

Received 5 June 2025, accepted 19 June 2025, date of publication 24 June 2025, date of current version 8 July 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3582756

RESEARCH ARTICLE

Regularization for Unconditional Image Diffusion Models via Shifted Data Augmentation

KENSUKE NAKAMURA¹, BONG-SOO SOHN², SIMON KORMAN³,
AND BYUNG-WOO HONG¹

¹Artificial Intelligence Department, Chung-Ang University, Seoul 06974, South Korea

²Computer Science Department, Chung-Ang University, Seoul 06974, South Korea

³Department of Computer Science, University of Haifa, Haifa 3498838, Israel

Corresponding author: Byung-Woo Hong (hong@cau.ac.kr)

This work was supported in part by the Korea Government [Ministry of Science and ICT (MSIT)], Korea under Grant IITP-RS-2021-II211341 and in part by the Artificial Intelligence Graduate School, Chung-Ang University under Grant NRF-RS-2023-00251366.

ABSTRACT Diffusion models are a powerful class of techniques in ML for generating realistic data, but they are highly prone to overfitting, especially with limited training data. While data augmentation such as image rotation can mitigate this issue, it often causes leakage, where augmented content appears in generated samples. In this paper, we propose a novel regularization framework, called shifted data-augmentation (SDA), for training unconditional diffusion models. SDA introduces an auxiliary diffusion path using transformed data and the noise-shift technique alongside the standard path with original data. This dual-path structure enables effective regularization while suppressing leakage through a trajectory shift in the diffusion process. We implement SDA with image rotation as its most basic and interpretable configuration. We also conduct synthetic and empirical analyses demonstrating that SDA effectively leverages the regularization benefit of image rotation. In particular, SDA yielded notable performance in training with limited data.

INDEX TERMS Data augmentation, diffusion model, noise-shift, regularization.

I. INTRODUCTION

Image generation is one of the fundamental tasks in Machine Learning, aiming to estimate the distribution of given training images and generate diverse and novel data. It has been widely utilized as an essential tool in various practical applications, e.g., image super-resolution [1], [2], image retrieval [3], image restoration [4], semantic segmentation [5], medical imaging [6], image synthesis [7], anomaly detection [8], and many others. Early works in this field saw remarkable success with generative adversarial networks (GANs) [9], [10], [11], which generate images using a generator trained against a discriminator network in an adversarial way.

Later, the diffusion model, or denoising diffusion probabilistic model (DDPM) [12], [13], [14], [15], which reconstructs realistic images from noisy inputs through a reverse diffusion process, demonstrated superior image

The associate editor coordinating the review of this manuscript and approving it for publication was Thomas Canhao Xu¹.

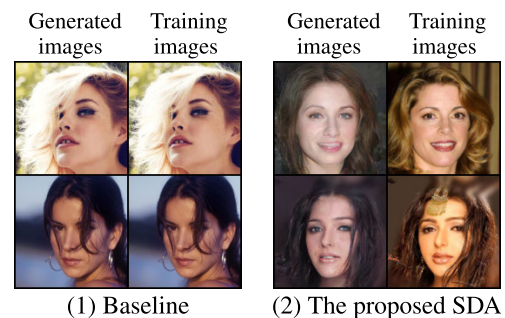


FIGURE 1. [The memorization problem of diffusion model] (1st column in each part) Generated images and (2nd column) nearest training images based on the cosine similarity using (left part) the baseline DDPM and (right) the proposed SDA trained using a limited set of $N = 256$ training examples from CelebA dataset [16]. More details are presented in the experimental section.

quality compared to GANs, establishing it as the mainstream for unconditional image generation. However, diffusion models are highly prone to overfitting due to their high capacity, which can degrade the quality of generated images.

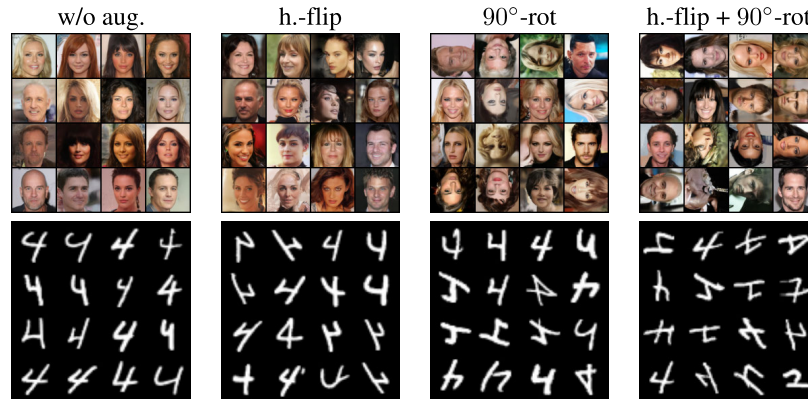


FIGURE 2. [The leakage of data augmentation] Generated images based on the baseline DDPM using (1st column) the training examples without data transformation, (2nd column) those using horizontal-flipping, (3rd column) 90°-rotations, and (4th column) their combination, based on (1st row) CeleBA [16] and (2nd row) '4'-images in MNIST dataset [17].

In extreme cases, the trained model may memorize and regenerate specific training samples, as demonstrated in Figure 1.

Data augmentation techniques, e.g., horizontal-flipping, image rotation, and hue changes, have been widely adopted in image classification and generation tasks [18], [19], [20], [21], [22] to enhance generalization under limited data. Applying such data augmentation to diffusion models is a promising direction for mitigating overfitting, but it introduces a major challenge known as the leakage problem, i.e., augmented images (e.g., rotated faces) can be directly reproduced by the model, as demonstrated in Figure 2. To prevent this, previous diffusion studies restrict augmentation to horizontal-flipping under the assumption of left-right symmetry.

In this paper, we propose a novel regularization method, called shifted data-augmentation (SDA), which incorporates transformed data into training while avoiding leakage through a noise-shifted auxiliary path. Specifically, SDA introduces two diffusion paths: one for the original images and one for augmented versions with noise shift. SDA enables mixing of the distributions of the original and transformed data in the fine stage, where the model learns fine details, while maintaining a sufficient trajectory shift between the two paths in the coarse stage. At inference, this trajectory shift ensures the generation follows the original data path, leading to diverse and realistic outputs without leakage. We implement the proposed SDA with image rotation as its most basic and interpretable configuration. Then, we present a synthetic experiment to illustrate the effect of the noise shift and its role in SDA. Quantitative and qualitative experiments across multiple datasets show that SDA successfully reduces overfitting and prevents leakage while maintaining generation quality. The code is available at *IEEE DataPort*.

We relate our method to prior works in Section II. We review diffusion models and augmentation strategies in Section III, followed by the proposed method in Section IV. Experiments are described in Section V, and conclusions are given in Section VI.

II. RELATED WORKS

A. DATA AUGMENTATION IN MACHINE LEARNING TASKS

Collecting a large number of high-quality images is often costly in many domains, such as human faces and medical imaging. To address this, various data augmentation techniques, including geometric and color-space transformations, have been applied to semi-supervised learning of networks for image classification tasks [18], [19], [20], [21], [22]. Additionally, data augmentation plays a crucial role in contrastive learning by enhancing the representation learning of training data [23], [24], [25], [26].

Considering its success in these Machine Learning tasks, it is natural to explore the application of data augmentation in generative models to improve generalization in an unsupervised manner at low cost.

B. DATA AUGMENTATION IN GAN

In order to train generative adversarial networks (GANs) with limited examples and further improve the quality of generated data, data transformation techniques such as image rotation, cropping [27], cut-out [28], [29], and noise injection [30], [31] have been applied during GAN training. In particular, StyleGAN2-Ada [10] adaptively adjusts the strength of data augmentation applied to training images using a heuristic metric based on the output confidence of the discriminator network. This ADA framework was later integrated into StyleGAN3 [11], which is considered a state-of-the-art GAN. However, these techniques are essentially ad hoc, as they rely on scheduled or adaptive annealing of data augmentation to prevent augmented data from being generated during inference.

Moreover, ADA [10] tunes the transformation probability based on the discriminator output, and is not applicable to diffusion models. In contrast, we propose a method that introduces an auxiliary diffusion path incorporating augmented data directly into the diffusion process, eliminating the need for augmentation annealing during training.

C. DIFFUSION MODELS

DDPMs [12], [13], [14], [15] are a Machine Learning framework for generating realistic data by reversing a noise injection process. The diffusion model has shown state-of-the-art performance compared to generative adversarial networks [32], and achieved remarkable success in various tasks, such as image super-resolution [33], [34], synthetic image generation [35], [36], medical imaging [37], image synthesis [38], [39], image restoration [40], and out-of-distribution detection [41].

However, the diffusion model is also prone to overfitting to the training examples, which compromises its generalization ability. In the worst case, the diffusion model directly replicates individual training images, a phenomenon known as memorization [42], [43], [44]. Unfortunately, the diffusion model also suffers from the leakage problem, which limits the practical use of data augmentation during training to horizontal-flipping [14]. This paper addresses the overfitting issue in diffusion models by introducing our approach that leverages the effect of data augmentation while preventing leakage during inference. Moreover, to the best of our knowledge, this is the first work that explicitly addresses the leakage problem in diffusion model training.

Denoising Diffusion Implicit Models (DDIM) [45] introduce a deterministic sampling process by reparameterizing the reverse diffusion trajectory, allowing for faster image generation without stochastic noise. However, since DDIM is primarily designed for efficient inference, it involves a trade-off between the number of sampling steps and generation fidelity. Moreover, it does not solve the overfitting issues of diffusion models in the training phase.

Furthermore, Stable Diffusion [46] is a representative example of pretrained diffusion models for text-to-image generation tasks. However, Stable Diffusion has also been reported to suffer from the issue of memorizing training images [47], [48]. Although several mitigation strategies have been proposed [49], [50], they are primarily designed to avoid generating outputs that closely resemble training examples during inference, rather than addressing the root cause of memorization during training. In contrast, while text-to-image generation is beyond the scope of this study, our method is novel in that it proactively suppresses overfitting during training via data transformations enabled by the proposed SDA framework.

D. NOISE SHIFT IN DIFFUSION MODELS

Noise shift is a data-driven approach for constructing class-conditional DDPMs without modifying the architecture. It defines, in practice, different mean values of noise data for different classes, and learns the trajectory from clean data to class-specific noise distributions. In inference, clean images are generated from noise data with the class-specific mean values.

For noise shift, two research groups independently proposed Shift-DDPM [51] and Shifted-Diffusion [52] at around the same time. Shift-DDPM is an explicit method that

defines the class-wise trajectories explicitly, and the diffusion model only predicts noise variance in noisy images. Shifted-Diffusion [52] is an implicit method in which the diffusion model is trained to predict not only the noise variance but also the mean trajectory. We observed that the explicit method, Shift-DDPM, is superior to Shifted-Diffusion in accuracy, as the implicit approach increases the computational complexity of model training.

We adapt the noise-shift technique, originally developed for conditional DDPMs, to unconditional image generation. In contrast to prior works [51], [52], which primarily focus on the separation of trajectories during the coarse stage of the diffusion process, we uniquely highlight an overlooked aspect in the fine stage. Specifically, we observe that noise-shifted trajectories tend to intersect in the fine stage, leading to information blending across different ‘classes.’ By leveraging this aspect, we design an auxiliary path in the diffusion process to leverage the regularization effect of data augmentation.

E. REGULARIZATION VIA CONTRASTIVE AND ADVERSARIAL LEARNING

In addition, we review recent regularization methods based on contrastive learning and adversarial learning frameworks. For example, it was demonstrated [53] that fusing features from two distinct models can enhance classification accuracy and suppress overfitting. On the other hand, data-driven regularization, including our approach, aim to provide diverse data through augmentation, enabling effective model training without requiring additional datasets or multiple models.

SimCLR [24] is a pioneering work in self-supervised contrastive learning, which identifies shared semantics from positive pairs generated via data augmentation, particularly image rotation. Xu et al. [54] extended this concept by combining contrastive learning with adversarial robustness, where negative pairs are generated by applying similar transformations to different samples. Although these studies represent novel forms of regularization via data augmentation, they cannot be directly applied to image generation tasks where augmented data are treated as negative samples.

Regarding diffusion sampling, Ouyang et al. [55] proposed a method to enhance the semantic alignment of generated images by incorporating contrastive loss based on positive pairs (e.g., same-class examples). While their method targets adversarial robustness, leveraging shared features among classes may also help prevent overfitting to individual training examples. In contrast, our method aims to prevent model overfitting without relying on additional positive/negative pair information.

Classifier-free guidance [56] leverages pretrained conditional diffusion models to guide the image generation toward a specific class without relying on an external classifier. This approach adopts a contrastive-like mechanism by treating class-unconditional predictions as implicit negatives. However, since this method operates solely at inference, it does not address the issue of training data memorization inherent to diffusion models.

III. PRELIMINARY

A. DIFFUSION MODELS

DDPM is a generative model that approximates the inverse of a forward diffusion process. Specifically, we focus on the unconditional DDPM for image generation tasks.

In the forward process, noise is added incrementally to a given clean image (x) over a series of steps, resulting in a sequence of data $\{x_0, x_1, x_2, \dots, x_T\}$, where T is the total number of diffusion steps. Formally, this forward diffusion process is the approximate posterior $q(x_{1:T}|x_0)$ which follows a Markov chain defined by a scheduled variances $\{\beta\}_{t=1}^T$ as

$$q(x_t|x_{t-1}) := \mathcal{N}(x_t; \sqrt{1 - \beta_t}x_{t-1}, \beta_t \mathbf{I}), \quad (1)$$

where $\mathbf{I}$ is the identity matrix. Eq.(1) allows us to sample any intermediate x_t directly from the clean image x_0 using

$$q(x_t|x_0) := \mathcal{N}(\sqrt{\bar{\alpha}_t}x_0, (1 - \bar{\alpha}_t)\mathbf{I}), \quad (2)$$

where $\bar{\alpha}_t := \prod_{s=1}^t \alpha_s$, $\alpha_s := 1 - \beta_s$, are constant and we obtain

$$x_t = \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon, \quad \epsilon \sim \mathcal{N}(0, \mathbf{I}). \quad (3)$$

The diffusion model ϵ_θ parameterized by θ predicts ϵ in x_t , and the model parameter θ is updated by gradient descent with a simplified loss given by

$$\mathcal{L}(x_0, \theta) := \mathbb{E} \left[\|\epsilon - \epsilon_\theta(\sqrt{\bar{\alpha}_t}x_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon, t)\|^2 \right]. \quad (4)$$

In the reverse process, given a pure Gaussian noise $x_T \sim \mathcal{N}(0, \mathbf{I})$, the model gradually predicts less noisy data as $\{x_T, x_{T-1}, x_{T-2}, \dots, x_0\}$ and obtains a clean image x_0 . Concretely, each reverse step is analogous to Langevin dynamics as

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}} \epsilon_\theta(x_t, t) \right) + \sigma_t z, \quad (5)$$

where ϵ_θ is the trained model, σ_t is the noise variance at step t scheduled using β , and $z \sim \mathcal{N}(0, \mathbf{I})$ is the Langevin noise.

B. DIFFUSION MODELS WITH NOISE SHIFT

Shift-DDPM [51] is a data-driven conditional DDPM that guides the diffusion forward and reverse processes whose trajectory of mean and variance is explicitly defined. Concretely,

$$q(x_t|x_0, c) := \mathcal{N}(\sqrt{\bar{\alpha}_t}x_0 + k_t \cdot E(c), (1 - \bar{\alpha}_t)\Sigma(c)), \quad (6)$$

where k_t is an arbitrary function that schedules the trajectory of noise mean over t , and $E(c)$ and $\Sigma(c)$ define the shift amount for each class c , respectively.

In the forward process of Shift-DDPM, a noisy image is sampled as

$$x_t = \sqrt{\bar{\alpha}_t}x_0 + s_t^{(c)} + \sqrt{1 - \bar{\alpha}_t}\epsilon, \quad \epsilon \sim \mathcal{N}(0, \Sigma(c)), \quad (7)$$

where $s_t^{(c)} = k_t \cdot E(c)$ is the added shift term.

The objective loss of Shift-DDPM is defined as

$$\mathcal{L}(x_0, \theta) := \mathbb{E} \left[\left\| \frac{x_t - \sqrt{\bar{\alpha}_t}x_0}{\sqrt{1 - \bar{\alpha}_t}} - \epsilon_\theta(x_t, t) \right\|_{\Sigma^{-1}}^2 \right]. \quad (8)$$

Consequently, the model is trained to estimate only the Gaussian noise ϵ , and the shifted mean with variance is given to the model for the conditioning.

In the reverse process of Shift-DDPM, the initial noise is sampled class-wise as $X_T \sim \mathcal{N}(k_T \cdot E(c), \Sigma(c))$, then less noisy data is predicted as

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}} \epsilon_\theta(x_t, t) \right) + S_t^{(c)} + \sigma_t z, \quad (9)$$

where S_t is the mean shift term given by

$$S_t^{(c)} = -\frac{\sqrt{\bar{\alpha}_t}(1 - \bar{\alpha}_{t-1})}{1 - \bar{\alpha}_t} S_t^{(c)} + s_{t-1}^{(c)}, \quad s_t^{(c)} = k_t \cdot E(c), \quad (10)$$

which determines the trajectory of the data mean. Although $\Sigma(c)$ in Shift-DDPM is learnable, it is known to make the training process unstable and lead to unfavorable generation results. Thus, the fixed variance $\Sigma(c) = \mathbf{I}$ is used [51].

Shifted-Diffusion [52] introduces another data-driven conditional DDPM for text-to-image generation in which the forward process is defined as

$$x_t = \sqrt{\bar{\alpha}_t}x_0 + s_t^{(c)} + \sqrt{1 - \bar{\alpha}_t}\epsilon, \quad (11)$$

where Shifted-Diffusion uses $s_t^{(c)} := (1 - \bar{\alpha}_t)E(c)$, and the loss is defined as

$$\mathcal{L}(x_0, \theta) := \mathbb{E} \left[\|\epsilon - \epsilon_\theta(x_t, t)\|^2 \right], \quad (12)$$

with $\epsilon \sim \mathcal{N}(0, \Sigma(c))$. This implies that the prediction model (ϵ_θ) is trained to estimate not only the Gaussian noise but also the mean-shift imposed in each time step. We observed that this property increases the computational complexity of model training and severely degenerates the quality of the generated images.

While our proposed SDA framework does not depend on a specific choice of noise shift, based on these observations, we adopt Shift-DDPM [51] with the fixed $\Sigma(c) = \mathbf{I}$ as one of the baselines, and integrate it with data augmentation to design an auxiliary path in the forward diffusion process, which provides a strong regularization effect while avoiding leakage of the applied transformation in inference.

C. LEAKAGE OF DATA AUGMENTATION IN DIFFUSION MODELS

Data augmentation is a powerful tool in machine learning that enhances the diversity of training data by introducing transformation functions (e.g., horizontal flipping, 90° rotations, and hue changes) to improve the generalization of the trained model to unseen data.

In training DDPMs, one may apply transformations to the clean data x_0 during the forward diffusion process. However, unlike prediction tasks such as classification, there is a risk

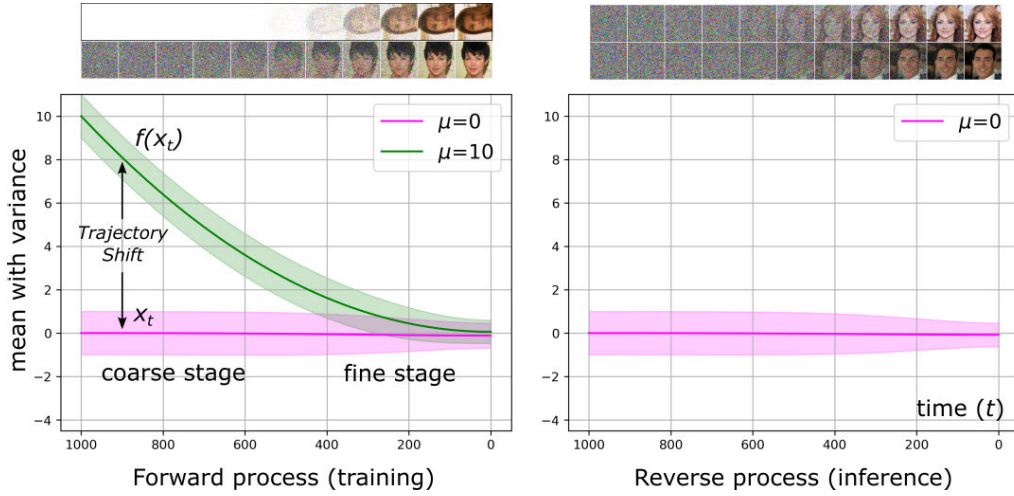


FIGURE 3. [The pipeline of the proposed shifted data-augmentation (SDA)](top) Sequence of the initial noise to clean images in (left) the forward process using data transformation, such as image rotation, and (right) the reverse process. (bottom) Their mean and variance across time step t (x-axis) in the auxiliary path (green line) and original path (magenta line). In the top-left image sequences, we visualize generated samples in the auxiliary path (1st row) and in the original path (2nd row). The noisy images in the auxiliary path appear bright due to high pixel intensities. We used parabolic shift with $\mu = 10$ and CelebA images [16] for this visualization.

of leakage when data transformation is applied in generative tasks as demonstrated earlier in Figure 2, where the DDPM produces images with unintended transformations that were applied to training data, such as horizontal flipping, rotation, and their combinations. Therefore, DDPMs typically adopt only horizontal flipping as data augmentation in practice [14].

IV. METHOD

A. PROPOSED SHIFTED DATA-AUGMENTATION (SDA)

We propose a regularization method, shifted data-augmentation (SDA), for training DDPM in unconditional image generation. SDA is based on a dual-path structure that leverages the regularization benefits of data augmentation while preventing the leakage problem, ensuring improved generalization without compromising the quality of generated images.

Figure 3 illustrates the pipeline of SDA. Our idea is to define an auxiliary diffusion path (green line) using transformed data in the forward diffusion process, in addition to the standard path using the untransformed original data (magenta line). Using the noise shift, we mix the two distributions of the original and transformed data in the fine stage where the model learns the detail of data, and also ensure a sufficient shift between the two diffusion trajectories in the coarse stage in order to prevent the leaking of data augmentation. In the inference, the initial noise is guided to the standard path due to the trajectory shift, and we obtain diverse data without the leakage. This SDA framework is controlled solely by the noise shift and does not need any modification of the model architecture.

Formally, the forward process of SDA follows a two-class conditional Shift-DDPM with data augmentation defined as

$$x_t^{(c)} = \sqrt{\bar{\alpha}_t} \mathcal{F}^{(c)}(x_0) + s_t^{(c)} + \sqrt{1 - \bar{\alpha}_t} \epsilon, \quad c \in \{0, 1\} \quad (13)$$

where $\epsilon \sim \mathcal{N}(0, \mathbf{I})$ and

$$s_t^{(c)} = k_t \cdot E(c), \quad E(0) = 0, \quad E(1) = \mu, \quad (14)$$

where μ is a hyper-parameter that determines the mean of the shifted trajectory. $\mathcal{F}^{(c)}$ transforms data (y) as

$$\mathcal{F}^{(c)}(y) = \begin{cases} y & \text{if } c = 0, \\ f(y) & \text{if } c = 1, \end{cases} \quad (15)$$

where f is a specific transform function, the class label $c = 0$ stands for the standard path using the original data, and $c = 1$ indicates the auxiliary path using transformed data. c is randomly assigned to each data in every training iterations with probability p .

To improve clarity, Eq. (13) can be reformulated using Eqs. (14) and (15) as

$$x_t^{(c)} = \begin{cases} \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1 - \bar{\alpha}_t} \epsilon & \text{if } c = 0, \\ \sqrt{\bar{\alpha}_t} f(x_0) + \mu \cdot k_t + \sqrt{1 - \bar{\alpha}_t} \epsilon & \text{if } c = 1, \end{cases} \quad (16)$$

with $\epsilon \sim \mathcal{N}(0, \mathbf{I})$, where $x_t^{(c=0)}$ denotes the standard diffusion process equivalent to Eq. (3). $x_t^{(c=1)}$ represents a noisy sample from the auxiliary path which follows Eq. (7), where the training data is first transformed using f .

Regarding the transformation probability p , the basic algorithm may assign labels c uniformly at random with a constant $p = 0.5$. However, as the diffusion time step (t) increases, the information of the transformation data decreases, making the transformation less meaningful. Therefore, we propose to apply the transformation with a higher probability p to more realistic data, i.e., lower t . Concretely, we use $p_t \sim \sqrt{\bar{\alpha}_t}$, following the proportion of clean data in the diffusion process.

The objective loss of SDA follows

$$\mathcal{L}(x_0, c, \theta) := \mathbb{E} \left[\left\| \frac{x_t^{(c)} - \sqrt{\bar{\alpha}_t} \mathcal{F}^{(c)}(x_0)}{\sqrt{1 - \bar{\alpha}_t}} - \epsilon_\theta(x_t^{(c)}, t) \right\|^2 \right] \quad (17)$$

where the same transformation (e.g., rotation angle) is applied to the clean image (x_0) in both $x_t^{(c)}$ using Eq.(13) and $\mathcal{F}^{(c)}(x_0)$ in Eq.(17) when $c = 1$, in order to guarantee the consistency of the information in the diffusion process. Eq.(17) can be rewritten as

$$\mathcal{L}(x_0, c, \theta) = \begin{cases} \mathbb{E} \left[\left\| \frac{x_t - \sqrt{\bar{\alpha}_t} x_0}{\sqrt{1 - \bar{\alpha}_t}} - \epsilon_\theta(x_t, t) \right\|^2 \right] & \text{if } c = 0, \\ \mathbb{E} \left[\left\| \frac{x_t - \sqrt{\bar{\alpha}_t} f(x_0)}{\sqrt{1 - \bar{\alpha}_t}} - \epsilon_\theta(x_t, t) \right\|^2 \right] & \text{if } c = 1. \end{cases} \quad (18)$$

Note that the $c = 0$ case reduces to the standard Shift-DDPM loss in Eq. (8), while the $c = 1$ case corresponds to a transformed sample x_t with noise shift using Eq. (16). Algorithm 1 shows the pseudo code of the training procedure using SDA.

In the reverse process, SDA samples the initial noise from the original distribution ($c = 0$) as $X_T \sim \mathcal{N}(0, \mathbf{I})$. Since $s_t^{(c=0)} = 0$ (i.e., $\mathcal{S}_t^{(c=0)} = 0$ from Eq.(10)) reduces Eq.(9) to Eq.(5). Therefore, as shown in Algorithm 2, SDA can reconstruct clean data using Eq.(5) or the standard diffusion-reverse process.

Algorithm 1 Training With SDA

1: **Input:** Diffusion steps T , noise scheduler $\bar{\alpha}_t$, shift scheduler k_t , shift magnitude μ , transform function $\mathcal{F}^{(c)}$
2: **Initialize:** Model parameter θ of noise estimator ϵ_θ
3: **repeat**
4: Given clean image x_0
5: Sample timestep $t \sim \{1, 2, \dots, T\}$ uniformly
6: Compute transformation probability $p_t := \sqrt{\bar{\alpha}_t}$
7: Sample class label $c \sim \text{Bernoulli}(p_t)$
8: Sample noise $\epsilon \sim \mathcal{N}(0, \mathbf{I})$
9: Compute shift $s_t^{(c)} := k_t \cdot c \cdot \mu$
10: Compute noisy image:
11: $x_t^{(c)} := \sqrt{\bar{\alpha}_t} \cdot \mathcal{F}^{(c)}(x_0) + s_t^{(c)} + \sqrt{1 - \bar{\alpha}_t} \cdot \epsilon$
12: Compute loss:
13: $\mathcal{L}(x_0, c, \theta) := \left\| \frac{x_t^{(c)} - \sqrt{\bar{\alpha}_t} \mathcal{F}^{(c)}(x_0)}{\sqrt{1 - \bar{\alpha}_t}} - \epsilon_\theta(x_t^{(c)}, t) \right\|^2$
14: Update θ via gradient descent
15: **until** Converged

B. IMPLEMENTATION OF SDA

The proposed SDA is characterized by the choice of the transform function (f) in Eq.(15), the mean scheduler (k_t), and the shift amount (μ) in Eq.(14).

Algorithm 2 Inference With SDA

1: Sampling $x_T \sim \mathcal{N}(0, \mathbf{I})$
2: **for** $t = T, \dots, 1$ **do**
3: $z \sim \mathcal{N}(0, \mathbf{I})$ if $t > 1$, else $z = 0$
4: $x_{t-1} = \frac{1}{\sqrt{\bar{\alpha}_t}} \left(x_t - \frac{1 - \bar{\alpha}_t}{\sqrt{1 - \bar{\alpha}_t}} \epsilon_\theta(x_t, t) \right) + \sigma_t z$
5: **end for**
6:
7: **return** x_0

Regarding transformation (f), SDA may adapt any transformation, such as scaling, flipping, or color jitter, as long as the consistency between the dual diffusion paths is preserved. We implement SDA using $\{90^\circ, 180^\circ, 270^\circ\}$ -rotations as its most basic and interpretable configuration. Image rotation has been widely used for data augmentation in conventional machine learning tasks [10], [27], making it a suitable initial instantiation to demonstrate the feasibility of SDA.

Unlike color transformations or cropping, image rotation preserves the full content of the image without introducing information loss, allowing for a clearer analysis of SDA's behavior. Moreover, this configuration allows the leakage behavior induced by rotation to be directly observed, further supporting its interpretability.

Regarding mean scheduler (k_t), Shift-DDPM [51] adopts any scheduler k_t and discusses three variants: prior-shift ($k_t = 1 - \sqrt{\bar{\alpha}_t}$), quadratic shift ($k_t = \sqrt{\bar{\alpha}_t}(1 - \sqrt{\bar{\alpha}_t})$), and data-normalization shift ($k_t = -\sqrt{\bar{\alpha}_t}$). As the data-normalization shift requires a modification of the model architecture for explicit conditioning, prior-shift and quadratic shift are practical choices.

However, k_t in the proposed SDA requires: $s_t^{(c=0)} \approx s_t^{(c=1)}$ in the fine stage in order to mix the distribution of the original and augmented data; and also $s_t^{(c=0)} \ll s_t^{(c=1)}$ in the coarse stage of the diffusion process to ensure a sufficient trajectory shift between the original and augmented data, guiding the successive generation of untransformed data in the reverse process. Therefore, we adopt $k_t := (t/T)^2$, referred to as the parabolic-shift, that satisfies these requirements. Figure 4 compares the standard path and the auxiliary path in SDA using prior-shift, quadratic-shift, and the proposed parabolic-shift.

Regarding the shift amount (μ), $\mu = 0$ generalizes the standard DDPM with the conventional data augmentation that can leak in the inference. Obviously, μ needs to be large enough to make the risk of leakage negligibly small. Since μ is a relative value with respect to the variance of the initial Gaussian noise $X_T \sim \mathcal{N}(0, \mathbf{I})$, it should satisfy $1 \ll \mu$. We observed that the proposed SDA is robust to this parameter and $\mu = 10$ is a practical choice. We verify these implementations in ablation studies using benchmarks in our experimental section.

C. ANALYSIS OF IMPLEMENTED SDA

We analyze the proposed SDA from a theoretical perspective, focusing on its behavior at the coarse and fine stages of

the diffusion process under the parabolic shift schedule ($k_t = (t/T)^2$). This analysis builds upon insights from Shift-DDPM [51] and Shifted-Diffusion [52] and clarifies the design rationales unique to our approach.

In the coarse stage, where $t \approx T$, the shift value k_t becomes large, and thus the mean shift $s_t^{(1)} = k_t \cdot \mu$ significantly separates the auxiliary trajectory ($c = 1$) from the standard trajectory ($c = 0$). This separation can also be theoretically justified by modeling the auxiliary and standard paths as Gaussian distributions with means $\mu \cdot k_t$ and 0, respectively. Following the approach in [52], the KL divergence between the two diffusion trajectories is

$$D_{\text{KL}}(\mathcal{N}(\mu \cdot k_t, \sigma^2) \| \mathcal{N}(0, \sigma^2)) = \frac{(\mu \cdot k_t)^2}{2\sigma^2} \approx \frac{(\mu)^2}{2\sigma^2} \quad (19)$$

When $\mu = 10$ and $\sigma^2 = 1$, this results in $D_{\text{KL}} = 50$, indicating that the two distributions are clearly distinguishable. Note that the initial noises on the standard path ($X_T \sim \mathcal{N}(0, 1)$) and the auxiliary path ($X_T \sim \mathcal{N}(\mu, 1)$) both evolve toward the distribution of x_0 through the diffusion process with the noise (mean) scheduler. Since the data distribution of x_0 is typically normalized (zero mean and unit variance), the choice of μ does not depend on the specific dataset. These behaviors are consistent with the findings of [51] and [52], where separate distributions in the latent space prevent the overlap of different trajectories. We emphasize that this property is a general feature of diffusion models with noise shift and not a unique contribution of our work.

In contrast, in the fine stage ($t \ll T$), the parabolic shift schedule ensures that $k_t \approx 0$, which makes the distance between the two diffusion trajectories as

$$D_{\text{KL}}(\mathcal{N}(\mu \cdot k_t, \sigma^2) \| \mathcal{N}(0, \sigma^2)) = \frac{(\mu \cdot k_t)^2}{2\sigma^2} \approx 0. \quad (20)$$

Also, substituting $s_t^{(1)} \approx 0$ into Eq. (13) yields

$$x_t^{(c)} = \begin{cases} \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon & \text{if } c = 0, \\ \sqrt{\bar{\alpha}_t}f(x_0) + \sqrt{1 - \bar{\alpha}_t}\epsilon & \text{if } c = 1. \end{cases} \quad (21)$$

Eq.(20) implies that the distributions of $x_t^{(c=0)}$ and $x_t^{(c=1)}$ on the standard and auxiliary paths, respectively, are virtually indistinguishable to the noise estimation model ϵ_θ . This enables the model to learn a richer representation from both the original and transformed data, as expressed in Eq.(21). These formulations in the fine stage of the diffusion process effectively approximate the behavior of baseline DDPMs trained with data transformation (f).

Importantly, in inference, the proposed SDA effectively guarantees that the parameter of data transformation (i.e., the angle in the image rotation) has negligible effect on reconstructed data. Specifically, our method benefits from two complementary properties: at the coarse stage, Eq.(19) ensures that the distributions are sufficiently separated to avoid leakage of augmented information during inference, and thus the initial noise on the standard path is reconstructed toward the untransformed data.

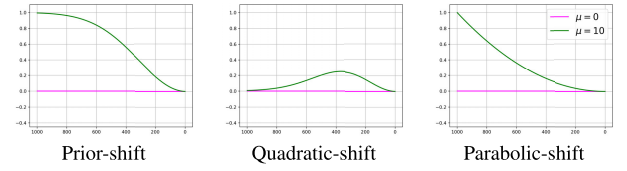


FIGURE 4. [Visualization of noise shift schedulers (k_t) under the SDA scheme.] (y-axis) the mean-shift trajectories for (magenta) the original data and (green) augmented data over (x-axis) diffusion steps (t), using (1st, 2nd, 3rd columns) prior-shift, quadratic-shift, and the proposed parabolic-shift.

In contrast, Eq.(20) suggests that, at the fine stage, the reconstructed image is further refined using the model at $t \ll T$, which reflects the distribution of both original and augmented data, and is thus expected to contribute to richer reconstructions during inference.

Therefore, SDA integrates the benefits of data augmentation during training while avoiding the influence of the applied transformation during the coarse stage of inference. This unique property cannot be achieved by either of our two baselines: the conventional rotational augmentation and DDPMs with shift techniques.

D. OBSERVATION ON THE DATA MIXTURE EFFECT BY NOISE-SHIFT

The proposed SDA is based on a key property of the noise shift in the DDPM training. Specifically, we observe that the noise shift enables [i] mixing of the distributions of the original training examples with augmented data, thereby introducing [ii] a strong regularization effect in training while ensuring that non-augmented data are generated during inference.

To first visualize property [i], we conduct an empirical analysis using synthetic data as follows. Suppose we have a set of *red* images with small Gaussian noise perturbations as the original training examples. We then apply a channel rotation to transform these images into *blue*, creating the augmented examples. Given the red images and the augmented blue images, we train the baseline DDPM using a noise shift with mean $\mu = \{0, 1, 10, 20\}$, where $\mu = 0$ represents the baseline DDPM with conventional data augmentation, and larger μ values are expected to enhance the mixing of the original and augmented sets. We generate images using the trained models during the training process and compute the channel-wise intensity of the generated images. Figure 5 shows the channel-wise intensity curves of the generated images over the training epochs, and Figure 6 visualizes examples of generated images during training.

1) LEAKING BEHAVIOR OF THE BASELINE

As demonstrated in Figure 5, $\mu = 0$ results in oscillations between the red and blue channels, indicating that the augmented data, i.e., blue images, are leaked into the generated outputs. In this case, the generated data consist of red and blue images with approximately equal probability, failing to preserve the original distribution.

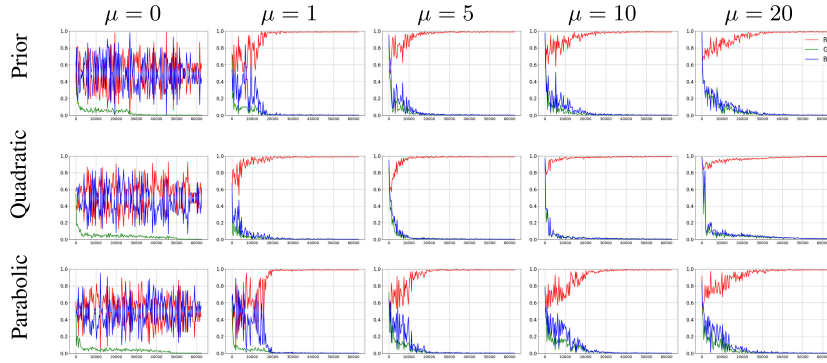


FIGURE 5. [Observation on the mixture effect due to noise-shift](y-axis) (red, blue, green lines) RGB channel-wise intensities of generated data in (x-axis) the training iterations, using DDPMs with (1st, 2nd, 3rd rows) prior-shift, quadratic shift, and parabolic shift with (1st, ..., 5th columns) different the initial noise mean μ .

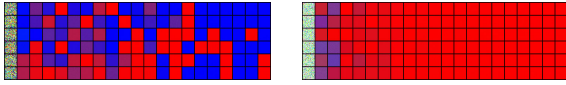


FIGURE 6. [Observation on the mixture effect due to noise-shift](1st, ..., 6th rows) Generated samples at (1st, ..., 21-th columns in each part) 0%, 5%, 10%, ..., 100%-epochs in training process, using DDPMs with SDA based on parabolic-shift with (left) $\mu = 0$ and (right) $\mu = 10$, respectively.

2) MIXTURE EFFECT BY THE NOISE SHIFT

With $\mu = 1$, the red and blue channels are mixed in the early training phase. Note that this is not due to an annealing of data augmentation over the training epochs, but an inherent property of SDA that emerges as the model learns the standard path with the original data and the auxiliary path with the augmented data. This controlled mixture of distributions is expected to introduce a regularization effect not only in the early phase but throughout the entire training process, thereby improving generalization effectively.

3) ROLE OF THE NOISE SCHEDULER

As we discussed, $s_0^{(c=1)} = k_0 \cdot \mu$ must be greater than the noise variance (i.e., 1) in order to guide the initial noise toward the standard diffusion path in the reverse process. However, as demonstrated in the 1st and 2nd rows in Figure 5, the prior-shift and quadratic-shift with a larger μ reduce the overlap between the two distributions in the early training phase, which may weaken the intended regularization effect. In contrast, the proposed parabolic-shift (shown in the 3rd row in Figure 5) successfully maintains the mixture effect between the two diffusion processes that appears in the early training phase, regardless of μ . Although the proposed SDA is robust to the choice of noise schedulers, we recommend the parabolic-shift with $\mu = 10$ as a practical choice, as it achieves satisfactory results as demonstrated in the experimental section.

It was reported that Shift-DDPM slightly improves model capability [51], but the mechanism behind it was not clearly

explained. Here, we visualized the distribution mixture effect of the noise shift using synthetic data.

Regarding [ii], the regularization effect introduced by the proposed SDA integrating the noise shift and data augmentation, we present qualitative and quantitative results in the following section.

V. EXPERIMENTAL RESULTS

A. EXPERIMENTAL SET-UP

We demonstrate the effectiveness and the robustness of the proposed SDA (shifted data-augmentation) through experiments under small- and large-scale scenarios using the major datasets. We also conduct a series of ablation experiments to further investigate the property of the proposed method.

1) SCENARIO

We conduct two scenarios: limited-data training and full-data training. In the limited-data scenario, we train the model using small subsets ($N = \{1,024, 512, 256\}$) of the available training examples, which allow us to evaluate the regularization effect of the proposed SDA in mitigating the memorization problem observed in the baseline DDPM. In the full-data scenario, we train the model using all available data to assess the quality of the generated images with SDA in comparison to state-of-the-art generative models. By considering both scenarios, we aim to demonstrate the effectiveness of SDA in reducing over-fitting in limited-data regimes while maintaining high-quality generation in the large-scale training.

2) DATASETS

We use three datasets for our experiments: CelebA [16], AFHQv2-Cats [57], and Oxford-Flowers [58] in order to validate the effectiveness of the proposed SDA across diverse data distributions, variations, and image resolutions. CelebA consists of approximately 200K images of highly structured and well-aligned human faces at a resolution of 128×128 pixels. This large-scale dataset serves as a primary

TABLE 1. [Limited-data training with $N = \{1,024, 512, 256\}$ examples] (1st ..., 4th rows in each part) Training FID, Test FID, Inception Score, and Replicate Percentage (RP) of generated images ($n = 10,000$) generated using DDPM with horizontal-flipping (Baseline), Baseline with early stopping where the model was selected just before Replicate Percentage exceeded zero (0%) (Early Stop), and Baseline with the proposed shifted data-augmentation (SDA). Models were trained with $N = \{1,024, 512, 256\}$ examples from CelebA, AFHQv2-Cats, and Oxford-Flowers datasets. Note that when the Replicate Percentage exceeds zero, Training/Test FIDs and IS no longer reliably reflect generalization.

$N = 1,024$		CelebA			AFHQv2-Cats			Oxford-Flowers		
Algorithm	Baseline	Early Stop	SDA	Baseline	Early Stop	SDA	Baseline	Early Stop	SDA	
Training FID (↓)	6.90	42.54	36.79	4.77	21.08	17.38	8.35	43.84	27.72	
Test FID (↓)	12.57	36.97	25.97	8.91	21.34	15.34	34.03	51.87	33.26	
Inception Score (↑)	3.39	2.71	2.94	1.44	1.65	1.58	2.78	2.90	2.18	
Replicate Percentage (ideally zero)	33.73%	0.00%	0.00%	15.98%	0.00%	0.00%	38.82%	0.00%	0.00%	
$N = 512$		CelebA			AFHQv2-Cats			Oxford-Flowers		
Algorithm	Baseline	Early Stop	SDA	Baseline	Early Stop	SDA	Baseline	Early Stop	SDA	
Training FID (↓)	8.75	62.71	48.63	6.28	35.60	23.74	7.63	49.24	41.20	
Test FID (↓)	19.42	49.46	29.05	9.62	27.07	13.67	47.64	56.29	50.46	
Inception Score (↑)	2.28	2.28	2.18	1.49	1.68	1.58	2.55	2.84	2.23	
Replicate Percentage (ideally zero)	41.37%	0.00%	0.00%	20.39%	0.00%	0.00%	51.96%	0.00%	0.00%	
$N = 256$		CelebA			AFHQv2-Cats			Oxford-Flowers		
Algorithm	Baseline	Early Stop	SDA	Baseline	Early Stop	SDA	Baseline	Early Stop	SDA	
Training FID (↓)	8.62	82.69	42.34	6.15	49.24	42.75	6.88	53.48	57.25	
Test FID (↓)	28.52	61.82	24.91	17.14	56.29	26.24	87.07	89.66	74.17	
Inception Score (↑)	3.11	3.45	1.82	1.82	2.84	1.78	2.19	2.27	1.85	
Replicate Percentage (ideally zero)	51.60%	0.00%	0.00%	49.20%	0.00%	0.00%	52.00%	0.00%	0.00%	

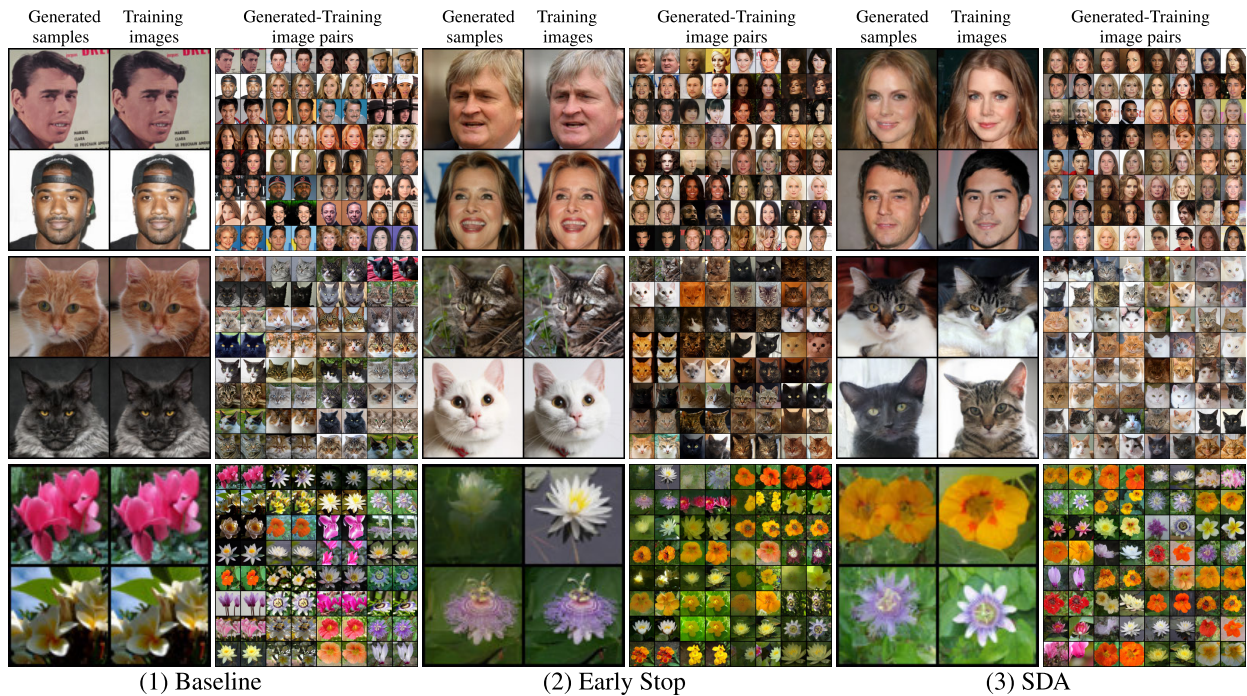


FIGURE 7. [Limited-data training with $N = 1,024$ examples](odd columns in each image grid) Generated samples and (even columns) their nearest training images based on the cosine similarity using (1) DDPM with the horizontal-flipping (Baseline), (2) Baseline with early stopping (Early Stop), and (3) Baseline with SDABased on (1st, 2nd, 3rd row parts) CelebA, AFHQv2-Cats, and Oxford-Flowers datasets: Each image grid consists of 4×8 samples, where odd-numbered columns show generated samples and even-numbered columns show their nearest training images. These nearest neighbors are selected from both the original N training images and the augmented N training images. The 8×8 image grids in the 2nd and 4th columns visualize these selected nearest neighbors. The top-left 2×2 images are highlighted in the 1st and 3rd column parts. For the early stopping, we used the baseline model obtained just before Replicate Percentage exceeded 0 in the training process.

benchmark for evaluating generative models. AFHQv2-Cats contains 5K images of cats at 128×128 resolution. Compared to CelebA, this dataset introduces pose variation and textures, making the image generation task more challenging. Oxford-Flowers includes about 8K images at 64×64 resolution and has intricate color patterns and diverse shapes.

3) MODELS

We employ a standard U-Net-based DDPM implemented using the Diffusers library from Hugging Face [59], the de-facto standard framework for diffusion models which ensures reproducibility and compatibility with existing research. Since SDA with image rotation intro-

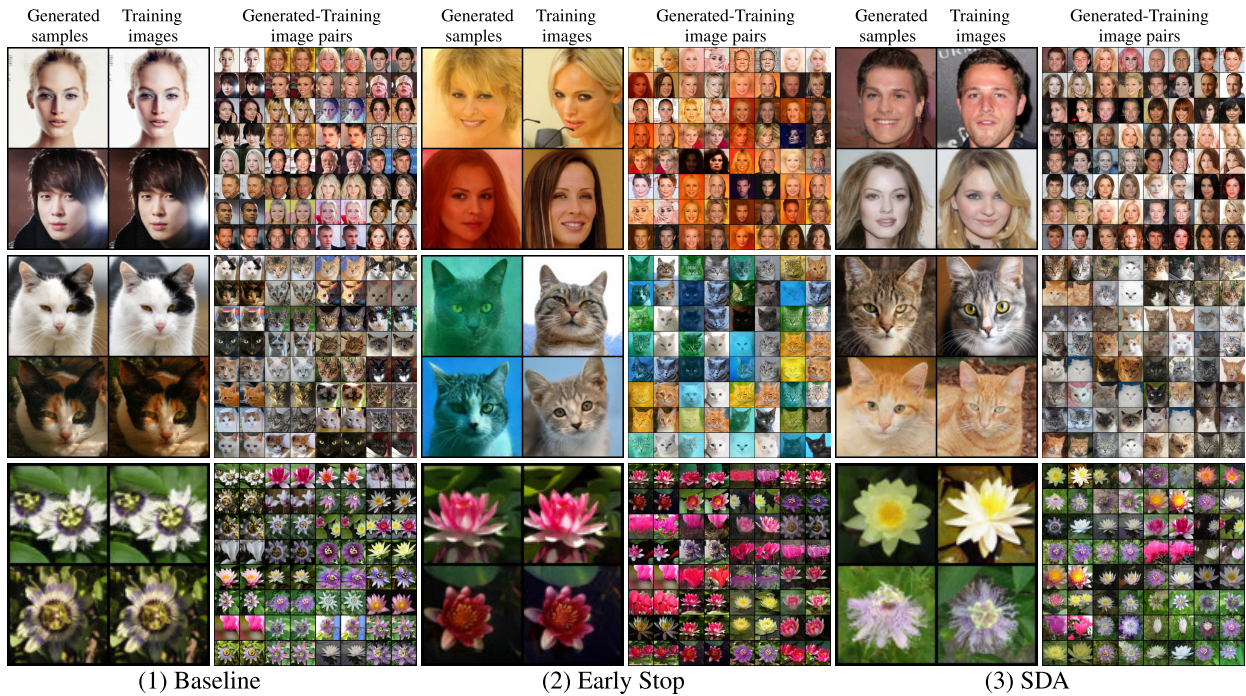


FIGURE 8. [Limited-data training with $N = 512$ examples](odd columns in each image grid) Generated samples and (even columns) their nearest training images based on the cosine similarity using (1) DDPM with the horizontal-flipping (Baseline), (2) Baseline with early stopping, and (3) Baseline with SDA based on (row parts) CelebA, AFHQv2-Cats, and Oxford-Flowers.

duces negligible additional computational and memory cost, its computational requirements remain essentially the same as those of the Hugging Face baseline implementation.

We treat DDPM with horizontal-flipping data augmentation that is recommended [14] as our baseline (denoted as ‘Baseline’ in this paper). We implement our algorithm (denoted as Ours) by incorporating the proposed SDA using the image rotations with $\{90^\circ, 180^\circ, 270^\circ\}$ into the baseline. To evaluate the regularization effect of SDA, we compare the baseline and our algorithm in the limited-data training scenario. Furthermore, we include vanilla DDPM (without data augmentation) and StyleGAN3 [11] as state-of-the-art generative models in the final experiment using the full-data for a comprehensive comparison.

4) TRAINING

In the training of DDPM, we use the common configurations (e.g., Adam-W [60], a learning rate of 0.0001, and the dropout probability of 0) recommended in Hugging Face, along with standard hyper-parameters for DDPMs (e.g., the number of diffusion steps $T = 1,000$ and the linear variance scheduler $\{\beta_t\}$). We train the models for 10,000 epochs to ensure that all tested algorithms can converge and select the best model based on the FID score during the training process so that our experimental results are not dependent on the epoch size. The total number of training epochs was fixed across all experimental conditions. StyleGAN3 is also trained using the recommended conditions [11].

5) EVALUATION METRICS

For quantitative evaluation of generated images, we employ two widely used metrics in the image generation task: the inception score (IS) [61] and the Fréchet inception distance (FID) [62]. IS measures the diversity of generated images in the feature space defined using a pre-trained network. FID measures the quality of generated images based on the distance between a training set with a generated image set in the feature space. Lower FID values and higher IS values indicate better quality and diversity of the generated data, respectively. Note that IS does not directly reflect image quality and can appear higher (i.e., better) even in the early stage of training. FID is the primary criterion, as it is widely recognized as the standard evaluation metric for image generation tasks [62], [63], [64], [65], [66], [67], and IS is only considered when FID values are comparable. Therefore, we primarily rely on FID in our experiments.

In the limited-data scenario, we compute both FID using the training set (denoted as Training FID) and FID using the remaining complementary test set (Test FID). We track FID (Training FID) throughout the model training process and select the final model based on the best Training FID score. Other metrics are computed using the final models. The metrics were averaged over three individual trials.

We also employ Replicate Percentage [42] to evaluate the extent of memorization. This is particularly important because FID and IS become unreliable if the generated images are mere copies of training examples, i.e., due to the memorization problem [42], [43], [44] in the image generation task, which is especially problematic when

the number of training samples is small. The Replicate Percentage measures the proportion of generated samples $X = \{x^{(1)}, x^{(2)}, \dots, x^{(N)}\}$ that satisfy $\frac{\|x - y_1\|_2}{\|x - y_2\|_2} < \frac{1}{3}$, where $x \in X$ is a generated sample, and y_1 and y_2 are the nearest and second-nearest images in the training set $\{y^{(1)}, y^{(2)}, \dots, y^{(N)}\}$ based on the L2 distance in the feature space. The threshold $\frac{1}{3}$ is a predefined constant, and a generated image is considered a copy of a training example (y_1) if it lies much closer to y_1 than to any other sample in the training set. Ideally, the Replicate Percentage should be zero in generative modeling tasks.

B. RESULTS ON LIMITED-DATA TRAINING

In this scenario, we train the baseline DDPM and DDPM with SDA using small subsets ($N = \{1,024, 512, 256\}$) of the available training examples and select the final models based on the best Training FID in the training process. In addition, we evaluate the baseline with early stopping, where the model is selected just before the Replicate Percentage exceeds zero (0%) during training.

Table 1 summarizes the evaluation metrics of the baseline DDPM, those with the early stopping, and SDA trained using different numbers of training examples ($N = 1,024, 512, 256$). Figure 7, Figure 8, and Figure 9 visualize pairs of generated images with their nearest training images based on the cosine similarity, for the $N = 1,024$, $N = 512$, and $N = 256$ cases, respectively. In these figures, we chose the nearest training example for each generated sample based on the cosine similarity, where the cosine similarity between a generated sample x and a training sample y is defined as $\text{CosineSim}(x, y) = (x \cdot y) / (\|x\|_2 \cdot \|y\|_2)$. This metric directly captures the pixel-wise visual structure of individual images, while being more robust to global intensity variations than the L2 distance, making it suitable for identifying the nearest neighbors in terms of perceptual similarity.

As shown in Table 1, the baseline yielded better image quality than the proposed SDA in most cases, as measured by Training FID, Test FID, and Inception Score. However, as indicated by Replicate Percentage in Table 1 and illustrated in Figures 7, 8, 9, the baseline DDPM suffers from the memorization problem, generating near-duplicate copies of individual training images that were memorized in the training process. In such cases, any image quality metric, including Test FID, becomes unreliable and loses its meaning. These results highlight the critical drawback of training baseline DDPMs under limited-data conditions. Moreover, for the baseline DDPM, there is a trade-off between the number of training iterations and the risk of memorization.

Therefore, we further focus on the baseline with early stopping, which ensures a Replicate Percentage of zero. Recall that Training FID is calculated from a limited number of training examples (compared to $N = 10,000$ generated samples) and may not fully reflect generalization performance in the limited-data scenario. Therefore, Test FID serves as a more reliable indicator of generation quality.

TABLE 2. [Ablation study on the mean shift (μ)] (1st, ..., 4th rows) Training FID, Test FID, Inception Score, and Replicate Percentage of generated samples ($n = 10,000$) created by the proposed SDA using the parabolic-shift with (1st, ..., 4th columns) varying mean shift values (μ): The models were trained using $N = 1,024$ examples from the CelebA dataset. $\mu=10$ is adopted as a practical setting.

Mean shift amount	$\mu=1$	$\mu=5$	$\mu=10$	$\mu=20$
Training FID ($\downarrow$)	39.81	40.53	36.79	41.41
Test FID ($\downarrow$)	29.03	29.69	25.97	30.81
Inception Score ($\uparrow$)	3.25	3.13	2.94	2.91
Replicate Percentage	0%	0%	0%	0%

Importantly, as shown in Table 1, the proposed SDA consistently outperforms the early-stopping baseline in terms of Test FID, demonstrating better generalization. In addition, as visualized in Figures 7, 8, 9, the image quality of the early-stopping baseline significantly degrades as the number of training examples decreases. Beyond the early-stopping point, the baseline continues to exhibit decreasing FID as training progresses, but this apparent improvement is misleading, as it results from copying the individual training examples. In contrast, DDPM with SDA successfully produced diverse and novel images due to its strong regularization effect, even in the extreme case where only $N = 256$ examples are available, as illustrated in Figure 9.

These empirical results under the limited-data conditions demonstrate that our SDA effectively mitigates memorization in DDPMs through its regularization effect due to the image rotation, while achieving superior image quality relative to the baseline DDPM with early stopping.

C. ABLATION STUDIES

We perform ablation studies in a greedy manner about the shift amount (μ) and the shift scheduler (k_t) in Eq.(14) in order to understand these components that contribute to the effectiveness of the proposed SDA. To evaluate models based on both Training FID and Test FID, we use $N = 1,024$ training examples from CelebA dataset as a case study.

1) RESULTS ON MEAN SHIFT AMOUNT (μ)

Table 2 summarizes the evaluation metrics for the proposed SDA using the parabolic-shift with varying values of $\mu = \{1, 5, 10, 20\}$. As discussed in Section IV-B, μ should be sufficiently large to prevent the auxiliary path from leaking into the original path. However, an excessively large μ may hinder the convergence of the algorithm. The results in Table 2 indicate that $\mu = 10$ yields the best performance in both Training FID and Test FID, while maintaining a zero Replicate Percentage. This indicates that $\mu = 10$ provides the best generalization performance among the tested μ values.

2) RESULTS ON MEAN SHIFT SCHEDULER (K_T)

In the implemented SDA, the parabolic-shift is adopted to align with the design principle discussed in Section IV-B. Table 3 summarizes the evaluation metrics using three shift

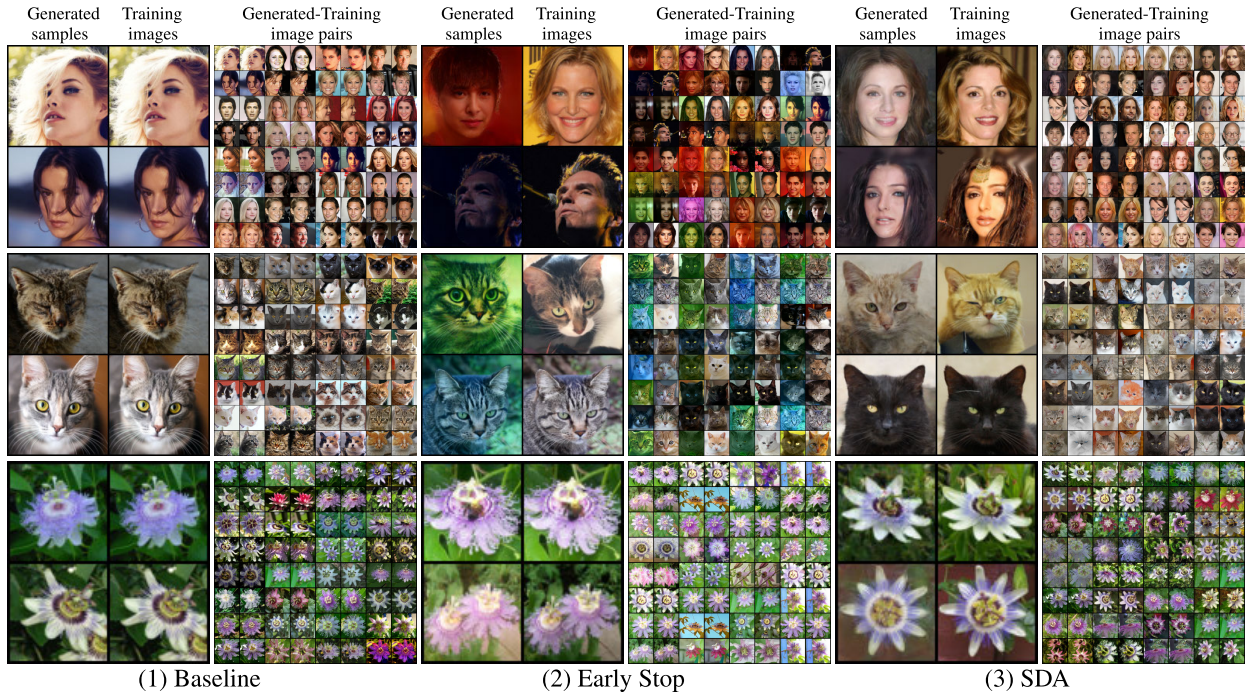


FIGURE 9. [Limited-data training with $N = 256$ examples](odd columns in each image grid) Generated samples and (even columns) their nearest training images based on the cosine similarity using (1) DDPM with the horizontal-flipping (Baseline), (2) Baseline with early stopping, and (3) Baseline with SDA based on (row parts) CelebA, AFHQv2-Cats, and Oxford-Flowers.

TABLE 3. [Ablation study on the shift scheduler (k_t)] (1st, ..., 4th rows) The metrics of generated images ($n = 10,000$) created by the proposed SDA using (1st, 2nd, 3rd columns) different mean schedulers with fixed $\mu = 10$: The models were trained using $N = 1,024$ examples from the CelebA dataset.

Shift scheduler	prior-shift	quadratic-shift	parabolic-shift
Training FID ($\downarrow$)	43.35	37.87	36.79
Test FID ($\downarrow$)	32.98	28.03	25.97
Inception Score ($\uparrow$)	2.92	2.67	2.94
Replicate Percentage	0%	0%	0%

schedulers: prior-shift, quadratic-shift, and parabolic-shift, visualized in Figure 4, with fixed $\mu = 10$. As shown in Table 3, the proposed parabolic-shift achieved better results across all metrics compared to the conventional prior-shift and quadratic-shift.

Taking these observations into account, we select the parabolic-shift with $\mu = 10$ as our recommended setting. Recall that μ is a relative value with respect to the variance of the initial noise $X_T \sim \mathcal{N}(0, I)$ and thus is independent of the training data distribution. Therefore, tuning μ for different target datasets is unnecessary, making the implemented SDA a practical and robust algorithm across datasets.

3) RESULTS OF SDA WITHOUT TRANSFORMATION (F)

We further demonstrate that the regularization effect originates from the applied transformation rather than from the noise shift originally introduced in Shift-DDPM [51]. To this end, we remove the image rotation f from SDA,

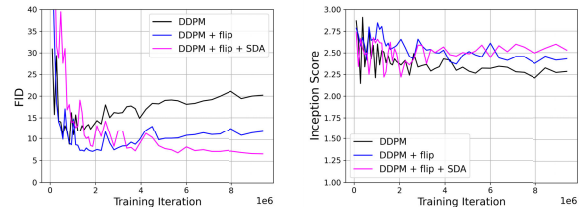


FIGURE 10. [Visualization of regularization effect] (left and right) FID ($\downarrow$) and IS ($\uparrow$) process of generated images ($n = 10,000$) using (black) DDPM, (blue) DDPM with horizontal flip (the baseline), and (magenta) the baseline with the proposed SDA trained for CelebA dataset.

where the original training images are fed into the prediction model via the standard diffusion path or the auxiliary path probabilistically during training.

Table 4 summarizes the training metrics of the baseline DDPM, the baseline with noise shift, and the proposed SDA. The results show that, for unconditional image generation, the noise shift itself does not yield a regularization effect in the absence of a data transformation. Combined with the leakage issue caused by conventional image rotation (i.e., rotation artifacts appearing in generated samples) as shown in Figure 2, this result underscores the key contribution of SDA: its regularization benefit cannot be achieved by the noise shift technique alone nor by conventional rotational augmentation.

D. RESULTS ON FULL-DATA TRAINING

We also evaluate our algorithm under the full-data scenario using all the available images in each dataset. We compare

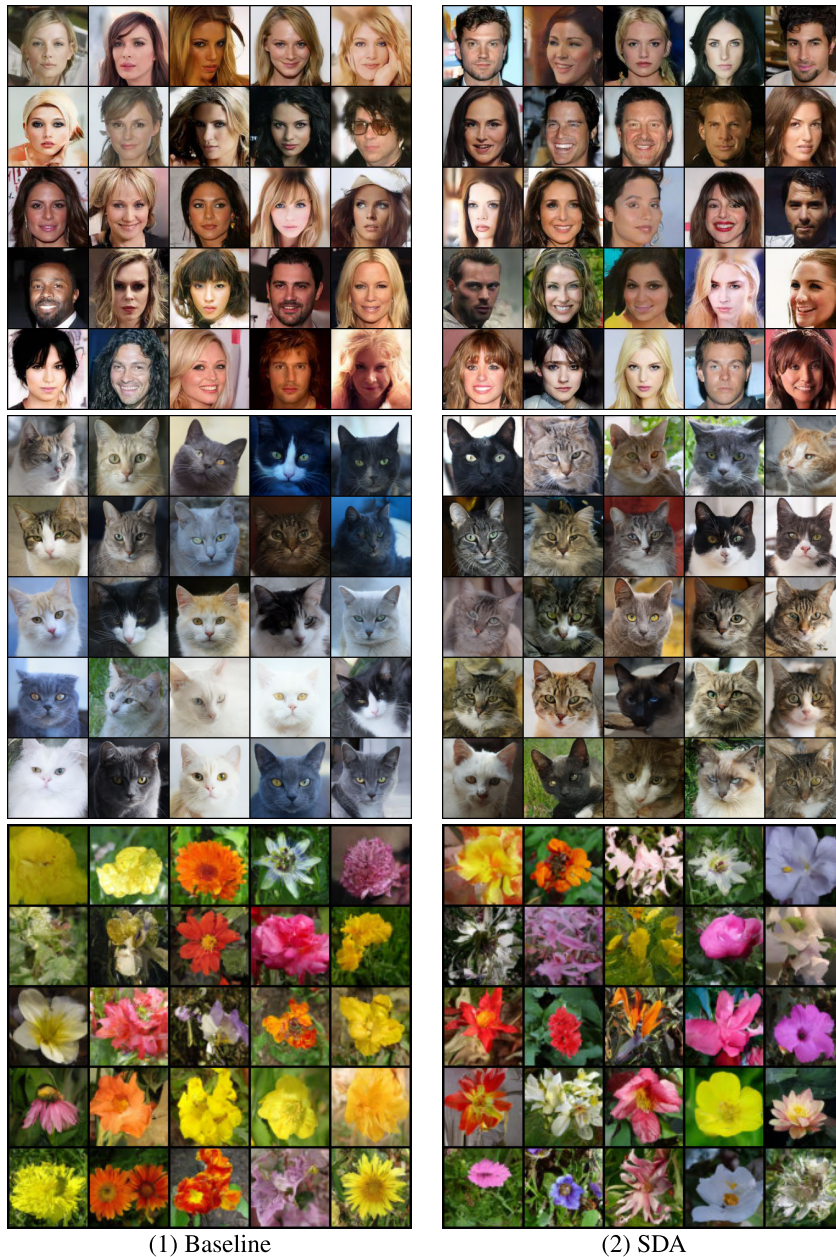


FIGURE 11. [Full-data training] Generated 5×5 images with the best FID in the training process using (left) DDPM with the horizontal-flipping data augmentation (Baseline), and (right) Baseline with the proposed SDA, based on (1st, 2nd, 3rd row parts) CelebA, AFHQv2-Cats, and Oxford-Flowers datasets, respectively.

five algorithms: StyleGAN3 [11] as the state-of-the-art GAN, standard DDPM without data augmentation, DDPM with horizontal-flipping data augmentation as the conventional baseline, and the same baseline combined with the proposed SDA (Ours). We also test DDIM [45], where we applied the DDIM inference with the sampling step of $\tau = 100$ to the baseline DDPM model trained with horizontal-flipping. In this experiment, results are presented based on a single trial due to the extremely high computational cost.

Table 5 summarizes the FID and Inception Score computed from the images generated by the tested algorithms on

CelebA, AFHQv2-Cats, and Oxford-Flowers, where we confirmed that the Replicate Percentage values of the tested algorithms were zero (0%) across the datasets, and no memorization of training examples was observed in this full-data training scenario. Figure 11 visualizes generated images using the baseline and the proposed SDA. Additionally, we visualize the training process of FID and IS for the three algorithms: DDPM, DDPM with flip, and SDA in Figure 10, using CelebA as an instance.

Figure 10 shows that the standard DDPM without augmentation (black line) exhibits an early drop in FID,

TABLE 4. [Ablation study on SDA without transformation (f), i.e., noise shift only] (rows) The metrics of generated images ($n = 10,000$) created using (columns) DDPM with horizontal flip (Baseline), Baseline with noise shift (Noise Shift), Noise Shift with the early stopping, and SDA using image rotation: The models were trained using $N = 1,024$ examples from the CelebA dataset.

Algorithm	Baseline	Noise Shift	SDA
DDPM with flip	✓	✓	✓
noise shift		✓	✓
transformation			✓
early stopping		✓	
Training FID ($\downarrow$)	6.90	6.98	41.29
Test FID ($\downarrow$)	12.57	12.51	34.22
Inception Score ($\uparrow$)	3.39	3.28	2.66
Replicate Percentage	33.74%	33.92%	0.00%

TABLE 5. [Full-data training] (top part) FID and (bottom part) Inception Score of generated images ($n = 10,000$) using (1st, ..., 5th columns) StyleGAN3, DDIM ($\tau = 100$) with horizontal-flipping data augmentation (flip), standard DDPM without data augmentation, DDPM with flip (Baseline), and Baseline with the proposed SDA based on (1st, 2nd, 3rd row parts) CelebA, AFHQv2-Cats, and Oxford-Flowers datasets. The reported FID corresponds to the best score achieved during training, and the IS is measured at that same training epoch.

FID ($\downarrow$)	StyleGAN3	DDIM + flip	DDPM	(Baseline)	
				DDPM + flip	DDPM + flip + SDA
CelebA	8.29	7.59	8.69	7.05	6.70
AFHQv2-Cats	9.32	6.99	12.34	6.08	4.32
Oxford-Flowers	16.49	20.69	39.64	16.93	15.15

FID ($\downarrow$)	StyleGAN3	DDIM + flip	DDPM	DDPM + flip	DDPM + flip + SDA
CelebA	2.93	3.19	2.15	2.27	2.16
AFHQv2-Cats	1.94	2.00	1.71	1.63	1.62
Oxford-Flowers	3.49	4.11	3.61	2.62	2.27

followed by a gradual rise, indicating that the model falls into a local minimum and subsequently begins to overfit. By contrast, the horizontal flip (blue line) helps the model escape from such local minima and stabilizes the FID curve, demonstrating a typical regularization effect of the simple data augmentation. Furthermore, SDA (magenta line) allows us to incorporate stronger augmentations, such as image rotation in our implementation, without the leakage. This further stabilizes the training dynamics and leads to improved accuracy, as reflected in a flatter and lower FID curve, while also enhancing the diversity of generated images, as indicated by the IS curves.

The results in Table 5 demonstrate that the proposed SDA improved FID across all tested datasets, which feature diverse image contents, shape complexities, and color variations. This consistent improvement indicates that the benefit of SDA is not limited to regularizing models trained on scarce data but also extends to enhancing generation quality under large-scale training conditions. Notably, the FID gains are observed not only on highly structured datasets like CelebA but also on more challenging domains such as

AFHQv2-Cats and Oxford-Flowers, where fine-grained textures and color diversity make generation more difficult. These results suggest that the proposed SDA serves as an effective and general-purpose regularization framework for diffusion models, capable of improving robustness and sample quality across varied data regimes and visual complexities.

Additionally, we included the results of DDIM and StyleGAN3 in Table 5 for reference. Since DDIM is an efficient sampling method, its generation quality is generally lower than that of DDPM. Note also that StyleGAN3 is not directly comparable to SDA, as it differs in architecture, loss, and training procedure.

VI. CONCLUSION

We proposed a regularization method, shifted data-augmentation (SDA), for training diffusion models in unconditional image generation. In the model training, SDA creates the auxiliary diffusion path with transformed data using the noise-shift method, alongside the standard diffusion path with the original data. The distribution mixture effect of the noise-shift mitigates the over-fitting of the diffusion models. The noise shift between the two diffusion paths guides the initial noisy images in inference toward the original distribution, enabling the generation of diverse and novel images without the leakage of augmented data.

We observed the distribution mixture effect of the noise-shift on synthetic data. By leveraging this effect, we proposed SDA and presented its implementation using the image rotation. We presented qualitative and quantitative results based on major datasets, in which SDA yielded significant improvements in the limited-data training, while also showing improvement over the baseline in the large-scale training.

A limitation of SDA is its assumption that the transformation in the secondary diffusion pass (i.e., image rotation) contributes to a data augmentation effect during training. We observed that the implemented SDA with image rotation did not lead to accuracy improvements on the LSUN datasets [68]. However, this is consistent with the earlier findings in [14], which reported that even the conventional horizontal flipping fails to enhance performance on this dataset. Given that LSUN contains images with high variability in orientation and spatial composition, it is natural that the spatial transformations do not function as effective augmentations. Importantly, this limitation arises from the specific transformation (i.e., rotation) used in our implementation, not from the SDA framework itself. SDA remains a general framework for incorporating various augmentations into diffusion models in a leakage-free manner.

Regarding the time complexity, the use of image rotation in SDA introduces negligible computational overhead compared to the baseline DDPM. Nevertheless, we observed that the number of training iterations required for FID convergence tends to increase. This reflects a common trade-off in regularization techniques, as also seen with

horizontal flipping, where improved generalization comes at the cost of slower convergence. While this may increase the overall training time, it enables a more thorough exploration of the solution space, potentially leading to better global optima.

Finally, as synthetic image generation becomes increasingly applicable across diverse domains, it is important to remain mindful of ethical considerations, particularly in sensitive areas such as facial imagery and medical data. These include potential misuse of generated content and the necessity of promoting its responsible deployment.

REFERENCES

- [1] C. Tian, X. Zhang, Q. Zhu, B. Zhang, and J. C.-W. Lin, "Generative adversarial networks for image super-resolution: A survey," 2022, *arXiv:2204.13620*.
- [2] J. Oh, D. Yoon, and I. Kim, "One-shot ultra-high-resolution generative adversarial network that synthesizes 16K images on a single GPU," *Image Vis. Comput.*, vol. 139, Nov. 2023, Art. no. 104815.
- [3] J. Song, T. He, L. Gao, X. Xu, A. Hanjalic, and H. T. Shen, "Unified binary generative adversarial network for image retrieval and compression," *Int. J. Comput. Vis.*, vol. 128, pp. 2243–2264, Sep. 2020.
- [4] Q. Liu, Q. Zhang, W. Liu, W. Chen, X. Liu, and X. Wang, "WSDS-GAN: A weak-strong dual supervised learning method for underwater image enhancement," *Pattern Recognit.*, vol. 143, Nov. 2023, Art. no. 109774.
- [5] X. Guo, Z. Wang, Q. Yang, W. Lv, X. Liu, Q. Wu, and J. Huang, "Gan-based virtual-to-real image translation for urban scene semantic segmentation," *Neurocomputing*, vol. 394, pp. 127–135, Jun. 2020.
- [6] J. Li, J. Dong, S. Huang, X. Li, J. Jiang, X. Fan, and Y. Zhang, "Virtual immunohistochemistry staining for histological images assisted by weakly-supervised learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 11259–11268.
- [7] Y. Chen, J. Xu, Z. An, and F. Zhuang, "Multi-scale conditional reconstruction generative adversarial network," *Image Vis. Comput.*, vol. 141, Jan. 2024, Art. no. 104885.
- [8] W. Lim, K. Y. S. Chek, L. B. Theng, and C. T. C. Lin, "Future of generative adversarial networks (gan) for anomaly detection in network security: A review," *Comput. Secur.*, vol. 139, Apr. 2024, Art. no. 103733.
- [9] 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.*, 2014, pp. 2672–2680.
- [10] 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.*, 2020, pp. 12104–12114.
- [11] T. Karras, M. Aittala, S. Laine, E. Härkönen, J. Hellsten, J. Lehtinen, and T. Aila, "Alias-free generative adversarial networks," in *Proc. NeurIPS*, 2021, pp. 852–863.
- [12] 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.
- [13] Y. Song and S. Ermon, "Generative modeling by estimating gradients of the data distribution," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 1–13.
- [14] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Proc. Adv. Neural Inf. Process. Syst.*, 2020, pp. 6840–6851.
- [15] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, "Score-based generative modeling through stochastic differential equations," 2020, *arXiv:2011.13456*.
- [16] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 3730–3738.
- [17] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
- [18] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2012, pp. 1097–1105.
- [19] S. C. Wong, A. Gatt, V. Stamatescu, and M. D. McDonnell, "Understanding data augmentation for classification: When to warp?" in *Proc. Int. Conf. Digit. Image Comput., Techn. Appl. (DICTA)*, Nov. 2016, pp. 1–6.
- [20] L. Perez and J. Wang, "The effectiveness of data augmentation in image classification using deep learning," 2017, *arXiv:1712.04621*.
- [21] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification," *Neurocomputing*, vol. 321, pp. 321–331, Dec. 2018.
- [22] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 1, pp. 1–48, Dec. 2019.
- [23] S. Gidaris, P. Singh, and N. Komodakis, "Unsupervised representation learning by predicting image rotations," 2018, *arXiv:1803.07728*.
- [24] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 1597–1607.
- [25] 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.*, 2020, pp. 18661–18673.
- [26] J. Zhang and K. Ma, "Rethinking the augmentation module in contrastive learning: Learning hierarchical augmentation invariance with expanded views," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 16629–16638.
- [27] N.-T. Tran, V.-H. Tran, N.-B. Nguyen, T.-K. Nguyen, and N.-M. Cheung, "On data augmentation for GAN training," *IEEE Trans. Image Process.*, vol. 30, pp. 1882–1897, 2021.
- [28] T. DeVries and G. W. Taylor, "Improved regularization of convolutional neural networks with cutout," 2017, *arXiv:1708.04552*.
- [29] M. Shi, F. Xie, J. Yang, J. Zhao, X. Liu, and F. Wang, "Cutout with patch-loss augmentation for improving generative adversarial networks against instability," *Comput. Vis. Image Understand.*, vol. 234, Sep. 2023, Art. no. 103761.
- [30] S. Jenni and P. Favaro, "On stabilizing generative adversarial training with noise," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 12137–12145.
- [31] R. Feng, D. Zhao, and Z.-J. Zha, "Understanding noise injection in GANs," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 3284–3293.
- [32] P. Dhariwal and A. Nichol, "Diffusion models beat GANs on image synthesis," in *Proc. Adv. Neural Inf. Process. Syst.*, 2021, pp. 8780–8794.
- [33] Y. Wang, W. Yang, X. Chen, Y. Wang, L. Guo, L.-P. Chau, Z. Liu, Y. Qiao, A. C. Kot, and B. Wen, "SinSR: Diffusion-based image super-resolution in a single step," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 25796–25805.
- [34] J. Kim and T.-K. Kim, "Arbitrary-scale image generation and upsampling using latent diffusion model and implicit neural decoder," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 9202–9211.
- [35] K. Islam, M. Z. Zaheer, A. Mahmood, and K. Nandakumar, "Diffusemix: Label-preserving data augmentation with diffusion models," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 27611–27620.
- [36] Z. Wang, L. Wei, T. Wang, H. Chen, Y. Hao, X. Wang, X. He, and Q. Tian, "Enhance image classification via inter-class image mixup with diffusion model," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 17223–17233.
- [37] C. Zhan, Y. Lin, G. Wang, H. Wang, and J. Wu, "MedM2G: Unifying medical multi-modal generation via cross-guided diffusion with visual invariant," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 11502–11512.
- [38] X. Yang, C. Ding, Z. Hong, J. Huang, J. Tao, and X. Xu, "Texture-preserving diffusion models for high-fidelity virtual try-on," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 7017–7026.
- [39] Y. Gu, H. Xu, Y. Xie, G. Song, Y. Shi, D. Chang, J. Yang, and L. Luo, "DiffPortrait3D: Controllable diffusion for zero-shot portrait view synthesis," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 10456–10465.
- [40] Z. Ding, X. Zhang, Z. Tu, and Z. Xia, "Restoration by generation with constrained priors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 2567–2577.
- [41] R. Gao, C. Zhao, L. Hong, and Q. Xu, "DiffGuard: Semantic mismatch-guided out-of-distribution detection using pre-trained diffusion models," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2023, pp. 1579–1589.

- [42] T. Yoon, J. Y. Choi, S. Kwon, and E. K. Ryu, "Diffusion probabilistic models generalize when they fail to memorize," in *Proc. ICML Workshop Struct. Probabilistic Inference Generative Modeling*, 2023, pp. 1–11.
- [43] X. Gu, C. Du, T. Pang, C. Li, M. Lin, and Y. Wang, "On memorization in diffusion models," 2023, *arXiv:2310.02664*.
- [44] Z. Kadkhodaie, F. Guth, E. P. Simoncelli, and S. Mallat, "Generalization in diffusion models arises from geometry-adaptive harmonic representations," 2023, *arXiv:2310.02557*.
- [45] J. Song, C. Meng, and S. Ermon, "Denoising diffusion implicit models," in *Proc. Int. Conf. Learn. Represent.*, 2021, pp. 1–20. [Online]. Available: <https://openreview.net/forum?id=StlgiaRCHLP>
- [46] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 10674–10685.
- [47] N. Carlini et al., "Extracting training data from diffusion models," in *Proc. 32nd USENIX Secur. Symp.*, 2023, pp. 5253–5270.
- [48] G. Somepalli, V. Singla, M. Goldblum, J. Geiping, and T. Goldstein, "Diffusion art or digital forgery? Investigating data replication in diffusion models," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 6048–6058.
- [49] G. Somepalli, V. Singla, M. Goldblum, J. Geiping, and T. Goldstein, "Understanding and mitigating copying in diffusion models," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 36, 2023, pp. 47783–47803.
- [50] C. Chen, D. Liu, and C. Xu, "Towards memorization-free diffusion models," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 8425–8434.
- [51] Z. Zhang, Z. Zhao, J. Yu, and Q. Tian, "Shiftdpms: Exploring conditional diffusion models by shifting diffusion trajectories," in *Proc. AAAI Conf. Artif. Intell.*, vol. 37, 2023, pp. 3552–3560.
- [52] Y. Zhou, B. Liu, Y. Zhu, X. Yang, C. Chen, and J. Xu, "Shifted diffusion for text-to-image generation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 10157–10166.
- [53] R. Cherukuru, K. Hussain, I. Kavati, A. M. Reddy, and K. S. Reddy, "Sentiment classification with modified RoBERTa and recurrent neural networks," *Multimedia Tools Appl.*, vol. 83, no. 10, pp. 29399–29417, Sep. 2023.
- [54] X. Xu, J. Zhang, F. Liu, M. Sugiyama, and M. S. Kankanhalli, "Enhancing adversarial contrastive learning via adversarial invariant regularization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2023, pp. 16783–16803.
- [55] Y. Ouyang, L. Xie, and G. Cheng, "Improving adversarial robustness through the contrastive-guided diffusion process," in *Proc. Int. Conf. Mach. Learn.*, 2023, pp. 26699–26723.
- [56] J. Ho and T. Salimans, "Classifier-free diffusion guidance," 2022, *arXiv:2207.12598*.
- [57] 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.
- [58] M.-E. Nilsback and A. Zisserman, "Automated flower classification over a large number of classes," in *Proc. 6th Indian Conf. Comput. Vis., Graph. Image Process.*, Dec. 2008, pp. 722–729.
- [59] Hugging Face. *Diffusers: Unconditional Training Documentation*. Accessed: May 29, 2025. [Online]. Available: https://huggingface.co/docs/diffusers/en/training/unconditional_training
- [60] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," 2017, *arXiv:1711.05101*.
- [61] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training GANs," in *Proc. 30th Int. Conf. Neural Inf. Process. Syst.*, D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, R. Garnett, Eds., 2016, pp. 2234–2242.
- [62] 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, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett, Eds., 2017, pp. 1–12.
- [63] I. Gulrajani, C. Raffel, and L. Metz, "Towards GAN benchmarks which require generalization," 2020, *arXiv:2001.03653*.
- [64] Y. Balaji, M. Sajedi, N. Mukund Kalibhat, M. Ding, D. Stöger, M. Soltanolkotabi, and S. Feizi, "Understanding overparameterization in generative adversarial networks," 2021, *arXiv:2104.05605*.
- [65] T. Kynkäänniemi, T. Karras, M. Aittala, T. Aila, and J. Lehtinen, "The role of ImageNet classes in Fréchet inception distance," 2022, *arXiv:2203.06026*.

- [66] K. Cui, Y. Yu, F. Zhan, S. Liao, S. Lu, and E. Xing, "KD-DLGAN: Data limited image generation via knowledge distillation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 3872–3882.
- [67] E. Liu, X. Ning, Z. Lin, H. Yang, and Y. Wang, "OMS-DPM: Optimizing the model schedule for diffusion probabilistic models," in *Proc. Int. Conf. Mach. Learn.*, 2023, pp. 21915–21936.
- [68] F. Yu, A. Seff, Y. Zhang, S. Song, T. Funkhouser, and J. Xiao, "LSUN: Construction of a large-scale image dataset using deep learning with humans in the loop," 2015, *arXiv:1506.03365*.



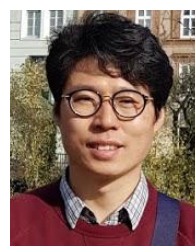
KENSUKE NAKAMURA received the M.Sc. and D.Phil. degrees in engineering from Kyoto Institute of Technology, Japan, in 2004 and 2012, respectively. He is currently a Research Associate with Chung-Ang University (CAU), Seoul, South Korea. He has participated in research projects on models and applications of the two-/three-dimensional human body shapes. His current research interests include computer vision, generative models, and their applications to fashion design.



BONG-SOO SOHN received the B.S. degree in computer science from Seoul National University, South Korea, and the M.S. and Ph.D. degrees in computer sciences from The University of Texas at Austin, in 2001 and 2005, respectively. He is currently an Associate Professor with the School of Computer Science and Engineering, Chung-Ang University (CAU). He was a Postdoctoral Scholar with Stanford University, supervised by Prof. Taylor. His research interests include multi-scale visualization, geometric modeling, and image processing from 3D/4D data with an emphasis on bio-molecular and medical applications.



SIMON KORMAN received the M.Sc. degree in computer science and mathematics from the Weizmann Institute supervised by Prof. Uriel Feige and the D.Phil. degree from Tel-Aviv University, in 2015, supervised by Prof. Shai Avidan. He is currently a Faculty Member with the Department of Computer Science, University of Haifa. He was a Postdoctoral Researcher with the Weizmann Institute supervised by Prof. Ronen Basri and a Postdoctoral Researcher with UCLA supervised by Prof. Stefano Soatto. His research interests include computer vision and machine learning, including unsupervised representation learning, generative models, and implicit data representations.



BYUNG-WOO HONG received the M.Sc. degree in computer vision from the Weizmann Institute of Science, in 2001, supervised by Prof. Shimon Ullman, and the D.Phil. degree in computer vision from the University of Oxford, in 2005, supervised by Prof. Sir Michael Brady. He joined the Computer Science Department, CAU, as a Faculty Member, in 2008, after his postdoctoral research at the Computer Science Department, University of California at Los Angeles, with Prof. Stefano Soatto. His research interests include image processing, computer vision, machine learning, and medical image analysis.

...