



# ICML

International Conference  
On Machine Learning

## Memory-Distilled Selection for Noise-Robust Anomaly Detection

Sirojbek Safarov\*<sup>1</sup>, Jaewoo Park\*<sup>1</sup>, Yoon Gyo Jung\*<sup>2</sup>, Kuan-Chuan Peng<sup>3</sup>,  
Wonchul Kim<sup>1</sup>, Seongdeok Bang<sup>1</sup>, and Octavia Camps<sup>2</sup>

<sup>1</sup>AIVEX

<sup>2</sup>Northeastern University

<sup>3</sup>Mitsubishi Electric Research Laboratories

**AIVEX**



**Northeastern  
University**



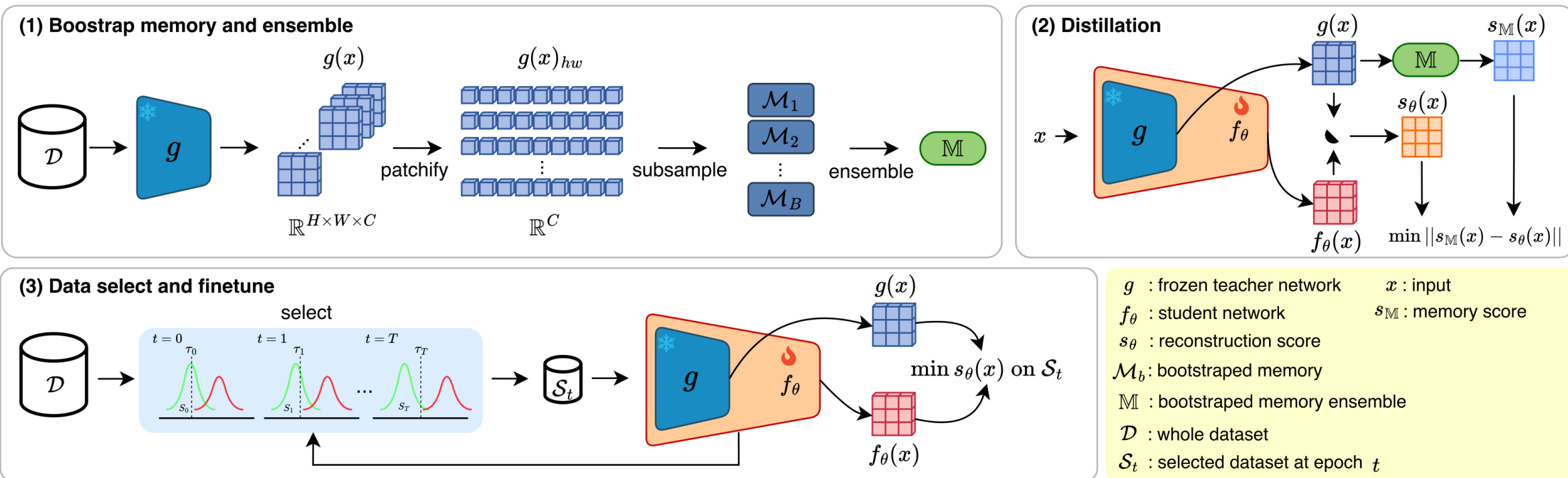
**MITSUBISHI ELECTRIC  
RESEARCH LABORATORIES**

# Memory-Distilled Selection for Noise-Robust Anomaly Detection

## Motivation :

- Data contamination is critical for anomaly detection methods assuming noise-free data.
- Previous methods have critical issues in nature which are
  - Noise robustness-Performance tradeoff:** underperform comparing with baselines trained on noise-free data.
  - Strong assumption:** requires noise in dataset and drastically fails on noise-free data

## MeDS Pipeline:



# Memory-Distilled Selection for Noise-Robust Anomaly Detection

## 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.

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

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

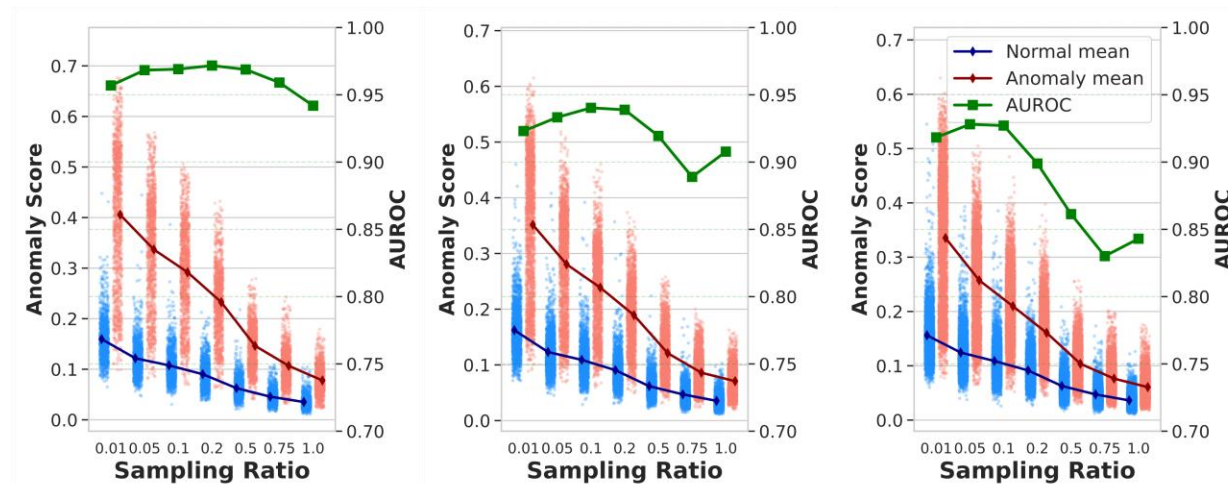
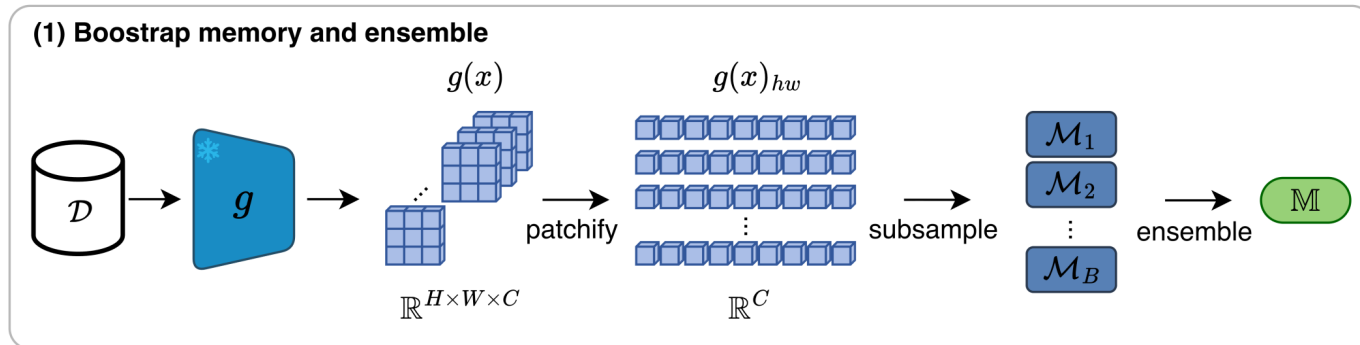
$$\Delta(m) := \mathbb{E}[D(q_{anom}, \mathcal{M})] - \mathbb{E}[D(q_{norm}, \mathcal{M})]$$

satisfies  $\Delta(m) > 0$  and is decomposed into a second-order Taylor approximation  $\Delta_0$  with remainder  $\epsilon_0(m)$ :  $\Delta(m) = \Delta_0(m) + \epsilon_0(m)$  where

$$\Delta_0(m) = \int_0^{\infty} \delta(r) \cdot \omega(m, r) dr$$

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  $|\mathcal{M}| = m$ .

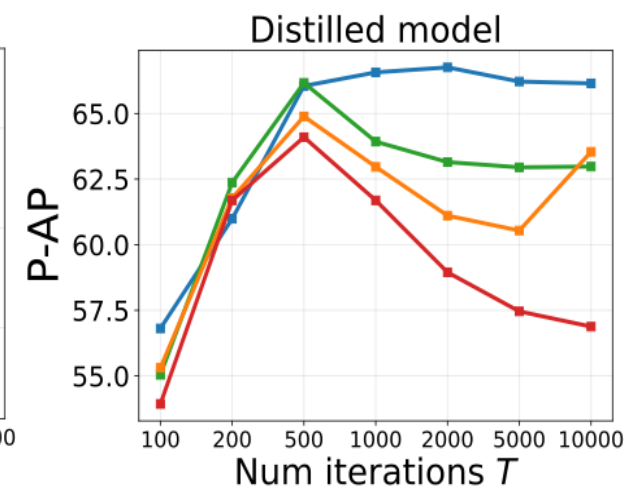
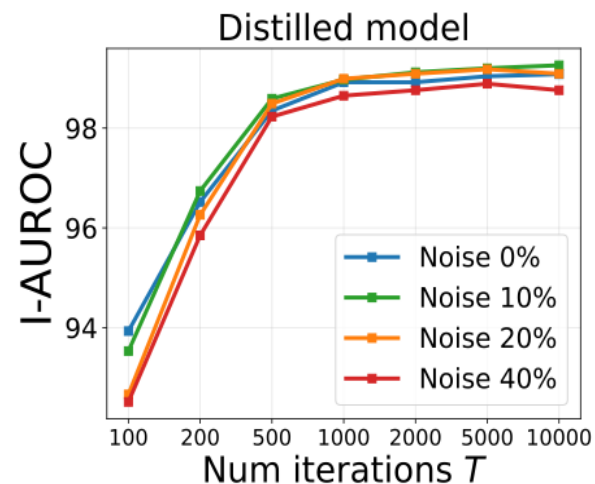
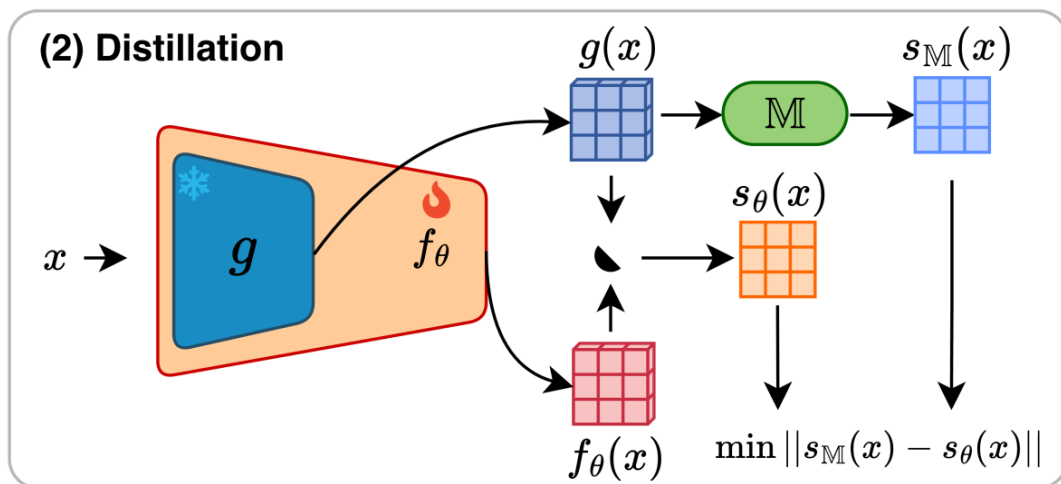
➔ There is an appropriate sampling ratio maximizing normal-abnormal gap



# Memory-Distilled Selection for Noise-Robust Anomaly Detection

## Step 2: Distillation

Using the memory scores from step 1, we distill these scores to a reconstruction-based AD model and amplify the difference between normal-abnormal features using each training data. However, prolonged training causes over-fitting in pixel level performance as shown in the graph on the right.



# Memory-Distilled Selection for Noise-Robust Anomaly Detection

## Step 3: Progressive Data Selection and Finetune

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

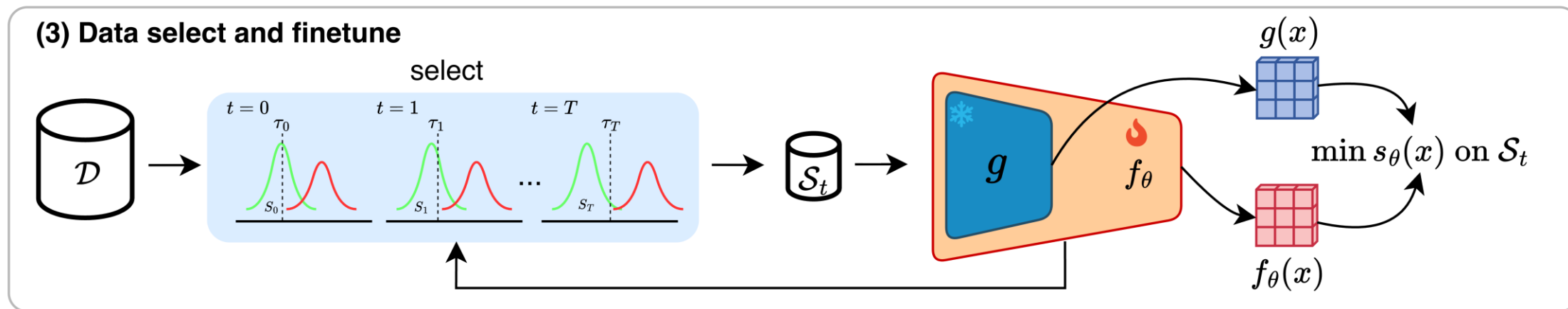
$$\min_{\theta} \frac{1}{|\mathcal{S}_t|} \sum_{x \in \mathcal{S}_t} s_\theta(x)$$

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

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

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

Where  $\max_{h,w} s(x)_{hw}$  is image level scores,  $s_{\theta_0}$  is the fixed initial model,  $s_\theta$  is current model,  $T$  is total iterations, and  $\text{MAD}$  is mean absolute deviation.



# Memory-Distilled Selection for Noise-Robust Anomaly Detection

## Results

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

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

# Memory-Distilled Selection for Noise-Robust Anomaly Detection

## Results

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

Metric	VisA															
	I-AUROC ( $\uparrow$ )				I-AP ( $\uparrow$ )				P-AUPRO ( $\uparrow$ )				P-AP ( $\uparrow$ )			
	Noise Ratio	0	2	5	10	0	2	5	10	0	2	5	10	0	2	5
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	<b>88.88</b>	87.92	87.12	86.10	<b>91.02</b>	90.02	89.34	88.33	<b>84.33</b>	<b>83.83</b>	<b>84.51</b>	<b>84.00</b>	<b>34.17</b>	<b>31.81</b>	<b>31.58</b>	<b>33.24</b>
HVQ + MeDS (ours)	88.36	<b>88.26</b>	<b>87.32</b>	<b>87.19</b>	90.25	<b>90.31</b>	<b>89.63</b>	<b>89.74</b>	83.19	83.29	83.27	83.40	30.26	31.00	31.38	31.99
Dinomaly	<b>98.47</b>	<b>97.35</b>	96.06	93.56	<b>98.63</b>	<b>97.67</b>	96.65	94.06	94.38	93.93	93.63	92.59	<b>52.80</b>	49.81	46.94	46.70
Dinomaly + MeDS (ours)	97.53	97.27	<b>97.51</b>	<b>97.43</b>	96.99	96.87	<b>97.12</b>	<b>97.01</b>	<b>94.39</b>	<b>94.06</b>	<b>94.10</b>	<b>94.39</b>	51.38	<b>50.93</b>	<b>51.27</b>	<b>51.46</b>
INP-Former	<b>98.15</b>	96.78	95.30	94.45	<b>98.37</b>	<b>96.96</b>	95.58	94.52	<b>94.70</b>	93.96	93.24	93.20	47.60	42.58	39.89	42.21
INP-Former + MeDS (ours)	98.02	<b>97.03</b>	<b>97.04</b>	<b>96.54</b>	96.66	96.78	<b>96.53</b>	<b>96.30</b>	94.19	<b>94.26</b>	<b>94.24</b>	<b>93.96</b>	<b>49.58</b>	<b>43.12</b>	<b>43.51</b>	<b>42.70</b>

# Memory-Distilled Selection for Noise-Robust Anomaly Detection

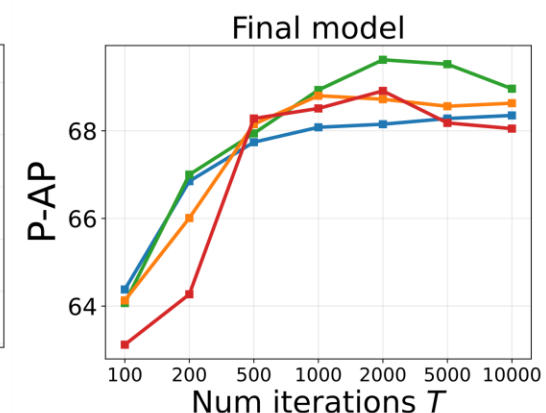
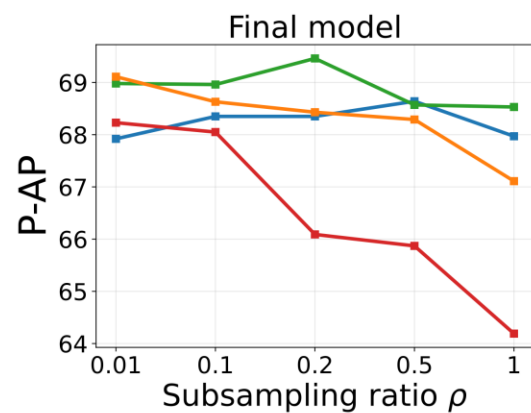
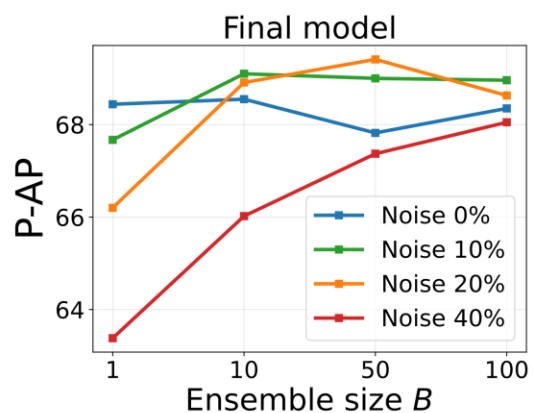
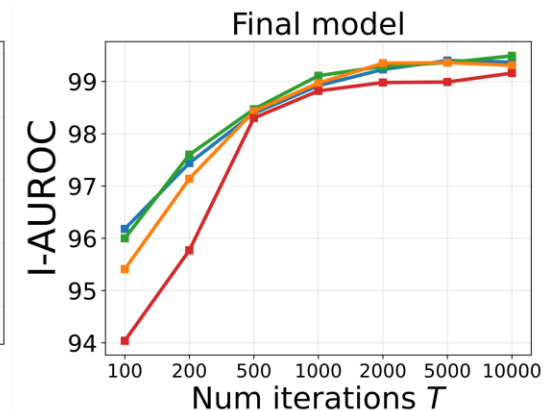
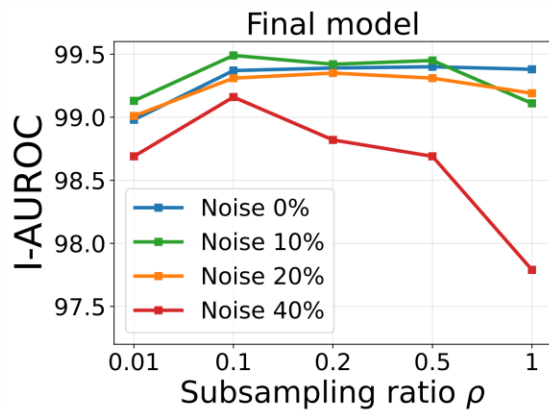
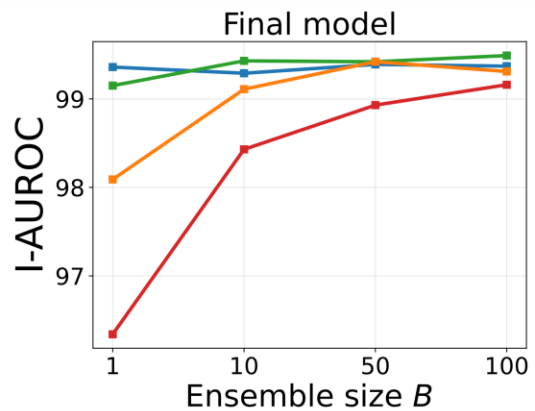
## Results

Table 3. Results on Real-IAD where **bold highlights best performance**

Metric	Real-IAD															
	I-AUROC ( $\uparrow$ )				I-AP ( $\uparrow$ )				P-AUPRO ( $\uparrow$ )				P-AP ( $\uparrow$ )			
	Noise Ratio	0	10	20	40	0	10	20	40	0	10	20	40	0	10	20
PatchCore	91.30	90.40	89.50	88.10	-	-	-	-	92.60	93.20	93.00	92.40	-	-	-	-
RD	88.10	88.10	87.30	84.50	-	-	-	-	95.10	95.10	94.90	94.70	-	-	-	-
UniAD	85.40	84.20	82.80	80.10	-	-	-	-	87.60	87.70	87.30	86.60	-	-	-	-
SoftPatch	91.48	90.81	<u>90.33</u>	<u>88.43</u>	<b>86.43</b>	<u>85.72</u>	<b>85.12</b>	<u>82.52</u>	91.30	91.43	91.22	90.96	35.78	34.78	<u>31.33</u>	28.16
Dinomaly	<b>92.39</b>	<u>91.38</u>	89.97	87.78	83.14	84.69	82.22	78.67	<b>95.97</b>	<b>96.16</b>	<u>95.88</u>	<u>95.47</u>	<u>40.04</u>	<b>40.58</b>	38.67	<u>35.41</u>
Dinomaly + MeDS (ours)	<u>92.34</u>	<b>92.36</b>	<b>92.08</b>	<b>90.99</b>	<u>84.76</u>	<b>84.89</b>	<u>85.11</u>	<b>83.98</b>	<u>95.53</u>	<u>95.60</u>	<b>95.90</b>	<b>95.67</b>	<b>40.41</b>	<u>40.43</u>	<b>41.15</b>	<b>41.22</b>

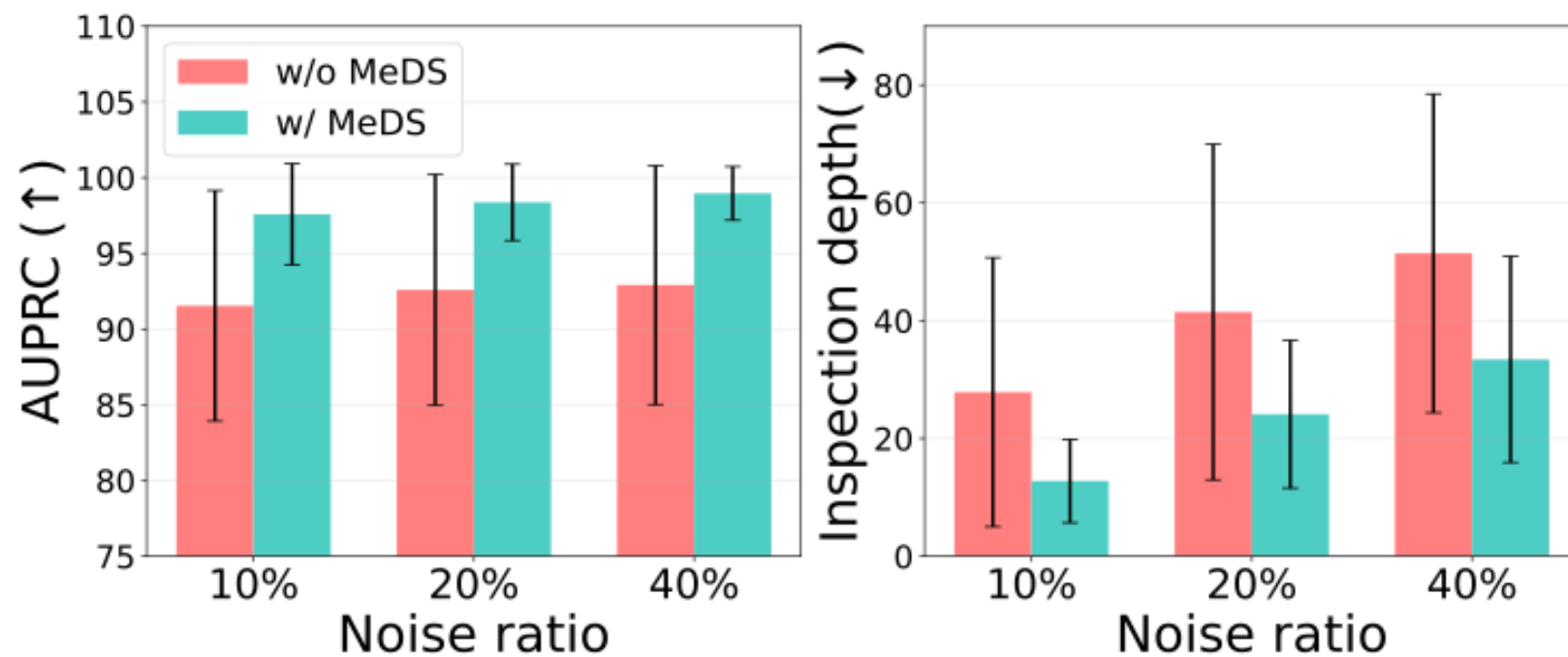
# Memory-Distilled Selection for Noise-Robust Anomaly Detection

## *Ablation*



# Memory-Distilled Selection for Noise-Robust Anomaly Detection

## *Active label correction*



# Memory-Distilled Selection for Noise-Robust Anomaly Detection

Thank You!

<https://github.com/SirojbekSafarov/MeDS>