Revisiting Reverse Distillation for Anomaly Detection

Tran Dinh Tien[1] Anh Tuan Nguyen[1] Nguyen Hoang Tran[1] Ta Duc Huy[1]
Soan T.M. Duong[1,3] Chanh D. Tr. Nguyen[1,2] Steven Q. H. Truong[1,2]
[1]VinBrain JSC, Vietnam [2]VinUniversity, Vietnam [3]Le Quy Don Technical University, Vietnam

Poster THU-PM-374

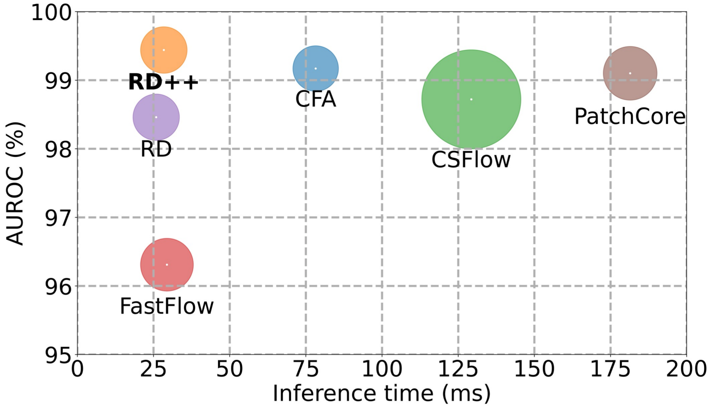
June 22, CVPR 2023

Motivation

- Recent SOTA methods (PatchCore, CFA, etc) for anomaly detection demonstrated excellent accuracy but come with a latency trade-off.
- How can we develop a method that achieves high accuracy and fast inference for **real-world applications**?
- We identify the limitations of the RD approach (Deng and Li) by examining feature compactness and anomalous signal suppression, then **introducing the multi-task learning design.**

Performance results of our RD++ method

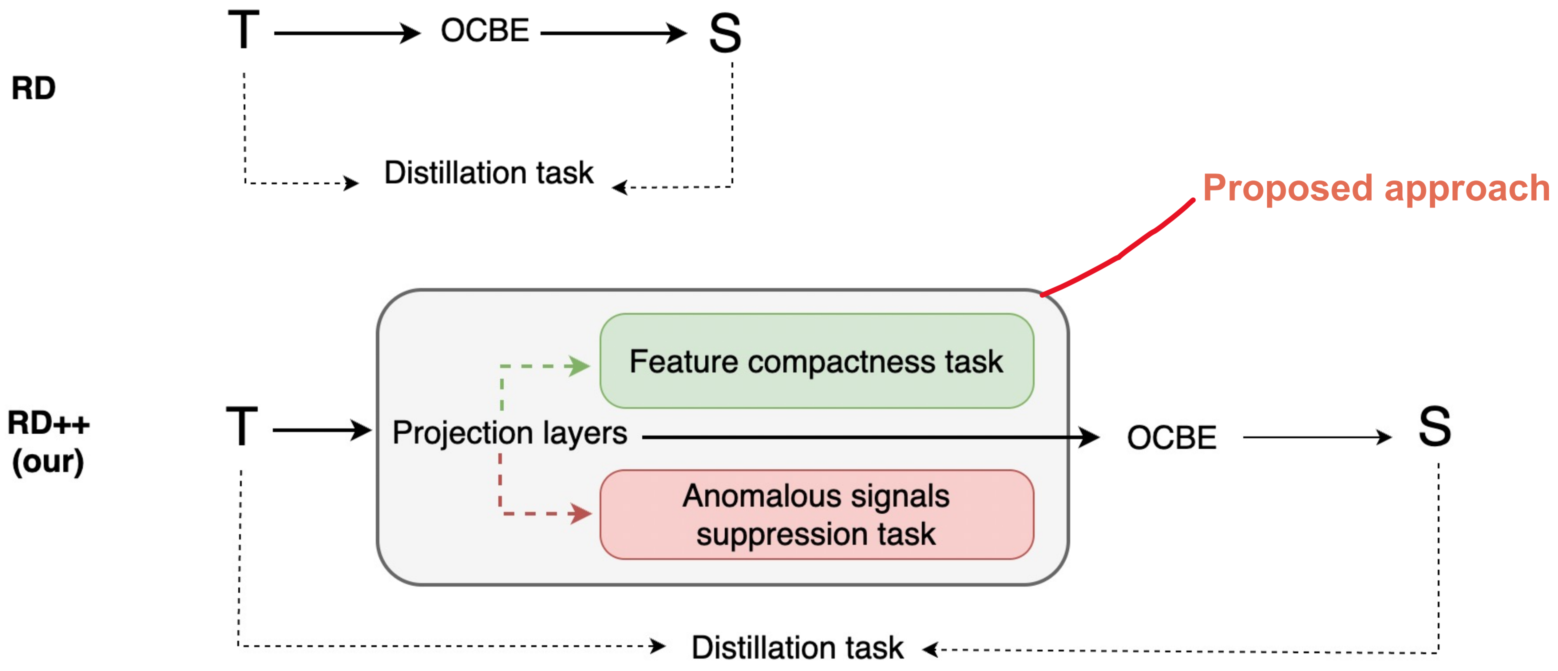
1 * New SOTA accuracy on MVTEC dataset
 * 6.x faster than PatchCore, 2.x faster than CFA



2. Strong generalizability in multiple datasets with different domains

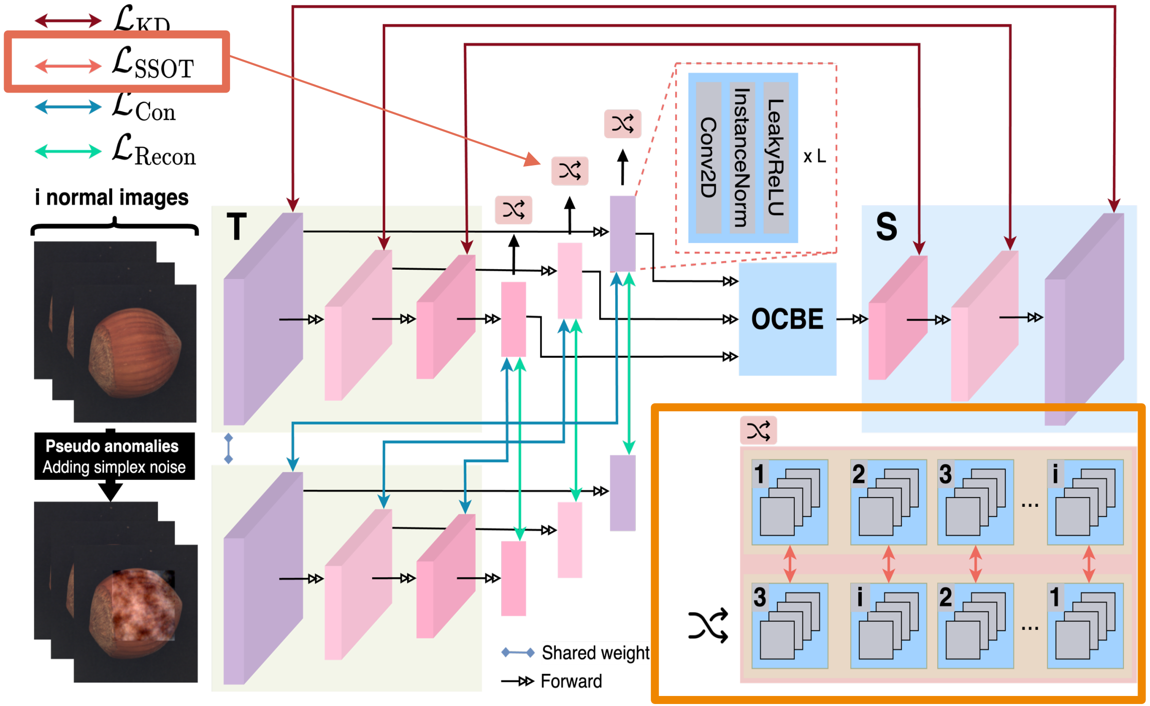
Datasets	Detection (AUROC sample)	Localization (AUROC pixel)	Localization (PRO)
MVTEC	Top 1	Top 1	Top 1
BTAD	Top 1	Top 2	Top 1
Retinal OCT	Top 1	Top 1	Top 1

Comparing architecture: RD++ (our) vs RD



1. **Feature compactness task**: by presenting a self-supervised optimal transport method
2. **Anomalous signal suppression task**: by simulating pseudo-abnormal samples with simplex noise and minimizing the reconstruction loss.

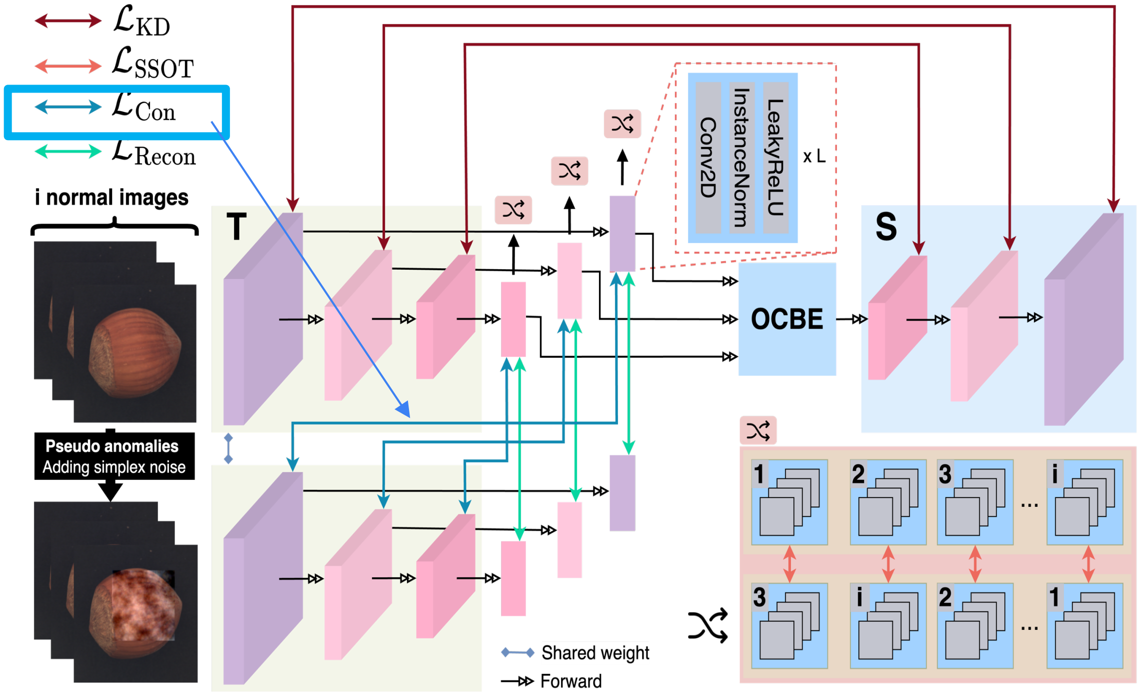
RD++ training process



Feature compactness task:

1. Self-supervised optimal transport loss (SSOT): ensuring pair-wise feature spaces in a mini-batch of normal images are close by minimizing the de-biased Sinkhorn divergence between their probability measures.

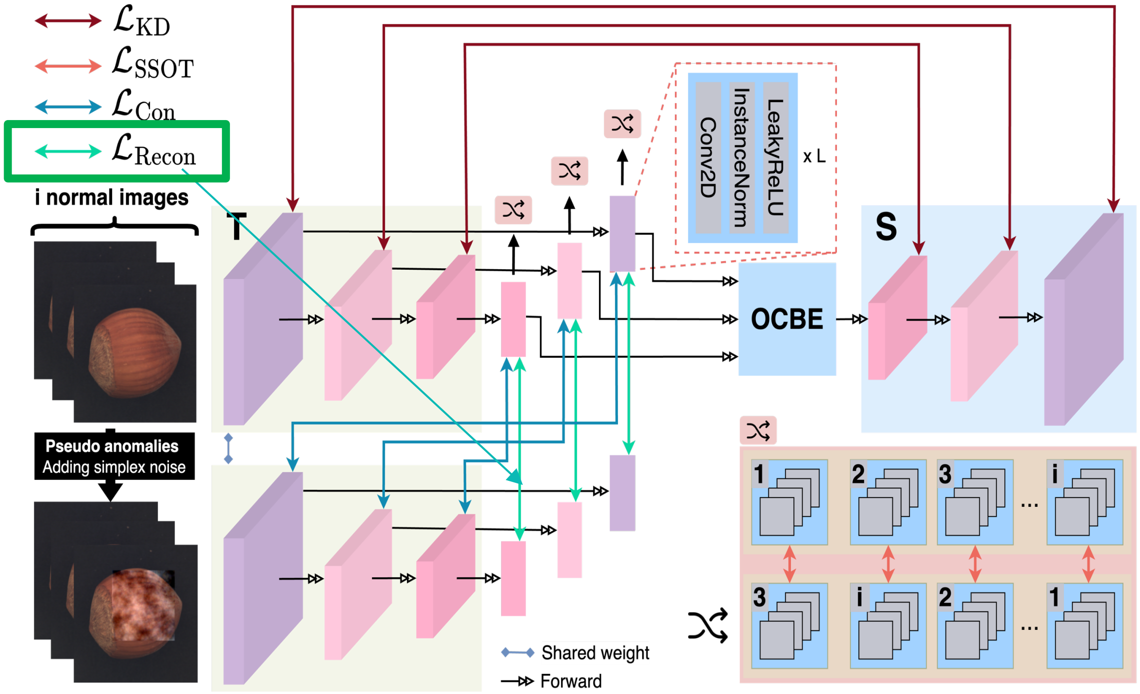
RD++ training process



Feature compactness task:

2. Contrast loss: supports projection layers learning compact embedding by setting projected normal features apart from abnormal features.

RD++ training process



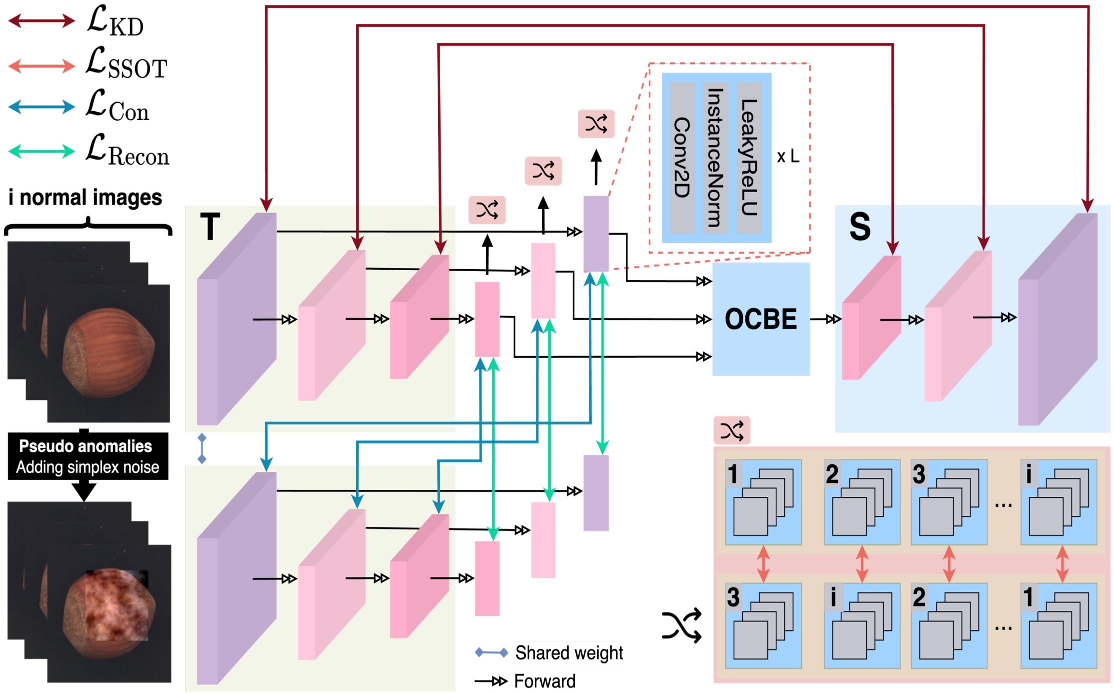
Anomalous signal suppression task

1. Simulating pseudo-abnormal samples via Simplex noise.

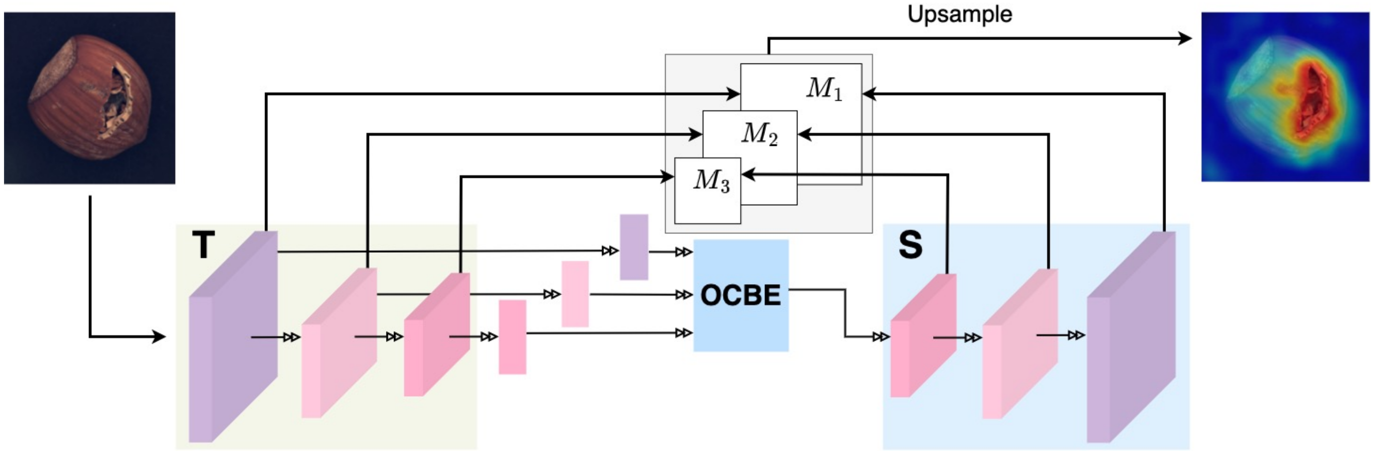
2. Reconstruction loss guide the projection layer to know how to reconstruct the normal feature space from the pseudo-abnormal feature.

RD++ training process

Overall Loss: $\mathcal{L} = \mathcal{L}_{\text{KD}} + \alpha\mathcal{L}_{\text{SSOT}} + \beta\mathcal{L}_{\text{Recon}} + \gamma\mathcal{L}_{\text{Con}}$



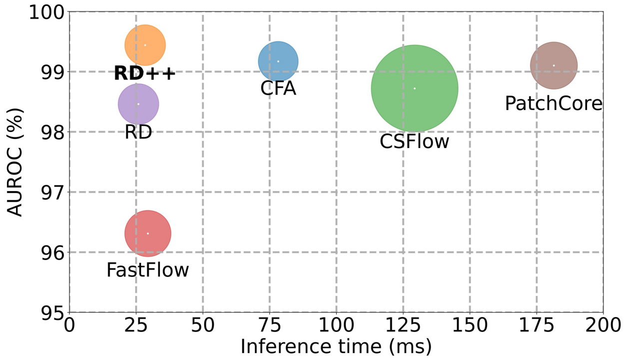
RD++ inference process



* Teacher's output embeddings are passed to their counterpart projection layer, then forward to student.
* Since the projection layer is lightweight, the inference time is almost the same as the baseline RD.

Method Results (%)

MVTEC dataset			
	Detection	Localization	
	AUROC sample	AUROC pixel	PRO
CSFLow	98.72	_	_
FastFlow	96.31	97.87	_
CFA	99.17	98.15	94.44
PatchCore	99.1	98.06	93.40
RD	98.46	97.81	93.93
RD++(Our)	**99.44**	**98.25**	**94.99**



BTAD dataset			
	Detection	Localization	
	AUROC sample	AUROC pixel	PRO
FastFlow	90.97	96.33	71.43
CFA	94.20	96.83	73.10
PatchCore	92.68	97.35	59.97
RD	94.30	97.67	77.10
RD++(Our)	95.63	97.43	77.30

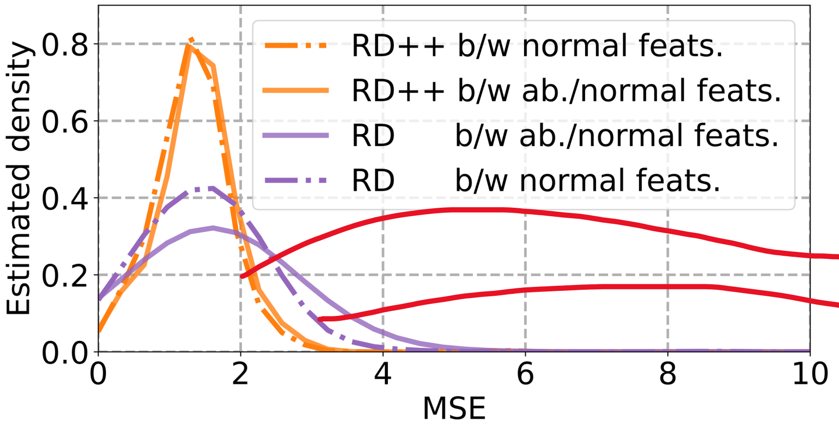
Retinal OCT dataset	
	AUROC sample
FastFlow	80.40
CFA	98.25
PatchCore	99.70
RD	99.36
RD++(Our)	99.73

1. RD++ achieves **highest accuracy** at both detection and localization, with **very fast inference**

2. RD++ indicates **strong generalizability** in multiple datasets with different domains

Method analysis

What can the student see?



RD++ feature compactness

RD feature compactness

We analyze two essential factors in student input features that play a vital role in the anomaly detection ability of the T/S architecture:

(i) Feature compactness

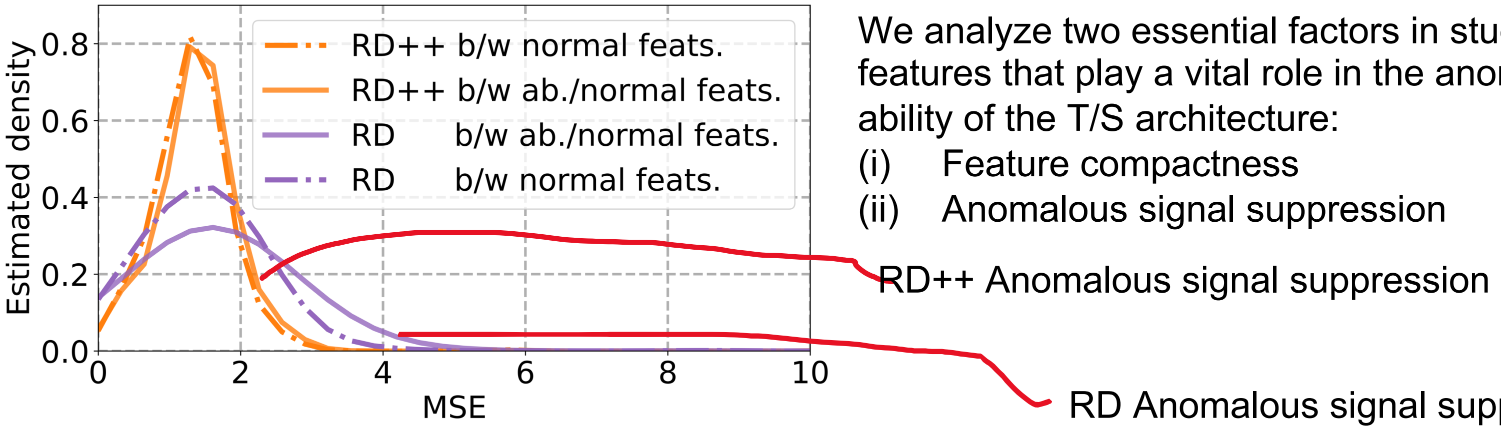
(ii) Anomalous signal suppression

(i) Feature compactness: by calculating MSE among features of the normal samples.

RD++ enjoys a much denser feature space while RD has a wider spread distance distribution

Method analysis

What can the student see?



We analyze two essential factors in student input features that play a vital role in the anomaly detection ability of the T/S architecture:

(i) Feature compactness

(ii) Anomalous signal suppression

(i) Feature compactness: by calculating MSE among features of the normal samples.
RD++ enjoys a much denser feature space while RD has a wider spread distance distribution

(ii) Anomalous signal suppression: by calculating the MSE between each of the abnormal samples with every normal sample. The inter-class MSE distribution of RD++ is also narrower than RD.
The feature distance between a normal and an abnormal sample **should be close** because the anomalous signals from the abnormal sample are suppressed, making it looks like a normal sample.

Conclusion and future work

Summary: Our approach obtains competitive accuracy for anomaly detection and real time inference. We hope the method will be **helpful in real applications** and pave the way for further advances in this field.

Future work: We aim to apply the method to other tasks besides anomaly detection, such as domain adaptation, where **invariant representation** is an essential factor. The method can benefit feature invariants through projection layers via compact learning and the ability to recover standard information when domains change.

Thank you for your listening

📜Paper

🌟Code