Modeling Multimodal Aleatoric Uncertainty in Segmentation with Mixture of Stochastic Experts

Zhitong Gao, Yucong Chen, Chuyu Zhang, Xuming He
ShanghaiTech University

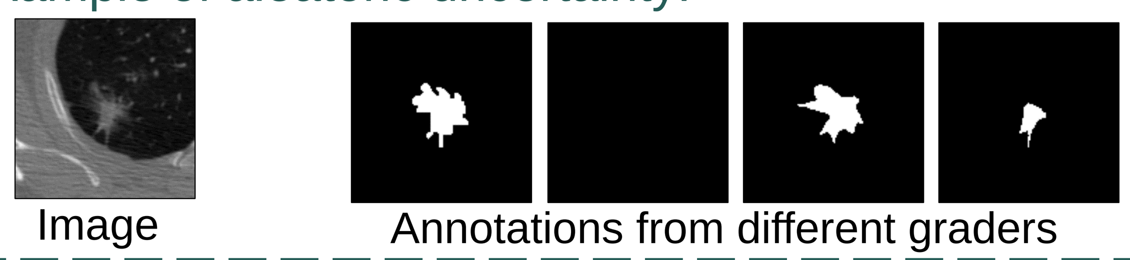
Paper

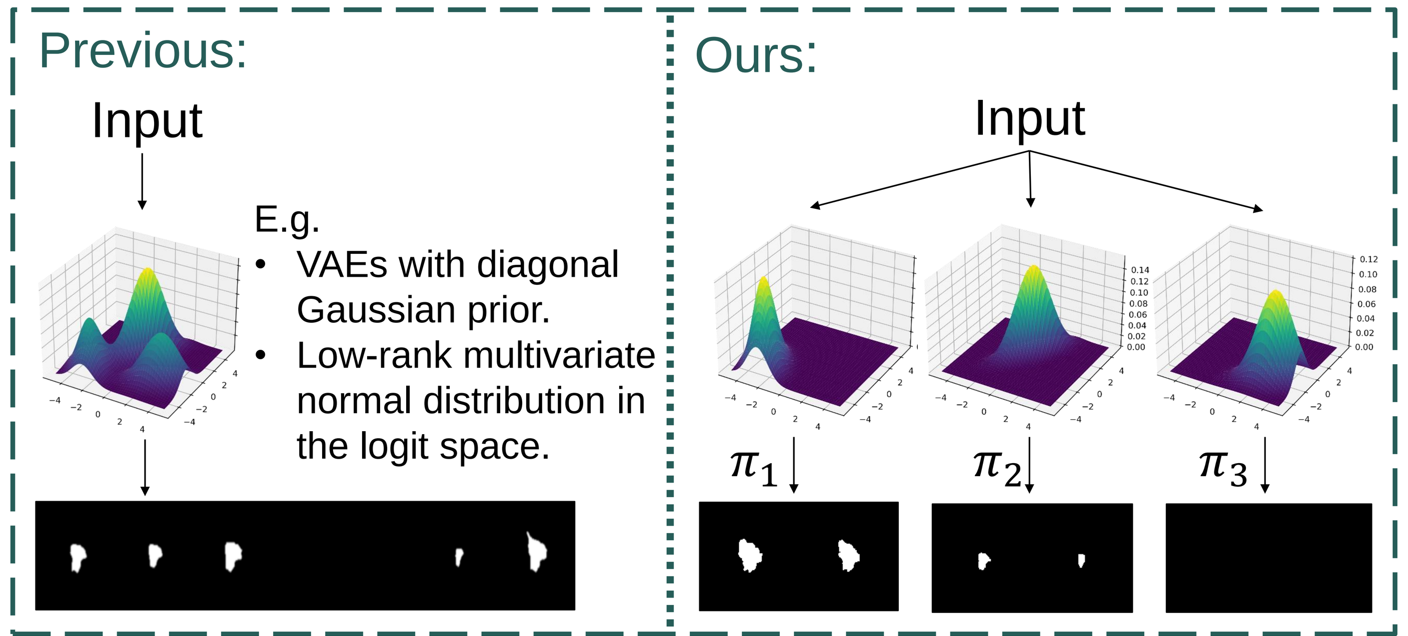
Code

INTRODUCTION

- **Problem:** Images are often ambiguous, leading to multiple plausible ground truth segmentation results.
- **Goal:** We aim to capture this data-inherent uncertainty (aka Aleatoric uncertainty) by learning the latent segmentation distribution.
- **Motivation:** The segmentation distribution is typically multi-modal. However, most previous methods have restricted capacity in capturing multi-modality, and rely on inefficient sampling to represent the predictive distribution.
- **Main idea:** We propose to explicitly model the multimodal characteristics of the distribution and provide a more efficient representation of the uncertainty.

An example of aleatoric uncertainty:





METHOD

- **Overview:** We introduce a conditional probabilistic model on segmentation $\mu_\theta(y|x)$, and design a distributional loss ℓ to represent and learn the aleatoric uncertainty.

$$\min_\theta \sum_n \ell(\mu_\theta(y|x_n), \nu_n)$$

where $\nu_n := \sum_{i=1}^{M} \mathrm{v}_n^{(i)} \delta(y_n^{(i)})$ is the empirical GT distribution ($M \geq 1$).

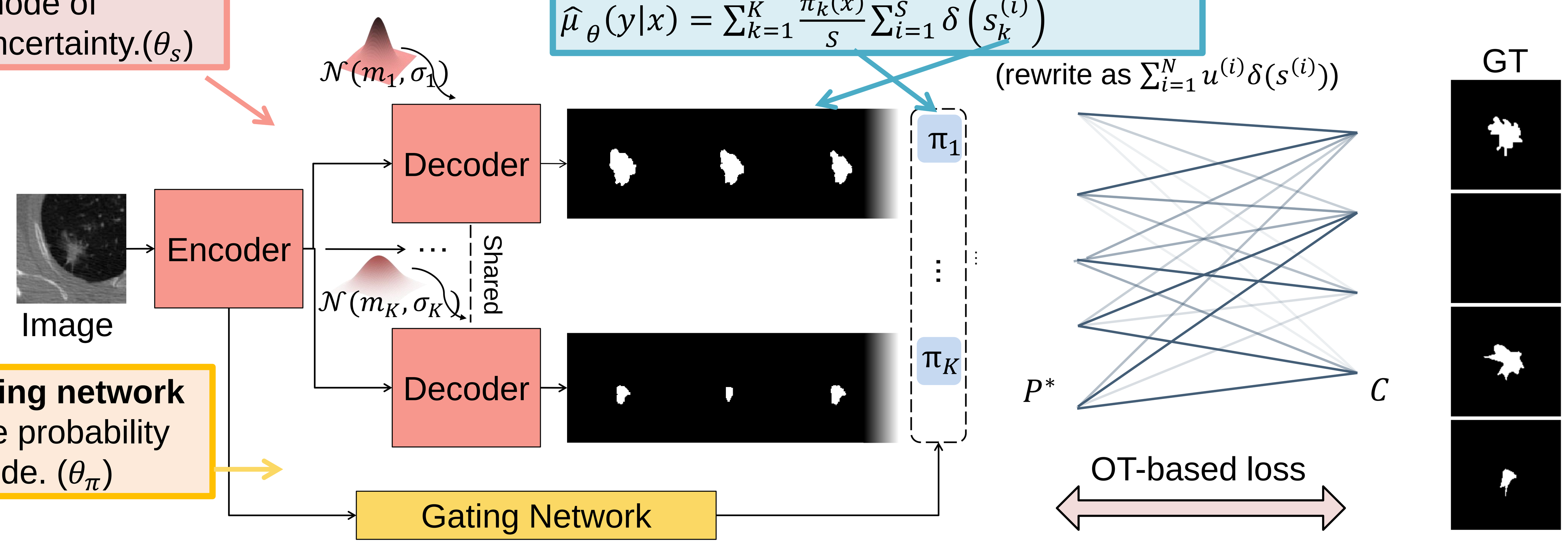
1. Mixture of Stochastic Experts (MoSE) Framework

A. Each expert encodes a distinct mode of aleatoric uncertainty.(θ_s)

B. The gating network predicts the probability of each mode. (θ_π)

C. Compact uncertainty representation
$\hat{\mu}_\theta(y|x) = \sum_{k=1}^{K} \frac{\pi_k(x)}{S} \sum_{i=1}^{S} \delta\left(s_k^{(i)}\right)$

(rewrite as $\sum_{i=1}^{N} u^{(i)} \delta(s^{(i)})$)



2. Optimal-Transport-Based (OT) Loss

- Formulate the model learning as an OT problem.

$$\min_{\theta_s, \theta_\pi} \sum_n \langle P^*, C_n(\theta_s) \rangle$$
$$P^* = \operatorname*{argmin}_{P \in R_+} \langle P, C_n \rangle \;\; s.t.\, P 1_M = u_n(\theta_\pi);\; P^T 1_N = \mathrm{v}_n$$

- Efficient bi-level optimization with constraint relaxation.

$$\min_{\theta_s, \theta_\pi} \sum_n \langle P^*, C_n(\theta_s) \rangle + \beta\, KL(P^* 1_M || u_n(\theta_\pi))$$
$$P^* = \operatorname*{argmin}_{P \in R_+} \langle P, C_n \rangle \;\; s.t.\, P 1_M \leq \gamma;\; P^T 1_N = \mathrm{v}_n$$

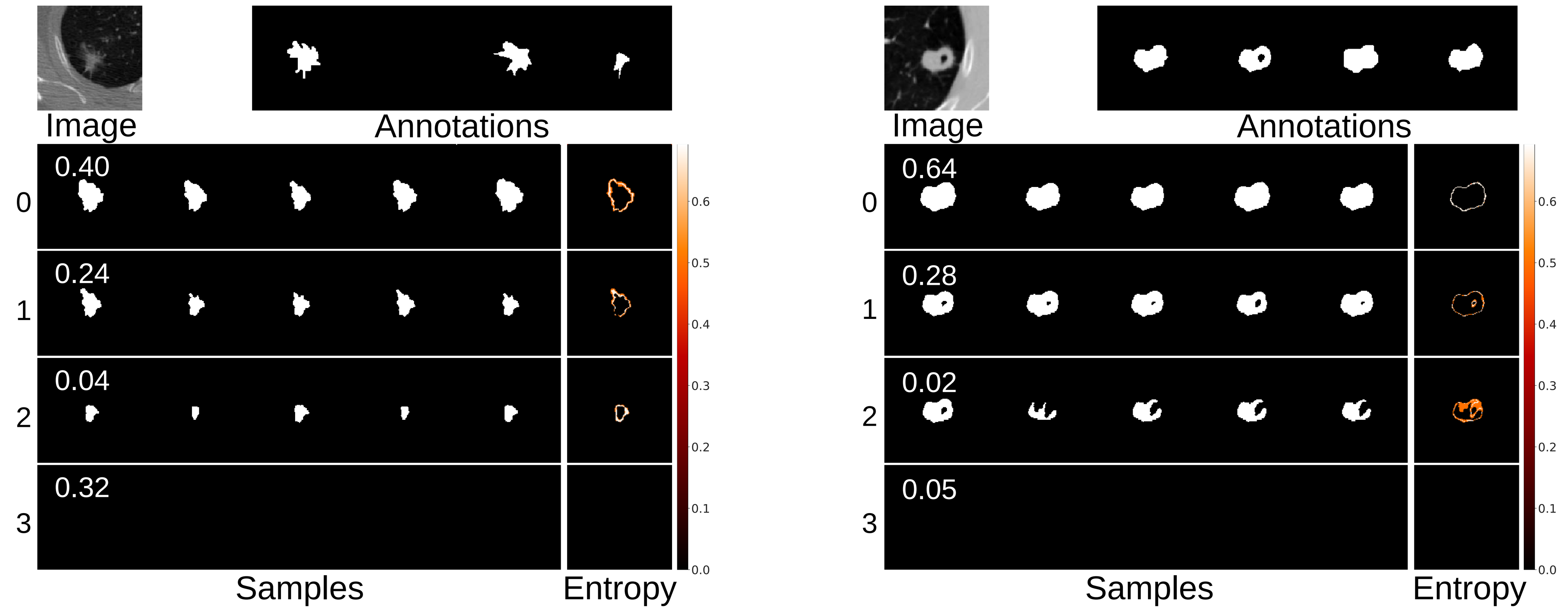
Annealing to 1.

RESULTS

1. Results on the LIDC dataset

- Compare with previous SOTA models, (.) denotes number of sampled outputs.

Method	# label	GED ↓ (16)	GED ↓ (100)	M-IoU ↑ (16)	ECE ↓ (%) (16)	# param.
Kohl et al. (2018)	All	0.320 ± 0.030	0.239 ± N/A †	0.500 ± 0.030	-	76.15M
Kohl et al. (2019)		0.270 ± 0.010	-	0.530 ± 0.010	-	87.51M
Baumgartner et al. (2019)		-	0.224 ± N/A	-	-	74.82M
Monteiro et al. (2020)			0.225 ± 0.002	-	-	41.28M
Kassapis et al. (2021)		0.264 ± 0.002	0.243 ± 0.004	0.592 ± 0.005	0.214 *	175.36M
Ours		**0.218 ± 0.003**	**0.189 ± 0.002**	**0.624 ± 0.004**	**0.064 ± 0.015**	41.60M
Ours - compact		**0.195 ± 0.005**	**0.186 ± 0.002**	**0.635 ± 0.003**	**0.054 ± 0.015**	41.60M
Kohl et al. (2018)	One	-	0.445 ± N/A †	-	-	76.15M
Baumgartner et al. (2019)		-	0.323 ± N/A	-	-	74.82M
Monteiro et al. (2020)		-	0.365 ± 0.005	-	-	41.28M
Ours		**0.252 ± 0.004**	**0.223 ± 0.005**	**0.596 ± 0.003**	**0.105 ± 0.009**	41.60M
Ours - compact		**0.228 ± 0.004**	**0.220 ± 0.005**	**0.605 ± 0.003**	**0.090 ± 0.011**	41.60M



2. Ablation study

- Evaluate the impact of each component on the full-labeled LIDC dataset. (16 samples)

Expert type	Expert weights	loss	GED ↓	M-IoU ↑	ECE ↓ (%)
stochastic	learnable / uniform	IoU loss	0.533 ± 0.001	0.533 ± 0.001	0.277 ± 0.017
stochastic	uniform	OT-based	0.282 ± 0.002	0.545 ± 0.007	0.215 ± 0.006
deterministic	learnable	OT-based	0.246 ± 0.006	0.591 ± 0.001	0.142 ± 0.003
stochastic	learnable	OT-based	**0.218 ± 0.003**	**0.624 ± 0.004**	**0.064 ± 0.015**

3. Results on the synthetic multimodal Cityscapes dataset

- Constructed by randomly flipping five classes with certain probabilities (GT distribution known).
- Quantitatively, our model achieves the SOTA or comparable performance on three metrics. (Please refer to our paper for more detailed information.)