IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 22 April 2023, accepted 28 April 2023, date of publication 1 May 2023, date of current version 8 May 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3272032

RESEARCH ARTICLE

Stabilizing and Improving Training of Generative Adversarial Networks Through Identity Blocks and Modified Loss Function

MOHAMED FATHALLAH[1], MOHAMED SAKR[2], AND SHERIF ELETRIBY[2]
[1]Department of Computer Science, Faculty of Computers and Information, Kafrelsheikh University, Kafrelsheikh 33516, Egypt
[2]Department of Computer Science, Faculty of Computers and Information, Menoufia University, Shebin El-Kom, Menoufia 32511, Egypt
Corresponding author: Mohamed Fathallah (mohamedfathallah777@gmail.com)

ABSTRACT Generative adversarial networks (GANs) are a powerful tool for synthesizing realistic images, but they can be difficult to train and are prone to instability and mode collapse. This paper proposes a new model called Identity Generative Adversarial Network (IGAN) that addresses these issues. This model is based on three modifications to the baseline deep convolutional generative adversarial network (DCGAN). The first change is to add a non-linear identity block to the architecture. This will make it easier for the model to fit complex data types and cut down on the time it takes to train. The second change is to smooth out the standard GAN loss function by using a modified loss function and label smoothing. The third and final change is to use minibatch training to let the model use other examples from the same minibatch as side information to improve the quality and variety of generated images. These changes help to stabilize the training process and improve the model's performance. The performance of the GAN models is compared using the inception score (IS) and the Fréchet inception distance (FID), which are widely used metrics for evaluating the quality and diversity of generated images. The effectiveness of our approach was tested by comparing an IGAN model with other GAN models on the CelebA and stacked MNIST datasets. Results show that IGAN outperforms all the other models, achieving an IS of 13.95 and an FID of 43.71 after traning for 200 epochs. In addition to demonstrating the improvement in the performance of the IGAN, the instabilities, diversity, and fidelity of the models were investigated. The results showed that the IGAN was able to converge to a distribution of the real data more quickly. Furthermore, the experiments revealed that IGAN is capable of producing more stable and high-quality images. This suggests that IGAN is a promising approach for improving the training and performance of GANs and may have a range of applications in image synthesis and other areas.

INDEX TERMS Generative adversarial network, deep learning, mode collapse, label smoothing, identity block.

I. INTRODUCTION

Generative adversarial networks, which originated by Goodfellow [1], are at the vanguard of efforts to generate high-fidelity and diversified images. In recent years, models learned directly from data, have significantly advanced the state of generative image modeling, including biomedical imaging [2], [3] and robotics [4], [5]. The GAN network involves generator (G) and discriminator (D) networks, whose purposes differ. The discriminator is also known as the critic (C). The first network G maps a random noise vector to a data distribution to make new (fake) data that looks like the original data it was trained on. The second network, D, is just a classifier network that discriminates between real and fake data. GANs involve finding a Nash equilibrium for the following two-player min-max formula:

$$\min_G \max_D \mathbb{E}_{x\sim q_{\text{data}}(x)}[\log D(x)] + \mathbb{E}_{z\sim p(z)}[\log(1 - D(G(z)))] \quad (1)$$

The associate editor coordinating the review of this manuscript and approving it for publication was Prakasam Periasamy.

where $z \in R^{d_z}$ is a latent variable sampled from normal distribution $p(z) \sim \mathcal{N}(0, I)$. The zero indicates that the mean of the normal distribution is zero, and I represents the identity matrix, indicating that the variance is one in all dimensions. The latent variable z usually has more than one random value to allow for more combinations of what z could be. Therefore, z has an important role in making sure that images generated from the same class don't all look the same. There is no special number that determines what works, but 100 is standard, and this paper uses the standard z vector. In most state-of-the-art GAN models whose objective is to synthesize images [6], [7], [8], [9], the D and the G are convolutional neural networks based on the DCGAN [10] and use the formula (1) as their loss function.

The DCGAN training is dynamic and sensitive to almost all setup factors, including optimization, hyperparameters, and the architecture you choose for the model. Depending on the details of the application, the interaction between the generator and discriminator may be seen as either a cooperative or a competitive game. For example, where D and G are collaborative to help improve performance, it is in data augmentation [11]. In this scenario, the generator's objective is to sample from a given data distribution by extracting important characteristics that can then be utilized to generate new data samples. The discriminator's objective is to make the generator match the real data distribution by providing useful feedback. In the context of healthcare applications, the GANs are employed to develop novel medications [12], [13]. In this scenario, the generator's objective is to come up with the ingredients for new medicines, either to improve existing treatments or to treat diseases that can't be cured. The discriminator's objective is to assist the generator in designing more effective drugs by analyzing their efficacy. Nevertheless, most of those architectures suffered in some way from model collapse while training. The IGAN architecture can help most of those state-of-the-art GANs achieve their objectives without mode collapse. Moreover, reduce the time needed for training.

The mode collapse problem has been observed in GAN training [14], [15], [16], [17]. The mode collapse problem causes the generator to stick to some distributions (modes) of the real data. These modes are the samples of data that the discriminator keeps recognizing as real distributions. The collapsed discriminator in this case sends back to the generator completely pointless weights. Because of these weights, the generator will keep generating the same data distributions (modes). As a result, only a small portion of the real data is sampled by the generator, and the GAN collapses to those distributions. This collapse results in the generator producing the same or similar images as shown in Fig.8, Fig.9, and Fig.5b.

The mode collapse is frequently sacrificed for more realistic individual samples [17], [18]. This trade-off of mode collapse for high-quality, realistic samples can lead to a biased model that produces a racial or gender-biased image [19].

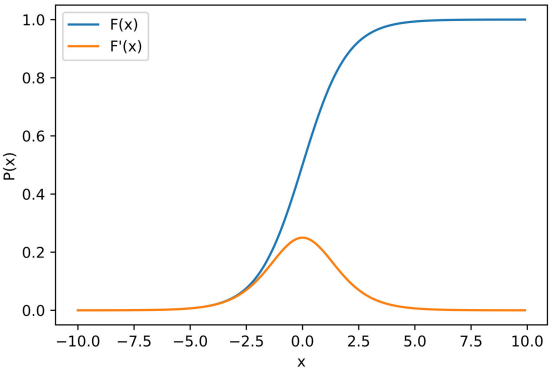


FIGURE 1. The gradient of the sigmoid function, which is used in DCGAN and other state-of-the-art GANs that are based on it.

Another issue with giving up on preventing mode collapse is that the model may be unstable and require considerable training time [20].

This paper observed the GAN training which tries to learn probability distribution, the model could simply learn probability density by defining a set of parametric densities $(P_\theta)_{\theta \in \mathbb{R}^d}$ and selecting the one that, given our data, had the highest likelihood. Thus, if there are real data examples $\{x^{(i)}\}_{i=1}^{m}$ and the $P_\theta\left(x^{(i)}\right)$ value is 1. This means the loss is nearly zero, indicating that the GAN model is converging. Nevertheless, in this case, the gradients are 0, as shown in Fig.1. For example, when $x^{(i)}$ is 7.5 the $P_\theta\left(x^{(i)}\right)$ is 1 and the gradients will be close to 0. In this case, if the discriminator continues to output 1s or 0s, the likelihood of the vanishing gradient will increase, implying that the model collapses into certain modes. Those modes are the ones that the generator produces to maximize the likelihood. As a result, decreasing the vanishing gradient can direct the model to generate more diverse images and reduce the likelihood of mode collapse. The Smoothing of $P_\theta\left(x^{(i)}\right)$ will ensure that the gradient never reaches zero. Therefore, the discriminator will keep providing the generator with useful feedback on its generated data. The smoothing of the loss function can be achieved by using smoothed labels, such as 0.9 and 0.1, instead of 0 and 1 in the discriminator targets.

$$\max_{\theta \in R^d} \frac{1}{m} \sum_{i=1}^{m} \log P_\theta\left(x^{(i)}\right) \tag{2}$$

In other words, when computing the backpropagation of GAN on the formula (1), the discriminator will push its output toward 0 and 1 as the model improves. The gradient at both ends is close to zero, as shown in Fig.1, which means the discriminator will not be able to give useful feedback. The generator will continue to generate bad samples, causing a vanishing gradient problem, and, from there, mode collapse happens. Smoothing the ends can help with the vanishing gradient and thus reduce mode collapse. The discriminator will then give the generator feedback at any point in the learning process. This will make the learning process more

stable. Moreover, this paper also shows that the addition of an identity block [21] improves G and D's ability to model complex structures and increases training speed.

The contributions of the paper are as follows:

1) The paper introduces a promising version of the GAN architecture that can be expanded to generate a high-resolution image with high accuracy compared with other available versions of GANs.
2) The modified model can improve the fidelity and diversity of GAN's generated images.
3) The model is the first to use label smoothing in the loss function and minibatch training to reduce mode collapse and stabilize the training.
4) The model reduces the number of trainable parameters and therefore reduces the training time of GAN using identity blocks.

This paper is organized as follows: Section II provides an overview of related works, Section III introduces the proposed model of the IGAN, Section IV introduces the experimental results and Section V introduces the conclusion and future work.

II. RELATED WORKS

This section provides the latest research about solving the mode collapse problem during the GAN training process and also discusses the problem of improving the quality of the images with regard to the stability of the model. This section begins with a discussion of the Nash equilibrium during GAN training, followed by a discussion of the low-dimensional manifolds of the distribution of real and fake data, the most recent research that introduces solutions to mitigate mode collapse, and finally, research that employs multiple discriminators and generator architectures. This section also discusses the limitations of the previous work.

A. NASH EQUILIBRIUM ACHIEVEMENT

The GANs models are designed to reach a Nash equilibrium [22], [23], [24], [25] As shown in equation (1), in which each player cannot minimize their cost without changing the parameters of the other player. A key aspect in the study of GANs is demonstrating that the minimum difference between the training and model distributions is achieved at this equilibrium point. Salimans et al. [26] introduced the GAN's gradient-descent-based training method. Two models are trained at the same time in this method to find a Nash equilibrium for both players in a non-cooperative game. However, each model adjusts its strategy independently from the other players in the game. Updating the gradients for both models at the same time cannot guarantee convergence.

B. LOW DIMENSIONAL MANIFOLDS

Arjovsky and Bottou [24] addressed the issue of the supports of P_r and P_g resting on low-dimensional manifolds and how it contributes significantly to the instability of GAN. Where P_r denotes data distribution over real sample x and P_g denote The generator's distribution over data x. The dimensions of many real-world datasets, as represented by P_r , appear to be artificially high. They have been shown to concentrate on a manifold with fewer dimensions. This is the core premise of Manifold Learning. When it comes to real-world images, a number of constraints must be fulfilled after the theme or the contained object is set. For instance, a dog with two ears and a tail, a skyscraper with a tall, straight body, etc. These constraints prevent images from having a high-dimensional, unconstrained shape, which we need to generate new, unseen data. Whenever the generator is asked to generate a much larger image, like 64×64. Given a low-dimension noise vector, for instance, consider a vector z with 100 random numbers. The z vector would define the distribution of colors over the 4096 pixels. The vector z in this case is low-dimensional, and it can hardly fill up the whole high-dimensional space of the image. Nevertheless, our model was able to have a high degree of diversity and fidelity with only 100 random numbers as the noise vector as shown in Section IV.

C. GAN MODE COLLAPSE

Along with improvements in the quality of the images generated by GAN [18], [27], [28], a large number of papers [14], [28], [29], [30], [31] mention and attempt to mitigate the problem of mode collapse in GANs. But the mode collapse problem was seen as a secondary goal to work on that would be taken care of automatically as GAN optimization became more stable [25], [31], [32].

In order to solve mode collapse. Unrolled GAN [33] proposed an unrolled discriminator optimization that best met the objective of the generator, thereby avoiding mode collapse. VEEGAN [34] makes use of the latent reconstruction loss in the latent space, and PacGAN [30] feeds the discriminator that determines if data is true or false with several examples of the same class. On the other hand, our method may be integrated with the most advanced GAN frameworks now available to produce significant speedups while preserving training stability.

D. MITIGATING MODE COLLAPSE USING MULTI-DISCRIMINATOR AND MULTI-GENERATOR

recently many papers discussed the possibility of using multiple adversarial networks in GANs to enhance outcomes and reduce mode collapse [14], [29], [35], [36]. Ignoring the instability, they added by using multiple loss functions. MicrobatchGAN [35] introduced a technique with more than two discriminators that may be provided as a training hyperparameter ahead. MGAN [36] uses a mixture of generators. D2GAN [29] uses two discriminators. Parallel W-CGAN [14] which is shown in Fig.2, is a parallel approach to training WGAN with conditional training. This simultaneous training decreases mode collapse and improves scalability by utilizing many generators that are concurrently trained, where each of them predicts a single data label. Those approaches have

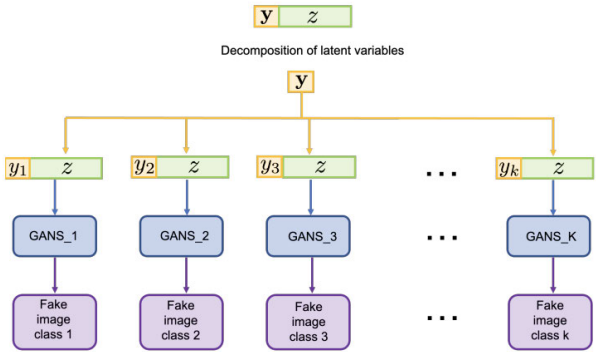


FIGURE 2. Illustration of parallel W-CGANs [14].

significant limitations as they all rely on multiple intricate factors, including initialization hyper-parameters, model design, data distribution complexity, and training dynamics. Also, it is costly and difficult to estimate, even with numerous runs of the model, as it requires huge computational power to train multi-adversarial networks. In comparison, we manage to employ a single discriminator and generator and yet achieve the same goal with less computing power and less training time.

E. SYNTHESIS HIGH QUALITY IMAGES

StyleGAN [37] and ProGAN [38] are two popular GAN architectures for generating high-quality images. StyleGAN excels at producing highly realistic and diverse images with fine-grained control over the generated output, but has a larger computational complexity. ProGAN's progressive training approach enables faster convergence and more stable training but may produce higher FID scores, indicating that the generated images are less similar to real images. The IGAN, on the other hand, reduces the computational complexity using an identity block and a fully convolutional network design; it also solves the FID score problem in ProGAN using a modified loss function to reduce mode collapse, stabilize the training, and give fine control over diversity and fidelity.The PanoHead model [39] is a 3D-aware generative approach capable of synthesizing high-quality, view-consistent images of entire human heads with diverse appearances and detailed geometry in 360°. Although the IGAN can potentially be adapted to generate 3D faces instead of 2D faces, it is currently beyond the scope of the research and will be considered in future work.

III. PROPOSED MODEL

In this section, we will present the architecture of IGAN as well as the improvements we have made to enhance and stabilize GAN's training while boosting training speed.

A. IGAN MODIFICATIONS

The phenomenon of "mode collapse" occurs when a deep learning generator collapses to a single setting of parameters that results in the repeated production of identical outputs. This is caused by the discriminator's gradients pointing in similar directions for multiple locations in the data distribution. Without coordination between these gradients, it becomes difficult for the generator to create diverse outputs. To mitigate this issue, minibatch training [26] can be implemented in the discriminator to allow it to analyze multiple data instances at once and understand the relationships between them. This approach has been shown to be effective in preventing mode collapse in deep learning. This paper estimates the proximity of every pair of samples in a single minibatch. Then, the overall summary of a single data point is computed by adding its proximity to other samples in the same batch $o(x_i) = \sum_j c(x_i, y_i)$. Finally, $o(x_i)$ is explicitly added to the model. The discriminator is still required to output a single number for each example. This number indicates the likelihood of the example originating from the training data. The single output of the discriminator with minibatch training is allowed to use other examples in the minibatch as side information.

The mode collapse occurs when the discriminator returns 1 or 0 as the classification result for the real and fake images. The gradients at both ends will be close to 0 As shown in Fig.1, and the discriminator will be unable to provide useful feedback, which will lead to a vanishing gradient problem. Therefore, this paper modifies the loss function of GAN shown in formula (1) by adding label smoothing [26], where instead of providing 1 and 0 labels for real and fake data while training the discriminator, we used softened values of 0.9 and 0.1, respectively. The loss function used in this paper is expressed in Eqs.(3),(4),(5), and (6).

$$L(G, D) = \min(L_G) + \max(L_D) \tag{3}$$

$$L_G = -\frac{1}{m}\sum_{i=1}^{m}[\log D(G(\boldsymbol{z}^i))] \tag{4}$$

$$L_D = \frac{1}{m}\sum_{i=1}^{m}[\log D(\boldsymbol{x}^i) + \log(1 - D(G(\boldsymbol{z}^i)))] \tag{5}$$

$$D(\boldsymbol{x}) = \begin{cases} 0.9 & \text{if } x \geq 0.5 \\ 0.1 & \text{if } x < 0.5 \end{cases}$$

$$D(G(z)) = \begin{cases} 0.9 & \text{if } G(z) \geq 0.5 \\ 0.1 & \text{if } G(z) < 0.5 \end{cases} \tag{6}$$

The last modification includes using learning rate decay, which we found to be very effective with the Adam optimizer and improves the speed of training when used with minibatch learning. This modification improves the model's stability (oscillation in losses) as well as the FID and Inception Score when compared to DCGAN without this modification. Also, using identity blocks [40] in our architecture shows an increasing ability of the IGAN to model much more complex distributions, even in the early stages of learning, which led to an increase in the fidelity of images.

Because GAN training is data-intensive, using datasets with fewer images per class, such as CIFAR100 or CIFAR10, makes it more difficult for the model to produce high-quality

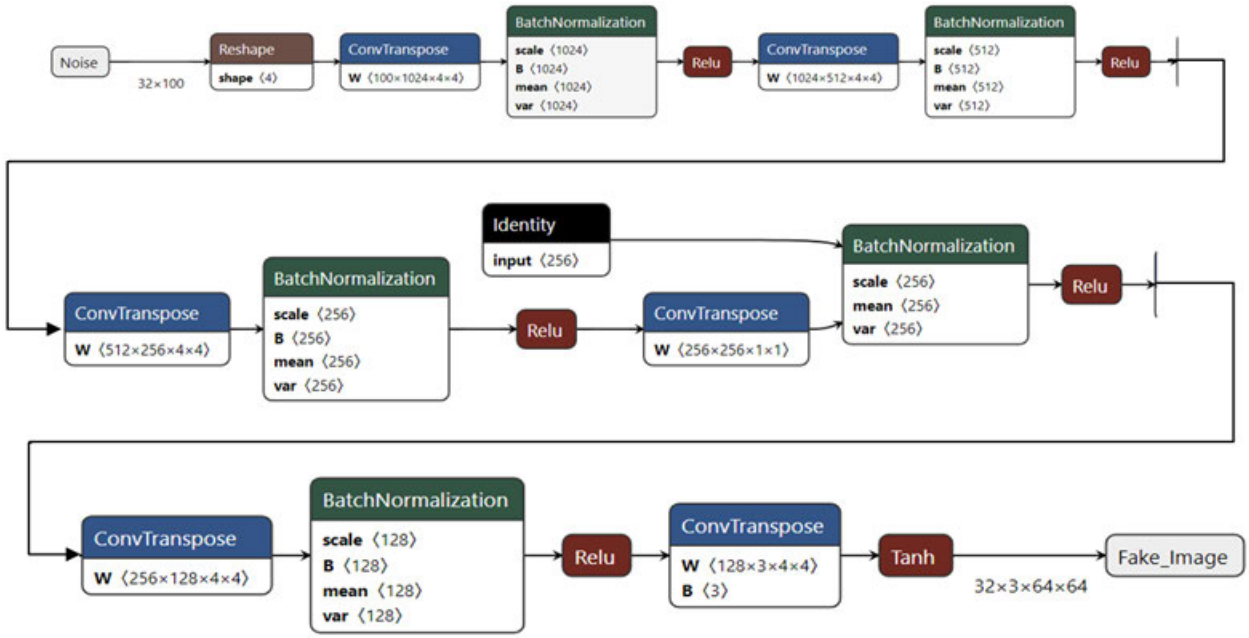


FIGURE 3. Illustration IGAN generator architecture.

images. This has an impact on the measurement of mode collapse. As a result, this paper uses the CelebA and MNIST datasets.

Despite the fact that the generator and discriminator designs can be tweaked through hyperparameter optimization, this is outside the scope of this work. The objective of this paper is to boost the performance of GANs with a fixed architecture. Thus, this paper introduces promising modifications to DCGAN [10] and compares the new architecture modifications with DCGAN, ProGAN and StyleGAN on the CelebA dataset.

Using the identity block, we were able to reduce training time by reducing the number of trainable parameters without affecting performance. The number of parameters was reduced by 200K with each identity block we added. The improved architecture can then be utilized as a foundation for any cutting-edge GAN models that need to overcome the mode collapse problem. This is possible despite limited computing power.

The Adam optimizer is used for IGAN training in both the generator and discriminator networks, with a total of 200 epochs, a learning rate initialization of 2E-3, and learning rate decay. The comparison between DCGAN and IGAN is performed on a quantitative level by measuring the IS [40] and FID [41]. The IS takes a set of images and outputs one floating-point value. The value indicates how realistic the output of a GAN is. The inception score is an automated alternative to having individuals rate image quality, which already exists thanks to HYPE [42].

B. DATASETS

This Paper uses CelebA [36] dataset. CelebA is a massive face characteristics collection with over 200K celebrity photos, each with 40 attribute annotations. This dataset contains images with a wide range of pose variants and background clutter, including 10,177 identities, 202,599 face images, and 5 landmark locations, with 40 binary attribute annotations per image. The dataset can be used to train and test computer vision systems for tasks like face attribute recognition, face recognition, face detection, landmark identification, and face synthesis. This paper uses the dataset for face synthesis. Additionally, this paper measured the mode collapses of multiple models on the Stacked MNIST dataset following [34].

C. HARDWARE AND SOFTWARE SPECIFICATION

The numerical experiments are carried out with an Intel® Core i7-10750H CPU running at 2.60 GHz on the 10th and an NVIDIA GeForce RTX 3050 Ti laptop GPU. The numerical assessments were carried out using Python 3.10.4 with the Pytorch Version 1.13.0 package. The GPU cuda version was py3.10_cuda11.3_cudann8_0.

D. EVALUATION METRIC

It is challenging to assess the effectiveness of generative models (e.g., GAN). The most effective method for assessing the visual quality of samples is to ask humans to determine the quality of the samples intuitively and reliably. This requires sufficient participants, and we did not have that. Also, we opt to use the inception score [40], a numerical assessment method, for quantitative evaluation.

$$I = exp(E_{x\sim p_g} D_{KL}(p(y|x) \,||\, p(y))). \tag{7}$$

In metric (7), x stands for one generated sample sampled from the generator, and y is the label predicted by the Inception model [43]. The idea behind this metric is that a good model should generate high-diversity and

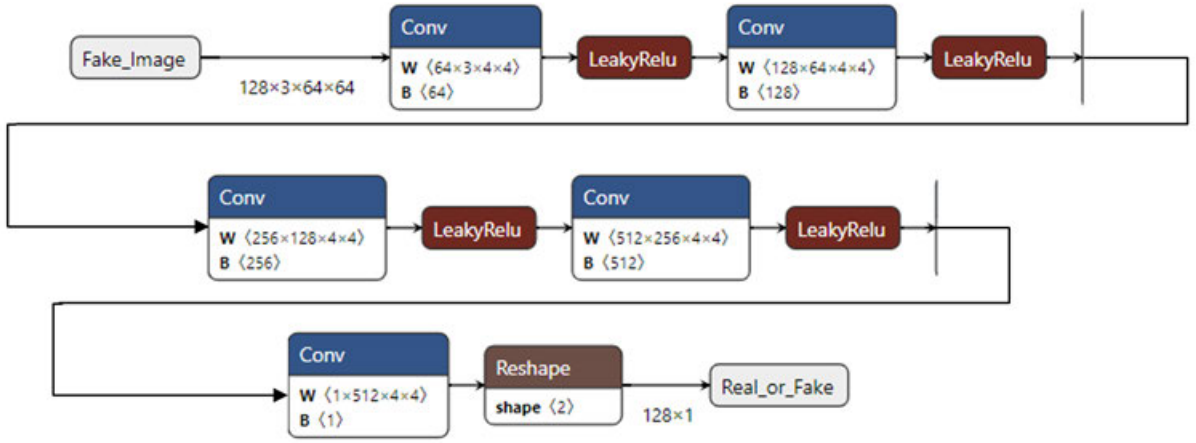


FIGURE 4. Illustration IGAN discriminator architecture.

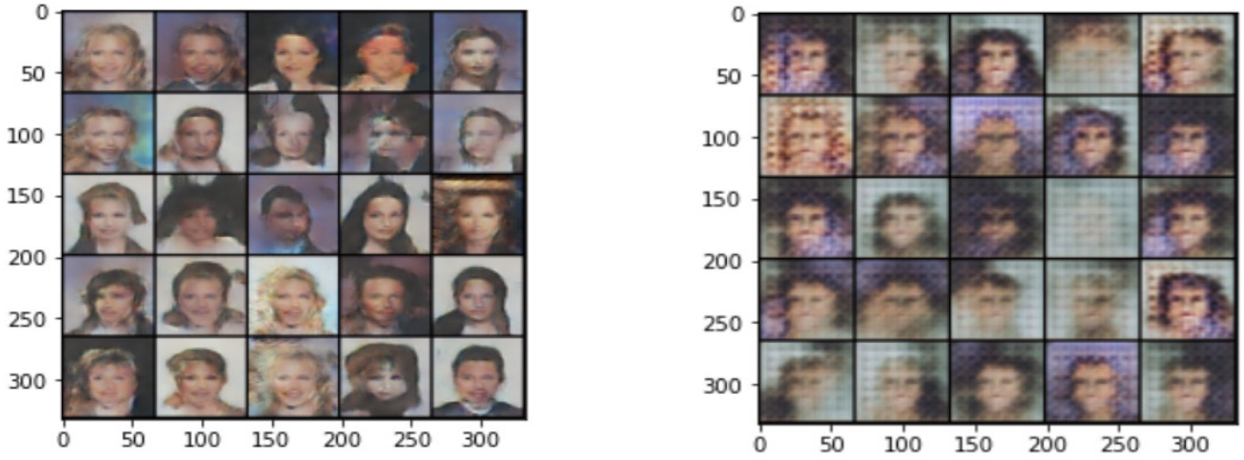


FIGURE 5. First epoch output for both models, IGAN and DCGAN. Figure (b) depicts the DCGAN, while Figure (a) depicts the IGAN.

high-fidelity images. Therefore, the divergence between the marginal distribution (real images) and the conditional distribution (generated images) should be large. This means the entropy of the conditional probability distribution is low, finding relevant objects and features, while the marginal probability distribution is high, finding a diverse set of features. We also adopt widely used Fréchet Inception Distance (FID) [44].

$$FID = ||\mu_r - \mu_g||^2 + Tr(\Sigma r + \Sigma g - 2(\Sigma r \Sigma g)^{1/2}). \quad (8)$$

Metric (8) is based on measuring the image's feature distance using a pre-trained Inception v3 network. This model is a pre-trained classifier that was trained on the ImageNet dataset. We used this model as a feature extractor. Those features are used to compare samples that are generated from the generator with the real data sample distributions, a feature-wise comparison. A lower FID shows that the generated images are closer to a realistic image distribution.

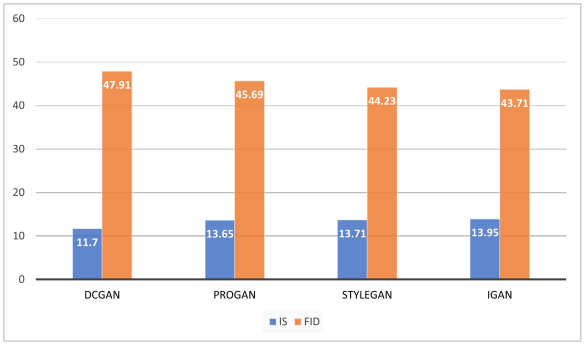


FIGURE 6. Comparison chart of generating images between GAN architectures.

IV. EXPERIMENTAL RESULTS

This section shows the results of comparing IGAN with baseline GAN architectures like DCGAN, ProGAN, and StyleGAN by comparing the fidelity and diversity of the generated images from each model. It also provides a more

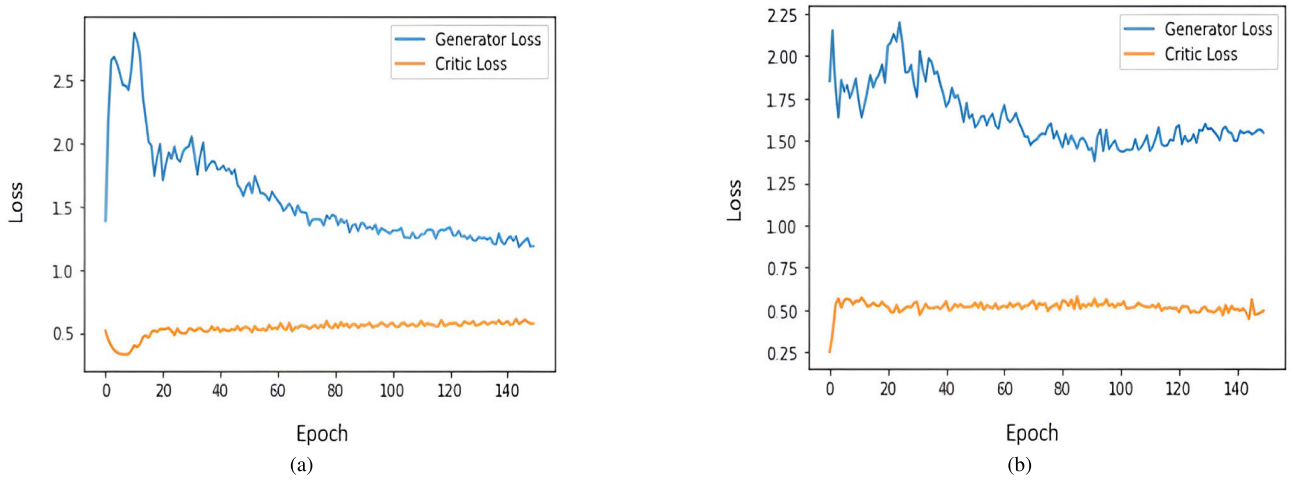


FIGURE 7. Visual results of generator and discriminator (Critic) training (a) shows IGAN losses; (b) shows DCGAN losses.

comprehensive comparison between IGAN and other recent GAN models that tend to solve the mode collapse problem.

A. RESULTS OF GENERATED IMAGES

Fig.5 provides a visual comparison of the images produced by DCGAN and IGAN in the first epoch. Fig.11 shows a sample of the generated images using IGAN after 200 epochs. IGAN images show a more accurate representation of the facial structure with no mode collapse, as shown in Fig.5a. In terms of facial features, IGAN images are significantly easier to identify than those that DCGAN, ProGAN, and StyleGAN try to sample in the early stages of learning. Table 1 provides a quantitative comparison between DCGAN,IGAN, StyleGAN, and ProGAN. The performance of the models is measured in terms of IS and FID. The IGAN model outperforms the other model regarding both indices. Given that the single epoch on the IGAN takes 39±2 minutes and the single epoch on the DCGAN, StyleGAN, and ProGAN takes 46±2,49± and 55±2 minutes, respectively, on our hardware specified in Section III-C. Fig.8, Fig.9 and Fig.10 show the comparison between the output of each model in the 10th epoch. Most of the model shows signs of mode collapse and less fidelity in the generated image compared to the IGAN output.

In Sections I and II-A we have described the Nash equilibrium between the generator and the discriminator (the critic). An exploratory experiment is performed to investigate this Nash equilibrium by observing the stability of the losses of the IGAN and the DCGAN. Fig.7a shows improved equilibrium between G and D during the training of the IGAN. On the other hand, Fig.7b shows instability in G and D losses. The D loss for real and fake images remains around 0.5, whereas for the G it is around 1.6. If the loss of G is way higher than the loss of D, this means G will not be able to deceive D. It will be hard to converge by then. In other words, if the D loss is way lower than the G loss, the D will return 0 for all generated fake images, causing the generator to get stuck with low-quality images. Also, if the G loss is way lower than the discriminator, this means the D will return 1 for all the generated images. Hence, the generator will not improve. The better GAN architecture must keep the D and G losses in the same range as shown in Fig.7a. Fig.7b shows that the DCGAN model was instable during training compared to IGAN. The modifications shows better capture of contours and distribution of the training dataset, maintaining stable training, reducing the time of training, and therefore reducing mode collapse.

TABLE 1. Inception score (IS) and Fréchet Inception Distance (FID) for the generated image from different models.

Model	Inception Score ↑	FID Score ↓
DC-GAN	11.70	47.91
ProGAN	13.65	45.69
StyleGAN	13.71	44.23
IGAN	**13.95**	**43.71**

The G should ideally produce a variety of data. If it's producing a single image or a similar set of images, there's a mode collapse. When a visually bad set of images is generated, this is a case of convergence failure. An idea of the performance of the model can be obtained by manually examining the images. Fig.10, Fig.9 and Fig.8 show the visual output of the IGAN, StyleGAN, and ProGAN at the early stage of learning, respectively. The mode collapse is obvious in both ProGAN and StyleGAN output. The IGAN output shows no sign of mode collapse at all. The output images from the IGAN show high diversity and fidelity, even at the early stages of the training. Due to the intricate nature of identifying mode collapse through visual inspection in the later stages of training, we provide the outcomes obtained during the initial phases of training.

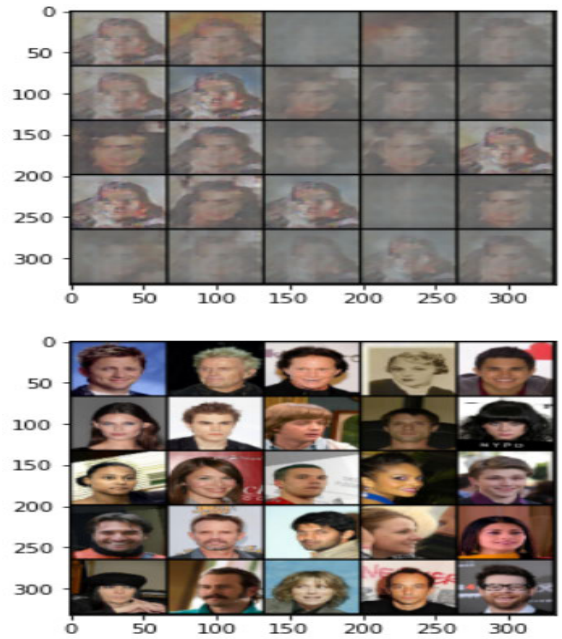


FIGURE 8. Visual output of the ProGAN in the 10th epoch.

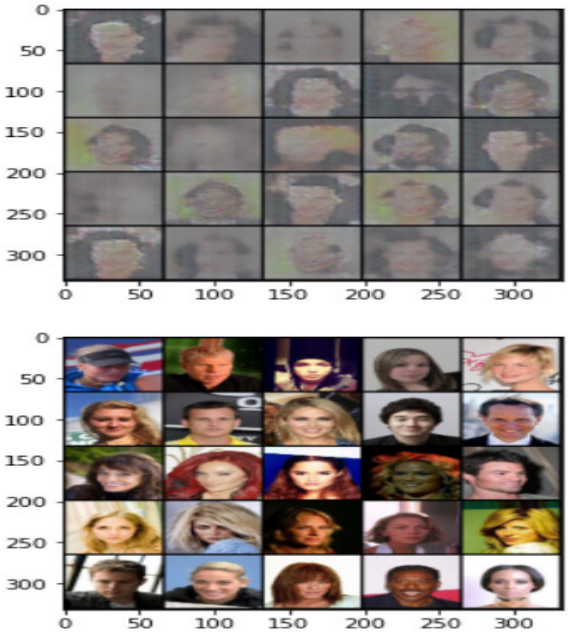


FIGURE 9. Visual output of the StyleGAN in the 10th epoch.

B. RESULTS OF MITIGATING MODE COLLAPSE

The result of comparing IGAN with other relative studies using the MNIST dataset is shown in Table 2. The MNIST dataset is considered simple, so this paper uses a set of transformations to create datasets with different levels of complexity. The transformations applied to the MNIST dataset provide flexibility in generating datasets with varying degrees of distribution complexity. Selecting $g(z)$, with progressively more complex transformations, can create synthetic datasets

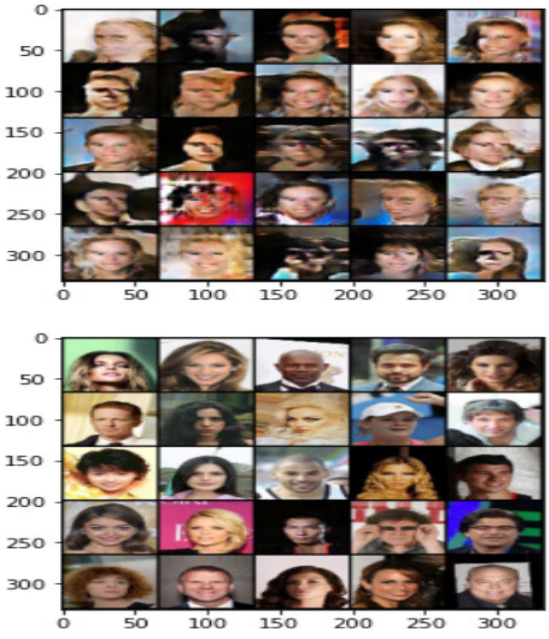


FIGURE 10. Visual output of the IGAN in the 10th epoch.

FIGURE 11. Generated images using IGANs.

TABLE 2. : T means that all of the data modes are learned by the generator, whereas F indicates the generator suffers from mode collapse. The results are shown with the SGD (left) and ADAM (right) optimizers. MNIST results using the ADAM optimizer are given as a reference. The MNIST is a relatively simple dataset, with a complexity level between Levels 1 and 5.

$g(z)$=	1		$A_{392\times2}$		z		MLP		MLP, $A_{392\times2}$		MNIST
Model Name	Level 1		Level 2		Level 3		Level 4		Level 5		-
DCGAN	F	T	F	T	F	F	F	F	F	F	T
WGAN	T	T	F	T	F	T	F	F	F	F	T
Unrolled GAN	T	T	T	T	T	T	T	T	F	F	T
D2GAN	T	T	T	T	T	T	T	T	F	F	T
GAN-NS+AMAT	T	T	T	T	T	T	T	T	F	F	T
W-CGAN	T	T	T	T	T	T	T	T	F	T	T
IGAN	**T**	**T**	**T**	**T**	**T**	**T**	**T**	**T**	**T**	**T**	T

that get more difficult and complicated. The first five datasets are classified as levels 1-5, depending on their complexity. The levels are produced using simple transforms such as

identity constant mapping (1 and z), small Multi-layer perceptrons (MLP), and well-conditioned linear transforms (A).

This benchmark investigates mode collapse on several GANs models using two optimizers, SGD and ADAM, as shown in Table 2. The results show that most of the GAN models were robust to mode collapse until level 4. The GANs model started to collapse in level 5 except for IGAN and W-CGAN [14] but W-CGAN collapsed when it used the SGD optimizer.

V. CONCLUSION

This paper presents a novel adaptive GAN training architecture that includes an identity block. In order to prevent mode collapse. Modifications to the loss function and hyperparameters by using label smoothing and mini-batch training are employed to stabilize the model. The model shows better results in stability, generated image quality, and diversity compared to other state-of-the-art GANs models. A higher IS score and a lower FID score validate our findings. In the future, we plan to use IGAN to reduce mode collapse in settings for high-resolution image generation. Additionally, we have a growing interest in extending the synthesised image dimension from 2D to 3D, specifically for the purpose of generating high-quality 3D facial images.

REFERENCES

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Commun. ACM*, vol. 63, pp. 139–144, Oct. 2020.

[2] D. Nie, R. Trullo, J. Lian, L. Wang, C. Petitjean, S. Ruan, Q. Wang, and D. Shen, "Corrections to 'medical image synthesis with deep convolutional adversarial networks' [Mar 18 2720-2730]," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 9, p. 2706, Sep. 2020.

[3] L. Lan, L. You, Z. Zhang, Z. Fan, W. Zhao, N. Zeng, Y. Chen, and X. Zhou, "Generative adversarial networks and its applications in biomedical informatics," *Frontiers Public Health*, vol. 8, p. 164, May 2020.

[4] K. Bousmalis, A. Irpan, P. Wohlhart, Y. Bai, M. Kelcey, M. Kalakrishnan, L. Downs, J. Ibarz, P. Pastor, K. Konolige, S. Levine, and V. Vanhoucke, "Using simulation and domain adaptation to improve efficiency of deep robotic grasping," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2018, pp. 4243–4250.

[5] K. Rao, C. Harris, A. Irpan, S. Levine, J. Ibarz, and M. Khansari, "RL-CycleGAN: Reinforcement learning aware simulation-to-real," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11157–11166.

[6] A. Kalpana and A. John, "Pix2Pix GAN image synthesis to detect electric vehicle license plate," pp. 439–443, 2022.

[7] X. Han, L. Zhang, K. Zhou, and X. Wang, "ProGAN: Protein solubility generative adversarial nets for data augmentation in DNN framework," *Comput. Chem. Eng.*, vol. 131, Dec. 2019, Art. no. 106533.

[8] C.-M. Huang, E. Wijanto, and H.-C. Cheng, "Applying a Pix2Pix generative adversarial network to a Fourier-domain optical coherence tomography system for artifact elimination," *IEEE Access*, vol. 9, pp. 103311–103324, 2021.

[9] S. Zheng, "Analysis of generating handwriting based on GAN model with different structures," in *Proc. IEEE Int. Conf. Comput. Sci., Electron. Inf. Eng. Intell. Control Technol. (CEI)*, Sep. 2021, pp. 654–658.

[10] J. Li, J. Jia, and D. Xu, "Unsupervised representation learning of image-based plant disease with deep convolutional generative adversarial networks," in *Proc. 37th Chin. Control Conf. (CCC)*, Jul. 2018, pp. 9159–9163.

[11] V. Sandfort, K. Yan, P. J. Pickhardt, and R. M. Summers, "Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks," *Sci. Rep.*, vol. 9, no. 1, p. 16884, Nov. 2019.

[12] A. E. Blanchard, C. Stanley, and D. Bhowmik, "Using GANs with adaptive training data to search for new molecules," *J. Cheminformatics*, vol. 13, no. 1, pp. 1–8, Dec. 2021.

[13] G. R. Padalkar, S. D. Patil, M. M. Hegadi, and N. K. Jaybhaye, "Drug discovery using generative adversarial network with reinforcement learning," in *Proc. Int. Conf. Comput. Commun. Informat. (ICCCI)*, Jan. 2021, pp. 1–3.

[14] M. L. Pasini and J. Yin, "Stable parallel training of Wasserstein conditional generative adversarial neural networks," *J. Supercomput.*, vol. 79, no. 2, pp. 1856–1876, Feb. 2023.

[15] D. Yang, S. Hong, Y. Jang, T. Zhao, and H. Lee, "Diversity-sensitive conditional generative adversarial networks," 2019, *arXiv:1901.09024*.

[16] H. Zhang, V. Sindagi, and V. M. Patel, "Image de-raining using a conditional generative adversarial network," 2017, *arXiv:1701.05957*.

[17] P. Zhou, L. Xie, X. Zhang, B. Ni, and Q. Tian, "Searching towards class-aware generators for conditional generative adversarial networks," 2020, *arXiv:2006.14208*.

[18] A. Brock, J. Donahue, and K. Simonyan, "Large scale GAN training for high fidelity natural image synthesis," 2019, *arXiv:1809.11096*.

[19] M. Hu and J. Li, "Exploring bias in GAN-based data augmentation for small samples," 2019, *arXiv:1905.08495*.

[20] S. Menon, A. Damian, S. Hu, N. Ravi, and C. Rudin, "PULSE: Self-supervised photo upsampling via latent space exploration of generative models," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 2437–2445.

[21] M. Lin, Q. Chen, and S. Yan, "Network in network," 2014, *arXiv:1312.4400*.

[22] Y.-P. Hsieh, C. Liu, and V. Cevher, "Finding mixed Nash equilibria of generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 2810–2819.

[23] Y. Jin, Y. Wang, M. Long, J. Wang, P. S. Yu, and J. Sun, "A multi-player minimax game for generative adversarial networks," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2020, pp. 1–6.

[24] M. Arjovsky and L. Bottou, "Towards principled methods for training generative adversarial networks," 2017, *arXiv:1701.04862*.

[25] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proc. 34th Int. Conf. Mach. Learn.*, 2017, pp. 214–223.

[26] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training GANs," vol. 29, 2016, pp. 2234–2242.

[27] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 12, pp. 4217–4228, Dec. 2021.

[28] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and improving the image quality of StyleGAN," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 8110–8119.

[29] C. A. Baca, W.-H. Chen, and W.-H. Chen, "Learning by competition: Dual discriminator generative adversarial networks," in *Proc. 6th IIAE Int. Conf. Intell. Syst. Image Process.*, 2018, pp. 2670–2680.

[30] Z. Lin, A. Khetan, G. Fanti, and S. Oh, "PacGAN: The power of two samples in generative adversarial networks," *IEEE J. Sel. Areas Inf. Theory*, vol. 1, no. 1, pp. 324–335, 2020.

[31] L. Mescheder, A. Geiger, and S. Nowozin, "Which training methods for GANs do actually converge?" in *Proc. 35th Int. Conf. Mach. Learn.*, vol. 8, 2018, pp. 3481–3490.

[32] D. Bau, J.-Y. Zhu, J. Wulff, W. Peebles, B. Zhou, H. Strobelt, and A. Torralba, "Seeing what a GAN cannot generate," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 4502–4511.

[33] L. Metz, J. Sohl-Dickstein, B. Poole, and D. Pfau, "Unrolled generative adversarial networks," 2017, *arXiv:1611.02163*.

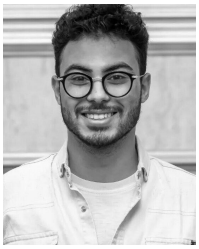
[34] A. Srivastava, L. Valkov, C. Russell, M. U. Gutmann, and C. Sutton, "VEEGAN: Reducing mode collapse in gans using implicit variational learning," vol. 30, 2017, pp. 3308–3318.

[35] G. Mordido, H. Yang, and C. Meinel, "MicrobatchGAN: Stimulating diversity with multi-adversarial discrimination," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2020, pp. 3061–3070.

[36] Q. Hoang, T. D. Nguyen, T. Le, and D. Phung, "MGAN: Training generative adversarial nets with multiple generators," in *Proc. Int. Conf. Learn. Represent.*, 2018, pp. 1–23.

[37] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4401–4410.

[38] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," 2017, *arXiv:1710.10196*.
[39] S. An, H. Xu, Y. Shi, G. Song, U. Ogras, and L. Luo, "PanoHead: Geometry-aware 3D full-head synthesis in 360°," 2023, *arXiv:2303.13071*.
[40] S. Barratt and R. Sharma, "A note on the inception score," 2018, *arXiv:1801.01973*.
[41] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs trained by a two time scale update rule converge to a local Nash equilibrium," vol. 30, 2017, pp. 6626–6637.
[42] S. Zhou, M. L. Gordon, R. Krishna, A. Narcomey, L. Fei-Fei, and M. S. Bernstein, "HYPE: A benchmark for human eye perceptual evaluation of generative models," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 3449–3461.
[43] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.
[44] A. Obukhov and M. Krasnyanskiy, "Quality assessment method for GAN based on modified metrics inception score and fréchet inception distance," in *Proc. Comput. Methods Syst. Softw.*, vol. 1294, 2020, pp. 102–114.



MOHAMED FATHALLAH received the B.S. degree in computer science from Kafrelsheikh University, Egypt, in 2019. He is currently a Teaching Assistant with the Department of Computer Science, Faculty of Computers and Information, Kafrelsheikh University. He is passionate about using technology to improve people's lives. He has been actively involved in various projects related to this goal. He is also an active member of the local tech community. He regularly attending events and workshops to stay up-to-date with the latest trends in technology. His research interests include artificial intelligence, machine learning, and generative models.

MOHAMED SAKR was born in Shebin El-Kom, Menoufia, Egypt, in 1990. He received the B.Sc., M.Sc., and Ph.D. degrees in computer science from Menoufia University, Egypt, in 2011, 2014, and 2019, respectively. He is currently a Lecturer of computer science with the Faculty of Computers and Information, Menoufia University. He has published many research papers in prestigious international conferences and reputable journals. His current research interests include big data analysis, anomaly detection, data mining, and machine learning.

SHERIF ELETRIBY was born in Tanta, Egypt, in 1970. He received the B.Sc. and M.Sc. degrees in computer science from Menoufia University, Egypt, in 1999, and the Ph.D. degree in computer science (computer vision) from the Institute for Electronics, Signal Processing and Communications (IESK), Otto-von-Guericke-University Magdeburg, Germany, in 2008. Since 2008, he has been an Assistant Professor (a Lecturer) with the Department of Computer Science, Faculty of Computers and Information, Menoufia University. Subsequently, he was the Chair of the Computer Science Department, College of Computer, Umm Al-Qura University, Saudi Arabia, from 2011 to 2015, where he has been an Associate Professor with the Computer Science Department, since 2016. Since 2021, he has been the Chair of the Computer Science Department, Menoufia University. His research interests include image processing, pattern recognition, computer vision, parallel-distributed systems, artificial intelligent, and machine-deep learning. He has served as a member of editorial board and a Program Committee Member for *International Journal of Computers and Information* (IJCI) and several international and national conferences/journals.