Adversarially Learned Anomaly Detection

IEEE International Conference on Data Mining, Singapore

Houssam Zenati [*,1,4] Manon Romain [*,2,4] Chuan Sheng Foo [*,4]
Bruno Lecouat [3,4] Vijay Chandrasekhar [4]

[*]Authors contributed equally to this work

[1]CentraleSupélec

[2]École Polytechnique

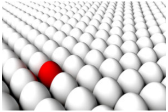
[3]Télécom ParisTech

[4]Institute for Infocomm Research, A*STAR

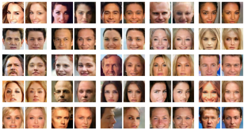
November 18, 2018

Context

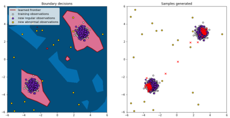
Problem: Perform anomaly detection



We wish to effectively model the data generating distribution ...



... and derive a statistically sound decision criteria for anomaly detection



Outline

Anomaly Detection methods

Brief overview of anomaly detection methods:

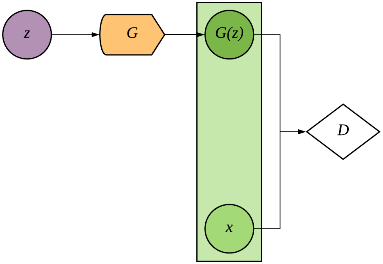
- Distance-based: NN [5]
- One-class classification: OCSVM [18]
- Reconstruction-based: PCA [10, 4]. Auto-encoders [21, 1]
- Energy or GMM based: DEBSM [19] and DAGMM [3]
- GANs method proposed in AnoGAN [17]

Shortcomings:

- For classic machine learning methods: curse of dimensionality in high dimensional data
- Neural network based: better modeling power with GANs
- GAN-based method: computationally highly expensive at inference time

What are GANs?

GANs [8]: two competing networks, generator G / discriminator D.



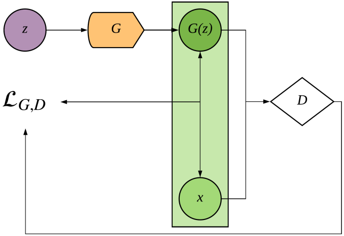
$$\min_G \max_D \mathbb{E}_{x \sim p_{\mathcal{X}}} \left[\log D(x) \right] + \mathbb{E}_{z \sim p_{\mathcal{Z}}} \left[\log \left(1 - D\left(G(z) \right) \right) \right]$$

Can model complex/high dimensional distributions of data [16]

Can we use GANs for Anomaly Detection?

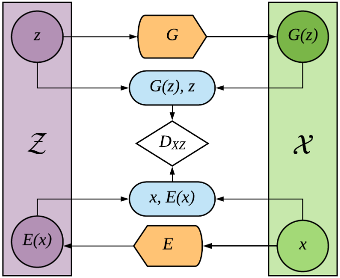
Reconstruction based: need to find latent representations.

AnoGAN [17]: standard GAN



$$z = \arg\min_z \mathcal{L}_{G,D}(z,x)$$

Bi-directionals GANs [7, 6]

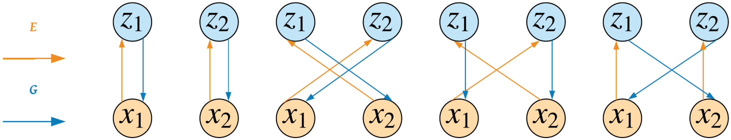


$$z = E(x)$$

500 forward/backward pass for AnoGAN, 1 forward pass for BiGANs

Cycle consistency

Non convergence to the saddle point optimization problem breaks cycle consistency $G(E(x)) \not\approx x$ (ALICE [12]).



Like ALICE [12], we impose a regularizer on the encoder E and the generator G with an additional discriminator network $D_{xx}(x, G(E(x)))$ to enhance $G(E(x)) \approx x$

Stabilizing the training

- Add another constraint on the latent space $D_{zz}(z, E(G(z)))$ to ensure $E(G(z)) \approx z$
- apply spectral normalization [14] on the discriminators of GANs [2, 9, 14, 20] and on the encoder

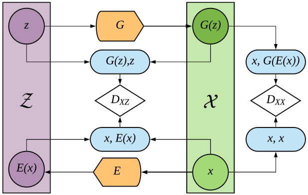


Figure: ALICE [17]

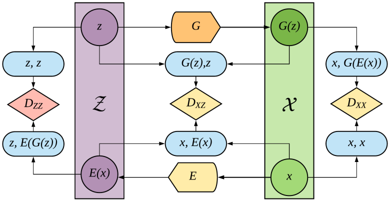
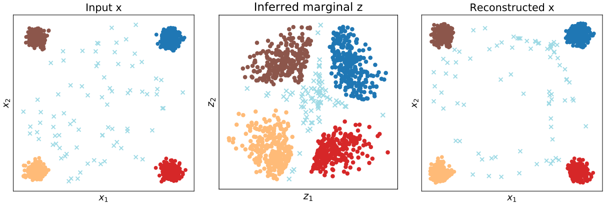


Figure: ALAD

Detecting anomalies

Train a model on the normal data to provide E, G.



Compute $A(x)$: L_1 reconstruction error between samples in the feature space of the cycle-consistency discriminator D_{xx}.

$$A(x) = \left|\left| f_{xx}(x, x) - f_{xx}(x, G(E(x))) \right|\right|_1$$

$f(\cdot, \cdot)$: activations of the layer before the logits (CNN codes) in the D_{xx} network for an input pair x, x'

Reconstructions

Normal
Reconstr.
Abnormal
Reconstr.

Figure: SVHN

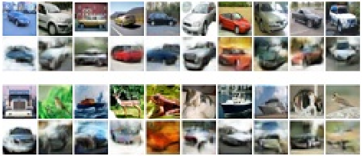


Figure: CIFAR10

Experiments

Experiments on publicly available tabular (KD99, Arrythmia [13]) and image datasets (CIFAR-10, SVHN [11, 15]).

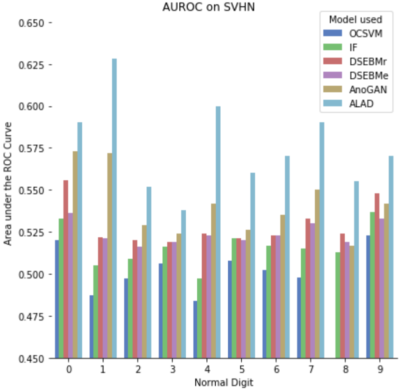


Figure: SVHN

Model	Prec.	Recall	F1
OC-SVM	0.746	0.852	0.795
IF	0.922	0.937	0.929
DSEBM-r	0.852	0.647	0.733
DSEBM-e	0.862	0.645	0.740
DAGMM	0.930	0.944	0.937
AnoGAN	0.879	0.830	0.887
ALAD	**0.943**	**0.958**	**0.950**

Table: KDD99

Ablation Study

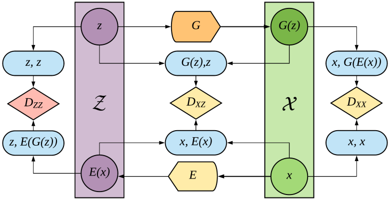


Figure: Spectral Norm (SN, yellow), Discriminator Latent (DL red)

Model	AUROC
Baseline	0.5701 ± 0.1282
Baseline + DL	0.5361 ± 0.1348
Baseline + SN	0.5991 ± 0.1308
Baseline + SN + DL	0.6072 ± 0.1201

Table: Ablation Study on CIFAR-10

Inference time

Dataset	Batch Size	AnoGAN	ALAD	Speed Up
KDD99	50	1235	1.4	~ 900
Arrhythmia	32	1102	41	~ 30
SVHN	32	10496	10.5	~ 1000
CIFAR-10	32	10774	10.5	~ 1000

Table: Average inference time (ms) on a GeForce GTX TITAN X

Conclusion

ALAD, a GAN-based anomaly detection method: not only effective, but also efficient at test time.

- Utilizes the bi-directional class of GANs [7, 6] to enable $\sim$ 1000X faster inference time.
- Better reconstructs samples with cycle-consistency technique [12].
- Uses training stabilization technique for GANs [14]
- Achieves state-of-the-art on a range of highdimensional tabular and image data.

J. An and S. Cho.
Variational autoencoder based anomaly detection using reconstruction probability.
Technical Report, 2015.

Martín Arjovsky, Soumith Chintala, and Léon Bottou.
Wasserstein gan.
Proceedings of Machine Learning Research, abs/1701.07875, 2017.

Zong Bo, Qi Song, Martin Renqiang Min, Wei Chengand Cristian Lumezanu, Daeki Cho, and Haifeng Chen.
Deep autoencoding gaussian mixture model for unsupervised anomaly detection.
International Conference on Learning Representations, 2018.

Emmanuel J. Candès, Xiaodong Li, Yi Ma, and John Wright.
Robust principal component analysis.
Journal of the ACM, 2009.

Varun Chandola, Arindam Banerjee, and Vipin Kumar.
Anomaly detection: A survey.
ACM Comp. Sur., 41:15:1:15:58, 2009.

Jeff Donahue, Philipp Krähenbühl, and Trevor Darrell.
Adversarial feature learning.
International Conference on Learning Representations, 2017.

Vincent Dumoulin, Ishmael Belghazi, Ben Poole, Alex Lamb, Martin Arjovsky, Olivier Mastropietro, and Aaron Courville.
Adversarially learned inference.
International Conference on Learning Representations, 2017.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio.
Generative adversarial nets.
In *Advances in Neural Information Processing Systems 27*, pages 2672–2680. 2014.

Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C Courville.
Improved training of wasserstein gans.
In *Advances in Neural Information Processing Systems 30*, pages 5767–5777. 2017.

S. Günter, Nicol N. Schraudolph, and S. V. N. Vishwanathan.
Fast iterative kernel principal component analysis.
Journal of Machine Learning Research, pages 1893–1918, 2007.

Alex Krizhevsky.
Learning multiple layers of features from tiny images.
Technical Report, 2012.

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

In *Advances in Neural Information Processing Systems 30*, pages 5495–5503. 2017.

M. Lichman.
UCI machine learning repository, 2013.

Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida.
Spectral normalization for generative adversarial networks.
In *International Conference on Learning Representations*, 2018.

Yuval Netzer, Tao Wang, Adam Coates, Alessandro Bissacco, Bo Wu, and Andrew Y Ng.
Reading digits in natural images with unsupervised feature learning.
In *NIPS Workshop on Deep Learning and Unsupervised Feature Learning*, 2011.

Alec Radford, Luke Metz, and Soumith Chintala.

Unsupervised representation learning with deep convolutional generative adversarial networks.
International Conference on Learning Representations, Workshop Track, 2016.

Thomas Schlegl, Philipp Seeböck, Sebastian M. Waldstein, Ursula Schmidt-Erfurth, and Georg Langs.
Unsupervised anomaly detection with generative adversarial networks to guide marker discovery.
International Conference on Information Processing in Medical Imaging, page pp. 146–157, 2017.

Bernhard Schölkopf, Robert Williamson, Alex Smola, John Shawe-Taylor, and John Platt.
Support vector method for novelty detection.
In *Proceedings of the 12th International Conference on Neural Information Processing Systems*, pages 582–588, 1999.

Shuangfei Zhai, Yu Cheng, Weining Lu, and Zhongfei Zhang.
Deep structured energy based models for anomaly detection.

International Conference on Machine Learning, pages 1100-1109, 2016.

H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena.
Self-Attention Generative Adversarial Networks.
ArXiv e-prints, May 2018.

Chong Zhou and Randy C. Paffenroth.
Anomaly detection with robust deep autoencoders.
Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 665-674, 2017.