Learning and Evaluating Representations for Deep One-class Classification

Google Cloud AI Research

Kihyuk Sohn*, Chun-Liang Li*, Jinsung Yoon, Minho Jin, Tomas Pfister

One-class Classification

- Finding decision boundary between “inlier” and “outlier” without knowing outlier distribution.

- Many real-world applications
    - Anomaly detection (e.g., fraud detection, manufacturing defect detection).
    - Out-of-distribution detection.

Existing methods

- Discriminative models
  - One-class Support Vector Machine [Scholkopf, 2000].
  - Support Vector Data Description [Tax and Duin, 2004].

- Generative models
  - Kernel Density Estimation (KDE) [Latecki, 2007].
  - Auto-encoder [Zong, 2018], flow-based density estimators [Kirichenko, 2020], ...

→ Require good representations.

Existing methods

- End-to-end learning of deep representation and classifier
  - Deep one-class classification [Ruff, 2018]
    - Hypersphere collapse
    - Require additional training technique to avoid degenerate solutions
  - Outlier exposure [Hendrycks, 2019]
    - Mismatch between proxy task and downstream one-class classification task.

➡ How do we achieve the best of both worlds?

Our contribution

- **A two-stage framework for deep one-class classification**
    - Improved learning of self-supervised representations.
    - Discriminative and generative one-class classifiers on deep representations.
- State-of-the-art performance on semantic one-class classification
    - CIFAR-10/100, Fashion MNIST, Cat-vs-Dog, CelebA.
    - Visual explanation of deep one-class classifier.
    - Evaluation on semi-supervised and unsupervised one-class classification settings.
- Application to manufacturing defect detection
    - Strong performance improvement on MVTec anomaly detection dataset on anomaly detection and localization.

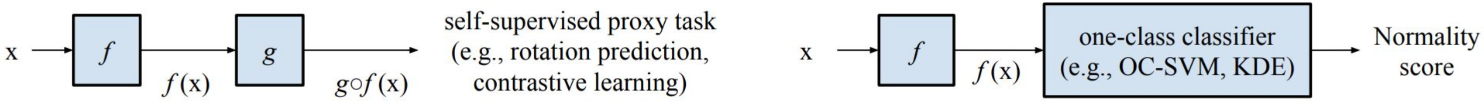
Our contribution

- A two-stage framework for deep one-class classification
  - Improved learning of self-supervised representations.
  - Discriminative and generative one-class classifiers on deep representations.
- **State-of-the-art performance on semantic one-class classification**
  - CIFAR-10/100, Fashion MNIST, Cat-vs-Dog, CelebA.
  - Visual explanation of deep one-class classifier.
  - Evaluation on semi-supervised and unsupervised one-class classification settings.
- Application to manufacturing defect detection
  - Strong performance improvement on MVTec anomaly detection dataset on anomaly detection and localization.

Our contribution

- A two-stage framework for deep one-class classification
  - Improved learning of self-supervised representations.
  - Discriminative and generative one-class classifiers on deep representations.
- State-of-the-art performance on semantic one-class classification
  - CIFAR-10/100, Fashion MNIST, Cat-vs-Dog, CelebA.
  - Visual explanation of deep one-class classifier.
  - Evaluation on semi-supervised and unsupervised one-class classification settings.
- **Application to manufacturing defect detection**
  - Strong performance improvement on MVTec anomaly detection dataset on anomaly detection and localization.

Two-stage framework for Deep One-class Classification



(a) Self-supervised representation learning (b) One-class classifier

a) Representation learning via self-supervision
    - Rotation prediction, contrastive learning, ...
b) Off-the-shelf one-class classifier on learned representations
    - One-class SVM, Kernel Density Estimation (KDE), ...

- Why two stage?
    - High-quality deep representation + faithful one-class classifier.

Improving Self-supervised Representations

- Rotation prediction [Gidaris, 2018, Golan and El-Yaniv, 2018]
  - Adopt MLP projection head at representation learning.

- Contrastive learning of representations [Chen, 2020]
  - Less compatible for OCC (e.g., class collision [Saunshi, 2019] and uniformity [Wang and Isola, 2020]).
  - ~~Large~~ Small batch training.
  - Distribution augmentation with geometric transformations (DA-CLR).

Experiments

- CIFAR-10/100, Fashion-MNIST, Cat-vs-Dog, CelebA
- Data from a single class as inlier, rest as outlier

Inlier

Outlier
(not available at training)

End-to-end vs Two-stage framework

AUC ↑	C10	C100	f-MNIST	Cat-vs-Dog	CelebA
Rotation classifier	86.8	80.3	87.4	86.1	51.4
KDE	**89.3**	**81.9**	**94.6**	**86.4**	**77.4**

Two-stage framework **>** End-to-end

MLP projection head

AUC ↑	C10	C100	f-MNIST	Cat-vs-Dog	CelebA
Rotation classifier	86.8	80.3	87.4	86.1	51.4
KDE	89.3	81.9	94.6	86.4	**77.4**
+ MLP head	**91.3**	**84.1**	**95.8**	**86.4**	69.5

MLP projection head ***improves*** the performance of Rotation prediction representations.

Contrastive Representation

AUC ↑	C10	C100	f-MNIST	Cat-vs-Dog	CelebA
Rotation classifier	86.8	80.3	87.4	86.1	51.4
KDE	89.3	81.9	94.6	86.4	77.4
+ MLP head	**91.3**	**84.1**	**95.8**	86.4	69.5
Contrastive	89.0	82.4	93.6	**87.7**	**84.6**

Contrastive representation is ***not as good as***
Rotation prediction representation with MLP head.

Distribution-augmented Contrastive Representation

AUC ↑	C10	C100	f-MNIST	Cat-vs-Dog	CelebA
Rotation classifier	86.8	80.3	87.4	86.1	51.4
KDE	89.3	81.9	94.6	86.4	77.4
+ MLP head	91.3	84.1	**95.8**	86.4	69.5
Contrastive	89.0	82.4	93.6	87.7	84.6
Contrastive (DA)	**92.4**	**86.5**	94.5	**89.6**	**85.6**

Distribution-augmented Contrastive representation performs the ***best***.

Visual Explanation

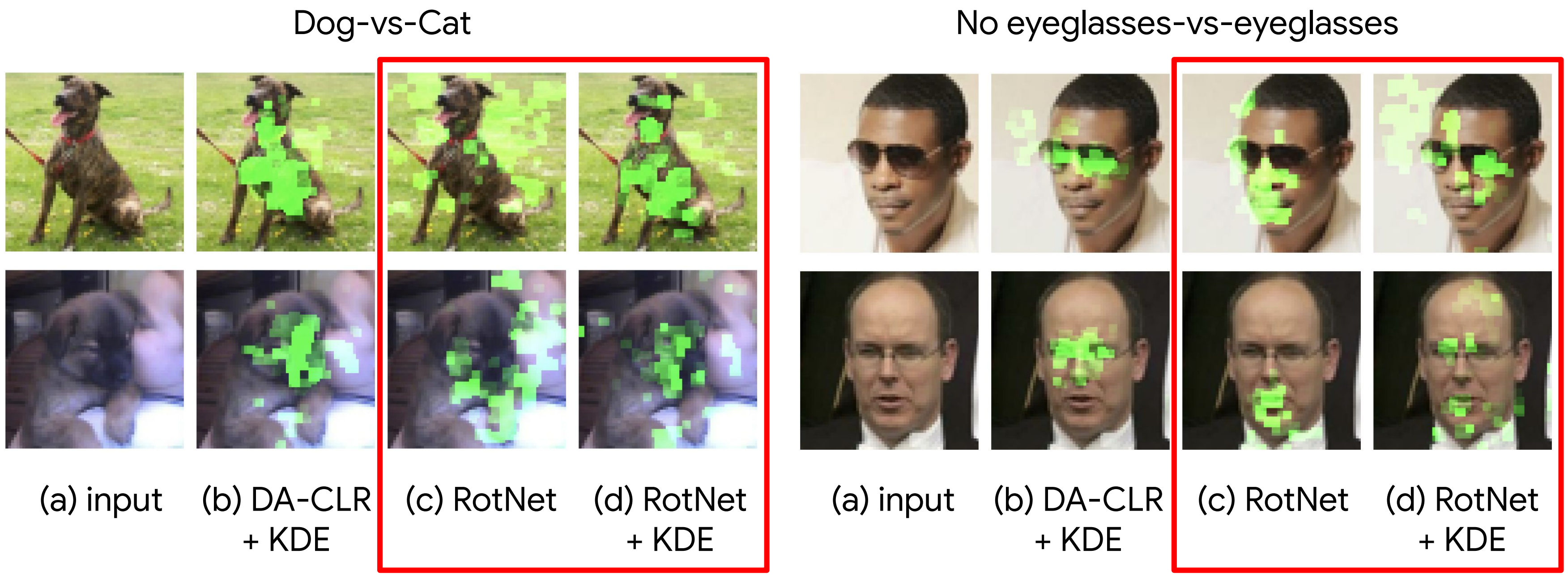
- Gradient-based visual explanation techniques (e.g., **Integrated Gradient**, GradCAM) can be used.

Dog-vs-Cat | No eyeglasses-vs-eyeglasses

(a) input (b) DA-CLR + KDE (c) RotNet (d) RotNet + KDE | (a) input (b) DA-CLR + KDE (c) RotNet (d) RotNet + KDE

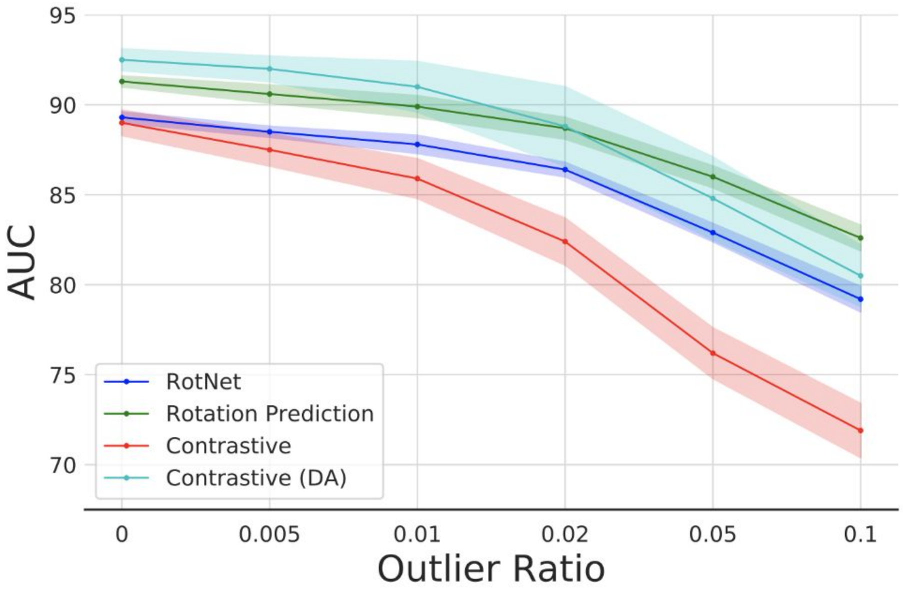
Visual Explanation

- Gradient-based visual explanation techniques (e.g., **Integrated Gradient**, GradCAM) can be used.



Realistic Evaluation of One-class Classification

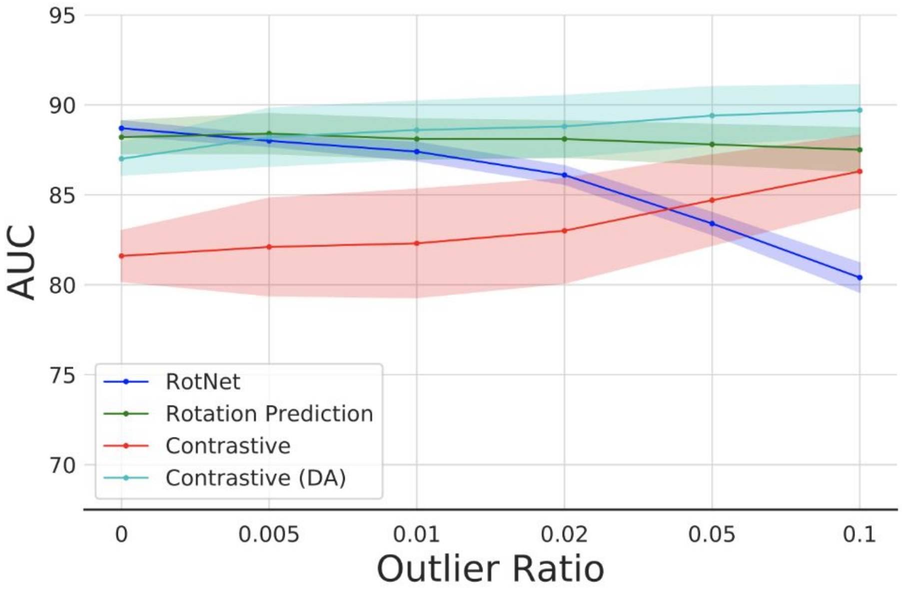
- Unsupervised: training data contains a small fraction of outliers



- Rotation prediction is more robust than contrastive representation.

Realistic Evaluation of One-class Classification

- Semi-supervised: unsupervised representation + OCC with **50** inlier data



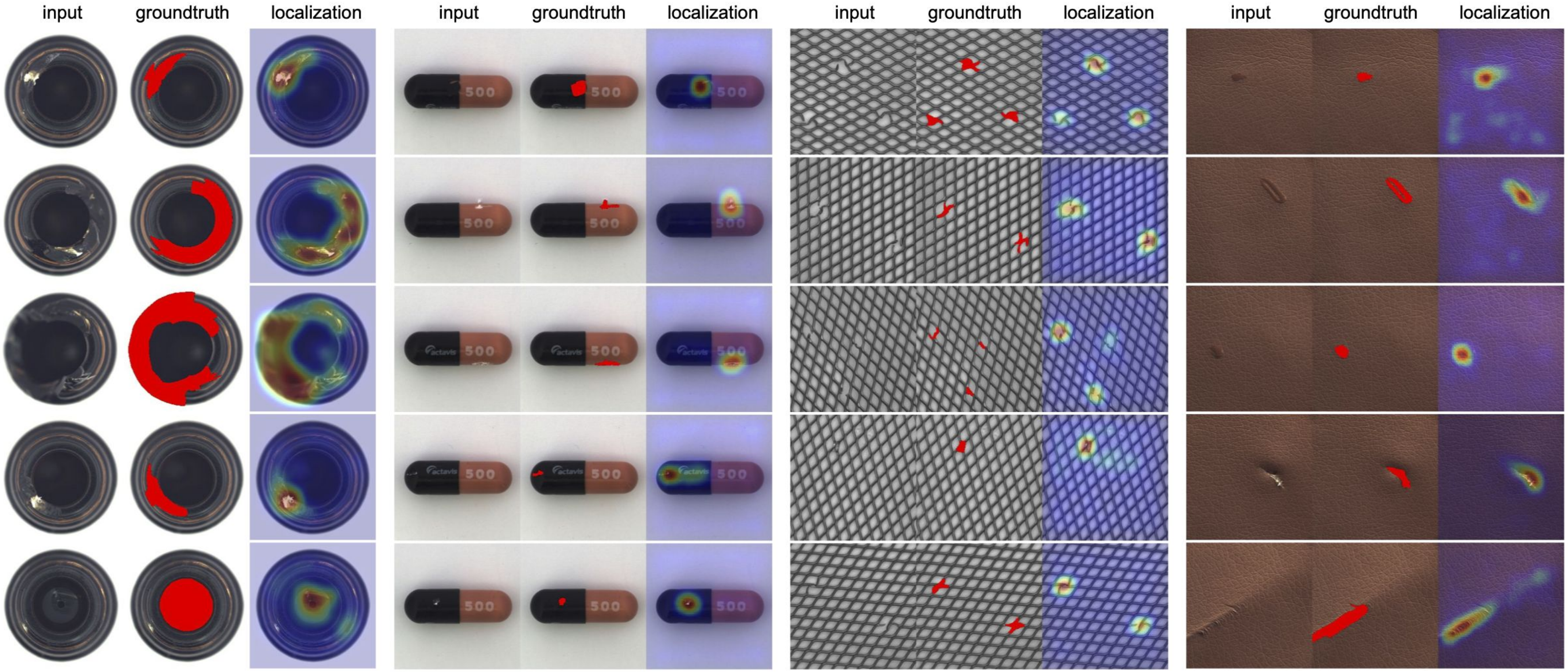
- contrastive representation outperforms rotation prediction.
- The more noisy, the better the contrastive representation.

Experiments on MVTec Anomaly Detection

	RotNet	RotNet + KDE	+ MLP head	Contrastive	Contrastive (DA)
Detection	71.0	83.5	**86.3**	80.2	**86.5**
Localization	75.6	**92.6**	**93.0**	85.6	90.4

- ***Effectiveness*** of two-stage framework, MLP head, distribution-augmented contrastive representation holds.
- ***No clear advantage*** of contrastive representation.

Visualization on MVTec Anomaly Localization



Conclusion

- A two-stage framework for deep one-class classification.
- State-of-the-art performance on semantic one-class classification.
- Strong improvement on manufacturing defect detection and localization.

Stay Tuned!

- CutPaste: Self-supervised learning for anomaly detection and localization.
  - To appear at CVPR 2021.
  - State-of-the-art defect detection and localization performance via two-stage framework + novel distribution augmentation strategy.