OMASGAN: Out-of-Distribution Minimum Anomaly Score GAN for Sample Generation on the Boundary

Nikolaos Dionelis [1] **Mehrdad Yaghoobi** [1] **Sotirios A. Tsaftaris** [1]

Abstract

Generative models trained in an unsupervised manner encounter the problem of setting high likelihood and low reconstruction loss to Out-of-Distribution (OoD) samples. This increases the Type II errors (false negatives, misses of anomalies) and decreases the Anomaly Detection (AD) performance. Also, deep generative models for AD suffer from the rarity of anomalies. To address these limitations, we propose the new OoD Minimum Anomaly Score GAN (OMASGAN) model. OMASGAN addresses the rarity of anomalies by generating strong abnormal samples on the boundary of the support of the data distribution, using data only from the normal class. OMASGAN improves the AD performance by retraining including the abnormal minimum-anomaly-score OoD samples generated by our negative sampling augmentation methodology. OMASGAN uses any f-divergence distribution metric in its variational representation, and explicit likelihood and invertibility are not needed. The proposed model uses a discriminator for inference and the evaluation of OMASGAN on images using the leave-one-out methodology shows that it achieves an improvement of at least 0.24 and 0.07 points in AUROC on average on MNIST and CIFAR-10, respectively, over recent state-of-the-art benchmarks.

1. Introduction

Anomaly Detection (AD) in high-dimensional spaces is challenging. In spite of progress ushered in by the use of generative models such as Generative Adversarial Networks (GAN), AD performance remains a challenge. Models learn to assign high probability to the seen data but are not trained to assign zero probability to Out-of-Distribution (OoD) samples. During inference, they assign non-zero probability to anomalies and this leads to a high number of false negatives (Nalisnick et al., 2019; Kirichenko et al., 2020). To address such limitations, we propose the OoD Minimum Anomaly Score GAN (OMASGAN), a new boundary-based model for training GANs for AD. OMASGAN uses data from the normal class and generates OoD samples as anomalies that are close to the boundary of the data distribution having a minimum anomaly score around the data. OMASGAN performs retraining by including the boundary samples, negative training, and AD with negative sampling (Sipple, 2020; Sinha et al., 2021). It can be incorporated with any GAN, including (Song & Ermon, 2020) and the f-divergence GAN (f-GAN) (Nowozin et al., 2016). Our contributions are:

- We propose OMASGAN, a model for AD to perform retraining by including the boundary samples created by our negative sampling augmentation methodology.
- To address the rarity of anomalies, we create abnormal samples using data only from the normal class and find the minimum-anomaly-score samples on the boundary of the support of the data distribution using any f-divergence without likelihood and/or invertibility.
- We train a discriminator to separate the data distribution from its complement and evaluate the use of inference for AD. The evaluation of OMASGAN using the leave-one-out (LOO) methodology shows that it achieves state-of-the-art performance in the Area Under the Receiver Operating Characteristics curve (AUROC) metric, outperforming recent benchmarks.

2. The Proposed OMASGAN Model

We propose an algorithm to address the problem of generative models setting high likelihood and low reconstruction loss to OoD samples which leads to Type II errors (false negatives, misses of anomalies) and decreases the AD performance. Figure 1 presents the flowchart of OMASGAN which has the following structure: (a) Train a f-divergence GAN using the data to obtain the implicit generator, $G(\mathbf{z})$. (b) Train the boundary model, $B(\mathbf{z})$, to find the boundary of G, where the boundary samples are the generated negative

[1]School of Engineering, University of Edinburgh, UK. Correspondence to: Nikolaos Dionelis <nikolaos.dionelis@ed.ac.uk>.

Corresponding Author: <dionelisnk@gmail.com>
Nikolaos Dionelis, <nikolaos.dionelis2021@gmail.com>

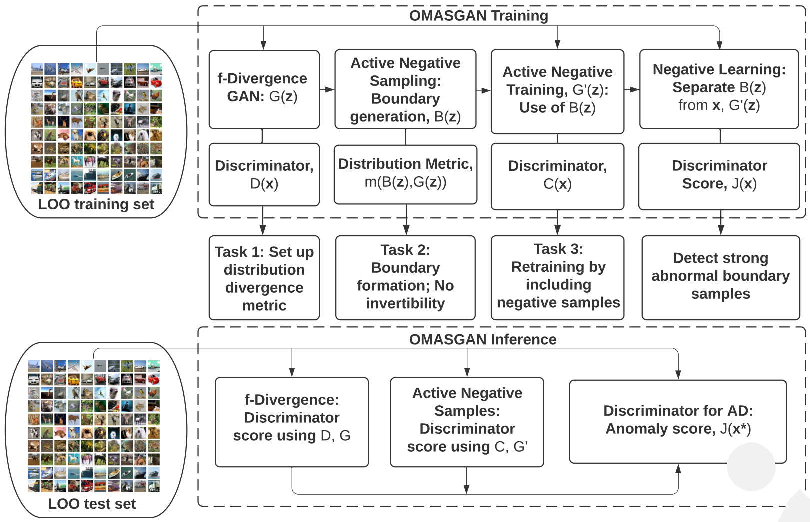


Figure 1. Flowchart of the OMASGAN model for AD which generates minimum-anomaly-score OoD samples on the boundary of the data distribution and subsequently uses these generated boundary samples to train a discriminative model to detect abnormal samples.

points and active negative sampling is performed using any f-divergence in its variational representation to compute the statistical divergence between the $B(\mathbf{z})$ and $G(\mathbf{z})$ samples. OMASGAN generates samples corresponding to a generalized notion of the boundary of the support of the data distribution, which is the set of points such that they are OoD and have a minimum anomaly score measured as the f-divergence metric or the Wasserstein distance. We name these points the OoD Minimum Anomaly Score (OMAS) samples and incorporate them in the proposed algorithm. (c) Perform active negative training for AD using the OMAS samples and the implicit distributions from (a) and (b), and train the generator $G'(\mathbf{z})$ using a discriminator $C(\mathbf{x})$. Then, we train the discriminator $J(\mathbf{x})$ for AD using active negative training, the OMAS samples, $B(\mathbf{z})$, and $G'(\mathbf{z})$. During inference, we compute the proposed anomaly score and detect anomalies using $J(\mathbf{x})$ and the trained OMASGAN model.

2.1. The OMASGAN Algorithm

Our algorithm comprises the following optimization tasks:

Task 1. Distribution metric. We train a f-divergence GAN to learn the data. Using $\mathbf{x} \sim p_{\mathbf{x}}$, and $G(\mathbf{z}) \sim p_g$,

$$\arg\min_G \max_D \mathbb{E}_{\mathbf{x}} \log(D(\mathbf{x})) + \mathbb{E}_{\mathbf{z}} \log(1 - D(G(\mathbf{z}))). \quad (1)$$

Output: The generator-discriminator pair, (G, D). *Output used in subsequent Tasks*: The implicit generator, $G(\mathbf{z})$.

Task 2. Formation of the boundary of the data distribution. To perform active negative sampling, we create and train the boundary model, $B(\mathbf{z})$. *Optimization problem*:

$$\begin{aligned} \arg\min_{\boldsymbol{\theta_b}} & -m(B(\mathbf{z};\boldsymbol{\theta_b}), G(\mathbf{z})) \\ & + \mu\, d(B(\mathbf{z};\boldsymbol{\theta_b}), G(\mathbf{z})) + \nu\, s(B(\mathbf{z};\boldsymbol{\theta_b}), \mathbf{z}) \end{aligned} \quad (2)$$

where $m(B,G)$ is the distribution metric from Task 1, i.e. any f-divergence in its variational representation expressed in terms of the convex conjugate function, $f^*(t)$, as in (7) in (Nowozin et al., 2016), where t is a variational function taking as input a sample and returning a scalar. The special cases of KL and Pearson are $f^*(t) = \exp(t-1)$ and $f^*(t) = 0.25t^2 + t$, respectively. The first term in (2) is a *strictly decreasing function* of a distribution metric. This divergence is between the boundary samples and the data where $m(B(\mathbf{z}), G(\mathbf{z}))$ is the distribution metric of choice, such as Jensen-Shannon of the original GAN, Kullback–Leibler (KL), and Pearson Chi-Squared of the Least Squares GAN to address saturation, or the Wasserstein distance to address mode collapse. Our boundary model does not need probability and invertibility obviating the rarity and sampling complexity problem by using a decreasing function of a distribution metric and any f-divergence to find the boundary of the data distribution.

Generation of the OMAS distribution and the OMAS samples: OMASGAN computes the minimum-anomaly-score OoD samples utilizing the loss in (2). It performs automatic negative data augmentation by eliminating the need for feature extraction and human intervention, and this strengthens the applicability of our model. We denote the parameters of the boundary model by $\boldsymbol{\theta_b}$ and its samples by $B(\mathbf{z};\boldsymbol{\theta_b})$. We use a f-divergence GAN discriminator to compute the distribution metric, $m(B,G)$. The combination of the first two terms leads the B samples to the boundary of $p_g(\mathbf{x})$. We compute the boundary with l_p-norm distance and dispersion regularization and we denote the l_p-norm distance between the point $B(\mathbf{z})$ and the set $\mathbf{x}$ by $d(B(\mathbf{z}), \mathbf{x})$. To capture all the modes of the data distribution, address mode collapse (Dionelis et al., 2020; Abbasnejad et al., 2020), and generate

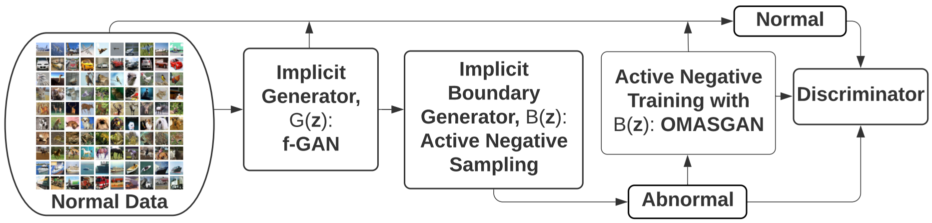


Figure 2. Training of the proposed OMASGAN model using active negative sampling and training by creating *strong* abnormal samples.

OMAS samples, we use a scattering measure for dispersion and regularization in (2). This scattering measure, $s(B(\mathbf{z}),\mathbf{z})$, and the l_p-norm distance, $d(B(\mathbf{z}),G(\mathbf{z}))$, are given by

$$d(B(\mathbf{z}),G(\mathbf{z})) = \min_{j=1,\dots,Q} ||B(\mathbf{z})-G(\mathbf{z}_j)||_2 \quad (3)$$

$$s(B(\mathbf{z}_i),\mathbf{z}_i) = \frac{1}{N-1}\sum_{j=1,\, j\neq i}^{N} \frac{||\mathbf{z}_i-\mathbf{z}_j||_2}{||B(\mathbf{z}_i)-B(\mathbf{z}_j)||_2} \quad (4)$$

where the batch size is denoted by N and the inference size by Q. OMASGAN achieves good modeling by finding the boundary of a model of $\mathbf{x}$ rather than of $\mathbf{x}$. In (2), the output of Task 1, $G(\mathbf{z})$, is used because it has established and set up a distribution metric in the data space, $\mathcal{X}$. OMASGAN finds the minimum-anomaly score OoD samples, performs active sampling of negative examples, and generates strong and specifically *adversarial* anomalies that lie close to the boundary of the data distribution and near high-probability data samples. The trainable model in Task 2 is the implicit boundary distribution $B(\mathbf{z};\boldsymbol{\theta_b})$, and $G(\mathbf{z})$ is non-trainable.

Output of Task 2 used in next Tasks: The boundary, $B(\mathbf{z})$. In summary, in Task 1, we use a f-divergence GAN for data distribution generation. In Task 2, OMASGAN performs active negative *sampling* and we train an implicit boundary generator to form the boundary. In Task 3, OMASGAN performs active negative *training* with the minimum-anomaly score OoD boundary samples, and trains a discriminator, $J(\mathbf{x})$, to classify normal and abnormal data as shown in Figure 2.

Task 3. Active negative training. Separation of generated and real normal from generated abnormal data for AD. To address the learning-OoD-samples problem of G (Nalisnick et al., 2019; Kirichenko et al., 2020), we perform model retraining for AD by including the negative samples created by our negative samples augmentation methodology in Task 2. OMASGAN introduces self-supervision using the boundary samples from Task 2. We perform *retraining* by including the abnormal $B(\mathbf{z})$ points. To train G', we use

$$\arg\min_{G'}\max_{C}\ \alpha\,\mathbb{E}_{\mathbf{x}}\log(C(\mathbf{x}))+\beta\,\mathbb{E}_{\mathbf{z}}\log(1-C(G'(\mathbf{z}))) + \gamma\mathbb{E}_{\mathbf{z}}\log(1-C(B(\mathbf{z})))+\delta\,\mathbb{E}_{\mathbf{z}}\log(C(G(\mathbf{z}))) \quad (5)$$

where $G'(\mathbf{z}) \sim p_{g'}$ lie in $\mathcal{X}$ and where C is a discriminator that computes distribution metrics and f-divergences, as in (Zaheer et al., 2020; Asokan & Seelamantula, 2020).

To calculate divergences between distributions, and in this case between $B(\mathbf{z})$ and $(\mathbf{x},G(\mathbf{z}))$, we use a discriminator and a *weighted sum* of f-divergences and probability metrics (Zaheer et al., 2020; Asokan & Seelamantula, 2020). The nested optimization in (5) comprises four terms and outputs the learned mappings $C:\mathcal{X}\rightarrow\mathbb{R}$ where the data space is denoted by $\mathcal{X}$, and $G':\mathscr{Z}\rightarrow\mathcal{X}$ where $\mathscr{Z}$ is the latent space. The trainable models are the implicit generator $G'(\mathbf{z})$ and the discriminator $C(\mathbf{x})$, and B and G are non-trainable. The first and fourth terms of the minimax optimization force the generated samples to the data, as in Rumi-GAN (Asokan & Seelamantula, 2020). The third term forces the generated samples away from our generated strong anomalies, which are near the support boundary of the data distribution and close to high-probability data. The discriminative model, $C(\mathbf{x})$, is trained to separate $B(\mathbf{z})$ from $(\mathbf{x},G(\mathbf{z}))$, while the implicit generative model, G', learns the data keeping away from and avoiding the generated abnormal $B(\mathbf{z})$ samples.

OMASGAN computes the optimal points for negative sampling in Task 2 and performs active negative training using the boundary in Task 3. The samples from $B(\mathbf{z};\boldsymbol{\theta_b})$ are the *closest points* to the data from the normal class and optimal retraining is performed. This leads to improvements with respect to the state-of-the-art (Zaheer et al., 2020; Pourreza et al., 2021; Sinha et al., 2021), as discussed in Sections 4-6. OMASGAN performs *automatic* negative samples augmentation and retraining for AD by eliminating the need for feature extraction (aim of deep learning), human intervention, and ad hoc methods because they do not scale, and this strengthens our model's applicability and generalization.

Inputs to the algorithm of Task 3: The inputs to the discriminative model, C, are samples from the data space, $\mathcal{X}$, and the inputs to the generator, G', are samples from the latent space, $\mathscr{Z}$. *Task 3 Output*: The implicit distribution $G'(\mathbf{z})$.

Detection of strong abnormal boundary samples. To address the learning-OoD-samples problem and to perform negative training for AD, we train the discriminator $J(\mathbf{x})$,

$$\arg\max_J\ \zeta\,\mathbb{E}_{\mathbf{x}}\log(1-J(\mathbf{x}))+(1-\zeta)\,\mathbb{E}_{\mathbf{z}}\log(1-J(G'(\mathbf{z})))+\mathbb{E}_{\mathbf{z}}\log(J(B(\mathbf{z}))). \quad (6)$$

Inputs to final active negative learning: The inputs to the discriminator, J, are $\mathbf{x}$ and samples from the data space. The inputs to $G'(\mathbf{z})$ and $B(\mathbf{z})$ are samples from the latent space. In (6), $J(\mathbf{x})$ is trainable while G' and B are non-learnable. *Output*: The discriminative model, $J(\mathbf{x})$, which learns to separate $B(\mathbf{z})$ from the normal samples, from $\mathbf{x}$ and $G'(\mathbf{z})$.

Inference mechanism for AD: The discriminator, J, is used for inference for AD. For a queried test sample, $\mathbf{x}^*$, the anomaly score is given by $J(\mathbf{x}^*)$. We leverage J to detect anomalous samples (Zaheer et al., 2020). The classification decision is: $\mathbf{x}^*$ is from the normal class if $J(\mathbf{x}^*) < \tau$, where τ is a predefined threshold, and $\mathbf{x}^*$ is abnormal otherwise.

By including the negative samples, J learns to discriminate between the data distribution and its *complement*. OMASGAN generates minimum-anomaly-score OoD samples and subsequently trains a discriminator for AD using the generated boundary samples. Our model uses negative training by generating abnormal OoD data on the boundary because $B(\mathbf{z})$ samples from the boundary of the data distribution.

3. Properties of OMASGAN

3.1. Global and Local Properties of Task 1

Let $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m\}$ be the data from the normal class containing m examples. Let $p_{\mathbf{x}}$ be the data distribution, $\mathbf{x} \sim p_{\mathbf{x}}$. Let $\mathbf{z} \sim p_{\mathbf{z}}$ lie in $\mathscr{Z}$ and $G(\mathbf{z}) \sim p_g$ in $\mathcal{X}$. Task 1 of OMASGAN attains a global optimum, $p_g = p_{\mathbf{x}}$. Using the alternating Gradient Descent, convergence to a local optimum is guaranteed (Nowozin et al., 2016; Goodfellow et al., 2014).

3.2. Global Properties of OMASGAN Task 2

Proposition 1. Let $L(\boldsymbol{\theta_b}, \mathbf{z}, B, G)$ be the objective loss in (2) as a function of the parameters of our trainable neural network, $\boldsymbol{\theta_b}$. Let the set over which these parameters vary be compact. Our loss is a continuous function of its parameters. Then, (2) attains a global minimum [illegible] that is, there exists a $\boldsymbol{\theta_b}^*$ such that $L(\boldsymbol{\theta_b}^*, \mathbf{z}, B, G)$ has the lowest value.

Using the optimization theorem that a continuous function defined on a compact set attains a global minimum and a global maximum, i.e. Weierstrass Extreme Value Theorem, the global properties of (2) are as follows. Task 2 attains a global minimum if the loss is continuous and its domain is compact (*sufficient* condition). A function composition of continuous functions is continuous. Our composite loss for negative data augmentation is continuous as a function of the parameters of B and [illegible]. The terms and f-divergences in (2)-(4) are continuous as functions of the model parameters. Also, the set over which the trainable $\boldsymbol{\theta_b}$ vary is compact.

Following our discussion in Section 2, we can set the Wasserstein distance as the *distribution metric* in Tasks 1-3. The Wasserstein distance (a) has a strong weakness property, (b) is continuous with respect to its parameters because of (a), and (c) is weaker than f-divergences (Arjovsky et al., 2017).

3.3. Local Properties of OMASGAN Task 2

Proposition 2. Let $\tilde{\boldsymbol{\theta}}_{\boldsymbol{b}}$ be the locality where the algorithm converges and terminates. Then, wherever our algorithm terminates, that point leads the implicit distribution $B(\mathbf{z}; \tilde{\boldsymbol{\theta}}_{\boldsymbol{b}})$ to be a distribution on the boundary of the data distribution.

We find $\boldsymbol{\theta_b}$ in (2) such that the distribution on $\mathcal{X}$ is evenly distributed on the boundary surface, i.e. a manifold which is defined as the set of parameters $A^* \triangleq \arg\min\{-m(c) + \mu\, d(c) \,|\, c\}$, and a *necessary* condition for this is $\nabla(-m(c) + \mu\, d(c)) = 0$, which defines the tangent plane to the manifold. If the loss is locally convex, then (2) attains a local minimum at least. We find a boundary distribution and [illegible] that the distribution is evenly distributed on the support boundary of the data distribution. Solving $A^* = \arg\min\{[illegible](c) \,|\, c \in [illegible]\}$ is a hard constrained optimization problem. We introduce a regularized loss in the optimization instead of trying to solve a much harder constrained optimization. We regularize and solve $\arg\min -m(c) + \mu\, d(c) + \nu$ [illegible]. The Gradient Descent algorithm finds a local minimum of this combined loss. By choosing the [illegible] parameters μ and ν appropriately, we find points in A^* [illegible] the Gradient Descent which terminates [illegible] the gradient is zero and the neighborhood points do not reduce the loss (Zhu et al., 2020), the solution to the first two terms in (2) and (3), enhanced with the [illegible] term [illegible] lies in the OMAS manifold and, as discussed in Section [illegible], this is our second contribution. When our algorithm converges at $\tilde{\boldsymbol{\theta}}_{\boldsymbol{b}}$, dispersion is maximized and mode collapse has been obviated because of the scattering measure in (2) and its combination with the other terms.

3.4. Global and Local Properties of Task 3

Let p_b be the implicit boundary distribution from Task 2 and B. Task 3 of OMASGAN in (5) attains a global optimum when $p_{g'} = p_{\mathbf{x}}$ and using samples from both the positive and negative classes, G' learns the distribution of the positive class (Asokan & Seelamantula, 2020; Zaheer et al., 2020). The global optimum at $p_{g'} = (1+\gamma)p_{\mathbf{x}} - \gamma p_b$ in (5) subject to $\gamma \geq \alpha + \delta - 1$ and $\alpha + \delta \in [0,1]$ subsumes the global optimum at $p_{g'} = p_{\mathbf{x}}$ as a special case and using the alternating Gradient Descent algorithm, convergence to a local optimum is guaranteed (Asokan & Seelamantula, 2020).

Proposition 3. Let $L_J(\boldsymbol{\theta_j}, \mathbf{z}, \mathbf{x}, J, B, G')$ be the *loss* in (6) as a function of the learnable parameters, $\boldsymbol{\theta_j}$. The set over which $\boldsymbol{\theta_j}$ vary is compact. Then, L_J is a continuous function of its parameters and attains a global *minimum* at $\boldsymbol{\theta_j}^*$, that is, there exists a $\boldsymbol{\theta_j}^*$ such that $L_J(\boldsymbol{\theta_j}^*, \mathbf{z}, \mathbf{x}, J, B, G')$ is lowest.

Global properties. Using the Extreme Value Theorem, (6) is continuous, its domain is compact, and it attains a global maximum (*sufficient* condition). The set over which $\boldsymbol{\theta_j}$ vary is compact, the domain is compact, and our composite objective is continuous as a function of the model parameters.

Table 1. Properties and architecture characteristics of the OMASGAN model and of the recent state-of-the-art benchmarks for AD.

	OMASGAN	MinLGAN	FenceGAN	EGBAD	AnoGAN	TailGAN	BDSG	OGNet	DeepSAD	ConAD	VAE	ADAE	GANomaly	AED
GAN-based	√	√	√	√	√	√								
AE-based								√	√	√	√	√	√	√
Active Negative Sampling & Training	√	√												
Negative Training	√	√						√	√					
Boundary Loss	√		√			√	√							
f-Divergence	√													
Discriminator Anomaly Score	√	√	√	√	√			√		√		√		
Reconstruction				√	√		√		√		√	√	√	√
Likelihood Loss		√				√	√			√	√			

Proposition 4. Regarding the local properties of (6), let $\tilde{\boldsymbol{\theta}}_{\boldsymbol{j}}$ be the point where the algorithm terminates. Then, wherever the algorithm converges, that $\tilde{\boldsymbol{\theta}}_{\boldsymbol{j}}$ leads $J(\mathbf{x};\tilde{\boldsymbol{\theta}}_{\boldsymbol{j}})$ to be a classifier separating the data distribution from its complement.

The solution to (6) using Stochastic Gradient Descent is a classifier that separates the data distribution from its complement because $B(\mathbf{z})$ samples from the boundary of the data distribution. $B(\mathbf{z})$ is an implicit distribution on the boundary producing *OMAS points for negative training*, for J, G', and C. As presented in Section 1, this is our third contribution.

4. Related Work

Table 1 presents the characteristics of OMASGAN and the benchmarks, the architecture (GAN or Autoencoder (AE)), the losses (active negative sampling and training, boundary loss), and the inference mechanism (discriminator anomaly score). The models are sorted with respect to the *number of properties*. In contrast to the benchmarks, OMASGAN performs active negative training, introduces self-generated labels and supervision, performs boundary and f-divergence loss training, and utilizes a discriminator anomaly score. It is a GAN rather than an AE and this is because (a) a distribution metric is established in AD, (b) GANs and distribution metrics have been shown to outperform AEs and distance metrics, and (c) contextual anomalies that have shared features with the data cannot be detected using a reconstruction anomaly score (near low probability data).

OMASGAN addresses the *rarity* of abnormal data and provides data augmentation by creating strong abnormal data on the distribution boundary, unlike (Sung et al., 2020; Sipple, 2020). Our negative augmentation methodology performs sampling of negative points, creates optimal points for negative training *closest to the data*, and does not need to know any data features, in contrast to (Sinha et al, 2021). We perform retraining using active negative sampling setting the boundary points as *strong* anomalies. As discussed in Section 1, this is our contribution. It differs from creating OoD samples by using (i) low-epoch reconstructions (Zaheer et al., 2020; Pourreza et al., 2021), (ii) mutated features (Sinha et al, 2021), and (iii) a CVAE (Bian et al., 2019). Old is Gold (OGNet) uses weak anomalies far from the boundary, low-quality reconstructions, as pseudo-anomalies generated in an *ad hoc way* not covering the OoD part of the space (Zaheer et al., 2020). In the second training stage, OGNet uses the reconstruction loss due to its pseudo-anomaly module which generates OoD points but uses a *restrictive* definition of anomaly as single-epoch blurry reconstructions. It m... of the discriminator (f-divergences) to distinguish good from bad quality reconstructions. For good and poor quality samples, a generator and an old state of the same generator are used, respectively. Anomalies far from the boundary are also created by (Bian et al., 2019).

The *rarity* of anomalies is not addressed in (Ruff et al., 2020; Asokan & Seelamantula, 2020) which however highlight the benefit of supervision. Minimum Likelihood GAN (MinLGAN) and FenceGAN generate samples on the boundary of the data distribution to subsequently use the discriminator score for OoD detection (Wang et al., 2018; Ngo et al., 2019). In contrast to the Boundary of Distribution Support Generator (BDSG) and Tail of distribution GAN (TailGAN) (Dionelis et al., 2020), OMASGAN uses any f-divergence, no likelihood or invertibility, and a discriminator for AD.

5. Evaluation of OMASGAN

We evaluate our model using the AUROC. The LOO methodology is used which is setting K classes of a dataset with $(K+1)$ classes as the normal class and the *leave-out class* as the abnormal class. It is more challenging than One Class Classification (OCC) used by MinLGAN, OGNet, and (Kim et al., 2020; Nguyen, 2019) which is setting a dataset class as the normal class and the remaining as the abnormal class.

Models. We train fully-connected networks for synthetic data and Convolutional Neural Networks (CNN) with batch normalization for image data. We utilize CNNs for $B(\mathbf{z})$ to create OMAS samples and we use the recently developed f-divergence-based KL-Wasserstein GAN (KLWGAN) (Song & Ermon, 2020) and the f-GAN (Nowozin et al., 2016).

AD using the f-divergence distribution metric: The discriminator computes f-divergences and for the distributions P and R, we write this metric as fD(P, R). The f-divergence metric is used for training and we also use it during *inference*. To find the AUROC using the f-divergence metric, we compute fD($\mathbf{x}^*$, G') for a queried test sample $\mathbf{x}^*$ where G' is the learned data distribution from training. We calculate fD($\delta_{\mathbf{x}}^*$, G') where $\delta_{\mathbf{x}}^*$ is a Dirac function centered at $\mathbf{x}^*$.

Benchmarks, Data. We compare OMASGAN to the GAN models AnoGAN (Schlegl et al., 2017), EGBAD (Zenati et al., 2018), and FenceGAN (Ngo et al., 2019), and to the AE models GANomaly (Akcay et al., 2018), VAE, and ADAE (Vu et al., 2019) on MNIST and CIFAR-10. We compare with the likelihood-based AD models, BDSG and TailGAN (Dionelis et al., 2020), and with the Autoencoder-Discriminator (AED) which uses reconstruction and latent losses, an encoder-decoder-encoder generator, and adversarial training (Table 1), developed by us for benchmarking. Also, we evaluate OMASGAN using out-of-dataset anomalies, specifically Fashion-MNIST, KMNIST, and SVHN.

5.1. Evaluation of OMASGAN on Synthetic Data

We train our f-divergence-based OMASGAN on synthetic data in Figure 3. *Sensitivity analysis:* For bimodal distributions, we use $Q = 4096$, $N = 100$, $\mu = 8$, and $\nu = 20$. The batch size, N, and Q affect the convergence speed of OMASGAN. For the successfully-converging $B(\mathbf{z})$ and $G(\mathbf{z})$, we obtain descending loss values in (2). Figure 3 depicts the OMAS samples, the green $B(\mathbf{z})$ points. The blue points are from the data distribution. OMASGAN works for multimodal distributions with *disconnected components*, in contrast to AE, FenceGAN, and unsupervised models (Nguyen et al., 2019; Ngo et al., 2019; Ruff et al., 2020). We evaluate our OMASGAN using histograms of anomaly scores and with retraining, we increase the AUROC and the Area Under the Precision-Recall Curve (AUPRC) from 0.91 to 0.99 and the F1, Precision, Recall, and Accuracy from 0.83 to 0.98.

5.2. Evaluation of OMASGAN on MNIST Data

Setup: We evaluate the f-divergence-based OMASGAN on MNIST using the LOO methodology and $p_{\mathbf{z}} = \mathrm{N}_{128}(\mathbf{0}, \mathbf{1})$. We train OMASGAN using the KLWGAN until convergence, 300 epochs, and utilize a CNN to generate the $G(\mathbf{z})$ distribution and our $B(\mathbf{z})$ to generate boundary samples. According to our sensitivity analyses, we use CNN model architectures for $G(\mathbf{z})$ and $B(\mathbf{z})$, $Q = 1024$, $N = 256$, $\mu = 0.2$, and $\nu = 0.3$. We train OMASGAN and obtain decreasing

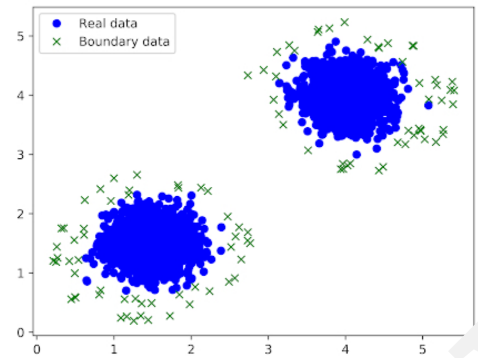


Figure 3. OMASGAN Task 2 for synthetic data. The blue points are real samples and the green points are $B(\mathbf{z})$ samples.

losses for the successfully-generating $G(\mathbf{z})$, and descending loss values for the successfully converging $B(\mathbf{z})$. We evaluate OMASGAN and our discriminator using histograms of the anomaly scores for normal and abnormal samples.

LOO Evaluation in AUROC: Figure 4 shows that on average and for all digits, OMASGAN outperforms the GAN benchmarks EGBAD, AnoGAN, BDSG, and TailGAN. Figure 5 shows that OMASGAN outperforms the AE benchmarks VAE and GANomaly in AUROC. We evaluate OMASGAN and compare it with GANomaly using the same inference conditions as those in OMASGAN, training statistics rather than test statistics for the batch normalization layers.

Figures 4 and 5 show that OMASGAN achieves on average an AUROC of 0.85 on MNIST data and outperforms the AD benchmarks by at least 0.24 points in AUROC, by a percentage of approximately 41%. It is robust and achieves the lowest standard deviation (SD) of 0.036 averaged over all digits, compared to EGBAD, AnoGAN, BDSG, TailGAN, GANomaly, and VAE for AD. These benchmarks have SDs 0.153, 0.093, 0.24, 0.059, 0.074, and 0.199 respectively.

The evaluation of OMASGAN trained on MNIST and tested on Fashion-MNIST and KMNIST, as in (Nalisnick et al., 2019), yields an AUROC of 0.83 and 0.71, respectively.

5.3. AUROC Evaluation of OMASGAN on CIFAR-10

Setup: We evaluate our model and according to sensitivity analyses, in (5) and (6), we use $\alpha + \delta = 0.7$, $\beta = 1$, $\gamma = 0.7$ in Section 3.4, and $\zeta = 0.5$ (Asokan & Seelamantula, 2020).

LOO Evaluation in AUROC: Figure 6 shows that the performance of the KLWGAN-based OMASGAN using LOO is better than that of the GAN models AnoGAN, EGBAD, FenceGAN, and BDSG on average and for all classes. In Figure 7, on average and for almost all classes, the proposed OMASGAN outperforms the AE benchmarks GANomaly, VAE, ADAE, and AED. According to Figures 6 and 7, OMASGAN outperforms the benchmarks in AUROC aver-

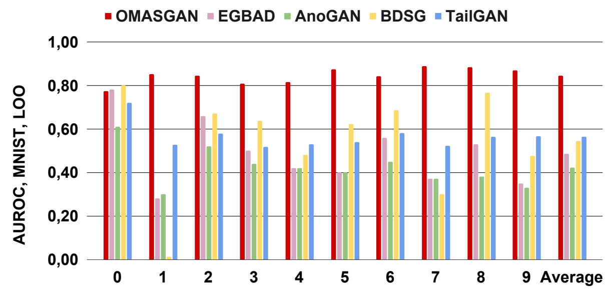


Figure 4. Performance of KLWGAN-based OMASGAN for AD on MNIST in AUROC using LOO compared to GAN benchmarks.

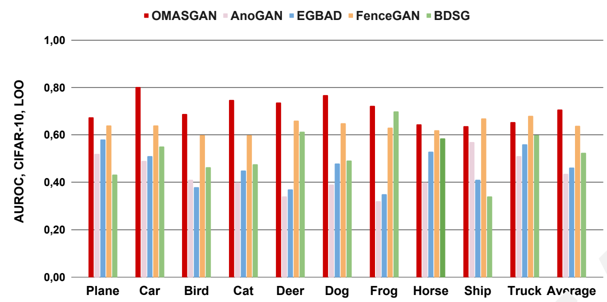


Figure 6. Performance of KLWGAN-based OMASG[illegible] ROC on CIFAR-10 using LOO compared to GAN [illegible]chmarks.

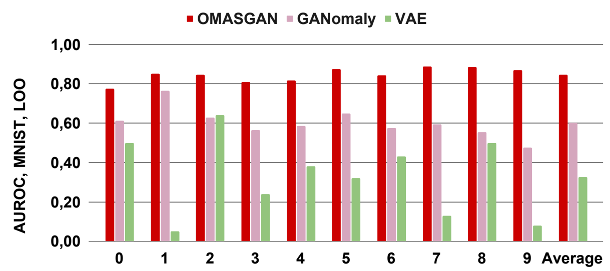
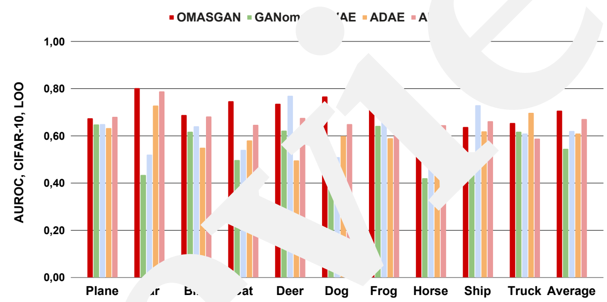


Figure 5. Performance of KLWGAN-based OMASGAN in AUROC on MNIST data using LOO compared to AE benchmarks.



[illegible] *7.* Per[illegible]nce of K[illegible]WGAN-based OMASGAN in AUROC [illegible] CIFA[illegible] [illegible]OO, compared to AE benchmarks.

aged over all classes. It is robust achieving the lowest SD, 0.056, compared to the AD benchmarks. It outperforms the benchmarks on average over all classes by at least 0.07 AUROC points, by a percentage increase of at lea[illegible]1%.

OMASGAN achieves on average an AUR[illegible] 0.71 on CIFAR-10 using LOO evaluation and *outperfo[illegible]* [illegible]onAD (Nguyen et al., 2019). Using OCC, C[illegible]D yield[illegible] AU[illegible]ROC improvement of up to 0.06 po[illegible]ts, i.[illegible] approxim[illegible]ly 9.6%, over previously [illegible]orted resu[illegible]: 0.[illegible] and 0.05 points improvement compar[illegible] [illegible]noGA[illegible] [illegible] VAE, r[illegible]pectively. It achieves an AUROC o[illegible] [illegible]7. OM[illegible] achieves on average AUROC values of 0.[illegible] using LOO evaluation and outperforms Deep Se[illegible]-supervi[illegible]AD (DeepSAD) (Ruff et al., 2020). Using O[illegible] DeepSA[illegible] yields a performance improvement e[illegible] whe[illegible] [illegible]ided with a few labeled data. It achieves an [illegible] impro[illegible]nt of up to 0.12 points over previous[illegible] reporte[illegible] [illegible]lts on CIFAR-10 using one known [illegible]bnorm[illegible]s, achiev[illegible] an AUROC of 0.73. OMASGAN, [illegible]ing LOO [illegible]ation on CIFAR-10, achieves an improveme[illegible] of appro[illegible]ely 7.5% compared to the benchmarks. Experi[illegible]ents pe[illegible]formed using OCC evaluation show that OGNet [illegible]tperforms state-of-the-art benchmarks by up to 3.6% im[illegible]ovement in AUROC, in (Zaheer et al., 2020).

[illegible]GAN trained on CIFAR-10 and tested on SVHN, as in (Kirichenko et al, 2020), achieves an AUROC of 0.76.

[illegible] Abla[illegible]dy/Analysis of OMASGAN

B[illegible] of retraining by including negative boundary sam[illegible] *OMASGAN trained on MNIST:* Figure 8 shows that, [illegible] average and for all the digits, OMASGAN improves the performance of the KLWGAN implemented in Task 1 for AD. Comparing the training loss in Task 1 to the loss in Task 3 and to the final loss, OMASGAN improves the performance of the base model. The base model KLWGAN achieves an AUROC of 0.59 averaged over all digits, increasing to 0.71 using Task 3, and then to 0.84 using our final model, and this is the contribution of Tasks 2 and 3.

Effect of base model and chosen f-divergence. In Figure 9, we compare the OMASGAN model using the KL-WGAN to the OMASGAN using the f-GAN in AUROC (Song & Ermon, 2020; Nowozin et al., 2016). The ablation study shows that OMASGAN boosts the performance of f-GAN by 0.26 in AUROC on average over all digits. The base model f-GAN achieves on average an AUROC of 0.51, which increases to 0.77 because of OMASGAN Task 3. Figures 8 and 9 show the benefit of the OMASGAN properties of *active negative sampling and training*, negative samples augmentation methodology, boundary loss training, and discriminator anomaly score, as presented in Table 1.

Improvement of OMASGAN compared to KLWGAN for AD on CIFAR-10. Figure 10 shows the ablation study

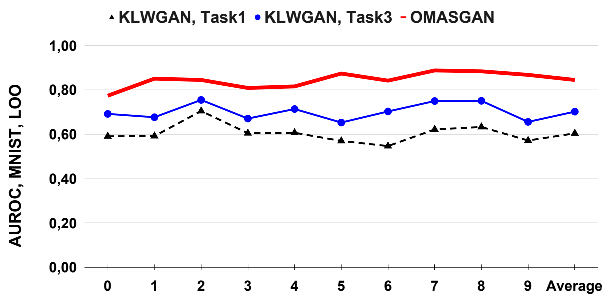


Figure 8. Ablation study of KLWGAN-based OMASGAN in AUROC on MNIST: Impact of the losses on the AD performance.

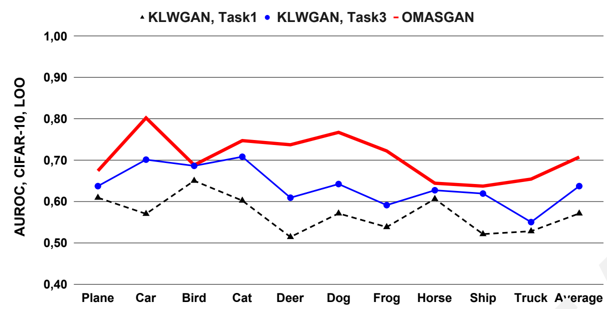


Figure 10. Ablation study of the KLWGAN-based OMASGAN in AUROC using LOO on CIFAR-10: Benefit of our loss functions.

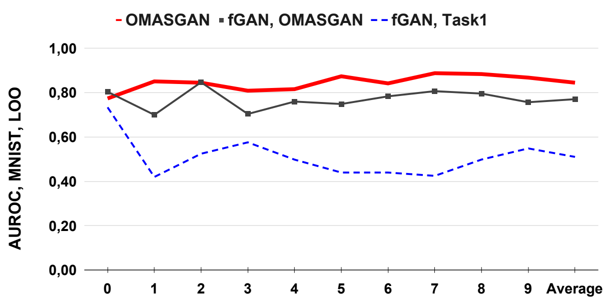


Figure 9. Comparison of KLWGAN-based OMASGAN to the f-GAN-based OMASGAN using the LOO evaluation on MNIST.

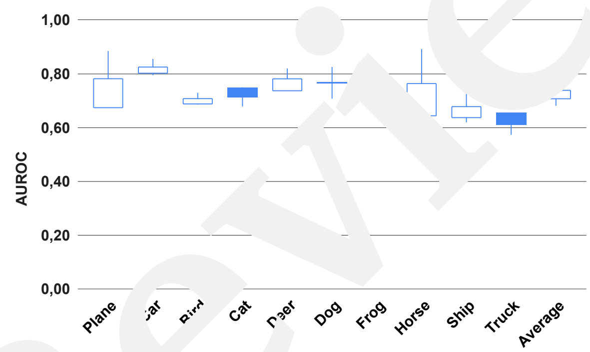


Figure 11. Performance of OMASGAN on CIFAR-10 for seeds 0, 1, and 2. The maximum value is the average AUROC over seeds, while the minimum value is the AUROC for seed 2 (opposite for blue box). Whiskers show the mean AUROC ± SD.

of OMASGAN in AUROC using LOO. It shows the impact of the losses on the performance. Our chosen base model, KLWGAN, yields an AUROC of 0.57 averaged over all classes and this increases to 0.64 using OMASGAN Task 3 and to 0.71 using the final OMASGAN. In reference to Table 1, Figure 10 shows that the 0.14 AUROC improvement is the contribution of our negative data augmentation and *retraining* methodology. This is the benefit of negative and boundary loss training. The mean SD over all classes is 0.05, 0.05, and 0.06 for Task1, Task3, and OMASGAN.

Effect of selected inference mechanism. Following our discussion in Section 5, the anomaly score for a test sample $\mathbf{x}^*$ is fD($\delta_{\mathbf{x}}^*$, G') and fD($\delta_{\mathbf{x}}^*$, G) when the OMASGAN model is stopped at Task 3 and Task 1, respectively, in the ablation study in Figures 8-10. Comparing OMASGAN to Tasks 1 and 3, the performance gradually improves and the use of the discriminator J, presented in Section 2, is beneficial.

Sensitivity Analysis to the Random Seed

Effect of initialization. Figure 11 shows the sensitivity of the KLWGAN-based OMASGAN on CIFAR-10 to changes to the seeds 0, 1, and 2. It shows the mean AUROC over all seeds per LOO class, the AUROC for seed 2, and the mean AUROC ± SD. The performance of OMASGAN in AUROC yields a difference of 0.05 between seeds and the average SD over all CIFAR-10 classes and seeds is 0.06. This SD is *lower* than that of (Ruff et al., 2020), which is 0.1. On average, we set the seed to 2 and obtain more robust AUROC values with lower SD compared to seeds 0 and 1.

6. Conclusion

We have proposed OMASGAN, a retraining methodology for AD with negative sampling. *Without likelihood*, we generate OMAS samples and strong anomalies leveraging any f-divergence, the KLWGAN divergence and the f-divergence (Song & Ermon, 2020; Nowozin et al., 2016). We address the *rarity of anomalies* problem and use data only from the normal class. The evaluation outcomes on MNIST and CIFAR-10, as well as on synthetic data, using the LOO methodology show that OMASGAN achieves state-of-the-art performance and outperforms the benchmarks. Using AUROC, OMASGAN yields on average (a) an improvement of at least 0.24 points on MNIST over the benchmarks, achieving values of 0.85, (b) an improvement of at least 0.07 points on CIFAR-10 data, achieving values of 0.71, and (c) high AUROC values for out-of-dataset anomalies.

References

Abati, D., Porrello, A., Calderara, S., and Cucchiara, R., "Latent Space Autoregression for Novelty Detection," in Proceedings IEEE/CVF International Conference on Computer Vision and Pattern Recognition (CVPR), pp. 481-490, Long Beach, California, USA, June 2019.
DOI: 10.1109/CVPR.2019.00057

Abbasnejad, M., Shi, J., van den Hengel, A., and Liu, L., "A Generative Adversarial Density Estimator," in Proceedings IEEE/CVF International Conference on Computer Vision and Pattern Recognition (CVPR), pp. 10774-10783, Long Beach, CA, USA, June 2019.
DOI: 10.1109/CVPR.2019.01104

Abbasnejad, M., Shi, J., van den Hengel, A., and Liu, L., "GADE: A Generative Adversarial Approach to Density Estimation and its Applications," International Journal of Computer Vision, pp. 2731-2743, vol. 128, no. 10, August 2020. DOI: 10.1007/s11263-020-01360-9

Akcay, S., Atapour-Abarghouei, A., and Breckon, T., "GANomaly: Semi-supervised Anomaly Detection via Adversarial Training", in Proceedings 14th Asian Conference on Computer Vision (ACCV), pp. 622-637, vol. 11363, Lecture Notes in Computer Science, Springer, Cham, Jawahar, C., Li, H., Mori, G., and Schindler, K. (eds), May 2019. DOI: 10.1007/978-3-030-20893-6_39

Akcay, S., Atapour-Abarghouei, A., and Breckon, T., "Skip-GANomaly: Skip Connected and Adversarially Trained Encoder-Decoder Anomaly Detection", in Proceedings IEEE International Joint Conference on Neural Networks (IJCNN), pp. 1-8, Budapest, Hungary, July 2019.
DOI: 10.1109/IJCNN.2019.8851808

Arjovsky, M. and Bottou, L., "Towards Principled Methods for Training Generative Adversarial Networks," in Proceedings 5th International Conference on Learning Representations (ICLR), Palais des Congrès Neptune, Toulon, France, April 2017.

Arjovsky, M., Chintala, S., and Bottou, L., "Wasserstein Generative Adversarial Networks," in Proceedings of the 34th International Conference on Machine Learning (ICML), pp. 214-223, vol. 70, Precup, D. and Teh, Y. (eds), Sydney, Australia, August 2017.

Asokan, S. and Seelamantula, C., "Teaching a GAN What Not to Learn," in Proceedings 34th Conference on Neural Information Processing Systems (NeurIPS 2020), Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., and Lin, H. (eds), Vancouver, Canada, December 2020.
Online: https://github.com/DarthSid95/RumiGANs
Online: https://papers.nips.cc/paper/2020/hash/29405e2a4c[illegible]05f557559c7fa4b-Abstract.html

Author, "OoD Minimum Anomaly Score GAN - Code for the Paper OMASGAN: Out-of-Distribution Minimum Anomaly Score GAN for Sample Generation on the Boundary," GitHub Repository, January 2021.
https://github.com/Anonymous-Author-2021/OMASGAN

Bian, J., Hui, X., Sun, S., Zhao, X., and Tan, M., "A Novel and Efficient CVAE-GAN-Based Approach With Informative Manifold for Semi-Supervised Anomaly Detection," in IEEE Access, vol. 7, pp. 88903-88916, June 2019. DOI: 10.1109/ACCESS.2019.2920251

Brock, A., Donahue, J., and Simonyan, K., "Large Scale GAN Training for High Fidelity Natural Image Synthesis," in Proceedings Seventh International Conference on Learning Representations (ICLR), New Orleans, Louisiana, USA, May 2019.

Clanuwat, T., Bober-Irizar, M., Kitamoto, A., Lamb, A., Yamamoto, K., and Ha, D., "Deep Learning for Classical Japanese Literature," arXiv preprint, arXiv:1812.01718 [cs.CV], December 2018. Online: https://github.com/rois-codh/kmnist

D'Amour, A., Heller, K., Moldovan, D., Adlam, B., Alipanahi, B., Beutel, A., Chen, C., Deaton, J., Eisenstein, J., Hoffman, M. et al., "Underspecification presents challenges for credibility in modern machine learning," arXiv preprint, arXiv:2011.03395 [cs.LG], November 2020.

Dionelis, N., Yaghoobi, M., and Tsaftaris, S., "Boundary of Distribution Support Generator (BDSG): Sample Generation on the Boundary," in Proceedings IEEE International Conference on Image Processing (ICIP), pp. 803-807, October 2020. DOI: 10.1109/ICIP40778.2020.9191341

Dionelis, N., Yaghoobi, M., and Tsaftaris, S., "Tail of Distribution GAN (TailGAN): Generative-Adversarial-Network-Based Boundary Formation," in Proceedings Sensor Signal Processing for Defence (SSPD) IEEE Conference, pp. 41-45, November 2020.
DOI: 10.1109/SSPD47486.2020.9272100

Gong, D., Liu, L., Le, V., Saha, B., Mansour, M., Venkatesh, S., and van den Hengel, A., "Memorizing Normality to Detect Anomaly: Memory-Augmented Deep Autoencoder for Unsupervised Anomaly Detection," in Proceedings IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South), pp. 1705-1714, November 2019. DOI: 10.1109/ICCV.2019.00179

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y., "Generative Adversarial Nets," in Proceedings Advances in Neural Information Processing Systems 27 (NIPS 2014), pp. 2672–2680, Montréal, Canada, December 2014.

Hendrycks, D., Mazeika, M., and Dietterich, T., "Deep

Anomaly Detection with Outlier Exposure," in Proceedings Seventh International Conference on Learning Representations (ICLR 2019), New Orleans, Louisiana, USA, May 2019.

Hjelm, R., Jacob, A., Che, T., Trischler, A., Cho, K., and Bengio, Y., "Boundary-Seeking Generative Adversarial Networks," in Proceedings Sixth International Conference on Learning Representations (ICLR), Vancouver, Canada, May 2018.

Jalal, A., Ilyas, A., Dimakis, A., and Daskalakis, C., "The Robust Defense: Adversarial Training using Generative Models," arXiv preprint, arXiv:1712.09196 [cs.CV], July 2019.

Kim, Y., Kwon, Y., Chang, H., and Paik, M., "Lipschitz Continuous Autoencoders in Application to Anomaly Detection," in Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics (AISTATS), pp. 2507-2517, vol. 108, Chiappa, S. and Calandra, R. (eds), Palermo, Italy, August 2020.

Kirichenko, P., Izmailov, P., and Wilson, A., "Why Normalizing Flows Fail to Detect Out-of-Distribution Data," in Proceedings 34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada, December 2020.

Krizhevsky, A., "Learning Multiple Layers of Features from Tiny Images," Technical Report, TR-2009, University of Toronto, Toronto, April 2009.
Online: https://www.cs.toronto.edu/ kriz/cifar.html

LeCun, Y. and Cortes, C., "MNIST Handwritten Digit Database," 2010.
Online: http://yann.lecun.com/exdb/mnist/

Lee, K., Lee, H., Lee, K., and Shin, J., "Training Confidence-Calibrated Classifiers for Detecting Out-of-Distribution Samples," in Proceedings 6th International Conference on Learning Representations (ICLR), Vancouver, Canada, May 2018.

Maaløe, L., Fraccaro, M., Liévin, V., and Winther, O., "BIVA: A Very Deep Hierarchy of Latent Variables for Generative Modeling," in Proceedings 33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada, December 2019.

Mao, X., Li, Q., Xie, H., Lau, R., Wang, Z., and Smolley, S., "Least Squares Generative Adversarial Networks," in Proceedings IEEE International Conference on Computer Vision (ICCV), pp. 2813-2821, Venice, Italy, October 2017. DOI: 10.1109/ICCV.2017.304

Mroueh, Y., Li, C., Sercu, T., Raj, A., and Cheng, Y., "Sobolev GAN," in Proceedings International Conference on Learning Representations (ICLR), Vancouver, Canada, May 2018.

Nalisnick, E., Matsukawa, A., Teh, Y., Gorur, D., and Lakshminarayanan, B., "Do Deep Generative Models Know What They Don't Know?," in Proceedings Seventh International Conference on Learning Representations (ICLR), New Orleans, USA, May 2019.

Netzer, Y., Wang, T., Coates, A., Bissacco, A., Wu, B., and Ng, A., "Reading Digits in Natural Images with Unsupervised Feature Learning," NIPS Workshop on Deep Learning and Unsupervised Feature Learning, 9 pages, December 2011. Online: http://ufldl.stanford.edu/housenumbers/

Ngo, P., Winarto, A., Li Kou, C., Park, S., Akram, F., and Lee, H., "Fence GAN: Towards Better Anomaly Detection", in Proceedings IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI), pp. 141-148, Portland, Oregon, USA, November 2019. DOI: 10.1109/ICTAI.2019.00028

Nguyen, D., Lou, Z., Klar, M., and Brox, T., "Anomaly Detection with Multiple-Hypotheses Predictions," in Proceedings International Conference on Machine Learning (ICML), pp. 4800-4809, vol. 97, Chaudhuri, K. and Salakhutdinov, R. (eds), Long Beach, California, USA, June 2019.

Nielsen, F., and Nock, R., "On the Chi Square and Higher-Order Chi Distances for Approximating f-Divergences," IEEE Signal Processing Letters, pp. 10-13, vol. 21, no. 1, January 2014. DOI: 10.1109/LSP.2013.2288355

Nowozin, S., Cseke, B., and Tomioka, R., "f-GAN: Training Generative Neural Samplers using Variational Divergence Minimization," in Proceedings Thirtieth Conference on Neural Information Processing Systems (NIPS), Barcelona, Spain, December 2016.

Pourreza, M., Mohammadi, B., Khaki, M., Bouindour, S., Snoussi, H., and Sabokrou, M., "G2D: Generate to Detect Anomaly," in Proceedings IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), pp. 2003-2012, Virtual Conference, January 2021.

Ruff, L., Vandermeulen, R., Görnitz, N., Binder, A., Müller, E., Müller, K., and Kloft, M., "Deep One-Class Classification," in Proceedings International Conference on Machine Learning (ICML), pp. 4393-4402, vol. 80, Dy, J. and Krause, A. (eds), Stockholm, Sweden, July 2018.

Ruff, L., Vandermeulen, R., Görnitz, N., Binder, A., Müller, E., Müller, K., and Kloft, M., "Deep Semi-Supervised Anomaly Detection," in Proceedings International Conference on Learning Representations (ICLR), Virtual Conference, Formerly Addis Ababa, Ethiopia, May 2020.

Sabokrou, M., Khalooei, M., Fathy, M., and Adeli, E., "Adversarially Learned One-Class Classifier for Novelty Detection," in Proceedings IEEE/CVF Conference Computer Vision and Pattern Recognition (CVPR), pp. 3379-3388, Salt Lake City, UT, USA, June 2018. DOI: 10.1109/CVPR.2018.00356

Sainburg, T., Migliori, B., Thielk, M., Theilman, B., and Gentner, T., "Generative Adversarial Interpolative Autoencoding," arXiv preprint, arXiv:1807.06650 [cs.LG], April 2019.

Schlegl, T., Seeböck, P., Waldstein, S., Schmidt-Erfurth, U., and Langs, G., "Unsupervised Anomaly Detection with GANs to Guide Marker Discovery," in Proceedings International Conference on Information Processing in Medical Imaging (IPMI), Appalachian State University, Boone, North Carolina, USA, June 2017.

Sinha, A., Ayush, K., Song, J., Uzkent, B., Jin, H., and Ermon, S., "Negative Data Augmentation," in Proceedings International Conference on Learning Representations (ICLR), May 2021.
Online: https://openreview.net/forum?id=Ovp8dvB8IBH
Online: https://openreview.net/pdf?id=Ovp8dvB8IBH
Online: https://anonymous.4open.science/r/99219ca9-ff6a-49e5-a525-c954080de8a7/

Sipple, J., "Interpretable, Multidimensional, Multimodal Anomaly Detection with Negative Sampling for Detection of Device Failure," in Proceedings Thirty-seventh International Conference on Machine Learning (ICML), pp. 9016-9025, vol. 119, Daumé III, H. and Singh, A. (eds), July 2020.
Online: https://icml.cc/media/Slides/icml/2020/virtual(no-parent)-14-21-00UTC-6171-interpretable_.pdf
http://proceedings.mlr.press/v119/sipple20a/sipple20a.pdf

Song, J. and Ermon, S., "Bridging the Gap Between f-GANs and Wasserstein GANs," in Proceedings International Conference on Machine Learning (ICML), pp. 9078-9087, vol. 119, Daumé III, H. and Singh, A. (eds), July 2020.

Sung, Y., Pei, S., Hsieh, S., and Lu, C., "Difference-Seeking GAN - Unseen Sample Generation," in Proceedings Eighth International Conference on Learning Representations (ICLR), Virtual Conference, Formerly Addis Ababa, Ethiopia, May 2020.

Sutherland, D., "Scalable, Flexible, and Active Learning on Distributions," Carnegie Mellon University, School of Computer Science, PhD Thesis, Pittsburgh, Pennsylvania, USA, September 2016.
http://reports-archive.adm.cs.cmu.edu/anon/2016/CMU-CS-16-128.pdf

Vu, H., Ueta, D., Hashimoto, K., Maeno, K., Pranata, S., and Shen, S., "Anomaly Detection with Adversarial Dual Autoencoders," arXiv preprint, arXiv:1902.06924 [cs.CV], February 2019.

Wang, C., Zhang, Y., and Liu, C., "Anomaly Detection via Minimum Likelihood Generative Adversarial Networks," in Proceedings 24th International Conference on Pattern Recognition (ICPR), pp. 1121-1126, Beijing, China, August 2018. DOI: 10.1109/ICPR.2018.8545381

Xiao, H., Rasul, K., and Vollgraf, R., "Fashion-MNIST: A Novel Image Dataset for Benchmarking Machine Learning Algorithms," arXiv preprint, arXiv:1708.07747 [cs.LG], September 2017. Online: https://www.kaggle.com/zalando-research/fashionmnist

Zaheer, M., Lee, J., Astrid, M., and Lee, S., "Old is Gold: Redefining the Adversarially Learned One-Class Classifier Training Paradigm," in Proceedings IEEE/CVF Conference Computer Vision and Pattern Recognition (CVPR 2020), pp. 14171-14181, Seattle, Washington, USA, June 2020.
DOI: 10.1109/CVPR42600.2020.01419

Zenati, H., Foo, C., Lecouat, B., Manek, G., and Chandrasekhar, V., "Efficient GAN-Based Anomaly Detection," Workshop Track in Sixth International Conference on Learning Representations (ICLR 2018), Vancouver, Canada, May 2018.

Zhu, Z., Soudry, D., Eldar, Y., and Wakin, M., "The Global Optimization Geometry of Shallow Linear Neural Networks," Journal of Mathematical Imaging and Vision, pp. 279–292, vol. 62, April 2020.
DOI: 10.1007/s10851-019-00889-w

Ma, Z., Soudry, D., Eldar, Y., and Wakin, M., "Implicit Regularization in Nonconvex Statistical Estimation: Gradient Descent Converges Linearly for Phase Retrieval, Matrix Completion, and Blind Deconvolution," Foundations of Computational Mathematics (FOCM 2020), pp. 451-632, vol. 20, no. 3, June 2020.
DOI: 10.1007/s10208-019-09429-9

Zisselman, E. and Tamar, A., "Deep Residual Flow for Out of Distribution Detection," in Proceedings IEEE/CVF Conference Computer Vision and Pattern Recognition (CVPR), pp. 13991-14000, June 2020.
DOI: 10.1109/CVPR42600.2020.01401

OMASGAN: Out-of-Distribution Minimum Anomaly Score GAN for Sample Generation on the Boundary

Supplementary Material

The discussions, explanations, experiments, and evaluations in this Supplementary Material are a continuation of the paper "OMASGAN: Out-of-Distribution Minimum Anomaly Score GAN for Sample Generation on the Boundary".

A. OMASGAN and Illustration of our Algorithm

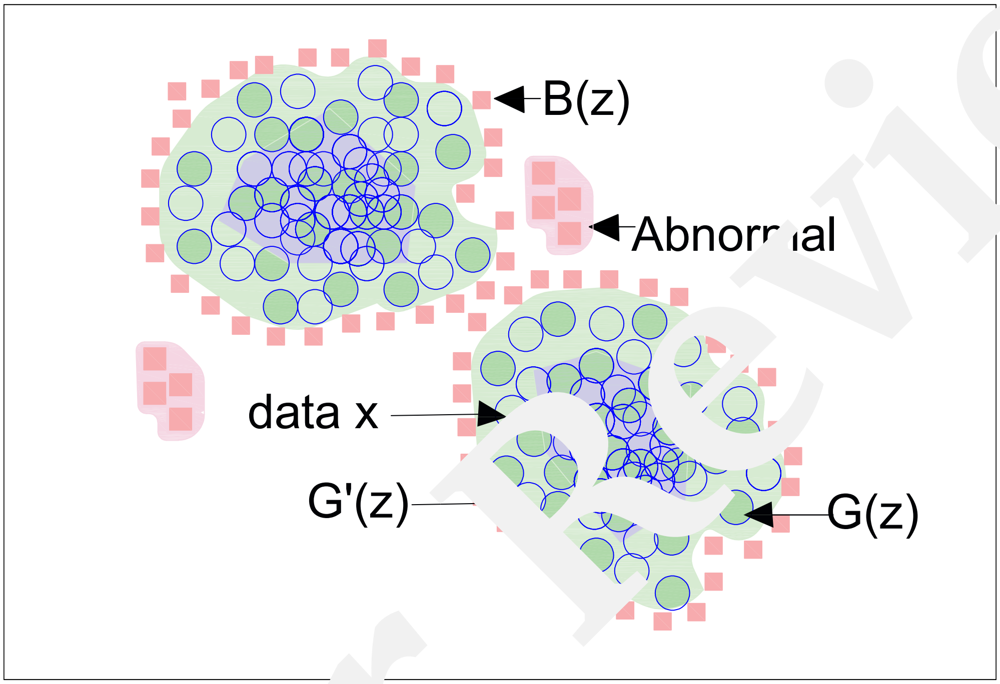


Figure 12. Illustration and pictorial representation of the proposed OMASGAN algorithm for AD.

In this section, we present an illustration of the proposed OMASGAN algorithm using the notation and the implicit generative models introduced in Section 3 of the paper. This section is a continuation of Sections 2 and 3 on pages 1-5 of the paper. Figure 12 depicts the main idea of the proposed OMASGAN algorithm. Following our discussion in Sections 1-3 of the paper, OMASGAN first generates minimum-anomaly-score OoD samples created by our negative sampling augmentation methodology and then performs active negative training for AD. OMASGAN performs model retraining and subsequently trains a discriminator for detecting the generated samples on the boundary of the support of the data distribution.

In Figure 12, the data are denoted by $\mathbf{x}$, the GAN-generated samples by $G(\mathbf{z})$, the OMAS samples by $B(\mathbf{z})$, and the active-negative-training-based GAN samples after model retraining by $G'(\mathbf{z})$. The samples $\mathbf{x}$, $G(\mathbf{z})$, $B(\mathbf{z})$, and $G'(\mathbf{z})$ all lie in the data space $\mathscr{X} \in \mathbb{R}^k$, and we denote the latent space by $\mathscr{Z} \in \mathbb{R}^l$, where $l << k$. Both Figure 1 in this section and Figure 3 of the paper illustrate the minimum-anomaly-score OoD $B(\mathbf{z})$ samples generated by the proposed OMASGAN model.

B. Implementation of OMASGAN

The proposed OMASGAN model is implemented in PyTorch, https://github.com/Anonymous-Author-2021/OMASGAN. The Code Repository can be found in https://anonymous.4open.science/r/7ca4da73-0380-4a24-a640-4c60d29db9f4/.

C. Evaluation of OMASGAN Trained on MNIST Data

The tables and figures in this section refer to Section 5 of the paper.

Table 2. Evaluation of the KLWGAN-based OMASGAN for AD using the AUROC metric, using abnormal out-of-dataset anomalies, where the normality is MNIST digits 0-9 and the anomalous cases are from the Fashion-MNIST and KMNIST datasets.

MNIST	Fashion-MNIST	KMNIST
AUROC	0.84	0.71

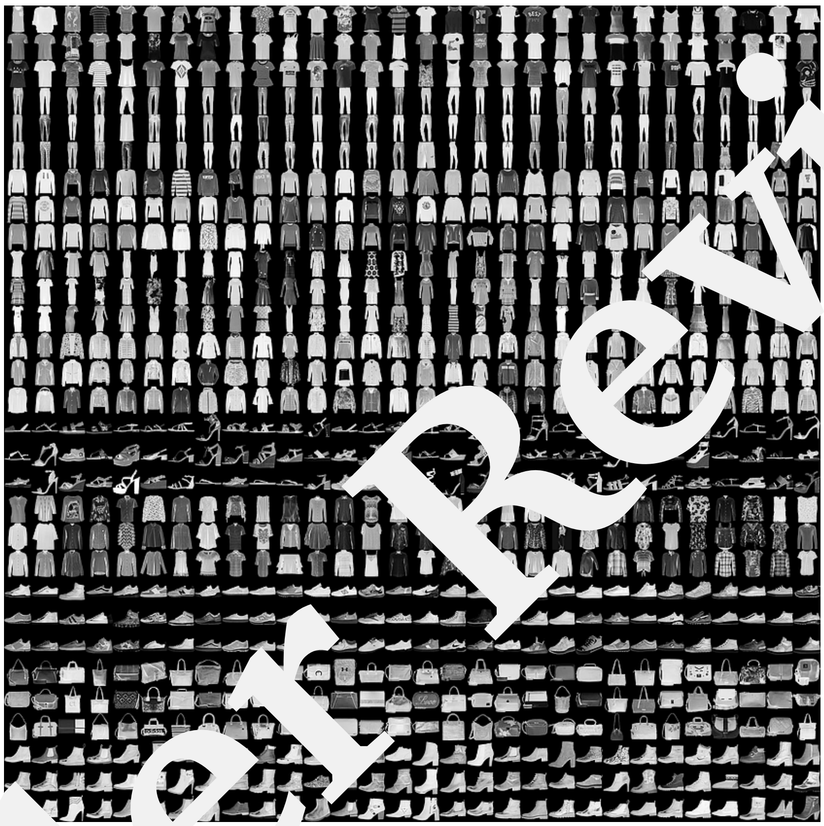


Figure 13. Samples from the Fashion-MNIST dataset (Xiao et al., 2017).

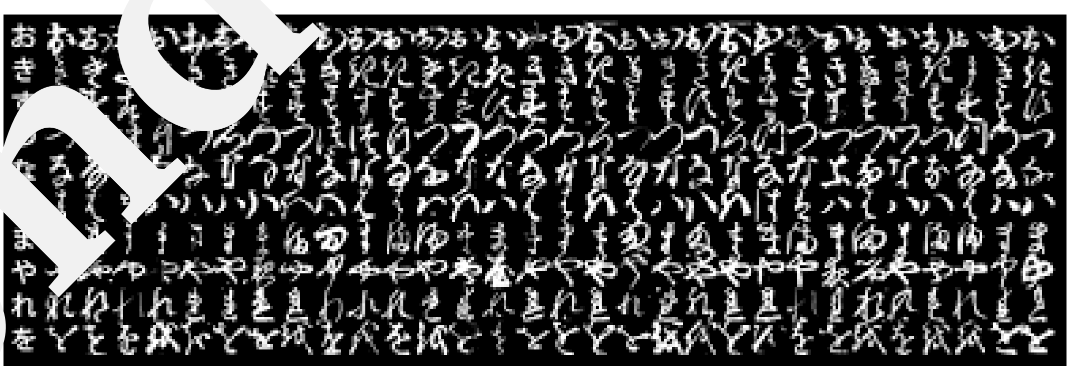


Figure 14. Samples from the KMNIST dataset (Clanuwat et al, 2018).

Figure 15. Samples from the MNIST image dataset (LeCun & Cortes, 2010).

D. Images from OMASGAN Task 1

Figure 16. Images from the normal class generated by the proposed f-GAN-based OMASGAN model from Task 1, trained on MNIST data using the LOO evaluation methodology.

E. Evaluation of OMASGAN Trained on CIFAR-10 Data

Table 3. Evaluation of the KLWGAN-based OMASGAN for AD using the AUROC metric, using OoD abnormal out-of-dataset anomalies, where the normal cases are from the CIFAR-10 dataset (classes 0-9) and the abnormal/anomalous cases are from the SVHN dataset.

CIFAR-10	SVHN
AUROC	0.76

Figure 17. Samples from the SVHN dataset (Netzer et al., 2011).

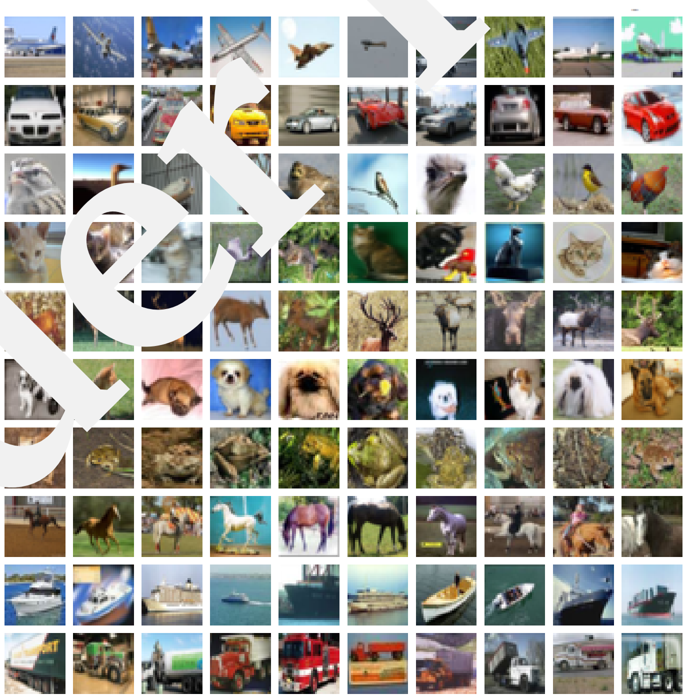


Figure 18. Samples from the CIFAR-10 dataset (Krizhevsky, 2009).

F. Images from OMASGAN Tasks 1 and 3

Figure 19. Images from the normal class generated by the proposed KLWGAN-based OMASGAN model from Task 1 (upper) and Task 3 (lower) trained on CIFAR-10 image data using the LOO evaluation methodology.