Latent Outlier Exposure for Anomaly Detection with Contaminated Data

Chen Qiu*[1,2], Aodong Li*[3], Marius Kloft[2], Maja Rudolph[1], Stephan Mandt[3]

1 BOSCH 2 TECHNISCHE UNIVERSITÄT KAISERSLAUTERN 3 UCIRVINE

Motivation & Problem Setup

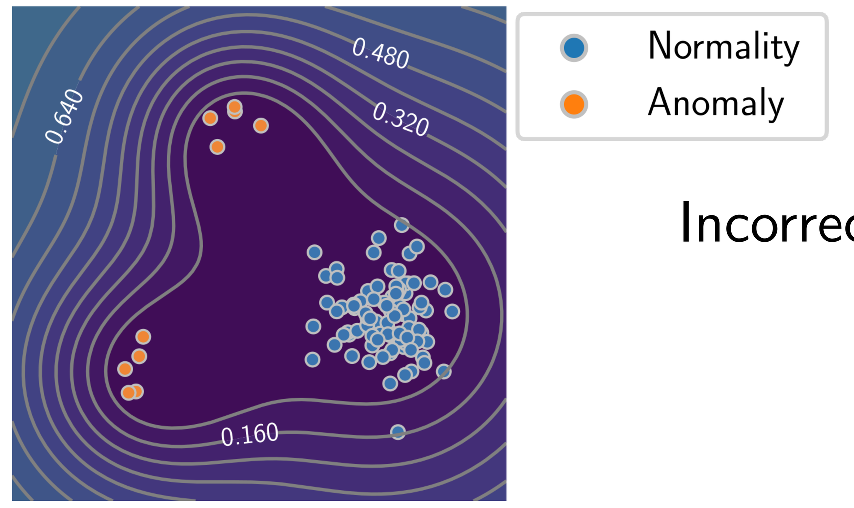
Anomaly Detection with Contaminated Training Data.

fraud transaction

Image taken from https://kunal3836.medium.com/fraud-detection-in-payments-db6d5fc89d13

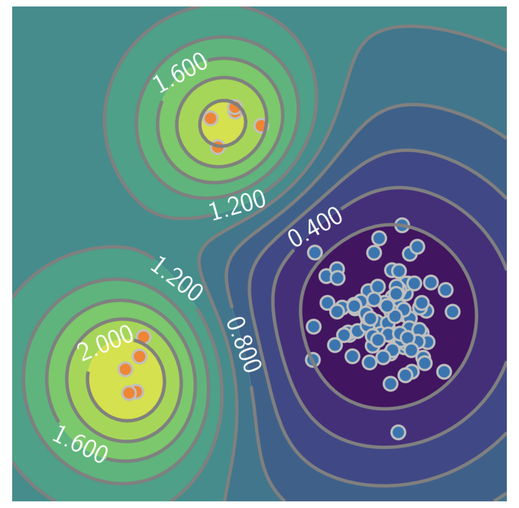
→ Common assumption: **clean** training data.
→ What if the training data contains unnoticed anomalies?

Anomaly score in input space



Incorrect normal region characterization.

→ A solution: exploit labels.



Supervised learning characterizes boundaries well.

→ However, labels are expensive. Can we have a cheaper way?
→ **Contribution:** Unsupervised latent outlier exposure.

Problem Setup.

→ Training sets contain many normal samples and a few anomalies.

Method: Latent Outlier Exposure

Proposed Loss.

$$\mathcal{L}(\theta, \mathbf{y}) = \sum_{i=1}^{N} (1-y_i)\mathcal{L}_n^{\theta}(\mathbf{x}_i) + y_i \mathcal{L}_a^{\theta}(\mathbf{x}_i)$$

→ $\mathbf{y}$ are variables to be optimized.
→ $\mathcal{L}_n^{\theta}(\mathbf{x})$: a normal loss that is designed to be minimized over normal data.
→ $\mathcal{L}_a^{\theta}(\mathbf{x})$: an abnormal loss that is designed to have the opposite effect.
→ E.g., for deep SVDD, $\mathcal{L}_n^{\theta}(\mathbf{x}) = ||f_\theta(\mathbf{x}) - \mathbf{c}||^2$ and $\mathcal{L}_a^{\theta}(\mathbf{x}) = 1/||f_\theta(\mathbf{x}) - \mathbf{c}||^2$.

Constrained Optimization Problem. *Hard* LOE.

$$\min_\theta \min_{\mathbf{y}\in\mathcal{Y}} \mathcal{L}(\theta, \mathbf{y}) \quad \text{s.t. } \mathcal{Y} = \left\{\mathbf{y} \in \{0,1\}^N : \sum_{i=1}^{N} y_i = \alpha N\right\}$$

→ α is an assumed contamination ratio.
→ Block coordinate descent:
  ▷ Update θ when $\mathbf{y}$ is fixed;
  ▷ Update $\mathbf{y}$ when θ is fixed and the constraint is satisfied. (Closed-form solution exists.)

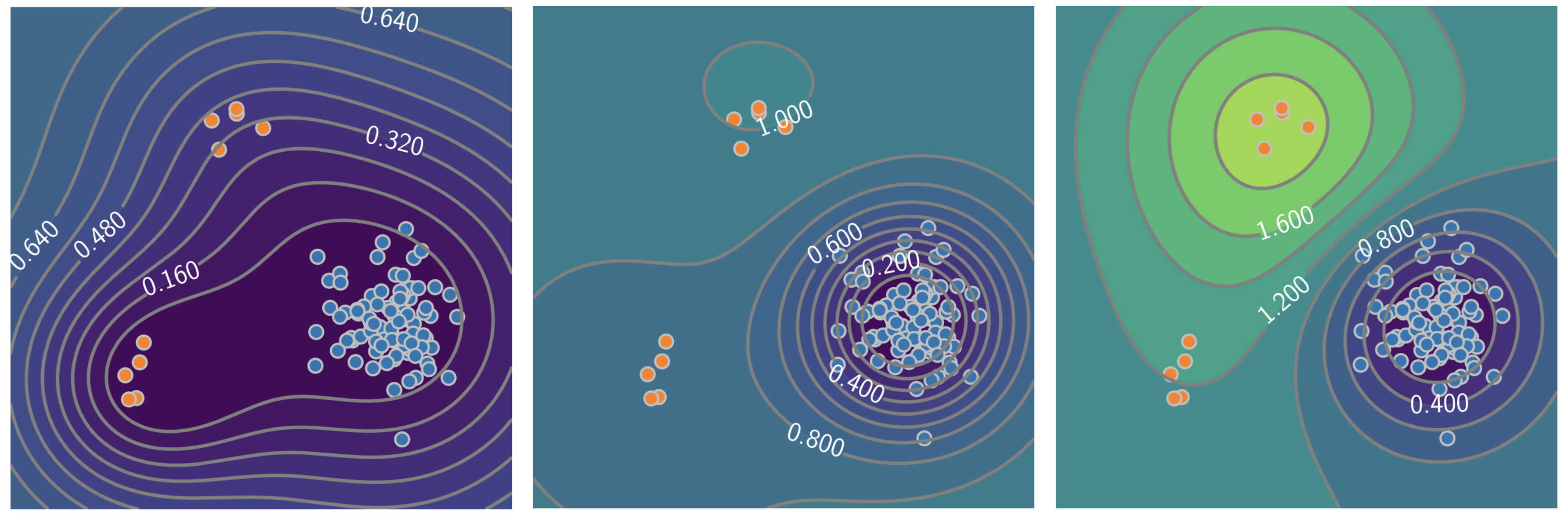
Model Extension. *Soft* LOE.

$$\min_\theta \min_{\mathbf{y}\in\mathcal{Y}'} \mathcal{L}(\theta, \mathbf{y}) \quad \text{s.t. } \mathcal{Y}' = \left\{\mathbf{y} \in \{0,0.5\}^N : \sum_{i=1}^{N} y_i = 0.5\alpha N\right\}$$

Anomaly Score.

$$S_i^{\text{test}} = \mathcal{L}_n^{\theta}(\mathbf{x}_i)$$

Experiments - Synthetic Data



Refine Soft LOE Hard LOE

Experiments - Real Data

Protocol.

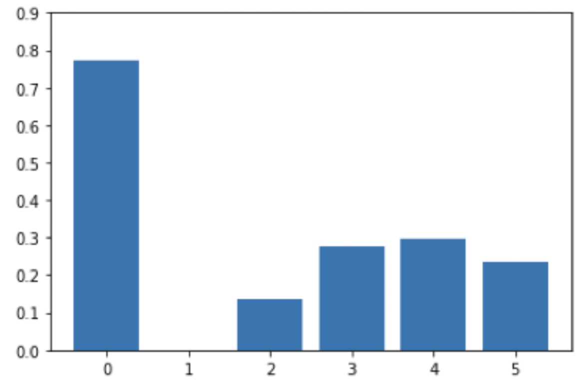
→ one vs. the rest.
→ Corruption of training set:
  ▷ Mix abnormal samples to have an anomaly ratio of α_0.

Data.

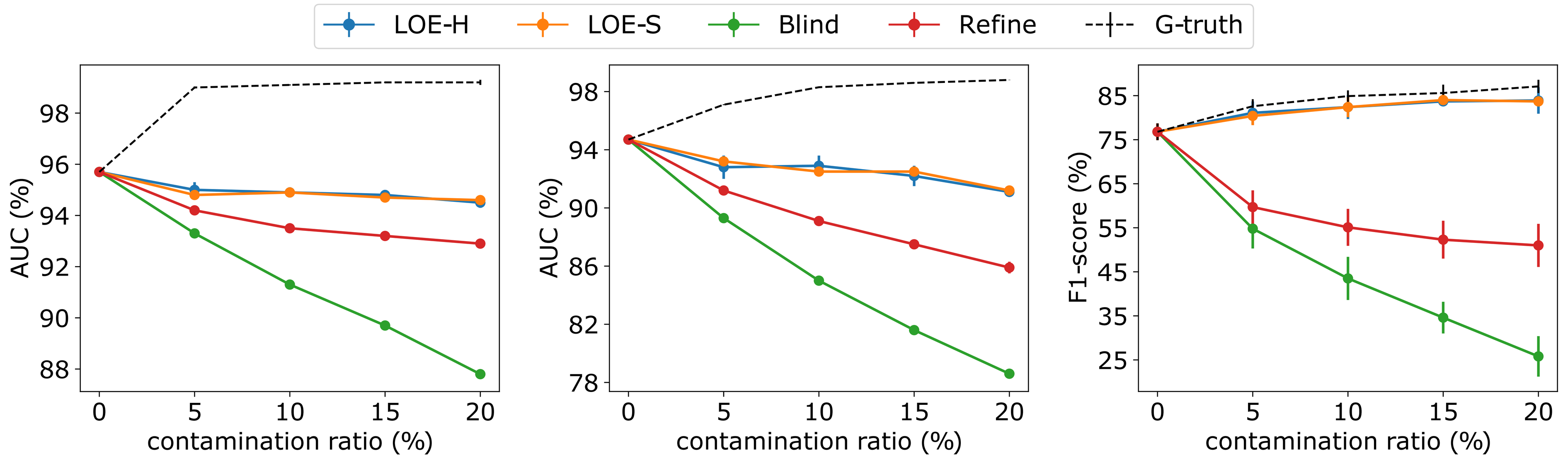
Image.

Video.

Tabular.



Results.

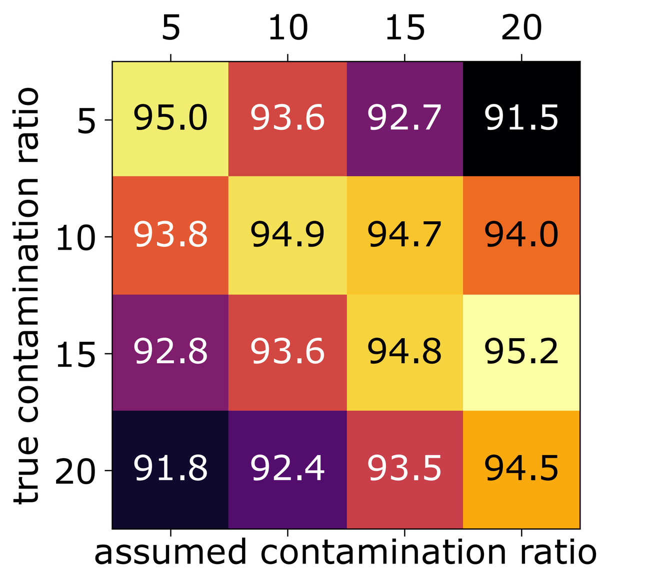


CIFAR-10 FMNIST Thyroid

Table. UCSD Peds1

Method	Contamination Ratio		
	10%	20%	30%*
(Tudor Ionescu et al., 2017)	-	-	68.4
(Liu et al., 2018)	-	-	69.0
(Del Giorno et al., 2016)	-	-	59.6
(Sugiyama & Borgwardt, 2013)	55.0	56.0	56.3
(Pang et al., 2020)	68.0	70.0	**71.7**
Blind	85.2±1.0	76.0±2.7	66.6±2.6
Refine	82.7±1.5	74.9±2.4	69.3±0.7
LOE$_H$ (ours)	82.3±1.6	59.6±3.8	56.8±9.5
LOE$_S$ (ours)	**86.8±1.2**	**79.2±1.3**	**71.5±2.4**

*Default setup in (Pang et al., 2020), corresponding to $\alpha_0 \approx 30\%$.



Sensitivity study: CIFAR-10