EGSDE: Unpaired Image-to-Image Translation via Energy-Guided Stochastic Differential Equations

Min Zhao, Fan Bao, Chongxuan Li and Jun Zhu

Outline

- Score-based diffusion models (SBDM)
- Energy-guided SDE (EGSDE) and its application in (Unpaired) image-to-image (I2I)
- EGSDE in a big picture

Denoising diffusion probabilistic models

Jonathon et al., NeurIPS 2020

Forward diffusion: a Markov chain with Gaussian kernel

$$q\left(\mathbf{x}^{(0)}\right) \longrightarrow q\left(\mathbf{x}^{(T)}\right) \approx \mathcal{N}\left(\mathbf{x}^{(T)}; 0, \mathbf{I}\right)$$

$$q\left(\mathbf{x}^{(t)}|\mathbf{x}^{(t-1)}\right) = \mathcal{N}\left(\mathbf{x}^{(t)}; \mathbf{x}^{(t-1)}\sqrt{1-\beta_t}, \mathbf{I}\beta_t\right)$$

Decay towards origin — Add small noise

Data distribution — Gaussian noise

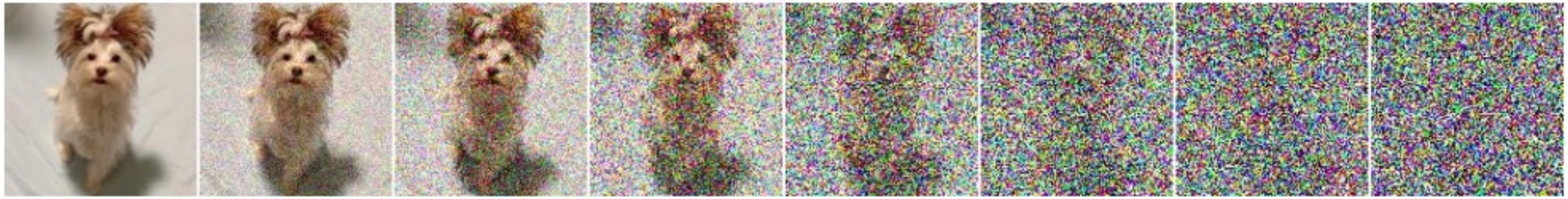


Image credit: *Sohl-Dickstein*

Denoising diffusion probabilistic models

Jonathon et al., NeurIPS 2020

Backward diffusion: a Markov chain with Gaussian kernel

$$p\left(\mathbf{x}^{(0)}\right) \approx q\left(\mathbf{x}^{(0)}\right) \longleftarrow p\left(\mathbf{x}^{(T)}\right) = \mathcal{N}\left(\mathbf{x}^{(T)}; 0, \mathbf{I}\right)$$

$$p\left(\mathbf{x}^{(t-1)}|\mathbf{x}^{(t)}\right) = \mathcal{N}\left(\mathbf{x}^{(t-1)}; \underline{f_\mu\left(\mathbf{x}^{(t)}, t\right), f_\Sigma\left(\mathbf{x}^{(t)}, t\right)}\right)$$

Learned drift and covariance functions

Data distribution

Gaussian noise

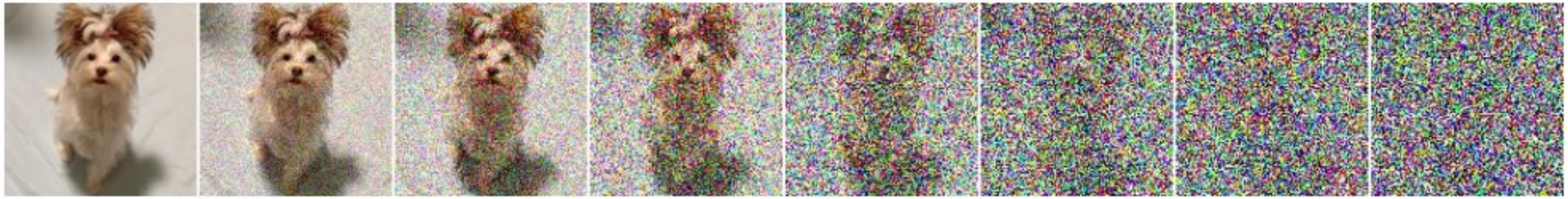


Image credit: *Sohl-Dickstein*

From finite steps to infinite steps

Song et al, ICLR 2021

$$q(x_i|x_{i-1}) = \mathcal{N}(\sqrt{1-\beta_i}x_{i-1}, \beta_i I)$$

Reparameterization

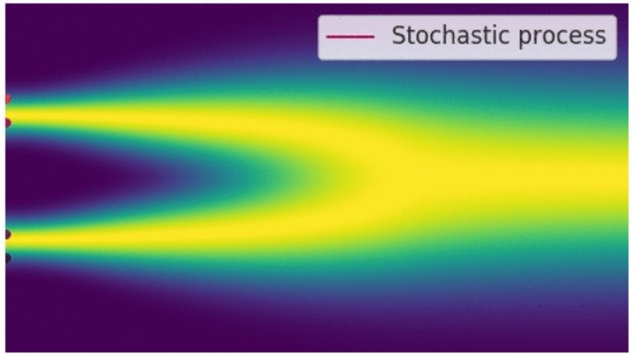
$$\mathbf{x}_i = \sqrt{1-\beta_i}\mathbf{x}_{i-1} + \sqrt{\beta_i}\mathbf{z}_{i-1}, \quad i = 1, \cdots, N, \quad \mathbf{z}_{i-1} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

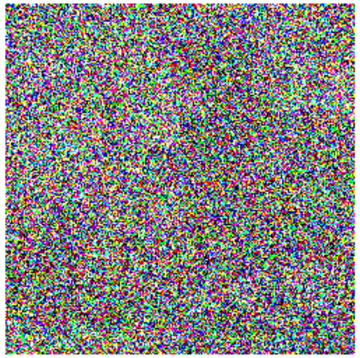
Rescaling by N $\left(\Delta t = \frac{1}{N}\right)$

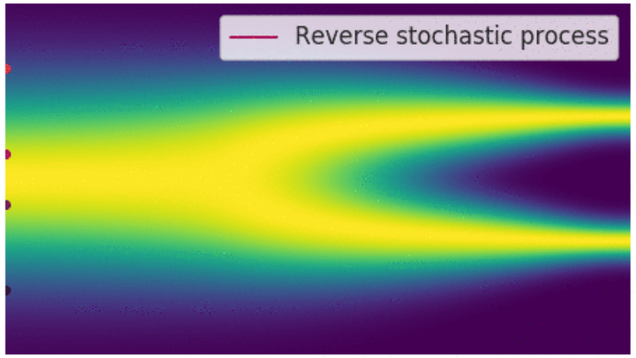
$$\mathbf{x}(t+\Delta t) = \sqrt{1-\beta(t+\Delta t)\Delta t}\,\mathbf{x}(t) + \sqrt{\beta(t+\Delta t)\Delta t}\,\mathbf{z}(t)$$

Limit of $\Delta t \to 0$

$$\mathrm{d}\mathbf{x} = -\frac{1}{2}\beta(t)\mathbf{x}\,\mathrm{d}t + \sqrt{\beta(t)}\,\mathrm{d}\mathbf{w}$$ VP-SDE







From finite steps to infinite steps

Song et al, ICLR 2021

- Forward time

$$\mathrm{d}\mathbf{x} = \mathbf{f}(\mathbf{x}, t)\mathrm{d}t + g(t)\mathrm{d}\mathbf{w}$$

- Reverse time

$$\mathrm{d}\boldsymbol{x} = [\boldsymbol{f}(\boldsymbol{x}, t) - g(t)^2 \boldsymbol{s}(\boldsymbol{x}, t)]\mathrm{d}t + g(t)\mathrm{d}\overline{\boldsymbol{w}}$$

$$\mathrm{d}\boldsymbol{x} = [\boldsymbol{f}(\boldsymbol{x}, t) - \frac{g(t)^2}{\sigma_t} \boldsymbol{\epsilon}_\theta(\boldsymbol{x}, t)]\mathrm{d}t + g(t)\mathrm{d}\overline{\boldsymbol{w}}$$

Energy-guided SDE (EGSDE) and its application in (Unpaired) image-to-image (I2I)

Unpaired Image-to-Image Translation

source

target

Training

source image

$\boldsymbol{x}_0$

$p(\boldsymbol{y}_0|\boldsymbol{x}_0)$

translated image

Inference

What is good translation?

$p(\boldsymbol{y}_0|\boldsymbol{x}_0)$

source image

$\boldsymbol{x}_0$

translated image

What is good translation?

source image

$\boldsymbol{x}_0$

translated image

- ➢ Be ***realistic*** for the target domain by changing the domain-specific features

What is good translation?

source image $\boldsymbol{x}_0$

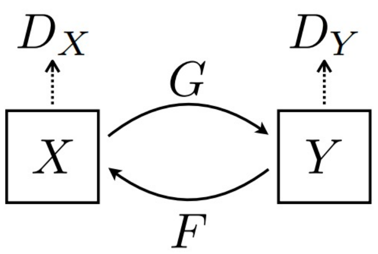
translated image

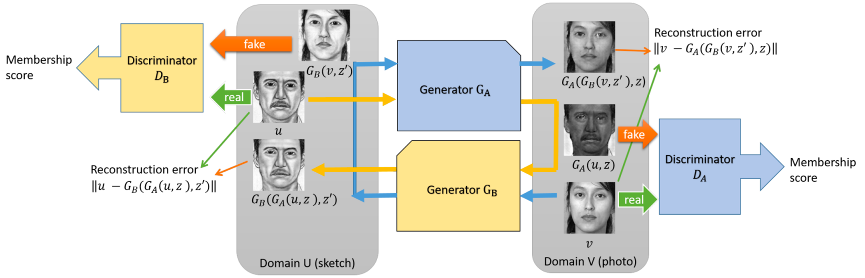
- Be ***realistic*** for the target domain by changing the domain-specific features
- Be ***faithful*** for the source image by preserving the domain-independent features

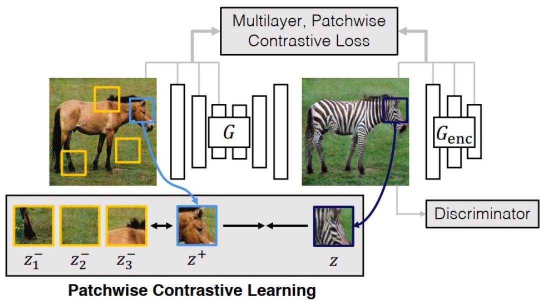
GANs-based methods dominated this field due to their ability to generate high-quality samples

GANs-based methods

- two-side mapping
  - CycleGAN (Zhu et al. , ICCV 2017)
  - DualGAN (Yi et al., ICCV 2017)
  - U-GAT-IT (Kim et al., ICLR 2019)
  - ……
- one-side mapping
  - GcGAN (Fu et al., CVPR 2019)
  - CUT (Park et al., ECCV 2020)
  - F-LSeSim (Zheng et al., CVPR 2021)
  - ……







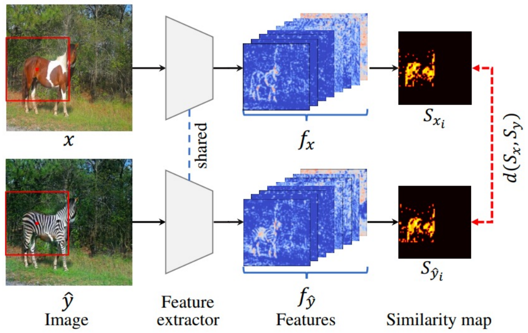
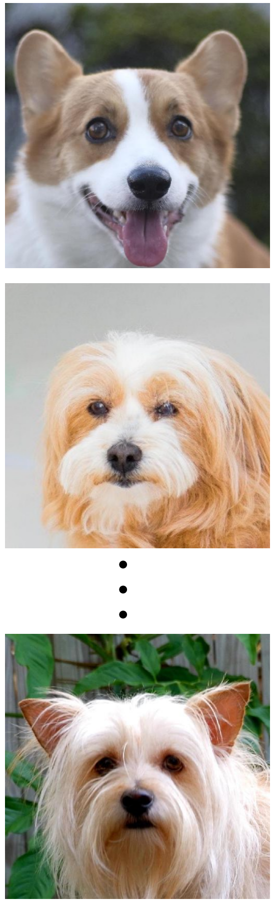
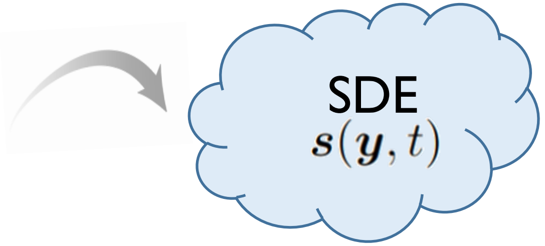


Image-to-Image Translation: Methods and Applications, Pang et al. 2021 ;The Spatially-Correlative Loss for Various Image Translation Tasks, Zheng et al. 2021

Motivation

Choi et al. ICCV 2021; Meng et al, ICLR 2022;



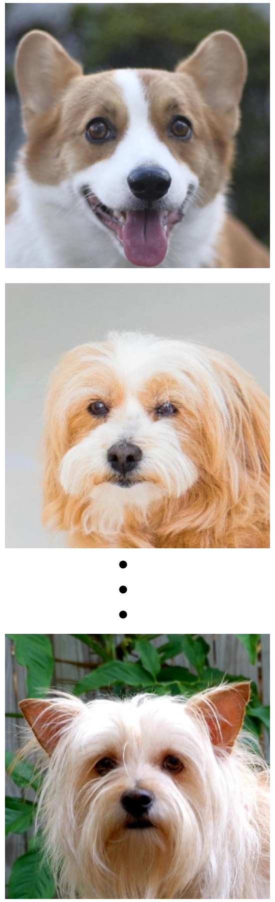


target

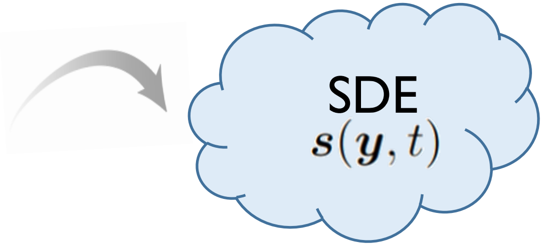
Training same as SDE

Motivation

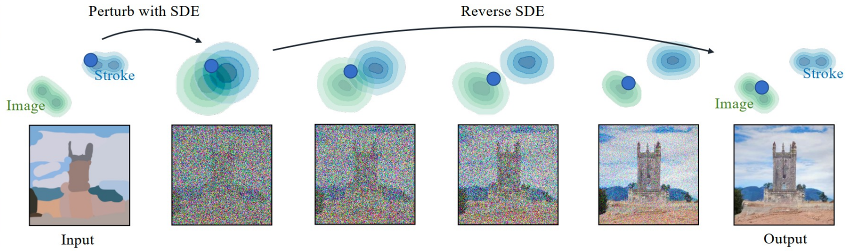
Choi et al. ICCV 2021; Meng et al, ICLR 2022;



target

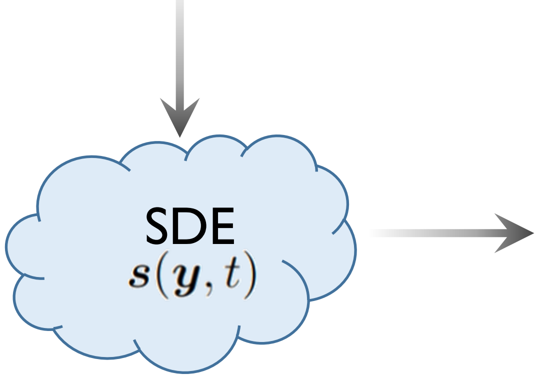


Training



source image

x_0

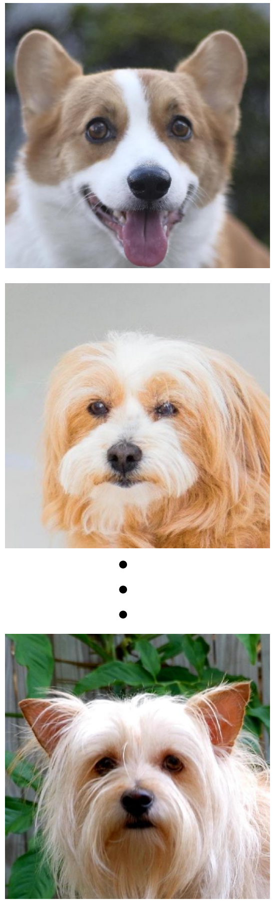


translated image

Translation only in inference

Motivation

Choi et al. ICCV 2021; Meng et al, ICLR 2022;



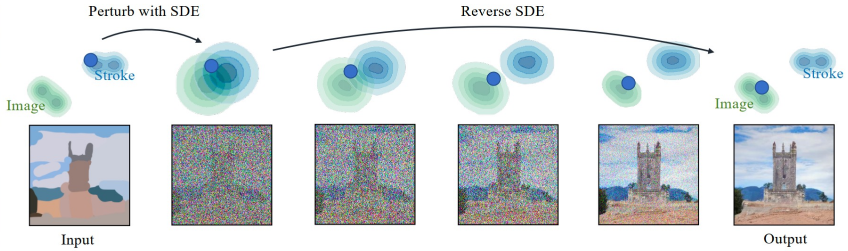
target

SDE
$s(\boldsymbol{y}, t)$

Training

source image

$\boldsymbol{x}_0$



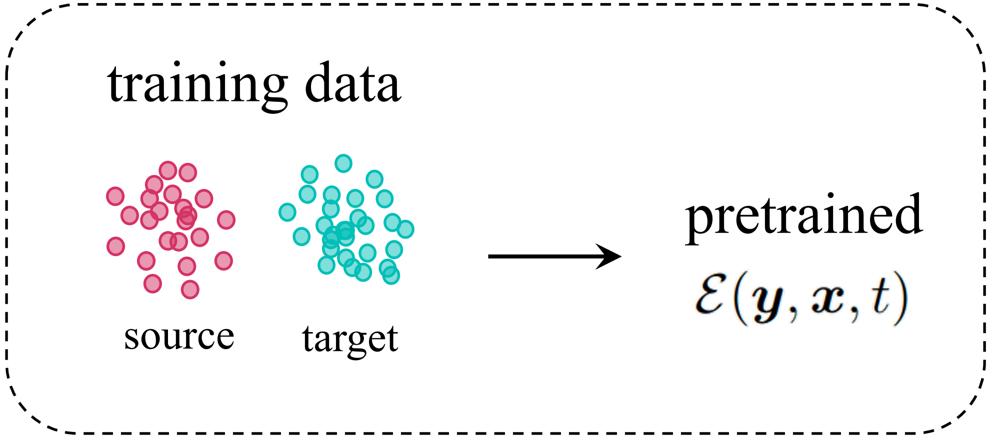
translated image

Did not leverage the training data in the source domain at all

Inference

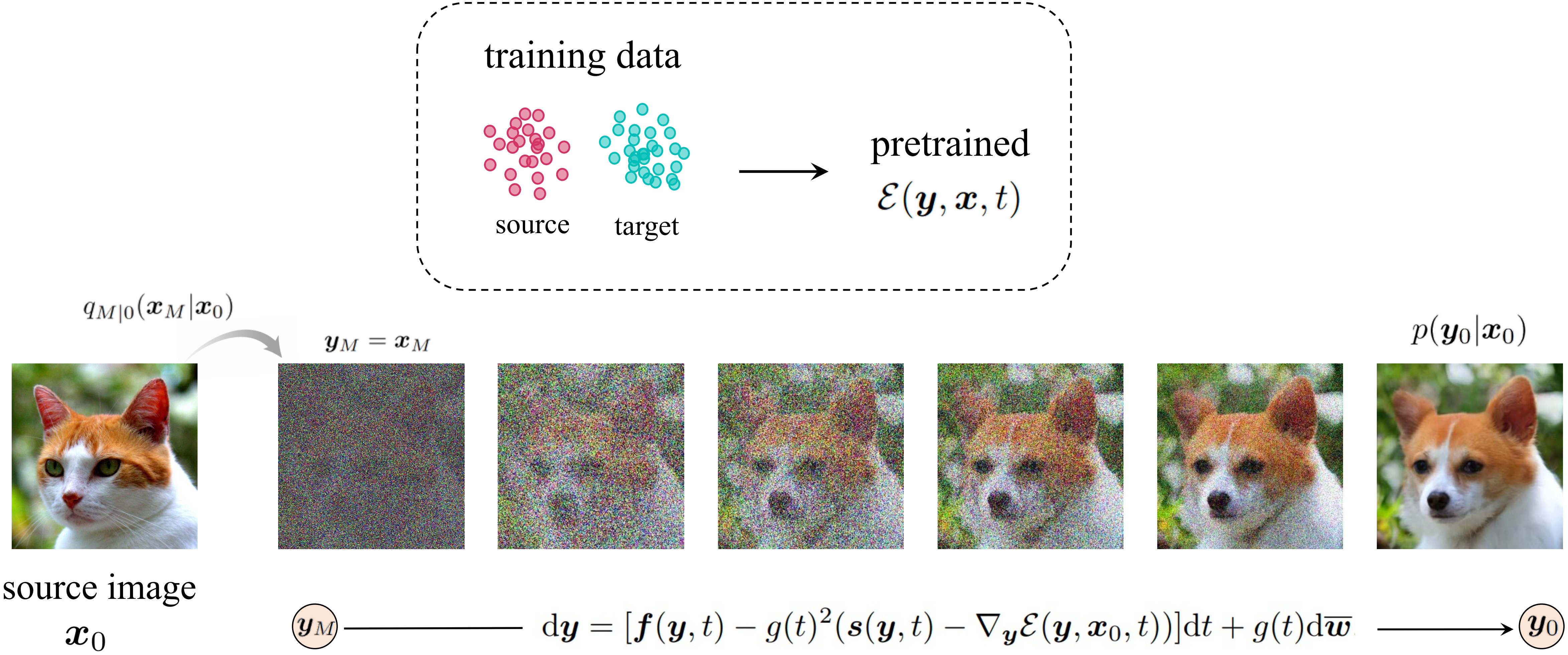
Energy-Guided Stochastic Differential Equations (EGSDE)

Zhao et al, NeurIPS 2022



Energy-Guided Stochastic Differential Equations (EGSDE)

Zhao et al, NeurIPS 2022



Following the SDE and decreasing the energy at the same time

Choice of Energy

Zhao et al, NeurIPS 2022

$$\boldsymbol{y}_M \longrightarrow \mathrm{d}\boldsymbol{y} = [\boldsymbol{f}(\boldsymbol{y}, t) - g(t)^2(\boldsymbol{s}(\boldsymbol{y}, t) - \nabla_{\boldsymbol{y}}\boxed{\mathcal{E}(\boldsymbol{y}, \boldsymbol{x}_0, t)})]\mathrm{d}t + g(t)\mathrm{d}\overline{\boldsymbol{w}} \longrightarrow \boldsymbol{y}_0$$

- Recall the goal of I2I:

Be ***realistic*** for the target domain by changing the domain-specific features
Be ***faithful*** for the source image by preserving the domain-independent features

Choice of Energy

Zhao et al, NeurIPS 2022

$$y_M \longrightarrow \mathrm{d}\boldsymbol{y} = [\boldsymbol{f}(\boldsymbol{y},t) - g(t)^2(\boldsymbol{s}(\boldsymbol{y},t) - \nabla_{\boldsymbol{y}}\boxed{\mathcal{E}(\boldsymbol{y},\boldsymbol{x}_0,t)})]\mathrm{d}t + g(t)\mathrm{d}\overline{\boldsymbol{w}} \longrightarrow y_0$$

- Recall the goal of I2I:

Be ***realistic*** for the target domain by changing the domain-specific features

Be ***faithful*** for the source image by preserving the domain-independent features

- Decompose the energy function $\mathcal{E}(y,x,t)$ as the sum of two log potential functions:

$$\mathcal{E}(y,x,t) = \boxed{\lambda_s\ \mathcal{E}_s(y,x,t)} + \boxed{\lambda_i\ \mathcal{E}_i(y,x,t)}$$

Choice of Energy

Zhao et al, NeurIPS 2022

$$\boldsymbol{y}_M \longrightarrow \mathrm{d}\boldsymbol{y} = [\boldsymbol{f}(\boldsymbol{y},t) - g(t)^2(\boldsymbol{s}(\boldsymbol{y},t) - \nabla_{\boldsymbol{y}}\mathcal{E}(\boldsymbol{y},\boldsymbol{x}_0,t))]\mathrm{d}t + g(t)\mathrm{d}\overline{\boldsymbol{w}} \longrightarrow \boldsymbol{y}_0$$

- Recall the goal of I2I:

Be ***realistic*** for the target domain by changing the domain-specific features

Be ***faithful*** for the source image by preserving the domain-independent features

- Decompose the energy function $\mathcal{E}(y,x,t)$ as the sum of two log potential functions:

$$\mathcal{E}(y,x,t) = \lambda_s\ \mathcal{E}_s(y,x,t) + \lambda_i\ \mathcal{E}_i(y,x,t)$$

$$= \lambda_s\ \mathbb{E}_{q_{t|0}(x_t|x)} S_s(y,x_t,t) - \lambda_i\ \mathbb{E}_{q_{t|0}(x_t|x)} S_i(y,x_t,t),$$

where $q_{t|0}(x_t|x)$ is the perturbation kernel from time 0 to time t in the forward SDE. $S_s(\cdot,\cdot,\cdot)$ and $S_i(\cdot,\cdot,\cdot)$ are two functions measuring the similarity between the sample and perturbed source image.

Choice of Energy

Zhao et al, NeurIPS 2022

$$y_M \longrightarrow \mathrm{d}\boldsymbol{y} = [\boldsymbol{f}(\boldsymbol{y},t) - g(t)^2(s(\boldsymbol{y},t) - \lambda_s \nabla_{\boldsymbol{y}} \mathcal{E}_s(\boldsymbol{y},\boldsymbol{x}_0,t) - \lambda_i \nabla_{\boldsymbol{y}} \mathcal{E}_i(\boldsymbol{y},\boldsymbol{x}_0,t))]\mathrm{d}t + g(t)\mathrm{d}\overline{\boldsymbol{w}} \longrightarrow y_0$$

- Suppose $E_s(\cdot,\cdot) \in R^{C\times H\times W}$ is a domain-specific feature extractor, $S_s(\cdot,\cdot,\cdot)$ is defined as the cosine similarity between the features extracted from the generated sample and the source image :

$$S_s(y, x_t, t) = \cos(E_s(y,t), E_s(x_t,t))$$

- Suppose $E_i(\cdot,\cdot) \in R^{C\times H\times W}$ is a domain-independent feature extractor , $S_i(\cdot,\cdot,\cdot)$ is defined as the negative squared L2 distance between the features extracted from the generated sample and the source image :

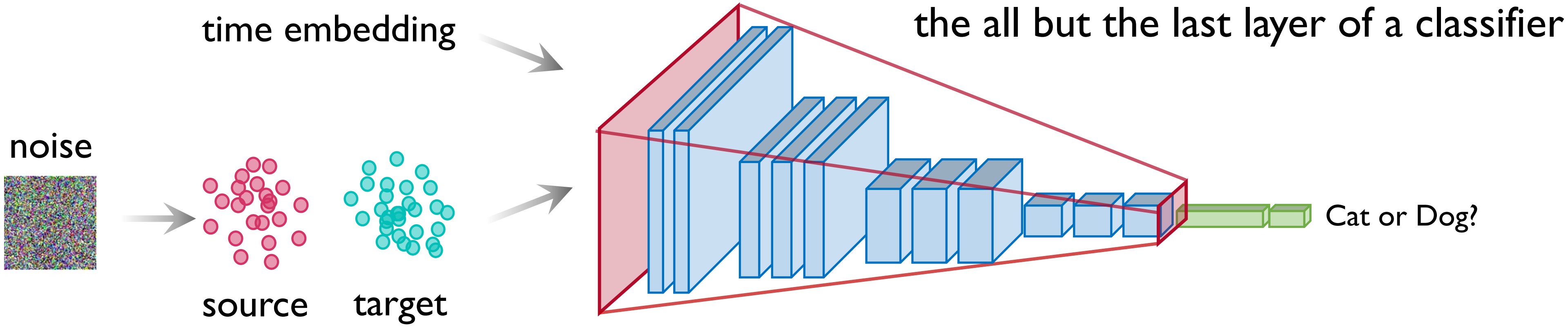
$$S_i(y, x_t, t) = -\|E_i(y,t) - E_i(x_t,t)\|_2^2$$

Choice of Energy

Zhao et al, NeurIPS 2022

$$y_M \longrightarrow \mathrm{d}\boldsymbol{y} = [\boldsymbol{f}(\boldsymbol{y}, t) - g(t)^2(s(\boldsymbol{y}, t) - \lambda_s \nabla_{\boldsymbol{y}} \mathcal{E}_s(\boldsymbol{y}, \boldsymbol{x}_0, t) - \lambda_i \nabla_{\boldsymbol{y}} \mathcal{E}_i(\boldsymbol{y}, \boldsymbol{x}_0, t))]\mathrm{d}t + g(t)\mathrm{d}\overline{\boldsymbol{w}} \longrightarrow y_0$$

➢ Domain-specific feature extractor:

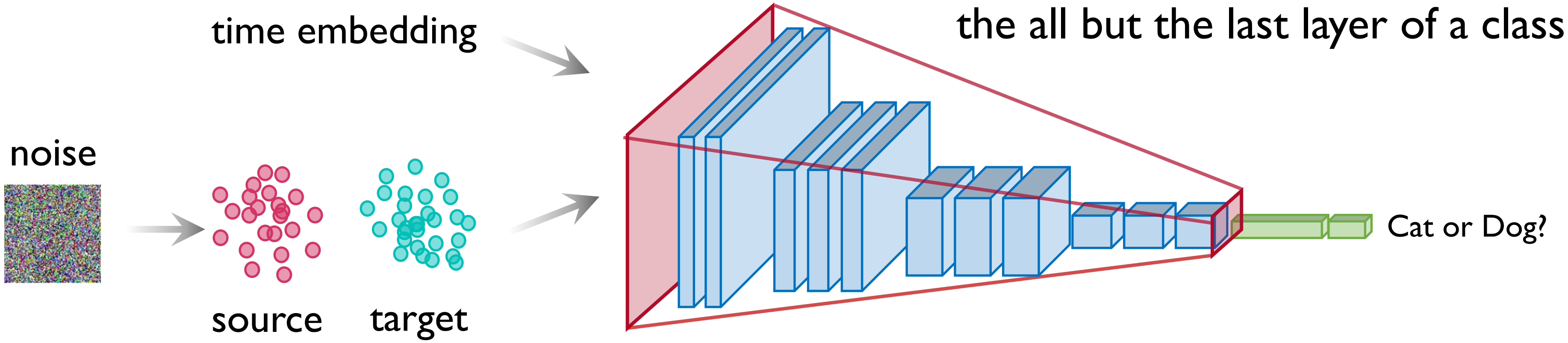


Choice of Energy

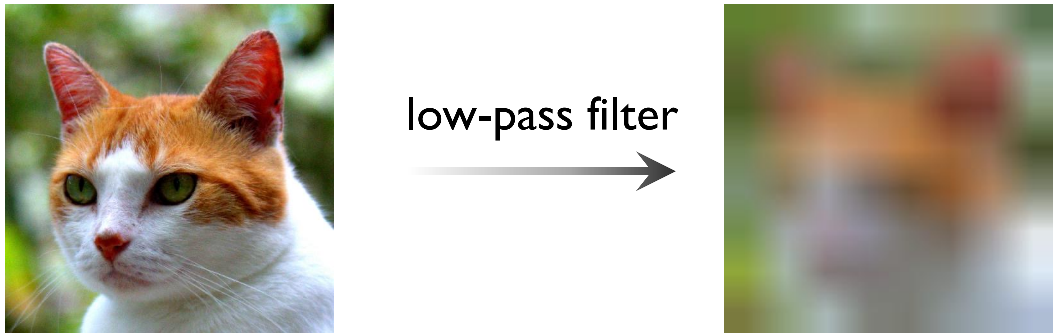
Zhao et al, NeurIPS 2022

$$y_M \longrightarrow \mathrm{d}\boldsymbol{y} = [\boldsymbol{f}(\boldsymbol{y}, t) - g(t)^2(s(\boldsymbol{y}, t) - \lambda_s \nabla_{\boldsymbol{y}} \mathcal{E}_s(\boldsymbol{y}, \boldsymbol{x}_0, t) - \lambda_i \nabla_{\boldsymbol{y}} \mathcal{E}_i(\boldsymbol{y}, \boldsymbol{x}_0, t))]\mathrm{d}t + g(t)\mathrm{d}\overline{\boldsymbol{w}} \longrightarrow y_0$$

- **Domain-specific feature extractor**:

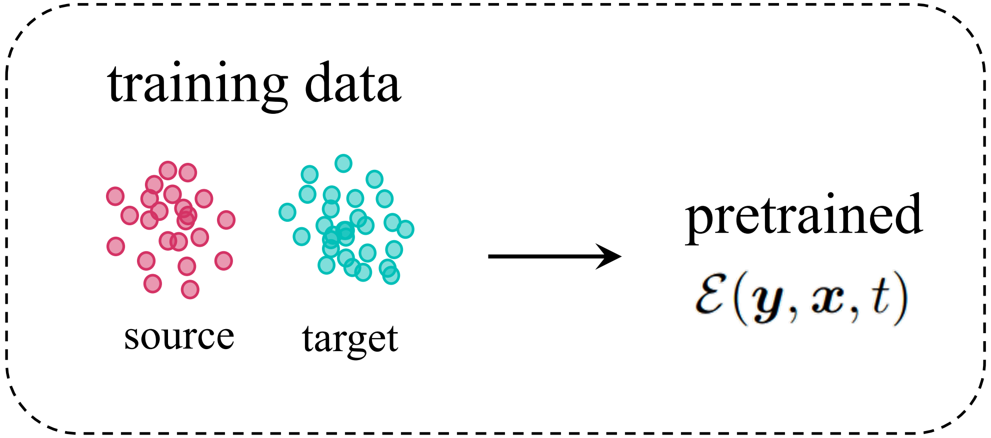


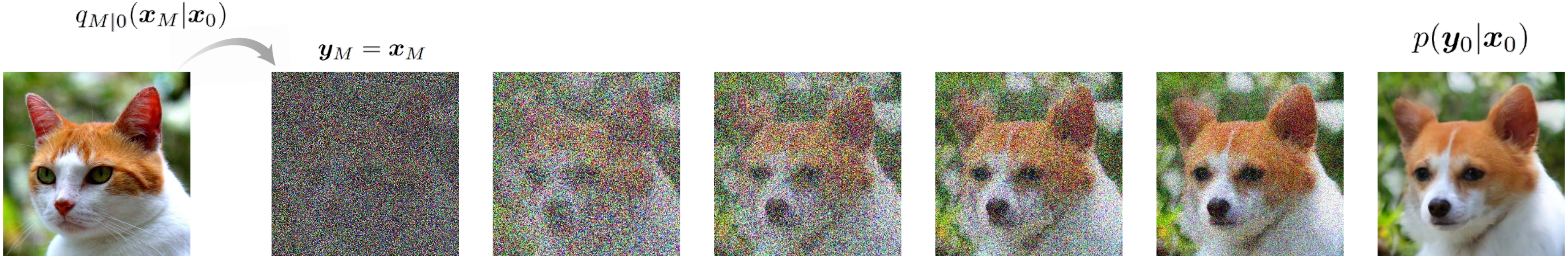
- **Domain-independent feature extractor** :



Energy-Guided Stochastic Differential Equations (EGSDE)

Zhao et al, NeurIPS 2022





source image
$\boldsymbol{x}_0$

$$\boldsymbol{y}_M \longrightarrow \mathrm{d}\boldsymbol{y} = [\boldsymbol{f}(\boldsymbol{y}, t) - g(t)^2(s(\boldsymbol{y}, t) - \lambda_s \nabla_{\boldsymbol{y}} \mathcal{E}_s(\boldsymbol{y}, \boldsymbol{x}_0, t) - \lambda_i \nabla_{\boldsymbol{y}} \mathcal{E}_i(\boldsymbol{y}, \boldsymbol{x}_0, t))]\mathrm{d}t + g(t)\mathrm{d}\overline{\boldsymbol{w}} \longrightarrow \boldsymbol{y}_0$$

known

Solving the Energy-guided Reverse-time SDE

Zhao et al, NeurIPS 2022

$$y_M \longrightarrow \mathrm{d}\boldsymbol{y} = [\boldsymbol{f}(\boldsymbol{y}, t) - g(t)^2(s(\boldsymbol{y}, t) - \lambda_s \nabla_{\boldsymbol{y}} \mathcal{E}_s(\boldsymbol{y}, \boldsymbol{x}_0, t) - \lambda_i \nabla_{\boldsymbol{y}} \mathcal{E}_i(\boldsymbol{y}, \boldsymbol{x}_0, t))]\mathrm{d}t + g(t)\mathrm{d}\overline{\boldsymbol{w}} \longrightarrow y_0$$

Algorithm 1 EGSDE for unpaired image-to-image translation

Require: the source image $\boldsymbol{x}_0$, the initial time M, denoising steps N, weighting hyper-parameters λ_s, λ_i, the similarity function $\mathcal{S}_s(\cdot,\cdot,\cdot), \mathcal{S}_i(\cdot,\cdot,\cdot)$, the score function $\boldsymbol{s}(\cdot,\cdot)$

$\boldsymbol{y} \sim q_{M|0}(\boldsymbol{y}|\boldsymbol{x}_0)$ # the start point

$h = \frac{M}{N}$

for $i = N$ to 1 **do**

  $s \leftarrow ih$

  $\boldsymbol{x} \sim q_{s|0}(\boldsymbol{x}|\boldsymbol{x}_0)$ # sample perturbed source image from the perturbation kernel

Euler-Maruyama solver

  $\mathcal{E}(\boldsymbol{y}, \boldsymbol{x}, s) \leftarrow \lambda_s \mathcal{S}_s(\boldsymbol{y}, \boldsymbol{x}, s) - \lambda_i \mathcal{S}_i(\boldsymbol{y}, \boldsymbol{x}, s)$ # compute energy with one Monte Carlo

  $\boldsymbol{y} \leftarrow \boldsymbol{y} - [\boldsymbol{f}(\boldsymbol{y}, s) - g(s)^2(\boldsymbol{s}(\boldsymbol{y}, s) - \nabla_{\boldsymbol{y}} \mathcal{E}(\boldsymbol{y}, \boldsymbol{x}, s))]h$ # the update rule in Eq. (12)

  $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ if $i > 1$, else $\boldsymbol{z} = \boldsymbol{0}$

  $\boldsymbol{y} \leftarrow \boldsymbol{y} + g(s)\sqrt{h}\boldsymbol{z}$

end for

$\boldsymbol{y}_0 \leftarrow \boldsymbol{y}$

return $\boldsymbol{y}_0$

EGSDE as Product of Experts

Zhao et al, NeurIPS 2022

Transition kernel

$$\tilde{p}(\boldsymbol{y}_t|\boldsymbol{x}_0) = \frac{p_{r1}(\boldsymbol{y}_t|\boldsymbol{x}_0)p_{r2}(\boldsymbol{y}_t|\boldsymbol{x}_0)p_f(\boldsymbol{y}_t|\boldsymbol{x}_0)}{Z_t} \longrightarrow \tilde{p}(\boldsymbol{y}_t|\boldsymbol{y}_s)$$

where $p_{r1}(\boldsymbol{y}_t|\boldsymbol{x}_0)$ is the marginal distribution defined by SDE conditioned on $\boldsymbol{x}_0$, $p_{r2}(\boldsymbol{y}_t|\boldsymbol{x}_0) \propto \exp(-\lambda_s \mathcal{E}_s(\boldsymbol{y}_t, \boldsymbol{x}_0, t))$, $p_f(\boldsymbol{y}_t|\boldsymbol{x}_0) \propto \exp(-\lambda_i \mathcal{E}_i(\boldsymbol{y}_t, \boldsymbol{x}_0, t))$.

EGSDE as Product of Experts

Zhao et al, NeurIPS 2022

Transition kernel

$$\tilde{p}(\boldsymbol{y}_t|\boldsymbol{x}_0) = \frac{p_{r1}(\boldsymbol{y}_t|\boldsymbol{x}_0)p_{r2}(\boldsymbol{y}_t|\boldsymbol{x}_0)p_f(\boldsymbol{y}_t|\boldsymbol{x}_0)}{Z_t} \longrightarrow \tilde{p}(\boldsymbol{y}_t|\boldsymbol{y}_s)$$

where $p_{r1}(\boldsymbol{y}_t|\boldsymbol{x}_0)$ is the marginal distribution defined by SDE conditioned on $\boldsymbol{x}_0$, $p_{r2}(\boldsymbol{y}_t|\boldsymbol{x}_0) \propto \exp(-\lambda_s \mathcal{E}_s(\boldsymbol{y}_t, \boldsymbol{x}_0, t))$, $p_f(\boldsymbol{y}_t|\boldsymbol{x}_0) \propto \exp(-\lambda_i \mathcal{E}_i(\boldsymbol{y}_t, \boldsymbol{x}_0, t))$.

EGSDE

Euler-Maruyama solver

$$\mathrm{d}\boldsymbol{y} = [\boldsymbol{f}(\boldsymbol{y}, t) - g(t)^2(s(\boldsymbol{y}, t) - \lambda_s \nabla_{\boldsymbol{y}} \mathcal{E}_s(\boldsymbol{y}, \boldsymbol{x}_0, t) - \lambda_i \nabla_{\boldsymbol{y}} \mathcal{E}_i(\boldsymbol{y}, \boldsymbol{x}_0, t))]\mathrm{d}t + g(t)\mathrm{d}\overline{\boldsymbol{w}} \longrightarrow p(\boldsymbol{y}_t|\boldsymbol{y}_s)$$

EGSDE as Product of Experts

Zhao et al, NeurIPS 2022

Transition kernel

$$\tilde{p}(\boldsymbol{y}_t|\boldsymbol{x}_0) = \frac{p_{r1}(\boldsymbol{y}_t|\boldsymbol{x}_0)p_{r2}(\boldsymbol{y}_t|\boldsymbol{x}_0)p_f(\boldsymbol{y}_t|\boldsymbol{x}_0)}{Z_t} \longrightarrow \tilde{p}(\boldsymbol{y}_t|\boldsymbol{y}_s)$$

where $p_{r1}(\boldsymbol{y}_t|\boldsymbol{x}_0)$ is the marginal distribution defined by SDE conditioned on $\boldsymbol{x}_0$,

$p_{r2}(\boldsymbol{y}_t|\boldsymbol{x}_0) \propto \exp(-\lambda_s \mathcal{E}_s(\boldsymbol{y}_t, \boldsymbol{x}_0, t))$, $p_f(\boldsymbol{y}_t|\boldsymbol{x}_0) \propto \exp(-\lambda_i \mathcal{E}_i(\boldsymbol{y}_t, \boldsymbol{x}_0, t))$.

Equivalent

EGSDE

Euler-Maruyama solver

$$\mathrm{d}\boldsymbol{y} = [\boldsymbol{f}(\boldsymbol{y}, t) - g(t)^2(s(\boldsymbol{y}, t) - \lambda_s \nabla_{\boldsymbol{y}} \mathcal{E}_s(\boldsymbol{y}, \boldsymbol{x}_0, t) - \lambda_i \nabla_{\boldsymbol{y}} \mathcal{E}_i(\boldsymbol{y}, \boldsymbol{x}_0, t))]\mathrm{d}t + g(t)\mathrm{d}\overline{\boldsymbol{w}} \longrightarrow p(\boldsymbol{y}_t|\boldsymbol{y}_s)$$

EGSDE as Product of Experts

Zhao et al, NeurIPS 2022

$$\tilde{p}(\boldsymbol{y}_t|\boldsymbol{x}_0) = \frac{p_{r1}(\boldsymbol{y}_t|\boldsymbol{x}_0)p_{r2}(\boldsymbol{y}_t|\boldsymbol{x}_0)p_f(\boldsymbol{y}_t|\boldsymbol{x}_0)}{Z_t}$$

t = 0

$$\tilde{p}(\boldsymbol{y}_0|\boldsymbol{x}_0) = \frac{p_{r1}(\boldsymbol{y}_0|\boldsymbol{x}_0)p_{r2}(\boldsymbol{y}_0|\boldsymbol{x}_0)p_f(\boldsymbol{y}_0|\boldsymbol{x}_0)}{Z_0}$$

source image

$\boldsymbol{x}_0$

$p(\boldsymbol{y}_0|\boldsymbol{x}_0)$

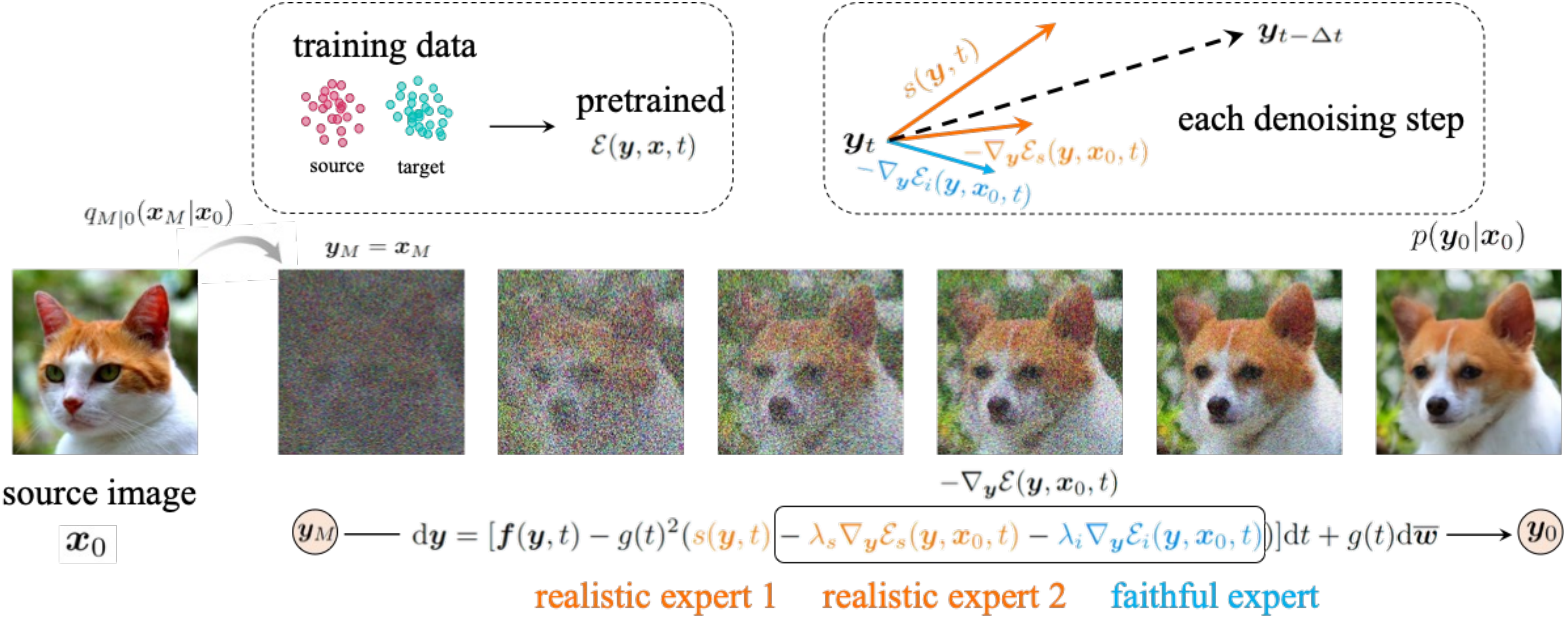
translated image

$$\boldsymbol{y}_M \longrightarrow \mathrm{d}\boldsymbol{y} = [\boldsymbol{f}(\boldsymbol{y},t) - g(t)^2(s(\boldsymbol{y},t) - \lambda_s \nabla_{\boldsymbol{y}} \mathcal{E}_s(\boldsymbol{y},\boldsymbol{x}_0,t) - \lambda_i \nabla_{\boldsymbol{y}} \mathcal{E}_i(\boldsymbol{y},\boldsymbol{x}_0,t))]\mathrm{d}t + g(t)\mathrm{d}\overline{\boldsymbol{w}} \longrightarrow \boldsymbol{y}_0$$

realistic expert 1 realistic expert 2 faithful expert

Energy-Guided Stochastic Differential Equations (EGSDE)

Zhao et al, NeurIPS 2022



Results

Zhao et al, NeurIPS 2022

Realistic **Faithful** **Human Evaluation, Both**

Model	FID ↓	L2 ↓	PSNR ↑	SSIM ↑	AMT ↑
		Cat → Dog			
CycleGAN* [54]	85.9	-	-	-	-
MUNIT* [17]	104.4	-	-	-	-
DRIT* [25]	123.4	-	-	-	-
Distance* [3]	155.3	-	-	-	-
SelfDistance* [3]	144.4	-	-	-	-
GCGAN* [10]	96.6	-	-	-	-
LSeSim* [52]	72.8	-	-	-	-
ITTR (CUT)* [53]	68.6	-	-	-	-
StarGAN v2 [8]	54.88 ± 1.01	133.65 ± 1.54	10.63 ± 0.10	0.27 ± 0.003	-
CUT* [34]	76.21	59.78	17.48	**0.601**	79.6%
ILVR [7]	74.37 ± 1.55	56.95 ± 0.14	17.77 ± 0.02	0.363 ± 0.001	75.4%
SDEdit [30]	74.17 ± 1.01	47.88 ± 0.06	19.19 ± 0.01	**0.423 ± 0.001**	65.2%
EGSDE	**65.82 ± 0.77**	**47.22 ± 0.08**	**19.31 ± 0.02**	0.415 ± 0.001	-
EGSDE†	**51.04 ± 0.37**	62.06 ± 0.10	17.17 ± 0.02	0.361 ± 0.001	-

EGSDE : $\lambda_s = 500, \lambda_i = 2, M = 0.5\text{T}$

EGSDE†: $\lambda_s = 700, \lambda_i = 0.5, M = 0.6\text{T}$

Results

Zhao et al, NeurIPS 2022

Model	FID ↓	L2 ↓	PSNR ↑	SSIM ↑	AMT ↑
		Cat → Dog			
CycleGAN* [54]	85.9	-	-	-	-
MUNIT* [17]	104.4	-	-	-	-
DRIT* [25]	123.4	-	-	-	-
Distance* [3]	155.3	-	-	-	-
SelfDistance* [3]	144.4	-	-	-	-
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
StarGAN v2 [8]	54.88 ± 1.01	133.65 ± 1.54	10.63 ± 0.10	0.27 ± 0.003	-
CUT* [34]	76.21	59.78	17.48	**0.601**	79.6%
ILVR [7]	74.37 ± 1.55	56.95 ± 0.14	17.77 ± 0.02	0.363 ± 0.001	75.4%
SDEdit [30]	74.17 ± 1.01	47.88 ± 0.06	19.19 ± 0.01	**0.423 ± 0.001**	65.2%
EGSDE	**65.82 ± 0.77**	**47.22 ± 0.08**	**19.31 ± 0.02**	0.415 ± 0.001	-
EGSDE†	**51.04 ± 0.37**	62.06 ± 0.10	17.17 ± 0.02	0.361 ± 0.001	-

EGSDE outperforms the SBDMs-based methods in almost all metrics

EGSDE $: \lambda_s = 500, \lambda_i = 2, M = 0.5\text{T}$

EGSDE† $: \lambda_s = 700, \lambda_i = 0.5, M = 0.6\text{T}$

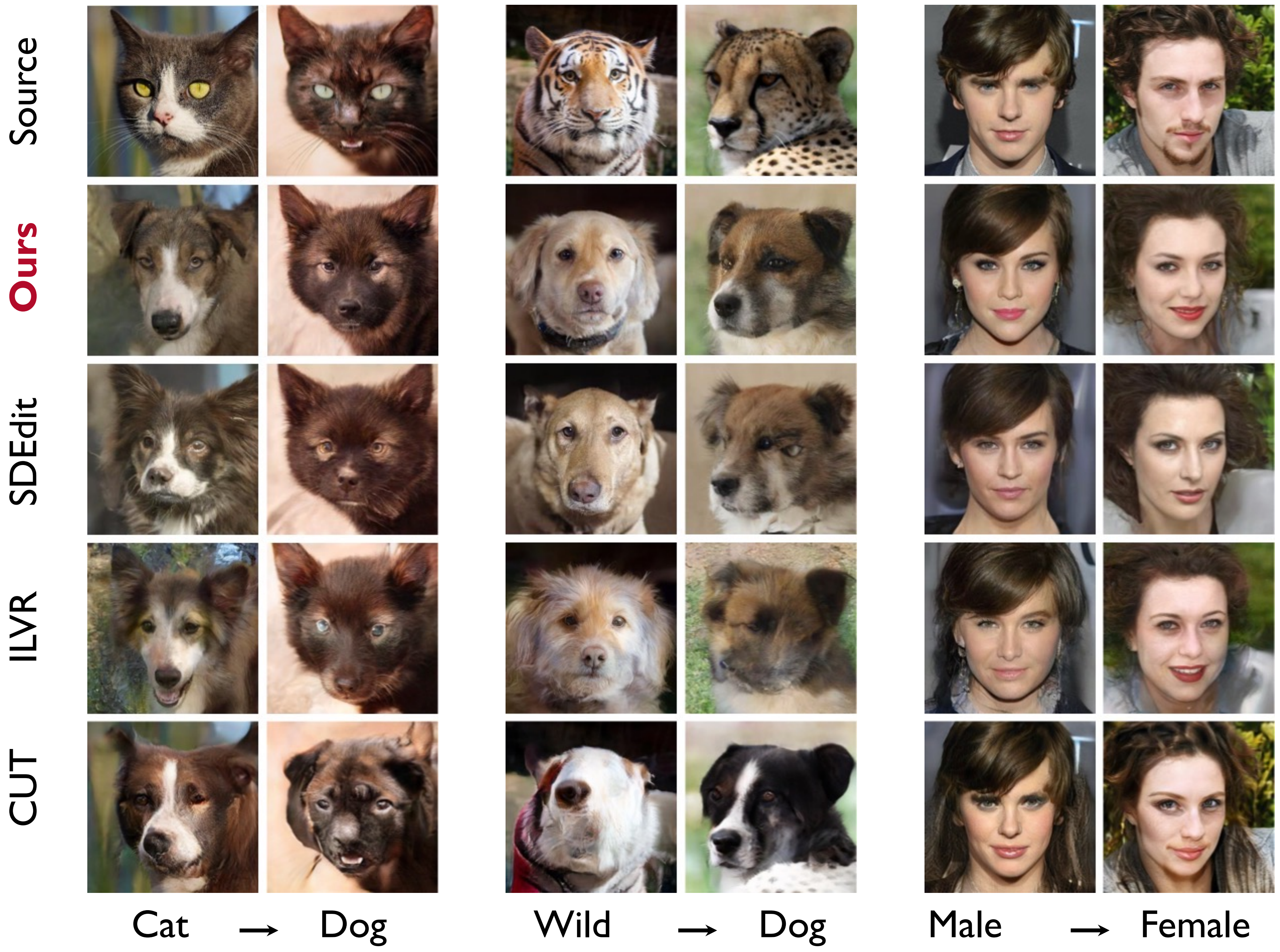
Results

Zhao et al, NeurIPS 2022

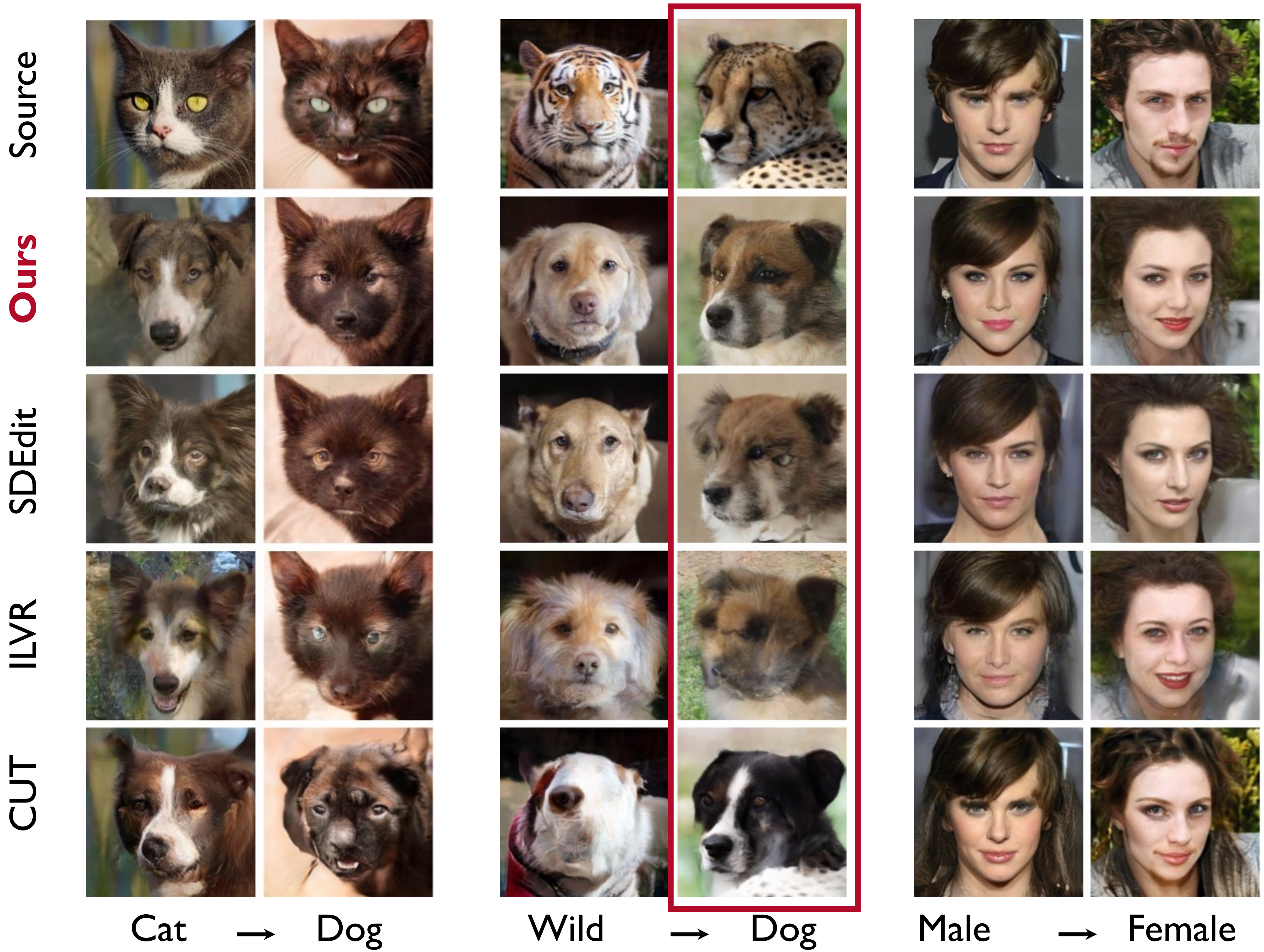
Model	FID ↓	L2 ↓	PSNR ↑	SSIM ↑	AMT ↑
		Cat → Dog			
CycleGAN* [54]	85.9	-	-	-	-
MUNIT* [17]	104.4	-	-	-	-
DRIT* [25]	123.4	-	-	-	-
Distance* [3]	155.3	-	-	-	-
SelfDistance* [3]	144.4	-	-	-	-
GCGAN* [10]	96.6	-	-	-	-
LSeSim* [52]	72.8	-	-	-	-
ITTR (CUT)* [53]	68.6	-	-	-	-
StarGAN v2 [8]	54.88 ± 1.01	133.65 ± 1.54	10.63 ± 0.10	0.27 ± 0.003	-
CUT* [34]	76.21	59.78	17.48	**0.601**	79.6%
ILVR [7]	74.37 ± 1.55	56.95 ± 0.14	17.77 ± 0.02	0.363 ± 0.001	75.4%
SDEdit [30]	74.17 ± 1.01	47.88 ± 0.06	19.19 ± 0.01	**0.423 ± 0.001**	65.2%
EGSDE	**65.82 ± 0.77**	**47.22 ± 0.08**	**19.31 ± 0.02**	0.415 ± 0.001	-
EGSDE†	**51.04 ± 0.37**	62.06 ± 0.10	17.17 ± 0.02	0.361 ± 0.001	-

EGSDE outperforms the current state-of-art GANs-based methods

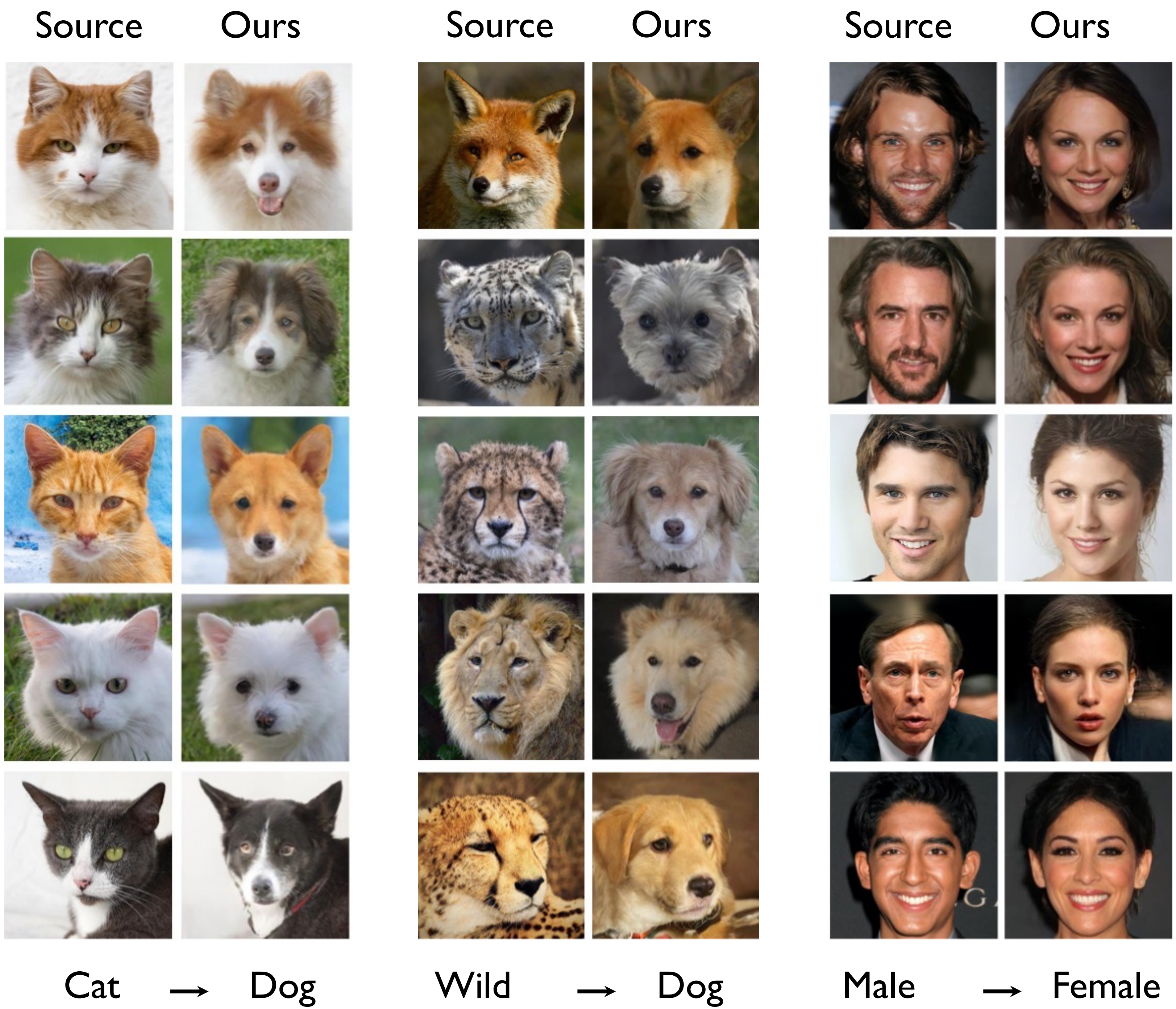
Results



Results

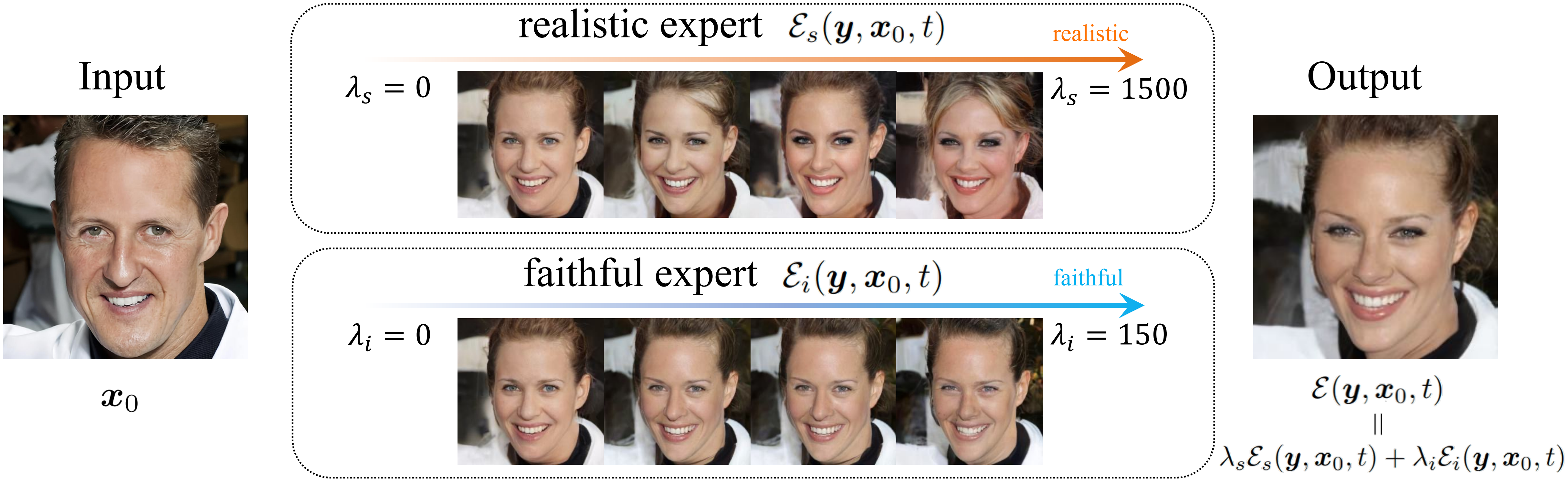


Results

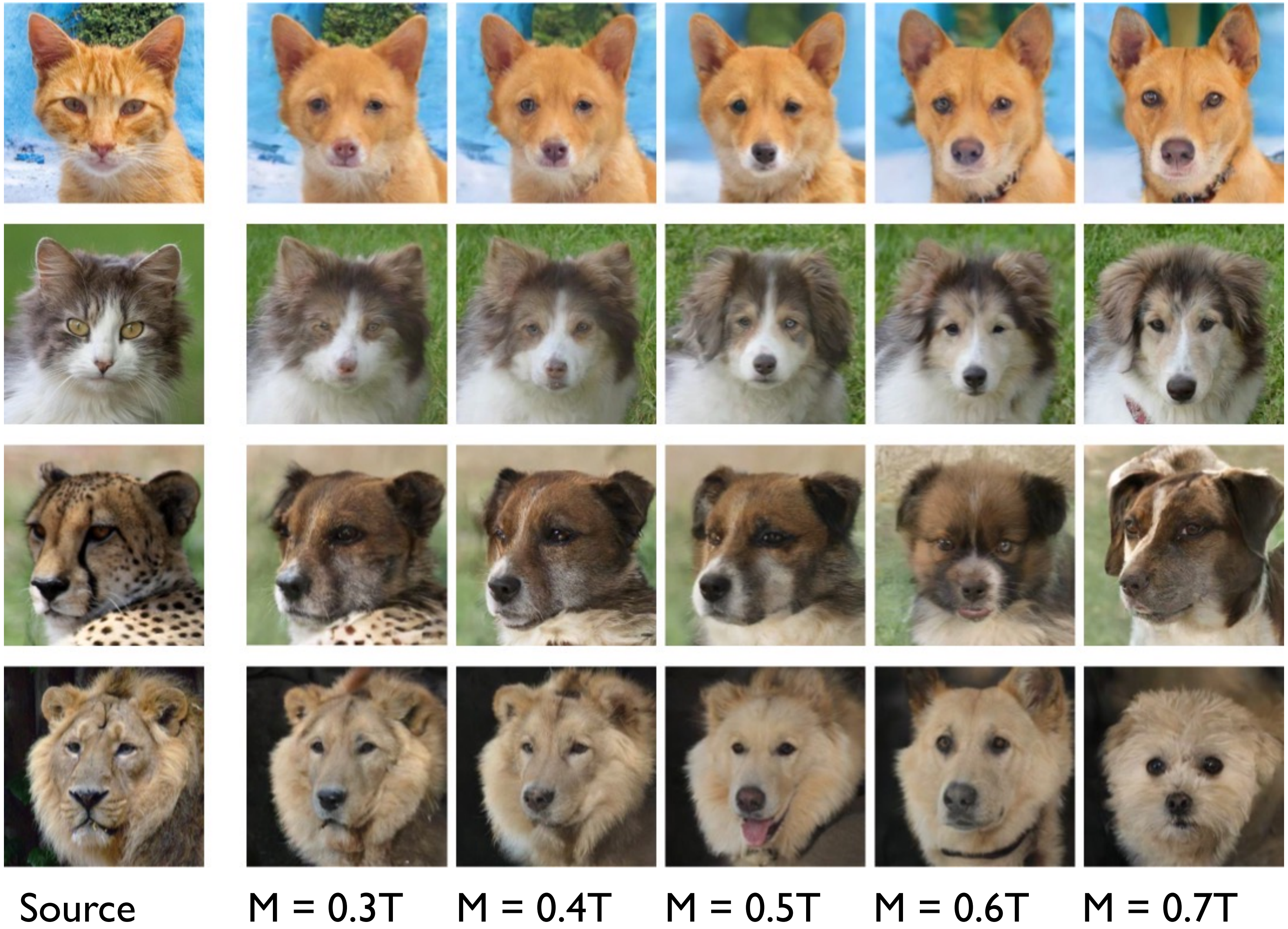

Source
Ours
Source
Ours
Source
Ours
Cat → Dog
Wild → Dog
Male → Female

The function of each expert

Zhao et al, NeurIPS 2022



The choice of initial time M



EGSDE in a big picture

The Connection with Classifier Guidance

Zhao et al, NeurIPS 2022

EGSDE: a general framework that employs an energy function with domain knowledge to guide the inference process

$$\mathrm{d}\boldsymbol{x} = [\boldsymbol{f}(\boldsymbol{x}, t) - g(t)^2(\boldsymbol{s}(\boldsymbol{x}, t) - \nabla_{\boldsymbol{x}}\mathcal{E}(\boldsymbol{x}, c, t))]\mathrm{d}t + g(t)\mathrm{d}\overline{\boldsymbol{w}}$$

The Connection with Classifier Guidance

Zhao et al, NeurIPS 2022

EGSDE: a general framework that employs an energy function with domain knowledge to guide the inference process

$$\mathrm{d}\boldsymbol{x} = [\boldsymbol{f}(\boldsymbol{x}, t) - g(t)^2(\boldsymbol{s}(\boldsymbol{x}, t) - \nabla_{\boldsymbol{x}}\mathcal{E}(\boldsymbol{x}, c, t))]\mathrm{d}t + g(t)\mathrm{d}\overline{\boldsymbol{w}}$$

$$\mathcal{E}(\boldsymbol{x}, c, t) \propto -\lambda \log p_t(c|\boldsymbol{x})$$

Classifier Guidance

$$\mathrm{d}\boldsymbol{x} = [\boldsymbol{f}(\boldsymbol{x}, t) - g(t)^2(\boldsymbol{s}(\boldsymbol{x}, t) + \lambda\nabla_{\boldsymbol{x}} \log p_t(c|\boldsymbol{x}))]\mathrm{d}t + g(t)\mathrm{d}\overline{\boldsymbol{w}}$$

The Connection with Classifier Guidance

Zhao et al, NeurIPS 2022

EGSDE: a general framework that employs an energy function with domain knowledge to guide the inference process

$$\mathrm{d}\boldsymbol{x} = [\boldsymbol{f}(\boldsymbol{x}, t) - g(t)^2(\boldsymbol{s}(\boldsymbol{x}, t) - \nabla_{\boldsymbol{x}}\mathcal{E}(\boldsymbol{x}, c, t))]\mathrm{d}t + g(t)\mathrm{d}\overline{\boldsymbol{w}}$$

$$\mathcal{E}(\boldsymbol{x}, c, t) \propto -\lambda \log p_t(c|\boldsymbol{x})$$

Classifier Guidance

$$\mathrm{d}\boldsymbol{x} = [\boldsymbol{f}(\boldsymbol{x}, t) - g(t)^2(\boldsymbol{s}(\boldsymbol{x}, t) + \lambda\nabla_{\boldsymbol{x}} \log p_t(c|\boldsymbol{x}))]\mathrm{d}t + g(t)\mathrm{d}\overline{\boldsymbol{w}}$$

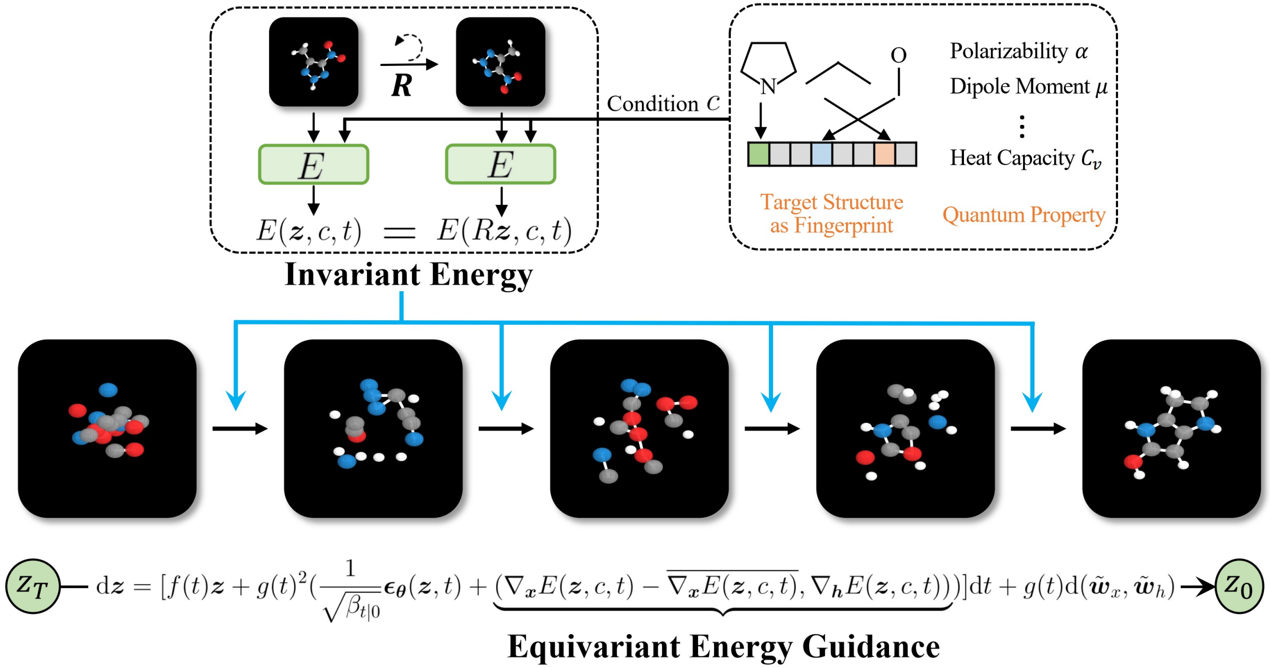
The classifier guidance can be regarded as a special design of energy function

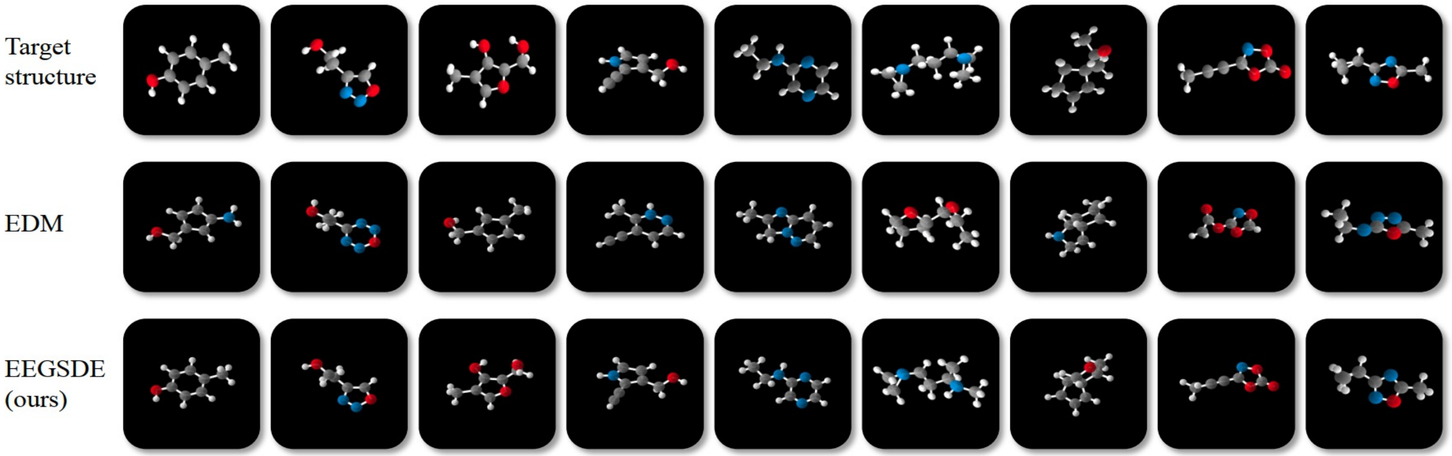
What's next?

Controllable 3D molecule generation

Bao & Zhao et al., arixv 2022

- Equivariant energy guidance
- Evaluated in inverse molecular design





Thank you!

References

- Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models[J]. Advances in Neural Information Processing Systems, 2020.
- Nichol, A., Dhariwal, P., Ramesh, A., Shyam, P., Mishkin, P., McGrew, B., Sutskever, I. and Chen, M., 2021. Glide: Towards photorealistic image generation and editing with text-guided diffusion models[J] ICML 2022.
- Zheng, Y., Wu, L., Liu, X., Chen, Z., Liu, Q. and Huang, Q., 2022. Neural Volumetric Mesh Generator. arXiv preprint arXiv:2210.03158.
- Hoogeboom, E., Satorras, V.G., Vignac, C. and Welling, M., 2022, June. Equivariant diffusion for molecule generation in 3d[J] ICML 2022
- Zhou, L., Du, Y. and Wu, J., 2021. 3d shape generation and completion through point-voxel diffusion[J] ICCV 2021
- Song Y, Sohl-Dickstein J, Kingma D P, et al. Score-based generative modeling through stochastic differential equations[J]. ICLR 2021.
- Taesung Park, Alexei A Efros, Richard Zhang, and Jun-Yan Zhu. Contrastive learning for unpaired image-to-image translation[J]. ECCV 2020.
- Chenlin Meng, Yutong He, Yang Song, Jiaming Song, Jiajun Wu, Jun-Yan Zhu, and Stefano Ermon. Sdedit: Guided image synthesis and editing with stochastic differential equations[J]. ICLR 2021.
- Jooyoung Choi, Sungwon Kim, Yonghyun Jeong, Youngjune Gwon, and Sungroh Yoon. Ilvr: Conditioning method for denoising diffusion probabilistic models[J]. ICCV 2021.
- Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis[J]. NeurIPS 2021.
- **Zhao M, Bao F, Li C, Zhu J. EGSDE: Unpaired Image-to-Image Translation via Energy-Guided Stochastic Differential Equations[J]. NeurIPS 2022.**
- **Bao F, Zhao M, Hao Z, Li P, Li C, Zhu J. Equivariant Energy-Guided SDE for Inverse Molecular Design. Arxiv preprint 2022.**