Table 5: Evaluation for generative models: ImageNet-1-mode, ImageNet-2-modes, ImageNet-5-modes, and ImageNet-10-modes.

Model	ImageNet-1-mode	ImageNet-2-modes	ImageNet-5-modes	ImageNet-10-modes
FID	58.30	57.34	57.78	57.26
AFD	0	8.14	12.84	14.47

A AFD VALIDATION

In this section, we thoroughly validate the effectiveness of our proposed metric, AFD, for measuring conditional diversity and demonstrate its role as a complementary metric to FID. In unconditional generation scenarios, the FID is widely used to evaluate the diversity of generated images. While low FID scores generally indicate high diversity across the entire dataset, they do not necessarily imply high conditional diversity. For instance, we observed that samples generated by the DDBM model often lack diversity when conditioned on edge images, despite achieving very low FID scores. To address this limitation, we introduce the concept of conditional diversity and propose a corresponding metric to quantify it.

The first question is why FID failed to measure the conditional diversity. To illustrate the limitations of FID in capturing conditional diversity, consider an extreme case: if the images generated by a generative model are identical to a set of baseline images, the FID score can be very low since the two distributions are indistinguishable. However, this scenario does not reflect diversity within the conditional outputs.

To further support our point, we designed two classes of pseudo-generative models capable of controlling the diversity of the generated images, which are further validated by FID and AFD. The experiments are evaluated on Imagenet dataset (Deng et al., 2009).

A.1 PSEUDO-GENERATIVE MODELS BY RANDOM SELECTION

We designed four pseudo-generative models: ImageNet-1-mode, ImageNet-2-modes, ImageNet-5-modes, and ImageNet-10-modes. The experimental setup is as follows:

- We selected 11,000 samples from the ImageNet validation dataset, randomly choosing 11 images per class.
- From these, we designated 1,000 images as the ”real” set, while the remaining images served as the source pool for the generative models.
- Each ImageNet-k-modes model simulates a generative process by randomly sampling images from a pool of k distinct images within a given class.

We present sampled images in Fig. 6, where it is evident that the ImageNet-10-modes model generates images with the highest conditional diversity. To quantify this, we conducted experiments to calculate both FID and AFD for the four generative models. The results are summarized in Table 5. While the FID scores are nearly identical across all models, the AFD values increase as the conditional diversity of the generative models improves. This highlights that AFD is a more effective metric for capturing conditional diversity than FID.

A.2 PSEUDO-GENERATIVE MODELS BY STRONG AUGMENTATION

Strong augmentation has been widely used in computer vision to generate synthetic data while preserving its underlying semantics (Chen et al., 2020; Zbontar et al., 2021; Sohn et al., 2020; Berthelot et al., 2019). The intensity of augmentation can be adjusted, with higher intensities producing more diverse images. To further validate our proposed metric, AFD, as a measure of diversity, we construct pseudo-generative models using strong augmentation.

We selected 1,000 images from the ImageNet-1k dataset, one from each category. These images were subjected to data augmentation, specifically using ColorJitter, with varying magnitudes to enhance diversity. For each image, the augmentation was applied 16 times, creating an augmented

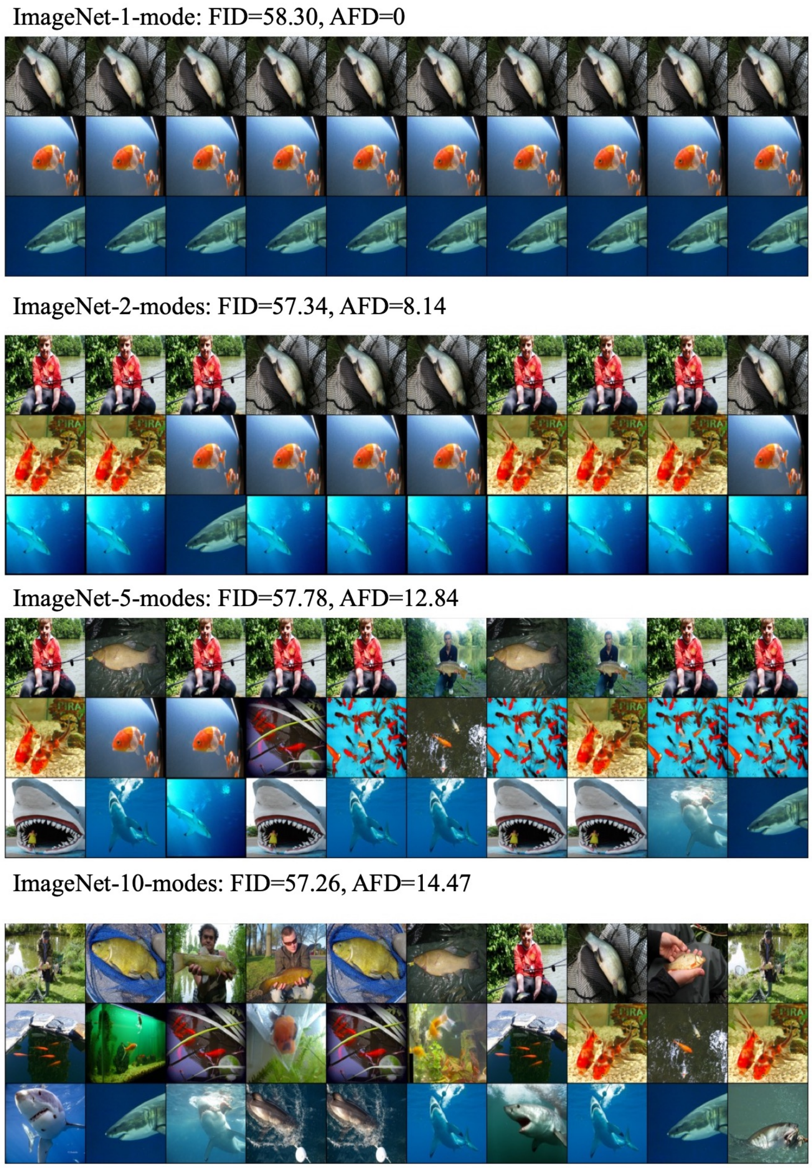


Figure 6: Sampled images from 4 generative models: ImageNet-1-mode, ImageNet-2-modes, ImageNet-5-modes, ImageNet-10-modes.

Table 6: AFD results across different augmentation magnitudes

Augmentation magnitude	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
AFD	2.16	3.77	5.13	6.16	6.98	7.63	8.22	9.01
FID	0.20	2.95	7.02	11.62	16.33	20.84	25.12	28.89

dataset for each magnitude setting. We then calculated the AFD for these augmented datasets to evaluate the relationship between dataset diversity (as influenced by augmentation magnitude) and the AFD value.

Table 6 summarizes the AFD results across various augmentation magnitude settings. The results show that as diversity increases, AFD values also rise, further confirming that the proposed AFD metric is a reliable indicator of image diversity.

B PROOFS

There are infinitely many pinned processes characterized by the Gaussian transition kernel $p_{t|0,T}(\mathbf{x}_t \mid \mathbf{x}_0, \mathbf{x}_T) = \mathcal{N}(\mathbf{x}_t; \alpha_t \mathbf{x}_0 + \beta_t \mathbf{x}_T, \gamma_t^2 \mathbf{I})$. Specifically, we formalize the pinned process as a linear Itô SDE, as presented in Lemma 3.

Lemma 3. *There exist a linear Itô SDE*

$$d\mathbf{X}_t = [f_t \mathbf{X}_t + s_t \mathbf{x}_T]dt + g_t d\mathbf{W}_t, \quad \mathbf{X}_0 = \mathbf{x}_0, \tag{18}$$

where $f_t = \frac{\dot{\alpha}_t}{\alpha_t}, \quad s_t = \dot{\beta}_t - \frac{\dot{\alpha}_t}{\alpha_t}\beta_t, \quad g_t = \sqrt{2(\gamma_t \dot{\gamma}_t - \frac{\dot{\alpha}_t}{\alpha_t}\gamma_t^2)}$, *that has a Gaussian marginal distribution* $\mathcal{N}\left(\mathbf{x}_t; \alpha_t \mathbf{x}_0 + \beta_t \mathbf{x}_T, \gamma_t^2 \mathbf{I}\right)$.

Given the pinned process (18), we can sample from the conditional distribution $p_{0|T}(\mathbf{x}_0|\mathbf{x}_T)$ by solving the reverse SDE or ODE from $t = T$ to $t = 0$:

$$d\mathbf{X}_t = \left[f_t \mathbf{X}_t + s_t \mathbf{x}_T - g_t^2 \nabla_{\mathbf{X}_t} \log p_t(\mathbf{X}_t|\mathbf{x}_T)\right] dt + g_t d\mathbf{W}_t, \quad \mathbf{X}_T = \mathbf{x}_T, \tag{19}$$

$$d\mathbf{X}_t = \left[f_t \mathbf{X}_t + s_t \mathbf{x}_T - \frac{1}{2} g_t^2 \nabla_{\mathbf{X}_t} \log p_t(\mathbf{X}_t|\mathbf{x}_T)\right] dt \quad \mathbf{X}_T = \mathbf{x}_T, \tag{20}$$

where the score $\nabla_{\mathbf{X}_t} \log p_t(\mathbf{X}_t|\mathbf{x}_T)$ can be estimated by score matching objective (8). To improve training stability, we introduced score reparameterization in Sec. 4.1.

Lemma 1. *There exist a linear Itô SDE*

$$d\mathbf{X}_t = [f_t \mathbf{X}_t + s_t \mathbf{x}_T]dt + g_t d\mathbf{W}_t, \quad \mathbf{X}_0 = \mathbf{x}_0, \tag{21}$$

where $f_t = \frac{\dot{\alpha}_t}{\alpha_t}, \quad s_t = \dot{\beta}_t - \frac{\dot{\alpha}_t}{\alpha_t}\beta_t, \quad g_t = \sqrt{2(\gamma_t \dot{\gamma}_t - \frac{\dot{\alpha}_t}{\alpha_t}\gamma_t^2)}$, *that has a Gaussian marginal distribution* $\mathcal{N}\left(\mathbf{x}_t; \alpha_t \mathbf{x}_0 + \beta_t \mathbf{x}_T, \gamma_t^2 \mathbf{I}\right)$.

Proof. Let $\mathbf{m}_t$ denote the mean function of the given Itô SDE, then we have $\frac{d\mathbf{m}_t}{dt} = f_t \mathbf{m}_t + s_t \mathbf{x}_T$. Given the transition kernel, the mean function $\mathbf{m}_t = \alpha_t \mathbf{x}_0 + \beta_t \mathbf{x}_T$, therefore,

$$\dot{\alpha}_t \mathbf{x}_0 + \dot{\beta}_t \mathbf{x}_T = f_t(\alpha_t \mathbf{x}_0 + \beta_t \mathbf{x}_T) + s_t \mathbf{x}_T. \tag{22}$$

Matching the above equation:

$$f_t = \frac{\dot{\alpha}_t}{\alpha_t}, s_t = \dot{\beta}_t - \beta_t \frac{\dot{\alpha}_t}{\alpha_t}. \tag{23}$$

Further, For the variance γ_t^2 of the process, the dynamics are given by:

$$\frac{d\gamma_t^2}{dt} = 2 f_t \gamma_t^2 + g_t^2. \tag{24}$$

Solving for g_t^2, we substitute $f_t = \frac{\dot{\alpha}_t}{\alpha_t}$:

$$g_t^2 = \frac{d\gamma_t^2}{dt} - 2\frac{\dot{\alpha}_t}{\alpha_t}\gamma_t^2 \tag{25}$$

Therefore,

$$g_t = \sqrt{2(\gamma_t \dot{\gamma}_t - \frac{\dot{\alpha}_t}{\alpha_t}\gamma_t^2)}. \tag{26}$$

□

For dynamics described by ODE $d\mathbf{X}_t = \mathbf{u}_t dt$, we can identify the entire class of SDEs that maintain the same marginal distributions, as detailed in Lemma 2. This enables us to control the stochasticity during sampling by appropriately designing ϵ_t.

Lemma 2. *Consider a continuous dynamics given by ODE of the form: $d\mathbf{X}_t = \mathbf{u}_t dt$, with the density evolution $p_t(\mathbf{X}_t)$. Then there exists forward SDEs and backward SDEs that match the marginal distribution p_t. The forward SDEs are given by: $d\mathbf{X}_t = (\mathbf{u}_t + \epsilon_t \nabla \log p_t)dt + \sqrt{2\epsilon_t} d\mathbf{W}_t, \epsilon_t > 0$. The backward SDEs are given by: $d\mathbf{X}_t = (\mathbf{u}_t - \epsilon_t \nabla \log p_t)dt + \sqrt{2\epsilon_t} d\mathbf{W}_t, \epsilon_t > 0$.*

Proof. For the forward SDEs, the Fokker-Planck equations are given by:

$$\begin{aligned}
\frac{\partial p_t(\mathbf{X}_t)}{\partial t} &= -\nabla \cdot \left[(\mathbf{u}_t + \epsilon_t \nabla \log p_t)\, p_t(\mathbf{X}_t) \right] + \epsilon_t \nabla^2 p_t(\mathbf{X}_t) && (27)\\
&= -\nabla \cdot \left[\mathbf{u}_t p_t(\mathbf{X}_t) \right] - \nabla \cdot \left[\epsilon_t (\nabla \log p_t) p_t(\mathbf{X}_t) \right] + \epsilon_t \nabla^2 p_t(\mathbf{X}_t) && (28)\\
&= -\nabla \cdot \left[\mathbf{u}_t p_t(\mathbf{X}_t) \right] - \epsilon_t \nabla \cdot \left[\nabla p_t(\mathbf{X}_t) \right] + \epsilon_t \nabla^2 p_t(\mathbf{X}_t) && (29)\\
&= -\nabla \cdot \left[\mathbf{u}_t p_t(\mathbf{X}_t) \right]. && (30)
\end{aligned}$$

This is exactly the Fokker-Planck equation for the original deterministic ODE $d\mathbf{X}_t = \mathbf{u}_t\, dt$. Therefore, the forward SDE maintains the same marginal distribution $p_t(\mathbf{X}_t)$ as the original ODE.

Now consider the backward SDEs, the Fokker-Planck equations become:

$$\begin{aligned}
\frac{\partial p_t(\mathbf{X}_t)}{\partial t} &= -\nabla \cdot \left[(\mathbf{u}_t - \epsilon_t \nabla \log p_t)\, p_t(\mathbf{X}_t) \right] - \epsilon_t \nabla^2 p_t(\mathbf{X}_t) && (31)\\
&= -\nabla \cdot \left[\mathbf{u}_t p_t(\mathbf{X}_t) \right] + \nabla \cdot \left[\epsilon_t (\nabla \log p_t) p_t(\mathbf{X}_t) \right] - \epsilon_t \nabla^2 p_t(\mathbf{X}_t) && (32)\\
&= -\nabla \cdot \left[\mathbf{u}_t p_t(\mathbf{X}_t) \right]. && (33)
\end{aligned}$$

This is again the Fokker-Planck equation corresponding to the original deterministic ODE $d\mathbf{X}_t = \mathbf{u}_t\, dt$. Therefore, the backward SDE also maintains the same marginal distribution $p_t(\mathbf{X}_t)$.

□

Theorem 3. *Suppose the transition kernel of a diffusion process is given by $p_{t|0,T}(\mathbf{x}_t \mid \mathbf{x}_0, \mathbf{x}_T) = \mathcal{N}(\mathbf{x}_t; \alpha_t \mathbf{x}_0 + \beta_t \mathbf{x}_T, \gamma_t^2 \mathbf{I})$, then the evolution of conditional probability $q(\mathbf{X}_t | \mathbf{x}_T)$ has a class of time reverse sampling SDEs of the form:*

$$d\mathbf{X}_t = \left[\dot{\alpha}_t \hat{\mathbf{x}}_0 + \dot{\beta}_t \mathbf{x}_T - (\dot{\gamma}_t \gamma_t + \epsilon_t) \nabla_{\mathbf{X}_t} \log p_t(\mathbf{X}_t | \mathbf{x}_T) \right] dt + \sqrt{2\epsilon_t} d\mathbf{W}_t \quad \mathbf{X}_T = \mathbf{x}_T. \tag{34}$$

Proof. Recall Eqs. (19) 20 and Lemma 2,

$$d\mathbf{X}_t = \left[\frac{\dot{\alpha}_t}{\alpha_t} \mathbf{x}_t + (\dot{\beta}_t - \frac{\dot{\alpha}_t}{\alpha_t} \beta_t) \mathbf{x}_T - (\gamma_t \dot{\gamma}_t - \frac{\dot{\alpha}_t}{\alpha_t} \gamma_t^2 + \epsilon_t) \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t | \mathbf{x}_T) \right] dt + \sqrt{2\epsilon_t} d\mathbf{w}_t. \tag{35}$$

□

Next we take the reparameterized score 12 into 35:

$$\begin{aligned}
d\mathbf{X}_t &= \left[\frac{\dot{\alpha}_t}{\alpha_t} \mathbf{X}_t + (\dot{\beta}_t - \frac{\dot{\alpha}_t}{\alpha_t} \beta_t) \mathbf{x}_T - (\gamma_t \dot{\gamma}_t - \frac{\dot{\alpha}_t}{\alpha_t} \gamma_t^2 + \epsilon_t) \frac{\alpha_t \hat{\mathbf{x}}_0 + \beta_t \mathbf{x}_T - \mathbf{X}_t}{\gamma_t^2} \right] dt + \sqrt{2\epsilon_t} d\mathbf{w}_t && (36)\\
&= \left[\dot{\alpha}_t \hat{\mathbf{x}}_0 + \dot{\beta}_t \mathbf{x}_T - (\gamma_t \dot{\gamma}_t + \epsilon_t) \frac{\alpha_t \hat{\mathbf{x}}_0 + \beta_t \mathbf{x}_T - \mathbf{X}_t}{\gamma_t^2} \right] dt + \sqrt{2\epsilon_t} d\mathbf{w}_t && (37)\\
&= \left[\dot{\alpha}_t \hat{\mathbf{x}}_0 + \dot{\beta}_t \mathbf{x}_T - (\dot{\gamma}_t + \frac{\epsilon_t}{\gamma_t}) \frac{\alpha_t \hat{\mathbf{x}}_0 + \beta_t \mathbf{x}_T - \mathbf{X}_t}{\gamma_t} \right] dt + \sqrt{2\epsilon_t} d\mathbf{w}_t && (38)\\
&= \left[\dot{\alpha}_t \hat{\mathbf{x}}_0 + \dot{\beta}_t \mathbf{x}_T - (\dot{\gamma}_t + \frac{\epsilon_t}{\gamma_t}) \hat{\mathbf{z}} \right] dt + \sqrt{2\epsilon_t} d\mathbf{w}_t. && (39)
\end{aligned}$$

Theorem 4. *Let* $(\mathbf{x}_0, \mathbf{x}_T) \sim \pi_0(\mathbf{x}_0, \mathbf{x}_T)$*,* $\mathbf{x}_t \sim p_t(\mathbf{x}|\mathbf{x}_0, \mathbf{x}_T)$*, Given the transition kernel:* $p(\mathbf{x}_t \mid \mathbf{x}_0, \mathbf{x}_T) = \mathcal{N}\left(\mathbf{x}_t; \alpha_t\mathbf{x}_0 + \beta_t\mathbf{x}_T, \gamma_t^2\mathbf{I}\right)$*, if* $\hat{\mathbf{x}}_0(\mathbf{x}_t, \mathbf{x}_T, t)$ *is a denoiser function that minimizes the expected* L_2 *denoising error for samples drawn from* $\pi_0(\mathbf{x}_0, \mathbf{x}_T)$*:*

$$\hat{\mathbf{x}}_0(\mathbf{x}_t, \mathbf{x}_T, t) = \arg\min_{D(\mathbf{x}_t, \mathbf{x}_T, t)} \mathbb{E}_{\mathbf{x}_0, \mathbf{x}_T, \mathbf{x}_t}\left[\lambda(t)\|D(\mathbf{x}_t, \mathbf{x}_T, t) - \mathbf{x}_0\|_2^2\right], \tag{40}$$

then the score has the following relationship with $\hat{\mathbf{x}}_0(\mathbf{x}_t, \mathbf{x}_T, t)$*:*

$$\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t|\mathbf{x}_T) = \frac{\alpha_t\hat{\mathbf{x}}_0(\mathbf{x}_t, \mathbf{x}_T, t) + \beta_t\mathbf{x}_T - \mathbf{x}_t}{\gamma_t^2}. \tag{41}$$

Proof.

$$\mathcal{L}(D) = \mathbb{E}_{(\mathbf{x}_0, \mathbf{x}_T)\sim\pi_0(\mathbf{x}_0, \mathbf{x}_T)}\mathbb{E}_{\mathbf{x}_t\sim p_t(\mathbf{x}_t|\mathbf{x}_0, \mathbf{x}_T)}\|D(\mathbf{x}_t) - \mathbf{x}_0\|_2^2 \tag{42}$$

$$= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\underbrace{\int_{\mathbb{R}^d} p_t(\mathbf{x}_t|\mathbf{x}_0, \mathbf{x}_T)\pi_0(\mathbf{x}_0, \mathbf{x}_T)\|D(\mathbf{x}_t) - \mathbf{x}_0\|_2^2 \,\mathrm{d}\mathbf{x}_0}_{=:\mathcal{L}(D;\mathbf{x}_t, \mathbf{x}_T)} \,\mathrm{d}\mathbf{x}_T\mathrm{d}\mathbf{x}_t, \tag{43}$$

$$\mathcal{L}(D; \mathbf{x}_t, \mathbf{x}_T) = \int_{\mathbb{R}^d} p_t(\mathbf{x}_t|\mathbf{x}_0, \mathbf{x}_T)\pi_0(\mathbf{x}_0, \mathbf{x}_T)\|D(\mathbf{x}_t) - \mathbf{x}_0\|_2^2 \,\mathrm{d}\mathbf{x}_0, \tag{44}$$

we can minimize $\mathcal{L}(D)$ by minimizing $\mathcal{L}(D; \mathbf{x}_t, \mathbf{x}_T)$independently for each $\{\mathbf{x}_t, \mathbf{x}_T\}$ pair.

$$D^*(\mathbf{x}_t, \mathbf{x}_T) = \arg\min_{D(\mathbf{x}_t)} \mathcal{L}(D; \mathbf{x}_t, \mathbf{x}_T) \tag{45}$$

$$\mathbf{0} = \nabla_{D(\mathbf{x}_t, \mathbf{x}_T)}[\mathcal{L}(D; \mathbf{x}_t, \mathbf{x}_T)] \tag{46}$$

$$= \int_{\mathbb{R}^d} p_t(\mathbf{x}_t|\mathbf{x}_0, \mathbf{x}_T)\pi_0(\mathbf{x}_0, \mathbf{x}_T)2[D(\mathbf{x}, \mathbf{x}_T) - \mathbf{x}_0] \,\mathrm{d}\mathbf{x}_0 \tag{47}$$

$$= 2[D(\mathbf{x}_t, \mathbf{x}_T)\int_{\mathbb{R}^d} p_t(\mathbf{x}_t|\mathbf{x}_0, \mathbf{x}_T)\pi_0(\mathbf{x}_0, \mathbf{x}_T) \,\mathrm{d}\mathbf{x}_0 - \int_{\mathbb{R}^d} p_t(\mathbf{x}_t|\mathbf{x}_0, \mathbf{x}_T)\pi_0(\mathbf{x}_0, \mathbf{x}_T)\mathbf{x}_0 \,\mathrm{d}\mathbf{x}_0] \tag{48}$$

$$= 2[D(\mathbf{x})p_t(\mathbf{x}_t, \mathbf{x}_T) - \int_{\mathbb{R}^d} p_t(\mathbf{x}_t|\mathbf{x}_0, \mathbf{x}_T)\pi_0(\mathbf{x}_0, \mathbf{x}_T)\mathbf{x}_0 \,\mathrm{d}\mathbf{x}_0], \tag{49}$$

$$D^*(\mathbf{x}_t, \mathbf{x}_T) = \int_{\mathbb{R}^d} \frac{p_t(\mathbf{x}_t|\mathbf{x}_0, \mathbf{x}_T)\pi_0(\mathbf{x}_0, \mathbf{x}_T)\mathbf{x}_0}{p_t(\mathbf{x}_t, \mathbf{x}_T)} \,\mathrm{d}\mathbf{x}_0, \tag{50}$$

$$\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t|\mathbf{x}_T) = \frac{\nabla_{\mathbf{x}_t} p_t(\mathbf{x}_t, \mathbf{x}_T)}{p_t(\mathbf{x}_t, \mathbf{x}_T)} \tag{51}$$

$$= \frac{\int \nabla_{\mathbf{x}_t} p_t(\mathbf{x}_t|\mathbf{x}_T, \mathbf{x}_0)\pi_0(\mathbf{x}_0, \mathbf{x}_T) d\mathbf{x}_0}{p_t(\mathbf{x}_t, \mathbf{x}_T)} \tag{52}$$

$$= -\int \frac{\mathbf{x}_t - \alpha_t\mathbf{x}_0 - \beta_t\mathbf{x}_T}{\gamma^2}\frac{p_t(\mathbf{x}_t|\mathbf{x}_0, \mathbf{x}_T)\pi_0(\mathbf{x}_0, \mathbf{x}_T)}{p_t(\mathbf{x}_t, \mathbf{x}_T)} d\mathbf{x}_0 \tag{53}$$

$$= \frac{\alpha_t D^*(\mathbf{x}_t, \mathbf{x}_T) + \beta_t\mathbf{x}_T - \mathbf{x}_t}{\gamma^2}. \tag{54}$$

Thus we conclude the proof.

□

C REFRAMING PREVIOUS METHODS IN OUR FRAMEWORK

We draw a link between our framework and the diffusion bridge models used in DDBM.

C.1 DDBM-VE

DDBM-VE can be reformulated in our framework as we set :

$$\alpha_t = s_t(1 - \frac{\sigma_t^2}{\sigma_T^2}), \beta_t = \frac{s_t \sigma_t^2}{s_1 \sigma_T^2}, \gamma_t = \sigma_t s_t \sqrt{(1 - \frac{\sigma_t^2}{\sigma_T^2})} \tag{55}$$

Proof. In the origin DDBM paper, the evolution of conditional probability $q(\mathbf{x}_t|\mathbf{x}_T)$ has a time reversed SDE of the form:

$$d\mathbf{X}_t = \left[\bar{\mathbf{f}}_t(\mathbf{X}_t) - \bar{g}_t^2 \bar{\mathbf{h}}_t(\mathbf{X}_t) - \bar{g}_t^2 \mathbf{s}_t(\mathbf{X}_t)\right] dt + \bar{g}_t d\hat{\mathbf{W}}_t, \tag{56}$$

and an associated probability flow ODE

$$d\mathbf{X}_t = \left[\bar{\mathbf{f}}_t(\mathbf{X}_t) - \bar{g}_t^2 \bar{\mathbf{h}}_t(\mathbf{X}_t) - \frac{1}{2} \bar{g}_t^2 \mathbf{s}_t(\mathbf{X}_t)\right] dt. \tag{57}$$

Compare Eqs. (56) and 57 with Lemma 3. We only need to prove:

$$\bar{\mathbf{f}}_t(\mathbf{X}_t) - \bar{g}_t^2 \bar{\mathbf{h}}_t(\mathbf{X}_t) = f_t \mathbf{X}_t + s_t \mathbf{x}_T, \bar{g}_t = g_t. \tag{58}$$

In the original paper,

$$\bar{\mathbf{f}}_t(\mathbf{X}_t) = 0, \bar{g}_t^2 = \frac{d}{dt}\sigma_t^2, \bar{\mathbf{h}}_t(\mathbf{X}_t) = \frac{\mathbf{x}_T - \mathbf{x}_t}{\sigma_T^2 - \sigma_t^2}. \tag{59}$$

Therefore,

$$\bar{\mathbf{f}}_t(\mathbf{X}_t) - \bar{g}_t^2 \bar{\mathbf{h}}_t(\mathbf{X}_t) = \frac{2\sigma_t \dot{\sigma}_t (\mathbf{x}_T - \mathbf{x}_t)}{\sigma_T^2 - \sigma_t^2}, \bar{g}_t^2 = 2\dot{\sigma}_t \sigma_t. \tag{60}$$

In our framework, f_t, s_t, g_t^2 can be calculated:

$$f_t = \frac{\dot{\alpha}_t}{\alpha_t} = \frac{d}{dt}\log \alpha_t = \frac{d}{dt}\log \frac{\sigma_T^2 - \sigma_t^2}{\sigma_T^2} = \frac{-2\sigma_t \dot{\sigma}_t}{\sigma_T^2 - \sigma_t^2}, \tag{61}$$

$$s_t = \dot{\beta}_t - \frac{\dot{\alpha}_t}{\alpha_t}\beta_t = \frac{2\sigma_t \dot{\sigma}_t}{\sigma_T^2} + \frac{2\sigma_t \dot{\sigma}_t}{\sigma_T^2 - \sigma_t^2} \cdot \frac{\sigma_t^2}{\sigma_T^2} = \frac{2\sigma_t \dot{\sigma}_t}{\sigma_T^2 - \sigma_t^2}. \tag{62}$$

$$g_t^2 = 2(\gamma_t \dot{\gamma}_t - \frac{\dot{\alpha}_t}{\alpha_t}\gamma_t^2) = 2\gamma_t^2 \left(\frac{\dot{\gamma}_t}{\gamma_t} - \frac{\dot{\alpha}_t}{\alpha_t}\right) = \gamma_t^2 \left(\frac{(\sigma_T^2 - 2\sigma_t^2)\dot{\sigma}_t}{(\sigma_T^2 - \sigma_t^2)\sigma_t} + \frac{2\dot{\sigma}_t \sigma_t}{\sigma_T^2 - \sigma_t^2}\right) = 2\sigma_t \dot{\sigma}_t. \tag{63}$$

Therefore,

$$f_t \mathbf{X}_t + s_t \mathbf{x}_T = \frac{2\sigma_t \dot{\sigma}_t (\mathbf{x}_T - \mathbf{x}_t)}{\sigma_T^2 - \sigma_t^2} = \bar{\mathbf{f}}_t(\mathbf{X}_t) - \bar{g}_t^2 \bar{\mathbf{h}}_t(\mathbf{X}_t), \quad \bar{g}_t = g_t, \tag{64}$$

which matches the formulation in DDBM.

□

C.2 DDBM-VP

DDBM-VP can be reformulated in our framework as we set :

$$\alpha_t = a_t(1 - \frac{\sigma_t^2 a_1^2}{\sigma_1^2 a_t^2}), \beta_t = \frac{\sigma_t^2 a_1}{\sigma_1^2 a_t}, \gamma_t = \sqrt{\sigma_t^2(1 - \frac{\sigma_t^2 a_1^2}{\sigma_1^2 a_t^2})}. \tag{65}$$

Proof. In the original DDBM-VP setting,

$$\bar{\mathbf{f}}_t(\mathbf{X}_t) = \frac{d \log a_t}{dt}\mathbf{x}_t, \tag{66}$$

$$\bar{g}_t^2 = 2\sigma_t\dot{\sigma}_t - 2\frac{\dot{a}_t}{a_t}\sigma_t^2 = \frac{2\sigma_t\dot{\sigma}_t a_t - 2\sigma_t^2\dot{a}_t}{a_t}, \tag{67}$$

$$\bar{\mathbf{h}}_t(\mathbf{X}_t) = \frac{(a_t/a_1)\mathbf{x}_T - \mathbf{x}_t}{\sigma_t^2(\mathrm{SNR}_t/\mathrm{SNR}_1 - 1)} = \frac{a_1 a_t \mathbf{x}_T - a_1^2\mathbf{x}_t}{\sigma_1^2 a_t^2 - \sigma_t^2 a_1^2}. \tag{68}$$

Therefore,

$$\bar{\mathbf{f}}_t(\mathbf{X}_t) - \bar{g}_t^2\bar{\mathbf{h}}_t(\mathbf{X}_t) = \left[\frac{\dot{a}_t}{a_t} - \frac{2\sigma_t a_1^2(\dot{\sigma}_t a_t - \sigma_t\dot{a}_t)}{a_t(\sigma_1^2 a_t^2 - \sigma_t^2 a_1^2)}\right]\mathbf{x}_t + \frac{2\sigma_t a_1(\dot{\sigma}_t a_t - \sigma_t\dot{a}_t)}{\sigma_1^2 a_t^2 - \sigma_t^2 a_1^2}\mathbf{x}_T. \tag{69}$$

In our framework, f_t, s_t, g_t^2 can be calculated:

$$\begin{aligned}
f_t = \frac{\dot{\alpha}_t}{\alpha_t} &= \frac{d}{dt}\log\alpha_t && (70)\\
&= \frac{d}{dt}\log\frac{\sigma_1^2 a_t^2 - \sigma_t^2 a_1^2}{\sigma_1^2 a_t} && (71)\\
&= \frac{2\sigma_1^2 a_t\dot{a}_t - 2a_1^2\sigma_t\dot{\sigma}_t}{\sigma_1^2 a_t^2 - \sigma_t^2 a_1^2} - \frac{\dot{a}_t}{a_t} && (72)\\
&= \frac{\dot{a}_t}{a_t} - \frac{2a_1^2\sigma_t(a_t\dot{\sigma}_t - \dot{a}_t\sigma_t)}{a_t(\sigma_1^2 a_t^2 - \sigma_t^2 a_1^2)}, && (73)
\end{aligned}$$

$$\begin{aligned}
s_t = \dot{\beta}_t - \frac{\dot{\alpha}_t}{\alpha_t}\beta_t &= \beta_t(\frac{\dot{\beta}_t}{\beta_t} - \frac{\dot{\alpha}_t}{\alpha_t}) && (74)\\
&= \frac{\sigma_t^2 a_1}{\sigma_1^2 a_t}\left(\frac{2\dot{\sigma}_t}{\sigma_t} - \frac{2\sigma_1^2 a_t\dot{a}_t - 2a_1^2\sigma_t\dot{\sigma}_t}{\sigma_1^2 a_t^2 - \sigma_t^2 a_1^2}\right) && (75)\\
&= \frac{2\sigma_t a_1(\dot{\sigma}_t a_t - \sigma_t\dot{a}_t)}{\sigma_1^2 a_t^2 - \sigma_t^2 a_1^2}, && (76)
\end{aligned}$$

$$\begin{aligned}
g_t^2 = \gamma_t\dot{\gamma}_t - \frac{\dot{\alpha}_t}{\alpha_t}\gamma_t^2 &= \gamma_t^2\left(\frac{\dot{\gamma}_t}{\gamma_t} - \frac{\dot{\alpha}_t}{\alpha_t}\right) && (77)\\
&= \gamma^2\frac{d}{dt}\log\frac{\gamma_t}{\alpha_t} && (78)\\
&= \gamma^2\frac{d}{dt}(\frac{1}{2}\log\frac{\sigma_t^2\sigma_1^2}{\sigma_1^2 a_t^2 - \sigma_t^2 a_1^2}) && (79)\\
&= \sigma_t^2\left(1 - \frac{\sigma_t^2 a_1^2}{\sigma_1^2 a_t^2}\right)\left(\frac{\dot{\sigma}_t}{\sigma_t} - \frac{\sigma_1^2 a_t\dot{a}_t - a_1^2\sigma_t\dot{\sigma}_t}{\sigma_1^2 a_t^2 - \sigma_t^2 a_1^2}\right) && (80)\\
&= \frac{\dot{\sigma}_t\sigma_t a_t - \sigma_t^2\dot{a}_t}{a_t}. && (81)
\end{aligned}$$

Therefore,

$$f_t \mathbf{X}_t + s_t \mathbf{x}_T == \bar{\mathbf{f}}_t(\mathbf{X}_t) - \bar{g}_t^2 \bar{\mathbf{h}}_t(\mathbf{X}_t), \bar{g}_t = g_t, \tag{82}$$

which matches the formulation in DDBM.

□

C.3 EDM

ODE formulation. The ODE formulation in EDM can be formlated in our framework as we set $\alpha_t = 1, \beta_t = 0, \gamma_t = \sigma_t$.

Proof. Recall 20, the ODE formulation is given by:

$$d\mathbf{X}_t = \left[f_t \mathbf{X}_t + s_t \mathbf{x}_T - \frac{1}{2} g_t^2 \nabla_{\mathbf{X}_t} \log p_t(\mathbf{X}_t|\mathbf{x}_T) \right] dt \quad \mathbf{X}_T = \mathbf{x}_T \tag{83}$$

where $f_t = \frac{\dot{\alpha}_t}{\alpha_t}, \quad s_t = \dot{\beta}_t - \frac{\dot{\alpha}_t}{\alpha_t}\beta_t, \quad g_t = \sqrt{2(\gamma_t \dot{\gamma}_t - \frac{\dot{\alpha}_t}{\alpha_t}\gamma_t^2)}$. As $\alpha_t = 1, \beta_t = 0, \gamma_t = \sigma_t$, The sampling ODE is given by:

$$d\mathbf{X}_t = -\sigma_t \dot{\sigma}_t \nabla_{\mathbf{x}_t} \log p_t(\mathbf{X}_t) dt \tag{84}$$

□

Denoising score matching. The score remarameterization in EDM is the same as ours in Eq. 12. Let $\alpha_t = 1, \beta_t = 0, \gamma_t = \sigma_t$, then the score reparameterization in Eq. 12 is given by:

$$\nabla_{\mathbf{x}_t} \log p_t(\mathbf{X}_t) \approx \frac{\hat{\mathbf{x}}_0 - \mathbf{x}_t}{\sigma_t^2}. \tag{85}$$

Sampling SDEs with stochasticity added. Recall Theorem 1, as $\alpha_t = 1, \beta_t = 0, \gamma_t = \sigma_t$, then the SDE has the form:

$$d\mathbf{X}_t = (-\sigma_t \dot{\sigma}_t + \epsilon_t) \nabla_{\mathbf{x}_t} \log p_t(\mathbf{X}_t) dt + \sqrt{2\epsilon_t} d\mathbf{W}_t. \tag{86}$$

Now we recover the stochastic sampling SDE in original EDM paper.

C.4 I2SB

I2SB can be reformulated in our framework as we let:

$$\alpha_t = 1 - \frac{\sigma_t^2}{\sigma_1^2}, \beta_t = \frac{\sigma_t^2}{\sigma_1^2}, \gamma_t = \sqrt{\sigma_t^2 (1 - \frac{\sigma_t^2}{\sigma_1^2})} \tag{87}$$

where $\sigma_t^2 := \int_0^t \beta_\tau d\tau$.

Using discretization 17:

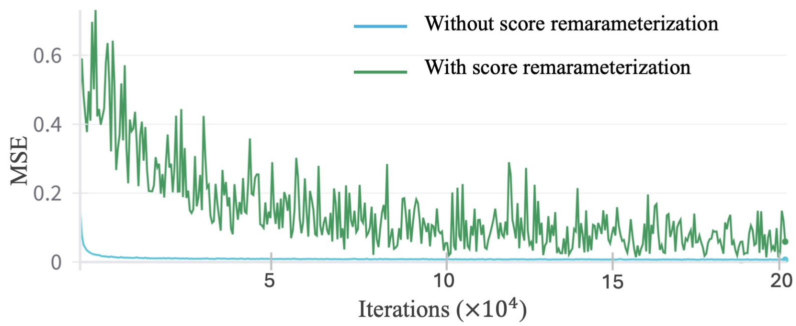
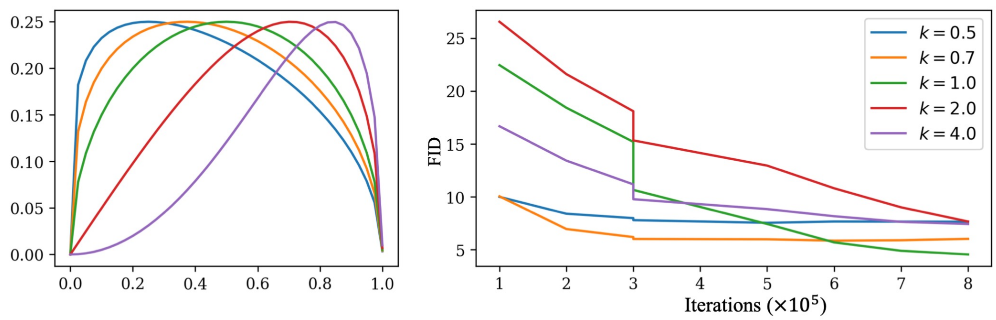


Figure 7: MSEs during training, where $\text{MSE} = \frac{1}{B}\sum_{i=1}^{B}\|\hat{x}_0 - x_0\|^2$.



(a) γ_t.

(b) Ablation study on the shape of γ_t.

Figure 8: Ablation study on the shape of γ_t.

$$\mathbf{x}_{t-\Delta t} = (\alpha_{t-\Delta t} - \alpha_t \frac{\beta_{t-\Delta t}}{\beta_t})\hat{\mathbf{x}}_0 + \frac{\beta_{t-\Delta t}}{\beta_t}\mathbf{x}_t + \sqrt{\gamma_{t-\Delta t}^2 - \frac{\beta_{t-\Delta t}^2 \gamma_t^2}{\beta_t^2}}\bar{\mathbf{z}}_t \tag{88}$$

$$= (1 - \frac{\beta_{t-\Delta t}}{\beta_t})\hat{\mathbf{x}}_0 + \frac{\beta_{t-\Delta t}}{\beta_t}\mathbf{x}_t + \sqrt{\gamma_{t-\Delta t}^2 - \frac{\beta_{t-\Delta t}^2 \gamma_t^2}{\beta_t^2}}\bar{\mathbf{z}}_t \tag{89}$$

$$= (1 - \frac{\sigma_{t-\Delta t}^2}{\sigma_t^2})\hat{\mathbf{x}}_0 + \frac{\sigma_{t-\Delta t}^2}{\sigma_t^2}\mathbf{x}_t + \sqrt{\frac{\sigma_{t-\Delta t}^2(1 - \frac{\sigma_{t-\Delta t}^2}{\sigma_1^2})\frac{\sigma_t^4}{\sigma_1^4} - \frac{\sigma_{t-\Delta t}^4}{\sigma_1^4}\sigma_t^2(1 - \frac{\sigma_t^2}{\sigma_1^2})}{\frac{\sigma_t^4}{\sigma_1^4}}}\bar{\mathbf{z}}_t \tag{90}$$

$$= (1 - \frac{\sigma_{t-\Delta t}^2}{\sigma_t^2})\hat{\mathbf{x}}_0 + \frac{\sigma_{t-\Delta t}^2}{\sigma_t^2}\mathbf{x}_t + \sqrt{\frac{\sigma_{t-\Delta t}^2(\sigma_t^2 - \sigma_{t-\Delta t}^2)}{\sigma_t^2}}\bar{\mathbf{z}}_t \tag{91}$$

In the I2SB paper, define $a_n^2 := \int_{t_n}^{t_{n+1}} \beta_\tau \mathrm{d}\tau$, $\sigma_n^2 := \int_0^{t_n} \beta_\tau \mathrm{d}\tau$. Therefore,

$$\mathbf{x}_n = \frac{a_n^2}{a_n^2 + \sigma_n^2}\hat{\mathbf{x}}_0 + \frac{\sigma_n^2}{a_n^2 + \sigma_n^2}\mathbf{x}_{n+1} + \sqrt{\frac{\sigma_n^2 a_n^2}{\alpha_n^2 + \sigma_n^2}}\bar{\mathbf{z}}_t \tag{92}$$

Thus, we reproduce the sampler of I2SB.

D Additional design guideline

Score reparameterization. We compared the training stability with and without score reparameterization using the DIODE (64×64) dataset, and the results are shown in Fig. 7. For training without

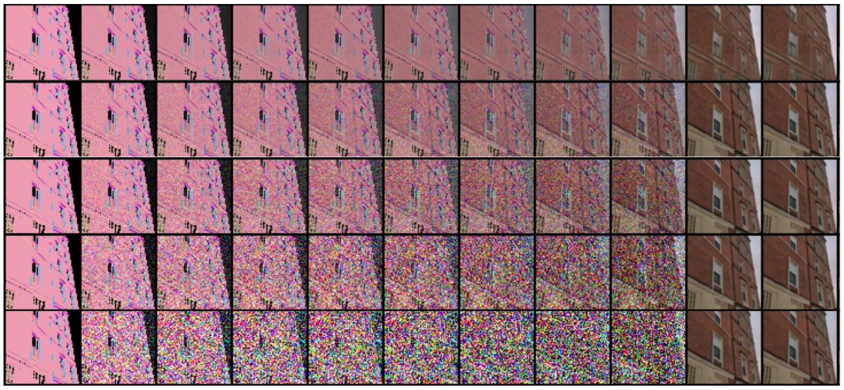


Figure 9: Sampling paths with dfferent choices of γ_t. As γ_t extreamly low, e.g, $\gamma_{\max} = 0.025$, the model will be failed to construct details of images.

score reparameterization, the score function $s_\theta(\mathbf{x}, \mathbf{x}_T, t)$ is parameterized by a neural network, and $\hat{\mathbf{x}}_0(\mathbf{x}, \mathbf{x}_T, t)$ is computed as: $\hat{\mathbf{x}}_0(\mathbf{x}, \mathbf{x}_T, t) = \frac{1}{\alpha_t}\left(\gamma_t^2 s_\theta(\mathbf{x}, \mathbf{x}_T, t) + \mathbf{x}_t - \beta \mathbf{x}_T\right)$. For training with score reparameterization, $\hat{\mathbf{x}}_0(\mathbf{x}, \mathbf{x}_T, t)$ is directly parameterized as a neural network. We then compared the mean squared error (MSE) between $\hat{\mathbf{x}}_0$ and $\mathbf{x}_0$ during training. The results in Fig. 7 indicate that score reparameterization helps reduce training instability.

α_t and β_t. Theoretically, α_t and β_t can be freely designed, and future work may explore alternative design choices. However, in this paper, we focus on the simple case where $\alpha_t = 1 - t$ and $\beta_t = t$. The rationale is as follows: consider the scenario where $\alpha_t = 1 - \beta_t$, which represents an interpolation along the line segment between x_0 and x_1. For the path $p_t^{(1)}(x) = \mathcal{N}((1 - \beta_t)x_0 + \beta_t x_1, \gamma_t^2 \mathbf{I})$, where β_t is invertible, it is straightforward to construct another path $p_t^{(2)}(x) = \mathcal{N}((1 - t)x_0 + t x_1, \gamma_{\beta_t^{-1}}^2 \mathbf{I})$, which achieves the same objective function but uses a different distribution of t during training. Based on this equivalence, setting $\alpha_t = 1 - t$ and $\beta_t = t$ is a reasonable choice.

The shape of γ_t. We conducted an ablation study on γ_t with different shapes. Specifically, we assumed γ_t has the form $\gamma_t = 2\gamma_{\max}\sqrt{t^k(1 - t^k)}$, as shown in Fig. 8, γ_t will have different shape as we set different k. The results indicate that the best performance is achieved when $k = 1$, which is the exact setting used in this paper.

$\gamma_{\max}$. Our ablation studies on $\gamma_{\max}$ demonstrate that the optimal values of $\gamma_{\max}$ are approximately 0.125 or 0.25. Furthermore, the sampling paths corresponding to different choices of γ_t are shown in Fig. 9. Adding an appropriate amount of noise to the transition kernel helps in constructing finer details.

ϵ_t. We use the setting $\epsilon_t = \eta\left(\gamma_t \dot{\gamma}_t - \frac{\dot{\alpha}_t}{\alpha_t}\gamma_t^2\right)$. The ablation studies on ϵ_t demonstrate that the optimal choice of η for the DDBM-VP model is approximately 0.3, while the best choice for the SDB model with a Linear Path is around 1.0. Additionally, we present sample paths and generated images under different η settings to illustrate heuristic parameter tuning techniques. The results are shown in Figures 11, 12, and 13. Too small a value of η results in the loss of high-frequency information, while too large a value of η produces over-sharpened and potentially noisy sampled images.

E Experiment Details

Architecture. We maintain the architecture and parameter settings consistent with (Linqi Zhou et al., 2023), utilizing the ADM model (Dhariwal & Nichol, 2021) for 64×64 resolution, modifying the channel dimensions from 192 to 256 and reducing the number of residual blocks from three to two. Apart from these changes, all other settings remain identical to those used for 64×64 resolution.

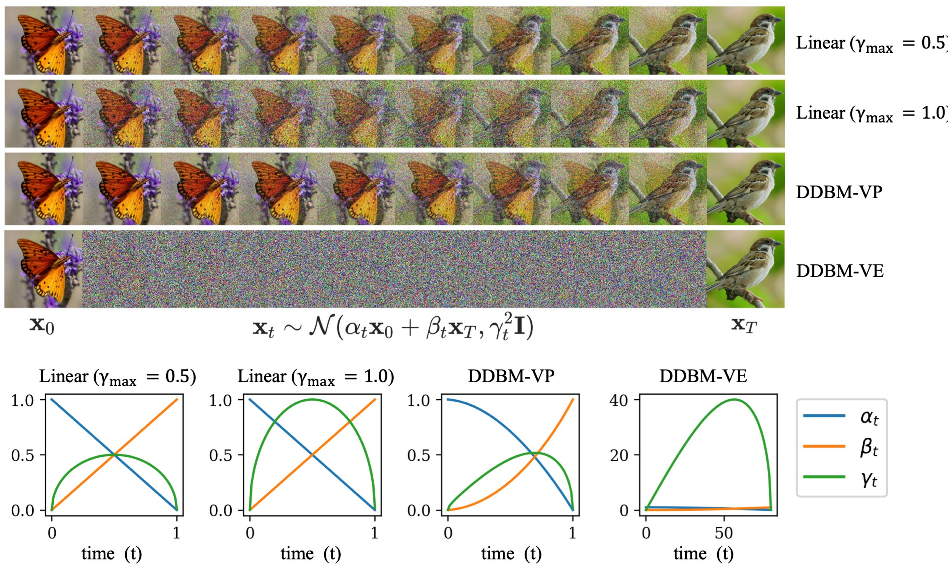


Figure 10: An illustration of design choices of transition kernels and how they affect the I2I translation process. α_t and β_t define the interpolation between two images, while γ_t controls the noise added to the process. ntuitively, the DDBM-VE model introduces excessive noise in the middle stages, which is unnecessary for effective image translation and may explain its poor performance. In contrast, our Linear path results in a symmetrical noise schedule, ensuring a more balanced process. On the other hand, the DDBM-VP path adds more noise near $\mathbf{x}_T$, , indicating that during training, more computational resources are focused around $\mathbf{x}_0$.

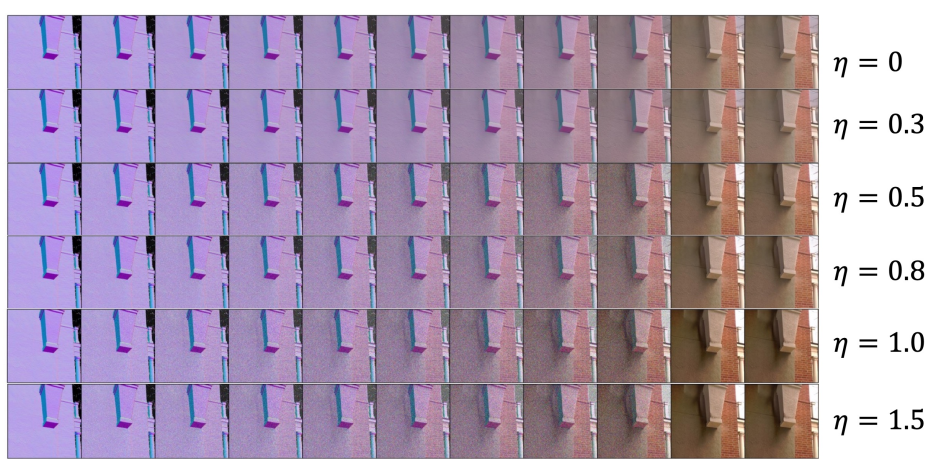


Figure 11: Sampling path with dfferent choices of ϵ_t. As $\epsilon_t = 0$, the generated images lack details, as ϵ_t too large, the sampled images are over-sharpening. The best choices of ϵ_t are around $\epsilon_t = 0.8$ and $\epsilon_t = 1.0$.

Training. We include additional pre- and post-processing steps: scaling functions and loss weighting, the same ingredient as (Karras et al., 2022). Let $D_\theta(\mathbf{x}_t, \mathbf{x}_T, t) = c_{\text{skip}}(t)\mathbf{x}_t + c_{\text{out}(t)}(t)F_\theta(c_{\text{in}}(t)\mathbf{x}_t, c_{\text{noise}}(t))$, where F_θ is a neural network with parameter θ, the effective training target with respect to the raw network F_θ is: $\mathbb{E}_{\mathbf{x}_t,\mathbf{x}_0,\mathbf{x}_T,t}\left[\lambda \| c_{\text{skip}}(\mathbf{x}_t + c_{\text{out}}F_\theta(c_{\text{in}}\mathbf{x}_t, c_{\text{noise}}) - \mathbf{x}_0\|^2\right]$. Scaling scheme are chosen by requiring network inputs and training targets to have unit variance ($c_{\text{in}}, c_{\text{out}}$), and amplifying errors in F_θ as little as possible. Following reasoning in (Linqi Zhou et al., 2023),

$$c_{\text{in}}(t) = \frac{1}{\sqrt{\alpha_t^2\sigma_0^2 + \beta_t^2\sigma_T^2 + 2\alpha_t\beta_t\sigma_{0T} + \gamma_t^2}}, \quad c_{\text{skip}}(t) = (\alpha_t\sigma_0^2 + \beta_t\sigma_{0T}) * c_{\text{in}}^2, \tag{93}$$

$$c_{\text{out}}(t) = \sqrt{\beta_t^2\sigma_0^2\sigma_1^2 - \beta_t^2\sigma_{0T}^2 + \gamma_t^2\sigma_0^2 c_{\text{in}}}, \quad \lambda = \frac{1}{c_{\text{out}}^2}, \quad c_{\text{noise}}(t) = \frac{1}{4}\log(t), \tag{94}$$

where σ_0^2, σ_T^2, and σ_{0T} denote the variance of $\mathbf{x}_0$, variance of $\mathbf{x}_T$ and the covariance of the two, respectively.

We note that TrigFlow (Lu & Song, 2024), a contemporaneous work, adopts the same score reparameterization and pre-conditioning techniques. It can be considered a special case of our framework by setting $\alpha_t = \cos(t)$, $\beta_t = 0$, $\gamma_t = \sigma_0 \sin(t)$, $t \in [0, \frac{\pi}{2}]$. In this case, $\sigma_T = 0$, $\sigma_{0T} = 0$,

$$c_{\text{in}}(t) = \frac{1}{\sqrt{\alpha_t^2\sigma_0^2 + \gamma_t^2}} = \frac{1}{\sqrt{\sin^2(t)\sigma_0^2 + \cos^2(t)\sigma_0^2}} = \frac{1}{\sigma_0}, \tag{95}$$

$$c_{\text{skip}}(t) = (\alpha_t\sigma_0^2)c_{in}^2 = \cos(t)\cdot\sigma_0^2\cdot\frac{1}{\sigma_0^2} = \cos(t), \tag{96}$$

$$c_{out}(t) = \sqrt{\gamma_t^2\sigma_0^2}\cdot c_{in} = \sin(t)\sigma_0, \tag{97}$$

$$D_\theta(x_t, t) = c_{\text{skip}}x_t + c_{\text{out}}F_\theta(c_{\text{in}}x_t, c_{\text{noise}}) = \cos(t)x_t + \sin(t)\sigma_0 F_\theta(\frac{1}{\sigma_0}, c_{\text{noise}}). \tag{98}$$

Then we recover TrigFlow.

In our implementation, we set $\sigma_0 = \sigma_T = 0.5, \sigma_{0T} = \sigma_0^2/2$ for all training sessions. Other setting are shown in Table 7.

Table 7: Training settings

Model	Dataset	edges→handbags	edges→handbags	edges→handbags
	η	0	0	0.5
	$\gamma_{\max}$	0.125	0.25	0.125
Setting	GPU	1 A6000 48G	1 H100 96G	1 H100 96G
	Batch size	32	128	200
	Learning rate	1×10^{-5}	5×10^{-5}	1×10^{-4}
	epochs	2078	2106	1443
	Training time	42 days	8 days	11 days
Model	Dataset	DIODE (256×256)	DOIDE (256×256)	
	η	0	0	
	$\gamma_{\max}$	0.125	0.25	
Setting	GPU	1 H100 96G	1 H100 96G	
	Batch size	16	16	
	Learning rate	2×10^{-5}	2×10^{-5}	
	epochs	2617	1745	
	Training time	17 days	25 days	

Sampling. We use the same timesteps distributed according to EDM (Karras et al., 2022): $(t_{\max}^{1/\rho} + \frac{i}{N}(t_{\min}^{1/\rho} - t_{\max}^{1/\rho}))^\rho$, where $t_{\min} = 0.001$ and $t_{\max} = 1 - 10^{-4}$. The best performance achieved by setting $\rho = 0.6$ for Edges2handbags and $\rho = 0.8$ for DIODE datasets.

Licenses

- Edges→Handbags Isola et al. (2017): BSD license.
- DIODE-Outdoor Vasiljevic et al. (2019): MIT license.

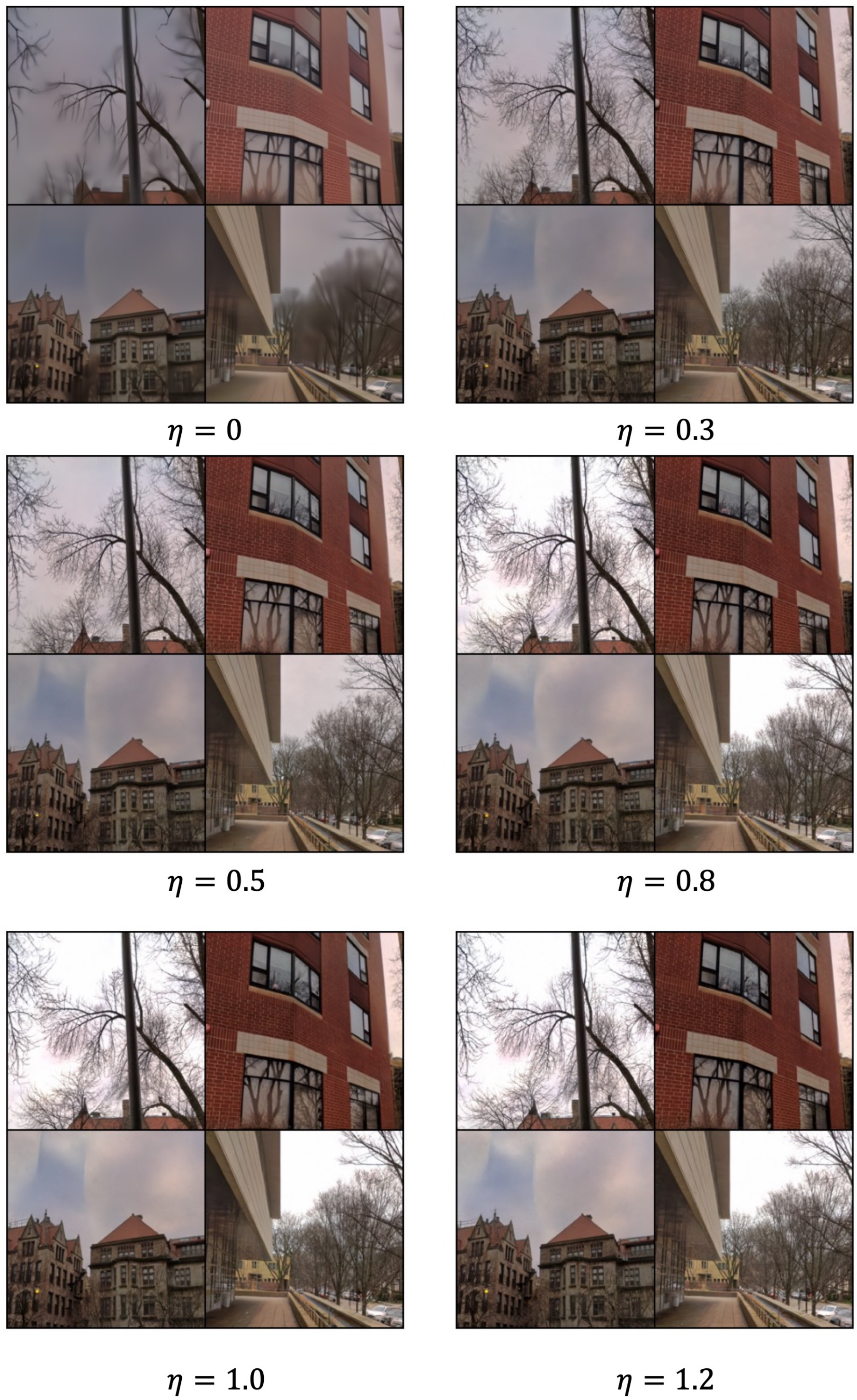


Figure 12: Comparison of sampled images with different ϵ_t for SDB model, where $\epsilon_t = \eta(\gamma_t \dot{\gamma}_t - \frac{\dot{\alpha}_t}{\alpha_t}\gamma_t^2)$, $\gamma_{\max} = 0.25$, $b = 0$.

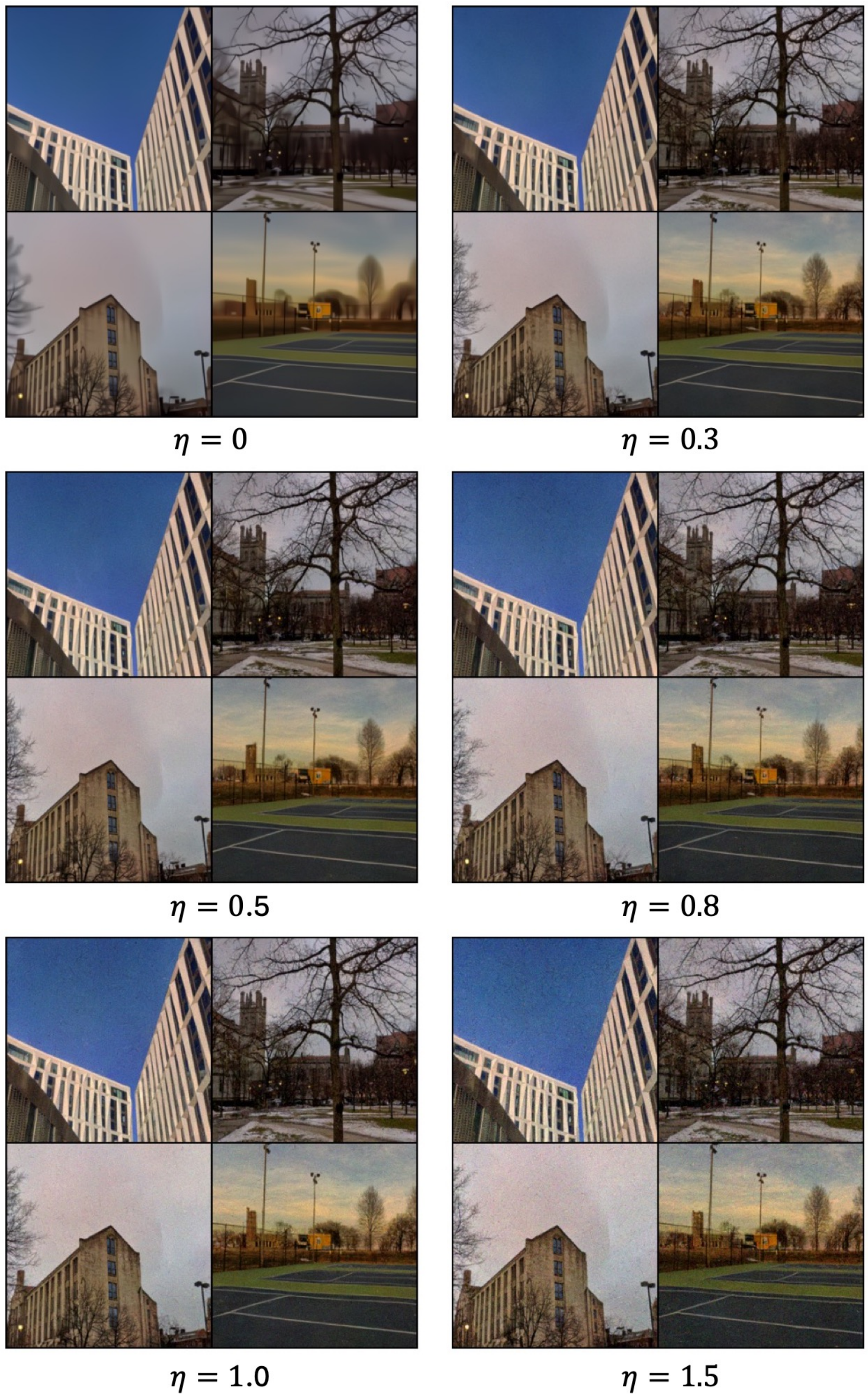


Figure 13: Comparison of sampled images with different ϵ_t for DDBM-VP pretrained model, where $\epsilon_t = \eta(\gamma_t\dot{\gamma}_t - \frac{\dot{\alpha}_t}{\alpha_t}\gamma_t^2)$.

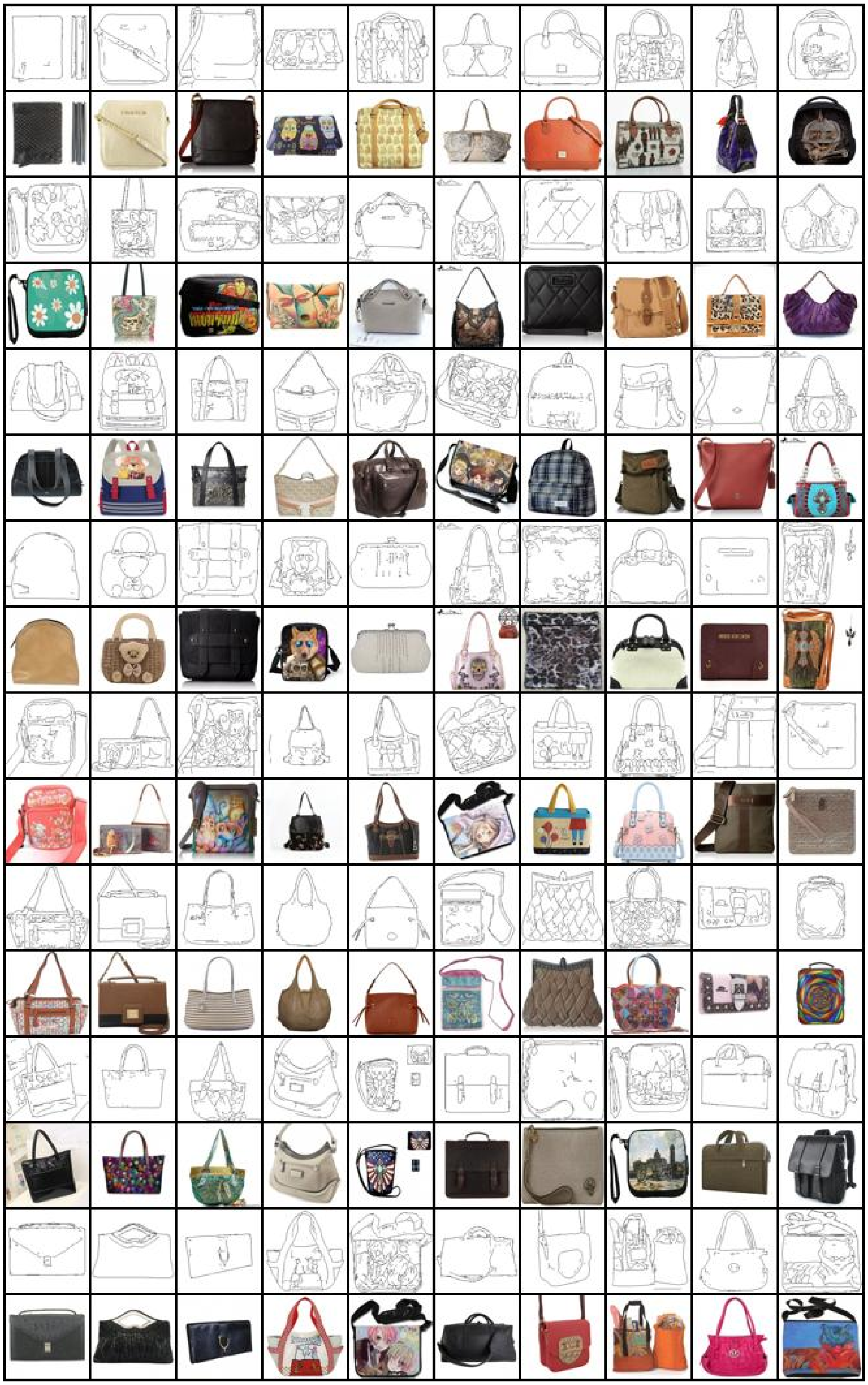


Figure 14: SDB model and sampler ($\gamma_{\max} = 0.125$, $\eta = 1$, $b = 0$, NFE=5, FID=0.89).

F Additional visualizations

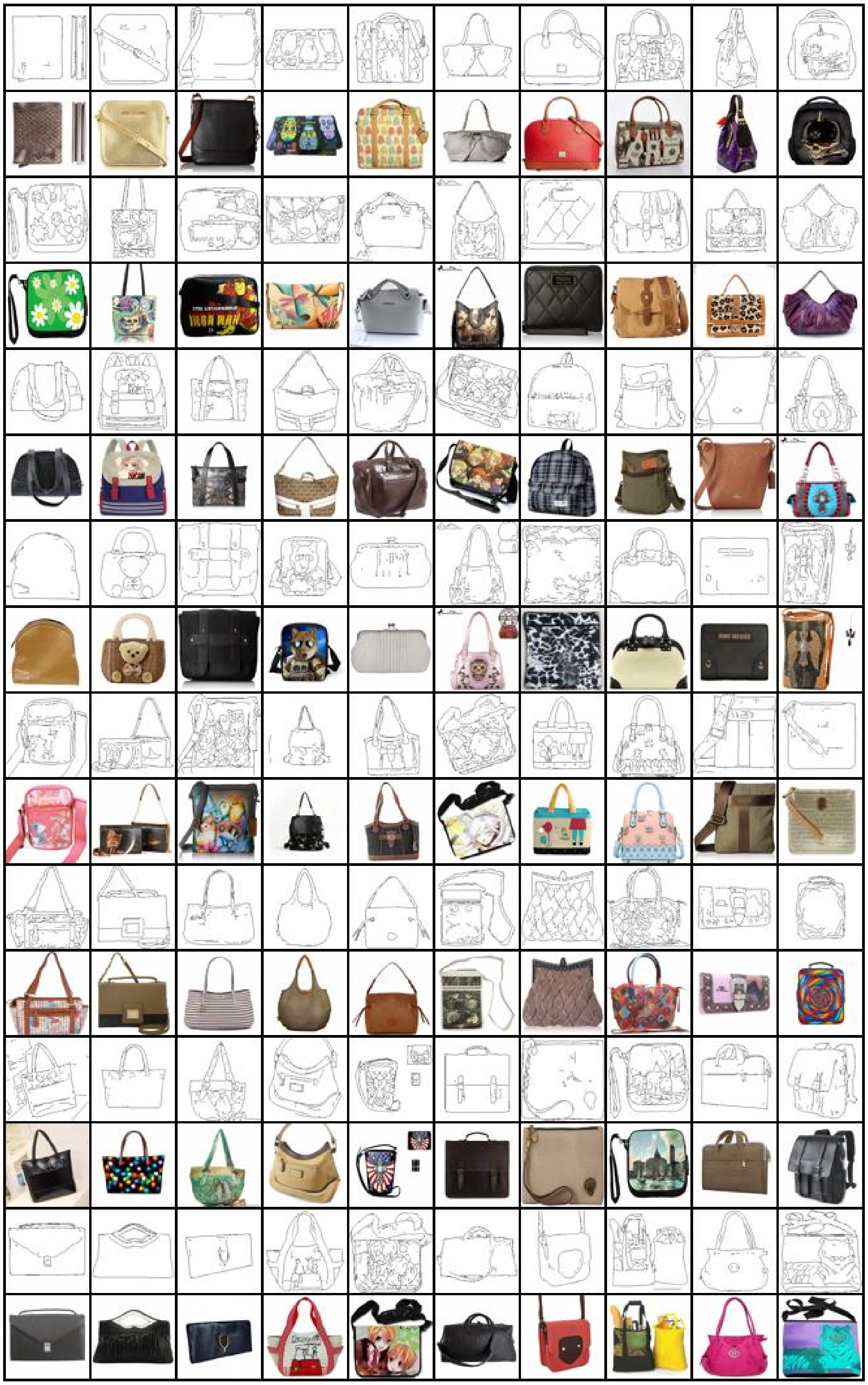


Figure 15: DDBM model and Our sampler (NFE=20, FID=1.53).

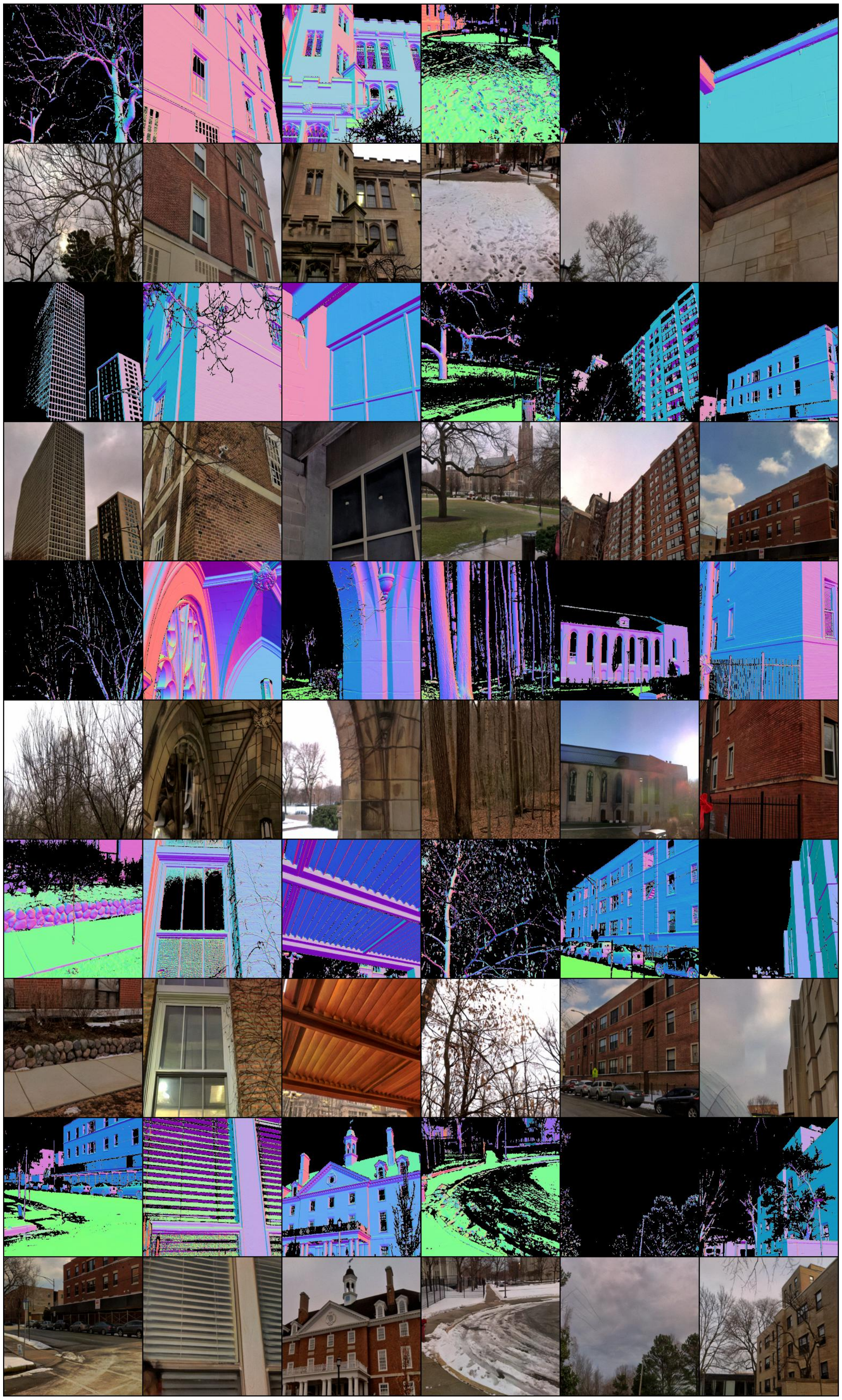


Figure 16: DDBM model and SDB sampler ($\eta = 0.3$, NFE=20, FID=4.12). Samples for DIODE dataset (conditoned on depth images).

Figure 17: SDB model and sampler ($\gamma_{\max} = 0.25$, $\eta = 1.0$, $b = 0$, NFE=5, FID = 4.16).

Figure 18: SDB model and sampler ($\gamma_{\max} = 0.25$, $\eta = 1.0$, $b = 0$, NFE=20, FID = 3.27).

Figure 19: DDBM model and DBIM sampler (NFE=10, FID = 2.46, AFD=5.20).

Figure 20: DDBM model and sampler (NFE=118, FID = 1.83, AFD=6.99).

Figure 21: SDB model and sampler ($\gamma_{\max} = 0.125$, $b = 1.0$, NFE=10, FID = 2.07, AFD=9.35).