Deep Anomaly Detection

Kang, Min-Guk

Mingukkang1994@gmail.com

Jan. 16, 2019

Contents

Introduction of Anomaly Detection

(Vision data only)

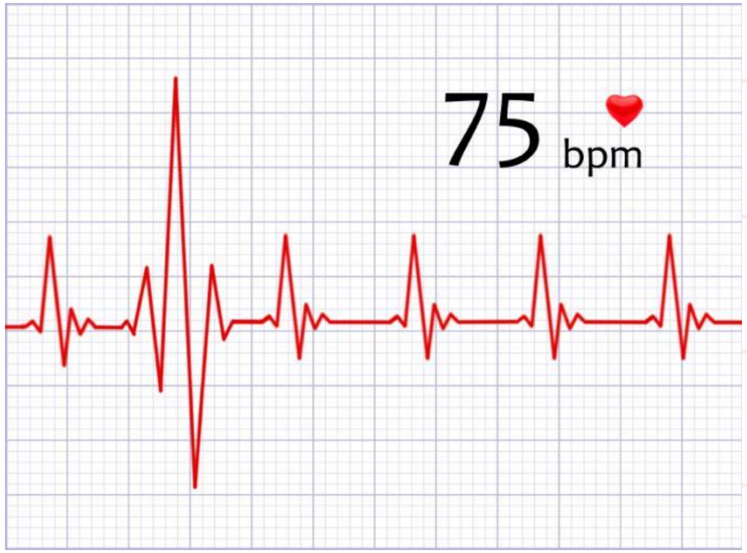
1. Introduction

Anomaly Detection is the process of identifying the new or unexplained set of data to determine if they are within the norm or outside of it.

1. Introduction

OODD: Out Of Distribution Detection!

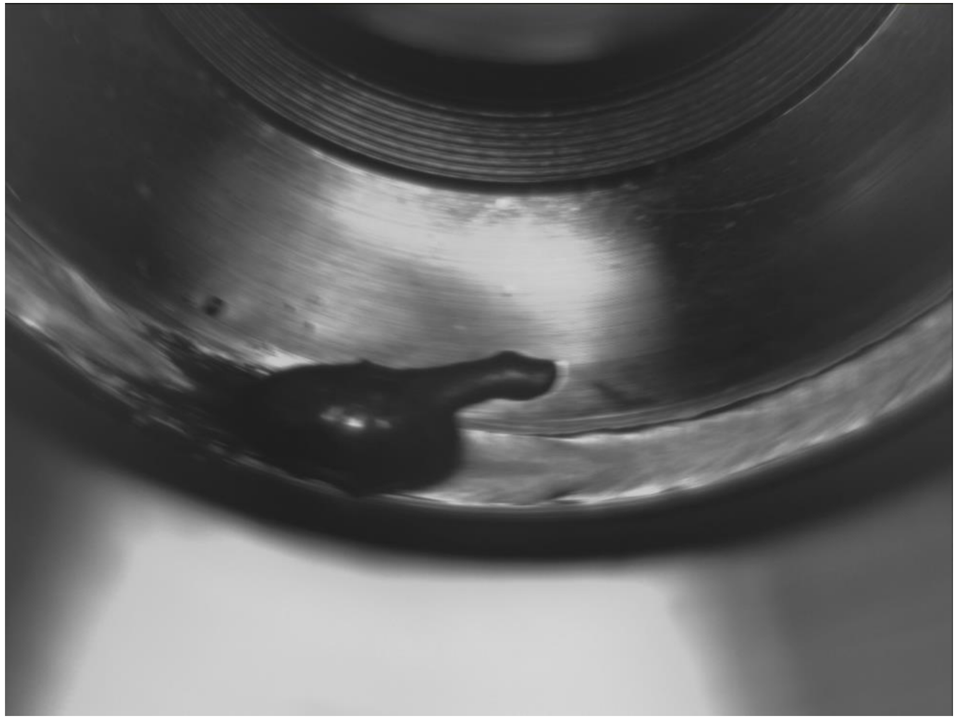
1. Introduction(Time Series Data)



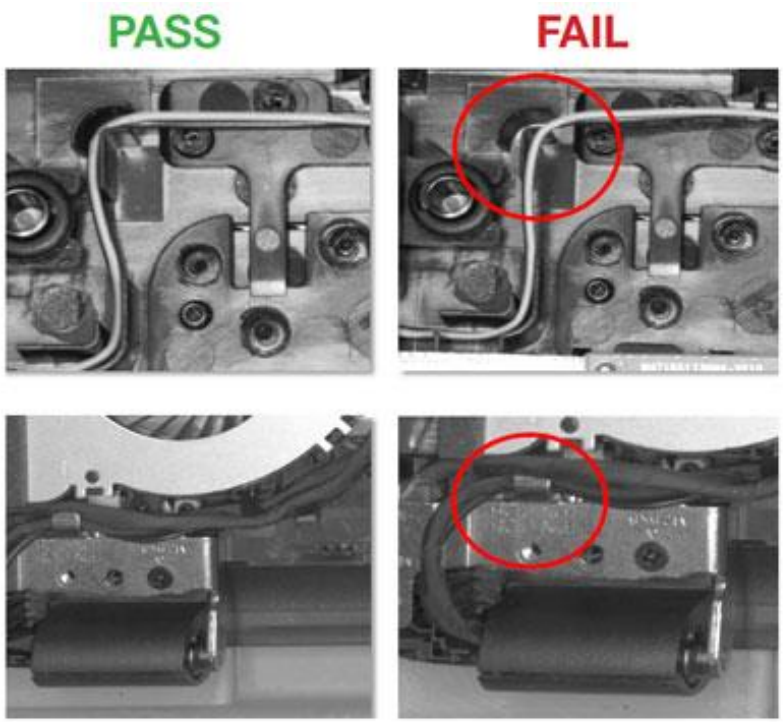
Patient's Heart rate data

QTUM Coin

1. Introduction(Vision Data)



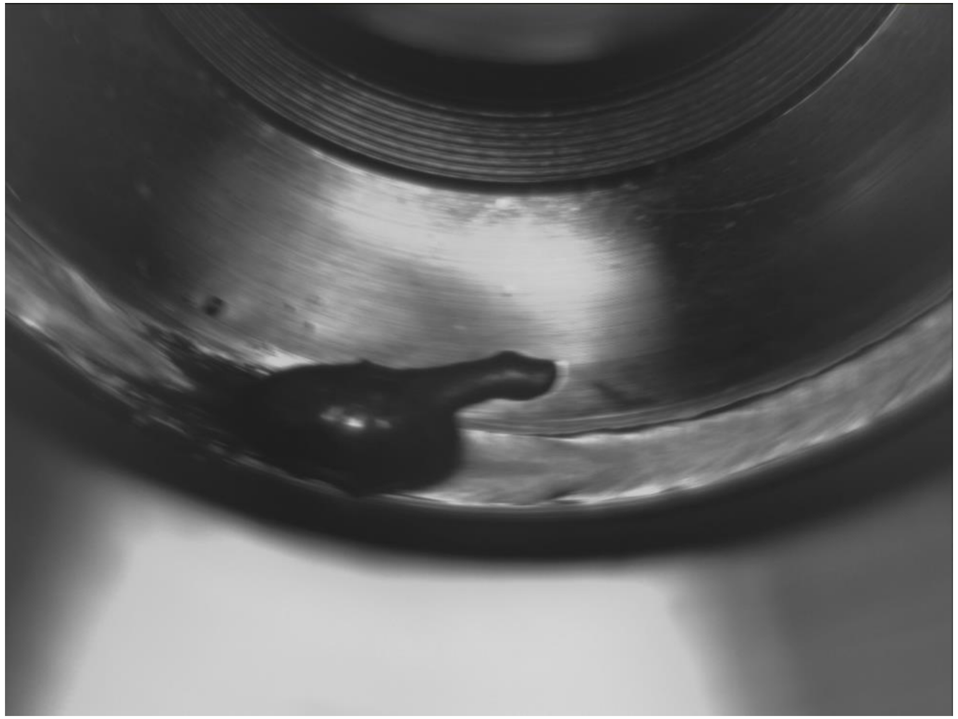
Welding Defect



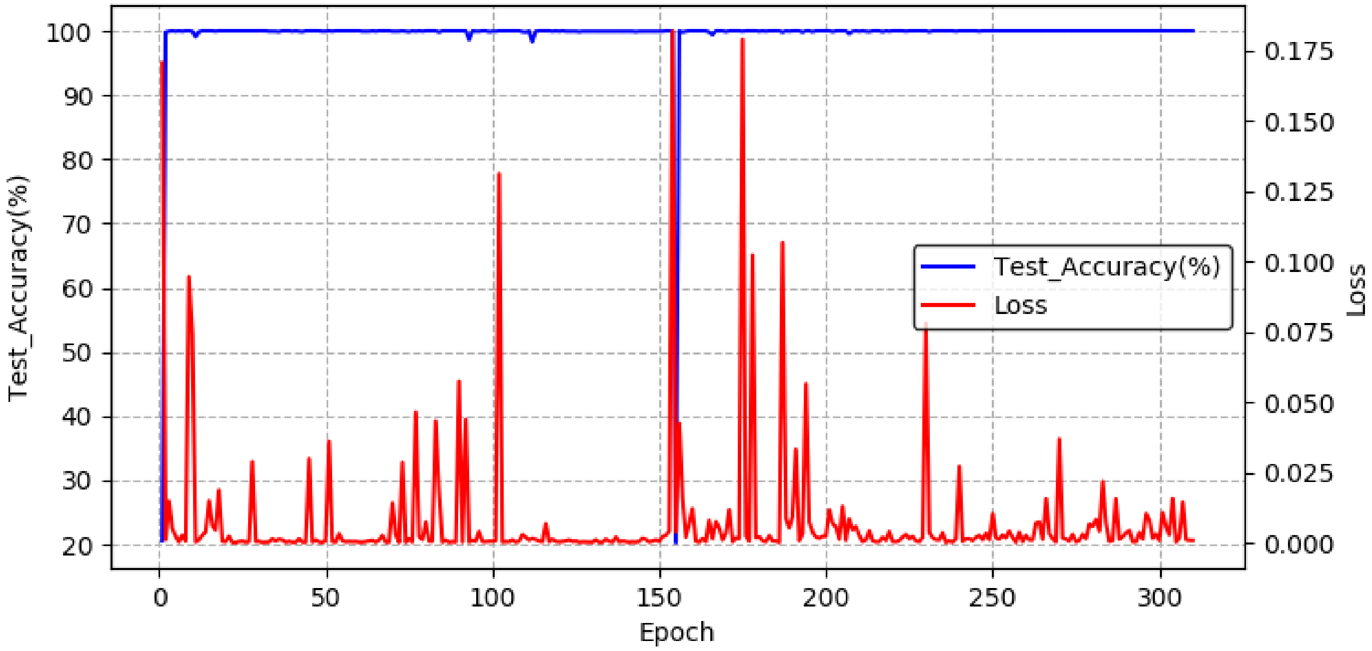
Assembly Inspection

1. Introduction(Vision Data)

SI ANALYTICS



Welding Defect



1. Introduction

Wasserstein GAN

C Proofs of things

Proof of Theorem 1. Let θ and θ' be two parameter vectors in $\mathbb{R}^d$. Then, we will first attempt to bound $W(\mathbb{P}_\theta, \mathbb{P}_{\theta'})$, from where the theorem will come easily. The main element of the proof is the use of the coupling γ, the distribution of the joint $(g_\theta(Z), g_{\theta'}(Z))$, which clearly has $\gamma \in \Pi(\mathbb{P}_\theta, \mathbb{P}_{\theta'})$.

By the definition of the Wasserstein distance, we have

$$\begin{aligned} W(\mathbb{P}_\theta, \mathbb{P}_{\theta'}) &\leq \int_{\mathcal{X}\times\mathcal{X}} \|x - y\| \, \mathrm{d}\gamma \\ &= \mathbb{E}_{(x,y)\sim\gamma}[\|x - y\|] \\ &= \mathbb{E}_z[\|g_\theta(z) - g_{\theta'}(z)\|] \end{aligned}$$

If g is continuous in θ, then $g_\theta(z) \to_{\theta\to\theta'} g_{\theta'}(z)$, so $\|g_\theta - g_{\theta'}\| \to 0$ pointwise as functions of z. Since $\mathcal{X}$ is compact, the distance of any two elements in it has to be uniformly bounded by some constant M, and therefore $\|g_\theta(z) - g_{\theta'}(z)\| \leq M$ for all θ and z uniformly. By the bounded convergence theorem, we therefore have

$$W(\mathbb{P}_\theta, \mathbb{P}_{\theta'}) \leq \mathbb{E}_z[\|g_\theta(z) - g_{\theta'}(z)\|] \to_{\theta\to\theta'} 0$$

Finally, we have that

$$|W(\mathbb{P}_r, \mathbb{P}_\theta) - W(\mathbb{P}_r, \mathbb{P}_{\theta'})| \leq W(\mathbb{P}_\theta, \mathbb{P}_{\theta'}) \to_{\theta\to\theta'} 0$$

proving the continuity of $W(\mathbb{P}_r, \mathbb{P}_\theta)$.

Now let g be locally Lipschitz. Then, for a given pair (θ, z) there is a constant $L(\theta, z)$ and an open set U such that $(\theta, z) \in U$, such that for every $(\theta', z') \in U$ we have

$$\|g_\theta(z) - g'_\theta(z')\| \leq L(\theta, z)(\|\theta - \theta'\| + \|z - z'\|)$$

VAE

$$\tilde{\mathbf{z}} = g_\phi(\epsilon, \mathbf{x}) \quad \text{with} \quad \epsilon \sim p(\epsilon) \tag{17}$$

where we choose a prior $p(\epsilon)$ and a function $g_\phi(\epsilon, \mathbf{x})$ such that the following holds:

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}, \phi; \mathbf{x}^{(i)}) &= \int q_\phi(\mathbf{z}|\mathbf{x}) \left(\log p_{\boldsymbol{\theta}}(\mathbf{x}^{(i)}|\mathbf{z}) + \log p_{\boldsymbol{\theta}}(\mathbf{z}) - \log q_\phi(\mathbf{z}|\mathbf{x})\right) dz \\ &= \int p(\epsilon) \left(\log p_{\boldsymbol{\theta}}(\mathbf{x}^{(i)}|\mathbf{z}) + \log p_{\boldsymbol{\theta}}(\mathbf{z}) - \log q_\phi(\mathbf{z}|\mathbf{x})\right) \Big|_{\mathbf{z}=g_\phi(\epsilon, \mathbf{x}^{(i)})} d\epsilon \end{aligned} \tag{18}$$

The same can be done for the approximate posterior $q_\phi(\boldsymbol{\theta})$:

$$\tilde{\boldsymbol{\theta}} = h_\phi(\zeta) \quad \text{with} \quad \zeta \sim p(\zeta) \tag{19}$$

where we, similarly as above, choose a prior $p(\zeta)$ and a function $h_\phi(\zeta)$ such that the following holds:

$$\begin{aligned} \mathcal{L}(\phi; \mathbf{X}) &= \int q_\phi(\boldsymbol{\theta}) \left(\log p_{\boldsymbol{\theta}}(\mathbf{X}) + \log p_{\boldsymbol{\alpha}}(\boldsymbol{\theta}) - \log q_\phi(\boldsymbol{\theta})\right) d\boldsymbol{\theta} \\ &= \int p(\zeta) \left(\log p_{\boldsymbol{\theta}}(\mathbf{X}) + \log p_{\boldsymbol{\alpha}}(\boldsymbol{\theta}) - \log q_\phi(\boldsymbol{\theta})\right) \Big|_{\boldsymbol{\theta}=h_\phi(\zeta)} d\zeta \end{aligned} \tag{20}$$

For notational conciseness we introduce a shorthand notation $f_\phi(\mathbf{x}, \mathbf{z}, \boldsymbol{\theta})$:

$$f_\phi(\mathbf{x}, \mathbf{z}, \boldsymbol{\theta}) = N \cdot (\log p_{\boldsymbol{\theta}}(\mathbf{x}|\mathbf{z}) + \log p_{\boldsymbol{\theta}}(\mathbf{z}) - \log q_\phi(\mathbf{z}|\mathbf{x})) + \log p_{\boldsymbol{\alpha}}(\boldsymbol{\theta}) - \log q_\phi(\boldsymbol{\theta}) \tag{21}$$

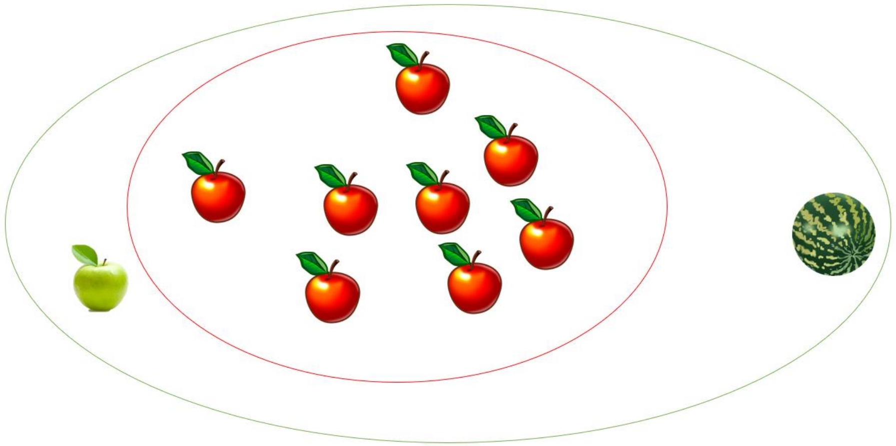
Using equations (20) and (18), the Monte Carlo estimate of the variational lower bound, given datapoint $\mathbf{x}^{(i)}$, is:

$$\mathcal{L}(\phi; \mathbf{X}) \simeq \frac{1}{L} \sum_{l=1}^{L} f_\phi(\mathbf{x}^{(l)}, g_\phi(\epsilon^{(l)}, \mathbf{x}^{(l)}), h_\phi(\zeta^{(l)})) \tag{22}$$

Conventional Anomaly Detection

(Vision data only)

2. Conventional Anomaly Detection



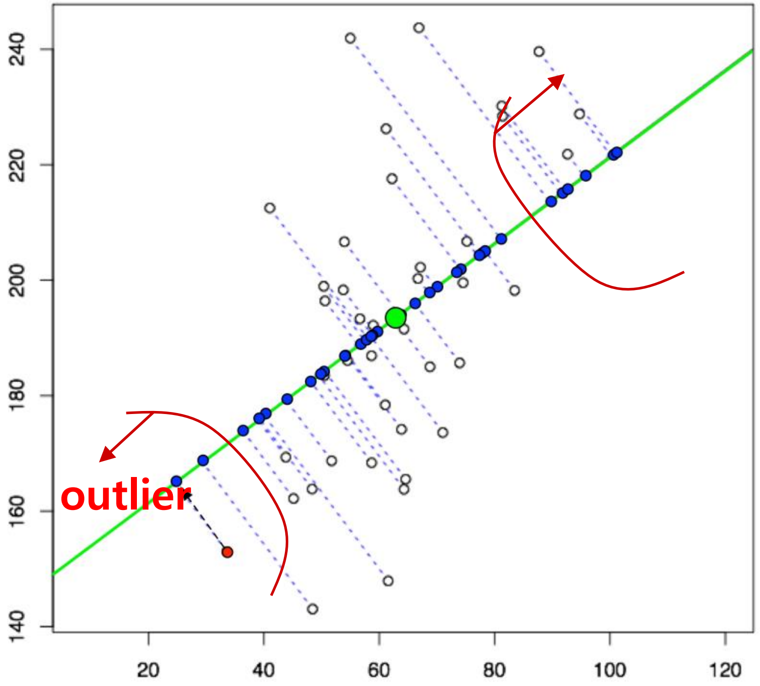
Before dealing with Anomaly Detection, It is essential to identify the definition of the problem.

→ **Domain Knowledge**

2. Conventional Anomaly Detection

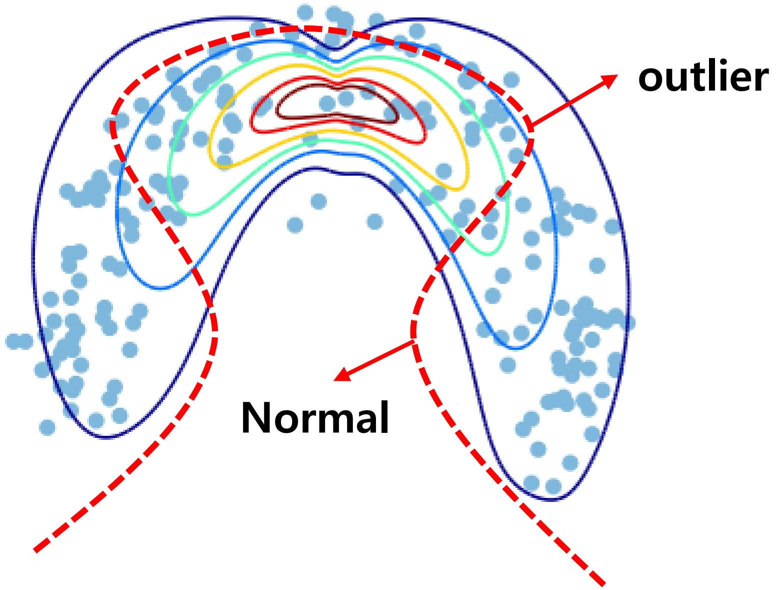
SI ANALYTICS

PCA



2. Conventional Anomaly Detection

KDE



$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(X - X_i) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{X - X_i}{h})$$

https://www.tableau.com/resource/top-10-big-data-trends-2017

2. Conventional Anomaly Detection

SI ANALYTICS

Support Vector Data Description(SVDD)

https://link.springer.com/article/10.1023/B:MACH.0000008084.60811.49

Isolation forests(IF)

https://cs.nju.edu.cn/zhouzh/zhouzh.files/publication/icdm08b.pdf

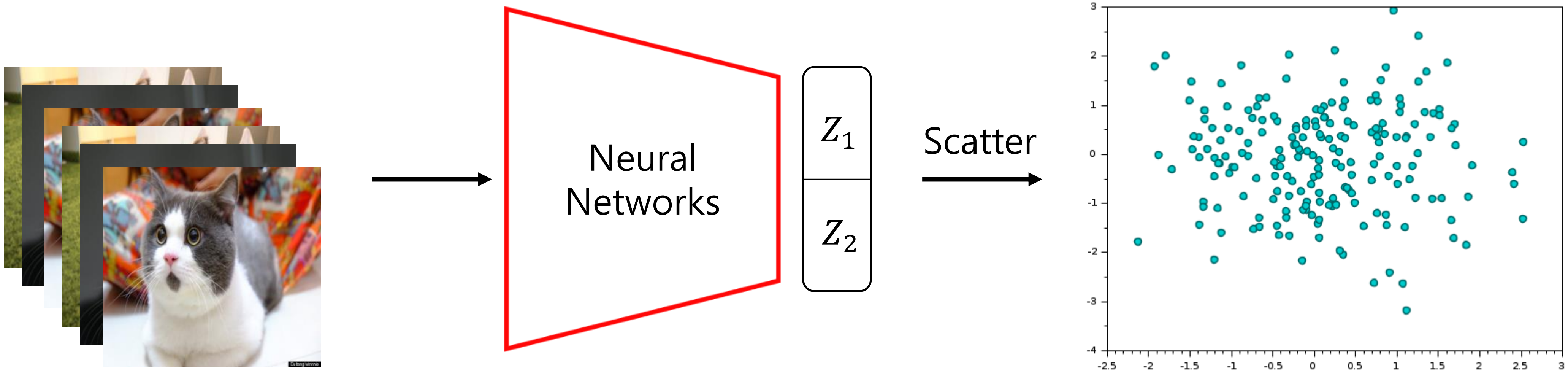
https://en.wikipedia.org/wiki/Anomaly_detection#cite_note-12

- Density-based techniques (k-nearest neighbor,[8][9][10] local outlier factor,[11] isolation forests,[12] and many more variations of this concept[13]).
- Subspace-[14] and correlation-based[15] outlier detection for high-dimensional data.[16]
- One-class support vector machines.[17]
- Replicator neural networks.[18]
- Bayesian Networks.[18]
- Hidden Markov models (HMMs).[18]
- Cluster analysis-based outlier detection.[19][20]
- Deviations from association rules and frequent itemsets.
- Fuzzy logic-based outlier detection.
- Ensemble techniques, using feature bagging,[21][22] score normalization[23][24] and different sources of diversity.[25][26]

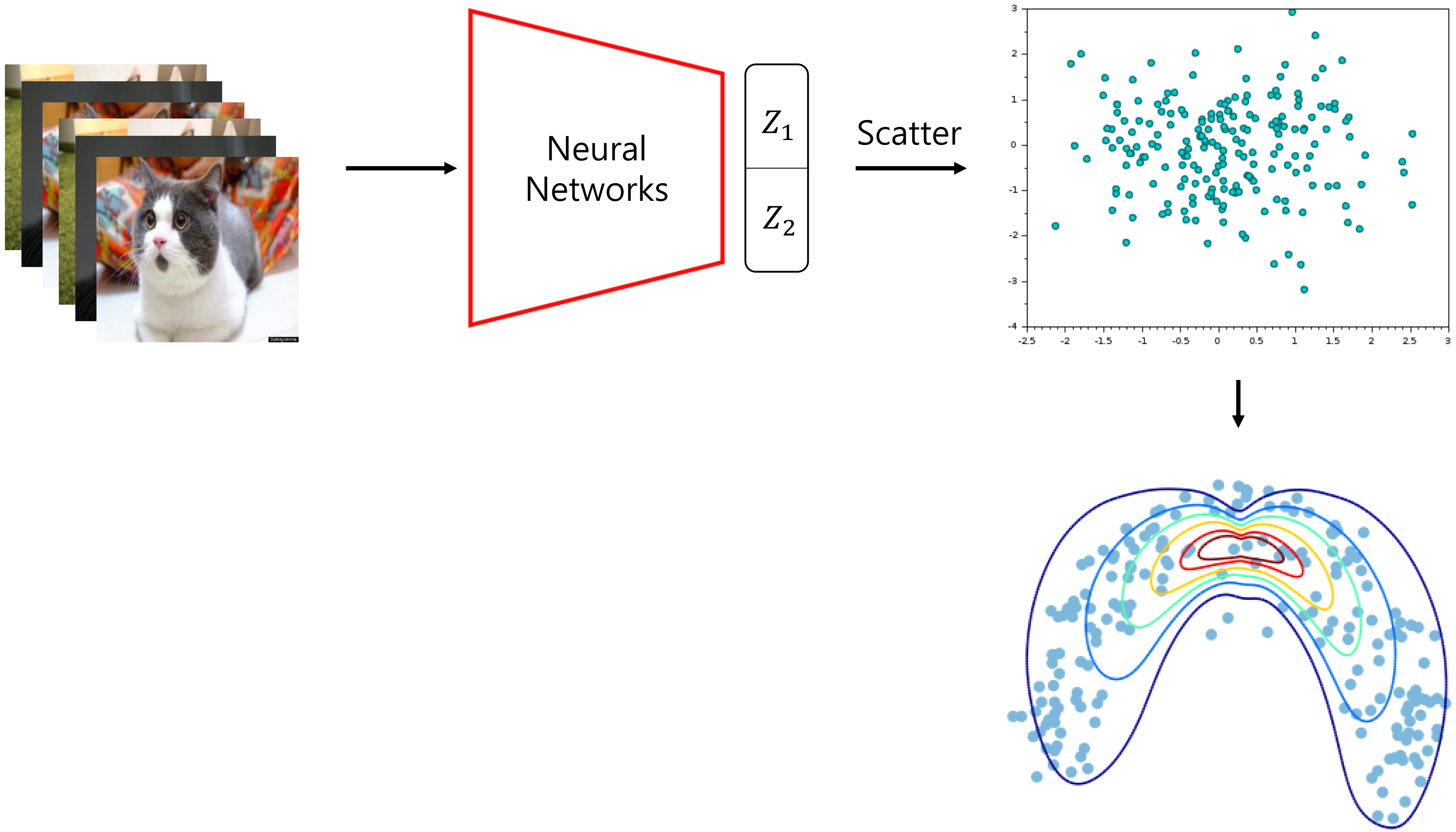
Deep Anomaly Detection

(Vision data only)

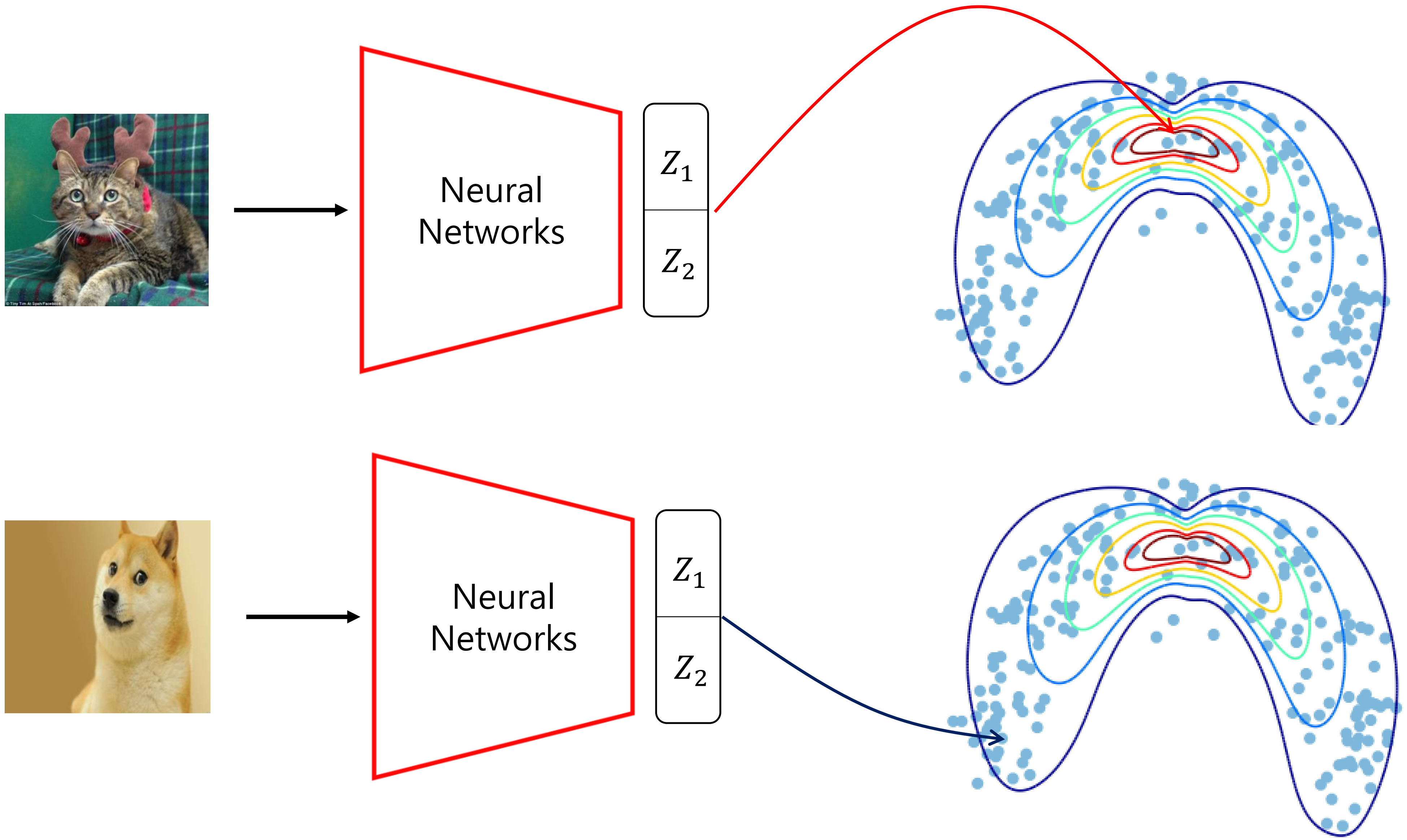
3. Deep Anomaly Detection(Representation Learning)



3. Deep Anomaly Detection(Representation Learning)



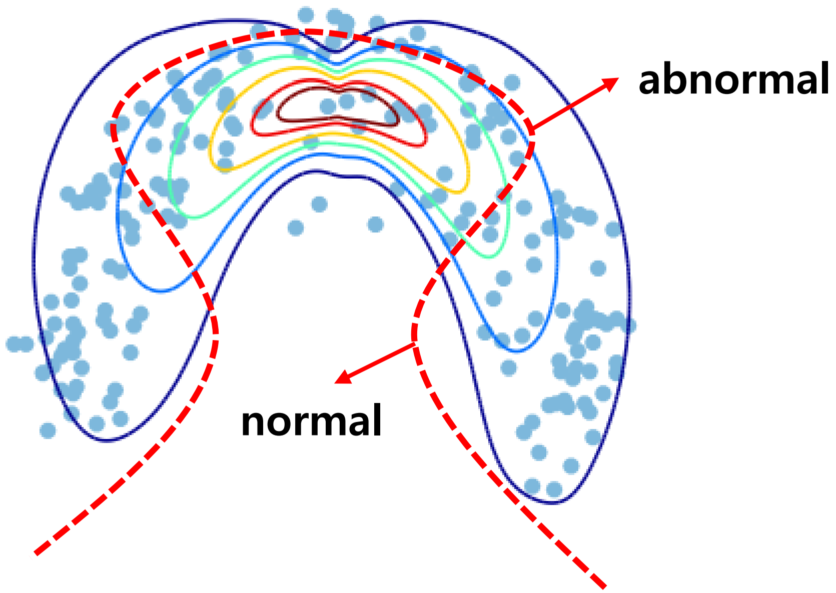
3. Deep Anomaly Detection(Representation Learning)



3. Deep Anomaly Detection(Representation Learning)

SI ANALYTICS

① Non linear Classifier



Deep One-Class Classification

http://proceedings.mlr.press/v80/ruff18a.html

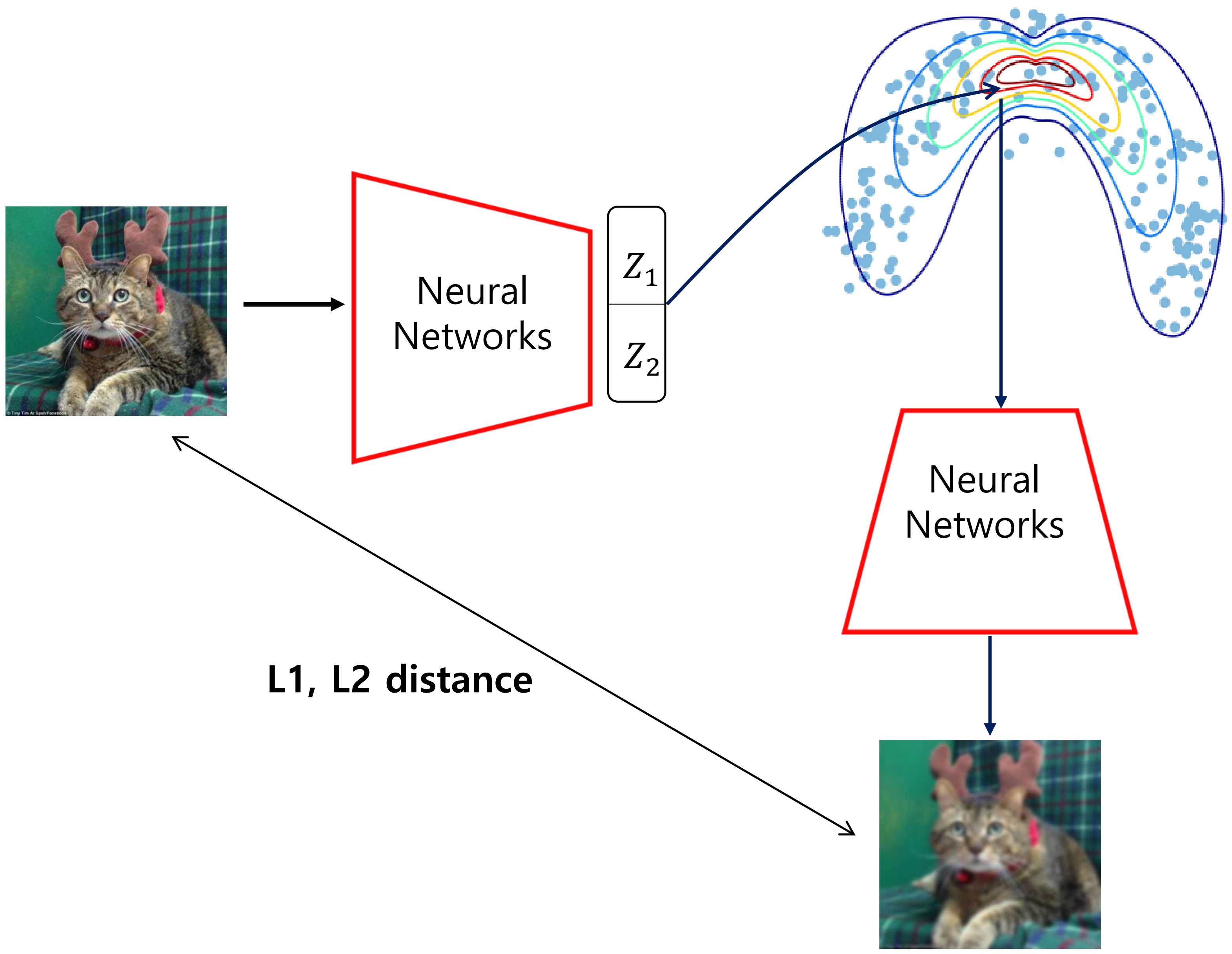
DCAE(Deep Convolutional Autoencoder)

Deep Anomaly Detection Using Geometric Transformation

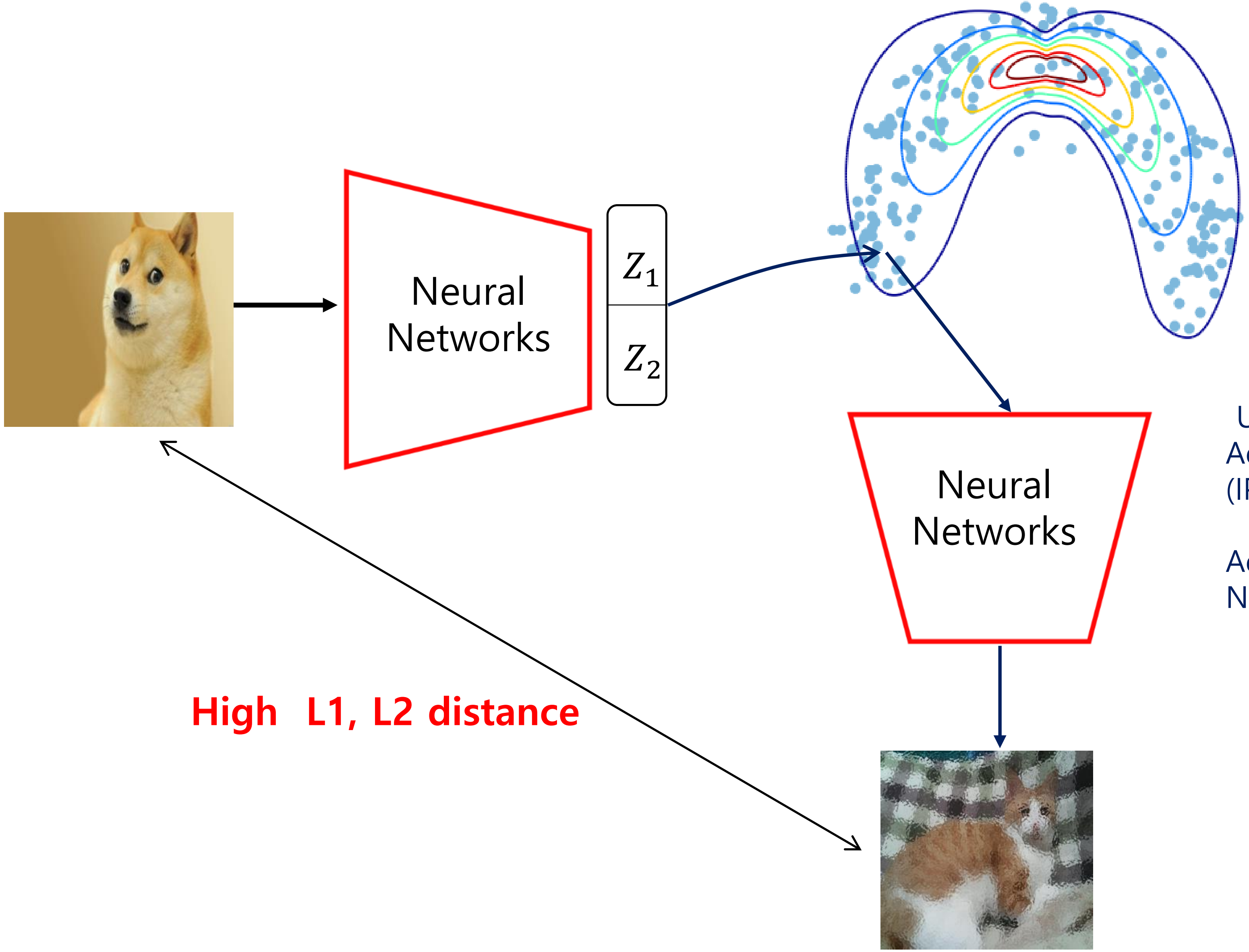
https://arxiv.org/pdf/1805.10917.pdf

3. Deep Anomaly Detection(Reconstruction Error)

SI ANALYTICS



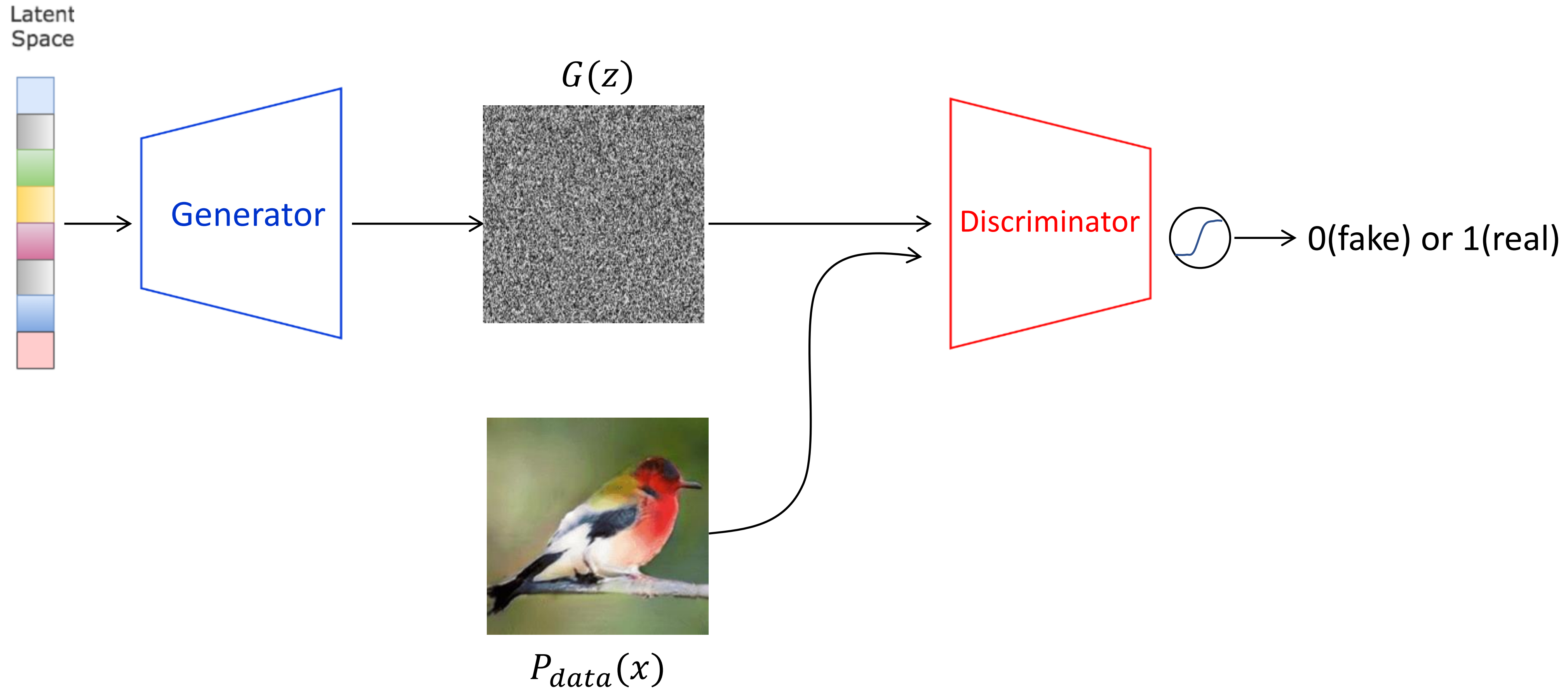
3. Deep Anomaly Detection(Reconstruction Error)



Unsupervsed Anomaly Detection with Generative Adversarial Networks to Guide marker discovery (IPML, 2017)

Adversarially Learned One-Class Classifier for Novelty Detection(CVPR, 2018)

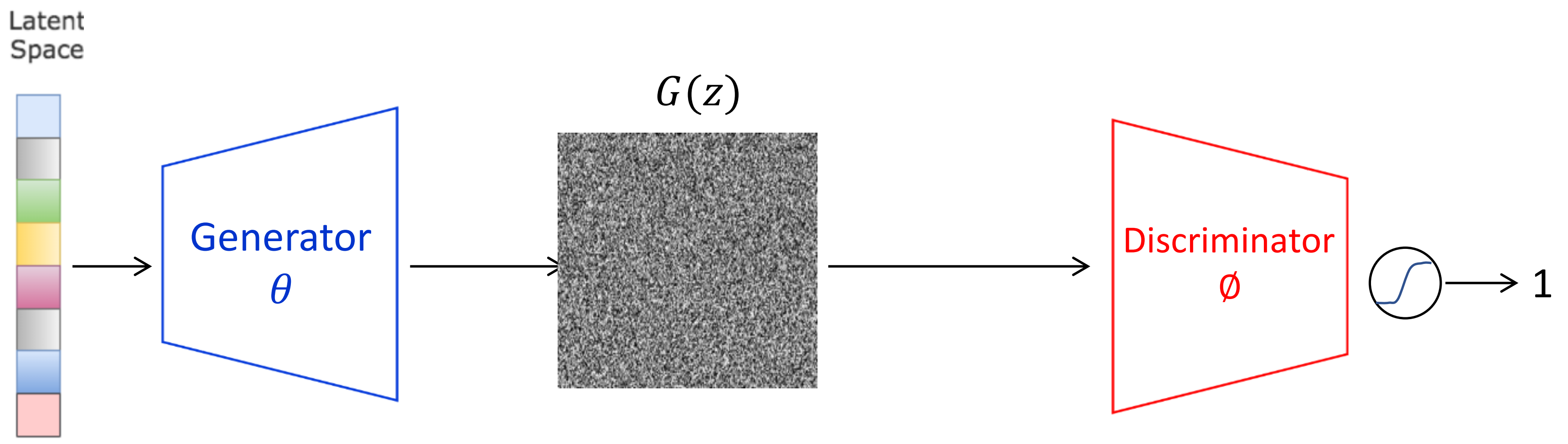
3. Deep Anomaly Detection(GANs)



SI ANALYTICS

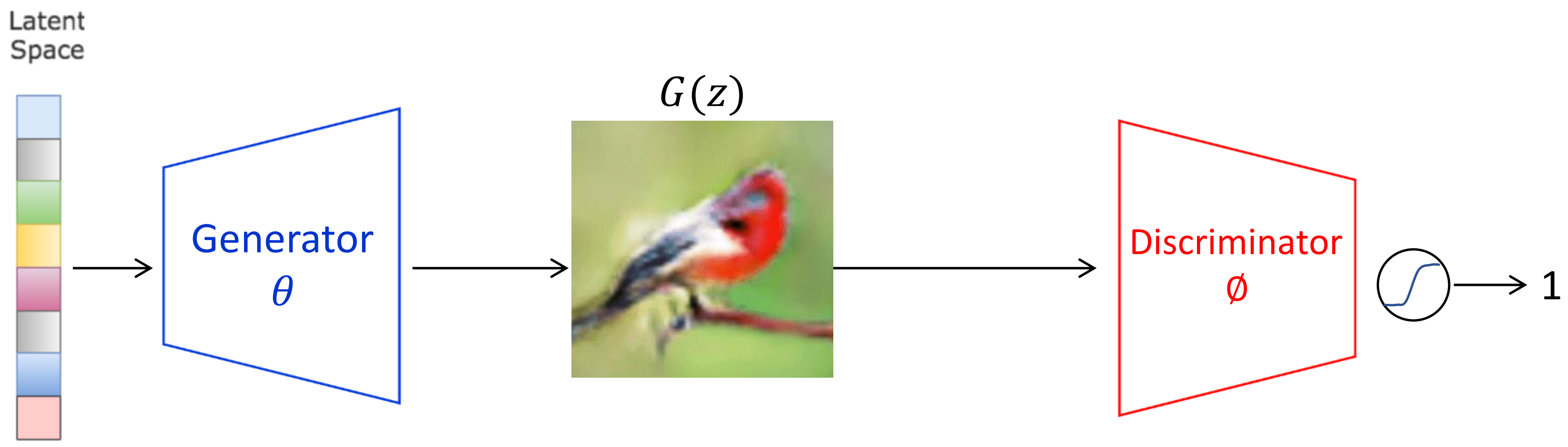
Generator

3. Deep Anomaly Detection(GANs)



$$G_{loss} = \ argmin_{\theta}\mathbb{E}_{z\sim p(z)}[\log(1 - D(G(z)))]$$

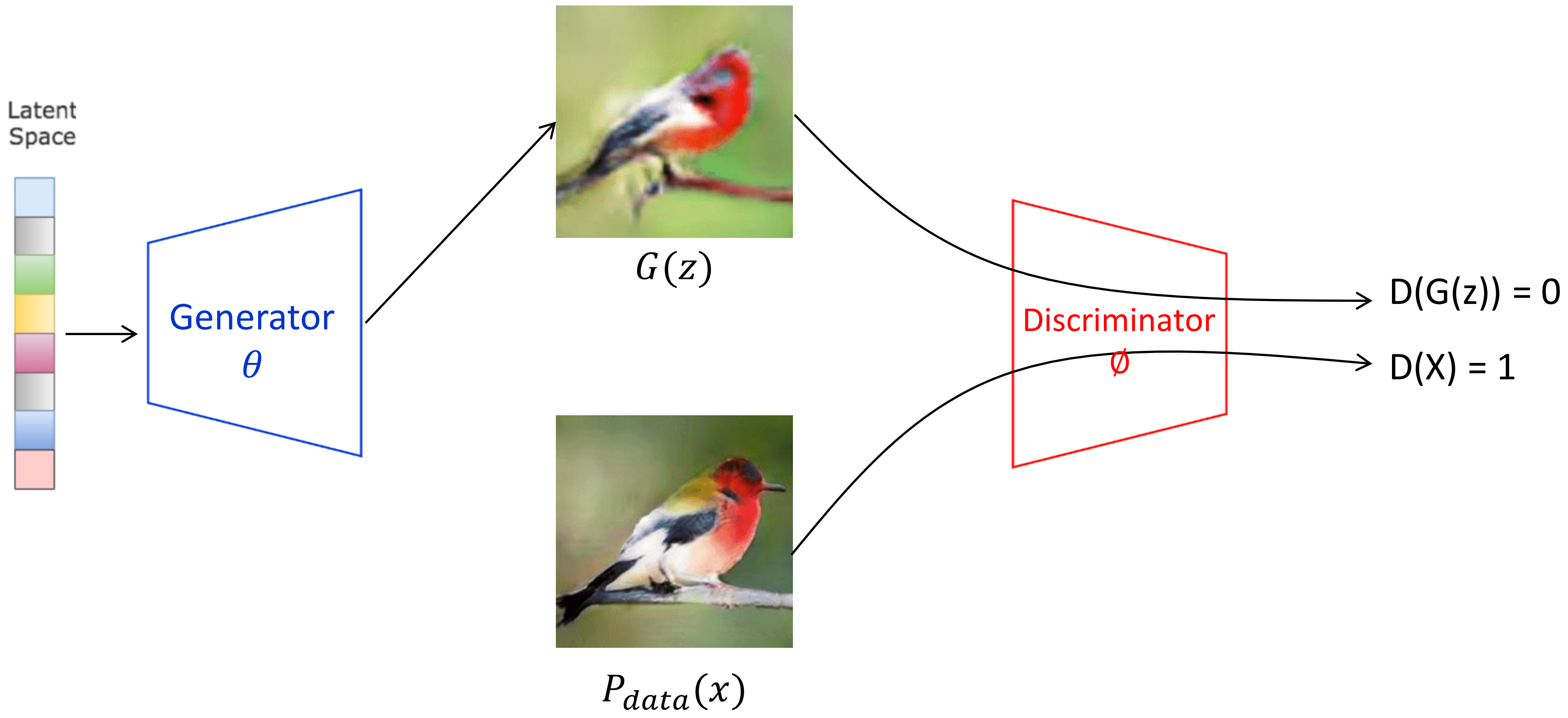
3. Deep Anomaly Detection(GANs)



$$G_{loss} = \ argmin_{\theta}\mathbb{E}_{z\sim p(z)}[\log(1 - D(G(z)))]$$

Discriminator

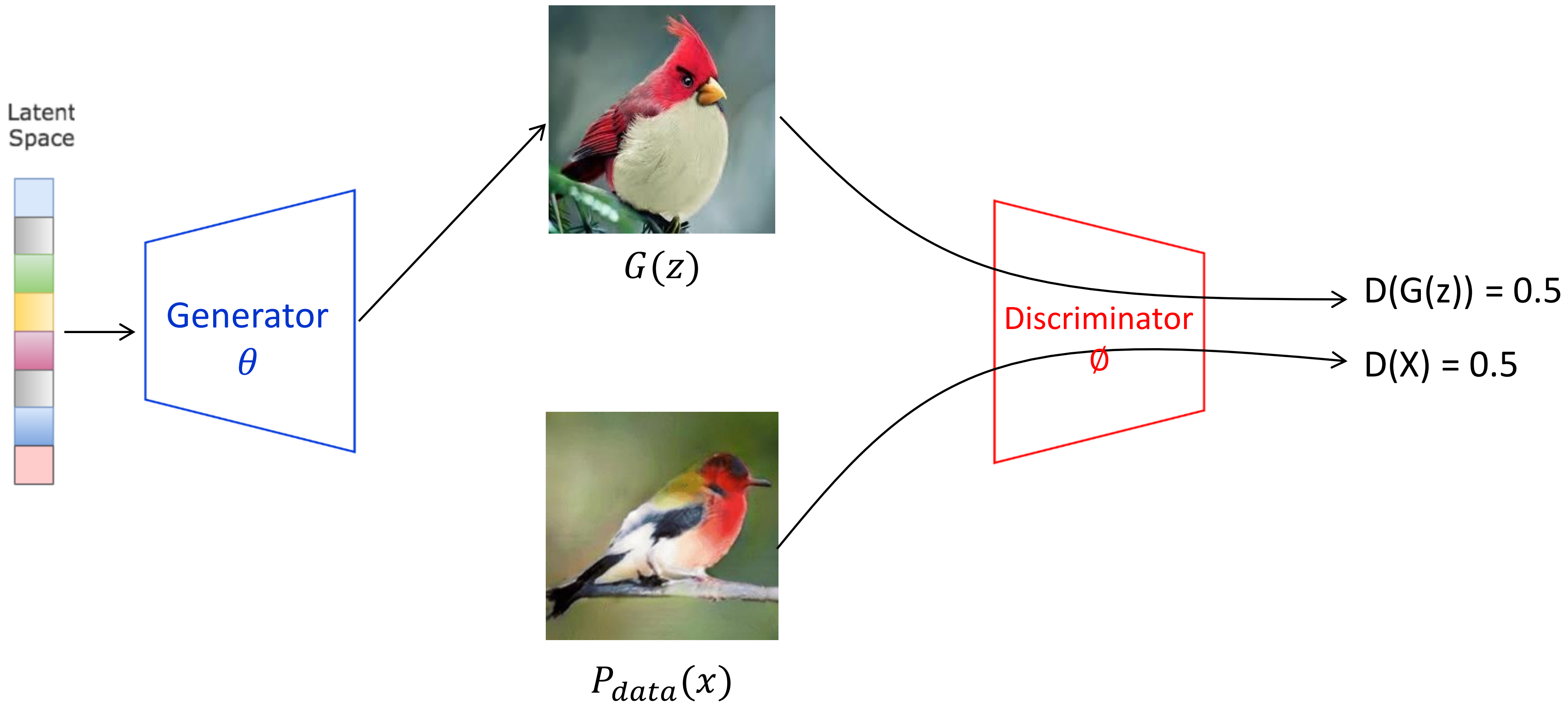
3. Deep Anomaly Detection(GANs)



$$D_{loss} = argmax_{\emptyset}\mathbb{E}_{x\sim p_{data}}[logD(x)] + \mathbb{E}_{z\sim p(z)}[\log(1-D(G(x)))]$$

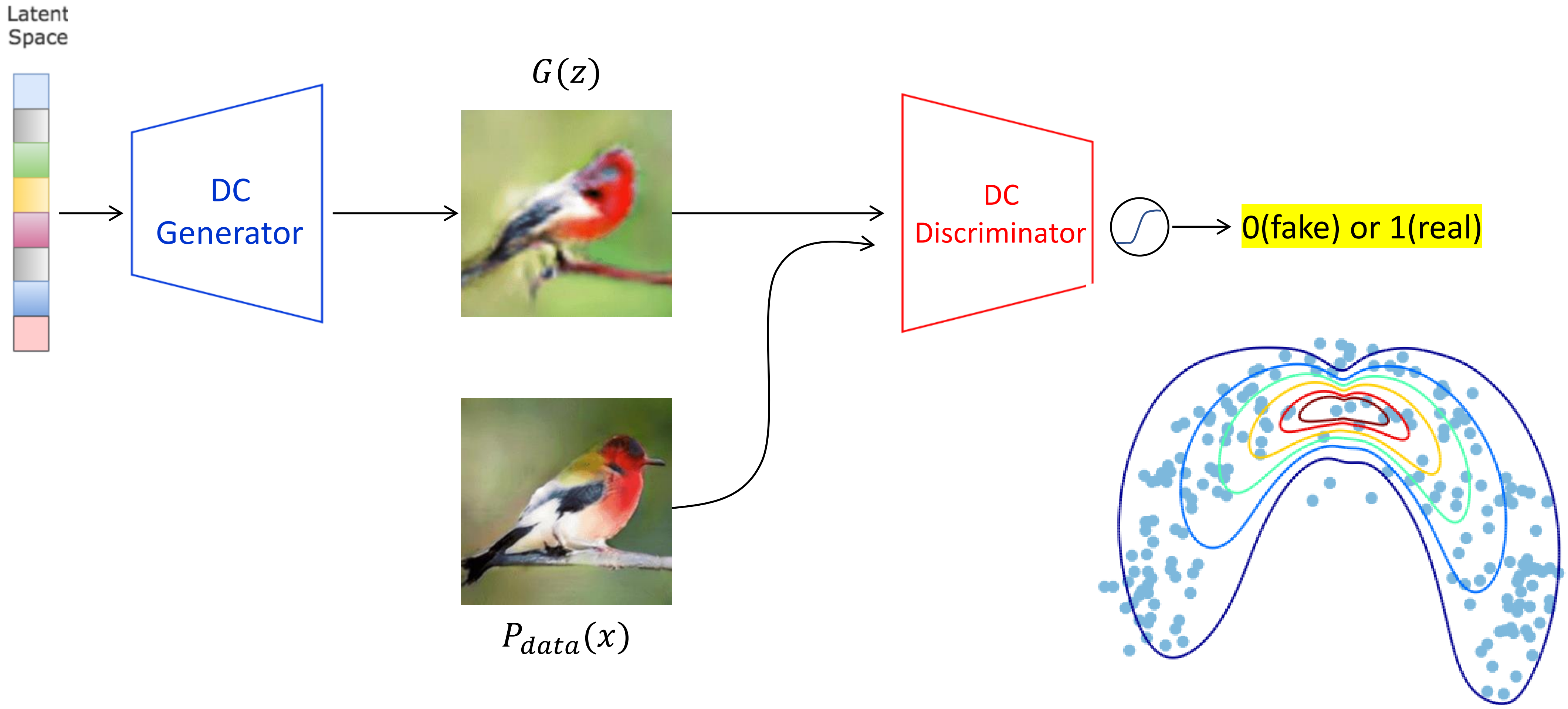
Repeat Over and Over

3. Deep Anomaly Detection(GANs)



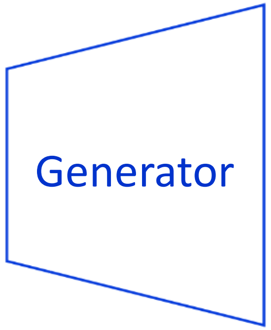
3. Deep Anomaly Detection(AnoGAN)

Unsupervsed Anomaly Detection with Generative Adversarial Networks to Guide marker discovery (IPML, 2017)



3. Deep Anomaly Detection(AnoGAN)

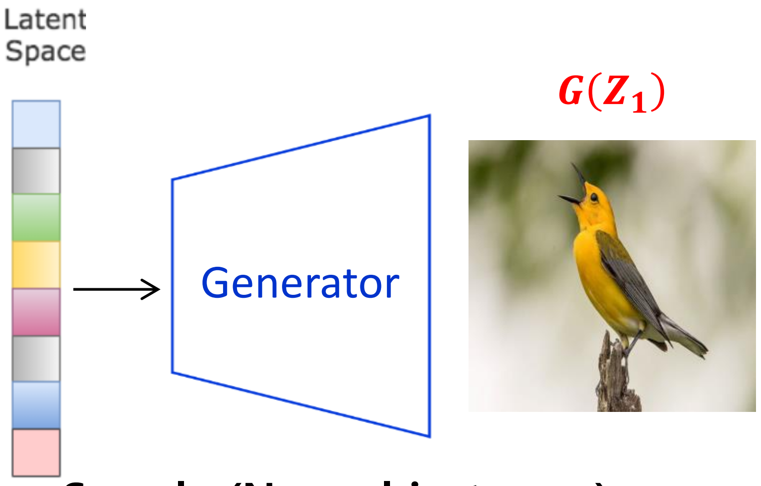
Unsupervsed Anomaly Detection with Generative Adversarial Networks to Guide marker discovery (IPML, 2017)



Sampler(Normal instances)

3. Deep Anomaly Detection(AnoGAN)

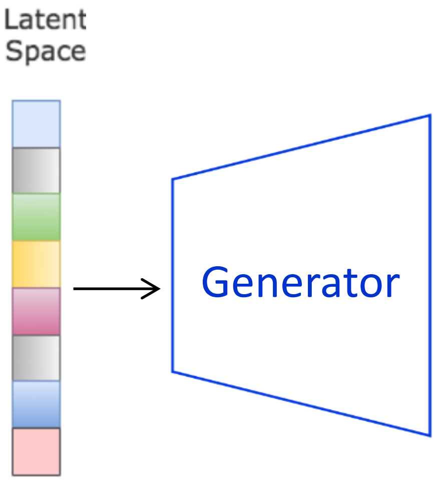
Unsupervsed Anomaly Detection with Generative Adversarial Networks to Guide marker discovery (IPML, 2017)



Sampler(Normal instances)

3. Deep Anomaly Detection(AnoGAN)

Unsupervsed Anomaly Detection with Generative Adversarial Networks to Guide marker discovery (IPML, 2017)



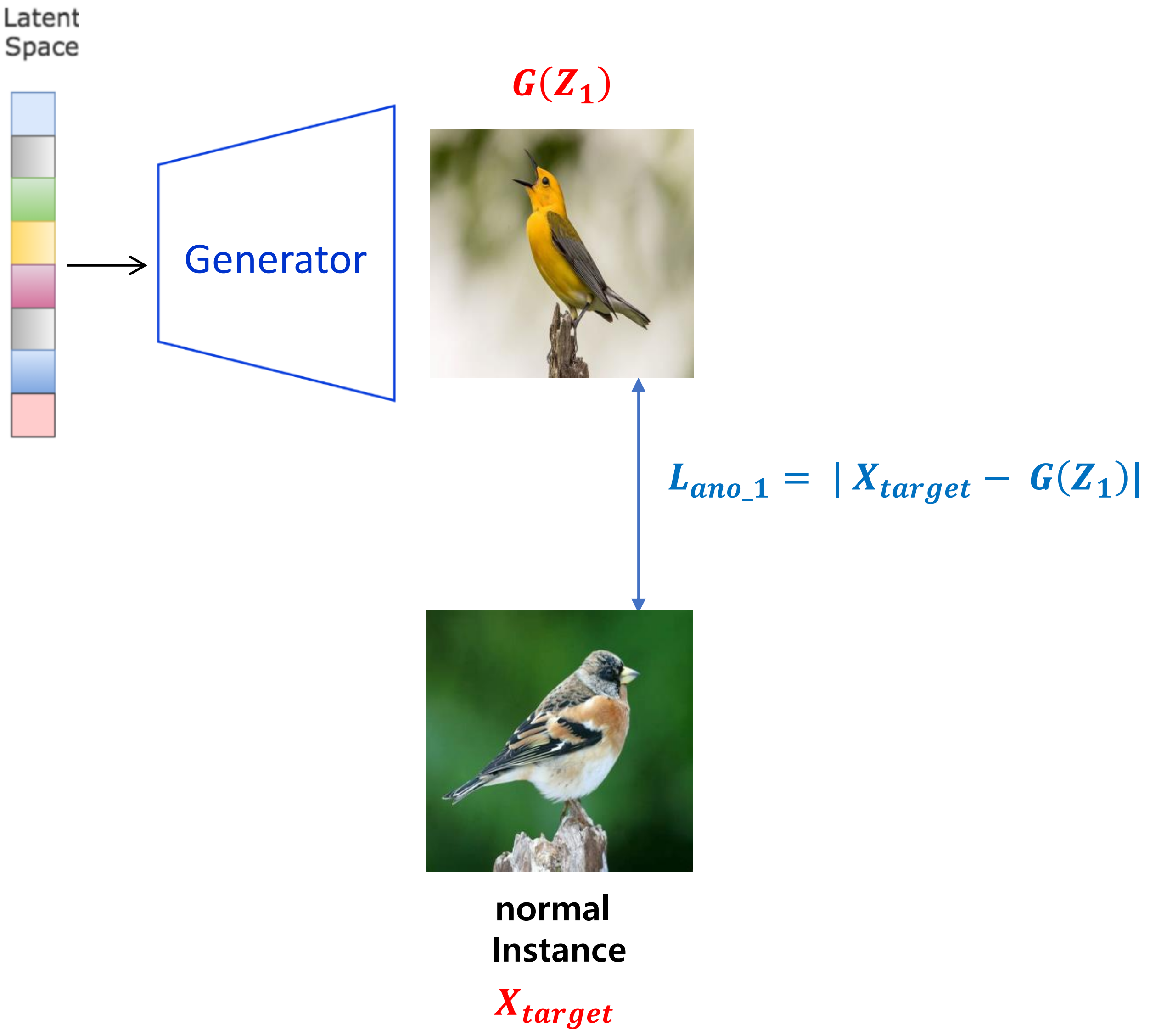
$G(Z_1)$

normal Instance

X_{target}

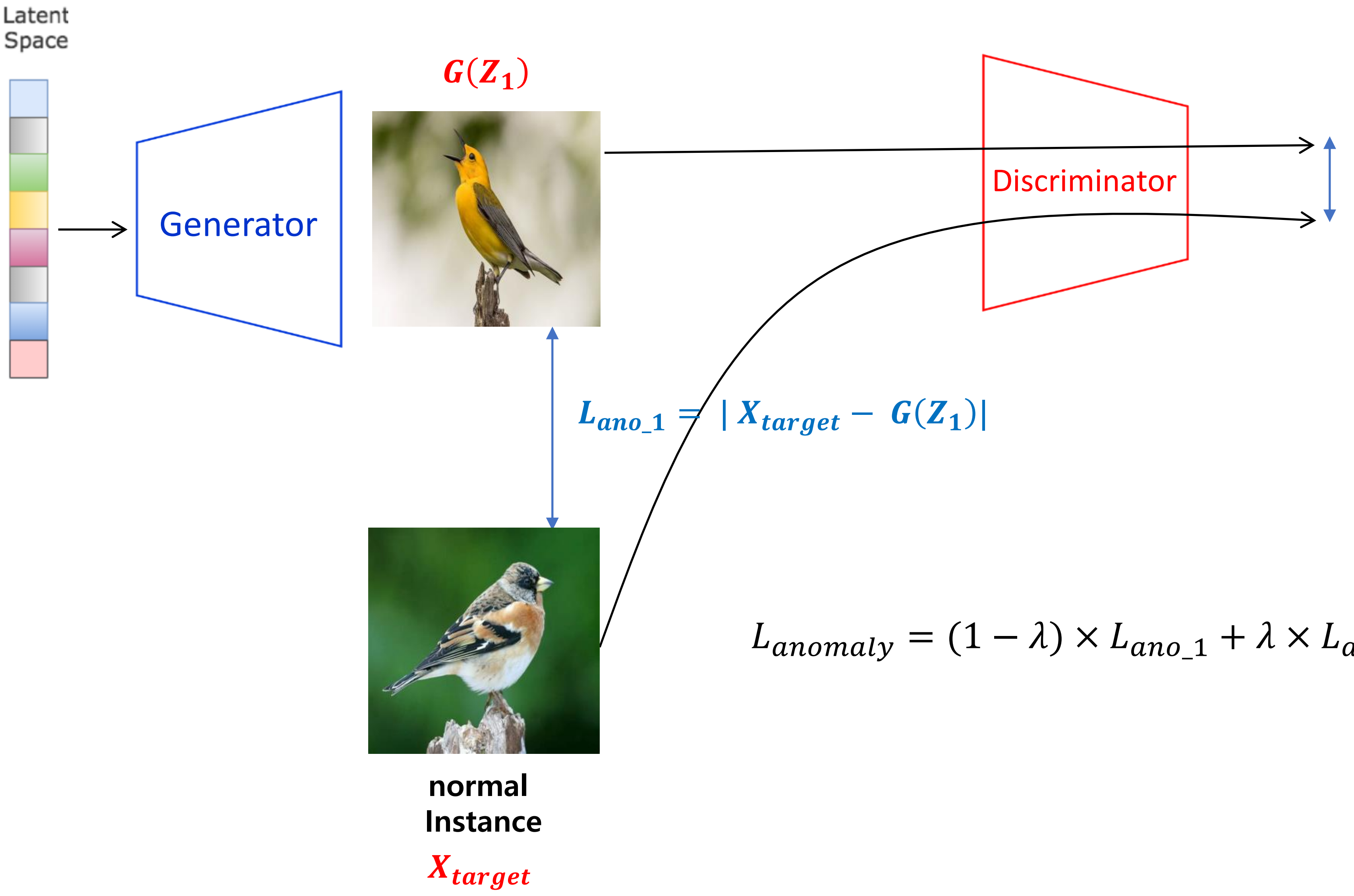
3. Deep Anomaly Detection(AnoGAN)

Unsupervsed Anomaly Detection with Generative Adversarial Networks to Guide marker discovery (IPML, 2017)



3. Deep Anomaly Detection(AnoGAN)

Unsupervsed Anomaly Detection with Generative Adversarial Networks to Guide marker discovery (IPML, 2017)



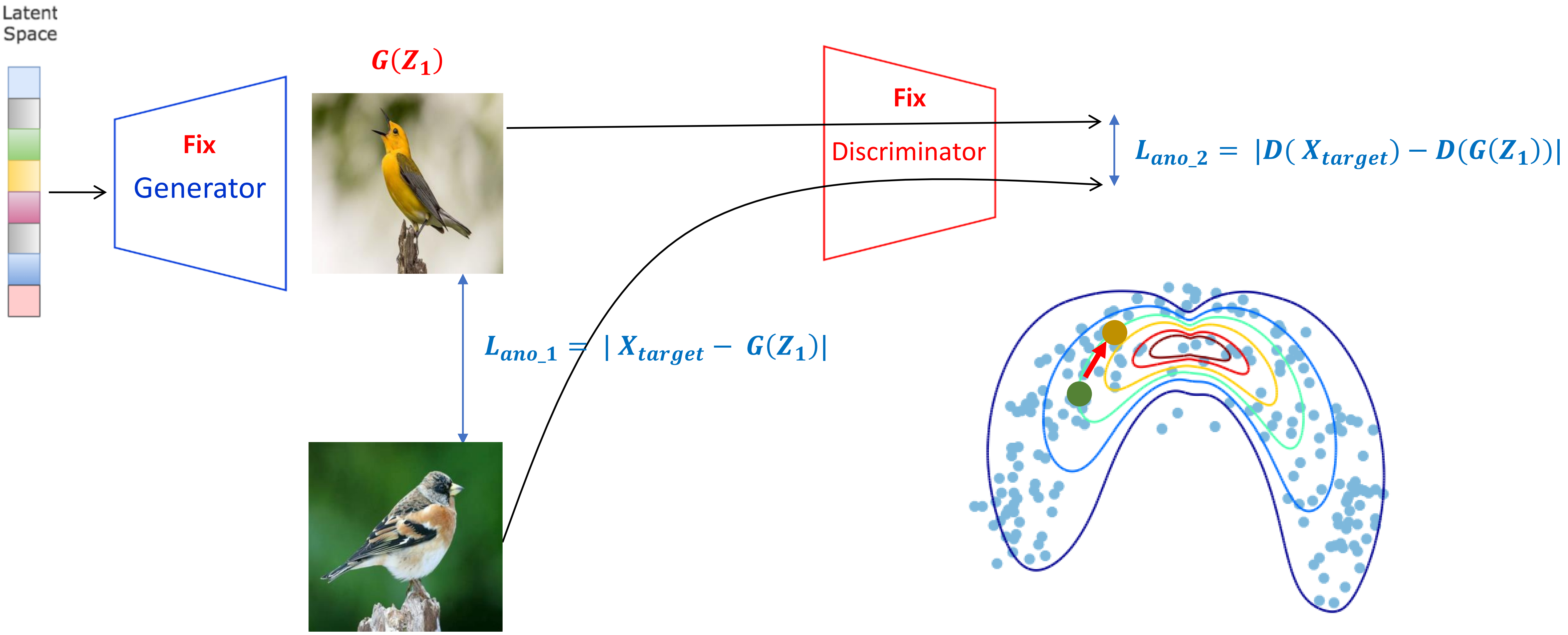
$$L_{ano\_1} = |\, X_{target} - G(Z_1)|$$

$$L_{ano\_2} = |D(\, X_{target}) - D(G(Z_1))|$$

$$L_{anomaly} = (1-\lambda) \times L_{ano\_1} + \lambda \times L_{ano\_2}$$

3. Deep Anomaly Detection(AnoGAN)

Unsupervsed Anomaly Detection with Generative Adversarial Networks to Guide marker discovery (IPML, 2017)

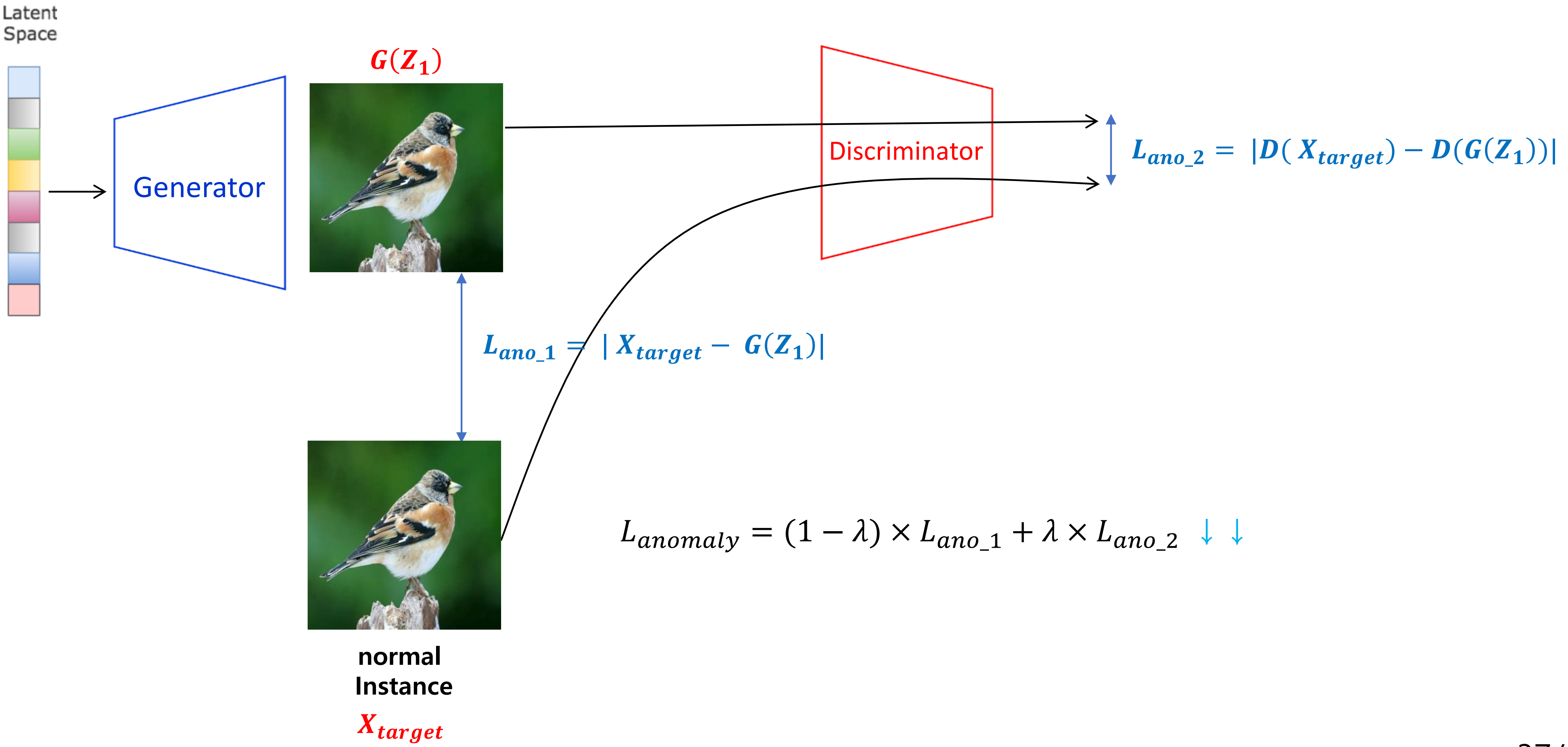


Update latent vector to reduce $L_{anomaly}$!

$$L_{anomaly} = (1-\lambda) \times L_{ano\_1} + \lambda \times L_{ano\_2}$$

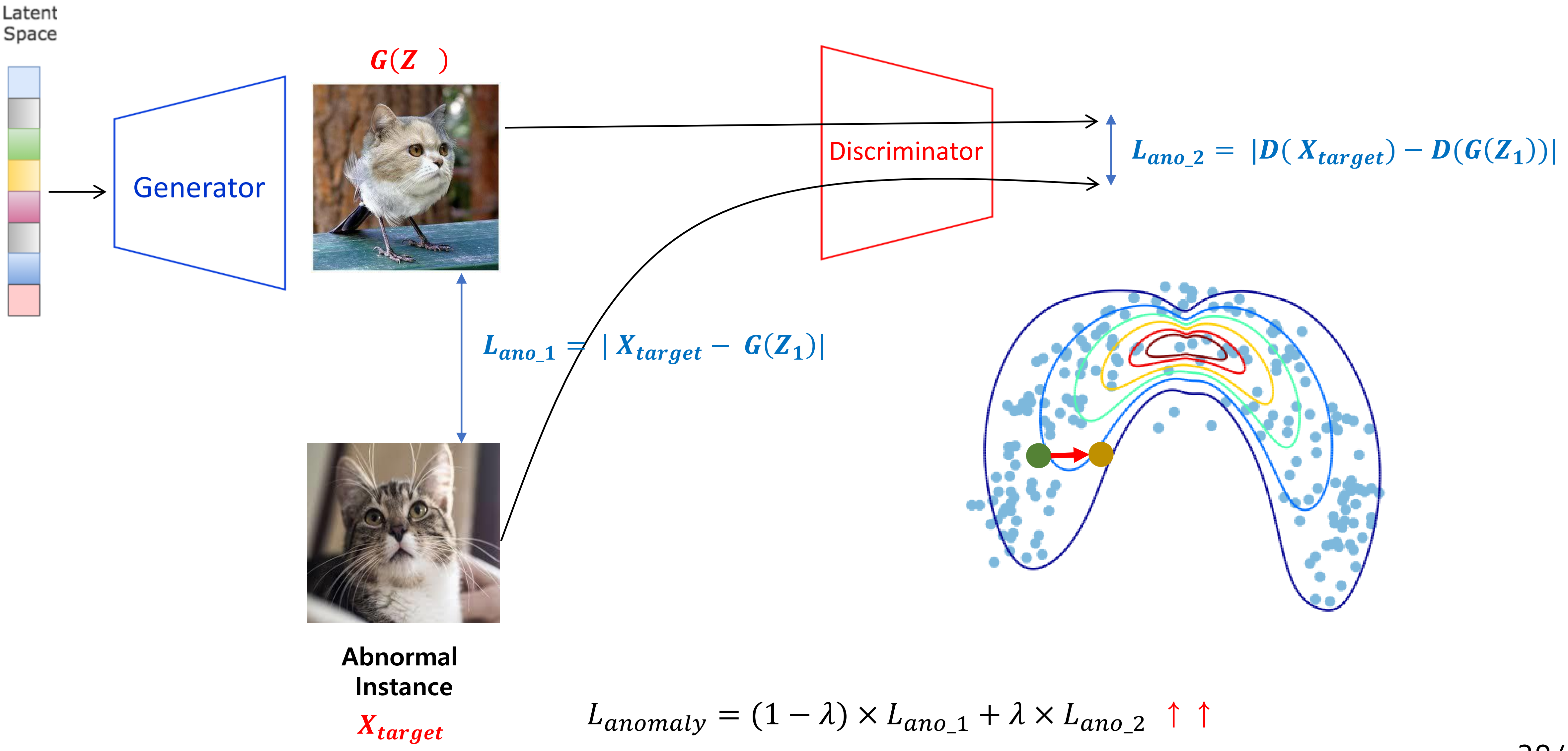
3. Deep Anomaly Detection(AnoGAN)

Unsupervsed Anomaly Detection with Generative Adversarial Networks to Guide marker discovery (IPML, 2017)



3. Deep Anomaly Detection(AnoGAN)

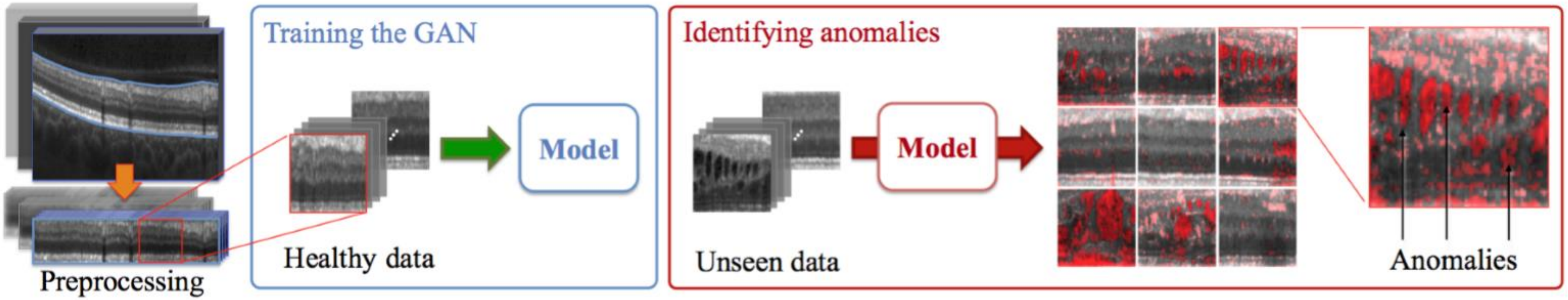
Unsupervsed Anomaly Detection with Generative Adversarial Networks to Guide marker discovery (IPML, 2017)

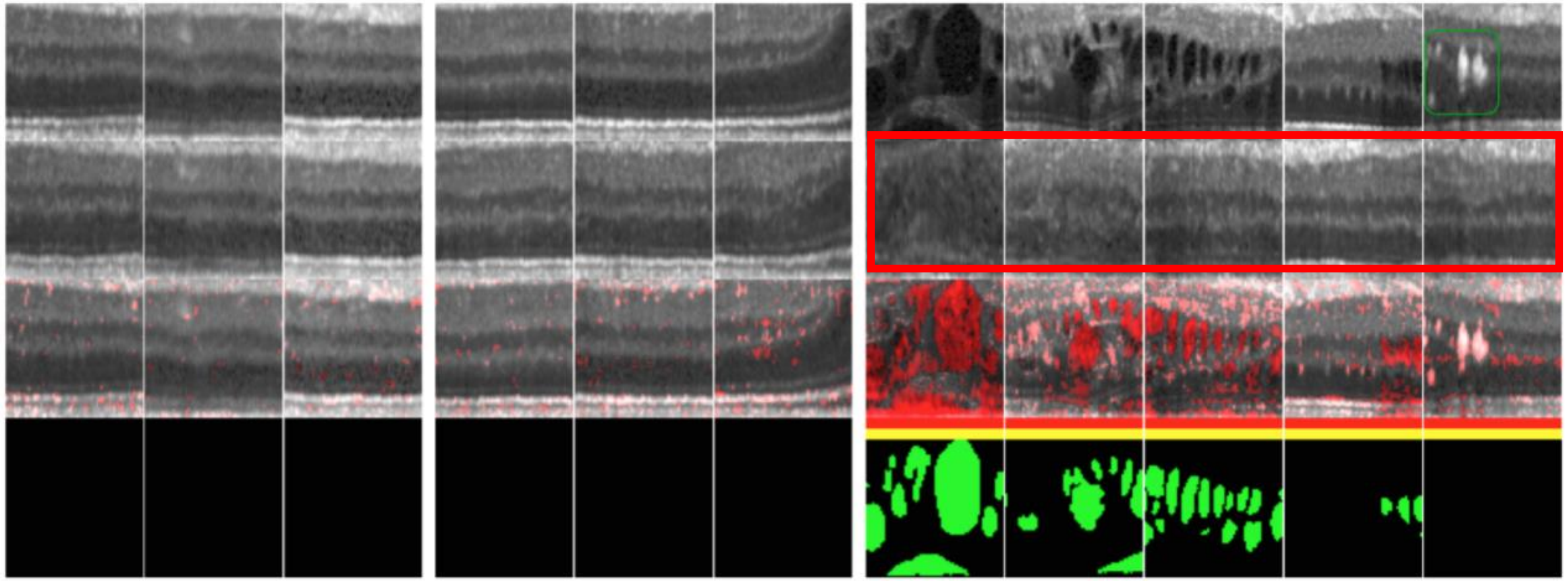


$$L_{anomaly} = (1-\lambda) \times L_{ano\_1} + \lambda \times L_{ano\_2} \uparrow \uparrow$$

3. Deep Anomaly Detection(AnoGAN)

Unsupervsed Anomaly Detection with Generative Adversarial Networks to Guide marker discovery (IPML, 2017)



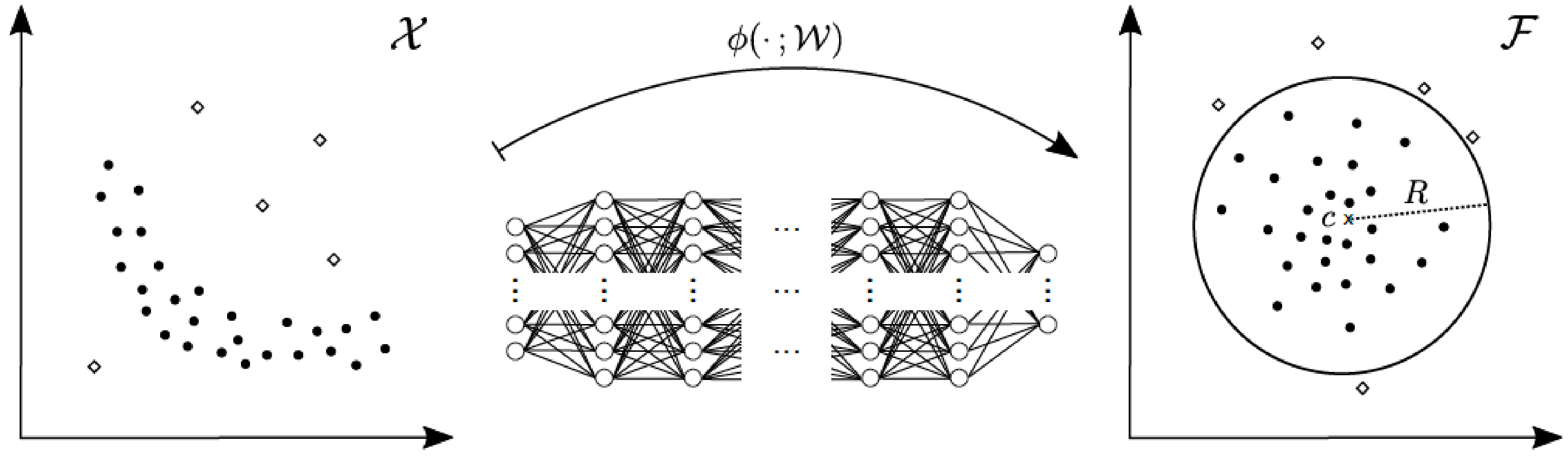


Generated Images are similar to the normal images.

3. Deep Anomaly Detection(DSVDD)

Deep One-Class Classification(ICML, 2018)

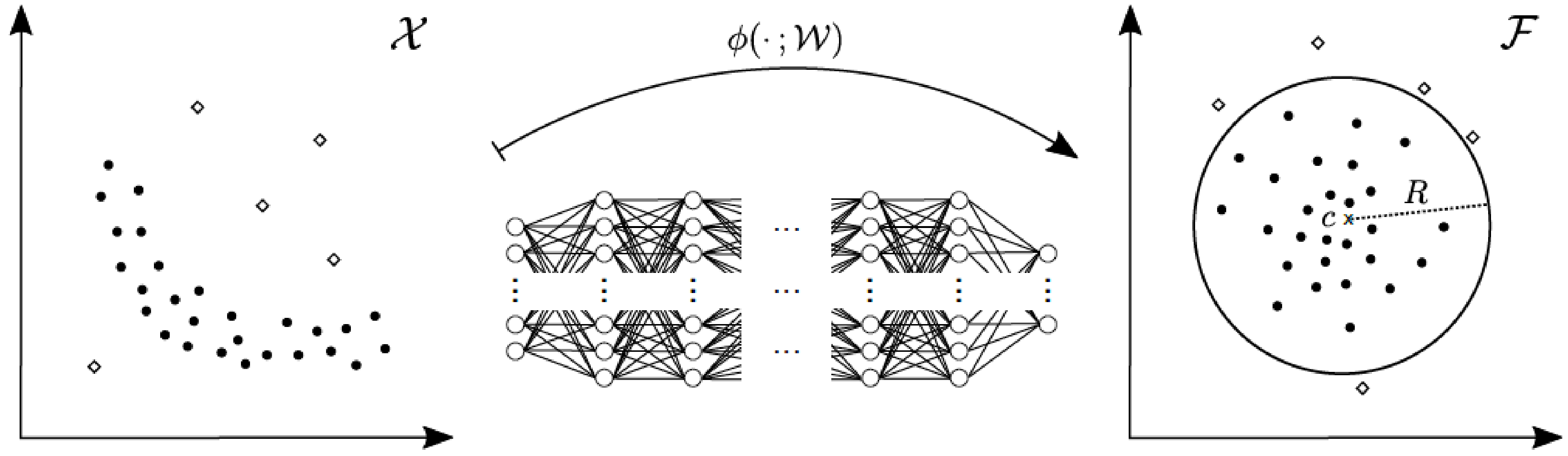
Anomaly Detection → **Feature Extraction** → **How do we extract representation of data well?**



3. Deep Anomaly Detection(DSVDD)

Deep One-Class Classification(ICML, 2018)

Anomaly Detection ⟶ Feature Extraction ⟶ How do we extract representation of data well?



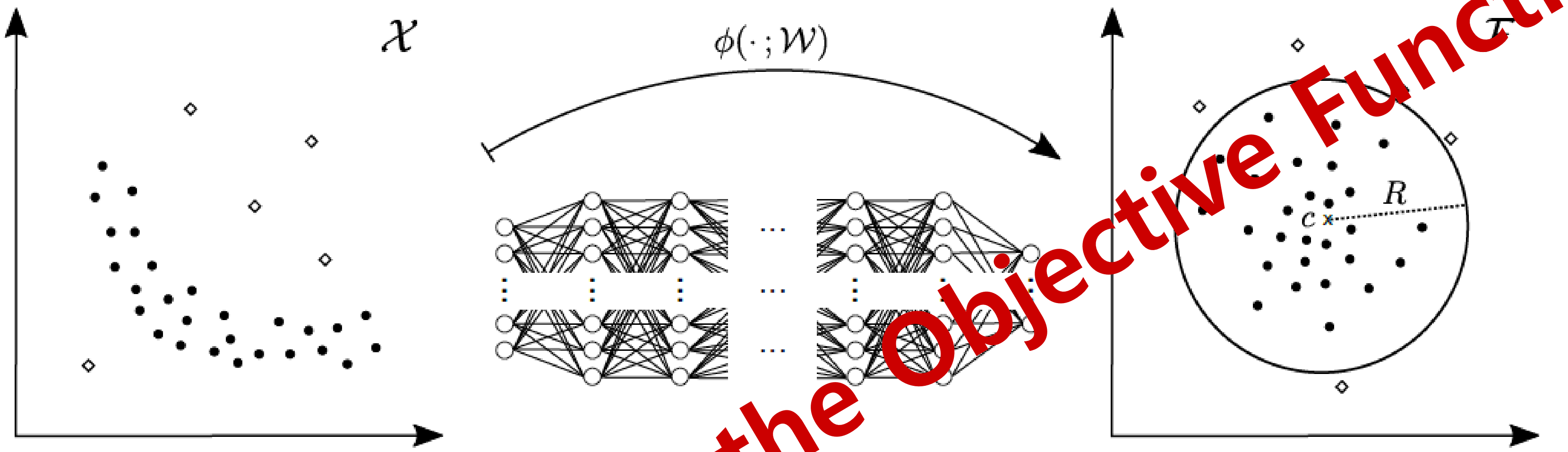
$$Soft\ SVDD = \ min_{R,W}\ R^2 + \frac{1}{\nu n}\sum_{i=1}^{n} \max\{0, ||\phi(x_i;W) - c||^2\} - R^2 \ + \frac{\lambda}{2}\sum_{l=1}^{L} ||W^l||_F^2$$

$$SVDD = \ min_W \sum_{i=1}^{n} ||\phi(x_i;W) - c^2|| + \frac{\lambda}{2}\sum_{l=1}^{L} ||W^l||_F^2$$

3. Deep Anomaly Detection(DSVDD)

Deep One-Class Classification(ICML, 2018)

Anomaly Detection → Feature Extraction → How do we extract representation of data well?



Just optimize the Objective Function using SGD

$$Soft\ SVDD = min_{R,W}\ R^2 + \frac{1}{\nu n}\sum_{i=1}^{n} \max\{0, ||\phi(x_i;W) - c||^2\} - R^2 + \frac{\lambda}{2}\sum_{l=1}^{L} ||W^l||_F^2$$

$$SVDD = min_W \sum_{i=1}^{n} ||\phi(x_i;W) - c^2|| + \frac{\lambda}{2}\sum_{l=1}^{L} ||W^l||_F^2$$

3. Deep Anomaly Detection(DSVDD)

Deep One-Class Classification(ICML, 2018)

Normal Class	OC-SVM/ SVDD	KDE	IF	DCAE	AnoGAN	Soft-Bound. Deep SVDD	One-Class Deep SVDD
0	**98.6**±0.0	97.1±0.0	98.0±0.3	97.6±0.7	96.6±1.3	97.8±0.7	98.0±0.7
1	99.5±0.0	98.9±0.0	97.3±0.4	98.3±0.6	99.2±0.6	99.6±0.1	**99.7**±0.1
2	82.5±0.1	79.0±0.0	88.6±0.5	85.4±2.4	85.0±2.9	89.5±1.2	**91.7**±0.8
3	88.1±0.0	86.2±0.0	89.9±0.4	86.7±0.9	88.7±2.1	90.3±2.1	**91.9**±1.5
4	**94.9**±0.0	87.9±0.0	92.7±0.6	86.5±2.0	89.4±1.3	93.8±1.5	**94.9**±0.8
5	77.1±0.0	73.8±0.0	85.5±0.8	78.2±2.7	88.3±2.9	85.8±2.5	**88.5**±0.9
6	96.5±0.0	87.6±0.0	95.6±0.3	94.6±0.5	94.7±2.7	98.0±0.4	**98.3**±0.5
7	93.7±0.0	91.4±0.0	92.0±0.4	92.3±1.0	93.5±1.8	92.7±1.4	**94.6**±0.9
8	88.9±0.0	79.2±0.0	89.9±0.4	86.5±1.6	84.9±2.1	92.9±1.4	**93.9**±1.6
9	93.1±0.0	88.2±0.0	93.5±0.3	90.4±1.8	92.4±1.1	94.9±0.6	**96.5**±0.3
AIRPLANE	61.6±0.9	61.2±0.0	60.1±0.7	59.1±5.1	**67.1**±2.5	61.7±4.2	61.7±4.1
AUTOMOBILE	63.8±0.6	64.0±0.0	50.8±0.6	57.4±2.9	54.7±3.4	64.8±1.4	**65.9**±2.1
BIRD	50.0±0.5	50.1±0.0	49.2±0.4	48.9±2.4	**52.9**±3.0	49.5±1.4	50.8±0.8
CAT	55.9±1.3	56.4±0.0	55.1±0.4	58.4±1.2	54.5±1.9	56.0±1.1	**59.1**±1.4
DEER	66.0±0.7	**66.2**±0.0	49.8±0.4	54.0±1.3	65.1±3.2	59.1±1.1	60.9±1.1
DOG	62.4±0.8	62.4±0.0	58.5±0.4	62.2±1.8	60.3±2.6	62.1±2.4	**65.7**±2.5
FROG	74.7±0.3	**74.9**±0.0	42.9±0.6	51.2±5.2	58.5±1.4	67.8±2.4	67.7±2.6
HORSE	62.6±0.6	62.6±0.0	55.1±0.7	58.6±2.9	62.5±0.8	65.2±1.0	**67.3**±0.9
SHIP	74.9±0.4	75.1±0.0	74.2±0.6	**76.8**±1.4	75.8±4.1	75.6±1.7	75.9±1.2
TRUCK	75.9±0.3	**76.0**±0.0	58.9±0.7	67.3±3.0	66.5±2.8	71.0±1.1	73.1±1.2

What is important?

4. What is important?

1. Domain Knowledge
2. GAN을 이용한 Anomaly Detection은 아직은 별루다.
   But Mode Problem을 잘 이용하면 재미있는 것을 할 수 있을지도…
3. Anomaly Detection은 Representation을 잘 추출하는게 관건이다.
   (Reconstruction, Classification)
4. 추출한 잠재변수를 잘 사용해서 비정상 점수를 잘 뽑아내야한다.

We are Hiring

https://github.com/MINGUKKANG/SIA

- **Selected as Deep Learning Best Practice at NVIDIA AI Conference 2018 Keynote.**
- **Use Cases for “Accelerate AI with Synthetic data using generative adversarial networks” at Strata Data Conference 2018 NY.**
- **ICPR 2018 Contests on Object Detection in Aerial Images 2위, 국방부 주관 제 20차 M&S 발전 세미나 우수 논문상 수상**
- **CVPR 2017 NTIRE Challenge Task 2,3,4,5 위 수상, Nips 2017 DeeoArt.io Poster Contest 1등상 수상**
- **Kaggle “DSTL Satellite Imagery Feature Detection” Silver medal 수상**
- **이외에도 SIA의 연구 내용은 CVPR, ICPR, ACML, ICML, ACM SIGSPATIAL 등 학회에서 발표되었습니다.**

Thank you!