Weekly Study Report

Wang Ma

2025-04-01

Electrical, Computer, and Systems Engineering Department
Rensselaer Polytechnic Institute

Outline

1. (ICML 2023) Probabilistic Contrastive Learning Recovers the Correct Aleatoric Uncertainty of Ambiguous Inputs

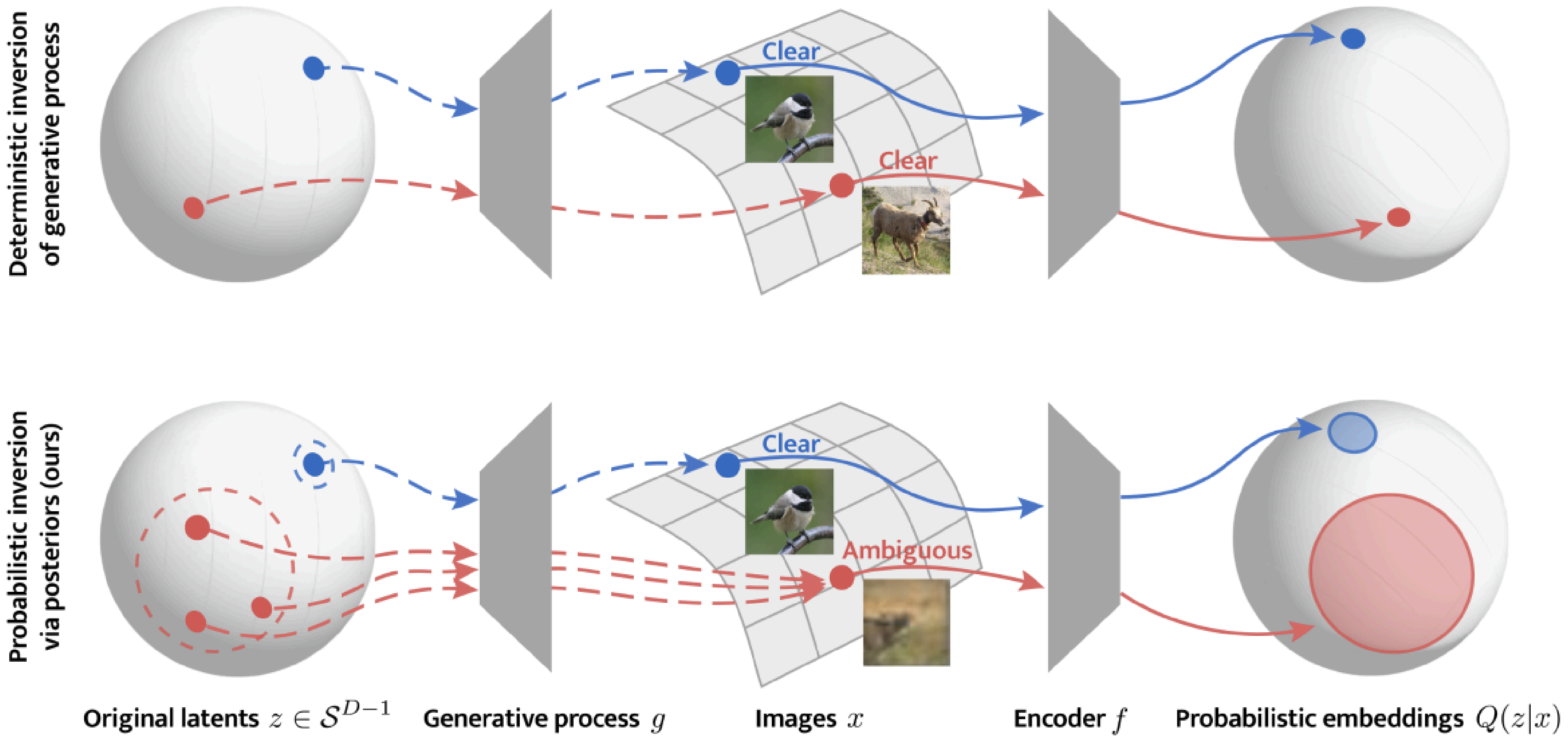


Figure 1. Deteministic encoders embed images to points in the latent space. This recovers the latent vectors that generated them (dashed), up to a rotation (top). However, if an image is ambiguous there are multiple possible latents that could have generated it (bottom). An encoder trained with MCInfoNCE correctly recovers this posterior of the generative process, up to a rotation, from contrastive supervision.

1.1 Probabilistic Embedding

Core Idea: When the input is ambiguous and blurred, the extracted feature is not longer a point, how to accurately recover the posterior distribution of the feature.

The posterior $p(z|x) \sim \text{vMF}\ (z; \mu(x), \kappa(x))$, where $\mu(x)$ represents the location/direction of the latent embedding in latent space, and $\kappa(x)$ represents the concentration.

The vMF distribution, *von Mises–Fisher Distribution*, can be seen as a Guassian Distribution which is defined on a Sphere. Since in the context of Contrastive Learning, researchers define the latent space on a sphere.

1.2 $\kappa(x)$, the concentration, is a realization of Aleatoric Uncertainty

Test on Rejecting: rejecting the images with low $\kappa(x)$(high AU), then perform the prediction. The accuracy improves.

But $\kappa(x)$ in definition is the inverse variance of the embedding, which is usually the total uncertainty in our setting.

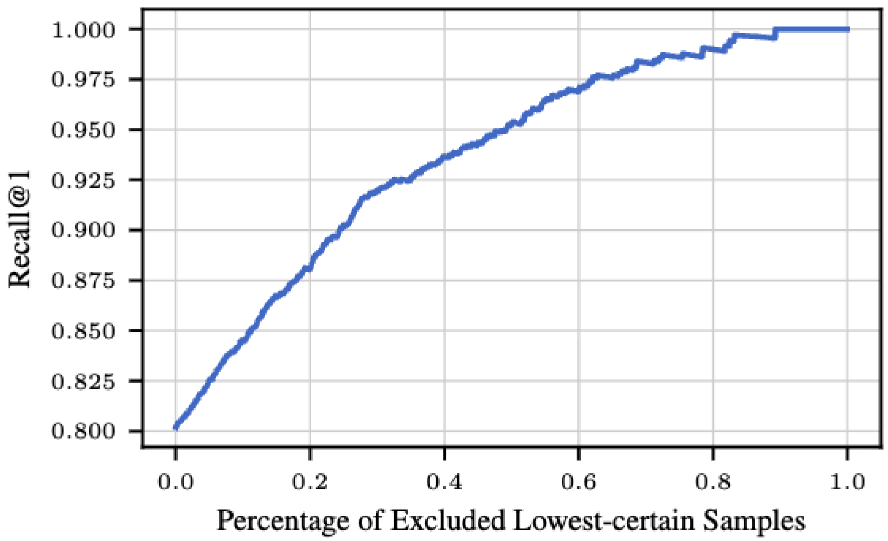


Figure 5. Rejecting images with low certainty values $\hat{\kappa}(x)$ improves the performance on the remaining data monotonically with the threshold. This shows that $\hat{\kappa}(x)$ is predictive of performance.

1.2 $\kappa(x)$, the concentration, is a realization of Aleatoric Uncertainty

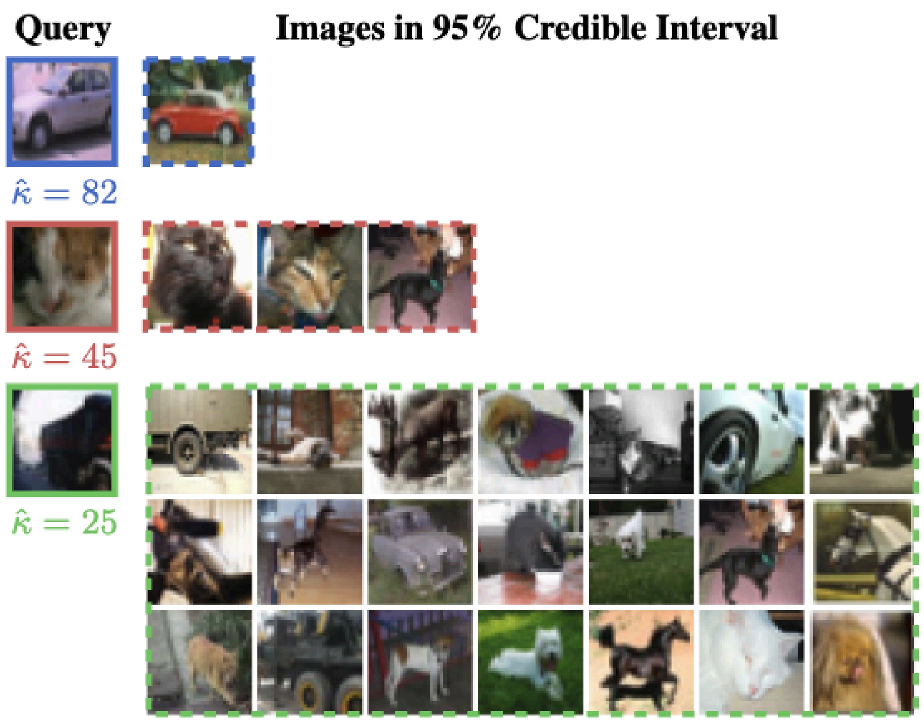


Figure 6. We use an image's posterior to define the credible interval that its latents lie in with a given probability. Clear query images (top) have small credible intervals containing images of the same class as the query. More ambiguous queries (bottom) return larger credible intervals with images from multiple possible classes.

1.3 An extension?

How about considering:

$$p(y|x,\theta) = \int_f p(y|f,\theta_2)p(f|x,\theta_1)df \approx \frac{1}{n}\sum_i^n p(y|f_i,\theta_2),$$

where f is the embedded feature, and

- $p(y|f_\theta,\theta_2)$ is the discriminative part which makes the prediction based on a given feature,
- $p(f|x,\theta_1)$ is the encoded feature distribution given the input image.

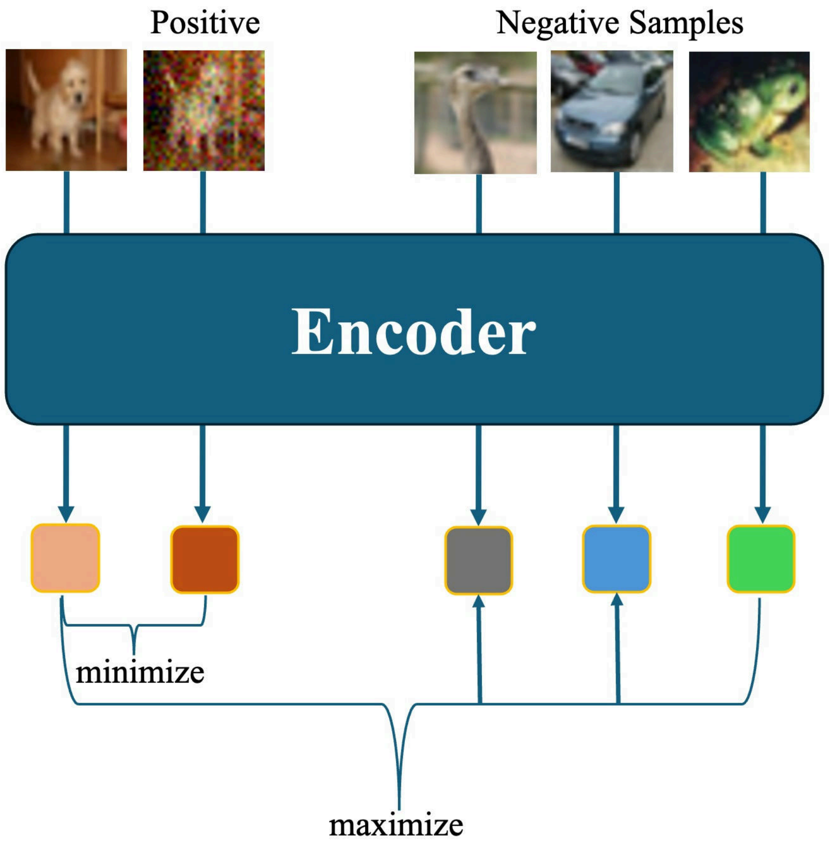
2. My Idea about Contrastive Learning

- Reference Sample: An Image X
- Positive Sample: Corrupted Image X_c
- Negative Samples: Other Images X_n

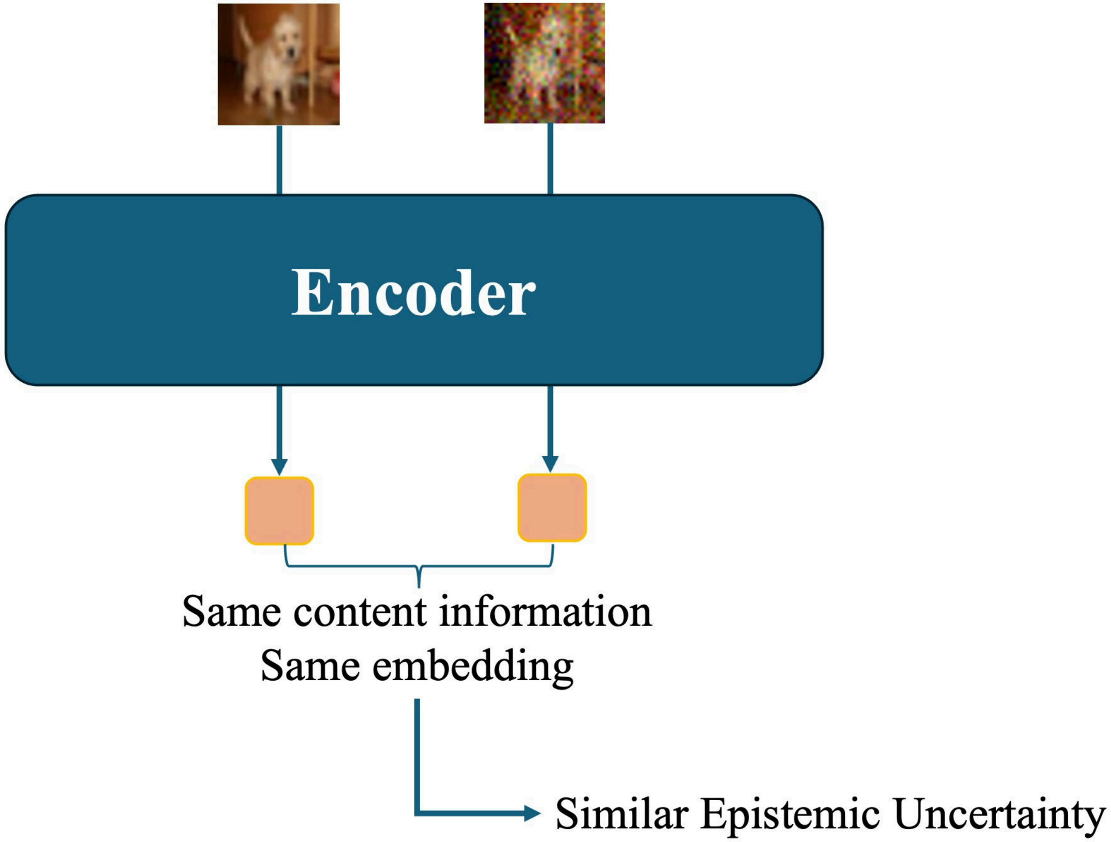
Then using contrastive learning methods to learn a robust latent embedding which is consistent in the clean image and corrupted image, and should be related to the "content".

This way, we may separate the content information and non-content information from the image.

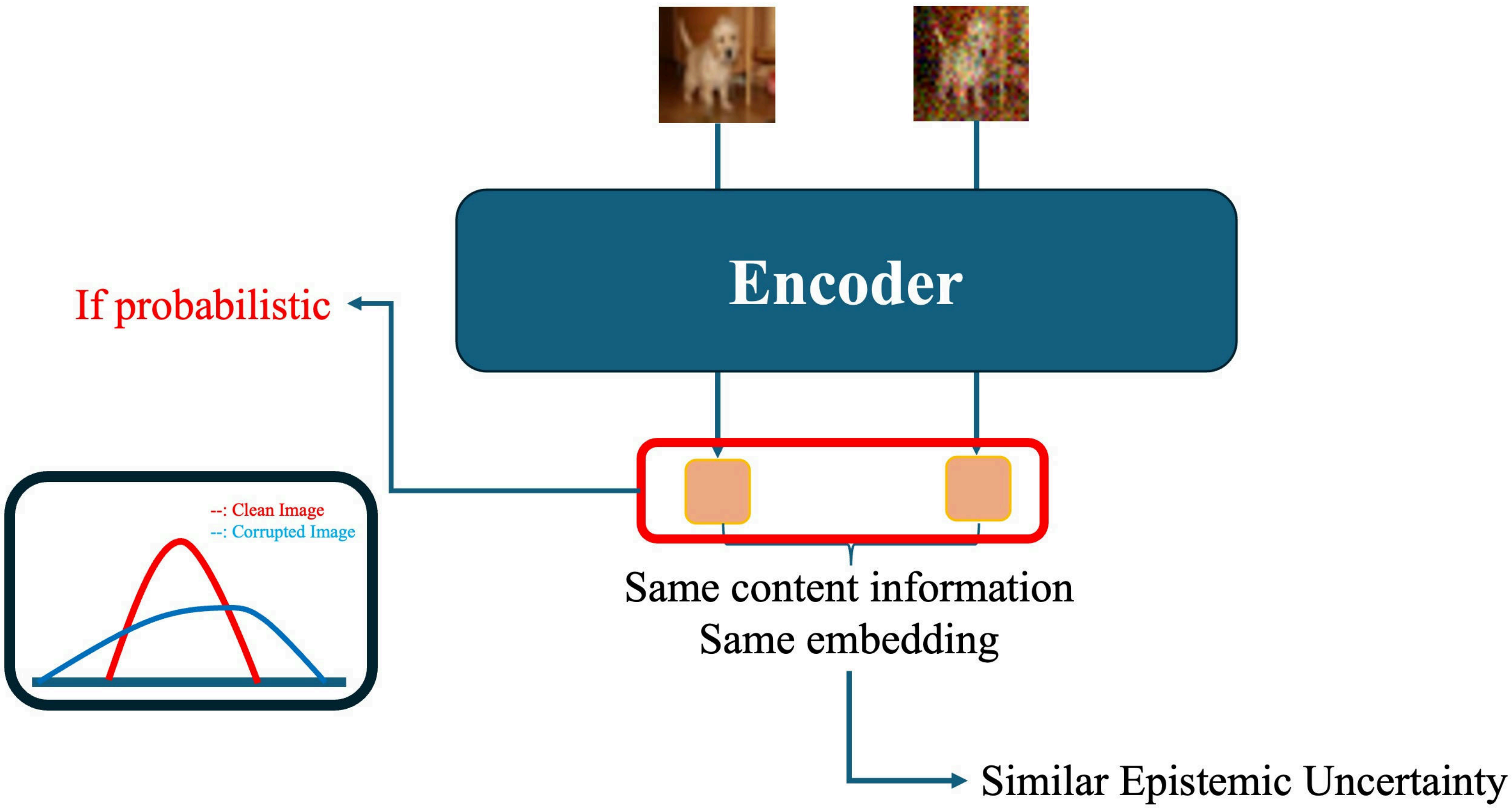
2. My Idea about Contrastive Learning



2. My Idea about Contrastive Learning



2. My Idea about Contrastive Learning



3. How disentangled are your classification uncertainties?

3.1 Experimental Design for Disentangled Uncertainty

3.1.1 Experiments

1. Changing the Dataset Size (Epistemic Uncertainty)
2. Testing on the Samples from an Unknown Class (Epistemic Uncertainty)
3. Training and Testing on Datasets with Label Noise

3.1.2 UQ Methods Used

1. Information Theoretic Way

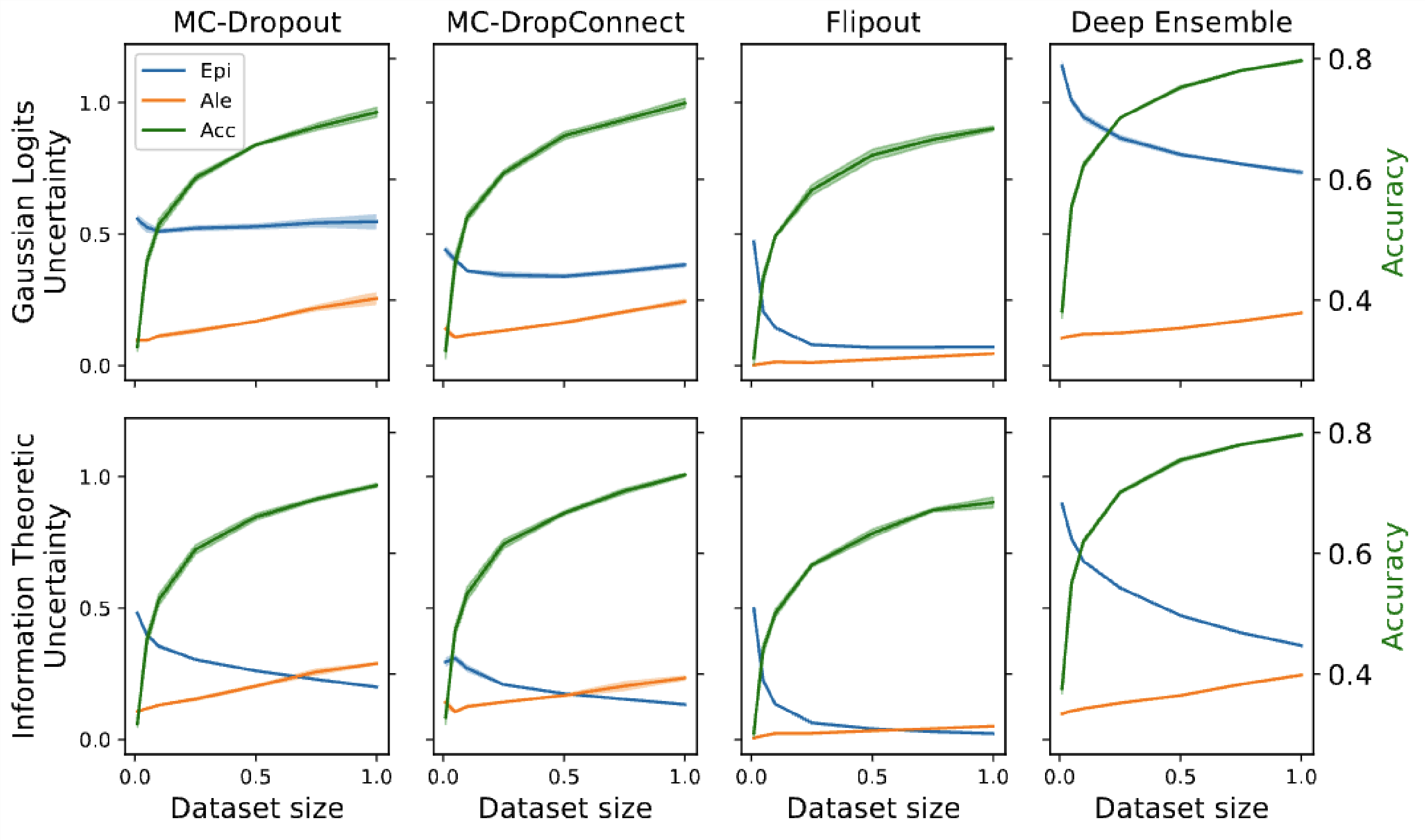
$$\mathbb{H}[y^*|x^*, D] = \mathbb{I}[y^*; \theta \mid x^*, D] + \mathbb{E}_{p(\theta|D)}\mathbb{H}[y^*|x^*, \theta]$$

1. Gaussian Logits Way

   Modeling the logits to follow Gaussian Distribution, then use the variance-based methods to quantify uncertainty.

$$\mathrm{Var}[y^*] = \mathrm{Var}[\mu(x)] + \mathbb{E}[\sigma^2(x)]$$

3.2.1 Dateset Size Experiment



3.2.2 Testing on the Samples from an Unknown Class (OOD Detection)

UQ Method	GL Ale	GL Epi	IT Ale	IT Epi
Dropout	0.644	0.642	**0.651**	0.649
DropConnect	0.650	0.657	0.657	**0.658**
Flipout	0.626	**0.629**	0.625	0.579
Deep Ens.	0.679	**0.709**	0.689	0.701

- GL = Gaussian Logits
- IT = Information Theoretic

3.2.3 Training and Testing on Dataset with Label Noise