Finding Clarity in Chaos: Leveraging Noisy Labels for Superior Anomaly Detection

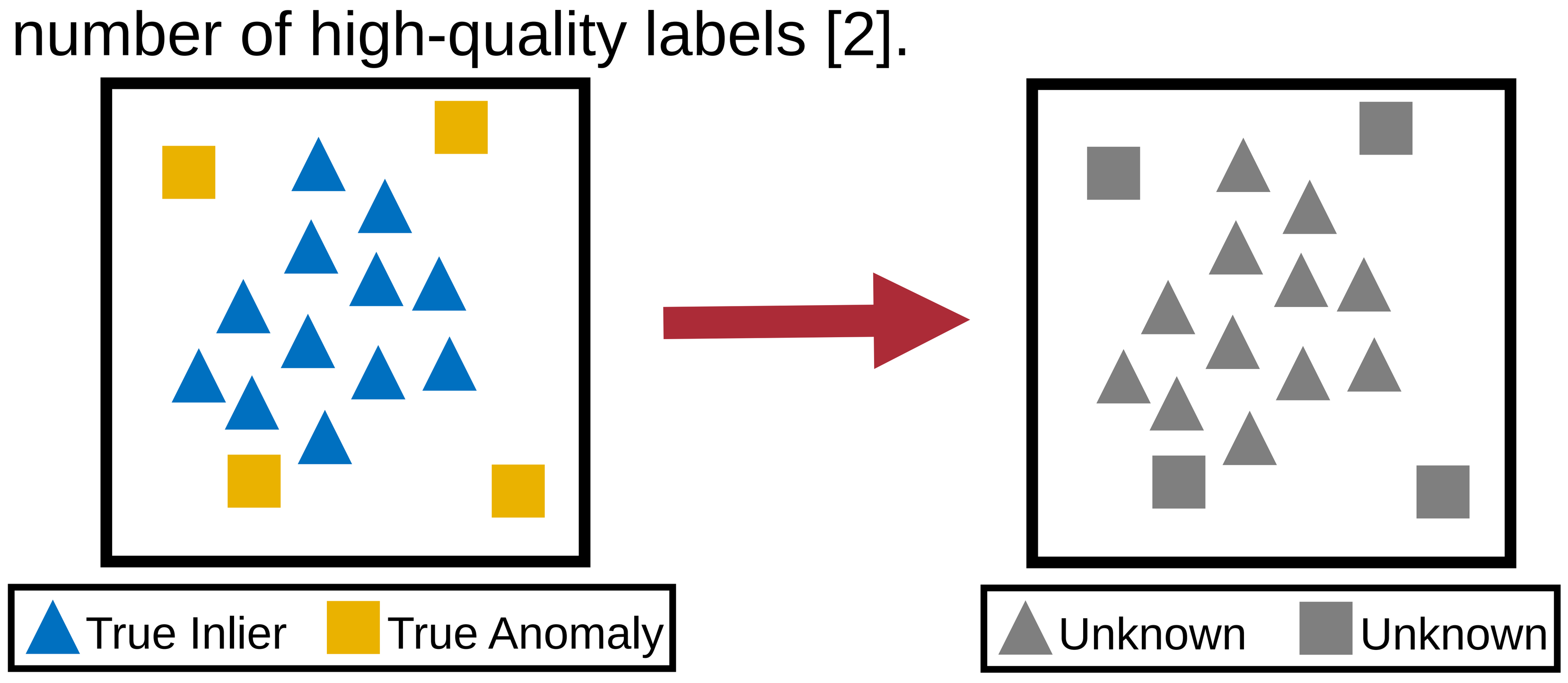
Dennis M. Hofmann, Peter M. VanNostrand, Lei Ma, Huayi Zhang, Joshua C. DeOliveira, Lei Cao, Elke A. Rundensteiner

Based on the paper “Agree to Disagree: Robust Anomaly Detection with Noisy Labels” [1] accepted at SIGMOD 2025.

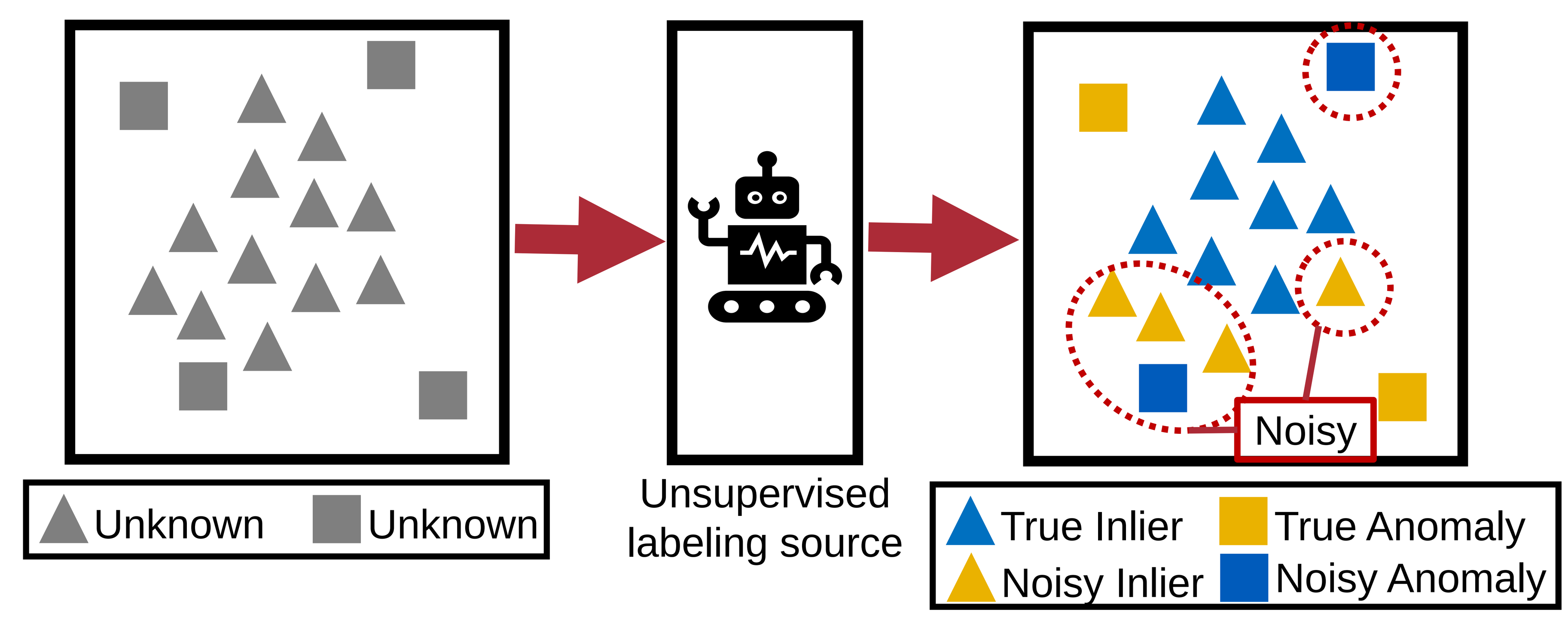
Background: Anomaly detection aims to identify objects that significantly differ from the expected behavior. Anomaly detection is critical for tasks such as detecting cybersecurity attacks, financial fraud, and life-threatening health conditions. [2]

1 Motivation

Since anomalies are rare, it is difficult to acquire large number of high-quality labels [2].

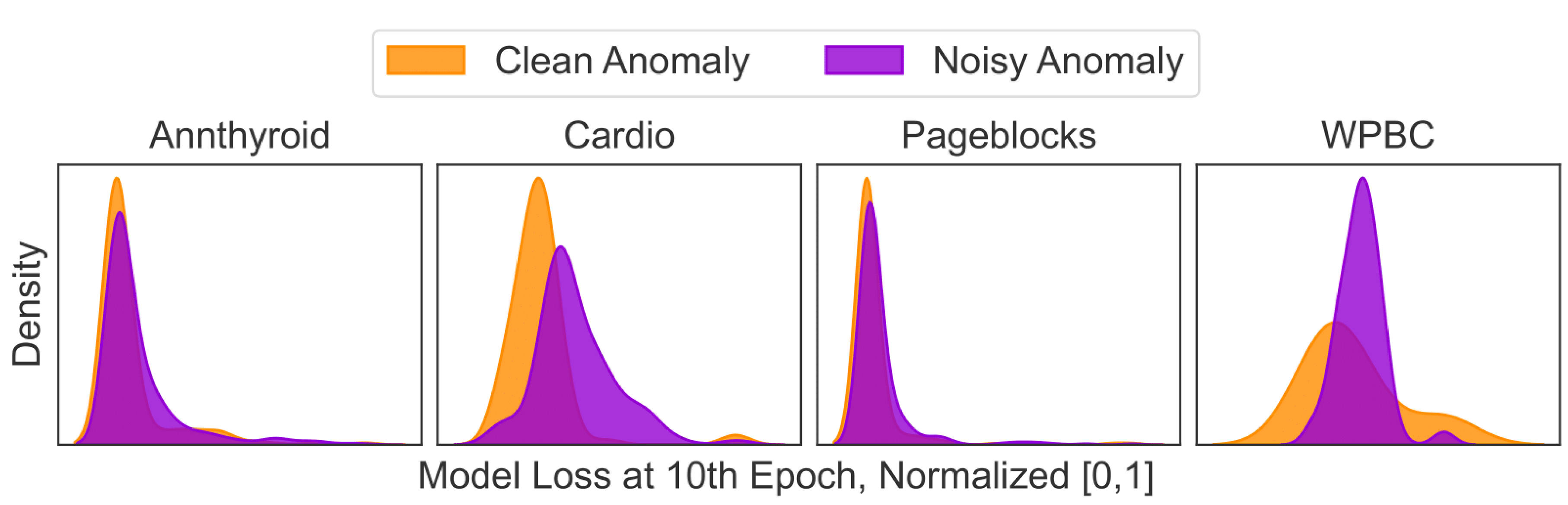


We can cheaply generate pseudo-labels.

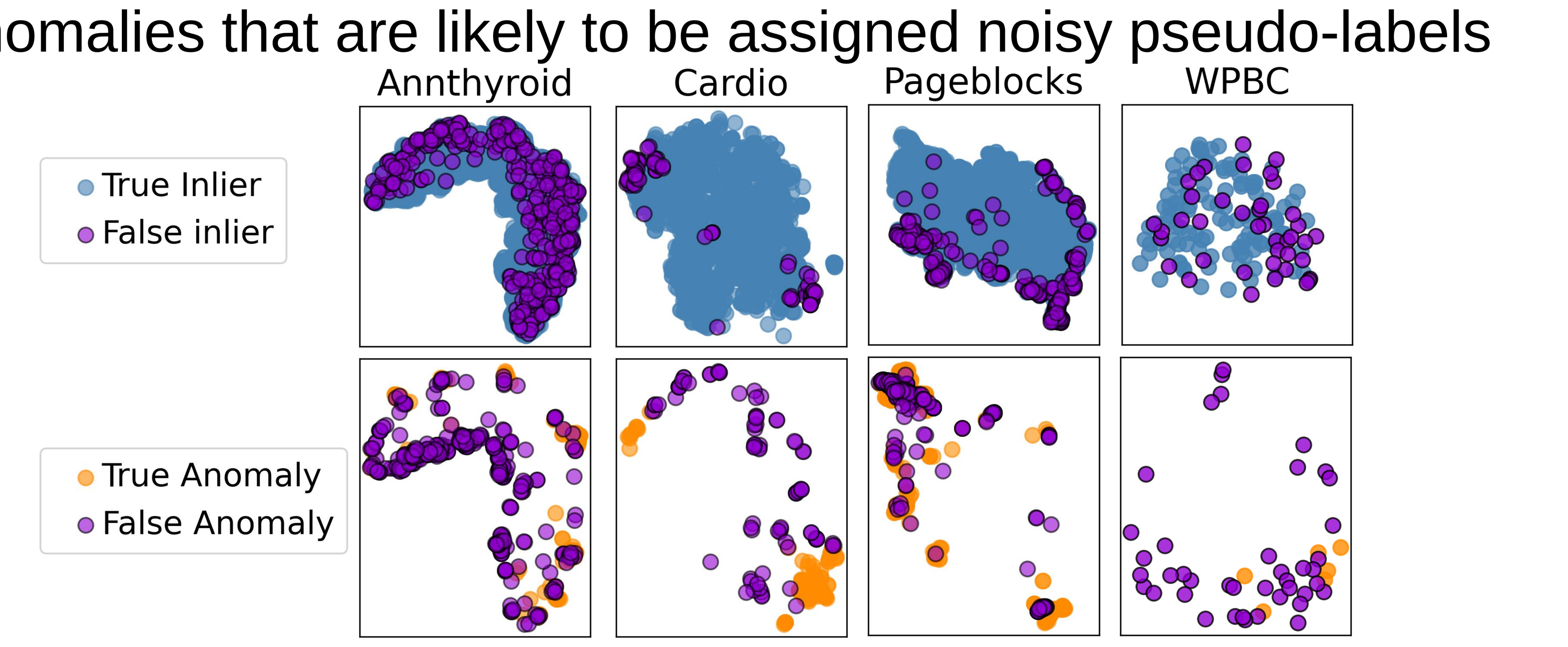


2 Challenges: Learning from Noisy Labels for Anomaly Detection

Challenge 1: The scarcity of anomalies invalidates common clean sample selection metrics (such as loss [3]).

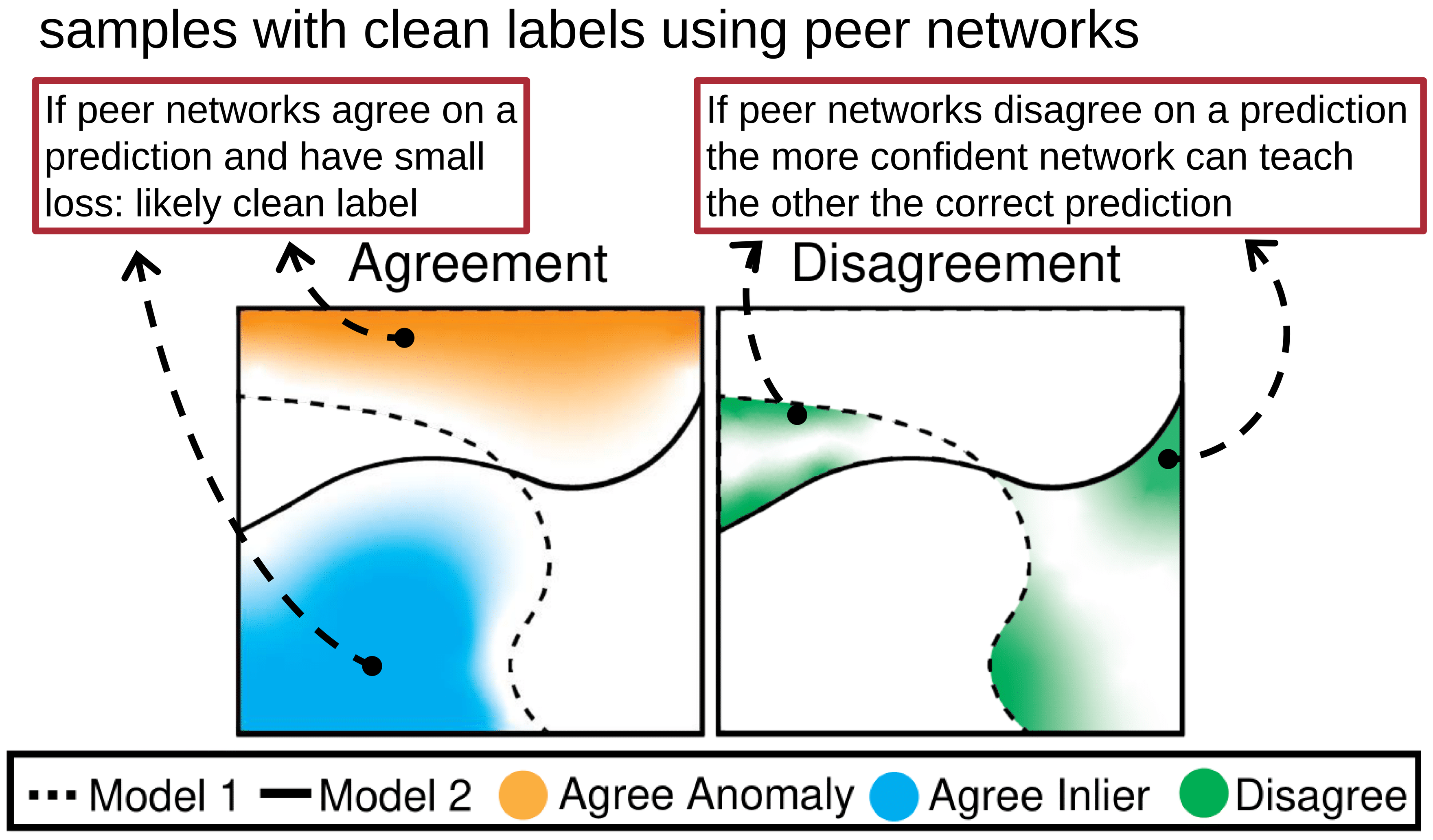


Challenge 2: The diversity of anomalies leads to many *marginal* anomalies that are likely to be assigned noisy pseudo-labels

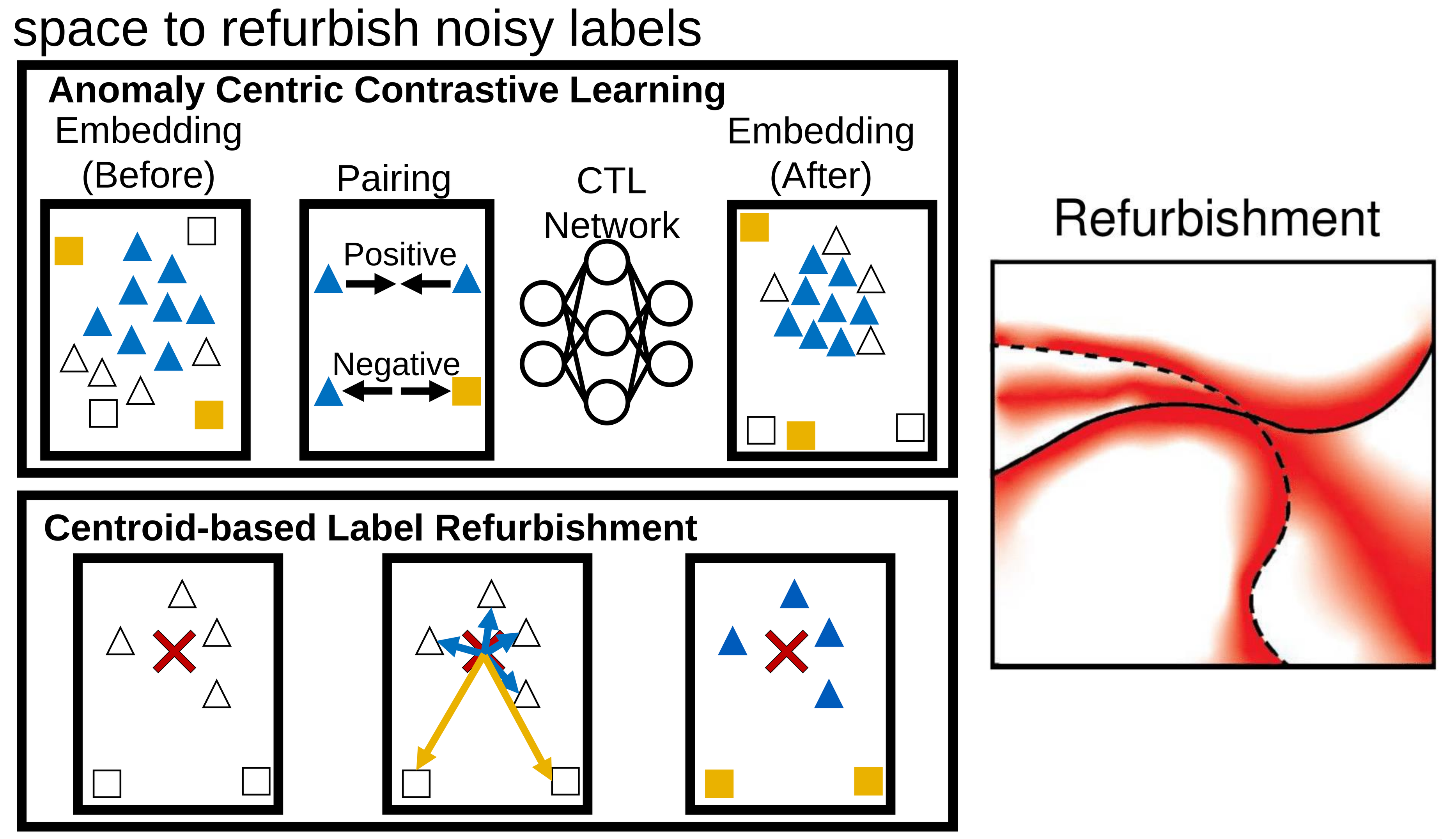


3 Solutions

Solution 1: Select a diverse set of easy and marginal samples with clean labels using peer networks

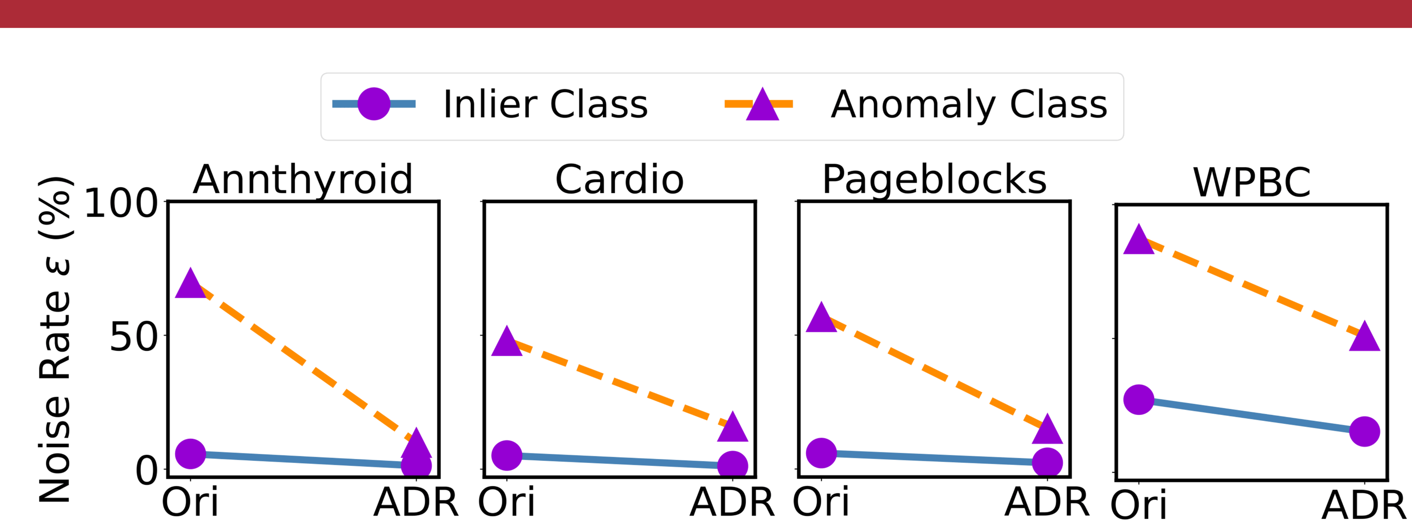


Solution 2: Learn a contrastive anomaly-aware embedding space to refurbish noisy labels



4 Experiments

Method	Datasets (F-1 score)			
	Annthyroid	Cardio	Pageblocks	WPBC
IF (Base)	0.34_{01}	0.54_{02}	0.44_{01}	0.17_{02}
AutoOD [4]	0.48_{04}	0.12_{04}	0.49_{06}	0.32_{03}
UADB [5]	0.39_{02}	0.61_{04}	0.25_{18}	0.17_{02}
UNITY (ours)	$\underline{0.78_{00}}$	$\mathbf{0.65_{09}}$	$\mathbf{0.63_{01}}$	$\mathbf{0.36_{02}}$



1. Unity outperforms SOTA LNL-AD Methods
2. Unity significantly decreases the noise rate

References

[1] D. Hofmann, et al. *Agree to Disagree: Robust Anomaly Detection with Noisy Labels.* SIGMOD 2025
[2] S. Han et al. *ADBench: Anomaly Detection Benchmark.* NeurIPS. 2022.
[3] H. Song, et al. Learning From Noisy Labels With Deep Neural Networks: A Survey. TNNLS 2023
[4] L. Cao et al. AutoOD: Automatic Outlier Detection. SIGMOD. 2023.
[5] H. Ye et al. UADB: Unsupervised Anomaly Detection Booster. ICDE. 2023.

Dennis M. Hofmann | PhD Student
Department of Data Science, Worcester Polytechnic Institute