Model / steps	1	2	4	8	16
Vanila GANs DDGAN [11] SIDDMs(ours) SIDDMs w/o AFD (ours)	3.641/4.598 - -	- 0.240/0.015 0.004/0.008 0.522/0.719	- 0.017/0.022 0.008/0.005 2.911/3.018	- 0.017/0.015 0.018/0.024 0.182/0.208	- 0.072/0.076 0.026/0.026 0.398/0.384

Table 6: $\overline{\text{MOG 5x5 results}}$, $\overline{\text{MMD}} \downarrow$ (left) / Unbaised FID \downarrow (right). This table shows the MMD distance and the Unbiased FID between the ground truth distribution and the generated data distribution.

Inference Steps	2	4	8	16
SSIDMs	6.78	3.11	14.706	41.27

Table 7: The results on conditional Imagenet1000 64×64 generation. We train the model with different step sizes and evaluate them with the corresponding step sizes. Step size number "4" denotes the paper results.



(5) Conditional examples on Imangenet1000 64×64 from our model with four steps training and inference without classifier guidance or classifier-free guidance.



(6) preliminary conditional examples on Laion4B from our model with four steps training and inference. The image resolution is 128×128 . Compared with our Imagenet1000 experiments, we have additional cross-attention for text input, and the input resolution for G is 128×128 . The left image is the ground truth, and the right side is the generated images from our model, which share the same text prompt condition.

$$\begin{split} \min_{\theta} \max_{D_{\phi}} \sum_{t>0} \mathbb{E}_{q(x_{0})q(x_{t-1}|x_{0})q(x_{t}|x_{t-1})} \Big[[-\log(D_{\phi}(x_{t-1},x_{t},t))] \\ &+ \mathbb{E}_{p_{\theta}(x_{t-1}|x_{t})} [-\log(1-D_{\phi}(x_{t-1},x_{t},t))] \Big] \\ &= \min_{\theta} \max_{D_{\phi}} \sum_{t>0} \mathbb{E}_{q(x_{t})q(x_{t-1}|x_{t})} [-\log(D_{\phi}(x_{t-1},x_{t},t))] \\ &+ \mathbb{E}_{q(x_{t})p_{\theta}(x_{t-1}|x_{t})} [-\log(1-D_{\phi}(x_{t-1},x_{t},t))] \\ &\iff \min_{\theta} \max_{D_{adv}} D_{adv}(q(x_{t-1},x_{t})) ||p_{\theta}(x_{t-1},x_{t})). \end{split}$$
(A)

$$\min_{\theta} - H(p_{\theta}(x_t|x_{t-1})) = \min_{\theta} \mathbb{E}_{p_{\theta}(x_t|x_{t-1})} \log p_{\theta}(x_t|x_{t-1}) \\
\approx \min_{\theta} \mathbb{E}_{p_{\theta}(x_t|x_{t-1})} \log p_{\psi}(x_t|x_{t-1}).$$
(B)