence (EICEEAI)*, Nov. 2022, pp. 1–6. doi: 10.1109/EICEEAI56378.2022.10050460.

[21] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Computer Vision – ECCV 2016*, 2016, pp. 694–711. doi: 10.1007/978-3-319-46475-6_43.

[22] Y. Huang and H. Yin, "Manifold-enhanced CycleGAN for facial expression synthesis," *The Imaging Science Journal*, vol. 70, no. 3, pp. 181–193, Apr. 2022, doi: 10.1080/13682199.2022.2164404.

[23] Y. Zhou, B. Yang, Z. Liu, Q. Wang, and P. Xiong, "Cross-domain facial expression recognition by combining transfer learning and face-cycle generative adversarial network," *Multimedia Tools and Applications*, vol. 83, no. 42, pp. 90289–90314, Mar. 2024, doi: 10.1007/s11042-024-18713-y.

[24] M. A. Deif, H. Attar, A. Alrosan, A. A. A. Solyman, and S. M. F. Abdelaliem, "Design and development of an intelligent neck and head support system based on eye blink recognition for cervical dystonia," *Discover Applied Sciences*, vol. 6, no. 11, Nov. 2024, doi: 10.1007/s42452-024-06291-5.

[25] J. Cho and K. Yoon, "Conditional activation GAN: Improved auxiliary classifier GAN," *IEEE Access*, vol. 8, pp. 216729–216740, 2020, doi: 10.1109/ACCESS.2020.3041480.

[26] J. R. Siddiqui, "Explore the expression: Facial expression generation using auxiliary classifier generative adversarial network," *Prepr. arXiv.2201.09061*, Jan. 2022.

[27] J. Gui, Z. Sun, Y. Wen, D. Tao, and J. Ye, "A review on generative adversarial networks: Algorithms, theory, and applications," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 4, pp. 3313–3332, Apr. 2023, doi: 10.1109/TKDE.2021.3130191.

[28] E. Schonfeld, B. Schiele, and A. Khoreva, "AU-Net based discriminator for generative adversarial networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 8207–8216.
[29] A. S. Saqlain, F. Fang, T. Ahmad, L. Wang, and Z. Abidin, "Evolution and effectiveness of loss functions in generative adversarial networks," *China Communications*, vol. 18, no. 10, pp. 45–76, Oct. 2021, doi: 10.23919/JCC.2021.10.004.
[30] L. Liao, H. Li, W. Shang, and L. Ma, "An empirical study of the impact of hyperparameter tuning and model optimization on the performance properties of deep neural networks," *ACM Transactions on Software Engineering and Methodology*, vol. 31, no. 3, pp. 1–40, Jul. 2022, doi: 10.1145/3506695.
[31] J. Cai *et al.*, "Identity-free facial expression recognition using conditional generative adversarial network," in *2021 IEEE International Conference on Image Processing (ICIP)*, Sep. 2021, pp. 1344–1348. doi: 10.1109/ICIP42928.2021.9506593.
[32] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, "StarGAN: Unified Generative Adversarial Networks for Multi-domain Image-to-Image Translation," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Jun. 2018, pp. 8789–8797. doi: 10.1109/CVPR.2018.00916.
[33] R. Wu, G. Zhang, S. Lu, and T. Chen, "Cascade EF-GAN: Progressive facial expression editing with local focuses," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2020, pp. 5020–5029. doi: 10.1109/CVPR42600.2020.00507.

BIOGRAPHIES OF AUTHORS

Mai Elbadwey is an associate lecturer at the Department of Electronics and Communication Engineering at Pyramids Higher Institute for Engineering and Technology. She received a B.Eng. degree in electronics and communication engineering from Fayoum University, Egypt, in 2011 and an M.S. in electronics and communication engineering from Fayoum University in 2021. Her research interests include computer science, image processing, artificial intelligence, fuzzy, genetic algorithms, CNN, and the GAN approach. She published her paper in the Fayoum University Journal of Engineering titled "Early classification of brain tumor on image histogram using fuzzy genetic algorithm." She published the search "Enhanced model for facial expression classification system" in the METSZET journal. She can be contacted at me1523@fayoum.edu.eg.

Hani Attar earned his Ph.D. in electrical and electronic engineering from the University of Strathclyde, United Kingdom, in 2011. Since then, he has been working as a researcher in electrical engineering and energy systems. Dr. Attar is currently a university lecturer at Zarqa University in Jordan. His research interests focus on network coding, wireless sensor networks, and wireless communications. He can be reached at Hattar@zu.edu.jo.

Ayman Amer received his Ph.D from Tennessee University in 1991. Dr. Amer is the Energy Engineering Head of the Department at Zarqa University, and his research interest topics are renewable energy, energy control, and energy storage. Moreover, smart design for energy-efficient designs is an interesting topic for Dr. Amer. He can be contacted at email: Aamer@zu.edu.jo.

Khalid H. Ibrahim is an assistant professor, Department of Electrical Engineering, Faculty of Engineering, Fayoum University, Fayoum, Egypt. He supervised several engineering projects as a representative of the College of Engineering. He trained the engineers' staff of the Komoshim industrial zone in June 2012. Training for mechatronics engineers in August 2015. Redesign of Komoshim substation to solve the power quality problem in some factories there, 2015. Head of Faculty of Engineering Training Center Fayoum University (2015 to 2017). Design of electrical distribution systems for many projects recently. Design of PV systems for different projects. Training for engineers in renewable energy. Supervision of many graduate projects in smart systems and renewable energy. He can be contacted at email: khi00@fayoum.edu.eg.

Somaya Ismail is a professor in the electronics and computer department at Nahda University. She was an associate professor and a head of the electrical department (electronics, communication, computer, and control engineering) at Obour Higher Institute for Engineering and Technology. She graduated in 1987 from Ain Shams University with a BSc in the electronics and communications department, with a general grade (very good), 80% of my order of merit, 13 of the successful students totaling (75). Her graduation project tackled distinction, and she was top-ranked as the 13th in her class. In 1997, she finished her Master of Science (MSc) in the same department. Afterwards, in 2000, she gained her Ph.D. from Ain Shams University under Prof. Hani Fikry Ragaie et al. She was an acting dean for the 2019 first term at Obour Higher Institute for Engineering and Technology. She has published in local and international conferences many scientific papers as well as multiple books in her related research interests (analog and digital VLSI design, current conveyor, nanoelectronics). She can be contacted at email: skayed@gmail.com.

Tamer Barakat received his B.Sc. in communications and computer engineering from Helwan University, Cairo, Egypt in 2000. He received his M.Sc. in cryptography and network security systems from Helwan University in 2004 and his Ph.D. in cryptography and network security systems from Cairo University in 2008. Currently, he is working as a lecturer and post-doctoral researcher and is also joining several network security projects in Egypt. His main interests are cryptography and network security. More specifically, he is working on designing efficient and secure cryptographic algorithms, particularly security in wireless sensor networks. Other things that interest him are number theory and the investigation of mathematics for designing secure and efficient cryptographic schemes. He can be contacted at email: tmb00@fayoum.edu.eg.