JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Ambiguous Medical Image Segmentation using Diffusion Models

Aimon Rahman[1] , *Jeya Maria Jose Valanarasu*[1] , *Ilker Hacihaliloglu*[2] , *Vishal M Patel*[1]

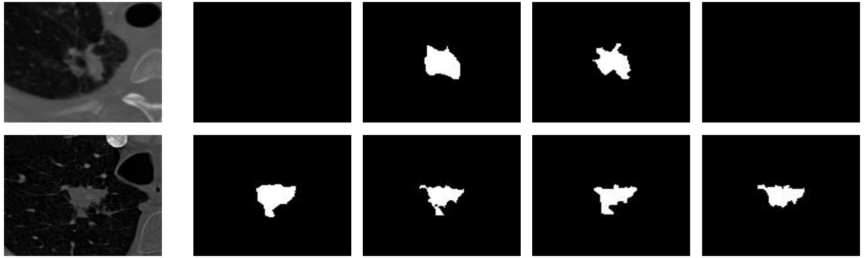
Johns Hopkins University[1] , University of British Columbia[2]

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

Background

Deterministic

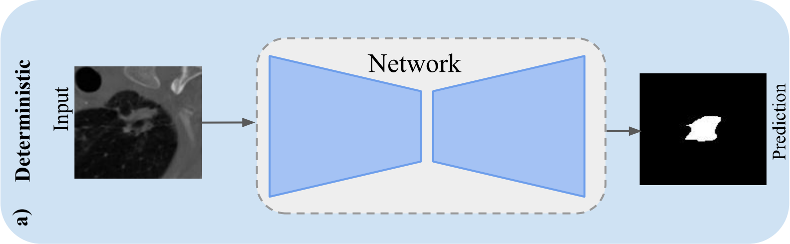
Probabilistic

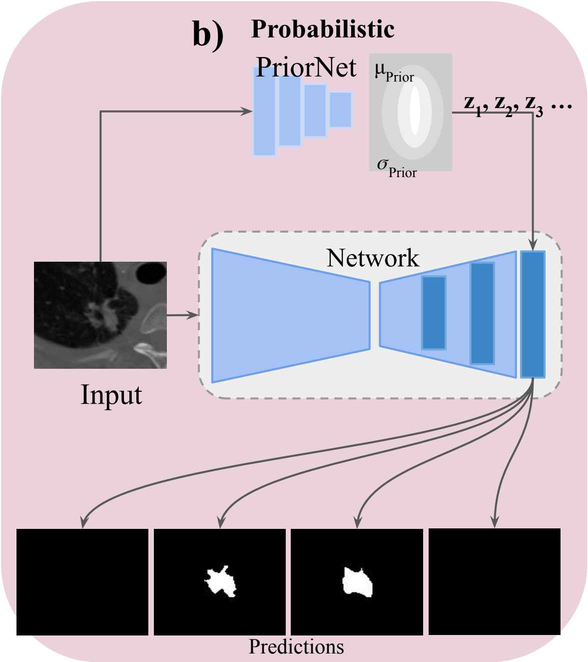


Unlike natural images, ground truths are not deterministic in medical images as different diagnosticians can have different opinions on the type and extent of an anomaly.

Background

Existing Probabilistic Networks

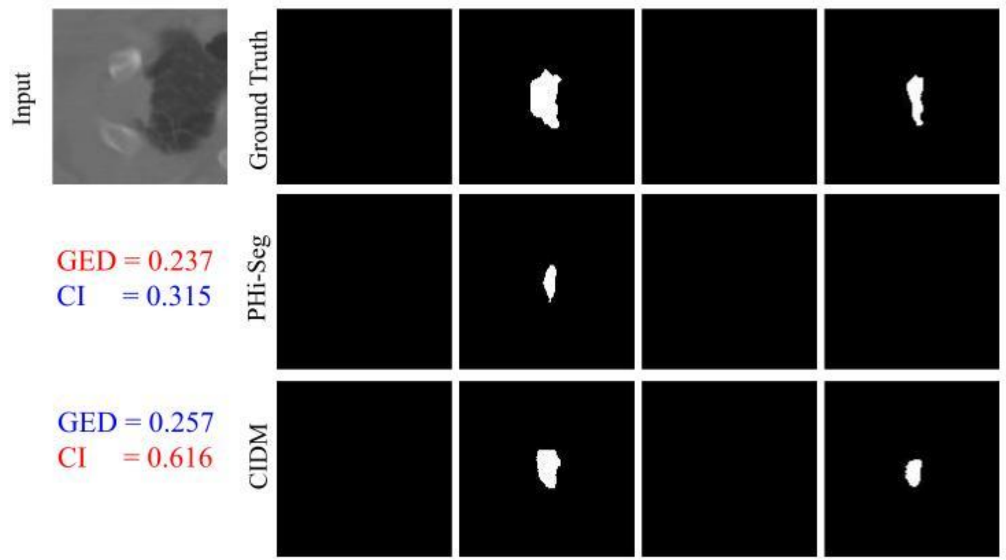




c-VAE-based methods incorporate prior information about the input image in a separate network and sample latent variables to produce stochastic segmentation masks

Background

Problem with Evaluation Metrics (GED)

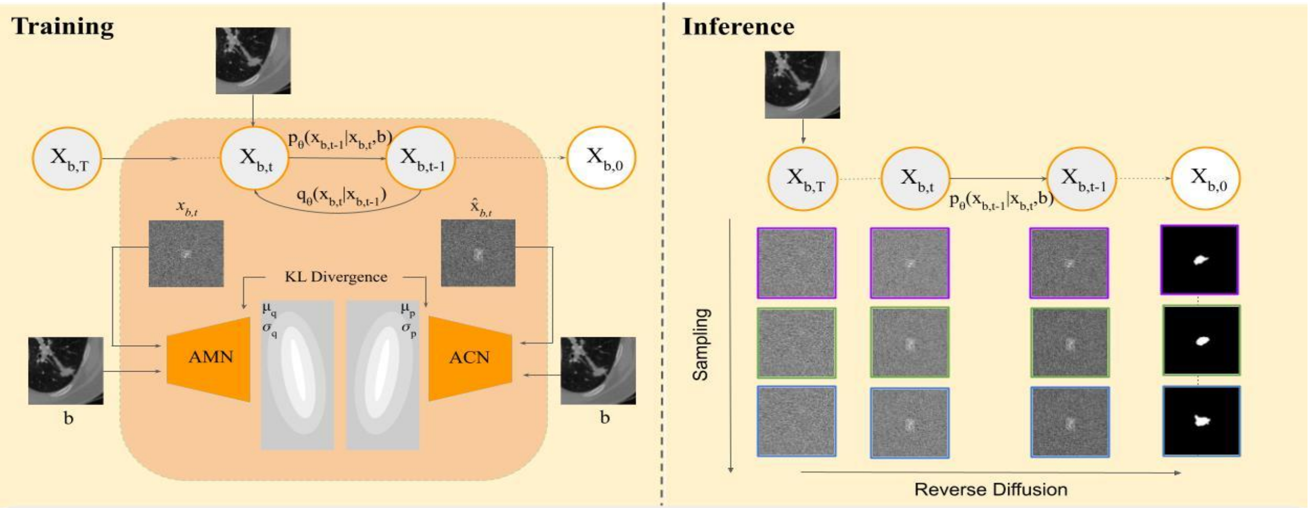


Generalized Energy Distance (GED) metric ***overly rewards sample diversity*** *without considering the match with ground truth samples*

Contribution

- We introduce a **diffusion model-based approach** that generates multiple plausible segmentation masks by learning a distribution over group insights.
- The proposed model utilizes diffusion's **stochastic sampling** process to produce diverse segmentation variants with minimal additional learning.
- The model's effectiveness is demonstrated on **CT, ultrasound, and MRI** images, outperforming existing state-of-the-art methods in accuracy and preserving natural variation.
- A **new metric** is proposed to evaluate both segmentation diversity and accuracy, catering to the interests of clinical practice and collective insights.

Method



b - Input image $x_{b,t}$ - Segmentation mask with added noise X_b - Segmentation masks of b $\hat{x}_{b,t}$ - Output from single time step

AMN - Ambiguity Modeling Network ACN - Ambiguity Controlling Network

CI Score - Intuition

- The proposed CI score (Collective Insight) addresses the limitations of GED and consists of three components: **Combined Sensitivity, Maximum Dice Matching, and Diversity Agreement.**

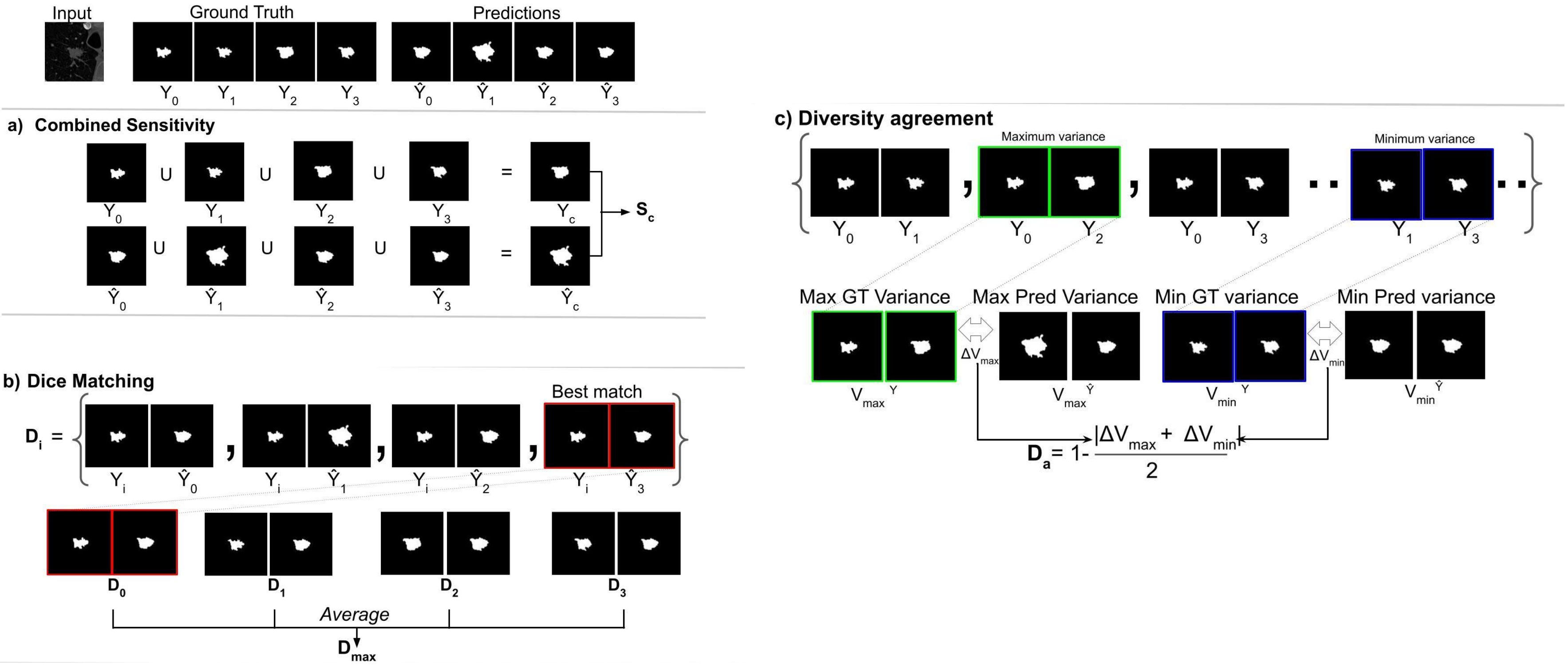
- Combined Sensitivity measures the true positive rate of the combined predictions and ground truths, aligning with clinical practice objectives.

- Maximum Dice Matching calculates the maximum Dice score between individual predictions and all ground truths, representing the comparison of student diagnoses with expert opinions.

- Diversity Agreement assesses the diversity of predicted outputs by comparing the variance between ground truth and prediction distributions.

- The CI score is defined as **the harmonic mean of the combined sensitivity, maximum Dice matching, and diversity agreement** components.

CI Score - Mechanism

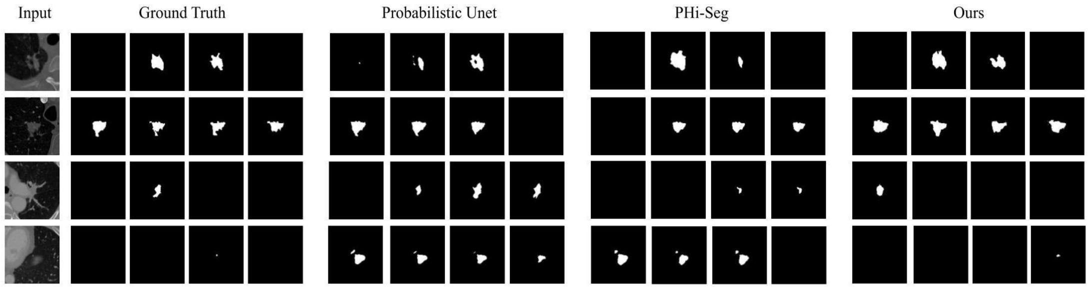


Results

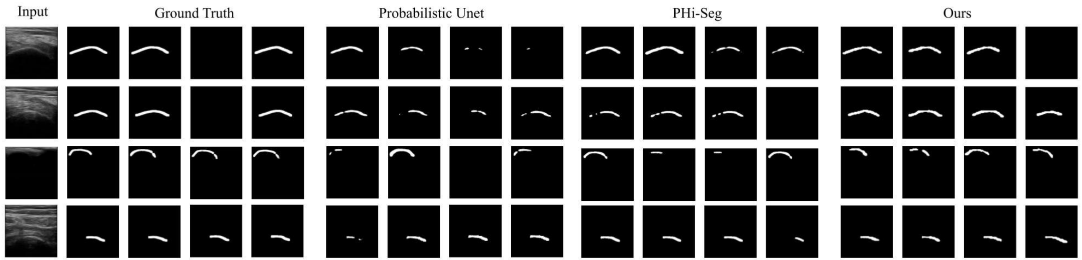
Table 1. Comparison of quantitative results in terms of GED, CI, and D_{max} for all the datasets with state-of-the-art ambiguous segmentation networks. The best results are in **Bold** and we achieve state-of-the-art results in terms of D_{max} and CI score across all datasets.

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**

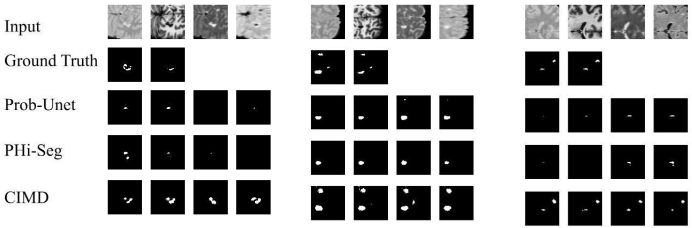
Results



Input
Ground Truth
Probabilistic Unet
PHi-Seg
Ours



Input
Ground Truth
Probabilistic Unet
PHi-Seg
Ours

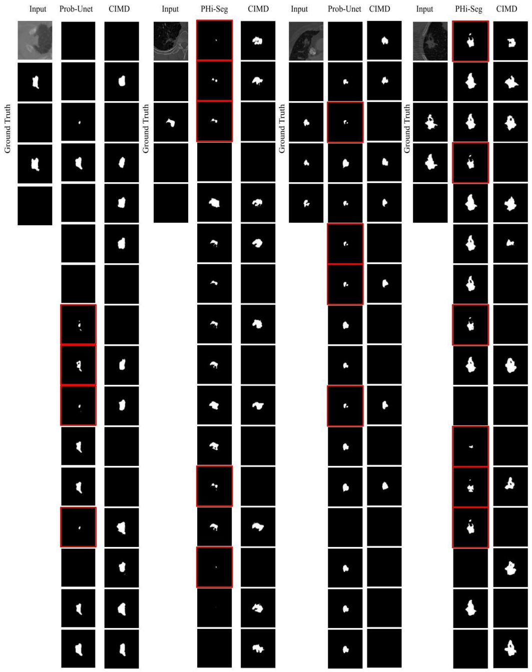


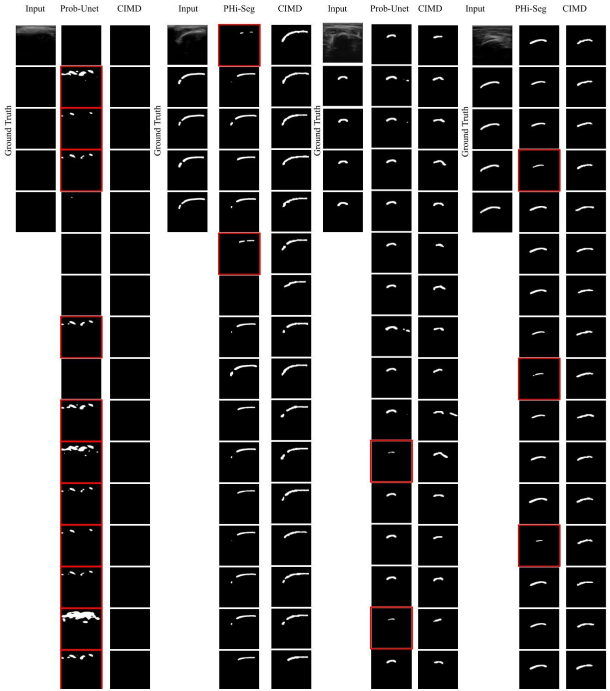
Input
Ground Truth
Prob-Unet
PHi-Seg
CIMD

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Results





JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Thank You