Generating High Fidelity Data from Low-density Regions using Diffusion Models

Vikash Sehwag† Caner Hazirbas‡ Albert Gordo‡ Firat Ozgenel‡ Cristian Canton Ferrer†
† Princeton University, ‡ Meta AI
`vvikash@princeton.edu, {hazirbas, agordo, firatozgenel, ccanton}@fb.com`

CVPR 2022

Preliminaries

Diffusion models:
Gradually add Gaussian noise and then reverse

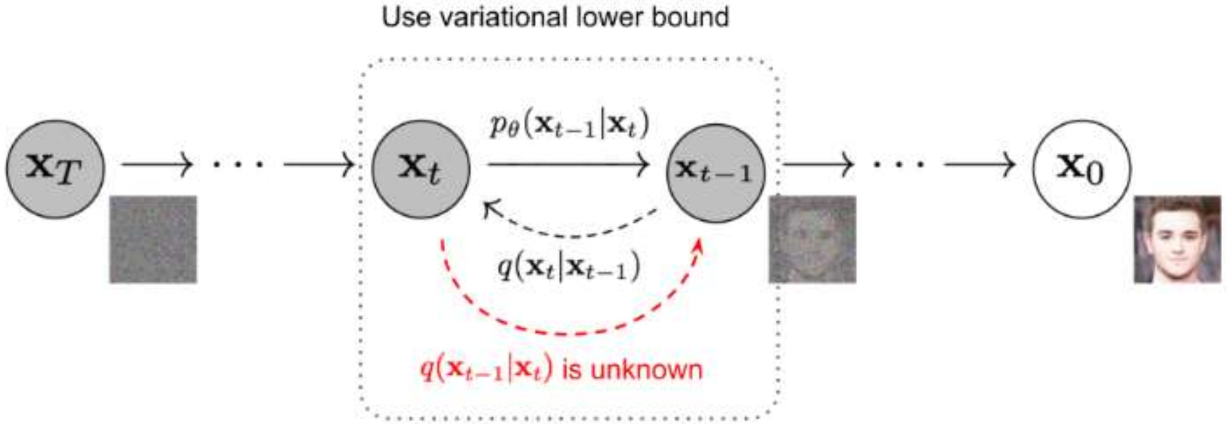


Fig. 2. The Markov chain of forward (reverse) diffusion process of generating a sample by slowly adding (removing) noise. (Image source: Ho et al. 2020 with a few additional annotations)

$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I}) \quad q(\mathbf{x}_{1:T}|\mathbf{x}_0) = \prod_{t=1}^{T} q(\mathbf{x}_t|\mathbf{x}_{t-1})$$

A nice property of the above process is that we can sample $\mathbf{x}_t$ at any arbitrary time step t in a closed form using reparameterization trick. Let $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i$:

$$\begin{aligned}
\mathbf{x}_t &= \sqrt{\alpha_t}\mathbf{x}_{t-1} + \sqrt{1-\alpha_t}\boldsymbol{\epsilon}_{t-1} && \text{;where } \boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \dots \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \\
&= \sqrt{\alpha_t\alpha_{t-1}}\mathbf{x}_{t-2} + \sqrt{1-\alpha_t\alpha_{t-1}}\bar{\boldsymbol{\epsilon}}_{t-2} && \text{;where } \bar{\boldsymbol{\epsilon}}_{t-2} \text{ merges two Gaussians (*).} \\
&= \dots \\
&= \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon} \\
q(\mathbf{x}_t|\mathbf{x}_0) &= \mathcal{N}(\mathbf{x}_t; \sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1-\bar{\alpha}_t)\mathbf{I})
\end{aligned}$$

$$\begin{aligned}
q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) &= q(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{x}_0)\frac{q(\mathbf{x}_{t-1}|\mathbf{x}_0)}{q(\mathbf{x}_t|\mathbf{x}_0)} \\
&\propto \exp\Big(-\frac{1}{2}\big(\frac{(\mathbf{x}_t - \sqrt{\alpha_t}\mathbf{x}_{t-1})^2}{\beta_t} + \frac{(\mathbf{x}_{t-1} - \sqrt{\bar{\alpha}_{t-1}}\mathbf{x}_0)^2}{1-\bar{\alpha}_{t-1}} - \frac{(\mathbf{x}_t - \sqrt{\bar{\alpha}_t}\mathbf{x}_0)^2}{1-\bar{\alpha}_t}\big)\Big) \\
&= \exp\Big(-\frac{1}{2}\big(\frac{\mathbf{x}_t^2 - 2\sqrt{\alpha_t}\mathbf{x}_t\mathbf{x}_{t-1} + \alpha_t\mathbf{x}_{t-1}^2}{\beta_t} + \frac{\mathbf{x}_{t-1}^2 - 2\sqrt{\bar{\alpha}_{t-1}}\mathbf{x}_0\mathbf{x}_{t-1} + \bar{\alpha}_{t-1}\mathbf{x}_0^2}{1-\bar{\alpha}_{t-1}} - \frac{(\mathbf{x}_t - \sqrt{\bar{\alpha}_t}\mathbf{x}_0)^2}{1-\bar{\alpha}_t}\big)\Big) \\
&= \exp\Big(-\frac{1}{2}\big((\frac{\alpha_t}{\beta_t} + \frac{1}{1-\bar{\alpha}_{t-1}})\mathbf{x}_{t-1}^2 - (\frac{2\sqrt{\alpha_t}}{\beta_t}\mathbf{x}_t + \frac{2\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_{t-1}}\mathbf{x}_0)\mathbf{x}_{t-1} + C(\mathbf{x}_t, \mathbf{x}_0)\big)\Big)
\end{aligned}$$

$$\tilde{\beta}_t = 1/(\frac{\alpha_t}{\beta_t} + \frac{1}{1-\bar{\alpha}_{t-1}}) = 1/(\frac{\alpha_t - \bar{\alpha}_t + \beta_t}{\beta_t(1-\bar{\alpha}_{t-1})}) = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t} \cdot \beta_t$$

$$\begin{aligned}
\tilde{\boldsymbol{\mu}}_t(\mathbf{x}_t, \mathbf{x}_0) &= (\frac{\sqrt{\alpha_t}}{\beta_t}\mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_{t-1}}\mathbf{x}_0)/(\frac{\alpha_t}{\beta_t} + \frac{1}{1-\bar{\alpha}_{t-1}}) \\
&= (\frac{\sqrt{\alpha_t}}{\beta_t}\mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_{t-1}}\mathbf{x}_0)\frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t} \cdot \beta_t \\
&= \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\mathbf{x}_0
\end{aligned}$$

Preliminaries-VLB

$$
\begin{aligned}
L_{\mathrm{CE}} &= -\mathbb{E}_{q(\mathbf{x}_0)} \log p_\theta(\mathbf{x}_0) \\
&= -\mathbb{E}_{q(\mathbf{x}_0)} \log \Big(\int p_\theta(\mathbf{x}_{0:T}) d\mathbf{x}_{1:T} \Big) \\
&= -\mathbb{E}_{q(\mathbf{x}_0)} \log \Big(\int q(\mathbf{x}_{1:T}|\mathbf{x}_0) \frac{p_\theta(\mathbf{x}_{0:T})}{q(\mathbf{x}_{1:T}|\mathbf{x}_0)} d\mathbf{x}_{1:T} \Big) \\
&= -\mathbb{E}_{q(\mathbf{x}_0)} \log \Big(\mathbb{E}_{q(\mathbf{x}_{1:T}|\mathbf{x}_0)} \frac{p_\theta(\mathbf{x}_{0:T})}{q(\mathbf{x}_{1:T}|\mathbf{x}_0)} \Big) \\
&\leq -\mathbb{E}_{q(\mathbf{x}_{0:T})} \log \frac{p_\theta(\mathbf{x}_{0:T})}{q(\mathbf{x}_{1:T}|\mathbf{x}_0)} \\
&= \mathbb{E}_{q(\mathbf{x}_{0:T})} \Big[\log \frac{q(\mathbf{x}_{1:T}|\mathbf{x}_0)}{p_\theta(\mathbf{x}_{0:T})} \Big] = L_{\mathrm{VLB}}
\end{aligned}
$$

$$
\begin{aligned}
L_{\mathrm{VLB}} &= \mathbb{E}_{q(\mathbf{x}_{0:T})} \Big[\log \frac{q(\mathbf{x}_{1:T}|\mathbf{x}_0)}{p_\theta(\mathbf{x}_{0:T})} \Big] \\
&= \mathbb{E}_q \Big[\log \frac{\prod_{t=1}^T q(\mathbf{x}_t|\mathbf{x}_{t-1})}{p_\theta(\mathbf{x}_T) \prod_{t=1}^T p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)} \Big] \\
&= \mathbb{E}_q \Big[-\log p_\theta(\mathbf{x}_T) + \sum_{t=1}^T \log \frac{q(\mathbf{x}_t|\mathbf{x}_{t-1})}{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)} \Big] \\
&= \mathbb{E}_q \Big[-\log p_\theta(\mathbf{x}_T) + \sum_{t=2}^T \log \frac{q(\mathbf{x}_t|\mathbf{x}_{t-1})}{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)} + \log \frac{q(\mathbf{x}_1|\mathbf{x}_0)}{p_\theta(\mathbf{x}_0|\mathbf{x}_1)} \Big] \\
&= \mathbb{E}_q \Big[-\log p_\theta(\mathbf{x}_T) + \sum_{t=2}^T \log \Big(\frac{q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)}{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)} \cdot \frac{q(\mathbf{x}_t|\mathbf{x}_0)}{q(\mathbf{x}_{t-1}|\mathbf{x}_0)} \Big) + \log \frac{q(\mathbf{x}_1|\mathbf{x}_0)}{p_\theta(\mathbf{x}_0|\mathbf{x}_1)} \Big] \\
&= \mathbb{E}_q \Big[-\log p_\theta(\mathbf{x}_T) + \sum_{t=2}^T \log \frac{q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)}{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)} + \sum_{t=2}^T \log \frac{q(\mathbf{x}_t|\mathbf{x}_0)}{q(\mathbf{x}_{t-1}|\mathbf{x}_0)} + \log \frac{q(\mathbf{x}_1|\mathbf{x}_0)}{p_\theta(\mathbf{x}_0|\mathbf{x}_1)} \Big] \\
&= \mathbb{E}_q \Big[-\log p_\theta(\mathbf{x}_T) + \sum_{t=2}^T \log \frac{q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)}{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)} + \log \frac{q(\mathbf{x}_T|\mathbf{x}_0)}{q(\mathbf{x}_1|\mathbf{x}_0)} + \log \frac{q(\mathbf{x}_1|\mathbf{x}_0)}{p_\theta(\mathbf{x}_0|\mathbf{x}_1)} \Big] \\
&= \mathbb{E}_q \Big[\log \frac{q(\mathbf{x}_T|\mathbf{x}_0)}{p_\theta(\mathbf{x}_T)} + \sum_{t=2}^T \log \frac{q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)}{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)} - \log p_\theta(\mathbf{x}_0|\mathbf{x}_1) \Big] \\
&= \mathbb{E}_q [\underbrace{D_{\mathrm{KL}}(q(\mathbf{x}_T|\mathbf{x}_0) \parallel p_\theta(\mathbf{x}_T))}_{L_T} + \sum_{t=2}^T \underbrace{D_{\mathrm{KL}}(q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) \parallel p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t))}_{L_{t-1}} \underbrace{- \log p_\theta(\mathbf{x}_0|\mathbf{x}_1)}_{L_0}]
\end{aligned}
$$

Preliminaries-DDPM

$$\tilde{\beta}_t = 1/(\frac{\alpha_t}{\beta_t} + \frac{1}{1-\bar{\alpha}_{t-1}}) = 1/(\frac{\alpha_t - \bar{\alpha}_t + \beta_t}{\beta_t(1-\bar{\alpha}_{t-1})}) = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\cdot\beta_t$$

$$\begin{aligned}
\tilde{\boldsymbol{\mu}}_t(\mathbf{x}_t,\mathbf{x}_0) &= (\frac{\sqrt{\alpha_t}}{\beta_t}\mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_{t-1}}\mathbf{x}_0)/(\frac{\alpha_t}{\beta_t} + \frac{1}{1-\bar{\alpha}_{t-1}}) \\
&= (\frac{\sqrt{\alpha_t}}{\beta_t}\mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_{t-1}}\mathbf{x}_0)\frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\cdot\beta_t \\
&= \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\mathbf{x}_0
\end{aligned}$$

$$q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_{t-1};\tilde{\boldsymbol{\mu}}(\mathbf{x}_t,\mathbf{x}_0),\sigma_t^2\mathbf{I})\, p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1};\boldsymbol{\mu}_\theta(\mathbf{x}_t,t),\sigma_t^2\mathbf{I})$$

$$\begin{aligned}
D_{\mathrm{KL}}(q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0)\parallel p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)) &= D_{\mathrm{KL}}(\mathcal{N}(\mathbf{x}_{t-1};\tilde{\boldsymbol{\mu}}(\mathbf{x}_t,\mathbf{x}_0),\sigma_t^2\mathbf{I})\parallel\mathcal{N}(\mathbf{x}_{t-1};\boldsymbol{\mu}_\theta(\mathbf{x}_t,t),\sigma_t^2\mathbf{I})) \\
&= \frac{1}{2}(n + \frac{1}{\sigma_t^2}\|\tilde{\boldsymbol{\mu}}_t(\mathbf{x}_t,\mathbf{x}_0) - \boldsymbol{\mu}_\theta(\mathbf{x}_t,t)\|^2 - n + \log 1) \\
&= \frac{1}{2\sigma_t^2}\|\tilde{\boldsymbol{\mu}}_t(\mathbf{x}_t,\mathbf{x}_0) - \boldsymbol{\mu}_\theta(\mathbf{x}_t,t)\|^2
\end{aligned}$$

$$L_{t-1} = \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}\left[\frac{1}{2\sigma_t^2}\|\tilde{\boldsymbol{\mu}}_t(\mathbf{x}_t,\mathbf{x}_0) - \boldsymbol{\mu}_\theta(\mathbf{x}_t,t)\|^2\right]$$

Preliminaries-DDPM

$$\tilde{\beta}_t = 1/\left(\frac{\alpha_t}{\beta_t} + \frac{1}{1-\bar{\alpha}_{t-1}}\right) = 1/\left(\frac{\alpha_t - \bar{\alpha}_t + \beta_t}{\beta_t(1-\bar{\alpha}_{t-1})}\right) = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\cdot\beta_t$$

$$\begin{aligned}\tilde{\boldsymbol{\mu}}_t(\mathbf{x}_t,\mathbf{x}_0) &= \left(\frac{\sqrt{\alpha_t}}{\beta_t}\mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_{t-1}}\mathbf{x}_0\right)/\left(\frac{\alpha_t}{\beta_t} + \frac{1}{1-\bar{\alpha}_{t-1}}\right) \\ &= \left(\frac{\sqrt{\alpha_t}}{\beta_t}\mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_{t-1}}\mathbf{x}_0\right)\frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\cdot\beta_t \\ &= \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\mathbf{x}_0\end{aligned}$$

$$\tilde{\boldsymbol{\mu}}_t = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_t\right)$$

$$\boldsymbol{\mu}_\theta(\mathbf{x}_t,t) = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t,t)\right)$$

Thus $\mathbf{x}_{t-1} = \mathcal{N}\left(\mathbf{x}_{t-1}; \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t,t)\right), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t,t)\right)$

$$\begin{aligned}L_t &= \mathbb{E}_{\mathbf{x}_0,\boldsymbol{\epsilon}}\left[\frac{1}{2\|\boldsymbol{\Sigma}_\theta(\mathbf{x}_t,t)\|_2^2}\|\tilde{\boldsymbol{\mu}}_t(\mathbf{x}_t,\mathbf{x}_0) - \boldsymbol{\mu}_\theta(\mathbf{x}_t,t)\|^2\right] \\ &= \mathbb{E}_{\mathbf{x}_0,\boldsymbol{\epsilon}}\left[\frac{1}{2\|\boldsymbol{\Sigma}_\theta\|_2^2}\left\|\frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_t\right) - \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t,t)\right)\right\|^2\right] \\ &= \mathbb{E}_{\mathbf{x}_0,\boldsymbol{\epsilon}}\left[\frac{(1-\alpha_t)^2}{2\alpha_t(1-\bar{\alpha}_t)\|\boldsymbol{\Sigma}_\theta\|_2^2}\|\boldsymbol{\epsilon}_t - \boldsymbol{\epsilon}_\theta(\mathbf{x}_t,t)\|^2\right] \\ &= \mathbb{E}_{\mathbf{x}_0,\boldsymbol{\epsilon}}\left[\frac{(1-\alpha_t)^2}{2\alpha_t(1-\bar{\alpha}_t)\|\boldsymbol{\Sigma}_\theta\|_2^2}\|\boldsymbol{\epsilon}_t - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_t, t)\|^2\right]\end{aligned}$$

$$\begin{aligned}L_t^{\text{simple}} &= \mathbb{E}_{t\sim[1,T],\mathbf{x}_0,\boldsymbol{\epsilon}_t}\left[\|\boldsymbol{\epsilon}_t - \boldsymbol{\epsilon}_\theta(\mathbf{x}_t,t)\|^2\right] \\ &= \mathbb{E}_{t\sim[1,T],\mathbf{x}_0,\boldsymbol{\epsilon}_t}\left[\|\boldsymbol{\epsilon}_t - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_t, t)\|^2\right]\end{aligned}$$

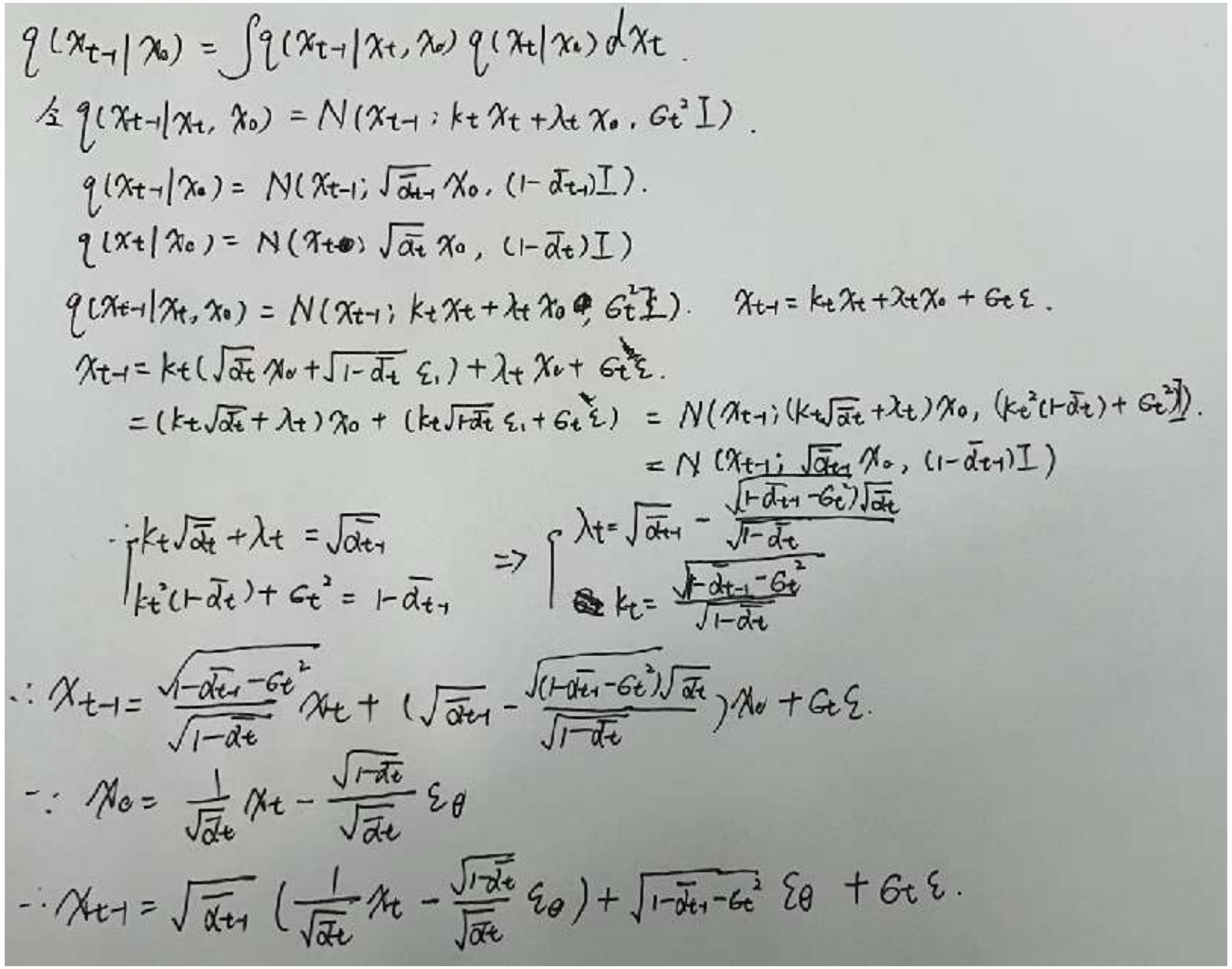
Algorithm 1 Training

```
repeat
    $\mathbf{x}_0 \sim q(\mathbf{x}_0)$
    $t \sim \mathrm{Uniform}(\{1,\dots,T\})$
    $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0},\mathbf{I})$
    Take gradient descent step on
        $\nabla_\theta \|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, t)\|^2$
until converged
```

Algorithm 2 Sampling

```
$\mathbf{x}_T \sim \mathcal{N}(\mathbf{0},\mathbf{I})$
for $t = T,\dots,1$ do
    $\mathbf{z} \sim \mathcal{N}(\mathbf{0},\mathbf{I})$ if $t > 1$, else $\mathbf{z} = \mathbf{0}$
    $\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t,t)\right) + \sigma_t\mathbf{z}$
end for
return $\mathbf{x}_0$
```

Preliminaries-DDIM



$$q(x_{t-1}|x_0) = \int q(x_{t-1}|x_t, x_0)\, q(x_t|x_0)\, dx_t .$$

$$令\ q(x_{t-1}|x_t, x_0) = N(x_{t-1}; k_t x_t + \lambda_t x_0, \sigma_t^2 I).$$

$$q(x_{t-1}|x_0) = N(x_{t-1}; \sqrt{\bar{\alpha}_{t-1}}\, x_0, (1-\bar{\alpha}_{t-1}) I).$$

$$q(x_t|x_0) = N(x_t; \sqrt{\bar{\alpha}_t}\, x_0, (1-\bar{\alpha}_t) I)$$

$$q(x_{t-1}|x_t, x_0) = N(x_{t-1}; k_t x_t + \lambda_t x_0, \sigma_t^2 I). \quad x_{t-1} = k_t x_t + \lambda_t x_0 + \sigma_t \varepsilon .$$

$$x_{t-1} = k_t(\sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1-\bar{\alpha}_t}\, \varepsilon_1) + \lambda_t x_0 + \sigma_t \varepsilon .$$

$$= (k_t\sqrt{\bar{\alpha}_t} + \lambda_t) x_0 + (k_t\sqrt{1-\bar{\alpha}_t}\, \varepsilon_1 + \sigma_t \varepsilon) = N(x_{t-1}; (k_t\sqrt{\bar{\alpha}_t} + \lambda_t) x_0, (k_t^2(1-\bar{\alpha}_t) + \sigma_t^2) I).$$

$$= N(x_{t-1}; \sqrt{\bar{\alpha}_{t-1}}\, x_0, (1-\bar{\alpha}_{t-1}) I)$$

$$\because \begin{cases} k_t\sqrt{\bar{\alpha}_t} + \lambda_t = \sqrt{\bar{\alpha}_{t-1}} \\ k_t^2(1-\bar{\alpha}_t) + \sigma_t^2 = 1-\bar{\alpha}_{t-1} \end{cases} \Rightarrow \begin{cases} \lambda_t = \sqrt{\bar{\alpha}_{t-1}} - \dfrac{\sqrt{(1-\bar{\alpha}_{t-1}-\sigma_t^2)}\sqrt{\bar{\alpha}_t}}{\sqrt{1-\bar{\alpha}_t}} \\ k_t = \dfrac{\sqrt{1-\bar{\alpha}_{t-1}-\sigma_t^2}}{\sqrt{1-\bar{\alpha}_t}} \end{cases}$$

$$\therefore x_{t-1} = \frac{\sqrt{1-\bar{\alpha}_{t-1}-\sigma_t^2}}{\sqrt{1-\bar{\alpha}_t}} x_t + \left(\sqrt{\bar{\alpha}_{t-1}} - \frac{\sqrt{(1-\bar{\alpha}_{t-1}-\sigma_t^2)}\sqrt{\bar{\alpha}_t}}{\sqrt{1-\bar{\alpha}_t}}\right) x_0 + \sigma_t \varepsilon .$$

$$\because x_0 = \frac{1}{\sqrt{\bar{\alpha}_t}} x_t - \frac{\sqrt{1-\bar{\alpha}_t}}{\sqrt{\bar{\alpha}_t}} \varepsilon_\theta$$

$$\therefore x_{t-1} = \sqrt{\bar{\alpha}_{t-1}} \left(\frac{1}{\sqrt{\bar{\alpha}_t}} x_t - \frac{\sqrt{1-\bar{\alpha}_t}}{\sqrt{\bar{\alpha}_t}} \varepsilon_\theta\right) + \sqrt{1-\bar{\alpha}_{t-1}-\sigma_t^2}\, \varepsilon_\theta + \sigma_t \varepsilon .$$

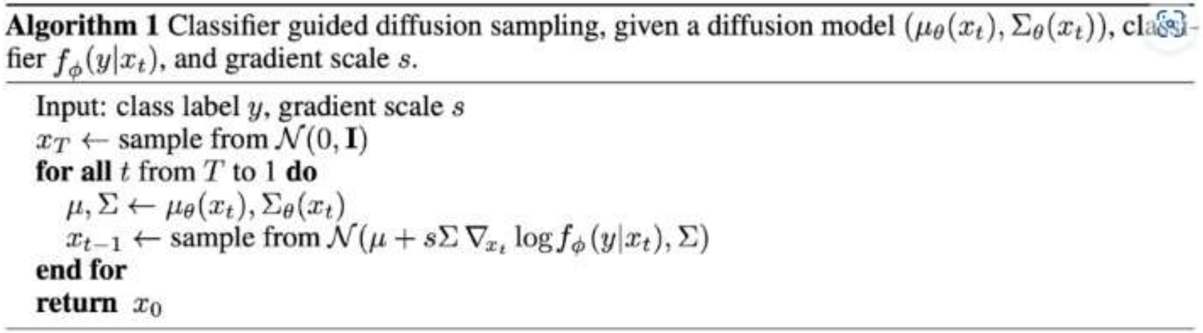
```python
class DDIMSampler:
    @torch.no_grad()
    def sample(S):
        # S 表示 DDIM 的迭代次数
        samples, intermediates = ddim_sampling(S)
        return samples, intermediates

    @torch.no_grad()
    def ddim_sampling(S):
        img = torch.randn((B, C, H, W))
        intermediates = []
        for step in [S - 1, ..., 0]:
            img, pred_x0 = p_sample_ddim(img, step) # DDIM 采样过程
            intermediates.append(pred_x0) # 保存中间结果
        return img, intermediates

    @torch.no_grad()
    def p_sample_ddim(x, t):
        sigma_t, a_t, a_prev = make_ddim_sampling_parameters(eta=0.0) # 返回 σ_t, α_t, α_{t-1}
        e_t = UNet(x, t) # 模型预估噪声 ε_θ^{(t)}(x_t)
        pred_x0 = (x - sqrt_one_minus_at * e_t) / a_t.sqrt() # (x_t - √(1-α_t) ε_θ^{(t)}(x_t)) / √α_t
        dir_xt = (1. - a_prev - sigma_t**2).sqrt() * e_t # √(1 - α_{t-1} - σ_t^2) · ε_θ^{(t)}(x_t)
        noise = sigma_t * noise_like(x.shape) # σ_t ε_t
        # x_{t-1} = √α_{t-1} ((x_t - √(1-α_t) ε_θ^{(t)}(x_t)) / √α_t) + √(1 - α_{t-1} - σ_t^2) · ε_θ^{(t)}(x_t) + σ_t ε_t
        x_prev = a_prev.sqrt() * pred_x0 + dir_xt + noise
        return x_prev, pred_x0

    def make_ddim_sampling_parameters(eta=0.0):
        # σ_{τ_i}(η) = η √((1 - α_{τ_{i-1}}) / (1 - α_{τ_i})) √(1 - α_{τ_i} / α_{τ_{i-1}})
        sigmas = eta * sqrt((1 - alphas_prev) / (1 - alphas) * (1 - alphas / alphas_prev))
        return sigmas, alphas, alphas_prev
```

Preliminaries-guideDM

$$\mathbf{s}_\theta(\mathbf{x}_t,t) \approx \nabla_{\mathbf{x}_t}\log q(\mathbf{x}_t) = \mathbb{E}_{q(\mathbf{x}_0)}[\nabla_{\mathbf{x}_t} q(\mathbf{x}_t|\mathbf{x}_0)] = \mathbb{E}_{q(\mathbf{x}_0)}\left[-\frac{\boldsymbol{\epsilon}_\theta(\mathbf{x}_t,t)}{\sqrt{1-\bar{\alpha}_t}}\right] = -\frac{\boldsymbol{\epsilon}_\theta(\mathbf{x}_t,t)}{\sqrt{1-\bar{\alpha}_t}}$$

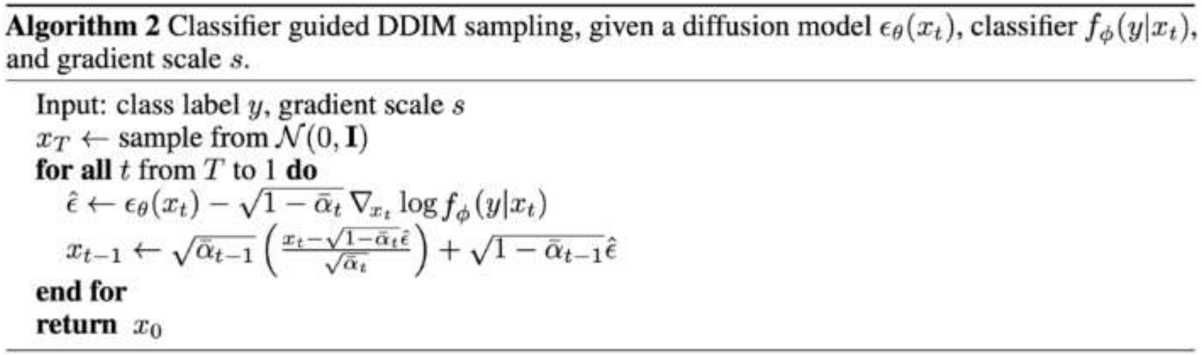
$$\begin{aligned}\nabla_{\mathbf{x}_t}\log q(\mathbf{x}_t,y) &= \nabla_{\mathbf{x}_t}\log q(\mathbf{x}_t) + \nabla_{\mathbf{x}_t}\log q(y|\mathbf{x}_t)\\ &\approx -\frac{1}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t,t) + \nabla_{\mathbf{x}_t}\log f_\phi(y|\mathbf{x}_t)\\ &= -\frac{1}{\sqrt{1-\bar{\alpha}_t}}\left(\boldsymbol{\epsilon}_\theta(\mathbf{x}_t,t) - \sqrt{1-\bar{\alpha}_t}\nabla_{\mathbf{x}_t}\log f_\phi(y|\mathbf{x}_t)\right)\end{aligned}$$

$$\bar{\boldsymbol{\epsilon}}_\theta(\mathbf{x}_t,t) = \boldsymbol{\epsilon}_\theta(x_t,t) - \sqrt{1-\bar{\alpha}_t}\, w\nabla_{\mathbf{x}_t}\log f_\phi(y|\mathbf{x}_t)$$



Algorithm 1 Classifier guided diffusion sampling, given a diffusion model $(\mu_\theta(x_t),\Sigma_\theta(x_t))$, classifier $f_\phi(y|x_t)$, and gradient scale s.

```
Input: class label $y$, gradient scale $s$
$x_T \leftarrow$ sample from $\mathcal{N}(0,\mathbf{I})$
for all $t$ from $T$ to 1 do
    $\mu,\Sigma \leftarrow \mu_\theta(x_t),\Sigma_\theta(x_t)$
    $x_{t-1} \leftarrow$ sample from $\mathcal{N}(\mu + s\Sigma\,\nabla_{x_t}\log f_\phi(y|x_t),\Sigma)$
end for
return $x_0$
```



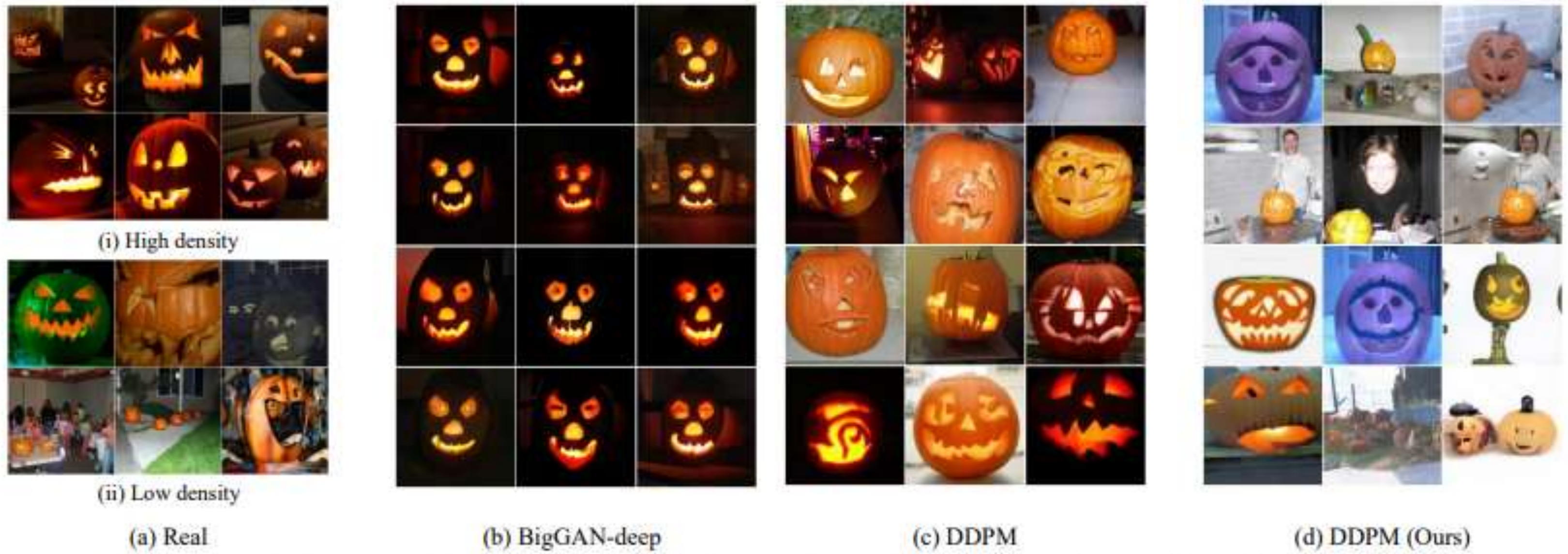
Algorithm 2 Classifier guided DDIM sampling, given a diffusion model $\epsilon_\theta(x_t)$, classifier $f_\phi(y|x_t)$, and gradient scale s.

```
Input: class label $y$, gradient scale $s$
$x_T \leftarrow$ sample from $\mathcal{N}(0,\mathbf{I})$
for all $t$ from $T$ to 1 do
    $\hat{\epsilon} \leftarrow \epsilon_\theta(x_t) - \sqrt{1-\bar{\alpha}_t}\,\nabla_{x_t}\log f_\phi(y|x_t)$
    $x_{t-1} \leftarrow \sqrt{\bar{\alpha}_{t-1}}\left(\frac{x_t-\sqrt{1-\bar{\alpha}_t}\hat{\epsilon}}{\sqrt{\bar{\alpha}_t}}\right) + \sqrt{1-\bar{\alpha}_{t-1}}\hat{\epsilon}$
end for
return $x_0$
```

Motivation

Troubles in Current Diffusion Model:

1. Generate those images lying in the high-density region with high pr
2. Those who wanna generate low-density images will



(a) Real (b) BigGAN-deep (c) DDPM (d) DDPM (Ours)

Method

$$H(\mathbf{x}, y) = \frac{1}{2}\Big[(f(\mathbf{x}) - \mu_y)^T \Sigma_y^{-1} (f(\mathbf{x}) - \mu_y) + \ln(\det(\Sigma_y)) + k\ \ln(2\pi)\Big] \quad (3)$$

$$L_{g_1}(\mathbf{x}_i, y_i) = \log\left[\frac{exp(H(\mathbf{x}_i, y_i)/\tau)}{\sum_{j=1}^{C} exp(H(\mathbf{x}_j, j)/\tau)}\right] \quad (4)$$

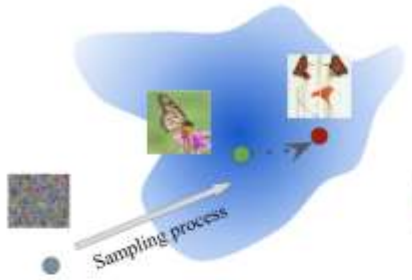
$$L_{g_2}(\mathbf{x}_i) = -\log\left[\frac{exp(H'(\mathbf{x}_i, 1)/\tau)}{\sum_{j=0}^{1} exp(H'(\mathbf{x}_j, j)/\tau)}\right] \quad (6)$$

$$\begin{aligned} \mathbf{x}_{t-1} = {} & \boldsymbol{\mu}_\theta(\mathbf{x}_t, t) + \boldsymbol{\Sigma}_\theta^{1/2}(\mathbf{x}_t, t)\,\mathbf{z} \\ & + \alpha\,\boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t)\,\nabla^* L_{g_1}(\mathbf{x}_t, y) \\ & + \beta\,\boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t)\,\nabla^* L_{g_2}(\mathbf{x}_t) \end{aligned} \quad (7)$$

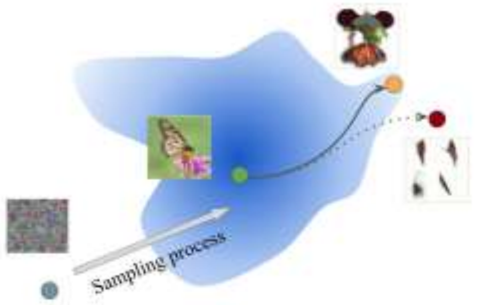
Algorithm 1: Sampling from low-density regions.

```
Input : Class label $(y), \alpha, \beta$
Function : $Normalize\ (\mathbf{u})$ : return $\mathbf{u}/\|\mathbf{u}\|$
$\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
for $i \leftarrow T$ to $1$ do
    if $t > 1$ then
        $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \mathbf{s} \leftarrow \mathbf{I}$
    else
        $\mathbf{z} \leftarrow \mathbf{0}, \mathbf{s} \leftarrow \mathbf{0}$
    end
    $\mathbf{u}_1 = \alpha\,\boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t)\ Normalize\ (\nabla L_{g_1}(\mathbf{x}_t, y))$
    $\mathbf{u}_2 = \beta\,\boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t)\ Normalize\ (\nabla L_{g_2}(\mathbf{x}_t))$
    $\mathbf{x}_{t-1} = \boldsymbol{\mu}_\theta(\mathbf{x}_t, t) + \boldsymbol{\Sigma}_\theta^{1/2}(\mathbf{x}_t, t)\,\mathbf{z} + \mathbf{s}(\mathbf{u}_1 + \mathbf{u}_2)$
end
return $\mathbf{x}_0$
```

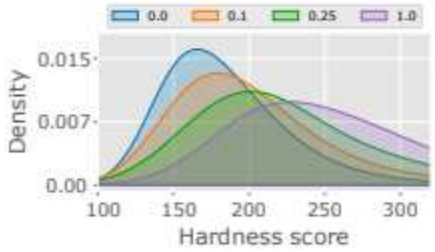
Experiment



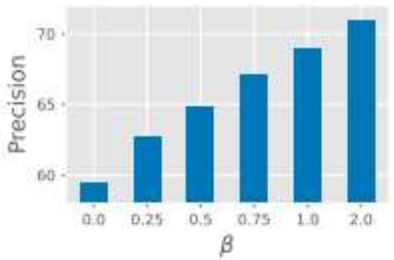
(a) Small α (●)



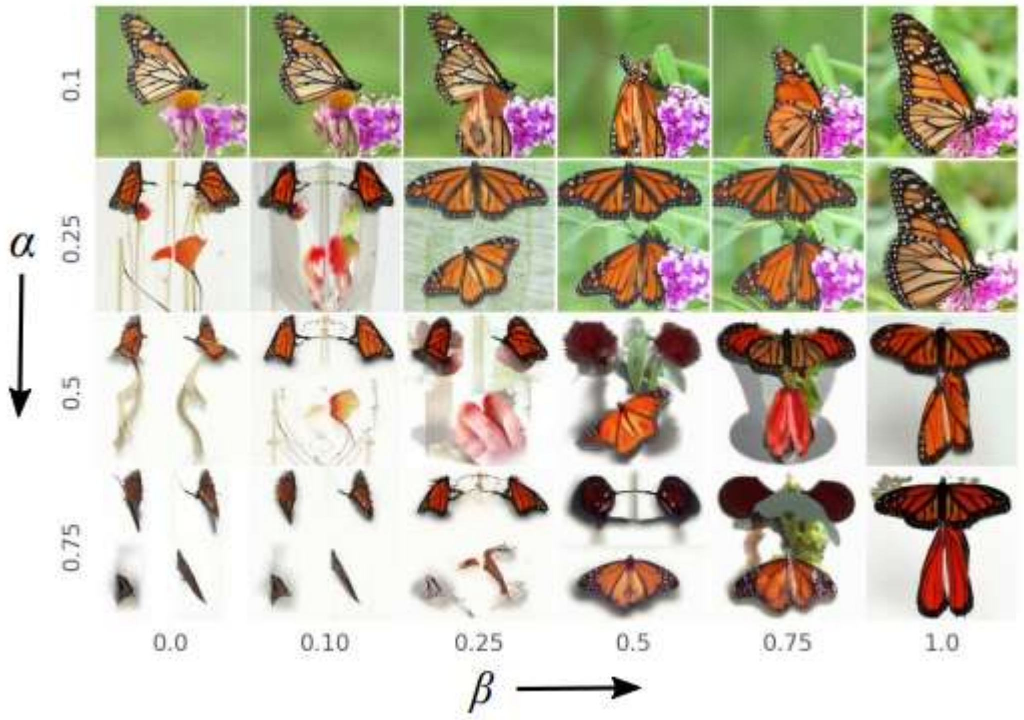
(b) Large α (●), With β (●)



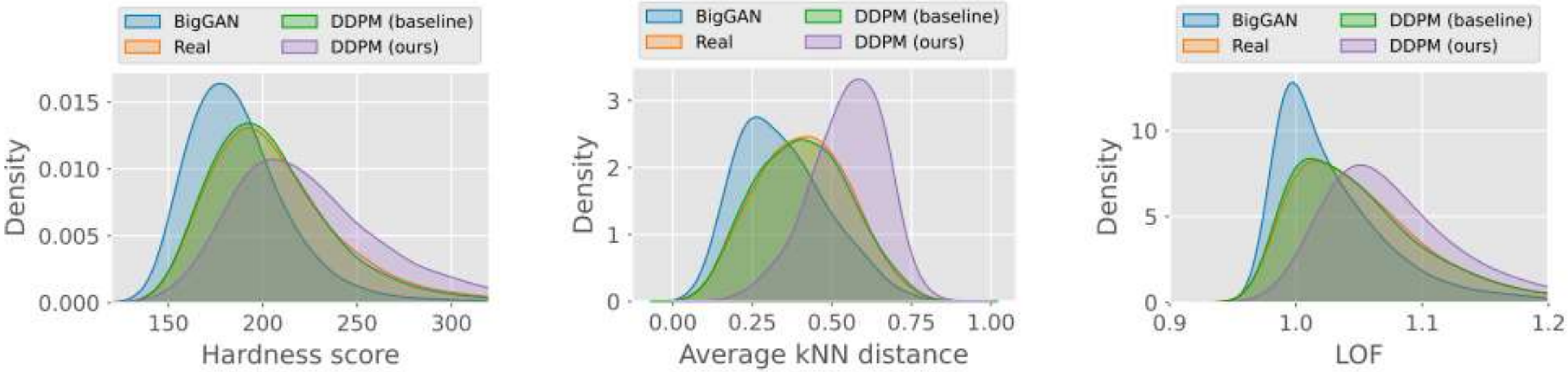
(a) α increases the hardness score.



(b) β improves the fidelity



Experiment

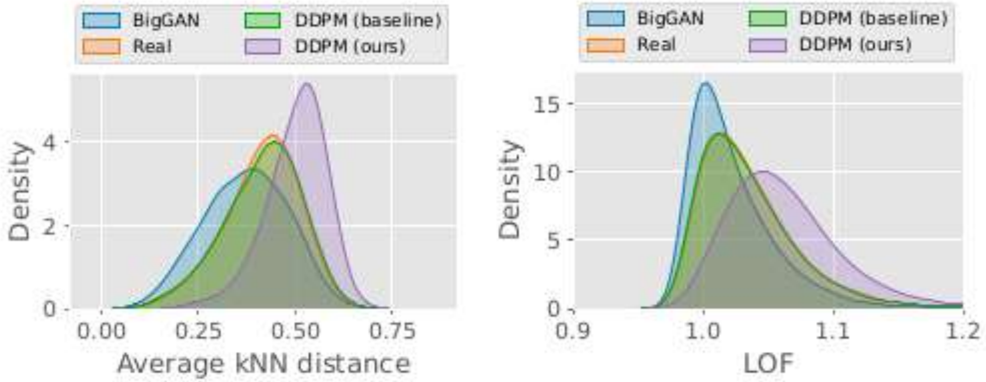


(a) Hardness score (b) AvgkNN distance (c) Local outlier factor

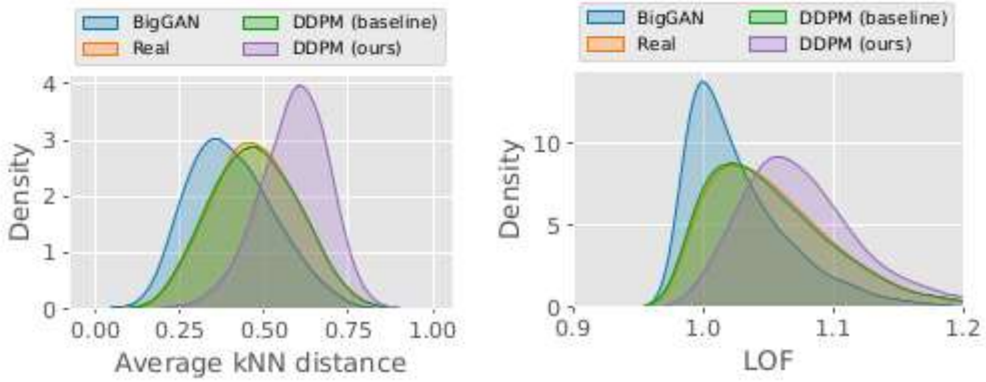
Figure 7. **Comparing neighborhood density.** We measure the density in the neighborhood of a given set of instances using three different metrics. All three metrics share a common trend: while baseline sampling generates synthetic samples that have similar density distribution as real data, our sampling process generates samples from low density neighborhoods with higher probability.

Score-range	200 – 240	240 – 280	280 – 320
Baseline	1.99	5.74	16.79
Ours	1.88 (×**1.1**)	2.03 (×**2.8**)	2.78 (×**6.0**)

Experiment



(a) Inception-v3



(b) VGG19

Figure 17. **Comparing neighborhood density across different choices of feature extractors.** We use two additional feature extractors, namely Inception-v3 and VGG19.

Experiment

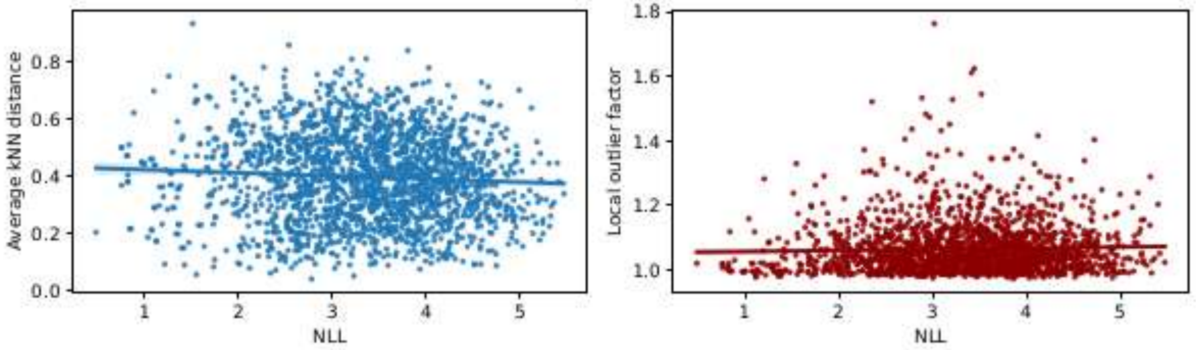
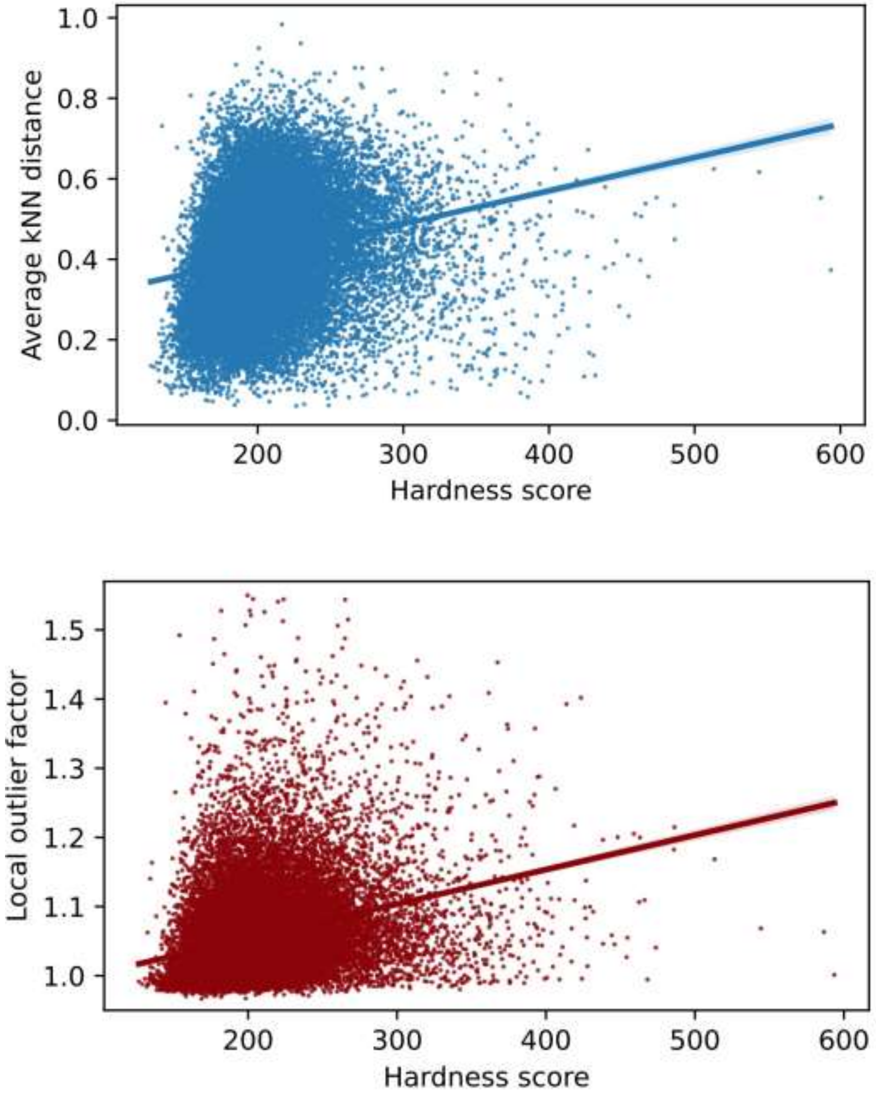
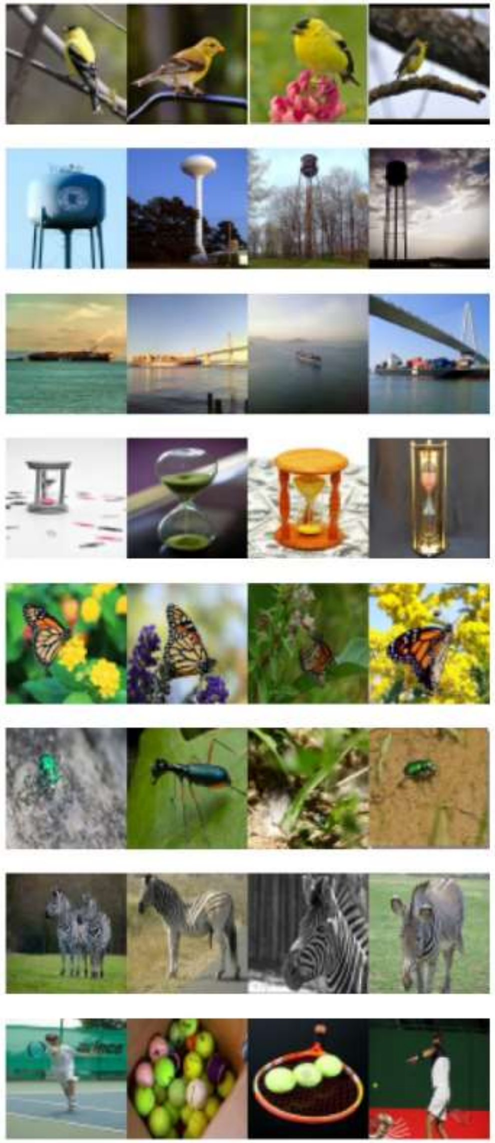


Figure 11. **Is NLL an effective measure of neighborhood density?** We compare the negative log-likelihood (NLL) estimates from the diffusion model with other commonly used metrics to measure data density. We find that NLL is poorly correlated with both of these metrics. Since NLL is computationally expensive to calculate for each image, we use 2K random images from the validation set of the ImageNet dataset for this analysis.

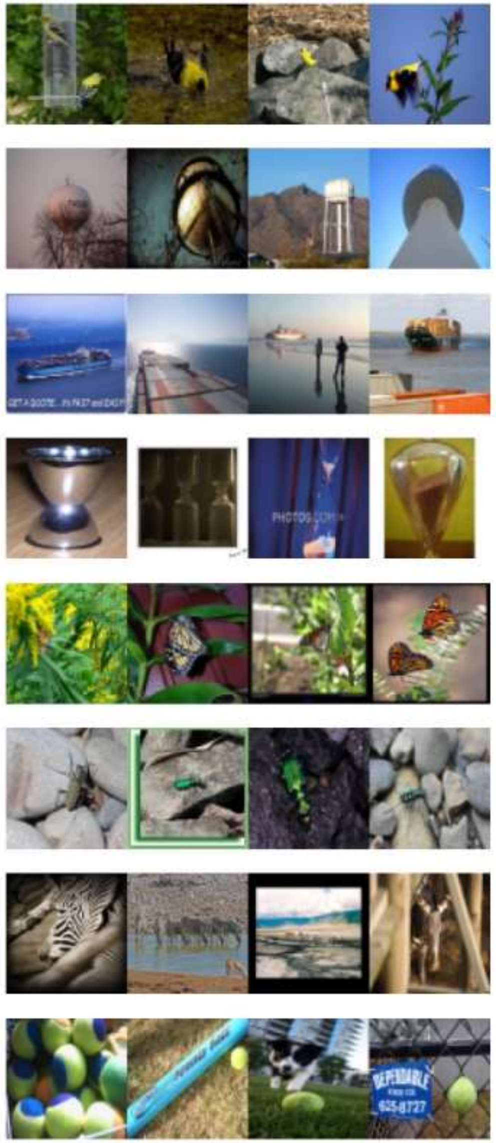


Experiment

$H(x) < H_{15}$

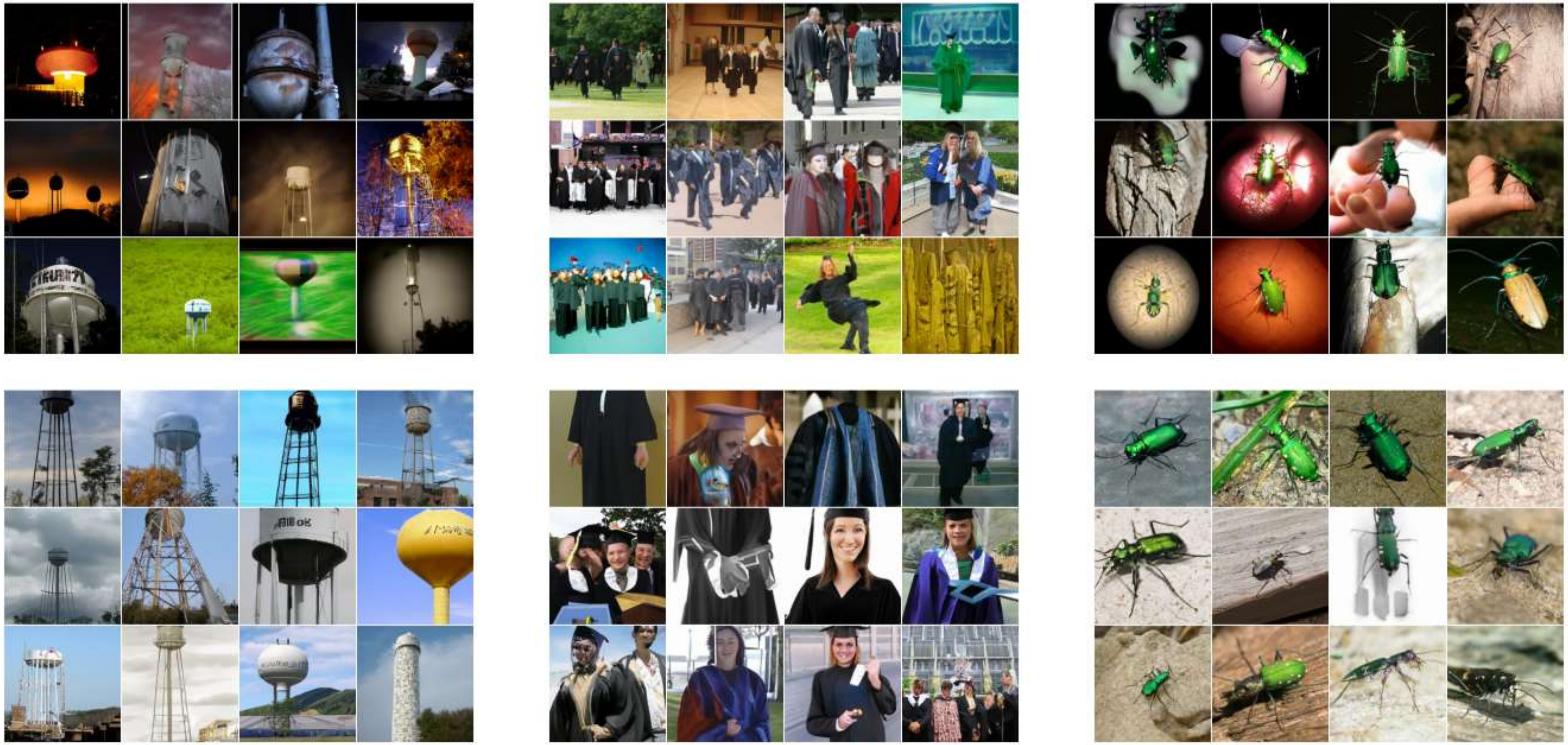


$H_{40} < H(x) < H_{60}$



$H(x) > H_{90}$

Experiment



(a) Water tower (b) Academic robe (c) Tiger beetle

Figure 5. **Comparing samples from proposed and baseline sampling process.** We compare synthetic images from our proposed sampling approach (top) with the baseline sampling process (bottom) on the ImageNet dataset. We use identical random seed for both stochastic sampling processes. Therefore, generation of each pair of images among the two approaches starts from the identical latent vectors and the only difference is the additional guidance terms in our approach.

Thanks