

---

# A Survey on Multimodal Large Language Models for Anomaly Detection

---

김규형

# CONTENTS

01

VMAD : Visual-enhanced Multimodal Large Language Model for Zero-Shot Anomaly Detection

02

Towards Zero-Shot Anomaly Detection and Reasoning with Multimodal Large Language Models

- Zero-shot anomaly detection(ZSAD)의 한계를 극복하기 위해 새로운 framework인 VMAD 제안
- Anomalous structure discriminability를 향상시키기 위해 DSSL(Defect-Sensitive Structure Learning) 방식 제안
- Anomaly detection에 맞춘 새로운 visual projector인 LTC(Locality-enhanced Token Compression) 제안
- 새로운 dataset인 RIAD(Real Industrial Anomaly Detection) 소개

## VMAD: Visual-enhanced Multimodal Large Language Model for Zero-Shot Anomaly Detection

Huilin Deng, Hongchen Luo, Wei Zhai, Yang Cao, *Member, IEEE*, Yu Kang, *Senior Member, IEEE*

**Abstract**—Zero-shot anomaly detection (ZSAD) recognizes and localizes anomalies in previously unseen objects by establishing feature mapping between textual prompts and inspection images, demonstrating excellent research value in flexible industrial manufacturing. However, existing ZSAD methods are limited by closed-world settings, struggling to unseen defects with predefined prompts. Recently, adapting Multimodal Large Language Models (MLLMs) for Industrial Anomaly Detection (IAD) presents a viable solution. Unlike fixed-prompt methods, MLLMs exhibit a generative paradigm with open-ended text interpretation, enabling more adaptive anomaly analysis. However, this adaption faces inherent challenges as anomalies often manifest in fine-grained regions and exhibit minimal visual discrepancies from normal samples. To address these challenges, we propose a novel framework VMAD (Visual-enhanced MLLM Anomaly Detection) that enhances MLLM with visual-based IAD knowledge and fine-grained perception, simultaneously providing precise detection and comprehensive analysis of anomalies. Specifically, we design a Defect-Sensitive Structure Learning scheme that transfers patch-similarities cues from visual branch to our MLLM for improved anomaly discrimination. Besides, we introduce a novel visual projector, Locality-enhanced Token Compression, which mines multi-level features in local contexts to enhance fine-grained de-

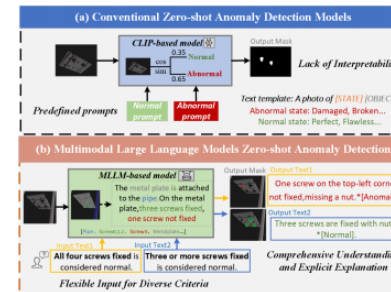


Fig. 1: Comparison between previous ZSAD methods and MLLMs-based ZSAD methods. (a) Previous ZSAD methods use fixed templates and generic descriptions, confined to closed-world anomaly detection. (b) MLLMs-based methods

## Industrial anomaly detection의 목표

- 비정상적인 패턴을 식별함으로써, 전반적인 생산성을 향상하기 위함

## Anomaly data

- 이상 데이터는 rarity하고 unpredictability 하기에 비지도 학습을 활용한 연구 진행
- 이상을 구별하기 위해 normal distribution으로부터 test sample feature와의 편차를 측정
- 이러한 방법은 많은 training sample이 필요하고, 새로운 클래스에 대한 일반화에 한계가 있음

## Zero-shot anomaly detection

- Text prompt를 사용해서 anomaly를 측정
- 이를 통해 보지 못한 객체에서의 탐지를 가능하게 하는 유연한 검사를 제공

## Introduction

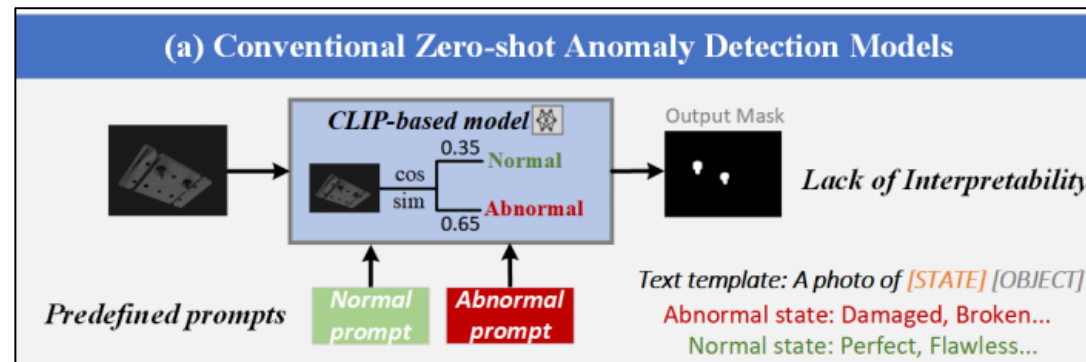
## Mainstream contrast-based method

- based on pre-trained CLIP
- Image feature와 정상 및 비정상을 나타내는 textual descriptors representing을 비교

|             |  |
|-------------|--|
| WinCLIP     | manually crafted prompt를 사용하여 처음에 CLIP을 조정                     |
| AnomalyCLIP | manual template을 object-agnostic text vector로 대체하여 일반적인 표현을 제공 |
| ClipSAM     | CLIP의 anomaly localization을 SAM에 활용                            |

## Limitations

- ZSAD가 closed-world setting에서 수행되고, predefined prompt로 제약된 의미 공간에서 binary classification을 실행하기에 unseen defect에 어려움을 겪고 있음
- 게다가 generic descriptor는 산업 제조의 다양한 이상을 포착하기에 부족함



## MLLMs

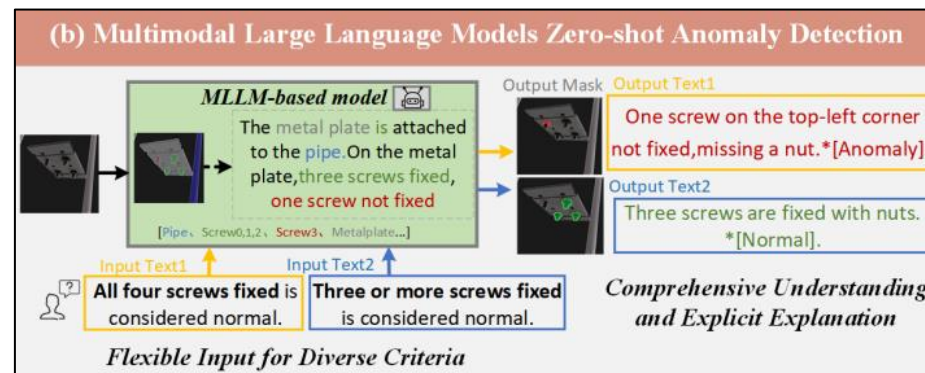
- Open-ended text interpretation을 통해, 보다 적응적인 이상 탐지를 가능하게 하는 generative paradigm을 나타냄
- 유연한 input-output 형식으로 뛰어난 image-text understanding을 입증
- 이러한 유연성 덕분에 임의의 textual prompt와 visual input을 처리할 수 있으며, 다양한 기준 및 시나리오에 따라 dynamic anomaly analysis를 촉진함

## Challenges

- 1) Anomalies are visually confusing
- 2) Normal and anomalous sample 사이의 최소한의 구별 가능한 특징 변화를 특성으로 함
- 3) 주로 객체가 국한된 영역에서 나타나는 특징

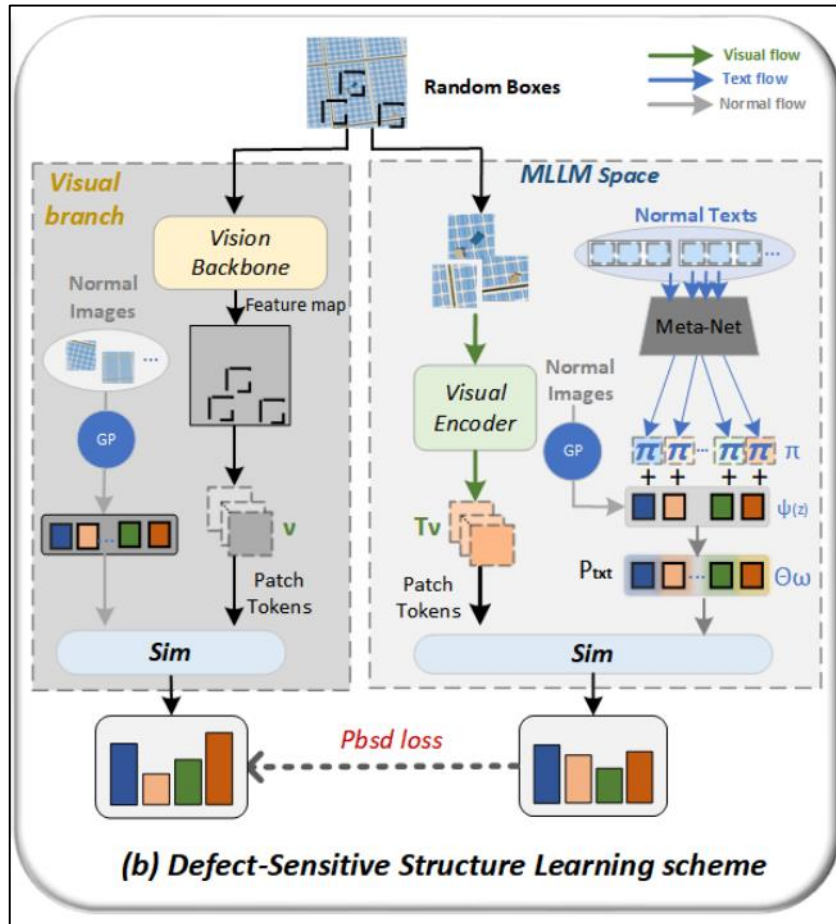


MLLMs often struggle with abnormality discriminability and fine-grained perception of defects



## DSSL

- MLLM의 약한 구별 능력의 원인으로, 종종 결함 있는 영역과 주변 정상 영역 사이의 경계를 흐릿하게 만드는 minor manufacturing flaw 또는 imperfection에서 기인한다고 함
- 이를 해결하기 위해 LLM 표현에서 abnormality distinction을 강화하기 위한 DSSL 제안
- 미세한 전역 시각적 차이에도 불구하고, 정상 sample과 비정상 sample은 patch-similarity distribution에서 뚜렷한 차이를 보임
- 구체적으로 normal patch는 여러 개의 unmarked sample에서 자주 visual parallel을 보이지만, anomalous area는 이러한 correspondence를 거의 찾기 힘들

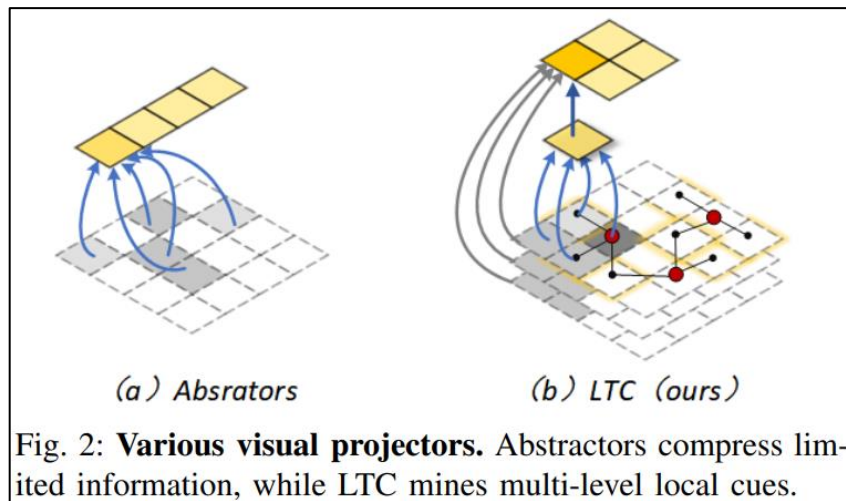


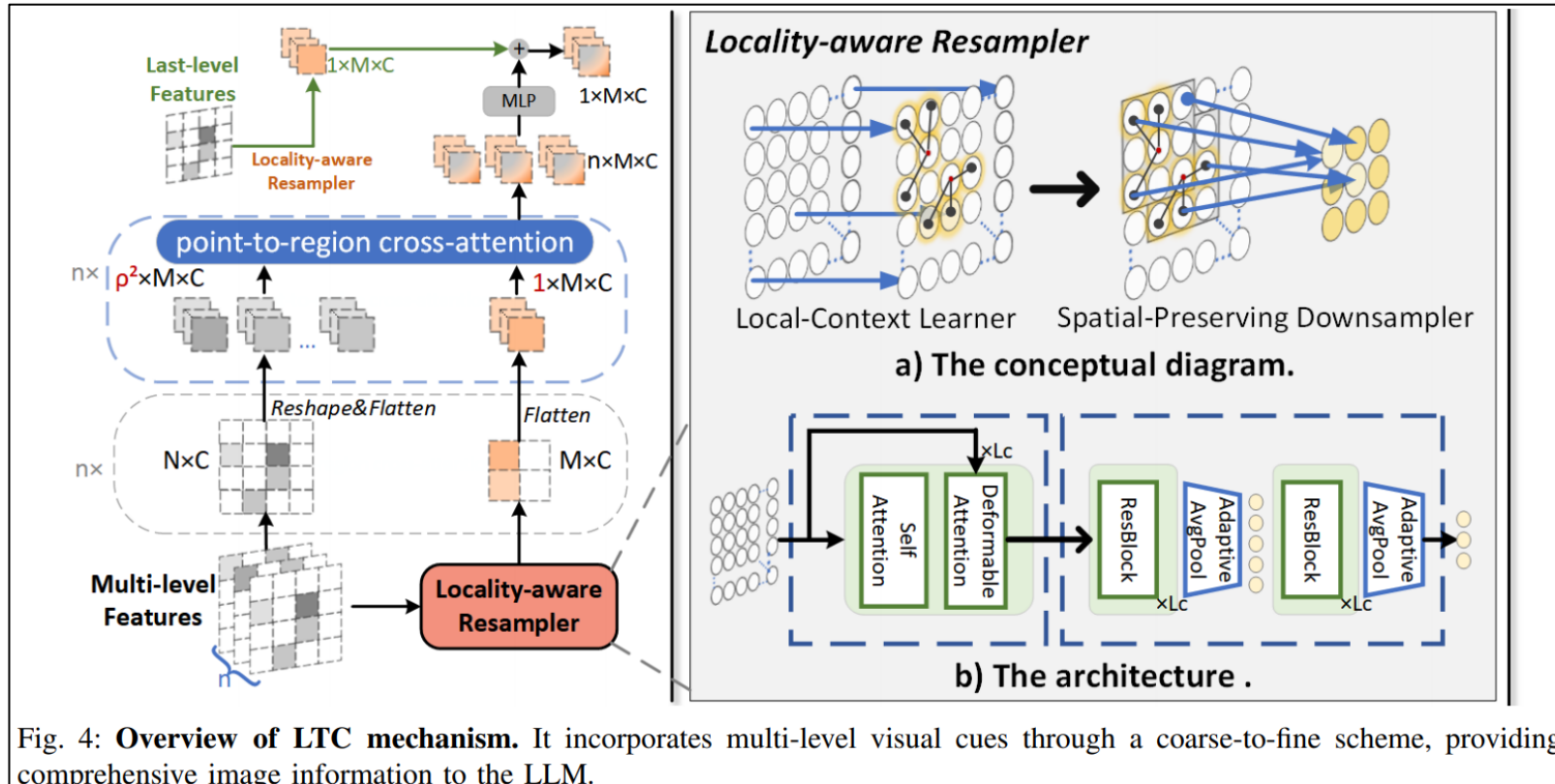
- Visual branch에서 DSSL은 patch embedding과 global normal feature 간의 local-global visual similarity를 계산
- MLLM 공간에서 similarity는 visual-patch token과 normal semantic token 간에 계산되며, 이는 언어적 단서와 시각적 단서를 모두 결합
- 마지막으로 DSSL은 contrastive loss를 사용하여 similarity distribution을 정렬해 patch-similarity 단서를 LLM 공간으로 효과적 전송

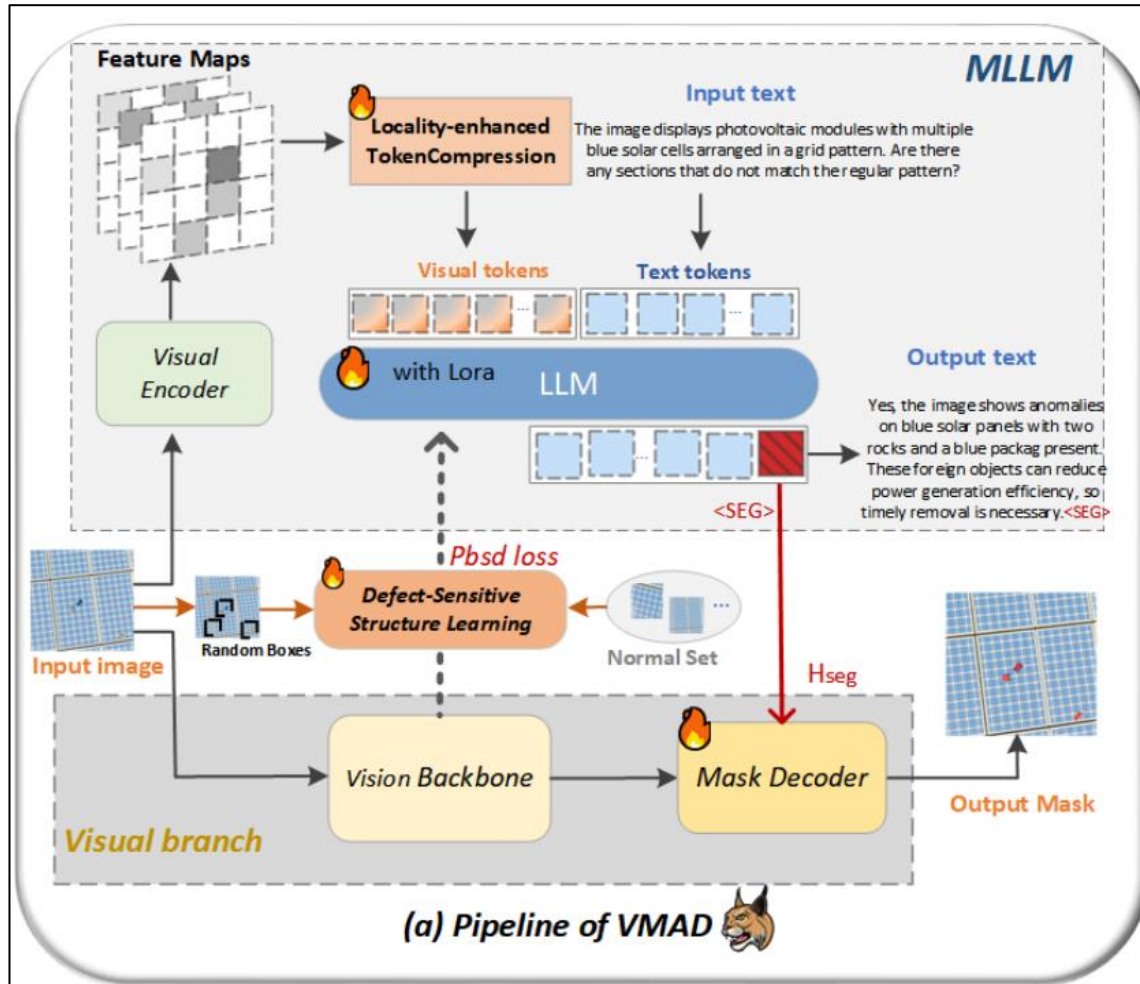


## LTC

- Visual signal을 LLM과 호환되는 visual token으로 변환시키는 visual projector가 MLLM의 미세한 이상 감지에 직접적인 영향을 미친다는 것을 관찰
- 효율성을 위해 visual projector로 abstractor를 사용하지만, 이는 제한된 영역에서 정보를 요약함으로써 visual feature의 무결성을 저해함
- 이를 해결하기 위해 LTC 제안
- Local context에서 multi-level feature를 추출하기 위해, multi-level feature integration과 coarse-to-fine injection 방식을 사용







## VMAD (Visual-enhanced MLLM Anomaly Detection)

- VMAD는 MLLM과 Visual branch로 구성
- MLLM은 image-text input을 처리하여 mask decoder의 prompt segmentation을 위한 <seg> token을 생성

## Two innovative modules

- DSSL**  
MLLM과 visual branch 간의 visual-similarity cue를 통합하여 향상된 anomaly discrimination을 제공
- LTC**  
MLLM 내에서 visual embedding을 처리하여 fine-grained defect perception을 위한 high-quality token을 생성

## Overall Architecture : MLLM framework

## (1) Visual encoder

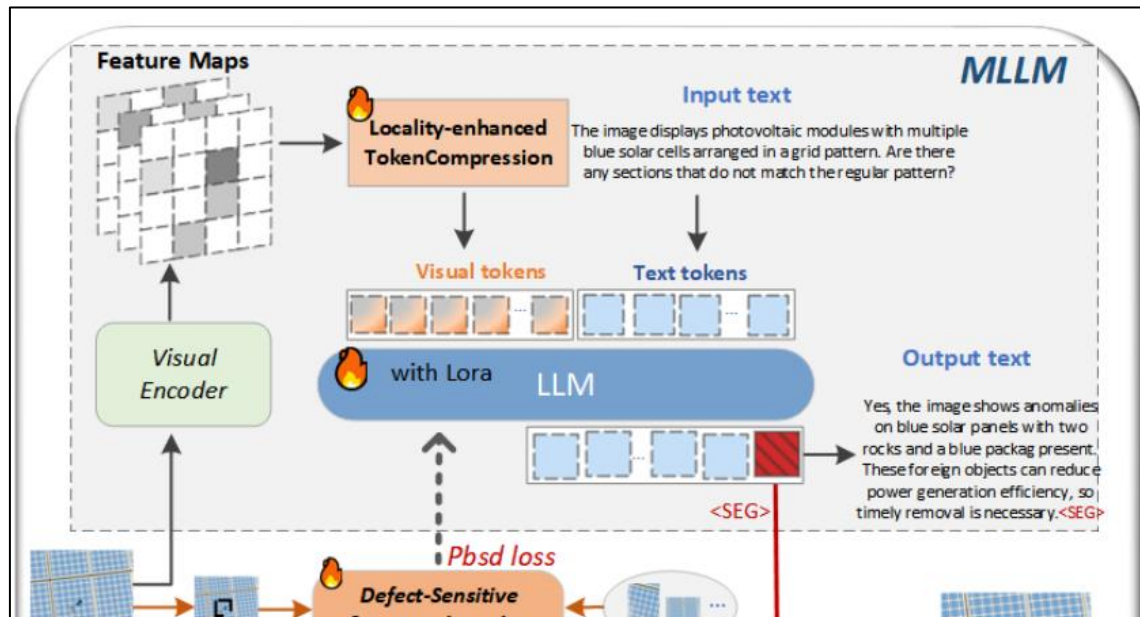
- 입력 이미지는 널리 사용되는 CLIP-ViT-L/14 vision encoder를 통해 visual embedding 집합으로 변환

## (2) Visual projector

- Visual embedding을 textual embedding space의 visual token으로 projecting 하는 역할
- 후속 language model에 적합한 차원을 보장함

## (3) LLM

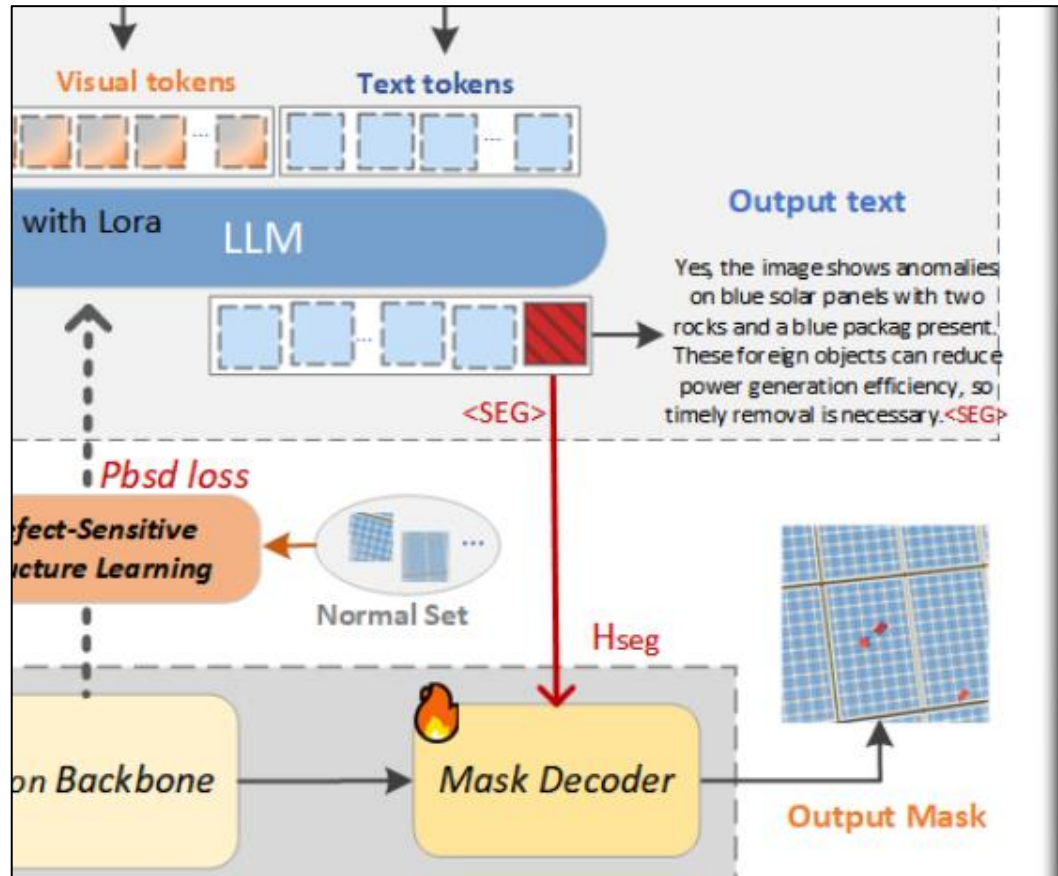
- Visual token과 textual token을 처리하고, 일관된 텍스트 응답을 auto-regressive 방식으로 생성



$$(1) \quad I_v = F_I(I_{\text{img}}),$$

$$(2) \quad T_v = \Psi_{I \rightarrow T}(I_v),$$

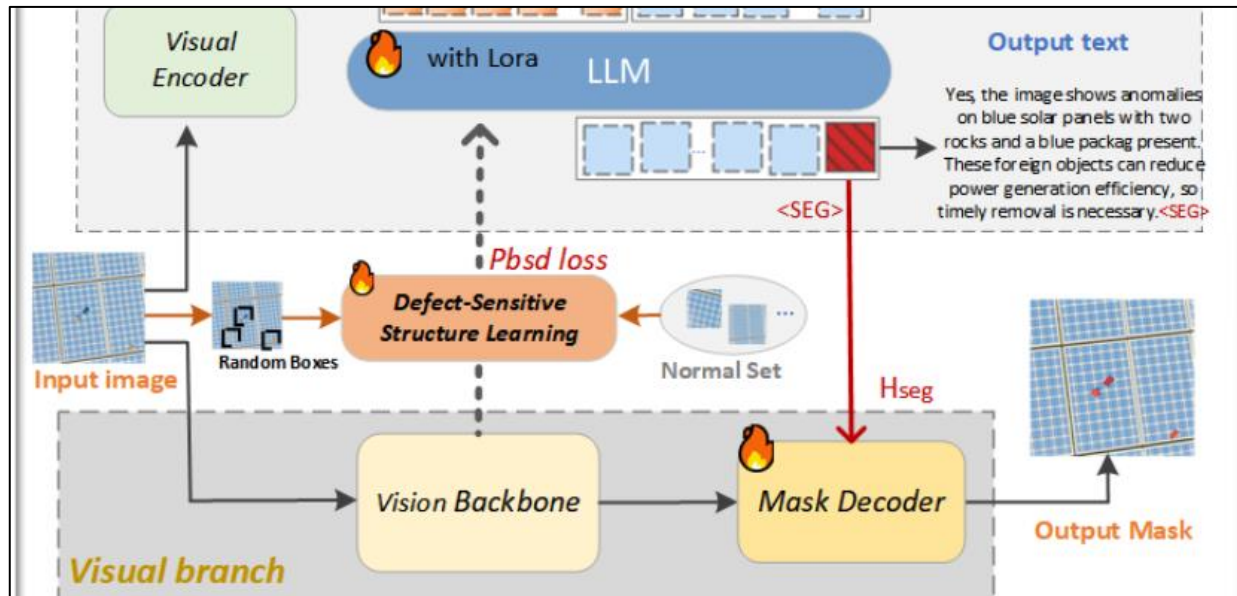
$$(3) \quad \hat{y}_{\text{txt}} = \Theta(T_v, T_t).$$



### Embedding-as-mask scheme

- MLLM에 anomaly localization을 위해 segmentation ability를 부여
- LLM의 vocabulary를 전문화된 <seg> token으로 확장하여 MLLM과 downstream mask decoder 간의 연결 역할을 함
- 추가로 text output이 <seg> token으로 끝나도록 보장

- LLM에서 <seg> token의 last-layer embedding  $\bar{H}_{seg}$ 가 추출되고, MLP projection layer를 통과하여  $H_{seg}$ 를 얻음
- $H_{seg}$ 는 mask decoder에게 anomaly discrimination을 위한 필수적인 semantic knowledge를 전달하는 역할
- $H_{seg}$ 와  $f$ 는 mask decoder로 전달되어 최종 segmentation mask를 얻음



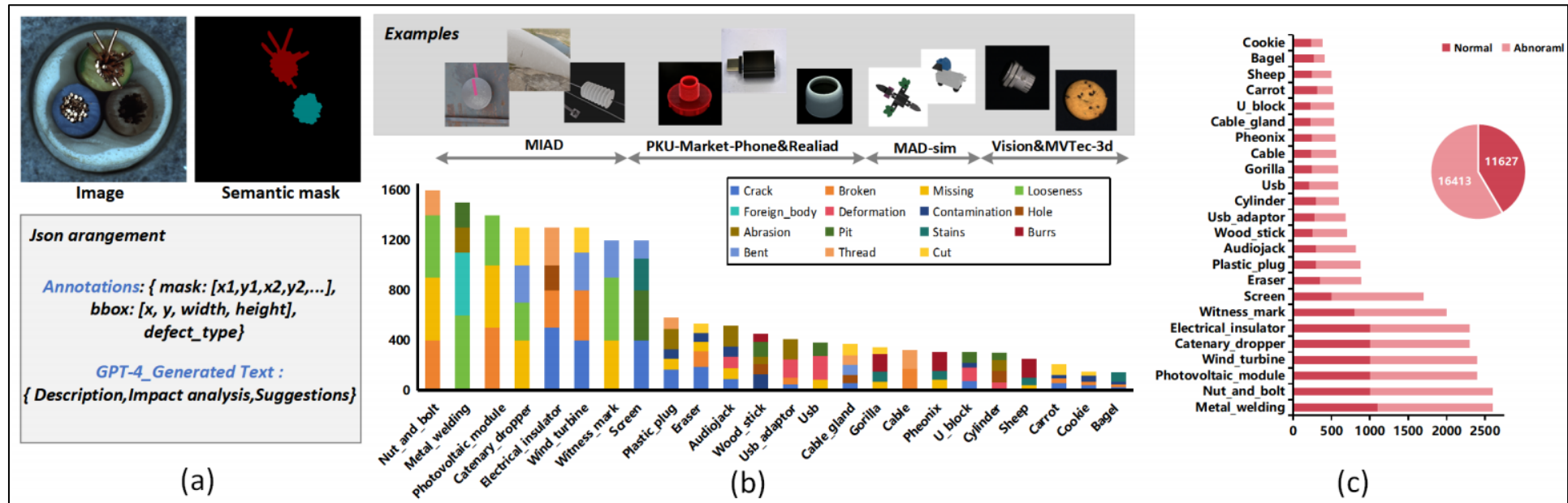
$$f = V_{backbone}(I_{img})$$

$$H_{seg} = \Gamma(\bar{H}_{seg})$$



## RIAD (Real Industrial Anomaly Detection Dataset)

- 24개의 object categories, 15개의 defect types, 28,040개의 이미지
- Anomaly mask가 있고, 쌍으로 된 image-text data를 특징으로 함
- Detailed anomaly descriptions, impact analyses, recommendations 제공
- (a) RIAD data pairs
- (b) categories of objects, quantity, defect types
- (c) The ratio of normal images and abnormal images



## Results : Cross-dataset evaluation

TABLE I: Zero-shot Anomaly Detection Performance on **Cross-dataset Evaluation** setting on MVTec-AD, Visa, and WFDD datasets. The best results are in **bold** while the second best are underlined. ‘Img’ and ‘Pixel’ represent the mean image-level and pixel-level AUROC.

| Methods         | MVTec-AD    |             |             |             | ViSA        |             |             |             | WFDD        |             |             |             |
|-----------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
|                 | Img         | Pixel       | PRO         | AP          | Img         | Pixel       | PRO         | AP          | Img         | Pixel       | PRO         | AP          |
| LISA [33]       | 89.1        | 91.6        | 82.2        | 81.2        | 81.3        | 85.2        | 77.8        | 79.3        | 92.1        | 93.5        | 86.0        | 87.2        |
| BGAD [42]       | 90.1        | 91.3        | 87.3        | 85.6        | <u>87.4</u> | 88.6        | 82.0        | 80.6        | 95.0        | 96.2        | 86.8        | 88.1        |
| AnomalyClip [7] | 91.5        | <u>94.7</u> | 86.2        | 83.4        | <u>82.3</u> | 87.1        | <u>84.2</u> | 80.3        | 94.1        | 95.3        | 84.9        | 83.8        |
| VAND [43]       | 91.8        | 92.5        | 88.2        | 86.8        | 83.5        | 90.7        | 81.9        | 82.8        | <u>94.6</u> | 96.1        | 84.3        | <u>88.5</u> |
| WinClip [6]     | <u>92.5</u> | 93.4        | <u>89.5</u> | 87.5        | 84.8        | 91.3        | 80.4        | 82.4        | 93.2        | 96.2        | 86.1        | 83.5        |
| AnomalyGPT [35] | 92.1        | 93.9        | 86.2        | 85.1        | 86.6        | <u>92.4</u> | 81.7        | 75.6        | 93.7        | <b>96.9</b> | 85.7        | 84.2        |
| Myriad [36]     | 93.2        | 92.3        | 87.9        | <b>88.2</b> | 85.9        | <u>90.5</u> | 81.3        | 83.5        | 91.8        | 93.9        | <u>87.3</u> | 82.7        |
| VMAD            | <b>95.8</b> | <b>96.1</b> | <b>91.2</b> | <u>87.6</u> | <b>89.7</b> | <b>93.8</b> | <b>84.6</b> | <b>85.2</b> | <b>95.3</b> | <u>96.7</u> | <b>89.2</b> | <b>90.1</b> |

- RIAD에서 훈련하고, 3개의 일반적인 공공 dataset에서 미세 조정 없이 평가
- 최고 결과는 **진하게** 표시되고, 두 번째 최고 결과는 **밑줄**
- MVTec-AD : 15개 클래스에서 5,354개의 고해상도 이미지
- ViSA : 색상이 있는 산업 부품 12개의 categories에서 10,821개의 이미지
- WFDD : 4개의 categories에서 다양한 질감과 패턴의 직물 이미지를 4,101개 사용

Img : mean image-level AUPRO  
 Pixel : pixel-level AUROC  
 PRO : Pixel-level Region Overlap  
 AP : Average Precision

AUPRO (Area Under the Receiver Operating Characteristic Curve)  
 - 이진 분류 모델의 성능을 평가하는 지표



TABLE II: Zero-shot Anomaly Detection Performance on Cross-category Evaluation across each class in RIAD.

| Methods         | metal_welding |             |             | u_block     |             |             | toy_brick   |             |             | ... | wind_turbine |             |             | Average     |             |             |
|-----------------|---------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-----|--------------|-------------|-------------|-------------|-------------|-------------|
|                 | Img           | Pixel       | PRO         | Img         | Pixel       | PRO         | Img         | Pixel       | PRO         |     | Img          | Pixel       | PRO         | Img         | Pixel       | PRO         |
| LISA [33]       | 84.1          | 90.5        | 81.0        | 89.1        | 91.2        | 86.6        | 80.4        | 82.3        | 76.1        |     | 81.8         | 89.5        | 82.7        | 79.2        | 82.7        | 78.9        |
| BGAD [42]       | 91.6          | 94.7        | 87.2        | 79.1        | 82.4        | 77.2        | 86.2        | 87.3        | 85.2        |     | 91.8         | 93.5        | 87.1        | 89.2        | 91.4        | 87.1        |
| AnomalyClip [7] | 90.2          | 92.2        | <u>91.1</u> | 87.4        | 90.7        | 89.2        | 89.1        | 93.7        | 89.2        | ... | 88.9         | 90.6        | 89.1        | 92.7        | <u>96.6</u> | <u>90.5</u> |
| VAND [43]       | <u>96.2</u>   | 95.4        | 94.2        | 91.2        | <u>96.3</u> | <u>89.7</u> | <u>91.7</u> | 92.3        | 86.9        |     | 78.8         | 82.6        | 88.1        | 92.1        | 94.7        | 89.4        |
| WinClip [6]     | 94.8          | <u>96.1</u> | 90.8        | <u>90.8</u> | 95.2        | 87.6        | 90.4        | 91.5        | 85.3        |     | 89.2         | <u>94.6</u> | <u>87.9</u> | <u>94.2</u> | 93.1        | 86.1        |
| AnomalyGPT [35] | 90.1          | 86.2        | 84.6        | 83.7        | 87.1        | 88.2        | 89.5        | 92.4        | 80.1        |     | 78.2         | 78.1        | 67.8        | 87.9        | 89.4        | 87.1        |
| Myriad [36]     | 95.2          | 93.1        | 88.1        | 84.6        | 84.4        | 89.4        | 93.3        | <u>95.8</u> | 90.8        |     | 89.1         | 81.9        | 87.4        | 91.1        | 92.4        | 85.2        |
| VMAD            | <b>97.2</b>   | <b>98.6</b> | <b>91.6</b> | <b>91.1</b> | <b>98.9</b> | <b>90.2</b> | <b>97.2</b> | <b>98.1</b> | <b>92.3</b> |     | <b>93.7</b>  | <b>98.1</b> | <b>89.5</b> | <b>94.9</b> | <b>98.9</b> | <b>92.3</b> |

- RIAD를 무작위로 겹치지 않는 하위 집합으로 나눔
- 훈련을 위한 16개의 클래스(15,274개)와 테스트를 위한 8개의 클래스(8,342개)

## Implementations

- Multi-modal LLM : LLaVA-7B-v1-1
- Visual backbone : ViT-H SAM
- LLM Fine-tuning : LoRA
- two MLLM-based anomaly detection models : **Myriad**, **AnomalyGPT**
- three zero-shot anomaly detection models : **WinClip**, **VAND**, **AnomalyClip**
- one supervised anomaly detection model : **BGAD**
- one multi-modal pixel-grounding model : **LISA**

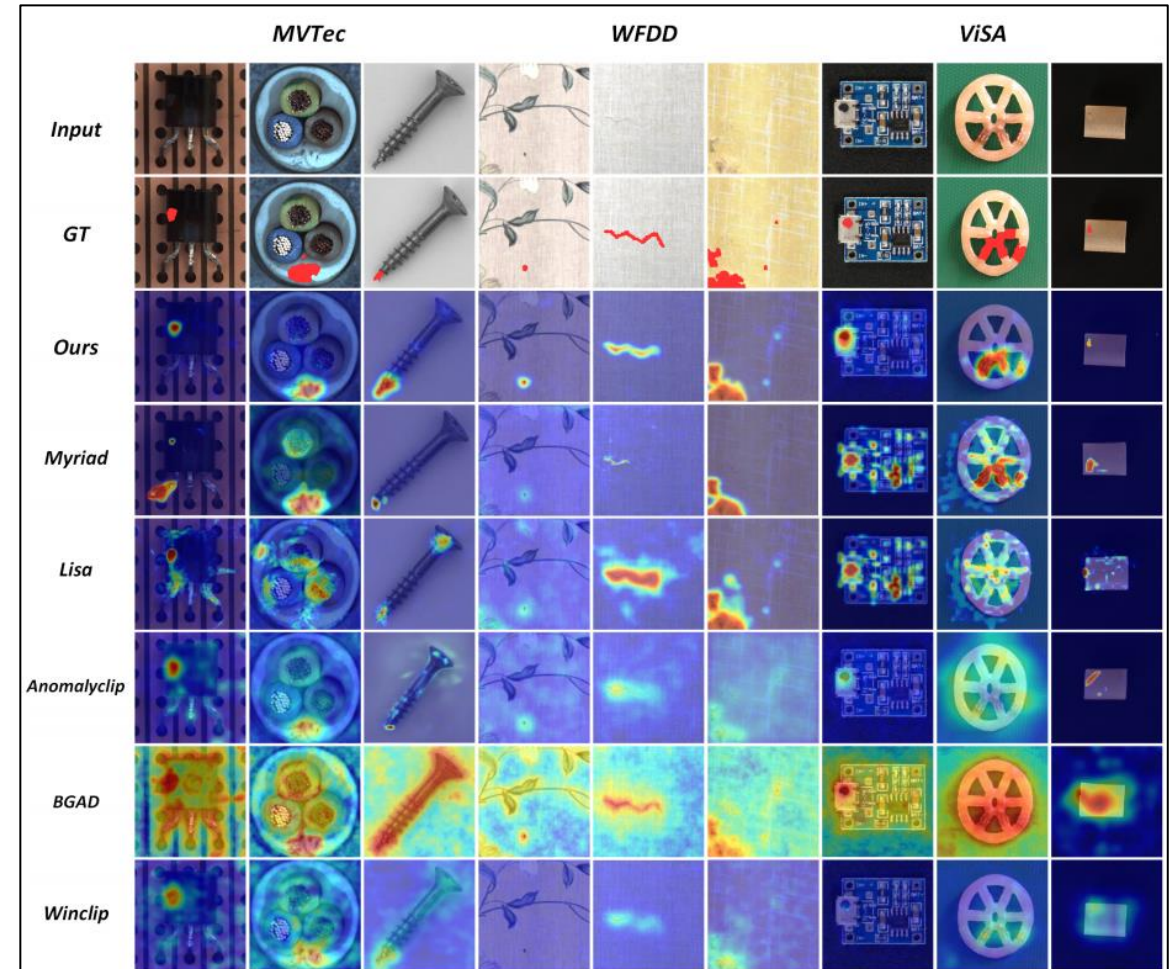


Fig. 6: Zero-shot anomaly segmentation on MVTec-AD, WFDD, and ViSA datasets.

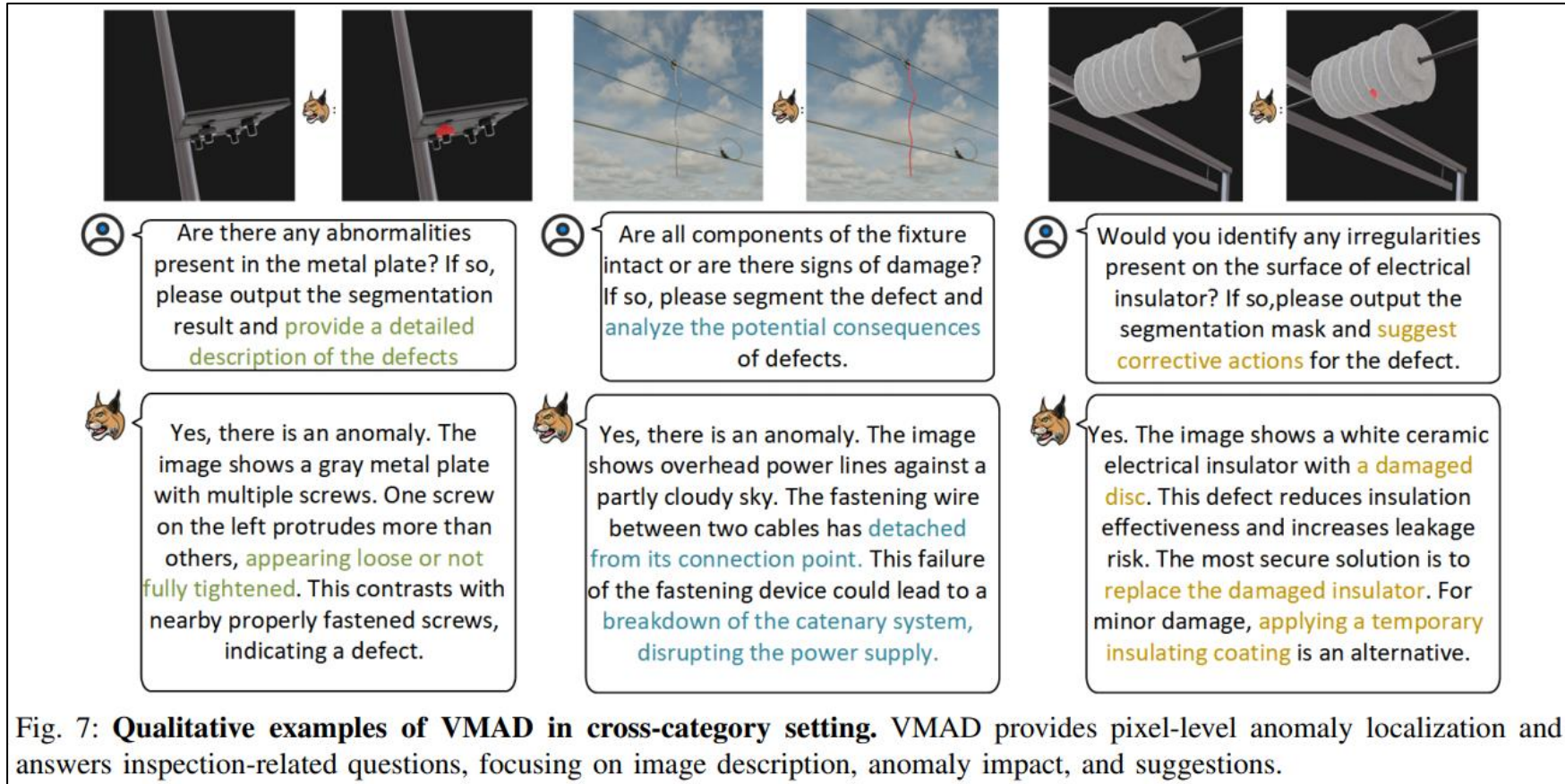


Fig. 7: **Qualitative examples of VMAD in cross-category setting.** VMAD provides pixel-level anomaly localization and answers inspection-related questions, focusing on image description, anomaly impact, and suggestions.

Image description

Image description

Image description

Anomaly impact

Anomaly impact

Suggestion

## Results

## Ablation Study

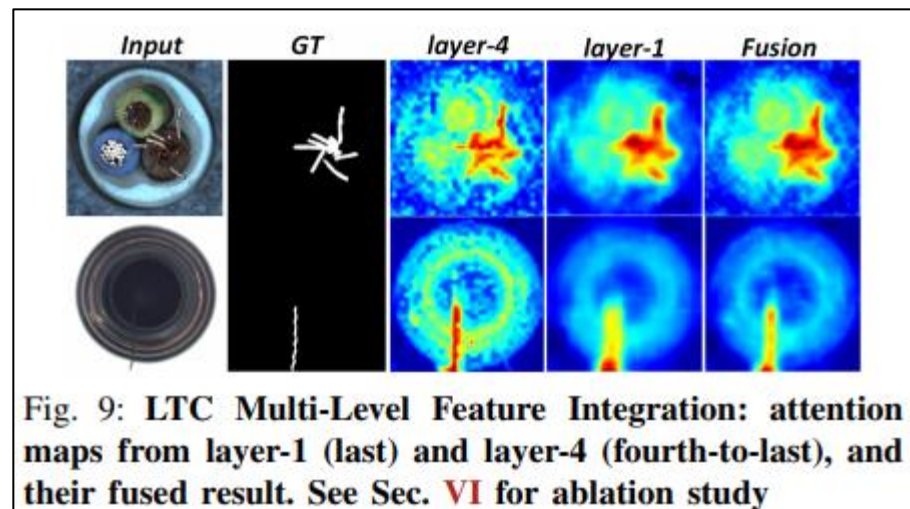
- LTC와 DSSL을 활성화 또는 비활성화했을 때의 성능을 비교
- 모두 활성화 되었을 때, 성능이 가장 높은 것을 확인할 수 있음

## LTC Multi-Level Feature Integration

- Layer-4에서는 주로 세부 사항을 포착하여 이상이 있는 부분을 강조
- Layer-1에서는 전반적인 구조에 대한 정보를 포함, 세부 사항보다는 큰 틀에서의 이상을 탐지
- 마지막 이미지는 layer-4와 layer-1의 출력을 결합한 결과
- 각각의 강점을 결합하여 보다 정교하고 정확한 이상 탐지를 가능하게 함

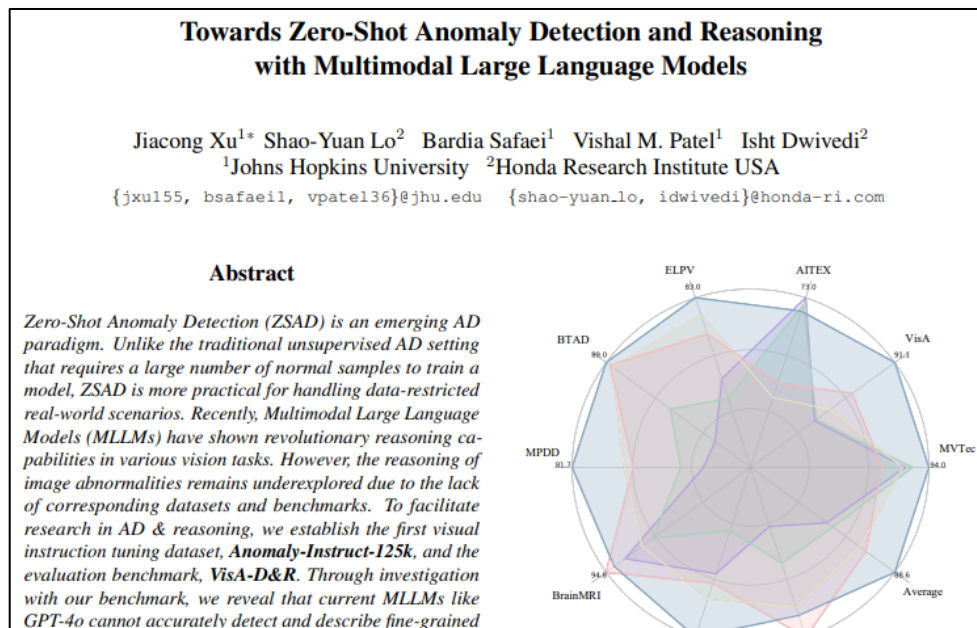
TABLE III: Component-wise experimental results.

| LTC | DSSL | MVTec       |             |             | RIAD        |             |             |
|-----|------|-------------|-------------|-------------|-------------|-------------|-------------|
|     |      | Img         | Pixel       | PRO         | Img         | Pixel       | PRO         |
| ✗   | ✗    | 90.5        | 91.3        | 87.3        | 91.3        | 93.4        | 89.5        |
| ✗   | ✓    | 93.6        | 94.3        | 89.1        | 92.1        | 94.5        | 90.8        |
| ✓   | ✗    | 92.9        | 93.7        | 90.5        | 93.2        | 96.1        | 91.5        |
| ✓   | ✓    | <b>95.8</b> | <b>96.1</b> | <b>91.2</b> | <b>94.9</b> | <b>98.9</b> | <b>92.3</b> |


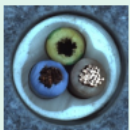
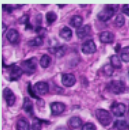
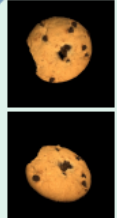




- 시각적 지침 조정 dataset인 **Anomaly-Instruct-125k** 소개
- Evaluation benchmark인 **VisA-D&R** 제안
- **Anomaly-OV(OneVision)** 모델 제안
- 사람의 시각 검사 행동에 영감을 받아 만든 **LTFM(Look-Twice Feature Matching)** mechanism 소개

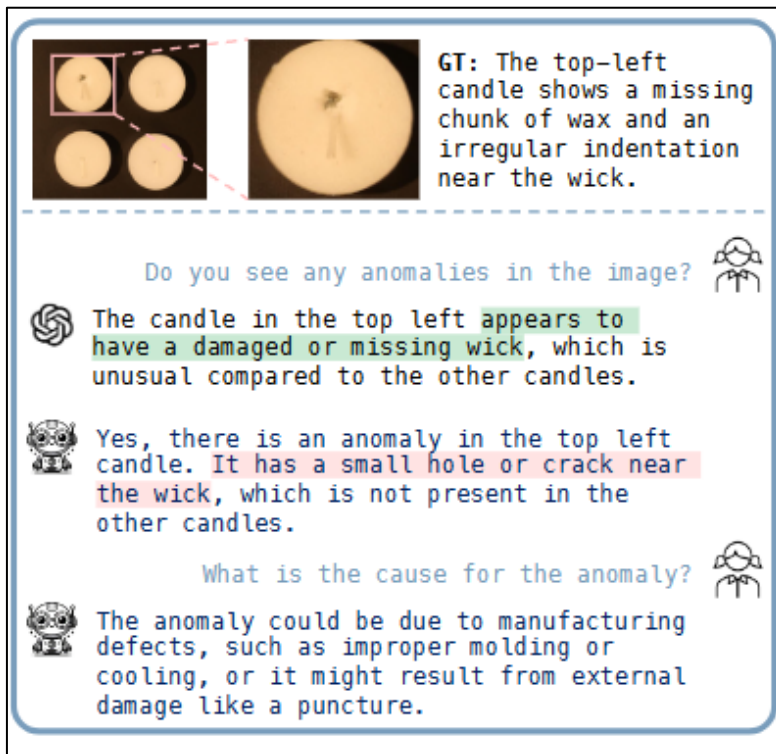


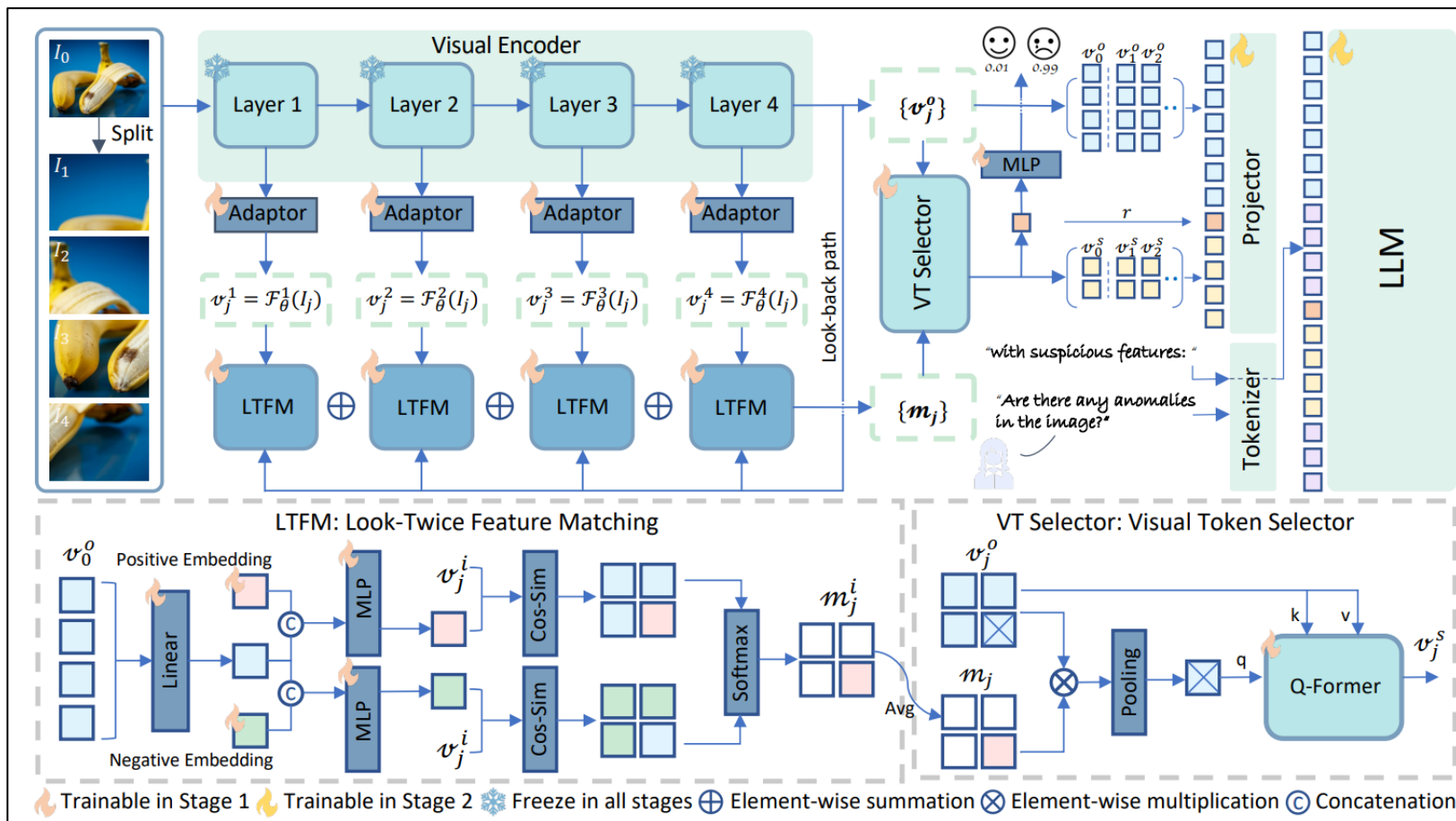
- 다양한 이미지와 그에 대한 자세한 설명을 포함
- MLLMs가 이미지 내 이상을 식별하고 설명할 수 있도록 도움
- Google 이미지 검색과 GPT-4o를 결합하고 자동화를 통해 데이터 수집
- In-the-wild, industrial, medical, and 3D
- 추론 단어는 파란색으로 강조 표시

|   |  |   |   |
|---|--|---|---|
| <p><b>In the wild</b></p>  <p><b>Q:</b> Do you notice any anomalies in the image?<br/> <b>A:</b> Yes, the anomaly in the image is a fine, dark crack visible on the ceramic cup, starting from the rim and extending downward.</p> <p><b>Q:</b> How can I prevent similar anomalies in the future?<br/> <b>A:</b> Avoid rapid temperature changes, handle the cup carefully to minimize impacts, and consider using cups made with thicker, more durable materials.</p> | <p><b>Industry</b></p>  <p><b>Q:</b> Are there any anomalies in this cable image?<br/> <b>A:</b> Yes, the green wire is missing its internal copper conductors, unlike the blue and brown wires, which contain complete sets of bundled copper strands.</p> <p><b>Q:</b> What could be the reason behind this anomaly?<br/> <b>A:</b> The absence of copper in the green wire could be due to manufacturing defects such as failure to insert the strands or a mechanical issue where the strands were removed or misplaced during production or handling.</p> | <p><b>Medical</b></p>  <p><b>Q:</b> What does the image show?<br/> <b>A:</b> The image shows a densely cellular tissue sample with features indicating potential metastatic breast cancer.</p> <p><b>Q:</b> What suggests that there might be malignant activity in the tissue sample?<br/> <b>A:</b> Malignant activity is suggested by irregularly shaped cells, hyperchromatic nuclei, and an increased nuclear-to-cytoplasmic ratio, indicative of abnormal and increased DNA content typical in cancerous cells.</p> | <p><b>3D</b></p>  <p><b>Q:</b> What kind of object are you examining in the images?<br/> <b>A:</b> I'm examining a round cookie with a golden-brown surface and visible chocolate chips embedded across its surface.</p> <p><b>Q:</b> Are there any significant anomalies detected on this cookie?<br/> <b>A:</b> Yes, there are two significant anomalies on the cookie: a central defect featuring a noticeable hole and an edge defect showing a gap or crack.</p> |
|---|--|---|---|

## 현재 MLLM의 한계

- 이미지 내의 fine-grained anomalous detail을 정확하게 탐지하고 설명할 수 없음







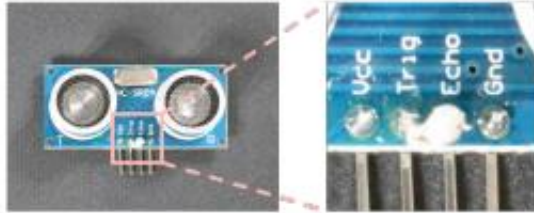
## Results

| Method             | MVTec       | VisA        | HeadCT      | BrainMRI    |
|--------------------|-------------|-------------|-------------|-------------|
| Full Model         | <b>94.0</b> | <b>91.1</b> | <b>97.6</b> | 93.9        |
| w/o. Look-back     | 92.8        | 90.5        | 96.6        | 93.5        |
| w/o. $e^+$ & $e^-$ | 92.1        | 90.1        | 94.7        | 92.9        |
| w/o. Q-Former      | 91.7        | 89.9        | 92.8        | <b>95.1</b> |
| w/o. WebAD         | 88.5        | 88.9        | 91.2        | 93.4        |

| Model             | Industrial Defects |             |             |             |             |             | Medical Anomalies |             |             | Average     |
|-------------------|--------------------|-------------|-------------|-------------|-------------|-------------|-------------------|-------------|-------------|-------------|
|                   | MVTec AD           | VisA        | AITEX       | ELPV        | BTAD        | MPDD        | BrainMRI          | HeadCT      | Br35H       |             |
| CLIP [73]         | 74.1               | 66.4        | 71.0        | 59.2        | 34.5        | 54.3        | 73.9              | 56.5        | 78.4        | 63.1        |
| CoOp [108]        | 88.8               | 62.8        | 66.2        | 73.0        | 66.8        | 55.1        | 61.3              | 78.4        | 86.0        | 70.9        |
| WinCLIP [38]      | 91.8               | 78.8        | <b>73.0</b> | 74.0        | 68.2        | 63.6        | 92.6              | 90.0        | 80.5        | 79.2        |
| APRIL-GAN [11]    | 86.2               | 78.0        | 57.6        | 65.5        | 73.6        | 73.0        | 89.3              | 89.1        | 93.1        | 78.4        |
| AnoVL [19]        | <u>92.5</u>        | 79.2        | <u>72.5</u> | 70.6        | 80.3        | 68.9        | 88.7              | 81.6        | 88.4        | 80.3        |
| AnomalyCLIP [110] | 91.5               | 82.1        | 62.2        | <u>81.5</u> | 88.3        | <u>77.0</u> | 90.3              | <u>93.4</u> | 94.6        | 84.5        |
| AdaCLIP [6]       | 89.2               | <u>85.8</u> | 64.5        | <u>79.7</u> | <u>88.6</u> | 76.0        | <b>94.8</b>       | 91.4        | <b>97.7</b> | <u>85.3</u> |
| Ours              | <b>94.0</b>        | <b>91.1</b> | 72.0        | <b>83.0</b> | <b>89.0</b> | <b>81.7</b> | <u>93.9</u>       | <b>97.6</b> | <u>95.5</u> | <b>88.6</b> |

## Results

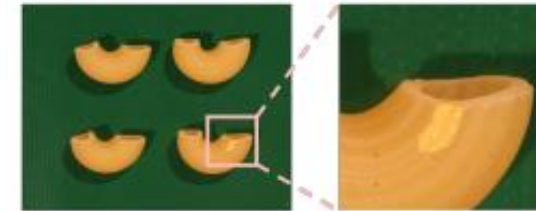
PCB Example: Excessive Solder



|          |   |
|----------|---|
| User     | Are there any anomalies visible in the image?   |
| GPT-4o   | The image shows an HC-SR04 ultrasonic sensor. There don't appear to be any visible anomalies with the sensor itself. The components and pins seem intact and properly aligned.                      |
| LLaVA-OV | The image shows an ultrasonic sensor module, specifically the HC-SR04 model. There are no visible anomalies in the image; it appears to be a standard representation of this type of sensor module. |
| Ours     | Yes, there is a noticeable anomaly in the image: a short circuit on the PCB, indicated by a white, crusty residue near the 'Echo' pin.  |

Table 3. Anomaly-OV presents more accurate anomaly detection.

Macaroni Example: Yellowish Spot



|          |   |
|----------|---|
| User     | Do you see any anomalies in the image?  |
| GPT-4o   | The image shows four pieces of elbow macaroni on a green background. The anomaly is that the macaroni pieces are not whole; they are cut in half.   |
| LLaVA-OV | The image shows four pieces of pasta, specifically macaroni shells, arranged on a green textured surface. The pasta appears to be uniformly colored and shaped, with no visible defects or anomalies. |
| Ours     | Yes, there is an anomaly in the image. The bottom right pasta piece has a noticeable yellowish discoloration on its surface.  |

Table 5. Anomaly-OV presents more precise anomaly reasoning.

