Discriminative-Generative Representation Learning for One-Class Anomaly Detection

Xuan Xia¹ xiaxuan@cuhk.edu.cn
Xizhou Pan¹ panxizhou@cuhk.edu.cn
Xing He¹ hexing@cuhk.edu.cn
Jingfei Zhang¹ 117010378@cuhk.edu.cn
Ning Ding¹,²,∗ dingning@cuhk.edu.cn
Lin Ma¹,∗ malin@cuhk.edu.cn

¹Shenzhen Institute of Artificial Intelligence and Robotics for Society, Shenzhen 518129, P. R. China.
²Institute of Robotics and Intelligent Manufacturing, Chinese University of Hong Kong, Shenzhen, 518172, P. R. China.
∗Corresponding author.

Abstract

As a kind of generative self-supervised learning methods, generative adversarial nets have been widely studied in the field of anomaly detection. However, the representation learning ability of the generator is limited since it pays too much attention to pixel-level details, and generator is difficult to learn abstract semantic representations from label prediction pretext tasks as effective as discriminator. In order to improve the representation learning ability of generator, we propose a self-supervised learning framework combining generative methods and discriminative methods. The generator no longer learns representation by reconstruction error, but the guidance of discriminator, and could benefit from pretext tasks designed for discriminative methods. Our discriminative-generative representation learning method has performance close to discriminative methods and has a great advantage in speed. Our method used in one-class anomaly detection task significantly outperforms several state-of-the-arts on multiple benchmark data sets, increases the performance of the top-performing GAN-based baseline by 6% on CIFAR-10 and 2% on MVTAD. What’s more, ablation studies show that absolute position information deteriorates representational learning ability of generative methods in geometric transformation tasks, and has different effects on the representational learning ability of discriminative methods in different geometric transformation tasks, which provides a criterion for the use of position information.

1. Introduction

Anomalies could be errors in the data or sometimes are previously unknown out-of-distribution samples. As a classic pattern recognition task, anomaly detection has made great progress in the era of deep learning. However, one-class anomaly detection (OCAD) in unsupervised scenarios is still a challenging task due to the difficulty of model establishment and the absence of labels.

OCAD aims to identify patterns that do not belong to the normal data distribution [5]. A typical solution is to map the normal data to a definite distribution in a latent space, thus identifying the data outside the distribution as anomalies. Among many methods, generative adversarial nets (GAN) has been widely used due to its advantages in distribution fitting. At the beginning, researchers usually use auto-encoder structure GAN for data encoding and reconstruction [2, 35, 34], and measure anomalies by reconstruction error. However, studies on representation learning show that generative learning methods have inherent defects in feature learning. Most GAN-based methods don’t have strong enough constraints on the learning of representations, which is crucial to distinguish abnormal from normal. Pixel-wise reconstruction may result in the loss of important semantic information, and leads to the degradation of anomaly detection performance.

Therefore, some scholars begin to explore how to learn features efficiently from the perspective of representation learning [13, 33, 36], and the research on self-supervised learning is particularly bright. Contemporary there are two types of self-supervised learning methods as shown in Figure 1. Most GAN-based methods can be naturally regarded as belonging to generative self-supervised learning methods, such as OCGAN [34] and Old-Is-Gold [47]. The generator tries to learn representations by reconstructing x from the x̂ transformed by t(x), where t(x) is the image transformation determined by the pretext task. In order to avoid models focusing too much on pixel details, discriminative
self-supervised learning methods for anomaly detection are gradually emerging. In these methods, representations are learned by predicting the labels $c_t$ provided by pretext task. For example, GEOM [13] performs efficient representation learning through the prediction of geometric transformations and improves the performance of anomaly detection as a pioneering work. After that, the potential of discriminative self-supervised learning in the field of anomaly detection began to show. For example, GOAD [3] and SLOOD [18] detect abnormalities by multiple transformation classification. Furthermore, GDFR [33] and ARnet [11] attempt to implement better representation learning by using generative method assisted discriminative method (called generative-discriminative representation learning). However, generator is difficult to learn abstract semantic representations from label prediction pretext tasks as effective as discriminator. Some samples produced by geometric transformation pretext tasks have the problem of label semantic ambiguity (shown in Figure 2), which means the generator cannot benefit from the pretext tasks designed for discriminative methods. On the other hand, discriminative methods mostly rely on the complex post-processing that require training sets in testing such as using Dirichlet score [13] and distribution normalization [11], which are high computation consumed.

In order to better representation learning for the generator, we try to combine generative methods with discriminative methods to mine as many meaningful representations as possible, and improve the anomaly detection performance of GAN in anomaly detection tasks. Figure 1(c) demonstrates the framework of our method. We reuse the discriminator as a predictor to guide the generator to generate samples that match the pretext task labels. The generator does not attempt to reconstruct the image on the pixel level, thus the representations learned do not focus too much on pixel details. On the other hand, the discriminator guides the generator to generate images of the correct categories and promotes the generator to learn accurate abstract semantic representations. Meanwhile, as other generative methods, we still use reconstruction error to represent the anomaly score, unlike some discriminative methods which require complicated post-processing.

Our method is different from the previous GAN-based approaches such as GDFR [33] and CompareGAN [41]. Their purpose is to make discriminator learn better representations assisted by generator, while our method is to make generator learn better representations guided by discriminator. We propose this discriminative-generative representation learning method for one-class anomaly detection task in this paper, named DGAD (discriminative-generative anomaly detection). The main contributions of us can be summarized as follows:

1. The generator of DGAD can learn abstract semantic representation more efficiently without label semantic ambiguity problem. This is the first attempt to benefit generator from pretext tasks designed for discriminative methods as we know.
2. Our method significantly outperforms several state-of-the-arts on multiple benchmark datasets, increases the performance of the top-performing GAN-based baseline by 6% on CIFAR-10 and 2% on MVTAD.
3. Our experiments prove that based on the representations learned from the same pretext prediction task, the generative methods can approach the performance of discriminative methods by simply taking the reconstruction error as the anomaly score, and has a great advantage in speed.
4. Ablation studies show that absolute position information can degrade the representation learning ability of generative methods in geometric transformation pretext tasks, and has different effects on the representational learning ability of discriminative methods in different geometric transformation tasks, suggesting that the use of position information should be carefully selected according to different pretext tasks.

2. Related Work
2.1. One-class anomaly detection (OCAD)
OCAD assumes that all training samples belong to one class, and strives to learn a classification boundary that sur-
rounds all normal samples. Thereby, any new sample that is not inside the decision boundary can be identified as an anomaly.

The performance of classical methods such as one-class SVM [40] and one-class nearest neighbors [39] usually limited by the dimensionality and complexity of the inputs, cannot be applied to high-dimensional or large scale datasets. In recent years, the development of deep learning improves the performance and practicability of OCAD [30] [1] [24]. An important kind of OCAD methods are based on auto-encoder [28]. The success of auto-encoder naturally attracts scholars to use GAN in order to obtain better detection performance.

2.2. GAN-based anomaly detection

According to the theory, the generated data distribution of GAN model will be consistent with the real data distribution under Nash equilibrium. Therefore, GAN can be used to detect anomalies outside the distribution. Based on this assumption, AnoGAN [37] was proposed in 2017 with the use of GAN by the first time. However, the foundation of AnoGAN is based on intuition instead of theory. In fact, before AnoGAN, both BiGAN [8] and ALI [10] have already used GAN for adversarial learning and inference, which is the theoretical basis of applying GAN to representation learning. Then, CIAFL [45] gives the research object of adversarial learning and inference: produce a data representation that maintains meaningful variations of data while eliminating noisy signals. This definition exactly meets the requirements of anomaly detection. Thus, researchers started to take advantage of theories related with adversarial representation learning and inference to modify the GAN-based anomaly detection. For example, Efficient GAN-Anomaly [48] adopts the theory of BiGAN. ALAD [49] adopts the theory of ALICE [25]. GANomaly [2] further modifies the network structure and loss function to constrain latent space.

With the improvement of the GAN theory, GAN-based anomaly detection methods are divided into two directions. One is the deeper insight into latent space such as manifold learning [35], sparse regularization [53] and informative-negative mining [34]. The other is more accurate anomaly location, such as the use of class activation maps [20] and attention expansion loss [43]. However, GAN-based representation learning has stagnated recently with only a few new research results [44, 50]. This situation limits the improvement of detection performance of GAN-based anomaly detection methods. Meanwhile, the rapid development of self-supervised learning provides some new inspiration for representation learning, promotes the use of pretext tasks in GAN-based anomaly detection method [33] [8] [18] [11].

2.3. Self-supervised learning

As shown in Figure 1, contemporary self-supervised learning methods can roughly be broken down into two classes of methods in the field of computer vision: generative methods and discriminative methods. As a more traditional approach, generative methods focus on reconstruction error in the pixel space to learn representations, such as colorization [51], super-resolution [23], inpainting [32], and cross-channel prediction [52]. However, using pixel-level losses can lead to overly focus on pixel-level details, rather than more abstract latent representations, thereby reducing their ability to model correlations or complex structure. On the contrary, discriminative methods create (pseudo) labels by pretext tasks and learn representations by label predictions, such as image jigsaw puzzle [29], context prediction [7], and geometric transformation recognition [12]. Most of these pretext tasks have been used for anomaly detection in recent years [13] [33] [3] [18] [11]. In particular, as a kind of discriminative methods, contrastive learning methods treat each instance as a category, learn representations by contrasting positive and negative examples. Contrastive learning methods have led to great empirical success in computer vision tasks recently, such as MoCo [16] and SimCLR [6]. Some of techniques of contrast learning are just beginning to be used in anomaly detection [38] [14] [31]. This shows that self-supervised learning has great potential in anomaly detection.

3. Our Method

3.1. Label semantic ambiguity

As shown in Figure 2, some images produced by the geometric transformations may have the same rationality as the original images, rendering the label meaningless.
ity. Most of the existing discriminative methods ignore this phenomenon because predictors are more or less tolerant to noise labels. Generative methods, however, are at a loss in the face of these pretext tasks designed for discriminative methods. Forced image generation based on meaningless labels will lead to confusion in model training and reduce the validity of representation. The problem of label semantic ambiguity prevents generators from learning representations from label prediction tasks. Hence, our method attempts to provide a solution that combines generative methods and discriminative methods to improve the representation learning ability of generative methods, while benefit from the pretext tasks designed for discriminative methods.

3.2. DGAD

We propose a self-supervised learning method for one-class anomaly detection task based on generative adversarial nets, called DGAD. Like most GAN-based anomaly detection methods, our method is based on the assumption that the reconstruction error of the abnormal sample is greater than that of the normal sample. Our model uses pretext task of geometric transformations for representation learning. Take rotation prediction for example, as shown in Figure 3, we first rotate the training samples randomly and generate labels. Secondly, an encoder and a decoder are used to learn the encoding of samples in latent space by the image reconstruction of $x_r$, meanwhile a discriminator is used to predict the rotation angle of the image. Finally, the rotated image needs to be restored by the encoder, and the authenticity and rotation angle of the restored image are checked by the discriminator.

Mathematically, defining $X$ to represent the domain of the data samples and $Z$ to represent the domain of the encoding. Given a sample $x_r \in X$ with label $c_r$ ($c_r$ is the same label that all the original samples have), we generate the transformed sample $x_t$ and transformation label $c_t$ by a transformer $T$, then the encoder $En$ converts them to encoded representation as $z_r$ and $z_t \in Z$; and the decoder $De$ is trained to reversely mapping them to $\hat{x}_r$ as follows

$$x_t = T(x_r)$$

$$z_r = En(x_r), z_t = En(x_t)$$

$$\hat{x}_r = De(z_r) = \hat{x}_t = De(z_t)$$

The discriminator $D$ is trained to distinguish between real and fake samples. Meanwhile, it contains a classifier $D_{cls}$ to predict the transformation label $c_t$ through the joint distributions of $x_t$ and $c_t$ as follows

$$c_t = \hat{c}_t = D_{cls}(x_t, z_t)$$

Therefore, this trained classifier can be used to guide the decoder to generate the restored image with label $c_r$.

$$c_r = \hat{c}_r = D_{cls}(\hat{x}_t, \hat{z}_t)$$

Please note that unlike other generative methods, we do not train the decoder to restore $x_r$ from the transformed image $x_t$ by reconstruction loss, but by the label classified by $D_{cls}$. Images that cannot be predicted by classifiers need not be forced to be reconstructed, thus it avoids the meaningless reconstruction. Through this training strategy, our method both utilizes the representation learning ability of the generative methods for pixel details and the discriminative methods for abstract semantics.

3.3. Losses

**Reconstruction Loss.** As a GAN-based method, it must have the reconstruction ability of normal images. We use the original image $x_r$ and its reconstruction for the training of encoder $En$ and decoder $De$.

$$L_{rec} = E_{x_r, \hat{x}_r} \| x_r - \hat{x}_r \|_1$$

Figure 3. The training pipeline of DGAD with the pretext task of rotation prediction. For images $x_t$, the discriminator is trained as a predictor for predicting rotation angles $c_t$. For fake images $\hat{x}_t$, the parameters of the discriminator are fixed and expected to output a normal angle label $c_r$, thus guiding the decoder to generate a normal angle image.
where $\theta_{en}$ and $\theta_{de}$ denote the parameters of encoder and decoder respectively, $\| \cdot \|_1$ means L1 norm. The transformed images $x_t$ are not included in this training, although we want $x_\hat{t}$ to be the same as $x_t$ after passing through the $En$ and $De$. The restoration of $x_t$ are guided by the discriminator.

**Classification Loss.** This objective has two terms: a loss of transformed images $x_\hat{t}$ and their labels $c_t$ used to optimize the discriminator $D$, and a loss of restored images used to optimize encoder and decoder. The former is defined as

$$L^D_{cls} = E_{\theta_d} [ - \log D_{cls} (c_t | (x_t, z_t))]$$  \hspace{1cm} (7)

where $\theta_d$ denotes the parameters of $D$, $D_{cls}(c(x, z))$ represents the joint probability distribution of $x$ and $z$ over labels $c$ computed by $D$. $D$ learns how to classify transformed images by this loss.

The effective learning of the discriminator for classification does not mean that encoder and decoder also learn the relevant semantic representation. They need to learn effective representations at the same time, so as to help the discriminator classify more accurately. That requires encoder and decoder try to generate images that can be classified as the target label $c_r$. Hence the loss used to optimize encoder and decoder is defined as

$$L^G_{cls} = E_{\theta_{en}} [ - \log D_{cls} (c_t | (x_t, z_t))]$$
$$+ E_{\theta_{en}, \theta_{de}} [ - \log D_{cls} (c_r | (x_t, z_t))]$$  \hspace{1cm} (8)

The first term of this loss improves the efficiency of representation learning of encoder, and thus encoder can provide representation to the discriminator for more accurate classification. The latter term of this loss enables encoder and decoder to learn to restore the image to the correct category under the guidance of the discriminator. The discriminator first learns the image classification, then instructs encoder and decoder to generate normal category image. Image restoration is no longer guided by reconstruction loss but this loss. Thus, the representation of encoder and decoder learning is avoided to pay too much attention to pixel details; on the other hand, the interference of meaningless image reconstruction on representation learning is avoided.

**Compactness Loss.** The encoding of normal samples should be a compact distribution in latent space since images of the same category should be encoded close to each other. To do this, we constrain the variance among all encodings in the training batch. However, it is not reasonable to try to have each component in the encodings close together, which limits the inner-class diversity and causes the abnormal sample encoding to move closer to the normal sample encoding. Therefore, we only limit the variance on the channel axis. The compactness loss is defined as

$$L_{cmp} = \sqrt{\text{var}_b(z_r)}$$  \hspace{1cm} (9)

where $\text{var}_b(.)$ means the variance among data from the same training batch,

$$\text{var}_b(z) = E_{\theta_{en}} \left( \text{ap}_c(z_i) - \frac{1}{N} \sum_{i=1}^{N} \text{ap}_c(z_i) \right)^2$$  \hspace{1cm} (10)

where $N$ is the batch size, $\text{ap}_c(.)$ means the average pooling on the channel axis.

The encoding of all normal samples has a similar mean on each channel by this limitation, but can have different values on width axis and height axis. In other words, the channel means of encodings of the normal samples $z_r$ are constrained in a compact high-dimensional distribution. The mean of the abnormal sample $z_e$ on the channels will no longer be the same as the normal sample, resulting in significant differences between the $z_r$ and $z_e$, and helping to distinguish abnormal samples.

**Adversarial Loss.** We use the discriminator of the same architecture as BiGAN [37] to train the model, and use hinge loss [9] to build the adversarial loss as follows

$$L^G_{adv} = E_{\theta_d} [ \max (0, 1 - D_{adv}(x_r, z_r))]$$
$$+ E_{\theta_d} [ \max (0, 1 + D_{adv}(x_\hat{t}, z_\hat{t}))]$$  \hspace{1cm} (11)

$$L^D_{adv} = -E_{\theta_{en}, \theta_{de}} [D_{adv}(x_\hat{t}, z_\hat{t})]$$  \hspace{1cm} (12)

where $D_{adv}(x, z)$ represents the discriminant value of joint probability distribution of $x$ and $z$ computed by $D$. Unlike receiving all transformed images in classification loss, the discriminator only considers $x_r$ to be true images in adversarial loss, forcing the distribution of generated images to approximate the distribution of original images.

**Total Loss.** The objective to optimize $En$, $De$ and $D$ are

$$L_D = L^{adv}_D + \lambda_{cls} L^{D}_{cls}$$  \hspace{1cm} (13)

$$L_{En, De} = L^{G}_{cls} + \lambda_{cls} L^{G}_{cls} + \lambda_{rec} L_{rec} + \lambda_{cmp} L_{cmp}$$  \hspace{1cm} (14)

where $\lambda_{cls}$, $\lambda_{rec}$ and $\lambda_{cmp}$ are hyper-parameters. We use $\lambda_{cls} = 10$, $\lambda_{rec} = 20$ and $\lambda_{cmp} = 100$ in all of our experiments.

### 3.4. Pretext task and absolute position information

In this section we describe the pretext tasks of geometric transformations we use in our study.

**Rotation.** An easy rotation mechanism that input images are rotated by $0^\circ$, $90^\circ$, $180^\circ$, $270^\circ$ is proposed by [12]. In our study, we use a 4-bits one-hot encoding as the (pseudo) label of a rotated image. So all the original samples are set the same real label $c_r$ (1, 0, 0, 0) and the rotated image are the special label $c_t$ as shown in Figure 3. Unlike GEOM [13], we don’t add flips and shifts to the transformations since sometimes they cannot be classified meaningfully as we explained in Figure 2.
**Jigsaw Puzzle.** For centrally symmetric images (e.g. metal nuts and bottles in MVTec AD dataset), the pretext task of rotating or flipping the whole image is meaningless. Hence we apply Jigsaw Puzzle to transform training task of rotating or flipping the whole image is meaningful (e.g. metal nuts and bottles in MVTec AD dataset), the pretext task of rotating or flipping the whole image is meaningful. We set $N = 4, 9, 16, \ldots, n^2, n \in Z^+$ partitions and each partition can be transformed independently. We set $N = 4$ and select to fixed the top-left partition and randomly permute other partitions with at least 2 displacements. The set of puzzles consist of 6 different permutations of the partitions in case of having 4 partitions. The split input image has 4 quadrants and each partition corresponds to each quadrant in original image.

There are three protocols in our study for one-class anomaly detection.

**Protocol 1:** Only using rotation on the original input images, every training samples is randomly rotated. So there are 4 different transformations and 4-bit one-hot label.

**Protocol 2:** Only using Jigsaw Puzzle on the split images. Randomly permuting the partition, and 6-bit one-hot (pseudo) labels are considered for the case of 6 possible transformations. In this protocol, the model need to learn the relative position of features.

**Protocol 3:** Based on Protocol 2, we select to add rotation on the partition after random permutation. So there are 6×64 different transformed possibilities. Considering that too many possibilities will increase the dimension of label, we suggest using multi-hot label instead of one-hot label in this protocol. In this protocol, the discriminator not only need to learn the relative position of features, but also the correct angle of each partition.

Since these geometric transformation pretext tasks are associated with location information, we explore the influence of absolute location information on the representational learning ability of our model We find the isolation of absolute position information can improve the performance of representation learning of generator, which will be proved in the ablation studies. Therefore, we use symmetry padding in accordance with reference [19] as the default of our model.

### 3.5. Anomaly score

As a GAN-based framework, we define the anomaly score based on the reconstruction error. It contains two parts, one is the error between test image $x$ and the restored image $\hat{x}$, the other one is the error between $z$ and $\tilde{z}$ correspond respectively the encoding of $x$ and $\tilde{x}$. Anomaly score is the weighted sum of them as follows

$$s_{rec} = \|x - \hat{x}\|_1 + \lambda_s\|z - \tilde{z}\|_1$$

where $\lambda_s=10$ is the weight we used.

During test phase, the joint distribution of restored image and its encoding are mapped into the distribution of the training set. The in-distribution sample gets a lower anomaly score, otherwise high anomaly score for anomaly. At last, for the evaluation of area under the curve (AUC) of the receiver operating characteristic (ROC), we normalize the anomaly scores based on the test results as follows

$$s_i' = \frac{s_i - \min(s_{rec})}{\max(s_{rec}) - \min(s_{rec})}$$ (16)

On the other hand, the trained discriminator enables us to use the predicted results for anomaly detection as well as other discriminative methods. Take Dirichlet score used in GEOM [13] as an example, the normality score of an image $x$ is

$$s_{dir} = \sum_{i=0}^{k-1} (\alpha_i - 1) \log D_{cls} (T_i(x), En (T_i(x)))$$ (17)

where $T_i(x)$ is the $i^{th}$ transformation of $x$, $\alpha_i$ is a constant determined by $T_i(x)$. The Dirichlet score calculates the degree of anomaly of each sample. Higher scores indicate a more normal sample. We find that $s_{dir}$ can get better indicators than $s_{rec}$ in our experiment, however it needs to use the training data set and greatly increases the computational complexity. Nevertheless, we will present both scores in the ablation studies to provide options for trade-off between speed and performance.

### 4. Experiments

#### 4.1. Datasets and baselines

We evaluated our model on MNIST [22], CIFAR-10 [21], and MVTAD [4] for one-class anomaly detection. We compare our method to several classical methods and state-of-the-art GAN-based methods for one-class anomaly detection. They are AnoGAN [37], DSVDD [36], GEOM [13], OCGAN [34], LSA [1], DROCC [15] and CA VGA [43]. Most of the results of them are the reported results in the original papers.

#### 4.2. Network structure and training details

The encoder consists of several convolution layers and residual blocks [17]. The input enters a 7x7 convolution layer with stride 1, two 4x4 convolution layers with stride 2 for down-sampling, three residual blocks layers and a 3x3 convolution layer with stride 1 in turn. Instance normalization (IN) [42] is used in all layers follows by a ReLU activation except the last output layer. A tanh activation was placed after the last convolution layer to restrict the output of the latent dimension. The decoder is the symmetry of the encoder. The difference is that following by a bilinear interpolation and 5x5 convolution layer for up-sampling after
### Table 1. One-class anomaly detection AUC results for MNIST dataset.

| Methods          | 0      | 1      | 2      | 3      | 4      | 5      | 6      | 7      | 8      | 9      | Mean   |
|------------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| AnoGAN/IPMI'17   | 0.966  | 0.992  | 0.850  | 0.887  | 0.894  | 0.883  | 0.947  | 0.935  | 0.849  | 0.924  | 0.9127 |
| DSVDD/CIFAR'18   | 0.980  | 0.997  | 0.917  | 0.919  | 0.949  | 0.885  | 0.983  | 0.946  | 0.939  | 0.965  | 0.9480 |
| OCGAN/CVPR'19    | 0.998  | 0.999  | 0.942  | 0.963  | 0.975  | 0.980  | 0.991  | 0.981  | 0.939  | 0.981  | 0.9750 |
| LSA/CVPR'19      | 0.993  | 0.999  | 0.999  | 0.966  | 0.956  | 0.964  | 0.994  | 0.980  | 0.953  | 0.984  | 0.9750 |
| CAVGA/ECCV'20    | 0.994  | 0.997  | 0.989  | 0.983  | 0.977  | 0.986  | 0.988  | 0.936  | 0.988  | 0.991  | 0.9860 |
| DGAD (Protocol 1)| 0.9934 | 0.9981 | 0.9876 | 0.9811 | 0.9853 | 0.988  | 0.9965 | 0.9837 | 0.9631 | 0.9872 | 0.9964 |
| DGAD (Protocol 2)| 0.9966 | 0.9987 | 0.9636 | 0.9549 | 0.9649 | 0.9663 | 0.9915 | 0.9668 | 0.9708 | 0.9735 | 0.9748 |

### Table 2. One-class anomaly detection AUC results for CIFAR-10 dataset.

| Methods          | Plane  | Car    | Bird   | Cat    | Deer   | Dog    | Frog   | Horse  | Ship   | Truck  | Mean   |
|------------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| AnoGAN/IPMI'17   | 0.671  | 0.547  | 0.529  | 0.545  | 0.651  | 0.603  | 0.585  | 0.625  | 0.758  | 0.665  | 0.6179 |
| LSA/CVPR'19      | 0.735  | 0.580  | 0.690  | 0.542  | 0.761  | 0.546  | 0.751  | 0.535  | 0.717  | 0.548  | 0.641  |
| DSVDD/CIFAR'18   | 0.617  | 0.659  | 0.508  | 0.591  | 0.609  | 0.637  | 0.677  | 0.673  | 0.739  | 0.731  | 0.6481 |
| OCGAN/CVPR'19    | 0.757  | 0.531  | 0.640  | 0.620  | 0.723  | 0.620  | 0.723  | 0.575  | 0.820  | 0.554  | 0.6566 |
| CAVGA/ECCV'20    | 0.653  | 0.784  | 0.761  | 0.747  | 0.775  | 0.552  | 0.813  | 0.745  | 0.801  | 0.741  | 0.737  |
| DROCC/CIFAR'20   | 0.817  | 0.767  | 0.667  | 0.671  | 0.736  | 0.744  | 0.744  | 0.714  | 0.800  | 0.762  | 0.7423 |
| DGAD (Protocol 1)| 0.746  | 0.876  | 0.732  | 0.711  | 0.766  | 0.804  | 0.760  | 0.884  | 0.862  | 0.872  | 0.8012 |
| DGAD (Protocol 2)| 0.800  | 0.855  | 0.688  | 0.658  | 0.662  | 0.761  | 0.713  | 0.805  | 0.854  | 0.833  | 0.7629 |

The discriminator is based on the BiGAN architecture. After two convolution layers for down-sampling, $z$ concatenates with the feature map of $x$ and into the next convolution layer. Leaky ReLU is used in all layers with a negative slope of 0.01 except the last output layer. The number of parameters of discriminator is about 3.5M.

Our model is trained using Adam with $\beta_1 = 0.5$ and $\beta_2 = 0.999$. The batch size is set to 64 for CIFAR-10 and MNIST dataset. We perform one by one update for generator and discriminator with an initial learning rate of 0.0001 for 10000 iterations. We use spectral normalization [27] in the discriminator. We use symmetry padding as defaults in convolution. The model is implemented by Tensorflow 2.1 and the training takes about 50 minutes per class on a single NVIDIA 2080TI GPU for CIFAR-10.

### 4.3. MNIST and CIFAR-10 for anomaly detection

In order to achieve one-class detection, one class at each time is considered as the normal class in training, and other class are regarded as anomaly in testing. We evaluate the performance by AUC which is commonly used for evaluating performance in anomaly detection tasks. The results of anomaly detection on MNIST and CIFAR-10 are presented in Table 1 and Table 2 respectively. They show the AUC value of each class and the total average AUC values. The proposed DGAD with Protocol 1 outperforms the compared methods for MNIST. DGAD with Protocol 1 and Protocol 2 both outperform the compared methods on CIFAR-10. DGAD increases the performance of the top-performing GAN-based baseline (CA VGA) by 6% on CIFAR-10.

However, the performance of Protocol 2 is weaker than that of Protocol 1, indicating that rotation is a more effective pretext task for our method. And the simplicity of rotation makes its advantages more obvious. However, the rotation-based pretext task is meaningless for a centrally-symmetric image, which limits its application. The existence of Protocol 2 and Protocol 3 is still necessary.

### 4.4. MVTec for anomaly detection

For MVTec AD dataset, we perform random zoom augmentation for each category and resize all the image to 128x128 in training and testing. Training is conducted for 6000 epochs with batch size 8 on normal data.

Considering that many images in the data set are either centrally symmetric or textured, we use Protocol 3 as the pretext task. As shown in Table 3, our method achieves the best average performance (2% higher than CAVGA). However, we can see from the table that our method is not suitable for detecting texture anomalies (e.g. carpet and grid), and some tiny anomalies (e.g. capsule and screw). This is the inherent defect of relying on the reconstruction error to measure the anomaly, which will be the focus of improvement in future work.

### 4.5. Ablation studies

The ablation studies have three purposes: verify the correctness of each conjecture in our method, compare the differences between generative method and discriminative method, and explore the influence of absolute location information on self-supervised representation learning.

All the studies are on the CIFAR-10 dataset as Table 4 shows. We compared Protocol 1 with Protocol 2 based on DGAD. There are two parts ($s_{rec}$ and $s_{dir}$) and eight variants both in Protocol 1 and Protocol 2. GAN represents learning representations only by pixel-level image re-

---

1. Our code and models are available at github.com/*.
construction, not by discriminator guidance. **DGAD - CL** represents the DGAD without compactness loss. **DGAD + zero-padding** represents using zero padding instead of symmetry padding in the convolution (introducing absolute location information). **DGAD + coord** represents using coordconv [26] for more direct absolute location information. **DGAD - CL** represents the DGAD without compactness loss.

As we can see in Table 4, DGAD with **Protocol 1** and **Protocol 2** both significantly increase AUC by 5.4~7.5% compared to **GAN**, which proves that the guidance of discriminator is more beneficial to representation learning than pixel-level reconstruction. And the absence of compactness loss reduces performance by 2~2.9%, which demonstrates its effectiveness.

In part one of **Protocol 1**, the experimental results of zero-padding and coordconv confirm the interference of absolute position information on representation learning. With the help of absolute location information for geometric transformation prediction, encoder and decoder become lazy to learn more semantic representations, resulting in a 1~2% performance reduction.

In part two of **Protocol 1**, on the contrary, the discriminator improves the performance of anomaly detection with the aid of absolute position information, which indicates that the absolute location information can make it easier for the discriminator to classify and recognize objects and scenes, but it will correspondingly weaken the generator’s learning of abstract representations.

However, the influence of absolute location information in **Protocol 2** is changed. The performance of both generator and discriminator is degraded with absolute location information. Experiments show that the effect of location information on performance is related to the choice of pretext tasks. Further experiments show that this is caused by the discriminator’s generalization capability under different pretext tasks (see supplementary materials). Therefore, we suggest making a careful analysis of the tasks to be faced before model design.

It is worth noting that \( s_{rec} \) is slightly lower than \( s_{dir} \) by 1~3\%. This shows that although the reconstruction loss is simple, it can be close to the score of discriminative method which relies on complex post-processing. And our approach shows a great advantage in speed as shown in Table 5. The calculation of \( s_{rec} \) is 3.6 to 8.5 times faster than that of \( s_{dir} \), even though the discriminator has a smaller number of parameters than the generator. The more kinds of transformation, the more time it takes to calculate \( s_{dir} \), while the calculation of \( s_{rec} \) has nothing to do with the pretext task and has a constant computation.

### 5. CONCLUSION

In this paper, we propose a novel GAN-based anomaly detection method with the help of discriminative-generative...
representation learning, named DGAD. Under the guidance of discriminator, the generator can better learn representation and avoid the problem of label semantic ambiguity. DGAD increases the performance of the top-performing GAN-based baseline by 6% on CIFAR-10 and 2% on MVTAD. DGAD approaches the performance of discriminative method and has a great advantage in speed. As an additional conclusion, we find that absolute position information has different effects on the representational learning ability of generative methods and discriminative methods, which indicates that the use of absolute position information should be carefully selected according to different pretext tasks. Considering the natural relationship between various pretext tasks and absolute location information, the interaction between them should arouse more attention of researchers. We believe that the representations learned by the generator should be better quantified to accurately measure the degree of abnormality, which will be the focus of our future research.

References

[1] Davide Abati, Angelo Porrello, Simone Calderara, and Rita Cucchiara. Latent space autoregression for novelty detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 481–490, 2019.

[2] Samet Akcay, Amir Atapour-Abarghouei, and Toby P Breckon. Ganomaly: Semi-supervised anomaly detection via adversarial training. In Asian conference on computer vision, pages 622–637. Springer, 2018.

[3] Liron Bergman and Yedid Hoshen. Classification-based anomaly detection for general data. arXiv preprint arXiv:2005.02359, 2020.

[4] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Mvtec ad—a comprehensive real-world dataset for unsupervised anomaly detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 9592–9600, 2019.

[5] Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. ACM computing surveys (CSUR), 41(3):1–58, 2009.

[6] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In International conference on machine learning, pages 1597–1607. PMLR, 2020.

[7] Carl Doersch, Abhinav Gupta, and Alexei A Efros. Unsupervised visual representation learning by contrast prediction. In Proceedings of the IEEE international conference on computer vision, pages 1422–1430, 2015.

[8] Jeff Donahue, Philipp Krähenbühl, and Trevor Darrell. Adversarial feature learning. arXiv preprint arXiv:1605.09782, 2016.

[9] Jeff Donahue and Karen Simonyan. Large scale adversarial representation learning. arXiv preprint arXiv:1907.02544, 2019.

[10] Vincent Dumoulin, Ishmael Belghazi, Ben Poole, Olivier Mastropietro, Alex Lamb, Martin Arjovsky, and Aaron Courville. Adversarially learned inference. arXiv preprint arXiv:1606.00704, 2016.

[11] Ye Fei, Chaoqin Huang, Cao Jinkun, Maosen Li, Ya Zhang, and Cewu Lu. Attribute restoration framework for anomaly detection. IEEE Transactions on Multimedia, 2020.

[12] Spyros Gidaris, Praveer Singh, and Nikos Komodakis. Unsupervised representation learning by predicting image rotations. arXiv preprint arXiv:1803.07728, 2018.

[13] Izhak Golan and Ran El-Yaniv. Deep anomaly detection using geometric transformations. arXiv preprint arXiv:1805.10917, 2018.

[14] Dong Gong, Lingqiao Liu, Vuong Le, Budhadiya Saha, Moussa Reda Mansour, Svetla 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, pages 1705–1714, 2019.

[15] Sachin Goyal, Aditi Raghunathan, Mohsh Jain, Harsha Vardhan Simhadri, and Prateek Jain. Drocc: Deep robust one-class classification. In International Conference on Machine Learning, pages 3711–3721. PMLR, 2020.

[16] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 9729–9738, 2020.

[17] 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, pages 770–778, 2016.

[18] Dan Hendrycks, Mantas Mazeika, Saurav Kadavath, and Dawn Song. Using self-supervised learning can improve model robustness and uncertainty. arXiv preprint arXiv:1906.12340, 2019.

[19] Md Amirul Islam, Sen Jia, and Neil DB Bruce. How much position information do convolutional neural networks encode? arXiv preprint arXiv:2001.08248, 2020.

[20] Daiki Kimura, Subhajit Chaudhury, Minori Narita, Asim Sapatnekar, and Ryuki Tachibana. Adversarial discriminative attention for robust anomaly detection. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pages 2172–2181, 2020.

[21] Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. Handbook of Systemic Autoimmune Diseases, 2009.

[22] Yann LeCun. The mnist database of handwritten digits. http://yann.lecun.com/exdb/mnist/, 1998.

[23] Christian Ledig, Lucas Theis, Ferenc Huszår, José Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 4681–4690, 2017.

[24] Kibok Lee, Kimin Lee, Kyle Min, Yuting Zhang, Jinwoo Shin, and Honglak Lee. Hierarchical novelty detection for
visual object recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1034–1042, 2018.

Chunyuan Li, Hao Liu, Changyou Chen, Yunchen Pu, Liqun Chen, Ricardo Henao, and Lawrence Carin. Alice: Towards understanding adversarial learning for joint distribution matching. arXiv preprint arXiv:1709.01215, 2017.

Rosanne Liu, Joel Lehman, Piero Molino, Felipe Petroski Such, Eric Frank, Alex Sergeev, and Jason Yosinski. An intriguing failing of convolutional neural networks and the coordconv solution. arXiv preprint arXiv:1807.03247, 2018.

Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida. Spectral normalization for generative adversarial networks. arXiv preprint arXiv:1802.05957, 2018.

Miguel Nicolau, James McDermott, et al. Learning neural representations for network anomaly detection. IEEE transactions on cybernetics, 49(8):3074–3087, 2018.

Mehdi Noroozi and Paolo Favaro. Unsupervised learning of visual representations by solving jigsaw puzzles. In European conference on computer vision, pages 69–84. Springer, 2016.

Poojan Oza and Vishal M Patel. One-class convolutional neural network. IEEE Signal Processing Letters, 26(2):277–281, 2018.

Hyunjong Park, Jonyoun Noh, and Bumsun Ham. Learning memory-guided normality for anomaly detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 14372–14381, 2020.

Deepak Pathak, Philipp Krahenbuhl, Jeff Donahue, Trevor Darrell, and Alexei A Efros. Context encoders: Feature learning by inpainting. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 2536–2544, 2016.

Pranathitha Perera, Vlad I Morariu, Rajiv Jain, Varun Manjunatha, Curtis Wigington, Vicente Ordonez, and Vishal M Patel. Generative-discriminative feature representations for open-set recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 11814–11823, 2020.

Pranathitha Perera, Ramesh Nallapatii, and Bing Xiang. Ogan: One-class novelty detection using gans with constrained latent representations. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 2898–2906, 2019.

Stanislav Pidhorskyi, Ranya Almohsen, Donald A Adjeroh, and Gianfranco Doretto. Generative probabilistic novelty detection with adversarial autoencoders. arXiv preprint arXiv:1807.02588, 2018.

 Lukas Ruff, Robert Vandermeulen, Nico Goernitz, Lucas Deecke, Shoaih Ahmed Siddiqui, Alexander Binder, Emanuel Müller, and Marius Kloft. Deep one-class classification. In International conference on machine learning, pages 4393–4402. PMLR, 2018.

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, pages 146–157. Springer, 2017.

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.

David MJ Tax and Robert PW Duin. Uniform object generation for optimizing one-class classifiers. Journal of machine learning research, 2(Dec):155–173, 2001.

David MJ Tax and Robert PW Duin. Support vector data description. Machine learning, 54(1):45–66, 2004.

Xiaohua Zhai Ting Chen, Mario Lucic Marvin Ritter, and Neil Houlsby. Self-supervised gans via auxiliary rotation loss. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, page 12154–12163. Springer, 2019.

Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Instance normalization: The missing ingredient for fast stylization. arXiv preprint arXiv:1607.08022, 2016.

Shashanka Venkataramanan, Kuan-Chuan Peng, Rajat Vikram Singh, and Abhijit Mahalanobis. Attention guided anomaly localization in images. In European Conference on Computer Vision, pages 485–503. Springer, 2020.

Jiayu Wang, Wengang Zhou, Guo-Jun Qi, Zhongqian Fu, Qi Tian, and Houqiang Li. Transformation gan for unsupervised image synthesis and representation learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 472–481, 2020.

Qizhe Xie, Zihang Dai, Yulun Du, Eduard Hovy, and Graham Neubig. Controllable invariance through adversarial feature learning. arXiv preprint arXiv:1705.11122, 2017.

Weiying Xie, Baozhu Liu, Yunsong Li, Jie Lei, Chein-I Chang, and Gang He. Spectral adversarial feature learning for anomaly detection in hyperspectral imagery. IEEE Transactions on Geoscience and Remote Sensing, 58(4):2352–2365, 2019.

Muhammad Zaigham Zaheer, Jin-ha Lee, Marcella Astrid, and Seung-Ilk Lee. Old is gold: Redefining the adversarially learned one-class classifier training paradigm. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 14183–14193, 2020.

Houssam Zenati, Chuan Sheng Foo, Bruno Lecouat, Gaurav Manek, and Vijay Chandrasekhar. Chancrasekhar. Efficient gan-based anomaly detection. arXiv preprint arXiv:1802.06222, 2018.

Houssam Zenati, Manon Romain, Chuan-Sheng Foo, Bruno Lecouat, and Vijay Chandrasekhar. Adversarially learned anomaly detection. In 2018 IEEE International conference on data mining (ICDM), pages 727–736. IEEE, 2018.

Shuangfei Zhai, Walter Talbott, Carlos Guestrin, and Joshua M Susskind. Adversarial fisher vectors for unsupervised representation learning. arXiv preprint arXiv:1910.13101, 2019.

Richard Zhang, Phillip Isola, and Alexei A Efros. Colorful image colorization. In European conference on computer vision, pages 649–666. Springer, 2016.
[52] Richard Zhang, Phillip Isola, and Alexei A Efros. Split-brain autoencoders: Unsupervised learning by cross-channel prediction. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1058–1067, 2017.

[53] Kang Zhou, Shenghua Gao, Jun Cheng, Zaiwang Gu, Huazhu Fu, Zhi Tu, Jianlong Yang, Yitian Zhao, and Jiang Liu. Sparse-gan: Sparsity-constrained generative adversarial network for anomaly detection in retinal oct image. In 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), pages 1227–1231. IEEE, 2020.
6. Supplementary Materials

In this document, we provide more experimental results and analyses that have not been presented in the original due to the space limitations.

6.1. Qualitative analysis of model performance

This section consists of two purposes: one is to verify that the encoder and decoder have indeed learned the abstract semantic representation; the other is to visualize the test results to verify our theory.

Figure 4 shows the reconstruction results of the rotated images by encoder and decoder. The figure shows the results of two models, one for airplane and the other for automobile. As we can see, our model can reconstruct the original image $x_r$ and restore the rotated image $x_t$ to the normal angle image $\hat{x}_t$. This confirms that our model achieves the goal of the pretext task. On the other hand, we can see that the restored image does not always match the original image, because we do not impose pixel-level consistency constraints. This confirms that our model indeed learned abstract semantic representations of geometric features, not just textures. In particular, the red box shows an example of a rotated image that remains unchanged before and after restoration. This phenomenon is consistent with our hypothesis. Since the rotated image still retains the visual rationality, the discriminator cannot predict its rotation angle. Therefore, the decoder only reconstructs the image without forcibly restoring it to the original image. Our model has better representation learning ability because it avoids the meaningless image restoration.

Figure 5 shows the reconstruction results of test images. We can see that the reconstruction results of out-of-distribution images tend to be in-distribution images, resulting in a huge difference from the original images. Meanwhile, the normal images can be reconstructed into similar images, which provides us with a tool to distinguish the anomaly by the reconstruction error.

6.2. Quantitative analysis of model performance

This section consists of two purposes: the first is to verify the authenticity of our experimental results through visualization, and the second is to verify the relevant conclusions in the paper through supplementary experiment results.

Figure 6 shows the visualization of anomaly detection results of automobile. As can be seen from the figure, there are two different distributions of anomaly scores of abnormal and normal in Figure 6(a). The anomaly scores of normal images are concentrated in low scores, while those of abnormal images are the opposite. However, the two distributions still have overlapping areas, leading to an error rate of 0.2 (Figure 6(b)), which needs to be improved in the future.

As the ablation studies and comparison experiments on CIFAR-10 dataset shown in Table 4 in the paper, absolute position information can degrade the representation learning ability of generator but benefit discriminator in Protocol 1. However, the Dirichlet scores of DGAD + zero-padding ($s_{d, dir}$) and DGAD + coord ($s_{d, dir}$) are also degraded in Protocol 2, even worse. To study the causes of this phenomenon, we compare the classification accuracy of discriminator as shown in Table 6 and Table 7.

We only use the normal samples in the test set to calculate the prediction accuracy of the discriminator in Table 6. For each image, we generate six kinds of transformed images and their corresponding labels according to the pretext task. We can see that absolute position information does improve the accuracy of the prediction by about 3%. However, when we use the whole test set to calculate the
Table 6. Classification accuracy of discriminator (only normal samples in test set).

| Methods          | Plane | Car | Bird | Cat | Deer | Dog | Frog | Horse | Ship | Truck | Mean |
|------------------|-------|-----|------|-----|------|-----|------|-------|------|-------|------|
| Protocol 2       |       |     |      |     |      |     |      |       |      |       |      |
| DGAD + zero-padding | 0.946 | 0.979 | 0.961 | 0.939 | 0.966 | 0.966 | 0.973 | 0.971 | 0.962 | 0.971 | 0.965 |
| DGAD + coord     | 0.957 | 0.976 | 0.943 | 0.971 | 0.941 | 0.974 | 0.966 | 0.968 | 0.967 | 0.965 | 0.963 |
| DGAD             | 0.920 | 0.969 | 0.906 | 0.916 | 0.920 | 0.937 | 0.935 | 0.941 | 0.946 | 0.946 | 0.934 |

Table 7. Classification accuracy of discriminator (all samples in test set).

| Methods          | Plane | Car | Bird | Cat | Deer | Dog | Frog | Horse | Ship | Truck | Mean |
|------------------|-------|-----|------|-----|------|-----|------|-------|------|-------|------|
| Protocol 2       |       |     |      |     |      |     |      |       |      |       |      |
| DGAD + zero-padding | 0.882 | 0.840 | 0.911 | 0.942 | 0.881 | 0.925 | 0.905 | 0.890 | 0.848 | 0.869 | 0.889 |
| DGAD + coord     | 0.854 | 0.832 | 0.928 | 0.929 | 0.905 | 0.925 | 0.904 | 0.889 | 0.830 | 0.870 | 0.887 |
| DGAD             | 0.409 | 0.507 | 0.433 | 0.355 | 0.254 | 0.516 | 0.397 | 0.405 | 0.605 | 0.768 | 0.465 |

The distribution of abnormal scores shows that the use of absolute position information greatly improves the prediction accuracy for abnormal samples. The absolute position information enables the discriminator to obtain good prediction accuracy for both normal samples and abnormal samples (more than 0.88), which leads to the reduction of the difference between their abnormal scores. On the contrary, DGAD without absolute position information has poor prediction accuracy for abnormal samples (lower than 0.47), which makes normal and abnormal can be well distinguished. The absolute position information makes the discriminator too generalizable, which is the reason for the performance degradation of anomaly detection.

Interestingly, we did not find this phenomenon in the rotation pretext task, which supports the point in our paper: the use of absolute position information should be carefully selected according to different pretext tasks.

6.3. Code

We provide the code of DGAD based on CIFAR-10. Reviewers can use our code to reproduce or test our model. Reviewers may need to upgrade Seaborn and Matplotlib to run the tests correctly.

Train: python main.py
This command will train each of the 10 categories of CIFAR-10 in turn.

Test: python main.py --phase test
This command will test each of the 10 categories of CIFAR-10 in turn. Or,

Test: python main.py --phase test --test_object 1
This command will test category 1 (automobile) of CIFAR-10.

The visual results (distribution diagram and ROC) of the test can be found in ./results.