

2026 겨울 세미나

2026.01.09



Sogang University

Vision & Display Systems Lab, Dept. of Electronic Engineering



Presented By

이준호

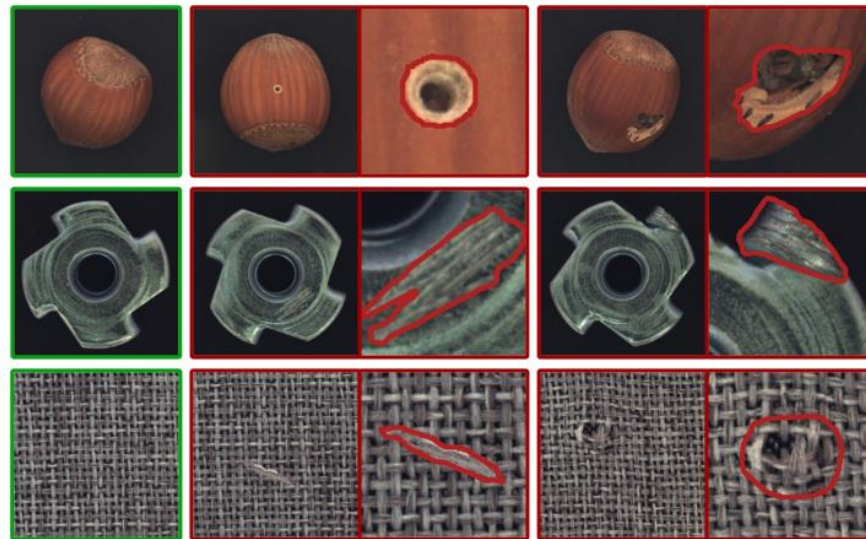
Outline

- Background
 - Anomaly detection
 - Video Anomaly Detection
- Paper 1
 - INFP: INdustrial Video Anomaly Detection via Frequency Prioritization (IJCAI 2025)
- Paper 2
 - Towards High-Resolution 3D Anomaly Detection: A Scalable Dataset and Real-Time Framework for Subtle Industrial Defects (AAAI 2026 oral)

Background

- Anomaly detection

- Normal(정상) sample과 abnormal(비정상, 이상치, 특이치) sample을 구별하는 task
- Anomaly detection은 제조업뿐만 아니라 CCTV, 의료 영상 등 다양한 분야에서 응용되면서 중요한 분야로 인식



< 산업용 anomaly dataset의 예시 >

Background

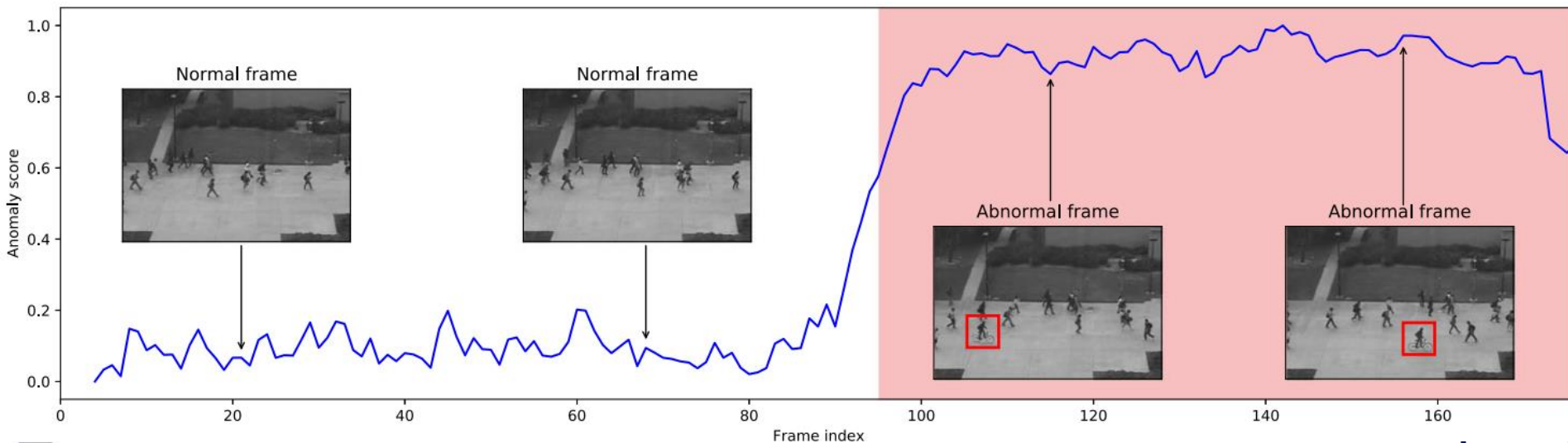
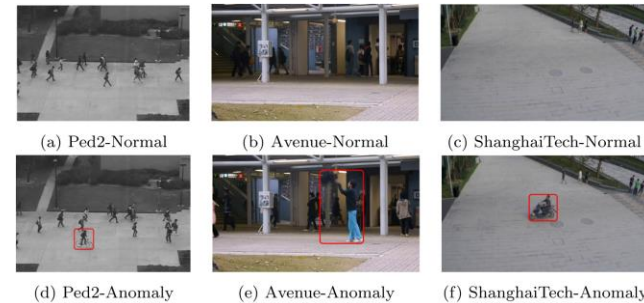
- Video Anomaly Detection

- Video의 각 frame에 대한 정상, 이상을 판별하는 task

- Input으로 sequential한 5개의 frame이 input으로 들어가며
첫 4개의 frame을 모델의 input으로 받아, 5번째 frame을 예측함
 - 예측한 frame과 gt frame간의 difference error를 최소화하는 방향으로 학습함

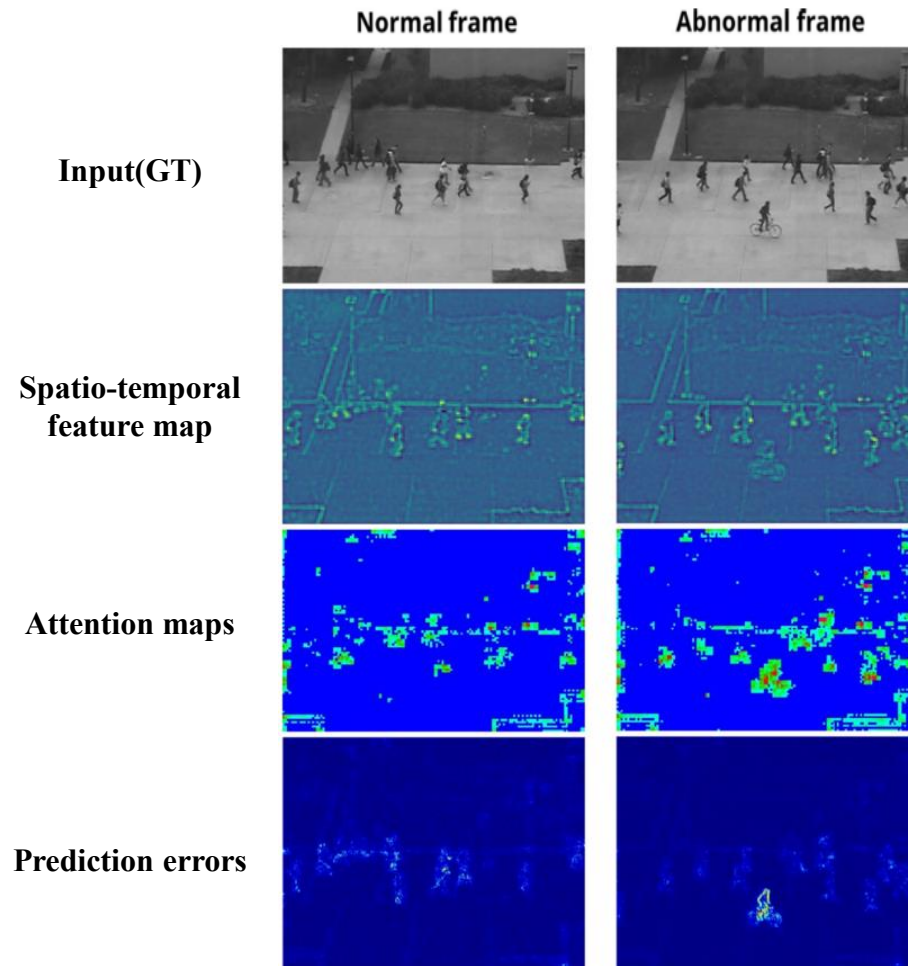
- 데이터셋은 CCTV 영상으로 구성되어 있으며, 각 영상 및 frame 단위의 gt label 공개되어 있음

- GT mask 또한 공개되어 있음



Background

- Video Anomaly Detection



INFP: INdustrial Video Anomaly Detection via Frequency Prioritization (IJCAI 2025)

Paper 1

- Contribution

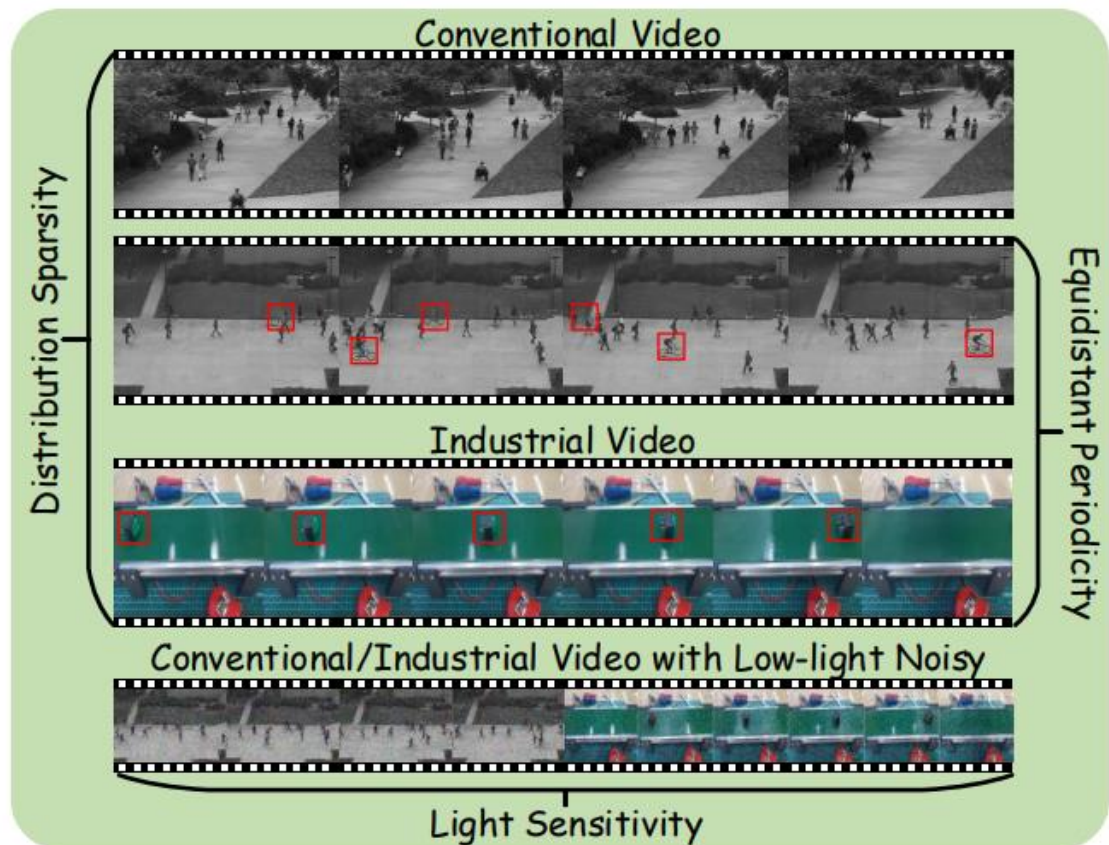
- Industrial Video Anomaly Detection (IVAD) task를 처음으로 제안함
 - 또한 IVAD 데이터셋에 optimize한 solution을 제공
- IVAD 환경에서 발생하는 다양한 noise에 대한 solution 제안
- Frequency-domain feature를 학습하는 framework 제안

Paper 1

- IVAD task란
 - Industrial setting의 생산 라인에서 수집되는 video 데이터를 다룸
- VAD vs IVAD
 - 기존 VAD는 여러 개의 정상 샘플이 동시에 등장하는 상황에서 이상을 식별하는 것을 목표로 하기 때문에, 영상 전체에 대한 정보를 다루는 경향이 있음
 - 반면 single object에 초점을 둔 IVAD는 fine-grained information과 특정한 pattern을 포착하는 것이 중요하기 때문에, local 정보를 처리하는 것이 중요함

Paper 1

- VAD to IVAD
 - 문제점 1 : Distribution Sparsity
 - 문제점 2 : Equidistant Periodicity
 - 문제점 3 : Light Sensitivity



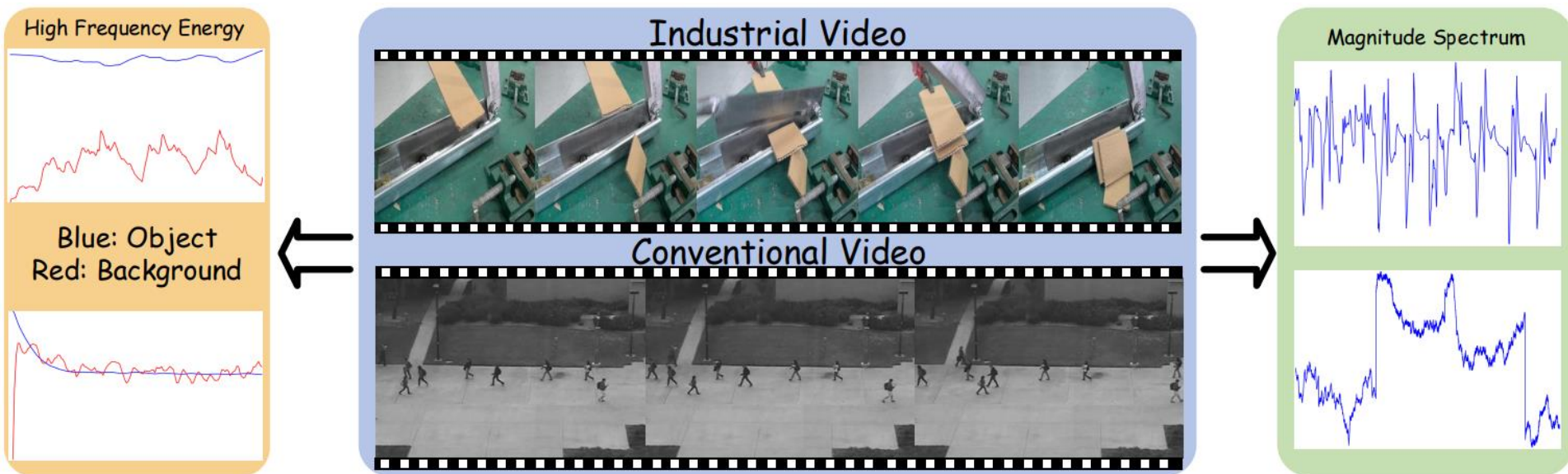
Paper 1

- IVAD solution

- 문제점 1 : Distribution Sparsity

- Industrial 환경에서 object와 background를 효과적으로 분리하여, detector가 object 부분을 더욱 효과적으로 집중할 수 있도록 함
 - Object와 background를 분리하기 위해, high-frequency energy 차이를 이용함

∴ Industrial video >>> Conventional video

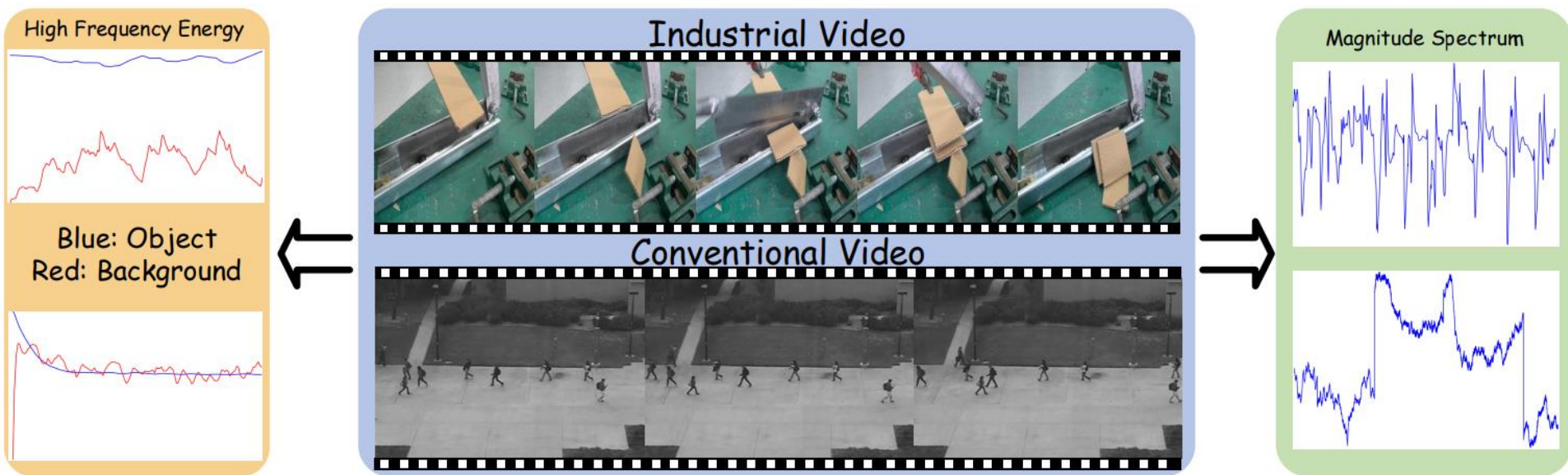


Paper 1

- IVAD solution

- 문제점 1 : Distribution Sparsity

- High-frequency energy가 높은 영역에 더 큰 weight를 부여하는 trajectory filter 제안
 - Train/Inference 단계에서 이동하는 object에 더 높은 할당함으로써, frequency-domain weight를 통해 IVAD setting에 optimize한 solution 제공



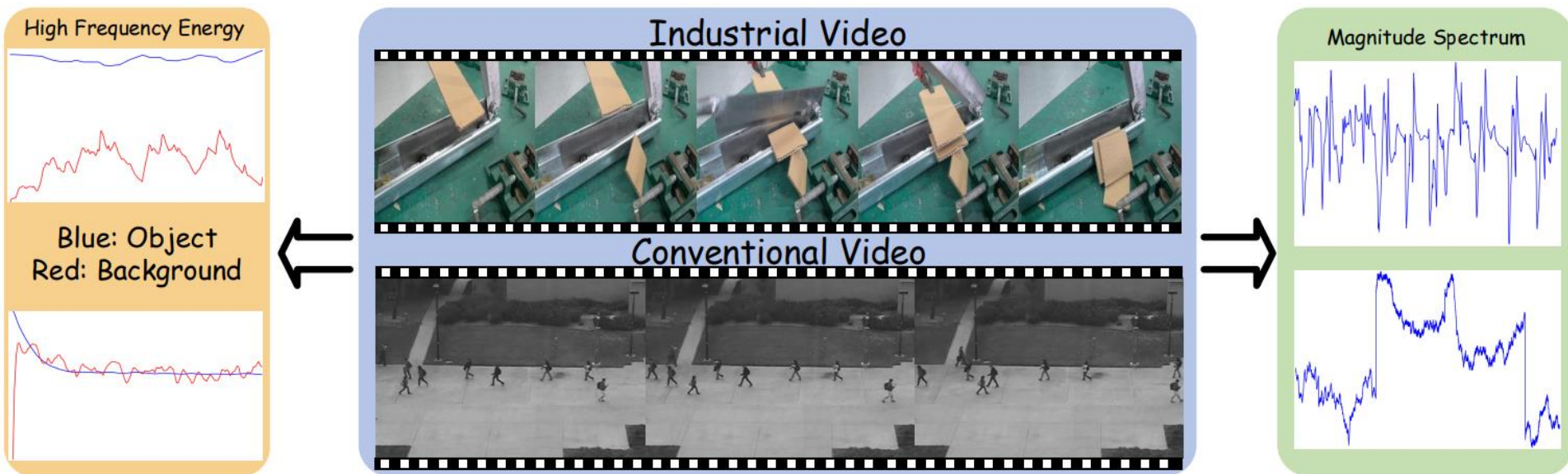
Paper 1

- IVAD solution

- 문제점 2 : Equidistant Periodicity & 문제점 3 : Light Sensitivity

- 뚜렷한 equidistant periodicity를 갖는 frequency-domain signal을 추출하기 위한 feature extractor를 설계함
 - 이후 frequency-domain feature와 spatio-temporal feature를 결합한 fusion feature는 마찬가지로, 명확한 equidistant periodicity를 유지하게 됨

※ 이는 decoder의 input에 구조적이고 강건한 input을 제공함



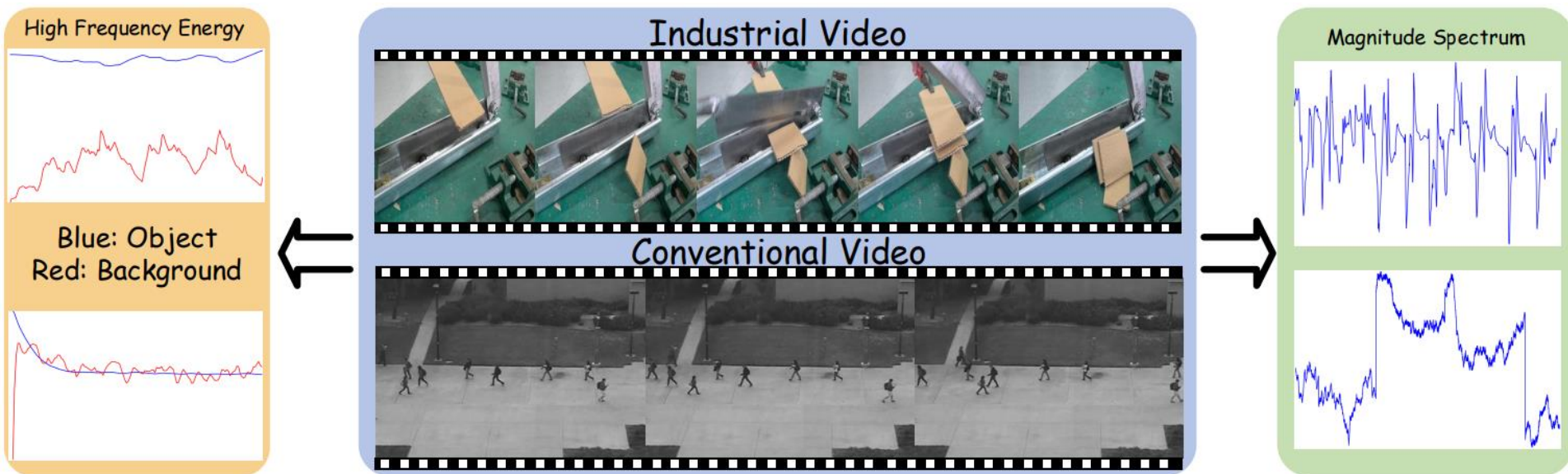
Paper 1

- IVAD solution

- 문제점 2 : Equidistant Periodicity & 문제점 3 : Light Sensitivity

- Frequency-domain feature & spatio-temporal feature를 융합한 fusion feature를 제공하면, IVAD에서 주로 발생하는 speed anomalies, product absence와 같은 이상 상황을 탐지하는데 도움을 줌

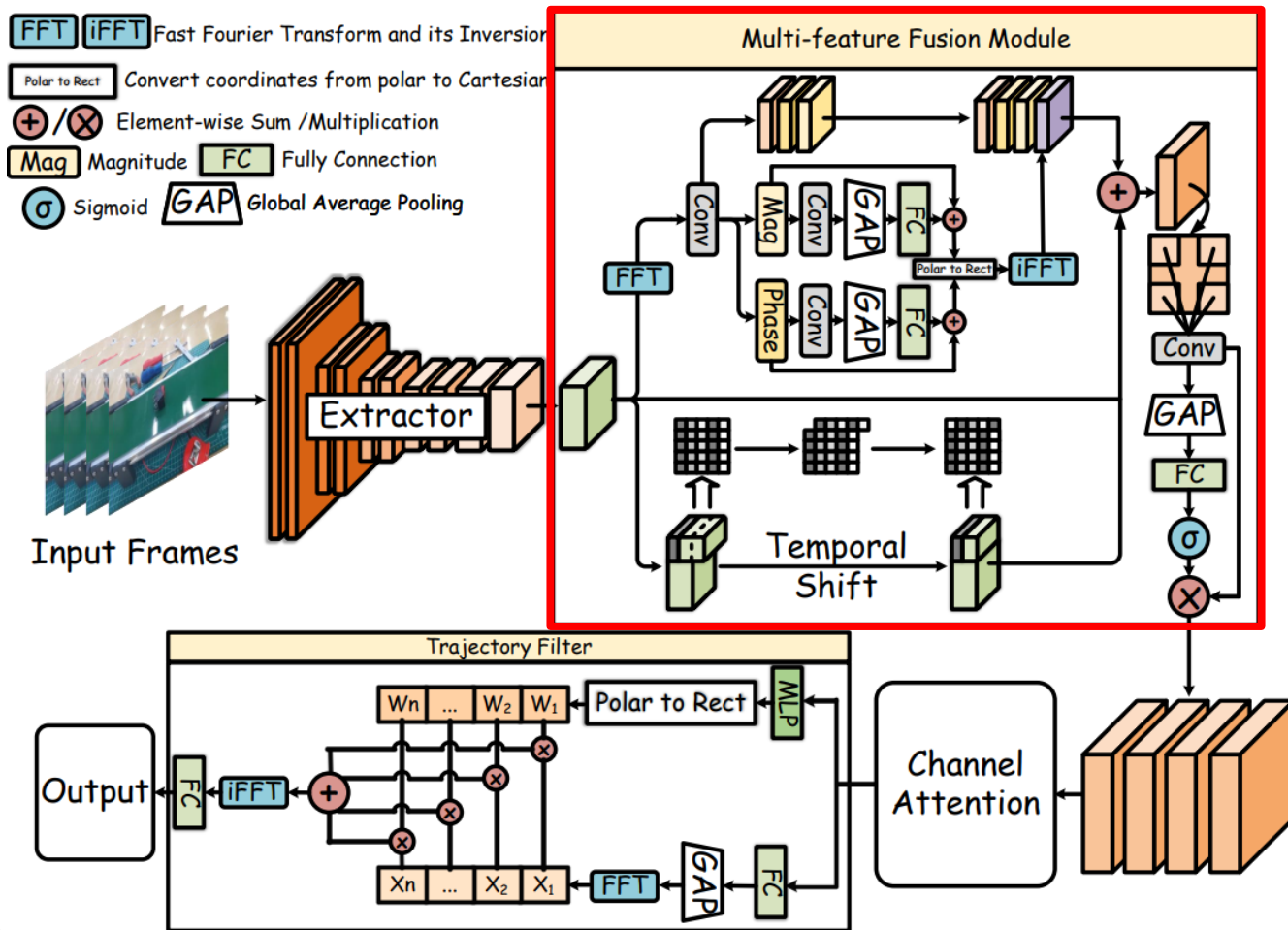
※ 조명(그림자) 변화로 인해 발생하는 영상 내의 큰 pixel variations를 효과적으로 완화



Paper 1

- Method

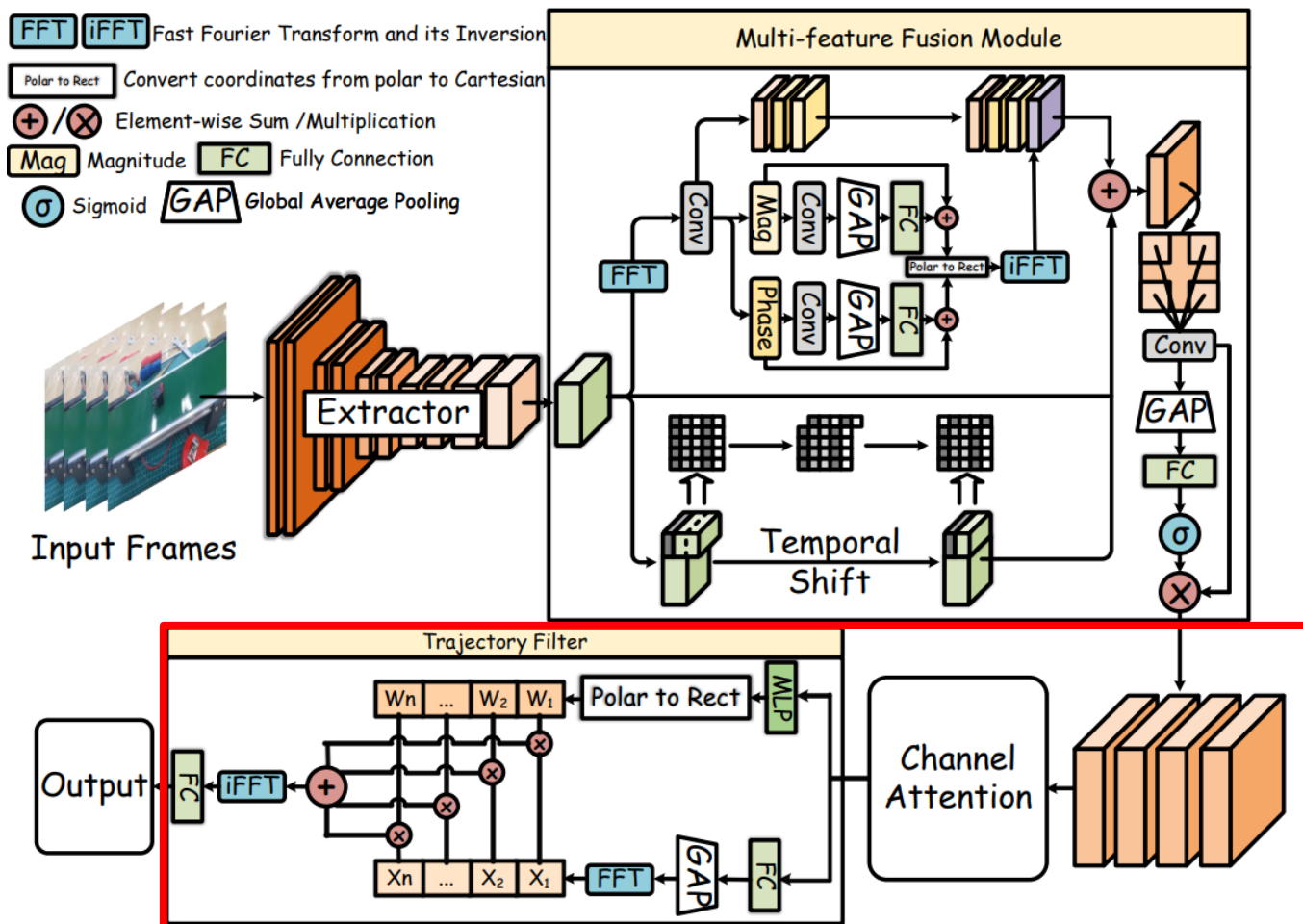
- Overview



Paper 1

- Method

- Overview



Paper 1

- Method

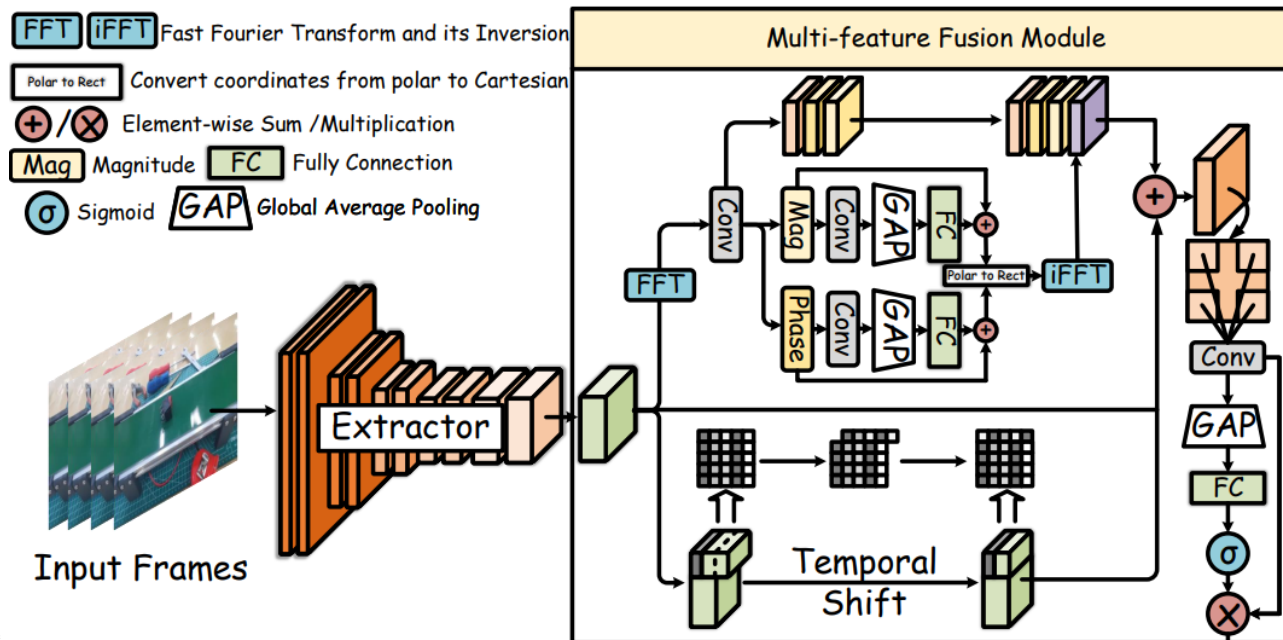
- Multi-feature Fusion Module (MFM)

- Extractor에서 추출한 feature를 3가지 branch로 나눠 처리함

- ⚙ Spatial feature

- ⚙ Temporal feature

- ⚙ Frequency feature



Paper 1

- Multi-feature Fusion Module (MFM)

- Periodic information과 lighting variation을 해결하기 위해, MFM 제안
- Extractor에서 추출한 feature에 대하여 Fourier transform(FFT)을 적용

$$z = \int_{-\infty}^{+\infty} f_e(I_v) e^{-j\omega t} dt$$

- FFT 결과는 복소수이기 때문에 아래와 같이 표현이 가능하며 $a^{\wedge} / b^{\wedge}$ 은 실수 / 허수부를 의미함
- 2가지로 분해하며, 각각 $\text{magnitude}(|z|)$ 와 $\text{phase}(\text{angle}=\arctan(b^{\wedge}/a^{\wedge}))$ 로 분해

$$z = \hat{a} + \hat{b}j$$

$$z = |z| e^{j \cdot \text{angle}(z)}$$

- Magnitude : 이 frequency가 얼마나 강한가

- Phase : 언제 반복되는가

$$a = \hat{a} + \delta(\overline{\omega}(f(\hat{a})))$$

$$b = \hat{b} + \delta(\overline{\omega}(f(\hat{b})))$$

Paper 1

- Multi-feature Fusion Module (MFM)

- 실수부와 허수부를 CNN, LeakyReLU, GAP, FC를 이용하여 보정

$$a = \hat{a} + \delta(\overline{\omega}(f(\hat{a})))$$

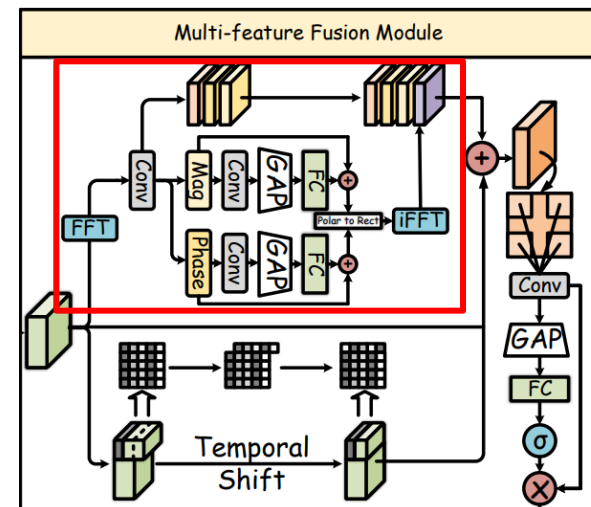
$$b = \hat{b} + \delta(\overline{\omega}(f(\hat{b})))$$

- 보정한 frequency를 iFFT를 이용하여 다시 spatio-temporal feature로 복원

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} z e^{j\omega t} d\omega$$

- Frequency-domain feature z 와 temporal-domain으로 복원한 feature $f(t)$ 를 결합

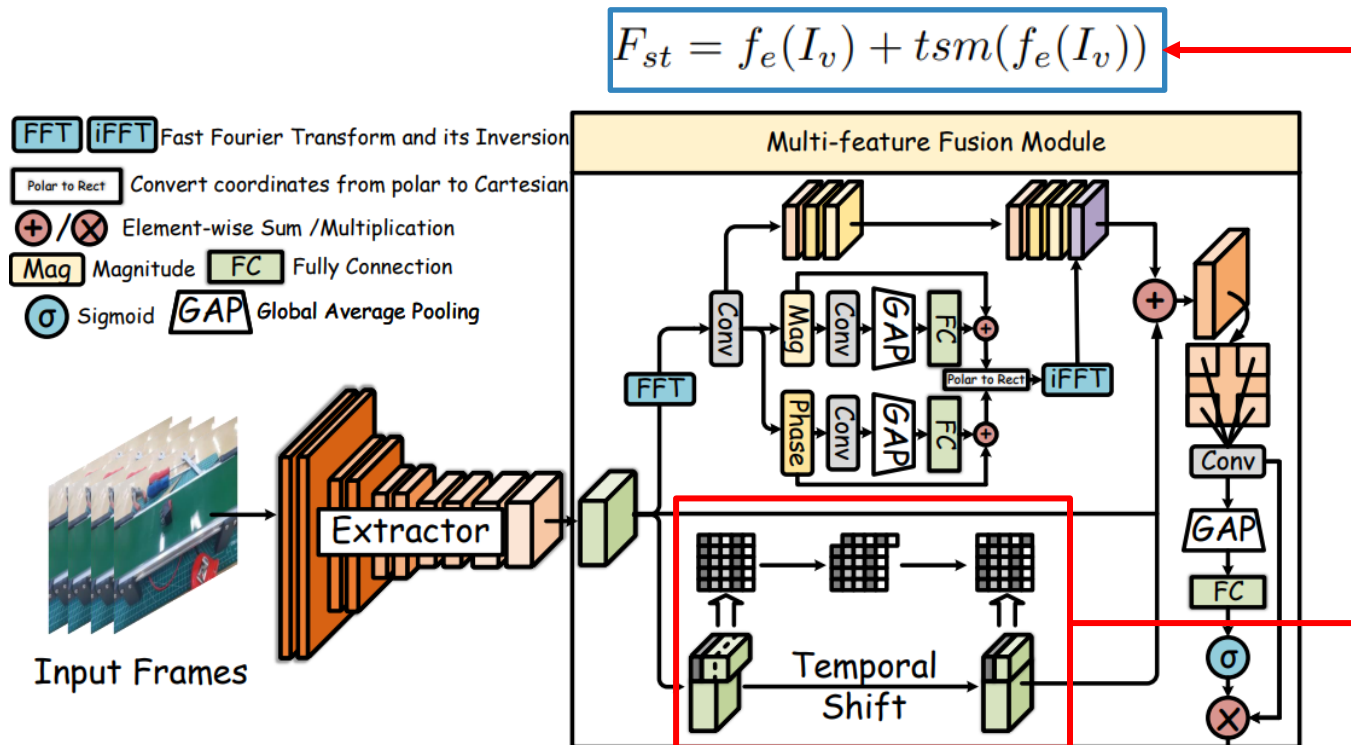
$$F_{freq} = [\gamma(z); f(t)]$$



Paper 1

- Multi-feature Fusion Module (MFM)

- Temporal Shift Module(TSM)을 사용하여 temporal 정보를 효과적으로 포착함



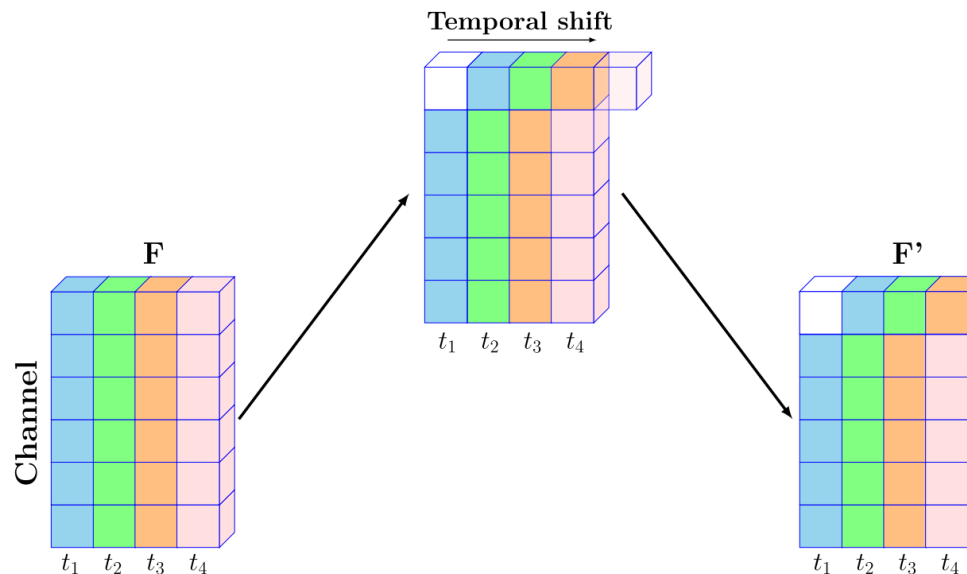
Paper 1

- Multi-feature Fusion Module (MFM)

- Temporal Shift Module(TSM)을 사용하여 temporal 정보를 효과적으로 포착함

- Input 4개의 frame의 feature map에 대하여 각 channel의 일부를 다음 frame으로 이동시킴
 - Channel을 임의의 fold로 나누어, 전체 channel 중 일부 fold를 time 축을 따라 인접한 frame에 이동을 수행함

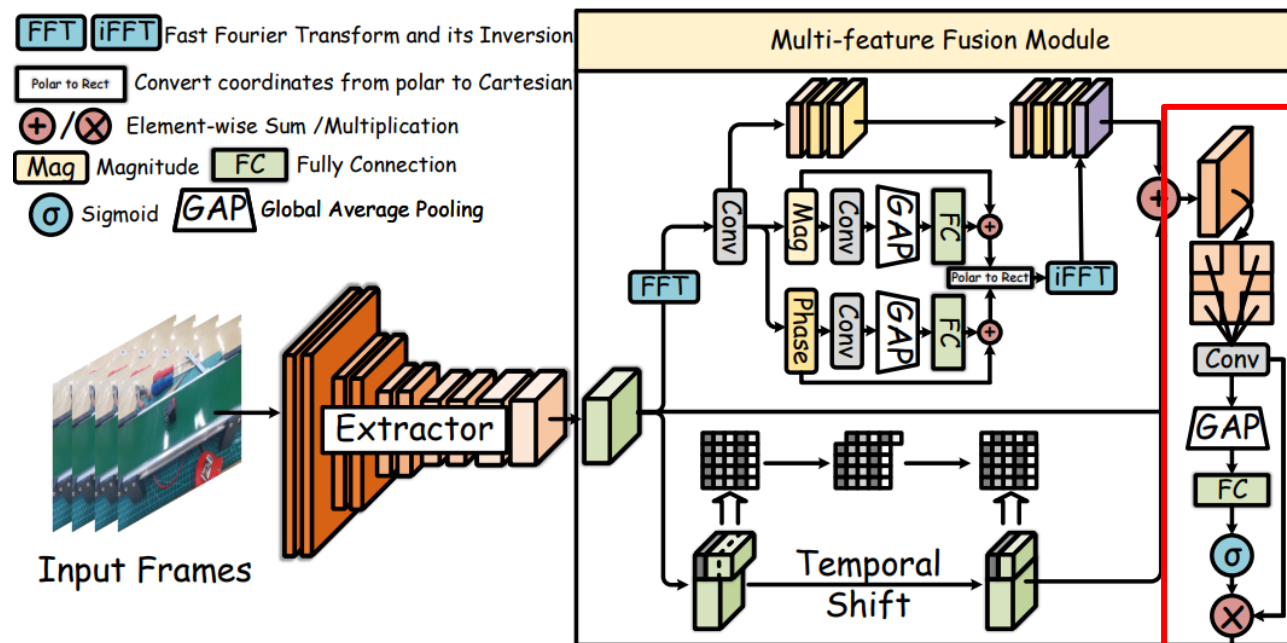
※ 추가적인 parameter나 computational cost 없이 이전 frame 정보를 다음 frame에 부여할 수 있음



Paper 1

- Multi-feature Fusion Module (MFM)

- Frequency & Temporal-domain 정보를 각각 처리한 뒤 summation으로 feature fusion
- Fusion한 feature에 대하여 조명 변화에 강건하고자, 추가적으로 처리하여 feature간의 context 정보를 결합함



Paper 1

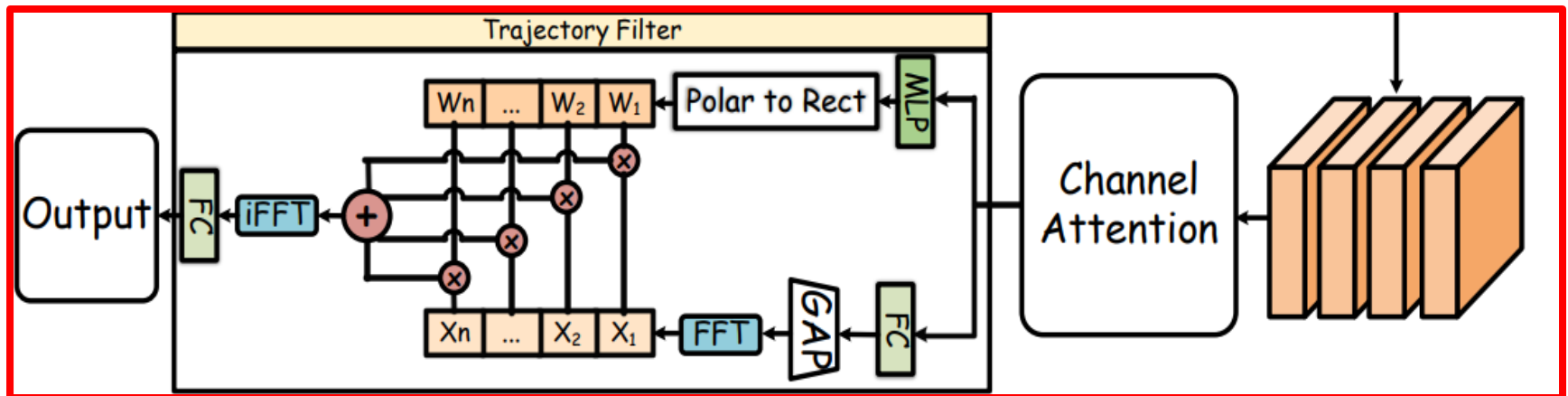
- Multi-feature Fusion Module (MFM)

- $F_{mul} \in R^{h \times w \times c}$ 내에 사전에 정의된 영역 $A \in R^{a \times a \times c}$ 를 설정하고, 영역의 4 모서리에 위치한 하위 영역 정의
 - 해당 하위 영역에 대해서만 conv 연산을 수행하여, 다른 영역의 정보는 무시하고 사전에 정의된 영역에 집중하도록 유도함
 - 각 영역의 conv 연산 값들은 영역 A의 값으로 사용됨
- 각 영역의 값들을 combined하는 과정에서 bias된 activation map이 생성될 수 있으므로, 이를 보정하기 위해 GAP, FC, ReLU, Sigmoid를 사용하여 보정함
- 최종 보정된 map은 decoder로 전달됨

Paper 1

- Trajectory Filter

- 고정된 trajectory를 따라 움직이는 object에 집중하고, 주변 background 영향을 줄이기 위해 제안
- 총 4개의 frame에 대한 MFM output은 channel attention (average pooling + 2개의 conv layer + ReLU + Sigmoid 로 구성)을 거친 feature map(F_d)이 input으로 전달됨



Paper 1

- Trajectory Filter

- Frequency feature를 추출하기 위해 아래와 같이, FFT 수행

- frequency feature를 GAP와 MLP로 processing한 뒤, trajectory weight를 계산함

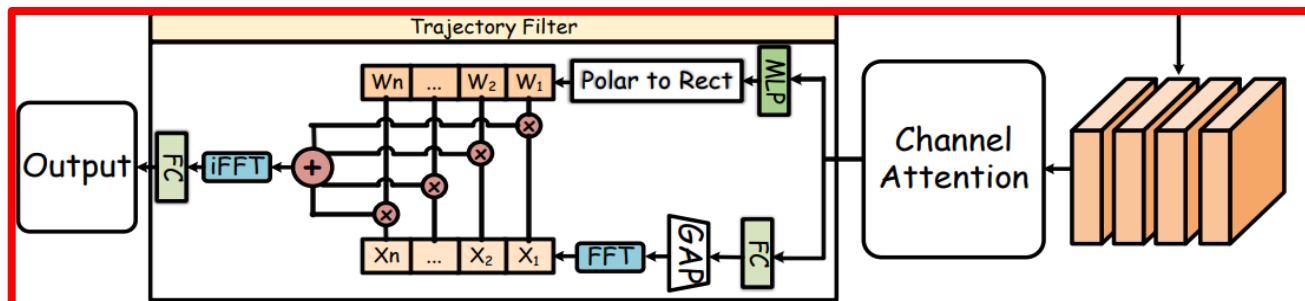
$$X = \{x_1, x_2, \dots, x_n\} = \int_{-\infty}^{+\infty} \delta(Z_3(F_d)) e^{-j\omega t} dt$$

$$R = \text{softmax}(Z_4(\text{GAP}(F_d)))$$

- 이후 임의의 복소수 weight $W = W_{real} + j * W_{imag}$ 를 생성하고, 이를 trajectory weight와 결합함

- Element-wise summation으로 결합

- 이후 iFFT를 통해 final output을 산출함



Paper 1

- Experiments

- Dataset : IPAD(arxiv 2024)

- Industrial setting에 focus한 유일한 industrial video anomaly detection 데이터셋
 - 4개의 real video와 12개의 simulation video로 구성됨

- 평가 지표

- AUROC

- ※ 모든 video frame을 연결한 뒤, AUROC 값을 계산

- 실험 환경

- Image size : 224 x 288
 - Train

- ※ 연속된 5개의 frame을 무작위로 선택한 뒤, 첫 4 frame은 입력 5번째 frame은 gt

Paper 1

- Experiments

- Main table

Method	S01	S02	S03	S04	S05	S06	S07	S08	S09	S10	S11	S12	R01	R02	R03	R04	Avg.
conAE[Hasan <i>et al.</i> , 2016]	63.1	47.7	53.0	34.7	82.9	46.6	58.3	69.1	53.0	55.3	39.8	50.4	77.6	64.3	40.7	70.1	56.7
memAE[Gong <i>et al.</i> , 2019]	63.2	50.6	65.6	49.5	78.8	45.9	57.9	84.7	65.7	59.9	49.4	50.7	77.9	65.0	41.6	70.7	61.1
AstNet[Le and Kim, 2023]	67.7	52.0	61.0	51.6	80.4	54.1	54.5	82.6	59.8	55.7	47.8	60.8	79.8	66.8	42.1	67.6	61.5
DMAD[Liu <i>et al.</i> , 2023]	55.9	55.3	47.9	47.9	69.3	61.0	<u>66.9</u>	87.5	<u>69.7</u>	<u>67.0</u>	56.0	55.8	79.5	68.5	43.1	63.1	62.2
V-Swin-T[Liu <i>et al.</i> , 2022]	68.2	60.0	<u>66.6</u>	54.7	85.6	53.3	59.5	<u>88.5</u>	<u>69.7</u>	60.5	54.8	69.1	81.1	74.1	42.3	<u>75.5</u>	66.5
IPAD_VAD[Liu <i>et al.</i> , 2024]	<u>69.5</u>	63.9	70.6	58.3	<u>86.2</u>	<u>61.2</u>	60.6	85.6	71.2	62.2	60.9	<u>67.1</u>	<u>84.4</u>	<u>75.4</u>	<u>43.5</u>	76.7	<u>68.6</u>
Ours	77.2	<u>62.8</u>	70.6	<u>56.0</u>	89.7	73.4	78.1	96.3	66.6	71.9	<u>56.9</u>	60.9	86.5	77.9	48.4	75.3	71.8

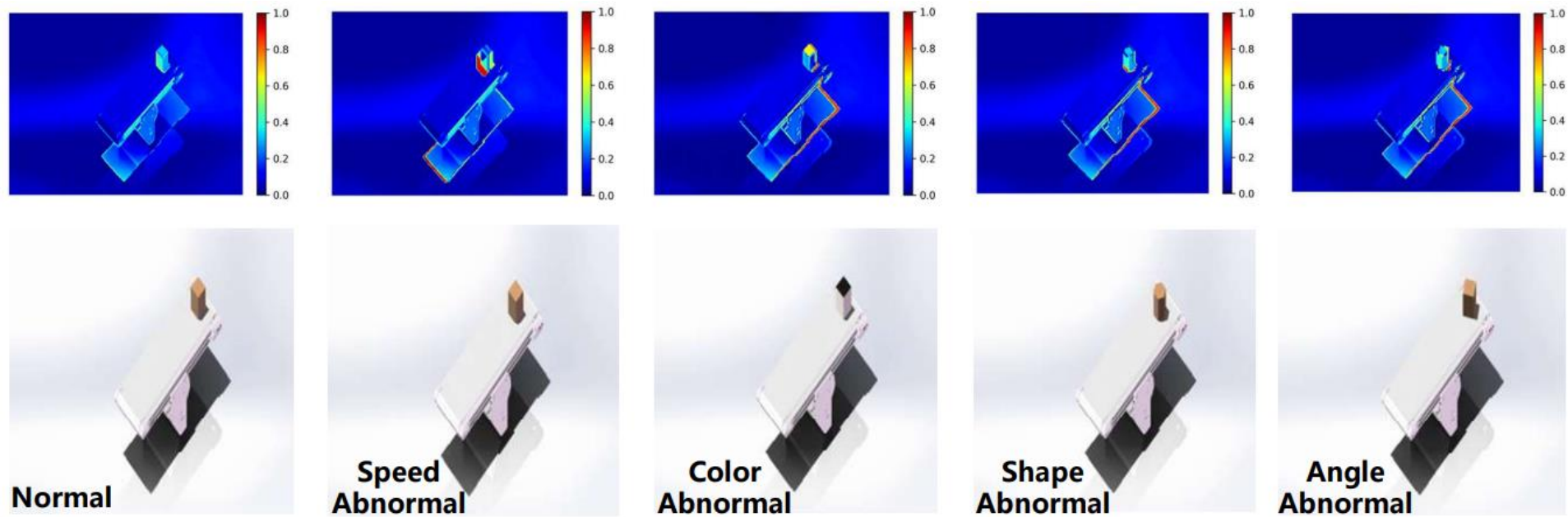
- 모듈 별 ablation

Category	Baseline	+Fusion	+Filter	+Fusion&Filter
Average	61.3	68.5	67.2	71.8

Paper 1

- Experiments

- 정성적 결과

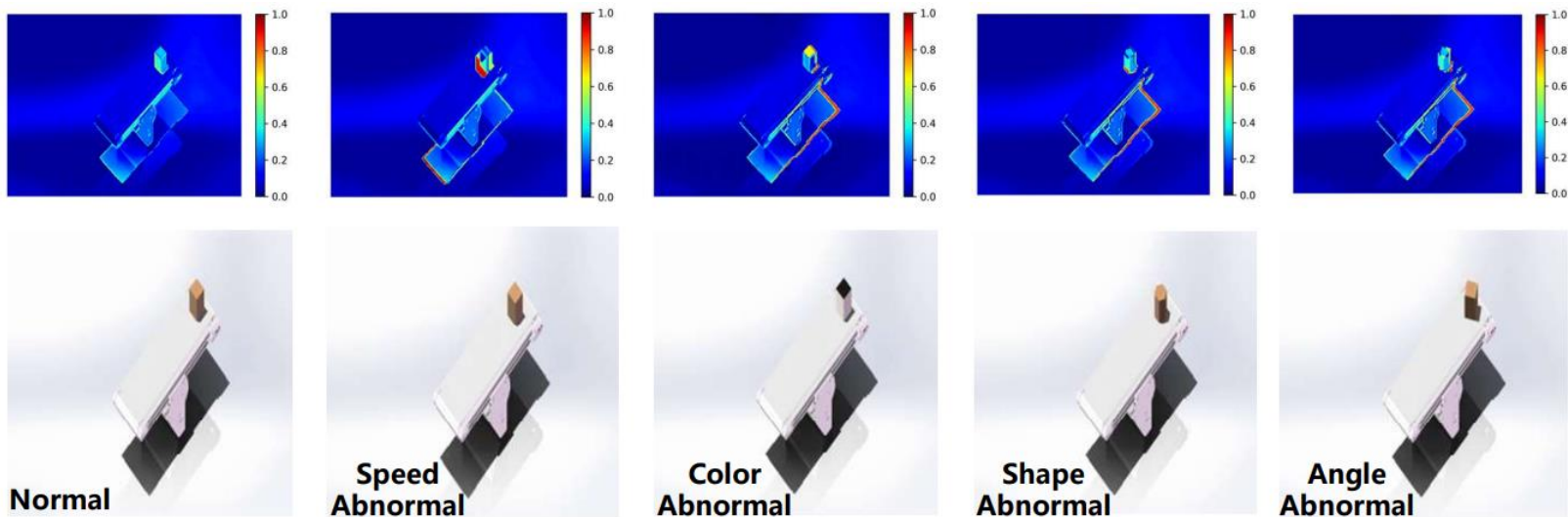


Paper 1

- Experiments

- 정상적 결과

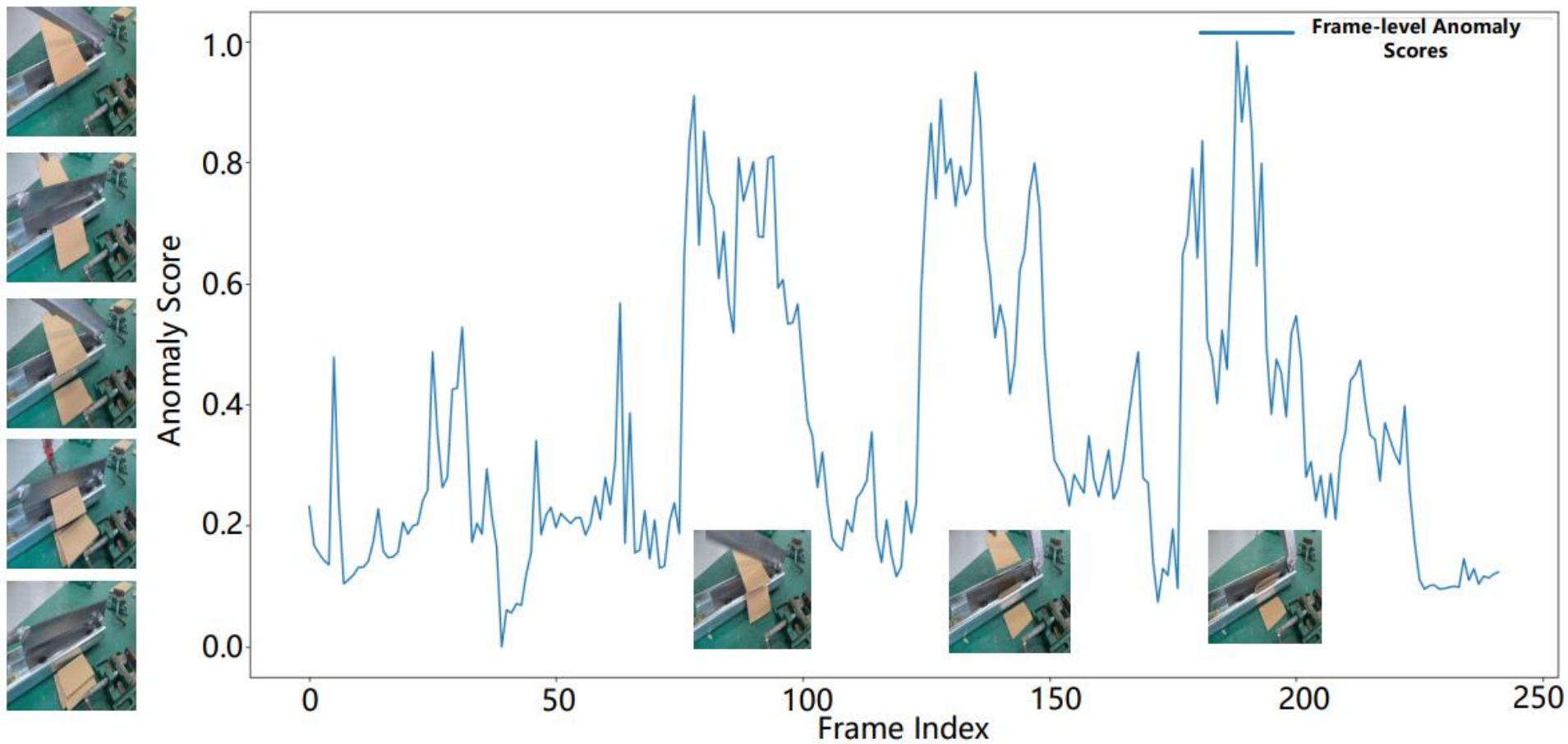
- Speed Abnormal : 움직임과 관련된 영역에서 큰 차이가 나타남
 - Color Abnormal : 색상 변화가 발생한 영역에서 집중적으로 차이가 나타남
 - Shape Abnormal : object의 boundary 주변에서 두드러진 차이가 나타남
 - Angle Abnormal : object의 회전에 의해 영향을 받는 영역에서 차이가 나타남



Paper 1

- Experiments

- 정성적 결과

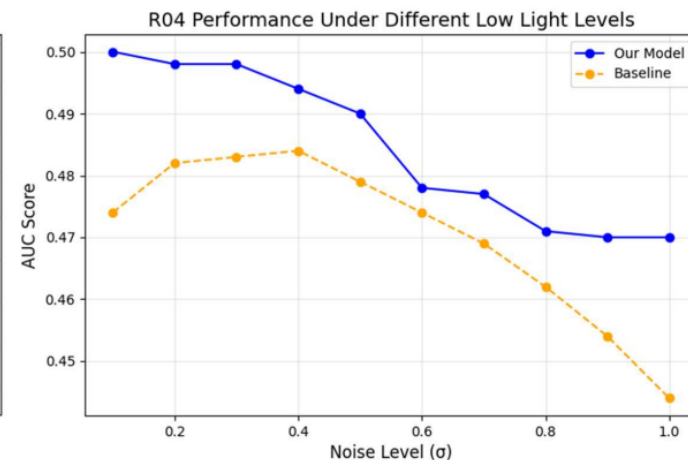
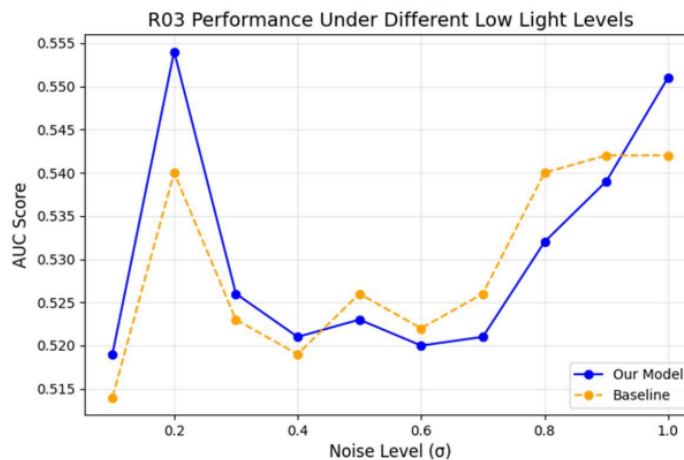
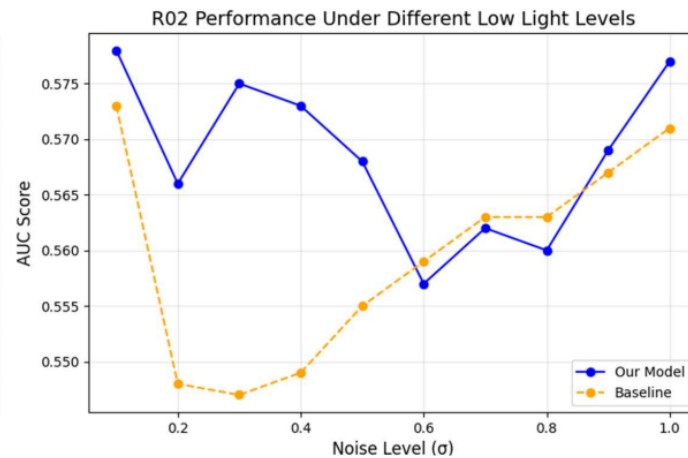
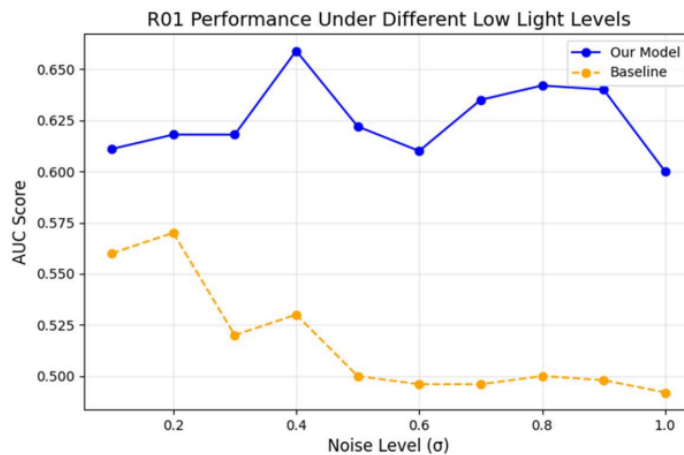


Paper 1

- Experiments

- 강건성 분석

- 조명 변화에 강건성을 평가하고자, low-light noise를 적용하여 성능 평가



Towards High-Resolution 3D Anomaly Detection: A Scalable Dataset and Real-Time Framework for Subtle Industrial Defects (AAAI 2026 Oral)

Paper 2

- Contribution

- High-resolution 3D anomaly detection dataset 제안
- Simple3D 제안
 - 최소한의 계산 비용
 - 초당 20frame을 초과하는 real-time inference 성능을 달성

Paper 2

- Motivation

- 기존 3D AD 방법론

- 원본 point cloud에서 적은 수의 point(보통 약 1000개)를 sampling
 - Local neighbor 정보를 encoding하여 point group을 구성
 - 각 group에 anomaly score를 할당

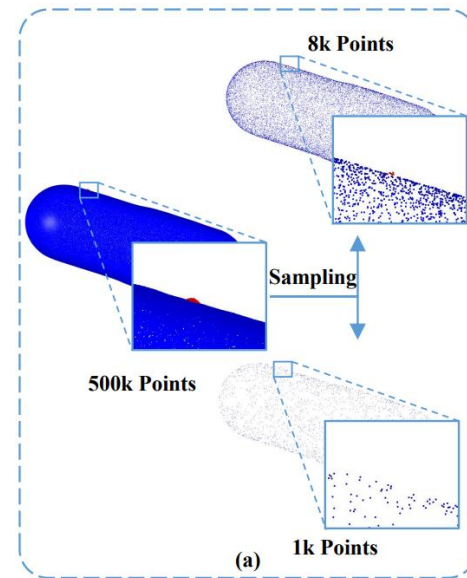
- Reconstruction-based 3D AD 방법론

- 각 group을 대표하는 token(prototype token)으로 변환
 - Transformer 구조를 통해 이를 reconstruction 수행
 - 원본 token과 reconstruction된 token 간의 차이를 계산하여 anomaly score 산출

- Sparse group-level score는 전체 point cloud로 interpolation되어 dense anomaly map를 생성함

- 하지만 이런 interpolation 과정은 미세한 결함을 탐지하는 데 필수적인 spatial granularity를 불가피하게 저하시킴

※ 20mm x 20mm 표면 위 1mm x 1mm 크기 결함은 point cloud를 축소하면 인식이 불가능해짐

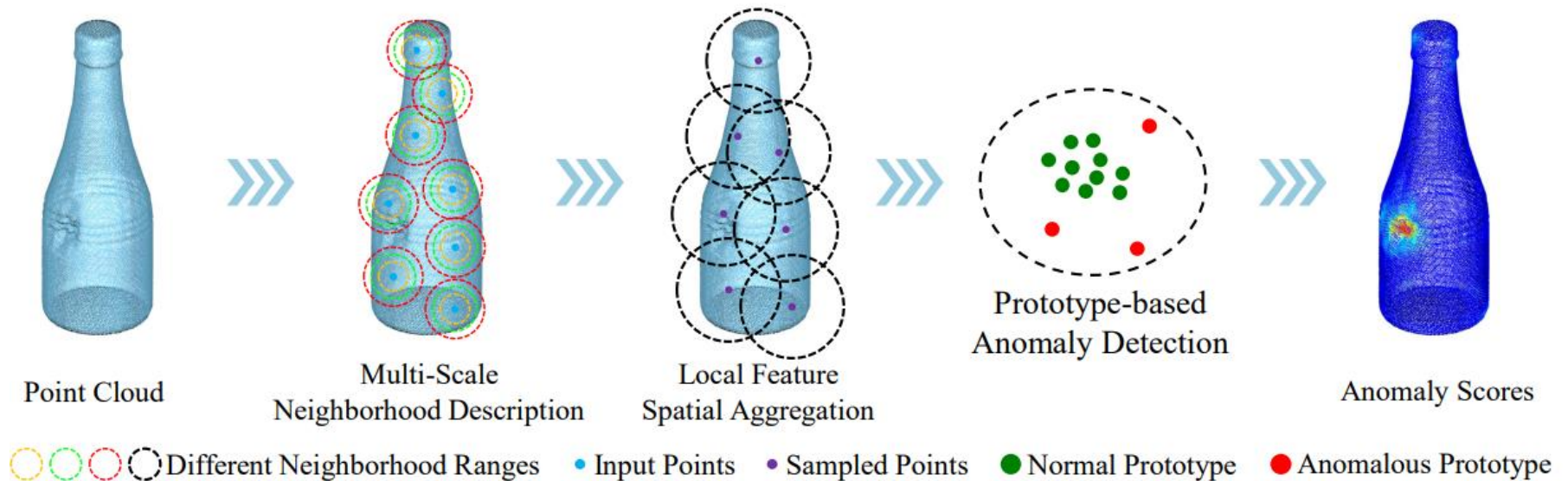


Paper 2

- Simple3D Method

- Overview

- CNN/ViT encoder를 활용하지 않고 handcrafted descriptors(ex. k-NN)를 사용하여 local point feature 추출
 - 추출한 feature들을 spatial hierarchies 구조에 따라 aggregate하여 anomaly scoring



Paper 2

- Simple3D Method

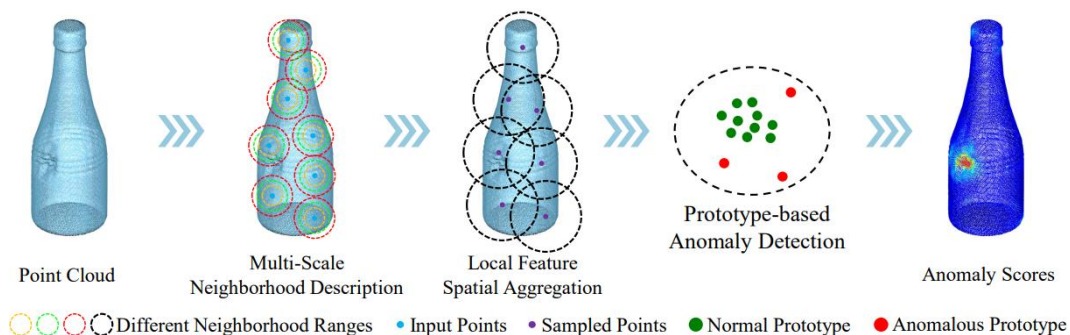
- Multi-Scale Neighborhood Description (MSND)

- Input point cloud($P \in \mathbb{R}^{n \times 3}$)에 대하여 각 point에 대해, k-NN search를 통해 point set(R_i)구성
 - Various neighborhood scale에서의 local point distribution이 가지는 complementary information를 활용하기 위해, 여러 개의 neighbor count를 설정

※ Multiple neighborhood counts를 생성

- 각 point set에 feature descriptors를 적용하여 feature operator 추출하고 concatenate하여 point의 multi-scale local feature representation을 구성

※ 이 과정을 통해 각 point는 다양한 기하학적 특성을 반영한 representation을 갖게 됨



Paper 2

- Simple3D Method

- Local Feature Spatial Aggregation (LFSA)

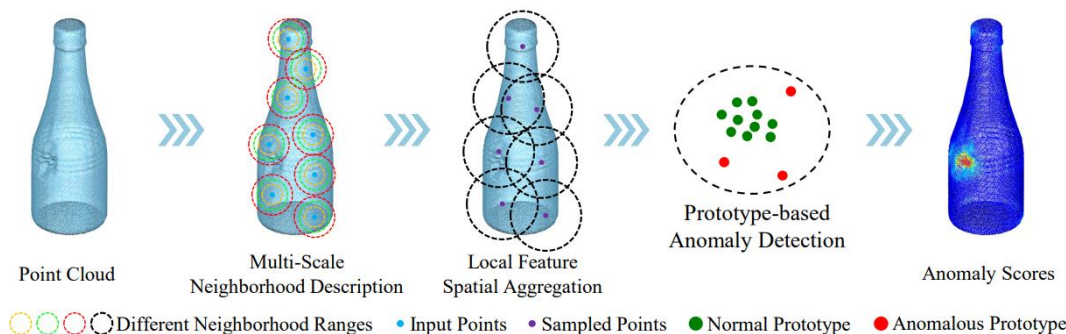
- Feature representation을 더욱 정교화하기 위해 제안
 - Point cloud에서 random하게 t 개의 point를 sampling하고 주변 neighbor의 MSND feature를 average aggregation 수행

- ※ 이 과정은 receptive field를 확장하고, 풍부한 fine-grained 정보를 갖는 feature 생성

- ※ 또한 공간적으로 안정적이며 표현력이 강화됨

- Anomaly Detection

- Normal point cloud로부터 얻어진 향상된 point feature들은 normal prototype set 구성
 - Test point cloud에서 추출된 feature에 대하여, normal prototype set과의 distance를 기반으로 anomaly scoring



Paper 2

- Simple3D Method

- Experiments

– Main table

Method → Real3D-AD	CPMF PR'24	Reg3D-AD NeurIPS'23	Group3AD ACM MM'24	IMRNet CVPR'24	ISMP AAAI'25	GLFM TASE'25	PO3AD CVPR'25	MC3D-AD IJCAI'25	Simple3D Ours
O-ROC/P-ROC	62.5/75.9	70.4/70.5	75.1/73.5	72.5/-	76.7/ <u>83.6</u>	75.0/76.7	76.5/-	<u>78.2</u> /76.8	80.4/92.3
Method → Anomaly-ShapeNet	M3DM-PM CVPR'23	CPMF PR'24	Reg3D-AD NeurIPS'23	IMRNet CVPR'24	ISMP AAAI'25	GLFM TASE'25	PO3AD CVPR'25	MC3D-AD IJCAI'25	Simple3D Ours
O-ROC/P-ROC	51.1/54.9	55.9/-	57.2/-	66.1/65.0	75.7/69.1	61.9/74.5	83.9/ <u>89.8</u>	<u>84.2</u> /74.8	86.0/92.9
Method → MulSen-AD	M3DM-PM CVPR'23	M3DM-PB CVPR'23	PatchCore-FP CVPR'22	PatchCore-FP-R CVPR'22	PatchCore-PM CVPR'22	IMRNet CVPR'24	Reg3D-AD NeurIPS'23	GLFM TASE'25	Simple3D Ours
O-ROC/P-ROC	62.8/58.7	70.5/61.1	<u>86.0</u> /64.0	83.3/62.0	84.0/60.5	60.1/46.7	74.9/64.1	78.5/66.5	88.2/80.3

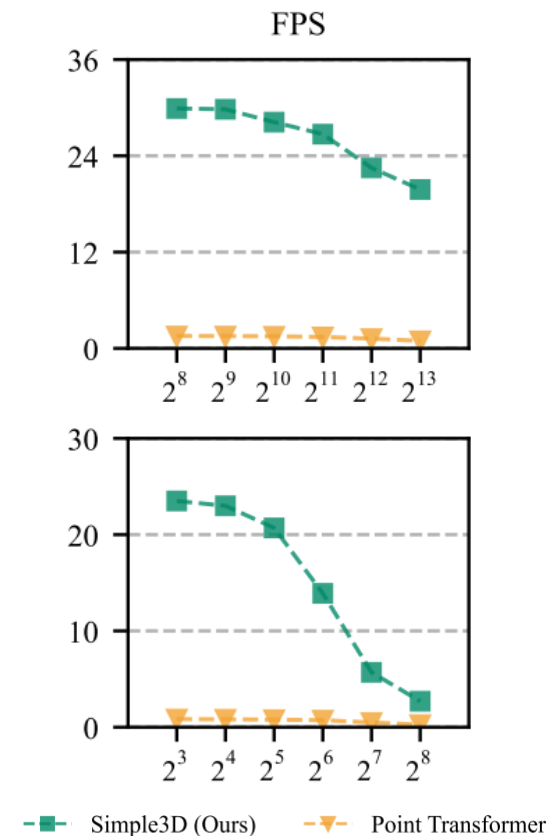
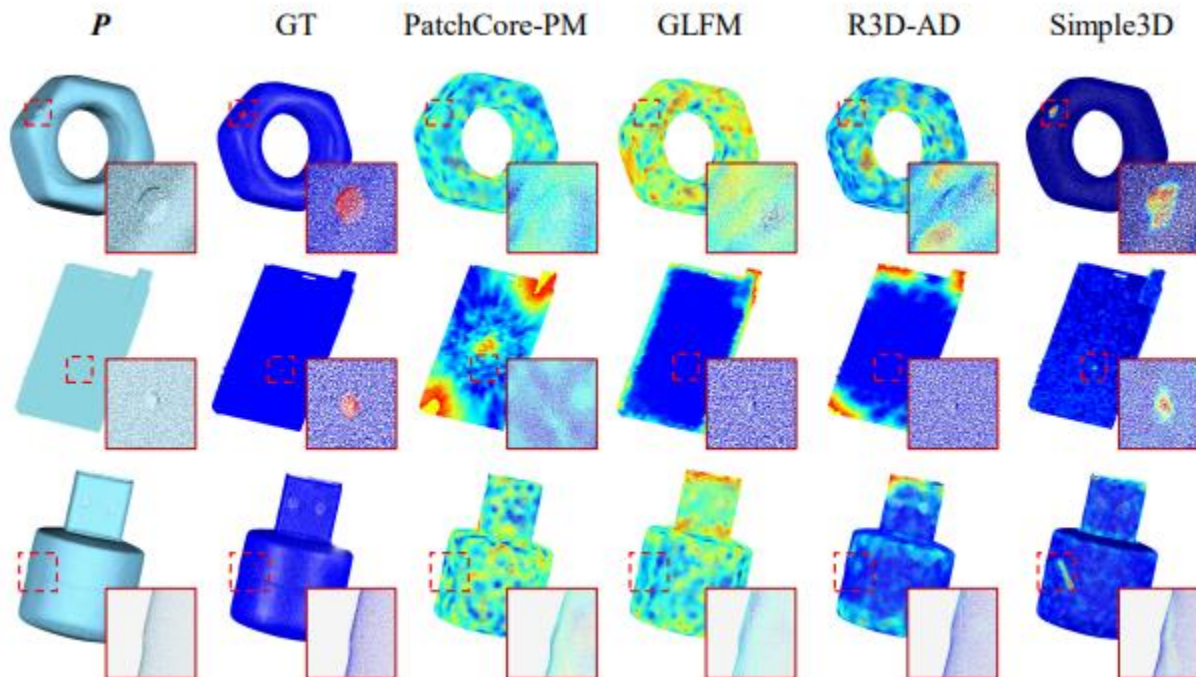
Method → Level ↓	PatchCore-FP CVPR'22	PatchCore-PM CVPR'22	PatchCore-PB CVPR'22	R3D-AD ECCV'24	GLFM TASE'25	Simple3D Ours
Easy	<u>68.3</u> /56.5	57.2/57.3	55.5/51.6	56.3/48.3	57.7/ <u>67.7</u>	75.6/77.3
Medium	<u>65.6</u> /54.1	55.8/52.6	46.6/50.3	50.8/50.7	55.0/ <u>57.3</u>	68.6/65.5
Hard	<u>61.4</u> /51.1	54.3/52.0	52.9/50.8	50.1/50.6	52.8/ <u>52.7</u>	61.6/56.3
ALL	<u>65.1</u> /53.7	55.9/54.1	50.0/51.4	53.6/49.8	55.8/ <u>58.7</u>	68.6/66.2

Paper 2

- Simple3D Method

- Experiments

- 정성적 결과



Thank You