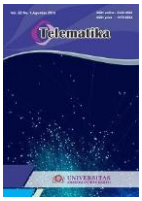


Available online at :
http://ejournal.amikompurwokerto.ac.id/index.php/telematika/

Telematika

Accreditated SINTA "2" Kemdiktisaintek, No.10/C/C3/DT.05.00/2025

Enhancing Image Generation with GANs: The Role of Mutual Information in Optimizing Generative Models

Sitaresmi Wahyu Handani[1,*], Pintusorn Suttiponpisarn[2], Gwan-yen Lin[3], Ruey-Feng Chang[4]

[1] Department of Informatics, Universitas Amikom Purwokerto, Central Java, Indonesia
[1, 2, 3,4]Department of Computer Science and Information Engineering, College of Electrical Engineering and Computer Science, National Taiwan University, Taipei, Taiwan

ARTICLE INFO

History of the article:
Received January 15, 2026
Revised February 25, 2026
Accepted April 2, 2026

Keywords:
Generative Adversarial Networks (GANs),
Mutual Information, InfoGAN,
Image Generation,
Latent Representation

Correspondence:

E-mail:
d09922018@csie.ntu.edu.tw

ABSTRACT

Generative Adversarial Networks (GANs) have become a prominent approach for image generation; however, they often suffer from training instability, mode collapse, and limited controllability of generated outputs. This study investigates the role of mutual information in improving generative modeling through a comparative analysis of Vanilla GAN, Conditional GAN (CGAN), and InfoGAN. Experiments were conducted using two image datasets with different levels of complexity, namely MNIST and Anime Face, under comparable training configurations. The evaluation focused on training behavior, convergence characteristics, generated image quality, and latent representation learning. The results revealed notable differences among the evaluated models. Vanilla GAN exhibited unstable convergence behavior at higher training epochs, while CGAN provided conditional control over generated outputs but did not fully mitigate training instability. In contrast, InfoGAN maintained more balanced generator and discriminator loss dynamics and produced visually consistent outputs across both datasets. Furthermore, latent code manipulation experiments showed that InfoGAN learned more structured and interpretable latent representations, enabling controllable feature variation in generated images. These findings indicate that incorporating mutual information improves representation learning, controllability, and training stability in GAN-based image generation. This study provides an empirical comparison of GAN, CGAN, and InfoGAN under a unified experimental framework and demonstrates that mutual information regularization contributes to improved training stability, controllable generation, and more interpretable latent representations. The findings highlight the potential of information-theoretic regularization for enhancing generative modeling performance.

1. INTRODUCTION

Generative Adversarial Networks (GANs) have emerged as one of the most influential frameworks in modern artificial intelligence, particularly for image generation tasks. By employing an adversarial learning mechanism between a generator and a discriminator, GANs are capable of producing highly realistic synthetic data that closely resemble real-world distributions (Goodfellow et al., 2014). This capability has enabled numerous applications, including image synthesis, style transfer, medical image analysis, and data augmentation (Golfe et al., 2023; Handani & Chang, 2025; Kuntalp & Düzyel, 2024; Su et al., 2024). Recent surveys further highlight the continued relevance of GANs in generative modeling despite the emergence of newer architectures (Chakraborty et al., 2024; Nayak et al., 2024). However, GANs continue to face several fundamental challenges, including training instability, mode collapse, and limited controllability over generated outputs. These problems often lead to unstable convergence and reduced

diversity in generated samples, particularly when training on complex datasets (Efatinasab et al., 2025; Luo & Yang, 2024).

To address these limitations, recent studies have explored the integration of information-theoretic concepts into generative modeling. One promising approach involves the use of Mutual Information (MI), which measures the dependency between variables and can be used to strengthen the relationship between latent variables and generated outputs (Lin et al., 2022; Najari et al., 2023; Zheng et al., 2026). By maximizing mutual information, generative models are encouraged to preserve meaningful latent features during image synthesis, resulting in more structured, interpretable, and controllable representations (Y. Li et al., 2022; Zhai et al., 2024)

Before the introduction of mutual information-based regularization, Conditional Generative Adversarial Networks (CGANs) were proposed to improve output controllability by incorporating auxiliary information such as class labels into both the generator and discriminator (Isola et al., 2017). While CGAN enables class-conditioned image generation, its effectiveness depends on the availability of labeled data and does not explicitly encourage disentangled latent representations. To overcome these limitations, InfoGAN extends the GAN framework by maximizing the mutual information between a subset of latent variables and generated outputs (Chen et al., 2016). This mechanism enables the model to learn interpretable and disentangled latent representations without requiring explicit supervision, allowing meaningful control over generated attributes. Several recent studies have further explored extensions of InfoGAN and representation learning for controllable image synthesis (Deng et al., 2023; Lv et al., 2022; Zhou et al., 2025).

Recent advances in generative modeling have introduced architectures beyond conventional GANs, including StyleGAN, StyleGAN2, and diffusion-based models, which have achieved remarkable performance in image synthesis tasks (Bermano et al., 2022; Cao et al., 2024; Croitoru et al., 2023; Trinh & Hamagami, 2024). Nevertheless, GAN-based approaches remain relevant due to their computational efficiency, controllable latent representations, and suitability for representation learning tasks. Furthermore, understanding the role of mutual information within GAN frameworks remains important for developing interpretable and controllable generative systems.

Although several GAN variants have been proposed, an important research gap remains. Most previous studies focus on proposing novel architectures or improving generation quality, whereas systematic investigations of how mutual information influences training stability, image quality, latent representation learning, and controllability across different GAN frameworks remain limited (Park & Shin, 2025; Yang & Lerch, 2018). In addition, many studies evaluate their methods using only a single dataset, making it difficult to assess whether the observed advantages generalize across different levels of data complexity.

To address this gap, this study conducts a comparative evaluation of Vanilla GAN, CGAN, and InfoGAN using two publicly available datasets with different characteristics: MNIST and Anime Face. MNIST represents a relatively simple and structured image generation problem consisting of grayscale handwritten digits (Lecun et al., 1998), whereas Anime Face contains more complex RGB images with higher variations in texture, color, and facial attributes (B. Li et al., 2022; Rao et al., 2025). The use of these

datasets enables the evaluation of model behavior under different levels of visual complexity and data diversity.

The main contribution of this study is an empirical investigation of the role of mutual information in GAN-based image generation under a consistent experimental framework. Specifically, this study compares Vanilla GAN, CGAN, and InfoGAN in terms of training behavior, convergence characteristics, generated image quality, and latent representation learning. Through this analysis, the study provides insights into how mutual information influences generative performance and demonstrates its effectiveness in improving controllability, interpretability, and training stability in image generation tasks.

2. RESEARCH METHODS

This study conducts a comparative analysis of Generative Adversarial Networks (GANs), Conditional GAN (CGAN), and InfoGAN to investigate the role of mutual information in improving image generation performance. The evaluation focuses on training behavior, image quality, latent representation learning, and model interpretability across different generative architectures.

2.1. Model Architectures

Three generative models were implemented and evaluated in this study.

2.1.1. Vanilla GAN (Baseline)

The Vanilla GAN (Goodfellow et al., 2014) represents the foundational adversarial framework consisting of a generator (G) and a discriminator (D). The model is trained through a minimax game in which (G) learns to generate realistic samples from a noise prior $Pz(z)$, while (D) learns to distinguish between real and generated samples. The objective function is defined as:

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim Pdata(x)}[\log D(x)] + \mathbb{E}_{z\sim Pz(z)}[\log(1 - D(G(z)))] \quad (1)$$

Despite its ability to generate realistic samples, the Vanilla GAN lacks control over the output attributes due to its highly entangled latent space, where individual dimensions of (z) lack explicit semantic meaning. This limitation serves as the primary motivation for more structured generative frameworks, such as CGAN and InfoGAN.

2.1.2. Conditional GAN (CGAN)

Conditional GAN (Mirza & Osindero, 2014; Zhang et al., 2020) extends the GAN framework by directing the data generation process through the incorporation of auxiliary information (y). Unlike the original GAN, which generates samples from a purely stochastic noise distribution (z), CGAN incorporates (y) (such as class labels, modality tags, or even data from other modalities) into both the generator and discriminator. This conditioning transforms the generative model into a conditional probabilistic model, where the generator (G) learns to map the joint distribution of ((z,y)) to the data space.

The objective function of a CGAN is expressed as a two-player minimax game with conditional probability:

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim Pdata(x)}[\log D(x|y)] + \mathbb{E}_{z\sim Pz(z)}[\log(1 - D(G(z|y)))] \quad (2)$$

In this architecture, the discriminator (D) is tasked not only with distinguishing real data from generated data but also with ensuring that the generated output ($G(z|y)$) aligns correctly with the provided label (y). While mutual information is not explicitly optimized as in InfoGAN, the conditioning mechanism implicitly encourages a dependency between the input labels and the resulting outputs. This structural constraint allows for greater control over the category or characteristics of the generated samples, making CGAN particularly effective for tasks such as image-to-image translation and category-specific synthesis.

2.1.3. InfoGAN

InfoGAN (Chen et al., 2016) addresses the limitation of standard GANs, where the generator often treats the latent noise space as a "black box," resulting in entangled representations where individual dimensions have no clear semantic meaning. To overcome this, InfoGAN decomposes the input into a traditional incompressible noise vector (z) and a latent control code (c), which is designed to capture meaningful structural variations in the data.

The core innovation of InfoGAN is the introduction of a Mutual Information regularization term, (I(c; G(z,c))), which encourages the information contained in the latent code (c) to be preserved in the generated output. The objective is formulated as a minimax game with an additional information-theoretic constraint:

$$\min_{G,Q} \max_D V_{InfoGAN}(D,G,Q) = V(D,G) - \lambda I\big(c; G(z,c)\big) \quad (3)$$

In practice, because the posterior (P(c|x)) is difficult to optimize directly, InfoGAN utilizes an auxiliary distribution (Q(c|x)) to approximate it. This leads to the use of a variational lower bound for mutual information, expressed as an Information Loss ((L_1)):

$$L_1(G,Q) = E_{c\sim P(c),x\sim G(z,c)}[\log Q(c|x)] + H(c) \quad (4)$$

By maximizing this lower bound, the model is encouraged to reduce the uncertainty of the latent code (c) given a generated image (G(z,c)). Consequently, the network learns disentangled representations, where specific dimensions of (c) may correspond to human-interpretable attributes such as stroke thickness or orientation in handwritten digits, or facial expressions and lighting conditions in portrait images. This mechanism provides a framework for learning more interpretable latent representations and investigating controllable image generation compared to the vanilla GAN framework.

2.2. Datasets

Two publicly available datasets with different levels of visual complexity were employed.

2.2.1 MNIST Dataset

The MNIST dataset consists of grayscale handwritten digit images representing ten classes (0–9) (Lecun et al., 1998). Due to its relatively simple structure, MNIST is widely used as a benchmark dataset for evaluating generative models.

2.2.2 Anime Face Dataset

The Anime Face dataset contains RGB facial images characterized by higher visual complexity, including variations in facial appearance, hairstyle, color distribution, and texture. Compared to MNIST, this dataset provides a more challenging environment for image generation tasks.

These datasets were selected to represent two distinct levels of data complexity, allowing the evaluation of mutual information effects under both simple and visually complex image-generation scenarios.

2.3. Experimental Setup

All models were implemented using publicly available source codes with minor modifications to ensure consistency across experimental settings. Experiments were conducted using Google Colab with NVIDIA Tesla T4 GPU acceleration. To ensure a fair comparison, all models were trained using identical hyperparameter settings whenever applicable. The image resolution was fixed at 64 × 64 pixels, the batch size was set to 64, and the latent dimension was fixed at 62. The Adam optimizer was employed with a learning rate of 0.0002 and momentum parameters $\beta1 = 0.5$ and $\beta2 = 0.999$. The generator and discriminator networks utilized convolutional layers with LeakyReLU activation functions and Batch Normalization where applicable. For InfoGAN, an additional latent code dimension of 2 was incorporated to facilitate disentangled representation learning through mutual information maximization. Prior to training, MNIST grayscale images were normalized to the range [-1, 1] using standard mean and standard deviation normalization, while Anime Face RGB images were scaled to a normalized range suitable for GAN training. These preprocessing steps were applied consistently across experiments to reduce input variability and improve training stability.

Each model was trained for 20 and 50 epochs to evaluate training behavior under different optimization durations. The MNIST dataset consists of grayscale images with a single channel and ten classes, whereas the Anime Face dataset consists of RGB images with three channels and no explicit class labels. A total of six experimental configurations were initially planned, covering combinations of three models (Vanilla GAN, CGAN, and InfoGAN) and two datasets (MNIST and Anime Face). The variation in the number of batches per epoch is attributed to differences in dataset size. Given a fixed batch size of 64, the MNIST dataset produced 938 batches per epoch from 60,032 images, while the Anime Face dataset produced 994 batches per epoch from 63,616 images.

A summary of the dataset characteristics and experimental configurations used in this study is presented in Table 1.

Table 1. Summary of Dataset-Specific Configurations and Experimental Setups

Model	Dataset	Classes	Channels	Image Size	Latent Dimension	Batch Size	Batches/ Epoch	Total Images
Vanilla GAN	MNIST	10	1	64x64	62	64	938	60032
Vanilla GAN	Anime Face	1	3	64x64	62	64	994	63616
CGAN	MNIST	10	1	64x64	62	64	938	60032
InfoGAN	MNIST	10	1	64x64	62	64	938	60032
InfoGAN	Anime Face	1	3	64x64	62	64	994	63616

Although an Anime Face configuration was initially planned for CGAN, the experiment could not be completed because the Anime Face dataset does not provide categorical labels required for conditional generation. Incorporating additional labeling procedures would alter the original experimental design and reduce comparability with the other evaluated models. Therefore, CGAN results are reported only for the MNIST dataset.

3. RESULTS AND DISCUSSION

3.1. Training Performance Analysis

The training performance of the evaluated models is summarized in Table 2, which reports the average generator loss and average discriminator loss after 20 and 50 training epochs. These metrics provide insights into adversarial training stability and convergence behavior across different model architectures. CGAN experiments were conducted on the MNIST dataset, where class labels are naturally available and align with the conditional learning mechanism employed by the model. The Anime Face dataset was primarily used to evaluate the performance of Vanilla GAN and InfoGAN under unlabeled and visually complex conditions.

Table 2. Average Generator Loss and Average Discriminator Loss Across Models After 20 and 50 Training Epochs

Epoch	Dataset	Vanilla GAN		InfoGAN		CGAN
		MNIST	ANIME	MNIST	ANIME	MNIST
Epoch 20	**Avg generator loss**	4.484460	6.105884	0.613792	0.251561	20.228795
	Avg discriminator loss	0.179973	0.095238	0.126677	0.250059	0.000000
Epoch 50	**Avg generator loss**	0.000000	6.105884	0.711919	0.251561	89.003324
	Avg discriminator loss	50.000000	0.095238	0.092265	0.250059	0.000000

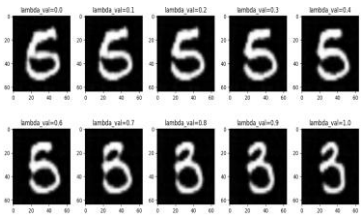
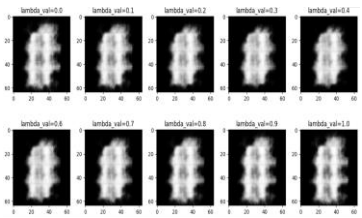
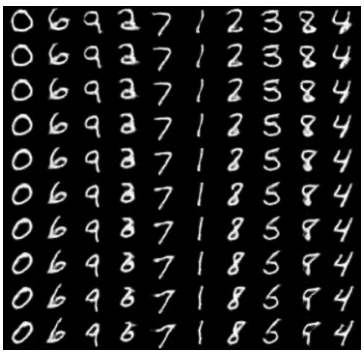
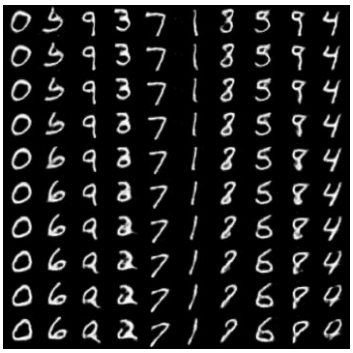
As shown in Table 2, InfoGAN demonstrated more balanced adversarial losses than the baseline models. In contrast, Vanilla GAN and CGAN exhibited larger fluctuations in generator and discriminator losses, particularly at higher training epochs. For the MNIST dataset, Vanilla GAN showed substantial degradation at 50 epochs, where generator and discriminator losses diverged considerably. Similar instability was observed in CGAN despite the incorporation of conditional information. By comparison, InfoGAN maintained relatively consistent loss values across training stages, suggesting that the mutual information regularization mechanism contributed to a more stable optimization process.

These observations are consistent with previous studies reporting that mutual information can improve latent-space structure and reduce instability in adversarial training (W. Li et al., 2021; Zhai et al., 2024). The more balanced loss behavior observed in InfoGAN indicates a healthier interaction between the generator and discriminator, resulting in more stable convergence throughout training.

3.2. Latent Space Analysis and Feature Disentanglement

To further investigate representation learning, latent-space interpolation and feature variation experiments were performed, as summarized in Table 3 for the MNIST dataset and Table 4 for the Anime Face dataset. These analyses provide qualitative insights into how latent representations are organized and utilized during image generation.

Table 3. Latent space interpolation and feature variation results on MNIST for Vanilla GAN and InfoGAN across 20 and 50 epochs

Analysis Type	20 Epochs	50 Epochs
Latent Space Interpolation (Vanilla GAN)		
Feature Variation 1 (InfoGAN)		
Feature Variation 2 (InfoGAN)		

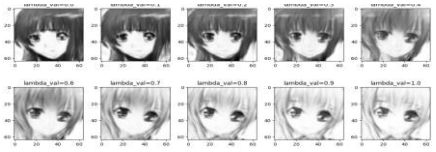
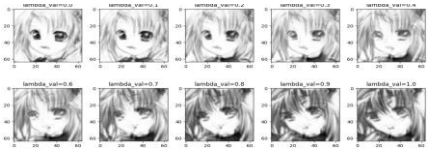
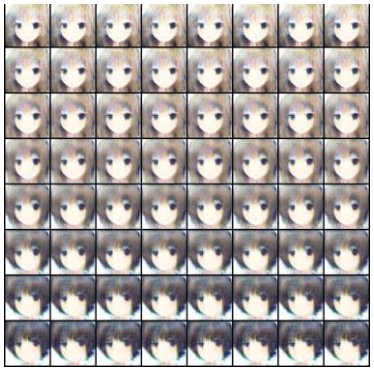
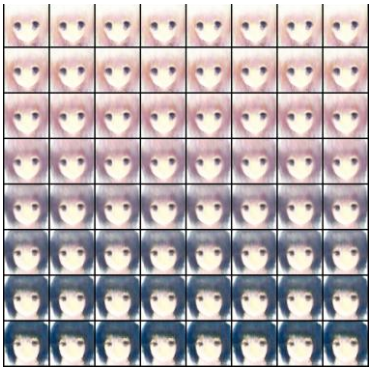
It is important to note that CGAN is not included in this latent-space analysis because its primary mechanism relies on externally provided class labels rather than disentangled latent representations. Consequently, latent interpolation and feature variation analyses are not directly comparable to those of Vanilla GAN and InfoGAN.

As illustrated in Tables 3 and 4, Vanilla GAN exhibits relatively smooth latent transitions during the early stages of training, particularly at 20 epochs. However, the interpolation quality deteriorates at higher epochs, indicating instability in the learned latent representation. This behavior is consistent with known challenges in traditional GAN training, including mode collapse and unstable convergence.

In contrast, InfoGAN enables explicit manipulation of latent variables through structured latent codes. The feature variation experiments demonstrate that specific attributes can be modified independently while

preserving the overall structure of the generated images. Such behavior indicates successful disentanglement of latent representations, where different latent dimensions capture distinct semantic characteristics.

Table 4. Latent space interpolation and feature variation results on the Anime Face dataset for Vanilla GAN and InfoGAN across 20 and 50 epochs

Analysis Type	20 Epochs	50 Epochs
Latent Space Interpolation (Vanilla GAN)		
Feature Variation 1 (InfoGAN)		
Feature Variation 2 (InfoGAN)		

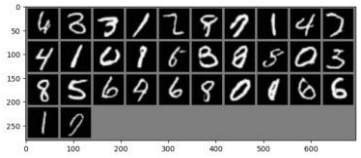
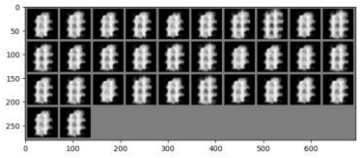
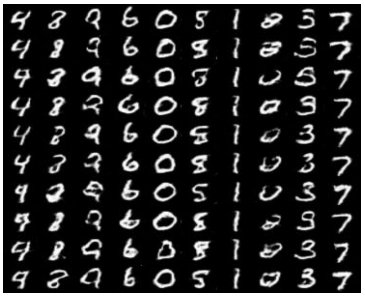
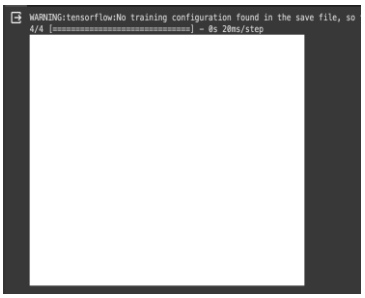
The observed differences highlight the role of mutual information in preserving meaningful relationships between latent variables and generated outputs. By maximizing this dependency, InfoGAN learns more structured and interpretable latent representations, resulting in improved controllability and feature disentanglement compared with Vanilla GAN.

Overall, the latent-space analysis demonstrates a progression from implicit and unstable latent representations in Vanilla GAN toward more structured and interpretable representations in InfoGAN through mutual information regularization.

3.3. Visual Quality of Generated Image

The generated image results are summarized in Table 5, which presents a comparative visualization across different models, datasets, and training epochs. The results are organized according to dataset complexity to facilitate visual comparison of model performance.

Table 5. Generated Image Results across Models and Datasets for Different Training Epochs

Model	20 Epochs	50 Epochs
	MNIST Dataset	
Vanilla GAN		
InfoGAN		
CGAN		
	Anime Face Dataset	
Vanilla GAN		
InfoGAN		

For the MNIST dataset, Vanilla GAN produces recognizable handwritten digits at 20 epochs, indicating that the model successfully captures the basic characteristics of the data distribution. However, image quality deteriorates at 50 epochs, suggesting increasing instability during training. A comparable

trend is observed for CGAN, where generated samples become less coherent at higher training epochs despite the incorporation of conditional information.

By comparison, InfoGAN produces visually consistent and structurally coherent images at both 20 and 50 epochs. The generated samples preserve recognizable digit structures while maintaining greater visual stability across training stages. This observation suggests that mutual information regularization contributes not only to latent-space organization but also to stable image synthesis.

For the Anime Face dataset, a similar pattern can be observed. Vanilla GAN demonstrates declining image quality as training progresses, whereas InfoGAN maintains relatively consistent facial structures and visual coherence. These results indicate that InfoGAN is more robust when handling visually complex datasets characterized by substantial variations in color, texture, and facial attributes.

Although CGAN introduces conditional control mechanisms, its applicability depends on the availability of explicit labels. Consequently, CGAN was not evaluated on the Anime Face dataset, highlighting a practical limitation of conditional generative approaches when applied to unlabeled image collections.

Overall, the visual results presented in Table 5 support the findings obtained from the training-performance and latent-space analyses. Within the experimental setting employed in this study, InfoGAN consistently exhibited more stable training behavior, improved latent-space interpretability, and visually coherent image generation compared with the baseline models.

4. DISCUSSION

Across all experiments, the results consistently reveal distinct performance differences among the evaluated models. Vanilla GAN exhibits significant training instability, particularly at higher epochs, leading to degraded image quality and less reliable latent representations. While CGAN introduces conditional information to guide image generation, the inclusion of labels alone does not fully eliminate optimization difficulties and training instability. In contrast, InfoGAN demonstrates the most balanced overall performance. The model achieves more stable adversarial learning, improved visual consistency, and more interpretable latent representations. These advantages can be attributed to the mutual information objective, which encourages the generator to preserve meaningful information from latent variables in the generated outputs. As a result, the model learns structured feature representations that facilitate controllable image generation and improved representation learning.

The findings are consistent with previous studies suggesting that mutual information serves as an effective regularization mechanism in generative modeling (Lin et al., 2022; Zhai et al., 2024). By strengthening the dependency between latent codes and generated samples, InfoGAN reduces ambiguity in latent representations and improves feature disentanglement, leading to more interpretable generation processes. It should be noted that the evaluation in this study primarily relies on training behavior, visual inspection, and latent-space analysis. Standard quantitative image-generation metrics such as Fréchet Inception Distance (FID), Inception Score (IS), and Precision–Recall were not included in the current experimental setup. Therefore, the findings should be interpreted as a comparative analysis of model behavior and representation learning rather than a comprehensive benchmark of image-generation quality.

Future work should incorporate standardized quantitative evaluation metrics and statistical validation procedures to further verify the observed performance differences.

5. CONCLUSIONS AND RECOMMENDATIONS

This study investigated the role of mutual information in enhancing the performance of Generative Adversarial Networks (GANs) for image generation tasks. Three generative models, namely Vanilla GAN, Conditional GAN (CGAN), and InfoGAN, were evaluated using the MNIST and Anime Face datasets to analyze their training behavior, image quality, and latent representation capabilities. The experimental results indicate that InfoGAN demonstrated more stable training dynamics, improved visual consistency, and more interpretable latent representations compared with the evaluated baseline models. The findings highlight the importance of incorporating mutual information into the GAN framework. By explicitly maximizing the dependency between latent codes and generated samples, InfoGAN is able to learn disentangled and meaningful feature representations, enabling greater controllability and diversity in image generation. These results support previous studies that emphasize the effectiveness of information-theoretic principles in improving generative modeling and representation learning.

Nevertheless, this study has several limitations. First, the evaluation was conducted on only two datasets with relatively limited diversity. Second, the comparison was restricted to three GAN variants and did not include more recent generative approaches such as StyleGAN, diffusion-based models, or transformer-based generative architectures. In addition, CGAN experiments on the Anime Face dataset could not be fully implemented due to the absence of class labels and compatibility constraints within the experimental framework. Future studies are recommended to evaluate mutual information-based approaches on larger and more complex datasets, incorporate quantitative image quality metrics such as FID, IS, and Precision-Recall, and compare InfoGAN with more recent state-of-the-art generative models. Furthermore, adaptive mutual information optimization strategies and hybrid architectures could be explored to further improve generation quality, stability, and controllability in complex image synthesis tasks.

REFERENCES

Bermano, A. H., Gal, R., Alaluf, Y., Mokady, R., Nitzan, Y., Tov, O., Patashnik, O., & Cohen-Or, D. (2022). State-of-the-Art in the Architecture, Methods and Applications of StyleGAN. *Computer Graphics Forum*, *41*(2), 591–611. https://doi.org/10.1111/CGF.14503

Cao, H., Tan, C., Gao, Z., Xu, Y., Chen, G., Heng, P. A., & Li, S. Z. (2024). A Survey on Generative Diffusion Models. *IEEE Transactions on Knowledge and Data Engineering*, *36*(7), 2814–2830. https://doi.org/10.1109/TKDE.2024.3361474

Chakraborty, T., Reddy K S, U., Naik, S. M., Panja, M., & Manvitha, B. (2024). Ten years of generative adversarial nets (GANs): a survey of the state-of-the-art. *Machine Learning: Science and Technology*, *5*(1), 011001. https://doi.org/10.1088/2632-2153/AD1F77

Chen, X., Duan, Y., Houthooft, R., Schulman, J., Sutskever, I., & Abbeel, P. (2016). InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets. *Advances in Neural Information Processing Systems*, *29*.

Croitoru, F. A., Hondru, V., Ionescu, R. T., & Shah, M. (2023). Diffusion Models in Vision: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(9), 10850–10869. https://doi.org/10.1109/TPAMI.2023.3261988

Deng, H., Wu, Q., Huang, H., Yang, X., & Wang, Z. (2023). InvolutionGAN: lightweight GAN with involution for unsupervised image-to-image translation. *Neural Computing and Applications 2023 35:22*, *35*(22), 16593–16605. https://doi.org/10.1007/S00521-023-08530-Z

Efatinasab, E., Brighente, A., Donadel, D., Conti, M., & Rampazzo, M. (2025). Towards robust stability prediction in smart grids: GAN-based approach under data constraints and adversarial challenges. *Internet of Things*, *33*, 101662. https://doi.org/10.1016/J.IOT.2025.101662

Golfe, A., del Amor, R., Colomer, A., Sales, M. A., Terradez, L., & Naranjo, V. (2023). ProGleason-GAN: Conditional progressive growing GAN for prostatic cancer Gleason grade patch synthesis. *Computer Methods and Programs in Biomedicine*, *240*, 107695. https://doi.org/10.1016/J.CMPB.2023.107695

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). *Generative Adversarial Networks*.

Handani, S. W., & Chang, R. F. (2025). A Comprehensive Review of Dual and Multi-Generator Architectures in GAN-Based Image-to-Image Translation. *Proceedings - 2025 9th International Conference on Information Technology, Information Systems and Electrical Engineering, ICITISEE 2025*, 439–444. https://doi.org/10.1109/ICITISEE68184.2025.11355114

Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1125-1134).

Kuntalp, M., & Düzyel, O. (2024). A new method for GAN-based data augmentation for classes with distinct clusters. *Expert Systems with Applications*, *235*, 121199. https://doi.org/10.1016/J.ESWA.2023.121199

Lecun, Y., Bottou, E., Bengio, Y., & Haffner, P. (1998). *Gradient-Based Learning Applied to Document Recognition*.

Li, B., Zhu, Y., Wang, Y., Lin, C. W., Ghanem, B., & Shen, L. (2022). AniGAN: Style-Guided Generative Adversarial Networks for Unsupervised Anime Face Generation. *IEEE Transactions on Multimedia*, *24*, 4077–4091. https://doi.org/10.1109/TMM.2021.3113786

Li, W., Liang, Z., Neuman, J., Chen, J., & Cui, X. (2021). Multi-generator GAN learning disconnected manifolds with mutual information. *Knowledge-Based Systems*, *212*. https://doi.org/10.1016/j.knosys.2020.106513

Li, Y., Chen, J., Zhang, Z., Xie, X., Xu, T., Ma, K., & Zheng, Y. (2022). Beyond mutual information: Generative adversarial network for domain adaptation using information bottleneck constraint. *IEEE Transactions on Medical Imaging*, *41*(3), 595–607. https://doi.org/10.1109/TMI.2021.3117996

Lin, C., Xiong, S., & Chen, Y. (2022). Mutual information maximizing GAN inversion for real face with identity preservation. *Journal of Visual Communication and Image Representation*, *87*, 103566. https://doi.org/10.1016/J.JVCIR.2022.103566

Luo, Y., & Yang, Z. (2024). DynGAN: Solving Mode Collapse in GANs With Dynamic Clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *46*(8), 5493–5503. https://doi.org/10.1109/TPAMI.2024.3367532

Lv, F., Liu, G., Wang, Q., Lu, X., Lei, S., Wang, S., & Ma, K. (2022). Pattern Recognition of Partial Discharge in Power Transformer Based on InfoGAN and CNN. *Journal of Electrical Engineering & Technology 2022 18:2*, *18*(2), 829–841. https://doi.org/10.1007/S42835-022-01260-7

Mirza, M., & Osindero, S. (2014). *Conditional Generative Adversarial Nets*. http://arxiv.org/abs/1411.1784

Najari, S., Salehi, M., Farahbakhsh, R., & Tyson, G. (2023). MidGAN: Mutual information in GAN-based dialogue models. *Applied Soft Computing*, *148*, 110909. https://doi.org/10.1016/J.ASOC.2023.110909

Nayak, A. A., Venugopala, P. S., & Ashwini, B. (2024). A Systematic Review on Generative Adversarial Network (GAN): Challenges and Future Directions. *Archives of Computational Methods in Engineering 2024 31:8*, *31*(8), 4739–4772. https://doi.org/10.1007/S11831-024-10119-1

Park, S., & Shin, Y. G. (2025). A Novel Generator with Auxiliary Branch for Improving GAN Performance. *IEEE Transactions on Neural Networks and Learning Systems*, *36*(3), 5818–5825. https://doi.org/10.1109/TNNLS.2024.3361087

Rao, X., Min, W., Deng, Z., & Liu, M. (2025). Facial expression transformation for anime-style image based on decoder control and attention mask. *Signal Processing: Image Communication*, *138*, 117343. https://doi.org/10.1016/J.IMAGE.2025.117343

Su, Q., Hamed, H. N. A., Isa, M. A., Hao, X., & Dai, X. (2024). A GAN-Based Data Augmentation Method for Imbalanced Multi-Class Skin Lesion Classification. *IEEE Access*, *12*, 16498–16513. https://doi.org/10.1109/ACCESS.2024.3360215

Trinh, L. T., & Hamagami, T. (2024). Latent Denoising Diffusion GAN: Faster Sampling, Higher Image Quality. *IEEE Access*, *12*, 78161–78172. https://doi.org/10.1109/ACCESS.2024.3406535

Yang, L. C., & Lerch, A. (2018). On the evaluation of generative models in music. *Neural Computing and Applications 2018 32:9*, *32*(9), 4773–4784. https://doi.org/10.1007/S00521-018-3849-7

Zhai, Y. K., Long, Z. H., Pan, W. F., & Chen, C. L. P. (2024). Mutual Information Compensation for High-Fidelity Image Generation with Limited Data. *IEEE Signal Processing Letters*, *31*, 2145–2149. https://doi.org/10.1109/LSP.2024.3439131

Zhang, H., Sindagi, V., & Patel, V. M. (2020). Image De-Raining Using a Conditional Generative Adversarial Network. *IEEE Transactions on Circuits and Systems for Video Technology*, *30*(11), 3943–3956. https://doi.org/10.1109/TCSVT.2019.2920407

Zheng, Z., Fan, C., Wang, C., Wang, M., He, X., & He, X. (2026). A GAN integrating CNN and transformer with mutual information and grayscale-based loss functions for modality translation in medical image. *Biomedical Signal Processing and Control*, *113*, 109062. https://doi.org/10.1016/J.BSPC.2025.109062

Zhou, Z., Li, Y., Liu, R., Xu, X., & Yan, Z. (2025). Unsupervised and controllable synthesizing for imbalanced energy dataset based on AC-InfoGAN. *Applied Energy*, *393*, 126107. https://doi.org/10.1016/J.APENERGY.2025.126107