Diffusion Models in Image Segmentation

Milan Szabo

ELTE

January 2024

Introduction

- First proposed by Sohl et al.[5]
- Inspired by non-equilibrium statistical physics
- They can achieve remarkable results in generative modelling
- Original purpose is image synthesis
- Various other computer vision tasks
- They are highly flexible and tractable
- The research is still in an early phase

DDPM: Based on Ho et al.[3]

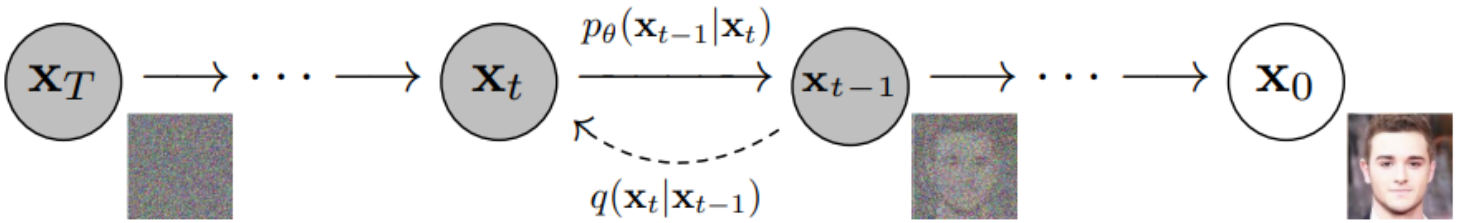

Figure: Diffusion Process from Ho et al.[3]

DDPM: Foward and Reverse Process

- $q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I)$
- $q(x_t|x_0) = \mathcal{N}(x_t; \sqrt{\overline{\alpha_t}}x_0, (1-\overline{\alpha_t})I)$
- $\alpha_t = 1 - \beta_t, \overline{\alpha}_t = \prod_{s=1}^{t} \alpha_t$
- The reverse process depends on the data distribution, so we have to estimate it.
- $p_\theta(x_t|x_{t-1}) = \mathcal{N}(x_t; \mu_\theta(x_{t-1}, t), \Sigma_\theta(x_{t-1}, t))$

DDPM: Loss function

$$\mathbb{E}\left[-\log p_\theta(\mathbf{x}_0)\right] \leq \mathbb{E}_q\left[-\log \frac{p_\theta(\mathbf{x}_{0:T})}{q(\mathbf{x}_{1:T}|\mathbf{x}_0)}\right] = \mathbb{E}_q\left[-\log p(\mathbf{x}_T) - \sum_{t\geq 1} \log \frac{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)}{q(\mathbf{x}_t|\mathbf{x}_{t-1})}\right] =: L$$

$$L_{\text{simple}}(\theta) := \mathbb{E}_{t,\mathbf{x}_0,\boldsymbol{\epsilon}}\left[\left\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, t)\right\|^2\right]$$

DDPM: Training and Sampling

Algorithm 1 Training

1: **repeat**
2: $\mathbf{x}_0 \sim q(\mathbf{x}_0)$
3: $t \sim \text{Uniform}(\{1, \dots, T\})$
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** $\mathbf{x}_0$

Noise Conditioned Score Networks [7]

- $\sigma_1 < \sigma_2 < \ldots < \sigma_T$ is a sequence of Gaussian noise scales
- $p_{\sigma_1}(x) \sim p(x_0), p_{\sigma_T}(x) \sim \mathcal{N}(0, I)$
- $p_{\sigma_t}(x_t|x) \sim \mathcal{N}(x_t; x, \sigma_t I)$
- Estimate $\nabla_{x_t} p_{\sigma_t}(x_t)$
- $\nabla_{x_t} p_{\sigma_t}(x_t|x) = \frac{x_t - x}{\sigma_t}$
- The following loss function is to be minimized:

$$\frac{1}{T}\sum_{t=1}^{T}\lambda(\sigma_T)E_{p(x)}E_{x_t\sim p_{\sigma_t}(x_t|x)}||s(x_t,\sigma_t) - \frac{x_t - x}{\sigma_t}||$$

- Sampling with Annealed Langevin dynamics

Stochastic Differential Equations [8]

- ω_t are standard normal variables
- σ is a time-dependent function that computes the diffusion coefficient
- f computes the drift coefficient.
- The following SDE gives us the diffusion process:

$$\frac{\partial x}{\partial t} = f(x,t) + \sigma(t)\omega_t \iff \partial x = f(x,t)\cdot\partial t + \sigma(t)\cdot\partial\omega$$

- And its reverse is:

$$\partial x = [f(x,t) - \sigma(t)^2 \cdot \nabla_x \log p_t(x)]\cdot\partial t + \sigma(t)\cdot\partial\hat{\omega}$$

- We aim to estimate $\nabla_x \log p_t(x)$, and sample with numerical SDE solvers.

The Dataset

- COVID-QU [2]
- 33,920 chest X-ray of which 11,956 has COVID-19, 11,263 has non-COVID infections (Viral or Bacterial Pneumonia), and 10,701 are X-rays of normal lungs
- 1,456 Normal, 1,457 non-COVID-19 chest X-rays with lung masks
- 2,913 COVID-19 chest X-rays with lung masks
- Image size: $64 \times 64 \times 1$

Model Architecture

- Training with 1000 diffusion steps
- $\beta_0 = 0.0001$, $\beta_{1000} = 0.02$ linear noising schedule
- Sampling with 50 DDIM [6] steps
- A modified U-Net as noise estimating model

The U-Net

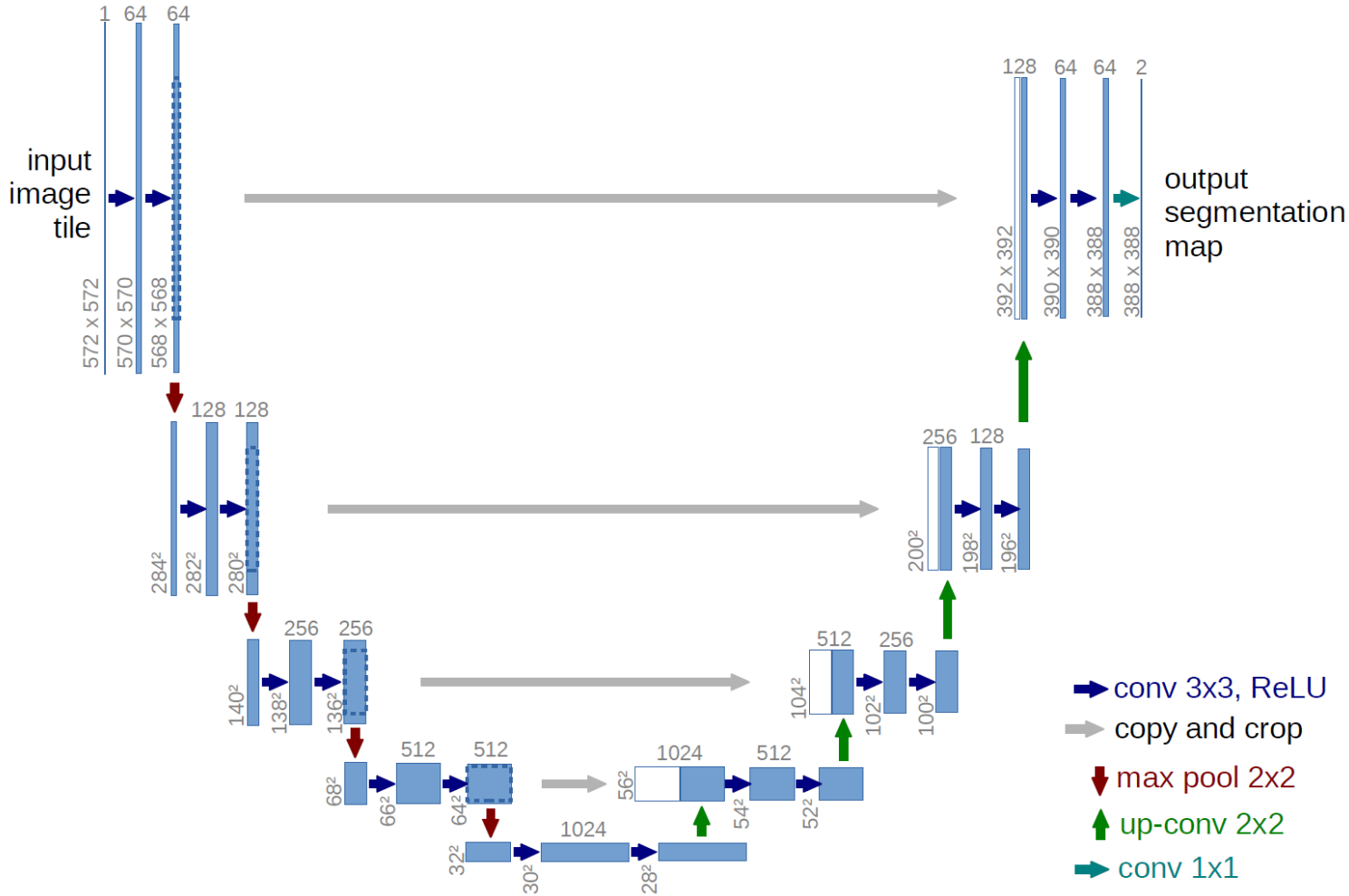

Figure: From [4]

The U-Net

- Two Residual Blocks on each level
- Down/Upsampling with Residual Blocks
- The timestep embedding is supplied to the Residual Blocks
- Linear Attention Layer at specific resolutions

Unconditional Image Synthesis

- Implement diffusion models into an experimental framework
- Test generative capabilities
- Identify and solve upcoming practical problems
- Optimize hyperparameters

Unconditional Image Synthesis

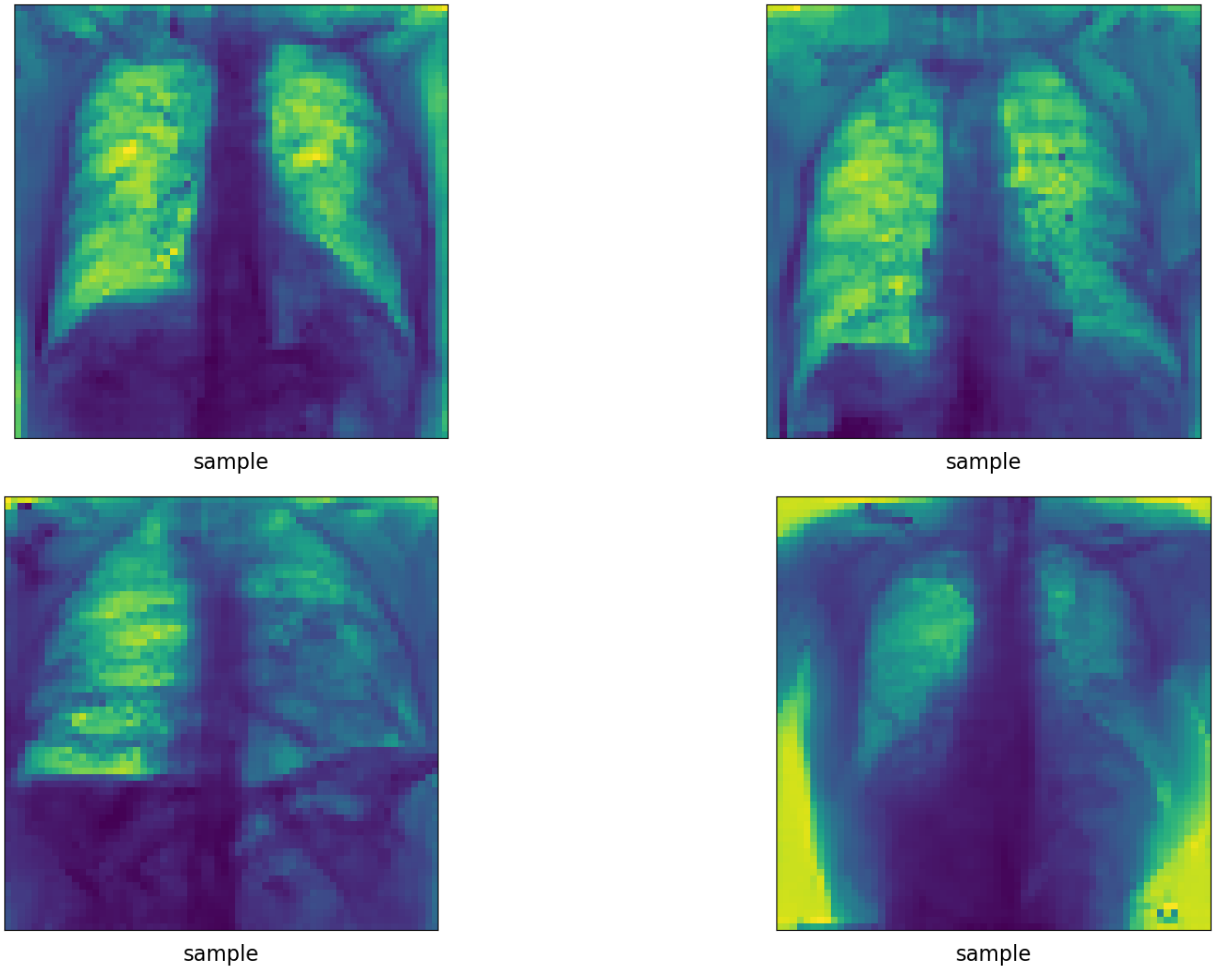

Figure: Synthetic Images of Chest X-Rays Generated from a Diffusion Model

Segmentation Results

- Classify each pixel according to its label
- Image Segmentation as Conditional Image Synthesis [1]
- Condition: Original Image to be segmented
- Generated Data: Segmentation Mask
- Estimate the conditional densities of the dataset
- How do we add conditioning to our noise estimating model?
- According to Amit et al. [1] addition should be superior to concatenation

Segmentation Results

- Segmentation metrics were measured 512 random data samples from each epoch since inference is slow.
- Confidence threshold of 0.5
- Same model architecture is used as in unconditional image synthesis
- The conditioning is first transformed before adding it to the U-Net.

Lung Segmentation

	Addition	Concatenation
DICE index	0.9529	0.9541
Jaccard index	0.9101	0.9123
Accuracy	0.9792	0.9793
Sensitivity	0.9455	0.9502
Specificity	0.9888	0.9878
Balanced Accuracy	0.9672	0.9793
Precision	0.9604	0.9581
MCC	0.9396	0.9408

Table: Measured metrics on the COVID-QU lung segmentation validation dataset from the last epoch.

Lung Segmentation

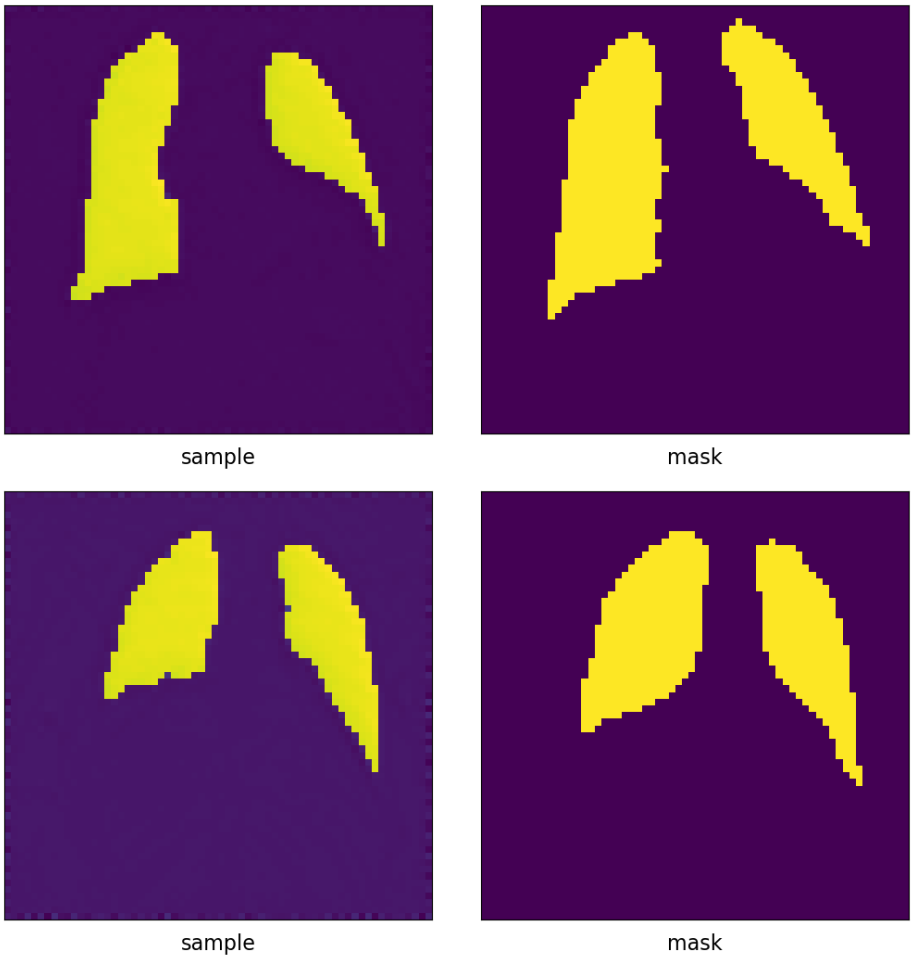

Figure: Examples of generated lung segmentation mask(left) and the original mask(right), the conditioning was supplemented via addition

Lung Segmentation

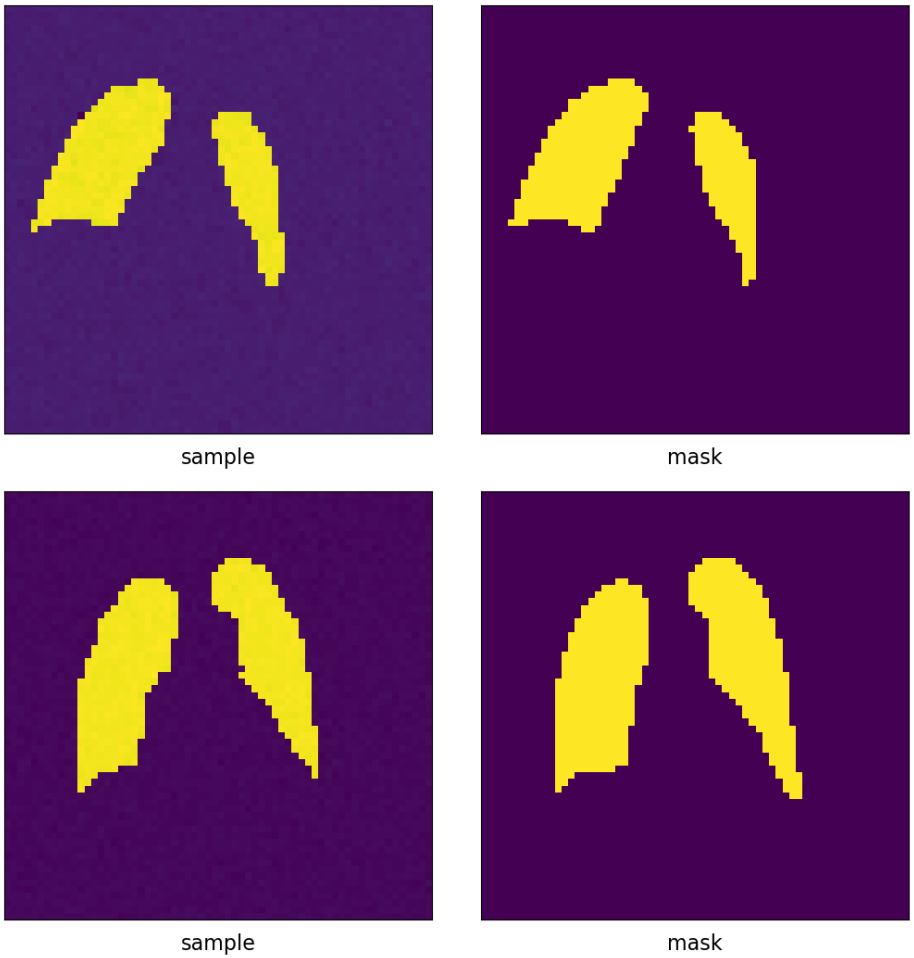

Figure: Examples of generated lung segmentation mask(left) and the original mask(right), the conditioning was supplemented via concatenation

Infection Segmentation Results

	Addition	Concatenation
DICE index	0.6755	0.4708
Jaccard index	0.5105	0.3087
Accuracy	0.9183	0.8967
Sensitivity	0.6687	0.3646
Specificity	0.9550	0.9736
Balanced Accuracy	0.8119	0.6691
Precision	0.6838	0.6674
MCC	0.6293	0.4431

Table: Measured metrics on the COVID-QU infection segmentation validation dataset from the last epoch.

Infection Segmentation

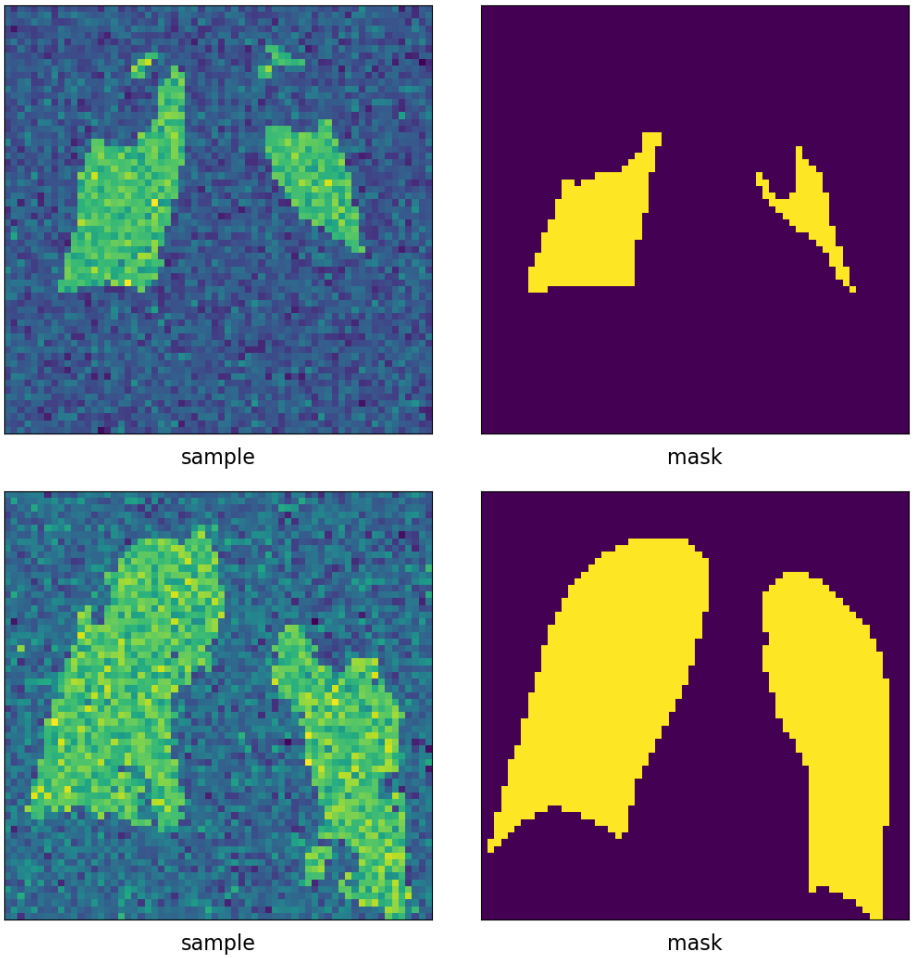

Figure: Examples of generated infection segmentation mask(left) and the original mask(right), the conditioning was supplemented via concatenation

Infection Segmentation

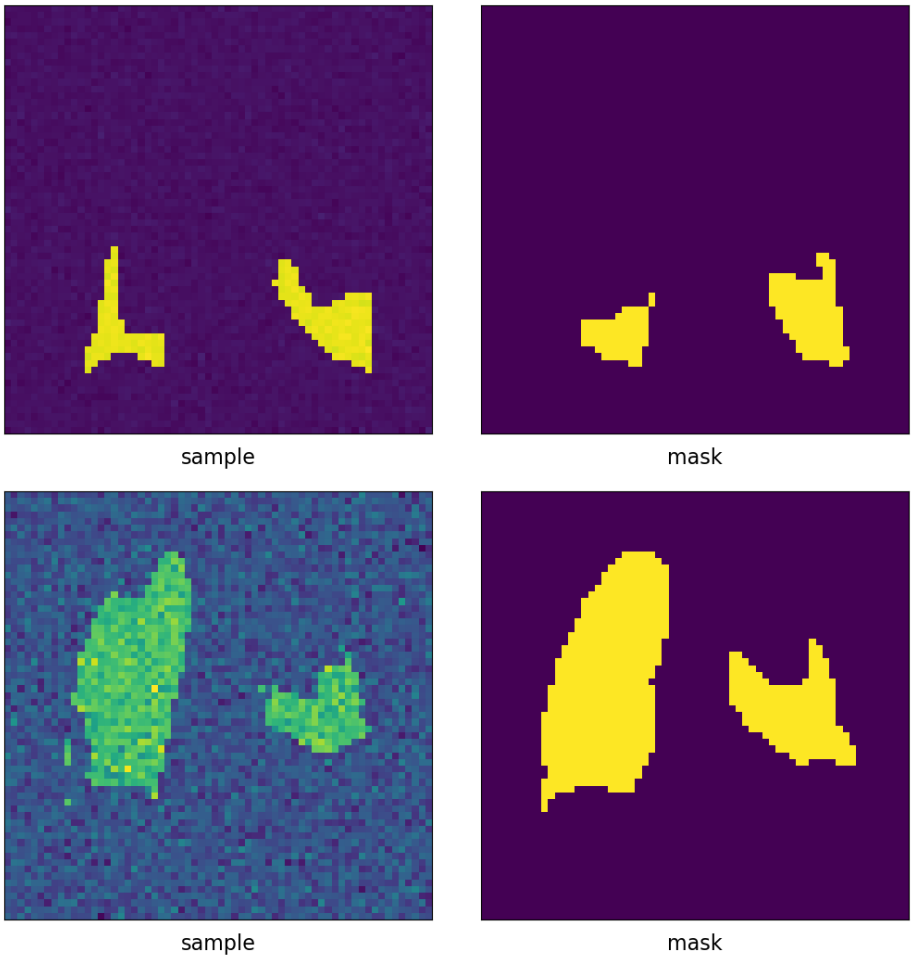

Figure: Examples of generated infection segmentation mask(left) and the original mask(right), the conditioning was supplemented via addition

References I

[1] Tomer Amit et al. "Segdiff: Image segmentation with diffusion probabilistic models". In: *arXiv preprint arXiv:2112.00390* (2021).

[2] Anas M. Tahir et al. *COVID-QU-Ex Dataset*. 2022. DOI: 10.34740/KAGGLE/DSV/3122958. URL: https://www.kaggle.com/dsv/3122958.

[3] Jonathan Ho, Ajay Jain, and Pieter Abbeel. "Denoising diffusion probabilistic models". In: *Advances in Neural Information Processing Systems* 33 (2020), pp. 6840–6851.

[4] O. Ronneberger, P.Fischer, and T. Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Vol. 9351. LNCS. (available on arXiv:1505.04597 [cs.CV]). Springer, 2015, pp. 234–241. URL: http://lmb.informatik.uni-freiburg.de/Publications/2015/RFB15a.

References II

[5] Jascha Sohl-Dickstein et al. "Deep unsupervised learning using nonequilibrium thermodynamics". In: *International Conference on Machine Learning*. PMLR. 2015, pp. 2256–2265.

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

[7] Yang Song and Stefano Ermon. "Generative Modeling by Estimating Gradients of the Data Distribution". In: *CoRR* abs/1907.05600 (2019). arXiv: 1907.05600. URL: http://arxiv.org/abs/1907.05600.

[8] Yang Song et al. "Score-Based Generative Modeling through Stochastic Differential Equations". In: *CoRR* abs/2011.13456 (2020). arXiv: 2011.13456. URL: https://arxiv.org/abs/2011.13456.