A Recycling Training Strategy for Medical Image Segmentation with Diffusion Denoising Models

Machine Learning for Biomedical Imaging Journal (MELBA) Symposium

Yunguan Fu[1,2], Yiwen Li[3], Shaheer U. Saeed[1], Matthew J. Clarkson[1], Yipeng Hu[1]

[1]University College London
[2]InstaDeep
[3]University of Oxford

June 11th 2024

What is denoising diffusion models?

Diffusion models takes Gaussian-noised input and learns to denoise it, by either predicting the noise-free data or noise.

denoise/reverse process (learned)

noise/forward process (pre-defined)

fully noised

$\mathbf{x}_T$

noise free

$\mathbf{x}_0$

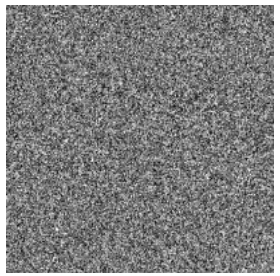
$$q(\mathbf{x}_t \mid \mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t \mathbf{I})$$

$\beta_t \in [0, 1]$ is pre-defined schedule

Special thanks to Wuhen , for graciously allowing us to use her picture for illustration.

What is denoising diffusion models?

Diffusion models takes Gaussian-noised input and learns to denoise it, by either predicting the noise-free data or noise.



denoise/reverse process (learned)

$$\mathbf{x}_T \rightleftharpoons \cdots \rightleftharpoons \mathbf{x}_t \underset{q(\mathbf{x}_t \mid \mathbf{x}_{t-1})}{\overset{p_\theta(\mathbf{x}_{t-1} \mid \mathbf{x}_t)}{\rightleftharpoons}} \mathbf{x}_{t-1} \rightleftharpoons \cdots \rightleftharpoons \mathbf{x}_0$$

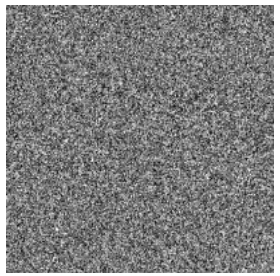
fully noised — noise/forward process (pre-defined) — noise free

Why denoising diffusion models can be generative?

Starting with different noises, the denoising paths will be different, thus generating different data samples.

What is denoising diffusion models?

Diffusion models takes Gaussian-noised input and learns to denoise it, by either predicting the noise-free data or noise.



denoise/reverse process (learned)

$$\mathbf{x}_T \rightleftharpoons \cdots \rightleftharpoons \mathbf{x}_t \underset{q(\mathbf{x}_t \mid \mathbf{x}_{t-1})}{\overset{p_\theta(\mathbf{x}_{t-1} \mid \mathbf{x}_t)}{\rightleftharpoons}} \mathbf{x}_{t-1} \rightleftharpoons \cdots \rightleftharpoons \mathbf{x}_0$$

fully noised — noise/forward process (pre-defined) — noise free

Would it be expensive to noise the image progressively?

Gaussian distribution is closed under convolution: the convolution of two Gaussian distributions results in another Gaussian distribution. This gives a closed form of x_t distribution, allow to the noised sample x_t can be direcly sampled:

$$\mathbf{x}_t \sim \mathcal{N}(\mathbf{x}_t; \sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1-\bar{\alpha}_t)\mathbf{I})$$

$\bar{\alpha}_t$ Is derived from β_t

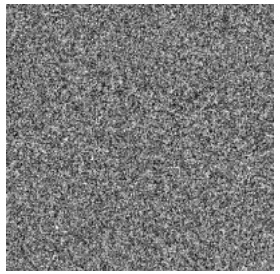
Why do we study generative segmentation models?

In this work, we aim to study the following questions:

1. For medical images, the segmentation masks may vary depends on the annotator.
   A. Could generative models provide various segmentation masks?
   B. As the inference takes multiple steps, would the model be able to improve the segmentation quality?

2. Generative models are relatively new compared to well-studied supervised models.
   A. Could diffusion model provide superior performance than standard UNet?

How is diffusion model applied to image segmentation?



denoise/reverse process (learned)

$$\mathbf{x}_T \rightleftharpoons \cdots \rightleftharpoons \mathbf{x}_t \underset{q(\mathbf{x}_t \mid \mathbf{x}_{t-1})}{\overset{p_\theta(\mathbf{x}_{t-1} \mid \mathbf{x}_t)}{\rightleftharpoons}} \mathbf{x}_{t-1} \rightleftharpoons \cdots \rightleftharpoons \mathbf{x}_0$$

fully noised

noise/forward process (pre-defined)

noise free

1. The ground truth segmentation masks are one-hot encoded, with values in {-1, 1}.
2. A linear interpolation with Gaussian noises is then performed.
3. The model takes image and noised mask to predict the ground truth.

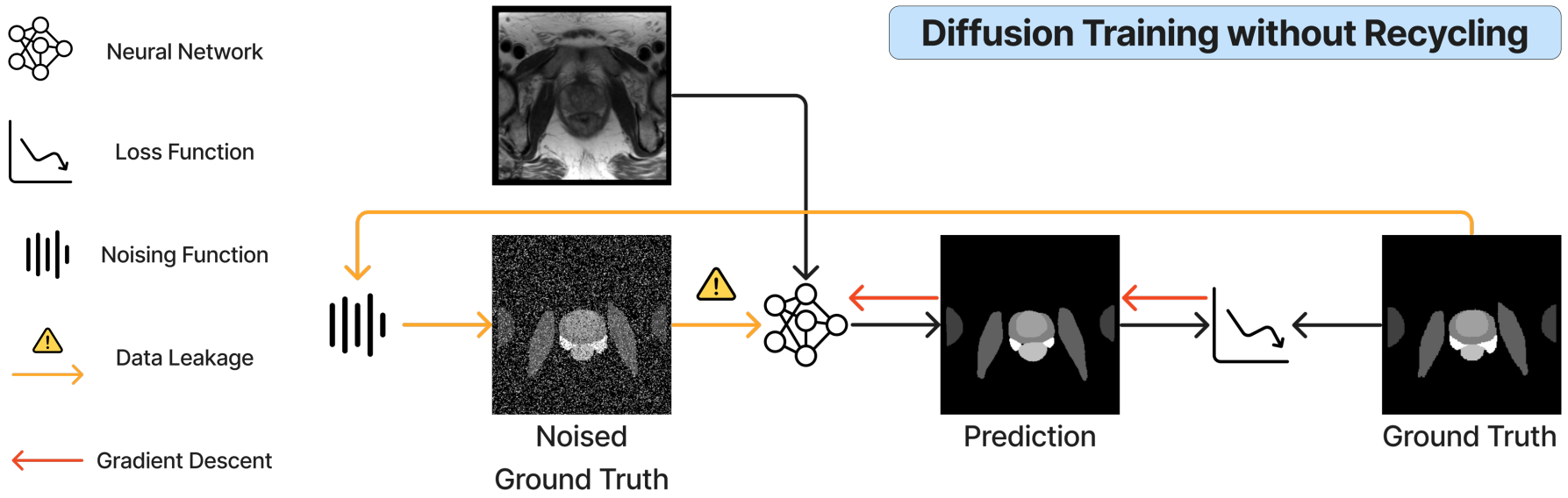
Our previous work[1] has shown that predicting segmentation mask is more effective than predicting noise. Dice loss is better than L2 loss. This trend has also been observed in recent work[2].

1. Importance of aligning training strategy with evaluation for diffusion models in 3d multiclass segmentation
2. MedSegDiff-V2: Diffusion based Medical Image Segmentation with Transformer

What is the issue in diffusion model for image segmentation?

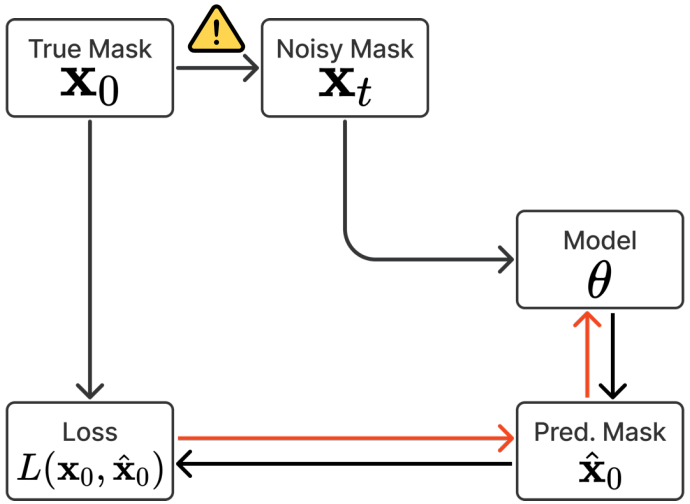
If segmentation masks and noise are mixed in raw pixel/voxel space, the ground truth object shape information remains recognizable at low noise level.

This means that the ground truth information is provided to the model during training, which poses data leakage risk.

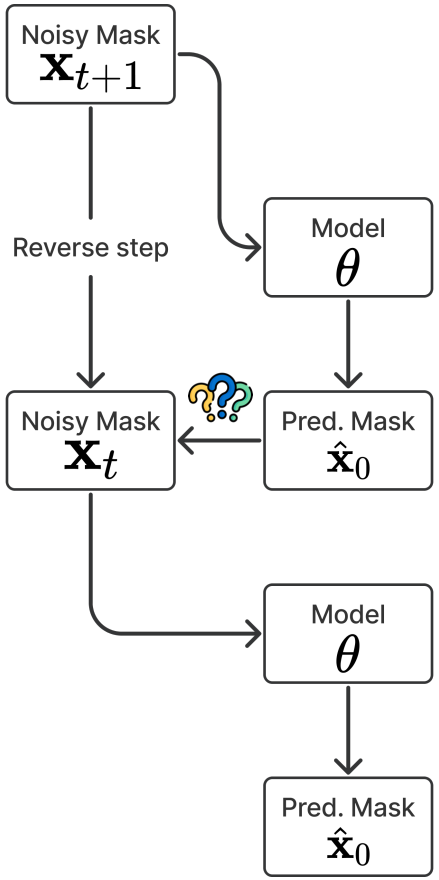


What is the issue in diffusion model for image segmentation?

During training, the noised mask is based on ground truth.
During inference, the noised mask is progressively generated by denoising a noise. No ground truth information is used.



diffusion training without recycling

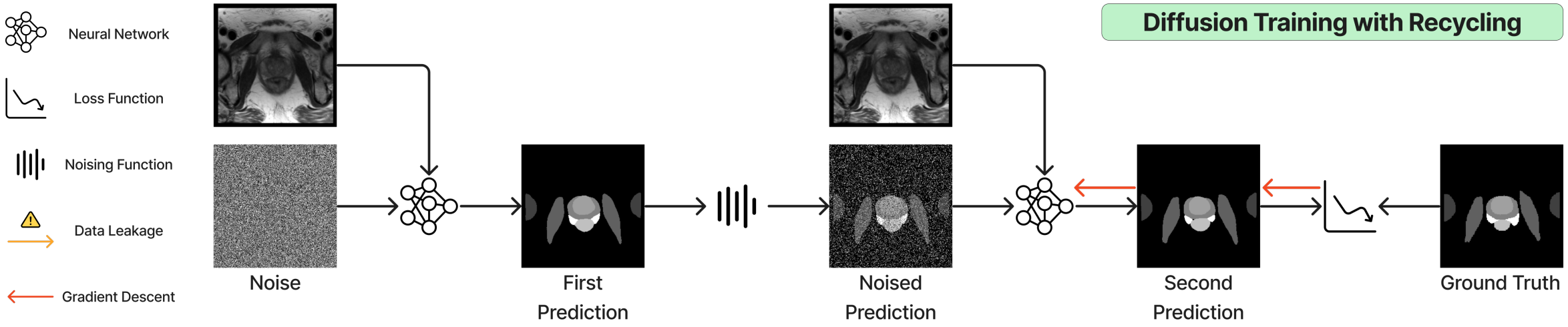


inference

image input is omitted for simplicity

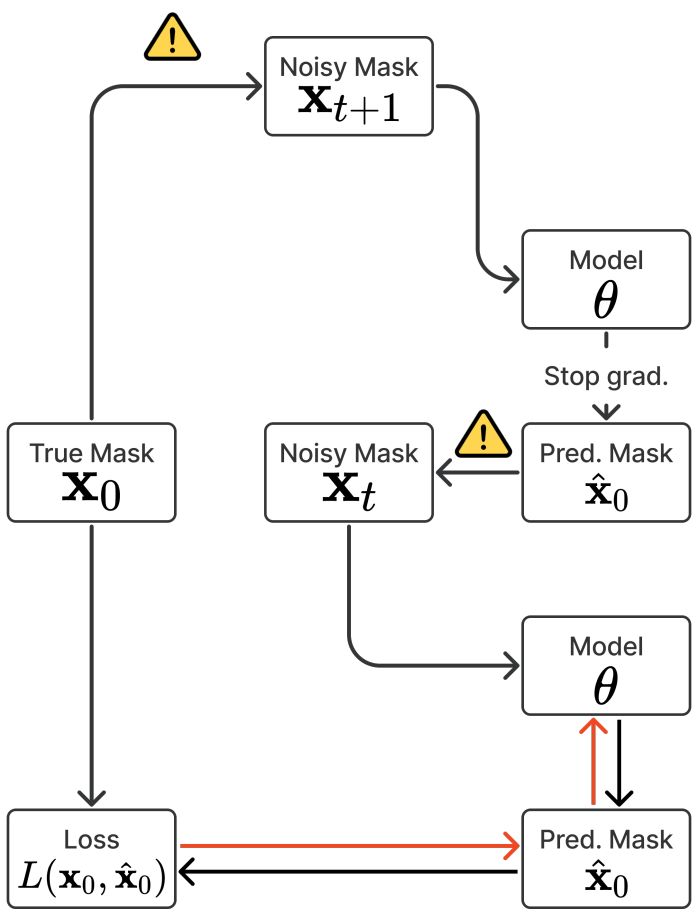
How to solve it?

Model takes the image and a noise to perform the first prediction (pre-segmentation).
This prediction is used instead of ground truth for noised mask generation.

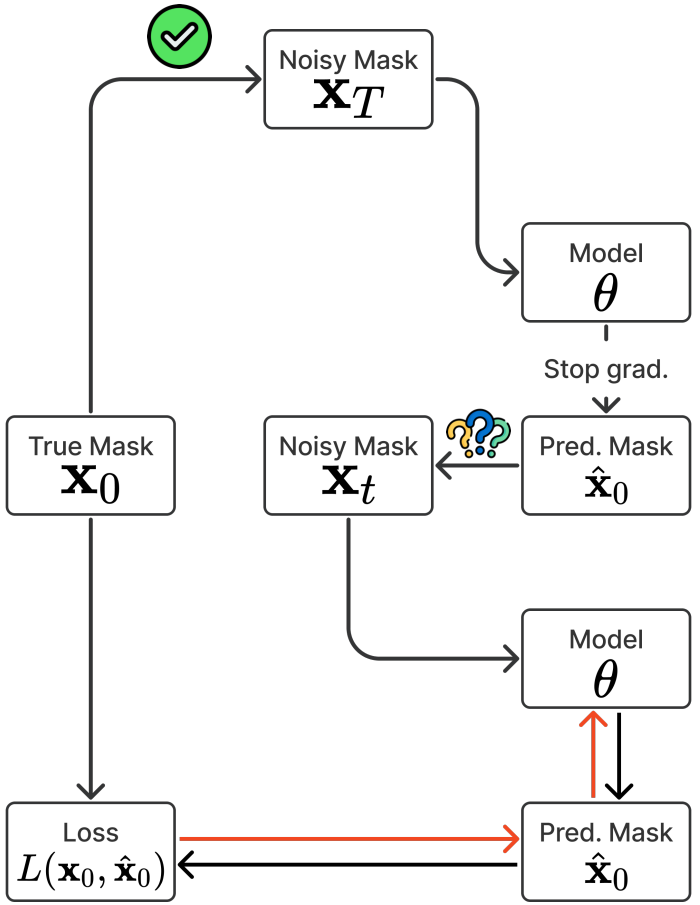


How to solve it?

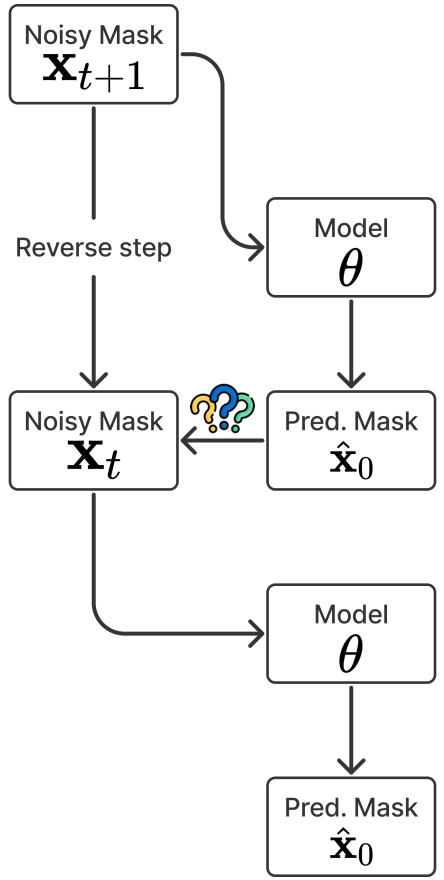
Use model's prediction instead of ground truth. This requires one additional forward.



recycling (MICCAI) Diff. rec. $\mathbf{x}_{t+1}$



recycling (MELBA) Diff. rec. $\mathbf{x}_T$

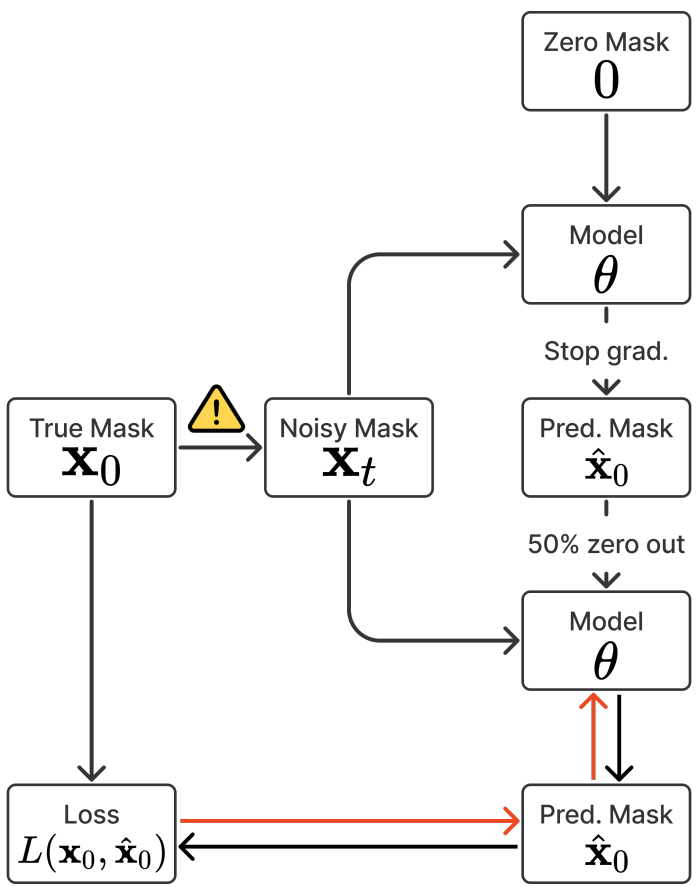


inference

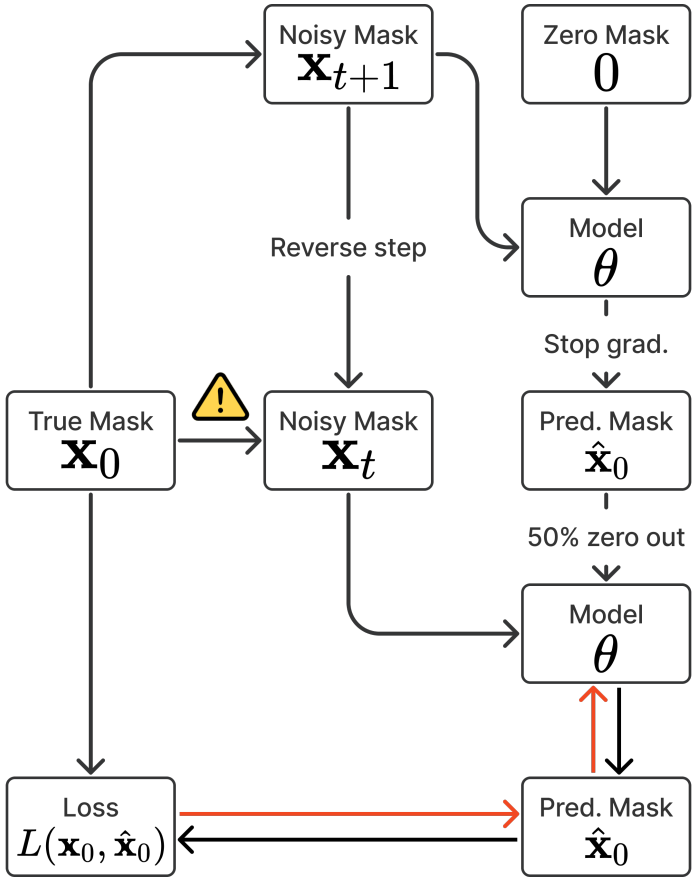
image input is omitted for simplicity

Any other methods?

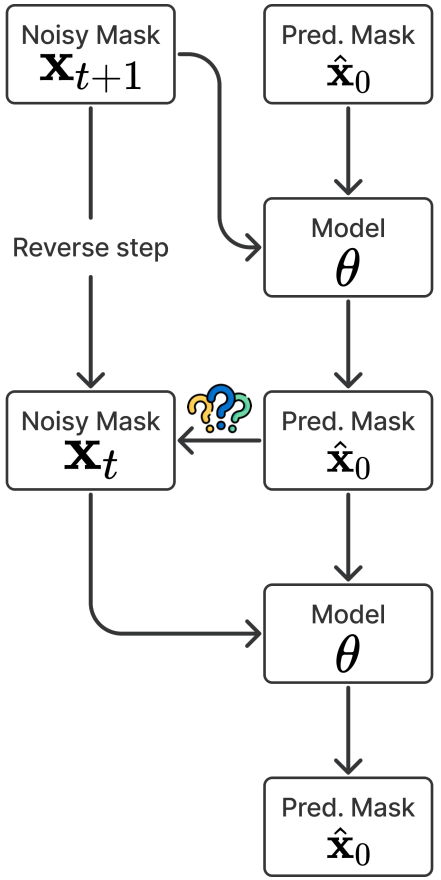
Self-conditioning methods were previously proposed. But they still takes noised ground truth and the model takes additional inputs, which further increases memory cost.



self-condition (Chen) Diff. sc. $\mathbf{x}_t$



self-condition (Watson) Diff. sc. $\mathbf{x}_{t+1}$



inference

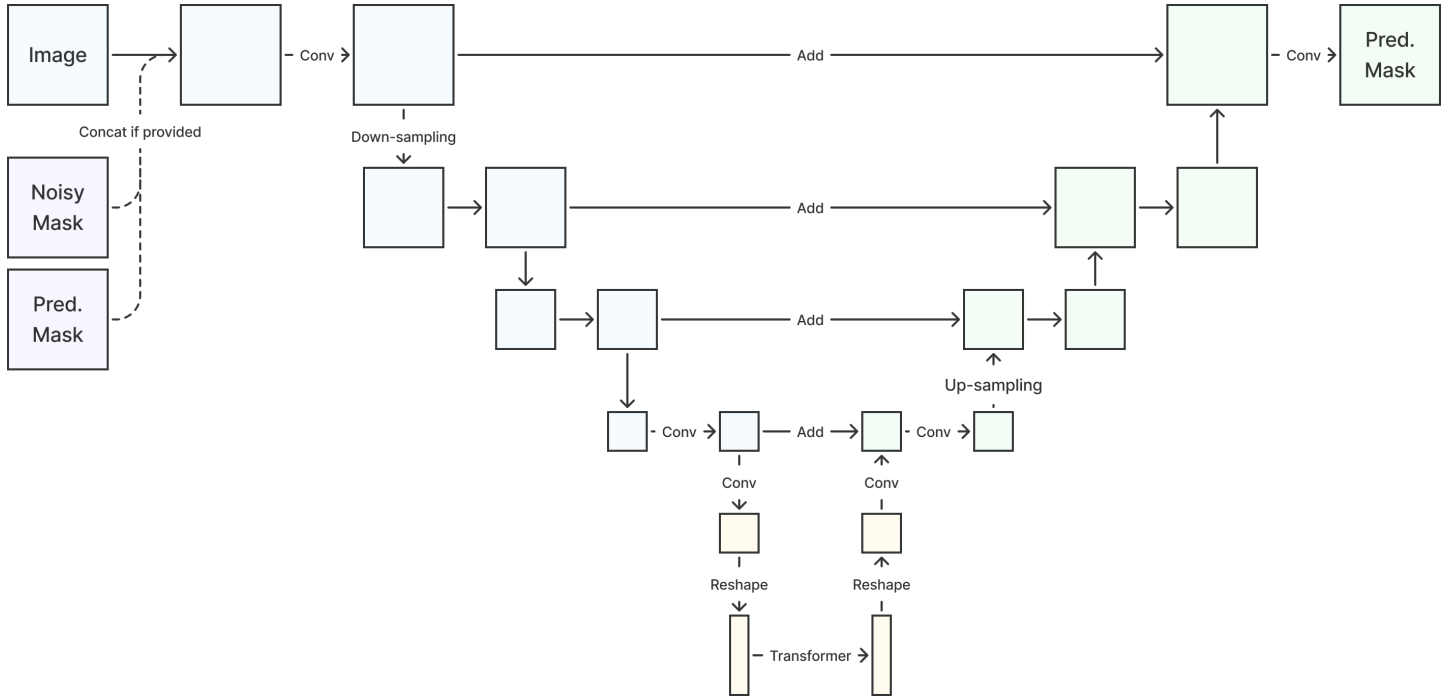
image input is omitted for simplicity

Does recycling work?

We tested the methods extensively on four different datasets. Used Unet with Transformer at bottom as network architecture, without other modules.

Dataset	#Samples	Image Size
Muscle Ultrasound	3910	(480,512)
Abdominal CT (AMOS)	300	(192,128,128)
Prostate MR	589	(256,256,32)
Brain MR	1251	(179,219,155)

dataset summary



UNet architecture

https://data.mendeley.com/datasets/3jykz7wz8d/1
https://zenodo.org/record/7155725#.ZAkbe-zP2r0
https://zenodo.org/record/7013610#.ZAkaXuzP2rM
https://www.kaggle.com/datasets/dschettler8845/brats-2021-task1

Does recycling work?

The proposed recycling method outperformed all other diffusion methods, with DDPM or DDIM sampling over 5 steps.

Method	DDPM		DDIM	
	DS ↑	HD ↓	DS ↑	HD ↓
Diff.	86.60 ± 12.38	41.11 ± 35.48	86.18 ± 12.41	42.31 ± 35.82
Diff. sc. $\mathbf{x}_t$	86.35 ± 14.14	40.42 ± 37.53	85.96 ± 13.78	42.00 ± 36.76
Diff. sc. $\mathbf{x}_{t+1}$	87.14 ± 11.48	39.24 ± 32.83	86.30 ± 11.49	41.89 ± 32.72
Diff. rec. $\mathbf{x}_{t+1}$	87.44 ± 12.39	39.68 ± 36.21	87.43 ± 12.25	39.82 ± 35.39
Diff. rec. $\mathbf{x}_T$	**88.23 ± 11.69**	**35.37 ± 31.79**	**88.21 ± 11.70**	**35.52 ± 31.91**

(a) Muscle Ultrasound

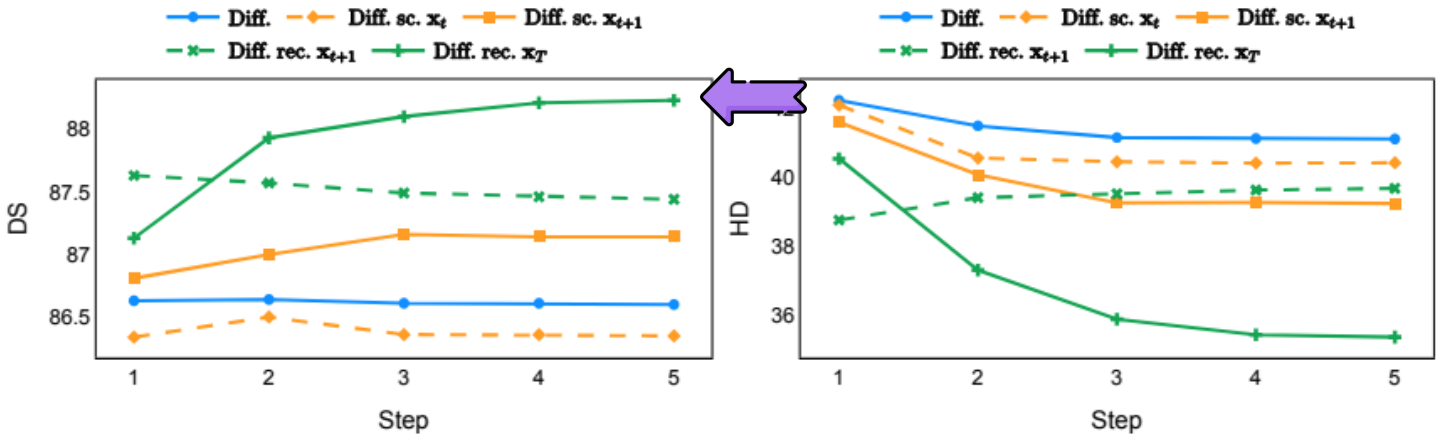
Method	DDPM		DDIM	
	DS ↑	HD ↓	DS ↑	HD ↓
Diff.	83.61 ± 4.87	5.10 ± 2.40	83.11 ± 4.81	5.00 ± 2.35
Diff. sc. $\mathbf{x}_t$	83.47 ± 4.85	5.17 ± 2.65	82.49 ± 4.88	5.42 ± 2.70
Diff. sc. $\mathbf{x}_{t+1}$	83.97 ± 4.85	4.93 ± 2.66	83.00 ± 4.89	5.10 ± 2.64
Diff. rec. $\mathbf{x}_{t+1}$	84.29 ± 5.12	4.59 ± 2.21	84.21 ± 4.89	4.96 ± 2.92
Diff. rec. $\mathbf{x}_T$	**85.54 ± 5.20**	**4.40 ± 1.96**	**85.54 ± 5.20**	**4.41 ± 1.96**

(c) Prostate MR

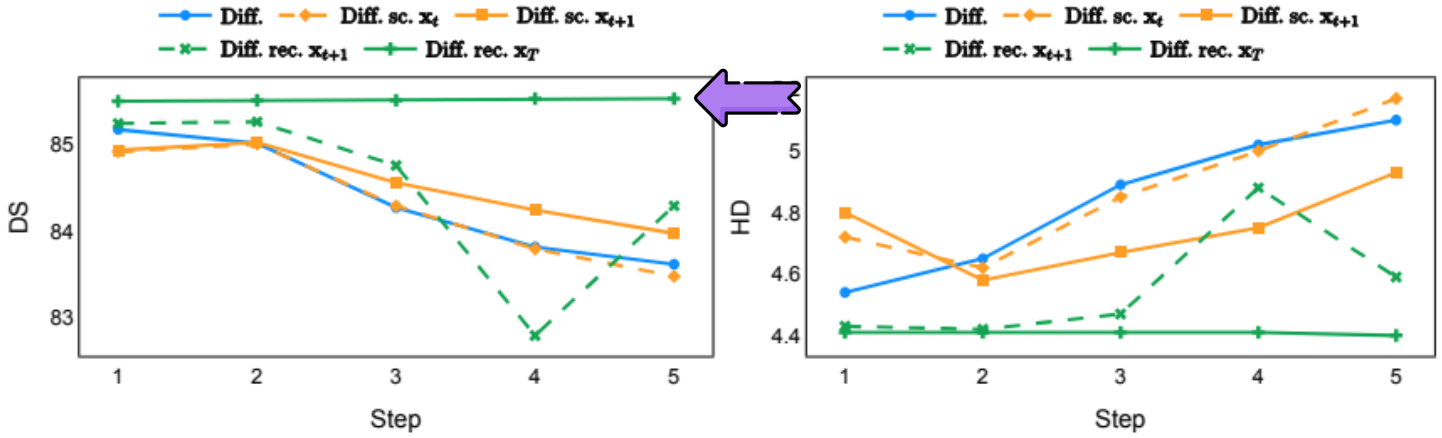
Results on Abdominal CT and brain MR are available in Table 1.

Does recycling work?

The proposed recycling method (green solid lines) maintains or improves segmentation quality.



(a) Dice Score for muscle ultrasound (b) Hausdorff distance for muscle ultrasound

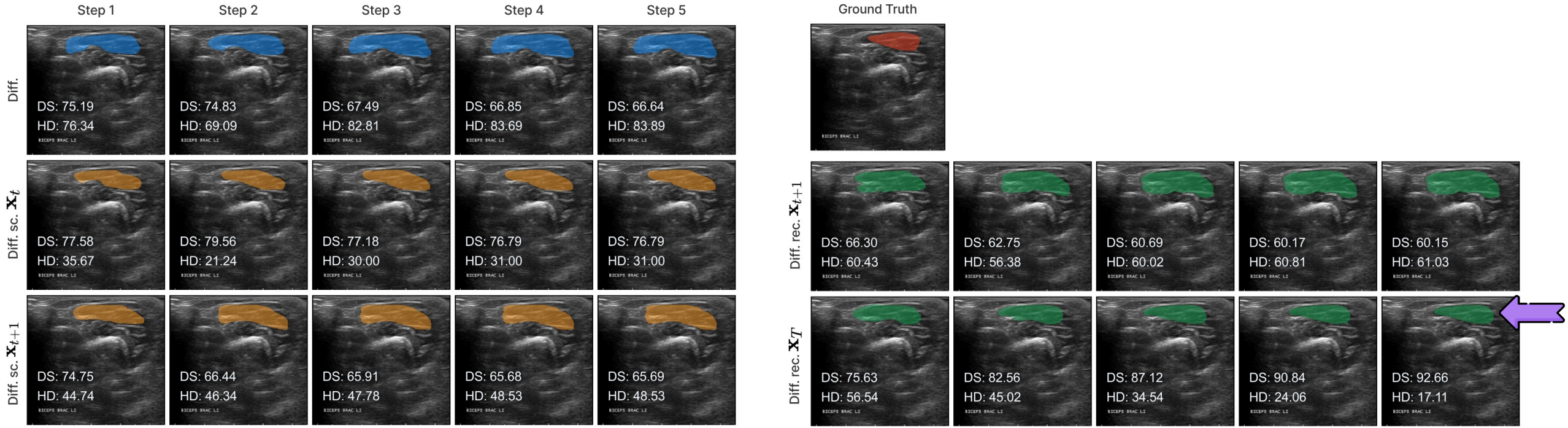


(e) Dice score for prostate MR (f) Hausdorff distance for prostate MR

Results on Abdominal CT *and brain MR are available in Figure 3.*

Does recycling work?

The proposed recycling method maintains or improves segmentation quality.



Does the model generate different segmentation masks?

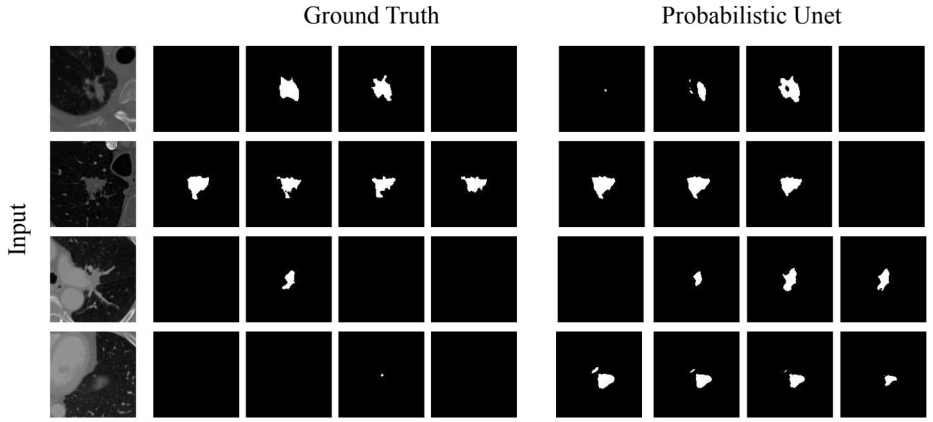
The model has large variance on Muscle Ultrasound data, but the variance decreases during inference steps: the segmentation masks converge. For other datasets, same trend has been observed but the variation is limited.

Table 4: **Diffusion model performance across different inference seeds.** For each sample, the maximum difference (Δ) across five random seeds is calculated. The average across all samples is reported.

Data Set	Mean Δ Dice Score				
	Step 1	Step 2	Step 3	Step 4	Step 5
Muscle Ultrasound	0.0212	0.0165	0.0122	0.0081	0.0051
Abdominal CT	0.0009	0.0010	0.0009	0.0008	0.0004
Prostate MR	0.0004	0.0004	0.0004	0.0004	0.0002
Brain MR	0.0005	0.0005	0.0005	0.0003	0.0001
Data Set	Mean Δ Hausdorff Distance				
	Step 1	Step 2	Step 3	Step 4	Step 5
Muscle Ultrasound	10.0582	7.0020	4.7440	3.1758	1.8164
Abdominal CT	0.1481	0.1339	0.1221	0.0751	0.0673
Prostate MR	0.0447	0.0426	0.0499	0.0431	0.0209
Brain MR	0.0758	0.0779	0.0678	0.0616	0.0197

Does the model generate different segmentation masks?

Future work can focus on ambiguous image segmentation[1], where datasets with multiple annotations per sample are used.



Examples on Lung lesion segmentation (LIDC-IDRI)

A potential question could be: by removing the data leakage in diffusion model training, could the resulted network achieve higher accuracy in segmentation while preserving the diversity.

1. Ambiguous Medical Image Segmentation using Diffusion Models

Is diffusion models better for segmentation?

Using same neural network architectures (UNet), the diffusion models with recycling methods are on par with standard supervised models.

Table 2: **Segmentation performance comparison to non-diffusion models.** "No diff." represents non-diffusion model. "Diff. rec. $\mathbf{x}_T$" represents the diffusion model with proposed recycling. "Ensemble" represents the model averaging the probabilities from "No diff." and "Diff. rec. $\mathbf{x}_T$". The inference sampler is DDPM. The best results are in bold and underline indicates the difference to non-diffusion model is significant with p-value < 0.05.

Data Set	Method	DS ↑	HD ↓
Muscle Ultrasound	No diff.	88.15 ± 10.77	36.86 ± 30.04
	Diff. rec. $\mathbf{x}_T$	88.23 ± 11.69	35.37 ± 31.79
	Ensemble	**88.88 ± 10.59**	**34.01 ± 28.75**
Abdominal CT	No diff.	87.59 ± 5.10	6.36 ± 3.86
	Diff. rec. $\mathbf{x}_T$	87.45 ± 5.43	6.56 ± 5.44
	Ensemble	**88.29 ± 5.21**	**5.60 ± 3.13**
Prostate MR	No diff.	85.22 ± 5.18	4.62 ± 2.37
	Diff. rec. $\mathbf{x}_T$	85.54 ± 5.20	4.40 ± 1.96
	Ensemble	**85.95 ± 5.12**	**4.32 ± 2.01**
Brain MR	No diff.	92.43 ± 9.10	5.20 ± 9.56
	Diff. rec. $\mathbf{x}_T$	92.29 ± 8.55	7.03 ± 13.48
	Ensemble	**92.67 ± 8.60**	**5.03 ± 8.41**

Is diffusion models better for segmentation?

The diffusion model in this work have multiple components that can be adjusted

1. Noising strategy
   - The noise can be added at latent space, using latent diffusion models[1].
   - Categorical noise (e.g. Bernoulli[2]) can be used in pixel/voxel space.
2. Network architecture
   - Diffusion-specific architecture can be used to align noise and image embedding[3].
3. Compute
   - Recycling forward pass can be combined with mean teacher/distillation.
   - Recycling can be used to fine-tune any pre-existing segmentation network[5].

1. High-Resolution Image Synthesis with Latent Diffusion Models
2. BerDiff: Conditional Bernoulli Diffusion Model for Medical Image Segmentation
3. MedSegDiff-V2: Diffusion based Medical Image Segmentation with Transformer
4. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results
5. Accelerating Diffusion Models Via Pre-Segmentation Diffusion Sampling for Medical Image Segmentation

A Recycling Training Strategy for Medical Image Segmentation with Diffusion Denoising Models

What's the task?
Image segmentation.

How to segment with diffusion models?
Progressively denoising a noised mask.

What's the concern?
Noised ground truth poses risks of data leakage.

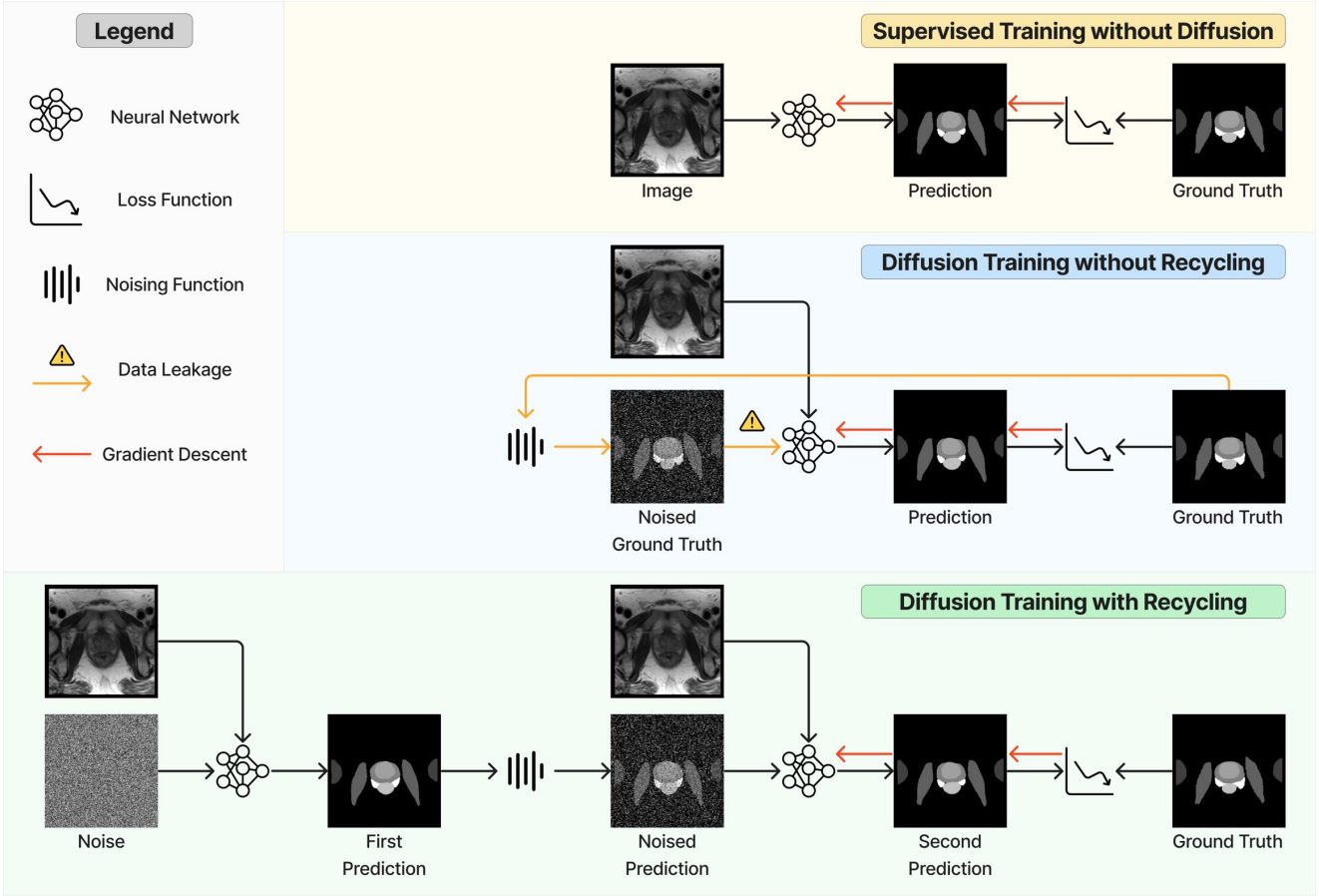
How to solve it?
Performs an extra inference step to use pre-segmentation to replace ground truth.

Does it work?
Outperforms self-condition on muscle ultrasound, abdominal CT, prostate MR, and brain MR.

Show me your code?
JAX based code is available.

Any catch?
Recycling needs extra compute.
Diffusion UNet did not surpass supervised UNet.

Future work?
Latent diffusion. Discrete diffusion. Mean-teacher. Diffusion-specific network.



ImgX-DiffSeg Public

A JAX-based deep learning framework for image segmentation using diffusion models.

Python 72 8

Acknowledgements

Supervisors

Yipeng Hu
Matthew J Clarkson

Coauthors

Yiwen Li
Shaheer U Saeed

Reviewers & Editors

MELBA Journal

MICCAI 2023 Vancouver CANADA

Funding

CANDI Computer Assisted Navigation, Diagnosis and Intervention

wellcome EPSRC centre interventional+ surgical sciences

cmic Centre for Medical Image Computing

Google's TPU Research Cloud (TRC)

Collaboration / Contact

yunguan.fu.18@ucl.ac.uk

@mathpluscode