Tutorial on

Diffusion Models for Medical Imaging

Pedro P. Sanchez

Dr. Walter Pinaya

Dr. Julia Wolleb

Prof. Dorit Merhof

Prof. Jorge Cardoso

Prof. Sotirios A. Tsaftaris





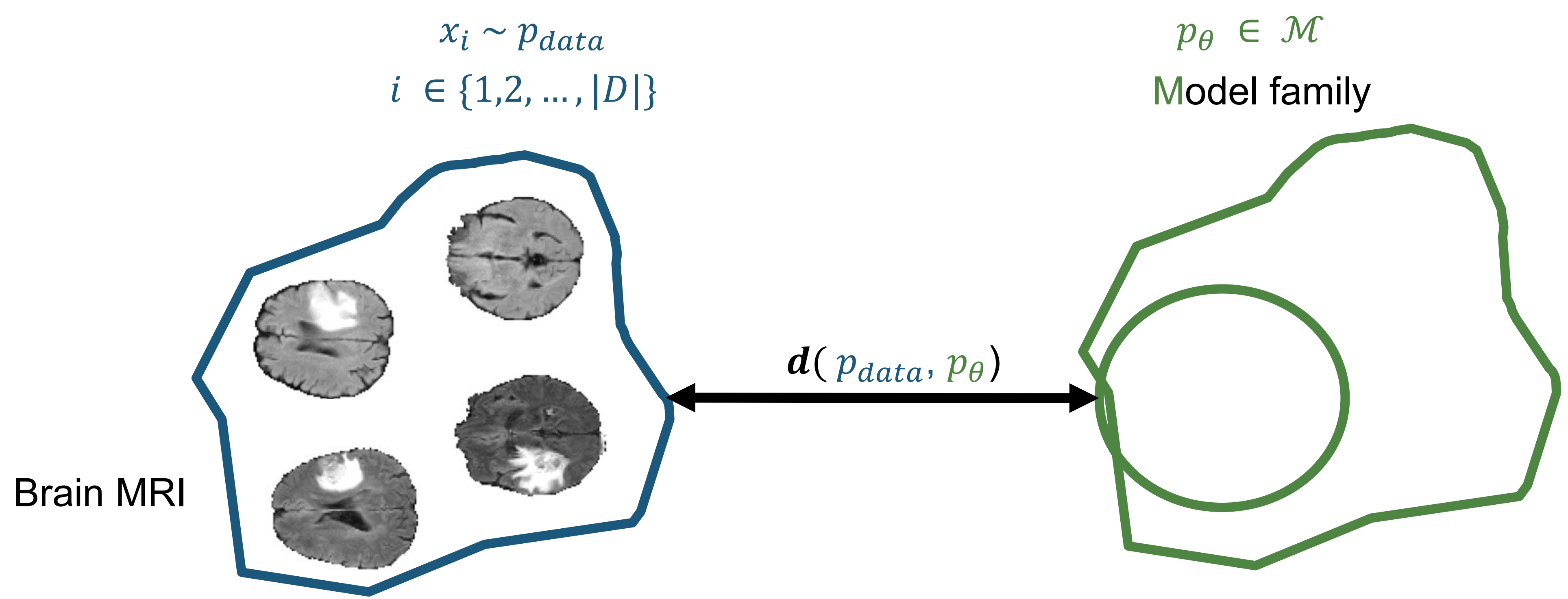

Tutorial Schedule

- ❑ Introduction [1:30pm-2:30pm]
  - ❑ What? Why? How?
  - ❑ Understanding and Intuition
  - ❑ DEMO - MONAI Generative Models
  - Coding tutorial on DDPM
- ❑ Advanced Topics [2:30pm-3:30pm]
  - ❑ Sampling Strategies
  - ❑ Inference-time Conditioning
  - ❑ Training-time Conditioning
  - ❑ DEMO - MONAI Generative Models
  - DDIM Inversion + Classifier-free guidance
- ❑ Applications in Medical Imaging [4pm-5pm]
  - ❑ Synthesis
  - ❑ Reconstruction
  - ❑ Segmentation
  - ❑ Registration
  - ❑ Inpainting
  - ❑ Anomaly Detection
  - ❑ Miscellaneous
- ❑ Panel Discussion [5pm-6pm]

Diffusion Models

What? Why? How?

What? Generative Models



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

What? Generative Models

Density Estimation

$p_\theta(x)$

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

Model family

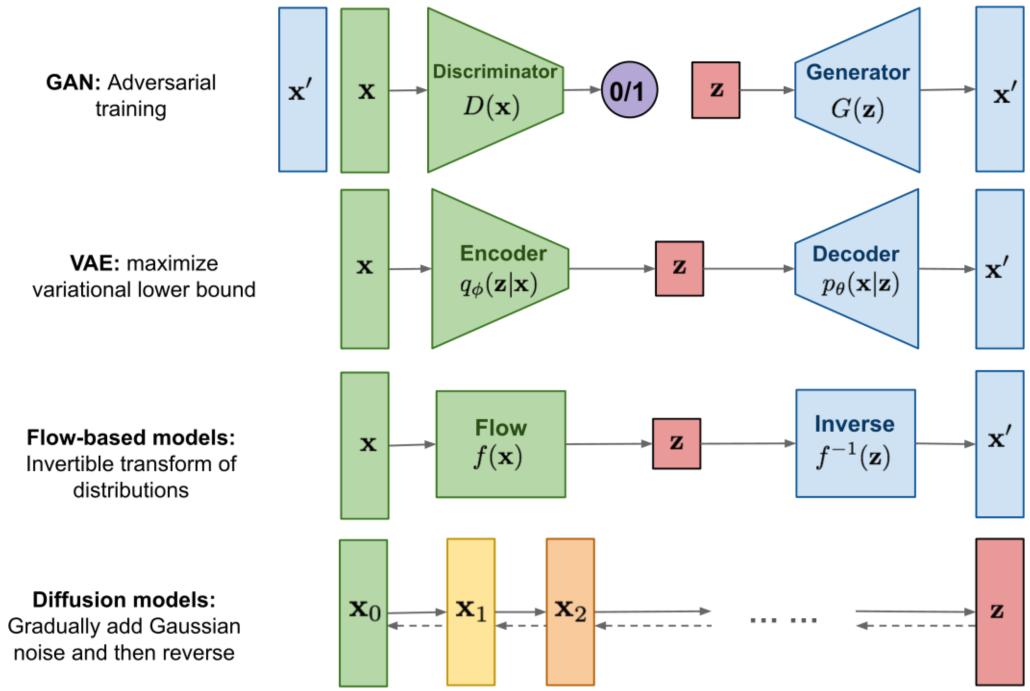
Sampling

$x_{new} \sim p_\theta$

Unsupervised Representation Learning

$z \leftarrow p_\theta(x)$

What? Generative models



likelihood-based models

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

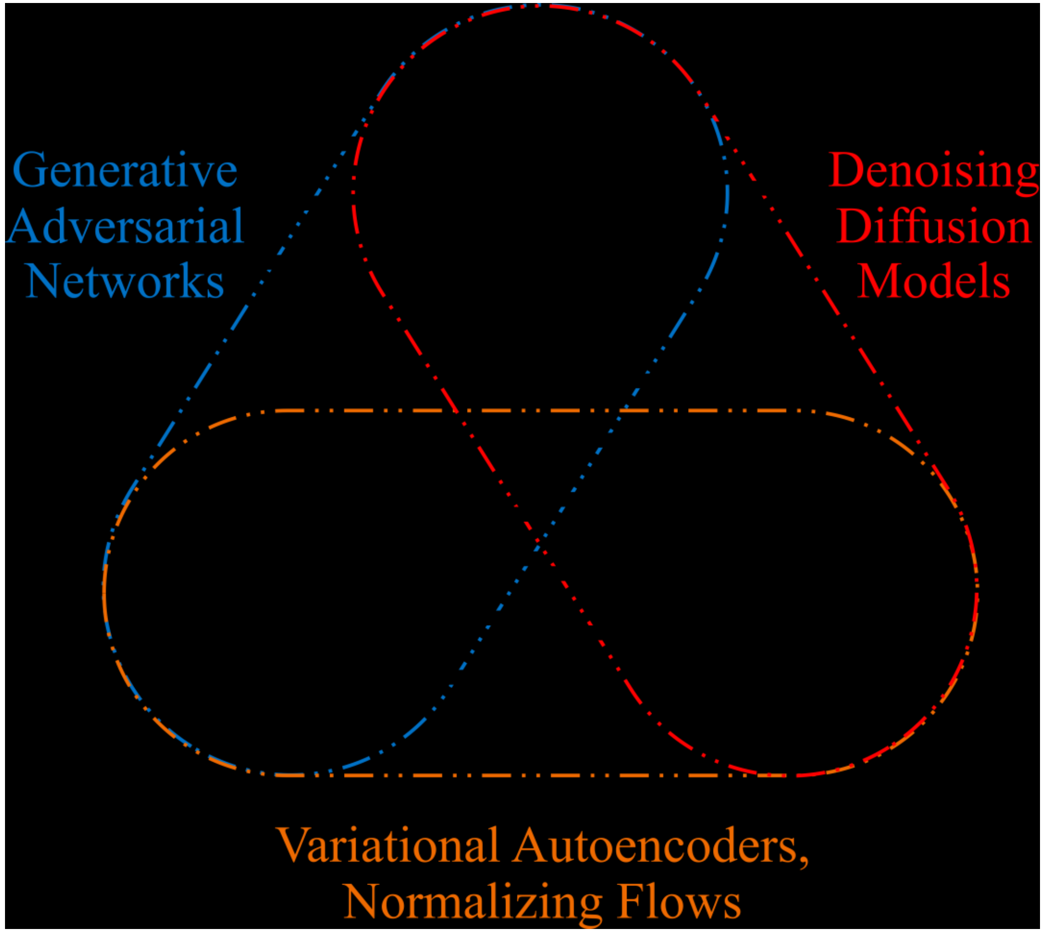
implicit generative models

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

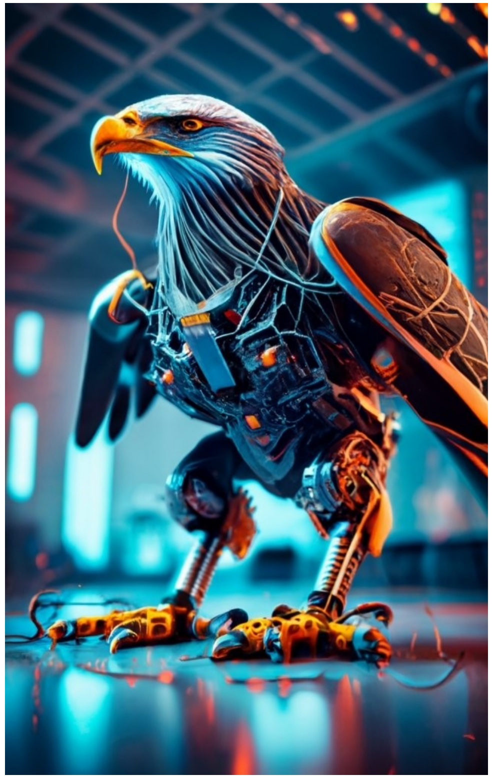


Xiao, Z., Kreis, K., & Vahdat, A. (2021). Tackling the generative learning trilemma with denoising diffusion gans. *ICLR*

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

Why? Community Push

Companies

Big models and data

OpenAI

Midjourney

stability.ai

IDEOGRAM AI

Open-Source

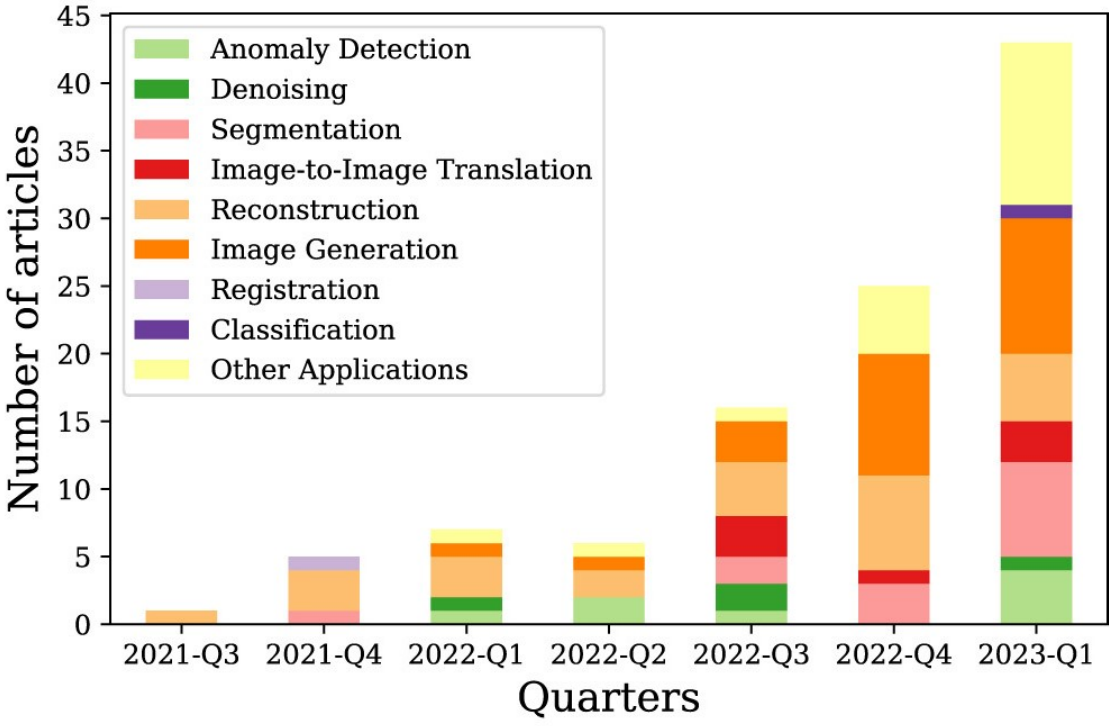
Ease of Use

Diffusers

MONAI

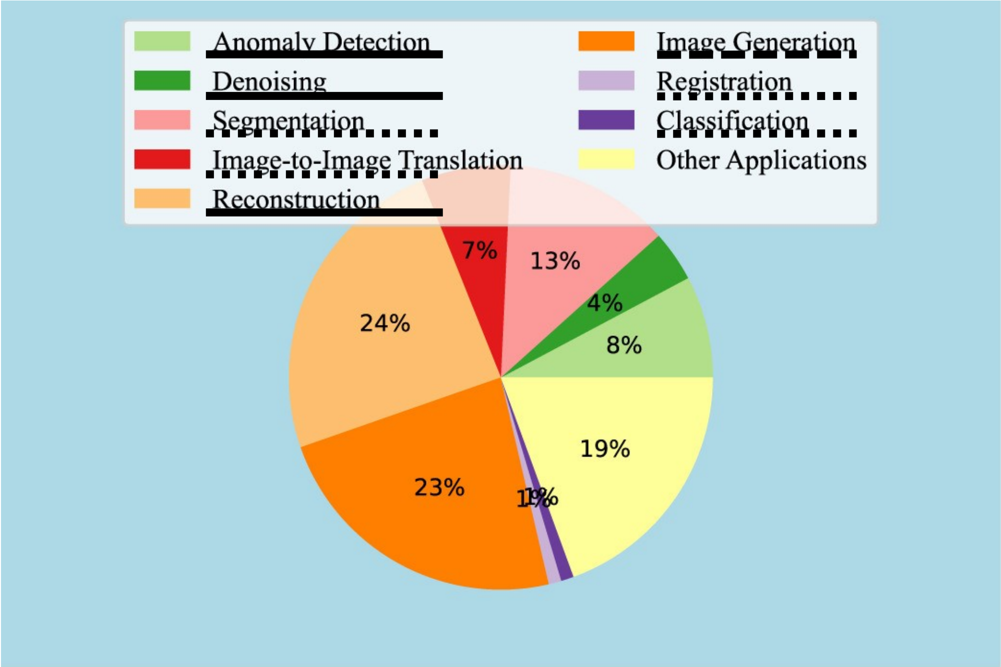
Generative Models

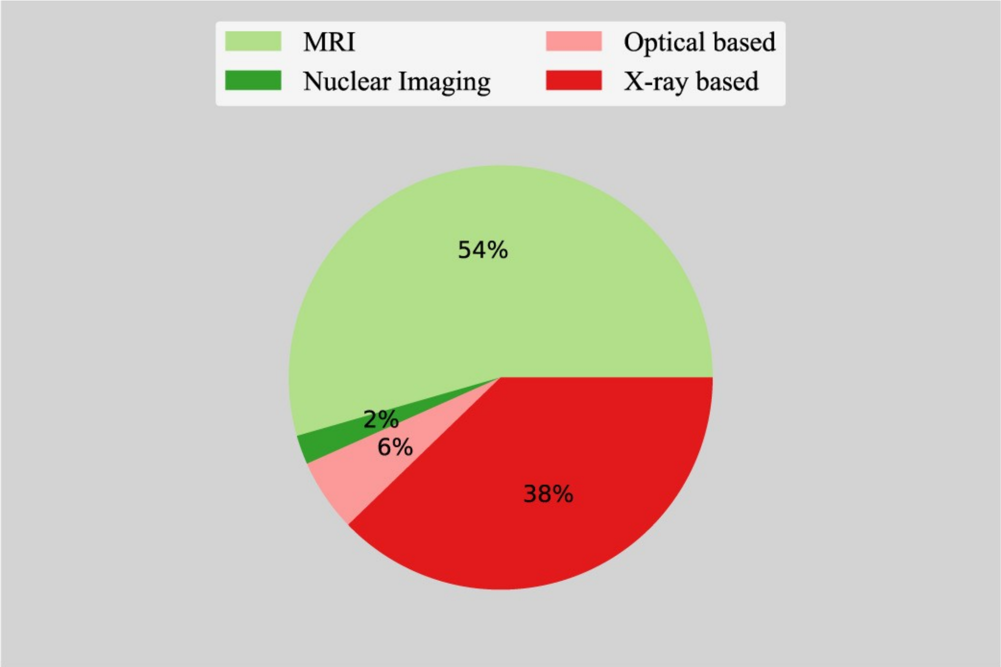
Why? Medical Imaging Popularity



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

Why? Medical Imaging Applications





Realism

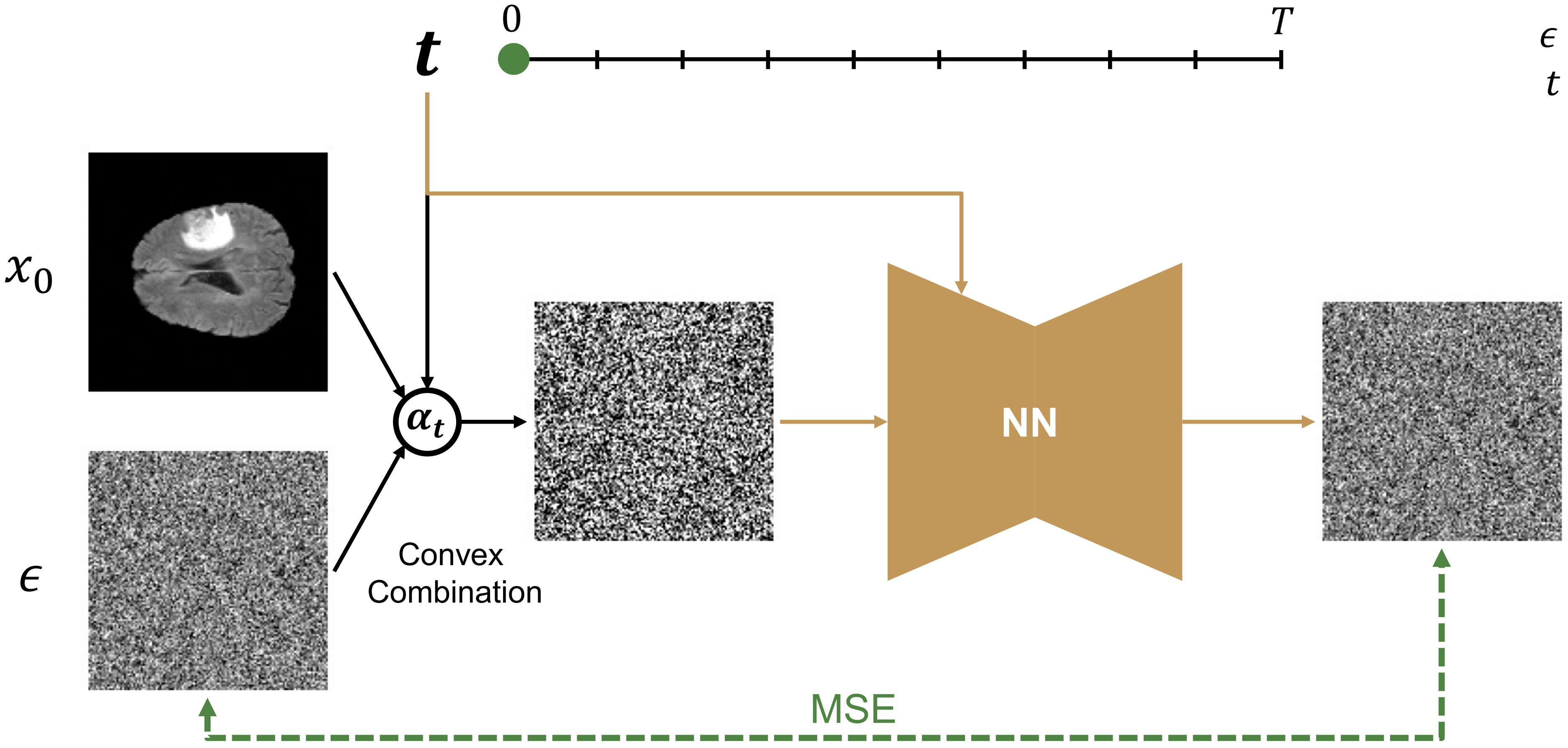
Control

Prior

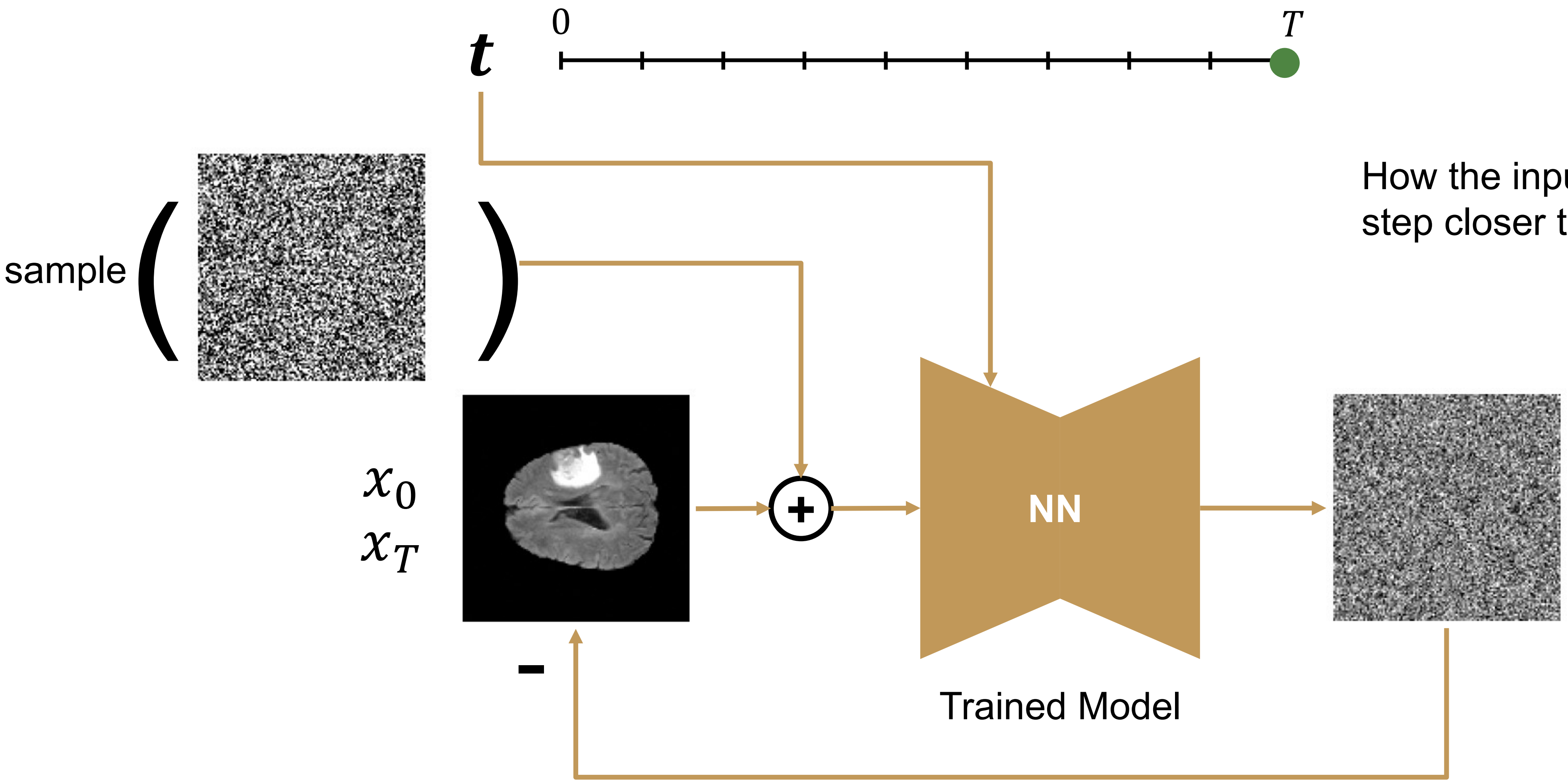
Kazerouni, Amirhossein, et al. "Diffusion models in medical imaging: A comprehensive survey." Medical Image Analysis (2023): 102846.

How? Training by Denoising

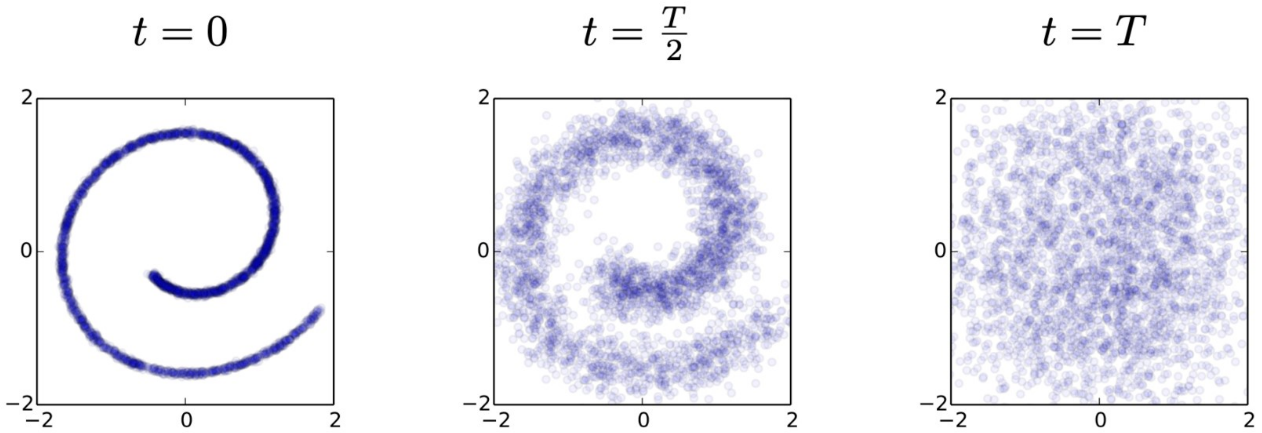
$\epsilon \sim \mathcal{N}(0, \mathbb{I})$
$t \sim \mathcal{U}(0, T)$



How? Inference



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



Understanding and Intuition

Score Function

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

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

$$\underbrace{\nabla_{\mathrm{x}} \log p_\theta(\mathrm{x})}_{\approx\, \boldsymbol{\epsilon}_\theta} = -\nabla_{\mathrm{x}} f_\theta(x) - \underbrace{\nabla_{\mathrm{x}} \log Z_\theta}_{0}$$

How to learn it?

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

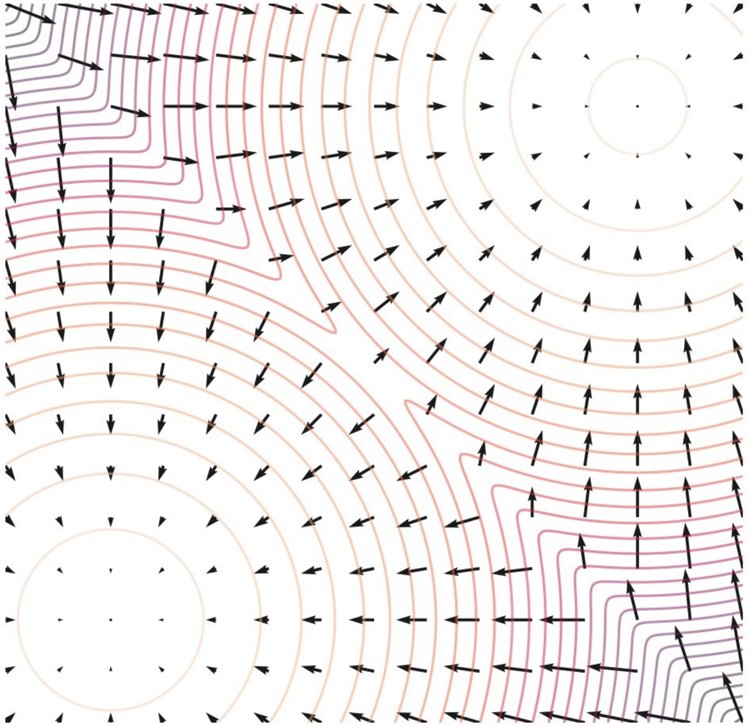


Image from blog post by Yang Song https://yang-song.net/blog/2021/score/

Denoising Score Matching

How to **learn** the score?

$$\mathbb{E}_{p(\mathrm{x})} \left\| \underbrace{\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\mathrm{x})}_{\text{Diffusion Model}} - \underbrace{\nabla_{\mathrm{x}} \log p_{\theta}(\mathrm{x})}_{\text{Score}} \right\|_2^2$$

$$\mathbb{E}_{p(\mathrm{x})} \| \boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\mathrm{x}) - \nabla_{\mathrm{x}} \log p_t(\mathrm{x}_t \mid \mathrm{x}) \|_2^2$$

$$\frac{\mathrm{x}_t - \mathrm{x}}{\sigma_t^2}$$

Forward Process

$$p_t(\mathrm{x}_t \mid \mathrm{x}) \approx p_{data}(\mathrm{x})$$

$$p_t(\mathrm{x}_t \mid \mathrm{x}) = \mathcal{N}\left(\sqrt{\alpha_t}\mathrm{x}, (1 - \alpha_t)\mathbf{I}\right)$$

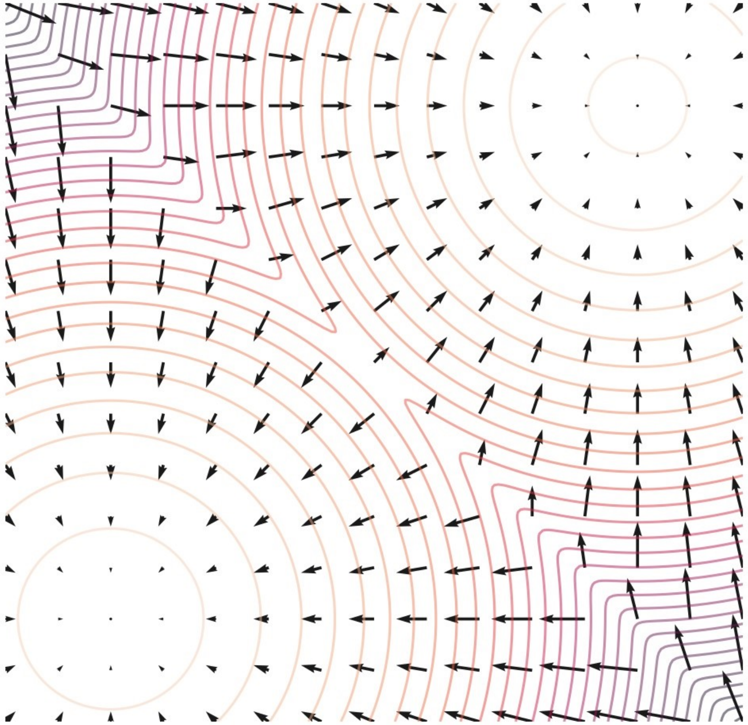
$$\mathrm{x}_t = \sqrt{\alpha_t}\,\mathrm{x} + \sqrt{1 - \alpha_t}\,\epsilon, \qquad \epsilon \sim \mathcal{N}(0, \mathbf{I})$$

Gaussian is a common perturbation

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

Learning the Score

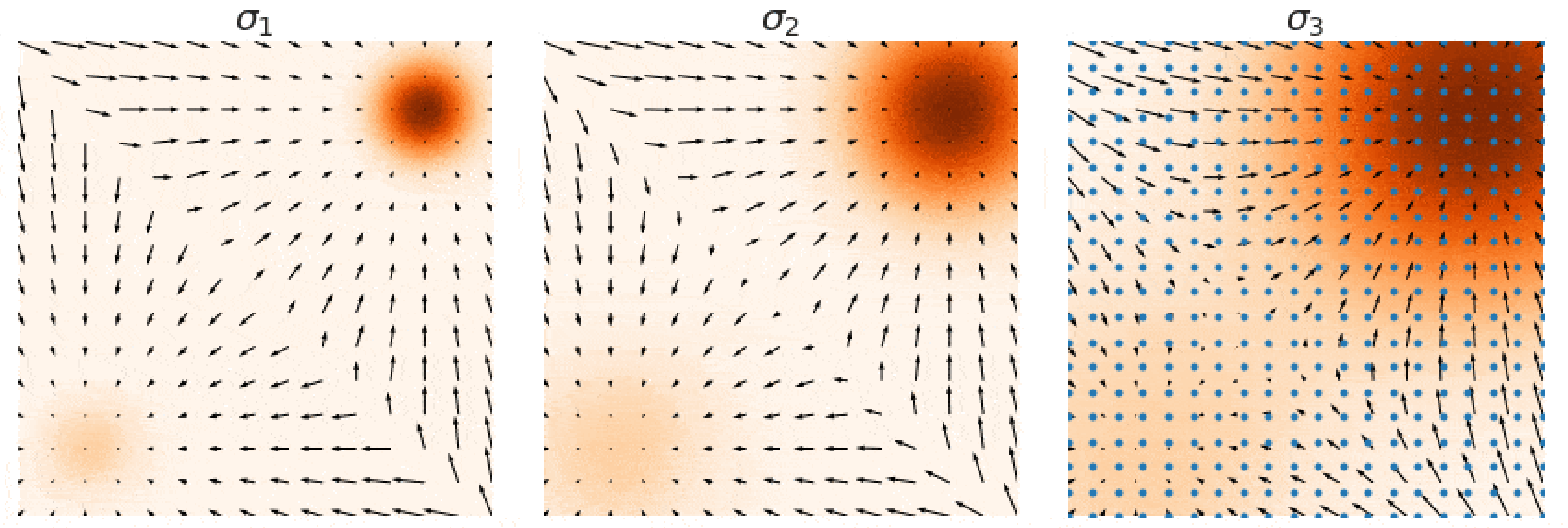
$$\mathbb{E}_{p(\mathrm{x})} \| \boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\mathrm{x}) - \nabla_{\mathrm{x}} \log p_t(\mathrm{x}_t \mid \mathrm{x}) \|_{\mathbf{2}}^{\mathbf{2}}$$



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

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

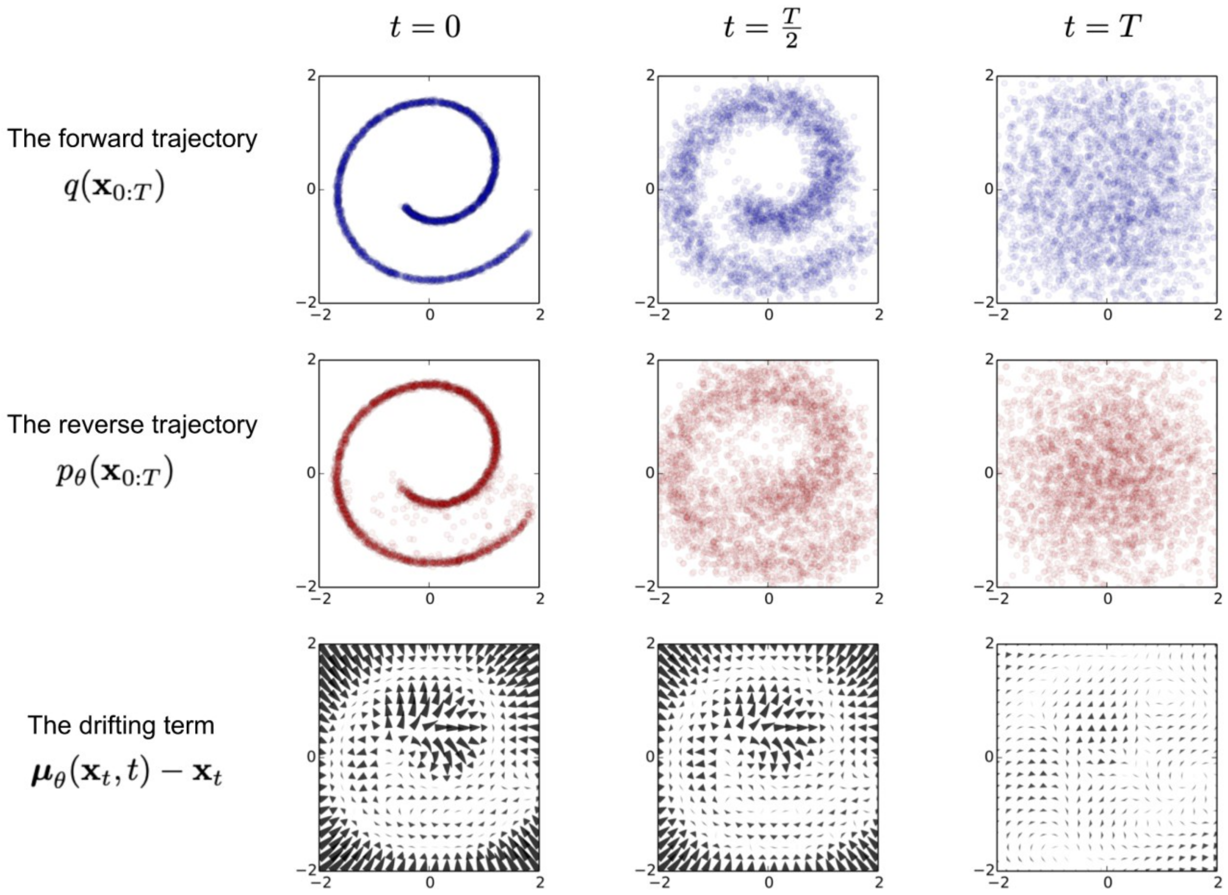
Perturbation at many scales



Learning in **low** density regions

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

Diffusion Models Learn the Gradient



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

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

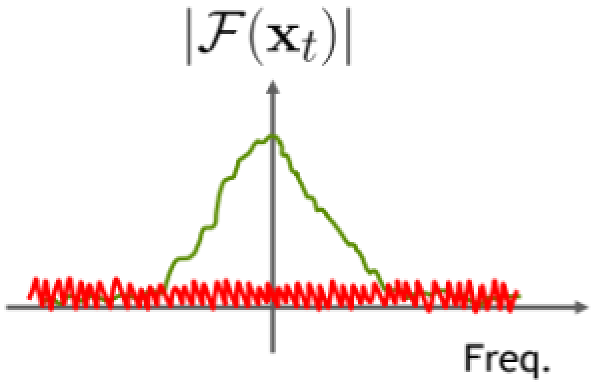
Fourier Transform

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

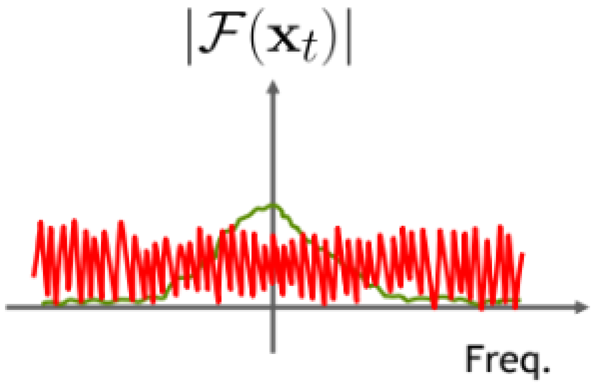
Fourier Transform

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

Small t

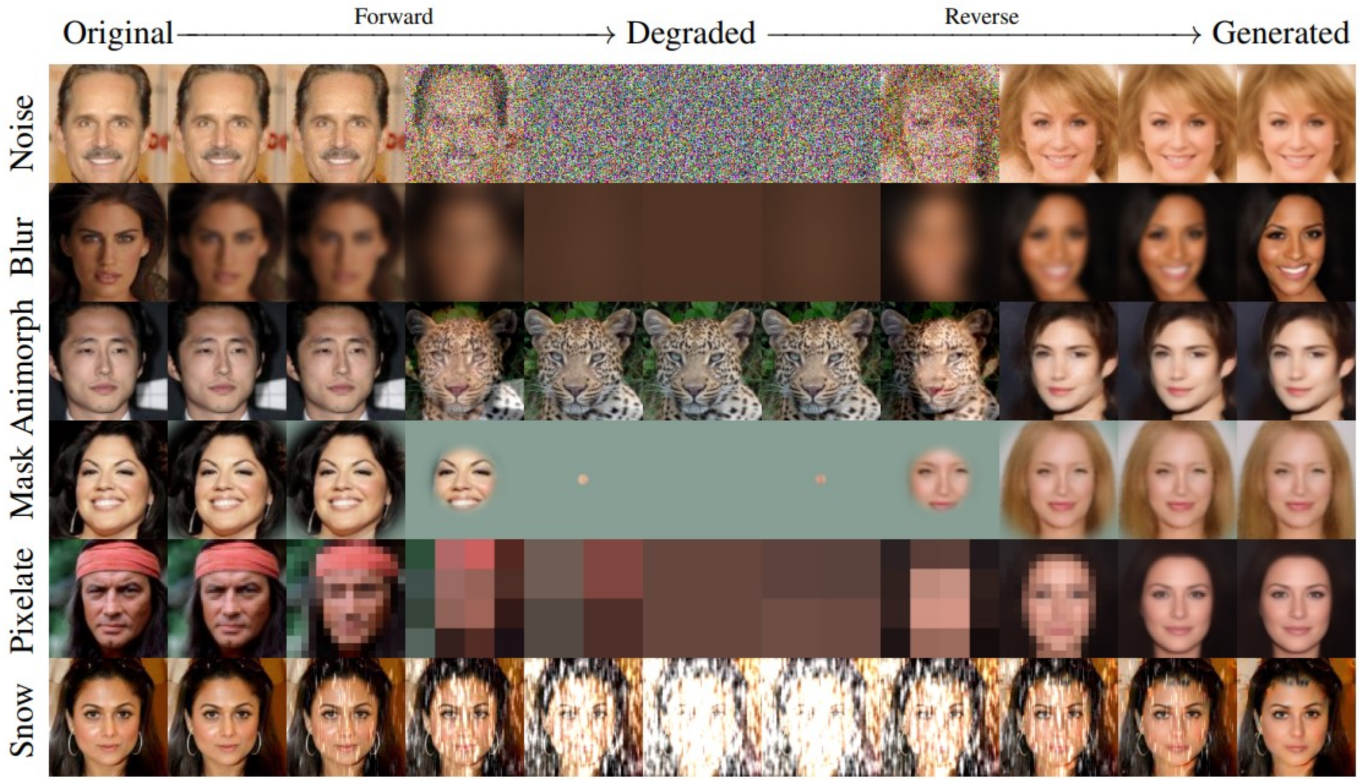


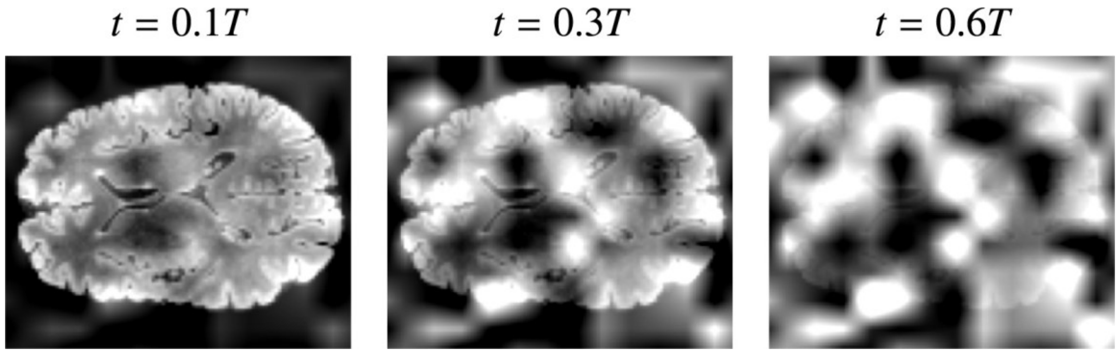
Big t



Slide inspired in CVPRs 2022 tutorial on diffusion models

Gaussian Perturbation?

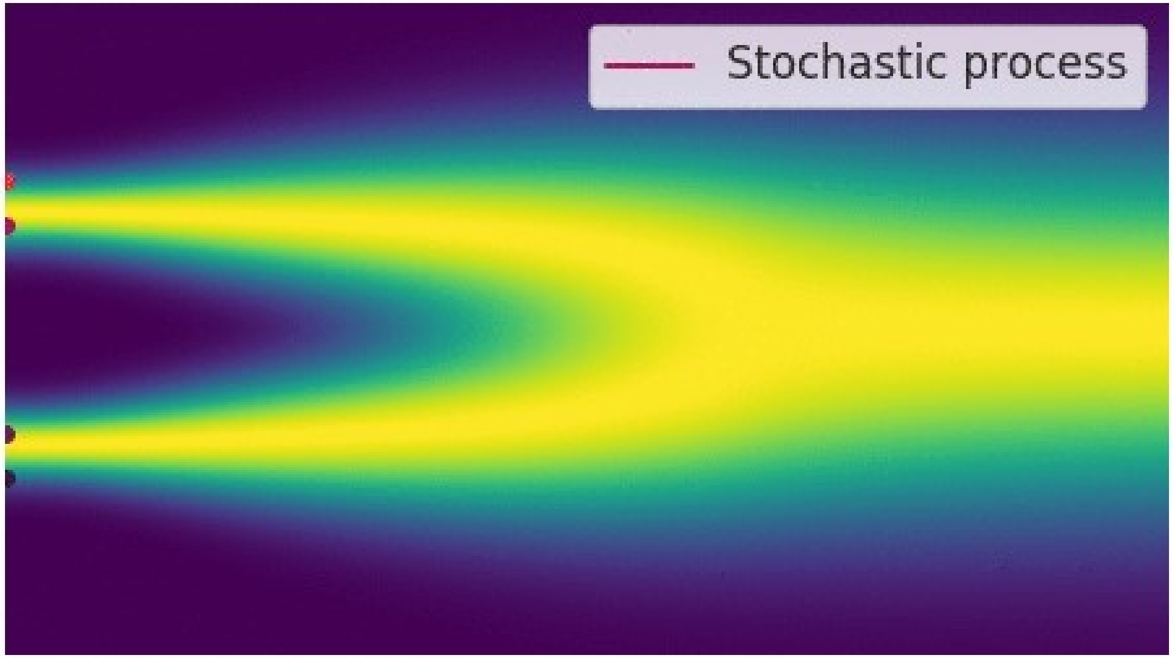




[1] Daras, Giannis, et al. "Soft diffusion: Score matching for general corruptions." arXiv preprint arXiv:2209.05442 (2022).
[2] Bansal, Arpit, et al. "Cold diffusion: Inverting arbitrary image transforms without noise." arXiv preprint arXiv:2208.09392 (2022).
[3] Kascenas, Antanas, et al. "The role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

Diffusion and Differential Equations

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

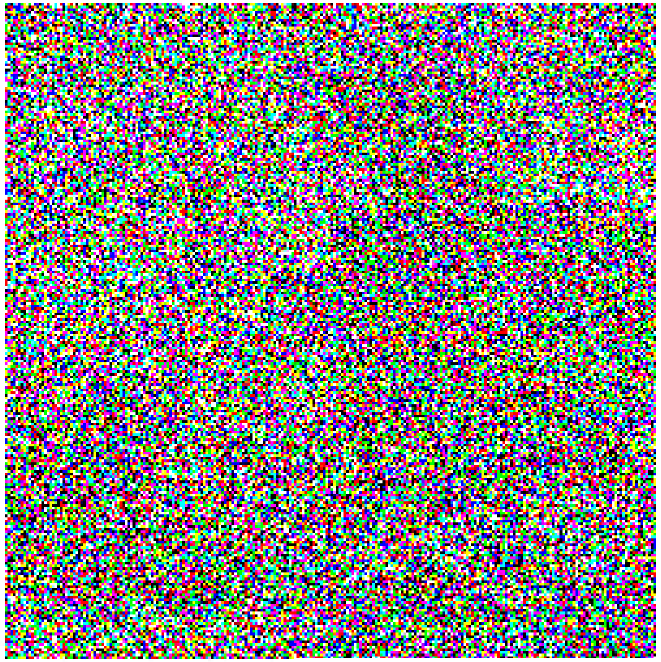


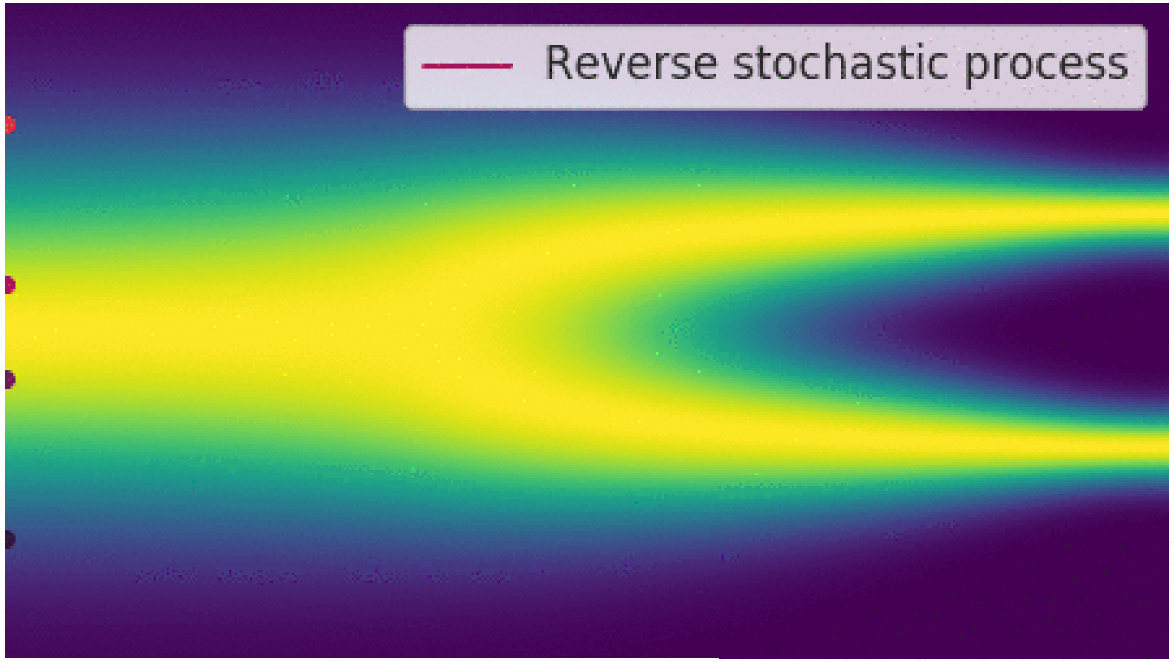
$$\mathrm{d}\mathbf{x} = \mathbf{f}(\mathbf{x}, t)\mathrm{d}t + g(t)\mathrm{d}\mathbf{w}.$$

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/

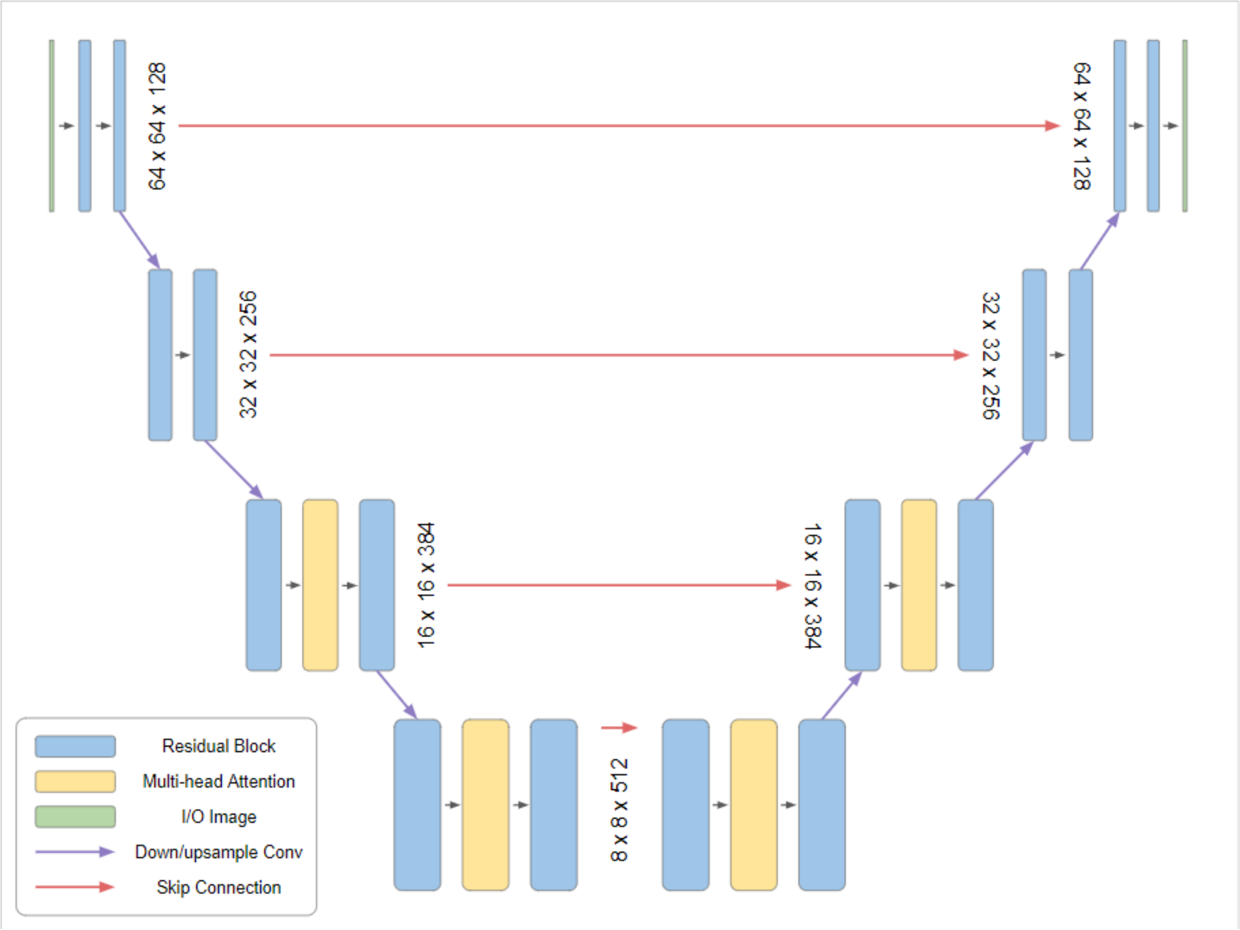
The Design Space

	VP [49]	VE [49]	iDDPM [37] + DDIM [47]	Ours ("EDM")
Sampling (Section 3)				
ODE solver	Euler	Euler	Euler	2^{nd} order Heun
Time steps $t_{i<N}$	$1+\frac{i}{N-1}(\epsilon_s-1)$	$\sigma_{\text{max}}^2\left(\sigma_{\text{min}}^2/\sigma_{\text{max}}^2\right)^{\frac{i}{N-1}}$	$u_{\lfloor j_0+\frac{M-1-j_0}{N-1}i+\frac{1}{2}\rfloor}$, where $u_M=0$ $u_{j-1}=\sqrt{\frac{u_j^2+1}{\max(\bar{\alpha}_{j-1}/\bar{\alpha}_j,C_1)}-1}$	$\left(\sigma_{\text{max}}{}^{\frac{1}{\rho}}+\frac{i}{N-1}(\sigma_{\text{min}}{}^{\frac{1}{\rho}}-\sigma_{\text{max}}{}^{\frac{1}{\rho}})\right)^{\rho}$
Schedule $\sigma(t)$	$\sqrt{e^{\frac{1}{2}\beta_{\text{d}}t^2+\beta_{\text{min}}t}-1}$	$\sqrt{t}$	t	t
Scaling $s(t)$	$1/\sqrt{e^{\frac{1}{2}\beta_{\text{d}}t^2+\beta_{\text{min}}t}}$	1	1	1
Network and preconditioning (Section 5)				
Architecture of F_θ	DDPM++	NCSN++	DDPM	(any)
Skip scaling $c_{\text{skip}}(\sigma)$	1	1	1	$\sigma_{\text{data}}^2/\left(\sigma^2+\sigma_{\text{data}}^2\right)$
Output scaling $c_{\text{out}}(\sigma)$	$-\sigma$	σ	$-\sigma$	$\sigma\cdot\sigma_{\text{data}}/\sqrt{\sigma_{\text{data}}^2+\sigma^2}$
Input scaling $c_{\text{in}}(\sigma)$	$1/\sqrt{\sigma^2+1}$	1	$1/\sqrt{\sigma^2+1}$	$1/\sqrt{\sigma^2+\sigma_{\text{data}}^2}$
Noise cond. $c_{\text{noise}}(\sigma)$	$(M-1)\,\sigma^{-1}(\sigma)$	$\ln(\frac{1}{2}\sigma)$	$M-1-\arg\min_j \lvert u_j-\sigma\rvert$	$\frac{1}{4}\ln(\sigma)$
Training (Section 5)				
Noise distribution	$\sigma^{-1}(\sigma)\sim\mathcal{U}(\epsilon_{\text{t}},1)$	$\ln(\sigma)\sim\mathcal{U}(\ln(\sigma_{\text{min}}),\ln(\sigma_{\text{max}}))$	$\sigma=u_j,\ \ j\sim\mathcal{U}\{0,M-1\}$	$\ln(\sigma)\sim\mathcal{N}(P_{\text{mean}},P_{\text{std}}^2)$
Loss weighting $\lambda(\sigma)$	$1/\sigma^2$	$1/\sigma^2$	$1/\sigma^2$ (note: *)	$\left(\sigma^2+\sigma_{\text{data}}^2\right)/(\sigma\cdot\sigma_{\text{data}})^2$
Parameters	$\beta_{\text{d}}=19.9, \beta_{\text{min}}=0.1$ $\epsilon_{\text{s}}=10^{-3}, \epsilon_{\text{t}}=10^{-5}$ $M=1000$	$\sigma_{\text{min}}=0.02$ $\sigma_{\text{max}}=100$	$\bar{\alpha}_j=\sin^2(\frac{\pi}{2}\frac{j}{M(C_2+1)})$ $C_1=0.001, C_2=0.008$ $M=1000, j_0=8^{\dagger}$	$\sigma_{\text{min}}=0.002, \sigma_{\text{max}}=80$ $\sigma_{\text{data}}=0.5, \rho=7$ $P_{\text{mean}}=-1.2, P_{\text{std}}=1.2$

* iDDPM also employs a second loss term L_{vlb} † In our tests, $j_0=8$ yielded better FID than $j_0=0$ used by iDDPM

Karras, T., Aittala, M., Aila, T., & Laine, S. (2022). Elucidating the design space of diffusion-based generative models. *NeurIPS*

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…

DEMO

MONAI Generative Models

Allows researchers and developers to easily train, evaluate, and deploy generative models on medical imaging.

King's College London

The University of Edinburgh

NVIDIA

Universitat de Barcelona

Icahn School of Medicine at Mount Sinai

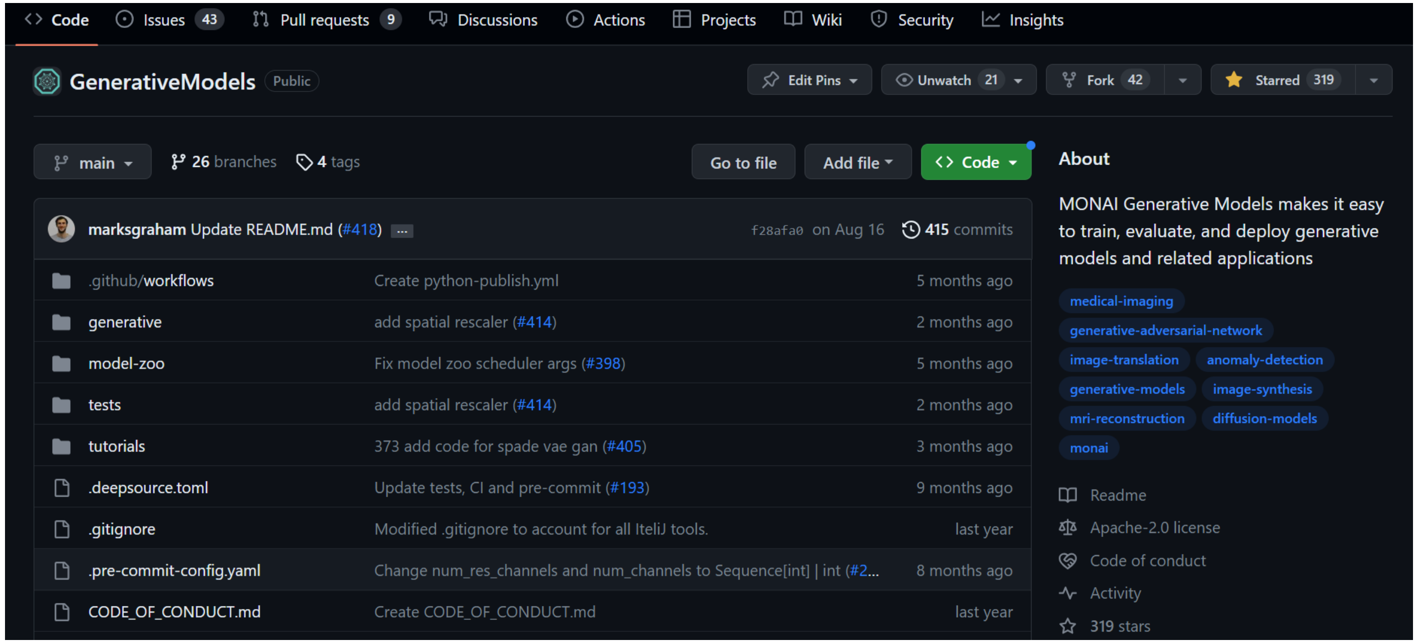
KAIST

Stanford University

NIH National Institutes of Health

Features

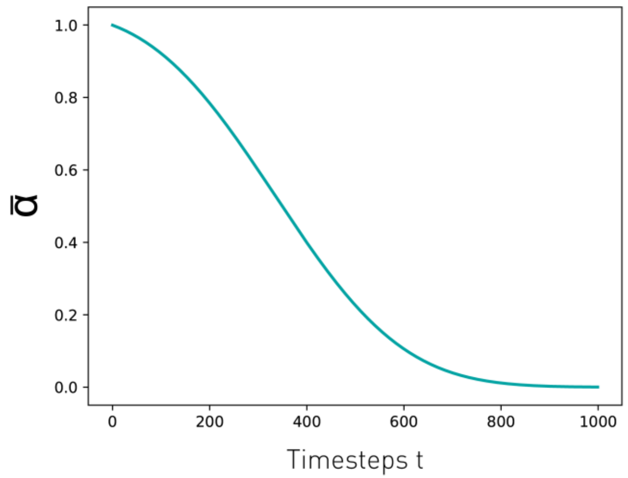
- ❑ State-of-the-art models
- ❑ Losses and supporting classes to train models
- ❑ Evaluation metrics
- ❑ Tutorials
- ❑ Pre-trained models



U-Net Architecture

```python
from generative.networks.nets import DiffusionModelUNet

model = DiffusionModelUNet(
    spatial_dims=3,
    in_channels=1,
    out_channels=1,
    num_channels=[256, 256, 512],
    attention_levels=[False, False, True],
    num_head_channels=[0, 0, 512],
    num_res_blocks=2,
)
```

Noise Schedulers

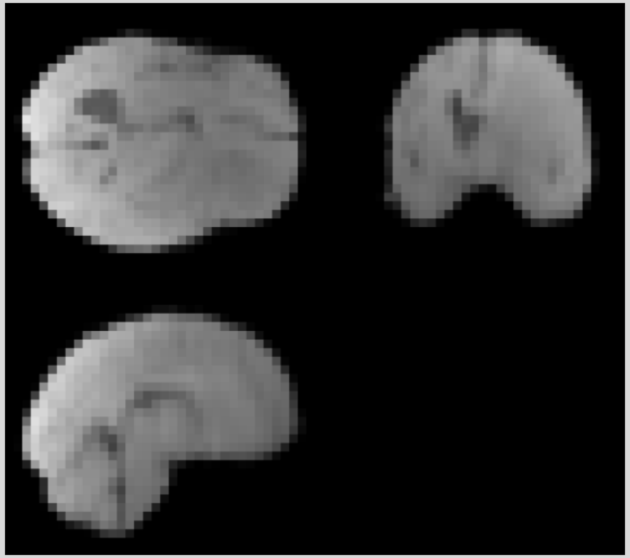


$$
\begin{aligned}
\mathbf{x}_t &= \sqrt{\alpha_t}\mathbf{x}_{t-1} + \sqrt{1-\alpha_t}\boldsymbol{\epsilon}_{t-1} \\
&= \sqrt{\alpha_t\alpha_{t-1}}\mathbf{x}_{t-2} + \sqrt{1-\alpha_t\alpha_{t-1}}\bar{\boldsymbol{\epsilon}}_{t-2} \\
&= \dots \\
&= \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon} \\
q(\mathbf{x}_t|\mathbf{x}_0) &= \mathcal{N}(\mathbf{x}_t; \sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1-\bar{\alpha}_t)\mathbf{I})
\end{aligned}
$$

```
from generative.networks.schedulers
import DDPMScheduler


scheduler = DDPMScheduler(
  num_train_timesteps=1000,
  beta_schedule="scaled_linear",
  beta_start=0.0005,
  beta_end=0.0195,
)
```

3D - Preprocessing

```
from monai import transforms
from monai.apps import DecathlonDataset
from monai.data import DataLoader

train_transform = transforms.Compose(
    [
transforms.LoadImaged(keys=["image"]),
transforms.Lambdad(keys=["image“], func=lambda x: x[:, :, :, 1]),
transforms.AddChanneld(keys=["image"]),
transforms.ScaleIntensityd(keys=["image"]),
transforms.CenterSpatialCropd(keys=["image"], roi_size=[160, 200, 155]),
transforms.Resized(keys=["image"], spatial_size=(32, 40, 32)),
    ]
)

train_ds = DecathlonDataset(
    root_dir="./data", task="Task01_BrainTumour", transform=train_transform, section="training", download=True
)
train_loader = DataLoader(train_ds, batch_size=8, shuffle=True, num_workers=8, persistent_workers=True)
```



Training

```
…
for batch in train_loader:
    model.train()
    images = batch["image"].to(device)

    optimizer.zero_grad(set_to_none=True)

    noise = torch.randn_like(images).to(device)
    timesteps = torch.randint(0, scheduler.num_train_timesteps,(images.shape[0],))
    noisy_image = scheduler.add_noise(original_samples=images,
                                      noise=noise,
                                      timesteps=timesteps)

    noise_pred = model(x=noisy_image, timesteps=timesteps)

    loss = F.mse_loss(noise_pred.float(), noise.float())
    …
```

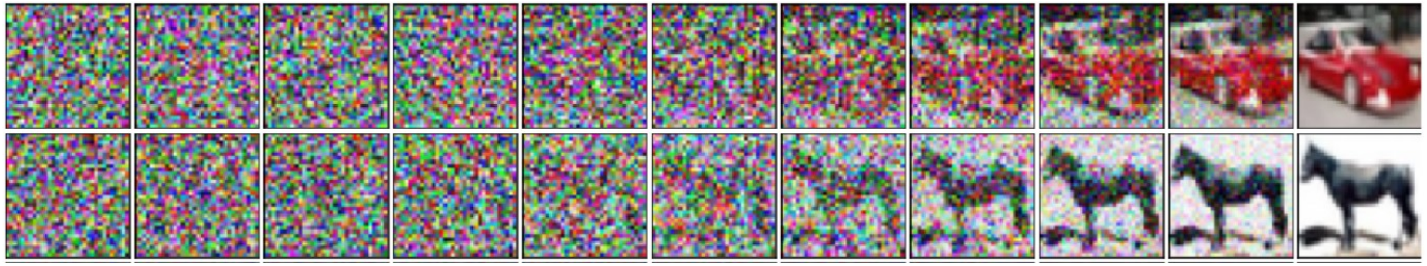
Sampling Images

```python
model.eval()
noise = torch.randn((1, 1, 32, 40, 32)) # BS, Channels, 3D
scheduler.set_timesteps(num_inference_steps=1000)

for t in iter(scheduler.timesteps):
    model_output = model(noise, timesteps=(t,))
    noise, _ = scheduler.step(model_output, t, noise)
image = noise
```

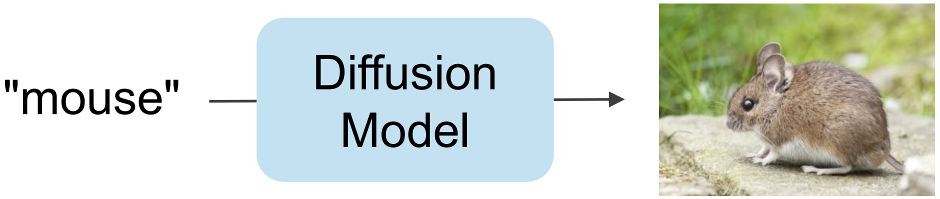

t=1000
Timestep [t]
t=0

Part 2 – Advanced Topics

- Sampling Strategies



- Conditioning Mechanisms



Basic Idea of Denoising Diffusion Models

x_0

x_0

Diffusion Models Beat GANs on Image Synthesis

Prafulla Dhariwal*
OpenAI
prafulla@openai.com

Alex Nichol*
OpenAI
alex@openai.com

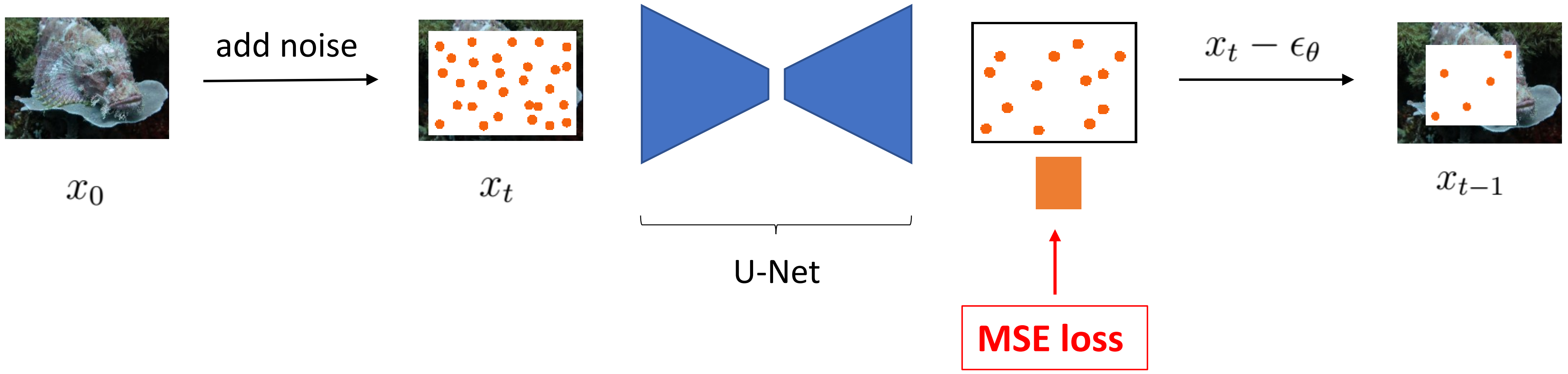
Abstract

We show that diffusion models can achieve image sample quality superior to the current state-of-the-art generative models. We achieve this on unconditional image synthesis by finding a better architecture through a series of ablations. For conditional image synthesis, we further improve sample quality with classifier guidance: a simple, compute-efficient method for trading off diversity for fidelity using gradients from a classifier. We achieve an FID of 2.97 on ImageNet 128×128, 4.59 on ImageNet 256×256, and 7.72 on ImageNet 512×512, and we match

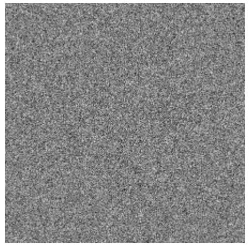
- The noising p … $x_{t-1}, \beta_t \mathbf{I})$
  with forward
- The
- For this, we n

Training Overview

- We choose a random step $t \in \{0,1,\dots,T\}$.
- We add t steps of noise to our input image x_0, and obtain a noisy image x_t.
- Our model predicts the noise pattern that needs to be subtracted from x_t, to predict a slightly denoised x_{t-1}.



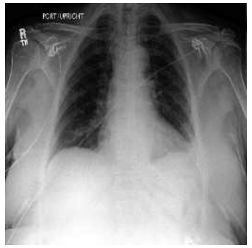
Fake Image Generation



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

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

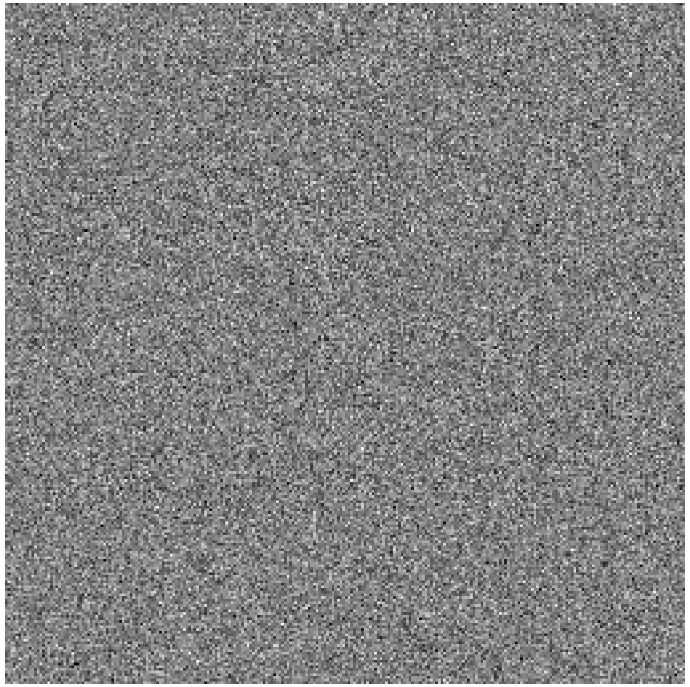
synthetic image



x_0

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\overline{\alpha}_t}} \overbrace{\qquad\qquad}^{\text{U-Net}}\right) + \sigma_t \quad \text{with}$$

Random component



DDPM Scheduler

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

Abstract

We present high quality image synthesis results using diffusion probabilistic models, a class of latent variable models inspired by considerations from nonequilibrium thermodynamics. Our best results are obtained by training on a weighted variational bound designed according to a novel connection between diffusion probabilistic

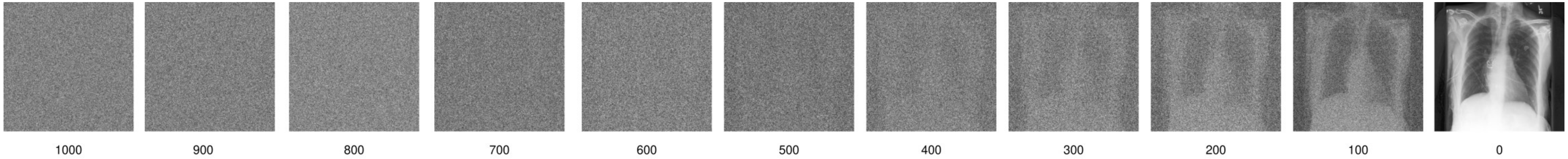
Algorithm 1 Training

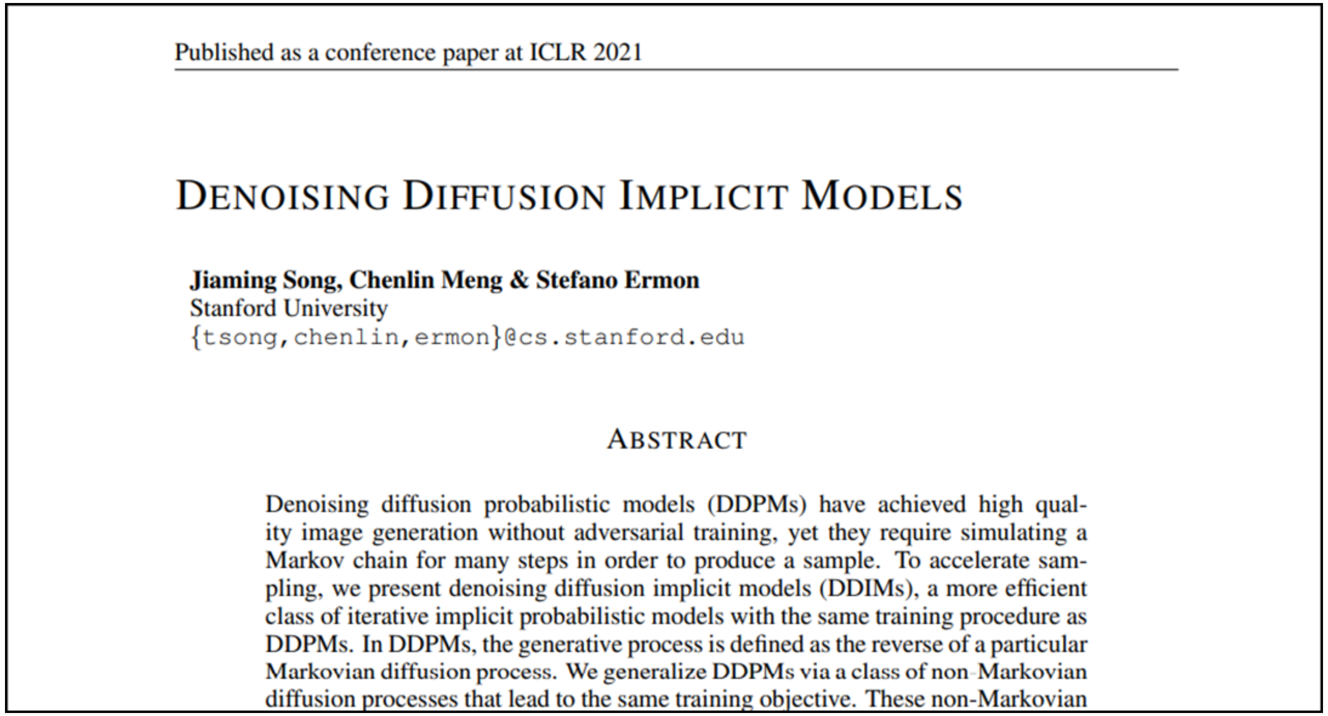
repeat
  $\mathbf{x}_0 \sim q(\mathbf{x}_0)$
  $t \sim \mathrm{Uniform}(\{1, \ldots, T\})$
  $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
  Take gradient descent step on
  $$\nabla_\theta \left\| \boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0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

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 $\sigma_t = \sqrt{(1-\alpha_{t-1})/(1-\alpha_t)}\sqrt{1-\alpha_t/\alpha_{t-1}}$

DDIM sampling scheme $\sigma_t = 0$

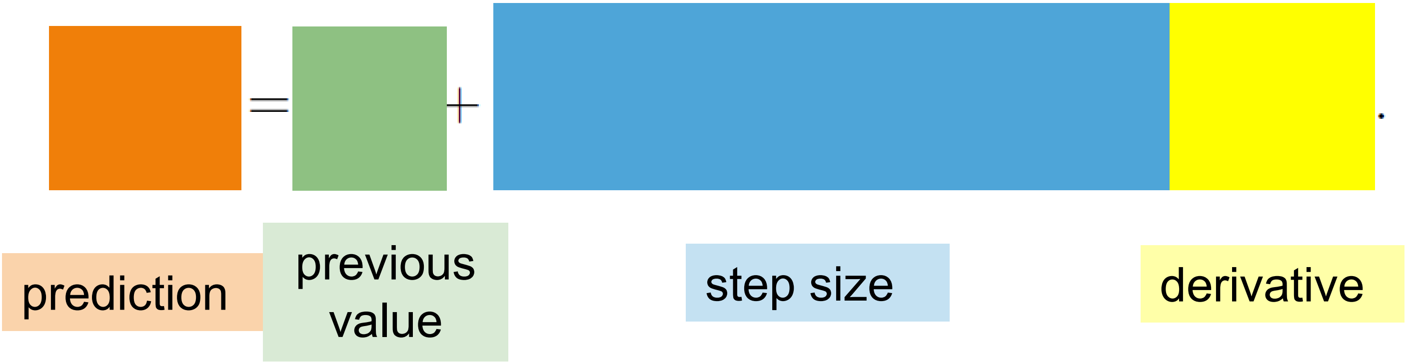
We remove the random component

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

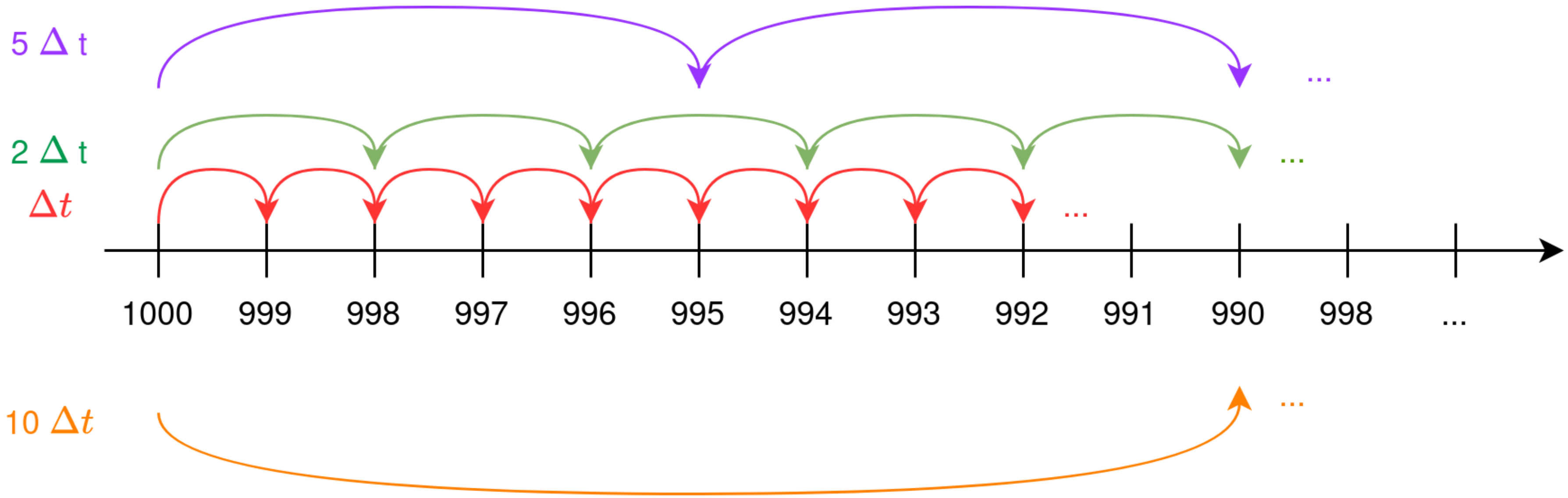


- This can be interpreted as the Euler approximation of an ODE.
- We can speed up the generation process by choosing a larger step size.
- DDIM is a **probability flow** ODE from a SDE [1].

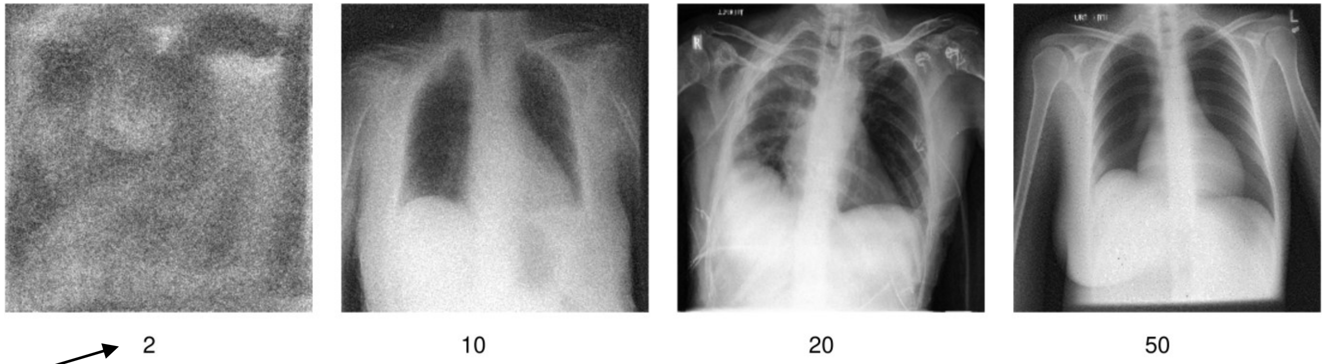
Faster, but less accurate

[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

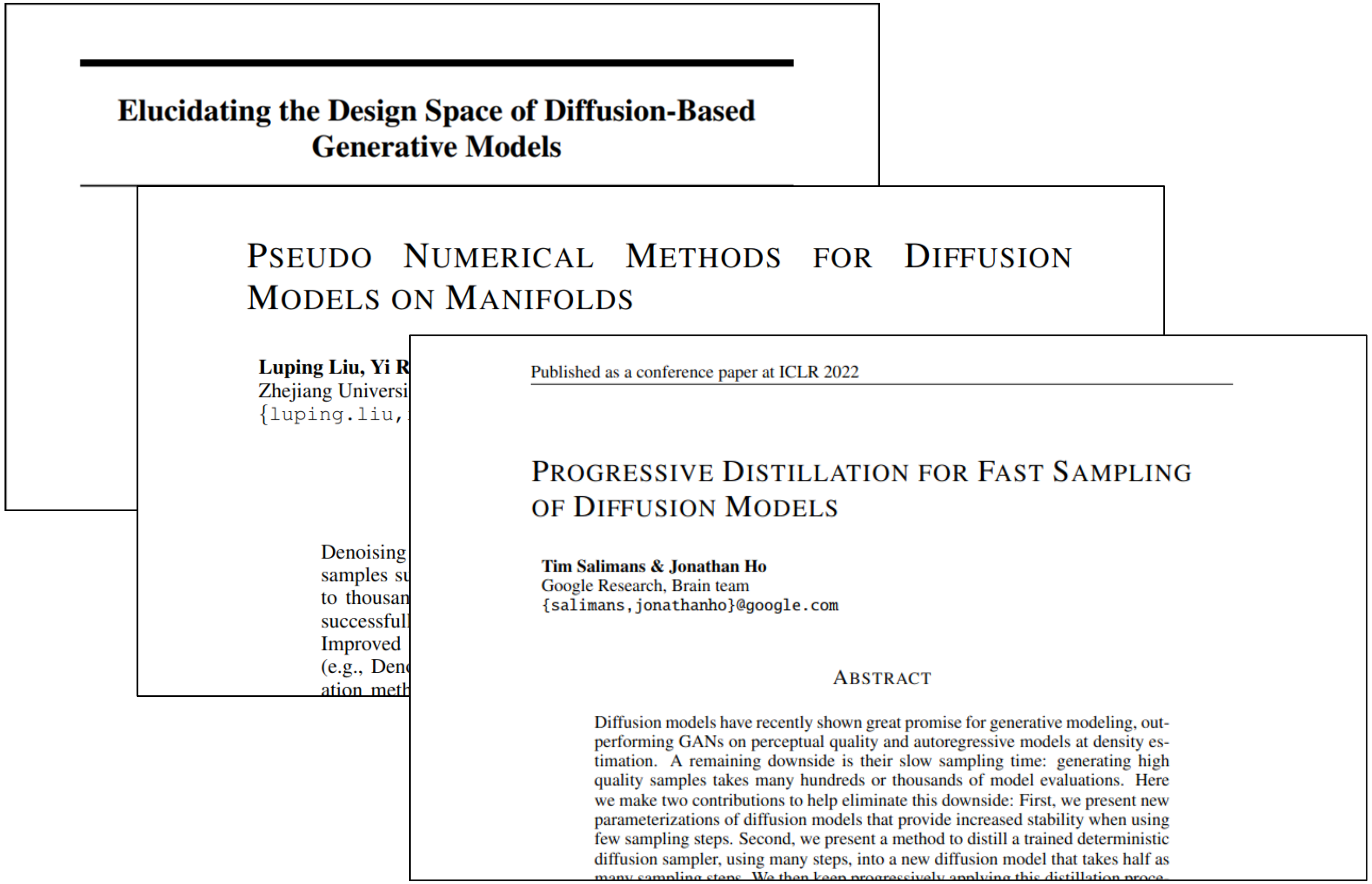


- By skipping k steps, we have a step size of $k\Delta t$.
- Sampling is k times faster.
- We trade image quality for speed.



Total amount of steps

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
successful
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

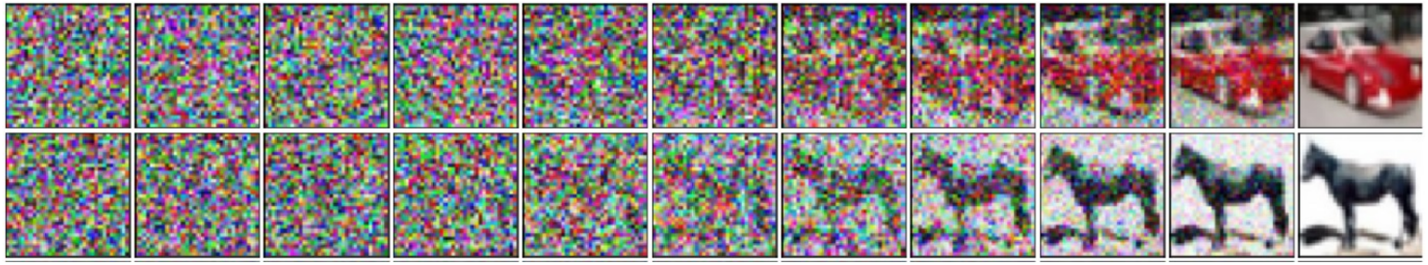
ABSTRACT

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

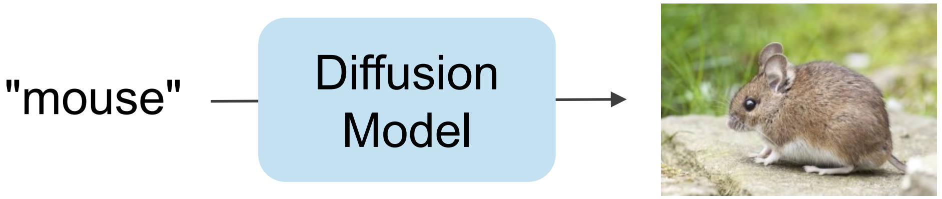
- Choosing a different solver for the given ODE can improve speed and image quality.
- Other numerical approaches such as **Heun's Method** or **Runge Kutta** solvers can be explored.
- Knowledge distillation techniques can be used for fast sampling.

Part 2 – Advanced Topics

- Sampling Strategies



- Conditioning Mechanisms



Conditioning

1. **Inference-time**
    1. An inverse problem view
        - ❑ Classifier guidance
    2. DDIM inversion
        - ❑ Interpolation
        - ❑ Gradient guidance

2. **Training-time**
    1. Scalar inputs
    2. Text
    3. Images
    4. ControlNet
    5. DreamBooth

Inverse Problem

- We consider two random variables x and y.
- Suppose we know the forward process of generating y from x, represented by the transition probability distribution $p(y|x)$.
- We aim to solve the inverse problem $p(x|y)$.
- With the Bayes' rule, we have

$$p(\mathbf{x} \mid \mathbf{y}) = p(\mathbf{x})p(\mathbf{y} \mid \mathbf{x}) / \int p(\mathbf{x})p(\mathbf{y} \mid \mathbf{x})\mathrm{d}\mathbf{x}.$$

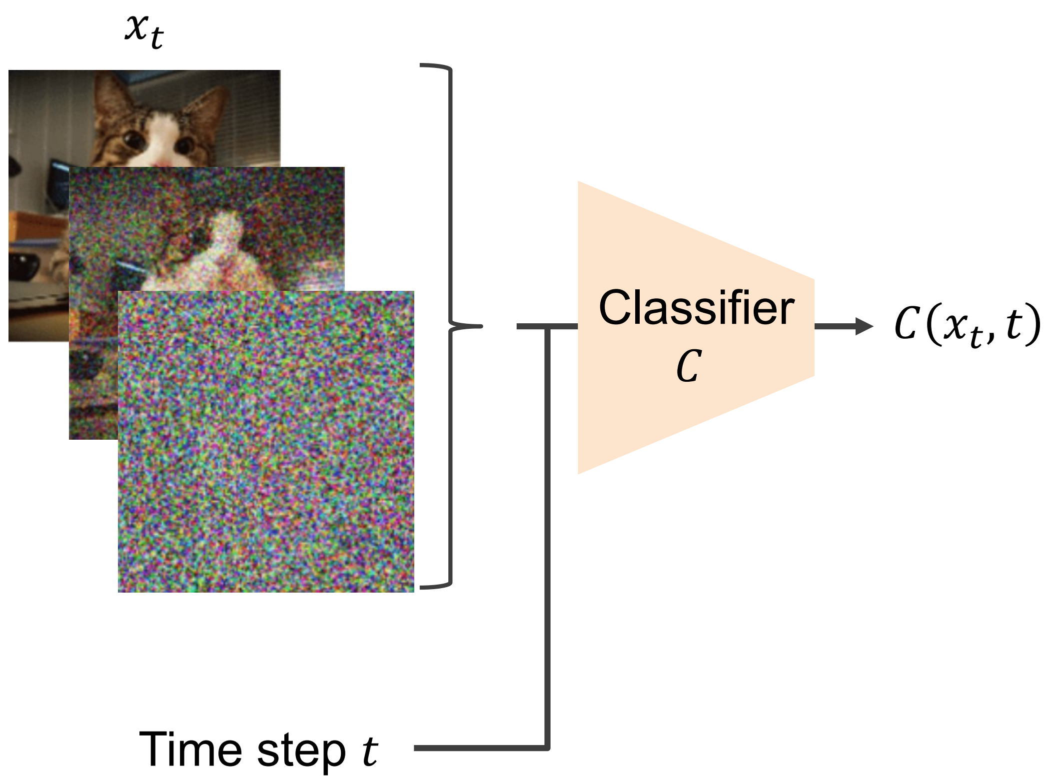
- Like in score-based models, we take the gradient of the log

$\nabla_{\mathbf{x}} \log$ [redacted] $= \nabla_{\mathbf{x}}$ [redacted] $+ \nabla_{\mathbf{x}}$ [redacted]

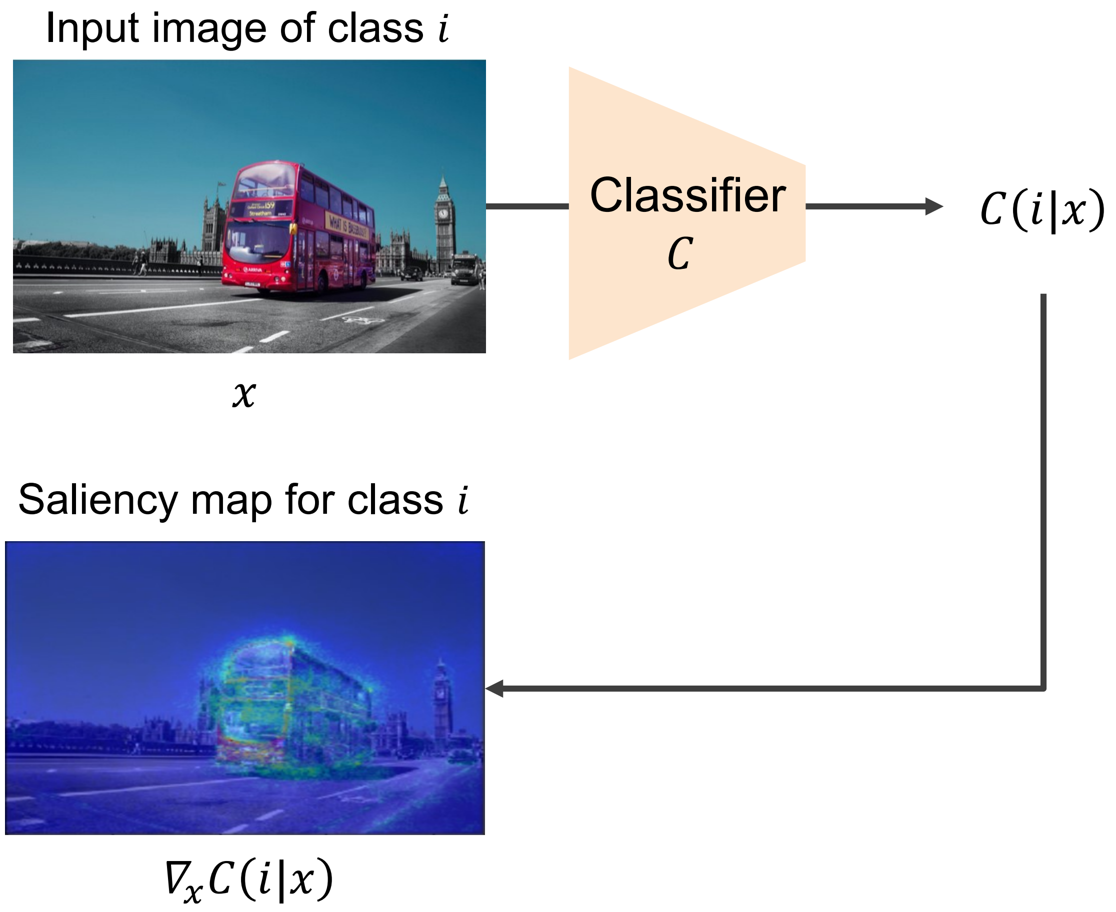
Image given the condition	This is known (diffusion model ϵ_θ)	This is the condition (classifier, …)

Example: Classifier Guidance

We want a class-conditional diffusion model.



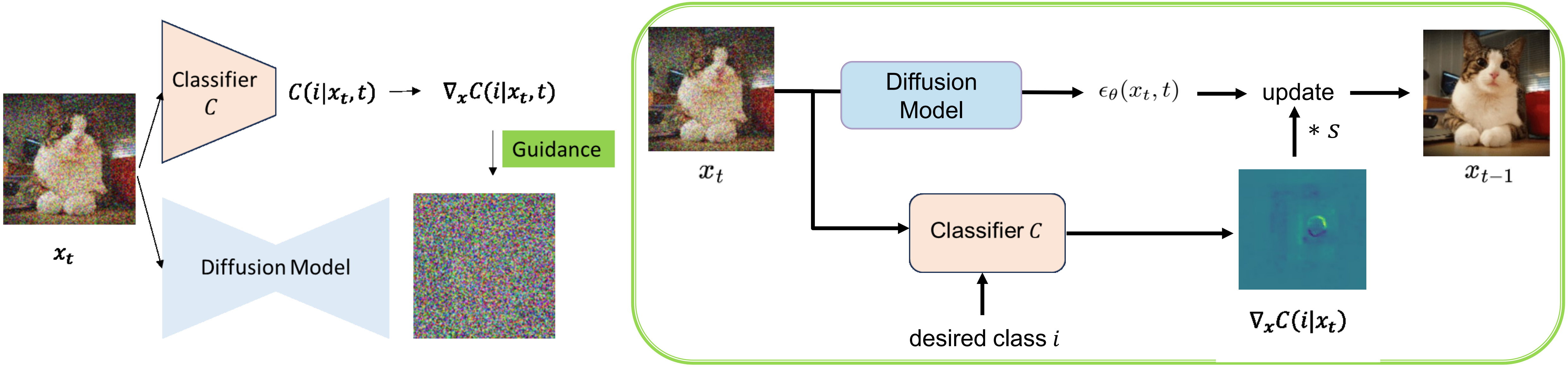
We consider the gradient with respect to the input pixels.



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.



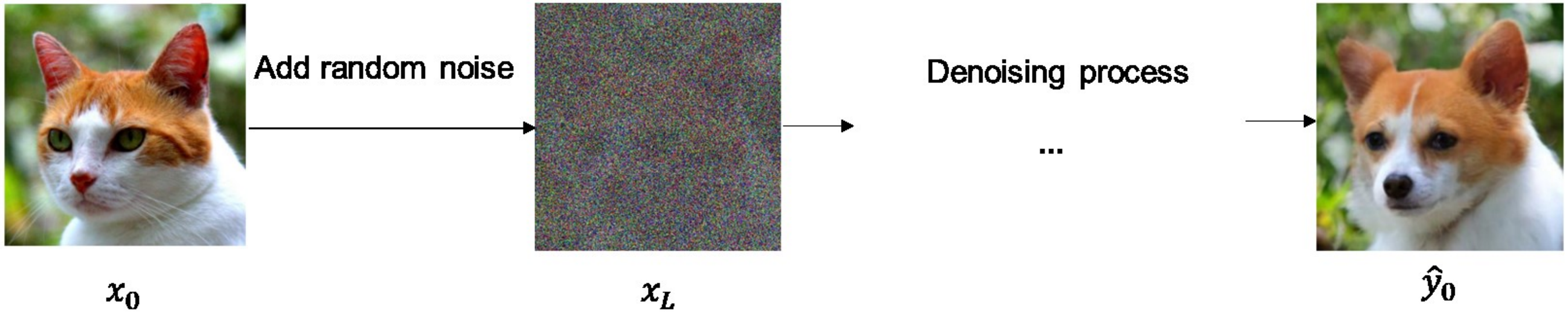
➡ Gradient guidance is not restricted to classification models. Other models (e.g., regression, segmentation, …) work just in the same way.

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

Classifier Guidance

How can we preserve information?

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

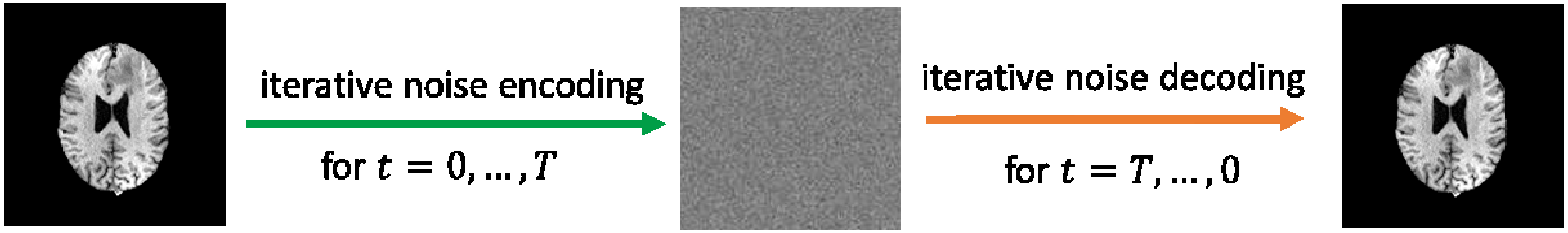
$$\frac{x_{t-1}}{\sqrt{\bar{\alpha}_{t-1}}} = \frac{x_t}{\sqrt{\bar{\alpha}_t}} + \left(\sqrt{\frac{1-\bar{\alpha}_{t-1}}{\bar{\alpha}_{t-1}}} - \sqrt{\frac{1-\bar{\alpha}_t}{\bar{\alpha}_t}}\right)\epsilon_\theta(x_t, t).$$

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

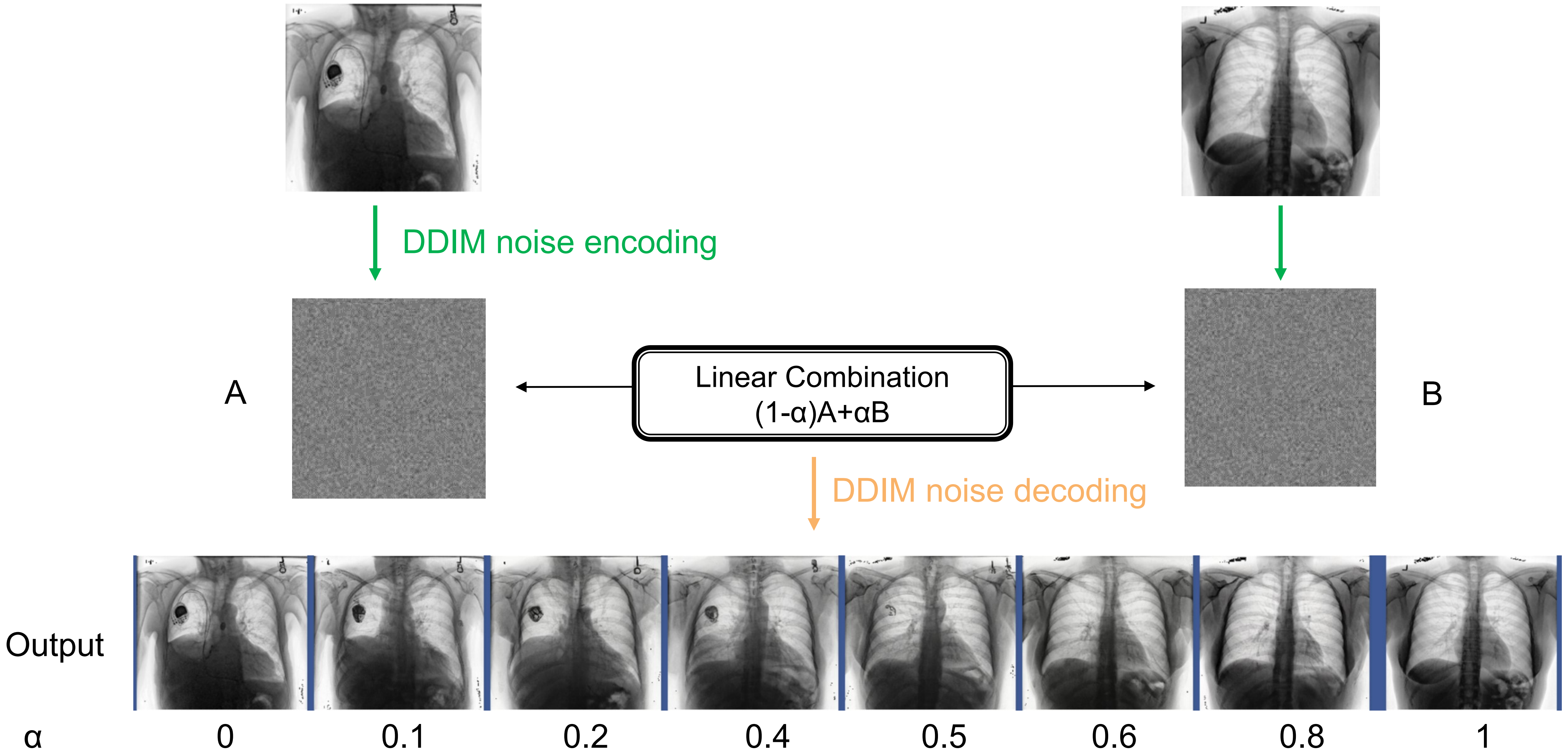
$$\frac{x_{t+1}}{\sqrt{\bar{\alpha}_{t+1}}} = \frac{x_t}{\sqrt{\bar{\alpha}_t}} + \left(\sqrt{\frac{1-\bar{\alpha}_{t+1}}{\bar{\alpha}_{t+1}}} - \sqrt{\frac{1-\bar{\alpha}_t}{\bar{\alpha}_t}}\right)\epsilon_\theta(x_t, t).$$

Noise encoding

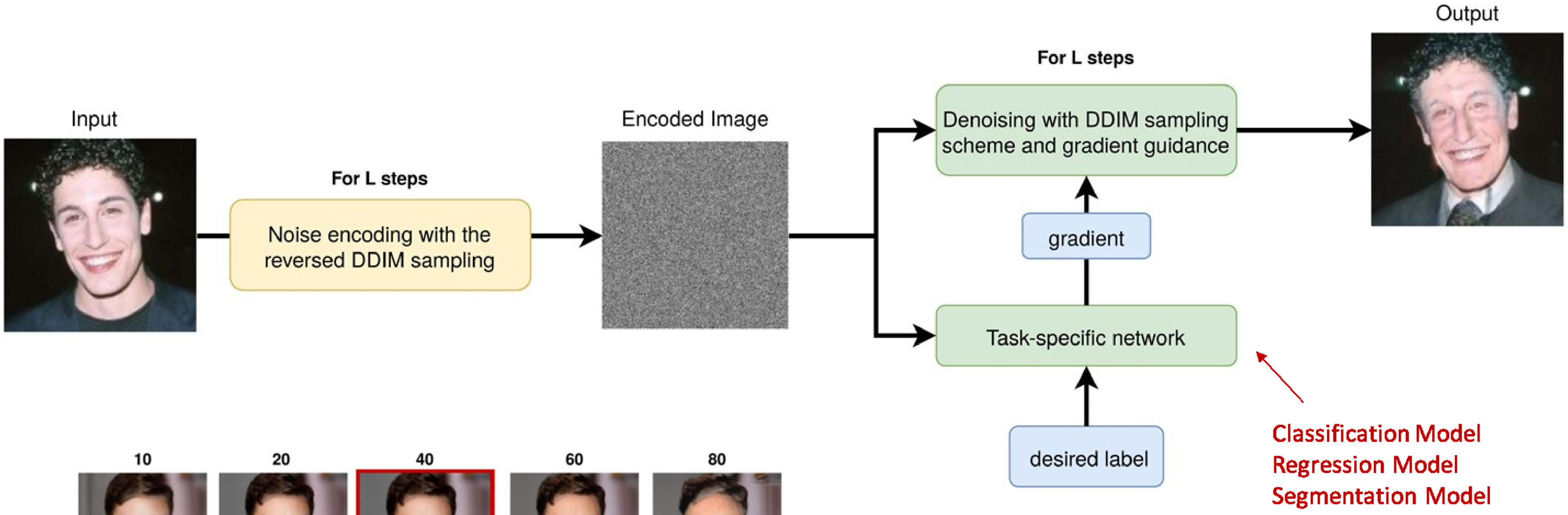


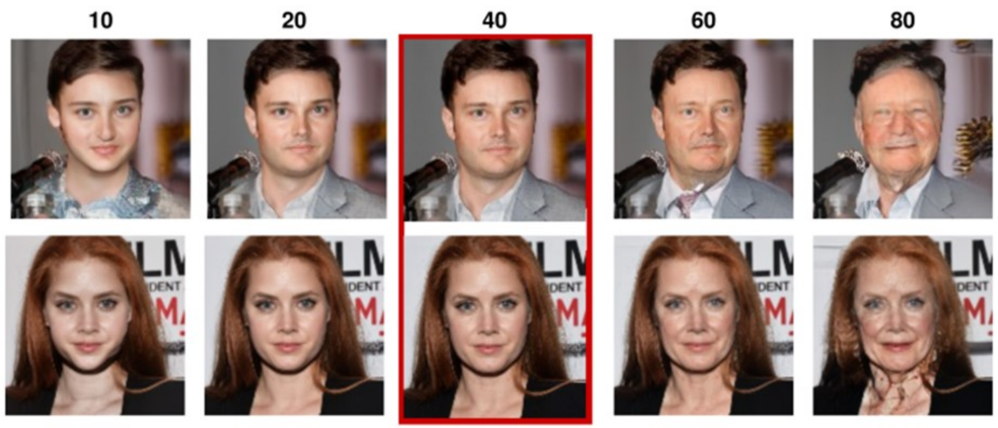
Song, Jiaming, Chenlin Meng, and Stefano Ermon. "Denoising diffusion implicit models." *arXiv preprint arXiv:2010.02502* (2020).

Image Interpolation



DDIM Inversion & Gradient Guidance





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).

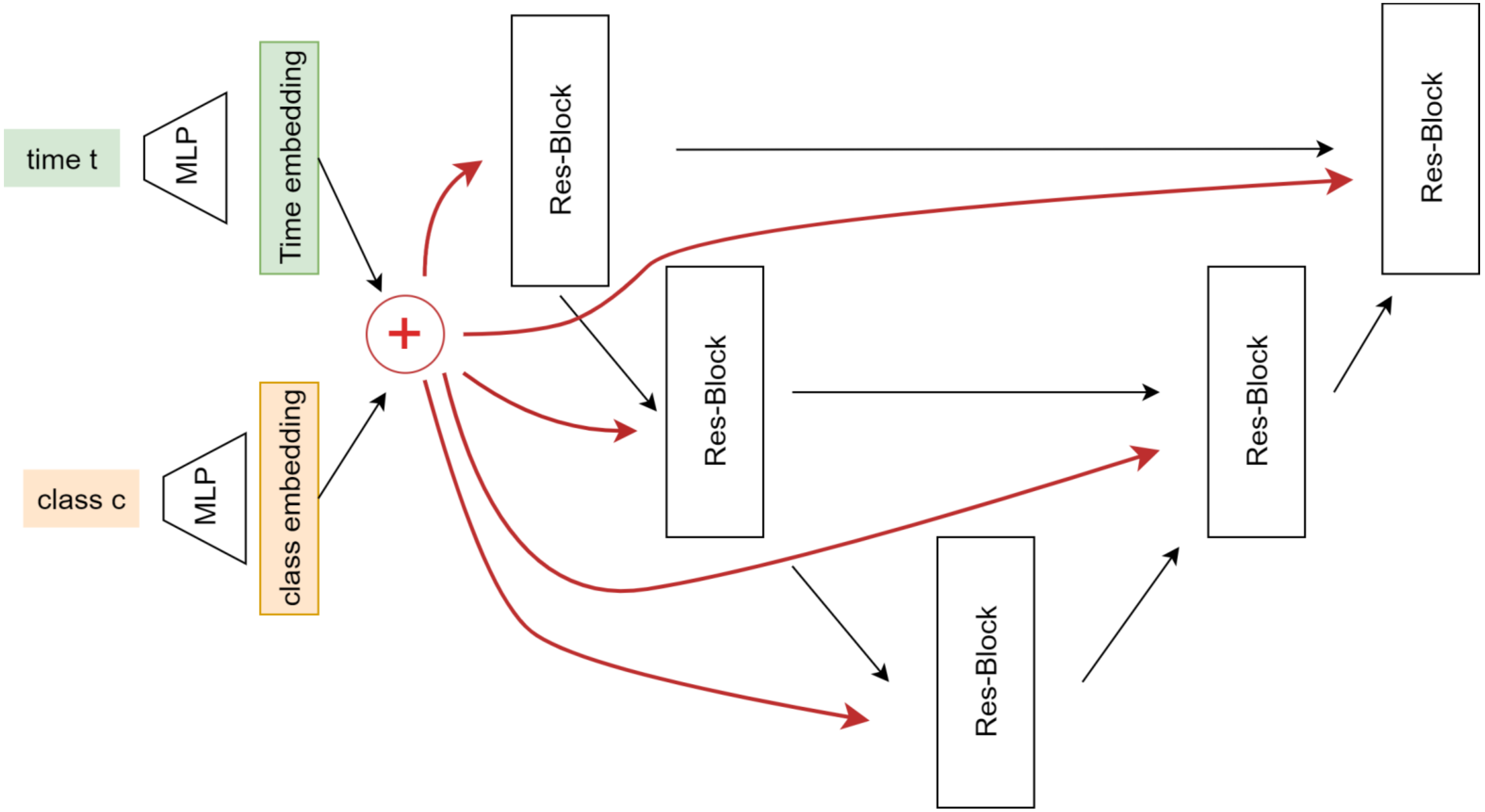
Conditioning

1. **Inference-time**
    1. An inverse problem view
        - ❑ Classifier guidance
    2. DDIM inversion
        - ❑ Interpolation
        - ❑ Gradient guidance

2. **Training-time**
    1. Scalar inputs
    2. Text
    3. Images
    4. ControlNet
    5. DreamBooth

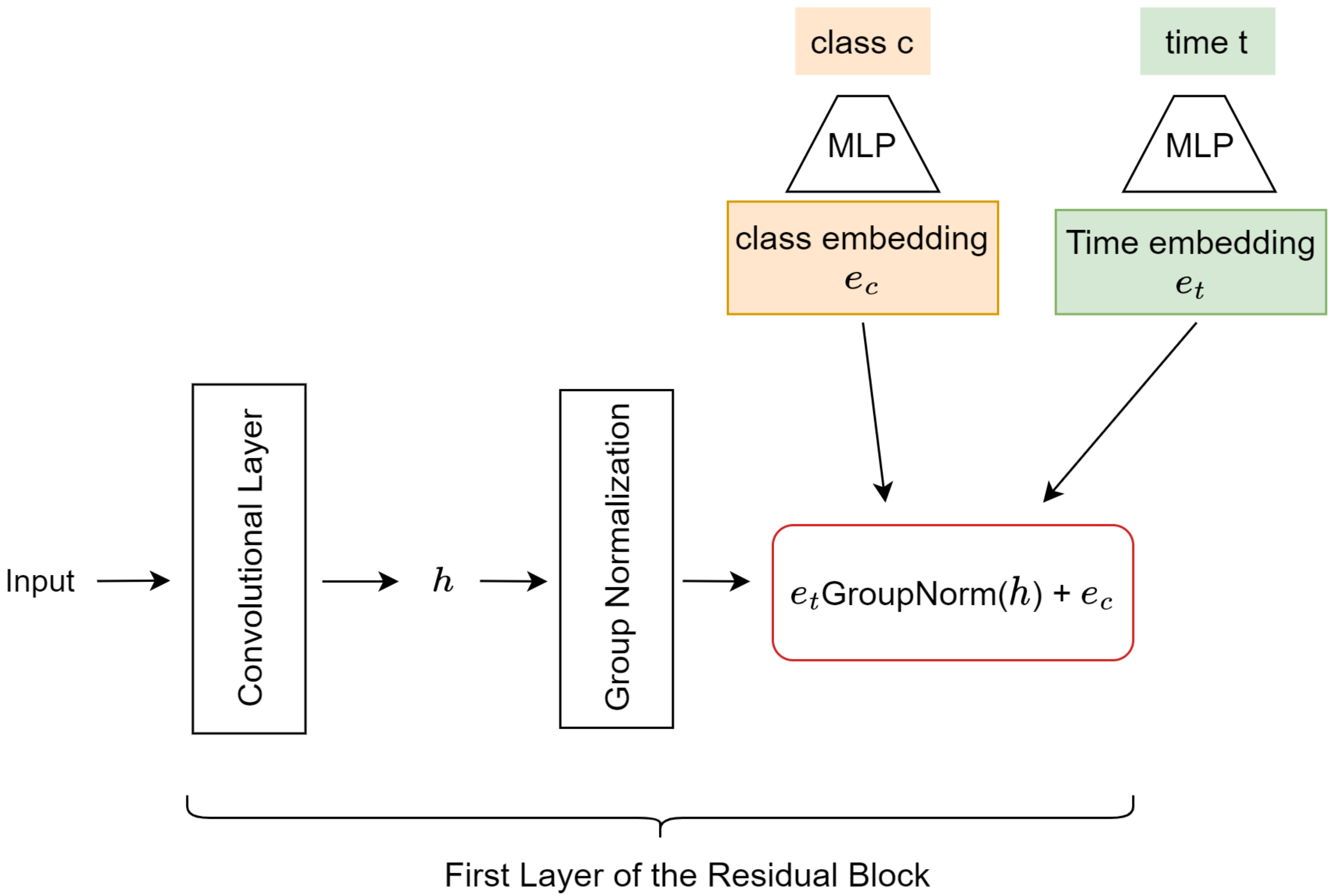
Scalar Conditioning via Spatial Addition

- 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.

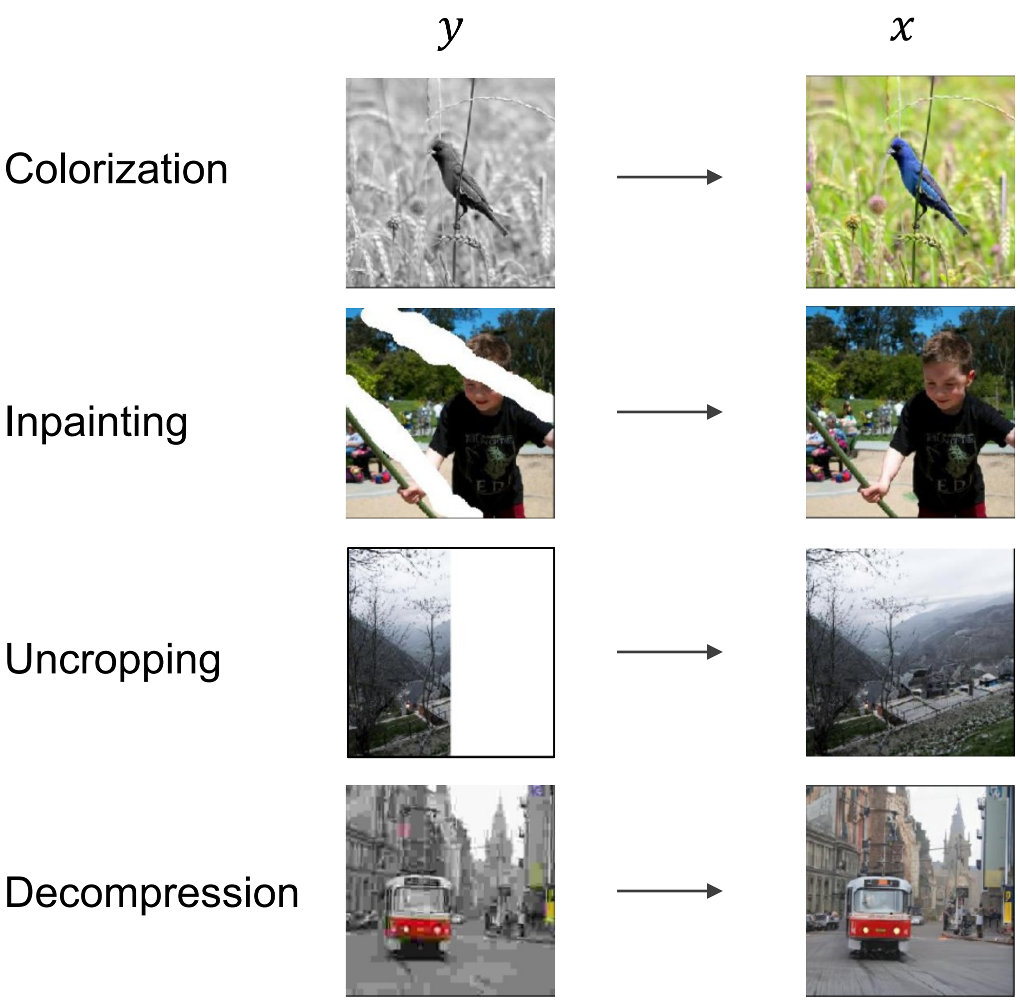
Scalar Conditioning via Adaptive Group Normalization



- 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.

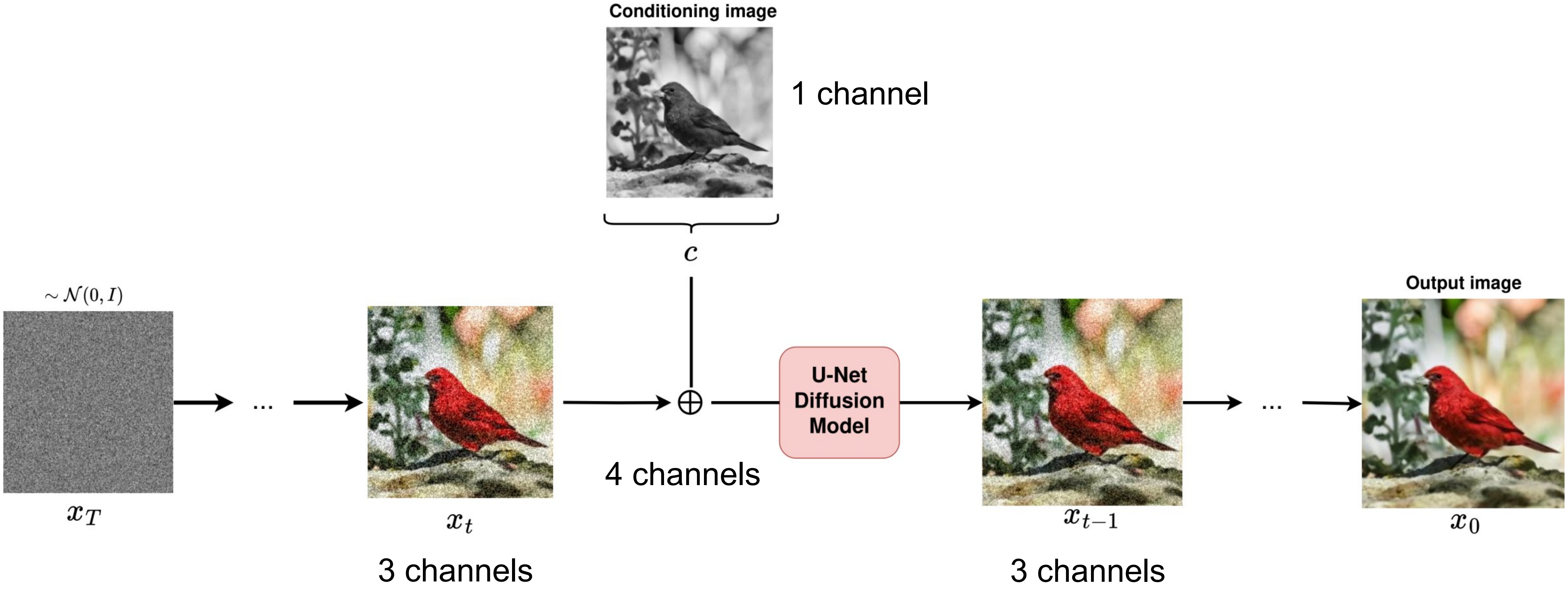
Image Conditioning through Concatenation



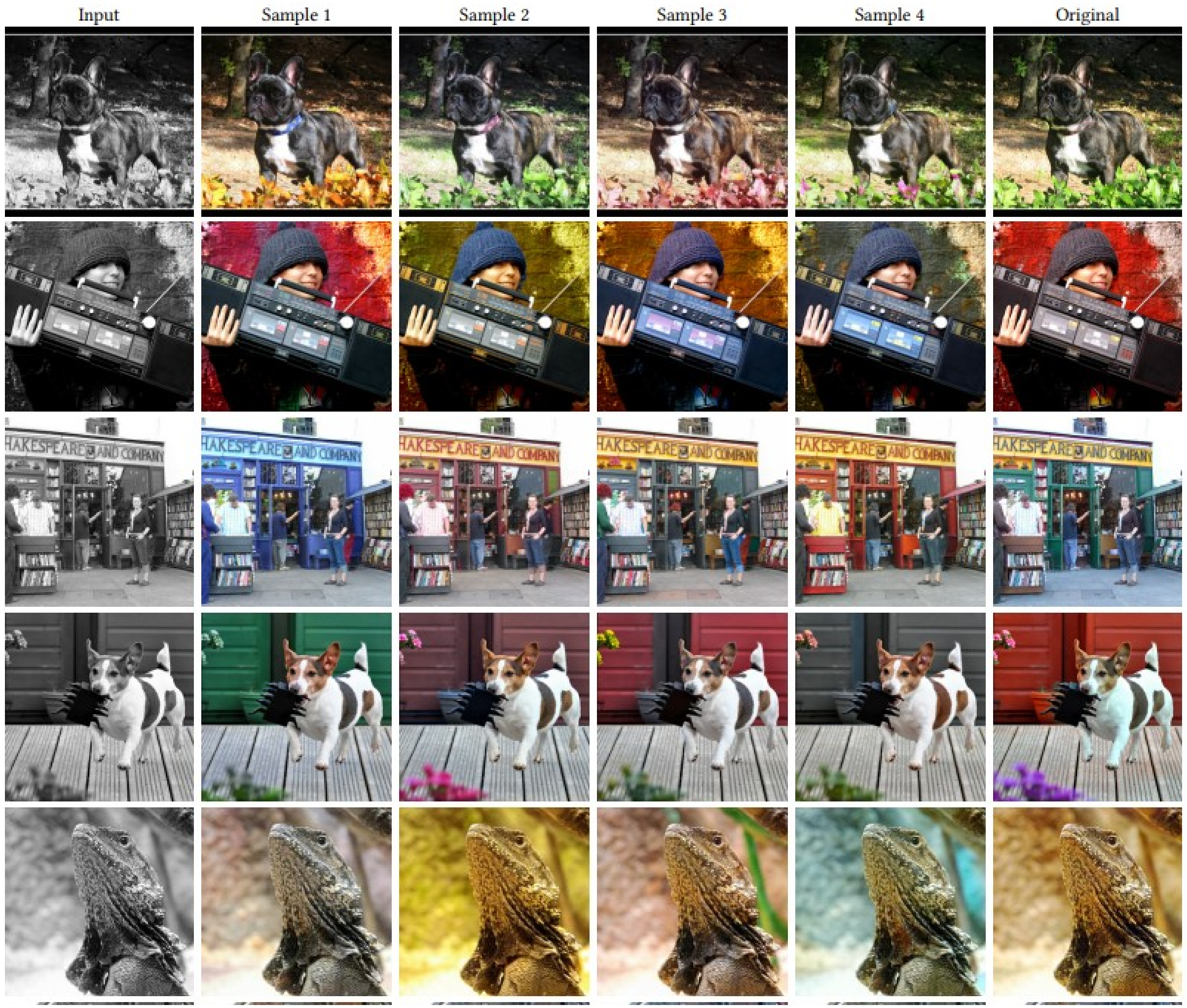
- For image generation of a fake image x, we can use a conditioning image y.
- This requires **paired** training.
- During training and sampling, we add information of the conditioning image x through **channel-wise concatenation.**

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

Image Conditioning through Concatenation



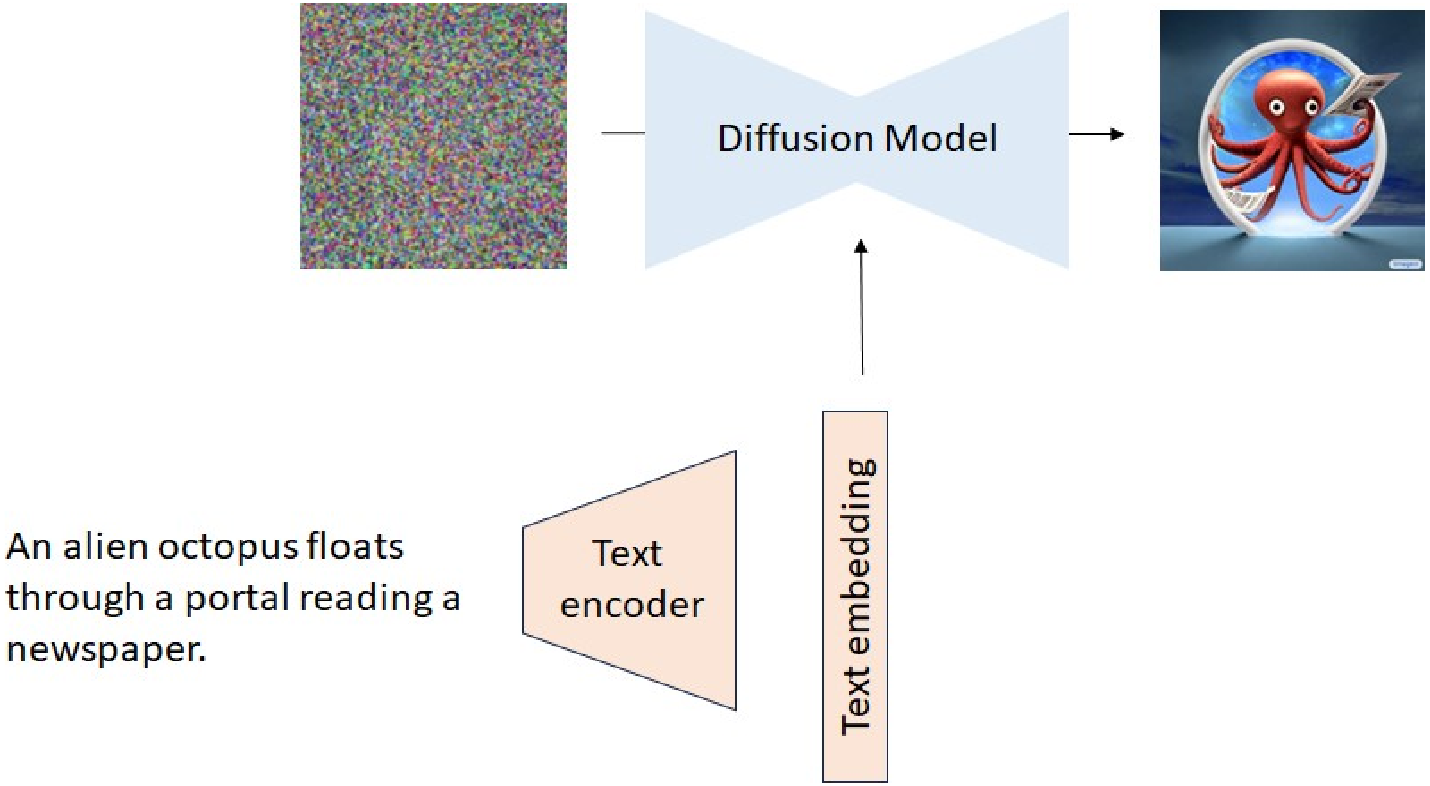
Palette: Image-to-Image Diffusion Models



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."

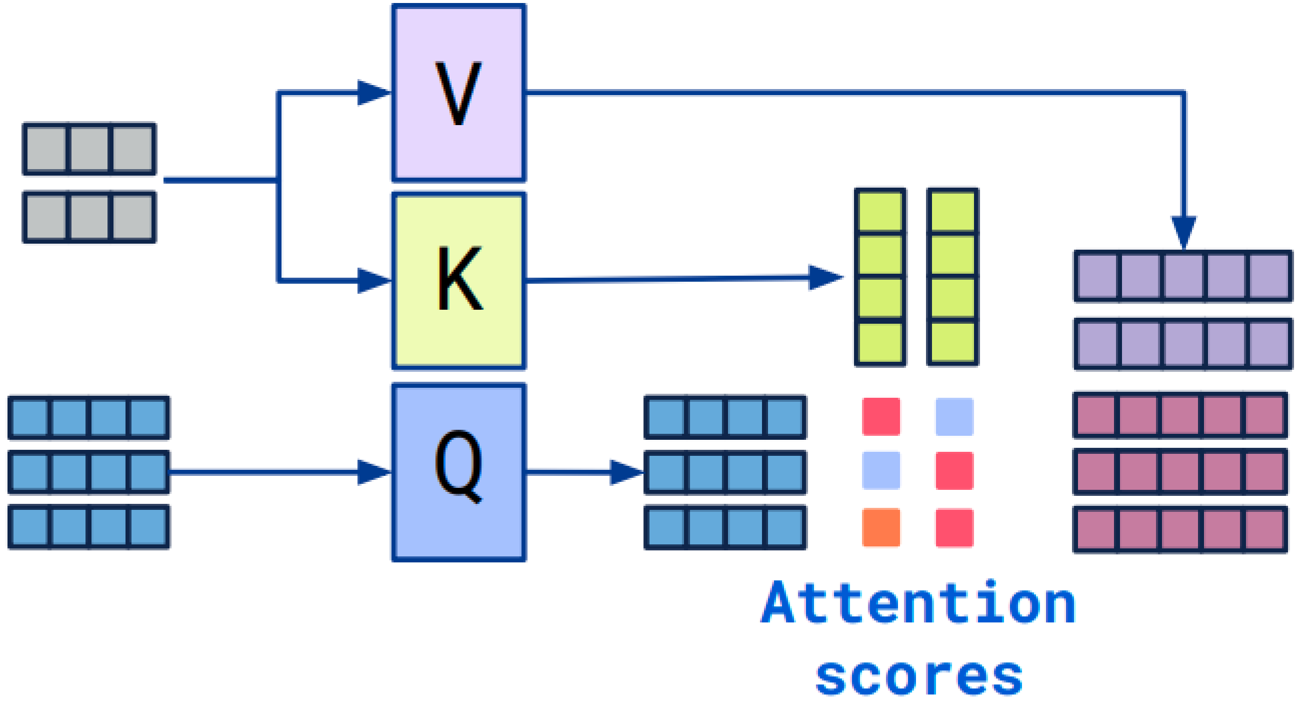


- CLIP
- Dall-E
- Stable Diffusion
- Imagen
- ...

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

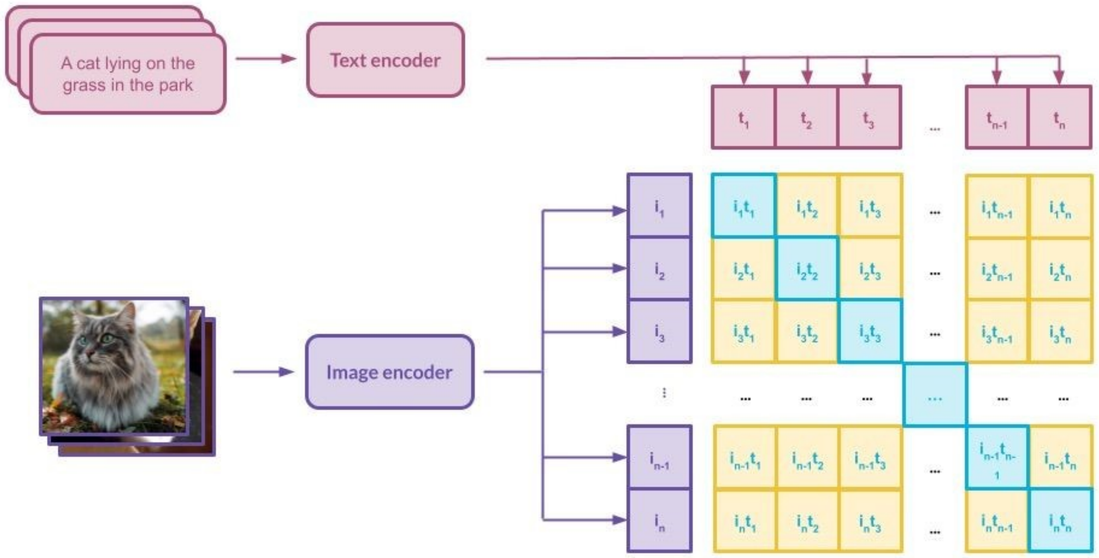
Architecture - Conditioning

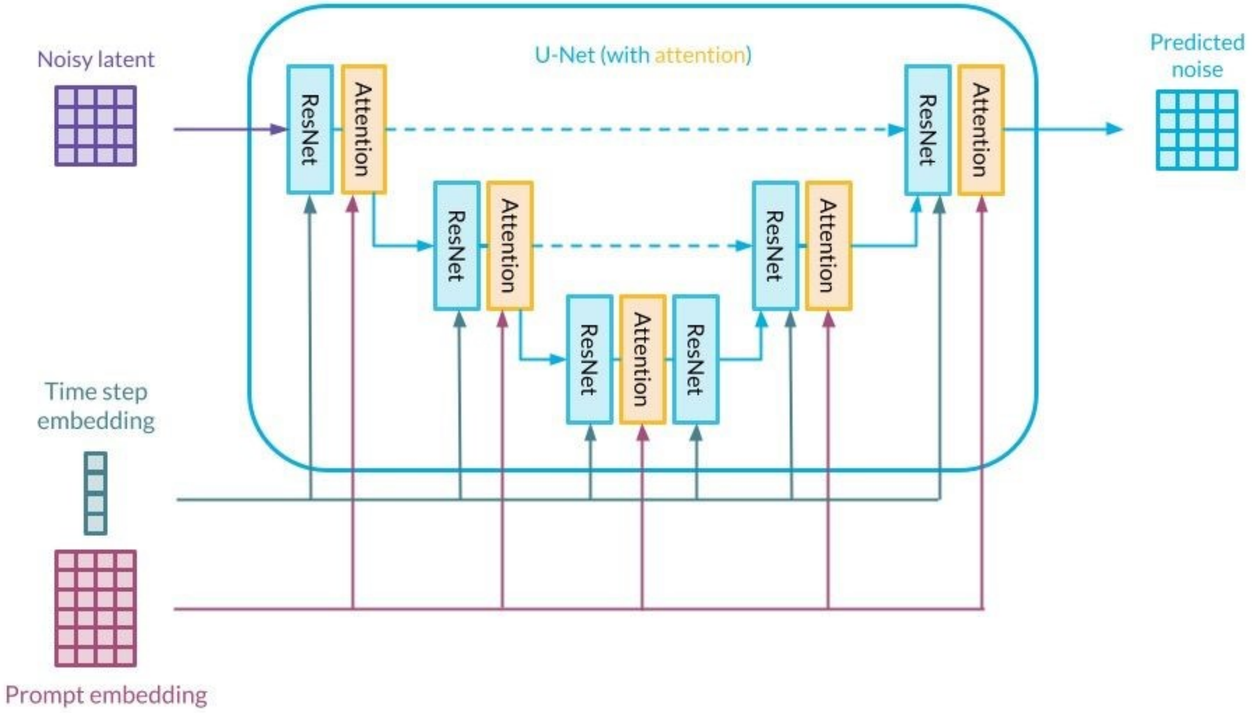
- Transformers **Cross-Attention**
- We use the text embedding to generate the key / value pair.
- We use the image embedding for the query.



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

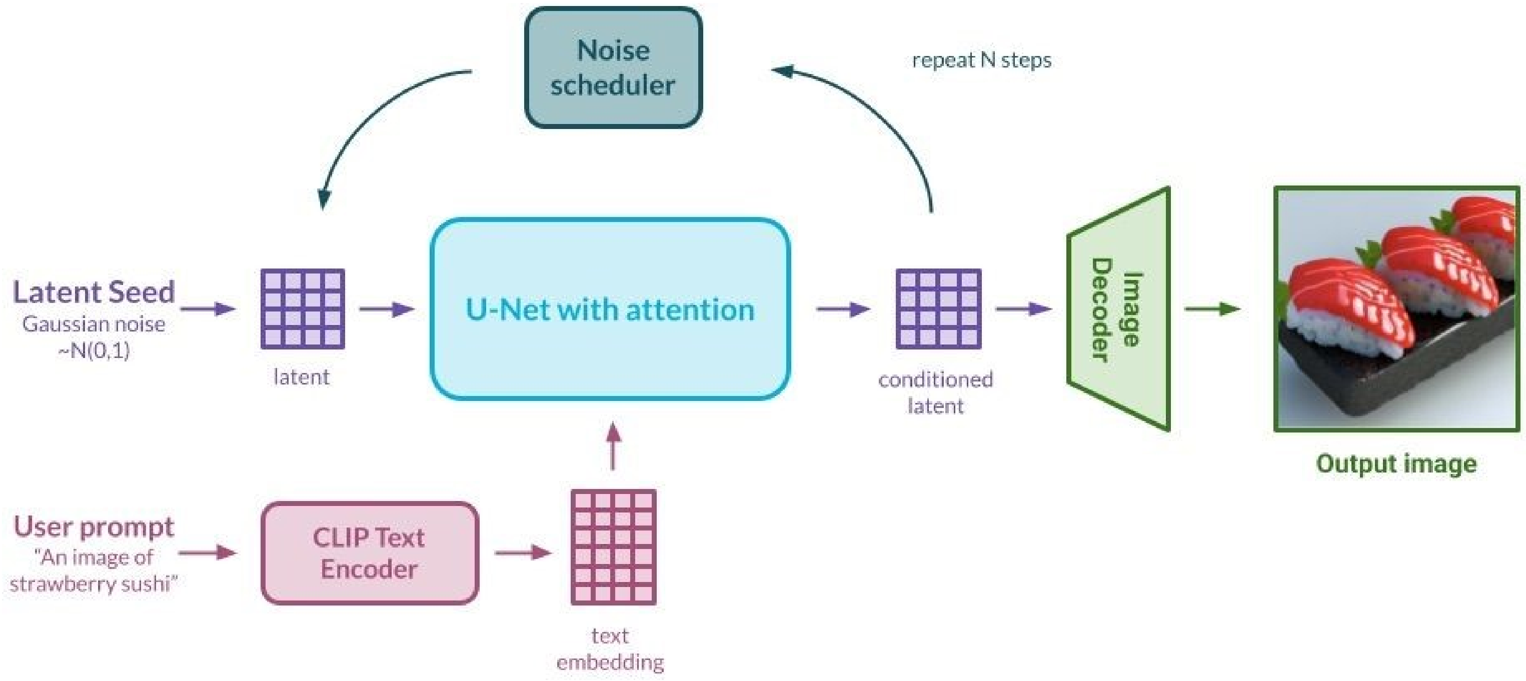
Text Conditioning

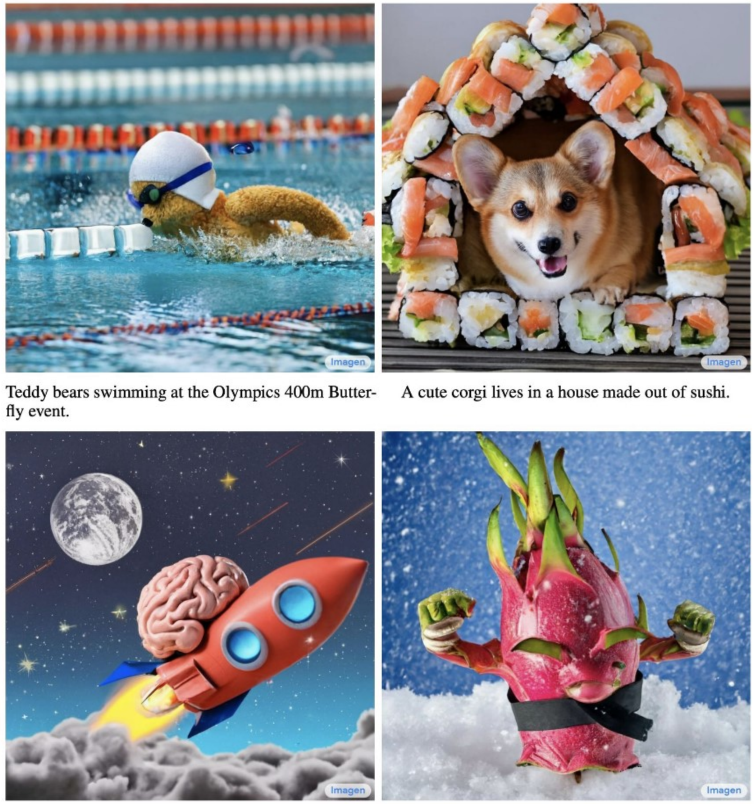




Hertz, Amir, et al. "Prompt-to-prompt image editing with cross attention control." *arXiv preprint arXiv:2208.01626* (2022).
https://deepsense.ai/diffusion-models-in-practice-part-1-the-tools-of-the-trade/

Text Conditioning





Teddy bears swimming at the Olympics 400m Butterfly event.

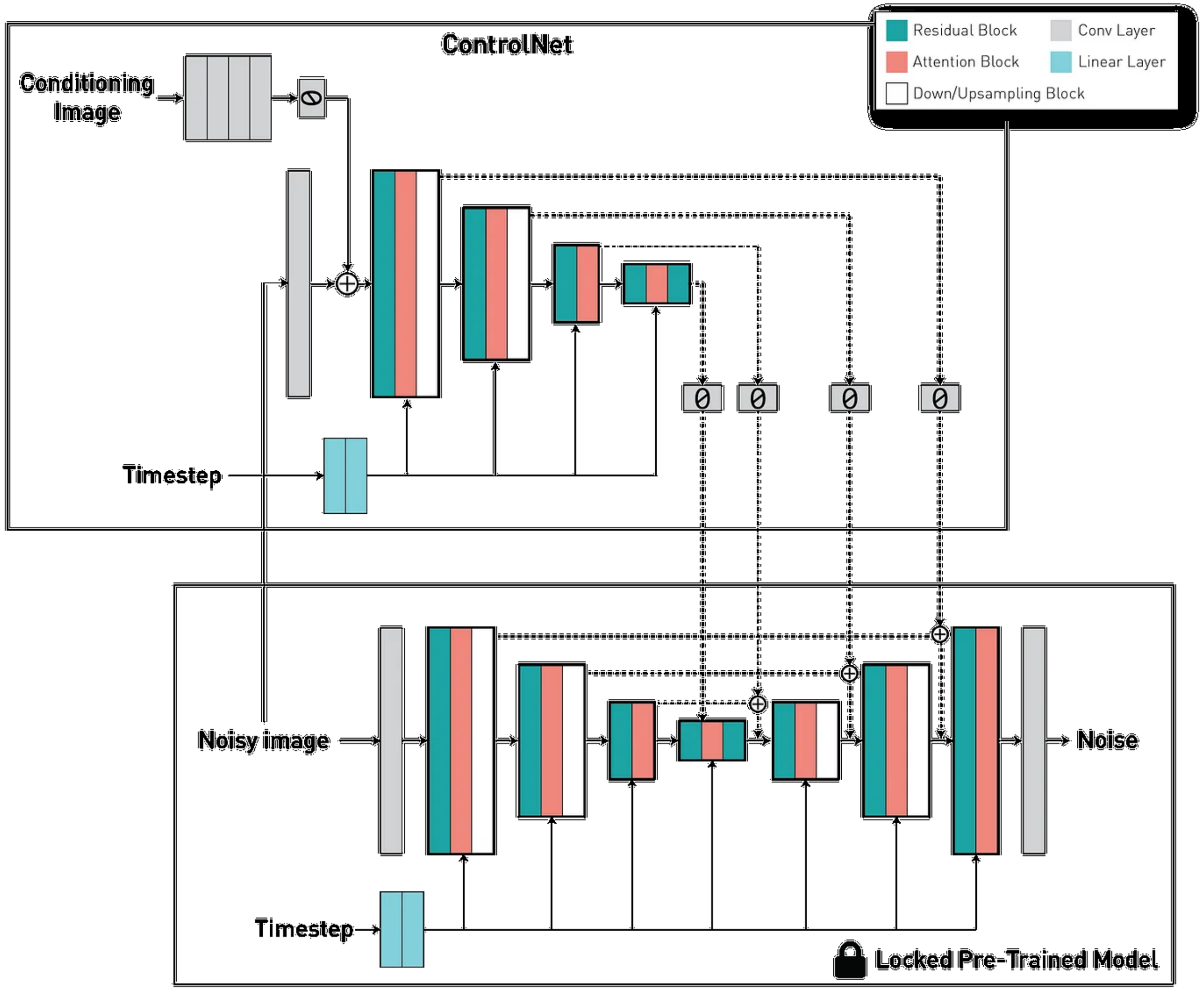
A cute corgi lives in a house made out of sushi.

A brain riding a rocketship heading towards the moon.

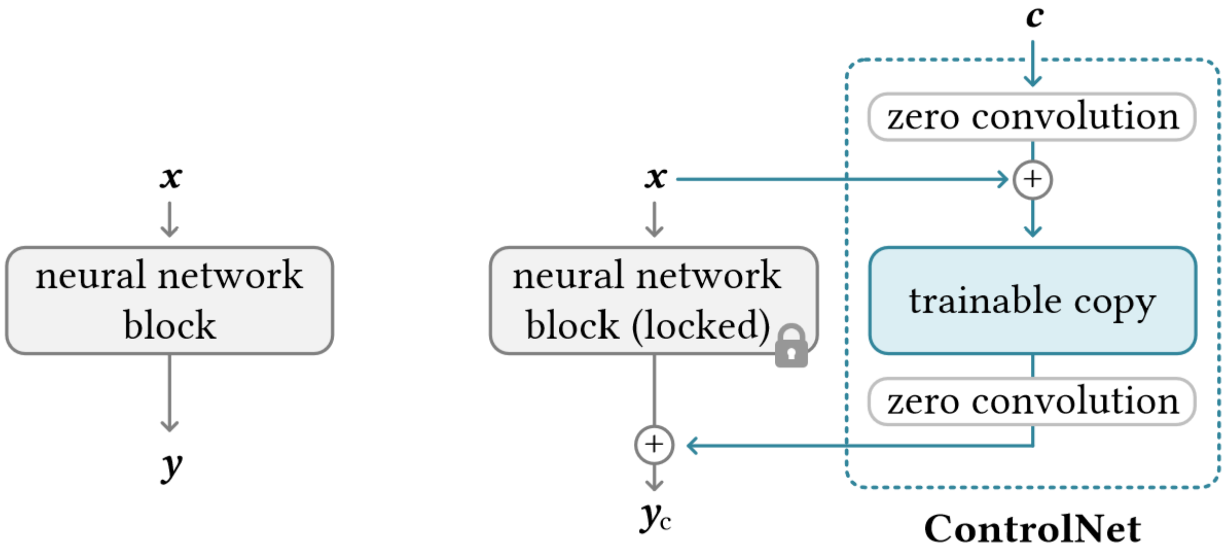
A dragon fruit wearing karate belt in the snow.

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

ControlNet

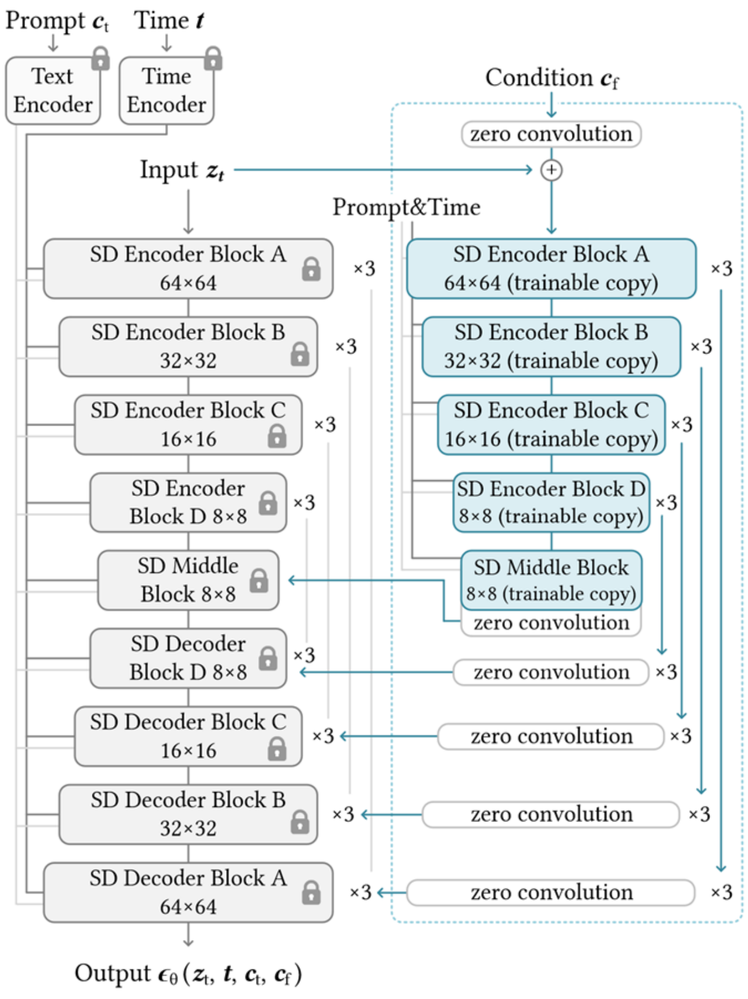


- We pretrain a diffusion model with text prompts.
- We freeze this model.
- We fine-tune a copy conditioned on c.
- We pass information through skip connections.



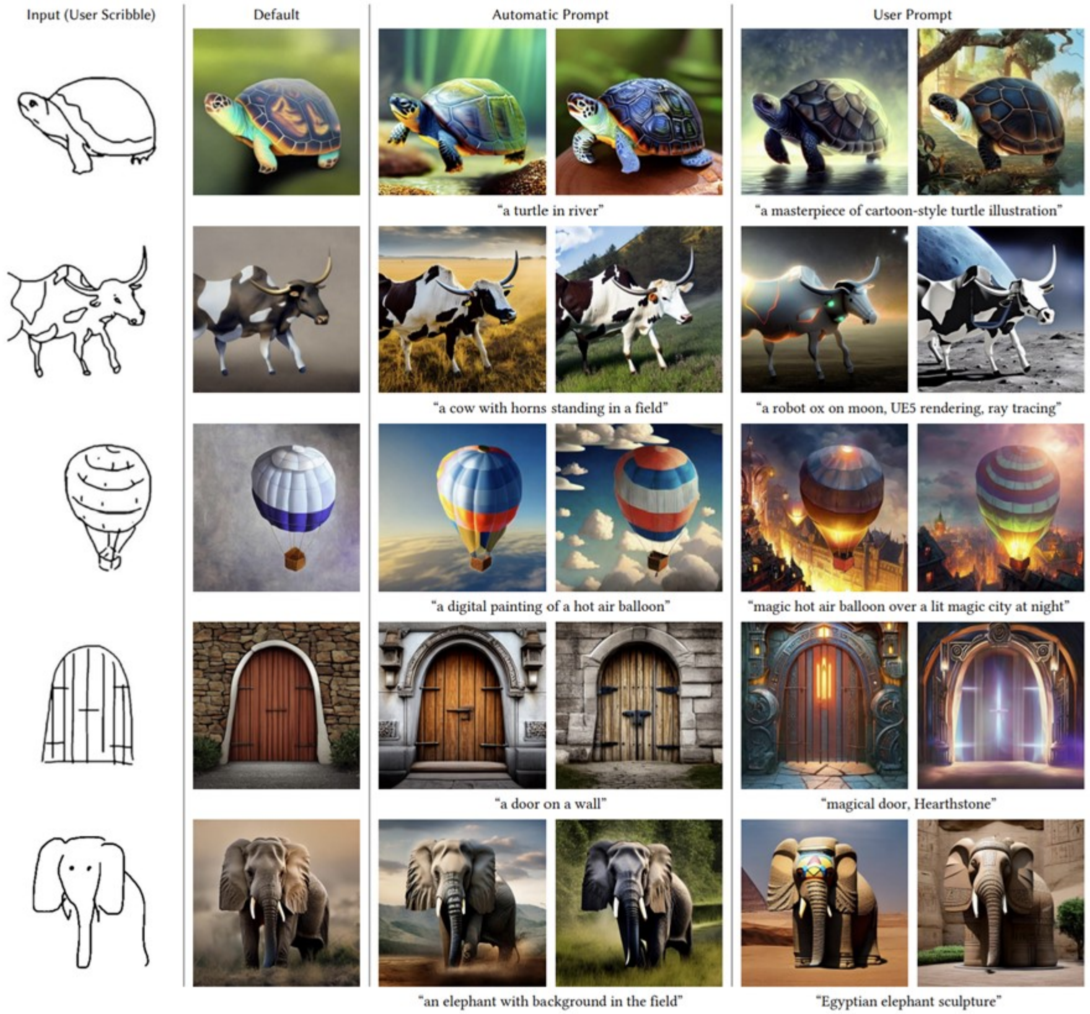
Zhang, Lvmin, Anyi Rao, and Maneesh Agrawala. "Adding conditional control to text-to-image diffusion models." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023.

ControlNet

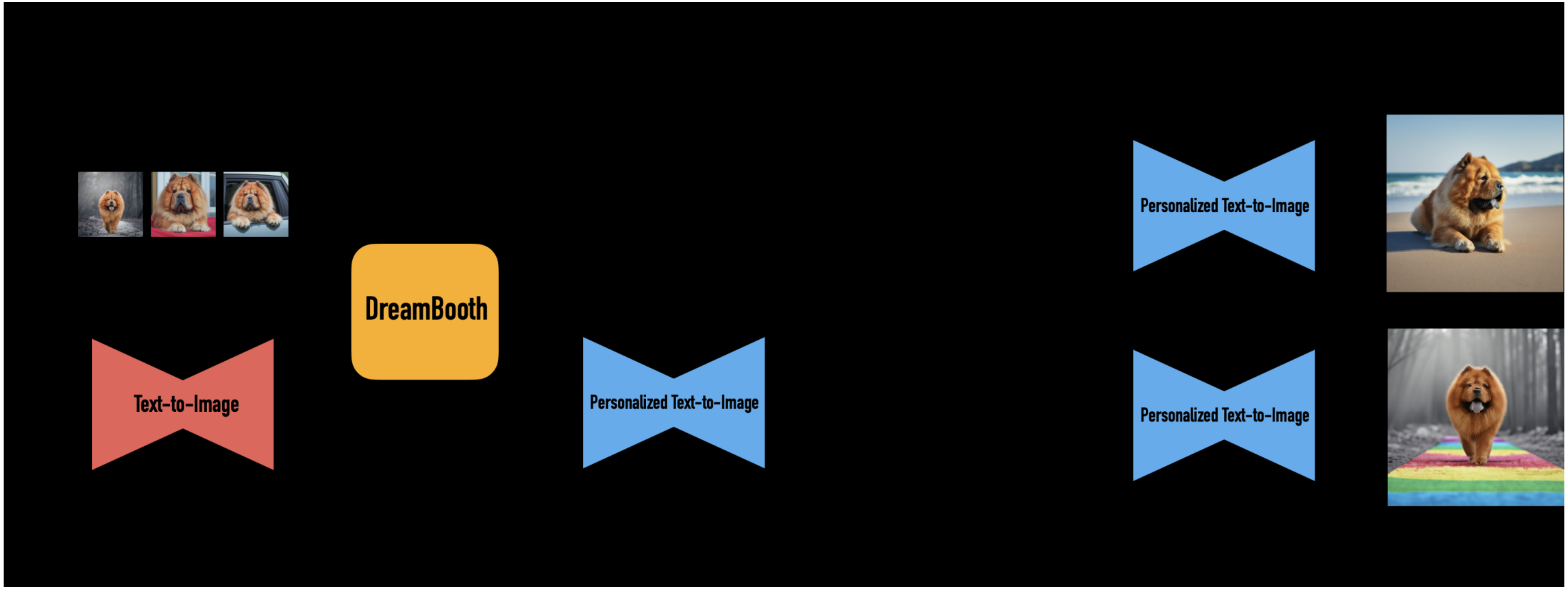


(a) Stable Diffusion (b) ControlNet

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

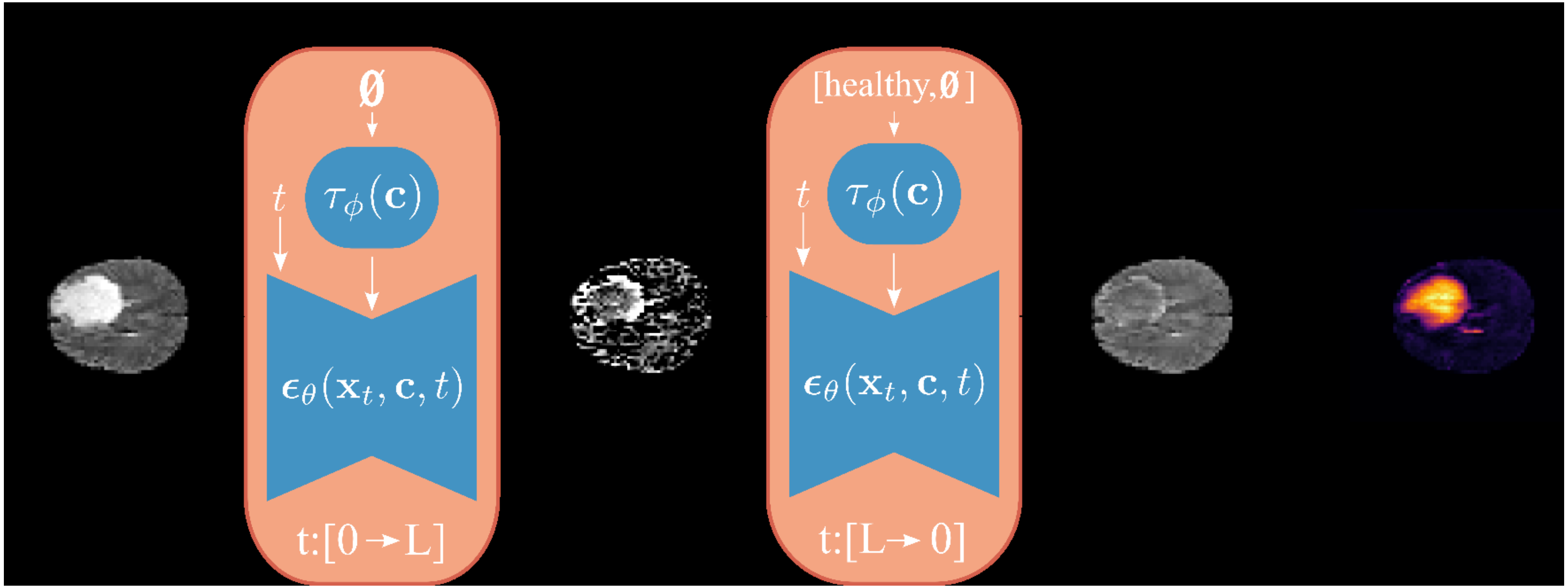
DEMO

MONAI+ Generative Models

DDIM Inversion + Classifier-Free Guidance

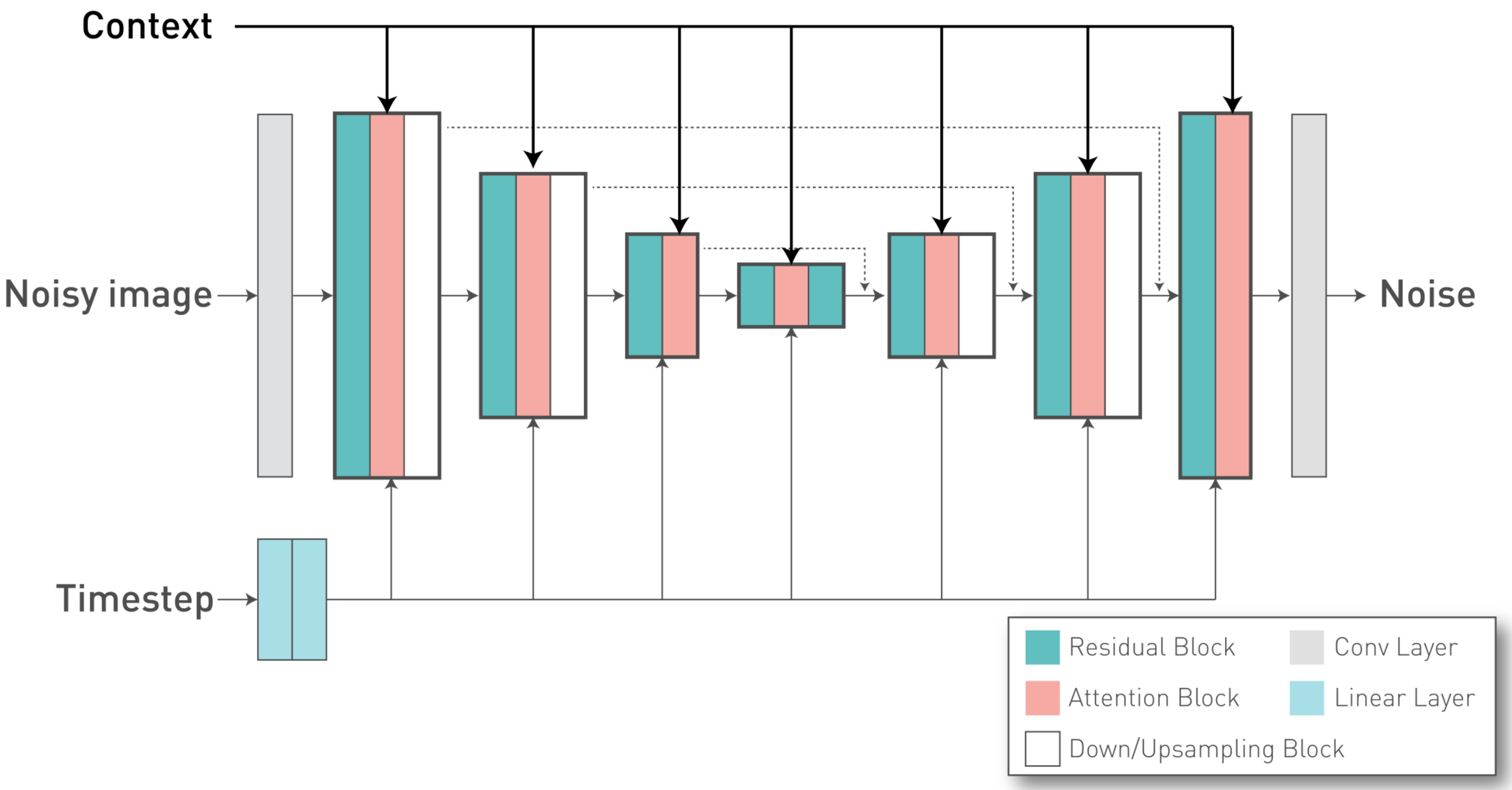
DEMO - Conditioning

1. Scalar Conditioning
2. Classifier-free guidance
3. DDIM Inversion



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

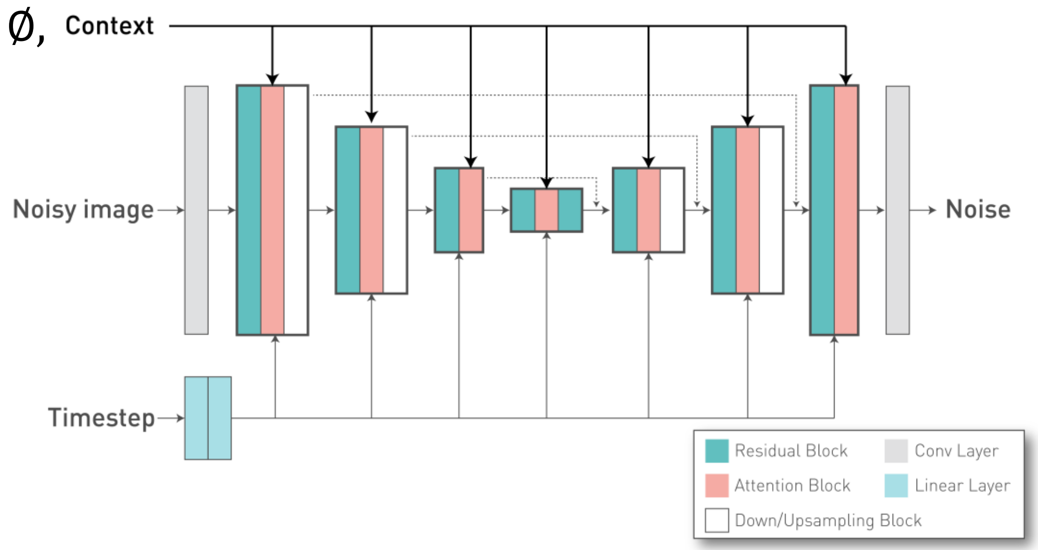
Conditional Unet



Conditional Unet

```python
from generative.networks.nets import DiffusionModelUNet

model = DiffusionModelUNet(
    …
    num_channels=[256, 256, 512],
    attention_levels=[False, True, True],
    num_head_channels=[0, 256, 512],
    with_conditioning=True,
    cross_attention_dim=768,
)
…
noise_pred = model(x=noisy_image,
                   timesteps=timesteps,
                   context=conditioning)
```

Classifier-free Guidance



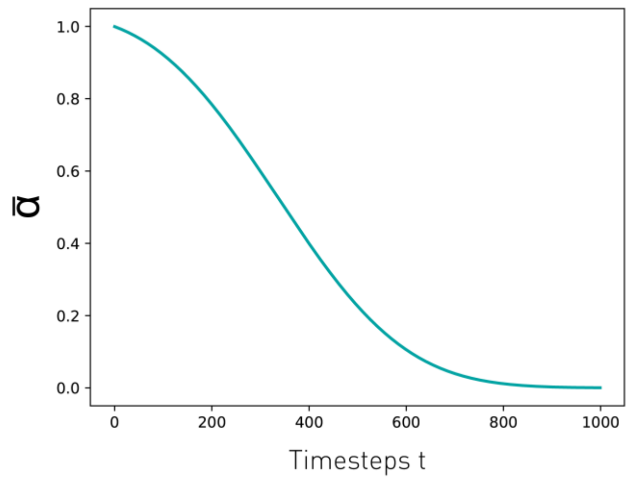
$$\widetilde{\epsilon_\theta}(x_t|y) = \epsilon_\theta(x_t|\emptyset) + w[\epsilon_\theta(x_t|y) - \epsilon_\theta(x_t|\emptyset)]$$

Ho, J., & Salimans, T. (2021). Classifier-Free Diffusion Guidance. In *NeurIPS 2021 Workshop*

Classifier-free Guidance

```
def classifier_free_guidance(noise, t, conditioning, w):

    conditioning = torch.cat([torch.zeros(1), conditioning])
    noise_input = torch.cat([noise] * 2)
    model_output = model(noise_input, timesteps=t, context=conditioning)
    noise_pred_uncond, noise_pred_text = model_output.chunk(2)

    noise_pred = noise_pred_uncond + w * (noise_pred_text - noise_pred_uncond)

    return noise_pred
```

Noise Schedulers




```python
from generative.networks.schedulers import DDIMScheduler


scheduler = DDIMScheduler(
  num_train_timesteps=1000,
  beta_schedule=“scaled_linear",
  beta_start=0.0005,
  beta_end=0.0195,
)
```

Training

```
…
for batch in train_loader:
    # classes {1: unhealthy, 2: unhealthy}
    images, classes = batch["image"], batch["classes"]
    # dropout classes 15% of the time
    classes = classes * (torch.rand_like(classes) > 0.15)
    optimizer.zero_grad(set_to_none=True)

    noise = torch.randn_like(images).to(device)
    timesteps = torch.randint(0, scheduler.num_train_timesteps,(images.shape[0],))
    noisy_image = scheduler.add_noise(original_samples=images,
                                      noise=noise,
                                      timesteps=timesteps,)

    noise_pred = model(x=noisy_image, timesteps=timesteps, context=classes)

    loss = F.mse_loss(noise_pred.float(), noise.float())
    …
```

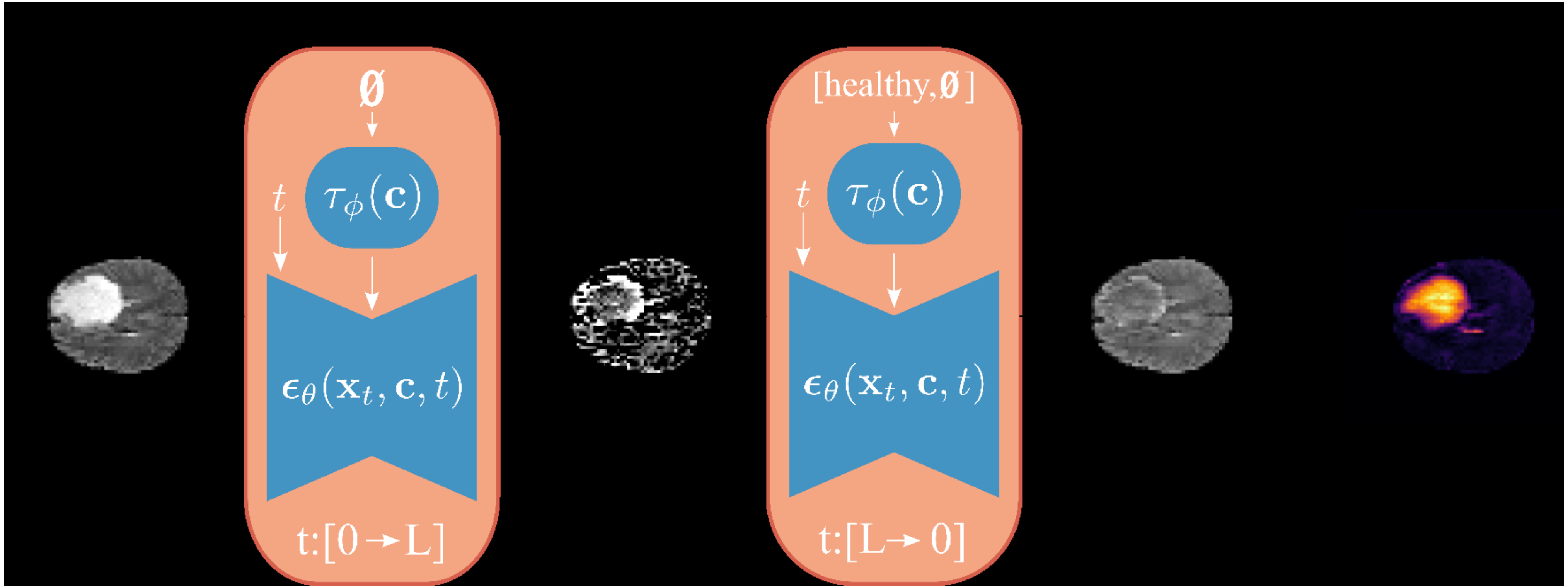
Sampling – DDIM Inversion + Guidance

```
L = 200
conditioning = torch.zeros(1)
scheduler.set_timesteps(num_inference_steps=1000)
current_img = batch["image"]
for t in range(L): # 0 -> L timesteps
    with torch.no_grad():
        model_output = model(current_img, timesteps=(t,), context=conditioning)
    current_img, _ = scheduler.reversed_step(model_output, t, current_img)
latent_space_L = current_img
```

```
conditioning = torch.ones(1) # Manipulate to be healthy
noise = latent_space_L
for i in range(L):
    t = L - i # t goes from L -> 0
    noise_pred = classifier_free_guidance(noise, t, conditioning, w)
    noise, _ = scheduler.step(noise_pred, t, noise)
image = noise
```

DEMO – Recap

1. Scalar Conditioning
2. Classifier-free guidance
3. DDIM Inversion



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

Part 2 – Q&A

Part 3 – Medical Image Applications

Image Reconstruction

Image Registration

Image Segmentation

Anomaly Detection

Inpainting

Image-to-Image Translation

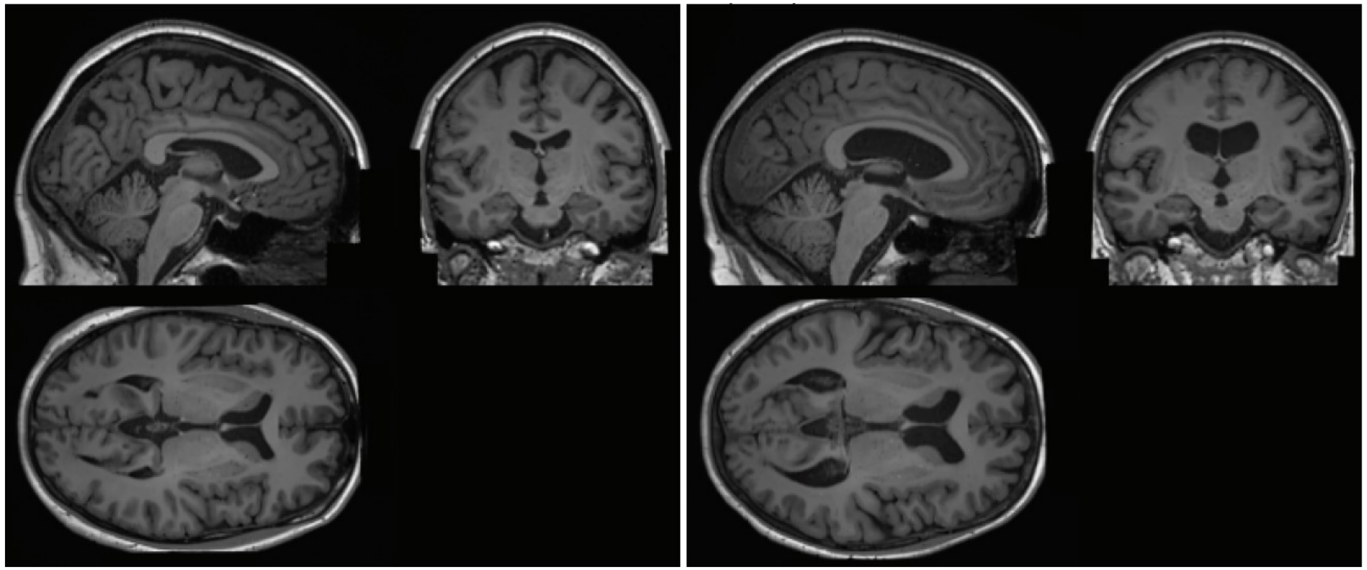
Image Synthesis

Image synthesis

Examples from the community

The simple setup of the problem

Real



Synthetic

PAPERS

Pinaya et al (2022) Brain Imaging Generation with Latent Diffusion Models. MICCAI workshop
Kim et al. (2022) Diffusion Deformable Model for 4D Temporal Medical Image Generation. MICCAI
Khader et al. (2022) Medical Diffusion -- Denoising Diffusion Probabilistic Models for 3D Medical Image Generation. arXiv:2211.03364
Packhäuser et al. (2022) Generation of Anonymous Chest Radiographs Using Latent Diffusion Models for Training Thoracic Abnormality Classification Systems. arXiv:2211.01323
Ali et al. (2022) Spot the fake lungs: Generating Synthetic Medical Images using Neural Diffusion Models. arXiv:2211.00902
Rouzrokh et al. (2022) Multitask Brain Tumor Inpainting with Diffusion Models: A Methodological Report. arXiv:2210.12113
Chambon et al (2022) Adapting Pretrained Vision-Language Foundational Models to Medical Imaging Domains. arXiv:2210.04133
Lyu et al. (2022) Conversion Between CT and MRI Images Using Diffusion and Score-Matching Models. arXiv:2209.12104
Ozbey et al. (2022) Unsupervised Medical Image Translation with Adversarial Diffusion Models. arXiv:2207.08208
Meng et al. (2022) A Novel Unified Conditional Score-based Generative Framework for Multi-modal Medical Image Completion. arXiv:2207.03430

Why? Medical Image Data is Scarce

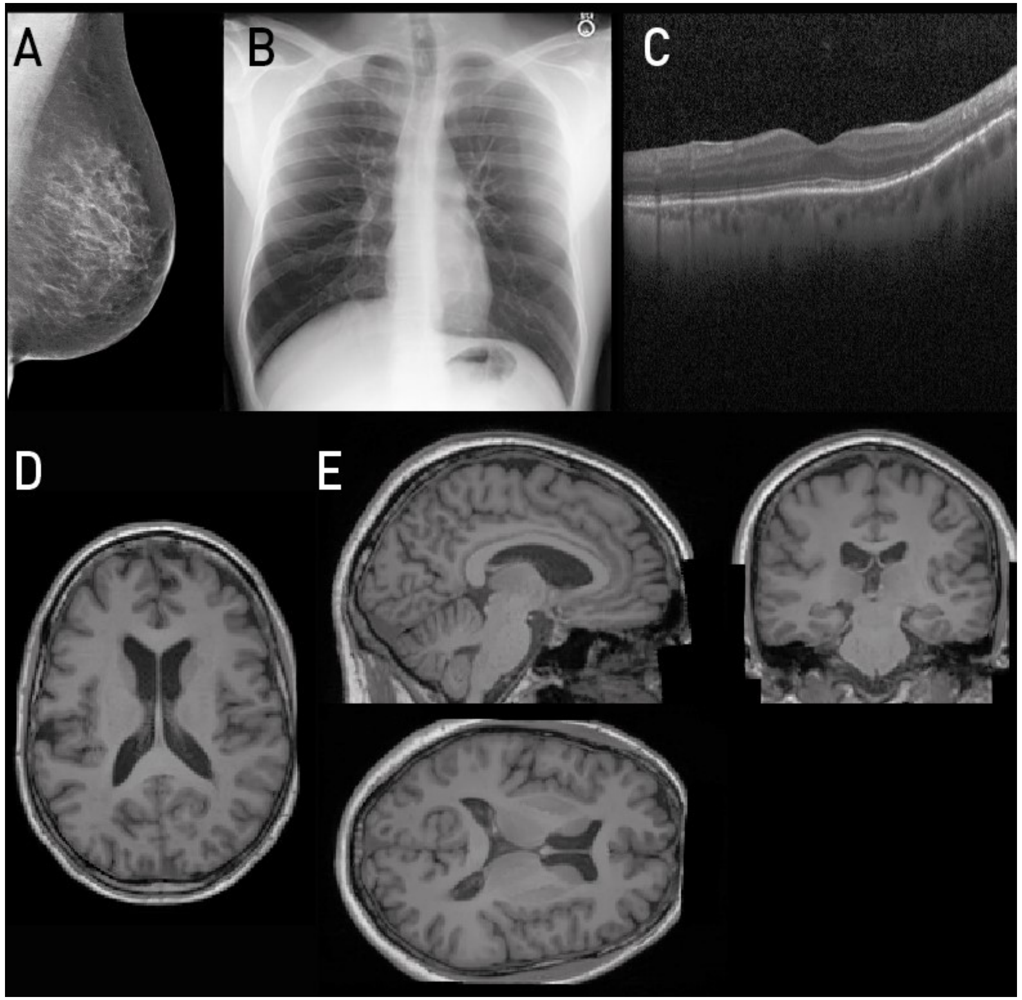
Privacy concerns → Limited data

Data Provider

Data Users

Use **of Synthetic Data**

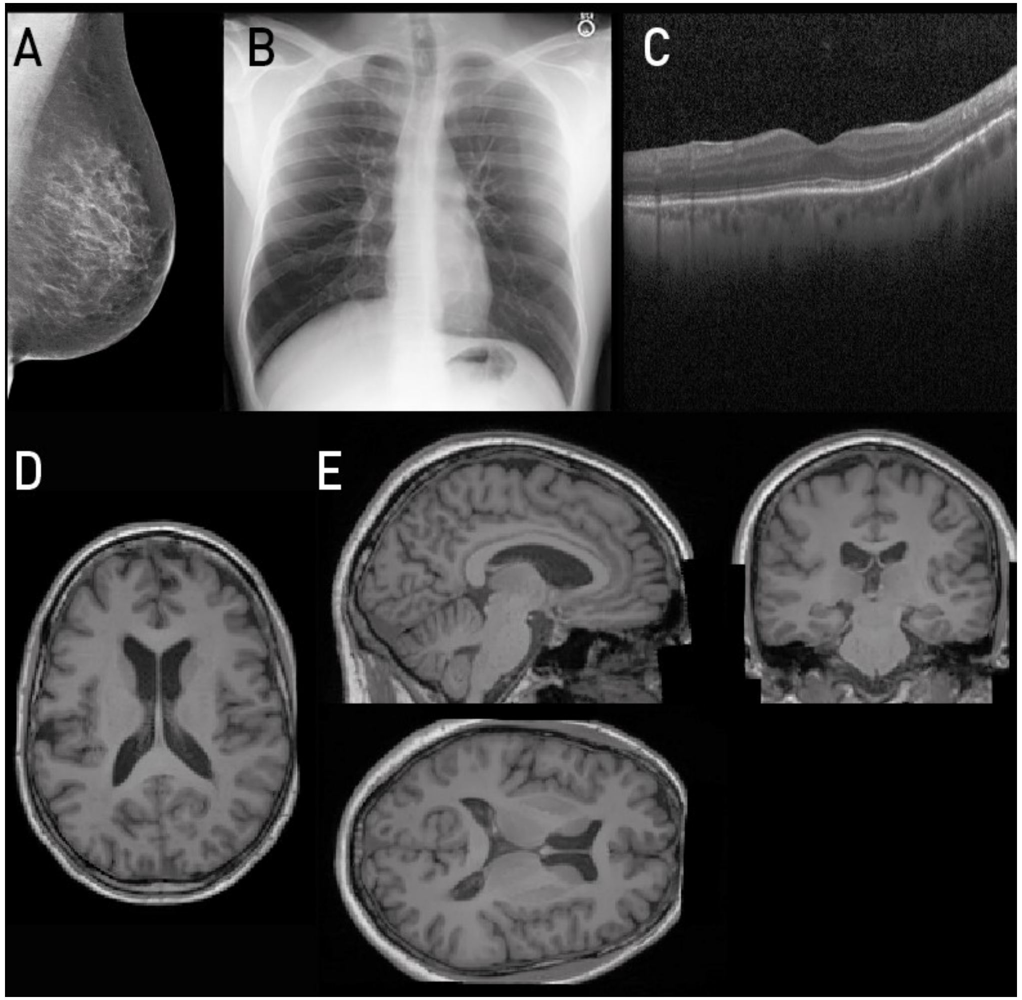
- ❑ Full "private" training
- ❑ Data augmentation
- ❑ Test-time augmentation
- ❑ Testing edge cases



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

Evaluation **of Synthetic Data**

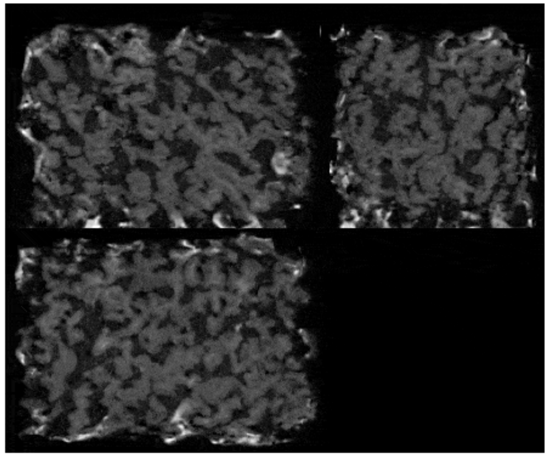
- ❑ Realism
- ❑ Diversity
- ❑ Privacy
- ❑ Benchmark

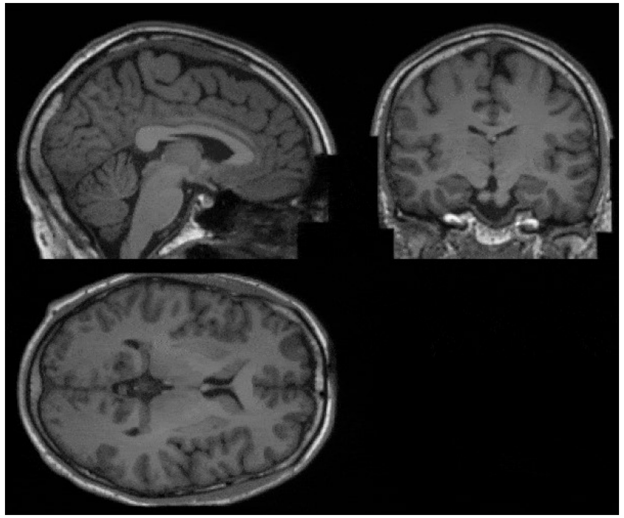


Pinaya, Walter HL, et al. "Generative AI for Medical Imaging: extending the MONAI Framework." arXiv preprint arXiv:2307.15208 (2023).

Generating high-resolution 3D brain data

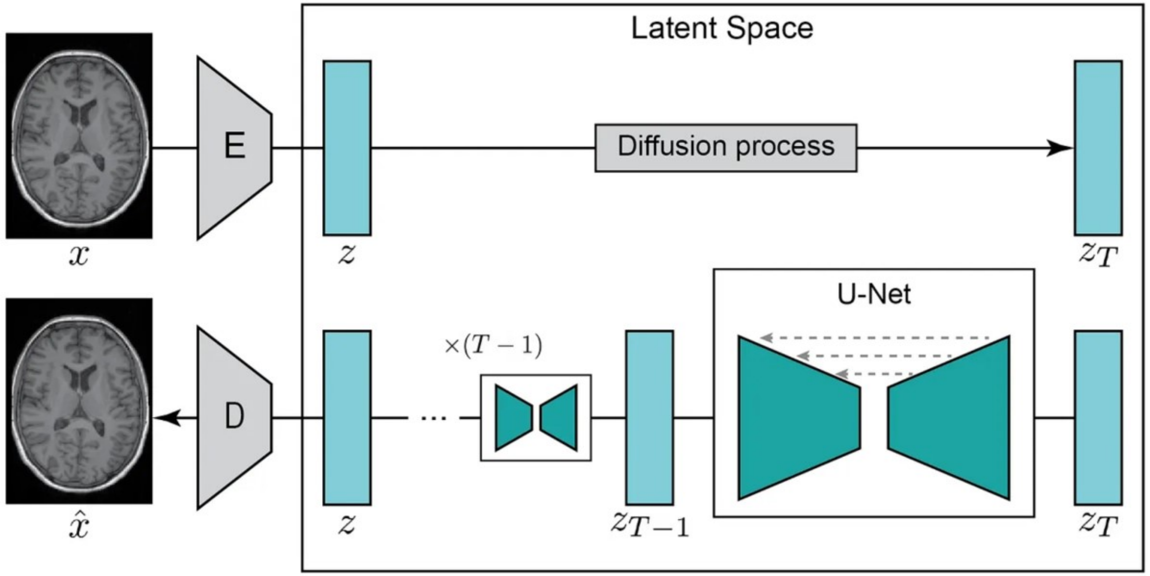
- Latent Diffusion Models trained on data from UK Biobank (N = 31,740)
  - T1 MRI brain images with 1 mm^3 voxel size (160 × 224 × 160 voxels)
- Conditioned on covariates, such as:
  - Age
  - Gender
  - Ventricular and Brain volumes





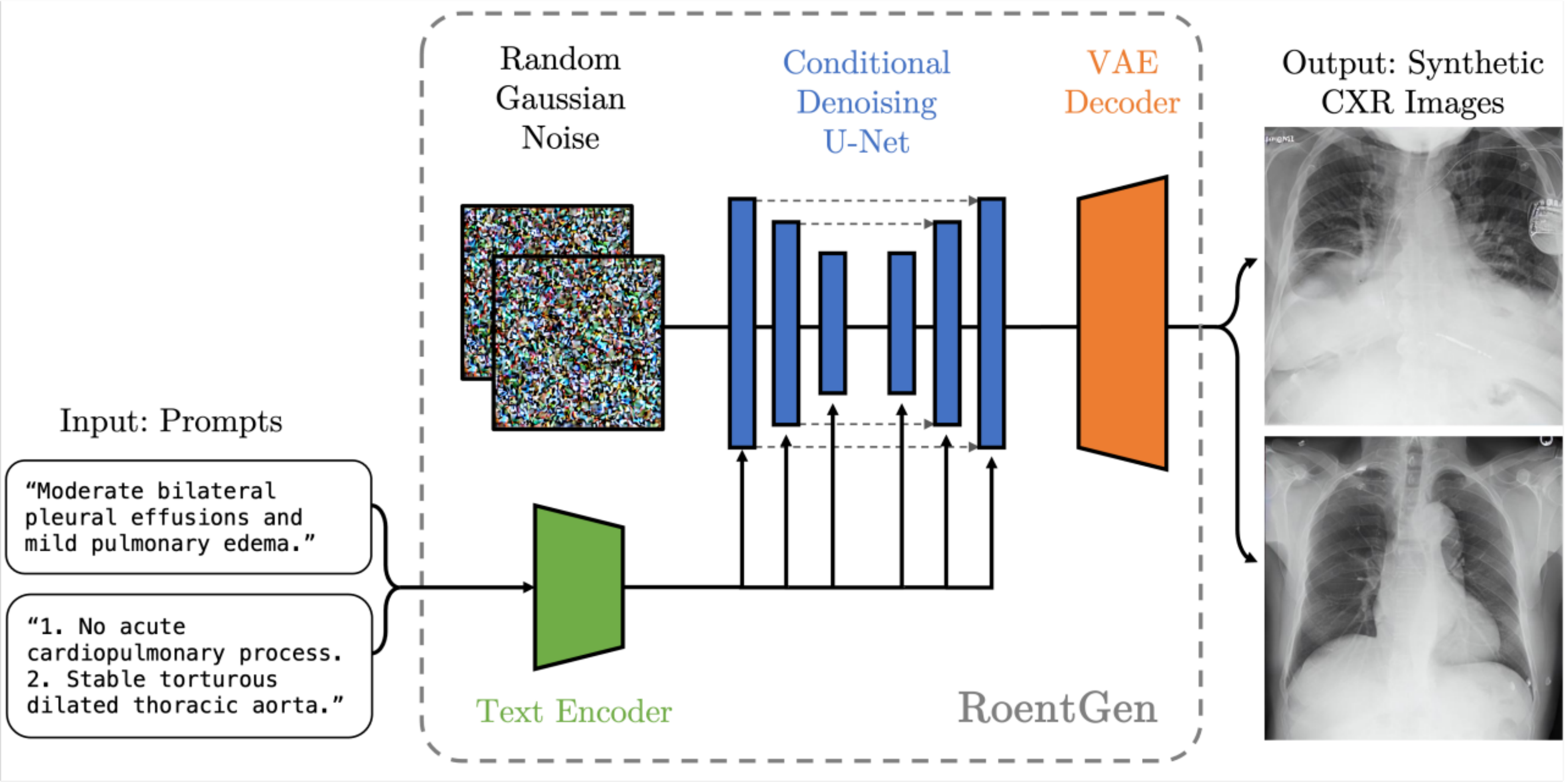
1. Pinaya et al (2022). Brain Imaging Generation with Latent Diffusion Models. DOI:10.1007/978-3-031-18576-2_12

Diffusion Model in the Latent Space



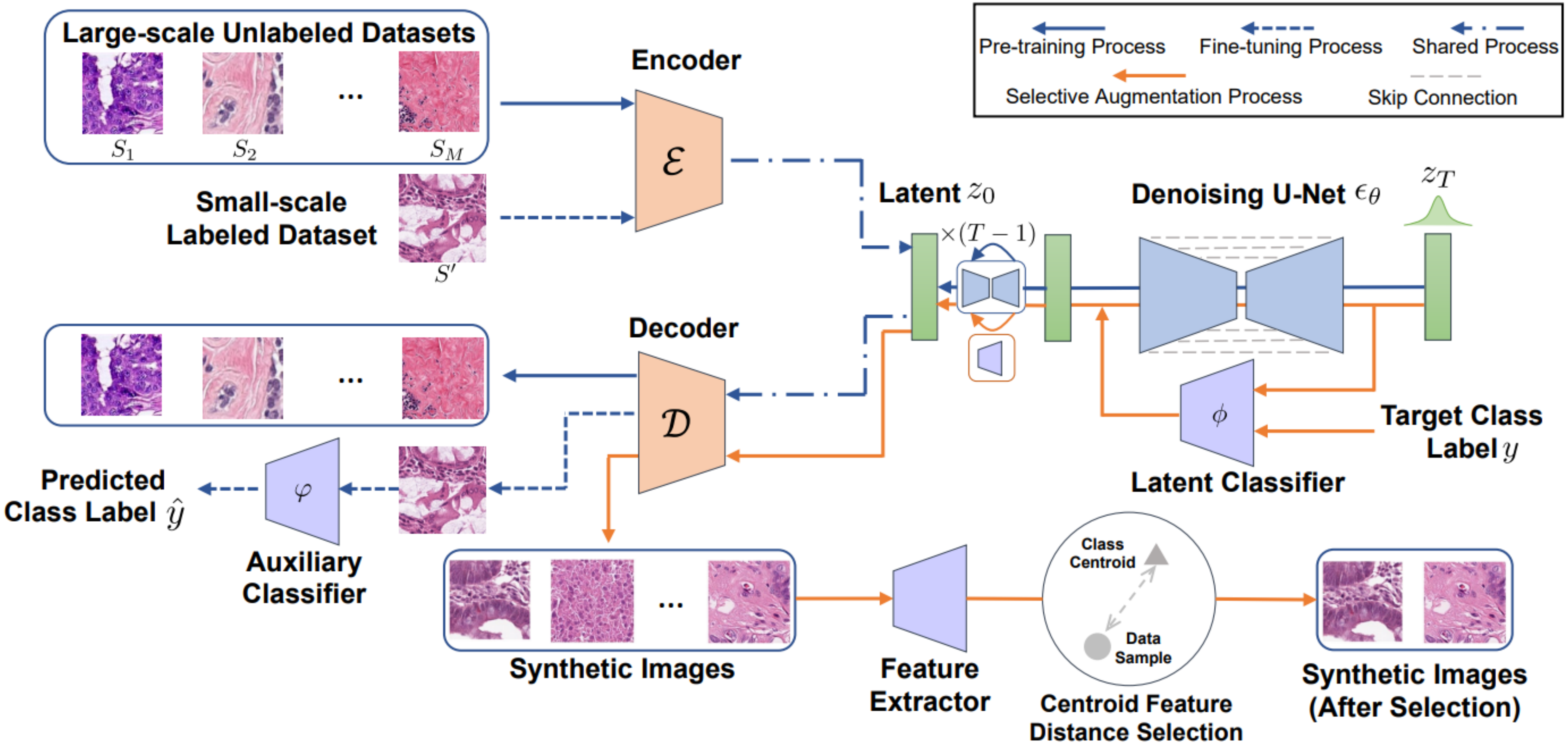
Rombach, Robin, et al. "High-resolution image synthesis with latent diffusion models." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2022.

Fine-tuning Stable Diffusion



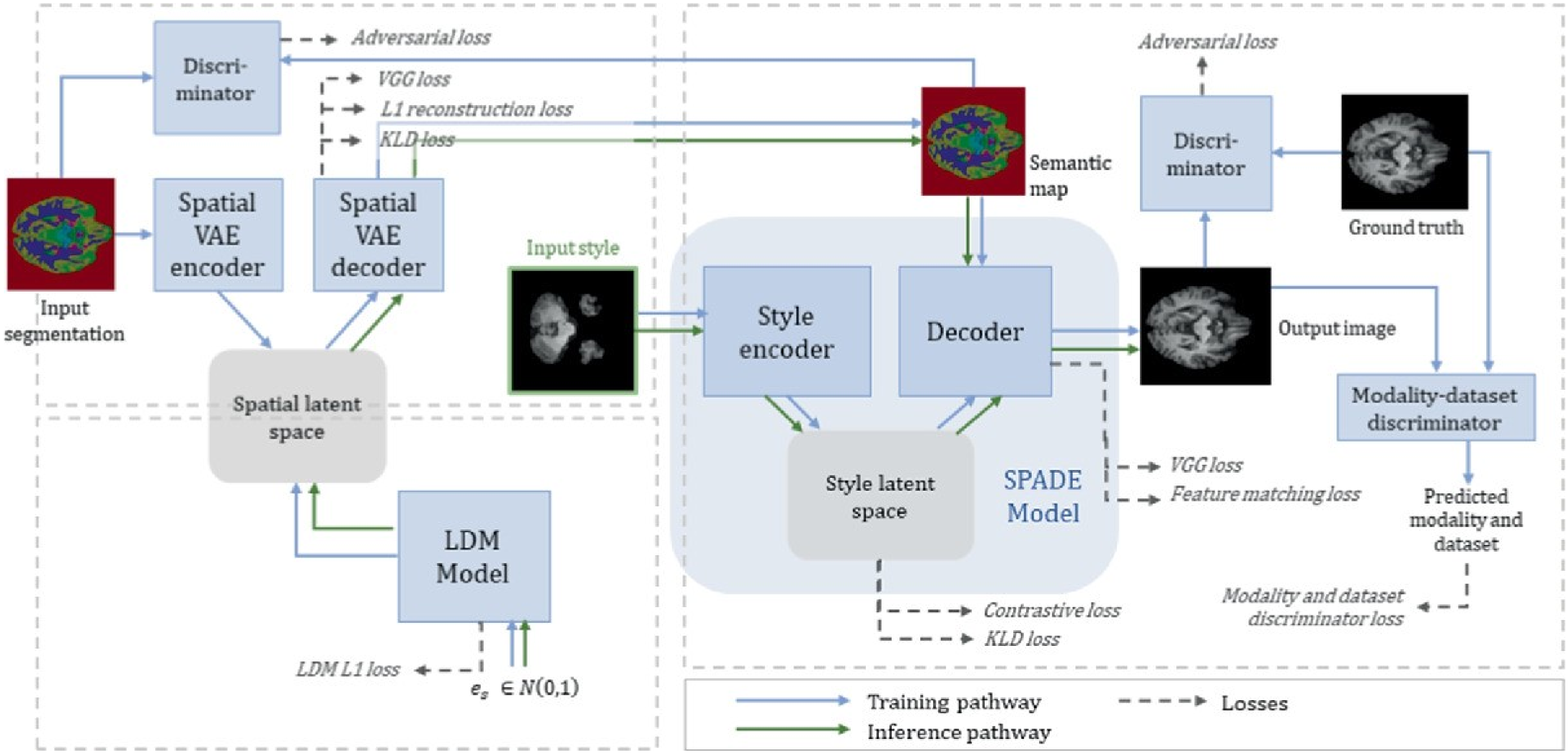
Chambon, Pierre, et al. (2022) RoentGen: vision-language foundation model for chest x-ray generation. *arXiv:2211.12737*

Unlabelled Pre-training



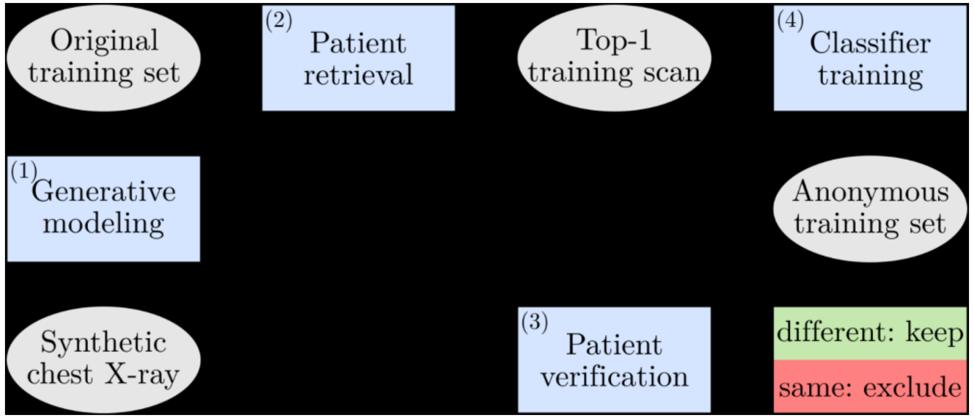
Ye, Jiarong, et al. (2023) Synthetic Augmentation with Large-scale Unconditional Pre-training. MICCAI

Generating Segmentation Masks



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

Generation of Anonymous Chest Radiographs



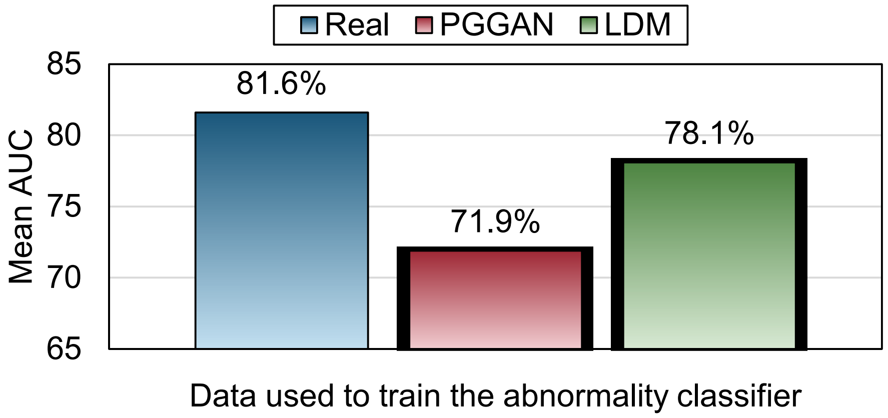
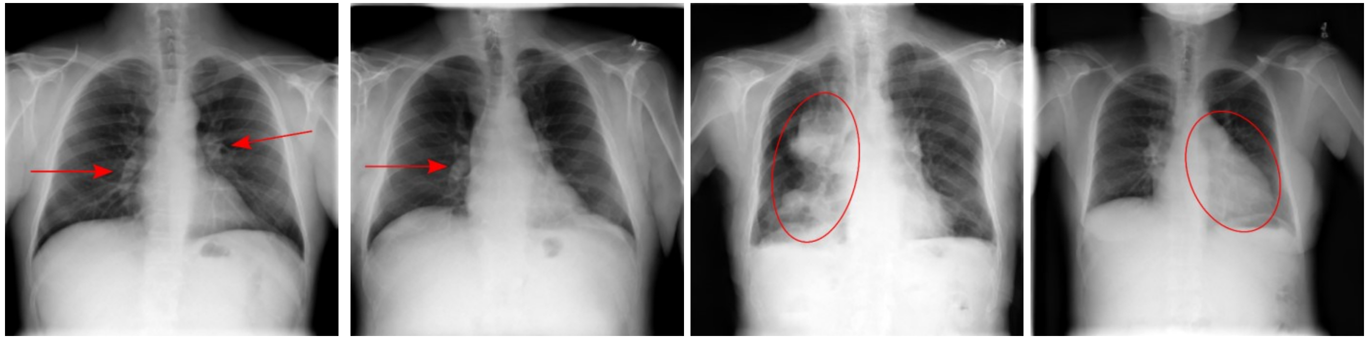


Fig. 1: Proposed privacy-enhancing image sampling strategy. Image taken from [1].

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



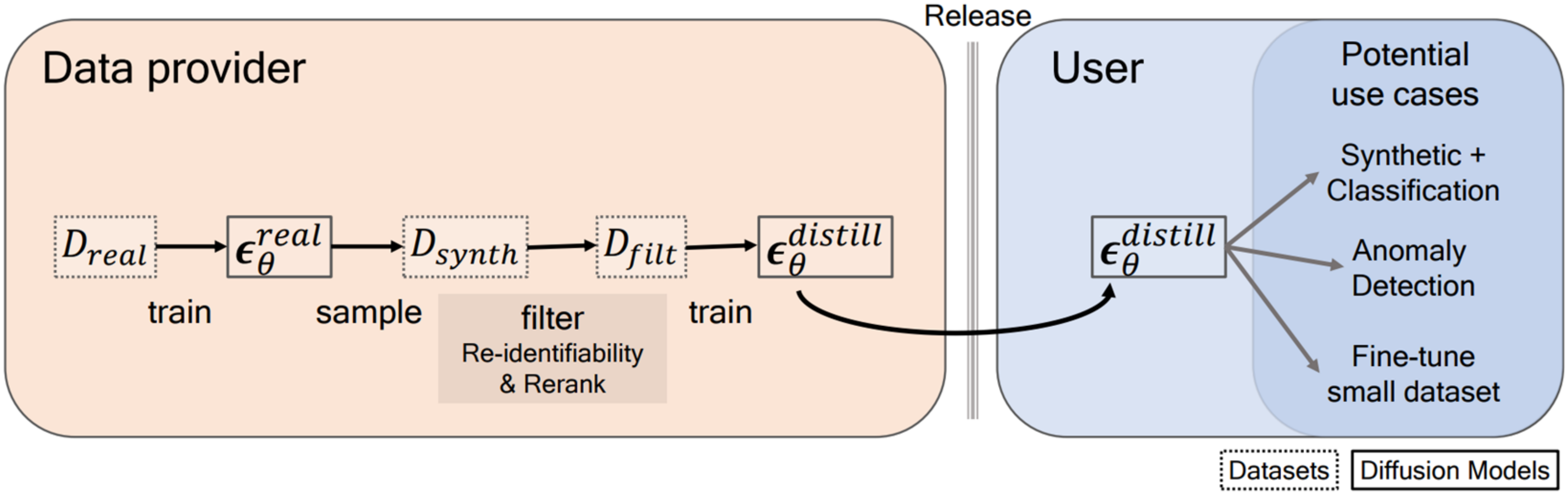
Infiltration Nodule Mass Cardiomegaly

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

Slides courtesy of Kai Packhäuser

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

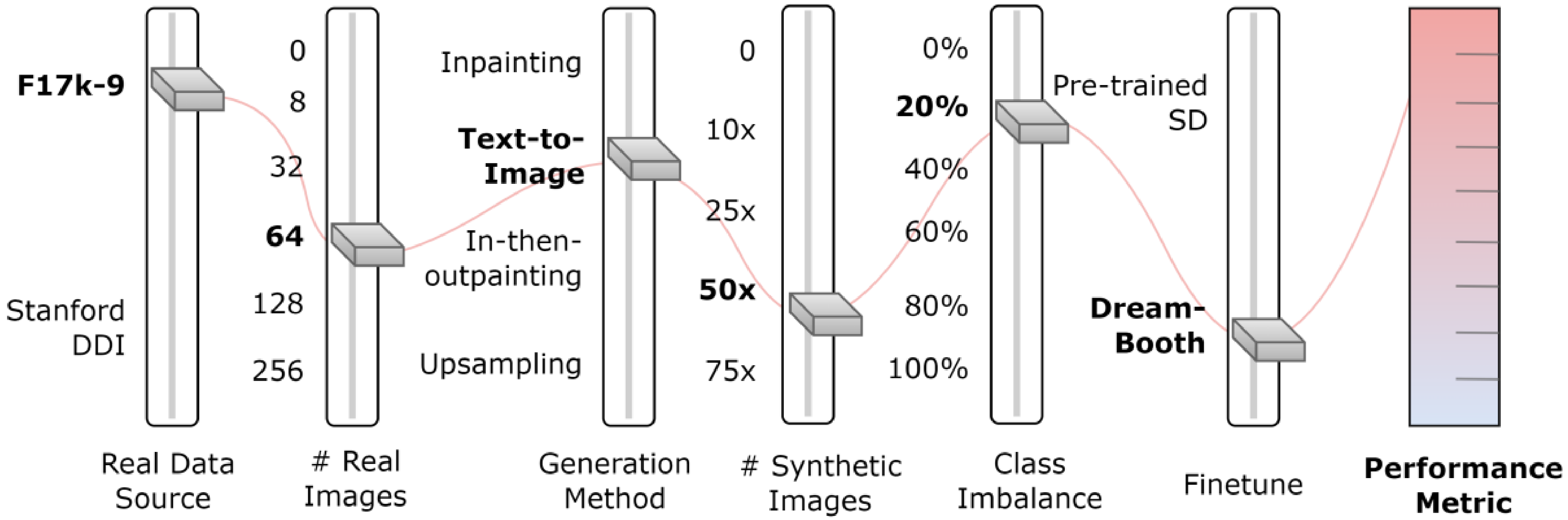
Privacy Distillation



Fernandez, V. et al (2023). Privacy Distillation: Reducing Re-identification Risk of Multimodal Diffusion Models. *MICCAI Workshop DGM4MICCAI*

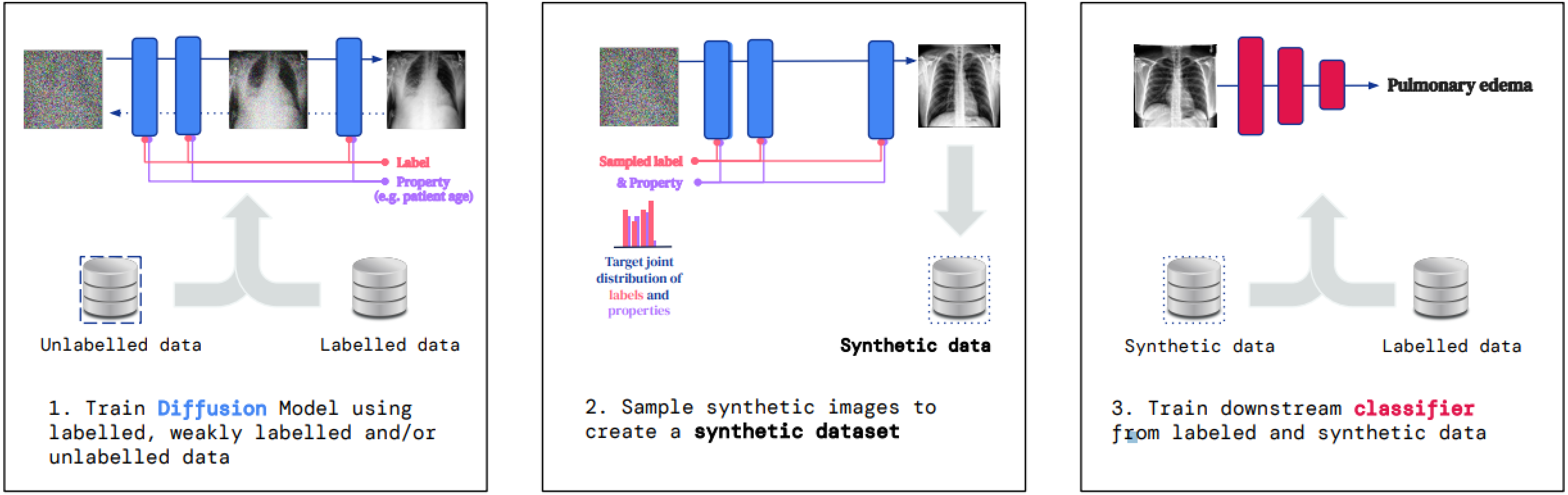
Synthetic Image Augmentation

Synthetic-to-real ratio of 10:1



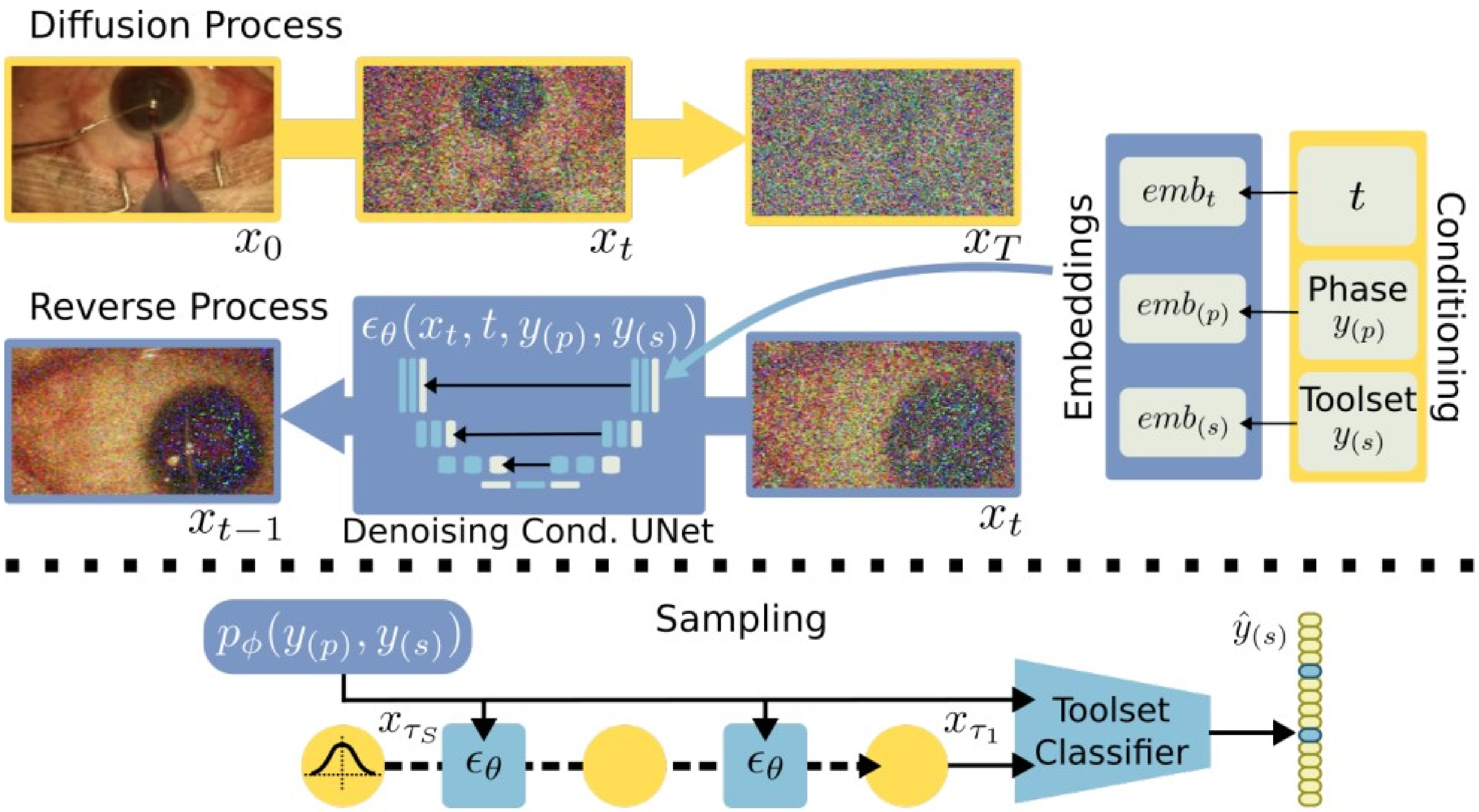
Sagers, Luke W., et al. (2023) Augmenting medical image classifiers with synthetic data from latent diffusion models. arXiv:2308.12453

Synthetic Data for Distribution Shifts



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

Synthesising Rare Samples

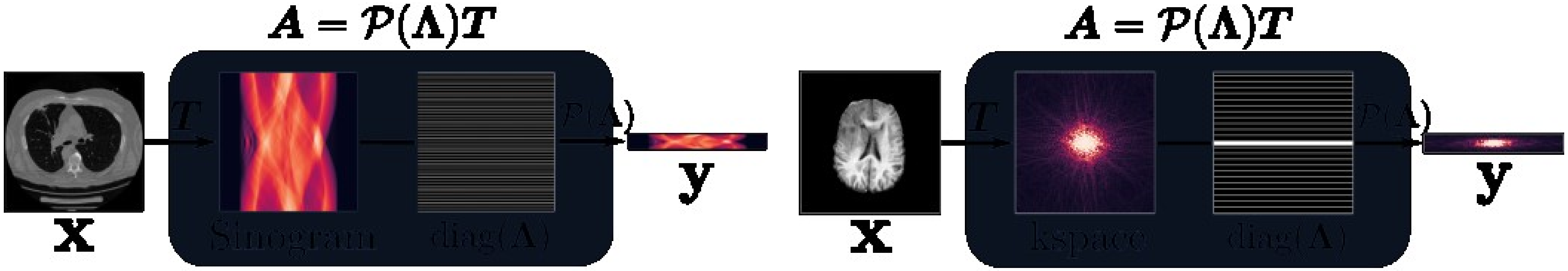


Frisch, Yannik, et al. (2023) Synthesising Rare Cataract Surgery Samples with Guided Diffusion Models. Miccai 2023.

Image reconstruction

Examples from the community

Setup



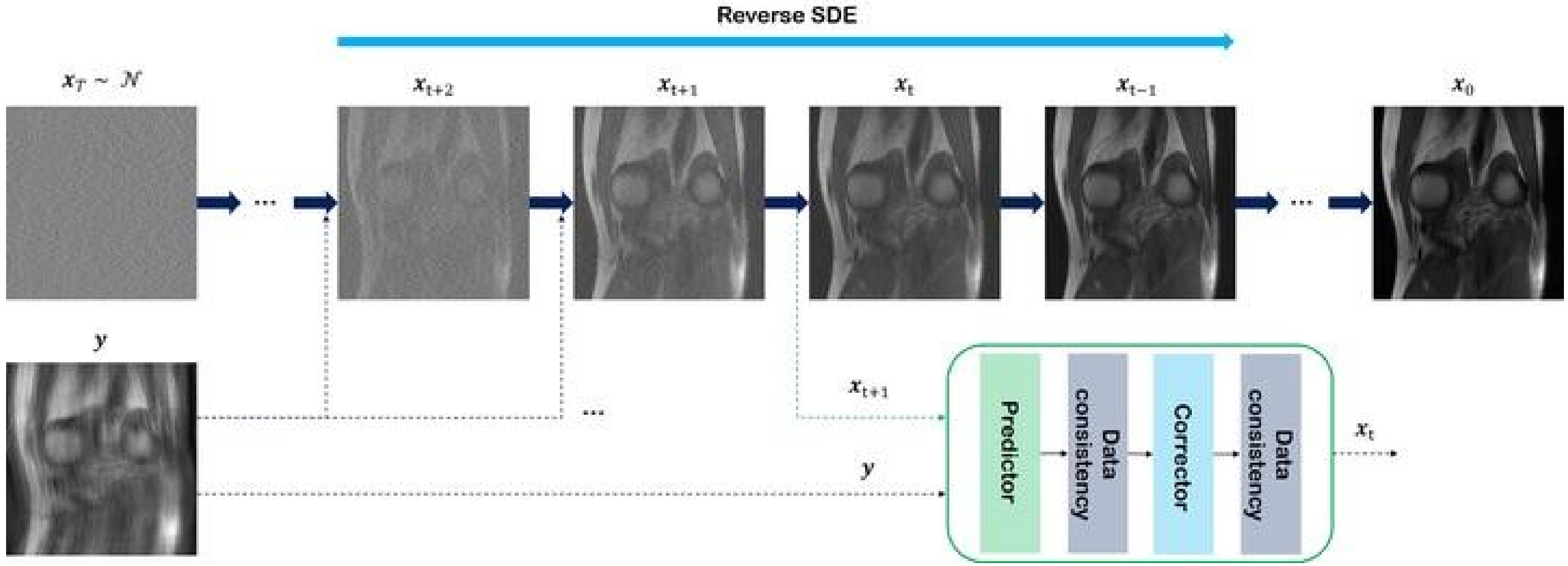
PAPERS

Song et al (2022) Solving Inverse Problems in Medical Imaging with Score-Based Generative Models. ICLR
Chung et al. (2022) Come-Closer-Diffuse-Faster: Accelerating Conditional Diffusion Models for Inverse Problems through Stochastic Contraction. CVPR
Luo et al. (2022) MRI Reconstruction via Data-Driven Markov Chains with Joint Uncertainty Estimation arxiv:2202.01479
Xie et al. (2022) Measurement-Conditioned Denoising Diffusion Probabilistic Model for Under-Sampled Medical Image Reconstruction. MICCAI
Peng et al. (2022) Towards Performant and Reliable Undersampled MR Reconstruction via Diffusion Model Sampling. MICCAI
Gungor et al. (2022) Adaptive Diffusion Priors for Accelerated MRI Reconstruction. arxiv:2207.05876
Cui et al. (2022) Self-Score: Self-Supervised Learning on Score-Based Models for MRI Reconstruction. Arxiv:2209.00835
Cao et al. (2022) High-Frequency Space Diffusion Models for Accelerated MRI. arxiv:2208.05481
Chung et al.(2022) Improving Diffusion Models for Inverse Problems using Manifold Constraints. arxiv:2206.00941
Chung et al. (2022) MR Image Denoising and Super-Resolution Using Regularized Reverse Diffusion. arxiv:2203.12621
Chung et al. (2021) Score-based diffusion models for accelerated MRI. MIA 2021
Hu et al. (2022) Unsupervised Denoising of Retinal OCT with Diffusion Probabilistic Model. arxiv:2201.11760
Gong et al (2022) PET image denoising based on denoising diffusion probabilistic models. arxiv:2209.06167

Figure by Song et al ICLR 2022.

Reconstruction with Data Consistency

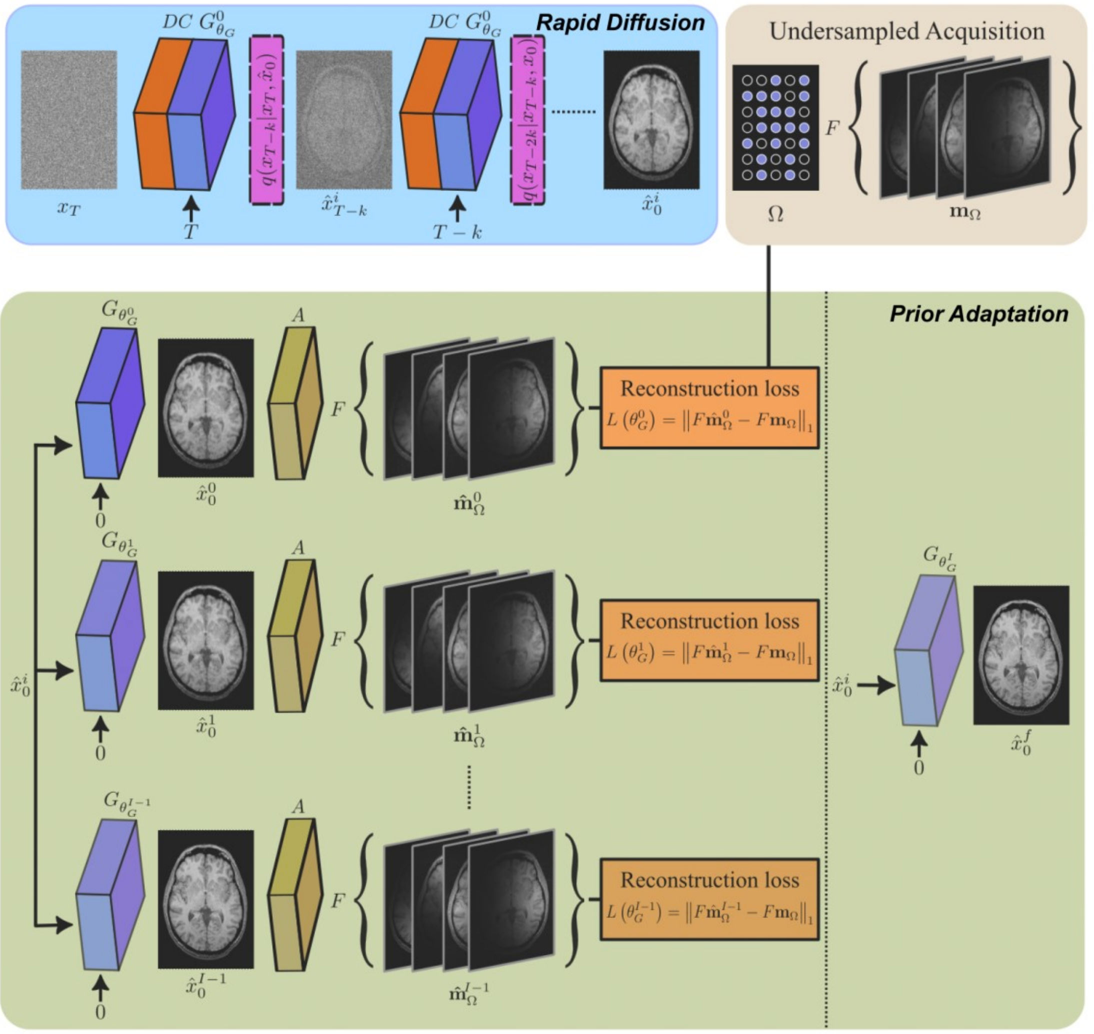
An unconditional diffusion prior is trained on fully-sampled MR acquisitions



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

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

MRI Reconstruction with Adaptive Diffusion Priors



Slides courtesy of Tolga Cukur

1. Gungor et al (2022). Adaptive Diffusion Priors for Accelerated MRI Reconstruction. arXiv:2207.05876. (https://github.com/icon-lab/AdaDiff)

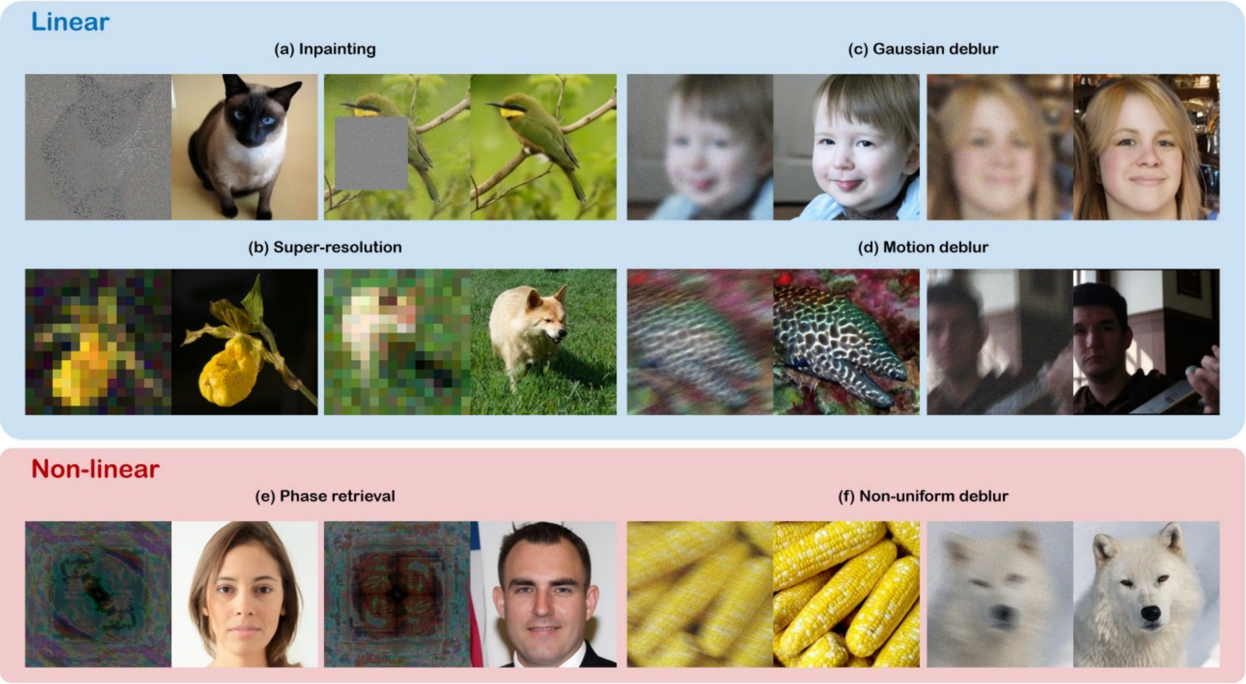
General Inverse Problems

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

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**
  $\hat{\boldsymbol{s}} \leftarrow \boldsymbol{s}_\theta(\boldsymbol{x}_i, i)$
  $\hat{\boldsymbol{x}}_0 \leftarrow \frac{1}{\sqrt{\bar{\alpha}_i}}(\boldsymbol{x}_i + (1 - \bar{\alpha}_i)\hat{\boldsymbol{s}})$
  $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$
  $\boldsymbol{x}'_{i-1} \leftarrow \frac{\sqrt{\alpha_i}(1-\bar{\alpha}_{i-1})}{1-\bar{\alpha}_i}\boldsymbol{x}_i + \frac{\sqrt{\bar{\alpha}_{i-1}}\beta_i}{1-\bar{\alpha}_i}\hat{\boldsymbol{x}}_0 + \tilde{\sigma}_i \boldsymbol{z}$
  $\boldsymbol{x}_{i-1} \leftarrow \boldsymbol{x}'_{i-1} - \zeta_i \nabla_{\boldsymbol{x}_i} \|\boldsymbol{y} - \mathcal{A}(\hat{\boldsymbol{x}}_0)\|_2^2$
end for
return $\hat{\mathbf{x}}_0$



Chung, et al. (2022). Diffusion posterior sampling for general noisy inverse problems. *ICLR*

Image registration

Examples from the community

DiffuseMorph

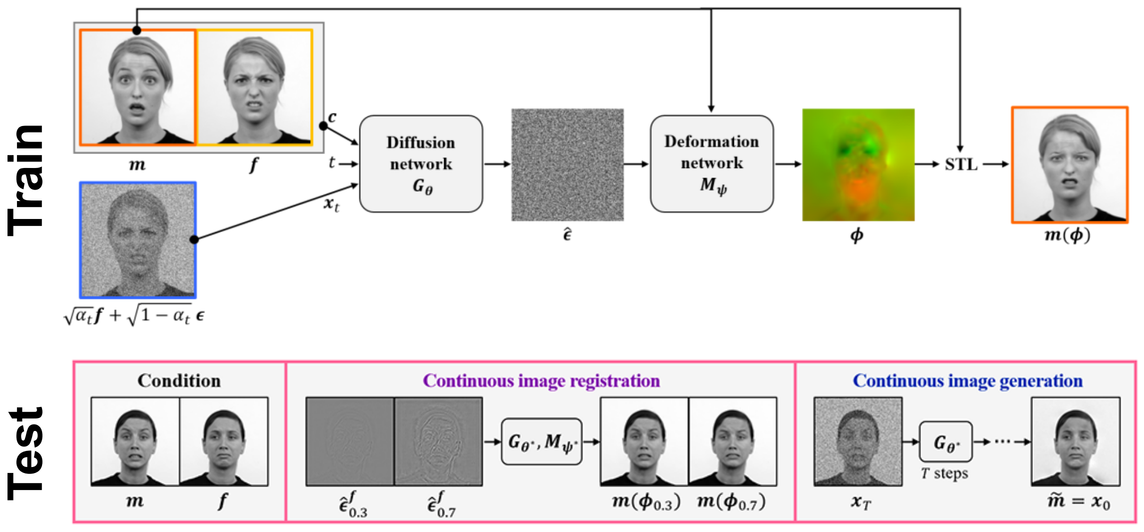
- To perform <u>image registration</u> along the continuous trajectory
- **Diffusion network**: To estimate a conditional score function
- **Deformation network**: To yield the registration fields & provide the deformed image

Loss function

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

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

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



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

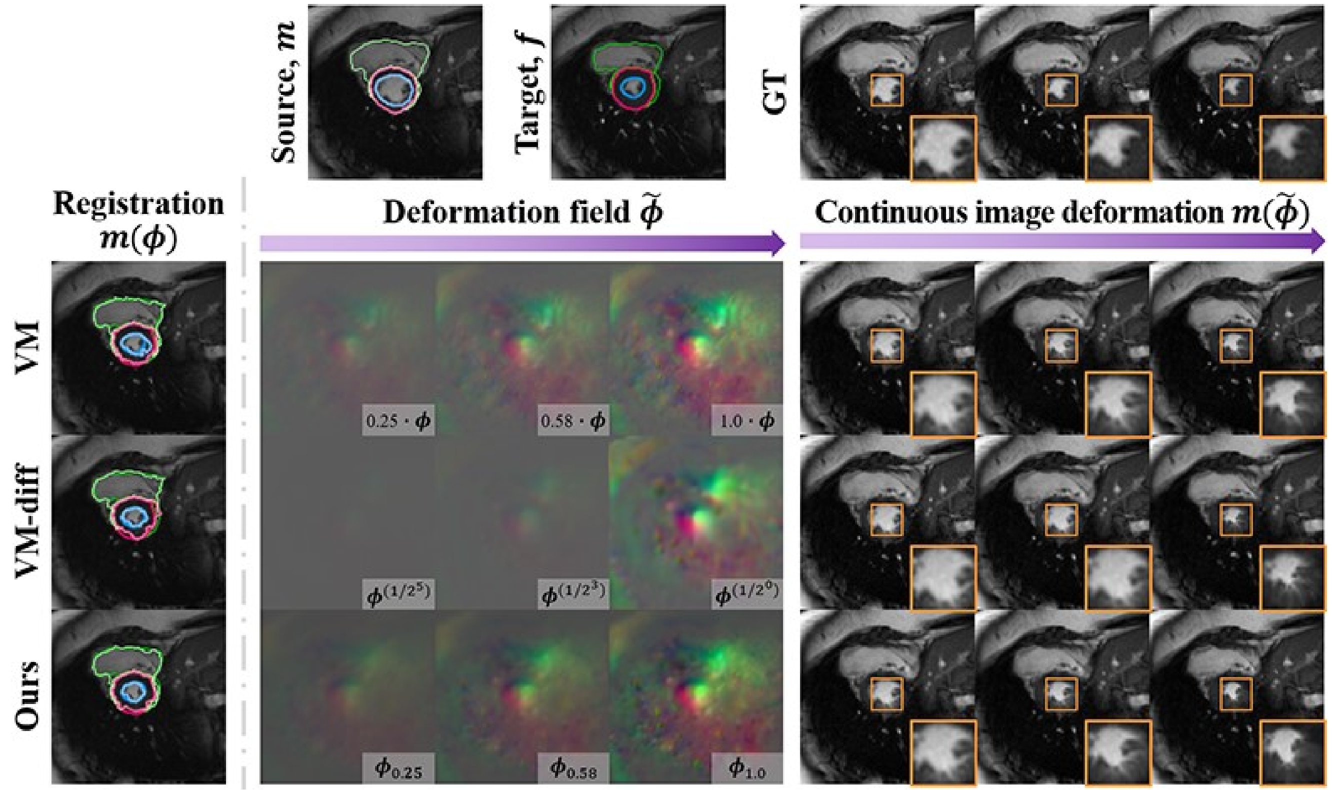
Input: Conditional images, $c = (m, f)$
Output: Synthetic deformed image, x
Set $T \in (0, T_{train})$
Sample $x_T = \sqrt{\alpha_T} m + \sqrt{1-\alpha_T}\epsilon$, where $\epsilon \sim \mathcal{N}(0, I)$
for $t = T, T-1, ..., 1$ **do**
  $z \sim \mathcal{N}(0, I)$
  $x_{t-1} \leftarrow \frac{1}{\sqrt{1-\beta_t}}(x_t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

DiffuseMorph

- Intra-subject 3D cardiac MR image registration

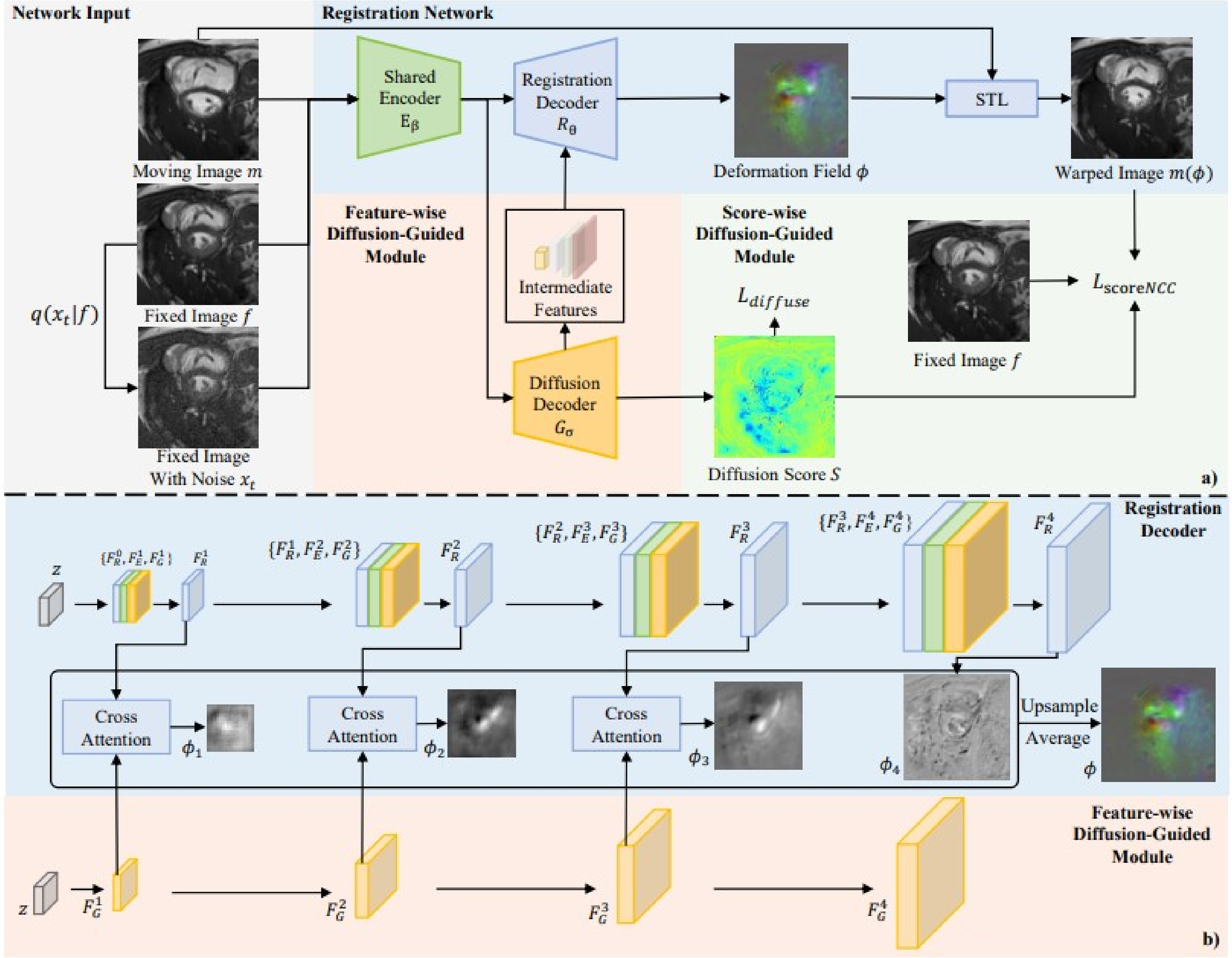


Methods	Dice	$\lvert J_\phi \rvert \leq 0$ (%)
Initial	0.642 (0.188)	-
VM [1]	0.787 (0.113)	0.169 (0.109)
VM-diff [2]	0.794 (0.104)	0.291 (0.188)
Ours	**0.802 (0.109)**	**0.161 (0.082)**

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

Feature-wise Diffusion-Guided

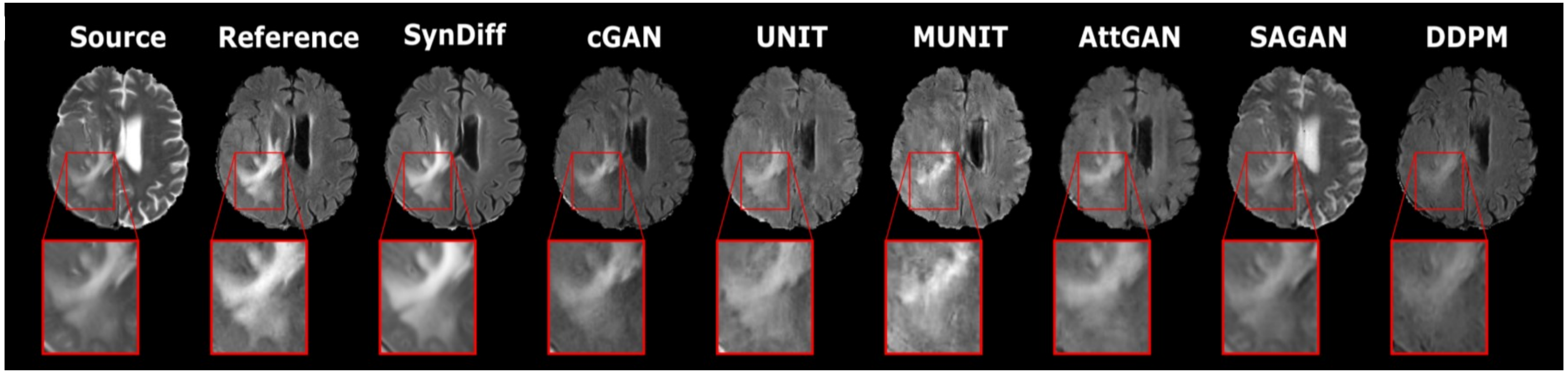


Qin et al. (2023) FSDiffReg: Feature-wise and Score-wise Diffusion-guided Unsupervised Deformable Image Registration for Cardiac Images. Miccai 2023

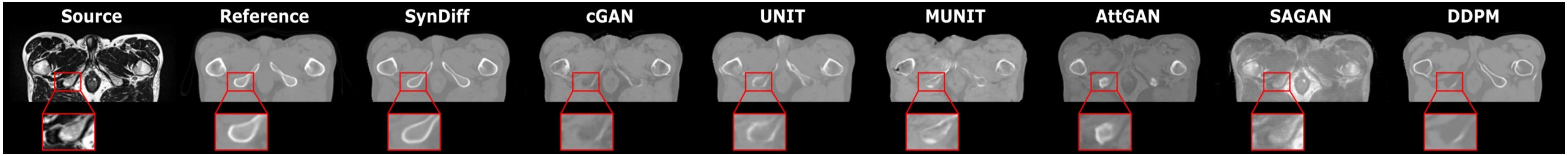
Image-to-Image translation

Setup

MRI Contrast Translation



MRI to CT Translation



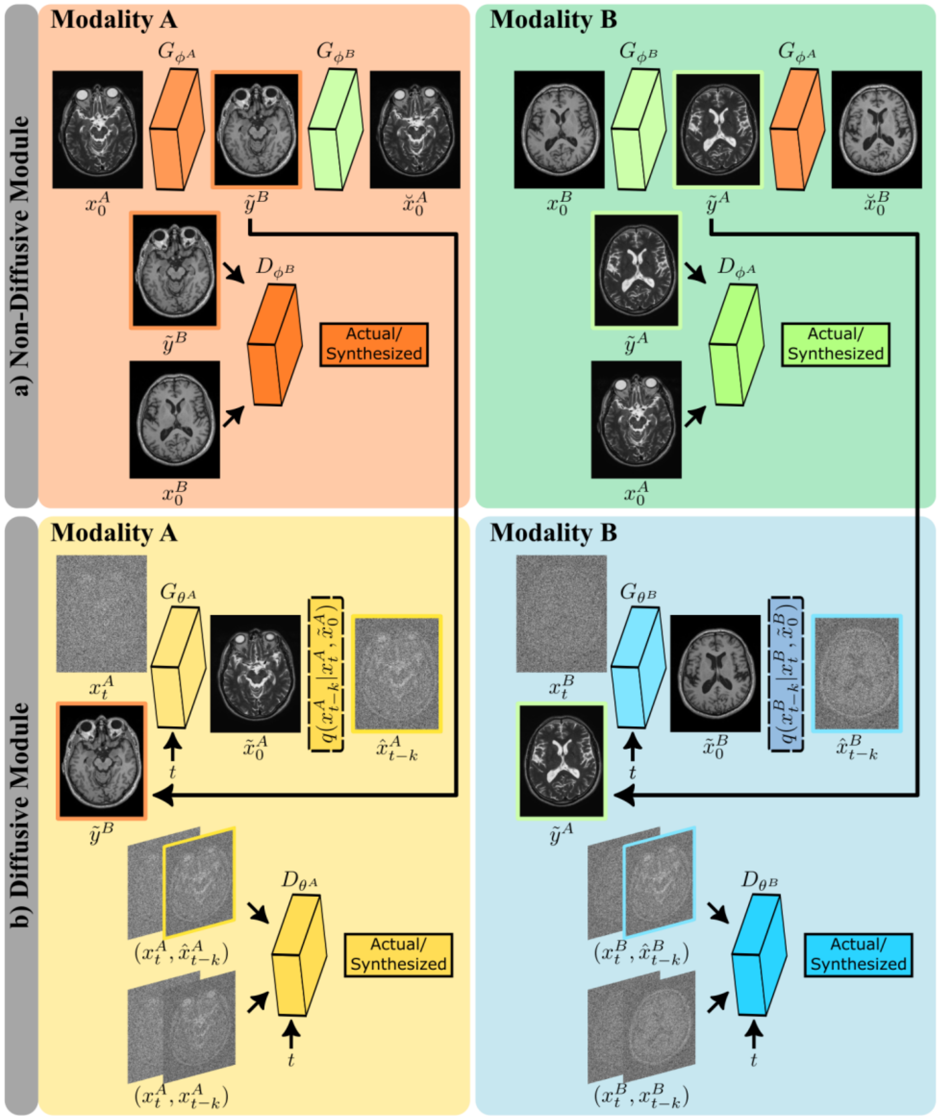
Slides courtesy of Tolga Cukur

Ozbey et al (2022). Unsupervised Medical Image Translation with Adversarial Diffusion Models. arXiv:2207.08208. (https://github.com/icon-lab/SynDiff)

Medical Image Translation with Adversarial Diffusion

<u>*SynDiff*</u>: an unsupervised diffusion model for medical image translation

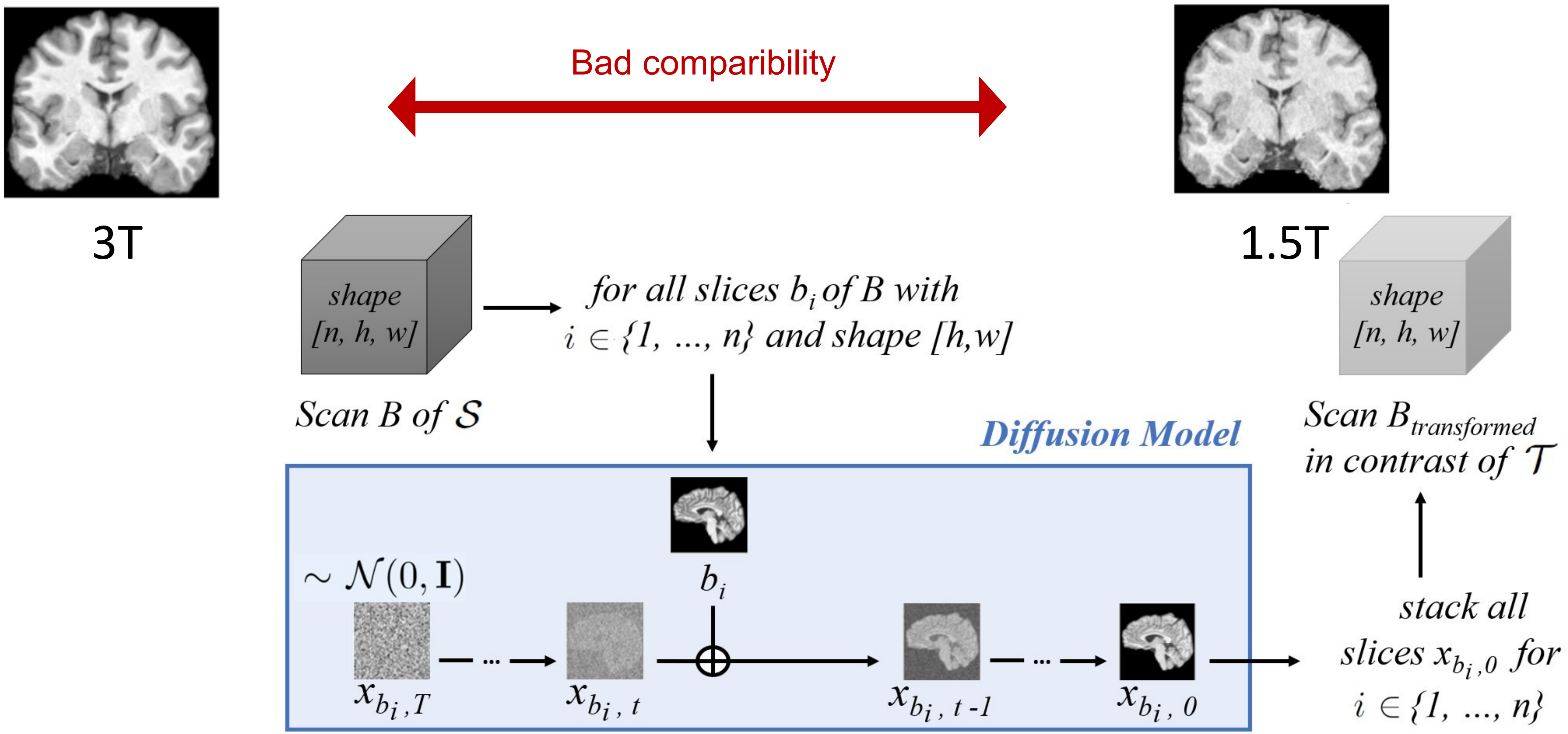
- An adversarial diffusive module maps fast source ➔ target
- A non-diffusive module with cycle-consistency loss enables training on unpaired datasets



Slides courtesy of Tolga Cukur

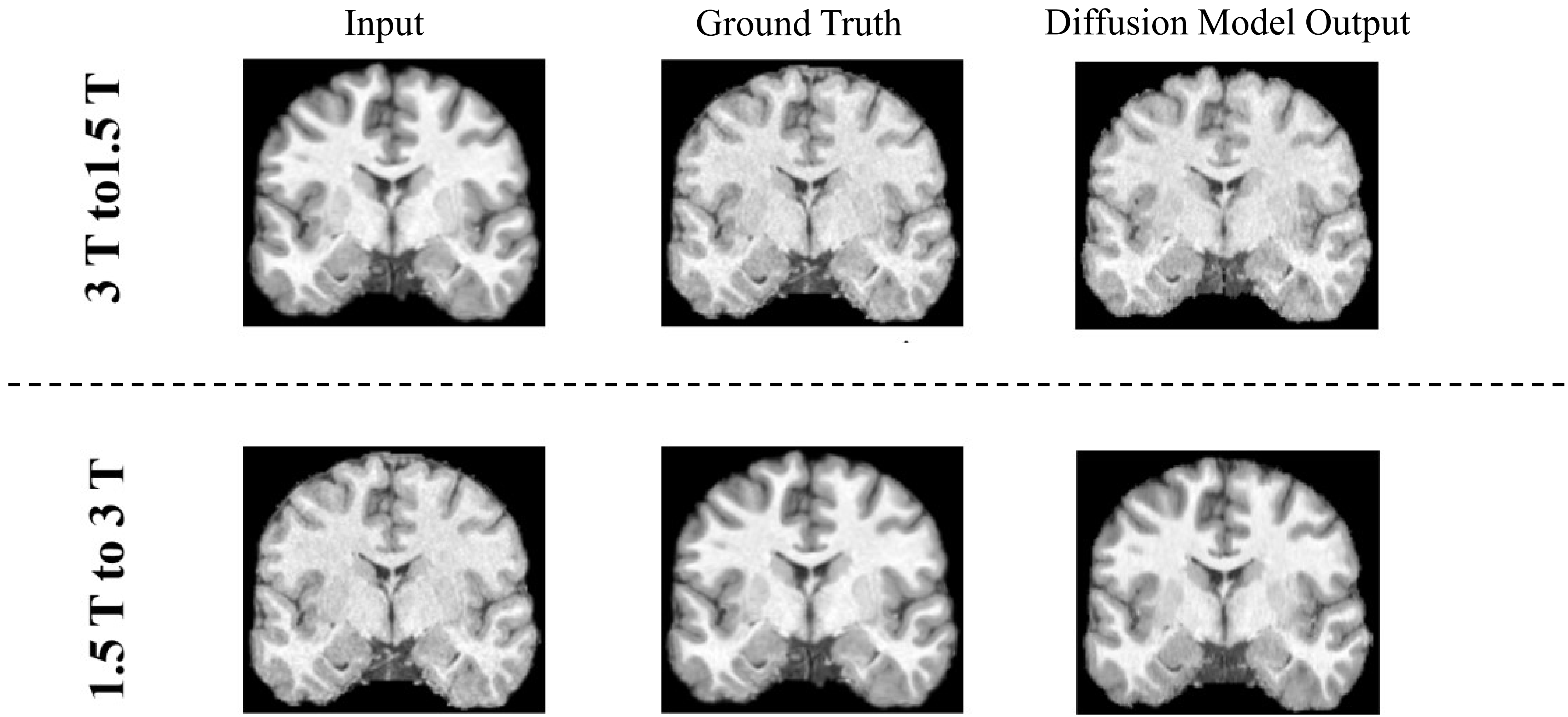
Ozbey et al (2022). Unsupervised Medical Image Translation with Adversarial Diffusion Models. arXiv:2207.08208. (https://github.com/icon-lab/SynDiff)

Diffusion Models for Contrast Harmonization



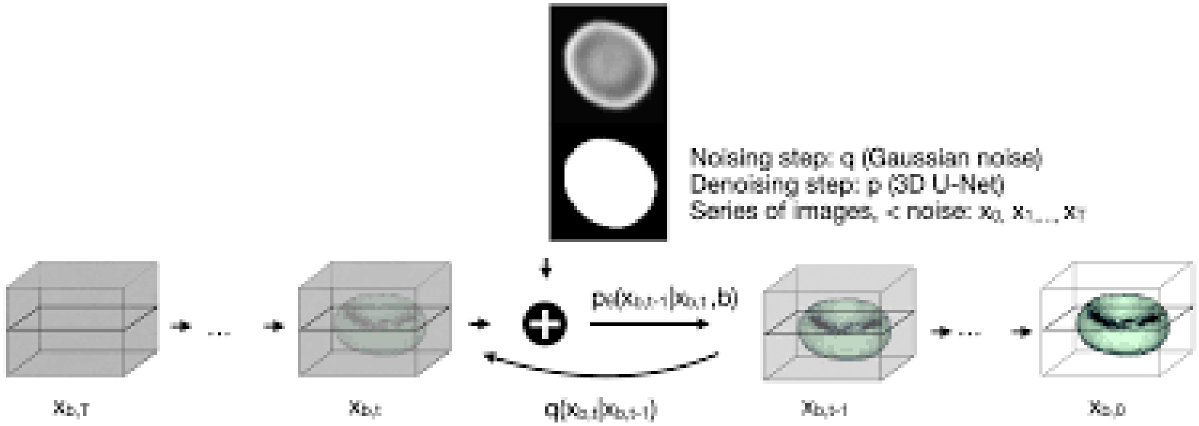
Durrer, Alicia, et al. "Diffusion Models for Contrast Harmonization of Magnetic Resonance Images." *arXiv preprint arXiv:2303.08189* (2023).

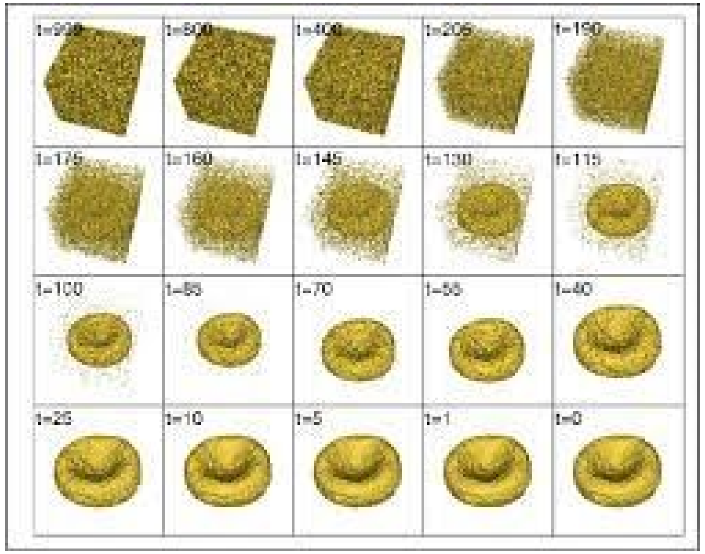
Contrast Harmonization Results

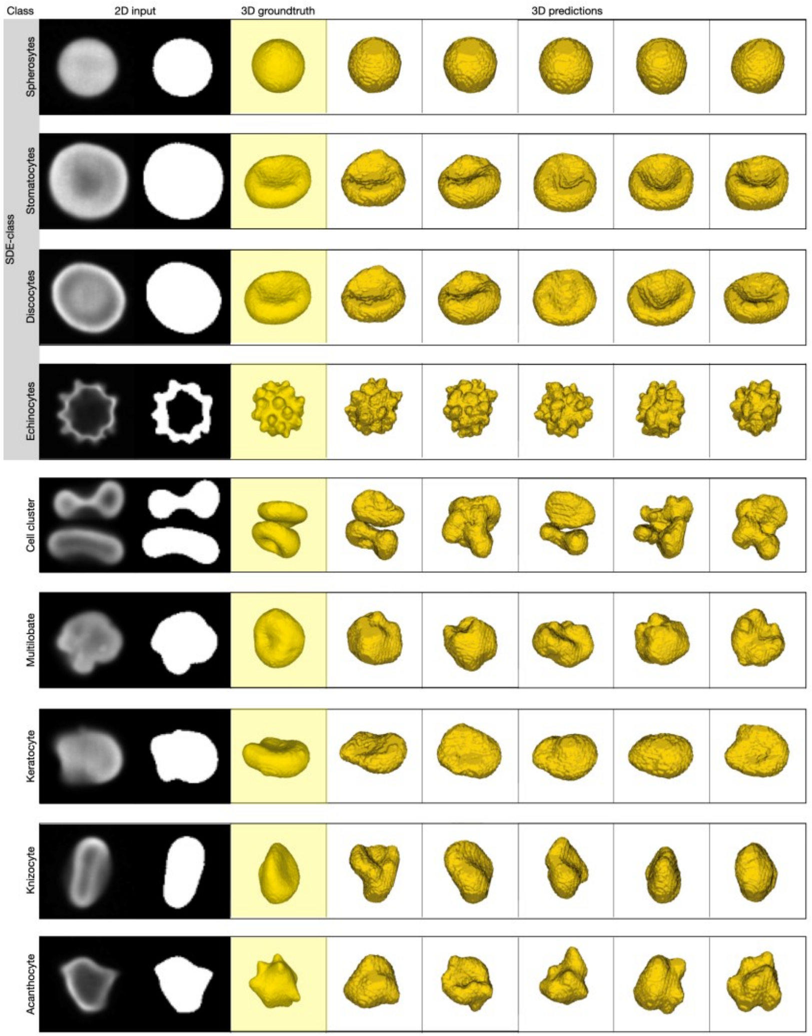


Durrer, Alicia, et al. "Diffusion Models for Contrast Harmonization of Magnetic Resonance Images." *arXiv preprint arXiv:2303.08189* (2023).

3D Shapes from 2D Microscopy Images

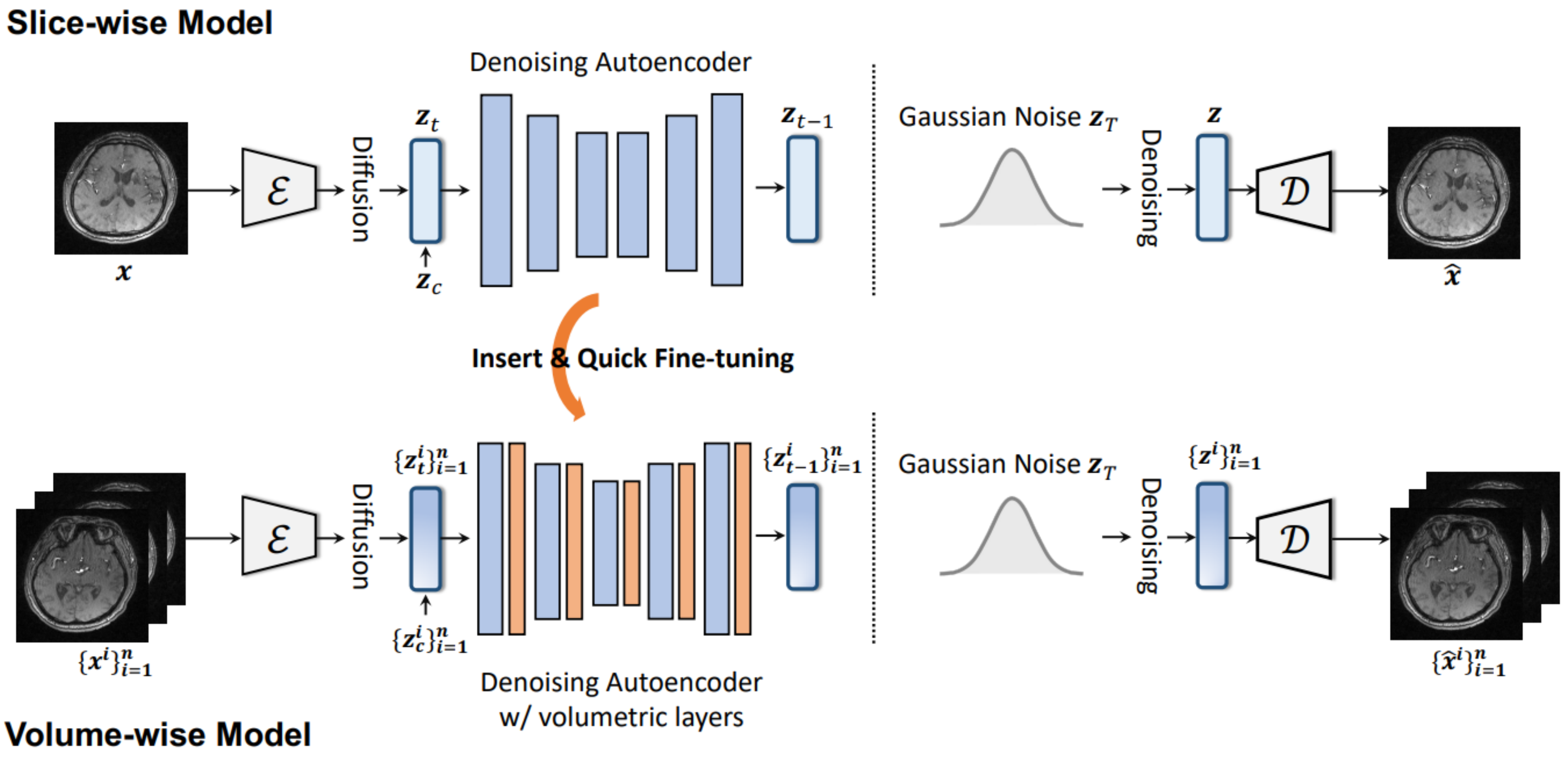






Waibel, D. J., Röell, E., Rieck, B., Giryes, R., & Marr, C. (2023, April). A diffusion model predicts 3d shapes from 2d microscopy images. In *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)* (pp. 1-5). IEEE.

3D with 2D model

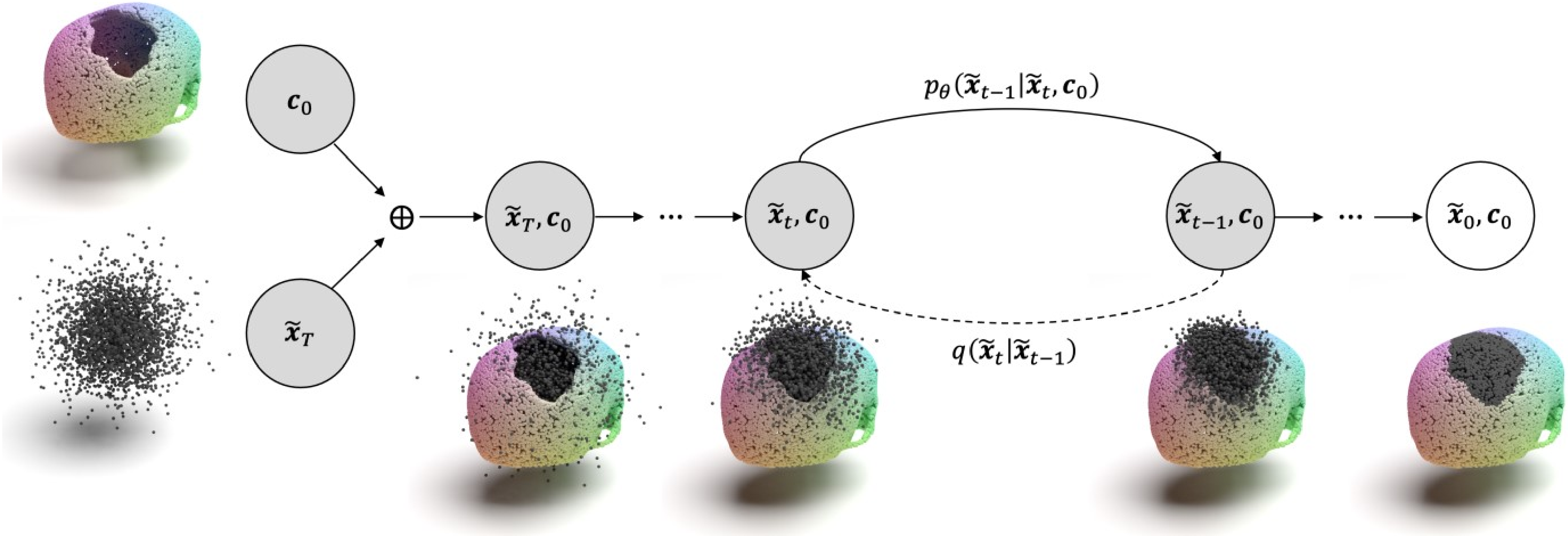


Zhu, Lingting, et al. (2023) Make-A-Volume: Leveraging Latent Diffusion Models for Cross-Modality 3D Brain MRI Synthesis. MICCAI

Inpainting

Examples from the community

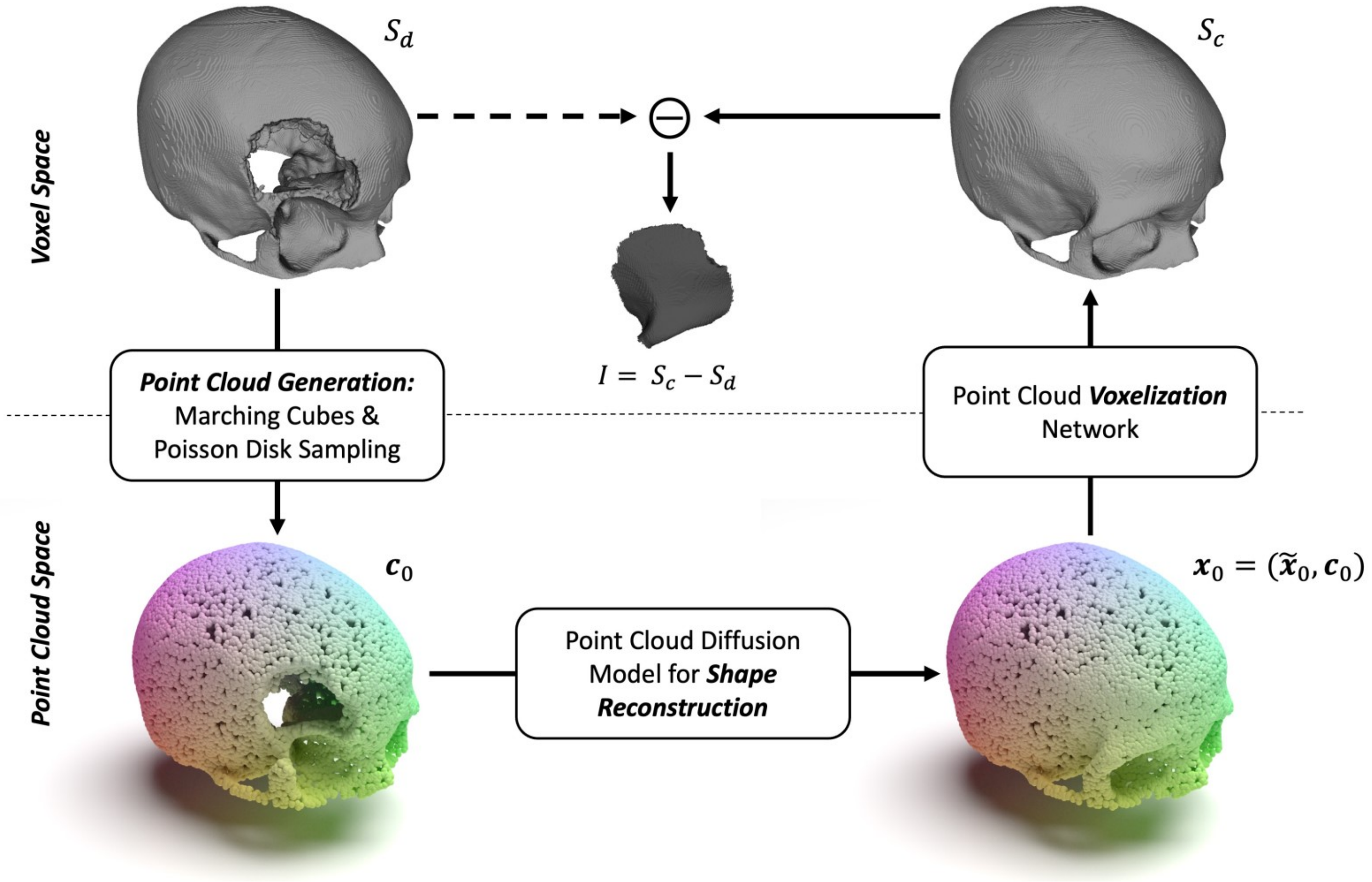
Point Cloud Diffusion Models for Implant Generation

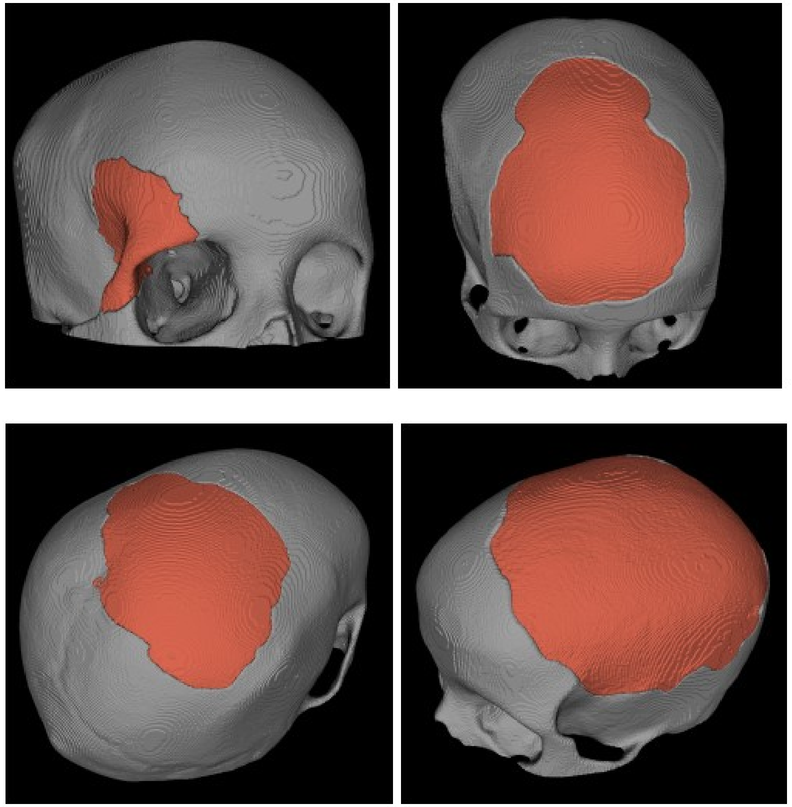


- For automatic implant generation, we aim to complete a defective skull.
- The diffusion process is applied on a **point cloud representation** due to memory and computation time restrictions.
- We condition the generation process on the skull with a defect.

Friedrich, Paul, et al. "Point cloud diffusion models for automatic implant generation." MICCAI 2023.

Point Cloud Completion



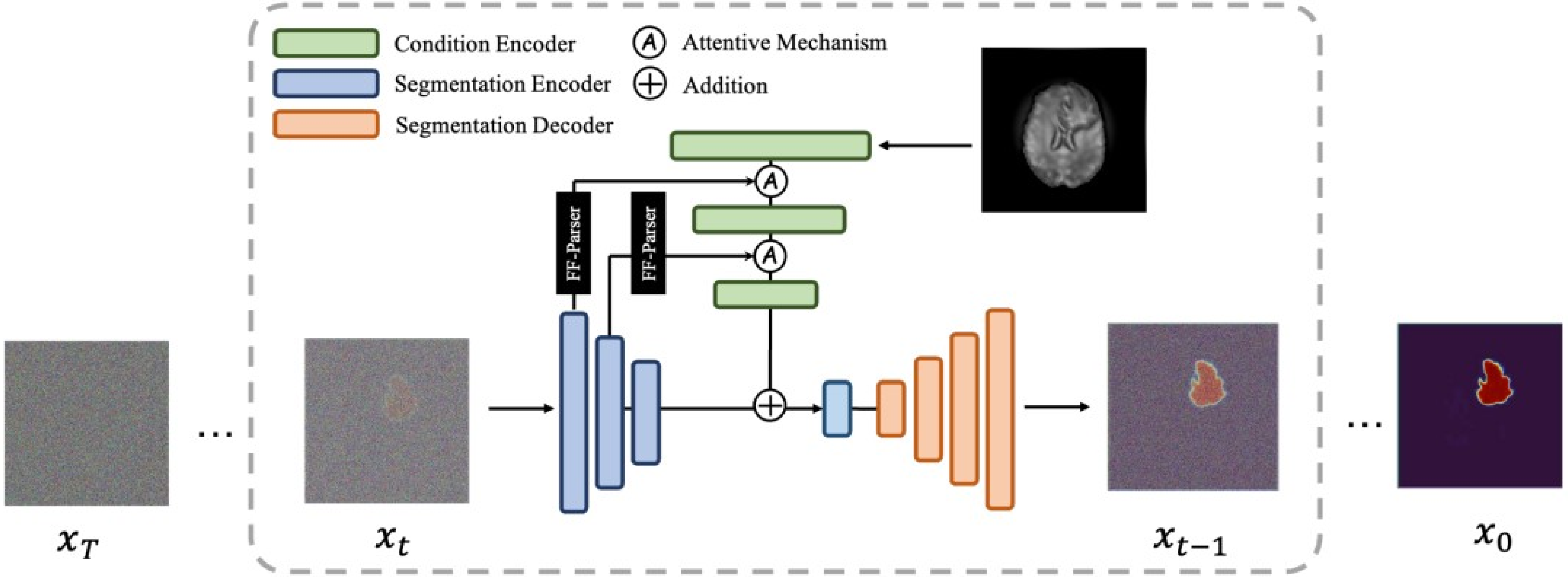


Friedrich, Paul, et al. "Point cloud diffusion models for automatic implant generation." MICCAI 2023.

Image segmentation

Examples from the community

Setup

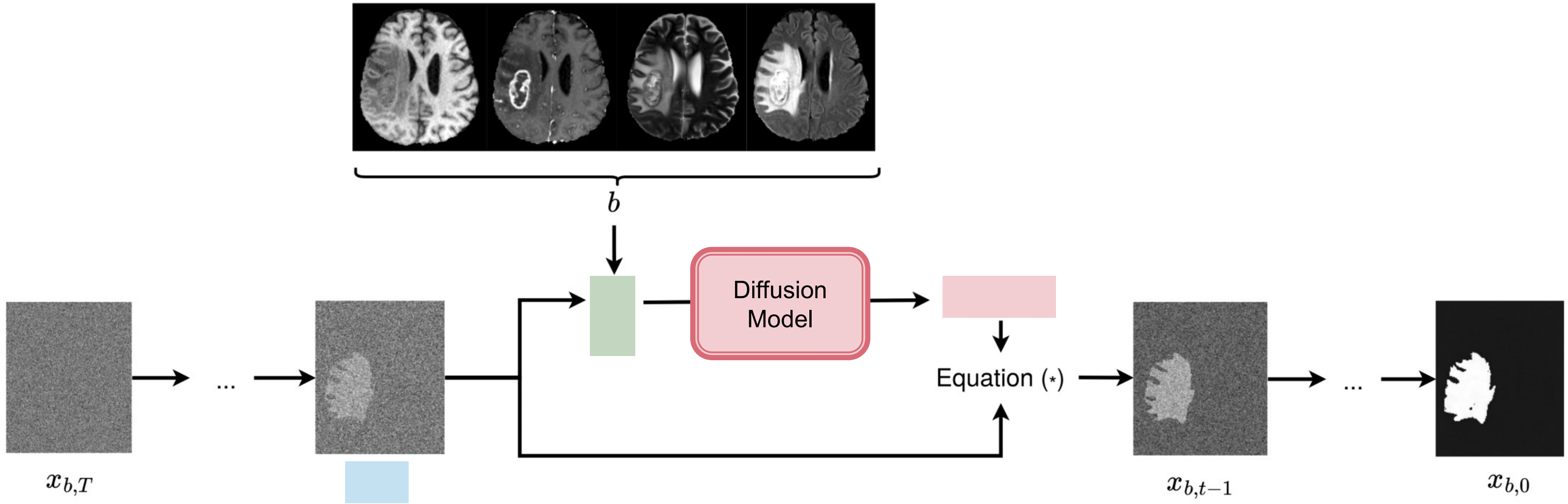


PAPERS

Wolleb et al (2022). Diffusion Models for Implicit Image Segmentation Ensembles, *MIDL 2022.* arXiv:2112.03145
Guo et al (2022) Accelerating Diffusion Models via Pre-segmentation Diffusion Sampling for Medical Image Segmentation. arXiv:2210.17408
La Barbera et al. (2022) Anatomically constrained CT image translation for heterogeneous blood vessel segmentation. arXiv:2210.01713
Kim et al. (2022) Diffusion Adversarial Representation Learning for Self-supervised Vessel Segmentation. arXiv:2209.14566
Wu et al (2022) MedSegDiff: Medical Image Segmentation with Diffusion Probabilistic Model. arXiv:2211.00611
Rahman, Aimon, et al. (2023) Ambiguous medical image segmentation using diffusion models. CVPR
Bieder et al. (2023) Memory-Efficient 3D Denoising Diffusion Models for Medical Image Processing. Medical Imaging with Deep Learning
Rousseau et al. (2023) Pre-Training with Diffusion models for Dental Radiography segmentation. Miccai 2023

Figure by Song et al ICLR 2022.

Diffusion Models for Segmentation Mask Generation

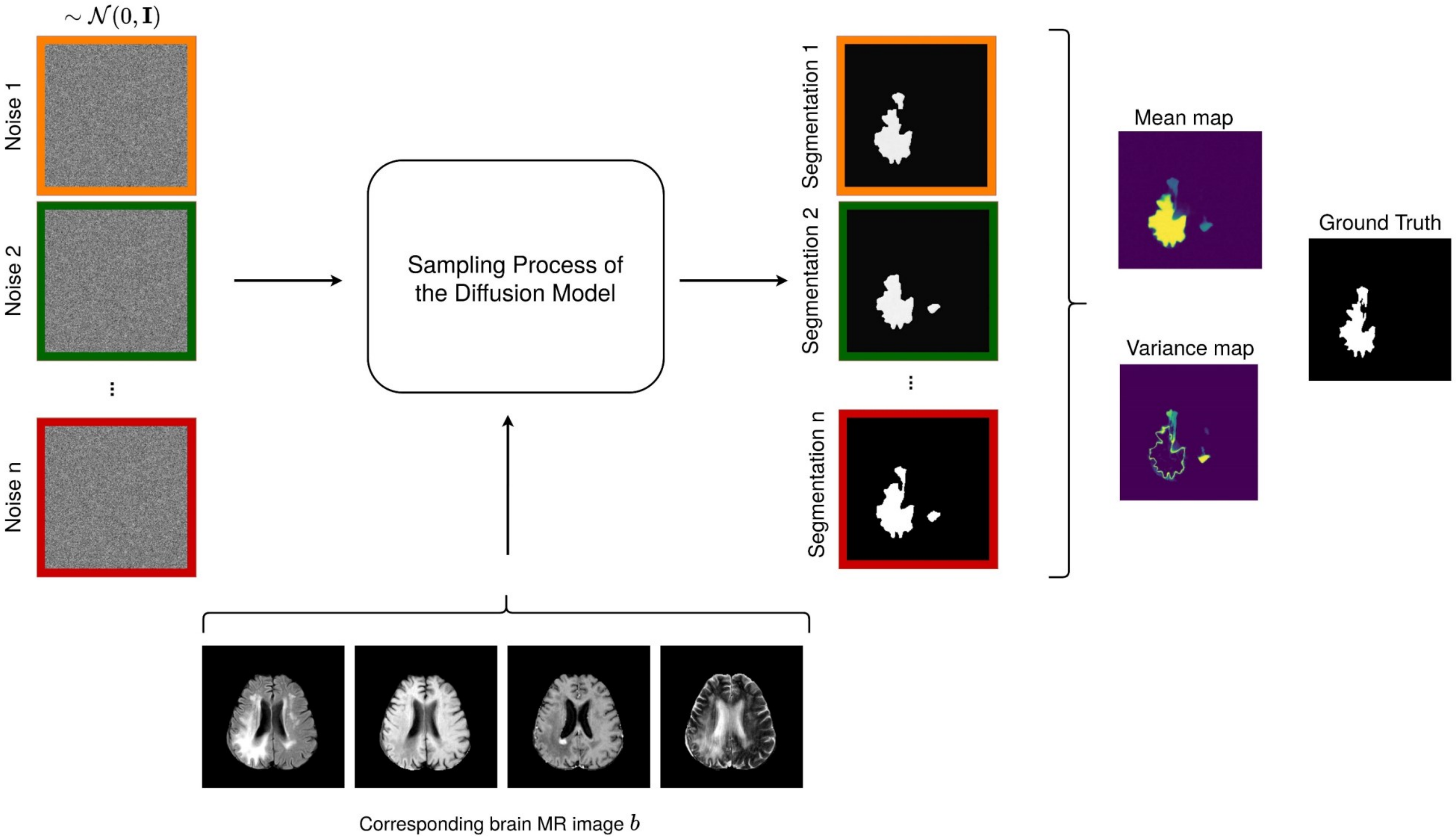


$$(*) \quad x_{b,t-1} = \frac{1}{\sqrt{\alpha_t}}\left(\quad - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}} \quad \right) + \sigma_t \mathbf{z}, \quad \text{with } \mathbf{z} \sim \mathcal{N}(0, \mathbf{I})$$

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

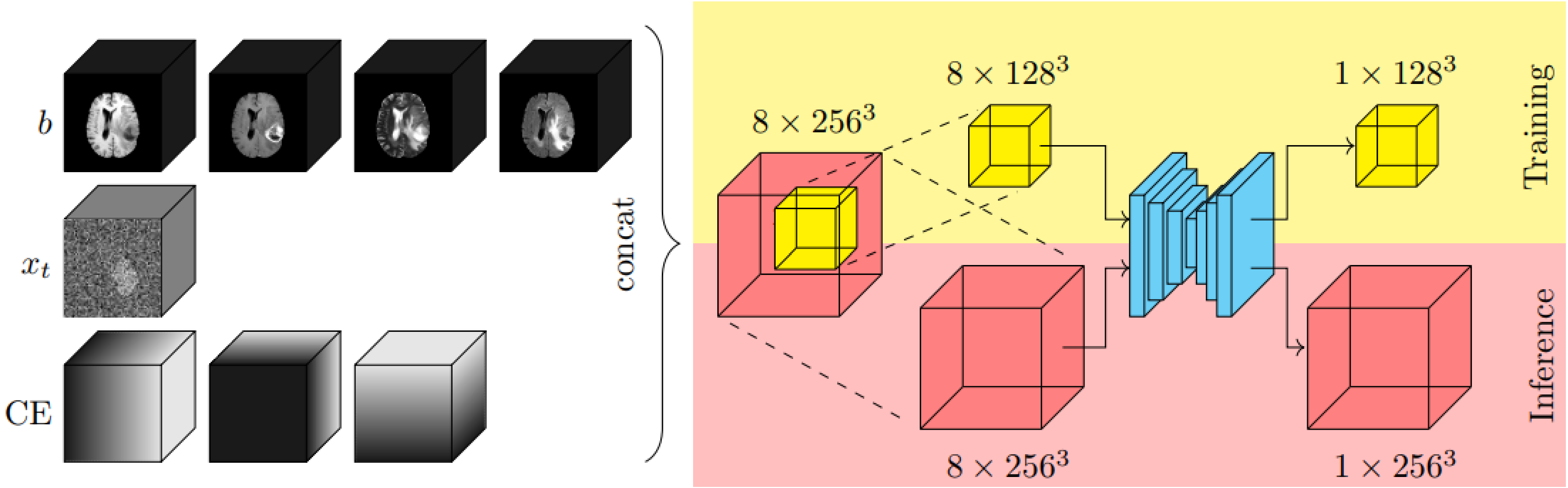
Wolleb et al (2022). Diffusion Models for Implicit Image Segmentation Ensembles, *MIDL 2022.* arXiv:2112.03145

Generation of Segmentation Ensembles



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

3D Segmentation with PatchDDM

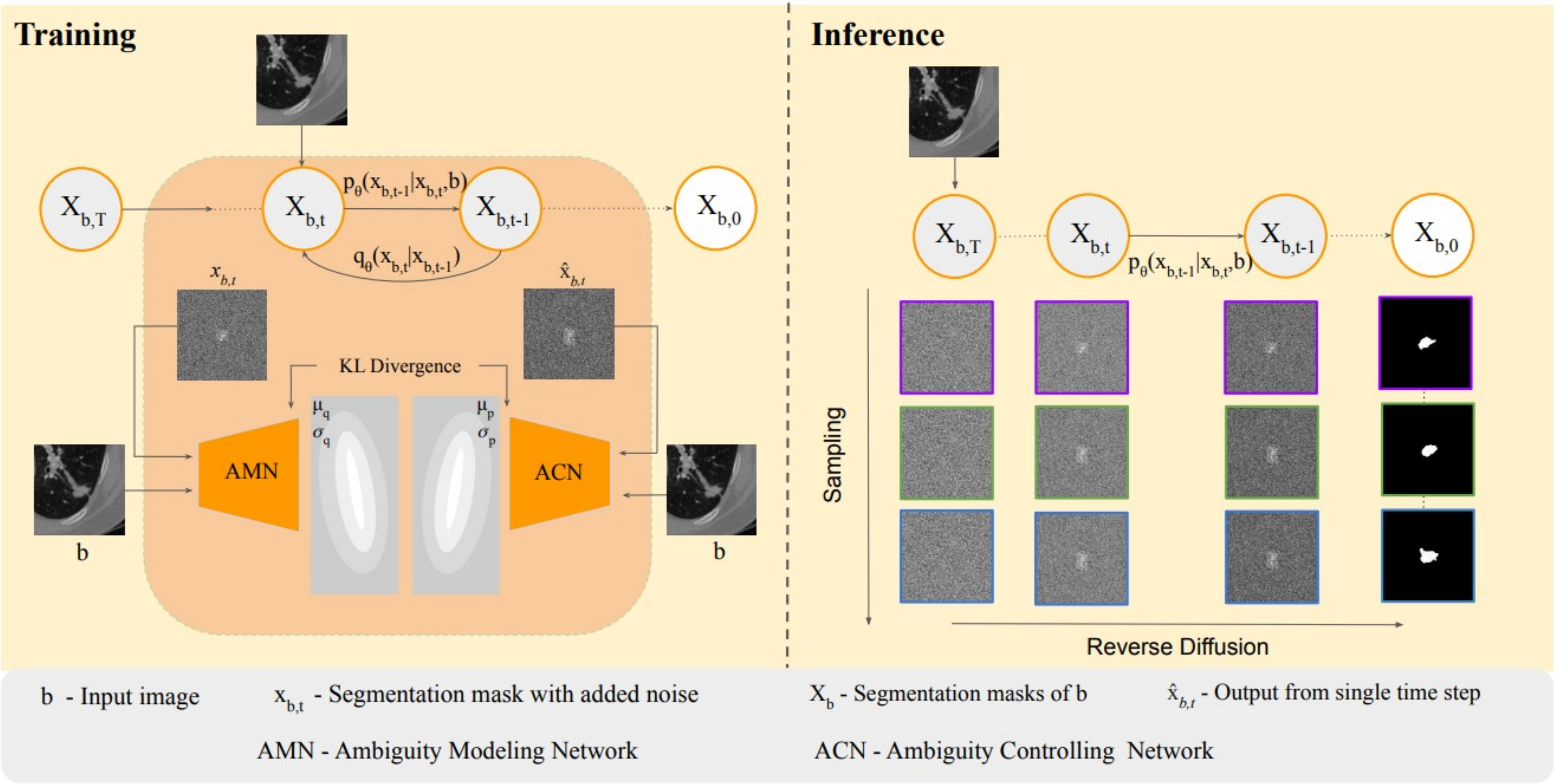


- We add a position encoding in all 3 spatial dimensions.
- Training is on patches only, and saves memory and training time.
- Inference runs over the whole 3D volume.

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

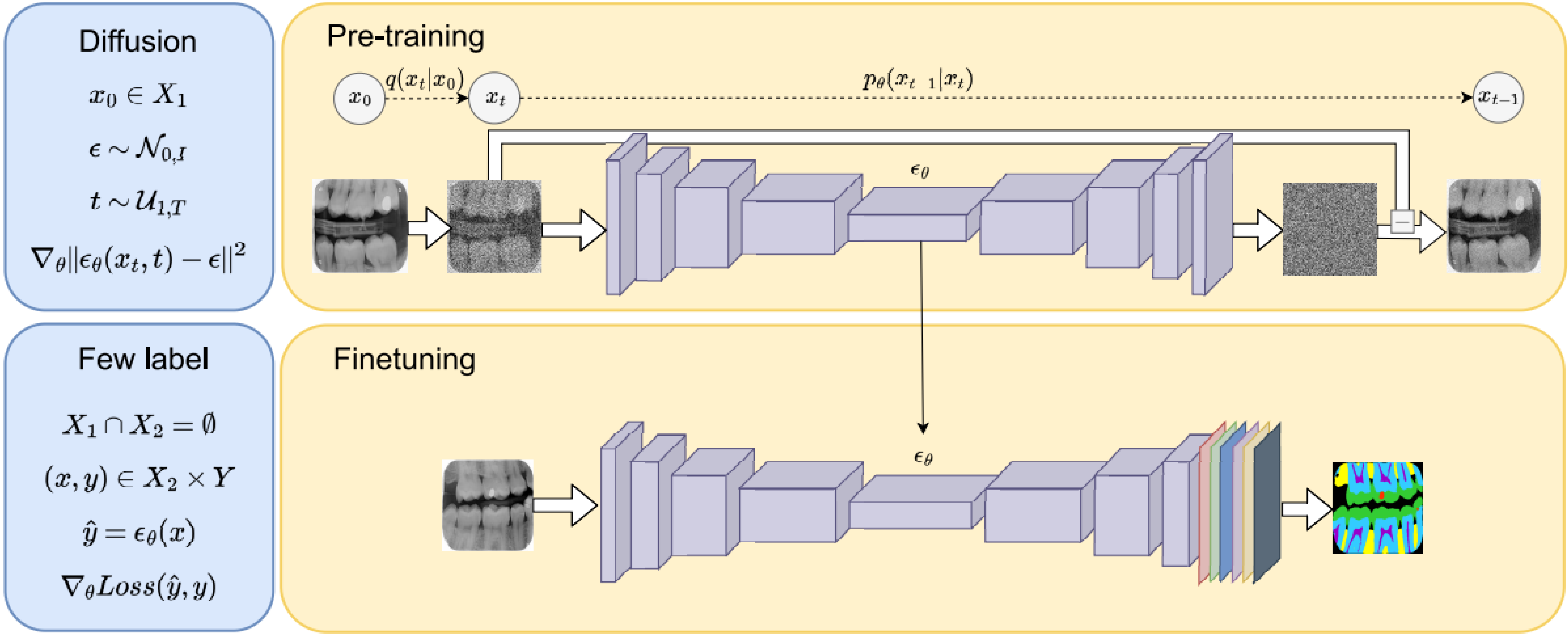
Ambiguous Segmentation

- 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

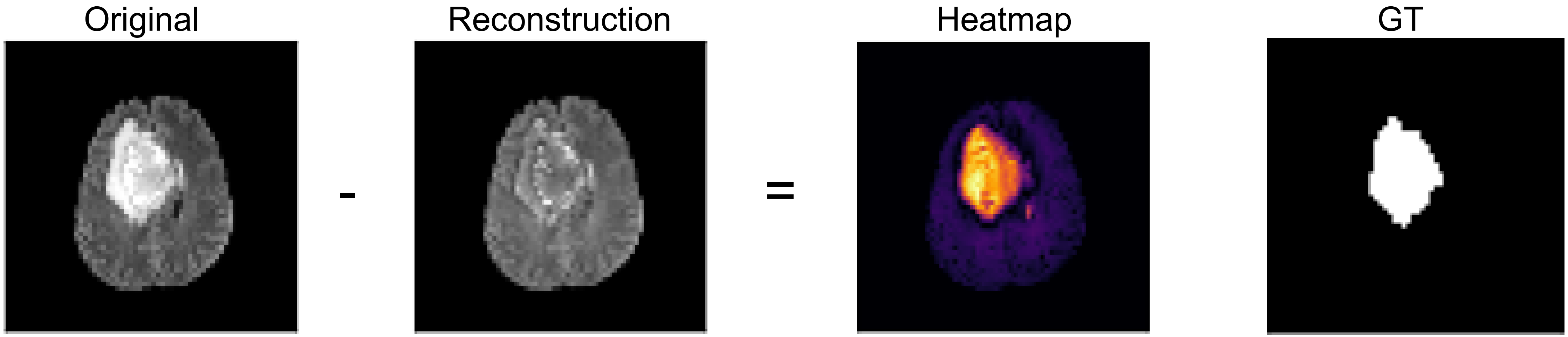


Rousseau et al. (2023) Pre-Training with Diffusion models for Dental Radiography segmentation. MICCAI 2023

Anomaly detection

Examples from the community

The simple setup of the problem

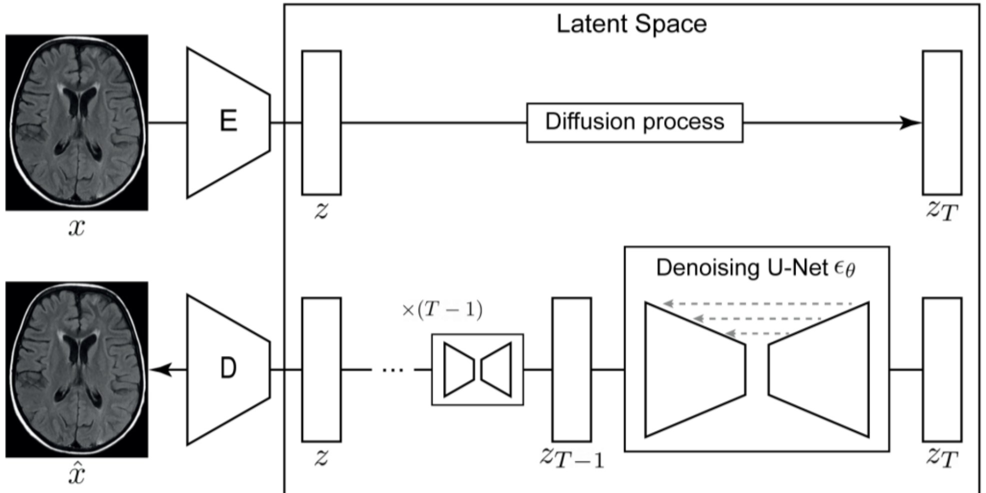


PAPERS

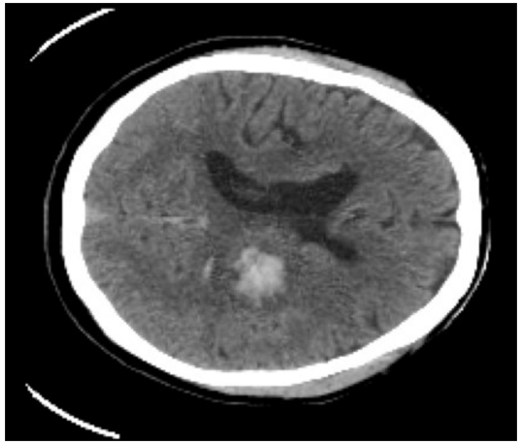
Sanchez et al. (2022) What is Healthy? Generative Counterfactual Diffusion for Lesion Localization. MICCAI workshop
Pinaya et al (2022) Fast Unsupervised Brain Anomaly Detection and Segmentation with Diffusion Models. MICCAI
Wolleb et al (2022) Diffusion Models for Medical Anomaly Detection. MICCAI
Wyatt et al (2022) AnoDDPM: Anomaly Detection with Denoising Diffusion Probabilistic Models using Simplex Noise. CVPR workshop
Kascenas et al (2023) The role of noise in denoising models for anomaly detection in medical images. Medical Image Analysis
Behrendt, Finn, et al. (2023) "Patched diffusion models for unsupervised anomaly detection in brain mri." Medical Imaging with Deep Learning
Liang, Ziyun, et al. (2023) "Modality Cycles with Masked Conditional Diffusion for Unsupervised Anomaly Segmentation in MRI." arXiv preprint arXiv:2308.16150.

Unsupervised Anomaly Segmentation

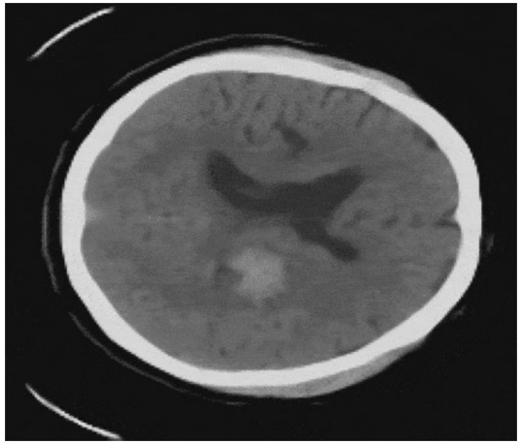
- Latent Diffusion Model (LDM) learns the distribution of healthy brain data
- Compression (Vector-Quantised VAE) scales for high-resolution images



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



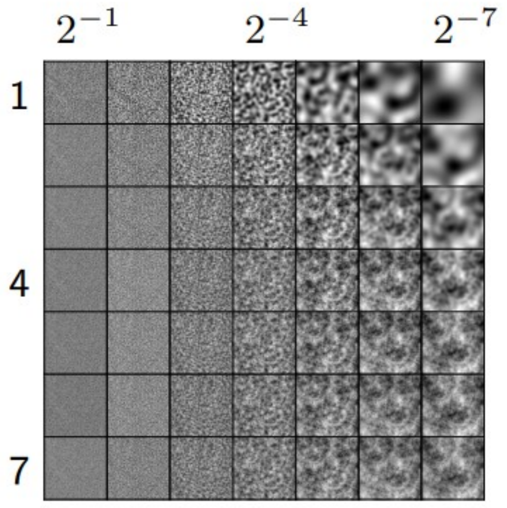
Reverse/denoising process is used to **inpaint** these regions and “**heal**” the possible anomalies



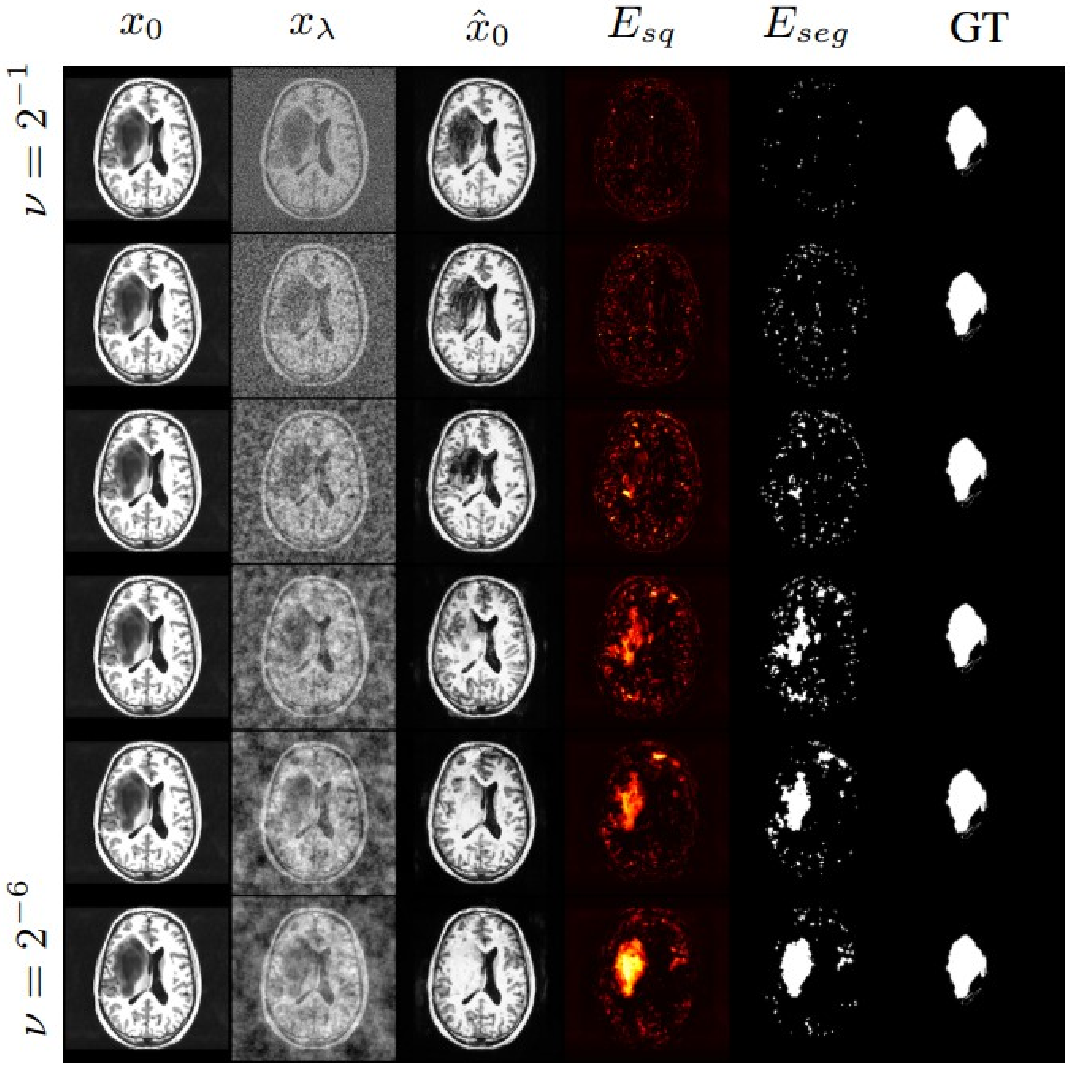
Pinaya et al (2022). Fast Unsupervised Brain Anomaly Detection and Segmentation with Diffusion Models. DOI:10.1007/978-3-031-16452-1_67

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.



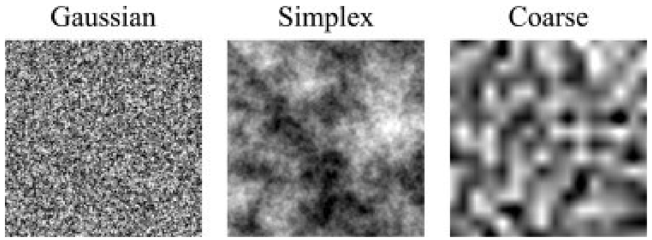
(a) Structures of simplex noise

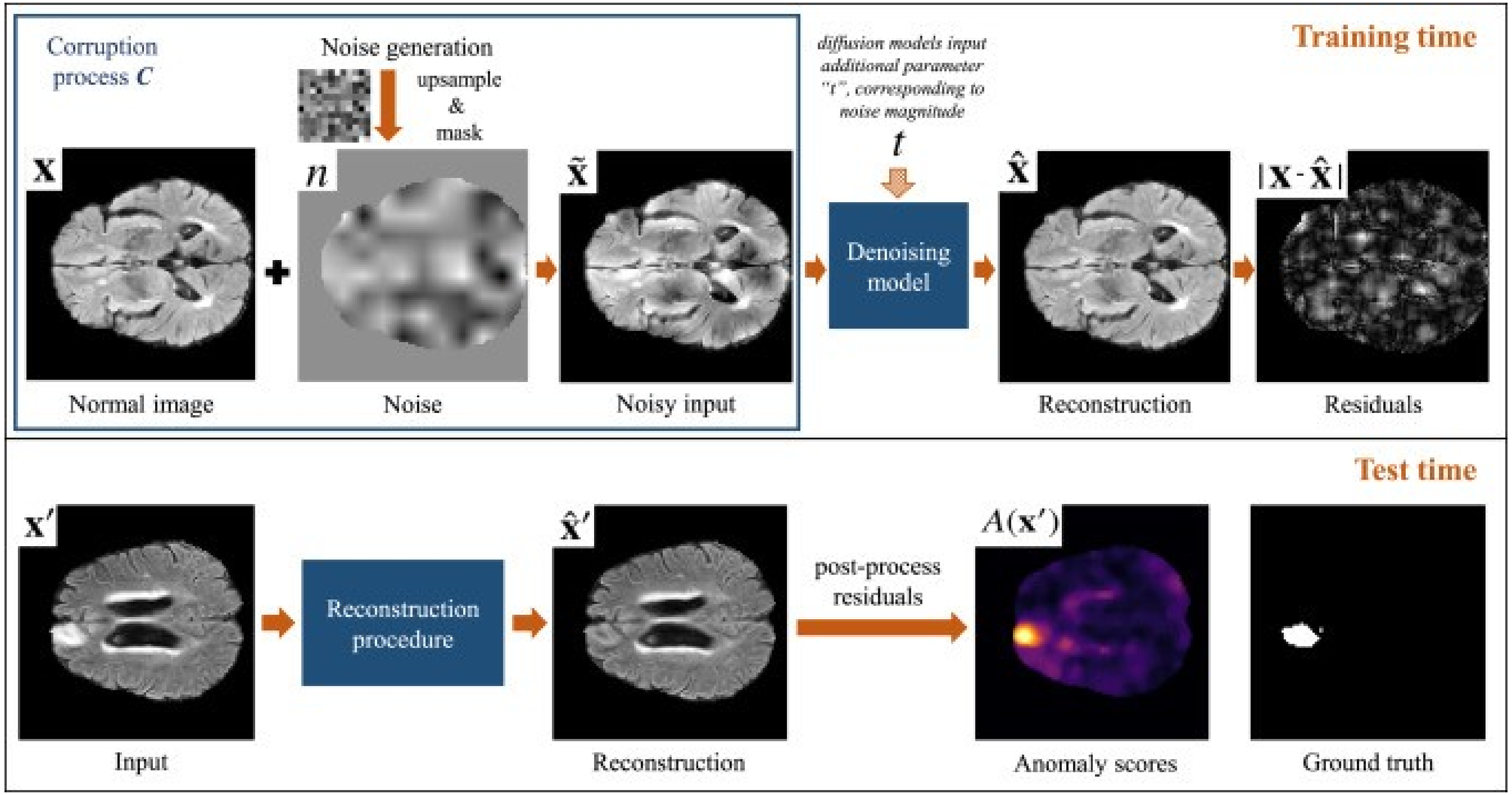


simplex noise scale *controls* target anomaly size

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

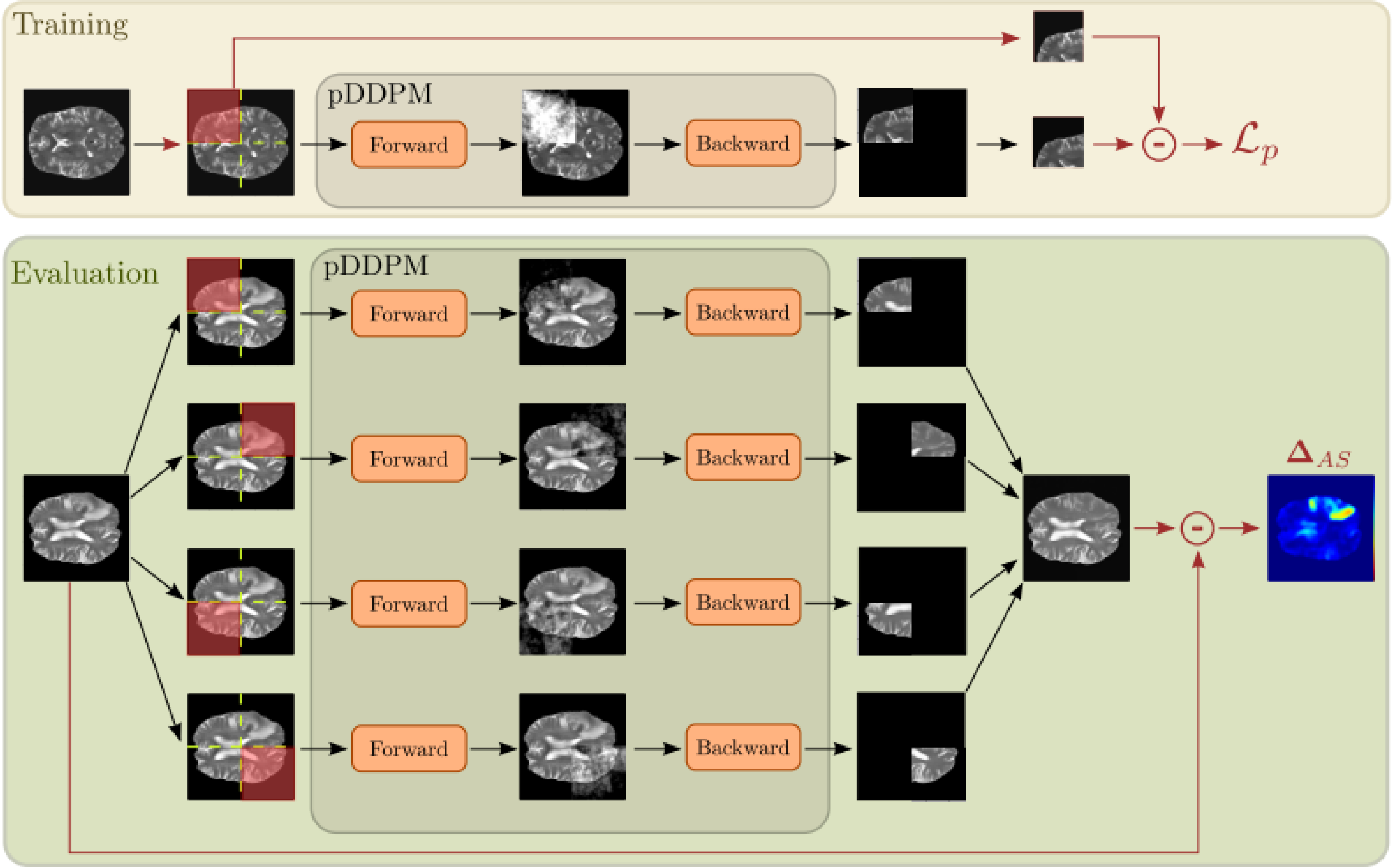
Anomaly Detection with Coarse Noise





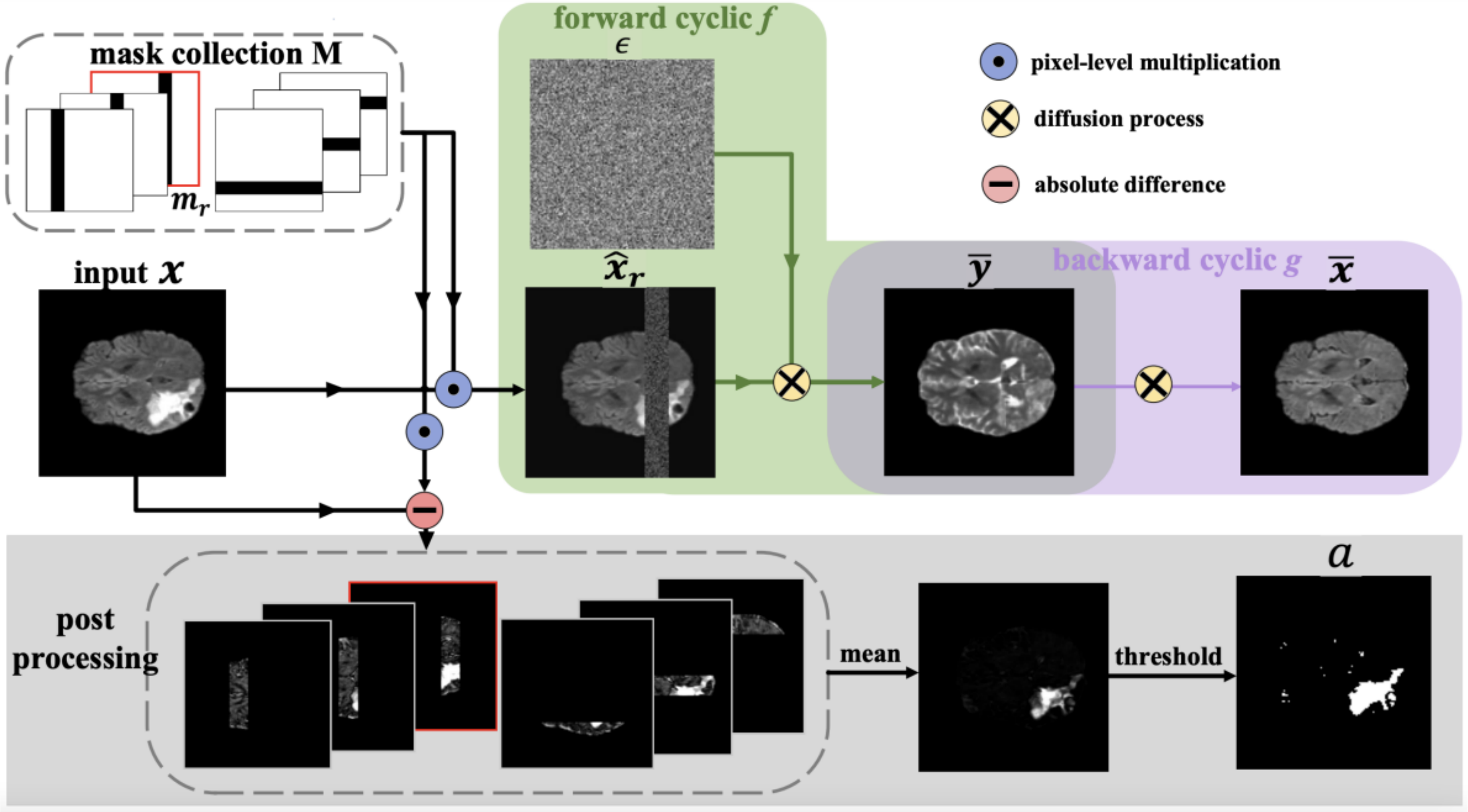
Kascenas et al (2023) The role of noise in denoising models for anomaly detection in medical images. Medical Image Analysis

Anomaly Detection from Patches



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

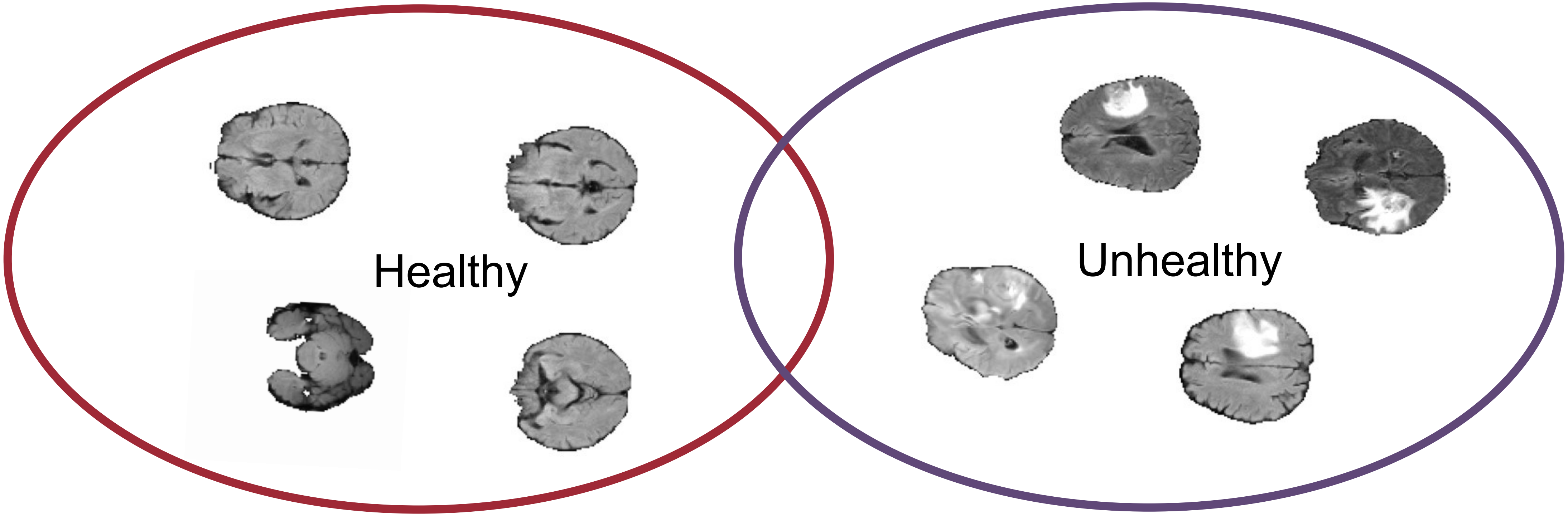
Anomaly Detection from Modality Cycles



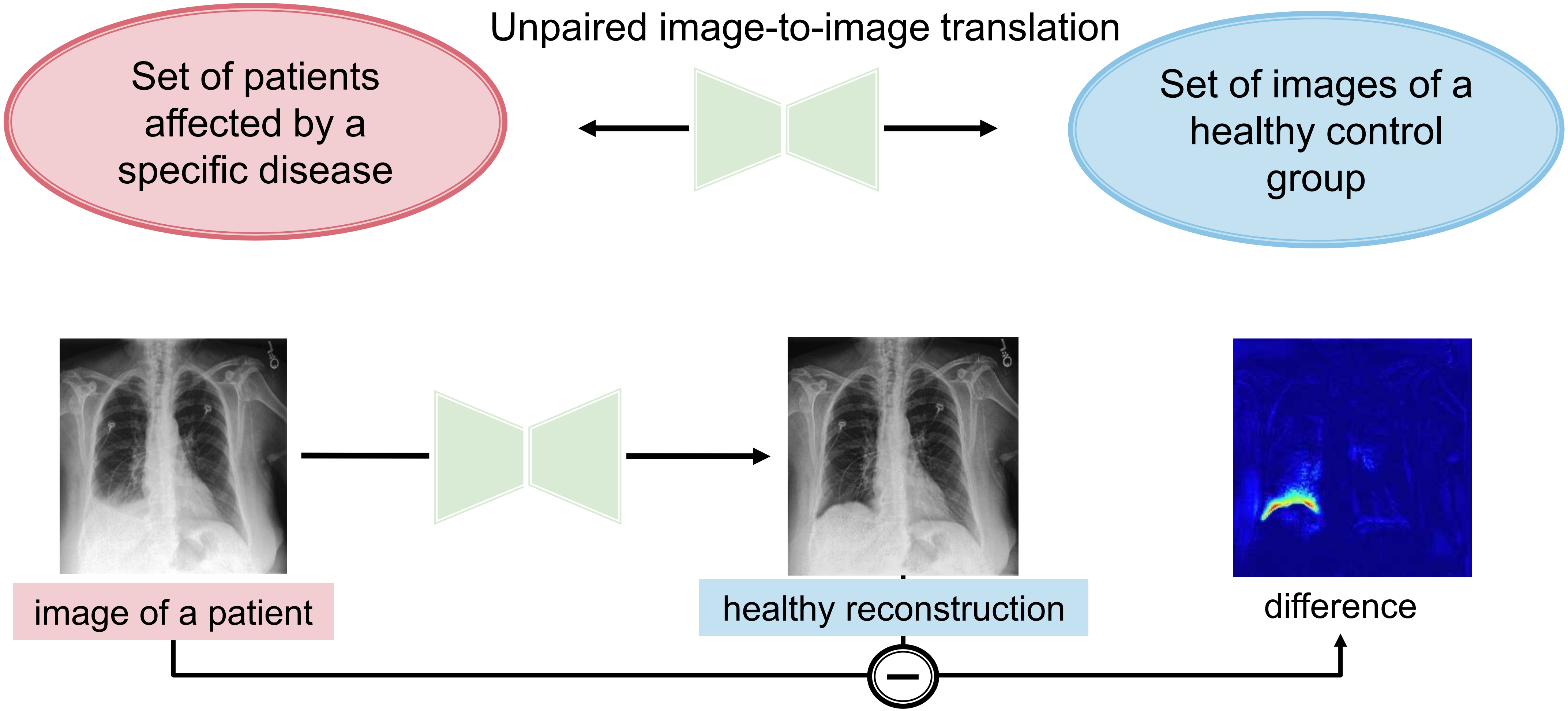
Liang, Ziyun, et al. (2023) "Modality Cycles with Masked Conditional Diffusion for Unsupervised Anomaly Segmentation in MRI." arXiv preprint arXiv:2308.16150.

Weakly Supervised Lesion Detection

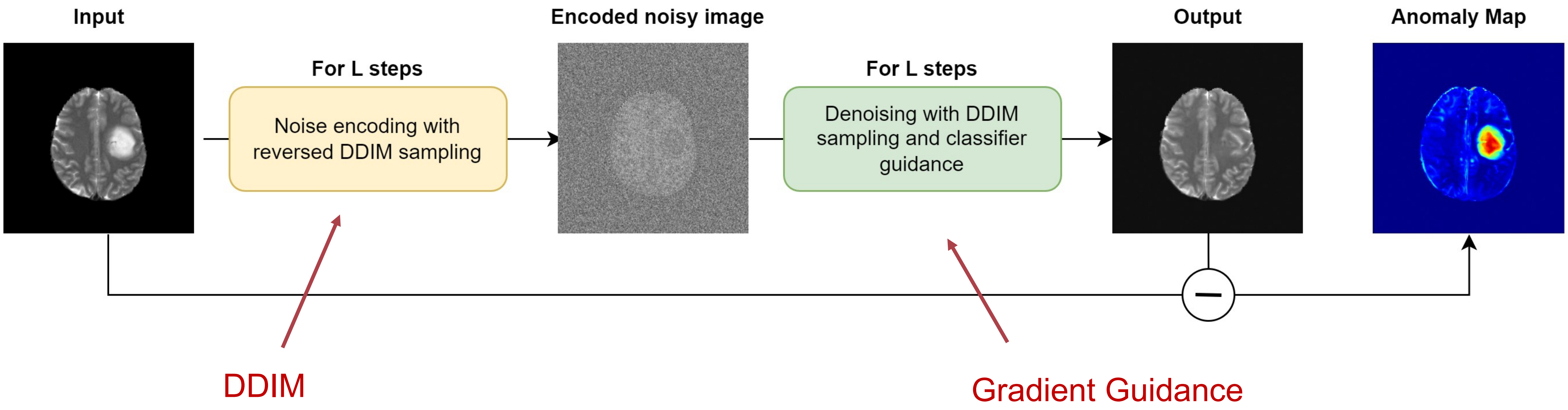
- **Goal:** Pixel-wise anomaly detection using image-level labels only

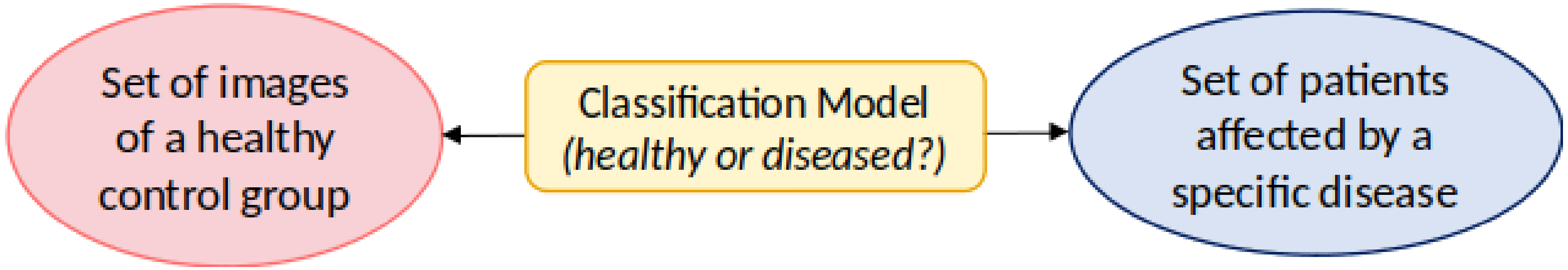


Weakly Supervised Lesion Detection



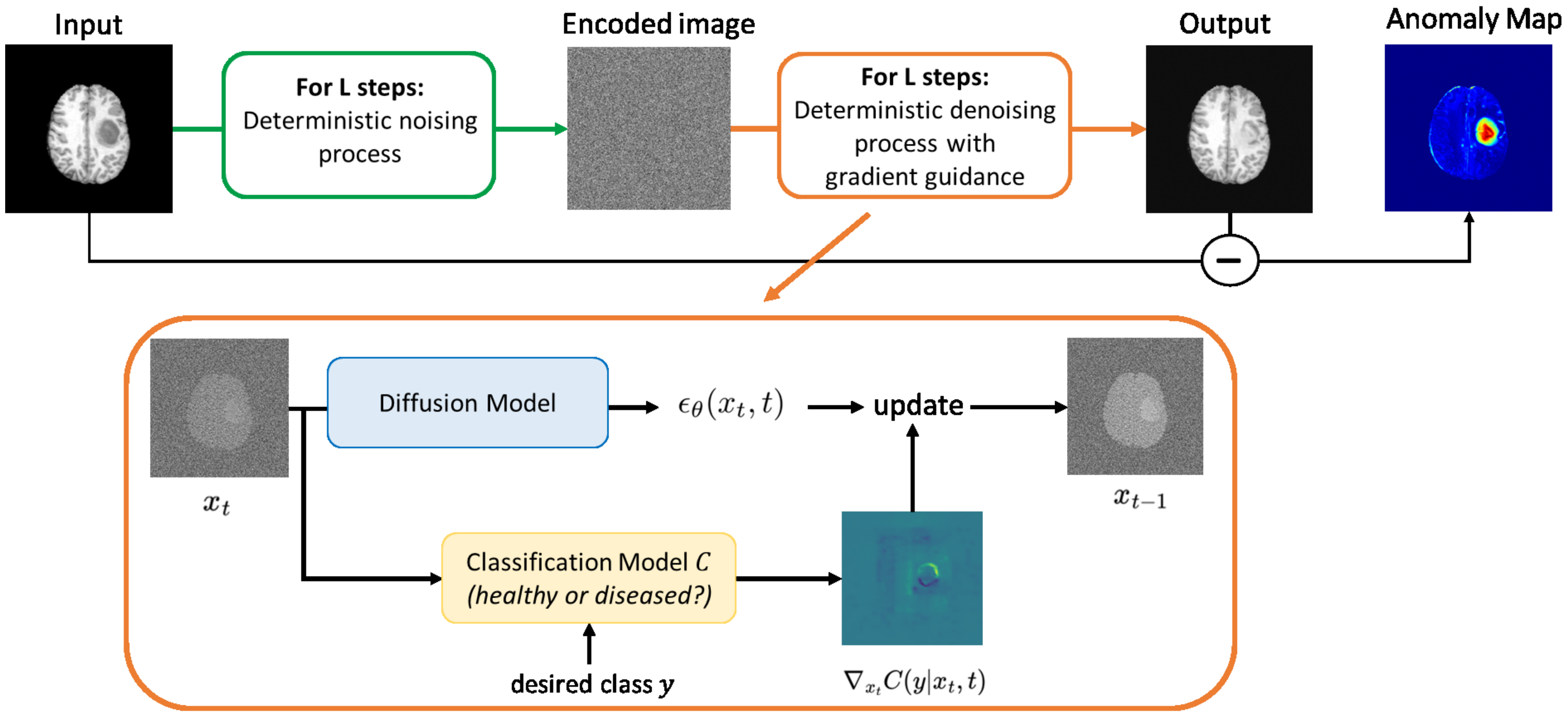
Weakly Supervised Lesion Detection





Wolleb et al (2022). Diffusion Models for Medical Anomaly Detection, *MICCAI 2022.* arXiv:2203.04306

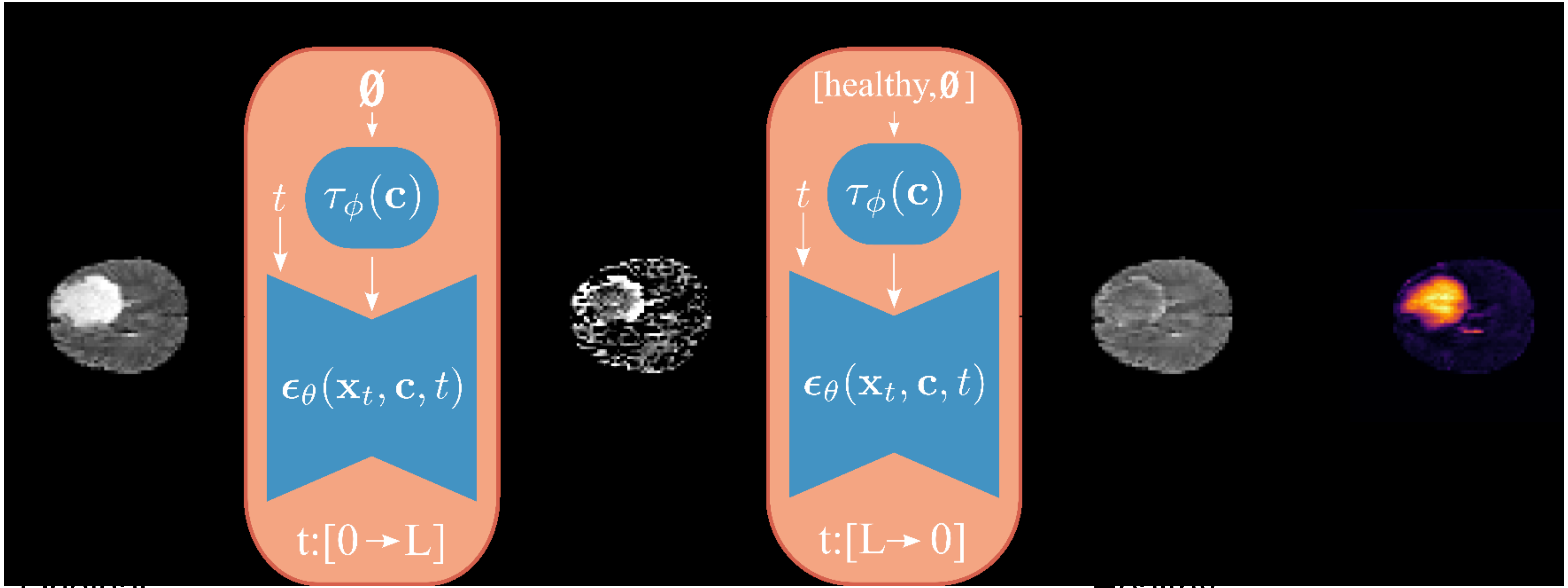
Gradient Guidance



Lesion Localization with Diffusion Models

Classifier-free guidance

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



Original Image

Latent Space

Healthy Counterfactual

Heatmap

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

Image Reconstruction

Image Registration

Anomaly Detection

Image Segmentation

Image-to-Image Translation

Image Synthesis

Inpainting

MICCAI Vancouver CANADA

Dr. Walter Pinaya

synthesia

Prof. Jong Chul Y

KAIST

Prof. Jorge Cardoso

KING'S College LONDON

Prof. Dorit Merhof

UR Universität Regensburg

Panel Discussion

Diffusion Models for Medical Images

Dr. Julia Wolleb

Universität Basel

Prof. Bernhard Kainz

FAU Friedrich-Alexander-Universität Erlangen-Nürnberg

Imperial College London

Dr. Alison O'Neil

Canon CANON MEDICAL

Prof. Sotirios A. Tsaftaris

THE UNIVERSITY of EDINBURGH

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