Diffusion Models Beat GANs on Image Synthesis

김기웅
kwkim02@g.skku.edu
CV Core
2024/11/05

TRAIN AND TEST

Contents

- Introduction
- Background
- Classifier Guidance
- Results
- Limitations

TNT
TRAIN AND TEST

Introduction

TRAIN AND TEST

Introduction

- Only GAN can be conditional with class input

  → Conditional diffusion model

- GAN's benefit : trade off diversity and fidelity

  → bring it to diffusion models

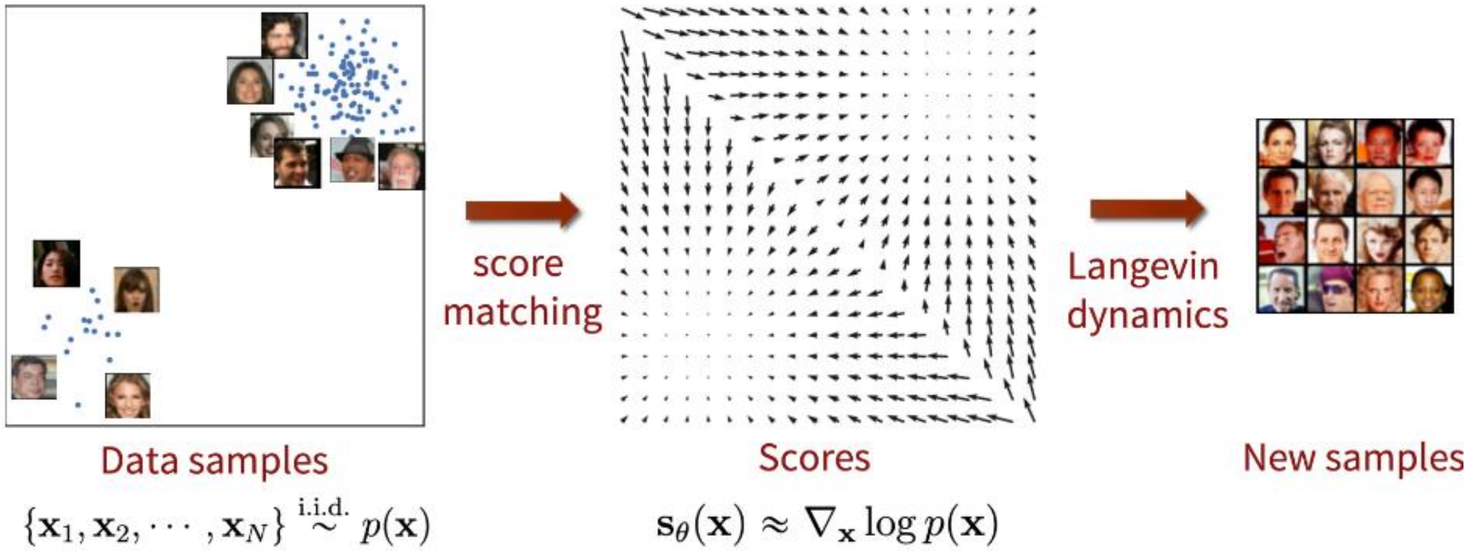
TRAIN AND TEST

Background

Background

Score-Matching

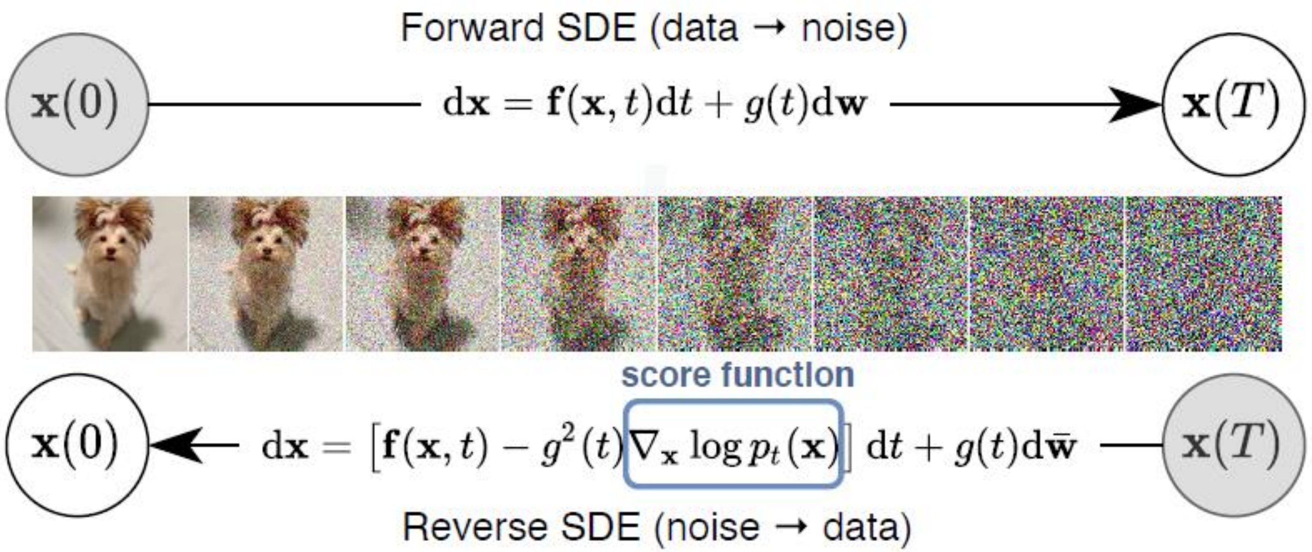
Score = Gradient of log(pdf)
$= \nabla_x logp(x)$



TRAIN AND TEST

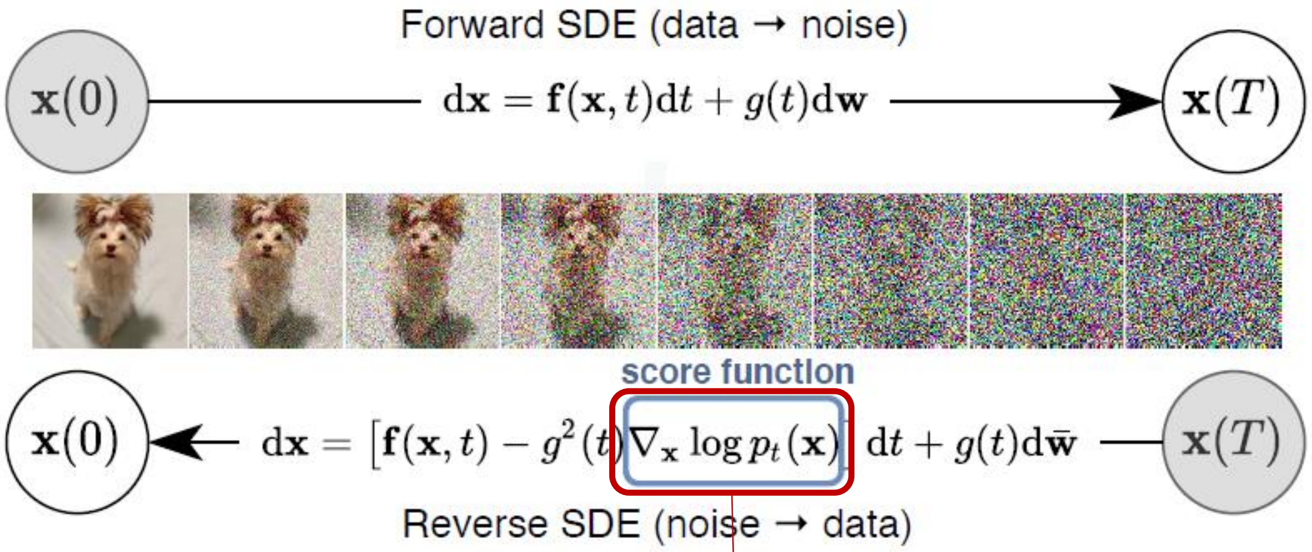
Background

Score based Generative Model through SDE



Background

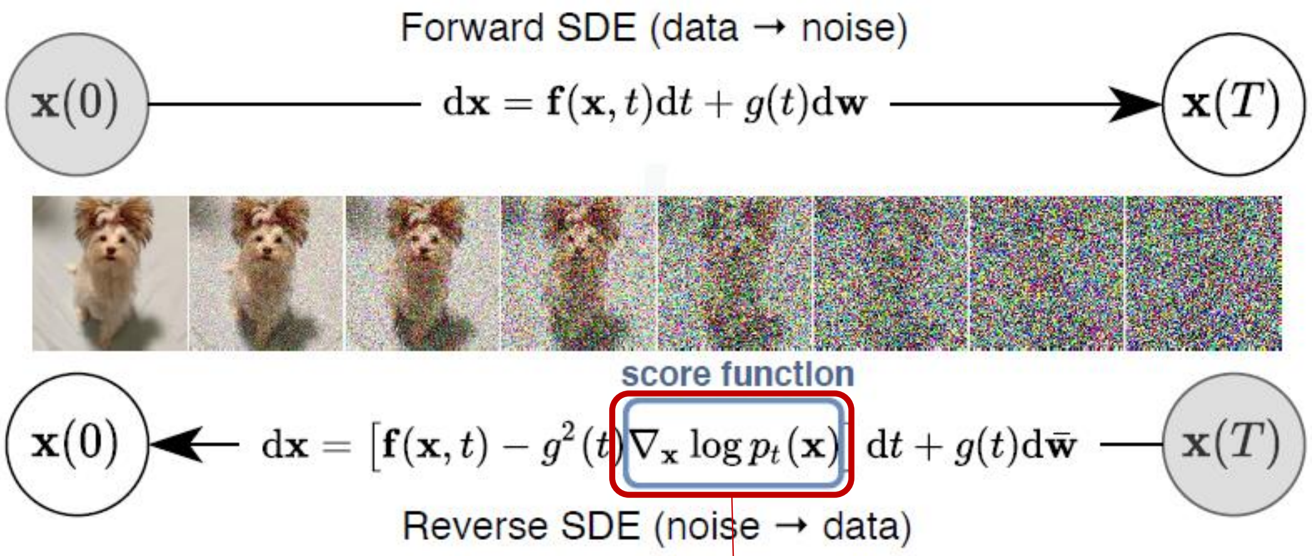
Classifier Guidance



Classifier Guidance $\boxed{\nabla_{x_t} \log(p_\theta(x_t|y))}$

Background

Classifier Guidance



Classifier Guidance $\boxed{\nabla_{x_t}\log(p_\theta(x_t|y))}$

class

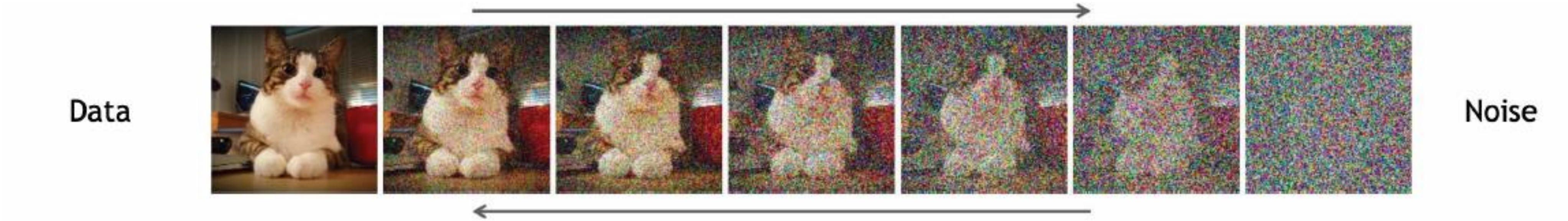
+ discrete time

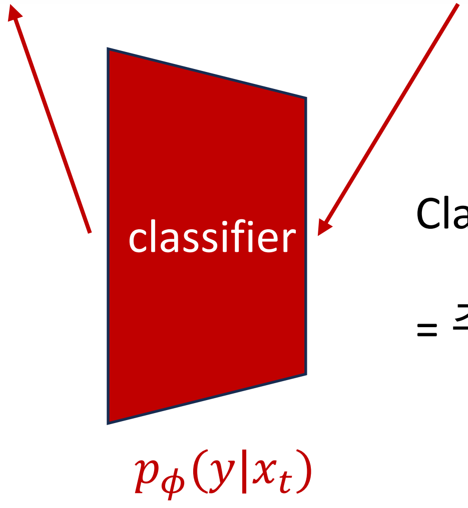
Classifier Guidance

TRAIN AND TEST

Classifier Guidance

Overview





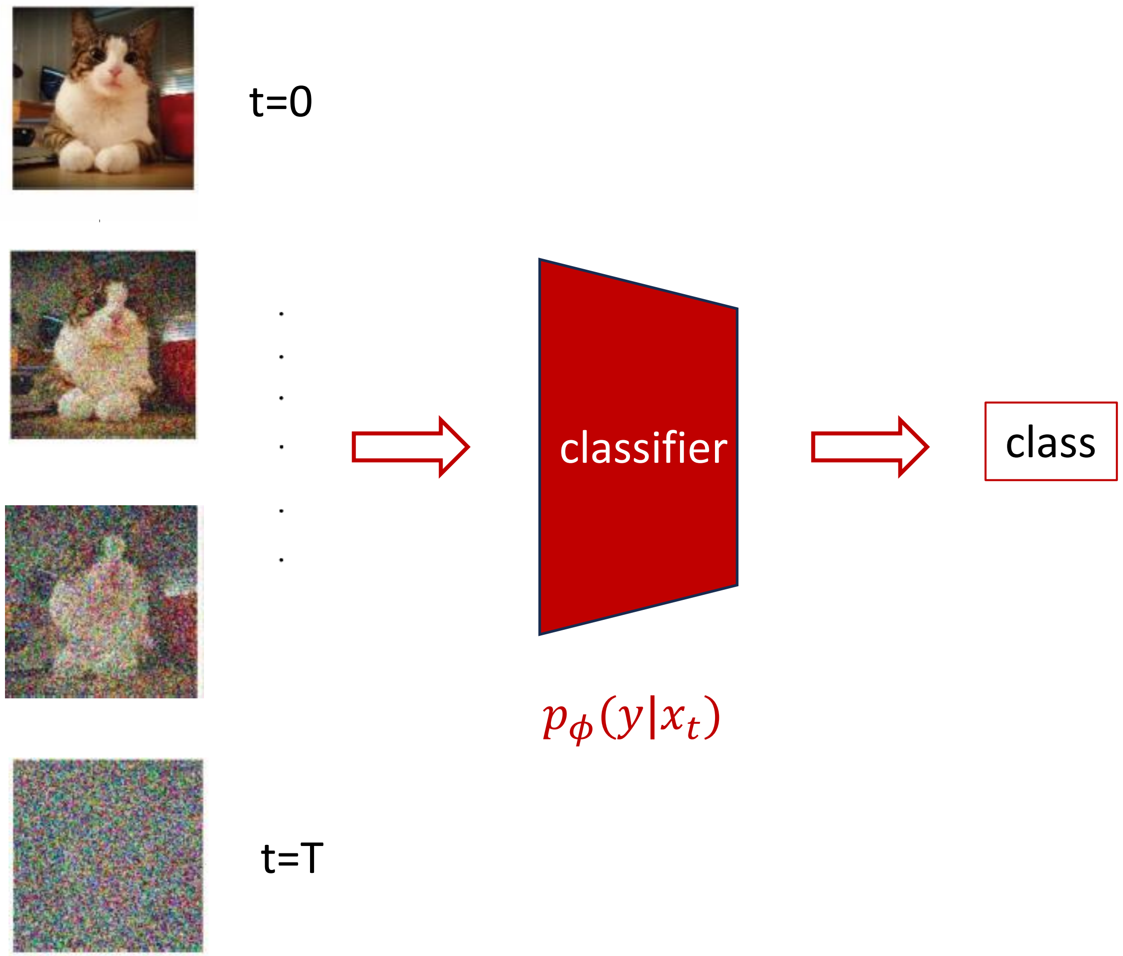
Classifier가 reverse process에 score 역할

= 주어진 class를 생성하도록 reconstruct 방향 조절

Classifier Guidance

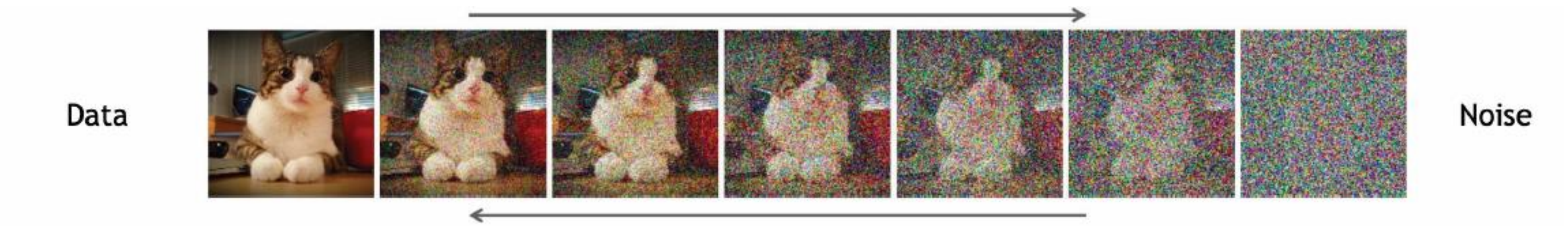
Training Classifier

- CNN based model
- Input noisy image
  with timestep embedding
- Predict class of given image



Classifier Guidance

Conditional Sampling for DDPM



x_t $\quad\quad$ x_{t+1}

$$\boxed{p_{\theta,\phi}(x_t|x_{t+1}, y)}$$

Classifier Guidance

Conditional Sampling for DDPM

$$\boxed{p_{\theta,\phi}(x_t|x_{t+1},y)} = Z\underbrace{p_\theta(x_t|x_{t+1})}_{\text{noise}}\underbrace{p_\phi(y|x_t)}_{\text{classifier}}$$

noise

$$p_\theta(x_t|x_{t+1}) = \mathcal{N}(\mu,\Sigma)$$
$$\log p_\theta(x_t|x_{t+1}) = -\frac{1}{2}(x_t-\mu)^T\Sigma^{-1}(x_t-\mu)+C$$

classifier

$$\log p_\phi(y|x_t) \approx \log p_\phi(y|x_t)|_{x_t=\mu} + (x_t-\mu)\nabla_{x_t}\log p_\phi(y|x_t)|_{x_t=\mu}$$
$$= (x_t-\mu)g + C_1$$

$g = \nabla_{x_t}\log p_\phi(y|x_t)$: Score of Classifier

$$\log(p_\theta(x_t|x_{t+1})p_\phi(y|x_t)) \approx -\frac{1}{2}(x_t-\mu)^T\Sigma^{-1}(x_t-\mu)+(x_t-\mu)g+C_2$$
$$= -\frac{1}{2}(x_t-\mu-\Sigma g)^T\Sigma^{-1}(x_t-\mu-\Sigma g)+C_3$$
$$= \log p(z) + C_4, z\sim\mathcal{N}(\mu+\Sigma g,\Sigma)$$

TNT TRAIN AND TEST

Classifier Guidance

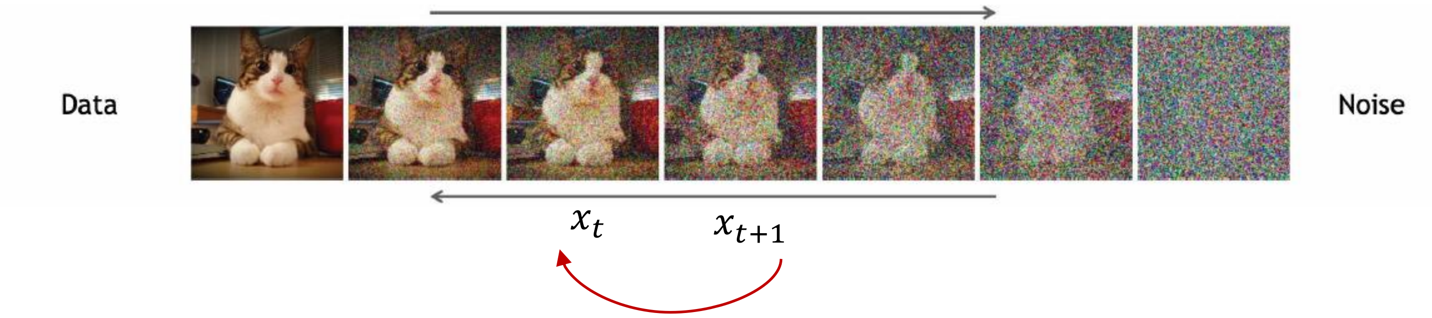
Conditional Sampling for DDPM

$$\boxed{\log(p_\theta(x_t|x_{t+1})p_\phi(y|x_t)) = \log p(z) + C_4, z \sim \mathcal{N}(\mu + \Sigma g, \Sigma)}$$

$$g = \nabla_{x_t} \log p_\phi(y|x_t)$$

Algorithm 1 Classifier guided diffusion sampling, given a diffusion model $(\mu_\theta(x_t), \Sigma_\theta(x_t))$, classifier $p_\phi(y|x_t)$, 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 $\mathcal{N}(\mu + s\Sigma\, \nabla_{x_t} \log p_\phi(y|x_t), \Sigma)$
end for
return x_0



Classifier Guidance

Conditional Sampling for DDIM

- Stochastic diffusion process cannot be applied to deterministic sampling methods like DDIM.

$$
\begin{aligned}
\nabla_{x_t} \log(p_\theta(x_t) p_\phi(y|x_t)) &= \nabla_{x_t} \log p_\theta(x_t) + \nabla_{x_t} \log p_\phi(y|x_t) \\
&= -\frac{1}{\sqrt{1-\bar{\alpha}_t}} \underbrace{\epsilon_\theta(x_t)}_{\text{Noise prediction model}} + \underbrace{\nabla_{x_t} \log p_\phi(y|x_t)}_{\text{Score of classifier}}
\end{aligned}
$$

TRAIN AND TEST

Classifier Guidance

Conditional Sampling for DDIM

Algorithm 2 Classifier guided DDIM sampling, given a diffusion model $\epsilon_\theta(x_t)$, classifier $p_\phi(y|x_t)$, 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**
  $\hat{\epsilon} \leftarrow \epsilon_\theta(x_t) - \sqrt{1-\bar{\alpha}_t}\,\nabla_{x_t} \log p_\phi(y|x_t)$
  $x_{t-1} \leftarrow \sqrt{\bar{\alpha}_{t-1}}\left(\frac{x_t - \sqrt{1-\bar{\alpha}_t}\hat{\epsilon}}{\sqrt{\bar{\alpha}_t}}\right) + \sqrt{1-\bar{\alpha}_{t-1}}\hat{\epsilon}$
end for
return x_0

TRAIN AND TEST

Classifier Guidance

Scaling Classifier Gradients

trade off

Conditional	Guidance	Scale	FID	sFID	IS	Precision	Recall
✗	✗		26.21	**6.35**	39.70	0.61	0.63
✗	✓	1.0	33.03	6.99	32.92	0.56	**0.65**
✗	✓	10.0	**12.00**	10.40	**95.41**	**0.76**	0.44
✓	✗		10.94	6.02	100.98	0.69	**0.63**
✓	✓	1.0	**4.59**	**5.25**	186.70	0.82	0.52
✓	✓	10.0	9.11	10.93	**283.92**	**0.88**	0.32

Table 4: Effect of classifier guidance on sample quality. Both conditional and unconditional models were trained for 2M iterations on ImageNet 256×256 with batch size 256.

TRAIN AND TEST

Classifier Guidance

Scaling Classifier Gradients

Figure 3: Samples from an unconditional diffusion model with classifier guidance to condition on the class "Pembroke Welsh corgi". Using classifier scale 1.0 (left; FID: 33.0) does not produce convincing samples in this class, whereas classifier scale 10.0 (right; FID: 12.0) produces much more class-consistent images.

TRAIN AND TEST

Result

TRAIN AND TEST

Result

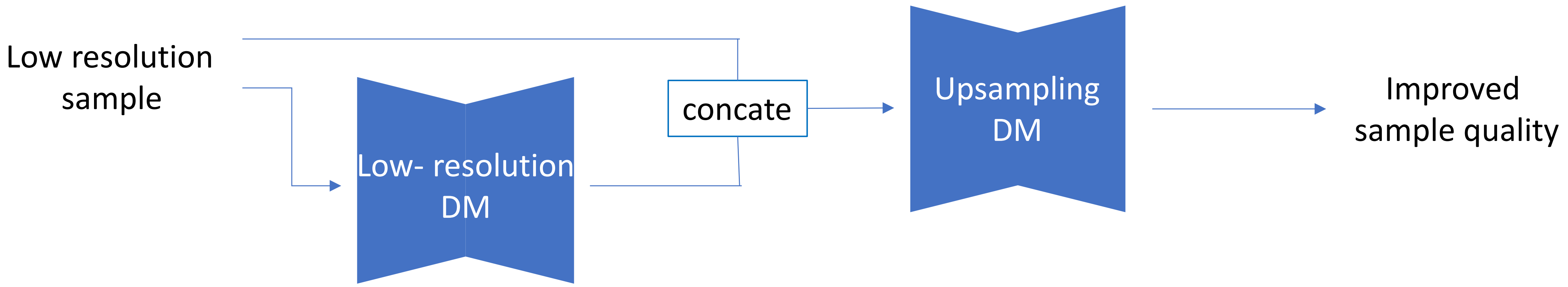
SOTA image Synthesis

Figure 6: Samples from BigGAN-deep with truncation 1.0 (FID 6.95, left) vs samples from our diffusion model with guidance (FID 4.59, middle) and samples from the training set (right).

TRAIN AND TEST

Result

Comparision to Upsampling



Limitation

TRAIN AND TEST

Limitation

- Slower than GAN at sampling time
  (multiple denoising step)

- Limited to labeled dataset

  → no effective strategy for trading off diversity and fidelity

TRAIN AND TEST

TRAIN AND TEST