OMASGAN: Out-of-distribution Minimum Anomaly Score GAN for Anomaly Detection

Nikolaos Dionelis, Sotirios A. Tsaftaris, Mehrdad Yaghoobi

SSPD 2022

University of Edinburgh, UK
School of Engineering
Electronics and Electrical Engineering
Institute for Digital Communications (IDCOM)

University Defence Research Collaboration (UDRC) in Signal Processing
Track 3 Deep Neural Networks and Machine Learning Methods
WP3.1 Robust Generative Neural Networks

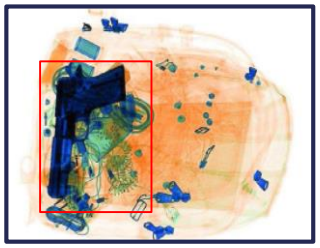
https://udrc.eng.ed.ac.uk/track-3-deep-neural-networks-and-machine-learning-methods

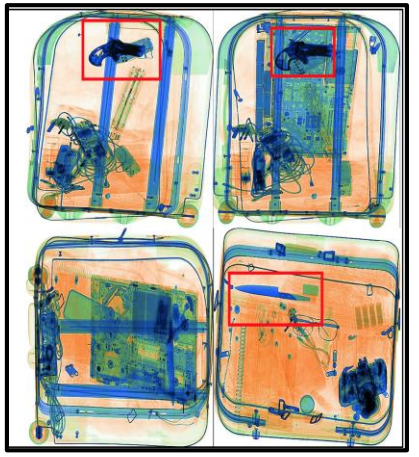
Contents

Introduction

Anomaly Detection/ OoD Detection in Images

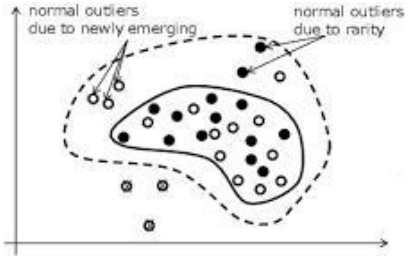
- **Anomaly detection**
  - **Example:** Pedestrians; Van
- **Out-of-Distribution** (**OoD**) detection
- **Identification** of samples that are different from **normal data**





- **Aim:** Provide decision mechanisms to **decide normal vs abnormal**
- **Anomalies** and **abnormal** data:
  - Are **rare**
  - **Might not be known** during training

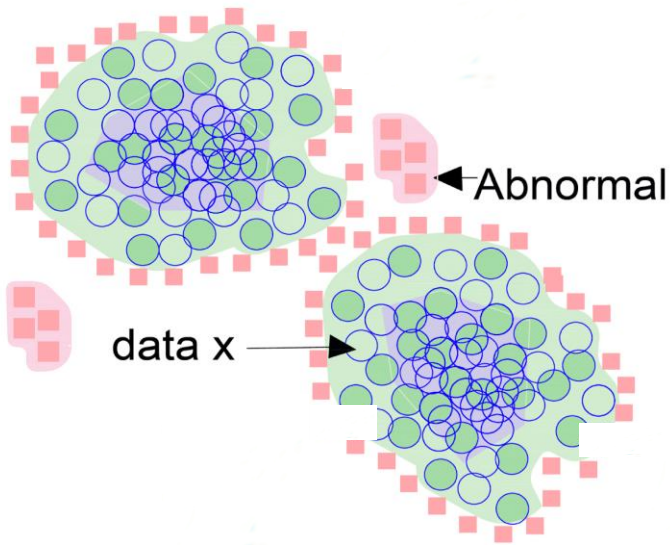
- **Application areas: Important critical systems**
  - **Security; Safety**; Autonomous systems
  - **Defence**; Medical imaging; Healthcare



OMASGAN - Main Contributions

- **Problem:**
  - Models may set **high likelihood** and **low reconstruction loss** to OoD samples
    - Leads to **failures to detect anomalies**
- **Aim:** **Accurate** and **robust** anomaly detection

- **Approach:**
  - Generative Adversarial Networks (**GAN**); **Generate** abnormal samples
  - Training with positive and **negative** data; Address: Learning-**OoD-samples** problem
- **Main contributions:**
  - Improved **OoD detection** performance
  - **Generation** of the **boundary** of the support of the **normal class distribution**:
    - **OoD minimum-anomaly score** samples
      - Include in **training**
  - Devise a **discriminator** for anomaly detection

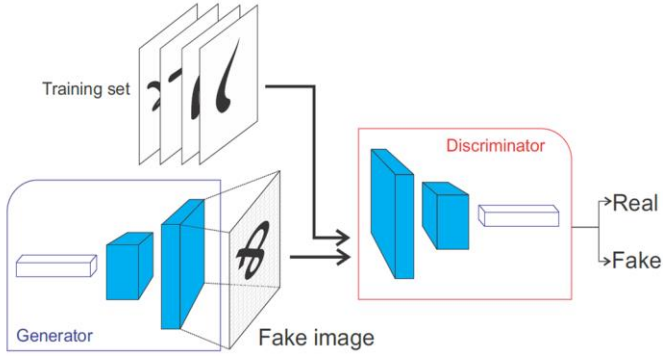


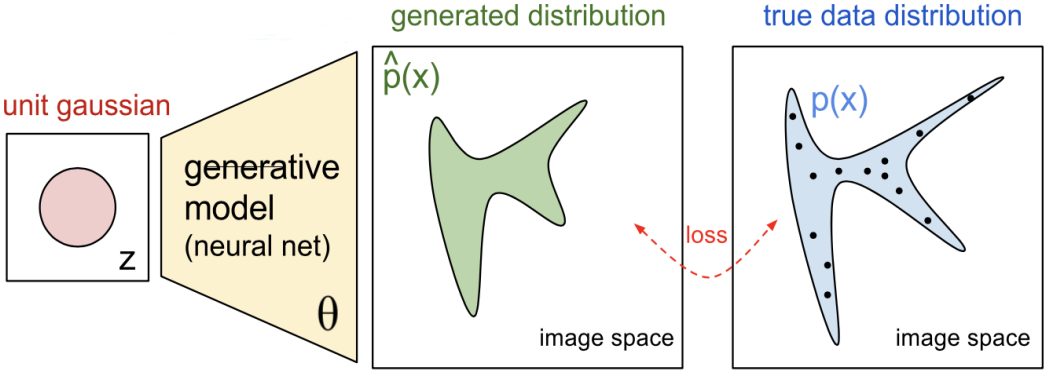
Related Work

Related Work and Main Challenges

- **Generative models:** Learn the underlying **distribution** of the normal class data
- **Problem: Models know what they do not know**
  - Set **high likelihood** and low reconstruction loss to **OoD data**
  - **This decreases** the anomaly detection performance

- Need to **improve the anomaly detection performance**
  - Eliminate **false negative errors**
    - **Failures to detect anomalies**
- Reduce **false positives**
  - **False alarms** of anomalies



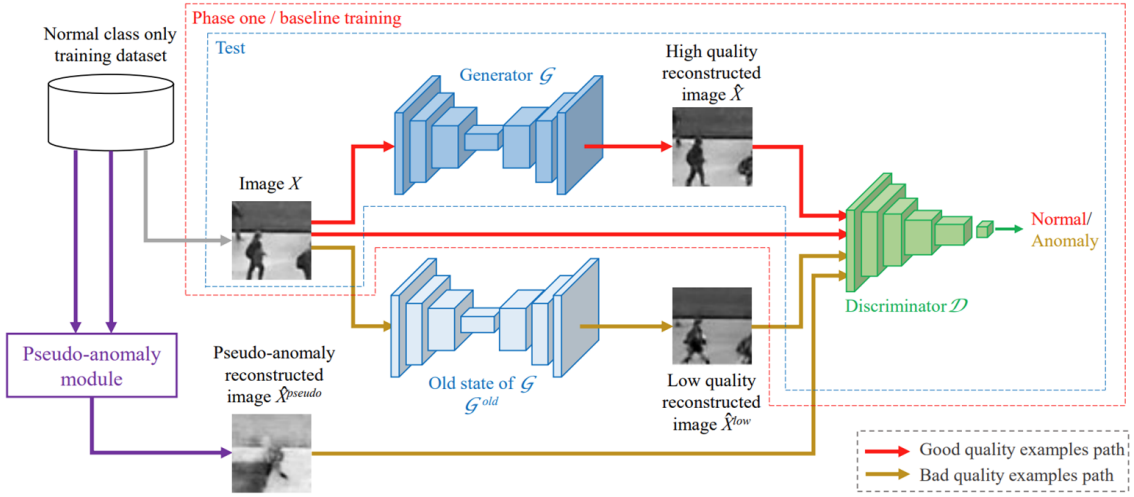


https://openai.com/blog/generative-models/

Related Research on Active Negative Training

- **Old is Gold (OGNet)** [1]: Uses a pseudo-anomaly module to produce **OoD samples**
  - **Restrictive definition of anomaly**
  - **Blurry reconstructions before convergence**
    - **Single-epoch** reconstructed images
- Active negative training but **not active negative sampling**
  - Does not cover the **OoD part** of the data space
- **Discriminator for AD:** Separates good from **bad quality** reconstructions

- **Deterministic: No sampling** is performed
- **AE**: Latent space is **learned** by the model

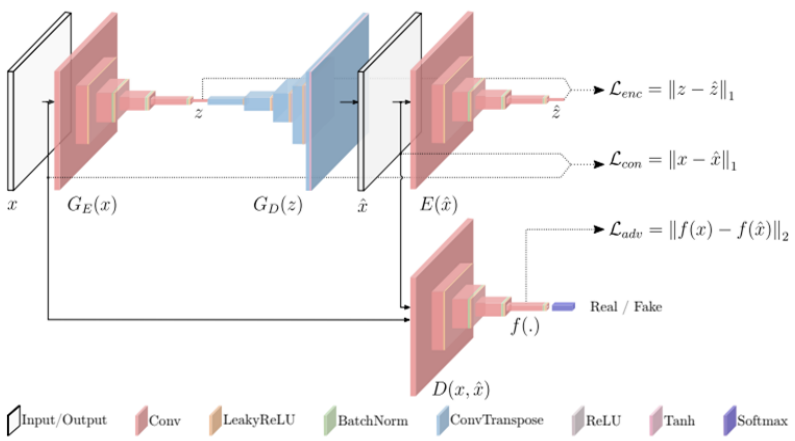


[1] M. Zaheer, et al. Old is Gold: Redefining the Adversarially Learned One-Class Classifier Training Paradigm. CVPR 2020

Related Work on Anomaly Detection

- **Anomaly detection using the reconstruction error**
  - **Unsupervised training** for OoD detection: **No class labels**
  - GAN for Anomaly detection (**GANomaly**) [2]: **Encoder-Decoder-Encoder**

- **Fence GAN** [3], **Minimum Likelihood GAN** [4]:
  - **Inference:** Discriminator
- Boundary of Data Distribution Support Generator (**BDSG**) [5]:
  - Flow-based invertible **generative models**
    - **Invertible** Residual Networks



[2] S. Akcay, et al., "GANomaly: Semi-Supervised Anomaly Detection via Adversarial Training," in Proc. ACCV, 2018

[3] P. Cuong Ngo, A. A. Winarto, C. K. Li Kou, S. Park, F. Akram, H. K. Lee. "Fence GAN: Towards Better Anomaly Detection", In Proc. 31st International Conference on Tools with Artificial Intelligence (ICTAI), 2019

[4] C. Wang, Y. Zhang, and C. Liu, "Anomaly Detection via Minimum Likelihood Generative Adversarial Networks". In Proc. 24th International Conference on Pattern Recognition (ICPR), 2018

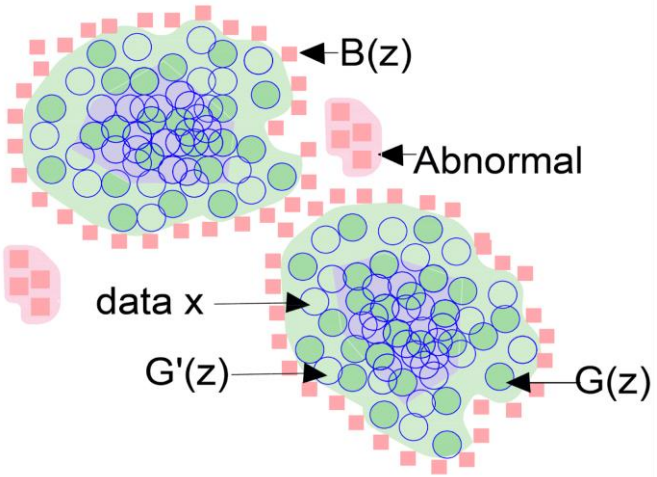
[5] N. Dionelis, M. Yaghoobi, and S. A. Tsaftaris, "Boundary of Distribution Support Generator (**BDSG**): Sample Generation on the Boundary," in Proc. IEEE International Conference on Image Processing (ICIP), 2020

OoD Minimum Anomaly Score GAN (OMASGAN)

OoD Minimum Anomaly Score GAN Model

- **Main problem:**
  - Models may set **high likelihood** and **low reconstruction loss** to OoD samples
    - Overall **decreases** Anomaly Detection (**AD**) performance
- AD models underperform due to the **rarity of anomalies**
- **We develop:** The proposed **OoD Minimum Anomaly Score GAN (OMASGAN)** model

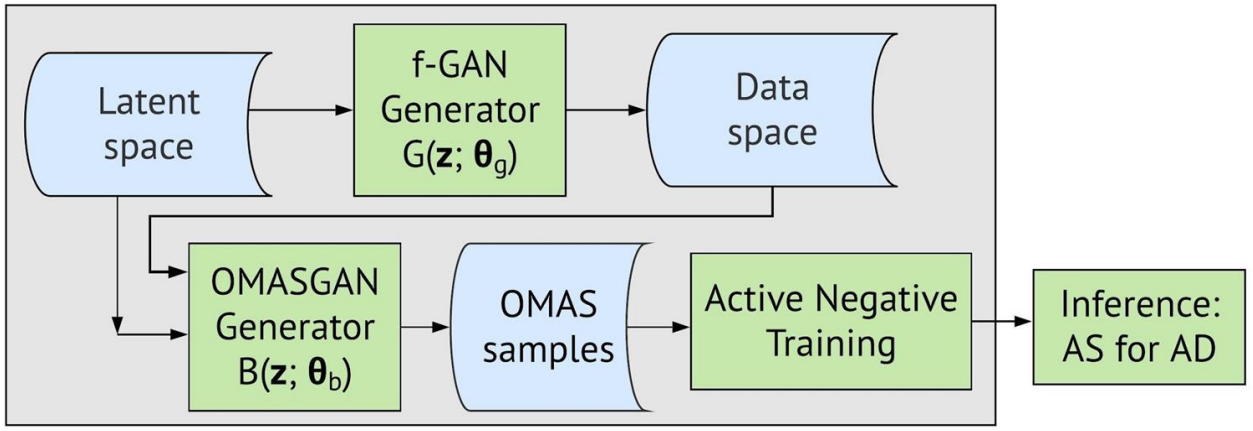
  - **Self-generate** abnormal/ OoD samples
    - **On the boundary** of the support of the normal class data distribution
  - **Invertibility** or probability density: **Not needed**
  - Impose **low likelihood** on this learned data distribution boundary
    - **Include** the proposed **OoD minimum-anomaly score samples** in training
  - Devise a **discriminator** for AD trained on both **negative** and positive data



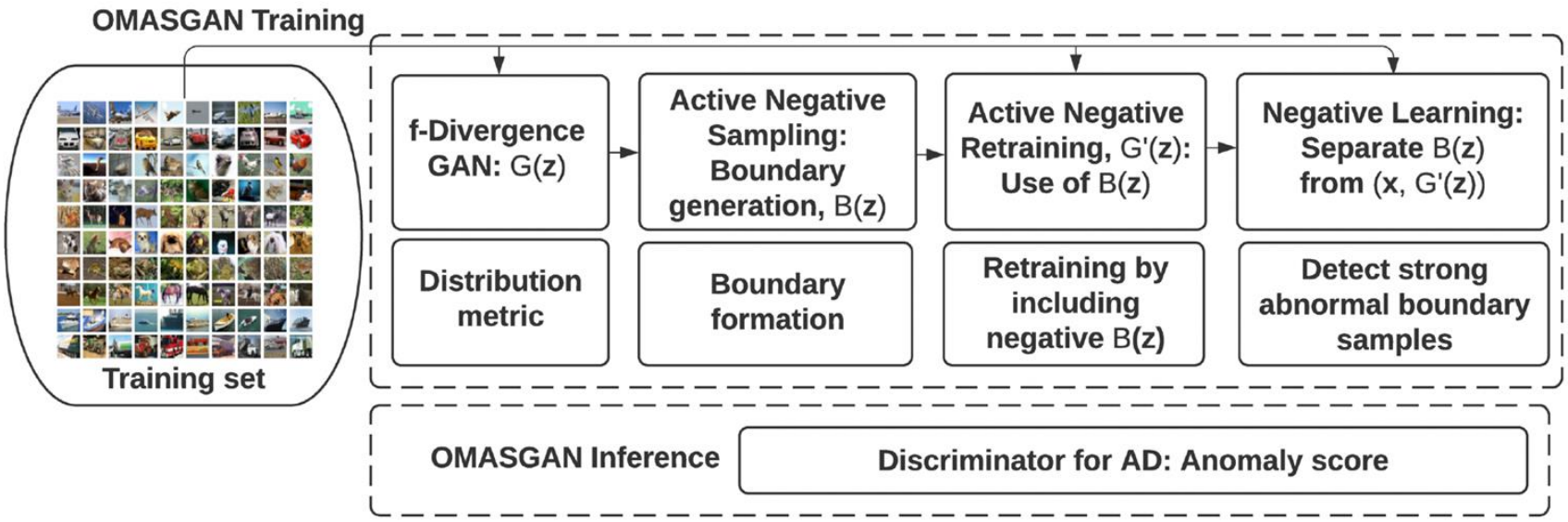
Contributions: OMASGAN Methodology

- **Focus on**: **GANs** to improve **anomaly detection performance**
- **GANs:** Able to learn complex data
  - Achieve convergence in **distribution metrics**

- **Our proposed methodology:** Use **only samples from the normal class**
  - Generate the **OoD boundary** of the support of the **data distribution**
- Find the **OoD minimum-anomaly-score samples** on the distribution boundary
  - With l_p-norm and **dispersion** regularisation
  - Perform **self-supervised learning**
- Integrate with any **f-divergence GAN** generative model



OMASGAN Flowchart



- Train a **f-divergence GAN**: Obtain the **generator**, G(**z**)
- Train a **boundary data generator**, B(**z**), to obtain **OoD distribution boundary samples**
  - **Active negative sampling**
- OMASGAN generates **OoD data**, B(**z**)
  - Incorporate **OoD Minimum Anomaly Score (OMAS) samples** in training
- **Active negative training** of our proposed generator, G'(**z**)
- Devise a **discriminator** for anomaly detection

Our Proposed OMASGAN Algorithm

OMASGAN - Our Proposed Algorithm

- Train a **f-divergence GAN** to establish **a distribution metric**
  - $\min_G \max_D E_{\mathbf{x}} \log(D(\mathbf{x})) + E_{\mathbf{z}} \log(1 - D(G(\mathbf{z})))$
    - where $z \sim p_{\mathbf{z}}(\mathbf{z})$, $G(\mathbf{z}) \sim p_g(\mathbf{x})$, $\mathbf{x} \sim p_{\mathbf{x}}(\mathbf{x})$
- Train the proposed **distribution boundary** model, B(**z**)
  - Perform **active negative sampling**
  - $\min_B -m(B(\mathbf{z}), G(\mathbf{z})) + \lambda\, d(B(\mathbf{z}), G(\mathbf{z})) + \mu\, s(B(\mathbf{z}), \mathbf{z})$
    - where m is a **f-divergence metric**
- **Active negative training: More accurately learn** the underlying data distribution
  - Train G'(**z**) to address **the learning-OoD-samples problem** of G(**z**)
  - $\min_{G'} \max_C \alpha\, E_{\mathbf{x}} \log(C(\mathbf{x})) + \beta\, E_{\mathbf{z}} \log(1\text{-}C(G'(\mathbf{z}))) + (1\text{-}\beta)\, E_{\mathbf{z}} \log(1\text{-}C(B(\mathbf{z}))) + (1\text{-}\alpha)\, E_{\mathbf{z}} \log(C(G(\mathbf{z})))$
    - where C is a discriminator
- Train a **discriminator**, J, to perform **active negative learning for AD**
  - $\max_J \gamma\, E_{\mathbf{x}} \log(1 - J(\mathbf{x})) + (1\text{-}\gamma)\, E_{\mathbf{z}} \log(1 - J(G'(\mathbf{z}))) + E_{\mathbf{z}} \log(J(B(\mathbf{z})))$
    - J learns to **separate** B(**z**) from G'(**z**) and **x**

Generation of OMAS Samples

- Generate **OoD minimum-anomaly-score samples**, B(**z**)
  - On the **boundary of the normal class data distribution**
  - arg $\min_B$ -m(B(**z**), G(**z**)) + λ d(B(**z**), G(**z**)) + μ s(B(**z**), **z**)

$$d(B(\mathbf{z};\boldsymbol{\theta_b}), G(\mathbf{z})) = \min_{j=1,\ldots,Q} ||B(\mathbf{z};\boldsymbol{\theta_b}) - G(\mathbf{z}_j)||_p^q$$

$$s(B(\mathbf{z}_i;\boldsymbol{\theta_b}), \mathbf{z}_i) = \frac{1}{N-1} \sum_{j=1,\, j\neq i}^{N} \frac{||\mathbf{z}_i - \mathbf{z}_j||_p^q}{||B(\mathbf{z}_i;\boldsymbol{\theta_b}) - B(\mathbf{z}_j;\boldsymbol{\theta_b})||_p^q}$$

- Train our proposed **GAN-based boundary formation** model, B(**z**), for AD:
  - Propose a **loss** with three terms
- B(**z**; **θ**); Run Gradient Descent on the proposed **loss**; Obtain **θ**
  - Batch size, N
  - **Inference sample** size, Q, for G(**z**)

- Force the samples to the **boundary of the data distribution**
- Effectively address **mode collapse**

- **First term:** Decreasing function of a **distribution divergence metric**
- **Second term:** l_p-norm distance
- **Third term:** Scattering, dispersion
  - Capture **all the modes**

OMASGAN Inference Mechanism

- **Anomaly score/ OoD score** based on the **Anomaly Discriminator**, J, and the **f-divergence distribution metric**
  - The **discriminator**, J, is trained to:
    - **Separate** the **normal class distribution** from its **complement**
- **f-divergence distribution metric:** Used for training and during **inference**

- **f-divergence** for probability distributions **P** and **Q**: **fD(P, Q)**
- For a **queried test** sample, **x*:**
  - **Calculate** fD(G', δ_{x^*})
    - where δ_{x^*} is a Dirac function centered at **x***
- **Our proposed anomaly/ OoD score:** AS(**x***; J, G') = J(**x***) + λ fD(G', δ_{x^*})
- Classification decision: **x*** is from the **normal class** if AS(**x***; J, G') < τ
  - where τ is a threshold
  - **x*** is **abnormal** otherwise

Evaluation of OMASGAN

AD Evaluation of the Proposed Model

- **Evaluation methods for anomaly/ OoD detection:**
  - Leave-one-out **(LOO)** evaluation
    - **Normal class: 9 classes** from a benchmark dataset with 10 classes
    - **Abnormal class:** Left-out class
- **Evaluation metrics:** Algorithm convergence criteria; Area Under the Receiver Operating Characteristics Curve **(AUROC)**

- **Examined datasets:**
  - Synthetic; **MNIST**; **CIFAR-10**

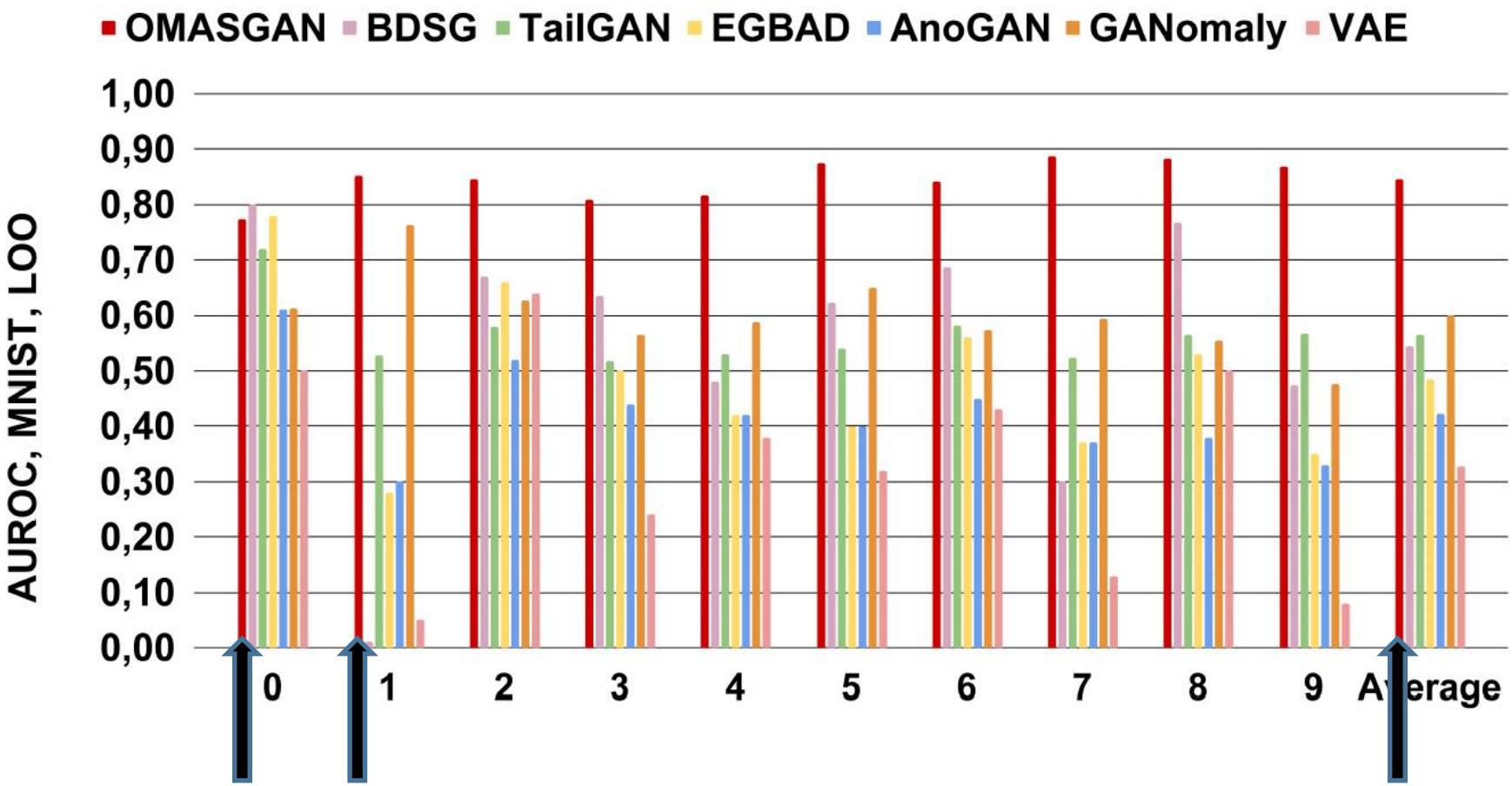
- **Baselines: GANomaly** [2]; VAE EGBAD; AnoGAN; **FenceGAN** [3]; MinLGAN [4] **BDSG** [5]; TailGAN

Our proposed model:

- **OMASGAN**
  - KLWGAN-based **OMASGAN**
  - f-GAN-based **OMASGAN**

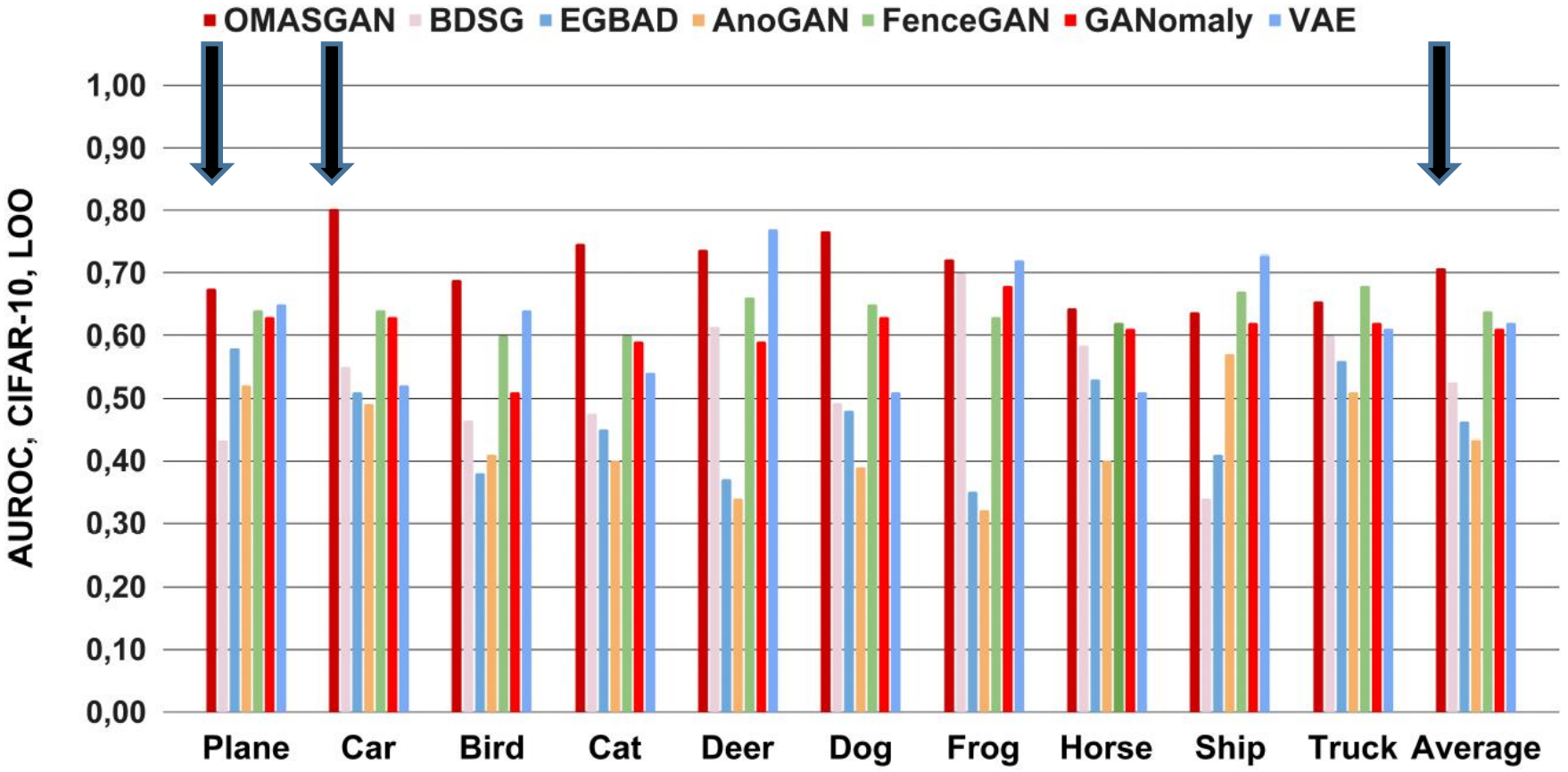
Evaluation of the OMASGAN Model on MNIST

- **Performance of OMASGAN** on MNIST data in **AUROC**
  - Compared to **GAN and AE baselines** using **LOO evaluation**

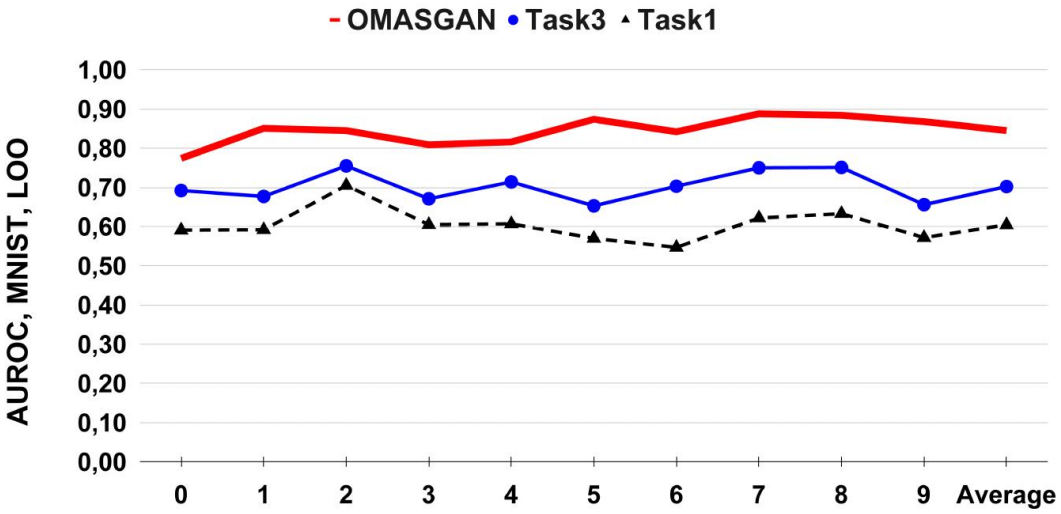


Evaluation of OMASGAN on CIFAR-10 Data

- **Performance of OMASGAN** in **AUROC** on the CIFAR-10 dataset
  - Comparison with **GAN and AE baselines** using **LOO evaluation**

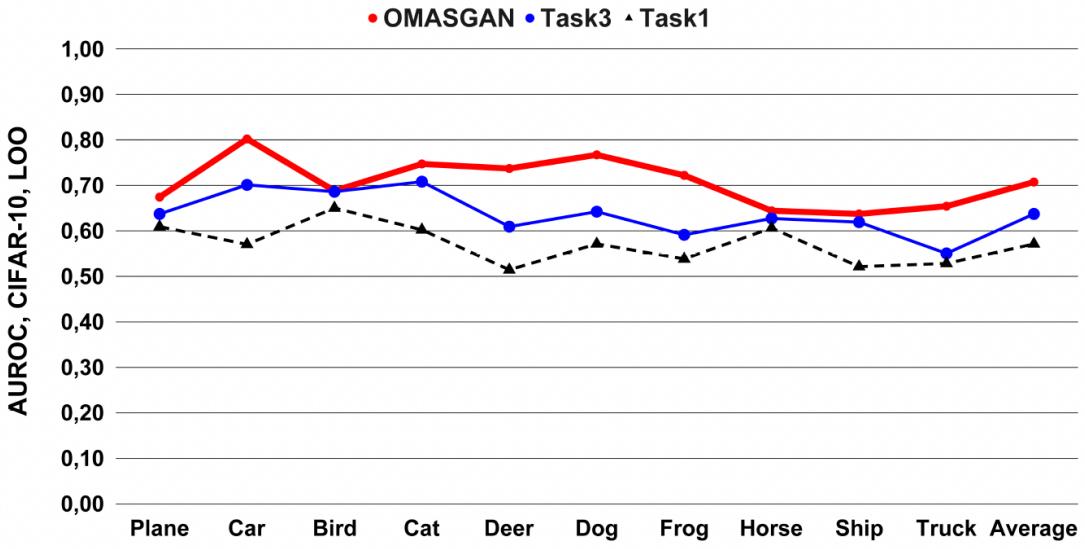


Evaluation of OMASGAN - Ablation Study



- **Ablation study** of OMASGAN in AUROC **on MNIST (left)** and on **CIFAR-10 (bottom)**
- We examine the **impact of our loss functions**
  - Using **LOO evaluation**

- **OMASGAN** outperforms:
  - Our chosen **base model**, KLWGAN
  - **Ablation study:** G'(z)



Conclusion

Conclusion

- **OMASGAN: GANs** to improve the **anomaly detection/ OoD detection performance**
- Use **only** data from the **normal class**
  - Perform **active negative sampling**
    - Generate abnormal **distribution boundary** samples
  - Perform **active negative training** for anomaly detection

- Address the **learning-OoD-samples problem** of generators
  - **Contrastive negative training** to alleviate the problem of deep **generative** models **knowing what they do not know** (setting **high likelihood** to **OoD data**)
- Effectively tackle the **rarity of anomalies problem**
- Propose a **discriminator-based anomaly score/ OoD score**
- **OMASGAN outperforms** the examined baselines
  - Using the **LOO evaluation** methodology
  - In **AUROC**, on the MNIST and CIFAR-10 datasets

Thank you very much for your attention!

Contact email: Nikolaos.Dionelis@ed.ac.uk

References

- [1] N. Dionelis, M. Yaghoobi, and S. A. Tsaftaris. "Boundary of Distribution Support Generator (**BDSG**): Sample Generation on the Boundary", in Proc. IEEE International Conference on Image Processing (ICIP), 2020
- [2] N. Dionelis, S. A. Tsaftaris, and M. Yaghoobi. OMASGAN: Out-of-Distribution Minimum Anomaly Score GAN for Sample Generation on the Boundary, https://github.com/nd1511/OMASGAN_
- [3] N. Dionelis, M. Yaghoobi, and S. A. Tsaftaris. "Tail of Distribution GAN (**TailGAN**): Generative Adversarial Network Based Boundary Formation", in Proc. Sensor Signal Processing for Defence (SSPD), 2020
- [4] S. Akcay, A. Atapour-Abarghouei, and T. Breckon. GANomaly: Semi-Supervised Anomaly Detection via Adversarial Training. arXiv preprint, arXiv:1805.06725 [cs.CV], 2018
- [5] H. Zenati, C. Foo, B. Lecouat, G. Manek, and V. Ramaseshan Chandrasekhar. Efficient GAN-Based Anomaly Detection. Workshop Track in International Conference on Learning Representations (ICLR), 2018
- [6] T. Schlegl, P. Seeböck, S. Waldstein, U. Schmidt-Erfurth, and G. Langs. Unsupervised Anomaly Detection with GANs to Guide Marker Discovery. In Proc. Information Processing in Medical Imaging (IPMI), 2017
- [7] P. Cuong Ngo, A. A. Winarto, C. K. Li Kou, S. Park, F. Akram, H. K. Lee. "Fence GAN: Towards Better Anomaly Detection", In Proc. 31st International Conference on Tools with Artificial Intelligence (ICTAI), 2019
- [8] C. Wang, Y. Zhang, and C. Liu, "Anomaly Detection via Minimum Likelihood Generative Adversarial Networks". In Proc. 24th International Conference on Pattern Recognition (ICPR), 2018
- [9] R. Devon Hjelm, A. Paul Jacob, T. Che, A. Trischler, K. Cho, and Y. Bengio. Boundary-Seeking Generative Adversarial Networks. arXiv preprint, arXiv:1702.08431 [stat.ML], 2018
- [10] L. Ruff, R. Vandermeulen, N. Gornitz, L. Deecke, S. Siddiqui, A. Binder, E. Muller, and M. Kloft. Deep One-Class Classification. In Proc. International Conference on Machine Learning (ICML), 2018
- [11] L. Ruff, R. A. Vandermeulen, et al., "Deep Semi-Supervised Anomaly Detection", arXiv:1906.02694, 2020