OMADA: On-manifold Adversarial Data Augmentation Improves Uncertainty Calibration

Kanil Patel[1,2], William Beluch[1], Dan Zhang[1], Michael Pfeiffer[1], Bin Yang[2]

[1]Bosch Center For Artificial Intelligence
[2]University of Stuttgart, Institute of Signal Processing and System Theory

BOSCH

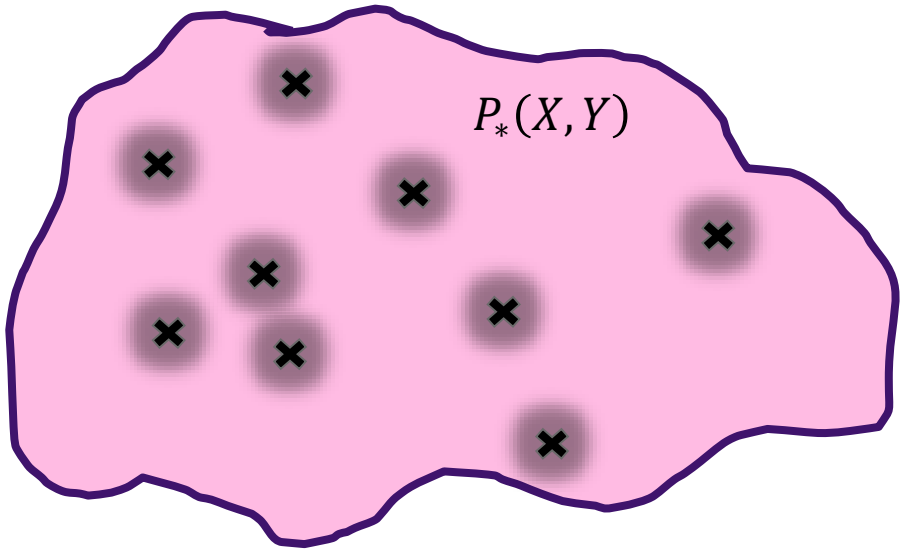
Motivation

- Deep learning predictions tend to be over-confident and mis-calibrated
- Improve DNN prediction confidences
  - Ultimate goal: Higher confidences for correct predictions and lower confidences for incorrect predictions
- Confidences should match up with the difficulty in predicting the samples
  - For a set of samples assigned p% confidence: accuracy should match the p%

Data Creation Process

- Unknown ground truth data distribution: $P_*(X, Y)$
- Sample dataset $\mathcal{D} = \{x_i, y_i\}_{i=1}^N$
- Empirical distribution is formed by assembling delta functions located on each example
- Replace the delta function by some estimate of the density in the vicinity (e.g. data augmentation)
- Target labels are hard (i.e. $\boldsymbol{y} \epsilon \{\boldsymbol{y} : \ \boldsymbol{y} \epsilon \{0,1\}^c, 1^T \boldsymbol{y} = 1\}$)
- Labels are sampled from $P_{\mathcal{D}}(y \mid x) = \text{argmax}_{\hat{y}} \, P_*(\hat{y} \mid x)$

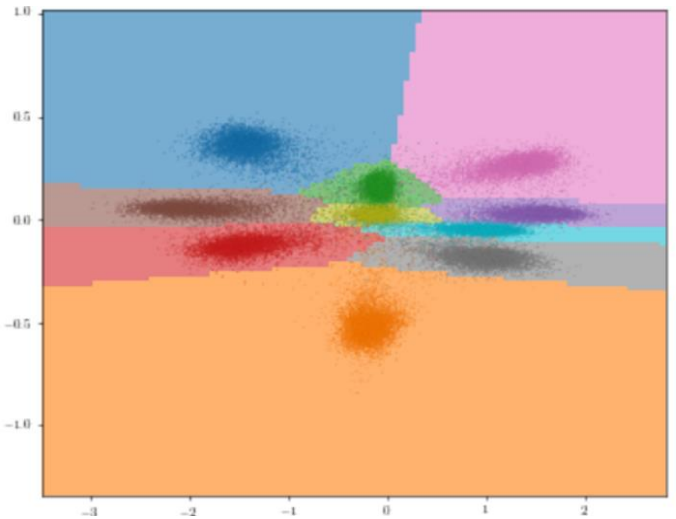
Image Source: Bright Side



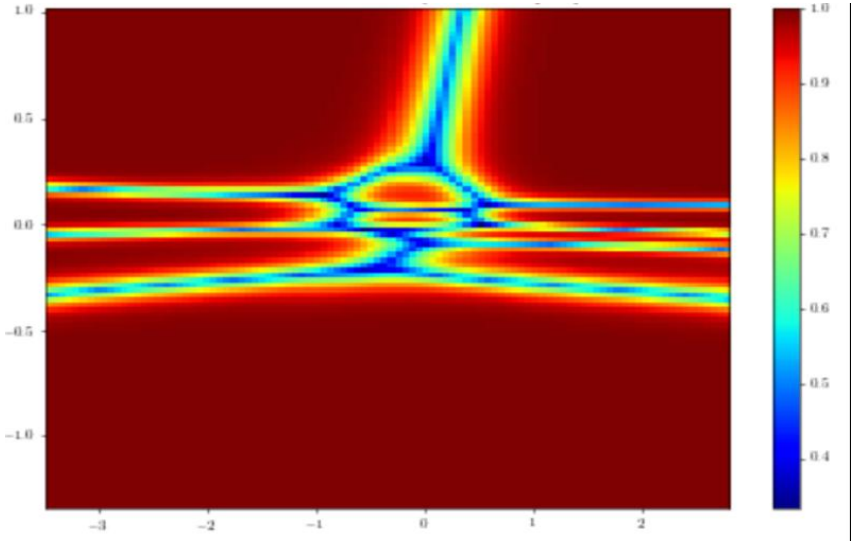
Cross-Entropy Loss with Hard Labels Can Encourage Over-Confidence

- One-hot vector forces network to be over confident
- Penalizes network for showing any level of uncertainty

Class Predictions



Max Confidence



- **Possible Solution:**

$X \sim P_*(X)$ (ambiguous samples)
$Y \sim P_*(Y \mid X)$ (soft label capturing uncertainty)

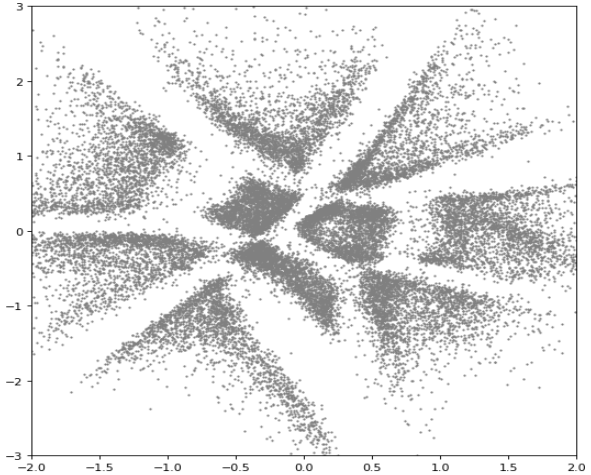
BOSCH

On-Manifold Data Augmentation

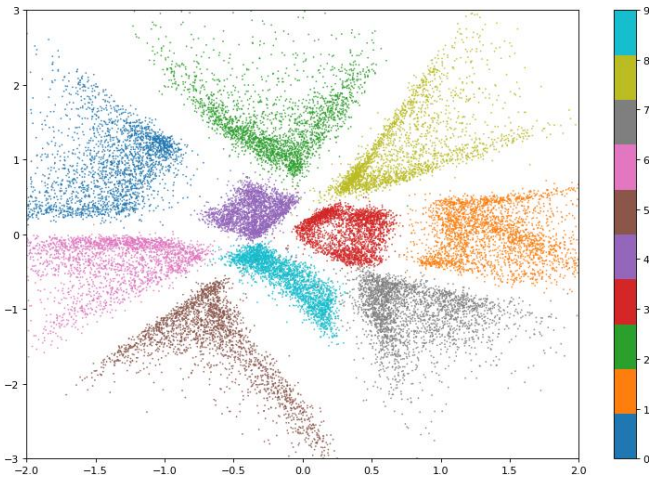
- Approx. Manifold: Encoder-Decoder Generative Model
- Semi-Supervised Training
- Interested in boundary regions constrained on manifold
- Inefficient to randomly sample these boundary regions on manifold

$X \sim P_*(X)$ (ambiguous samples)
$Y \sim P_*(Y \mid X)$ (label capturing uncertainty)

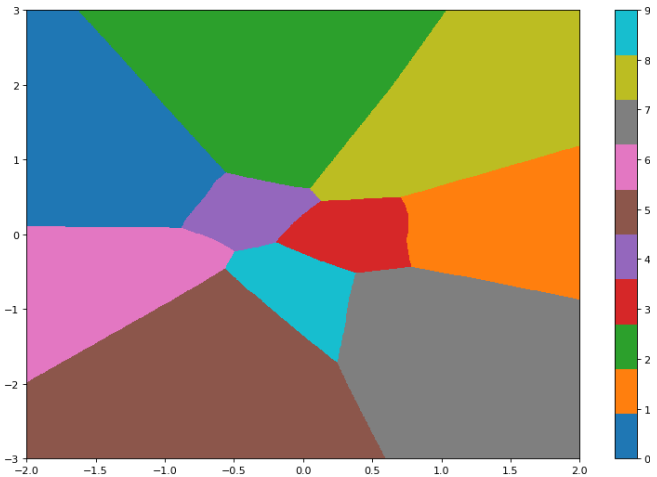
(Unsupervised) Manifold



(Semi-supervised) Manifold



Predictions Map on Manifold



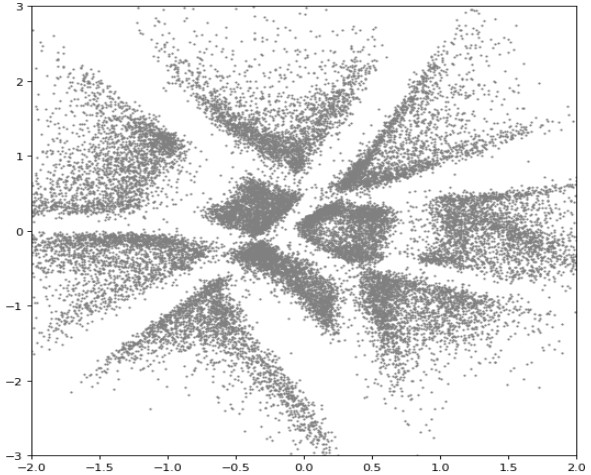
BOSCH

On-Manifold Data Augmentation

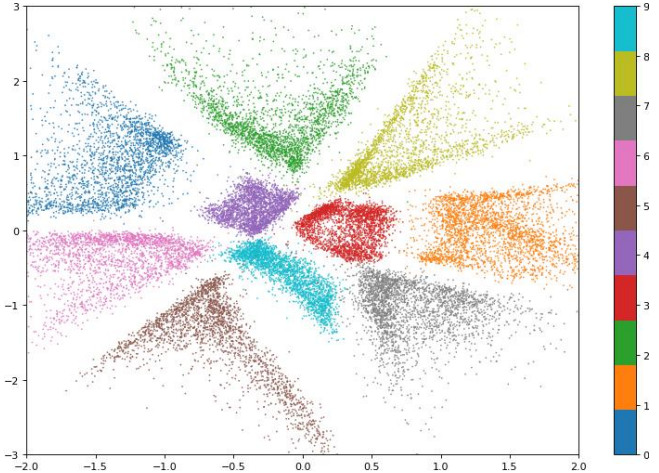
- Approx. Manifold: Encoder-Decoder Generative Model
- Semi-Supervised Training
- Interested in boundary regions constrained on manifold
- Inefficient to randomly sample these boundary regions on manifold

$X \sim P_*(X)$ (ambiguous samples)
$Y \sim P_*(Y \mid X)$ (label capturing uncertainty)

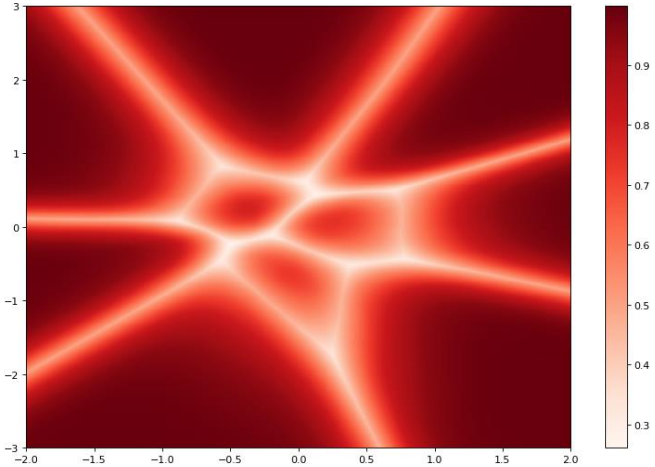
(Unsupervised) Manifold



(Semi-supervised) Manifold



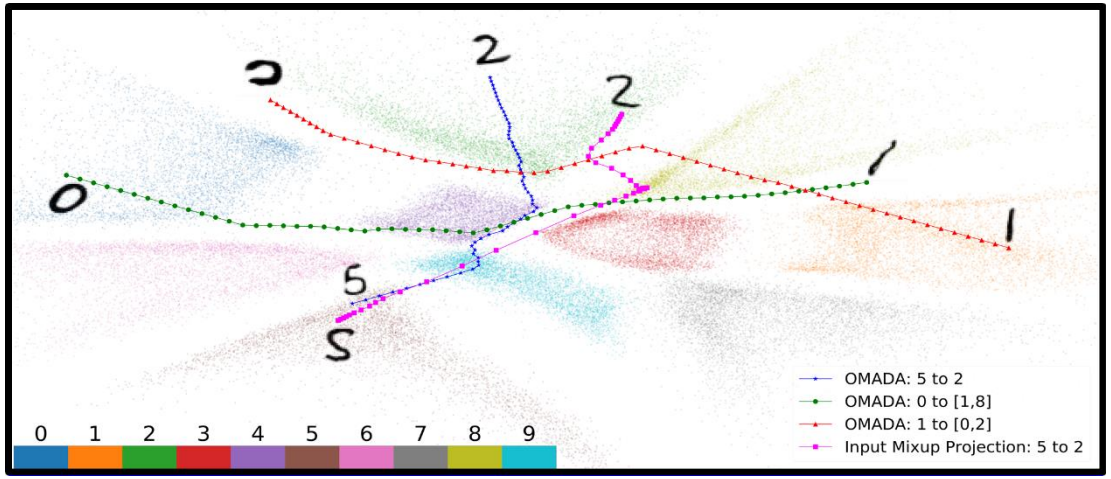
Confidence Map on Manifold

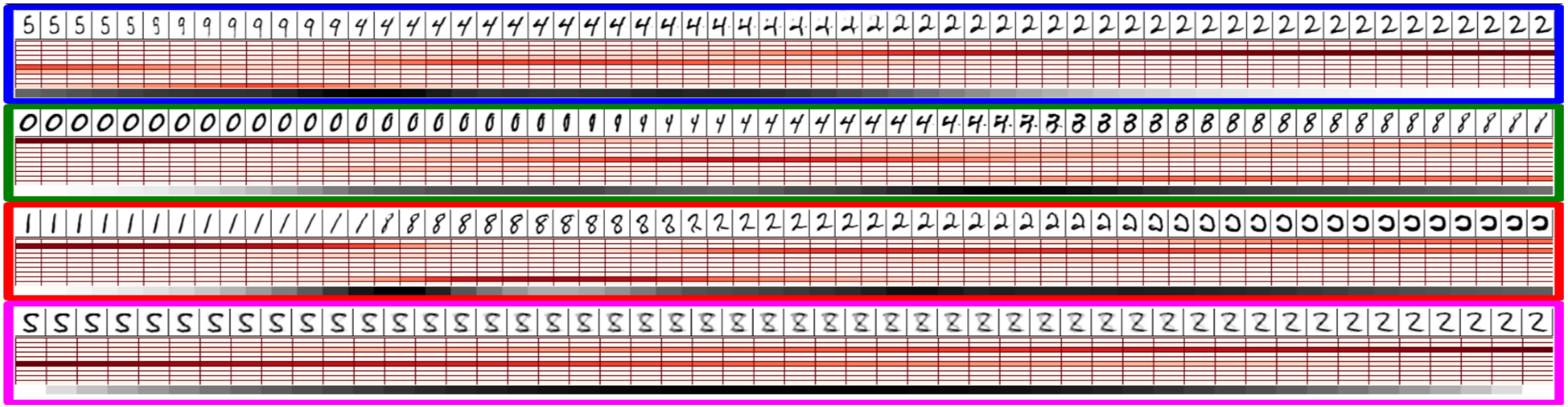


BOSCH

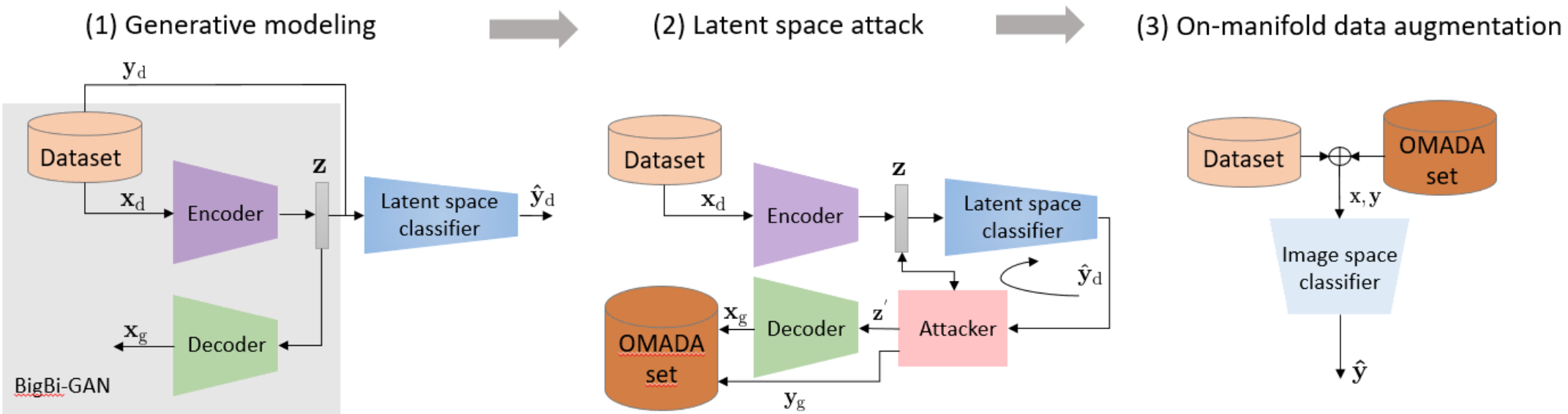
On-Manifold Adversarial Attacks

$X \sim P_*(X)$ (ambiguous samples)
$Y \sim P_*(Y \mid X)$ (label capturing uncertainty)





OMADA – On Manifold Adversarial Data Augmentation



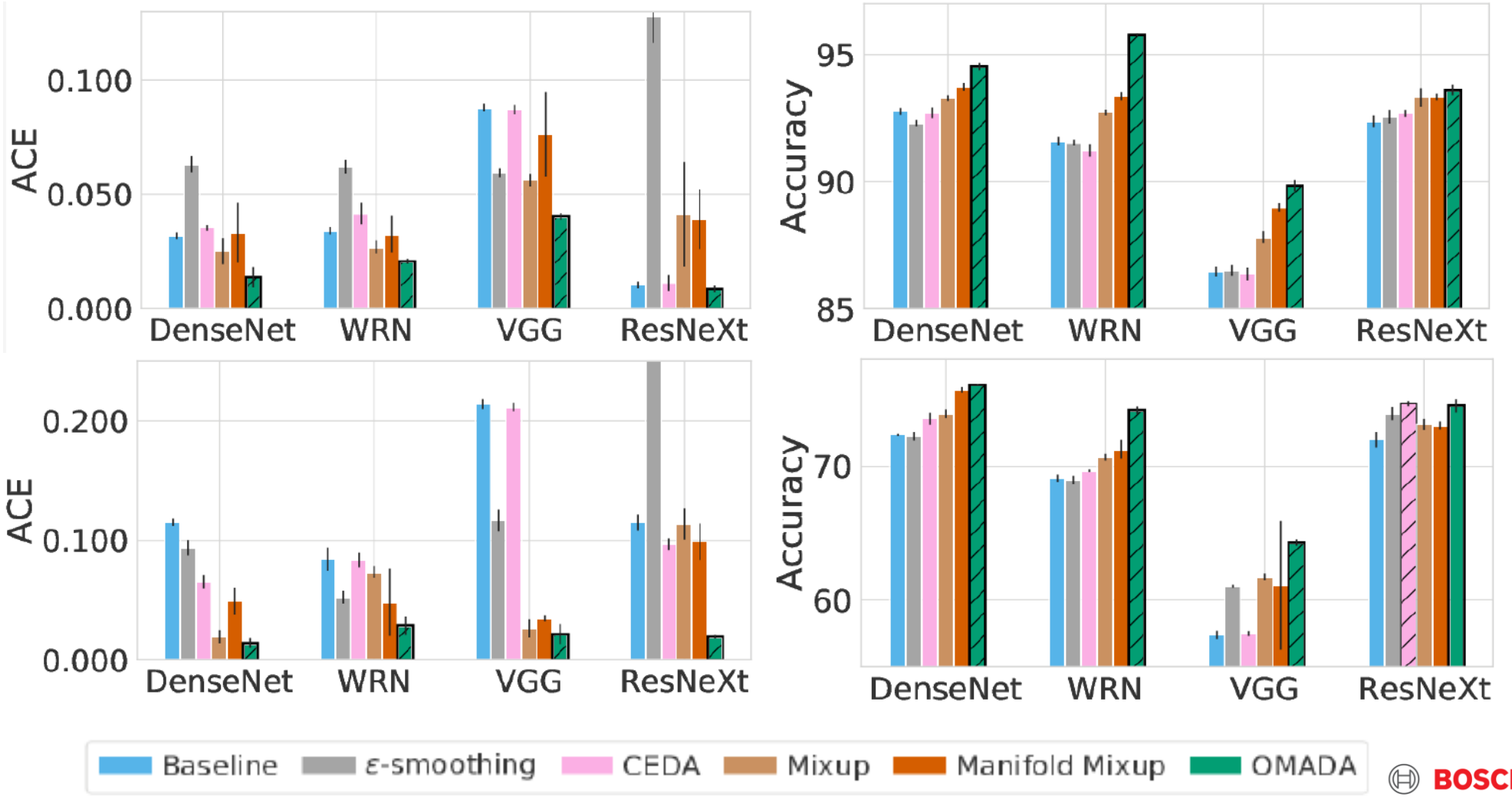
Experimental Setup

- Datasets: CIFAR-10, CIFAR-100 and SVHN
- Models: DenseNet (L=100, k=12), WRN-28-10, VGG-16, ResNext-29
- Compare against multiple related methods
- Evaluation Metrics:
  - Accuracy
  - Calibration Error (ACE)
  - Sparsification
  - Outlier detection performance (AUC)
  - Outlier MMC (Mean Max. Confidence)

Calibration Error and Accuracy

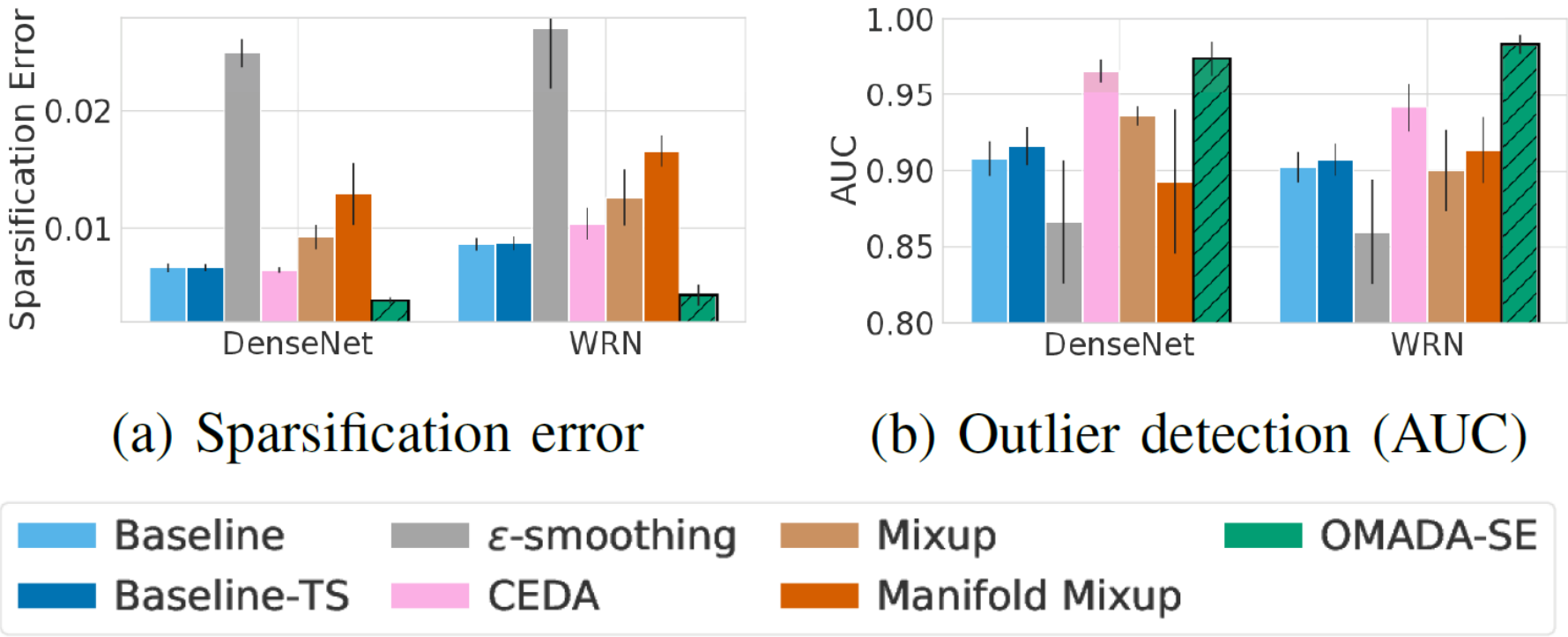
CIFAR-10

CIFAR-100



BOSCH

Sparsification Error and Outlier Detection



Conclusion

- Introduced concept of **on-manifold adversarial data augmentation** for uncertainty estimation
- Leverage recent advances in **generative modeling**
- Latent space classifier **to approximate decision boundaries**
- **Adversarial attack** on latent space for sampling ambiguous samples
- Show improvements on accuracy and calibration performance

Thank you