Weekly Study Report

Wang Ma

2025-03-04

Electrical, Computer, and Systems Engineering Department
Rensselaer Polytechnic Institute

Outline

1. More Experiments on Credal Ensemble and Credal Wrapper

1.1 Overall Results

OOD detection.

	Cifar 10 (Resnet 18) vs SVHN			Cifar 100 (Resnet 50) vs SVHN		
Models	Test Accuracy	AUROC	AUPRC	Test Accuracy	AUROC	AUPRC
Deep Ensemble	**95.95%**	0.9683	0.9832	85.35%	**0.8567**	**0.9227**
Credal Ensemble	**96.43%**	**0.9813**	**0.9924**	**86.02%**	0.852	0.8838
Single Credal Nets	95.78%	0.8646	0.9495	83.44%	0.7354	0.8186
Credal Wrapper	95.51%	**0.9761**	**0.9877**	84.93%	**0.8956**	**0.9488**

- The models used by Credal Wrapper are exactly the same as the models used in Deep Ensemble.
- The test accuracy of credal methods are using the mid-points to do prediction.
- The original paper uses pre-trained models and tune them for 5 epochs, where the models can be underfitting (their test accuracies are lower). I fine-tuned the models for 100 epochs with early stop.
- Using Epistemic Uncertainty to do OOD detection.

1.2 Original Credal Ensemble Results

	CIFAR10		CIFAR100		ImageNet	
	Test Accuracy	ECE	Test Accuracy	ECE	Test Accuracy	ECE
DEs-5	93.32±0.13	0.0131±0.0010	75.80±0.28	0.0392±0.0027	77.92±0.02	0.2415±0.0009
CreDEs-5 $\hat{i}_{\min}$	**93.75±0.11**	**0.0092±0.0016**	**79.54±0.21**	**0.0366±0.0025**	**78.41±0.02**	0.5930±0.0006
(Ours) $\hat{i}_{\max}$	**93.74±0.11**	**0.0108±0.0017**	**79.65±0.19**	**0.0268±0.0023**	**78.51±0.02**	**0.1685±0.0004**

ID Samples	CIFAR10				CIFAR100				ImageNet	
OOD Samples	SVHN		Tiny-ImageNet		SVHN		Tiny-ImageNet		ImageNet-O	
Performance Indicator	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC
DEs-5 $H(\tilde{\boldsymbol{q}})-\tilde{H}(\boldsymbol{q})$	89.58±0.93	92.29±1.00	86.87±0.20	83.02±0.16	73.83±1.97	84.96±1.25	78.80±0.20	74.68±0.27	65.03±0.53	62.77±0.38
CreDEs-5 $\overline{H}(\mathbb{Q})-\underline{H}(\mathbb{Q})$	**96.55±0.25**	**98.17±0.17**	**88.10±0.26**	**87.85±0.35**	**78.55±1.15**	**86.57±0.65**	**82.54±0.26**	**77.60±0.44**	**67.82±0.06**	**62.80±0.12**

1.3 Original Credal Wrapper Results

Uncertain Metrics			CIFAR10 (ID)				CIFAR100 (ID)				ImageNet (ID)	
			SVHN (OOD)		Tiny-ImageNet (OOD)		SVHN (OOD)		Tiny-ImageNet (OOD)		ImageNet-O (OOD)	
			AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC
TU	Baseline	$H(\tilde{\boldsymbol{p}})$	94.80±0.43	97.26±0.29	88.80±0.19	87.21±0.29	78.53±1.94	88.83±1.01	80.75±0.15	77.65±0.19	50.20±0.07	50.44±0.06
	Ours	$\overline{H}(\mathbb{P})$	**95.44±0.37**	**97.57±0.23**	**89.30±0.17**	**87.97±0.25**	**80.71±1.96**	**89.97±0.99**	**81.46±0.14**	**78.29±0.17**	**54.87±0.08**	**52.27±0.05**
EU	Baseline	$H(\tilde{\boldsymbol{p}}) - \tilde{H}(\boldsymbol{p})$	89.58±0.93	92.29±1.00	86.87±0.20	83.02±0.16	73.83±1.97	84.96±1.25	78.80±0.20	74.68±0.27	65.70±0.41	63.20±0.35
	Ours	$\overline{H}(\mathbb{P}) - \underline{H}(\mathbb{P})$	**93.77±0.60**	**96.06±0.46**	**88.78±0.15**	**86.83±0.23**	**80.22±1.96**	**89.40±1.03**	**81.00±0.16**	**77.16±0.23**	**66.20±0.38**	***63.23±0.34***

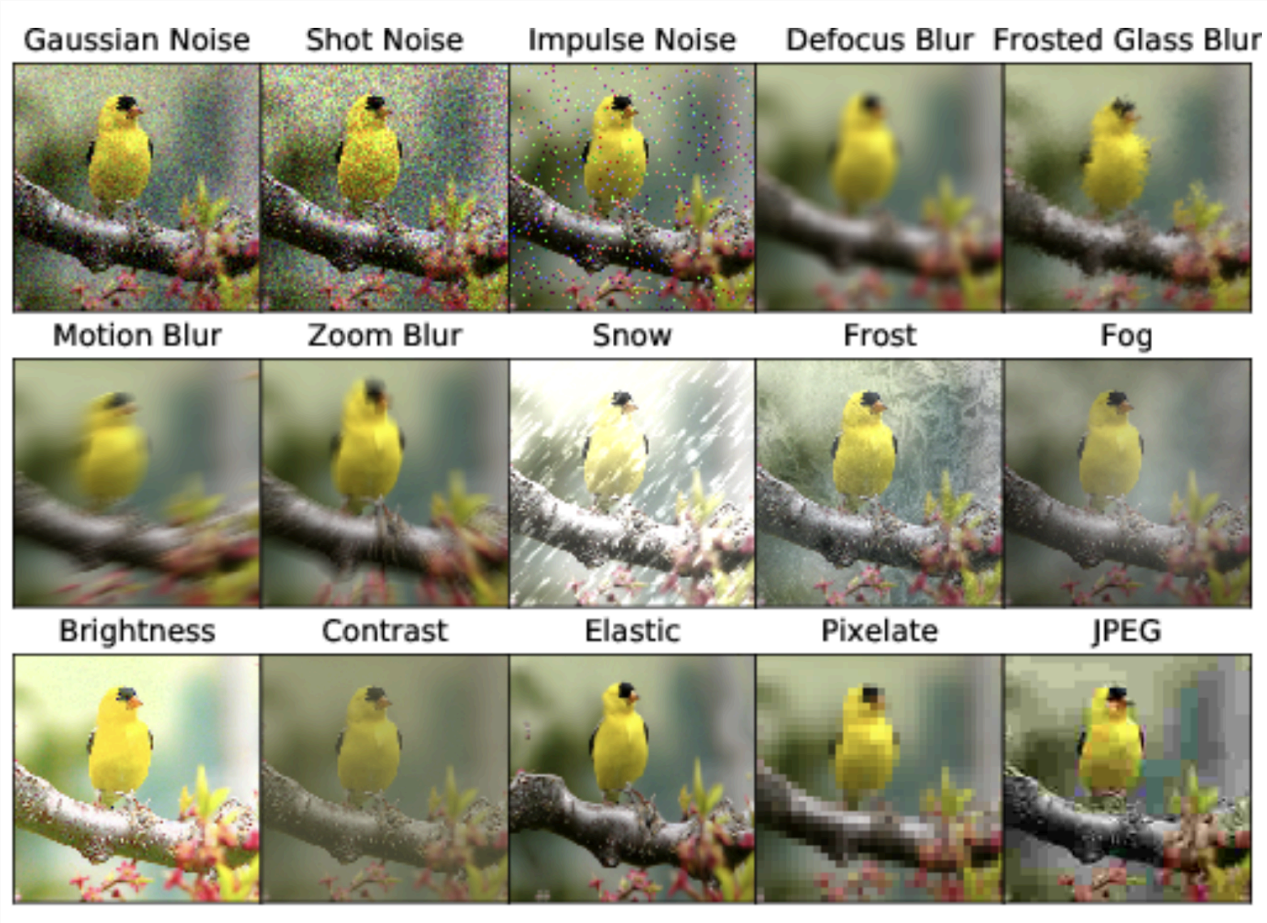
So, actually, their results also supports that Credal Ensemble is better on Cifar10, and Credal Wrapper is better on Cifar100.

2. Experiments of EU/AU on Corrupted Images

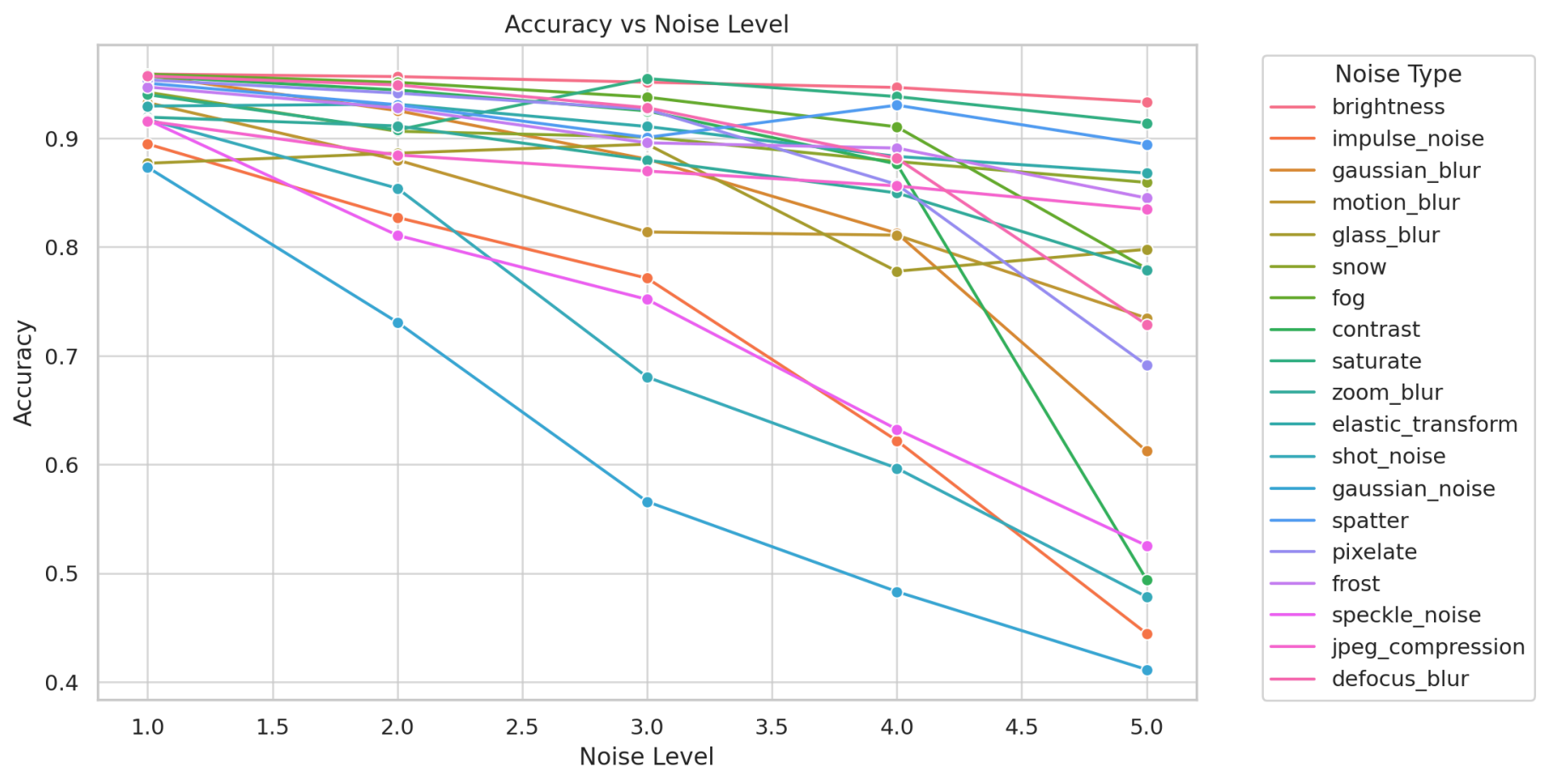
2.1 Basic Settings

- Standard Model: Resnet-18 on Cifar-10.
- Comparison: Cifar-10-c and Occluded Cifar-10

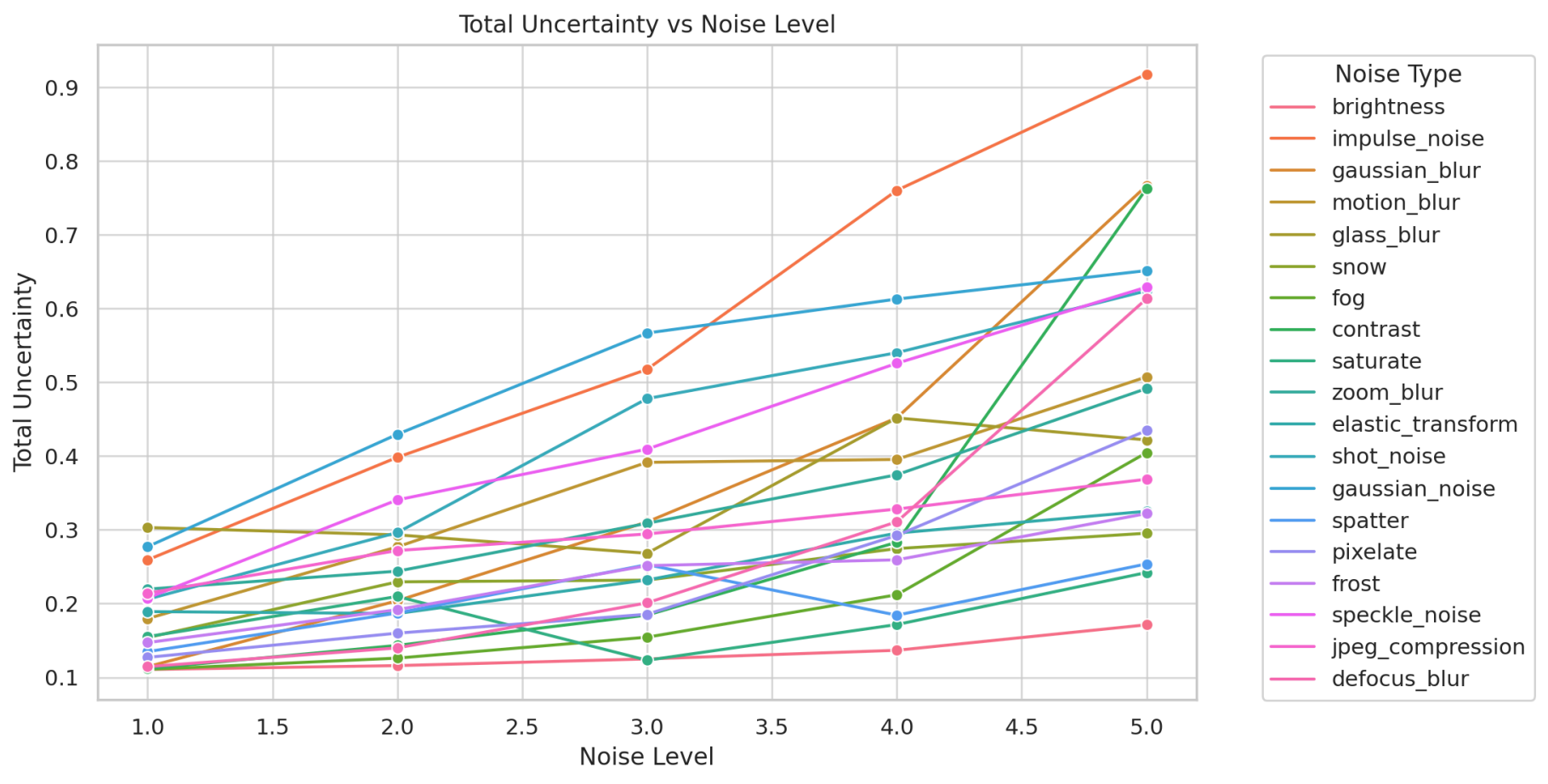
2.2.1 Cifar-10-c



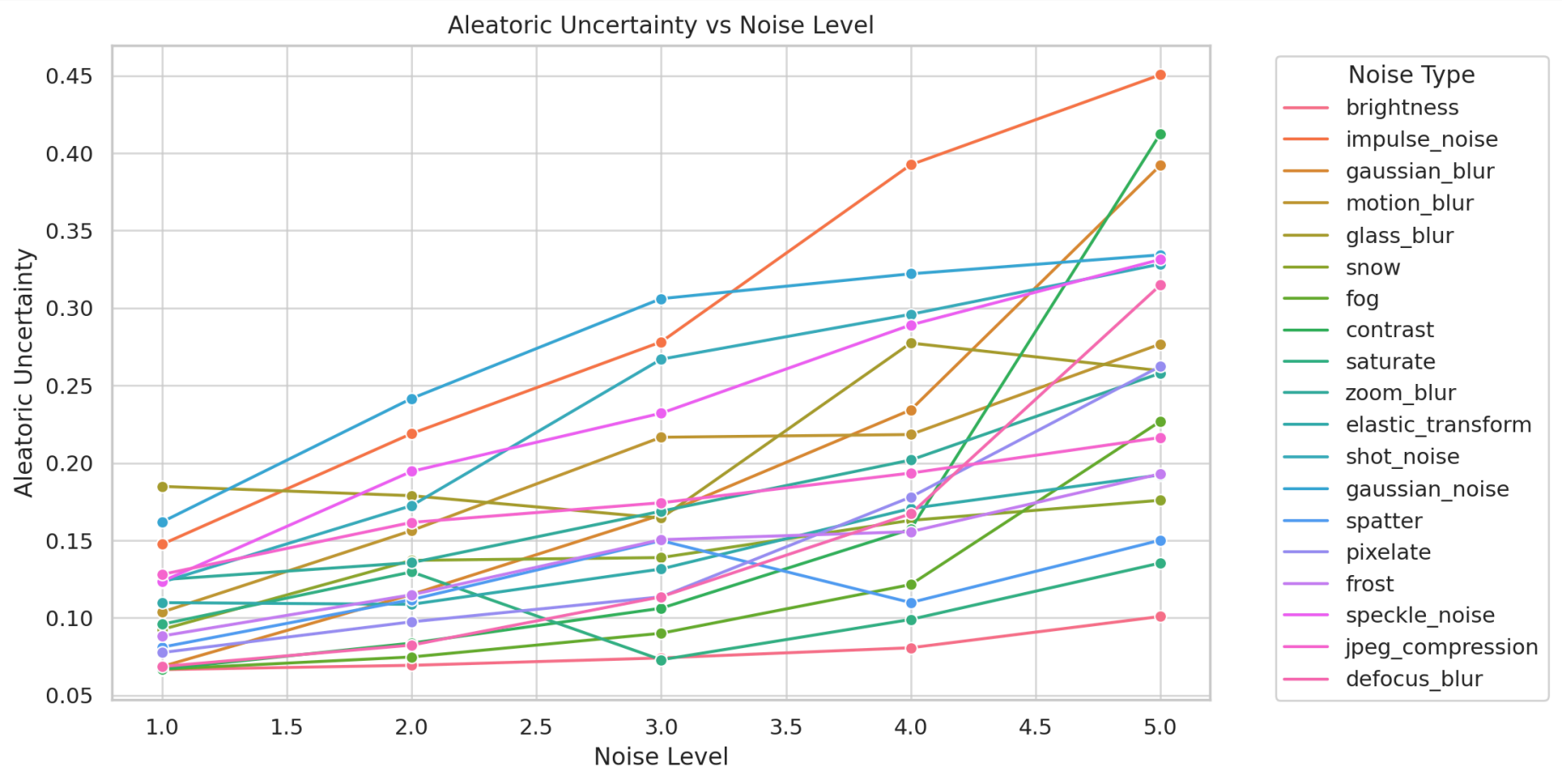
2.2.2 Accuracy v.s. Noise Level



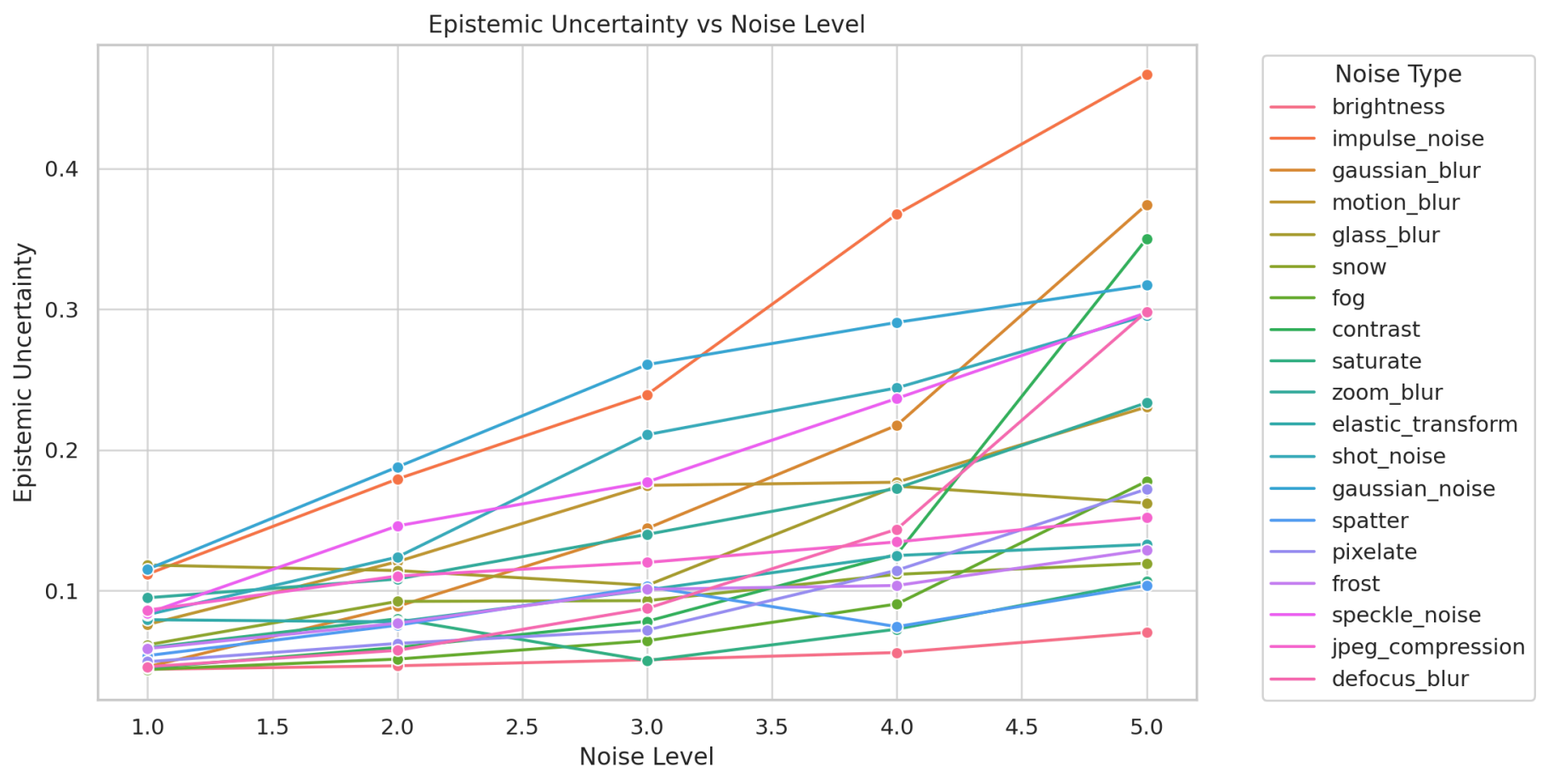
2.2.3 Total Uncertainty v.s. Noise Level



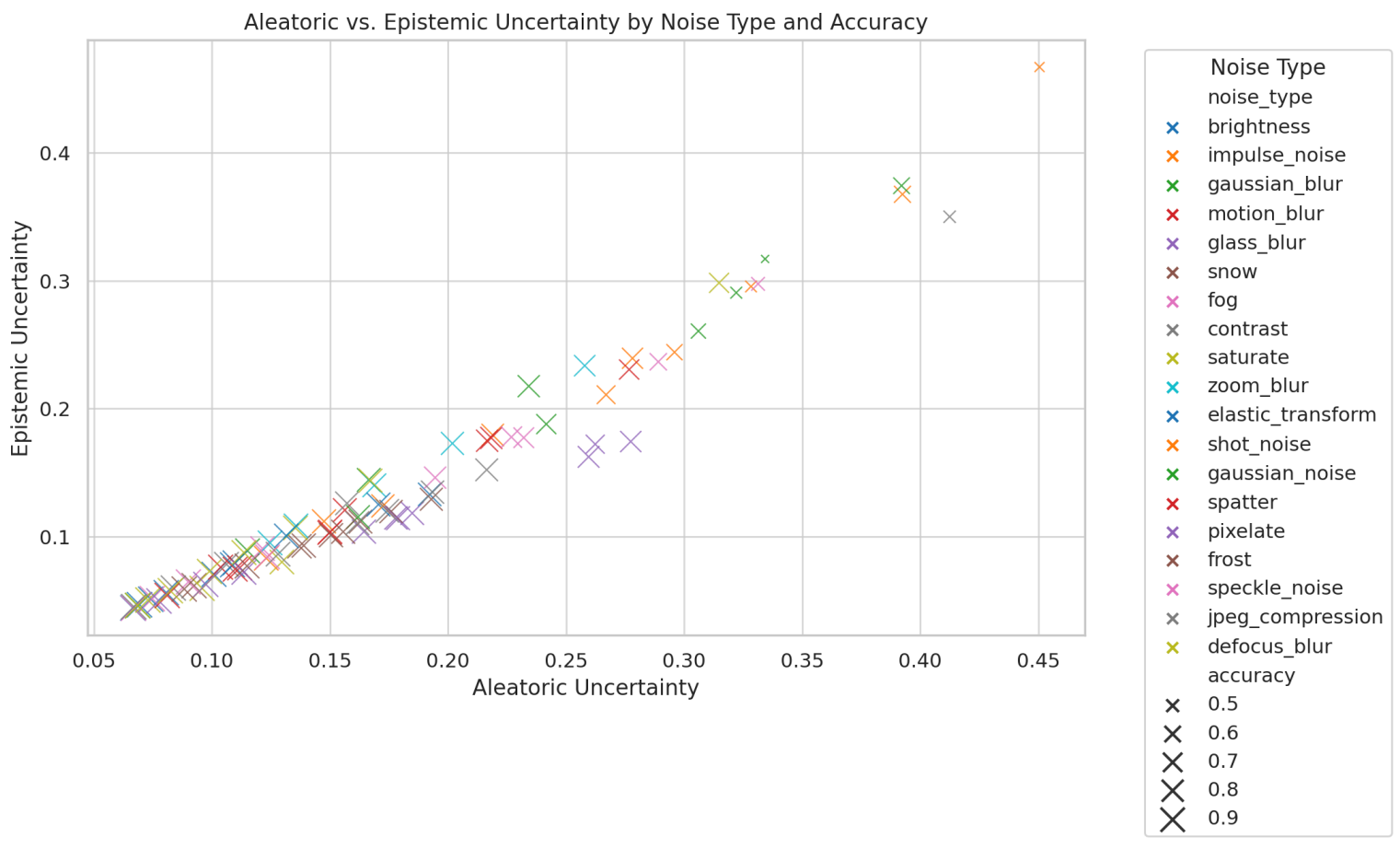
2.2.4 Aleatoric Uncertainty v.s. Noise Level



2.2.5 Epistemic Uncertainty v.s. Noise Level



2.2.6 Aleatoric Uncertainty v.s. Epistemic Uncertainty



2.2 Results on Cifar-10-c

The relationship between aleatoric and epistemic uncertainty appears highly linear:

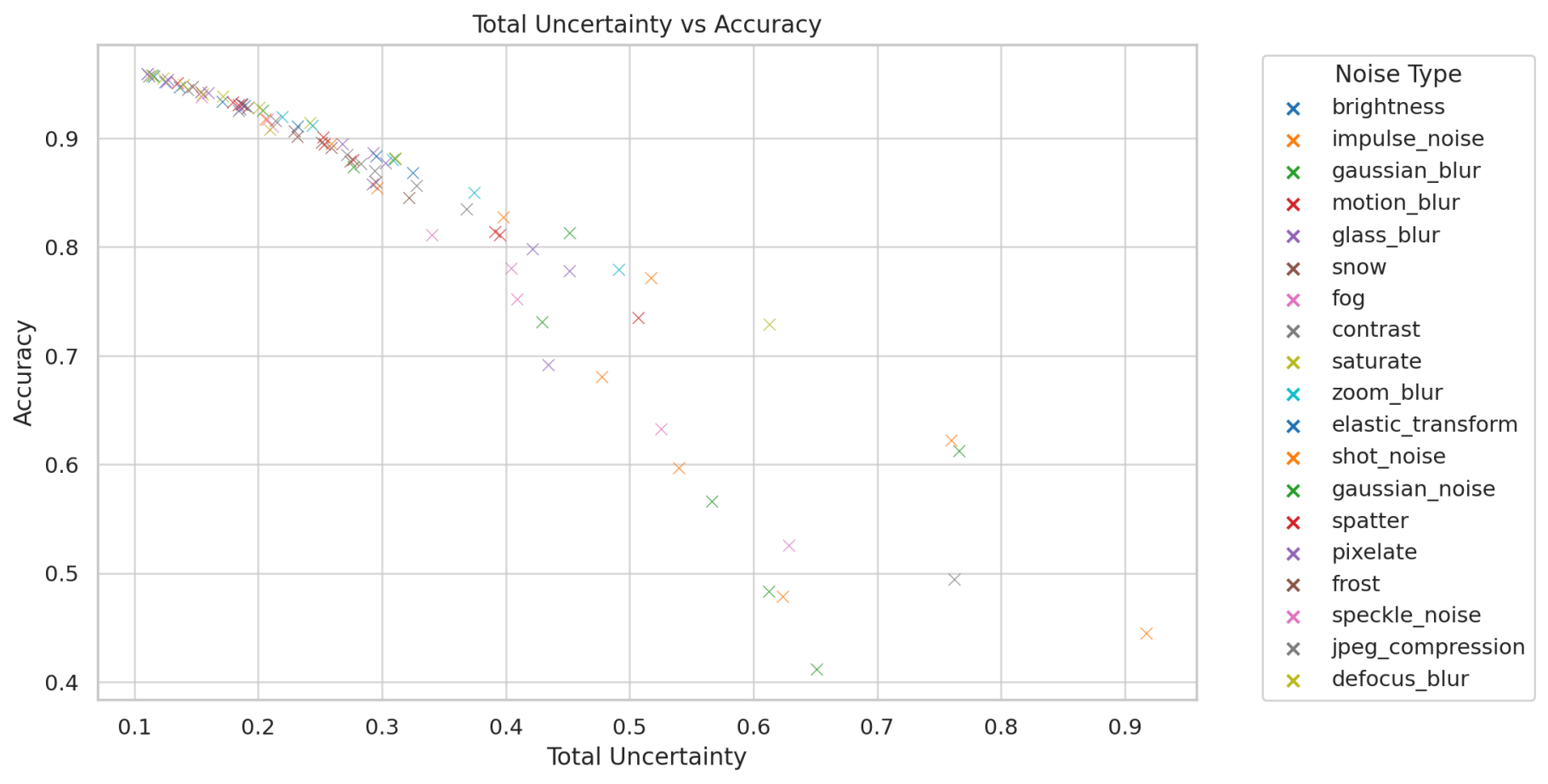
- Pearson correlation coefficient: 0.979 (very strong positive correlation)
- p-value: $2.01 \times 10^{\{-66\}}$ (indicating statistical significance)
- Regression equation:

$$\text{Epistemic Uncertainty} = -0.0323 + 0.9607 \times \text{Aleatoric Uncertainty}$$

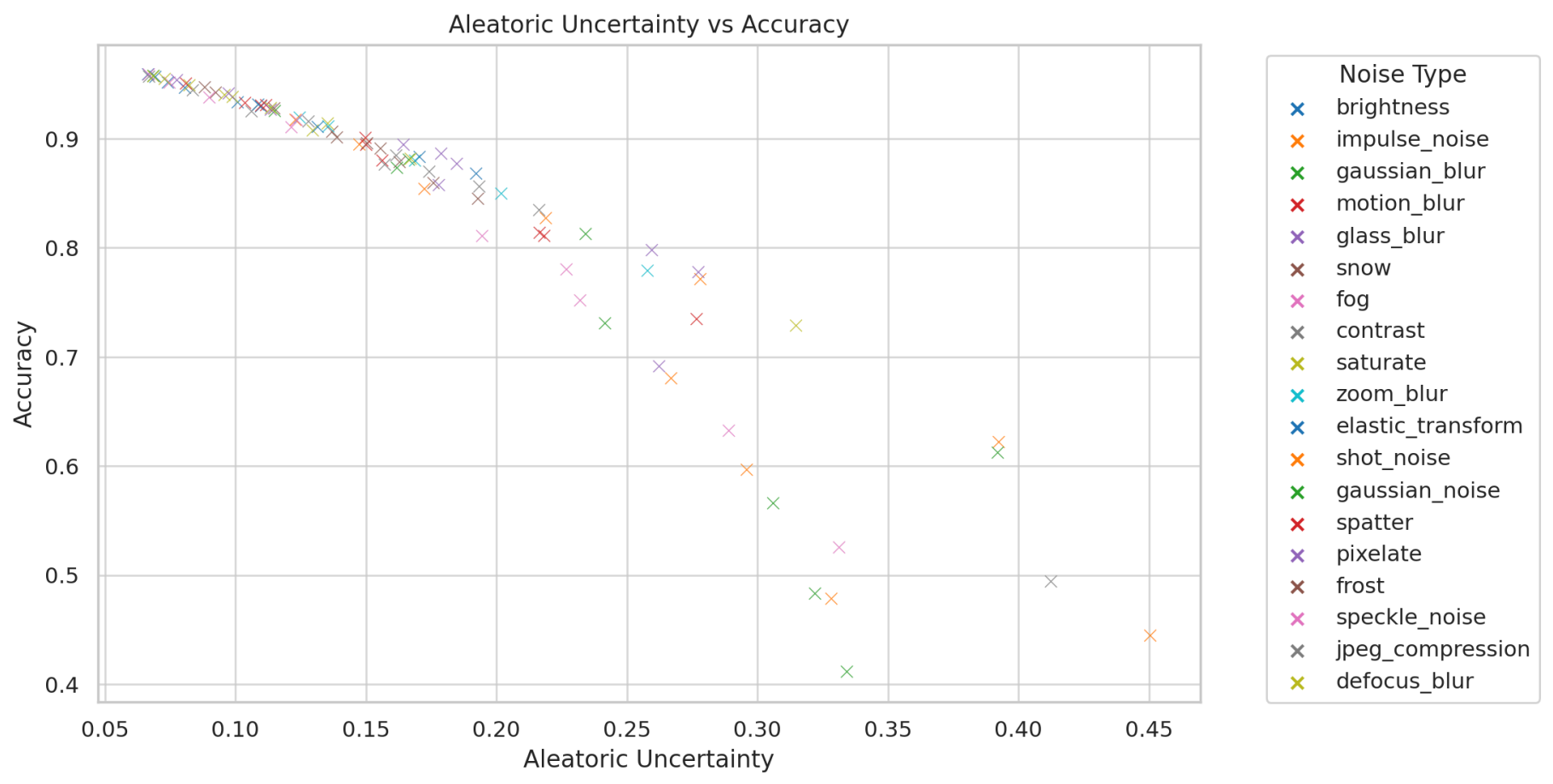
- The slope (0.9607) suggests a nearly one-to-one relationship.
- The R^2 value of 0.959 indicates that 95.9% of the variance in epistemic uncertainty is explained by aleatoric uncertainty.

This confirms a strong linear relationship between the two types of uncertainty in the dataset. The epistemic uncertainty increases as the aleatoric uncertainty increases, with the same scale of values.

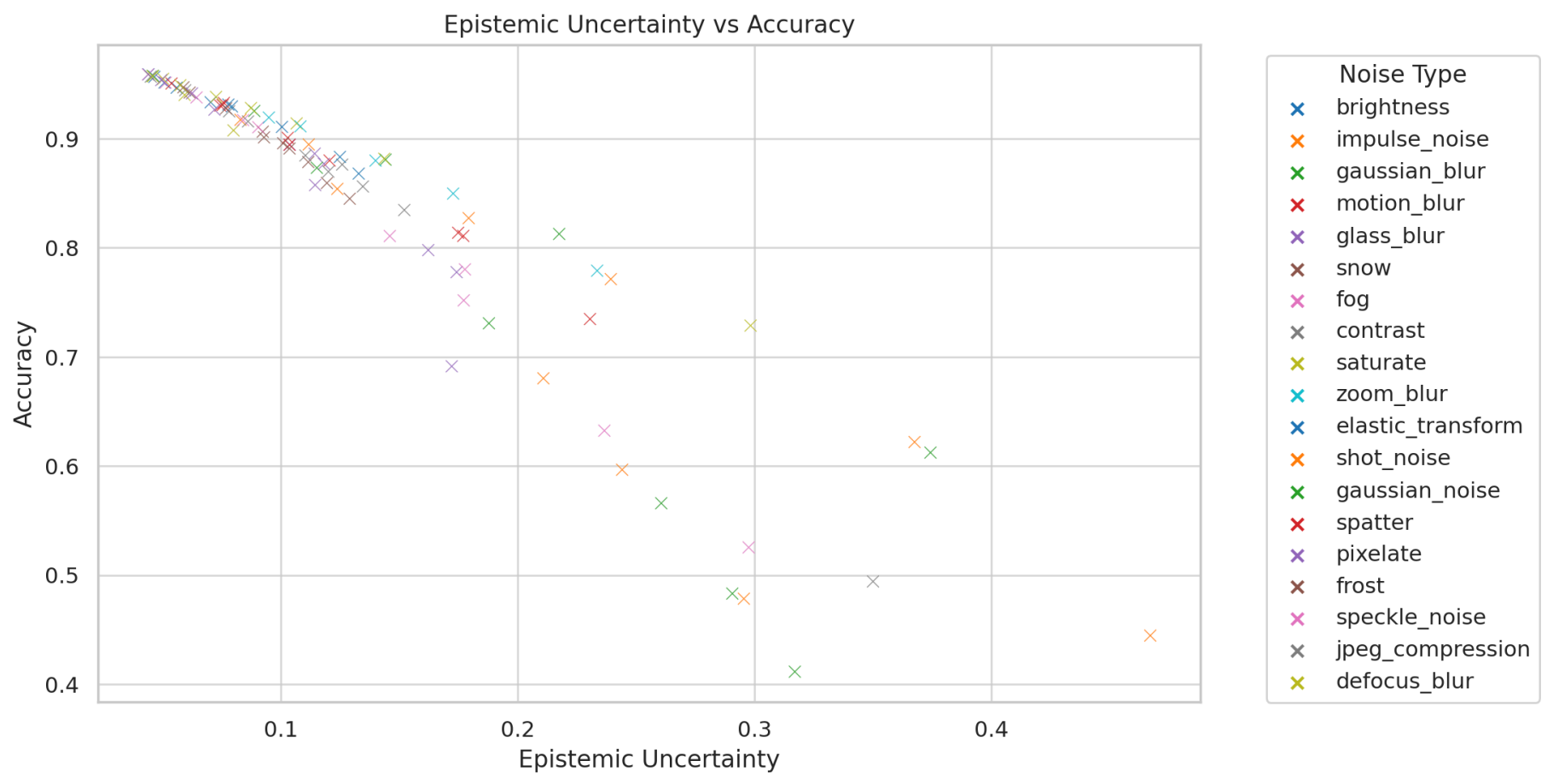
2.2.7 Total Uncertainty v.s. Accuracy



2.2.8 Aleatoric Uncertainty v.s. Accuracy

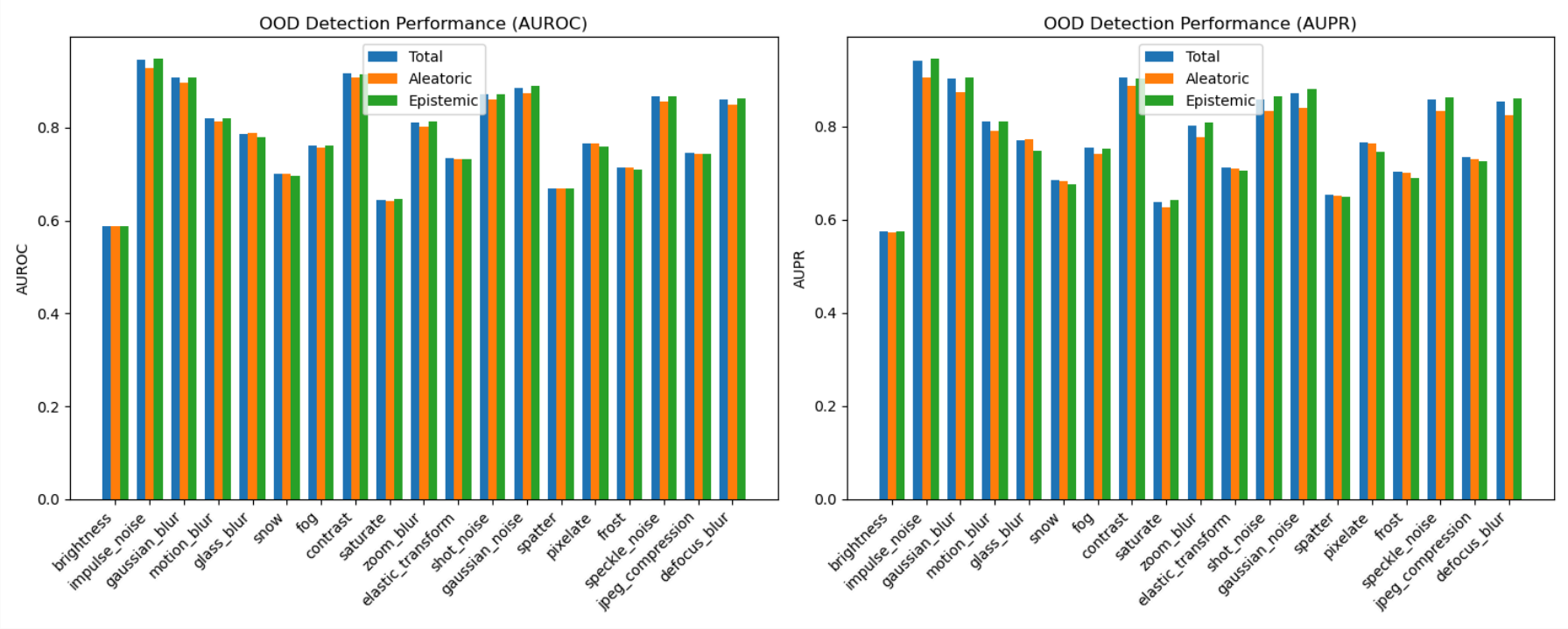


2.2.9 Epistemic Uncertainty v.s. Accuracy



2.2.10 OOD Detection Results

Let the Cifar-10-c be the OOD dataset and let model recognize them. Lower value means the data is more like the uncorrupted version.



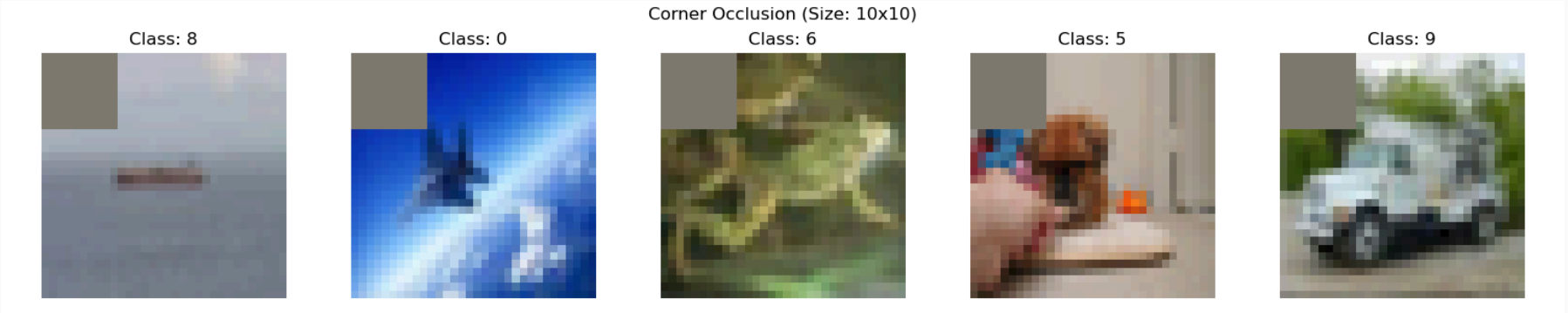
2.3.1 Occlusion Approach

- Occlusion Type: Center, Random Corner, Horizontal and Vertical
- Occlusion Size: 1*1, 2*2, 4*4, 8*8, 10*10, 16*16

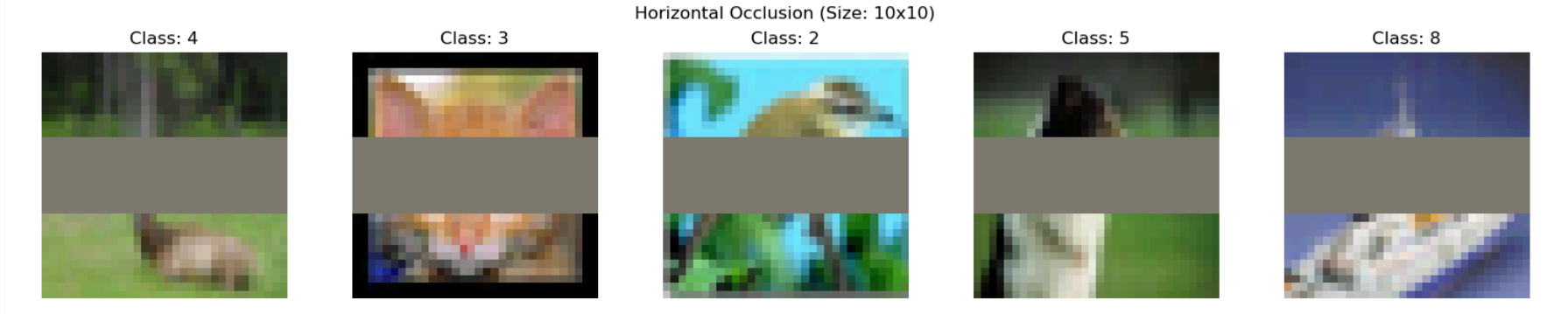
2.3.1.1 Center Occlusion

2.3.1.2 Random Occlusion

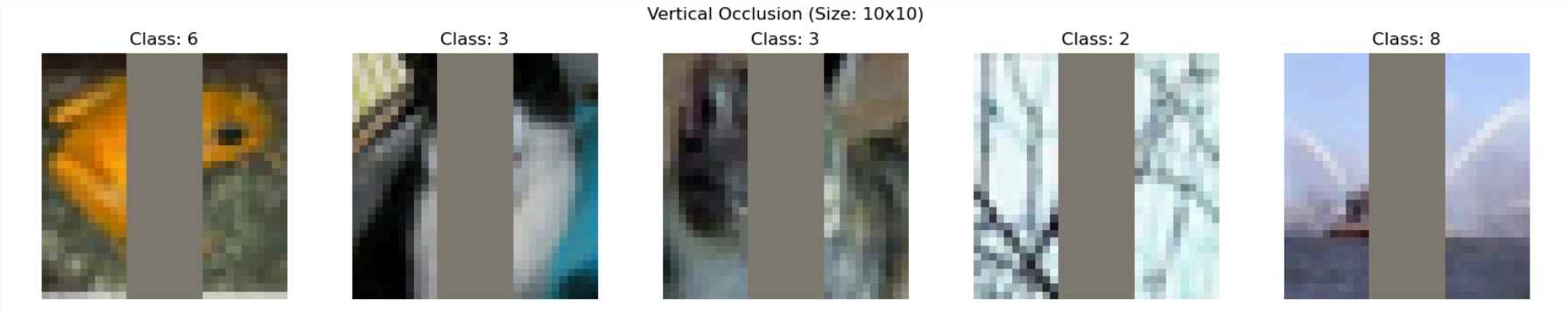
2.3.1.3 Corner Occlusion



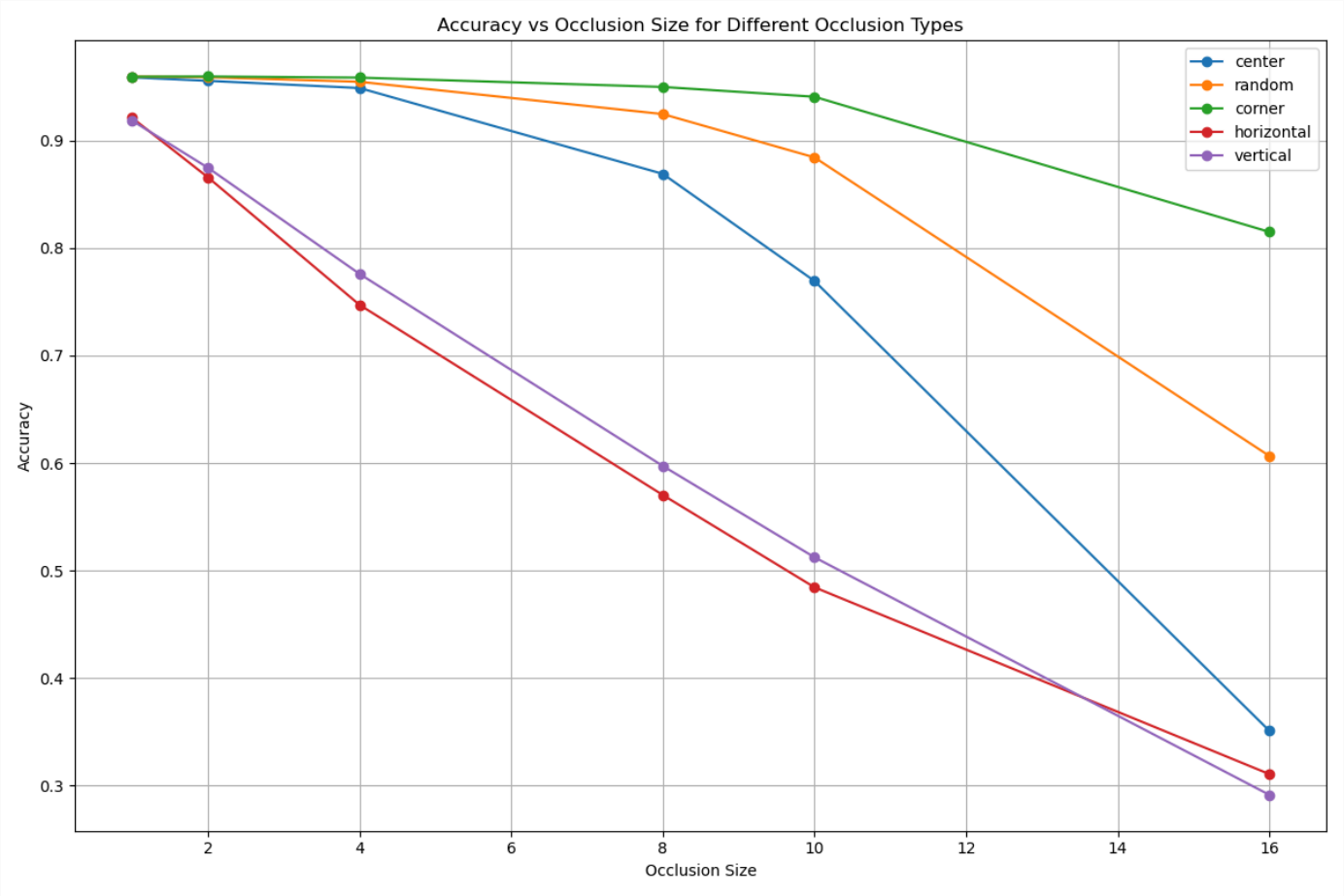
2.3.1.4 Horizontal Occlusion



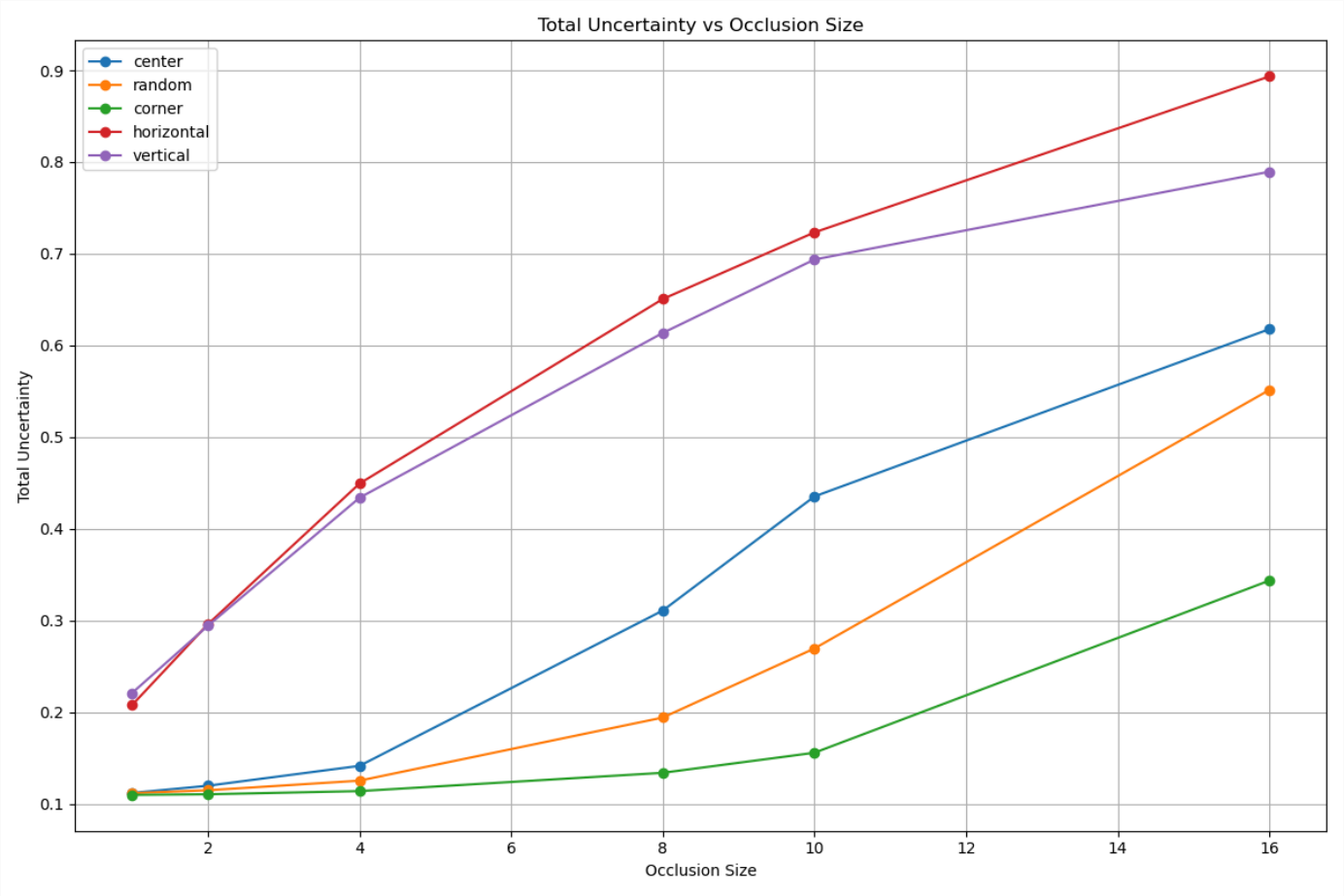
2.3.1.5 Vertical Occlusion



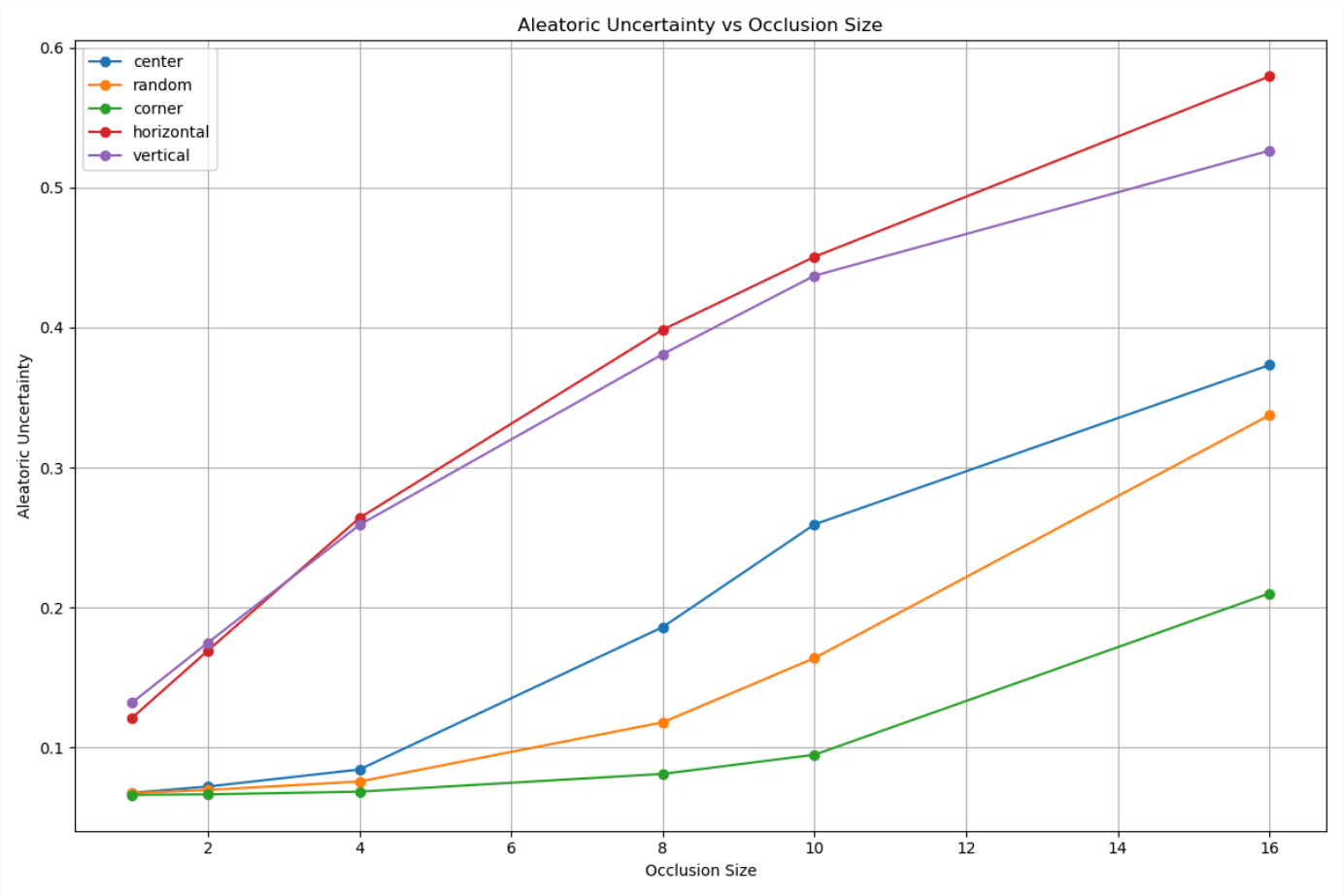
2.3.2 Accuracy v.s. Occlusion Size



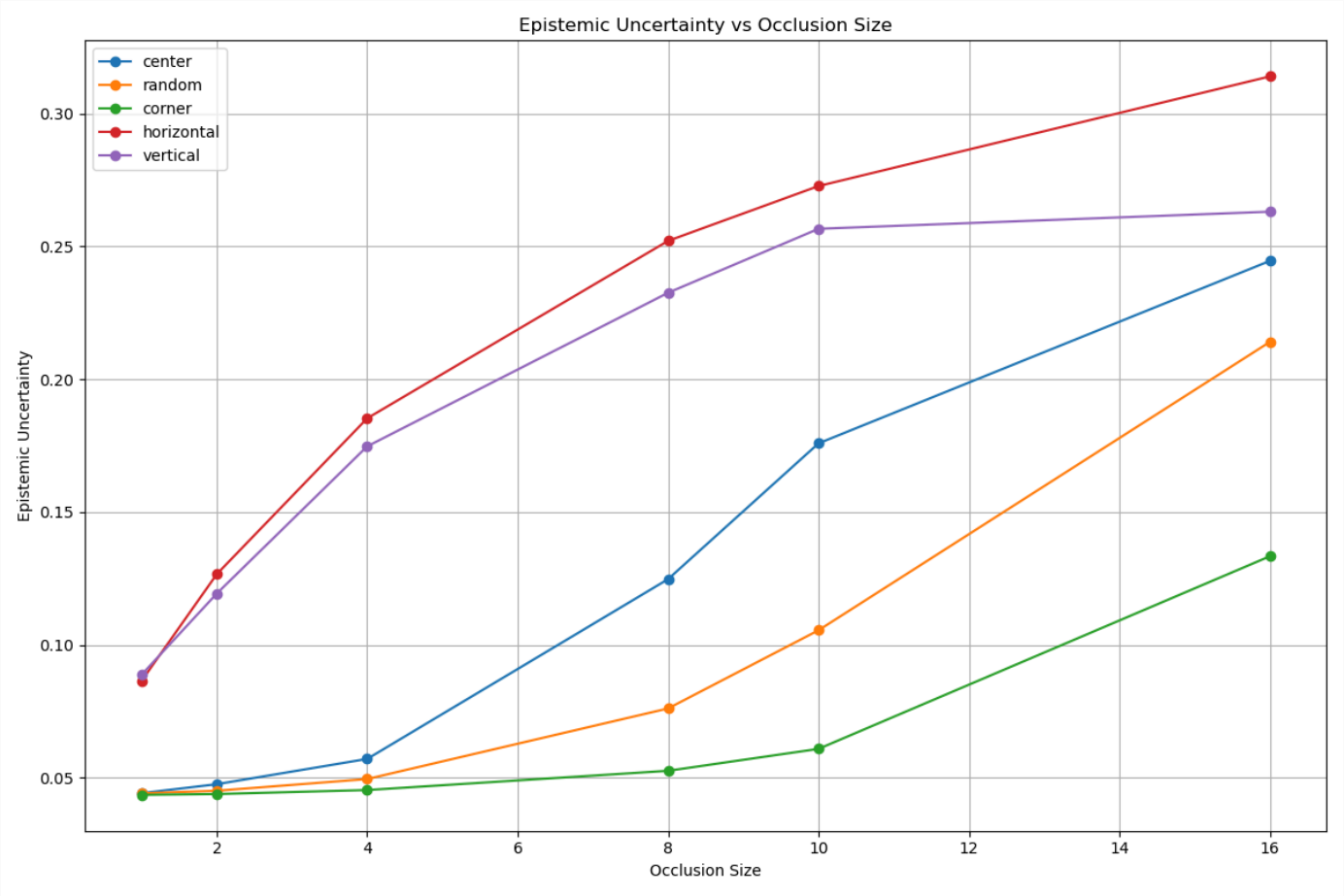
2.3.3 Total Uncertainty v.s. Occlusion Size



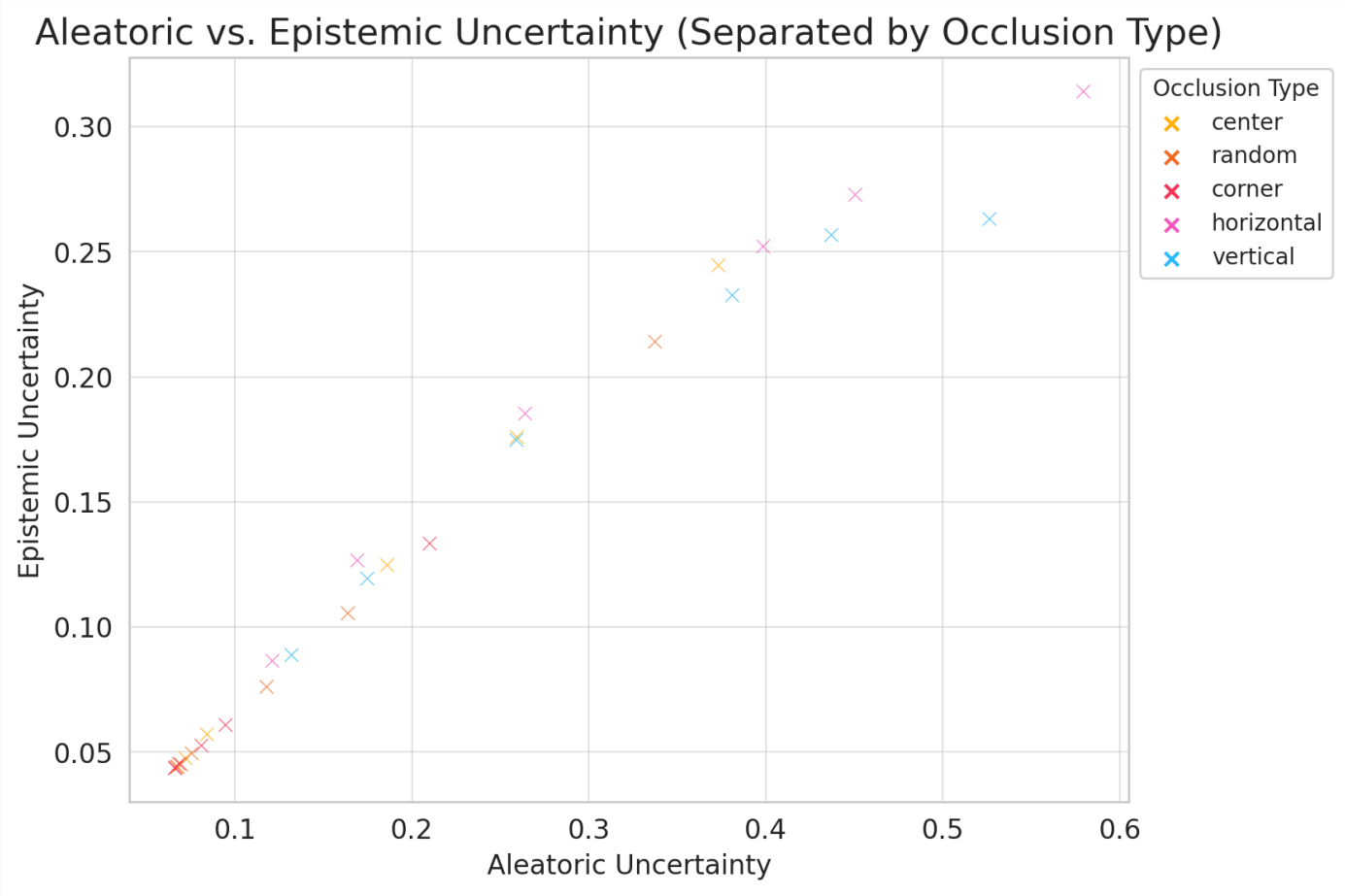
2.3.4 Aleatoric Uncertainty v.s. Occlusion Size



2.3.5 Epistemic Uncertainty v.s. Occlusion Size



2.3.6 Aleatoric Uncertainty v.s. Epistemic Uncertainty



2.3 Results on Occluded Cifar-10

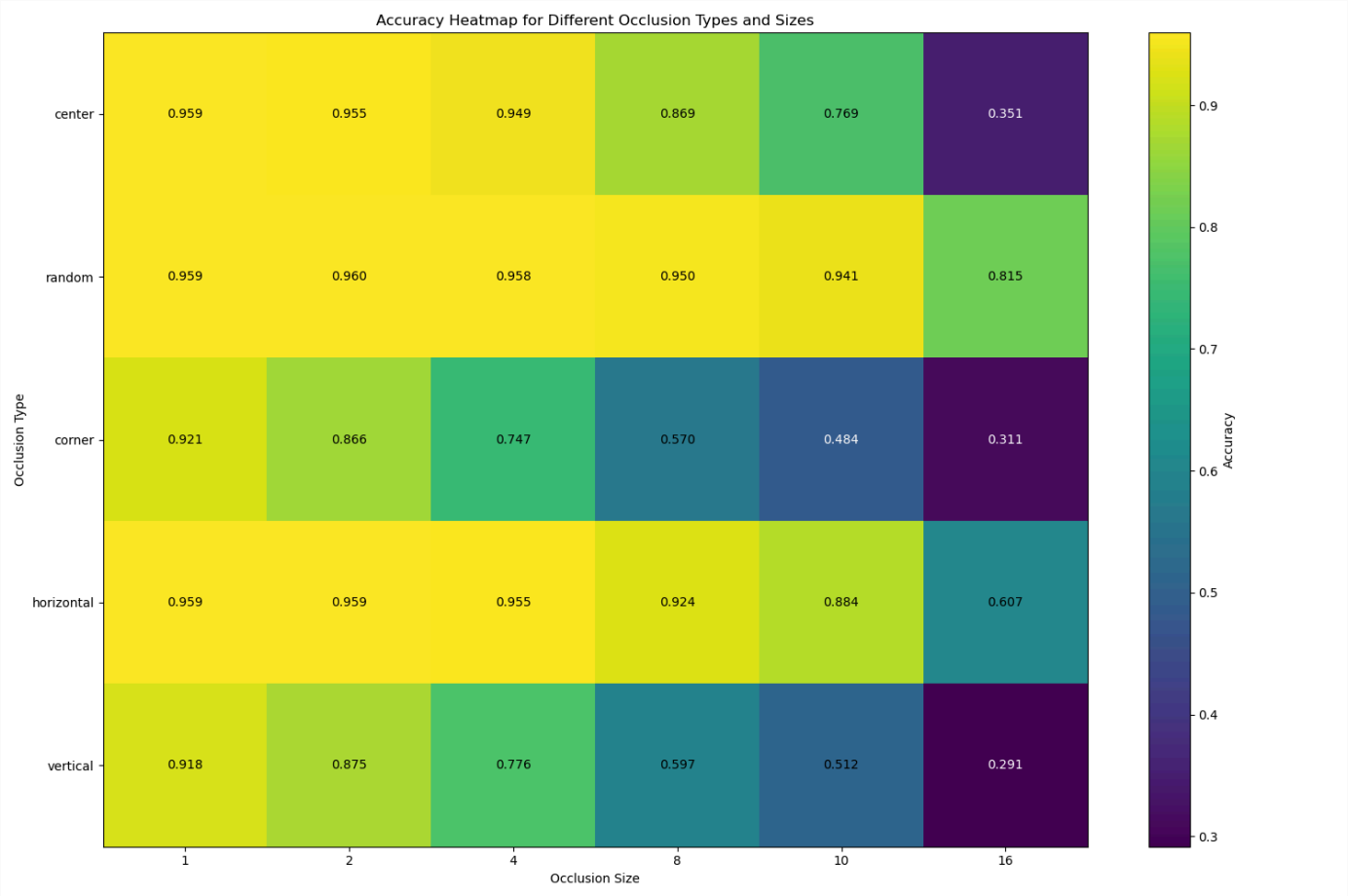
- Pearson Correlation: 0.987 (very high), indicating a strong positive linear relationship.
- Linear Model Results:

$$\text{Epistemic Uncertainty} = 0.0146 + 0.5574 \times \text{Aleatoric Uncertainty}$$

- Intercept (const): 0.0146
- Slope: 0.5574, meaning that for every unit increase in aleatoric uncertainty, epistemic uncertainty increases by 0.56.
- R^2 Value: 0.975, meaning that 97.5% of the variance in epistemic uncertainty is explained by aleatoric uncertainty.
- p-value: Extremely small ($7.27 \times 10^{\{-24\}}$), indicating the relationship is statistically significant.

This confirms a strong linear relationship between the two types of uncertainty.

2.3.7 Accuracy Heatmap



2.3.8 OOD Detection Results

Let the Occluded cifar-10 be the OOD dataset and let model recognize them. Lower value means the data is more like the uncorrupted version.