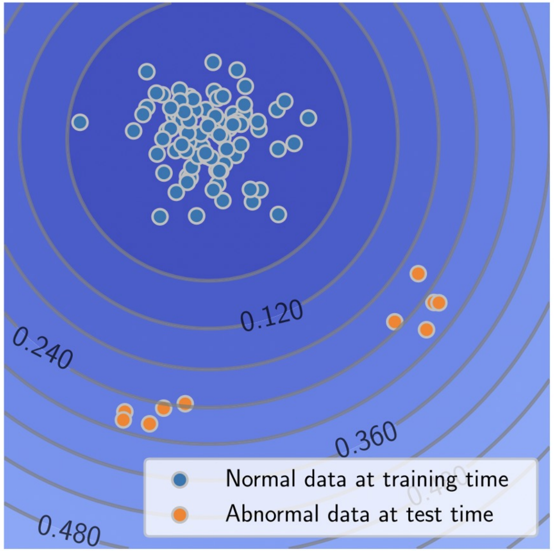
Zero-Shot Anomaly Detection via Batch Normalization

Aodong Li*, Chen Qiu*,
Marius Kloft, Padhraic Smyth, Maja Rudolph†, Stephan Mandt†

* Equal contribution † Joint Supervision Correspondence to: aodongl1@uci.edu

Anomaly detection (AD)

The task of **anomaly detection (AD)** is to train an anomaly detector $S_\theta(x) : \mathcal{X} \rightarrow \mathbb{R}$ to assign low anomaly scores to normal data and high scores to abnormal data.



Train with only normal data distribution P_s

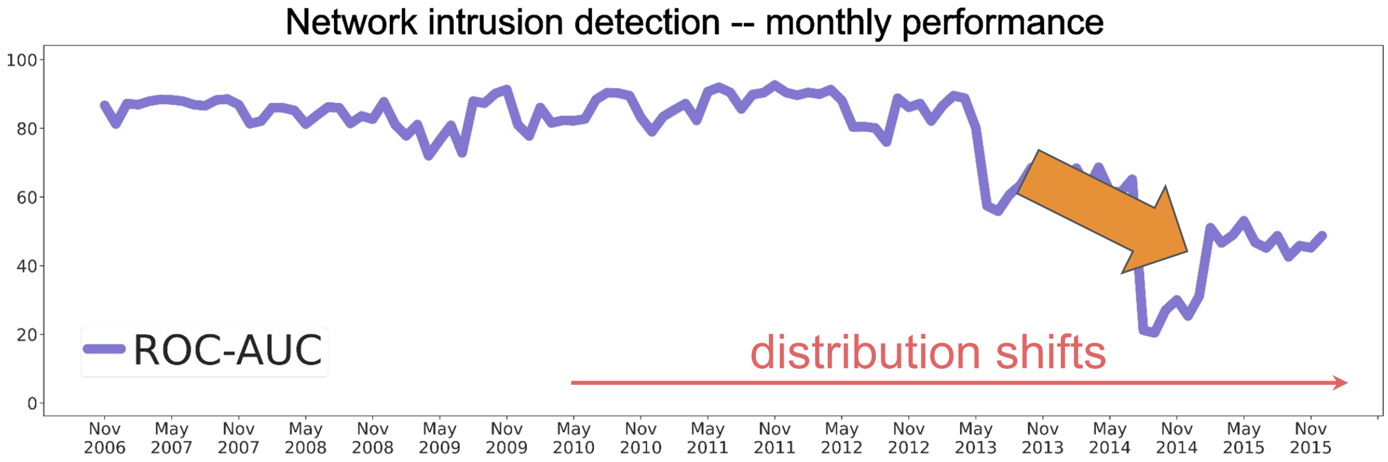
$$\theta_s^* = \arg\min_\theta \mathbb{E}_{x \sim P_s} \underbrace{L[S_\theta(x)]}_{\text{loss function}}$$

At **test** time, evaluate $S_{\theta_s^*}(x)$ on both normal and abnormal data.

However, an anomaly detector will fail under distribution shifts.

AD under distribution shifts

In the applications of AD, distribution shifts make an anomaly detector ineffective.



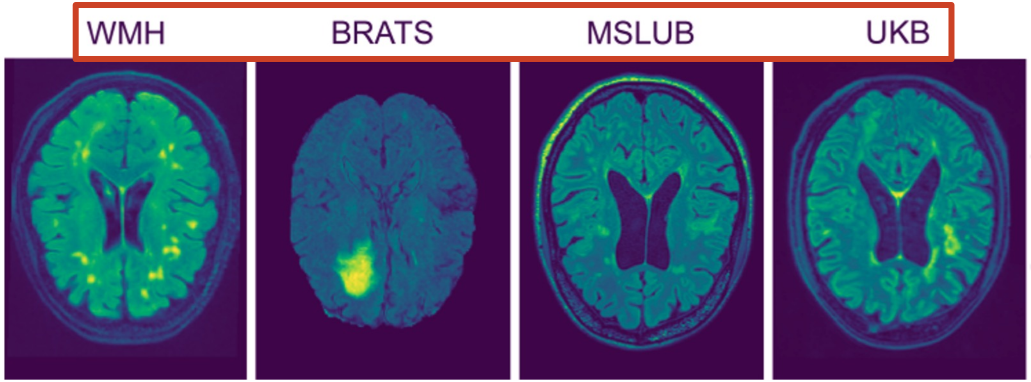
Dragoi, Marius, et al. "AnoShift: A distribution shift benchmark for unsupervised anomaly detection." *Advances in Neural Information Processing Systems* 35 (2022): 32854-32867.

AD under distribution shifts

In the applications of AD, distribution shifts make an anomaly detector ineffective.

Brain tumor detection

from different hospitals



AUPRC: 0.320 0.555 0.272 0.159

Pinaya, Walter Hugo Lopez, et al. "Unsupervised Brain Anomaly Detection and Segmentation with Transformers." *Medical Imaging with Deep Learning*. PMLR, 2021.

An intuition for zero-shot AD

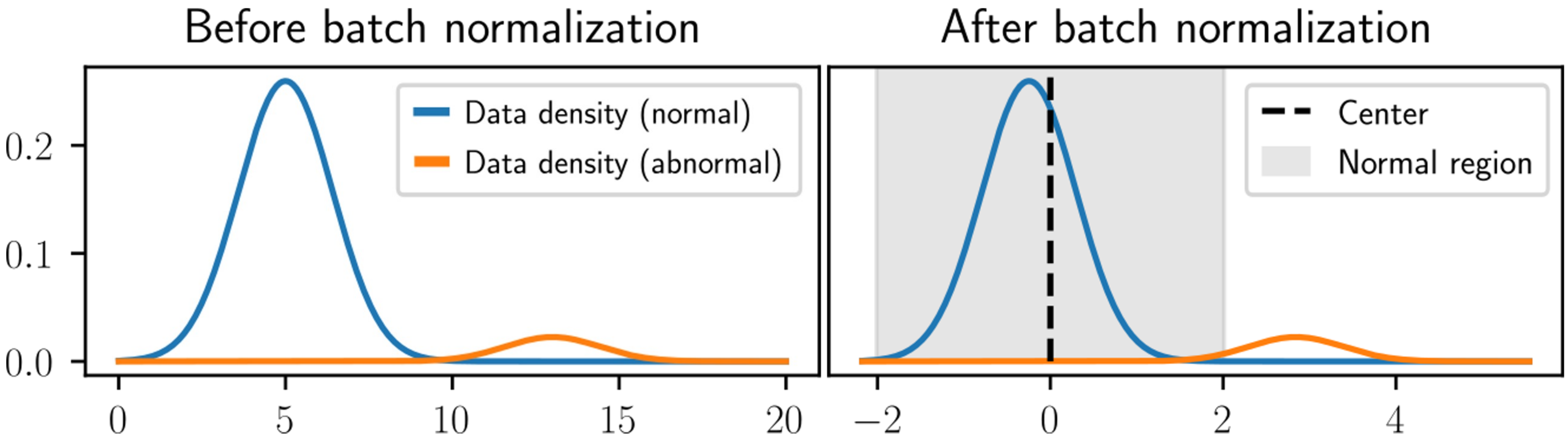
batch-level prediction

How do we implement this intuition?

Key idea: batch normalization

Batch normalization automatically centers the normal data of a mini-batch $\mathbf{x}_{\mathcal{B}}$

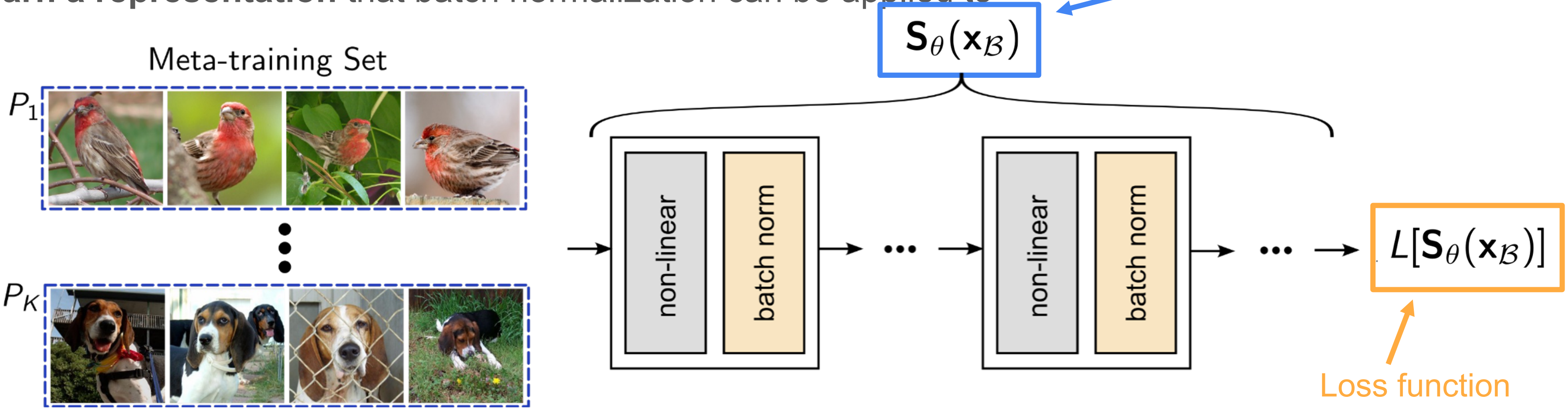
$$\mathbf{x}_i^{\text{BN}} = (\mathbf{x}_i - \bar{\mu}_{\mathbf{x}_{\mathcal{B}}})/\bar{\sigma}_{\mathbf{x}_{\mathcal{B}}}$$



Batch normalization as a zero-shot method does not directly apply for complex data such as images.

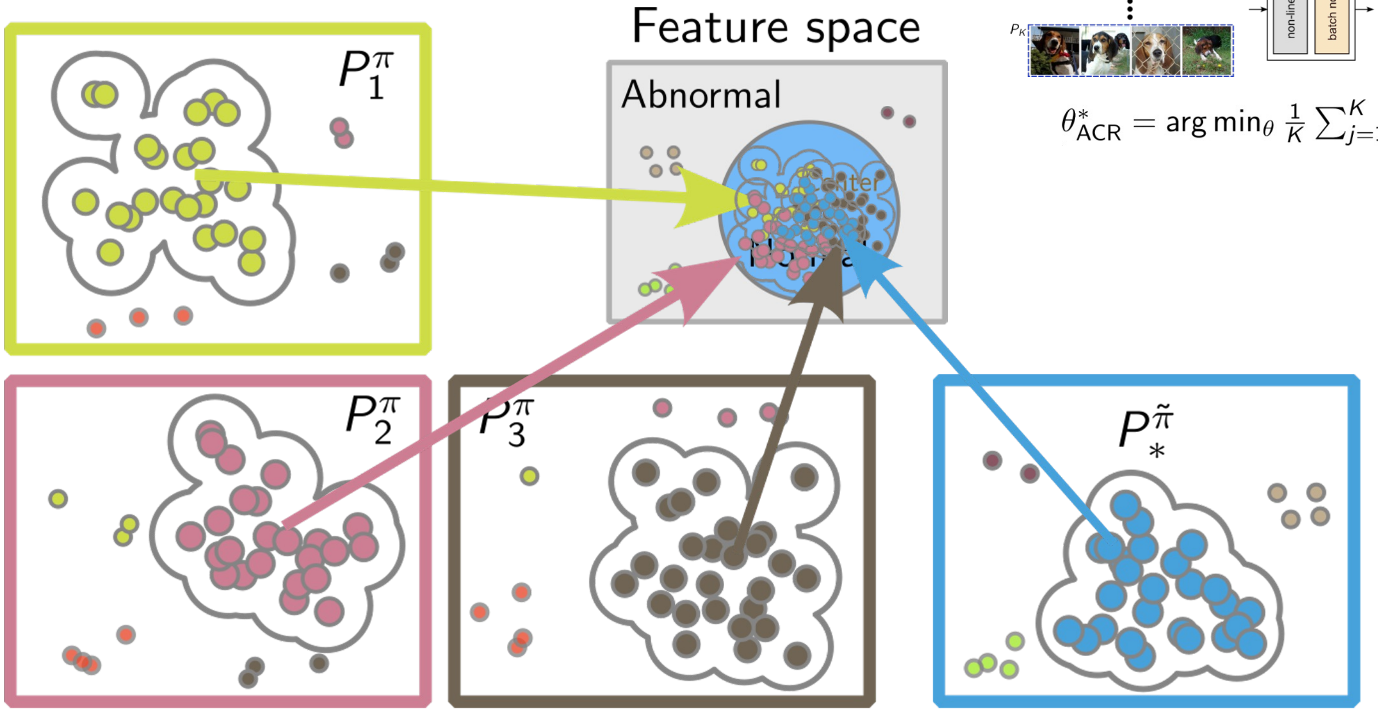
Proposed Method - Automatic Centered Representations (ACR)

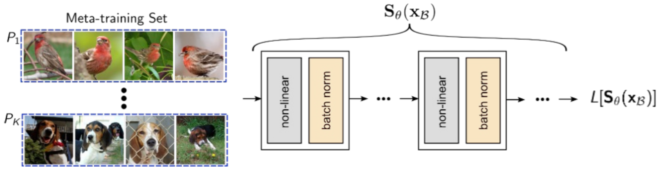
Learn a representation that batch normalization can be applied to



$$\theta^*_{\text{ACR}} = \arg\min_\theta \frac{1}{K} \sum_{j=1}^{K} \mathbb{E}_{\mathbf{x}_\mathcal{B} \sim P_j} L[\mathbf{S}_\theta(\mathbf{x}_\mathcal{B})]$$

Proposed Method - Automatic Centered Representations (ACR)





$$\theta^*_{\mathrm{ACR}} = \arg\min_\theta \frac{1}{K}\sum_{j=1}^{K} \mathbb{E}_{\mathbf{x}_\mathcal{B}\sim P_j} L[\mathbf{S}_\theta(\mathbf{x}_\mathcal{B})]$$

Comparison against stationary anomaly detection

Zero-Shot Anomaly Detection (ours)

Batch-mode training

Batch-norm layers in training mode

$$\theta^*_{\text{ACR}} = \arg\min_\theta \frac{1}{K} \sum_{j=1}^{K} \mathbb{E}_{\mathbf{x}_\mathcal{B} \sim P_j} L[\mathbf{S}_\theta(\mathbf{x}_\mathcal{B})]$$

Batch-mode test

Batch-norm layers in training mode

$$\mathbf{x}_\mathcal{B} \sim \pi P_* + (1-\pi)\bar{P}_*$$

$$\mathbf{S}_{\theta^*_{\text{ACR}}}(\mathbf{x}_\mathcal{B}) = \{S^1_{\theta^*_{\text{ACR}}}(\mathbf{x}_\mathcal{B}), \dots, S^{|\mathcal{B}|}_{\theta^*_{\text{ACR}}}(\mathbf{x}_\mathcal{B})\}$$

Stationary Anomaly Detection

Batch-mode training

Batch-norm layers in training mode

$$\theta^*_s = \arg\min_\theta \mathbb{E}_{\mathbf{x}_\mathcal{B} \sim P_s} L[\mathbf{S}_\theta(\mathbf{x}_\mathcal{B})]$$

Batch-mode test

Batch-norm layers in inference mode

$$\mathbf{x}_\mathcal{B} \sim \pi P_s + (1-\pi)\bar{P}_s$$

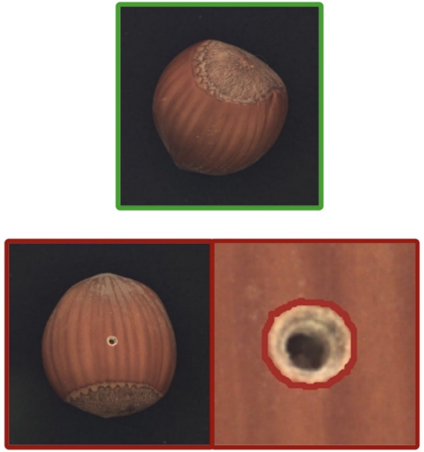
$$\mathbf{S}_{\theta^*_s}(\mathbf{x}_\mathcal{B}) = \{S^1_{\theta^*_s}(\mathbf{x}_\mathcal{B}), \dots, S^{|\mathcal{B}|}_{\theta^*_s}(\mathbf{x}_\mathcal{B})\}$$

Experiments

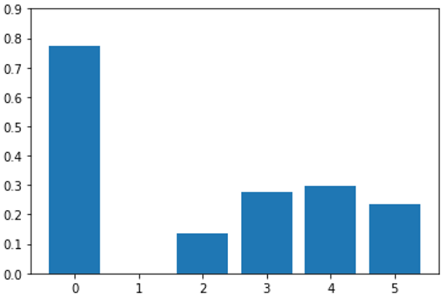
Tasks

Object-level anomaly detection

Industrial inspection



Tabular data



We ensure the test distributions are not encountered at training time.

Poster

Zero-Shot Anomaly Detection via Batch Normalization

Aodong Li · Chen Qiu · Marius Kloft · Padhraic Smyth · Maja Rudolph · Stephan Mandt

Great Hall & Hall B1+B2 #1017

[Abstract]

[OpenReview]

Tue 12 Dec 8:45 a.m. PST — 10:45 a.m. PST (Bookmark)

Abstract

Anomaly detection (AD) plays a crucial role in many safety-critical application domains. The challenge of adapting an anomaly detector to drift in the normal data distribution, especially when no training data is available for the "new normal", has led to the development of zero-shot AD techniques. In this paper, we propose a simple yet effective method called Adaptive Centered Representations (ACR) for zero-shot batch-level AD. Our approach trains off-the-shelf deep anomaly detectors (such as deep SVDD) to adapt to a set of inter-related training data distributions in combination with batch normalization, enabling automatic zero-shot generalization for unseen AD tasks. This simple recipe, batch normalization plus meta-training, is a highly effective and versatile tool. Our theoretical results guarantee the zero-shot generalization for unseen AD tasks; our empirical results demonstrate the first zero-shot AD results for tabular data and outperform existing methods in zero-shot anomaly detection and segmentation on image data from specialized domains. Code is at `https://github.com/aodongli/zero-shot-ad-via-batch-norm`