SC395:

Image Generative Models in Computer Vision

Viraj Shah

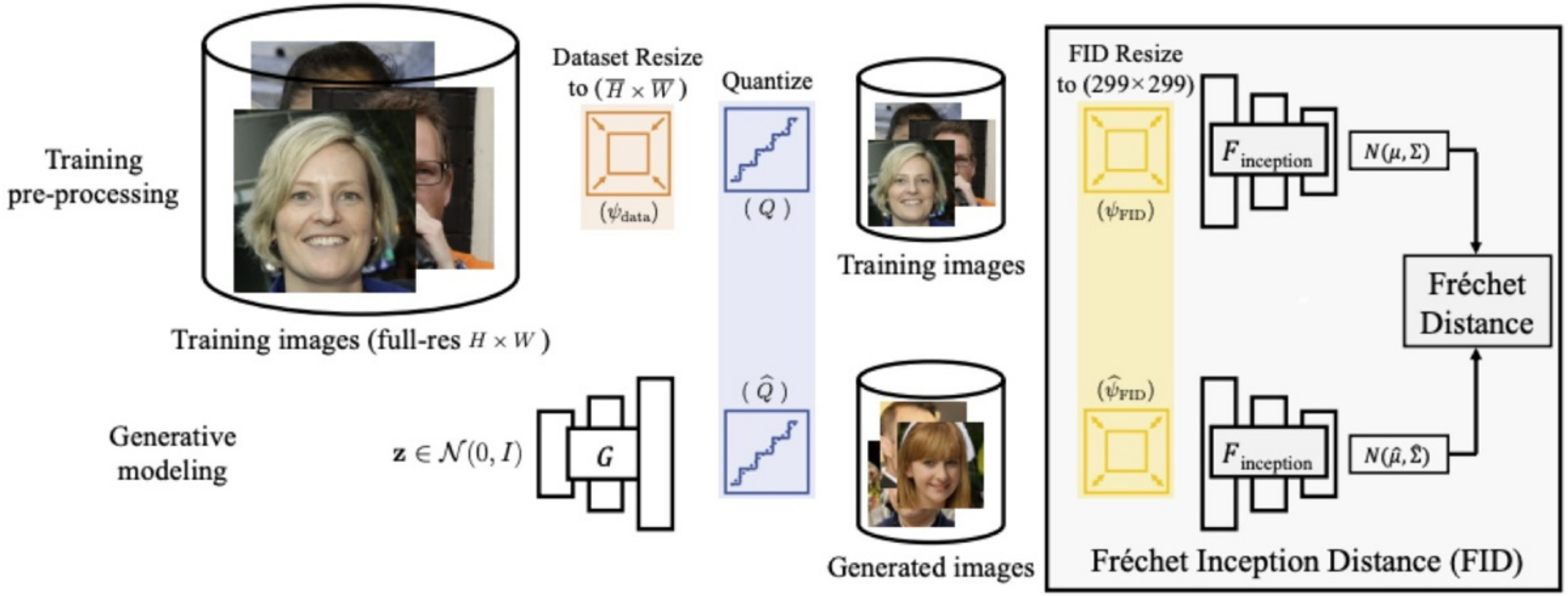
Lecture 3

Jan 9th, 2026

sc395.virajshah.com

GANs Applications

How to evaluate performance of GAN?



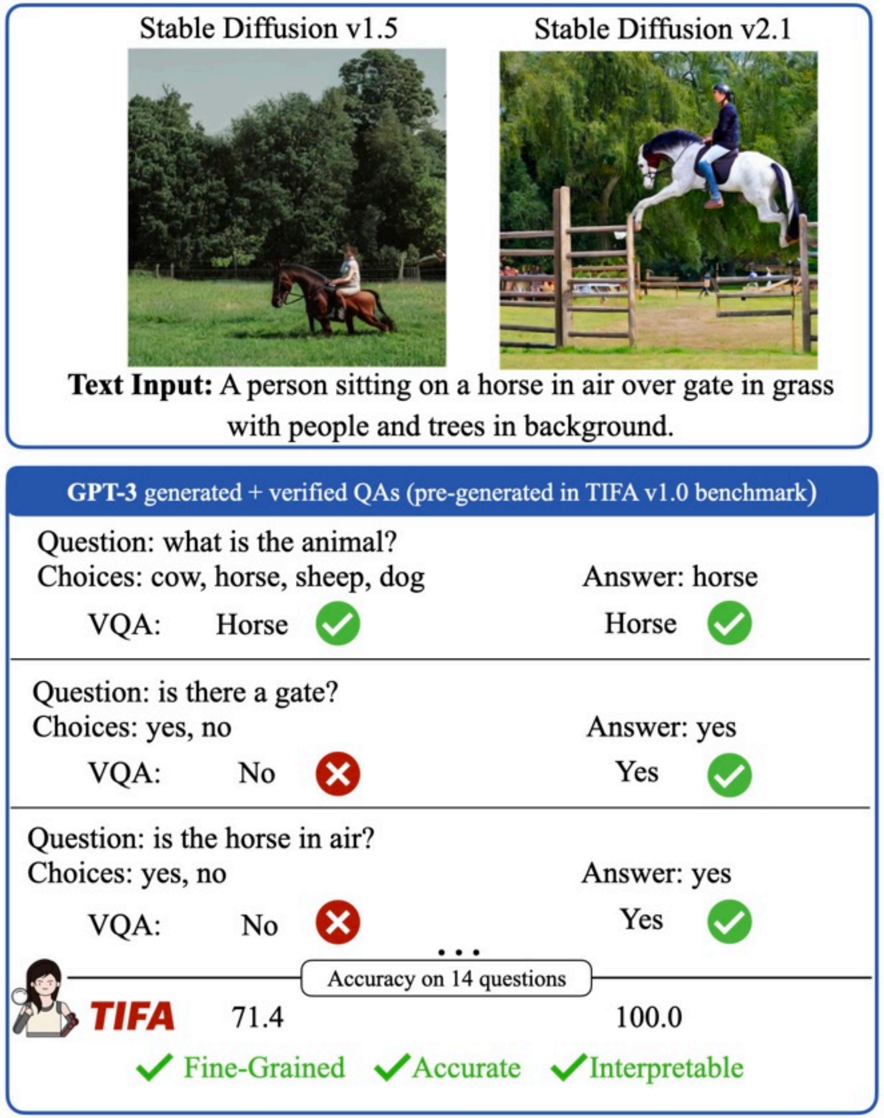
Fréchet Inception Distance (FID)

$$\mathbf{FID} = ||\mu - \hat{\mu}||_2^2 + \mathbf{Tr}(\Sigma + \widehat{\Sigma} - 2(\Sigma\widehat{\Sigma})^{1/2})$$

Slide from Jun-Yun Zhu, CMU

Recent Trends in Evaluation of Generative Models

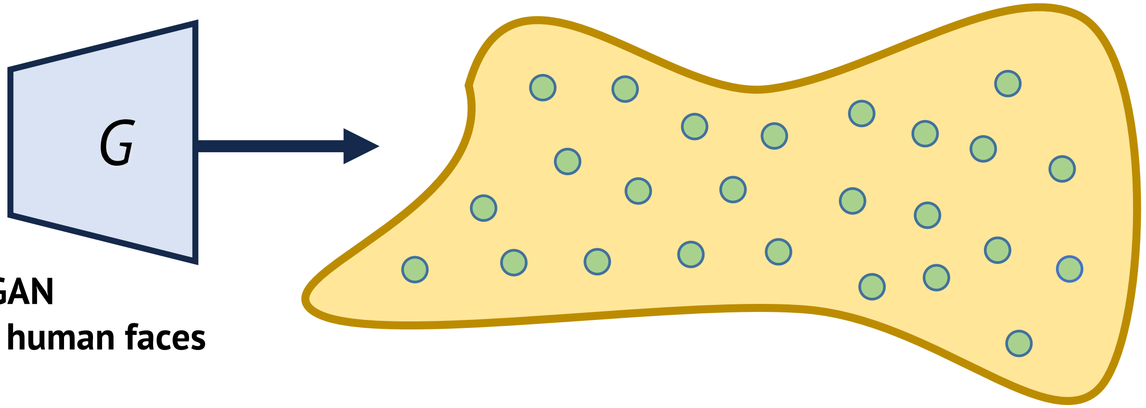
- Metrics that align human preferences
  - VQAScore
  - TIFA
- Human-in-the-loop Studies
  - LMArena:
  - lmarena.ai/leaderboard/text-to-image



GANs: Applications

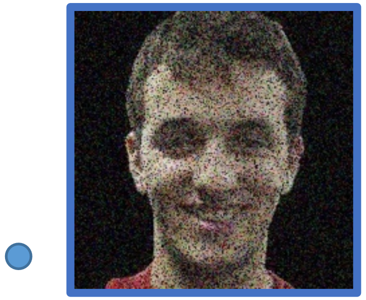
- Rich representation of the image data
- Smooth latent space with interpolation capabilities
- Fast, single step inference

Image Restoration using GANs



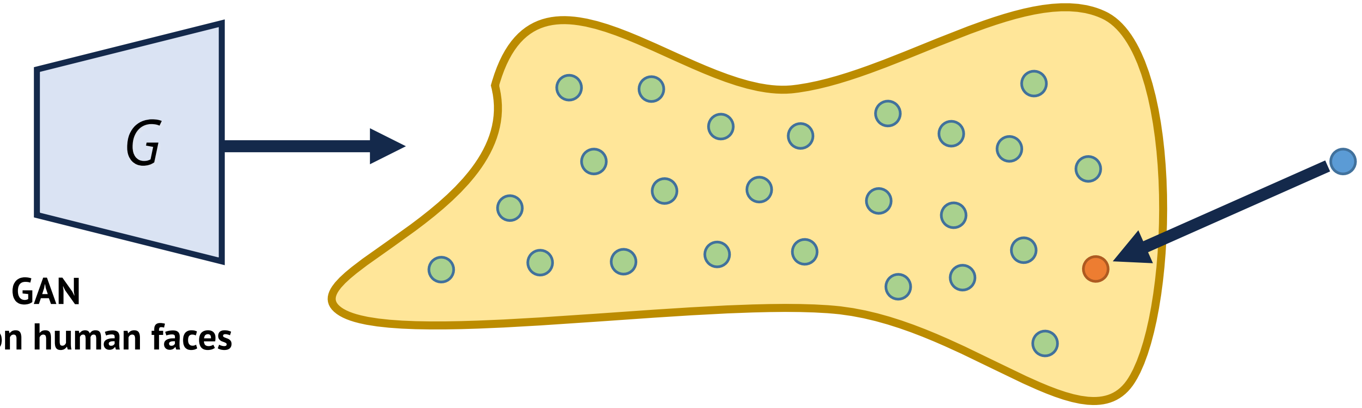
GAN
trained on human faces

GAN Image Space ≈ Set of all human faces



Corrupted Image

Image Restoration using GANs



GAN
trained on human faces

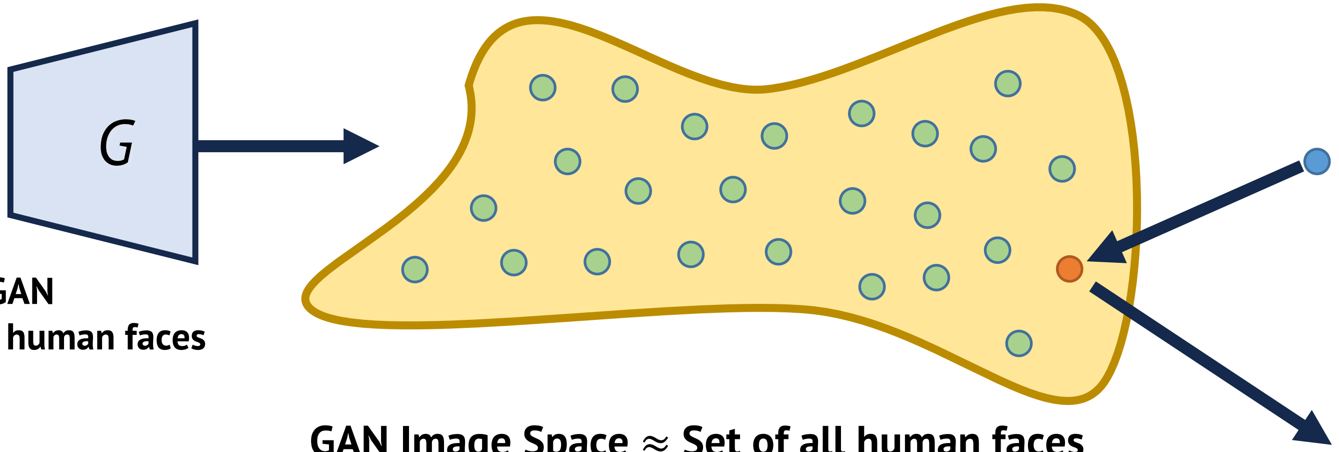
GAN Image Space ≈ Set of all human faces

Corrupted Image

Find the closest image in the GAN Image Space

Solving Inverse Problems using GAN priors, V. Shah et al., ICASSP '18

Image Restoration using Generative Models



GAN
trained on human faces

GAN Image Space ≈ Set of all human faces

Corrupted Image

Find the closest image in the GAN Image Space

$$w^* = \underset{z}{\operatorname{argmin}} \|y - G(w)\|_2$$

$\sim G(w^*)$

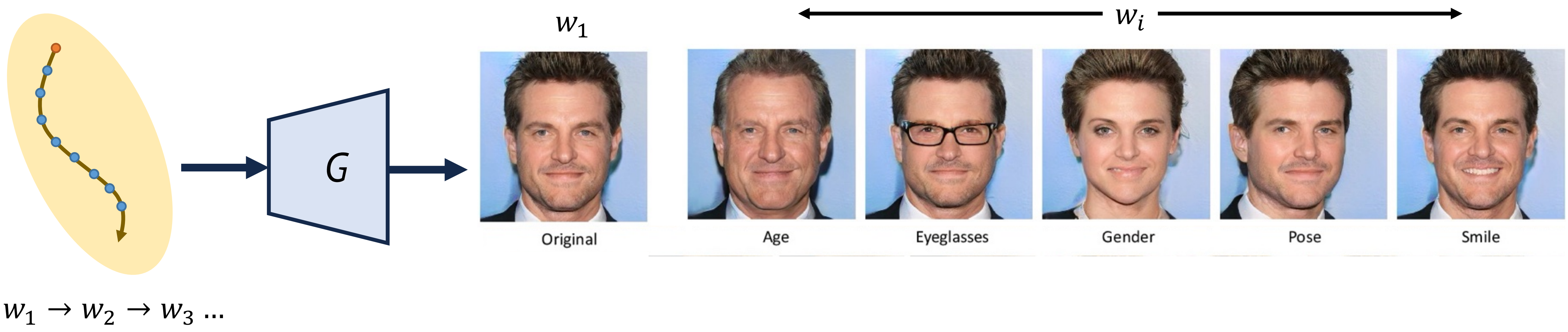
Restored Image

Provable algorithms for solving:

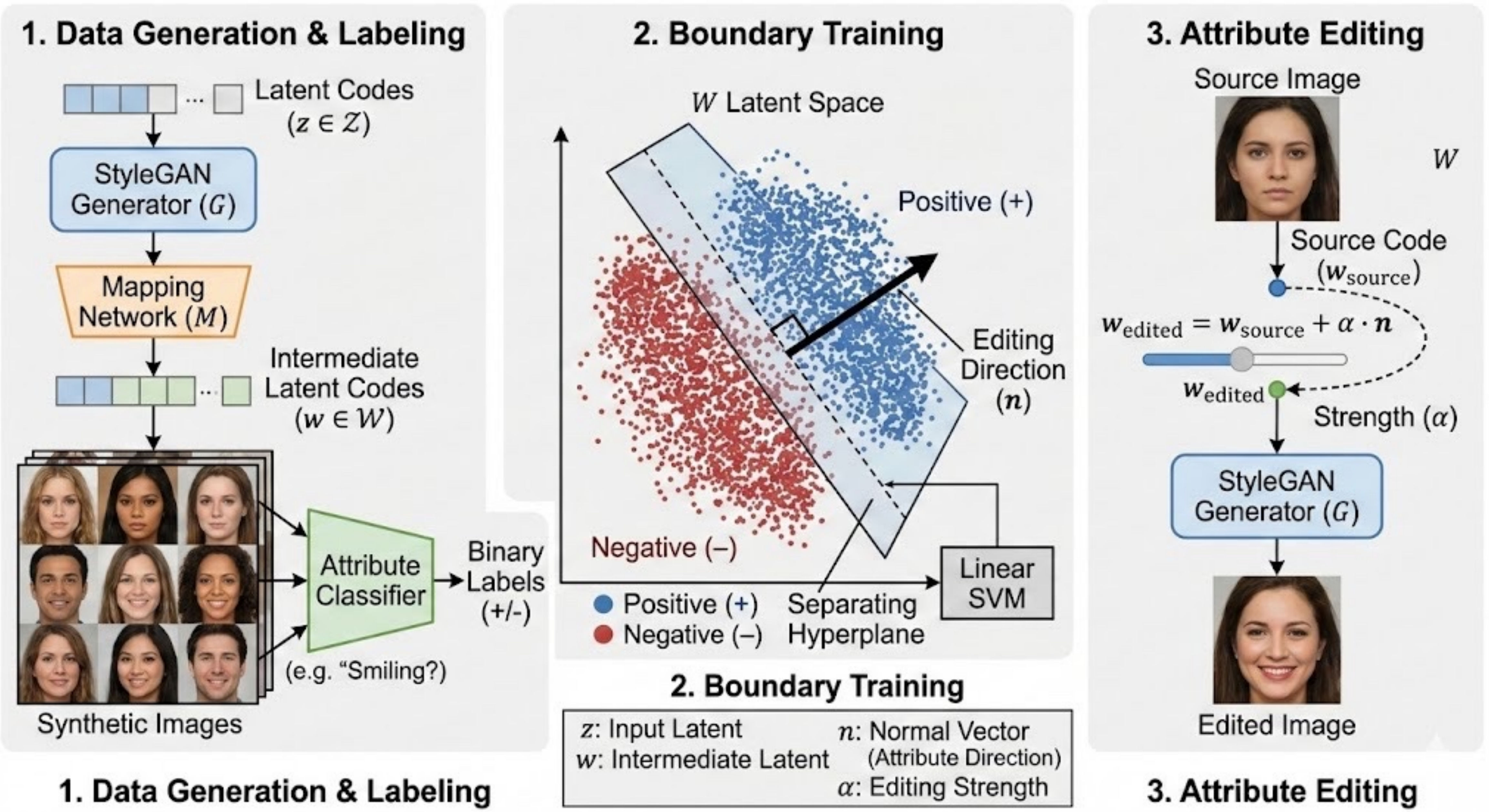
- Image Restoration
- Image Denoising, Inpainting etc.

Solving Inverse Problems using GAN priors, V. Shah et al., ICASSP '18

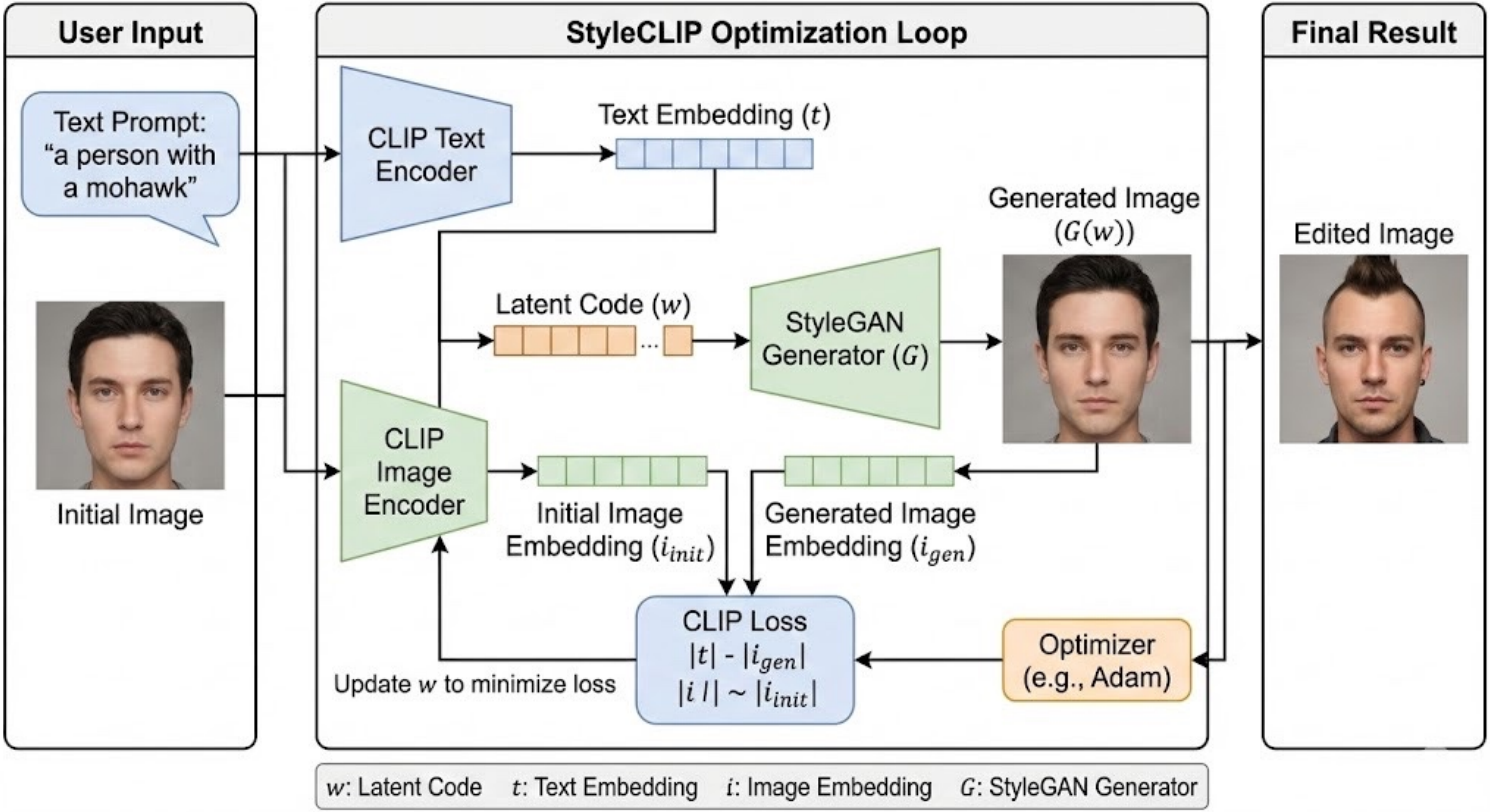
GAN Inversion is necessary for Image Editing



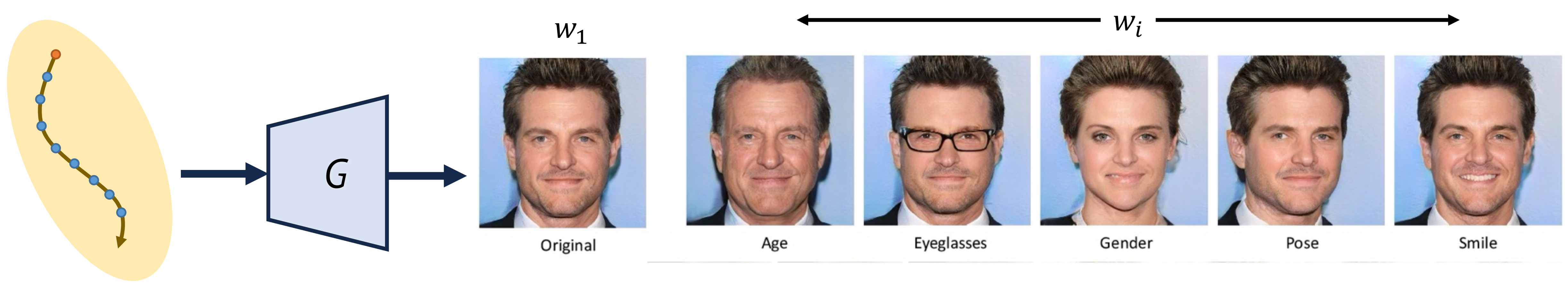
Ref: InterfaceGAN, Shen et al.. CVPR '20



1. Data Generation & Labeling
Latent Codes
$(z \in \mathcal{Z})$
StyleGAN Generator (G)
Mapping Network (M)
Intermediate Latent Codes
$(w \in \mathcal{W})$
Synthetic Images
Attribute Classifier
(e.g. "Smiling?)
Binary Labels
(+/-)
2. Boundary Training
W Latent Space
Positive (+)
Negative (–)
Editing Direction
(n)
Linear SVM
Positive (+)
Negative (–)
Separating Hyperplane
2. Boundary Training
z: Input Latent
w: Intermediate Latent
n: Normal Vector
(Attribute Direction)
α: Editing Strength
3. Attribute Editing
Source Image
W
Source Code
(w_{source})
$w_{\text{edited}} = w_{\text{source}} + \alpha \cdot n$
w_{edited}
Strength (α)
StyleGAN Generator (G)
Edited Image
1. Data Generation & Labeling
3. Attribute Editing


User Input
StyleCLIP Optimization Loop
Final Result
Text Prompt: "a person with a mohawk"
CLIP Text Encoder
Text Embedding (t)
Generated Image (G(w))
Edited Image
Latent Code (w)
StyleGAN Generator (G)
Initial Image
CLIP Image Encoder
Initial Image Embedding (i_{init})
Generated Image Embedding (i_{gen})
CLIP Loss
$|t|$ - $|i_{gen}|$
$|i\,/|$ ~ $|i_{init}|$
Optimizer (e.g., Adam)
Update w to minimize loss
w: Latent Code
t: Text Embedding
i: Image Embedding
G: StyleGAN Generator

GAN Inversion is necessary for Image Editing



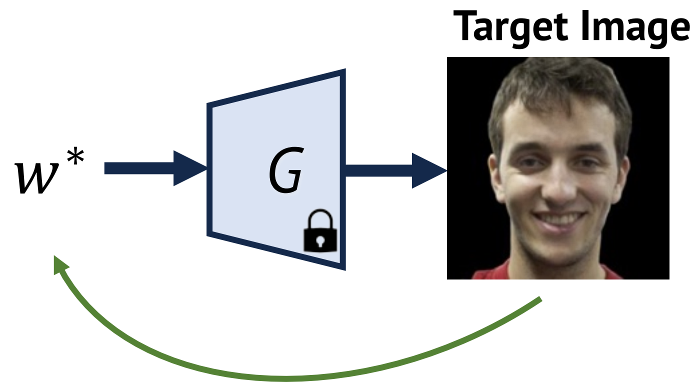
$w_1 \rightarrow w_2 \rightarrow w_3 \ldots$

Edit a real image?

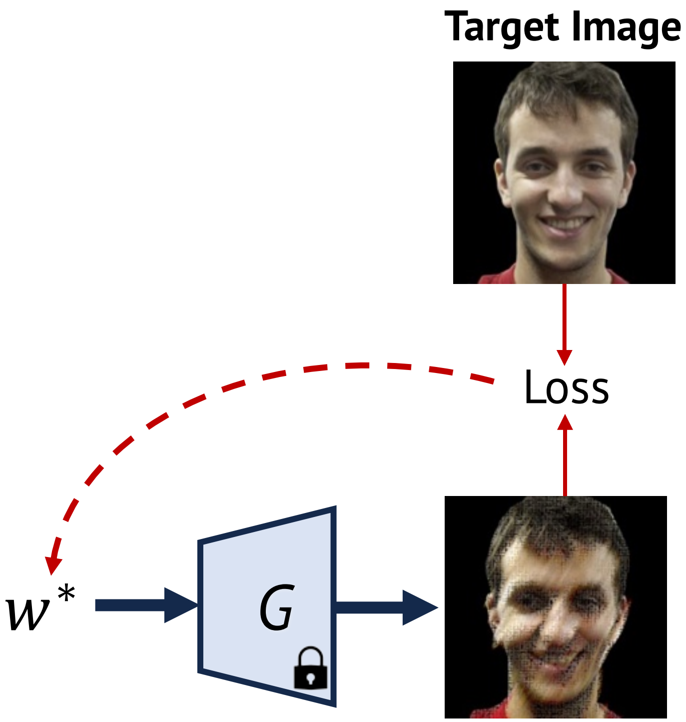
Corresponding Latent code must be known!

Ref: InterfaceGAN, Shen et al.. CVPR '20

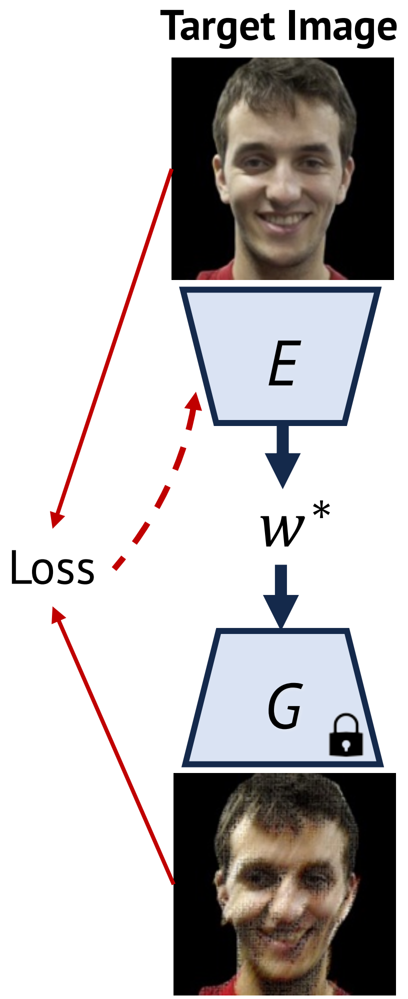
GAN Inversion



GAN Inversion



Optimization-based

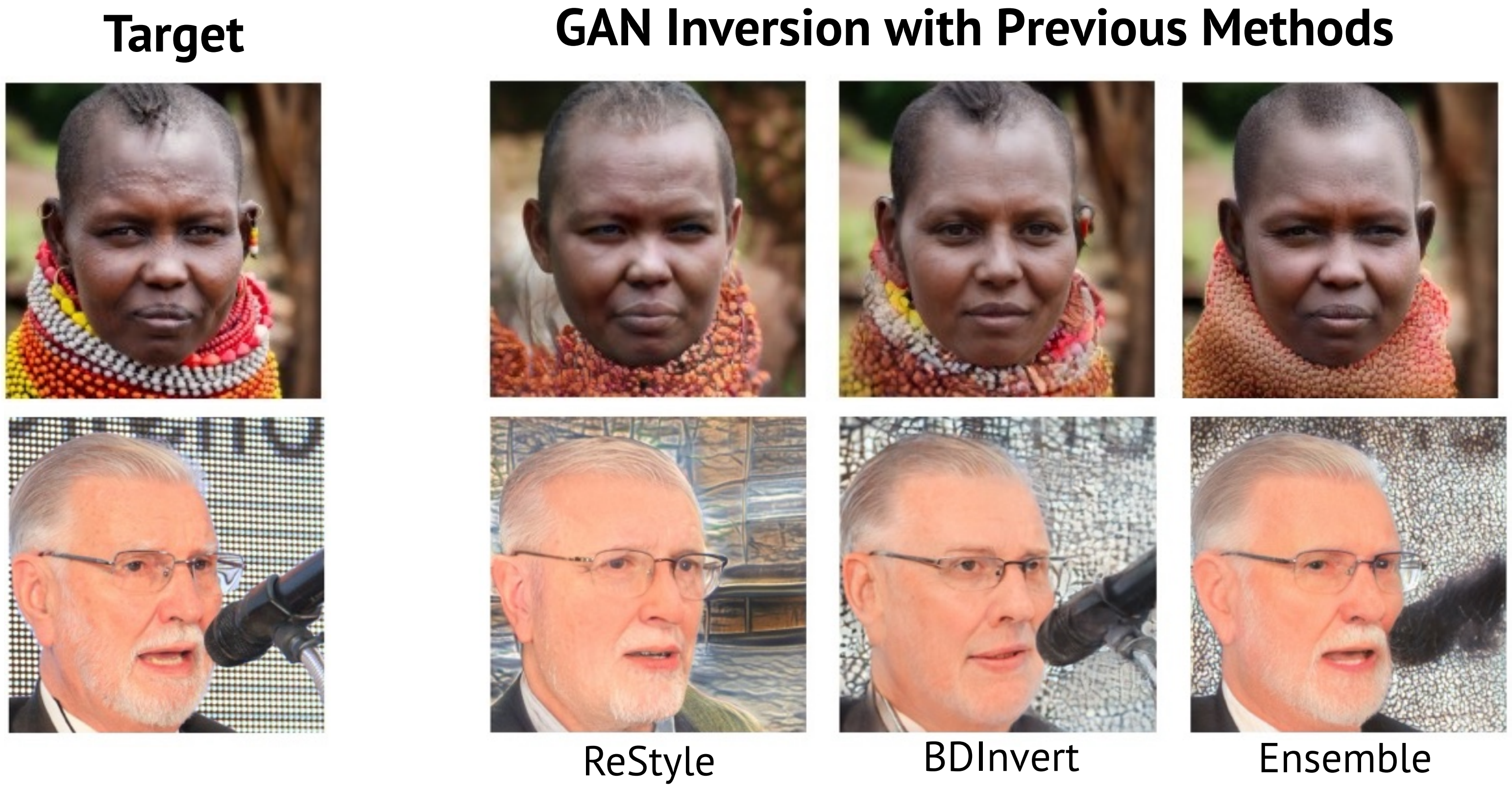


Encoder-based

backprop

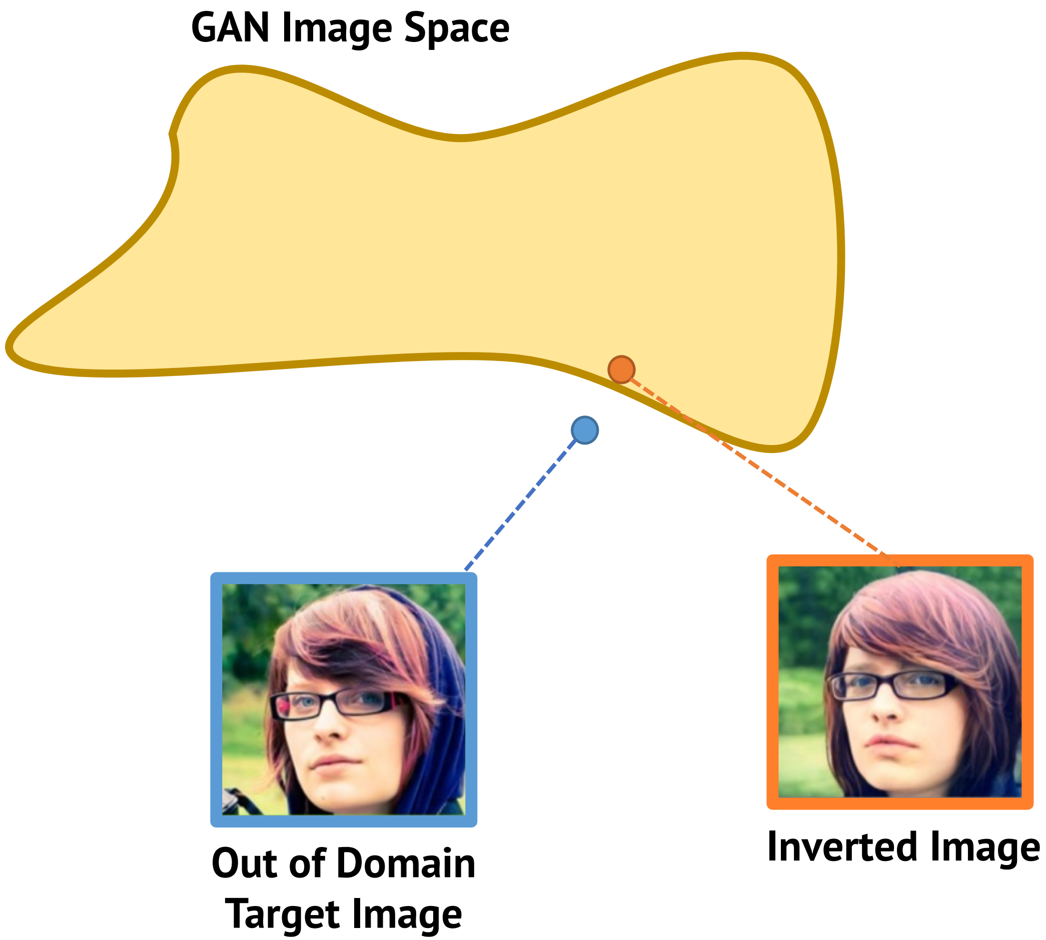
Ref: GAN Inversion Survey, Xia et al. ’21; Image2StyleGAN, Abdal et al. ICCV ’19; e4e, Tov et al. SIGGRAPH 2020

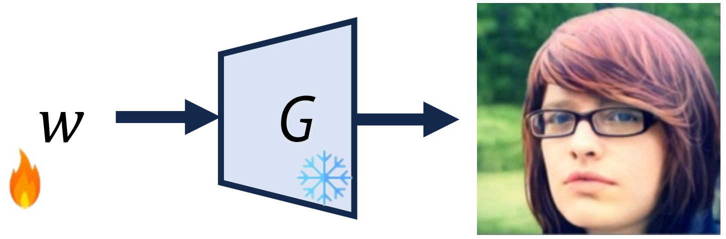
GAN Inversion: Most Techniques Fail on out-of-domain Images



Ref: ReStyle, Alauf et al., ICCV '21
BDInvert, Kang et al., ICCV '21
Ensemble, Cai et al., CVPR '21

Key Challenge in Inverting a Frozen GAN

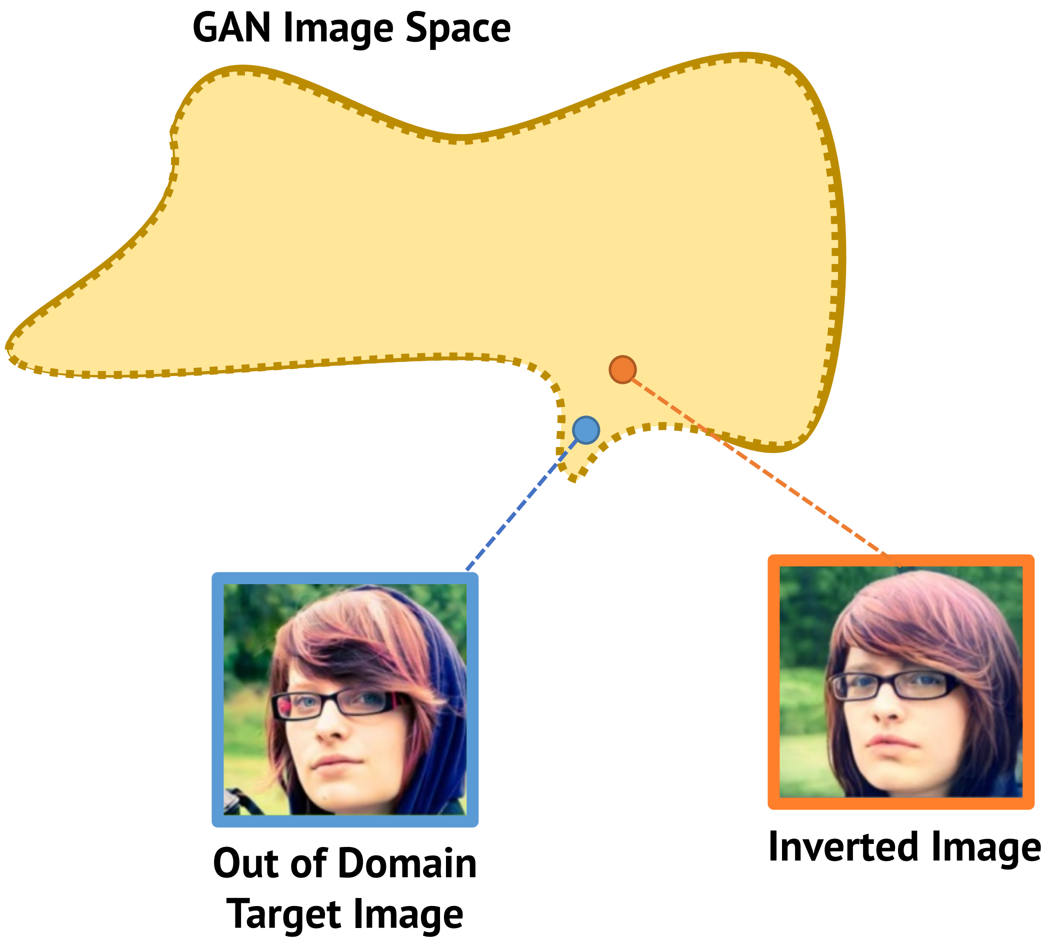


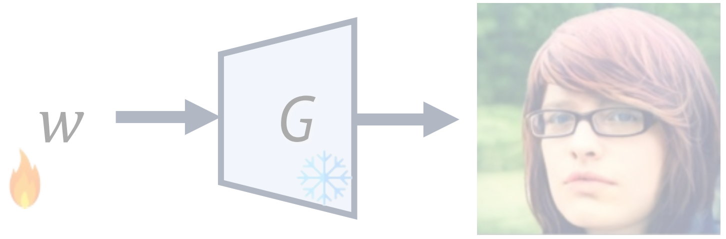


Generator is *Fixed*, Latent is *Flexible*

Near-perfect GAN Inversion, Q. Feng, V. Shah et al., ‘22; **Make It So**, A. Bhattad, V. Shah et al., ‘23

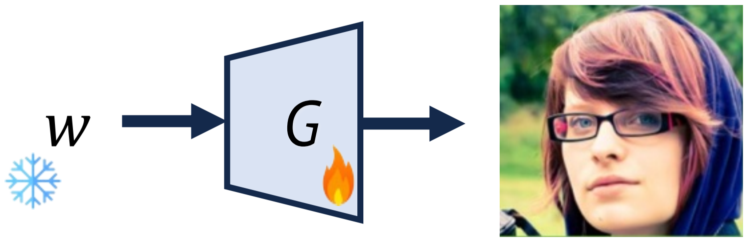
Key Challenge in Inverting a Frozen GAN





Generator is *Frozen,* Latent is *Trainable*

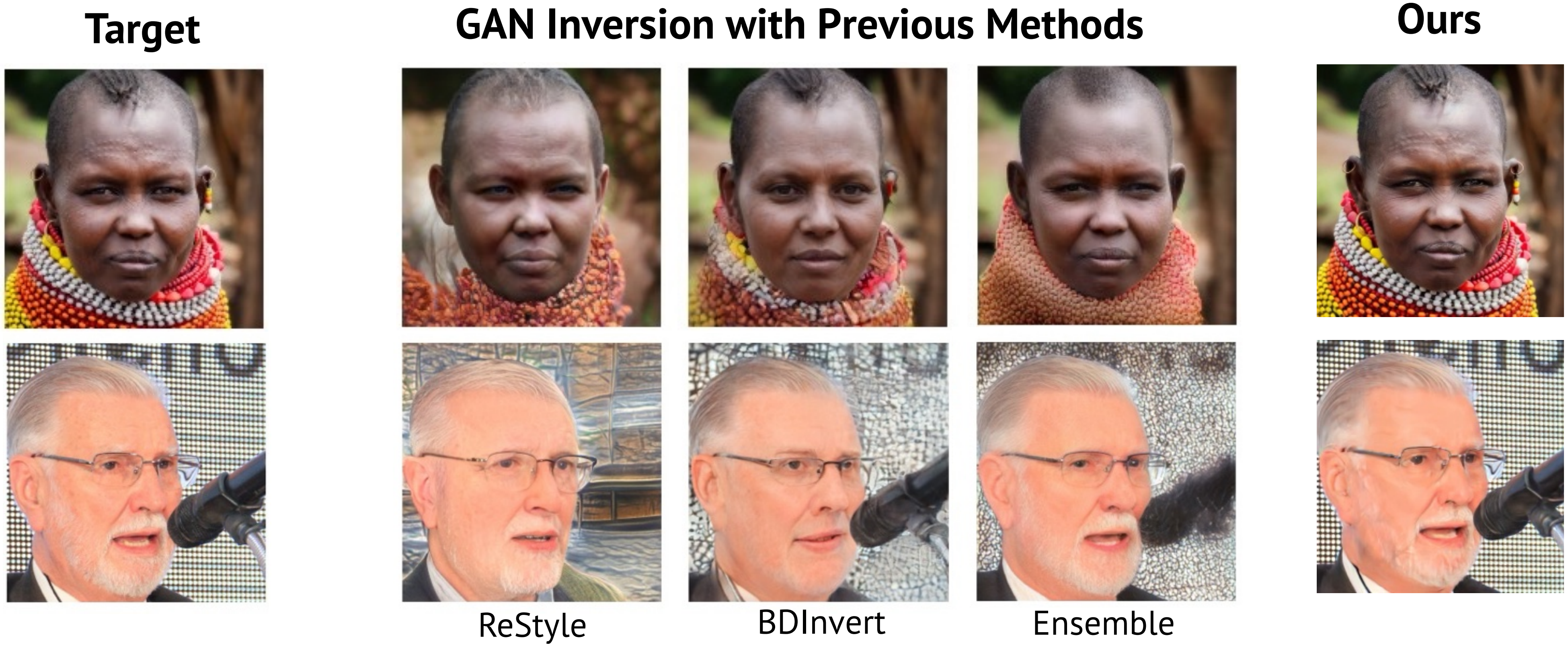
Proposed Idea:



Generator is *Flexible*, Latent is *Fixed*

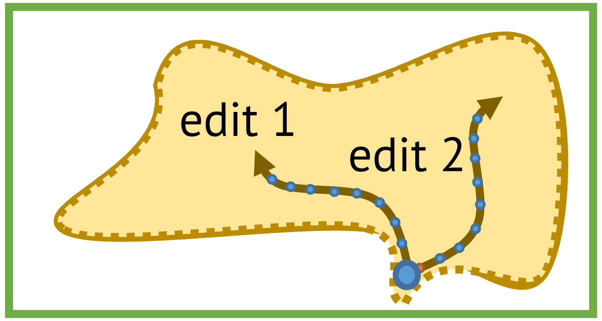
Near-perfect GAN Inversion, Q. Feng, V. Shah et al., '22; **Make It So**, A. Bhattad, V. Shah et al., '23

Our Method Achieves Near-perfect GAN Inversion



Ref: ReStyle, Alauf et al., ICCV '21
BDInvert, Kang et al., ICCV '21
Ensemble, Cai et al., CVPR '21

Most off-the-shelf Editing methods work!



Eyes

Mouth

Lipstick

Hair

using **StyleSpace Editing Directions**

Ref: StyleSpace, CVPR 2021

One-shot Image Stylization

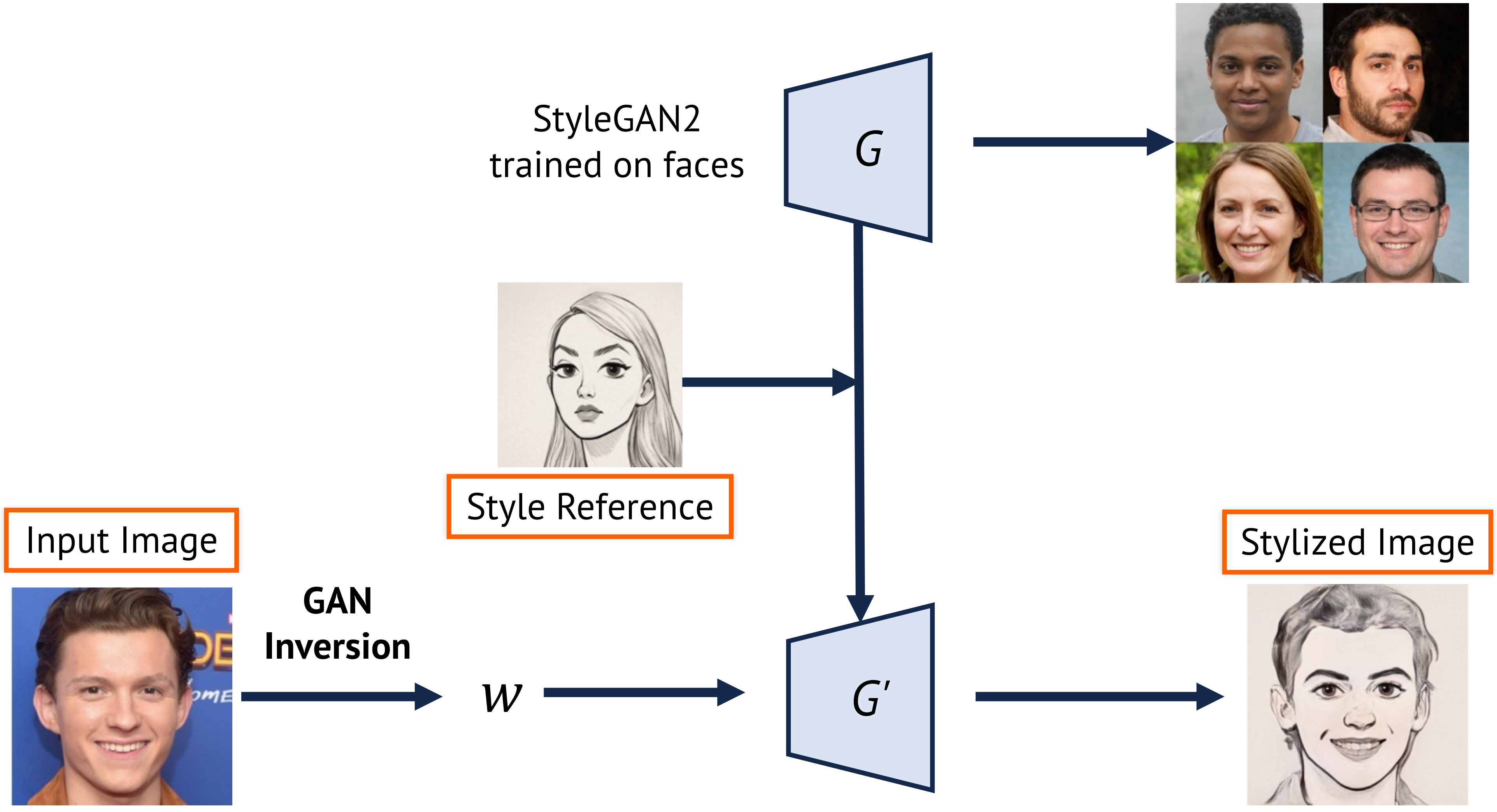
Style Reference

Input Image

Stylized Image

Stylizing an Input Image in the style of a Reference Style using only *one* example

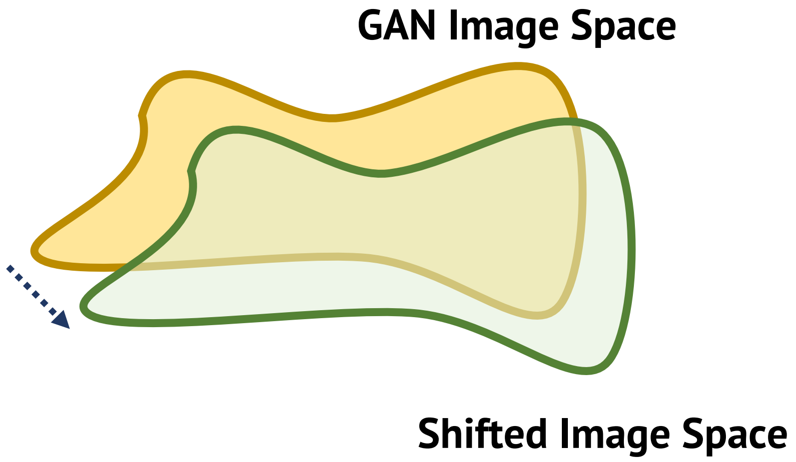
Pre-trained GAN for one-shot Stylization

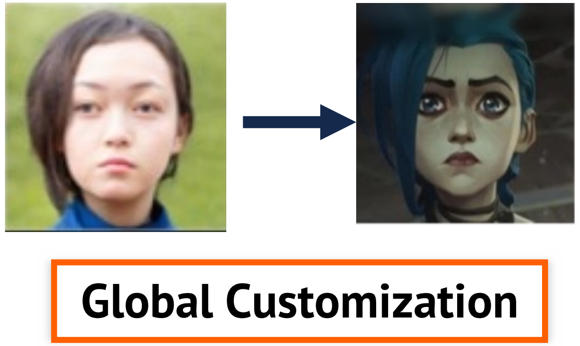


Fine-tuned StyleGAN2

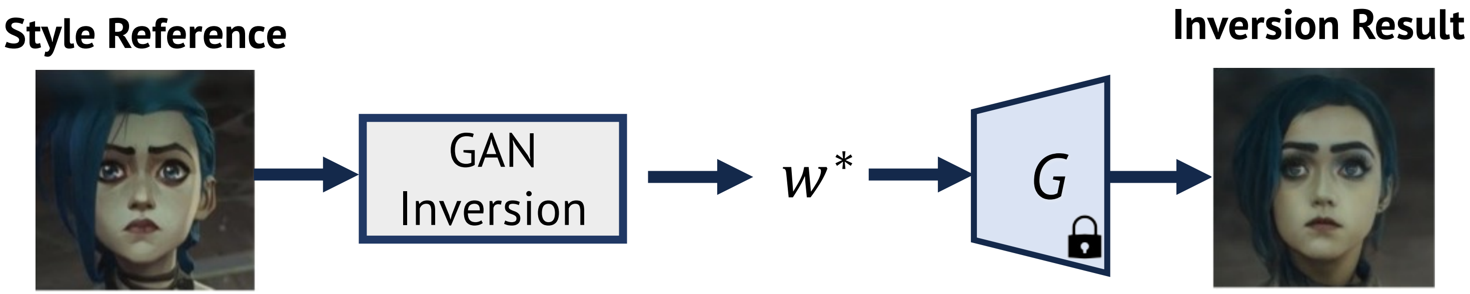
JoJoGAN, Chong et al. ECCV '22

How to perform global customization of StyleGAN2?





Step 1: GAN Inversion

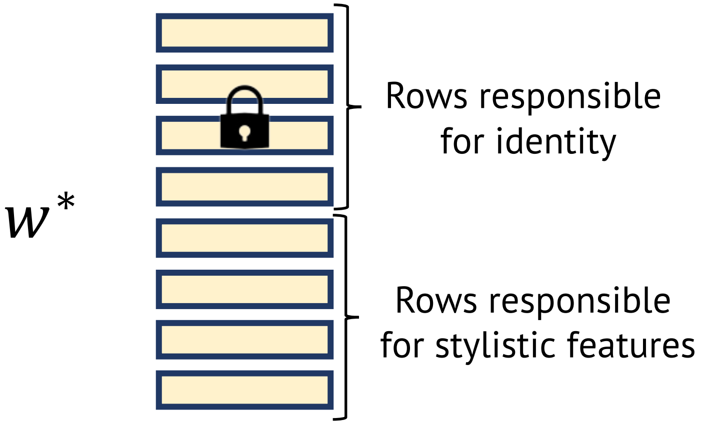


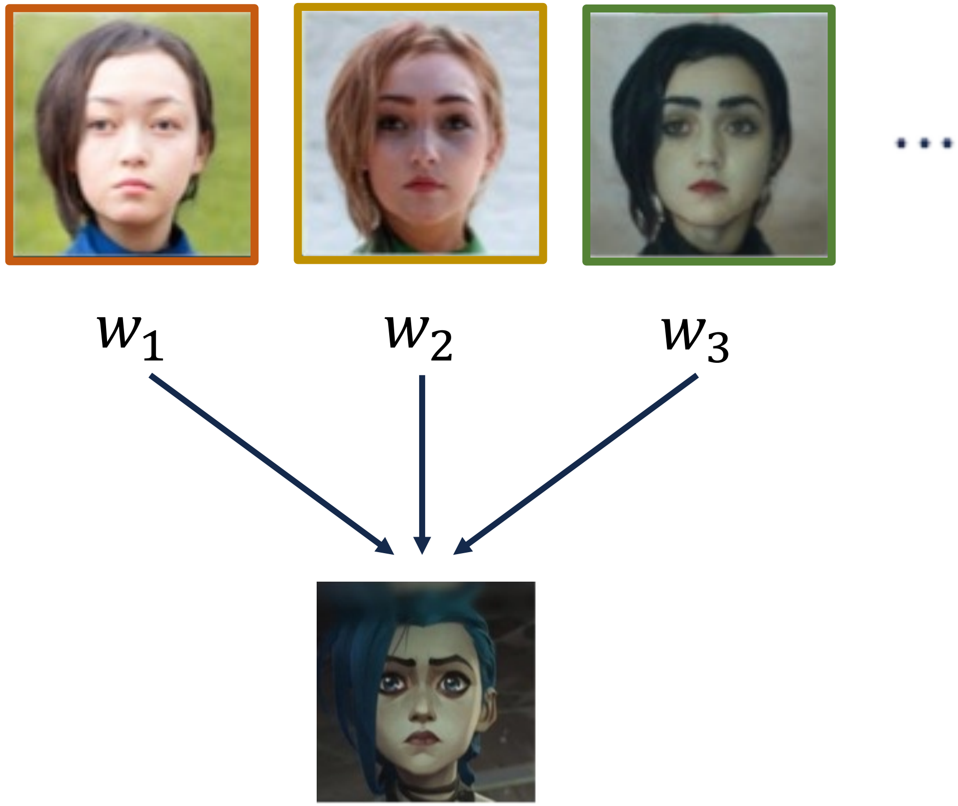
Ref: GAN Inversion Survey, Xia et al.'21; e4e, Tov et al. SIGGRAPH 2020

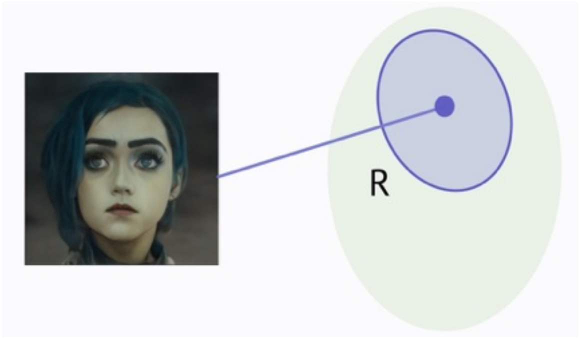
JoJoGAN: Style-mixing

Step 2. Use Style-mixing property to Create a Training Set

Inversion result





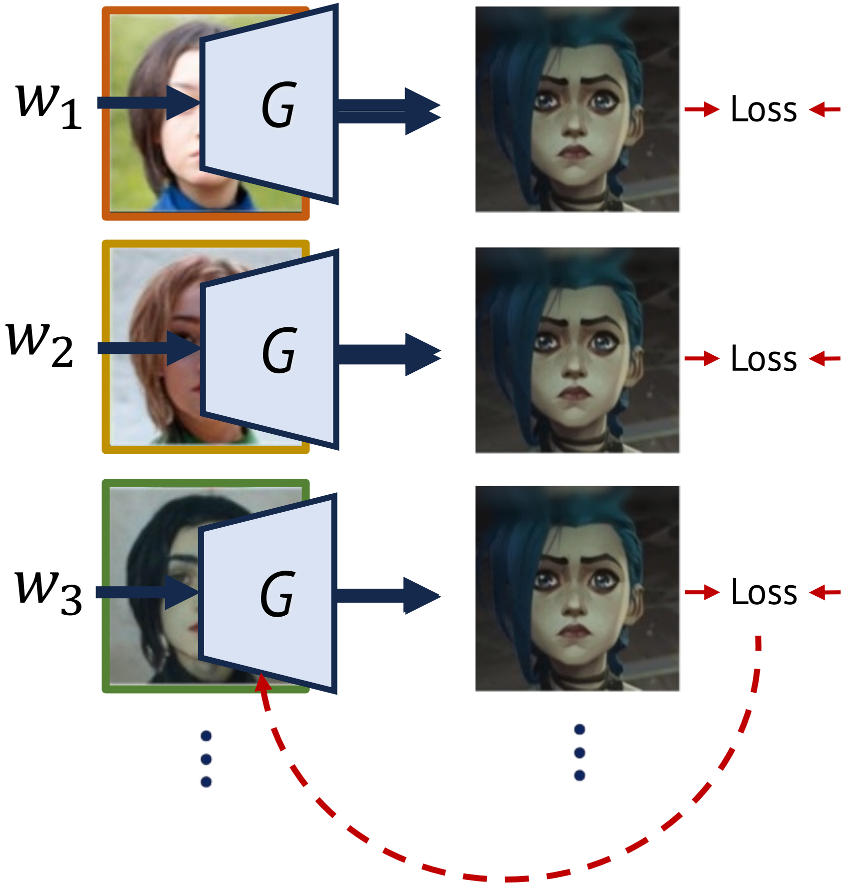


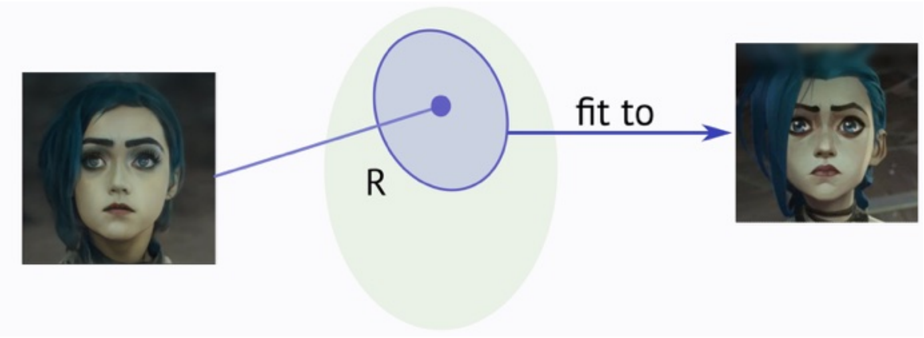
S or W space

Ref: StyleGAN, Karras et al. CVPR '19; StyleGAN2, Karras et al. CVPR '20

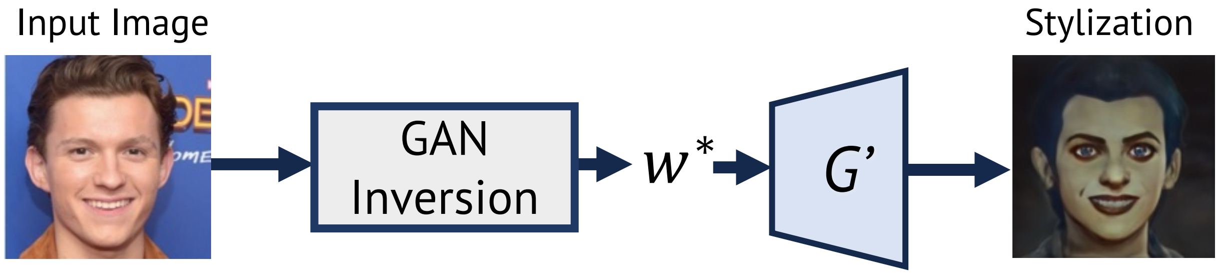
How to Fine-tune StyleGAN2 ? : JoJoGAN

Fine-tune *G* using pixel-level losses



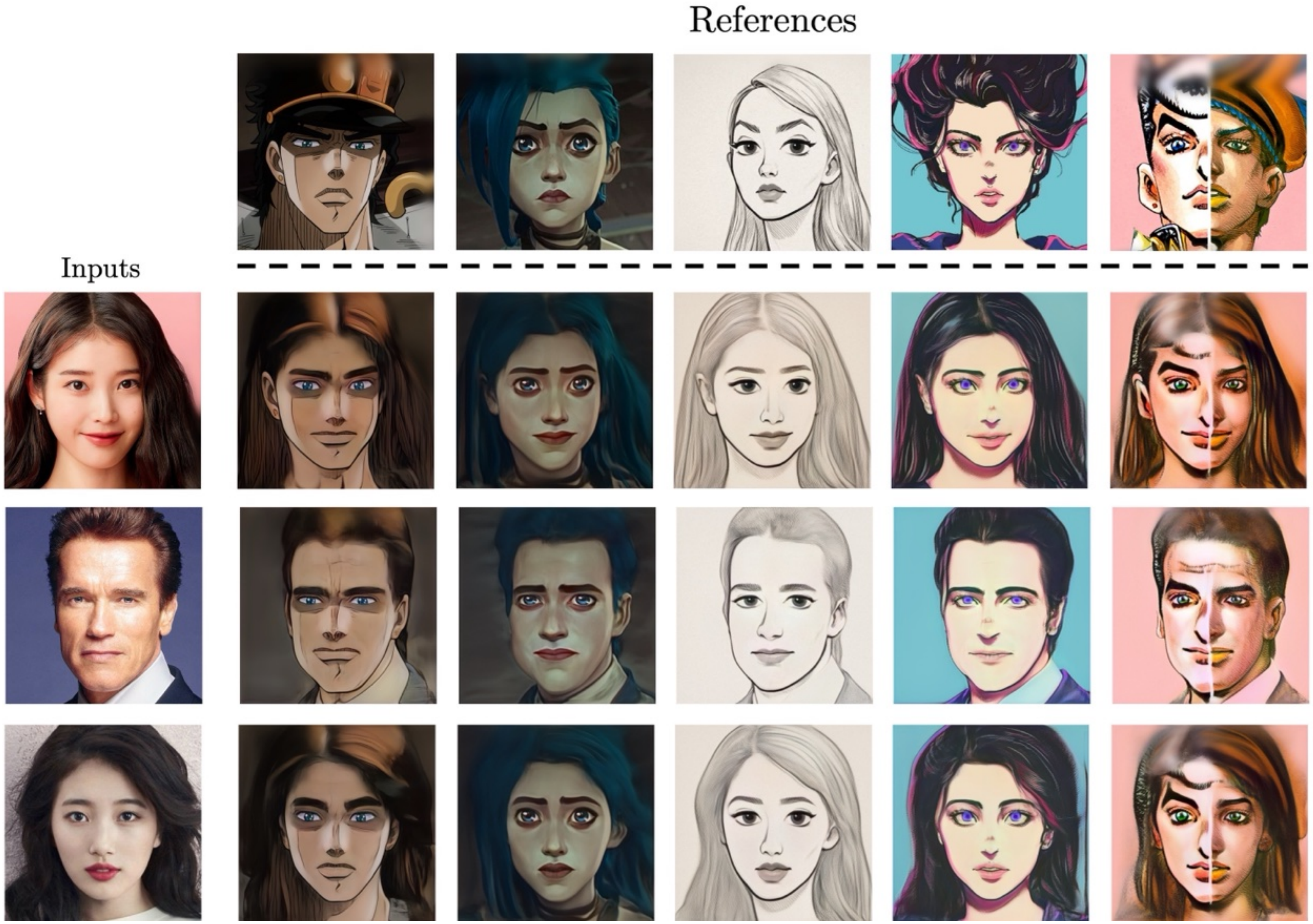


Inference on fine-tuned *G*



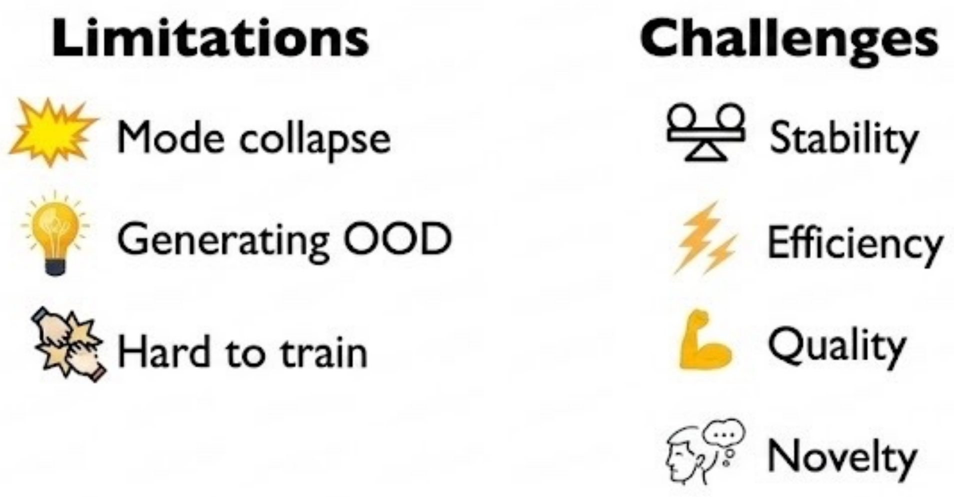
Ref: JoJoGAN, Chong et al. ECCV '22

JoJoGAN: Results



GANs: Limitations and Challenges

- Tricky to train due to two-player dynamics
- Thus, difficult to scale
- Mostly limited to single domain datasets
  - Exceptions: StyleGAN-XL, GigaGAN
- Mode Collapse is common, limited quality
- Are GANs still relevant?! : [current and future trends]

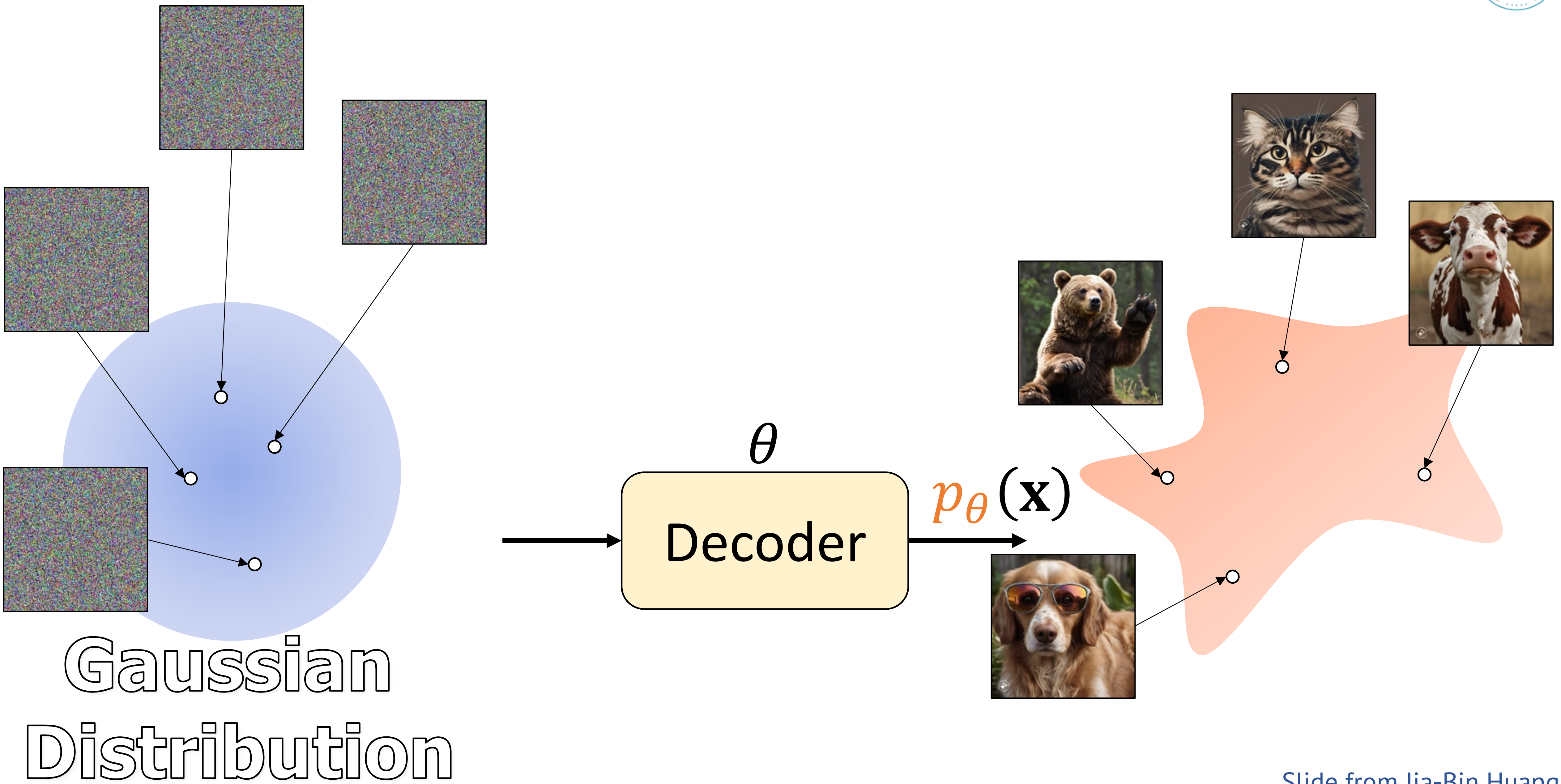


- **GANs don't use "direct supervision"**
  - **The loss relies on the indirect "real / fake" signal coming from Discriminator**

Important GAN concepts

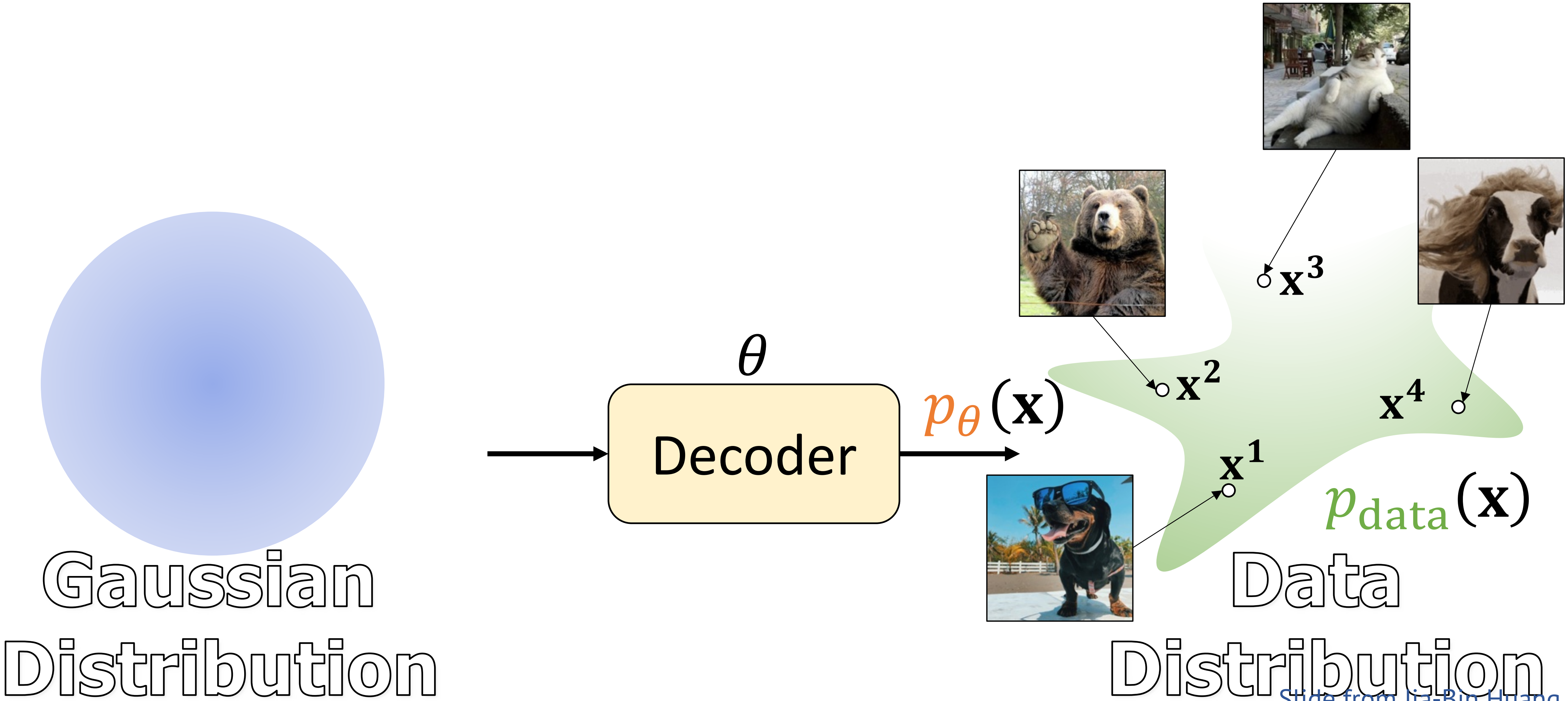
- The idea of two player game and minimax objective
- Standard objective minimizes JS divergence, but other divergences could potentially be more effective
  - E.g. Wasserstein Distance (aka Earth Mover's Distance)
- Scaling the model architecture is important
- Adaptive Instance Normalization introduced by StyleGAN is super-important -- StyleGAN architecture is the cornerstone of "image editing with GAN" literature
- The main advantage of GAN is smooth latent space with intuitive behavior
- Editing a real image requires GAN Inversion; once the inverted latent code is obtained, thousands of latent space traversal techniques are available
- GANs require the dataset to be aligned/single domain
- GANs are difficult to scale with respect to data due to unstable training dynamics

Lecture 3: Diffusion Models



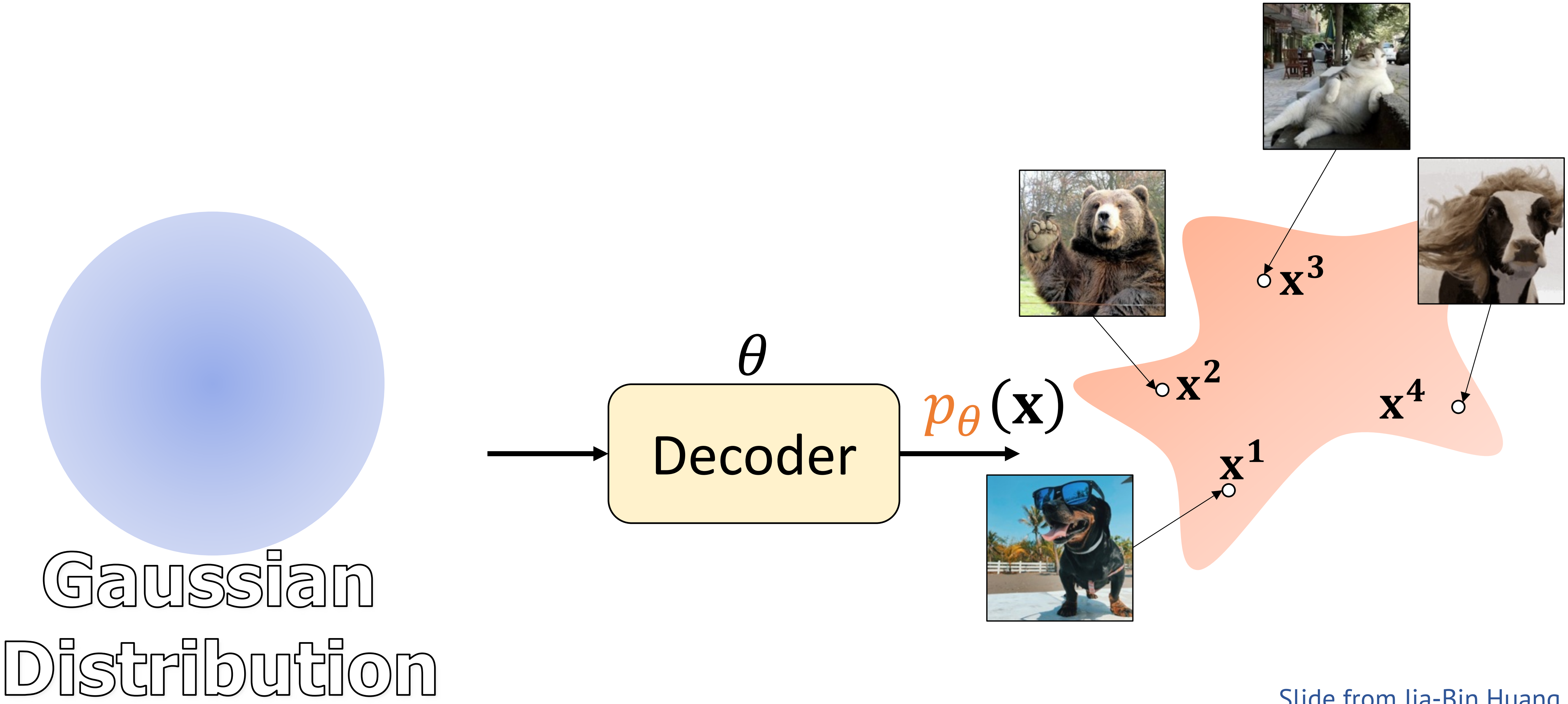
Slide from Jia-Bin Huang

Dataset: $\{\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^m\} \sim p_{\text{data}}(\mathbf{x})$



Slide from Jia-Bin Huang

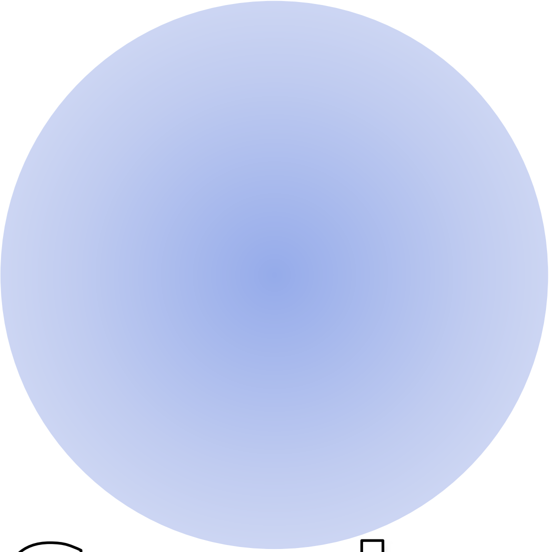
Dataset: $\{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^m\} \sim p_{\text{data}}(\mathbf{x})$



Dataset: $\{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^m\} \sim p_{\text{data}}(\mathbf{x})$

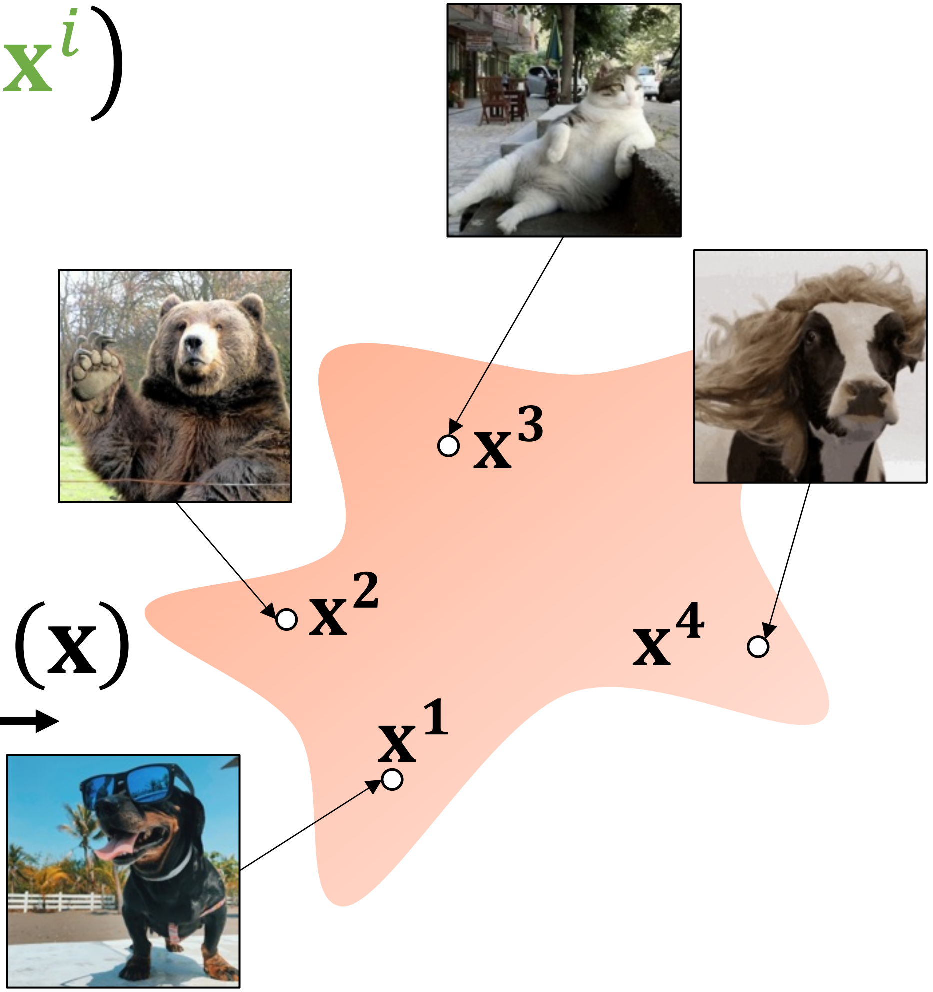
Maximum Likelihood

$$\theta^* = \operatorname{argmax} \prod_{i=1}^{m} p_\theta(\mathbf{x}^i)$$



Gaussian Distribution

$\longrightarrow$ Decoder (θ) $\longrightarrow$ $p_\theta(\mathbf{x})$

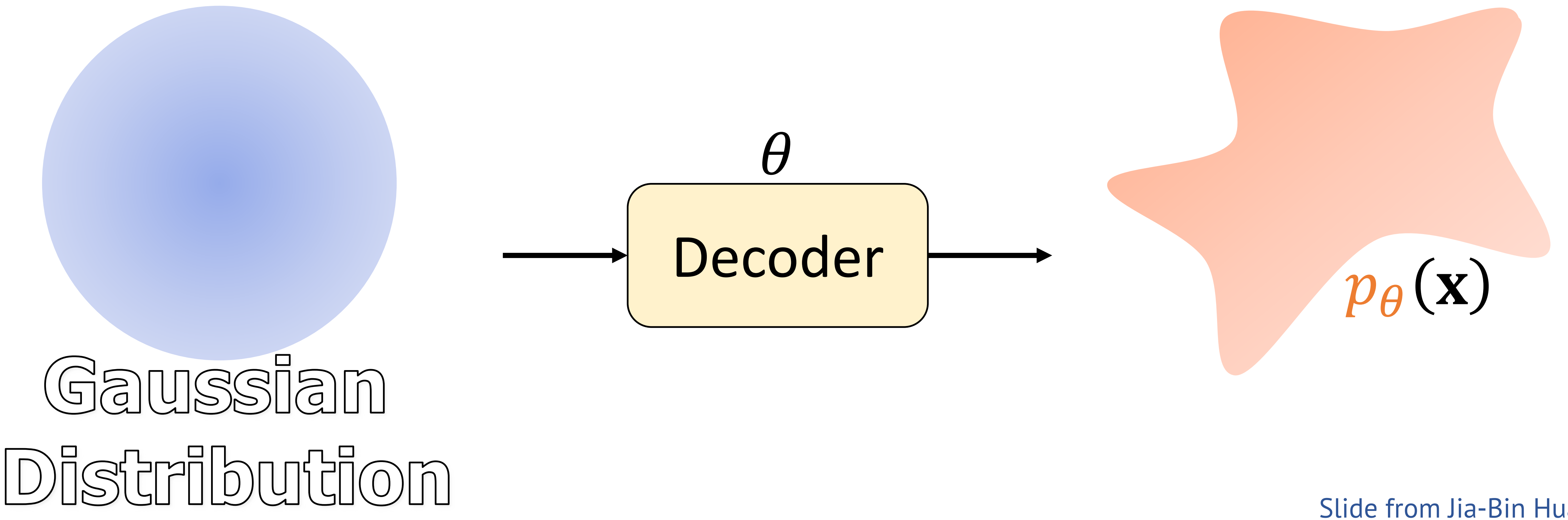


Slide from Jia-Bin Huang

Dataset: $\{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^m\} \sim p_{\text{data}}(\mathbf{x})$

Maximum Likelihood

$$\theta^* = \operatorname{argmax} \prod_{i=1}^{m} p_\theta(\mathbf{x}^i)$$

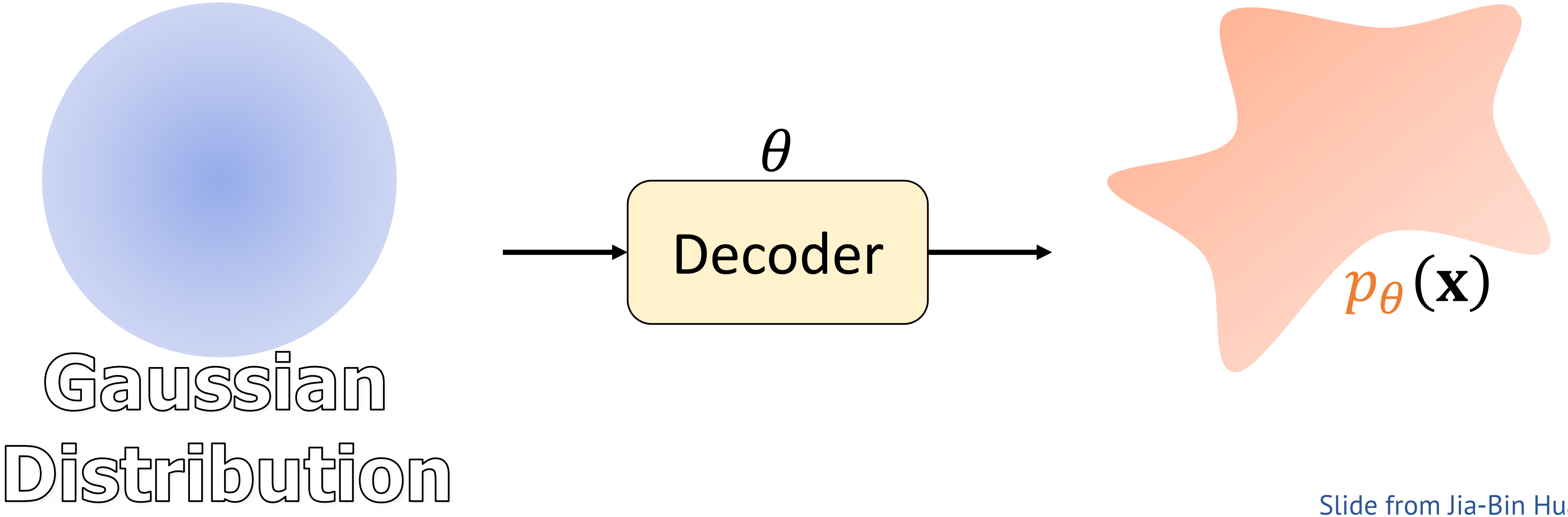


Slide from Jia-Bin Huang

Dataset: $\{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^m\} \sim p_{\text{data}}(\mathbf{x})$

Maximum Likelihood

$$\theta^* = \operatorname{argmax} \prod_{i=1}^{m} p_\theta(\mathbf{x}^i)$$

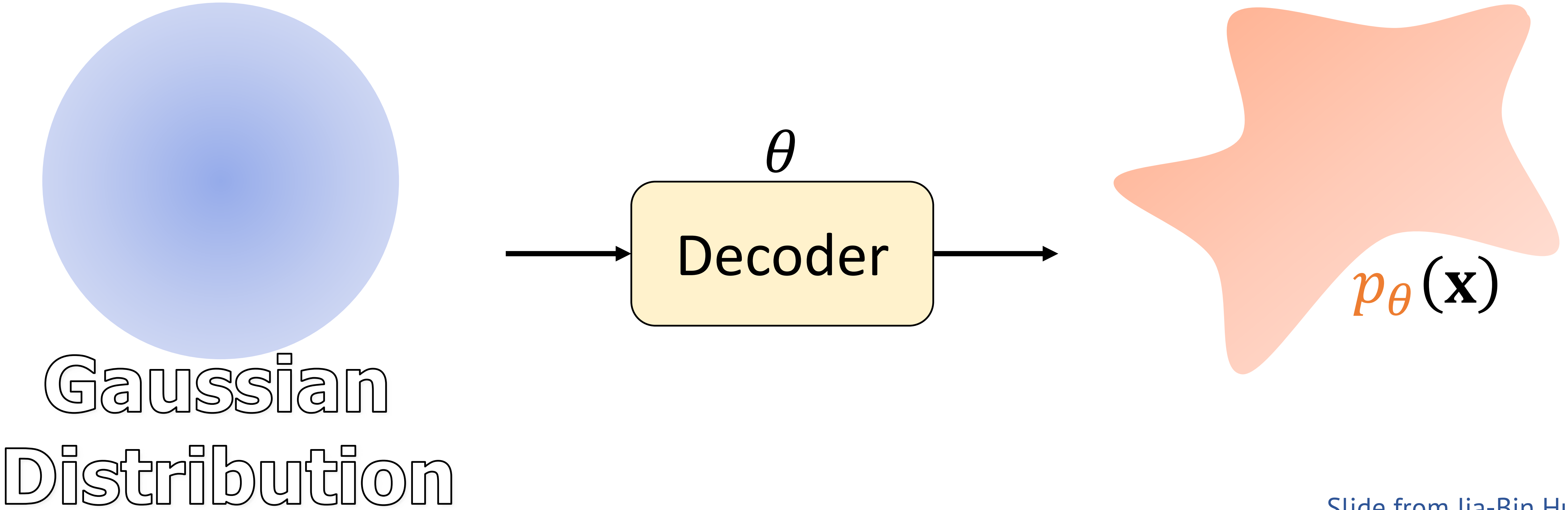


Slide from Jia-Bin Huang

Dataset: $\{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^m\} \sim p_{\text{data}}(\mathbf{x})$

Maximum Likelihood

$$\theta^* = \operatorname{argmax} \log \left(\prod_{i=1}^{m} p_\theta(\mathbf{x}^i) \right)$$

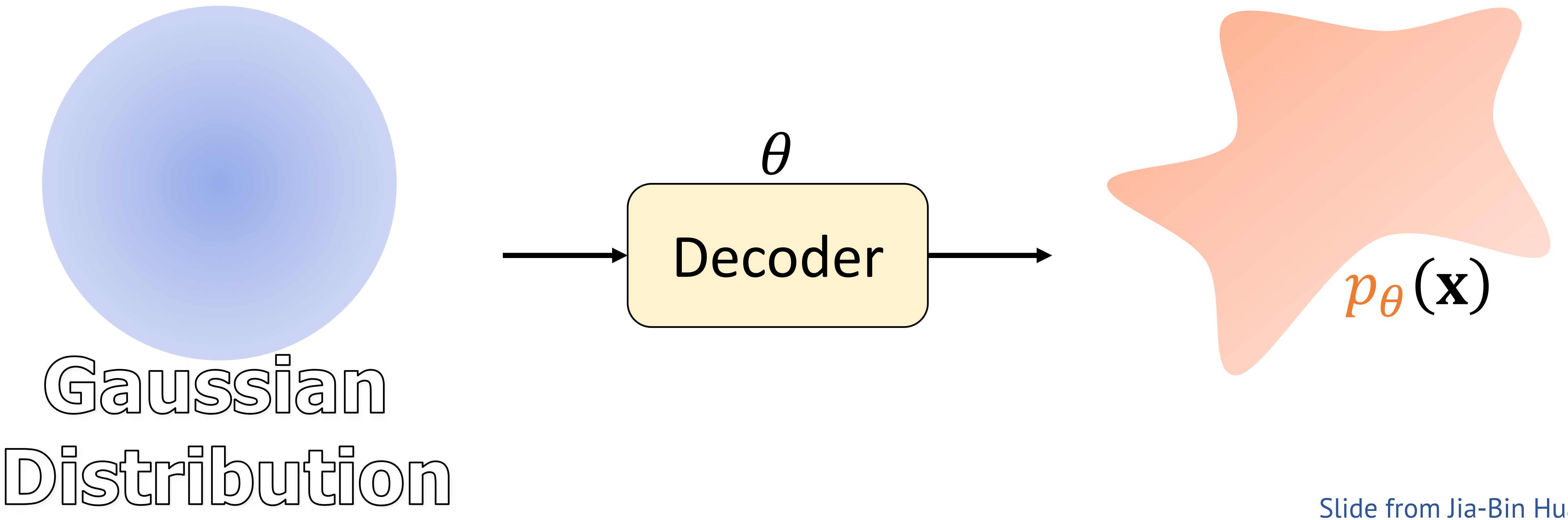


Slide from Jia-Bin Huang

Dataset: $\{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^m\} \sim p_{\text{data}}(\mathbf{x})$

Maximum Likelihood

$$\theta^* = \operatorname{argmax} \sum_{i=1}^{m} \log p_\theta(\mathbf{x}^i)$$

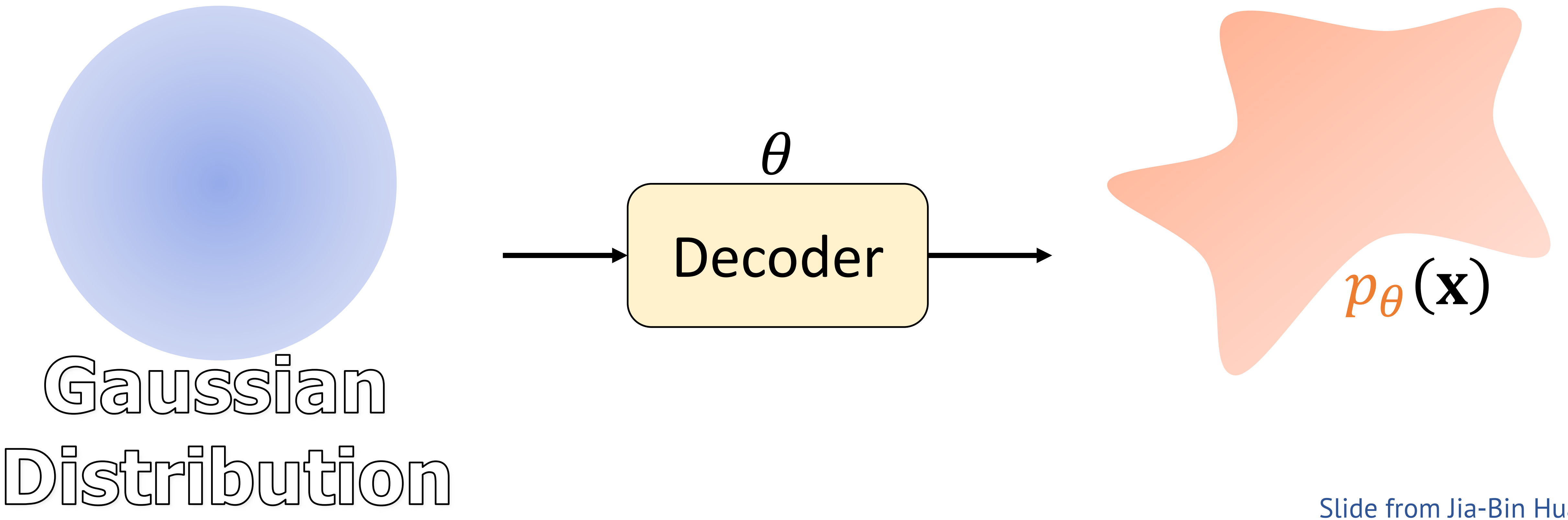


Slide from Jia-Bin Huang

Dataset: $\{\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^m\} \sim p_{\text{data}}(\mathbf{x})$

Maximum Likelihood

$$\theta^* \approx \operatorname{argmax} \; \mathbb{E}_{\mathbf{x} \sim p_{\text{data}}}[\log p_\theta(\mathbf{x})]$$

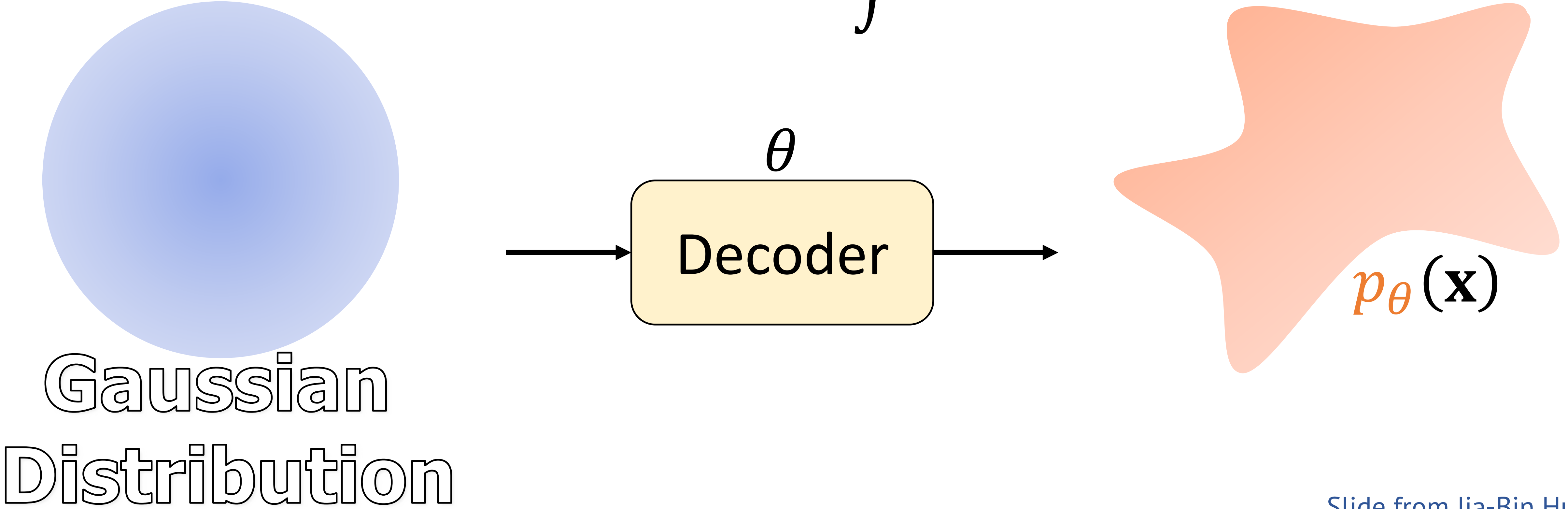


Slide from Jia-Bin Huang

Dataset: $\{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^m\} \sim p_{\text{data}}(\mathbf{x})$

Maximum Likelihood

$$\theta^* \approx \operatorname{argmax} \int p_{\text{data}}(\mathbf{x}) \log p_\theta(\mathbf{x}) d\mathbf{x} - \int p_{\text{data}}(\mathbf{x}) \log p_{\text{data}}(\mathbf{x}) d\mathbf{x}$$

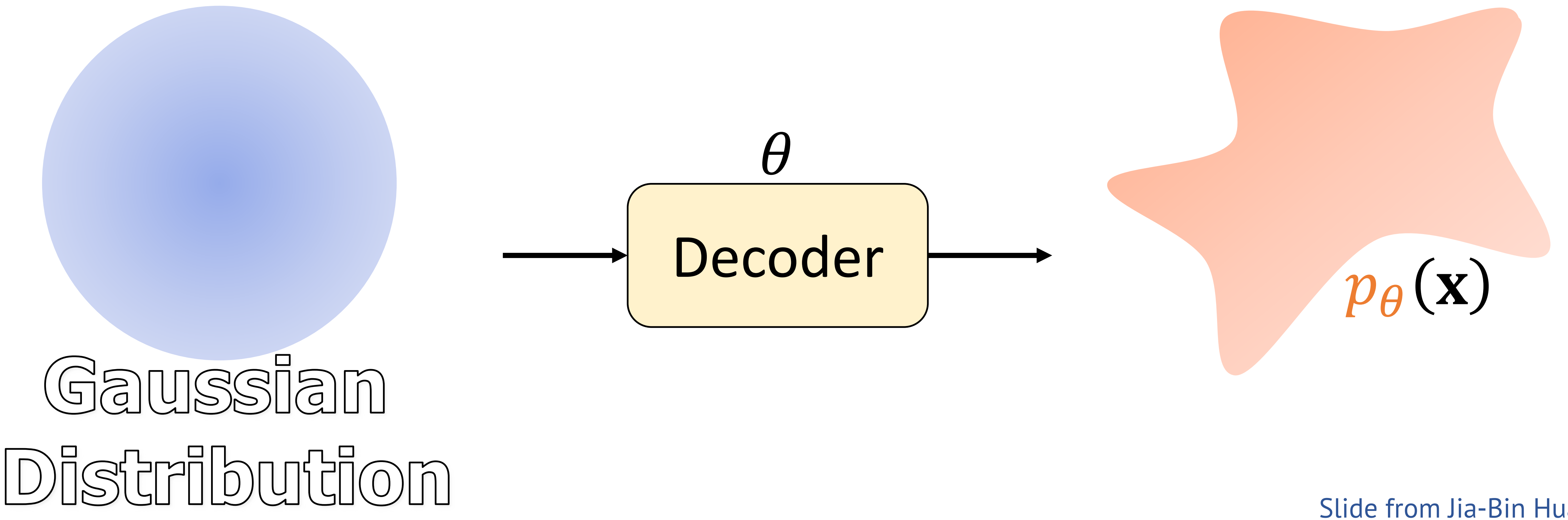


Slide from Jia-Bin Huang

Dataset: $\{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^m\} \sim p_{\text{data}}(\mathbf{x})$

Maximum Likelihood

$$\theta^* \approx \operatorname{argmax} \int p_{\text{data}}(\mathbf{x}) \log \frac{p_\theta(\mathbf{x})}{p_{\text{data}}(\mathbf{x})} \mathrm{d}\mathbf{x}$$

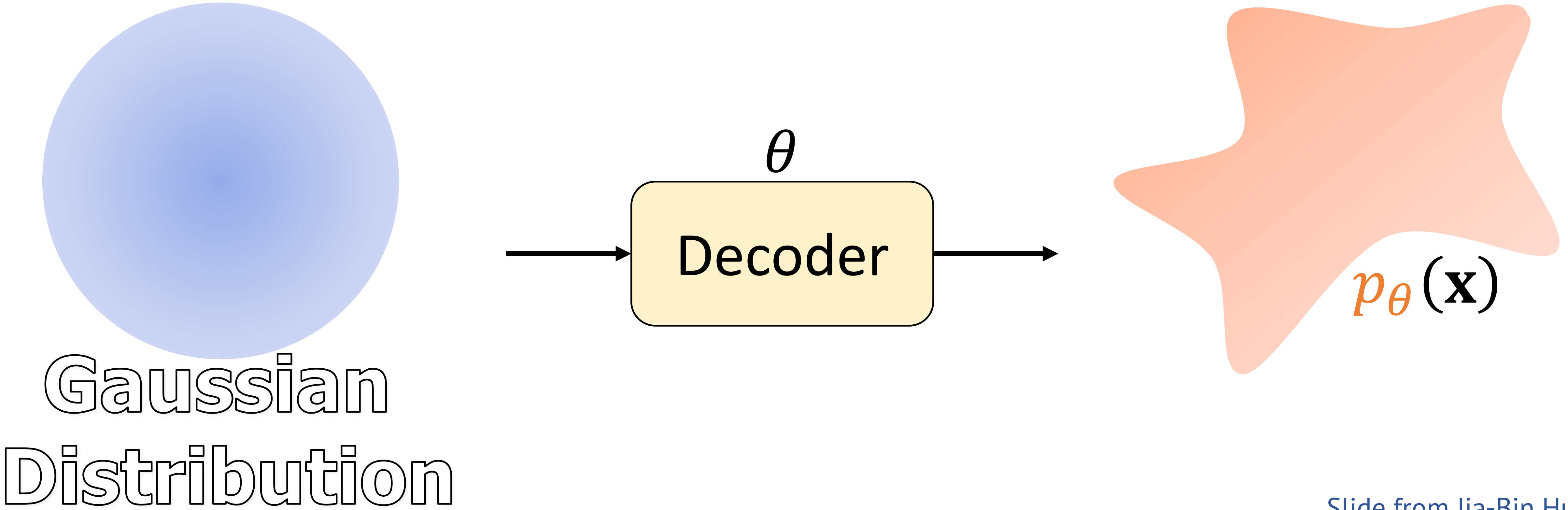


Dataset: $\{\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^m\} \sim p_{\text{data}}(\mathbf{x})$

Maximum Likelihood

$$\theta^* \approx \operatorname{argmin} \int p_{\text{data}}(\mathbf{x}) \log \frac{p_{\text{data}}(\mathbf{x})}{p_\theta(\mathbf{x})} d\mathbf{x}$$

$$= \operatorname{argmin} \mathrm{D}_{\mathrm{KL}}(p_{\text{data}} || p_\theta)$$

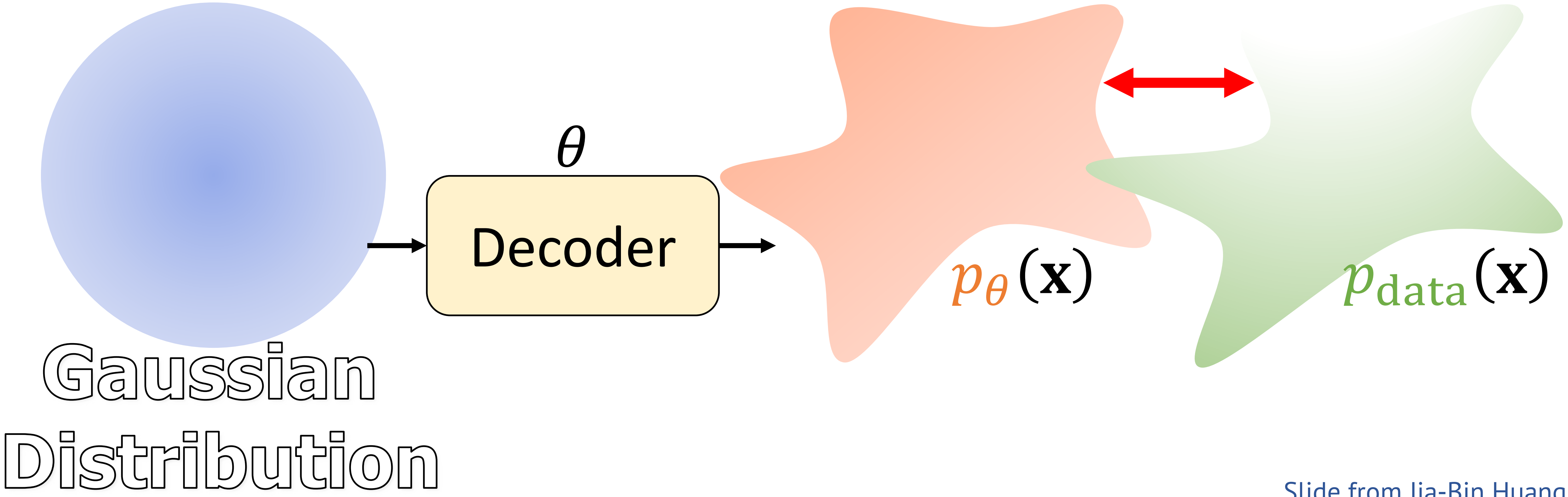


Slide from Jia-Bin Huang

Maximum Likelihood

$$\theta^* \approx \operatorname{argmin} \int p_{\text{data}}(\mathbf{x}) \log \frac{p_{\text{data}}(\mathbf{x})}{p_\theta(\mathbf{x})} d\mathbf{x}$$

$$= \operatorname{argmin} D_{\text{KL}}(p_{\text{data}} || p_\theta)$$



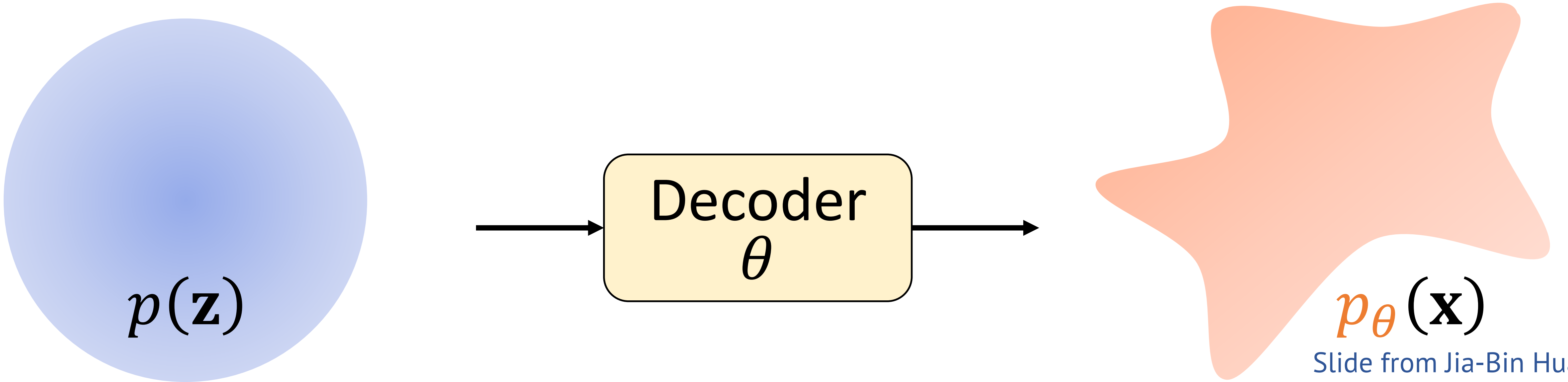
Slide from Jia-Bin Huang

$$\log p(\mathbf{x}) = \log \int p(\mathbf{x}, \boldsymbol{z})\, \mathrm{d}\mathbf{z}$$

Integrating out all latent variables $\mathbf{z}$

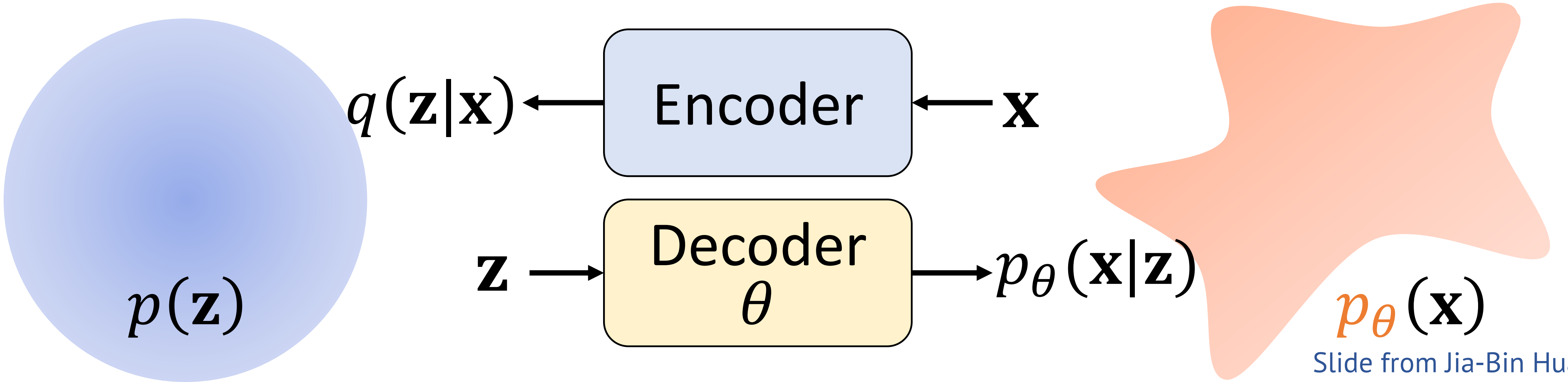
$$\log p(\mathbf{x}) = \log \frac{p(\mathbf{x}, \mathbf{z})}{p(\mathbf{z}|\mathbf{x})}$$

Assuming we know latent encoder $p(\mathbf{z}|\mathbf{x})$



Slide from Jia-Bin Huang

$$\log p(\mathbf{x}) = \log p(\mathbf{x}) \boxed{\int q(\mathbf{z}|\mathbf{x})\, \mathrm{d}\mathbf{z}}$$



Slide from Jia-Bin Huang

$$\log p(\mathbf{x}) = \log p(\mathbf{x}) \int q(\mathbf{z}|\mathbf{x})\, \mathrm{d}\mathbf{z}$$

Slide from Jia-Bin Huang

$$\log p(\mathbf{x}) = \int q(\mathbf{z}|\mathbf{x}) \log p(\mathbf{x}) \, d\mathbf{z}$$

$$\log p(\mathbf{x}) = \mathbb{E}_{q(\mathbf{z}|\mathbf{x})}[\log p(\mathbf{x})]$$

Slide from Jia-Bin Huang

$$\log p(\mathbf{x}) = \mathbb{E}_{q(\mathbf{z}|\mathbf{x})}\left[\log \frac{p(\mathbf{x},\mathbf{z})}{p(\mathbf{z}|\mathbf{x})}\right]$$

Slide from Jia-Bin Huang

$$\log p(\mathbf{x}) = \mathbb{E}_{q(\mathbf{z}|\mathbf{x})}\left[\log \frac{p(\mathbf{x}, \mathbf{z})}{p(\mathbf{z}|\mathbf{x})} \frac{q(\mathbf{z}|\mathbf{x})}{q(\mathbf{z}|\mathbf{x})}\right]$$

$$\log p(\mathbf{x}) = \mathbb{E}_{q(\mathbf{z}|\mathbf{x})}\left[\log \frac{p(\mathbf{x},\mathbf{z})}{q(\mathbf{z}|\mathbf{x})} \frac{q(\mathbf{z}|\mathbf{x})}{p(\mathbf{z}|\mathbf{x})}\right]$$

$$\log p(\mathbf{x}) = \mathbb{E}_{q(\mathbf{z}|\mathbf{x})}\left[\log \frac{p(\mathbf{x}, \mathbf{z})}{q(\mathbf{z}|\mathbf{x})}\right] + \mathbb{E}_{q(\mathbf{z}|\mathbf{x})}\left[\log \frac{q(\mathbf{z}|\mathbf{x})}{p(\mathbf{z}|\mathbf{x})}\right]$$

Slide from Jia-Bin Huang

$$\log p(\mathbf{x}) = \mathbb{E}_{q(\mathbf{z}|\mathbf{x})}\left[\log \frac{p(\mathbf{x},\mathbf{z})}{q(\mathbf{z}|\mathbf{x})}\right] + \mathrm{D}_{\mathrm{KL}}\left(q(\mathbf{z}|\mathbf{x}) \,||\, p(\mathbf{z}|\mathbf{x})\right)$$

Slide from Jia-Bin Huang

$$\log p(\mathbf{x}) = \mathbb{E}_{q(\mathbf{z}|\mathbf{x})}\left[\log \frac{p(\mathbf{x},\mathbf{z})}{q(\mathbf{z}|\mathbf{x})}\right] + \boxed{\mathrm{D_{KL}}\left(q(\mathbf{z}|\mathbf{x}) \,||\, p(\mathbf{z}|\mathbf{x})\right)}$$

$$\geq 0$$

Slide from Jia-Bin Huang

$$\log p(\mathbf{x}) \geq \mathbb{E}_{q(\mathbf{z}|\mathbf{x})}\left[\log \frac{p(\mathbf{x}, \mathbf{z})}{q(\mathbf{z}|\mathbf{x})}\right]$$

Slide from Jia-Bin Huang

$$\log p(\mathbf{x}) \geq \mathbb{E}_{q(\mathbf{z}|\mathbf{x})}\left[\log \frac{p(\mathbf{x}, \mathbf{z})}{q(\mathbf{z}|\mathbf{x})}\right]$$

Evidence **Evidence Lower Bound**

Slide from Jia-Bin Huang

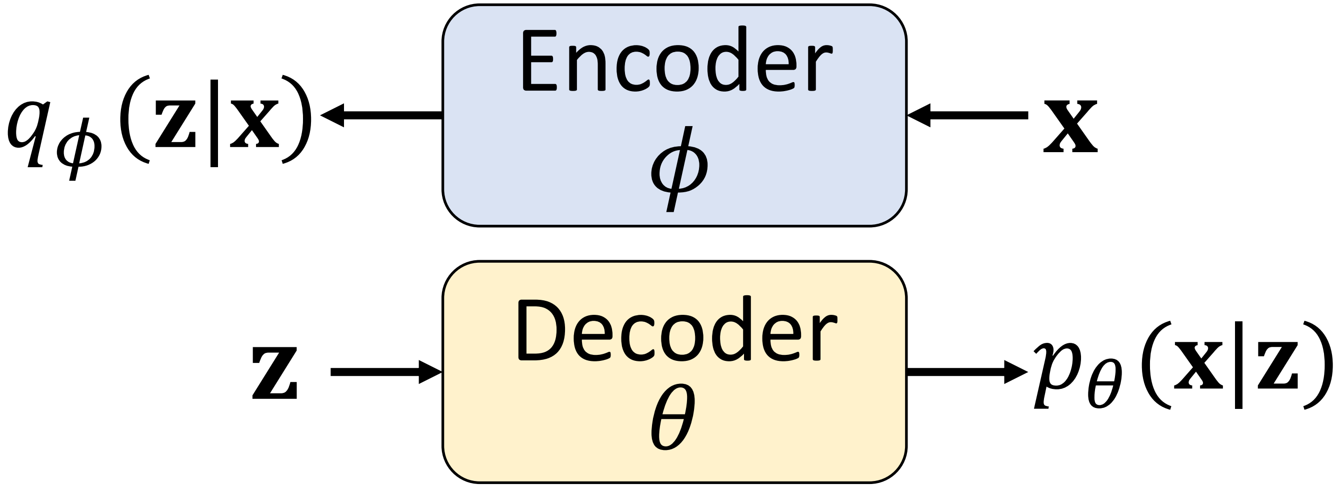
$$\log p(\mathbf{x}) \geq \mathbb{E}_{q(\mathbf{z}|\mathbf{x})}\left[\log \frac{p(\mathbf{x},\mathbf{z})}{q(\mathbf{z}|\mathbf{x})}\right]$$

Evidence **Evidence Lower Bound (ELBO)**

Slide from Jia-Bin Huang

$$\log p(\mathbf{x}) \geq \mathbb{E}_{q(\mathbf{z}|\mathbf{x})}\left[\log \frac{p(\mathbf{x}, \mathbf{z})}{q(\mathbf{z}|\mathbf{x})}\right]$$

Evidence **Evidence Lower Bound (ELBO)**



Slide from Jia-Bin Huang

$$\log p(\mathbf{x}) \geq \mathbb{E}_{q(\mathbf{z}|\mathbf{x})}\left[\log \frac{p(\mathbf{x},\mathbf{z})}{q(\mathbf{z}|\mathbf{x})}\right]$$ **(ELBO)**

$$\begin{aligned}
\mathbb{E}_{q_\phi(\boldsymbol{z}|\boldsymbol{x})}\left[\log \frac{p(\boldsymbol{x},\boldsymbol{z})}{q_\phi(\boldsymbol{z}\mid\boldsymbol{x})}\right] &= \mathbb{E}_{q_\phi(\boldsymbol{z}|\boldsymbol{x})}\left[\log \frac{p_{\boldsymbol{\theta}}(\boldsymbol{x}\mid\boldsymbol{z})p(\boldsymbol{z})}{q_\phi(\boldsymbol{z}\mid\boldsymbol{x})}\right] \\
&= \mathbb{E}_{q_\phi(\boldsymbol{z}|\boldsymbol{x})}[\log p_{\boldsymbol{\theta}}(\boldsymbol{x}\mid\boldsymbol{z})] + \mathbb{E}_{q_\phi(\boldsymbol{z}|\boldsymbol{x})}\left[\log \frac{p(\boldsymbol{z})}{q_\phi(\boldsymbol{z}\mid\boldsymbol{x})}\right] \\
&= \underbrace{\mathbb{E}_{q_\phi(\boldsymbol{z}|\boldsymbol{x})}[\log p_{\boldsymbol{\theta}}(\boldsymbol{x}\mid\boldsymbol{z})]}_{\text{Reconstruction term}} - \underbrace{D_{\mathrm{KL}}\left(q_\phi(\boldsymbol{z}\mid\boldsymbol{x}) \parallel p(\boldsymbol{z})\right)}_{\text{Regularization term}}
\end{aligned}$$

$$\|x - \hat{x}\|_2 = \|x - DE(EN(z))\|_2 \qquad -\frac{1}{2}\sum_{j=0}^{k-1}\left(\sigma_j + \mu_j^2 - 1 - \log \sigma_j\right)$$

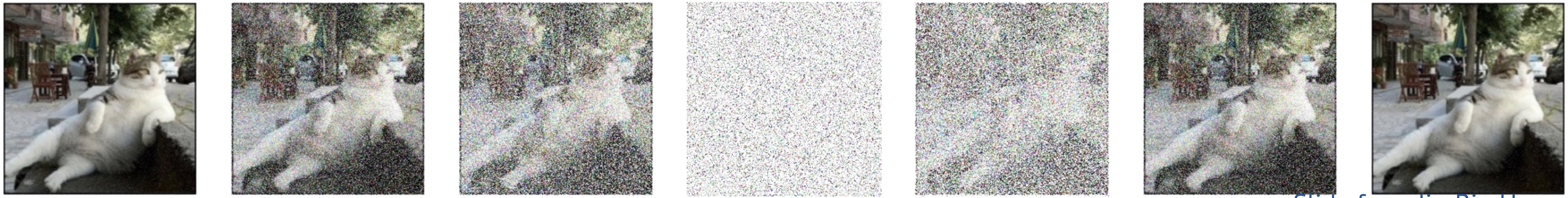
Slide from Jia-Bin Huang

Variational Autoencoder (VAE)

$\mathbf{x} \rightarrow$ Encoder ϕ $\rightarrow q_\phi(\mathbf{z}|\mathbf{x}) \xrightarrow{\text{sample}} \mathbf{z} \rightarrow$ Decoder θ $\rightarrow p_\theta(\mathbf{x}|\mathbf{z}) \xrightarrow{\text{sample}} \hat{\mathbf{x}}$

Diffusion models

$\mathbf{x}_0 \xrightarrow{\text{Encoding}} \mathbf{x}_1 \xrightarrow{\text{Encoding}} \mathbf{x}_2 \cdots \xrightarrow{\text{Encoding}} \mathbf{x}_T \xrightarrow{\text{Decoding } \theta} \mathbf{x}_{T-1} \cdots \xrightarrow{\text{Decoding } \theta} \hat{\mathbf{x}}_0$



Slide from Jia-Bin Huang

Variational Autoencoder (VAE)

$$\mathbf{x} \rightarrow \boxed{\text{Encoder } \phi} \rightarrow q_\phi(\mathbf{z}|\mathbf{x}) \xrightarrow{\text{sample}} \mathbf{z} \rightarrow \boxed{\text{Decoder } \theta} \rightarrow p_\theta(\mathbf{x}|\mathbf{z}) \xrightarrow{\text{sample}} \hat{\mathbf{x}}$$

Observed $\mathbf{x}$

Latent $\mathbf{z}$

Maximize $\mathbb{E}_{q(\mathbf{z}|\mathbf{x})}\left[\log \frac{p(\mathbf{x},\mathbf{z})}{q(\mathbf{z}|\mathbf{x})}\right]$

Diffusion models

$$\mathbf{x}_0 \xrightarrow{\text{Encoding}} \mathbf{x}_1 \xrightarrow{\text{Encoding}} \mathbf{x}_2 \cdots \xrightarrow{\text{Encoding}} \mathbf{x}_T \xrightarrow{\text{Decoding } \theta} \mathbf{x}_{T-1} \quad \cdots \quad \xrightarrow{\text{Decoding } \theta} \hat{\mathbf{x}}_0$$

Observed $\mathbf{x}_0$

Latent $\mathbf{x}_1, \dots, \mathbf{x}_T$

Maximize $\mathbb{E}_{q(\mathbf{x}_1:\mathbf{x}_T|\mathbf{x}_0)}\left[\log \frac{p(\mathbf{x}_0:\mathbf{x}_T)}{q(\mathbf{x}_1:\mathbf{x}_T|\mathbf{x}_0)}\right]$

Slide from Jia-Bin Huang

$$\mathbf{x}_0 \xrightarrow{\text{Encoding}} \mathbf{x}_1 \xrightarrow{\text{Encoding}} \mathbf{x}_2 \cdots \xrightarrow{\text{Encoding}} \mathbf{x}_T \xrightarrow{\text{Decoding } \theta} \mathbf{x}_{T-1} \quad \cdots \quad \xrightarrow{\text{Decoding } \theta} \hat{\mathbf{x}}_0$$

Observed $\mathbf{x}_0$

Latent $\mathbf{x}_1, \ldots, \mathbf{x}_T$

Maximize $\mathbb{E}_{q(\mathbf{x}_1:\mathbf{x}_T|\mathbf{x}_0)}\left[\log \frac{p(\mathbf{x}_0:\mathbf{x}_T)}{q(\mathbf{x}_1:\mathbf{x}_T|\mathbf{x}_0)}\right]$

Slide from Jia-Bin Huang

$$\mathbf{x}_0 \xrightarrow{\text{Encoding}} \mathbf{x}_1 \xrightarrow{\text{Encoding}} \mathbf{x}_2 \cdots \xrightarrow{\text{Encoding}} \mathbf{x}_T \xrightarrow{\text{Decoding } \theta} \mathbf{x}_{T-1} \quad \cdots \quad \xrightarrow{\text{Decoding } \theta} \hat{\mathbf{x}}_0$$

$$\text{Maximize} \quad \mathbb{E}_{q(\mathbf{x}_1:\mathbf{x}_T|\mathbf{x}_0)}\left[\log \frac{p(\mathbf{x}_0:\mathbf{x}_T)}{q(\mathbf{x}_1:\mathbf{x}_T|\mathbf{x}_0)}\right]$$

Slide from Jia-Bin Huang

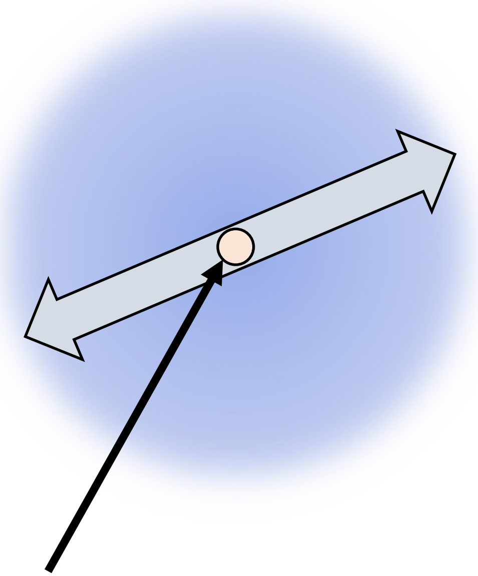
$$\mathbf{x}_0 \xrightarrow{\text{Encoding}} \mathbf{x}_1 \xrightarrow{\text{Encoding}} \mathbf{x}_2 \cdots \xrightarrow{\text{Encoding}} \mathbf{x}_T \xrightarrow{\text{Decoding } \theta} \mathbf{x}_{T-1} \quad \cdots \quad \xrightarrow{\text{Decoding } \theta} \hat{\mathbf{x}}_0$$

$$\text{Maximize} \quad \mathbb{E}_{q(\mathbf{x}_1:\mathbf{x}_T|\mathbf{x}_0)}\left[\log \frac{p(\mathbf{x}_0:\mathbf{x}_T)}{q(\mathbf{x}_1:\mathbf{x}_T|\mathbf{x}_0)}\right]$$

$$\mathbf{x}_0 \xrightarrow{\text{Encoding}} \mathbf{x}_1 \xrightarrow{\text{Encoding}} \mathbf{x}_2 \cdots \xrightarrow{\text{Encoding}} \mathbf{x}_T \xrightarrow{\text{Decoding } \theta} \mathbf{x}_{T-1} \quad \cdots \quad \xrightarrow{\text{Decoding } \theta} \hat{\mathbf{x}}_0$$

Maximize $\mathbb{E}_{q(\mathbf{x}_1:\mathbf{x}_T|\mathbf{x}_0)}\left[\log \frac{p(\mathbf{x}_0:\mathbf{x}_T)}{q(\mathbf{x}_1:\mathbf{x}_T|\mathbf{x}_0)}\right]$

$$q(\mathbf{x}_1:\mathbf{x}_T|\mathbf{x}_0) = \prod_{t=1}^{T} \underbrace{q(\mathbf{x}_t|\mathbf{x}_{t-1})}_{\text{Encoding}}$$



$$\underbrace{q(\mathbf{x}_t|\mathbf{x}_{t-1})}_{\text{Encoding}} = \mathcal{N}(\mathbf{x}_t; \underbrace{\sqrt{\alpha_t}\mathbf{x}_{t-1}}_{\text{Mean}}, \underbrace{(1-\alpha_t)\mathbf{I}}_{\text{Variance}})$$

Slide from Jia-Bin Huang

$$\mathbf{x}_0 \xrightarrow{\text{Encoding}} \mathbf{x}_1 \xrightarrow{\text{Encoding}} \mathbf{x}_2 \cdots \xrightarrow{\text{Encoding}} \mathbf{x}_T \xrightarrow{\text{Decoding } \theta} \mathbf{x}_{T-1} \quad \cdots \quad \xrightarrow{\text{Decoding } \theta} \hat{\mathbf{x}}_0$$

$$\mathbb{E}_{q(\mathbf{x}_1:\mathbf{x}_T|\mathbf{x}_0)}\left[\log \frac{p(\mathbf{x}_0:\mathbf{x}_T)}{q(\mathbf{x}_1:\mathbf{x}_T|\mathbf{x}_0)}\right]$$

1 $-\mathrm{D}_{\mathrm{KL}}(q(\mathbf{x}_T|\mathbf{x}_0)||p(\mathbf{x}_T))$ **Prior matching**

2 $\mathbb{E}_{q(\mathbf{x}_1|\mathbf{x}_0)}[\log p_\theta(\mathbf{x}_0|\mathbf{x}_1)]$ **Reconstruction**

3 $-\sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}[\mathrm{D}_{\mathrm{KL}}(q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0)||p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t))]$ **Denoising matching**

Slide from Jia-Bin Huang

$$\mathbf{x}_0 \xrightarrow{\text{Encoding}} \mathbf{x}_1 \xrightarrow{\text{Encoding}} \mathbf{x}_2 \cdots \xrightarrow{\text{Encoding}} \mathbf{x}_T \xrightarrow{\text{Decoding } \theta} \mathbf{x}_{T-1} \quad \cdots \quad \xrightarrow{\text{Decoding } \theta} \hat{\mathbf{x}}_0$$

$$\mathbb{E}_{q(\mathbf{x}_1:\mathbf{x}_T|\mathbf{x}_0)}\left[\log \frac{p(\mathbf{x}_0:\mathbf{x}_T)}{q(\mathbf{x}_1:\mathbf{x}_T|\mathbf{x}_0)}\right]$$

1 $\boxed{-\mathrm{D}_{\mathrm{KL}}(q(\mathbf{x}_T|\mathbf{x}_0)||p(\mathbf{x}_T))}$ **Prior matching**

2 $\boxed{\mathbb{E}_{q(\mathbf{x}_1|\mathbf{x}_0)}[\log p_\theta(\mathbf{x}_0|\mathbf{x}_1)]}$ **Reconstruction**

3 $-\sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}[\mathrm{D}_{\mathrm{KL}}(q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0)||p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t))]$ **Denoising matching**

Slide from Jia-Bin Huang

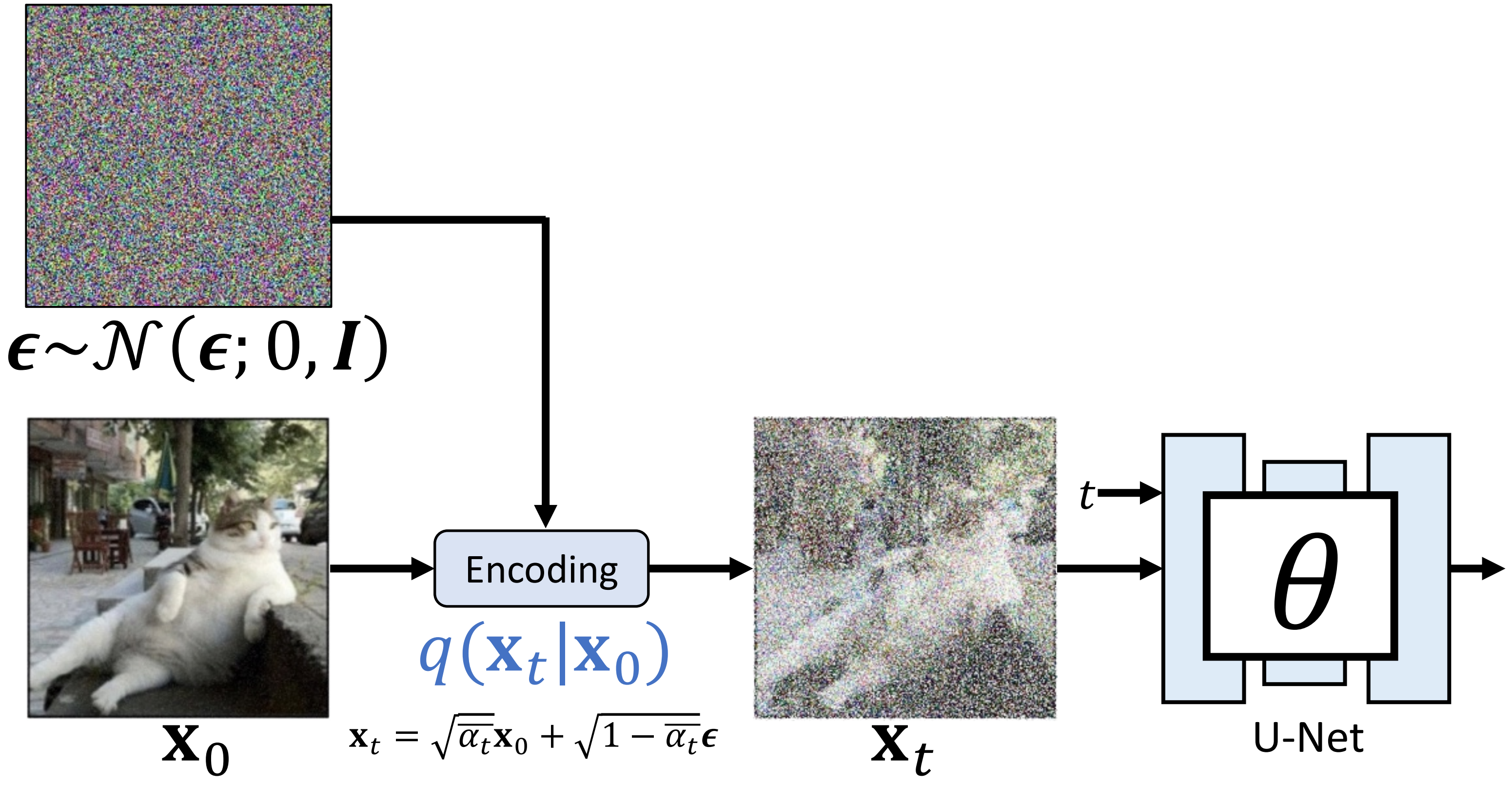
$$\mathbf{x}_0 \xrightarrow{\text{Encoding}} \mathbf{x}_1 \xrightarrow{\text{Encoding}} \mathbf{x}_2 \cdots \xrightarrow{\text{Encoding}} \mathbf{x}_T \xrightarrow{\text{Decoding } \theta} \mathbf{x}_{T-1} \quad \cdots \quad \xrightarrow{\text{Decoding } \theta} \hat{\mathbf{x}}_0$$

$$\sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}\left[\frac{1}{2\boldsymbol{\sigma}_q^2(t)} \left[|| \mu_\theta(\mathbf{x}_t, t) - \mu_q(\mathbf{x}_t, \mathbf{x}_0) ||_2^2 \right] \right]$$

Slide from Jia-Bin Huang

$$\mathbf{x}_0 \xrightarrow{\text{Encoding}} \mathbf{x}_1 \xrightarrow{\text{Encoding}} \mathbf{x}_2 \cdots \xrightarrow{\text{Encoding}} \mathbf{x}_T \xrightarrow{\text{Decoding } \theta} \mathbf{x}_{T-1} \quad \cdots \quad \xrightarrow{\text{Decoding } \theta} \hat{\mathbf{x}}_0$$

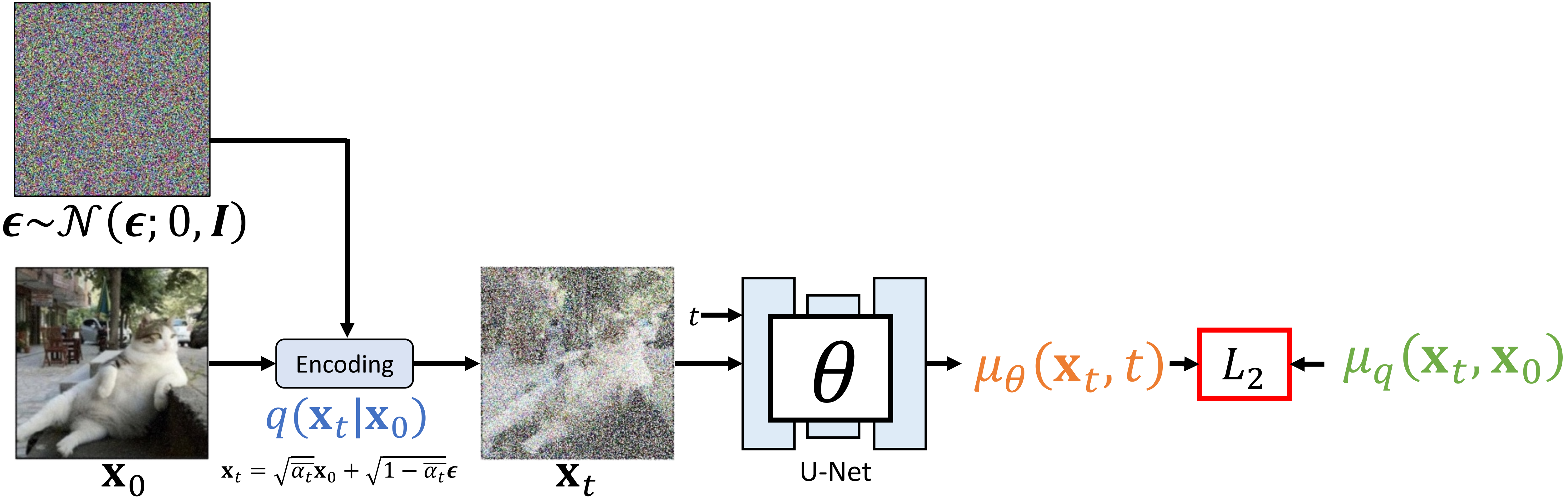
$$\sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}\left[\frac{1}{2\boldsymbol{\sigma}_q^2(t)} \left[\| \mu_\theta(\mathbf{x}_t, t) - \mu_q(\mathbf{x}_t, \mathbf{x}_0) \|_2^2 \right] \right]$$



Slide from Jia-Bin Huang

$$\mathbf{x}_0 \xrightarrow{\text{Encoding}} \mathbf{x}_1 \xrightarrow{\text{Encoding}} \mathbf{x}_2 \cdots \xrightarrow{\text{Encoding}} \mathbf{x}_T \xrightarrow{\text{Decoding } \theta} \mathbf{x}_{T-1} \quad \cdots \quad \xrightarrow{\text{Decoding } \theta} \hat{\mathbf{x}}_0$$

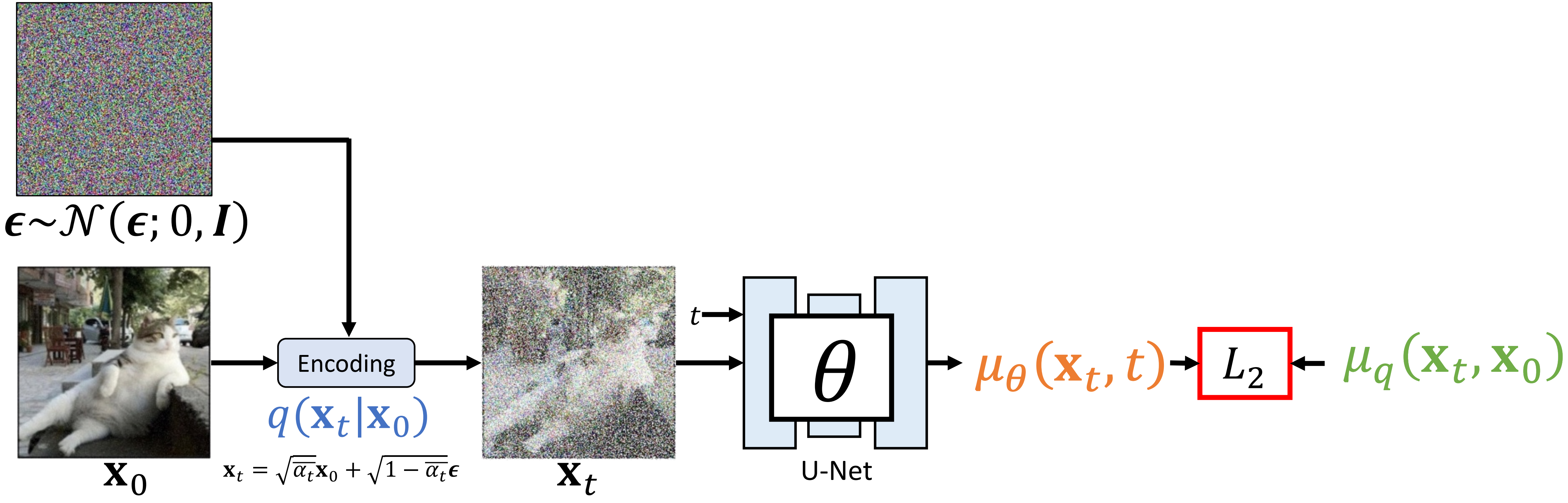
$$\sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}\left[\frac{1}{2\boldsymbol{\sigma}_q^2(t)} \left[\| \mu_\theta(\mathbf{x}_t, t) - \mu_q(\mathbf{x}_t, \mathbf{x}_0) \|_2^2 \right] \right]$$



Slide from Jia-Bin Huang

$$\mathbf{x}_0 \xrightarrow{\text{Encoding}} \mathbf{x}_1 \xrightarrow{\text{Encoding}} \mathbf{x}_2 \cdots \xrightarrow{\text{Encoding}} \mathbf{x}_T \xrightarrow{\text{Decoding } \theta} \mathbf{x}_{T-1} \quad \cdots \quad \xrightarrow{\text{Decoding } \theta} \hat{\mathbf{x}}_0$$

$$\sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}\left[\frac{1}{2\boldsymbol{\sigma}_q^2(t)} \left[\| \mu_\theta(\mathbf{x}_t, t) - \mu_q(\mathbf{x}_t, \mathbf{x}_0) \|_2^2 \right] \right]$$

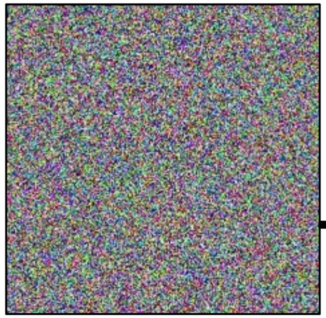


Slide from Jia-Bin Huang

$$\mathbf{x}_0 \xrightarrow{\text{Encoding}} \mathbf{x}_1 \xrightarrow{\text{Encoding}} \mathbf{x}_2 \cdots \xrightarrow{\text{Encoding}} \mathbf{x}_T \xrightarrow{\text{Decoding } \theta} \mathbf{x}_{T-1} \quad \cdots \quad \xrightarrow{\text{Decoding } \theta} \hat{\mathbf{x}}_0$$

$$\sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}\left[\frac{1}{2\boldsymbol{\sigma}_q^2(t)} \left[|| \mu_\theta(\mathbf{x}_t, t) - \mu_q(\mathbf{x}_t, \mathbf{x}_0) ||_2^2 \right] \right]$$

$$\mu_q(\mathbf{x}_t, \mathbf{x}_0) = \frac{\sqrt{\alpha_t}(1-\overline{\alpha}_{t-1})}{1-\overline{\alpha}_t}\mathbf{x}_t + \frac{\sqrt{\overline{\alpha}_{t-1}}(1-\alpha_t)}{1-\overline{\alpha}_t}\mathbf{x}_0$$



$\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{\epsilon}; 0, \boldsymbol{I})$

$\mathbf{x}_0$

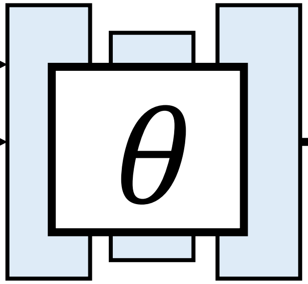
Encoding

$q(\mathbf{x}_t|\mathbf{x}_0)$

$\mathbf{x}_t = \sqrt{\overline{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\overline{\alpha}_t}\boldsymbol{\epsilon}$

$\mathbf{x}_t$

t



U-Net

$\mu_\theta(\mathbf{x}_t, t) \rightarrow \boxed{L_2} \leftarrow \mu_q(\mathbf{x}_t, \mathbf{x}_0)$

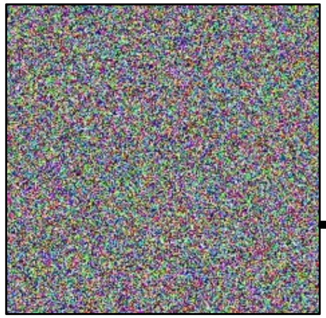
Slide from Jia-Bin Huang

$$\mathbf{x}_0 \xrightarrow{\text{Encoding}} \mathbf{x}_1 \xrightarrow{\text{Encoding}} \mathbf{x}_2 \cdots \xrightarrow{\text{Encoding}} \mathbf{x}_T \xrightarrow{\text{Decoding } \theta} \mathbf{x}_{T-1} \quad \cdots \quad \xrightarrow{\text{Decoding } \theta} \hat{\mathbf{x}}_0$$

$$\sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}\left[\frac{1}{2\boldsymbol{\sigma}_q^2(t)} \left[\left\| \mu_\theta(\mathbf{x}_t, t) - \mu_q(\mathbf{x}_t, \mathbf{x}_0) \right\|_2^2 \right] \right]$$

$$\mu_q(\mathbf{x}_t, \mathbf{x}_0) = \frac{\sqrt{\alpha_t}(1-\overline{\alpha}_{t-1})}{1-\overline{\alpha}_t}\mathbf{x}_t + \frac{\sqrt{\overline{\alpha}_{t-1}}(1-\alpha_t)}{1-\overline{\alpha}_t}\mathbf{x}_0$$

$$\mu_\theta(\mathbf{x}_t, t) = \frac{\sqrt{\alpha_t}(1-\overline{\alpha}_{t-1})}{1-\overline{\alpha}_t}\mathbf{x}_t + \frac{\sqrt{\overline{\alpha}_{t-1}}(1-\alpha_t)}{1-\overline{\alpha}_t}\hat{\mathbf{x}}_\theta(\mathbf{x}_t, t)$$



$\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{\epsilon}; 0, \boldsymbol{I})$

$\mathbf{x}_0$

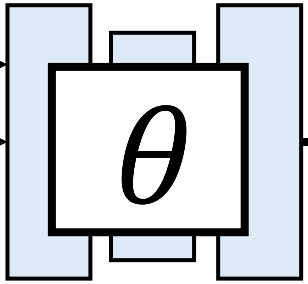
Encoding

$q(\mathbf{x}_t|\mathbf{x}_0)$

$\mathbf{x}_t = \sqrt{\overline{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\overline{\alpha}_t}\boldsymbol{\epsilon}$

$\mathbf{x}_t$

t



U-Net

$\mu_\theta(\mathbf{x}_t, t) \rightarrow \boxed{L_2} \leftarrow \mu_q(\mathbf{x}_t, \mathbf{x}_0)$

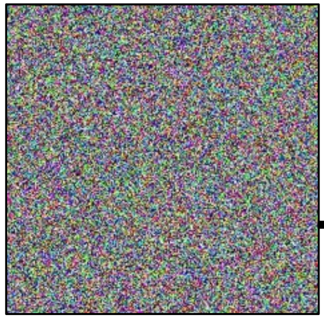
Slide from Jia-Bin Huang

$$\mathbf{x}_0 \xrightarrow{\text{Encoding}} \mathbf{x}_1 \xrightarrow{\text{Encoding}} \mathbf{x}_2 \cdots \xrightarrow{\text{Encoding}} \mathbf{x}_T \xrightarrow{\text{Decoding } \theta} \mathbf{x}_{T-1} \quad \cdots \quad \xrightarrow{\text{Decoding } \theta} \hat{\mathbf{x}}_0$$

$$\sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}\left[\frac{1}{2\boldsymbol{\sigma}_q^2(t)} \left[|| \mu_\theta(\mathbf{x}_t, t) - \mu_q(\mathbf{x}_t, \mathbf{x}_0) ||_2^2 \right] \right]$$

$$\mu_q(\mathbf{x}_t, \mathbf{x}_0) = \frac{\sqrt{\alpha_t}(1-\overline{\alpha}_{t-1})}{1-\overline{\alpha}_t}\mathbf{x}_t + \frac{\sqrt{\overline{\alpha}_{t-1}}(1-\alpha_t)}{1-\overline{\alpha}_t}\mathbf{x}_0$$

$$\mu_\theta(\mathbf{x}_t, t) = \frac{\sqrt{\alpha_t}(1-\overline{\alpha}_{t-1})}{1-\overline{\alpha}_t}\mathbf{x}_t + \frac{\sqrt{\overline{\alpha}_{t-1}}(1-\alpha_t)}{1-\overline{\alpha}_t}\hat{\mathbf{x}}_\theta(\mathbf{x}_t, t)$$



$\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{\epsilon}; 0, \boldsymbol{I})$

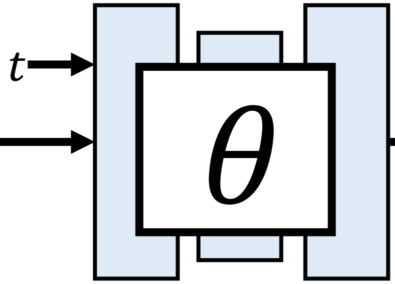
$\mathbf{x}_0$

Encoding

$q(\mathbf{x}_t|\mathbf{x}_0)$

$\mathbf{x}_t = \sqrt{\overline{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\overline{\alpha}_t}\boldsymbol{\epsilon}$

$\mathbf{x}_t$



U-Net

$\mu_\theta(\mathbf{x}_t, t) \rightarrow \boxed{L_2} \leftarrow \mu_q(\mathbf{x}_t, \mathbf{x}_0)$

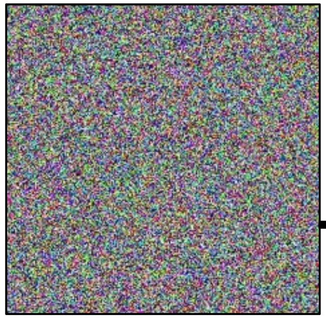
Slide from Jia-Bin Huang

$$\mathbf{x}_0 \xrightarrow{\text{Encoding}} \mathbf{x}_1 \xrightarrow{\text{Encoding}} \mathbf{x}_2 \cdots \xrightarrow{\text{Encoding}} \mathbf{x}_T \xrightarrow{\text{Decoding } \theta} \mathbf{x}_{T-1} \quad \cdots \quad \xrightarrow{\text{Decoding } \theta} \hat{\mathbf{x}}_0$$

$$\sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}\left[\; w(t) \; \left[\|\, \hat{\mathbf{x}}_\theta(\mathbf{x}_t, t) \; - \; \mathbf{x}_0 \; \|_2^2 \right] \right]$$

$$\mu_q(\mathbf{x}_t, \mathbf{x}_0) = \frac{\sqrt{\alpha_t}(1-\overline{\alpha}_{t-1})}{1-\overline{\alpha}_t}\mathbf{x}_t + \frac{\sqrt{\overline{\alpha}_{t-1}}(1-\alpha_t)}{1-\overline{\alpha}_t}\mathbf{x}_0$$

$$\mu_\theta(\mathbf{x}_t, t) = \frac{\sqrt{\alpha_t}(1-\overline{\alpha}_{t-1})}{1-\overline{\alpha}_t}\mathbf{x}_t + \frac{\sqrt{\overline{\alpha}_{t-1}}(1-\alpha_t)}{1-\overline{\alpha}_t}\hat{\mathbf{x}}_\theta(\mathbf{x}_t, t)$$



$\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{\epsilon}; 0, \boldsymbol{I})$

$\mathbf{x}_0$

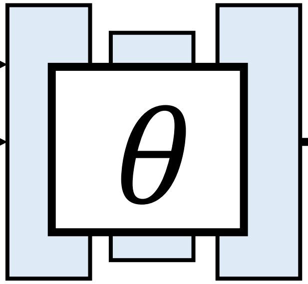
Encoding

$q(\mathbf{x}_t|\mathbf{x}_0)$

$\mathbf{x}_t = \sqrt{\overline{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\overline{\alpha}_t}\boldsymbol{\epsilon}$

$\mathbf{x}_t$

t



U-Net

$\mu_\theta(\mathbf{x}_t, t) \rightarrow \boxed{L_2} \leftarrow \mu_q(\mathbf{x}_t, \mathbf{x}_0)$

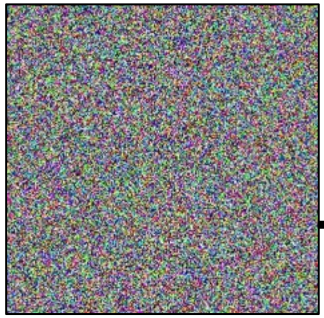
Slide from Jia-Bin Huang

$$\mathbf{x}_0 \xrightarrow{\text{Encoding}} \mathbf{x}_1 \xrightarrow{\text{Encoding}} \mathbf{x}_2 \cdots \xrightarrow{\text{Encoding}} \mathbf{x}_T \xrightarrow{\text{Decoding } \theta} \mathbf{x}_{T-1} \quad \cdots \quad \xrightarrow{\text{Decoding } \theta} \hat{\mathbf{x}}_0$$

$$\sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}\left[\; w(t) \;\left[\, \| \; \hat{\mathbf{x}}_\theta(\mathbf{x}_t, t) \; - \; \mathbf{x}_0 \; \|_2^2 \,\right]\right]$$

$$\mu_q(\mathbf{x}_t, \mathbf{x}_0) = \frac{\sqrt{\alpha_t}(1-\overline{\alpha}_{t-1})}{1-\overline{\alpha}_t}\mathbf{x}_t + \frac{\sqrt{\overline{\alpha}_{t-1}}(1-\alpha_t)}{1-\overline{\alpha}_t}\mathbf{x}_0$$

$$\mu_\theta(\mathbf{x}_t, t) = \frac{\sqrt{\alpha_t}(1-\overline{\alpha}_{t-1})}{1-\overline{\alpha}_t}\mathbf{x}_t + \frac{\sqrt{\overline{\alpha}_{t-1}}(1-\alpha_t)}{1-\overline{\alpha}_t}\hat{\mathbf{x}}_\theta(\mathbf{x}_t, t)$$



$\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{\epsilon}; 0, \boldsymbol{I})$

$\mathbf{x}_0$

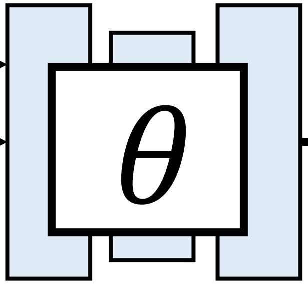
Encoding

$q(\mathbf{x}_t|\mathbf{x}_0)$

$\mathbf{x}_t = \sqrt{\overline{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\overline{\alpha}_t}\boldsymbol{\epsilon}$

$\mathbf{x}_t$

t



U-Net

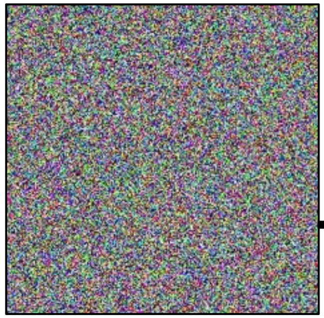
$\hat{\mathbf{x}}_\theta(\mathbf{x}_t, t) \rightarrow \boxed{L_2}$

Slide from Jia-Bin Huang

$$\mathbf{x}_0 \xrightarrow{\text{Encoding}} \mathbf{x}_1 \xrightarrow{\text{Encoding}} \mathbf{x}_2 \cdots \xrightarrow{\text{Encoding}} \mathbf{x}_T \xrightarrow{\text{Decoding } \theta} \mathbf{x}_{T-1} \quad \cdots \quad \xrightarrow{\text{Decoding } \theta} \hat{\mathbf{x}}_0$$

$$\sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}\left[\; w(t) \;\; \left[\, || \, \mu_\theta(\mathbf{x}_t, t) - \mu_q(\mathbf{x}_t, \mathbf{x}_0) \, ||_2^2 \, \right] \right]$$

$$\mu_q(\mathbf{x}_t, \mathbf{x}_0) = \frac{\sqrt{\alpha_t}(1-\overline{\alpha}_{t-1})}{1-\overline{\alpha}_t}\mathbf{x}_t + \frac{\sqrt{\overline{\alpha}_{t-1}}(1-\alpha_t)}{1-\overline{\alpha}_t}\mathbf{x}_0$$



$\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{\epsilon}; 0, \boldsymbol{I})$

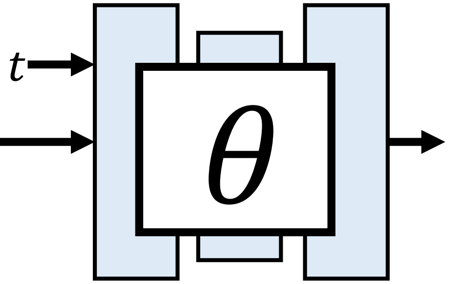
$\mathbf{x}_0$

Encoding

$q(\mathbf{x}_t|\mathbf{x}_0)$

$\mathbf{x}_t = \sqrt{\overline{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\overline{\alpha}_t}\boldsymbol{\epsilon}$

$\mathbf{x}_t$



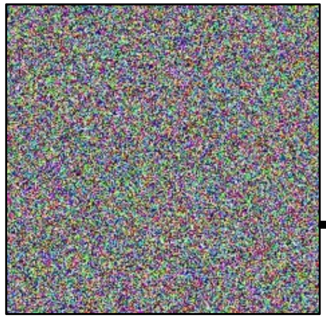
U-Net

Slide from Jia-Bin Huang

$$\mathbf{x}_0 \xrightarrow{\text{Encoding}} \mathbf{x}_1 \xrightarrow{\text{Encoding}} \mathbf{x}_2 \cdots \xrightarrow{\text{Encoding}} \mathbf{x}_T \xrightarrow{\text{Decoding } \theta} \mathbf{x}_{T-1} \quad \cdots \quad \xrightarrow{\text{Decoding } \theta} \hat{\mathbf{x}}_0$$

$$\sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}\left[\ w(t)\ \left[\, \| \mu_\theta(\mathbf{x}_t, t) - \mu_q(\mathbf{x}_t, \mathbf{x}_0) \|_2^2 \,\right]\right]$$

$$\mu_q(\mathbf{x}_t, \mathbf{x}_0) = \frac{\sqrt{\alpha_t}(1-\overline{\alpha}_{t-1})}{1-\overline{\alpha}_t}\mathbf{x}_t + \frac{\sqrt{\overline{\alpha}_{t-1}}(1-\alpha_t)}{1-\overline{\alpha}_t}\mathbf{x}_0$$



$\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{\epsilon}; 0, \boldsymbol{I})$

$$\mathbf{x}_t = \sqrt{\overline{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\overline{\alpha}_t}\boldsymbol{\epsilon}$$

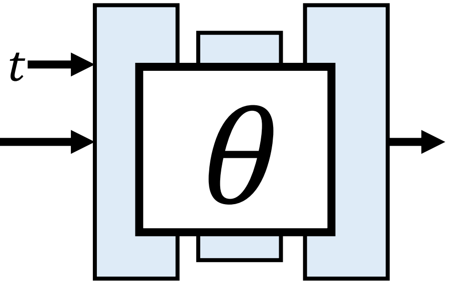
$\mathbf{x}_0$

Encoding

$q(\mathbf{x}_t|\mathbf{x}_0)$

$\mathbf{x}_t = \sqrt{\overline{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\overline{\alpha}_t}\boldsymbol{\epsilon}$

$\mathbf{x}_t$



U-Net

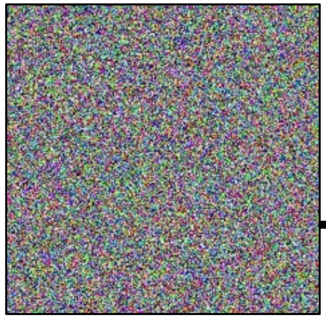
Slide from Jia-Bin Huang

$$\mathbf{x}_0 \xrightarrow{\text{Encoding}} \mathbf{x}_1 \xrightarrow{\text{Encoding}} \mathbf{x}_2 \cdots \xrightarrow{\text{Encoding}} \mathbf{x}_T \xrightarrow{\text{Decoding } \theta} \mathbf{x}_{T-1} \quad \cdots \quad \xrightarrow{\text{Decoding } \theta} \hat{\mathbf{x}}_0$$

$$\sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}\left[\; w(t) \; \left[\left\| \mu_\theta(\mathbf{x}_t, t) - \mu_q(\mathbf{x}_t, \mathbf{x}_0) \right\|_2^2 \right] \right]$$

$$\mu_q(\mathbf{x}_t, \mathbf{x}_0) = \frac{\sqrt{\alpha_t}(1 - \overline{\alpha}_{t-1})}{1 - \overline{\alpha}_t}\mathbf{x}_t + \frac{\sqrt{\overline{\alpha}_{t-1}}(1 - \alpha_t)}{1 - \overline{\alpha}_t}\mathbf{x}_0$$

$$\sqrt{\overline{\alpha}_t}\mathbf{x}_0 = \mathbf{x}_t - \sqrt{1 - \overline{\alpha}_t}\boldsymbol{\epsilon}$$



$\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{\epsilon}; 0, \boldsymbol{I})$

$\mathbf{x}_0$

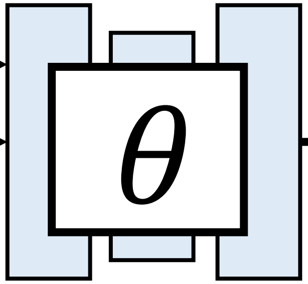
Encoding

$q(\mathbf{x}_t|\mathbf{x}_0)$

$\mathbf{x}_t = \sqrt{\overline{\alpha}_t}\mathbf{x}_0 + \sqrt{1 - \overline{\alpha}_t}\boldsymbol{\epsilon}$

$\mathbf{x}_t$

t

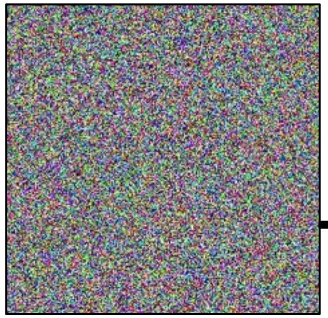


U-Net

Slide from Jia-Bin Huang

$$\mathbf{x}_0 \xrightarrow{\text{Encoding}} \mathbf{x}_1 \xrightarrow{\text{Encoding}} \mathbf{x}_2 \cdots \xrightarrow{\text{Encoding}} \mathbf{x}_T \xrightarrow{\text{Decoding } \theta} \mathbf{x}_{T-1} \quad \cdots \quad \xrightarrow{\text{Decoding } \theta} \hat{\mathbf{x}}_0$$

$$\sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}\left[\ w(t) \ \left[|| \mu_\theta(\mathbf{x}_t, t) - \mu_q(\mathbf{x}_t, \mathbf{x}_0) ||_2^2 \right] \right]$$



$\epsilon \sim \mathcal{N}(\epsilon; 0, \boldsymbol{I})$

$$\mu_q(\mathbf{x}_t, \mathbf{x}_0) = \frac{\sqrt{\alpha_t}(1 - \overline{\alpha}_{t-1})}{1 - \overline{\alpha}_t}\mathbf{x}_t + \frac{\sqrt{\overline{\alpha}_{t-1}}(1 - \alpha_t)}{1 - \overline{\alpha}_t}\mathbf{x}_0$$

$$\mathbf{x}_0 = \frac{\mathbf{x}_t - \sqrt{1 - \overline{\alpha}_t}\epsilon}{\sqrt{\overline{\alpha}_t}}$$

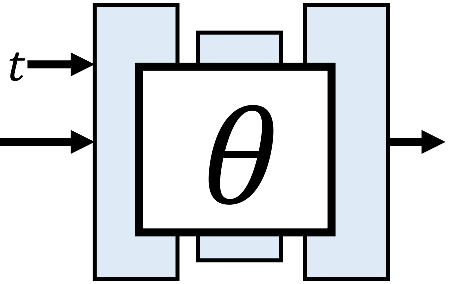
$\mathbf{x}_0$

Encoding

$q(\mathbf{x}_t|\mathbf{x}_0)$

$\mathbf{x}_t = \sqrt{\overline{\alpha}_t}\mathbf{x}_0 + \sqrt{1 - \overline{\alpha}_t}\epsilon$

$\mathbf{x}_t$

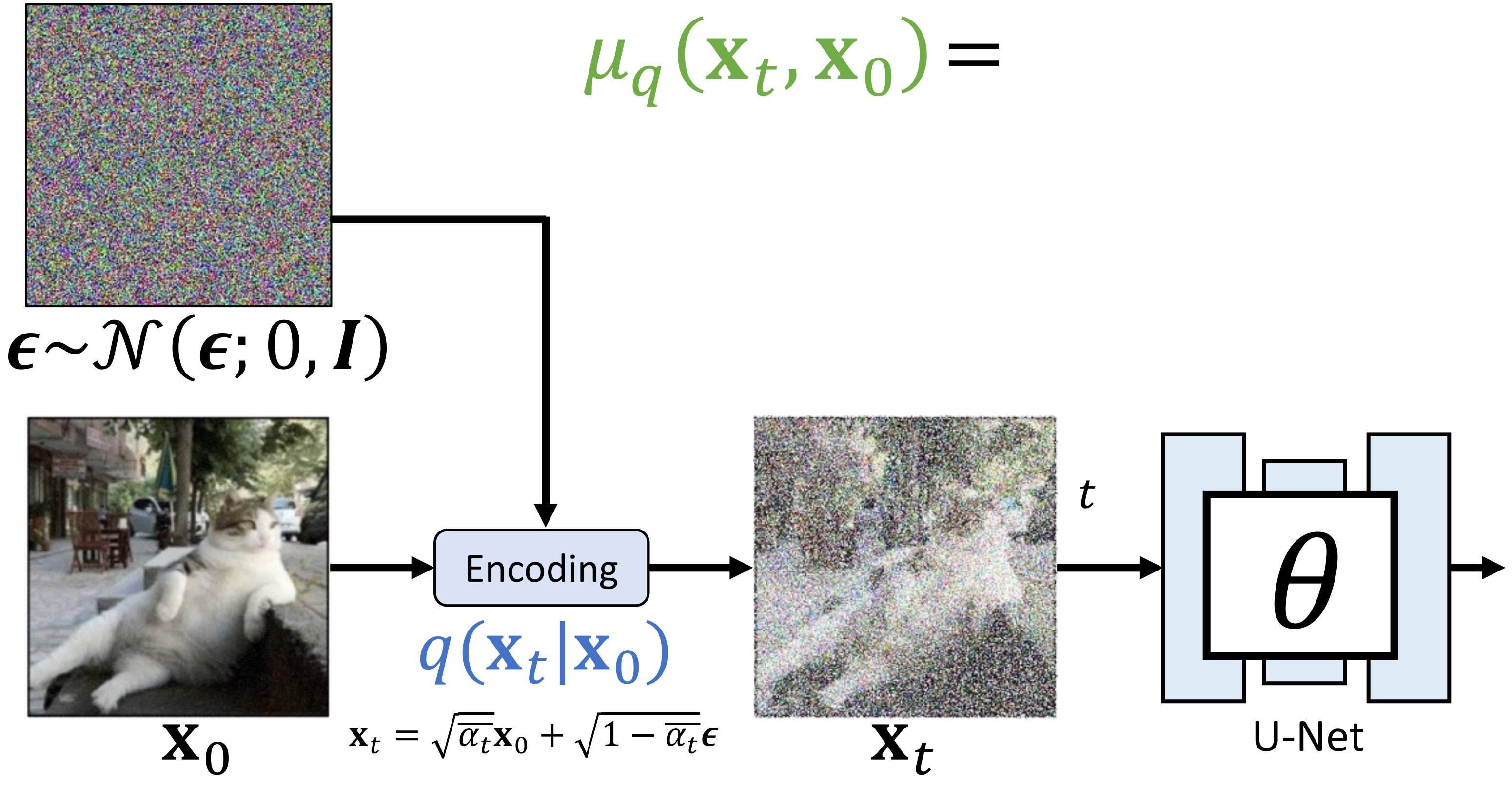


Slide from Jia-Bin Huang

$$\mathbf{x}_0 \xrightarrow{\text{Encoding}} \mathbf{x}_1 \xrightarrow{\text{Encoding}} \mathbf{x}_2 \cdots \xrightarrow{\text{Encoding}} \mathbf{x}_T \xrightarrow{\text{Decoding } \theta} \mathbf{x}_{T-1} \quad \cdots \quad \xrightarrow{\text{Decoding } \theta} \hat{\mathbf{x}}_0$$

$$\sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}\left[\quad w(t) \quad \left[|| \mu_\theta(\mathbf{x}_t, t) - \mu_q(\mathbf{x}_t, \mathbf{x}_0) ||_2^2 \right] \right]$$

$$\mu_q(\mathbf{x}_t, \mathbf{x}_0) =$$



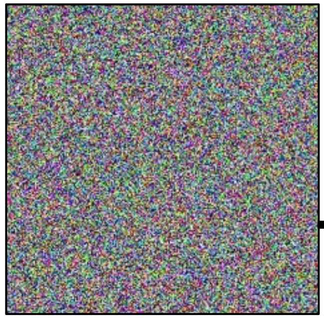
Slide from Jia-Bin Huang

$$\mathbf{x}_0 \xrightarrow{\text{Encoding}} \mathbf{x}_1 \xrightarrow{\text{Encoding}} \mathbf{x}_2 \cdots \xrightarrow{\text{Encoding}} \mathbf{x}_T \xrightarrow{\text{Decoding } \theta} \mathbf{x}_{T-1} \quad \cdots \quad \xrightarrow{\text{Decoding } \theta} \hat{\mathbf{x}}_0$$

$$\sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}\left[\quad w(t) \quad \left[\left\| \mu_\theta(\mathbf{x}_t, t) - \mu_q(\mathbf{x}_t, \mathbf{x}_0) \right\|_2^2 \right] \right]$$

$$\mu_q(\mathbf{x}_t, \mathbf{x}_0) = \frac{1}{\sqrt{\alpha_t}}\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{(1-\bar{\alpha}_t)\alpha_t}}\boldsymbol{\epsilon}$$

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



$\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{\epsilon}; 0, \boldsymbol{I})$

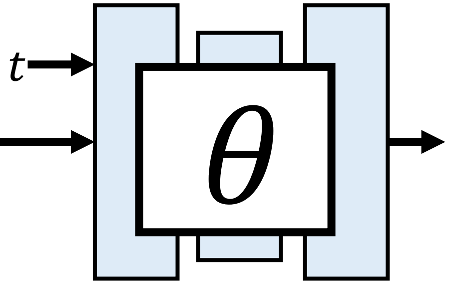
$\mathbf{x}_0$

Encoding

$q(\mathbf{x}_t|\mathbf{x}_0)$

$\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}$

$\mathbf{x}_t$

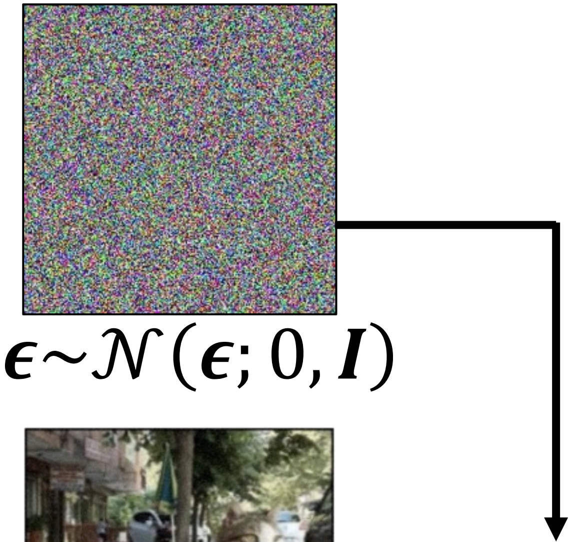


U-Net

Slide from Jia-Bin Huang

$$\mathbf{x}_0 \xrightarrow{\text{Encoding}} \mathbf{x}_1 \xrightarrow{\text{Encoding}} \mathbf{x}_2 \cdots \xrightarrow{\text{Encoding}} \mathbf{x}_T \xrightarrow{\text{Decoding } \theta} \mathbf{x}_{T-1} \quad \cdots \quad \xrightarrow{\text{Decoding } \theta} \hat{\mathbf{x}}_0$$

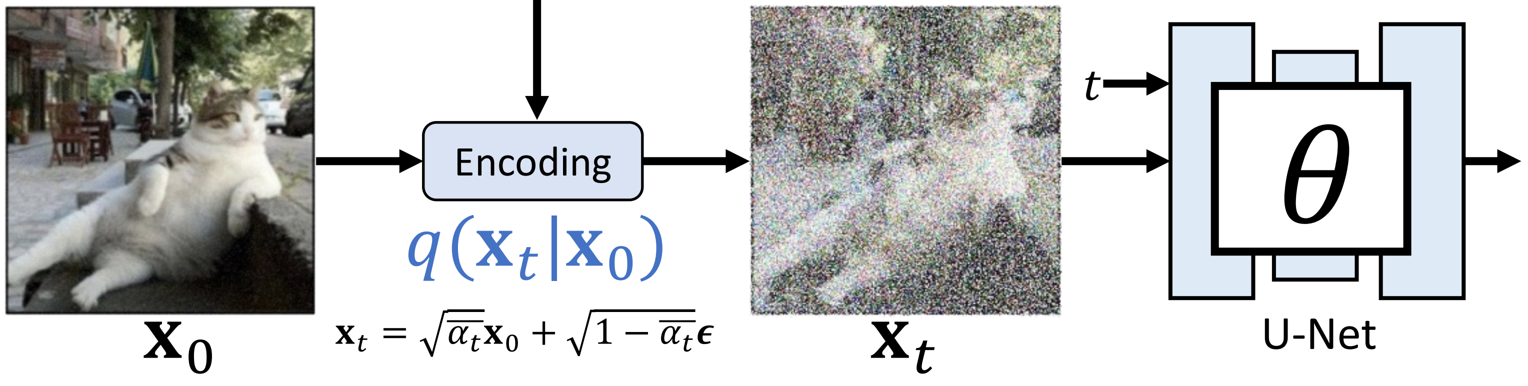
$$\sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}\Big[\; w(t) \; \big[\, || \mu_\theta(\mathbf{x}_t, t) - \mu_q(\mathbf{x}_t, \mathbf{x}_0) ||_2^2 \,\big]\Big]$$



$\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{\epsilon}; 0, \boldsymbol{I})$

$$\mu_q(\mathbf{x}_t, \mathbf{x}_0) = \frac{1}{\sqrt{\alpha_t}}\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{(1-\bar{\alpha}_t)\alpha_t}}\boldsymbol{\epsilon}$$

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



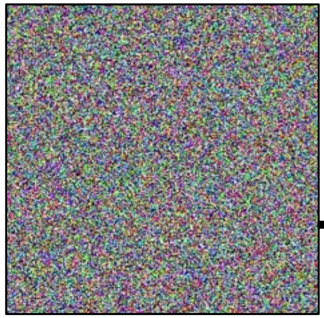
Slide from Jia-Bin Huang

$$\mathbf{x}_0 \xrightarrow{\text{Encoding}} \mathbf{x}_1 \xrightarrow{\text{Encoding}} \mathbf{x}_2 \cdots \xrightarrow{\text{Encoding}} \mathbf{x}_T \xrightarrow{\text{Decoding } \theta} \mathbf{x}_{T-1} \quad \cdots \quad \xrightarrow{\text{Decoding } \theta} \hat{\mathbf{x}}_0$$

$$\sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}\left[w(t) \left[||\hat{\boldsymbol{\epsilon}}_\theta(\mathbf{x}_t, t) - \boldsymbol{\epsilon} ||_2^2 \right] \right]$$

$$\mu_q(\mathbf{x}_t, \mathbf{x}_0) = \frac{1}{\sqrt{\alpha_t}}\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{(1-\bar{\alpha}_t)\alpha_t}}\boldsymbol{\epsilon}$$

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



$\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{\epsilon}; 0, \boldsymbol{I})$

$\mathbf{x}_0$

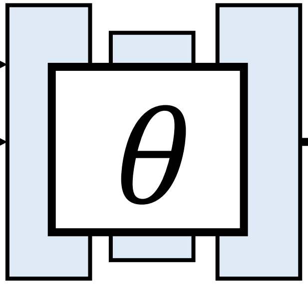
Encoding

$q(\mathbf{x}_t|\mathbf{x}_0)$

$\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}$

$\mathbf{x}_t$

t



U-Net

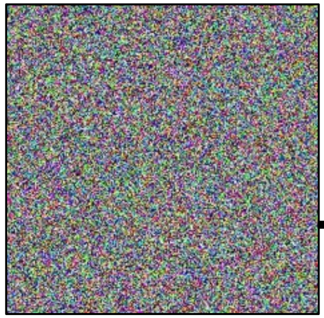
Slide from Jia-Bin Huang

$$\mathbf{x}_0 \xrightarrow{\text{Encoding}} \mathbf{x}_1 \xrightarrow{\text{Encoding}} \mathbf{x}_2 \cdots \xrightarrow{\text{Encoding}} \mathbf{x}_T \xrightarrow{\text{Decoding } \theta} \mathbf{x}_{T-1} \quad \cdots \quad \xrightarrow{\text{Decoding } \theta} \hat{\mathbf{x}}_0$$

$$\sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}\left[\; w(t) \; \left[\| \hat{\boldsymbol{\epsilon}}_{\boldsymbol{\theta}}(\mathbf{x}_t, t) \; - \; \boldsymbol{\epsilon} \; \|_2^2 \right] \right]$$

$$\mu_q(\mathbf{x}_t, \mathbf{x}_0) = \frac{1}{\sqrt{\alpha_t}}\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{(1-\bar{\alpha}_t)\alpha_t}}\boldsymbol{\epsilon}$$

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



$\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{\epsilon}; 0, \boldsymbol{I})$

$\mathbf{x}_0$

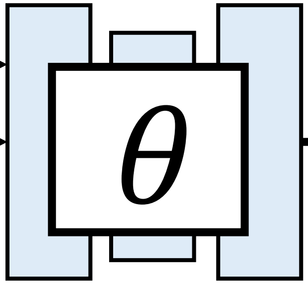
Encoding

$q(\mathbf{x}_t|\mathbf{x}_0)$

$\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}$

$\mathbf{x}_t$

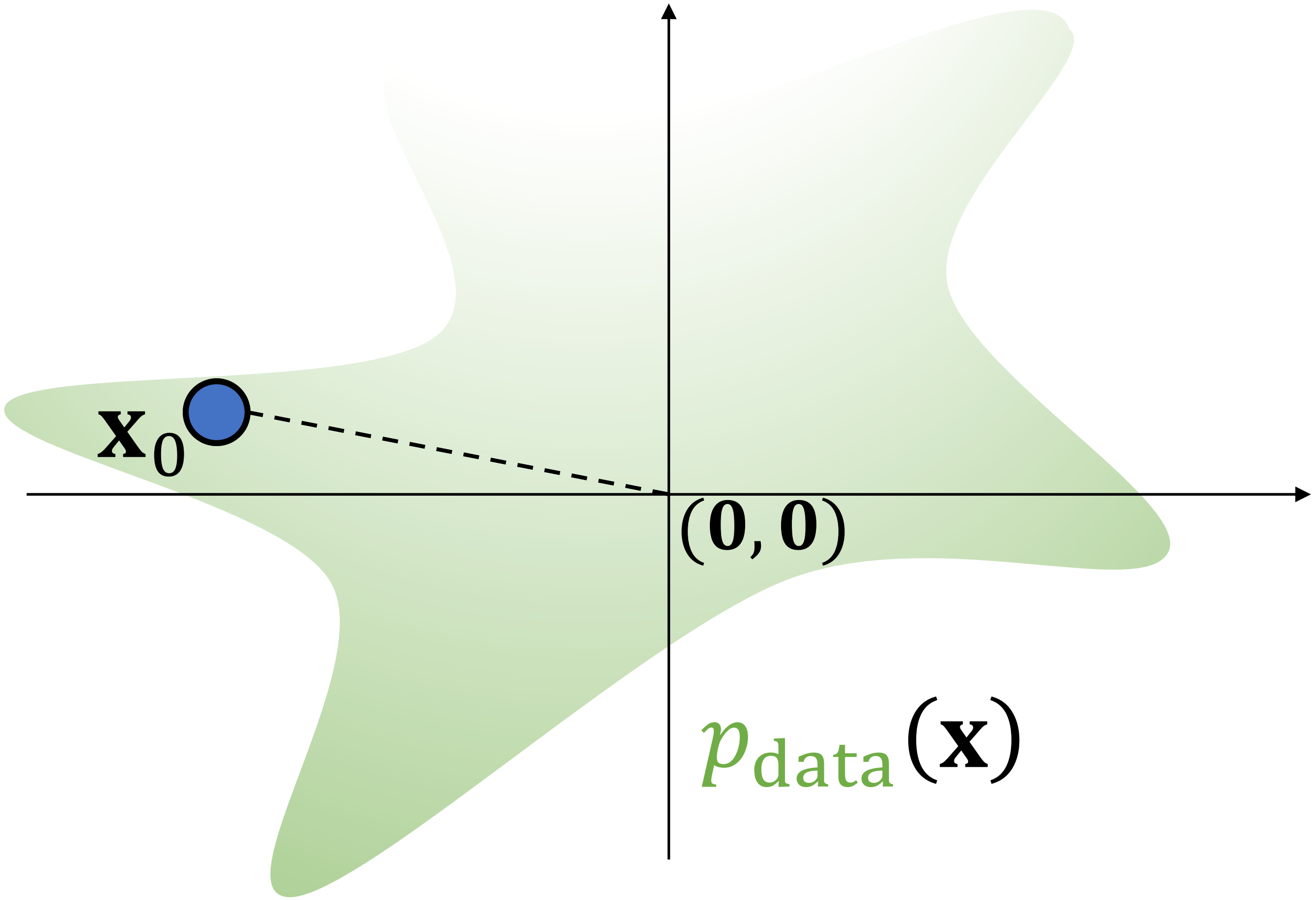
t



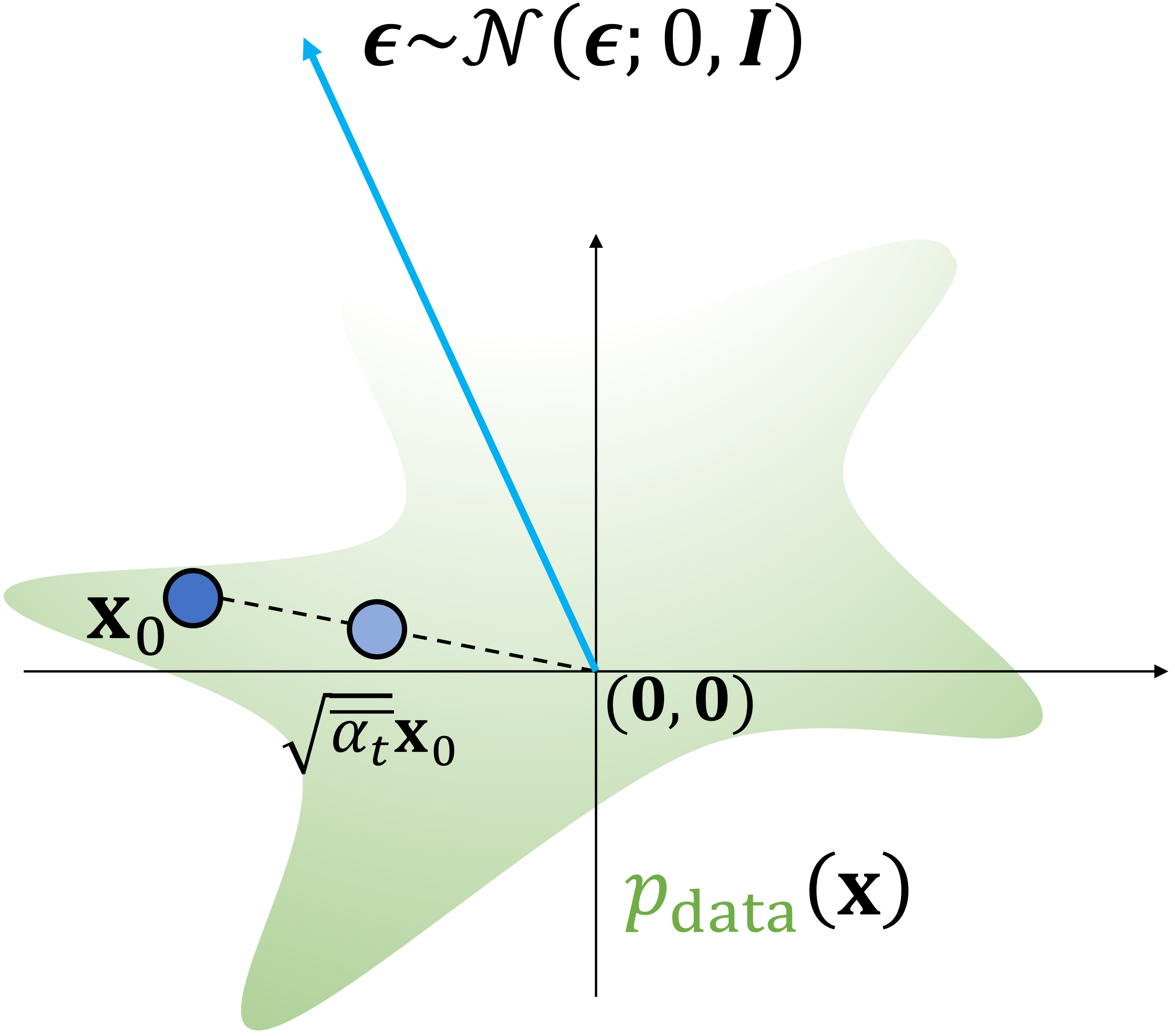
U-Net

$\hat{\boldsymbol{\epsilon}}_{\boldsymbol{\theta}}(\mathbf{x}_t, t) \rightarrow L_2$

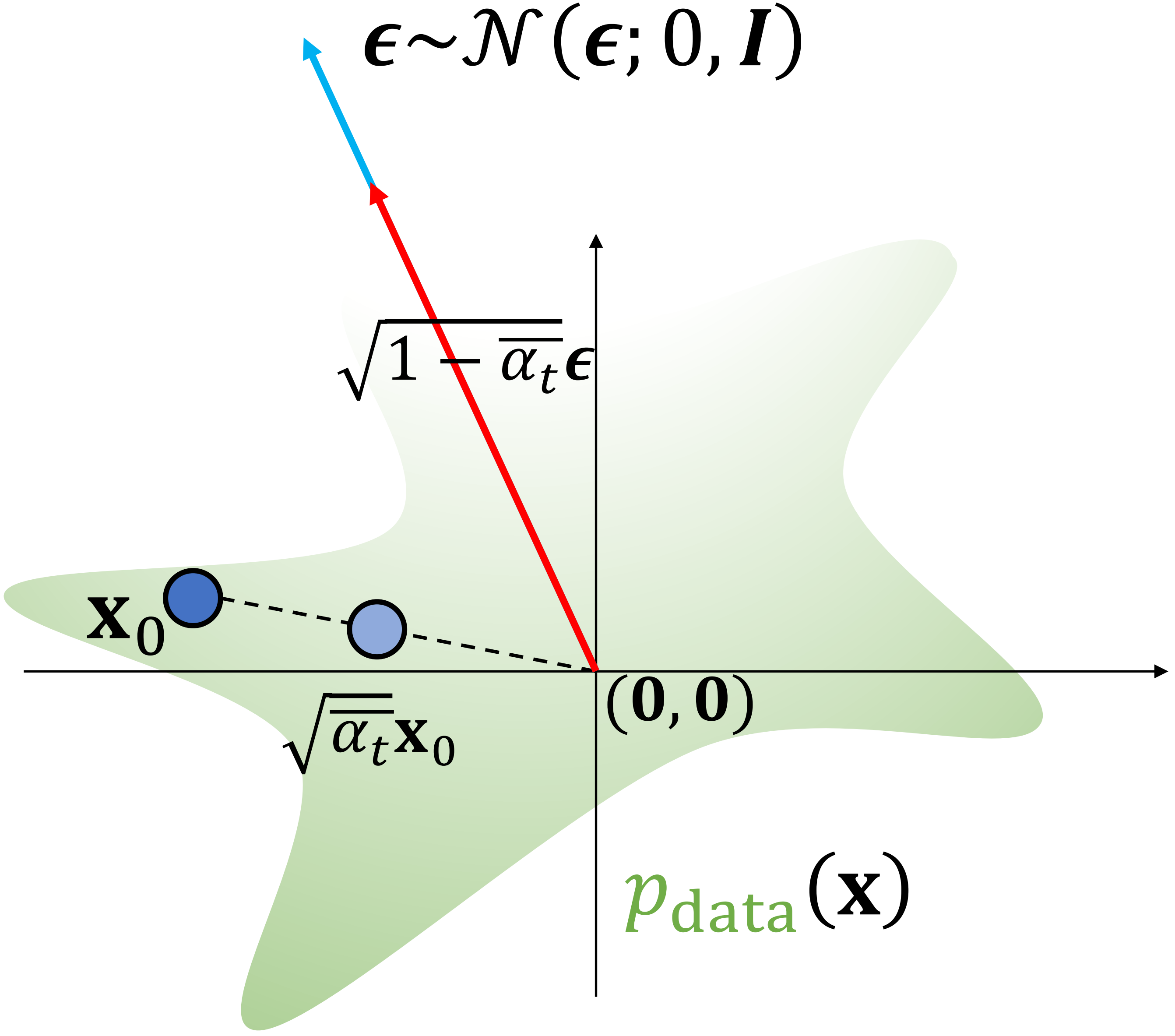
Slide from Jia-Bin Huang



$\mathbf{x}_0$
$(\mathbf{0}, \mathbf{0})$
$p_{\text{data}}(\mathbf{x})$

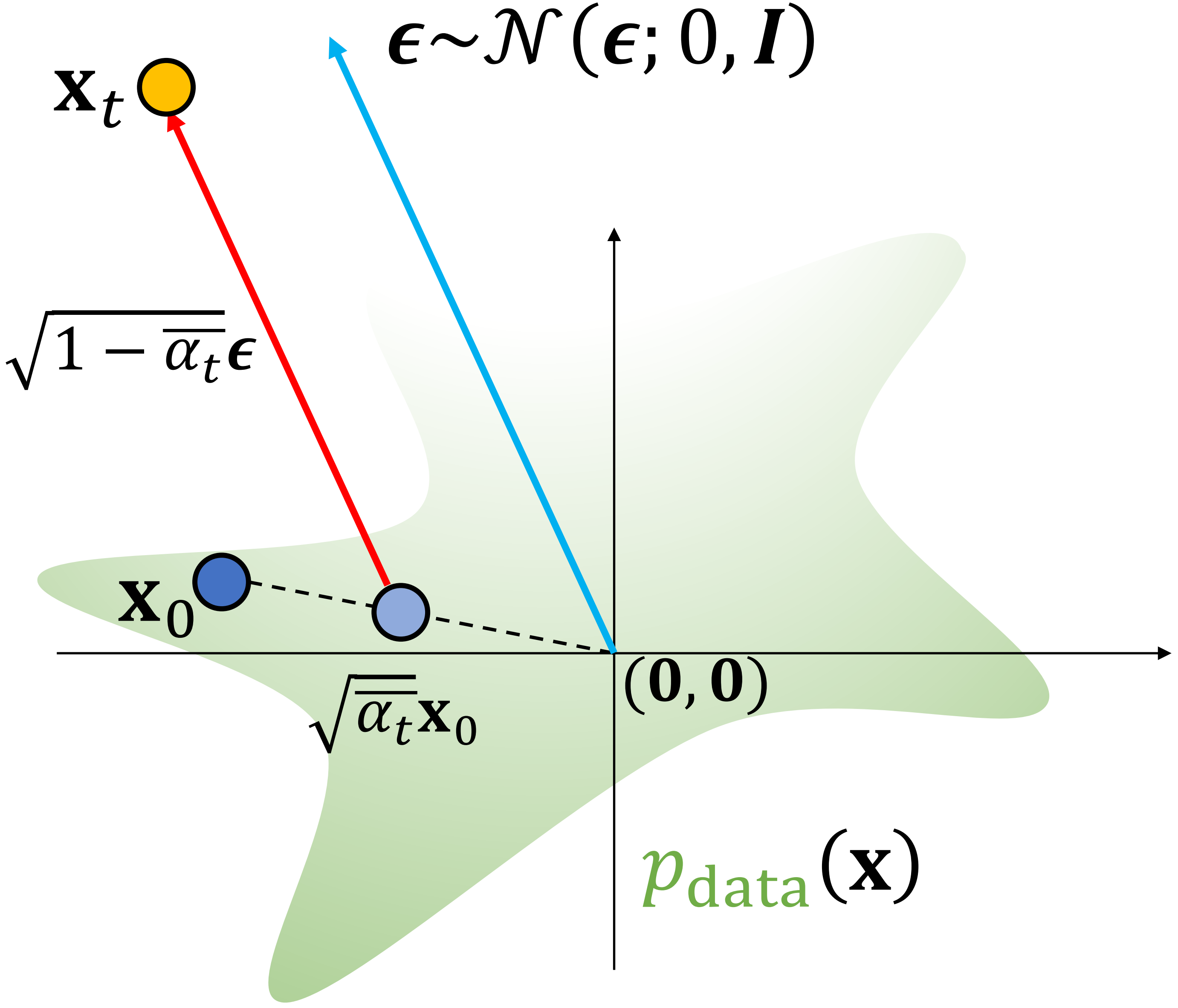


Slide from Jia-Bin Huang



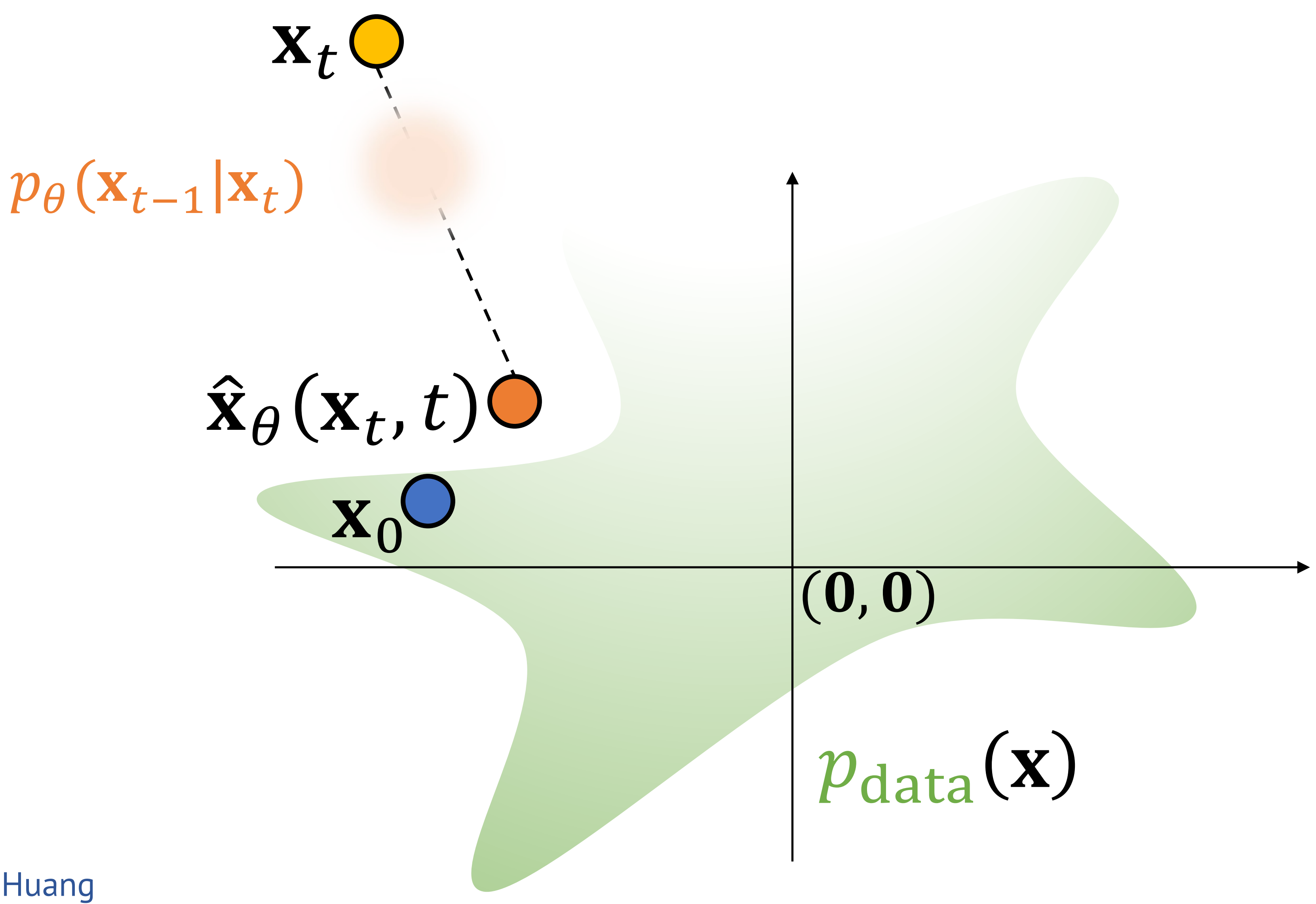
Slide from Jia-Bin Huang

$$\mathbf{x}_t = \sqrt{\overline{\alpha_t}}\mathbf{x}_0 + \sqrt{1 - \overline{\alpha_t}}\boldsymbol{\epsilon}$$



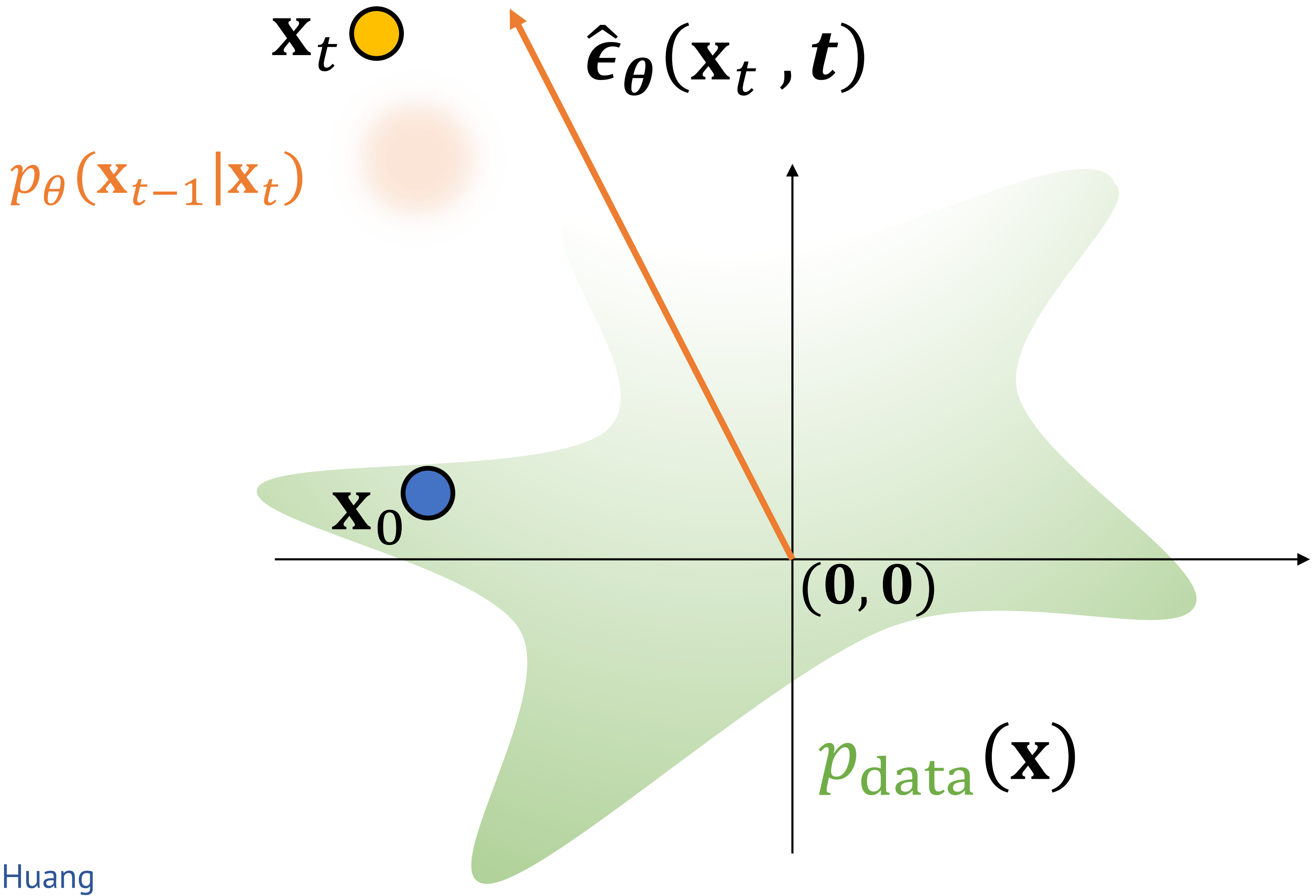
Slide from Jia-Bin Huang

$$\mu_\theta(\mathrm{x}_t, t) = \frac{\sqrt{\alpha_t}(1-\overline{\alpha}_{t-1})}{1-\overline{\alpha}_t}\mathbf{x}_t + \frac{\sqrt{\overline{\alpha}_{t-1}}(1-\alpha_t)}{1-\overline{\alpha}_t}\hat{\mathbf{x}}_\theta(\mathbf{x}_t, t)$$



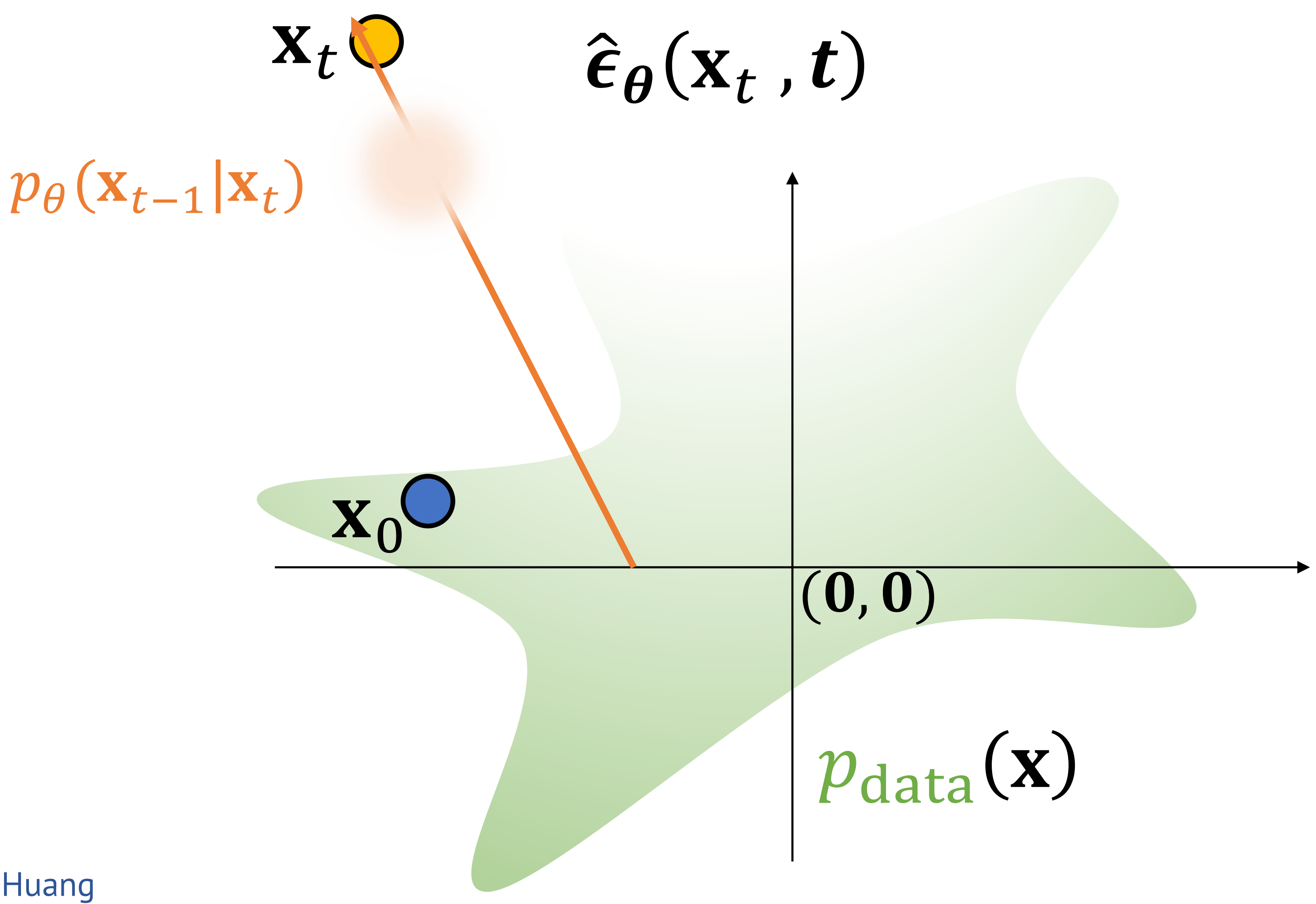
Slide from Jia-Bin Huang

$$\mu_\theta(\mathrm{x}_t, t) = \frac{1}{\sqrt{\alpha_t}}\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{(1-\bar{\alpha}_t)\alpha_t}}\hat{\boldsymbol{\epsilon}}_\theta(\mathbf{x}_t, t)$$



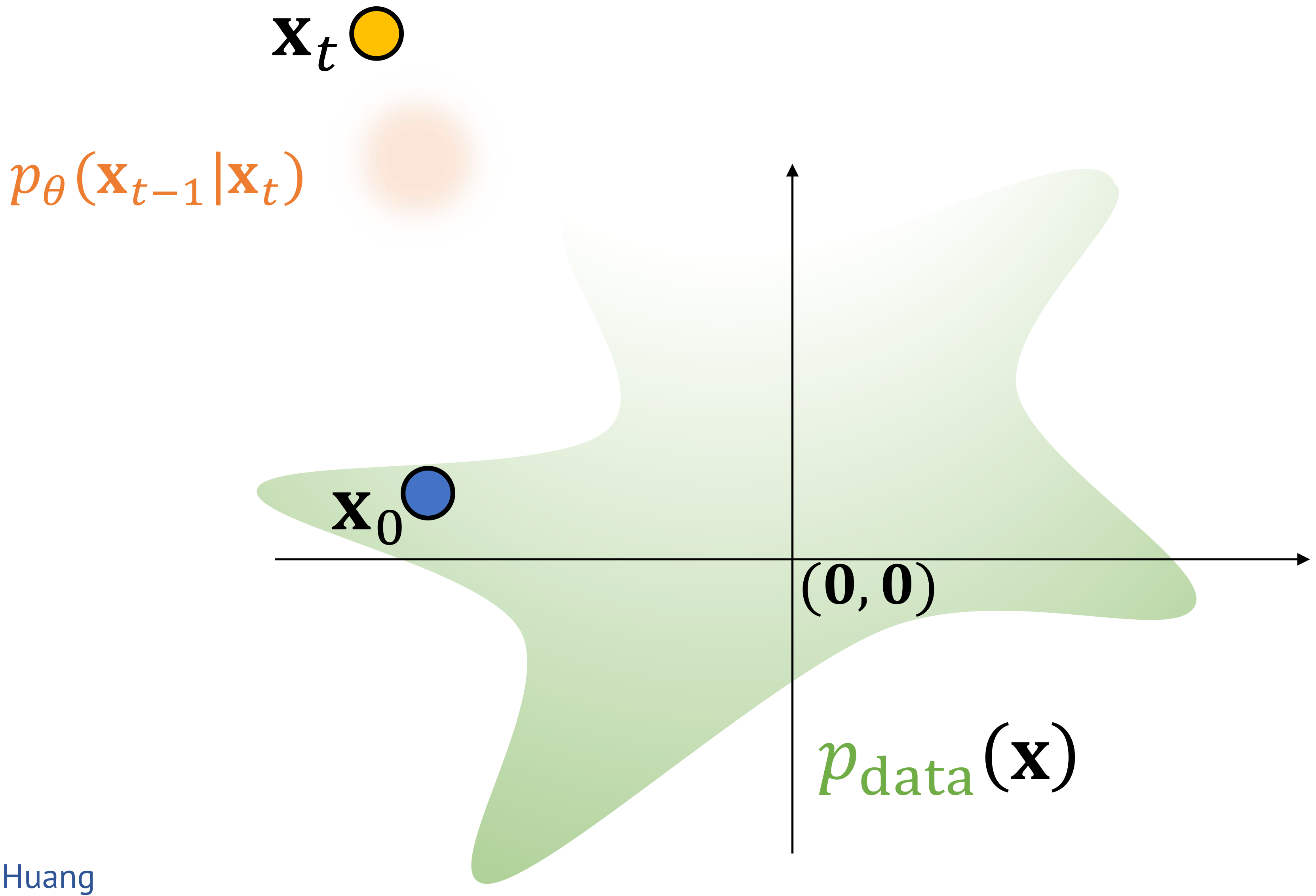
Slide from Jia-Bin Huang

$$\mu_\theta(\mathrm{x}_t, t) = \frac{1}{\sqrt{\alpha_t}}\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{(1-\bar{\alpha}_t)\alpha_t}}\hat{\epsilon}_\theta(\mathbf{x}_t, t)$$

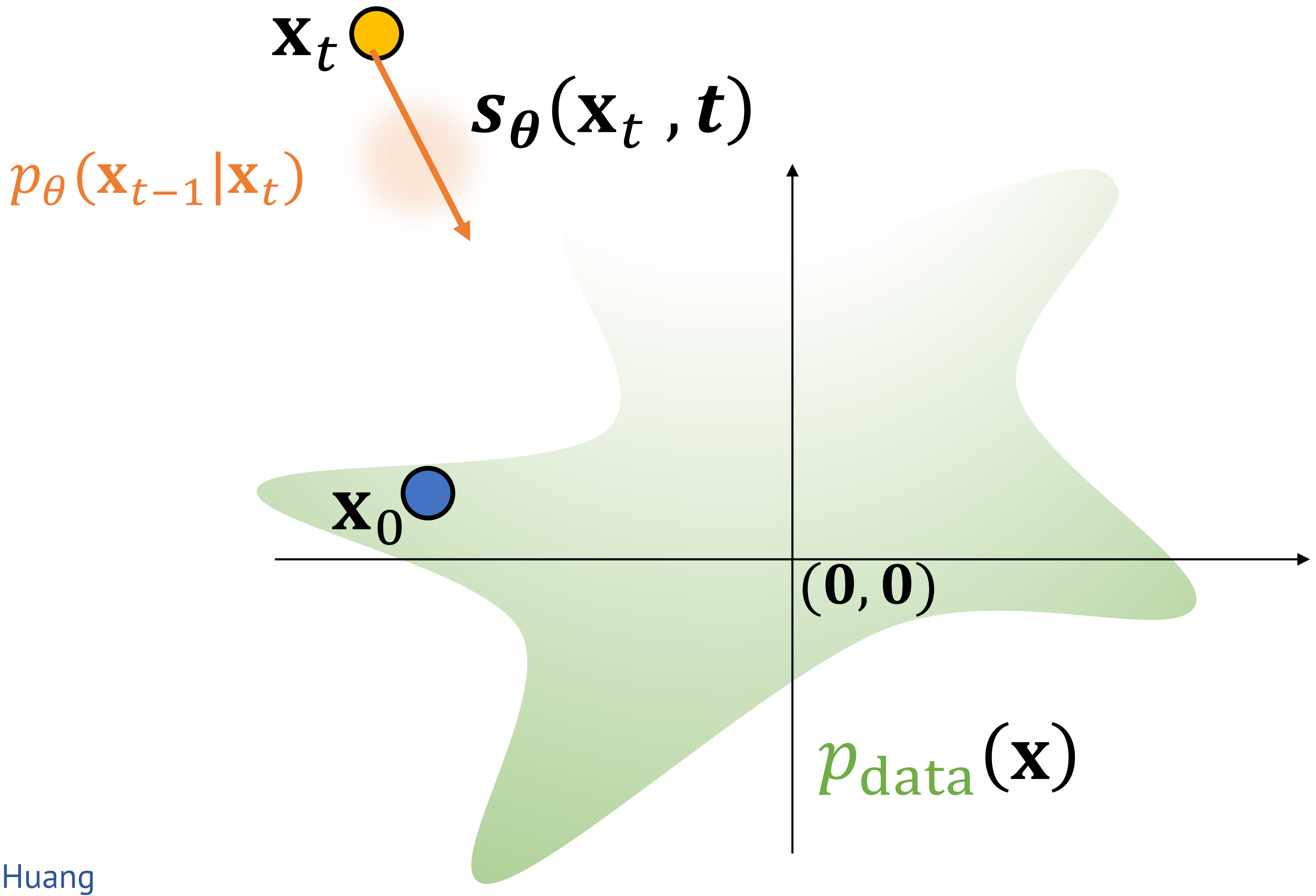


Slide from Jia-Bin Huang

$$\mu_\theta(\mathrm{x}_t, t) =$$

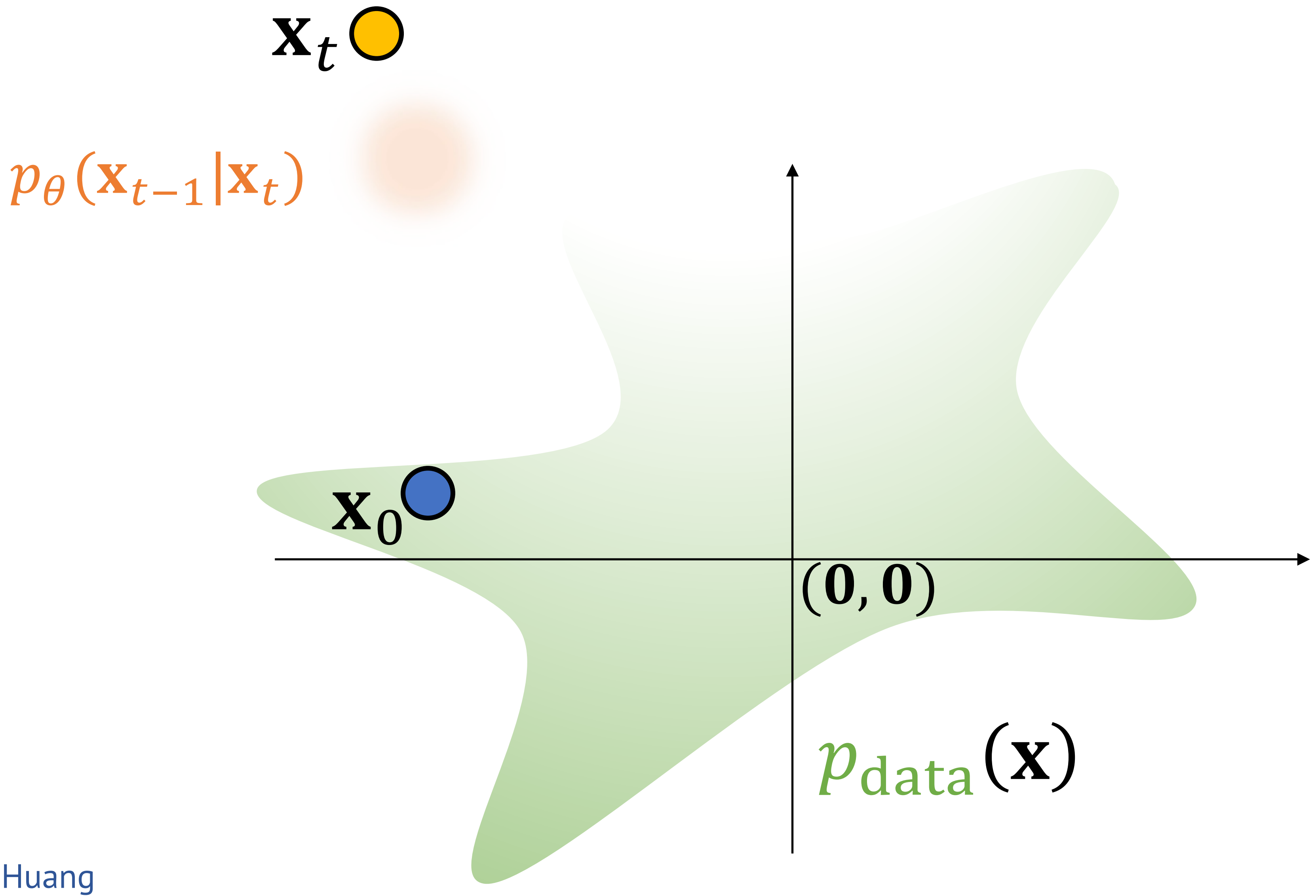


$$\mu_\theta(\mathrm{x}_t, t) = \frac{1}{\sqrt{\alpha_t}}\mathbf{x}_t + \frac{1-\alpha_t}{\sqrt{\alpha_t}}\boldsymbol{s}_\theta(\mathbf{x}_t, t)$$



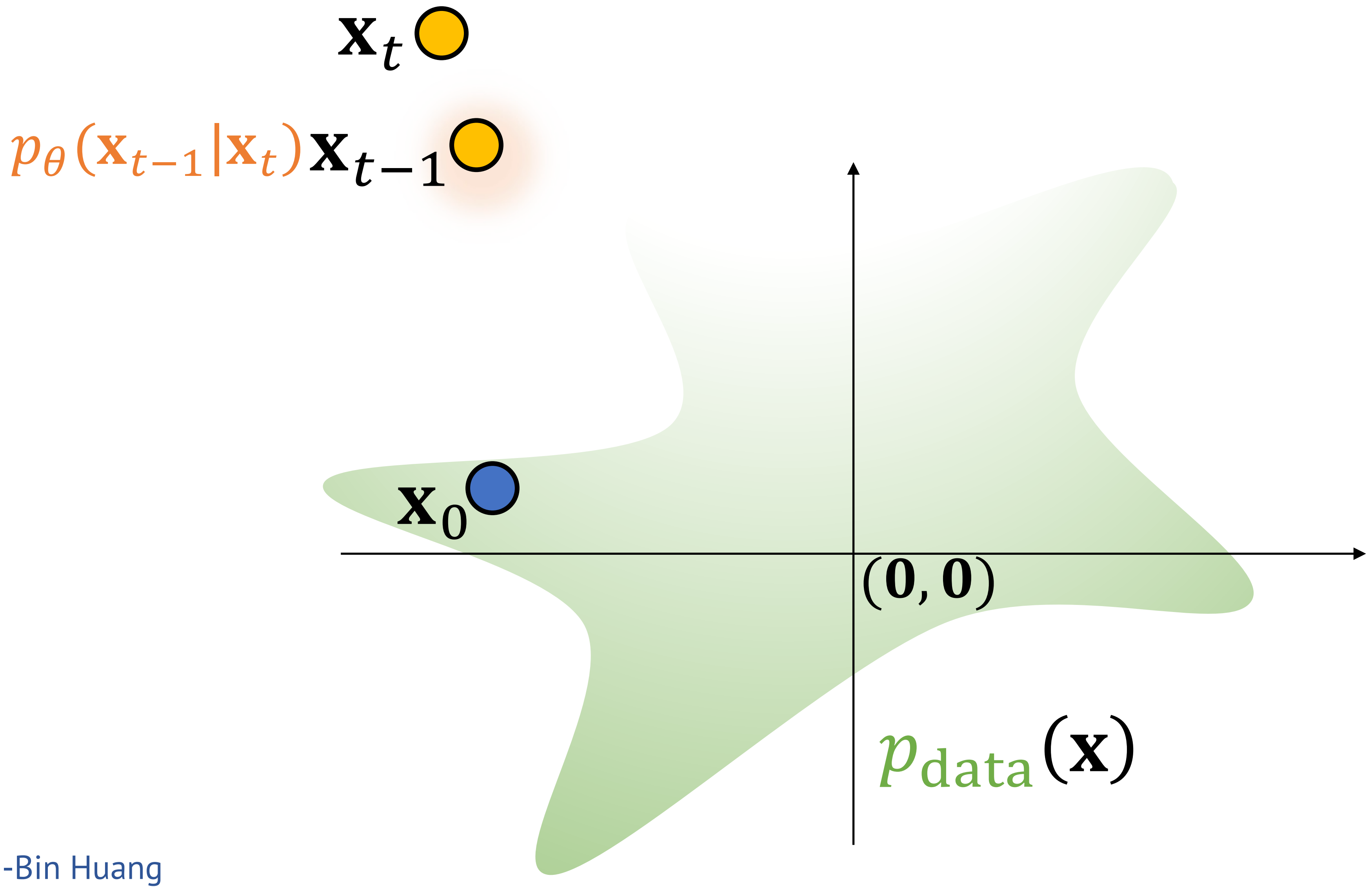
Slide from Jia-Bin Huang

$$\mu_\theta(\mathrm{x}_t, t) = \frac{1}{\sqrt{\alpha_t}}\mathbf{x}_t + \frac{1-\alpha_t}{\sqrt{\alpha_t}}\boldsymbol{s}_{\boldsymbol{\theta}}(\mathbf{x}_t, t)$$

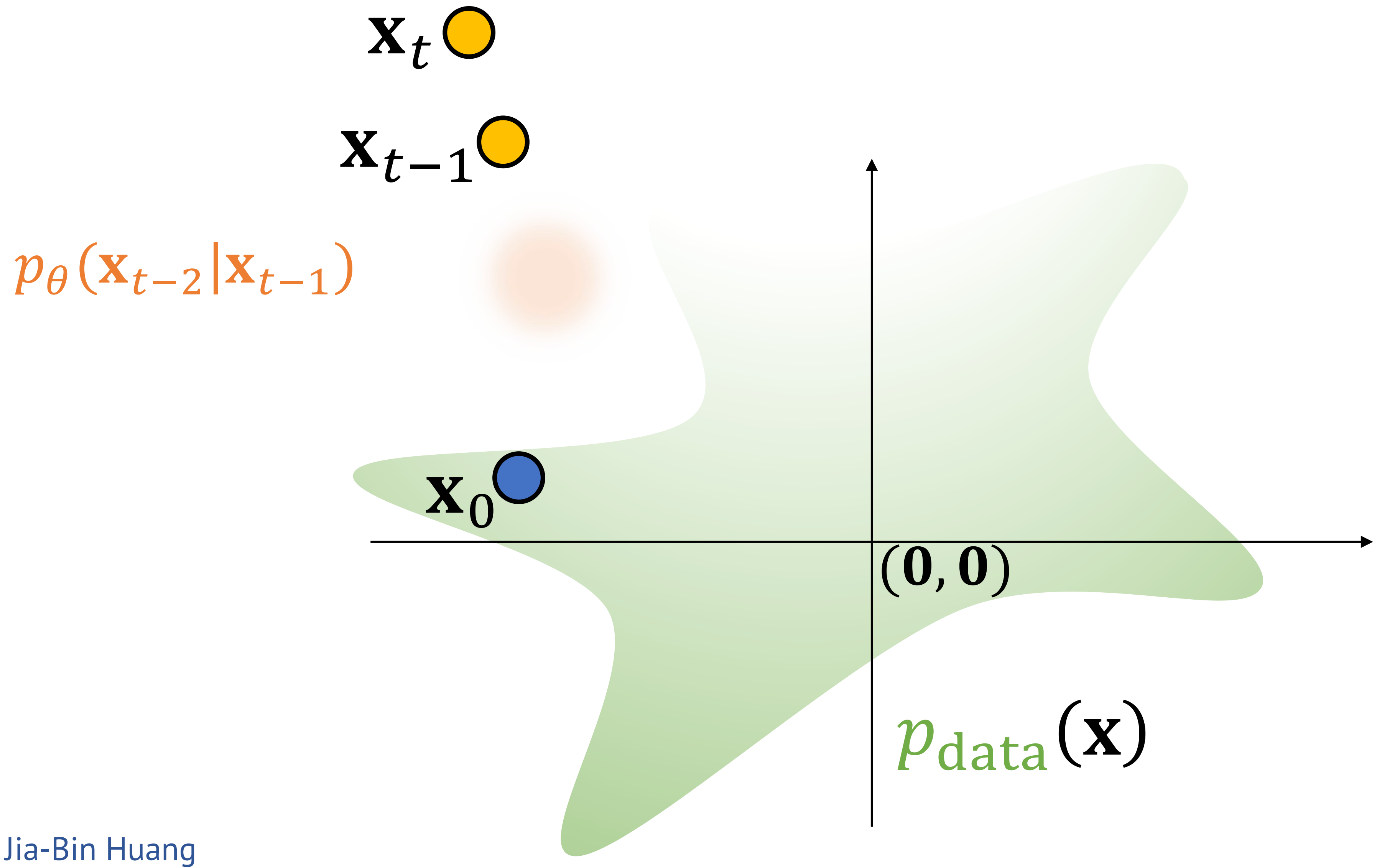


Slide from Jia-Bin Huang

$$\mu_\theta(\mathrm{x}_t, t) = \frac{1}{\sqrt{\alpha_t}}\mathbf{x}_t + \frac{1-\alpha_t}{\sqrt{\alpha_t}}\boldsymbol{s}_\theta(\mathbf{x}_t, t)$$

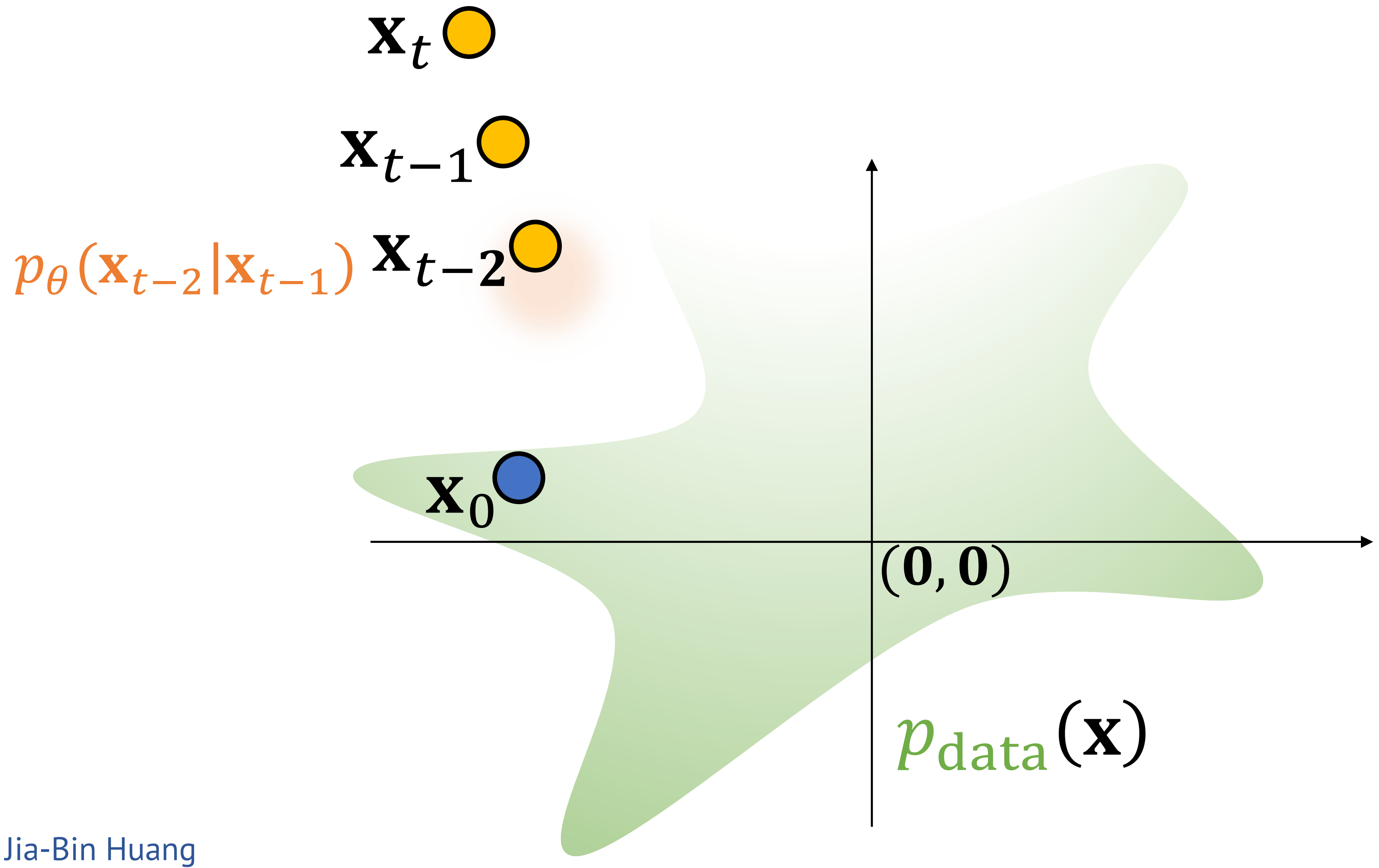


$$\mu_\theta(\mathrm{x}_t, t) = \frac{1}{\sqrt{\alpha_t}}\mathbf{x}_t + \frac{1-\alpha_t}{\sqrt{\alpha_t}}\boldsymbol{s}_{\boldsymbol{\theta}}(\mathbf{x}_t, t)$$

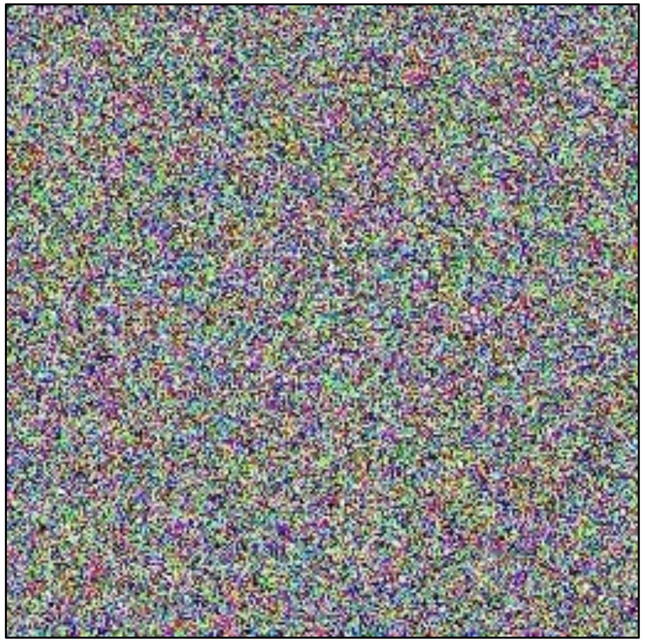


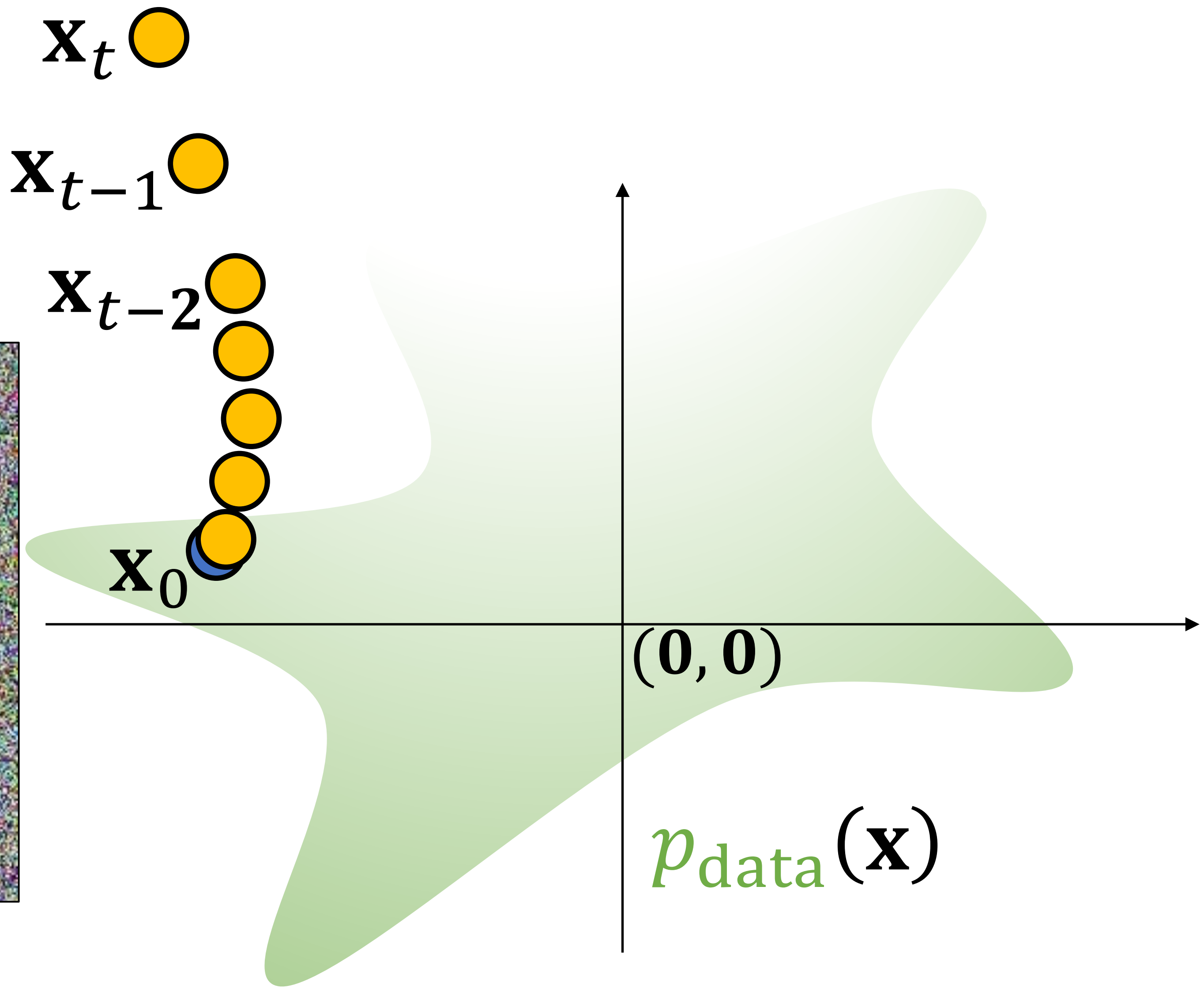
Slide from Jia-Bin Huang

$$\mu_\theta(\mathrm{x}_t, t) = \frac{1}{\sqrt{\alpha_t}}\mathbf{x}_t + \frac{1-\alpha_t}{\sqrt{\alpha_t}}\boldsymbol{s}_\theta(\mathbf{x}_t, t)$$



$$\mu_\theta(\mathrm{x}_t, t) = \frac{1}{\sqrt{\alpha_t}}\mathbf{x}_t + \frac{1-\alpha_t}{\sqrt{\alpha_t}}\boldsymbol{s}_{\boldsymbol{\theta}}(\mathbf{x}_t, t)$$





Slide from Jia-Bin Huang

Please fill in the Feedback Form at:

https://virajshah.com/sc395-feedback3

Thank You!