technische universität
dortmund

Anomaly Detection using GANs

Presenter: Dhanunjaya Elluri Thimmaraju (231767)

Prof. Dr. Emmanuel Müller and M. Sc. Benedikt Böing

26.01.2022

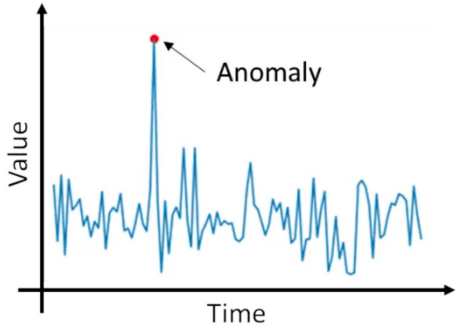
Overview

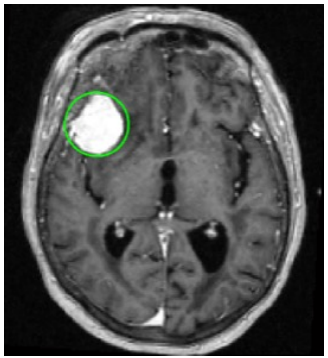
- Introduction
- Motivation
- Introduction to GANs
- BiGANs
- GANs for Anomaly Detection
- Experiments and Results
- Summary
- Reference

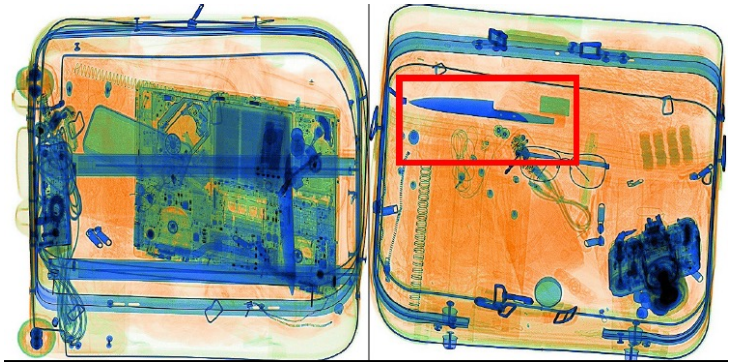
Introduction

Anomalies

- Abnormal patterns in data
- Often referred as Noise or Outliers
- Anomaly detection is applicable in various domains
  - Fraud Detection (Tabular data, Time series)
  - Medical Imaging (Images)
  - Baggage Scanning (Images) etc.







Motivation

Disadvantages of Supervised Anomaly Detection

- High annotation effort
- Class imbalance
- Cannot detect unseen anomalies during test time

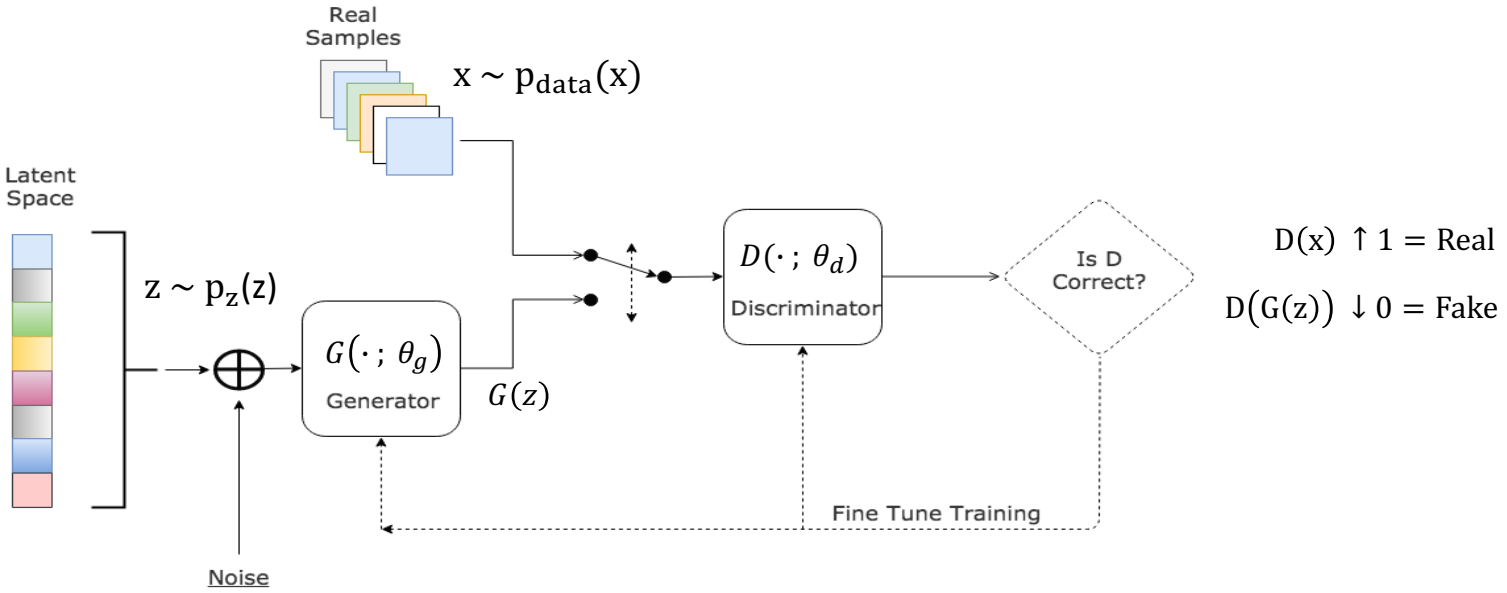
Introduction to Unsupervised Anomaly Detection

- AnoGAN (GANs)
- EGBAD (BiGANs)
- GANomaly (Modified GANs)

Intuition on Generative Adversarial Networks (GANs)

GANs are Neural Networks trained in adversarial manner.
It has two parts:

1. Discriminator (D): Acts as a binary classifier. Classify between real and fake data.
2. Generator (G): Generate data (X') as close as possible to true distribution (X).



Source: https://bolster.ai/blog/gans-in-real-world-can-bad-actors-use-gans-to-beat-ai/

Understanding Loss Function

Discriminator point of view → $\uparrow D(x)$ $\downarrow D(G(z))$

Generator point of view → $\uparrow D(G(z))$

$$\log(D(x)) \quad \log(1 - D(G(z)))$$

$$\mathbb{E}_{x \sim P_{data}(X)}[\log(D(x)] \quad \mathbb{E}_{z \sim P_z(Z)}[\log(1 - D(G(z)))]$$

$$V(D, G) = \mathbb{E}_{x \sim P_{data}(X)}[\log(D(x)] + \mathbb{E}_{z \sim P_z(Z)}[\log(1 - D(G(z)))]$$

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim P_{data}(X)}[\log(D(x)] + \mathbb{E}_{z \sim P_z(Z)}[\log(1 - D(G(z)))]$$

Optimization

Algorithm 1 Minibatch stochastic gradient descent training of generative adversarial nets. The number of steps to apply to the discriminator, k, is a hyperparameter. We used $k = 1$, the least expensive option, in our experiments.

for number of training iterations **do**

  for k steps **do**

  - Sample minibatch of m noise samples $\{\boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(m)}\}$ from noise prior $p_g(\boldsymbol{z})$.
  - Sample minibatch of m examples $\{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(m)}\}$ from data generating distribution $p_{\text{data}}(\boldsymbol{x})$.
  - Update the discriminator by ascending its stochastic gradient:

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^{m} \left[\log D\left(\boldsymbol{x}^{(i)}\right) + \log\left(1 - D\left(G\left(\boldsymbol{z}^{(i)}\right)\right)\right) \right].$$

  end for

  - Sample minibatch of m noise samples $\{\boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(m)}\}$ from noise prior $p_g(\boldsymbol{z})$.
  - Update the generator by descending its stochastic gradient:

$$\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^{m} \log\left(1 - D\left(G\left(\boldsymbol{z}^{(i)}\right)\right)\right).$$

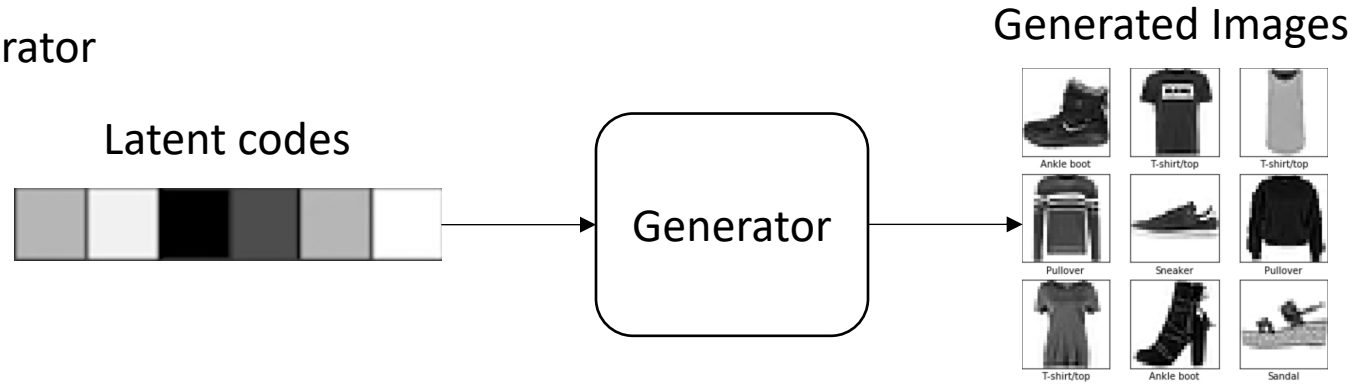
end for

The gradient-based updates can use any standard gradient-based learning rule. We used momentum in our experiments.

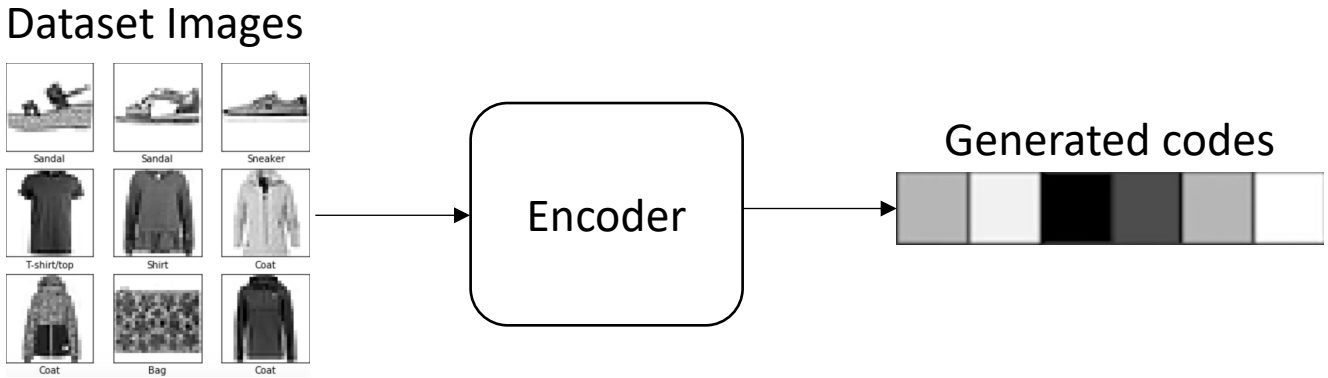
Source: Generative Adversarial Networks, 2014

Bi-Directional GANs (BiGANs)

1. Generator

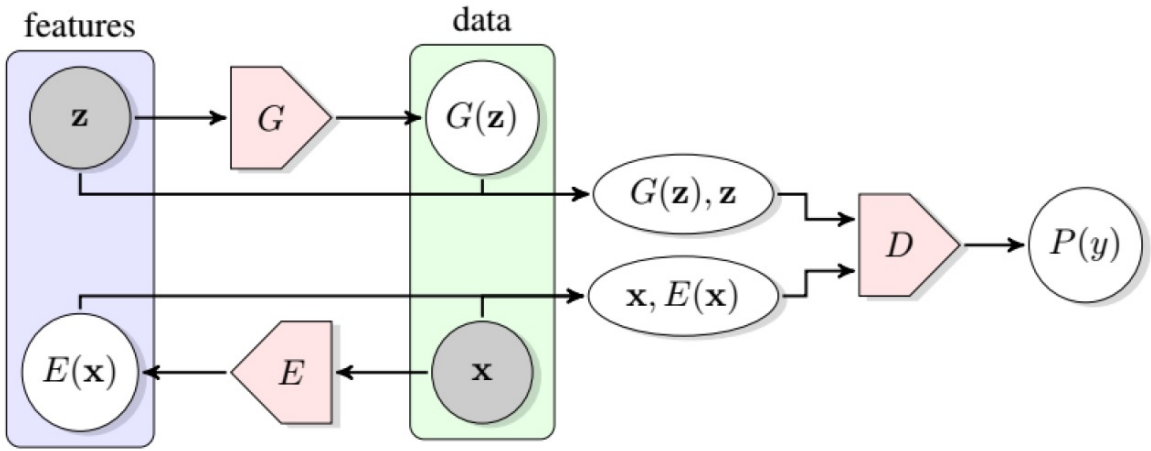


2. Encoder



Bi-Directional GANs (BiGANs)

Architecture:



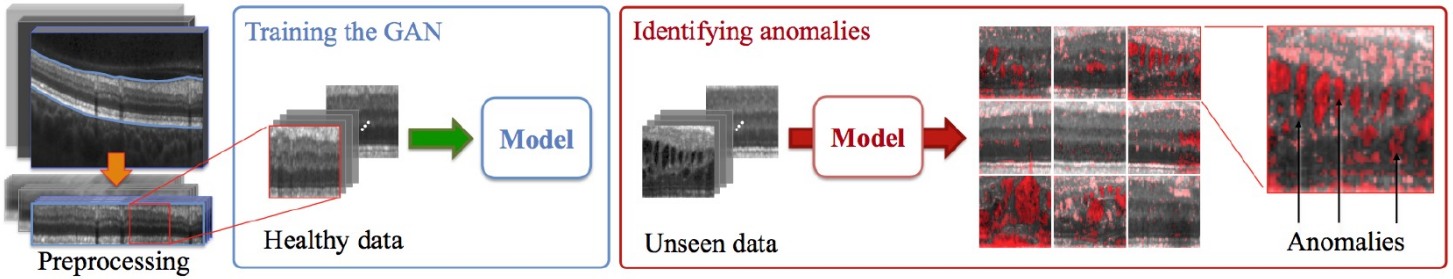
Loss Function:

$$\min_{G,E} \max_{D} V(D,E,G) = \mathbb{E}_{x \sim P_{data}(X)}[\log(D(x,E(x))] + \mathbb{E}_{z \sim P_z(Z)}[\log(1 - D(G(z),z))]$$

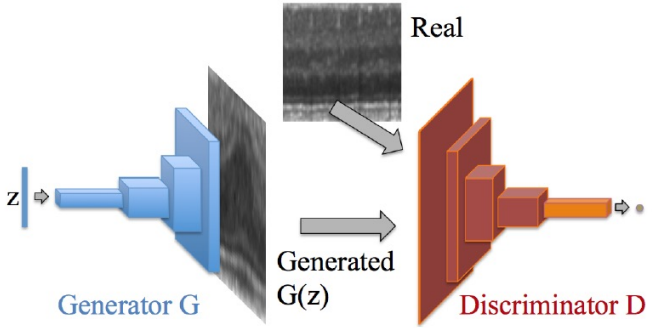
Source: Adversarial Feature Learning, 2017

AnoGAN

Anomaly Detection Technique



Uses DCGANs



Source: Unsupervised Anomaly Detection with GANs to guide Marker Discovery, 2017

AnoGAN

1. $G(\mathbf{z}) = \mathbf{z} \mapsto \mathbf{x}$

$$\min_{G} \max_{D} V(D,G) = \mathbb{E}_{x \sim P_{data}(X)}[\log(D(x)] + \mathbb{E}_{z \sim P_z(Z)}[\log(1 - D(G(z)))]$$

2. $\mu(\mathbf{x}) = \mathbf{x} \mapsto \mathbf{z}$

3. Loss Function

   a. Residual Loss:

$$\mathcal{L}_R(\mathbf{z}_\gamma) = \sum |\mathbf{x} - G(\mathbf{z}_\gamma)|$$

$$\Longrightarrow \quad \mathcal{L}(\mathbf{z}_\gamma) = (1-\lambda) \cdot \mathcal{L}_R(\mathbf{z}_\gamma) + \lambda \cdot \mathcal{L}_D(\mathbf{z}_\gamma)$$

   b. Discriminator Loss:

$$\mathcal{L}_D(\mathbf{z}_\gamma) = \sum |\mathbf{f}(\mathbf{x}) - \mathbf{f}(G(\mathbf{z}_\gamma))|$$

4. Anomaly Score

$$A(\mathbf{x}) = \mathcal{L}(\mathbf{z}_\Gamma)$$

Source: Unsupervised Anomaly Detection with GANs to guide Marker Discovery, 2017

Efficient GAN-Based Anomaly Detection (EGBAD)

- AnoGANs require Γ optimization steps
- Enables encoder E to map input samples to their latent representation
- Anomaly Score:

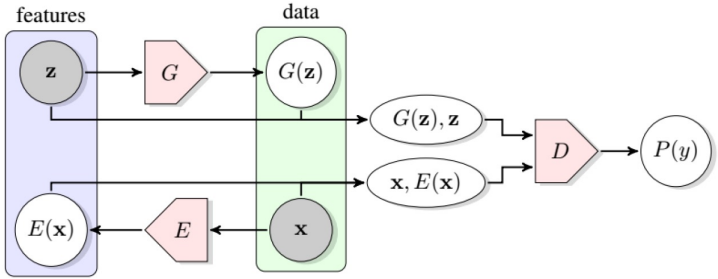
$$A(x) = \alpha L_G(x) + (1 - \alpha) L_D(x),$$

where

$$L_G(x) = \left|\left| x - G(E(x)) \right|\right|_1$$

And

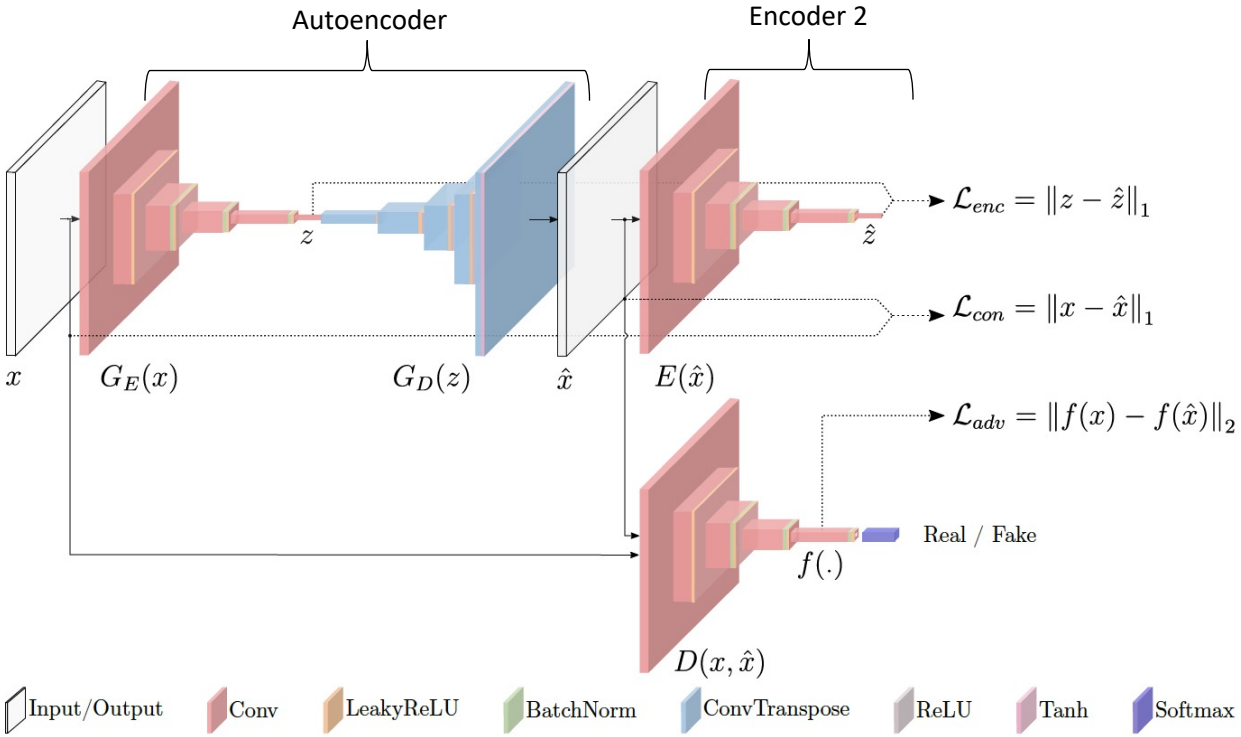
$$L_D(x) = \left|\left| f_D(x, E(x)) - f_D\left(G(E(x)), E(x)\right) \right|\right|_1$$



Source: Efficient GAN-Based Anomaly Detection, 2018

GANomaly

Model: Autoencoder + Encoder + Discriminator



Source:GANomaly: Semi-Supervised Anomaly Detection via Adversarial Training, 2018

GANomaly

Contextual Loss: Make the reconstruction image similar to input image

$$\mathcal{L}_{con} = \mathbb{E}_{\mathbf{x} \sim p_X} ||x - G(\mathbf{x})||_1$$

Encoder Loss: Let the Encoder 2 learn encoding the normal image

$$\mathcal{L}_{enc} = \mathbb{E}_{\mathbf{x} \sim p_X} ||G_E(\mathbf{x}) - E(G(\mathbf{x}))||_2$$

Adversarial Loss: Feature matching

$$\mathcal{L}_{adv} = \mathbb{E}_{\mathbf{x} \sim p_X} ||f(\mathbf{x}) - \mathbb{E}_{\mathbf{x} \sim p_X} f(G(\mathbf{x}))||_2$$

Overall Loss:

$$\mathcal{L} = w_{adv}\mathcal{L}_{adv} + w_{con}\mathcal{L}_{con} + w_{enc}\mathcal{L}_{enc}$$

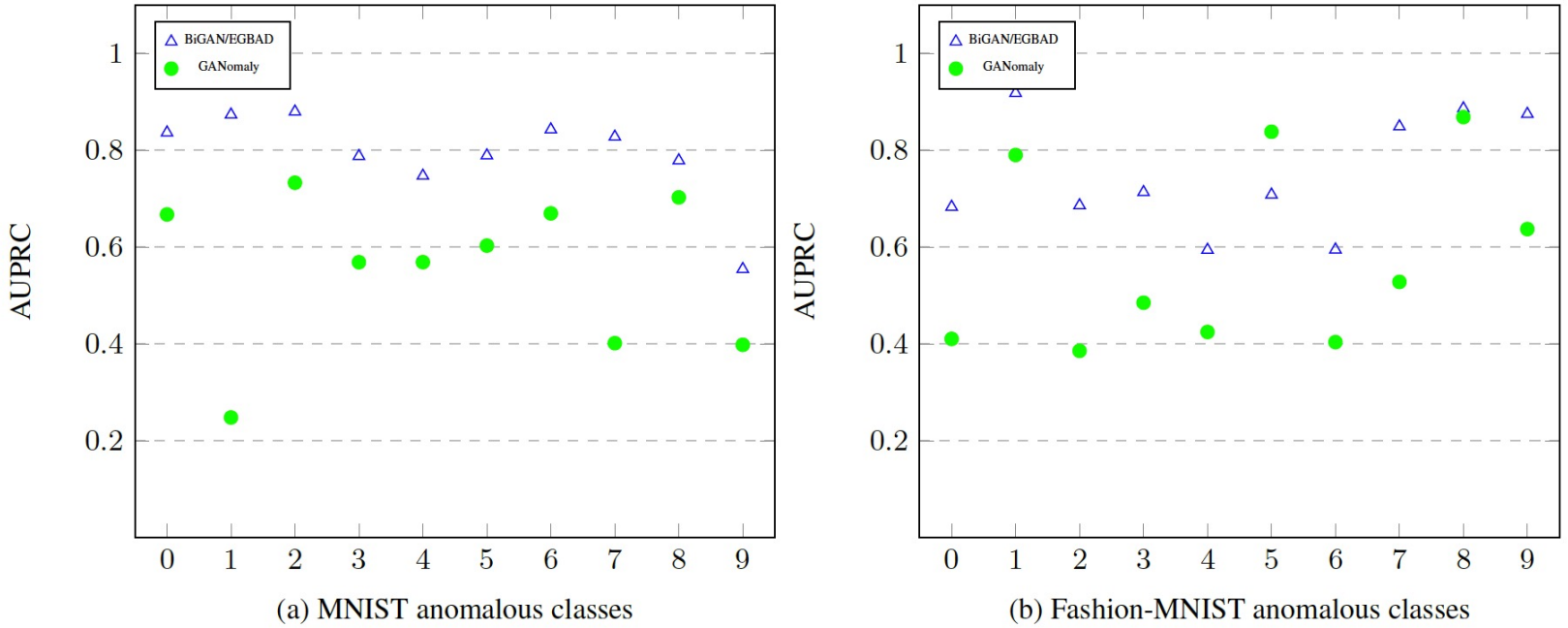
Anomaly Score:

$$\mathcal{A}(\mathbf{x}) = ||G_E(\mathbf{x}) - E(G(\mathbf{x}))||_2$$

Source: GANomaly: Semi-Supervised Anomaly Detection via Adversarial Training, 2018

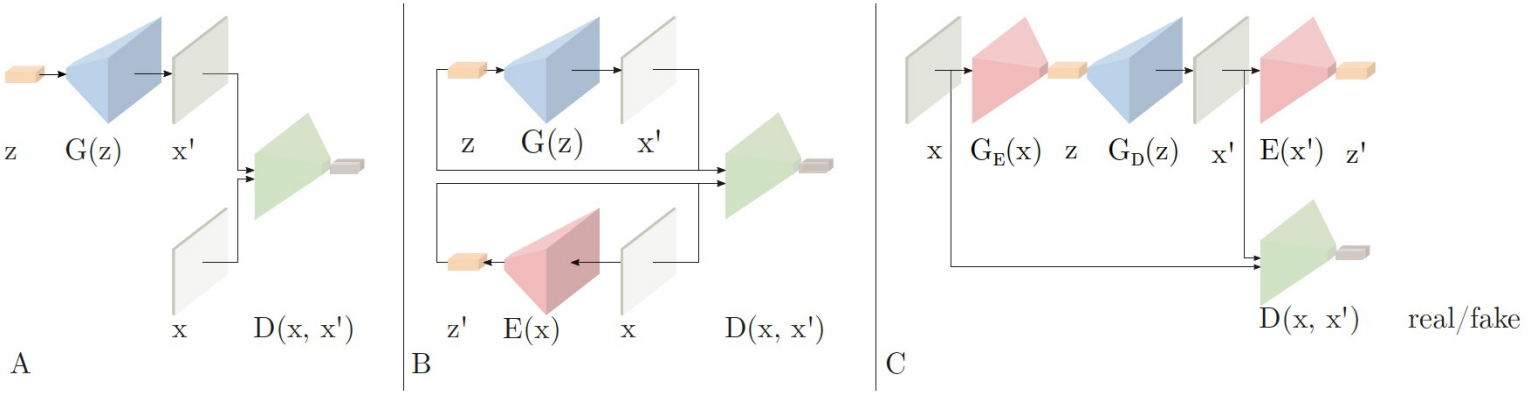
Experiments and Results

BiGAN/EGBAD and GANomaly performance comparison on MNIST and Fashion-MNIST datasets



Summary

- Discussion of drawbacks in Supervised Anomaly Detection
- Introduction to Unsupervised Anomaly Detection
- Discussed GANs and BiGANs architectures
- Discussed AnoGANs, EGBAD and GANomaly



A. AnoGAN

B. EGBAD

C. GANomaly

References

- Federico Di Mattia, Paolo Galeone, Michele De Simoni, Emanuele Ghelfi. A Survey on GANs for Anomaly Detection. abs/1906.11632, 2019. URL https://arxiv.org/abs/1906.11632
- Akcay, S., Abarghouei, A. A., and Breckon, T. P. GANomaly: Semi-Supervised Anomaly Detection via Adversarial Training. abs/1805.06725, 2018. URL http://arxiv.org/abs/1805.06725.
- Zenati, H., Foo, C. S., Lecouat, B., Manek, G., and Chandrasekhar, V. R. Efficient GAN-Based Anomaly Detection. abs/1802.06222, 2018. URL http://arxiv.org/abs/1802.06222.
- Schlegl, T., Seeb ock, P., Waldstein, S. M., Schmidt-Erfurth, U., and Langs, G. Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery. abs/1703.05921, 2017. URL http://arxiv.org/abs/1703.05921.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. Generative Adversarial Nets. pp. 2672–2680, 2014. URL http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf.