

# **Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery**

2023.02.15

## **I. INTRODUCTION**

## **II. Generative Adversarial Representation Learning to Identify Anomalies**

- 2.1 Unsupervised Manifold Learning of Normal Anatomical Variability
- 2.2 Mapping new Images to the Latent Space
- 2.3 Detection of Anomalies

## **III. Experiments**

- 3.1 Results

## **IV. Conclusion**

# <Abstract>

(unsupervised learning based DCGAN + novel anomaly scoring scheme) → anomaly detection  
anomaly detection framework AnoGAN 제시

## about AnoGAN

- Anomaly Detection
  1. unsupervised learning(manifold of normal data 를 학습) 방식의 a deep convolutional generative adversarial network (DCGAN)
  2. (image space → a latent space) 로 매핑하기 위한 새로운 anomaly scoring scheme 추가
    - ➡ (1+2) anomaly score 계산
- training 후 새로운 데이터가 들어오면,
  1. anomalies 에 레이블 지정
  2. normal data 로만 학습된 distribution 과 얼마나 유사한지 image patches 에 anomaly score 나타냄.
- 실험 결과  
OCT 로 촬영한 망막 이미지로 실험했을 때, AnoGAN 은 anomalous images(retinal fluid or hyperreflective foci) 를 정확히 식별해냄.

# I. INTRODUCTION

- imaging data 에서 disease markers 의 detection and quantification 는 disease progression 과 treatment response 의 진단 및 모니터링에 중요
    - 하지만, vocabulary of known markers 에 의존하는 것은 더 많은 관련 정보를 포함하는 imaging data 의 사용이 제한
  - 기존 연구의 한계점
    1. supervised training
      - automating detection을 위해 labeled lesions 과 같은 많은 양의 annotated data 가 필요한 supervised training 수행
      - 새로운 유형의 정상, 비정상이 나타나면 감지 x
    2. Data Imbalanced problem
      - many diseases 은 (normal data ↑ abnormal data ↓)
- unsupervised learning 으로 large-scale imaging data 에 대한 relevant anomalies 식별 제안

# I. INTRODUCTION

## + AnoGAN 의 등장 배경

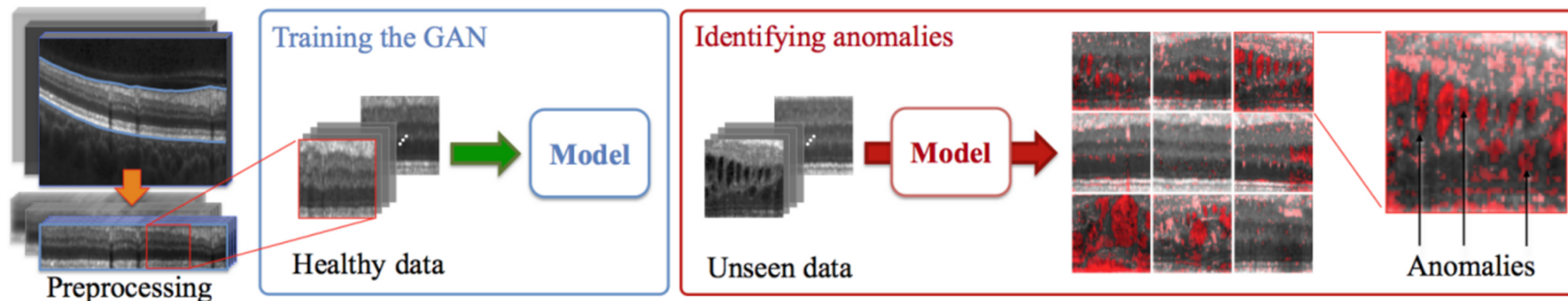
- Anomaly detection : training 중 봤던 normal data distribution 와 맞지 않는 test data 식별 작업
- GAN 의 Discriminator : input image 가 fake(0)/real(1) 인지 판별해주는 classifier
- idea: medical data 기준, Discriminator 를 train 할 때 훈련하지 않은 데이터(abnormal data)들은 다 Fake 라고 판별하지 않을까?  
→ Normal Distribution 의 outlier detection 을 할 수 있는 AnoGAN 제시
- outlier : 해당 anomalies 의 정확한 질환 설명 X  
해당 이미지는 정상은 아니고 정확히 어떤 질환이라고 판단할 수 는 없지만 정상 데이터는 아닌듯 하다.  
해당 이미지의 어느 부분이 의심스러운지, Anomaly 한 부분이 있는지 탐지

# I. INTRODUCTION

- model

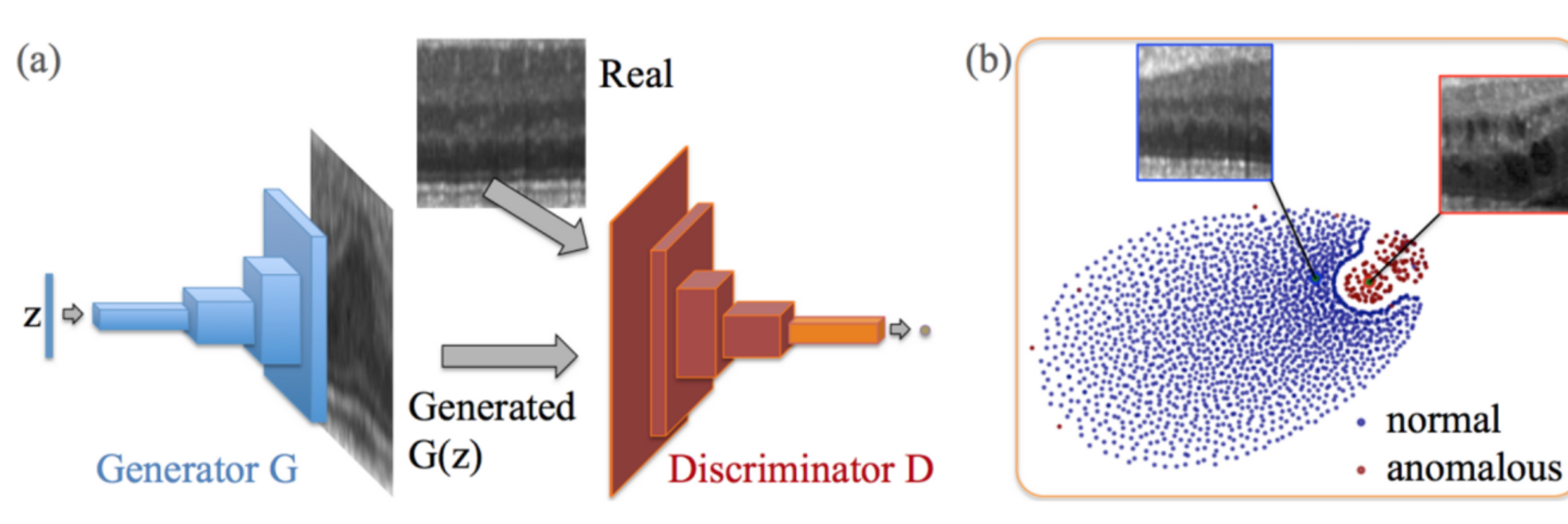
1. Data Imbalanced problem 해결을 위해 normal data(healthy local anatomical appearance) 를 만들고 + (normal/abnormal) 의 차이를 학습하는 unsupervised learning 의 generative adversarial networks (GANs) 을 제안
2. [image space  $\rightarrow$  latent space] 로 매핑하는 기술 제안  
 $\rightarrow$  2 가지 요소를 사용해서 anomaly detection 수행

## Anomaly detection framework



- preprocessing step
  - retinal area 의 extraction and flattening + patch extraction + intensity normalization.
- Generative adversarial training
  - healthy data(normal data) 로만 수행
- testing
  - unseen healthy cases and anomalous data 에 대해 모두 수행

## 2.1 Unsupervised Manifold Learning of Normal Anatomical Variability



(a) Deep convolutional generative adversarial network. (DCGAN)

(b) t-SNE embedding

D의 last convolution layer feature representation 기준  
normal (blue) and anomalous (red)

- train data(normal)

- normal medical images ( $I_m$ )

- $I_m \in \mathbb{R}^{a \times b}$ ,  $m = 1, 2, \dots, M$  : size  $a \times b$  의  $M$  개 medical images (normal = healthy anatomy)

- normal medical images( $I_m$ )  $\rightarrow$  image patches ( $x$ )

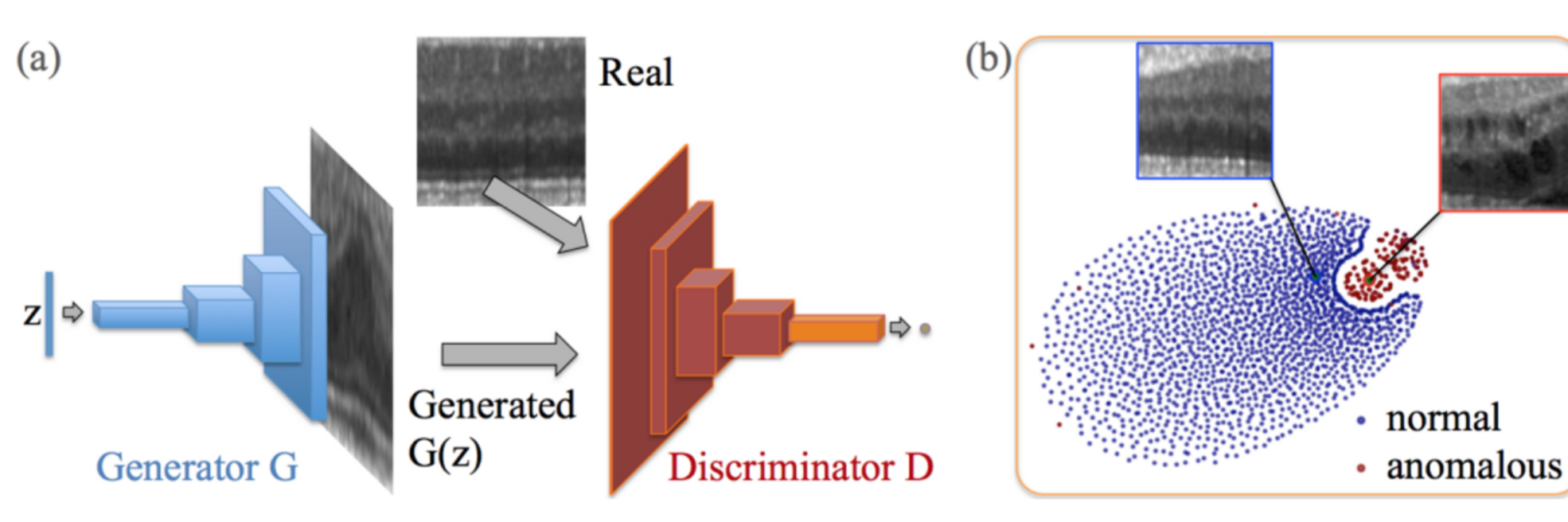
- From each image  $I_m$  에서 랜덤으로 샘플링된 size  $c \times c$  의  $K$  개 2D image patches( $x$ )추출

$$\mathbf{x} = x_{k,m} \in \mathcal{X}, \text{ with } k = 1, 2, \dots, K.$$

$\rightarrow$  큰 메디컬 사진( $I_m$ ) 안에서  $c \times c$  크기의 일부 구간을 랜덤으로 잘라내고, 잘라진 구간(image patch) 라고 말하며,  
해당 image patch 는 healthy manifold  $\mathcal{X}$  를 학습

manifold : embedding 된 데이터가 원래 데이터를 잘 표현하는 저차원 공간 학습

## 2.1 Unsupervised Manifold Learning of Normal Anatomical Variability



(a) Deep convolutional generative adversarial network. (DCGAN)

(b) t-SNE embedding

D의 last convolution layer feature representation 기준  
normal (blue) and anomalous (red)

- test data(normal/abnormal)

- $[y_n, l_n]$  - test data 와 label

- $y_n$  : new testing data  $J$  추출된 size  $c \times c$  의 unseen images
    - $l_n \in \{0, 1\}$  , where  $1 \leq n \leq N$  : an array of binary image-wise ground-truth labels

→ anomaly detection performance 를 평가하기 위해 testing 중일 때만 제공



# Encoding Anatomical Variability with a Generative Adversarial Network

## AnoGAN - GAN 구조

→ 일반적인 DCGAN 과 같음

- samples  $z$  : 1D vector, Latent space  $Z$ 로부터 추출
- Image patch  $x$  : normal medical images 에서 랜덤으로 샘플링된 size  $c \times c$  이미지이며, image space manifold  $X$  안에 속함
- generator  $G : z \rightarrow x$  (1D  $\Rightarrow$  2D), stack strided convolutions 구조로 Deconvolution 방식
  - samples  $z$  의 mapping 함수  $G(z) : z \rightarrow x$  를 통해 data  $x$  distribution 인  $p_g$  학습  
 $\Rightarrow$  latent space  $Z \rightarrow$  image space manifold  $X$  (healthy examples) 의 2D images 로 매핑하여 fake normal image 생성
- discriminator  $D : a \text{ 2D image} \rightarrow a \text{ single scalar value } D(\cdot)$  매핑하는 standard CNN 으로 Convolution 연산 이용
  - $D(\cdot)$ :  $D$  의 output = probability  
 $\Rightarrow$  training data  $X$  에 주어진 (a real image  $x$  / generator  $G$  가 생성한  $G(z)$ ) 구분

# Encoding Anatomical Variability with a Generative Adversarial Network

## GAN 학습 Loss function

$$\min_G \max_D V(D, G) = \mathbb{E}_{\mathbf{x} \sim p_{data}(\mathbf{x})} [\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z} \sim p_{\mathbf{z}}(\mathbf{z})} [\log(1 - D(G(\mathbf{z})))] . \quad (1)$$

- D and G 는 minmax function  $V(G, D)$  을 통해 동시에 optimized  
→  $V(G, D)$  : 잘 구분하는 D 에 대해 잘 구분하지 못하는 G를 생성
- D : V가 maximum이 되도록 학습 ( $D(x) = 1, D(G(z)) = 0$  일때 maximum = 0) [ $D(\text{fake}) = 0$  /  $D(\text{real}) = 1$ ]
- G : V가 minimum이 되도록 학습 ( $D(G(z)) = 1$ 일때 minimum =  $-\infty$ )

adversarial training is completed...

- generator G : mapping  $G(z)$  [latent space  $z \rightarrow$  realistic (normal) images  $x$ ] train 완료

## 2.2 Mapping new Images to the Latent Space

adversarial training is completed...

- generator  $G$  : mapping  $G(z) = [\text{latent space } z \rightarrow \text{realistic (normal) images } x]$  train 완료
- 문제점 : mapping  $G^{-1}(z) = [\text{real (normal) images } x \rightarrow \text{latent space } z]$  이 어려움.
  - GAN 의 Generator :  $z \rightarrow x$  생성
  - AnoGAN 의 Generator : real image  $x$  와 최대한 유사한 latent space  $Z$  를 찾아서 point  $z$  로부터 fake image 생성  
→ real image 와 유사한 fake image 를 생성해야 하기 때문
  - AnoGAN 의 Generator 는 GAN 과 다르게  $X$  의 분포를  $Z$  로 매핑하는 과정이 필요  
→ 결국 Real Image 의 분포를 참고한 Fake Image 를 만들어내게 된다.
- AnoGAN : query image  $x$  가 주어지면, latent space 에서 a point  $z$  를 찾음.
  - 최적의 point  $z$  : query image  $x$  와 시각적으로 유사함 + 비슷한 분포를 가지는 latent space 에서의 한 지점

## 2.2 Mapping new Images to the Latent Space

- To find the best  $z$ ,
  1. G와 D 를 normal dataset 으로 학습을 완료  $\rightarrow$  더 이상 G 와 D 의 파라미터는 업데이트 되지 않음.
  2. Random sampling 으로 얻은  $z_1$ 을 trained generator 에 input으로  $\rightarrow$  a generated image  $G(z_1)$  을 얻음.
    - 1 = update iteration 의 횟수
  3. novel loss function 을 기반으로 gradients를 계산하고 backpropagation을 통해서  $z$  를 업데이트 ( $z_1, z_2, \dots, z_\Gamma$ )

$\Rightarrow$  latent space  $Z$  에서 point  $z$  의 위치는 backpropagation steps (  $\gamma = 1, 2, \dots, \Gamma (500)$ ) 를 통해 optimized
- [new images  $\rightarrow$  the latent space] 로 매핑하기 위한 2가지 요소로 구성된 loss function 정의
  1. a residual loss : generated image  $G(z_\gamma) \leftrightarrow$  query image  $x$  사이의 visual similarity 을 높임.
  2. a discrimination loss : generated image  $G(z_\gamma)$  를 learned manifold  $X$  에 속하도록 함.

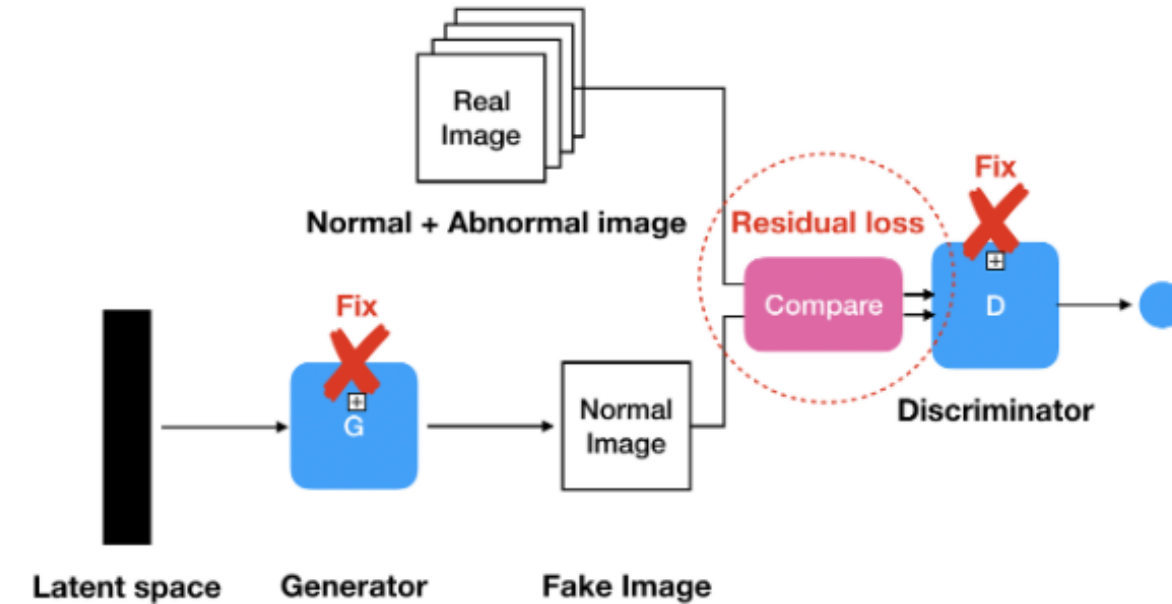
## 2.2 Mapping new Images to the Latent Space

### *Residual Loss*

→ query image  $x$  와 생성된 이미지  $G(z_\gamma)$  의 시각적 차이 측정

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

|real(normal) image - generated image|



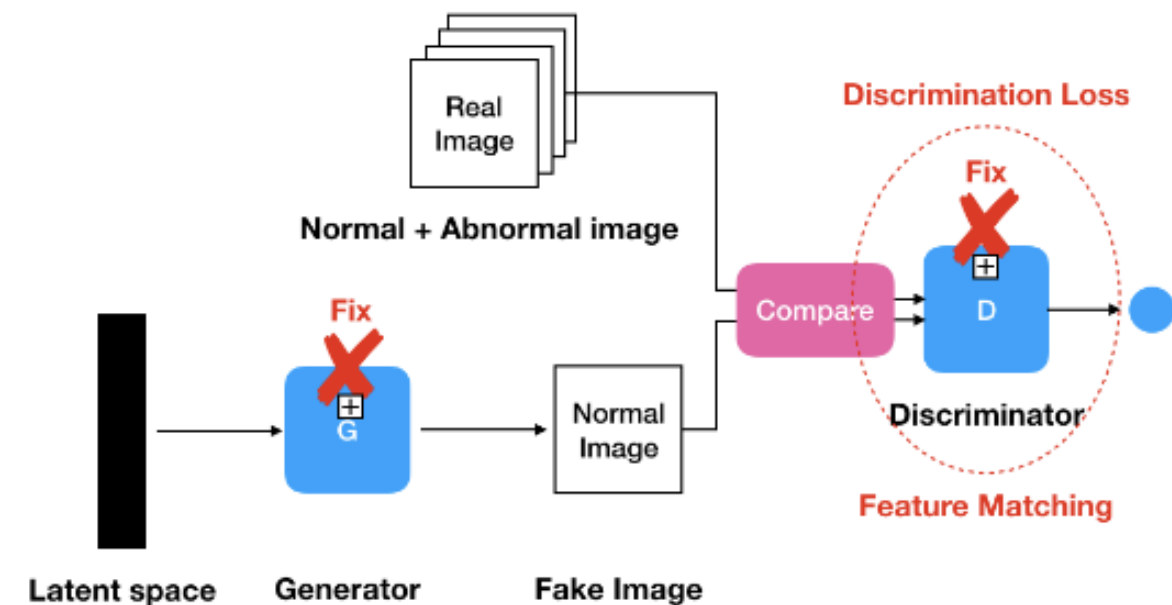
### *discrimination loss - Feature Mapping*

→ generated image  $G(z_\gamma)$  가 manifold  $X$  에 잘 매핑되도록

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

|real(normal) image feature - generated image feature|

- $f(\cdot)$ : input image 의 feature vector 반환
- discriminator 의 scalar output 을 사용하는 대신, discriminator intermediate layer  $f(\cdot)$ 의 feature 를 넣어서 계산  
→ 중간층이 더 많은 표현력 가짐.



## 2.3 Detection of Anomalies

*For the mapping to the latent space : residual loss + discrimination loss*

→ 입력 이미지와 가장 유사한 이미지를 생성해내는 latent vector  $z$  를 업데이트 하는 loss function

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

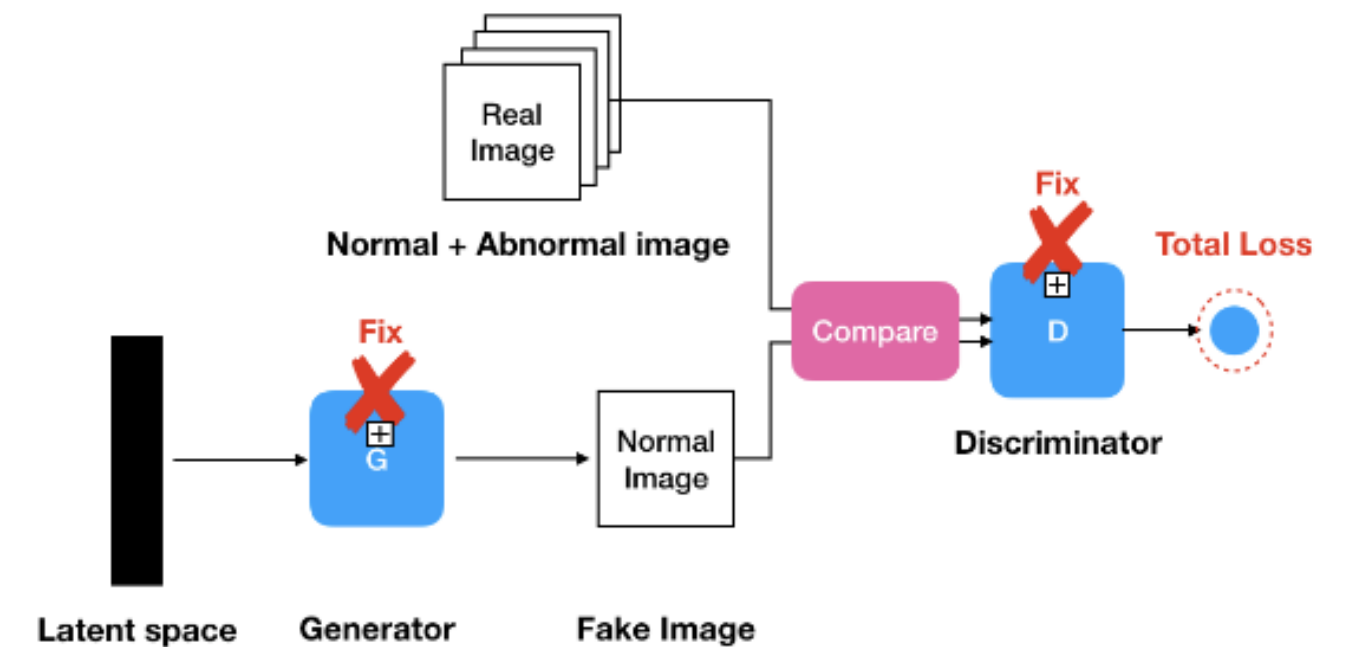
- $\lambda$  : 가중치 (0.1)
- generator and discriminator 의 trained parameters 는 고정
- $z$  의 coefficients 만 backpropagation 을 통해 조정
- $L(z_\gamma)$  : 매 update iteration  $\gamma$  마다  $G(z_\gamma)$  과 query image  $x$  비교

### Anomaly Score

→ new query image  $x$  를 normal or anomalous image 로 판단

$$A(\mathbf{x}) = (1 - \lambda) \cdot R(\mathbf{x}) + \lambda \cdot D(\mathbf{x}),$$

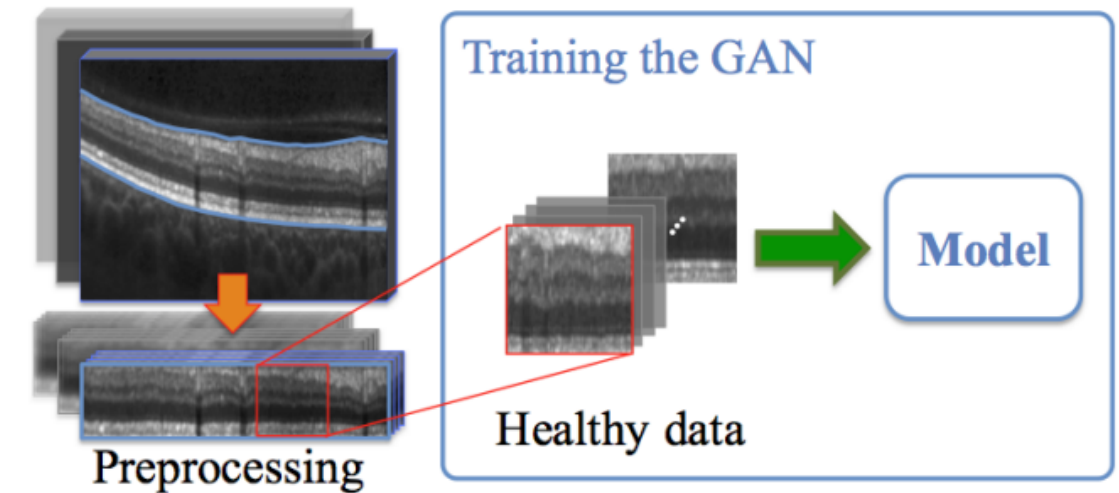
- $A(x)$  : Anomaly Score,  $A(x) \uparrow$  anomaly,  $A(x) \downarrow$  학습한 데이터와 유사한 이미지
- $R(x)$  : Residual Score  $L_R(z_\gamma)$ , 이미지의 어느 부분이 anomalous regions 인지 판단
- $D(x)$  : Discriminator Score  $L_D(z_\gamma)$



# 3 Experiments

## *Data, Data Selection and Preprocessing*

- Data : SD-OCT라는 의료 영상 이미지(망막)를 사용
- OCT volumes Preprocessing
  - a. gray values : (-1~1) 사이 범위로 normalize
  - b. volumes : 22 $\mu$ m 크기로 x 방향으로 조정  $\rightarrow$  약 256 columns 생성
  - c. retinal area 추출 후 평평하게 하여 방향, 모양, 두께 조정
  - d. automatic layer segmentation algorithm 사용하여 z-direction 으로 망막의 상단 및 하단 layer 찾음.
- train data
  - 위 데이터에서 randomly sampled positions 에서 64 $\times$ 64 pixels 의 1,000,000 2D training patches 추출
- test data
  - retinal fluid 가 포함된 10 additional healthy cases + 10 pathological cases 에서 patches 추출
- GAN : healthy subjects 의 망막에서 추출한 270 개의 2D image patches 에 대해 train 됨.



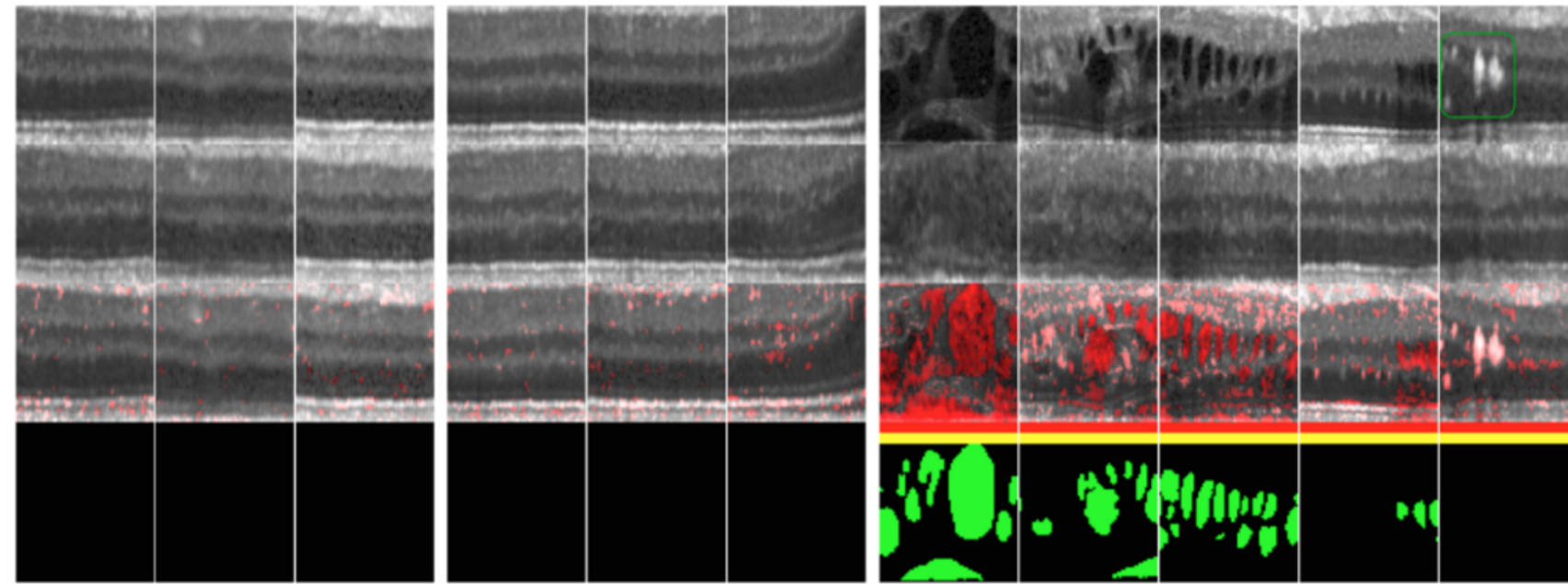
# 3 Experiments

## *Implementation details*

- 이미지 size :  $64 \times 64$  pixels  $\rightarrow$  가장 안정적으로 train 됨.
- Discriminator와 Generator의 구조 : 기존의 DCGAN architecture 와 같음
- generator
  - 4 fractionally-strided convolution layers
- discriminator
  - 4 convolution layers
- channel : 512—256—128—64 channels (DCGAN 의 절반으로)  $\rightarrow$  gray-scale images 을 사용하기 때문
- filter size :  $5 \times 5$
- new images  $\rightarrow$  latent space : 500 backpropagation steps
- $\lambda = 0.1$  in Equations (5) and (6)



# 3.1 Results



Real input images. (x)

generated images. (g(z))

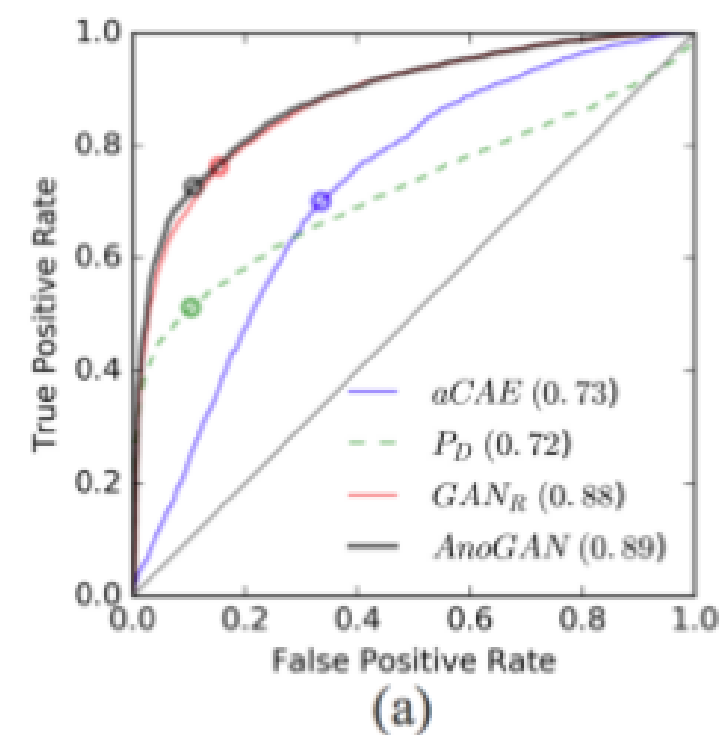
Residual overlay - **Red bar**: Anomaly identification by  $R(x)$   
- **Yellow bar**: Anomaly identification by  $D(x)$

Pixel-level annotations of retinal fluid.

- 1 block and 2 block: training set and test set 에서 추출된 Normal images
- 3 block: test set 에서 추출된 diseased cases
  - input images ↔ generated images : intensity or textural differences 차이를 보여줌
- 마지막 열: Hyperreflective foci (within **green box**)
  - additional retinal lesions 식별 가능

# 3.1 Results

각 모델의 *image level* 의 (ROC) curves + Clinical performance



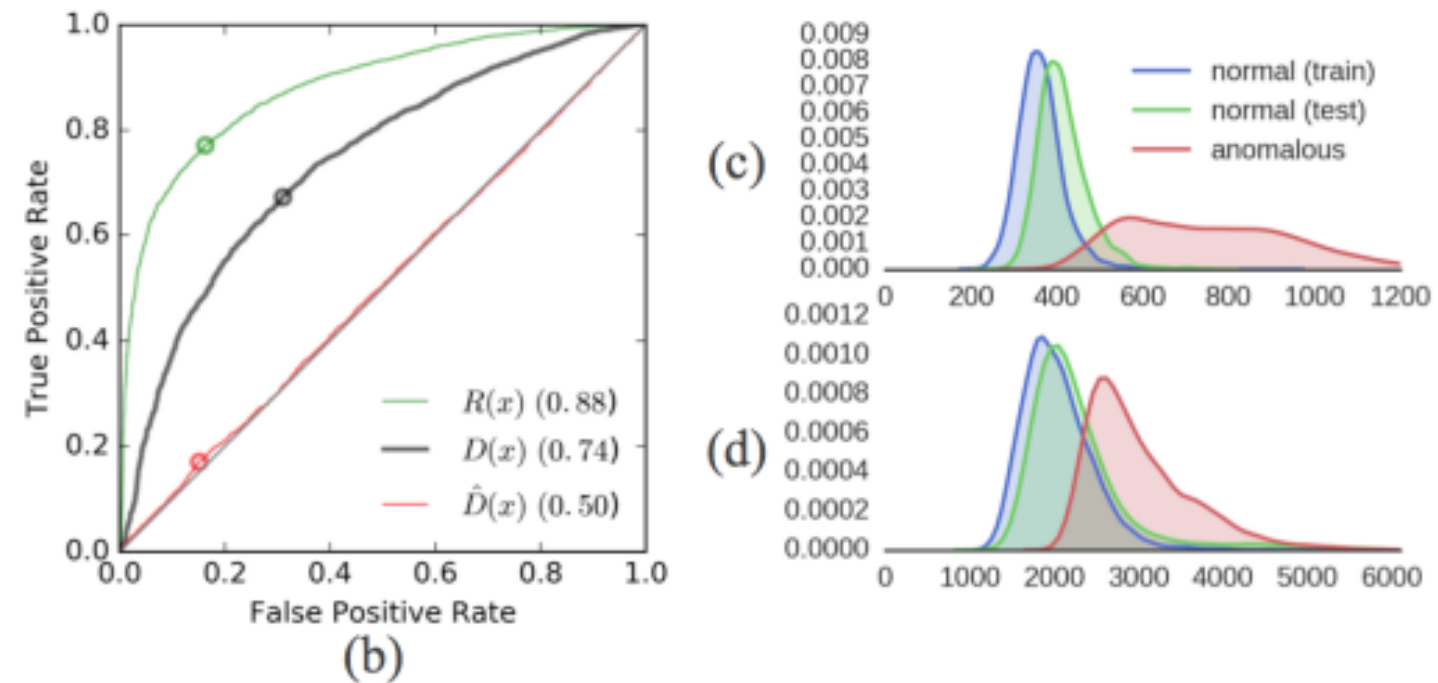
|         | Precision | Recall | Sensitivity | Specificity | AUC  |
|---------|-----------|--------|-------------|-------------|------|
| aCAE    | 0.7005    | 0.7009 | 0.7011      | 0.6659      | 0.73 |
| $P_D$   | 0.8471    | 0.5119 | 0.5124      | 0.8970      | 0.72 |
| $GAN_R$ | 0.8482    | 0.7631 | 0.7634      | 0.8477      | 0.88 |
| AnoGAN  | 0.8834    | 0.7277 | 0.7279      | 0.8928      | 0.89 |

- sensitivity: 질병이 있는 사람이 검사에서 질병으로 판정될 확률
- specificity : 정상인 사람이 검사에서 정상으로 판정될 확률

- (a) : 각 모델의 ROC curve
  - aCAE: Adversarial Convolution AutoEncoder 구조로 Encoder-Decoder Based
  - $P_D$ : Discriminator 가 판단한 확률 값을 Anomaly score 로 사용한 모델
  - $GAN_R$  : Referenced Adversarial score를 Anomaly score 로 사용한 것
  - AnoGAN → 가장 성능 좋음
- Table : 각 모델의 Clinical performance
  - aCAE:  $A(x)$
  - $P_D$ : DCGAN 의 output
  - $GAN_R$ :  $A^{\wedge}(x)$
  - AnoGAN :  $A(x)$  → 가장 성능 좋음

# 3.1 Results

각 모델의 *image level* 의 (ROC) curves



- (b) score 별 ROC curve
  - $R(x)$  : residual score → 가장 성능 좋음
  - $D(x)$  : discriminator score
  - $\hat{D}(x)$  : reference discriminator score
- (c) : Residual score 의 분포 → 더 정확히 normal/abnormal 구별
- (d) : Discrimination score 의 분포

# 4 Conclusion

- We propose anomaly detection based on deep generative adversarial networks.
- healthy data 에 대한 unsupervised training 기반 학습 중 보지 못했던 unseen data 인 retinal fluid and HRF 와 같은 anomalies 를 판별할 수 있음.
  - model is expected to be capable to discover novel anomalies.
- In contrast to prior work,
  - residual loss 만으로도 from image to latent space 의 mapping 에 좋은 결과를 제공함.