

BeatGAN : 적대적으로 생성된 시계열을 이용한 비정상적인 리듬 감지

BeatGAN: Anomalous Rhythm Detection using Adversarially Generated Time Series

INDEX

1. 서론

- 1) 이상 탐지 기술
- 2) 비지도 이상 탐지
- 3) 생성모델 그리고 GAN

2. 본론

- 1) 시계열 데이터
- 2) BeatGAN
- 3) 실험 및 결과

3. 결론

1. 서론

이상 현상

이상(anomaly)현상

의도하지 않은 정상적인 시스템의 동작

→ 발생한 현상이 '이상 상태'임을 인지

이상 현상→탐지하는 기술

이상 탐지 기술
(Anomaly
Detection)

이상 탐지 기술 (Anomaly Detection)

학습 시 '비정상 sample' & 'label' 의 유무에 따른 분류

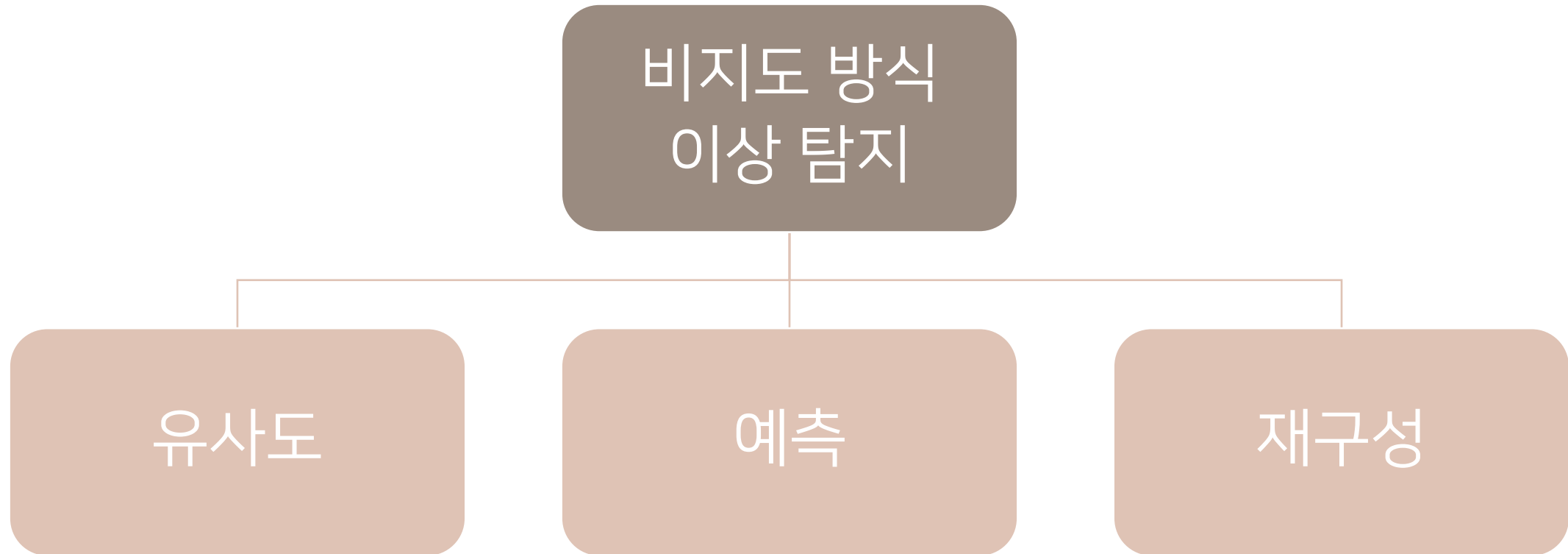
지도 (Supervised) 이상 탐지

- Labeled data
- Direct feedback
- 장점 : 다른 방법 대비 정확도가 높다.
- 단점 : 비정상 **sample** 취득->시간과 비용이 많이 든다.

비지도 (Unsupervised) 이상 탐지

- No labeled data
- No feedback
- “숨겨진 패턴과 특성을 발견하여 분석”
- 장점 : **Labeling** 과정 필요X
- 단점 : 다른 방법 대비 정확도가 낮다.

비지도(Unsupervised) 이상 탐지



재구성 기반 판별 방식

재구성 기반 판별 방식
(Reconstruction)

재구성 기반 판별 방식

이상 판별할 데이터를 **저차원** 형태의 잠재 구조(latent structure)를 획득하고
인위적으로 **재구성**한 데이터를 생성하기 위한 모델 사용

Why? → **저차원**으로 매핑하면 정보가 손실되어 **효과적으로 재구성 가능**

대표적인 방법

주어진 데이터에 대해 **PCA**(Principal Component Analysis, 주성분 분석)을 이용하여
차원을 축소하고 **복원**을 하는 과정을 통해 비정상 sample 검출

→ 단점 : Linear 재구성으로 제한

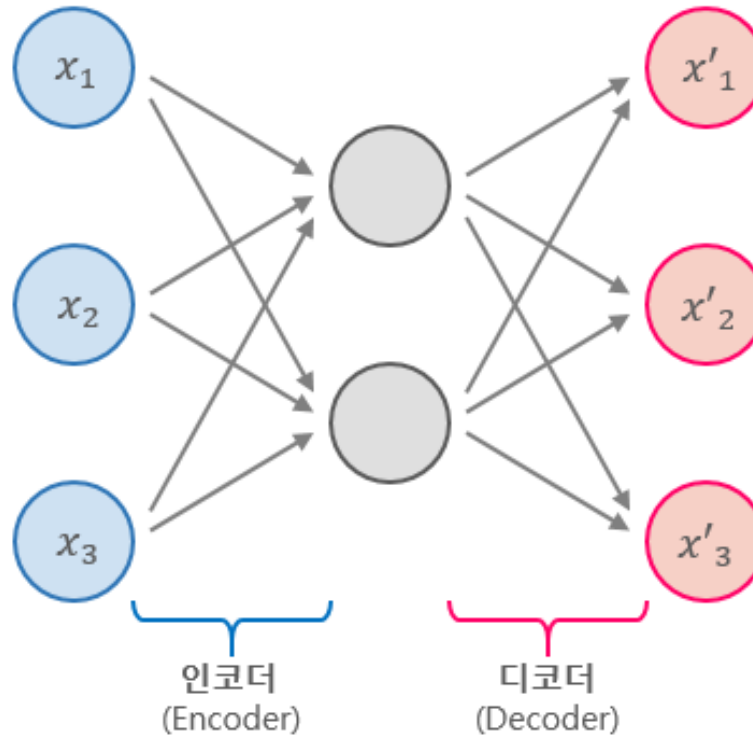
딥러닝 기반 기술

주로 **Autoencoder** 기반의 방법론이 자주 사용

AE(Auto-Encoder), **VAE**(Variational Auto-Encoder), **LSTM 기반 인코더-디코더 구조**

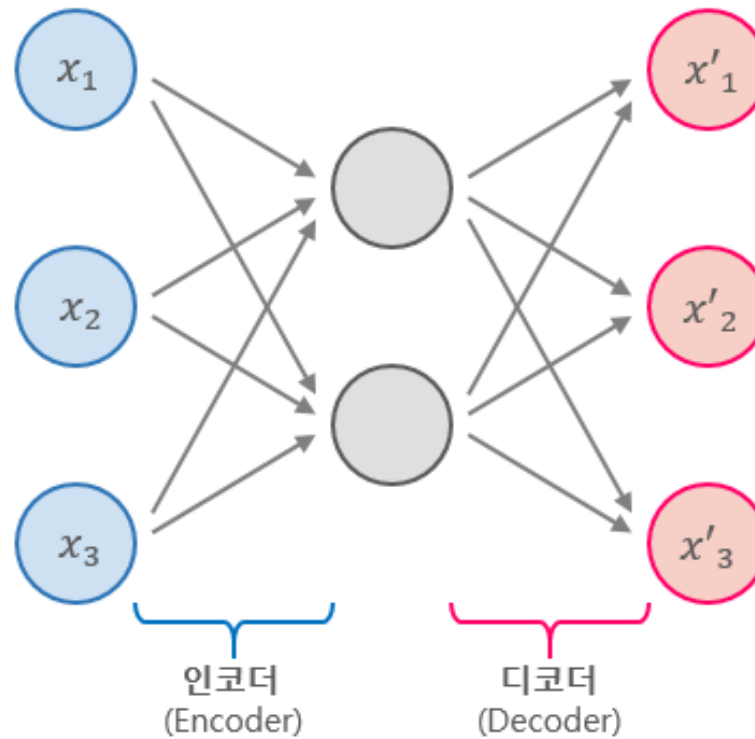
→ PCA와 달리 non-linear 차원 축소를 다룬다.

오토인코더



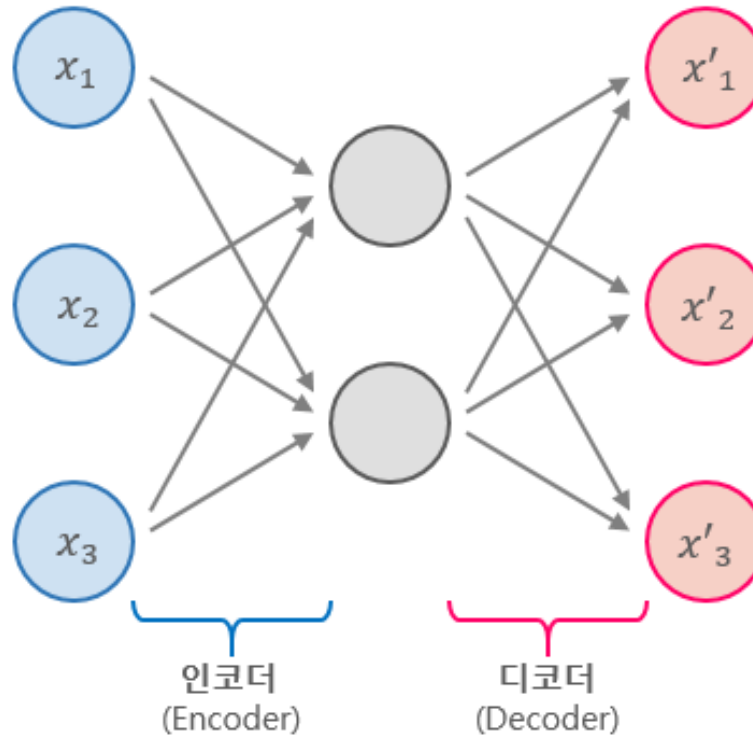
입력을 code 또는 latent variable로 압축하는 **Encoding**과
이를 다시 원본에 가깝게 복원해내는 **Decoding** 과정으로 진행

오토인코더



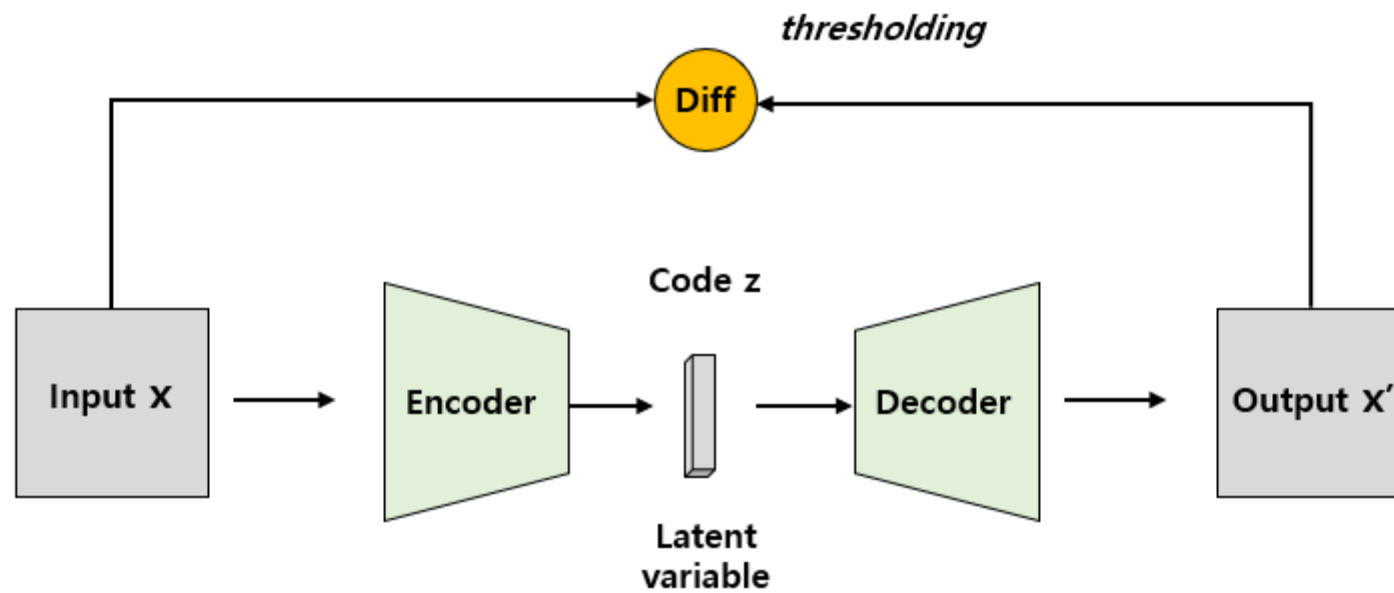
- Hidden layer의 뉴런이 입력층보다 작으므로 입력이 **저차원**으로 표현됨
- Autoencoder는 입력을 그대로 출력으로 복사할 수 없어서
출력이 입력과 같은 것을 출력하기 위해 학습해야 함
 - 입력 데이터에서 가장 중요한 **특성**을 **학습**하도록 만듦

오토인코더

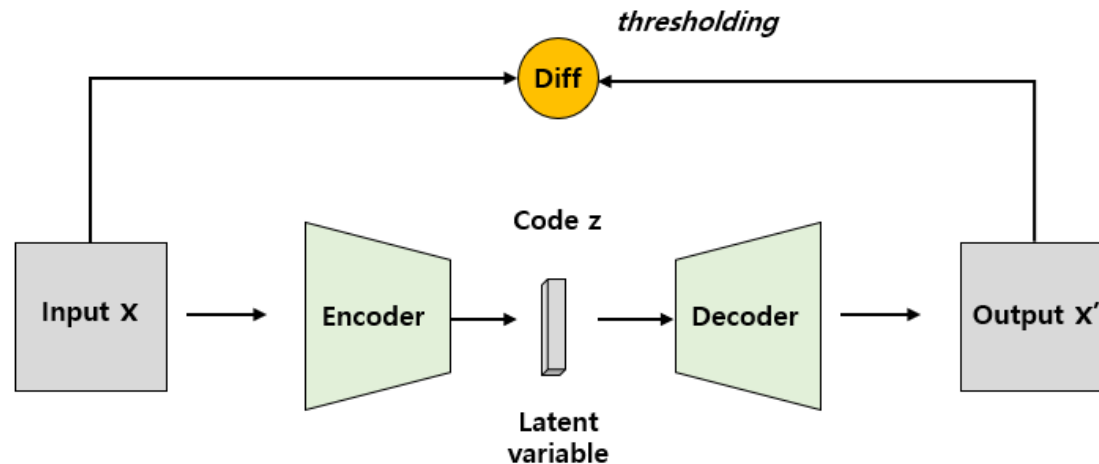


데이터에 대한 labeling을 하지 않아도
데이터의 중요한 정보들만 압축적으로 배울 수 있다.
(데이터 주성분을 배우는 PCA와 유사한 동작)

Autoencoder(AE)

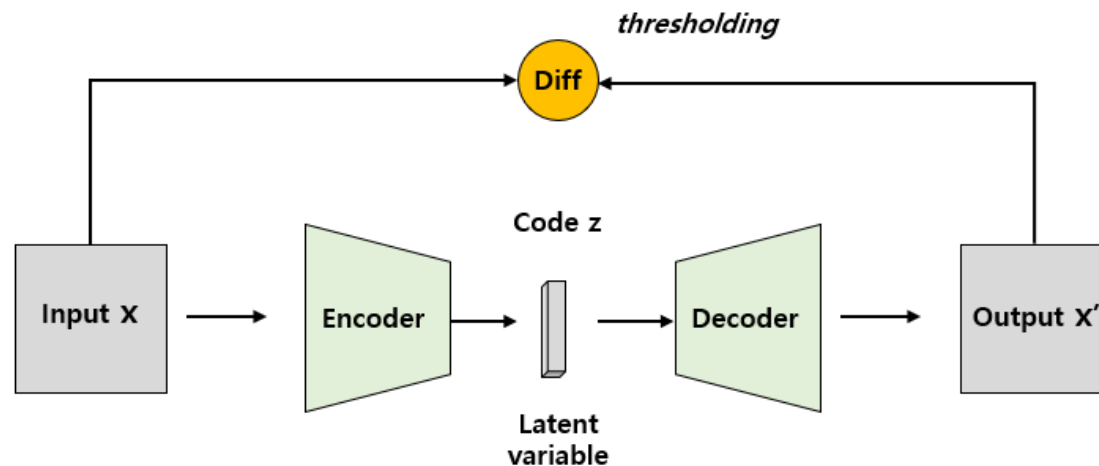


Autoencoder(AE)



- 입력 샘플을 encoder 통해 저차원으로 **압축**
- 압축된 샘플을 decoder에 통과시켜 다시 원래 차원으로 **복원**
- 입력 샘플과 복원 샘플의 **복원 오차**(reconstruction error)을 구함
- 복원 오차는 **이상 점수**(anomaly score)가 되어 threshold와 비교로 이상 여부를 결정
 - Threshold보다 크면 → 이상
 - Threshold보다 작으면 → 정상

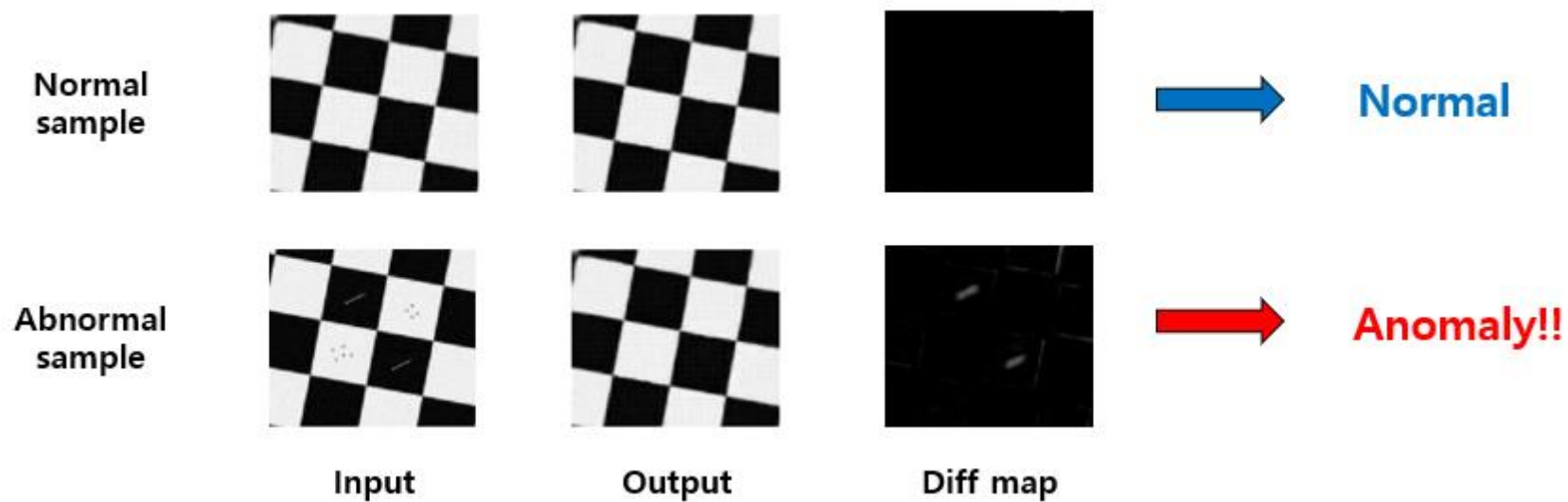
Autoencoder(AE)



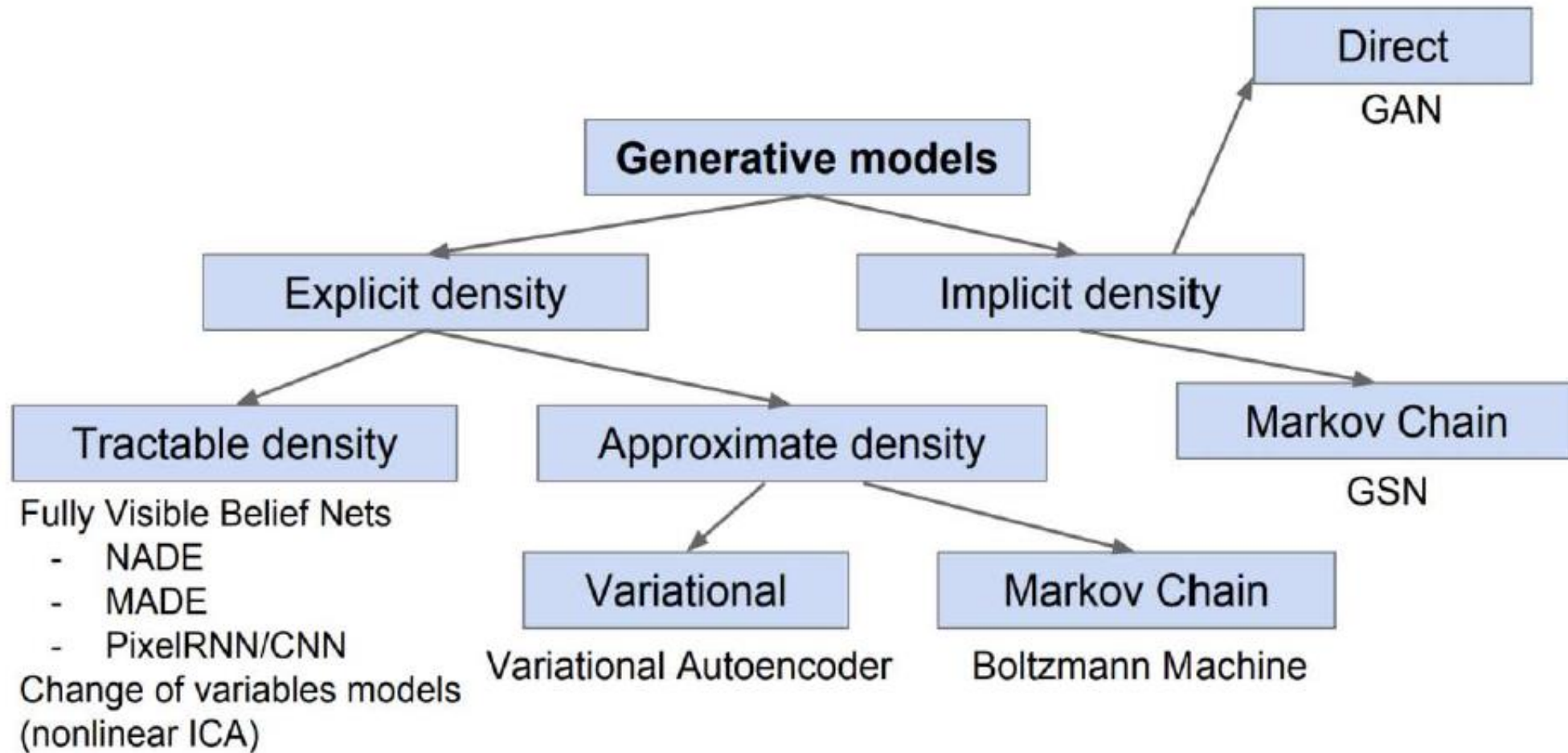
비정상 sample이 input으로 주어진다면?

- Autoencoder은 효과적으로 압축과 복원을 수행하지 못함
- 주어진 sample의 특징을 잘 추출하지 못하여 **복원 오차가 커짐**
- 비정상 sample로 판정

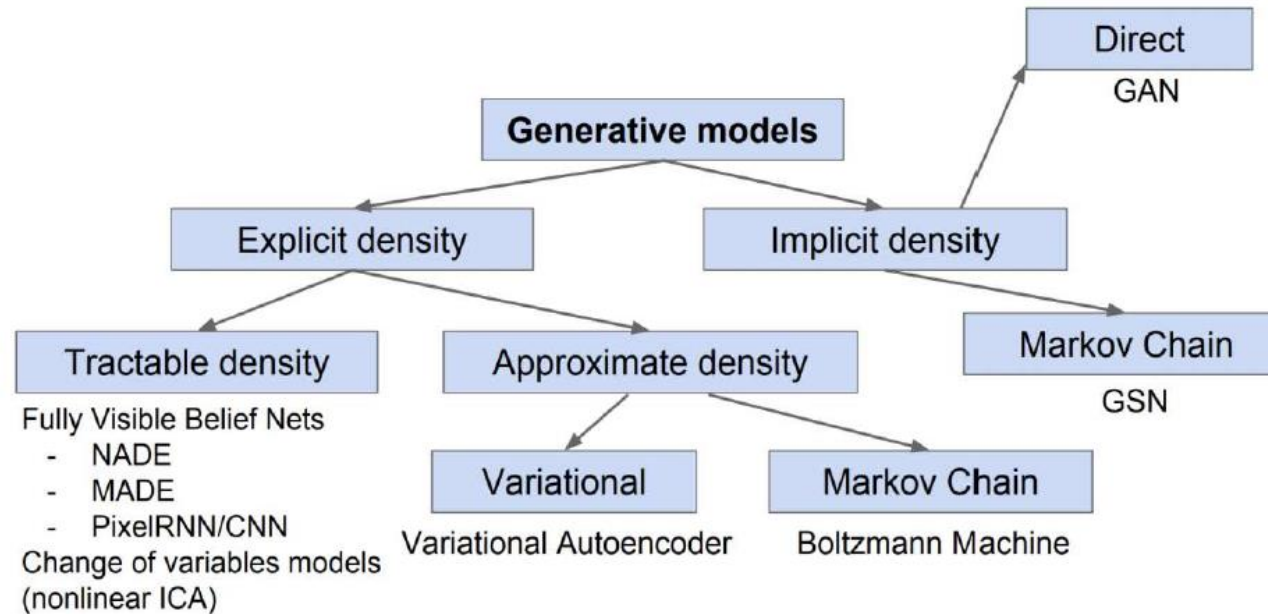
Autoencoder(AE)



생성모델(Generative model)



생성모델(Generative model)



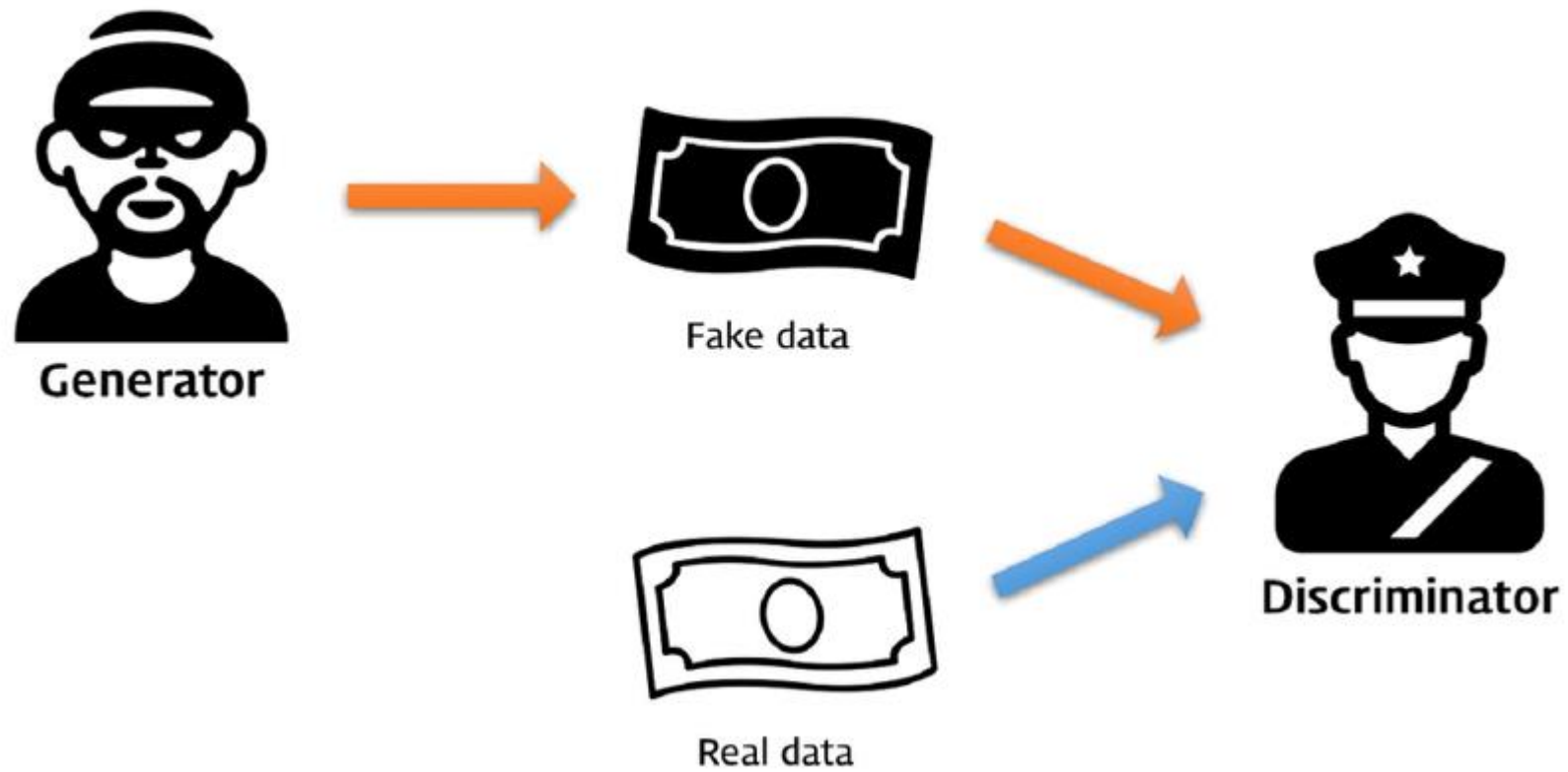
생성모델은 주어진 **학습 데이터를 학습**하여 학습데이터의 분포를 따르는 **유사한 데이터를 생성**하는 모델

- Explicit : 학습 데이터의 분포를 안 상태에서 생성
- Implicit : 학습 데이터의 분포를 잘 모르지만 생성

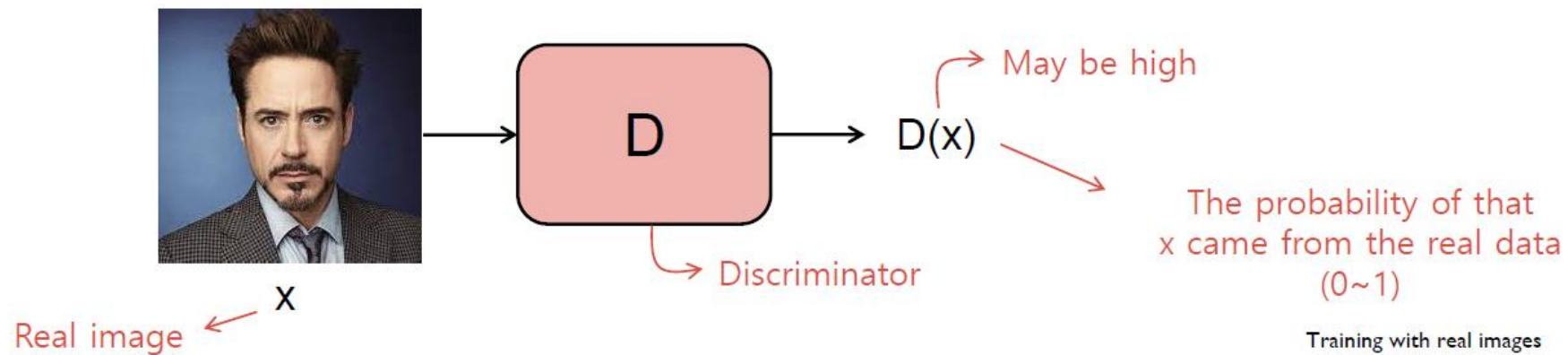
Generative Adversarial Networks

생성자(Generator)와 구분자(Discirimiator) 두 네트워크를
적대적(Adversarial)으로 학습시키는
비지도 학습 기반의 생성모델(Unsupervised Generative model)

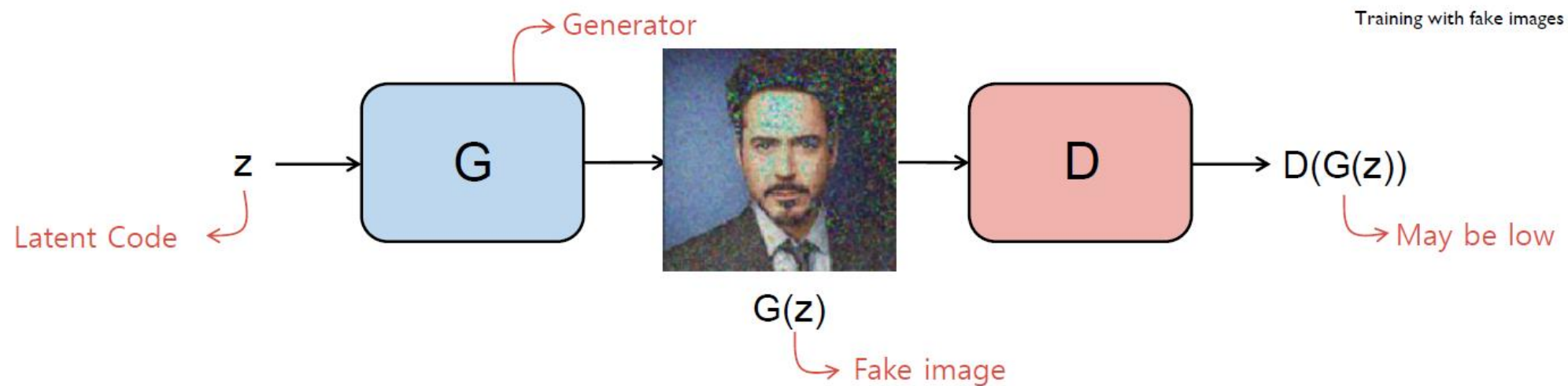
GAN



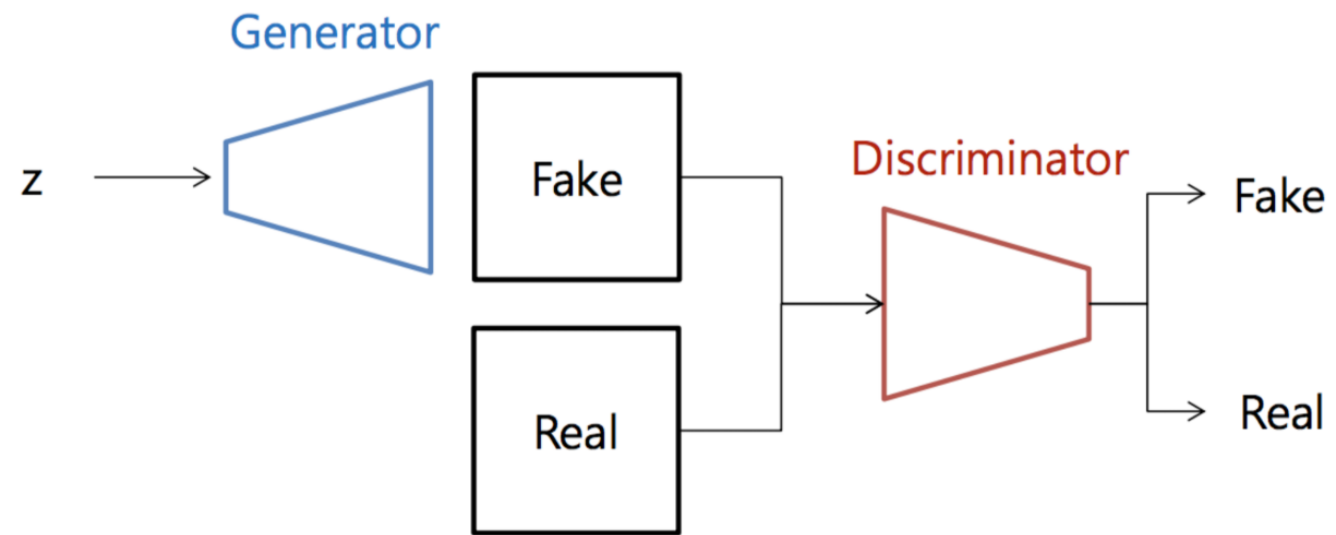
Discriminator의 역할/목표



Generator의 역할/목표



GAN



GAN의 value function

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)}[\log D(x)] + E_{z \sim p_z(z)}[\log(1 - D(G(z)))]$$

GAN의 value function

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log(1 - D(G(z)))]$$

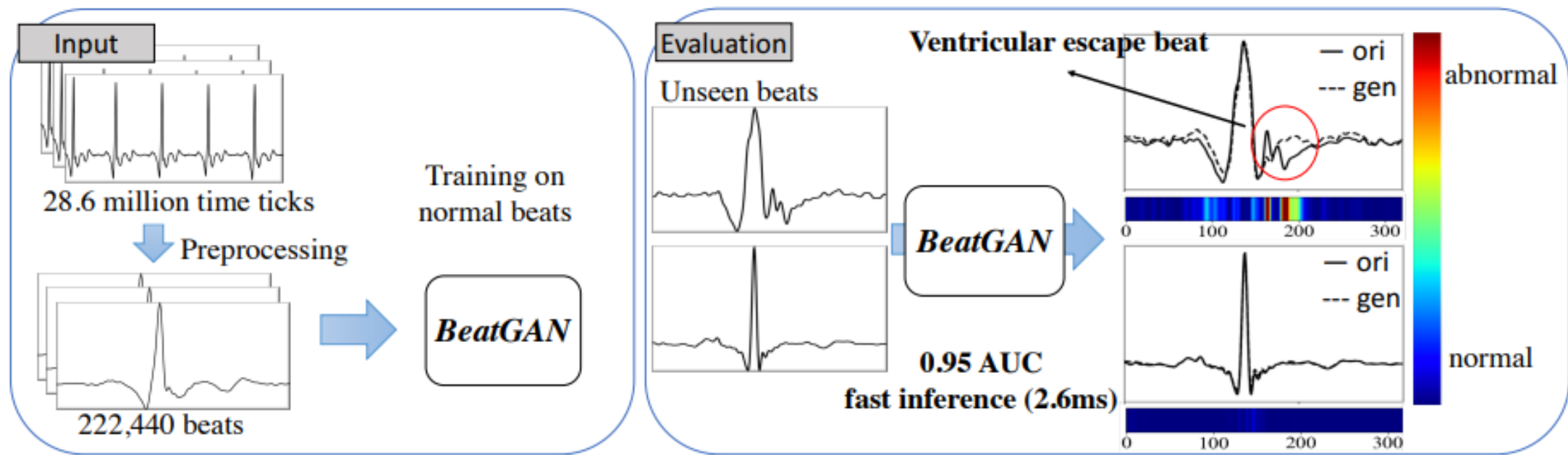
D는 $V(D, G)$ 를 **최대화** ← $\log D(x)$ 와 $\log(1 - D(G(z)))$ 가 1이 되어야 한다,
즉 실제 데이터는 '진짜'로, 가짜 데이터는 '가짜'로 분류하도록 학습

GAN의 value function

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log(1 - D(G(z)))]$$

G는 $V(D, G)$ 를 최소화 ← $\log(1 - D(G(z)))$ 가 최소가 되어야 한다, 즉 $D(G(z))$ 는 1이 되어야 한다
판별자가 '진짜'로 구별할 만큼 완벽한 가짜 데이터를 생성하도록 학습

General Framework



Preprocessing한 ECG data → train the model with normal heartbeats
→ unseen data x 를 학습된 모델에 feed → generated beat x' 를 얻는다
→ x 와 x' 를 비교하여 anomalies를 포착

The optimization objective & The anomalousness score

$$\begin{aligned} L &= \|X - G(X)\|_2 + R(G) \\ &= \sum_x \|x - G(x)\|_2 + R(G) \end{aligned} \tag{1}$$

$$A(x) = \|x - G(x)\|_2 \tag{2}$$

General Framework

Method	$G(\cdot)$	$R(G)$
SVD/ FBOX	$\sum_{k=1}^p \sigma_k u_k v_k^T$	$\lambda_1 I - U^T U _2 +$ $\lambda_2 I - V^T V _2$
AE	$G_D(G_E(x))$	/
VAE	$G_D(G_E(x))$	$\lambda D_{KL}[Q(z x) P(z)]$
BeatGAN	$G_D(G_E(x))$	adversarial regularization

Table 2: Unifying the reconstruction-based methods for anomaly detection. $G(\cdot)$ is the reconstruction function, $R(G)$ is the regularization loss

재구성 기반 이상 탐지 방법 ... SVD, AE, VAE, BeatGAN
모두 특정한 형태의 재구성 함수 $G(\cdot)$ 와 정규화 손실 $R(G)$ 을 갖는다

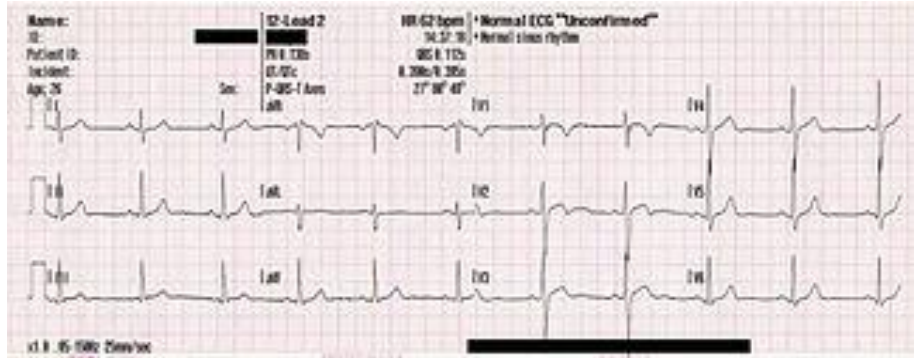
2. 본론

심전도(ECG, electrocardiogram)

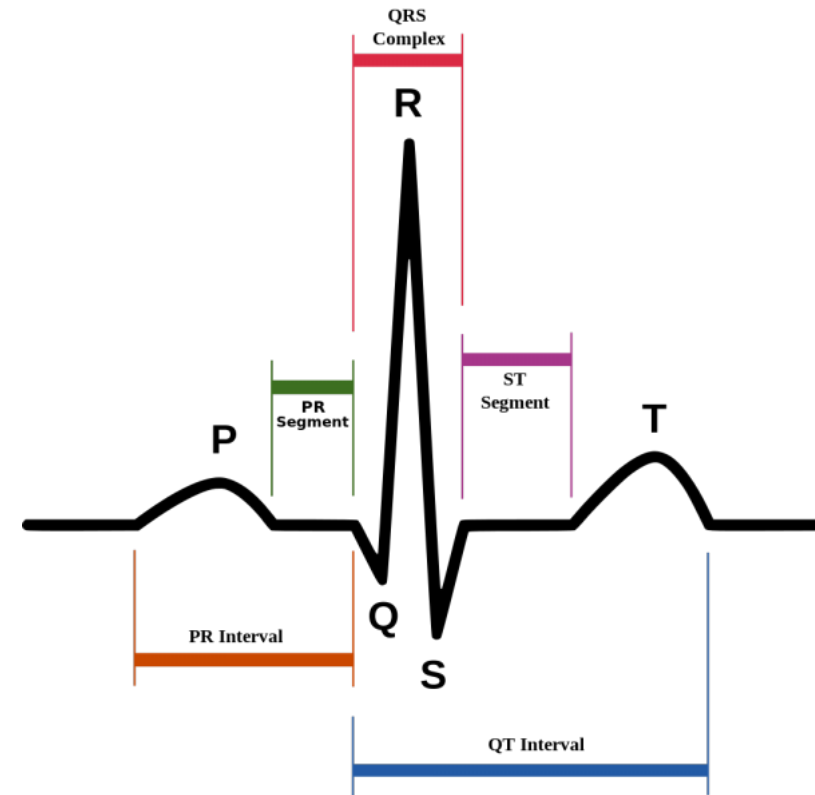
심전도(心電圖)

- 정해진 시간에 심장의 전기적 활동을 해석하는 것이다.
- 심전도는 피부에 부착된 전극과 신체 외부의 장비에 의해 기록된다.
- 심장의 비정상적인 리듬을 측정하고 진단하는 가장 좋은 방법이다.

심전도



가로축: 시간
세로축: 전압
→ 대규모 리듬 시계열 데이터

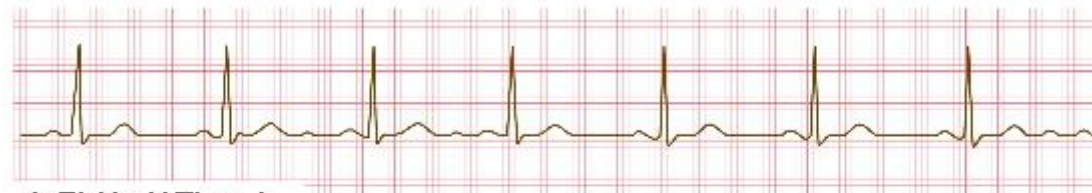


부정맥(cardiac dysrhythmia, heart arrhythmia)

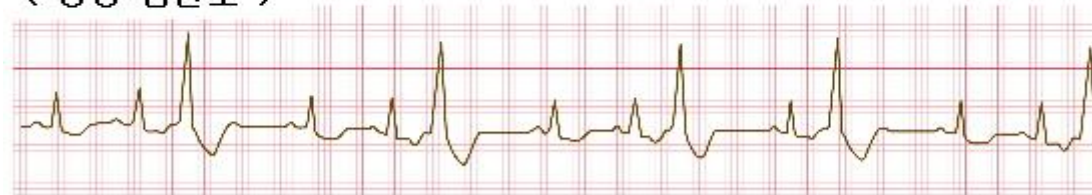
부정맥(不整脈)

- 심장박동이 고르지 못하고 **불규칙한 상태**를 말한다.
- 심장은 어른은 1분에 60회 정도 규칙적으로 뛴다.
- 지나치게 빠르거나, 늦거나 혹은 맥박이 불규칙하게 뛰는 것

부정맥



< 정상 심전도 >



< 부정맥 >

→ 비정상적인 리듬

BeatGAN

대규모의 리듬 시계열 데이터,
예를 들어 심전도(ECG) 판독값에서
효과적이고 효율적인 방법으로
비정상적인 beats를 감지하는 방법



BeatGAN

적대적으로 생성된 beats와 비교하여
입력 beat의 비정상적인 시간 틱을
정확히 찾아내는 설명 가능한 결과를 출력

BeatGAN

an unsupervised anomaly detection algorithm for time series data

The logo consists of the word "BeatGAN" in white, bold, sans-serif font, centered within a rounded rectangular box with a light brown or tan background.

BeatGAN

적대적 생성 접근 방식을 사용한 재구성 오류의 정규화와
시계열 워핑을 사용한 데이터 보강으로
견고성이 보장, 직관적인 접근 방식 제공

BeatGAN

Therefore, we propose BeatGAN,
an **unsupervised anomaly detection algorithm** for **time series data**.
BeatGAN outputs **explainable results** to pinpoint the anomalous time ticks of an input beat,
by comparing them to **adversarially generated beats**.

시계열 데이터 (Time Series Data)

- 데이터에 **시간**이라는 차원이 포함되어 있다
- 대표적으로 주가데이터, 일별 온도, ECG 데이터 등이 있다.
- 특정한 시간을 간격으로 관측된 데이터들의 집합을 모두 **Time Series**이라고 할 수 있다.
- 이러한 **시계열 데이터**를 **Classification**하는 방법 중 가장 널리 이용되는 것 : 1-NN DTW

Time Warping

DTW is not differentiable → 재구성 오류에 직접 사용 불가능

$$L_G = \|x - x'\|_2 + \lambda \|f_D(x) - f_D(x')\|_2$$

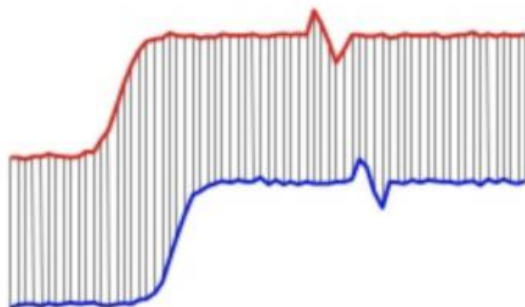
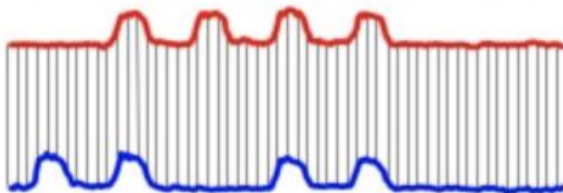
동적 시간 워핑 (DTW, Dynamic Time Warping)



Dynamic Time Warping?

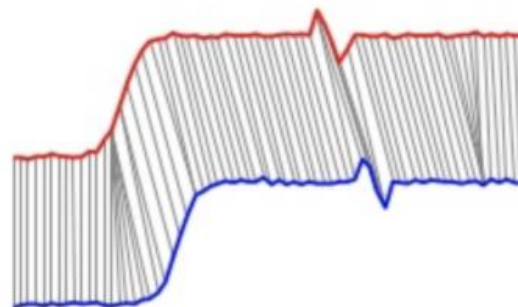
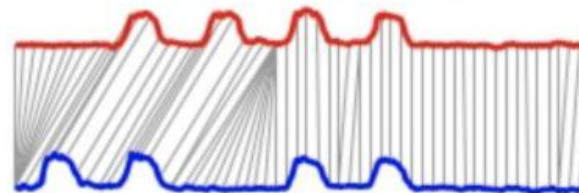
Euclidean Distance

Sequences are aligned "one to one"



DTW

Nonlinear alignments are possible



동적 시간 워핑 (DTW, Dynamic Time Warping)

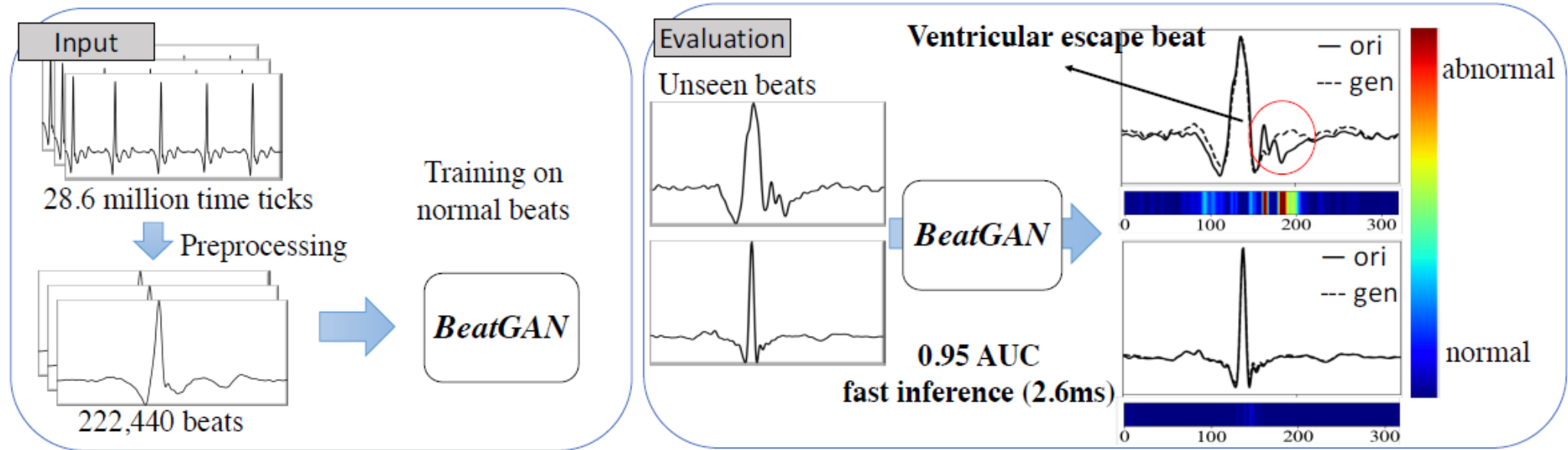
- 시계열 데이터 분석에서 널리 이용되는 거리 척도 중 하나
- 다른 속도, 움직임을 가진 서로 다른 신호의 시간축에 대한 파장의 유사성을 측정하는 알고리즘
- 그래픽, 비디오, 오디오 분야에서 자주 사용되며 의료분야에서 보행 유사성 등에서 자주 사용
- 데이터셋이 많을 경우 정확도가 높아진다.
- 단점 :
 - 자체 연산 비용이 비싸다.
 - 시계열 데이터의 길이가 길어질수록 시간이 오래 걸리는 단점이 존재
 - Not differentiable everywhere

Data Augmentation Using Time Warping

DTW is not differentiable → 재구성 오류에 직접 사용 불가능

$$L_G = \|x - x'\|_2 + \lambda \|f_D(x) - f_D(x')\|_2$$

BeatGAN



Preprocessing한 ECG data → train the model with normal heartbeats
→ unseen data x 를 학습된 모델에 feed → generated beat x' 를 얻는다
→ x 와 x' 를 비교하여 anomalies를 포착

Confusion matrix

학습을 통해 모델이 얼마나 예측을 잘 하는지 평가

		Predict	
		Positive	Negative
Actual	Positive	TP	FN
	Negative	FP	TN

- Actual(Positive) : 실제 환자
- Actual(Negative) : 실제 환자가 아님
- Predict(Positive) : 실제 환자로 예측
- Predict(Negative) : 환자가 아닌 것으로 예측

Confusion matrix

학습을 통해 모델이 얼마나 예측을 잘 하는지 평가

		Predict	
		Positive	Negative
Actual	Positive	TP	FN
	Negative	FP	TN

- TP(True Positive) : 긍정예측을 성공 즉, 환자라고 예측해서 실제 환자임을 맞춤
- TN(True Negative) : 부정예측을 성공 즉, 비환자라고 예측하여 실제 비환자임을 맞춤
- FP(False Positive) : 긍정예측을 실패 즉, 환자라고 예측했지만 비환자임
- FN(False Negative) : 부정예측을 실패 즉, 비환자라고 예측했지만 실제 환자임

AUC, ROC

AUC : Area under the Curve
= ROC curve 아래쪽의 면적

ROC : Receiver Operating Characteristic

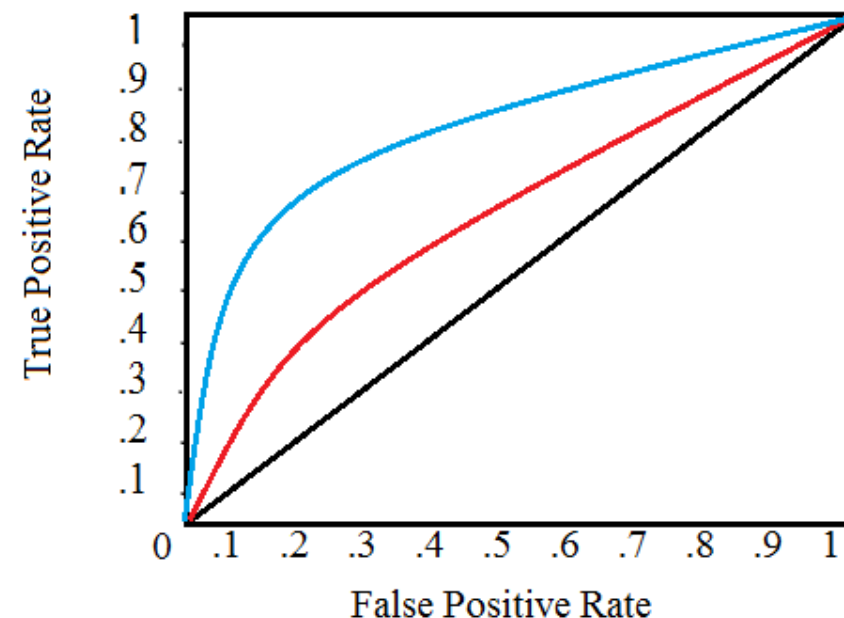
Y축 : True Positive Rate

X축 : False Positive Rate

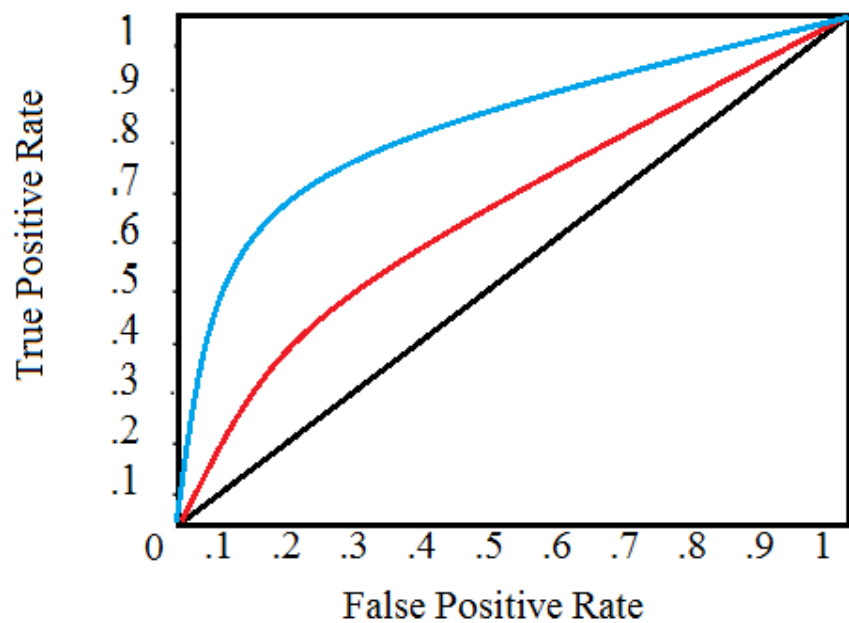
→ 왼쪽 상단(0,1)방향으로 더 휘어질 수록 예측성능이 뛰어난 모델

→ AUC가 클수록 예측을 잘하는 모델

ROC 곡선을 통해 모델의 성능 평가 or 최적의 분류기준(threshold)을 찾음



AUC, ROC



왼쪽 하단(0,0) : 모두 0(음성/부정)으로 예측
오른쪽 상단(1,1) : 모두 1(양성/긍정)으로 예측
왼쪽 상단(0,1): 잘못 예측한 것 없이 모두 맞춘 경우

BeatGAN : 재구성의 정규화

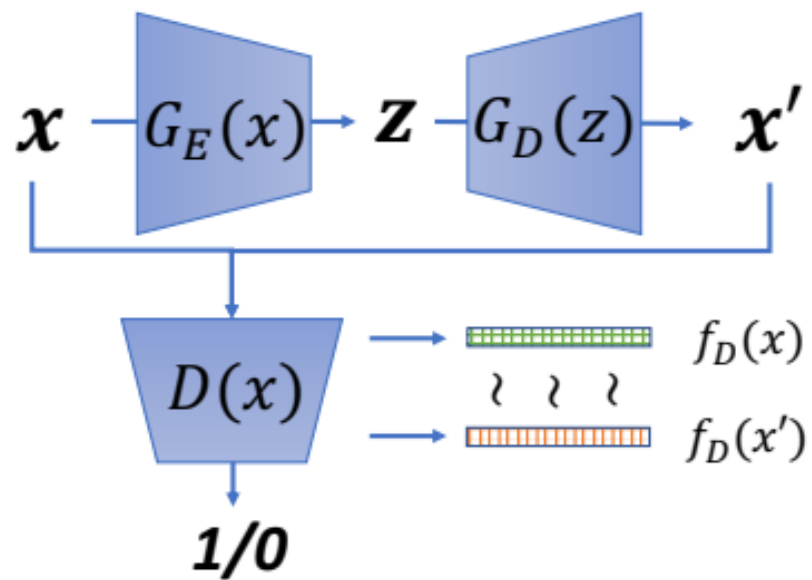


Figure 2: Illustration of our network structure

BeatGAN 실험 기준

001 >> Accuracy

얼마나 **정확**한가?

Data augmentation using time warping 을 했을 때와 안했을 때의 BeatGAN 차이는?

002 >> Explainability

입력의 비정상적인 부분을 잘 찾아내고 사람들의 주의를 잘 유도할 수 있는가?

설명 가능한 결과를 보여주는가?

003 >> Efficiency

추론은 얼마나 **빠른가**?

MIT-BIH ECG dataset

CMU Motion Capture dataset

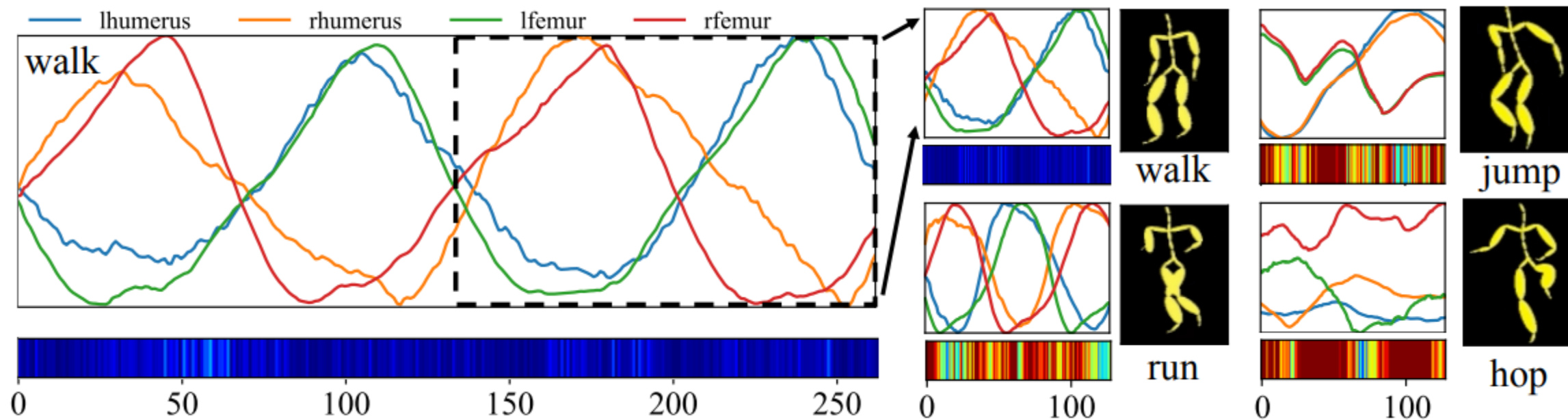


Figure 3: Example of anomaly detection on motion capture time series(4-dimensions). The right side shows the original time series and heatmaps to pinpoint the anomalies of jumping/running/hopping from walking motions.

Q1. Accuracy

BeatGAN gives the anomalousness score for each time series, i.e. $\mathcal{S} = \{s_i : A(x_i), x_i \in \mathcal{Z}\}$ for a given evaluation set \mathcal{Z} . To calculate metrics, we first standardize the scores between 0 and 1 by min-max scaling. Then we calculate the two metrics, AUC (Area Under ROC Curve) and AP (Average Precision) [Davis and Goadrich, 2006].

Q1. Accuracy

Method	AUC	AP
PCA	0.8164 ± 0.0037	0.6522 ± 0.0061
OCSVM	0.7917 ± 0.0018	0.7588 ± 0.0027
AE	0.8944 ± 0.0128	0.8415 ± 0.0163
VAE	0.8316 ± 0.0025	0.7882 ± 0.0024
AnoGAN	0.8642 ± 0.0100	0.8035 ± 0.0069
Ganomaly	0.9083 ± 0.0122	0.8701 ± 0.0141
BeatGAN	0.9447 ± 0.0053	0.9108 ± 0.0049
BeatGAN _{aug}	0.9475 ± 0.0037	0.9143 ± 0.0047
BeatGAN ^{0.1%} _{aug}	0.9425 ± 0.0022	0.8973 ± 0.0042

Table 3: BeatGAN performs the best for anomalous rhythm detection in ECG data. In BeatGAN_{aug}, we augment the training data size to $3\times$ with time warping. In BeatGAN^{0.1%}_{aug}, we add the 0.1% anomalous time series to the training data for evaluating robustness. 5-fold cross-validations are run, and mean and std are given.

Q1. Accuracy

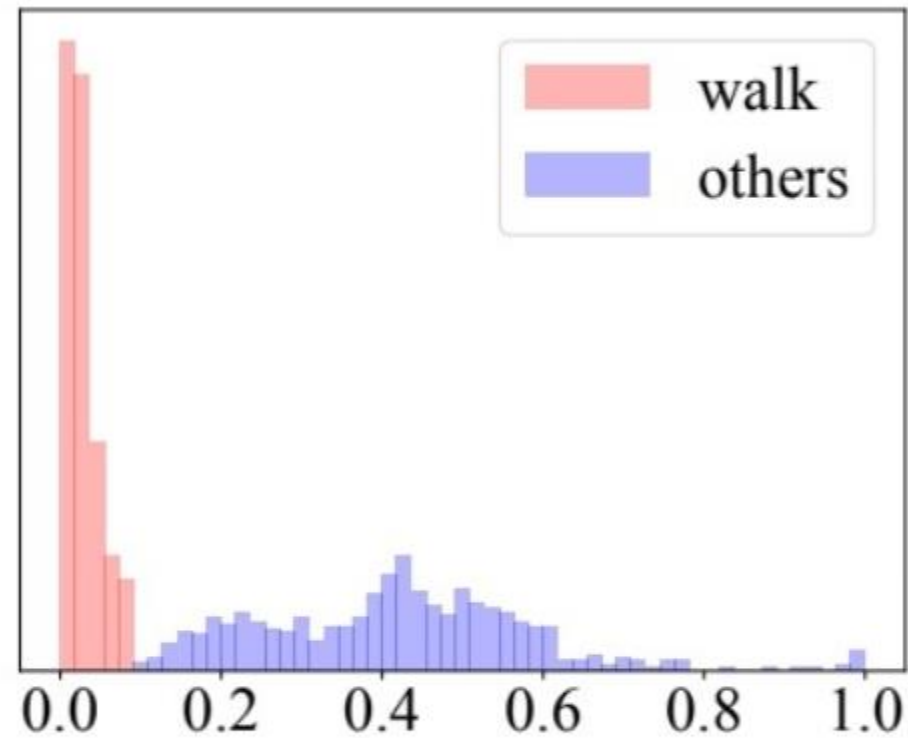


Figure 4: Normalized anomalousness score distributions.

Q2. Explainability

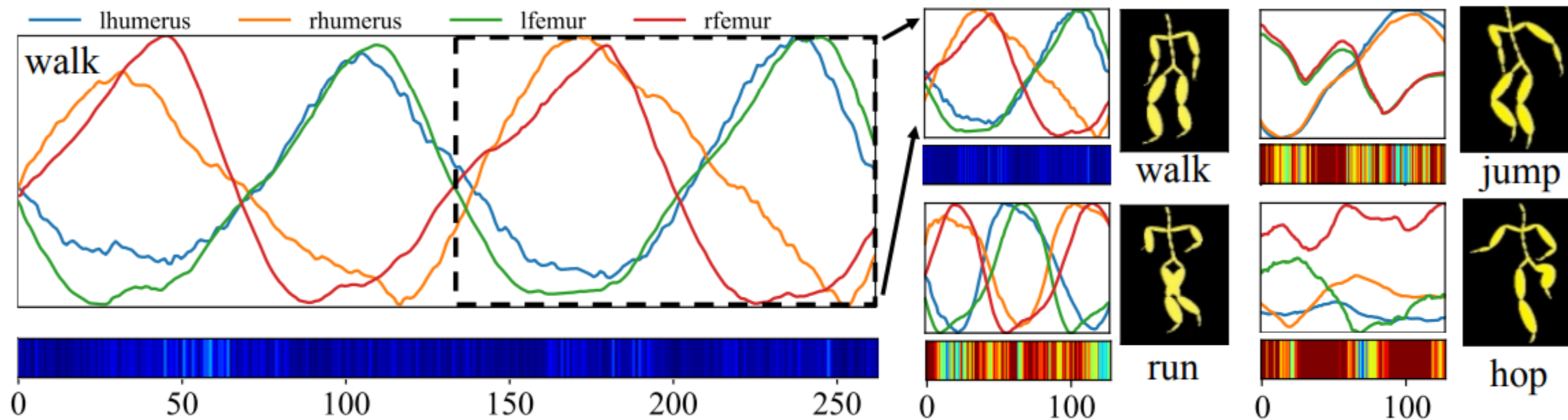
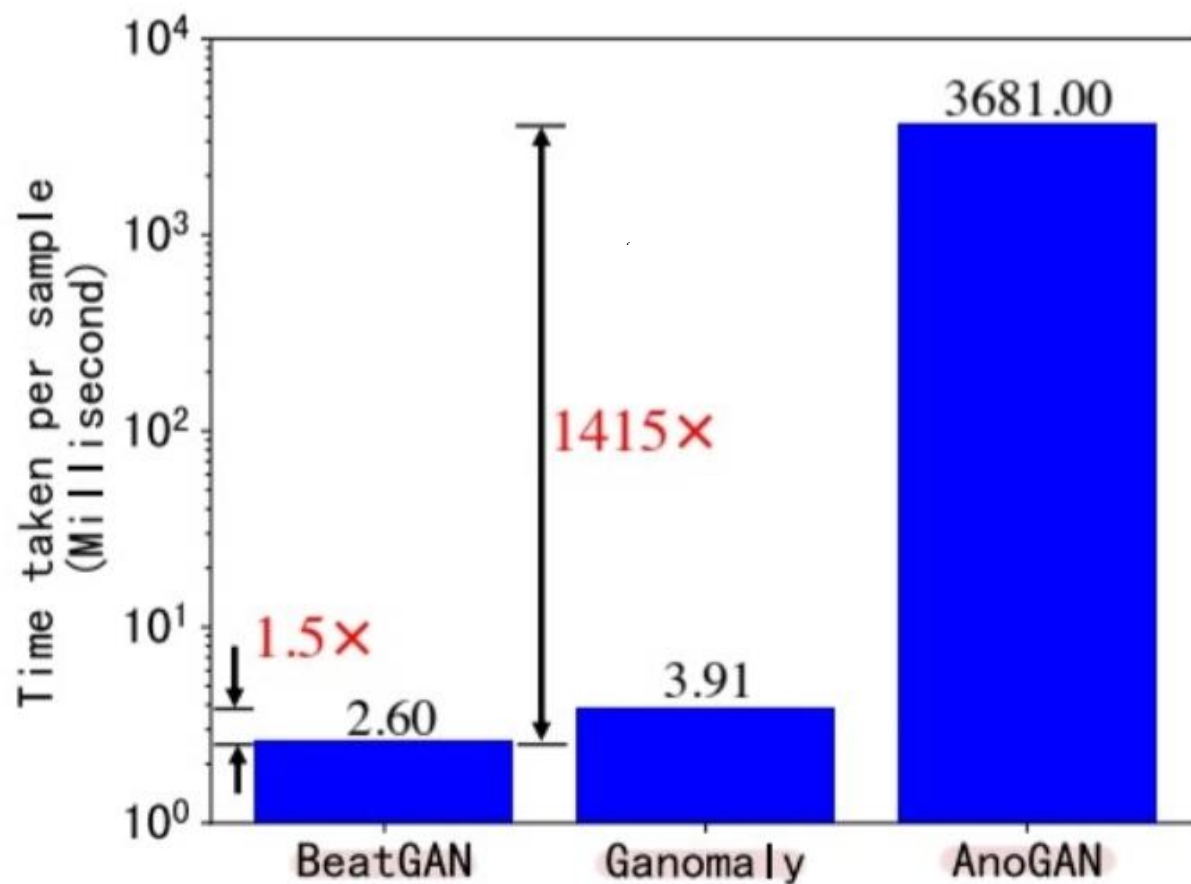


Figure 3: Example of anomaly detection on motion capture time series(4-dimensions). The right side shows the original time series and heatmaps to pinpoint the anomalies of jumping/running/hopping from walking motions.

Q3. Efficiency



BeatGAN only takes 2.6ms per beat, which is 1.5x faster than Ganomaly and 1415x faster than AnoGAN.

Figure 5: BeatGAN has fast inference (2.6ms).

3. 결론

Conclusions

We propose an anomaly detection algorithm for anomalous beats based on adversarially generated time series. BeatGAN has the following advantages: 1) **Unsupervised**: it is applicable even when labels are unavailable; 2) **Effectiveness**: BeatGAN outperforms baselines in both accuracy and inference speed, achieving accuracy of nearly 0.95 AUC on ECG data and very fast inference (2.6 ms per beat). 3) **Explainability**: BeatGAN pinpoints the anomalous ticks as shown in Fig 1; 4) **Generality**: BeatGAN also successfully detects unusual motions in multivariate motion-capture database.

References

ROC curve와 AUC 이해하기 (YeEunOh) <https://dsdoris.medium.com/roc-curve%EC%99%80-auc-%EC%9D%B4%ED%95%B4%ED%95%98%EA%B8%B0-126978d80a9e>
희소하고 긴 시계열 데이터의 동적 시간 워핑 거리 상계값 개선 서장혁 /Janghyuk Seo 1, 정우환 /Woohwan Jung 2, 심규석 /Shim, Kyuseok 3
최윤제 연구원 Naver D2 발표자료 (GAN) <https://www.slideshare.net/NaverEngineering/1-gangenerative-adversarial-network>
Ki's blog, Autoencoder based Anomaly Detection <https://kh-kim.github.io/blog/2019/12/15/Autoencoder-based-anomaly-detection.html>
ITT's tech-blog, GAN(Generative Adversarial Networks) https://kjhov195.github.io/2020-03-09-generative_adversarial_network/
SPRi, 인공지능 기반 이상감지 기술 https://spri.kr/posts/view/23193?code=industry_trend