Weekly Study Report

Wang Ma

2025-02-04

Electrical, Computer, and Systems Engineering Department
Rensselaer Polytechnic Institute

Outline

[UAI 2024] Shedding Light on Large Generative Networks: Estimating Epistemic Uncertainty in Diffusion Models

1. Epistemic Uncertainty Quantification in Diffusion Model

Ensemble Approach:

$$p(y|D, x) = \sum_{j=1}^{M} \pi_j p\left(y|x, \theta_j\right) \text{ , where} \sum_{j=1}^{M} \pi_j = 1$$

Standard Way.

$$I(y, \theta \mid D, x) = H(y|D, x) - E_{p(\theta)}[H(y|x, \theta)]$$

However, in the context of high-dimensional continuous outputs, the above equation is hard and expensive to compute. Here, the author relied on *Pairwise-Distance Estimators* (PaiDEs) [Kolchinsky and Tracey, 2017], which has been shown to accurately capture epistemic uncertainty for high-dimensional continuous outputs.

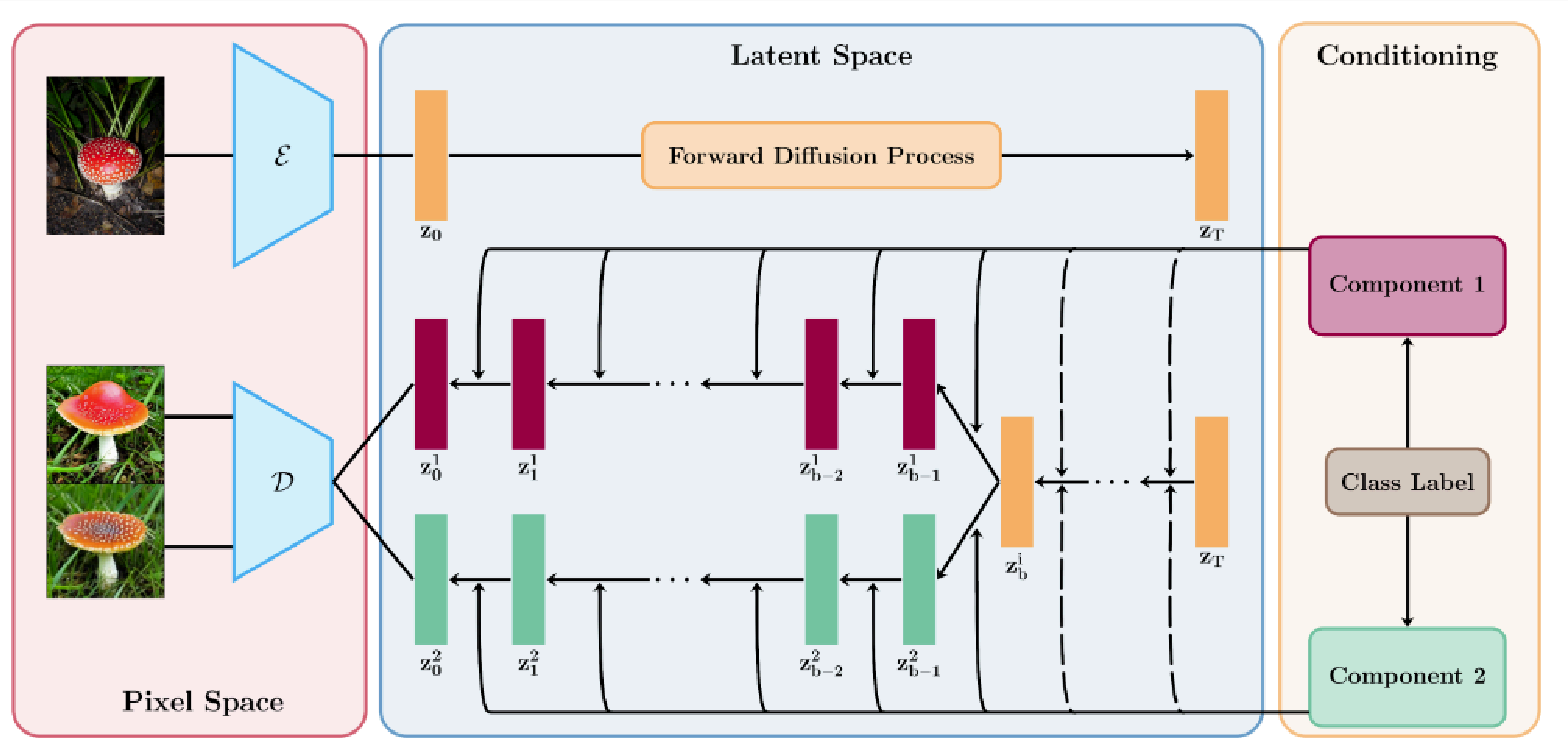
1.2 Pairwise-Distance Esitmators (PaiDEs)

Let $D(p_i \| p_j)$ denote a generalized distance function between two probability distributions, satisfying $D(p_i \| p_j) \geq 0$ and $D(p_i \| p_j) = 0 \ \text{ if } \ p_i = p_j$.

Then PaiDEs can be defined as follows:

$$\hat{I}_\rho(y_{t-1}, \theta \mid y_t, x) = -\sum_{i=1}^{M} \pi_i \ln\left(\sum_{j=1}^{M} \pi_j \exp\left(-D\left(p_i \| p_j\right)\right)\right),$$

PaiDEs offer a variety of options for $D(p_i \| p_j)$, such as KL divergence, Wasserstain distance, etc..



Latent Space
Conditioning
Forward Diffusion Process
$\mathcal{E}$
$\mathcal{D}$
z_0
z_T
Component 1
Class Label
Component 2
z_0^1
z_1^1
z_{b-2}^1
z_{b-1}^1
z_b^i
z_T
z_0^2
z_1^2
z_{b-2}^2
z_{b-1}^2
Pixel Space

Ensemble on Conditional Encoder.

The choice of D: 2-Wasserstein Distance:

$$W_2(p_i \| p_j) = \| \mu_i - \mu_j \|_2^2 + \mathrm{tr}\left[\left(\Sigma_i + \Sigma_j - 2\left(\Sigma_i^{\frac{1}{2}} \Sigma_j \Sigma_i^{\frac{1}{2}}\right)^{\frac{1}{2}}\right)\right].$$

When applying DDIM, $\Sigma_i = \Sigma_j = \mathbf{0}$, KL divergence and Bhattacharyya distance are undefined, so 2-Wasserstein Distance is a more general choice.

In DDIM setting:

$$\hat{I}(z_{t-1}, \theta \mid z_t, x, b) = -\sum_{i=1}^{M} \pi_i \ln\left(\sum_{j=1}^{M} \pi_j \exp(-W_2(p_i \| p_j))\right),$$

where $W_2(p_i \| p_j) = \| \mu_i - \mu_j \|_2^2$.

Main Idea: Epistemic Uncertainty == Model's Knowledge based on Training Data

Split ImageNet into four bins and train model on them:

- 100 Classes: 1 random image per class from bin 1.
- 100 Classes: 10 random images per class from bin 10.
- 100 Classes: 100 random images per class from bin 100.
- 700 Classes: All 1300 images per class from bin 1300 were utilized.

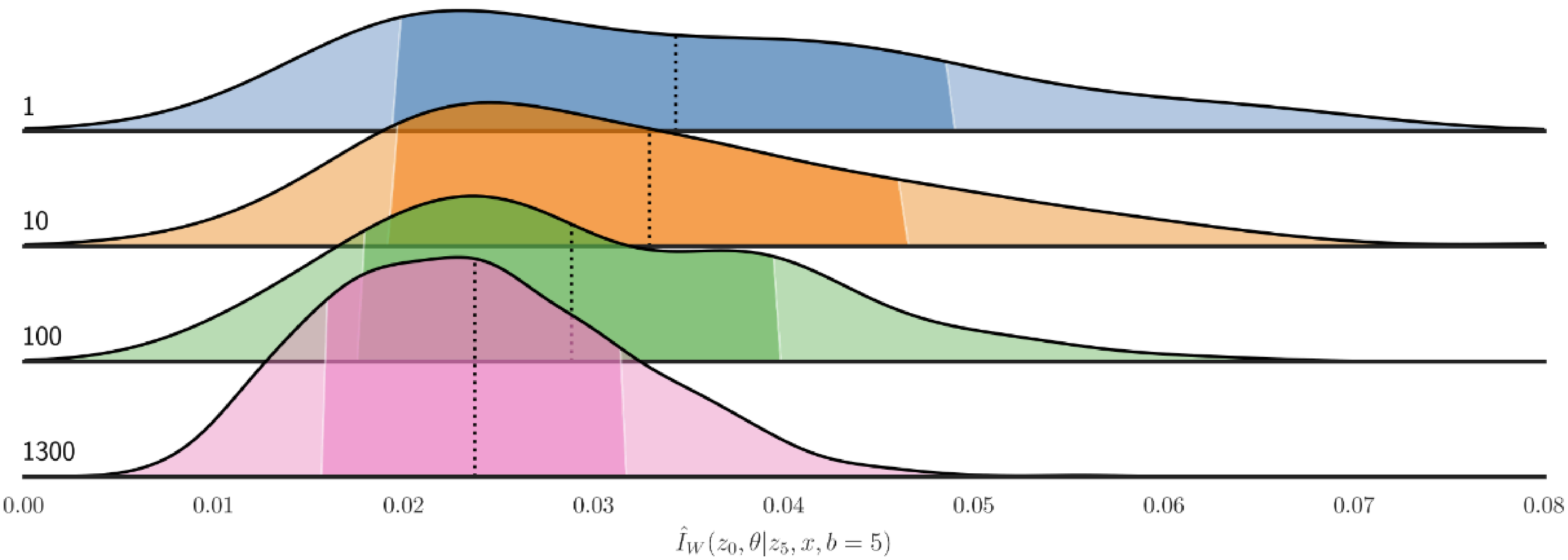
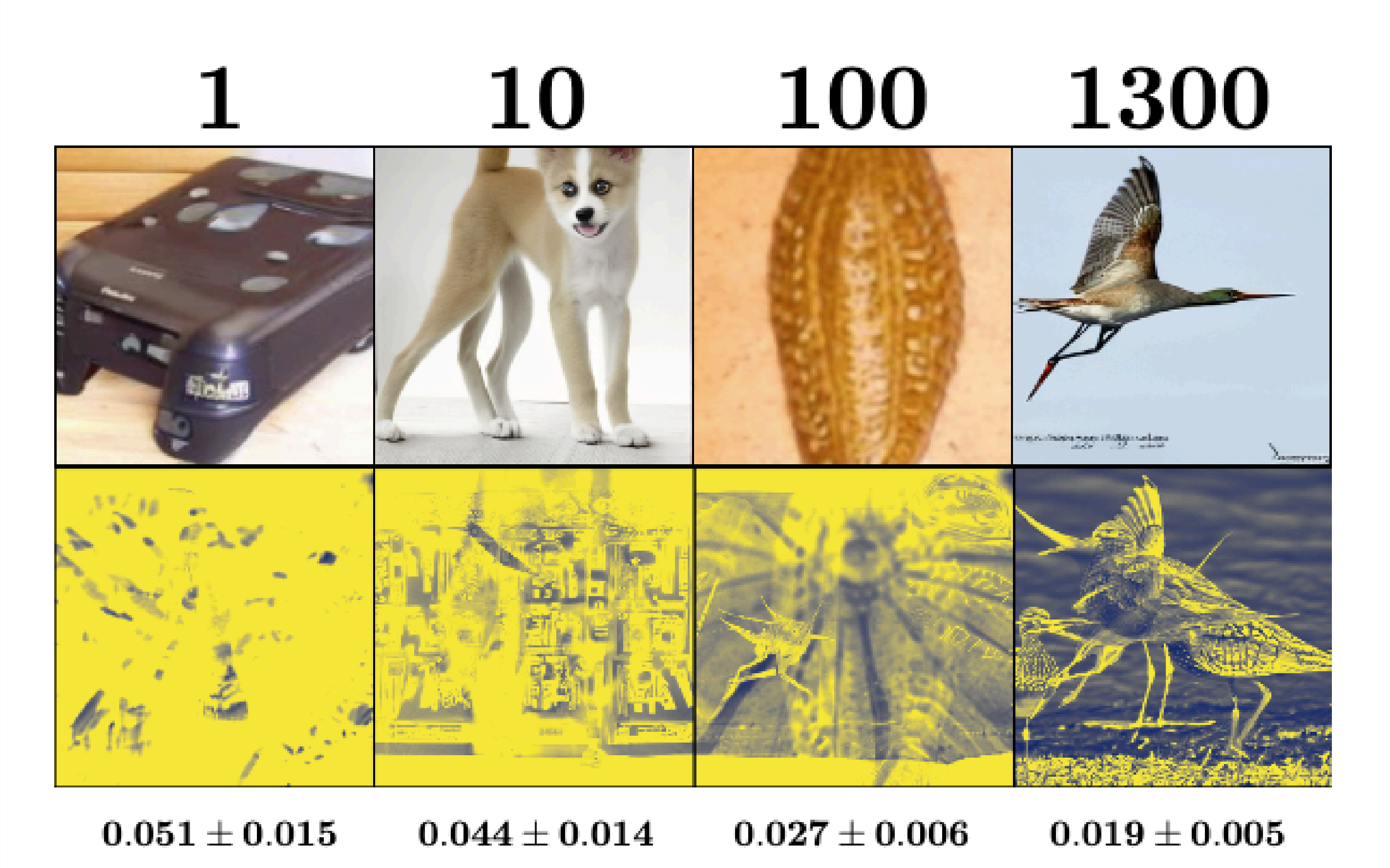


Figure 5: This figure displays uncertainty distributions for each bin, derived from corresponding class uncertainty estimates.

1.4 Experiment Design



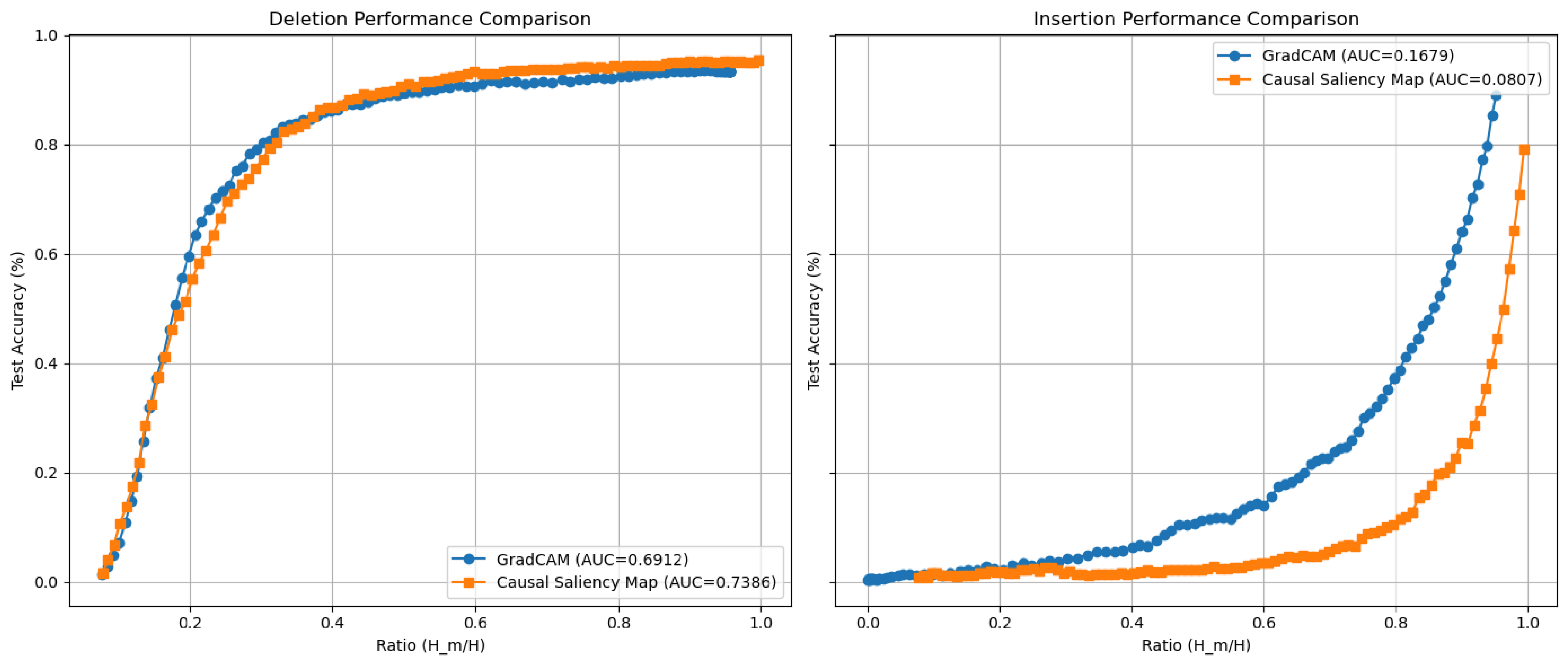
Yellow: high EU ; Blue: low EU

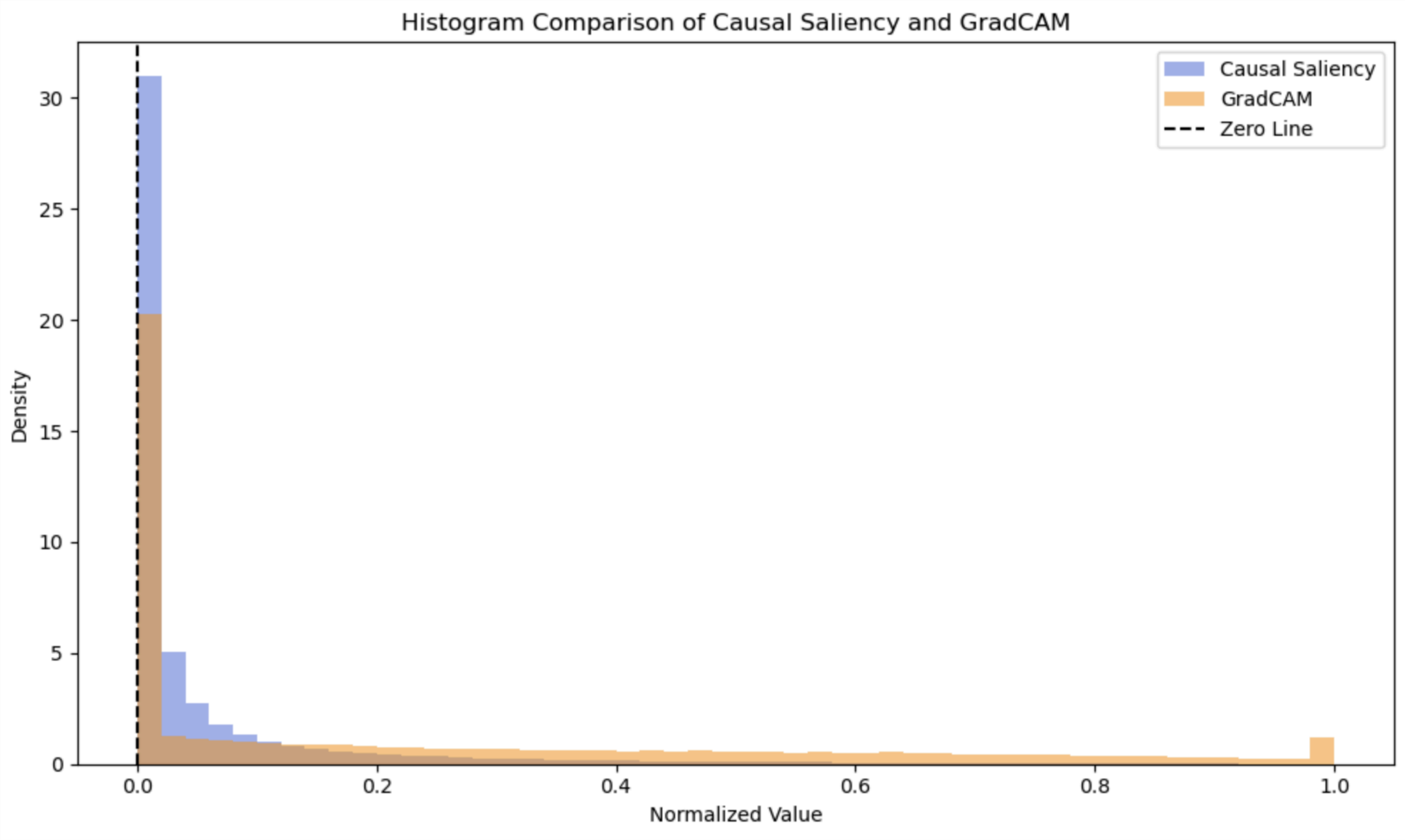
2. Preliminary Results of Causal Saliency Map

2. Preliminary Results of Causal Saliency Map

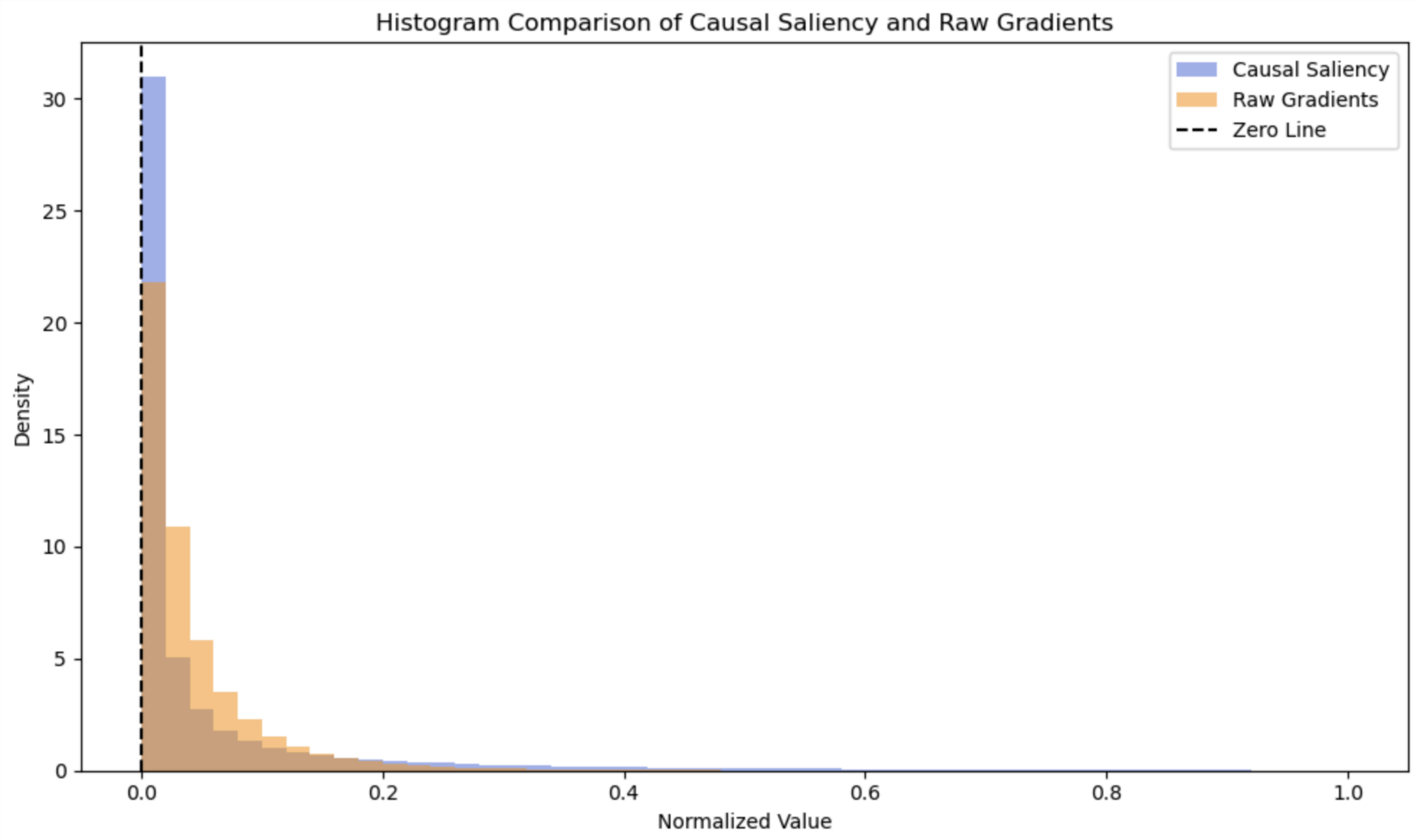
	GradCAM	Gradients	Lime	Occlusion	Causal Saliency Map
Deletion ⬆	0.7772	0.3730	0.6682	0.4814	**0.7862**
Insertion ⬇	0.1911	**0.0442**	**0.0514**	0.1374	0.0807

2. Preliminary Results of Causal Saliency Map

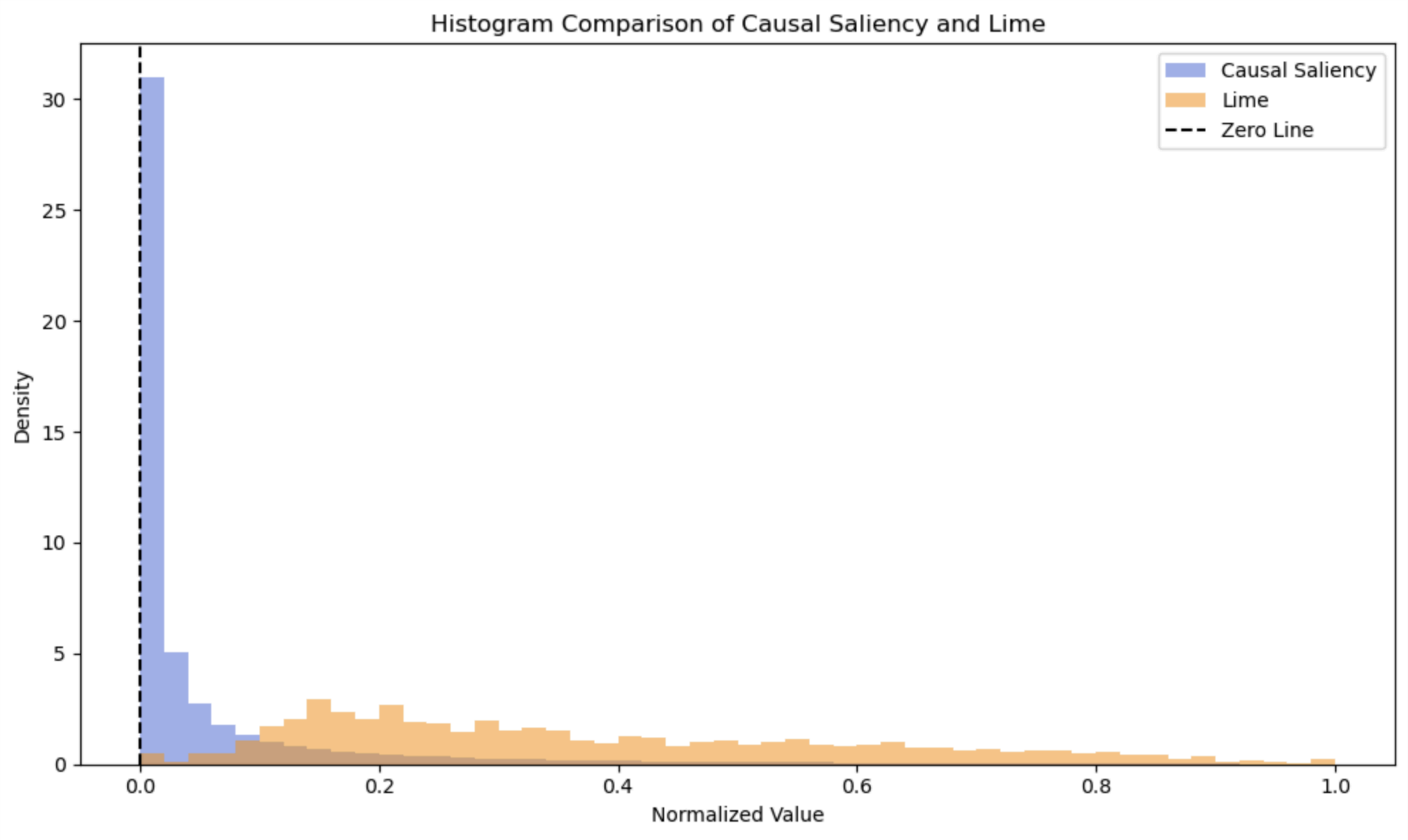



Histogram Comparison of Causal Saliency and GradCAM
Causal Saliency
GradCAM
Zero Line
Density
Normalized Value
0
5
10
15
20
25
30
0.0
0.2
0.4
0.6
0.8
1.0

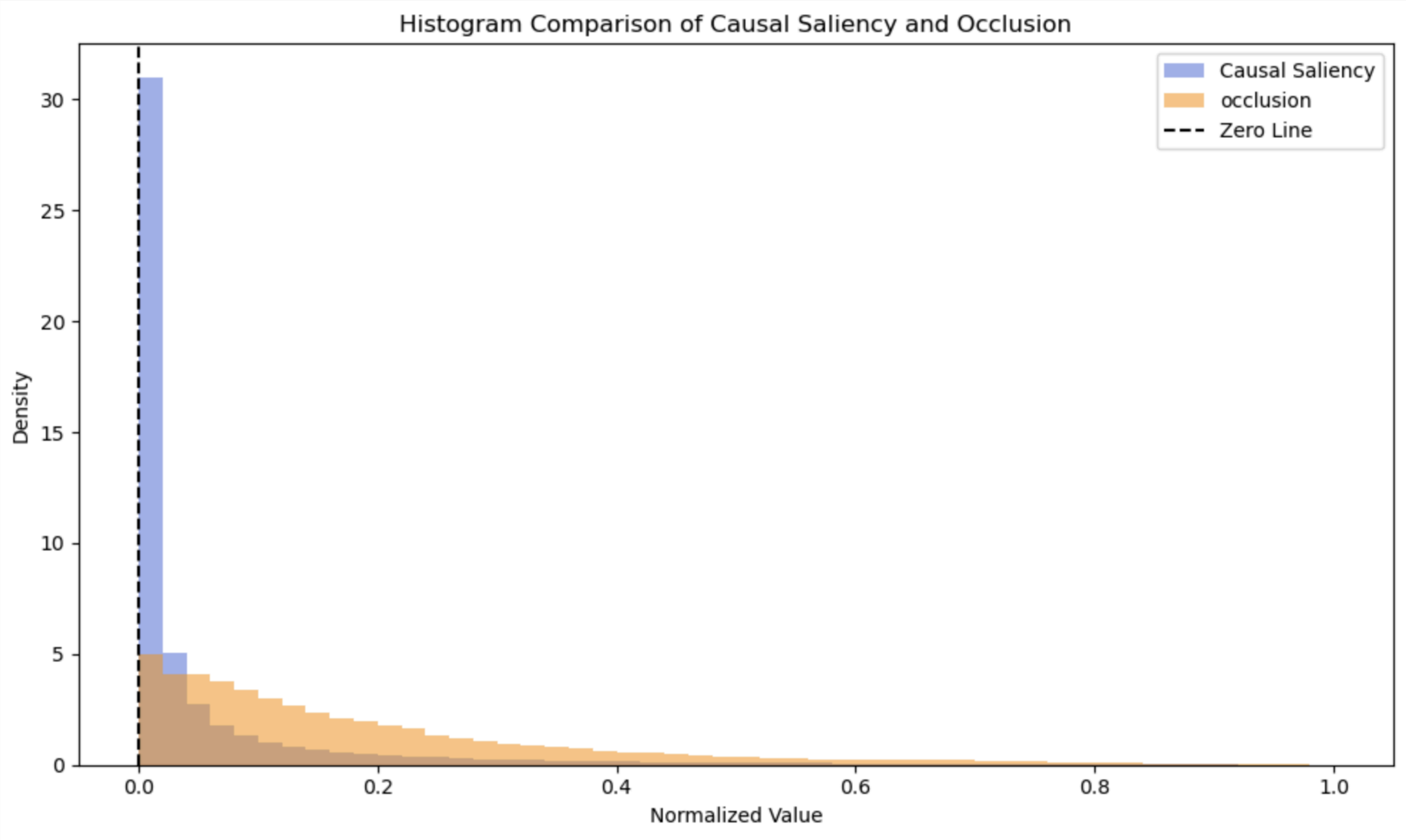
2.1 Check the Saliency Map



2.1 Check the Saliency Map



2.1 Check the Saliency Map



3. About experiments on Credal Nets

1. The training of credal nets is really unstable

base model: Res50	Accuracy ↑	SVH vs Cifar 10 OOD ↑
Deep Ensemble	0.9027	0.7835
Credal Nets	0.9273	0.9123

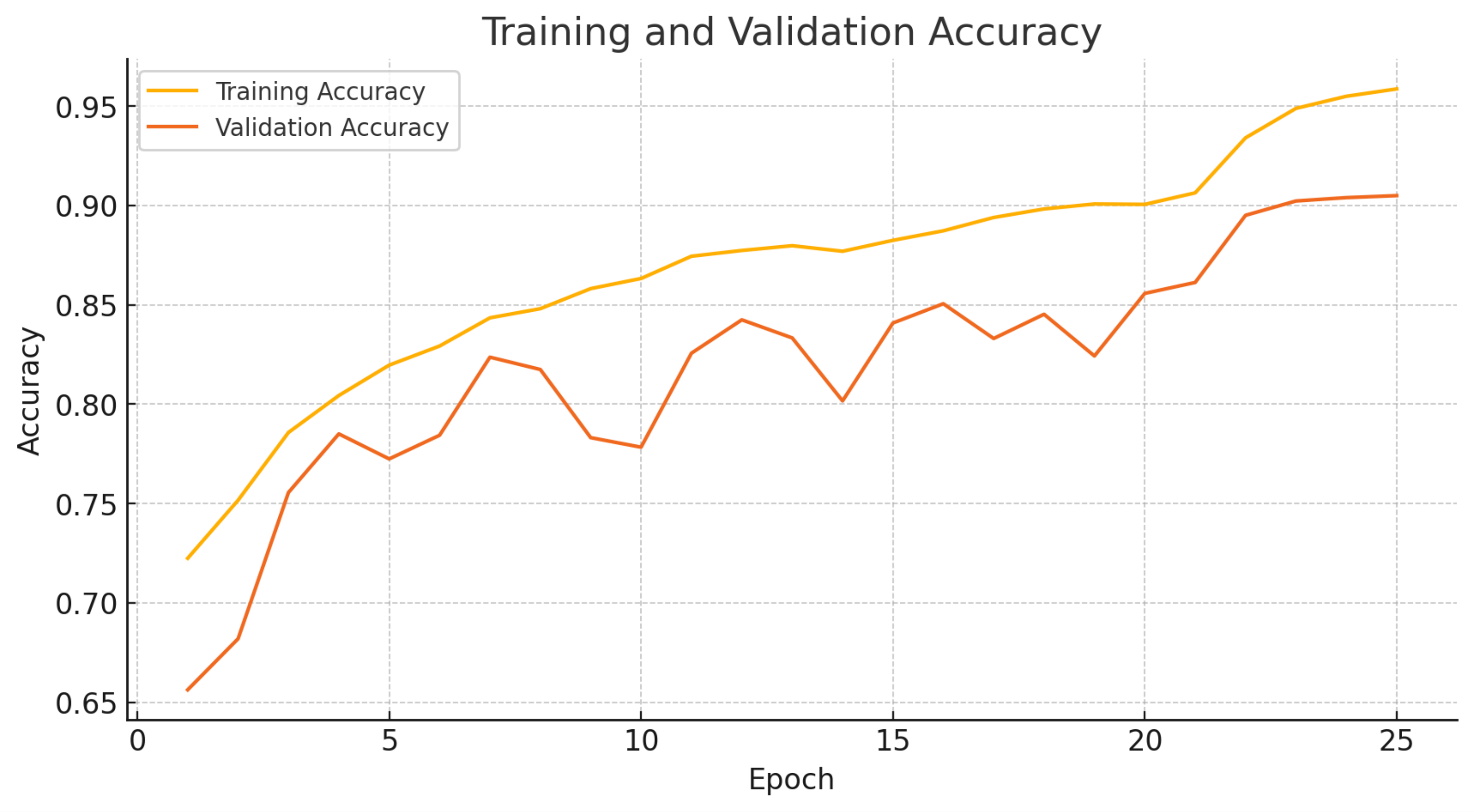
2. In the original paper, they even did not mention the result on Res50

Table 3: Test accuracy (%, ↑) and ECE (↓) of DEs-5 and CreDEs-5 on CIFAR10 as ID dataset (left). AUROC and AUPRC scores (%, ↑) for OOD detection on CIFAR10 vs SVHN/Tiny-ImageNet (right). Results averaged over 15 runs. The Best results are in bold.

			CIFAR10 (ID)			CIFAR10 vs SVHN		CIFAR10 vs Tiny-ImageNet	
			Test Accuracy	ECE		AUROC	AUPRC	AUROC	AUPRC
VGG16	DEs-5		85.53±0.10	0.0815±0.0011	$H(\tilde{\boldsymbol{q}})-\tilde{H}(\boldsymbol{q})$	82.19±0.82	87.52±0.81	78.58±0.15	73.28±0.23
	CreDEs-5 (Ours)	$\hat{i}_{\min}$	**87.94±0.11**	**0.0203±0.0014**	$\overline{H}(\mathbb{Q})-\underline{H}(\mathbb{Q})$	**87.68±0.73**	**93.47±0.57**	**82.56±0.28**	**80.81±0.52**
		$\hat{i}_{\max}$	**87.92±0.11**	**0.0611±0.0012**					
ViT Base	DEs-5		90.43±0.97	0.0181±0.0019	$H(\tilde{\boldsymbol{q}})-\tilde{H}(\boldsymbol{q})$	77.71±1.67	88.73±0.32	82.27±0.79	78.85±0.81
	CreDEs-5 (Ours)	$\hat{i}_{\min}$	**93.60±0.40**	**0.0107±0.0014**	$\overline{H}(\mathbb{Q})-\underline{H}(\mathbb{Q})$	**88.57±2.08**	**93.24±1.25**	**88.73±0.32**	**87.84±0.52**
		$\hat{i}_{\max}$	**93.59±0.39**	**0.0104±0.0012**					

But actually simple NN can achieve 92%+ accuracy and 0.90+ AUROC score.

3. About experiments on Credal Nets



3. About experiments on Credal Nets