Siamese U-Net for Image Anomaly Detection and Segmentation with Contrastive Learning

Chia-Ying Lin[1], Shang-Hong Lai[1,2]

[1]Institute of Information Systems and Applications, National Tsing Hua University, Hsinchu, Taiwan
[2]Department of Computer Science, National Tsing Hua University, Hsinchu, Taiwan

Motivation

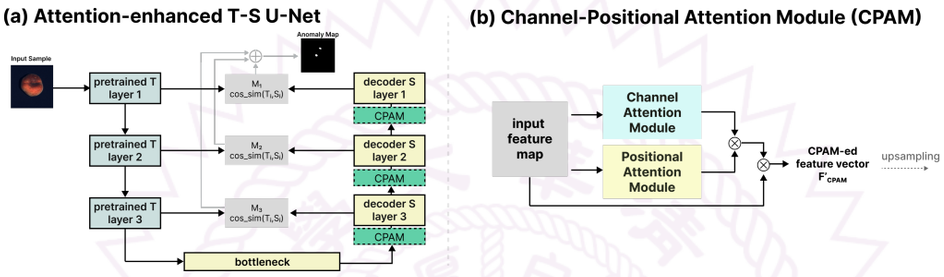
Mainstream approaches for anomaly detection and segmentation task in smart manufacturing usually suffer from two main issues:

- **High FPR/FNR: M**odels trained with normal data only in an unsupervised manner are more likely to produce high false positive/ negative rates as they have no access to true anomalies.
- **Indiscriminative image-level anomaly score: S**egmentation-dominated models are prone to produce indistinguishable image-level anomaly scores for subtle anomalies. Slight discrepancy between pixel anomaly scores for normal and anomalous features often results in indiscriminative image-level anomaly scores, leading to degraded anomaly detection performance.

To tackle the aforementioned issues, we propose a novel Siamese U-Net model trained with contrastive learning and deviation finetuning, incorporating a few anomalous samples, either real-world anomalies or synthetic samples, into model training.

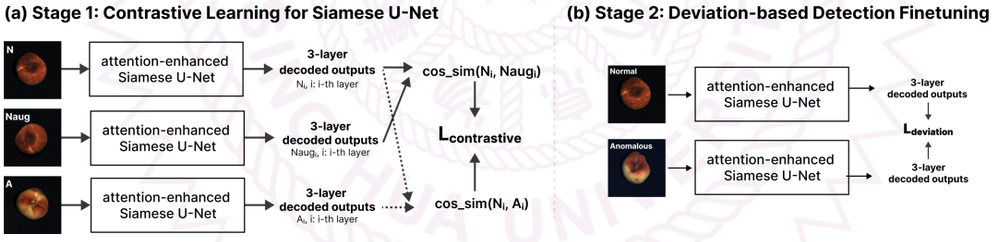
Proposed Method

Model Architecture



(a) Overview of the T-S U-Net block embedded with channel-positional attention module (CPAM). Layerwise cosine similarity maps are aggregated as anomaly map. (b) CPAM comprises two attention submodules, applying attention from channel and spatial aspects.

Model Flow: Two-Stage Training



(a) Stage 1: Contrastive Learning (b) Stage 2: Deviation- based Detection Finetuning

Experimental Results

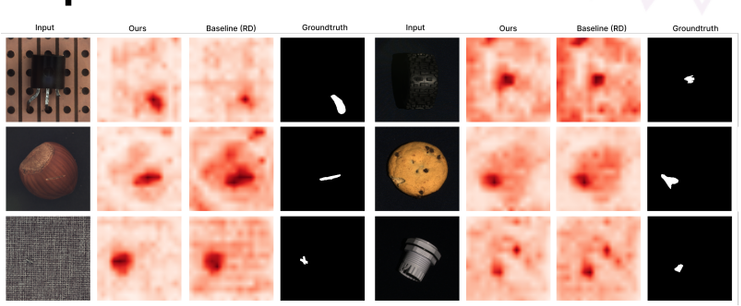


Fig 1: Qualitative comparison between our method and the baseline Reverse Distillation on MVTecAD (left) and MVTec3D-AD (right). For challenging case "tire" on the top row of the right half, our model can still generate a reliable anomaly map due to the aid of the deviation finetuning. Without deviation finetuning, the baseline tends to falsely predict the background with a high anomaly score.

	PatchCore	FastFlow	CS-Flow	Reverse Distillation	DRAEM+SSPCAB	RSTPM	CPC-AD	Ours(r)	Ours(s)
carpet	(-, -)	(**100.0, 99.4**)	(99.0, -)	(98.9, 97.0)	(98.2, 95.5)	(-, 99.0)	(80.9, -)	(**100.0**, 99.2)	(99.8, 99.3)
grid	(-, -)	(99.7, 98.3)	(100.0, -)	(100.0, 99.3)	(100.0, 99.7)	(-, 99.3)	(98.3, -)	(98.0, 99.3)	(**100.0**, 97.7)
leather	(-, -)	(100.0, 99.5)	(100.0, -)	(100.0, 99.4)	(100.0, 98.6)	(-, 99.0)	(99.0, -)	(98.0, 99.4)	(**100.0**, 99.5)
tile	(-, -)	(**100.0, 96.3**)	(100.0, -)	(99.3, 95.6)	(100.0, 99.2)	(-, 96.8)	(95.7, -)	(99.5, 95.9)	(99.1, 99.5)
wood	(-, -)	(**100.0, 97.0**)	(100.0, -)	(99.2, 95.3)	(99.5, 96.4)	(-, 96.4)	(80.3, -)	(**100.0**, 95.6)	(98.3, 95.7)
bottle	(-, -)	(100.0, 97.7)	(99.8, -)	(100.0, 98.7)	(98.4, 99.1)	(-, 98.9)	(99.8, -)	(**100.0**, 98.8)	(99.0, 98.6)
cable	(-, -)	(**100.0, 98.4**)	(97.1, -)	(95.0, 97.4)	(96.9, 94.7)	(-, 97.6)	(88.0, -)	(**100.0**, 97.7)	(96.8, 97.8)
capsule	(-, -)	(**100.0, 99.1**)	(98.6, -)	(96.3, 98.7)	(99.3, 94.3)	(-, 98.9)	(64.1, -)	(**100.0**, 98.9)	(97.6, 98.7)
hazelnut	(-, -)	(100.0, 99.1)	(99.3, -)	(99.9, 98.9)	(100.0, 99.7)	(-, 99.1)	(99.6, -)	(97.4, 99.1)	(**100.0**, 99.0)
metal_nut	(-, -)	(100.0, 98.5)	(99.7, -)	(100, 97.3)	(100.0, 99.5)	(-, 98.6)	(84.5, -)	(98.9, 97.7)	(**100.0**, 97.4)
pill	(-, -)	(**99.4, 99.2**)	(99.1, -)	(96.6, 98.2)	(99.8, 97.6)	(-, 97.1)	(92.1, -)	(99.6, 98.5)	(96.9, 98.0)
screw	(-, -)	(97.8, 99.4)	(99.6, -)	(97.0, 99.6)	(97.9, 97.6)	(-, 99.4)	(89.7, -)	(**100.0, 99.7**)	(97.5, 99.4)
toothbrush	(-, -)	(94.4, 98.9)	(99.1, -)	(99.5, 99.1)	(100.0, 98.1)	(-, 99.0)	(87.8, -)	(98.0, **99.1**)	(99.2, 99.0)
transistor	(-, -)	(**99.8, 97.3**)	(97.6, -)	(96.7, 92.5)	(92.9, 90.9)	(-, 88.1)	(92.5, -)	(99.5, 93.2)	(97.1, 92.3)
zipper	(-, -)	(**99.5**, 98.7)	(91.9, -)	(98.5, 98.2)	(100.0, 98.8)	(-, 98.5)	(99.3, -)	(98.7, 98.8)	(97.8, 98.1)
Total AVG	(**99.6, 98.4**)	(99.4, 98.5)	(98.7, -)	(98.5, 97.9)	(98.9, 97.3)	(96.9, 97.7)	(90.1, -)	(99.2, 98.1)	(98.6, 98.6)

Tab 1: **Comparison on MVTecAD with full-shot training setting.** AUROC% is reported in the format of (image-level, pixel-level). Our results include two variants: trained with a few (10-shot) real anomalies and with synthetic data, denoted as Ours(r) and Ours(s), respectively.

	PaDim	PatchCore	FastFlow	CFlow	DifferNet	CSFlow	Reverse Distillation	Ours(r)	Ours(s)
bagel	(97.5, 98.0)	(91.2, 89.9)	(89.3, 88.0)	(88.0, 85.5)	(81.9, -)	(89.4, -)	(96.2, 99.1)	(**99.2, 99.4**)	(99.0, 98.3)
cable gland	(77.5, 94.4)	(90.2, 95.3)	(62.0, 75.2)	(85.8, 91.9)	(67.0, -)	(91.7, -)	(88.7, **99.4**)	(**100.0, 99.4**)	(91.9, 97.8)
carrot	(69.8, 94.5)	(88.5, 95.7)	(79.5, 92.3)	(82.8, 95.8)	(61.2, -)	(74.9, -)	(94.2, **99.5**)	(**96.0, 99.5**)	(91.0, 98.9)
cookie	(58.2, 92.5)	(70.9, 91.8)	(42.6, 81.2)	(56.3, 86.7)	(48.4, -)	(66.8, -)	(62.4, 98.0)	(**83.2, 98.4**)	(65.8, 98.9)
dowel	(95.9, 96.1)	(95.2, 93.0)	(88.0, 92.9)	(98.6, 96.9)	(63.4, -)	(93.8, -)	(98.1, 99.5)	(**98.5, 99.6**)	(98.4, 99.2)
foam	(66.3, 79.2)	(73.3, 71.9)	(72.8, 64.6)	(73.8, 50.0)	(68.9, -)	(89.7, -)	(85.2, **96.7**)	(**93.5**, 94.9)	(83.1, 82.9)
peach	(85.8, 96.6)	(72.7, 92.0)	(65.1, 78.2)	(75.7, 88.9)	(65.5, -)	(60.3, -)	(**90.2, 99.4**)	(81.3, **99.4**)	(89.0, 97.8)
Potato	(53.5, 94.0)	(56.2, 93.7)	(56.0, 61.5)	(62.8, 93.5)	(60.0, -)	(41.9, -)	(68.2, 99.2)	(**90.3, 99.3**)	(66.8, 98.3)
Rope	(83.2, 93.7)	(96.2, 93.8)	(98.2, 91.3)	(97.0, 90.4)	(72.9, -)	(97.1, -)	(**98.4**, 99.4)	(92.4, **99.6**)	(**98.8**, 96.6)
Tire	(76.0, 91.2)	(76.8, 92.9)	(61.3, 55.0)	(72.0, 91.9)	(53.6, -)	(72.6, -)	(73.0, 98.5)	(**98.6, 98.9**)	(81.2, 94.1)
AVG	(76.4, 93.0)	(81.1, 91.0)	(71.5, 78.0)	(79.3, 87.1)	(64.3, -)	(77.8, -)	(85.5, **98.8**)	(**93.3, 98.8**)	(86.5, 95.4)

Tab 2: **Quantitative comparison with SOTA methods on MVTec3D-AD under 2D setup.** Our results include trained with a few (10-shot) real-world anomalies and with synthetic data, denoted as Ours(r) and Ours(s), respectively.