DSV: An Alignment Validation Loss for Self-supervised Outlier Model Selection

Jaemin Yoo
KAIST
jaemin@kaist.ac.kr

Yue Zhao
University of Southern California
yzhao010@usc.edu

Lingxhao Zhao
Carnegie Mellon University
lingxiao@cmu.edu

Leman Akoglu
Carnegie Mellon University
lakoglu@andrew.cmu.edu

Overview

Motivation: How can we search hyperparameters for self-supervised anomaly detection (SSAD)?

Why is it difficult? No labels are given at training
→ No validation data can be used for HP search
→ Unsupervised model selection problem

Proposed method: DSV
Validation loss for unsupervised model selection
Focus on HPs in a data augmentation function
E.g., the patch size in CutOut

Problem Definition

Problem: Unsupervised model selection for SSAD

Given
1. Data augmentation function f_{aug}
2. Set Φ of different detector models that are trained with f_{aug} with different HPs
3. Training data $\mathcal{D}_{\mathrm{trn}}$ (w/o labels)
4. Test data $\mathcal{D}_{\mathrm{test}}$ (as a transductive setting)

Find
Detector $\phi^* \in \Phi$ with best accuracy on $\mathcal{D}_{\mathrm{test}}$

Main Ideas

Proposition: Functional alignment
Let f_{gen} be the anomaly-generating function
Detector ϕ will be accurate if
it is trained with f_{aug} that is aligned w/ f_{gen}

Q: How can we define the alignment?
Idea 1: Compute the embedding distance btw. data augmented by $f_{\mathrm{aug}}(\mathcal{D}_{\mathrm{trn}})$ and true anomalies in $\mathcal{D}_{\mathrm{test}}$

Q: We don't know which are the true anomalies…
Idea 2: Use $\mathcal{D}_{\mathrm{test}}$ as a whole, not only anomalies, comparing $\mathcal{D}_{\mathrm{trn}} \cup \mathcal{D}_{\mathrm{aug}}$ and $\mathcal{D}_{\mathrm{test}}$

Steps for Approximation

Step 1: Decompose alignment into two terms: **Discordance** h_d and **Separability** h_s

Step 2: Design approximate **surrogate losses**
$h_d \rightarrow \mathcal{L}_{\mathrm{dis}}$ and $h_s \rightarrow \mathcal{L}_{\mathrm{sep}}$

Step 3: Propose the DSV loss w/ $\mathcal{L}_{\mathrm{dis}}$ and $\mathcal{L}_{\mathrm{sep}}$

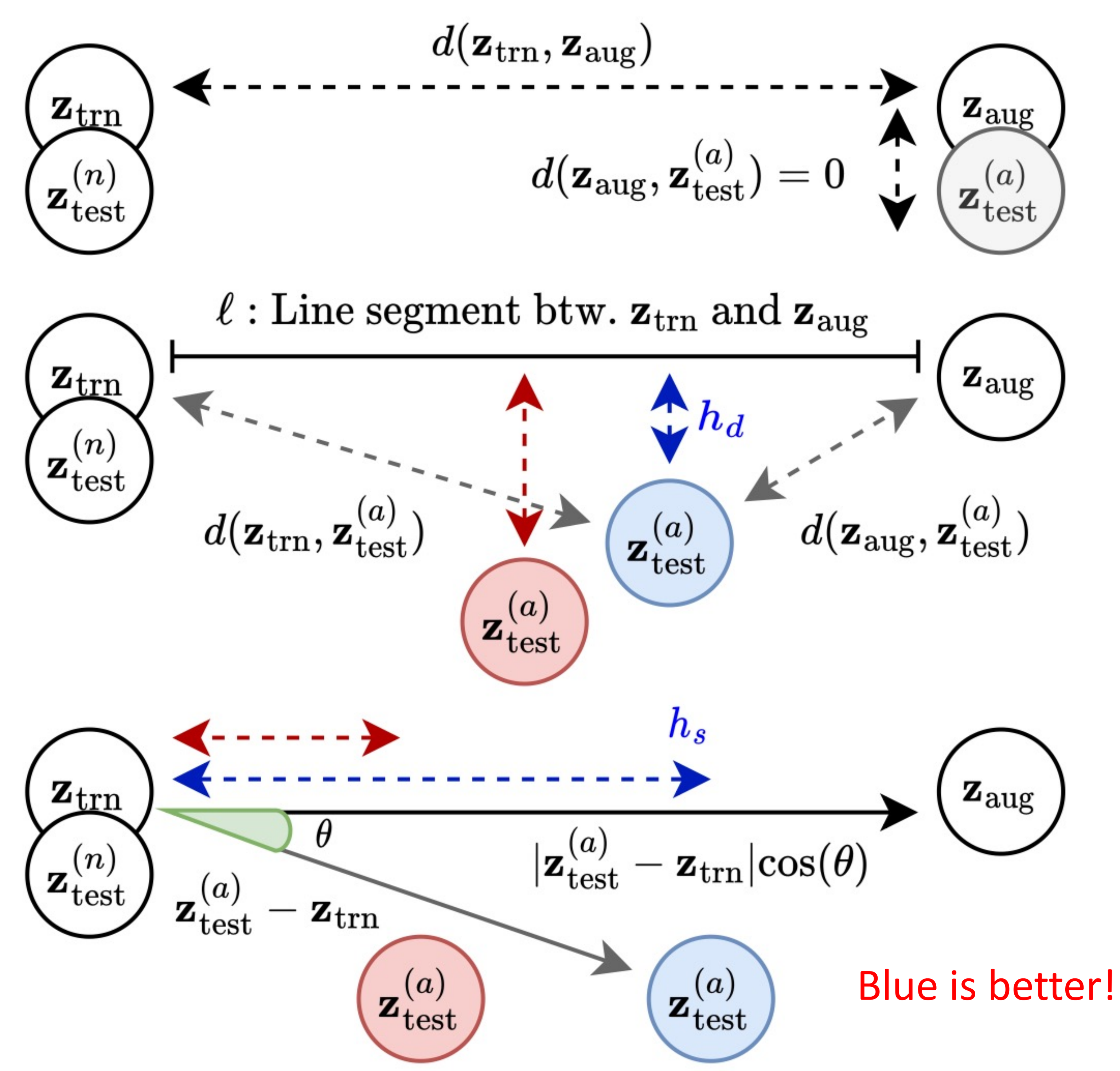
Proposed Method

Discordance: The distance between ℓ and $\mathcal{Z}_{\mathrm{test}}^{(a)}$
Separability: The dist. between $\mathcal{Z}_{\mathrm{trn}}$ and $\mathcal{Z}_{\mathrm{test}}^{(a)}$ on ℓ

Surrogate losses:

$$\mathcal{L}_{\mathrm{dis}} = \frac{d(\mathcal{Z}_{\mathrm{trn}} \cup \mathcal{Z}_{\mathrm{aug}}, \mathcal{Z}_{\mathrm{test}})}{d(\mathcal{Z}_{\mathrm{trn}}, \mathcal{Z}_{\mathrm{aug}})} \qquad \mathcal{L}_{\mathrm{sep}} = \frac{\mathrm{std}(\mathrm{proj}(\mathcal{Z}_{\mathrm{test}}))}{d(\mathcal{Z}_{\mathrm{trn}}, \mathcal{Z}_{\mathrm{aug}})}$$

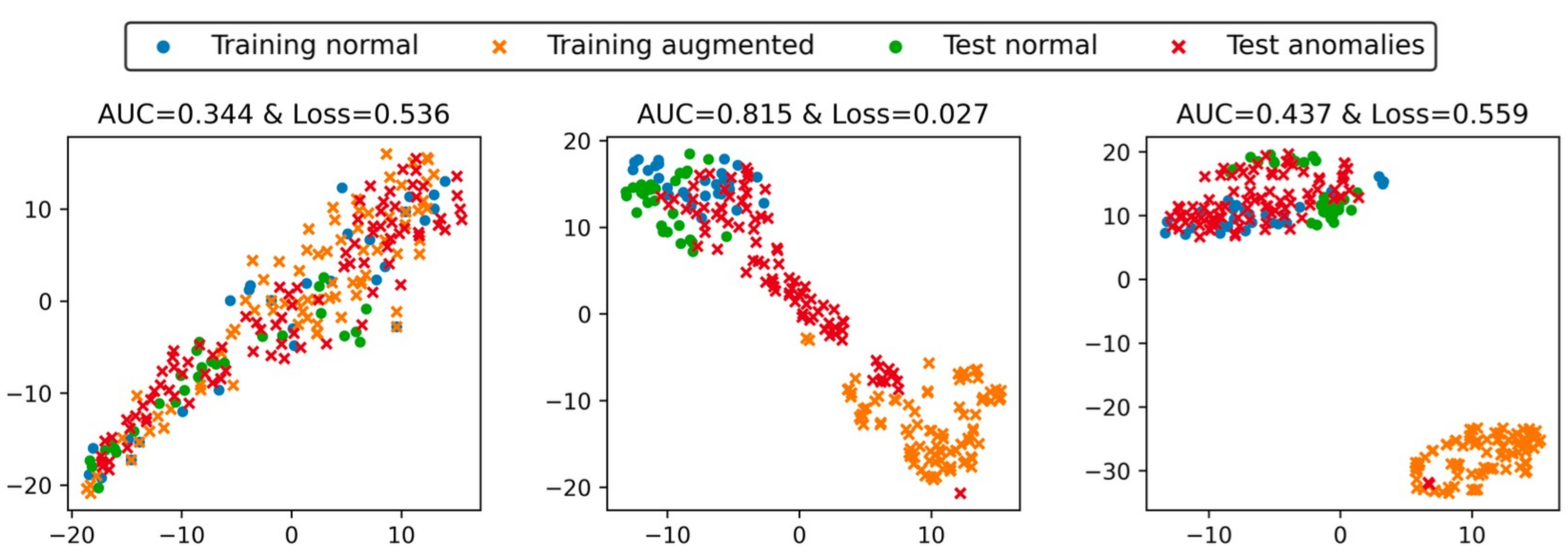
Simplified final loss: $\mathcal{L}_{\mathrm{DSV}} = \mathcal{L}_{\mathrm{dis}} - \mathcal{L}_{\mathrm{sep}} / \mathcal{L}_{\mathrm{dis}}$



Experiments

Task: Industrial image anomaly detection
Detector: ResNet18-based classifier model
Trained to minimize the cross entropy between normal and augmented data
Augmentation functions: CutOut, CutPaste, etc.
Target HPs: Patch size of augmentation

f_{aug}	Avg.	Rand.	Base	MMD	STD	MC	SEL	HITS	**DSV**
CutOut	0.739	0.776	0.741	0.735	0.739	0.749	0.727	0.757	**0.813**
CutAvg	0.739	**0.817**	0.721	0.692	0.745	0.751	0.744	0.742	0.806
CutDiff	0.743	0.711	0.739	0.730	0.744	0.747	0.741	0.777	**0.811**
CutPaste	0.788	0.841	0.694	0.756	0.818	0.862	0.830	0.850	**0.884**



(a) CutOut (0.00004) (b) CutOut (0.00128) (c) CutOut (0.04)