DiMEDIA:
Diffusion Models in Medical Imaging and Analysis

https://vios.science/tutorials/DiMEDIA-2024

Nefeli Gkouti
Archimedes/Athena RC,
University of Athens

Prof. Sotirios A. Tsaftaris
Archimedes/Athena RC,
University of Edinburgh

Tutorial Schedule

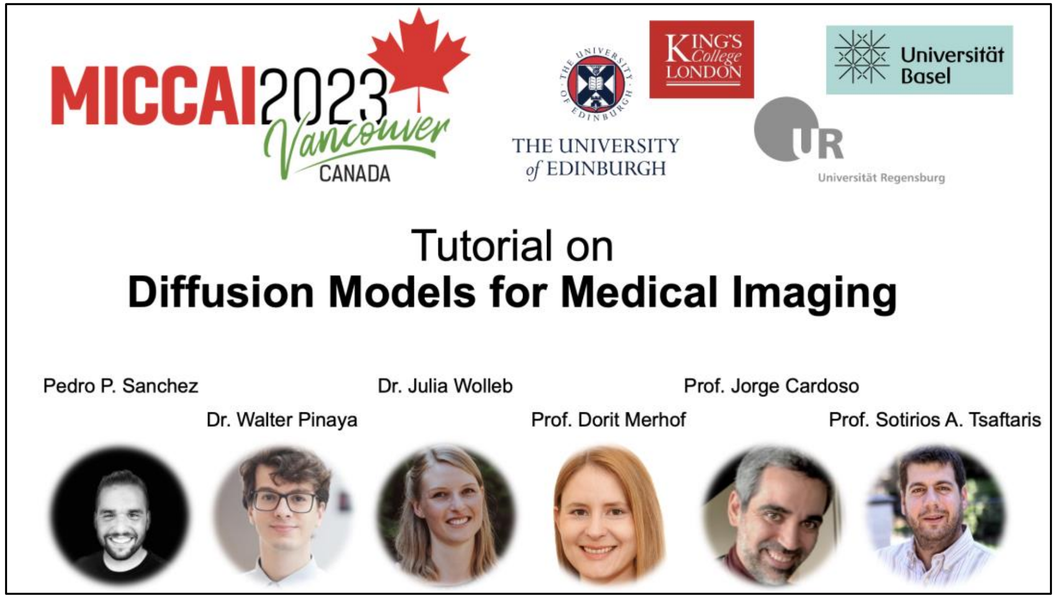
First Half

- ❑ Part 1: Introduction (30 mins)
  - ❖ What? Why? How?
  - ❖ Denoising Diffusion Models
  - ❖ Understanding and Intuition
- ❑ Part 2: Advanced Topics (30 mins)
  - ❖ Sampling Strategies
  - ❖ Conditioning

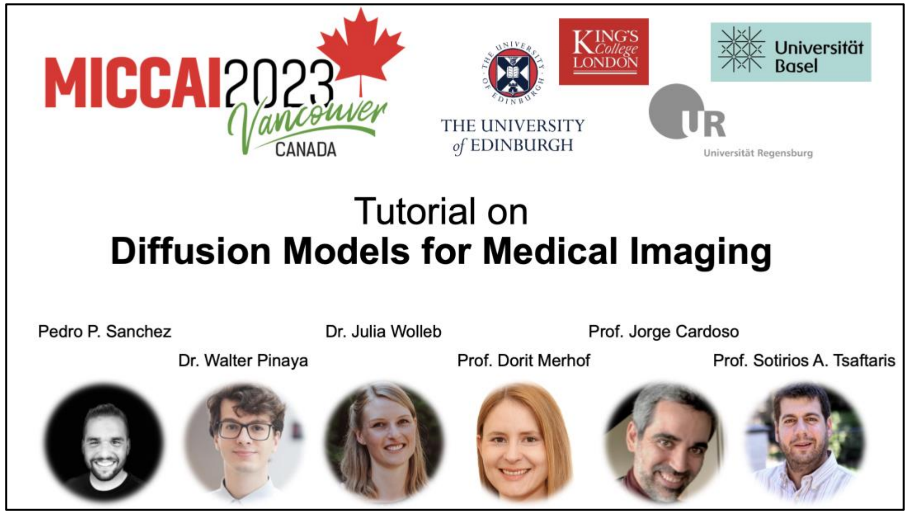
Second Half

- ❑ Q&A
- ❑ Medical Imaging Applications (30 mins)
  - ❖ Synthesis
  - ❖ Segmentation
  - ❖ Anomaly Detection
  - ❖ Reconstruction
  - ❖ Registration

This tutorial is an abridged version of the successful diffusion tutorials in MICCAI 2023 in Vancouver, and ISBI 2024 in Athens

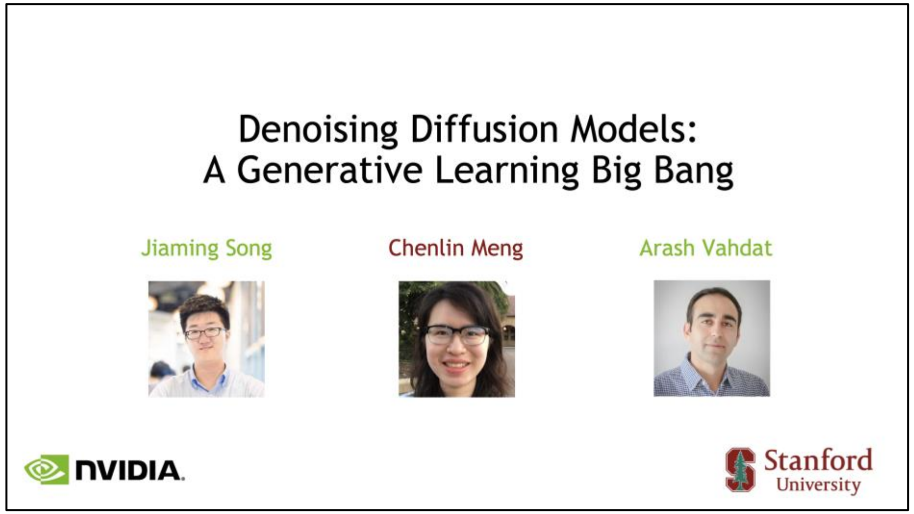


Acknowledgement



MICCAI 2023 - Tutorial

Some slides are adopted/inspired/borrowed from these very nice tutorials!



CVPR 2023 - Tutorial

ISBI 2024 - Tutorial

We went to learn a few things about you

We will use wooclap:

- Are you familiar with diffusion models
- Are you familiar with generative model theory?
- What are your primary objectives for today?

More will come...

How to participate?

1 Go to wooclap.com

2 Enter the event code in the top banner

Event code
FROCKY

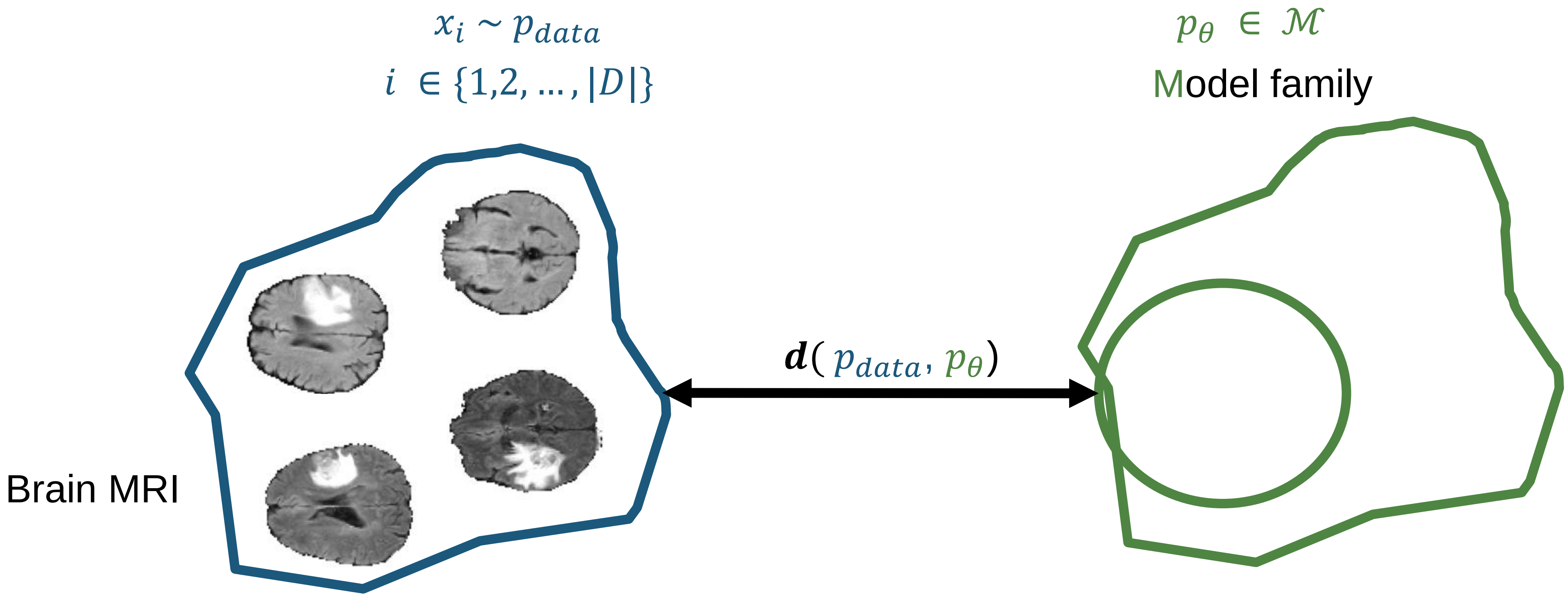
Enable answers by SMS

Copy participation link

Diffusion Models

What? Why? How?

What? Generative Models



$\mathcal{M}$ = {VAE, GAN, NF, **Diffusion Models**}

What? Generative Models

$p_\theta \in \mathcal{M}$

Model family

Density Estimation

$p_\theta(x)$

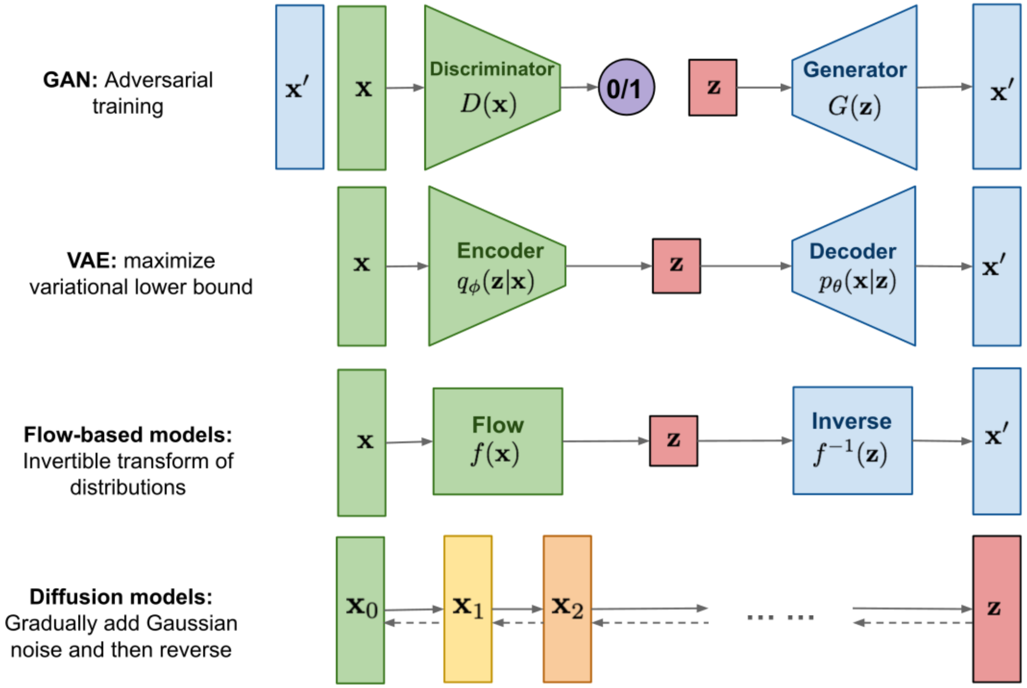
Sampling

$x_{new} \sim p_\theta$

Unsupervised Representation Learning

$z \leftarrow p_\theta(x)$

What? Generative models



likelihood-based models

Require
- <u>inductive bias</u> to ensure a tractable normalizing constant for likelihood computation; or
- <u>surrogate objectives</u> to approximate ML training.

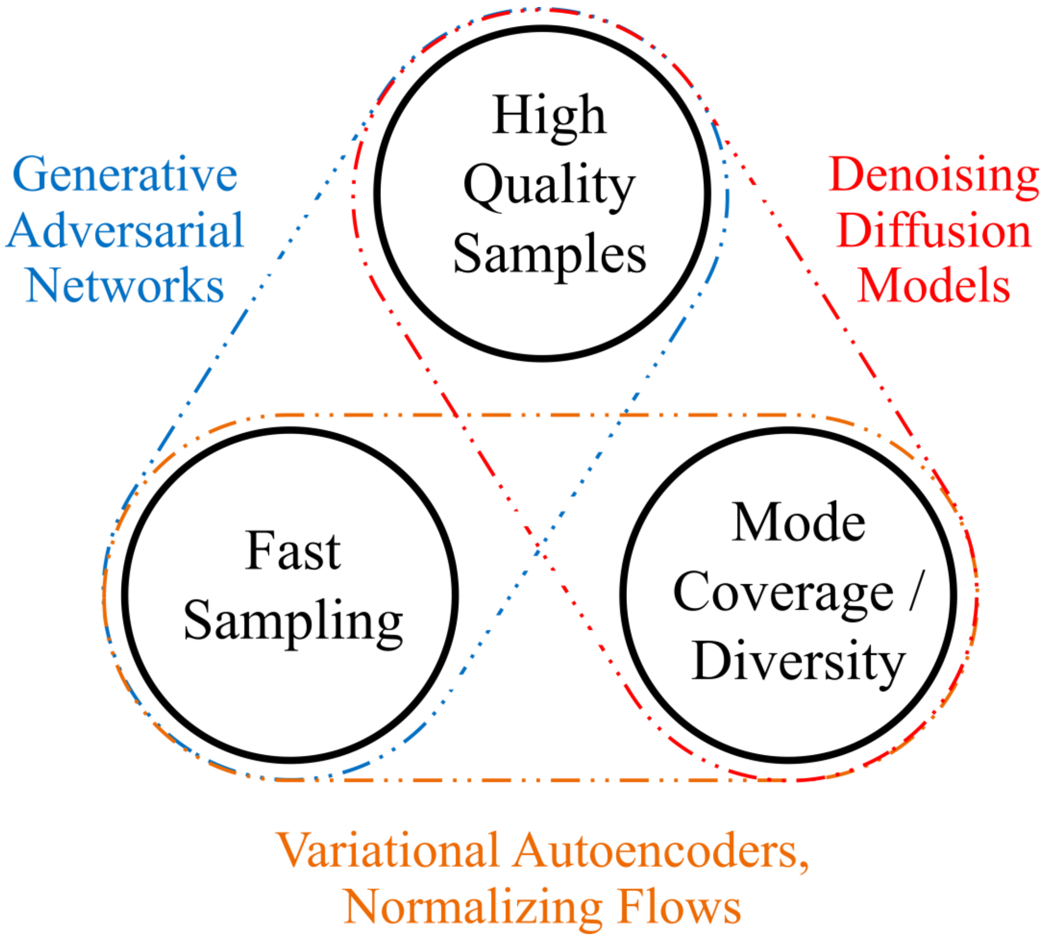
implicit generative models

Require adversarial training:
- notoriously unstable; leading to
- mode collapse

diffusion models bypass both with neat tricks

Figure by Lilian Weng
https://lilianweng.github.io/posts/2021-07-11-diffusion-models/

Sampling Trilemma

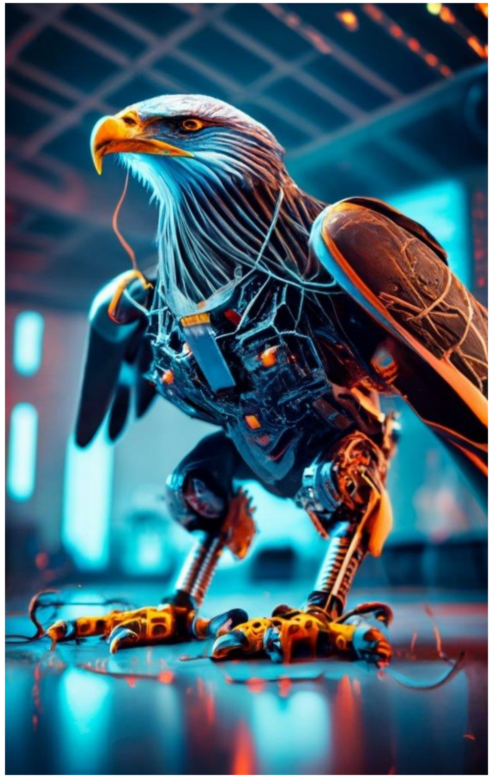


Xiao, Z., Kreis, K., & Vahdat, A. (2021). Tackling the generative learning trilemma with denoising diffusion gans. *ICLR*

Why? Unprecedented Quality

1. Realism
2. Control
3. Prior

“realistic photo of a cybernetic Eagle”



“…
Tribe taking a selfie …”

“A dystopian male face made of volcanic lava, mysterious, image containing secret codes”

Images from ideogram.ai and Midjourney

Images generated by these engines or taken from

Why? Community Push

Companies
Big models and data

OpenAI

Midjourney

stability.ai

IDEOGRAM AI

Open-Source
Ease of Use

Diffusers

MONAI

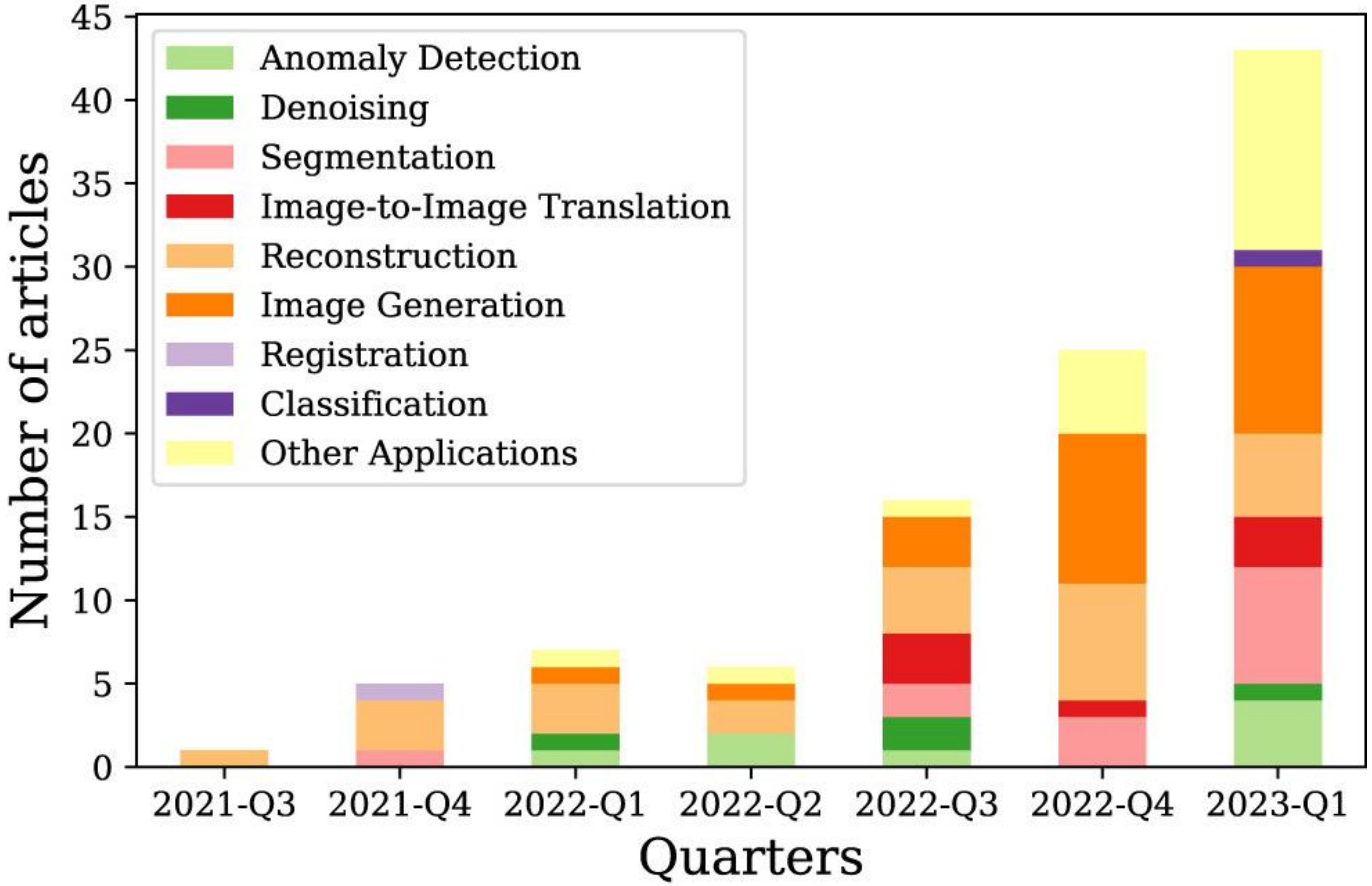
Generative Models

Why? Research Community Push

	ISBI	MICCAI	ICLR	CVPR	ICML	NeurIPS
2024	-	-	**7%**	**12%**	**-**	**-**
2023	**7%**	**7%**	3%	4%	**4%**	**6%**
2022	4%	2%	1%	>1%	1%	1%
2021	1%	>1%	>1%	≈0%	>1%	>1%

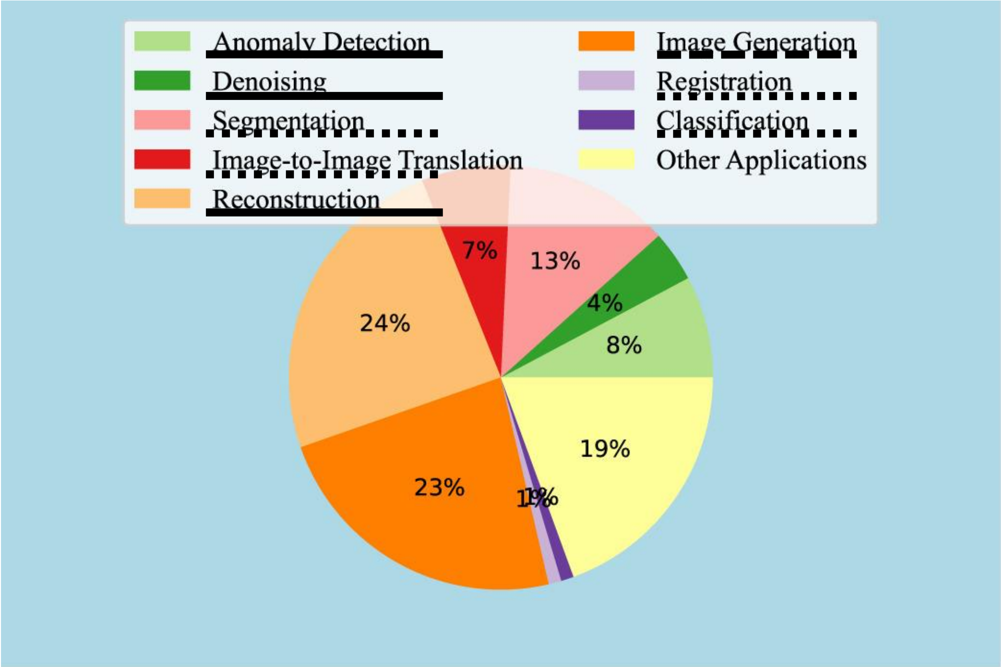
Percentage of accepted papers in great conferences that include diffusion in their title is growing

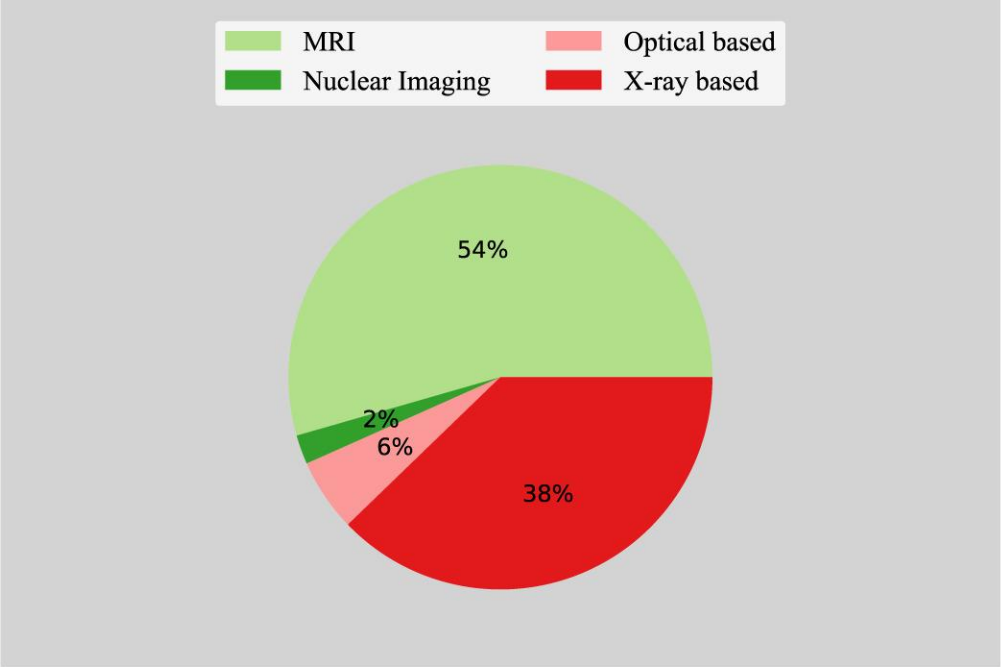
Why? Medical Imaging Popularity



Kazerouni, Amirhossein, et al. "Diffusion models in medical imaging: A comprehensive survey." Medical Image Analysis (2023): 102846.

Why? Medical Imaging Applications





Realism

Control

Prior

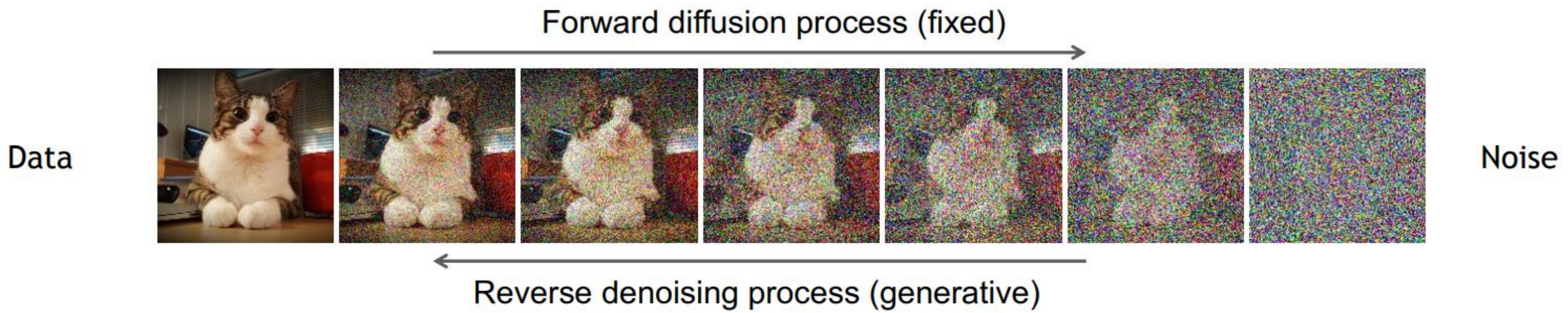
Kazerouni, Amirhossein, et al. "Diffusion models in medical imaging: A comprehensive survey." Medical Image Analysis (2023): 102846.

Diffusion Models
How?

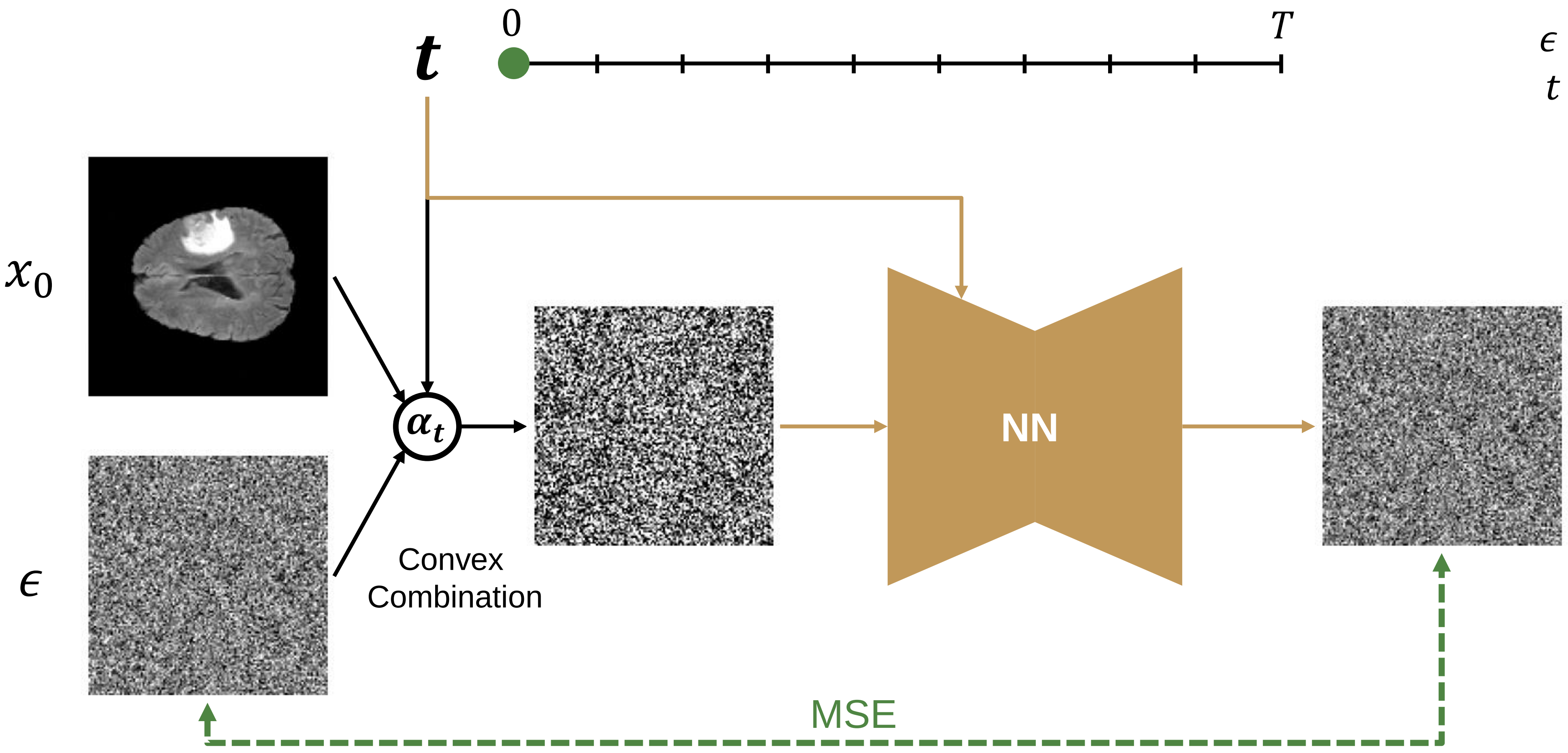
Training by Denoising

How? Training by Denoising

- Add gradually noise
- **Training by denoising**

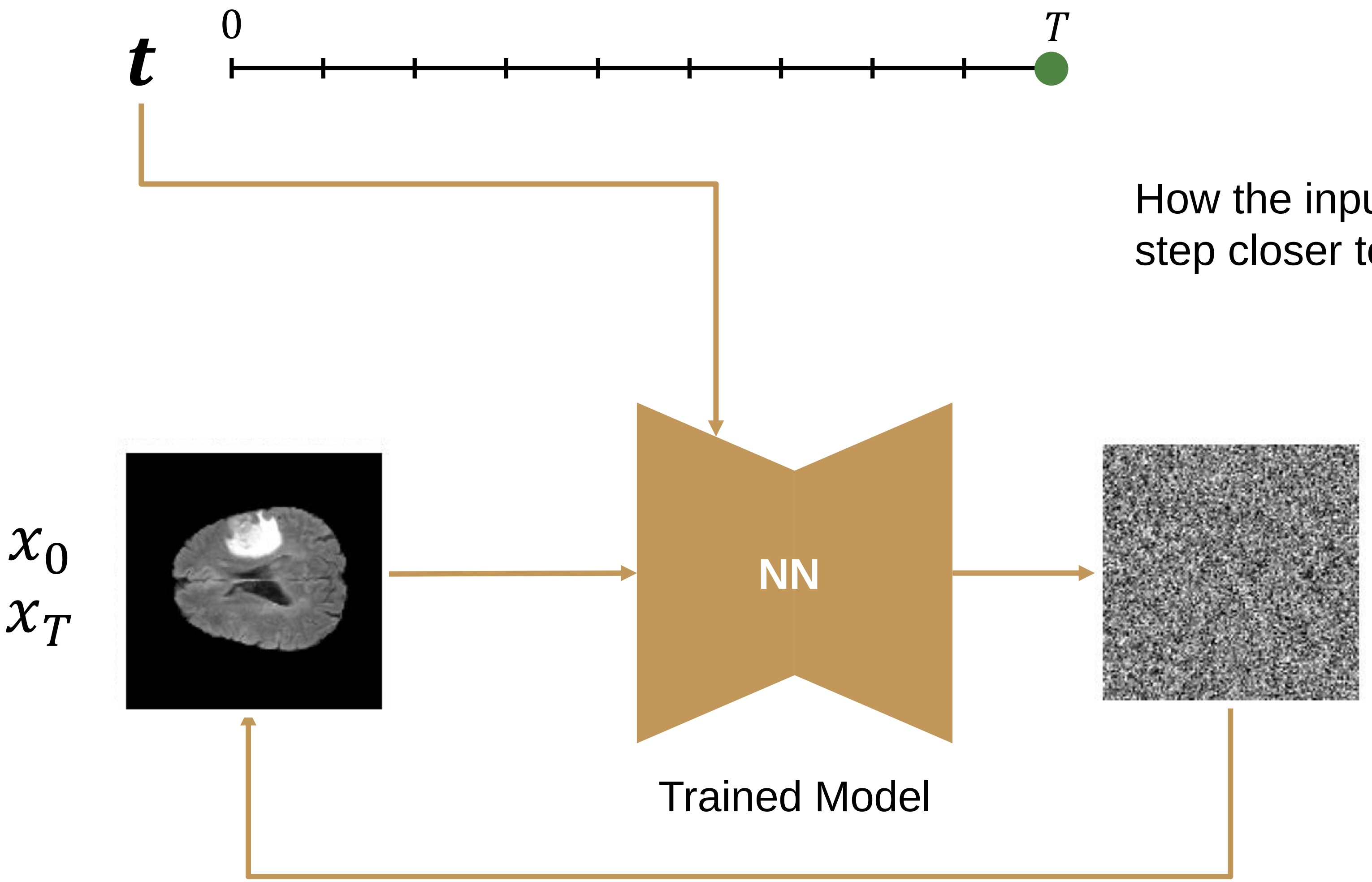


How? Training by Denoising



$\epsilon \sim \mathcal{N}(0, \mathbb{I})$
$t \sim \mathcal{U}(0, T)$

How? Inference



How the input can be a step closer to an image?

Tutorial Schedule

First Half

- ❑Part 1: Introduction (30 mins)
  - ❖ What? Why? How?
  - ❖ Denoising Diffusion Models
  - ❖ Understanding and Intuition
- ❑Part 2: Advanced Topics (30 mins)
  - ❖ Sampling Strategies
  - ❖ Conditioning

We are here!

Second Half

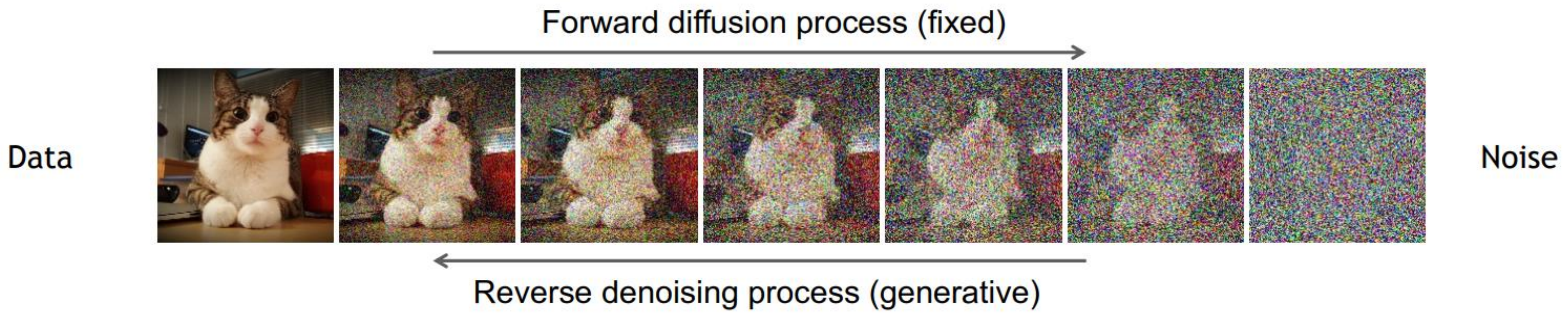
- ❑ Q&A
- ❑ Medical Imaging Applications (30 mins)
  - ❖ Synthesis
  - ❖ Segmentation
  - ❖ Anomaly Detection
  - ❖ Reconstruction
  - ❖ Registration

Denoising Diffusion Model

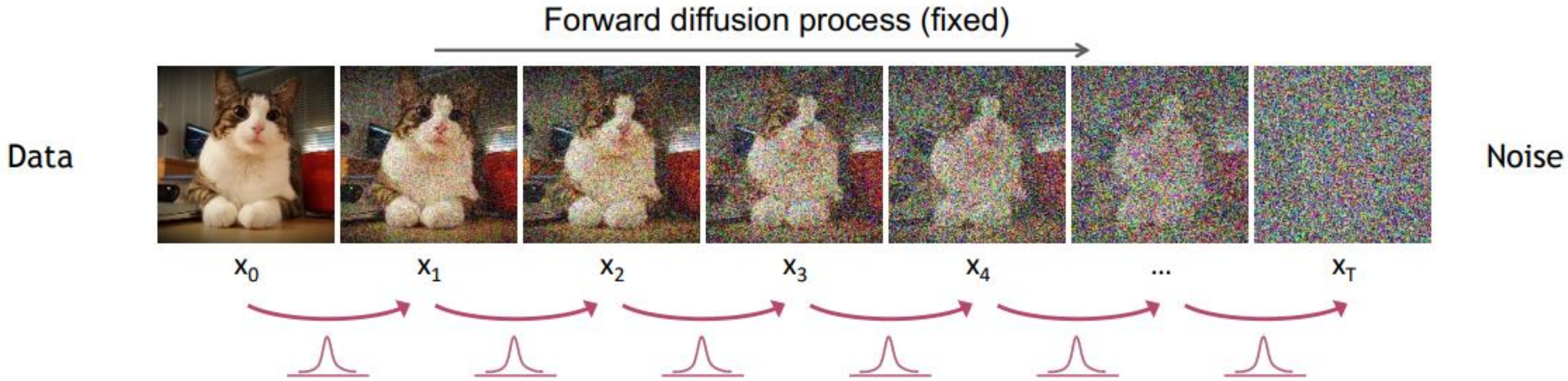
Denoising Diffusion Models

Denoising diffusion models consist of two processes:

- Forward diffusion process that gradually adds noise to input
- Reverse denoising process that learns to generate data by denoising



Forward diffusion process



We use normal distribution to generate a noisy image conditioned on previous image

$$q(x_t|x_{t-1}) = N\left(x_t; \sqrt{\beta_t} x_{t-1}, (1-\beta_t)I\right)$$

Forward diffusion process

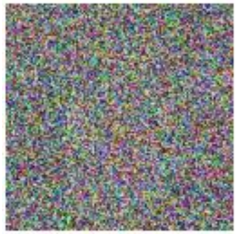
joint distribution → $q(x_{1:T}|x_0) = \prod_{t=1}^{T} q(x_t|x_{t-1})$

Given the step β_s, we can generate a particular step t

$q(x_t|x_0) = N\left(x_t; \sqrt{\bar{a}_t}x_0, (1-\bar{a}_t)I\right)$ where $\bar{a}_t = \prod_{s=1}^{t}(1-\beta_s)$

diffusion kernel

For sampling at timestep t : $x_t = \sqrt{\bar{a}_t}x_0 + (1-\bar{a}_t)\varepsilon$ where $\epsilon \sim N(0, I)$



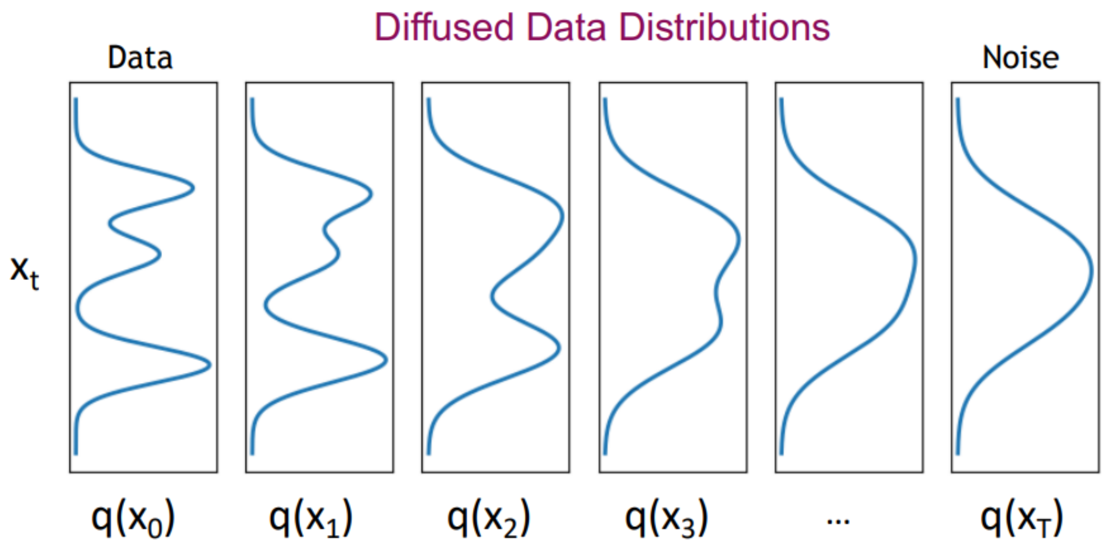
β_t is the noise schedule such that $\bar{a}_t \to 0 \quad \Rightarrow \quad q(x_T|x_0) \approx N(x_T; 0, I)$

The diffusion kernel at timestep T can be approximated using **standard normal distribution**

Forward diffusion process

The diffusion kernel $q(x_t|x_0)$ is Gaussian convolution

$$q(x_t) = \int q(x_0, x_t)\,\mathrm{d}x_0 = \int q(x_0)\underbrace{q(x_t|x_0)}_{\text{diffusion kernel}}\,\mathrm{d}x_0$$



We can sample $x_t \sim q(x_t)$
by first sampling $x_0 \sim q(x_0)$ and then sampling $x_t \sim q(x_t|x_0)$

Generation by Denoising

Recall $q(x_T) \sim N(x_T; 0, I)$

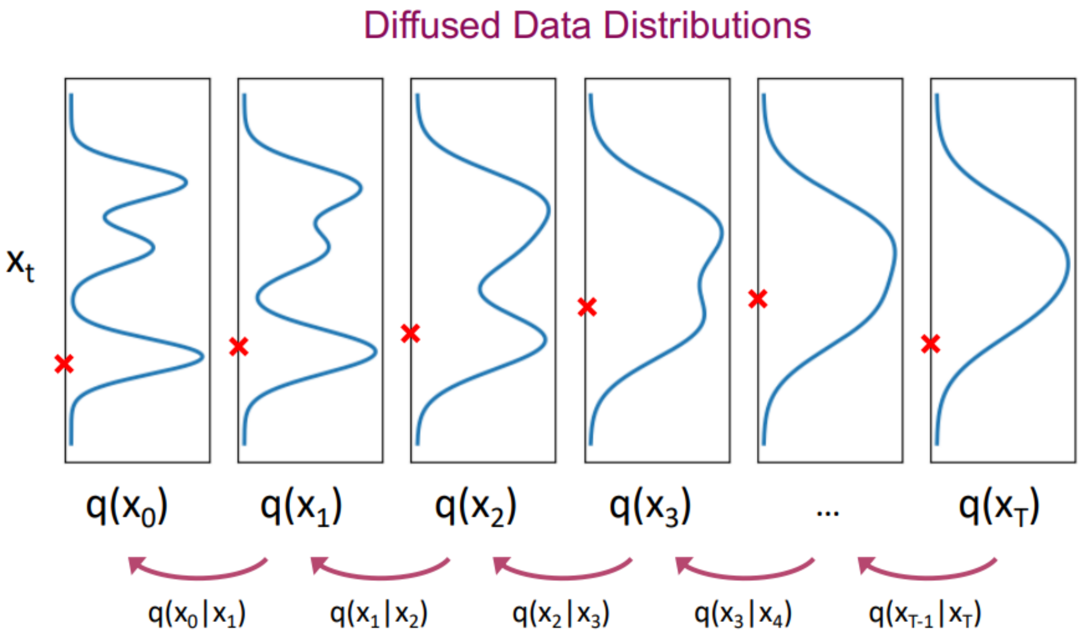
Start from $x_T \sim N(x_T; 0, I)$

Iteratively $x_{t-1} \sim q(x_{t-1}|x_t)$

But $q(x_{t-1}|x_t)$ is **intractable**!

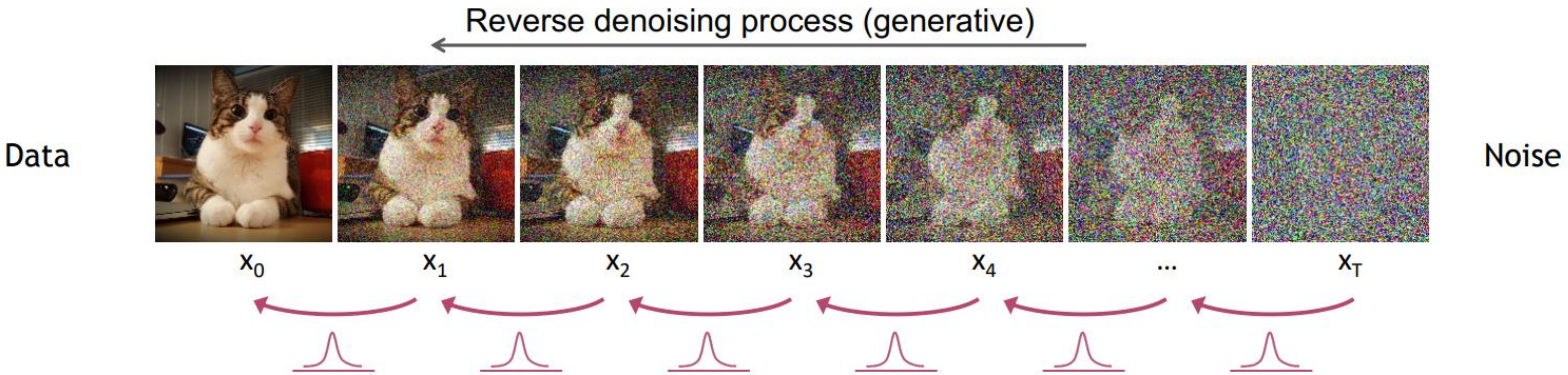
We can approximate it with a normal distribution.

if β_t is small in forward process



We need a **parametric model** to mimic $q(x_{t-1}|x_t)$

Reverse diffusion process



We define a denoising distribution $p_\theta(x_{t-1}|x_t)$

Reverse diffusion process

$$p(x_T) = N(x_T; 0, I)$$

$$p_\theta(x_{t-1}|x_t) = N(x_{t-1}; \underbrace{\mu_\theta}(x_t, t), \sigma_t^2 I) \quad \rightarrow \quad \underbrace{p_\theta(x_{0:T})} = p(x_T) \prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)$$

μ_θ is a trainable network (like U-net)

joint distr. of full reverse trajectory

For training, we form **variational upper bound (ELBO)** that is commonly used for training variational autoencoders:

$$E_{q(x_0)}[-\log p_\theta(x_0)] \leq E_{q(x_0)q(x_{1:\mathrm{T}}|x_0)}\left[-\log \frac{p_\theta(x_{0:T})}{q(x_{1:\mathrm{T}}|x_0)}\right] =: ELBO$$

Reverse diffusion process

$$E_{q(x_0)}[-\log p_\theta (x_0)] \leq E_{q(x_0)q(x_{1:\mathrm{T}}|x_0)}[-\log \frac{p_\theta(x_{0:T})}{q(x_{1:\mathrm{T}}|x_0)}] =: ELBO$$

recall $x_t = \sqrt{\bar{a}_t} x_0 + (1 - \bar{a}_t)\epsilon$

After some simple arithmetic operations, the variational objective is:

$$ELBO = E_{x_o \sim q(x_0), t \sim \cup\{1,\mathrm{T}\}, \epsilon \sim N(0,I)}[\lambda_\mathrm{t} || \epsilon - \epsilon_\theta(\overbrace{\sqrt{\bar{a}_t} x_0 + (1 - \bar{a}_t)\epsilon}^{x_t}, \mathrm{t})||^2]$$

λ_t is a function of β_t

What does the diffusion model optimise?

How to participate?

Copy participation link

1 Go to wooclap.com

2 Enter the event code in the top banner

Event code
FROCKY

Enable answers by SMS

Tutorial Schedule

First Half

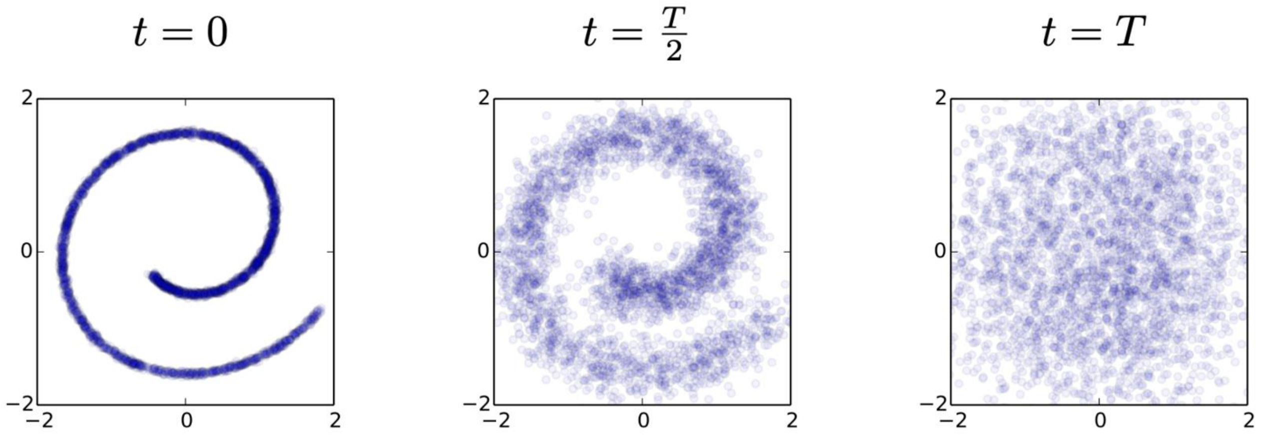
- ❑ Part 1: Introduction (30 mins)
  - ❖ What? Why? How?
  - ❖ Denoising Diffusion Models
  - ❖ Understanding and Intuition

We are here!

- ❑ Part 2: Advanced Topics (30 mins)
  - ❖ Sampling Strategies
  - ❖ Conditioning

Second Half

- ❑ Q&A
- ❑ Medical Imaging Applications (30 mins)
  - ❖ Synthesis
  - ❖ Segmentation
  - ❖ Anomaly Detection
  - ❖ Reconstruction
  - ❖ Registration



Understanding and Intuition

Score Function

Likelihood models want to learn $p_\theta(\mathrm{x})$ directly

Diffusion models want to learn the score $\nabla_{\mathrm{x}} \log p_\theta(\mathrm{x})$

(i.e. the gradient with respect to the input of the loglikelihood)

a distribution can be written as:

$$p_\theta(\mathrm{x}) = \frac{e^{-f_\theta(x)}}{Z_\theta}$$

Score Function

a distribution can be written as:

$$p_\theta(\mathrm{x}) = \frac{e^{-f_\theta(x)}}{Z_\theta}$$

$$\log p_\theta(\mathrm{x}) = \log e^{-f_\theta(x)} - \log Z_\theta$$

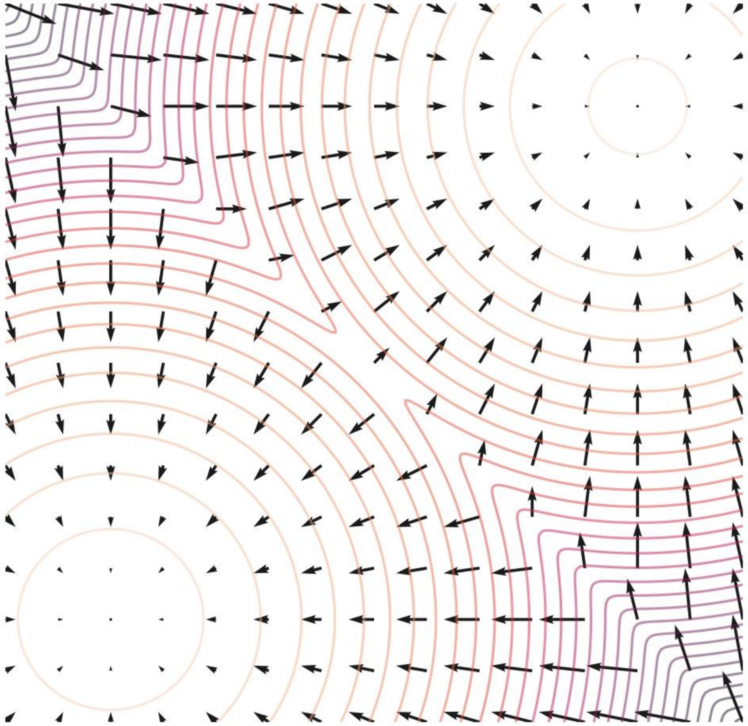
$$\nabla_{\mathrm{x}} \log p_\theta(\mathrm{x}) = -\nabla_{\mathrm{x}} f_\theta(x) - \cancelto{0}{\nabla_{\mathrm{x}} \log Z_\theta}$$

$$\approx$$

$$\boldsymbol{\epsilon_\theta}$$

How to learn it?

Mixture of two Gaussians
Score function (the vector field)
Density function (contours)



The score is pointing to the areas of biggest mass

Image from blog post by Yang Song https://yang-song.net/blog/2021/score/

Denoising Score Matching

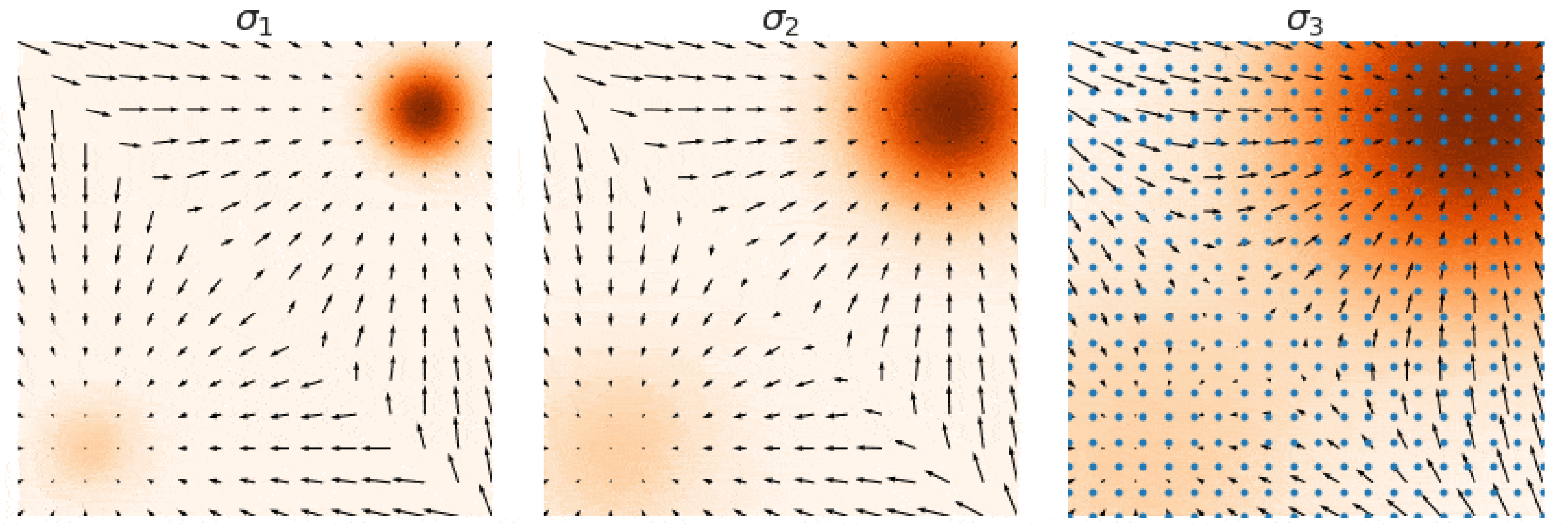
How to **learn** the score?

$$\mathbb{E}_{q(\mathrm{x})} \left\| \underbrace{\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\mathrm{x})}_{\text{Diffusion Model}} - \underbrace{\nabla_{\mathrm{x}} \log p_{\theta}(\mathrm{x})}_{\text{Score}} \right\|_{\mathbf{2}}^{\mathbf{2}}$$

$$\mathbb{E}_{q(\mathrm{x})} \| \boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\mathrm{x}) - \nabla_{\mathrm{x}} \log q(\mathrm{x}_t \mid \mathrm{x}) \|_{\mathbf{2}}^{\mathbf{2}}$$

Vincent, Pascal. "A connection between score matching and denoising autoencoders." Neural computation 23.7 (2011): 1661-1674.

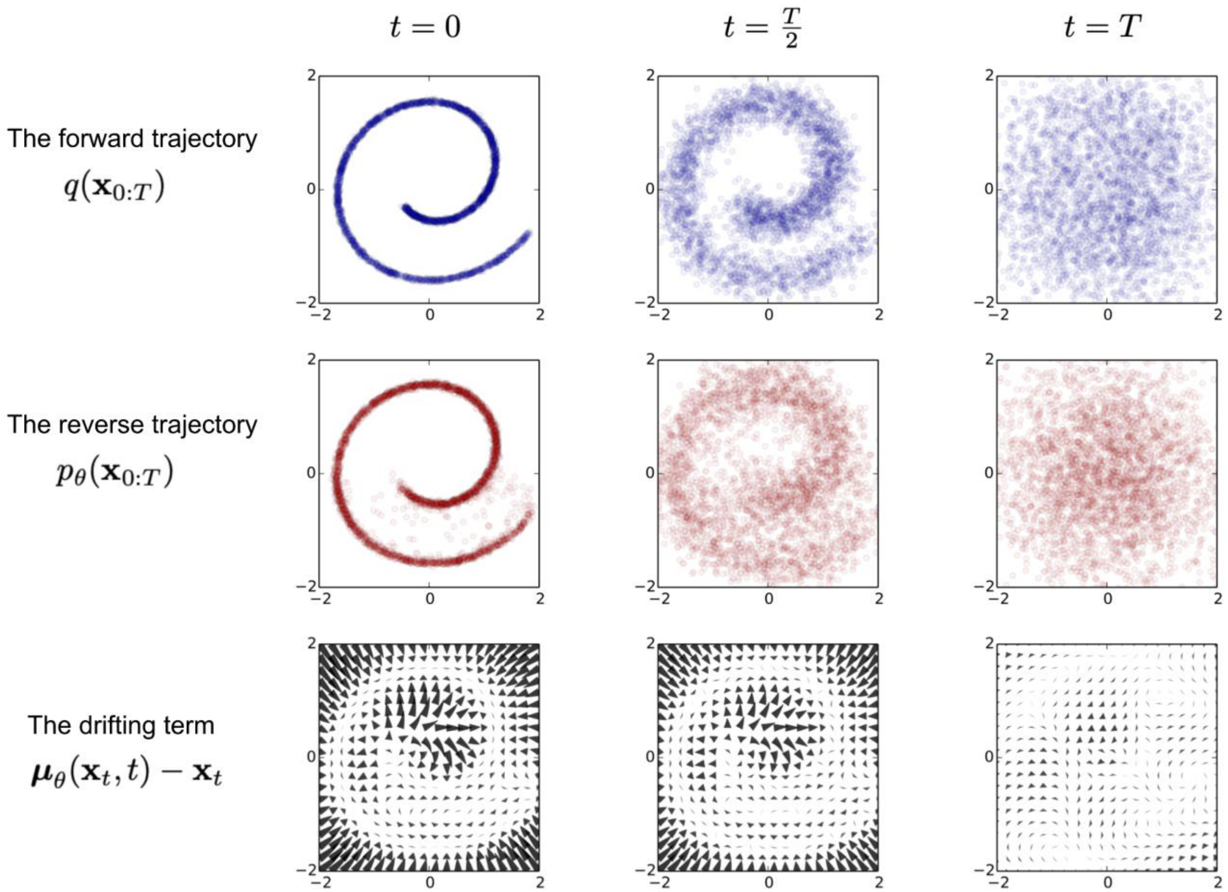
Perturbation at many scales



Learning in **low** density regions

Image from blog post by Yang Song https://yang-song.net/blog/2021/score/

Diffusion Models Learn the Gradient

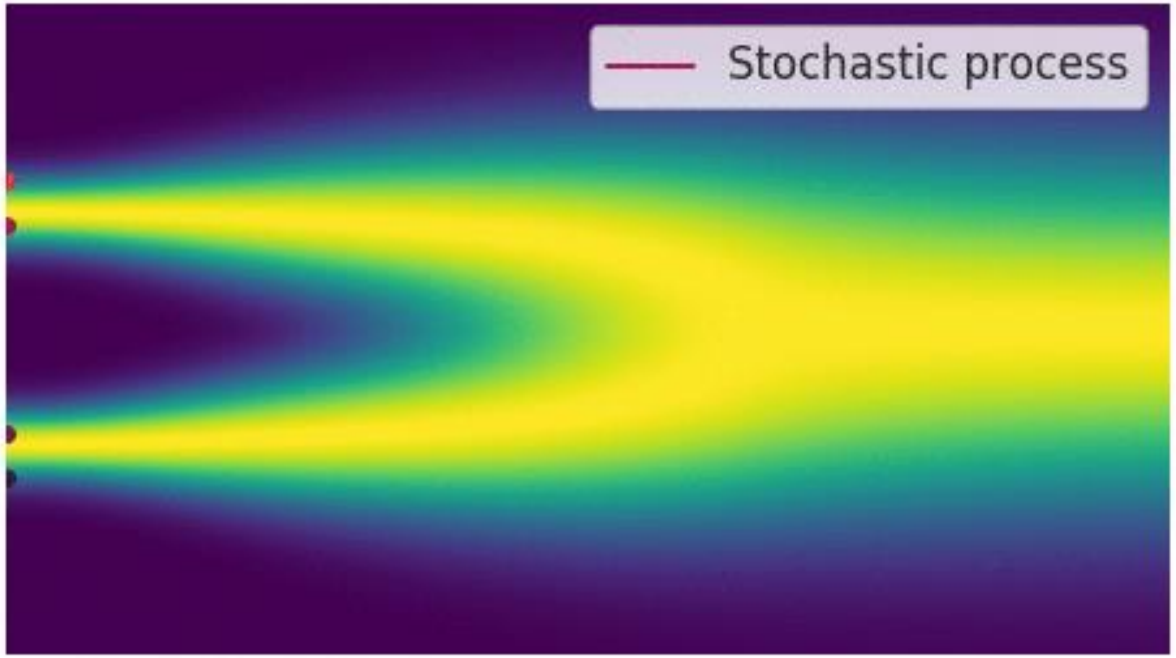


$$\nabla_x log\ \mathrm{p}(\mathrm{x})$$

Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., & Ganguli, S. (2015). Deep unsupervised learning using nonequilibrium thermodynamics. In International Conference on Machine Learning.

Diffusion and Differential Equations

- Perturbation process is a Stochastic Differential Equation (SDE)
- From complex to simple
- Allow different values for SDE modelling

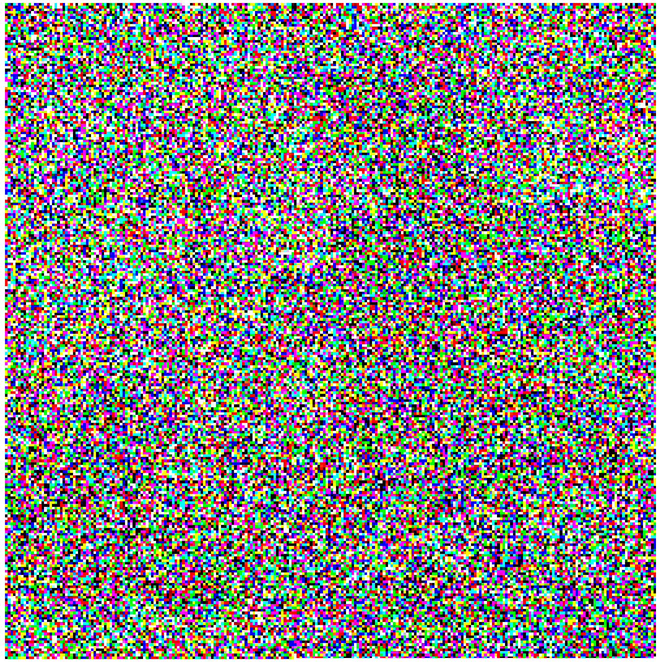


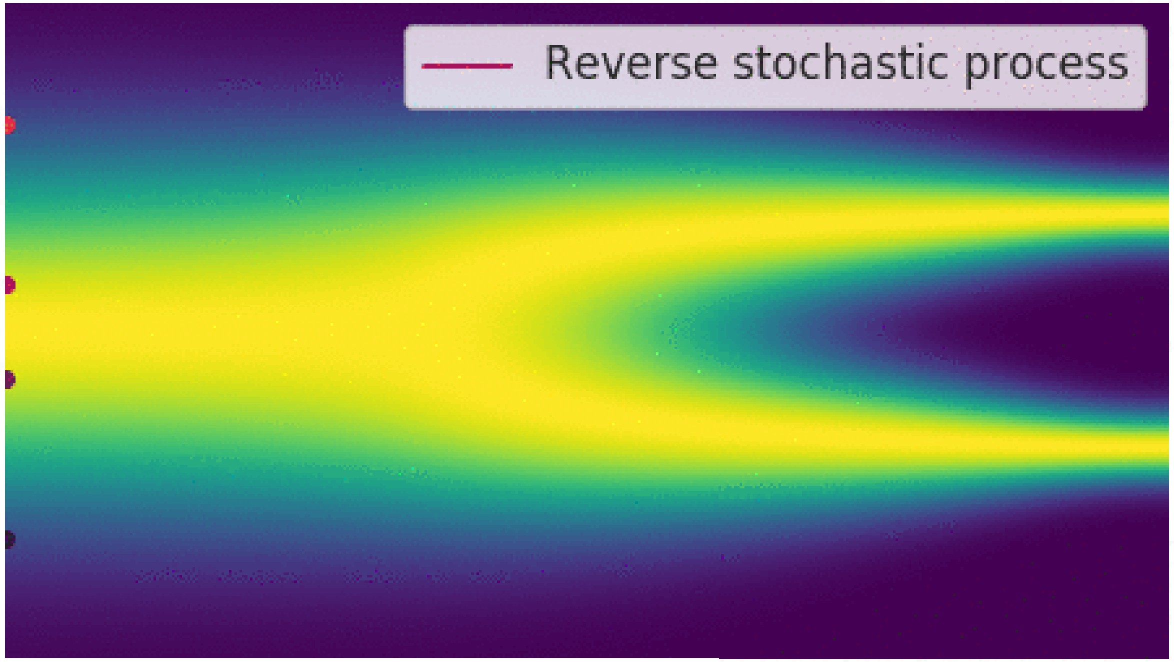
$$\mathrm{d}\mathbf{x} = \mathbf{f}(\mathbf{x}, t)\mathrm{d}t + g(t)\mathrm{d}\mathbf{w}.$$

Image from blog post by Yang Song https://yang-song.net/blog/2021/score/

Reversing the Process is Generation

- Samplers are discrete solutions of the reverse-time SDE





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

Image from blog post by Yang Song https://yang-song.net/blog/2021/score/

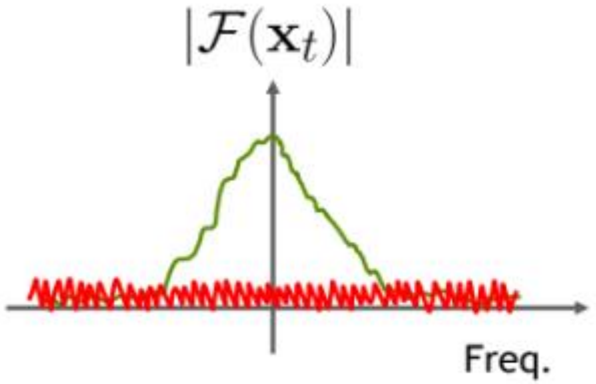
Fourier Transform

$$x_t = \sqrt{\alpha_t}\, x + \sqrt{1 - \alpha_t}\, \epsilon, \qquad \epsilon \sim \mathcal{N}(0, \mathbf{I})$$

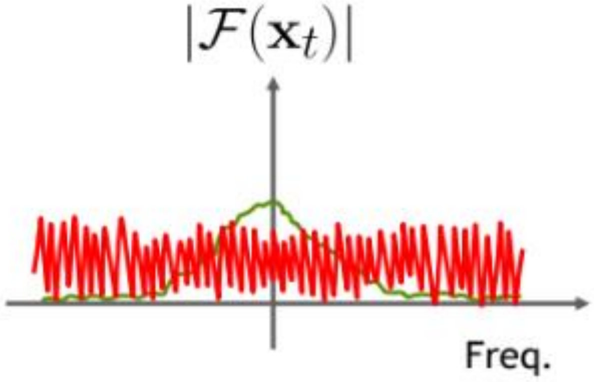
$$\Big\downarrow \text{ ***Fourier Transform***}$$

$$\mathcal{F}(x_t) = \sqrt{\alpha_t}\mathcal{F}(x) + \sqrt{1 - \alpha_t}\mathcal{F}(\epsilon)$$

Small t
$\bar{a}_t \sim 1$

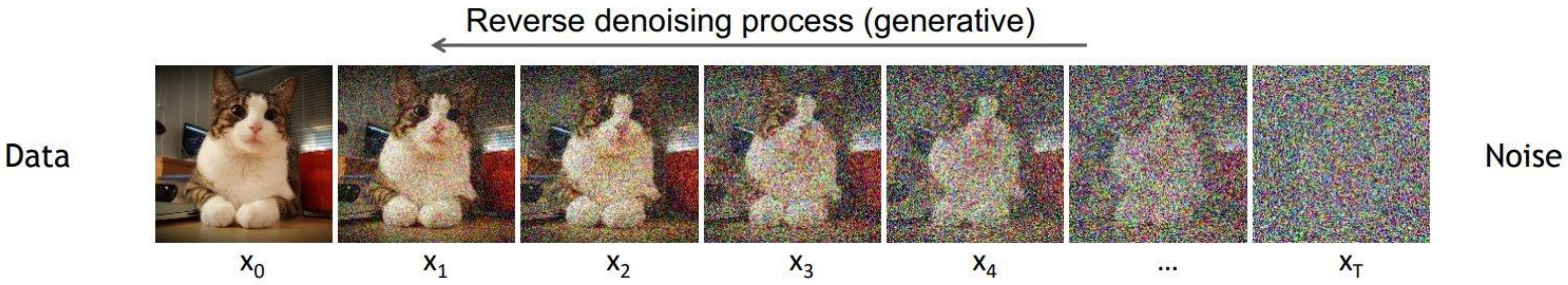


Big t
$\bar{a}_t \sim 0$



Slide inspired in CVPRs 2022 tutorial on diffusion models

Content – Detail Tradeoff

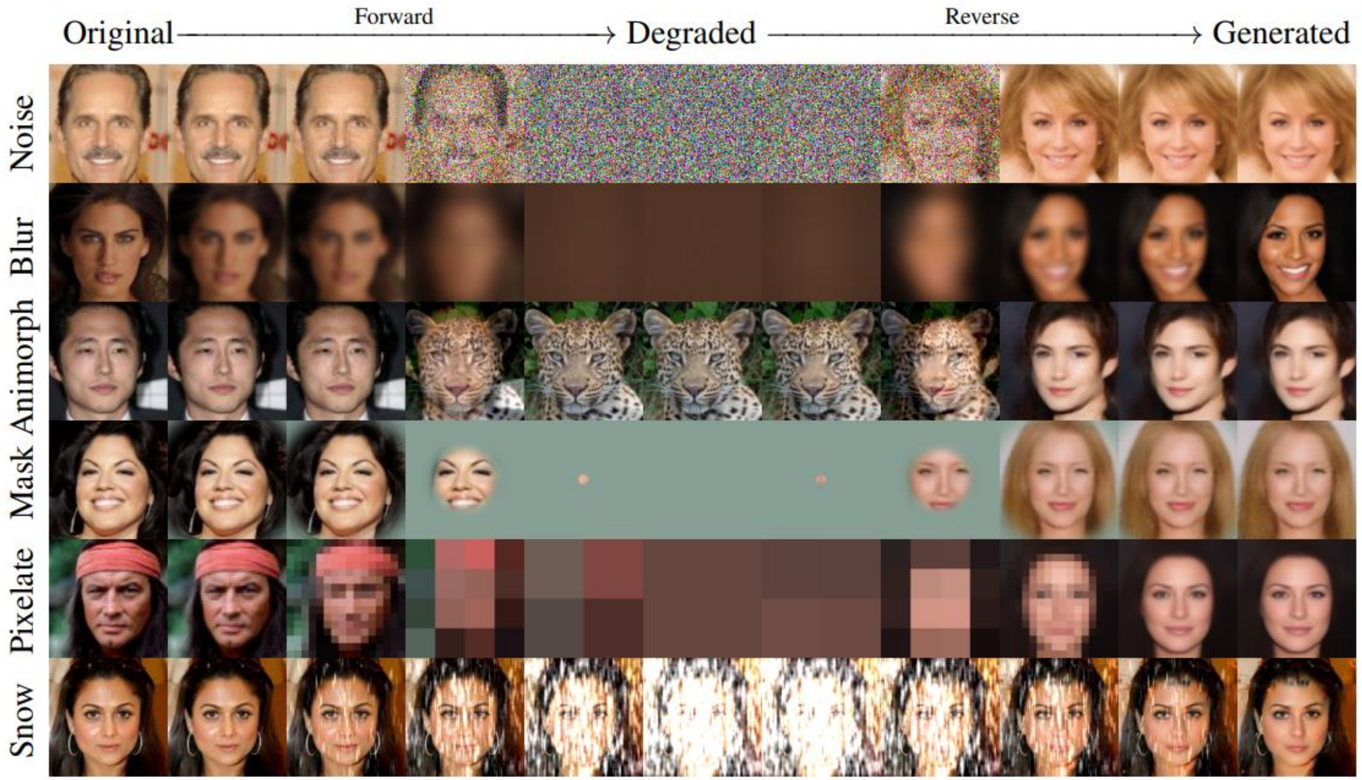


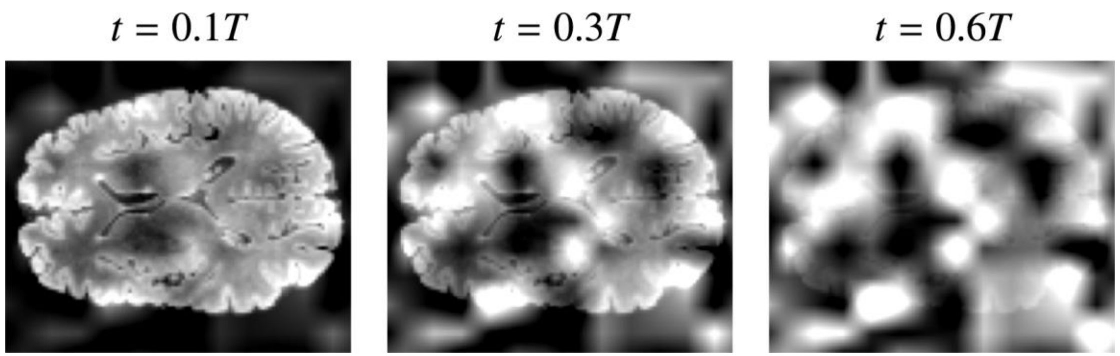
$\underbrace{\hphantom{x_0 \quad x_1 \quad x_2}}_{}$

Small t
The model generates high frequency content
(low-level details)

Big t
The model generates low frequency content
(coarse details)

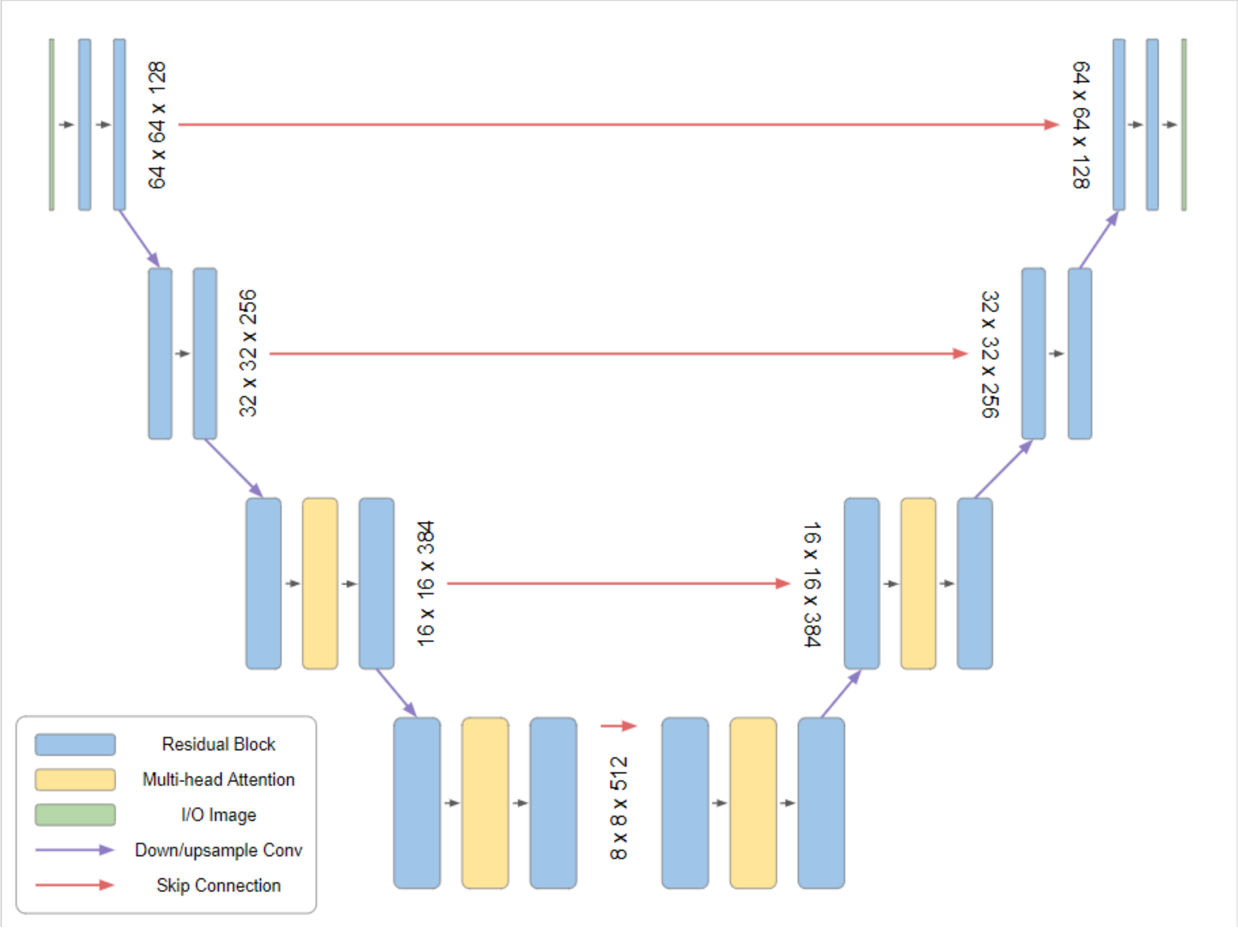
Gaussian Perturbation?





[1] Daras, Giannis, et al. "Soft diffusion: Score matching for general corruptions." arXiv preprint arXiv:2209.05442 (2022).
[2] Bansal, Arpit, et al. "Cold diffusion: Inverting arbitrary image transforms without noise." arXiv preprint arXiv:2208.09392 (2022).
[3] Kascenas, Antanas, et al. "The role of noise in denoising models for anomaly detection in medical images." Medical Image Analysis (2023): 102963.

Architecture – Reusing the *classics*, and the *SoTA*



Jonathan Ho, Ajay Jain, Pieter Abbeel (2020) Denoising Diffusion Probabilistic Models. NeuriPS

Unet!

Or transformers
Or VQ-VAEs
Or…

Figure by the author of the papers.

How do you think a diffusion model might compare to GANs in terms of output quality and stability?

How to participate?

1. Go to wooclap.com
2. Enter the event code in the top banner

Event code
FROCKY

Enable answers by SMS

Copy participation link

Tutorial Schedule

First Half

- ❑ Part 1: Introduction (30 mins)
  - ❖ What? Why? How?
  - ❖ Denoising Diffusion Models
  - ❖ Understanding and Intuition
- ❑ Part 2: Advanced Topics (30 mins)
  - ❖ Sampling Strategies
  - ❖ Conditioning

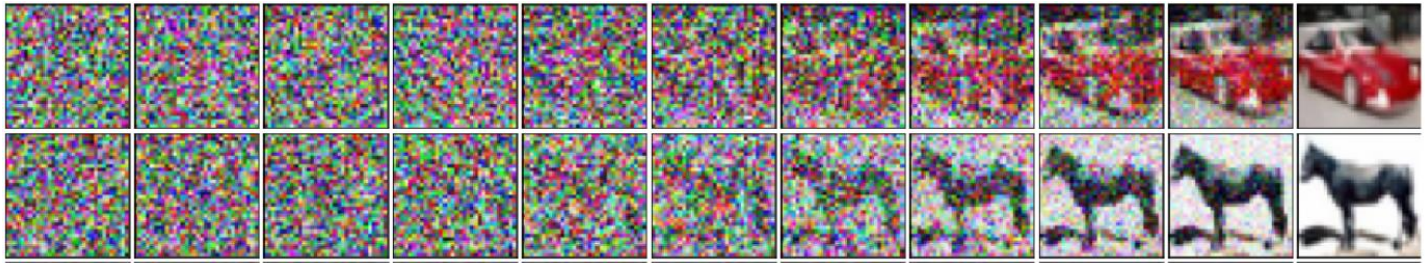
We are here!

Second Half

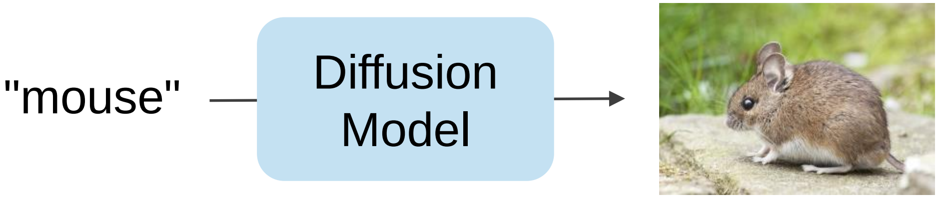
- ❑ Q&A
- ❑ Medical Imaging Applications (30 mins)
  - ❖ Synthesis
  - ❖ Segmentation
  - ❖ Anomaly Detection
  - ❖ Reconstruction
  - ❖ Registration

Part 2 – Advanced Topics

- Sampling Strategies

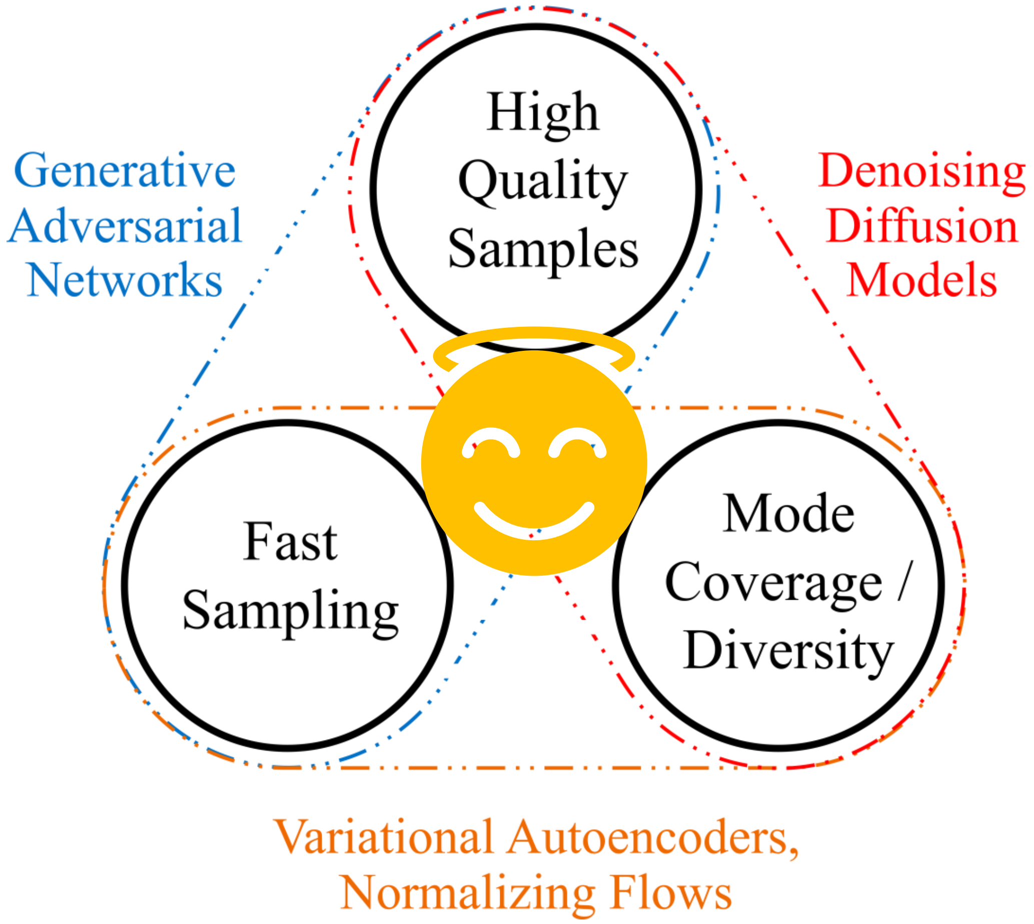


- Conditioning Mechanisms



An inherent drawback: slowness

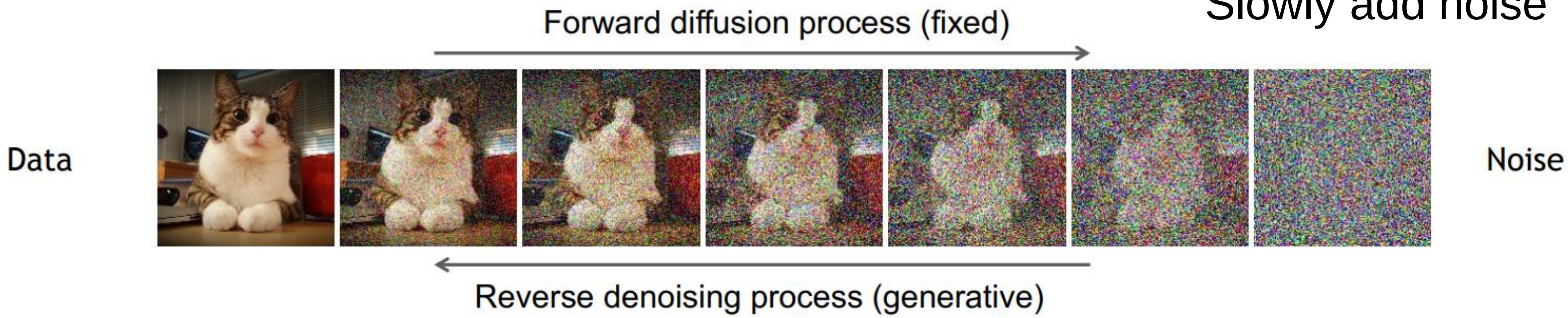
- Facts about DMs
  - Training
    - 150-1000 V100 GPU days
  - Sampling
    - 50k samples, 5 A100 GPU days



Xiao, Z., Kreis, K., & Vahdat, A. (2021). Tackling the generative learning trilemma with denoising diffusion gans. *ICLR*

How to accelerate diffusion models?

Slowly add noise



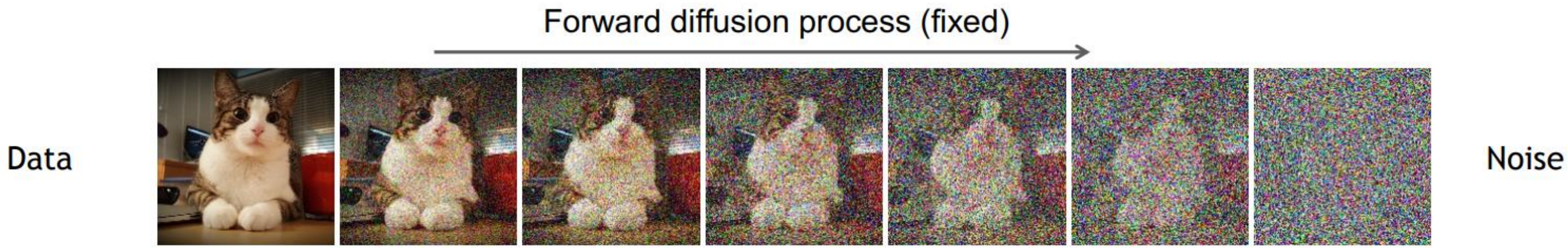
Map noise back to data

Naïve acceleration methods

- such as reducing diffusion time steps in training or sampling
    - Leading to immediate worse performance.

We need something clever!

Acceleration: Forward Process



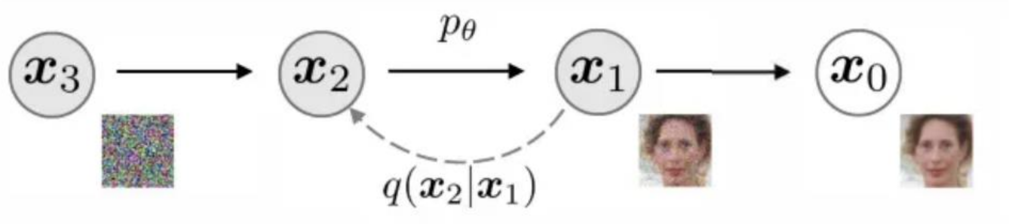
$$\mathbf{x}_0 \rightarrow \cdots \mathbf{x}_t \rightarrow \mathbf{x}_{t+1} \rightarrow \cdots \rightarrow \mathbf{x}_T$$

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

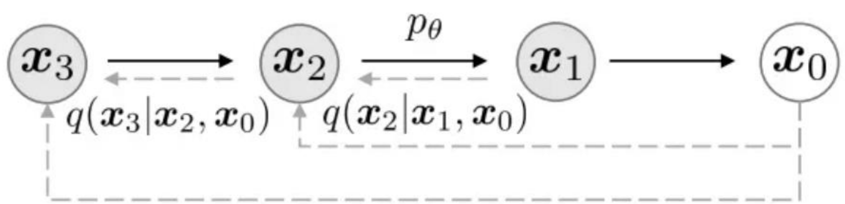
- Does it have to be a Markovian process?
- Is there any faster diffusion process?

Denoising Diffusion Implicit Models (DDIM)

- Diffusion Model does not need to be Markovian!
  - DDPM forward process: $q_\sigma(\mathbf{x}_t|\mathbf{x}_{t-1})$
- The loss function for DDPM: $L(\theta) = [||\epsilon - \epsilon_\theta(\mathbf{x}_t, t)||^2]$

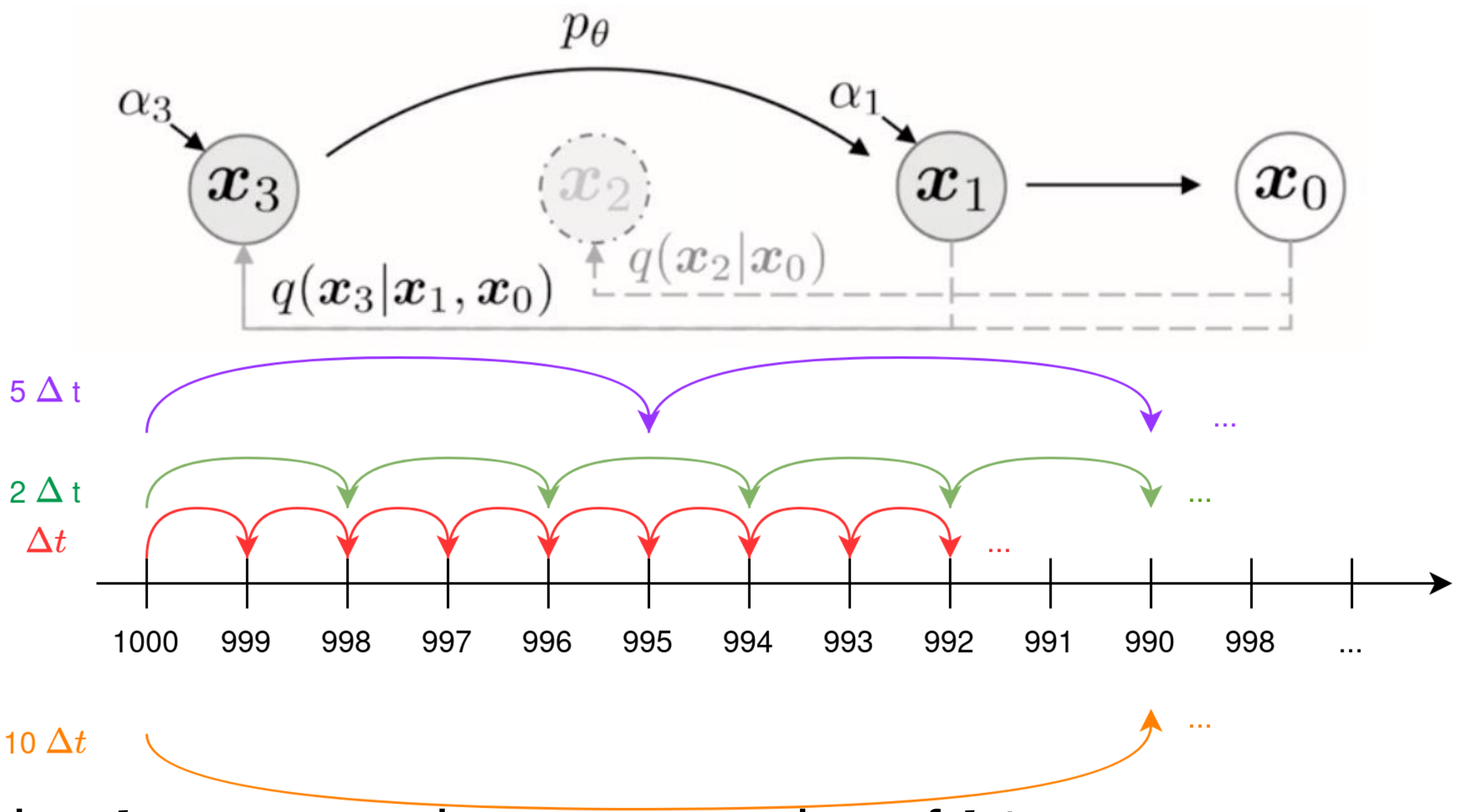


- Define a new forward process: $q_\sigma(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{x}_0)$



Song, J., Meng, C., & Ermon, S. (2020). Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*.

Acceleration: Skip steps



- By skipping k steps, we have a step size of $k\Delta t$.
- Sampling is k times faster.
- We trade image quality for speed.

Song, J., Meng, C., & Ermon, S. (2020). Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*.

DDIM Result

- DDIM: Non-Markovian process but 10-50x faster!!
  - We can directly use a pretrained DDPM to do sampling acceleration

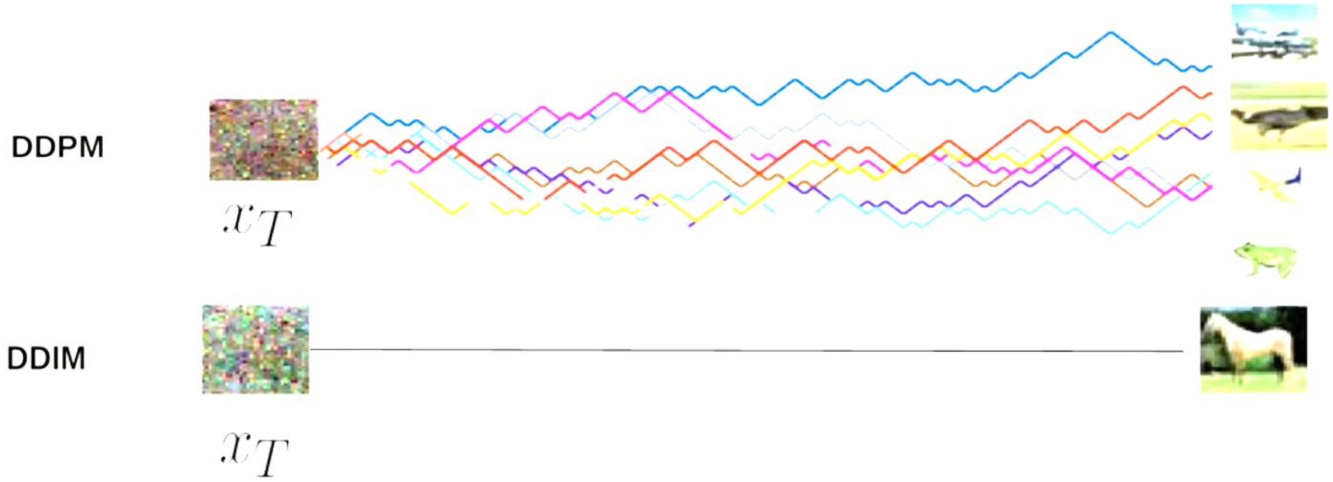
Table 1: CIFAR10 and CelebA image generation measured in FID. $\eta = 1.0$ and $\hat{\sigma}$ are cases of DDPM (although Ho et al. (2020) only considered $T = 1000$ steps, and $S < T$ can be seen as simulating DDPMs trained with S steps), and $\eta = 0.0$ indicates DDIM.

		CIFAR10 (32 × 32)					CelebA (64 × 64)				
	S	10	20	50	100	1000	10	20	50	100	1000
η	0.0	**13.36**	**6.84**	**4.67**	**4.16**	4.04	**17.33**	**13.73**	**9.17**	**6.53**	3.51
	0.2	14.04	7.11	4.77	4.25	4.09	17.66	14.11	9.51	6.79	3.64
	0.5	16.66	8.35	5.25	4.46	4.29	19.86	16.06	11.01	8.09	4.28
	1.0	41.07	18.36	8.01	5.78	4.73	33.12	26.03	18.48	13.93	5.98
$\hat{\sigma}$		367.43	133.37	32.72	9.99	**3.17**	299.71	183.83	71.71	45.20	**3.26**

Song, J., Meng, C., & Ermon, S. (2020). Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*.

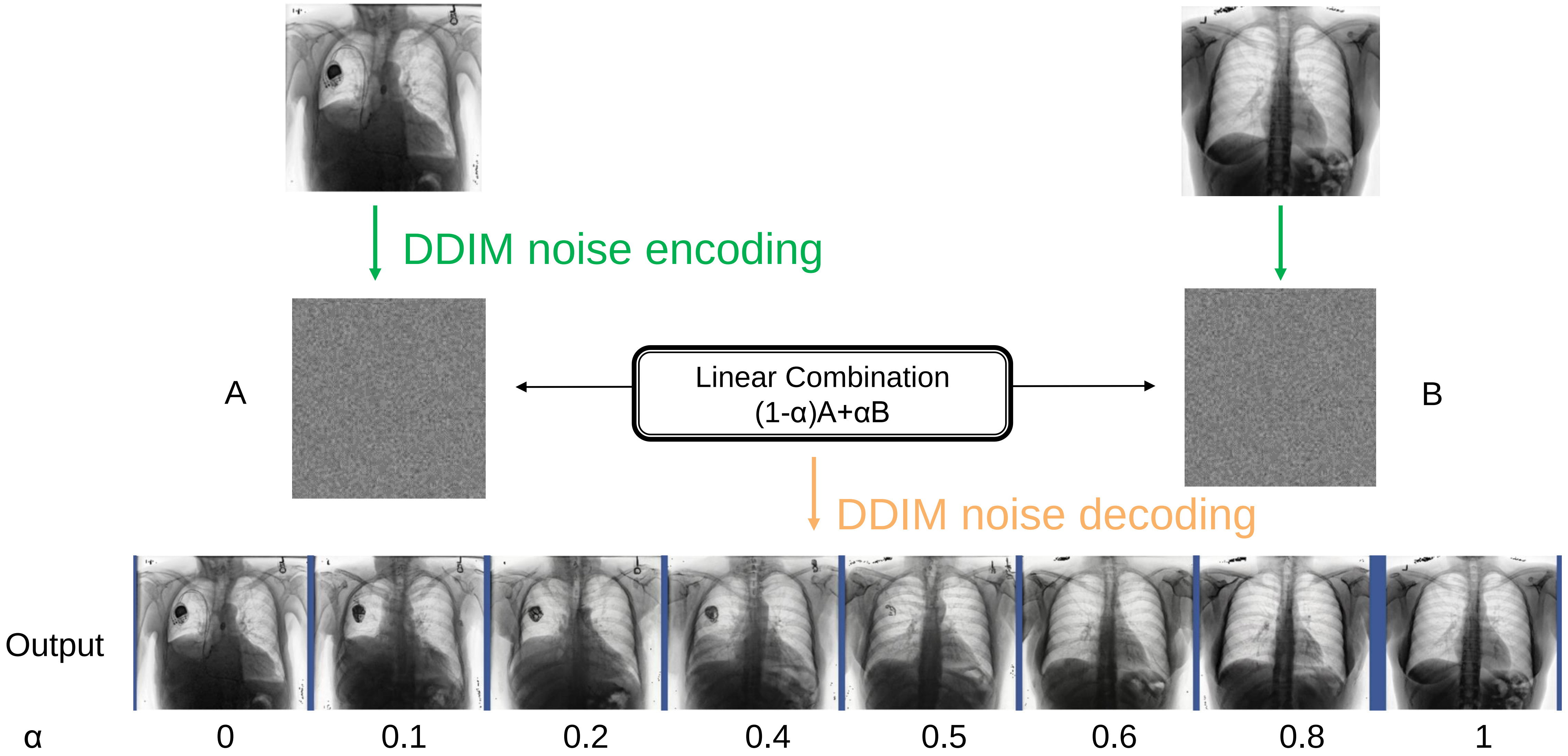
Denoising Diffusion Implicit Models (DDIM)

$$\boldsymbol{x}_{t-1} = \sqrt{\alpha_{t-1}} \underbrace{\left(\frac{\boldsymbol{x}_t - \sqrt{1-\alpha_t}\,\epsilon_\theta^{(t)}(\boldsymbol{x}_t)}{\sqrt{\alpha_t}} \right)}_{\text{" predicted } \boldsymbol{x}_0\text{"}} + \underbrace{\sqrt{1-\alpha_{t-1}-\sigma_t^2}\cdot\epsilon_\theta^{(t)}(\boldsymbol{x}_t)}_{\text{"direction pointing to } \boldsymbol{x}_t\text{"}} + \boxed{\underbrace{\sigma_t\epsilon_t}_{\text{random noise}}}$$

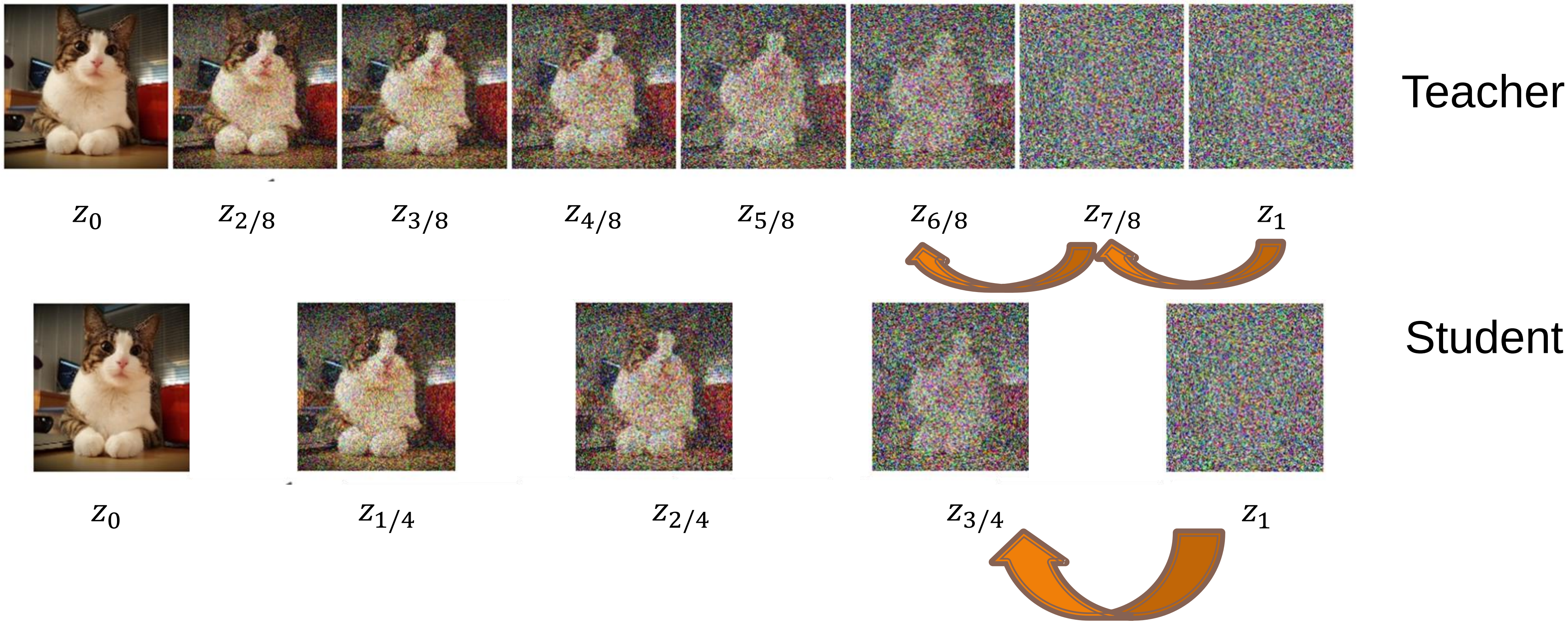


Song, J., Meng, C., & Ermon, S. (2020). Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*.

Image Interpolation

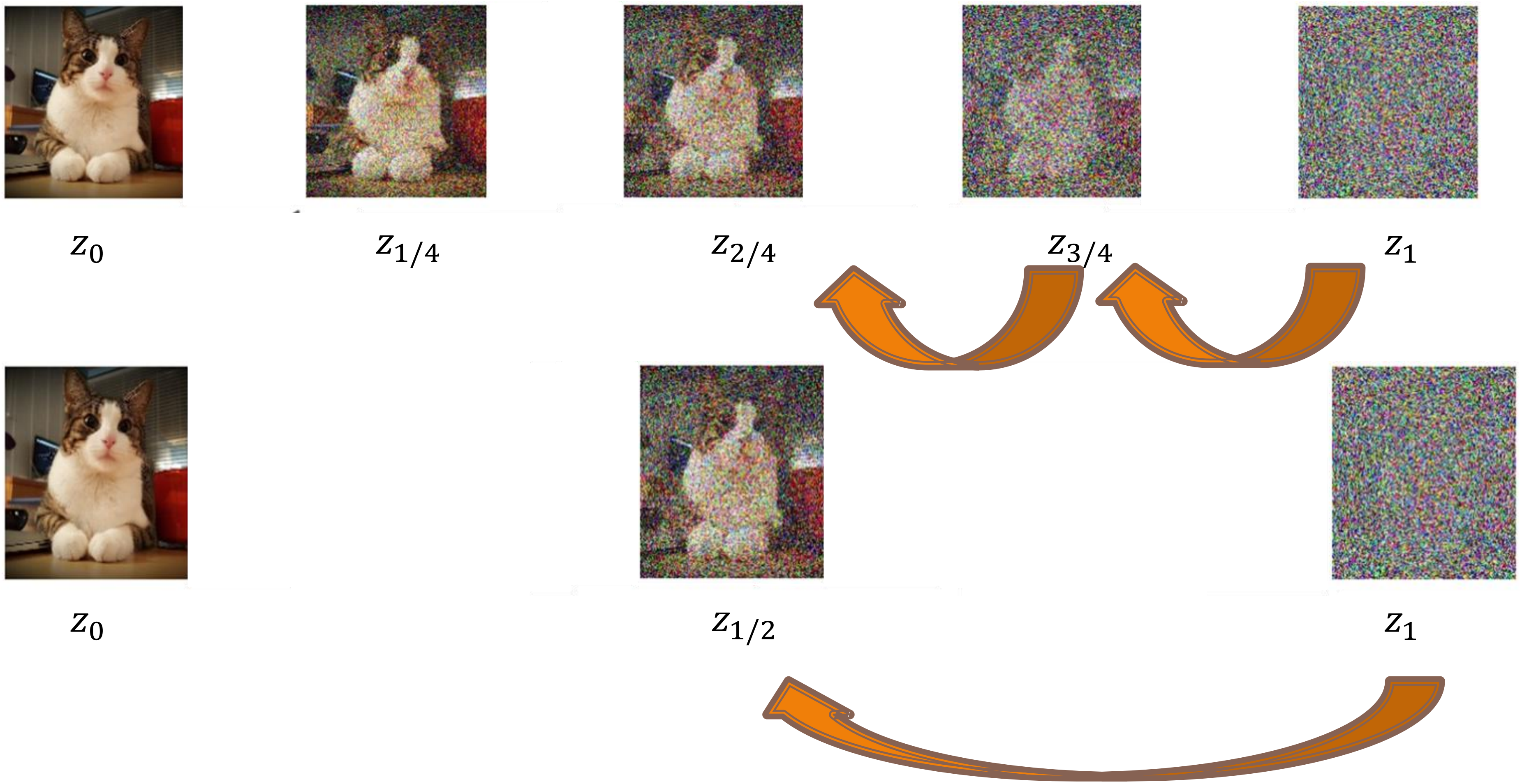


Progressive Distillation for Fast Sampling



Salimans T, Ho J. Progressive distillation for fast sampling of diffusion models[J]. arXiv preprint arXiv:2202.00512, 2022.

Progressive Distillation for Fast Sampling

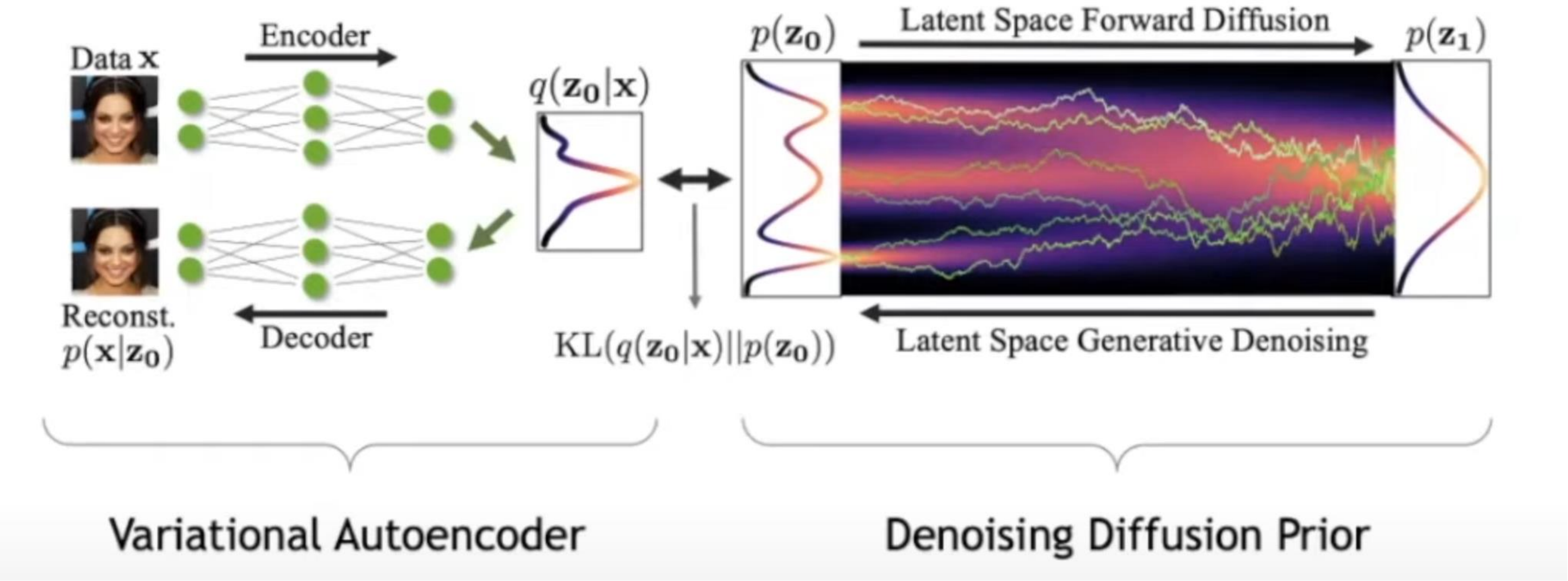


Student = New Teacher

New Student

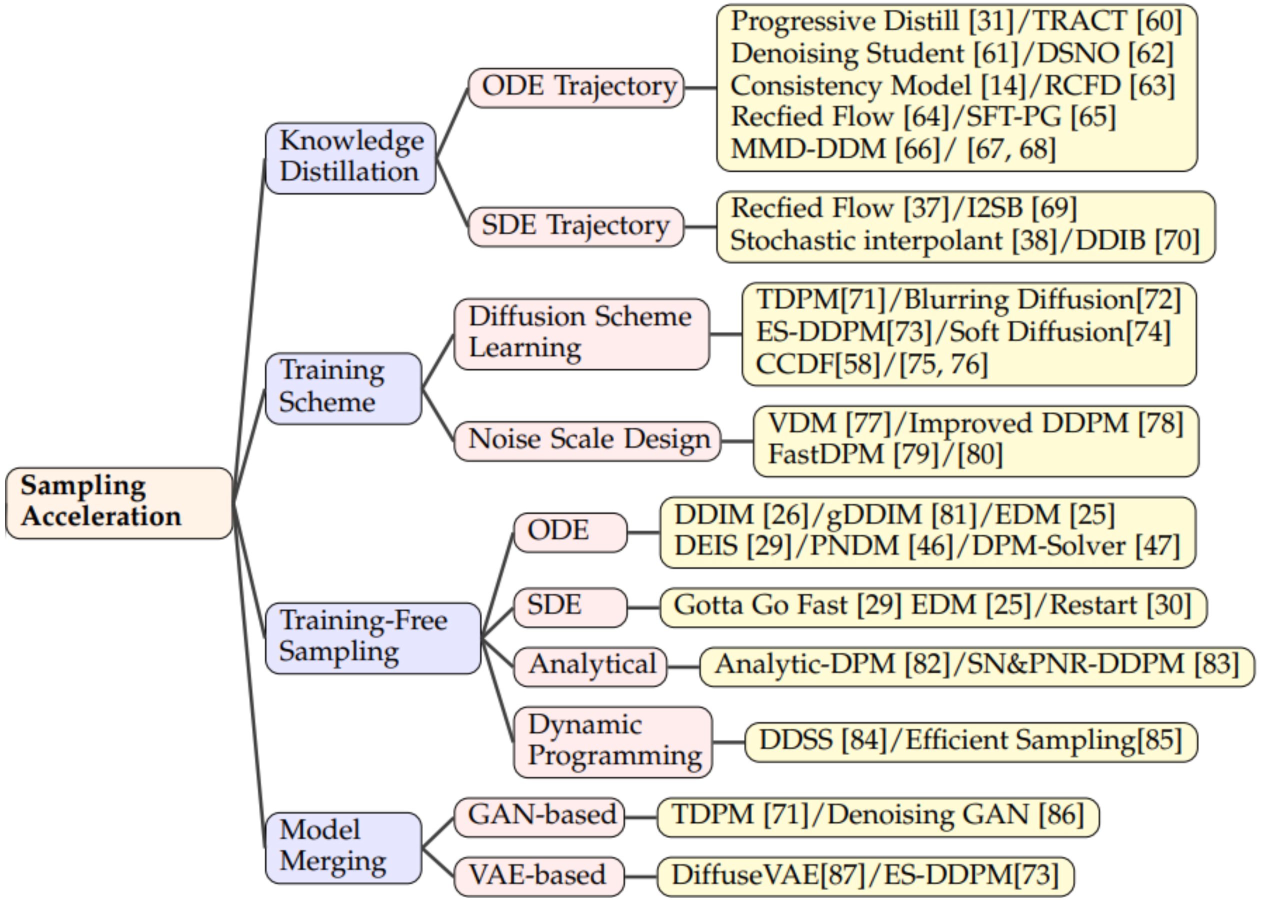
Salimans T, Ho J. Progressive distillation for fast sampling of diffusion models[J]. arXiv preprint arXiv:2202.00512, 2022.

Latent-space Diffusion Models



Rombach R, Blattmann A, Lorenz D, et al. High-resolution image synthesis with latent diffusion models[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 10684-10695.

Taxonomy of Sampling Acceleration



Cao, Hanqun, et al. "A survey on generative diffusion models." *IEEE Transactions on Knowledge and Data Engineering* (2024).

Tutorial Schedule

First Half

- ❑ Part 1: Introduction (30 mins)
  - ❖ What? Why? How?
  - ❖ Denoising Diffusion Models
  - ❖ Understanding and Intuition
- ❑ Part 2: Advanced Topics (30 mins)
  - ❖ Sampling Strategies
  - ❖ Conditioning

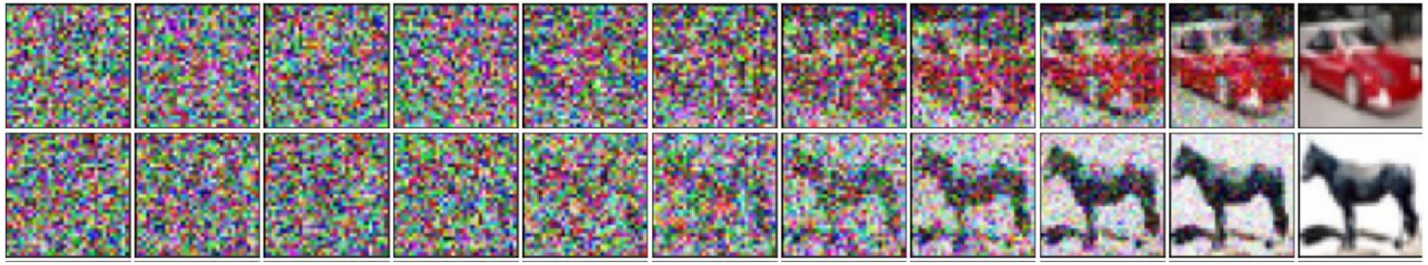
We are here!

Second Half

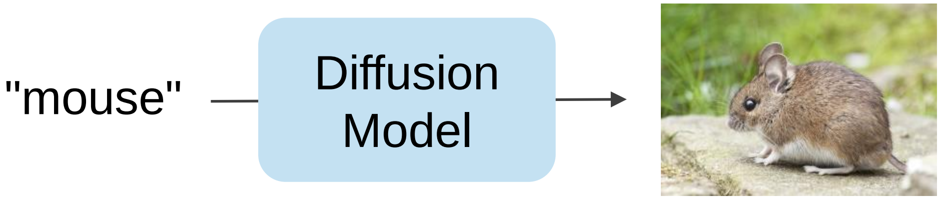
- ❑ Q&A
- ❑ Medical Imaging Applications (30 mins)
  - ❖ Synthesis
  - ❖ Segmentation
  - ❖ Anomaly Detection
  - ❖ Reconstruction
  - ❖ Registration

Part 2 – Advanced Topics

- Sampling Strategies



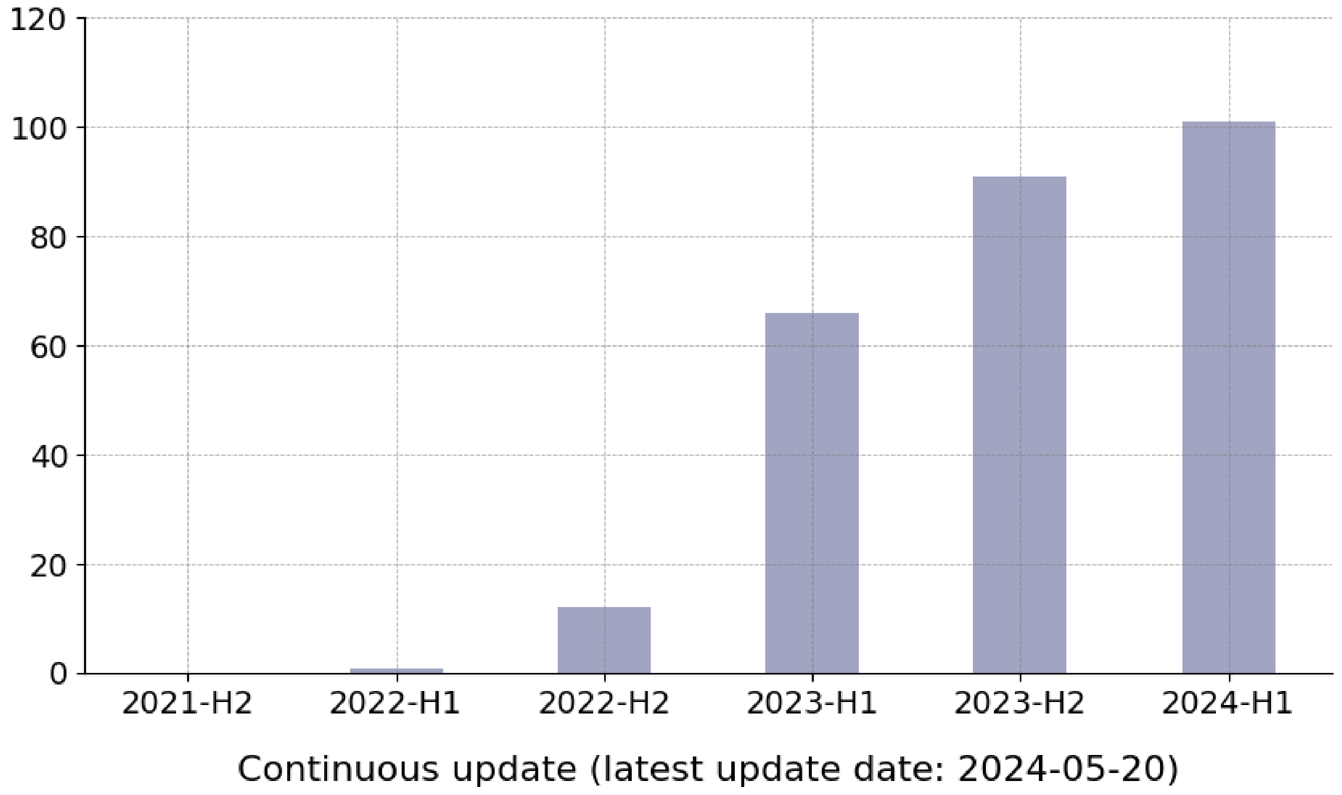
- Conditioning Mechanisms



Conditioning example: Image-to-image generation

Saharia, Chitwan, et al. "Photorealistic text-to-image diffusion models with deep language understanding." *Advances in neural information processing systems* 35 (2022): 36479-36494.

Trends in Conditioning



The number of papers on controllable generation based on T2I diffusion models.

Cao P, Zhou F, Song Q, et al. Controllable Generation with Text-to-Image Diffusion Models: A Survey[J]. arXiv preprint arXiv:2403.04279, 2024.

How to evaluate generated samples?

- How to evaluate a generative model?
  - Quality
  - Diversity

- Low FID
  - Not real enough
- High diversity
  - Means hard to control

- We want to introduce condition so that the generation can be controllable

Dhariwal, Prafulla, and Alexander Nichol. "Diffusion models beat gans on image synthesis." *Advances in neural information processing systems* 34 (2021): 8780-8794.

Generation Direction

- $\nabla_{x_t} \log p(\mathbf{x}_t)$: the direction of generation
  - SDE, generate with randomness
- What about a condition y?
  - $\nabla_{x_t} \log p(\mathbf{x}_t|y)$: the direction of generation, conditioned on y
- We have $p(x_t|y) = \frac{p(y|x_t)\,p(x_t)}{p(y)}$, then we have
  - $\nabla_{x_t} \log p(\mathbf{x}_t|y) = \boxed{\nabla_{x_t} \log p(y|\mathbf{x}_t)} + \boxed{\nabla_{x_t} \log p(\mathbf{x}_t)}$

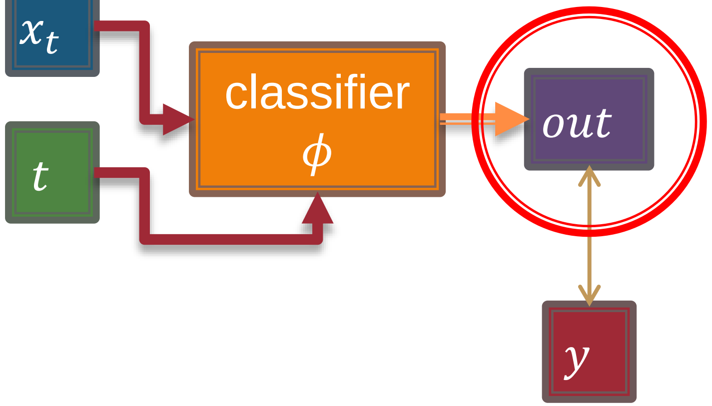
Direction of the condition

Direction of unconditional generation

Dhariwal, Prafulla, and Alexander Nichol. "Diffusion models beat gans on image synthesis." *Advances in neural information processing systems* 34 (2021): 8780-8794.

Classifier Guidance - Post Editing

- Let's create a classifier $p_\phi(y|x_t)$

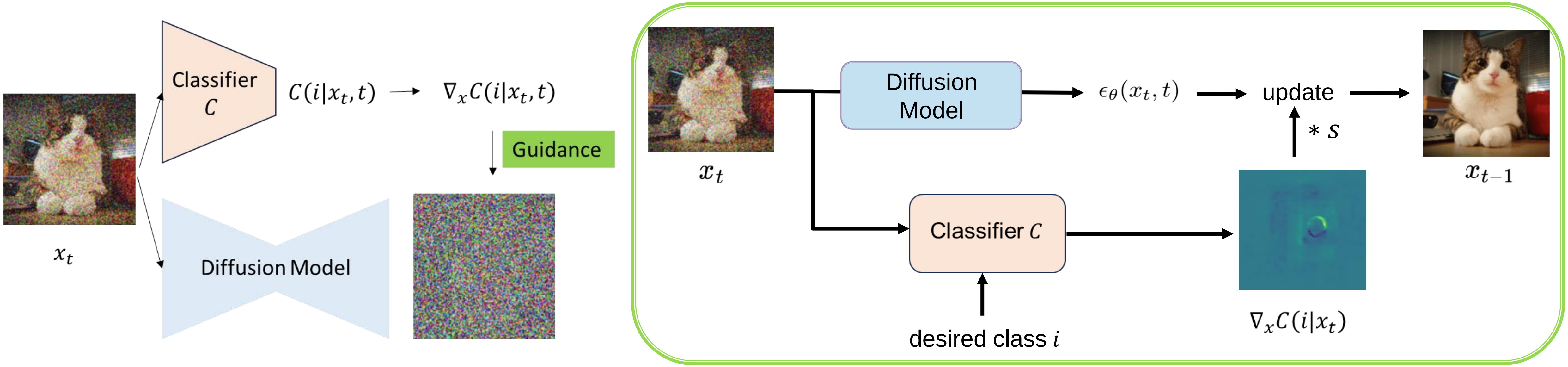


Take derivate to x_t and get
$\nabla_{x_t} \log p(\hat{y}|\mathbf{x}_t)$

Dhariwal, Prafulla, and Alexander Nichol. "Diffusion models beat gans on image synthesis." *Advances in neural information processing systems* 34 (2021): 8780-8794.

Classifier Guidance

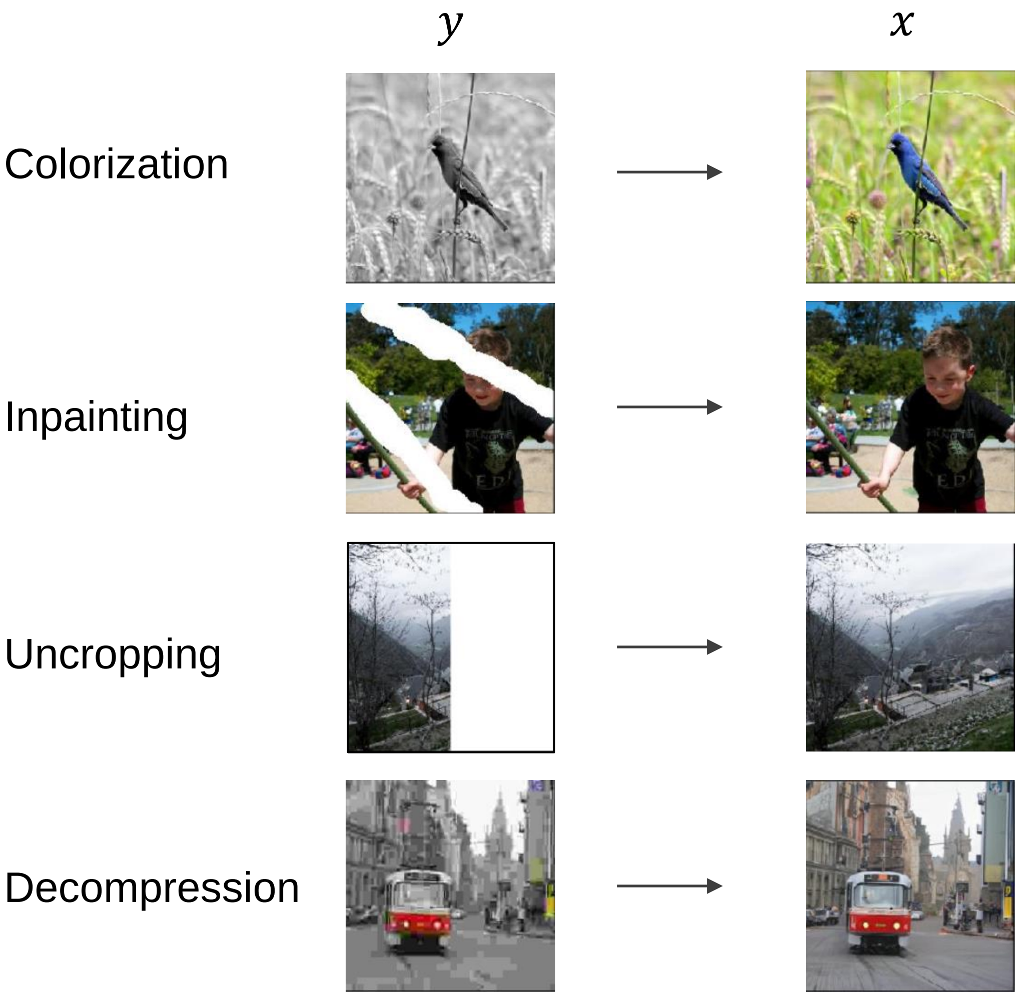
We use the gradient to guide the generation process towards a desired class.



Gradient guidance is not restricted to classification models. Other models (e.g., regression, segmentation, …) work just in the same way.

Dhariwal, Prafulla, and Alexander Nichol. "Diffusion models beat gans on image synthesis." *Advances in neural information processing systems* 34 (2021): 8780-8794.

Classifier Free Conditioning – Pre-training



- For image generation of a fake image x, we can use a conditioning image y.

- This requires **paired** training.

- During training and sampling, we add information of the conditioning image x through **channel-wise concatenation.**

Saharia, Chitwan, et al. "Palette: Image-to-image diffusion models." *ACM SIGGRAPH 2022 Conference Proceedings*. 2022.

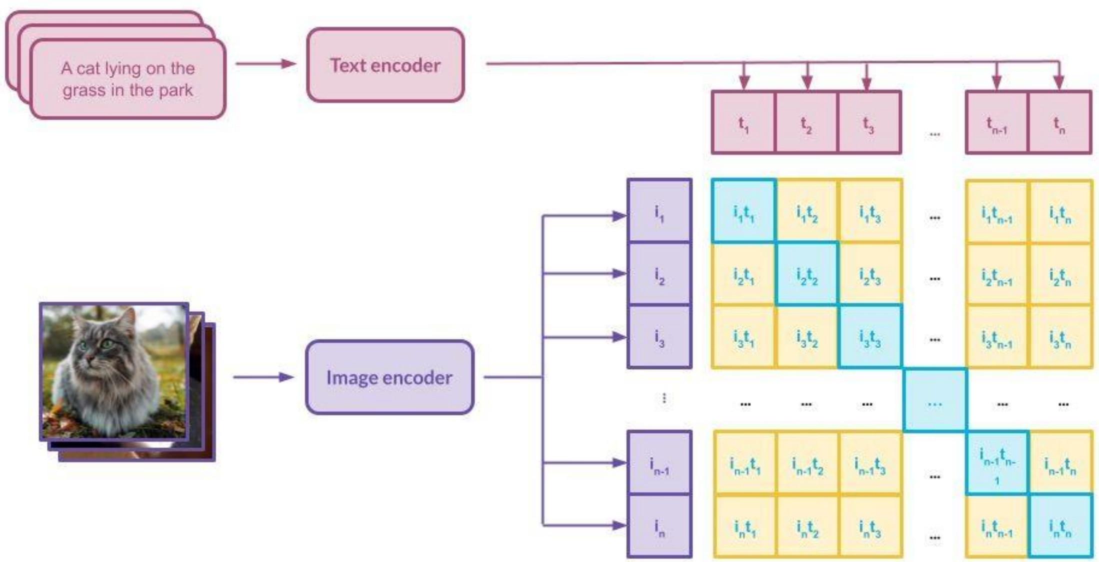
Comparison

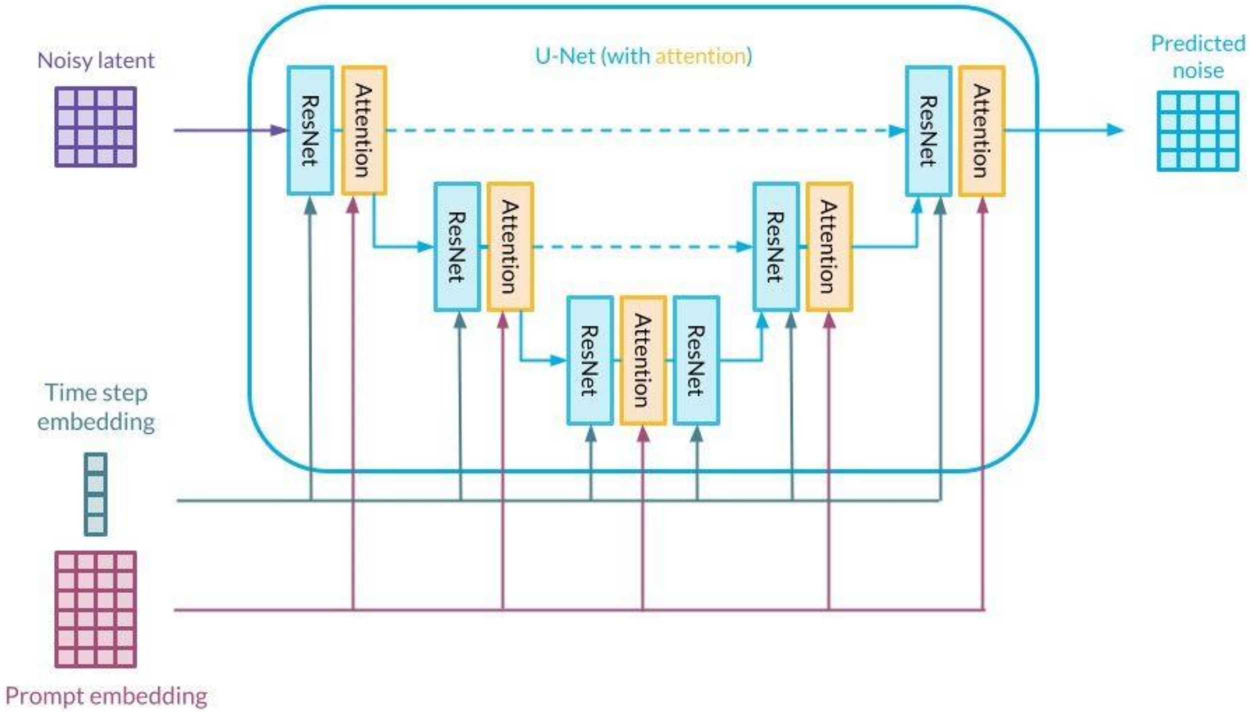
- Conditioning
  - Pre-training: classifier free
  - Post-editing: classifier guidance
- Pros and Cons
  - Classifier guidance
    - Low training costs
    - Less detailed generation
  - Classifier free
    - Great details; Training with y, the more input samples the easier to train
    - Every time you want to add another signal/condition, needs to retrain the model

Evolutions of Text-to-Image Diffusions

- Pixel Space Diffusion Models
  - GLIDE (ICML 2022)
    - Transforms the input text c into a token sequence via a transformer
    - Replace the *class-embedding* with the *pooled text feature* and **concatenate** to the attention layers
  - Imagen (NeurIPS 2022)
    - Use a pre-trained LLM as its text encoder
    - "Cross-attention is **the most effective technique**"
  - DALL.E 2, 3(arXiv 2022)
    - Bridges the gap between CLIP text and the image latent space $p(z_i|z_t)$
- Latent Space Diffusion Models
  - LDM (CVPR 2022)
    - Enhance the underlying Unet with cross-attention
  - Stable Diffusion (SD) v1, v2(CVPR 2022)
  - SD XL (ICLR 2024)

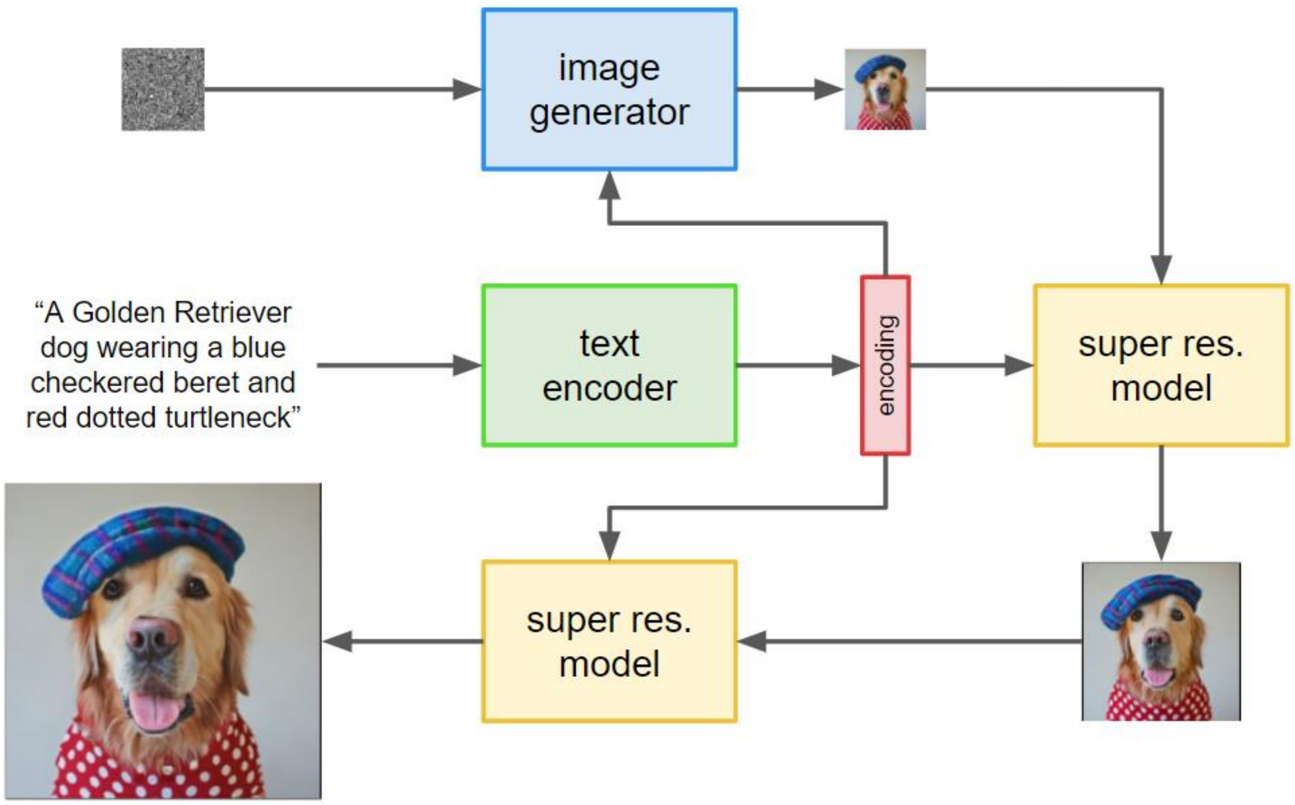
Glide: Text Conditioning

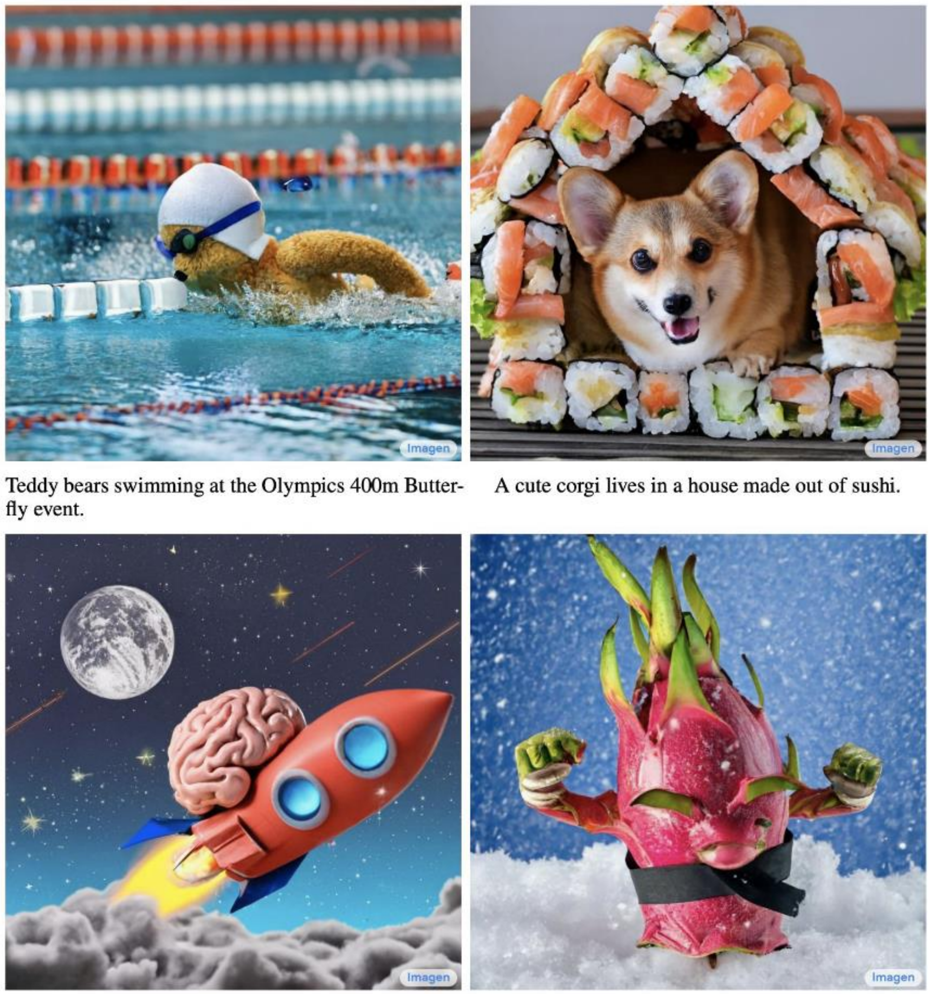




Nichol, Alex, et al. "Glide: Towards photorealistic image generation and editing with text-guided diffusion models." arXiv preprint arXiv:2112.10741 (2021).

Imagen: Text Conditioning





Teddy bears swimming at the Olympics 400m Butterfly event.

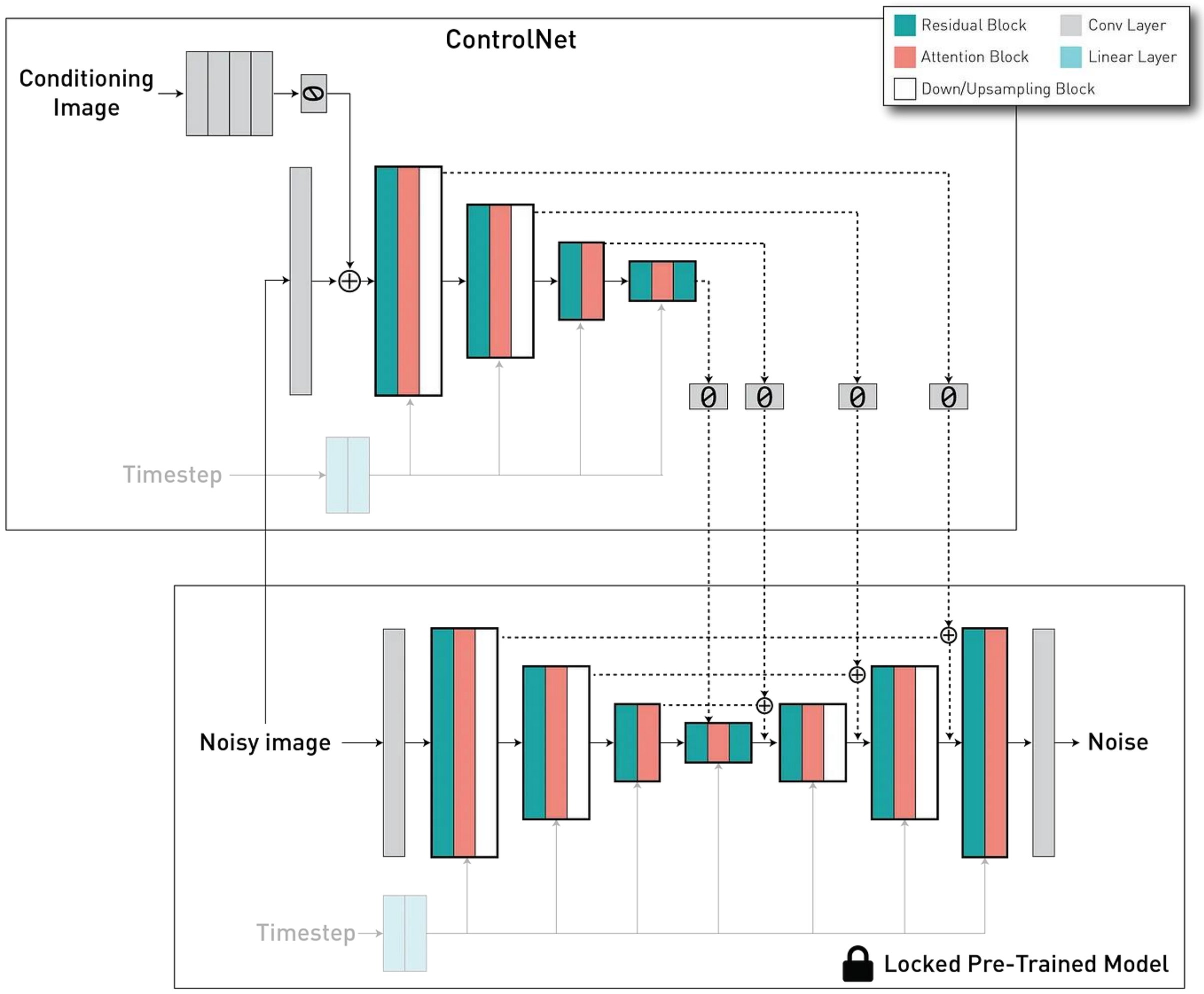
A cute corgi lives in a house made out of sushi.

A brain riding a rocketship heading towards the moon.

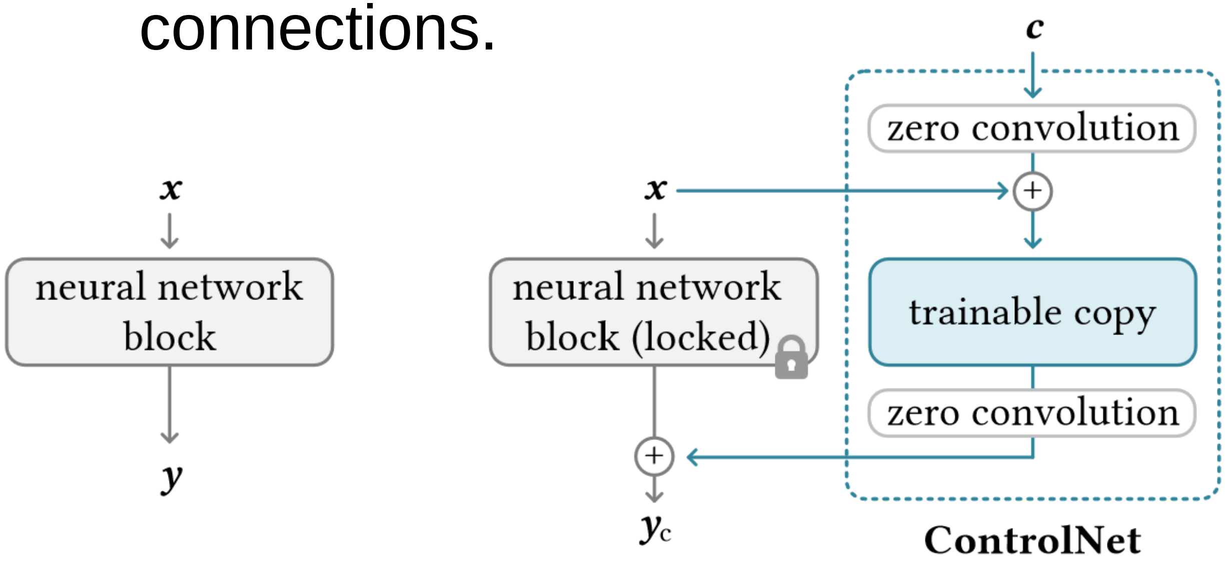
A dragon fruit wearing karate belt in the snow.

Saharia, Chitwan, et al. "Photorealistic text-to-image diffusion models with deep language understanding." *Advances in neural information processing systems* 35 (2022): 36479-36494.

ControlNet



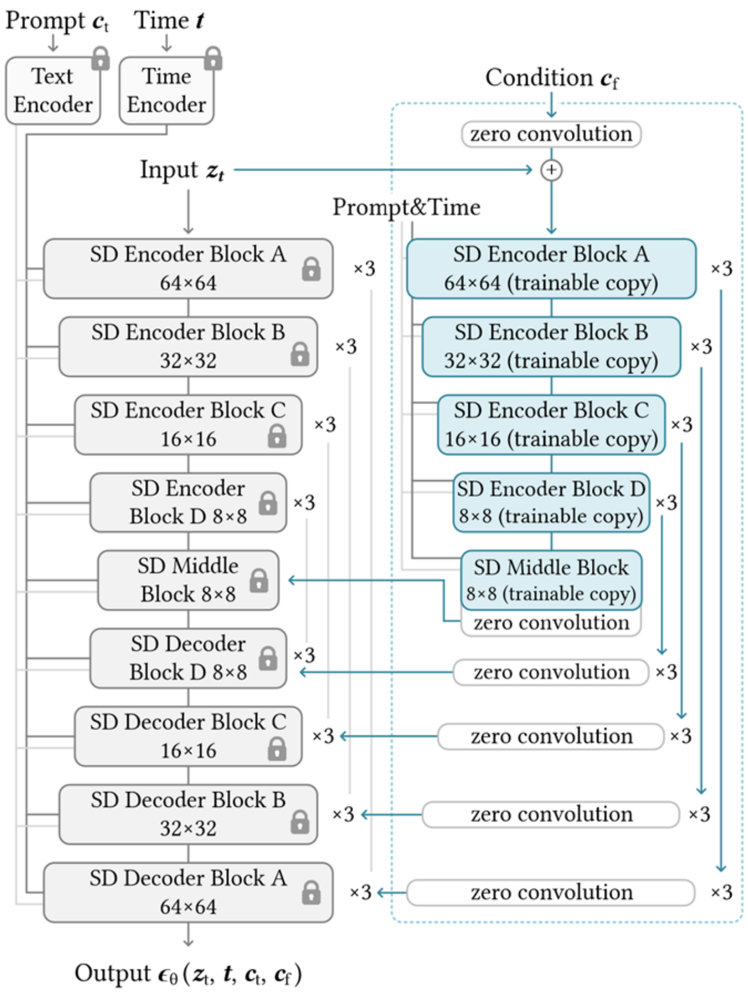
- We pretrain a diffusion model with text prompts.
- We freeze this model.
- We fine-tune a copy conditioned on $\boldsymbol{c}$.
- We pass information through skip connections.



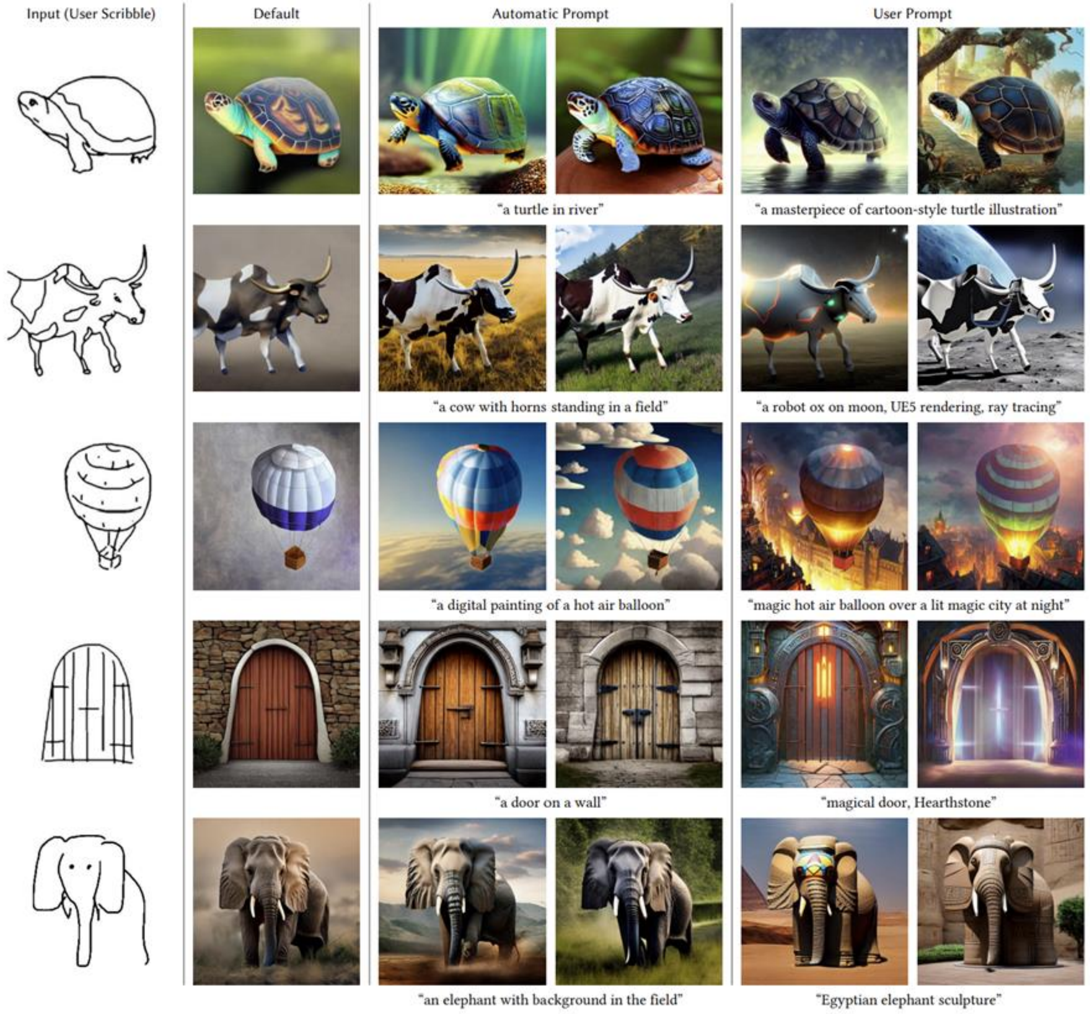
Zhang, Lvmin, Anyi Rao, and Maneesh Agrawala. "Adding conditional control to text-to-image diffusion models." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023.

ControlNet

conditioning image

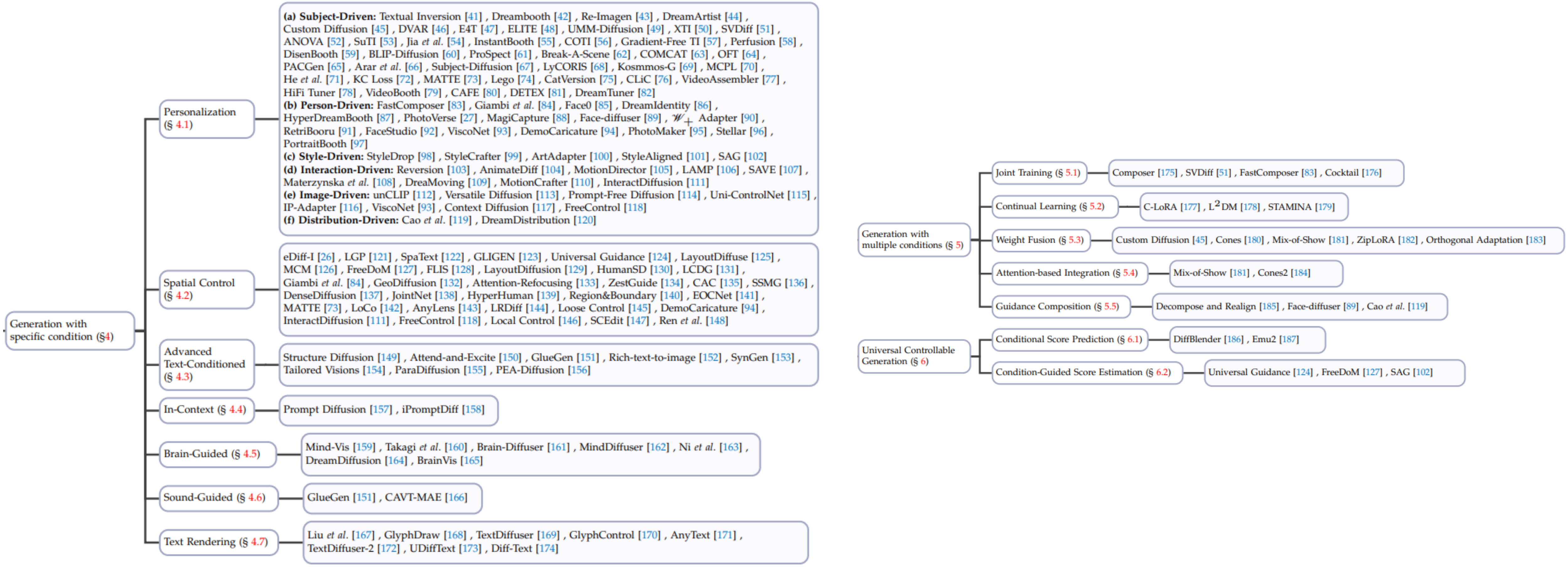


(a) Stable Diffusion (b) ControlNet



Zhang, Lvmin, Anyi Rao, and Maneesh Agrawala. "Adding conditional control to text-to-image diffusion models." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023.

Taxonomy of Controllable Generations



Cao P, Zhou F, Song Q, et al. Controllable Generation with Text-to-Image Diffusion Models: A Survey[J]. arXiv preprint arXiv:2403.04279, 2024.

Why do you think diffusion models exhibit such performance?

How to participate?

1 Go to wooclap.com

2 Enter the event code in the top banner

Event code
FROCKY

Enable answers by SMS

Copy participation link

What types of applications in Medical Imaging are you most interested in applying generative models to?

How to participate?

1. Go to wooclap.com
2. Enter the event code in the top banner

Event code
FROCKY

Enable answers by SMS

Copy participation link

Q&A

Generated by DALL.E 3

Tutorial Schedule

First Half

- ❑ Part 1: Introduction (30 mins)
  - ❖ What? Why? How?
  - ❖ Denoising Diffusion Models
  - ❖ Understanding and Intuition
- ❑ Part 2: Advanced Topics (30 mins)
  - ❖ Sampling Strategies
  - ❖ Conditioning

Second Half

- ❑ Q&A
- ❑ Medical Imaging Applications (30 mins)
  - ❖ Synthesis
  - ❖ Segmentation
  - ❖ Anomaly Detection
  - ❖ Reconstruction
  - ❖ Registration

We are here!

Part 2 – Medical Image Applications

Image Reconstruction

Image Registration

Image Segmentation

Anomaly Detection

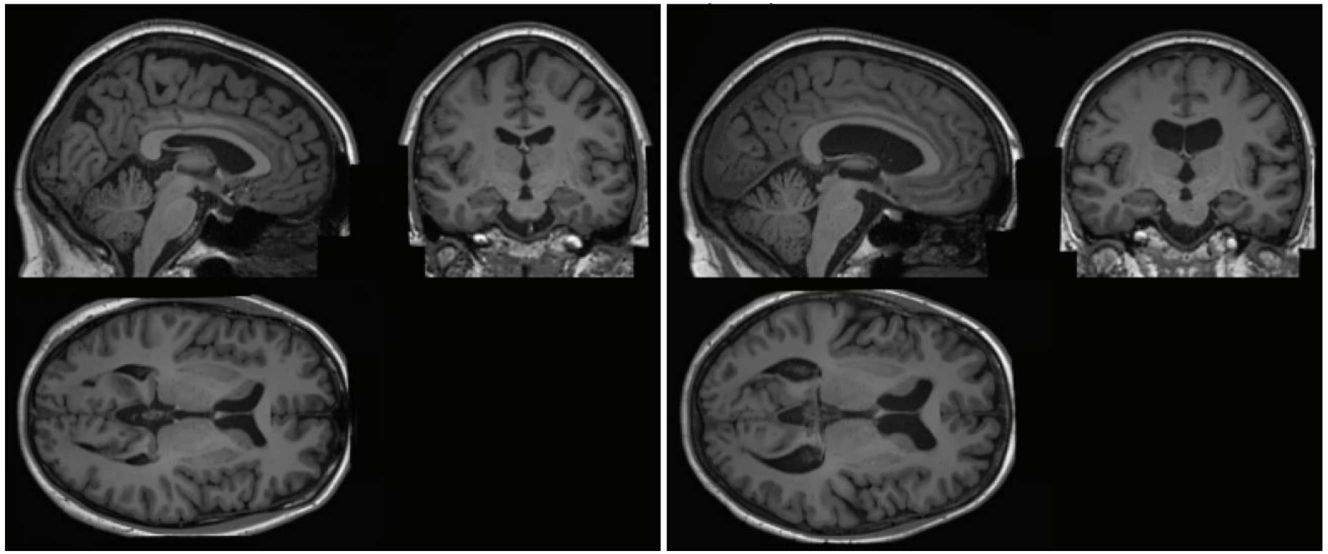
Image Synthesis

Image synthesis

Examples from the community

The simple setup of the problem

Real



Synthetic

Figure by Song et al ICLR 2022.

PAPERS
Pinaya et al (2022) Brain Imaging Generation with Latent Diffusion Models. MICCAI 2022 workshop
Khader et al. (2022) Medical Diffusion -- Denoising Diffusion Probabilistic Models for 3D Medical Image Generation. Scientific Reports, 2023
Chambon, Pierre, et al. (2022) RoentGen: vision-language foundation model for chest x-ray generation.
Ye, Jiarong, et al. (2023) Synthetic Augmentation with Large-scale Unconditional Pre-training. MICCAI 2023
Fernandez, V.et al. (2022). Can segmentation models be trained with fully synthetically generated data? MICCAI 2022 Workshop
Packhäuser et al. (2022) Generation of Anonymous Chest Radiographs Using Latent Diffusion Models for Training Thoracic Abnormality Classification Systems. ISBI 2023
Fernandez, V. et al (2023). Privacy Distillation: Reducing Re-identification Risk of Multimodal Diffusion Models. MICCAI 2023 Workshop
Sagers, Luke W., et al. (2023) Augmenting medical image classifiers with synthetic data from latent diffusion models.
Frisch, Yannik, et al. (2023) Synthesising Rare Cataract Surgery Samples with Guided Diffusion Models. MICCAI 2023.
Kim et al. (2022) Diffusion Deformable Model for 4D Temporal Medical Image Generation. MICCAI 2022
Ali et al. (2022) Spot the fake lungs: Generating Synthetic Medical Images using Neural Diffusion Models. AICS 2022
Rouzrokh et al. (2022) Multitask Brain Tumor Inpainting with Diffusion Models: A Methodological Report.
Chambon et al (2022) Adapting Pretrained Vision-Language Foundational Models to Medical Imaging Domains. NeurIPS2022
Lyu et al. (2022) Conversion Between CT and MRI Images Using Diffusion and Score-Matching Models.
Ozbey et al. (2023) Unsupervised Medical Image Translation with Adversarial Diffusion Models. IEEE Transactions on Medical Imaging 2023
Ktena, Ira, et al. (2024) Generative models improve fairness of medical classifiers under distribution shifts. Nature Medicine, 2024
Dominik J. E. Waibel, et al. (2023) Diffusion Model Predicts 3D Shapes from 2D Microscopy Images. ISBI 2023.
P.Huy, et al. (2023) Denoising Diffusion Medical Models. ISBI 2023.

Why? Medical Image Data is Scarce

Privacy concerns → Limited data

Data Provider

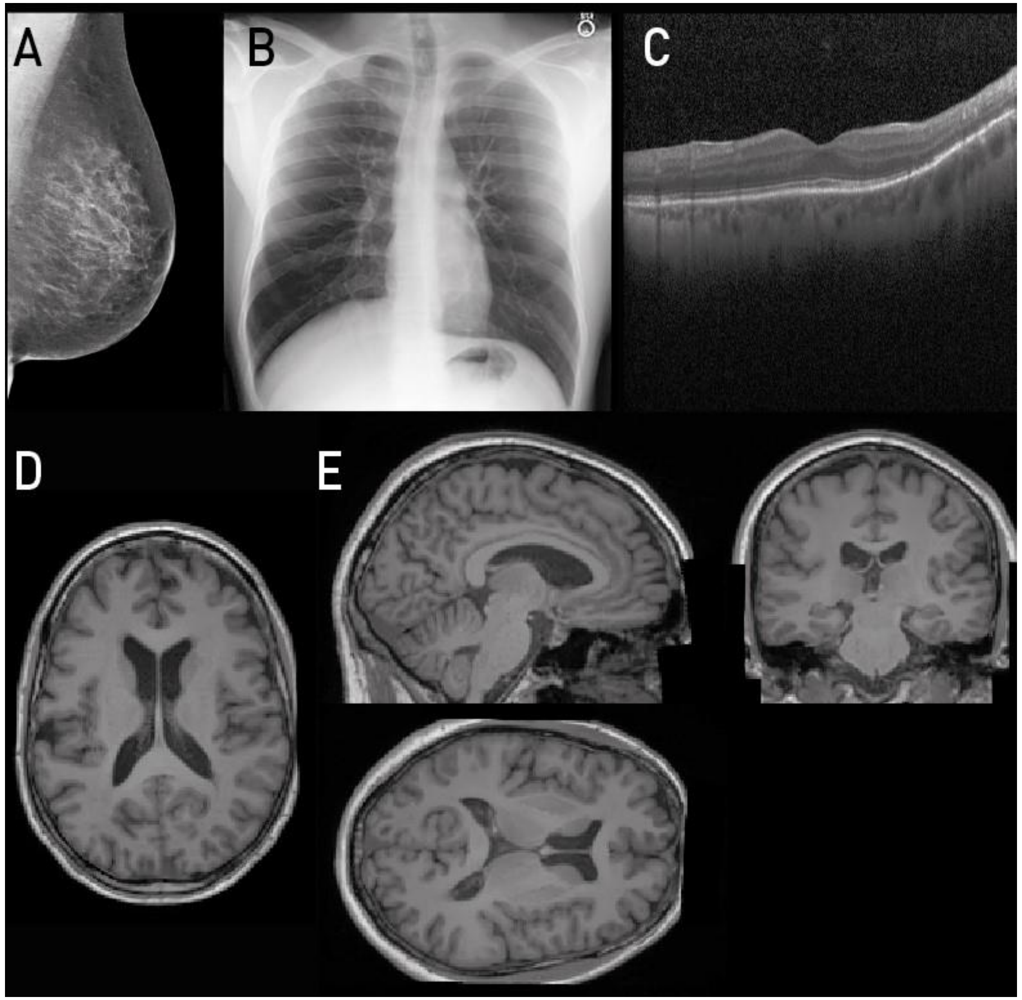
Data Users

Use **of Synthetic Data**

- ❑ Full “private” training
- ❑ Data augmentation
- ❑ Test-time augmentation
- ❑ Testing edge cases

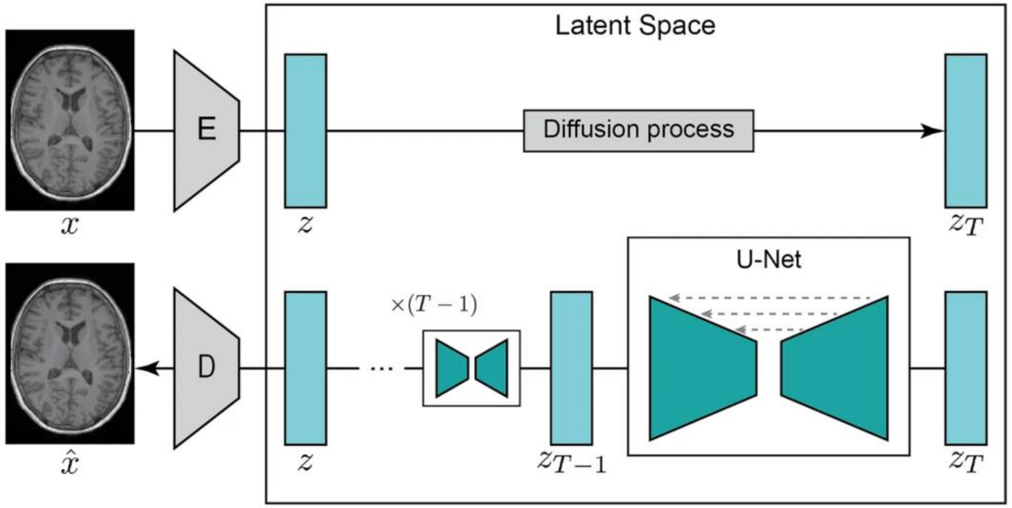
Evaluation Criteria

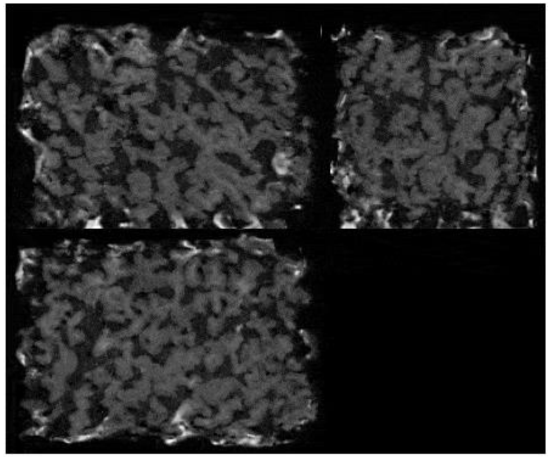
- ❑ Realism
- ❑ Diversity
- ❑ Privacy



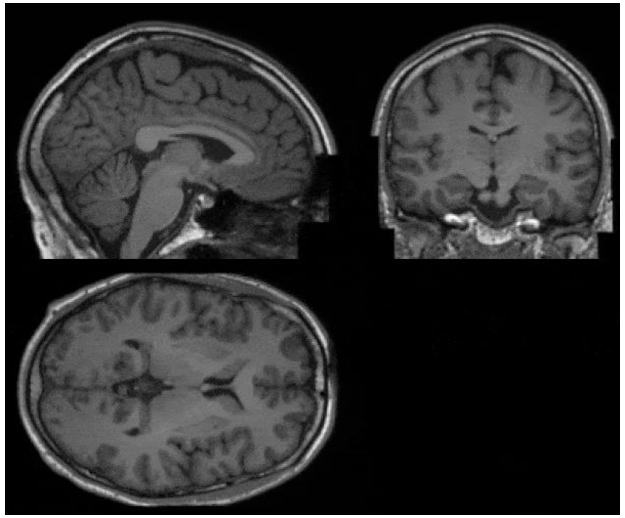
Pinaya, Walter HL, et al. "Generative AI for Medical Imaging: extending the MONAI Framework." arXiv preprint arXiv:2307.15208 (2023).

Generating high-resolution 3D brain data





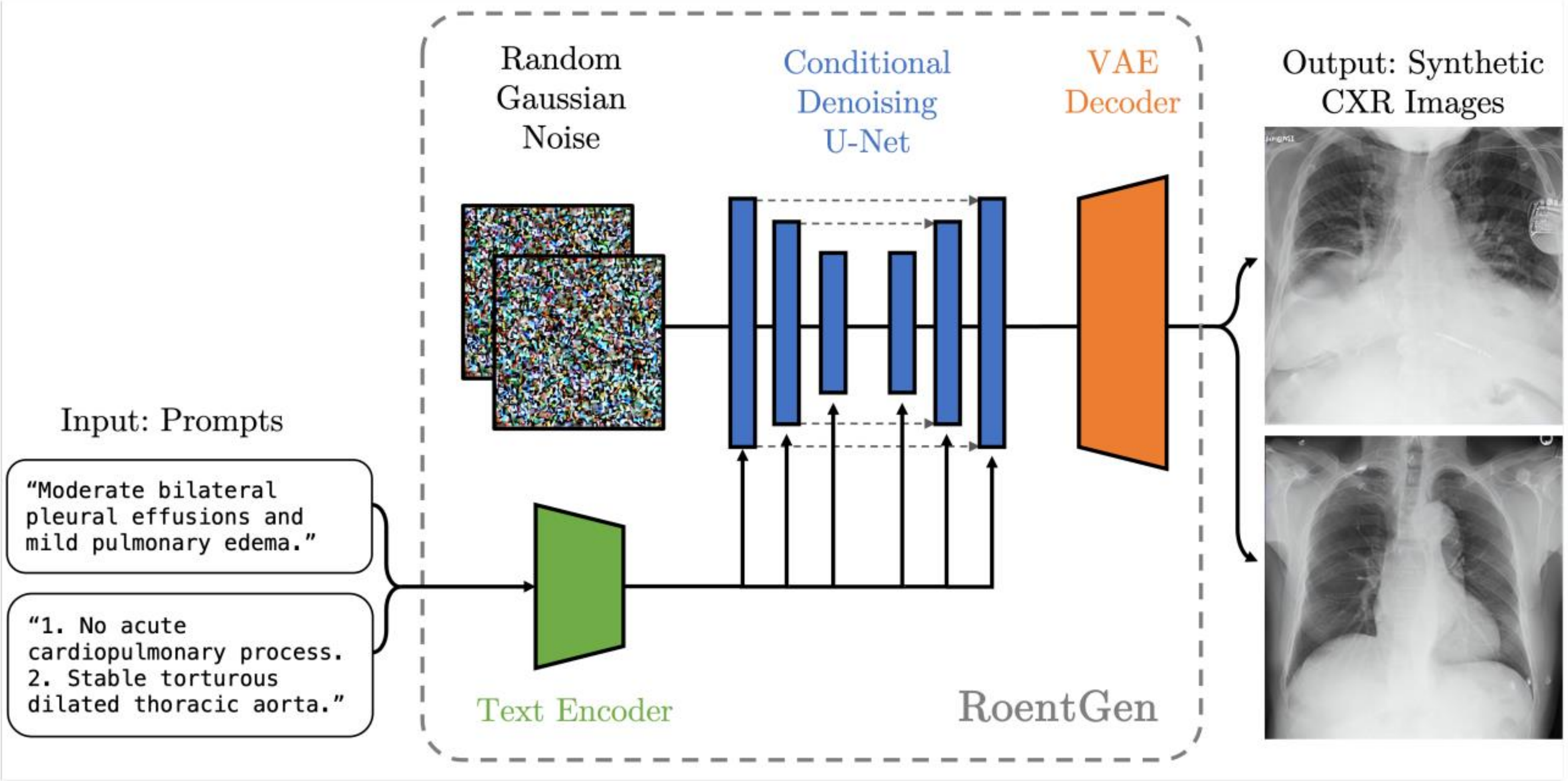
Synthetic



- Latent Diffusion Models trained on data from UK Biobank (N = 31,740)
  - T1 MRI brain images with 1 mm^3 voxel size (160 × 224 × 160 voxels)
- Conditioned on covariates, such as:
  - Age
  - Gender
  - Ventricular and Brain volumes

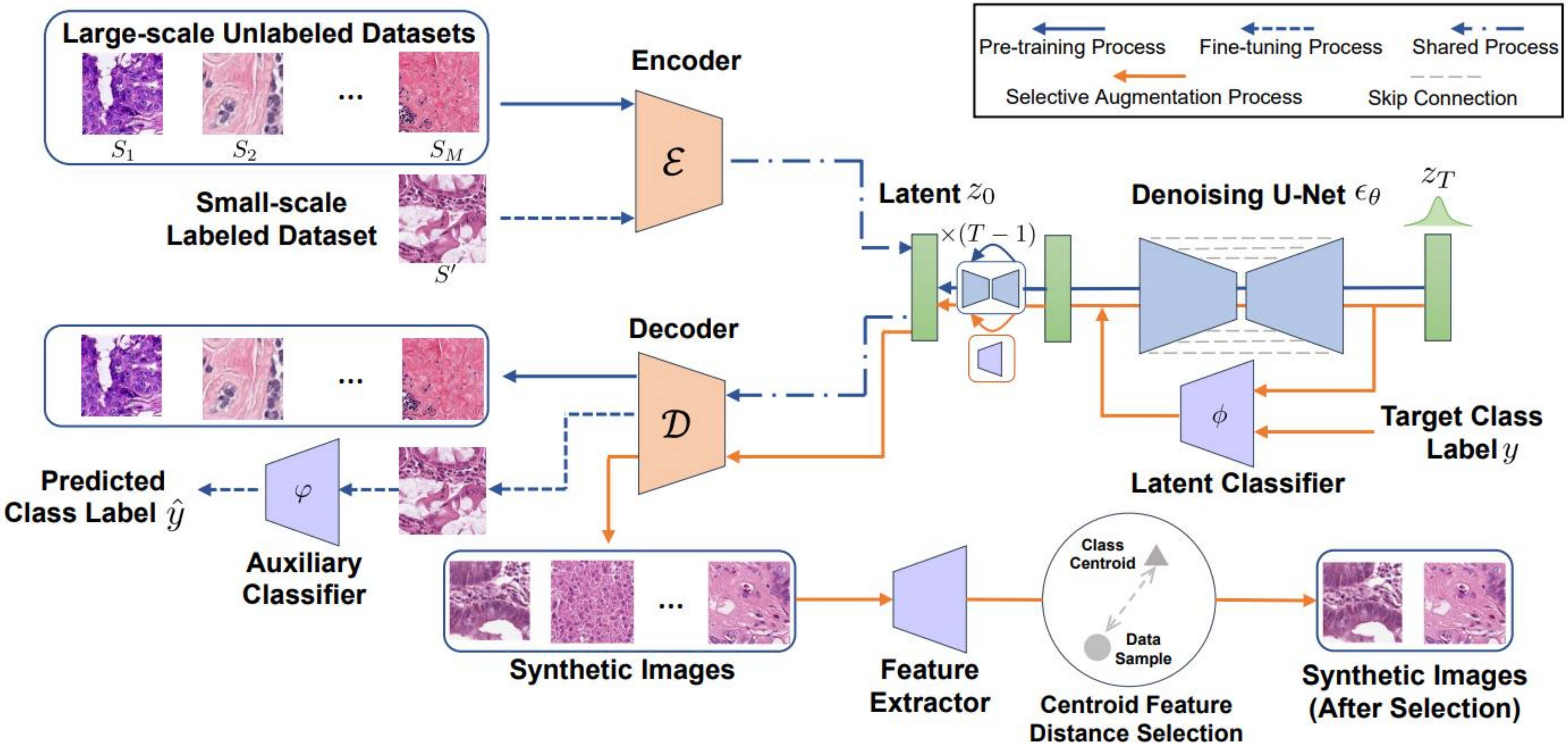
Pinaya et al (2022). Brain Imaging Generation with Latent Diffusion Models. MICCAI 2022 workshop
Khader et al. (2023) Medical Diffusion - Denoising Diffusion Probabilistic Models for 3D Medical Image Generation. Scientific Reports, 2023

Fine-tuning Stable Diffusion



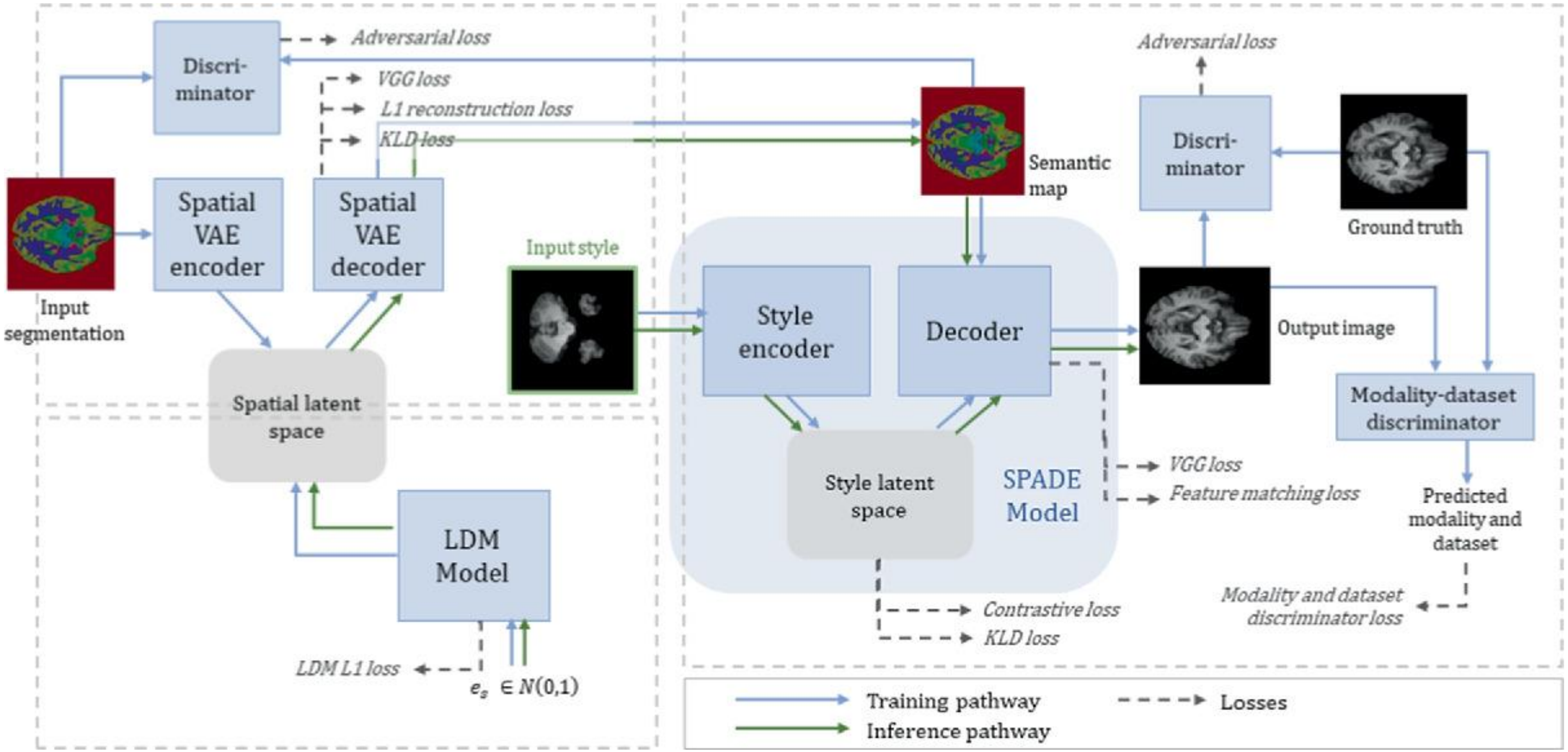
Chambon, Pierre, et al. (2022) RoentGen: vision-language foundation model for chest x-ray generation.

Unlabelled Pre-training



Ye, Jiarong, et al. (2023) Synthetic Augmentation with Large-scale Unconditional Pre-training. MICCAI 2023

Generating Segmentation Masks



Fernandez, V.et al. (2022). Can segmentation models be trained with fully synthetically generated data? MICCAI 2022 Workshop

Generation of Anonymous Chest Radiographs

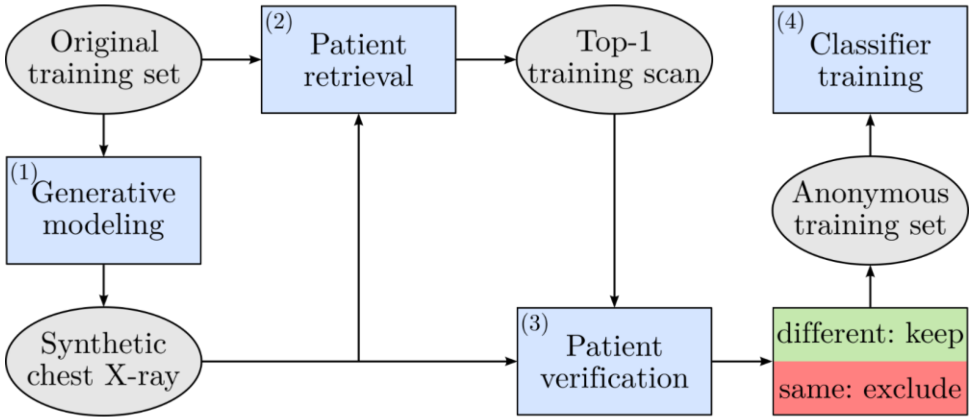


Fig. 1: Proposed privacy-enhancing image sampling strategy. Image taken from [1].

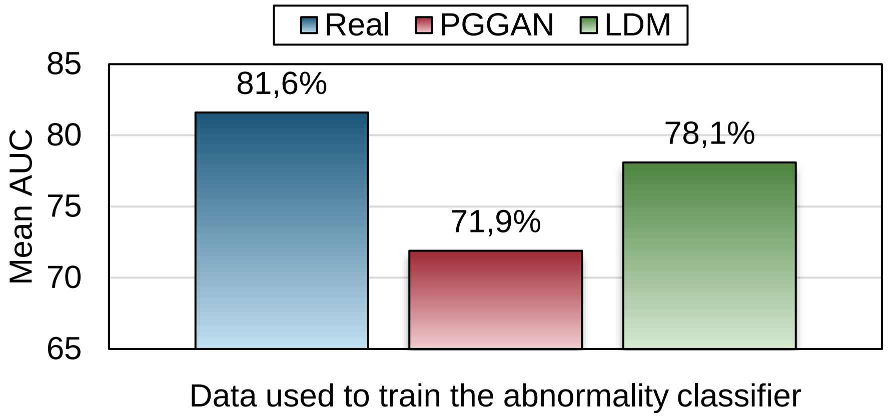


Fig. 2: Comparison of the classification performance of CheXNet.

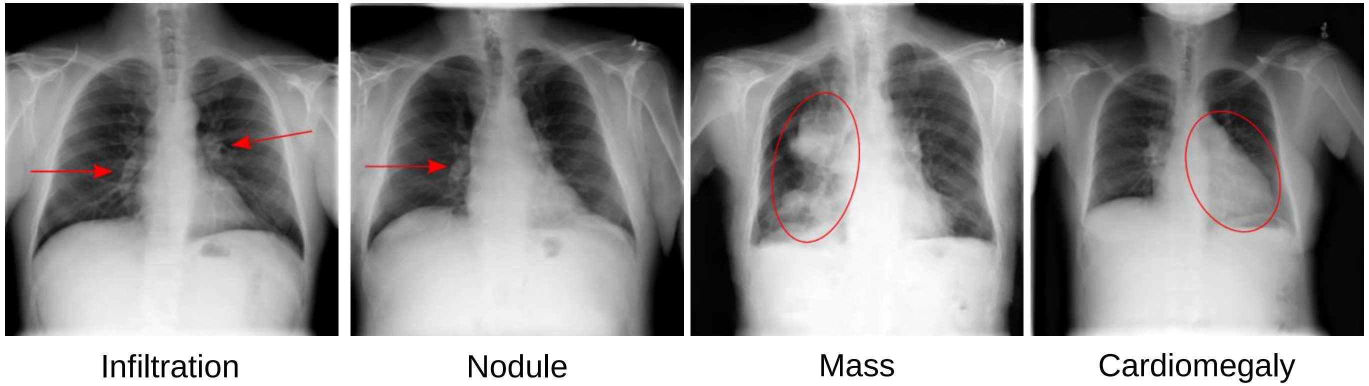
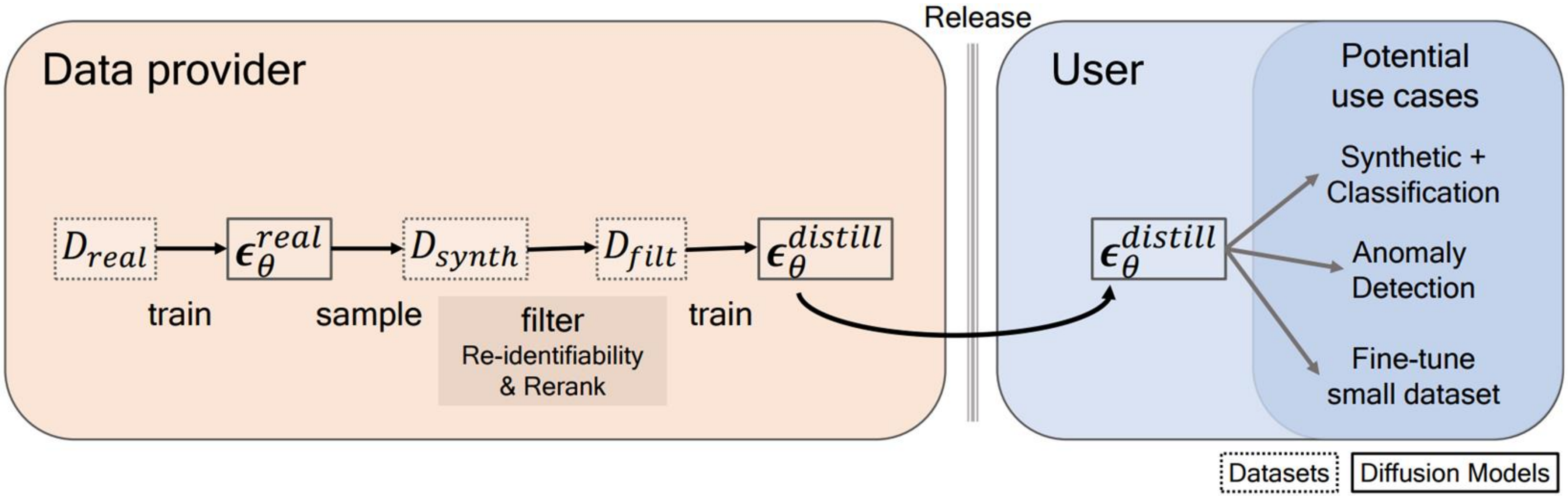


Fig. 3: Randomly selected images generated by the trained LDM. Images taken from [1].

Slides courtesy of Kai Packhäuser

Packhäuser et al. (2022) Generation of Anonymous Chest Radiographs Using Latent Diffusion Models for Training Thoracic Abnormality Classification Systems. ISBI 2023

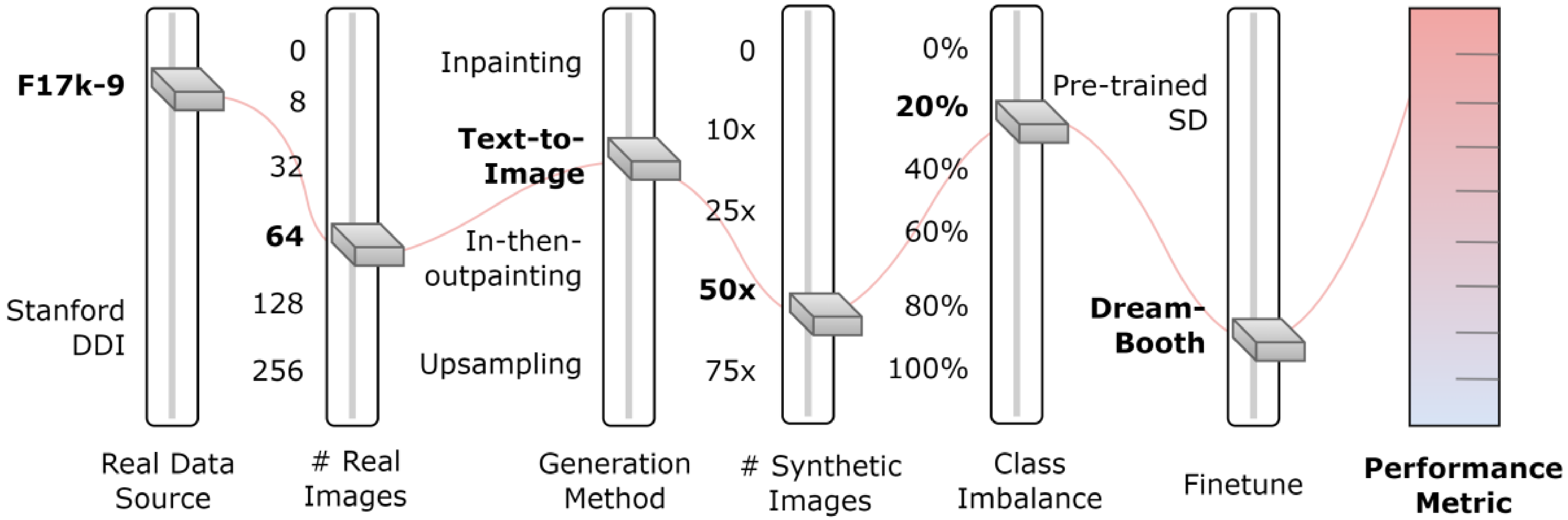
Privacy Distillation



Fernandez, V. et al (2023). Privacy Distillation: Reducing Re-identification Risk of Multimodal Diffusion Models. MICCAI 2023 Workshop

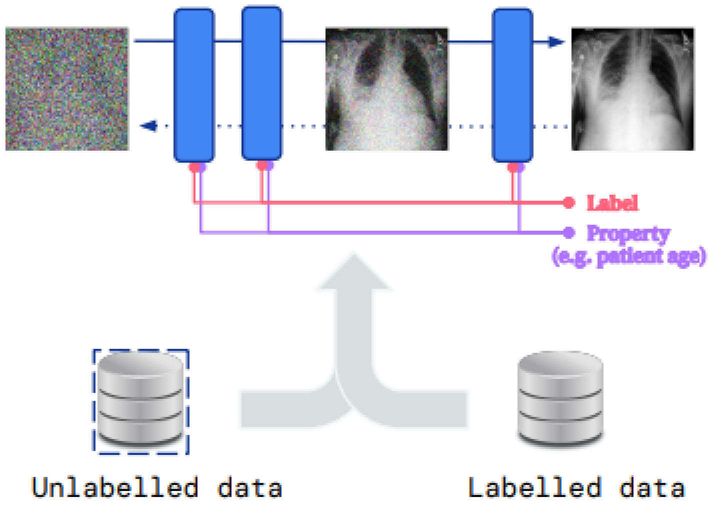
Synthetic Image Augmentation

Synthetic-to-real ratio of 10:1

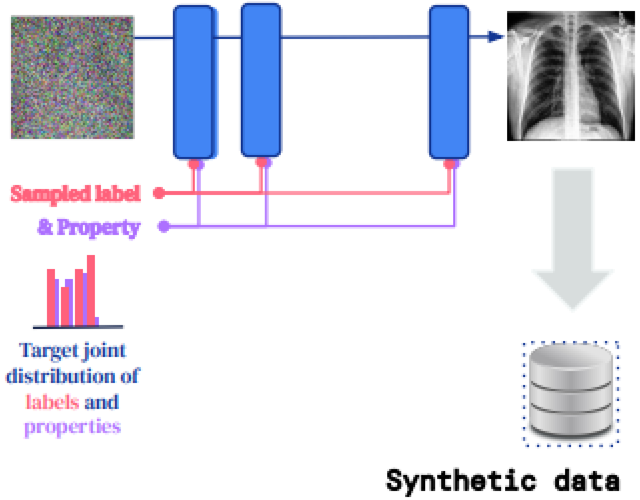


Sagers, Luke W., et al. (2023) Augmenting medical image classifiers with synthetic data from latent diffusion models.

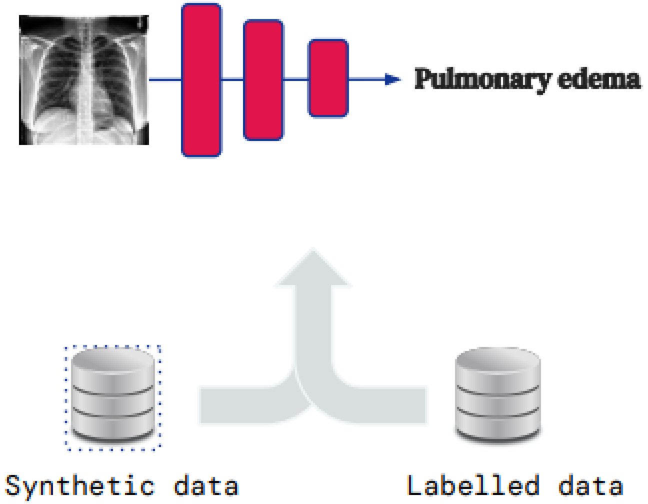
Synthetic Data for Distribution Shifts



1. Train **Diffusion** Model using labelled, weakly labelled and/or unlabelled data



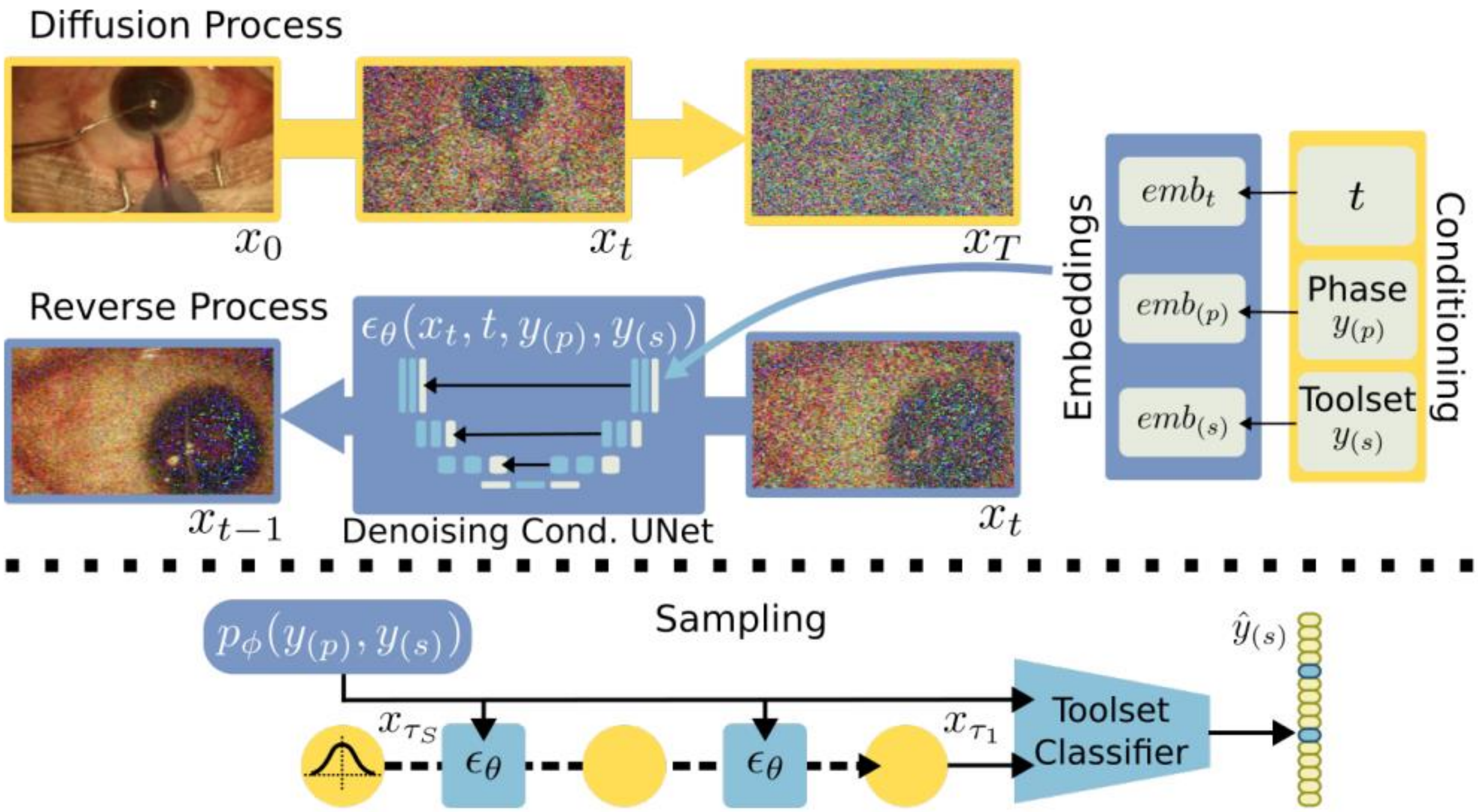
2. Sample synthetic images to create a **synthetic dataset**



3. Train downstream **classifier** from labeled and synthetic data

Ktena, Ira, et al. (2024) Generative models improve fairness of medical classifiers under distribution shifts. Nature Medicine, 2024

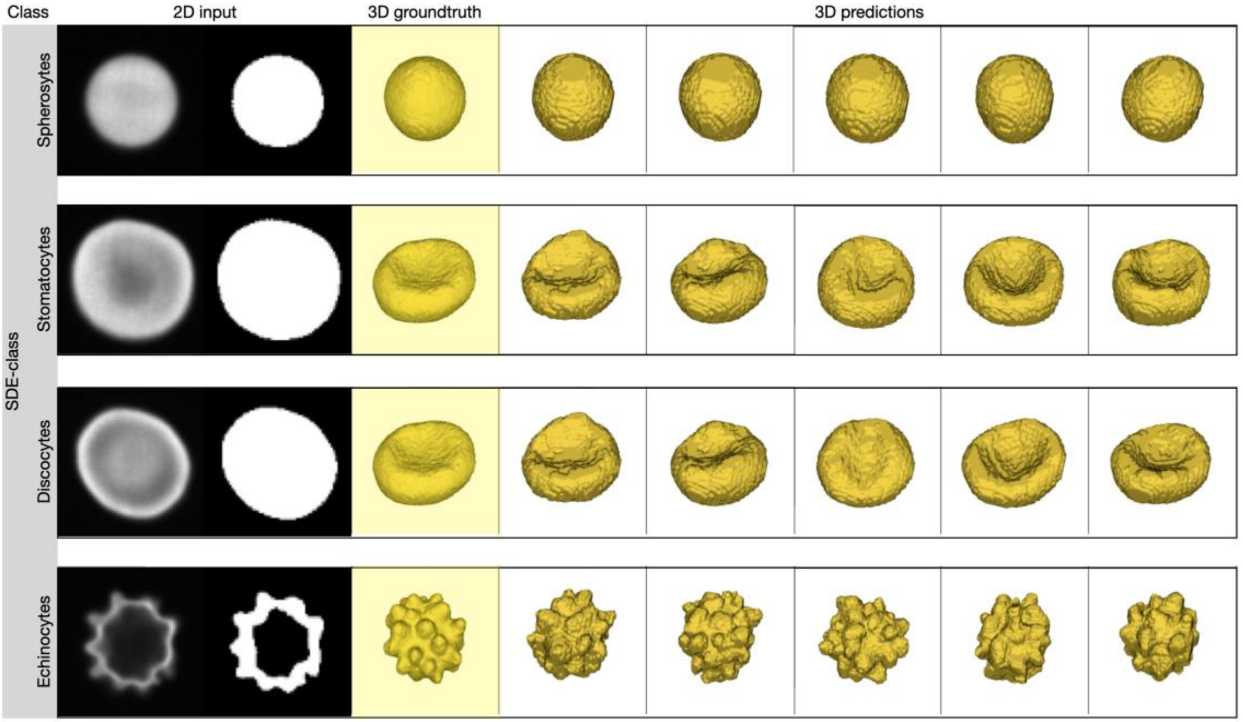
Synthesising Rare Samples

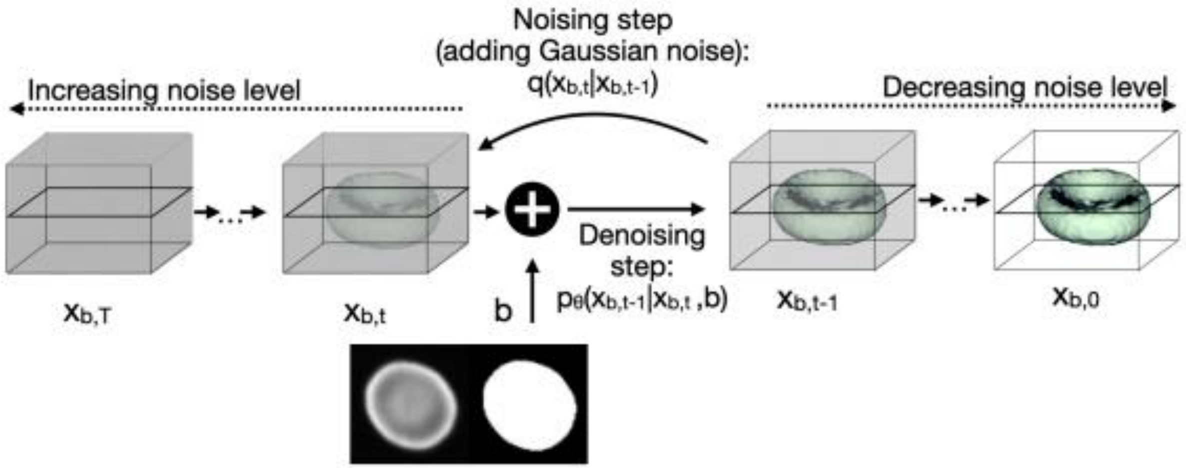


Frisch, Yannik, et al. (2023) Synthesising Rare Cataract Surgery Samples with Guided Diffusion Models. MICCAI 2023.

DIffusion based Shape PRediction, **DISPR**

3D shell shapes generation conditioned on 2D microscopy images

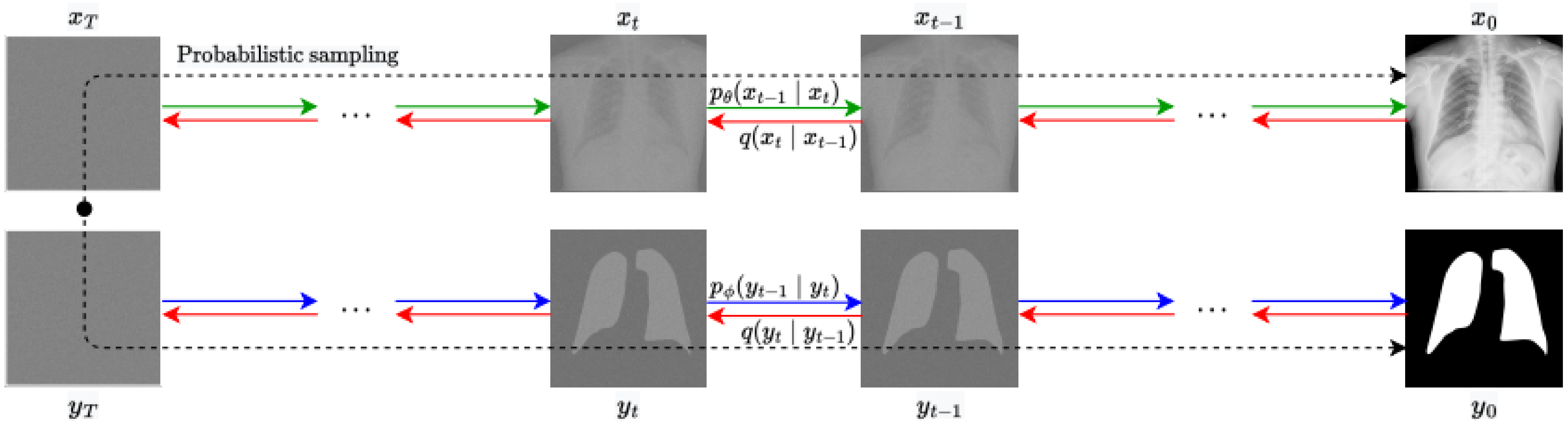




the 2D image is concatenated in each timestep of diffusion process

Dominik J. E. Waibel, et al. (2023) Diffusion Model Predicts 3D Shapes from 2D Microscopy Images. ISBI 2023.

Denoising Diffusion Medical Models DDMM



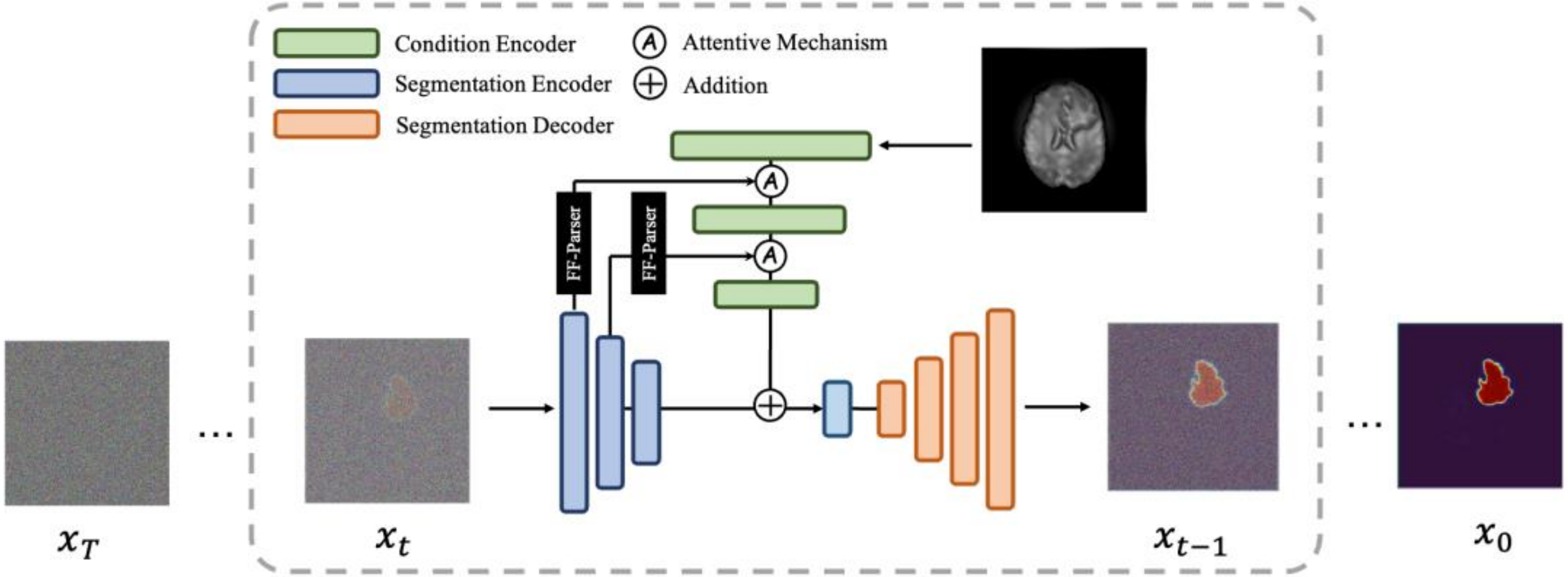
The DDMM model is built from one or more branches of DDPM (**radiographs** and **segmentation branches**) that share the same noise scheduler and latent code, which enforce **semantic consistency**

P.Huy, et al. (2023) Denoising Diffusion Medical Models. ISBI 2023.

Image segmentation

Examples from the community

Setup

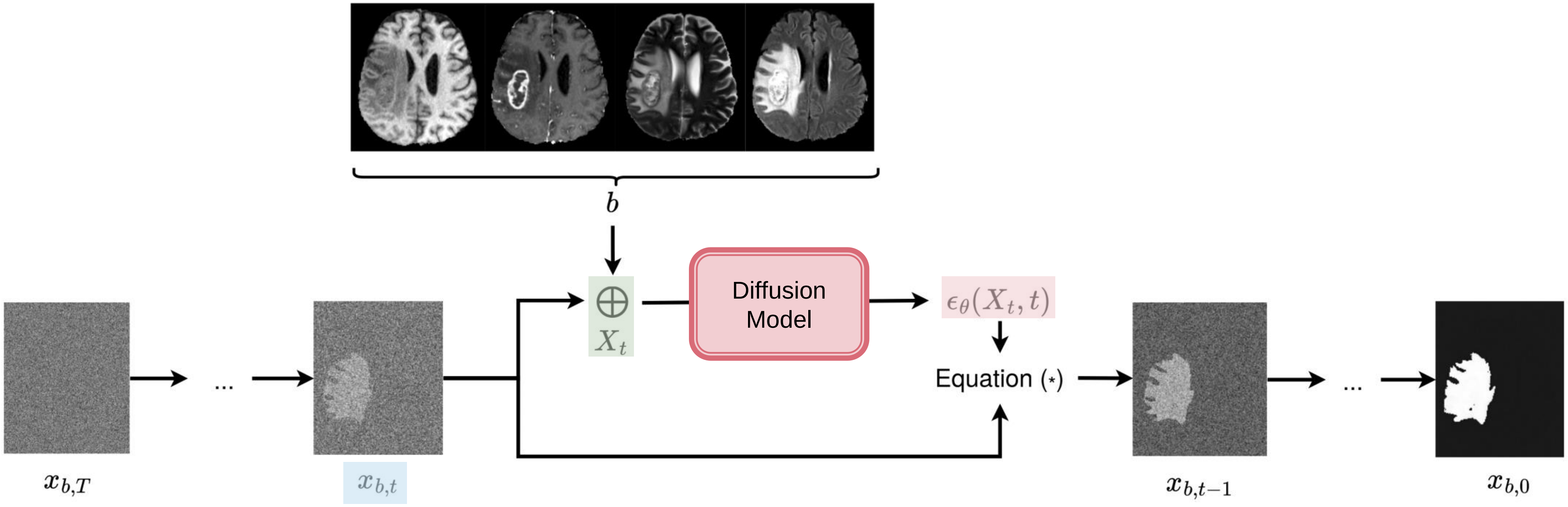


PAPERS

Wolleb et al (2022). Diffusion Models for Implicit Image Segmentation Ensembles, MIDL 2022
Bieder et al. (2023) Memory-Efficient 3D Denoising Diffusion Models for Medical Image Processing. MIDL 2023
Rahman, Aimon, et al. (2023) Ambiguous medical image segmentation using diffusion models. CVPR 2023
Rousseau et al. (2023) Pre-Training with Diffusion models for Dental Radiography segmentation. MICCAI 2023
Guo et al. (2023) Accelerating Diffusion Models Via Pre-Segmentation Diffusion Sampling for Medical Image Segmentation. ISBI 2023
La Barbera et al. (2022) Anatomically constrained CT image translation for heterogeneous blood vessel segmentation. BMVC 2022
Kim et al. (2022) Diffusion Adversarial Representation Learning for Self-supervised Vessel Segmentation. ICLR 2023
Wu et al (2022) MedSegDiff: Medical Image Segmentation with Diffusion Probabilistic Model. MIDL 2023

Figure by Song et al ICLR 2022.

Diffusion Models for Segmentation Mask Generation

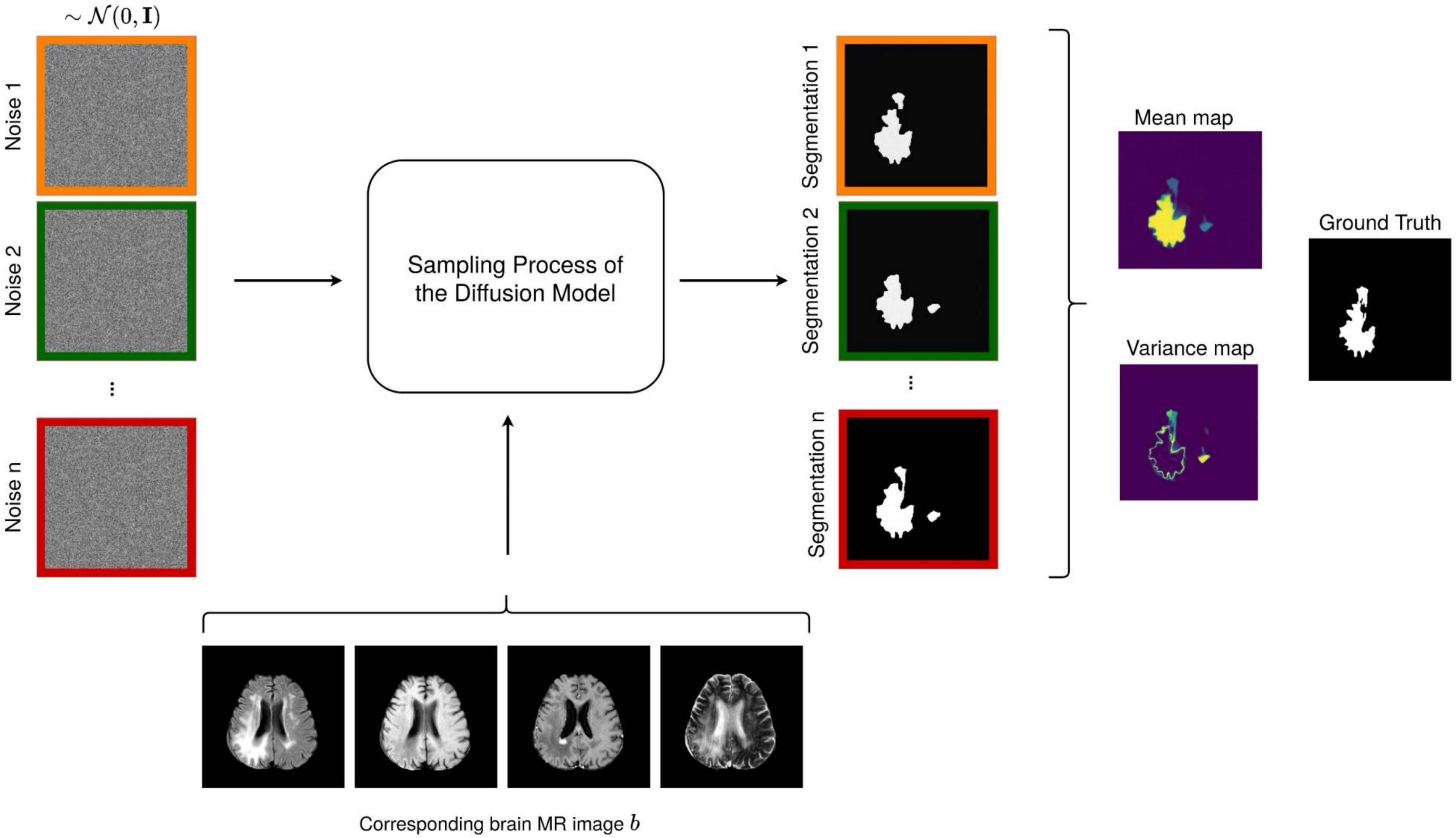


$$(*) \quad x_{b,t-1} = \frac{1}{\sqrt{\alpha_t}}\left(x_{b,t} - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(X_t, t)\right) + \sigma_t \mathbf{z}, \quad \text{with } \mathbf{z} \sim \mathcal{N}(0, \mathbf{I})$$

The anatomical information is added by concatenating the input images b to the noisy segmentation mask $x_{b,t}$ *in* every step *t.*

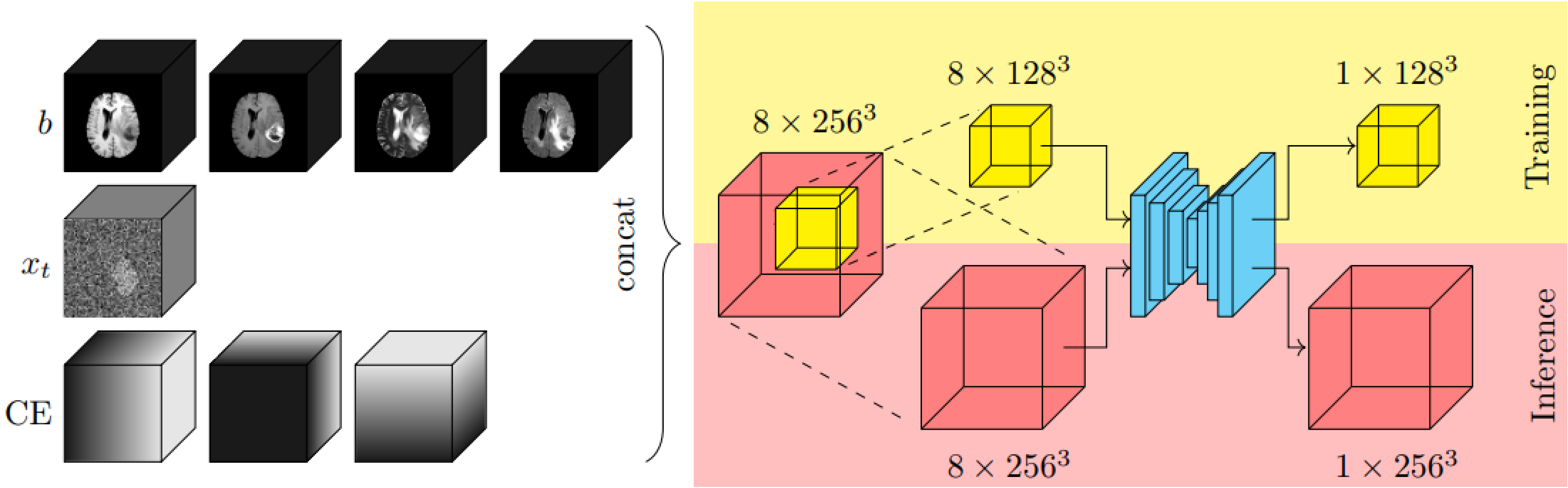
Wolleb et al (2022). Diffusion Models for Implicit Image Segmentation Ensembles, MIDL 2022

Generation of Segmentation Ensembles



Wolleb et al (2022). Diffusion Models for Implicit Image Segmentation Ensembles, MIDL 2022

3D Segmentation with PatchDDM



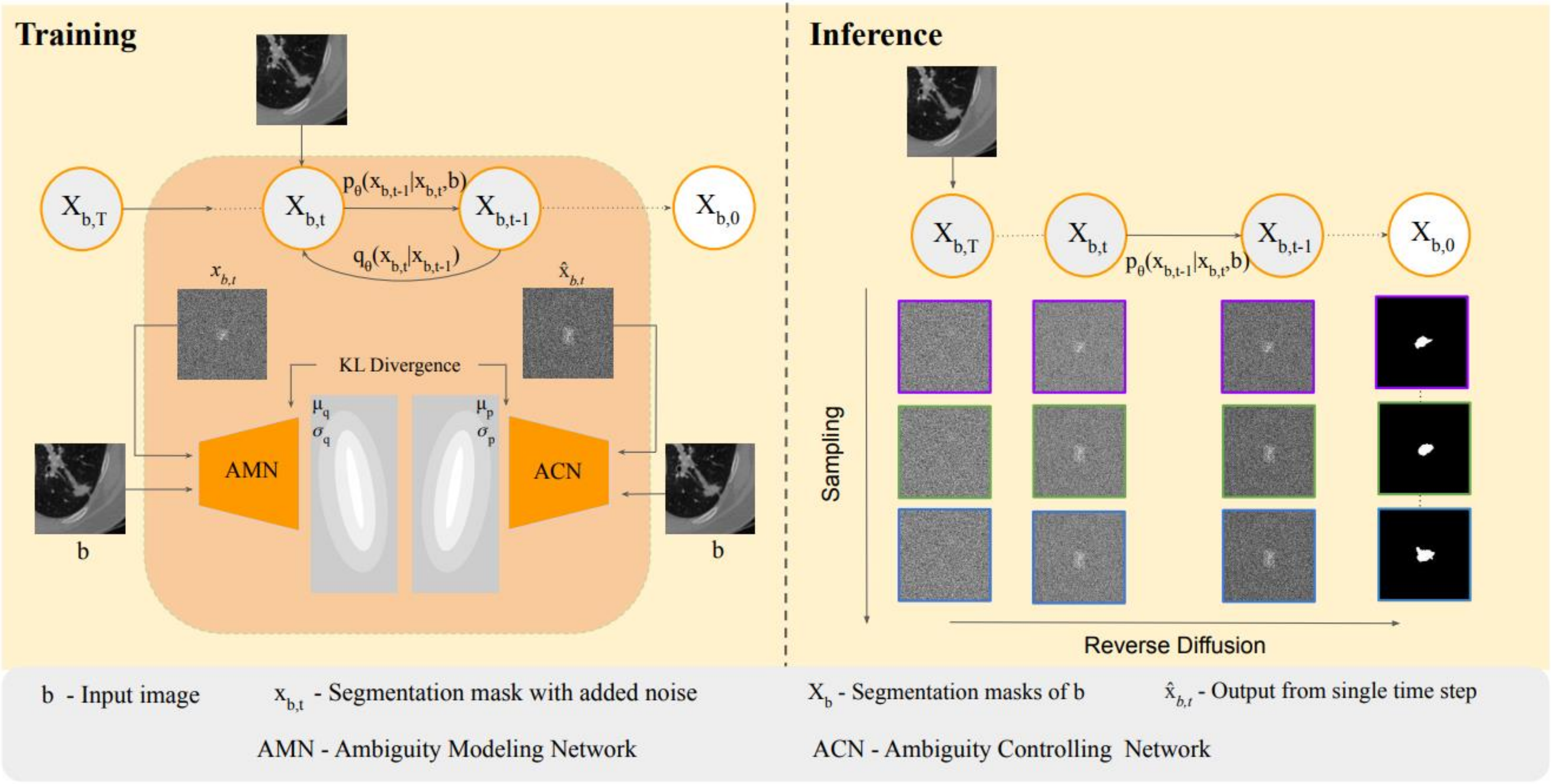
- We add a position encoding in all 3 spatial dimensions.
- Training is on patches only, and saves memory and training time.
- Inference runs over the whole 3D volume.

Bieder et al. (2023) Memory-Efficient 3D Denoising Diffusion Models for Medical Image Processing. MIDL 2023

Ambiguous Segmentation

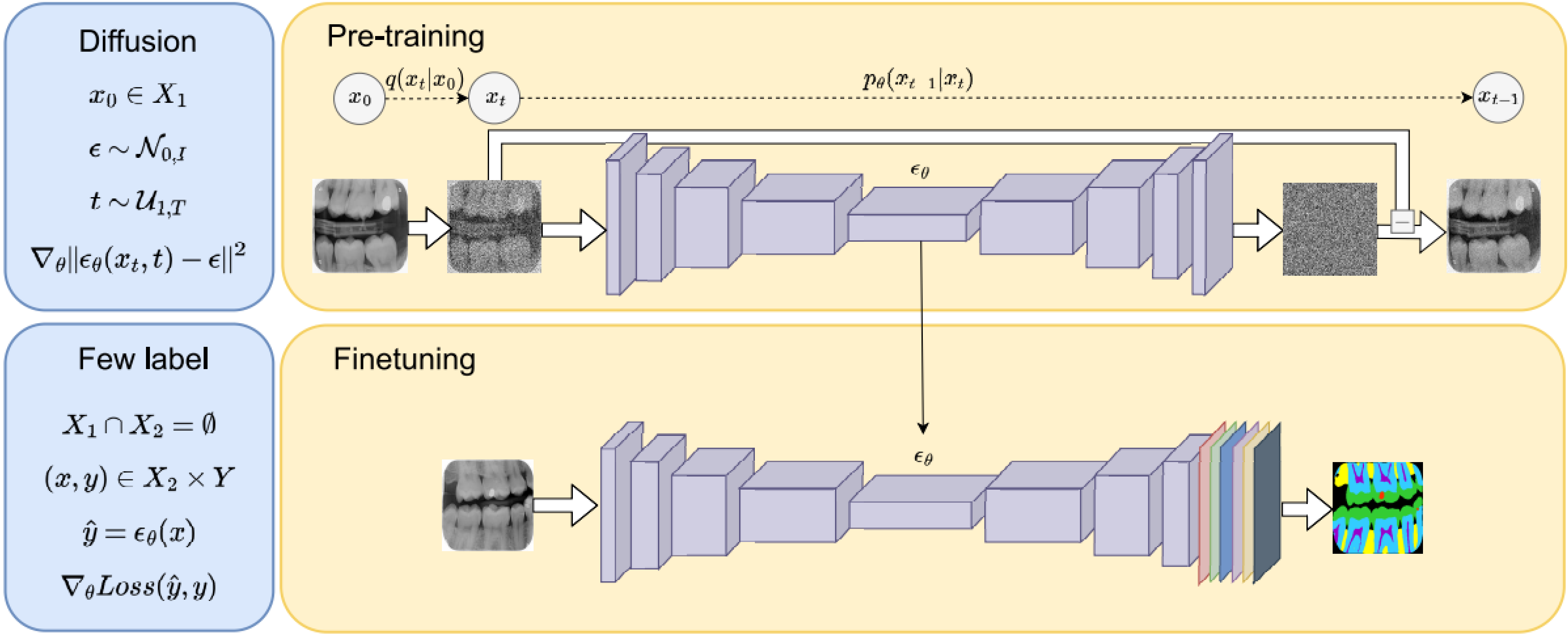
- Ambiguity Modelling Network (AMN) models the distribution of ground truth masks given an input image.

- Ambiguity Controlling Network (ACN) models the noisy output from the diffusion model conditioning on an input image.



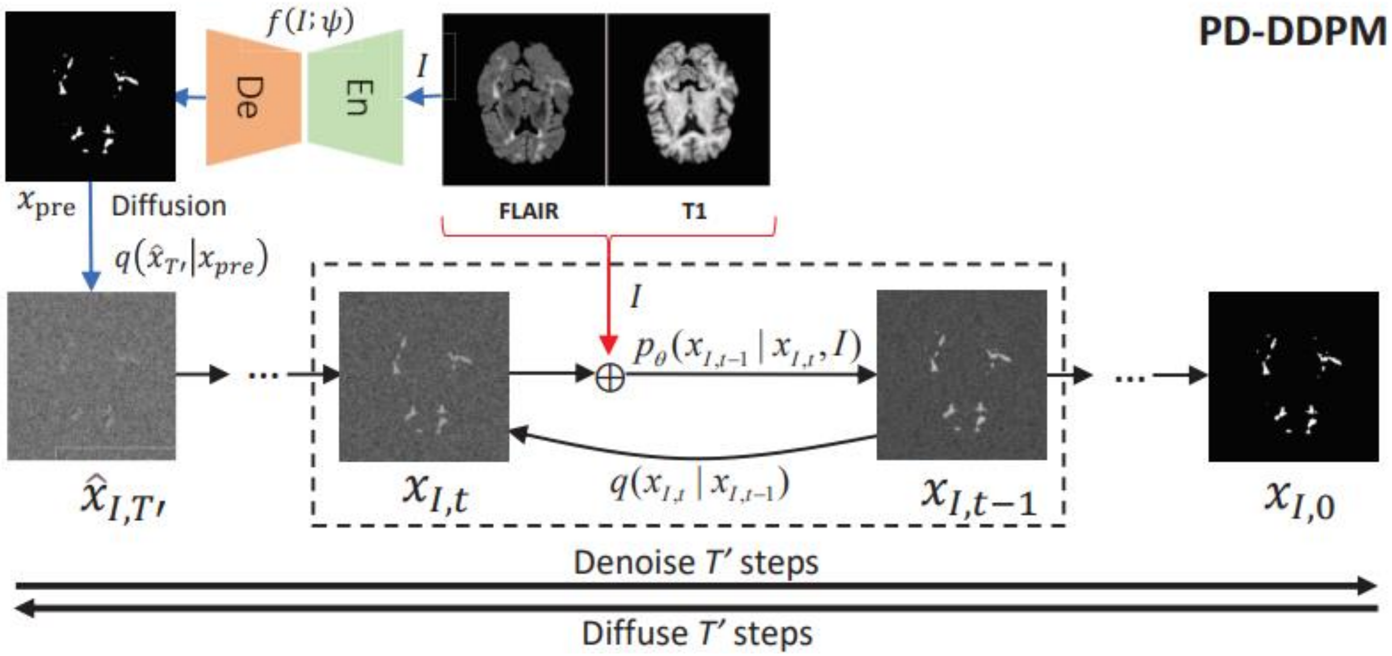
Rahman, Aimon, et al. (2023) Ambiguous medical image segmentation using diffusion models. CVPR

Segmentation with Diffusion Pre-training



Rousseau et al. (2023) Pre-Training with Diffusion models for Dental Radiography segmentation. MICCAI 2023

Acceleration of Diffusion Segmentation



- They obtain pre-segmentation predictions x_{pre} based on a separately trained segmentation network.
- They start with noisy predictions and use fewer reverse steps T' to generate segmentation results.

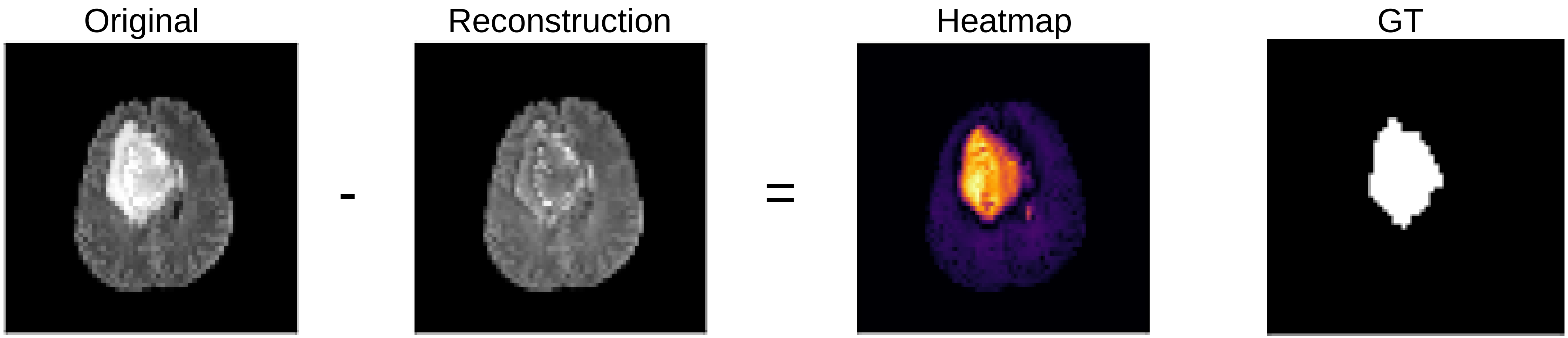
With a significantly smaller number of reverse sampling steps, PD-DDPM outperforms the vanilla DDPM

Guo et al. (2023) Accelerating Diffusion Models Via Pre-Segmentation Diffusion Sampling for Medical Image Segmentation. ISBI 2023

Anomaly detection

Examples from the community

The simple setup of the problem

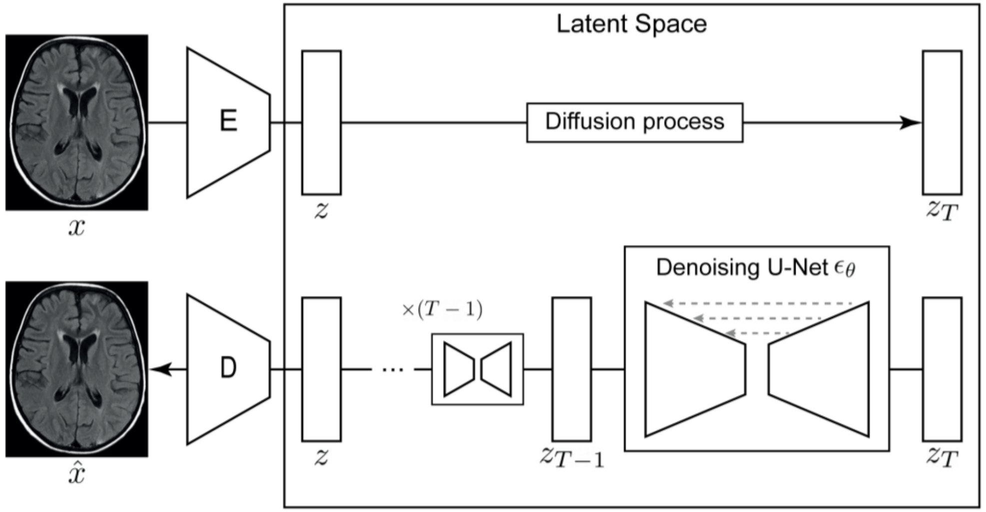


PAPERS

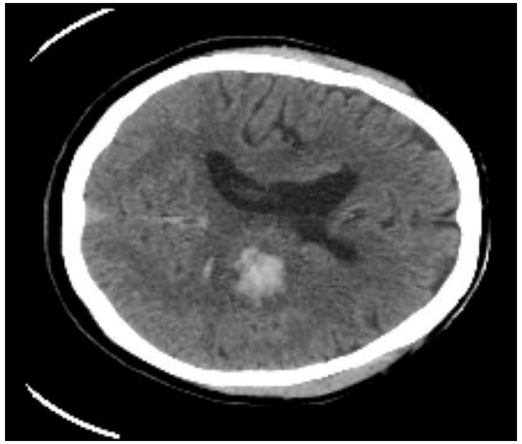
Pinaya et al (2022) Fast Unsupervised Brain Anomaly Detection and Segmentation with Diffusion Models. MICCAI 2022
Wyatt et al (2022) AnoDDPM: Anomaly Detection with Denoising Diffusion Probabilistic Models using Simplex Noise. CVPR 2022 workshop
Kascenas et al (2023) The role of noise in denoising models for anomaly detection in medical images. Medical Image Analysis 2023
Behrendt, Finn, et al. (2023) Patched diffusion models for unsupervised anomaly detection in brain mri. MIDL 2023
Liang, Ziyun, et al. (2023) Modality Cycles with Masked Conditional Diffusion for Unsupervised Anomaly Segmentation in MRI. MICCAI 2023
Wolleb et al (2022). Diffusion Models for Medical Anomaly Detection, MICCAI 2022
Sanchez et al. (2022) What is Healthy? Generative Counterfactual Diffusion for Lesion Localization. MICCAI 2022 workshop

Unsupervised Anomaly Segmentation

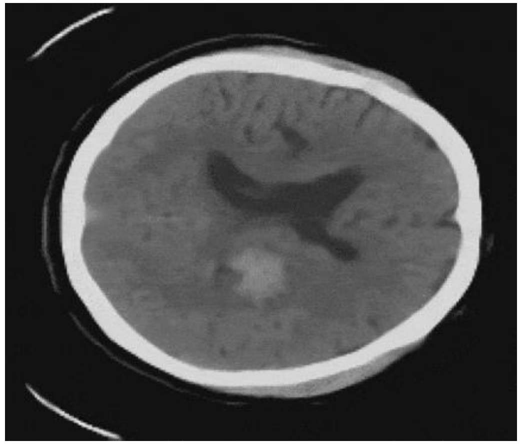
- Latent Diffusion Model (LDM) learns the distribution of healthy brain data
- Compression (Vector-Quantised VAE) scales for high-resolution images



LDM identify regions with a low likelihood of being part of the healthy dataset



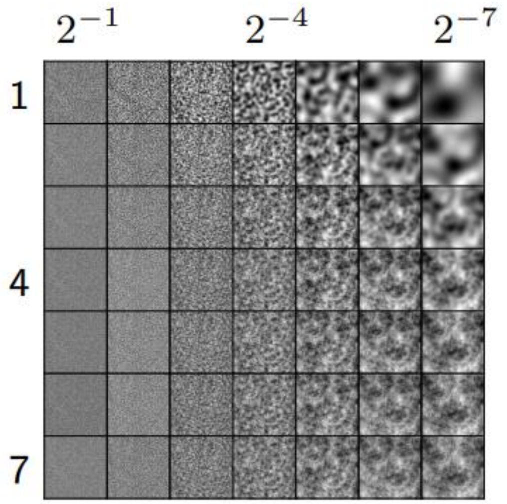
Reverse/denoising process is used to **inpaint** these regions and “**heal**” the possible anomalies



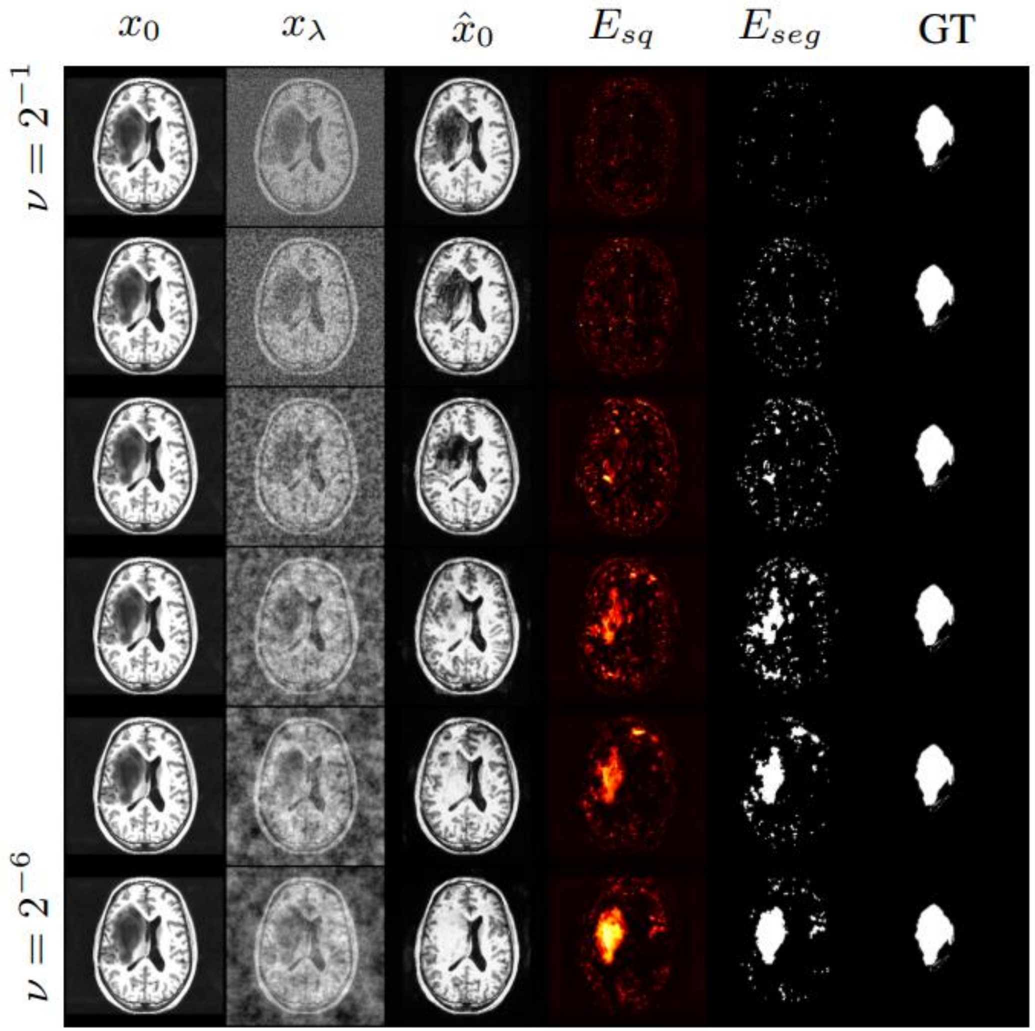
Pinaya et al (2022) Fast Unsupervised Brain Anomaly Detection and Segmentation with Diffusion Models. MICCAI 2022

Anomaly Detection with Simplex Noise

- Typical Gaussian noise is found to be insuffient for anomaly detection.
- Therefore, we explore the use of simplex noise for the corruption and sample generation of medical images.



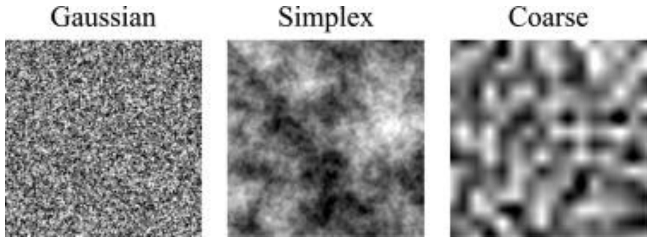
(a) Structures of simplex noise

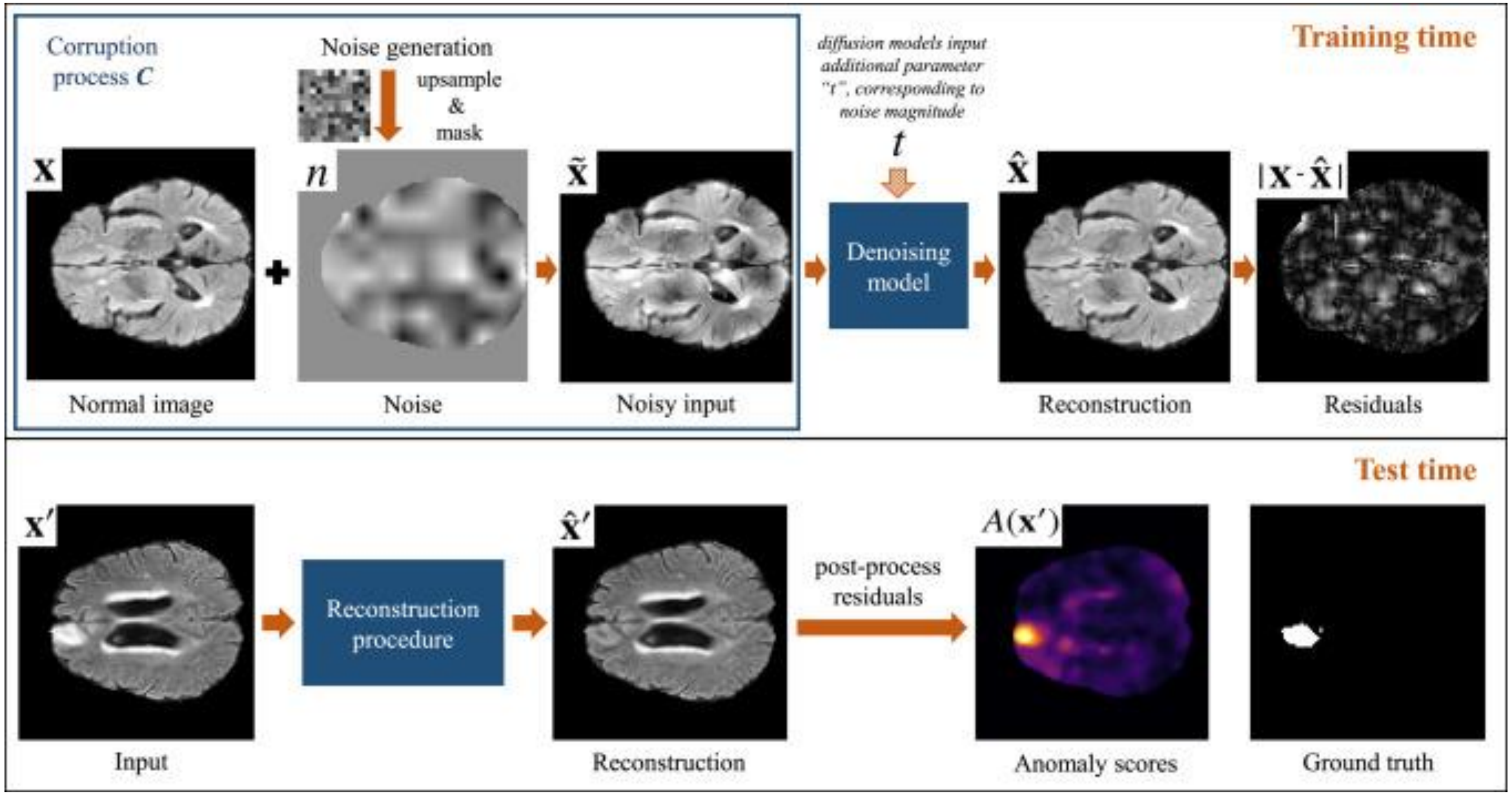


simplex noise scale *controls* target anomaly size

Wyatt et al (2022) AnoDDPM: Anomaly Detection with Denoising Diffusion Probabilistic Models using Simplex Noise. CVPR workshop

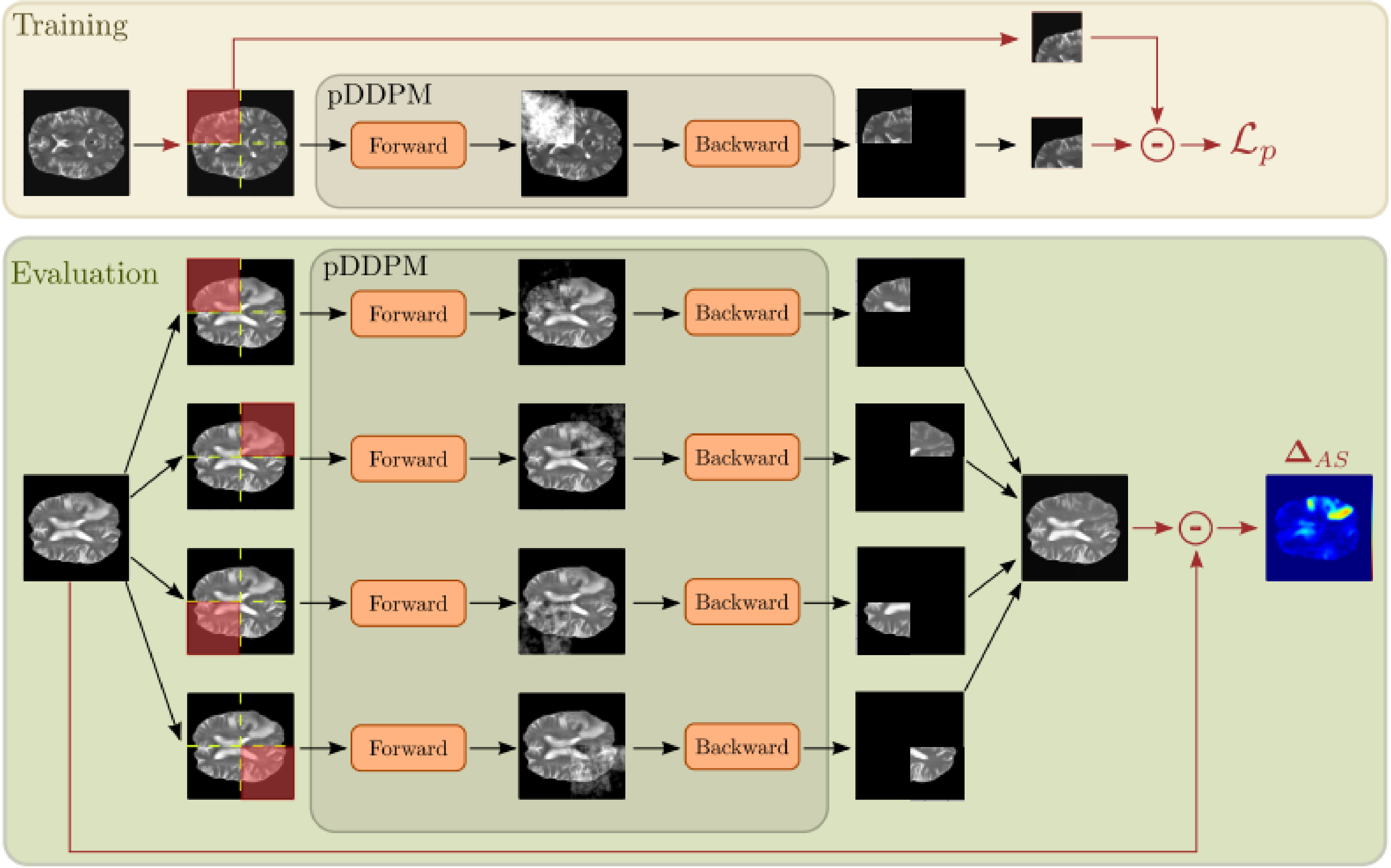
Anomaly Detection with Coarse Noise





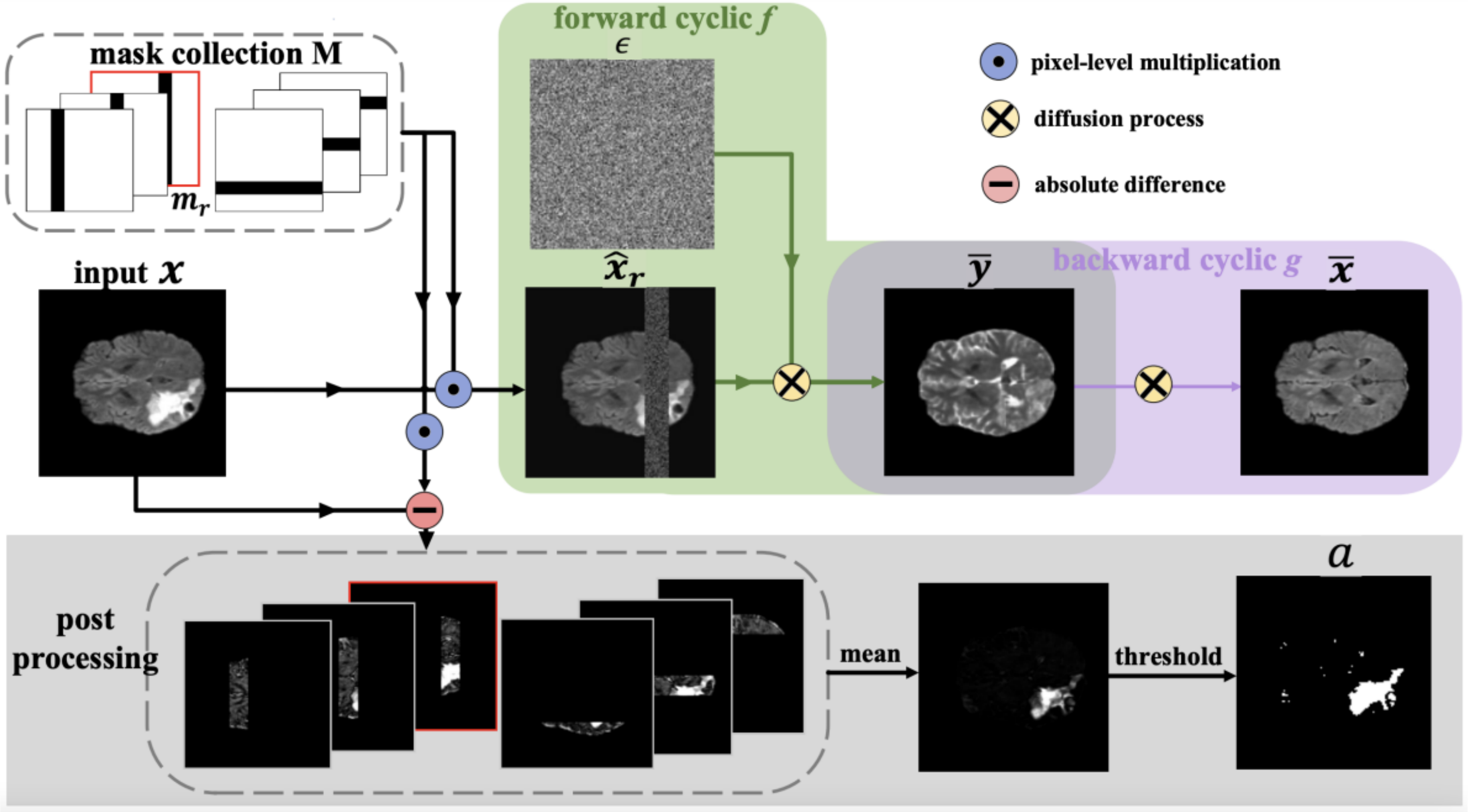
Kascenas et al (2023) The role of noise in denoising models for anomaly detection in medical images. Medical Image Analysis 2023

Anomaly Detection from Patches



Behrendt, Finn, et al. (2023) Patched diffusion models for unsupervised anomaly detection in brain mri. MIDL 2023

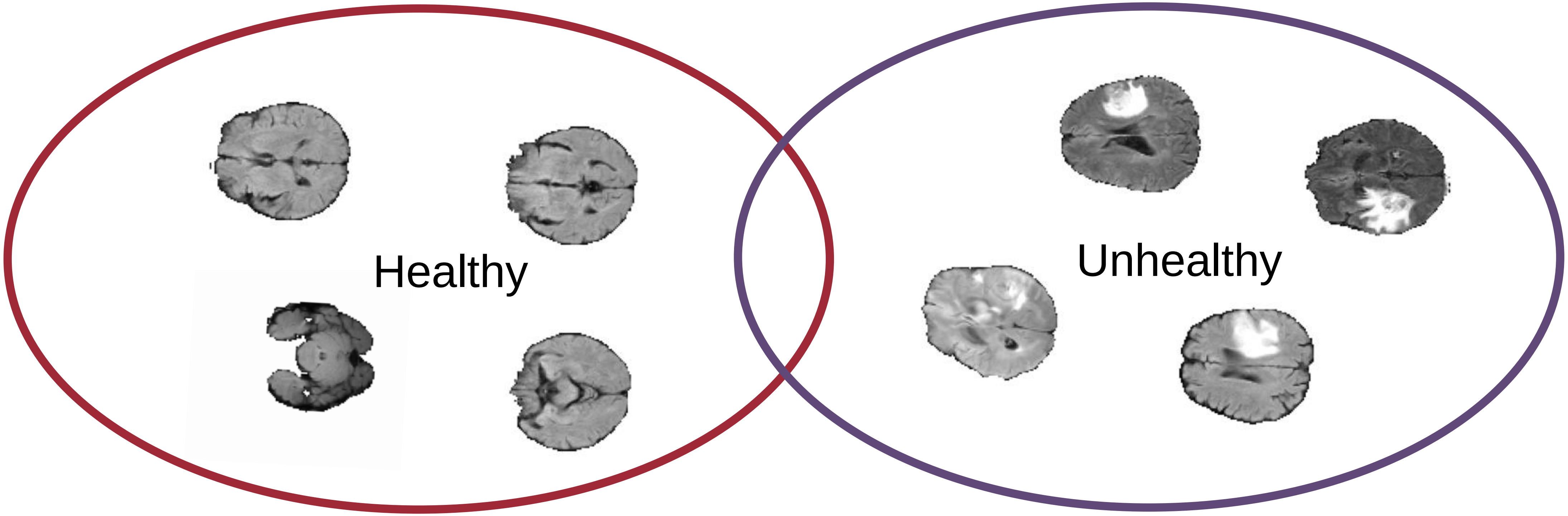
Anomaly Detection from Modality Cycles



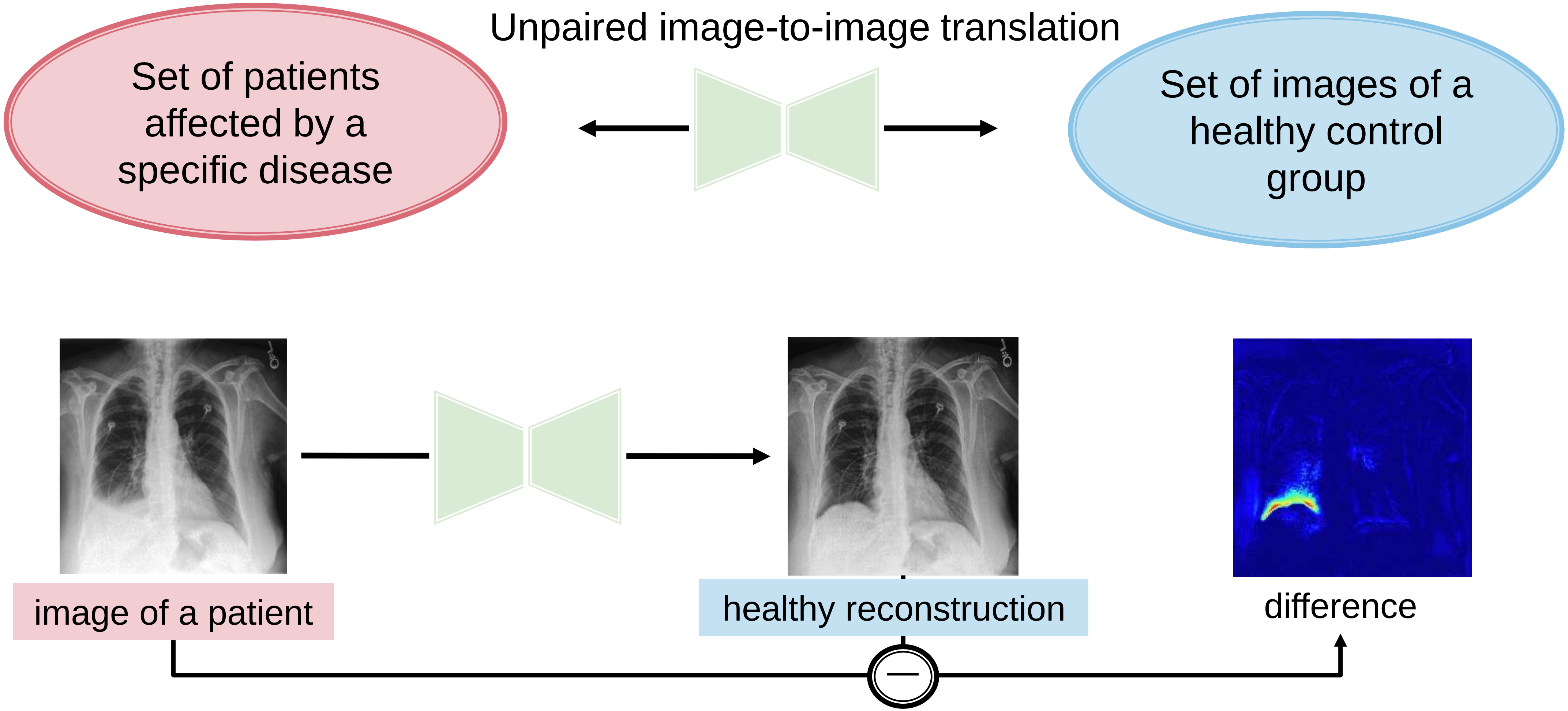
Liang, Ziyun, et al. (2023) Modality Cycles with Masked Conditional Diffusion for Unsupervised Anomaly Segmentation in MRI. MICCAI 2023

Weakly Supervised Lesion Detection

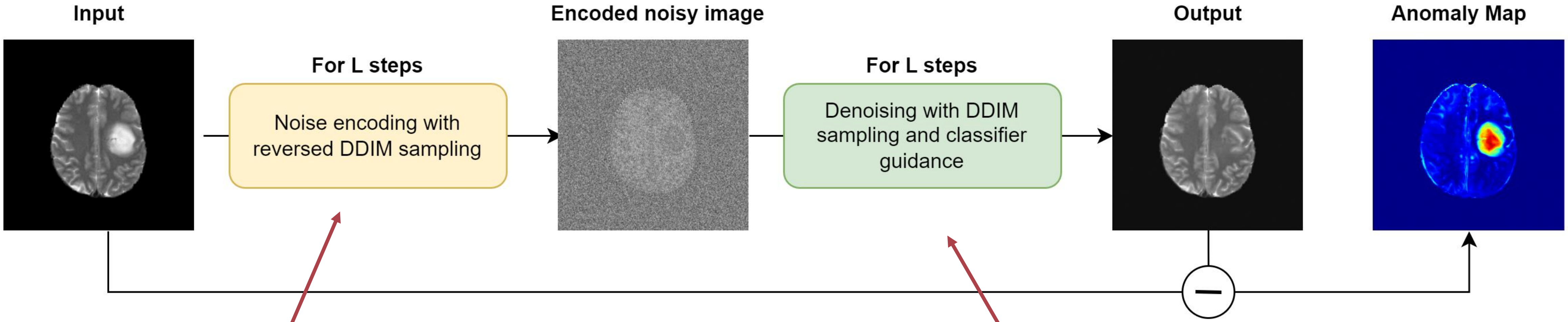
- **Goal:** Pixel-wise anomaly detection using image-level labels only



Weakly Supervised Lesion Detection

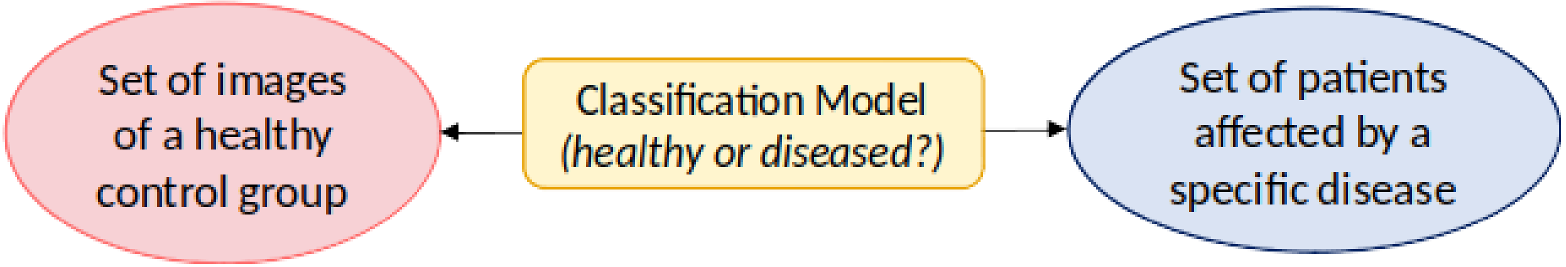


Weakly Supervised Lesion Detection



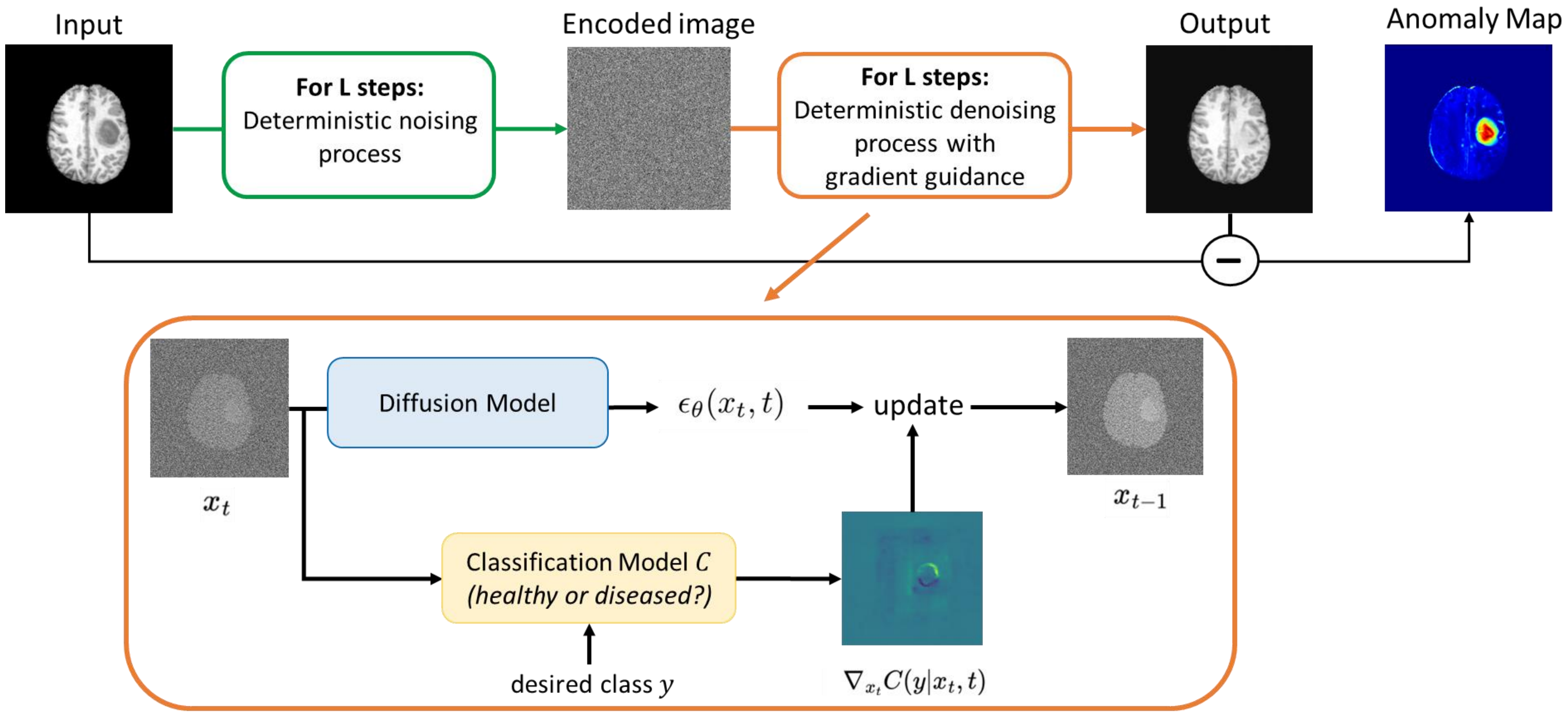
DDIM

Gradient Guidance



Wolleb et al (2022). Diffusion Models for Medical Anomaly Detection, *MICCAI 2022*

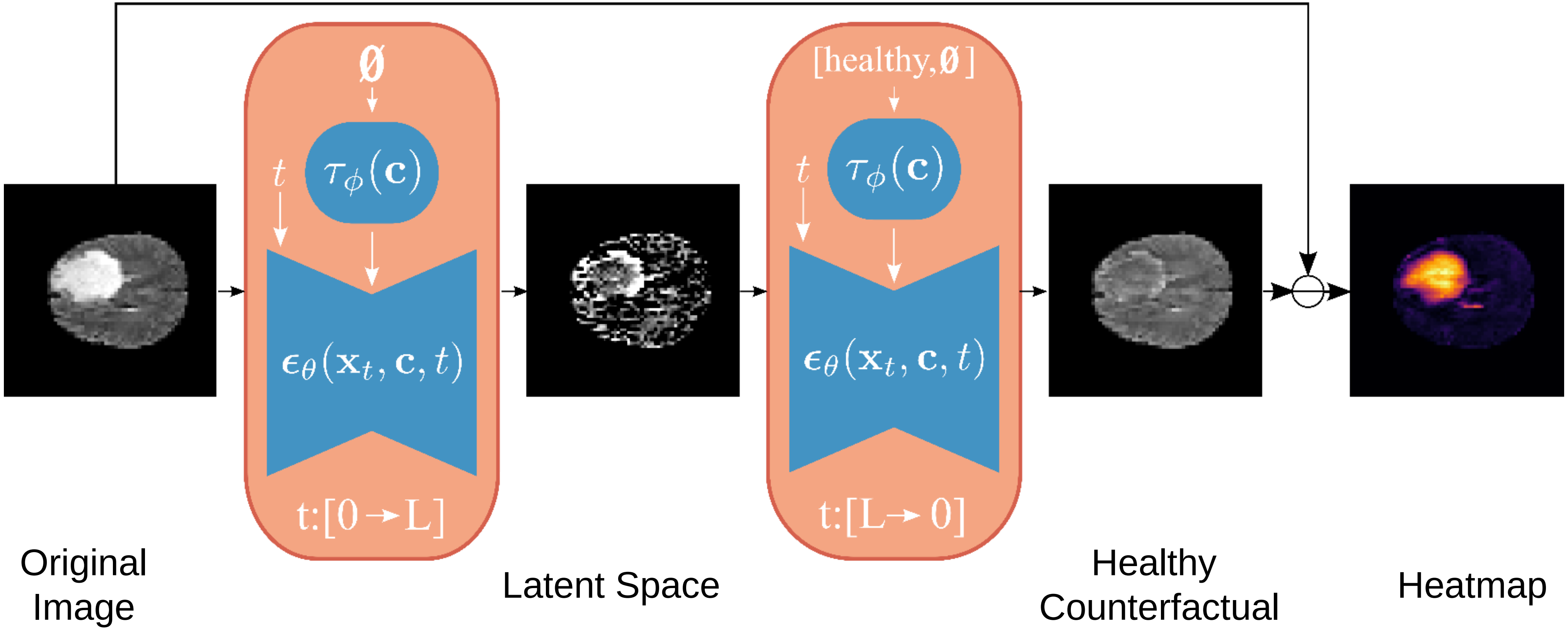
Gradient Guidance



Lesion Localization with Diffusion Models

Classifier-free guidance

1. DDIM Encoding - Empty condition
2. DDIM Decoding - Target class



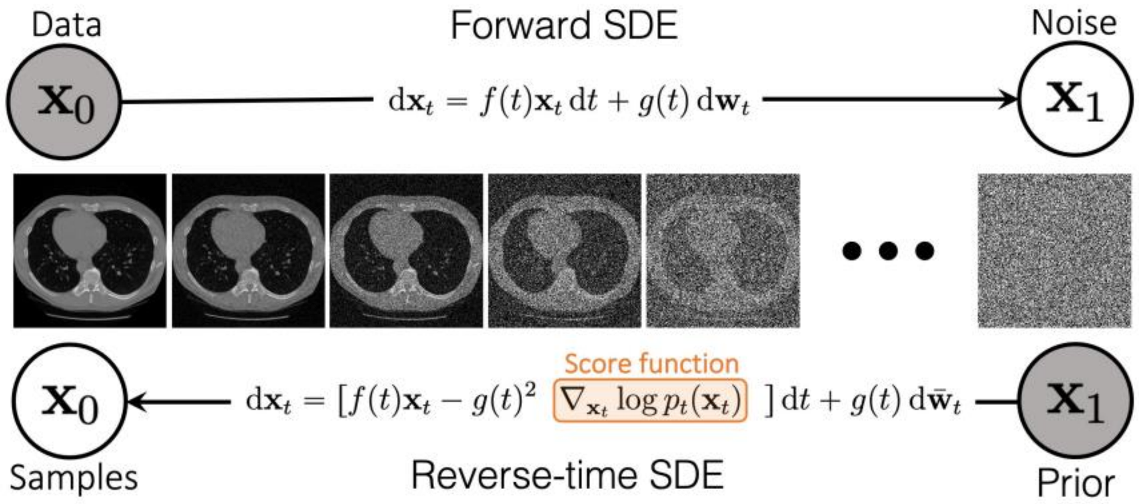
Sanchez et al (2022). What is Healthy? Generative Counterfactual Diffusion for Lesion Localization. MICCAI 2022 workshop

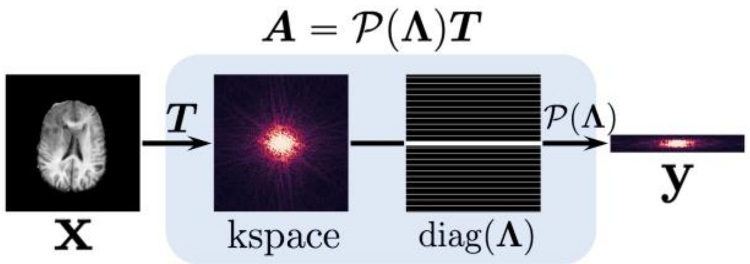
Image reconstruction

Examples from the community

Solving Inverse Problems in Medical Imaging with Score-Based Generative Models

- No need for paired data
- Learn from the prior
  - using a dataset of complete medical images to learn the underlying data distribution
- Conditional Sampling for inverse problem
  - $y = Ax$, y is the measurement and x is the image.
  - Equivalent results for supervised methods

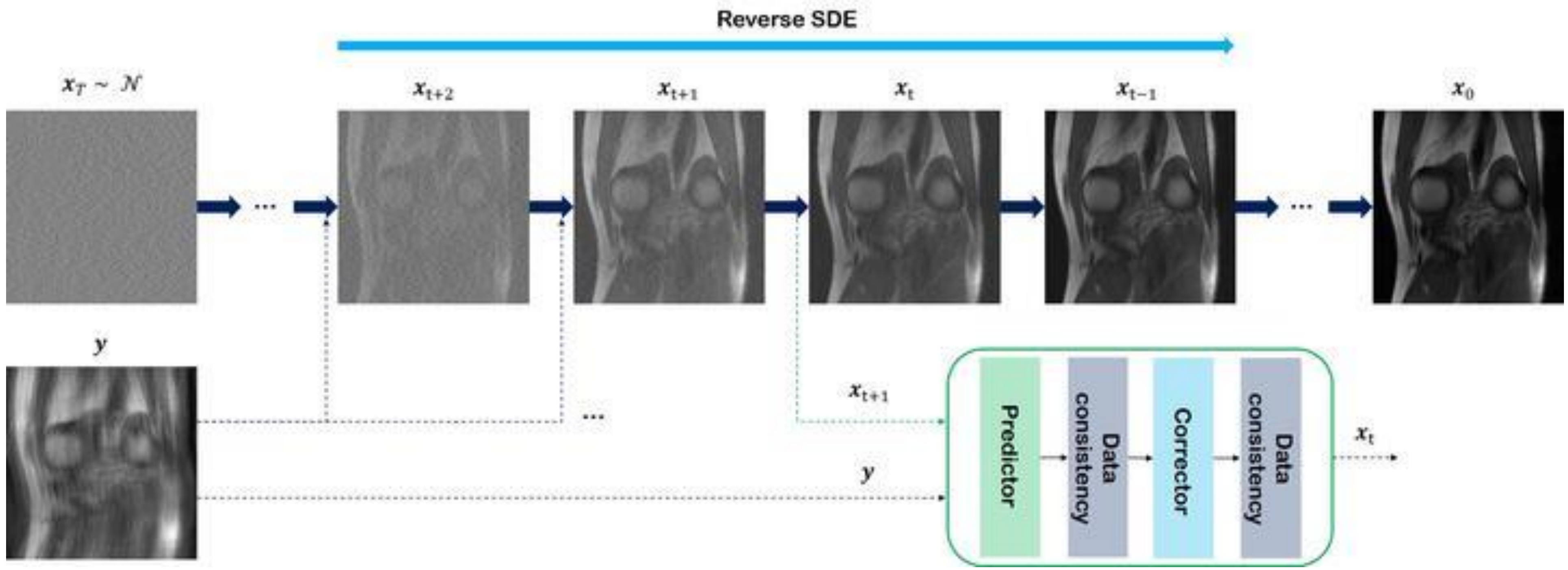




Song Y, Shen L, Xing L, et al. Solving inverse problems in medical imaging with score-based generative models[J]. arXiv preprint arXiv:2111.08005, 2021.

Score-based diffusion models for accelerated MRI

An unconditional diffusion prior is trained on fully-sampled MR acquisitions



Add a **data consistency** term at each sampling step:

$$\boldsymbol{x}_i \leftarrow \boldsymbol{x}_i + \lambda A^*(\boldsymbol{y} - A\boldsymbol{x}_i)$$

Chung H, Ye J C. Score-based diffusion models for accelerated MRI[J]. Medical image analysis, 2022, 80: 102479.

General Inverse Problems

$$\boldsymbol{y} = \mathcal{A}(\boldsymbol{x}_0) + \boldsymbol{n}, \quad \boldsymbol{y}, \boldsymbol{n} \in \mathbb{R}^n,\ \boldsymbol{x} \in \mathbb{R}^d$$

Algorithm 1 DPS - Gaussian

Require: $N, \boldsymbol{y}, \{\zeta_i\}_{i=1}^N, \{\tilde{\sigma}_i\}_{i=1}^N$

$\boldsymbol{x}_N \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$

for $i = N-1$ **to** 0 **do**

$\quad \hat{\boldsymbol{s}} \leftarrow \boldsymbol{s}_\theta(\boldsymbol{x}_i, i)$

$\quad \hat{\boldsymbol{x}}_0 \leftarrow \frac{1}{\sqrt{\bar{\alpha}_i}}(\boldsymbol{x}_i + (1-\bar{\alpha}_i)\hat{\boldsymbol{s}})$

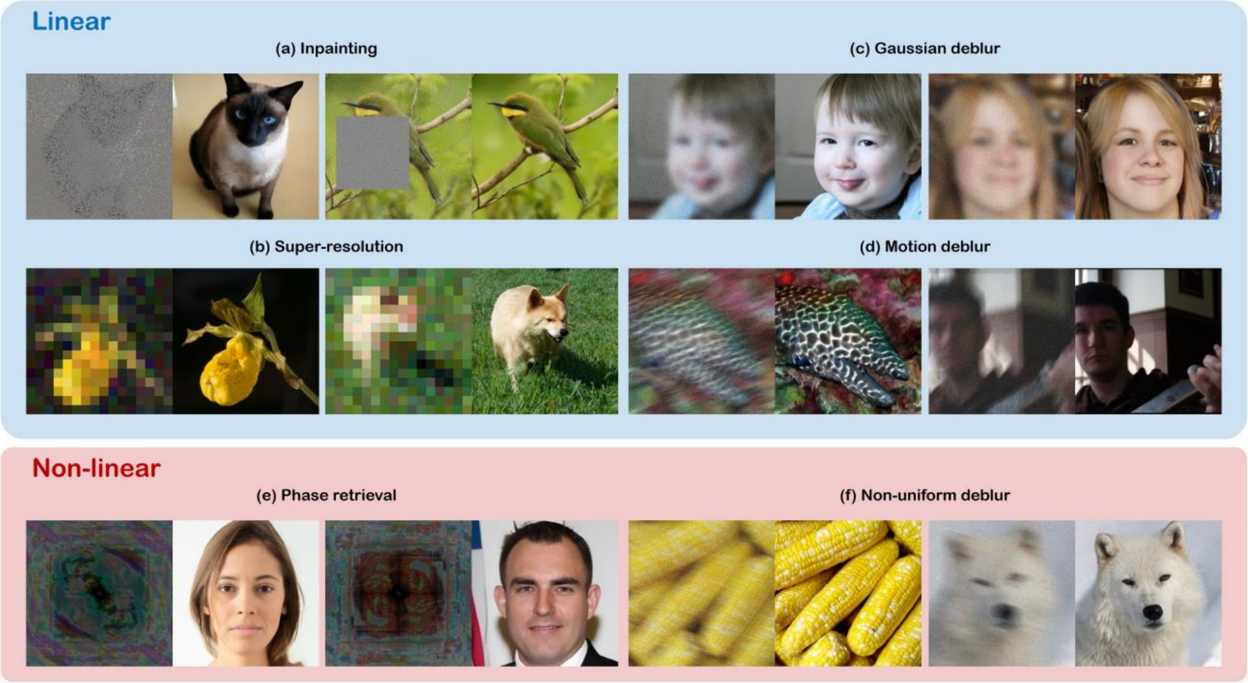
$\quad \boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$

$\quad \boldsymbol{x}'_{i-1} \leftarrow \frac{\sqrt{\alpha_i}(1-\bar{\alpha}_{i-1})}{1-\bar{\alpha}_i}\boldsymbol{x}_i + \frac{\sqrt{\bar{\alpha}_{i-1}}\beta_i}{1-\bar{\alpha}_i}\hat{\boldsymbol{x}}_0 + \tilde{\sigma}_i \boldsymbol{z}$

$\quad \boldsymbol{x}_{i-1} \leftarrow \boldsymbol{x}'_{i-1} - \zeta_i \nabla_{\boldsymbol{x}_i} \|\boldsymbol{y} - \mathcal{A}(\hat{\boldsymbol{x}}_0)\|_2^2$

end for

return $\hat{\mathbf{x}}_0$



Chung H, Kim J, Mccann M T, et al,[J]. arXiv preprint arXiv:2209.14687, 2022.

Image registration

Examples from the community

DiffuseMorph

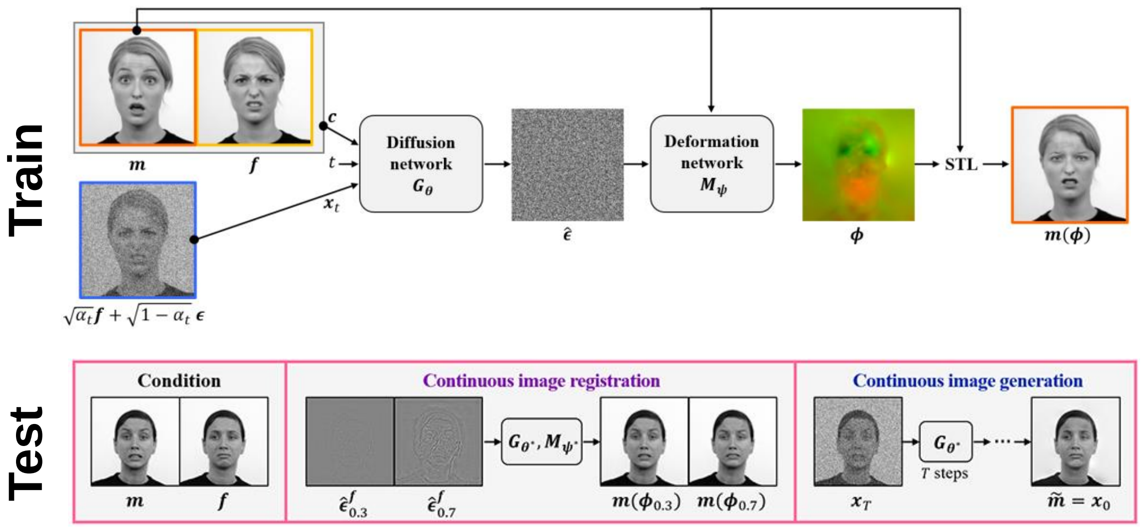
- To perform image registration along the continuous trajectory
- **Diffusion network**: To estimate a conditional score function
- **Deformation network**: To yield the registration fields & provide the deformed image

Loss function

$$\min_{G_\theta, M_\psi} \boldsymbol{L_{diffusion}}(c, x_t, t) + \lambda \boldsymbol{L_{regist}}(m, f)$$

$$\boldsymbol{L_{diffusion}}(c, x_t, t) = \mathbb{E}_{\epsilon, x_t, t} \| G_\theta(c, x_t, t) - \epsilon \|_2^2$$

$$\boldsymbol{L_{regist}}(m, f) = -(m(\phi) \otimes f) + \lambda_\phi \Sigma \| \nabla \phi \|^2$$



Algorithm 1 Continuous image registration

Input: Conditional images, $c = (m, f)$
Output: Deformed moving image, $m(\phi_\eta)$
Set the latent feature $\hat{\epsilon}^f = G_{\theta^*}(c, f, 0)$
for $\eta \in [0, 1]$ **do**
  $\hat{\epsilon}^f_\eta \leftarrow \eta \cdot \hat{\epsilon}^f$
  $\phi_\eta \leftarrow M_{\psi^*}(m, \hat{\epsilon}^f_\eta)$
end for=
return $m(\phi_\eta)$

Algorithm 2 Synthetic image generation process

Input: Conditional images, $c = (m, f)$
Output: Synthetic deformed image, x
Set $T \in (0, T_{train})$
Sample $x_T = \sqrt{\alpha_T} m + \sqrt{1 - \alpha_T}\epsilon$, where $\epsilon \sim \mathcal{N}(0, I)$
for $t = T, T-1, ..., 1$ **do**
  $z \sim \mathcal{N}(0, I)$
  $x_{t-1} \leftarrow \frac{1}{\sqrt{1-\beta_t}}(x_t - \frac{\beta_t}{\sqrt{1-\alpha_t}} G_{\theta^*}(c, x_t, t)) + \sigma_t z$
end for
return x_0

Slides courtesy of Boah Kim & Jong Chul Ye

1. Kim et al (2022). DiffuseMorph: Unsupervised Deformable Image Registration Along Continuous Trajectory Using Diffusion Models. ECCV 2022

Useful key references, gits to watch etc

- Surveys
  - https://arxiv.org/abs/2404.07771
  - https://arxiv.org/abs/2209.02646
  - https://arxiv.org/abs/2209.00796
- Github
  - https://github.com/heejkoo/Awesome-Diffusion-Models
- Tutorial
  - https://arxiv.org/pdf/2403.18103.pdf
  - https://cvpr2022-tutorial-diffusion-models.github.io
  - https://huggingface.co/blog/annotated-diffusion
  - https://huggingface.co/docs/diffusers

DiMEDIA:
Diffusion Models in Medical Imaging and Analysis

https://vios.science/tutorials/DiMEDIA-2024

Nefeli Gkouti
Archimedes/Athena RC,
University of Athens

Prof. Sotirios A. Tsaftaris
Archimedes/Athena RC,
University of Edinburgh

DiMEDIA 2024