IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 30 December 2023, accepted 17 February 2024, date of publication 4 March 2024, date of current version 20 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3372996

RESEARCH ARTICLE

DuDGAN: Improving Class-Conditional GANs via Dual-Diffusion

TAESUN YEOM[1], CHANHOE GU[2], AND MINHYEOK LEE[2,3], (Member, IEEE)
[1]Department of Mechanical Engineering, Chung-Ang University, Seoul 06974, South Korea
[2]Department of Intelligent Semiconductor Engineering, Chung-Ang University, Seoul 06974, South Korea
[3]School of Electrical and Electronics Engineering, Chung-Ang University, Seoul 06974, South Korea

Corresponding author: Minhyeok Lee (mlee@cau.ac.kr)

This work was supported in part by the National Research Foundation of Korea (NRF) grant funded by Korea government (MSIT) (No. RS-2023-00251528), and in part by Korea Institute for Advancement of Technology (KIAT) grant funded by Korea Government (MOTIE) (P0020967, Advanced Training Program for Smart Sensor Engineers).

ABSTRACT Class-conditional image generation using generative adversarial networks (GANs) has been investigated through various techniques; however, it continues to face challenges such as mode collapse, training instability, and low-quality output in cases of datasets with high intra-class variation. Furthermore, most GANs often converge in larger iterations, resulting in poor iteration efficacy in training procedures. While Diffusion-GAN has shown potential in generating realistic samples, it has a critical limitation in generating class-conditional samples. To overcome these limitations, we propose a novel approach for class-conditional image generation using GANs called DuDGAN, which incorporates a dual diffusion-based noise injection process. DuDGAN consists of three unique networks: a discriminator, a generator, and a classifier. During the training process, Gaussian-mixture noises are injected into the two noise-aware networks, the discriminator and the classifier, in distinct ways. This noisy data helps to prevent overfitting by gradually introducing more challenging tasks, leading to improved model performance. As a result, DuDGAN outperforms state-of-the-art conditional GAN models for image generation in terms of performance. We evaluated DuDGAN using the AFHQ, Food-101, CIFAR-10, and BAAT datasets and observed superior results across metrics such as FID, KID, Precision, and Recall score compared with comparison models; FID decreases 12.9% and 5.1% on average for AFHQ and CIFAR-10, respectively, highlighting the effectiveness of the proposed approach.

INDEX TERMS Conditional image generation, deep learning, diffusion-based probabilistic models, generative adversarial networks.

I. INTRODUCTION

Generative adversarial networks (GANs) and their numerous variations have demonstrated significant success within the realm of computer vision. These networks have shown impressive performance in a wide array of tasks, such as image generation [1], [2], [3], [4], [5], image-to-image translation [6], [7], [8], [9], [10], video generation [11], [12], [13], 3D reconstruction [14], [15], [16], [17], and GAN inversion [18], [19], [20]. The field of image generation, in particular, has experienced considerable advancements in both quality and diversity, largely attributed to the development of style-based architectures [21], [22], [23], [24].

Typically, image generation using GANs can be classified into two categories: unconditional and conditional image generation. While unconditional image generation does not require any additional information, the conditional approach necessitates supplementary input, such as a specific image, text prompt, or class label. A majority of conditional GAN models aim to control the output image through auxiliary supervision during the training process. Consequently, numerous studies, including [3], [25], [26], [27], have been conducted to enhance the quality of generated images with class-conditional information. Despite the notable results achieved thus far, class-conditional image generation remains

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

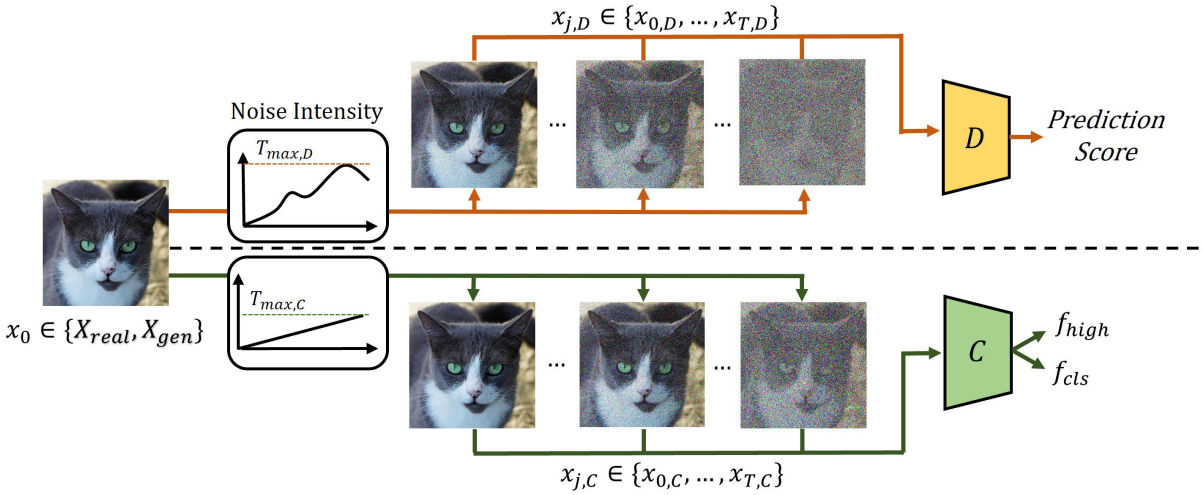


FIGURE 1. Structural overview of DuDGAN. An arbitrary input image is first sampled from the training dataset, which is then diffused by an independent noise schedule for both the discriminator and the classifier. Subsequently, the generator produces high-quality and class-specific images with the aid of the supervision of the discriminator and classifier.

more challenging due to the need for learning over smaller intra-class data distributions. Additionally, these methods are hindered by the necessity of a vast, labeled dataset and plenty of iterations to ensure stable training.

Nevertheless, the process of collecting and curating a large, class-specific dataset is both labor-intensive and time-consuming. Moreover, in this case, conditional GANs often encounter several issues during the training phase, such as mode collapse and gradient explosion problems [28], [29], [30]. Consequently, it is crucial to explore suitable techniques for training conditional GANs with limited data.

While some GAN training methods for handling the issues have been proposed [23], [31], [32], these approaches predominantly focus on training within an unconditional data regime rather than using class-labeled images. Indeed, some recent research has aimed to enhance conditional image generation with small datasets. For instance, Transitional-CGAN [28] introduced a novel training strategy that combines unconditional and conditional training to address condition-induced mode collapse. However, this method primarily concentrates on reducing supervision for conditions during the early training stage and thus may not be an effective solution for preventing collapse in later stages of training. Moreover, this approach is not efficient in terms of iteration efficacy due to the extensive scales involved in the transitional process.

In response to these challenges, we present DuDGAN (Fig. 1), a robust method for class-conditional image generation that excels in data and iteration efficient training. DuDGAN comprises three distinct networks: a generator, a discriminator, and a classifier. Drawing inspiration from previous work [32] that trains a discriminator using noise injection, the objective is for a timestep-dependent classifier to learn and output class-conditional information during training while incorporating a diffusion-based noise injection process. Concurrently, the timestep-dependent discriminator acquires prior knowledge from the classifier to discern whether images are real or fake.

The classifier's output consists of two types: high-dimensional class information for calculating contrastive loss [33] and class-dimensioned logits for classification loss. Throughout the training process, we employ an appropriate diffusion intensity for both the discriminator and the classifier, determined by each network's status. Thanks to dual-diffusion process, DuDGAN generates high-fidelity and diverse images in datasets with class imbalances and high intra-class variation.

Our key contributions are as follows:

- We investigate the impact of using an additional classifier trained with a diffusion-based noise injection process for class-conditional image generation.
- We propose a novel training approach termed *dual-diffusion*, which signifies the collaboration between the discriminator and the classifier, both of which are trained using diffusion-based noise injection.
- DuDGAN achieves fast convergence within a limited number of iterations, thereby accomplishing both high-quality generation and iteration-efficient training under limited class-wise data and an inter-class imbalanced condition.
- As a result, DuDGAN achieves superior performance in compared to state-of-the-art GAN models on the AFHQ [5], Food-101 [34], CIFAR-10 [35], and BAAT [36].

II. RELATED WORK

A. CLASS-CONDITIONAL GENERATIVE ADVERSARIAL NETWORKS

GANs [37] are deep generative models designed to produce realistic data by approximating a real data distribution $p(x)$. Two primary neural networks, the discriminator and the

generator, undergo simultaneous training to achieve their objectives. The discriminator learns to differentiate between real and fake data, while the generator strives to generate data that can deceive the discriminator. In this context, the objective function for Vanilla GAN [37] can be expressed as:

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

where $z \sim p(z)$ is a noise vector randomly sampled from a particular distribution (e.g., Gaussian distribution) and $x \sim p(x)$ is sampled from the real data distribution. Under ideal conditions, the discriminator outputs a probability of one-half for any given input.

However, unconditional GAN models are unable to generate the desired images as they train over the entire data distribution, regardless of class-wise information. To address this issue, Mirza and Osindero introduced CGAN [25], which generates conditional images by incorporating a class label into both the generator and discriminator. The basic form of the objective function for conditional GANs with discrete class-conditional information c is as follows:

$$\min_G \max_D V(G, D) = E_{x\sim p(x)}[\log(D(x, c))] + E_{z\sim p(z)}[\log(1 - D(G(z, c)))]. \quad (2)$$

Several studies have been conducted in this area. ACGAN [26] enhances the performance of conditional image generation by employing an auxiliary classifier to output class information for backpropagation. Transitional-CGAN [28] uses a linear transition function for each network during the training phase, which takes unconditional training first and then conditional training to prevent mode-collapse from limited data while class-conditioning. Rebooting-ACGAN [29] projects a vector onto a hypersphere to mitigate mode collapse caused by gradient explosion.

B. DIFFUSION-BASED PROBABILISTIC MODELS

Particularly within the domain of computer vision, the diffusion probabilistic model is considered the general form of denoising diffusion probabilistic models (DDPM) [38]. It consists of a two-way Markov chain, known as the forward and reverse processes. In the forward process, Gaussian noise is gradually injected into the data at discrete timesteps $t \in \{0, 1, \ldots, T-1, T\}$. As a result, the data becomes random noise $\mathcal{N}(0,1)$ after the final T steps. Note that the predefined variance schedule β_t and variance I in the equations below do not have any learnable parameters. The equation for the forward noising process is as follows:

$$F(x_t \mid x_{t-1}) := \mathcal{N}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I), \beta_t \in (0, 1). \quad (3)$$

In contrast, the reverse process entails denoising the data from noise to target data. This process is governed by a set of parameters θ. The reverse process can be represented as the model of latent variables:

$$R_\theta\left(x_{t-1} \mid x_t\right) := \mathcal{N}\left(x_{t-1}; \mu_\theta\left(x_t, t\right), \Sigma_\theta\left(x_t, t\right)\right). \quad (4)$$

The set of parameters θ at each denoising step can be calculated by parameterizing a specific neural network within the model. Diffusion-GAN [32] demonstrates that diffusion-based data augmentation is effective for mode-catching and provides non-leaking augmentation for the discriminator. In this paper, we explore the efficacy of the forward noise injection process in improving the quality of image generation in a class-conditional setting.

From the perspective of the diffusion-based model itself, DuDGAN has a similar direction with a few recent works, such as classifier diffusion guidance [39], which aims to pass an informative gradient of the classifier to the diffusion network for high-quality conditional image generation, but DuDGAN aims to jointly optimize components of GAN networks (not a diffusion network) and thus does not require a two-step training stage with a frozen classifier.

III. METHOD

A. NOISE INJECTION THROUGH FORWARD DIFFUSION PROCESS

The primary training objective of a class-conditional GAN is to generate high-quality and photorealistic conditional samples by training through the real data distribution $p(x)$, while predicting modes over a limited class-wise distribution $p(x|c)$, which is a subset of $p(x)$. In this process, Gaussian noise is injected into both the discriminator's and classifier's inputs using a forward diffusion chain. As mentioned in Section II, a distribution component derived from the noise injection process of an arbitrary noisy sample $x_j \in \{x_0, \ldots, x_{T_k}\}$ at specific iteration k through the forward Markov chain can be expressed as a closed-form Gaussian distribution [38], [40]:

$$F_{T_k}(x_j|x_0) = \mathcal{N}(x_j; \sqrt{\bar{\alpha}_j}x_0, (1 - \bar{\alpha}_j)\sigma^2 I), \quad (5)$$

where the distribution depends on timestep and T_k, a maximum timestep at the iteration. In the equation, $\bar{\alpha}_j := \prod_{k=1}^{j} 1 - \beta_k$, and x_0 is the real or generated image that is not perturbed by the Gaussian noise. Furthermore, by applying the reparameterization trick [38], [41], the noisy sample x_j can be summarized as the linear combination of original data and noise:

$$x_j = \sqrt{\bar{\alpha}_j}x_0 + \sqrt{1 - \bar{\alpha}_j}\sigma\epsilon, \quad j \in \mathcal{T} := \{0, 1, \ldots, T_k - 1, T_k\}. \quad (6)$$

In discrete timestep j, Gaussian noise is injected into real or generated images by the equation. As the timestep increases, more information loss occurs in the sample.

However, since DuDGAN focuses on training with class-conditional images, each class-conditioned image x_j is affected by a conditional distribution defined as F_{T_k} under maximum timestep $T_k \in \{T_0, T_1, \ldots T_{max}\}$ at iteration k, and class $c \in C := \{c_1, c_2, \ldots c_{max}\}$ [32]:

$$(x_j, c_{x_0}) \sim F_{T_k}(x_j \in c_{x_0} \mid x_0, j) = \mathcal{N}(x_j; \sqrt{\bar{\alpha}_j}x_0, (1 - \bar{\alpha}_j)\sigma^2 I). \quad (7)$$

While the distribution represents information at an arbitrary iteration, it should also contain information about the timestep variable j, which appears to be noise in images. In this context, we define a discrete mapping f, which is monotonically increasing and sum of the elements in the range of f is one. The output of mapping behaves a weight factor in the total mixture distribution:

$$\forall j \in \mathcal{T},\ f : (j, T_{max}) \longrightarrow w_j. \tag{8}$$

Extending this process to predefined total iterations, Eq. 7, can be generalized as additional summation on class set and maximum iteration during the training process; a Gaussian-mixture distribution:

$$F_{T_k}(x_j \mid x_0) := \sum_{\forall k \in K} \sum_{\forall c \in C} \{w_j \cdot F_{T_k}(x_j \in c_{x_0} \mid x_0, j)\}. \tag{9}$$

For both real or generator-produced images, we can sample from noised image in this distribution; $\forall x_0 \in \{X_{real}, X_{gen}\}$. Implementation of DuDGAN diffusion process is based on the DDPM [38], which injects Gaussian noise at the pixel-level of the image.

B. ADDITIONAL CLASSIFIER FOR CONDITIONAL IMAGE GENERATION

To achieve high-fidelity and diversity of generated images, models for class-conditional image generation must have the ability to handle extensive class-wise distribution as much as whole distribution. This necessitates an additional network that deals with class information. Inspired by previous work [10], which demonstrates the effectiveness of an independent classifier network intending to increase class-wise and class-aware capacity for GAN training, DuDGAN includes an independent classifier that receives real or generated images with Gaussian-mixture noise and outputs class information. Note that the classifier input consists only of the real or generated image and does not contain class labels. Consequently, the classifier can predict distribution beyond the bounded information configured by class-wise images for training.

This procedure prevents overfitting on the training set and enables learning broadly of the class information. Furthermore, classifier outputs comprise two-level conditional information represented as f_{high} and f_{cls}. f_{high} consists of a high-dimensional latent code that contains high-frequency class-conditional features, while f_{cls}, class logits for domain classification on the class of the input image and the class predicted by the network, are formed of a vector whose dimension is the same as class labels. For accurate training on the classifier, the classifier is trained only with real images, not with images generated by the generator. With an arbitrary noisy image x_j, classifier outputs can be written as:

$$(f_{\text{high}}, f_{cls}) = C\left(x_j\right). \tag{10}$$

Moreover, the multi-dimensional features are able to feed auxiliary information to the generator and the classifier in the training process. which is not considered in typical GANs. This serves to optimize the balance of training, and as a result, fast convergence can be achieved within a limited number of iterations.

C. DUAL-DIFFUSION PROCESS

DuDGAN targets training two neural networks simultaneously, the discriminator and the classifier, through an independent diffusion-based noise injection process. For both networks, DuDGAN is based on [32], which employs the procedure of gradually presenting the discriminator with a more challenging task by first showing clear samples and then introducing noisy samples.

The discriminator, which undertakes the bi-classification task of predicting the realness score by taking real and generated images as input during the training process, aims to self-supervise the noise intensity by leveraging a predefined hyperparameter r_d, which indicates the extent to which the discriminator is overfitted to the training set [23], [32]. Considering that the intensity of noise is determined by the process of iteration k, which is a multiple of 4, is summarized as follows:

$$P_{k,D} = P_{k-4,D} + \text{sign}\left(r_d - D_{opt}\right) * \text{const}. \tag{11}$$

Note that $P_{k,D} \in (0, 1)$ represents the maximum intensity of the noise injection process in iteration k, and r_d is determined to be 0.6 by the experiment in [23].

The classifier aims to perform classification by labels according to the input image. Similarly to Eq. 11, the classifier receives a noisy sample with an independent noise schedule. To improve classification, we predefined the noise intensity for each iteration by dividing the total number of predefined maximum iteration k_{max}. This can be interpreted as a linear increase in noise intensity from the original image to the fully noised image in proportion to the number of iterations. Furthermore, we bound the maximum diffusion intensity in training the classifier for better classification. The noise intensity for the independent classifier is written as follows:

$$P_{k,C} = P_{k-4,C} + \frac{4}{k_{\max}}, \qquad P_{k,C} \in (0, 0.3). \tag{12}$$

As in Eq. 11 and Eq. 12, the diffusion intensities of the discriminator and the classifier are updated every 4 iterations. The maximum timestep $T_k \in \{T_{k,D}, T_{k,C}\}$ is now adjusted by computed $P_k \in \{P_{k,D}, P_{k,C}\}$ for both parameterized networks $N_\theta \in \{D, C\}$:

$$T_{k,N_\theta} = T_{k-4,N_\theta} + T_{max,N_\theta} \cdot P_{k,N_\theta} \leq T_{max,N_\theta}. \tag{13}$$

Algorithm 1 presents the dual-diffusion noise intensity adjustment, a central feature of the DuDGAN.

D. OVERALL TRAINING WITH DIFFUSION

The outline of the training procedure of DuDGAN is displayed in Fig. 2. To enhance quality and prevent collapse in class-conditional image generation, we propose a new form of overall loss functions. Three different networks, the

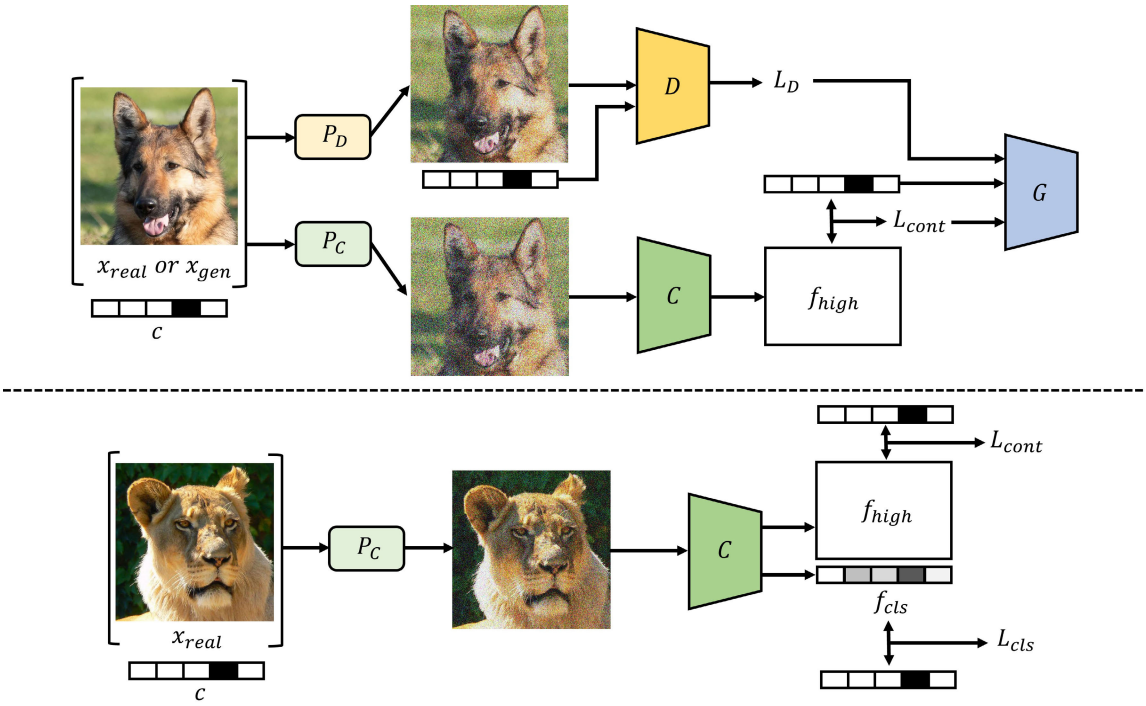


FIGURE 2. DuDGAN training procedure. The procedure has two components: (1) losses used during generator training, which include the discriminator's prediction score and a contrastive loss in a high-dimensional feature space with original class labels, and (2) independent training of the classifier, including both a contrastive loss and cross-entropy for classification loss. Input images undergo a separate Gaussian-mixture noise injection process, causing information loss, which is accounted for by a predefined parameter for high-quality generative results.

Algorithm 1 Dual-Diffusion Noise Intensity Adjustment

for All iterations; $k = 1$ to $k_{\max}$ **do** ▷ Training over total iterations
  if $k \mod 4 = 0$ **then**
    $P_{k,D} \leftarrow P_{k-4,D} + \text{sign}(r_d - D_*) \cdot \text{const}$
    $P_{k,C} \leftarrow P_{k-4,C} + \frac{4}{k_{\max}}$
    if $P_{k,D} \geq P_{\max,D}$ **then**
      $P_{k,D} \leftarrow P_{\max,D}$
    end if
    if $P_{k,C} \geq P_{\max,C}$ **then**
      $P_{k,C} \leftarrow P_{\max,C}$
    end if
  end if
  return $P_{k,D}$, $P_{k,C}$
  Calculate $T_{k,D}$ and $T_{k,C}$
  Training D and C by $T_{k,D}$ and $T_{k,C}$
end for ▷ Training ends
return Best checkpoint

generator, the discriminator, and the classifier, are jointly trained with loss functions to achieve their objectives. First, during the classifier's training with real images, L_{cont}^{real} and L_{cls}^{real} are calculated from the two-level outputs, f_{high} and f_{cls}, respectively. L_{cont}^{real} represents the supervised contrastive loss [33] derived from the high-dimensional latent space, while L_{cls}^{real} denotes the simple classification error between the predicted and given labels.

Aiming to produce photorealistic and diverse images within class-wise distribution, the generator receives additional information from the classifier for the generated images. Thus, similar to the classifier, the generator's loss function consists of the contrastive loss of generated images, which guides the generator to produce high-fidelity images, while the original loss L_G^{gen} remains.

Finally, the discriminator does not receive any informative gradient from the classifier, so the loss function remains the same as in the baseline model [32], which is the non-saturating GAN loss. Summarizing this section, the following loss functions constitute the full training objective:

$$L_C = \lambda_C \cdot L_{\text{cont}}^{\text{real}}\left(f_{\text{high}}, c_r\right) + (1-\lambda_C) \cdot L_{\text{cls}}^{\text{real}}, \tag{14}$$

$$L_G = \lambda_G \cdot L_G^{\text{gen}} + (1-\lambda_G) \cdot L_{\text{cont}}^{\text{gen}}\left(f_{\text{high}}, c_f\right), \tag{15}$$

$$L_D = L_D^{NS}, \tag{16}$$

where λ_C and λ_G are hyperparameters to modulate the training of the classifier and generator, respectively. We experimentally set the value of both parameters to 0.95.

IV. EXPERIMENTS

A. DATASET

For class-conditional image generation with GANs, the dataset for training must contain label information. In this

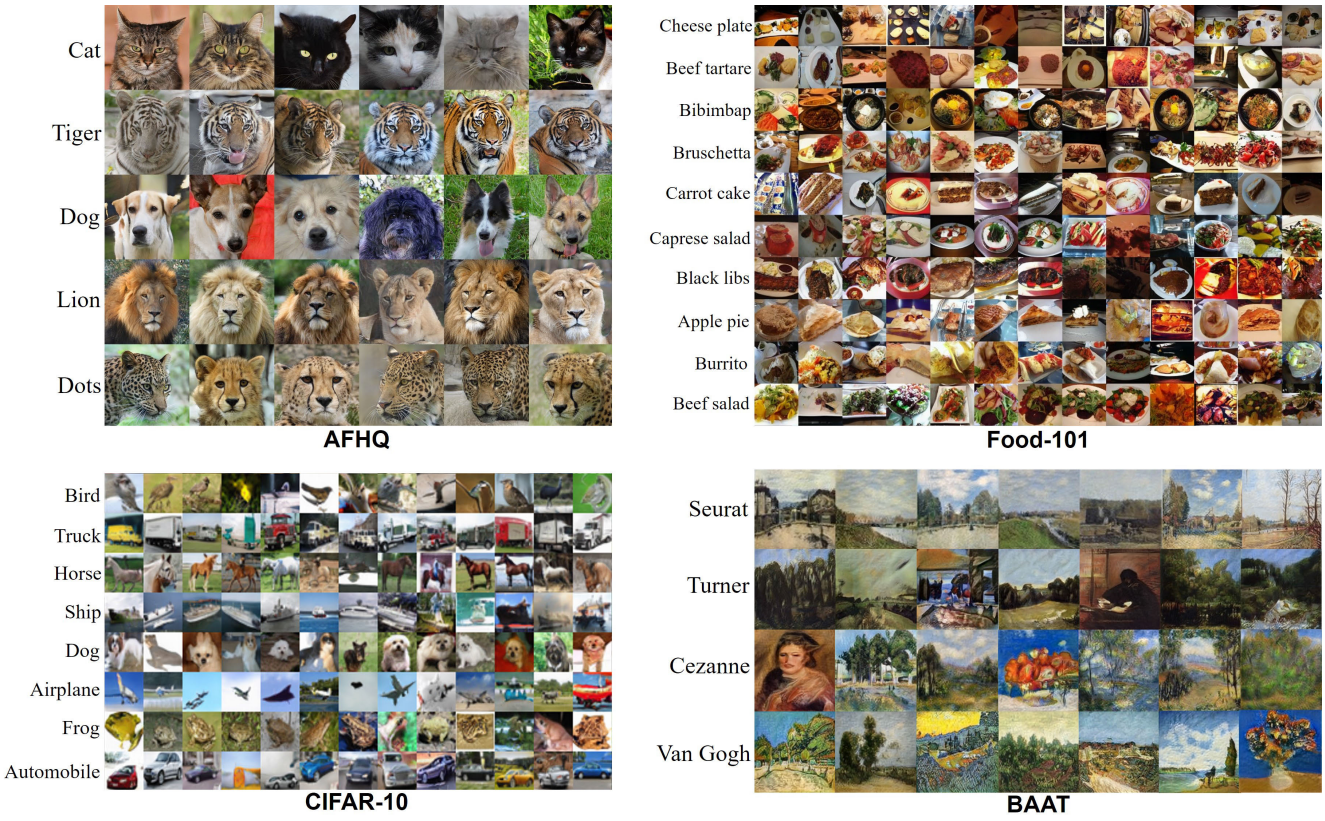


FIGURE 3. Generated images for qualitative result on AFHQ, Food-101, CIFAR-10, and BAAT datasets. Images within the same row belong to the same class. The resolutions of the generated images are 128 × 128 for AFHQ, Food-101, and BAAT, and 32 × 32 for CIFAR-10.

regard, we train and evaluate DuDGAN using four different datasets, each with a different domain and properties. The preprocessing steps and specifics for each dataset are described below:

- AFHQ (512 × 512) [5]: AFHQ is a dataset originally consisting of three different categories: dogs, cats, and wild animals. To demonstrate the effectiveness of DuDGAN method across various domains, we use a recreated version of the dataset [10], which increases the number of classes from 3 to 7. In such a setting, the number of images in each class is imbalanced. For instance, the number of images in 'cat' class is 4739; the largest, while the number of images in 'fox' class is 433; the smallest.
- Food-101 (128 × 128) [34]: Food-101 contains 101 different categories of food, where each class consists of 1k different images. We use a portion of the dataset consisting of 20 labels, without reducing class-wise data size. Additionally, due to the variability of the size of images, we preprocess each training image to a resolution of 128 × 128.
- CIFAR-10 (32 × 32) [35]: CIFAR-10 is divided into 10 different classes, each containing 50k training images and 10k test images.
- BAAT (128 × 128) [36]: 'Best Artworks for All Time' (BAAT) dataset includes various artistic paintings by 50 different artists. Due to the unconventional characteristic of art paintings, the dataset has a relatively high intra-class variation compared to other datasets. Similar to AFHQ, class-wise images in BAAT are also imblanced. For example, the number of image in 'Vincent Van Gogh' is 877, while the number of images in 'Claude Monet' is 73. Additionally, for specific artists, painting style varies, as some of the paintings are portraits, and some are landscape paintings. This characteristic critically hinders the quality of the generated samples due to the extremely large intra-class variation in the small dataset.

B. EXPERIMENT DETAILS

To evaluate DuDGAN and compare it with other models, we employ Fréchet inception distance (FID) [42] and kernel inception distance (KID) [43] to measure the generation quality and assess whether the generation adheres to the distribution of the training data. Additionally, we utilize the Precision and Recall score [44] to gauge the fidelity and diversity of the generated samples. To compare comprehensively, we additionally adopt F1 score [45]; the harmonic mean of Precision and Recall, which demonstrate overall robustness over the distribution. As for computational equipment, we utilized 1 or 2 NVIDIA GeForce 3090 RTX

TABLE 1. Quantitative result on AFHQ, Food-101, CIFAR-10 and BAAT datasets. In the table, bold indicates best result and underline indicates second best result.

Method	AFHQ (512 × 512)					Food-101 (128 × 128)					CIFAR-10 (32 × 32)					BAAT (128 × 128)				
	FID↓	KID↓	Pr↑	Re↑	F1↑	FID↓	KID↓	Pr↑	Re↑	F1↑	FID↓	KID↓	Pr↑	Re↑	F1↑	FID↓	KID↓	Pr↑	Re↑	F1↑
CStyleGAN-ADA	5.11	**0.0010**	**0.75**	**0.31**	**0.44**	13.29	0.0067	0.57	**0.18**	0.27	3.81	0.0011	**0.64**	0.56	0.59	21.18	0.0086	0.65	0.15	0.24
CDiffusion-GAN	5.31	**0.0010**	0.64	0.28	0.39	21.07	0.0130	0.64	0.09	0.16	3.77	0.0011	0.63	0.57	0.60	44.10	0.0213	**0.66**	0.03	0.06
Transitional-CGAN	7.80	0.0011	0.63	0.20	0.30	**10.37**	**0.0034**	0.63	0.17	0.27	4.25	0.0013	**0.64**	0.52	0.57	20.18	0.0071	0.64	0.22	0.33
DuDGAN	**5.10**	**0.0010**	0.68	0.29	0.41	10.71	0.0051	**0.73**	**0.18**	**0.29**	**3.73**	**0.0009**	**0.64**	**0.58**	**0.61**	**19.95**	**0.0051**	**0.66**	**0.31**	**0.42**

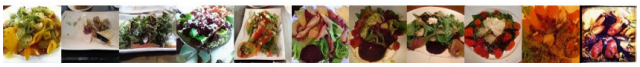
(with 24GB or 48GB memory) or 1 NVIDIA RTX A6000 (with 48GB memory) GPU for all experiments.

To demonstrate that DuDGAN exhibits strength in fast convergence, all models are trained until the discriminator processes 10,000k images, a 60% smaller than those used in experiments with comparative models. Furthermore, especially for the classifier, we adopt AdamW optimizer [46] instead of Adam optimizer [47] to prevent class-induced overfitting. The default setting of each model with 64 batch sizes was adopted for fair comparison. We used the 'cifar' configuration setting for the CIFAR-10 dataset [35], and the 'paper256' configuration setting for the remaining datasets, as specified in [22]. Furthermore, we employed diffusion-based noise intensity by leveraging the priority sampling scheme [32] in all experiments.

Additionally, under the computation equipment, DuDGAN's training time only increased by 12% compared to Diffusion-GAN, due to additional parameters of classifier network, while consistently achieving superior performance and training stability.

For fair comparison, the main experiment is built upon comparison with three different baseline methods, which are based on StyleGAN2-ADA [23]: 1) class-conditional training of StyleGAN2-ADA (CStyleGAN2-ADA), 2) class-conditional training of Diffusion-StyleGAN2 (CDiffusion-GAN) [32], and 3) default setting of Transitional-CGAN [28]. Note that, while CDiffusion-GAN is one of the baselines, the model is first introduced in this study, as a class-conditional version of Diffusion-GAN.

(a) Transitional-CGAN



(b) Ours

FIGURE 4. Analysis of 'Caprese Salad' label generation on the Food-101 dataset. The figure presents the images generated using (a) Transitional-CGAN and (b) novel DuDGAN model. Although Transitional-CGAN outperforms DuDGAN for FID and KID on Food-101 dataset, we observe that DuDGAN outperforms Transitional-CGAN for generation quality in terms of human perception.

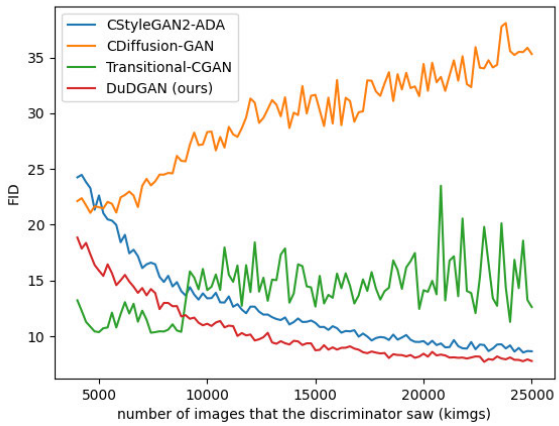


FIGURE 5. FID on the Food-101 dataset post 4,000k image exposure to the discriminator. This depicts the FID after the discriminator has been exposed to 4,000k images, marking the end of the transition division in Transitional-CGAN. DuDGAN consistently maintains training stability throughout the training procedure and exhibits the best FID, compared to other models.

C. QUALITATIVE RESULT

As demonstrated in Table 1, DuDGAN surpasses the comparison models with respect to FID on the AFHQ and CIFAR-10 datasets, indicating superior generation quality. In particular, on the CIFAR-10 dataset, DuDGAN outperforms all other models, including the main baseline, CDiffusion-GAN, across all datasets. The result implies adopting basic diffusion process in class-conditional GAN training is not a effective approach. Notably, in the case of the AFHQ dataset, FID is reduced by 4.0%. Furthermore, DuDGAN attains the highest Recall score on the CIFAR-10 dataset, signifying enhanced diversity in the generated samples. Although DuDGAN exhibits a marginally lower Precision and Recall score on the AFHQ dataset, it remains competitive with the top-performing model. In the case of BAAT dataset, DuDGAN achieves best results compared with other models.

Additionally, we observe that while Transitional-CGAN seems to produce high-quality images for the Food-101 dataset, due to comparatively lower values of FID and KID. However, it fails to learn accurate class-wise distribution, leading to degenerate qualitative results. The reason is that the unique traning scheme of Transitional-CGAN which unconditionally trained first and the metrics used bias the total distribution, neglecting the sub-distribution segregated by class labels, as shown in Fig. 4.

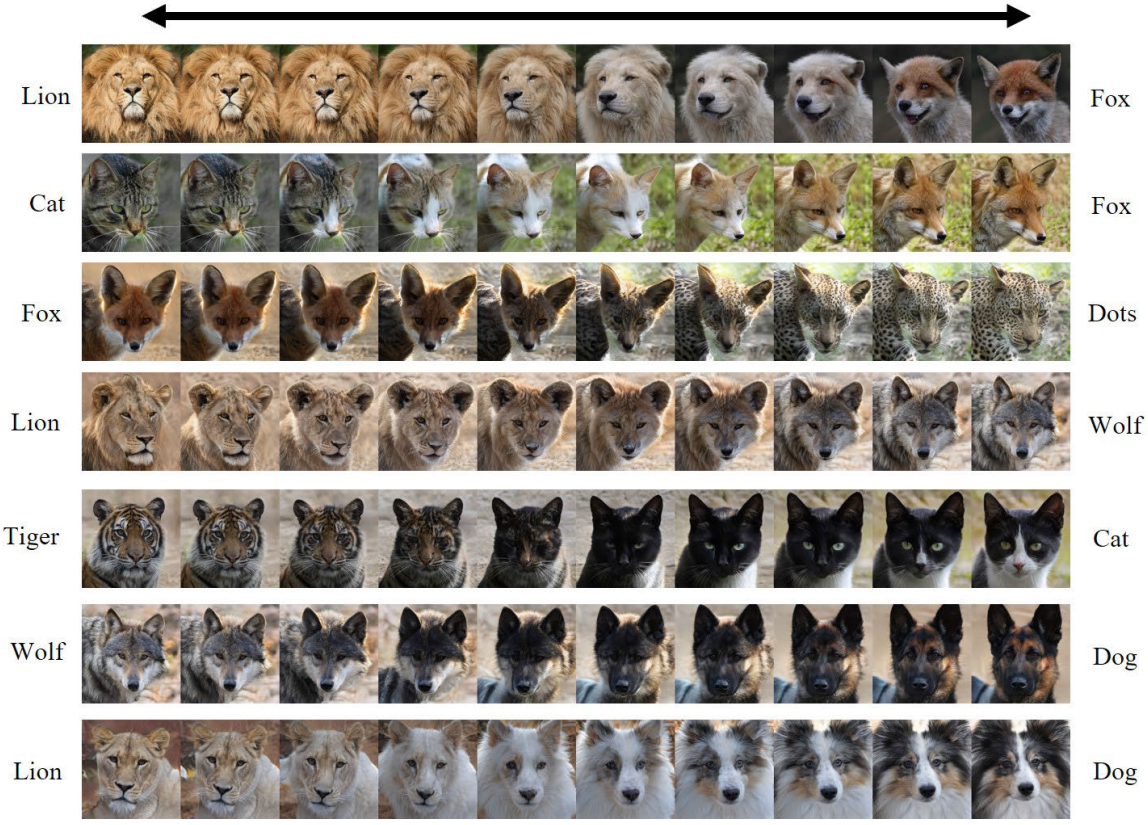


FIGURE 6. Inter-Class interpolation on AFHQ dataset. This figure shows the smooth transition of DuDGAN between different classes on the AFHQ dataset.

TABLE 2. Outcomes of prolonged training iterations. This table details the results following further iterations, until the discriminator has processed 25,000 images.

Method (for prolonged iterations)	Food-101(128 × 128)				
	FID↓	KID↓	Pr↑	Re↑	F1↑
CStyleGAN2-ADA	8.71	0.0037	0.69	0.22	0.33
CDiffusion-GAN	21.07	0.0130	0.64	0.09	0.16
Transitional-CGAN	10.37	0.0034	0.63	0.17	0.27
DuDGAN	**7.66**	**0.0030**	**0.70**	**0.23**	**0.35**

TABLE 3. Ablation study: Classifier loss formulations.

Experiment of L_C	CIFAR-10 (32 × 32)				
	FID↓	KID↓	Pr↑	Re↑	F1↑
$L_C = L_{\text{cont}}^{\text{real}}(f_{\text{high}}, c_r)$	68.14	0.0315	0.63	0.01	0.02
$L_C = L_{\text{cls}}^{\text{real}}$	3.96	0.0011	0.63	0.57	0.60
L_C (DuDGAN)	**3.73**	**0.0009**	**0.64**	**0.58**	**0.61**

D. QUANTITATIVE RESULT

In addition to the quantitative results, we examine the visual quality of the generated images by DuDGAN. As shown in Fig. 3, the generated samples exhibit photo-realistic characteristics, demonstrating the efficacy of DuDGAN. The images possess fine details, accurate colors, and clear textures, which contribute to their overall photo-realistic appearance. These results further validate the superiority of DuDGAN in generating high-quality, diverse, and visually appealing images within class-wise distributions.

E. EXTENDING THE TRAINING ITERATIONS

While DuDGAN manifests robustness in low-iteration training, we further the experimentation by comparing results over extended iterations until the discriminator is exposed to 25,000k images (Table. 2, Fig. 5) [23], [28], [32]. The experiment show that FID of CDiffusion-GAN continuously diverge and Transitional-CGAN's FID oscilates after unconditional training stage. This is probably because mode collapse occurs after a certain number of iterations during the class-conditioning process due to an imbalance in learning between the discriminator and generator. Moreover, CStyleGAN2-ADA's convergence speed is slower than DuDGAN. This experiment underscores DuDGAN's ability to provide consistent, high-quality results even over prolonged training, a testament to the model's endurance and adaptability.

F. INTER-CLASS INTERPOLATION

Perhaps one of the most captivating demonstrations of DuDGAN's abilities is depicted in Fig. 6, where we observe

TABLE 4. Ablation study: Examination of generator's loss formulation.

Experiment of L_G		CIFAR-10 (32 × 32)				
		FID↓	KID↓	Pr↑	Re↑	F1↑
contrastive + classification	$L_G = \lambda_G \cdot L_G^{\text{gen}} + (1-\lambda_G) \cdot L_{\text{cont}}^{\text{gen}}\left(f_{\text{high}}, c_f\right) + 0.3 \cdot L_{\text{cls}}^{\text{gen}}$	32.22	0.0101	0.61	0.04	0.08
	$L_G = \lambda_G \cdot L_G^{\text{gen}} + (1-\lambda_G) \cdot L_{\text{cont}}^{\text{gen}}\left(f_{\text{high}}, c_f\right) + 0.5 \cdot L_{\text{cls}}^{\text{gen}}$	5.36	0.0015	0.63	0.48	0.54
only contrastive (DuDGAN)	$\boldsymbol{L_G = \lambda_G \cdot L_G^{\text{gen}} + (1-\lambda_G) \cdot L_{\text{cont}}^{\text{gen}}\left(f_{\text{high}}, c_f\right)}$	**3.73**	**0.0009**	**0.64**	**0.58**	**0.61**

TABLE 5. Ablation study: Variation in λ_G.

Experiment of λ_G ($\lambda_C = 0.95$)	CIFAR-10 (32 × 32)				
	FID↓	KID↓	Pr↑	Re↑	F1↑
$\lambda_G = 0.5$	4.55	0.0015	0.63	0.53	0.58
$\lambda_G = 0.8$	54.88	0.0217	**0.67**	0.02	0.04
$\boldsymbol{\lambda_G = 0.95}$ (DuDGAN)	**3.73**	**0.0009**	0.64	**0.58**	**0.61**

TABLE 6. Ablation study: Variation in λ_C.

Experiment of λ_C ($\lambda_G = 0.95$)	CIFAR-10 (32 × 32)				
	FID↓	KID↓	Pr↑	Re↑	F1↑
$\lambda_C = 0.5$	4.02	0.0012	0.63	0.56	0.59
$\lambda_C = 0.8$	4.24	0.0013	**0.64**	0.54	0.59
$\boldsymbol{\lambda_C = 0.95}$ (DuDGAN)	**3.73**	**0.0009**	**0.64**	**0.58**	**0.61**

TABLE 7. Ablation study: Variation in diffusion process.

Experiment on diffusion	Diffusion		CIFAR-10 (32 × 32)				
	on D	on C	FID↓	KID↓	Pr↑	Re↑	F1↑
No Diffusion	X	X	3.89	0.0011	0.64	0.56	0.59
Mono-Diffusion	O	X	4.41	0.0013	0.64	0.53	0.58
	X	O	4.23	0.0017	**0.65**	0.54	0.59
Dual-Diffusion (DuDGAN)	**O**	**O**	**3.73**	**0.0009**	0.64	**0.58**	**0.61**

the seamless class interpolation on the AFHQ dataset. Fig. 6 underscores DuDGAN's deftness in synthesizing smooth transitions between different classes, reinforcing the model's robustness in dealing with bounded label information. It bears testament to the ingenious design choices behind DuDGAN and the profound impact they have on the model's performance.

V. ABLATION STUDY

In the ablation study, we conduct three distinct experiments by modifying the formulation of the classifier's loss function, adjusting the hyperparameter λ in both the generator and the classifier, and changing diffusion procedure for the noise-aware networks; the discriminator and the classifier. The experiments in this section are based on the class-conditional training of the CIFAR-10 dataset. Note that since GANs have a multi-network training scheme, an imbalance in the losses directly leads to training instability and mode collapse problems, resulting in an explosion in metrics, as in some cases in Tables 3, 4, and 5.

A. TWO-LEVEL OUTPUT OF THE CLASSIFIER

DuDGAN's classifier generates a two-level loss, computed as the label-dimensioned logits and contrastive loss, with the aim of providing informative guidance to the generator. To verify the effectiveness of this formulation in generating high-quality and diverse images, we evaluate two different loss formulations. As demonstrated in Table. 3, our two-level loss significantly contributes to the training process, resulting in superior performance across various metrics.

B. LOSS FORMULATION OF THE GENERATOR

Table. 4 illustrates an important element of the exploration into the generator's loss formulation as it pertains to DuDGAN. We experimented with two different weights for the classification loss (0.3 and 0.5). The numerical results indicate that there is a marked decrease in the FID, KID, Precision, and Recall scores when the weight of the classification loss is increased from 0.3 to 0.5.

Despite these improvements, the scenario wherein we utilize only the contrastive loss (i.e., the absence of the classification loss) in the generator's loss formulation exhibits the best overall performance. This configuration led to the lowest FID and KID values (3.73 and 0.0009, respectively), and the highest Precision and Recall scores (0.64 and 0.58, respectively). This outcome underscores the effectiveness of our novel approach, demonstrating that by employing only the contrastive loss, we have managed to avoid the potential pitfalls associated with early-stage divergence from incorrect class prediction, a common issue prevalent in other models.

C. HYPERPARAMETER SETTING

As described in Eq. 14 and Eq. 15, the primary role of the predefined hyperparameter λ is to balance the influence of each network, namely the classifier and the generator, during the training process. Following a similar approach, we assess the metrics while varying λ to investigate the optimal balance in the network. In Table. 5 and Table. 6, the metrics are computed with different λ values in the target network, while other parameters remain constant. DuDGAN proposes

a suitable setting for both λ_G and λ_C at 0.95, as evidenced by the best values across all pairings in Table. 5 and Table. 6.

D. DIFFUSION PROCEDURE

We perform additional experiments involving three distinct diffusion mechanisms, in addition to a baseline model trained without any noising steps. As evidenced in Table 7, DuDGAN's dual-diffusion network outperforms competing methods across various metrics. This superior performance is attributable to the dual-diffusion mechanism's ability to effectively utilize class information and enhance generation quality, thereby facilitating network equilibrium during the training phase.

VI. CONCLUSION

In this paper, we propose novel approaches to the class-conditional GAN training procedure via dual-diffusion, which entails diffusion-based noise injection using Gaussian-mixture noise. Throughout the training process, the discriminator and the classifier are trained with gradually noisy images, mitigating overfitting within the networks. DuDGAN's independent classifier generates a two-level loss comprising the label-supervised contrastive loss and the classification loss, which guides the generator by providing informative feedback. With the assistance of both the discriminator and the classifier, the generator successfully produces high-quality and diverse images corresponding to specific labels. Moreover, DuDGAN facilitates iteration-efficient training, as demonstrated by rapid convergence within a limited number of iterations. Consequently, DuDGAN achieves superior results in both quantitative and qualitative evaluations, outperforming state-of-the-art class-conditional GAN models.

Despite DuDGAN's novel approach to class-conditional GAN training, there are inherent limitations and avenues for improvement. While DuDGAN excels in dataset-agnostic training, it struggles with classes containing fewer than 50 images, compromising the quality of generated conditional samples.

We propose separate noising schedules for the discriminator and classifier, which operate independently. This independence can complicate optimization. A unified noising schedule, incorporating feedback from both components, could enhance both generation quality and training efficiency.

The current framework offers potential extensions to other tasks, such as diffusion-based 3D scene generation and GAN inversion, which involve multi-network training. These extensions are subjects for future investigation.

REFERENCES

[1] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," in *Proc. Int. Conf. Learn. Represent.*, 2018, pp. 1–26.

[2] A. Brock, J. Donahue, and K. Simonyan, "Large scale GAN training for high fidelity natural image synthesis," in *Proc. Int. Conf. Learn. Represent.*, 2019, pp. 1–35. [Online]. Available: https://openreview.net/forum?id=B1xsqj09Fm

[3] M. Lee and J. Seok, "Controllable generative adversarial network," *IEEE Access*, vol. 7, pp. 28158–28169, 2019.

[4] A. Sauer, K. Chitta, J. Müller, and A. Geiger, "Projected GANs converge faster," in *Proc. Neural Inf. Process. Syst. (NIPS)*, vol. 34, 2021, pp. 17480–17492.

[5] Y. Choi, Y. Uh, J. Yoo, and J.-W. Ha, "StarGAN v2: Diverse image synthesis for multiple domains," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 8185–8194.

[6] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2242–2251.

[7] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, "StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8789–8797.

[8] H. Emami, M. M. Aliabadi, M. Dong, and R. B. Chinnam, "SPA-GAN: Spatial attention GAN for image-to-image translation," *IEEE Trans. Multimedia*, vol. 23, pp. 391–401, 2021.

[9] S. Xie, M. Gong, Y. Xu, and K. Zhang, "Unaligned image-to-image translation by learning to reweight," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 14154–14164.

[10] K. Ko, T. Yeom, and M. Lee, "SuperstarGAN: Generative adversarial networks for image-to-image translation in large-scale domains," *Neural Netw.*, vol. 162, pp. 330–339, May 2023.

[11] S. Tulyakov, M.-Y. Liu, X. Yang, and J. Kautz, "MoCoGAN: Decomposing motion and content for video generation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1526–1535.

[12] Y. Wang, P. Bilinski, F. Bremond, and A. Dantcheva, "ImaGINator: Conditional spatio-temporal GAN for video generation," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2020, pp. 1149–1158.

[13] M. Chu, Y. Xie, J. Mayer, L. Leal-Taixé, and N. Thuerey, "Learning temporal coherence via self-supervision for GAN-based video generation," *ACM Trans. Graph.*, vol. 39, no. 4, pp. 1–75, Aug. 2020.

[14] K. Schwarz, Y. Liao, M. Niemeyer, and A. Geiger, "GRAF: Generative radiance fields for 3D-aware image synthesis," in *Proc. 34th Int. Conf. Neural Inf. Process. Syst.*, 2020, pp. 20154–20166.

[15] Q. Meng, A. Chen, H. Luo, M. Wu, H. Su, L. Xu, X. He, and J. Yu, "GNeRF: GAN-based neural radiance field without posed camera," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 6331–6341.

[16] M. Niemeyer and A. Geiger, "GIRAFFE: Representing scenes as compositional generative neural feature fields," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 11448–11459.

[17] E. R. Chan, C. Z. Lin, M. A. Chan, K. Nagano, B. Pan, S. de Mello, O. Gallo, L. Guibas, J. Tremblay, S. Khamis, T. Karras, and G. Wetzstein, "Efficient geometry-aware 3D generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 16123–16133.

[18] E. Richardson, Y. Alaluf, O. Patashnik, Y. Nitzan, Y. Azar, S. Shapiro, and D. Cohen-Or, "Encoding in style: A styleGAN encoder for image-to-image translation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 2287–2296.

[19] O. Tov, Y. Alaluf, Y. Nitzan, O. Patashnik, and D. Cohen-Or, "Designing an encoder for StyleGAN image manipulation," *ACM Trans. Graph.*, vol. 40, no. 4, pp. 1–14, Aug. 2021.

[20] T. Wang, Y. Zhang, Y. Fan, J. Wang, and Q. Chen, "High-fidelity GAN inversion for image attribute editing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 11369–11378.

[21] 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. 4396–4405.

[22] 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. 8107–8116.

[23] T. Karras, M. Aittala, J. Hellsten, S. Laine, J. Lehtinen, and T. Aila, "Training generative adversarial networks with limited data," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 12104–12114.

[24] T. Karras, M. Aittala, S. Laine, E. Harkonen, J. Hellsten, J. Lehtinen, and T. Aila, "Alias-free generative adversarial networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 852–863.

[25] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, *arXiv:1411.1784*.

[26] A. Odena, C. Olah, and J. Shlens, "Conditional image synthesis with auxiliary classifier GANs," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 2642–2651.
[27] M. Kang and J. Park, "ContraGAN: Contrastive learning for conditional image generation," in *Proc. Adv. Neural Inf. Process. Syst.*, 2020, pp. 21357–21369.
[28] M. Shahbazi, M. Danelljan, D. P. Paudel, and L. V. Gool, "Collapse by conditioning: Training class-conditional GANs with limited data," in *Proc. Int. Conf. Learn. Represent.*, 2022, pp. 1–21.
[29] M. Kang, W. J. Shim, and M. Cho, "Rebooting ACGAN: Auxiliary classifier GANs with stable training," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 23505–23518.
[30] H.-Y. Tseng, L. Jiang, C. Liu, M.-H. Yang, and W. Yang, "Regularizing generative adversarial networks under limited data," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 7917–7927.
[31] S. Zhao, Z. Liu, J. Lin, J.-Y. Zhu, and S. Han, "Differentiable augmentation for data-efficient GAN training," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 7559–7570.
[32] Z. Wang, H. Zheng, P. He, W. Chen, and M. Zhou, "Diffusion-GAN: Training GANs with diffusion," in *Proc. 11th Int. Conf. Learn. Represent.*, 2022, pp. 1–26.
[33] P. Khosla, P. Teterwak, C. Wang, A. Sarna, Y. Tian, P. Isola, A. Maschinot, C. Liu, and D. Krishnan, "Supervised contrastive learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 18661–18673.
[34] L. Bossard, M. Guillaumin, and L. van Gool, "Food-101-mining discriminative components with random forests," in *Proc. 13th Eur. Conf. Comput. Vis.*, Zurich, Switzerland. Cham, Switzerland: Springer, 2014, pp. 446–461.
[35] A. Krizhevsky, "Learning multiple layers of features from tiny images," Univ. Toronto, Toronto, ON, Canada, Tech. Rep., 2009.
[36] ICARO. *Best Artworks of All Time, Kaggle*. Accessed: Oct. 1, 2023. [Online]. Available: https://www.kaggle.com/datasets/ikarus777/best-artworks-of-all-time
[37] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014, pp. 1–9.
[38] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 6840–6851.
[39] P. Dhariwal and A. Nichol, "Diffusion models beat GANs on image synthesis," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 8780–8794.
[40] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, and S. Ganguli, "Deep unsupervised learning using nonequilibrium thermodynamics," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 2256–2265.
[41] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," 2013, *arXiv:1312.6114*.
[42] 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," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–12.
[43] M. Bińkowski, D. J. Sutherland, M. Arbel, and A. Gretton, "Demystifying MMD GANs," in *Proc. Int. Conf. Learn. Represent.*, 2018, pp. 1–36.
[44] T. Kynkäänniemi, T. Karras, S. Laine, J. Lehtinen, and T. Aila, "Improved precision and recall metric for assessing generative models," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 1–10.
[45] Y. Sasaki, "The truth of the F-measure," *Teach Tutor mater*, vol. 1, no. 5, pp. 1–5, 2007.
[46] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *Proc. Int. Conf. Learn. Represent.*, 2018, pp. 1–18.
[47] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

TAESUN YEOM is currently pursuing the bachelor's degree in mechanical engineering with Chung-Ang University, Seoul, South Korea. His current research interests include implicit neural representations, model merging, and the properties of the weight space of neural networks.

CHANHOE GU received the B.S. degree in physics from Chung-Ang University, Seoul, South Korea, in 2021, where he is currently pursuing the M.S. degree in intelligent semiconductor engineering. His research interests include AI-based inverse design and generative adversarial networks.

MINHYEOK LEE (Member, IEEE) received the bachelor's and Ph.D. degrees in electrical engineering from Korea University, in 2015 and 2020, respectively. He has been an Assistant Professor with the School of Electrical and Electronics Engineering, Chung-Ang University, since 2021, and concurrently holds a position with the Department of Intelligent Semiconductor Engineering, since 2023. Prior to these appointments, he was a Research Professor with Korea University, from 2020 to 2021. Over the past three years, he has published over 40 papers in international journals and conferences, primarily focusing on artificial intelligence and generative AI. His research interests include generative artificial intelligence, generative adversarial networks (GANs), diffusion models, neural radiance fields (NeRF), AI-based finance engineering, and bioinformatics.