Weekly Study Report

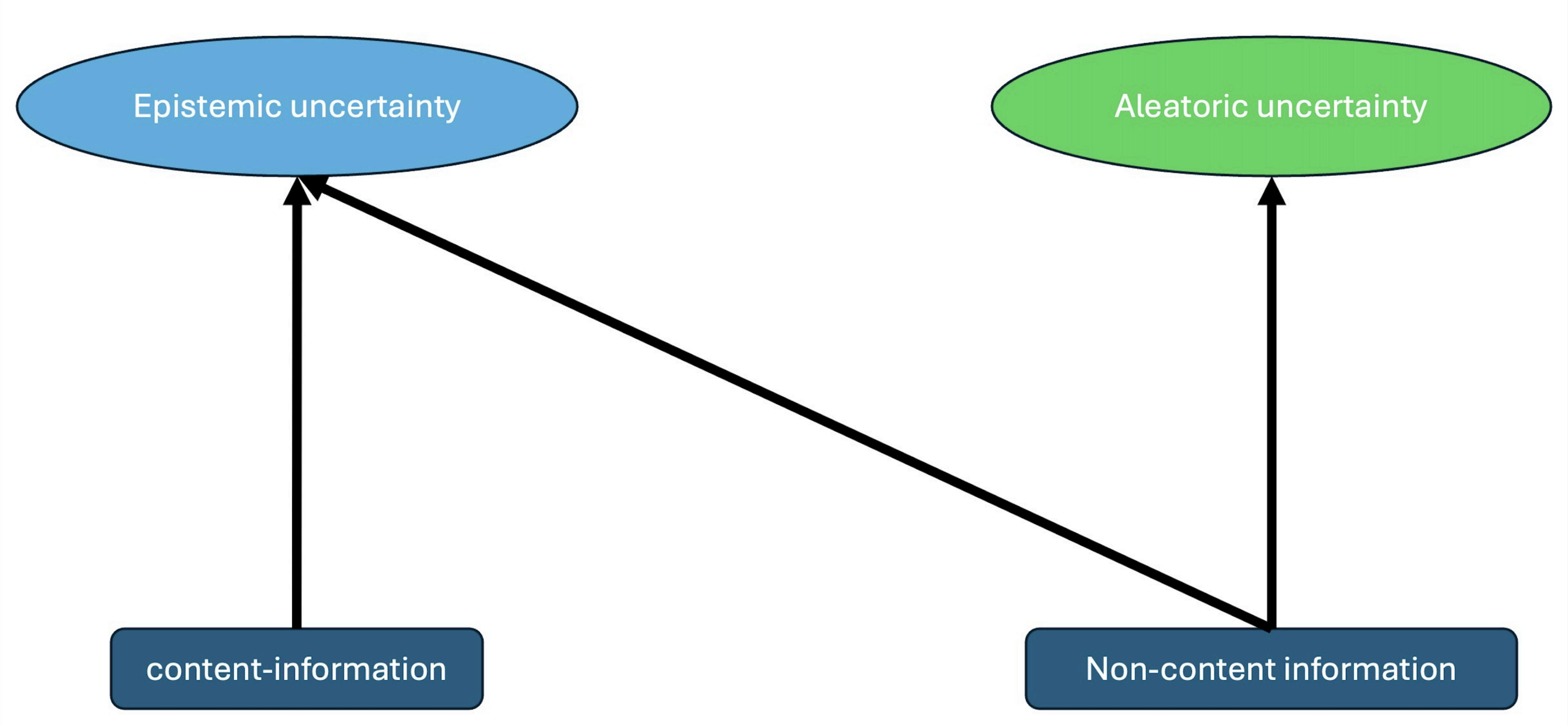
Wang Ma

2025-04-08

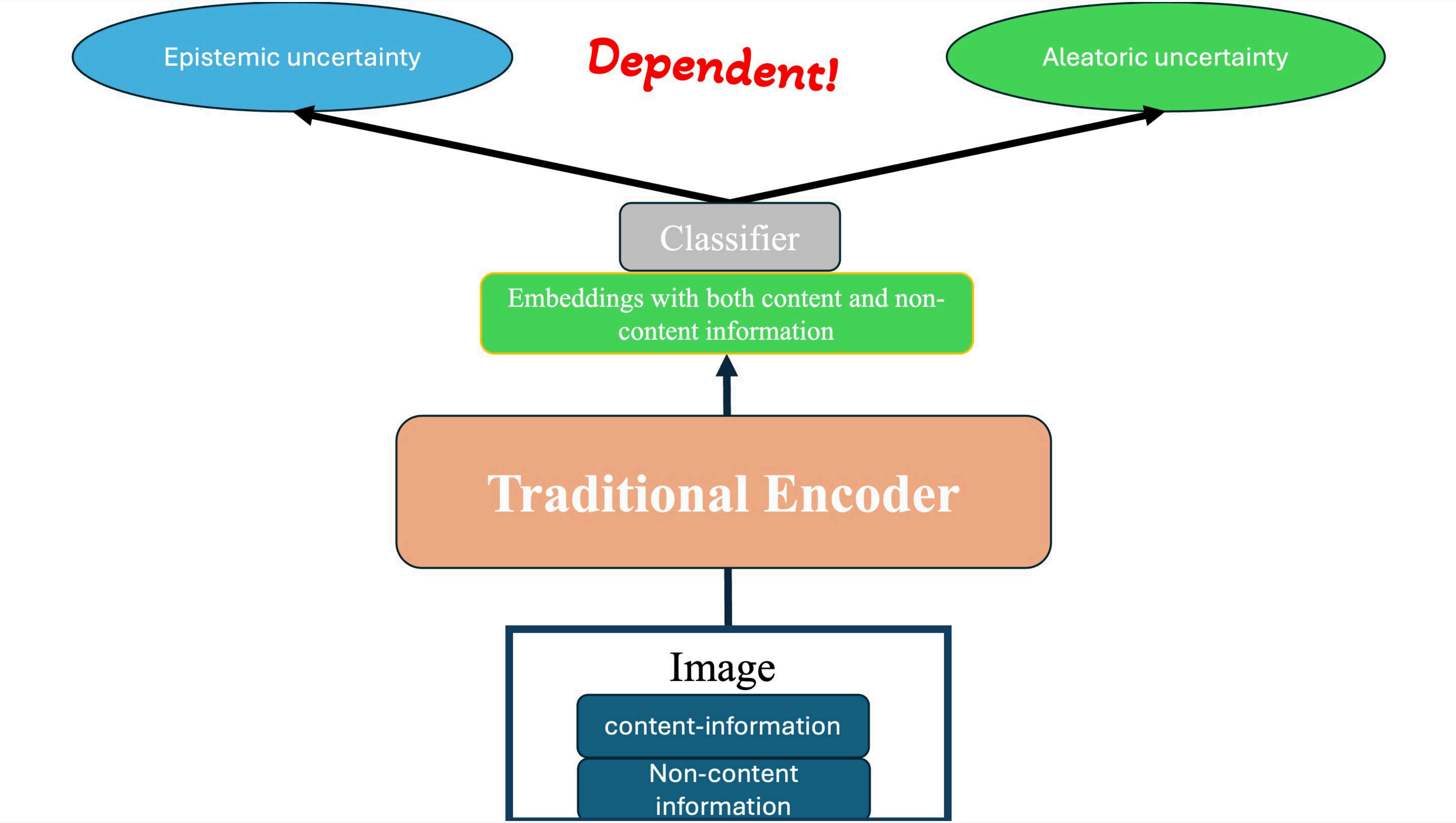
Electrical, Computer, and Systems Engineering Department
Rensselaer Polytechnic Institute

1. Using Contrastive Learning to Extract the content-related Epistemic Uncertainty

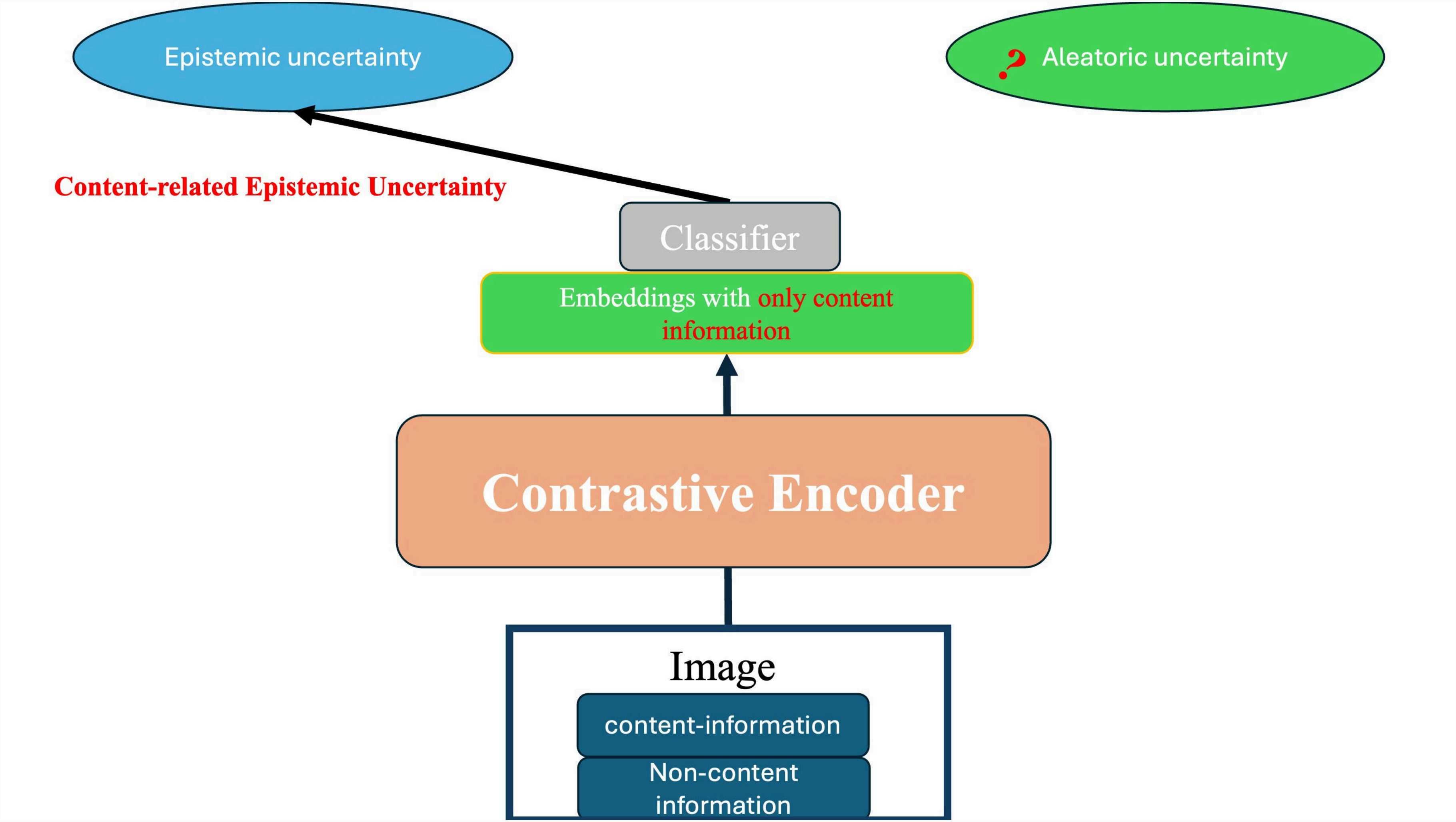
1.1 Background



1.1 Background



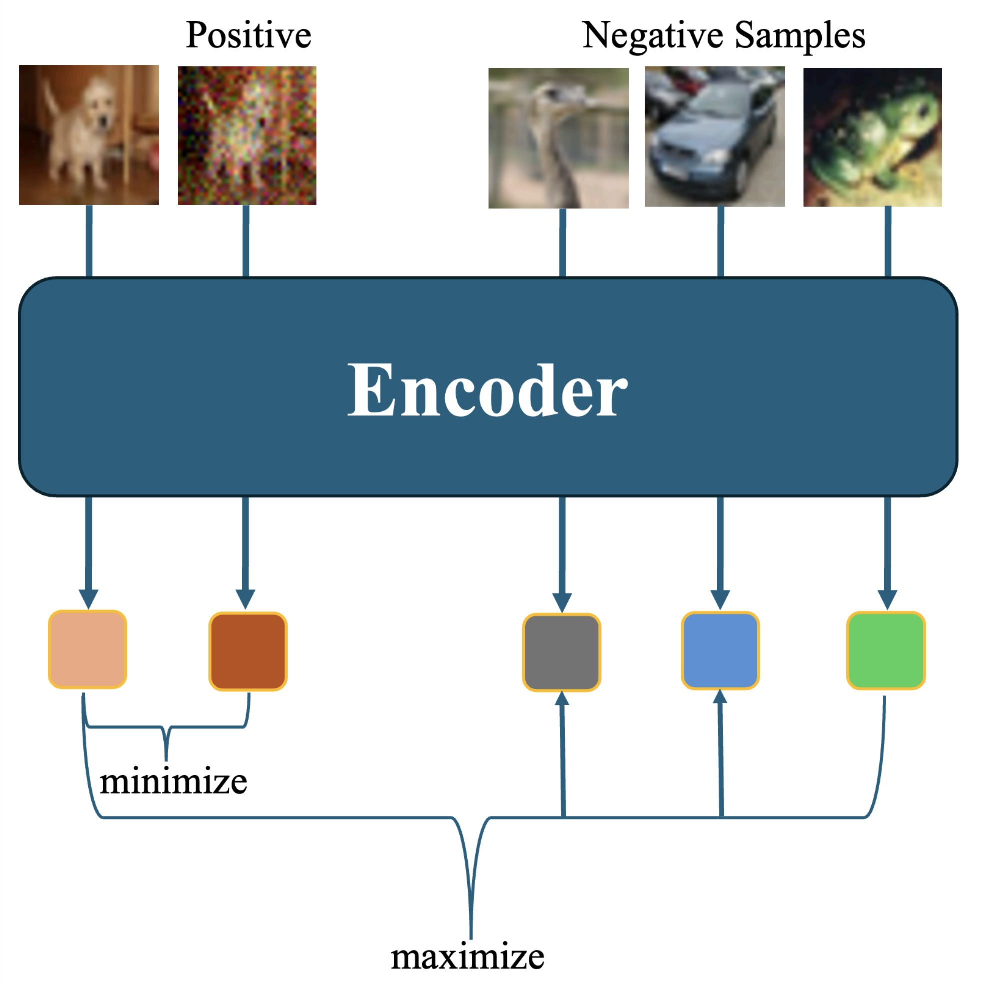
1.1 Background



GOAL.

1. Training a contrastive encoder to learn consistent features/embeddings from high- and low- quality input.
2. We want to minimize the influence of non-content information to the epistemic uncertainty.
3. The final goal is to obtain True content-related Epistemic Uncertainty, which can be used to detect in-lier data when both AU and EU are high.

1.3 The Contrastive Learning Model (Encoder Learning)



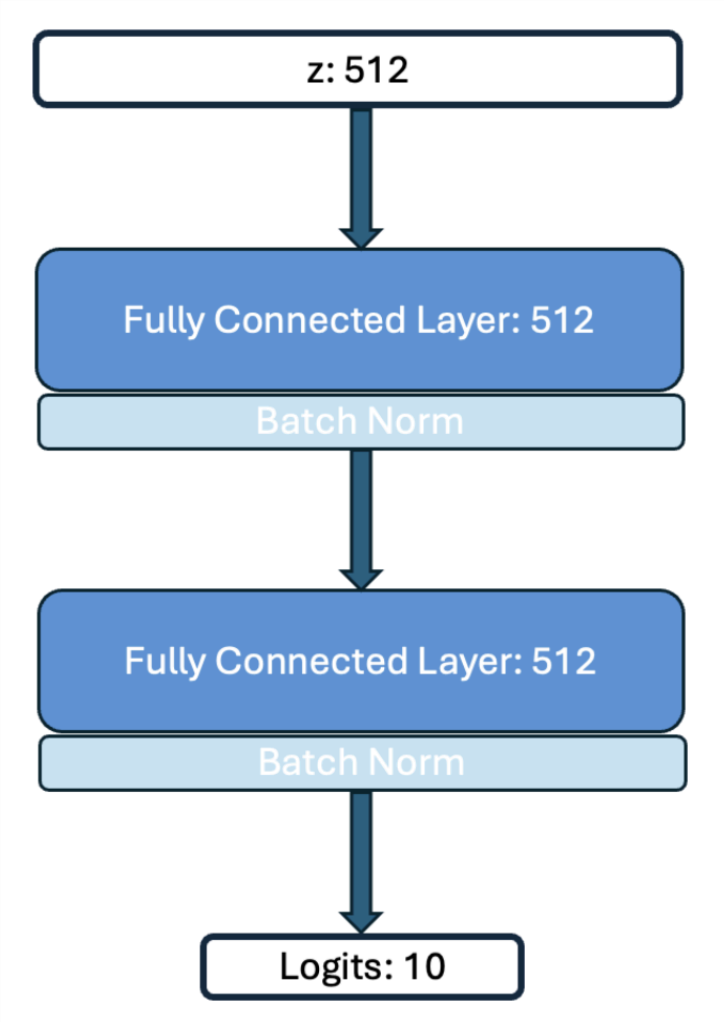
- Anchor: the clean (high-quality) image
- Positive Samples: the corrupted image
- Negative Samples: other images in batch

Output of the Encoder: the embedding z

The loss function:

$$L_i = -\log \frac{\exp(z_i z_i^+)}{\sum_{k,k \in z_i^-} \exp(z_i z_k)},$$

where $\|z\|^2 = 1$, so $z_i z_i^j$ is the cosine similarity of the two embeddings.



During the training of the Classifier head, we freeze the encoder part and solely update the parameters in the MLP head.

This training follows a standard Classification task training with cross entropy loss.

1.5 Design of the Experiments

- Training.
  1. Train a contrastive encoder, where the positive samples are corrupted images. (Original contrastive learning uses augmentations as the positive samples, and learns consistent embeddings from the augmentations.)
  2. Add a classification head to the encoder to get prediction results and do uncertainty quantification. (Here we train ensemble models of size 10)
- Testing. We want to test on
  1. Original Clean high-quality images (get uncertainty results)
  2. Generated low-quality images
     - Corrupted images with same corruption types and severity as training (get uncertainty results)
     - Corrupted images with different corruption types
  3. OOD data (SVHN)

1.6 Expectation of the Test Results

1. Original clean high-quality images

   Low uncertainties
2. Generated low-quality images

   - Corrupted images with same corruption types and severity as training. (Severity: 2)

     Low uncertainties as the clean images, hard to use uncertainty to separate from clean images (they are not ood, so can not be separated)

   - **Corrupted images with different corruption types**. (Severity: 4)

     Low uncertainties as before, hard to be separated. Our goal is that the contrastive encoder can learn consistent embeddings from both high- and low- quality data, the experimental results on this setting is the most important.

   - OOD data (SVHN)

     High uncertainty, can be easily separated.

1.7.1 Uncertainty Quantification Performance

Contrastive Leared Encoder		Clean_id	Corrupted_trained	Corrupted_not_trained	OOD
Total Uncertainty	mean	0.6045	0.708	0.8869	1.2616
	std	0.5204	0.5294	0.5272	0.356
Aleatoric Uncertainty	mean	0.5138	0.6094	0.7513	1.079
	std	0.444	0.4582	0.4488	0.2997
Epistemic Uncertainty	mean	0.0908	0.0986	0.1357	0.1826
	std	0.0924	0.0917	0.106	0.1007

ResNet18 Results		Clean_id	Corrupted_trained	Corrupted_not_trained	OOD
Total Uncertainty	mean	0.7137	0.8148	0.9463	0.9289
	std	0.5042	0.5082	0.5175	0.3823
Aleatoric Uncertainty	mean	0.3273	0.3854	0.4602	0.4641
	std	0.2491	0.2521	0.2657	0.2081
Epistemic Uncertainty	mean	0.3863	0.4294	0.4861	0.4648
	std	0.2979	0.2888	0.2958	0.2352

1.7 Current Results

Test Accuracy for the two models:

- Contrastive Encoder: 82%
- Resnet18: 88%

1. The Contrastive Learned Encoder seems to disentangle AU and EU from the strong Linear Relationship.
   - AU increases as the corruption severity increases
   - EU is consistent for Clean Data and Corrupted_id data
2. For resnet-18, the measured AU and EU are in strong linear relationship. And it seems the model fails to identify SVHN (?)

1.7.2 Detecting OOD and Low-quality Data

Contrastive Leared Encoder		Corrupted_trained	Corrupted_not_trained	OOD
Total Uncertainty	AUROC	0.5599	0.6528	0.835
	AUPR	0.5453	0.6515	0.9001
Aleatoric Uncertainty	AUROC	0.5633	0.6515	0.8376
	AUPR	0.5516	0.6217	0.9
Epistemic Uncertainty	AUROC	0.5367	0.6376	0.7679
	AUPR	0.519	0.6155	0.8697

Resnet18 Classifier		Corrupted_trained	Corrupted_not_trained	OOD
Total Uncertainty	AUROC	0.5576	0.6273	**0.6302**
	AUPR	0.5488	0.6233	**0.7686**
Aleatoric Uncertainty	AUROC	0.5664	0.6438	**0.6682**
	AUPR	0.5607	0.6422	**0.808**
Epistemic Uncertainty	AUROC	0.5427	0.5981	**0.5923**
	AUPR	0.5333	0.5752	**0.743**

1.7 Current Results

Test Accuracy for the two models:

- Contrastive Encoder: 82%
- Resnet18: 88%

Conclusion:

1. The Contrastive Learned Encoder is consistent to Corrupted_id data. And for Corrupted_ood data, the AUROC is low meaning the model does not really see difference in highly-corrupted data.
2. Meanwhile, the model can identify true OOD data well, even the test accuracy is not good. But it seems the EU does not work well on SVHN detection.
3. The Resnet-18 almost failed on the OOD detection, which means it cannot capture significant features. This means I need to check the model training and retrain the resnet18 models.

1.8 TO DO

1. Adjust the training of Resnet18 way to get a more correct baseline.
2. Try to add label information into the contrastive encoder learning
3. Since the training of contrastive learning is really expensive, use Last-layer laplace method to quantify uncertainty and see the results.