Memory-Distilled Selection for Noise-Robust Anomaly Detection

Sirojbek Safarov*[1], Jaewoo Park*[1], Yoon Gyo Jung*[2], Kuan-Chuan Peng[3], Wonchul Kim[1], Seongdeok Bang[1], and Octavia Camps[2]

[1]AIVEX [2]Northeastern University [3]Mitsubishi Electric Research Laboratories

QR Code

ICML International Conference On Machine Learning

Introduction

Q) Is data contamination critical on Anomaly Detection where most methods leverage the normal data manifold for their solutions?
A) Yes, very much.

Q) Can we trust human-annotated data to be 100% noiseless?
A) Probably not.

Q) Are there methods that solve this?
A) Yes, however, previous methods 1) have tradeoff between noise robustness and performance and underperform comparing with baselines trained on noise-free data or 2) contradictively, underperform on noise-free setup.

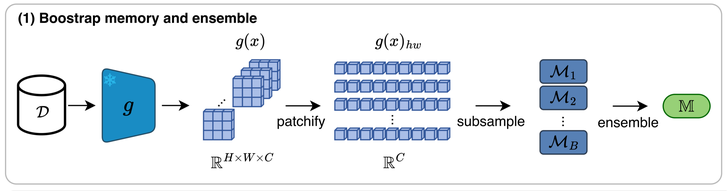
Q) Can we resolve these current limitations?
A) Yes, we propose a noise-robust anomaly detection framework that achieves performance comparable to noise-free baselines on any noise ratio including 0%
MeDS: **Me**mory-**D**istilled **S**election consisted of

1.***Memory Score Construction***: ensemble sub-sampled memory scores
2.***Memory Score Distillation***: distill the memory scores to AD network.
3.***Progressive Data Selection***: progressively self-learn by select-and-train on normal samples.

Contributions

- We introduce a noise-robust AD method exploiting the sparsity of bootstrapped memory ensemble to isolate nominal patterns and subsequently refine pixel-level detection via distillation and iterative self-selection. MeDS achieves SOTA performance on none to high ratio contamination level.
- We provide theoretical and empirical insights on how small-ratio memory subsampling yields high recall under heavy contamination.
- We conduct extensive empirical analysis and demonstrate the effectiveness of MeDS for active label correction.

Method



Step 1: Memory Score Construction

Subsample $\rho = 0.1$ of feature set $g(\mathcal{D})$ B times resulting $\{\mathcal{M}_1, \dots, \mathcal{M}_B\}$ and ensemble them. With *appropriate subsampling ratio*, the sparse nature of memory ensemble acts as a low-pass filter explained in Theorem 1 and separates normal and anomalous features. However, these features aren't directly trained by the AD dataset and show limited feature representation.

$$S_{\mathbb{M}}(x) = \frac{1}{\mathcal{B}}\sum_{b=1}^{\mathcal{B}} s_{\mathcal{M}_b}(x)$$

Theorem 1. Under a regularity condition, for any anomaly patch features q_{anom} and normal patch features q_{norm}, the expected gap

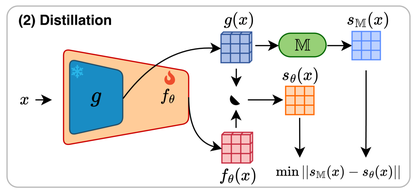
$$\triangle(m) \coloneqq \mathbb{E}[D(q_{anom}, \mathcal{M})] - \mathbb{E}[D(q_{norm}, \mathcal{M})]$$

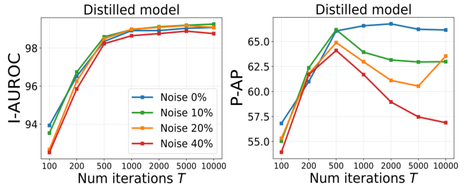
satisfies $\triangle(\mathbf{m}) > \mathbf{0}$ and is decomposed into a second-order Taylor approximation $\triangle_0$ with remainder $\epsilon_0(m)$: $\triangle(m) = \triangle_0(m) + \epsilon_0(m)$ where

$$\triangle_0(m) = \int_0^{\infty} \delta(r)\cdot\omega(m,r)dr$$

with a weight function $\omega(m,r)$ is unimodal with respect to m. The expectation $\mathbb{E}$ is over the memory $\mathcal{M}$ that is randomly subsampled from the extracted feature set $g(\mathcal{D})$ with the constraint $|\mathcal{M}| = m$.
➔ ***There is an appropriate sampling ratio maximizing normal-abnormal gap***

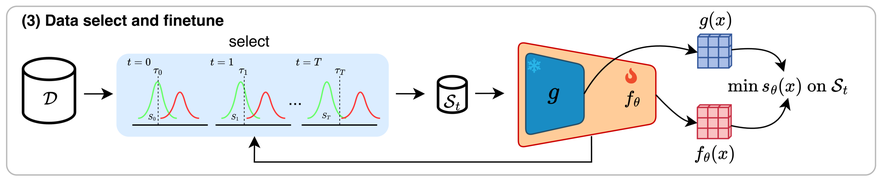
Method





Step 2: Distillation

Distill the memory scores to a reconstruction AD model s_θ which has features where normal-abnormal difference are amplified by directly using training dataset. However, prolonged training causes over-fitting in pixel level performance as shown in the graph on the right.



Step 3: Progressive Data Selection and Finetune

With the distilled model from Step 2, we use the image level representations to select normal samples progressively and repeat (train – selection) iteratively. Specifically, finetune is done on the AD model s_θ with progressively selected subset $\mathcal{S}_t$ by

$$\min_\theta \frac{1}{|\mathcal{S}_t|}\sum_{x\in\mathcal{S}_t} s_\theta(x)$$

and $\mathcal{S}_t$ selection is done by criterion $\eta_t(x)$ and threshold τ_t:

$$\eta_t(x) = (1-\alpha_t)\max_{h,w} s_{\theta_0}(x)_{hw} + \alpha_t \max_{h,w} s_\theta(x)_{hw}, \alpha_t = \min(1, \frac{2t}{T})$$

$$\tau_t = Median\big(\eta_t(x)\big) + k_t MAD\big(\eta_t(x)\big), k_t = k\left(\frac{t}{T}\right)$$

Where $\max_{h,w} s(x)_{hw}$ is image level scores, s_{θ_0} is the fixed initial model, s_θ is current model, T is total iterations, and MAD is mean absolute deviation.

Experiments

Dataset Setup

We use MVTecAD and VisA with **0%, 10%, 20%**, and **40%** noise ratios with 4 different seeds. For Real-IAD, we follow the protocol provided from the authors.

Table 1. Results on MVTecAD where underline highlights the best noisy AD baseline performance and **bold emphasizes the better performance between baseline and MeDS.**

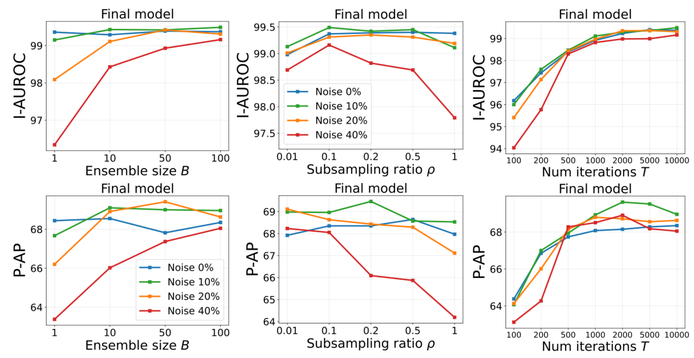
MVTecAD	I-AUROC (↑)				I-AP (↑)				P-AUPRO (↑)				P-AP (↑)			
Noise Ratio	0	10	20	40	0	10	20	40	0	10	20	40	0	10	20	40
SoftPatch	98.80	98.10	96.89	93.58	99.60	99.30	98.87	97.66	92.80	86.40	77.75	69.33	66.30	57.70	49.38	35.25
InReach	92.00	87.41	78.80	73.57	97.06	95.14	91.78	89.34	86.14	81.81	75.28	72.06	52.87	49.65	45.04	39.67
FUN-AD	81.89	95.49	96.06	97.70	89.18	97.73	98.04	98.90	61.49	78.39	78.21	73.91	42.94	58.52	58.86	61.38
HVQ	**96.71**	91.08	91.49	92.14	**98.83**	96.53	96.69	96.69	**91.31**	87.76	88.41	88.69	**47.71**	42.47	42.67	41.88
HVQ + MeDS (ours)	95.89	**94.95**	**94.76**	**94.26**	98.53	**98.09**	**98.01**	**97.80**	91.18	**90.40**	**90.19**	**90.13**	47.42	**46.13**	**44.65**	**44.82**
Dinomaly	**99.64**	95.19	92.16	87.38	**99.80**	97.34	95.50	93.28	94.62	91.04	90.21	89.26	68.19	58.22	54.60	53.00
Dinomaly + MeDS (ours)	99.37	**99.49**	**99.31**	**99.16**	99.72	**99.76**	**99.71**	**99.63**	**94.74**	**94.69**	**94.59**	**94.54**	**68.35**	**68.96**	**68.63**	**68.05**
INP-Former	**99.66**	95.13	91.21	85.85	**99.88**	97.34	94.31	91.14	94.88	91.06	89.64	88.85	**70.55**	59.88	54.39	51.26
INP-Former + MeDS (ours)	99.45	**99.39**	**99.41**	**99.17**	99.78	**99.79**	**99.78**	**99.68**	**95.13**	**95.25**	**95.22**	**95.21**	67.15	**67.79**	**67.51**	**67.39**

Table 2. Results on VisA where underline highlights the best noisy AD baseline performance and **bold emphasizes the better performance between baseline and MeDS.**

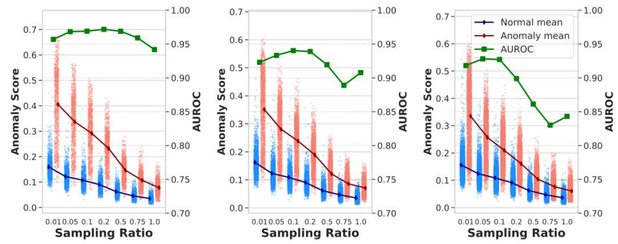
VisA	I-AUROC (↑)				I-AP (↑)				P-AUPRO (↑)				P-AP (↑)			
Noise Ratio	0	2	5	10	0	2	5	10	0	2	5	10	0	2	5	10
SoftPatch	93.85	93.37	92.47	89.91	94.88	94.74	94.12	92.86	88.39	86.77	82.88	75.09	47.50	45.56	44.73	37.03
InReach	83.99	79.34	73.40	64.15	86.82	84.09	80.64	74.06	78.80	73.78	65.45	51.30	31.37	29.54	27.76	24.15
FUN-AD	82.57	90.69	92.27	94.79	82.71	90.31	92.40	95.50	51.22	60.93	64.59	66.48	29.36	37.04	45.41	46.55
HVQ	**88.88**	87.92	87.12	86.10	**91.02**	90.02	89.34	88.33	**84.33**	**83.83**	**84.51**	**84.00**	**34.17**	**31.81**	**31.58**	**33.24**
HVQ + MeDS (ours)	88.36	**88.26**	**87.32**	**87.19**	90.25	**90.31**	**89.63**	**89.74**	83.19	83.29	83.27	83.40	30.26	31.00	31.38	31.99
Dinomaly	**98.47**	**97.35**	96.06	93.56	**98.63**	**97.67**	96.65	94.06	94.38	93.93	93.63	92.59	**52.80**	49.81	46.94	46.70
Dinomaly + MeDS (ours)	97.53	97.27	**97.51**	**97.43**	96.99	96.87	**97.12**	**97.01**	**94.39**	**94.06**	**94.10**	**94.39**	51.38	**50.93**	**51.27**	**51.46**
INP-Former	**98.15**	96.78	95.30	94.45	**98.37**	**96.96**	95.58	94.52	**94.70**	93.96	93.24	93.20	47.60	42.58	39.89	42.21
INP-Former + MeDS (ours)	98.02	**97.03**	**97.04**	**96.54**	96.66	96.78	**96.53**	**96.30**	94.19	**94.26**	**94.24**	**93.96**	**49.58**	**43.12**	**43.51**	**42.70**

Ablation & Analysis

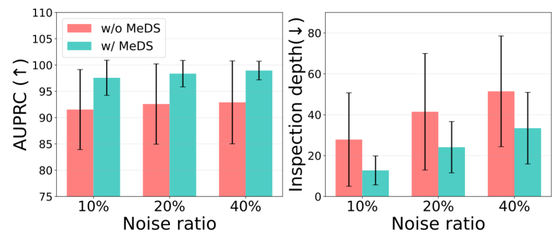
- Ablation results



- Empirical results of ***Theorem 1***



- Active label correction



Conclusion

- We suggest Memory-Distilled Selection framework to enable robust anomaly detection in contaminated data without any information of noise ratio or perform specific hyperparameter tuning.
- We leverage the sparse subsampling of memory banks to act as a low-pass filter, which theoretically and empirically amplifies the separation between normal and anomalous features.
- Distilling these scores into a reconstruction network exploits the early-learning bias of neural networks, allowing a progressive self-selection and fine-tune mechanism to achieve precise pixel-level localization without overfitting to noisy samples.
- We demonstrate consistent improvements across various datasets: MVTecAD, VisA, and Real-IAD, and prove the framework also serves as a highly practical tool for active label correction.