

2025 여름 세미나

2025.06.27



Sogang University

Vision & Display Systems Lab, Dept. of Electronic Engineering



Presented By

이준호

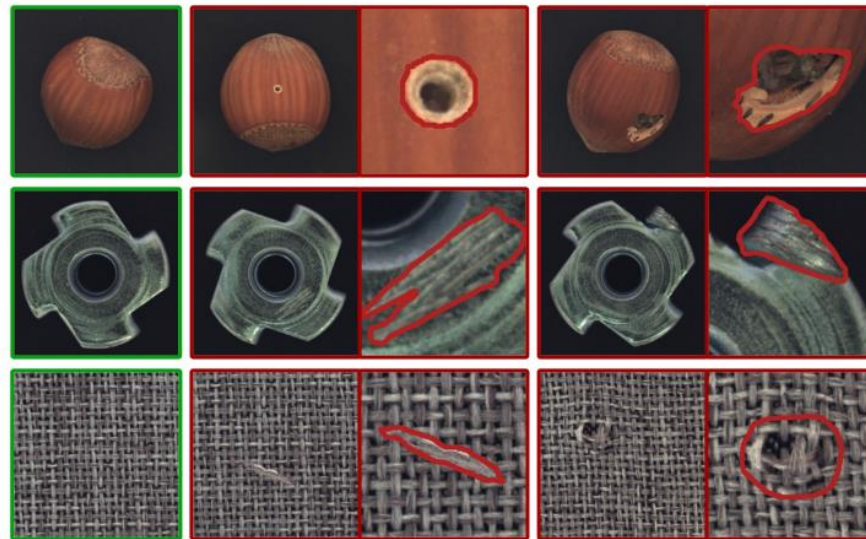
Outline

- Background
 - Anomaly detection1
 - Few-shot anomaly detection
 - Zero-shot anomaly detection
- Paper 1
 - AnomalyGPT: Detecting Industrial Anomalies Using Large Vision-Language Models (AAAI 2024)
- Paper 2
 - Towards Zero-Shot Anomaly Detection and Reasoning with Multimodal Large Language Models (CVPR 2025 Highlight)

Background

- 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

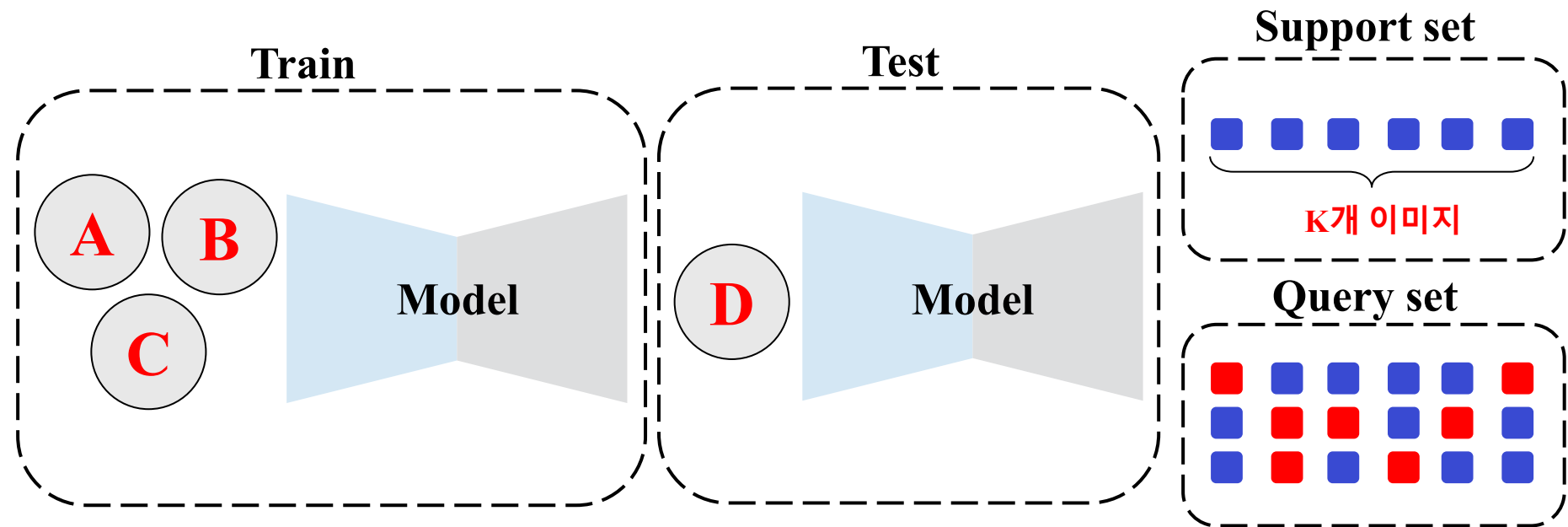
- Few-shot learning in anomaly detection

- Train

- A, B, C class로 normality, anomaly 학습

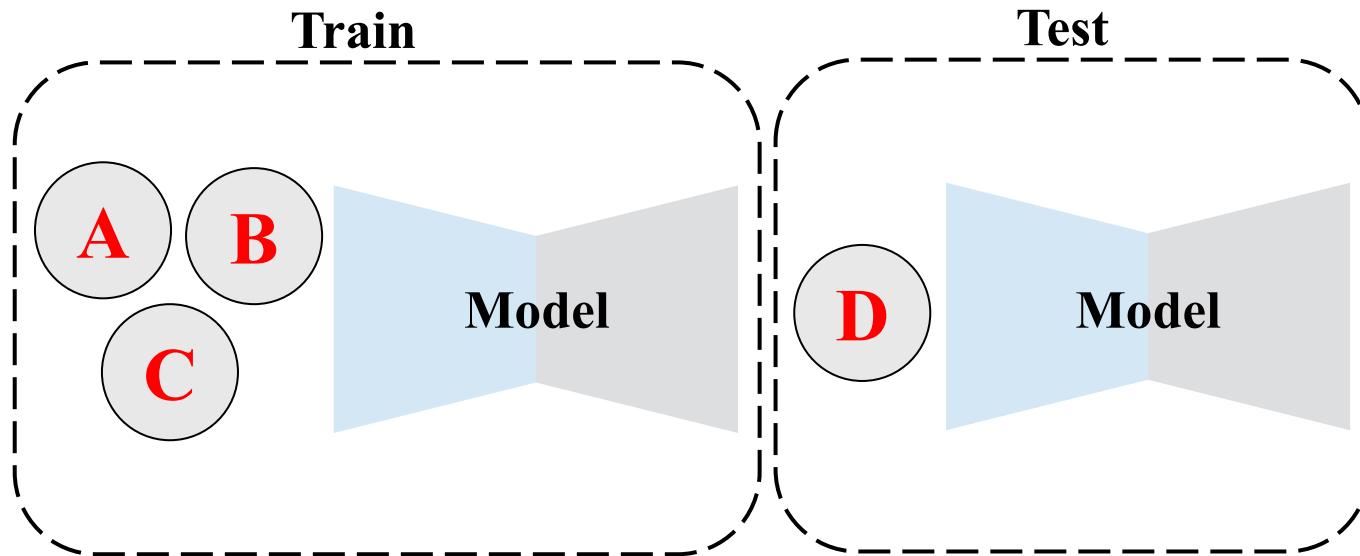
- Test

- A, B, C class로 학습된 weight를 기반으로 D class의 support set(k개의 normal sample)으로 재학습한 뒤, D class의 Query set(normal, anomaly 데이터 공존)로 inference



Background

- Zero-shot learning in anomaly detection
 - Train
 - A, B, C class로 normality, anomaly 학습
 - Test
 - A, B, C class로 학습된 weight를 기반으로 D class로 inference



Background

- Vision-Language Model in anomaly detection

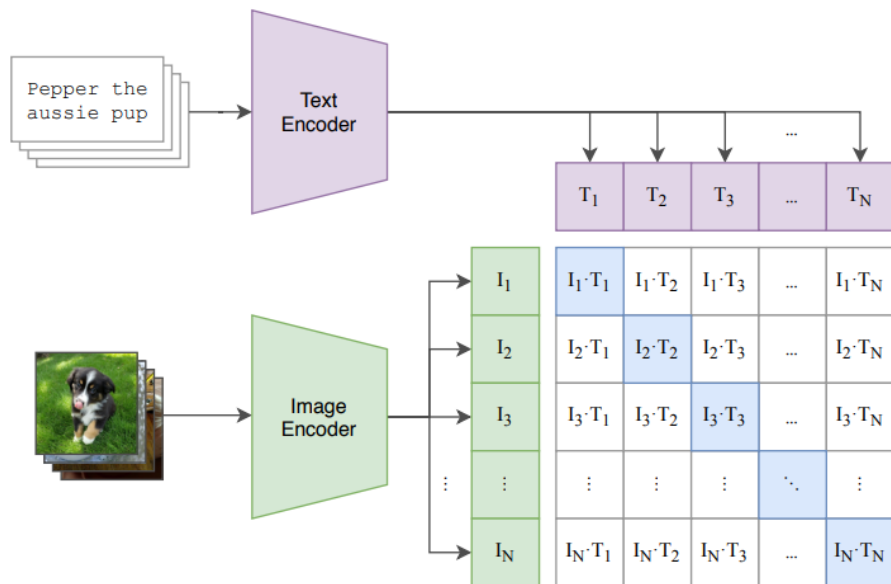
- CLIP 모델

- Text encoder와 image encoder로 구성

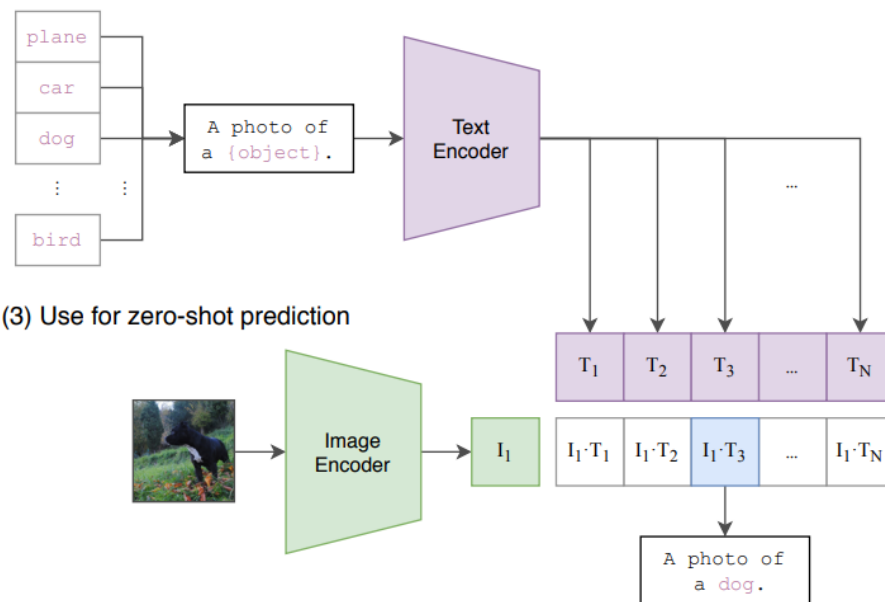
- ※ ImageNet으로 pretrain되어 있음

- 두 encoder의 output을 fusion하여 anomaly score 산출

(1) Contrastive pre-training



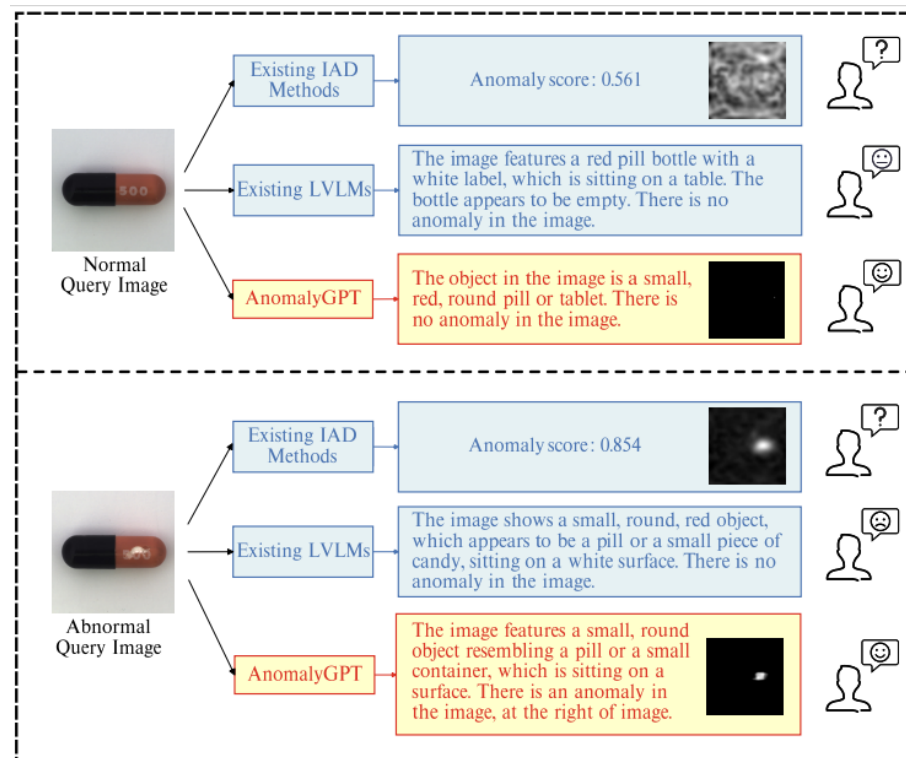
(2) Create dataset classifier from label text



AnomalyGPT: Detecting Industrial Anomalies Using Large Vision-Language Models

Paper 1

- LLM with anomaly detection task 설명
 - 일반적인 AD 모델 : anomaly score
 - 일반적인 LLM 모델 : image description (not correct)
 - LLM with AD : image description (correct)



Paper 1

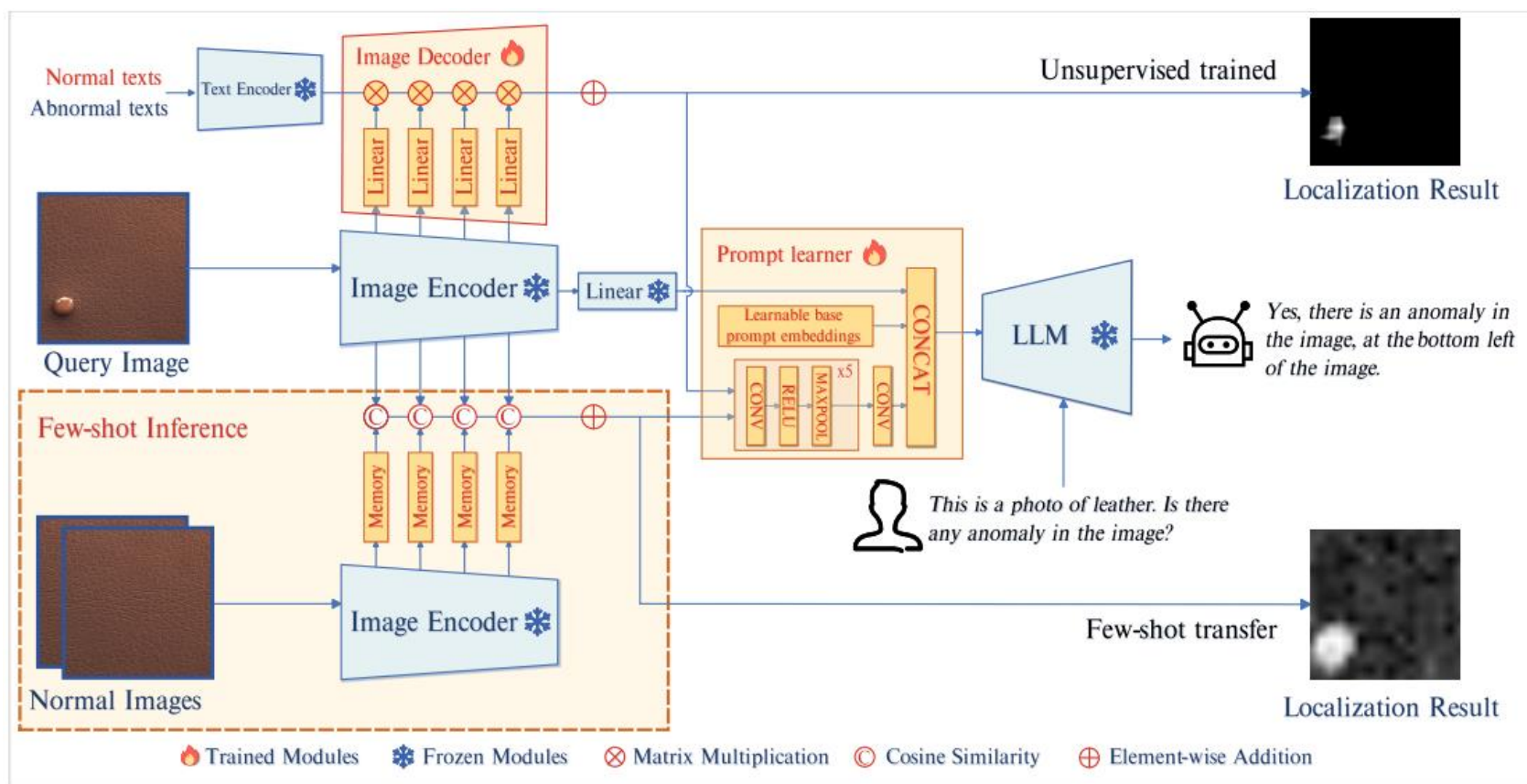
- LLM with anomaly detection task 설명
 - 일반적인 AD 모델 : anomaly score
 - Question & Answer 기능 X
 - 일반적인 LLM 모델 : image description (not correct)
 - Normality, anomalyity 학습 X
 - LLM with AD : image description (correct)

Methods	Few-shot learning	Anomaly score	Anomaly localization	Anomaly judgement	Multi-turn dialogue
Traditional IAD methods		✓	✓		
Few-shot IAD methods	✓	✓	✓		
LVLMS	✓				✓
AnomalyGPT (ours)	✓	✓	✓	✓	✓

Paper 1

• Overview

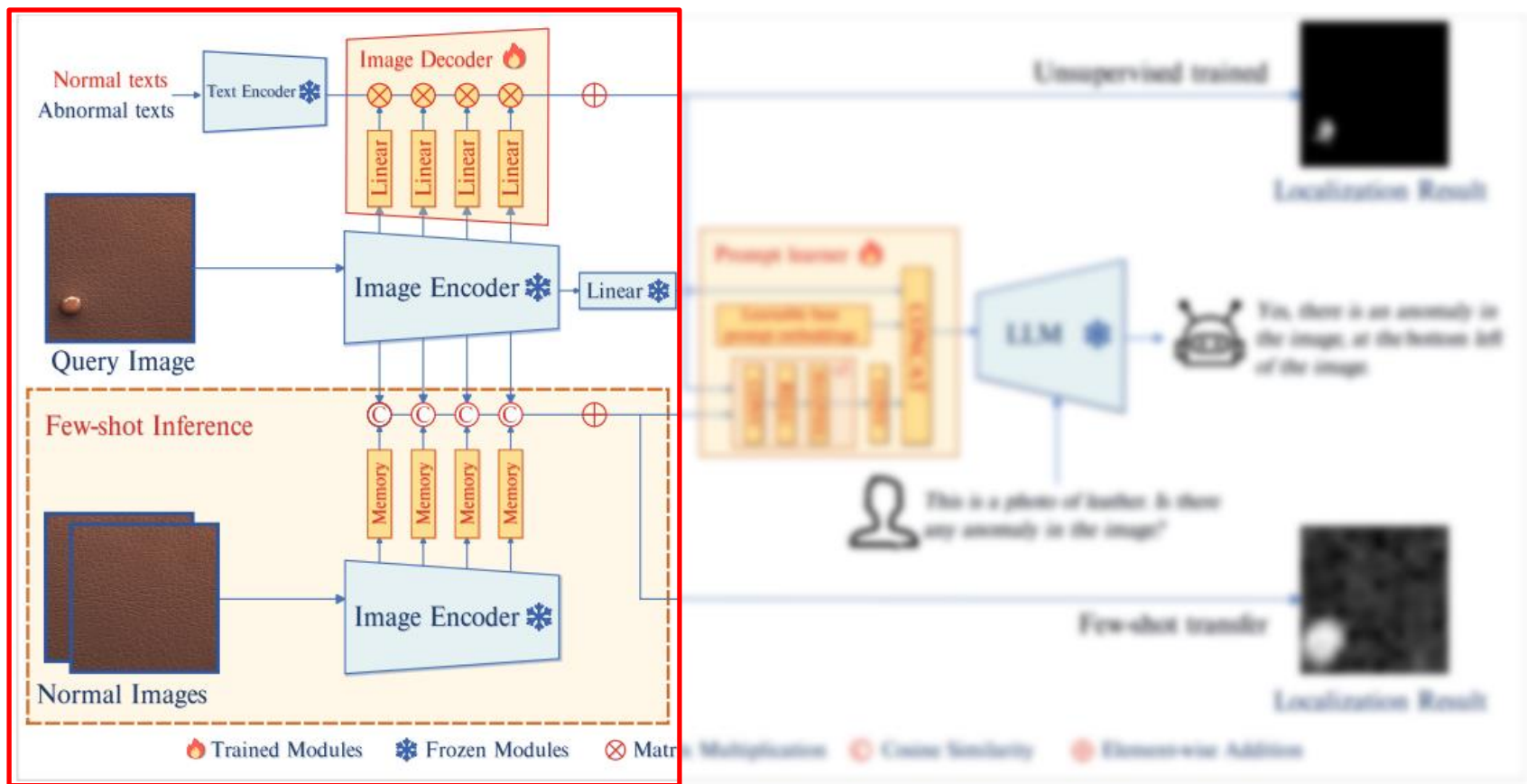
- 목표 1 : pretrained LLM을 AD task에 finetuning 시키기
- 목표 2 : normality, anomaly 학습



Paper 1

- Method

- Image Decoder



Paper 1

• Method

▪ Image decoder

- Image, text features

⌘ Pretrained encoder의 각 4개의 stage에서 patch-level feature를 추출한 뒤, aggregate

⌘ Image, text feature의 align(차원)을 맞추기 위해 learnable한 linear layer 사용하여 align을 맞춤

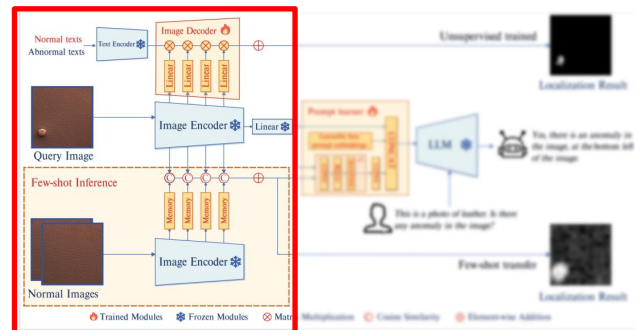
- Text feature는 normal prompt, abnormal prompt로 구성되어 있어, 각 prompt에 대한 text feature에 visual feature를 아래와 같이 연산함

$$M = Upsample \left(\sum_{i=1}^4 softmax(\tilde{F}_{patch}^i F_{text}^T) \right)$$

- 동일한 image encoder를 사용하여 normal 샘플의 중간 patch-level feature를 추출한 뒤, 이를 memory bank (B)에 저장함

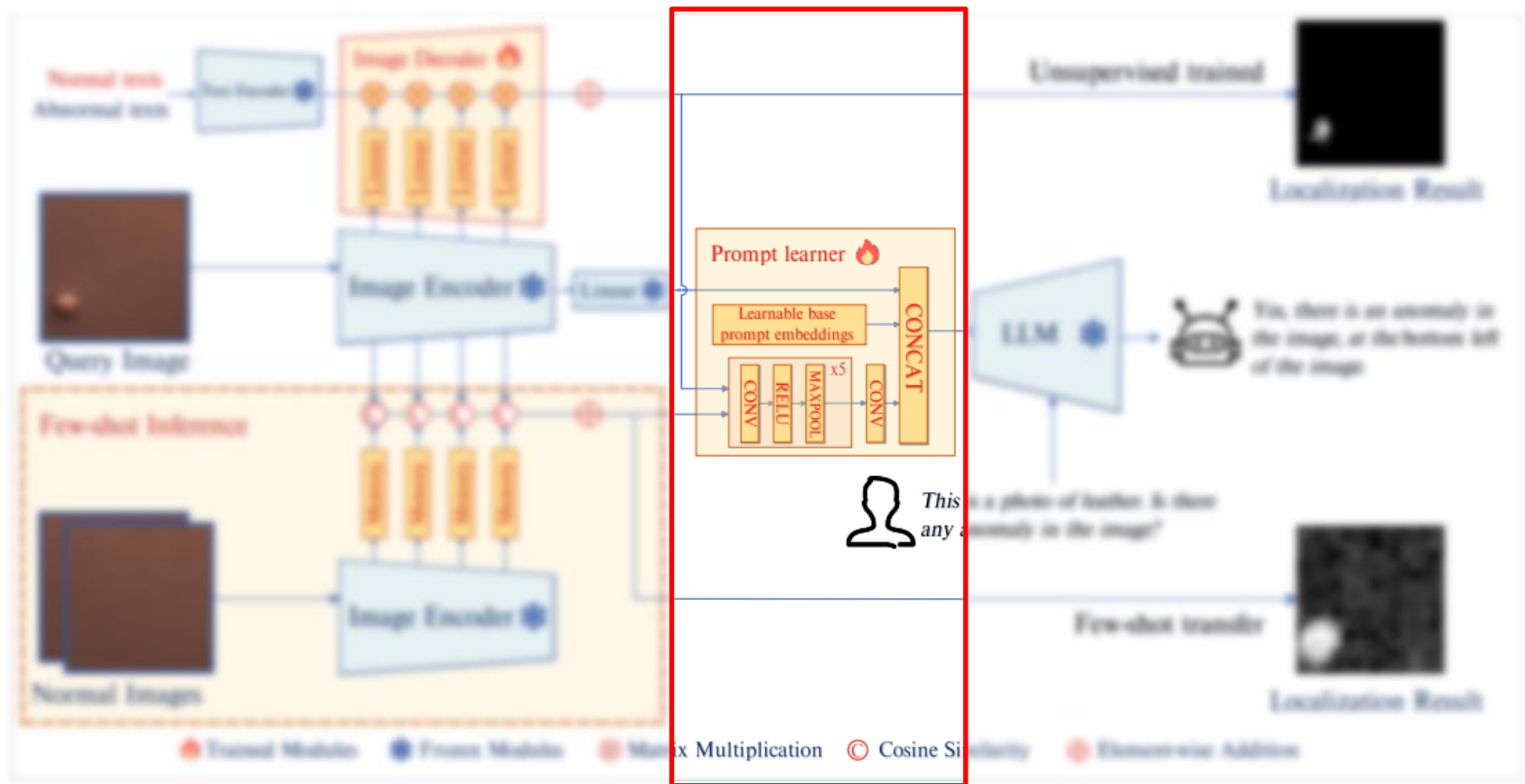
⌘ 각 patch와 memory bank내에 가장 유사한 feature간의 거리를 계산하여 localization results를 얻음

$$M = Upsample \left(\sum_{i=1}^4 \left(1 - max(F_{patch}^i \cdot B^{iT}) \right) \right)$$



Paper 1

- Method
 - Prompt Learner



Paper 1

- Method

- Prompt Learner

- 이미지 내의 fine-grained semantic를 활용하고, LLM과 Decoder 간의 의미적 consistency를 유지하기 위해 사용

※ Localization results를 prompt embeddings로 변환함

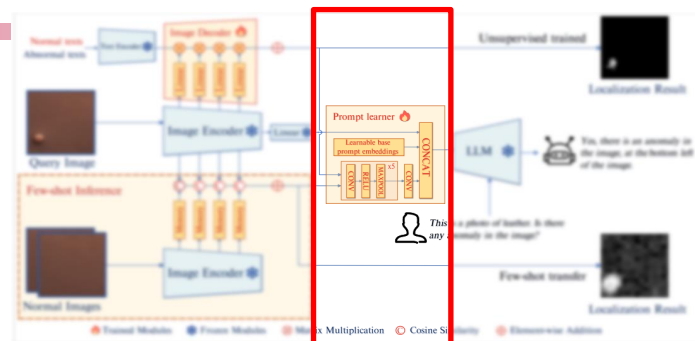
- Decoder의 output과 무관하게 learnable한 base prompt embedding을 prompt learner에 통합하여 추가적인 정보를 제공함

- 모듈 구성

- Learnable base prompt embeddings
 - Convolution neural network

※ Localization results M 을 prompt embeddings로 변환

- 두 prompt embeddings는 concat하여 LLM의 input으로 들어감



Paper 1

- Method

- Data for Image-Text Alignment

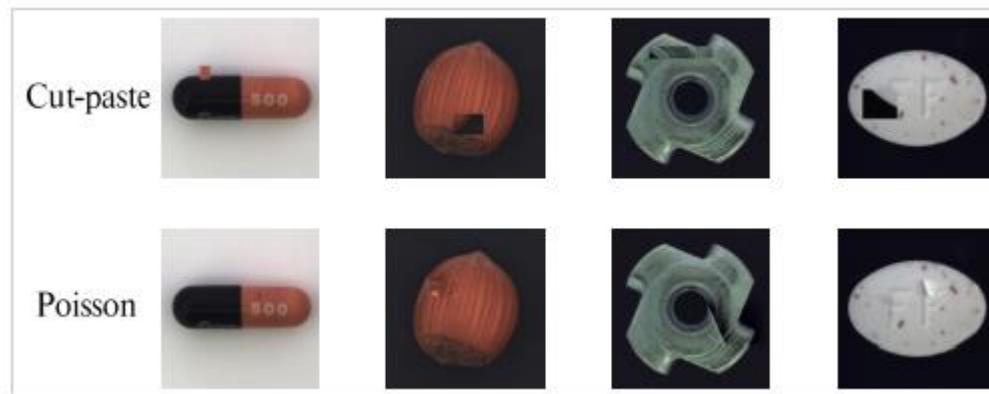
- 정상 이미지에 augmentation을 적용하여 임의의 anomalous 이미지 생성

- ※ Cutpaste, Poisson 기법을 사용함

- ※ Cutpaste는 임의의 구역을 잘라 랜덤하게 붙여넣는 기법으로, 해당 기법을 적용한 이미지는 부자연스러운 경향을 보임

- ※ 위 문제를 해결하기 위해 Poisson 기법을 추가로 적용하여 해결함

- ✓ Poisson은 Poisson partial differential equations를 적용하여 자연스러운 이미지를 만드는데 기여함



Paper 1

- Method

- Question and Answer Content

- LLM의 prompt tuning을 위해 anomalous image에 기반한 textual queries를 생성함
 - 각 query는 2가지 요소로 구성됨

- ⌘ Input image의 description

- ✓ This is a photo of bottle, which should be brown and without any damage, flaw, defect, scratch, hole or broken part

- ⌘ Object내의 anomalies 존재 여부

- ✓ Is there any anomaly in the image?

- LLM은 2가지 유형의 대답을 하도록 구성됨

- ⌘ Anomaly가 있다면 :

- Yes, there is an anomaly in the image, at the bottom left of the image.*

- ⌘ Anomaly가 없다면 :

- No, there are no anomalies in the image.*

Paper 1

- Method

- Detailed Image Description

Class	Image description
Bottle	This is a photo of a bottle for anomaly detection, which should be round and without any damage, flaw, defect, scratch, hole or broken part.
Cable	This is a photo of three cables for anomaly detection, they are green, blue and grey, which cannot be missed or swapped and should be without any damage, flaw, defect, scratch, hole or broken part.
Capsule	This is a photo of a capsule for anomaly detection, which should be black and orange, with print '500' and without any damage, flaw, defect, scratch, hole or broken part.
Carpet	This is a photo of carpet for anomaly detection, which should be without any damage, flaw, defect, scratch, hole or broken part.
Grid	This is a photo of grid for anomaly detection, which should be without any damage, flaw, defect, scratch, hole or broken part.
Hazelnut	This is a photo of a hazelnut for anomaly detection, which should be without any damage, flaw, defect, scratch, hole or broken part.
Leather	This is a photo of leather for anomaly detection, which should be brown with patterns and without any damage, flaw, defect, scratch, hole or broken part.
Metal nut	This is a photo of a metal nut for anomaly detection, which should be without any damage, flaw, defect, scratch, hole or broken part, and shouldn't be flipped.

Paper 1

- Method

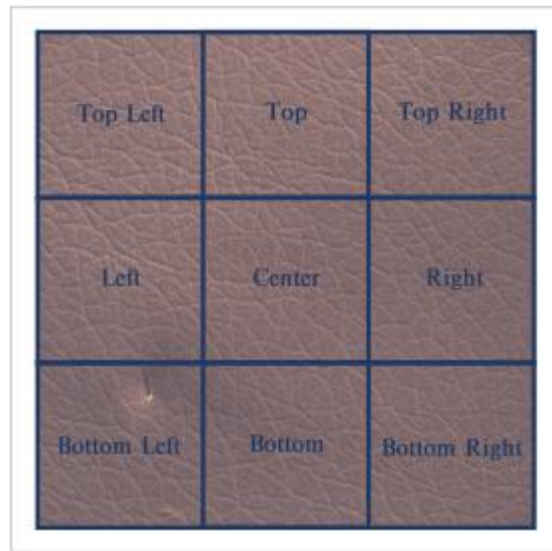
- Question and Answer Content

- Anomaly가 있다면 :

- Yes, there is an anomaly in the image, at the bottom left of the image.*

- 이미지 내의 anomaly 위치를 표현하고자, 이미지를 3x3 grid로 나눠 LLM이 각 구역을 언어로 나타낼 수 있도록 구성함

- 설명형 문장은 LLM이 이미지 내용을 더 잘 이해하도록 돕는 역할을 함



Paper 1

- Method

- Loss functions

- Decoder와 prompt learner를 학습시키기 위해, 3가지 loss를 사용함

- Cross-entropy loss

- ⌘ Language model을 학습시키기 위해 사용

- ⌘ 모델이 생성한 text sequence와 target text sequence간의 disparity를 정량화

- Focal loss

- ⌘ Object detection, semantic segmentation task에서 class imbalance 문제를 해결하기 위해 사용

- ⌘ Cross-entropy loss의 weight distribution을 조절

- Dice loss

- ⌘ Semantic segmentation task에서 자주 사용하며 dice coefficient에 기반함

Paper 1

- Experimental results

- 정량적 결과

Setup	Method	MVTec-AD			VisA		
		Image-AUC	Pixel-AUC	Accuracy	Image-AUC	Pixel-AUC	Accuracy
1-shot	SPADE	81.0 ± 2.0	91.2 ± 0.4	-	79.5 ± 4.0	95.6 ± 0.4	-
	PaDiM	76.6 ± 3.1	89.3 ± 0.9	-	62.8 ± 5.4	89.9 ± 0.8	-
	PatchCore	83.4 ± 3.0	92.0 ± 1.0	-	79.9 ± 2.9	95.4 ± 0.6	-
	WinCLIP	93.1 ± 2.0	95.2 ± 0.5	-	83.8 ± 4.0	96.4 ± 0.4	-
	AnomalyGPT (ours)	94.1 ± 1.1	95.3 ± 0.1	86.1 ± 1.1	87.4 ± 0.8	96.2 ± 0.1	77.4 ± 1.0
2-shot	SPADE	82.9 ± 2.6	92.0 ± 0.3	-	80.7 ± 5.0	96.2 ± 0.4	-
	PaDiM	78.9 ± 3.1	91.3 ± 0.7	-	67.4 ± 5.1	92.0 ± 0.7	-
	PatchCore	86.3 ± 3.3	93.3 ± 0.6	-	81.6 ± 4.0	96.1 ± 0.5	-
	WinCLIP	94.4 ± 1.3	96.0 ± 0.3	-	84.6 ± 2.4	96.8 ± 0.3	-
	AnomalyGPT (ours)	95.5 ± 0.8	95.6 ± 0.2	84.8 ± 0.8	88.6 ± 0.7	96.4 ± 0.1	77.5 ± 0.3
4-shot	SPADE	84.8 ± 2.5	92.7 ± 0.3	-	81.7 ± 3.4	96.6 ± 0.3	-
	PaDiM	80.4 ± 2.5	92.6 ± 0.7	-	72.8 ± 2.9	93.2 ± 0.5	-
	PatchCore	88.8 ± 2.6	94.3 ± 0.5	-	85.3 ± 2.1	96.8 ± 0.3	-
	WinCLIP	95.2 ± 1.3	96.2 ± 0.3	-	87.3 ± 1.8	97.2 ± 0.2	-
	AnomalyGPT (ours)	96.3 ± 0.3	96.2 ± 0.1	85.0 ± 0.3	90.6 ± 0.7	96.7 ± 0.1	77.7 ± 0.4

Paper 1

- Experimental results
 - Question and Answering



User input

Is there any anomaly in the image?



Yes, there is an anomaly in the image, at the bottom of the image.

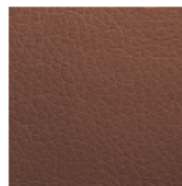


Model output

What's this in the image?



In the image, there is a white pill or medication on a black surface.

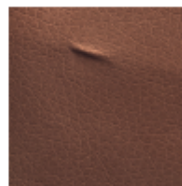


Normal input

Is there any anomaly in the image?



Yes, there is an anomaly in the image, at the top of the image.



Query input

What is the material in the picture?



The material in the picture is leather.



Model output

Tell me something about the leather in the image.



The leather in the image is described as brown and rough, which suggests that it is made from a high-quality, durable leather material.

Paper 1

• Experimental results

▪ Question and Answering

- 다양한 LLM과 비교실험 결과

☼ 이상, 정상 여부

- ✓ 일반적인 LLM은 전혀 다른 답변을 제공하거나 일부만 맞는 답변을 제공

☼ 이미지 설명

- ✓ 일반적인 LLM도 올바른 답변을 제공함

(Normal)		This is a top-view photo of a bottle. Is there any anomaly in the image?	
PandaGPT		Yes, there is an anomaly in the image. The bottle appears to be upside down, which is not a typical orientation for a bottle. This could be due to the angle of the photo or a creative choice to present the image in a unique way.	✗
LLaVA		Yes, there is an anomaly in the image. The bottle is upside down, which is unusual for a bottle that is typically placed upright on a shelf or table.	✗
MiniGPT-4		No, there is no anomaly in the image. It appears to be a clear glass bottle with a black cap and a label that reads "Bordeaux."	✓
 AnomalyGPT		No, there is no anomaly in the image.	✓
		What's the shape and material of the bottle in the image?	
PandaGPT		The bottle in the image is round and made of glass.	✓
LLaVA		The bottle in the image is round and made of glass.	✓
MiniGPT-4		The bottle in the image appears to be made of glass and has a round, cylindrical shape.	✓
 AnomalyGPT		The bottle in the image is round and made of glass.	✓

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

Paper 2

- 기존 LLM 기반 anomaly detection의 한계

- 기존 방법들은 단순히 이상 여부만 예측할 뿐, 왜 이상인지에 대한 설명이 부족함

- Reasoning

- 사용자 입장에서 모델이 이상이라고 판단한 이유를 시각적 근거와 함께 설명해주기를 기대함

- 최신 MLLM인 GPT-4o조차도 anomaly detection은 수행하지만 그 원인에 대한 설명은 정확성이 떨어짐

GPT_4o →



The candle in the top left appears to have a damaged or missing wick, which is unusual compared to the other candles.

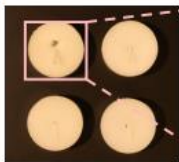
Ours →



Yes, there is an anomaly in the top left candle. It has a small hole or crack near the wick, which is not present in the other candles.



The anomaly could be due to manufacturing defects, such as improper molding or cooling, or it might result from external damage like a puncture.



GT: The top-left candle shows a missing chunk of wax and an irregular indentation near the wick.

Do you see any anomalies in the image?

The candle in the top left appears to have a damaged or missing wick, which is unusual compared to the other candles.

Yes, there is an anomaly in the top left candle. It has a small hole or crack near the wick, which is not present in the other candles.

What is the cause for the anomaly?

The anomaly could be due to manufacturing defects, such as improper molding or cooling, or it might result from external damage like a puncture.

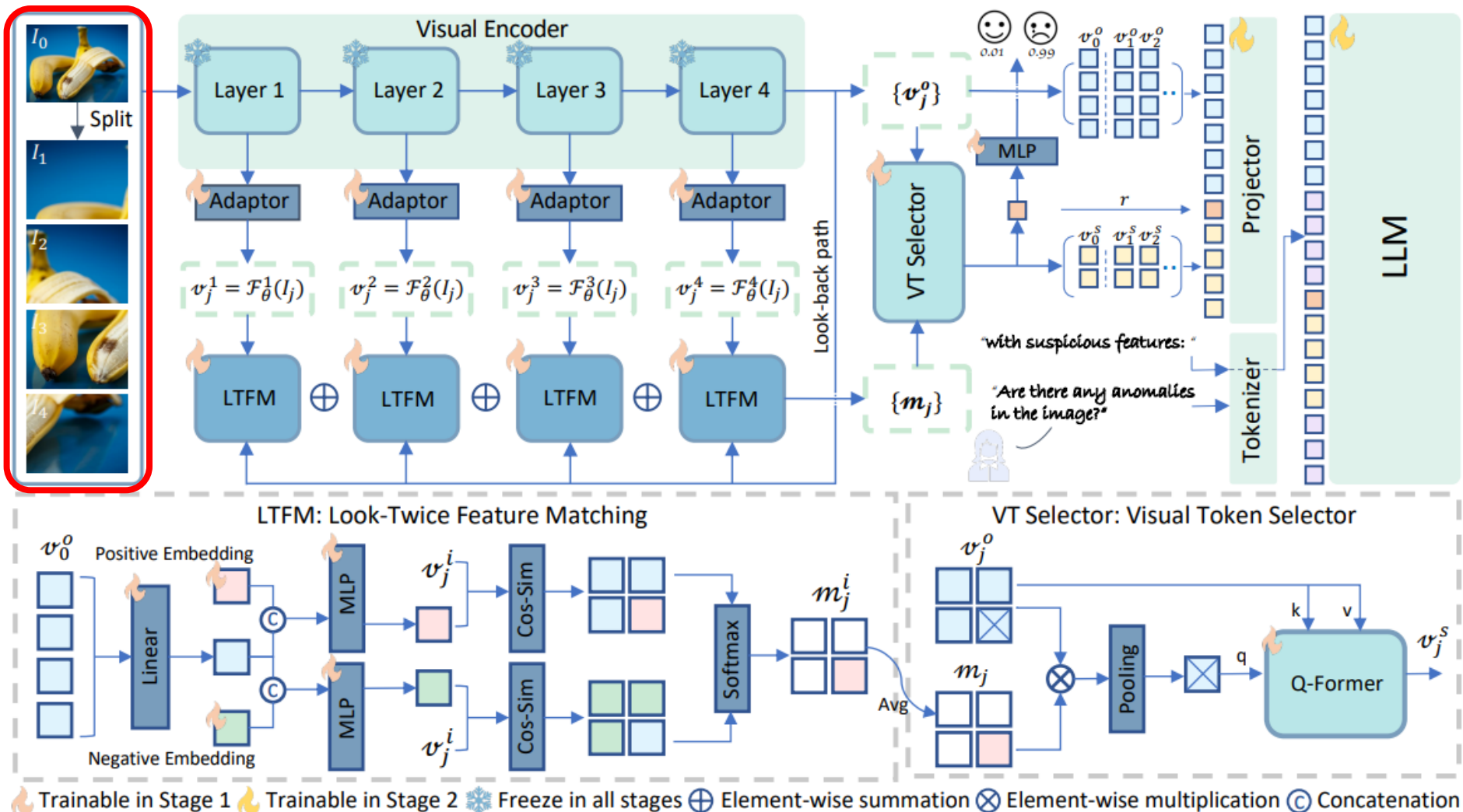
Paper 2

- Preliminary

- MLLM을 처음부터 학습하는 것은 방대한 데이터와 계산 자원이 필요함
 - 즉 visual, textual embedding space를 align하고 robust한 instruction-following capabilities를 develop하는 데 큰 비용이 필요함
- Pretrained MLLM은 일반적인 domain에선 잘 작동하지만, 특화된 domain에선 열등한 성능을 보임
 - 이 문제를 해결하기 위해, auxiliary specialist or expert model을 design해서 pretrained MLLM이 중요한 visual tokens를 잘 선택하고 활용할 수 있도록 유도함
- Base MLLM으로는 LLaVA-OneVision 모델 선택
 - Visual encoder : 이미지로부터 visual 정보를 추출
 - Projector : visual feature space와 textual feature space를 align
 - LLM : textual instruction 처리와 복잡한 추론
- CLIP은 input image resolution이 고정되어있기 때문에, LLaVA-OneVision은 AnyRes와 pooling 기법을 사용하여 input image resolution을 up scaling함
 - High resolution image를 prototype수의 crop으로 나뉘며, visual encoder는 image crop을 독립적으로 처리한 후 spatial pooling 적용함

Paper 2

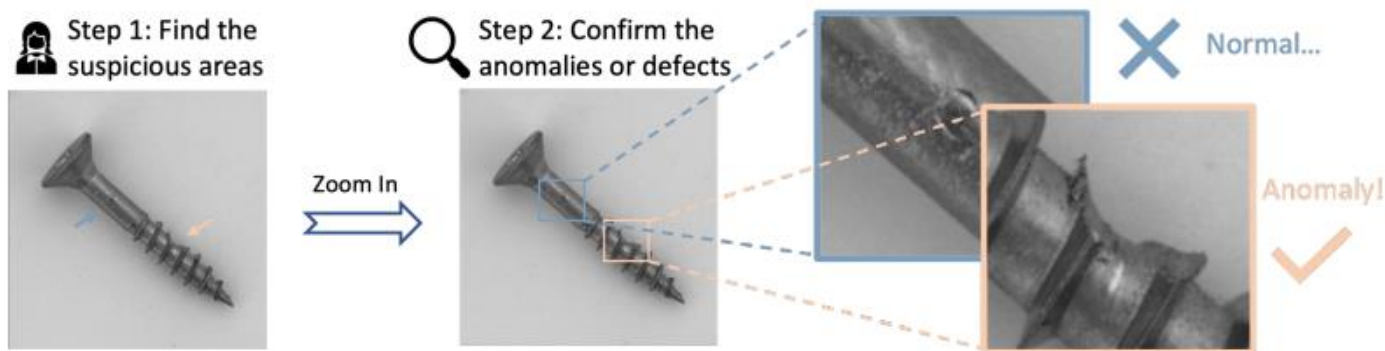
• Overview



Paper 2

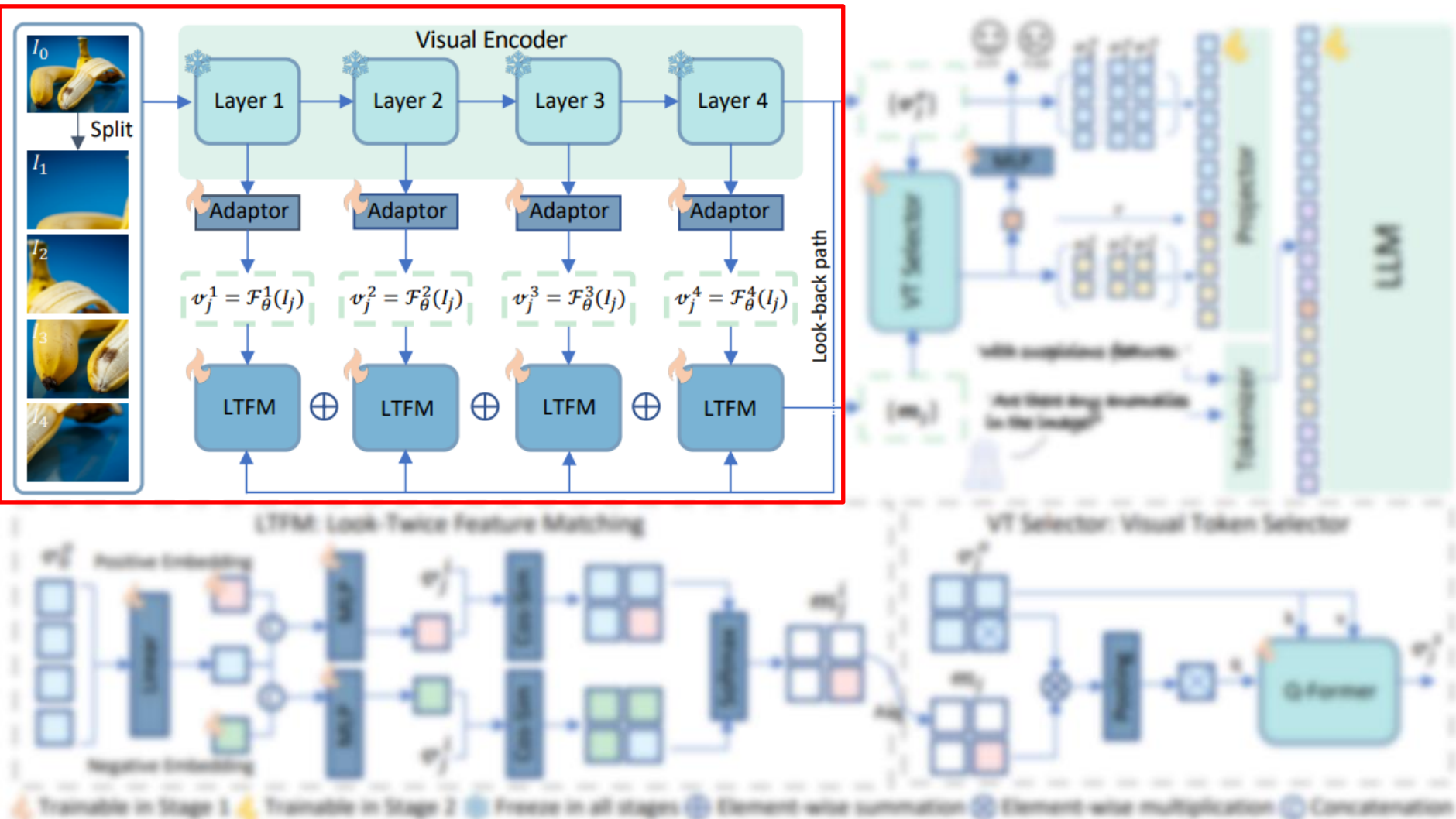
• Overview

- Human involvement를 피하고 complexity를 줄이기 위해, text encoder를 제거하고 visual model 자체가 의심스러운 class나 object에 대한 정보를 해석하도록 함
 - Visual feature는 look-back path에서 global description을 제공하는 데 활용됨
 - Multi-level feature와 global embedding을 바탕으로, LTFM 모듈은 의심스러운 token을 인식하고 localization 수행
 - 사람의 시각적 inspection 과정에서 의심스러운 객체나 영역을 먼저 식별하고 그 후에 자세히 살펴보는 행동에서 영감을 받아, VT selector 모듈을 설계함
- ※ 중요한 visual token을 aggregate하고, LLM이 instruction을 처리하는데 도움
- ✓무수히 많은 visual token 중 중요한 token을 찾도록 도움



Paper 2

- Method



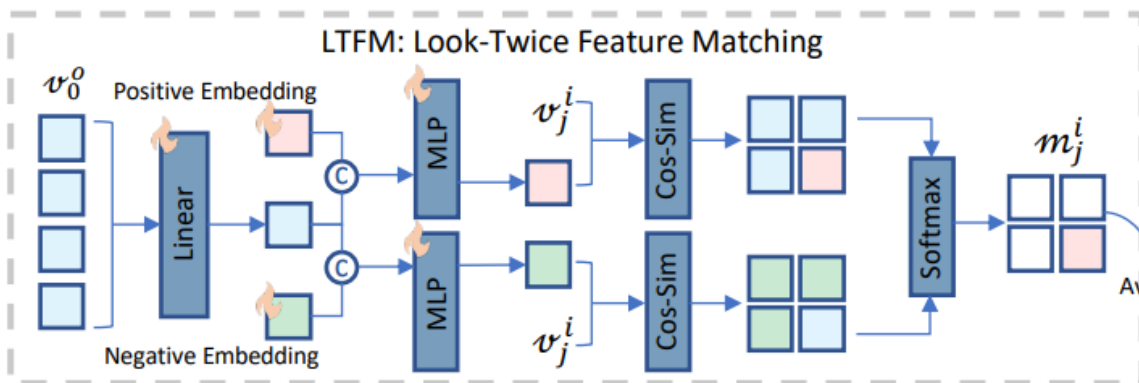
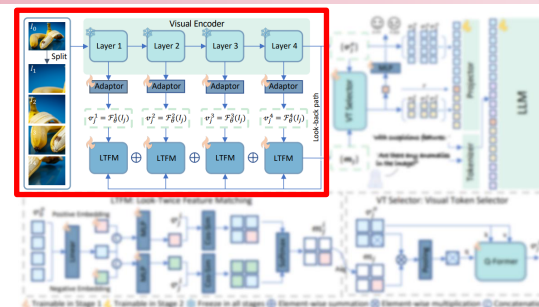
Paper 2

• Method

▪ Look-Twice Feature Matching (LTFM)

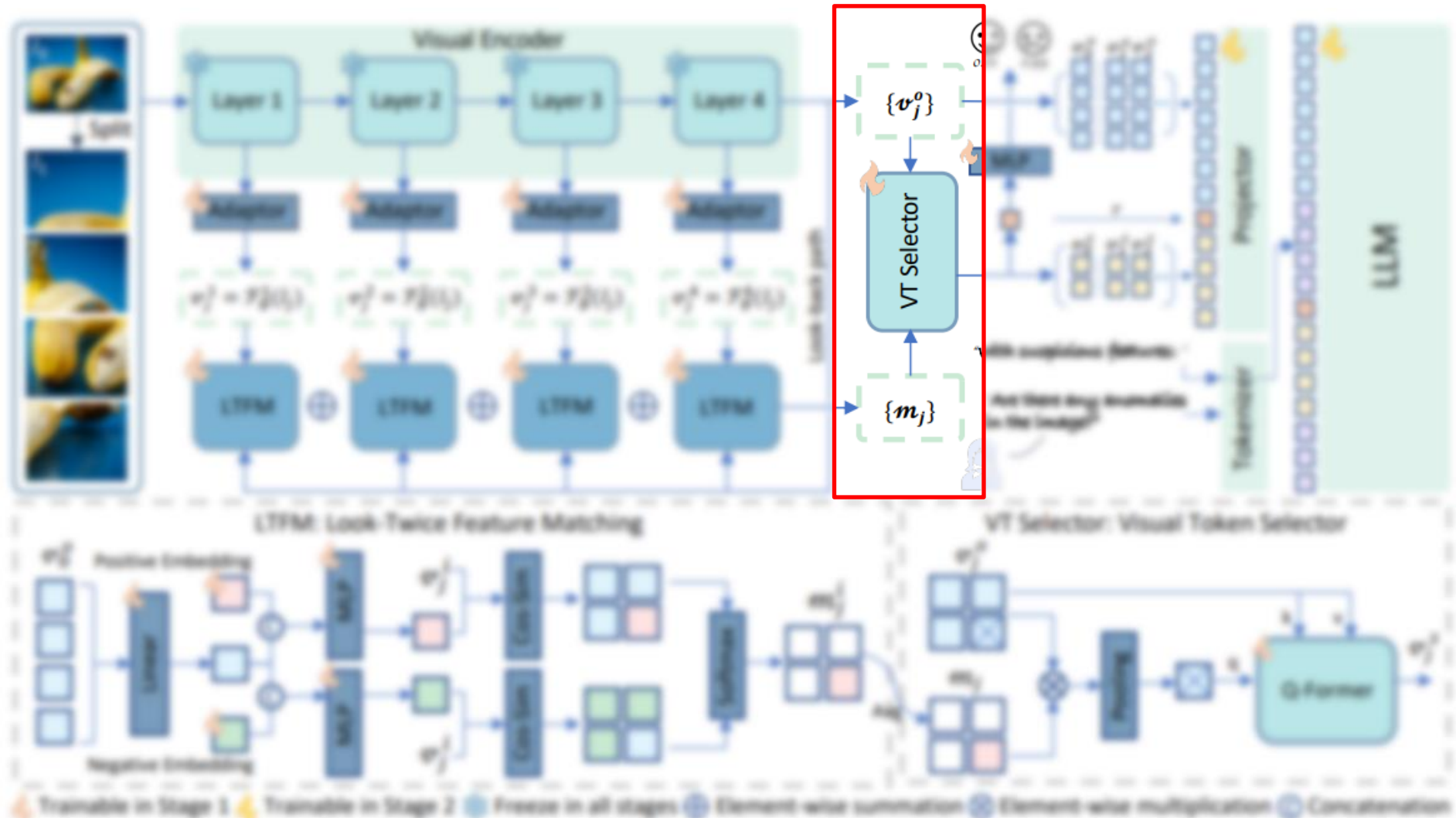
- Look-back path에서 제공된 global object information(v_j^i)을 기반으로, class-awareness abnormality description을 생성함
 - ※ v_j^i 의 중요한 token을 select하고 fuse하며, Positive/Negative embedding과 concat함
 - ※ 2개의 MLP를 통해 normality/abnormality description을 생성하도록 함
- Cosine similarity와 softmax 연산을 결합해서 각 patch token이 anomalous pattern에 속할 가능성을 계산함
- 이후 모든 map은 low-level ~ high-level까지 token 중요도를 포착하기 위해 평균화됨

▪ Visual feature는 forward, look-back path 경로에서 2번 사용됨



Paper 2

- Method

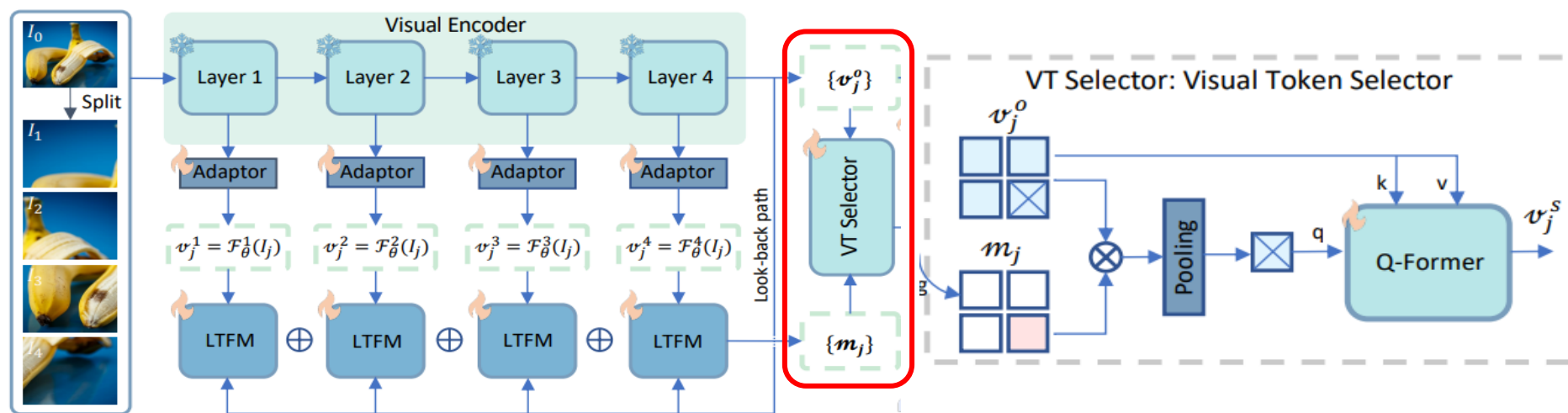


Paper 2

• Method

▪ Visual Token Selector (VT selector)

- Visual token은 풍부한 visual information을 제공하지만, LLM은 specific task에 adapt시키기 위해선 가장 유용한 information을 선택해야함
- LLM이 해당 domain에 충분한 지식이 없는 경우, token 선택 과정은 복잡해짐
따라서 어느 token이 중요한지를 아는 expert를 도입함



Paper 2

• Method

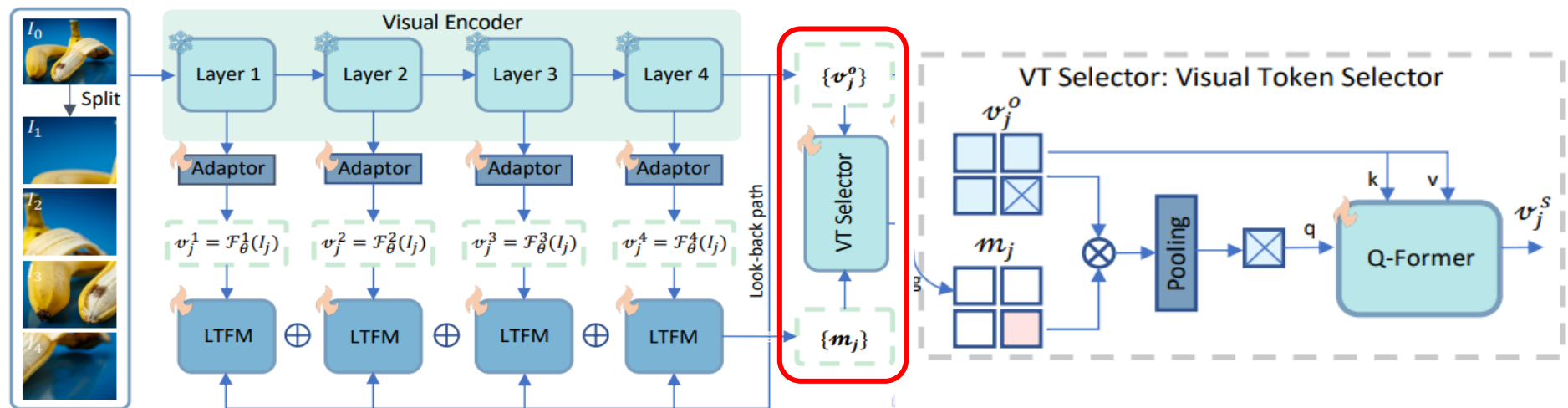
▪ Visual Token Selector (VT selector)

- Image에 대한 visual token과 LTFM에서 얻은 중요도 map이 주어졌을 때, 의심스러운 token은 두 tensor의 직접 곱하여 강조됨

⚡ 정상 token은 0으로 scaling

⚡ 이상 token은 유지됨

⚡ 이후 spatial average pooling이 token 수를 줄이기 위해 적용됨



Paper 2

• Method

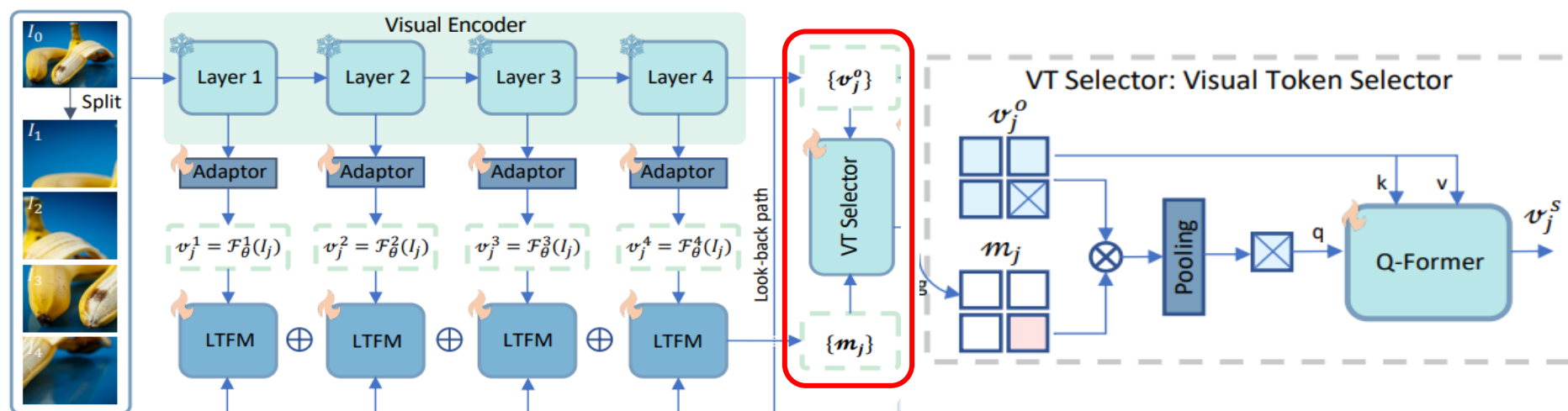
▪ Visual Token Selector (VT selector)

- 이후 Q-Former를 통해 가장 의심스러운 visual token을 aggregate함

※ Q-Former는 BLIP-2에서 처음 제안된 모듈

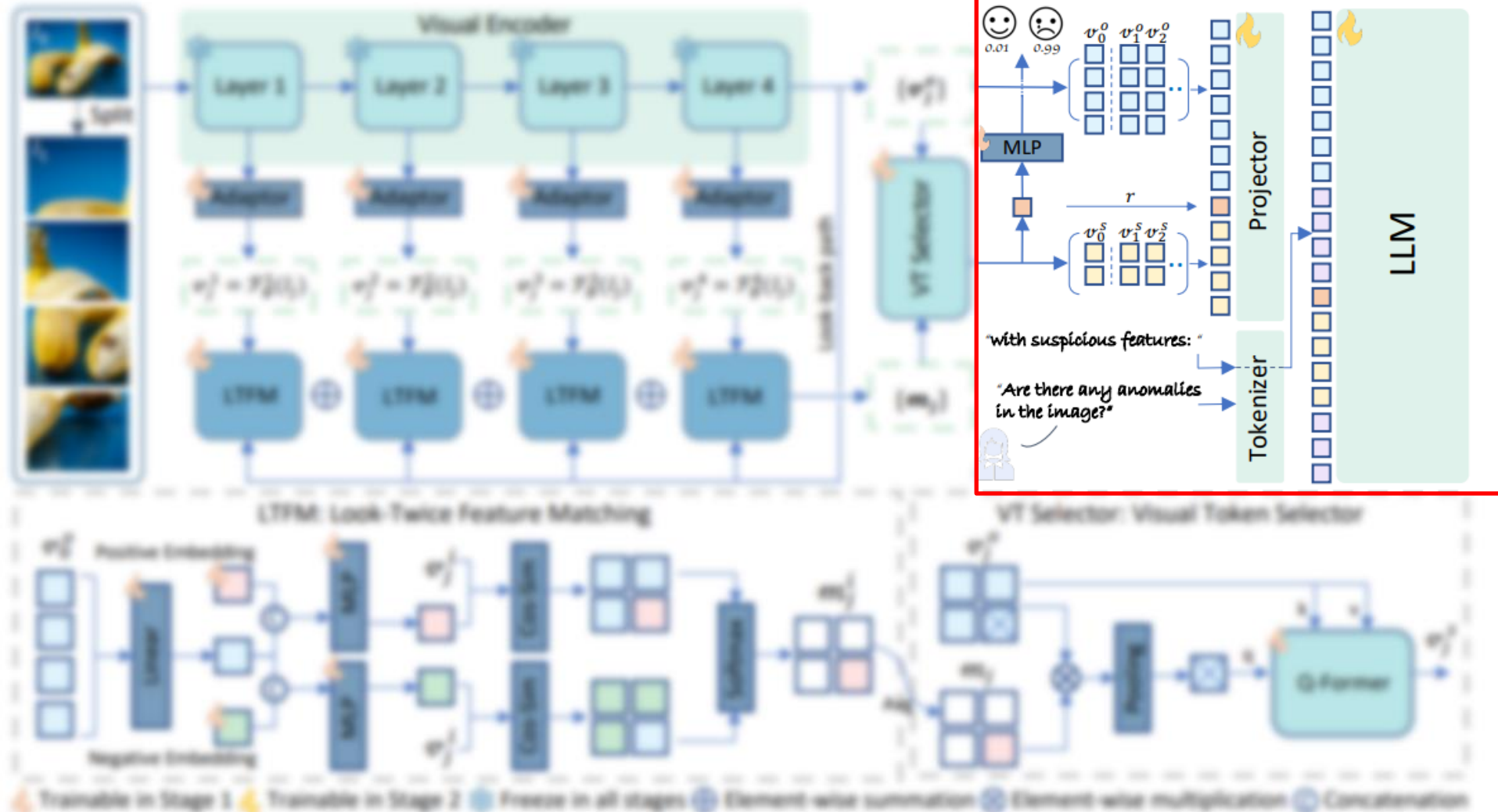
※ Query 기반으로 중요한 정보를 뽑아내는 구조임

- Pooling된 token(q)을 기반으로 원본 visual feature (v_j^i)에서 가장 관련있는 token만 aggregate함



Paper 2

- Inference



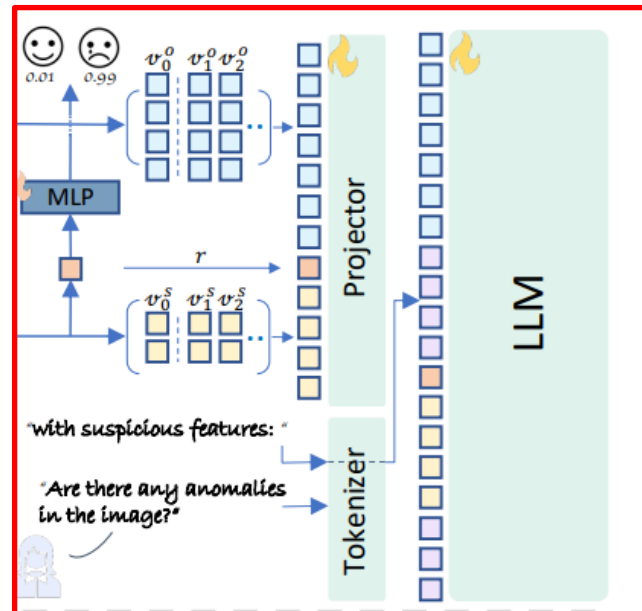
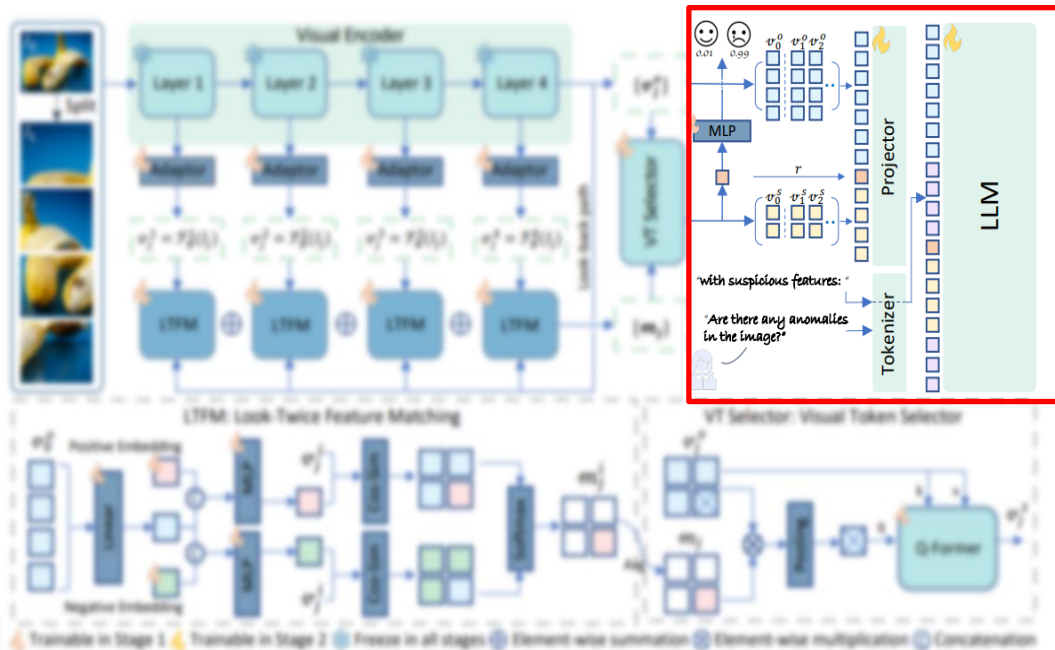
Paper 2

• Inference

▪ Anomaly Prediction

- 모든 crop된 이미지의 anomaly information를 중요도 map에 따라 가중 평균하는 방식으로 aggregate함
- Aggregate된 정보($r(I)$)에 대하여 MLP를 통과시켜 normal/abnormal semantics를 식별함
- Sigmoid로 anomaly score 계산

$$\text{score}(I) = \text{Sigmoid}(\mathcal{G}^o(\mathbf{r}(I)))$$



Paper 2

- Inference

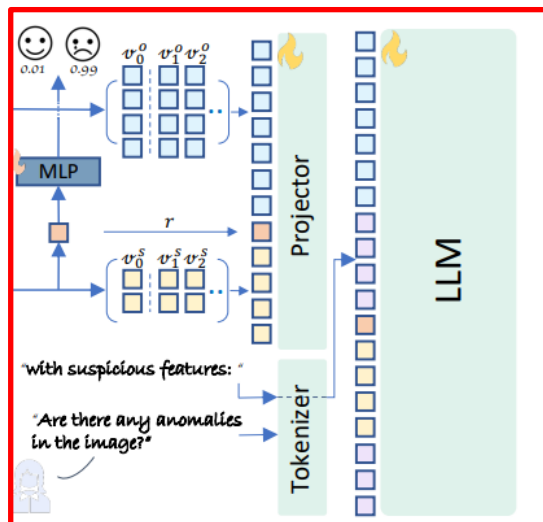
- Text Generation

- VT Selector로부터 얻은 $(r(I), v_j^s)$ 와 v_j^i 를 concat하여 LLM에 전달하는 대신
두 token 사이에 “<adv> suspicious feature:” 라는 indication prompt를 삽입

- ⌘ LLM이 anomaly-related instruction를 처리할 때 선택된 token을 강조할 수 있도록 함

- ⌘ “<adv>”는 {highly, moderately, slightly} 중에서 선택되며, 이는 score(I)와 미리 정의된 threshold에 따라 결정됨

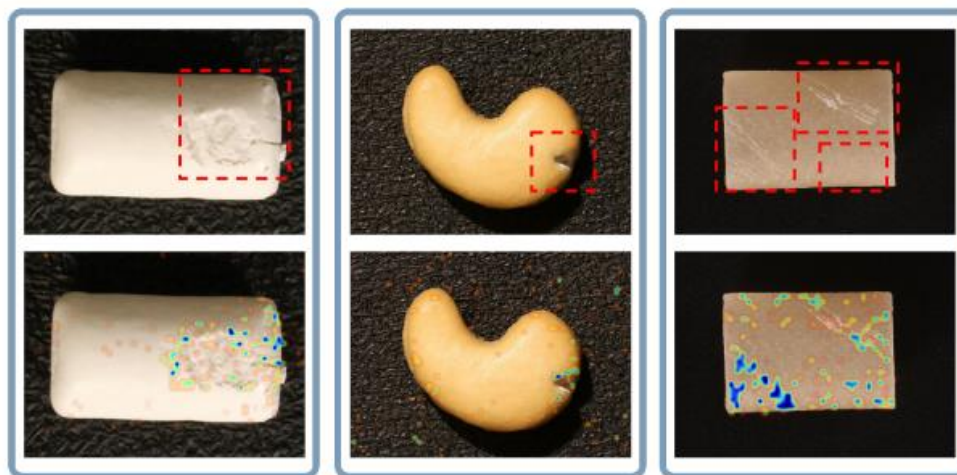
- ✓ Input image I가 높은 anomaly를 갖는 경우, LLM은 선택된 token에 더 큰 중요도를 부여함



Paper 2

- Experiments

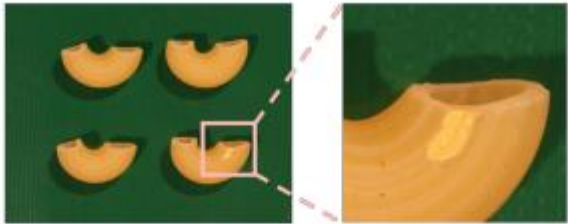
Model	Industrial Defects						Medical Anomalies			Average
	MVTec AD	VisA	AITEX	ELPV	BTAD	MPDD	BrainMRI	HeadCT	Br35H	
CLIP [73]	74.1	66.4	71.0	59.2	34.5	54.3	73.9	56.5	78.4	63.1
CoOp [108]	88.8	62.8	66.2	73.0	66.8	55.1	61.3	78.4	86.0	70.9
WinCLIP [38]	91.8	78.8	73.0	74.0	68.2	63.6	92.6	90.0	80.5	79.2
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	79.7	<u>88.6</u>	76.0	94.8	91.4	97.7	<u>85.3</u>
Ours	94.0	91.1	72.0	83.0	89.0	81.7	<u>93.9</u>	97.6	<u>95.5</u>	88.6



Paper 2

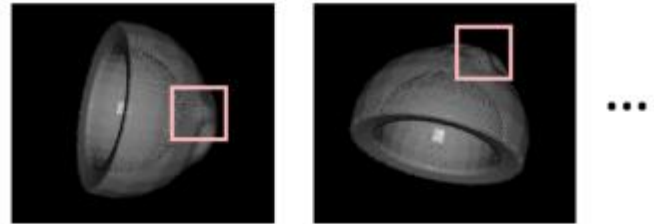
- Experiments

Macaroni Example: Yellowish Spot



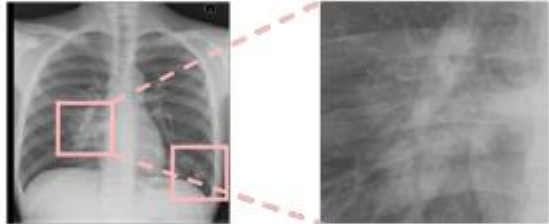
User	Do you see any anomalies in the image?
GPT-4o	The image shows four pieces of elbow macaroni on a green background. The anomaly is that the macaroni pieces are not whole; they are cut in half.
LLaVA-OV	The image shows four pieces of pasta, specifically macaroni shells, arranged on a green textured surface. The pasta appears to be uniformly colored and shaped, with no visible defects or anomalies.
Ours-7B	Yes, there is an anomaly in the image. The bottom right pasta piece has a noticeable yellowish discoloration on its surface.

3D Example: Bulge



User	Is there any anomaly present in the bowl?
Ours-7B	Yes, there is a bulge on the side of the bowl, which disrupts its otherwise smooth and symmetrical surface.

Medical Example: Pneumonia



User	What is the most likely diagnosis for this patient?
Ours-7B	Based on the chest X-ray, the most likely diagnosis for this patient is pneumonia. The presence of bilateral patchy opacities in the lower lung fields is a common finding in pneumonia, which is an infection that causes inflammation in the air sacs of the lungs.
User	What should the patient do next?
Ours-7B	The patient should consult a healthcare professional for further evaluation and appropriate treatment. This may include antibiotics to treat the infection, supportive care such as oxygen therapy, and monitoring of the patient's condition.

Thank You