O1-1-1

17th International Conference on Machine Vision Applications (MVA)
Fully Online, July 25–27, 2021.

Boosting Semi-Supervised Anomaly Detection via Contrasting Synthetic Images

Sheng-Feng Yu[†‡]
[†]Macronix International Co., Ltd., Taiwan
robertyu1@mxic.com.tw

Wei-Chen Chiu[‡]
[‡]National Chiao Tung University, Taiwan
walon@cs.nctu.edu.tw

Abstract

In this paper we propose to tackle the problem of semi-supervised anomaly detection, which aims to learn the outlier detector from the training set composed of only inliers. Built upon the recent advances of introducing contrastive learning to achieve the state-of-the-art of anomaly detection, we propose a simple but effective extension to further boost the performance via integrating the contrastive learning and the generative model of inliers into a unified framework. On one hand, the contrastive learning amongst the real samples and synthetic ones produced by the generative model improves the representation learning; on the other hand, the generative model learning is also benefited from the contrastive learning. We conduct extensive experiments to demonstrate the efficacy of our proposed method to advance anomaly detection, its superiority against several baselines, and the contribution of our model designs.

1 Introduction

Anomaly detection, which is also known as outlier detection or out-of-distribution (OOD) detection, is one of the most important tasks in general field of artificial intelligence and has diverse applications from network intrusion detection to automatic manufacturing via robotics. The task of anomaly detection is to identify the outliers which are referred to the samples obviously deviating from the normal distribution of a given dataset (noting that the normal samples are inliers). From the literature and research works of anomaly detection, there are three main scenarios, in accordance with the existence of outlier observations in the training dataset for learning outlier detector/classifier: supervised, unsupervised, and semi-supervised ones.

The supervised anomaly detection assumes that the training dataset contains both the labeled inliers and outliers. For instance, Görnitz *et al.* [10] label outliers by using active learning, and train the outlier detector in a supervised manner; Pang *et al.* [25] propose to leverage a few outliers for end-to-end learning both feature representation and a scoring function which enforce the deviation between outliers and inliers in terms of the anomaly scores; Hendrycks *et al.* [15] present an outlier exposure approach which particularly utilizes an auxiliary dataset of outliers to better learn the representation for anomaly detection. However, such supervised scenario is typically considered to be impractical due to the fact that outliers are often scarce and diverse thus being hard to collect.

The unsupervised anomaly detection has an assumption different from the supervised one, in which the training set still contains (mostly) inliers and (small fraction of) outliers at the same time but they are all unlabeled. The learning objective now thus turns to single out outliers from the training set. For instance, [34] uses the reconstruction errors of an autoencoder for finding a discriminative separation between inliers and outliers; [35] jointly optimizes for the reconstruction errors and the density estimation on the latent representation of inliers in order to improve distiguishing outliers; and the robust PCA technique are adopted by [5, 20] together with the autoencoder reconstruction to enforce the linear structure in the latent embedding thus achieving better robustness against outliers. However, as the learning of unsupervised anomaly detection aims to separate outliers from inliers particularly for the given dataset, the learned model usually becomes less generalizable to other datasets or new observations thus limiting its applications.

When it comes to the semi-supervised scenario of anomaly detection (also known as one-class classification), there exists an assumption which is typically considered to be more practical than the supervised scenario and more feasible than the unsupervised one: the entire training dataset includes only inliers (due to the rare and diverse nature of outliers). The common solution of semi-supervised anomaly detection is learning to model the observed inliers and estimate the novelty scores of samples at the test time. Such semi-supervised scenario attracts more focuses in recent years, where the research works can be roughly further grouped into four categories: **(1) One-class classifiers**, which aim to learn a decision boundary that encloses the inliers. For instance, [31, 33] adopts support vector machines (SVMs) for such setting, while [29] uses a neural network to approximate the kernel of one-class SVM; **(2) Reconstruction-based approaches**, which typically follow the framework of firstly modelling the generative procedure of inliers via the architectures based on autoencoder then adopting the reconstruction errors to identify the outliers, with the assumption that the generative model usually generalizes worse on the outliers which are quite different from the normal inlier samples. For instance, [30, 2, 27] utilize the adversarial learning scheme to learn the generative model of inliers. [26] further imposes stronger constraints on the latent space in order to limit the generalization ability of generator and enlarge the possible reconstruction error for the outliers. [12] in turn additionally stores the prototypical patterns of inliers to better distinguish outliers. **(3) Density-based approaches**, which target to fit distribution for the inliers, and perform the detection based on the likelihood. For example, [24] proposes a hypothesis testing scheme to determine whether a new sample resides in the typical set or not. [8, 13] estimate energy based models (EBMs) of inliers and find outliers based on its energy function. And [28] proposes a new likelihood-ratio method which removes the impact of background statistics. **(4) Self-supervised approaches**, which aim to learn good representations for the training inliers by the self-supervised learning, and detect outliers by measuring the distance between test samples and training ones on the representation space. For example, [11, 16] are amongst the first for discovering the benefits of self-supervised representation learning to the task of anomaly detection. [3] uses metric learning technique to improve the classification-based self-supervised representation learning. In particular, a recent work published by [32] leverages the recent advance in the contrastive learning [6] (which is a popular branch of self-supervised learning) to reach the state-of-the-art performance for anamoly detection.

In this paper, we discover that the generative models and the idea of contrastive learning behind [32] can be seamlessly integrated into a unified framework to further improve the ability of anomaly detection. Basically, we train the generative model on the inliers built upon the architecture of adversarial autoencoder [22] and expect it only capable of generating the samples that are similar to the training inliers, while the contrastive learning is simultaneously applied on both the real and synthesized samples. We experimentally demonstrate that the generation is beneficial for the contrastive learning and inversely the contrastive learning also contributes to better quality of generated samples. With careful and holistic design of learning objectives, our proposed method is effective to boost the performance of anomaly detection in comparison to the state-of-the-art baseline CSI [32] by a clear margin on various datasets, which are detailed in the following.

2 Methodology

2.1 Preliminaries

In this paper, we mainly consider the semi-supervised anomaly detection in terms of image data. Let $\mathcal{U} = \mathcal{X} \cup \mathcal{X}^c$ be the population of inliers and outliers. Given a set of inliers $X \subseteq \mathcal{X}$ as training data, we want to learn a detection function such that the output score differentiate the outliers from inliers. Typically, the detection function is the composition of two components (i.e. $\mathcal{M} \circ \mathcal{F}$): a feature extractor $\mathcal{F} : \mathcal{U} \to \mathcal{Z}$ which extracts the representation z of a given sample x, and a mapping function $M : \mathcal{Z} \to \mathbb{R}$ which maps the representation z to the output score.

As stated previously, the contrastive learning framework adopted by CSI [32] largely benefits the anomaly detection and our framework is stemmed on it, we therefore briefly review its main idea here. Given a sample $x_i \in X$, its two counterparts $x_i^{(1)}$ and $x_i^{(2)}$ are firstly produced by applying the transformations T_1 and T_2 respectively sampled from $\mathcal{T}$, where $\mathcal{T}$ includes random resized cropping, random color distortions, random gray scale, and random Gaussian blur, as following the setting in SimCLR [6] work. The objective of contrastive learning aims to perform the instance discrimination where $x_i^{(1)}$ and $x_i^{(2)}$ should have similar representations but different from all the other samples $X \backslash x_i$, the SimCLR loss is hence defined as:

$$\begin{aligned}\mathcal{L}_{\text{SimCLR}}(X) = &\sum_i \mathcal{L}_{\text{NT}-\text{Xent}}(x_i^{(1)}, x_i^{(2)}, X \backslash x_i) \\ &+ \sum_i \mathcal{L}_{\text{NT}-\text{Xent}}(x_i^{(2)}, x_i^{(1)}, X \backslash x_i) \qquad (1) \\ \text{in which} \quad &\mathcal{L}_{\text{NT}-\text{Xent}}(x, x_+, \{x_-\}) \\ = -\log &\frac{\exp(\text{sim}(\mathcal{F}(x), \mathcal{F}(x_+))/\tau)}{\sum_{x' \in x_+ \cup \{x_-\}} \exp(\text{sim}(\mathcal{F}(x), \mathcal{F}(x'))/\tau)}\end{aligned}$$

where sim indicates the cosine similarity and τ is the temperature parameter.

In [6], they discover that not all transformations benefit the contrastive learning, instead, some other transformations (e.g. rotation) actually deteriorate the performance of learning, where the resultant images after performing these transformations are named as shifted instances. Moreover, when each of the shifted instances is independently treated as a new sample, they can inversely contribute to better representation learning thus boosting the anomaly detection. With denoting the transformations related to shifted instances as $\mathcal{S} = \{S_1, S_2, ..., S_K\}$ (assuming there are K transformations), the contrastive objective $\mathcal{L}_{\text{con}-\text{SI}}$ built upon shifted instances is written as:

$$\begin{aligned}&\mathcal{L}_{\text{con}-\text{SI}}(X) = \mathcal{L}_{\text{SimCLR}}(\{X, X^{\mathcal{S}}\}) \\ &\text{where} \quad X^{\mathcal{S}} = \bigcup_i \bigcup_{S \in \mathcal{S}} \{S(x_i)\}\end{aligned} \qquad (2)$$

Moreover, the objective $\mathcal{L}_{\text{cls}-\text{SI}}$ based on the auxiliary classification task [9, 11] to predict the type of transformation of a shifted instance also helps the learning:

$$\mathcal{L}_{\text{cls}-\text{SI}}(X) = -\sum_i \log p(y = S | S(x_i)) \qquad (3)$$

The overall objective of CSI [32] approach hence is:

$$\mathcal{L}_{\text{CSI}}(X) = \mathcal{L}_{\text{con}-\text{SI}}(X) + \mathcal{L}_{\text{cls}-\text{SI}}(X) \qquad (4)$$

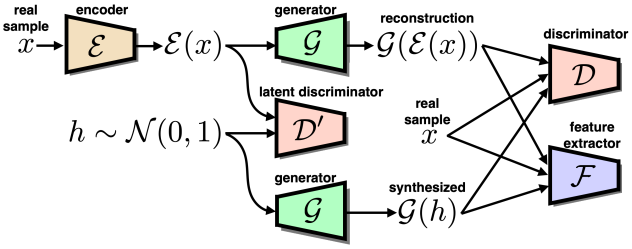


Figure 1: Illustration of our proposed method, which seamlessly integrate generative model built upon adversarial autoencoder and contrastive learning into a unified framework. Please refer to Sec. 2.2 for details.

2.2 Our Proposed Method

As motivated previously, our proposed method integrates the generative model and contrastive learning into a unified framework, as illustrated in Figure 1, which includes five sub-networks: encoder $\mathcal{E}$, generator $\mathcal{G}$, latent discriminator $\mathcal{D}'$, discriminator $\mathcal{D}$, and feature extractor $\mathcal{F}$. The learning objectives of our proposed method can be categorized into two groups: **Joint Generative and Contrastive Learning** and **Learning Adversarial Autoencoder**, in which we detail these two groups sequentially in the following.

• **Joint Generative and Contrastive Learning**. Assume that the generator $\mathcal{G}$ is able to well model the generative procedure of real samples of inliers $x \in X$ in the training set, the synthesized samples $\tilde{x} \in \tilde{X}$ (where $\tilde{x} = \mathcal{G}(h)$ and $h \sim \mathcal{N}(0,1)$), which ideally are similar to real inliers [26], could help to enrich the training set of contrastive learning thus benefiting the representation learning. We hence propose to have $\mathcal{L}_{\text{JOINT}}$ to train the feature extractor $\mathcal{F}$:

$$\mathcal{L}_{\text{JOINT}} = \mathcal{L}_{\text{CSI}}(X \cup \tilde{X}) \tag{5}$$

Moreover, we propose to utilize the contrastive learning amongst the synthesized samples $\tilde{X}$ which in turn helps to improve the learning of generator $\mathcal{G}$:

$$\mathcal{L}_{\text{CSI-S}} = \mathcal{L}_{\text{CSI}}(\tilde{X}) \tag{6}$$

• **Learning Adversarial Autoencoder**. The generative model in our proposed method is built upon the architecture of adversarial autoencoder [22]. First, we adopt the adversarial learning amongst real samples x, synthesized samples $\tilde{x}$, and the reconstructed samples $\hat{x} = \mathcal{G}(\mathcal{E}(x))$ as inspired by VAEGAN [21]:

$$\mathcal{L}_{\text{D}} = \log \mathcal{D}(x) + \log(1 - \mathcal{D}(\tilde{x})) + \log(1 - \mathcal{D}(\hat{x})) \tag{7}$$

Second, the reconstruction loss between x and $\hat{x} = \mathcal{G}(\mathcal{E}(x))$ is adopted on both pixel and feature domains, where the latter additionally includes the contrastive learning idea inside (λ is experimentally set to 100):

$$\mathcal{L}_{\text{REC}} = \lambda \|x - \hat{x}\|_2^2 - \text{sim}(\mathcal{F}(T_1(x)), \mathcal{F}(T_2(\hat{x}))) \tag{8}$$

Third, we adopt latent discriminator $\mathcal{D}'$ to impose prior distribution on the latent space, with $h \sim \mathcal{N}(0,1)$:

$$\mathcal{L}_{\text{LD}} = \log \mathcal{D}'(h) + \log(1 - \mathcal{D}'(\mathcal{E}(x))) \tag{9}$$

Lastly, we introduce the latent reconstruction loss as [1, 4] to better regularize the learning of generative model:

$$\mathcal{L}_{\text{LREC}} = \|\mathcal{E}(\mathcal{G}(h)) - h\|_2^2 \tag{10}$$

The overall training procedure of our proposed method is then summarized in the algorithm below:

Algorithm 1: Training our proposed method

```
Result: $\theta_{\mathcal{E}}, \theta_{\mathcal{G}}, \theta_{\mathcal{D}'}, \theta_{\mathcal{D}}$, and $\theta_{\mathcal{F}}$
Initialize $\theta_{\mathcal{E}}, \theta_{\mathcal{G}}, \theta_{\mathcal{D}'}, \theta_{\mathcal{D}}$, and $\theta_{\mathcal{F}}$
for next image batch $X$ do
    Update $\theta_{\mathcal{F}}$ by minimizing $\mathcal{L}_{\text{CSI}}(X)$
    if every $N$ batches then
        Sample plenty $h \sim \mathcal{N}(0, I)$
        $\tilde{X}, \hat{X} \leftarrow \mathcal{G}(h), \mathcal{G}(\mathcal{E}(X))$
        $\theta_{\mathcal{D}}, \theta_{\mathcal{G}}, \theta_{\mathcal{E}} \leftarrow \arg\min_{\theta_{\mathcal{D}}} \max_{\theta_{\mathcal{G}}, \theta_{\mathcal{E}}} \mathcal{L}_{\text{D}}(X, \tilde{X}, \hat{X})$
        $\theta_{\mathcal{G}} \leftarrow \arg\min_{\theta_{\mathcal{G}}} \mathcal{L}_{\text{CSI-S}}(\tilde{X})$
        $\theta_{\mathcal{D}'}, \theta_{\mathcal{E}} \leftarrow \arg\min_{\theta_{\mathcal{D}'}} \max_{\theta_{\mathcal{E}}} \mathcal{L}_{\text{LD}}(h, X)$
        Sample $T_1, T_2 \sim \mathcal{T}$
        $\theta_{\mathcal{G}}, \theta_{\mathcal{E}} \leftarrow \arg\min_{\theta_{\mathcal{G}}, \theta_{\mathcal{E}}} \mathcal{L}_{\text{REC}}(X, \hat{X}, T_1, T_2)$
        $\theta_{\mathcal{E}} \leftarrow \arg\min_{\theta_{\mathcal{E}}} \mathcal{L}_{\text{LREC}}(h)$
        $\theta_{\mathcal{F}} \leftarrow \arg\min_{\theta_{\mathcal{F}}} \mathcal{L}_{\text{JOINT}}(X \cup \tilde{X})$
    end
end
```

2.3 Outlier Score Function

We use a simplified score function $\mathcal{M}$ from [32] to measure the outlier score by comparing the average feature similarity between a given test sample x_{test} and its nearest neighbor $\breve{x}$ from the training set X, under various transformations $\mathcal{S}$ related to shifted instances:

$$\begin{aligned} &\mathcal{M}(x_{\text{test}}, \breve{x}) = \sum_{S \in \mathcal{S}} m(S(x_{\text{test}}), S(\breve{x})) \\ &\text{where} \quad m(x, x') = \text{sim}(\mathcal{F}(x), \mathcal{F}(x')) \cdot \|\mathcal{F}(x)\| \\ &\text{and} \quad \breve{x} = \arg\max_{x' \in X} \sum_{S \in \mathcal{S}} m(x_{\text{test}}, x') \end{aligned} \tag{11}$$

Moreover, as indicated in [32], the detection performance can be further improved by applying multiple random transformations $\{T_r\}_{r=1}^{R} \sim \mathcal{T}$ in addition to $\mathcal{S}$ on the score function $\mathcal{M}$, which results in $\mathcal{M}_{\text{ens}}$.

3 Experimental Results

Setup. The implementation of our proposed method basically follows [32] to adopt ResNet-18 [14] as our feature extractor $\mathcal{F}$ and uses rotation 0°, 90°, 180°, 270°

Table 1: Average AUROC (%) of anomaly detection on CIFAR-10, CIFAR-100, and ImageNet-30

Dataset	DeepSVDD [29]	OCGAN [26]	Geom [11]	Rot+Trans [16]	GOAD [3]	CSI [32]	Ours
CIFAR-10	64.8	65.7	86.0	90.1	88.2	94.3	**95.1**
CIFAR-100	-	-	78.7	79.8	74.5	89.6	**90.9**
ImageNet-30	-	-	-	85.7	-	91.6	**93.0**

for augmentation $\mathcal{S}$, as well as follows [23] to design the network for generator $\mathcal{G}$ and discriminator $\mathcal{D}$, while the architecture of encoder $\mathcal{E}$ is symmetric to $\mathcal{G}$. Most of our optimization settings follows [32]. We use Adam optimizer [18] for model training with learning rate and momentum set to 0.0002 and $(0.9, 0.999)$ respectively. Moreover, we set N (refer to Algorithm 1) to 5.

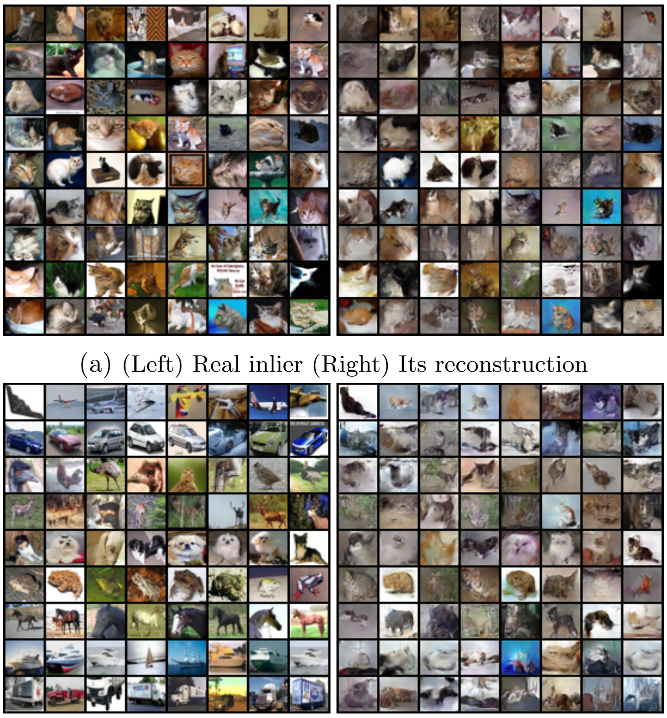
Dataset and Metrics. Three datasets are leveraged for running our evaluation, i.e. CIFAR-10 [19], CIFAR-100 [19], and ImageNet [7], where we consider the one-class setup in which each task chooses a single class as the inlier and other classes are outliers, and the overall performance is average over all tasks. In particular, for CIFAR-100, we adopt the pre-defined superclass (i.e. 5 classes belonging to similar object category) as the unit of a class for our experiments, while for ImageNet we adopt only 30 classes (i.e. ImageNet-30 as following [16, 32]). The model is trained on only inliers, and its performance is evaluated on full testing set. The performance metric is the area under the Receiver Operating Characteristic (AUROC) curve.

3.1 Quantitative and Qualitative Results

We compare our proposed method with respect to several baselines, including DeepSVDD [29], OCGAN [26], Geom [11], Rot+Trans [16], GOAD [3] and the state-of-the-art CSI [32]. Table 1 provides the semi-supervised anomaly detection results for CIFAR-10, CIFAR-100, and ImageNet-30 datasets. It is significant to see that our proposed method achieves superior performance in comparison to all the baselines on all the three datasets. In particular, the outperformance of our proposed method with respect to the the state-of-the-art CSI method successfully verifies the our contribution on integrating contrastive learning and generative model into a unified framework, where the synthesized samples benefit contrastive learning for learning better representations while the constrastive learning in turn helps to improve the generation. In Figure 2 we also provide some qualitative examples of our method, where the inlier class is CIFAR-10 Cat and other classes are outliers. We can see that, all the outliers are reconstructed as inliers, and inliers remain alike themselves, thus verifying the effectiveness of our proposed method.

3.2 Ablation Study

We further conduct ablation study on our model designs, based on the CIFAR-10 dataset, where we start from CSI method and sequentially integrate our generator $\mathcal{G}$ and encoder $\mathcal{E}$ onto it to reach the full model of our proposed method. With only having generator $\mathcal{G}$ integrated with CSI method, the AUROC score improves from 94.3 (i.e. CSI only) to 94.8, while further including encoder $\mathcal{E}$ (i.e. our full model) advances the AUROC score to 95.1, thus verifying again the benefit of generative model made for contrastive learning. Moreover, we adopt the FID score [17] (the lower the better), which is widely adopted for quantifying the quality of synthesized images, to perform the ablation study. We observe that the variant of having only generator integrated with CSI has FID score 69.5 while our full model achieves 54.1, thus implicitly verifying that better constrastive learning helps the generation.



(a) (Left) Real inlier (Right) Its reconstruction

(b) (Left) Real outlier (Right) Its reconstruction

Figure 2: Qualitative examples (inlier: CIFAR-10 Cat)

4 Conclusion

In this work, we propose an effective approach to boost the performance semi-supervised anomaly detection via having contrastive learning and generative model integrated in a unified framework. The extensive experiments on various datasets and the ablation studies successfully verify the mutual benefits between contrastive learning and generative model as well as the contribution of our model designs.

Acknowledgment We would like to express our gratitude to Tzu-Hsiang Su, Yi-Hao Jhu, Min-Liang Wei, Wei-Chen Wang, Hsiang-Pang Li, and Keh-Chung Wang for helpful feedback and suggestions.

References

[1] S. Akash, V. Lazar, R. Chris, G. M. U., and S. Charles. Veegan: Reducing mode collapse in gans using implicit variational learning. *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

[2] S. Akcay, A. Atapour-Abarghouei, and T. P. Breckon. Ganomaly: Semi-supervised anomaly detection via adversarial training. In *Asian Conference on Computer Vision (ACCV)*, 2018.

[3] L. Bergman and Y. Hoshen. Classification-based anomaly detection for general data. In *International Conference on Learning Representations (ICLR)*, 2020.

[4] A. T. Cemgil, S. Ghaisas, K. Dvijotham, S. Gowal, and P. Kohli. Autoencoding variational autoencoder. *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

[5] R. Chalapathy, A. K. Menon, and S. Chawla. Robust, deep and inductive anomaly detection. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases (ECML-PKDD)*, 2017.

[6] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton. A simple framework for contrastive learning of visual representations. In *International Conference on Machine Learning (ICML)*, 2020.

[7] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009.

[8] Y. Du and I. Mordatch. Implicit generation and modeling with energy based models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

[9] S. Gidaris, P. Singh, and N. Komodakis. Unsupervised representation learning by predicting image rotations. In *International Conference on Learning Representations (ICLR)*, 2018.

[10] N. Goernitz, M. Kloft, K. Rieck, and U. Brefeld. Toward supervised anomaly detection. *Journal of Artificial Intelligence Research (JAIR)*, 2013.

[11] I. Golan and R. El-Yaniv. Deep anomaly detection using geometric transformations. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2018.

[12] D. Gong, L. Liu, V. Le, B. Saha, M. R. Mansour, S. Venkatesh, and A. v. d. Hengel. Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection. In *IEEE International Conference on Computer Vision (ICCV)*, 2019.

[13] W. Grathwohl, K.-C. Wang, J.-H. Jacobsen, D. Duvenaud, M. Norouzi, and K. Swersky. Your classifier is secretly an energy based model and you should treat it like one. *International Conference on Learning Representations (ICLR)*, 2020.

[14] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[15] D. Hendrycks, M. Mazeika, and T. Dietterich. Deep anomaly detection with outlier exposure. In *International Conference on Learning Representations (ICLR)*, 2019.

[16] D. Hendrycks, M. Mazeika, S. Kadavath, and D. Song. Using self-supervised learning can improve model robustness and uncertainty. *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

[17] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

[18] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *International Conference on Learning Representations (ICLR)*, 2015.

[19] A. Krizhevsky. Learning multiple layers of features from tiny images. Technical report, 2009.

[20] C.-H. Lai, D. Zou, and G. Lerman. Robust subspace recovery layer for unsupervised anomaly detection. In *International Conference on Learning Representations (ICLR)*, 2020.

[21] A. B. L. Larsen, S. K. Sønderby, H. Larochelle, and O. Winther. Autoencoding beyond pixels using a learned similarity metric. In *International Conference on Machine Learning (ICML)*, 2016.

[22] A. Makhzani, J. Shlens, N. Jaitly, and I. Goodfellow. Adversarial autoencoders. In *International Conference on Learning Representations (ICLR)*, 2016.

[23] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida. Spectral normalization for generative adversarial networks. In *International Conference on Learning Representations (ICLR)*, 2018.

[24] E. Nalisnick, A. Matsukawa, Y. W. Teh, and B. Lakshminarayanan. Detecting out-of-distribution inputs to deep generative models using a test for typicality. *ArXiv:1906.02994*, 2019.

[25] G. Pang, C. Shen, and A. van den Hengel. Deep anomaly detection with deviation networks. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2019.

[26] P. Perera, R. Nallapati, and B. Xiang. OCGAN: One-class novelty detection using gans with constrained latent representations. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[27] S. Pidhorskyi, R. Almohsen, D. A. Adjeroh, and G. Doretto. Generative probabilistic novelty detection with adversarial autoencoders. *Advances in Neural Information Processing Systems (NeurIPS)*, 2018.

[28] J. Ren, P. J. Liu, E. Fertig, J. Snoek, R. Poplin, M. A. DePristo, J. V. Dillon, and B. Lakshminarayanan. Likelihood ratios for out-of-distribution detection. *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

[29] L. Ruff, R. Vandermeulen, N. Goernitz, L. Deecke, S. A. Siddiqui, A. Binder, E. Müller, and M. Kloft. Deep one-class classification. In *International Conference on Machine Learning (ICML)*, 2018.

[30] T. Schlegl, P. Seeböck, S. M. Waldstein, U. Schmidt-Erfurth, and G. Langs. Unsupervised anomaly de-

tection with generative adversarial networks to guide marker discovery. In *International Conference on Information Processing in Medical Imaging*, 2017.

[31] B. Schölkopf, R. C. Williamson, A. J. Smola, J. Shawe-Taylor, J. C. Platt, et al. Support vector method for novelty detection. In *Advances in Neural Information Processing Systems (NeurIPS)*, 1999.

[32] J. Tack, S. Mo, J. Jeong, and J. Shin. CSI: Novelty detection via contrastive learning on distributionally shifted instances. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

[33] D. M. Tax and R. P. Duin. Support vector data description. *Machine Learning*, 2004.

[34] Y. Xia, X. Cao, F. Wen, G. Hua, and J. Sun. Learning discriminative reconstructions for unsupervised outlier removal. In *IEEE International Conference on Computer Vision (ICCV)*, 2015.

[35] B. Zong, Q. Song, M. R. Min, W. Cheng, C. Lumezanu, D. Cho, and H. Chen. Deep autoencoding gaussian mixture model for unsupervised anomaly detection. In *International Conference on Learning Representations (ICLR)*, 2018.