

2025 동계 세미나

Reconstruction-based Anomaly Detection



Sogang University

Vision & Display Systems Lab, Dept. of Electronic Engineering



Presented By

김예슬

Contents

- Background
 - Anomaly detection (AD)
 - Supervised and Unsupervised approaches
 - Reconstruction-based Anomaly Detection
 - CutPaste paper review
- Zavrtnik, et al. "Draem-a discriminatively trained reconstruction embedding for surface anomaly detection." *Proceedings of the IEEE/CVF international conference on computer vision*. 2021.
 - Introduction
 - Method
 - Experiments
 - Conclusions

Background

- Anomaly Detection (AD)

- Binary classification problem

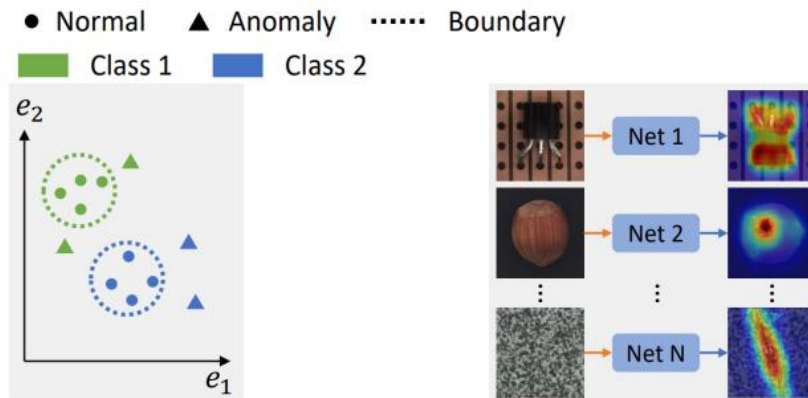
- Input 이미지의 anomaly 포함 여부를 판단하는 문제

- ※ 일반적으로 Abnormal 샘플은 normal 샘플 수 대비 소수이기 때문에 적절한 distribution을 학습하기 어려움

- ※ 따라서, normal 샘플만을 활용하여 해당 클래스의 특징적 distribution를 학습하는 one-class classification 방식이 주로 사용됨

- 여러 개의 object나 class를 다루는 상황에서는 각 클래스별로 normal 데이터를 별도로 학습하여 개별적인 decision boundary를 형성하는 one-class-one-model 접근법이 주를 이룸

- ※ 즉, 각 클래스당 하나의 모델을 통해 normal distribution를 학습하고, 새로운 입력이 이 normal distribution에서 벗어나면 해당 입력을 anomaly로 판단하는 방식



< One-class data distribution >

< One-class-one-model scheme >

Background

- Anomaly Detection (AD)

- 전체 이미지 (Image-level)

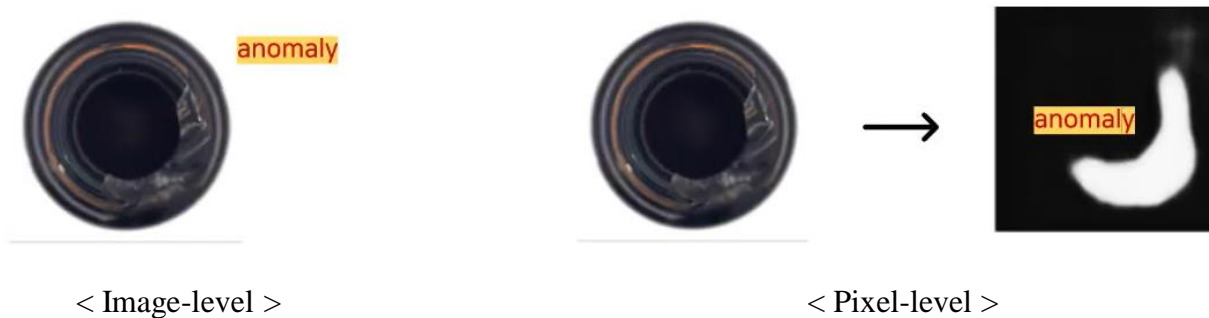
- 정상 데이터와 분명한 차이를 전체적인 이미지 구성을 기준으로 이상치를 판단함

- ※ 예시: 이미지 전체의 색상, 모양 등이 명확히 다른 경우

- 영역 (Pixel-level) – Anomaly segmentation

- 정상 데이터와의 미세한 차이를 픽셀 단위의 세부적인 변화를 기준으로 이상치를 판단함

- ※ 예시: 이미지 내부의 특정 영역에서 발생하는 작은 이상 신호



Background

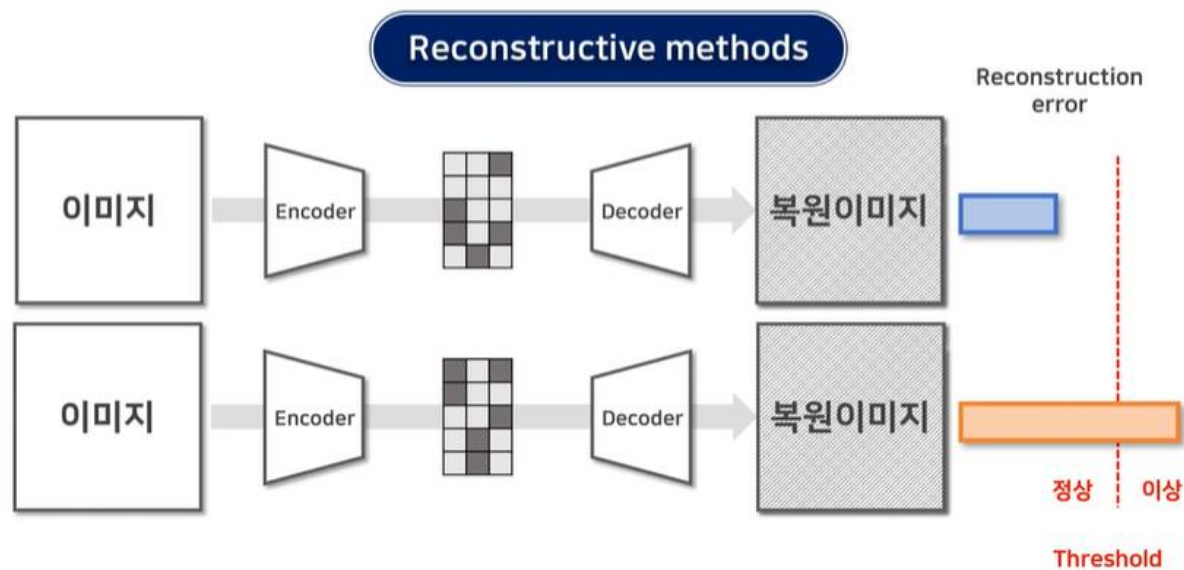
- Reconstruction-based Anomaly Detection

- Auto-encoders, GANs 등을 사용한 방법론들이 다수 존재함

- 정상(anomaly-free) 이미지만을 사용하여 reconstruction 하도록 network를 학습함

- Input 이미지와 reconstruction된 이미지 간의 차이를 threshold로 구분하여 anomaly detection을 수행함

- ※ Reconstruction error가 threshold보다 작으면 정상으로, 크면 이상으로 구분함



Background

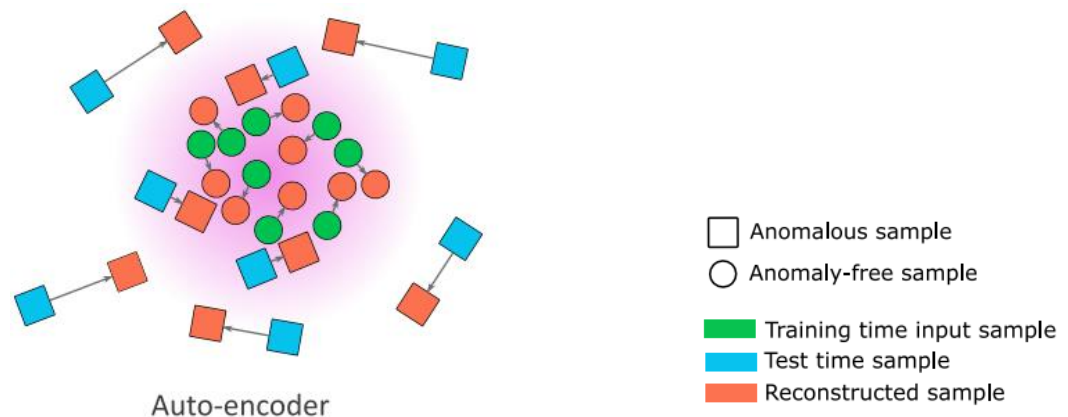
- Reconstructive methods의 한계

- 원인

- 정상 이미지와 차이가 크지 않은 이상치 이미지
 - 정상 이미지 (anomaly free) 데이터만을 사용하여 학습을 진행함

- 문제 발생 1: Over-generalize

- 모델이 특정 데이터를 지나치게 포괄적으로 처리하여, 정상 데이터와 이상 데이터를 제대로 구별하지 못하는 상황을 의미함
 - Autoencoder는 정상 데이터만으로 학습되기 때문에 이상 데이터도 정상 데이터처럼 잘 재구성해버려, 정상과 이상 간의 차이를 효과적으로 탐지하지 못하게 됨



Background

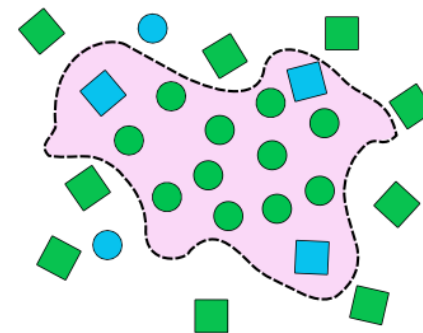
- Reconstructive methods의 한계

- 원인

- 정상 이미지와 차이가 크지 않은 이상치 이미지
 - 정상 이미지 (anomaly free) 데이터만을 사용하여 학습을 진행함

- 문제 발생 2: Over-fitting

- Discriminative network: 입력 데이터의 특징을 학습하여 정상 데이터와 이상 데이터를 명확히 구분하는 데 초점을 맞춘 네트워크
 - 실제로 이상 데이터가 부족하기 때문에, Discriminative network는 주로 합성 이상 데이터(Synthetic anomalies)를 사용하여 학습함
 - ※ 합성 데이터는 실제 이상 데이터와 다를 수 있으며, 이로 인해 네트워크가 합성 데이터의 특정 특징에만 overfitting 될 가능성이 높음
 - ※ 또한, 합성 이상 데이터는 데이터의 다양성이 낮을 수 있어, 네트워크가 다양한 실제 이상 데이터를 처리하는 데 실패할 수 있음
 - 결과적으로 모델이 실제 이상 데이터(real anomalies)에 대해 일반화 성능(generalization ability)을 상실하게 됨



Discriminative segmentation

Background

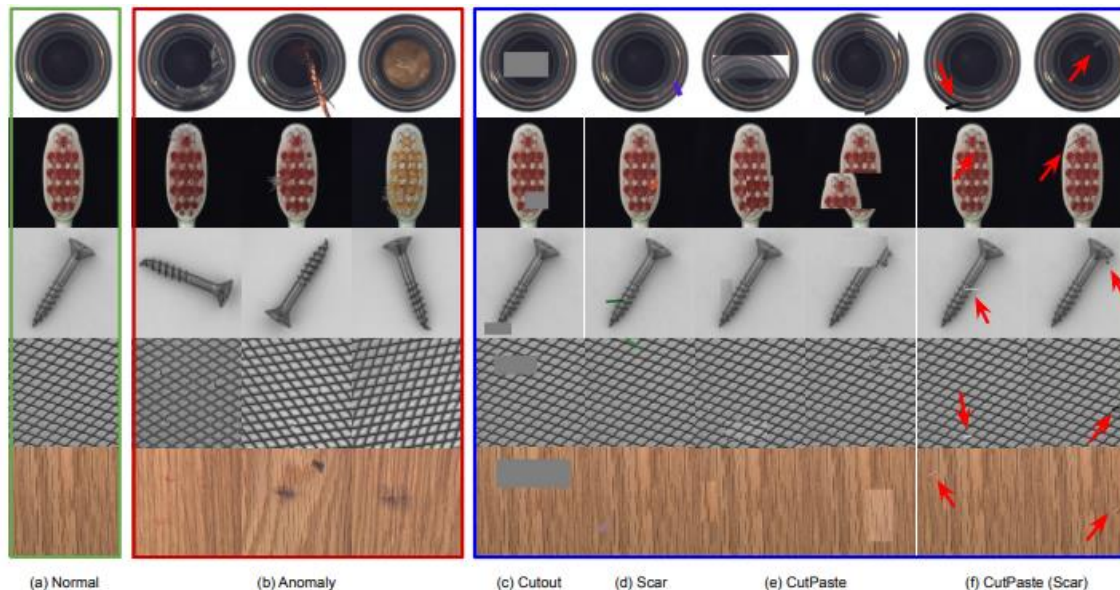
• CutPaste [1]

▪ Introduction

- 일반적인 Unsupervised AD는 정상 데이터만을 사용해 모델을 학습한 후, 새로운 이미지와의 feature를 비교하여 비정상 데이터를 탐지하는 방식임
- 그러나 이러한 접근법에는 한계가 존재하며, 이를 보완하기 위해 CutPaste는 비정상 데이터를 인위적으로 생성하여 학습하는 방법을 제안함

▪ Data Augment Method

- 직사각형 크기의 patch를 정하여 자른(cut) 뒤에 무작위로 위치를 선정하여 붙여 넣는(paste) 방법
- 추가로, 얇은 직사각형(scar) patch를 정하여 Cut and Paste 하는 방법도 함께 적용함



(a) Normal

(b) Anomaly

(c) Cutout

(d) Scar

(e) CutPaste

(f) CutPaste (Scar)

Background

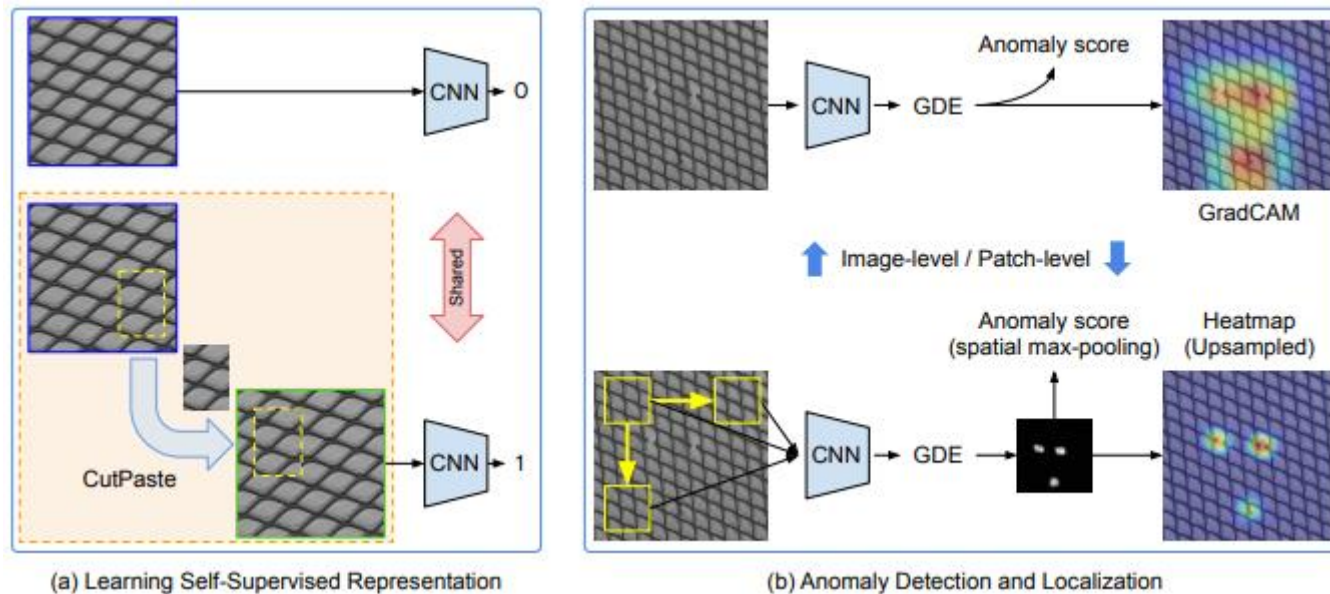
- CutPaste [1]

- Method

- Self-Supervised Learning (SSL)을 사용하여 정상 데이터의 feature를 학습하고, Gaussian Density Estimation (GDE)을 사용하여 정상 데이터의 분포를 기준으로 비정상 데이터를 탐지함

- ※ (a) 단계: 원본 정상 이미지 (0), CutPaste (1), CutPaste-Scar (2)로 classification을 수행하여 학습함

- ※ (b) 단계: GDE 방법을 사용하여 정상 데이터의 분포를 학습함



DRÆM

**A discriminatively trained reconstruction embedding
for surface anomaly detection**

Introduction

- Key Contributions

- Anomaly Generation

- 기존의 Discriminative segmentation 방식이 합성 데이터에 over-fitting되는 문제를 극복함
 - 합성 데이터가 실제 데이터와 정확히 일치할 필요가 없도록 설계하여 generalization 성능을 향상시킴

- Reconstructive Sub-network

- 정상 데이터만을 학습하므로 이상 데이터 없이도 학습이 가능하며, 재구성된 데이터는 Discriminative sub-network에 이상 탐지의 근거를 제공함

- Discriminative Sub-network

- 기존 방식에서 합성 이상 데이터에 over-fitting되는 문제를 극복하기 위해, DRAEM은 원본-재구성 간의 구조적 차이를 학습하도록 설계함
 - 이는 실제 이상 데이터에도 잘 일반화할 수 있는 모델을 제공함

- Surface Anomaly Localization & Detection

- 기존 방식은 이상 여부만을 탐지하거나 localization가 부정확한 경우가 많았지만, DRAEM은 픽셀 단위로 정확한 localization를 수행할 수 있음

Method

- Overview

- Anomaly generation

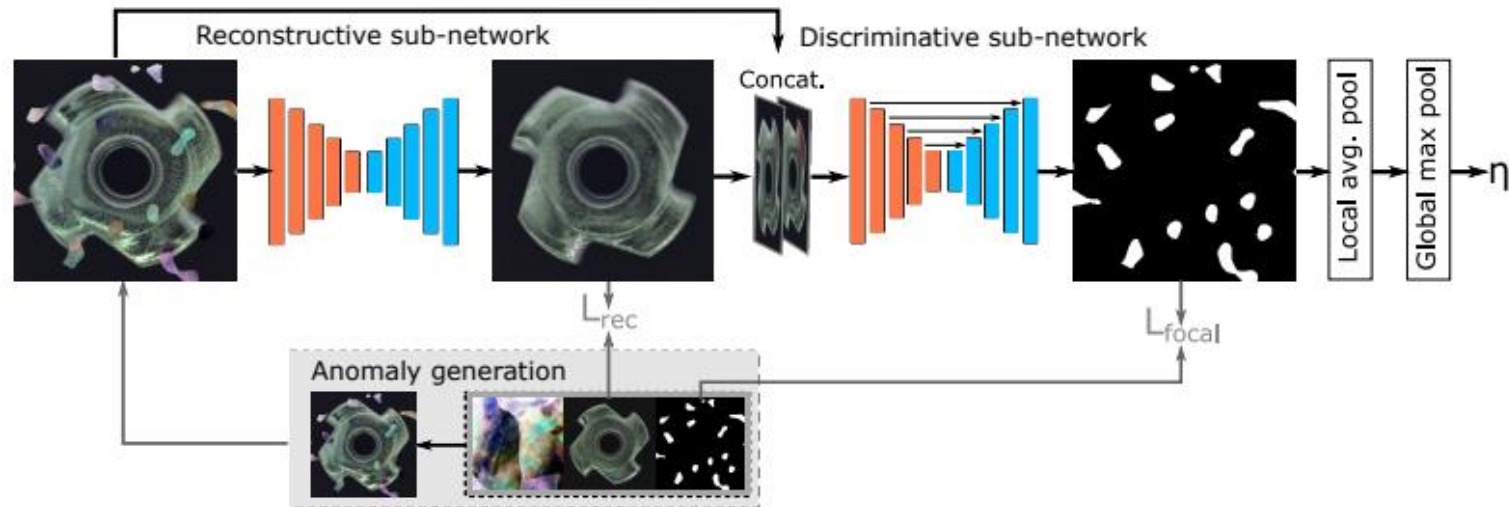
- 합성 이상 데이터 생성하는 역할을 수행함

- Reconstructive sub-network

- 입력 이미지를 정상 이미지로 재구성하는 역할을 수행함

- Discriminative sub-network

- Joint Representation을 기반으로 이상 영역을 탐지하고 Segmentation Map을 생성하는 역할을 수행함

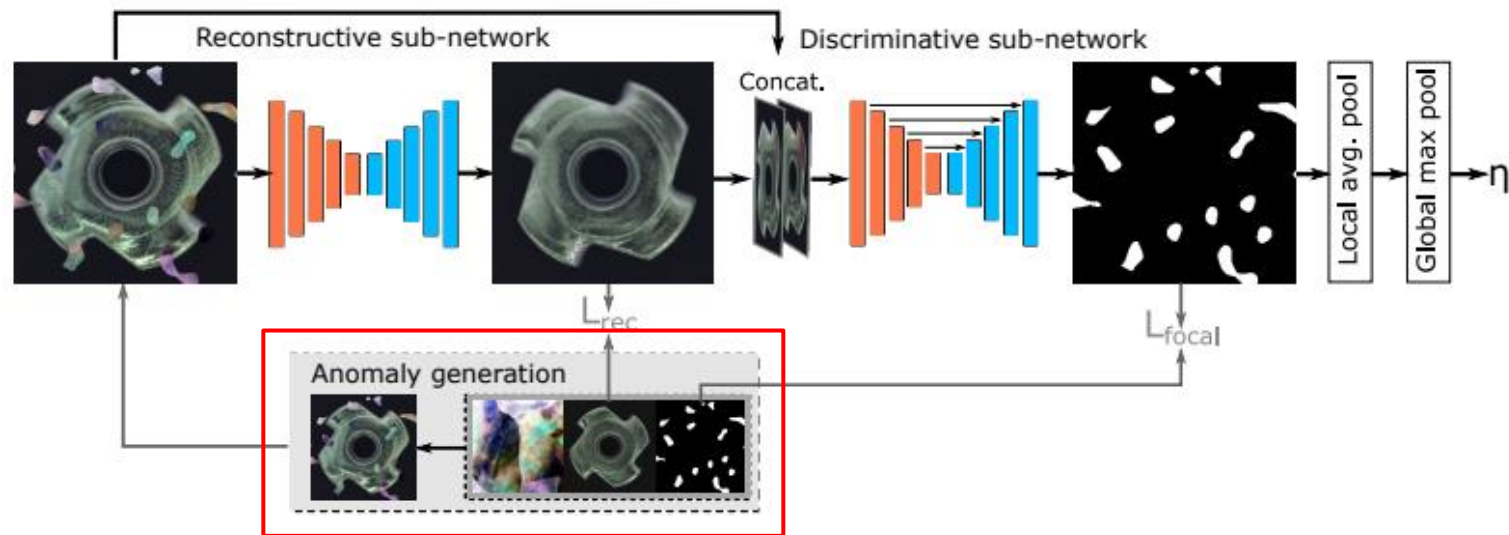


Method

- Anomaly generation

- Simulated anomaly generation

- DRAEM에서는 실제 이상치의 외형을 정확히 재현할 필요가 없음
 - 따라서, just-out-of-distribution appearance을 사용하여 이상 데이터를 생성함

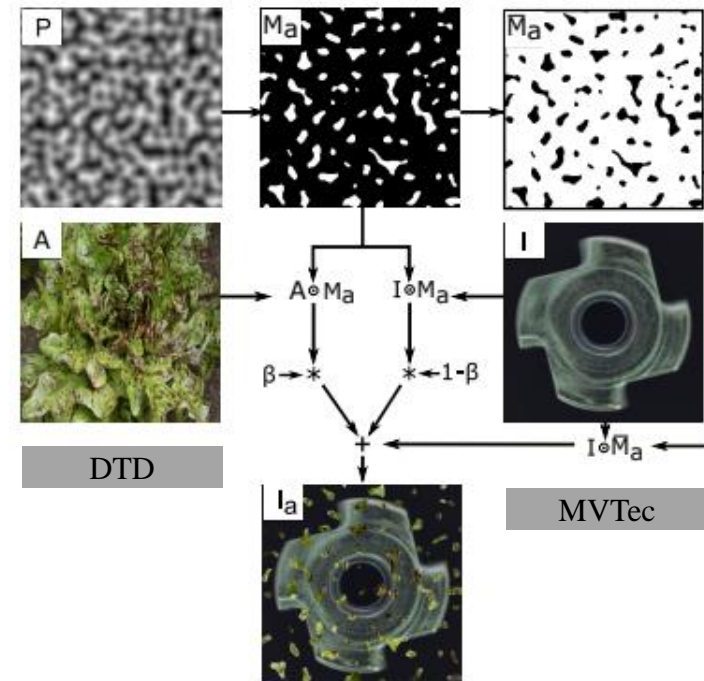


Method

- Anomaly generation

- Simulated anomaly generation

- P : Perlin Noise generator로 noise 이미지 생성
 - ※ Perlin Noise: 난수 sequence 생성
 - ✓ 다양한 패턴이 있는 texture를 표현할 때 이용 가능
 - ※ 다양한 anomaly shape 포착 가능
 - M_a : Anomaly map 이미지 masking
 - ※ Threshold를 이용한 binarize
 - $\overline{M_a}$: M_a 의 inverse
 - A : Anomaly texture source 이미지 (DTD 데이터셋 사용)
 - ※ RandAugment
 - ✓ {posterize, sharpness, solarize, equalize, brightness change, color change, auto-contrast}
 - ✓ 3개 랜덤하게 sampling하여 적용
 - ※ Input 이미지 분포와 다른 데이터셋으로부터 sampling



$$I_a = \overline{M_a} \odot I + (1 - \beta)(M_a \odot I) + \beta(M_a \odot A)$$

Method

- Anomaly generation

- Simulated anomaly generation

- A : Augmented texture 이미지

- ※ M_a 로 masking

- ※ I 와 함께 blending

- just-out-of distribution anomaly 생성

- β : Opacity 파라미터

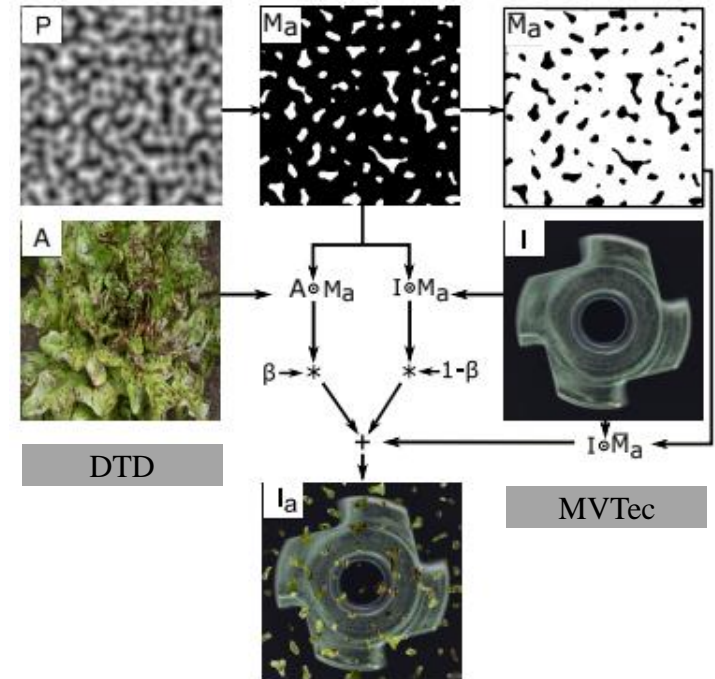
- ※ $\beta \in [0.1, 1.0]$ 에서 uniform하게 sampling

- I_a : Augmented training 이미지

- ※ Simulated anomaly를 포함하는 augmented 이미지



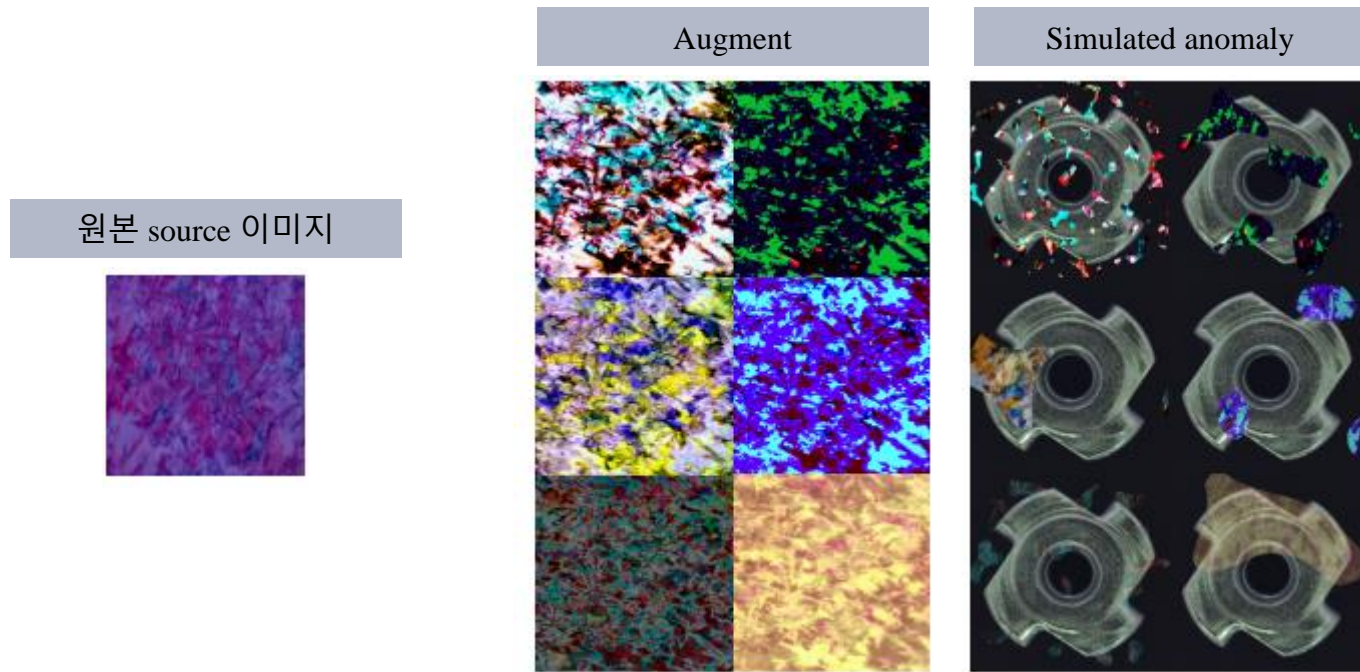
Training sample Triplet 생성



$$I_a = \overline{M}_a \odot I + \underbrace{(1 - \beta)(M_a \odot I)} + \underbrace{\beta(M_a \odot A)}$$

Method

- Anomaly generation
 - Simulated anomaly generation

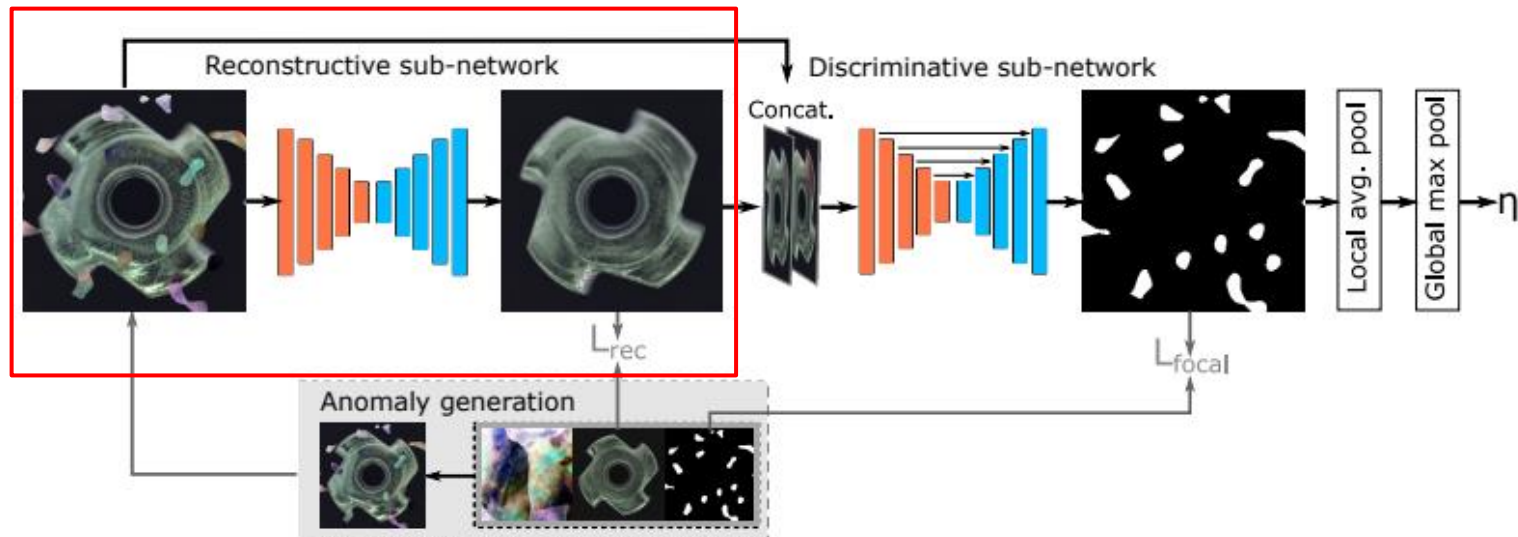


< Simulated anomaly generation 결과 >

Method

- Reconstructive sub-network

- Simulator에 의해 얻어진 corrupted version I_a 를 원본 이미지 I 로 재구성함
- Encoder-Decoder 구조 사용
- Input 이미지의 local pattern을 정상 sample의 분포에 가까운 pattern으로 변환하는 역할을 수행함



Method

- Reconstructive sub-network

- Loss function: L2 loss와 patch 기반의 SSIM loss를 결합하여 사용함

$$L_{rec}(I, I_r) = \lambda L_{SSIM}(I, I_r) + l_2(I, I_r)$$

- L2 loss (l_2)

- ※ 각 픽셀 값의 차이를 제공하여 평균을 구하는 방식

- ※ 즉, 원본 이미지와 복원된 이미지가 완전히 똑같다면 L2 loss는 0이 됨

- SSIM loss (L_{SSIM})

- ※ 이미지의 구조적 특성(밝기, 대비, 구조/패턴)을 비교하는 방식

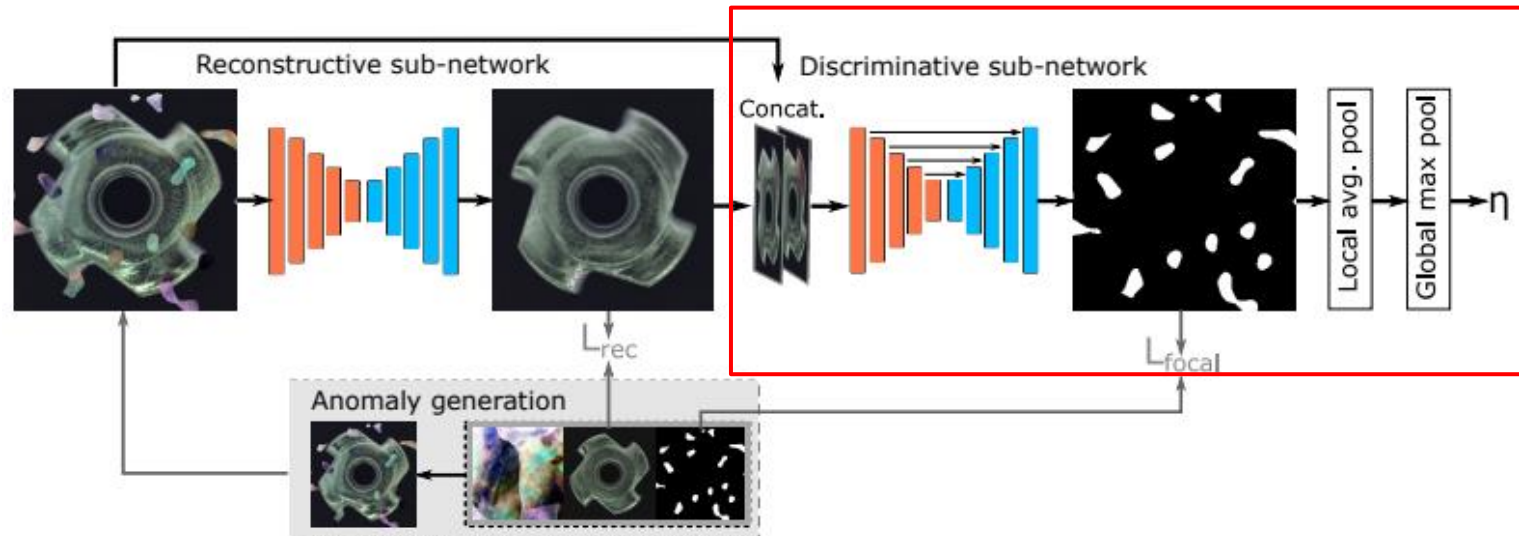
- ※ 즉, 단순히 픽셀 값의 차이를 비교하는 것이 아닌 전체적인 구조적 패턴의 유사도를 평가함

- λ : Loss balancing hyper-parameter

- 위의 2가지 loss를 함께 사용하면, 세부적인 차이와 전체적인 구조적 차이를 모두 고려 가능

Method

- Discriminative sub-network
 - Joint reconstruction anomaly embedding 학습함
 - U-Net과 유사한 구조 사용
 - Anomaly segmentation map 생성하는 역할을 수행함



Method

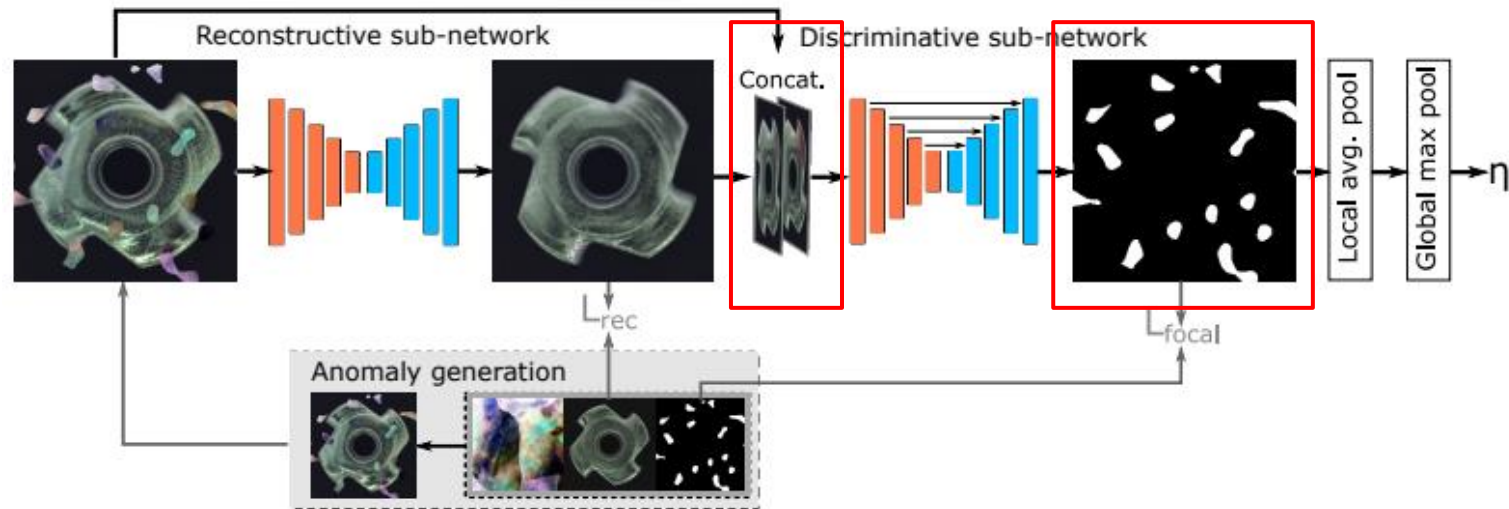
- Discriminative sub-network

- Input: $I_c = I + I_r$

- Reconstructive sub-network의 출력인 I_r + 원본 이미지 I
- Channel-wise concatenation

- Output: M_o

- Pixel-level anomaly detection mask (anomaly score map)
- 해당 mask는 I 와 동일한 크기



Method

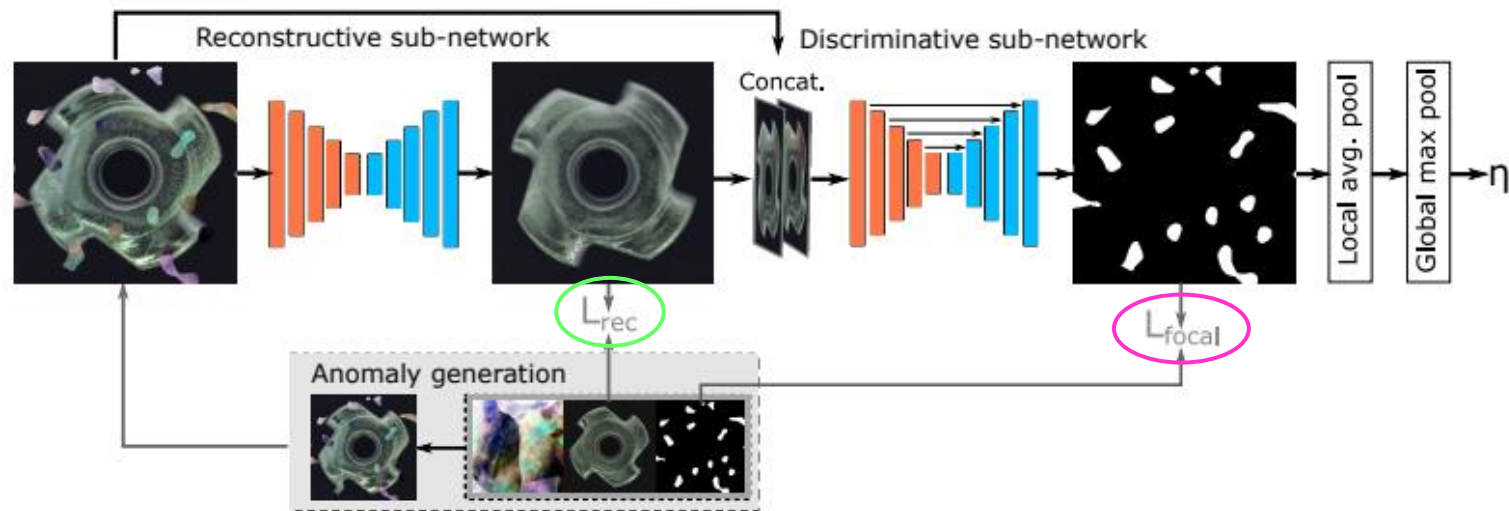
- Train

- Total loss: Reconstruction loss와 segmentation loss를 결합하여 사용함

$$L(I, I_r, M_a, M) = L_{rec}(I, I_r) + L_{seg}(M_a, M)$$

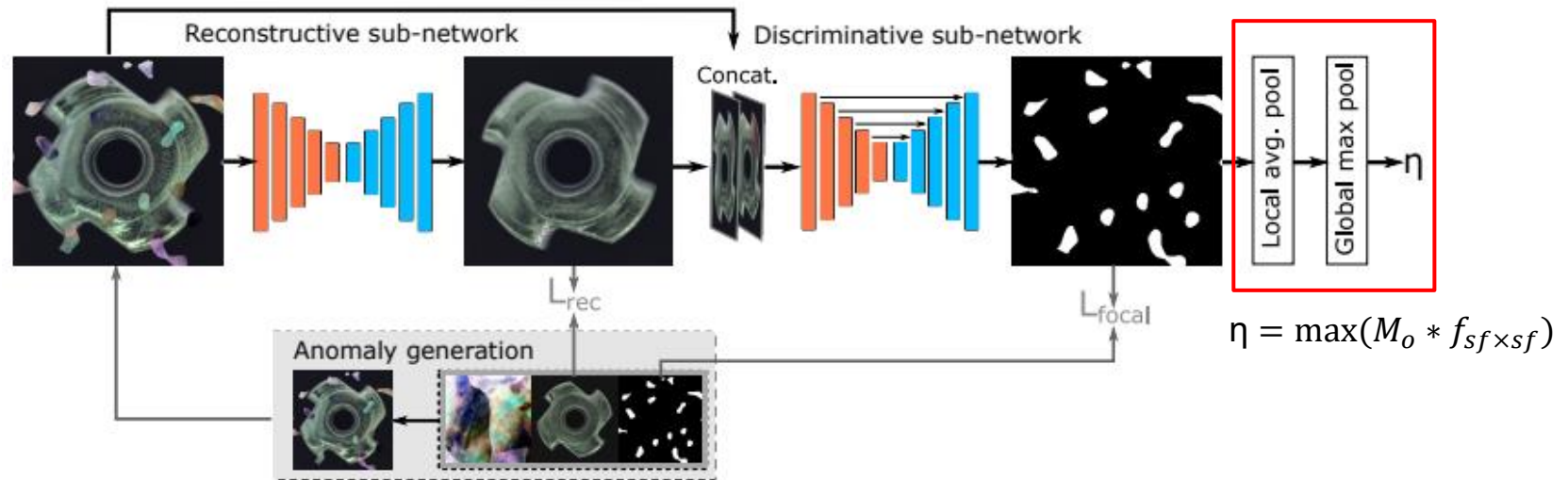
Focal loss

- M_a : Anomaly segmentation mask (anomaly binary mask)
- M : Ground Truth anomaly segmentation mask



Method

- Surface anomaly localization & detection
 - Discriminative sub-network의 Output: M_o
 - Pixel-level anomaly detection mask (anomaly score map)
 - Image-level anomaly score 예측
 - M_o smoothing: Local average pooling, Global max pooling 적용
 - 최종 image-level anomaly score (η) 산출



Method

- 기존의 한계 해결

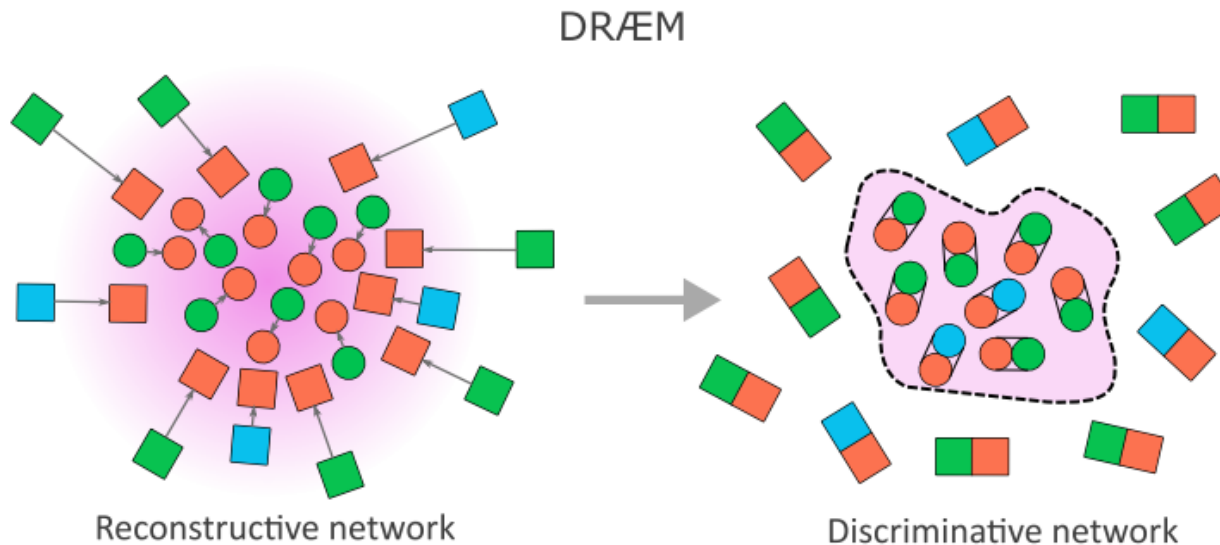
- 기존의 한계

- Auto-encoder: Anomaly에 over-generalize

- Discriminative segmentation: Synthetic anomaly에 over-fit이 되어, 실제 data에 generalize하지 못함

- DRAEM의 해결 방안

- Reconstructive sub-network와 Discriminative sub-network를 결합하여, Joint Representation을 학습하는 접근법을 제안함



Experiments

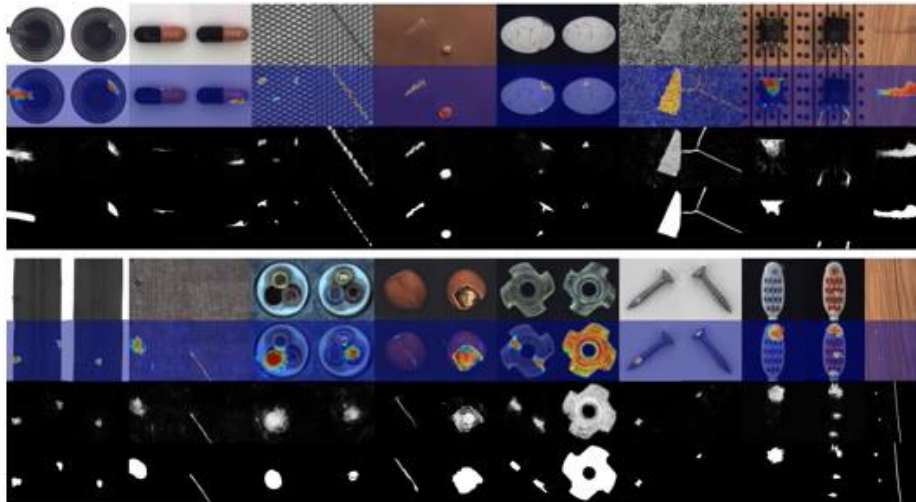
- Dataset

- MVTec

- Benchmark dataset
 - 15개의 object class를 포함함

- DTD (Describable Textures Dataset)

- Anomaly source dataset
 - Texture database: 5640장의 이미지를 포함함
 - 47개의 category로 나누어져 있음



< MVTec >



< DTD >

Experiments

- Metric

- AUROC

- Anomaly detection

- ※ Image-level

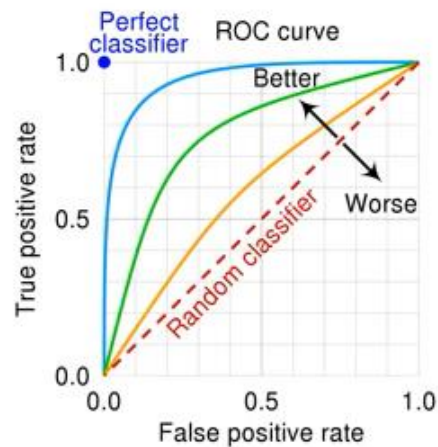
- Anomaly localization

- ※ Pixel-level

- Average Precision (AP)

- Class 불균형이 심한 경우에 사용하기 적절한 평가 지표

- ※ Class 불균형이 심한 surface anomaly detection에 적절함



< ROC Curve >

Experiments

- Surface anomaly detection

- 15개의 MVTEC 데이터셋 클래스 중 9개에서, 다른 모델보다 높은 성능을 보임

Class	[1]	[26]	[4]	[31]	[20]	[11]	DRÆM
bottle	79.4	98.3	99.0	99.9	100	99.9	99.2
capsule	72.1	68.7	86.1	88.4	92.3	91.3	98.5
grid	74.3	86.7	81.0	99.6	92.9	96.7	99.9
leather	80.8	94.4	88.2	100	100	100	100
pill	67.1	76.8	87.9	83.8	83.4	93.3	98.9
tile	72.0	96.1	99.1	98.7	97.4	98.1	99.6
transistor	80.8	79.4	81.8	90.9	95.9	97.4	93.1
zipper	74.4	78.1	91.9	98.1	97.9	90.3	100
cable	71.1	66.5	86.2	81.9	94.0	92.7	91.8
carpet	82.1	90.3	91.6	84.2	95.5	99.8	97.0
hazelnut	87.4	100	93.1	83.3	98.7	92.0	100.0
metal nut	69.4	81.5	82.0	88.5	93.1	98.7	98.7
screw	100	100	54.9	84.5	81.2	85.8	93.9
toothbrush	70.0	95.0	95.3	100	95.8	96.1	100
wood	92.0	97.9	97.7	93.0	97.6	99.2	99.1
<i>avg</i>	78.2	87.3	87.7	91.7	94.4	95.5	98.0

- [1]: GANomaly
- [26]: DAGAN
- [4]: Student-Teacher AD
- [31]: RIAD
- [20]: Gaussian AD
- [11]: Padim

Experiments

- Anomaly Localization

- 15개의 MVTec 데이터셋 클래스 중 11개에서, 다른 모델보다 높은 AP score을 보임

Class	US[4]	RIAD[31]	PaDim[11]	DRÆM
bottle	97.8 / 74.2	98.4 / 76.4	98.2 / 77.3	99.1 / 86.5
capsule	96.8 / 25.9	92.8 / 38.2	98.6 / 46.7	94.3 / 49.4
grid	89.9 / 10.1	98.8 / 36.4	97.1 / 35.7	99.7 / 65.7
leather	97.8 / 40.9	99.4 / 49.1	99.0 / 53.5	98.6 / 75.3
pill	96.5 / 62.0	95.7 / 51.6	95.7 / 61.2	97.6 / 48.5
tile	92.5 / 65.3	89.1 / 52.6	94.1 / 52.4	99.2 / 92.3
transistor	73.7 / 27.1	87.7 / 39.2	97.6 / 72.0	90.9 / 50.7
zipper	95.6 / 36.1	97.8 / 63.4	98.4 / 58.2	98.8 / 81.5
cable	91.9 / 48.2	84.2 / 24.4	96.7 / 45.4	94.7 / 52.4
carpet	93.5 / 52.2	96.3 / 61.4	99.0 / 60.7	95.5 / 53.5
hazelnut	98.2 / 57.8	96.1 / 33.8	98.1 / 61.1	99.7 / 92.9
metal nut	97.2 / 83.5	92.5 / 64.3	97.3 / 77.4	99.5 / 96.3
screw	97.4 / 7.8	98.8 / 43.9	98.4 / 21.7	97.6 / 58.2
toothbrush	97.9 / 37.7	98.9 / 50.6	98.8 / 54.7	98.1 / 44.7
wood	92.1 / 53.3	85.8 / 38.2	94.1 / 46.3	96.4 / 77.7
<i>avg</i>	93.9 / 45.5	94.2 / 48.2	97.4 / 55.0	97.3 / 68.4

AUROC / AP

Conclusions

- DRAEM: Discriminatively trained reconstruction anomaly embedding model
 - Anomaly 이미지와 anomaly-free reconstruction 이미지의 특징을 함께 학습하여, 효과적인 joint representation을 생성함
 - 이를 통해, 이상 탐지 성능을 향상시킴
 - 간단하고 general한 anomaly simulation을 사용하여 모델을 학습함
 - 이를 통해, 다양한 이상 패턴을 탐지할 수 있음

감사합니다!