Structure-preserving machine learning and data-driven structure discovery

Wei Zhu

Georgia Institute of Technology

Machine Learning and PDEs Workshop

FAU

April 30, 2025

Structure-preserving generative models ~~and data-driven structure discovery~~

Wei Zhu

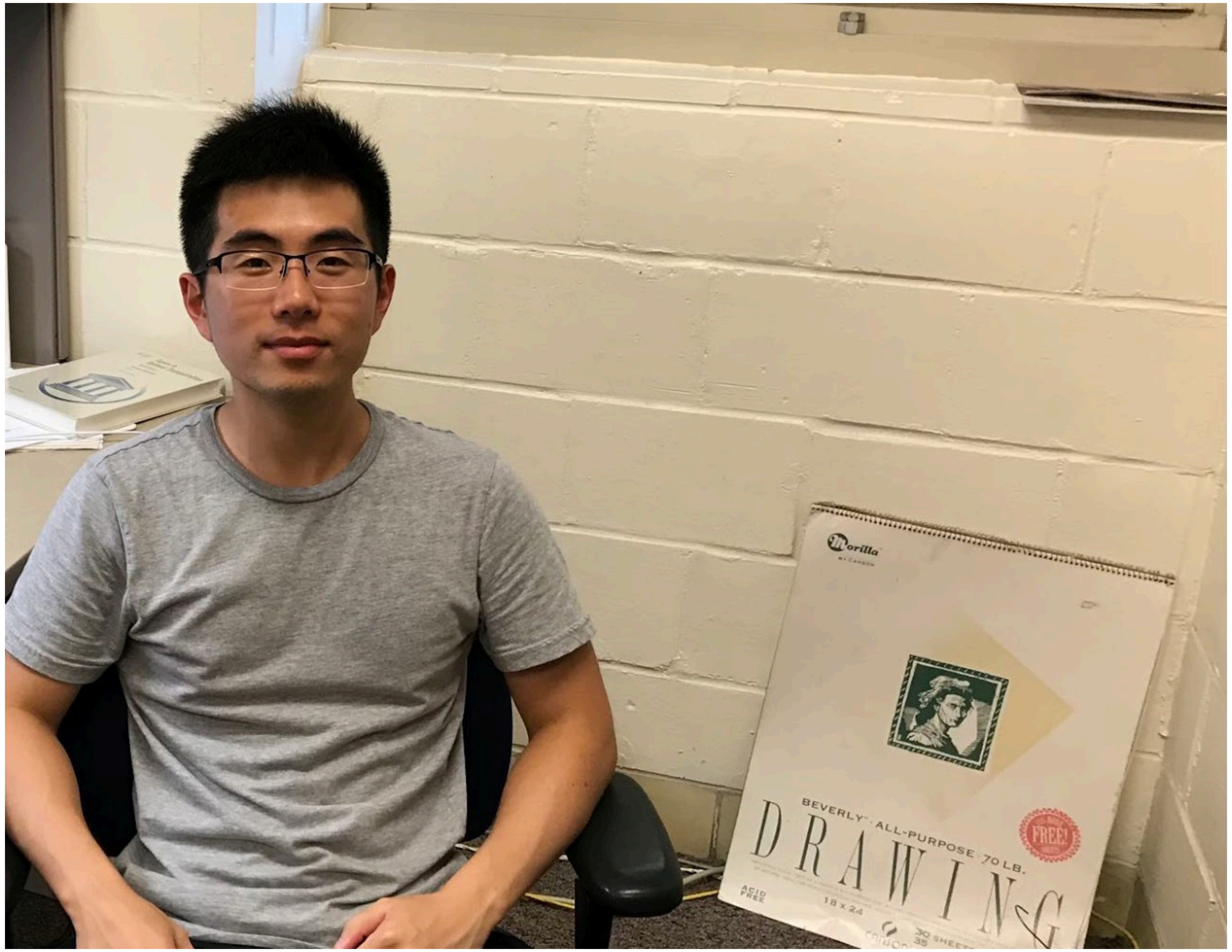
Georgia Institute of Technology

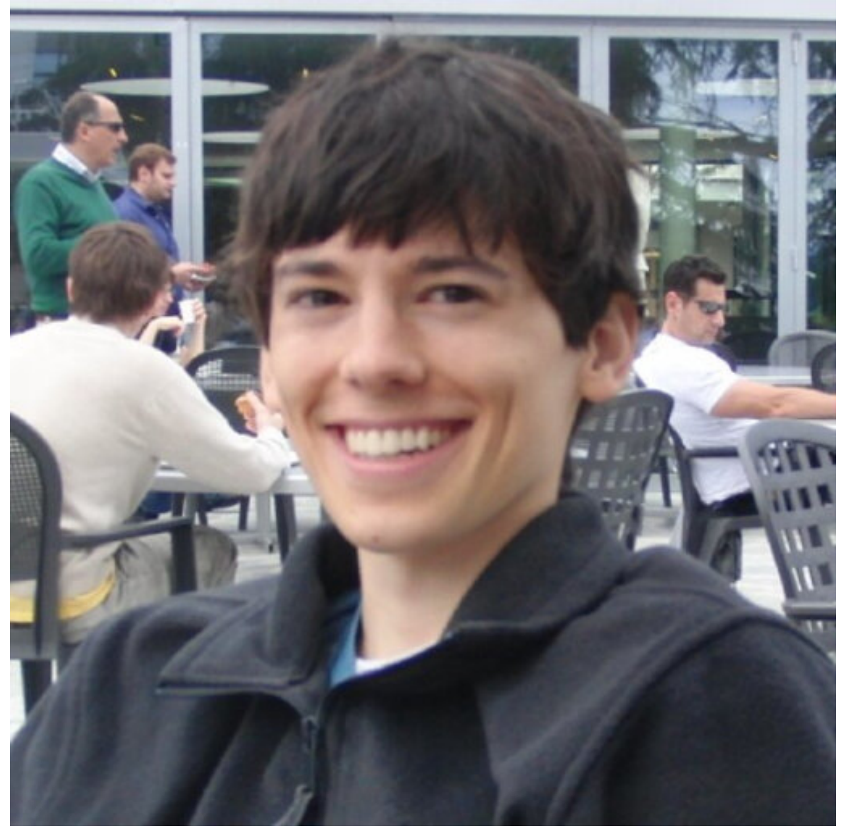
Machine Learning and PDEs Workshop

FAU

April 30, 2025

Joint with Jeremiah Birrell, **Ziyu Chen**, Markos Katsoulakis, Luc Rey-Bellet, Ben Zhang.





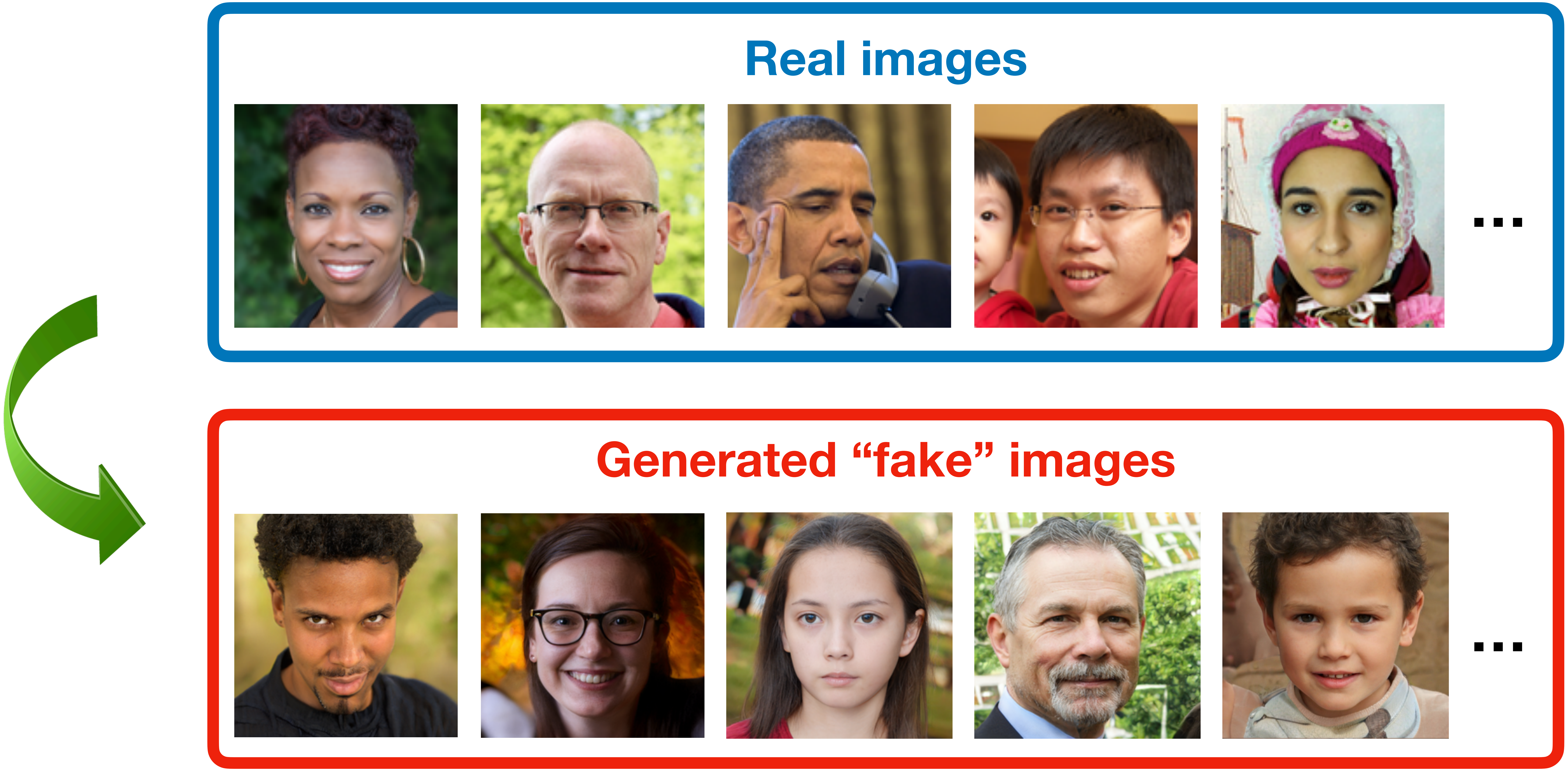
Generative models

Real images

...

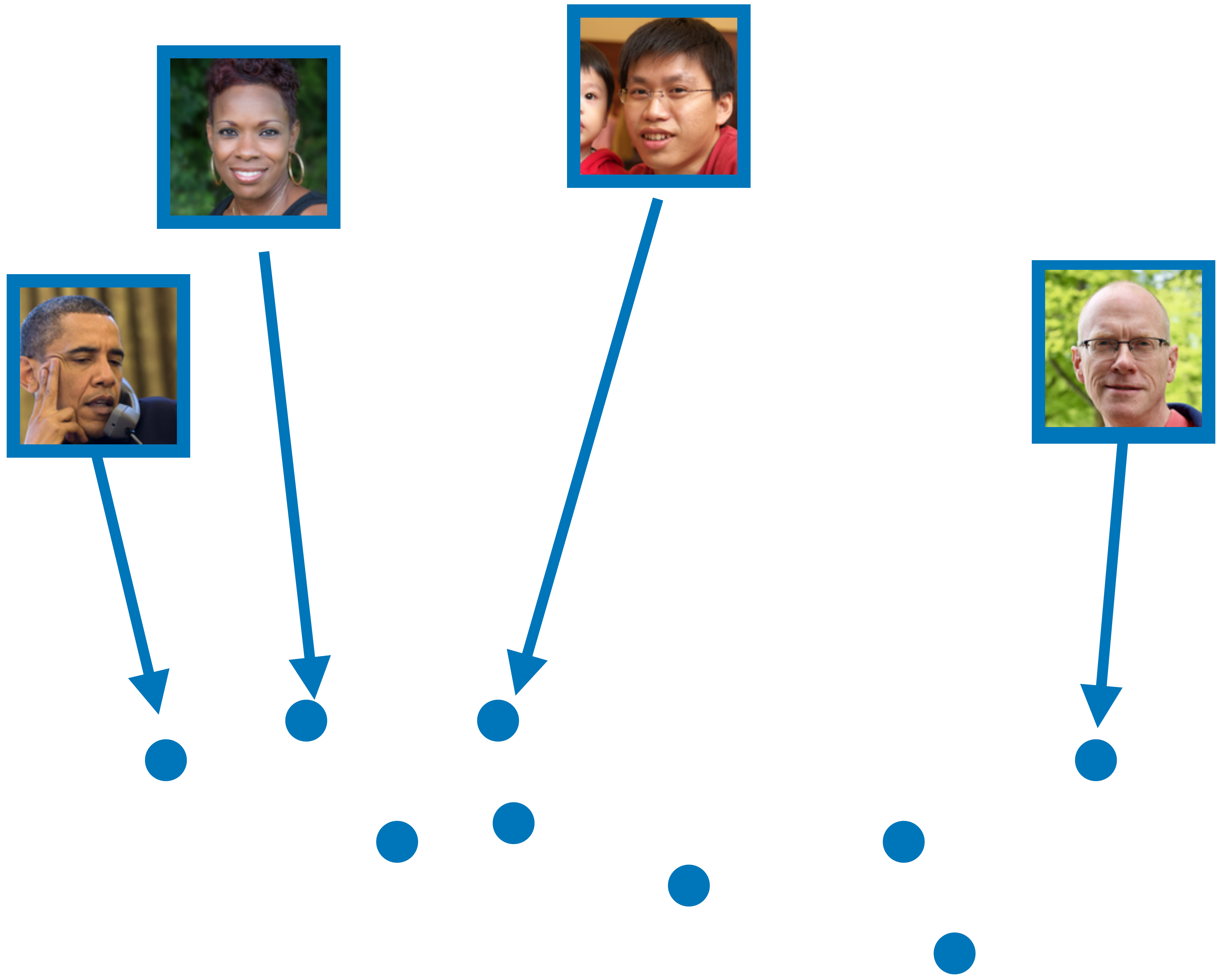
StyleGAN2, Karras et al., *CVPR* 2020

Generative models



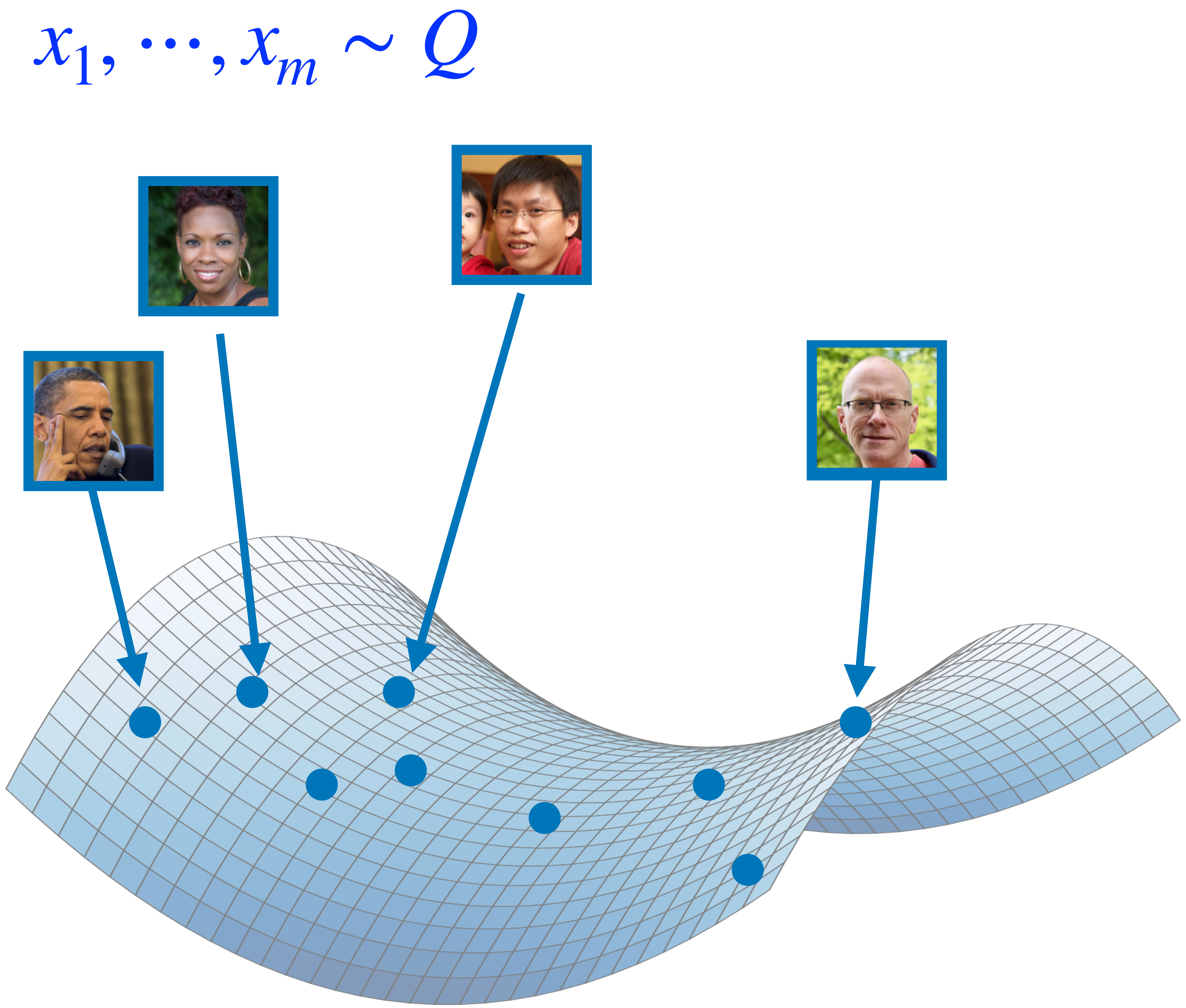
StyleGAN2, Karras et al., *CVPR* 2020

Generative models



Generative models

- $(x_1, \cdots, x_m) \sim Q$—**unknown target** distribution.

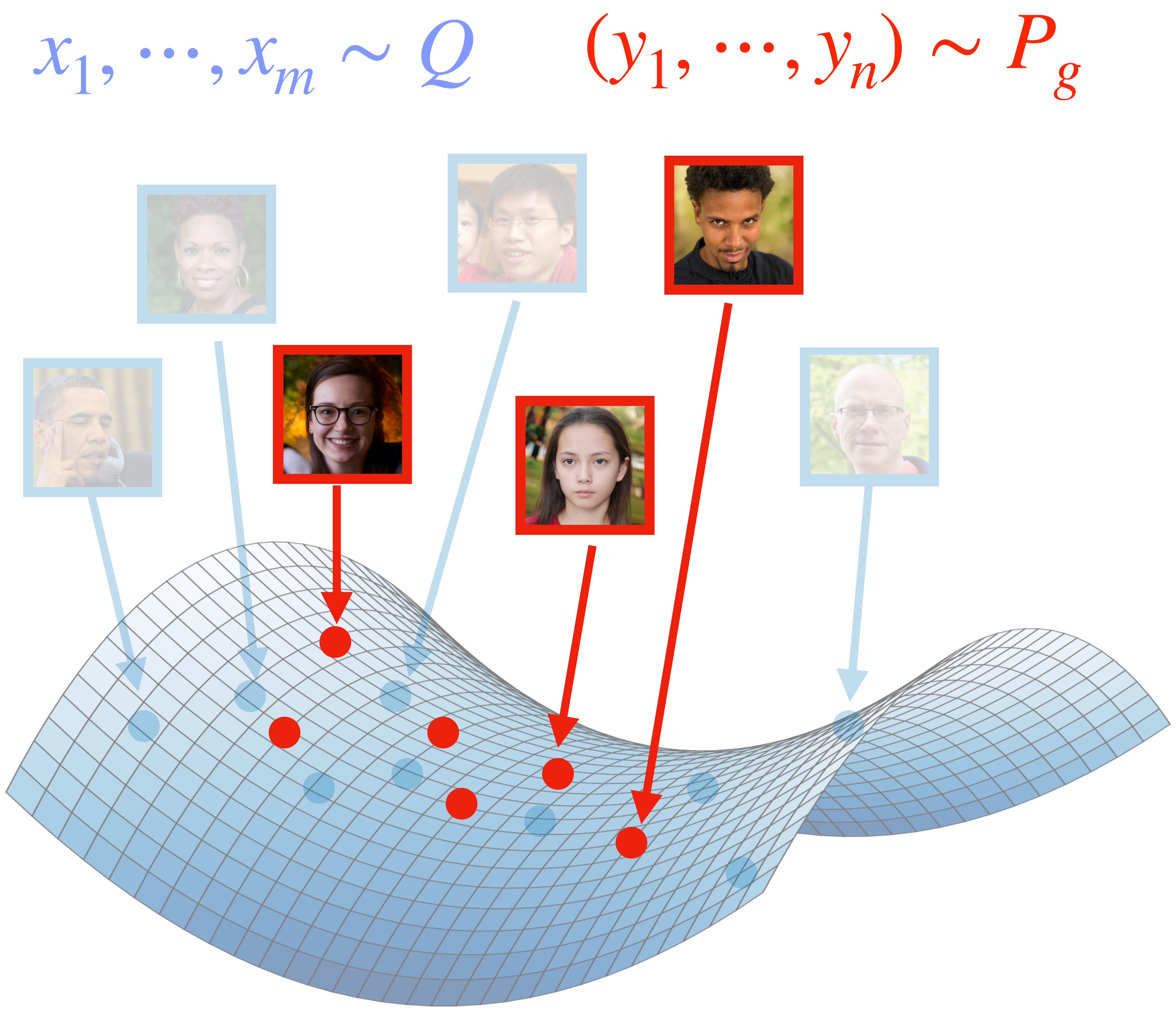


Generative models

- $(x_1, \cdots, x_m) \sim Q$—**unknown target** distribution.
- **Learn** a "mechanism" to generate *new* samples

$$(y_1, \cdots, y_n) \sim P_g$$

  - P_g is the **generated** distribution.
  - $Q \approx P_g$



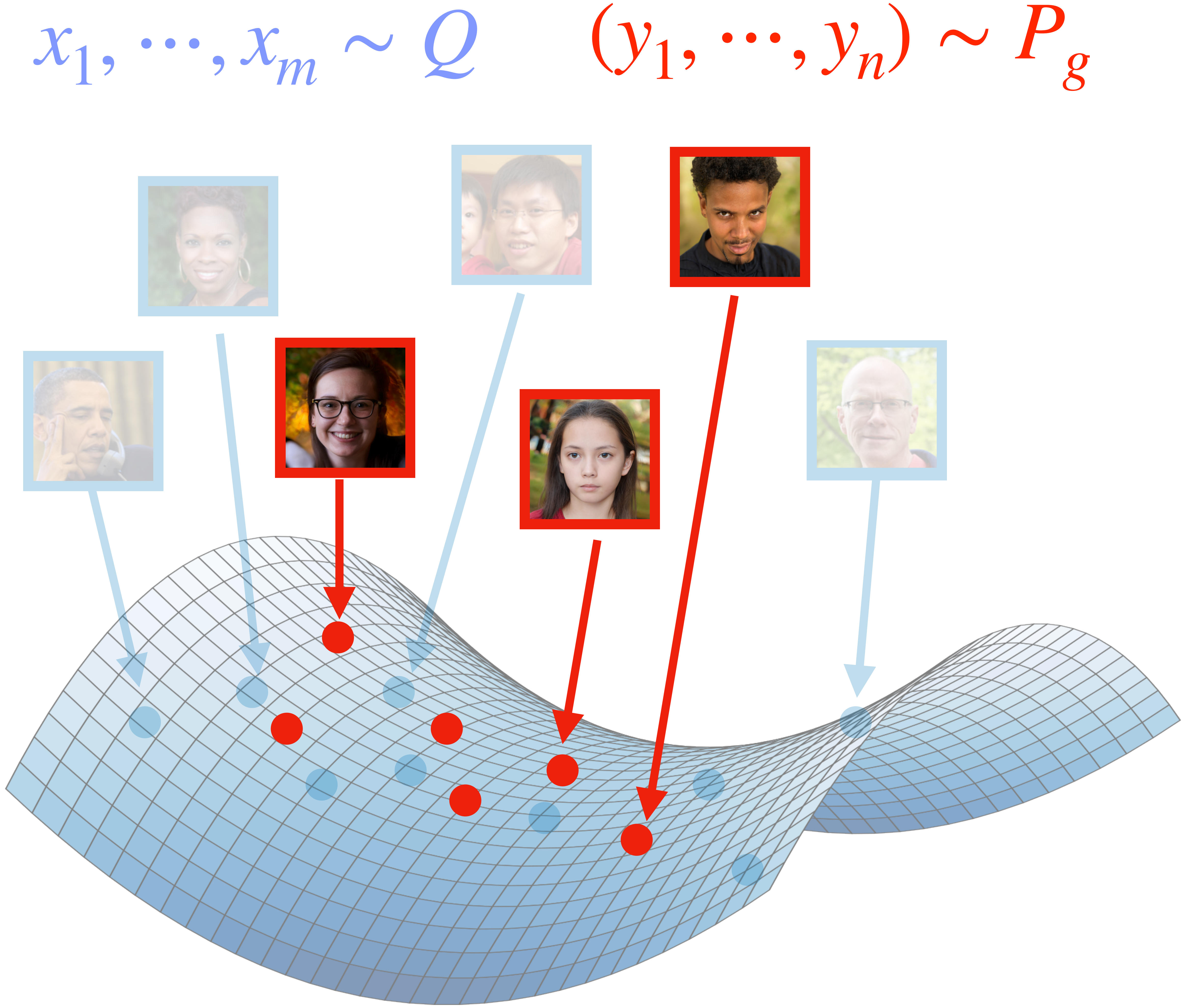
Generative models

- $(x_1, \cdots, x_m) \sim Q$—**unknown target** distribution.
- **Learn** a "mechanism" to generate *new* samples

$$(y_1, \cdots, y_n) \sim P_g$$

  - P_g is the **generated** distribution.
  - $Q \approx P_g$

Question: How to achieve this?



Generative Adversarial Networks (GANs)

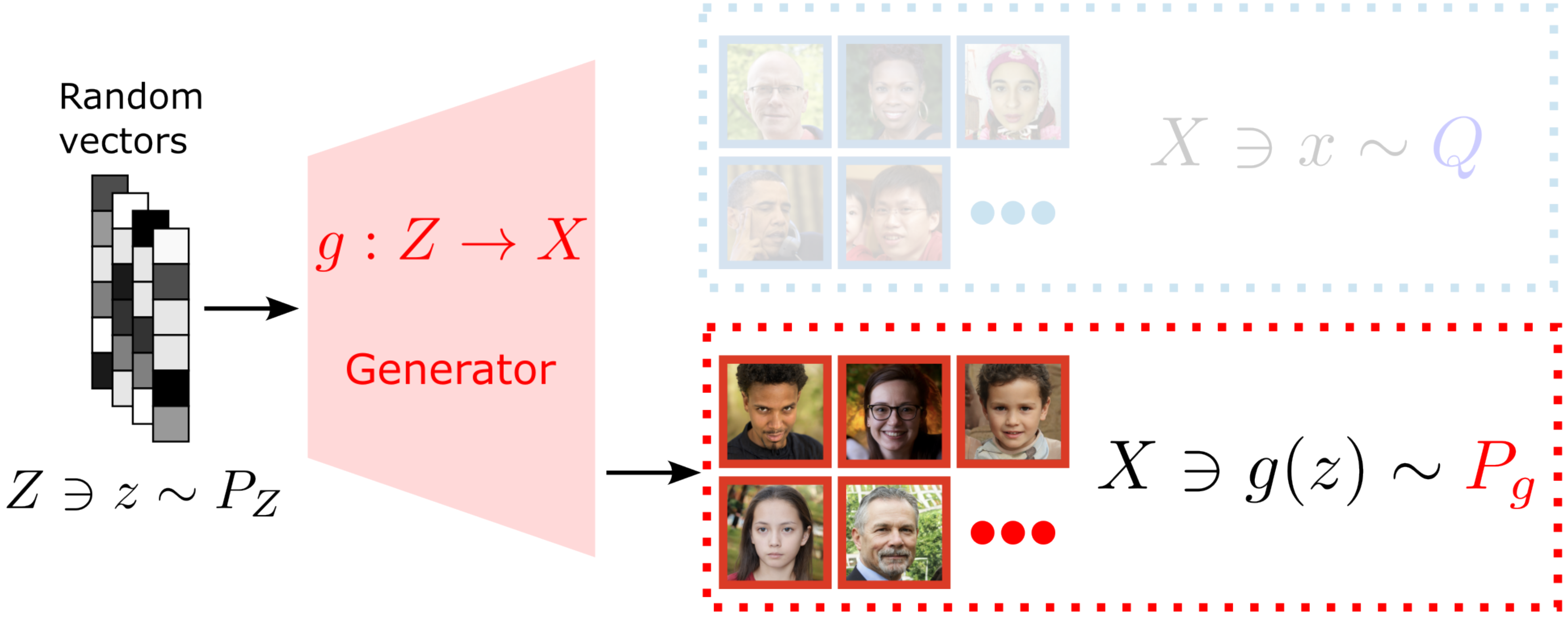
$$X \ni x \sim Q$$

Generative Adversarial Networks (GANs)

$X \ni x \sim Q$

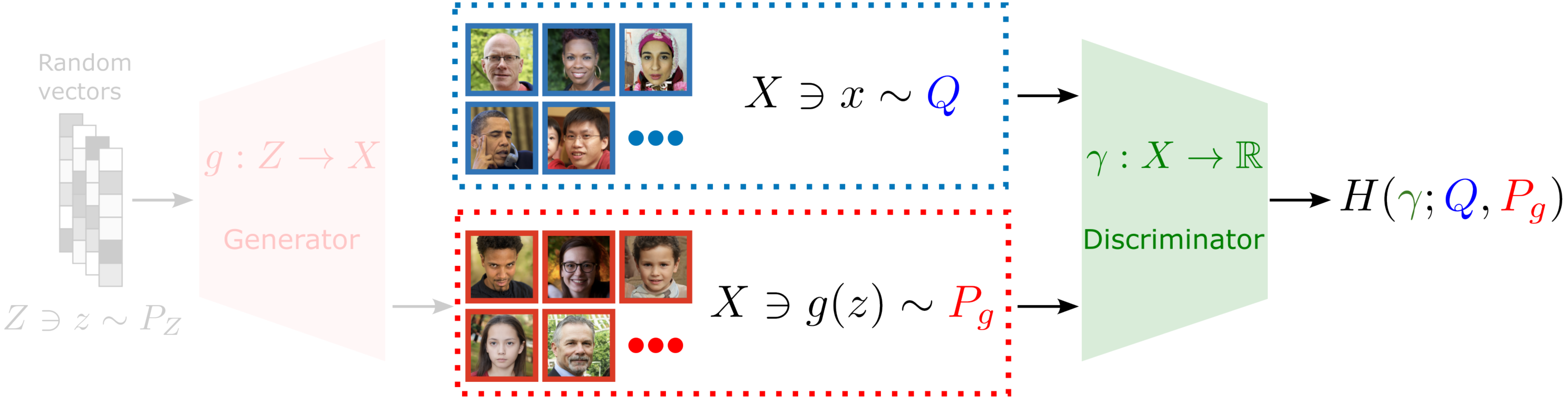
- **GAN:** two-player game for learning Q.

Generative Adversarial Networks (GANs)



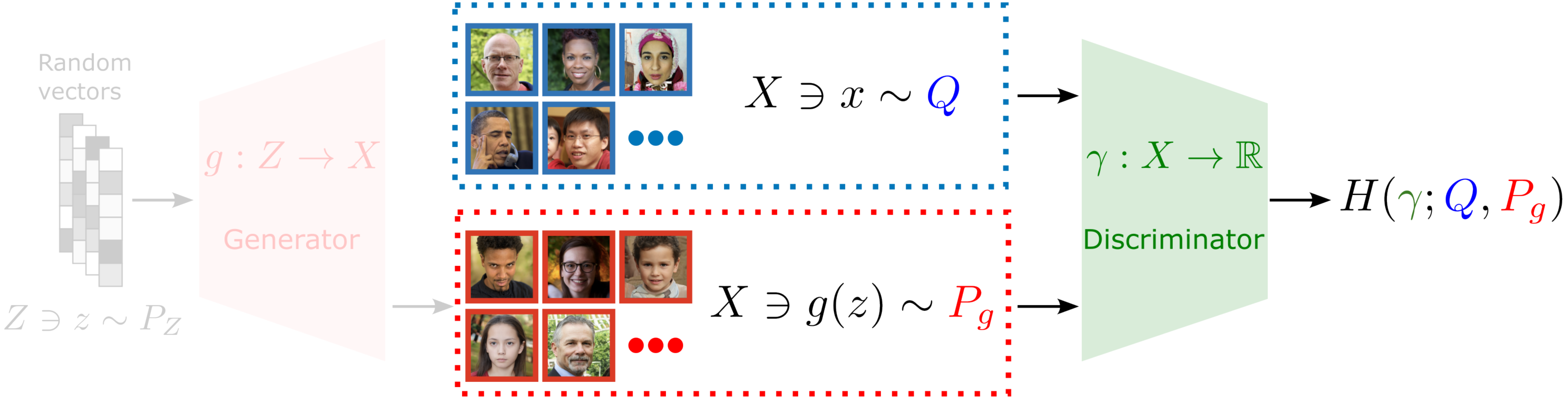
- **GAN:** <u>two-player game</u> for <u>learning</u> Q.
- **Generator** $g : Z \to X$

Generative Adversarial Networks (GANs)



- **GAN:** two-player game for learning Q.
- **Generator** $g : Z \to X$
- **Discriminator** $\gamma : X \to \mathbb{R}$
- Discrepancy $H(\gamma; Q, P_g)$
  - e.g., $H(\gamma; Q, P_g) = \mathbb{E}_Q[\gamma] - \mathbb{E}_{P_g}[\gamma]$

Generative Adversarial Networks (GANs)



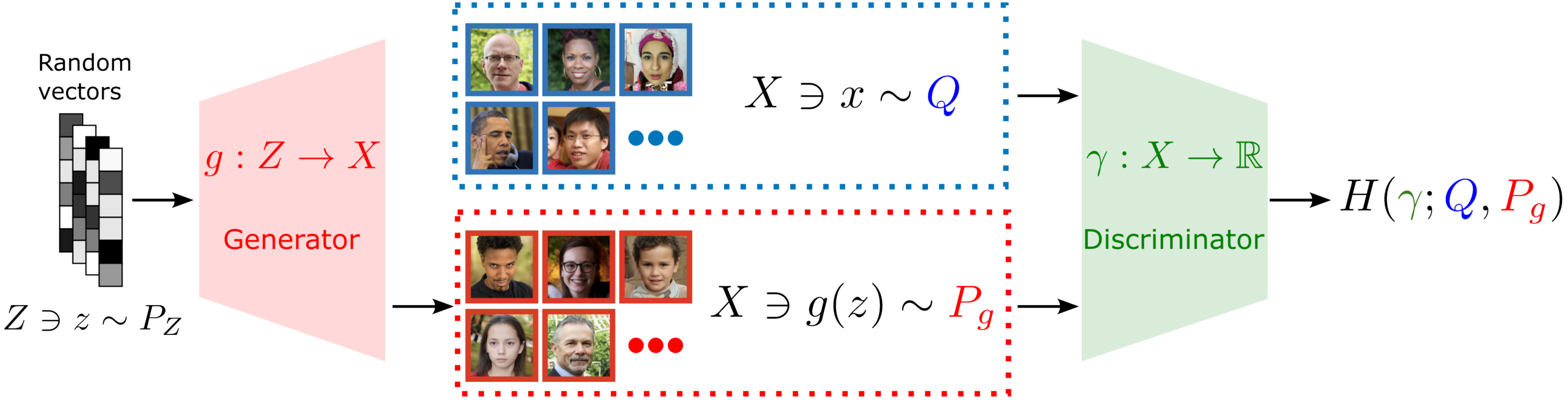
- **GAN:** two-player game for learning Q.
- **Generator** $g : Z \to X$
- **Discriminator** $\gamma : X \to \mathbb{R}$
- Discrepancy $H(\gamma; Q, P_g)$
  - e.g., $H(\gamma; Q, P_g) = \mathbb{E}_Q[\gamma] - \mathbb{E}_{P_g}[\gamma]$

- **Probability divergence**

$$D_H^\Gamma(Q \| P_g) = \max_{\gamma \in \Gamma} H(\gamma; Q, P_g)$$

  - e.g., Wasserstein distance: $\Gamma = \mathrm{Lip}_1(X)$

Generative Adversarial Networks (GANs)



- **GAN:** <u>two-player game</u> for <u>learning</u> Q.
- **Generator** $g : Z \to X$
- **Discriminator** $\gamma : X \to \mathbb{R}$
- Discrepancy $H(\gamma; Q, P_g)$
  - e.g., $H(\gamma; Q, P_g) = \mathbb{E}_Q[\gamma] - \mathbb{E}_{P_g}[\gamma]$

- **Probability divergence**

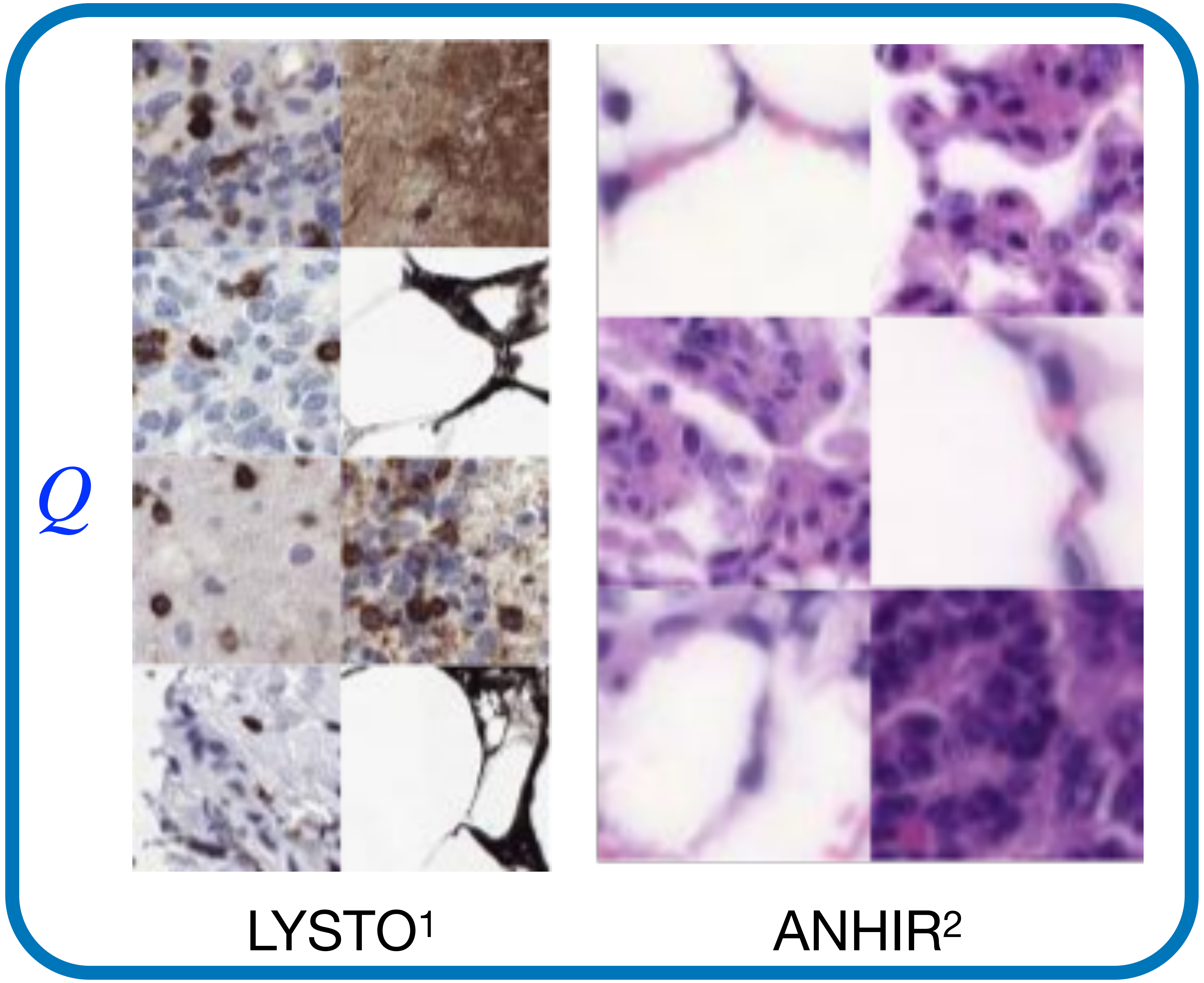
$$D_H^\Gamma(Q \| P_g) = \max_{\gamma \in \Gamma} H(\gamma; Q, P_g)$$

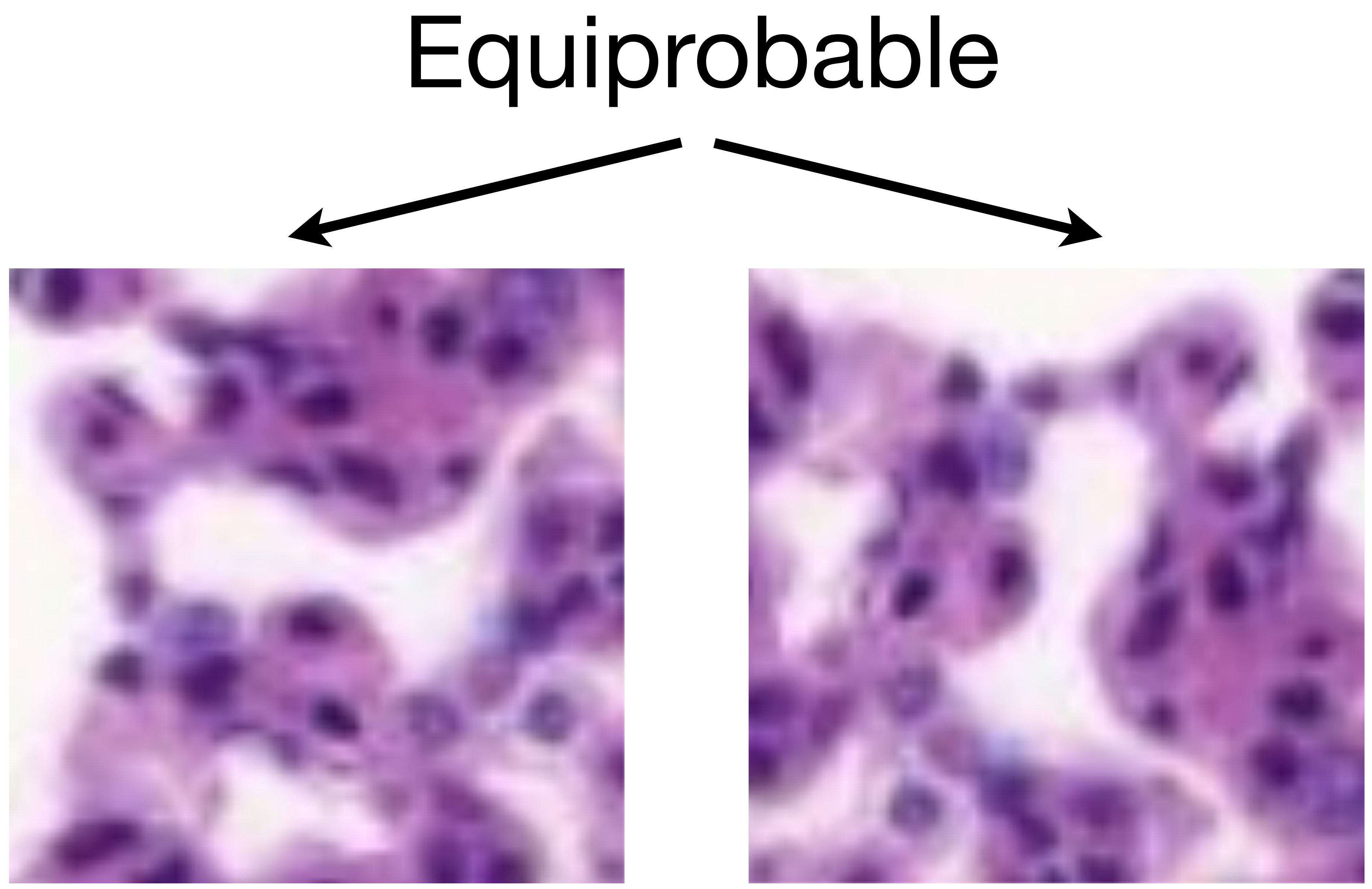
  - e.g., <u>Wasserstein distance</u>: $\Gamma = \mathrm{Lip}_1(X)$

- **GAN:** $\min_{g \in G} D_H^\Gamma(Q \| P_g) = \min_{g \in G} \max_{\gamma \in \Gamma} H(\gamma; Q, P_g)$

Many real-world data have intrinsic structures

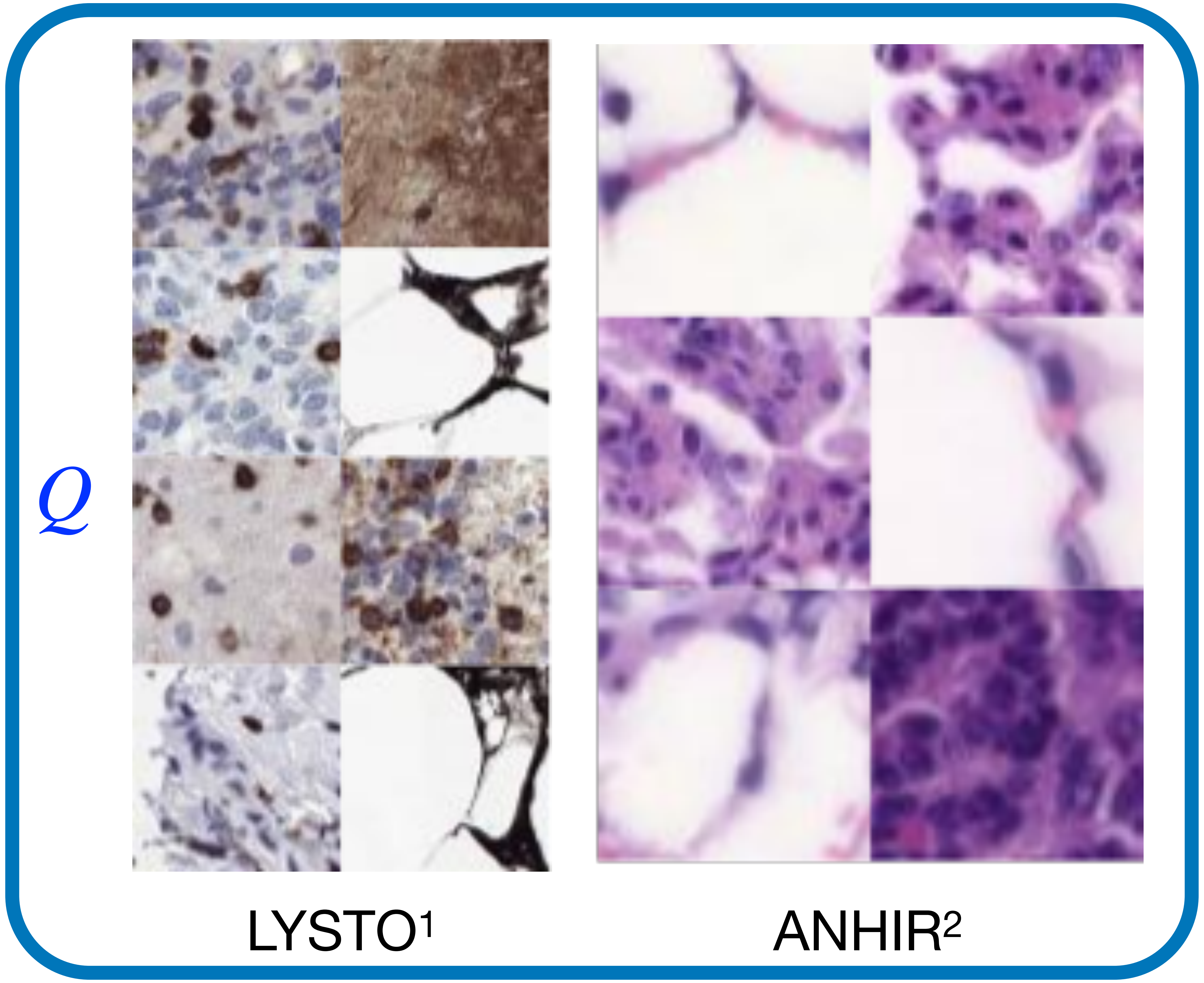
Many real-world data have intrinsic structures

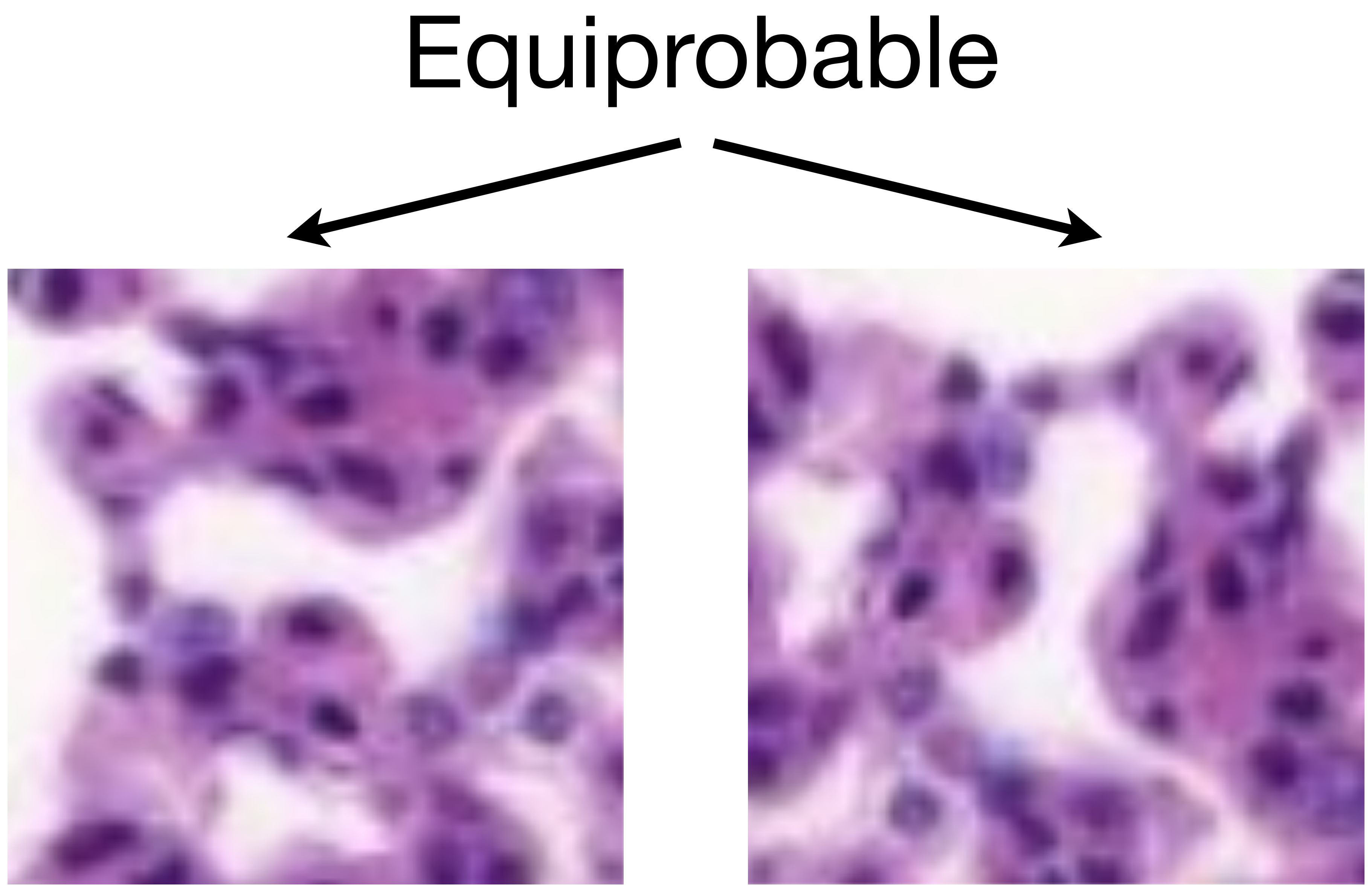




1. Ciompi et al., Zenodo 2019; 2. Borovec et al., IEEE Transactions on Medical Imaging 2020

Many real-world data have intrinsic structures





- An image and its rotated copies should have the same **likelihood/probability**.
- Q has group symmetry: **invariant** under rotation.

1. Ciompi et al., Zenodo 2019; 2. Borovec et al., IEEE Transactions on Medical Imaging 2020

Topics

Topics

- **How can we impose symmetry in generative models (GANs)?**

Topics

- How can we impose symmetry in generative models (GANs)?

- How can we quantify the performance gain?

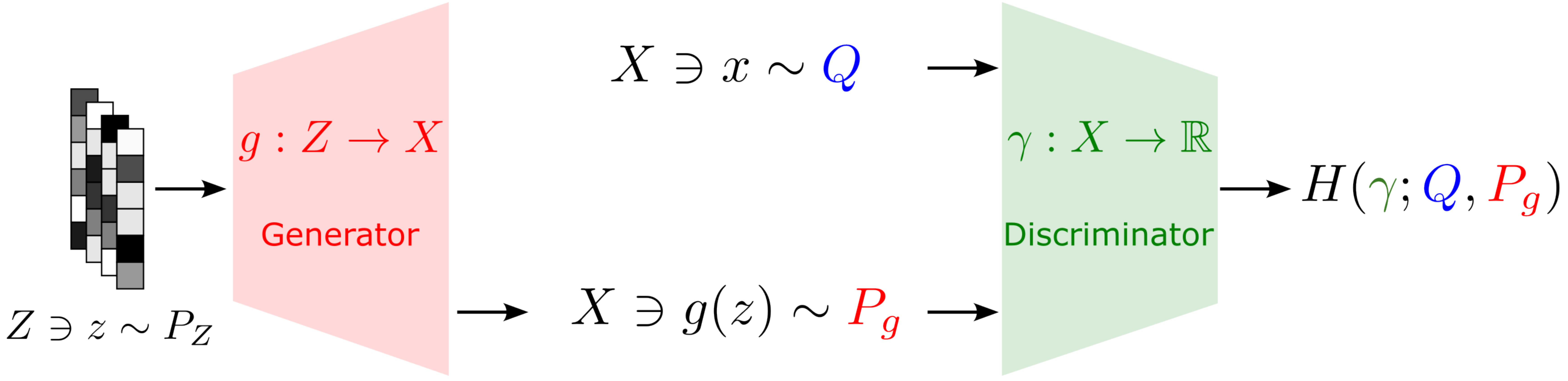
Topics

- How can we impose symmetry in generative models (GANs)?
- How can we quantify the performance gain?
- How can we flexibly impose general structures in diffusion models?

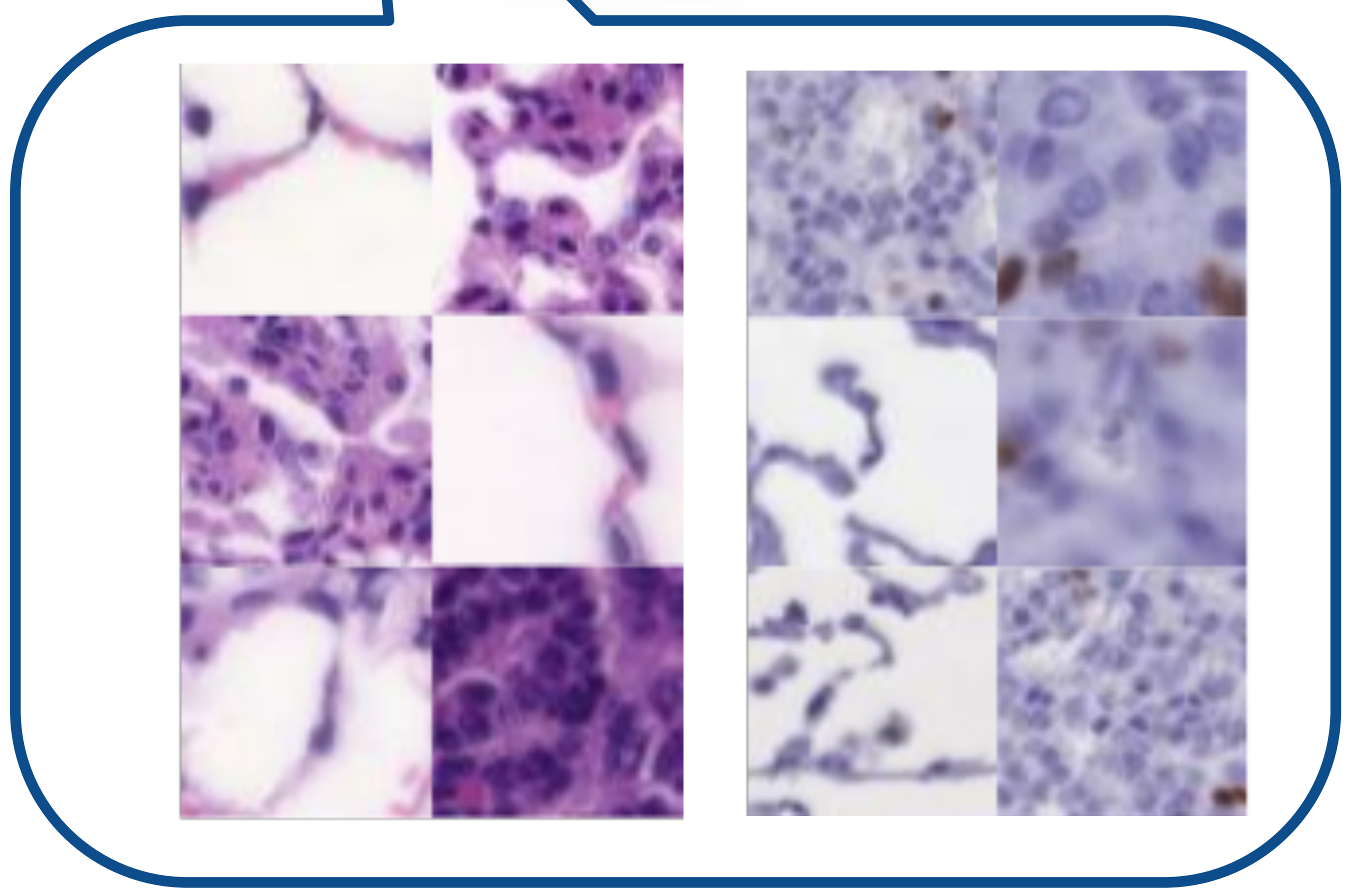
Topics

- How can we impose symmetry in generative models (GANs)?
- How can we quantify the performance gain?
- How can we flexibly impose general structures in diffusion models?

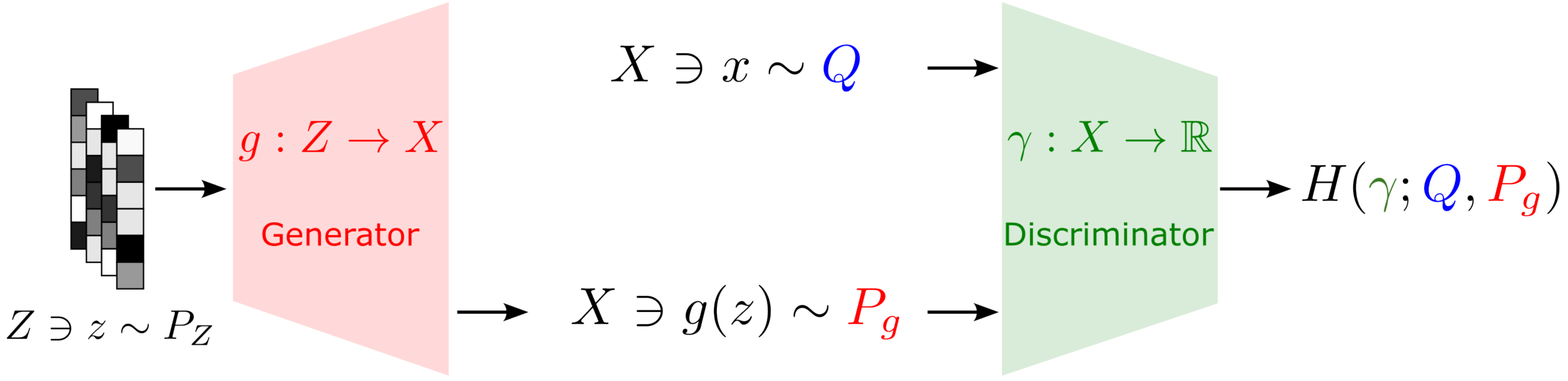
GAN with embedded structure



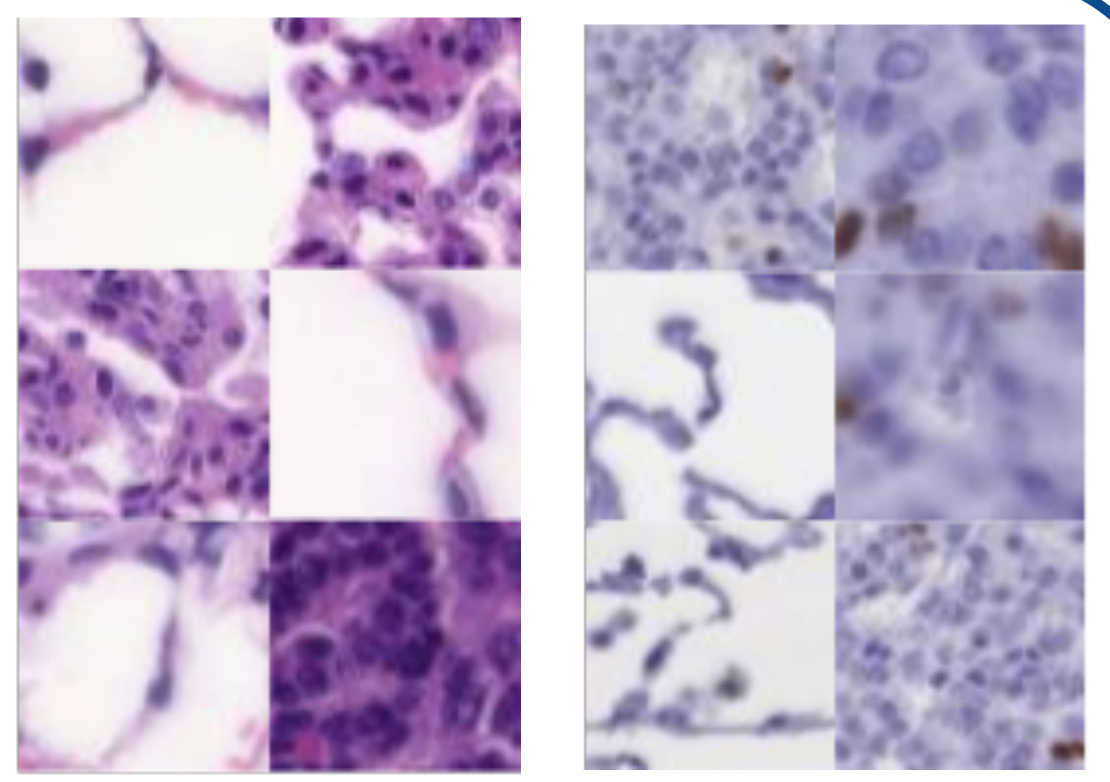
$$\min_{g\in G} D_H^\Gamma(Q\|P_g) = \min_{g\in G}\max_{\gamma\in\Gamma} H(\gamma; Q, P_g), \quad \underline{Q \text{ is } \Sigma\text{-invariant}}$$



GAN with embedded structure

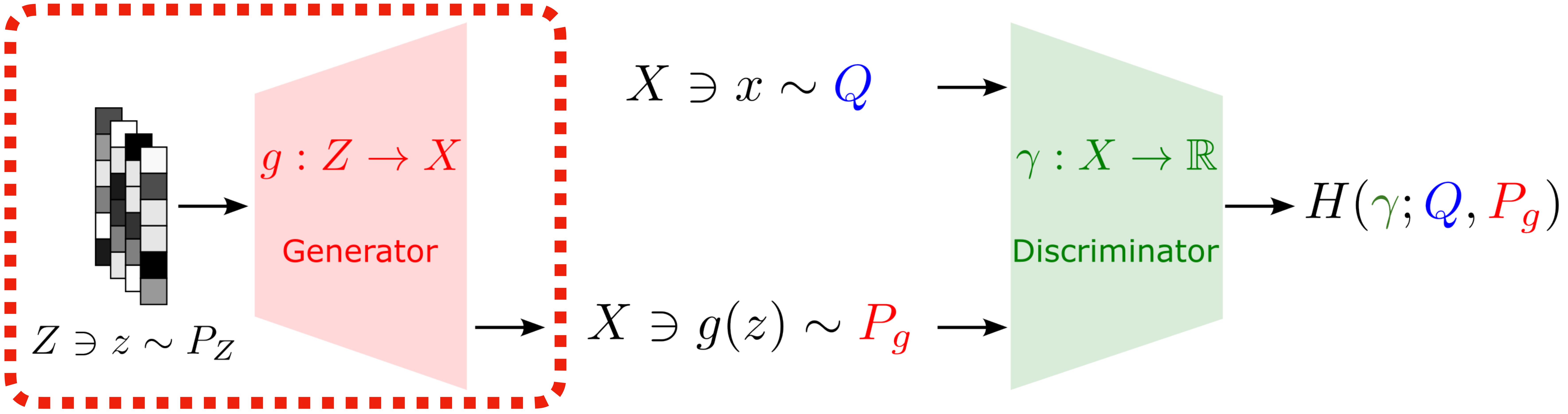


$$\min_{g \in G} D_H^{\Gamma}(Q \| P_g) = \min_{g \in G} \max_{\gamma \in \Gamma} H(\gamma; Q, P_g), \quad \underline{Q \text{ is } \Sigma\text{-invariant}}$$

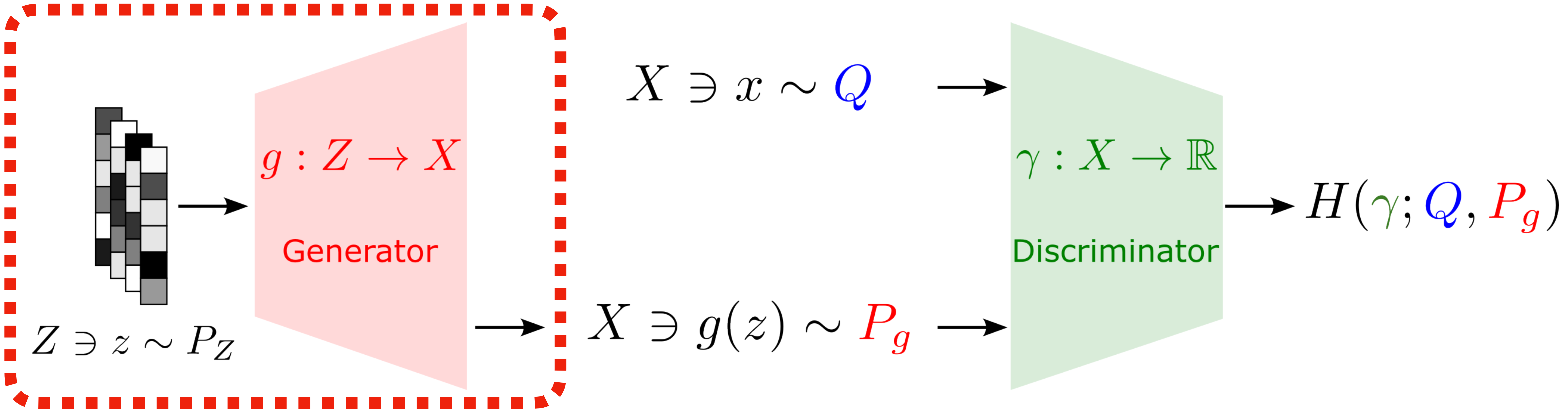


- Target distribution Q is invariant to a group Σ.
- **How to incorporate such structure into g and γ?**

Theorem 1: “smarter” generator



Theorem 1: “smarter” generator

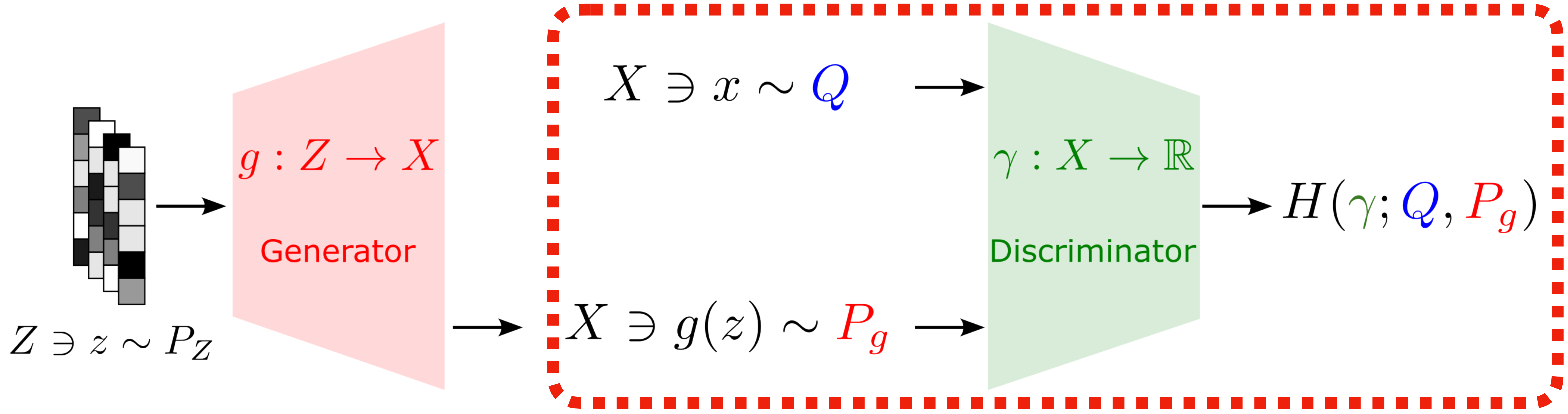


Theorem [Birrell, Katsoulakis, Rey-Bellet, **Z.**, *ICML* 2022]

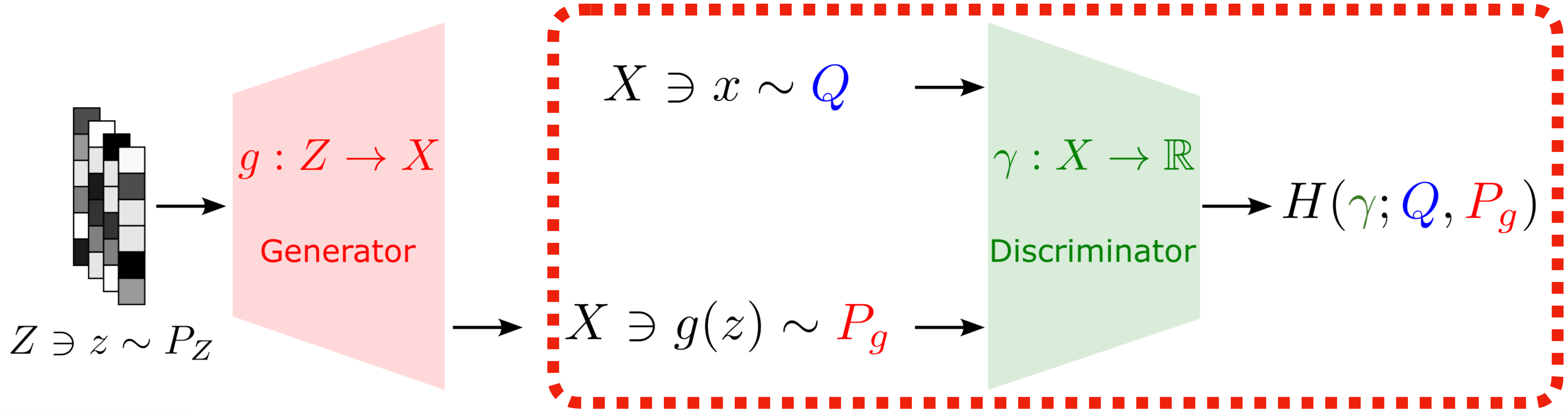
If P_Z is Σ-invariant and $g : Z \to X$ is Σ-equivariant, the generated measure P_g is Σ-invariant.

- Symmetry information of Q can be built into a “smarter” generator.

Theorem 2: “smarter” discriminator



Theorem 2: “smarter” discriminator



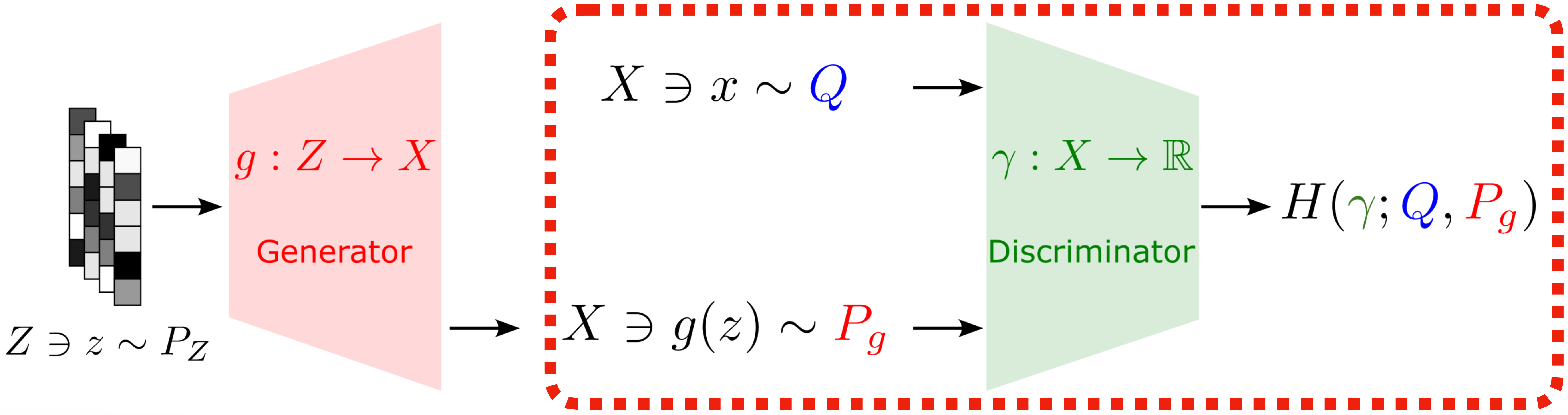
Theorem [Birrell, Katsoulakis, Rey-Bellet, **Z.**, *ICML* 2022]

Under mild assumptions on Σ and Γ, if the distributions P, Q are **Σ-invariant**, then

$$\sup_{\gamma\in\Gamma} H(\gamma; Q, P) = D^{\Gamma}(Q\|P) = D^{\Gamma_\Sigma^{\text{inv}}}(Q\|P) = \sup_{\gamma\in\Gamma_\Sigma^{\text{inv}}} H(\gamma; Q, P),$$

- $\Gamma_\Sigma^{\text{inv}}$ is the subset of Γ that includes only **Σ-invariant** “smarter” discriminators.

Theorem 2: “smarter” discriminator



Theorem [Birrell, Katsoulakis, Rey-Bellet, **Z.**, *ICML* 2022]

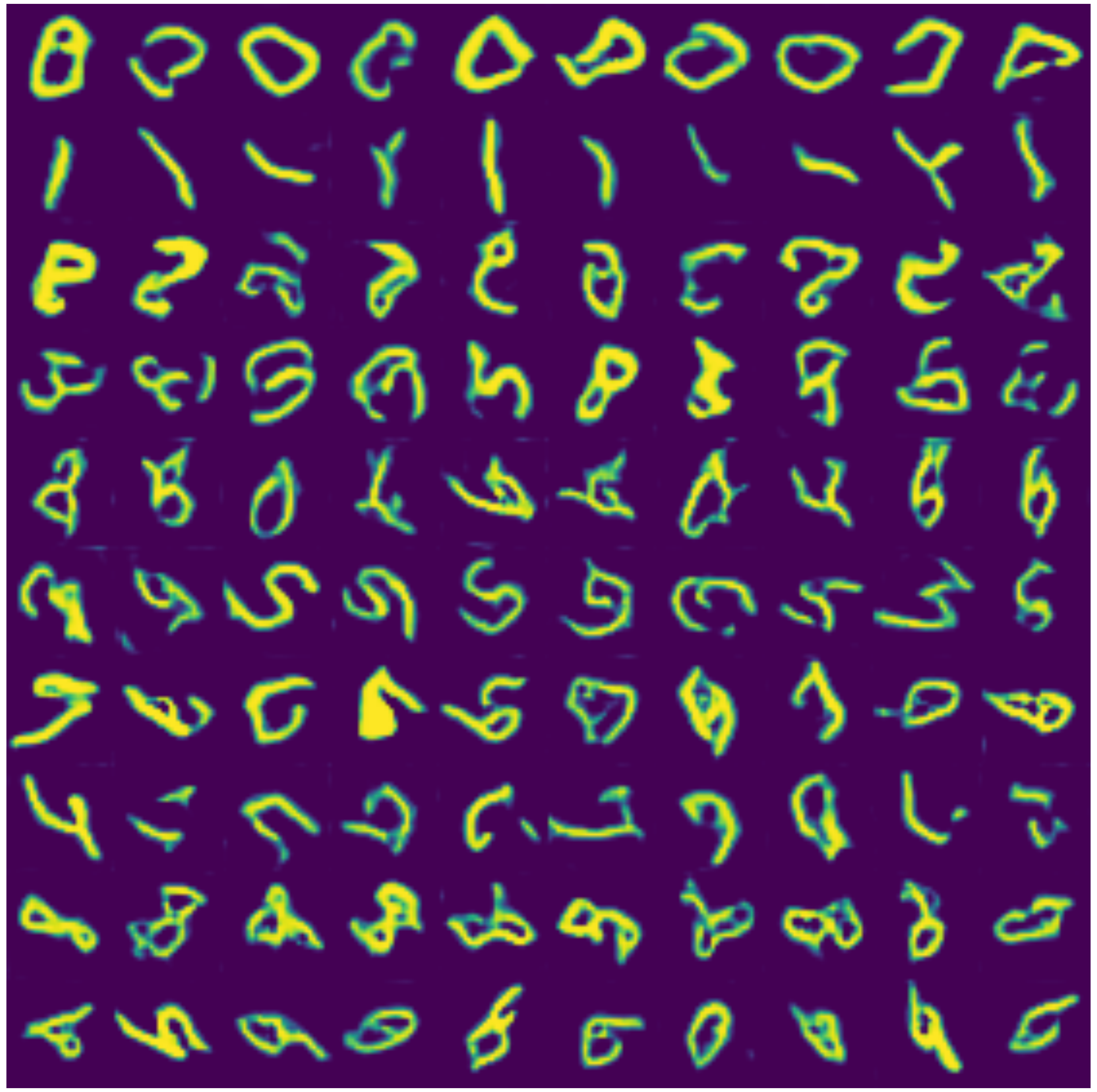
Under mild assumptions on Σ and Γ, if the distributions P, Q are **Σ-invariant**, then

$$\sup_{\gamma\in\Gamma} H(\gamma; Q, P) = D^{\Gamma}(Q\|P) = D^{\Gamma_{\Sigma}^{\text{inv}}}(Q\|P) = \sup_{\gamma\in\Gamma_{\Sigma}^{\text{inv}}} H(\gamma; Q, P),$$

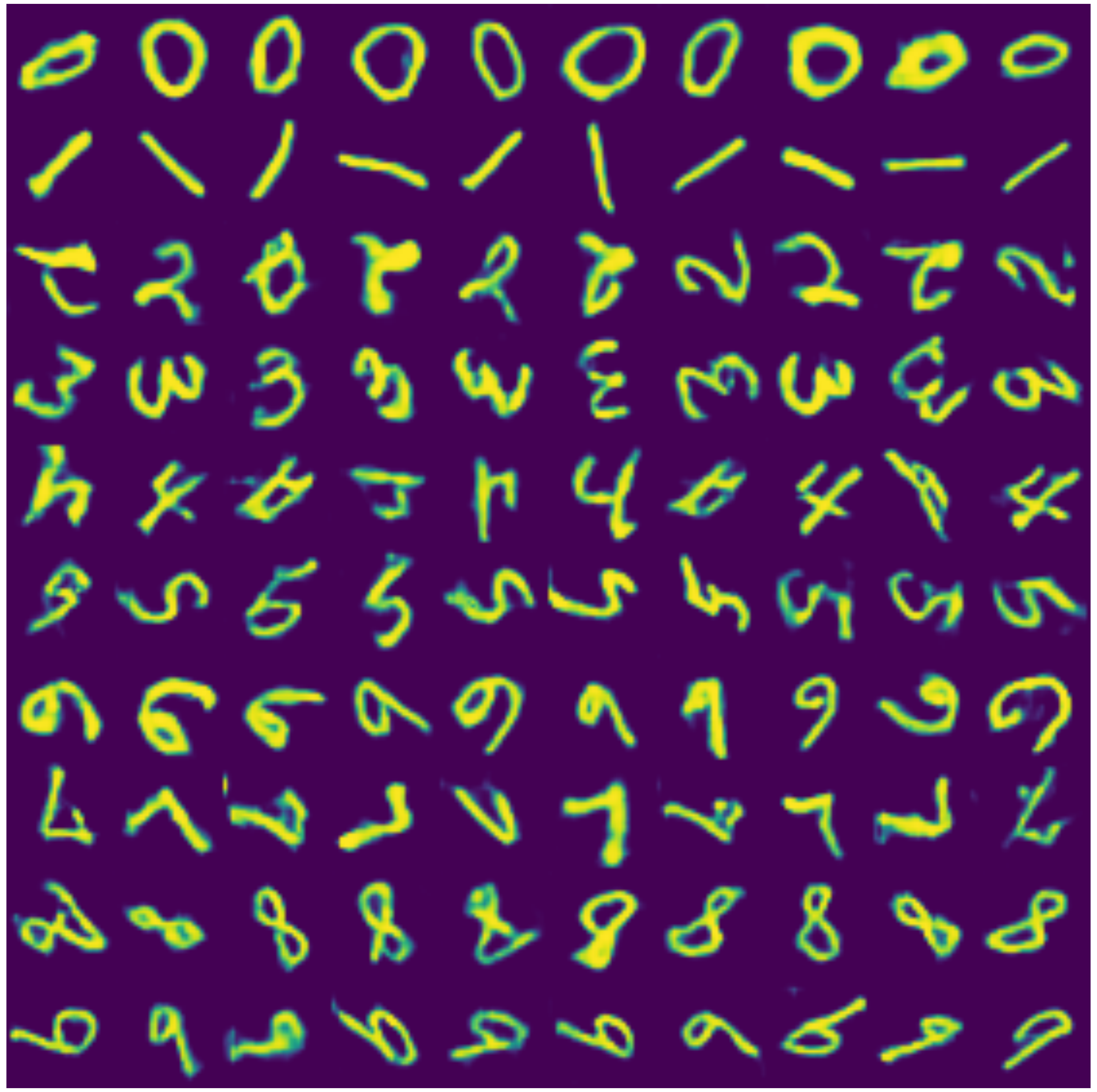
- $\Gamma_{\Sigma}^{\text{inv}}$ is the subset of Γ that includes only **Σ-invariant** “smarter” discriminators.
- We only need to optimize over the **smaller** subset $\Gamma_{\Sigma}^{\text{inv}}$ to “compute” the divergence $D^{\Gamma}(Q\|P)$.

Rotated MNIST with 1% training data

“Ignorant” players

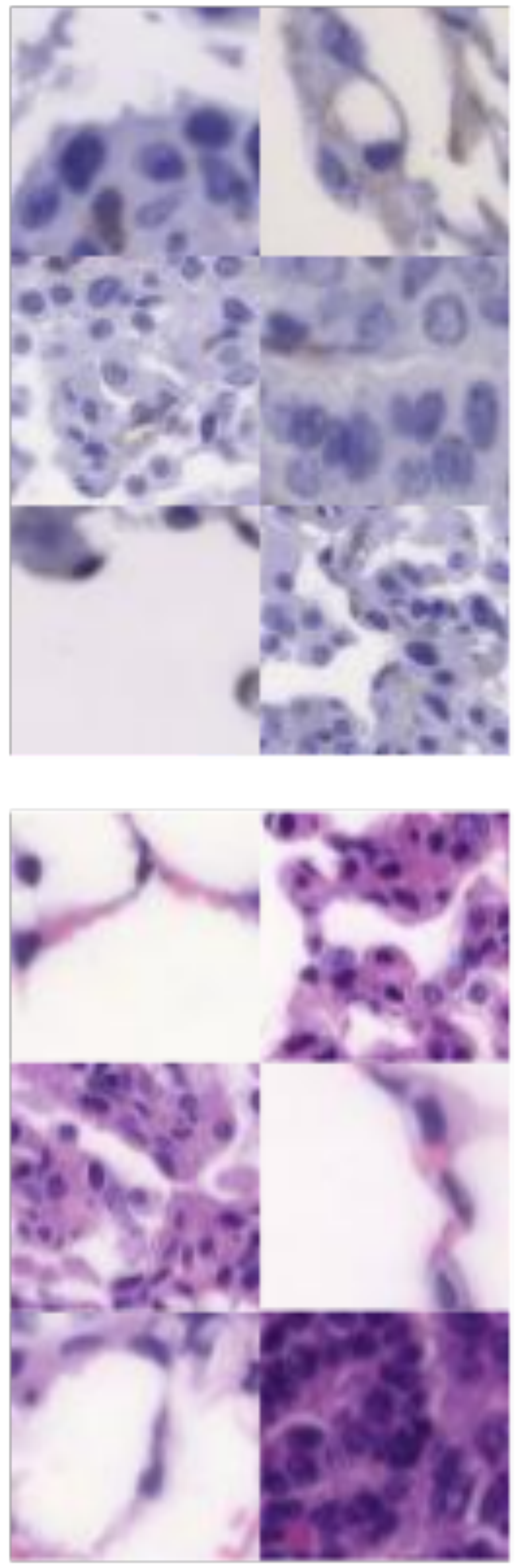


“Smart” players

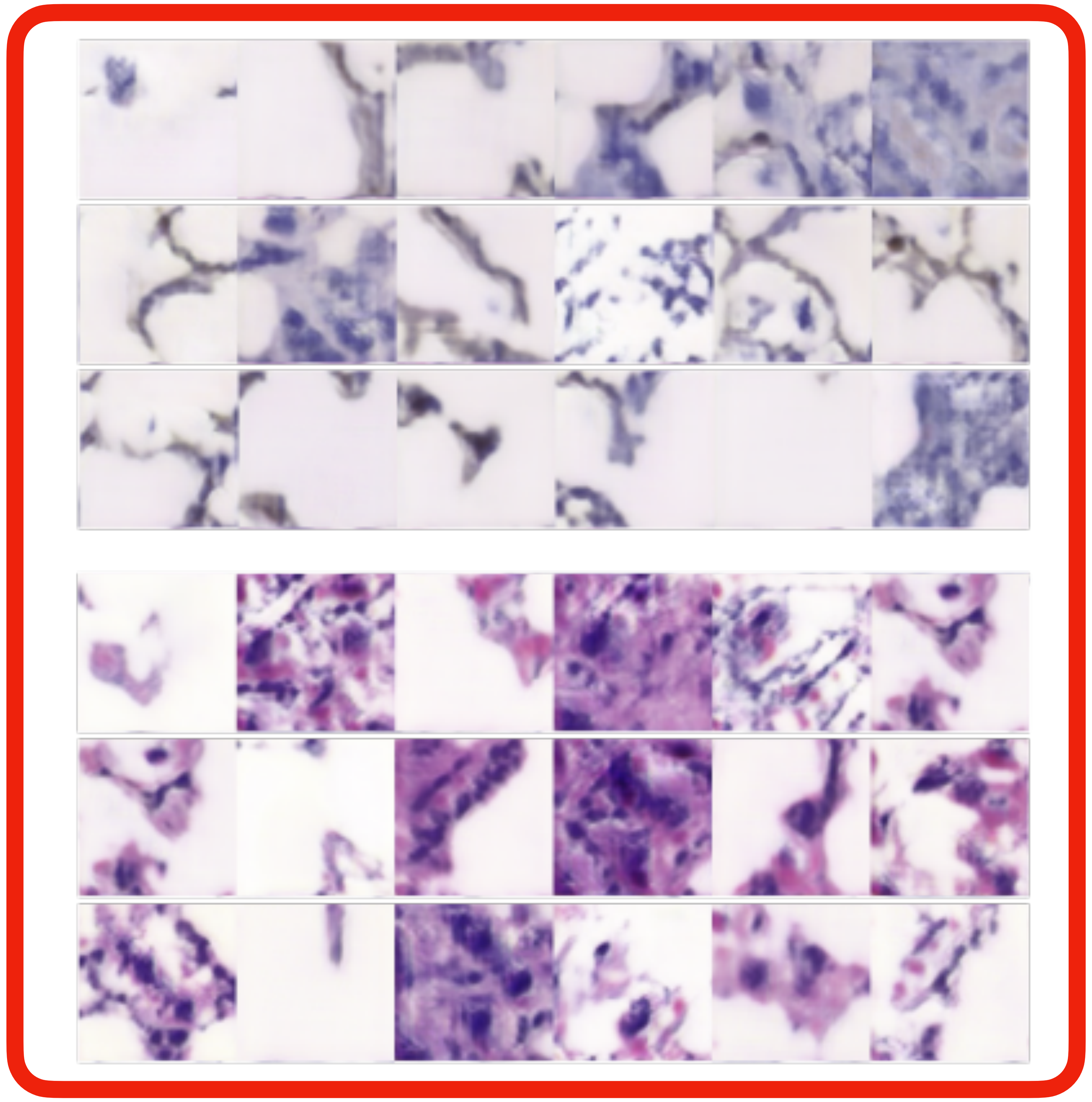


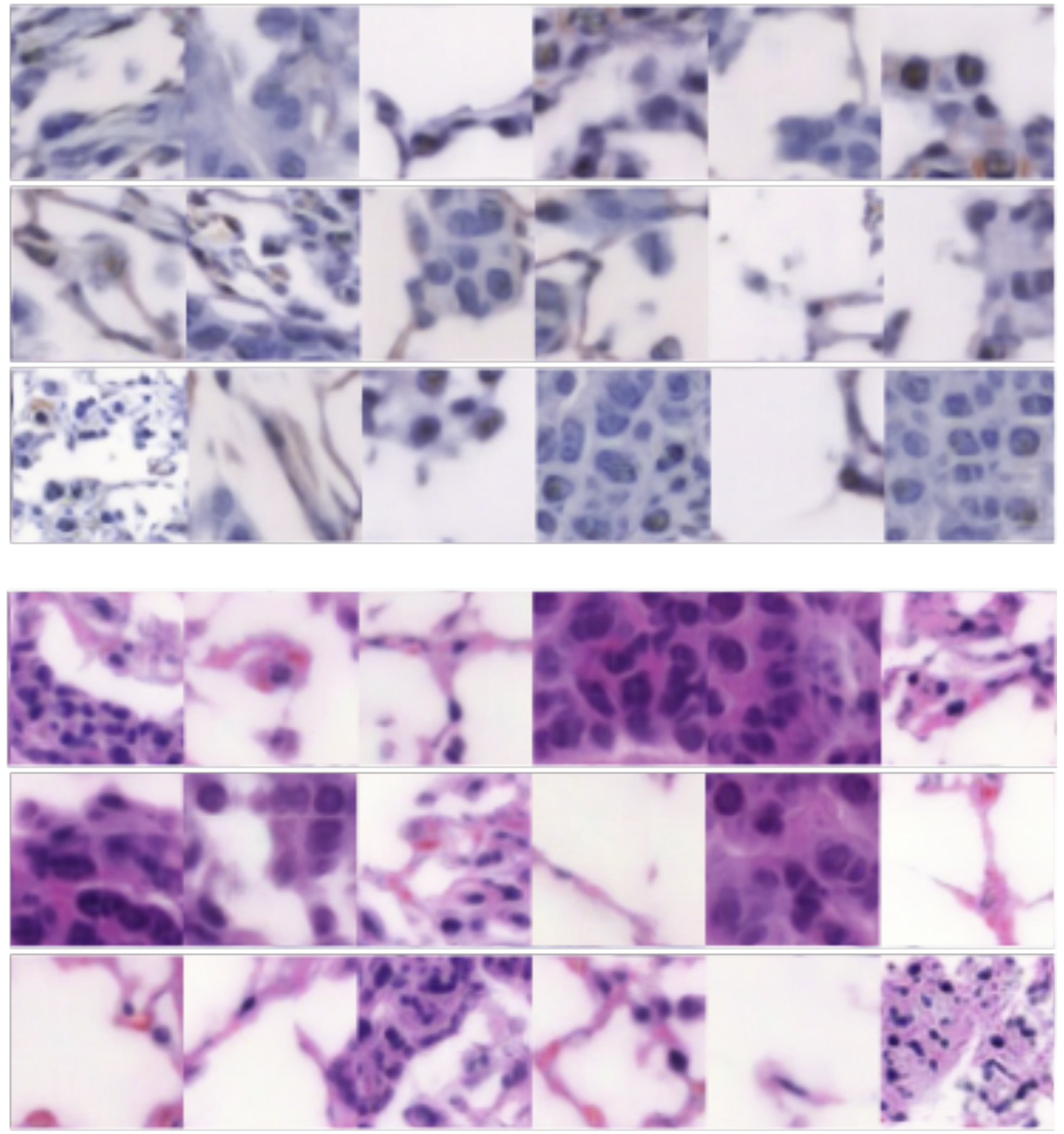
Medical images (ANHIR)

Real Samples



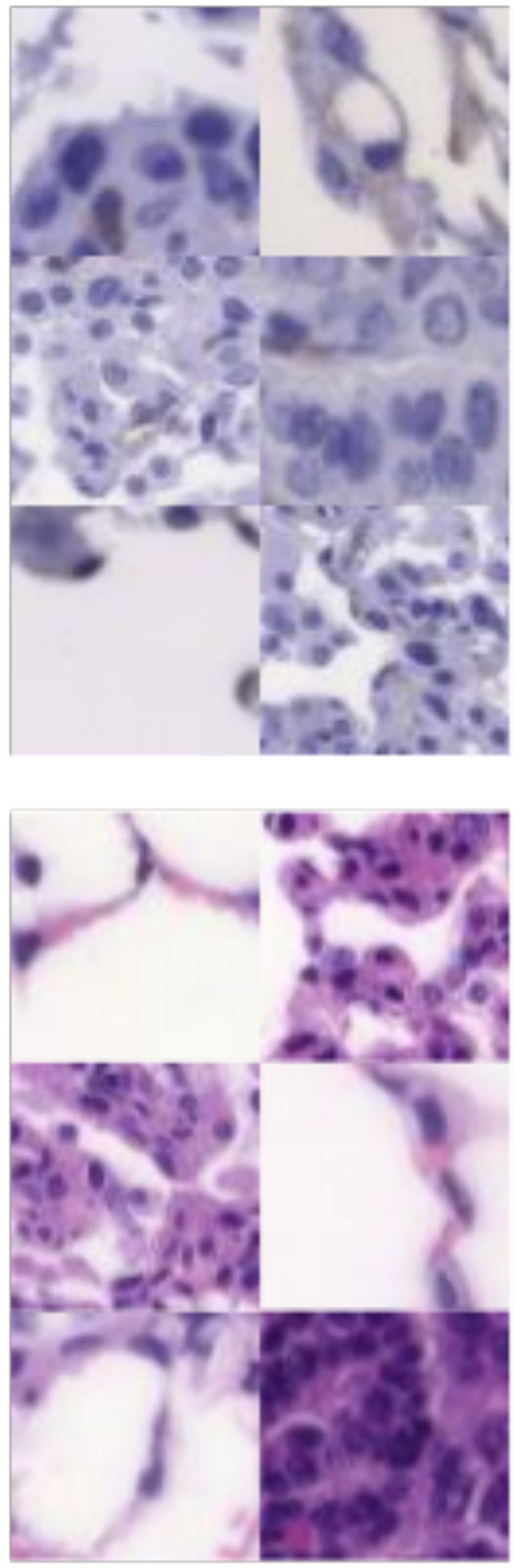
“Ignorant” players



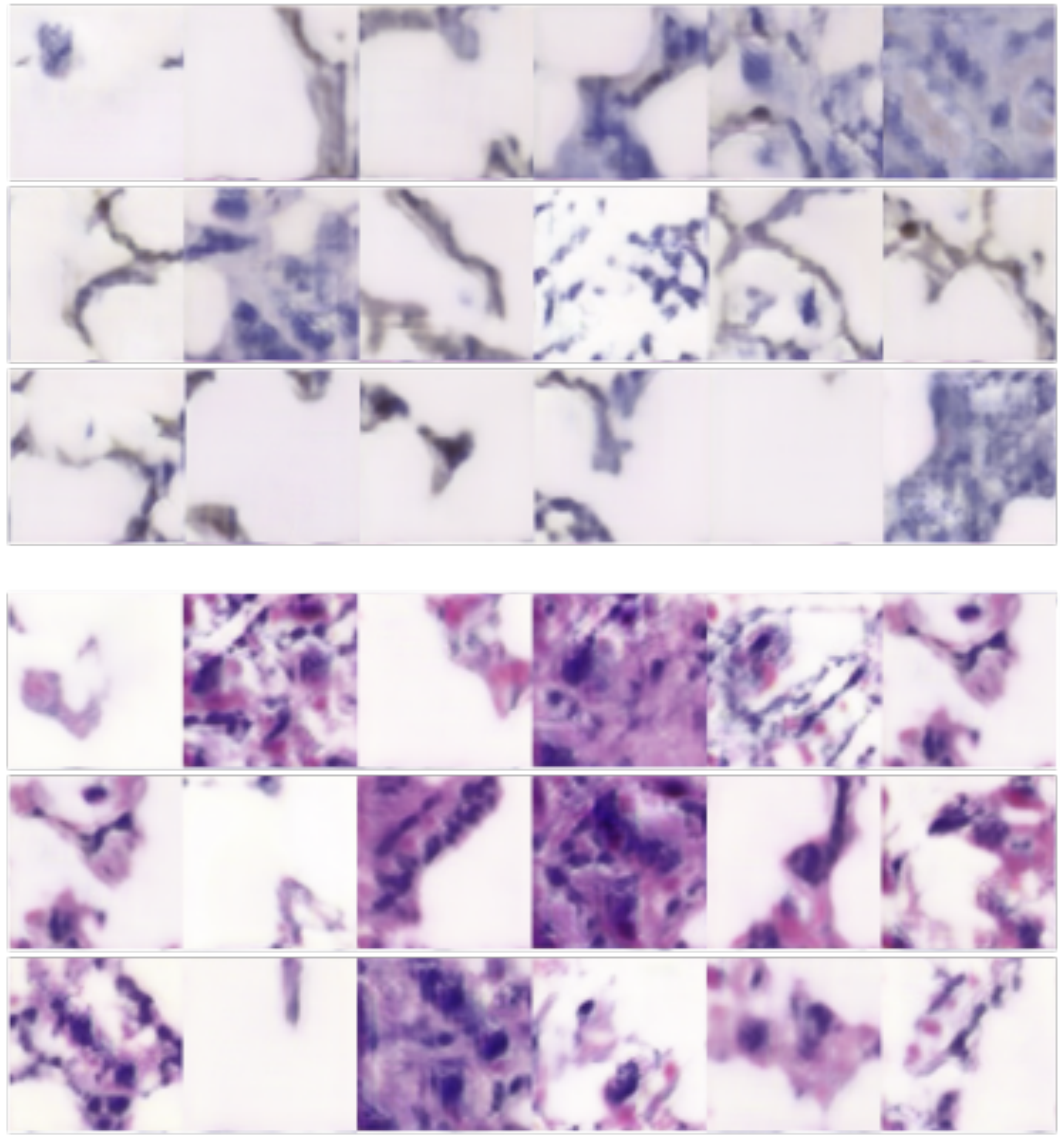


Medical images (ANHIR)

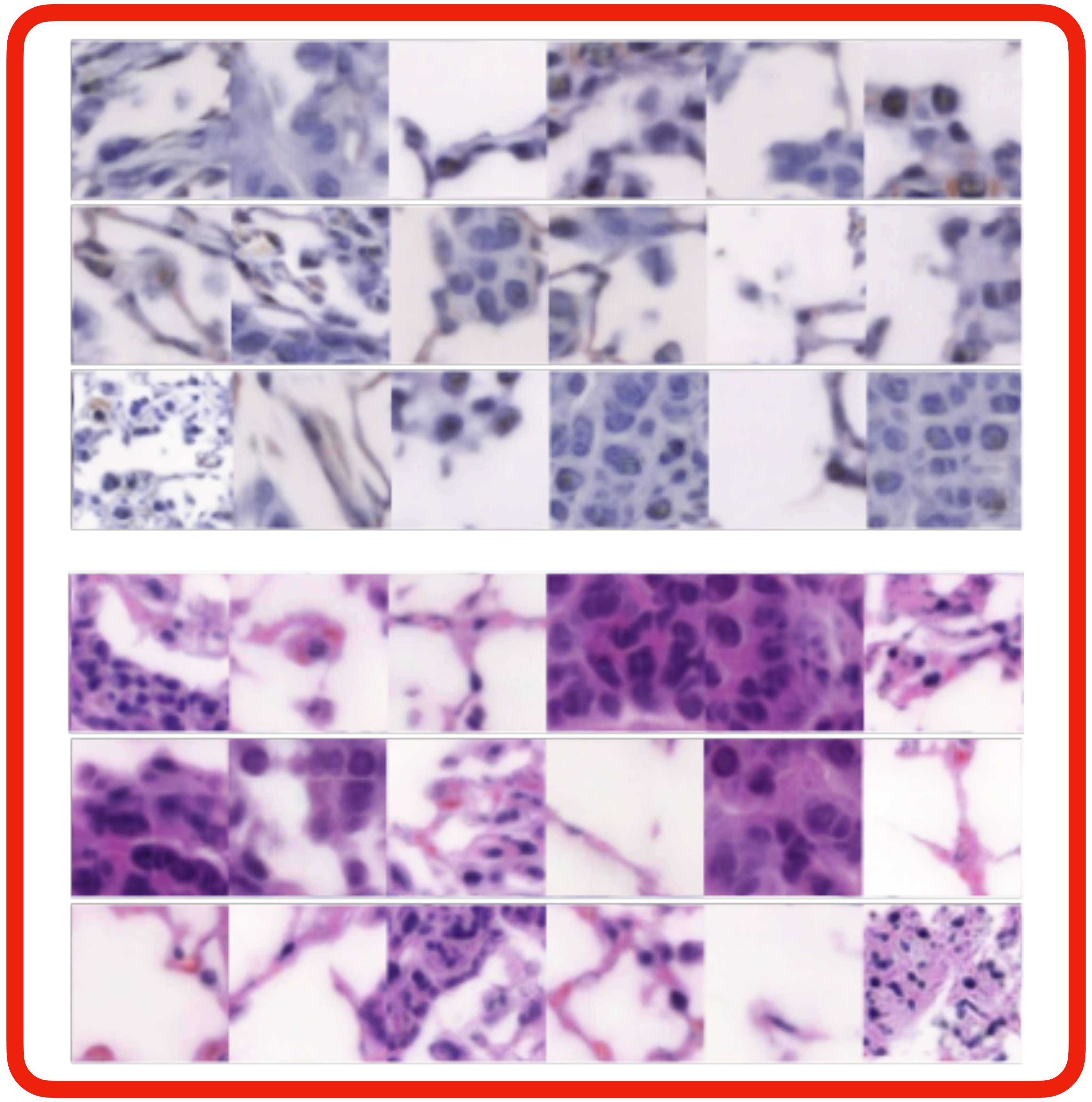
Real Samples



“Ignorant” players



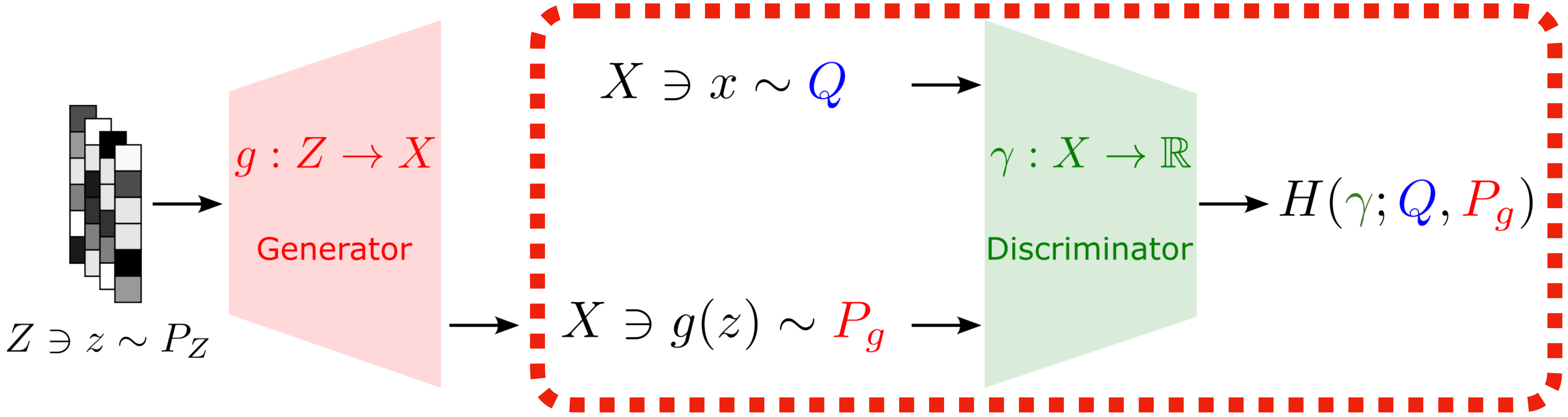
“Smart” players



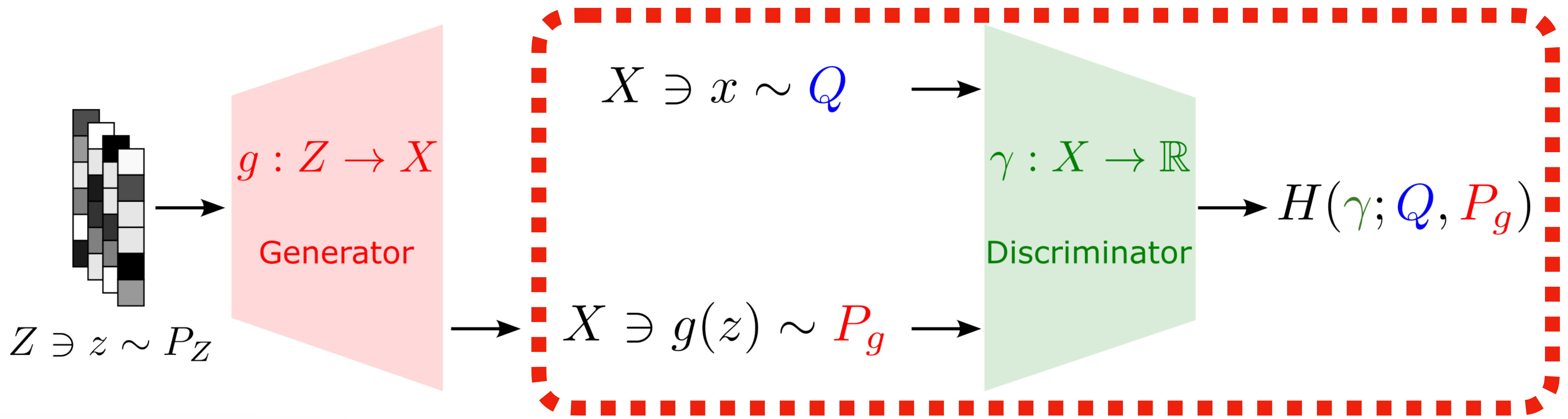
Topics

- How can we impose symmetry in generative models (GANs)?
- How can we quantify the performance gain?
- How can we flexibly impose general structures in diffusion models?

What is the reason behind the improvement?



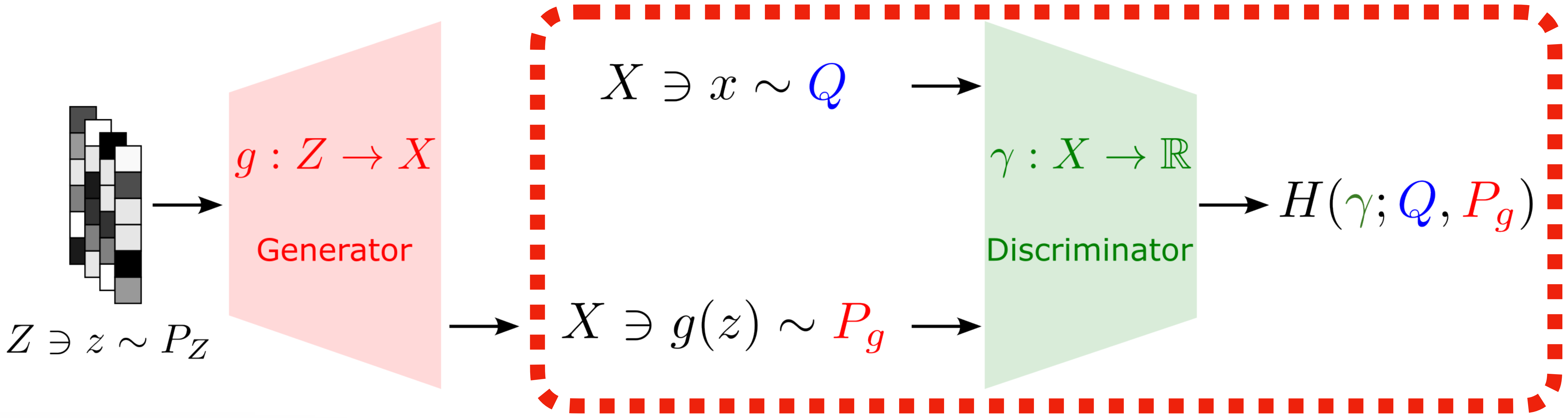
What is the reason behind the improvement?



P, Q are **Σ-invariant** $\Longrightarrow D^{\Gamma}(Q\|P) = D^{\Gamma_{\Sigma}^{\text{inv}}}(Q\|P)$

- Reducing Γ to $\Gamma_{\Sigma}^{\text{inv}}$ can provide a better **empirical estimation** for $D^{\Gamma}(Q\|P)$.

What is the reason behind the improvement?

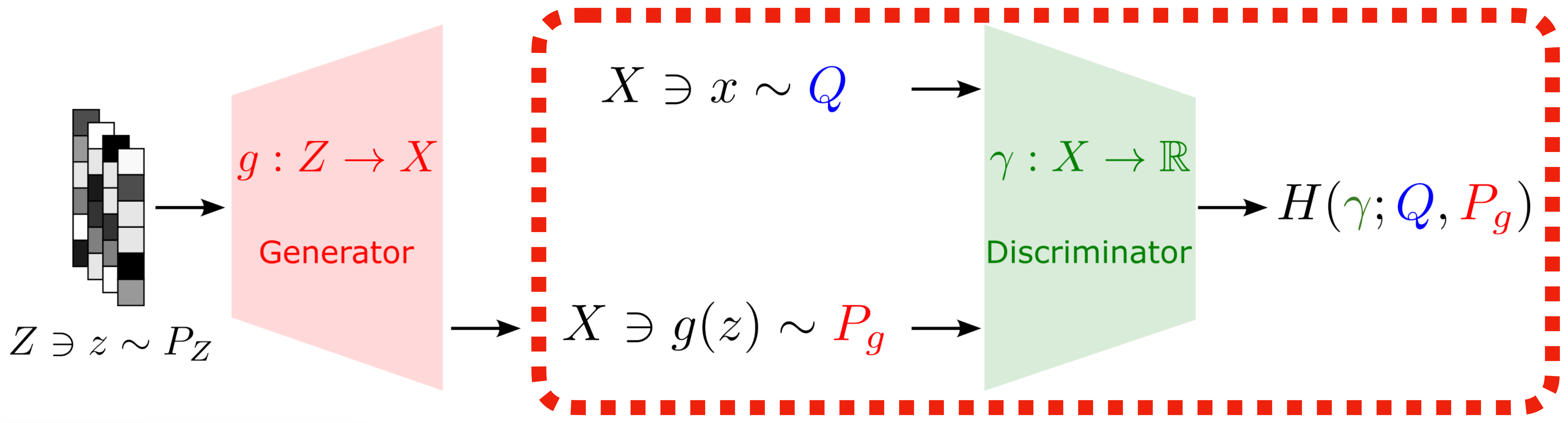


P, Q are Σ-**invariant** $\Longrightarrow D^{\Gamma}(Q\|P) = D^{\Gamma_{\Sigma}^{\mathrm{inv}}}(Q\|P)$

- Reducing Γ to $\Gamma_{\Sigma}^{\mathrm{inv}}$ can provide a better **empirical estimation** for $D^{\Gamma}(Q\|P)$.

$$\sup_{\gamma \in \Gamma} H(\gamma; Q, P) = D^{\Gamma}(Q\|P)$$

What is the reason behind the improvement?



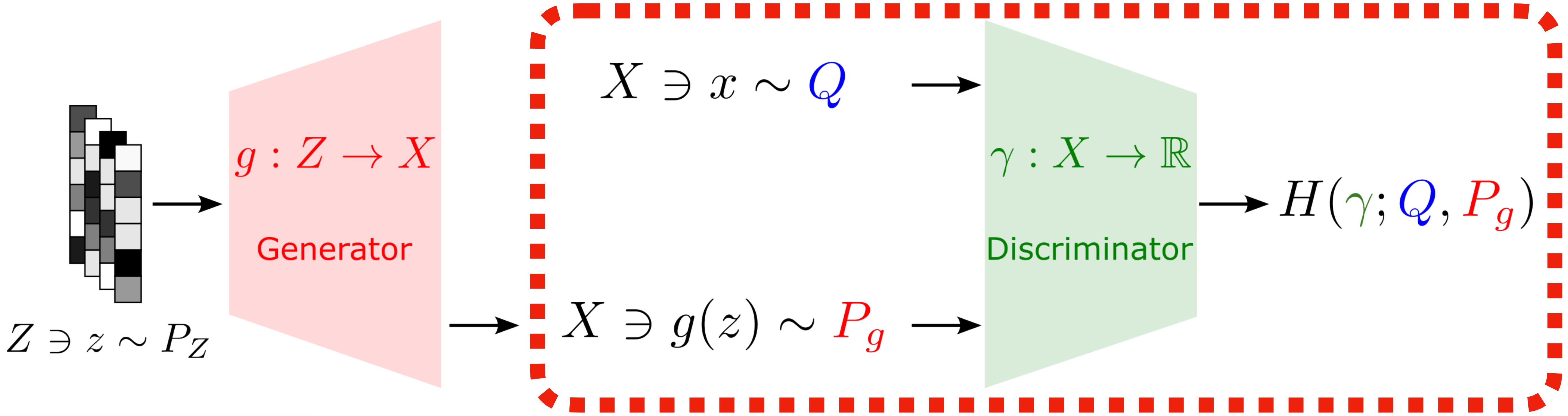
P, Q are Σ**-invariant** $\Longrightarrow D^{\Gamma}(Q\|P) = D^{\Gamma_{\Sigma}^{\text{inv}}}(Q\|P)$

- Reducing Γ to $\Gamma_{\Sigma}^{\text{inv}}$ can provide a better **empirical estimation** for $D^{\Gamma}(Q\|P)$.

$$\sup_{\gamma \in \Gamma} H(\gamma; Q, P) = D^{\Gamma}(Q\|P) \approx D^{\Gamma}(Q_n\|P_m)$$

- $(x_1, \cdots, x_m) \sim P$, $(y_1, \cdots, y_n) \sim Q \Longrightarrow$ **Empirical measures** $P_m = \frac{1}{m}\sum_{i=1}^{m} \delta_{x_i}$, $Q_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{y_i}$

What is the reason behind the improvement?



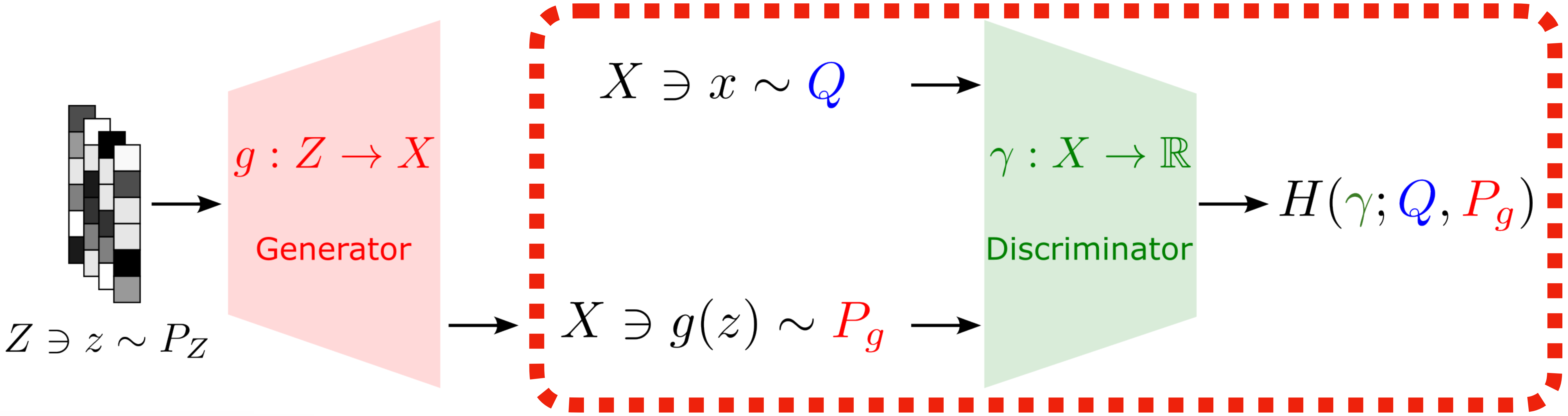
P, Q are Σ**-invariant** $\implies D^{\Gamma}(Q\|P) = D^{\Gamma_{\Sigma}^{\text{inv}}}(Q\|P)$

- Reducing Γ to $\Gamma_{\Sigma}^{\text{inv}}$ can provide a better **empirical estimation** for $D^{\Gamma}(Q\|P)$.

$$\sup_{\gamma \in \Gamma} H(\gamma; Q, P) = D^{\Gamma}(Q\|P) \approx \cancel{D^{\Gamma}(Q_n\|P_m)} \approx \boxed{D^{\Gamma_{\Sigma}^{\text{inv}}}(Q_n\|P_m)}$$

- $(x_1, \cdots, x_m) \sim P, (y_1, \cdots, y_n) \sim Q \implies$ **Empirical measures** $P_m = \frac{1}{m}\sum_{i=1}^{m} \delta_{x_i}, Q_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{y_i}$

What is the reason behind the improvement?



P, Q are **Σ-invariant** $\Longrightarrow D^{\Gamma}(Q\|P) = D^{\Gamma_{\Sigma}^{\text{inv}}}(Q\|P)$

- Reducing Γ to $\Gamma_{\Sigma}^{\text{inv}}$ can provide a better **empirical estimation** for $D^{\Gamma}(Q\|P)$.

$$\sup_{\gamma \in \Gamma} H(\gamma; Q, P) = D^{\Gamma}(Q\|P) \approx \cancel{D^{\Gamma}(Q_n\|P_m)} \approx \boxed{D^{\Gamma_{\Sigma}^{\text{inv}}}(Q_n\|P_m)}$$

- $(x_1, \cdots, x_m) \sim P, (y_1, \cdots, y_n) \sim Q \Longrightarrow$ **Empirical measures** $P_m = \frac{1}{m}\sum_{i=1}^{m} \delta_{x_i}, Q_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{y_i}$

- **Question:** How much **more accurate** is the new estimation?

Maximum Mean Discrepancy (MMD)

- $\mathrm{MMD}(Q, P) = \sup_{\gamma \in \Gamma} \{\mathbb{E}_Q[\gamma] - \mathbb{E}_P[\gamma]\}$. Γ is the unit ball in some **RKHS** $\mathscr{H}$ with kernel $k(x, y)$.
- **Estimator:** $\mathrm{MMD}^{\Sigma}(Q_n, P_m) = \sup_{\gamma \in \Gamma_{\Sigma}^{\text{inv}}} \{\mathbb{E}_{Q_n}[\gamma] - \mathbb{E}_{P_m}[\gamma]\} \approx \mathrm{MMD}(Q, P)$

Maximum Mean Discrepancy (MMD)

- $\mathrm{MMD}(Q,P) = \sup_{\gamma\in\Gamma}\{\mathbb{E}_Q[\gamma] - \mathbb{E}_P[\gamma]\}$. Γ is the unit ball in some **RKHS** $\mathscr{H}$ with kernel $k(x,y)$.
- **Estimator:** $\mathrm{MMD}^{\Sigma}(Q_n,P_m) = \sup_{\gamma\in\Gamma_\Sigma^{\text{inv}}}\{\mathbb{E}_{Q_n}[\gamma] - \mathbb{E}_{P_m}[\gamma]\} \approx \mathrm{MMD}(Q,P)$

Theorem [Chen, Katsoulakis, Rey-Bellet, **Z.**, *ICML* 2023]

Let Σ be a finite group acting on X, and P, Q are Σ-invariant. With high probability,

$$\left|\mathrm{MMD}(Q,P) - \mathrm{MMD}^{\Sigma}(Q_n,P_m)\right| = O\left(A_{\Sigma,k}\left(\frac{1}{\sqrt{m}} + \frac{1}{\sqrt{n}}\right)\right)$$

where $A_{\Sigma,k} = \sqrt{a_{\Sigma,k} + \frac{1-a_{\Sigma,k}}{|\Sigma|}}$, and $a_{\Sigma,k} \in (0,1)$ depends on Σ and the **kernel** $k(x,y)$.

Maximum Mean Discrepancy (MMD)

- $\text{MMD}(Q, P) = \sup_{\gamma \in \Gamma} \{\mathbb{E}_Q[\gamma] - \mathbb{E}_P[\gamma]\}$. Γ is the unit ball in some **RKHS** $\mathscr{H}$ with kernel $k(x, y)$.
- **Estimator:** $\text{MMD}^{\Sigma}(Q_n, P_m) = \sup_{\gamma \in \Gamma_{\Sigma}^{\text{inv}}} \{\mathbb{E}_{Q_n}[\gamma] - \mathbb{E}_{P_m}[\gamma]\} \approx \text{MMD}(Q, P)$

Theorem [Chen, Katsoulakis, Rey-Bellet, **Z.**, *ICML* 2023]

Let Σ be a finite group acting on X, and P, Q are Σ-invariant. With high probability,

$$\left| \text{MMD}(Q, P) - \text{MMD}^{\Sigma}(Q_n, P_m) \right| = O\left(A_{\Sigma,k} \left(\frac{1}{\sqrt{m}} + \frac{1}{\sqrt{n}} \right) \right)$$

where $A_{\Sigma,k} = \sqrt{a_{\Sigma,k} + \frac{1 - a_{\Sigma,k}}{|\Sigma|}}$

if $a_{\Sigma,k} \approx 0$

Maximum Mean Discrepancy (MMD)

- $\mathrm{MMD}(Q,P) = \sup_{\gamma\in\Gamma}\{\mathbb{E}_Q[\gamma] - \mathbb{E}_P[\gamma]\}$. Γ is the unit ball in some **RKHS** $\mathscr{H}$ with kernel $k(x,y)$.
- **Estimator:** $\mathrm{MMD}^{\Sigma}(Q_n,P_m) = \sup_{\gamma\in\Gamma_{\Sigma}^{\mathrm{inv}}}\{\mathbb{E}_{Q_n}[\gamma] - \mathbb{E}_{P_m}[\gamma]\} \approx \mathrm{MMD}(Q,P)$

Theorem [Chen, Katsoulakis, Rey-Bellet, **Z.**, *ICML* 2023]

Let Σ be a finite group acting on X, and P, Q are Σ-invariant. With high probability,

$$\left|\mathrm{MMD}(Q,P) - \mathrm{MMD}^{\Sigma}(Q_n,P_m)\right| = O\left(A_{\Sigma,k}\left(\frac{1}{\sqrt{m}} + \frac{1}{\sqrt{n}}\right)\right)$$

where $A_{\Sigma,k} = \sqrt{a_{\Sigma,k} + \frac{1 - a_{\Sigma,k}}{|\Sigma|}}$ $\approx \sqrt{0 + \frac{1-0}{|\Sigma|}} = \sqrt{\frac{1}{|\Sigma|}}$, if $a_{\Sigma,k} \approx 0$

Maximum Mean Discrepancy (MMD)

- $\mathrm{MMD}(Q,P) = \sup_{\gamma\in\Gamma}\{\mathbb{E}_Q[\gamma] - \mathbb{E}_P[\gamma]\}$. Γ is the unit ball in some **RKHS** $\mathscr{H}$ with kernel $k(x,y)$.
- **Estimator:** $\mathrm{MMD}^{\Sigma}(Q_n, P_m) = \sup_{\gamma\in\Gamma_{\Sigma}^{\text{inv}}}\{\mathbb{E}_{Q_n}[\gamma] - \mathbb{E}_{P_m}[\gamma]\} \approx \mathrm{MMD}(Q,P)$

Theorem [Chen, Katsoulakis, Rey-Bellet, **Z.**, *ICML* 2023]

Let Σ be a finite group acting on X, and P, Q are Σ-invariant. With high probability,

$$\left|\mathrm{MMD}(Q,P) - \mathrm{MMD}^{\Sigma}(Q_n,P_m)\right| = O\left(A_{\Sigma,k}\left(\frac{1}{\sqrt{m}} + \frac{1}{\sqrt{n}}\right)\right) \approx O\left(\frac{1}{\sqrt{|\Sigma|m}} + \frac{1}{\sqrt{|\Sigma|n}}\right)$$

where $A_{\Sigma,k} = \sqrt{a_{\Sigma,k} + \frac{1-a_{\Sigma,k}}{|\Sigma|}} \approx \sqrt{0+\frac{1-0}{|\Sigma|}} = \sqrt{\frac{1}{|\Sigma|}}$, if $a_{\Sigma,k} \approx 0$

Maximum Mean Discrepancy (MMD)

- $\text{MMD}(Q, P) = \sup_{\gamma \in \Gamma} \{\mathbb{E}_Q[\gamma] - \mathbb{E}_P[\gamma]\}$. Γ is the unit ball in some **RKHS** $\mathcal{H}$ with kernel $k(x, y)$.
- **Estimator:** $\text{MMD}^{\Sigma}(Q_n, P_m) = \sup_{\gamma \in \Gamma_{\Sigma}^{\text{inv}}} \{\mathbb{E}_{Q_n}[\gamma] - \mathbb{E}_{P_m}[\gamma]\} \approx \text{MMD}(Q, P)$

Theorem [Chen, Katsoulakis, Rey-Bellet, **Z.**, *ICML* 2023]

Let Σ be a finite group acting on X, and P, Q are Σ-invariant. With high probability,

$$\left| \text{MMD}(Q, P) - \text{MMD}^{\Sigma}(Q_n, P_m) \right| = O\left(A_{\Sigma,k} \left(\frac{1}{\sqrt{m}} + \frac{1}{\sqrt{n}} \right) \right)$$

where $A_{\Sigma,k} = \sqrt{a_{\Sigma,k} + \frac{1 - a_{\Sigma,k}}{|\Sigma|}}$

if $a_{\Sigma,k} \approx 1$

Maximum Mean Discrepancy (MMD)

- $\text{MMD}(Q,P) = \sup_{\gamma\in\Gamma}\{\mathbb{E}_Q[\gamma] - \mathbb{E}_P[\gamma]\}$. Γ is the unit ball in some **RKHS** $\mathscr{H}$ with kernel $k(x,y)$.
- **Estimator:** $\text{MMD}^{\Sigma}(Q_n, P_m) = \sup_{\gamma\in\Gamma_\Sigma^{\text{inv}}}\{\mathbb{E}_{Q_n}[\gamma] - \mathbb{E}_{P_m}[\gamma]\} \approx \text{MMD}(Q,P)$

Theorem [Chen, Katsoulakis, Rey-Bellet, **Z.**, *ICML* 2023]

Let Σ be a finite group acting on X, and P, Q are Σ-invariant. With high probability,

$$\left|\text{MMD}(Q,P) - \text{MMD}^{\Sigma}(Q_n,P_m)\right| = O\left(A_{\Sigma,k}\left(\frac{1}{\sqrt{m}} + \frac{1}{\sqrt{n}}\right)\right)$$

where $A_{\Sigma,k} = \sqrt{a_{\Sigma,k} + \frac{1 - a_{\Sigma,k}}{|\Sigma|}}$ $\approx \sqrt{1 + \frac{1-1}{|\Sigma|}} = 1,$ if $a_{\Sigma,k} \approx 1$

Maximum Mean Discrepancy (MMD)

- $\mathrm{MMD}(Q,P) = \sup_{\gamma\in\Gamma}\{\mathbb{E}_Q[\gamma] - \mathbb{E}_P[\gamma]\}$. Γ is the unit ball in some **RKHS** $\mathscr{H}$ with kernel $k(x,y)$.
- **Estimator:** $\mathrm{MMD}^{\Sigma}(Q_n,P_m) = \sup_{\gamma\in\Gamma_{\Sigma}^{\text{inv}}}\{\mathbb{E}_{Q_n}[\gamma] - \mathbb{E}_{P_m}[\gamma]\} \approx \mathrm{MMD}(Q,P)$

Theorem [Chen, Katsoulakis, Rey-Bellet, **Z.**, *ICML* 2023]

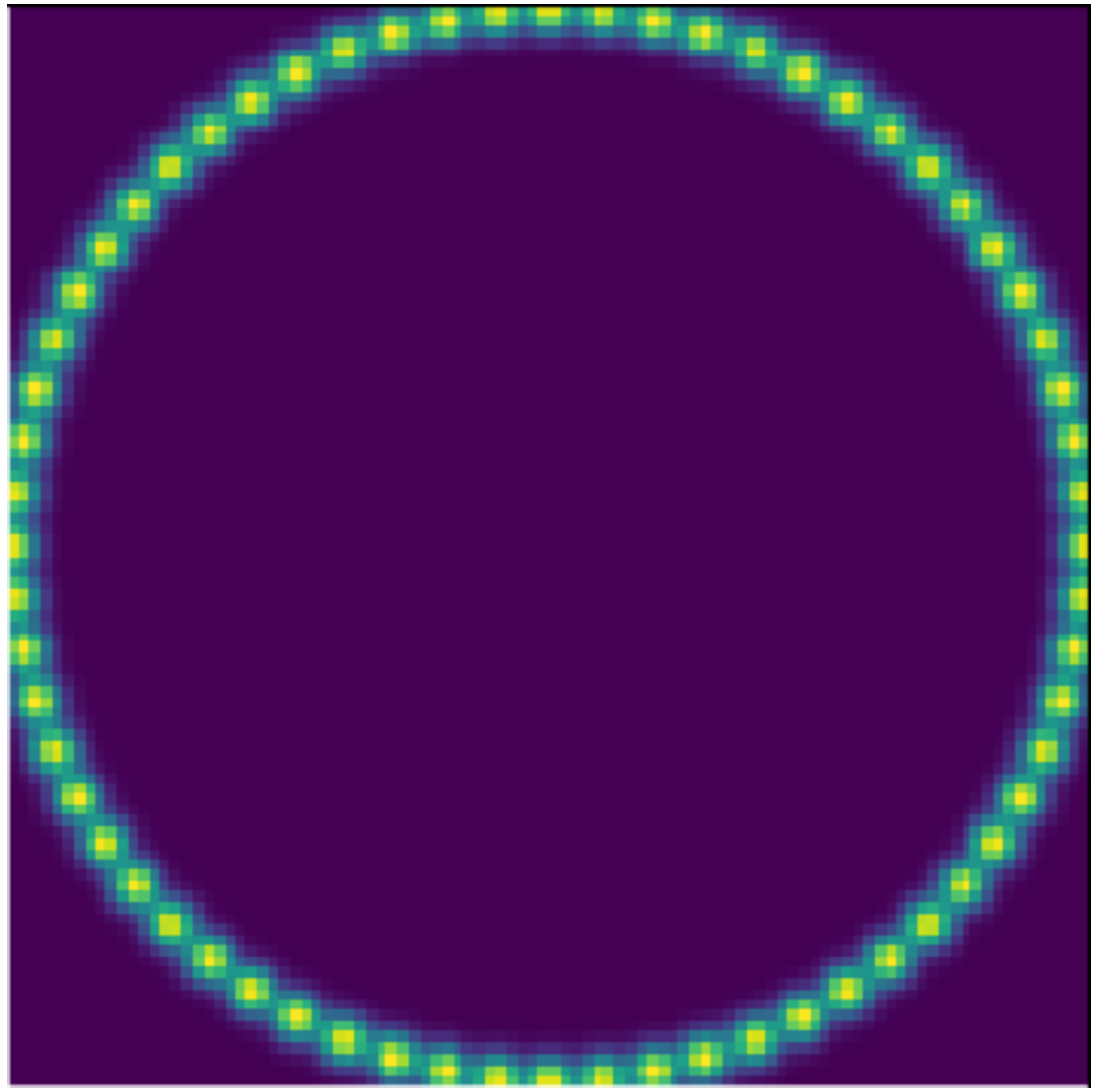
Let Σ be a finite group acting on X, and P, Q are Σ-invariant. With high probability,

$$\left|\mathrm{MMD}(Q,P) - \mathrm{MMD}^{\Sigma}(Q_n,P_m)\right| = O\left(A_{\Sigma,k}\left(\frac{1}{\sqrt{m}} + \frac{1}{\sqrt{n}}\right)\right) \approx O\left(\frac{1}{\sqrt{m}} + \frac{1}{\sqrt{n}}\right)$$

where $A_{\Sigma,k} = \sqrt{a_{\Sigma,k} + \frac{1-a_{\Sigma,k}}{|\Sigma|}}$ $\approx \sqrt{1+\frac{1-1}{|\Sigma|}} = 1,$ if $a_{\Sigma,k} \approx 1$

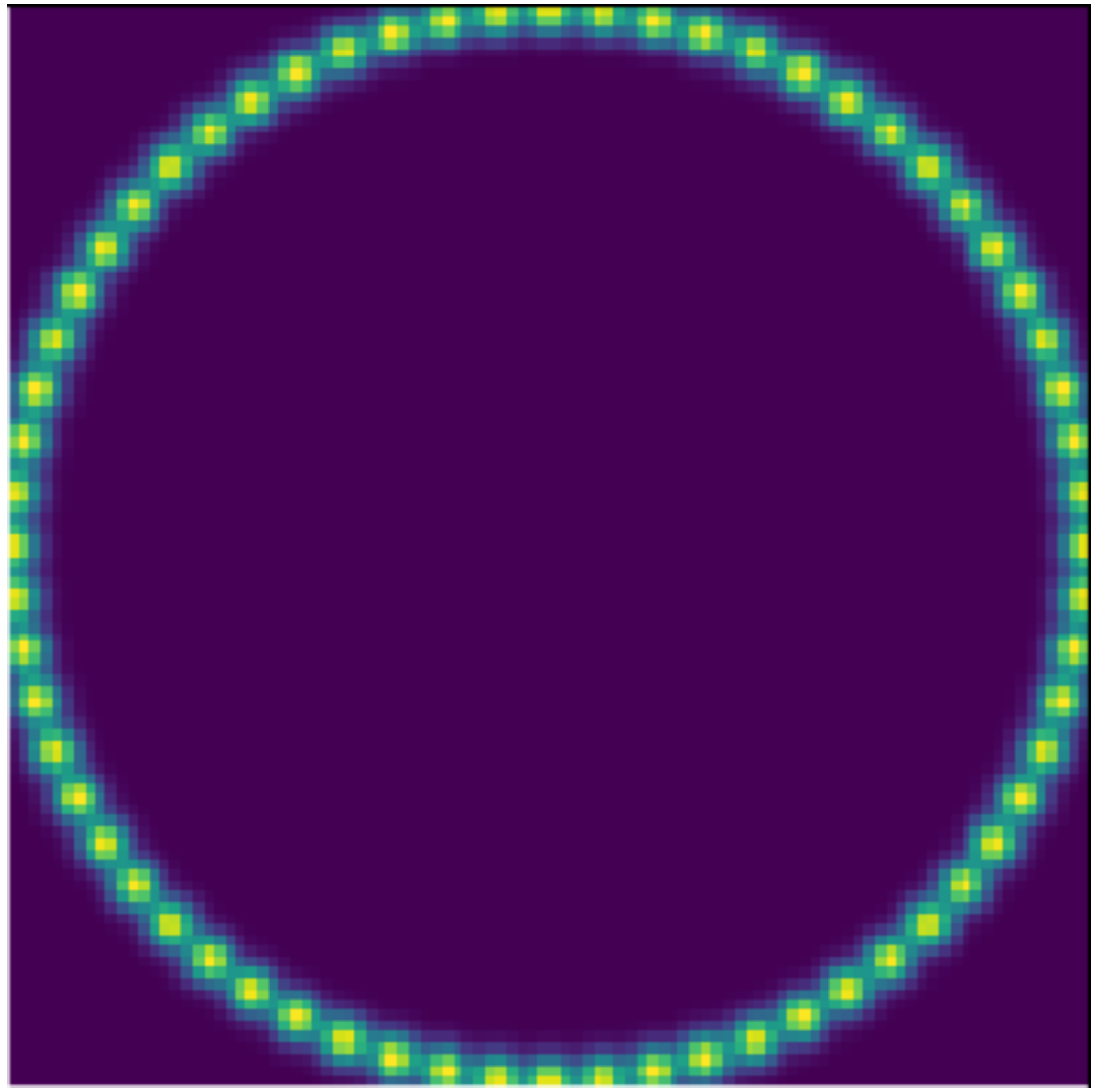
Maximum Mean Discrepancy (MMD)

$$\left| \mathrm{MMD}(Q, P) - \mathrm{MMD}^{\Sigma}(Q_n, P_m) \right| = O\left(A_{\Sigma,k} \left(\frac{1}{\sqrt{m}} + \frac{1}{\sqrt{n}} \right) \right), \quad A_{\Sigma,k} = \sqrt{a_{\Sigma,k} + \frac{1 - a_{\Sigma,k}}{|\Sigma|}}$$



Maximum Mean Discrepancy (MMD)

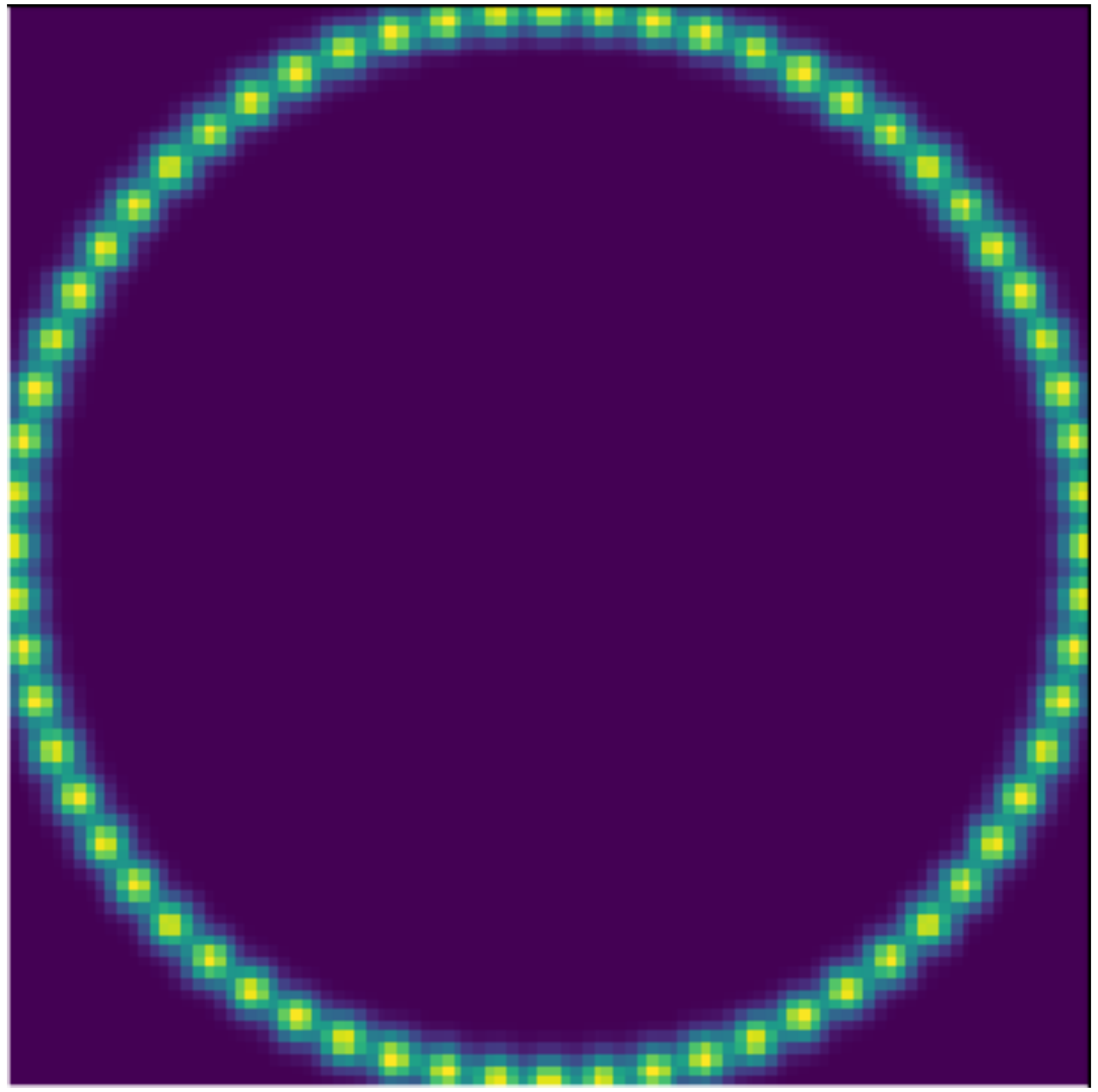
$$\left|\mathrm{MMD}(Q,P) - \mathrm{MMD}^{\Sigma}(Q_n,P_m)\right| = O\left(A_{\Sigma,k}\left(\frac{1}{\sqrt{m}}+\frac{1}{\sqrt{n}}\right)\right), \quad A_{\Sigma,k} = \sqrt{a_{\Sigma,k} + \frac{1-a_{\Sigma,k}}{|\Sigma|}}$$



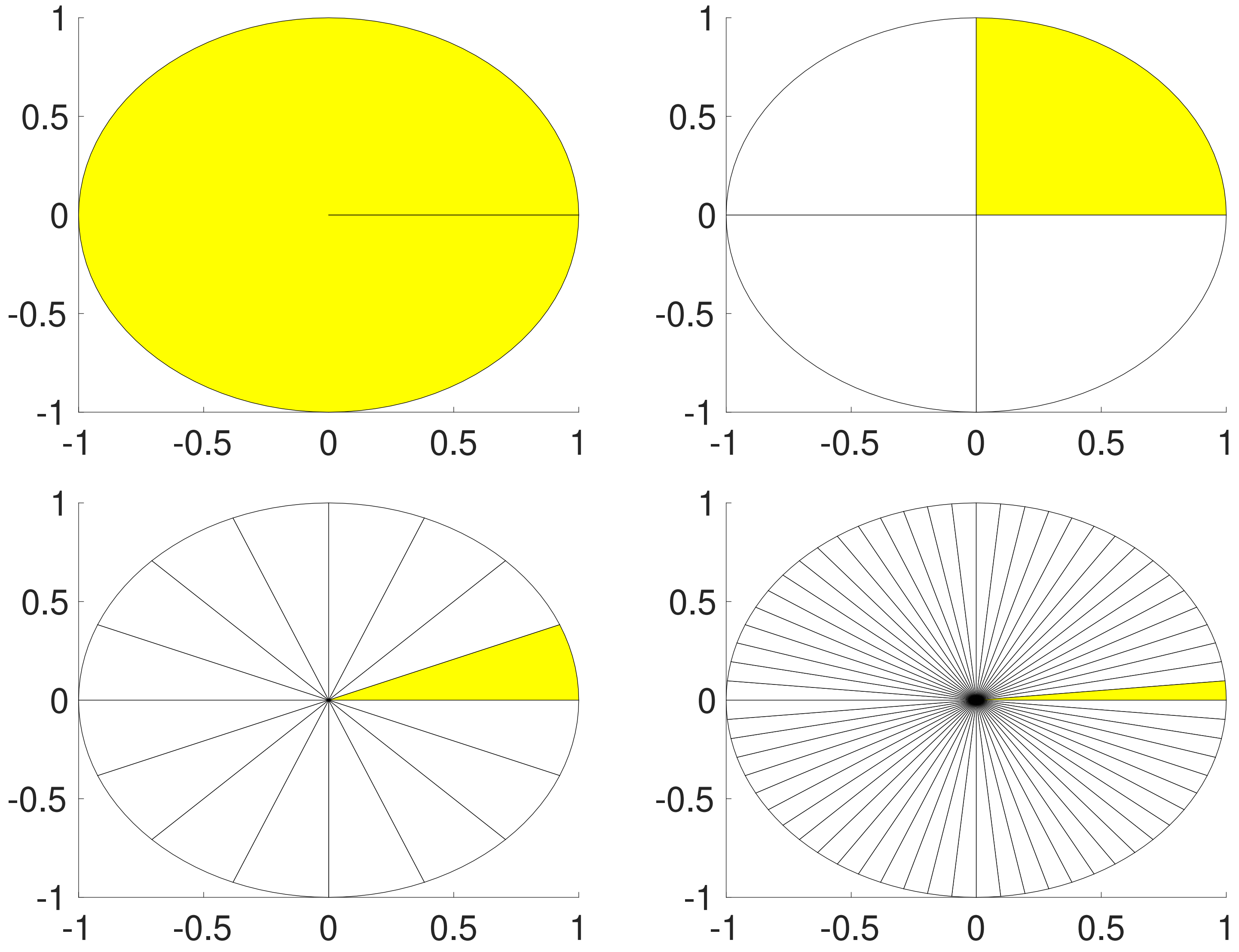
- $P = Q$ is a mixture of 64 Gaussians.

Maximum Mean Discrepancy (MMD)

$$\left|\mathrm{MMD}(Q,P) - \mathrm{MMD}^{\Sigma}(Q_n,P_m)\right| = O\left(A_{\Sigma,k}\left(\frac{1}{\sqrt{m}}+\frac{1}{\sqrt{n}}\right)\right), \qquad A_{\Sigma,k} = \sqrt{a_{\Sigma,k} + \frac{1-a_{\Sigma,k}}{|\Sigma|}}$$

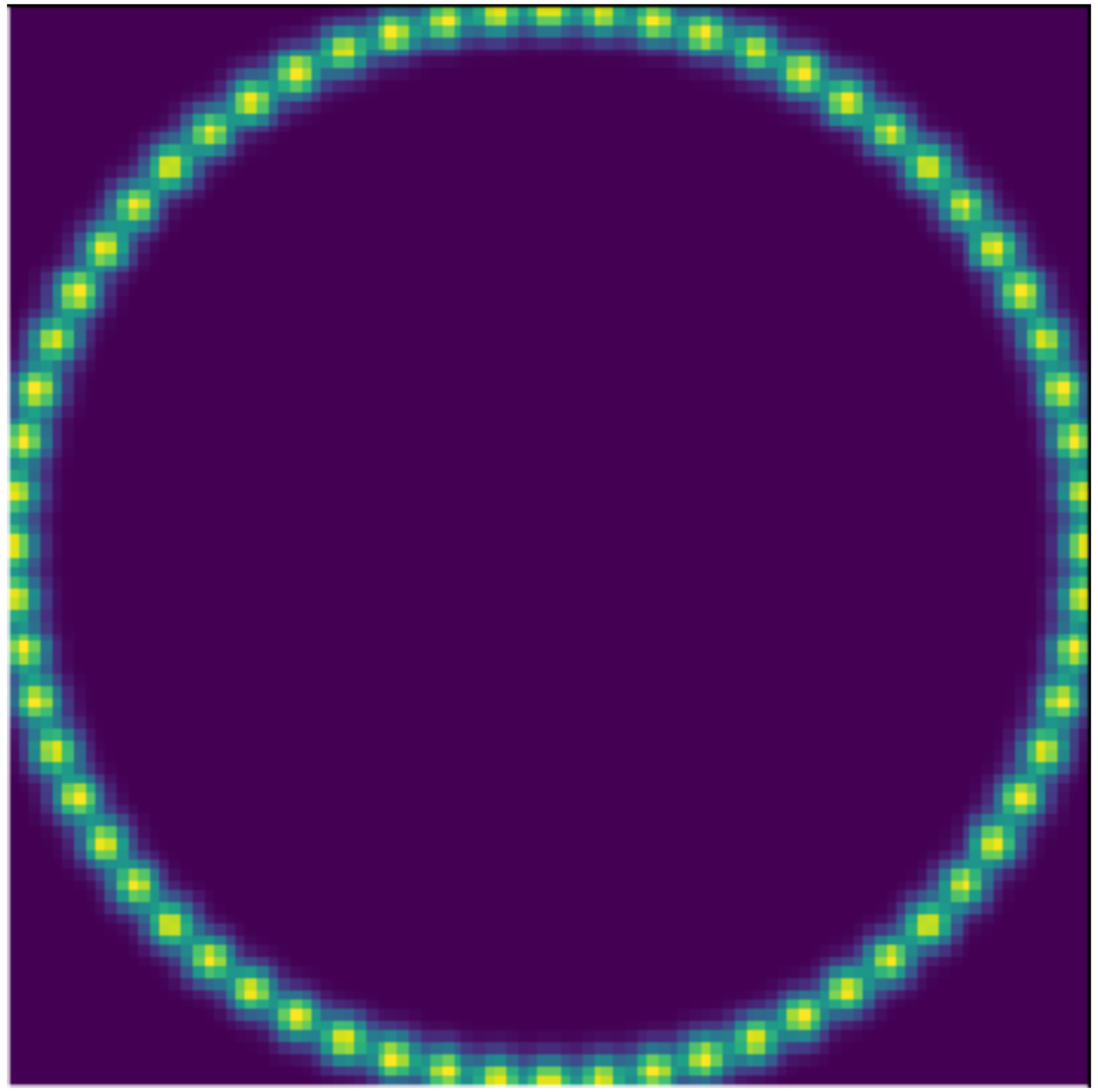


- $P = Q$ is a mixture of 64 Gaussians.
- Then P, Q are **Σ-invariant**, for $\Sigma = C_1, C_4, C_{16}, C_{64}$.



Maximum Mean Discrepancy (MMD)

$$\left|\mathrm{MMD}(Q,P) - \mathrm{MMD}^{\Sigma}(Q_n,P_m)\right| = O\left(A_{\Sigma,k}\left(\frac{1}{\sqrt{m}} + \frac{1}{\sqrt{n}}\right)\right), \qquad A_{\Sigma,k} = \sqrt{a_{\Sigma,k} + \frac{1 - a_{\Sigma,k}}{|\Sigma|}}$$

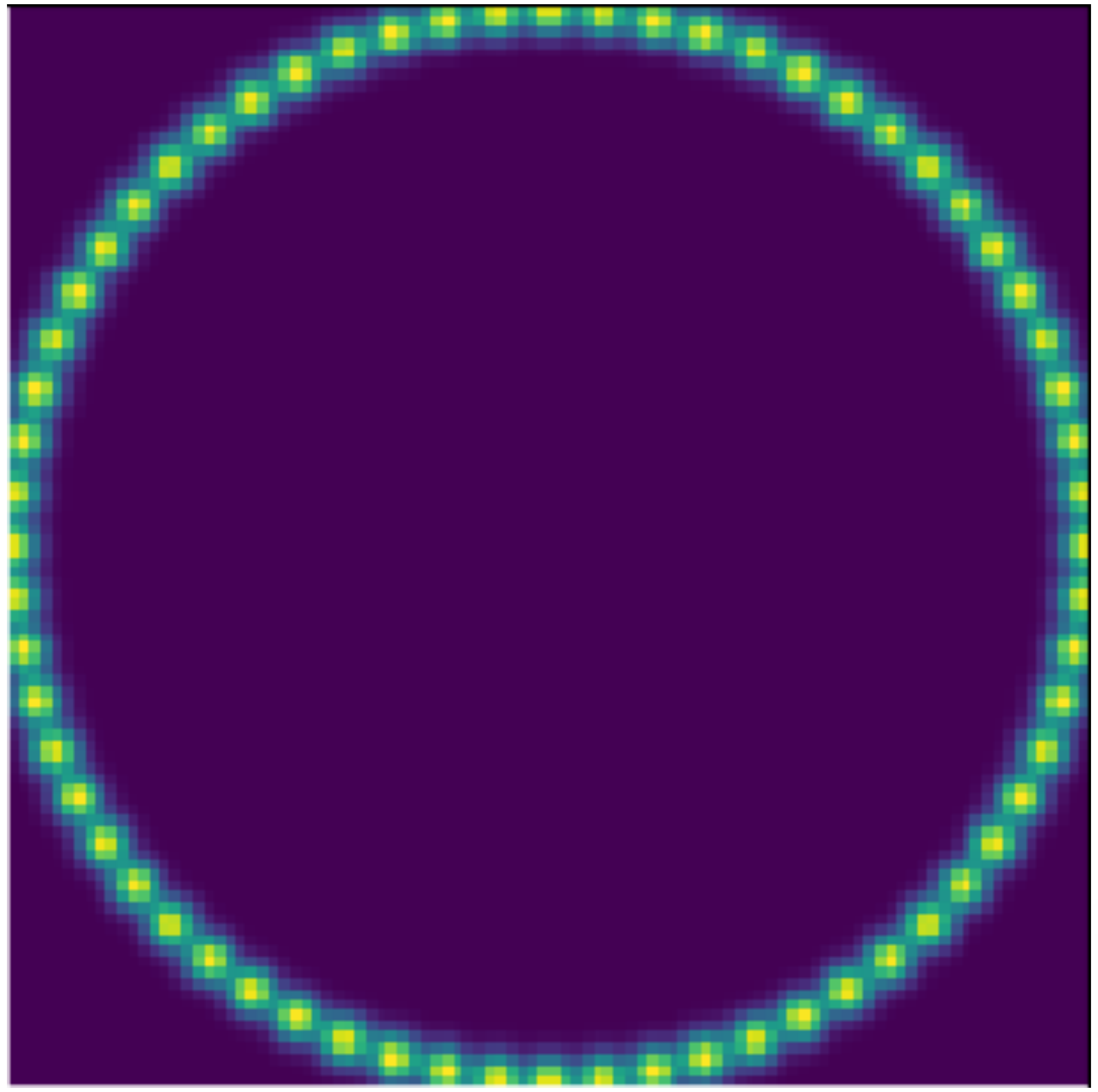


- $P = Q$ is a mixture of 64 Gaussians.
- Then P, Q are Σ**-invariant**, for $\Sigma = C_1, C_4, C_{16}, C_{64}$.
- Consider the Gaussian kernel with bandwidth s

$$k_s(x,y) = \exp\left(-\frac{\|x-y\|^2}{2s^2}\right)$$

Maximum Mean Discrepancy (MMD)

$$\left|\mathrm{MMD}(Q,P) - \mathrm{MMD}^{\Sigma}(Q_n,P_m)\right| = O\left(A_{\Sigma,k}\left(\frac{1}{\sqrt{m}}+\frac{1}{\sqrt{n}}\right)\right), \quad A_{\Sigma,k} = \sqrt{a_{\Sigma,k} + \frac{1-a_{\Sigma,k}}{|\Sigma|}}$$



- $P = Q$ is a mixture of 64 Gaussians.
- Then P, Q are **Σ-invariant**, for $\Sigma = C_1, C_4, C_{16}, C_{64}$.
- Consider the Gaussian kernel with bandwidth s

$$k_s(x,y) = \exp\left(-\frac{\|x-y\|^2}{2s^2}\right)$$

- **Question:** is the estimation more accurate as $|\Sigma| \nearrow$

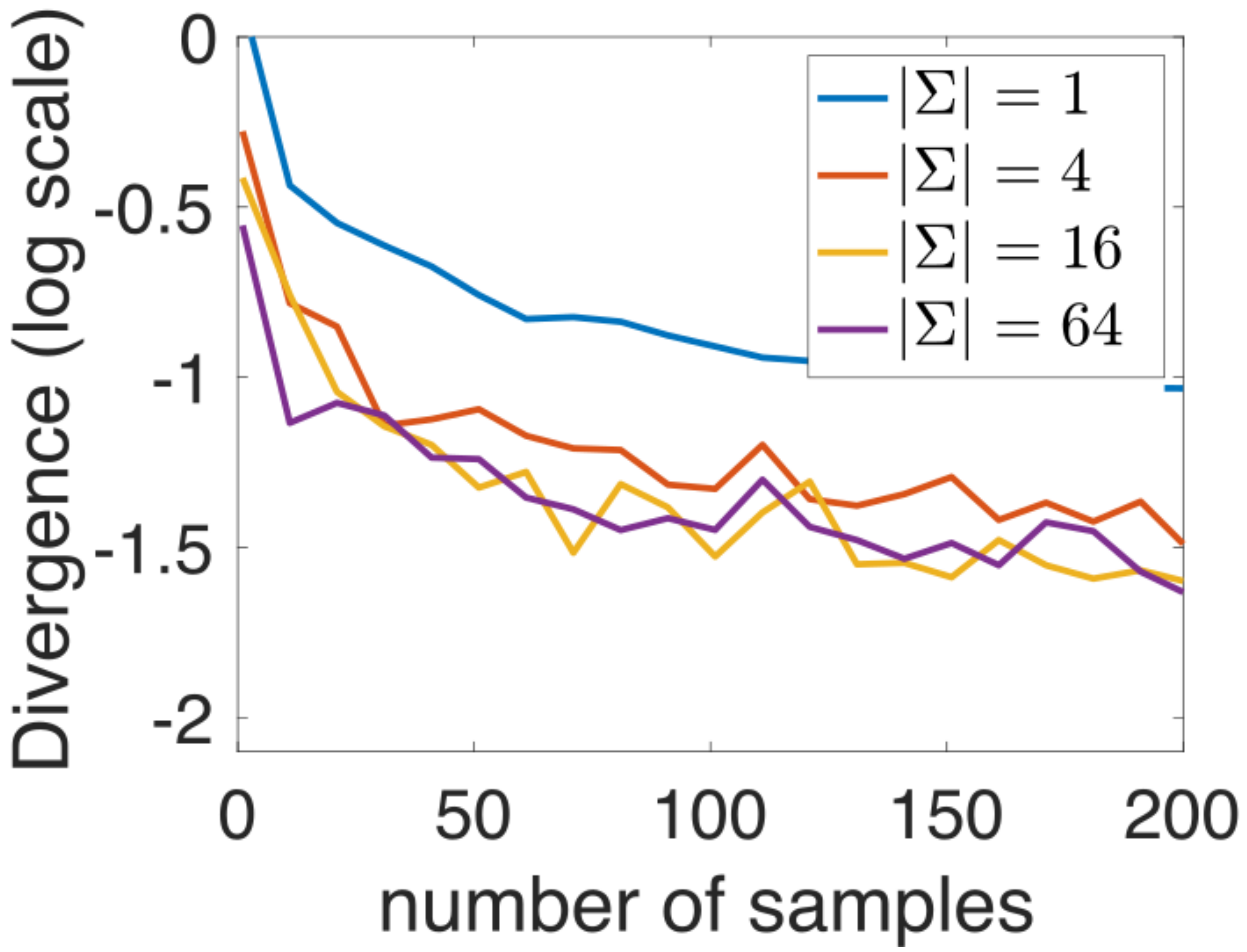
Maximum Mean Discrepancy (MMD)

$$\left|\mathrm{MMD}(Q,P) - \mathrm{MMD}^{\Sigma}(Q_n, P_m)\right| = O\left(A_{\Sigma,k}\left(\frac{1}{\sqrt{m}} + \frac{1}{\sqrt{n}}\right)\right), \quad A_{\Sigma,k} = \sqrt{a_{\Sigma,k} + \frac{1 - a_{\Sigma,k}}{|\Sigma|}}$$

Maximum Mean Discrepancy (MMD)

$$\left|\mathrm{MMD}(Q,P) - \mathrm{MMD}^{\Sigma}(Q_n,P_m)\right| = O\left(A_{\Sigma,k}\left(\frac{1}{\sqrt{m}} + \frac{1}{\sqrt{n}}\right)\right), \quad A_{\Sigma,k} = \sqrt{a_{\Sigma,k} + \frac{1 - a_{\Sigma,k}}{|\Sigma|}}$$

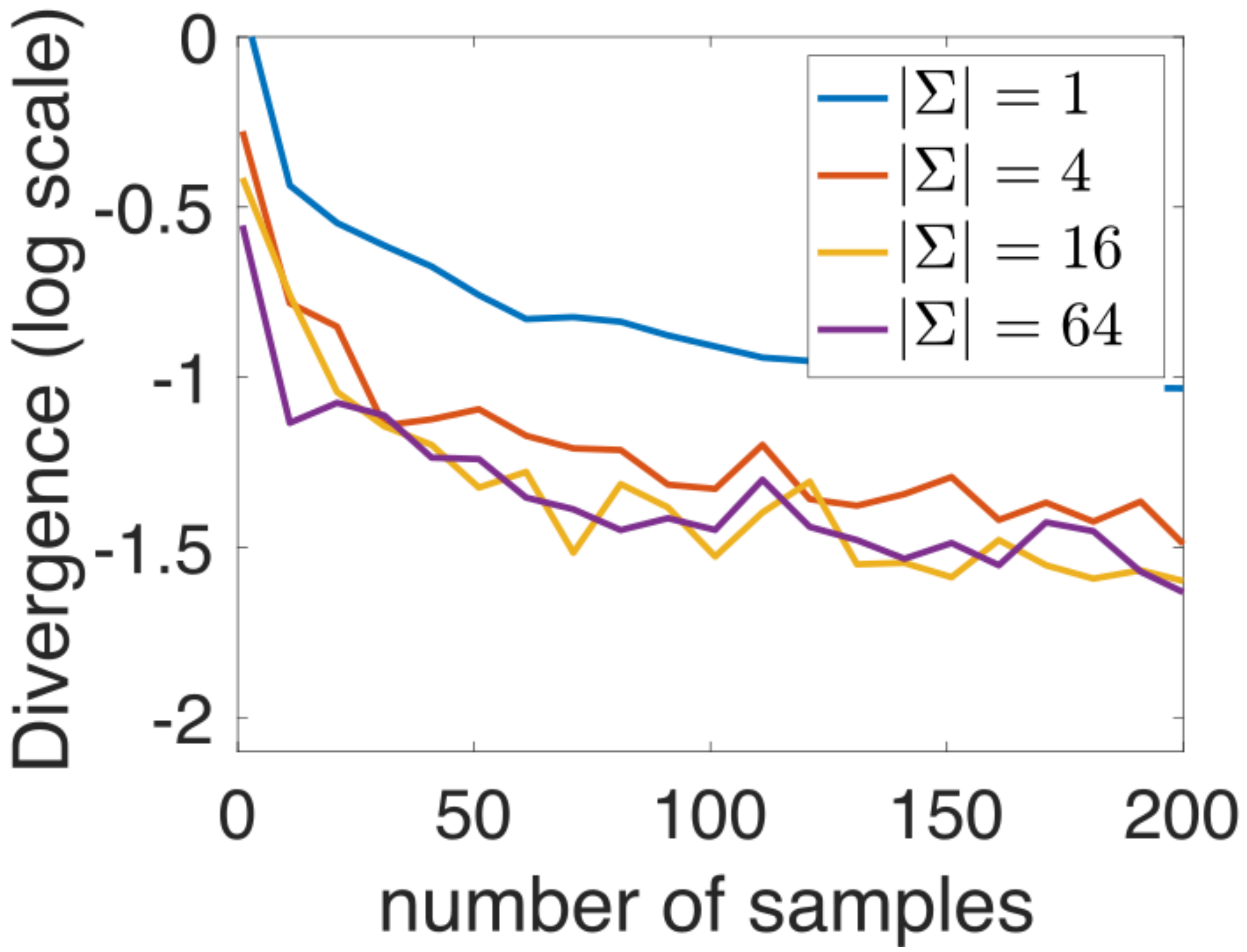
$$s = \frac{2\pi}{24}$$



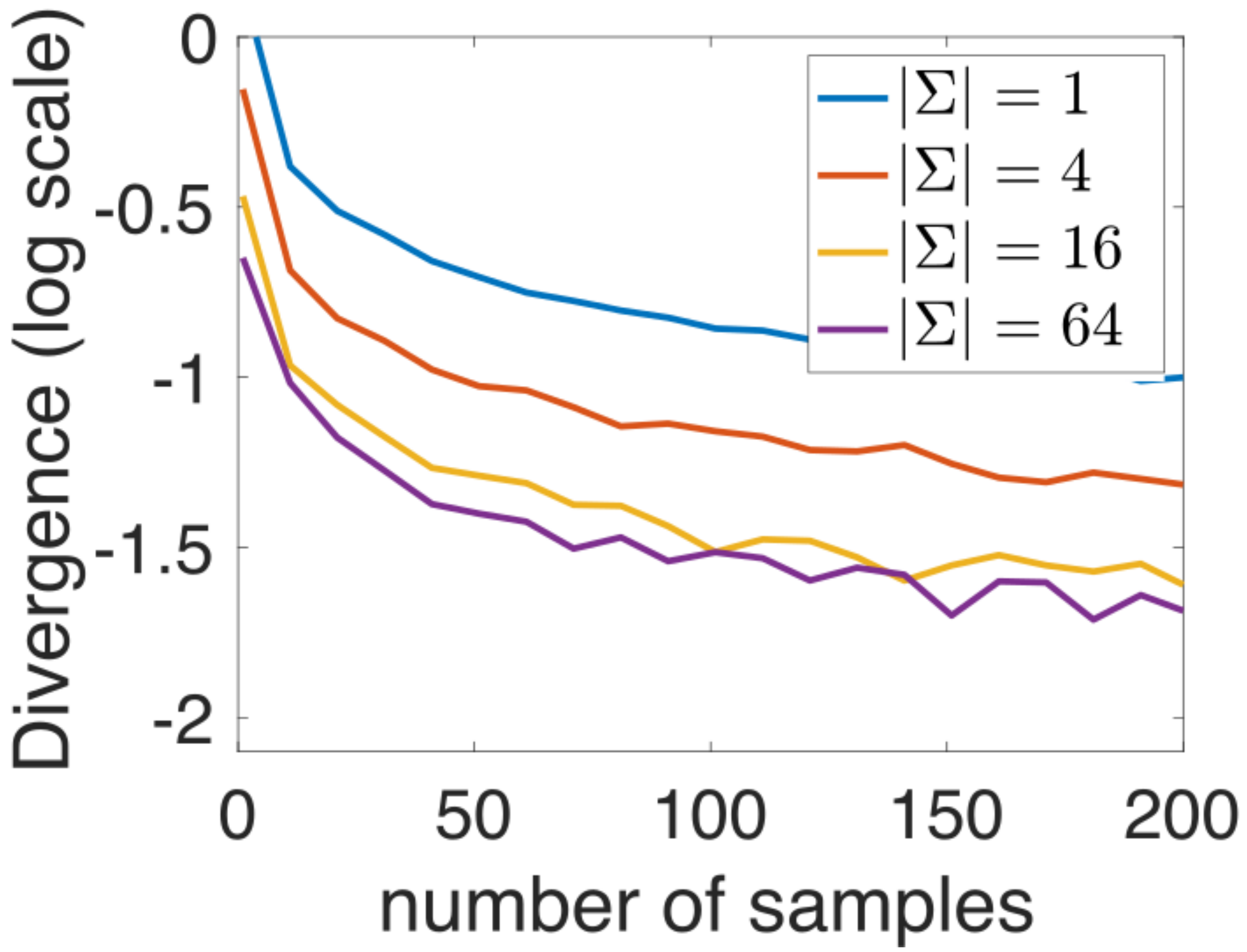
Maximum Mean Discrepancy (MMD)

$$\left|\mathrm{MMD}(Q,P) - \mathrm{MMD}^{\Sigma}(Q_n,P_m)\right| = O\left(A_{\Sigma,k}\left(\frac{1}{\sqrt{m}}+\frac{1}{\sqrt{n}}\right)\right), \quad A_{\Sigma,k} = \sqrt{a_{\Sigma,k} + \frac{1-a_{\Sigma,k}}{|\Sigma|}}$$

$s = \frac{2\pi}{24}$



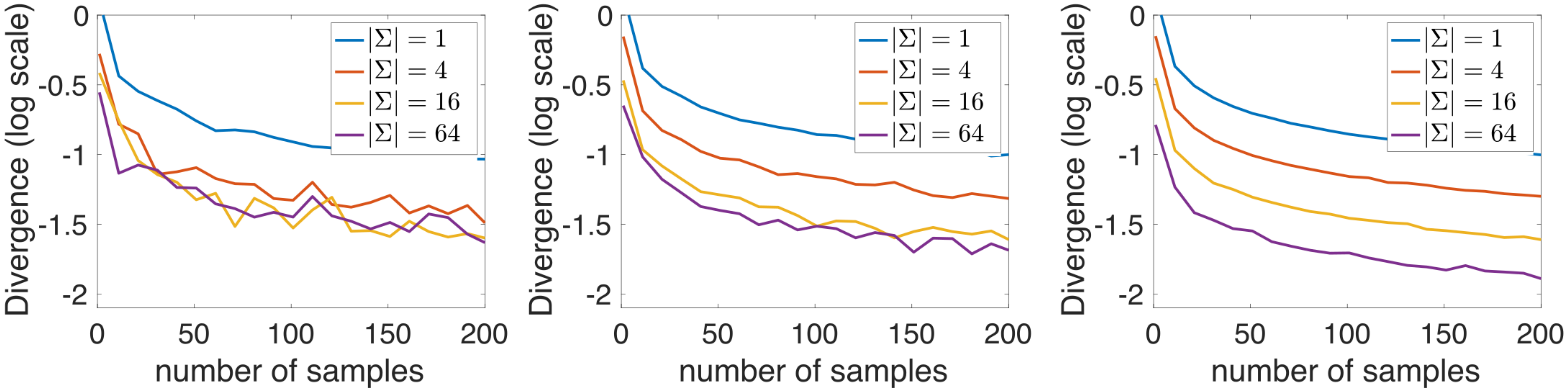
$s = \frac{2\pi}{96}$



Maximum Mean Discrepancy (MMD)

$$\left|\mathrm{MMD}(Q,P) - \mathrm{MMD}^{\Sigma}(Q_n,P_m)\right| = O\left(A_{\Sigma,k}\left(\frac{1}{\sqrt{m}} + \frac{1}{\sqrt{n}}\right)\right), \qquad A_{\Sigma,k} = \sqrt{a_{\Sigma,k} + \frac{1 - a_{\Sigma,k}}{|\Sigma|}}$$

$s = \frac{2\pi}{24}$ $\qquad$ $s = \frac{2\pi}{96}$ $\qquad$ $s = \frac{2\pi}{384}$

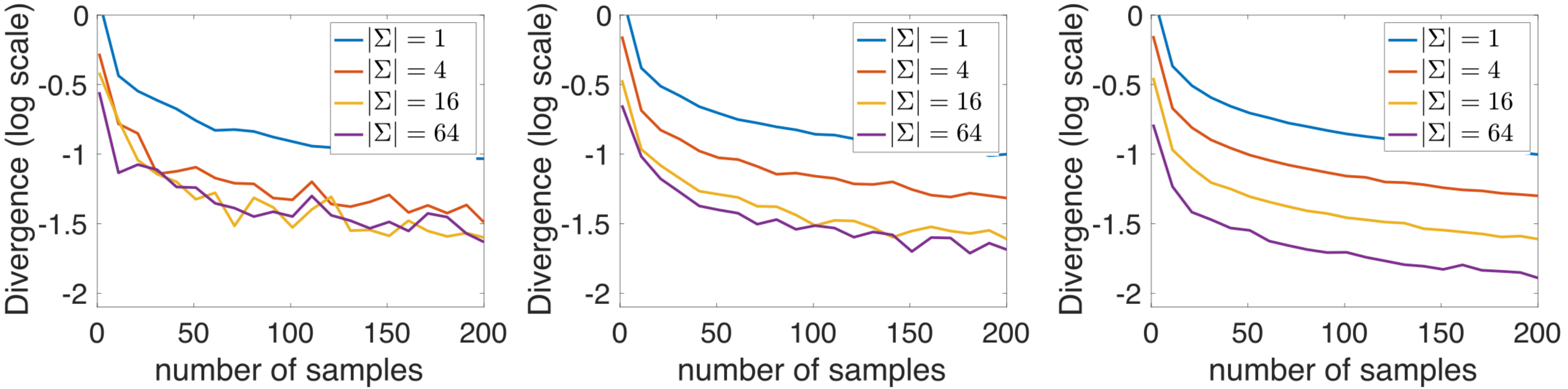


Maximum Mean Discrepancy (MMD)

- Chen, Katsoulakis, Rey-Bellet, **Z.**, "Statistical Guarantees of Group-Invariant GANs", *SIAM J. on Uncertainty Quantification* (2025)

- Chen, **Z.**, "On the implicit bias of linear equivariant steerable networks", *NeurIPS* (2023).

$s = \frac{2\pi}{24}$ $\qquad$ $s = \frac{2\pi}{96}$ $\qquad$ $s = \frac{2\pi}{384}$

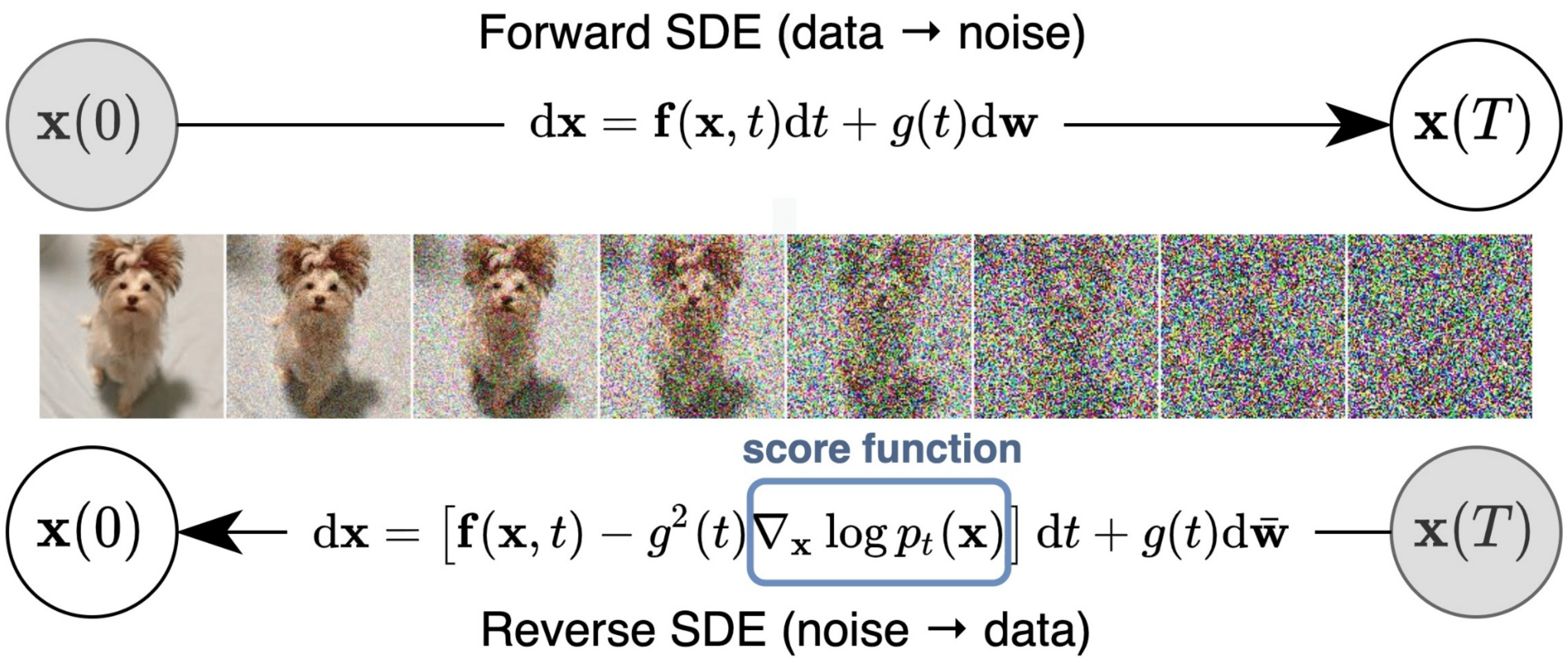


Topics

- How can we impose symmetry in generative models (GANs)?
- How can we quantify the performance gain?
- How can we flexibly impose general structures in diffusion models?

General structures? Diffusion models?

General structures? Diffusion models?



Song et al., *ICLR 2021*

- $\mathbf{x}(0) \sim Q$ (data distribution)
- $\mathbf{x}(T) \sim \pi$ (tractable prior/reference distribution)

General structures? Diffusion models?

Forward SDE (data → noise)

$$\mathbf{x}(0) \xrightarrow{\mathrm{d}\mathbf{x} = \mathbf{f}(\mathbf{x}, t)\mathrm{d}t + g(t)\mathrm{d}\mathbf{w}} \mathbf{x}(T)$$

score function

$$\mathbf{x}(0) \xleftarrow{\mathrm{d}\mathbf{x} = \left[\mathbf{f}(\mathbf{x}, t) - g^2(t)\boxed{\nabla_{\mathbf{x}} \log p_t(\mathbf{x})}\right] \mathrm{d}t + g(t)\mathrm{d}\bar{\mathbf{w}}} \mathbf{x}(T)$$

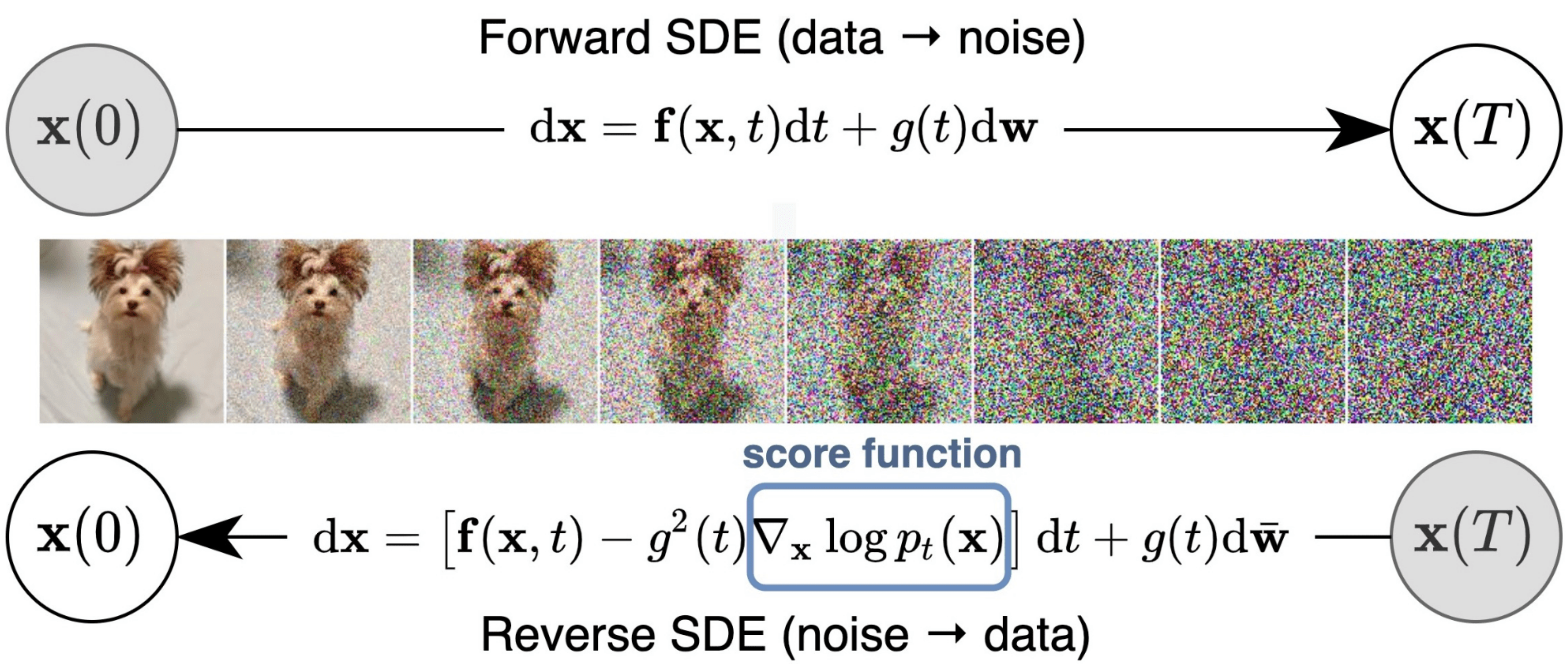
Reverse SDE (noise → data)

Song et al., *ICLR 2021*

- $\mathbf{x}(0) \sim Q$ (data distribution)
- $\mathbf{x}(T) \sim \pi$ (tractable prior/reference distribution)
- Embedded structure in **score function**

$$S(x, t) = \nabla_x \log p(x, t)$$

General structures? Diffusion models?



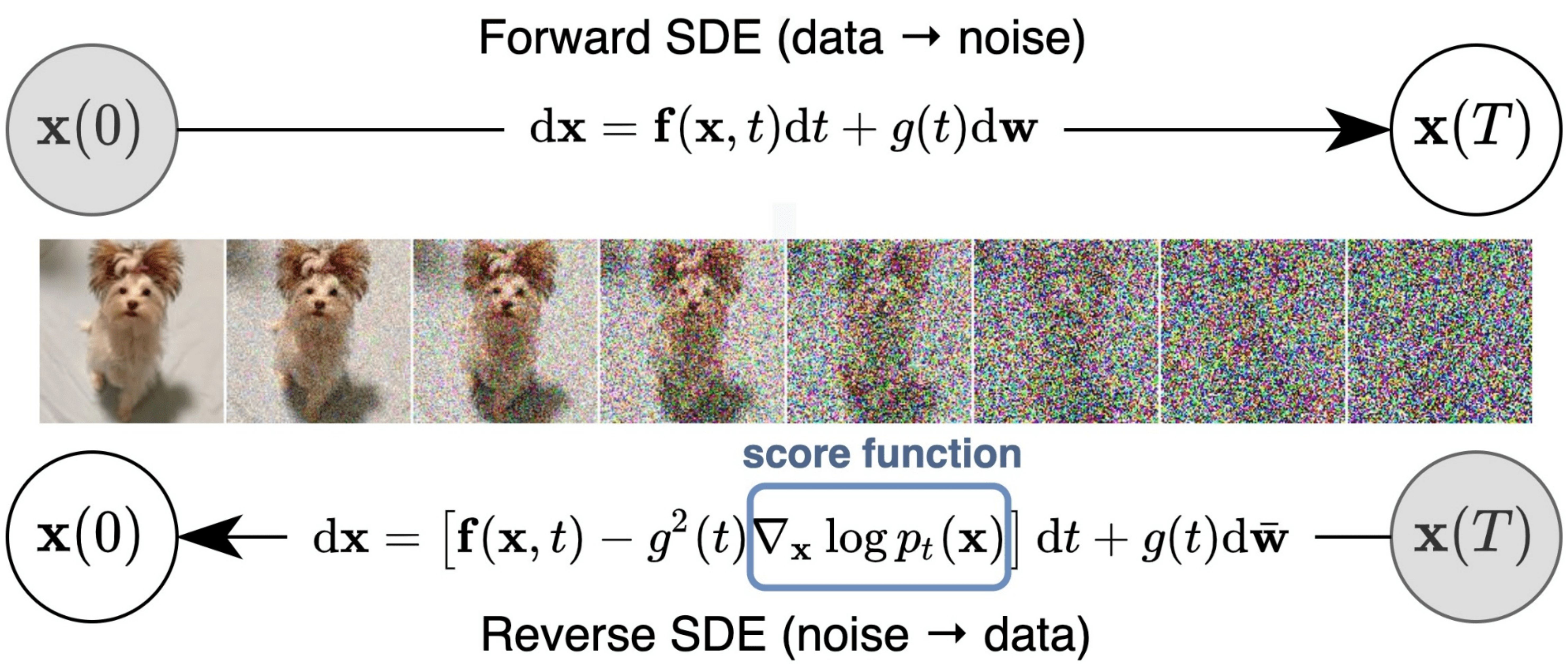
Song et al., *ICLR 2021*

- $\mathbf{x}(0) \sim Q$ (data distribution)
- $\mathbf{x}(T) \sim \pi$ (tractable prior/reference distribution)
- Embedded structure in **score function**

$$S(x, t) = \nabla_x \log p(x, t)$$

- Rigid…

General structures? Diffusion models?



Song et al., *ICLR 2021*

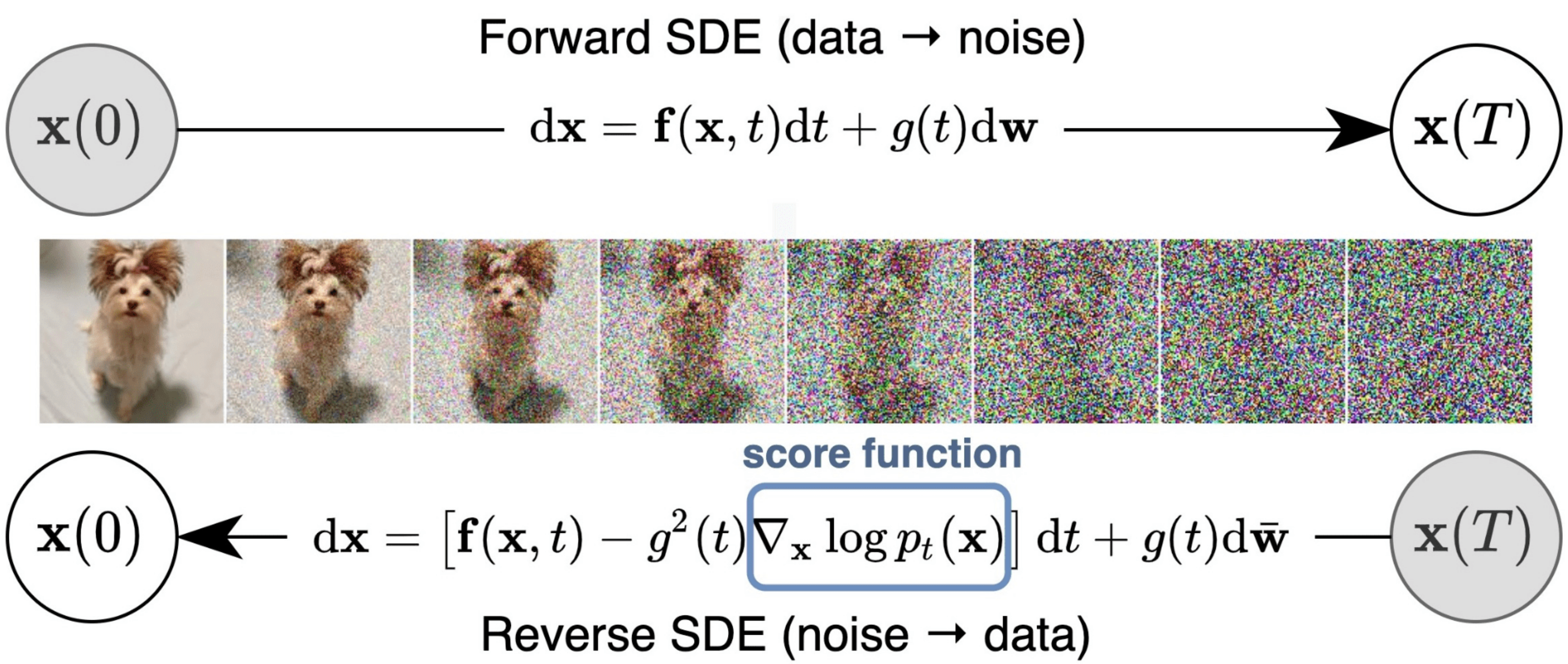
- $\mathbf{x}(0) \sim Q$ (data distribution)
- $\mathbf{x}(T) \sim \pi$ (tractable prior/reference distribution)
- Embedded structure in **score function**

$$S(x, t) = \nabla_x \log p(x, t)$$

- Rigid…

- Embed structure **(flexibly!)** in π
  - Multimodel, approximate group symmetry, low-dimensionality, etc.
  - Tractable **reference distribution** close to Q

General structures? Diffusion models?



Song et al., *ICLR 2021*

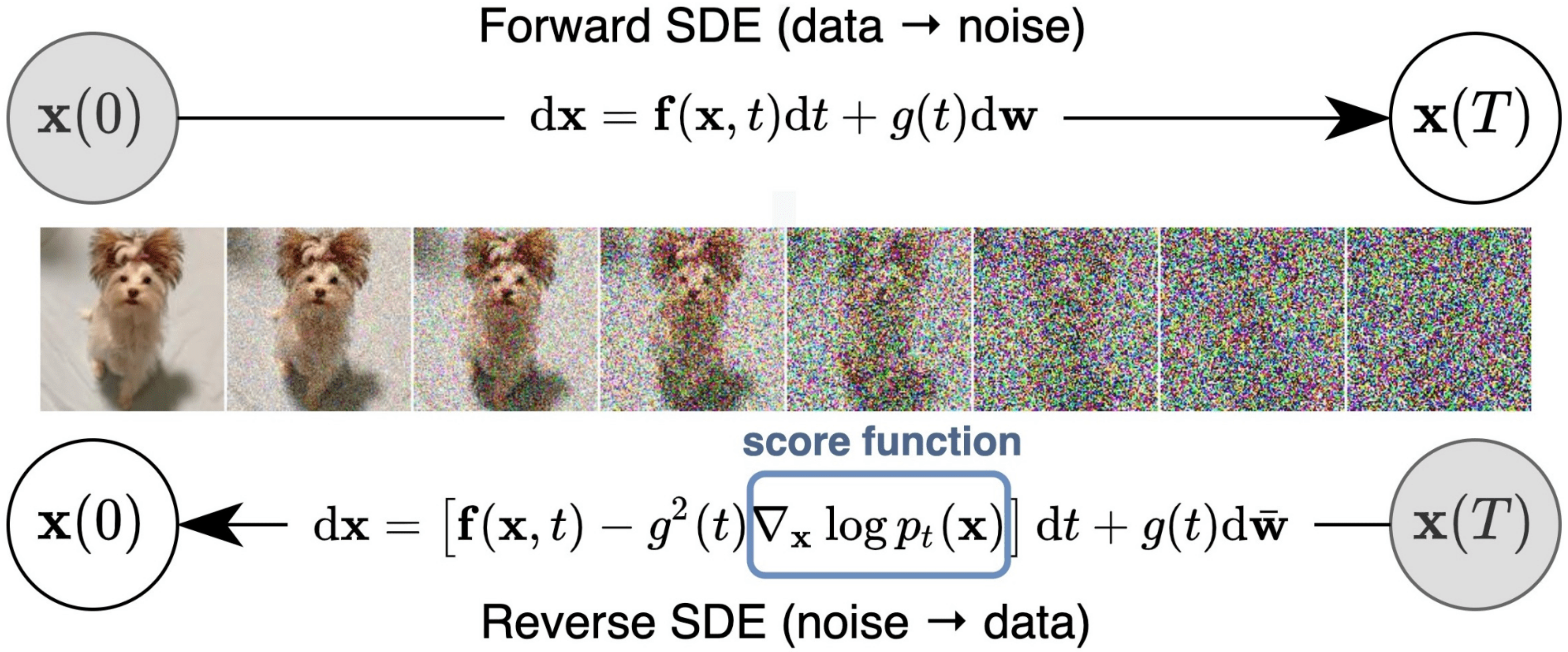
- $\mathbf{x}(0) \sim Q$ (data distribution)
- $\mathbf{x}(T) \sim \pi$ (tractable prior/reference distribution)
- Embedded structure in **score function**

$$S(x, t) = \nabla_x \log p(x, t)$$

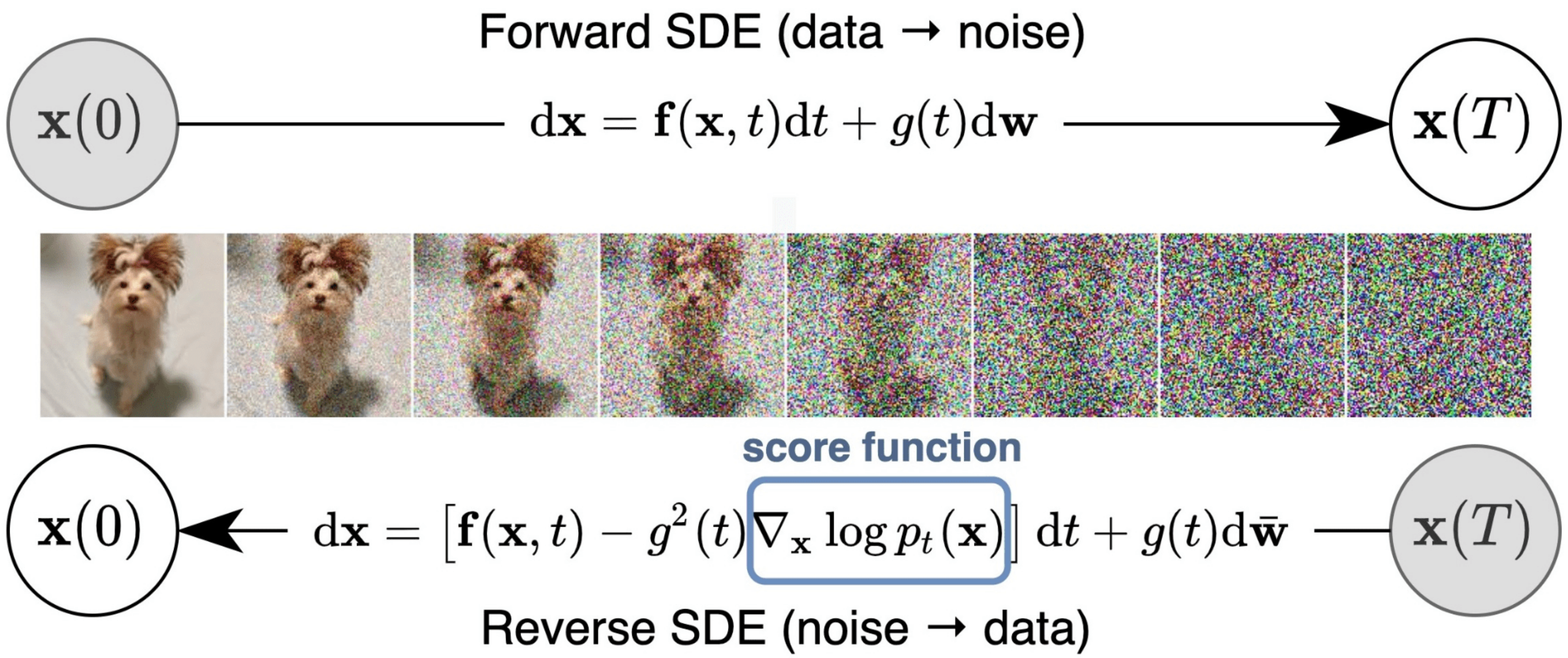
- Rigid…

- Embed structure **(flexibly!)** in π
  - Multimodel, approximate group symmetry, low-dimensionality, etc.
  - Tractable **reference distribution** close to Q
- Requires **nonlinear** drift $f(x, t)$, special training.

Nonlinear denoising score matching (NDSM)



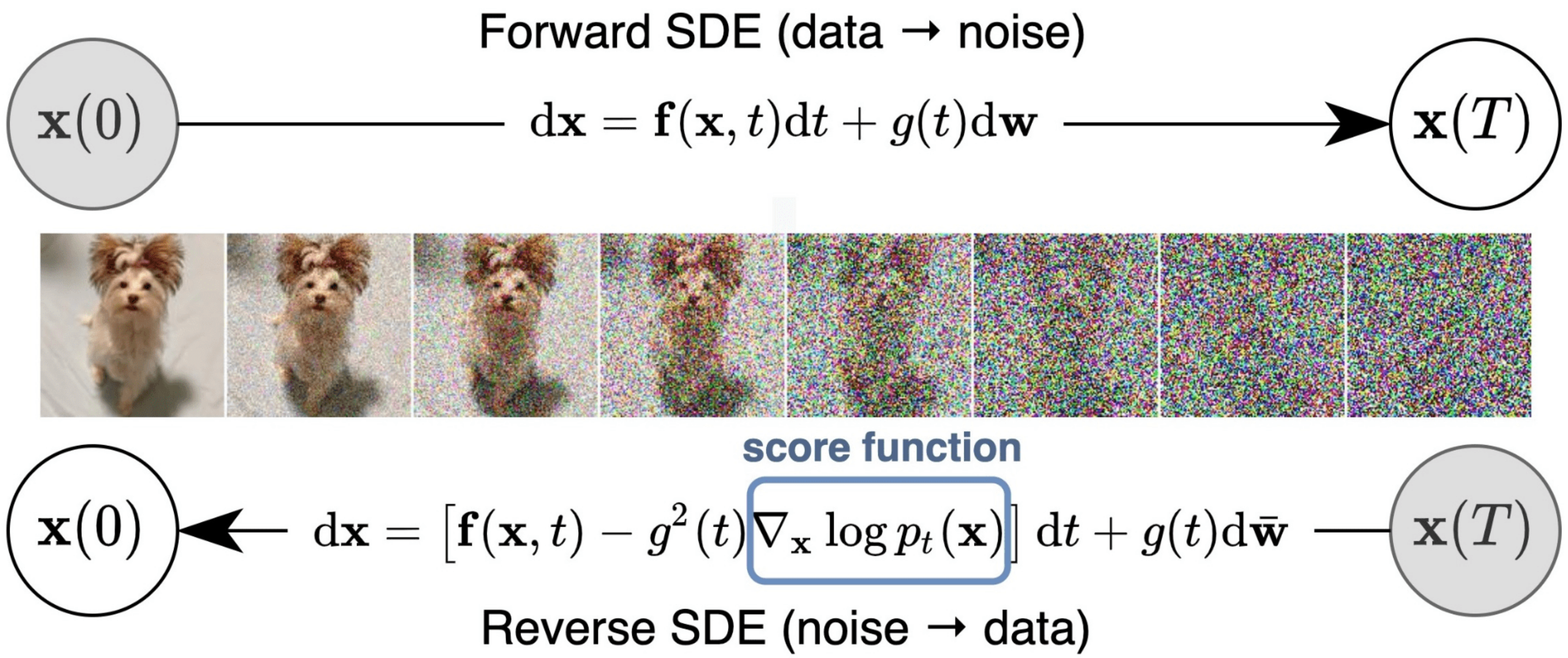
Nonlinear denoising score matching (NDSM)



- Denoising score matching

$$\min_{\theta} \mathbb{E} \left\{ \lambda(t)\mathbb{E}_{x(0)}\mathbb{E}_{x(t)|x(0)} \left[S_\theta(x(t), t) - \nabla_{x(t)} \log p_{t|0} \left(x(t) \,|\, x(0) \right) \right] \right\}$$

Nonlinear denoising score matching (NDSM)

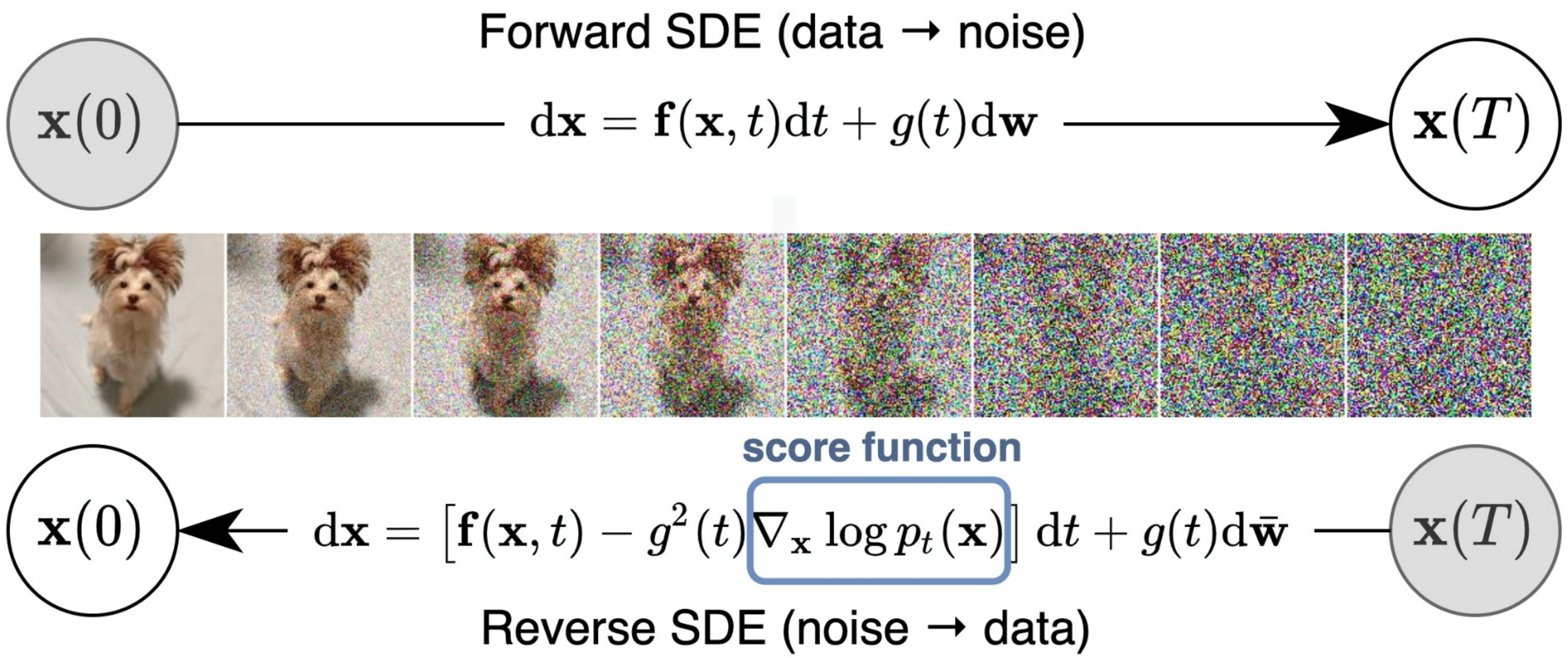


- Denoising score matching

$$\min_{\theta} \mathbb{E}\left\{\lambda(t)\mathbb{E}_{x(0)}\mathbb{E}_{x(t)|x(0)}\left[S_\theta(x(t), t) - \nabla_{x(t)} \log p_{t|0}\left(x(t) \,|\, x(0)\right)\right]\right\}$$

- $p_{t|0}\left(x(t) \,|\, x(0)\right)$ has **closed form formula** for **Ornstein-Uhlenbeck (OU)** processes.

Nonlinear denoising score matching (NDSM)



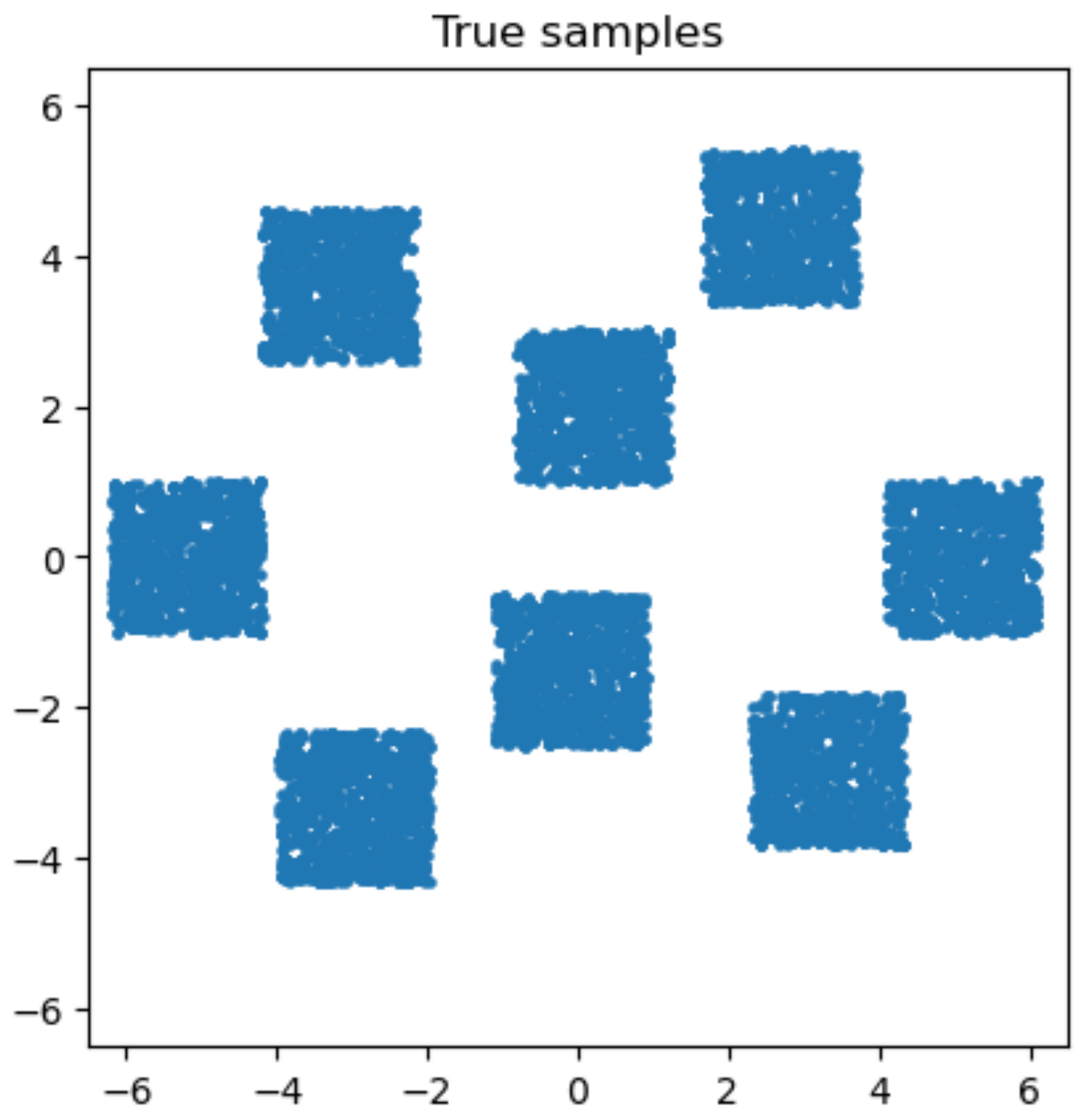
- Denoising score matching

$$\min_{\theta} \mathbb{E}\left\{\lambda(t)\mathbb{E}_{x(0)}\mathbb{E}_{x(t)|x(0)}\left[S_\theta(x(t), t) - \nabla_{x(t)}\log p_{t|0}\left(x(t)\,|\,x(0)\right)\right]\right\}$$

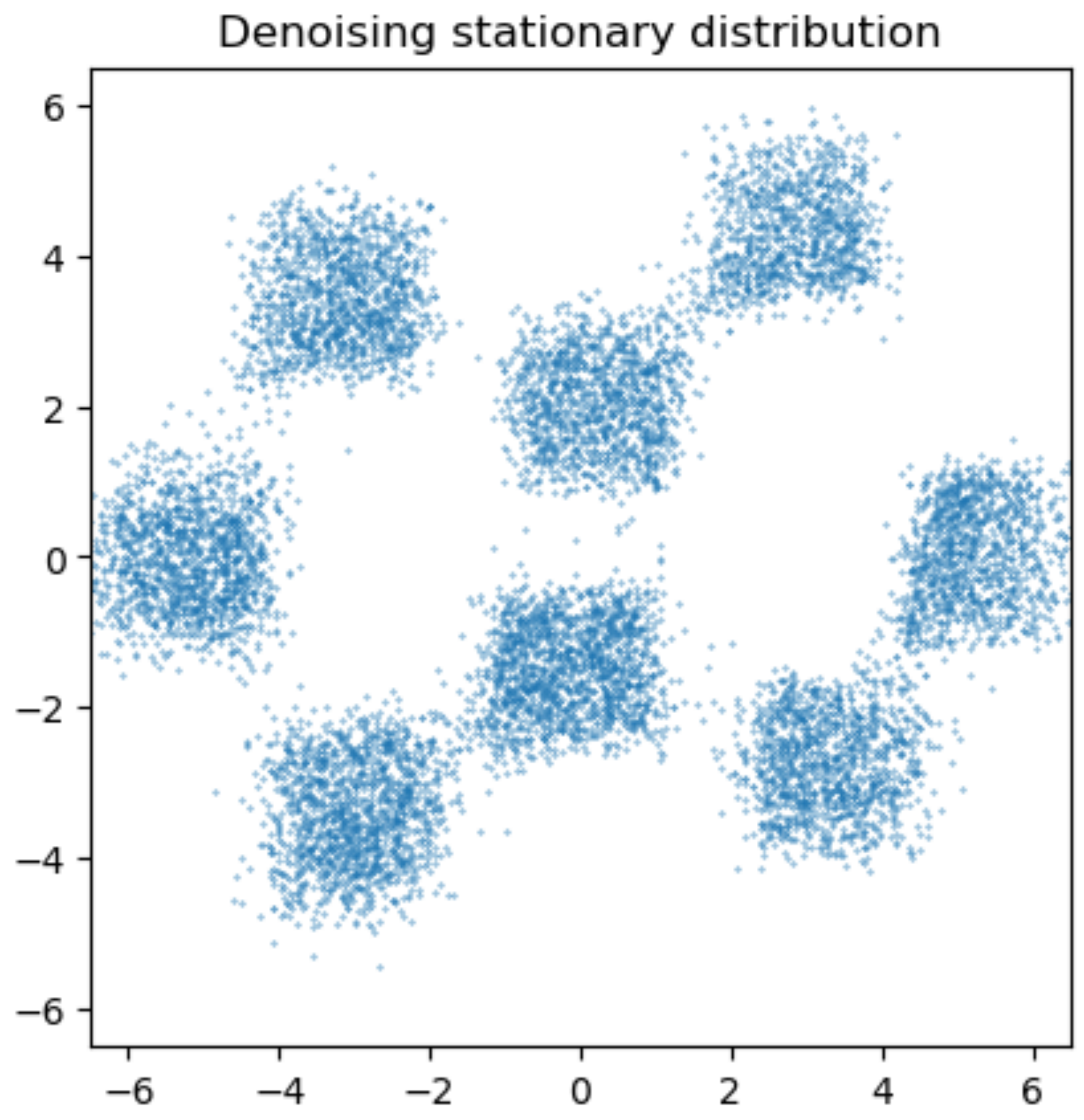
- $p_{t|0}\left(x(t)\,|\,x(0)\right)$ has **closed form formula** for **Ornstein-Uhlenbeck (OU)** processes.
- For nonlinear forward SDE, e.g., Langevin dynamics, $p_{t|0}\left(x(t)\,|\,x(0)\right)$ is not known.

Multimodal distribution

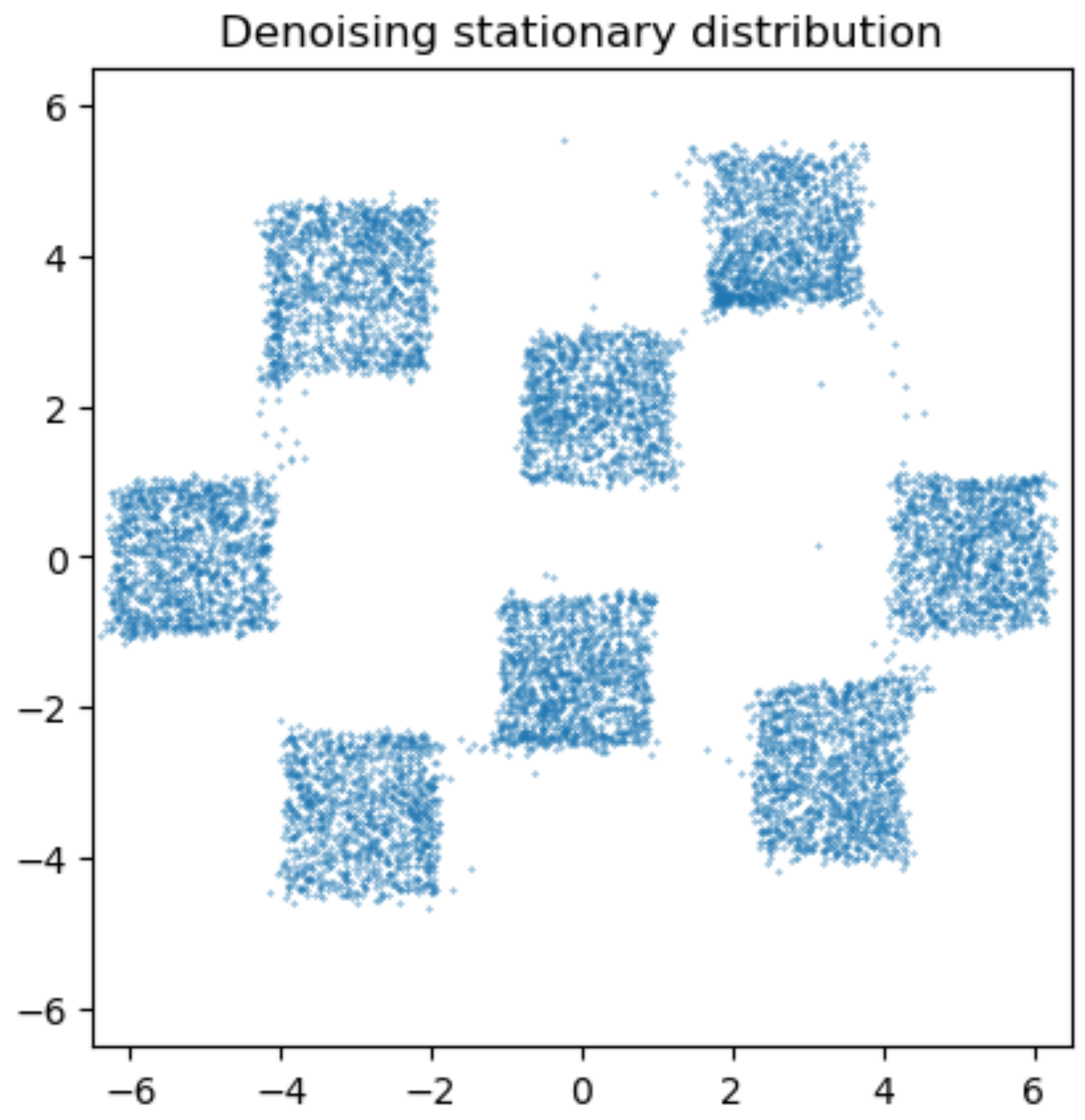
True Samples



OU+DSM



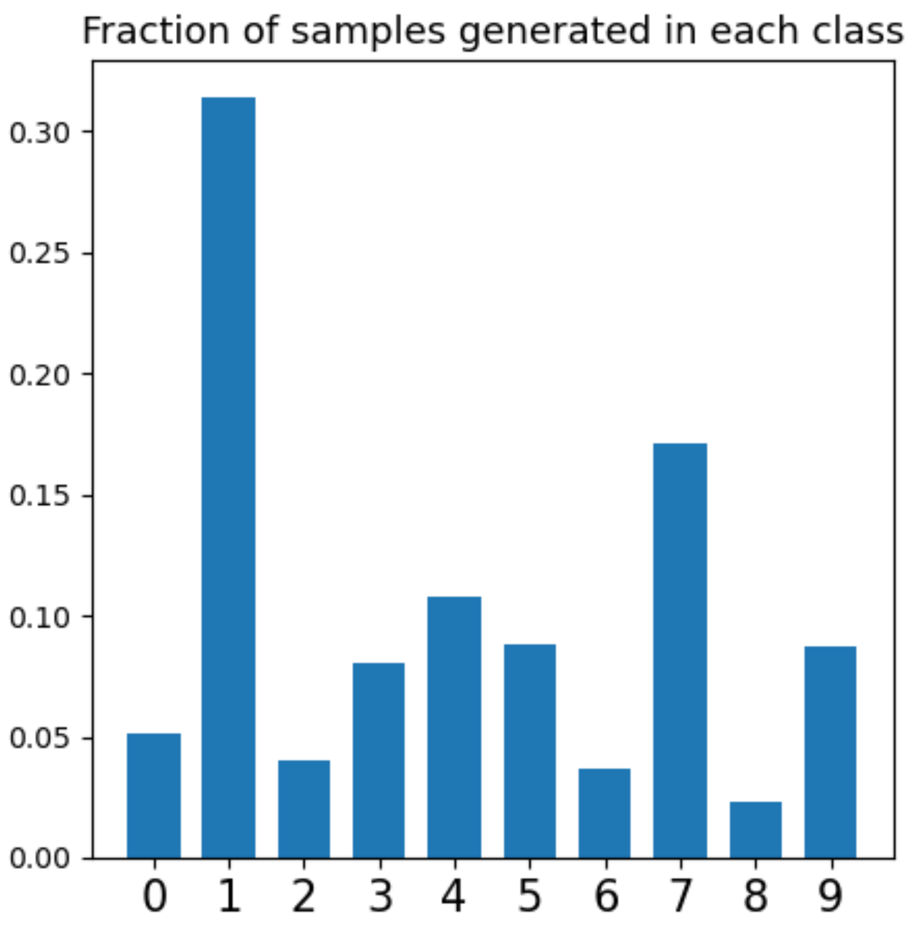
GM+NDSM



Multimodal distribution

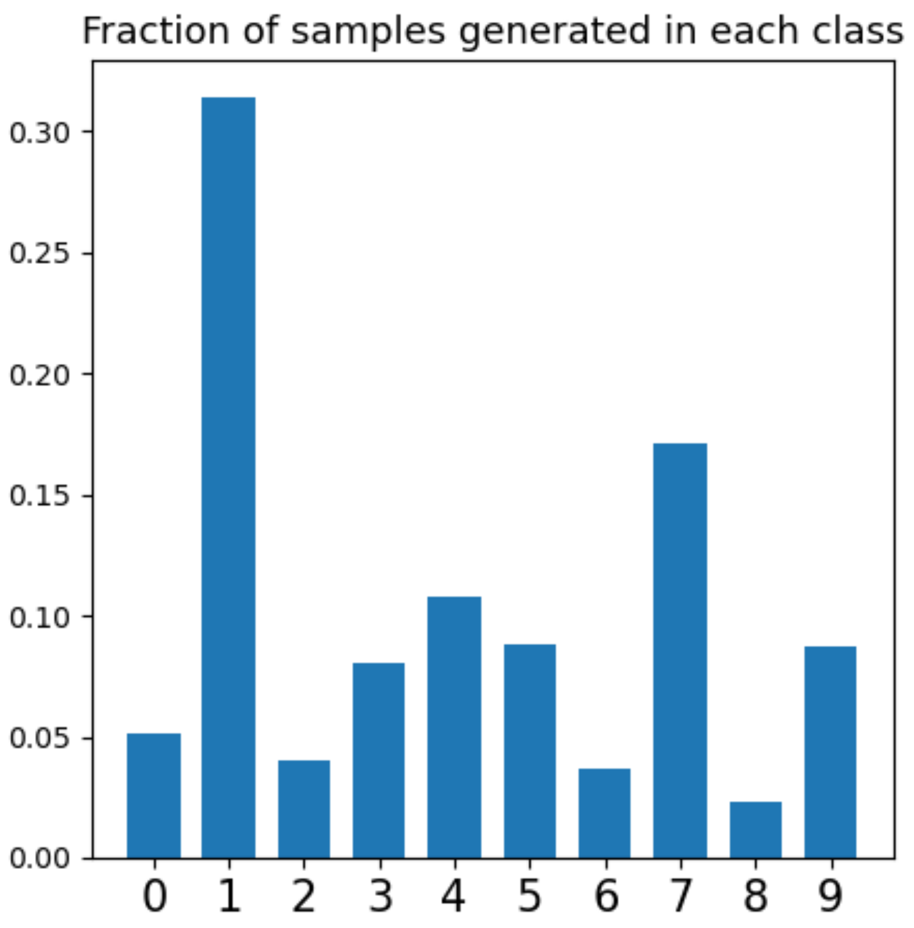
Multimodal distribution

OU+DSM

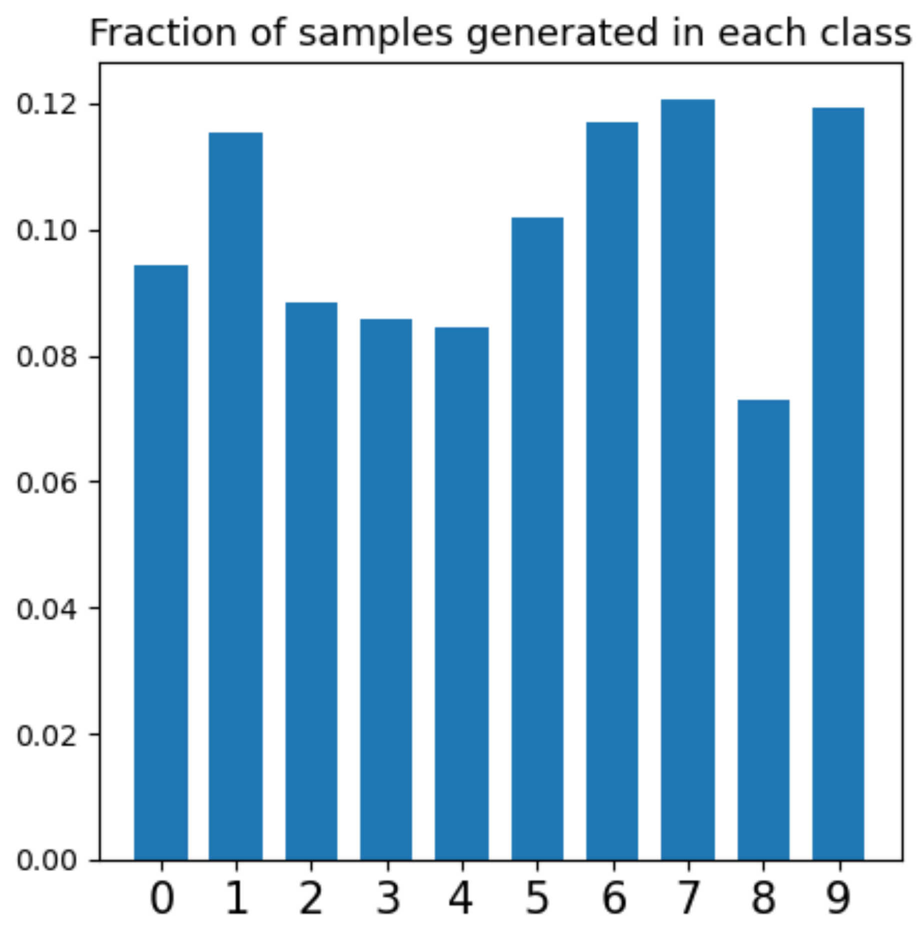


Multimodal distribution

OU+DSM



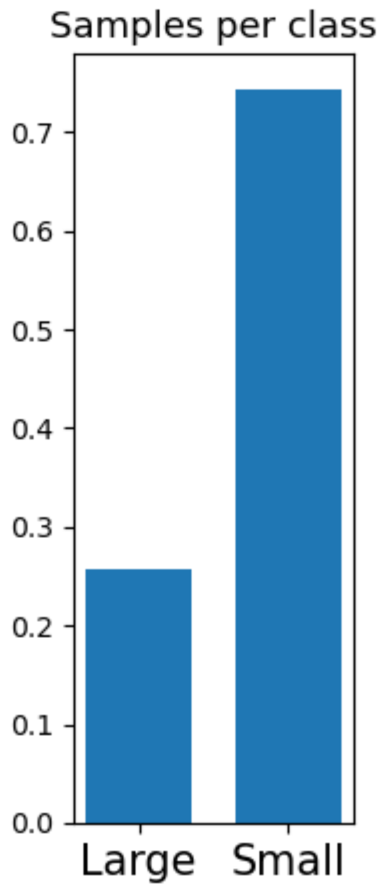
GM+NDSM



Approximate group symmetry

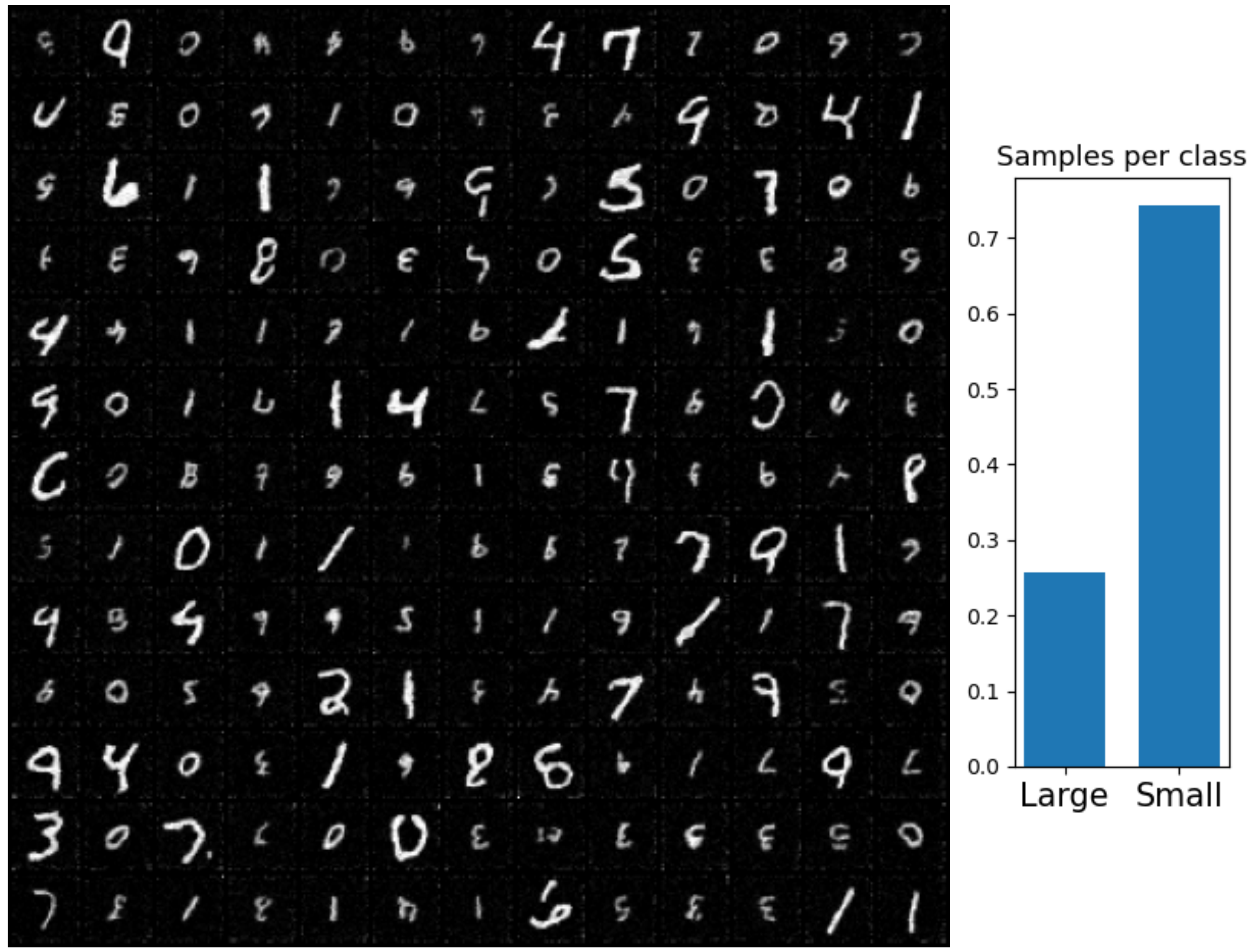
Approximate group symmetry

OU+DSM

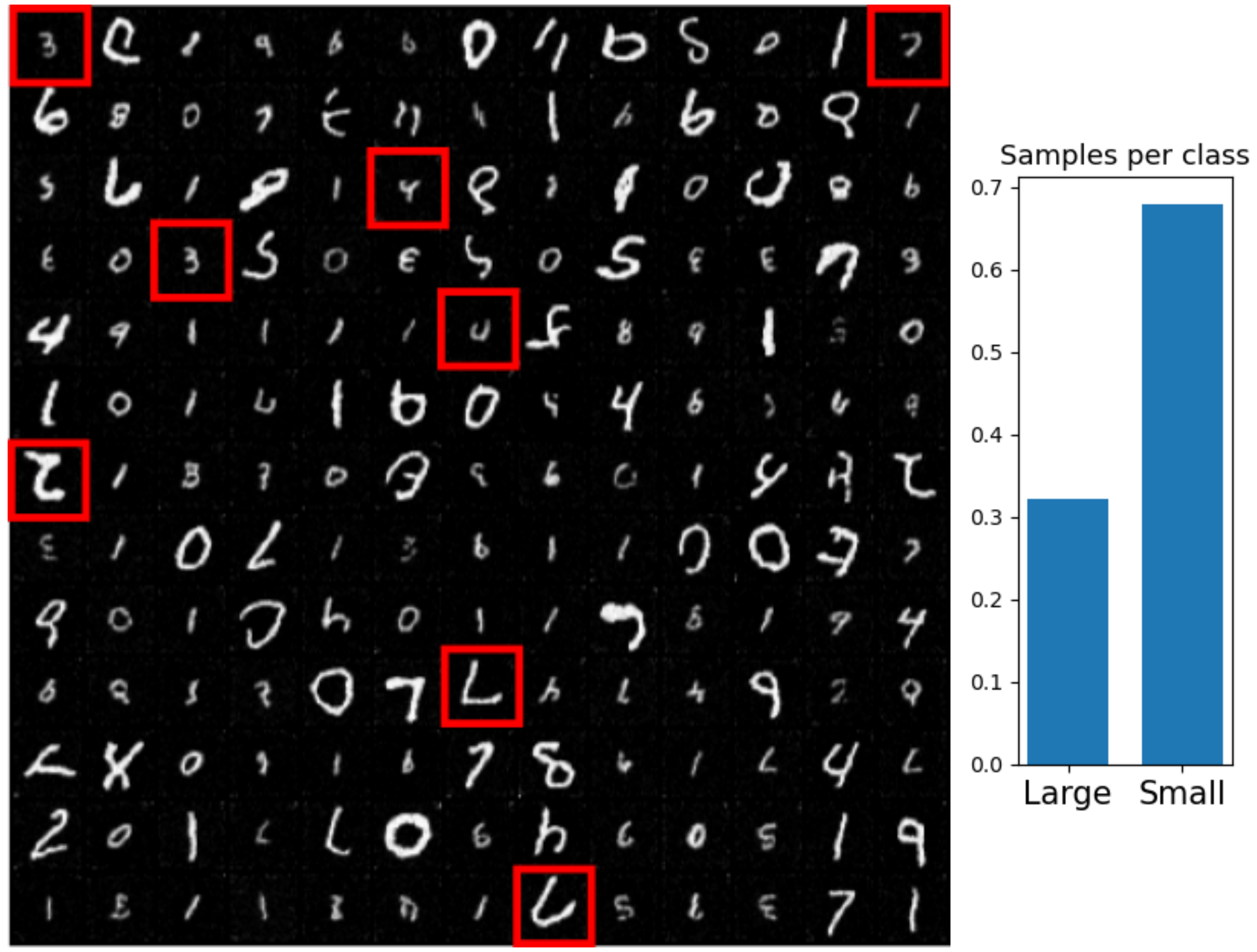


Approximate group symmetry

OU+DSM

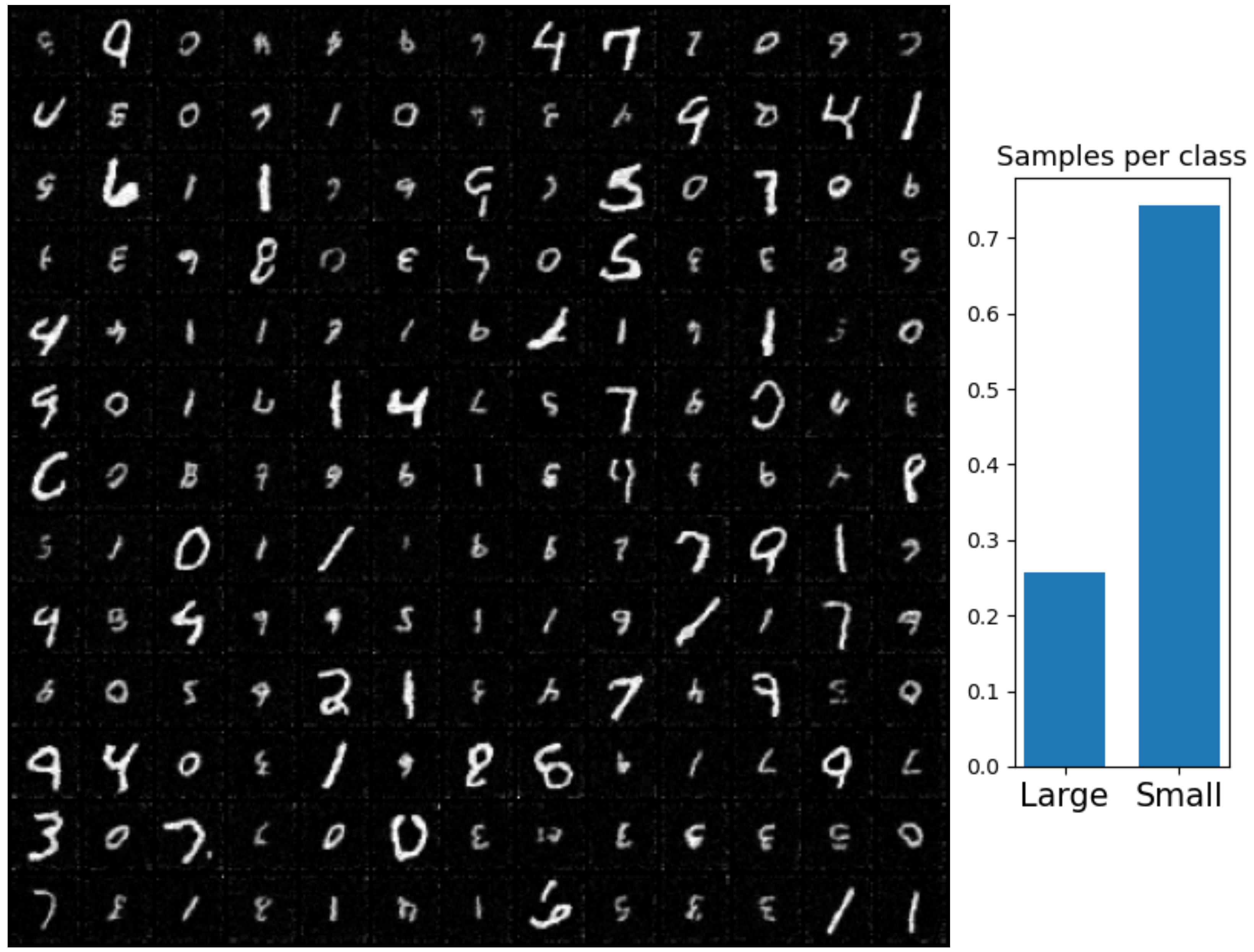


OU+DSM, Sym scores

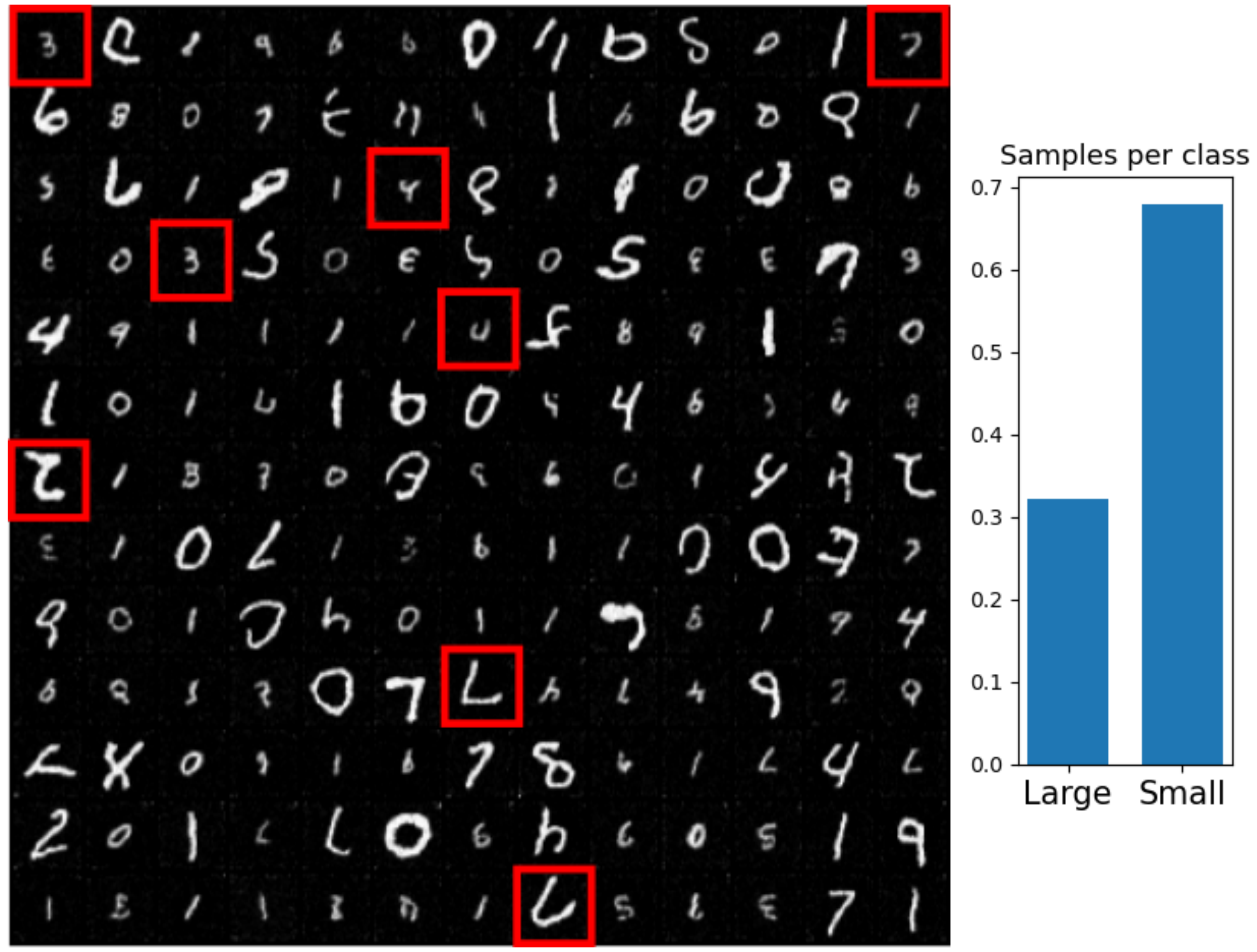


Approximate group symmetry

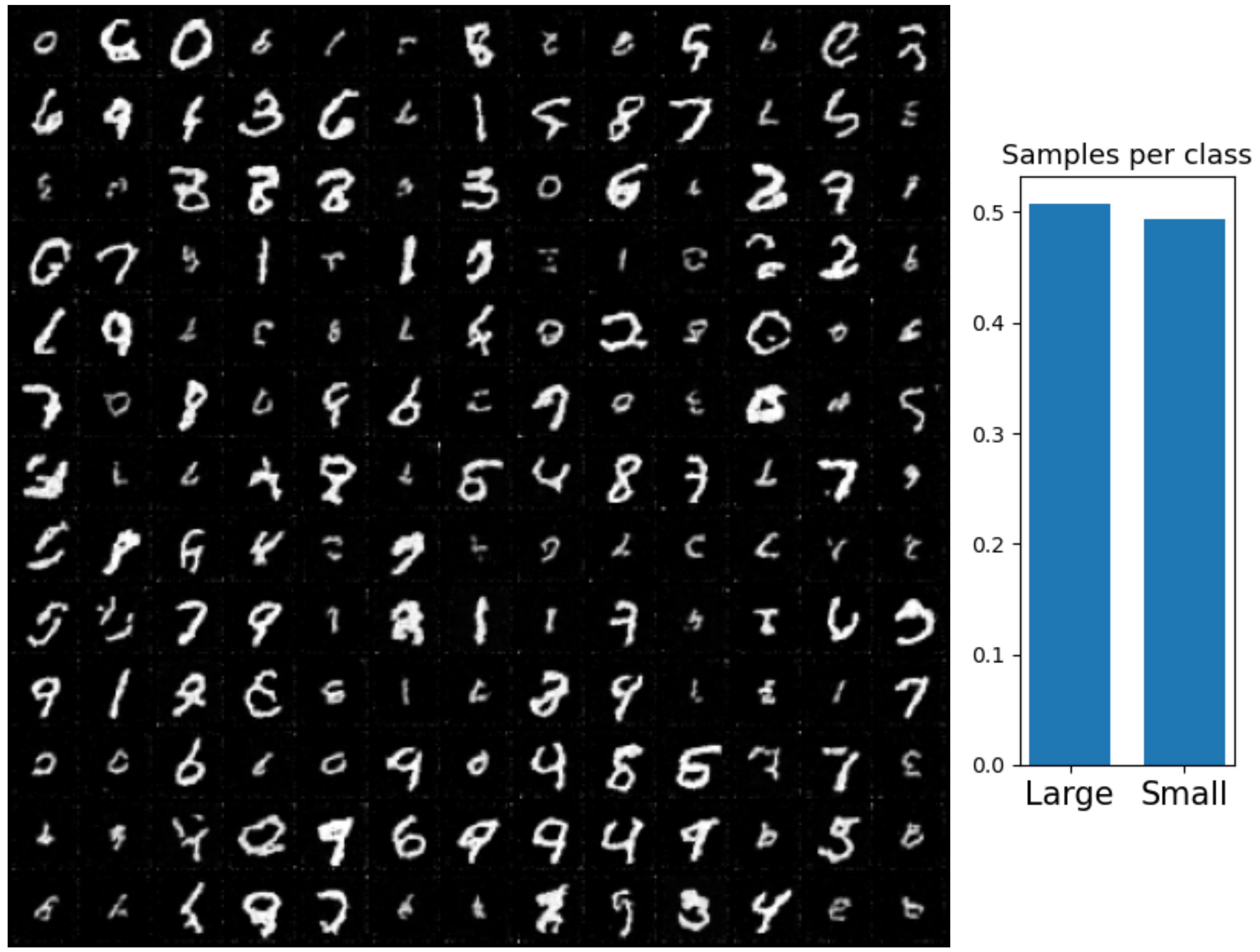
OU+DSM



OU+DSM, Sym scores



GM+NDSM



Data-driven structure discovery

PHYSICAL REVIEW E **108**, L022301 (2023)

Letter

Machine learning of independent conservation laws through neural deflation

Wei Zhu ,* Hong-Kun Zhang, and P. G. Kevrekidis
Department of Mathematics and Statistics, University of Massachusetts Amherst, Amherst, Massachusetts 01003-4515, USA

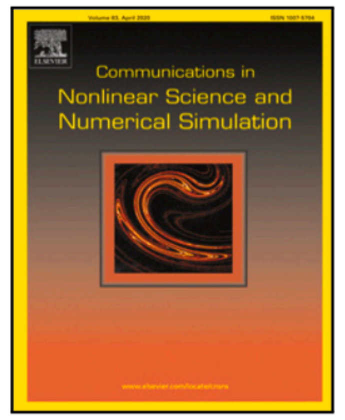
(Received 5 April 2023; accepted 24 July 2023; published 18 August 2023)

ELSEVIER

Contents lists available at ScienceDirect

Communications in Nonlinear Science and Numerical Simulation

journal homepage: www.elsevier.com/locate/cnsns



Data-driven discovery of conservation laws from trajectories via neural deflation

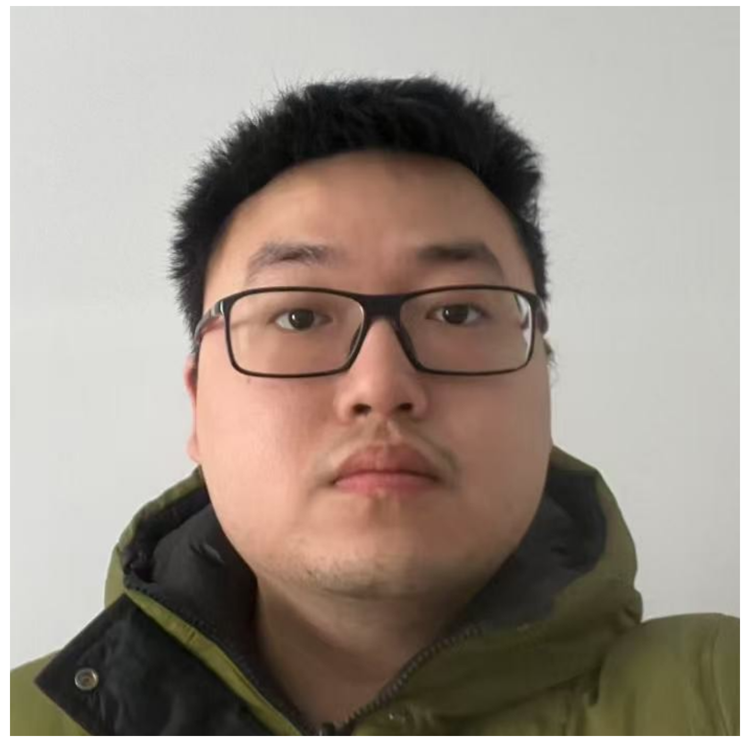
Shaoxuan Chen [a],*, Panayotis G. Kevrekidis [a], Hong-Kun Zhang [a], Wei Zhu [b]
[a] *Department of Mathematics and Statistics, University of Massachusetts Amherst, 710 N Pleasant St, Amherst, 01003-4515, MA, USA*
[b] *School of Mathematics, Georgia Institute of Technology, 686 Cherry St NW, Atlanta, 30332, GA, USA*

Data-driven structure discovery

Identification of moment equations via data-driven approaches in nonlinear Schrödinger models

Su Yang[1*], Shaoxuan Chen[1], Wei Zhu[1,2] and P. G. Kevrekidis[1]

[1]Department of Mathematics and Statistics, University of Massachusetts Amherst, Amherst, MA, United States, [2]School of Mathematics, Georgia Institute of Technology, Atlanta, GA, United States



COMPUTER ASSISTED DISCOVERY OF INTEGRABILITY VIA SILO: SPARSE IDENTIFICATION OF LAX OPERATORS

JIMMIE ADRIAZOLA*, WEI ZHU†, PANAYOTIS KEVREKIDIS‡, AND ALEJANDRO ACEVES§

Thank you for listening

Acknowledgement

- Air Force Young Investigator Award.
- NIH 1R01AI180243.
- NSF DMS-2052525, DMS-2140982, and DMS-2244976.