

2025 겨울 세미나

2025.01.17



Sogang University

Vision & Display Systems Lab, Dept. of Electronic Engineering



Presented By

이준호

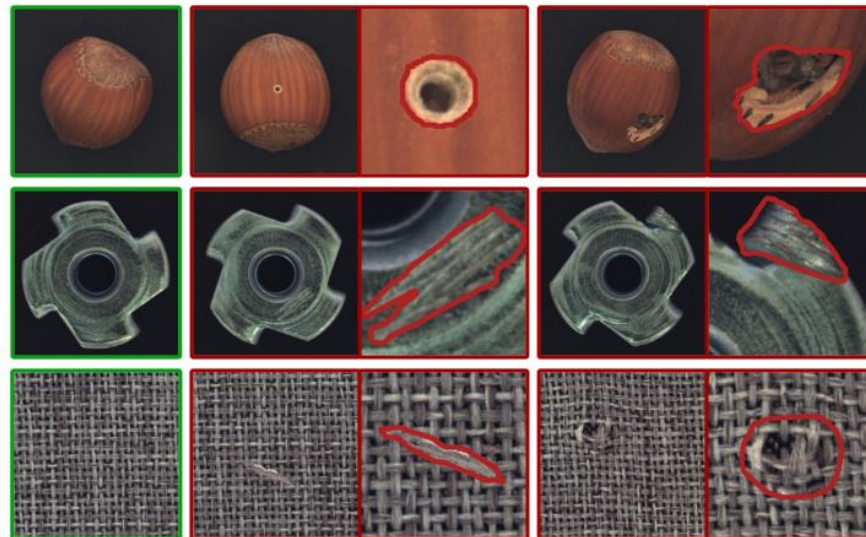
Outline

- Background
 - Anomaly detection
 - Unsupervised learning
 - Zero-shot anomaly detection
- Paper 1
 - AnomalyCLIP: Object-agnostic Prompt Learning for Zero-shot Anomaly Detection (2024-ICLR)
- Paper 2
 - AdaCLIP: Adapting CLIP with Hybrid Learnable Prompts for Zero-Shot Anomaly Detection (2024-ECCV)

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

- Unsupervised learning in anomaly detection

- Train

- Normal 데이터로만 학습

- Test

- Normal과 Anomal 데이터가 섞여 있으며, 어떤 데이터가 normal인지 명시되어 있지 않음

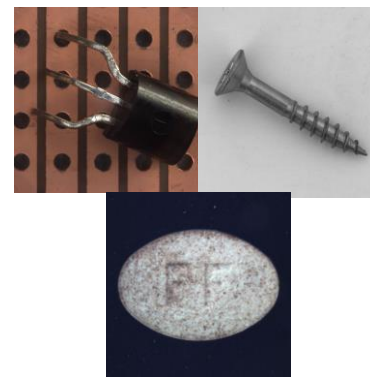
- 학습된 정보를 기반으로 normal, anomal 데이터 classification, segmentation

Train



Model

Test



Model

Background

- Zero-shot learning in anomaly detection

- Train

- 외부 데이터셋으로만 학습
- 외부 데이터셋에는 target 데이터셋이 전혀 속해있지 않음

- Test

- 학습한 정보를 기반으로 target 데이터셋 inference

Train



Model

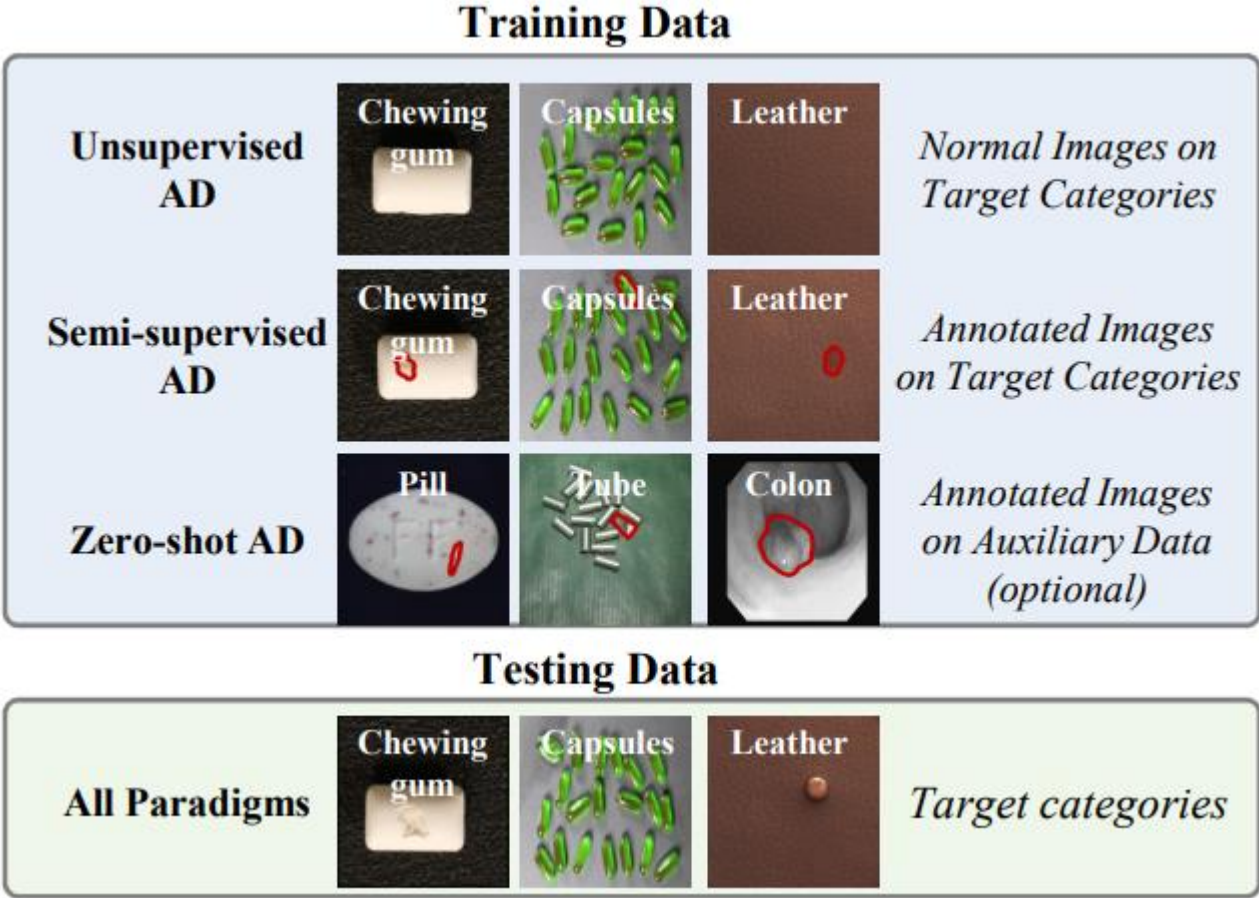
Test



Model

Background

- Zero-shot learning in anomaly detection



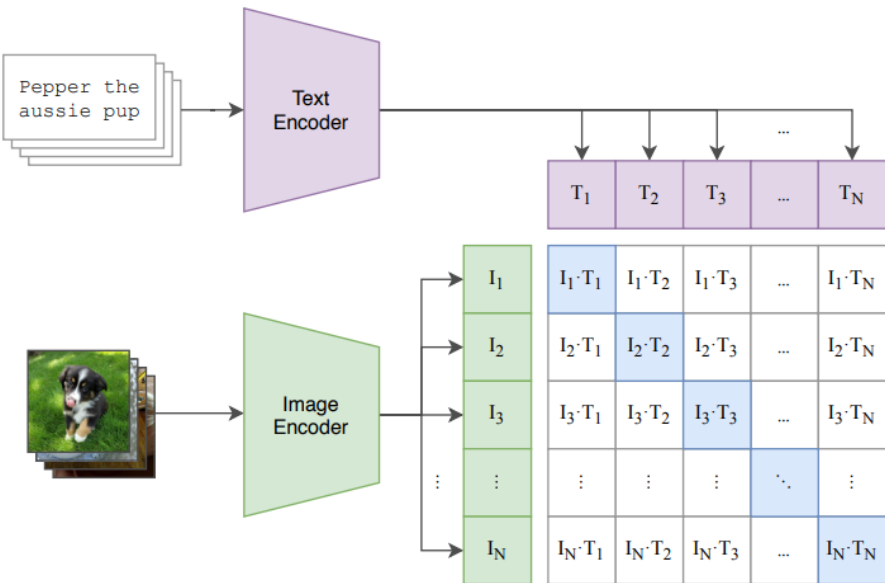
Background

- Zero-shot learning in anomaly detection

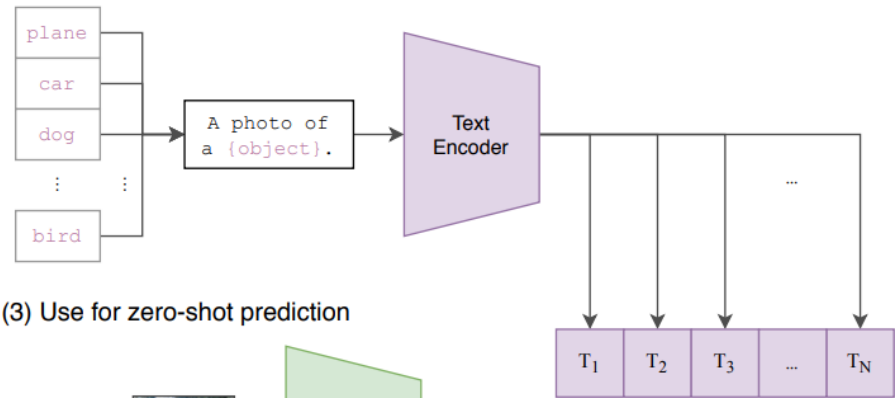
- CLIP 모델

- Text encoder와 Visual encoder로 구성
 - 두 encoder의 output을 fusion하여 anomaly scoring
 - ImageNet으로 pretrain된 weight를 사용하여 anomaly detection task에서 feature 추출

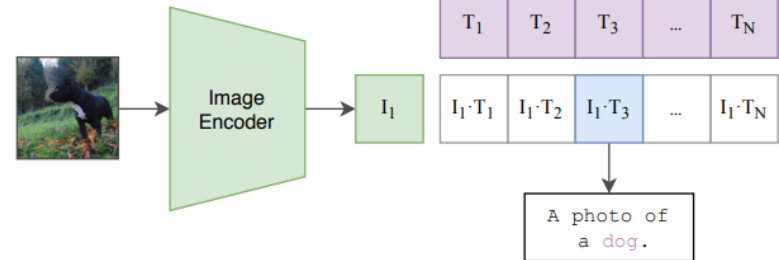
(1) Contrastive pre-training



(2) Create dataset classifier from label text



(3) Use for zero-shot prediction



AnomalyCLIP: Object-agnostic Prompt Learning for Zero-shot Anomaly Detection

Paper 1

- Contribution

- Problem 1

- ImageNet으로 pretrain된 CLIP의 encoder를 사용함

- ※ Normality/abnormality에 집중하기 보다, object의 semantics에 집중하는 경향이 있음

- Solution 1

- Object-agnostic text prompt를 사용하여 normality/abnormality에 집중하도록 유도

- Problem 2

- 현재 prompting 접근법들은 수동으로 정의된 text prompt 또는 learnable prompt를 사용하지만, global 정보에 의존하는 prompt를 생성하는 경향이 있어, local 정보에 집중하지 못함

- Solution 2

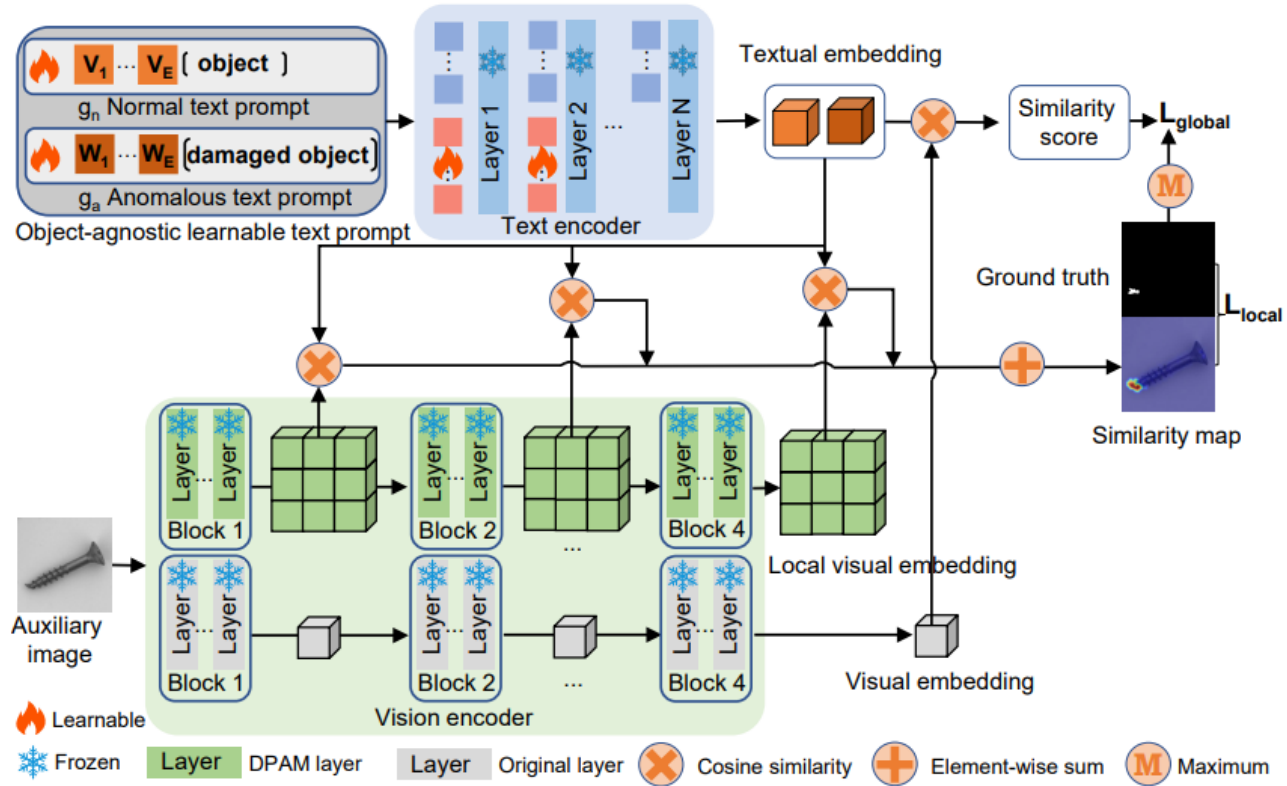
- Normality/abnormality를 대표하는 간단하고 보편적인 learnable prompt template 설계

- Image/pixel-level loss 함수를 활용해서 외부 데이터에 대한 global/local normality/abnormality 학습

Paper 1

- Overview

- Object-agnostic text prompt templates (Problem 1)
- Learning generic abnormality and normality prompts (Problem 2)



Paper 1

- Object-agnostic text prompt templates

- Motivation

- 일반적인 CLIP의 text prompt templates

- ※ A photo of a [class name]

- ※ 위 template은 object semantics에 focus하기 때문에 anomaly/normality에 대한 text embedding을 생성하기 힘들

- Anomaly를 더 잘 표현하기 위해 prompt에 defect 유형을 추가함

- ※ A photo of a [class name] **with scratches**

- ※ 하지만 defect 유형은 다양하고 defect 정보가 없기 때문에 모든 유형을 추가하는 건 불가능

- 따라서 아래와 같이 범용적으로 사용할 수 있는 template 설계

- ※ A photo of a [**damaged**] [class name]

- ※ 하지만 generic anomaly-discriminativg textual embedding을 생성하기 힘들

- ※ CLIP이 image의 normal/abnormality보다 object semantics에 집중하기 때문

Paper 1

- Object-agnostic text prompt templates

- Solution

- Learnable text prompt template 설계 + auxiliary dataset을 사용하여 fine tuning

- ※ Fine tuning과정에서 learnable template은 global, local 정보를 모두 다룰 수 있으며, normality/abnormality를 더 잘 표현할 수 있음

- Motivation

- 다양한 데이터셋을 다루기 때문에 object semantics의 variation이 다양할 수 있음

- ※ 하지만 defect의 근본적인 패턴은 유사할 수 있음

- ✓ 따라서 object semantics에 대한 정보는 불필요함

- Object semantics 정보를 제외하면 abnormality 자체에 집중할 수 있다고 예측

- Solution

- Object-agnostic prompt learning을 도입하여 object semantics와 무관하게 normality/abnormality 학습

- A photo of a [damaged] [object]

Paper 1

- Object-agnostic text prompt templates
 - Summary
 - A photo of a [class name]
 - A photo of a [class name] **with scratches**
 - A photo of a [**damaged**] [class name]
 - A photo of a [**damaged**] [**object**]

Paper 1

- Learning generic abnormality and normality prompts
 - Object-agnostic text prompts를 효과적으로 학습하기 위해 제안
 - Joint optimization approach
 - Global/Local 관점에서 normality/abnormality을 학습할 수 있음

※ Global context optimization

- ✓ Object-agnostic textual embedding이 다양한 object에 대한 global visual embedding과 일치하도록 유도함
- ✓ Cross entropy loss 사용
 - Text, visual embedding 간의 cosine similarity를 일치시키기 위해 사용

※ Local context optimization

- ✓ Visual encoder의 M개의 intermediate layer에서 추출된 local abnormal region에 object-agnostic text prompt가 focus하도록 유도
- ✓ Focal, Dice loss를 사용
 - Class imbalanced 문제를 해결하기 위해 focal loss 사용
 - 모델이 정확한 decision boundary를 형성하기 위해 dice loss 사용

$$L_{total} = L_{global} + \lambda \sum_{M_c \in \mathcal{M}} L_{local}^{M_c}$$

Paper 1

- Refinement of the textual space

- Motivation

- Discriminative textual space를 더 정확히 형성하기 위해 제안
 - CLIP의 text encoder에 additional learnable token embedding을 추가하여 original textual space가 refine되도록 유도
 - ※ Text prompt tuning 역할

- Method

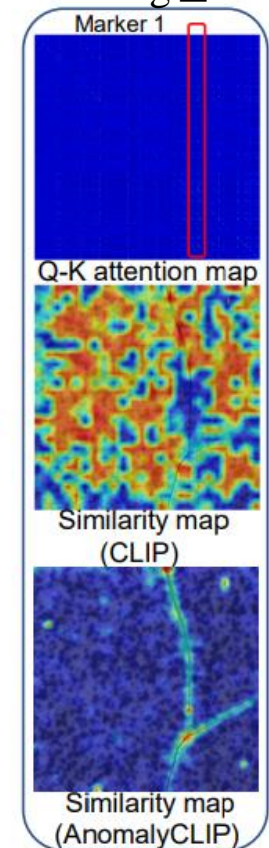
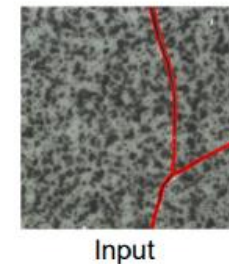
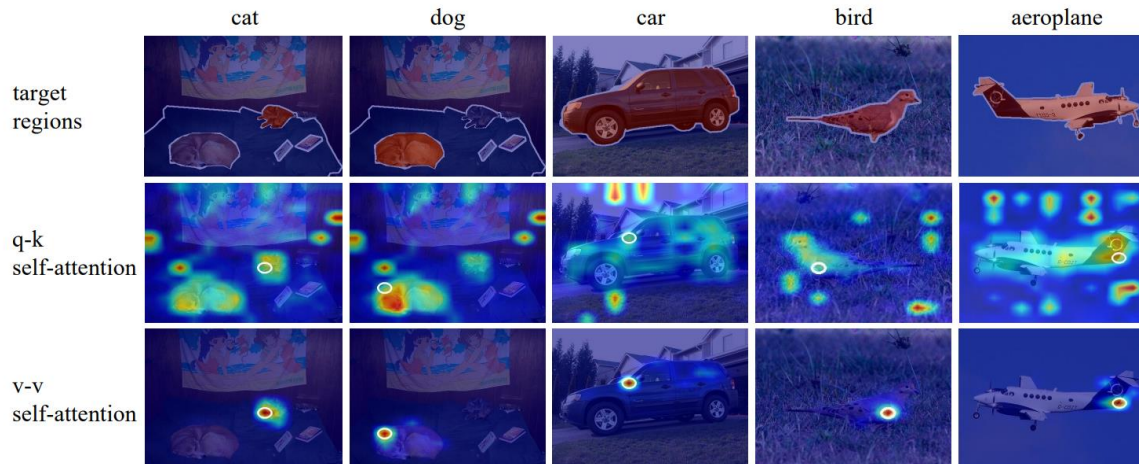
- Pretrained된 CLIP의 m-th layer에 random init된 learnable token embedding t'_m 을 추가함
 - Learnable token embedding t'_m 과 원래 CLIP의 token embedding을 concat(channel)
 - Concat된 token embedding을 m+1-th layer에 전달하여 output r'_{m+1} 과 t_{m+1} 을 얻음
 - ※ r'_{m+1} 을 제거하고 새로 random init된 t'_{m+1} 설정
 - ※ r'_{m+1} 이 제거되더라도 self-attention 덕분에 update된 gradient는 t'_m 을 최적화하기 위해 backprop됨
 - 위 과정을 다음 layer까지 반복

Paper 1

• Refinement of the local visual space

▪ Motivation

- CLIP은 global object semantics에 focus하도록 pretrain 됨
- CLIP에서 사용하는 contrastive loss는 classification task를 위한 global embedding을 생성하도록 유도함
- CLIP visual encoder의 self-attention을 거쳐 나온 attention map은 빨간 블록에서 강조된 특정 token에 focus함
- 이 token들은 global object recognition에 효과적일 수 있지만, local visual semantics에 악영향을 줌

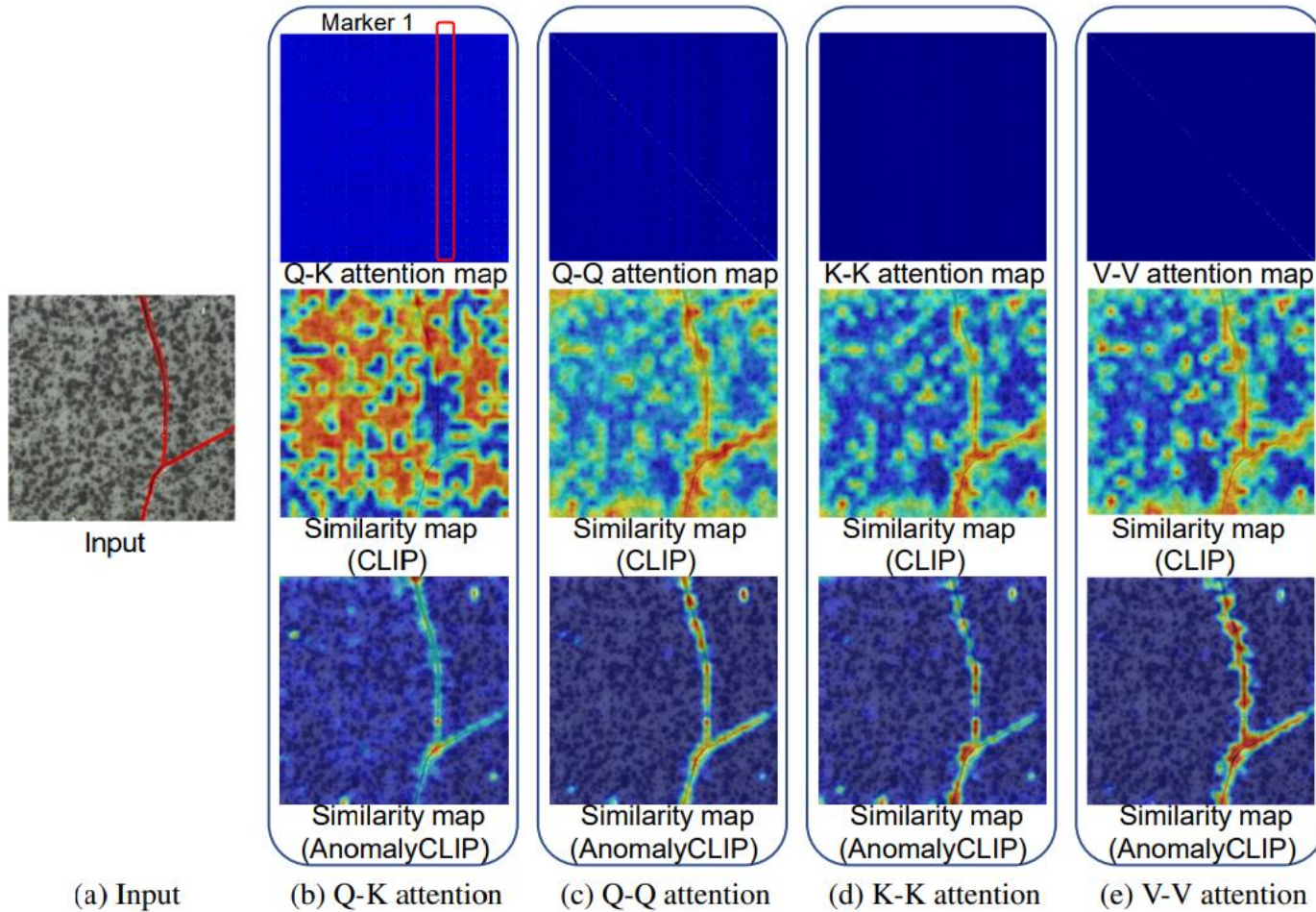


(a) Input

(b) Q-K attention

Paper 1

- Refinement of the local visual space



Paper 1

- Experiments

- 정량적 결과

- Industrial domain

Task	Category	Datasets	$ \mathcal{C} $	CLIP	CLIP-AC	WinCLIP	VAND	CoOp	AnomalyCLIP
Image-level (AUROC, AP)	Obj & texture	MVTec AD	15	(74.1, 87.6)	(71.5, 86.4)	(91.8, 96.5) [†]	(86.1, 93.5) [†]	(88.8, 94.8)	(91.5, 96.2)
		VisA	12	(66.4, 71.5)	(65.0, 70.1)	(78.1, 81.2) [†]	(78.0, 81.4) [†]	(62.8, 68.1)	(82.1, 85.4)
	Obj	MPDD	6	(54.3, 65.4)	(56.2, 66.0)	(63.6, 69.9)	(73.0, 80.2)	(55.1, 64.2)	(77.0, 82.0)
		BTAD	3	(34.5, 52.5)	(51.0, 62.1)	(68.2, 70.9)	(73.6, 68.6)	(66.8, 77.4)	(88.3, 87.3)
		SDD	1	(65.7, 45.2)	(65.2, 45.7)	(84.3, 77.4)	(79.8, 71.4)	(74.9, 65.1)	(84.7, 80.0)
	Texture	DAGM	10	(79.6, 59.0)	(82.5, 63.7)	(91.8, 79.5)	(94.4, 83.8)	(87.5, 74.6)	(97.5, 92.3)
		DTD-Synthetic	12	(71.6, 85.7)	(66.8, 83.2)	(93.2, 92.6)	(86.4, 95.0)	(-, -)	(93.5, 97.0)
	Pixel-level (AUROC, PRO)	Obj & texture	MVTec AD	15	(38.4, 11.3)	(38.2, 11.6)	(85.1, 64.6) [†]	(87.6, 44.0) [†]	(33.3, 6.7)
VisA			12	(46.6, 14.8)	(47.8, 17.3)	(79.6, 56.8) [†]	(94.2, 86.8) [†]	(24.2, 3.8)	(95.5, 87.0)
Obj		MPDD	6	(62.1, 33.0)	(58.7, 29.1)	(76.4, 48.9)	(94.1, 83.2)	(15.4, 2.3)	(96.5, 88.7)
		BTAD	3	(30.6, 4.4)	(32.8, 8.3)	(72.7, 27.3)	(60.8, 25.0)	(28.6, 3.8)	(94.2, 74.8)
		SDD	1	(39.0, 8.9)	(32.5, 5.8)	(68.8, 24.2)	(79.8, 65.1)	(28.9, 7.1)	(90.6, 67.8)
Texture		DAGM	10	(28.2, 2.9)	(32.7, 4.8)	(87.6, 65.7)	(82.4, 66.2)	(17.5, 2.1)	(95.6, 91.0)
		DTD-Synthetic	12	(33.9, 12.5)	(23.7, 5.5)	(83.9, 57.8)	(95.3, 86.9)	(-, -)	(97.9, 92.3)

Paper 1

- Experiments

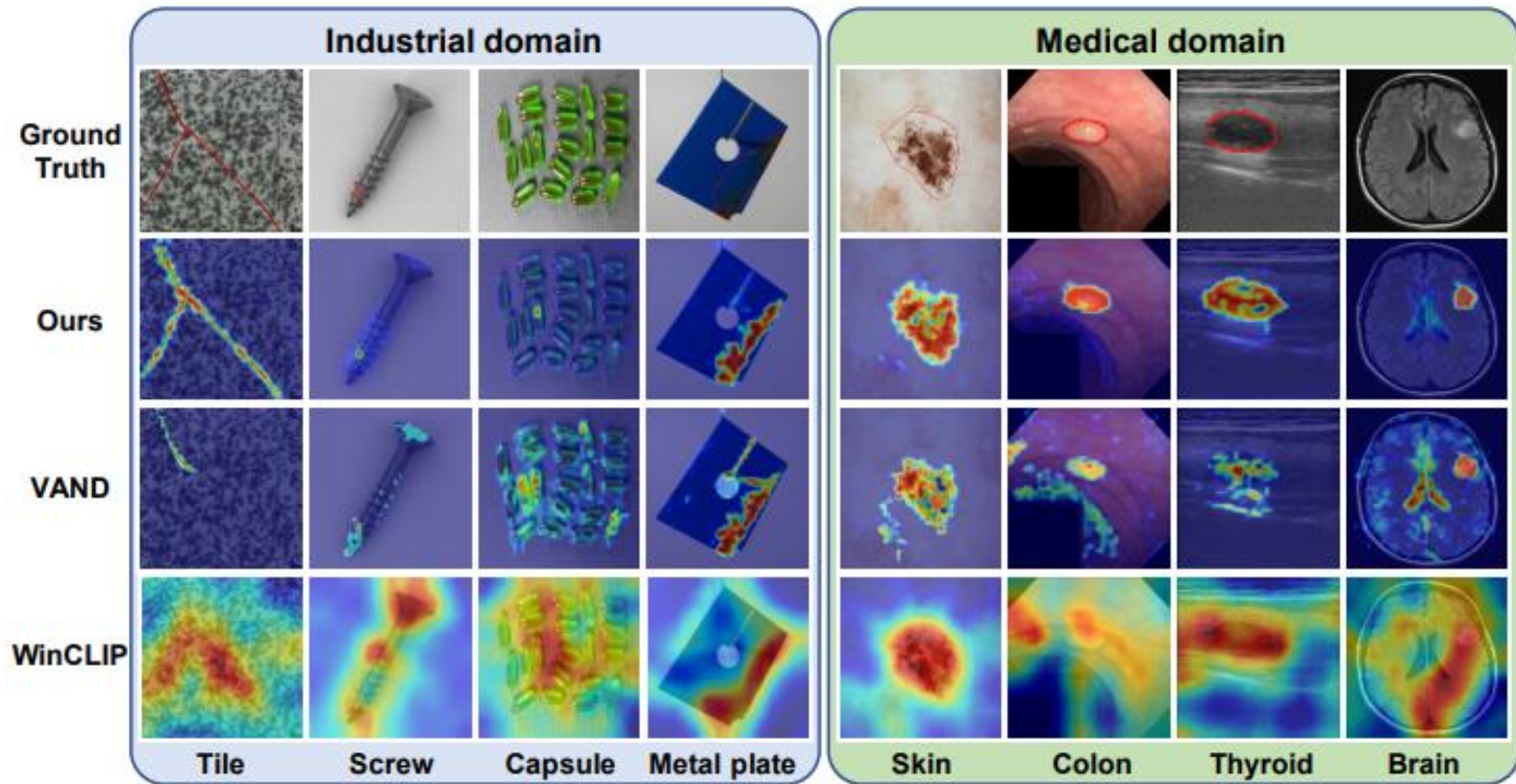
- 정량적 결과

- Medical domain

Task	Category	Datasets	$ C $	CLIP	CLIP-AC	WinCLIP	VAND	CoOp	AnomalyCLIP
Image-level (AUROC, AP)	Brain	HeadCT	1	(56.5, 58.4)	(60.0, 60.7)	(81.8, 80.2)	(89.1, 89.4)	(78.4, 78.8)	(93.4, 91.6)
		BrainMRI	1	(73.9, 81.7)	(80.6, 86.4)	(86.6, 91.5)	(89.3, 90.9)	(61.3, 44.9)	(90.3, 92.2)
		Br35H	1	(78.4, 78.8)	(82.7, 81.3)	(80.5, 82.2)	(93.1, 92.9)	(86.0, 87.5)	(94.6, 94.7)
	Chest	COVID-19	1	(73.7, 42.4)	(75.0, 45.9)	(66.4, 42.9)	(15.5, 8.5)	(25.3, 9.2)	(80.1, 58.7)
Pixel-level (AUROC, PRO)	Skin	ISIC	1	(33.1, 5.8)	(36.0, 7.7)	(83.3, 55.1)	(89.4, 77.2)	(51.7, 15.9)	(89.7, 78.4)
		CVC-ColonDB	1	(49.5, 15.8)	(49.5, 11.5)	(70.3, 32.5)	(78.4, 64.6)	(40.5, 2.6)	(81.9, 71.3)
	Colon	CVC-ClinicDB	1	(47.5, 18.9)	(48.5, 12.6)	(51.2, 13.8)	(80.5, 60.7)	(34.8, 2.4)	(82.9, 67.8)
		Kvasir	1	(44.6, 17.7)	(45.0, 16.8)	(69.7, 24.5)	(75.0, 36.2)	(44.1, 3.5)	(78.9, 45.6)
	Thyroid	Endo	1	(45.2, 15.9)	(46.6, 12.6)	(68.2, 28.3)	(81.9, 54.9)	(40.6, 3.9)	(84.1, 63.6)
		TN3K	1	(42.3, 7.3)	(35.6, 5.2)	(70.7, 39.8)	(73.6, 37.8)	(34.0, 9.5)	(81.5, 50.4)

Paper 1

- Experiments
 - 정성적 결과



AdaCLIP: Adapting CLIP with Hybrid Learnable Prompts for Zero-Shot Anomaly Detection

Paper 2

- Contribution

- 2가지 형태의 lightweight learnable parameter 도입

- Projection layer

- ※ Patch token과 text embedding 간의 dimension을 align하도록 구성

- Prompting layer

- ※ CLIP내의 vision-transformer layer를 대체함

- ※ 추가적인 prompting token과 layer input을 concat

- Hybrid learnable prompt

- Static prompt

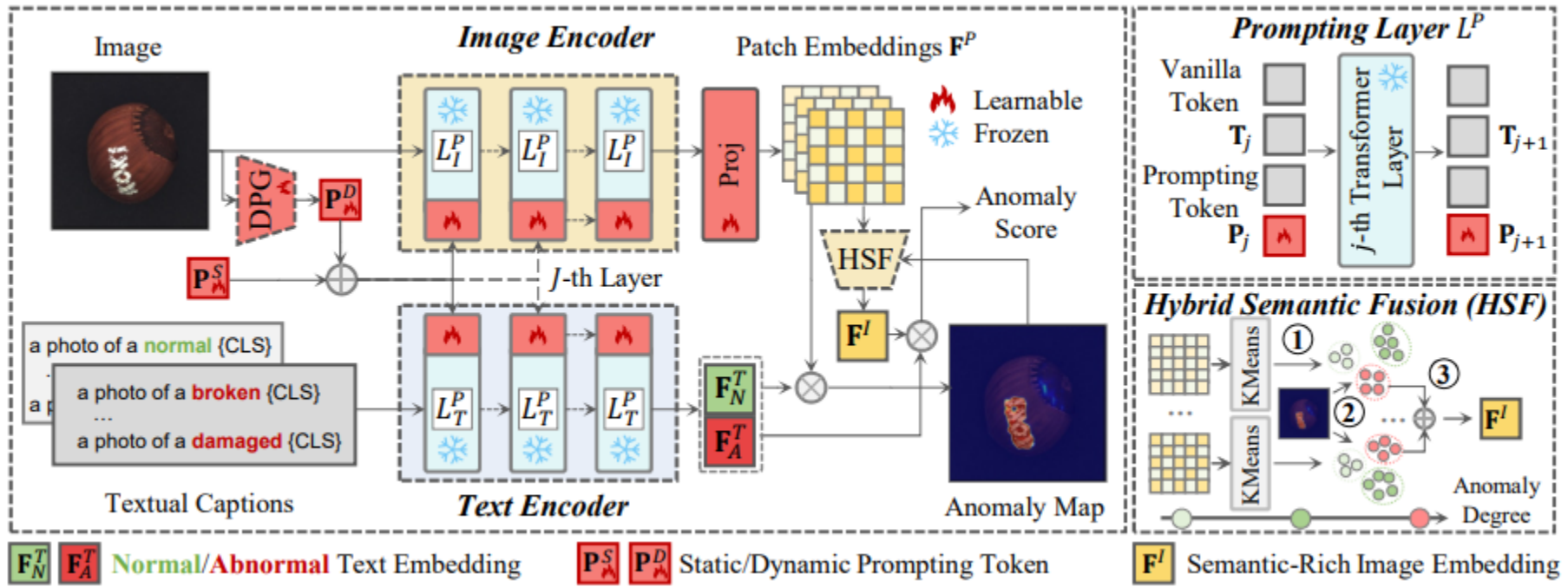
- Dynamic prompt

- Hybrid semantic fusion module

- Anomaly regions에 대한 region-level context를 추출하여 성능 개선

Paper 2

- Overview



Paper 2

- Lightweight learnable parameters

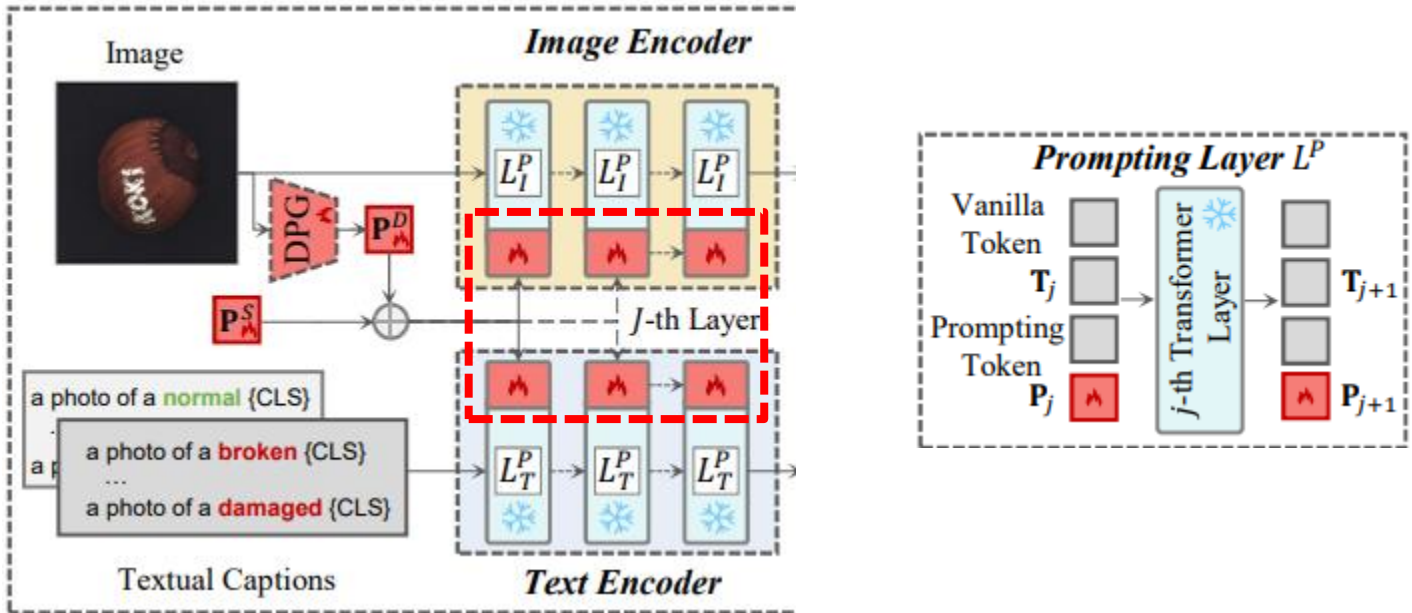
- Prompting layers (L_j^P)

- CLIP encoder의 transformer layer 교체

- Transformer의 weight를 보존하여 generalization capability를 유지함

- ※ Transformer layer의 Self-attention 덕분에, learnable prompt token은 output token에도 contribute할 수 있음

- Input image, text에서 얻어진 vanilla tokens에 learnable prompt token을 concat하는 구조



Paper 2

- Hybrid learnable prompts

- Motivation

- 효율적으로 외부 데이터를 활용하기 위함

- Static prompts (P^S)

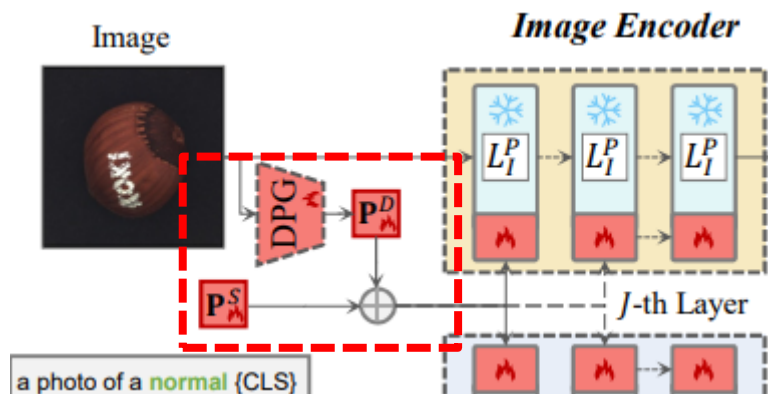
- P^S 는 모든 이미지에 공유되는 learnable token

- Train 과정에서 외부 데이터로부터 명시적으로 학습됨

- Dynamic prompts (P^D)

- 다양한 distribution을 모델링할 수 있는 capacity 강화하기 위해 P^D 를 추가적으로 제안

- P^S 와 달리 Dynamic Prompt Generator (DPG)에 의해 test image마다 생성됨



$$P^D = \text{DPG}(I).$$

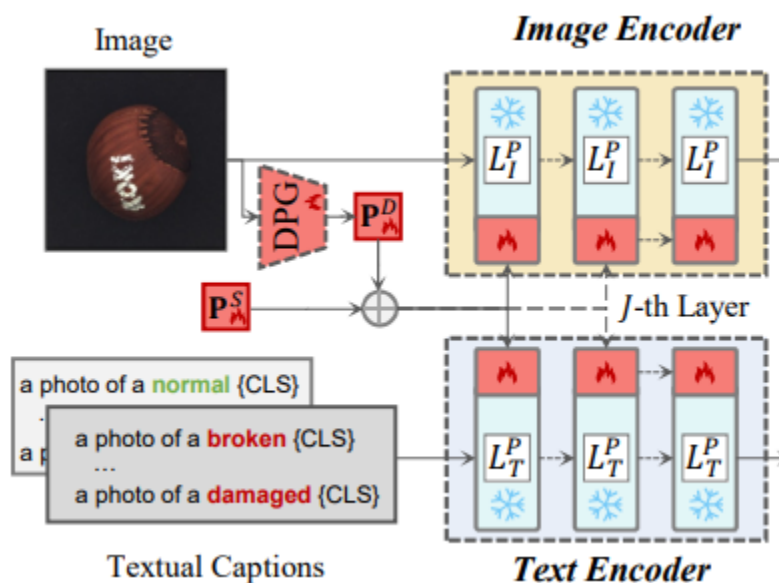
Paper 2

- Hybrid learnable prompts

- Dynamic prompts (P^D)

- DPG는 CLIP을 사용하여 추출한 class token을 P^D 로 projection하기 위한 learnable linear layer로 구성

- L_I^P, L_T^P 에 input으로 넣기 위한 dynamic prompt는 모두 test image에서 생성됨



Paper 2

- Hybrid learnable prompts

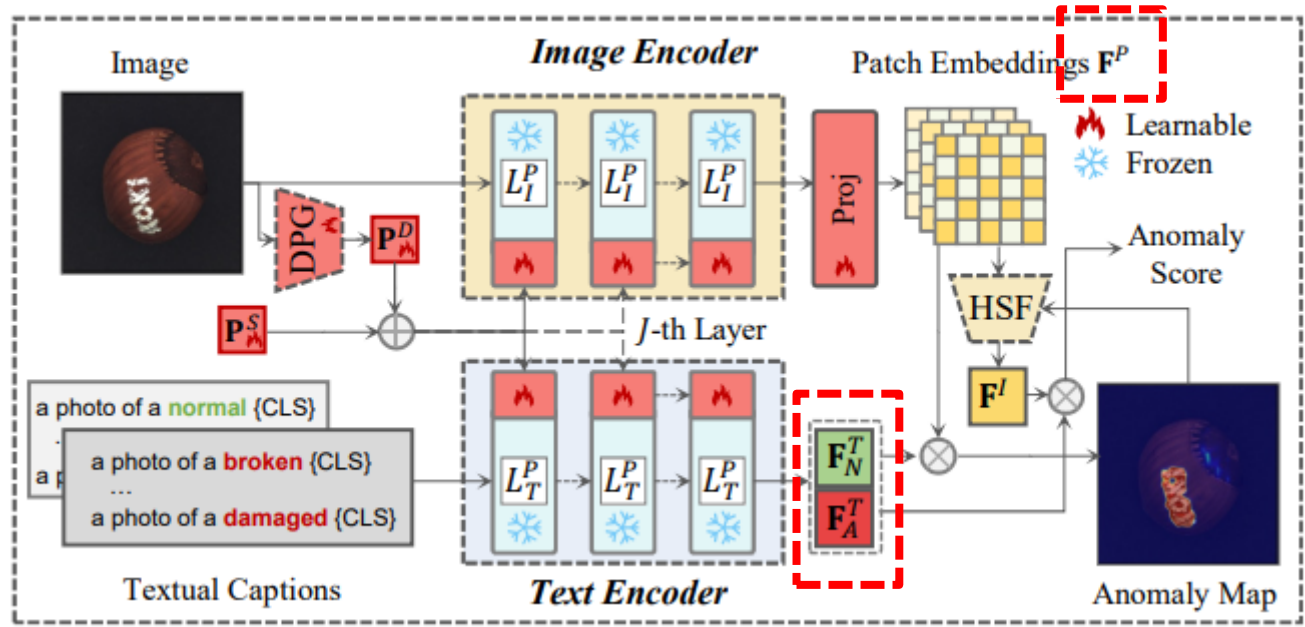
- Hybrid prompts

- Static, dynamic prompts를 합쳐서 hybrid prompts 생성

- ※ 이는 L_I^P, L_T^P 두 prompt layer에서 사용됨

- 기존 ViT의 layer를 prompt layer로 대체하여 image encoder는 input image의 patch embedding F^P 추출

- Text encoder는 text caption에 대한 normal/abnormal text embedding F_N^T, F_A^T 생성



Paper 2

- Lightweight learnable parameters

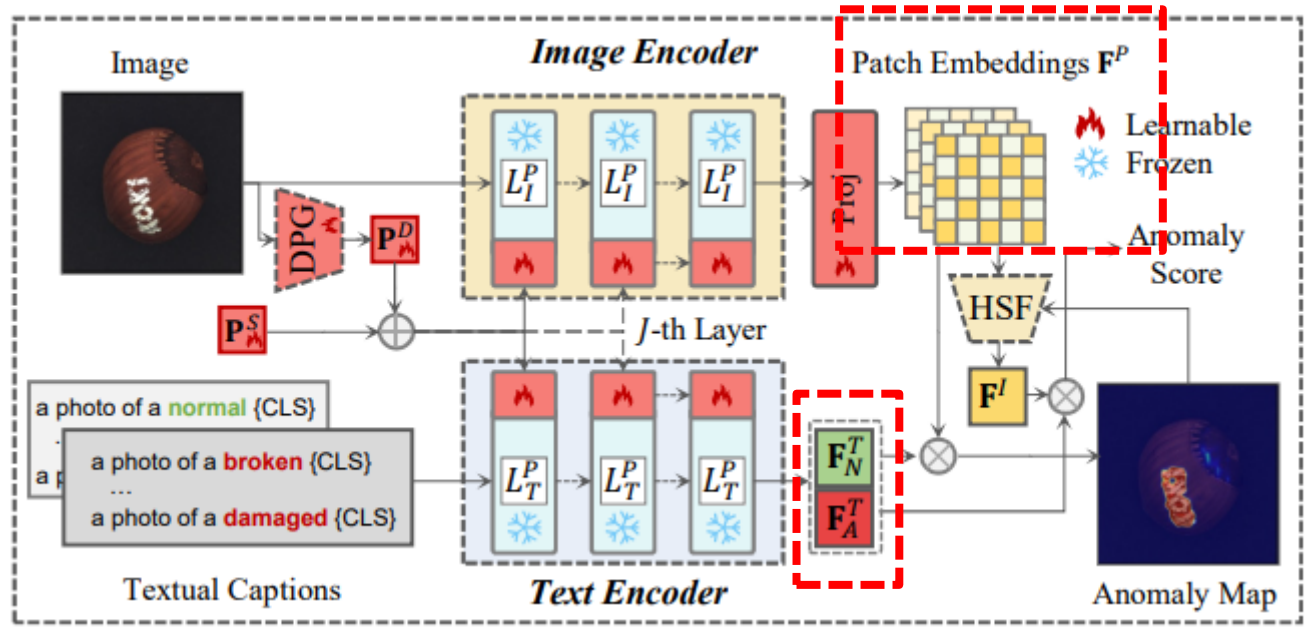
- Projection layers

- Motivation

※ 기존 CLIP 구조는 patch embedding과 text embedding의 dimension이 같지 않음

- Solution

※ Image encoder에 linear layer를 사용하여 patch embedding F^P 과 F_N^T, F_A^T 간의 dimension을 일치시킴



Paper 2

- Anomaly scoring

- Pixel-level Anomaly Localization

- Patch embedding \mathbf{F}^P 과 text embedding $\mathbf{F}_N^T, \mathbf{F}_A^T$ 간의 cosine similarity 측정하여 anomaly score 계산
 - 아래 수식과 같이 anomaly map을 정의

$$\mathbf{M}_i = \phi \left(\frac{\exp(\cos(\mathbf{F}_i^P, \mathbf{F}_A^T))}{\exp(\cos(\mathbf{F}_i^P, \mathbf{F}_N^T)) + \exp(\cos(\mathbf{F}_i^P, \mathbf{F}_A^T))} \right),$$

- 여러 layer에서의 anomaly map을 multi-hierarchy 방식으로 결합하여 최종 prediction M 생성
 - Train 중에는 최종 prediction M 을 optimization하기 위해 외부 데이터를 사용하여 Dice loss, Focal loss 사용

Paper 2

- Anomaly scoring

- Image-level Anomaly Detection

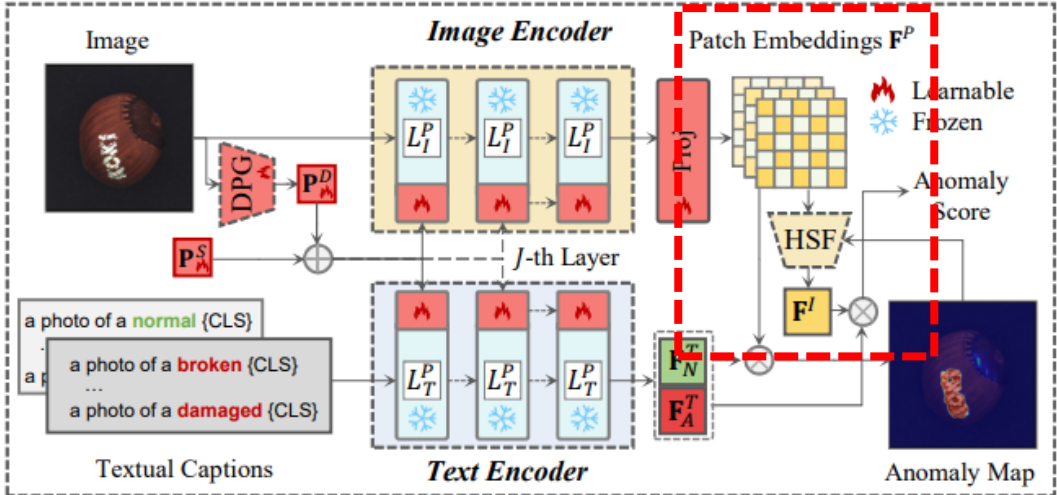
- Hybrid Semantic Fusion (HSF) Module

※ 기존 image-level AUROC 계산법

- ✓ Anomaly map의 최대 값을 선택하여 anomaly score 계산
 - ✓ 이는 noise가 많은 예측에 민감함

※ HSF module은 anomaly를 나타낼 가능성이 높은 patch embedding을 aggregate하여 region-level의 정보를 통합함

- ✓ 강건한 image-level anomaly detection 가능함

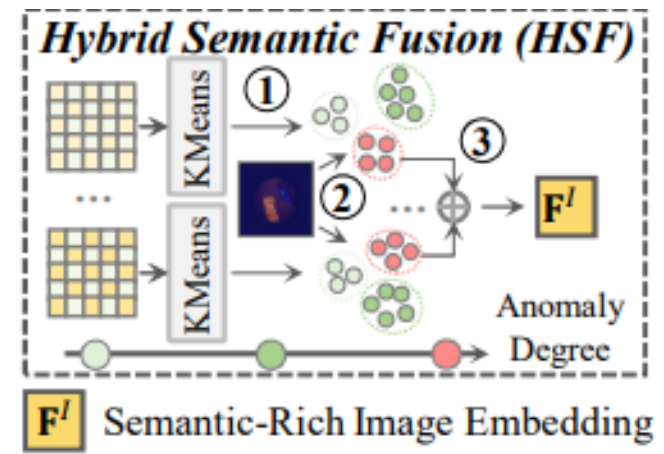
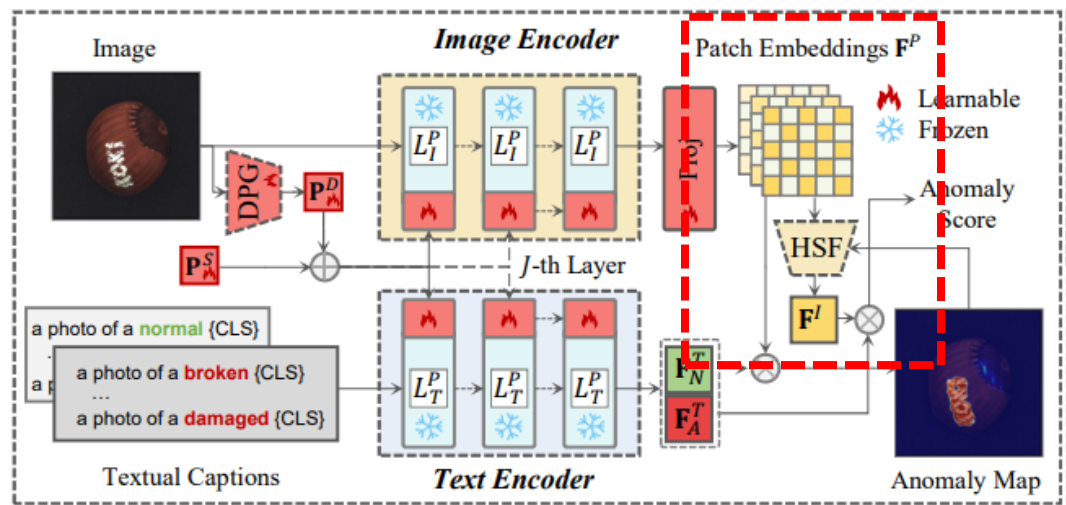


Paper 2

- Anomaly scoring
 - Image-level Anomaly Detection

- Hybrid Semantic Fusion (HSF) Module
- 3가지 step으로 작동됨

- ※ Kmeans를 사용하여 patch embedding을 K개의 group으로 clustering
- ※ Anomaly map M 에서 해당 위치의 score를 평균 내어 각 cluster의 anomaly score 계산
- ※ 가장 높은 anomaly score를 가진 cluster를 선택하고, 해당 cluster의 중심값을 계산한 뒤 최종 image embedding F^I 에 aggregate



Paper 2

- Experiments

- 정량적 결과

Metric	Dataset	w/o supervised training		w/ supervised training			
		SAA [10]	WinCLIP [24]	DINOv2 [39]	SAM [30]	APRIL-GAN [14]	AdaCLIP
Image-level (AUROC, max-F1)	MVTec AD	(63.5, 87.4)	(91.8, 92.9) [†]	(74.4, 87.4)	(70.8, 86.0)	(82.3, 88.9)	(89.2, 90.6)
	VisA	(67.1, 75.9)	(78.1, <u>80.7</u>) [†]	(75.2, 78.5)	(61.9, 73.9)	(81.7, 80.7)	(85.8, 83.1)
	MPDD	(42.7, 73.9)	(61.4, <u>77.5</u>)	(62.4, 74.9)	(63.0, 77.0)	(<u>66.0</u> , 76.0)	(76.0, 82.5)
	BTAD	(59.0, 89.7)	(68.2, 67.6)	(79.3, 69.3)	(89.4, 85.7)	(85.2, 82.0)	(<u>88.6</u> , <u>88.2</u>)
	KSDD	(68.6, 37.6)	(93.3, 79.0)	(94.9, 77.5)	(65.8, 37.9)	(<u>95.7</u> , <u>85.2</u>)	(97.1, 90.7)
	DAGM	(87.1, 88.8)	(91.7, 87.6)	(90.7, 89.2)	(82.7, 83.6)	(<u>93.5</u> , <u>91.8</u>)	(99.1, 97.5)
	DTD-Synthetic	(94.4, 93.5)	(95.1, 94.1)	(85.8, 93.5)	(81.9, 91.1)	(98.1, 96.8)	(<u>95.5</u> , <u>94.7</u>)
	Average Rank	(68.9, 78.1) (5.3, 4.4)	(82.8, 82.8) (3.4, 3.4)	(80.4, 81.5) (4.0, 4.1)	(73.6, 76.4) (4.7, 5.0)	(<u>86.1</u> , <u>85.9</u>) (<u>2.1</u> , <u>2.6</u>)	(90.2, 89.6) (1.4, 1.4)
Pixel-level (AUROC, max-F1)	MVTec AD	(75.5, 38.1)	(85.1, 31.6) [†]	(<u>85.9</u> , 39.6)	(85.4, 29.4)	(83.7, <u>39.8</u>)	(88.7, 43.4)
	VisA	(76.5, 31.6)	(79.6, 14.8) [†]	(95.0, 30.3)	(92.6, 18.2)	(<u>95.2</u> , <u>32.3</u>)	(95.5, 37.7)
	MPDD	(81.7, 18.9)	(71.2, 15.4)	(<u>95.6</u> , <u>31.1</u>)	(94.8, 22.1)	(95.1, 30.6)	(96.1, 34.9)
	BTAD	(65.8, 14.8)	(72.6, 18.5)	(91.9, 43.4)	(93.8, 46.9)	(89.5, 38.4)	(<u>92.1</u> , 51.7)
	KSDD	(78.8, 6.6)	(95.8, 21.3)	(99.3, 50.6)	(91.2, 18.4)	(<u>98.2</u> , 56.2)	(97.7, <u>54.5</u>)
	DAGM	(62.7, 32.6)	(81.3, 13.9)	(<u>90.9</u> , 52.0)	(88.6, 40.7)	(90.3, 57.9)	(91.5, 57.5)
	DTD-Synthetic	(76.7, 60.6)	(79.5, 16.1)	(97.0, 63.4)	(95.0, 56.7)	(<u>97.8</u> , 72.7)	(97.9, 71.6)
	Average Rank	(73.9, 29.0) (5.9, 4.7)	(80.7, 18.8) (4.9, 5.6)	(<u>93.7</u> , 44.3) (<u>2.3</u> , 3.0)	(91.7, 33.2) (3.6, 4.3)	(92.8, 46.9) (3.0, <u>2.0</u>)	(94.2, 50.2) (1.4, 1.4)

Paper 2

- Experiments

- 정량적 결과

Metric	Dataset	w/o supervised training			w/ supervised training		
		SAA [10]	WinCLIP [24]	DINOv2 [39]	SAM [30]	APRIL-GAN [14]	AdaCLIP
Image-level (AUROC, max-F1)	HeadCT	(46.8, 68.0)	(84.1, 79.8)	(71.4, 72.4)	(78.4, 76.4)	(93.6, 86.4)	(91.4, 85.2)
	BrainMRI	(34.4, 76.7)	<u>(89.8, 86.3)</u>	(78.3, 82.7)	(71.5, 78.9)	(89.7, <u>89.5</u>)	(94.8, 91.2)
	Br35H	(33.2, 67.3)	(81.6, 74.4)	(69.1, 70.5)	(59.0, 67.2)	<u>(95.6, 91.0)</u>	(97.7, 92.4)
	Average	(38.1, 70.7)	(85.2, 80.2)	(72.9, 75.2)	(69.7, 74.1)	<u>(93.0, 89.0)</u>	(94.6, 89.6)
	Rank	(6.0, 5.7)	(2.7, 3.0)	(4.3, 4.3)	(4.7, 5.0)	<u>(2.0, 1.7)</u>	(1.3, 1.3)
Pixel-level (AUROC, max-F1)	ISIC	(83.8, 74.2)	(67.1, 48.5)	(94.2, 79.6)	<u>(94.2, 81.0)</u>	(92.1, 77.4)	(89.3, 71.4)
	ColonDB	(71.8, 31.5)	(61.1, 19.6)	(87.3, <u>56.5</u>)	(86.1, 45.7)	<u>(88.7, 52.6)</u>	(90.4, 58.2)
	ClinicDB	(66.2, 29.1)	(67.1, 24.4)	(83.3, <u>56.2</u>)	<u>(83.5, 43.0)</u>	(82.5, 51.8)	(84.4, 58.2)
	TN3K	(66.8, 32.6)	(67.2, 30.0)	(73.3, 35.7)	(70.1, 32.5)	<u>(75.9, 36.4)</u>	(77.2, 41.9)
	Average	(72.1, 41.8)	(65.6, 30.6)	(84.5, <u>57.0</u>)	(83.5, 50.5)	<u>(84.8, 54.6)</u>	(85.3, 57.4)
Rank	(5.5, 4.5)	(5.5, 6.0)	<u>(2.5, 2.3)</u>	(3.0, 3.5)	(2.8, 2.8)	(1.8, 2.0)	

Paper 2

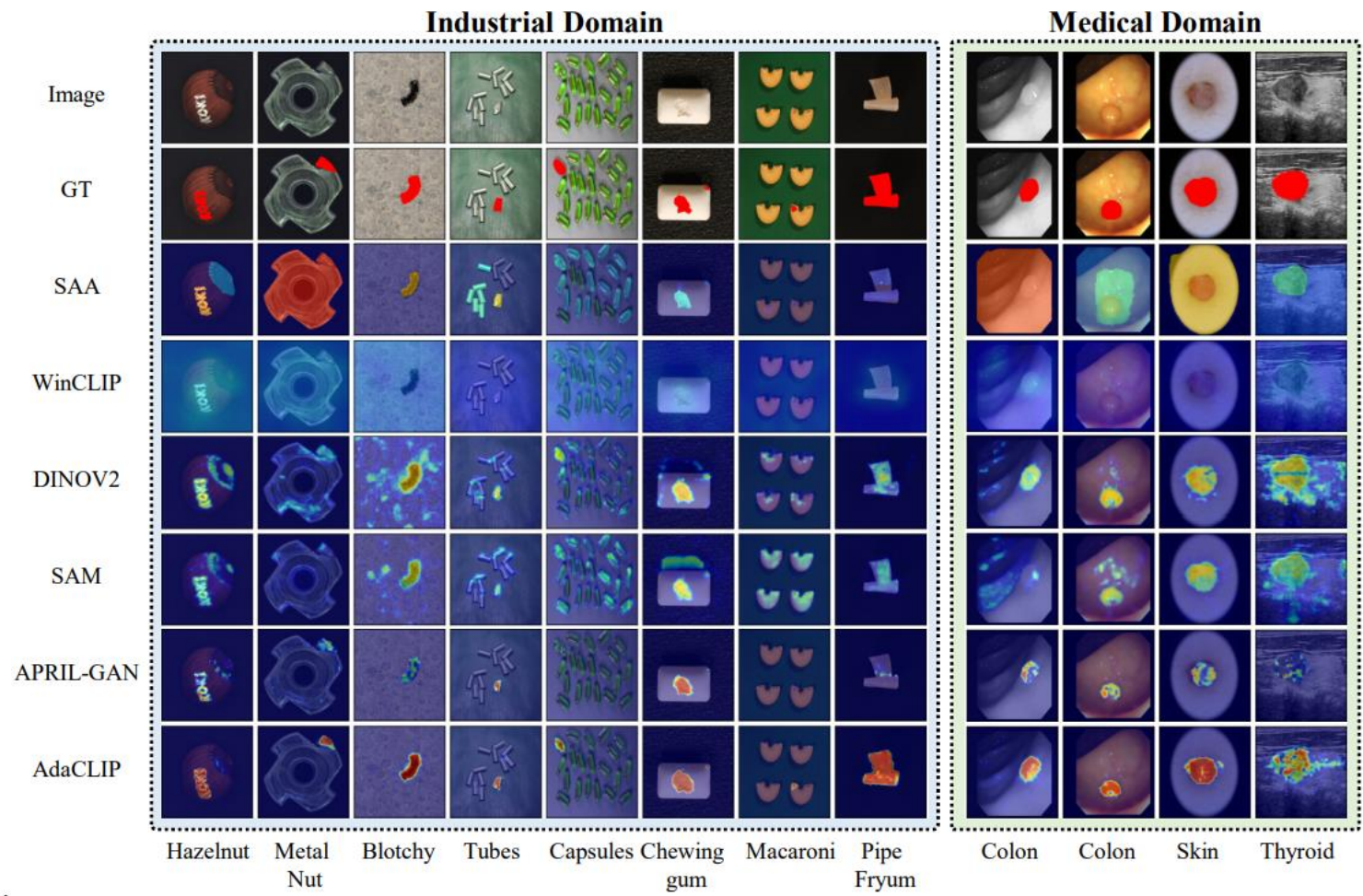
- Experiments

- 정량적 결과

Task	Category	Datasets	C	CLIP	WinCLIP	CoOp	AnomalyCLIP	AdaCLIP	GlocalCLIP
Image-level (AUROC, AP)	Obj & texture	MVTec AD	15	(83.3, 92.4)	(90.4, 95.6)	(82.1, 91.4)	(91.5, 96.2)	(91.2, 95.9)	(91.7, 96.4)
		VisA	12	(71.7, 76.6)	(75.6, 78.8)	(77.7, 81.5)	(81.4, 84.9)	(81.7, 84.0)	(83.7, 86.2)
	Obj	MPDD	6	(71.2, 78.2)	(61.5, 69.2)	(76.0, 78.3)	(76.9, 81.4)	(72.1, 76.0)	(77.6, 82.0)
		BTAD	3	(82.7, 86.5)	(68.2, 70.9)	(77.7, 77.7)	(87.5, 90.7)	(90.2, 90.6)	(89.8, 92.2)
		SDD	1	(74.0, 57.5)	(84.3, 77.4)	(80.8, 71.0)	(85.3, 81.6)	(81.2, 72.6)	(86.6, 84.5)
	Texture	DTD-Synthetic	12	(73.7, 89.7)	(95.1, 97.7)	(96.2, 98.1)	(93.7, 97.3)	(97.8, 99.0)	(93.7, 97.3)
		Mean		(76.1, 80.2)	(79.2, 81.6)	(81.8, 83.0)	(86.1, 88.7)	(85.7, 86.4)	(87.2, 89.8)
Pixel-level (AUROC, PRO)	Obj & texture	MVTec AD	15	(38.2, 8.8)	(82.3, 61.9)	(44.4, 11.1)	(91.0, 81.9)	(89.4, 37.8)	(91.4, 82.8)
		VisA	12	(47.9, 16.1)	(73.2, 51.1)	(42.1, 12.2)	(95.3, 85.1)	(95.5, 77.8)	(95.9, 87.5)
	Obj	MPDD	6	(42.5, 19.8)	(71.2, 40.5)	(33.7, 14.1)	(96.2, 87.5)	(96.4, 62.2)	(96.6, 89.0)
		BTAD	3	(39.5, 7.8)	(72.7, 27.3)	(28.1, 6.5)	(94.5, 73.6)	(94.8, 32.5)	(96.1, 77.9)
		SDD	1	(38.7, 10.1)	(68.8, 24.2)	(24.4, 8.3)	(90.6, 67.0)	(71.7, 17.6)	(93.1, 72.4)
	Texture	DTD-Synthetic	12	(37.6, 15.0)	(79.5, 51.5)	(14.8, 3.0)	(97.8, 91.1)	(98.7, 80.0)	(98.2, 92.5)
		Mean		(40.7, 12.9)	(74.6, 42.8)	(31.3, 8.5)	(94.2, 81.0)	(91.1, 51.3)	(95.2, 83.7)

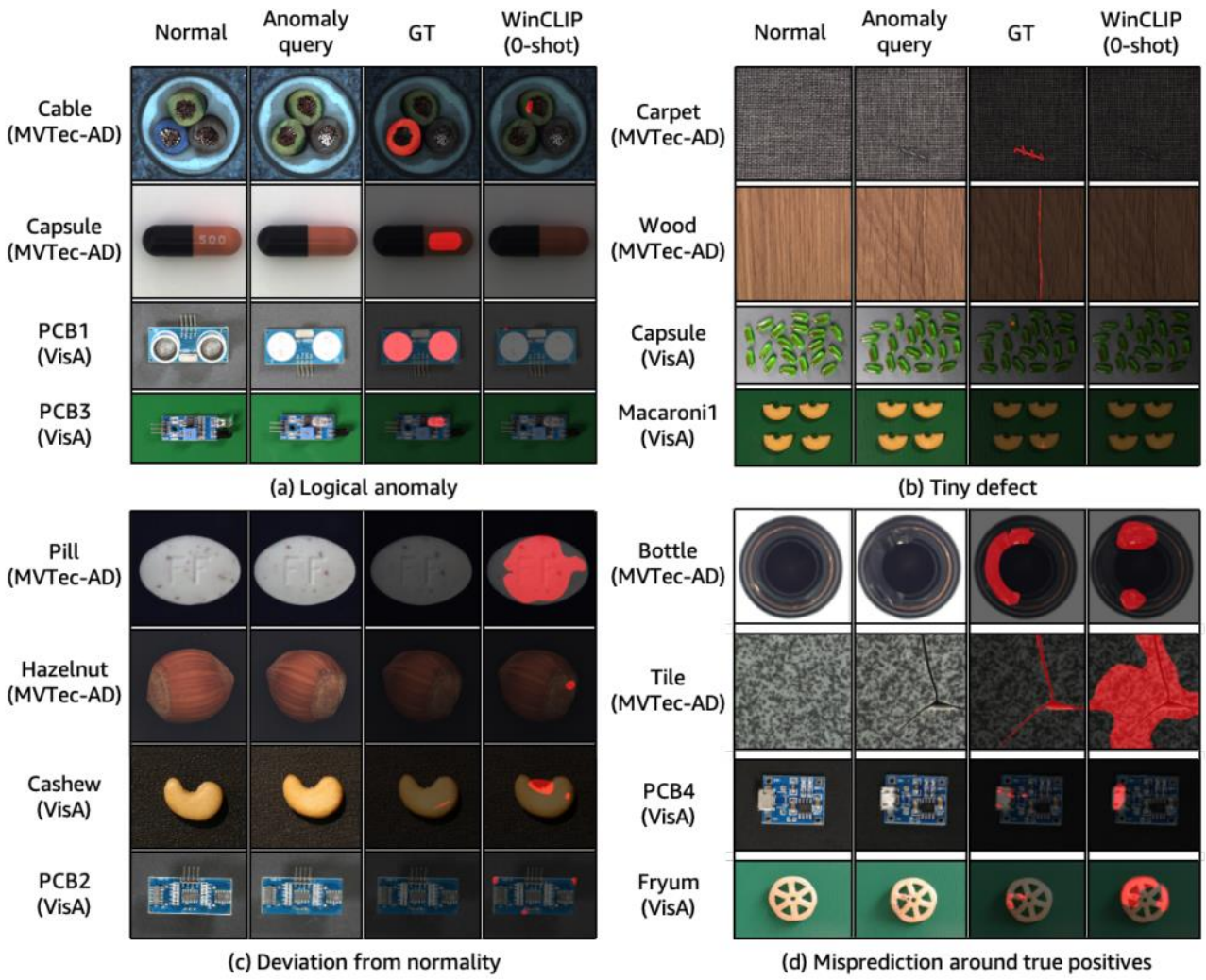
Paper 2

- Experiments
 - 정성적 결과



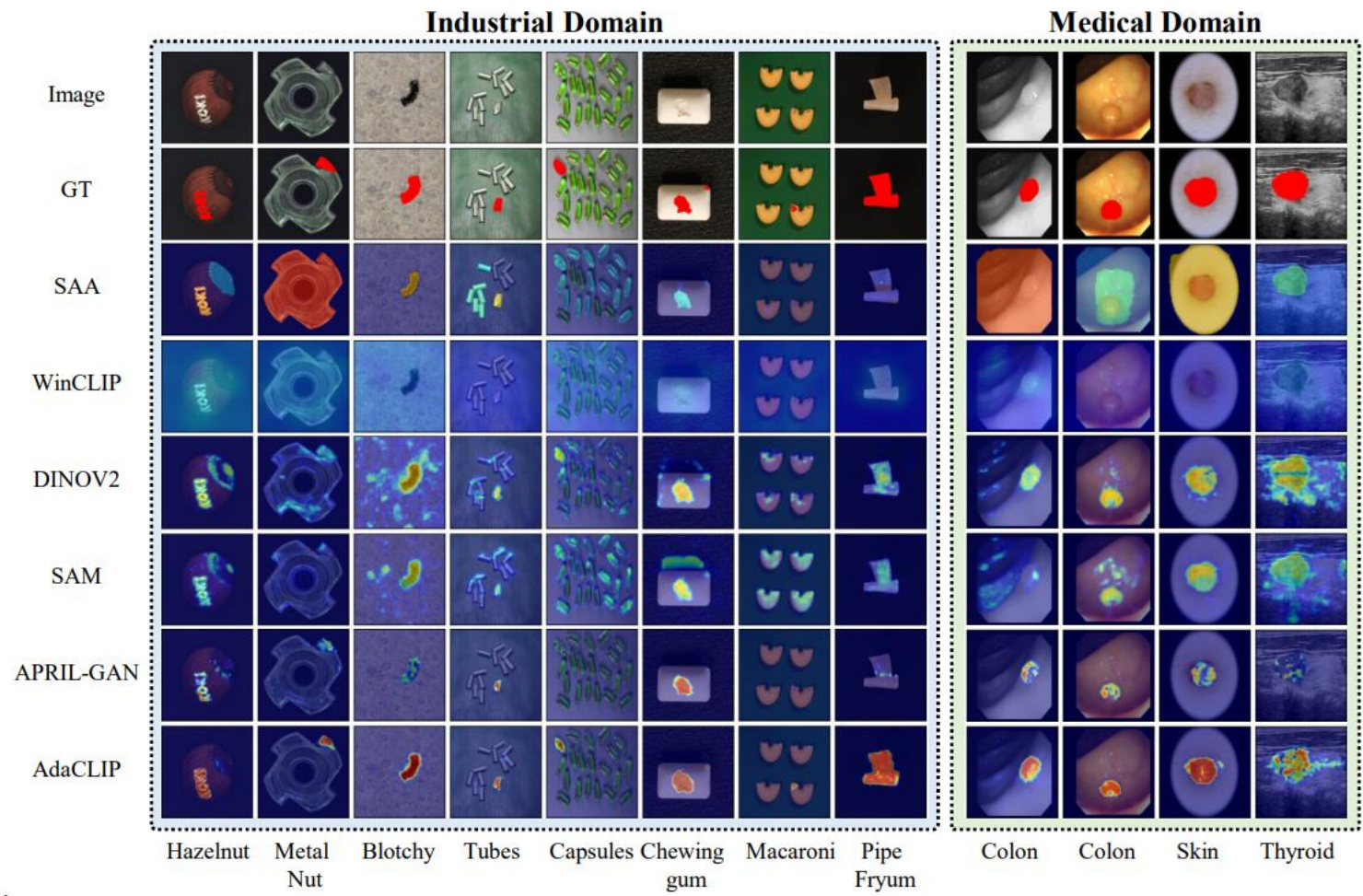
Paper 2

- Experiments in WinCLIP



Paper 2

- Experiments
 - 정성적 결과



Thank You