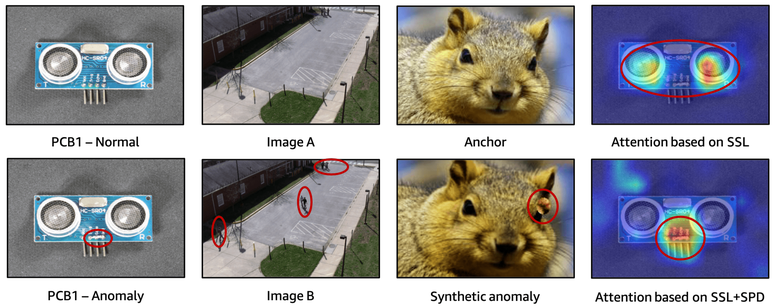
Spot-the-Difference: Self-Supervised Pre-training for Anomaly Detection and Segmentation

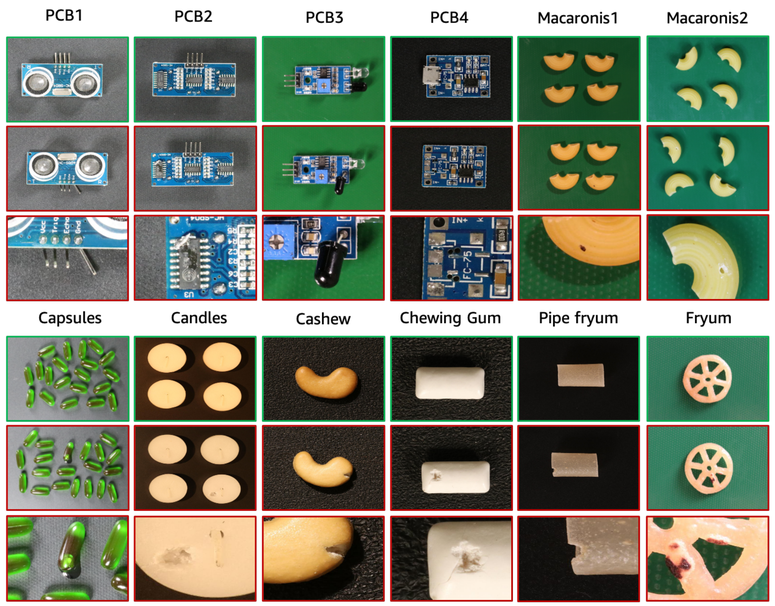
Yang Zou, Jongheon Jeong*, Latha Pemula, Dongqing Zhang, Onkar Dabeer

* PhD student at KAIST, work done during internship at AWS AI Labs

Anomaly detection and Spot-the-difference

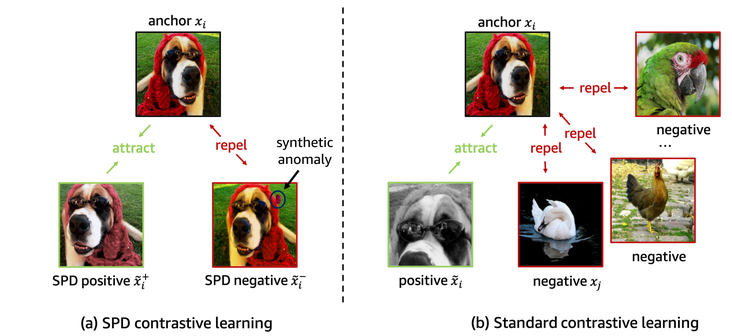


Visual Anomaly (VisA) Dataset

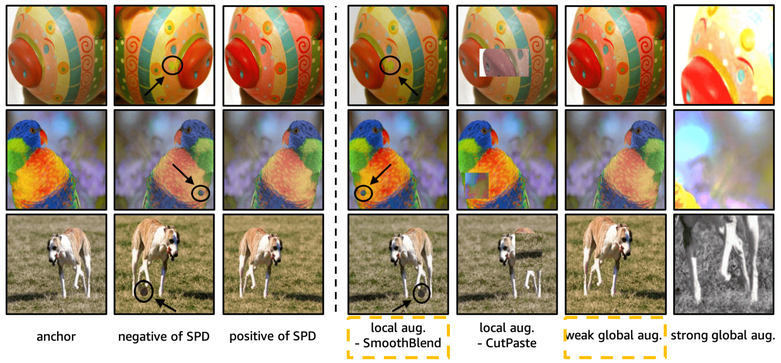


9,621 normal+1,200 anomalous images with multi-class image/pixel-level labels

Overview of SPot-Diff (SPD) Contrastive Learning



SmoothBlend and Augmentations for SPD



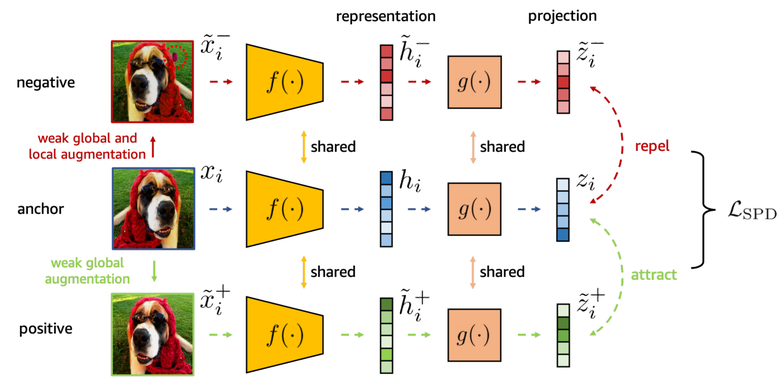
Preliminaries on Contrastive Learning

SimCLR [1]

InfoNCE loss $$\mathcal{L}_{\text{NCE}}(x_i, \hat{x}_i) = -\log \frac{\exp(z_i \cdot \hat{z}_i/\tau)}{\exp(z_i \cdot \hat{z}_i/\tau) + \sum_{j=1}^{N} \mathbb{1}_{j \neq i} \exp(z_i \cdot z_j/\tau)}$$

x_i: anchor image
x_j: other image in the same batch
N: batch size
$\hat{x}_i$: image with strongly global aug.
z_i: projection
τ: temperature

SPD Learning



SPD loss $$\mathcal{L}_{\text{SPD}}(x_i, \tilde{x}_i^-, \tilde{x}_i^+) = \cos(z_i, \tilde{z}_i^-) - \cos(z_i, \tilde{z}_i^+)$$

$\tilde{x}_i^-$: image with strongly local aug. (e.g. SmoothBlend)
$\tilde{x}_i^+$: image with weakly global aug.
h_i: representation
z_i: projection

SPD as Regularization

Regularized various contrastive SSL, such as SimCLR, MoCo [2], SimSiam [3]

SimCLR/MoCo with SPD $$\mathcal{L}(x_i, \hat{x}_i, \tilde{x}_i^-, \tilde{x}_i^+) = \mathcal{L}_{\text{NCE}}(x_i, \hat{x}_i) + \eta \cdot \mathcal{L}_{\text{SPD}}(x_i, \tilde{x}_i^-, \tilde{x}_i^+)$$

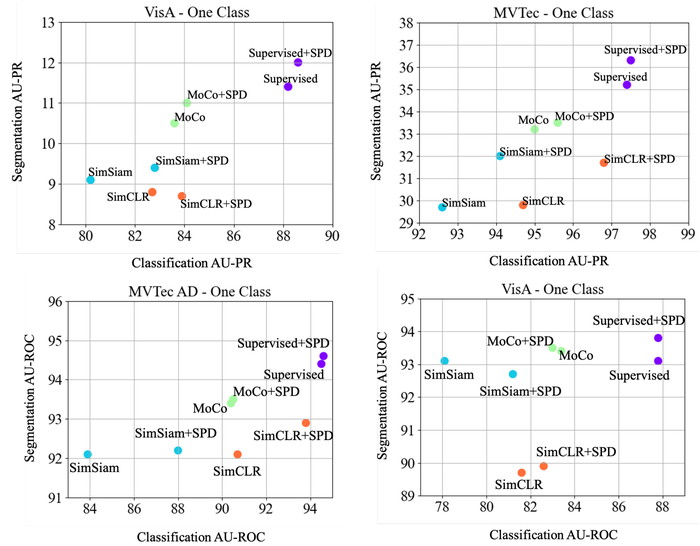
η: loss weight

Regularized Supervised Pre-training: Supervised pre-training with an auxiliary classifier trained by *xent* loss to classify if an image is with SmoothBlend or not.

Experiments on VisA/MVTec-AD

Anomaly classification and segmentation on 1-class full-shot setups

The anomaly detection method: PaDiM [4]



The SOTA anomaly detection method: PatchCore [5]

Backbone: Wide ResNet50	VisA (1-class)				MVTec-AD (1-class)			
	Classification		Segmentation		Classification		Segmentation	
	AU-PR	AU-ROC	AU-PR	AU-ROC	AU-PR	AU-ROC	AU-PR	AU-ROC
Sup. pre-train	93.3	92.4	38.4	98.4	99.2	99.8	48.8	97.6
Sup. pre-train+SPD	93.8 (+0.5)	92.5 (+0.1)	39.3 (+0.9)	98.1 (-0.3)	99.0 (-0.2)	99.7 (-0.1)	49.3 (+0.5)	97.5 (-0.1)

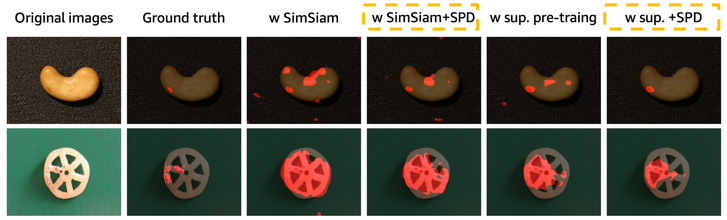
Anomaly classification and segmentation on 2-class few-shot setups of VisA

	ImageNet labels	Classification (2-class, low-shot)				Segmentation (2-class, low-shot)			
		5-shot		10-shot		5-shot		10-shot	
		AU-PR	AU-ROC	AU-PR	AU-ROC	AU-PR	AU-ROC	AU-PR	AU-ROC
Sup. pre-train	✓	59.2	85.5	70.4	91.7	17.8	74.6	28.3	81.8
SimSiam	✗	51.9	82.3	65.0	89.4	17.3	75.2	28.5	81.6
+SPD	✗	56.1 (+4.2)	84.0 (+1.7)	67.6 (+2.6)	90.8 (+1.4)	18.2 (+0.9)	**76.0** (+0.8)	29.7 (+1.2)	83.2 (+1.6)
MoCo	✗	56.1	83.8	68.7	90.6	21.5	80.5	32.3	**85.7**
+SPD	✗	56.4 (+0.3)	83.9 (+0.1)	68.0 (-0.7)	90.1 (-0.5)	**22.1** (+0.6)	78.5 (-2.0)	**32.8** (+0.5)	84.9 (-0.8)
SimCLR	✗	48.4	79.6	58.2	86.0	18.4	71.2	23.0	75.1
+SPD	✗	47.4 (-1.0)	79.9 (+0.3)	59.0 (+0.8)	86.1 (+0.1)	18.9 (+0.5)	74.5 (+3.3)	25.1 (+2.1)	78.2 (+3.1)
Sup. pre-train+SPD	✓	**59.8** (+0.6)	**85.9** (+0.4)	**71.2** (+0.8)	**92.1** (+0.4)	18.7 (+0.9)	75.9 (+1.3)	30.6 (+2.3)	81.8 (+0.0)

Ablation study

	VisA (1-class)				MVTec-AD (1-class)			
	Classification		Segmentation		Classification		Segmentation	
	AU-PR	AU-ROC	AU-PR	AU-ROC	AU-PR	AU-ROC	AU-PR	AU-ROC
SimSiam w/ Res50	80.2	78.1	9.1	93.1	92.6	83.9	29.7	92.1
+SPD ($\eta = 0.1$)	82.8	81.2	9.4	92.7	94.1	88.0	32.0	92.2
+SPD ($\eta = 0.5$)	80.5	79.3	8.7	93.0	93.3	84.9	30.1	91.9
+SPD ($\eta = 1.0$)	81.5	79.8	9.4	92.8	93.4	85.8	30.0	92.0
+SPD w/ CutPaste	78.8	77.0	9.7	93.1	93.5	85.2	28.2	91.3
+SPD w/ Xent	71.4	66.6	2.7	84.8	86.3	71.0	15.2	82.6
SimSiam w/ WideRes50	80.3	77.7	9.9	93.6	93.0	84.7	31.3	92.2
+SPD	81.9	80.4	10.5	93.7	93.4	85.4	32.5	92.8

Qualitative results



Take-away messages

- Improving sensitivity to local variation improves both self-supervised and supervised ImageNet pre-training for anomaly det/seg
- Supervised ImageNet pre-training generally outperforms self-supervised representation while SSL outperforms supervised pre-training in few-shot anomaly segmentation

[1] A Simple Framework for Contrastive Learning of Visual Representations, ICML 2020
[2] Momentum Contrast for Unsupervised Visual Representation Learning, CVPR 2020
[3] Exploring Simple Siamese Representation Learning, CVPR 2021
[4] PaDiM: a Patch Distribution Modeling Framework for Anomaly Detection and Localization, ICPR 2020
[5] Towards Total Recall in Industrial Anomaly Detection, CVPR 2022