Ambiguous Medical Image Segmentation Using Diffusion Models

Aimon Rahman[1] Jeya Maria Jose Valanarasu[1] Ilker Hacihaliloglu[2] Vishal M. Patel[1]
[1]Johns Hopkins University [2]University of British Columbia
arahma30@jhu.edu

CVPR 2023

Motivation: Instead of Single Diagnosis

- Clinical Practice:

  Collective insights from a group of experts have always proven to outperform an individual's best diagnostic for clinical tasks.

- Current Medical Image Segmentation Models:

Single Expert

Deterministic

$$mask = f(image)$$

- one mask for one image

- This paper

Group of Experts

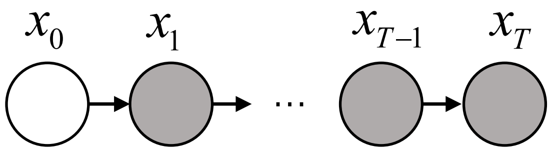
(Conditional) Generative

$$\mu,\sigma = f(image)$$
$$p(mask) = N(\mu,\sigma)$$

- multiple masks for one image
- one mask one expert

Diffusion Models

- Nonequilibrium Thermodynamics [Sohl-Dickstein et al. ICML'15]



diffusion kernel:

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

Objective:

$$\boxed{\begin{aligned} &\max \quad E_{x_0 \sim q} \log p(x_0;\, \boldsymbol{\theta}) \\ &\geq E_{x_{0:T} \sim q} D_{KL}[q(x_{t-1} \mid x_t, x_0) \,||\, p(x_{t-1} \mid x_t;\, \boldsymbol{\theta})] \end{aligned}}$$

forward posterior:

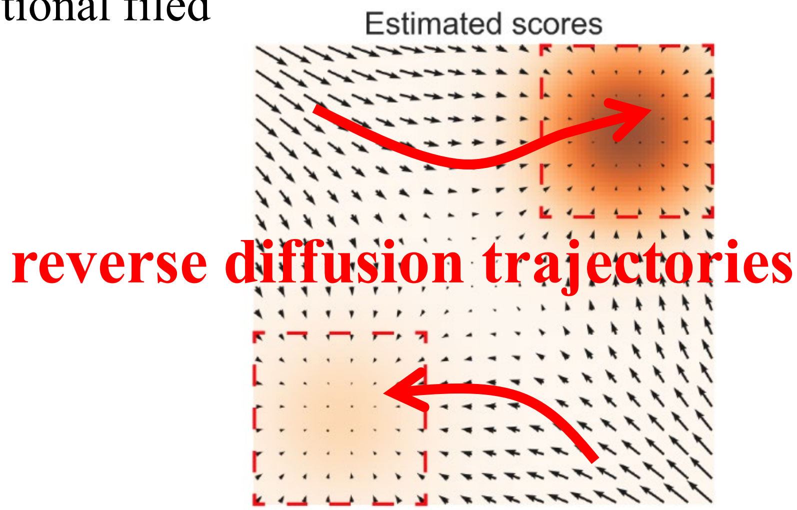
$$q(x_{t-1} \mid x_t, x_0) = N(x_{t-1} \mid \tilde{\mu}(x_t, x_0, \beta_t), \tilde{\beta}_t I)$$

information transfer →

parameterized model:

$$p(x_{t-1} \mid x_t;\, \boldsymbol{\theta}) = N(x_{t-1} \mid \tilde{\mu}(x_t, t;\, \boldsymbol{\theta}), \tilde{\beta}_t)$$

- Same dimension between latent variable and output variable
  - directional filed



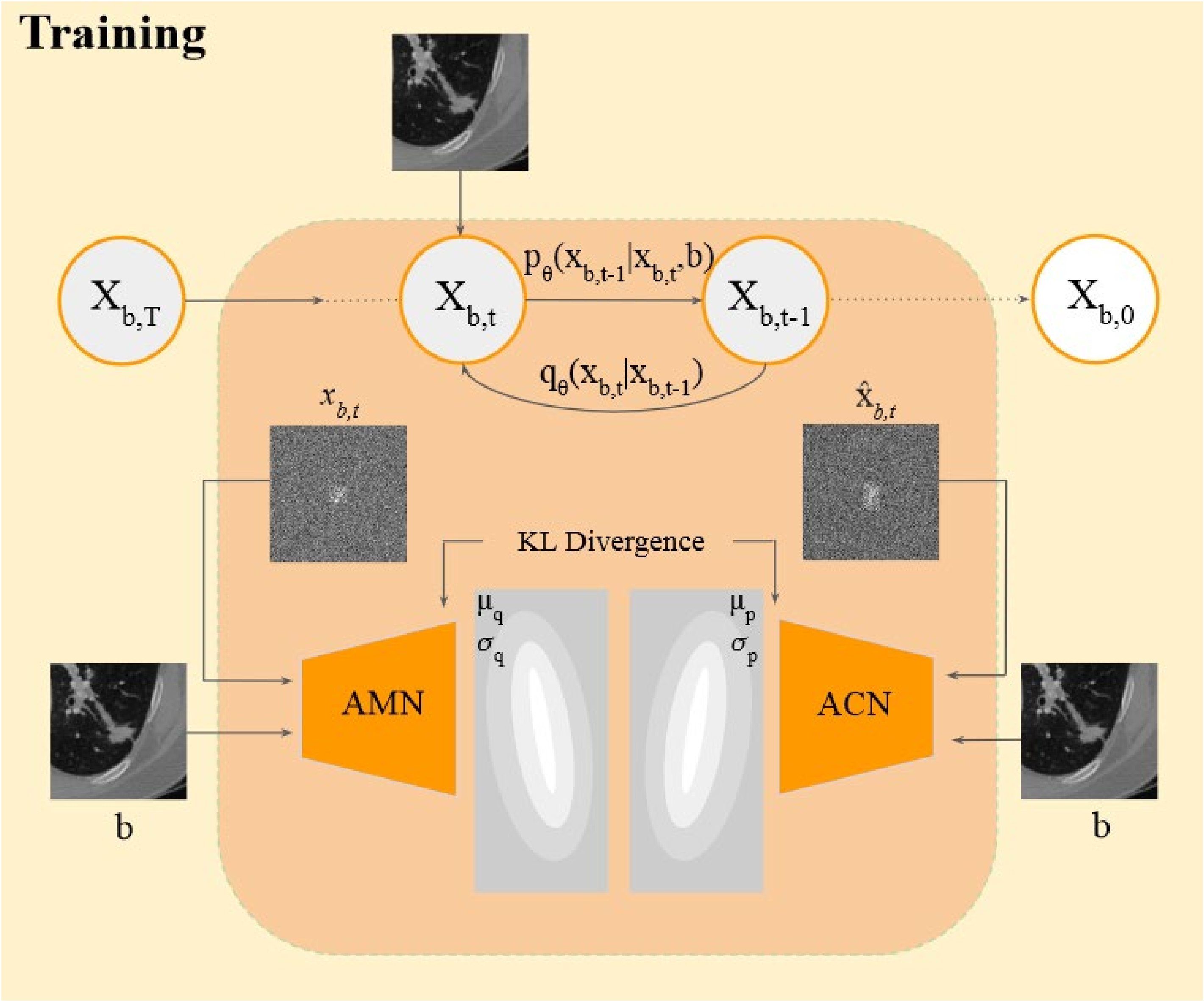
Algorithm 2 Sampling

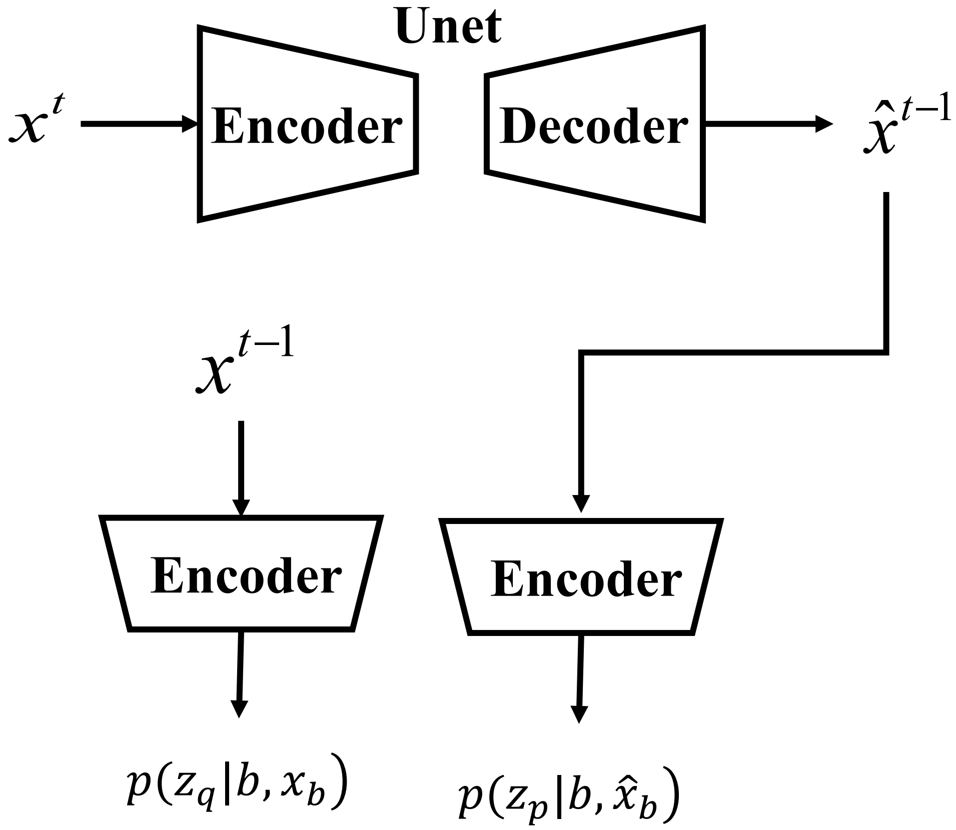
1: $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
2: **for** $t = T, \ldots, 1$ **do**
3: $\quad \mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ if $t > 1$, else $\mathbf{z} = \mathbf{0}$
4: $\quad \mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}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

Langevin Dynamics

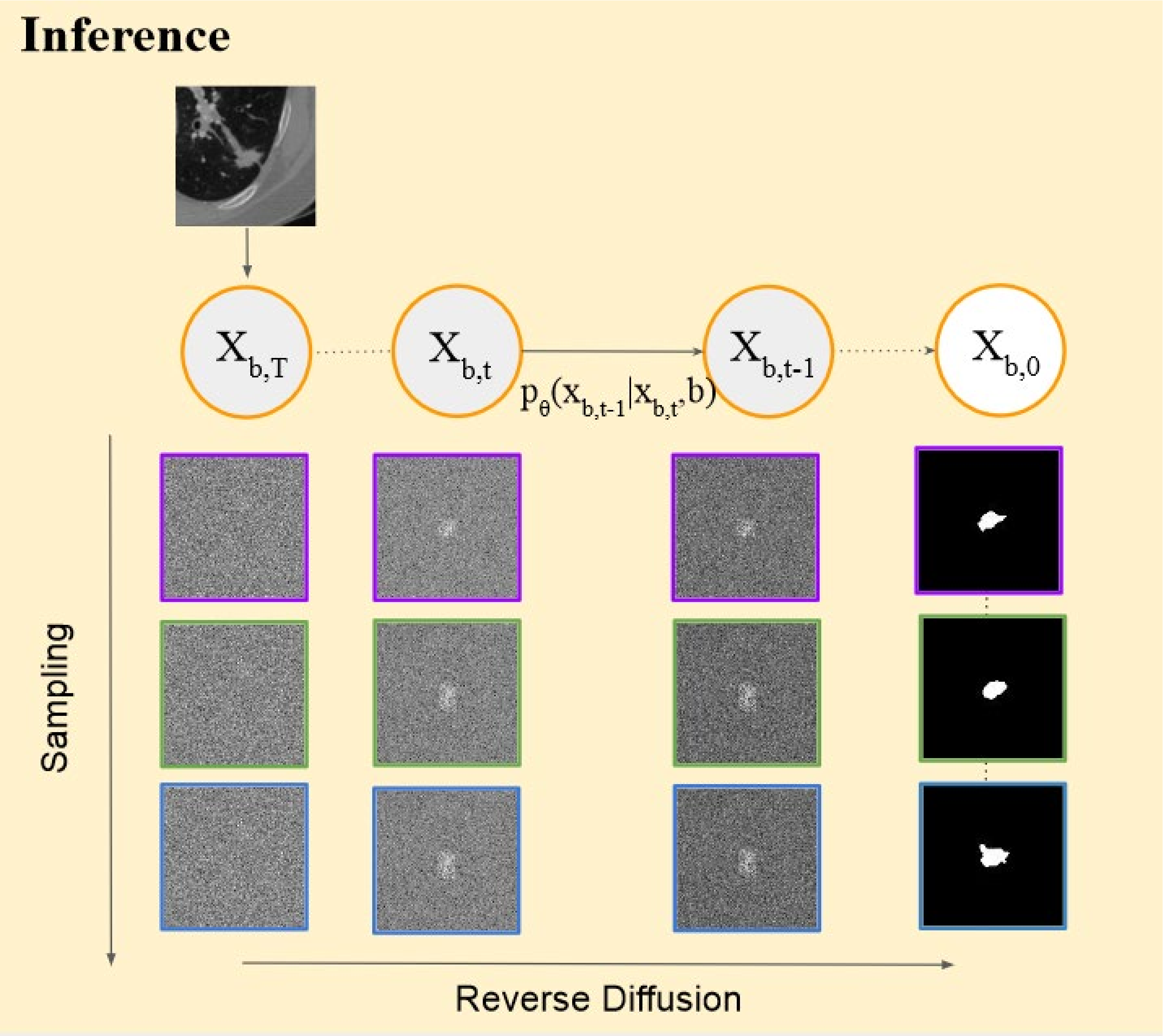
[Song et al. NIPS'19]

Collectively Intelligent Medical Diffusion (CIMD)





Collectively Intelligent Medical Diffusion (CIMD)



Datasets

- Lung Lesion Segmentation
    - 15096 CT images from 1010 subjects
    - 4 experts
- Bone Surface Segmentation
    - 1980 bone Ultrasound scans from 30 subjects
    - 4 annotations
- Multiple Sclerosis Lesion Segmentation
    - 7423 MRI scans from 5 subjects
    - 2 experts

New Metric for Ambiguous Segmentation

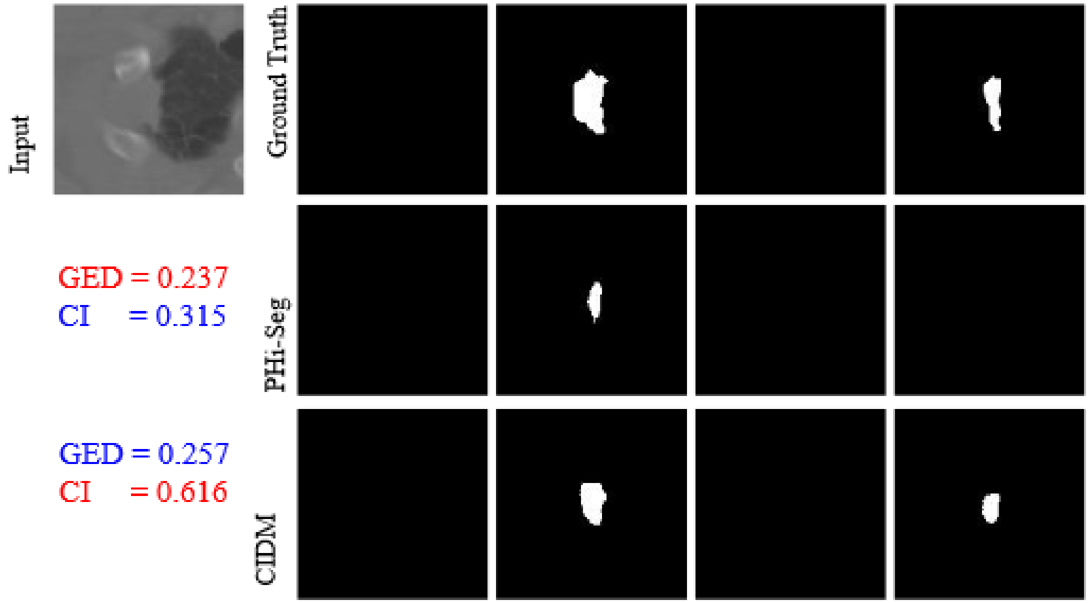
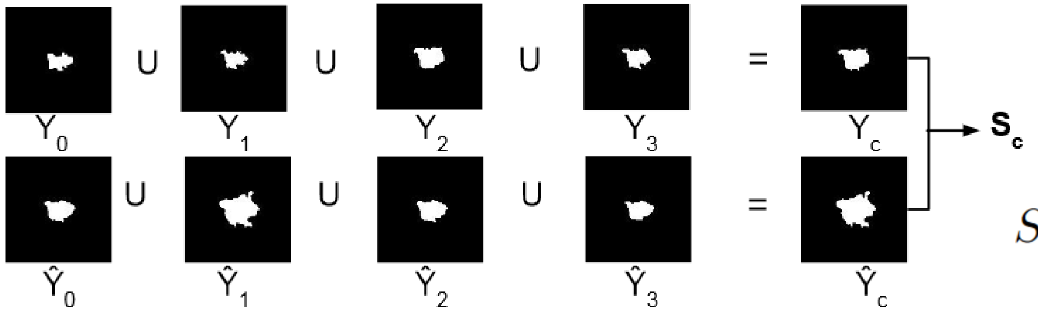


Figure 4. Visual analysis of the GED vs. the CI score for the LIDC-IDRI lung CT dataset. It can be observed that GED is lower for PHi-Seg even though it failed to segment most of the lesions. However, the combined sensitivity penalizes under segmentation hence the CI score is lower in that case. Red corresponds to better and blue corresponds to a lower score.

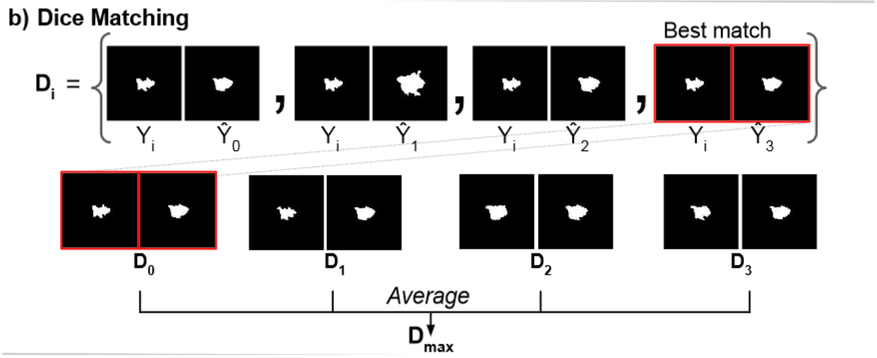
New Metric for Ambiguous Segmentation

- Combined Sensitivity



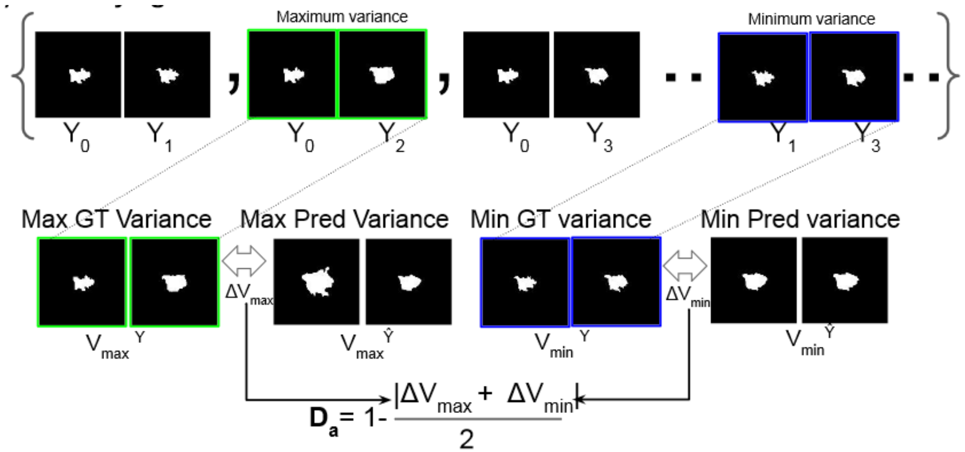
$$S_c(\hat{Y}_c, Y_c) = \begin{cases} \frac{TP}{TP+FN}, & \text{if } \hat{Y}_c \cup Y_c \neq \emptyset \\ 1, & \text{if otherwise.} \end{cases}$$

- Dice Matching



$$Dice(\hat{Y}, Y) = \begin{cases} \frac{2|Y \cap \hat{Y}|}{|Y|+|\hat{Y}|}, & \text{if } Y \cup \hat{Y} \neq \emptyset \\ 1, & \text{otherwise.} \end{cases}$$

- Diversity Agreement



$$CI = \frac{3 \times S_c \times D_{max} \times D_a}{S_c + D_{max} + D_a},$$

Qualitative Results on Lung Lesion Segmentation

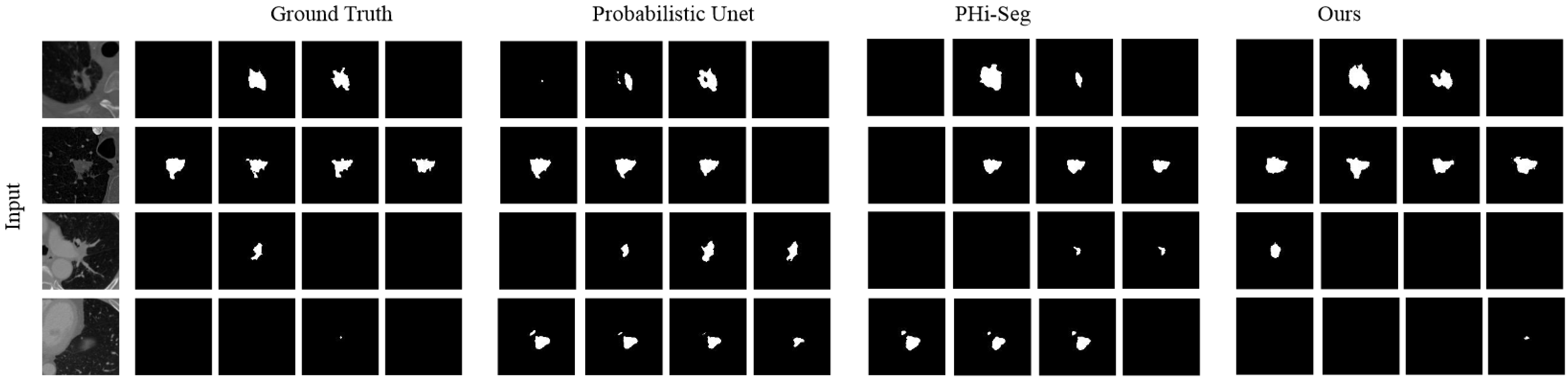


Figure 3. Comparative qualitative analysis with the two baseline methods – Probabilistic U-net [29] and PHi-Seg [8]. Sample images from the LIDC-IDRI dataset with 4 available expert gradings are shown on the left. Note that empty segmentation masks are also valid grading. For a fair comparison, we visualize only the first 4 sampled segmentation masks from the segmentation networks.

Qualitative Results on Bone Surface Segmentation

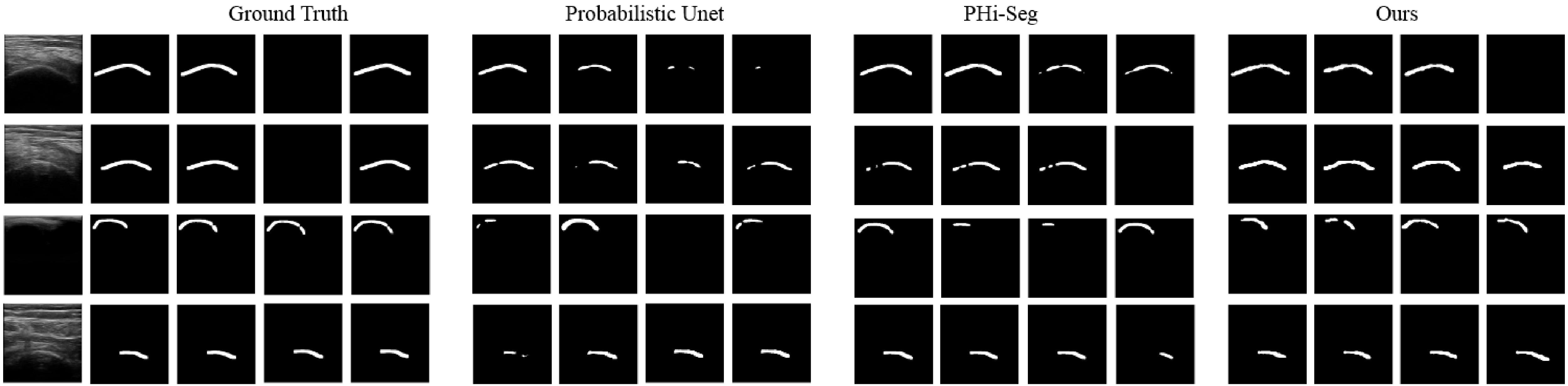


Figure 5. Comparative qualitative analysis with the two baseline methods Probabilistic U-net [29] and PHi-Seg [8]. Examples of the Bone-US dataset with 1 expert and 3 novice gradings are shown on the left. The bone-US dataset has a comparatively small inter-rater disagreement. We sample the first 4 segmentation masks from prediction distribution.

Qualitative Results on Sclerosis Lesion Segmentation

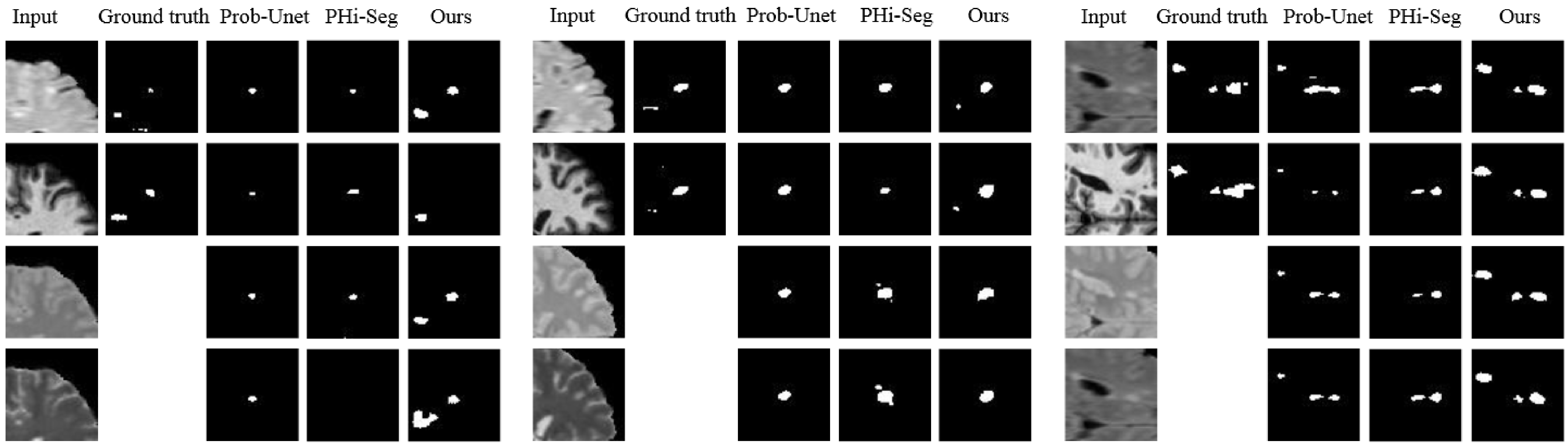


Figure 6. Comparative qualitative analysis with the two baseline methods Probabilistic U-net [29] and PHi-Seg [8]. Examples of the MS-MRI dataset with 2 expert gradings are shown here. We sample the first 4 segmentation masks from the prediction distribution.

Performance

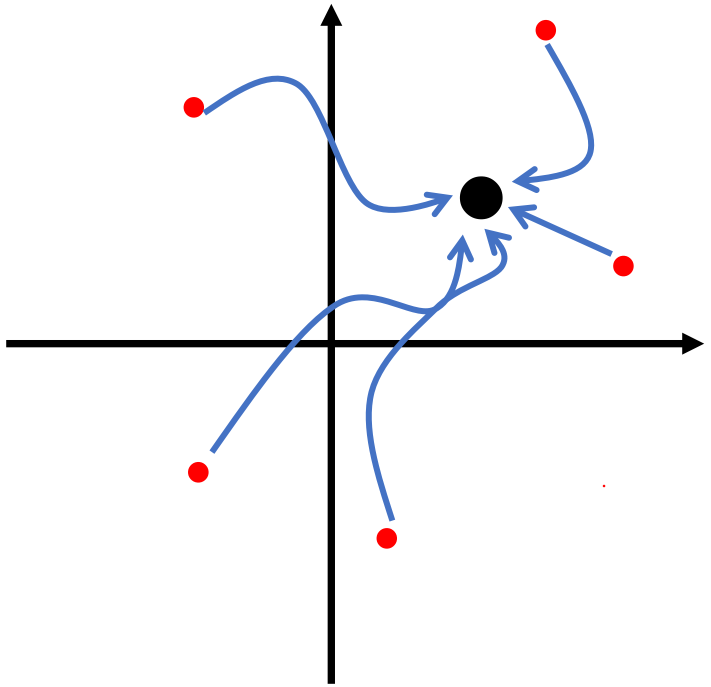
Table 1. Comparison of quantitative results in terms of GED, CI, and D_{max} for all the datasets with state-of-the-art ambiguous segmentation networks. The best results are in **Bold** and we achieve state-of-the-art results in terms of D_{max} and CI score across all datasets.

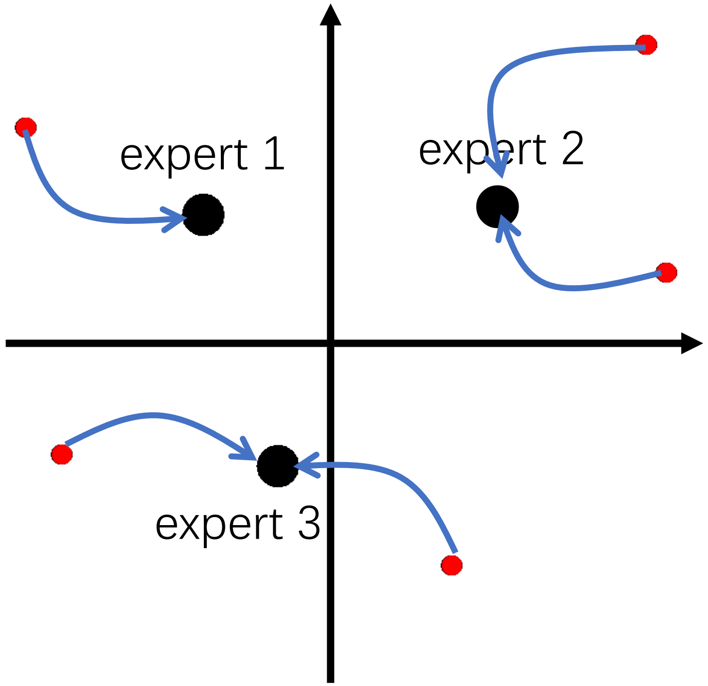
Method	LIDC-IDRI [4]			Bone Segmentation			MS-Lesion [12]		
	GED (↓)	CI(↑)	D_{max}(↑)	GED (↓)	CI(↑)	D_{max}(↑)	GED (↓)	CI(↑)	D_{max}(↑)
Probabilistic Unet [29]	0.353	0.731	0.892	0.390	0.738	0.844	0.749	0.514	0.502
PHi-Seg [8]	**0.270**	0.736	0.904	0.312	0.7544	0.848	0.681	0.518	0.506
Generalized Probabilistic U-net [10]	0.299	0.707	0.905	**0.289**	0.7501	0.863	**0.678**	0.522	0.513
$CIMD$ (Ours)	0.321	**0.759**	**0.915**	0.295	**0.7578**	**0.889**	0.733	**0.560**	**0.562**

Conclusions

Other papers
Medical Image Segmentation Using Diffusion Models

This paper
Ambiguous Medical Image Segmentation Using Diffusion Models





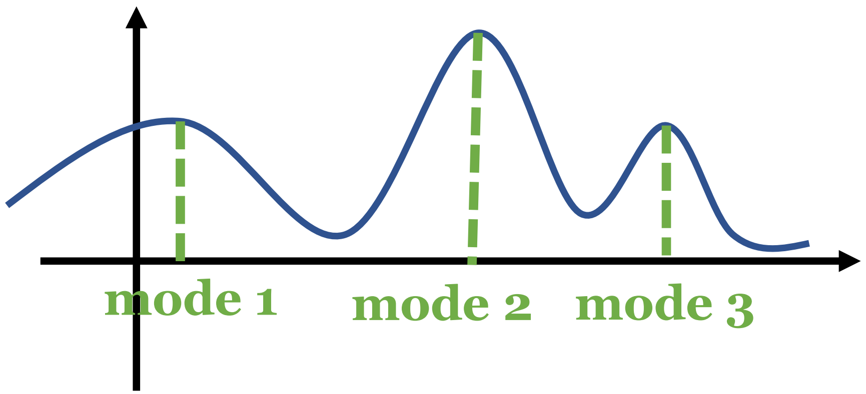
Sampled Gaussian points

Ground Truth

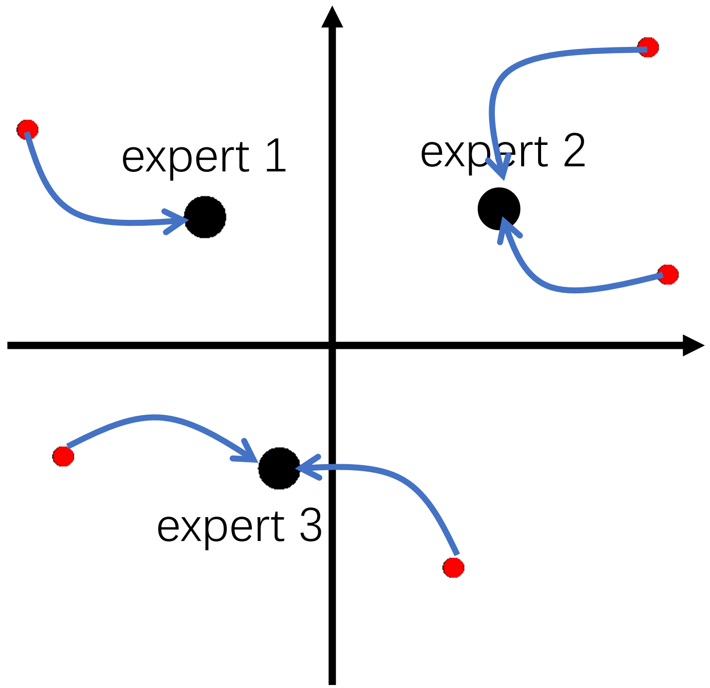
The directional field learned by diffusion model

Insights

- One mask for one medical image ?



This paper
Ambiguous Medical Image Segmentation Using Diffusion Models



- Diffusion is able to learn distributions of multiple modes