Introduction to Anomaly Detection

2021. 10. 15

Data Mining & Quality Analytics Lab.

발표자 : 김서연

발표자 소개

- **김서연 (Seo-Yeon Kim)**
  - 고려대학교 산업경영공학부 재학 중
  - Data Mining & Quality Analytics Lab
  - 석사과정 (2020.03 ~)

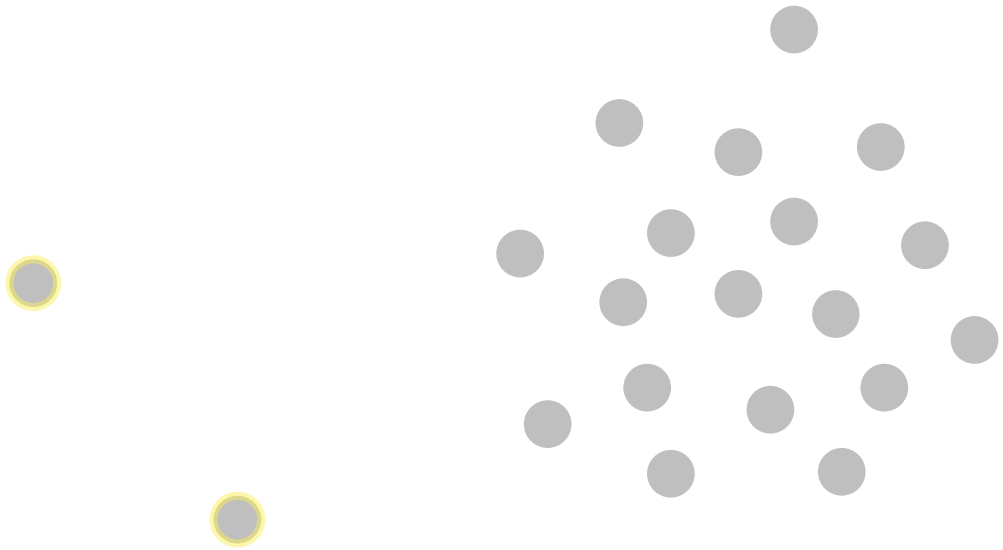
- 관심 연구 분야
  - Machine Learning & Deep Learning Algorithms
  - Time Series Data Analysis
  - Explainable AI

- E-mail : joanne2kim@korea.ac.kr

Contents

Introduction to Anomaly Detection

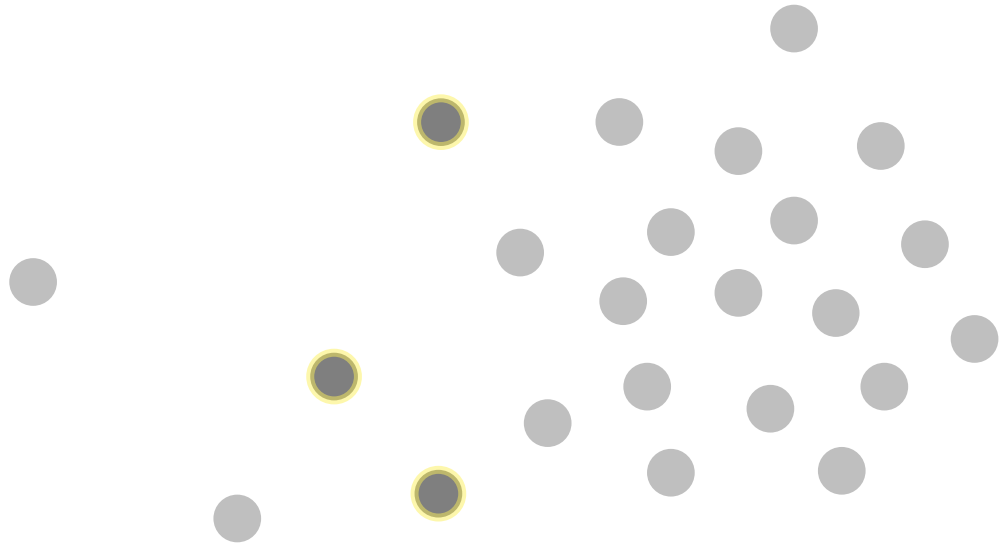
이상치란?



Outlier : 데이터의 전체적인 패턴에서 벗어난 관측치

Introduction to Anomaly Detection

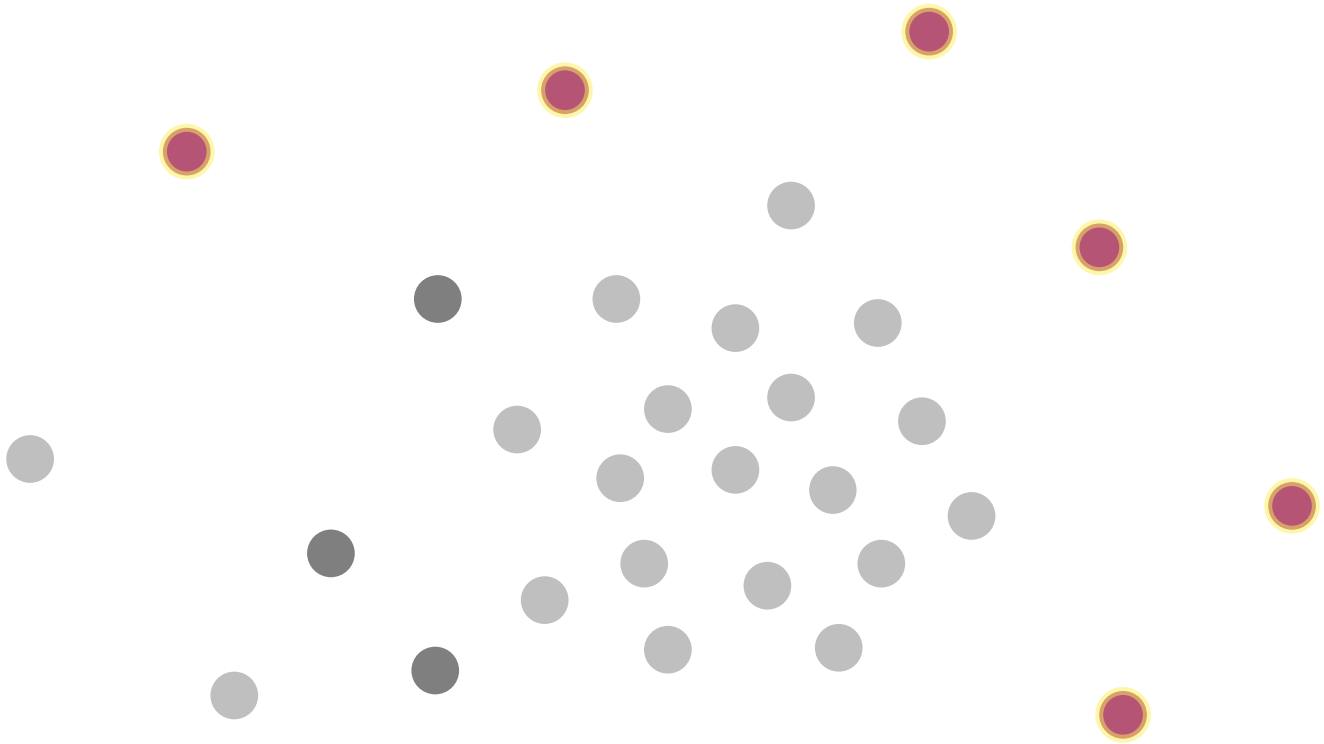
이상치란?



Novelty : 본질적인 데이터는 같지만 유형이 다른 관측치

Introduction to Anomaly Detection

이상치란?



Anomaly : 대부분의 데이터와 본질적인 특성이 다른 관측치
전혀 다른 방식으로 생성되었을 것으로 추정되는 관측치

Introduction to Anomaly Detection

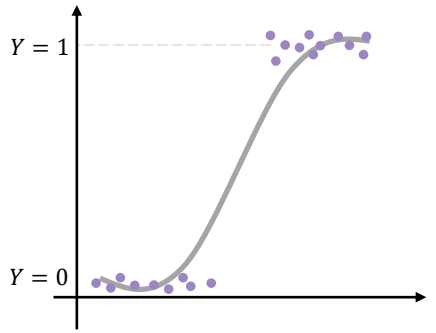
지도 학습 기반 이상치 탐지

X_1	X_2	X_3	···	X_{n-1}	X_n	Y
0.1	0.4	0.2	···	0.3	0.7	0
0.25	0.5	0.3	···	0.4	0.8	1
···	···	···	···	···	···	···
0.2	0.3	0.5	···	0.6	0.9	1

Introduction to Anomaly Detection

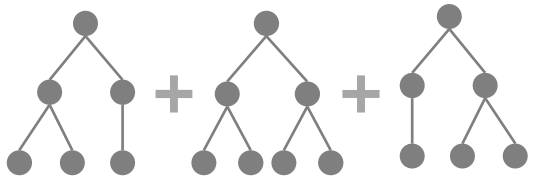
지도 학습 기반 이상치 탐지

X_1	X_2	X_3	···	X_{n-1}	X_n
0.1	0.4	0.2	···	0.3	0.7
0.25	0.5	0.3	···	0.4	0.8
···	···	···	···	···	···
0.2	0.3	0.5	···	0.6	0.9



로지스틱 회귀 분석

그래디언트 부스팅



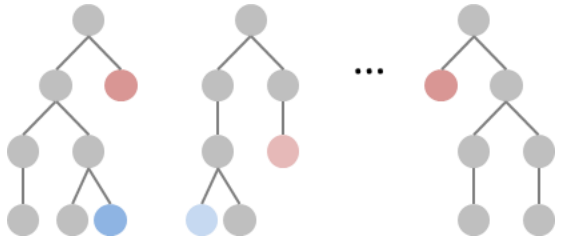
Y
0
1
···
1

Introduction to Anomaly Detection

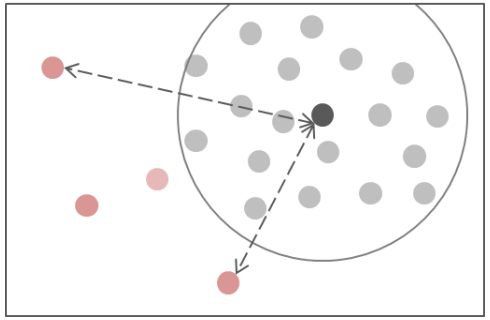
비지도 학습 기반 이상치 탐지

X_1	X_2	X_3	···	X_{n-1}	X_n
0.1	0.4	0.2	···	0.3	0.7
0.25	0.5	0.3	···	0.4	0.8
···	···	···	···	···	···
0.2	0.3	0.5	···	0.6	0.9

Isolation Forest



Local Outlier Factor



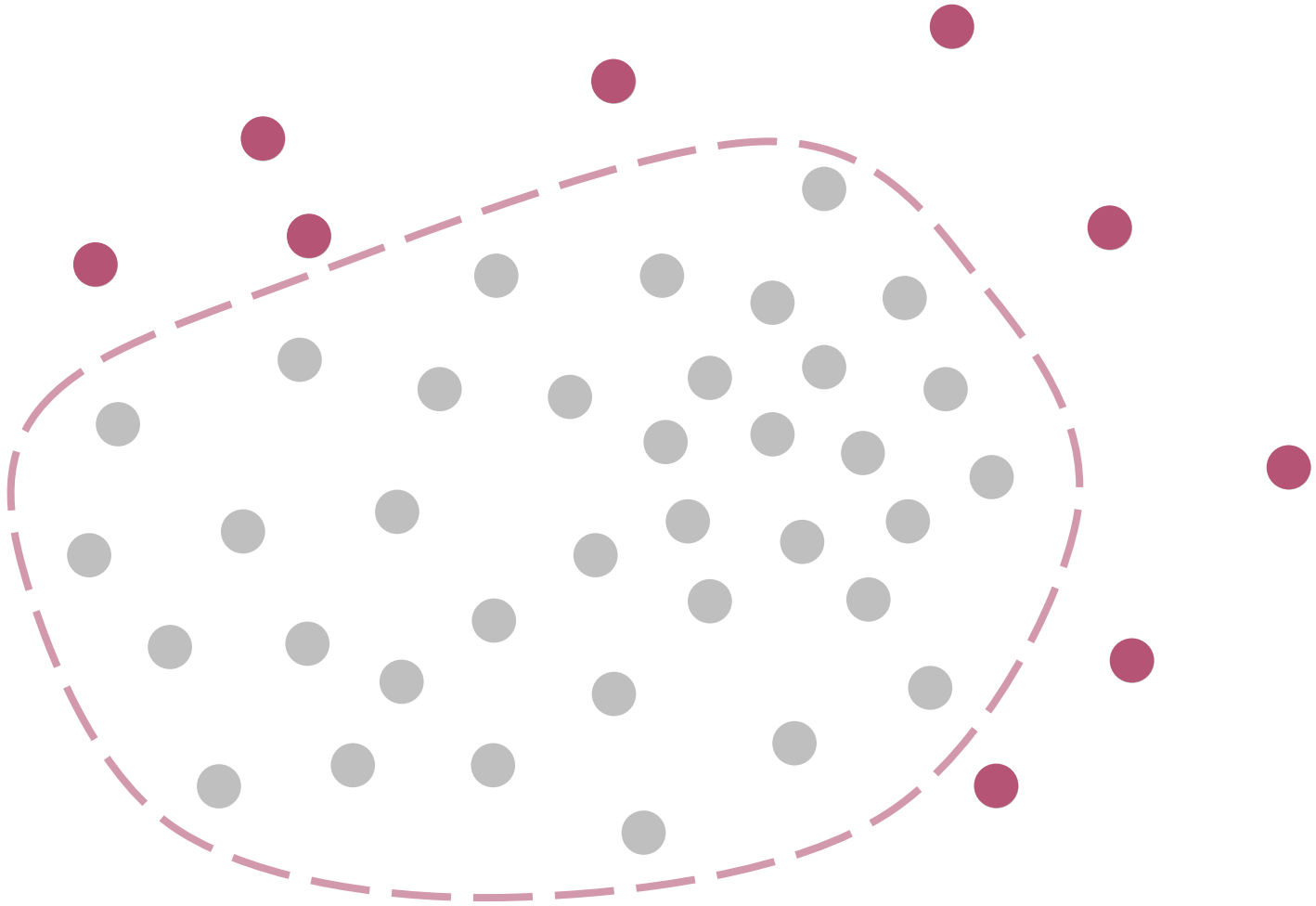
Introduction to Anomaly Detection

딥러닝 기반 이상치 탐지 (Deep Anomaly Detection)

딥러닝 기반 방법론 제안

Introduction to Anomaly Detection

딥러닝 기반 이상치 탐지 (Deep Anomaly Detection)

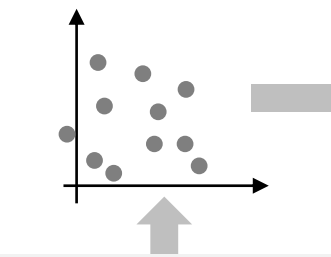


Introduction to Anomaly Detection

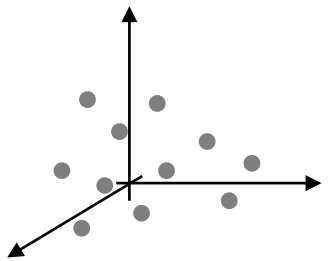
딥러닝 기반 이상치 탐지 (Deep Anomaly Detection)

- 딥러닝 기반 이상치 탐지의 세 가지 접근

차원 축소 기반 이상치 탐지



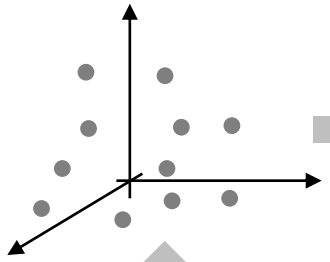
차원 축소 방법론
ex) 주성분 분석
(principal component analysis)



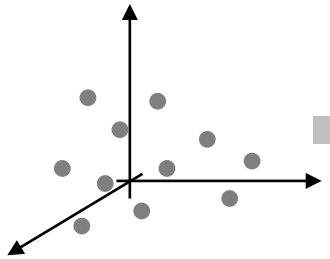
이상치 점수 함수

관측치 번호	이상치 점수
1	#
2	#
…	…
N-1	#
N	#

데이터 특성 기반 이상치 탐지



데이터 특성을 학습하는 방법론
ex) Autoencoder, GAN, 등



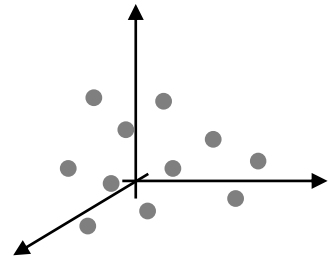
데이터
특성을
기반으로
이상치 판단

End-to-end 이상치 탐지

관측치 번호	이상치 점수
1	#
2	#
…	…
N-1	#
N	#

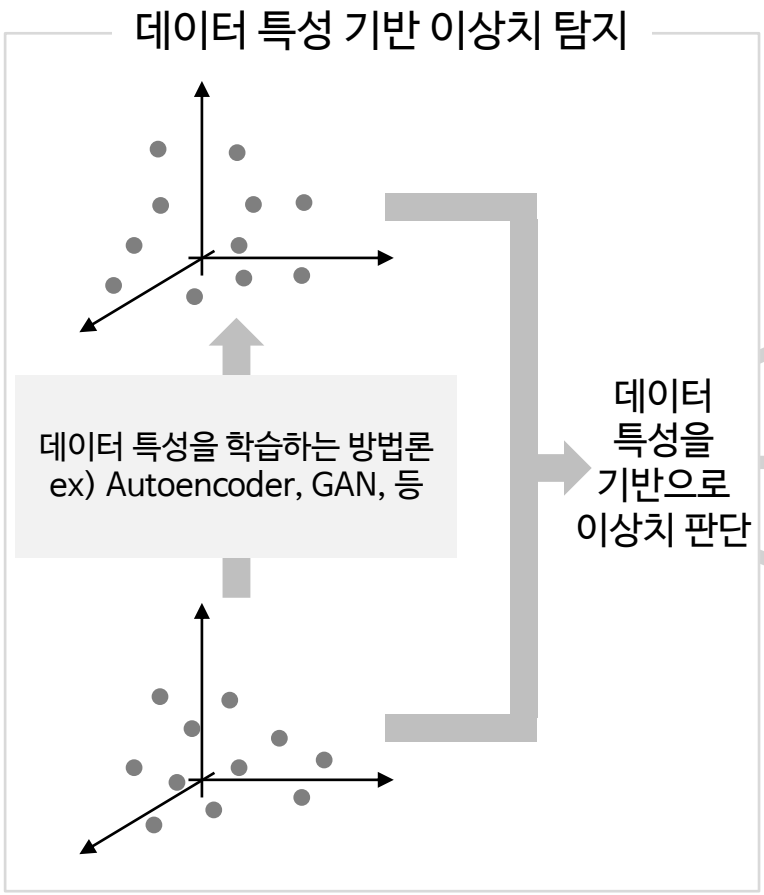
이상치 점수 예측

데이터 특성 파악



Introduction to Anomaly Detection

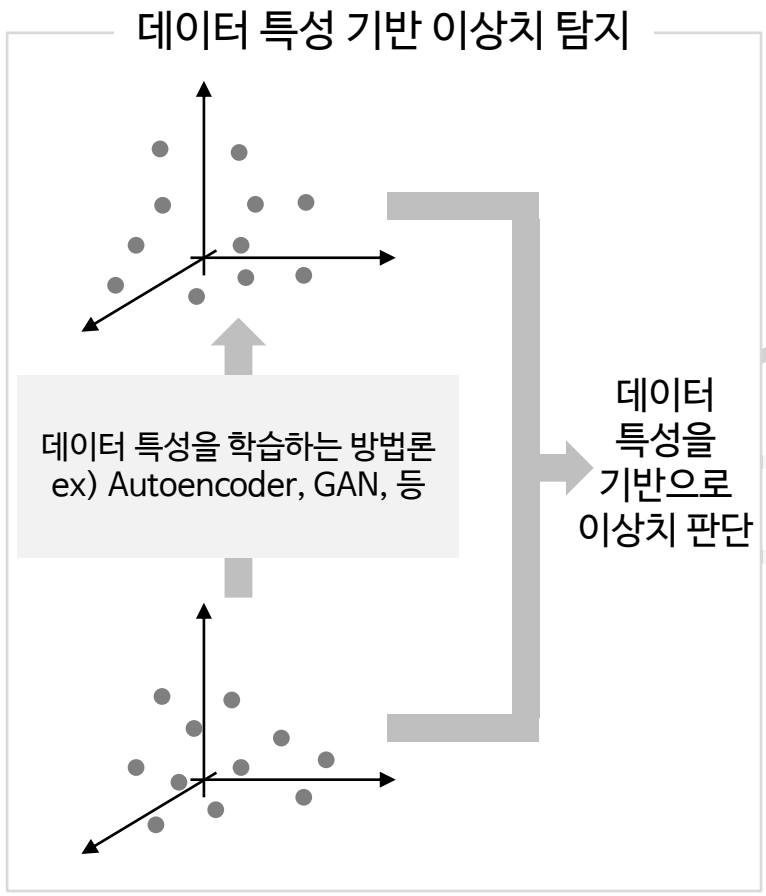
딥러닝 기반 이상치 탐지 (Deep Anomaly Detection)



Autoencoder based Anomaly Detection

GAN based Anomaly Detection

Self-Supervised Learning based Anomaly Detection



Autoencoder based Anomaly Detection

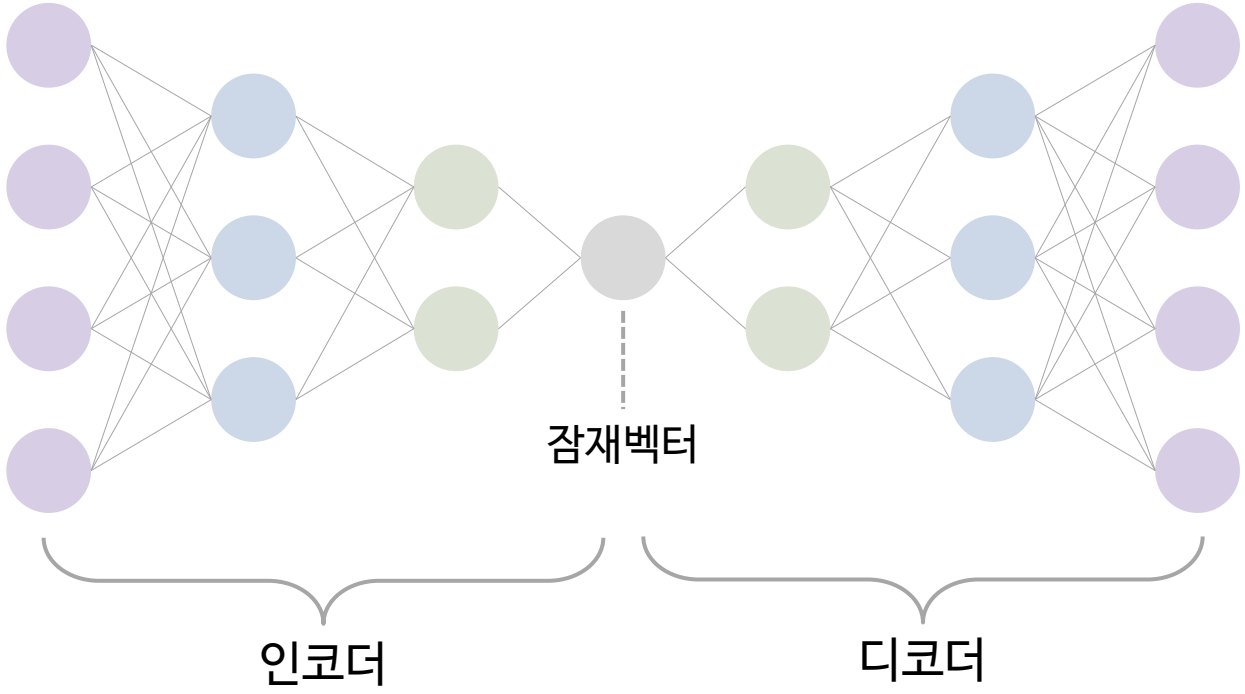
GAN based Anomaly Detection

Self-Supervised Learning based Anomaly Detection

Autoencoder based Anomaly Detection

오토인코더란?

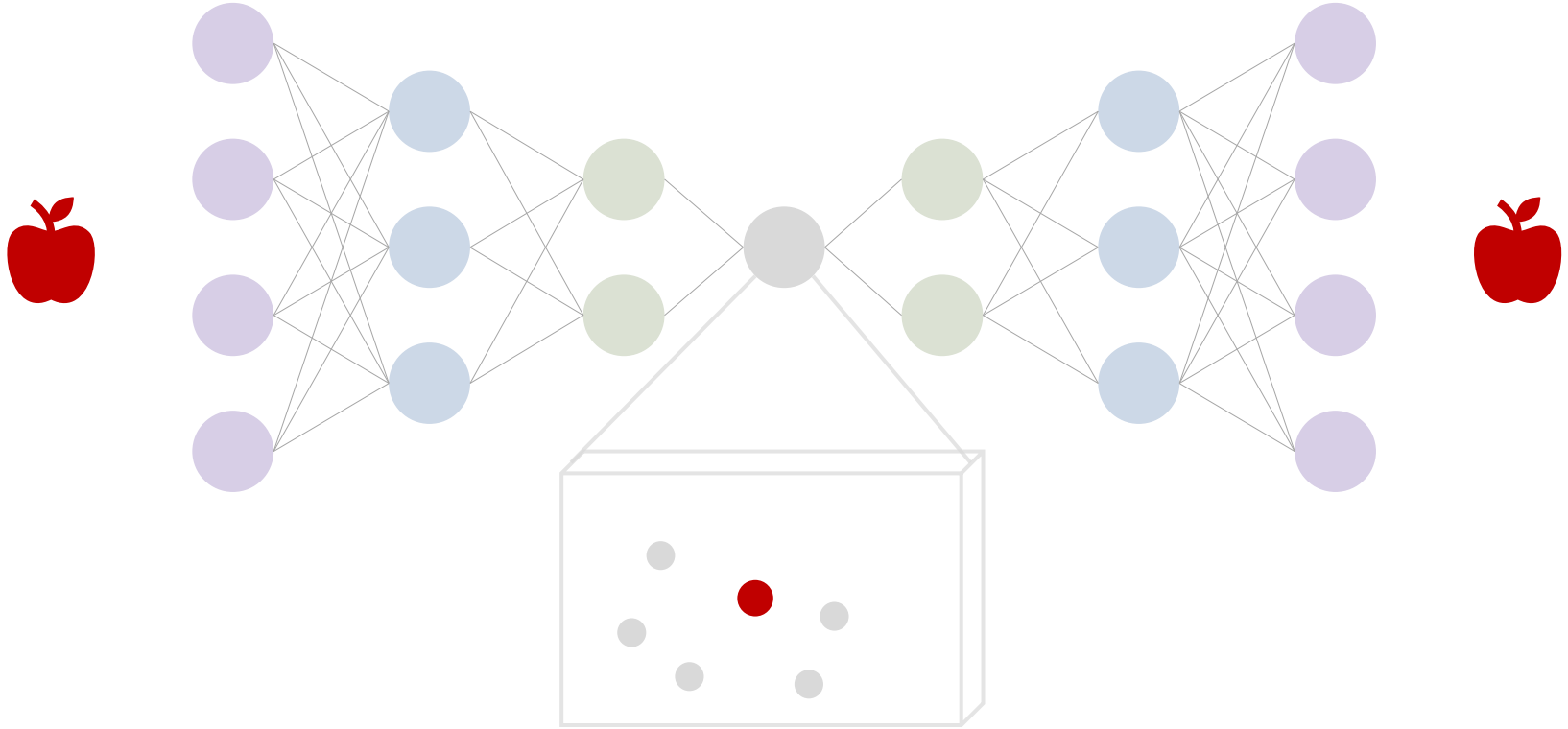
- 입력된 데이터의 특성을 요약하는 **인코더**와 요약된 정보를 복원하는 **디코더**의 형태로 구성
- 데이터가 잘 복원된 경우 저차원의 데이터 특성 공간을 파악할 수 있음



Autoencoder based Anomaly Detection

오토인코더란?

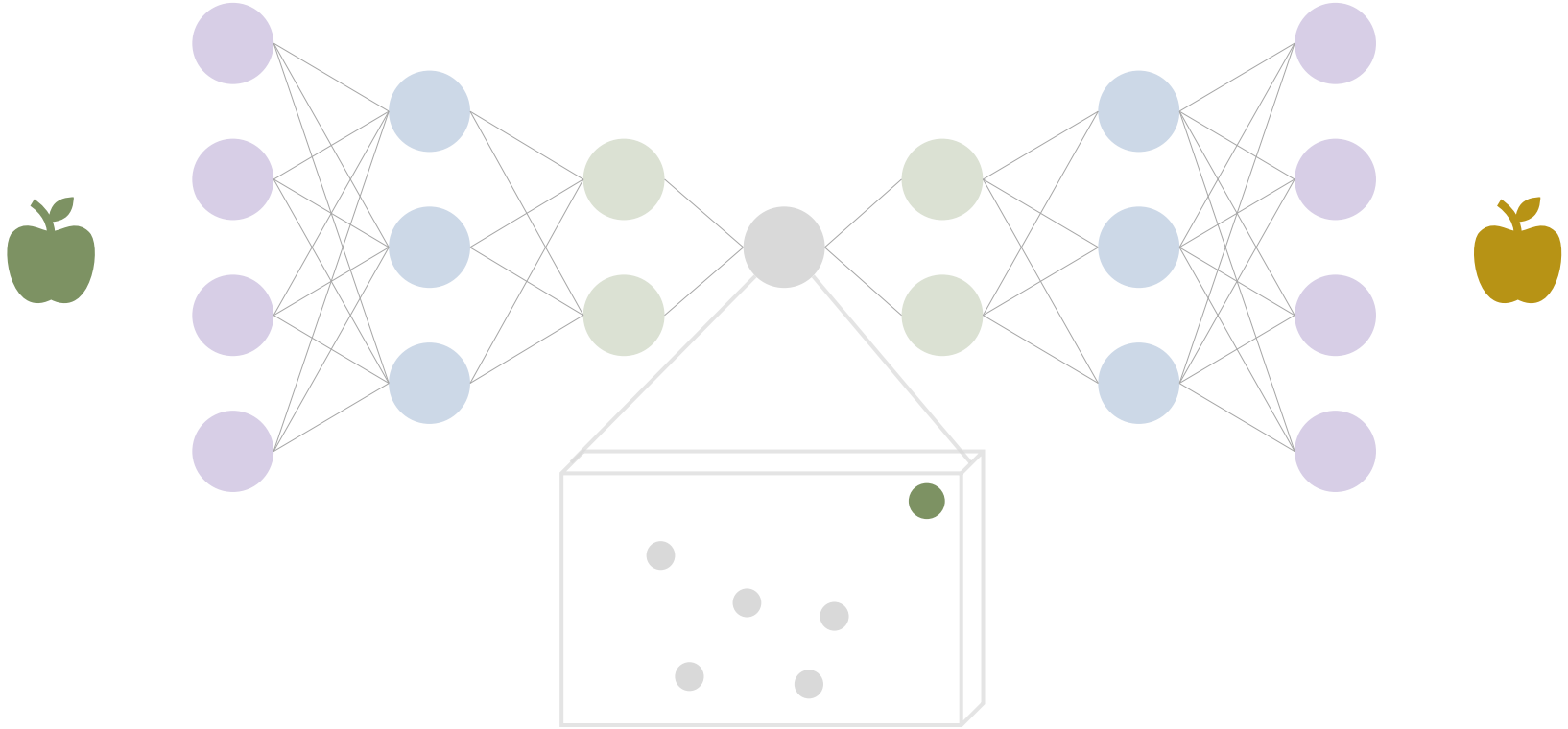
- 입력된 데이터의 특성을 요약하는 **인코더**와 요약된 정보를 복원하는 **디코더**의 형태로 구성
- 가정 : 정상 관측치들은 불량 관측치보다 더 잘 복원될 것이다.



Autoencoder based Anomaly Detection

오토인코더란?

- 입력된 데이터의 특성을 요약하는 **인코더**와 요약된 정보를 복원하는 **디코더**의 형태로 구성
- 가정 : 정상 관측치들은 불량 관측치보다 더 잘 복원될 것이다.



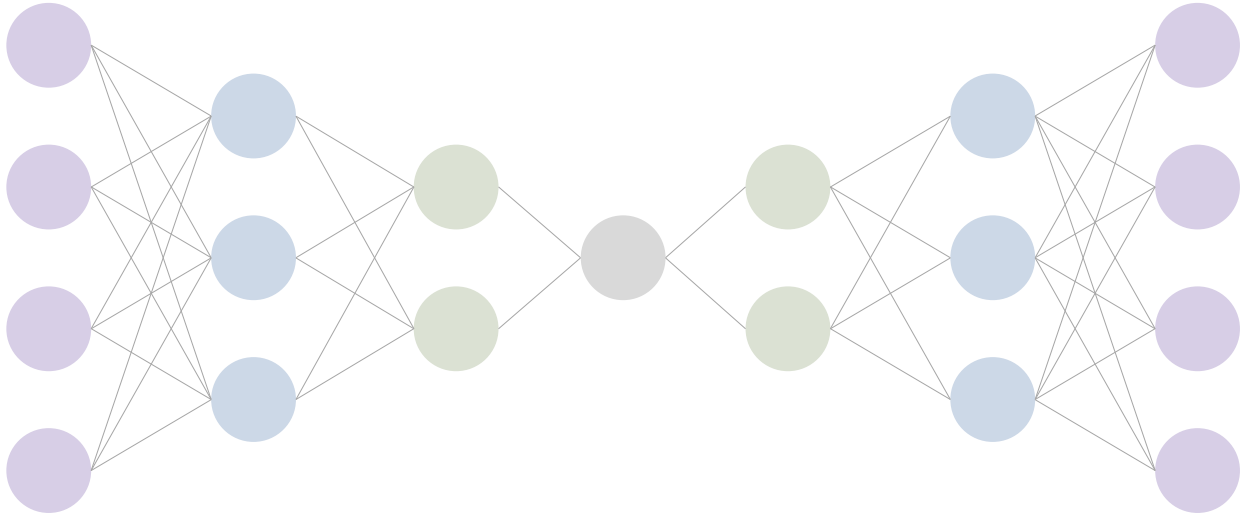
Autoencoder based Anomaly Detection

오토인코더로 이상치 탐지하기

- 입력 데이터와 복원된 데이터 사이의 차이 계산

X_1	X_2	X_3
1	4	2
2.5	5	3
1.5	5	4
2	3	5

X



X_1'	X_2'	X_3'
1	3	2
2	4	2
0	4	2
2	3	0.5

X’

Autoencoder based Anomaly Detection

오토인코더로 이상치 탐지하기

- 입력 데이터와 복원된 데이터 사이의 차이 계산 (재구축 오차, Reconstruction Error)

$\left\| \right.$

X_1	X_2	X_3
1	4	2
2.5	5	3
1.5	5	4
2	3	5

X

$-$

X_1'	X_2'	X_3'
1	3	2
2	4	2
0	4	2
2	3	0.5

X'

$\left\| \right\|_2^2 = \sum ($

X_1-X_1'	X_2-X_2'	X_3-X_3'
0	1	0
0.5	1	1
1.5	1	2
0	0	0

$)^2 =$

$X - X'$

$\|\|X - X'\|\|_2^2$
1
2.25
7.25
0

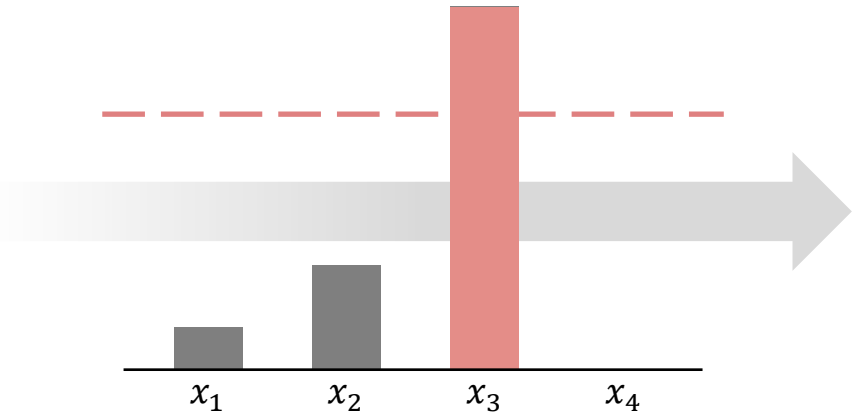
재구축 오차
이상치 점수

Autoencoder based Anomaly Detection

오토인코더로 이상치 탐지하기

- 재구축 오차가 특정 임계값보다 큰 경우 불량 관측치로 판단

$\|X - X'\|_2^2$
1
2.25
7.25
0

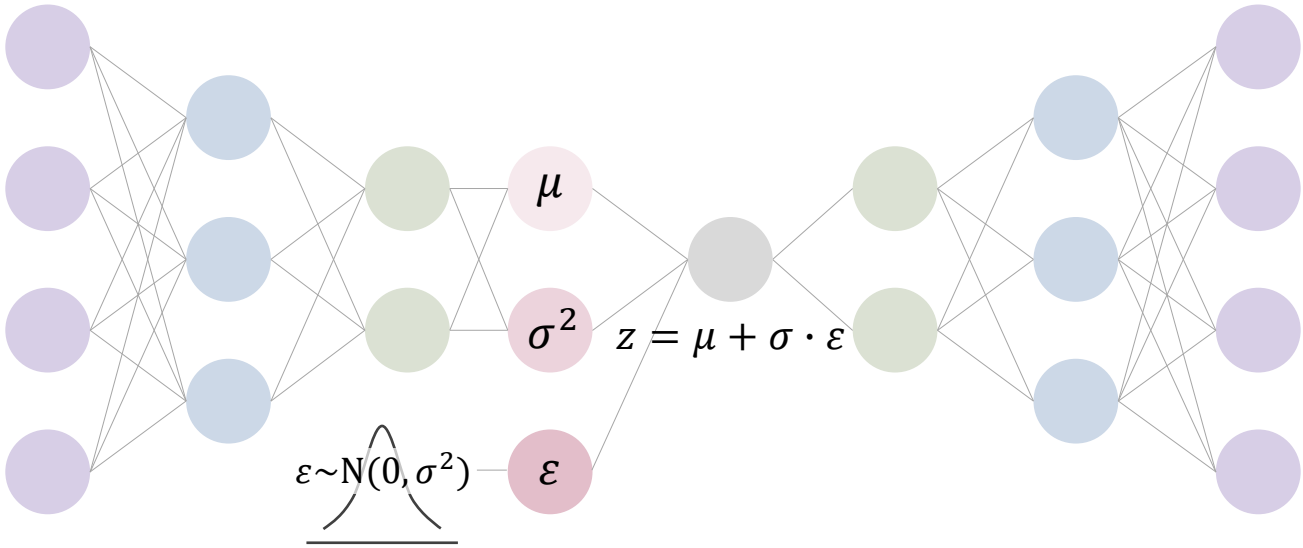


예측결과
정상
정상
불량
정상

Autoencoder based Anomaly Detection

오토인코더 계열 모델

- **변이형 오토인코더 (Variational Autoencoder)**
- 잠재 벡터(z)가 정규분포와 유사해지도록 제약식을 추가한 후, 원본 데이터로 복원하는 과정 진행
- 제약식 추가 과정을 위하여 요약된 데이터의 평균과 표준편차를 계산함

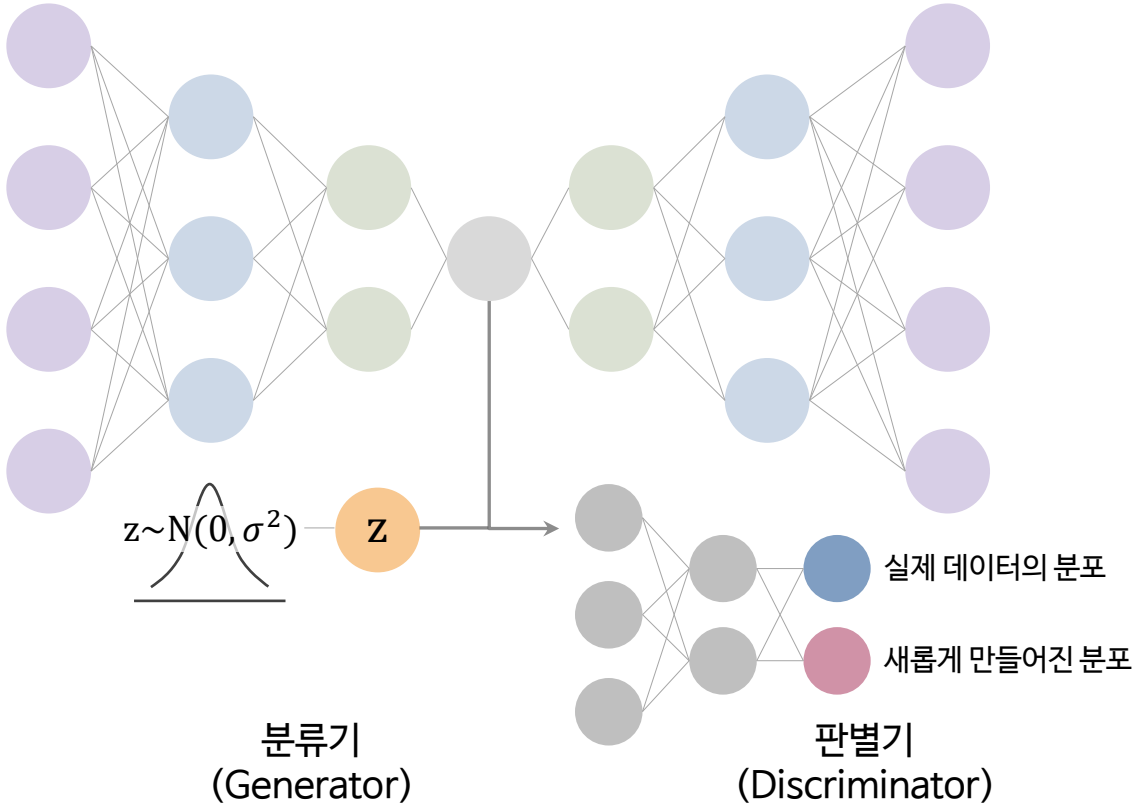


Autoencoder based Anomaly Detection

오토인코더 계열 모델

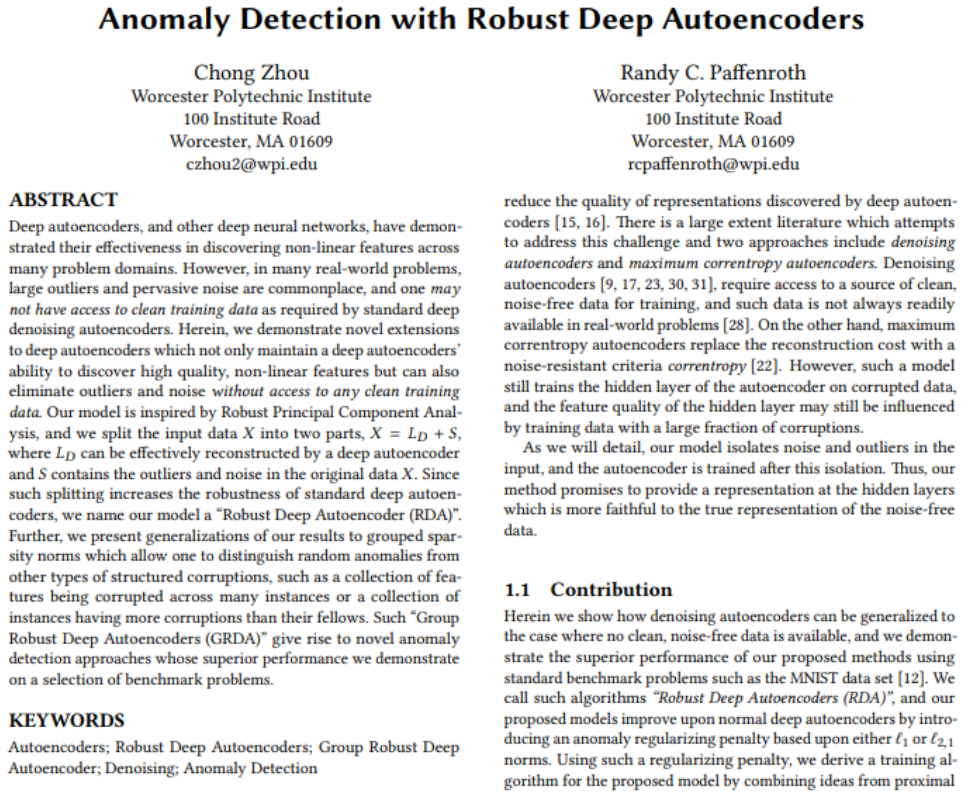
- **적대적 오토인코더 (Adversarial Autoencoder)**

- 요약된 잠재 벡터와 정규 분포에서 임의로 샘플링한 데이터를 구분하는 판별기를 추가하여 학습에 도움을 주는 구조



Autoencoder based Anomaly Detection

- Anomaly Detection with Robust Deep Autoencoders
- ACM SIGKDD International Conference에서 2017년 발표된 논문
- 2021년 10월 13일 기준 596회 인용



Anomaly Detection with Robust Deep Autoencoders

Chong Zhou
Worcester Polytechnic Institute
100 Institute Road
Worcester, MA 01609
czhou2@wpi.edu

Randy C. Paffenroth
Worcester Polytechnic Institute
100 Institute Road
Worcester, MA 01609
rcpaffenroth@wpi.edu

ABSTRACT

Deep autoencoders, and other deep neural networks, have demonstrated their effectiveness in discovering non-linear features across many problem domains. However, in many real-world problems, large outliers and pervasive noise are commonplace, and one *may not have access to clean training data* as required by standard deep denoising autoencoders. Herein, we demonstrate novel extensions to deep autoencoders which not only maintain a deep autoencoders' ability to discover high quality, non-linear features but can also eliminate outliers and noise *without access to any clean training data*. Our model is inspired by Robust Principal Component Analysis, and we split the input data X into two parts, $X = L_D + S$, where L_D can be effectively reconstructed by a deep autoencoder and S contains the outliers and noise in the original data X. Since such splitting increases the robustness of standard deep autoencoders, we name our model a "Robust Deep Autoencoder (RDA)". Further, we present generalizations of our results to grouped sparsity norms which allow one to distinguish random anomalies from other types of structured corruptions, such as a collection of features being corrupted across many instances or a collection of instances having more corruptions than their fellows. Such "Group Robust Deep Autoencoders (GRDA)" give rise to novel anomaly detection approaches whose superior performance we demonstrate on a selection of benchmark problems.

KEYWORDS

Autoencoders; Robust Deep Autoencoders; Group Robust Deep Autoencoder; Denoising; Anomaly Detection

reduce the quality of representations discovered by deep autoencoders [15, 16]. There is a large extent literature which attempts to address this challenge and two approaches include *denoising autoencoders* and *maximum correntropy autoencoders*. Denoising autoencoders [9, 17, 23, 30, 31], require access to a source of clean, noise-free data for training, and such data is not always readily available in real-world problems [28]. On the other hand, maximum correntropy autoencoders replace the reconstruction cost with a noise-resistant criteria *correntropy* [22]. However, such a model still trains the hidden layer of the autoencoder on corrupted data, and the feature quality of the hidden layer may still be influenced by training data with a large fraction of corruptions.

As we will detail, our model isolates noise and outliers in the input, and the autoencoder is trained after this isolation. Thus, our method promises to provide a representation at the hidden layers which is more faithful to the true representation of the noise-free data.

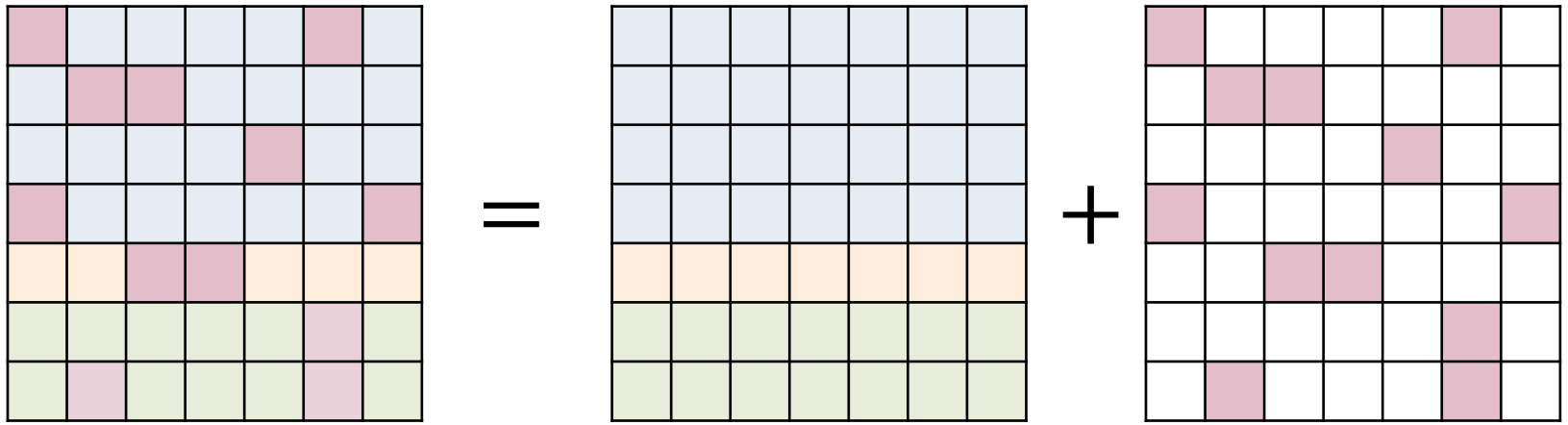
1.1 Contribution

Herein we show how denoising autoencoders can be generalized to the case where no clean, noise-free data is available, and we demonstrate the superior performance of our proposed methods using standard benchmark problems such as the MNIST data set [12]. We call such algorithms *"Robust Deep Autoencoders (RDA)"*, and our proposed models improve upon normal deep autoencoders by introducing an anomaly regularizing penalty based upon either ℓ_1 or $\ell_{2,1}$ norms. Using such a regularizing penalty, we derive a training algorithm for the proposed model by combining ideas from proximal

Autoencoder based Anomaly Detection

Robust Principal Component Analysis (RPCA)

- 주성분 분석 (PCA)의 이상치에 민감한 특징을 반영하여 제안된 방법론
- 주어진 데이터를 최저 계수 행렬 (Low Rank Matrix)과 희소 행렬 (Sparse Matrix)로 분할하는 과정



$\hat{X}$ = L + S

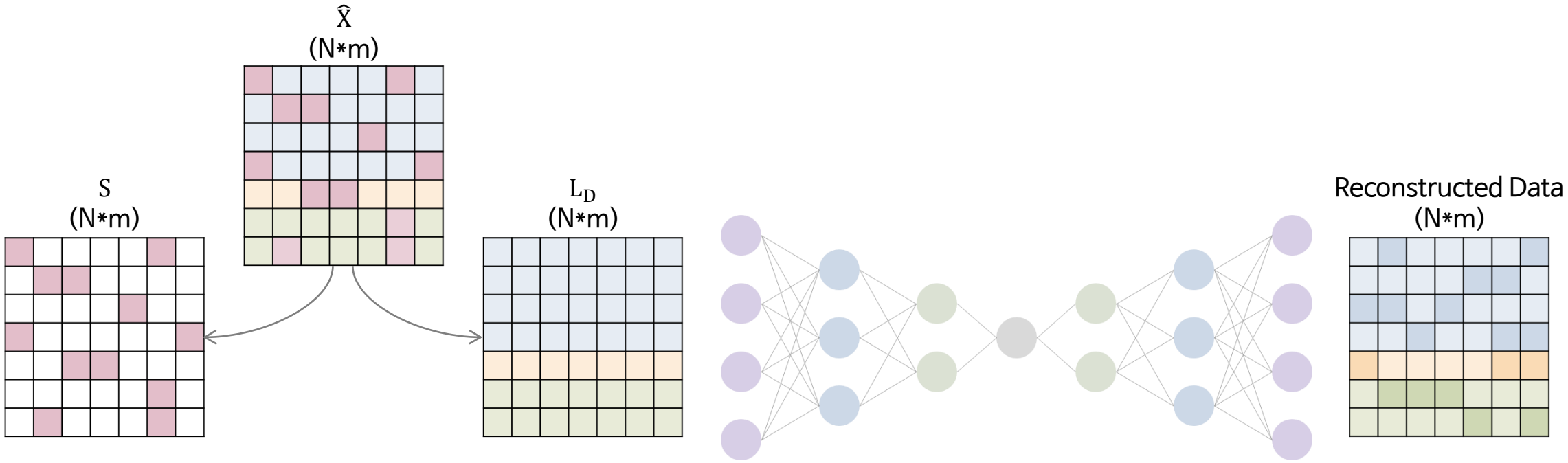
L: 저차원으로 특징이 요약된 행렬

S: 제거해야하는 이상치

Autoencoder based Anomaly Detection

Robust Deep Autoencoders (RDA)

- Autoencoder + Robust Principal Component Analysis (RPCA)
- RPCA를 통하여 얻은 깨끗해진 데이터로 오토인코더 모델 학습
- 이상치를 위하여 추가로 학습해야 하는 데이터 및 모델이 없음

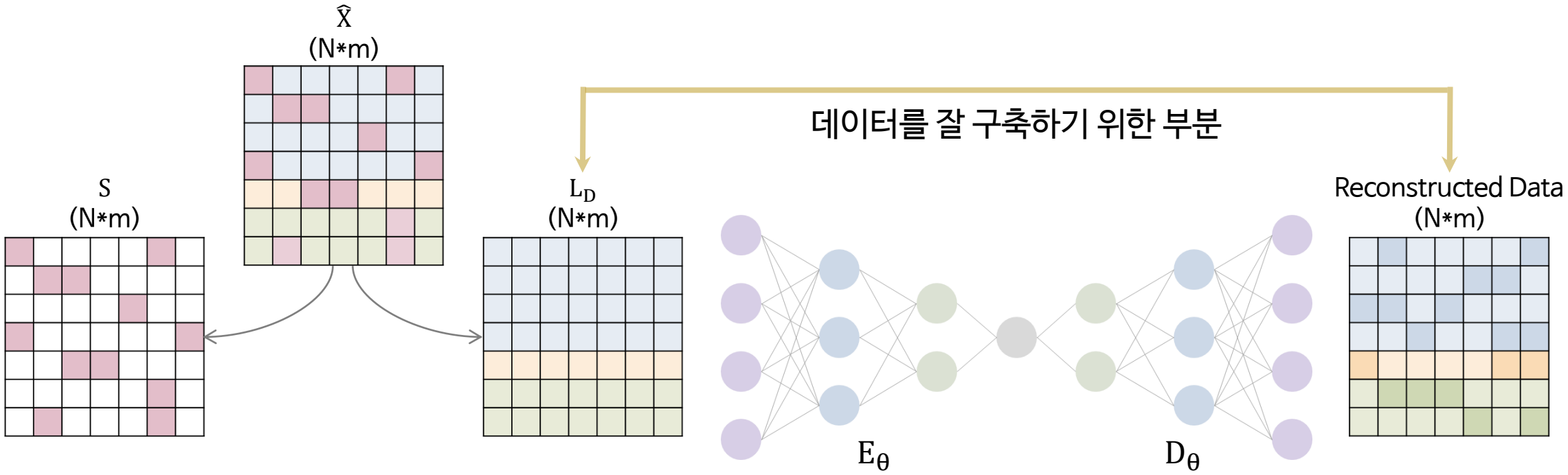


Autoencoder based Anomaly Detection

Robust Deep Autoencoders (RDA)

- Autoencoder + Robust Principal Component Analysis (RPCA)

$$\min_{\theta} \|\mathrm{L_D} - \mathrm{D_\theta}(\mathrm{E_\theta}(\mathrm{L_D}))\|_2 + \lambda\|\mathrm{S}\|_1$$
$$s.t.\ \mathrm{X} - \mathrm{L_D} - \mathrm{S} = 0$$



Autoencoder based Anomaly Detection

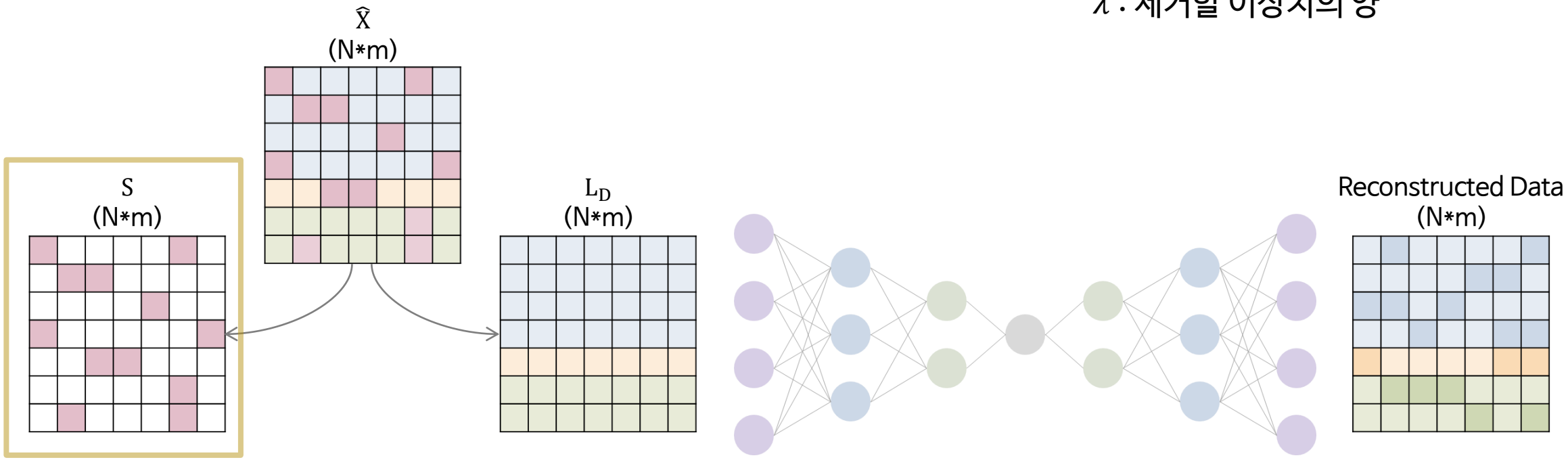
Robust Deep Autoencoders (RDA)

- Autoencoder + Robust Principal Component Analysis (RPCA)

$$\min_{\theta} \|\mathrm{L_D} - \mathrm{D_\theta}(\mathrm{E_\theta}(\mathrm{L_D}))\|_2 + \lambda\|\mathrm{S}\|_1$$

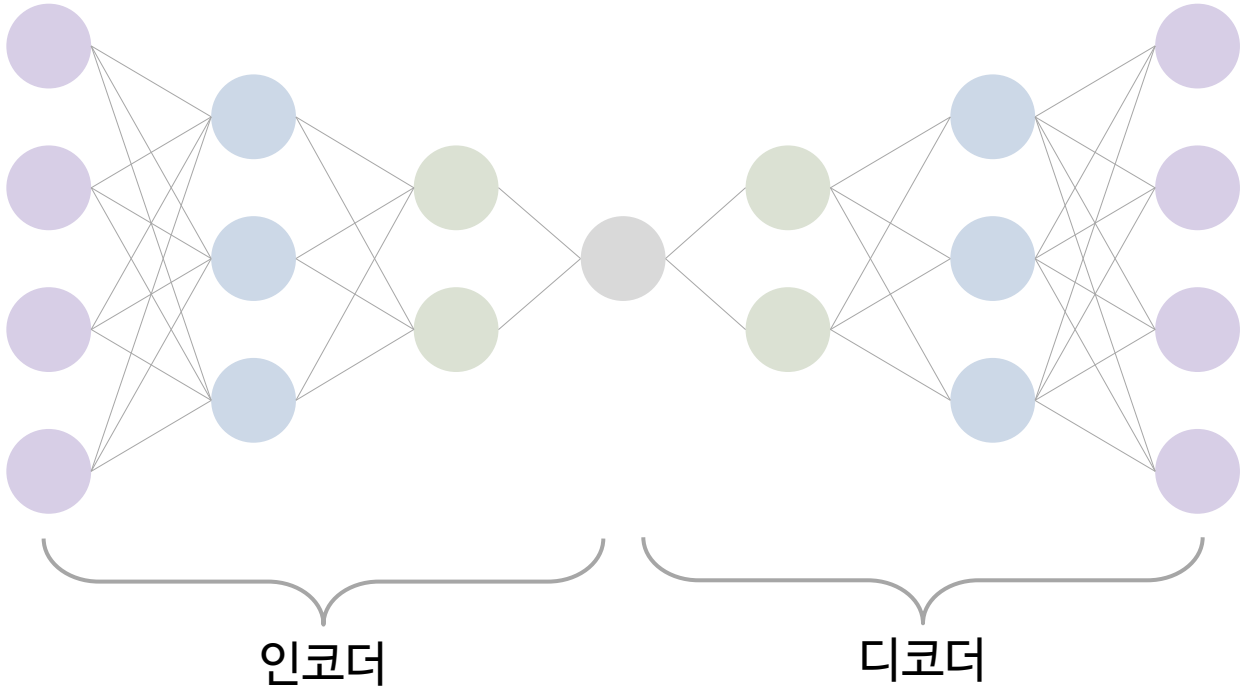
$$s.t.\ \mathrm{X} - \mathrm{L_D} - \mathrm{S} = 0$$

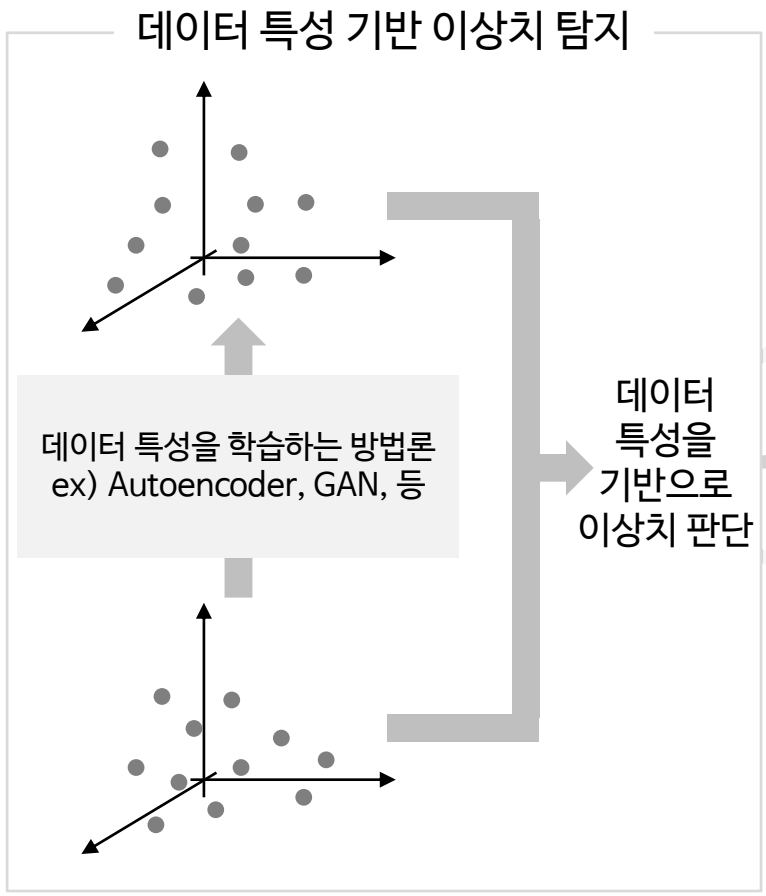
이상치를 제거하기 위한 부분
λ : 제거할 이상치의 양



Autoencoder based Anomaly Detection

- 저차원 공간에서 데이터의 특성을 파악하여 다시 잘 재구축 시키는 오토인코더 모델
- 다양한 오토인코더 기반 모델들을 활용하여 이상치 탐지 가능
  - 변이형 오토인코더, 적대적 오토인코더, Robust Deep Autoencoders (RDA)





Autoencoder based Anomaly Detection

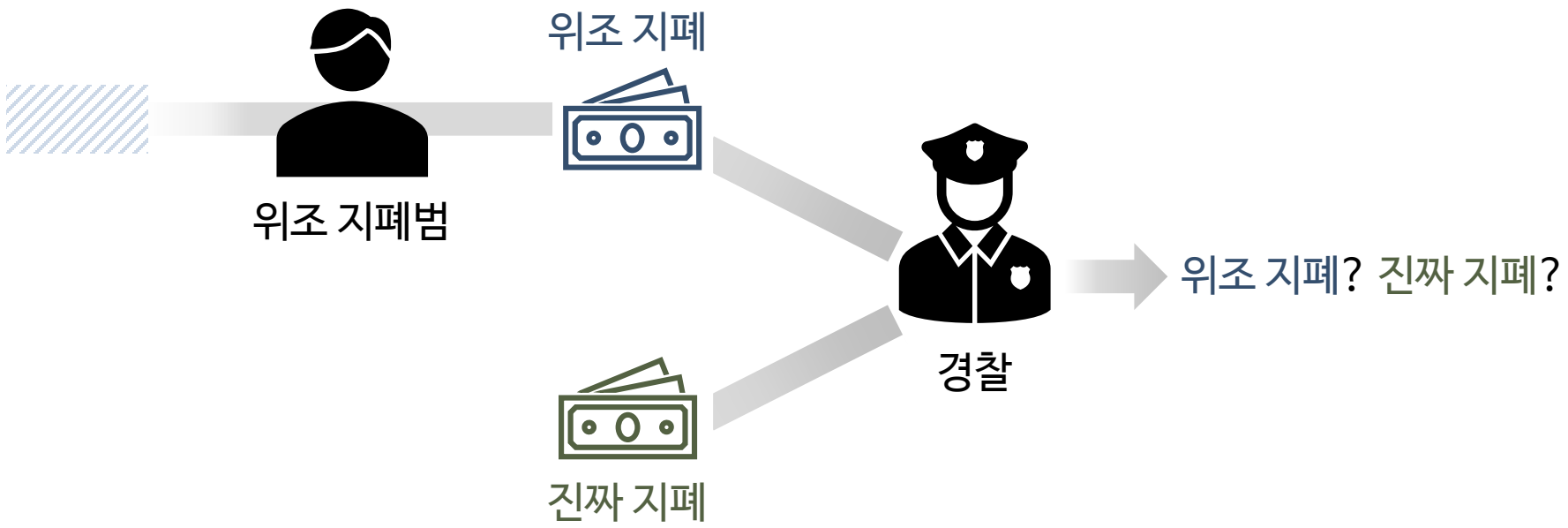
GAN based Anomaly Detection

Self-Supervised Learning based Anomaly Detection

GAN based Anomaly Detection

GAN이란?

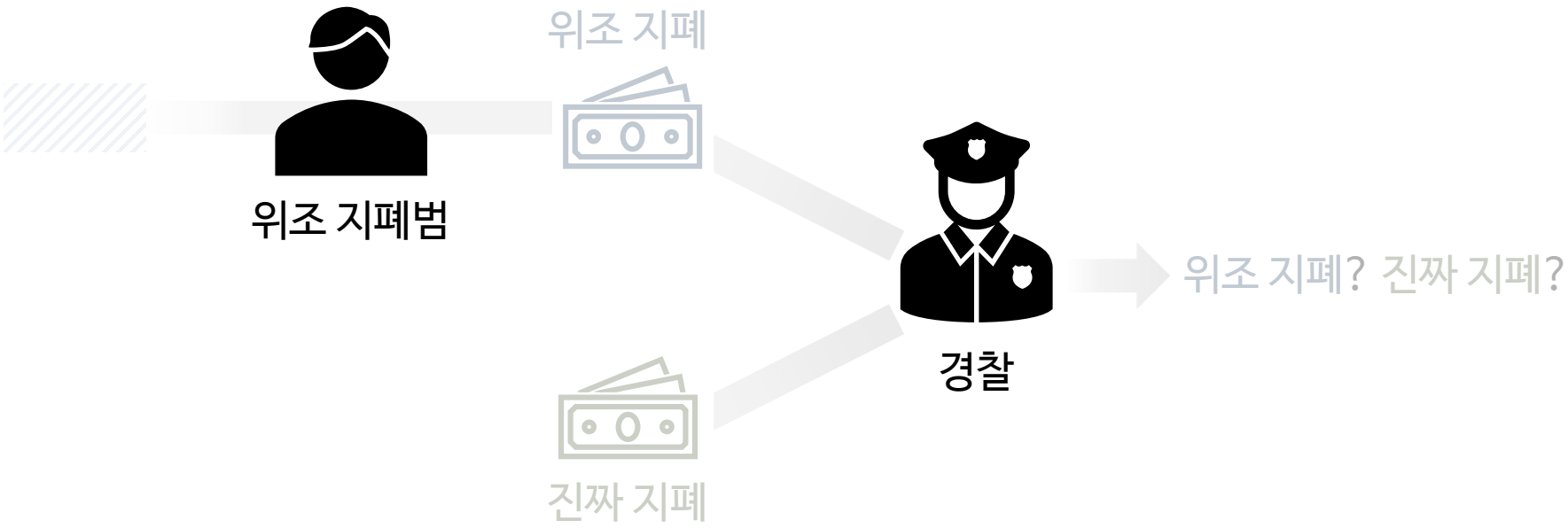
- 생성적 적대 신경망 (Generative Adversarial Network, GAN)



GAN based Anomaly Detection

GAN이란?

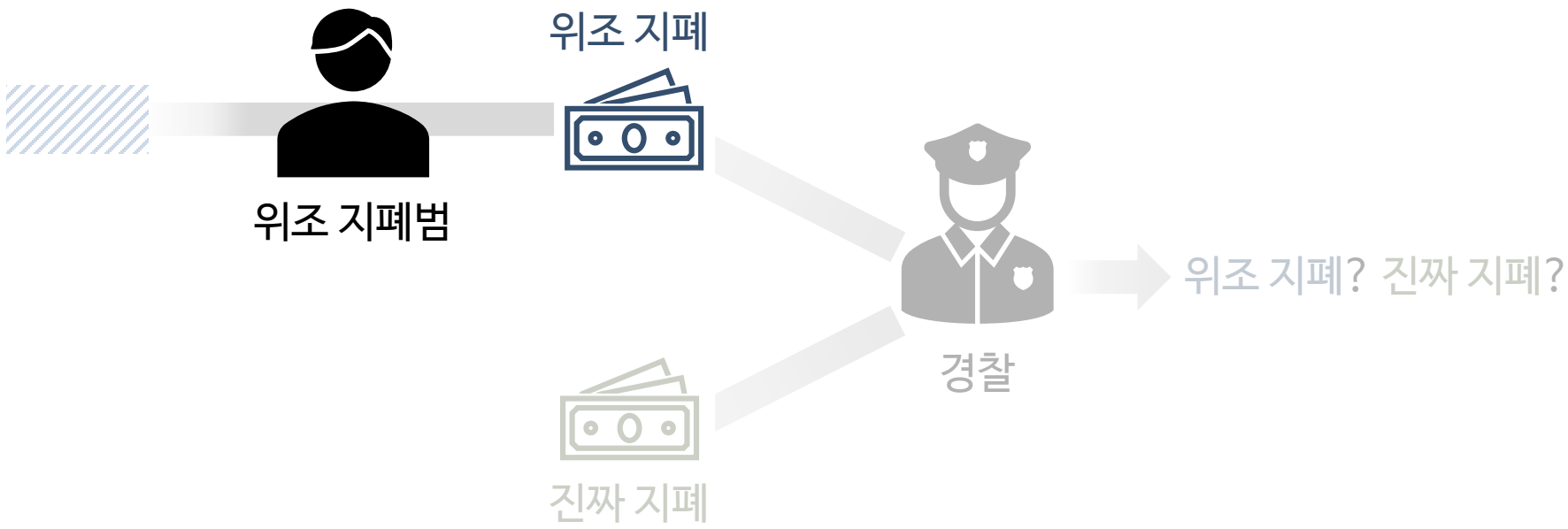
- 생성적 적대 신경망 (Generative Adversarial Network, GAN)



GAN based Anomaly Detection

GAN이란?

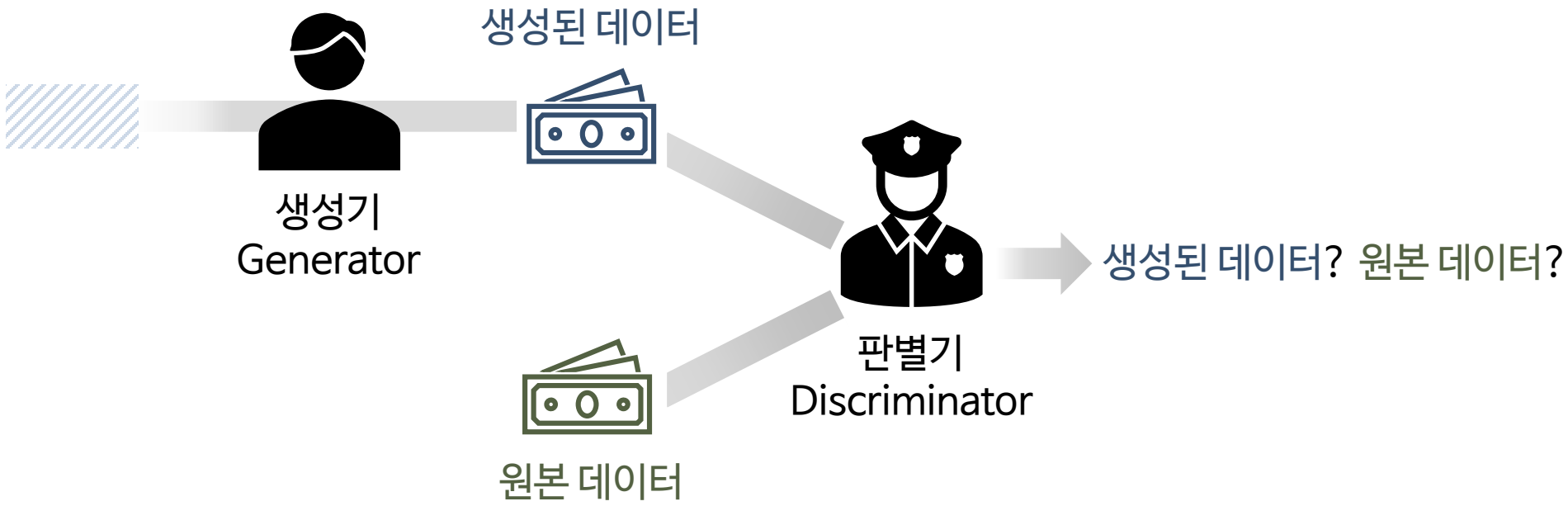
- **생성적** 적대 신경망 (**Generative** Adversarial Network, GAN)



GAN based Anomaly Detection

GAN이란?

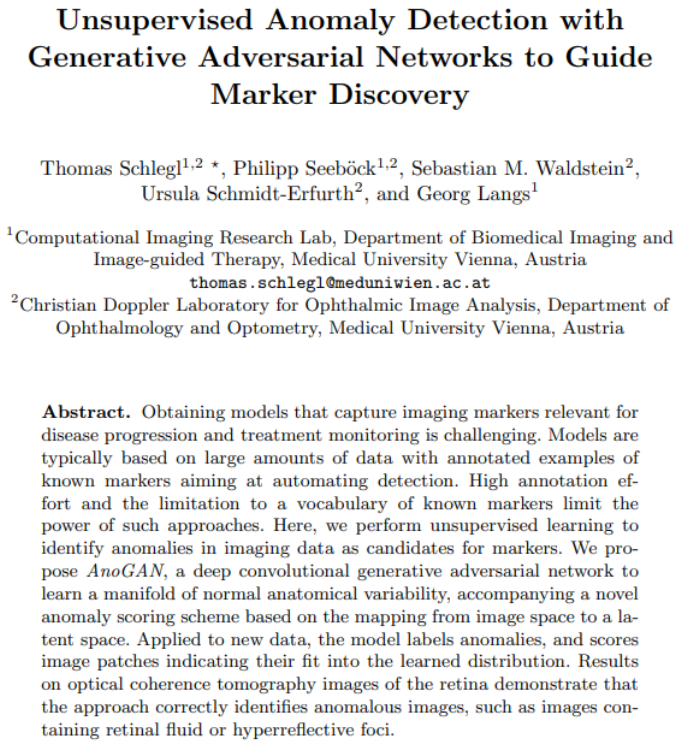
- 생성적 적대 신경망 (Generative Adversarial Network, GAN)



GAN based Anomaly Detection

AnoGAN

- Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery
- Information Processing in Medical Imaging에서 2017년 발표된 논문
- 2021년 10월 13일 기준 1137회 인용



Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery

Thomas Schlegl[1,2] *, Philipp Seeböck[1,2], Sebastian M. Waldstein[2], Ursula Schmidt-Erfurth[2], and Georg Langs[1]

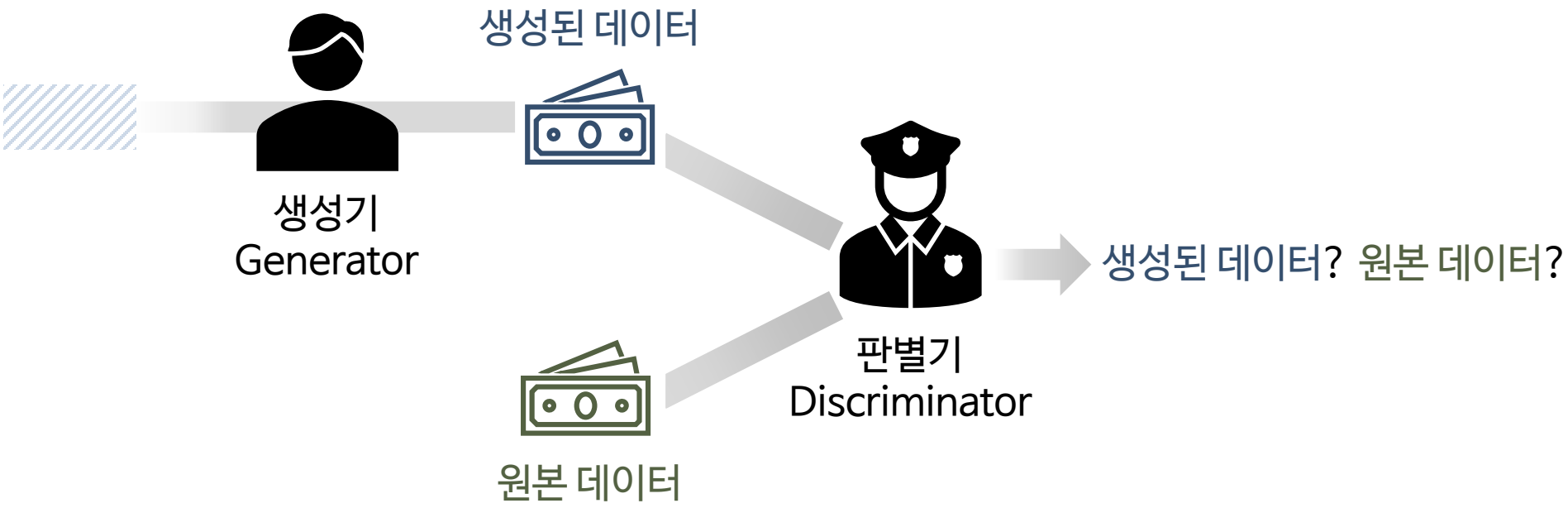
[1]Computational Imaging Research Lab, Department of Biomedical Imaging and Image-guided Therapy, Medical University Vienna, Austria
thomas.schlegl@meduniwien.ac.at
[2]Christian Doppler Laboratory for Ophthalmic Image Analysis, Department of Ophthalmology and Optometry, Medical University Vienna, Austria

Abstract. Obtaining models that capture imaging markers relevant for disease progression and treatment monitoring is challenging. Models are typically based on large amounts of data with annotated examples of known markers aiming at automating detection. High annotation effort and the limitation to a vocabulary of known markers limit the power of such approaches. Here, we perform unsupervised learning to identify anomalies in imaging data as candidates for markers. We propose *AnoGAN*, a deep convolutional generative adversarial network to learn a manifold of normal anatomical variability, accompanying a novel anomaly scoring scheme based on the mapping from image space to a latent space. Applied to new data, the model labels anomalies, and scores image patches indicating their fit into the learned distribution. Results on optical coherence tomography images of the retina demonstrate that the approach correctly identifies anomalous images, such as images containing retinal fluid or hyperreflective foci.

GAN based Anomaly Detection

Deep Convolutional Generative Adversarial Networks (DCGAN)

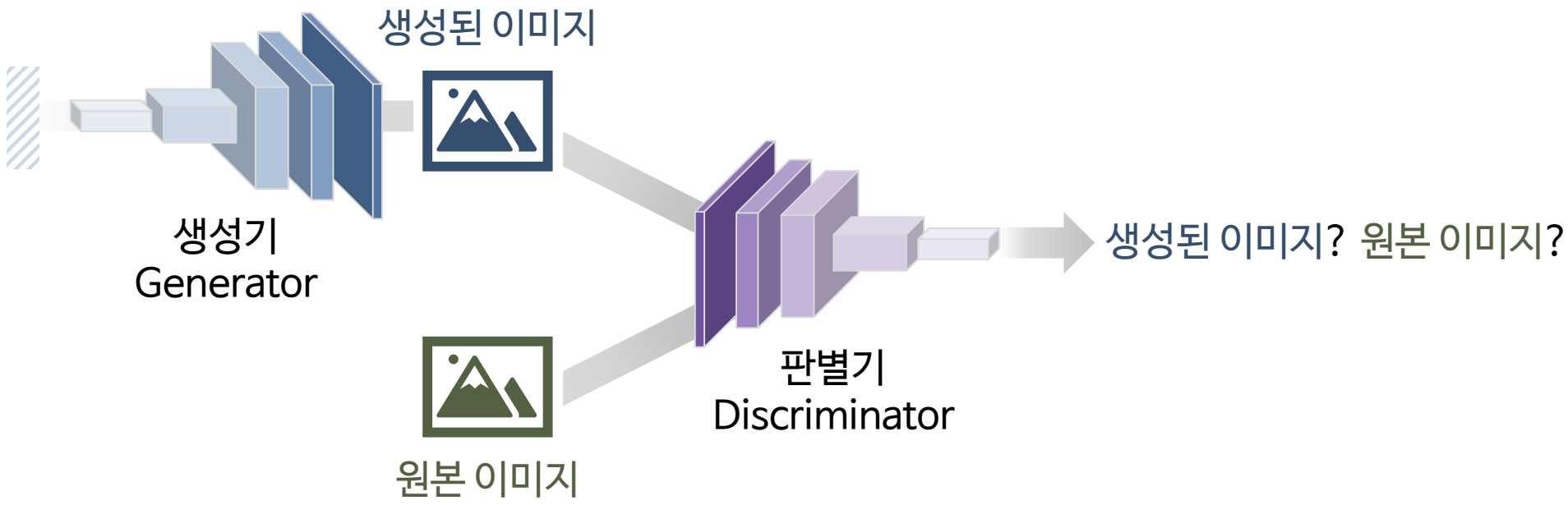
- GAN의 구조에 Convolutional Neural Network (CNN)을 적용한 것
- 생성기가 생성한 이미지와 원본 이미지를 판별하도록 학습됨



GAN based Anomaly Detection

Deep Convolutional Generative Adversarial Networks (DCGAN)

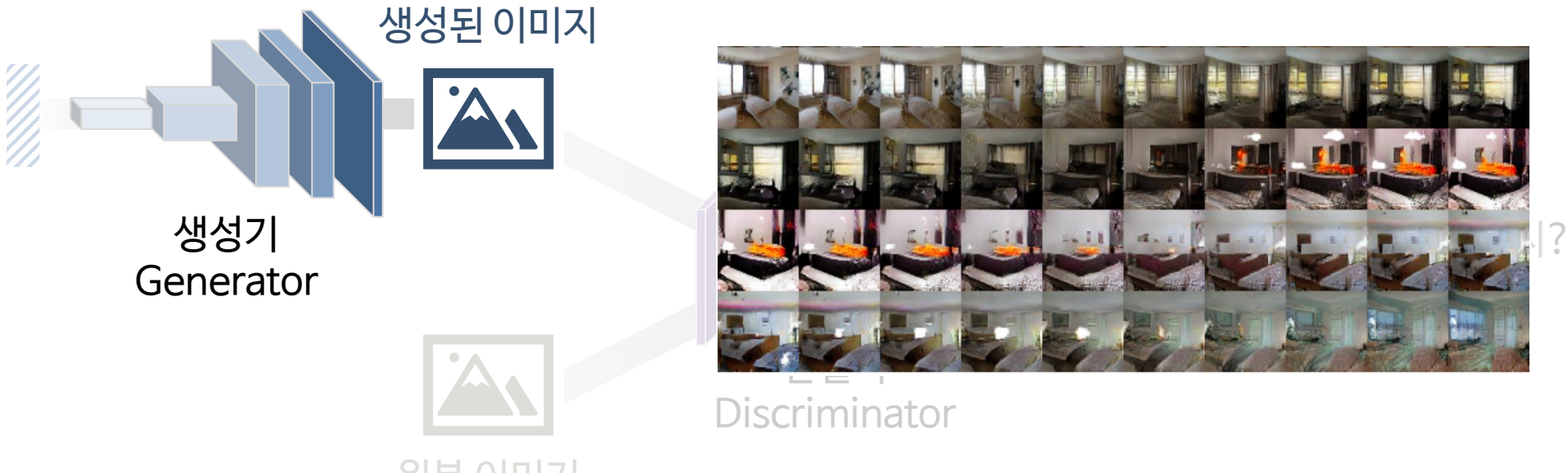
- GAN의 구조에 Convolutional Neural Network (CNN)을 적용한 것
- 생성기가 생성한 이미지와 원본 이미지를 판별하도록 학습됨



GAN based Anomaly Detection

Deep Convolutional Generative Adversarial Networks (DCGAN)

- GAN의 구조에 Convolutional Neural Network (CNN)을 적용한 것
- 생성기가 생성한 이미지와 원본 이미지를 판별하도록 학습됨



→ 인풋 벡터의 값을 조금씩 변화시키면 생성되는 이미지도 조금씩 변화가 생김

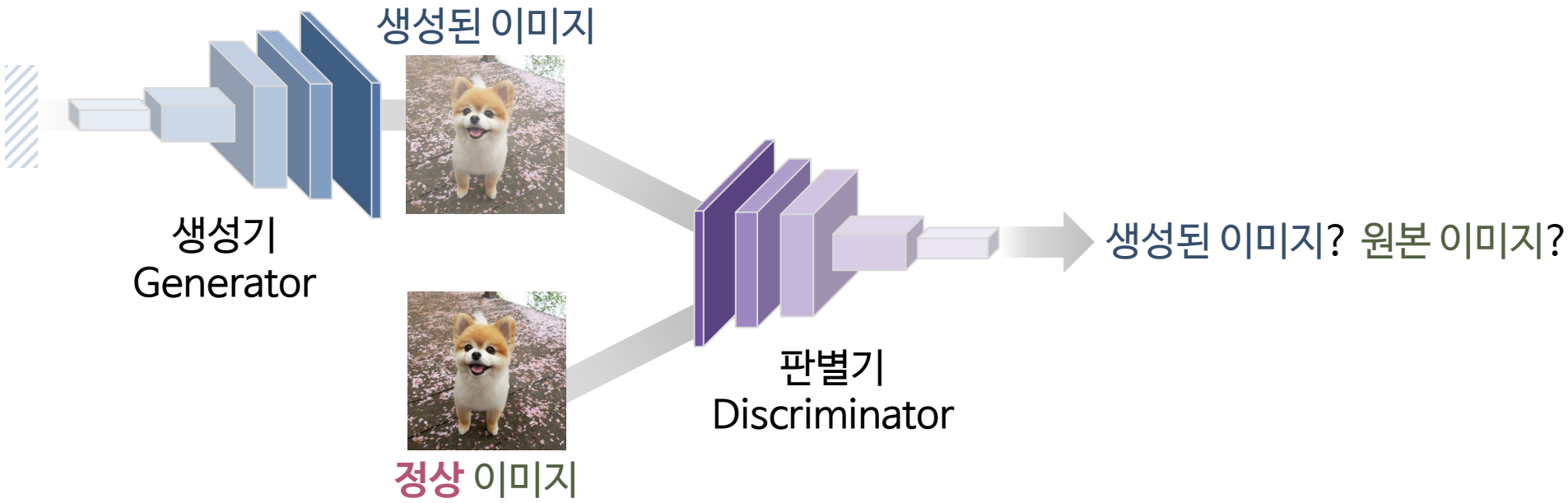
“Walking In the latent space"

GAN based Anomaly Detection

AnoGAN

- 불량 이미지 (이상치)를 탐지하는 AnoGAN 구조 제안

1) 정상 이미지 만으로 학습 진행



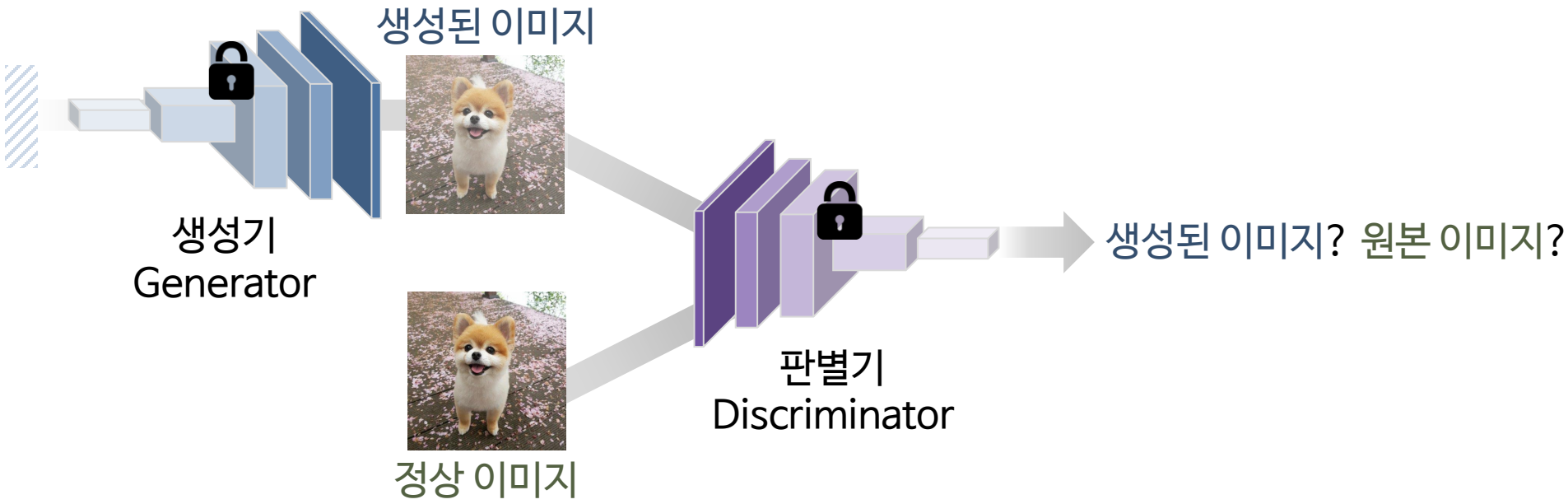
GAN based Anomaly Detection

AnoGAN

- 불량 이미지 (이상치)를 탐지하는 AnoGAN 구조 제안

2) 생성기와 판별기의 학습은 멈춘 상태로, 어떤 인풋을 넣어야 원본 이미지와 유사해질지 임의의 인풋을 학습

→ 정상 이미지로부터 맞는 인풋 벡터를 생성하는 과정, 정상 이미지의 잠재 공간(latent space)을 찾는 것

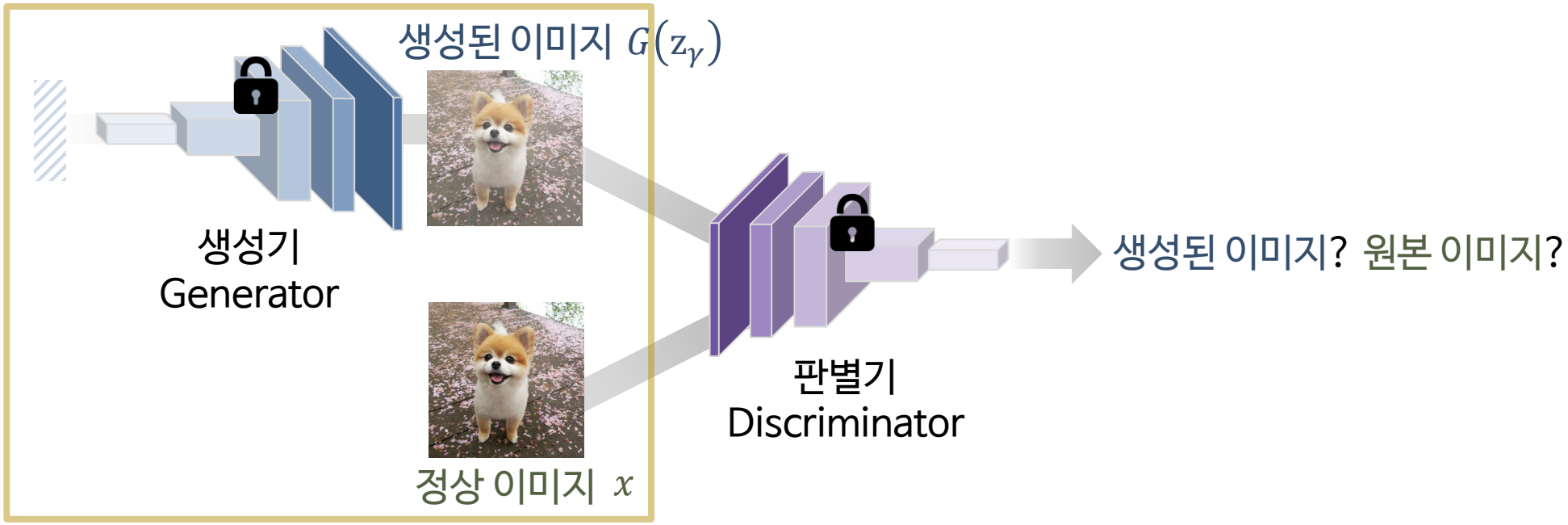


GAN based Anomaly Detection

AnoGAN

- 불량 이미지 (이상치)를 탐지하는 AnoGAN 구조 제안

$$\mathcal{L}(z_\gamma) = (1-\lambda)\cdot\sum|x - G(z_\gamma)| + \lambda\cdot\sum\left|f(x) - f\left(G(z_\gamma)\right)\right|$$



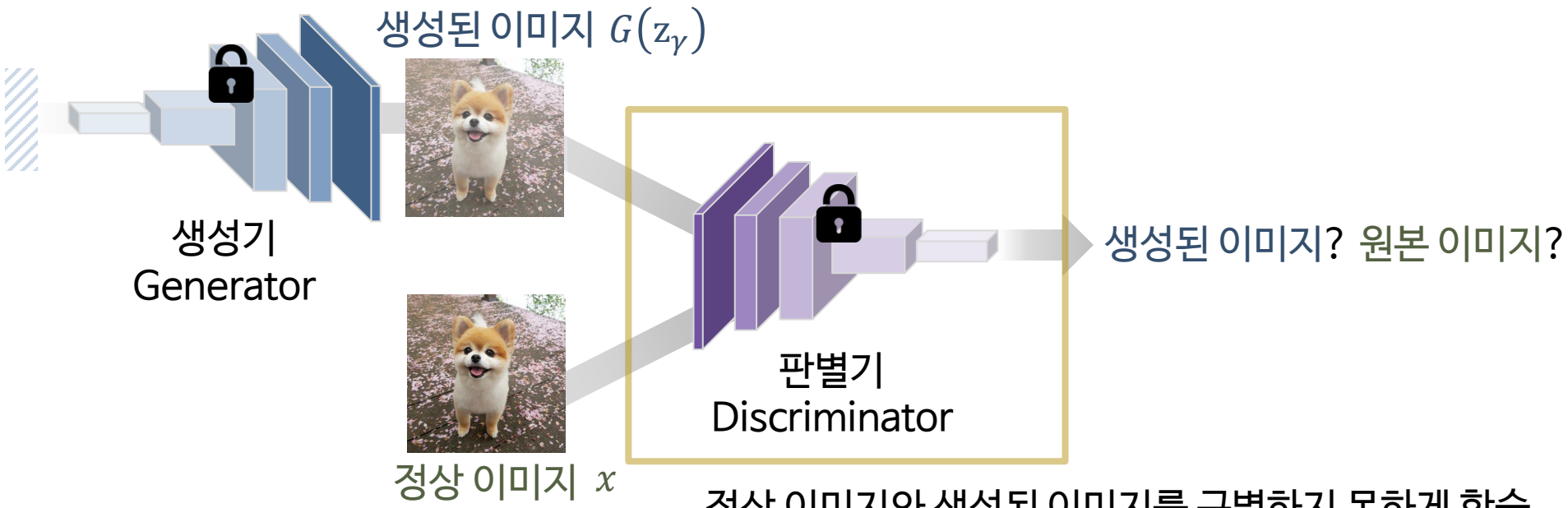
최대한 정상 이미지와 유사해지도록 학습

GAN based Anomaly Detection

AnoGAN

- 불량 이미지 (이상치)를 탐지하는 AnoGAN 구조 제안

$$\mathcal{L}(z_\gamma) = (1-\lambda)\cdot\sum|x - G(z_\gamma)| + \lambda\cdot\sum\left|f(x) - f\left(G(z_\gamma)\right)\right|$$



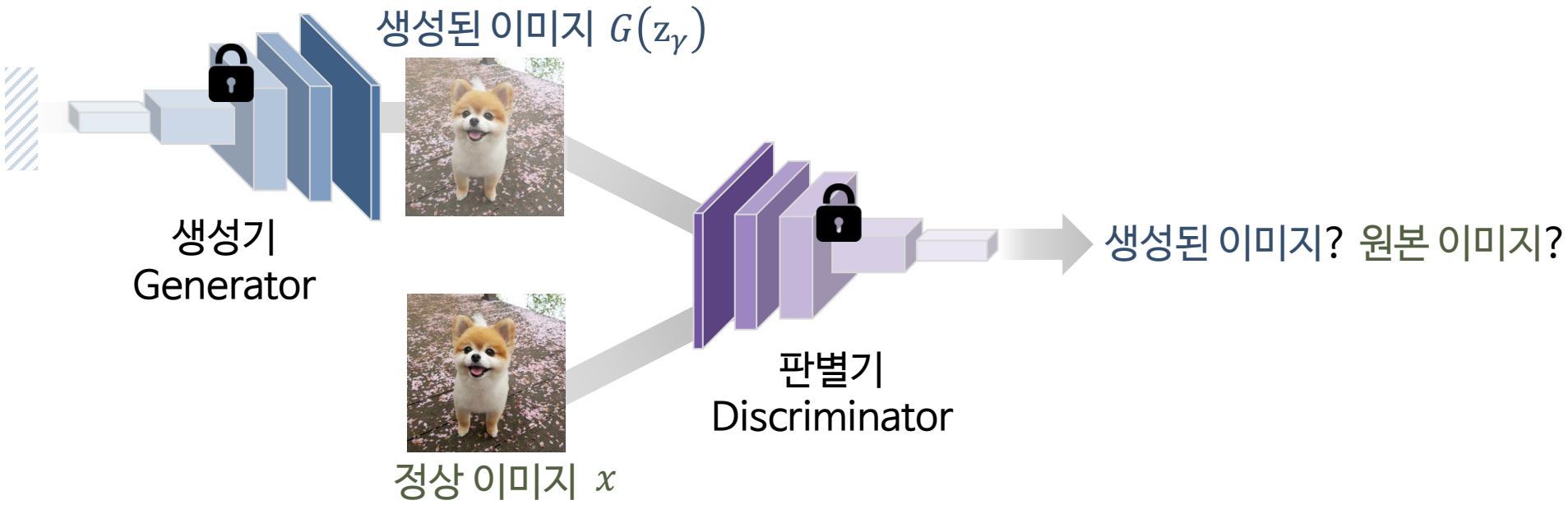
GAN based Anomaly Detection

AnoGAN

- 불량 이미지 (이상치)를 탐지하는 AnoGAN 구조 제안

3) 새로 들어온 이미지로 이상치 점수 (Anomaly Score) 계산

$$Anomaly\ Score = (1-\lambda)\cdot\sum|x - G(z)| + \lambda\cdot\sum|f(x) - f(G(z))|$$



→ 정상 이미지가 들어오면, 생성되는 이미지와 큰 차이가 없기 때문에 ($x \approx G(z)$) 작은 이상치 점수 확인

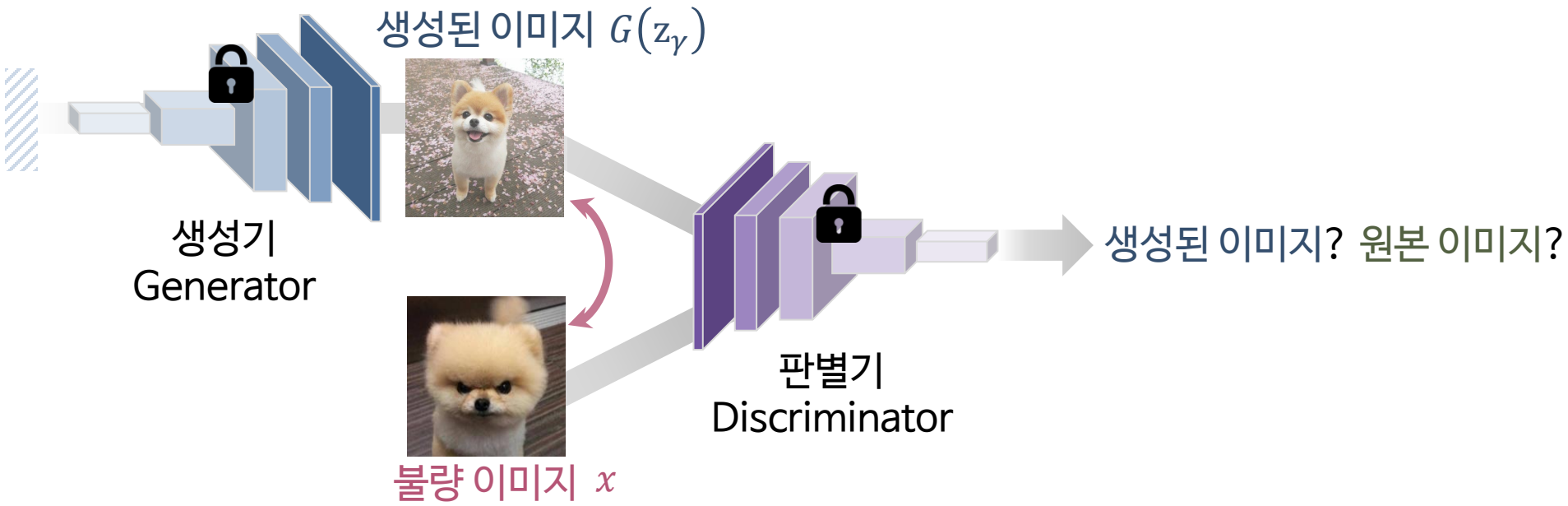
GAN based Anomaly Detection

AnoGAN

- 불량 이미지 (이상치)를 탐지하는 AnoGAN 구조 제안

3) 새로 들어온 이미지로 이상치 점수 (Anomaly Score) 계산

$$Anomaly\ Score = (1-\lambda)\cdot\sum|x-G(z)| + \lambda\cdot\sum|f(x)-f(G(z))|$$



→ 불량 이미지가 들어오면, 정상을 기반으로 생성되는 이미지와 큰 차이를 보이기 때문에 $(x \neq G(z))$ 큰 이상치 점수 확인

GAN based Anomaly Detection

- GANomaly: Semi-Supervised Anomaly Detection via Adversarial Training
- Asian Conference on Computer Vision에서 2018년 발표된 논문
- 2021년 10월 13일 기준 439회 인용

GANomaly: Semi-Supervised Anomaly Detection via Adversarial Training

Samet Akcay[1], Amir Atapour-Abarghouei[1], and Toby P. Breckon[1,2]

Department of {Computer Science[1], Engineering[2]}, Durham University, UK
{ samet.akcay, amir.atapour-abarghouei, toby.breckon }@durham.ac.uk

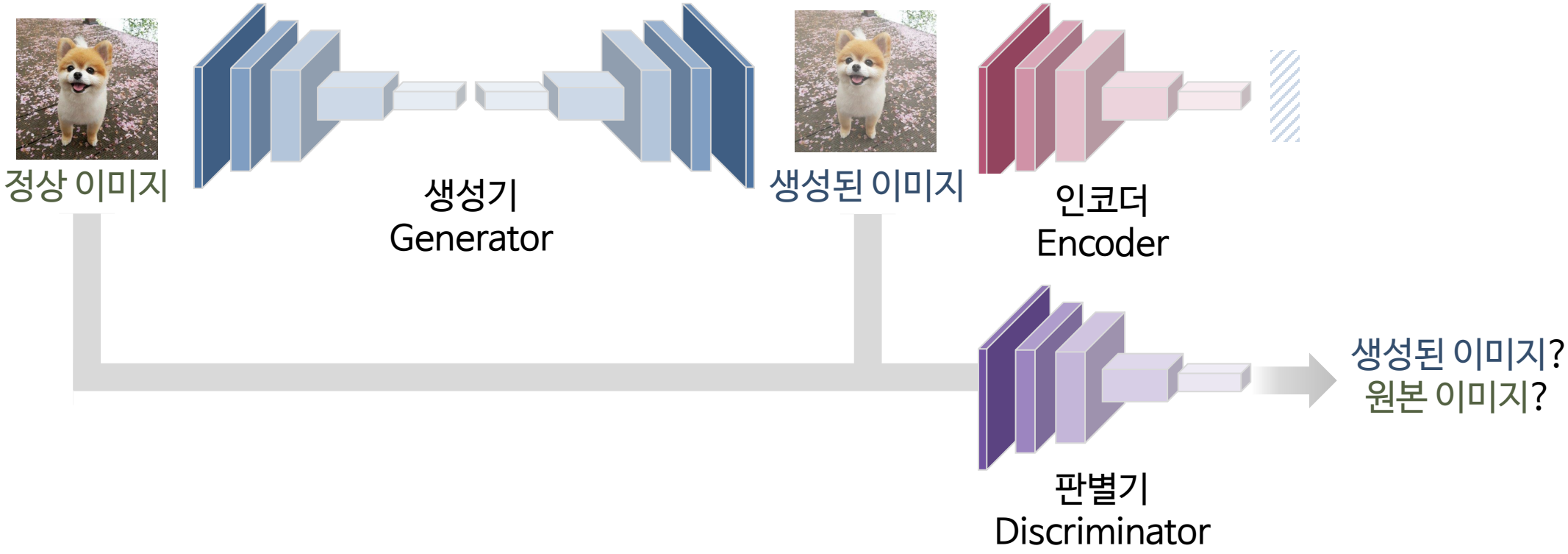
Abstract. Anomaly detection is a classical problem in computer vision, namely the determination of the *normal* from the *abnormal* when datasets are highly biased towards one class (normal) due to the insufficient sample size of the other class (abnormal). While this can be addressed as a supervised learning problem, a significantly more challenging problem is that of detecting the unknown/unseen anomaly case that takes us instead into the space of a one-class, semi-supervised learning paradigm. We introduce such a novel anomaly detection model, by using a conditional generative adversarial network that jointly learns the generation of high-dimensional image space and the inference of latent space. Employing encoder-decoder-encoder sub-networks in the generator network enables the model to map the input image to a lower dimension vector, which is then used to reconstruct the generated output image. The use of the additional encoder network maps this generated image to its latent representation. Minimizing the distance between these images and the latent vectors during training aids in learning the data distribution for the normal samples. As a result, a larger distance metric from this learned data distribution at inference time is indicative of an outlier from that distribution — *an anomaly*. Experimentation over several benchmark datasets, from varying domains, shows the model efficacy and superiority over previous state-of-the-art approaches.

Keywords: Anomaly Detection · Semi-Supervised Learning · Generative Adversarial Networks · X-ray Security Imagery.

GAN based Anomaly Detection

GANomaly

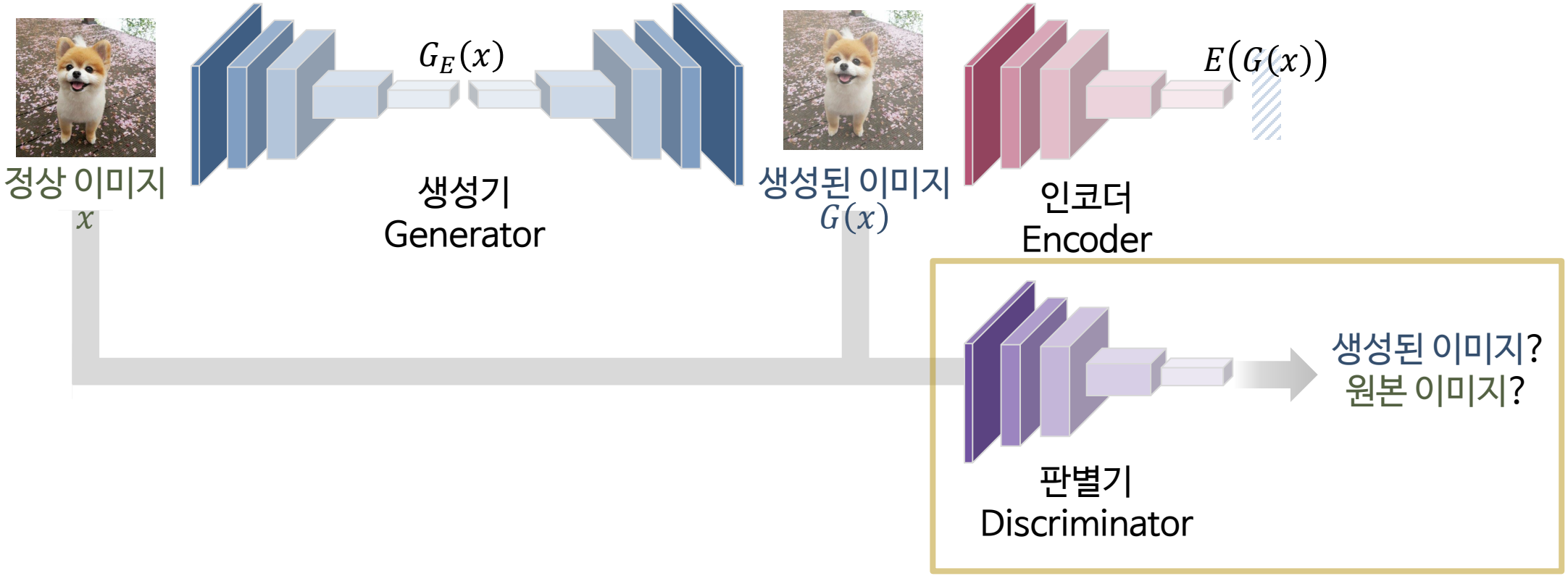
- 이미지에 대한 학습과 잠재 공간에 대한 학습을 한 번에 진행하기 위하여 제안된 모델



GAN based Anomaly Detection

GANomaly

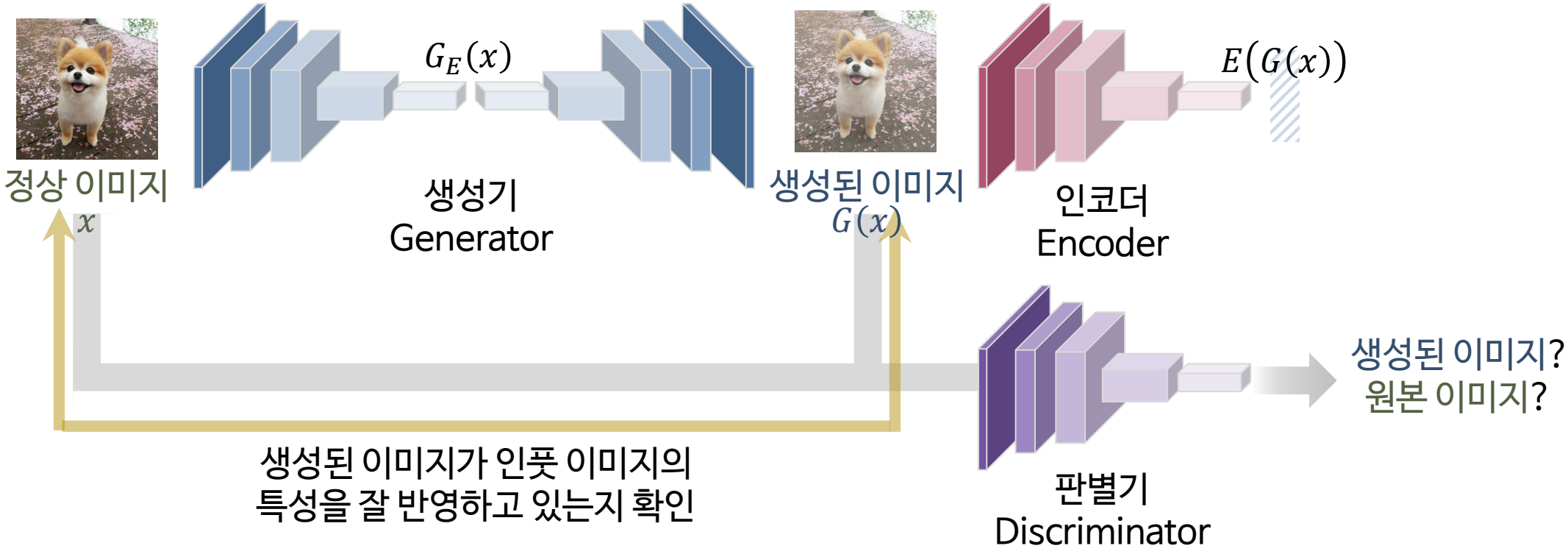
$$\mathcal{L} = w_{adv} E_{x \sim p_x} \left\| f(x) - E_{x \sim p_x} f\left(G(x)\right) \right\|_2 + w_{con} E_{x \sim p_x} \|x - G(x)\|_1 + w_{enc} E_{x \sim p_x} \left\| G_E(x) - E\left(G(x)\right) \right\|_2$$



GAN based Anomaly Detection

GANomaly

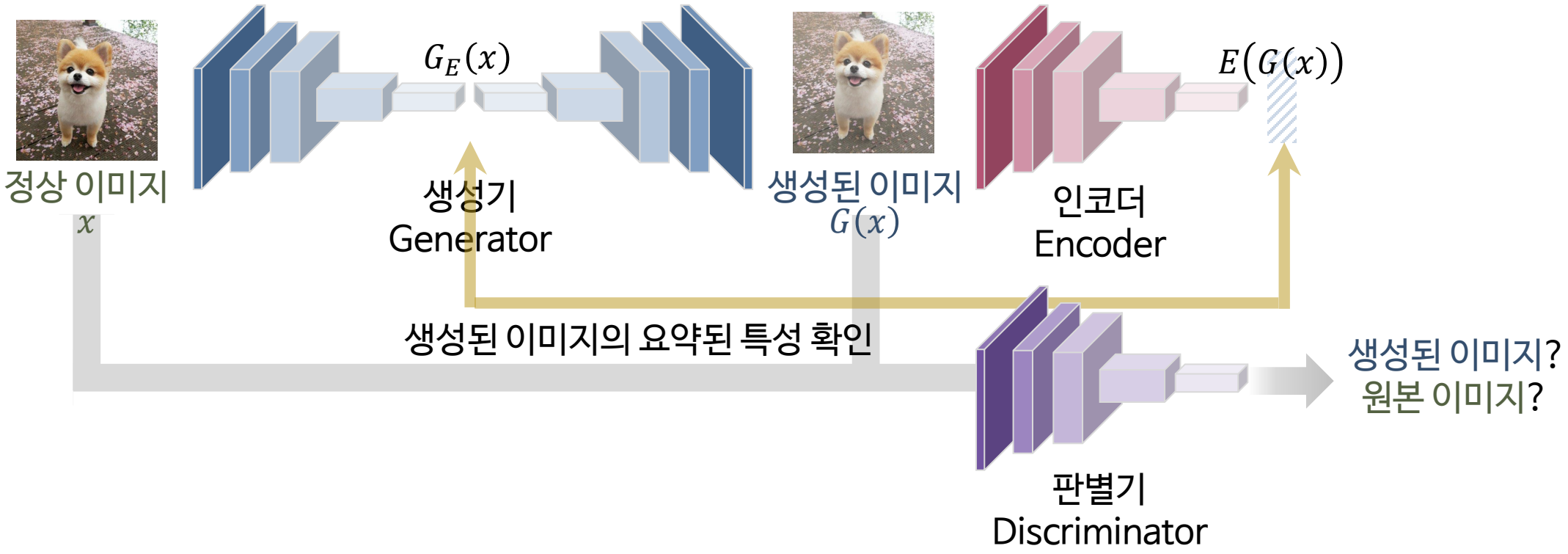
$$\mathcal{L} = w_{adv} E_{x \sim p_x} \left\| f(x) - E_{x \sim p_x} f\left(G(x)\right) \right\|_2 + w_{con} E_{x \sim p_x} \|x - G(x)\|_1 + w_{enc} E_{x \sim p_x} \left\| G_E(x) - E\left(G(x)\right) \right\|_2$$



GAN based Anomaly Detection

GANomaly

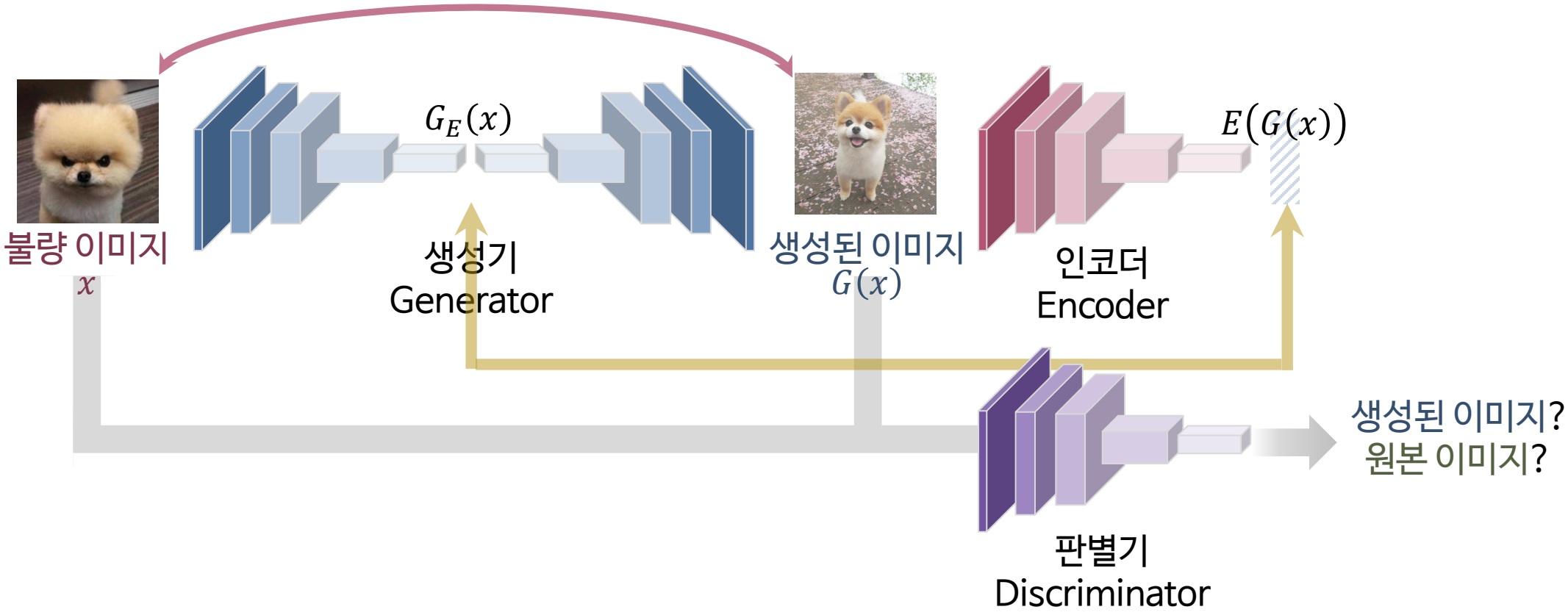
$$\mathcal{L} = w_{adv}E_{x\sim p_x}\left\|f(x) - E_{x\sim p_x}f\left(G(x)\right)\right\|_2 + w_{con}E_{x\sim p_x}\|x - G(x)\|_1 + w_{enc}E_{x\sim p_x}\left\|G_E(x) - E\left(G(x)\right)\right\|_2$$



GAN based Anomaly Detection

GANomaly

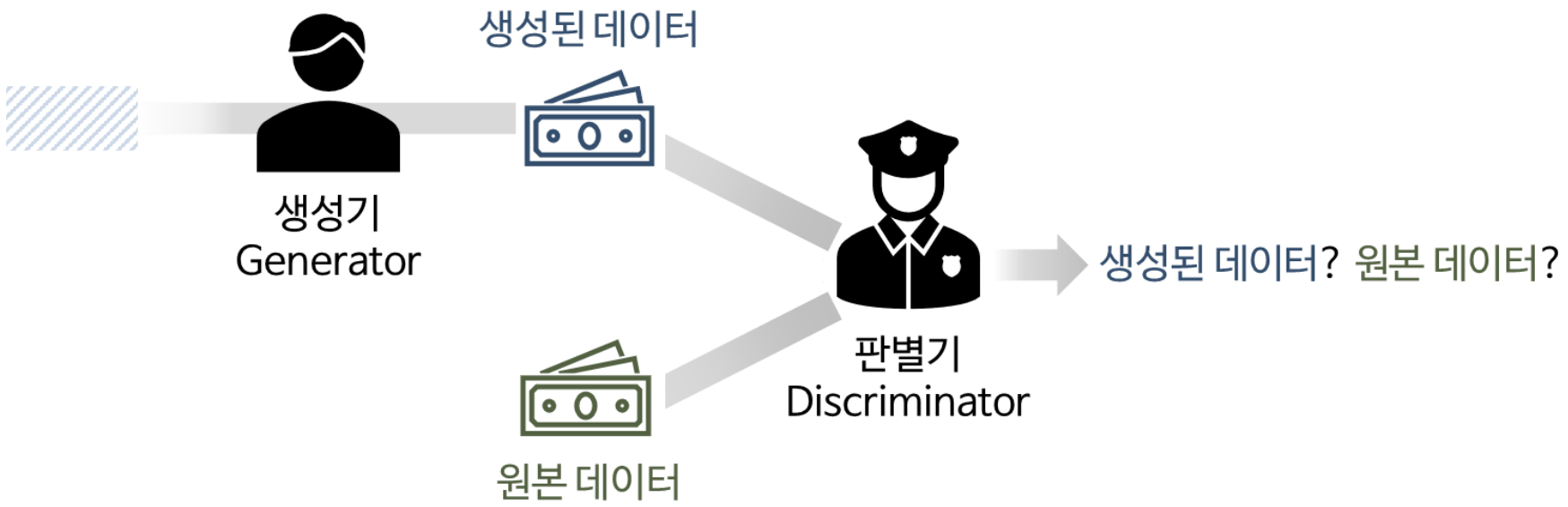
$$Anomaly\ Score = \left\| G_E(x) - E(G(x)) \right\|_1$$

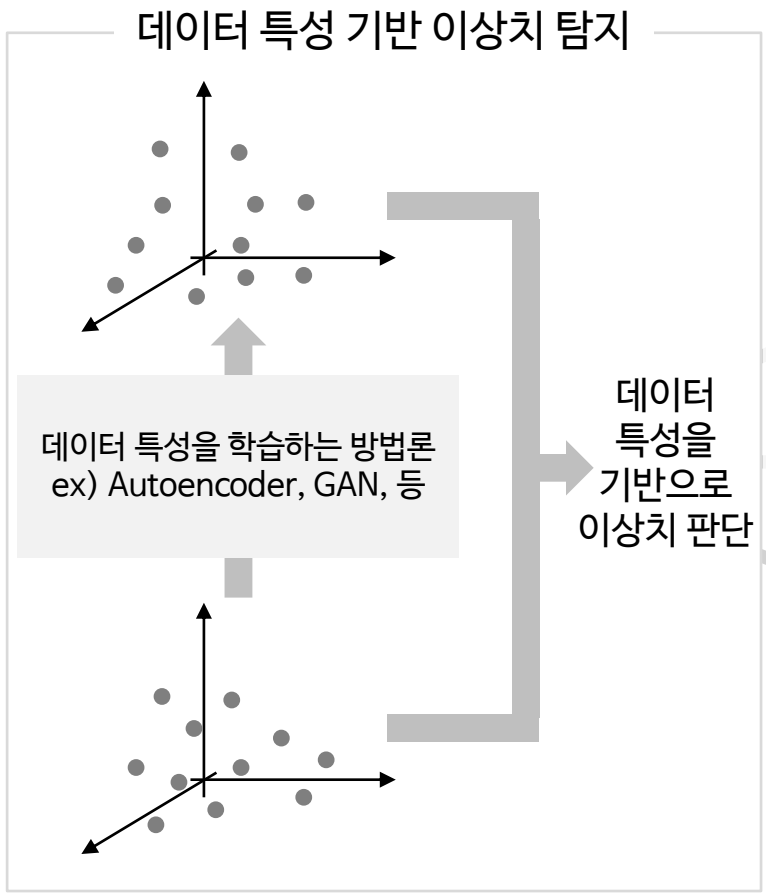


→ 불량 이미지가 들어오면, 정상을 기반으로 생성되는 이미지와 큰 차이를 보이기 때문에 $(x \neq G(z))$ 큰 이상치 점수 확인

GAN based Anomaly Detection

- 임의의 인풋으로부터 원본 데이터를 잘 생성해내는 GAN 모델
- 다양한 GAN 기반 모델들을 활용하여 이상치 탐지 가능
  - AnoGAN, GANomaly





Autoencoder based Anomaly Detection

GAN based Anomaly Detection

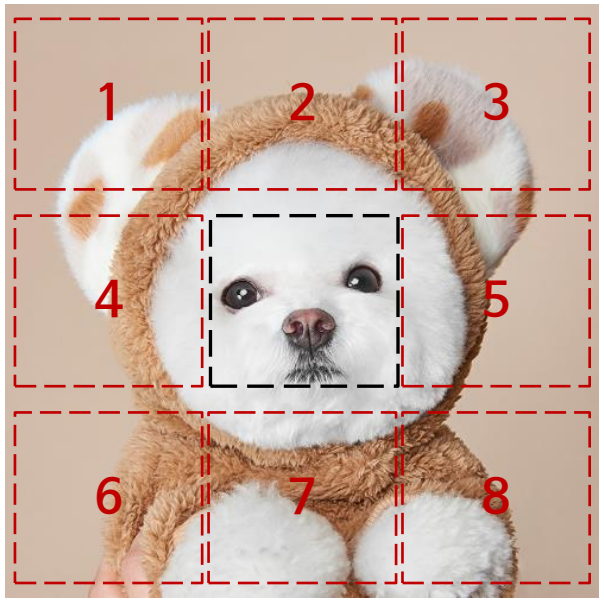
Self-Supervised Learning based Anomaly Detection

Self-Supervised Learning based Anomaly Detection

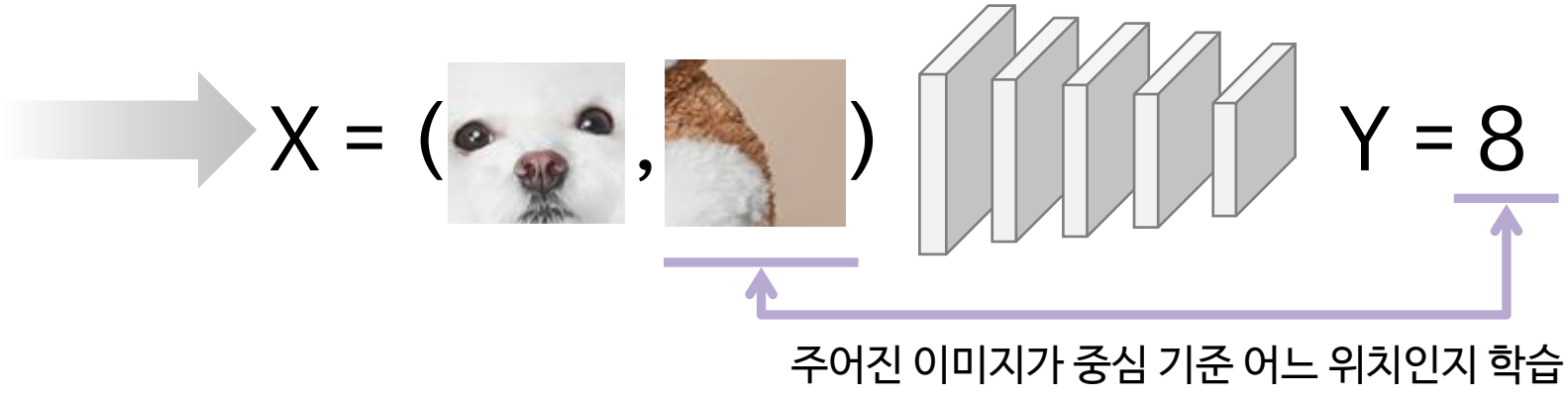
Self-Supervised Learning이란?

- 딥러닝 계열 모델들을 잘 학습시키기 위해서는 많은 양의 레이블링 된 데이터가 필요
- **레이블링이 없는 상황**에서도 데이터의 특성을 잘 배우기 위한 연구

1) 데이터의 특성을 파악할 수 있는 모델의 사전 학습 (pre-training) 단계



레이블이 없는 데이터



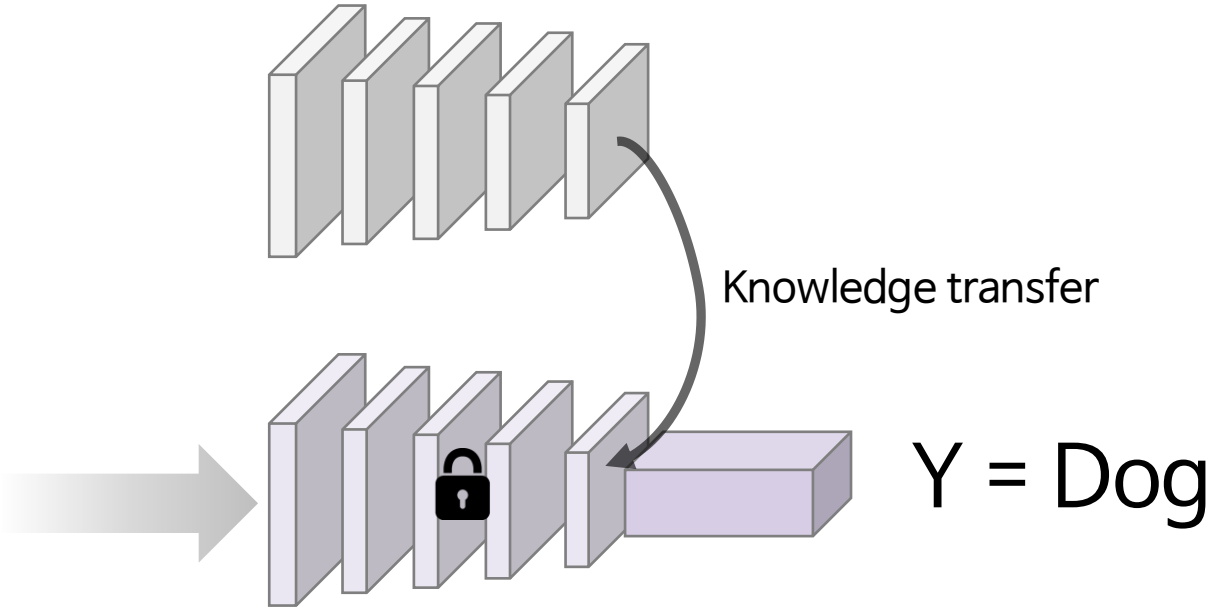
Self-Supervised Learning based Anomaly Detection

Self-Supervised Learning이란?

- 딥러닝 계열 모델들을 잘 학습시키기 위해서는 많은 양의 레이블링 된 데이터가 필요
- **레이블링이 없는 상황**에서도 데이터의 특성을 잘 배우기 위한 연구

2) 학습된 모델을 fine tuning하여 원하는 문제를 해결하는 단계

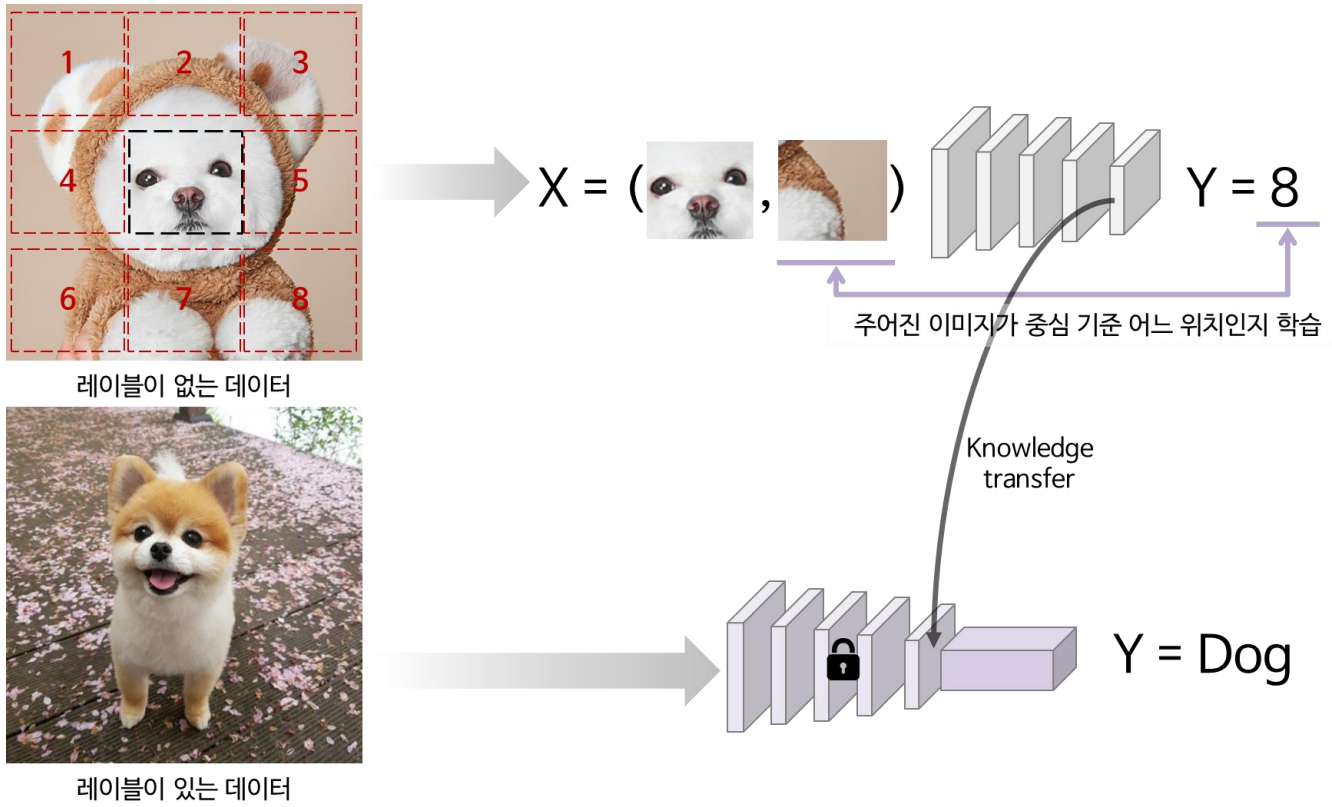
레이블이 있는 데이터



Self-Supervised Learning based Anomaly Detection

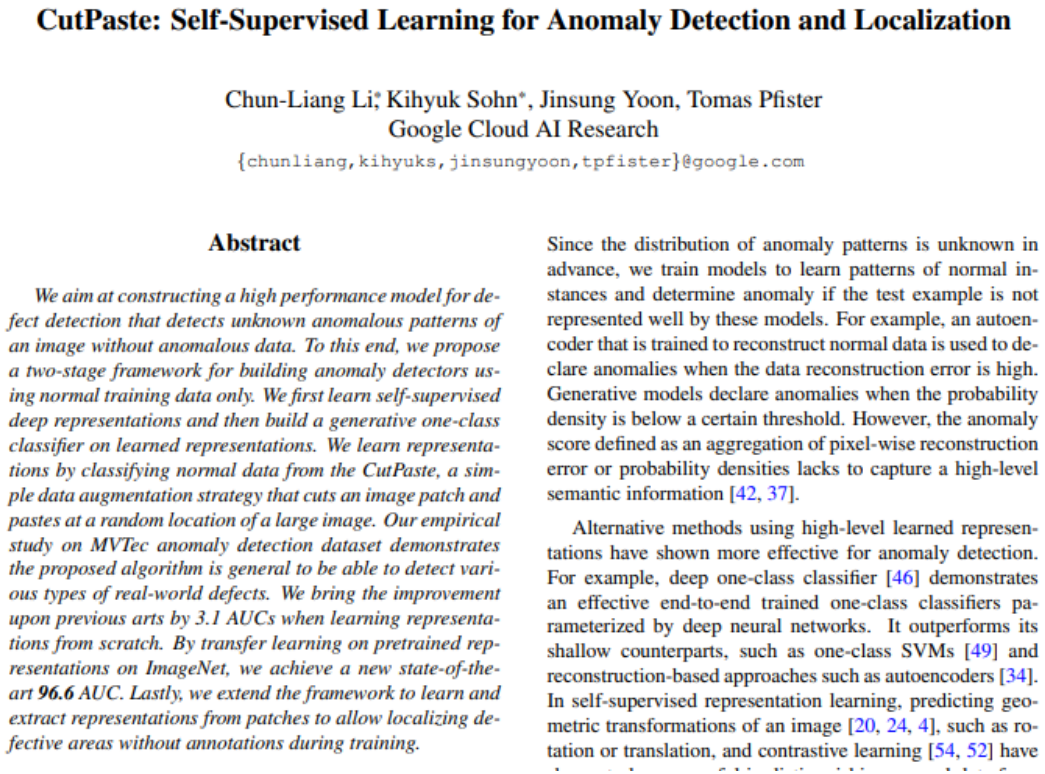
Self-Supervised Learning for Anomaly Detection?

- 사용자가 정의한 임의의 task를 학습시키는 과정을 통하여 데이터의 특성 파악 가능
- 가정 : 정상 데이터와 불량 데이터 사이에 파악된 특성이 서로 다를 것이다.



Self-Supervised Learning based Anomaly Detection

- CutPaste: Self-Supervised Learning for Anomaly Detection and Localization
- Computer Vision and Pattern Recognition에서 2021년 발표된 논문
- 2021년 10월 13일 기준 9회 인용



CutPaste: Self-Supervised Learning for Anomaly Detection and Localization

Chun-Liang Li*, Kihyuk Sohn*, Jinsung Yoon, Tomas Pfister
Google Cloud AI Research
{chunliang,kihyuks,jinsungyoon,tpfister}@google.com

Abstract

*We aim at constructing a high performance model for defect detection that detects unknown anomalous patterns of an image without anomalous data. To this end, we propose a two-stage framework for building anomaly detectors using normal training data only. We first learn self-supervised deep representations and then build a generative one-class classifier on learned representations. We learn representations by classifying normal data from the CutPaste, a simple data augmentation strategy that cuts an image patch and pastes at a random location of a large image. Our empirical study on MVTec anomaly detection dataset demonstrates the proposed algorithm is general to be able to detect various types of real-world defects. We bring the improvement upon previous arts by 3.1 AUCs when learning representations from scratch. By transfer learning on pretrained representations on ImageNet, we achieve a new state-of-the-art **96.6** AUC. Lastly, we extend the framework to learn and extract representations from patches to allow localizing defective areas without annotations during training.*

Since the distribution of anomaly patterns is unknown in advance, we train models to learn patterns of normal instances and determine anomaly if the test example is not represented well by these models. For example, an autoencoder that is trained to reconstruct normal data is used to declare anomalies when the data reconstruction error is high. Generative models declare anomalies when the probability density is below a certain threshold. However, the anomaly score defined as an aggregation of pixel-wise reconstruction error or probability densities lacks to capture a high-level semantic information [42, 37].

Alternative methods using high-level learned representations have shown more effective for anomaly detection. For example, deep one-class classifier [46] demonstrates an effective end-to-end trained one-class classifiers parameterized by deep neural networks. It outperforms its shallow counterparts, such as one-class SVMs [49] and reconstruction-based approaches such as autoencoders [34]. In self-supervised representation learning, predicting geometric transformations of an image [20, 24, 4], such as rotation or translation, and contrastive learning [54, 52] have

Self-Supervised Learning based Anomaly Detection

CutPaste

- 불량 이미지의 경우 이미지의 작은 부분에서 다름을 보이는 아이디어 활용
- 이미지의 일부분을 떼어 다른 위치에 붙여 기존 이미지와 큰 차이가 없어 보이도록 하는 데이터 증강 기법 CutPaste 제안
- CutPaste를 통하여 얻는 새로운 이미지를 가상 불량 이미지로 설정

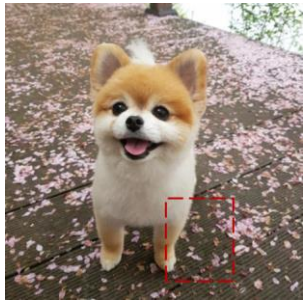
정상 이미지 가상 불량 이미지

Self-Supervised Learning based Anomaly Detection

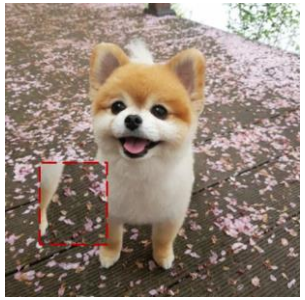
CutPaste

- 불량 이미지의 경우 이미지의 작은 부분에서 다름을 보이는 아이디어 활용
- 이미지의 일부분을 떼어 다른 위치에 붙여 기존 이미지와 큰 차이가 없어 보이도록 하는 데이터 증강 기법 CutPaste 제안

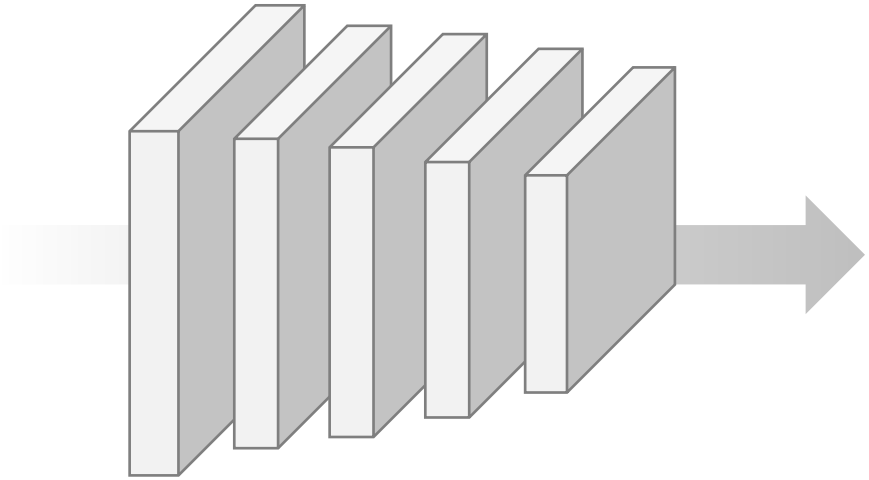
1) 정상과 가상 불량 이미지를 분류하는 분류기 모델 학습



정상 이미지



가상 불량 이미지



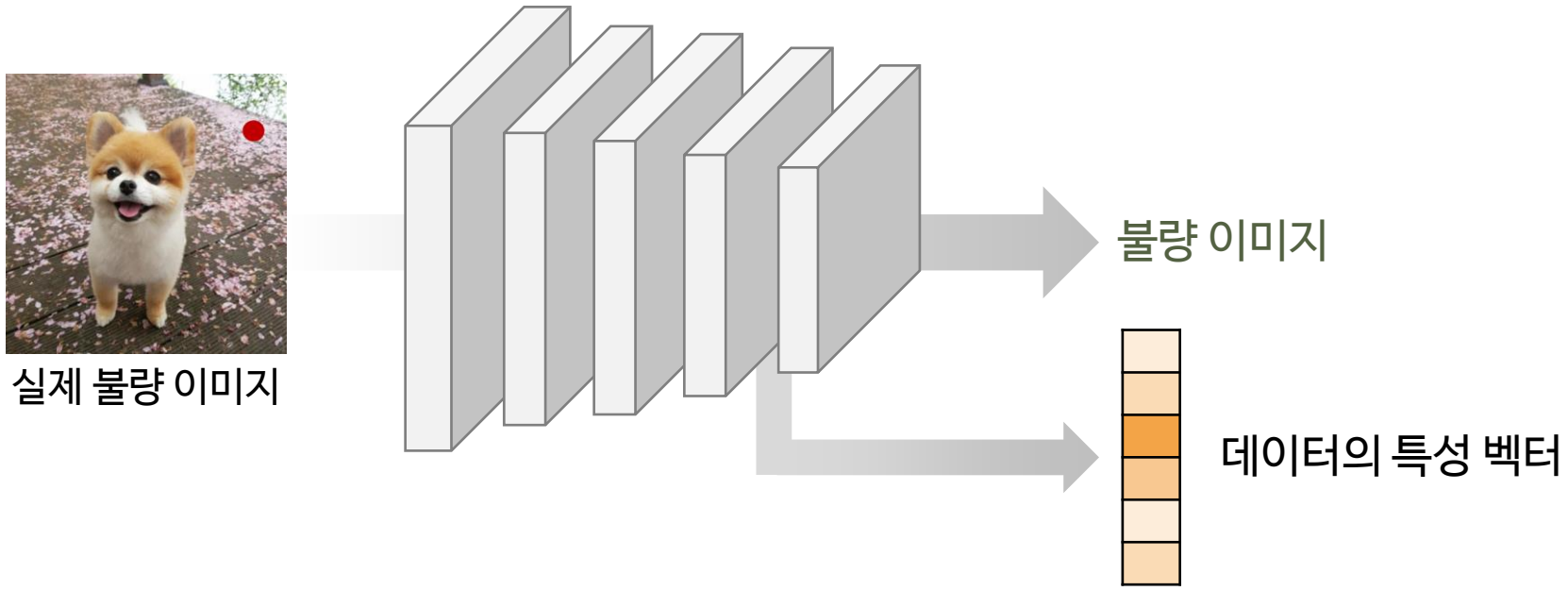
정상 이미지? 불량 이미지?

Self-Supervised Learning based Anomaly Detection

CutPaste

- 불량 이미지의 경우 이미지의 작은 부분에서 다름을 보이는 아이디어 활용
- 이미지의 일부분을 떼어 다른 위치에 붙여 기존 이미지와 큰 차이가 없어 보이도록 하는 데이터 증강 기법 CutPaste 제안

2) 새로 들어온 데이터를 학습된 모델에 적용하면 데이터의 특성 벡터 확인 가능

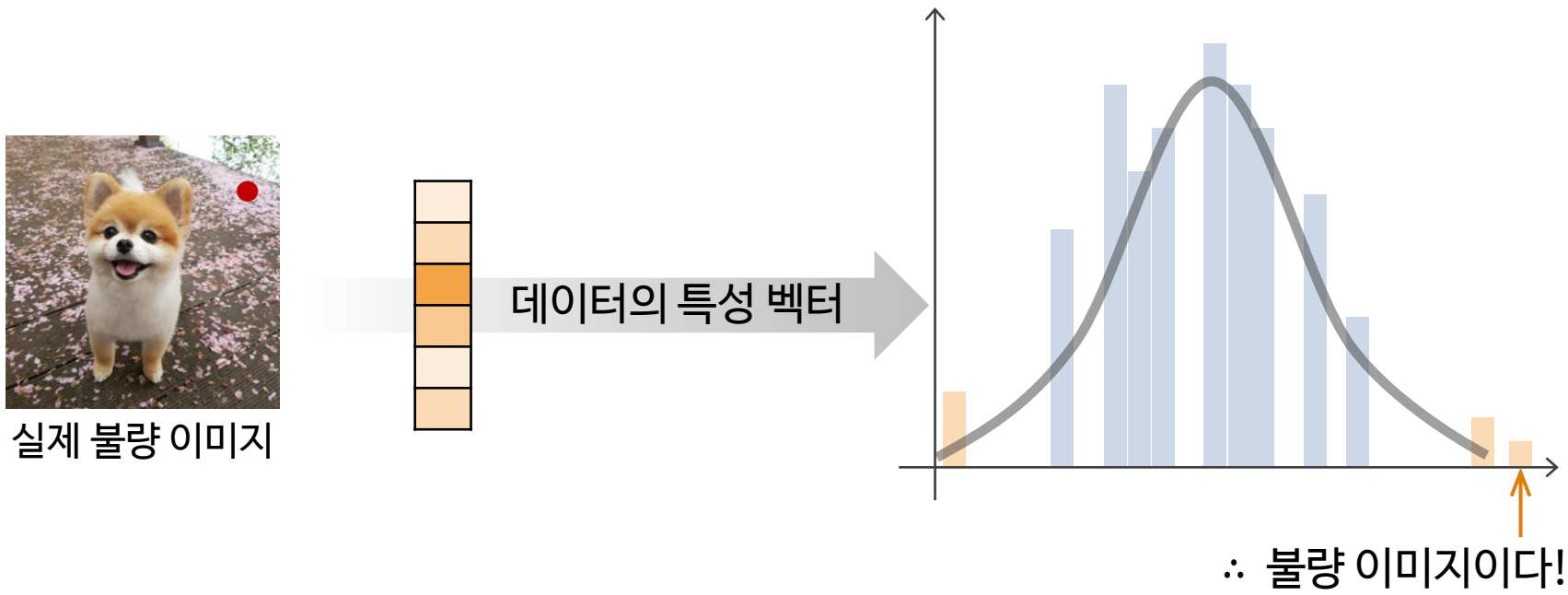


Self-Supervised Learning based Anomaly Detection

CutPaste

- 불량 이미지의 경우 이미지의 작은 부분에서 다름을 보이는 아이디어 활용
- 이미지의 일부분을 떼어 다른 위치에 붙여 기존 이미지와 큰 차이가 없어 보이도록 하는 데이터 증강 기법 CutPaste 제안

3) 데이터의 특성 벡터에 가우시안 밀도 추정법 적용하여 이상치 점수 계산 및 불량 탐지

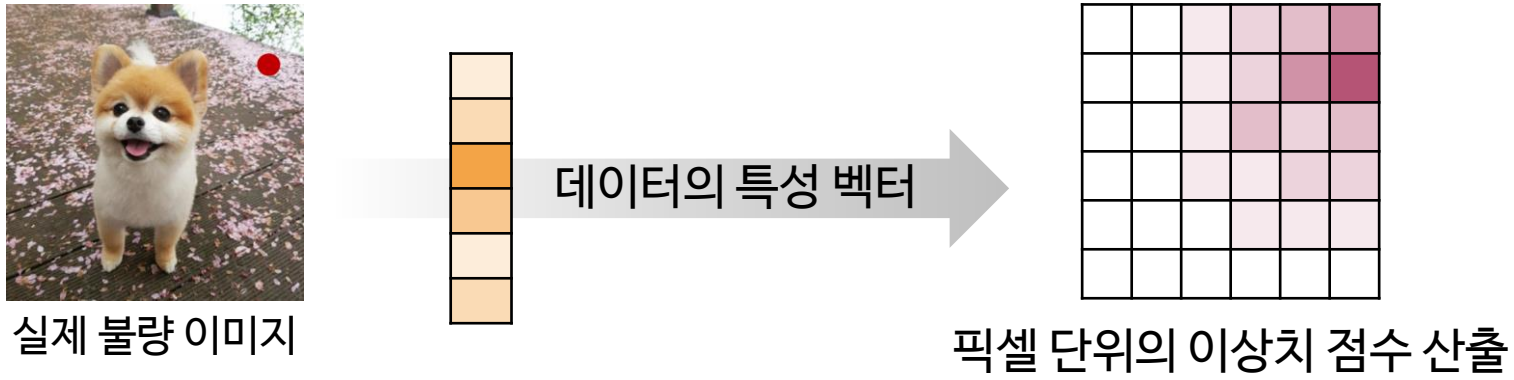


Self-Supervised Learning based Anomaly Detection

CutPaste

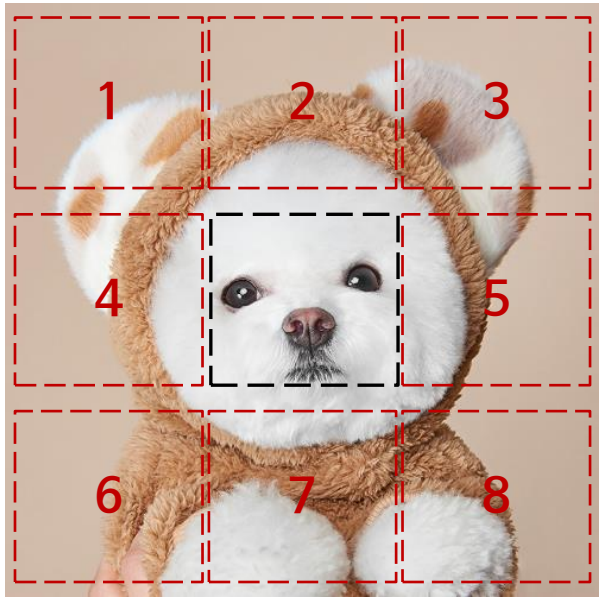
- 불량 이미지의 경우 이미지의 작은 부분에서 다름을 보이는 아이디어 활용
- 이미지의 일부분을 떼어 다른 위치에 붙여 기존 이미지와 큰 차이가 없어 보이도록 하는 데이터 증강 기법 CutPaste 제안

3) 데이터의 특성 벡터에 가우시안 밀도 추정법 적용하여 이상치 점수 계산 및 불량 탐지

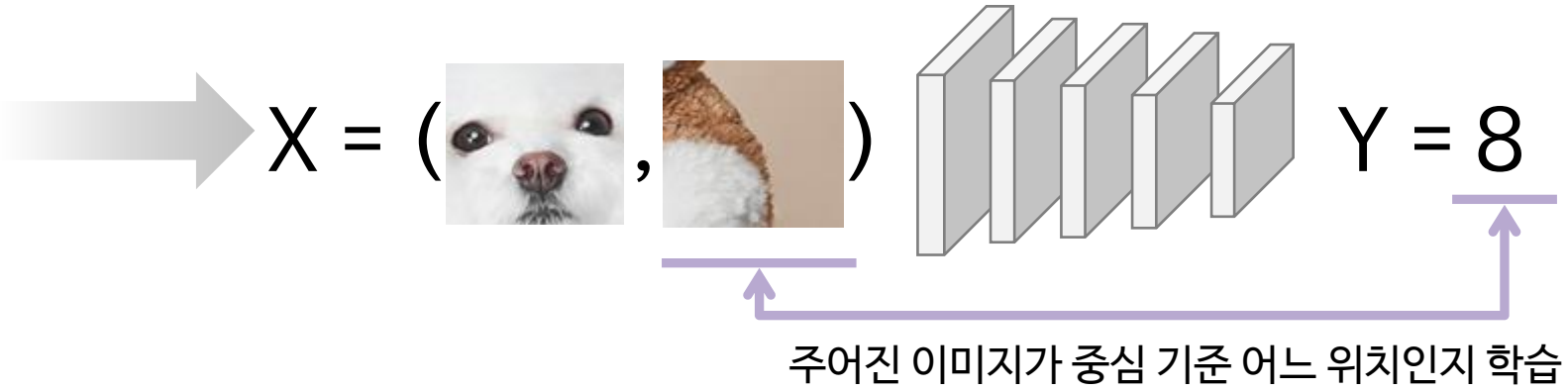


Self-Supervised Learning based Anomaly Detection

- 데이터의 레이블이 없는 상황에서도 데이터의 특성을 잘 학습하기 위한 방법

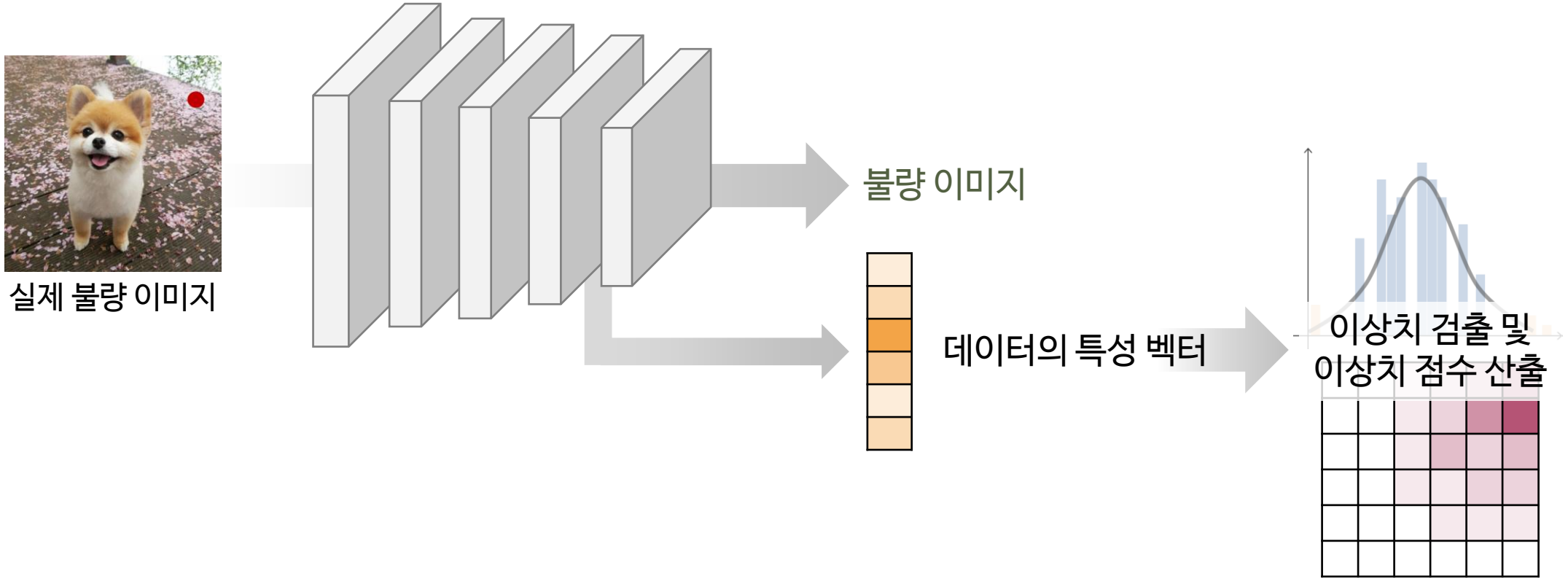


레이블이 없는 데이터



Self-Supervised Learning based Anomaly Detection

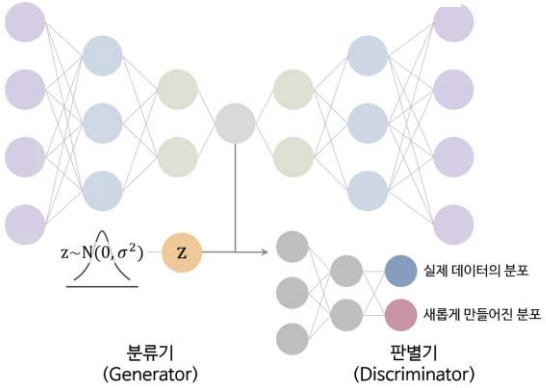
- 학습된 데이터의 특성을 활용하여 이상치 탐지 진행



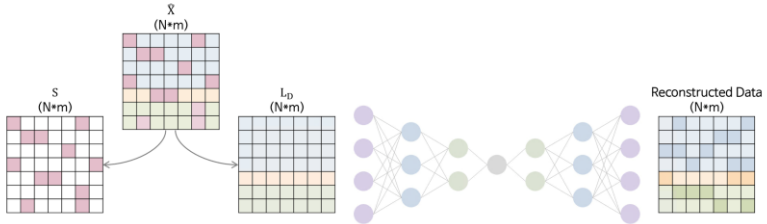
Concusion

- **이상치 탐지 알고리즘** : 정상 데이터만의 분포와 특징을 파악한 이후에,
  새로 들어온 데이터 중 서로 다른 불량 데이터를 찾는 과정

Autoencoder based

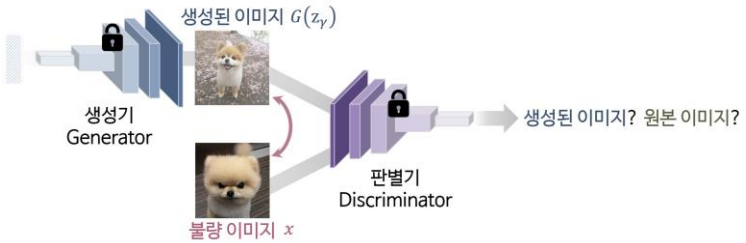


Adversarial Autoencoder

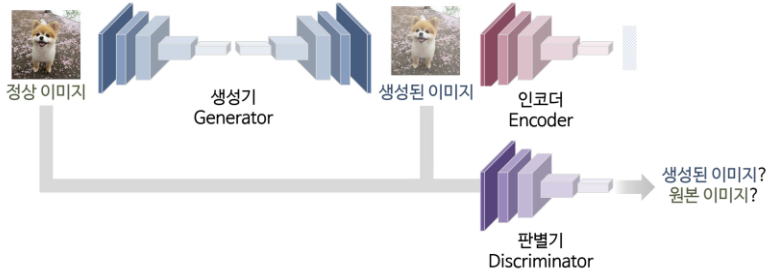


Robust Deep Autoencoders (RDA)

GAN based

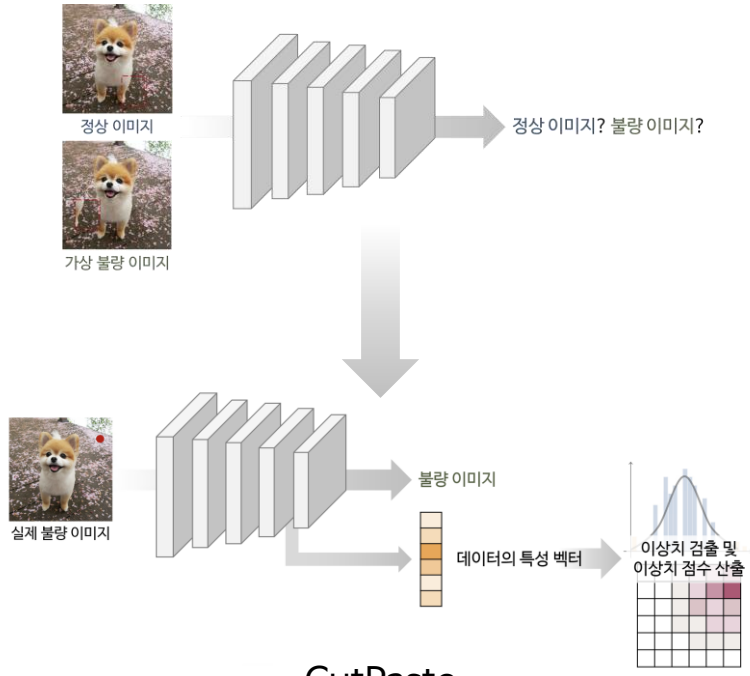


AnoGAN



GANomaly

Self-Supervised Learning based



CutPaste

감사합니다.

참고자료

- Chalapathy, R., & Chawla, S. (2019). Deep learning for anomaly detection: A survey. arXiv preprint arXiv:1901.03407.
- Pang, G., Shen, C., Cao, L., & Hengel, A. V. D. (2021). Deep learning for anomaly detection: A review. ACM Computing Surveys (CSUR), 54(2), 1-38.
- Zhou, C., & Paffenroth, R. C. (2017, August). Anomaly detection with robust deep autoencoders. In Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 665-674).
- Schlegl, T., Seeböck, P., Waldstein, S. M., Schmidt-Erfurth, U., & Langs, G. (2017, June). Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In International conference on information processing in medical imaging (pp. 146-157). Springer, Cham.
- Akcay, S., Atapour-Abarghouei, A., & Breckon, T. P. (2018, December). Ganomaly: Semi-supervised anomaly detection via adversarial training. In Asian conference on computer vision (pp. 622-637). Springer, Cham.
- Li, C. L., Sohn, K., Yoon, J., & Pfister, T. (2021). CutPaste: Self-Supervised Learning for Anomaly Detection and Localization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 9664-9674).