Dicle University
Journal of Engineering
https://dergipark.org.tr/tr/pub/dumf
duje.dicle.edu.tr

Research Article

The Effect of Latent Space Vector on Generating Animal Faces in Deep Convolutional GAN: An Analysis

İsa ATAŞ*

Dicle University, Computer Technologies Department, isa_atas@dicle.edu.tr, Orcid No: 0000-0003-4094-9598

ARTICLE INFO

Article history:

Received 21 November 2023
Received in revised form 7 March 2024
Accepted 19 March 2024
Available online 29 March 2024

Keywords:

deep learning, FID, GAN, image generation, IS, latent space

Doi: 10.24012/dumf.1393797

* Corresponding author

ABSTRACT

Researchers are showing great interest in Generative Adversarial Networks (GANs), which use deep learning techniques to mimic the content of datasets and are particularly adept at data generation. Despite their impressive performance, there is uncertainty about how GANs precisely map latent space vectors to realistic images and how the chosen dimensionality of the latent space affects the quality of the generated images. In this paper, we explored the potential of generative models in generating animal face images. For this purpose, we used the Deep Convolutional Generative Adversarial Network (DCGAN) model as a reference. To analyze the impact of selected latent space vectors, we synthesized animal face images by training data representations in the DCGAN model with the well-known AFHQ dataset from the literature. We compared the quantitative evaluation of the produced images using Fréchet Inception Distance (FID) and Inception Score (IS). As a result, we demonstrated that generative models can produce images with latent sizes significantly smaller and larger than the standard size of 100.

Introduction

With the advancement of artificial intelligence, deep learning have successfully used in various fields, including healthcare [1], hyperspectral imaging [2], and remote sensing [3], [4]. One of the main challenges in deep learning is collecting data and scaling the data for training purposes. GANs [5], which consist of two competing players: a Generator model that produces image or audio data from random noise input and a Discriminator model that classifies real and generated data, can help address this problem by generating synthetic data for data augmentation. Generative models focus on capturing the generalization ability of the model and allow sampling from the learned distribution from training, and the statistical distribution, respectively. The Generator maps from a latent space Z (a hidden space data) to X (real image data), while the Discriminator identifies whether the generated output is real or fake. GANs are very interesting due to their ability to explore the structure of the hidden space and to generate realistic data in the problem area and can help develop an intuition for what the generator model has learned. In the proposed approach, they provided a better solution than the known StyleMapping network while maintaining state-of-the-art image quality. In this study, the DCGAN [6] model, a variant of GANs known to produce realistic and high quality images, is taken as a reference. An empirical comparison of commonly used quantitative and qualitative evaluation techniques is made, considering various latent dimensions and learned data distributions. The relationship between the semantic features exhibited in images generated by GANs and the latent codes of different generator models has been studied in many works. Hwang [7] analyzed the impact of the selected latent space dimension of an automatic encoder on the final performance. Ayvaz and Baytaş [8] proposed a deep learning model to analyze variables that are determinants of the transformation from Mild Cognitive Impairment (MCI) to Alzheimer's disease. Using a generative decoder and Alzheimer's dementia-inducing dimensions, they created synthetic dementia patient images from MCI patients were generated. They obtained promising quantitative and qualitative results with their data sets. In their study, Shimizu et al. [9] investigated the connection between latent vectors and human perception or cognition through psycho-visual analysis affecting the latent vectors of human faces. In the perception study, they examined whether subjects could perceive visuals in facial images before and after changes in the latent space. In the cognition study, they analyzed whether participants could recognize a face as similar even after latent field changes. In their experiments, they showed that the distance between facial images in latent space is related to human perception and cognition of visual changes. Zhang and Schomaker [10] focused on the problems of how to ensure that the samples produced are realistic, how to take advantage of the latent space of the generator to adjust a synthesized image, and how to explain the text to image conversion layers. They used Good/Bad human face and bird images as data sets. They presented an algorithm that identifies semantically identifiable sides in the latent space of a conditional text-to-image GAN architecture by performing unbiased component analysis using the generator's pre-trained weight values. They achieved better than 98% accuracy in predicting good/bad classes for synthetic samples. Ntavelis et al. [11] proposed a separate latent space propagation for Generative Adversarial Networks (GANs). Instead of drawing latent vectors, they sampled from a finite set of elements. They developed this approach inspired by the coding of data in biological organisms.

Our main contribution in this research is as follows:
• Firstly, we conduct comparative analyses on animal face image datasets using the DCGAN model, which is a popular variant of GANs for synthetic image generation.
• Secondly, we provide a detailed explanation of the DCGAN model in the field of image synthesis and critically evaluate its performance.
• Thirdly, we qualitatively compare the impact of different dimensions of the latent space on the image quality produced by DCGAN by analyzing datasets with different latent space dimensions.
• Finally, we evaluate the potential use of optimized FID and IS metrics for assessing the quality of natural animal face images.
This paper is structured as follows: Section 2 presents the characteristics of the datasets, the GAN model used, the experimental setups and the analysis of the experimental results. Then, section 3 presents the evaluation metrics we use throughout the paper. Finally, in section 4, we discuss the results of the proposed model.

Methodology

Dataset

We used the publicly available AFHQ (Animal Faces High Quality) [12] dataset from Kaggle to evaluate the DCGAN model. This dataset includes a wide range of high-quality animal face photographs with diversity in breeds, ages, and genders to provide variation in online animal faces. It consists of over 15,000 images in total, with each of the three different animal species (cat, dog, and wildlife) contributing 5,000 images. All the images are aligned vertically and horizontally, with the eyes centered.

GAN and variant DCGAN

1. Generative Adversarial Network (GAN)

Goodfellow at al. [5] proposed GANs as a class of generative models. GANs are deep learning models that utilize two different artificial neural networks to learn by competing with each other and mimic the content of data sets. During the training phase, they engage in a mutual competition and collaboration. The generator produces detailed synthetic (fake) data that completely fools the discriminator, while the discriminator generates penalty scores to distinguish between fake and original data. Both the generator and the discriminator strive to maximize their own success while minimizing the success of their opponent. This means that both networks are optimized to achieve their own objectives: the G-Generator must create realistic examples, while the D-Discriminator must be experienced in rejecting generator samples and accepting real examples. The generator aims to maximize the likelihood of its outputs (i.e., the fake data) being recognized as ”real,” while the discriminator aims to minimize the same value [5].

The training process of a GAN model is divided into two stages: Generator stage: The generator takes z latent noise data as input and generates some fake example data, which is then passed to the discriminator. At the beginning, the generator does not know how to generate the actual data because it has never seen it before. However, the discriminator updates its parameters and computes the cost, which then backpropagates the gradients to update the generator's parameters. Discriminator stage: The discriminator is structurally a binary classifier. It receives input data (real and fake) without prior knowledge about the quality of correct or incorrect data. As output, it calculates the probability of the data generated by the generator being true or false [13]. The GAN architecture is shown in Fig. 1.

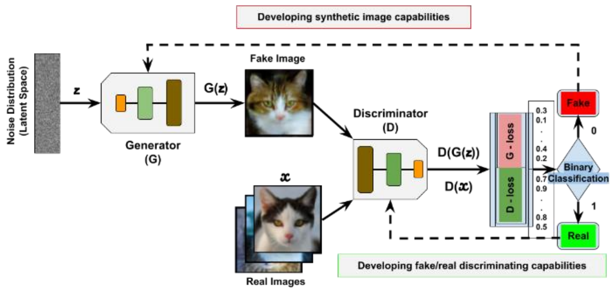

Figure 1. Architecture of the GAN.

The generator aims to map the latent space data Z to the real image data X, while the discriminator tries to determine whether the generated output is real or fake. These two neural networks, with the objective function V(D, G) given in equation 1, compete with each other during training to optimize their opposing loss functions. The objective function helps balance the competition between these two networks and assists the GAN in producing desired results.

$$\min_G \max_D V(D,G) = E_{x\sim P_{data(x)}}[logD(x)] + E_{z\sim P_{data(z)}}[\log(1-D(G(z)))] \quad (1)$$

As in the simultaneous training and learning process where two adversarial networks compete in a minmax game, the model can become noisy and unstable due to vanishing gradients, convergence error and mode collapse. For these reasons, GAN variants have been developed to overcome these problems.

2. Deep convolution GAN (DCGAN)

Fully connected layers tend to degrade the quality of generated images in GAN models. To address this issue, Radford et al. [6] proposed the DCGAN model, which introduces certain constraints to the topology of convolutional networks to enable the generation of high-quality images. They demonstrated that DCGAN contributes significantly to unsupervised learning and achieves state-of-the-art results on various image classification tasks. The DCGAN architecture replaces all pooling layers with strided convolutions in the discriminator model and transposed convolutions in the generator model. Additionally, it uses ReLU activation in all layers of the generator model, except for the last layer that uses TanH. However, it employs LeakyReLU activation function in all layers of the discriminator. The key components of the DCGAN model, which are simple and transposed convolutions used during the training phase, enable the GAN to learn excellent down sampling and up sampling operations. These upsampling operations help improve image synthesis [14]. Fig. 2, shows the important components in the generator and discriminator layers of the DCGAN model.

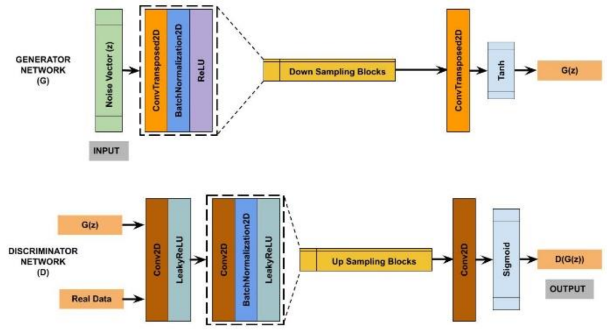

Figure 2. DCGAN model diagram.

Evaluation Techniques

Assessing the quality of GANs is important in deep learning research since it can potentially help make informed decisions such as which model to use, when to terminate training, and how to improve the model. Therefore, quantitative and qualitative evaluations are carried out to identify unintended training problems and to analyze how successfully the generator and discriminator achieve their respective goals. Quantitative evaluation involves using metrics such as IS, FID to assess the quality, diversity, and similarity of generated samples compared to real data. These metrics provide numerical measures that can be used to compare different GAN models or track the progress of a single model during training. Qualitative evaluation includes visual inspection of produced samples and human judgment. The visual realism, clarity and consistency angle of the created images are examined carefully. Factors such as the presence of artifacts, mode collapse (where the generator produces only limited variations of the same sample) and overall image quality are also considered.

The Inception Score [15] and Fr´echet Inception Distance [16] are widely preferred quantitative evaluation metrics that utilize the pre-trained InceptionV3 [17] network on the ImageNet dataset [18].

1. Inception Score (IS)

Inception Score (IS) is a metric used in machine learning to evaluate the quality of images created by generative models. Observing a produced image and making a visual assessment of the image is subjective and can vary greatly depending on the preferences and biases of the human viewer. IS is used to objectively measure the consistent performance of generated images, to determine the quality and capability of the generative model. The IS algorithm focuses on two factors; the quality of the created image and the variety of images produced. The IS algorithm equation is given in (2).

Here; *p(y|x)* denotes the conditional probability distribution, the *y* index identifies the labeling sets, the *x* index identifies the generator sampled image. The score calculated with the IS algorithm can range from zero (worst) to infinite (best) [19].

$$\uparrow \text{IS (G)} = \exp\left(E_{x\sim p_g} D_{KL}[\text{p(y|x)} \,||\, \text{p(y)}]\right) \quad (2)$$

2. Fr´echet Inception Distance (FID)

Fr´echet Initial Distance (FID) is a metric used to evaluate the performance of generative models, as the IS algorithm. It measures the similarity of the generated images to the training images. The smaller the measurement index value, the better the structural consistency. The FID equation is provided in equation (3).

$$\downarrow \text{FID (x, y)} = \left|\left|\mu_x - \mu_y\right|\right|^2 + T_r(C_x + C_y - 2(C_x C_y))^{\frac{1}{2}} \quad (3)$$

where (μ_x, C_x) and (μ_y, C_y) indices represent the mean and linear dependence measure values of the real and produced image. A low FID score is desirable [16].

While quantitative approaches in GAN evaluation provide less subjective measures, they may not always align with human perception of the quality of generated images. To complement quantitative evaluation results and assess overfitting, qualitative evaluation methods can be used to gain a better understanding of the learned data representations and generalization ability of the model.

One such method is the nearest neighbors approach, which involves classifying or predicting samples based on their similarities to known data points in the training set. This provides insights into how well the generated samples align with the characteristics of the real data and can offer a more intuitive assessment of image quality.

3. Nearest Neighbors

One way to check for overfitting in synthetic generated images is to visualize the generated images along with their nearest neighbors, which are the training images that are most similar to them. Nearest neighbors can be computed at the pixel level using similarity metrics such as Euclidean distance. However, one disadvantage of Euclidean distance is its sensitivity to small perturbations, such as shifting an image by a few pixels. As a result, GANs that produce transformed training data can pass such overfitting tests [20].

Experimental Results

Application Details

For the model training, we set the hyperparameters as follows: batch size = 64, initial weight = 0.02, image size = 64 × 64, latent space noise vector = 1, 2, 4, 8, 100, 512, 1000, and epoch numbers = 100, 200, 350. We used the open CEZERİ Library (OCL) for image preprocessing [21]. Additionally, we used a learning rate of 0.0002 and the gradient-based Adam optimizer [22] with momentum $\beta 1$ = 0.5, $\beta 2$ = 0.999, known to converge faster during training. All experiments were conducted using the PyTorch open-source deep learning framework and implemented in the Python programming language. We performed the training and testing procedures in a server-based Google Colab environment equipped with 13,342 RAM and Tesla K80 and NVIDIA T4 GPUs for accelerated processing.

Quantitative Results

Initially, we evaluated the DCGAN empirically for the widely used [23] 100-dimensional hidden field by training it on the AFHQ dataset. Fig. 3 shows the results obtained for two quantitative assessment metrics, IS and FID, respectively. In Fig. 3, it can be observed that the AFHQ Wild dataset performs better in terms of the FID metric, while the AFHQ Dog dataset achieves better results in the IS metric when using the DCGAN model in a 100-dimensional latent space.

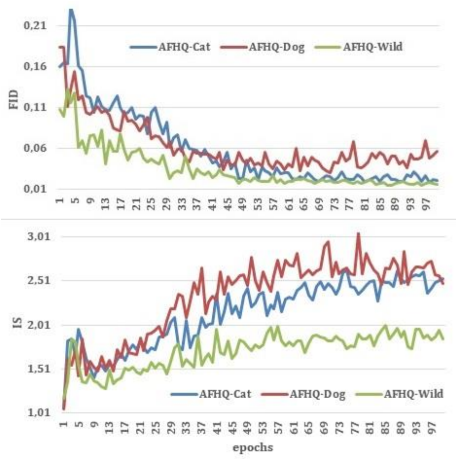

Figure 3. DCGAN evaluation scores for dim(z) = 100.

The FID metric compares the similarity of synthetic images to real images, indicating that the DCGAN model produces more realistic images with the AFHQ Wild dataset. On the other hand, the IS metric evaluates image quality and diversity, and it is observed that the DCGAN model performs well with the AFHQ Dog dataset. To analyze the effect of different latent space dimensions on the DCGAN model, the AFHQ Cat dataset was used as a reference, and the DCGAN model was trained with latent dimensions ranging from 1 to 1000. In Fig. 4, we can observe how the FID and IS values change during training for seven different latent dimensions: dim(z) ∈ 1, 2, 4, 8, 100, 512, 1000. The FID value corresponding to the smallest latent dimension, dim(z) = 1, is higher than the other six dimensions at all iterations, indicating poorer performance. Similarly, the IS value is lower than the other six dimensions at all iterations, also indicating poorer performance. The dimensionality of the latent space should typically be at a certain value that captures the complexity of the target data distribution [24].

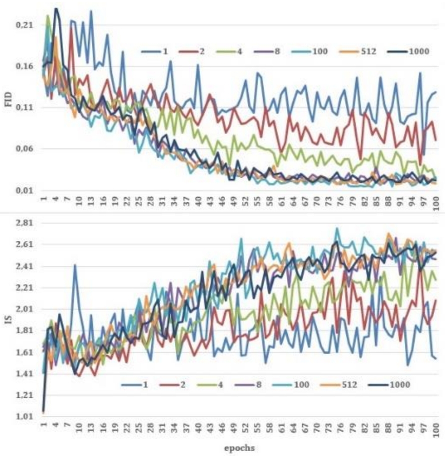

Figure 4. Quantitative results of AFHQ Cats.

In Fig. 4, it can also be observed that the AFHQ Cat dataset fails to capture the complexity at dim(z) values of 1, 2, and 4. A similar situation is observed for the AFHQ Dog and AFHQ Wild Animals datasets as well. Despite the significant difference between dimensions 8 and 1000, the FID and IS values intertwined during training, performing well as the Nash equilibrium was achieved. Higher dimensional latent spaces (100, 512, 1000) have the potential to capture more complex variations and details in the generated data, but they can also make training more challenging and computationally expensive. On the contrary, lower-dimensional latent spaces (1, 2, 4) may limit the complexity and diversity of the generated samples [24]. In the DCGAN model with the smallest latent space dimension trained on three datasets, instability and mode collapse were observed during GAN training due to insufficient latent space provided to the generator to generate features and limited ability to capture image features. As a result of mode collapse, the quantitative metric graphs of the generated synthetic images are shown in Fig. 5.

Mode collapse disrupts the stability of the DCGAN model, preventing the generation of high-quality and diverse images, resulting in FID and IS values exhibiting unstable fluctuations within a certain range.

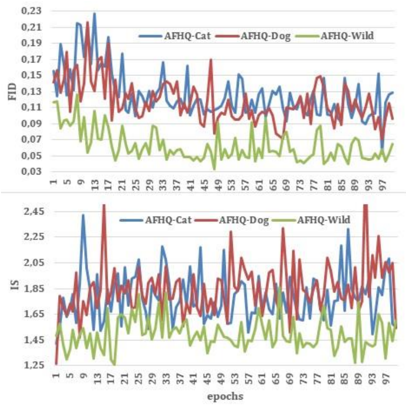

Figure 5. Evaluation scores for dim(z) = 1.

Consequently, the generated images suffer from distortions or low diversity. In our experiments, we observed evident and pronounced errors in latent space dimensions of 1, 2, and 4. Taking into consideration the best FID and IS results obtained from Table 1, and Table 2 it can be said that the 4-dimensional model performs better overall compared to the 1-dimensional and 2-dimensional models.

Comparing different latent dimensions fairly and avoiding potential misleading conclusions that could arise from either mode collapse or convergence errors can help determine the final model that may be affected by one of these issues. Furthermore, for a more qualitative analysis of the latent space's impact on the image quality, the FID value with the lowest reference point and the IS value with the highest reference point can be chosen as representatives for a specific dimension.

Table 1. TOP-5 Documented FID Results Per Latent Dimension- AFHQ Cat.

	dim(z)						
FID	1	2	4	8	100	512	1024
1.	0.148	0.083	0.033	0.025	0.021	0.024	0.043
2.	0.154	0.085	0.038	0.034	0.024	0.037	0.051
3.	0.177	0.087	0.040	0.043	0.027	0.045	0.057
4.	0.198	0.090	0.041	0.050	0.029	0.048	0.062
5.	0.203	0.096	0.047	0.063	0.032	0.053	0.066

Table 2. TOP-5 Documented IS Results Per Latent Dimension- AFHQ Cat.

	dim(z)						
IS	1	2	4	8	100	512	1024
1.	1.551	2.091	2.287	2.475	2.586	2.537	2.324
2.	1.530	2.083	2.269	2.460	2.550	2.516	2.305
3.	1.493	2.045	2.244	2.451	2.523	2.498	2.288
4.	1.475	2.021	2.217	2.429	2.498	2.473	2.265
5.	1.453	2.004	2.189	2.391	2.462	2.460	2.246

Qualitative Results

Fig. 6 shows the generated images and the corresponding loss graphs produced by the DCGAN model trained with a 100-dimensional latent space over 100 epochs using 5000 images from each of the AFHQ Cat, AFHQ Dog, and AFHQ Wild Animals datasets. The process based on the generator and discriminator loss functions oscillates within a narrow range due to the effort to reach a Nash equilibrium, which is a situation where the discriminator cannot distinguish between real and fake images in a competitive system. However, achieving this balance is challenging due to the simultaneous training of the two adversarial networks. It can be observed in the loss graphs that the discriminator's loss decreases faster towards zero compared to the generator's loss. Upon examining the generated images (left) in Fig. 6, it can be observed that as the epochs increase, the Generator manages to capture somewhat realistic images and deceive the Discriminator to some extent. When examining the loss functions (right)for all three shapes, it is evident that the Generator achieves the Nash equilibrium with the Discriminator earlier in the Cat and Dog datasets compared to the Wild dataset, demonstrating a more stable progress.

The delayed attainment of the Nash equilibrium in training the DCGAN with the Wild dataset suggests that the generator network struggles to capture the features of real data by utilizing the feature maps of the discriminative network based on images of four different animal species. Therefore, the loss function of the generator network oscillates in a wider range. In the training of DCGAN with the smallest latent space dimension, the generator lacks sufficient latent space to generate meaningful features, leading to limited ability in capturing image features.

The generated images from the DCGAN model trained on three different datasets are shown in Fig. 7. While it is evident that the model starts to capture some fine details, it remains weak in terms of image diversity as it continues to produce similar-looking images.

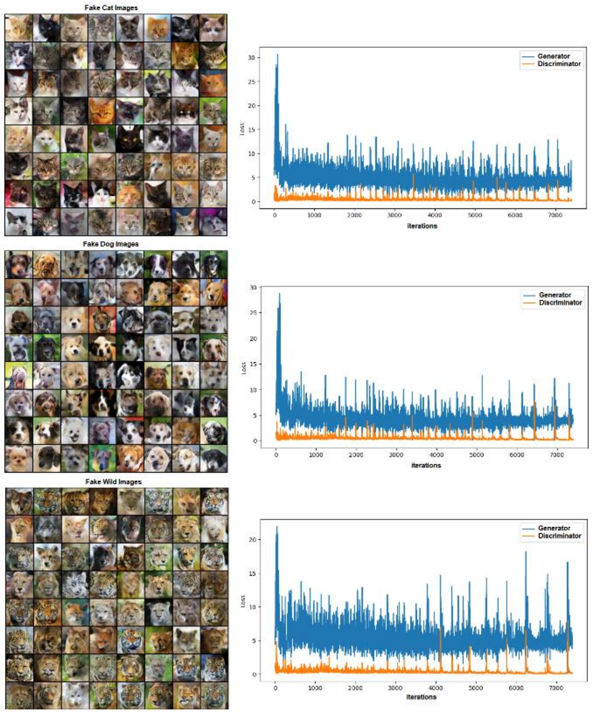

Figure 6. Images and loss plots produced in the DCGAN model trained with AFHQ Cat, AFHQ Dog and AFHQ Wild Animals datasets.

This occurs when the generative model becomes too specialized in capturing training data models but fails to generalize well to unseen data.

Figure 7. Mode collapse observed with synthetic images generated the latent space dimension z=1 of DCGAN trained with three separate datasets.

To determine whether some visually generating ”fake” images were simply memorized training data or suffered from overfitting, we compared them with the training images. Since manual comparison was not feasible, we used a distance metric, namely the Euclidean distance network, to measure the similarity between the feature representations of the generated images and the training images. As we extracted the clearest images with the cat dataset in our experiments, the cat dataset was utilized as a reference. We selected the closest neighbors (from the training data) for the generated images based on their feature representations. Fig. 8 illustrates the closest neighbors for three selected images per latent space dimension dim(z) ∈ 8, 100, 512, 1000. We can observe the similarities between the generated images and their closest neighbors. However, no model produced images that were identical to the training images, indicating that the generators were able to learn meaningful and generalizable data representations without significant overfitting.

Figure 8. Nearest neighbor of three chosen generated images per each latent dimension dim(z) ∈ 8, 100, 512, 1000.

Conclusion

The choice of noise vector size used in the latent space domain is usually made by empirical experiments to question the ultimate effect of the performance of the generative model on a given data set. This article explores the effect of the latent field dimension on DCGAN’s ability to synthesize plausible and diverse animal facial images and learn a semantically interpretable latent representation of the data. Visual inspection of the synthesized images showed that increasing or decreasing the commonly preferred common latent size (100) still enables the generative model to produce new compelling animal facial images. In our experiments, despite going as high as 512 and 1000, the complex variations and details in the generated images could not be captured. However, a significant improvement in DCGAN performance has been seen in recent increments of the 4th latent dimension. Further increase in latent size (8 - 100- 512 -1000) after initial improvements resulted in less degradation of learned mapping on DCGAN performance in quantitative graphs and a milder positive impact on qualitative results. Considering both the quantitative and qualitative results, dimensions 8, 100, and 512 seem to be the most prominent in our settings. Considering the performance of the hidden area dimension 8, FID and IS values at the end of the training, sample diversity and image quality are similar to the common hidden dimension. Nevertheless, the overall performance of the model trained with a 100-dimensional hidden space field shows remarkable results. Possible mode crashes and instability in the DCGAN training procedure are related to many factors as well as to the chosen hidden dimension value. Ultimately, the latent space dimension is a factor that affects the performance of GANs. In future studies, research on automatic hidden area size selection for GANs can be investigated. These studies may attempt to automatically determine the optimal size by analyzing the characteristics of the dataset or by other methods. This can enable the user to use GANs more easily and effectively without having to adjust the size by trial and error. As a result, this study can lead to advances in generative modeling and better understanding of the effects of hidden area size on GAN performance and designing better GAN models to optimize. Second, by combining quantitative and qualitative assessment methods, researchers can obtain a more comprehensive assessment of GAN models, taking into account both objective measures and subjective human perception. This helps assess the strengths and weaknesses of the model's performance and make informed decisions about its further development.

"There is no need to obtain permission from the ethics committee for the article prepared."

"There is no conflict of interest with any person / institution in the article prepared."

References

[1] M. Ucan, B. Kaya and M. Kaya, "Comparison of deep learning models for body cavity fluid cytology images classification." *In 2022 International Conference on Data Analytics for Business and Industry (ICDABI)* 2022, pp. 151–155.

[2] M. Atas, Y. Yardimci and A. Temizel, "Aflatoxin contaminated chili pepper detection by hyperspectral imaging and machine learning." *Sensing for Agriculture and Food Quality and Safety III.* vol. 8027. SPIE, 2011.

[3] M. Ataş et al., "Use of interactive multisensor snow and ice mapping system snow cover maps (IMS) and artificial neural networks for simulating river discharges in Eastern Turkey." *Arabian Journal of Geosciences 9*, 2016, pp. 1-17.

[4] H. Acar, M. S. Özerdem and E. Acar, "Soil moisture inversion via semiempirical and machine learning methods with full-polarization radarsat-2 and polarimetric target decomposition data: A comparative study." *IEEE Access*, vol. 8, pp. 197896–197907, 2020.

[5] I. Goodfellow et al., "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, Nov. 2020. DOI: 10.1145/3422622.

[6] A. Radford, L. Metz and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks." *arXiv preprint arXiv:1511.06434*, 2015.

[7] K. N. Mai, and M. Hwang, "Finding the best k for the dimension of the latent space in autoencoders," in *Computational Collective Intelligence: 12th International Conference*, 2020, pp. 453–464.

[8] D. S. Ayvaz, and M. I. Baytas, "Investigating conversion from mild cognitive impairment to alzheimer's disease using latent space manipulation." *arXiv preprint arXiv:2111.08794* 2021.

[9] K. Shimizu et al. "Human Latent Metrics: Perceptual and Cognitive Response Correlates to Distance in GAN Latent Space for Facial Images." *ACM Symposium on Applied Perception*, 2022, pp. 1–10.

[10] Z. Zhang and L. Schomaker, Optimizing and interpreting the latent space of the conditional text-to-image GANs. *Neural Comput & Applic*, Springer 2023, vol. 36, pp. 2549–2572. https://doi.org/10.1007/s00521-023-09185-6.

[11] E. Ntavelis et al. "StyleGenes: Discrete and efficient latent distributions for GANs." *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 2024, pp. 4077 – 4086.

[12] AFHQ, Animal faces, 2020. [Online]. Available: https://www.kaggle.com/datasets/undrewmvd/animal-faces.

[13] N. Nekamiche et al., "A deep convolution generative adversarial network for the production of images of human faces," in *Intelligent Information and Database Systems: 14th Asian 19 Conference,* ACIIDS, 2022, pp. 313–326.

[14] K. Cheng et al., "An analysis of generative adversarial networks and variants for image synthesis on MNIST dataset," *Multimedia Tools and Applications*, vol. 79, no. 19, pp. 13725–13752, Feb. 2020. DOI: 10.1007/s11042-019-08600-2.

[15] T. Salimans et al., "Improved techniques for training GANs," *NIPS'16: Proceedings of the 30th International Conference on Neural Information Processing Systems*, 2016, pp. 2234–2242.

[16] M. Heusel et al., "GANs trained by a two time-scale update rule converge to a local Nash equilibrium," *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems,* 2017, pp. 6629–6640.

[17] C. Szegedy et al., "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition,* 2016, pp. 2818–2826.

[18] J. Deng et al., "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition,* 2009, pp. 248–255.

[19] S. Barratt and R. Sharma, "A note on the inception score," *arXiv preprint arXiv:1801.01973*, 2018.

[20] O. Kramer, "*Dimensionality reduction with unsupervised nearest neighbors,*" Berlin, Springer, 2013, vol. 51, pp. 13-23.

[21] M. Ataş, "Open Cezeri Library: A novel java based matrix and computer vision framework." *Computer Applications in Engineering Education*, pp. 736-743, 2016.

[22] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[23] I. Marin, S. Gotovac and M. Russo, "Evaluation of generative adversarial network for human face image synthesis," in *International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, IEEE, 2020, pp. 1–6.

[24] I. Marin et al., "The effect of latent space dimension on the quality of synthesized human face images," *Journal of Communications Software and Systems*, vol. 17, no. 2, pp. 124–133, May 2021. DOI: 10.24138/jcomss-2021-0035.