LEARN DISCRIMINATION FROM CONTAMINATED DATA: MULTI-INSTANCE LEARNING WITH CONVEX REPRESENTATION FOR UNSUPERVISED ANOMALY DETECTION

Paper ID: 702

ABSTRACT

Anomaly detection aims at identifying deviant samples from the normal data distribution. Much progress has been made in recent years for anomaly detection with a variety of methods with self-supervised representation learning. However, most existing approaches assume the training set contains either only clean normal samples or some labeled abnormal samples. With a contaminated unlabeled training set, the performance is degraded with unclear discrimination between normal and abnormal samples. To address the above challenge, in this paper, we propose a novel unsupervised representation learning framework that takes advantage of the extra information provided by the anomalies in the unlabeled contaminated data for anomaly detection. Specifically, anomaly discrimination learning with pseudo normality score generation and multi-instance contrastive learning is proposed for distinguishing abnormal samples from the unlabeled set. Meanwhile, we combine the discrimination learning with mean-shifted contrastive learning and distribution-shifting transformation classification into a multi-task representation learning framework to improve the stability and robustness of the training. The proposed method achieves state-of-the-art performance on CIFAR-10 and MNIST datasets. Extensive ablation studies further demonstrate the effectiveness of different components of the proposed framework.

***Index Terms*—** Anomaly detection, unsupervised, multi-instance learning, contaminated data

1. INTRODUCTION

Anomaly detection, also referred to as outlier detection, is a task of identifying samples that significantly differ from the majority. As a general but essential problem, anomaly detection has a wide range of applications, including medical image analysis [1, 2], video surveillance [3, 4], and fraud detection [5]. Some self-supervised anomaly detection methods [6, 7, 8] assume the training set only contains clean normal data, while some works [9, 10, 11, 12] combine labeled abnormal samples in the training. Under such settings, much progress has been made in recent years for anomaly detection with a variety of methods, such as reconstruction-based [13, 14, 15] and contrastive learning-based approaches [6, 16, 12, 17].

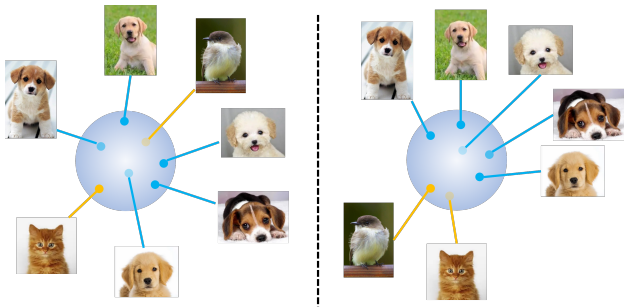

Fig. 1. Challenges of representation learning for unsupervised anomaly detection with contaminated data. Light blue and yellow color represent normal and abnormal classes, respectively. Existing representation learning approaches fail to learn discriminative representation between normal and abnormal samples, while our method aims to separate anomalies from contaminated data.

However, obtaining a set of clean normal data or labeled abnormal data for training is difficult in practical scenarios since labeling is usually expensive. Although some methods utilize a large pre-training dataset to generate good features [18], the result degrades greatly with the presence of abnormal samples in the training set. In real-world scenarios, it is more practical to access a large amount of contaminated unlabeled samples where normal samples are dominant. With such an extreme imbalanced set without labels, it is difficult to design proper unsupervised objective functions that are suitable for clustering, representation learning, and insensitive to cluster densities at the same time, leading to sub-optimal solutions without clear discrimination between normal and abnormal samples, as shown in Fig. 1.

To address the challenges raised from anomaly detection with unlabeled contaminated samples, in this paper, we propose a novel unsupervised representation learning framework that takes account of the information from both unlabeled normal and abnormal samples in the training. The framework is shown in Fig. 2. Firstly, anomaly discrimination learning with pseudo normality score generation and multi-instance contrastive learning is proposed for distinguishing abnormal samples from the unlabeled set. Meanwhile, we combine the discrimination learning with mean-shifted contrastive learning and distribution-shifting transformation classification into

a multi-task representation learning framework to improve the stability and robustness of the training. We achieve state-of-the-art (SOTA) performance on CIFAR-10 and MNIST datasets.

Our main contributions are summarized as follows: 1) We present a novel unsupervised anomaly detection method that combines pseudo normality score generation and multi-instance contrastive learning with convex representation with unlabeled contaminated data. 2) We formulate a multi-task representation learning framework that improves the discrimination, robustness, and stability of the learning. 3) We conduct extensive experiments with systematic analysis. The SOTA performance on CIFAR-10 and MNIST is achieved.

2. RELATED WORK

Unsupervised Clustering. Unsupervised clustering methods group samples without ground truth labels. It is a challenging task to learn sample representation and conduct clustering simultaneously. To this end, much progress has been made recently. For example, [19] groups the features with k-means and update the weights of the network based on the clustering result. [20] avoids degenerated solutions which are very common in most existing clustering methods by maximizing mutual information between augmented data pairs. [21] combines contrastive learning into clustering by regarding class prediction as cluster representation, then optimizes the instance- and cluster-level contrastive loss simultaneously. [22] proposes a novel method of prototype pseudo-labeling and reliable pseudo-labeling, which exploits semantic information to help the training of the clustering head. However, most works in recent unsupervised clustering assume that the dataset is perfectly balanced across different categories. They are sensitive to different data densities when the training set is extremely imbalanced.

Anomaly Detection. Learning informative sample representations is critical for anomaly detection. With the fast progress in self-supervised learning, many works have been proposed in recent years. For example, [16] presents a novel approach with distribution augmentation to reduce the uniformity of normal distribution on a hypersphere and ensures the compactness of the sample representation. [6] further extends the concept of distribution augmentation with contrastive learning and raises an effective score function for anomaly detection. [17] proposes a re-centering method that changes the origin to the center of features extracted from a pre-trained dataset in the angular distance measurement. [23] maximizes the distance between normal and anomalous samples in the joint distribution based on information theory. [24] utilizes adversarial interpolated samples to train a one-class Gaussian classifier that differentiates the anomalies based on their distance to the Gaussian center as well as the standard deviation of these distances. However, training with contaminated data for anomaly detection still needs more exploration.

3. PROPOSED METHOD

In this section, we firstly describe the unsupervised contaminated setting of anomaly detection. Then we demonstrate how to generate pseudo normality scores based on the similarity measures among neighbor samples. In addition, we illustrate how we design a soft sampling strategy and multi-instance contrastive learning with convex representation for robust discriminative learning to distinguish abnormal samples. Meanwhile, we combine the existing mean-shifted loss and distribution-shifting classification loss into a multi-task learning framework to improve the robustness of the representation.

3.1. Problem Description

Let $\mathcal{X} = \{\boldsymbol{x}_i\}_{i=1}^{N}$ denote the contaminated training set, consisting of a majority of normal samples and a small portion of abnormal samples without access to ground truth labels. Denote γ as the contamination ratio which represents the abnormal sample ratio in the unlabeled training set. Let $f_{\boldsymbol{\Theta}}(\cdot)$ denote the encoder with learnable parameters $\boldsymbol{\Theta}$ for image representation. We aim at obtaining a discriminative encoder for representation learning that can distinguish the anomalies in the testing data with the contaminated training set.

3.2. Anomaly Discriminative Learning

Distinguishing abnormal samples from the unlabeled contaminated set is a difficult task. To learn the discrimination of abnormal samples, we first present a pseudo normality score generation approach with the averaged similarity among first K nearest neighbors to measure the density of each sample. Then we sample the instances with the distribution of designed pseudo normality scores and use a convex combination of multiple samples to learn the contrastive relations between normal and abnormal samples. Such multi-instance learning with convex representation can largely reduce the impact from the sampling noise and stabilize the training to improve the discrimination of anomalies.

3.2.1. Pseudo Normality Score Generation

Denote $\boldsymbol{Z} = f_{\boldsymbol{\Theta}}(\mathcal{X}_{\mathcal{B}}) \in \mathbb{R}^{N \times D}$ as the representation of a batch data $\mathcal{X}_{\mathcal{B}}$, where N and D are the number of the batch size and feature dimension, respectively. Assume $\boldsymbol{Z}$ is already normalized with Euclidean distance. To measure the relationships among batch samples, we define the similarity matrix as follows,

$$\boldsymbol{M} = \boldsymbol{Z}\boldsymbol{Z}^T, \tag{1}$$

where $\boldsymbol{M}_{i,j}$ represents the cosine similarity between samples $\boldsymbol{z}_i$ and $\boldsymbol{z}_j$.

Based on the assumption that the majority of samples are from normal class, then the isolated samples are more likely

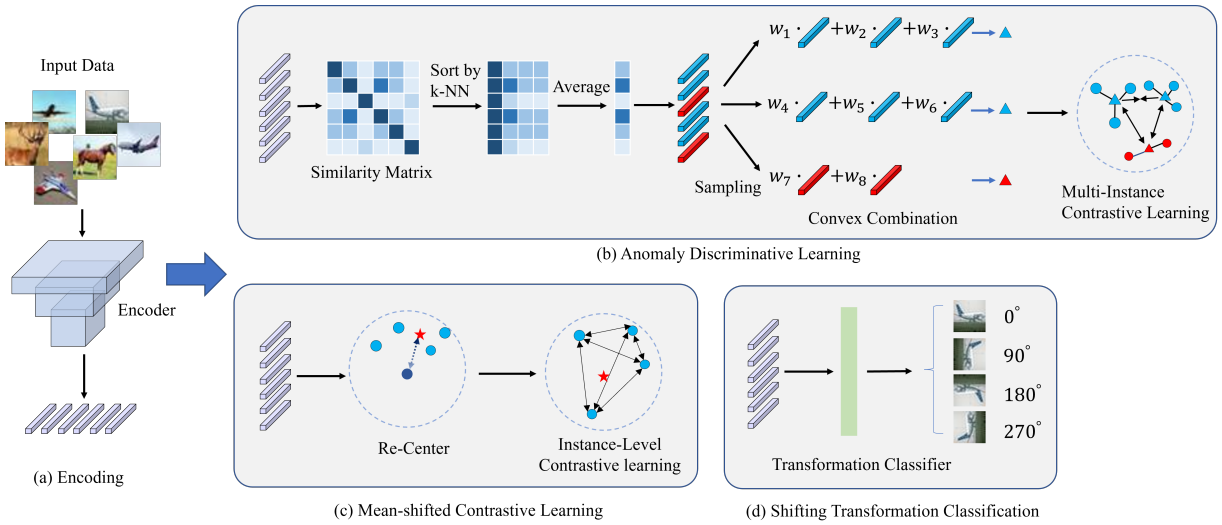

Fig. 2. The proposed framework for unsupervised anomaly detection with contaminated data. The encoder and representations are learned based on a multi-task learning framework with the designed anomaly discriminative learning component, as well as another two components from existing works, *i.e.*, mean-shifted contrastive learning and shifting transformation classification. More details are demonstrated in Sec. 3.

to be abnormal samples. This is also a widely used assumption in density-based clustering [25]. As a result, we can use the similarity of the first K nearest neighbors to generate the normality scores. The normality score is defined as follows,

$$s_i = \frac{1}{K} \sum_{j \in \mathcal{N}_K(i)} \boldsymbol{M}_{i,j}, \tag{2}$$

where $\mathcal{N}_K(i)$ represents the K nearest neighbors of sample i.

3.2.2. Multi-Instance Contrastive Learning with Convex Representation

Based on the generated normality scores, we can divide samples into normal and abnormal sets, then followed by the InfoNCE loss [26, 27, 6],

$$\mathcal{L}_{\text{InfoNCE}} = -\log \frac{\sum_{\boldsymbol{x}' \in \mathcal{X}^+} \exp\left(\text{sim}(f_{\boldsymbol{\Theta}}(\boldsymbol{x}), f_{\boldsymbol{\Theta}}(\boldsymbol{x}'))/\tau\right)}{\sum_{\boldsymbol{x}' \in (\mathcal{X}^+ \cup \mathcal{X}^-)} \exp\left(\text{sim}(f_{\boldsymbol{\Theta}}(\boldsymbol{x}), f_{\boldsymbol{\Theta}}(\boldsymbol{x}'))/\tau\right)}, \tag{3}$$

where $\mathcal{X}^+$ and $\mathcal{X}^-$ represents the sampled normal and abnormal instances, respectively; τ is the temperature parameter; cosine similarity is commonly used for $\text{sim}(\cdot,\cdot)$. However, the normality scores can be very noisy, leading to large false positives and false negatives in the sampling of $\mathcal{X}^+$ and $\mathcal{X}^-$.

To address such issues, we end up with a soft sampling strategy. We divide the batch data into a normal candidate set $\mathcal{S}_n$ and abnormal candidate set $\mathcal{S}_a$ according to the normality score s. In practice, we put 10% samples with the lowest normality scores in $\mathcal{S}_a$ and the rest in $\mathcal{S}_n$. Then we define the sampling probability as follows,

$$\boldsymbol{p}_i^+ = \frac{s_{i \in \mathcal{S}_n}}{\sum_{i \in \mathcal{S}_n} s_i}, \quad \boldsymbol{p}_j^- = \frac{s_{j \in \mathcal{S}_a}}{\sum_{j \in \mathcal{S}_a} s_j}, \tag{4}$$

and sample the positive and negative instances with $\mathcal{X}^+{}_{\{\boldsymbol{x} \sim \boldsymbol{p}^+\}}$ and $\mathcal{X}^-{}_{\{\boldsymbol{x} \sim \boldsymbol{p}^-\}}$, respectively.

Meanwhile, multi-instance learning with convex representation is employed to improve learning stability. Denote the weighted averaged representation $\boldsymbol{z}^+$ and $\boldsymbol{z}^-$ as follows,

$$\begin{aligned} \boldsymbol{z}^+ &= \frac{1}{C} \sum_{k=1}^{C} w_k f_{\boldsymbol{\Theta}}(\boldsymbol{x}_k | \boldsymbol{x}_k \sim \boldsymbol{p}^+), \\ \boldsymbol{z}^- &= \frac{1}{C} \sum_{k=1}^{C} w_k f_{\boldsymbol{\Theta}}(\boldsymbol{x}_k | \boldsymbol{x}_k \sim \boldsymbol{p}^-), \end{aligned} \tag{5}$$

where w_k is a non-negative random variable sampled from a uniform distribution with the constraint $\sum_{k=1}^{C} w_k = 1$. In other words, $\boldsymbol{z}^+$ and $\boldsymbol{z}^-$ lie in the convex hull of the sampled instances, and such representations are more likely to contain information from normal and abnormal samples, which is more resistant to noise than single samples. With the newly defined convex representation with multi-instance learning, we update the loss from Eq. (3) to the following,

$$\mathcal{L}_{\text{Conv}} = -\log \frac{\exp(\text{sim}(\boldsymbol{z}_i^+, \boldsymbol{z}_j^+))}{\exp(\text{sim}(\boldsymbol{z}_i^+, \boldsymbol{z}_j^+)) + \exp(\text{sim}(\boldsymbol{z}_i^+, \boldsymbol{z}_k^-))}. \tag{6}$$

3.3. Multi-Task Representation Learning

Apart from the discriminative learning to distinguish abnormal samples from unlabeled contaminated data, we also combine two extra components, *i.e.*, mean-shifted contrastive loss [17] and distribution-shifting transformation classification loss [6] into a multi-task learning framework, which also stabilizes the training.

Specifically, the mean-shifted contrastive loss $\mathcal{L}_{\text{Mean}}$ [17] is defined as follows,

$$\mathcal{L}_{\text{Mean}} = -\log \frac{\sum_{\boldsymbol{x}' \in \mathcal{X}^+} \exp\left(\text{sim}(\tilde{f}_{\boldsymbol{\Theta}}(\boldsymbol{x}), \tilde{f}_{\boldsymbol{\Theta}}(\boldsymbol{x}'))/\tau\right)}{\sum_{\boldsymbol{x}' \in (\mathcal{X}^+ \cup \mathcal{X}^-)} \exp\left(\text{sim}(\tilde{f}_{\boldsymbol{\Theta}}(\boldsymbol{x}), \tilde{f}_{\boldsymbol{\Theta}}(\boldsymbol{x}'))/\tau\right)} - f_{\boldsymbol{\Theta}}(\boldsymbol{x})^T \bar{f}_{\boldsymbol{\Theta}}, \tag{7}$$

where $\mathcal{X}^+$ and $\mathcal{X}^-$ represent augmented copies from the same and different instances, respectively. $\bar{f}_{\boldsymbol{\Theta}}$ is the mean representation of all samples, and $\tilde{f}_{\boldsymbol{\Theta}}(\cdot)$ is the re-centered representation defined as follows,

$$\tilde{f}_{\boldsymbol{\Theta}}(\boldsymbol{x}) = \frac{f_{\boldsymbol{\Theta}}(\boldsymbol{x}) - \bar{f}_{\boldsymbol{\Theta}}}{||f_{\boldsymbol{\Theta}}(\boldsymbol{x}) - \bar{f}_{\boldsymbol{\Theta}}||}. \tag{8}$$

The first term in $\mathcal{L}_{\text{Mean}}$ learns the contrastive relationship based on re-centered representations, while the second term encourages representations should be close to the center of representations. More details are demonstrated in [17].

Moreover, we include the rotation with degrees in $\{0°, 90°, 180°, 270°\}$ as the distribution-shifting transformation in the data augmentation [6]. A single-layer classifier $g_{\boldsymbol{\Phi}}(\cdot)$ is added with cross-entropy loss adopted for the rotation prediction,

$$\mathcal{L}_{\text{Rot}} = -\sum_{k \in \{0°, 90°, 180°, 270°\}} y^k \log g^k_{\boldsymbol{\Phi}}(\boldsymbol{z}), \tag{9}$$

where y^k is the one-hot label with the corresponding rotation.

With the multi-task representation learning, the total loss is defined as

$$\mathcal{L} = \mathcal{L}_{\text{Conv}} + \lambda_1 \mathcal{L}_{\text{Mean}} + \lambda_2 \mathcal{L}_{\text{Rot}}, \tag{10}$$

where λ_1 and λ_2 are the weights of individual losses.

3.4. Detection Score for Inference

Given a testing sample, we compare its similarity with all training samples and choose the largest one as the detection score to represent the normality of the sample. The detection score is defined as follows,

$$d(\boldsymbol{x}) = \max_m f_{\boldsymbol{\Theta}}(\boldsymbol{x}_m)^T f_{\boldsymbol{\Theta}}(\boldsymbol{x}), \tag{11}$$

where $\boldsymbol{x}$ is the testing sample and $\boldsymbol{x}_m$ is the m-th training sample.

4. EXPERIMENTAL RESULTS

4.1. Implementation Details

Datasets. We run our experiments on two datasets, CIFAR-10 [28] and MNIST [29]. CIFAR-10 consists of 50,000 training and 10,000 testing images with 10 image classes. MNIST has a training set of 60,000 examples, and a testing set of 10,000 examples, with 10 handwriting-digit classes. We treat anomaly detection as one-class classification problem. For each dataset, samples from a one class are regarded as normal data and those from the remaining classes are regarded as anomalous data.

Training Details. We use ResNet-18 [30] architecture in the experiments. For data augmentations $\mathcal{T}$, we use inception crop , horizontal flip, color jitter, and grayscale. Random rotation in $\{0°, 90°, 180°, 270°\}$ is used for shifting transformations. we train the proposed method for each one-class classification model with 200 epochs under Adam optimizer with an initial learning rate of 1e-3. For the learning rate scheduler, we use linear warmup for the early 10 epochs, followed by the cosine decay schedule. The mean-shifted representation from Eq. (8) is conducted after 15 epochs. The model is learned from scratch without any large-dataset pre-training. We use one Nvidia RTX 3090 GPU for training.

Evaluation Metrics. Following [6], we use the area under the receiver operating characteristics (AUROC) score as the evaluation metric. The ROC represents the true positive rate against the false-positive rate, while AUROC is the area under the curve. It is a common summary statistic for the goodness of a predictor in a binary classification task. The higher the score, the better the performance.

4.2. Results

Different Contamination Ratios. We compare our proposed method with the SOTA methods MSC [17], CC [21], and CSI [6] with different contamination ratios γ on CIFAR-10 and MNIST in Table 1 and Table 2, respectively, where the best results are marked in bold. Different from the original paper of MSC that uses a pre-trained model, we learn the encoder of MSC from scratch for a fair comparison with other methods, denoted as MSC* in the table. When only a small number of abnormal samples are available in the training ($\gamma = 2\%$), the results are roughly the same between our method and CSI on CIFAR-10. This is because extra information provided by abnormal samples is very limited and the power of anomaly discriminative learning is not significant. We can see that with the increase of γ, the performance of CC slightly increases since the extreme imbalanced issue between normal and abnormal samples is alleviated. Moreover, the performance of CSI has a significant degradation with the increase of γ, while our proposed method is much more robust with different levels of contamination.

Ablation Study on Different Components. We also conduct ablation study on different components of the loss, with respect to rotation transformation classification (Rot), mean-shifted contrastive learning, and anomaly discriminative learning (ADL). The AUROC results are reported in Table 3 with $\gamma = 5\%$. Without the proposed anomaly discrimination learning, the performance has a significant degradation. The best performance is achieved with the combination

Table 1. AUROC on CIFAR-10 with Different Contamination Ratios

Method (γ)	Plane	Car	Bird	Cat	Deer	Dog	Frog	Horse	Ship	Truck	Mean
MSC* [17] (2%)	63.3	66.4	57.4	59.6	66.5	46.9	55.7	55.3	83.5	67.4	62.2
CC [21] (2%)	66.7	59.8	82.1	65.1	71.8	65.7	57.8	50.5	67.8	70.7	65.8
CSI [6] (2%)	84.1	98.6	87.9	78.8	91.8	87.5	85.0	97.8	96.3	93.3	**90.1**
Ours (2%)	80.5	95.0	83.5	82.5	92.6	89.8	95.1	88.4	94.8	93.5	89.6
MSC* [17] (5%)	61.7	62.6	57.2	49.3	36.2	54.5	35.2	55.1	77.4	76.6	56.6
CC [21] (5%)	70.5	69.5	63.0	62.5	72.5	59.9	65.0	60.8	71.5	72.0	66.7
CSI [6] (5%)	81.9	97.8	81.3	71.2	86.3	81.6	76.0	96.5	93.6	91.3	85.8
Ours (5%)	82.7	97.3	81.6	79.5	91.0	87.5	92.9	95.9	95.2	94.8	**89.8**
MSC* [17] (10%)	63.1	37.5	64.1	54.5	57.5	60.2	51.4	46.7	85.8	68.5	58.9
CC [21] (10%)	73.1	76.6	73.6	66.0	78.4	56.5	73.9	50.9	70.5	71.5	69.1
CSI [6] (10%)	75.0	97.0	74.9	62.9	79.7	73.7	68.5	93.7	90.7	89.0	80.5
Ours (10%)	83.5	97.1	74.7	62.7	89.8	84.0	93.0	88.2	94.3	94.7	**86.2**

Table 2. AUROC on MNIST with Different γ

Method	γ (2%)	γ (5%)	γ (10%)
MSC* [17]	66.3	72.6	52.1
CC [21]	57.7	58.2	60.1
CSI [6]	85.0	76.9	70.2
Ours	**94.1**	**92.3**	**90.1**

Table 3. Component Analysis with γ (5%)

Rot	Mean-shifted	ADL	AUROC
✓	-	-	85.8
✓	✓	-	87.2
✓	-	✓	85.1
✓	✓	✓	**89.8**

of all individual components, demonstrating the effectiveness of our multi-task representation learning framework.

Visualization. We also show one visualization example of our proposed method using t-SNE in Fig. 3. The left and right sub-figures represent the dimension reduced representations before and after training, respectively. After training, abnormal samples tend to be closer to each other and further away from normal samples, which verifies the success of our framework on distinguishing anomalies with unlabeled contaminated data.

5. CONCLUSIONS

In this paper, we improve the performance of unsupervised anomaly detection on contaminated training data with the proposed anomaly discriminative learning with a multi-task learning framework. Our method demonstrates outstanding performance under various contamination ratios and shows great discrimination ability and stability.

Initial Representation Learned Representation

Fig. 3. A visualization example with a comparison between the representation before and after learning using t-SNE.

6. REFERENCES

[1] Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Ursula Schmidt-Erfurth, and Georg Langs, "Unsupervised anomaly detection with generative adversarial networks to guide marker discovery," in *International conference on information processing in medical imaging*. Springer, 2017, pp. 146–157.

[2] Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Georg Langs, and Ursula Schmidt-Erfurth, "f-anogan: Fast unsupervised anomaly detection with generative adversarial networks," *Medical image analysis*, vol. 54, pp. 30–44, 2019.

[3] Ziming Wang, Yuexian Zou, and Zeming Zhang, "Cluster attention contrast for video anomaly detection," in *Proceedings of the 28th ACM International Conference on Multimedia*, 2020, pp. 2463–2471.

[4] Waqas Sultani, Chen Chen, and Mubarak Shah, "Real-world anomaly detection in surveillance videos," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 6479–6488.

[5] Xun Zhou, Sicong Cheng, Meng Zhu, Chengkun Guo, Sida Zhou, Peng Xu, Zhenghua Xue, and Weishi Zhang, "A state of the art survey of data mining-based fraud detection and credit scoring," in *MATEC Web of Conferences*. EDP Sciences, 2018, vol. 189, p. 03002.

[6] Jihoon Tack, Sangwoo Mo, Jongheon Jeong, and Jinwoo Shin, "Csi: Novelty detection via contrastive learning on distributionally shifted instances," *arXiv preprint arXiv:2007.08176*, 2020.

[7] Junbong Kim, Kwanghee Jeong, Hyomin Choi, and Kisung Seo, "Gan-based anomaly detection in imbalance problems," in *European Conference on Computer Vision*. Springer, 2020, pp. 128–145.

[8] Mohammadreza Salehi, Atrin Arya, Barbod Pajoum, Mohammad Otoofi, Amirreza Shaeiri, Mohammad Hossein Rohban, and Hamid R Rabiee, "Arae: Adversarially robust training of autoencoders improves novelty detection," *arXiv preprint arXiv:2003.05669*, 2020.

[9] Nico Görnitz, Marius Kloft, Konrad Rieck, and Ulf Brefeld, "Toward supervised anomaly detection," *Journal of Artificial Intelligence Research*, vol. 46, pp. 235–262, 2013.

[10] Diederik P Kingma, Shakir Mohamed, Danilo Jimenez Rezende, and Max Welling, "Semi-supervised learning with deep generative models," in *Advances in neural information processing systems*, 2014, pp. 3581–3589.

[11] Lukas Ruff, Robert A Vandermeulen, Nico Görnitz, Alexander Binder, Emmanuel Müller, Klaus-Robert Müller, and Marius Kloft, "Deep semi-supervised anomaly detection," *arXiv preprint arXiv:1906.02694*, 2019.

[12] Sungwon Han, Hyeonho Song, Seungeon Lee, Sungwon Park, and Meeyoung Cha, "Elsa: Energy-based learning for semi-supervised anomaly detection," *arXiv preprint arXiv:2103.15296*, 2021.

[13] Chong Zhou and Randy C Paffenroth, "Anomaly detection with robust deep autoencoders," in *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, 2017, pp. 665–674.

[14] Yiru Zhao, Bing Deng, Chen Shen, Yao Liu, Hongtao Lu, and Xian-Sheng Hua, "Spatio-temporal autoencoder for video anomaly detection," in *Proceedings of the 25th ACM international conference on Multimedia*, 2017, pp. 1933–1941.

[15] Dong Gong, Lingqiao Liu, Vuong Le, Budhaditya Saha, Moussa Reda Mansour, Svetha Venkatesh, and Anton van den Hengel, "Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 1705–1714.

[16] Kihyuk Sohn, Chun-Liang Li, Jinsung Yoon, Minho Jin, and Tomas Pfister, "Learning and evaluating representations for deep one-class classification," *arXiv preprint arXiv:2011.02578*, 2020.

[17] Tal Reiss and Yedid Hoshen, "Mean-shifted contrastive loss for anomaly detection," *arXiv preprint arXiv:2106.03844*, 2021.

[18] Tal Reiss, Niv Cohen, Liron Bergman, and Yedid Hoshen, "Panda: Adapting pretrained features for anomaly detection and segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 2806–2814.

[19] Mathilde Caron, Piotr Bojanowski, Armand Joulin, and Matthijs Douze, "Deep clustering for unsupervised learning of visual features," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 132–149.

[20] Xu Ji, Joao F Henriques, and Andrea Vedaldi, "Invariant information clustering for unsupervised image classification and segmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 9865–9874.

[21] Yunfan Li, Peng Hu, Zitao Liu, Dezhong Peng, Joey Tianyi Zhou, and Xi Peng, "Contrastive clustering," in *2021 AAAI Conference on Artificial Intelligence (AAAI)*, 2021.

[22] Chuang Niu, Hongming Shan, and Ge Wang, "Spice: Semantic pseudo-labeling for image clustering," *arXiv preprint arXiv:2103.09382*, 2021.

[23] Fei Ye, Huangjie Zheng, Chaoqin Huang, and Ya Zhang, "Deep unsupervised image anomaly detection: An information theoretic framework," in *2021 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2021, pp. 1609–1613.

[24] Yuanhong Chen, Yu Tian, Guansong Pang, and Gustavo Carneiro, "Unsupervised anomaly detection with multiscale interpolated gaussian descriptors," *arXiv preprint arXiv:2101.10043*, 2021.

[25] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al., "A density-based algorithm for discovering clusters in large spatial databases with noise.," in *kdd*, 1996, vol. 96, pp. 226–231.

[26] Aaron van den Oord, Yazhe Li, and Oriol Vinyals, "Representation learning with contrastive predictive coding," *arXiv preprint arXiv:1807.03748*, 2018.

[27] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton, "A simple framework for contrastive learning of visual representations," in *International conference on machine learning*. PMLR, 2020, pp. 1597–1607.

[28] Alex Krizhevsky, Geoffrey Hinton, et al., "Learning multiple layers of features from tiny images," 2009.

[29] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[30] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.