CSI: Novelty Detection via Contrastive Learning on Distributionally Shifted Instances 👮

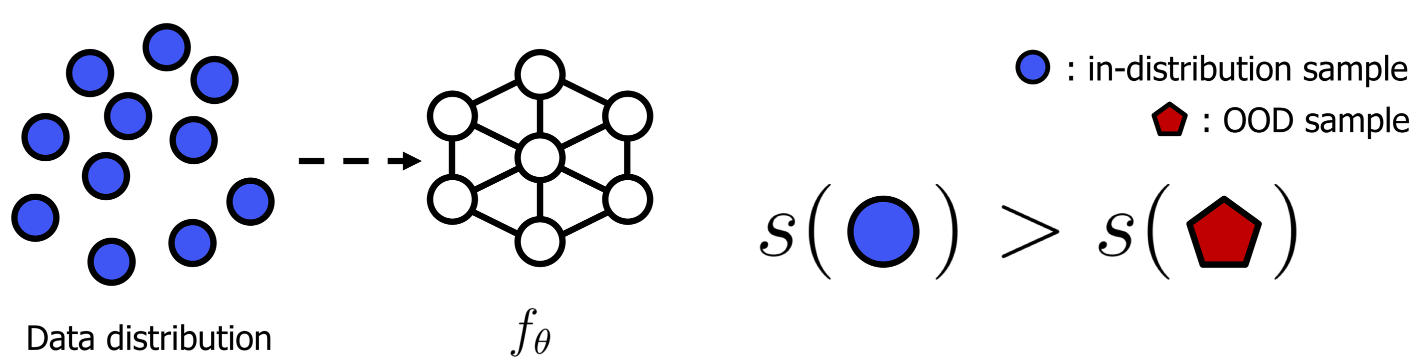
Jihoon Tack*, Sangwoo Mo*, Jongheon Jeong, Jinwoo Shin (* equal contribution)
Korea Advanced Institute of Science and Technology (KAIST)

Paper: https://arxiv.org/abs/2007.08176
Code: https://github.com/alinlab/CSI

TL;DR. We propose a novel contrastive learning scheme for out-of-distribution (OOD) detection, which contrasts hard (distribution-shifting) augmentations to improve in-vs-out discriminability

Introduction

Out-of-distribution (OOD) (novelty, or anomaly) detection is a task of identifying whether a given sample belongs to the data distribution



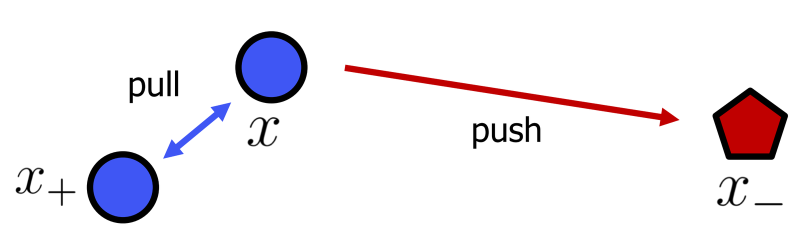
General approach. Most recent approaches tackle the problem
- by learning a representation $f_\theta(\cdot)$ from the data distribution
- then define a detection score $s(\cdot)$ upon the learned representation

Motivation. Inspired by the recent success of self-supervised learning for OOD detection [1], we aim to utilize the power of contrastive learning, the state-of-the-art method for representation learning [2]

Contribution. We propose (a) new contrastive learning scheme and (b) new detection score which utilizes the learned contrastive representation

Contrastive Learning

Contrastive learning encodes the inductive bias of data by pulling similar samples (positives) and pushing the dissimilar samples (negatives)



We consider simple contrastive learning (SimCLR) [2]:
- pull the same samples but with different augmentations (x_i, x_j)
- push the different samples in the batch $\{x_k\}$ for $k \neq i$

For representation $z(x)$ of a sample x, SimCLR loss is given by:

$$-\log \frac{\exp(\mathrm{sim}(z(x_i), z(x_j))/\tau)}{\sum_{k\neq i}\exp(\mathrm{sim}(z(x_i), z(x_k))/\tau)}$$

τ: temperature hyperparameter

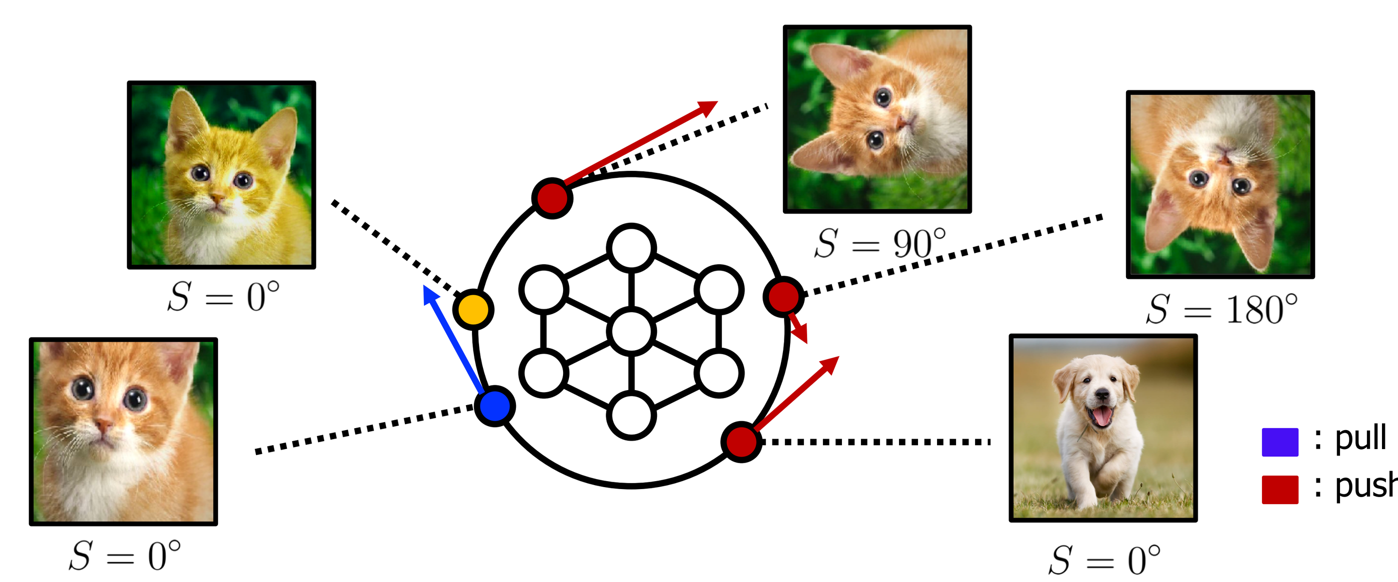
[1] Hendrycks et al. "Using Self-Supervised Learning Can Improve Model Robustness and Uncertainty". NeurIPS 2019.
[2] Chen et al. "A Simple Framework for Contrastive Learning of Visual Representations". ICML 2020.
[3] Khosla et al. "Supervised Contrastive Learning". NeurIPS 2020.

Contrasting Shifted Instances (CSI)

Hard (distribution-shifting) augmentations (e.g., rotation)
- …was known to be harmful and unused for standard contrastive learning
- …turns out to be effective for OOD detection!

Intuition. Distributionally-shifted samples are "nearby, but not too nearby" outliers, hence help the model to discriminate in- vs. out-of-distribution

Representation learning. Contrast the distributionally-shifted samples of itself in addition to the different samples
- contrast: use shifted samples a negative for contrastive learning
- classify: train an auxiliary classifier for transformations (as in [1])



Detection score. For a given sample x, we define the detection score s_{con} for contrastive representation as a combination of two features:
- cosine similarity to the nearest training sample in $\{x_m\}$
- norm of the representation $z(x)$

$$s_{\text{con}}(x; \{x_m\}) := \max_m \ \mathrm{sim}(z(x_m), z(x)) \cdot \| z(x) \| .$$

We further improve the score by incorporating shifting transformations:
- $s_{\text{con-SI}}$: Ensemble s_{con} over the shifting transformations
- $s_{\text{cls-SI}}$: Confidence of the auxiliary transformation classifier

$$s_{\texttt{CSI}}(x; \{x_m\}) := s_{\texttt{con-SI}}(x; \{x_m\}) + s_{\texttt{cls-SI}}(x)$$

OOD-ness: How to choose the shifting transformation? We choose the most OOD-like yet semantically meaningful transformation, measured by the AUROC between original vs transformed samples

Extension to confident-calibrated classifiers. We also adapt CSI for supervised contrastive learning (SupCLR) [3] to calibrate classifiers

Main Results

CSI achieves the state-of-the-art performance for all tested scenarios:
(1) unlabeled one-class, (2) unlabeled multi-class, (3) labeled multi-class

(a) One-class CIFAR-10

Method	Network	Plane	Car	Bird	Cat	Deer	Dog	Frog	Horse	Ship	Truck	Mean
OC-SVM* [64]	-	65.6	40.9	65.3	50.1	75.2	51.2	71.8	51.2	67.9	48.5	58.8
DeepSVDD* [60]	LeNet	61.7	65.9	50.8	59.1	60.9	65.7	67.7	67.3	75.9	73.1	64.8
AnoGAN* [63]	DCGAN	67.1	54.7	52.9	54.5	65.1	60.3	58.5	62.5	75.8	66.5	61.8
OCGAN* [55]	OCGAN	75.7	53.1	64.0	62.0	72.3	62.0	72.3	57.5	82.0	55.4	65.7
Geom* [17]	WRN-16-8	74.7	95.7	78.1	72.4	87.8	87.8	83.4	95.5	93.3	91.3	86.0
Rot* [27]	WRN-16-4	71.9	94.5	78.4	70.0	77.2	86.6	81.6	93.7	90.7	88.8	83.3
Rot+Trans* [27]	WRN-16-4	77.5	96.9	87.3	80.9	92.7	90.2	90.9	96.5	95.2	93.3	90.1
GOAD* [2]	WRN-10-4	77.2	96.7	83.3	77.7	87.8	87.8	90.0	96.1	93.8	92.0	88.2
Rot [27]	ResNet-18	78.3±0.2	94.3±0.3	86.2±0.4	80.8±0.6	89.4±0.5	89.0±0.4	88.9±0.4	95.1±0.2	92.3±0.3	89.7±0.3	88.4
Rot+Trans [27]	ResNet-18	80.4±0.3	96.4±0.2	85.9±0.3	81.1±0.5	91.3±0.3	89.6±0.3	89.9±0.3	95.9±0.1	95.0±0.1	92.6±0.2	89.8
GOAD [2]	ResNet-18	75.5±0.3	94.1±0.3	81.8±0.5	72.0±0.3	83.7±0.9	84.4±0.3	82.9±0.8	93.9±0.3	92.9±0.3	89.5±0.2	85.1
CSI (ours)	ResNet-18	**89.9**±0.1	**99.1**±0.0	**93.1**±0.2	**86.4**±0.2	**93.9**±0.1	**93.2**±0.2	**95.1**±0.1	**98.7**±0.0	**97.9**±0.0	**95.5**±0.1	**94.3**

(b) One-class CIFAR-100 (super-class)

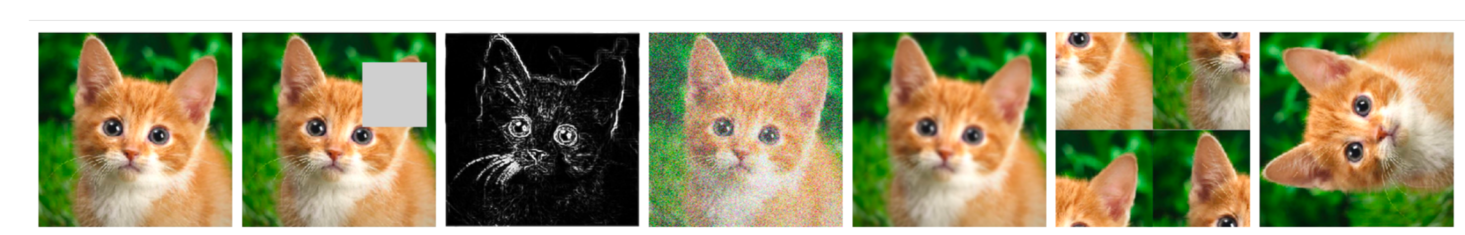
Method	Network	AUROC
OC-SVM* [64]	-	63.1
Geom* [17]	WRN-16-8	78.7
Rot [27]	ResNet-18	77.7
Rot+Trans [27]	ResNet-18	79.8
GOAD [2]	ResNet-18	74.5
CSI (ours)	ResNet-18	**89.6**

(c) One-class ImageNet-30

Method	Network	AUROC
Rot* [27]	ResNet-18	65.3
Rot+Trans* [27]	ResNet-18	77.9
Rot+Attn* [27]	ResNet-18	81.6
Rot+Trans+Attn* [27]	ResNet-18	84.8
Rot+Trans+Attn+Resize* [27]	ResNet-18	85.7
CSI (ours)	ResNet-18	**91.6**

Effects of Shifting Transformations

Measure the (a) OOD-ness and (b) OOD detection performance applied on CSI for various transformations (rotation is the best for CIFAR-10)



(a) Original (b) Cutout (c) Sobel (d) Noise (e) Blur (f) Perm (g) Rotate

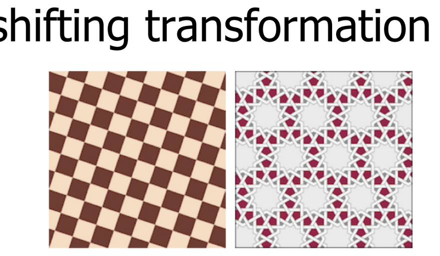
(a) OOD-ness of various transformations

	Cutout	Sobel	Noise	Blur	Perm	Rotate
OOD-ness	79.5	69.2	74.4	76.0	83.8	85.2

(b) OOD detection performance of various transformations, applied on CSI

Base		Cutout	Sobel	Noise	Blur	Perm	Rotate
87.9	+Align	84.3	85.0	85.5	88.0	73.1	76.5
	+Shift	88.5	88.3	89.3	89.2	90.7	94.3

The best shifting transformation depends on the datasets (e.g., for textile)



(a) OOD-ness		(b) AUROC		
Rot.	Noise	Base	CSI(R)	CSI(N)
50.6	**75.7**	70.3	65.9	**80.1**

Open question. Which transformation should be (or not be) contrasted?