University of Basel

Department of
Biomedical Engineering

A gentle introduction to **Diffusion Models for Medical Image Analysis**

30.01.2024

Paul Friedrich, MSc

Julia Wolleb, PhD

Center for medical Image Analysis and Navigation (CIAN), University of Basel

A Quick Recap on Generative Modelling



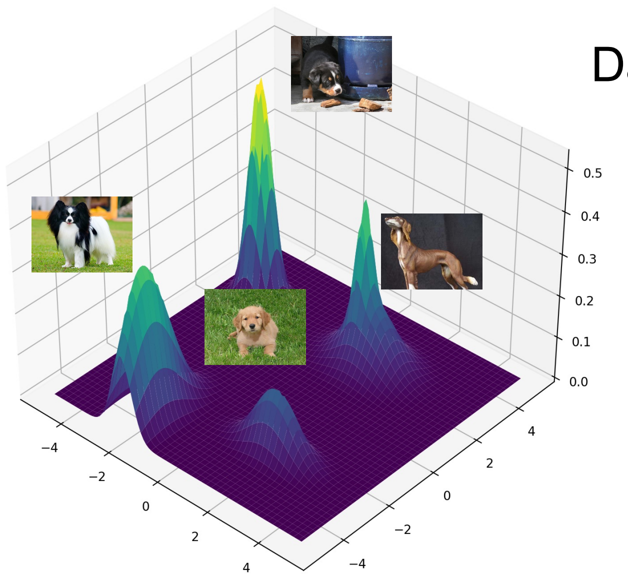
...



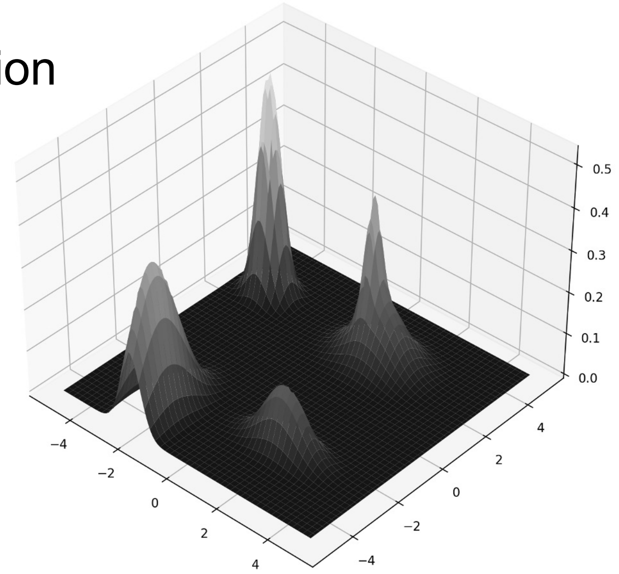
Data samples

(e.g. Stanford Dogs)

A Quick Recap on Generative Modelling



Data distribution **(unknown)** $\longleftrightarrow^{\approx}$ Model distribution

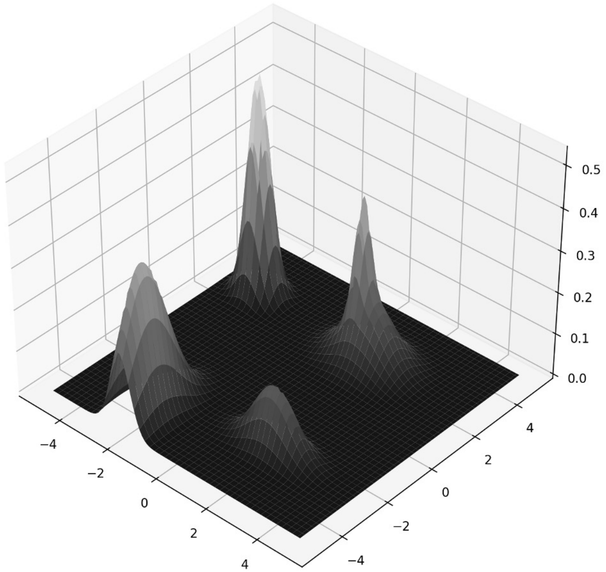


Datapoints are *i.i.d.* samples of this underlying data distribution

We can define a parameterized distribution that we tune to be close to the data distribution

A Quick Recap on Generative Modelling

Sampling



Probability evaluation

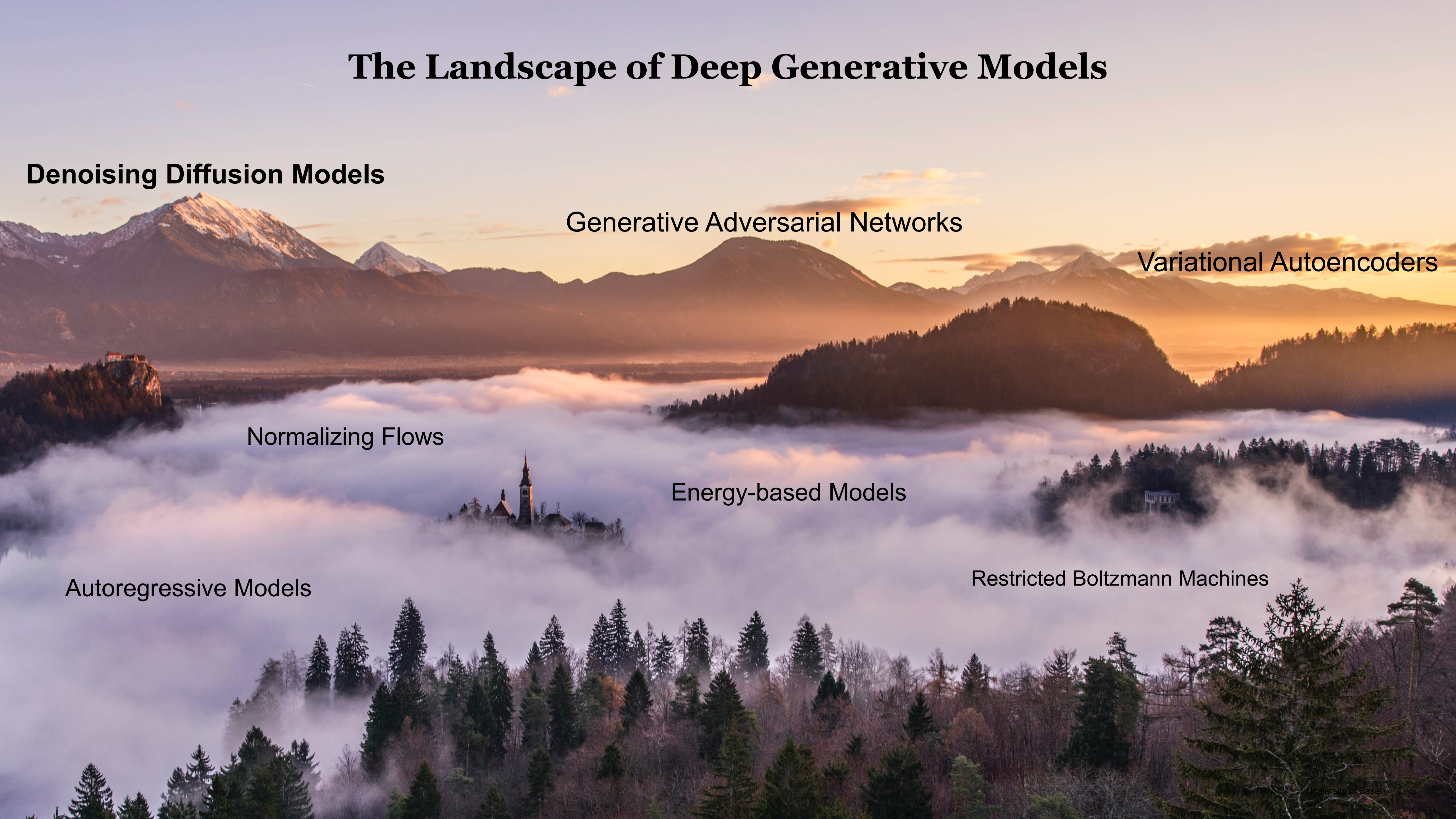
Model distribution
= Generative Model


Novel data points

High probability

Low probability

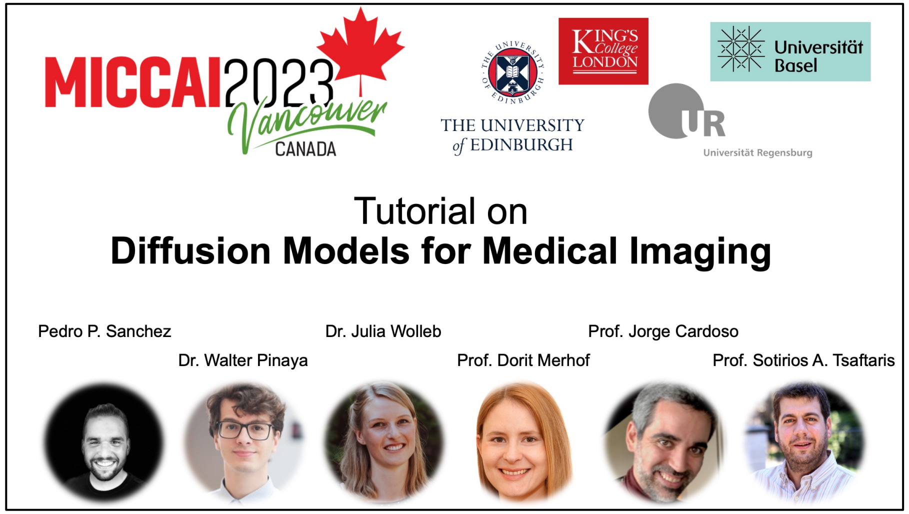
The Landscape of Deep Generative Models



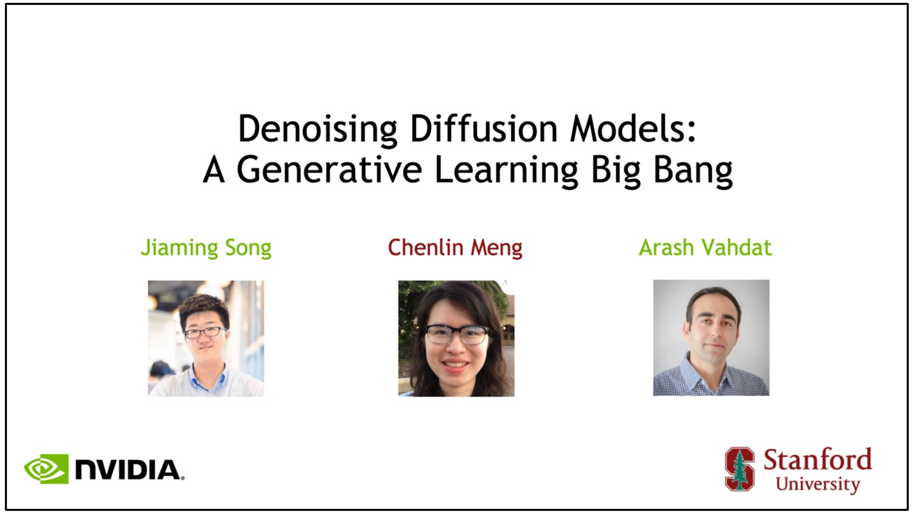
Agenda

- Introduction to Diffusion Models [~30 min]
  - Physical Intuition & General Concepts
  - Denoising Diffusion Probabilistic Models
  - A Score-Based View on Diffusion Models
- Advanced Topics [~30 min]
  - Sampling Strategies
  - Inference-time Conditioning
  - Training-time Conditioning
- Applications in Medical Imaging [~30 min]
  - Synthesis
  - Inpainting
  - Segmentation
  - Anomaly Detection
  - Reconstruction
  - Registration

Acknowledgement



MICCAI 2023 - Tutorial



CVPR 2023 - Tutorial

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

The Physical Intuition behind Diffusion Models

Deep Unsupervised Learning using
Nonequilibrium Thermodynamics

Jascha Sohl-Dickstein JASCHA@STANFORD.EDU
Stanford University

Eric A. Weiss EWEISS@BERKELEY.EDU
University of California, Berkeley

Niru Maheswaranathan NIRUM@STANFORD.EDU
Stanford University

Surya Ganguli SGANGULI@STANFORD.EDU
Stanford University

PMLR (2015)

The Physical Intuition behind Diffusion Models (Macroscopic)

The dye density represents our probability density.

Goal: We want to *learn this probability density*

Observing this diffusion process:

- Original data distribution is perturbed over time
- Data distribution → Uniform distribution (mapping to a simple distribution)

The Physical Intuition behind Diffusion Models (Macroscopic)

The dye density represents our probability density.

Goal: We want to *learn this probability density*

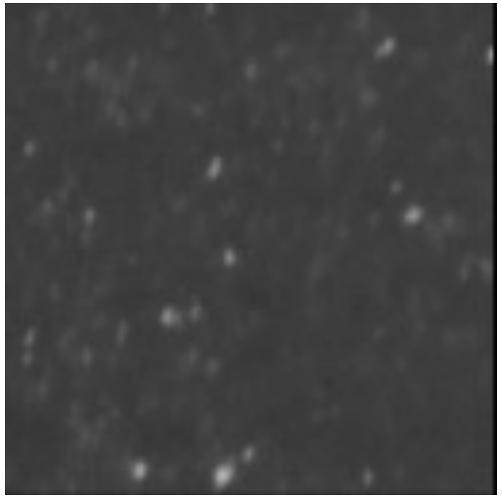
Observing this diffusion process:

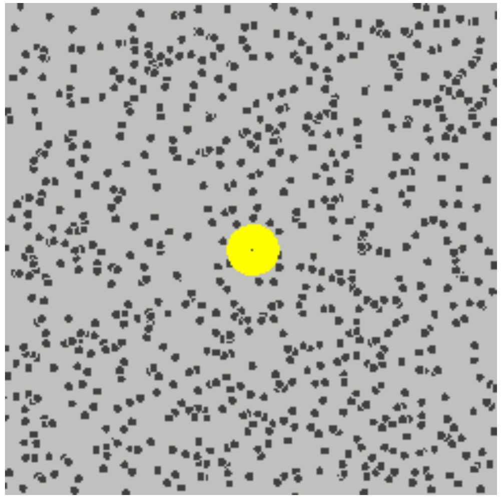
- Original data distribution is perturbed over time
- Data distribution → Uniform distribution (mapping to a simple distribution)

Can we learn to revert this process (run it backwards)?

- Uniform distribution → Data distribution
- **Yes**, but we first need a way to model the system.

The Physical Intuition Behind Diffusion Models (Microscopic)





We can try to model the diffusion process by **modelling the Brownian motion** of single particles.

- We can observe that the position updates follow **small Gaussians**
- This holds true for the **forward as well as the reverse process** (for small enough Δt)
- We can define a known diffusion process with a chain of Gaussian position updates
- We try to learn the reverse process by estimating the mean and the covariance of the backward steps

→ Same mechanism is used in the Diffusion Models we will see today!

Denoising Diffusion Probabilistic Models

Denoising Diffusion Probabilistic Models

Jonathan Ho
UC Berkeley
jonathanho@berkeley.edu

Ajay Jain
UC Berkeley
ajayj@berkeley.edu

Pieter Abbeel
UC Berkeley
pabbeel@cs.berkeley.edu

NeurIPS (2020)

Improved Denoising Diffusion Probabilistic Models

Alex Nichol [*1] **Prafulla Dhariwal** [*1]

ICML (2021)

Diffusion Models Beat GANs on Image Synthesis

Prafulla Dhariwal[*]
OpenAI
prafulla@openai.com

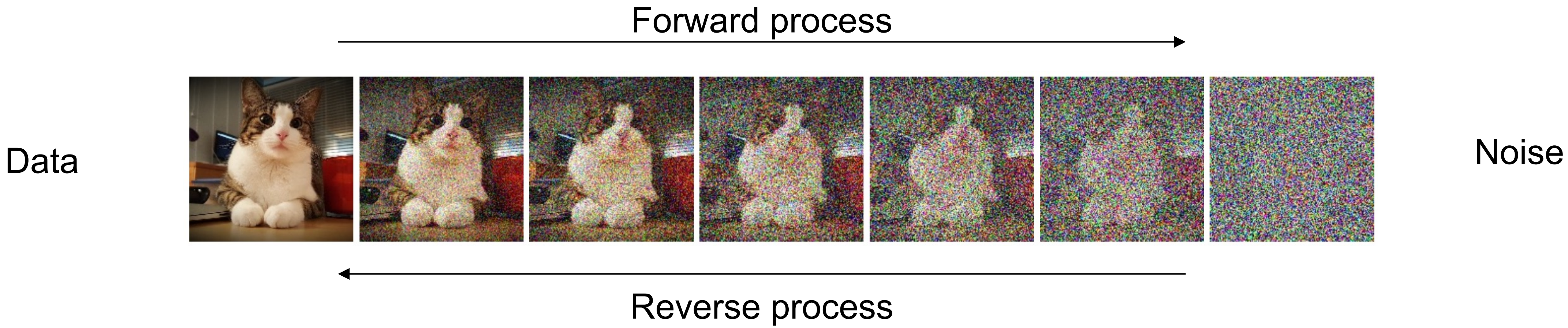
Alex Nichol[*]
OpenAI
alex@openai.com

NeurIPS (2021)

How do Diffusion Models work?

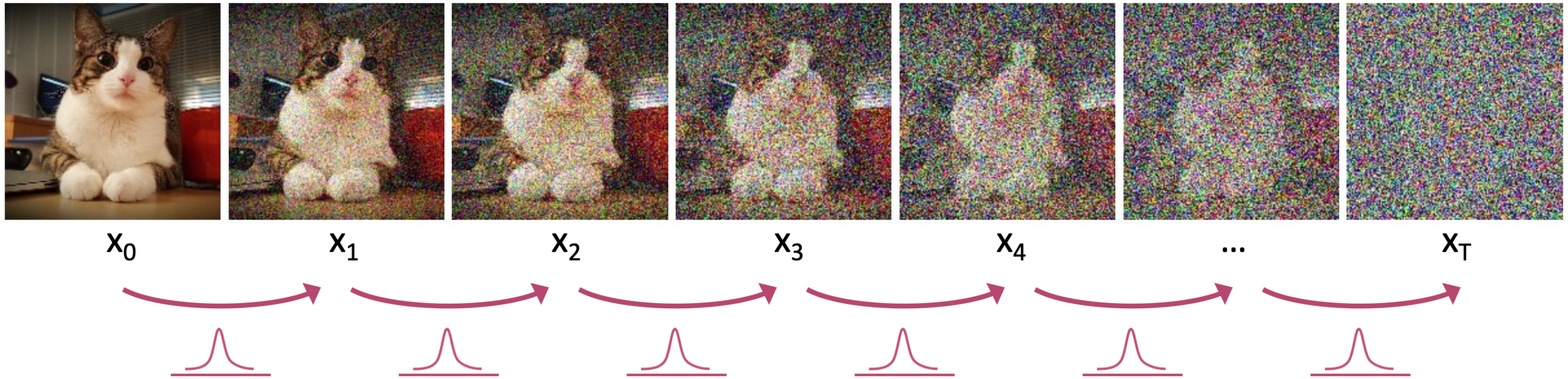
Diffusion Models consist of two main components:

- A ***fixed forward diffusion process*** that gradually adds noise to the image
- A ***learned reverse diffusion process*** that gradually removes noise from the image



The Forward Diffusion Process

Data



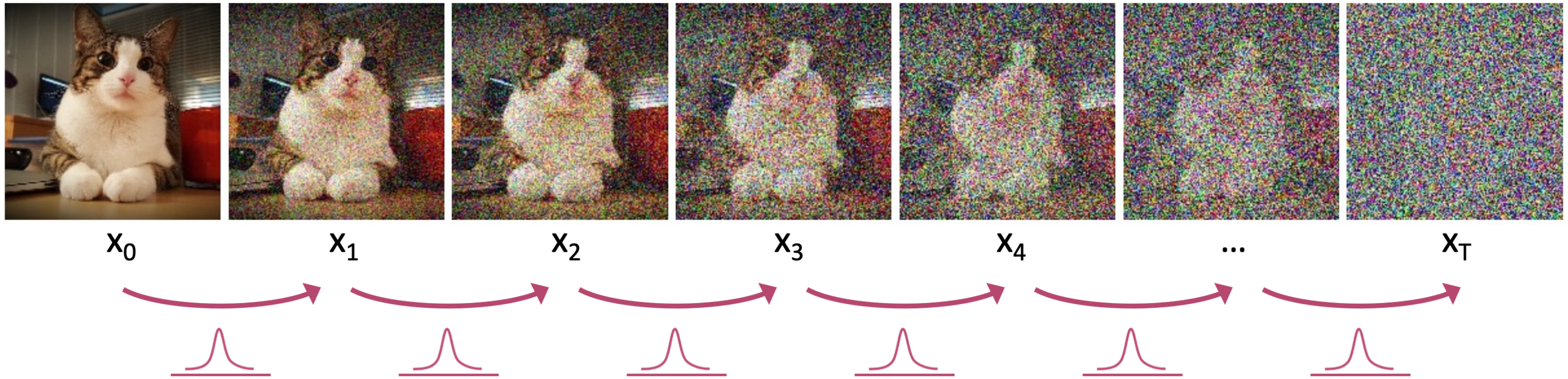
Noise

We model the **forward process** as a Markov chain: $q(x_{1:T}|x_0) = \prod_{t=1}^{T} q(x_t|x_{t-1})$

with each transition being a parameterized Gaussian: $q(x_t|x_{t-1}) = N(\sqrt{1-\beta_t}\, x_{t-1}, \beta_t \boldsymbol{I})$

The Forward Diffusion Process

Data



Noise

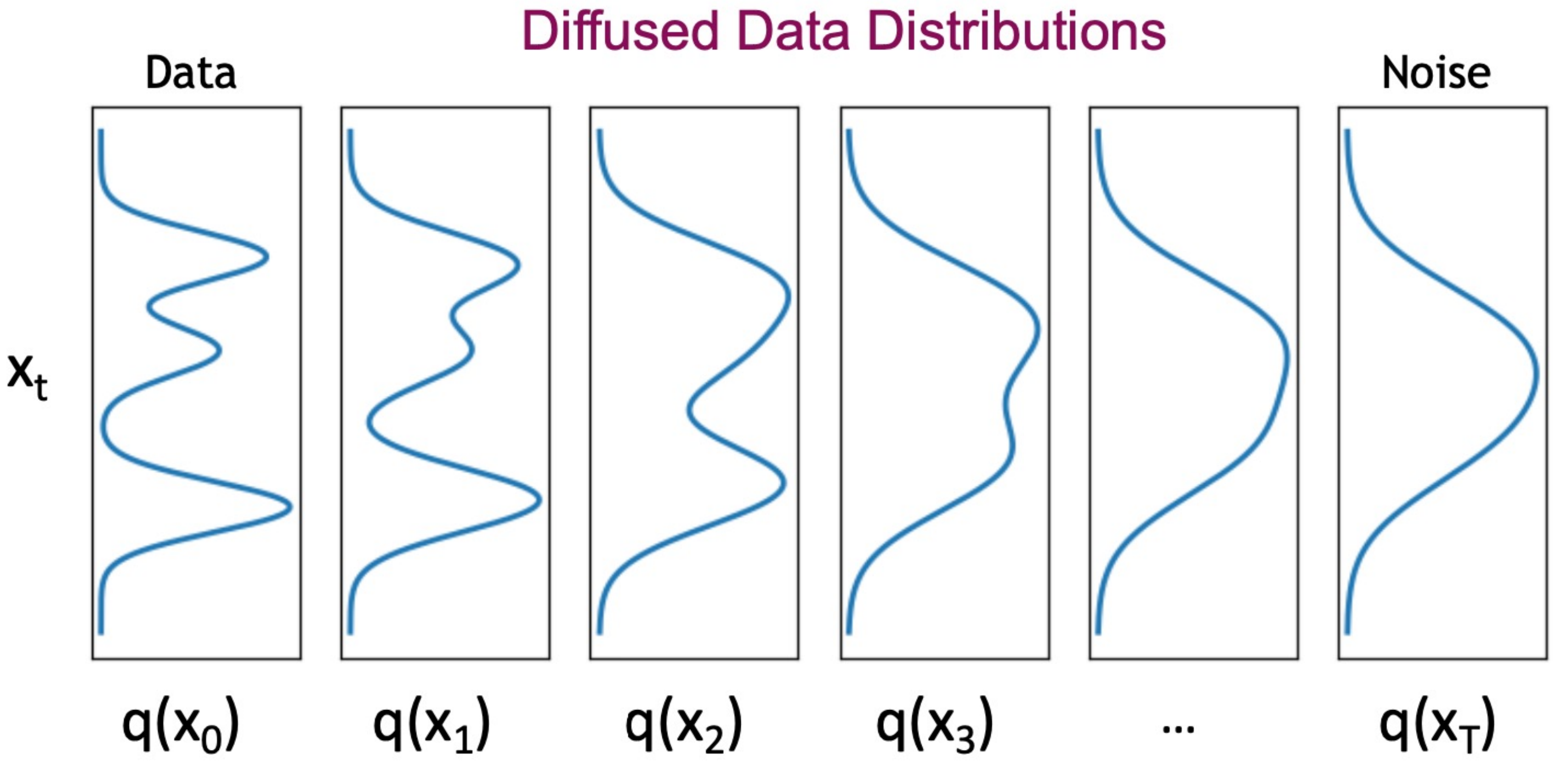
$$q(x_t|x_{t-1}) = N(\sqrt{1-\beta_t}\, x_{t-1}, \beta_t \boldsymbol{I})$$

We define a variance schedule for β_t

$$q(x_T|x_0) \approx N(0, \boldsymbol{I})$$

We usually choose $T \approx 1000$ (this remains a design choice)

The Forward Diffusion Process



We defined a forward process that transforms our data distribution to noise.

The Reverse Diffusion Process

Recall, the diffusion process is designed in a way that:

$$q(x_T) \approx N(0, \boldsymbol{I})$$

We could generate new samples by:

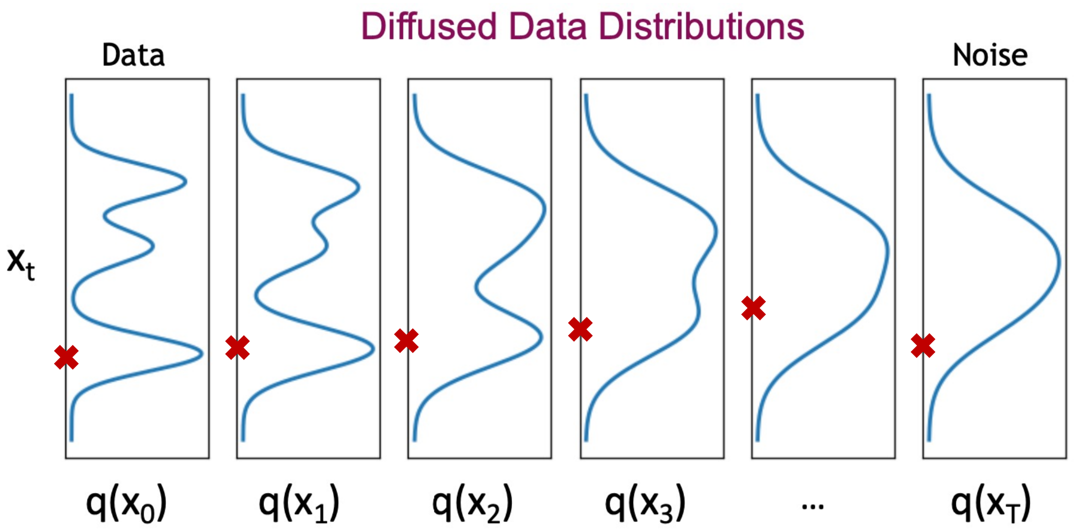
- Sampling x_T:

$$x_T \sim N(0, \boldsymbol{I})$$

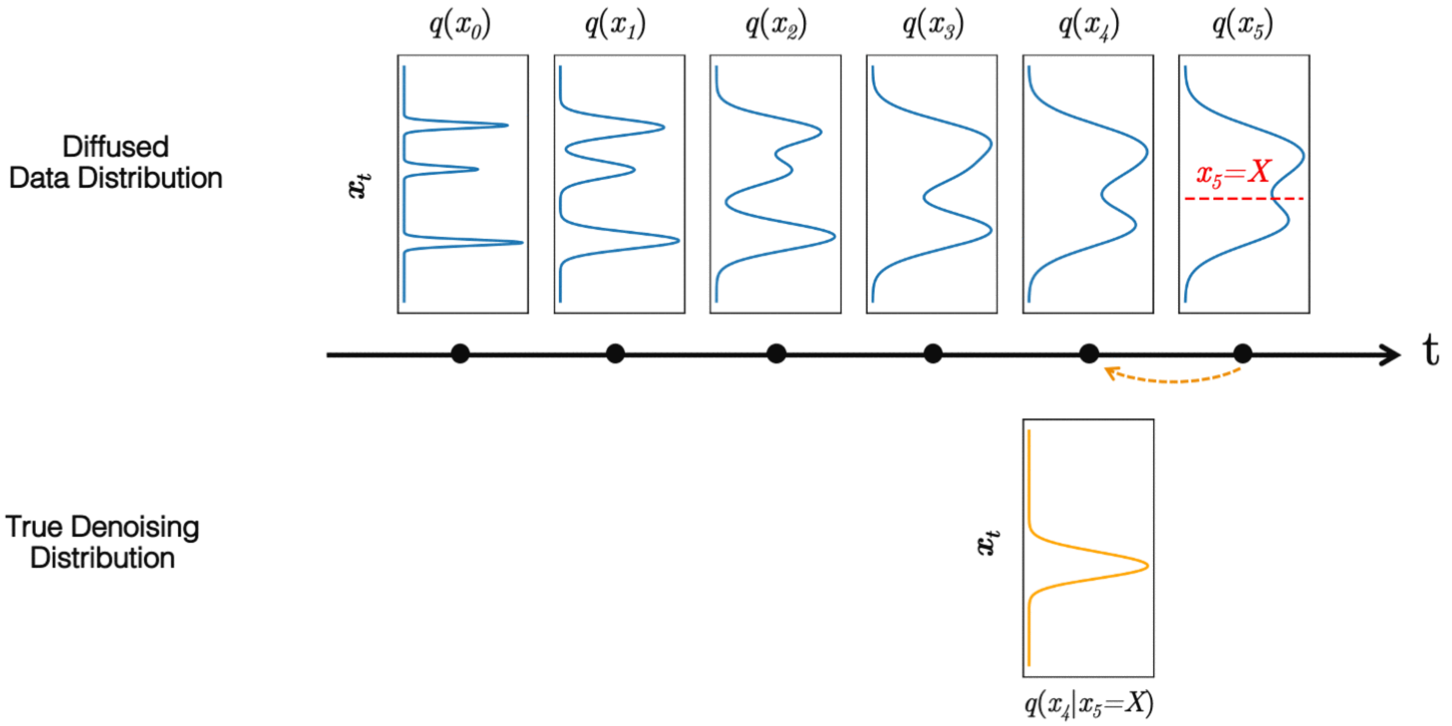
- Iteratively sample x_{t-1} for T timesteps:

$$x_{t-1} \sim \underbrace{q(x_{t-1}|x_t)}_{\text{True denoising distribution}}$$

True denoising distribution → This distribution is unknown. Can we estimate it?



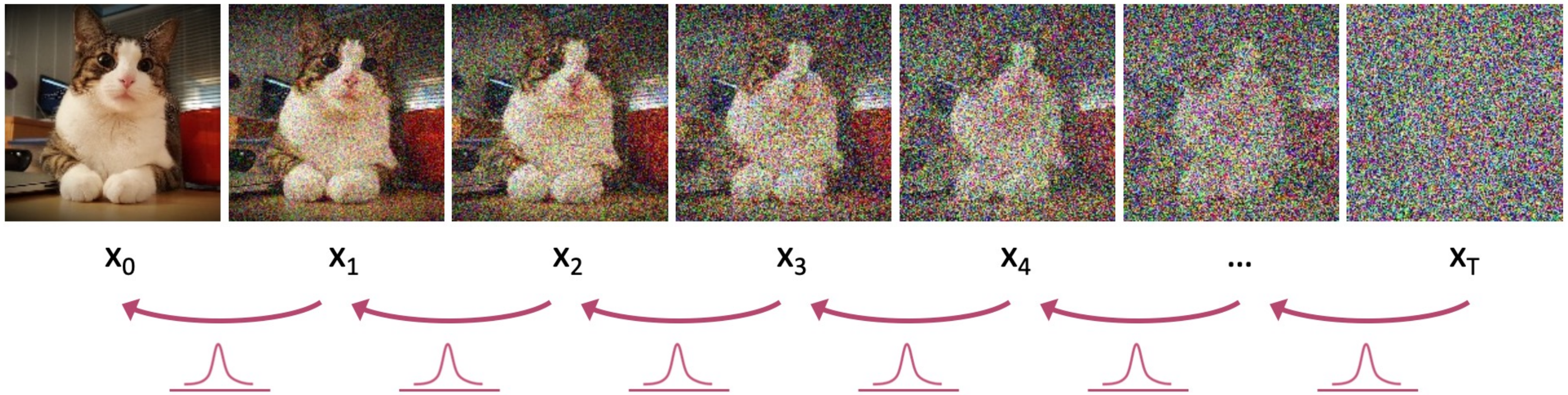
The Reverse Diffusion Process



YES! We can approximate the true denoising distribution (as a normal distribution) for small steps.

The Reverse Diffusion Process

Data



Noise

We approximate the true denoising distribution $q(x_{t-1}|x_t)$ as being normal distributed:

$$p_\theta(x_{t-1}|x_t) = N(\underbrace{\mu_\theta(x_t, t)}_{\text{Mean is estimated by a neural network}}, \sigma_t^2 \boldsymbol{I})$$

How can we train such a model?

Keep in mind: $p_\theta(x_{t-1}|x_t) = N(\mu_\theta(x_t,t), \sigma_t^2 \boldsymbol{I})$

Ho et al. (2020) found that we can parameterize $\mu_\theta(x_t,t)$ as follows:

$$\mu_\theta(x_t,t) = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}} \underbrace{\epsilon_\theta(x_t,t)}_{\text{the noise to be removed}}\right)$$

with $\alpha_t \coloneqq 1 - \beta_t$ and $\bar{\alpha}_t \coloneqq \prod_{s=1}^{t} \alpha_s$.

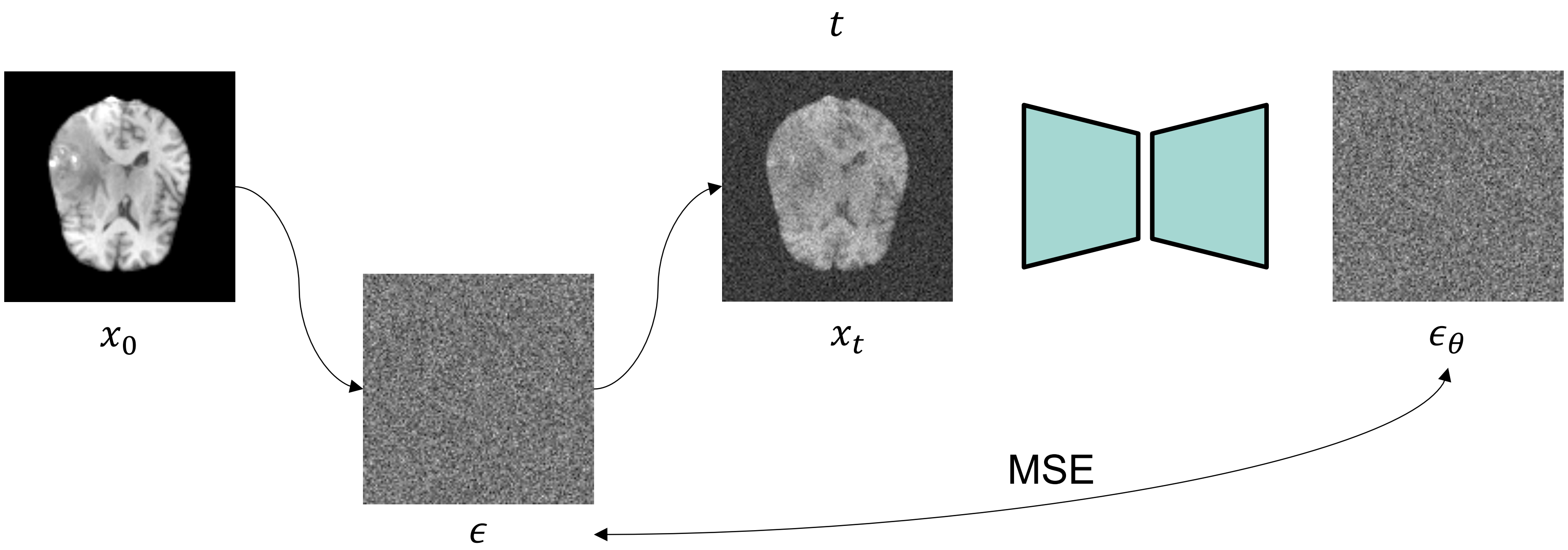
Algorithm 1 Training

1: **repeat**
2: $\mathbf{x}_0 \sim q(\mathbf{x}_0)$
3: $t \sim \mathrm{Uniform}(\{1,\ldots,T\})$
4: $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0},\mathbf{I})$
5: Take gradient descent step on
$$\nabla_\theta \left\| \boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, t) \right\|^2$$
6: **until** converged

Algorithm 2 Sampling

1: $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0},\mathbf{I})$
2: **for** $t = T,\ldots,1$ **do**
3: $\mathbf{z} \sim \mathcal{N}(\mathbf{0},\mathbf{I})$ if $t > 1$, else $\mathbf{z} = \mathbf{0}$
4: $\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t,t)\right) + \sigma_t \mathbf{z}$
5: **end for**
6: **return** $\mathbf{x}_0$

How can we train such a model?



Algorithm 1 Training

repeat
  $\mathbf{x}_0 \sim q(\mathbf{x}_0)$
  $t \sim \mathrm{Uniform}(\{1, \ldots, T\})$
  $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
  Take gradient descent step on
  $$\nabla_\theta \left\| \boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, t) \right\|^2$$
until converged

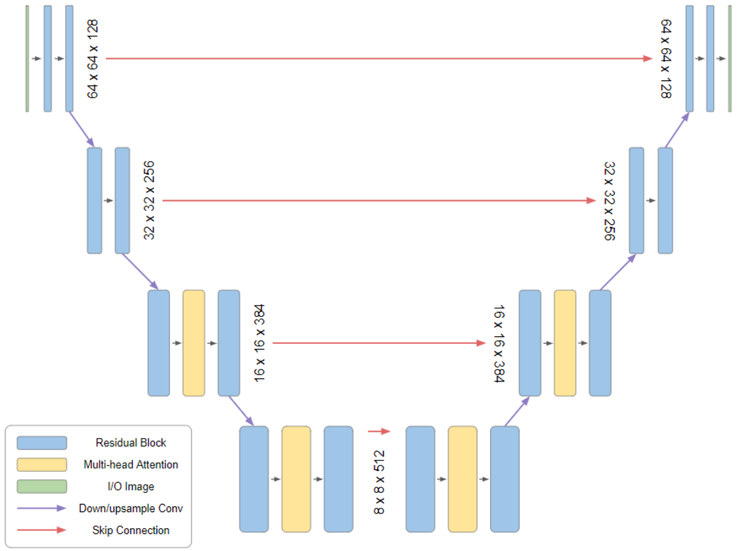
Algorithm 2 Sampling

$\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
for $t = T, \ldots, 1$ **do**
  $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ if $t > 1$, else $\mathbf{z} = \mathbf{0}$
  $\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)\right) + \sigma_t \mathbf{z}$
end for
return $\mathbf{x}_0$

Implementation Details

Keep in mind: We want to predict the noise to be removed from a corrupted image.

$\epsilon_\theta(x_t, t)$ is usually implemented as a U-Net:

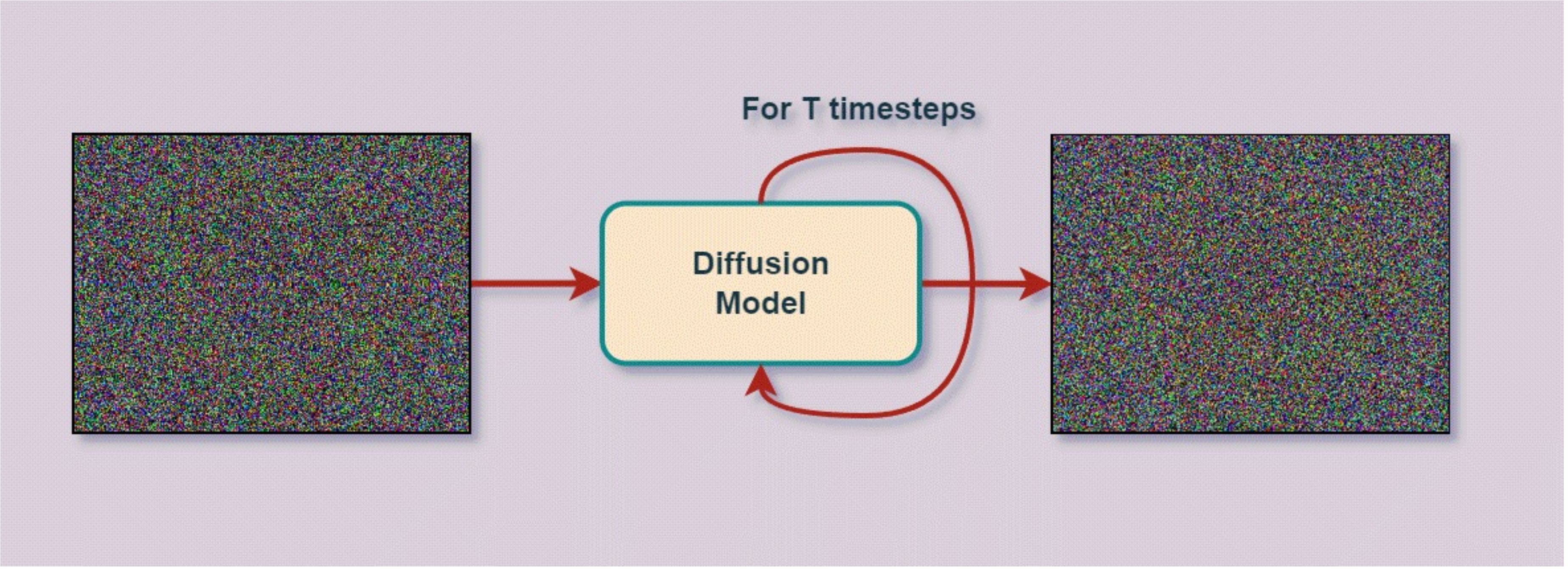


BUT, we could also use:

- Transformers
- VQ-VAEs
- …

This remains a design choice.

Generating Samples



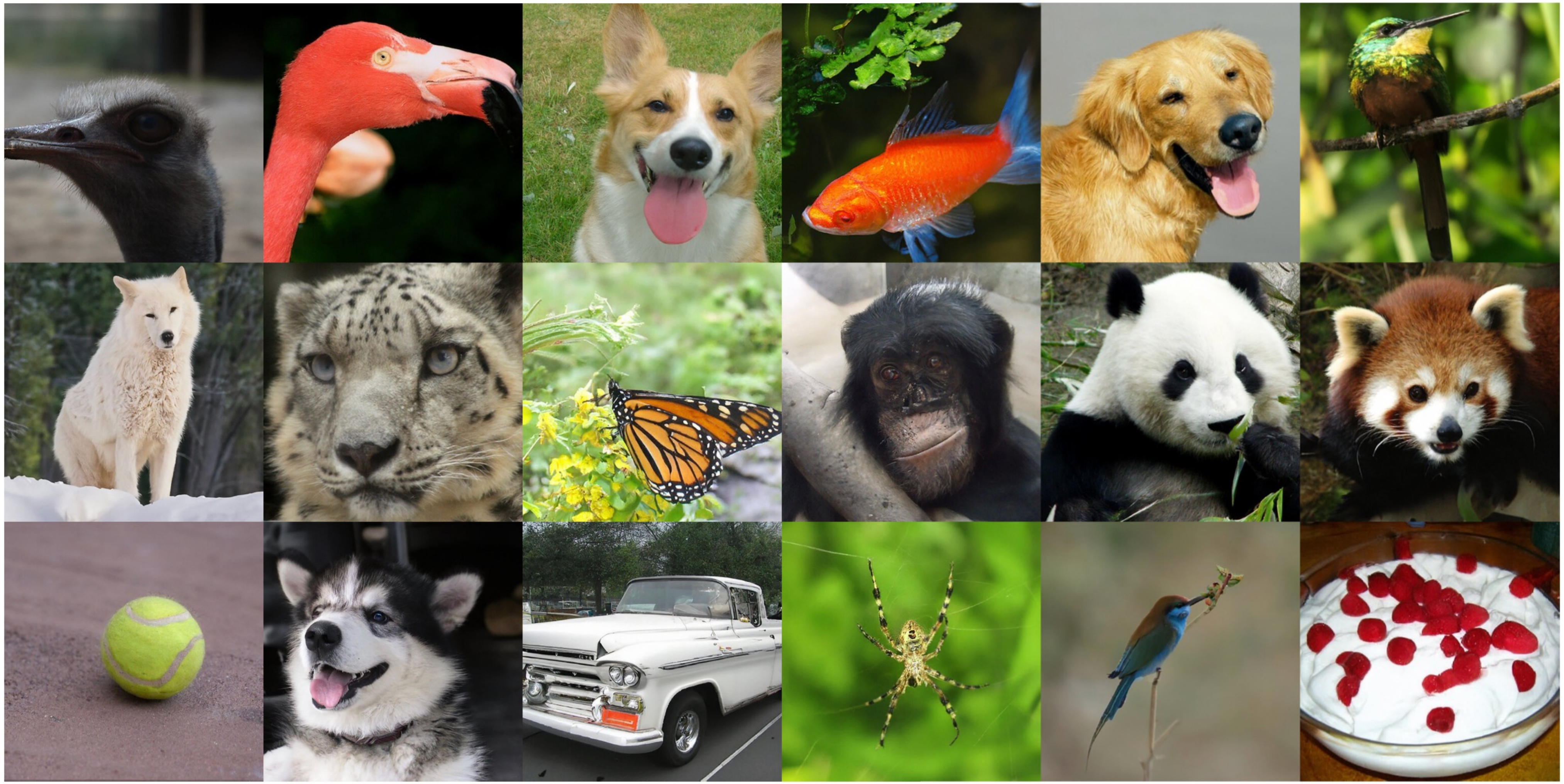
Algorithm 1 Training

1: **repeat**
2: $\mathbf{x}_0 \sim q(\mathbf{x}_0)$
3: $t \sim \mathrm{Uniform}(\{1, \dots, T\})$
4: $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
5: Take gradient descent step on
$$\nabla_\theta \left\| \boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, t) \right\|^2$$
6: **until** converged

Algorithm 2 Sampling

1: $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
2: **for** $t = T, \dots, 1$ **do**
3: $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ if $t > 1$, else $\mathbf{z} = \mathbf{0}$
4: $\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)\right) + \sigma_t \mathbf{z}$
5: **end for**
6: **return** $\mathbf{x}_0$

Sample Quality



Samples from model trained on ImageNet (512 x 512)

Pros & Cons

Samples from model trained on CelebA-HQ (256 x 256)

Pros:

+ High sample quality & diversity
+ Build on a strong theoretical foundation
+ Easy and stable to train (just a simple MSE Loss + just one network)

Cons:

– Very slow (sampling usually requires multiple model evaluations)
  → we will see some strategies to speed this process up

Playing Around with Diffusion Models

A Score-Based View on Diffusion Models

Generative Modeling by Estimating Gradients of the Data Distribution

Yang Song
Stanford University
yangsong@cs.stanford.edu

Stefano Ermon
Stanford University
ermon@cs.stanford.edu

NeurIPS (2019)

SCORE-BASED GENERATIVE MODELING THROUGH STOCHASTIC DIFFERENTIAL EQUATIONS

Yang Song*
Stanford University
yangsong@cs.stanford.edu

Jascha Sohl-Dickstein
Google Brain
jaschasd@google.com

Diederik P. Kingma
Google Brain
durk@google.com

Abhishek Kumar
Google Brain
abhishk@google.com

Stefano Ermon
Stanford University
ermon@cs.stanford.edu

Ben Poole
Google Brain
pooleb@google.com

ICLR (2021)

Maximum Likelihood Training of Score-Based Diffusion Models

Yang Song*
Computer Science Department
Stanford University
yangsong@cs.stanford.edu

Conor Durkan*
School of Informatics
University of Edinburgh
conor.durkan@ed.ac.uk

Iain Murray
School of Informatics
University of Edinburgh
i.murray@ed.ac.uk

Stefano Ermon
Computer Science Department
Stanford University
ermon@cs.stanford.edu

NeurIPS (2021)

What do we mean when talking about a score function?

$p(\boldsymbol{x})$

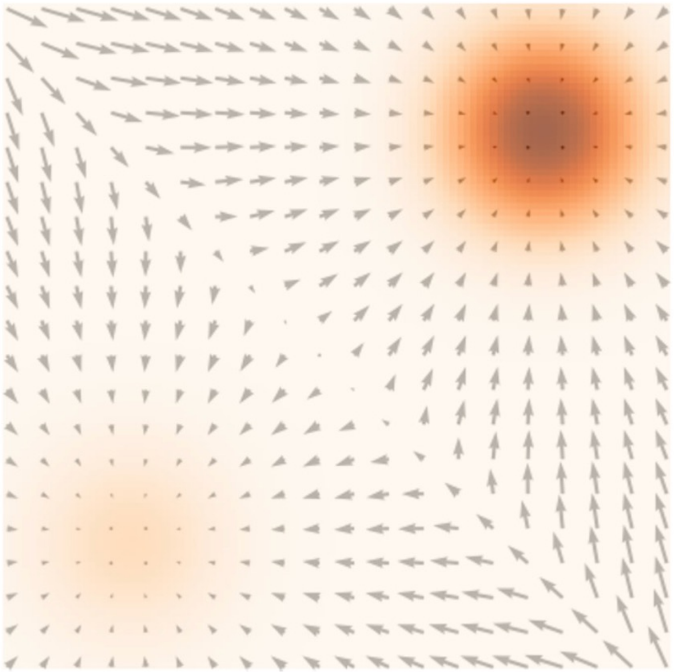
Probability density function

Take derivative ↓ ↑ Solve integral

$\boxed{\nabla_{\boldsymbol{x}} \log p(\boldsymbol{x})}$

Score function

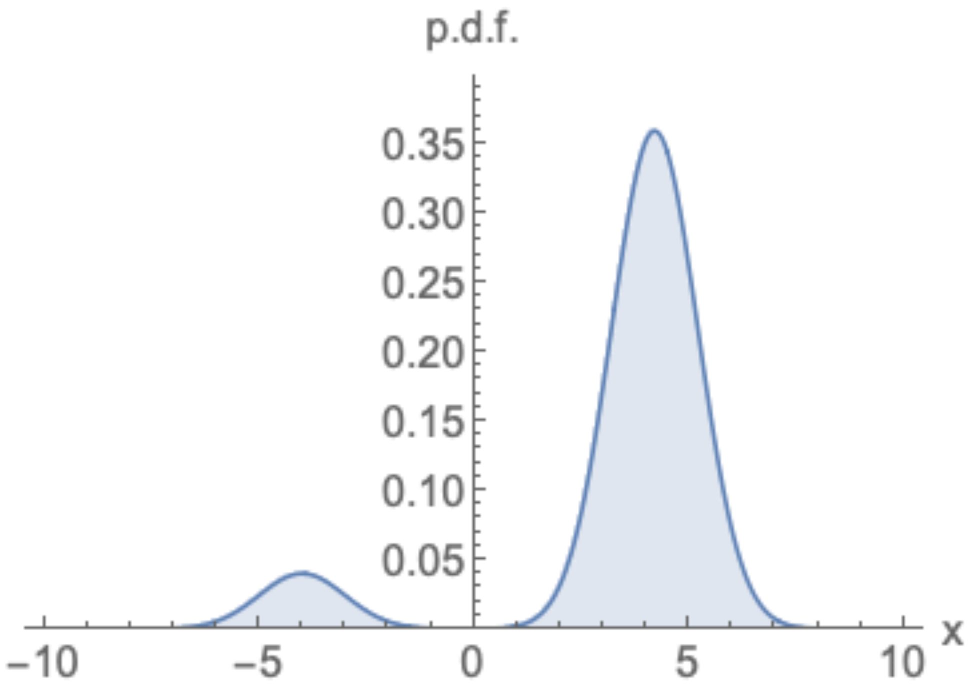
Score function represents the probability distribution



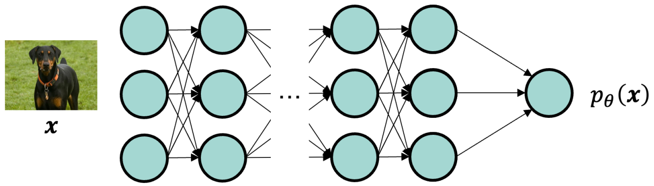
Probability density function (color coded)
Score function (vector field)

The score function **preserves all information** of the density function, but is much **easier to handle** → **why?**

Score functions bypass the normalizing constant

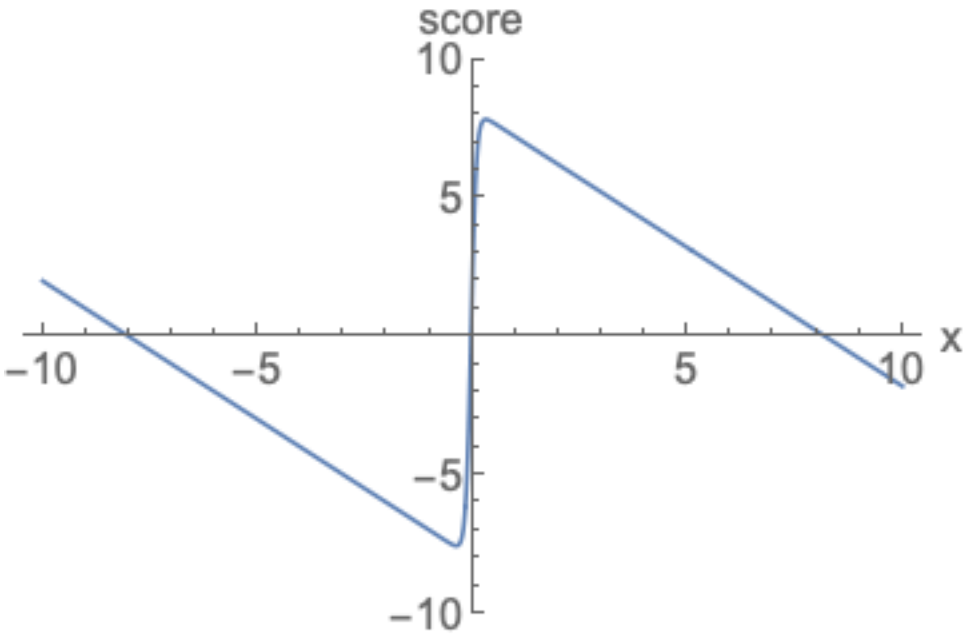


Probability density function



$$\frac{e^{f_\theta(\boldsymbol{x})}}{\cancel{Z_\theta}} = p_\theta(\boldsymbol{x})$$

We always need to ensure normalization.



Score function

$$\nabla_{\boldsymbol{x}} \log p_\theta(\boldsymbol{x}) = \nabla_{\boldsymbol{x}} f_\theta(\boldsymbol{x}) - \boxed{\nabla_{\boldsymbol{x}} \log Z_\theta}$$

$$= \nabla_{\boldsymbol{x}} f_\theta(\boldsymbol{x}) - 0$$

Score function doesn't rely on normalizing constant.

Can we estimate such a score function from data?

We know it's possible to **train a properly normalized statistical model to estimate the data density function using maximum likelihood** → can we do something similar to estimate a score function/model?

Given (Data): $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_N\} \sim p_{data}(\boldsymbol{x})$

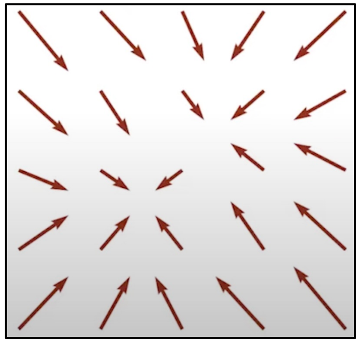
Goal (Score function): $\nabla_{\boldsymbol{x}} \log p_{data}(\boldsymbol{x})$

We can't compute this as we don't know $p_{data}(x)$

Score model: $s_\theta(\boldsymbol{x}) \colon \mathbb{R}^d \to \mathbb{R}^d \approx \nabla_{\boldsymbol{x}} \log p_{data}(\boldsymbol{x})$

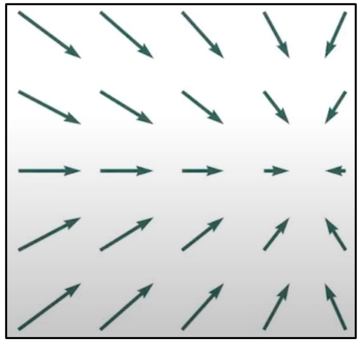
Objective: $\mathbb{E}_{p_{data}(x)}[\|\nabla_x \log p_{data}(x) - s_\theta(x)\|_2^2]$ (Fisher divergence to compare the vector fields)

$\nabla_{\boldsymbol{x}} \log p_{data}(\boldsymbol{x})$



?

$s_\theta(\boldsymbol{x})$



Score matching

There exists a so-called **score matching objective**, that is similar to the Fisher divergence (up to a constant):

$$\mathbb{E}_{p_{data(x)}}\left[\frac{1}{2}\|s_\theta(x)\|_2^2 + trace(\nabla_x s_\theta(x))\right]$$

doesn’t rely on the ground truth score

As a constant doesn’t matter for optimization, this score matching objective defines the same optimum as the Fisher divergence and can **effectively be estimated by the empirical mean over the training data set**:

$$\approx \frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{2}\|s_\theta(x_i)\|_2^2 + trace(\nabla_x s_\theta(x_i))\right]$$

Journal of Machine Learning Research 6 (2005) 695–709 Submitted 11/04; Revised 3/05; Published 4/05

Estimation of Non-Normalized Statistical Models by Score Matching

Aapo Hyvärinen AAPO.HYVARINEN@HELSINKI.FI
Helsinki Institute for Information Technology (BRU)
Department of Computer Science
FIN-00014 University of Helsinki, Finland

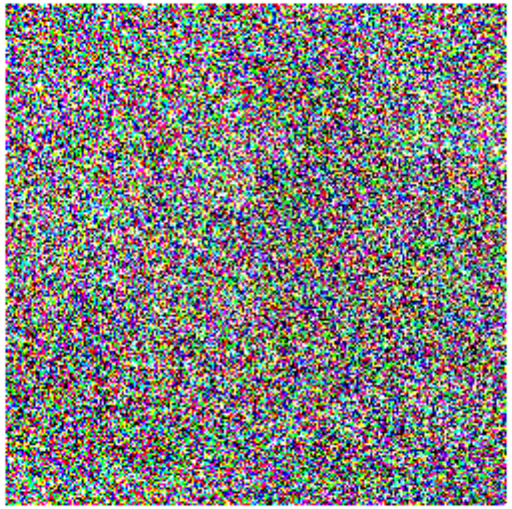
https://andrewcharlesjones.github.io/journal/21-score-matching.html

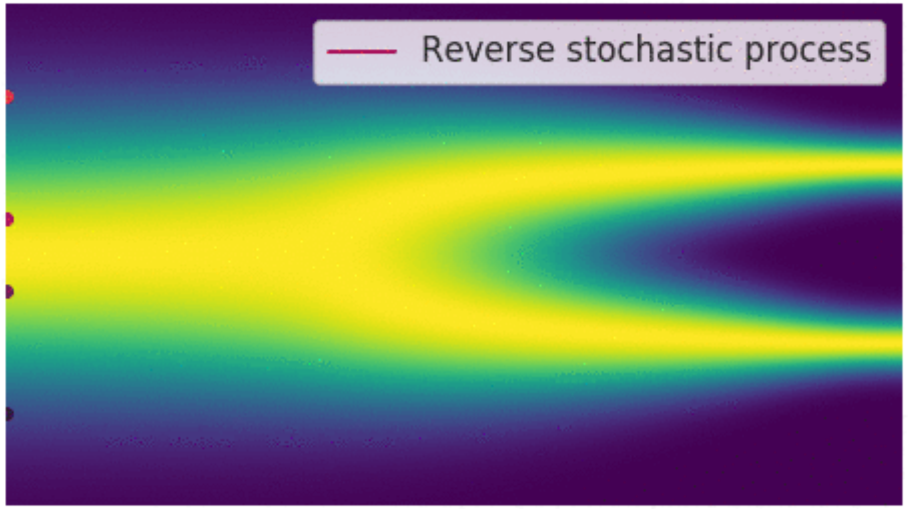
Modelling the Diffusion Process with SDEs

Forward SDE (data → noise)

$$\mathbf{x}(0) \longrightarrow \mathrm{d}\mathbf{x} = \underbrace{\mathbf{f}(\mathbf{x}, t)\mathrm{d}t}_{\text{Drift coefficient}} + \underbrace{g(t)\mathrm{d}\mathbf{w}}_{\text{Diffusion coefficient}} \longrightarrow \mathbf{x}(T)$$

Modelling the Diffusion Process with SDEs





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

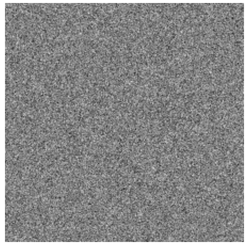
If we can model this score function, we can solve the reverse SDE using Euler, Milstein or Runge-Kutta method.

https://yang-song.net/blog/2021/score/ (More information on this very nice blog post)

Agenda

- ❑ Introduction to Diffusion Models [~30 min]
  - ❑ Physical Intuition & General Concepts
  - ❑ Denoising Diffusion Probabilistic Models
  - ❑ A Score-Based View on Diffusion Models
- ❑ Advanced Topics [~30 min]
  - ❑ Sampling Strategies
  - ❑ Inference-time Conditioning
  - ❑ Training-time Conditioning
- ❑ Applications in Medical Imaging [~30 min]
  - ❑ Synthesis
  - ❑ Inpainting
  - ❑ Segmentation
  - ❑ Anomaly Detection
  - ❑ Reconstruction
  - ❑ Registration

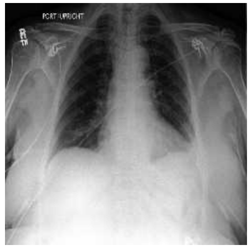
Fake Image Generation



$x_T \sim N(0, \mathbf{I})$

$\rightarrow x_{T-1} \rightarrow x_{T-2} \rightarrow \ldots \rightarrow$

synthetic image

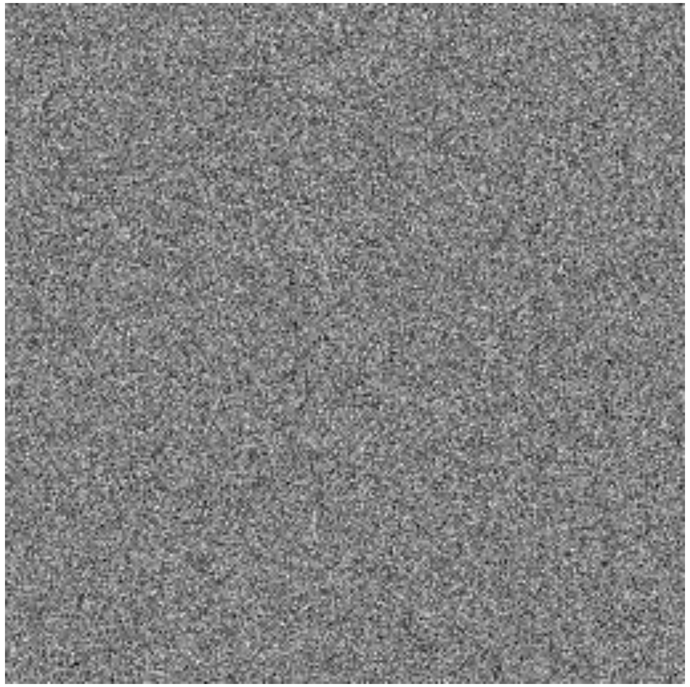


x_0

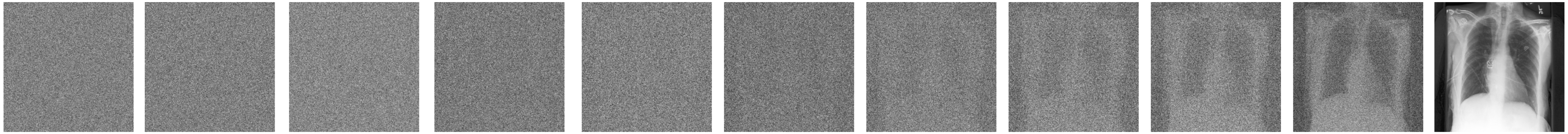
U-Net

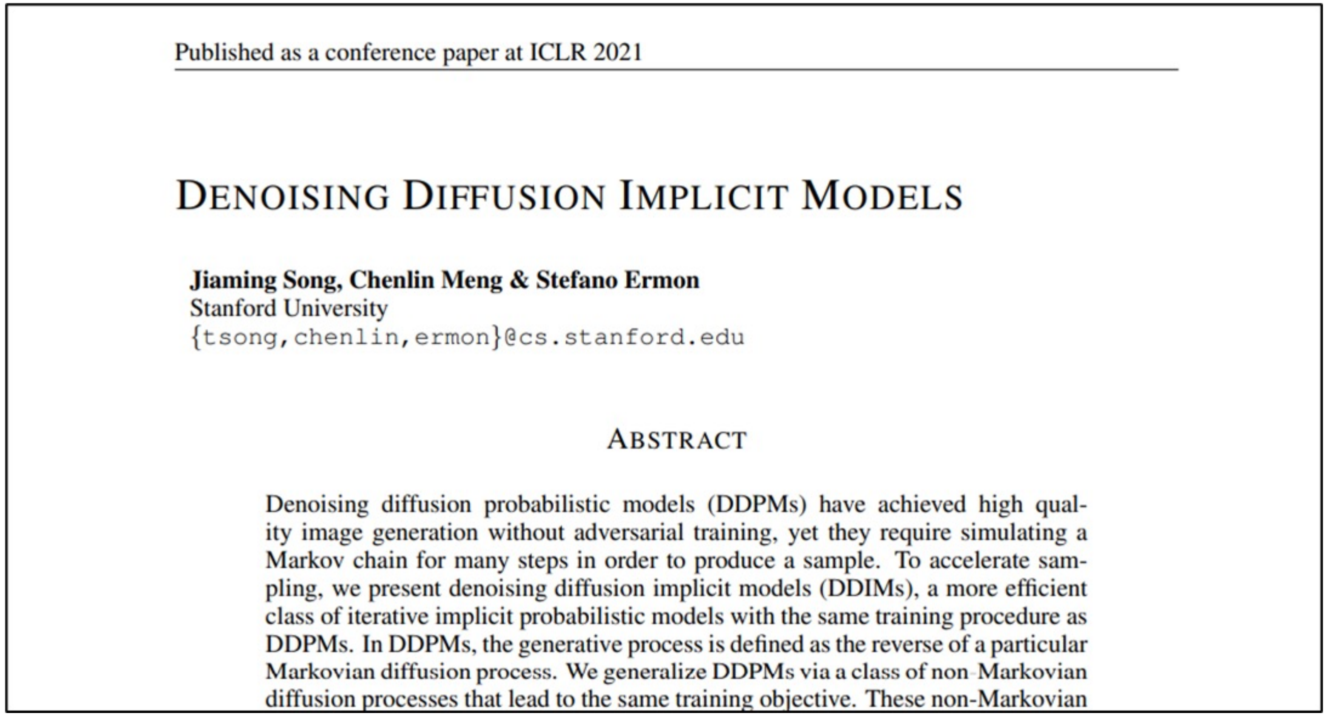
$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}} \epsilon_\theta(x_t, t) \right) + \sigma_t \epsilon, \quad \text{with } \epsilon \sim \mathcal{N}(0, \mathbf{I}).$$

Random component



Schedulers: How to Accelerate Sampling?




Published as a conference paper at ICLR 2021

DENOISING DIFFUSION IMPLICIT MODELS

Jiaming Song, Chenlin Meng & Stefano Ermon
Stanford University
{tsong,chenlin,ermon}@cs.stanford.edu

ABSTRACT

Denoising diffusion probabilistic models (DDPMs) have achieved high quality image generation without adversarial training, yet they require simulating a Markov chain for many steps in order to produce a sample. To accelerate sampling, we present denoising diffusion implicit models (DDIMs), a more efficient class of iterative implicit probabilistic models with the same training procedure as DDPMs. In DDPMs, the generative process is defined as the reverse of a particular Markovian diffusion process. We generalize DDPMs via a class of non-Markovian diffusion processes that lead to the same training objective. These non-Markovian

"Denoising diffusion probabilistic models (DDPMs) have achieved high quality image generation, yet they require simulating a Markov chain for many steps in order to produce a sample."

We need to make the generation process faster.

From DDPMs to DDIMs

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

DDPM sampling scheme $\quad \sigma_t = \sqrt{(1-\alpha_{t-1})/(1-\alpha_t)}\sqrt{1-\alpha_t/\alpha_{t-1}}$

DDIM sampling scheme $\quad \sigma_t = 0$ $\Rightarrow$ We remove the random component

The training process stays the same.

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

An Excursion into ODEs

- The connection to ordinary differential equations (ODEs) can be seen when we rewrite the DDIM denoising step as

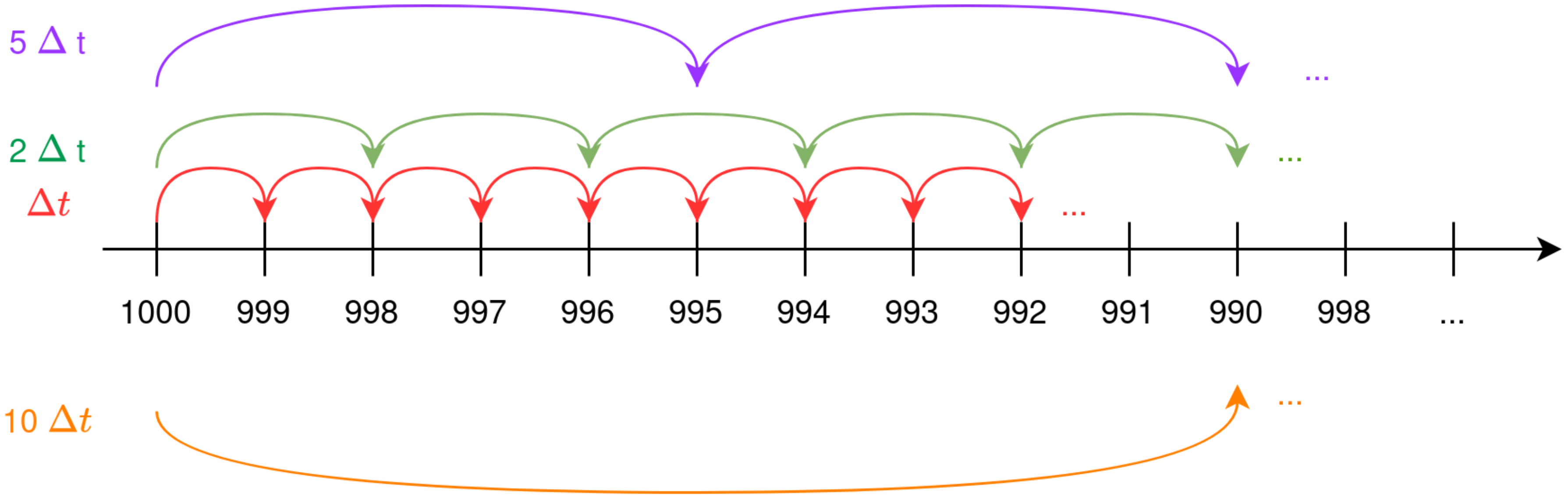
$$\underbrace{\frac{x_{t-1}}{\sqrt{\bar{\alpha}_{t-1}}}}_{\text{prediction}} = \underbrace{\frac{x_t}{\sqrt{\bar{\alpha}_t}}}_{\text{previous value}} + \underbrace{\left(\sqrt{\frac{1-\bar{\alpha}_{t-1}}{\bar{\alpha}_{t-1}}} - \sqrt{\frac{1-\bar{\alpha}_t}{\bar{\alpha}_t}}\right)}_{\text{step size}} \underbrace{\epsilon_\theta(x_t, t)}_{\text{derivative}}.$$

- This can be interpreted as the Euler approximation of an ODE.
- DDIM is a **probability flow** ODE from a SDE [1].
- We can speed up the generation process by choosing a larger step size.

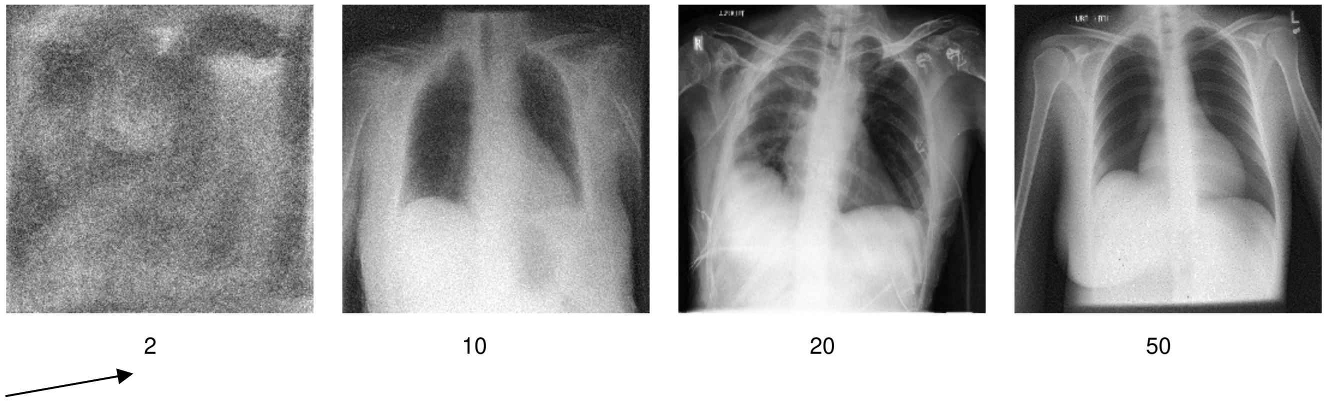
→ Faster, but less accurate

[1] Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., & Poole, B. (2020). Score-based generative modeling through stochastic differential equations. arXiv preprint arXiv:2011.13456.

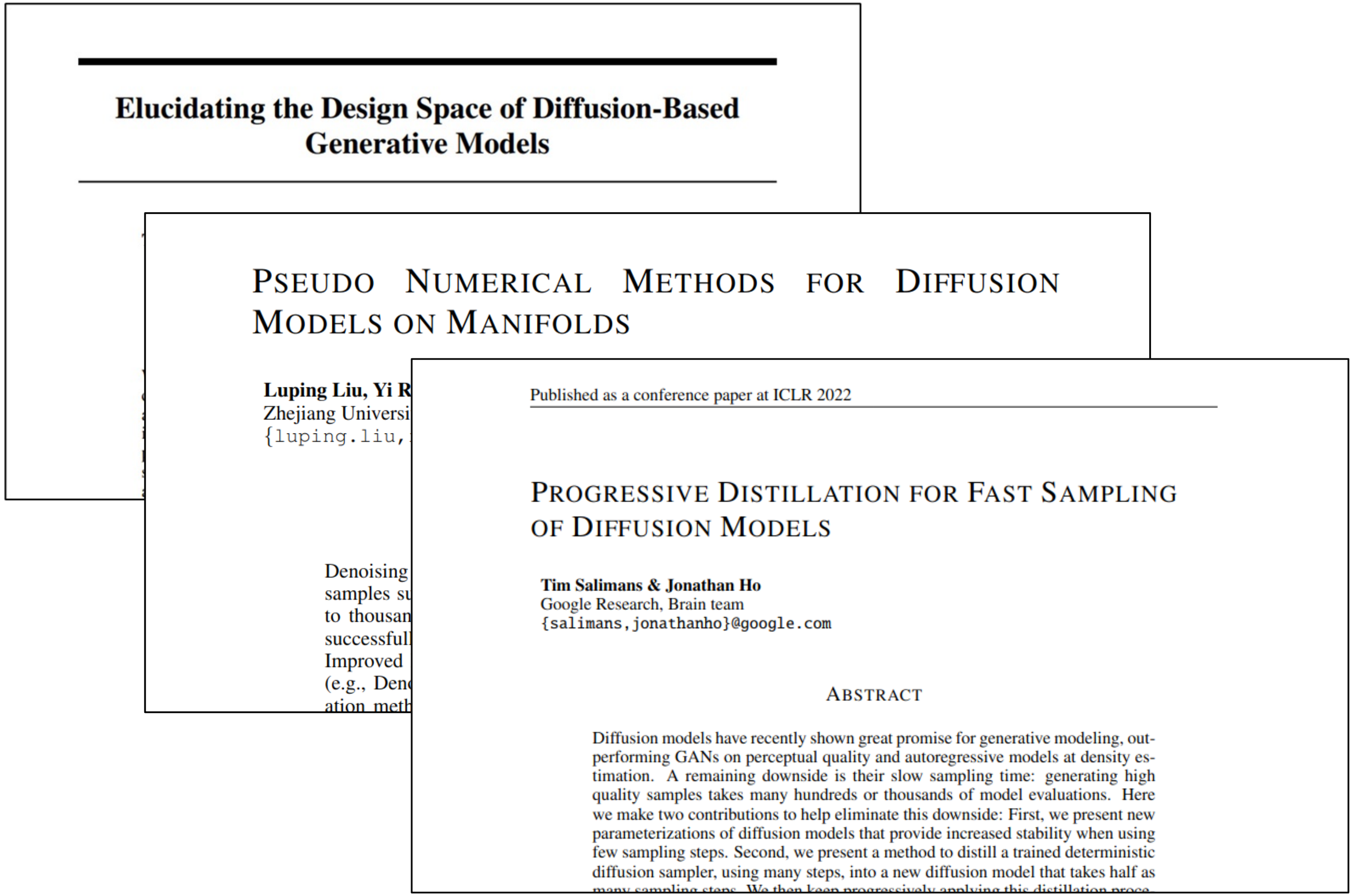
DDIM Accelerated Sampling



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



Various Schedulers…


Elucidating the Design Space of Diffusion-Based Generative Models

PSEUDO NUMERICAL METHODS FOR DIFFUSION MODELS ON MANIFOLDS

Luping Liu, Yi R
Zhejiang Universi
{luping.liu,

Denoising
samples su
to thousan
successfull
Improved
(e.g., Den
ation meth

Published as a conference paper at ICLR 2022

PROGRESSIVE DISTILLATION FOR FAST SAMPLING OF DIFFUSION MODELS

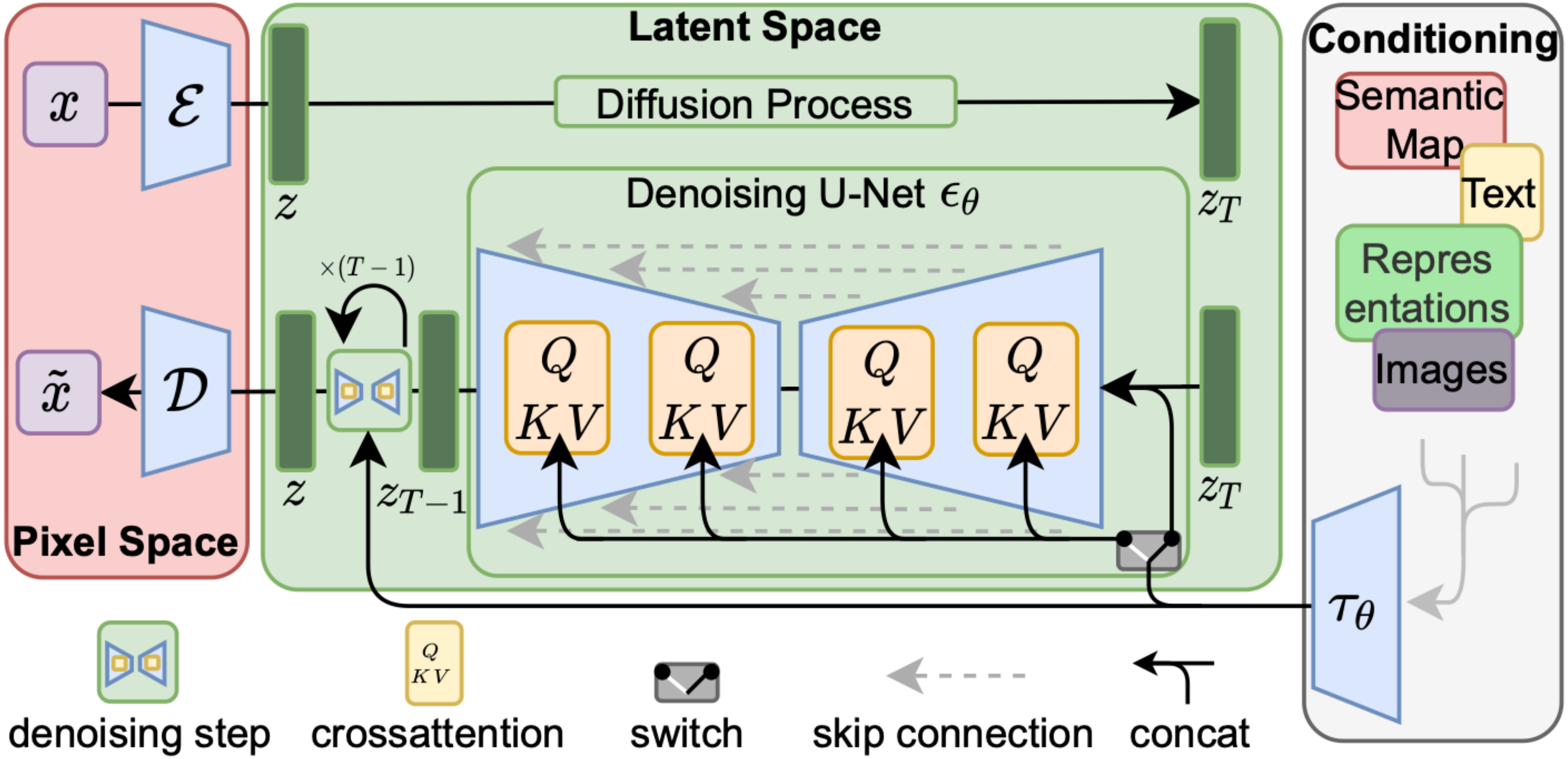
Tim Salimans & Jonathan Ho
Google Research, Brain team
{salimans,jonathanho}@google.com

ABSTRACT

Diffusion models have recently shown great promise for generative modeling, outperforming GANs on perceptual quality and autoregressive models at density estimation. A remaining downside is their slow sampling time: generating high quality samples takes many hundreds or thousands of model evaluations. Here we make two contributions to help eliminate this downside: First, we present new parameterizations of diffusion models that provide increased stability when using few sampling steps. Second, we present a method to distill a trained deterministic diffusion sampler, using many steps, into a new diffusion model that takes half as

- Choosing a different solver for the given ODE can improve speed and image quality.
- Other numerical approaches such as **Heun's Method** or **Runge Kutta** solvers can be explored.
- Knowledge distillation techniques can be used for fast sampling.

Diffusion Models in the Latent Space



Rombach, Robin, et al. "High-resolution image synthesis with latent diffusion models." *CVPR 2022*

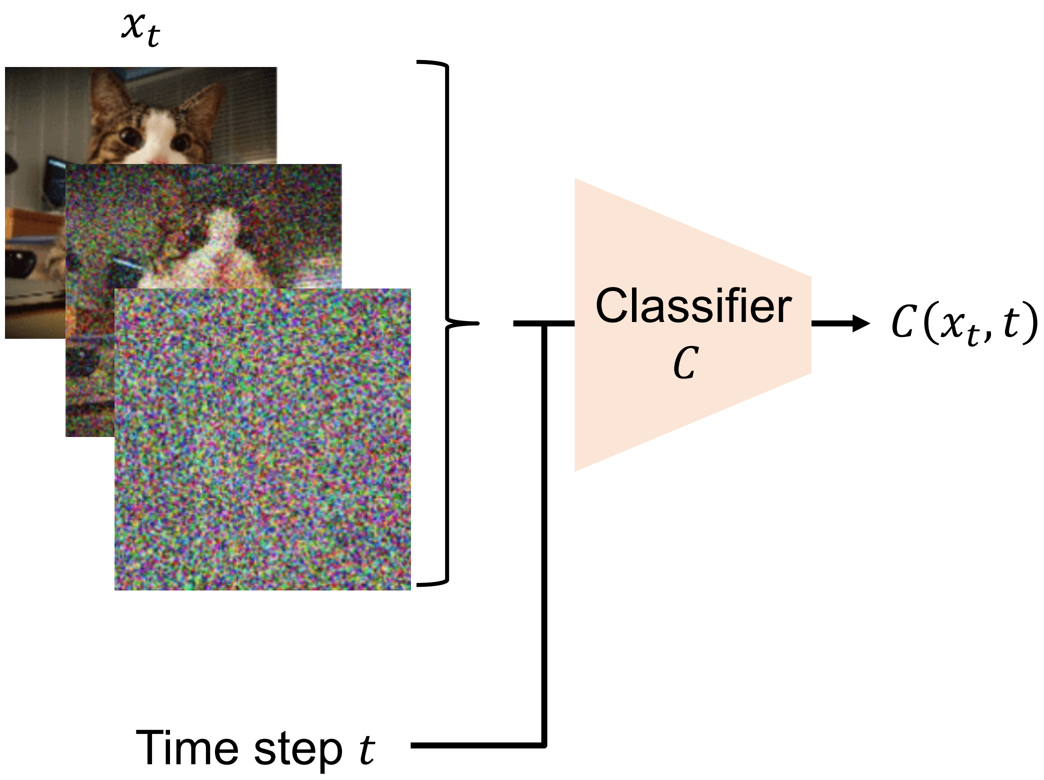
Agenda

- ❑ Introduction to Diffusion Models [~30 min]
  - ❑ Physical Intuition & General Concepts
  - ❑ Denoising Diffusion Probabilistic Models
  - ❑ A Score-Based View on Diffusion Models
- ❑ Advanced Topics [~30 min]
  - ❑ Sampling Strategies
  - ❑ Inference-time Conditioning
  - ❑ Training-time Conditioning
- ❑ Applications in Medical Imaging [~30 min]
  - ❑ Synthesis
  - ❑ Inpainting
  - ❑ Segmentation
  - ❑ Anomaly Detection
  - ❑ Reconstruction
  - ❑ Registration

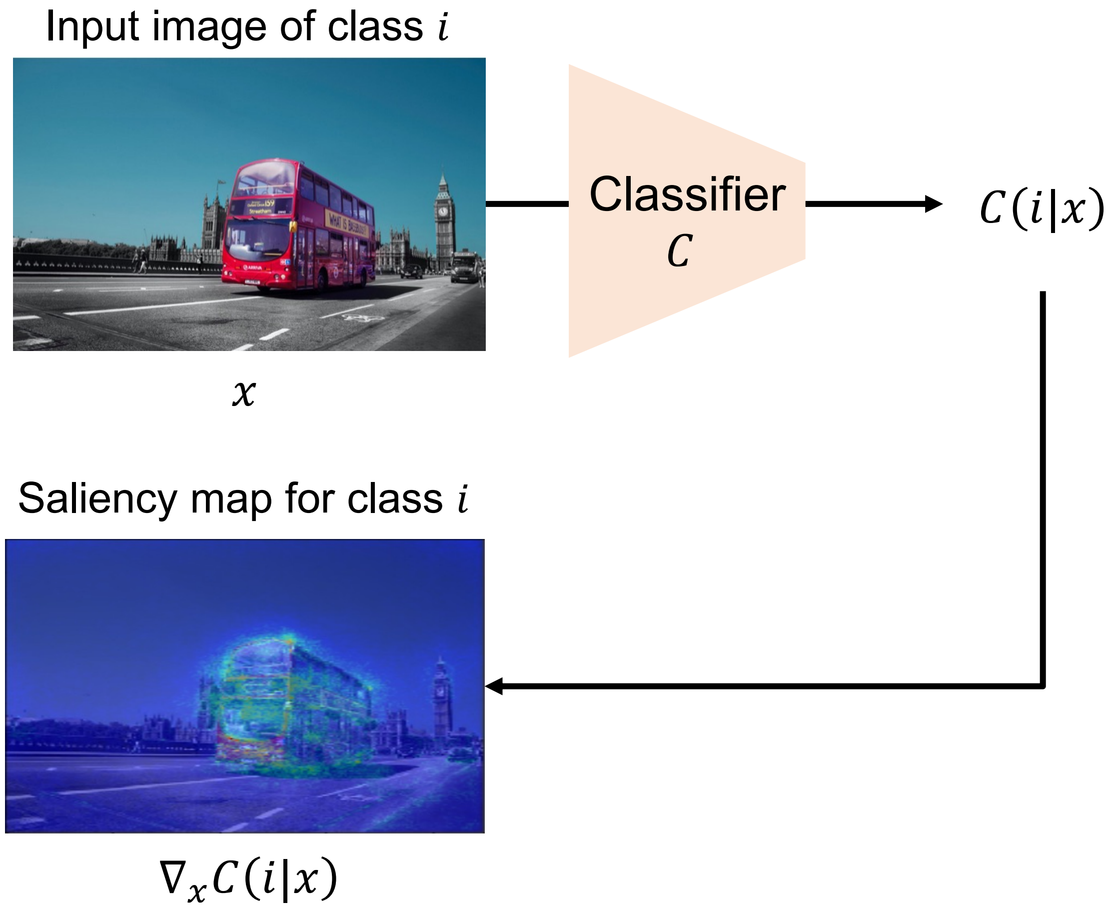
"mouse" → Diffusion Model →

Example: Classifier Guidance

We want a class-conditional diffusion model.



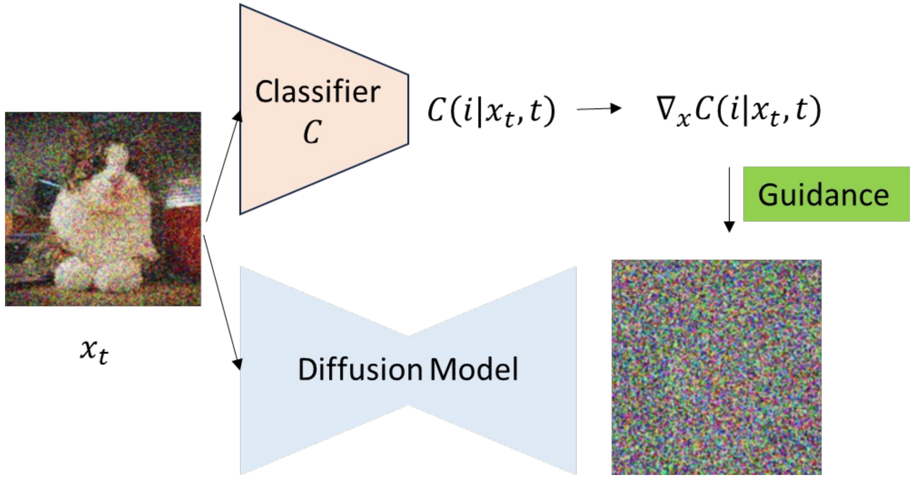
We consider the gradient with respect to the input pixels.

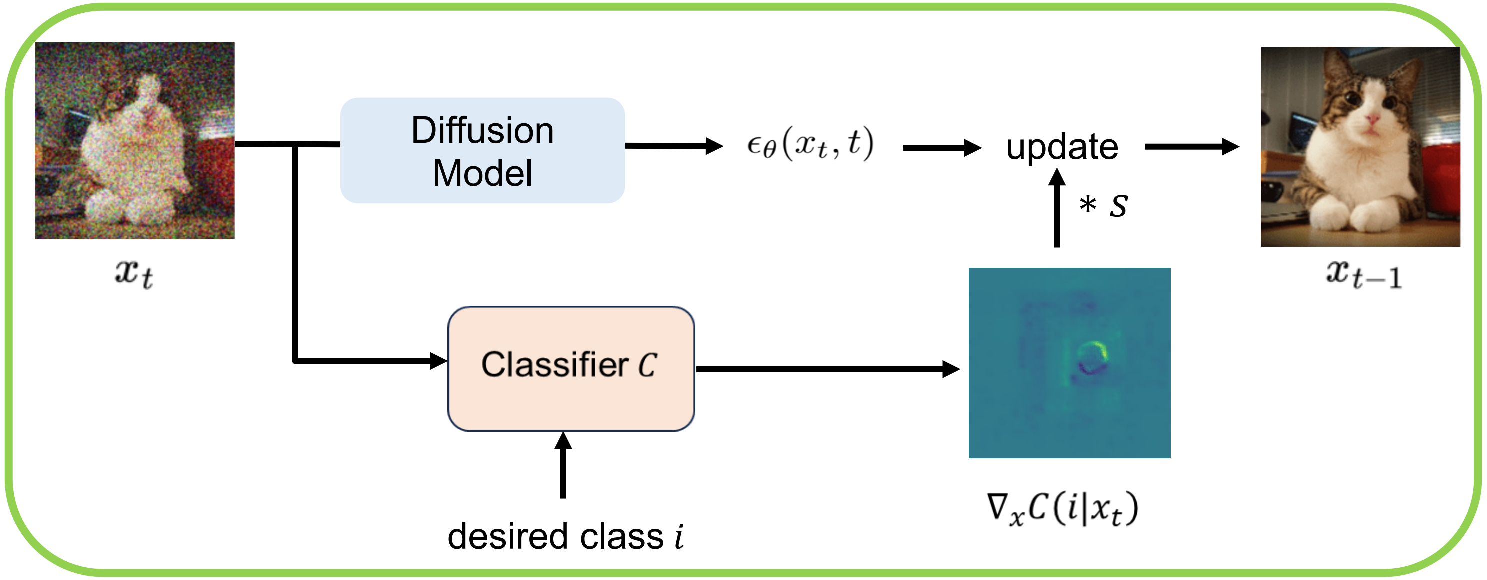


https://corochann.com/library-release-visualize-saliency-map-of-deep-neural-network-644/

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 Guidance

goldfish

arctic fox

butterfly

African elephant

flamingo

tennis ball

cheeseburger

fountain

balloon

tabby cat

lorikeet

agaric

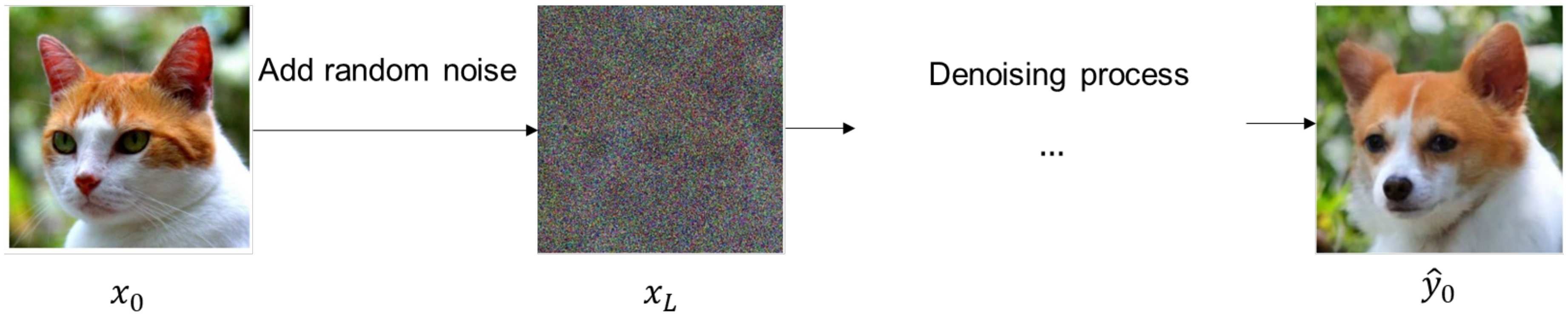
Check out this very nice tutorial: https://sander.ai/2022/05/26/guidance.html

---

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

Image-to-image translation

We might want to translate an image to another…



- We add L steps of noise to an input image x_0.
- The smaller L, the less the image can be changed.
- The higher L, the more information is destroyed.

We need to find a way to keep the information of x_0.

We consider Denoising Diffusion Implicit Models (DDIMs).

DDIM Inversion

- Under the DDIM sampling scheme, we remove the random component.
- The connection to ordinary differential equations (ODEs) can be seen when we rewrite the denoising step as

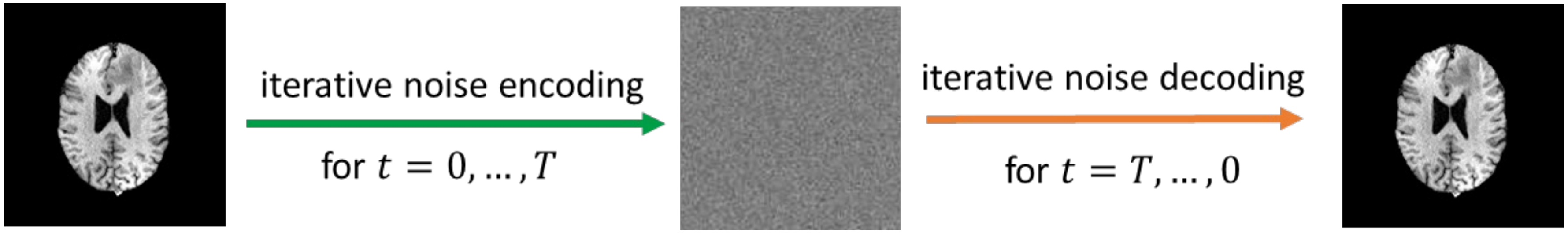
$$\frac{x_{t-1}}{\sqrt{\bar{\alpha}_{t-1}}} = \frac{x_t}{\sqrt{\bar{\alpha}_t}} + \left(\sqrt{\frac{1-\bar{\alpha}_{t-1}}{\bar{\alpha}_{t-1}}} - \sqrt{\frac{1-\bar{\alpha}_t}{\bar{\alpha}_t}}\right)\epsilon_\theta(x_t, t).$$

noise decoding

- This can be interpreted as the Euler approximation of an ODE.
- Given infinitely small steps t, the reversed ODE can then be solved with

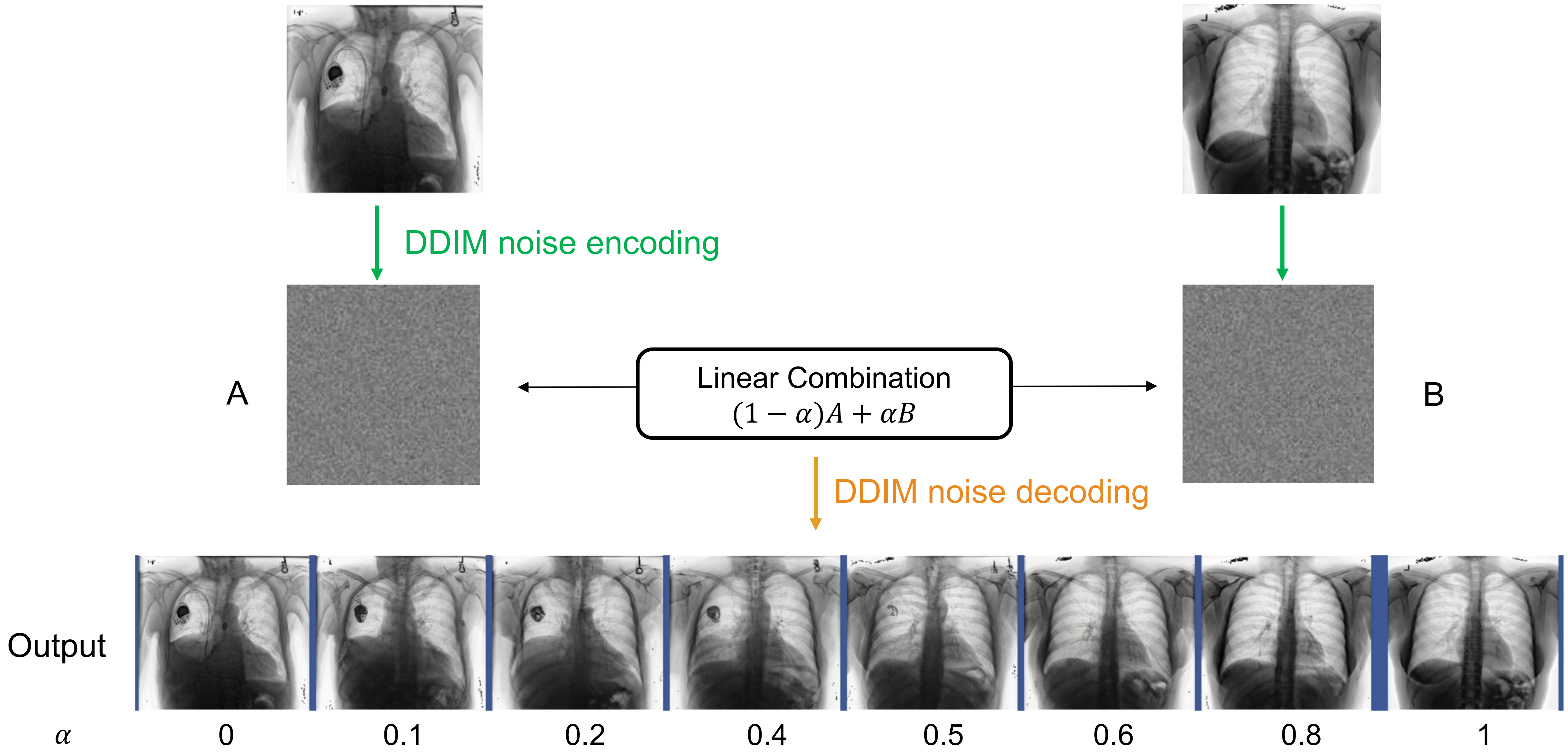
$$\frac{x_{t+1}}{\sqrt{\bar{\alpha}_{t+1}}} = \frac{x_t}{\sqrt{\bar{\alpha}_t}} + \left(\sqrt{\frac{1-\bar{\alpha}_{t+1}}{\bar{\alpha}_{t+1}}} - \sqrt{\frac{1-\bar{\alpha}_t}{\bar{\alpha}_t}}\right)\epsilon_\theta(x_t, t).$$

noise encoding

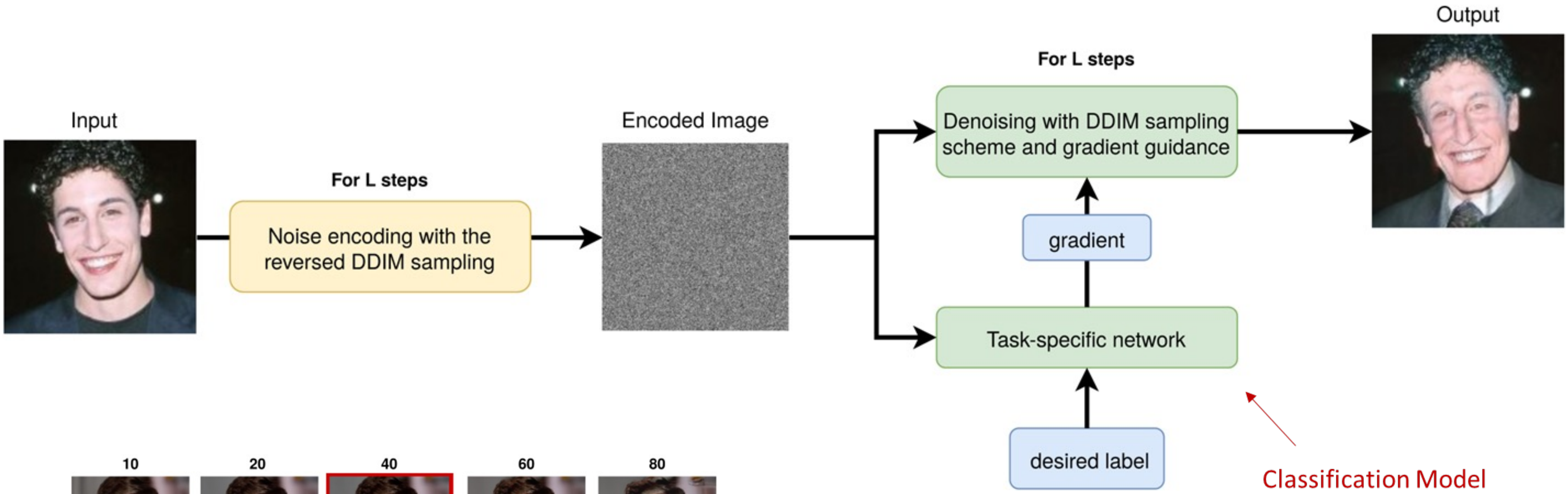


Song, Jiaming, Chenlin Meng, and Stefano Ermon. "Denoising diffusion implicit models." *arXiv preprint arXiv:2010.02502* (2020).

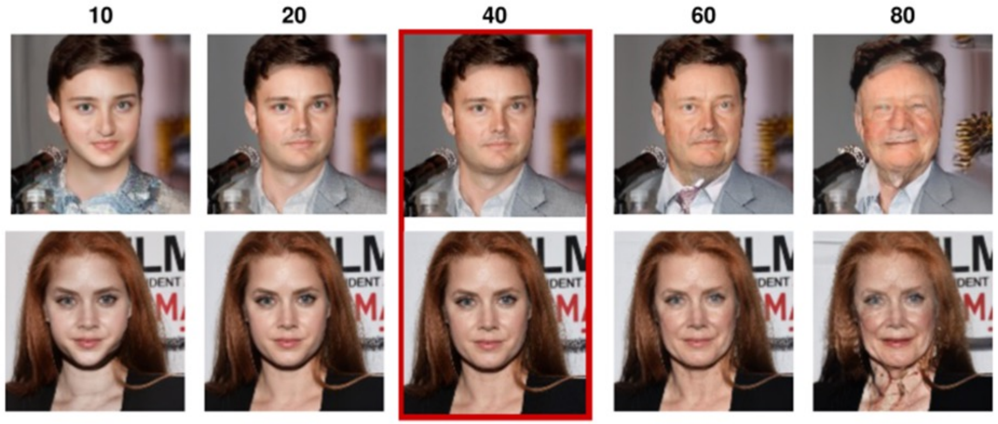
Image Interpolation



DDIM Inversion & Gradient Guidance



Classification Model
Regression Model
Segmentation Model



Example: age regression

Wolleb, Julia, et al. "The swiss army knife for image-to-image translation: Multi-task diffusion models." *arXiv preprint arXiv:2204.02641* (2022).

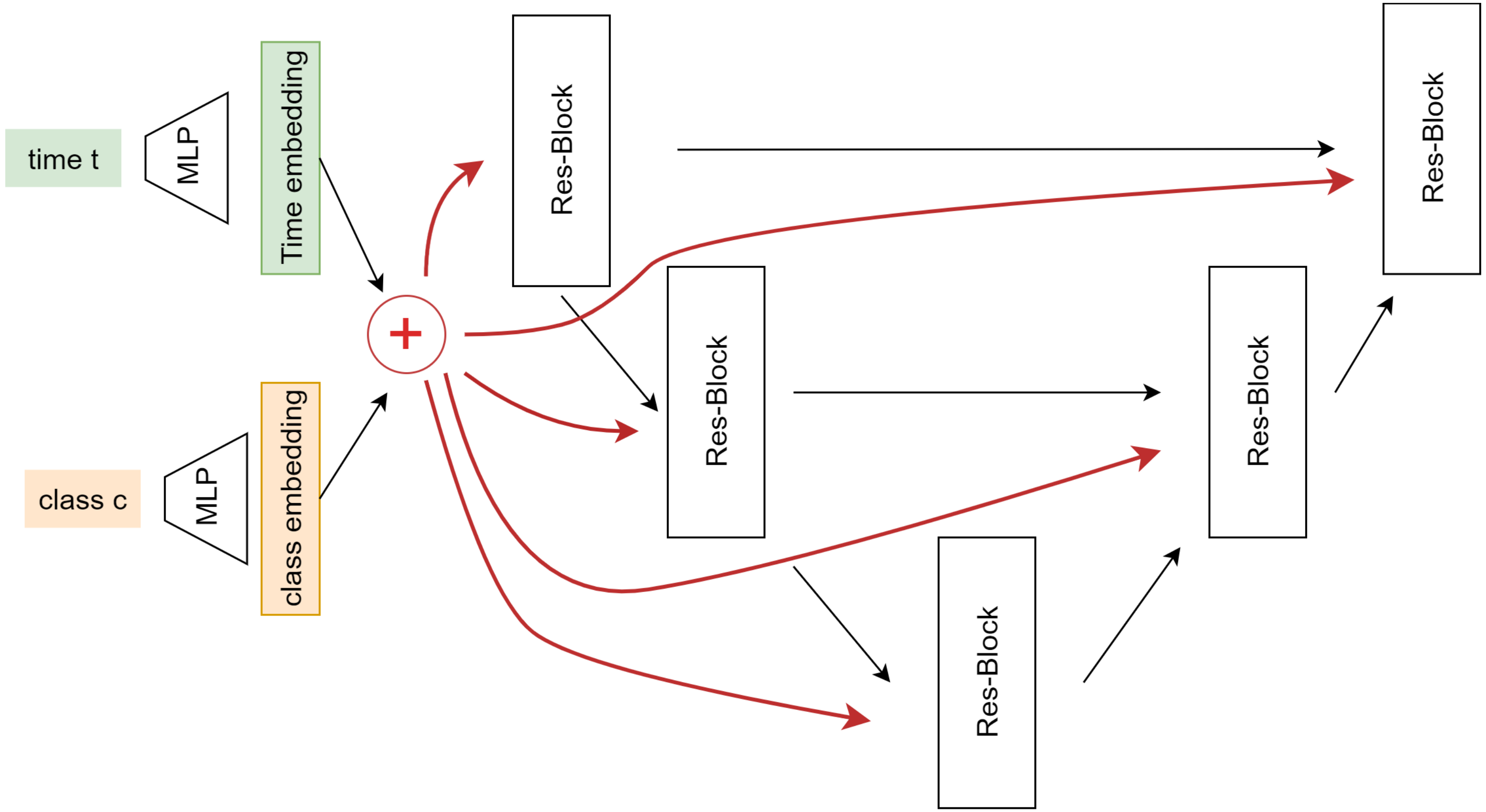
Agenda

- ❑ Introduction to Diffusion Models [~30 min]
  - ❑ Physical Intuition & General Concepts
  - ❑ Denoising Diffusion Probabilistic Models
  - ❑ A Score-Based View on Diffusion Models
- ❑ Advanced Topics [~30 min]
  - ❑ Sampling Strategies
  - ❑ Inference-time Conditioning
  - ❑ Training-time Conditioning
- ❑ Applications in Medical Imaging [~30 min]
  - ❑ Synthesis
  - ❑ Inpainting
  - ❑ Segmentation
  - ❑ Anomaly Detection
  - ❑ Reconstruction
  - ❑ Registration

"mouse" → Diffusion Model →

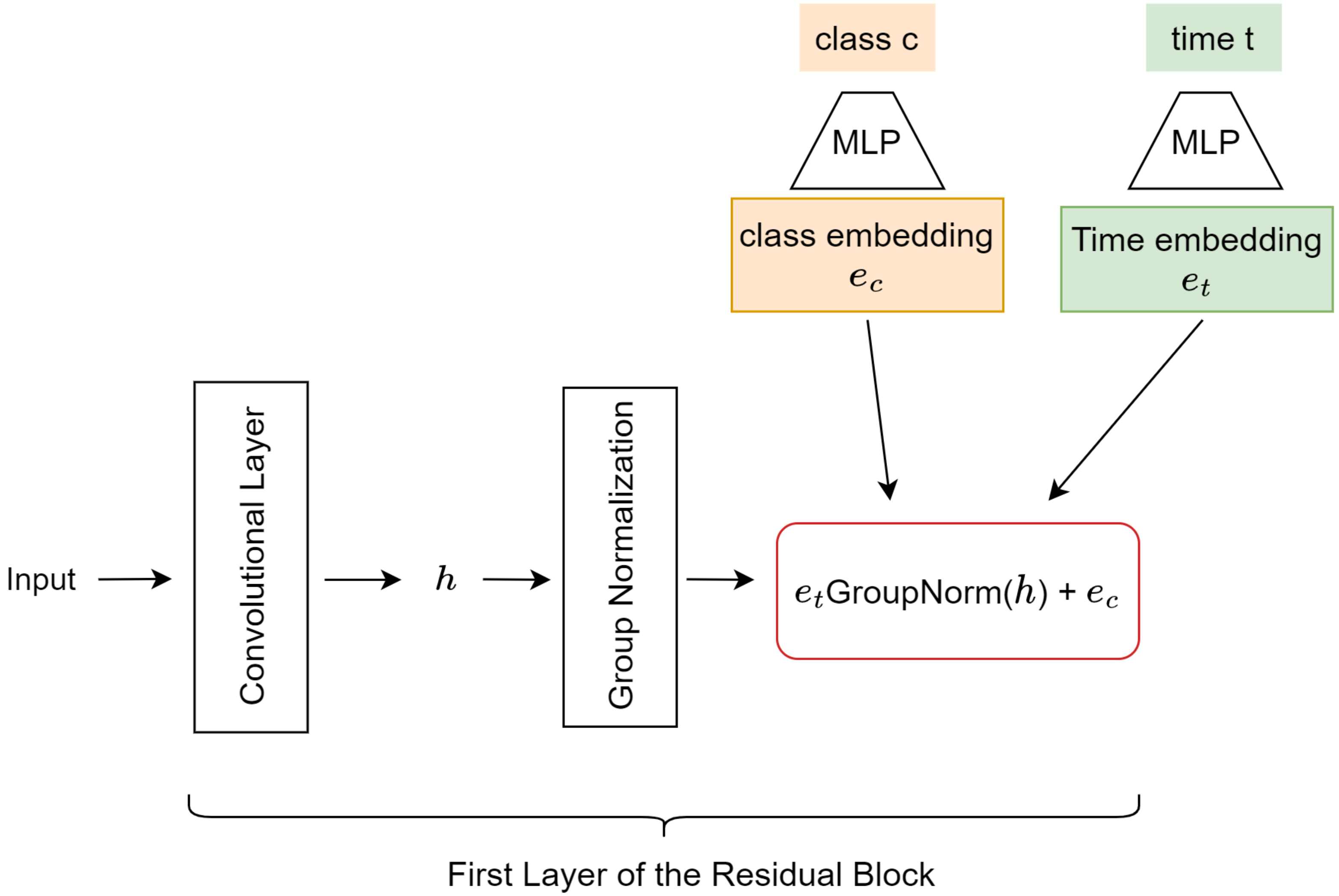
Scalar Conditioning via Spatial Addition

- We train a class-conditional diffusion model by including a class label c.
- We compute a class embedding, and pass it to the residual blocks by spatial addition.



Nichol, Alexander Quinn, and Prafulla Dhariwal. "Improved denoising diffusion probabilistic models." *International Conference on Machine Learning*. PMLR, 2021.

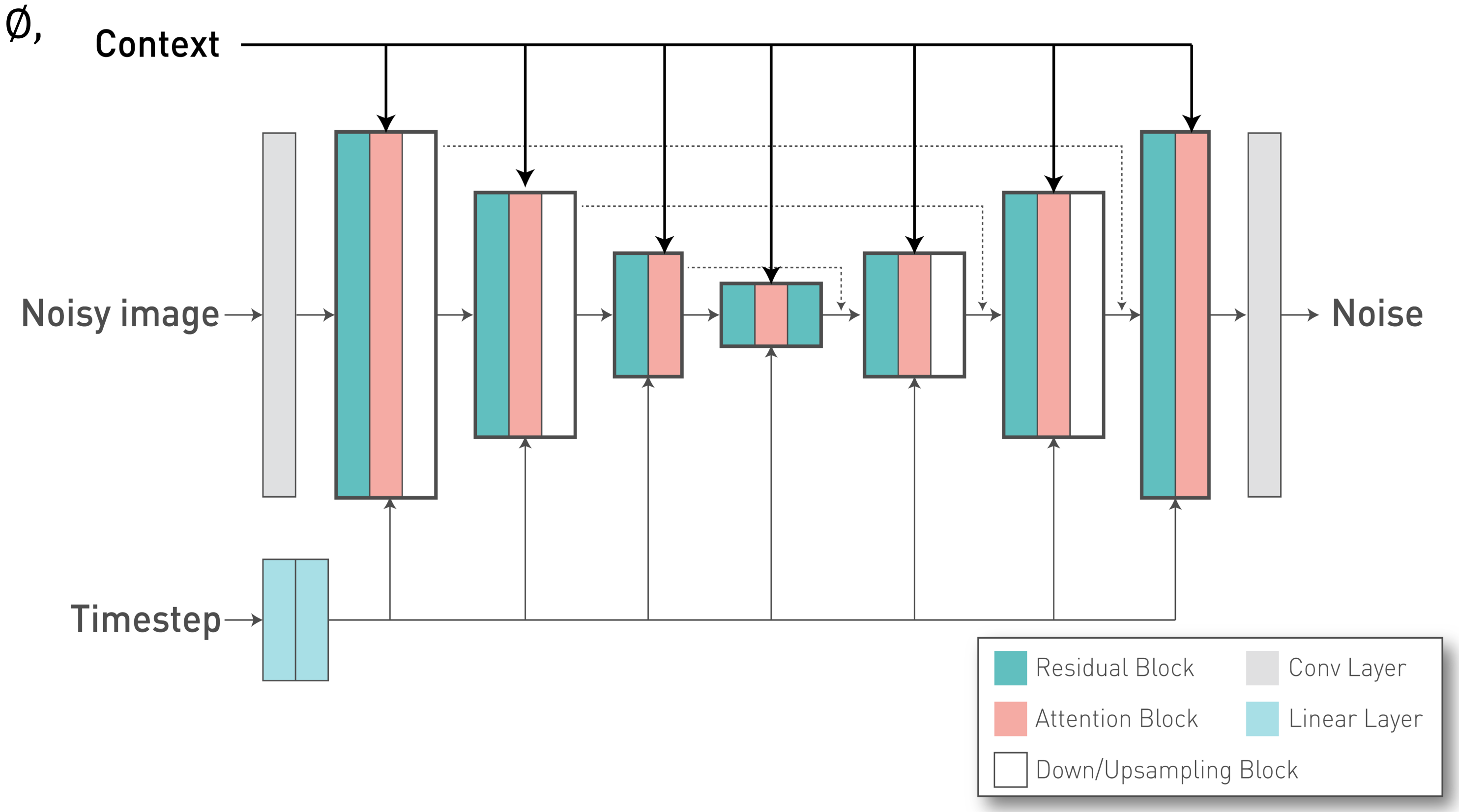
Scalar Conditioning via Adaptive Group Normalization



- Similar to StyleGAN, we add time and class information using a group normalization layer.
- This happens in all residual blocks of the U-Net.

Dhariwal, P., & Nichol, A. (2021). Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, *34*, 8780-8794.

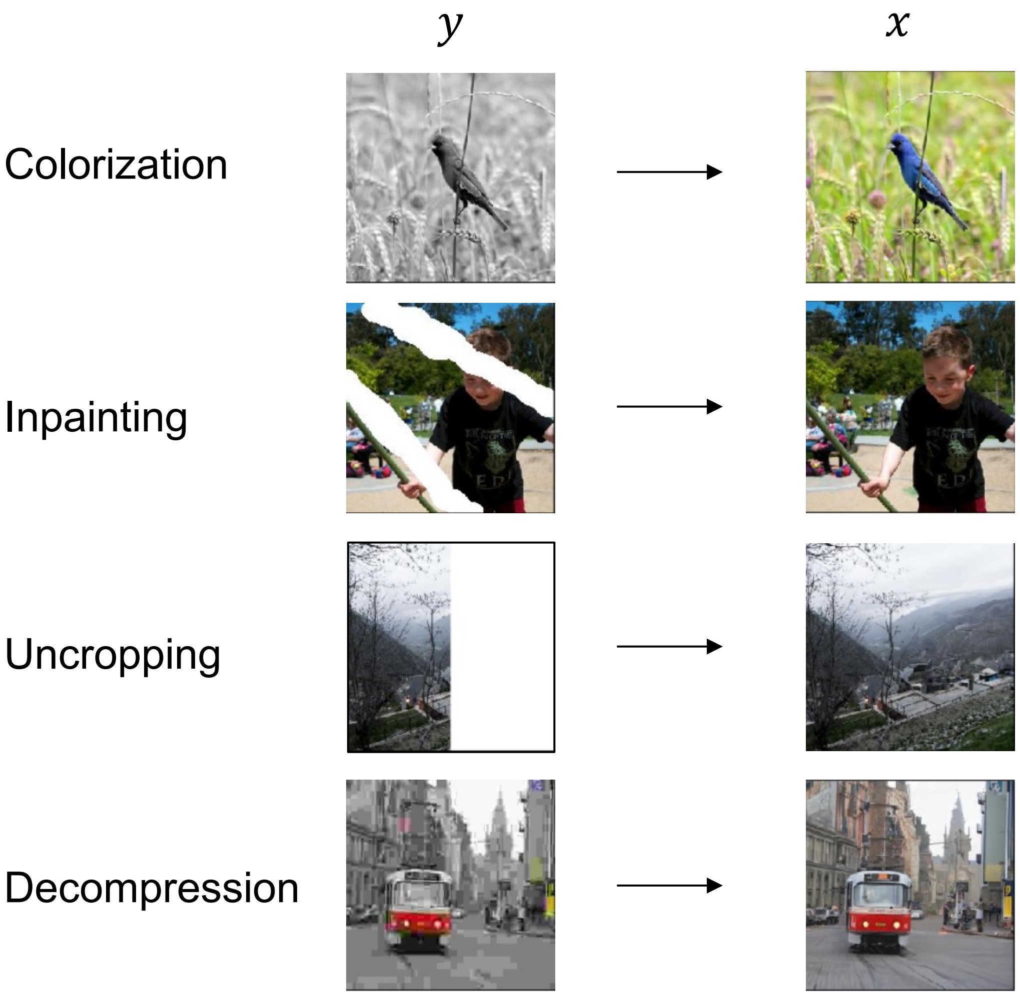
Classifier-free Guidance



$$\widetilde{\epsilon_\theta}(x_t|y) = \epsilon_\theta(x_t|\emptyset) + w[\epsilon_\theta(x_t|y) - \epsilon_\theta(x_t|\emptyset)]$$

Check out this very nice tutorial: https://sander.ai/2022/05/26/guidance.html

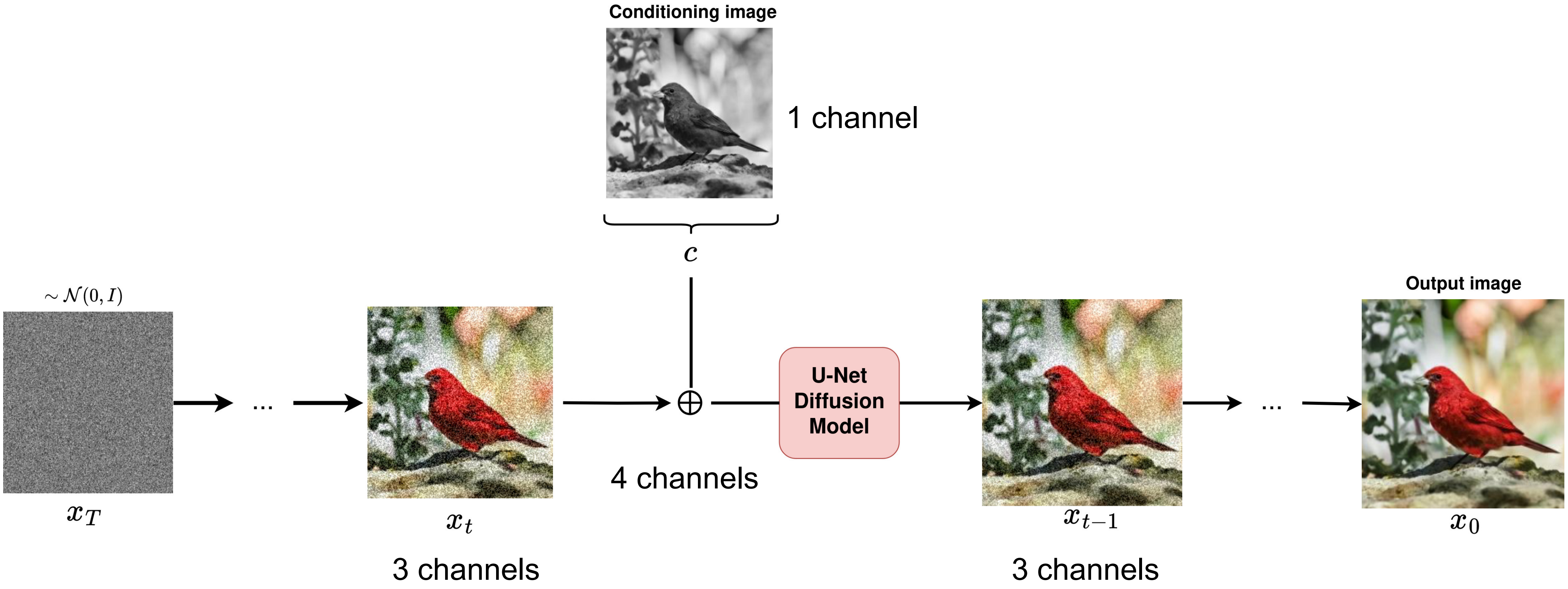
Image Conditioning through Concatenation



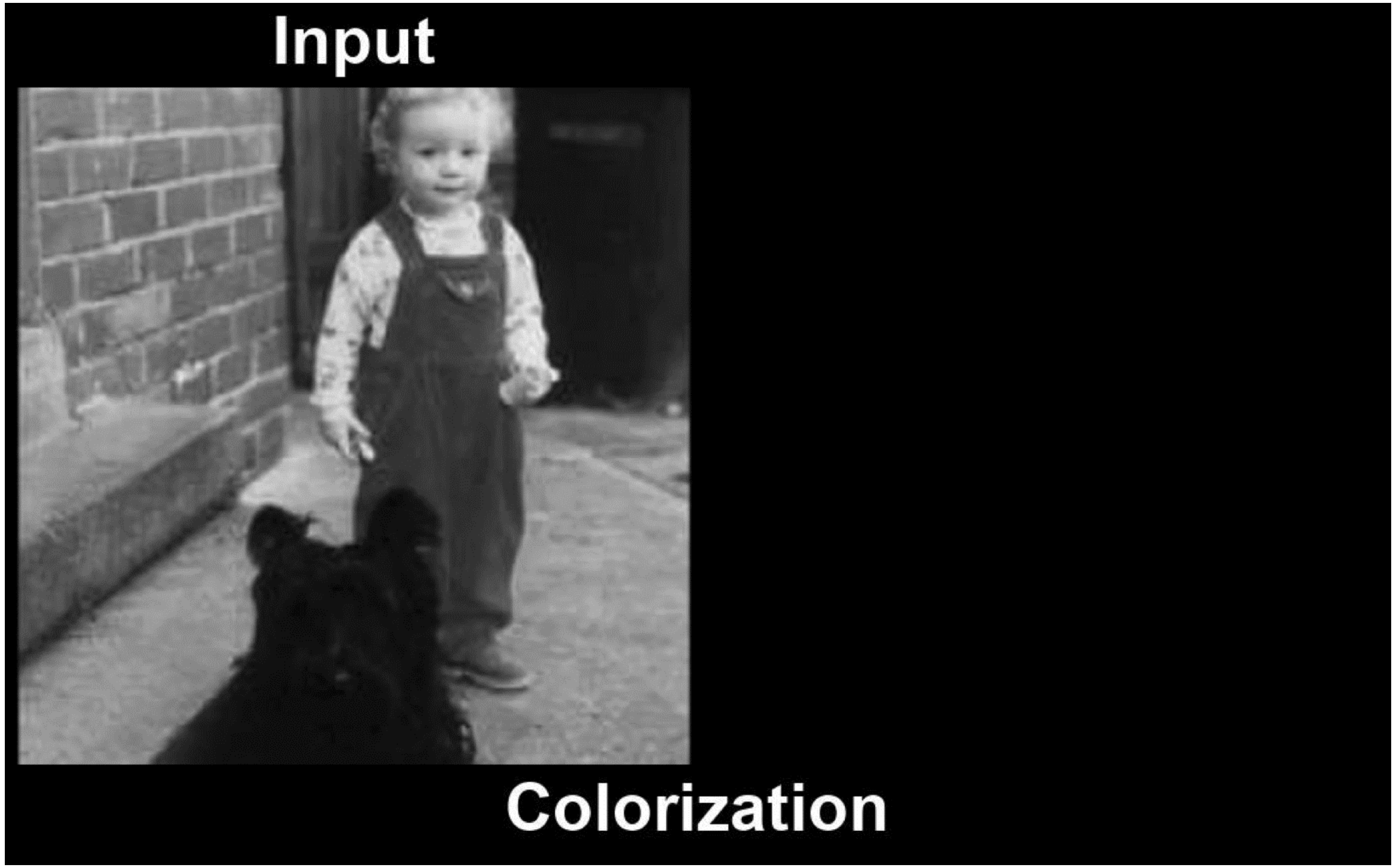
- For image generation of a fake image x, we can use a conditioning image y.
- This requires **paired** training.
- During training and sampling, we add information of the conditioning image x through **channel-wise concatenation.**

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

Image Conditioning through Concatenation



Palette: Image-to-Image Diffusion Models

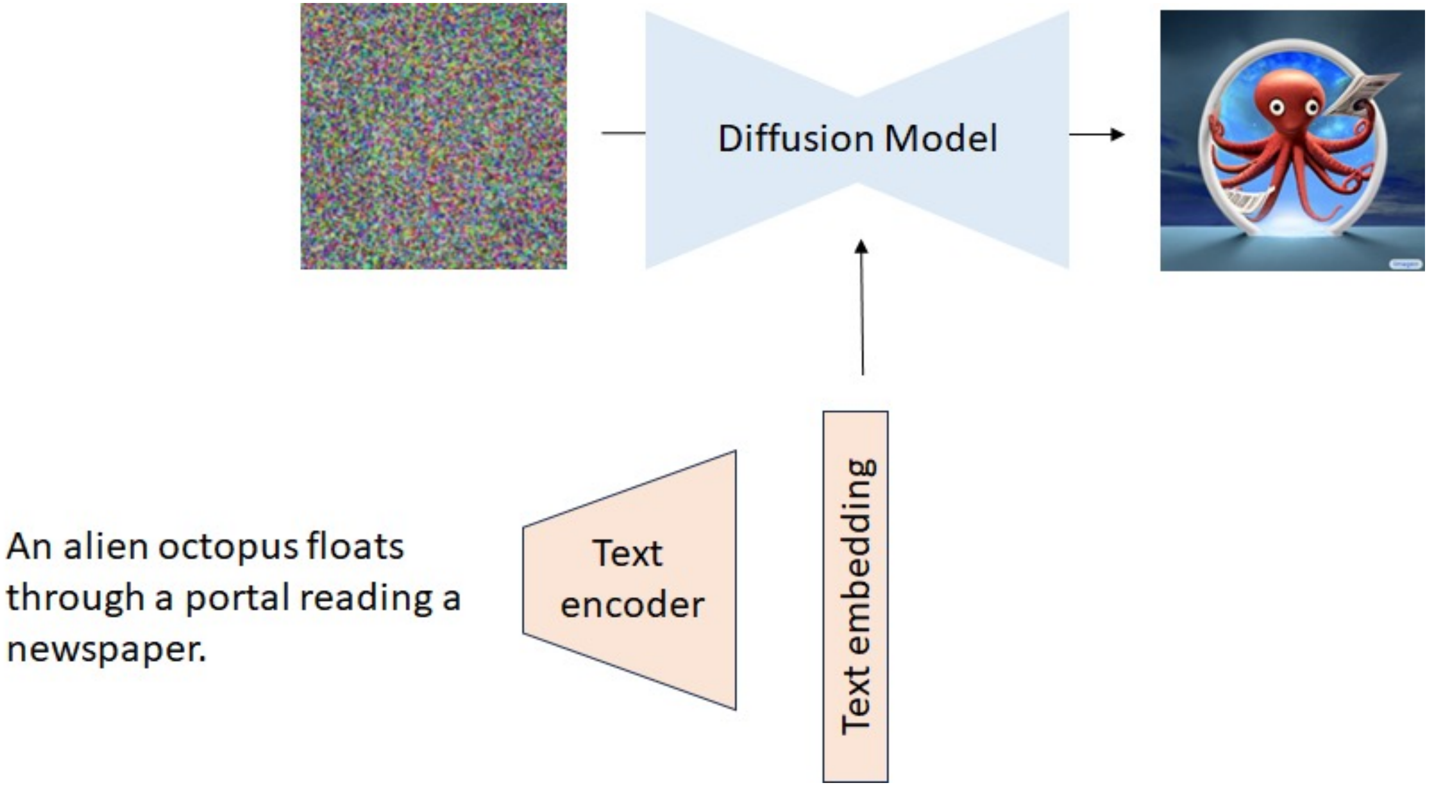


Saharia, Chitwan, et al. "Palette: Image-to-image diffusion models." *ACM SIGGRAPH 2022 Conference Proceedings*. 2022.

Text Conditioning

"A small cactus wearing a straw hat and neon sunglasses in the Sahara desert."

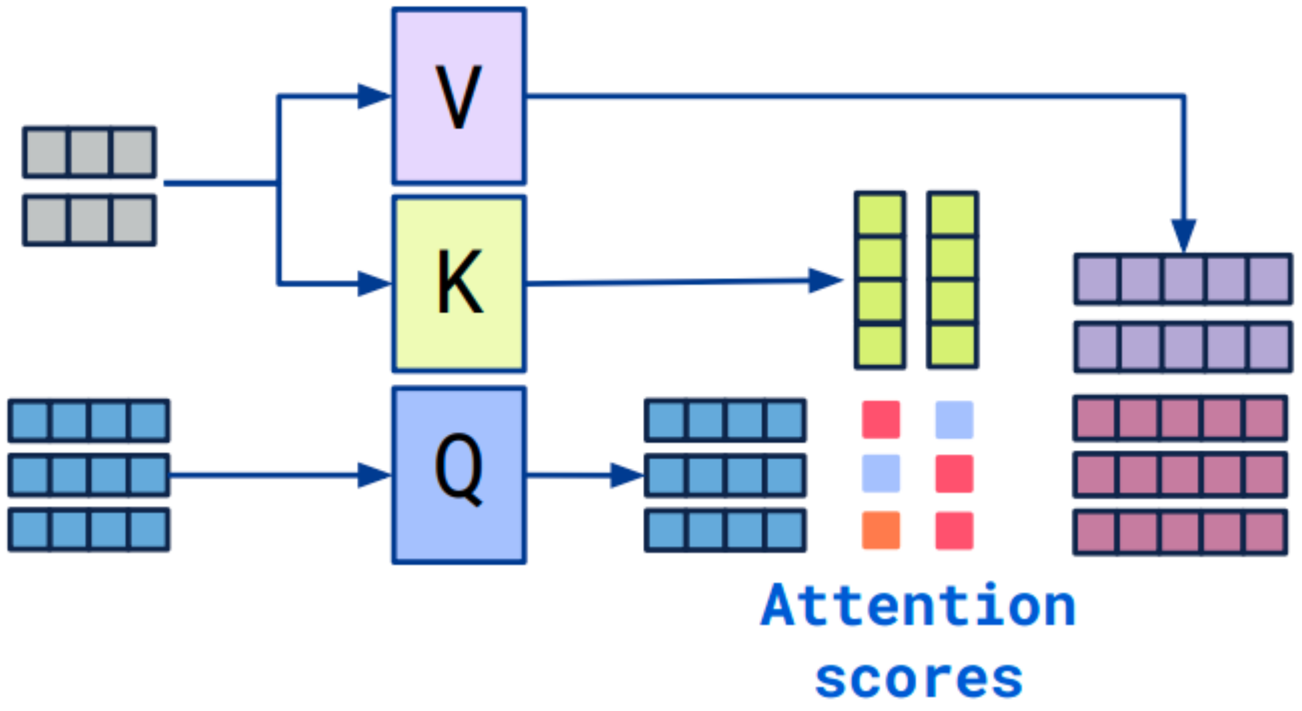
CLIP
Dall-E
Stable Diffusion
Imagen
...



Saharia, Chitwan, et al. "Photorealistic text-to-image diffusion models with deep language understanding." *Advances in Neural Information Processing Systems* 35 (2022): 36479-36494.

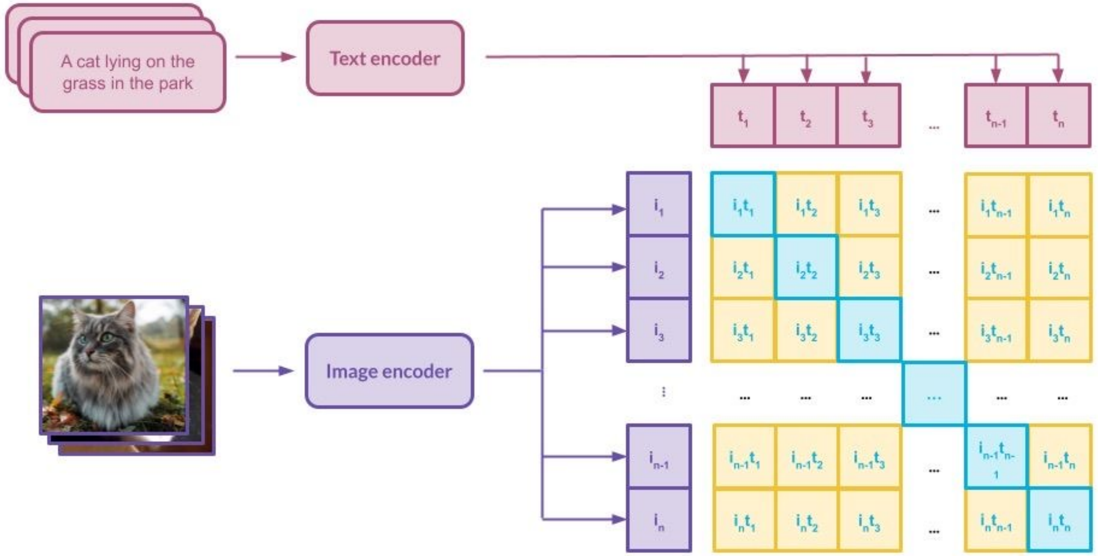
Architecture - Conditioning

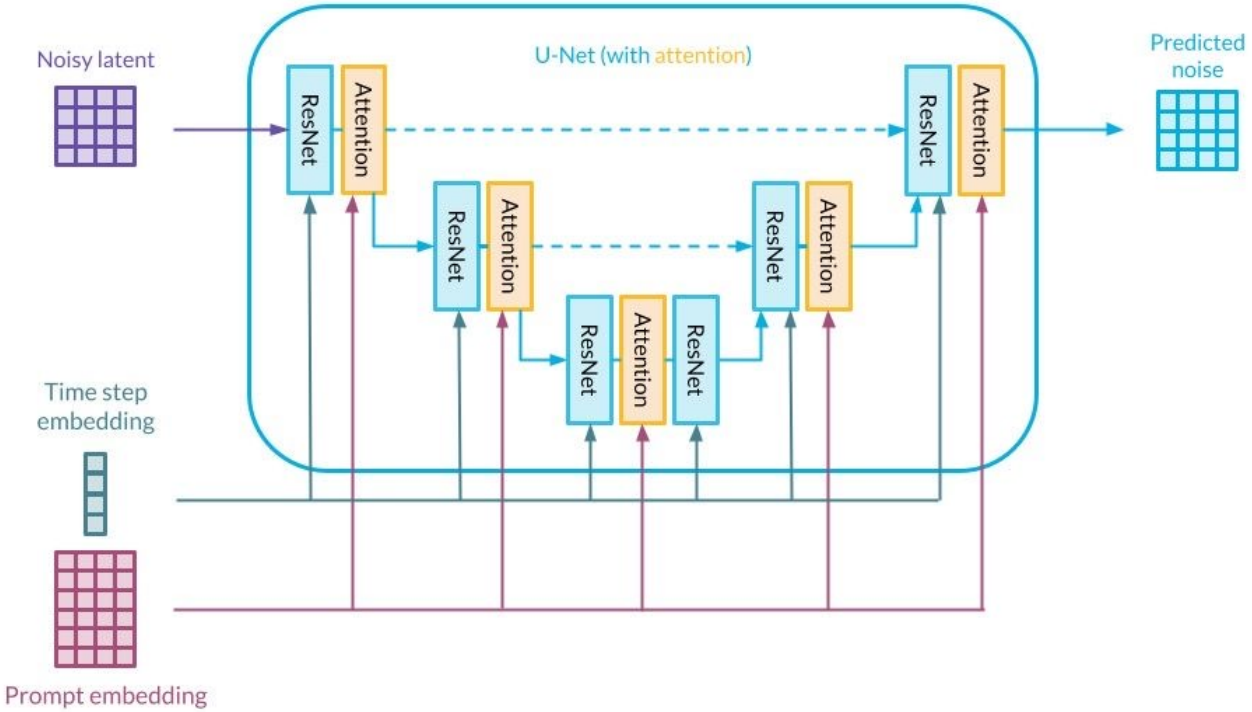
- Transformers **Cross-Attention**
- We use the text embedding to generate the key / value pair.
- We use the image embedding for the query.



Jaegle et al (2022). Perceiver IO: A General Architecture for Structured Inputs & Outputs. In ICL.

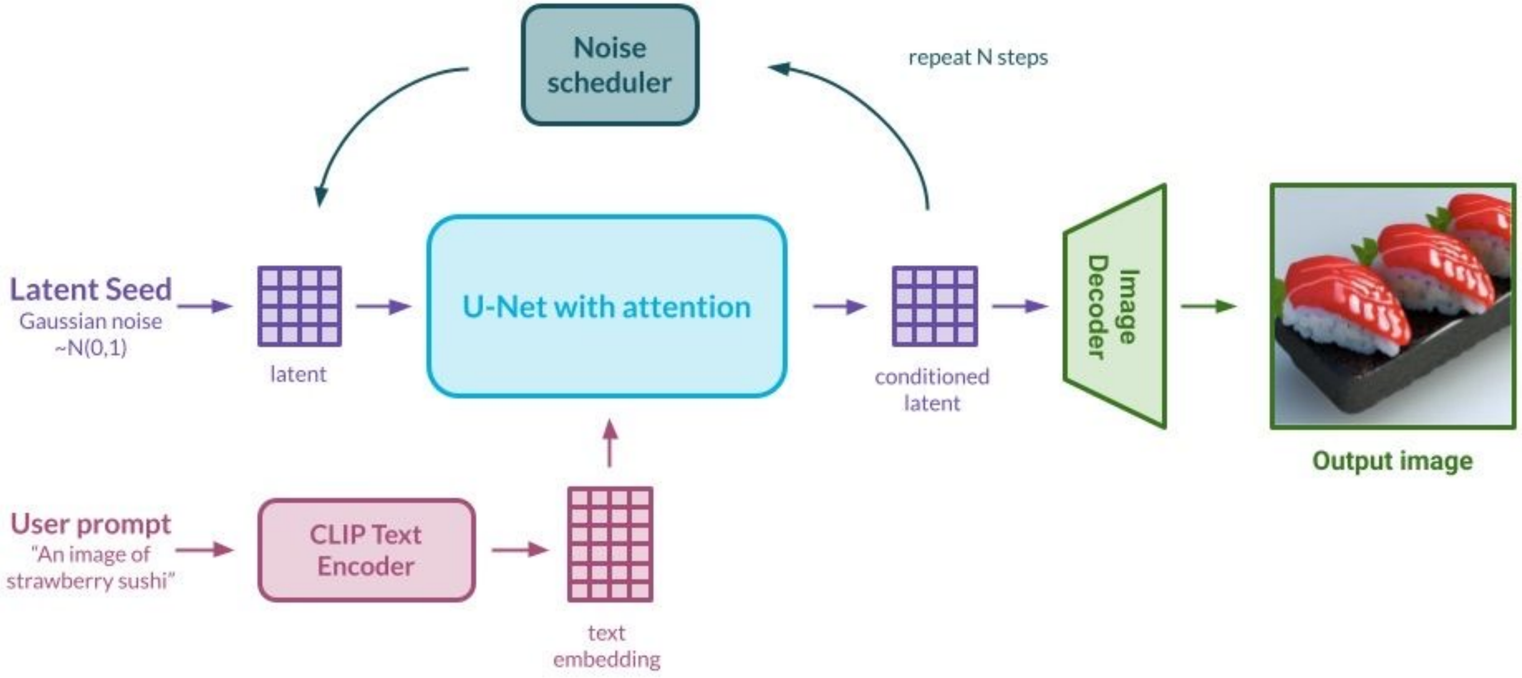
Text Conditioning

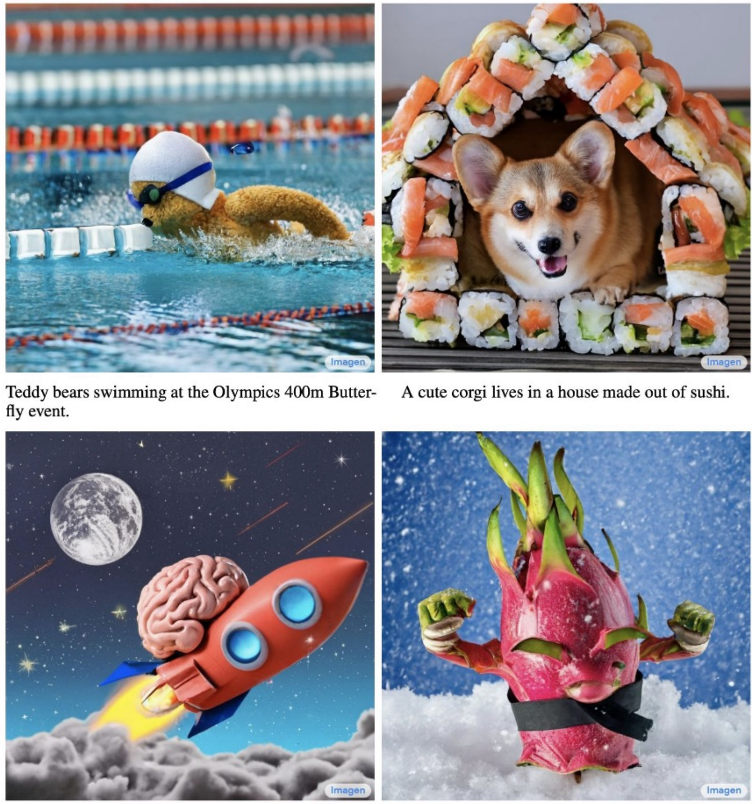




Hertz, Amir, et al. "Prompt-to-prompt image editing with cross attention control." *arXiv preprint arXiv:2208.01626* (2022).
https://deepsense.ai/diffusion-models-in-practice-part-1-the-tools-of-the-trade/

Text Conditioning





Teddy bears swimming at the Olympics 400m Butter-fly event.

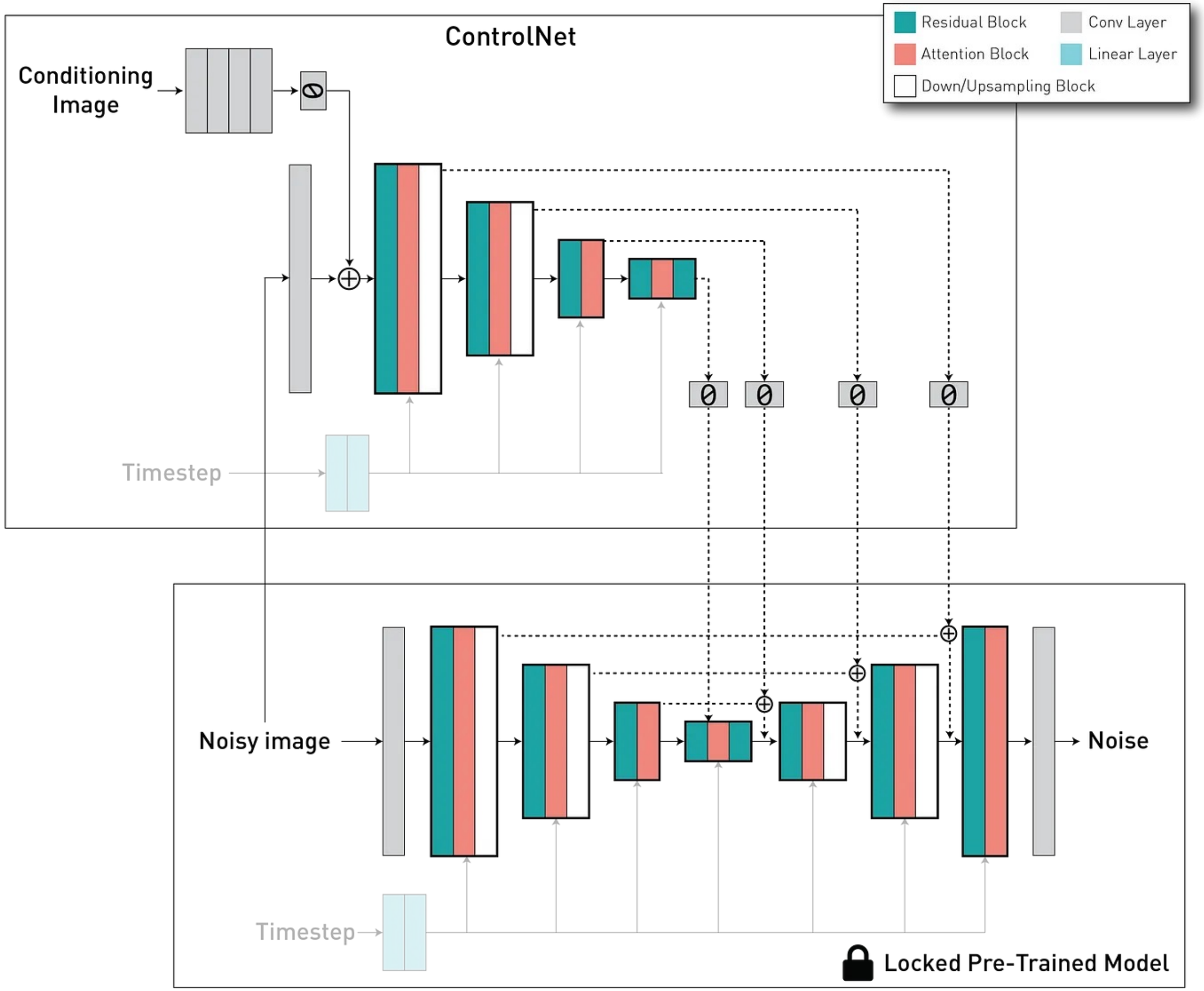
A cute corgi lives in a house made out of sushi.

A brain riding a rocketship heading towards the moon.

A dragon fruit wearing karate belt in the snow.

https://deepsense.ai/diffusion-models-in-practice-part-1-the-tools-of-the-trade/

ControlNet

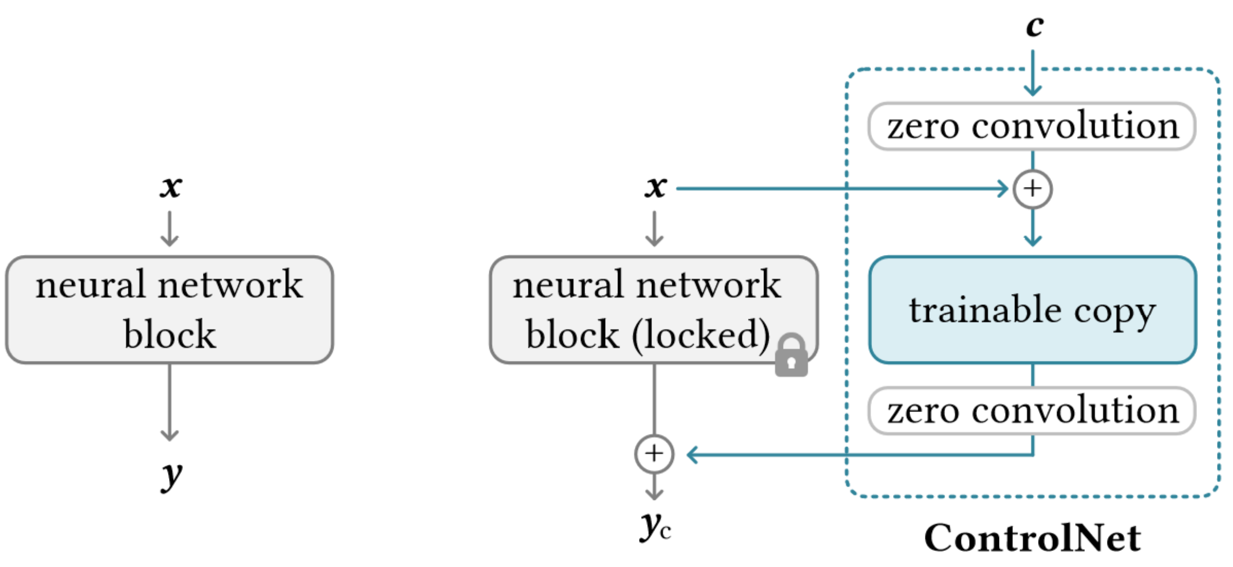


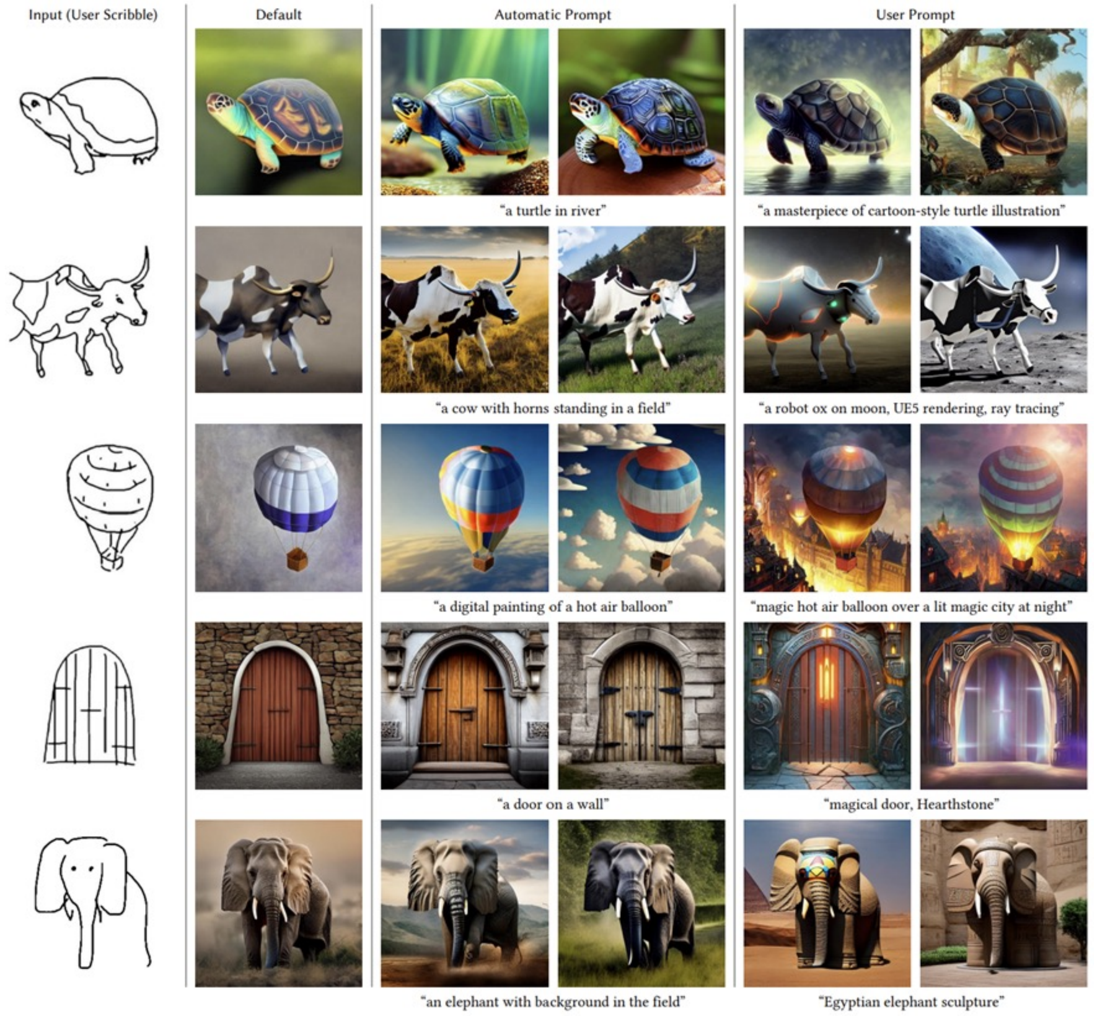
- We pretrain a diffusion model with text prompts.
- We freeze this model.
- We fine-tune a copy conditioned on $\boldsymbol{c}$.
- We pass information to the frozen model through skip connections.

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.

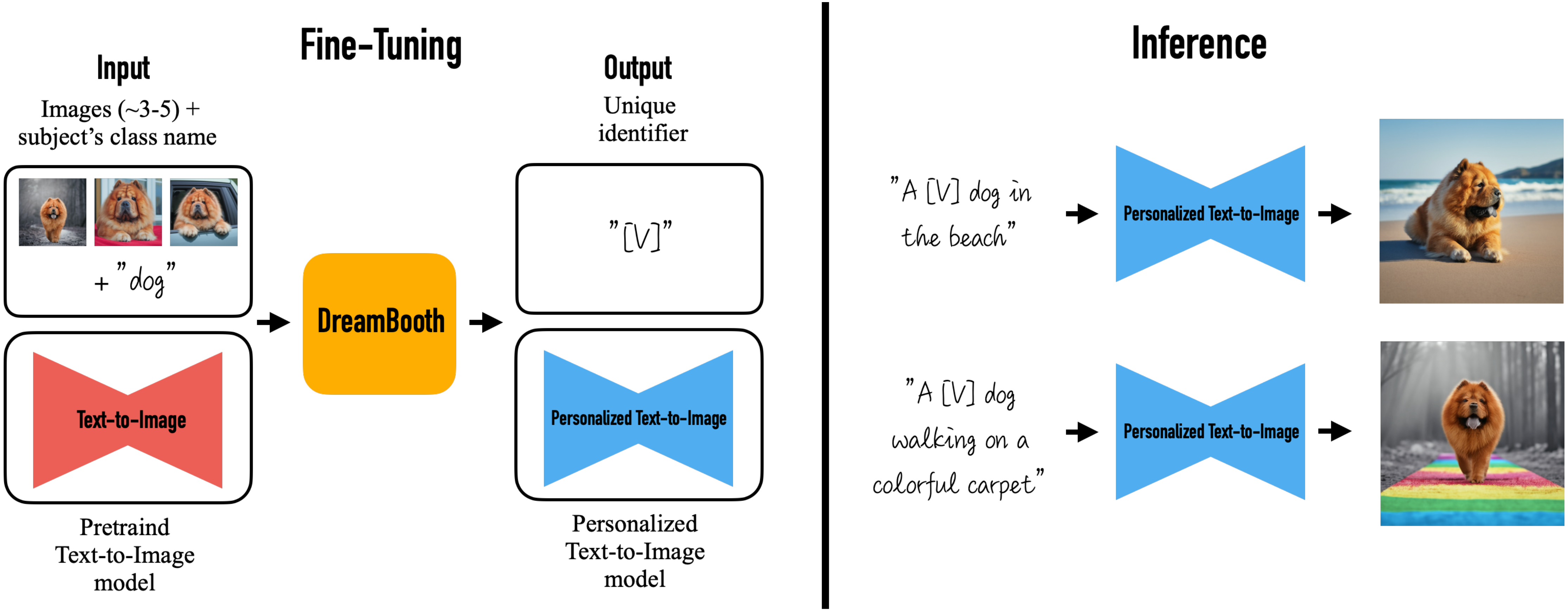
ControlNet

conditioning image





DreamBooth



Ruiz, N., Li, Y., Jampani, V., Pritch, Y., Rubinstein, M., & Aberman, K. (2023). Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation. CVPR

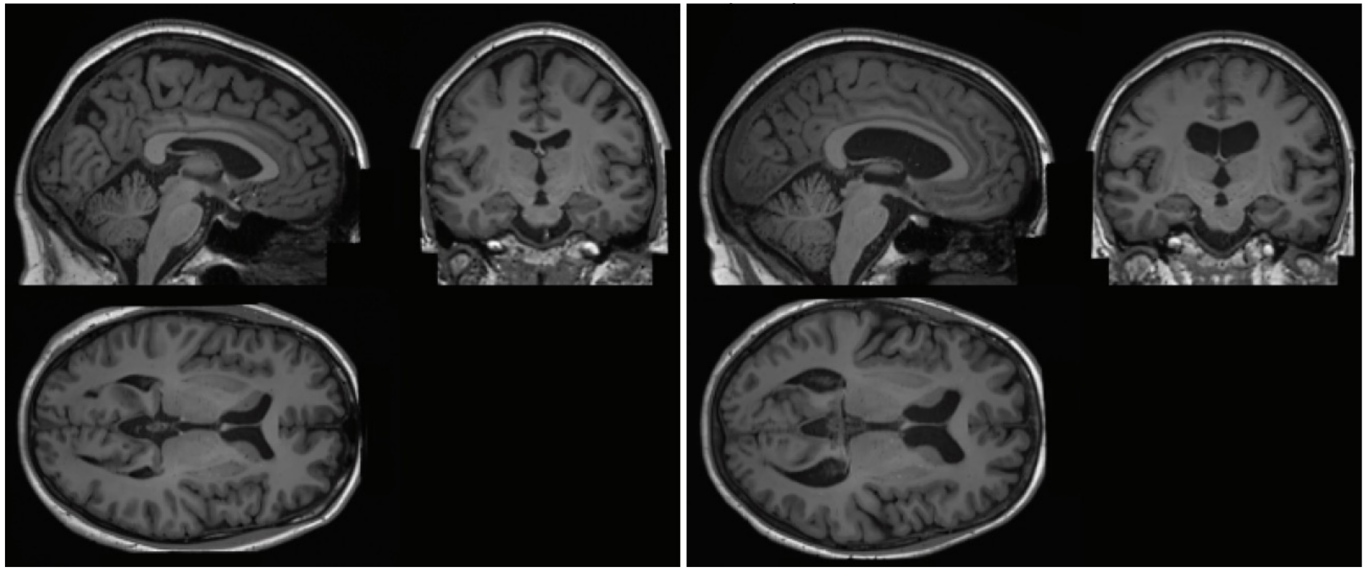
Agenda

- ❑ Introduction to Diffusion Models [~30 min]
  - ❑ Physical Intuition & General Concepts
  - ❑ Denoising Diffusion Probabilistic Models
  - ❑ A Score-Based View on Diffusion Models
- ❑ Advanced Topics [~30 min]
  - ❑ Sampling Strategies
  - ❑ Inference-time Conditioning
  - ❑ Training-time Conditioning
- ❑ Applications in Medical Imaging [~30 min]
  - ❑ Synthesis
  - ❑ Inpainting
  - ❑ Segmentation
  - ❑ Anomaly Detection
  - ❑ Reconstruction
  - ❑ Registration

Image Synthesis

Examples from the Community

Real



Synthetic

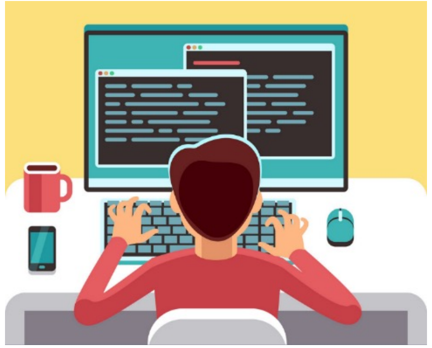
Why? Medical Images are Rare!

Privacy concerns
Costly image acquisition

Limited data

Data Provider

Data Users



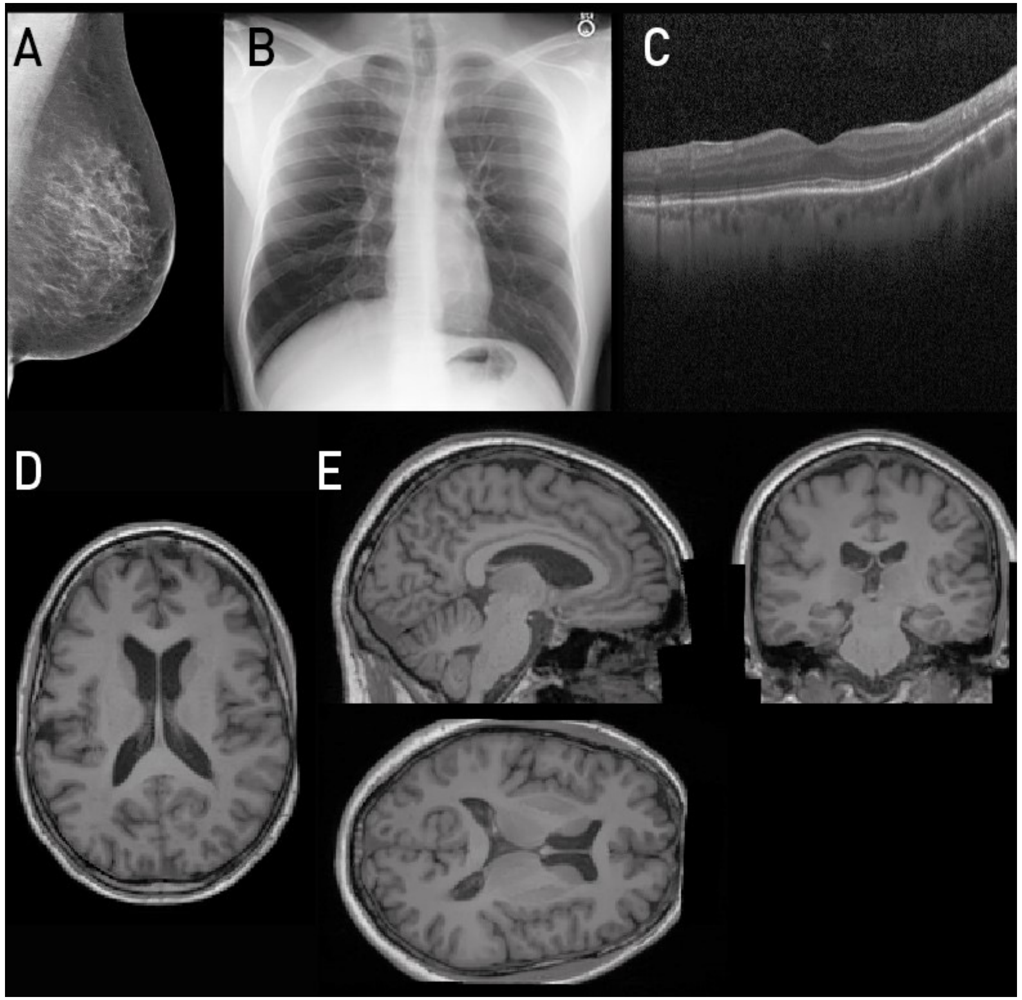
Use of Synthetic Data

What can we use generated images for?

- Full private training
- Data augmentation
- Testing edge cases

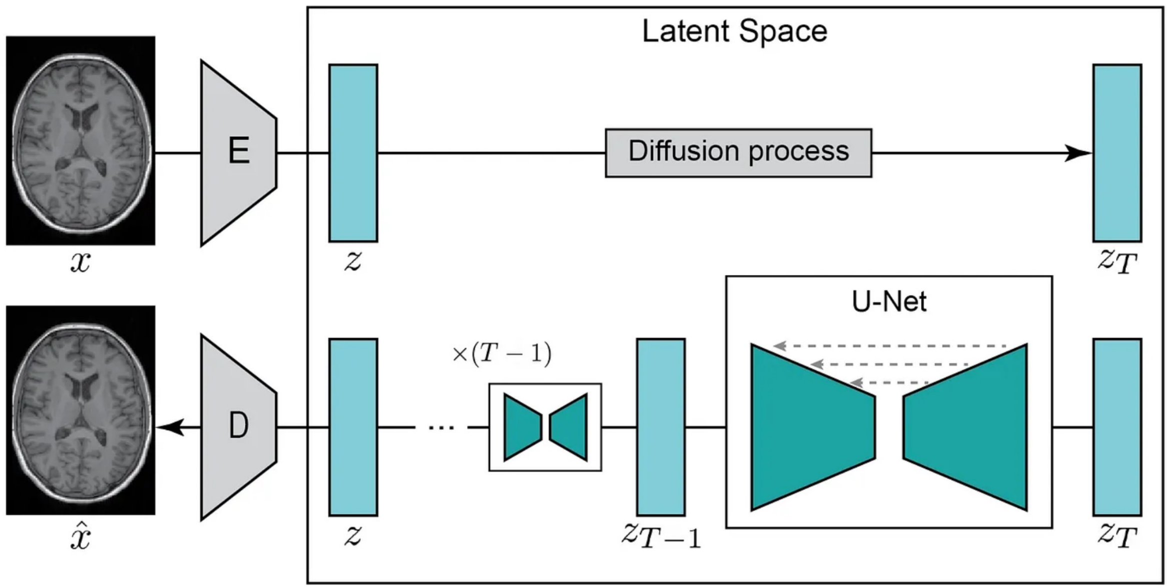
Evaluation Criteria:

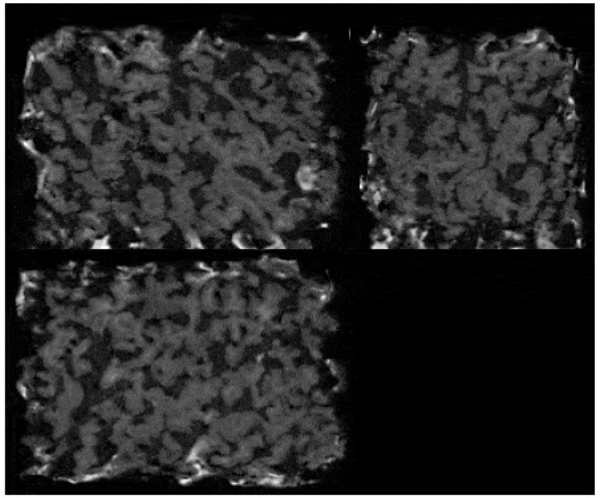
- Realism
- Diversity
- Privacy

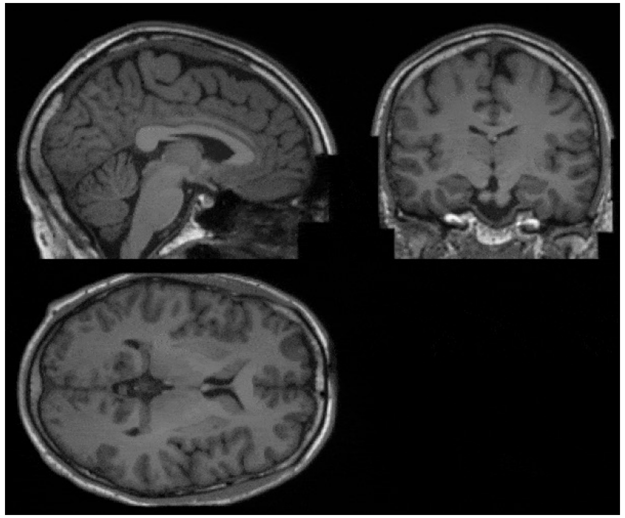


Pinaya, Walter HL, et al. "Generative AI for Medical Imaging: extending the MONAI Framework." arXiv preprint arXiv:2307.15208 (2023).

Generating High-Resolution 3D Brain Data







Latent Diffusion Model trained on data from UK Biobank (N = 31,740)

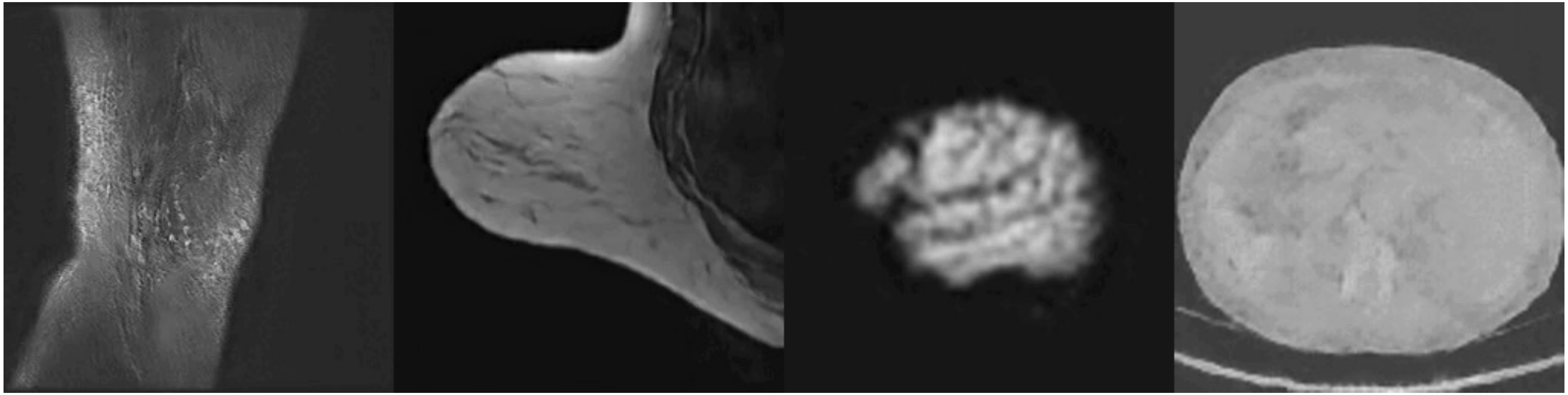
- Allows for the generation of T1-weighted brain MR images with a resolution of 160 x 224 x 160

Can be conditioned on covariates like:

- Age
- Gender
- Ventricular and Brain volumes

Pinaya, W. H., et al. (2022) Brain image generation with latent diffusion models. In *MICCAI Workshop on Deep Generative Models*

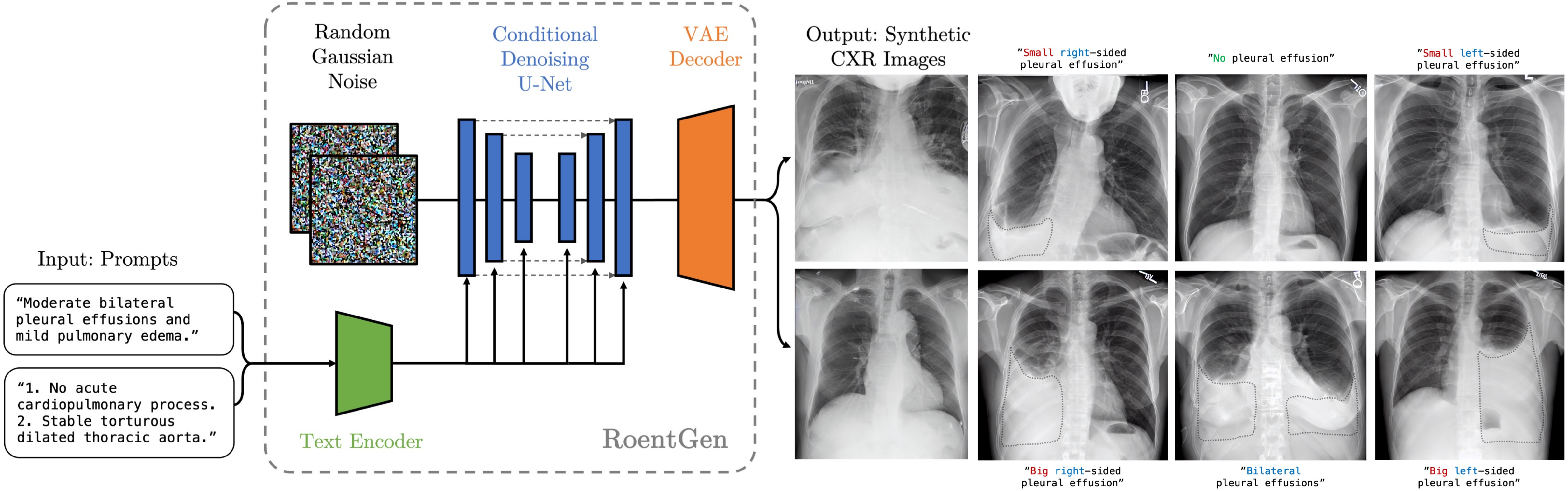
High-Resolution 3D Medical Image Generation with LDMs



DUKE (256 x 256 x 32) | MRNet (256 x 256 x 32) | ADNI (64 x 64 x 64) | LIDC-IDRI (128 x 128 x 128)

Khader, F., *et al.* (2023) Denoising diffusion probabilistic models for 3D medical image generation. In *Scientific Reports*

Fine-Tuning Stable Diffusion

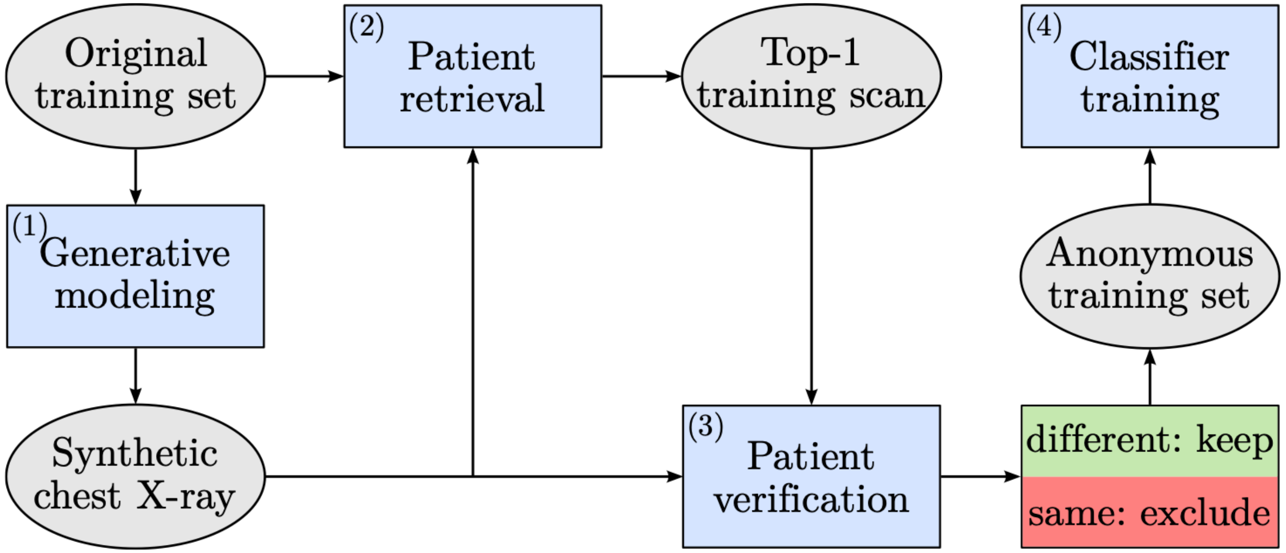


Fine Tuned Stable Diffusion pipeline for text-conditioned chest X-Ray generation

- U-Net and Text Encoder are jointly fine-tuned on the MIMIC-CXR dataset

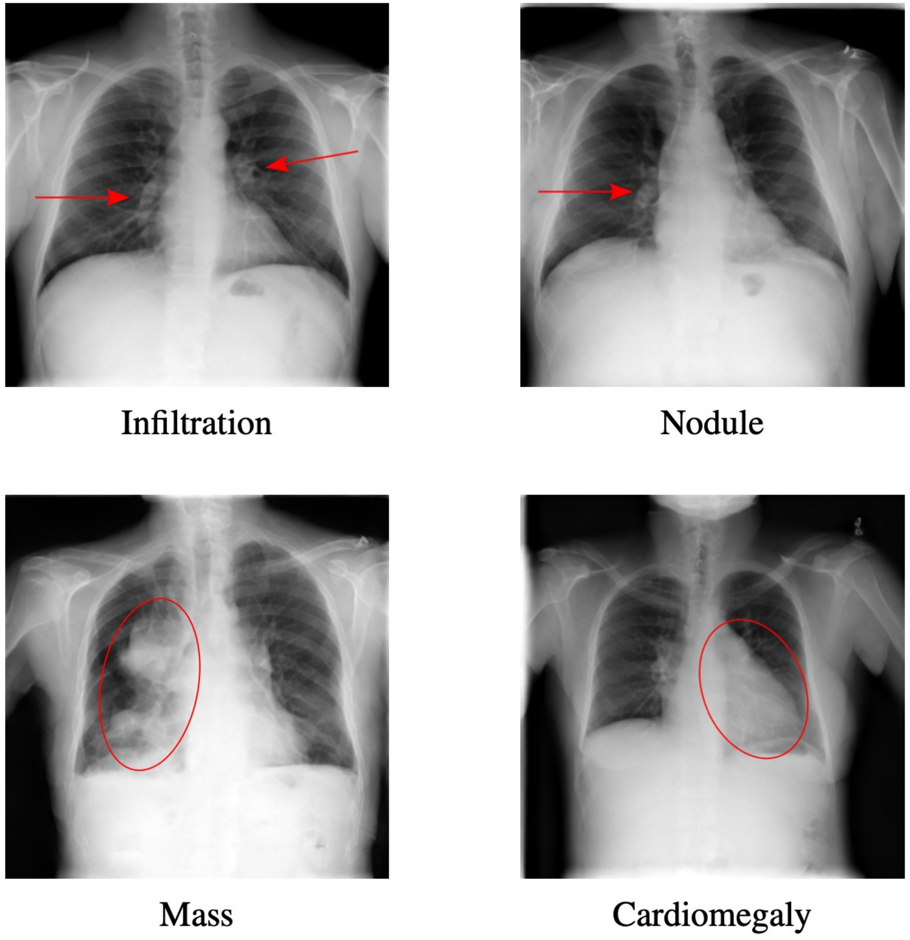
Chambon, Pierre, et al. (2022) RoentGen: vision-language foundation model for chest x-ray generation. *arXiv:2211.12737*

Generation of Anonymous Chest Radiographs



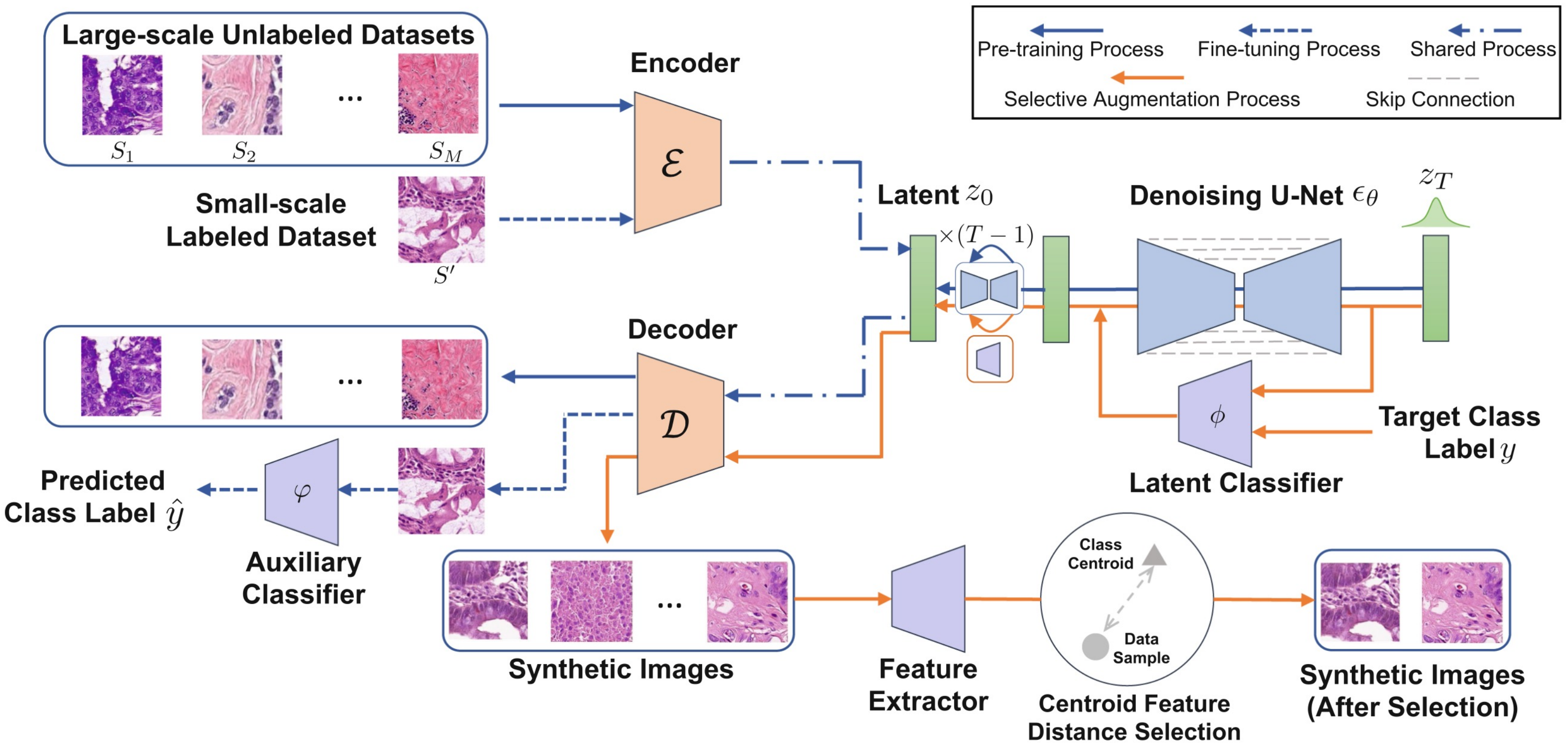
Privacy enhanced sampling strategy for conditional chest X-Ray generation

- Generated images are not per se private, as generative models may memorize training examples
- 1) Generate Image 2) Find most similar one from training data 3) Verify if patient is the same → include/exclude



Packhäuser et al (2022). Generation of Anonymous Chest Radiographs Using Latent Diffusion Models for Training Thoracic Abnormality Classification Systems. arXiv:2211.01323

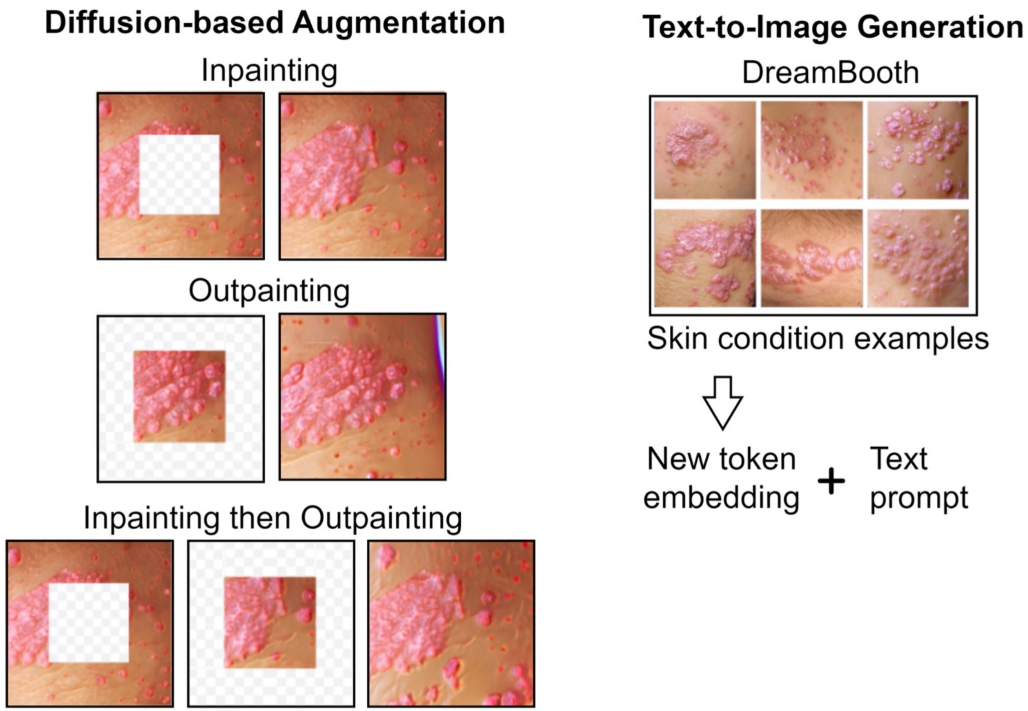
Synthetic Data Augmentation



1) **Pre-training an unconditional Latent Diffusion Model** on a large unlabeled dataset
2) **Conditional fine-tuning** on an unseen labeled (small) dataset through a latent classifier
3) **Selection of the highly-confident synthetic samples** based on feature similarity with real data

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

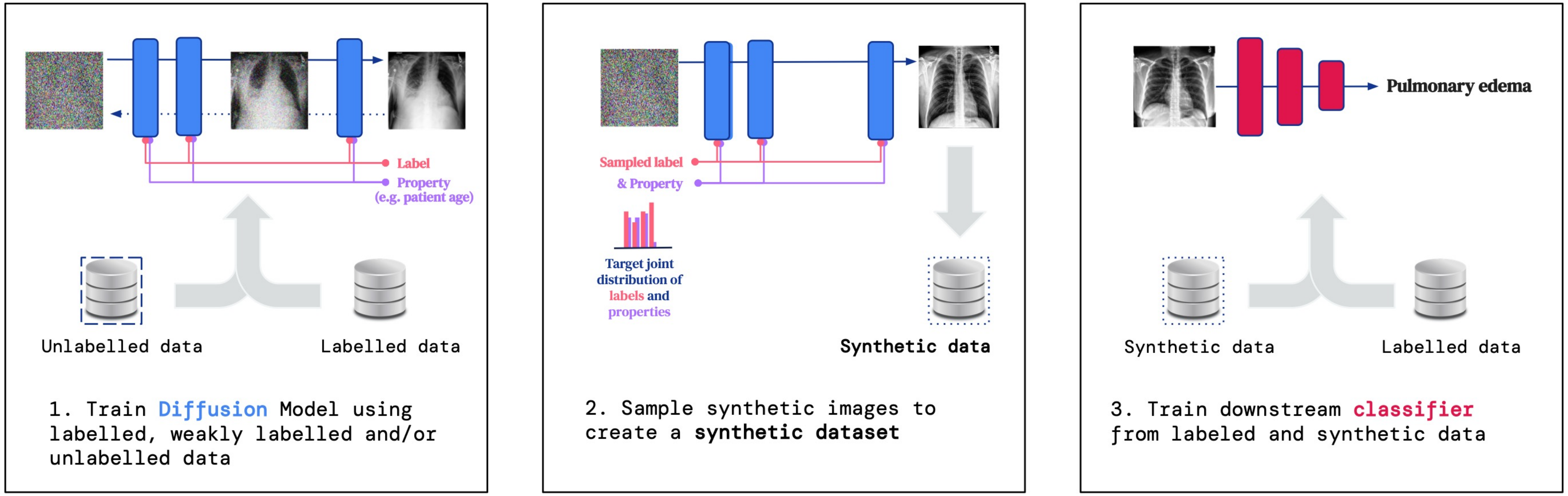
Synthetic Image Augmentation – Performance Gain



- Diffusion models can scalably generate images of skin disease and training models with that augmented data **improves performance in data-limited settings**
- Performance gains (in this setting) saturate at a **synthetic-to-real ratio of 10:1** and is substantially smaller than the gains obtained by adding real images → **collection of diverse real data remains more important**

Sagers, L. W., et al. (2023) Augmenting medical image classifiers with synthetic data from latent diffusion models. arXiv:2308.12453

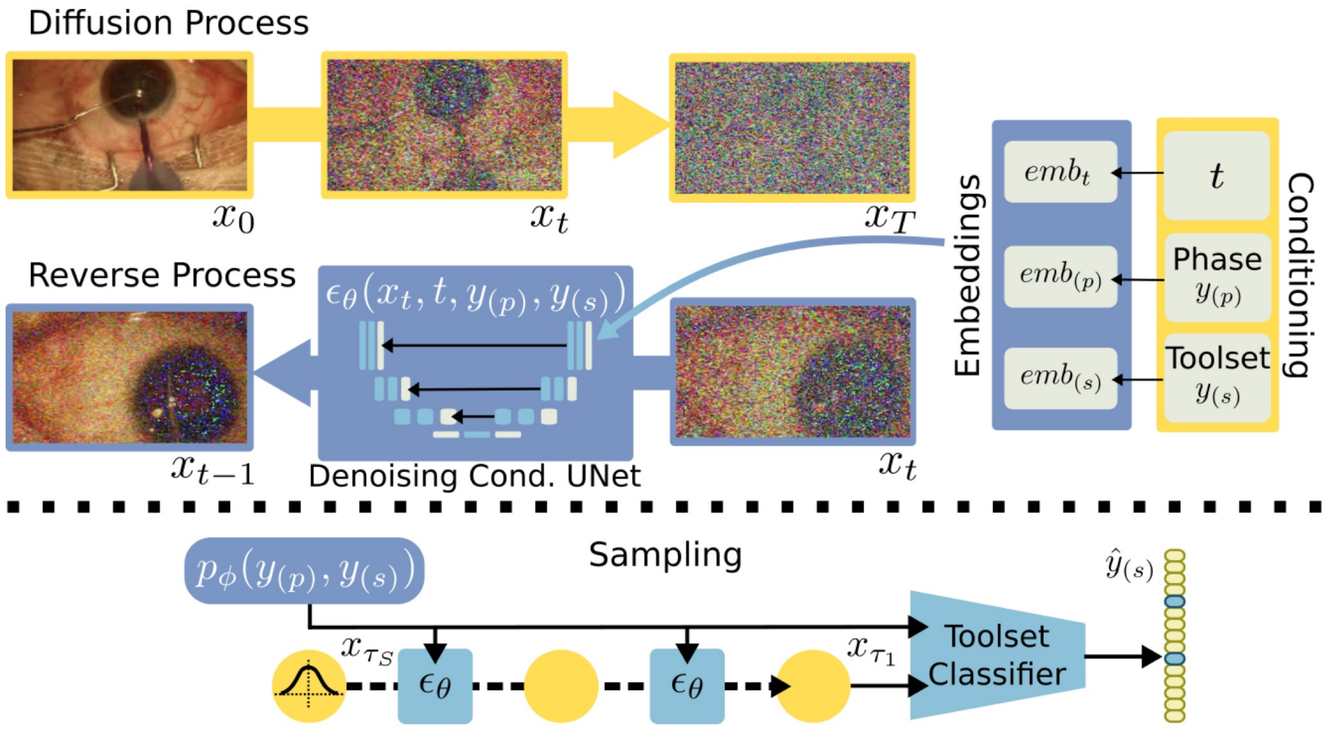
Synthetic Data for Distribution Shifts – Improving Fairness in AI



- Learning realistic augmentations from data is possible in a label-efficient manner using diffusion models
- Unlabeled data can be used to capture the data distribution of different conditions and subgroups → **steer the distribution of synthetic examples according to specific requirements**
- Learned augmentations can surpass heuristic, manually implemented ones by making **models more robust and statistically fair in- and out- of-distribution**

Ktena, I., et al. (2023) Generative models improve fairness of medical classifiers under distribution shifts. arXiv:2304.09218.

Synthesising Rare Samples – Testing Edge Cases

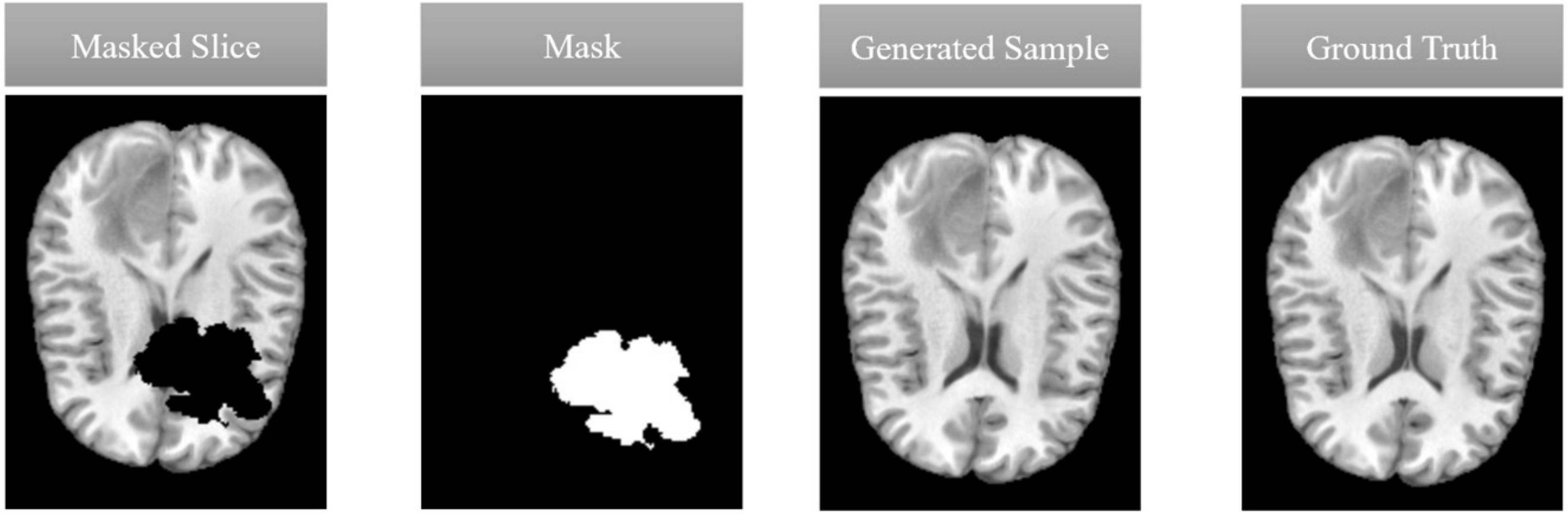


- Diffusion model is used to **synthesize rare cases** in instrument classification in cataract surgery
- Performance gain of > 10% for rare cases

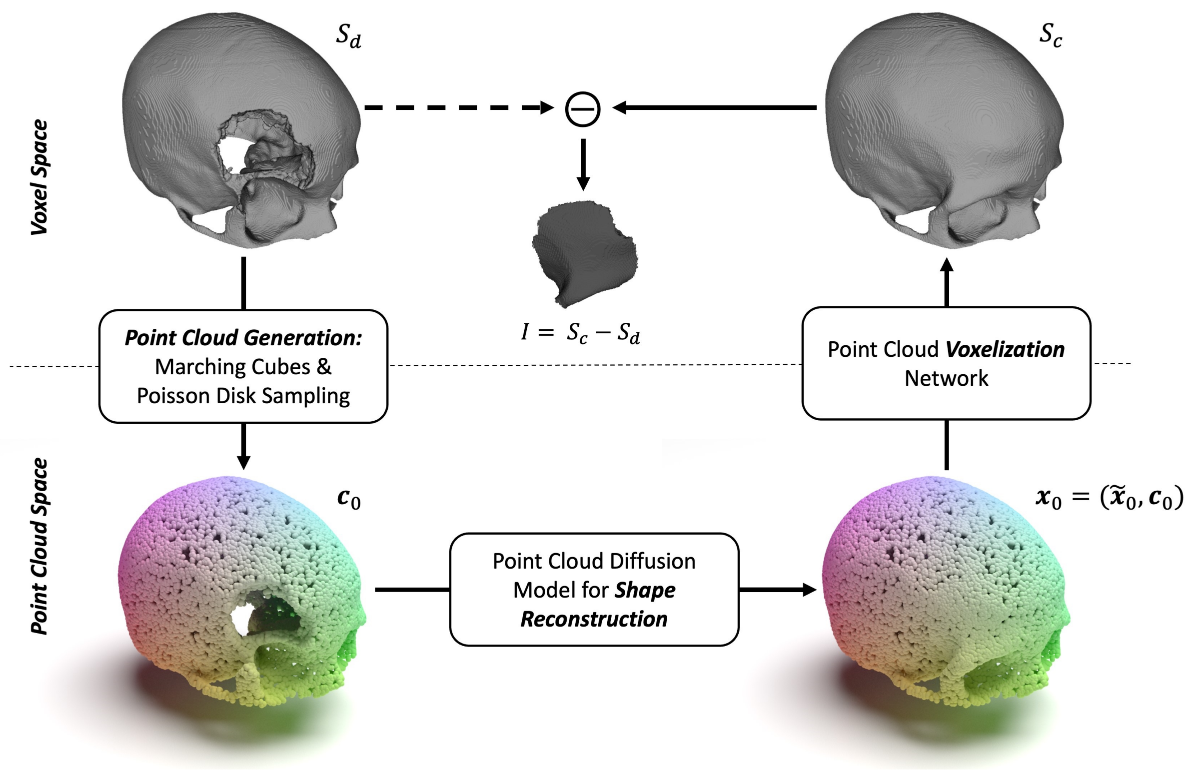
Frisch, Y., et al. (2023) Synthesising Rare Cataract Surgery Samples with Guided Diffusion Models. MICCAI 2023

Inpainting

Examples from the Community



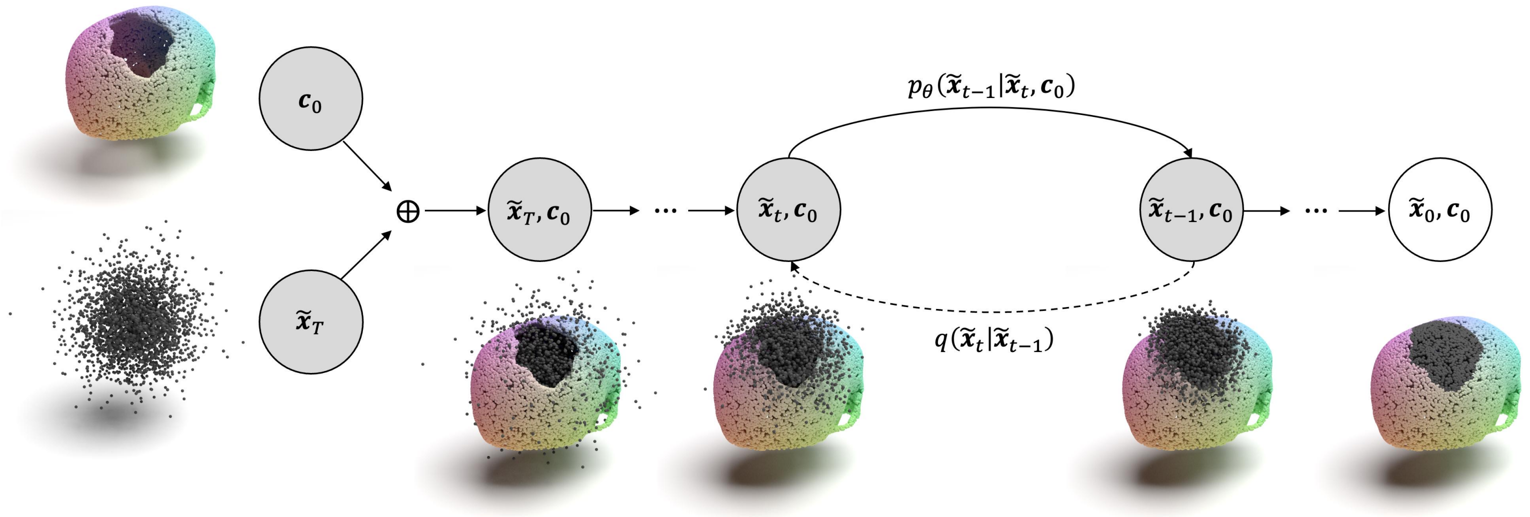
Point Cloud Diffusion Models for Automatic Implant Generation

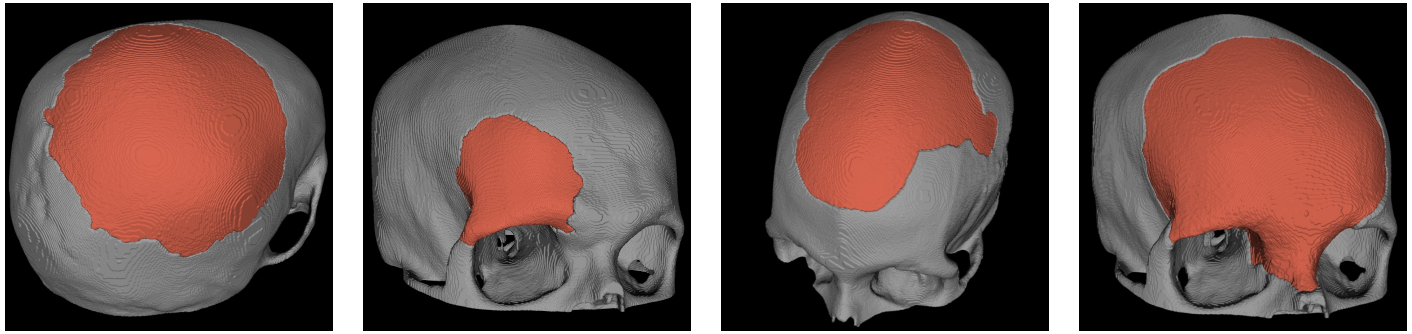


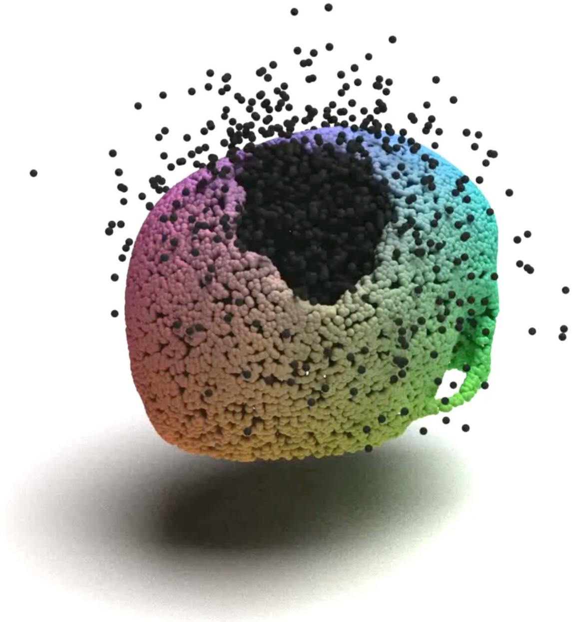
- Implant generation as shape completion task (basically 3D inpainting)
- Applying the diffusion model to high-resolution volumes (512 x 512 x 512) is impossible due to limited GPU memory
- Diffusion models **can be applied to a wide variety of different data types** (like point clouds)

Friedrich, P., et al. (2023) "Point Cloud Diffusion Models for Automatic Implant Generation." MICCAI 2023.

Point Cloud Completion with Diffusion Models



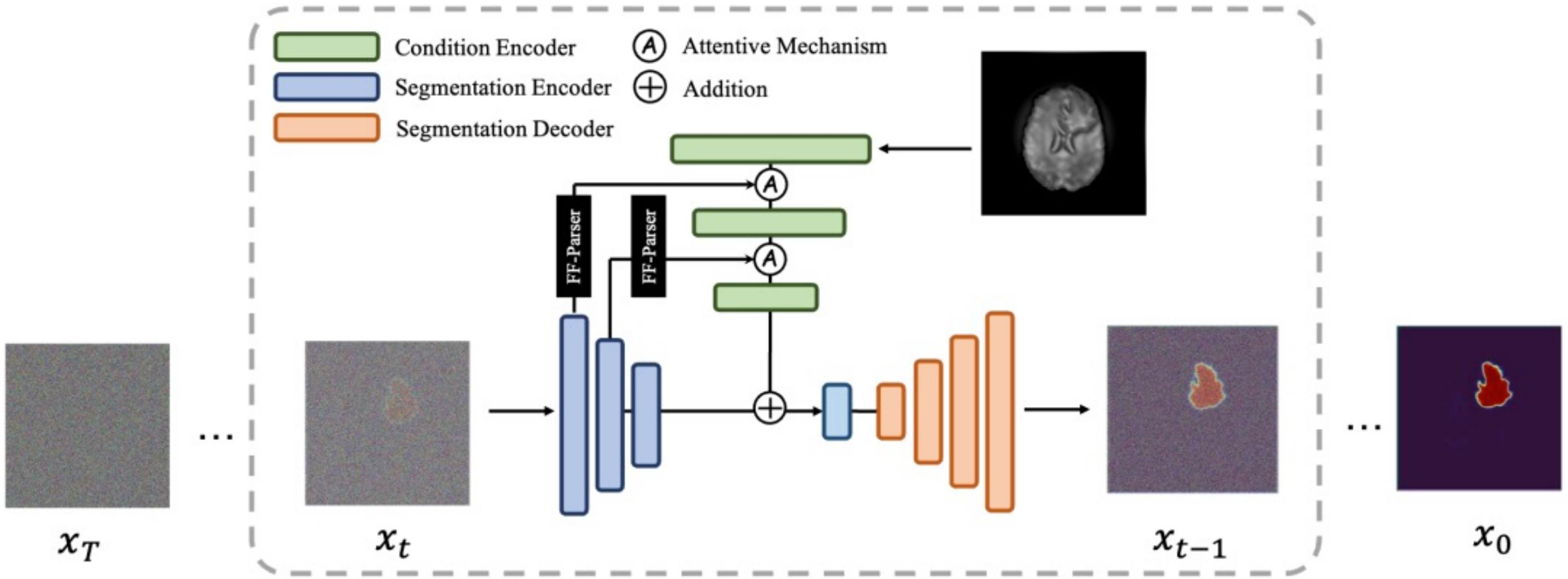




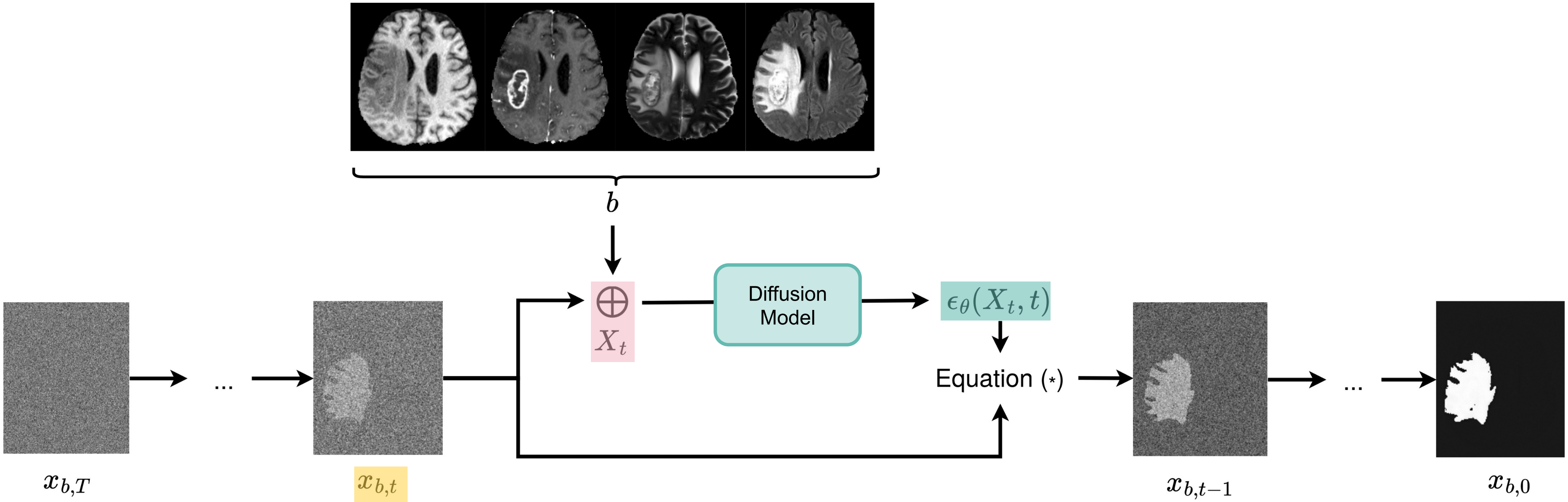
Friedrich, P., et al. (2023) "Point Cloud Diffusion Models for Automatic Implant Generation." MICCAI 2023.

Image Segmentation

Examples from the Community



Diffusion Models for Segmentation Mask Generation

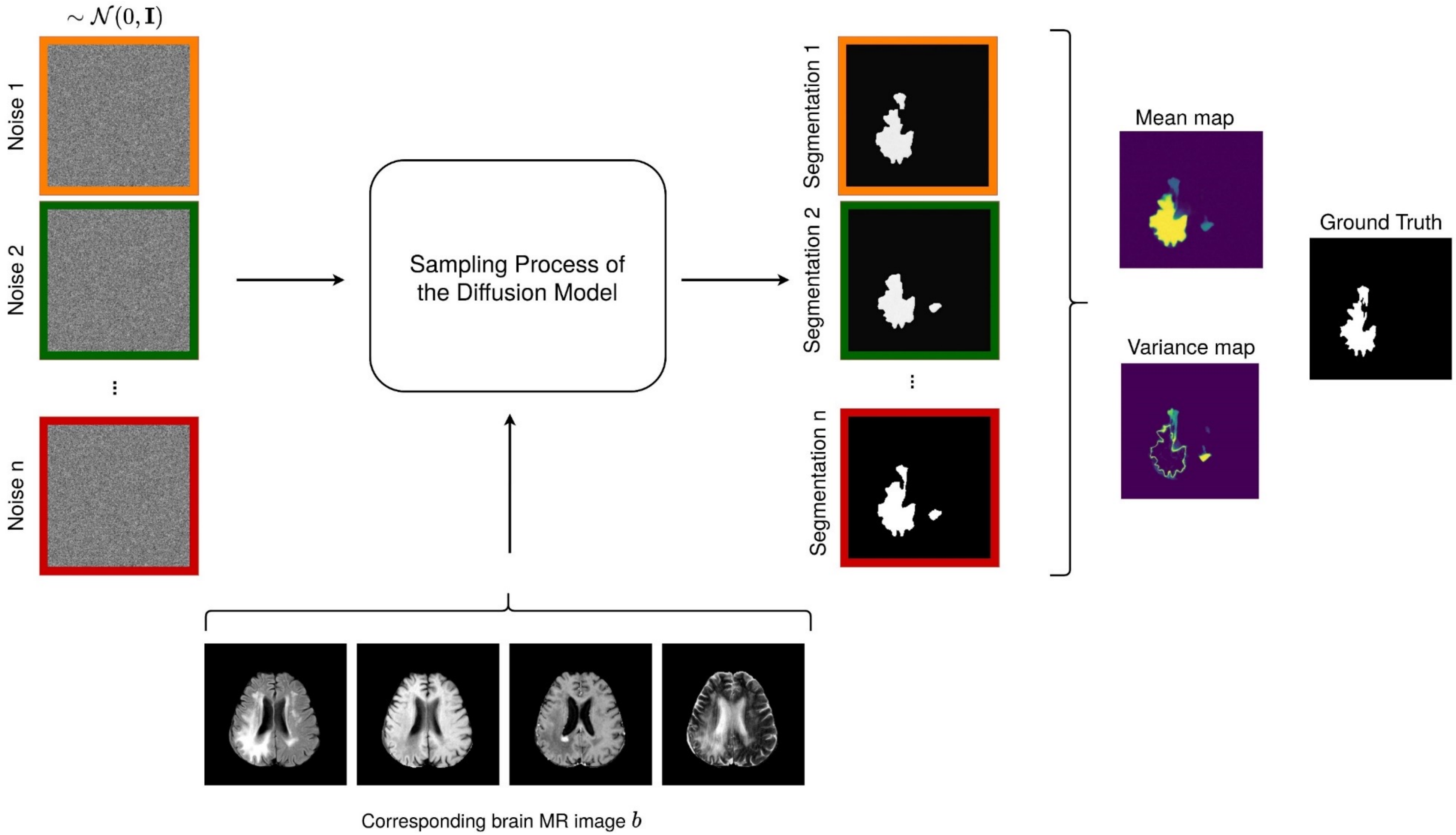


$$(*) \quad x_{b,t-1} = \frac{1}{\sqrt{\alpha_t}}\left(x_{b,t} - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(X_t, t)\right) + \sigma_t \mathbf{z}, \quad \text{with } \mathbf{z} \sim \mathcal{N}(0, \mathbf{I})$$

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

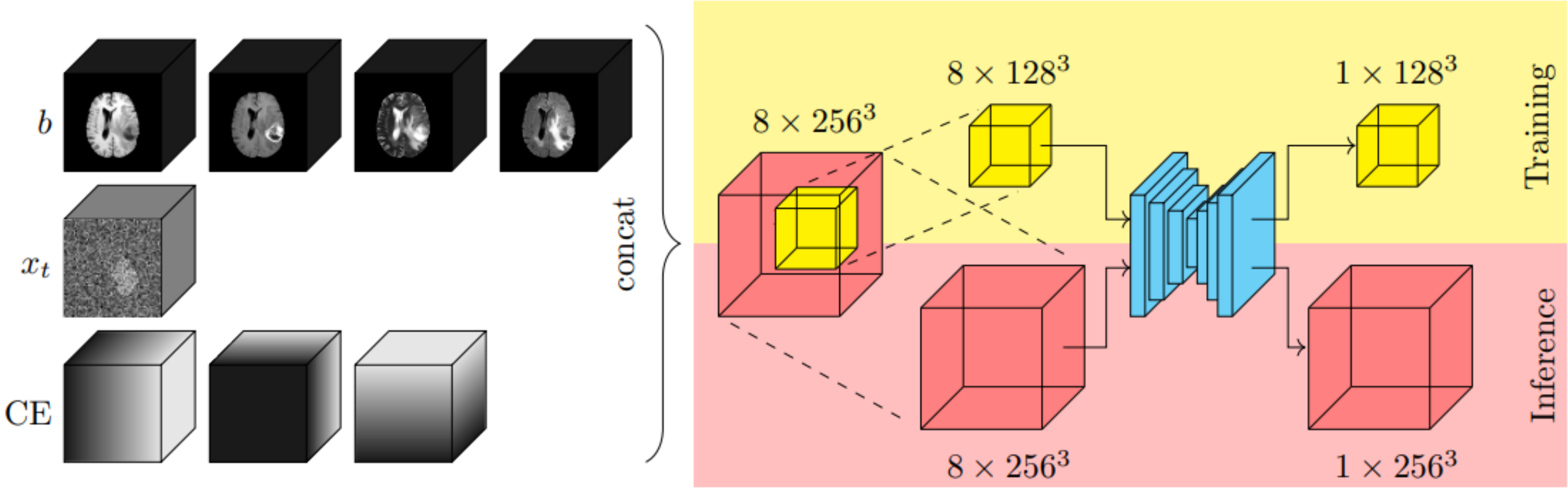
Wolleb et al (2022). Diffusion Models for Implicit Image Segmentation Ensembles, *MIDL 2022.* arXiv:2112.03145

Generation of Segmentation Ensembles



Wolleb et al (2022). Diffusion Models for Implicit Image Segmentation Ensembles, *MIDL 2022.* arXiv:2112.03145

3D Segmentation with PatchDDM

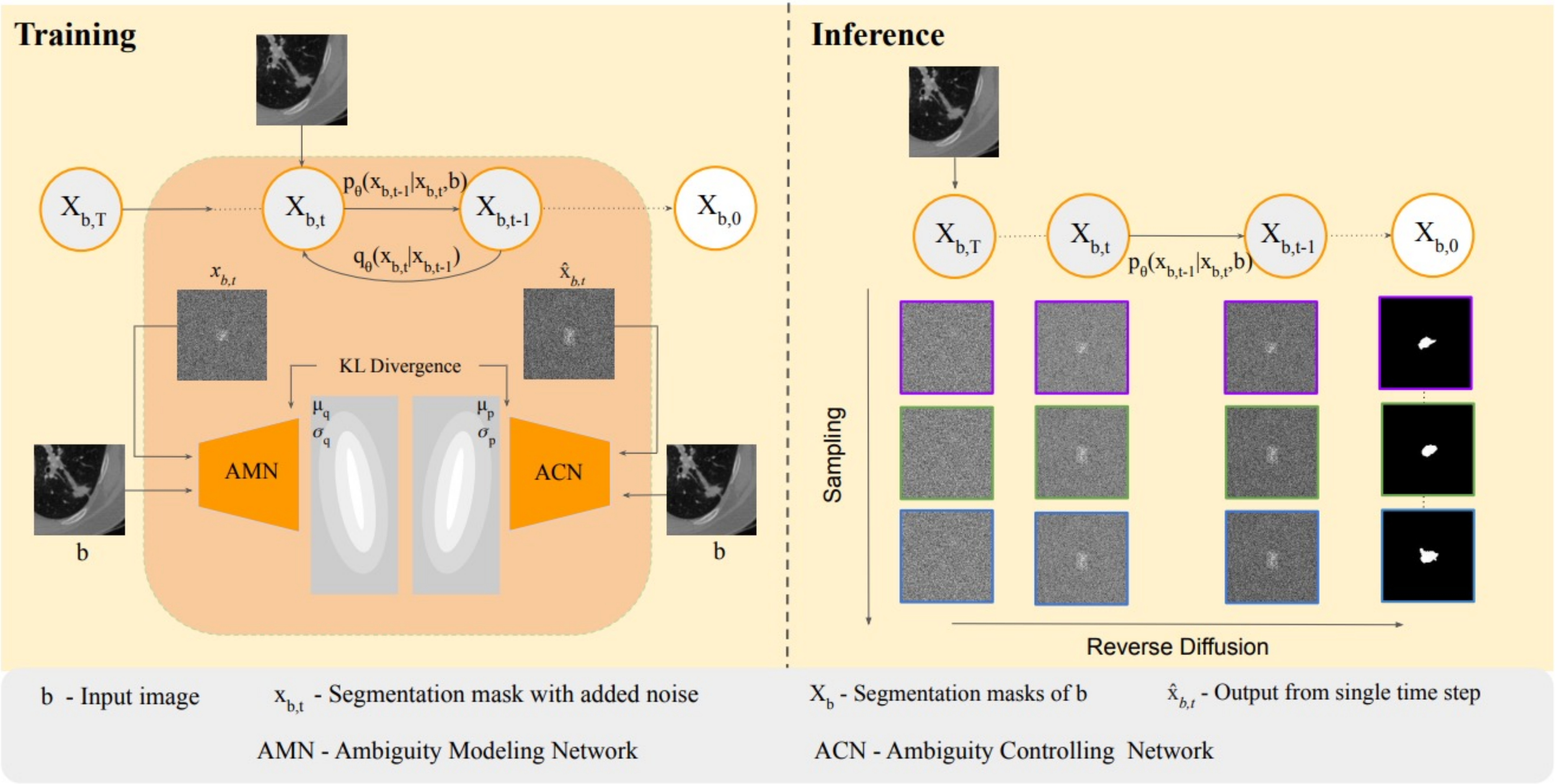


- We add a position encoding in all 3 spatial dimensions.
- Training is on patches only, and saves memory and training time.
- Inference runs over the whole 3D volume.

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

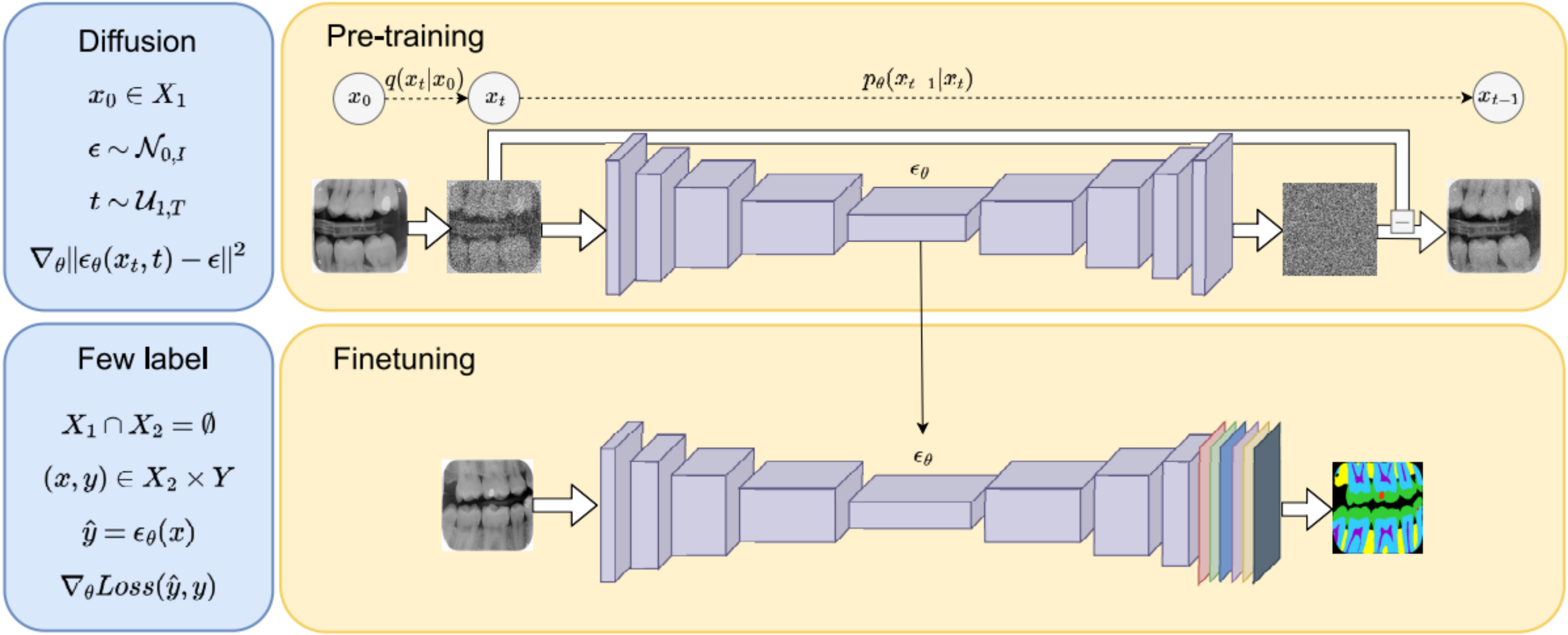
Ambiguous Segmentation

- 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.



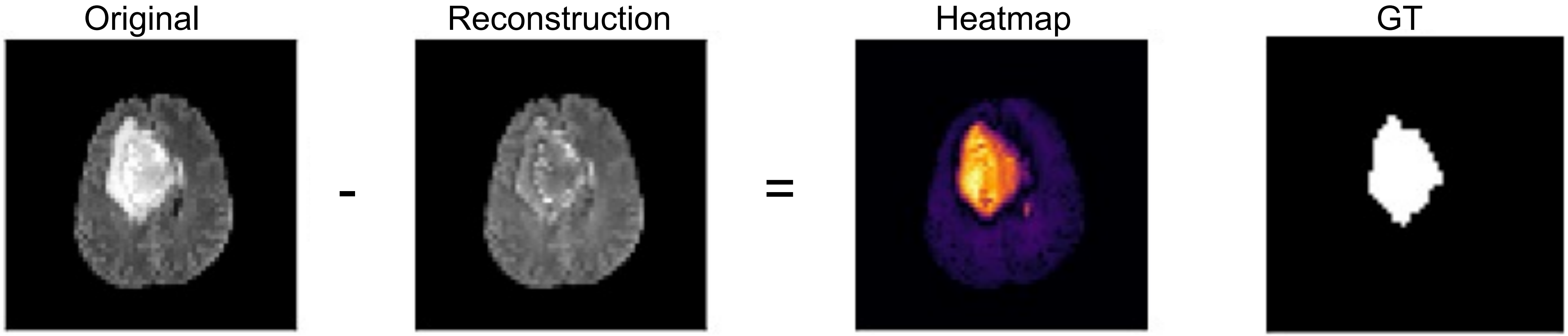
Rahman, Aimon, et al. (2023) Ambiguous medical image segmentation using diffusion models. CVPR

Segmentation with Diffusion Pre-training



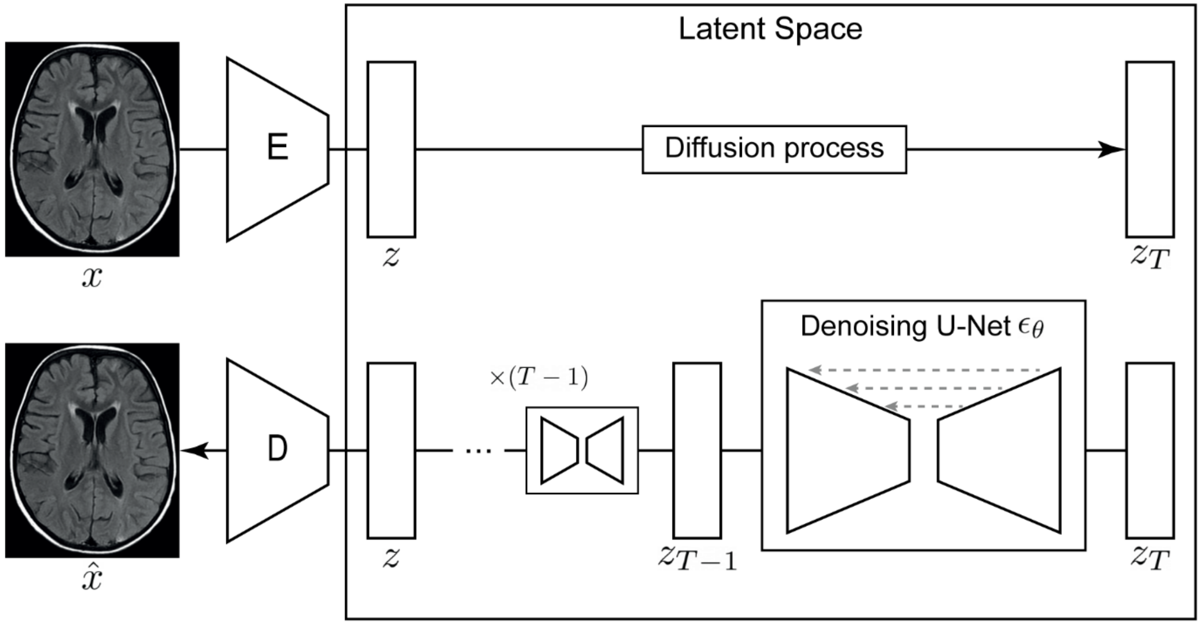
Anomaly Detection

Examples from the Community

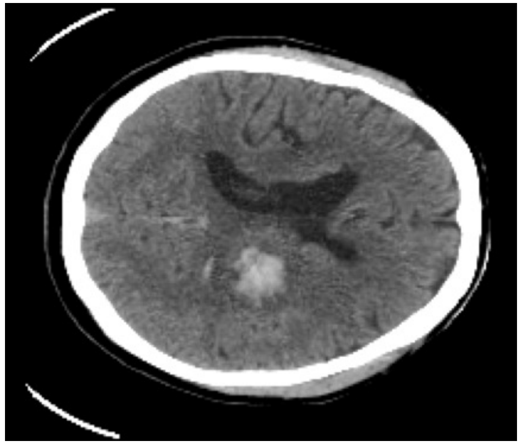


Unsupervised Anomaly Segmentation

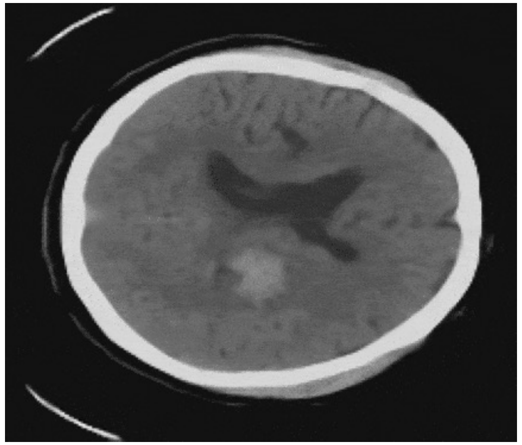
Latent Diffusion Model (LDM) learns the distribution of healthy brain data
Compression (VQ-VAE) scales for high-resolution images



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



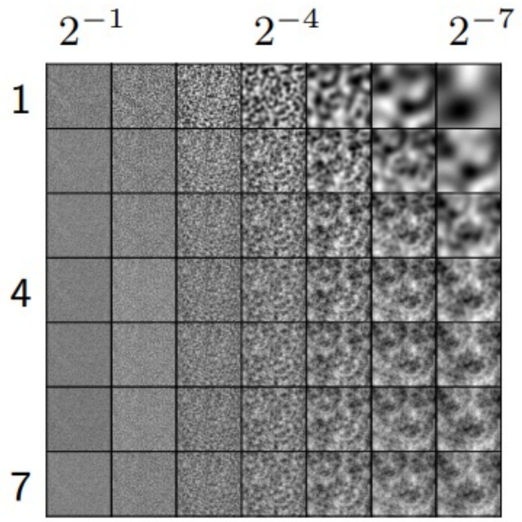
Reverse/denoising process is used to **inpaint** these regions and “**heal**” the possible anomalies



Pinaya et al (2022). Fast Unsupervised Brain Anomaly Detection and Segmentation with Diffusion Models. DOI:10.1007/978-3-031-16452-1_67

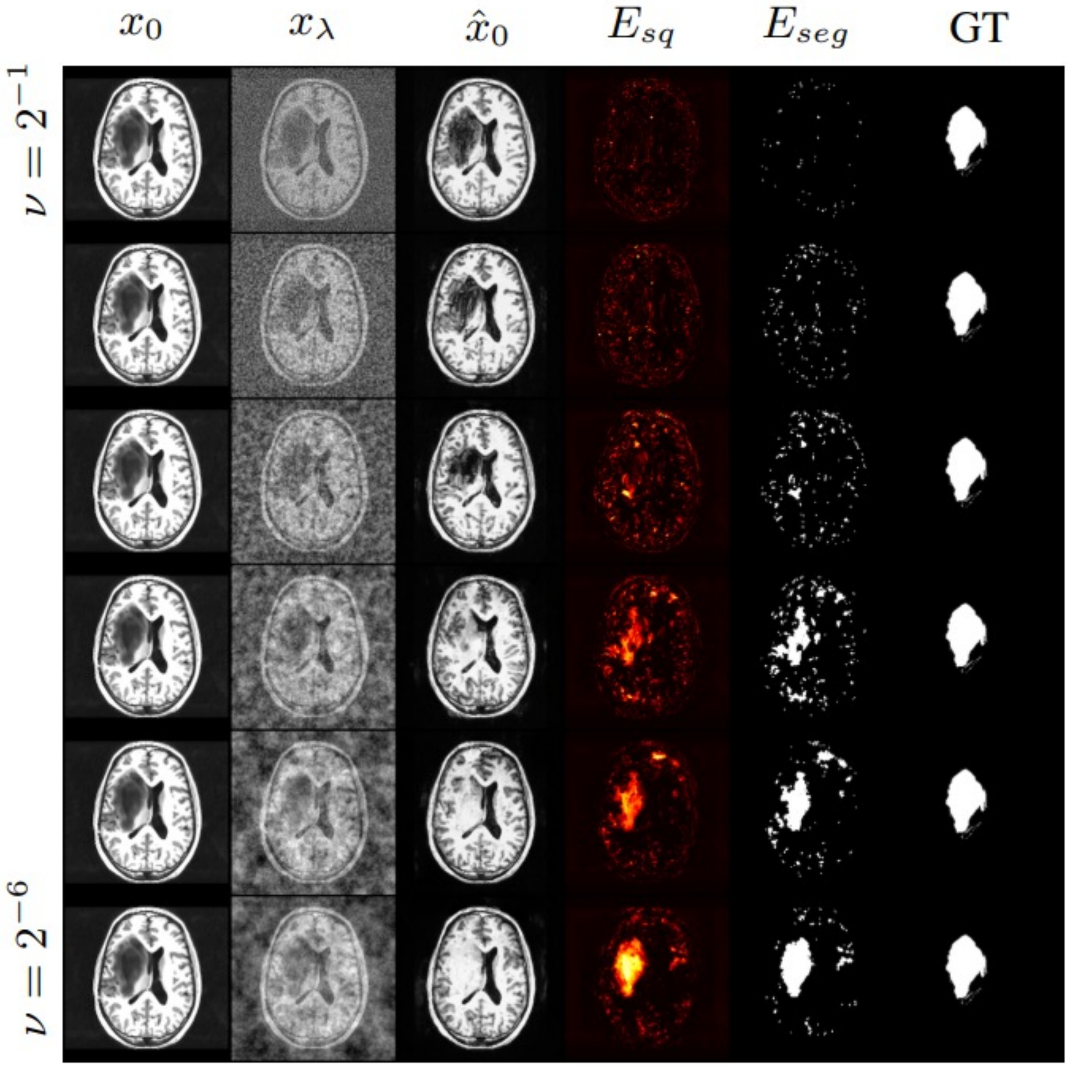
Anomaly Detection with Simplex Noise

- 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.



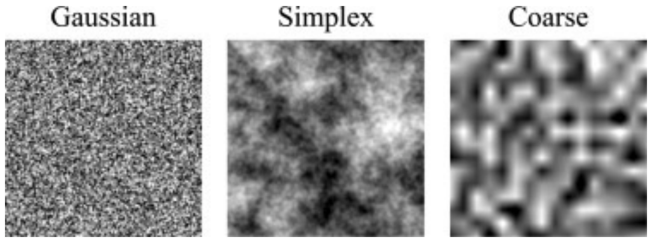
(a) Structures of simplex noise

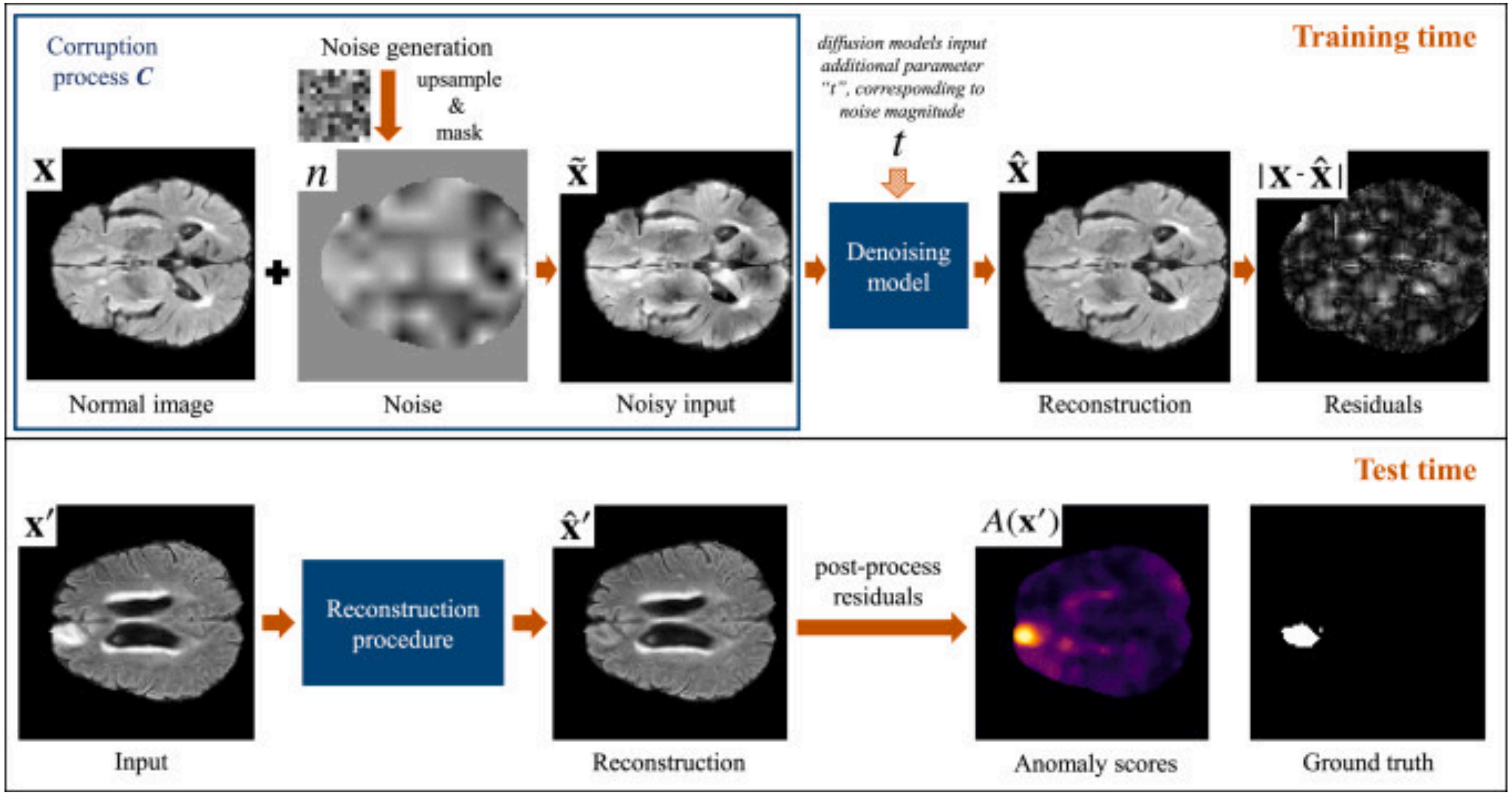
simplex noise scale *controls* target anomaly size



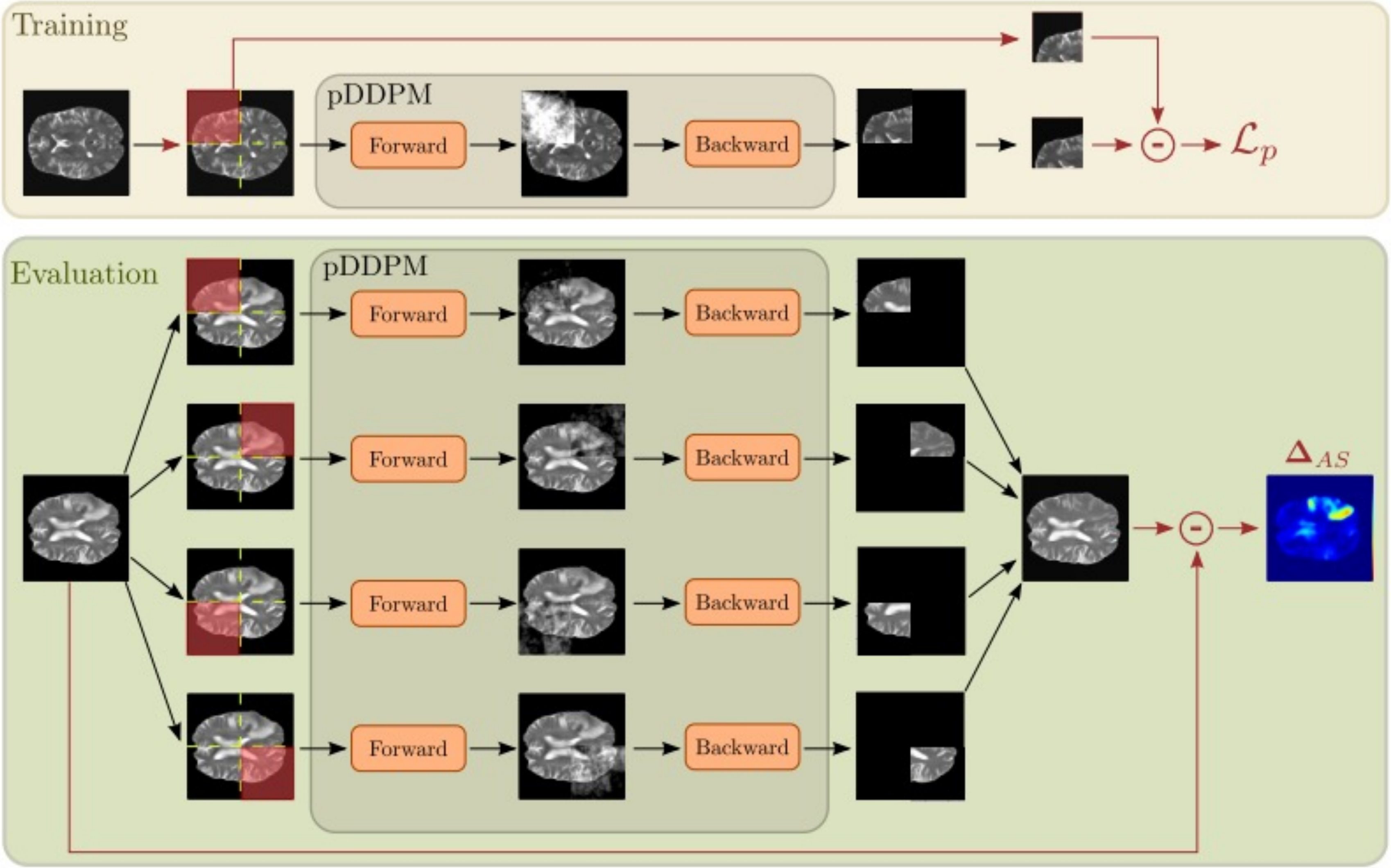
Wyatt et al (2022) AnoDDPM: Anomaly Detection with Denoising Diffusion Probabilistic Models using Simplex Noise. CVPR workshop

Anomaly Detection with Coarse Noise



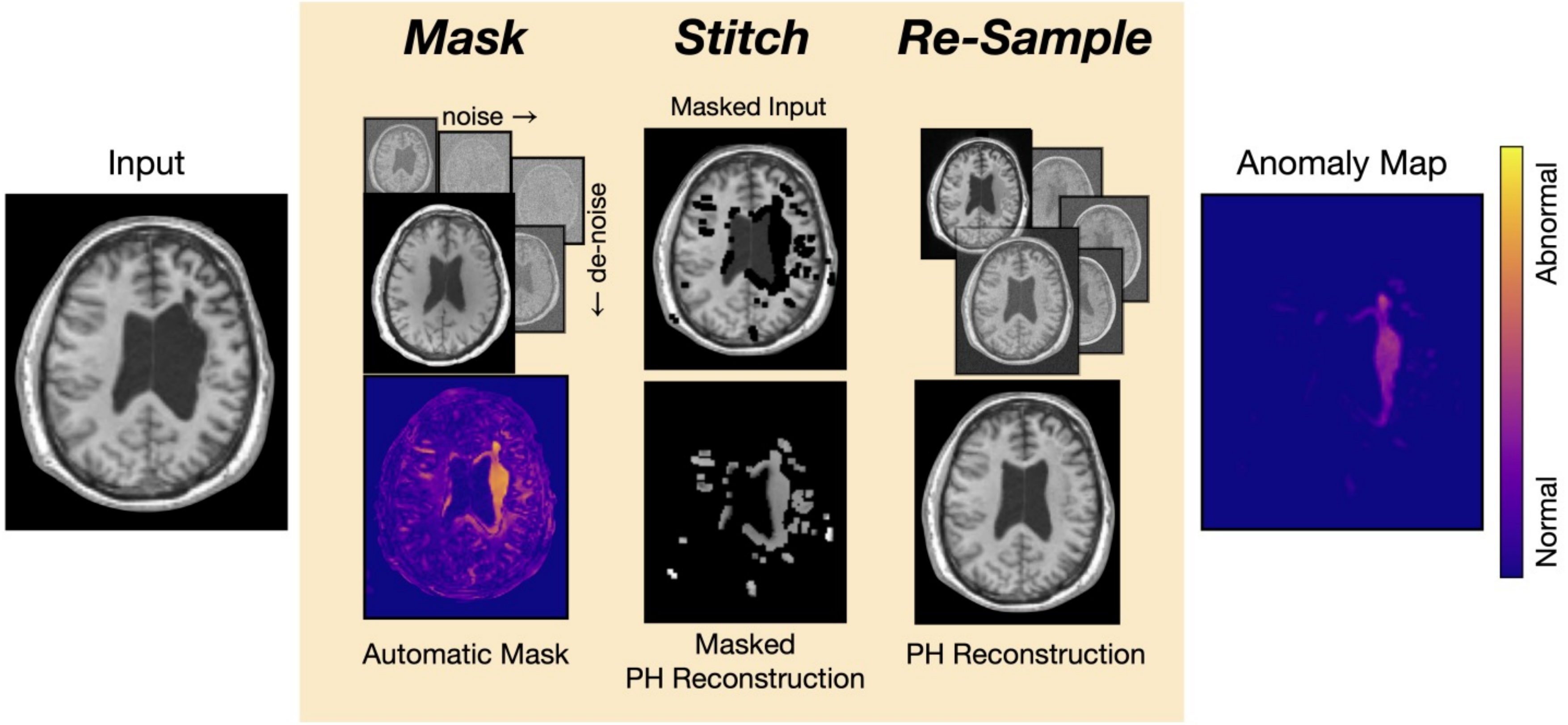


Anomaly Detection from Patches



Behrendt, Finn, et al. (2023) "Patched diffusion models for unsupervised anomaly detection in brain mri." Medical Imaging with Deep Learning

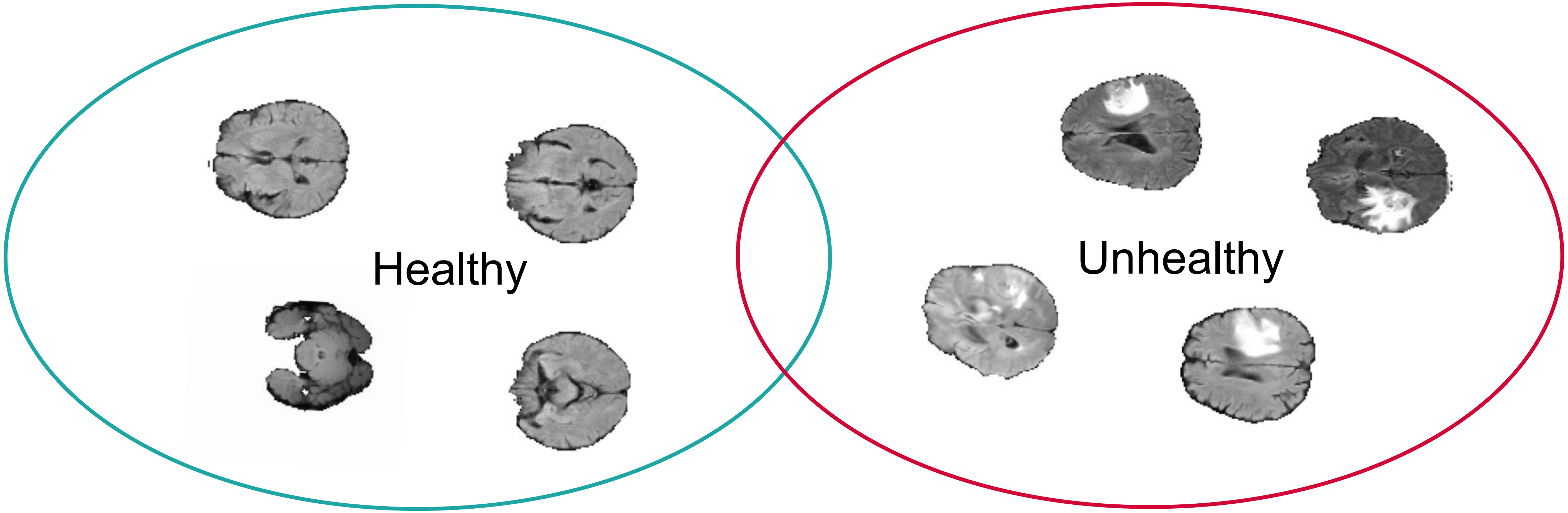
Mask, Stitch, and Re-Sample



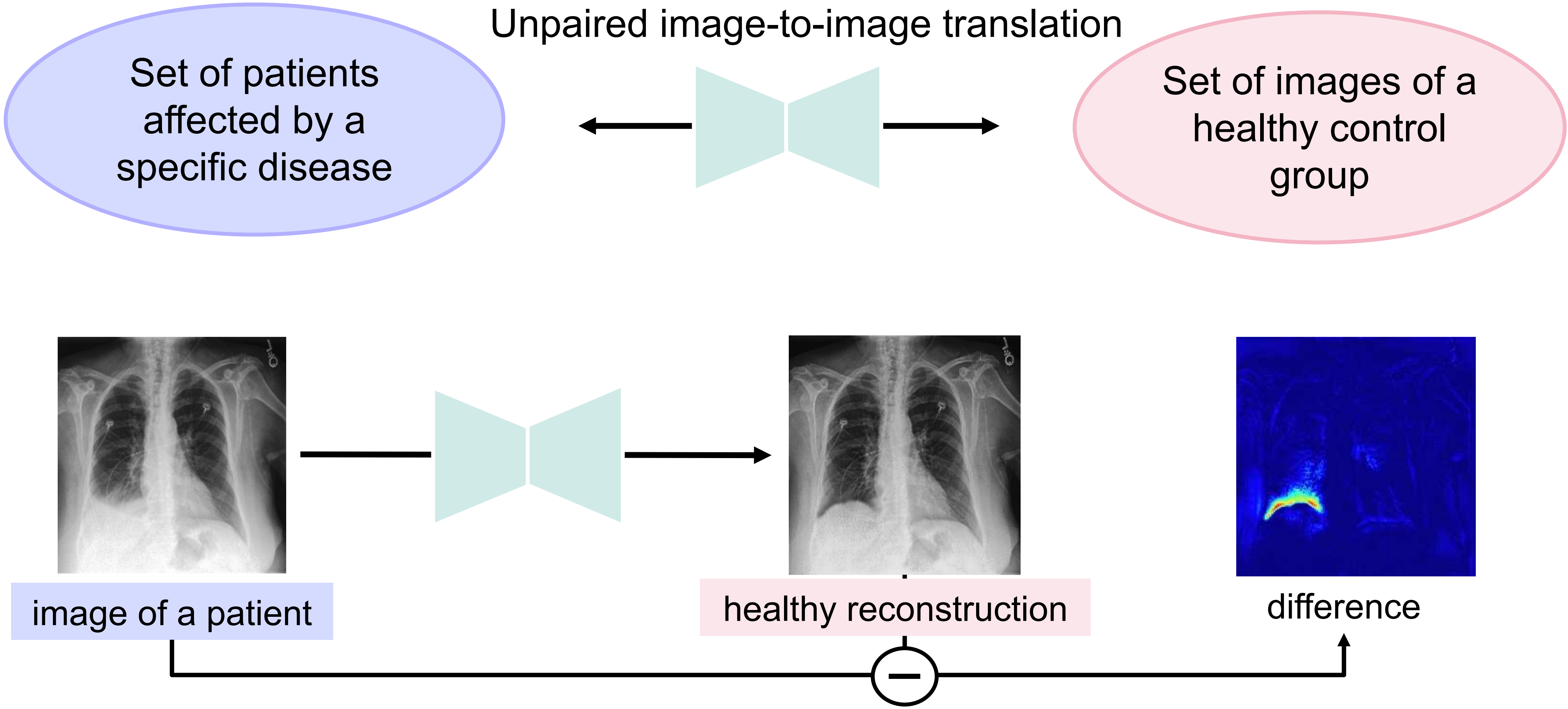
Bercea, C et al. (2023). Mask, Stitch, and Re-Sample: Enhancing Robustness and Generalizability in Anomaly Detection through Automatic Diffusion Models. *arXiv preprint arXiv:2305.19643*.

Weakly Supervised Lesion Detection

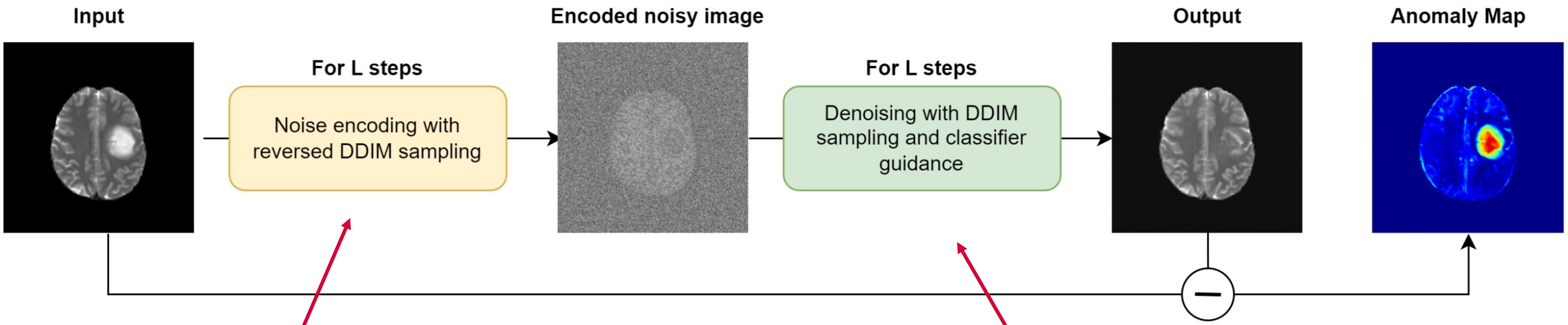
Goal: Pixel-wise anomaly detection using image-level labels only



Weakly Supervised Lesion Detection

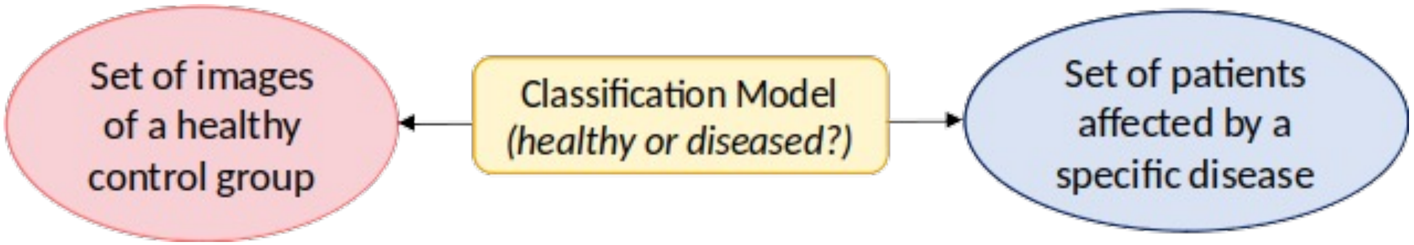


Weakly Supervised Lesion Detection

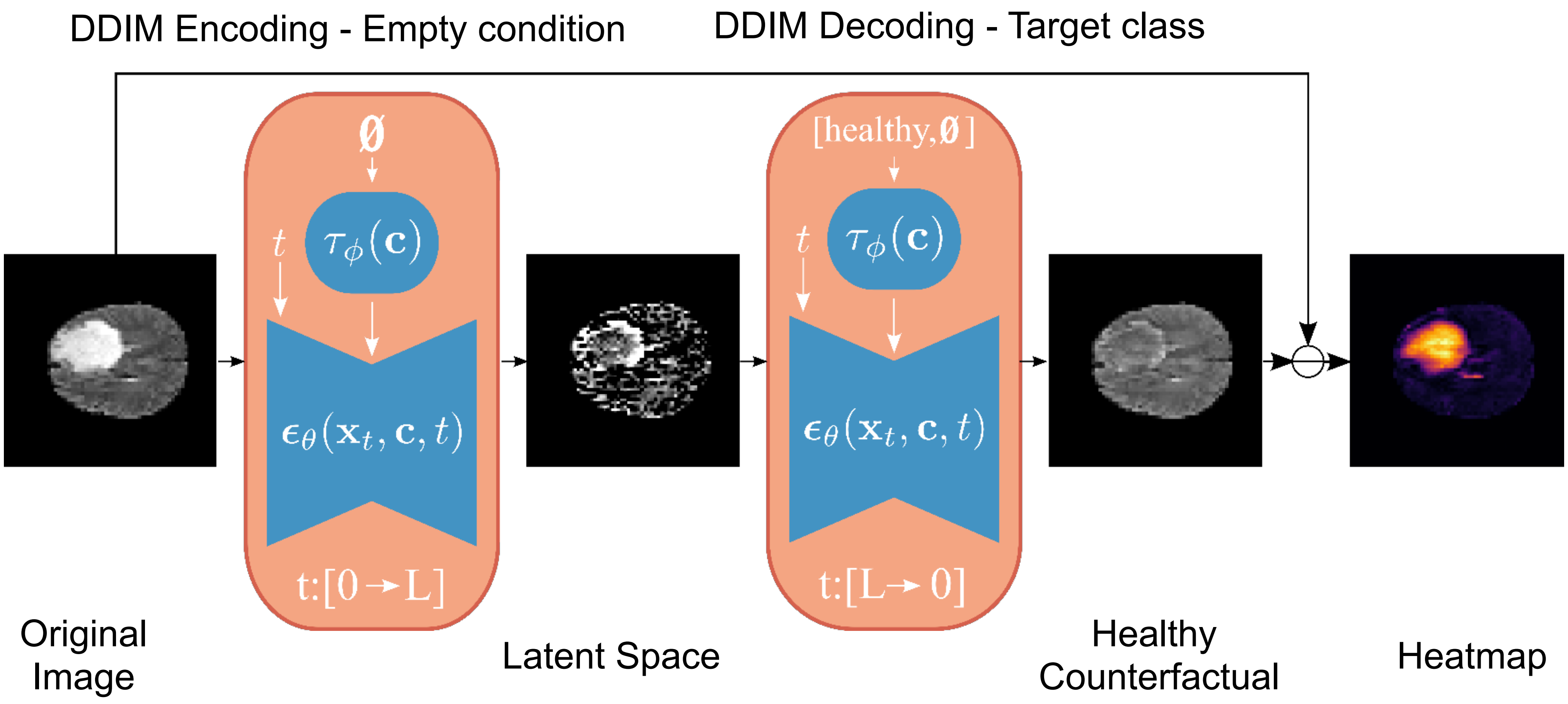


DDIM

Gradient Guidance



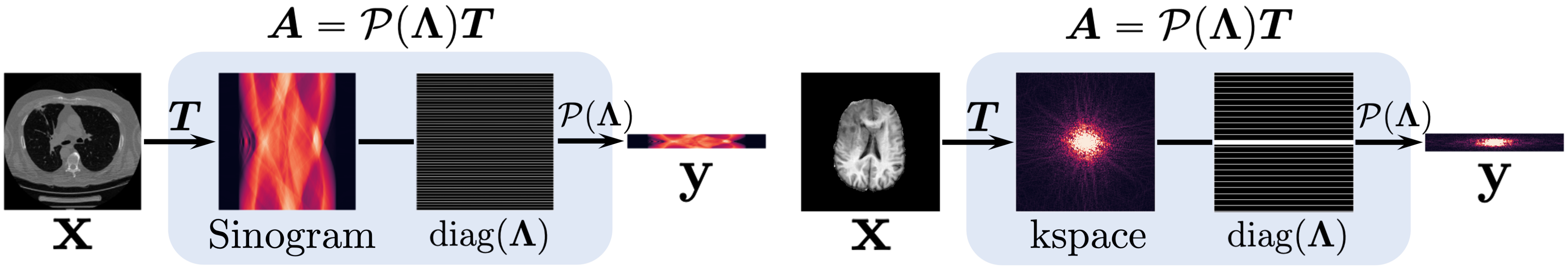
Lesion Localization with Classifier-free Guidance



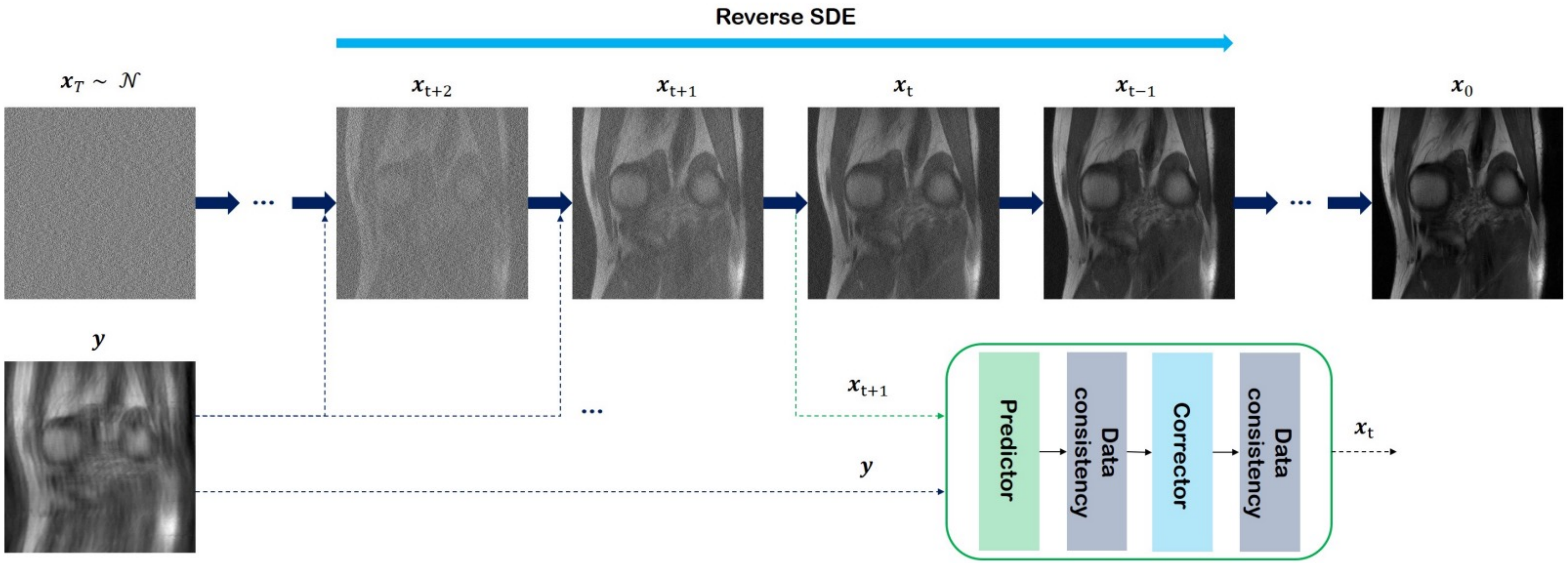
Sanchez et al (2022). What is Healthy? Generative Counterfactual Diffusion for Lesion Localization. DGM4*MICCAI 2022.* arXiv:2207.12268

Image Reconstruction

Examples from the Community

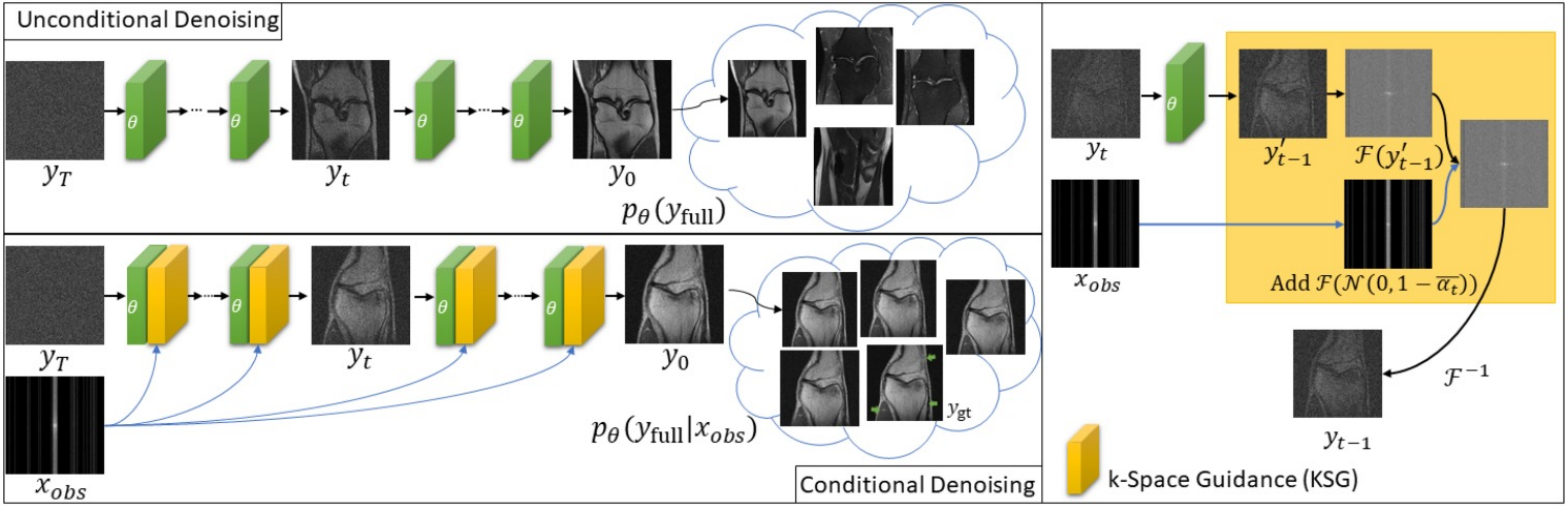


Score-based Diffusion Models for Accelerated MRI



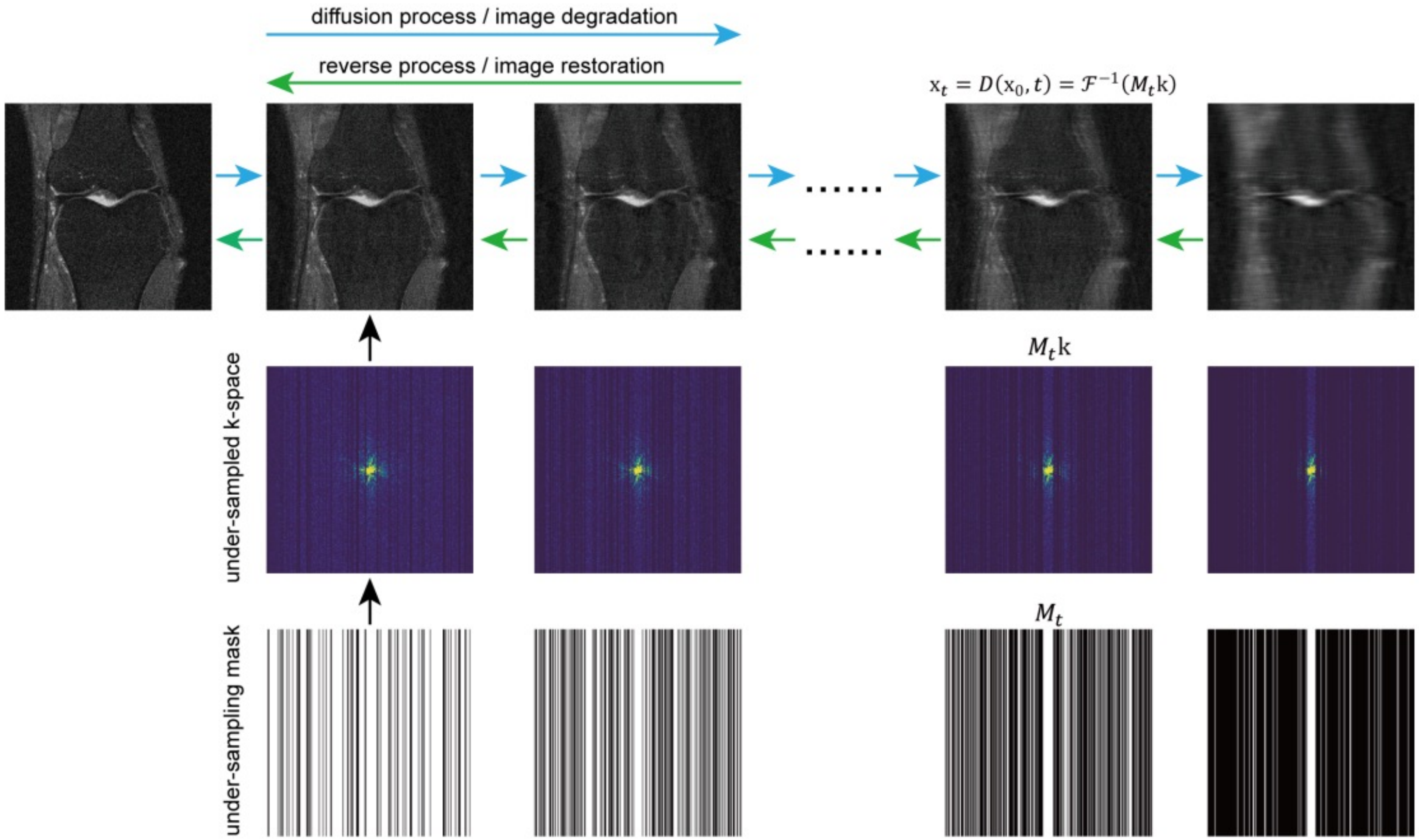
Chung et al. (2021) Score-based diffusion models for accelerated MRI. MIA 2021

Undersampled MR Reconstruction via Diffusion Model Sampling



Peng, C., Guo, P., Zhou, S. K., Patel, V. M., & Chellappa, R. (2022). Towards performant and reliable undersampled MR reconstruction via diffusion model sampling. *MICCAI 2022*

K-space Cold Diffusion

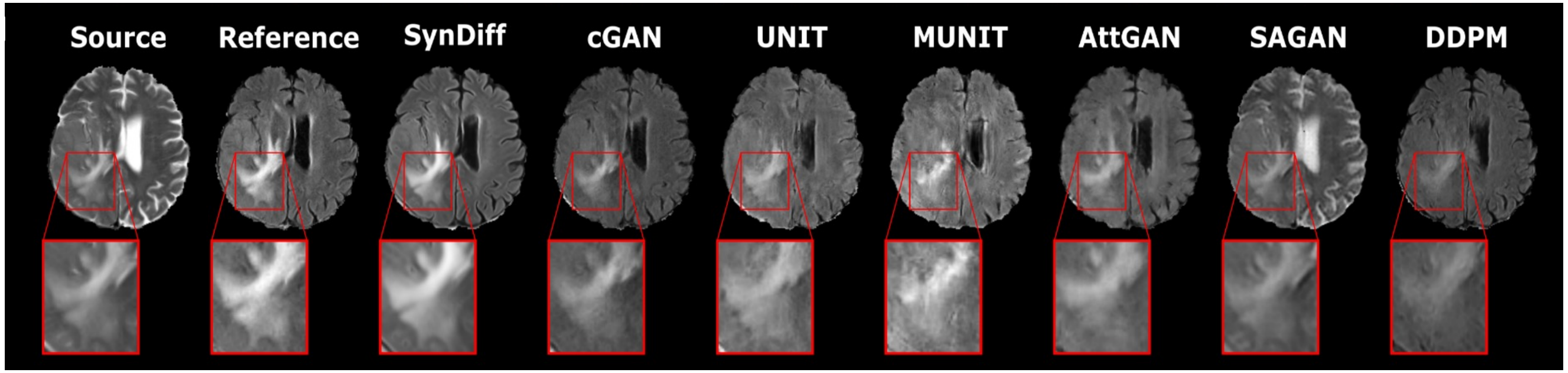


Shen, G., Li, M., Farris, C. W., Anderson, S., & Zhang, X. (2023). K-space Cold Diffusion: Learning to Reconstruct Accelerated MRI without Noise. *arXiv preprint arXiv:2311.10162*.

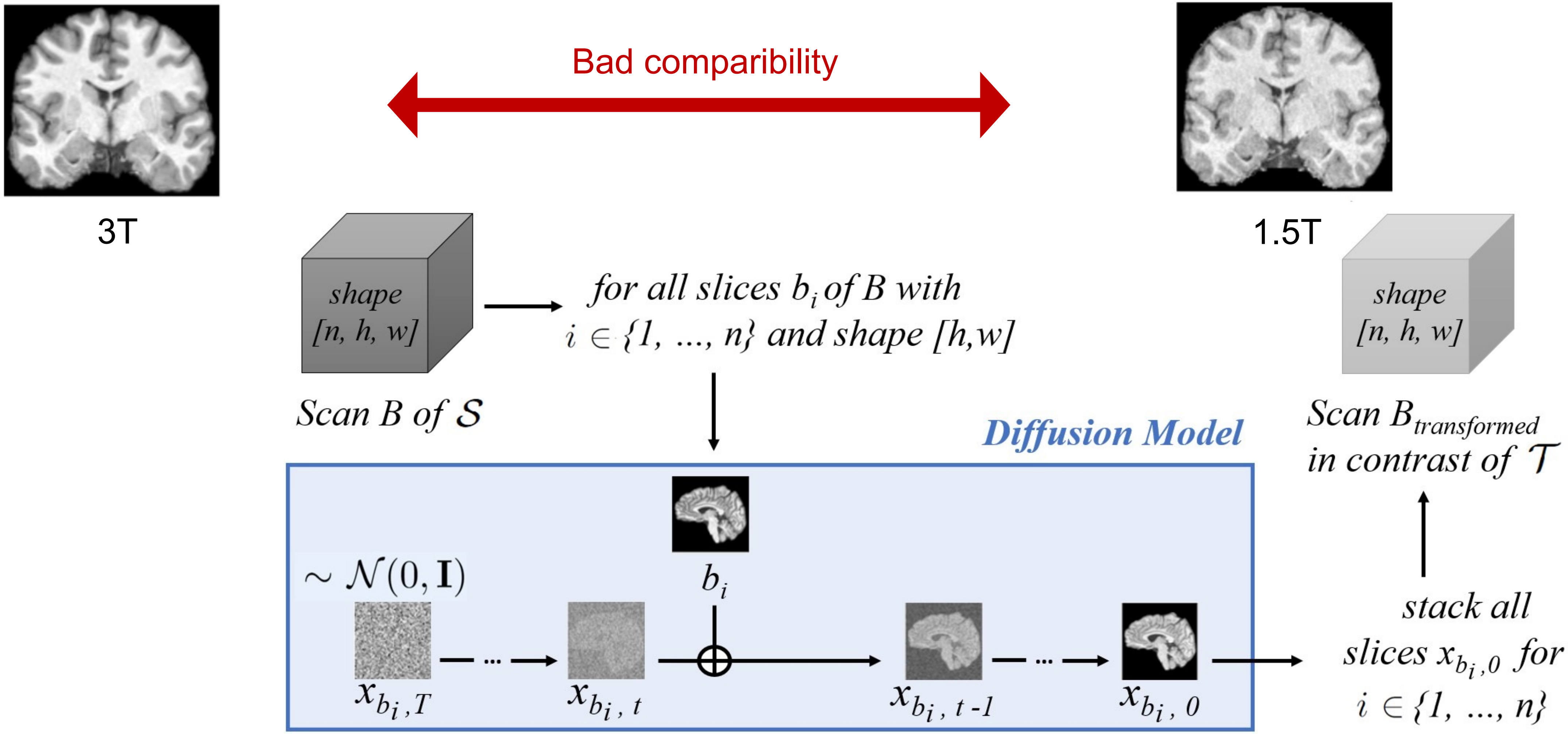
Image-to-Image Translation

Examples from the Community

MRI Contrast Translation

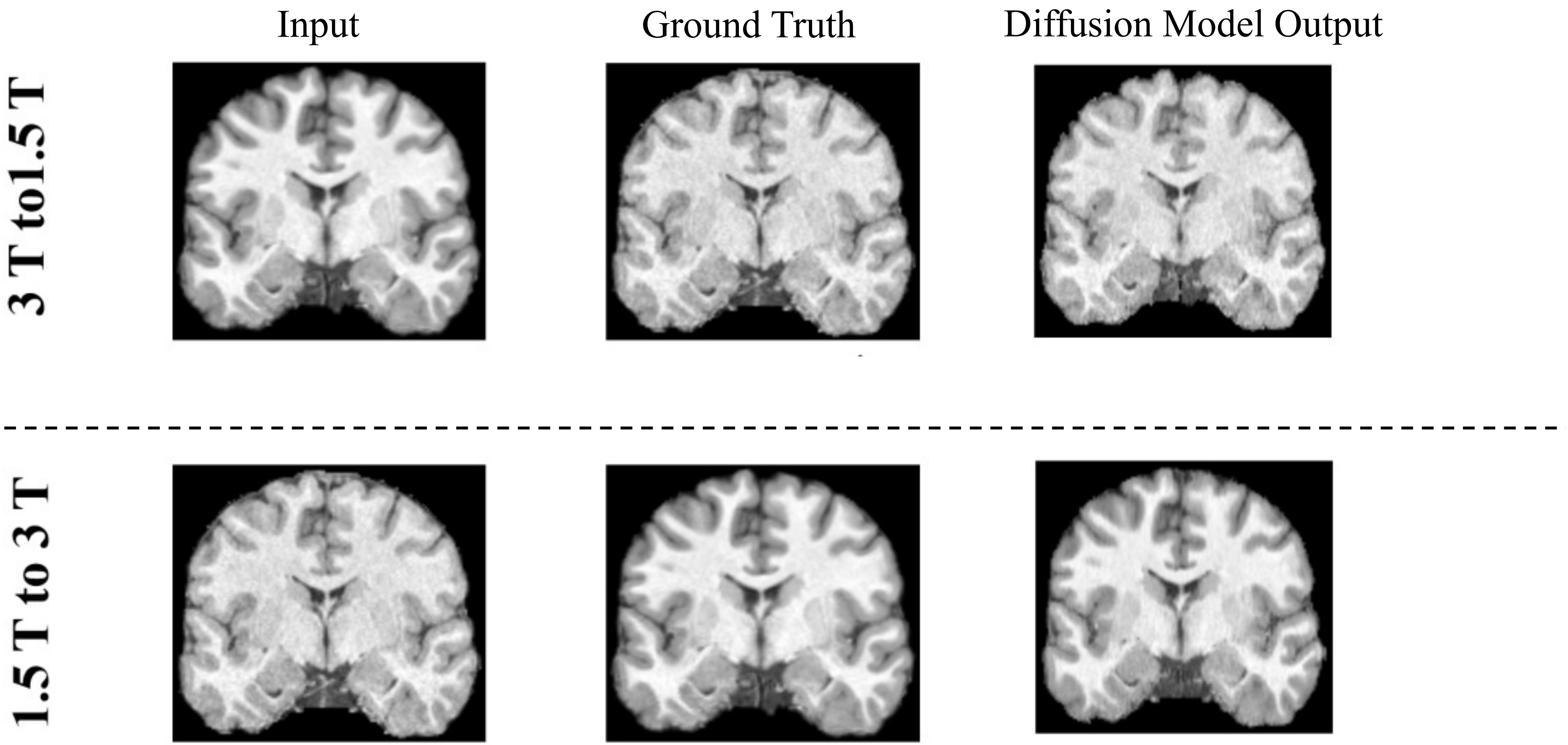


Diffusion Models for Contrast Harmonization



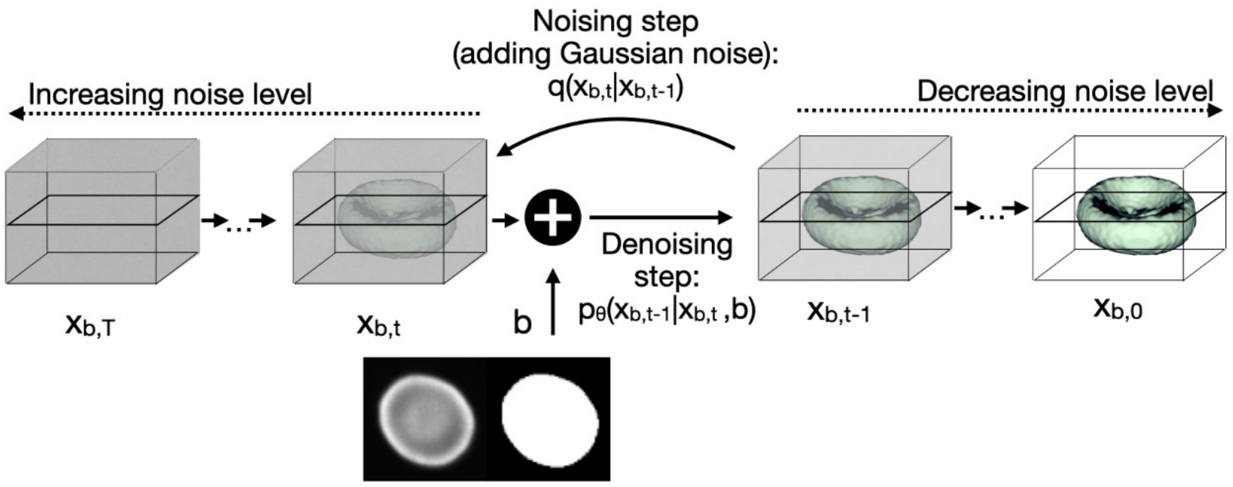
Durrer, Alicia, et al. "Diffusion Models for Contrast Harmonization of Magnetic Resonance Images." *arXiv preprint arXiv:2303.08189* (2023).

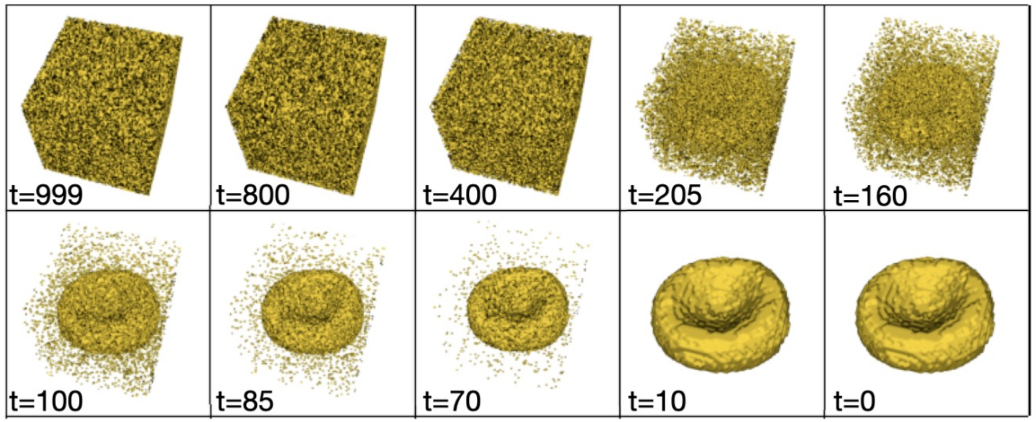
Contrast Harmonization Results

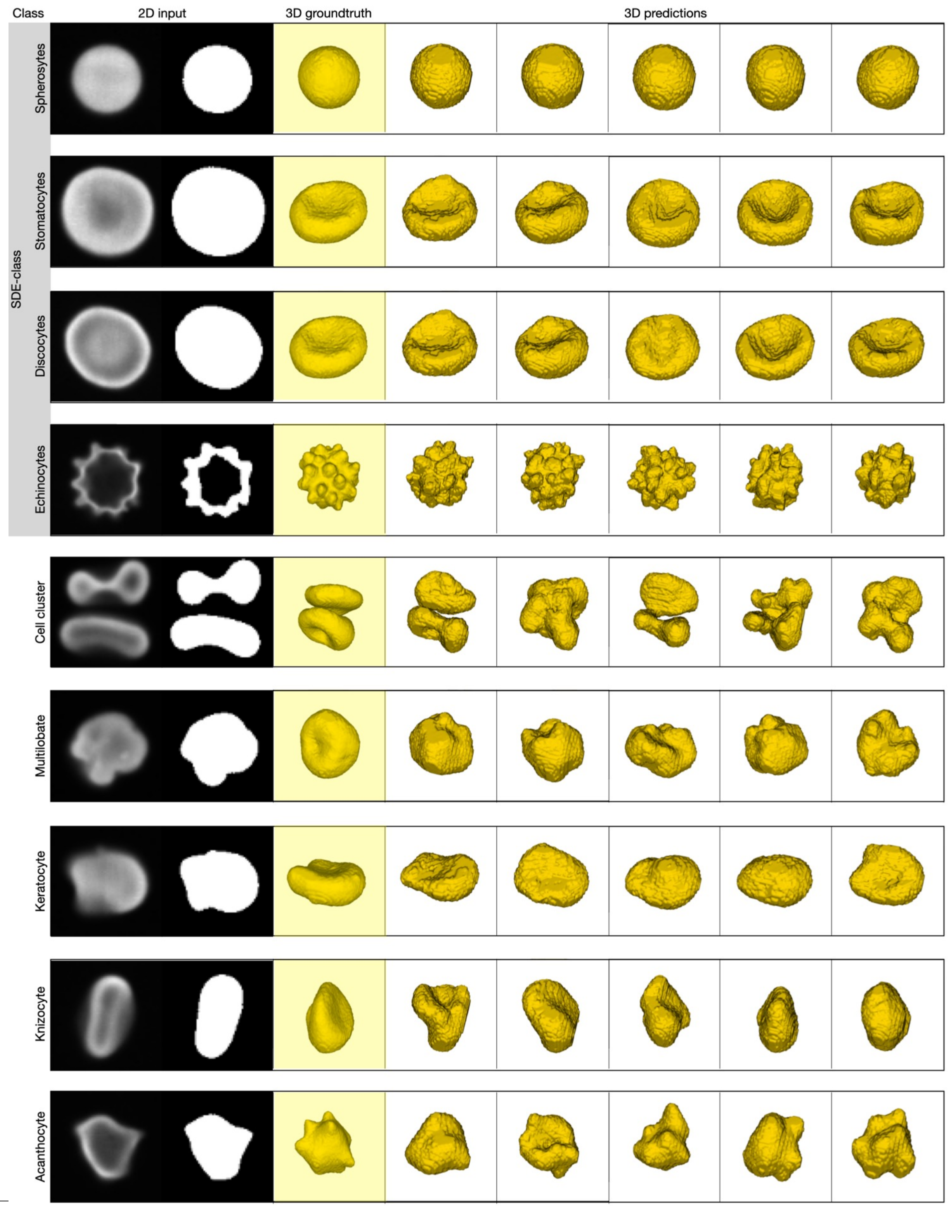


Durrer, Alicia, et al. "Diffusion Models for Contrast Harmonization of Magnetic Resonance Images." *arXiv preprint arXiv:2303.08189* (2023).

3D Shapes from 2D Microscopy

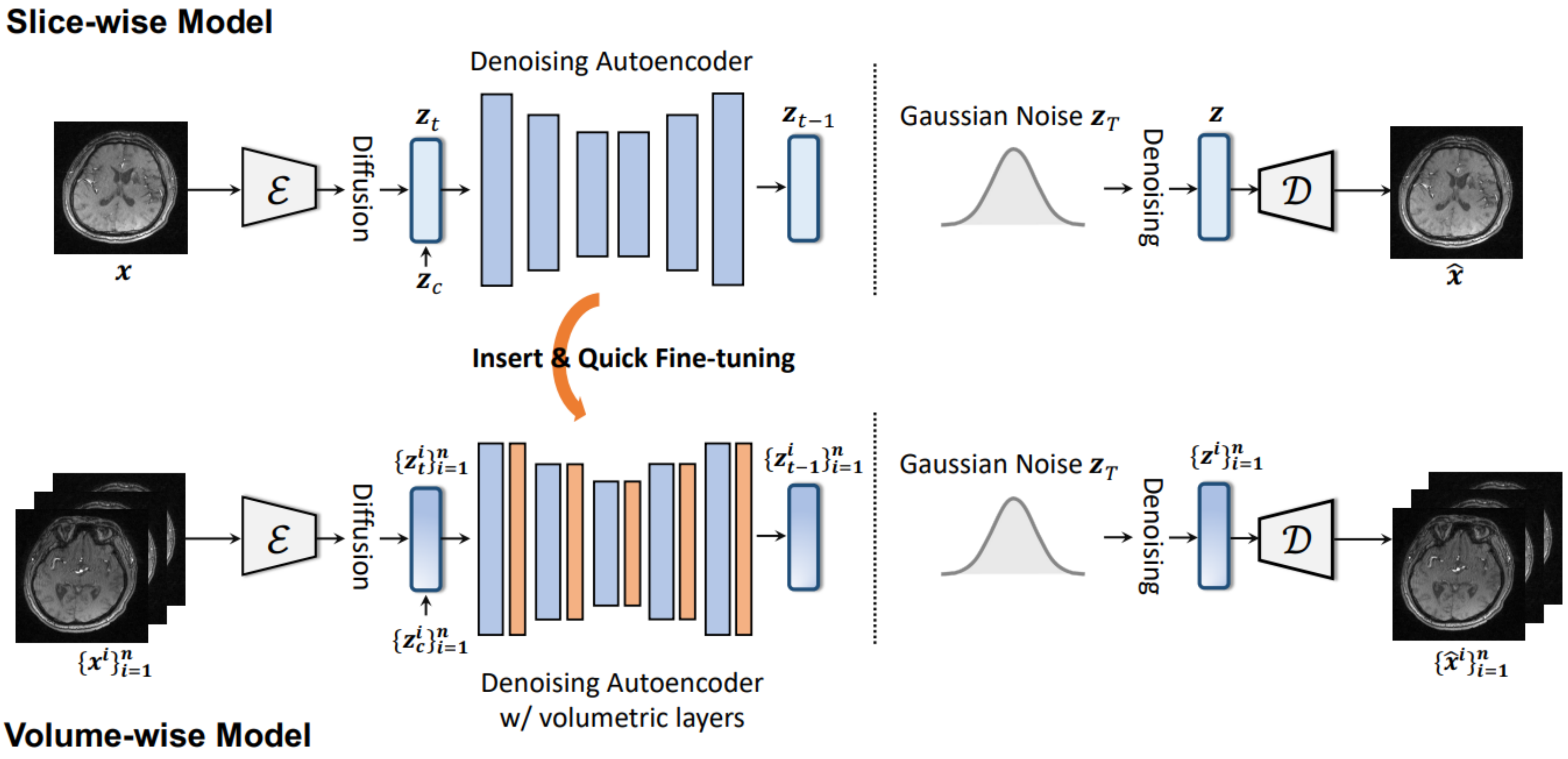






Waibel, D. J., Röell, E., Rieck, B., Giryes, R., & Marr, C. (2023, April). A diffusion model predicts 3d shapes from 2d microscopy images. In *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)* (pp. 1-5). IEEE.

3D with 2D model

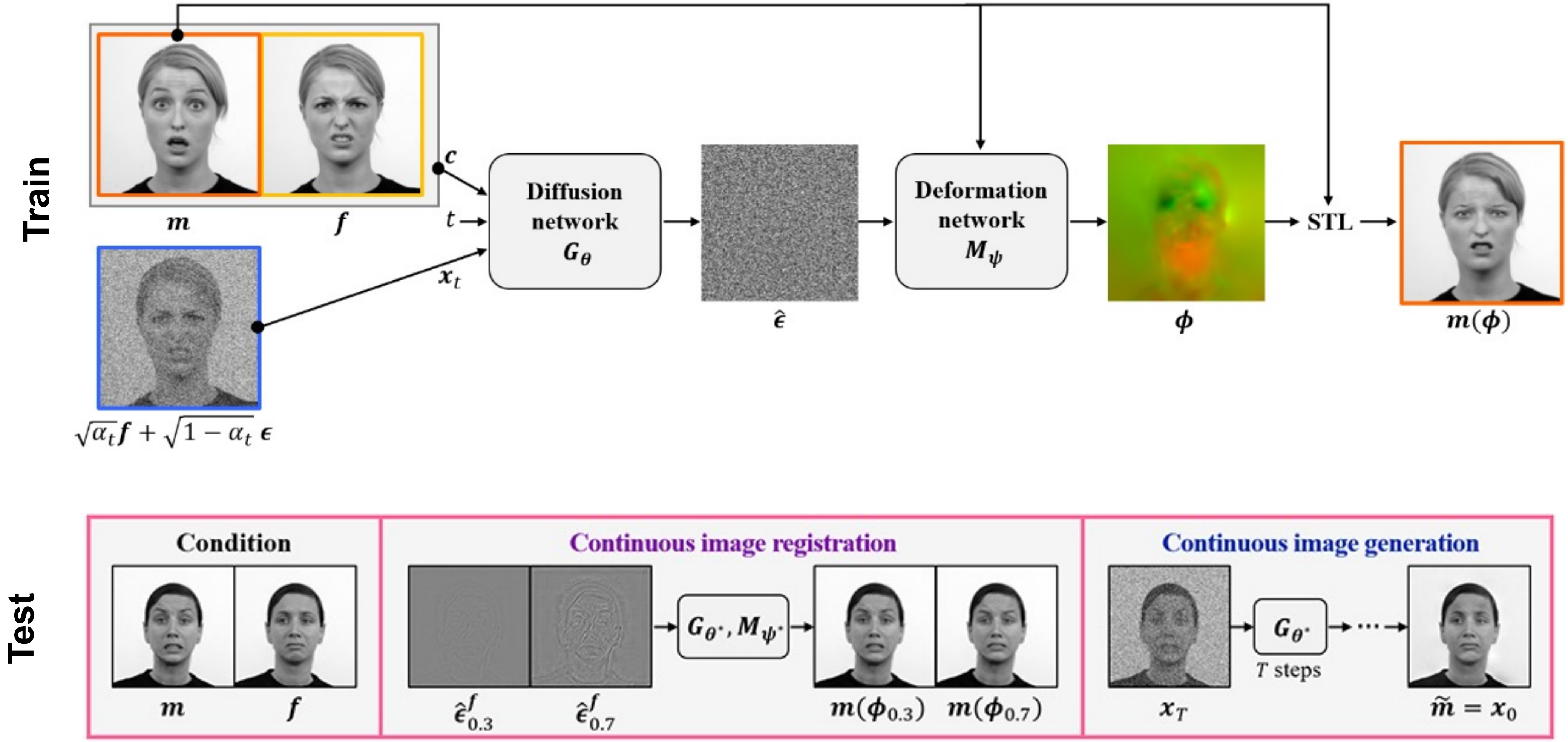


Zhu, Lingting, et al. (2023) Make-A-Volume: Leveraging Latent Diffusion Models for Cross-Modality 3D Brain MRI Synthesis. MICCAI

Image Registration

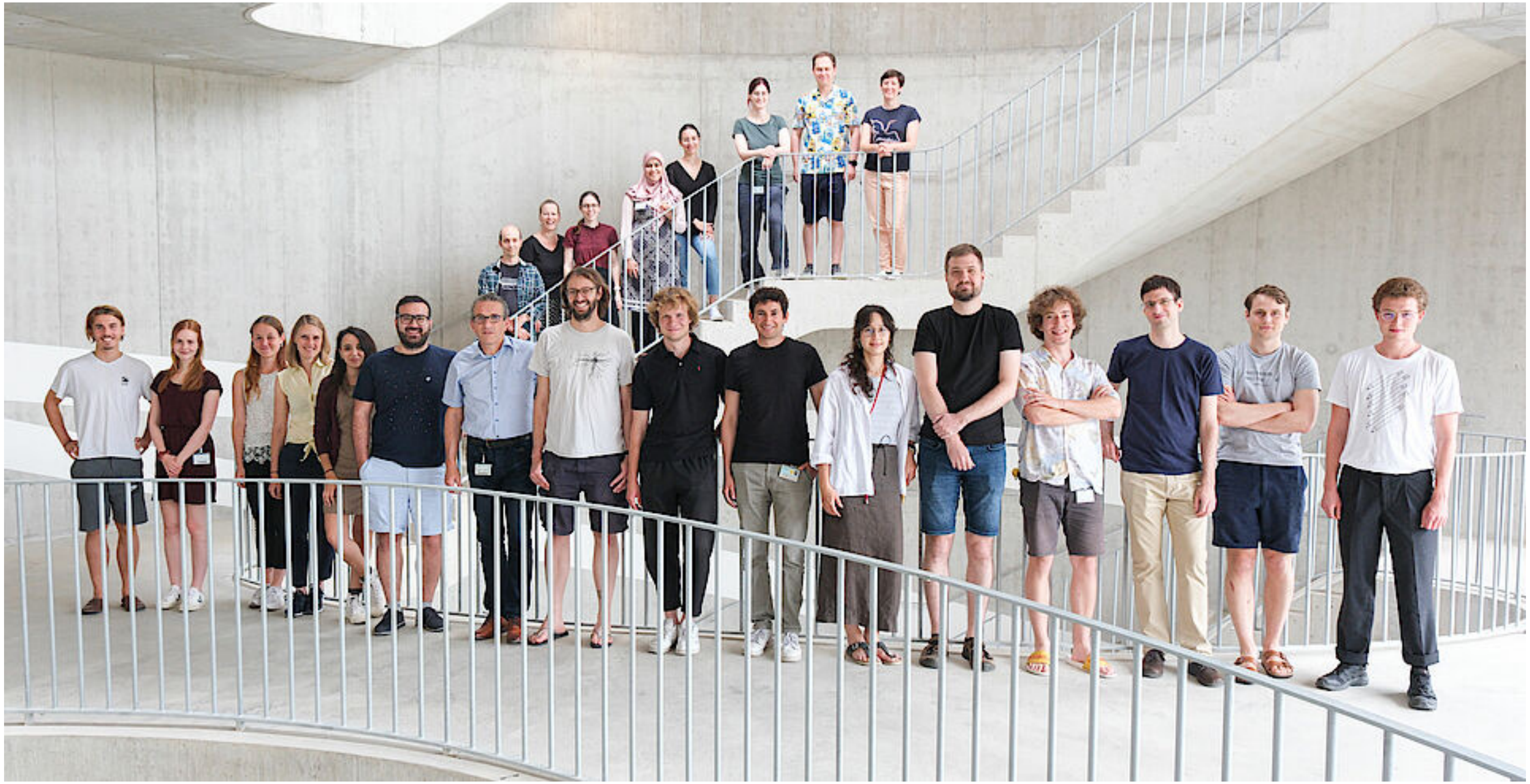
Examples from the Community

DiffuseMorph



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

Thank you!



Center for medical Image Analysis & Navigation (CIAN), Prof. Philippe C. Cattin
Universität Basel

Useful Key References, Gits to Watch, etc.

Surveys

- https://arxiv.org/abs/2209.02646
- https://arxiv.org/abs/2209.00796

Github

- https://github.com/heejkoo/Awesome-Diffusion-Models

Tutorials

- https://cvpr2022-tutorial-diffusion-models.github.io
- https://huggingface.co/blog/annotated-diffusion
- https://huggingface.co/docs/diffusers