

2024 여름 세미나

2024.07.19



Sogang University

Vision & Display Systems Lab, Dept. of Electronic Engineering



Presented By

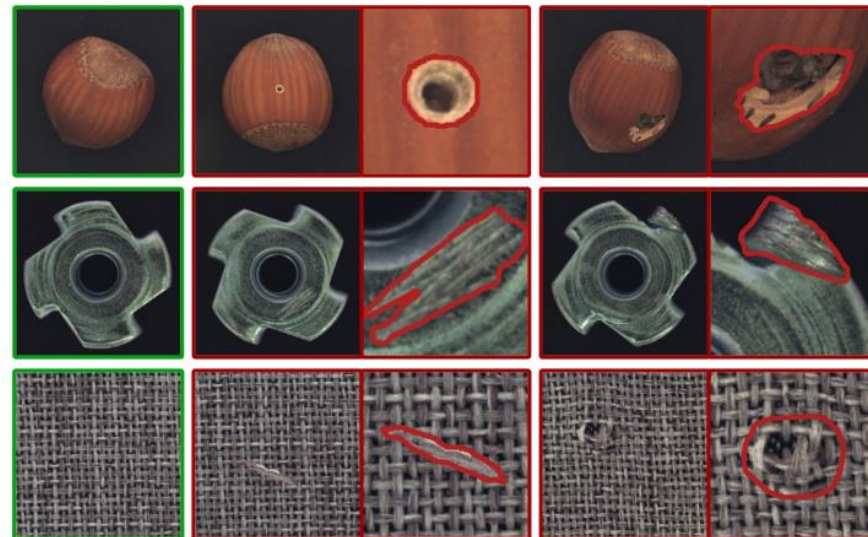
이준호

Outline

- Background
 - What is anomaly detection
 - Reconstruction-based anomaly detection
 - Text-based anomaly detection
- Paper 1
 - Text-Guided Variational Image Generation for Industrial Anomaly Detection and Segmentation (2024-CVPR)
- Paper 2
 - RealNet: A Feature Selection Network with Realistic Synthetic Anomaly for Anomaly Detection (2024-CVPR)

Background

- What is anomaly detection
 - Normal(정상) sample과 abnormal(비정상, 이상치, 특이치) sample을 구별하는 문제
 - Anomaly detection은 제조업뿐만 아니라 CCTV, 의료 영상, Social Network 등 다양한 분야에서 응용되면서 중요한 분야로 인식



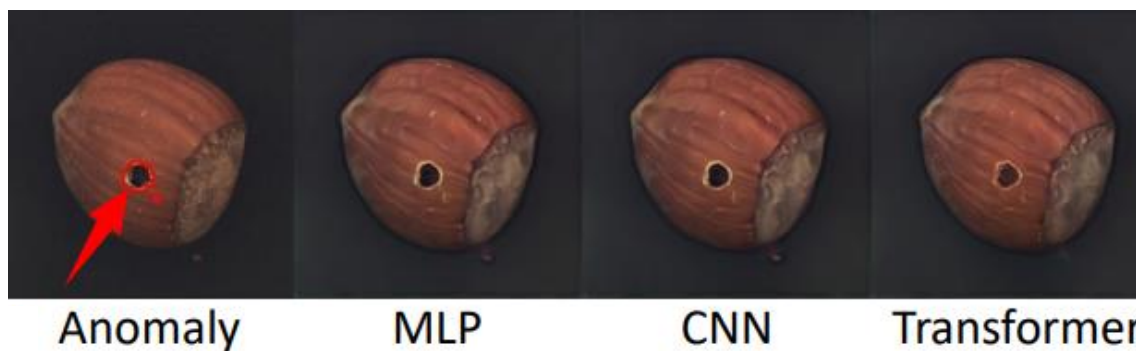
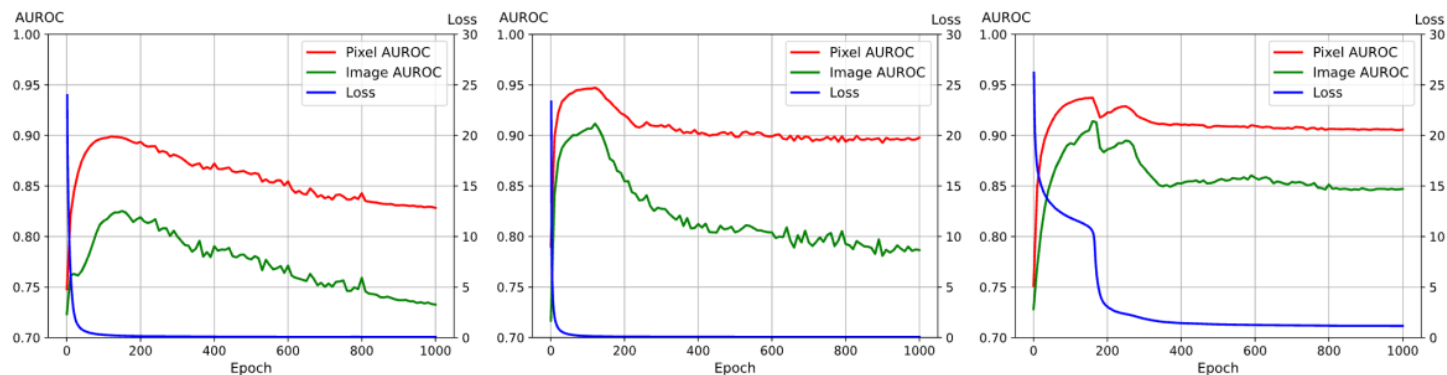
< 산업용 anomaly dataset의 예시 >

Background

- Reconstruction-based anomaly detection
 - Input 이미지를 reconstruct하는 과정을 통해 normal 데이터의 distribution을 train하는 방법론들이 trend
 - Normal 데이터의 distribution을 학습한 model은 normal 데이터의 distribution으로부터 멀리 떨어진 데이터의 reconstruct capability가 좋지 않을 것이라는 가정 하에 reconstruction error가 큰 sample을 anomaly로 판별함
 - 하지만 reconstruction-based 방법론들은 identical shortcut 문제가 발생함
 - Identical shortcut : normal/ abnormal sample을 모두 잘 복원하고 anomaly detection에 실패하는 문제
 - ※ Fully-connected layer in MLP
 - ※ Convolutional layer in CNN
 - ※ Transformer with query embedding

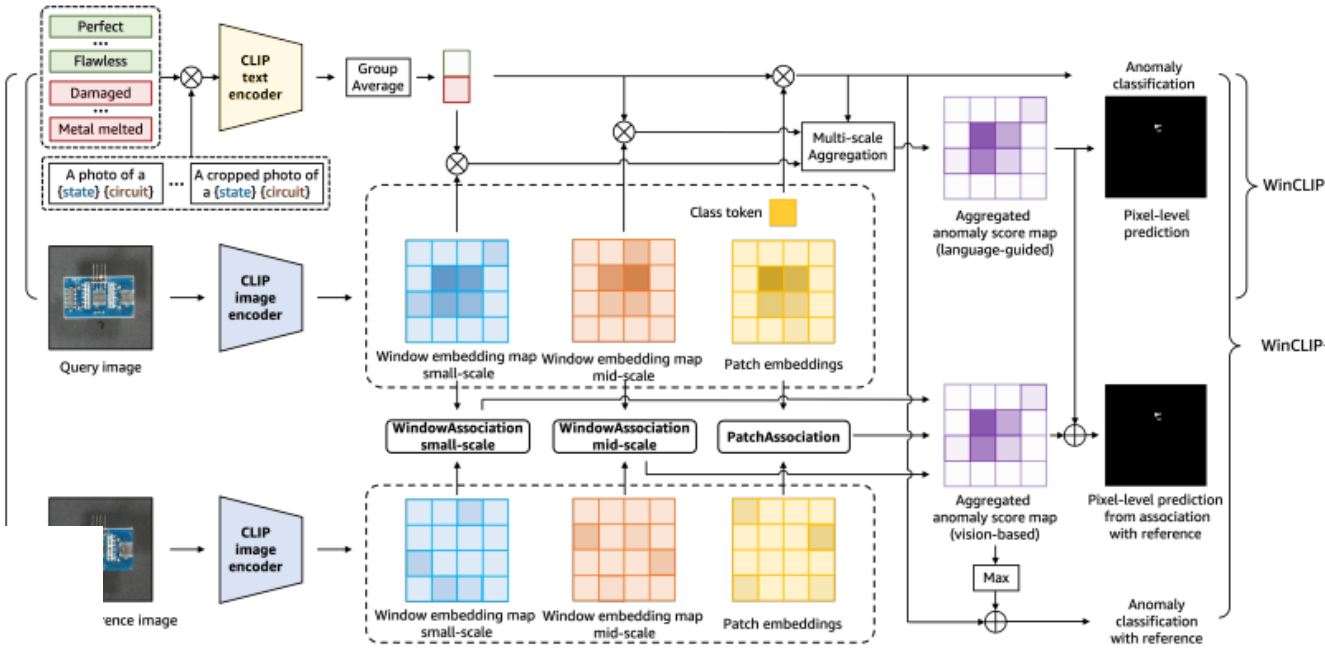
Background

- Reconstruction-based anomaly detection
 - Identical shortcut : normal/ abnormal sample을 모두 잘 복원하고 anomaly detection에 실패하는 문제



Background

- Text-based anomaly detection
 - Reconstruction-base와 달리 text-based method는 다른 setting과 method를 사용함
 - WinCLIP은 state words와 text candidate를 mix하여 input image에 대해 pre-text prompts를 생성함
 - Binary mask를 통해 input image에 대한 multi-scale features를 추출하고 생성된 prompt 간의 similarity score를 계산하여 anomaly를 판별함



< WinCLIP 모델 구조 >

Text-Guided Variational Image Generation for Industrial Anomaly Detection and Segmentation

Paper 1

- Abstract & Introduction

- Problem

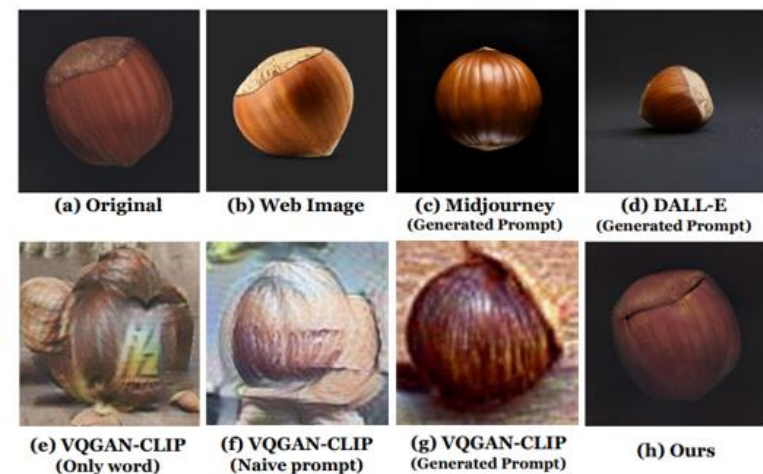
- Imbalanced non-defective(normal) data

- ※ Normal data가 대부분 서로 매우 유사하거나 uniform한 특성을 보이는데, 매우 작은 defect에 대한 detection 성능에 악영향

- Defect의 정도는 촬영하는 기계의 종류와 관계가 있음

- Wrong labeling

- ※ Anomaly data를 normal data로 잘못 labeling함



Method	Detection	Segmentation
Original Image	88.3	94.3
Web Image	84.8	93.1
Midjourney (by Generated Prompt)	82.7	93.3
DALL-E (by Generated Prompt)	85.7	91.9
VQGAN-CLIP (by Only word)	64.8	86.8
VQGAN-CLIP (by Naïve Prompt)	71.9	88.2
VQGAN-CLIP (by Generated Prompt)	81.4	91.7
Ours	93.3	94.6

Paper 1

- Abstract & Introduction

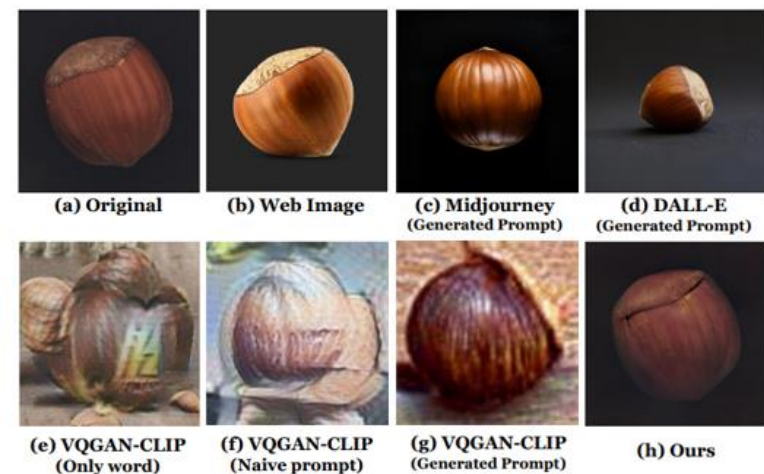
- Motivation

- Industrial manufacturing에서 anomaly detection을 위한 clean data를 얻기 위해 제안

- Summary

- Target object에 대한 text 정보를 extensive text library에서 학습하여, input 이미지와 유사한 normal 이미지를 생성하는 데 활용함

- 생성된 normal 이미지가 text & image-based knowledge에서 파생된 예상 distribution과 일치하도록 보장하여 stability, generality 향상



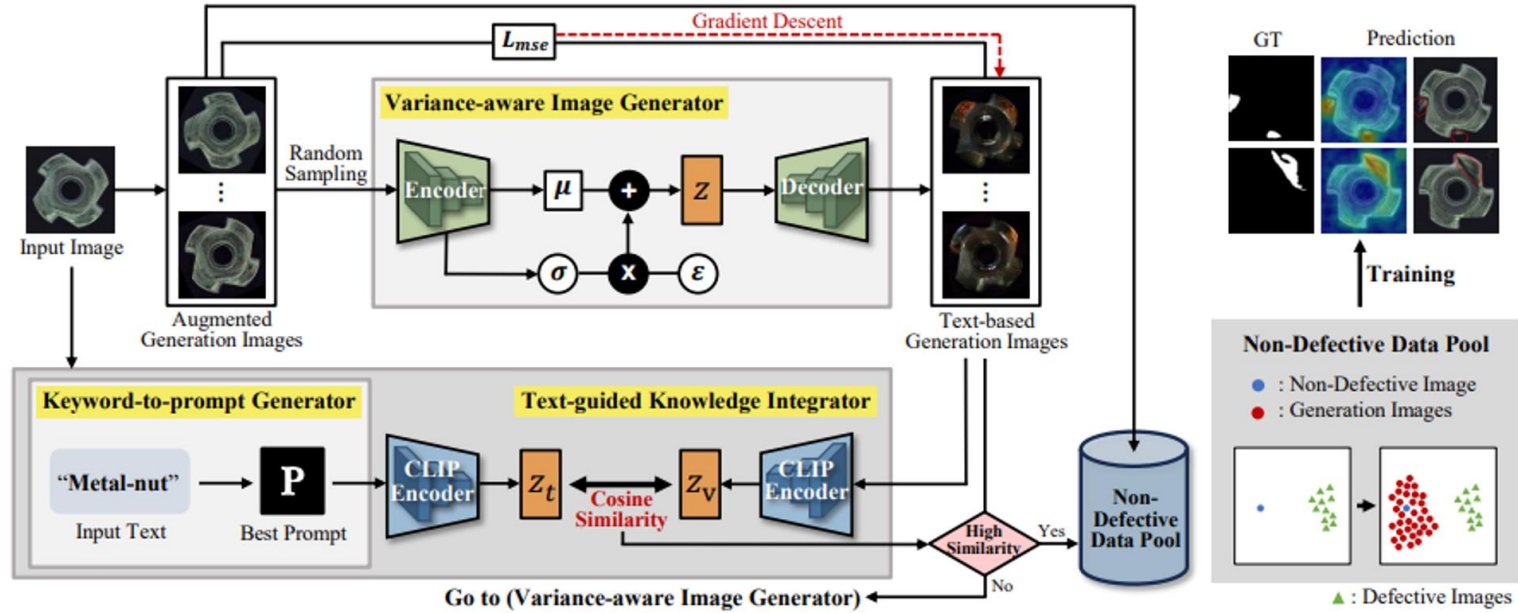
Method	Detection	Segmentation
Original Image	88.3	94.3
Web Image	84.8	93.1
Midjourney (by Generated Prompt)	82.7	93.3
DALL-E (by Generated Prompt)	85.7	91.9
VQGAN-CLIP (by Only word)	64.8	86.8
VQGAN-CLIP (by Naïve Prompt)	71.9	88.2
VQGAN-CLIP (by Generated Prompt)	81.4	91.7
Ours	93.3	94.6

Paper 1

- Method

- Overview

- Keyword-to-prompt generator
 - Variance-aware image generator
 - Text-guided knowledge integrator



< 제안 모델 구조 >

Paper 1

- Method

- Keyword-to-prompt Generator

- Object name W 와 original training image I 의 가장 적절한 prompt P 를 제공함
 - WordNet에 기반하여, T-different word를 설계하여 candidate prompts를 얻음
 - 획득한 candidate prompts에서 최적의 prompts를 찾음

※ 찾는 기준은 2가지가 있음

- ✓ Distance-based outlier removal

- Candidate prompts의 latent features와 original image간의 L2 distance를 계산하여 outlier prompts를 제거함

- ✓ Embedding similarity estimation

- CLIP encoder를 사용하여 image와 text embedding vectors를 얻음
 - Cosine similarity를 계산한 뒤, argmax 취함

Table 4. Prompt examples by Keyword-to-prompt Generator.

Best prompt	Worst prompt
“a {hazelnut} with {cobnut}”	“a {hazelnut} with {decantherous}”
“a {metalnud} with {metallical}”	“a {metalnud} with {predegenerate}”
“a {zipper} with {metallization}”	“a {zipper} with {Echinops}”
“a {capsule} with {incapsulation}”	“a {capsule} with {perceptible}”
“a {toothbrush} with {parazoan}”	“a {toothbrush} with {chaetopod}”

Paper 1

- Method

- Variance-aware Image Generator

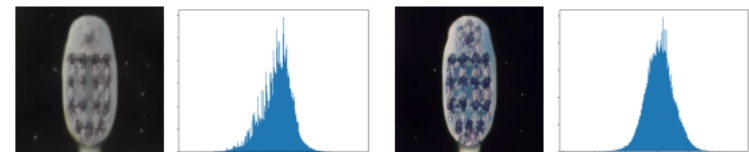
- 원본 이미지의 visual variance와 유사한 이미지를 생성해야 하기 때문에, latent vector의 distribution을 표준 정규분포로 근사하는 VQGAN을 training image 생성 모델로 사용

- ※ 하지만 VQGAN은 latent vector를 사전에 정의된 code로 근사하여 patch들의 diversity를 해치는 문제가 있음

- ※ 이를 방지하기 위해, 각 patch의 latent vector를 code가 아닌 augmented image로부터 얻은 latent vector의 평균을 사용하여 decoder에 넣음

- ✓ 이 과정으로 diversity를 부여함

Component	Image-Level (%)	Gain (%)	Pixel-Level (%)	Gain (%)
Baseline	74.5(±1.6)	-	97.0(±0.1)	-
1) Text prompt				
w/o Text	77.6(±0.6)	(+3.1%)	97.1(±0.0)	(+0.1%)
w/ Text (ours)	79.1(±0.6)	(+4.6%)	96.9(±0.0)	(-0.1%)
2) N-Generated images				
1-copy	78.6(±0.4)	(+4.1%)	97.0(±0.0)	(+0.0%)
10-copy	77.9(±0.4)	(+3.4%)	96.9(±0.0)	(-0.1%)
20-copy	78.4(±0.5)	(+3.9%)	97.0(±0.0)	(+0.0%)
30-copy	79.0(±0.7)	(+4.5%)	97.0(±0.0)	(+0.0%)
50-copy	78.5(±0.3)	(+4.0%)	96.9(±0.1)	(-0.1%)
100-copy	78.4(±0.5)	(+3.9%)	96.9(±0.0)	(-0.1%)
3) Variance-aware				
w/o Variance	74.3(±0.4)	(-0.2%)	97.0(±0.0)	(+0.0%)
w/ Variance (ours)	78.9(±0.5)	(+4.3%)	97.3(±0.0)	(+0.3%)



w/o Variance

w/Variance (ours)

Paper 1

- Method

- Text-guided Knowledge Integrator (1/2)

- 가장 적합한 prompt P와 잘 맞는 non-defective images를 생성하여, 이를 anomaly detection model을 위한 normal data pool에 추가함
 - Keyword-to-prompt generator에서 선택한 P로부터, CLIP text encoder를 통해 text clip feature를 추출
 - 동시에 variance-aware image generator를 이용하여 image를 생성함
 - ※ 모든 image feature를 평균화하여 visual clip feature가 추정됨
 - Image set을 생성하는 것을 반복하며, image feature expectation과 best prompt feature 사이의 cosine similarity에 따라 image set을 scoring함
 - ※ Cosine similarity에 기반하여, 생성된 image의 best set을 선택하고, 이후 anomaly detection training에 사용함

Paper 1

- Method

- Text-guided Knowledge Integrator (2/2)

- 새로운 image set을 생성할 때마다, 동시에 variance-aware image generator를 update함

- ※ 위 작업은 2가지 목적이 있음

- ✓ Normal input과 유사한 image를 generator가 생성할 수 있도록 하기 위함

- ✓ 생성된 image set의 diversity를 향상시킬 수 있음

- Encoder, decoder, codebook vector parameters를 포함한 variance-aware image generator의 모든 parameter를 update함

- Input과 유사한 image를 생성하기 위해 MSE loss를 활용하며, Adam optimizer로 update함

Paper 1

- Experiments

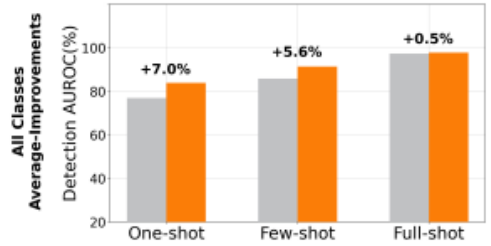
- 정량적 실험 결과

- 일반화 성능을 검증하기 위해, one-shot, few-shot, full-shot로 실험 진행

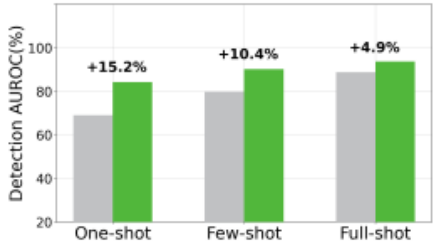
- ※ 모든 시나리오에서 성능 향상

- One-shot에서 성능 향상이 가장 높은 5가지 class에 대한 분석이 포함됨

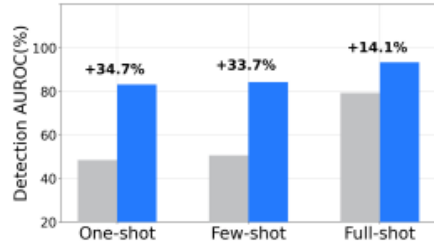
Patchcore vs Ours



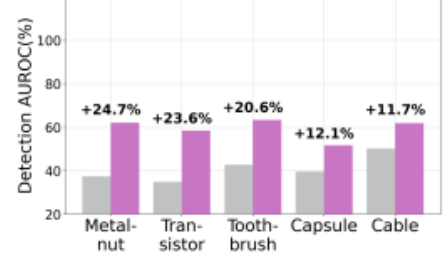
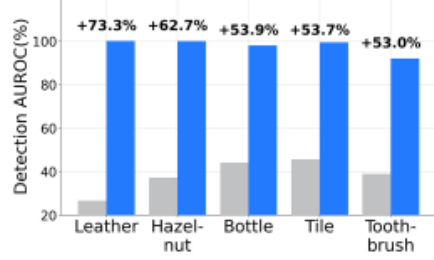
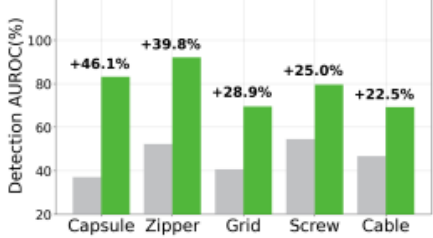
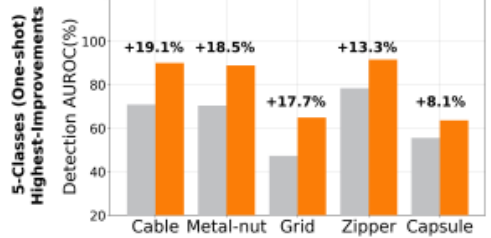
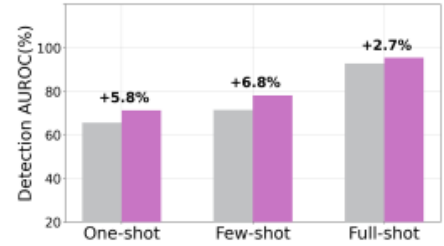
Cflow vs Ours



Reverse Distillation vs Ours



Efficient-AD vs Ours



< MVTec-AD 데이터셋에 대한 실험 결과 >

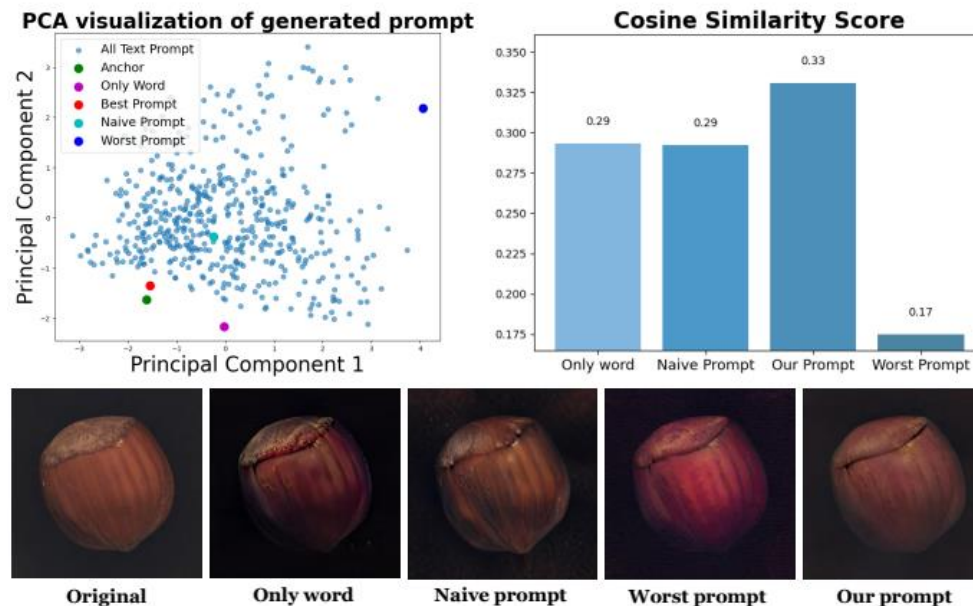
Paper 1

- Experiments

- Optimizing results between our prompt and generated images (1/2)

- Keyword-to-prompt Generator가 최적의 prompt를 성공적으로 찾는 결과와 variance-aware image generator 및 text-guided knowledge integrator가 최적의 prompt 기반으로 생성한 이미지를 시각화 함
- 최악의 prompt와 최적의 prompt가 뚜렷한 차이를 보임

※ 두 prompt간의 유사성을 cosine similarity score로 측정함



Original Only word Naive prompt Worst prompt Our prompt

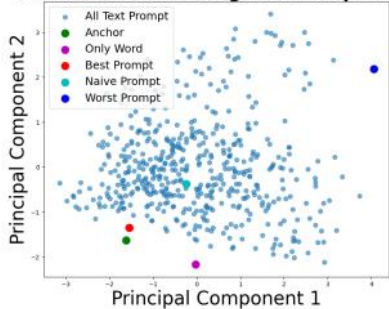
Paper 1

• Experiments

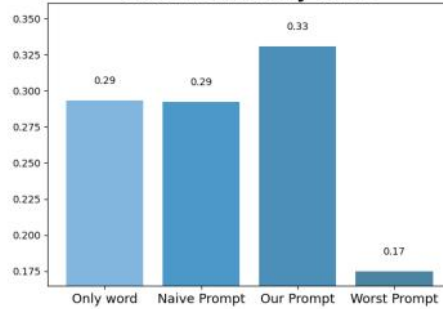
▪ Optimizing results between our prompt and generated images (2/2)

- 최악의 prompt와 최적의 prompt 사이에는 뚜렷한 결과가 보임
- 아래 실험은 최적의 prompt가 normal 이미지와 유사한 특성을 가지고 있어 text-guided knowledge integrator의 성능을 향상시킬 수 있음을 증명함
- Table은 생성한 이미지가 원본 이미지와 가장 유사함을 정량적으로 증명함

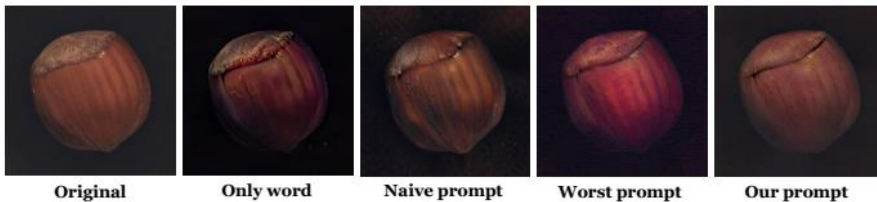
PCA visualization of generated prompt



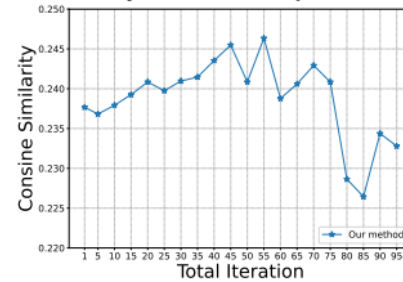
Cosine Similarity Score



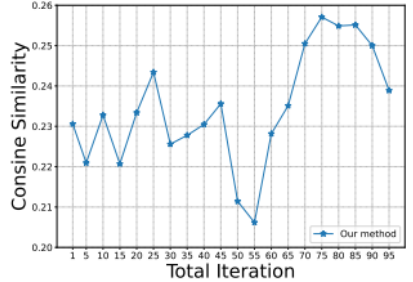
Method	Prompt	IS(↑)	SSIM(↑)	PSNR(↑)	LPIPS(↓)
Original	N/A	1.00	1.00	-	0.00
Only word	"a hazelnut"	1.62	0.50	19.01	984.15
Naïve prompt	"a photo of hazelnut"	1.34	0.78	23.46	764.48
Worst prompt	"a {hazelnut} with {decantherous}"	4.56	0.79	23.57	782.75
Our prompt	"a {hazelnut} with {cobnut}"	22.59	0.90	28.39	701.84



Similarity between Prompt and Images



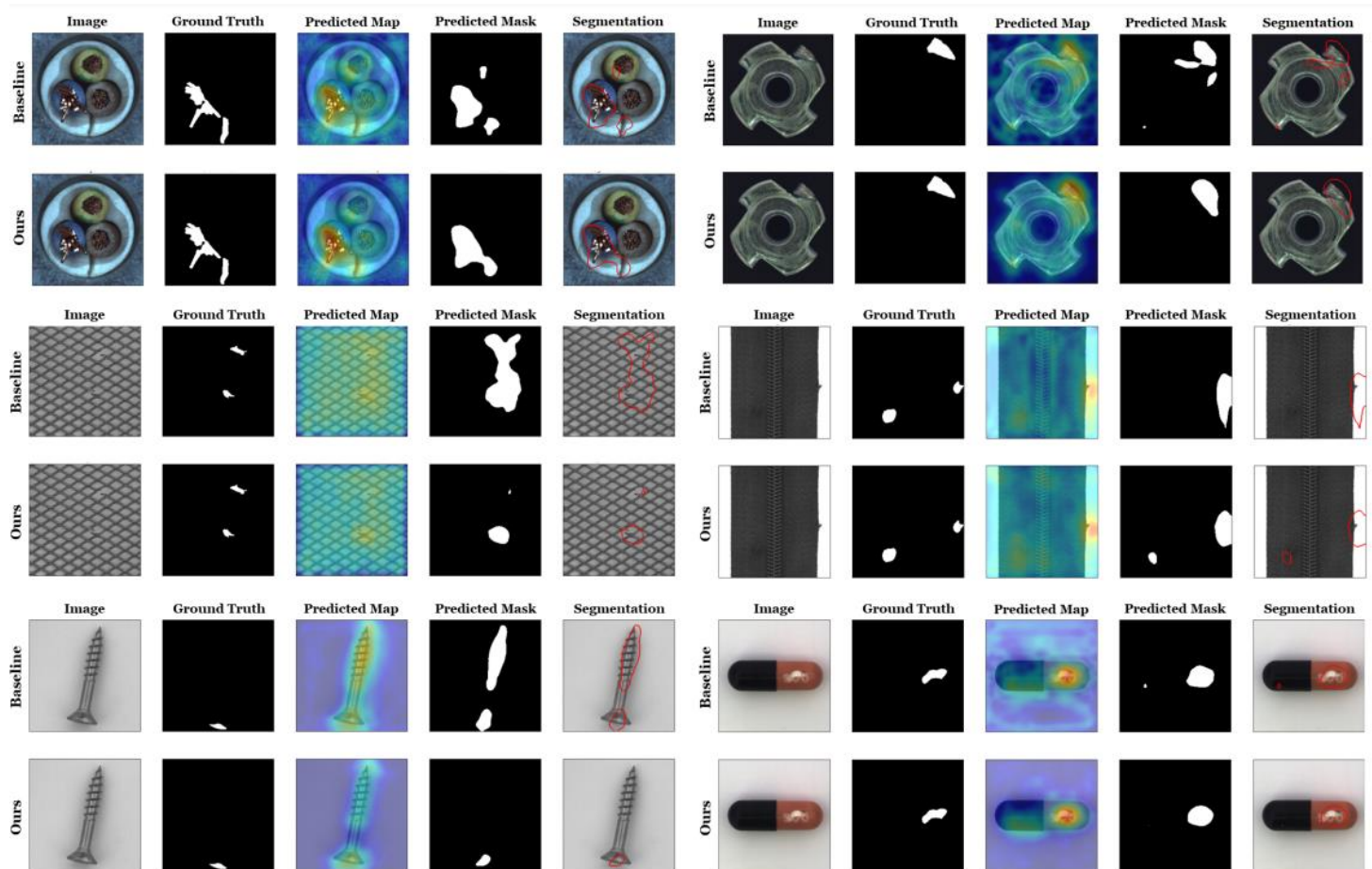
Similarity between Prompt and Images



Paper 1

- Experiments

- 정성적 실험 결과



RealNet: A Feature Selection Network with Realistic Synthetic Anomaly for Anomaly Detection

Paper 2

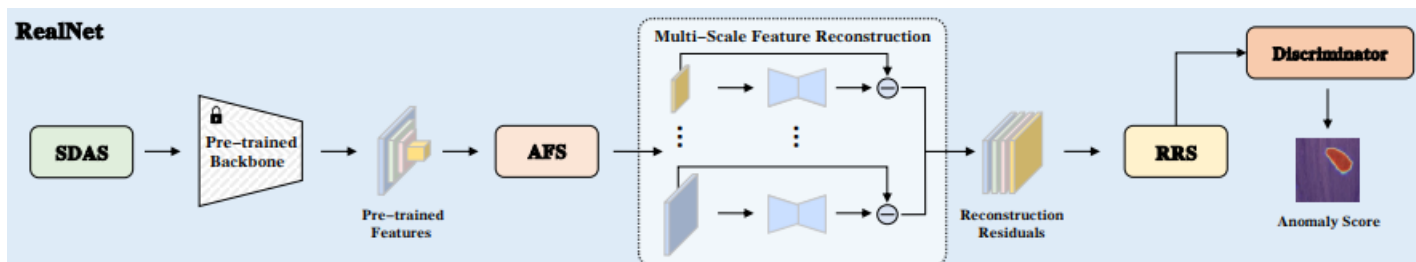
• Abstract & Introduction

▪ Problem

- 기존의 방법들은 anomaly 이미지와 anomaly categories에 대한 prior knowledge가 부족함
 - ※ Crafted data augmentation strategies 또는 external data에 의존하여 합성해야 함
 - ✓ 이는 실제 이미지와 합성 이미지의 distribution간의 상당한 차이가 발생함
 - ✓ 따라서 real-world에서 응용이 어려움

▪ Solution (Contribution)

- 위 문제를 해결하기 위해 Strength-controllable Diffusion Anomaly Synthesis (SDAS) 제안
 - ※ Natural distribution에 더 가까운 다양한 sample을 생성하고 anomaly strength를 제어
 - ※ Normal sample의 distribution을 모델링하기 위해, DDPM을 사용
- Anomaly 이미지를 normal 이미지의 feature로 reconstruct하고 차이점을 이용해 anomaly detection, localization 수행



Paper 2

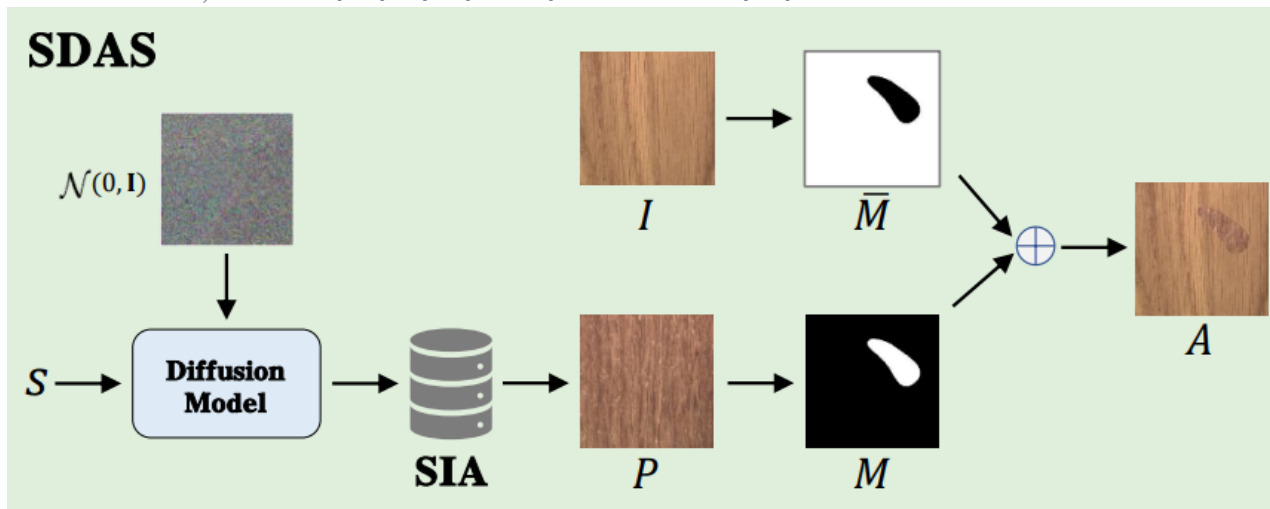
- Method

- Strength-controllable Diffusion Anomaly Synthesis (SDAS)

- DDPM (forward process), Markov chain (reverse process)을 사용하여 normal 이미지의 distribution 학습
- 다양한 anomaly 형태를 위해, perlin noise generator 사용하고 이를 binarization하여 mask M 을 생성
- 마지막으로 아래 알고리즘을 수행하여 최종 합성 이미지 생성

$$A = \bar{M} \odot I + (1 - \delta)(M \odot I) + \delta(M \odot P)$$

※ $\bar{M} = 1 - M$, δ 는 이미지의 불투명도를 의미

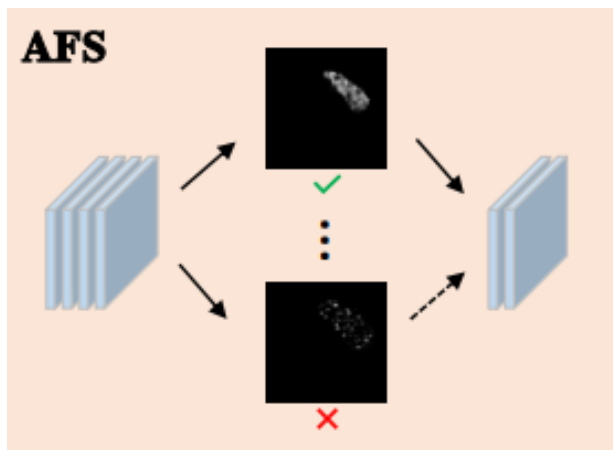


Paper 2

- Method

- Anomaly-aware Features Selection (AFS)

- Reconstruction에 사용하기 위해 SDAS에서 생성한 이미지를 pretrained network에 넣어 k번째 layer의 output을 선택함으로써 feature를 추출함
 - Pretrained network에 원본 이미지와 생성한 anomalous 이미지를 넣고 각 output 간의 차이를 구한 뒤, mask M과 같은 resolution을 갖도록 하여 M과의 차이를 loss로 사용
 - ※ 이때 layer는 다양한 layer를 선택하기 때문에 multi-scale임
 - feature dimensionality를 줄이고, pre-training bias를 제거하며, reconstruction cost를 관리하는 self-supervised method



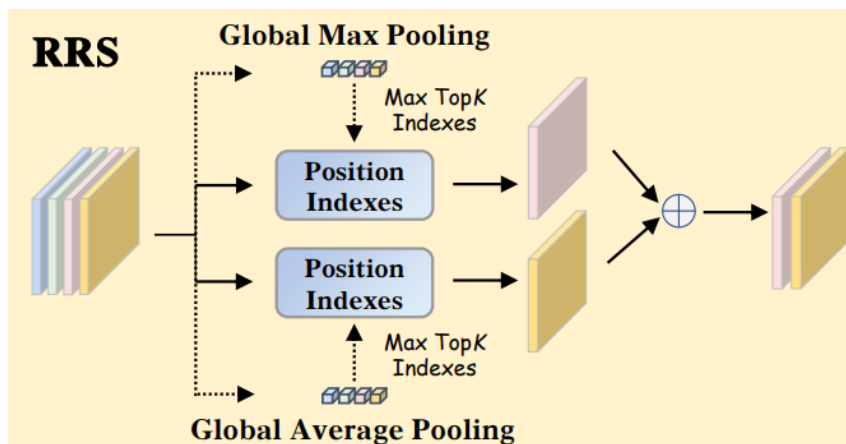
$$\mathcal{L}_{AFS}(\phi_{k,i}) = \frac{1}{N} \sum_{n=1}^N \|F([\phi_{k,i}(A_n) - \phi_{k,i}(I_n)]^2) - M_n\|_2^2$$

Paper 2

• Method

▪ Reconstruction Residuals Selection (RRS)

- 가장 anomalous information을 갖는 reconstruction residuals의 subset을 선택함
- Reconstruction residuals에 global max pooling (GMP), global average pooling (GAP)을 사용
 - ※ 이는 각각 top k reconstruction residuals를 의미
 - ※ GMP, GAP는 각각 local, global properties를 의미함
 - ✓ GMP, GAP는 각각 local anomalies, global anomalies를 포착하는데 효과적
 - ※ 두 case를 합쳐 various scales의 anomalies를 포착함
- 마지막으로 image-level resolution으로 reconstruction한 뒤, anomaly scores를 계산



Paper 2

- Experiments

- SIA와 anomaly synthesis method를 사용하여 realnet 학습

- 학습에 사용한 anomaly synthesis method

- ☼ DTD

- ✓ DTD 데이터셋을 사용하여 생성된 anomaly texture와 이미지를 혼합하며, 혼합 과정에서 이미지 편집을 수행

- ☼ NSA

- ✓ 푸아송 이미지 편집을 사용

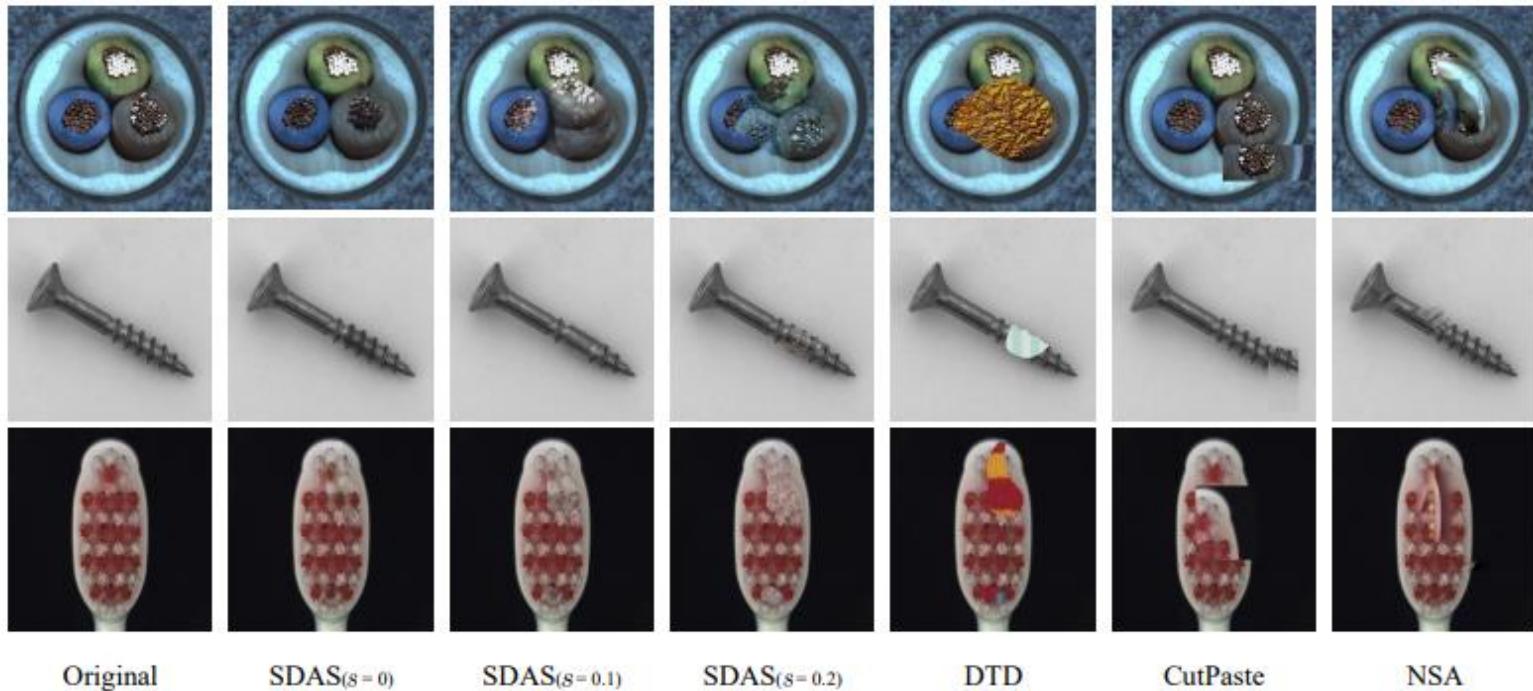
- ☼ Cutpaste

- ✓ random하게 이미지 자르기, 붙여넣기

Category	SIA	DTD [5]	NSA [32]	CutPaste [20]	
Texture	Carpet	(99.84, 99.19, 96.41)	(100.0, 99.27, 96.96)	(99.80, 98.60, 88.77)	(99.24, 98.42, 93.85)
	Grid	(100.0, 99.51, 97.28)	(100.0, 99.57, 97.14)	(100.0, 99.32, 91.31)	(100.0, 99.18, 92.53)
	Leather	(100.0, 99.76, 96.22)	(100.0, 99.77, 96.41)	(100.0, 99.24, 96.85)	(100.0, 99.41, 92.12)
	Tile	(99.96, 99.44, 97.70)	(100.0, 99.35, 95.27)	(100.0, 97.40, 86.45)	(99.86, 97.63, 84.39)
	Wood	(99.21, 98.22, 90.54)	(99.65, 98.28, 91.23)	(97.63, 93.30, 87.20)	(98.95, 95.29, 81.47)
	AVG	(99.80, 99.22, 95.63)	(99.93, 99.25, 95.40)	(99.49, 97.57, 90.11)	(99.61, 97.99, 88.87)
Object	Bottle	(100.0, 99.30, 95.62)	(100.0, 99.35, 95.57)	(100.0, 99.37, 93.49)	(100.0, 99.14, 91.41)
	Cable	(99.19, 98.10, 93.38)	(98.95, 97.84, 90.36)	(99.33, 97.62, 93.26)	(96.35, 96.23, 86.05)
	Capsule	(99.56, 99.32, 84.48)	(99.32, 99.19, 82.28)	(99.04, 99.27, 85.77)	(98.48, 99.10, 79.55)
	hazelnut	(100.0, 99.68, 93.14)	(100.0, 99.46, 93.46)	(100.0, 99.25, 94.41)	(100.0, 99.03, 91.51)
	Metal Nut	(99.76, 98.58, 94.39)	(99.90, 98.58, 96.49)	(100.0, 99.11, 93.27)	(99.90, 98.03, 89.69)
	Pill	(99.13, 99.02, 91.04)	(98.36, 98.88, 84.44)	(97.19, 98.28, 95.15)	(97.22, 98.96, 86.48)
	Screw	(98.83, 99.45, 87.90)	(97.72, 99.36, 85.22)	(98.79, 99.62, 93.74)	(92.74, 98.53, 79.63)
	Toothbrush	(99.44, 98.71, 91.57)	(99.44, 98.69, 90.87)	(100.0, 99.18, 89.20)	(99.17, 98.85, 78.48)
	Transistor	(100.0, 98.00, 92.92)	(99.71, 97.15, 86.56)	(98.54, 95.67, 79.09)	(99.38, 96.32, 76.52)
	zipper	(99.82, 99.17, 93.43)	(99.68, 99.02, 88.77)	(99.90, 98.91, 93.05)	(99.61, 98.03, 92.26)
	AVG	(99.57, 98.93, 91.79)	(99.31, 98.75, 89.40)	(99.28, 98.63, 91.04)	(98.29, 98.22, 85.16)
	AVG	(99.65, 99.03, 93.07)	(99.52, 98.92, 91.40)	(99.35, 98.28, 90.73)	(98.73, 98.14, 86.40)

Paper 2

- Experiments
 - Comparison anomaly synthesis method



Paper 2

- Experiments

- Comparison related models

- 동일한 과정을 다른 model에서 train한 결과

- ※ Realnet이 가장 효과적이었음

- SIA를 사용했을때, 다른 augmentation method보다 효과적인 것을 증명한 실험

Methods	Image AUROC	Pixel AUROC
DRAEM [48]	88.1	87.2
PaDiM [7]	84.2	89.5
UniAD [44]	96.5	96.8
OmniAL [54]	97.2	98.3
RealNet	97.3	98.4

Methods	FID [15] ↓	LPIPS [51] ↑
DTD [5]	120.52 ± 0.63	0.16 ± 0.00
CutPaste [20]	77.34 ± 0.09	0.11 ± 0.00
NSA [32]	68.76 ± 0.16	0.09 ± 0.01
SIA	60.39 ± 1.26	0.18 ± 0.01

Thank You