華中科技大學
Huazhong University of Science and Technology
武汉精测电子集团股份有限公司
WUHAN JINGCE ELECTRONIC GROUP CO.,LTD

CVPR Nashville JUNE 11-15, 2025

AnomalyNCD: Towards Novel Anomaly Class Discovery in Industrial Scenarios

Ziming Huang*, Xurui Li*, Haotian Liu*, Feng Xue, Yuzhe Wang, Yu Zhou

Table of Contents

Table of Contents

Characteristics of Industrial Anomaly Data

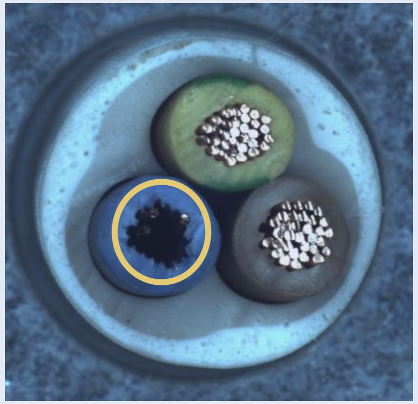
"industrial defects", "industrial anomalies" denote defects in industrial products.

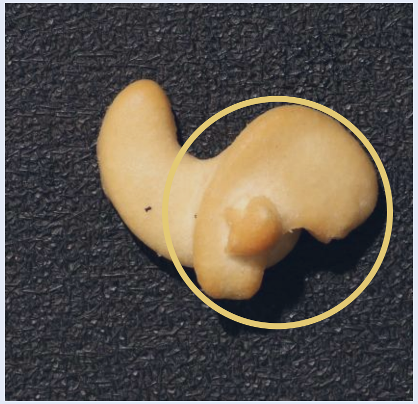
Visual Defects (Additive Defects)

Visually distinct from normal conditions, such as stains, cracks, etc.

Logical Defects (Subtractive Defects)

Partial loss of local information on normal products

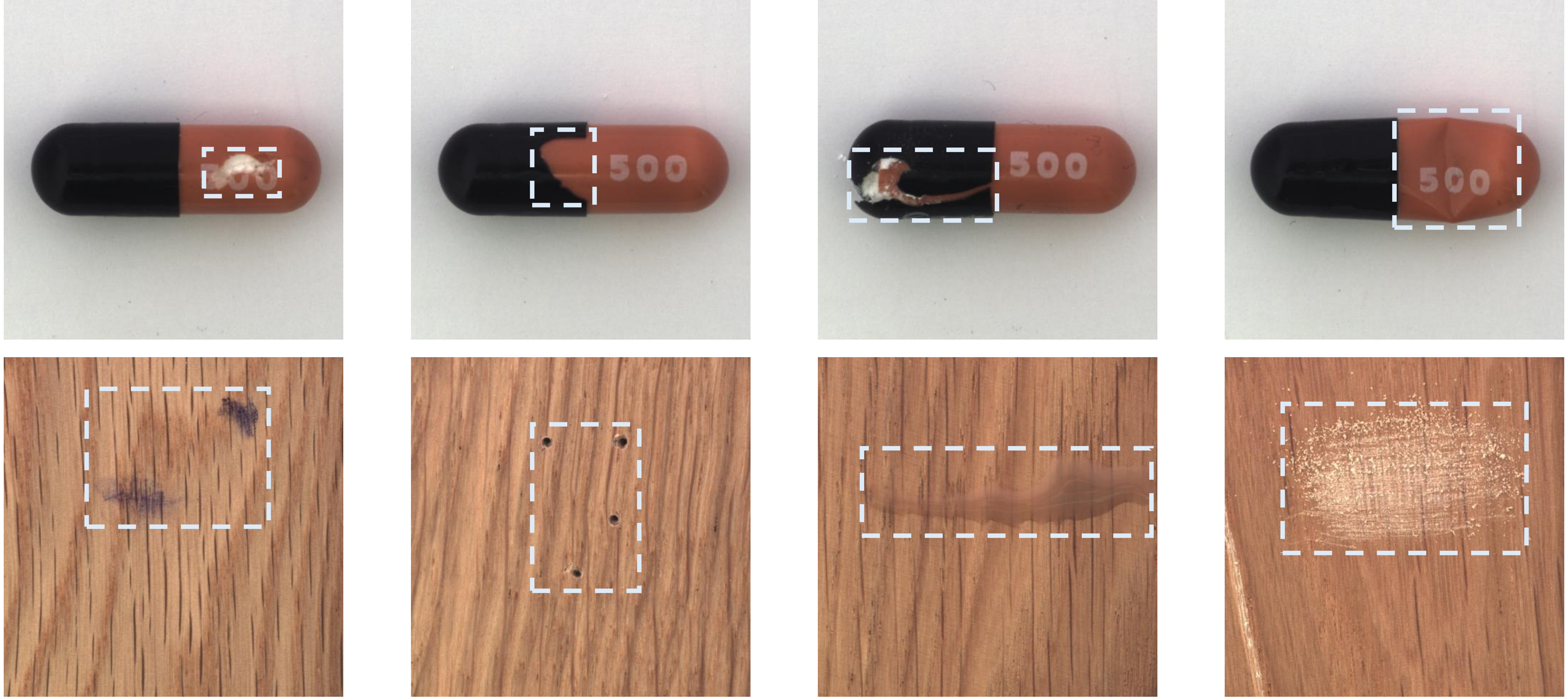




Based on normal industrial products, parts that violate the rules of normal products are determined as defects (anomalies).

Characteristics of Industrial Anomaly Data

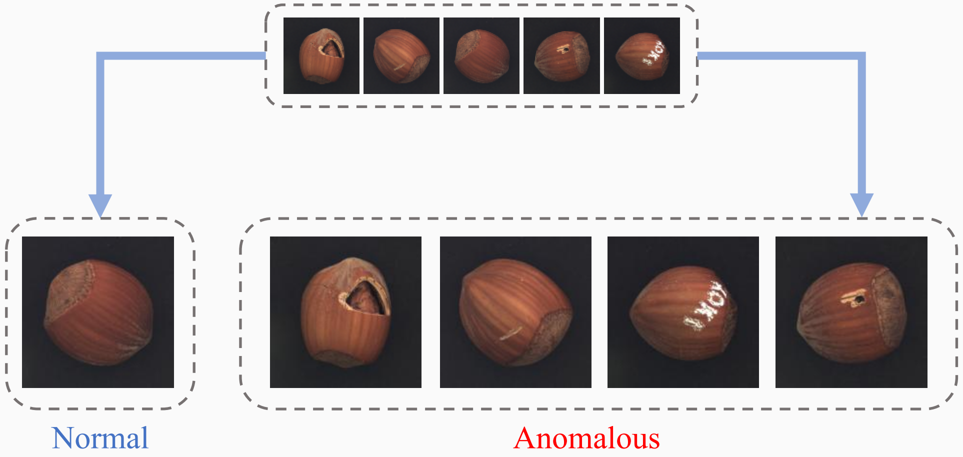
Visual Randomness | **Shape Randomness** | **Location Randomness**



Industrial Anomaly Detection and Novel Class Discovery

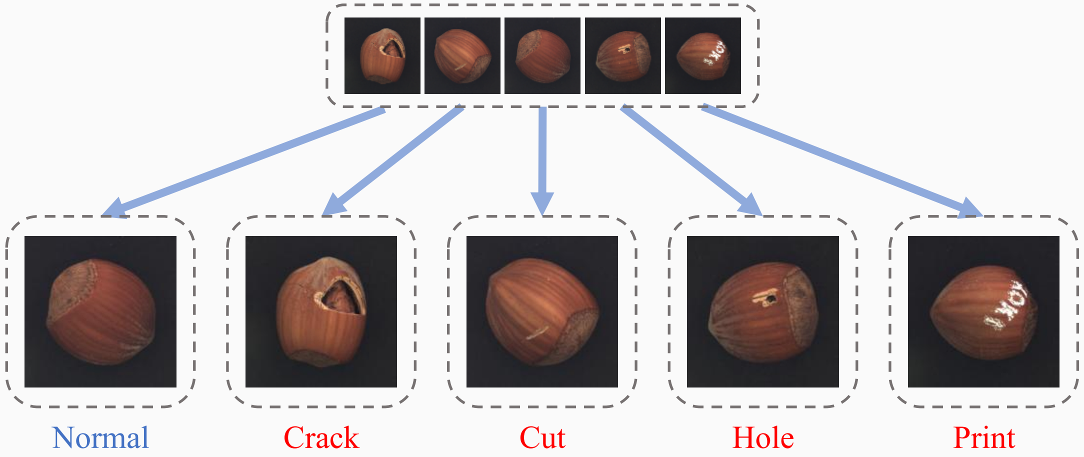
Research Tasks

Industrial Anomaly Detection:
Binary Classification Task



Only distinguish between: Normal or Anomalous

Industrial Novel Class Discovery :
Multi-class Classification Task



Distinguish between: Normal, Anomaly Type 1, Anomaly Type 2, …

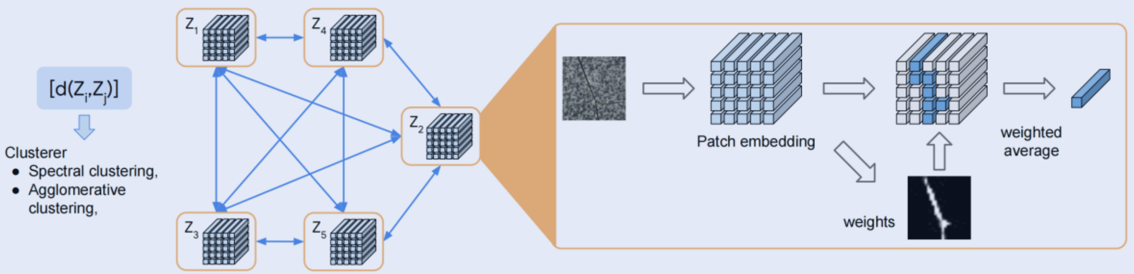
Supervised defect classification methods have become relatively mature and are widely applied.

Table of Contents

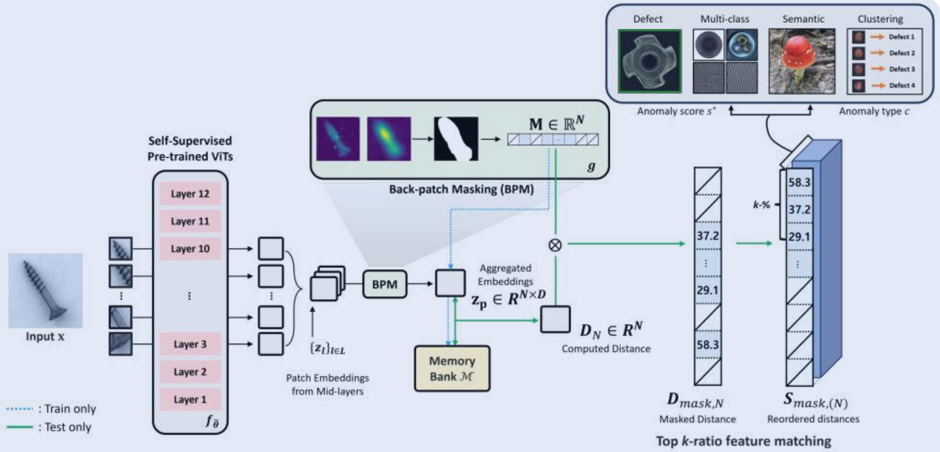
Anomaly Clustering Methods

Industrial Defect Clustering Method Based on Anomaly Region Focusing,

such as Anomaly Clustering,Uniformaly, etc.



Anomaly Clustering



Uniformaly

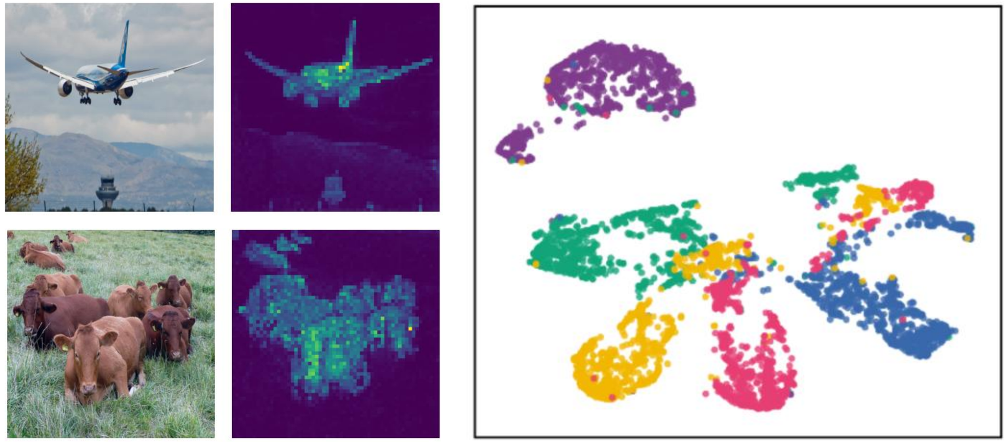
Model the features of abnormal regions and perform clustering to achieve classification.

Limitations of existing methods

Challenges in anomaly clustering methods: low feature discriminability

Natural scene images

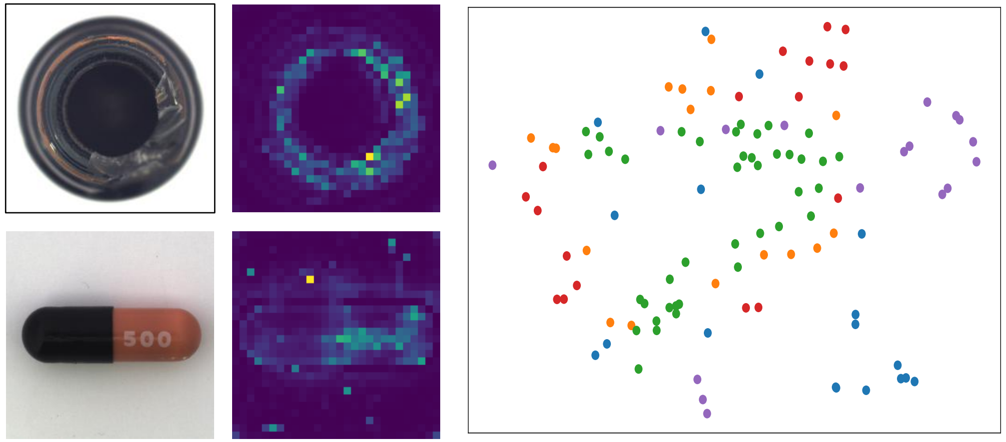
High
feature discriminability



Strong semantics, easy to focus on

Industrial scene images

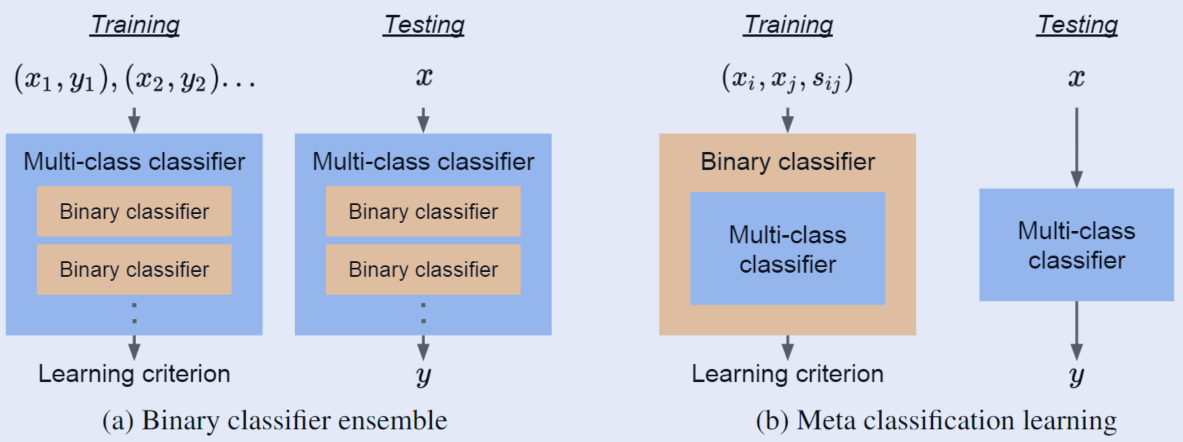
Low
feature discriminability



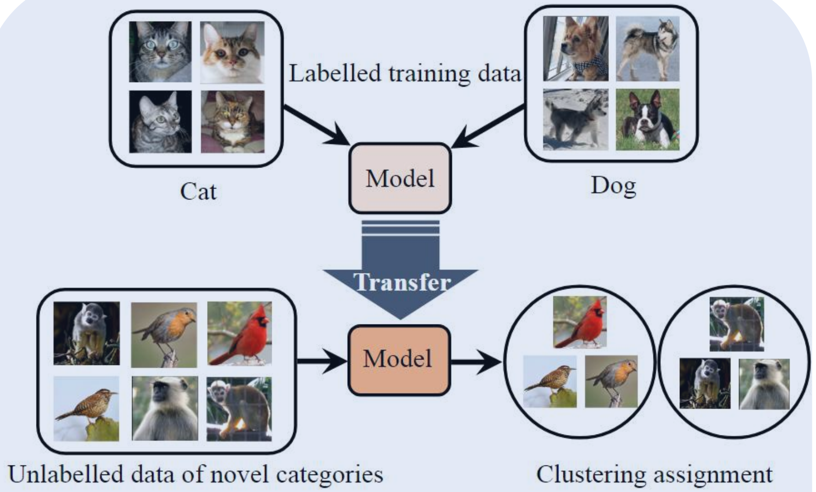
Weak semantics, difficult to focus on

Novel Class Discovery

Two-stage knowledge transfer methods based on shared networks,
such as CCN,MCL,DTC, etc.



MCL



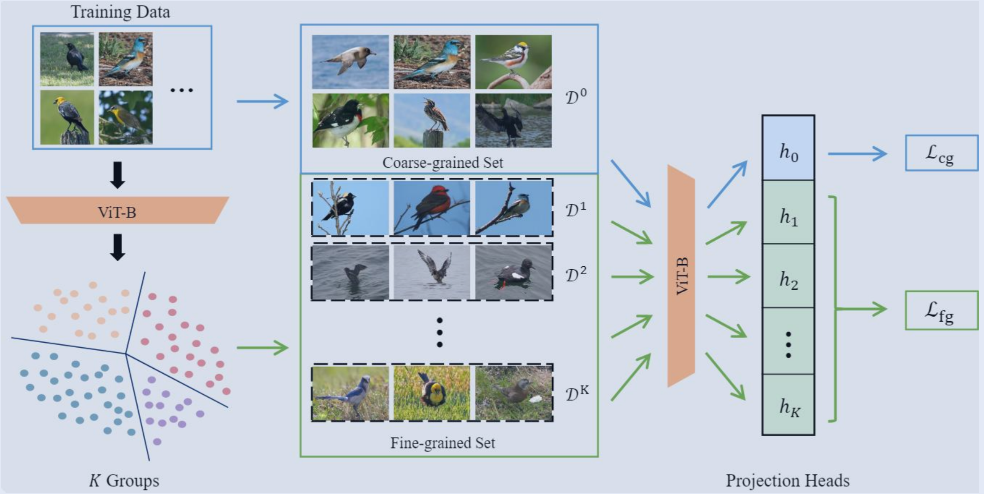
DTC

Leverage labeled data to learn classification prior knowledge,

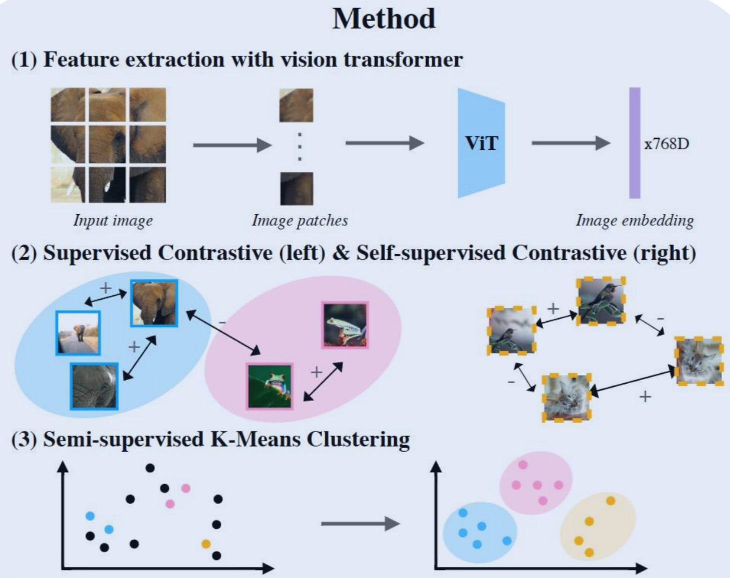
thereby assisting in the classification of unlabeled data.

Novel Class Discovery

Clustering methods based on self-supervised contrastive learning of image features, such as GCD,XCon,etc.



XCon

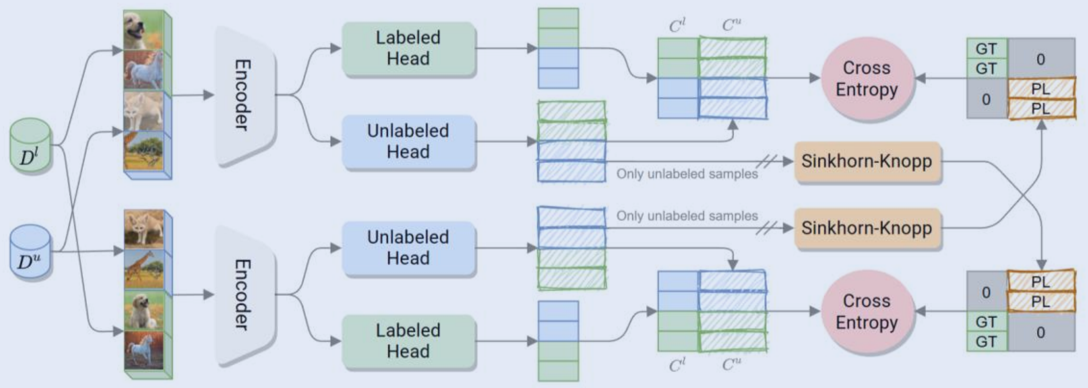


GCD

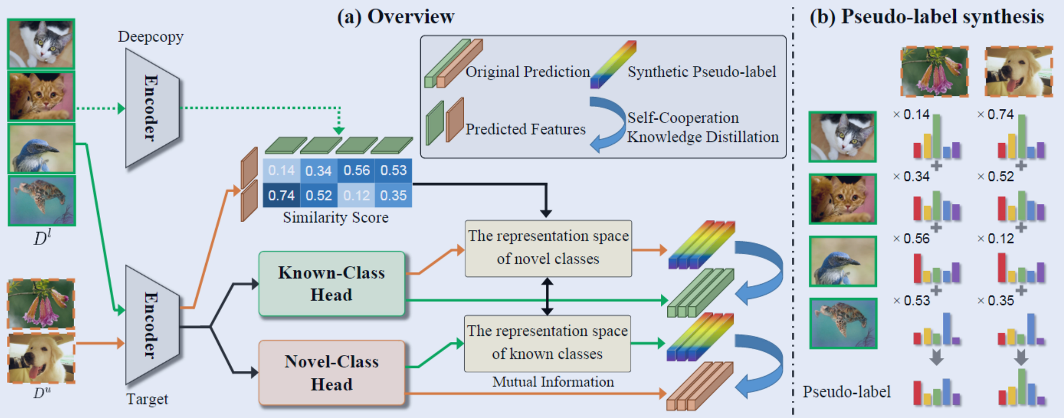
Utilize the differences between image features for contrastive learning and clustering to achieve classification.

Novel Class Discovery

Classification methods based on pseudo-label generation and parameterized classifiers, such as UNO, SCKD, BYOP, etc.



UNO



SCKD

Generate pseudo-labels for unlabeled data to train parameterized classifiers to achieve classification.

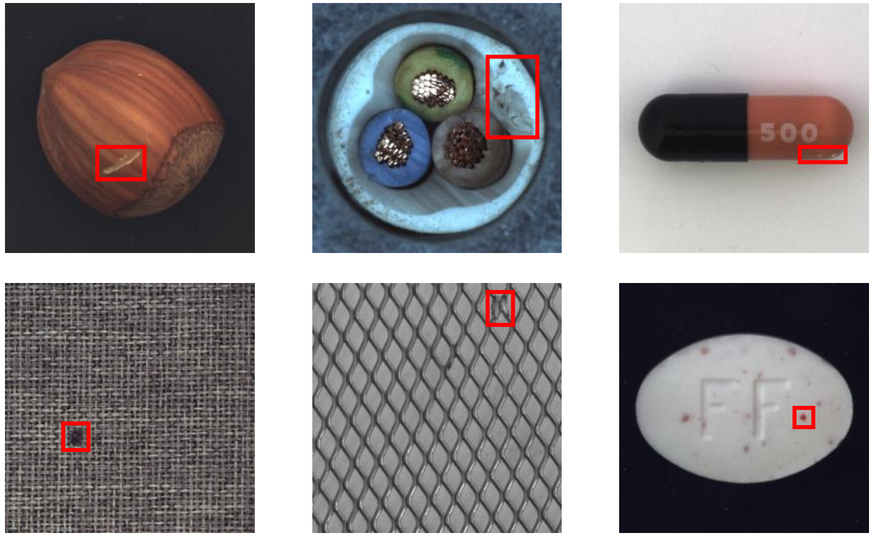
Limitations of existing methods

Challenges in novel category discovery methods:
difficulty in focusing on local minor defects.

Natural scene images

The target object is centrally positioned with a large area.

Industrial scene images

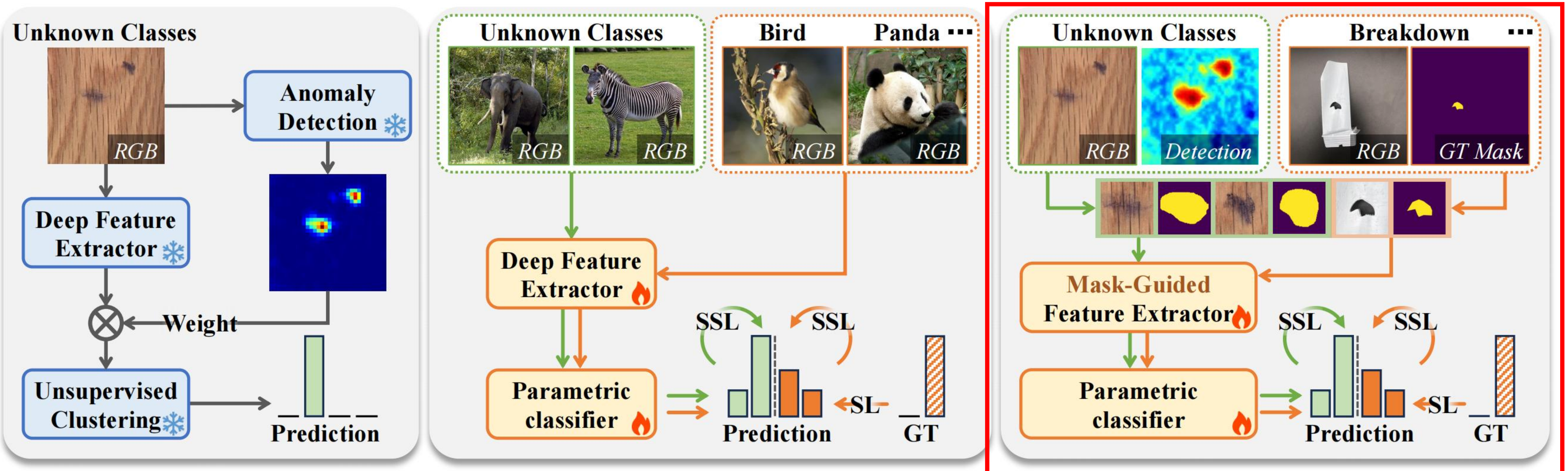


The target defect is randomly positioned with a small area.

Table of Contents

Research motivation

Motivation

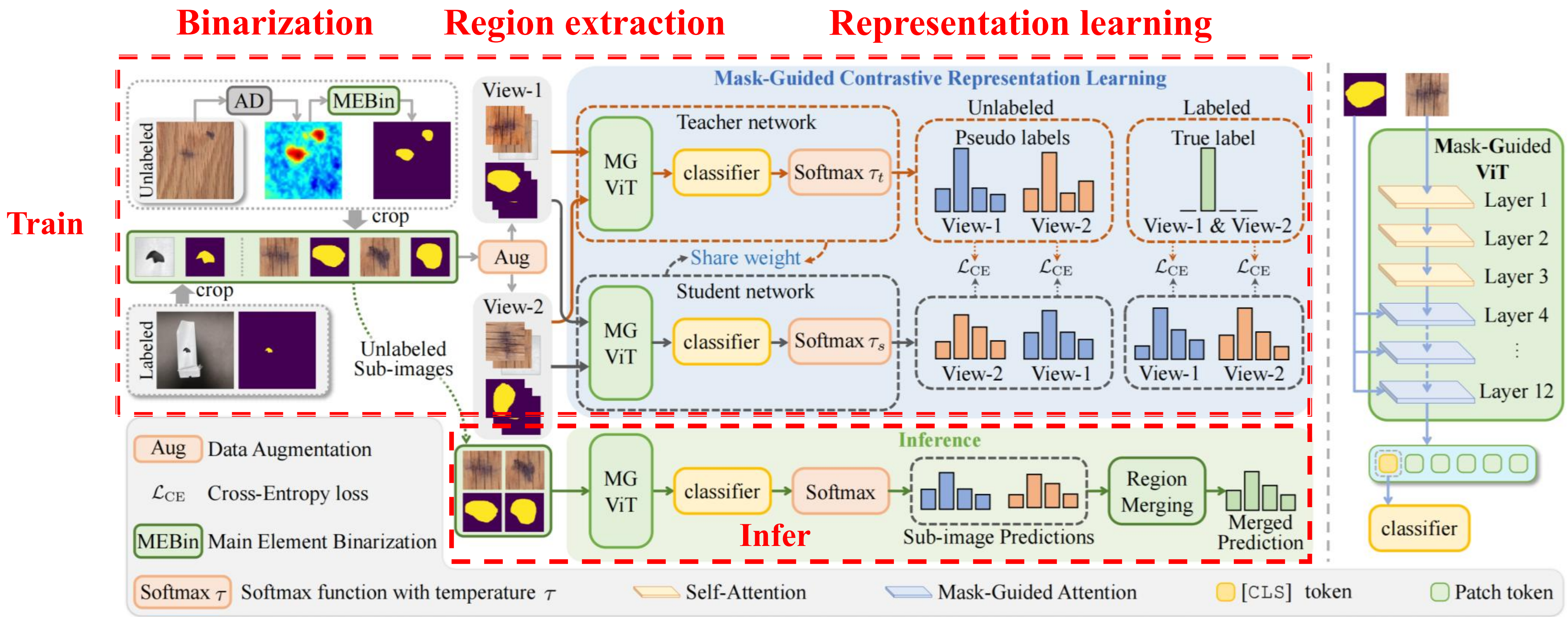


(a) Anomaly Clustering (b) NCD in natural scenario (c) AnomalyNCD

Adaptive iterative optimization, focusing on abnormal regions for classification.

Method

Framework



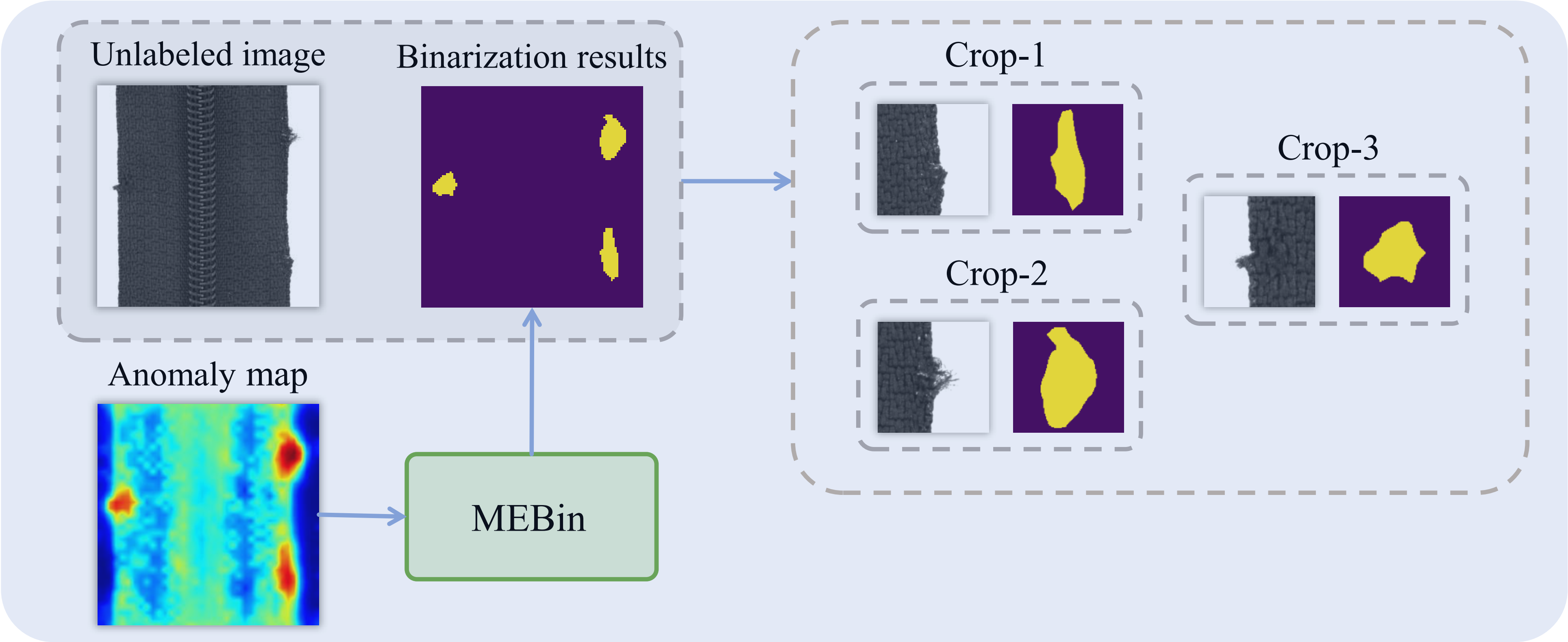
Input test images, perform adaptive binarization and defect region extraction,

and enhance feature discriminability of different defect categories via self-supervised training.

Input the test images into the above-trained model again.

Method

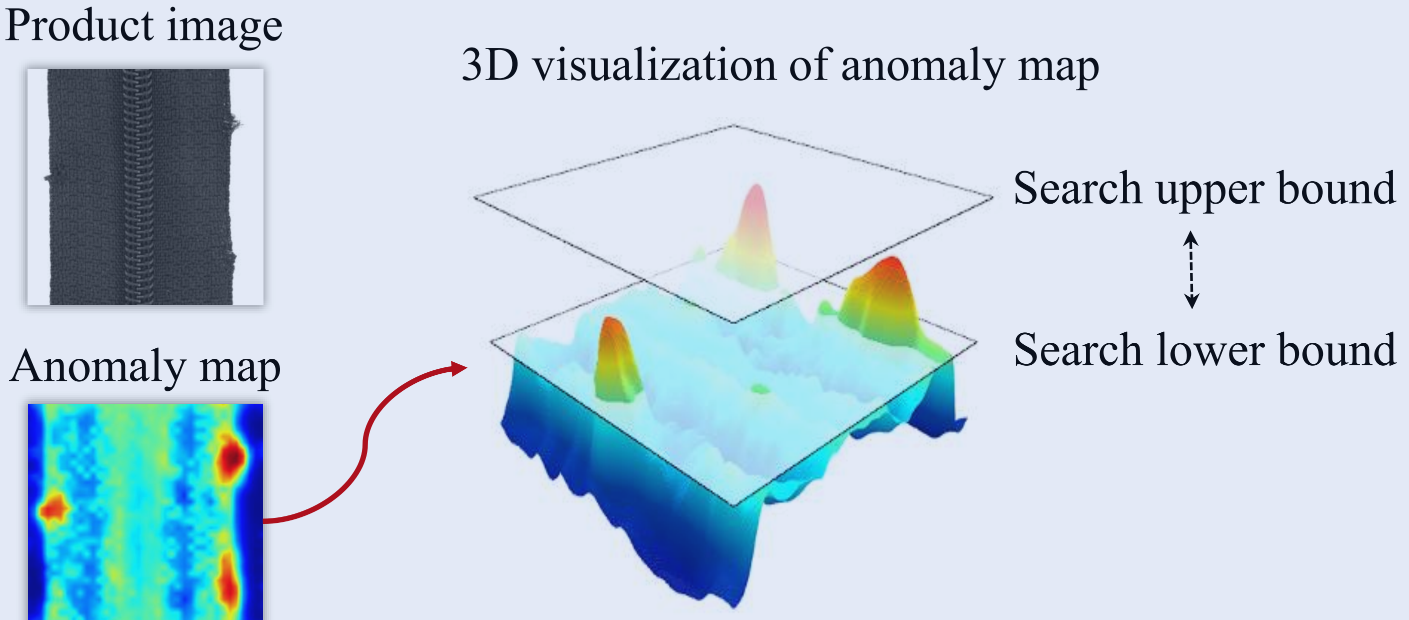
MEBin — Overview



Method

MEBin—— Setting threshold search range

Given a batch of anomaly maps to be binarized, calculate the maximum and minimum values of the anomaly scores for the images in this batch, and adaptively determine the threshold search interval based on these extreme values.

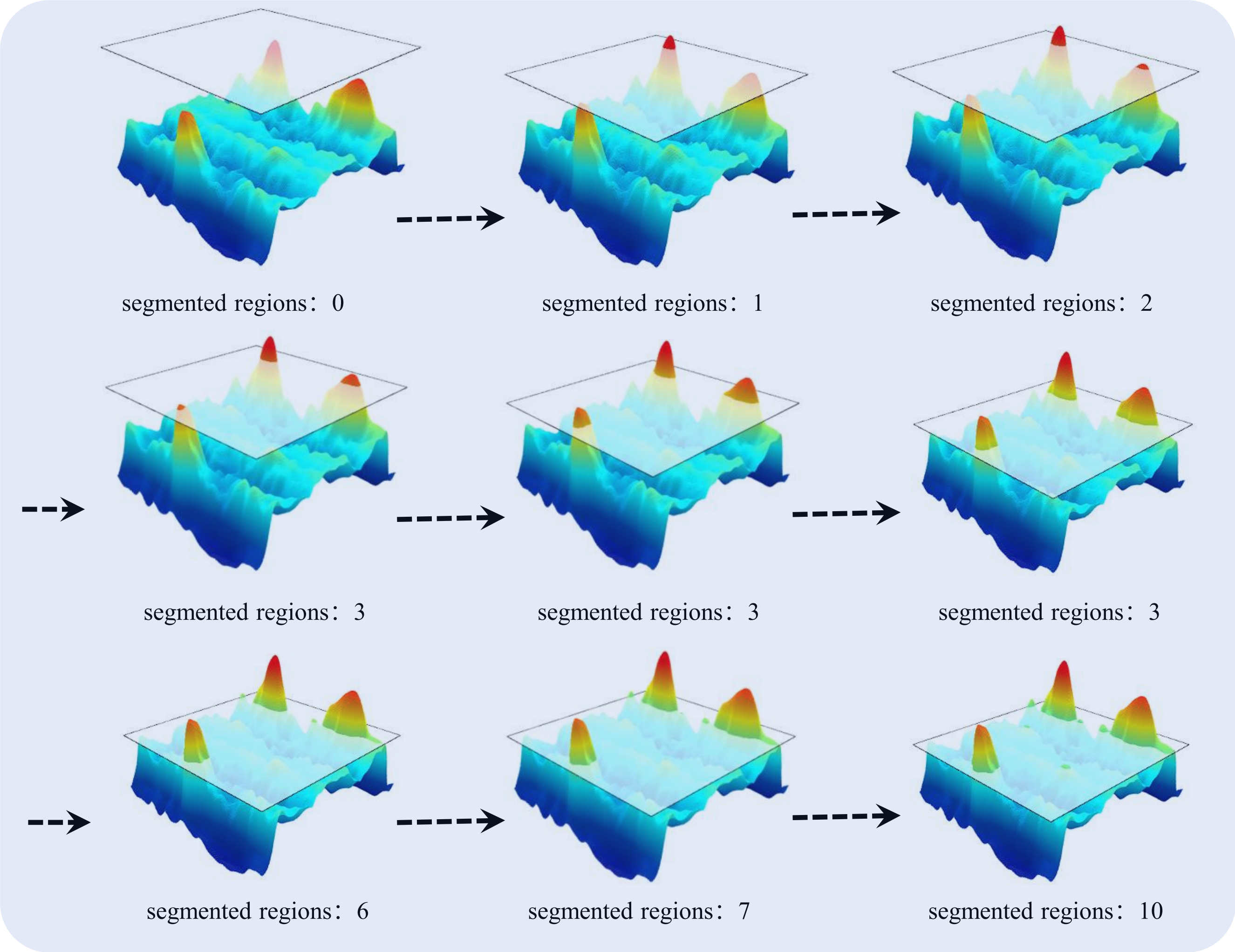


The maximum value of anomaly scores （1）

The minimum value of abnormal scores

Method

MEBin——Traverse the threshold search interval and count the number of segmented regions



Defect regions start to segment — the number of segmented regions increases

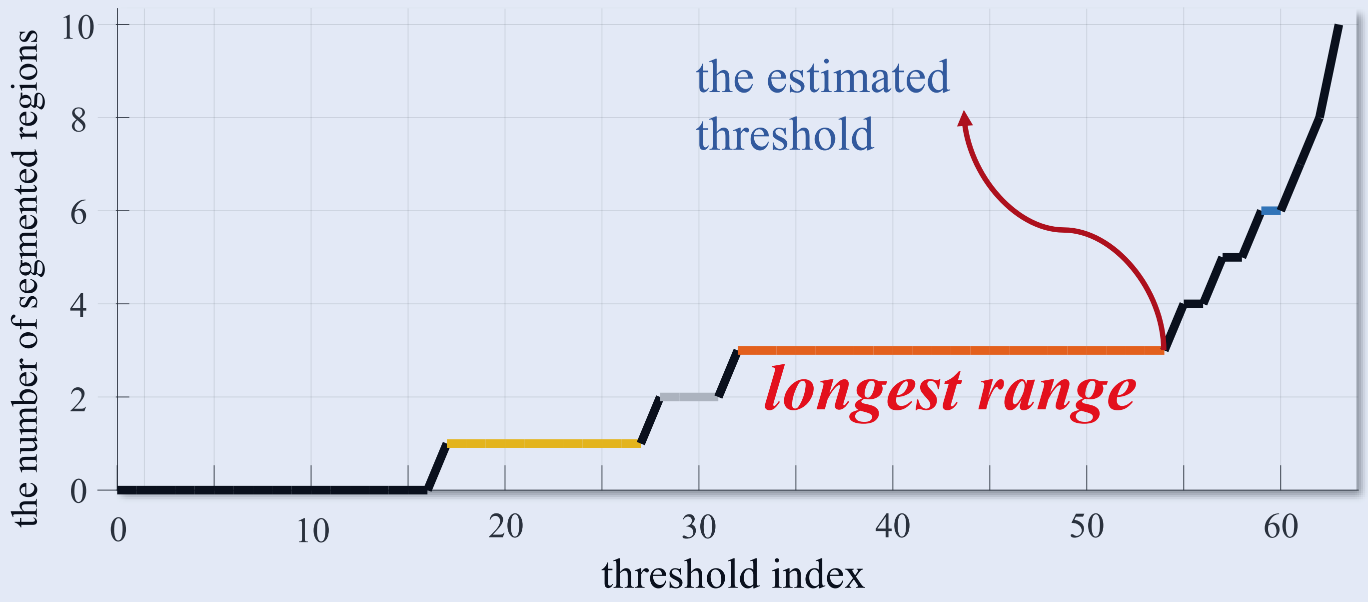
Defect regions are gradually and completely segmented — the number of segmented regions stabilizes

Noise appears in normal regions — the number of segmented regions changes

Method

MEBin——Calculate the stable interval and adaptively estimate the threshold

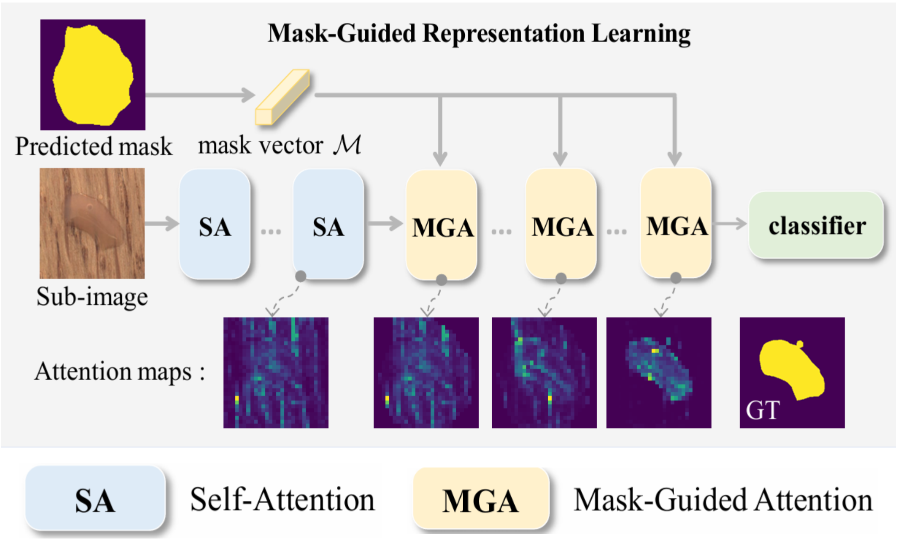
Stable interval: A threshold interval where the number of segmented regions remains constant and exceeds a predefined minimum length. Calculate all stable intervals and use the minimum threshold of the longest stable interval as the estimated threshold.



By selecting the 'minimum' threshold within the longest stable interval, we aim to reduce false negatives (missed detections).

Method

MGViT



ViT: Self Attention

$$Attn = \mathrm{softmax}\left(\mathrm{concat}\left(\boxed{\mathbf{Q}_{l-1}^{\mathrm{cls}} \mathbf{K}_{l-1}^{\top}}, \mathbf{Q}_{l-1}^{\mathrm{patch}} \mathbf{K}_{l-1}^{\top}\right)\right) \mathbf{V}_{l-1}$$

MGViT: Mask-Guided Attention

$$Attn = \mathrm{softmax}\left(\mathrm{concat}\left(\boxed{\mathbf{Q}_{l-1}^{\mathrm{cls}} \mathbf{K}_{l-1}^{\top} + \overline{\mathcal{M}}}, \mathbf{Q}_{l-1}^{\mathrm{patch}} \mathbf{K}_{l-1}^{\top}\right)\right) \mathbf{V}_{l-1}$$

Leverage mask constraints on the [CLS] token

to guide the model to focus exclusively on patches containing defects for classification.

$$\overline{\mathcal{M}}(i) = \begin{cases} 0, & \text{if } \mathcal{M}(i) > 0.5 \\ -\infty, & \text{otherwise} \end{cases}$$

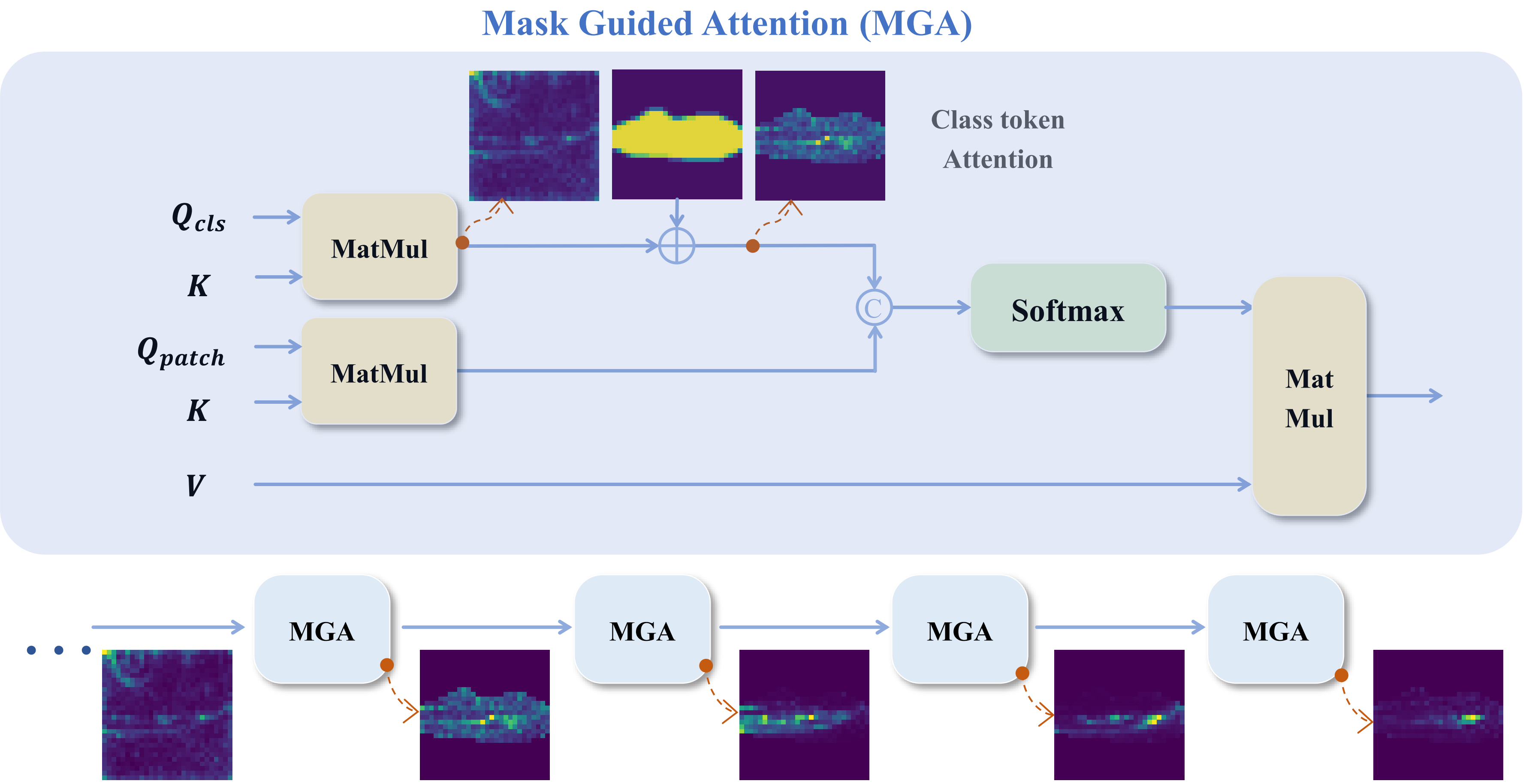
Patches containing defects retain their original attention.

For patches without defects, the corresponding attention is set to 0.

Leverage mask constraints to guide the model to focus on defects and extract features from defective regions.

Method

MGViT



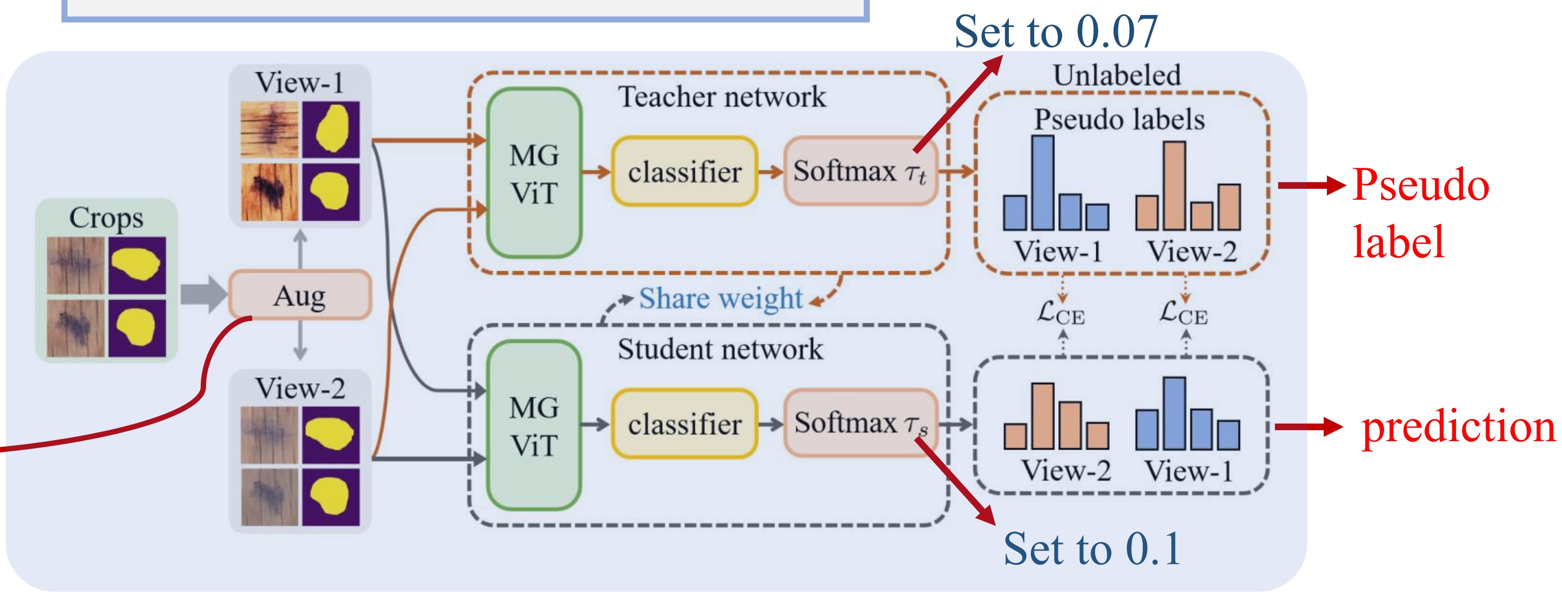
In cases where the predicted masks are rough, the features gradually focus on the anomalous regions as the guidance deepens.

Method

Representation learning

Unlabeled dataset——pseudo label generation

Data augmentation methods
1.random crop
2.horizontal flip, hflip
3.vertical flip, vflip
4.colorjitter
5.rotate
6.adjust sharpness
7.gaussian blur
8. posterize



Logits output by the linear classifier $\tilde{l}_j = \mathcal{H}(f(\tilde{x}_j))$ x_j : input image $\mathcal{H}(\cdot)$: linear classifier

Pseudo label $$\hat{q}_j^{(k)} = \frac{\exp(\hat{l}_j^{(k)}/\tau_t)}{\sum_{k=1}^{\mathcal{C}^l+\mathcal{C}^u} \exp(\hat{l}_j^{(k)}/\tau_t)}$$

Prediction $$\hat{p}_j^{(k)} = \frac{\exp(\hat{l}_j^{(k)}/\tau_s)}{\sum_{k=1}^{\mathcal{C}^l+\mathcal{C}^u} \exp(\hat{l}_j^{(k)}/\tau_s)}$$

Method

loss

Cross-entropy loss

$$\mathcal{L}_{\mathrm{CE}}(p, q) = -\sum_{c=0}^{\mathcal{C}^{\mathbf{l}}+\mathcal{C}^{\mathbf{u}}-1} p^c \log q^c$$

Classification loss for unlabeled dataset

$$\mathcal{L}_{\mathrm{cls}}^{\mathbf{u}} = \frac{1}{|B^{\mathbf{u}}|} \sum_{j \in B^{\mathbf{u}}} \left(\mathcal{L}_{\mathrm{CE}}(\boxed{\hat{q}_j}, \tilde{p}_j) + \mathcal{L}_{\mathrm{CE}}(\boxed{\tilde{q}_j}, \hat{p}_j)\right)$$

Pseudo label calculation loss

Classification loss for auxiliary dataset

$$\mathcal{L}_{\mathrm{cls}}^{\mathbf{l}} = \frac{1}{|B^{\mathbf{l}}|} \sum_{j \in B^{\mathbf{l}}} \left(\mathcal{L}_{\mathrm{CE}}(\boxed{\hat{y}_j}, \tilde{p}_j) + \mathcal{L}_{\mathrm{CE}}(\boxed{\tilde{y}_j}, \hat{p}_j)\right)$$

Ground truth calculation loss

Regularization loss

$$\mathcal{L}_{\mathrm{reg}}^{\mathbf{u}} = \mathcal{L}_{\mathrm{CE}}(\bar{p}_j, \bar{p}_j) \qquad \bar{p}_j = \frac{1}{2|B^{\mathbf{u}}|} \sum_{j \in B^{\mathbf{u}}} (\hat{p}_j + \tilde{p}_j)$$

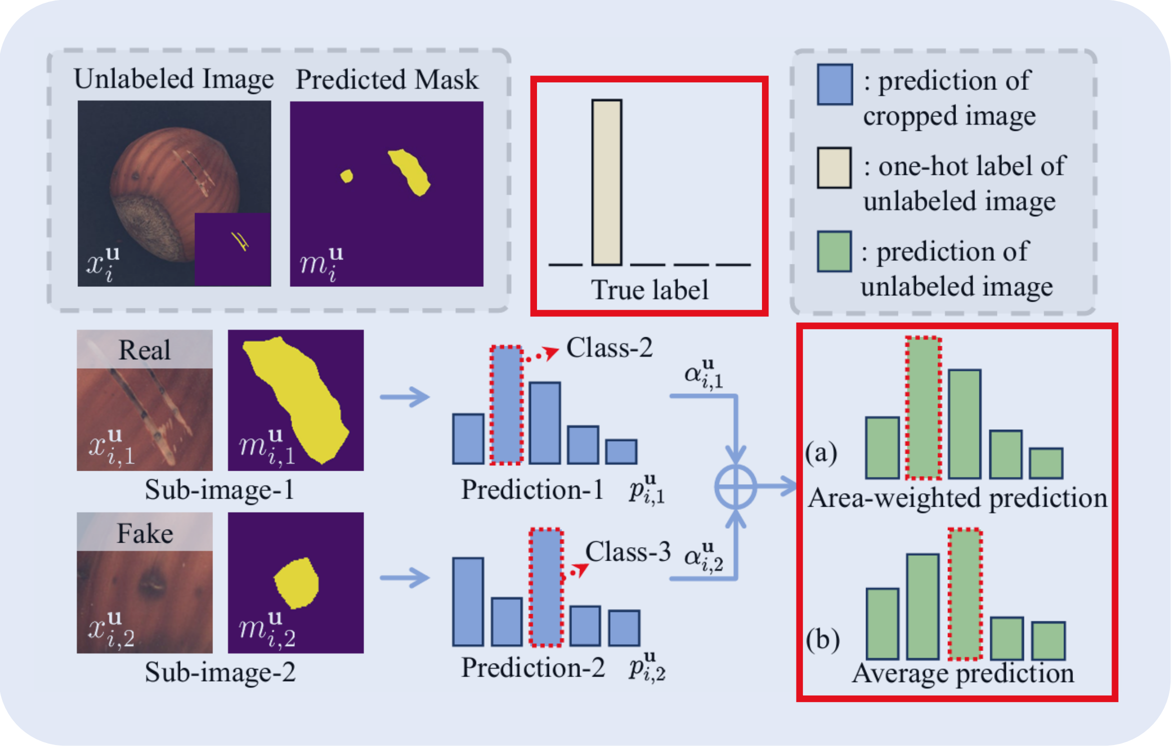
Overall training objective

$$\mathcal{L} = \lambda(\mathcal{L}_{\mathrm{rep}}^{\mathbf{l}} + \mathcal{L}_{\mathrm{cls}}^{\mathbf{l}}) + (1-\lambda)(\mathcal{L}_{\mathrm{rep}} + \mathcal{L}_{\mathrm{cls}}^{\mathbf{u}} + \mu \mathcal{L}_{\mathrm{reg}}^{\mathbf{u}})$$

Contrastive loss + Classification loss

Method

Region Merging



The prediction result of the original image is the weighted sum of the prediction results of sub-images.

$$p_i^{\mathbf{u}} = \sum_{k=1}^{N} \alpha_{i,k}^{\mathbf{u}} p_{i,k}^{\mathbf{u}}$$

Sub-image weights

$$\alpha_{i,k}^{\mathbf{u}} = \frac{\exp(\sqrt{\boxed{\mathbf{a}_{i,k}^{\mathbf{u}}}}/\tau_\alpha)}{\sum_{k=1}^{\bar{\delta}_i} \exp(\sqrt{\boxed{\mathbf{a}_{i,k}^{\mathbf{u}}}}/\tau_\alpha)}$$

Sub-image defect area (number of defective pixels)

The larger the area, the greater the weight, which suppresses the influence of small-area false detections on the final classification results.

Experiment

Quantitative experimental results（MVTec AD、MTD）

Datasets	Metric	IIC[23]	GATCluster[31]	SCAN[42]	UNO[13]	GCD[43]	SimGCD[45]	AMEND[3]	AC[38] (Unsup.)	MuSc [28] +AnomalyNCD
MVTec AD [5]	NMI	0.093	0.136	0.210	0.146	0.417	0.452	0.431	0.525	**0.613**
	ARI	0.020	0.053	0.103	0.052	0.302	0.346	0.333	0.431	**0.526**
	F_1	0.285	0.264	0.335	0.342	0.553	0.569	0.542	0.604	**0.712**
MTD [21]	NMI	0.064	0.028	0.041	0.034	0.211	0.105	0.138	0.179	**0.268**
	ARI	0.020	0.009	0.029	0.011	0.115	0.048	0.067	0.120	**0.228**
	F_1	0.252	0.243	0.282	0.221	0.381	0.293	0.324	0.346	**0.509**

Without using normal samples from the training set

MVTec AD
8.8%NMI improvement
9.5% ARI improvement
10.8% F1 improvement

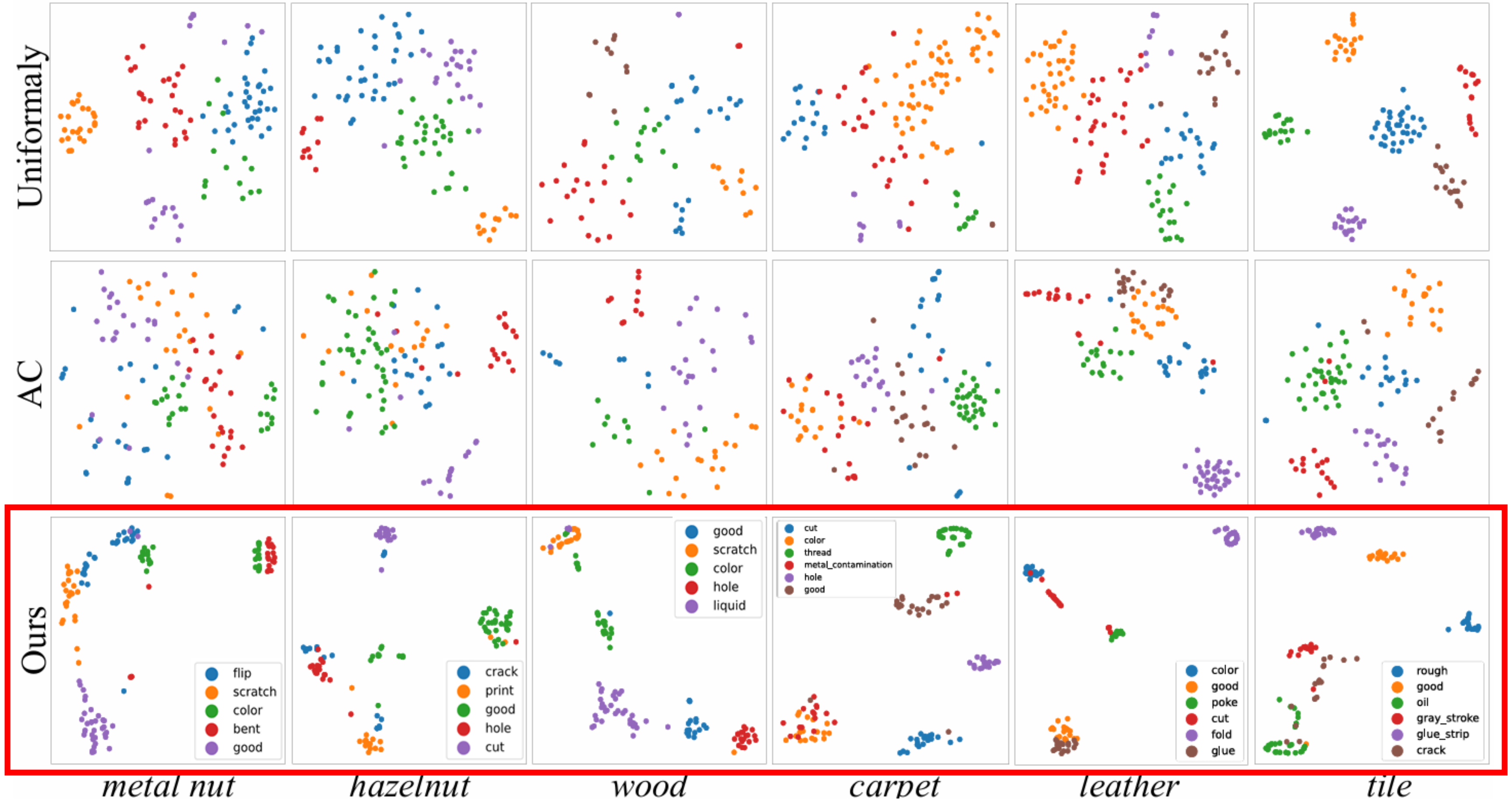
Datasets	Metric	AC[38] (Semi-sup.)	UniFormaly[26] (0.953 / 0.837)	PatchCore[36] (0.938 / 0.729) +AnomalyNCD	RD++[41] (0.950 / 0.741) +AnomalyNCD	EfficientAD[4] (0.917 / 0.731) +AnomalyNCD	PNI[2] (0.942 / 0.516) +AnomalyNCD	CPR[27] (0.964 / -) +AnomalyNCD
MVTec AD [5]	NMI	0.608	0.547	0.670	0.631	0.516	0.675	**0.736**
	ARI	0.489	0.433	0.601	0.542	0.394	0.609	**0.674**
	F_1	0.652	0.645	0.769	0.721	0.641	0.769	**0.805**
MTD [21]	NMI	0.390	**0.421**	0.380	0.368	0.220	0.181	-
	ARI	0.314	0.322	**0.390**	0.361	0.188	0.219	-
	F_1	0.490	0.609	**0.617**	0.600	0.467	0.465	-

Using normal samples from the training set

MVTec AD
12.8%NMI improvement
18.5% ARI improvement
15.3% F1 improvement

Experiment

Qualitative experimental results--t-SNE visualization of features

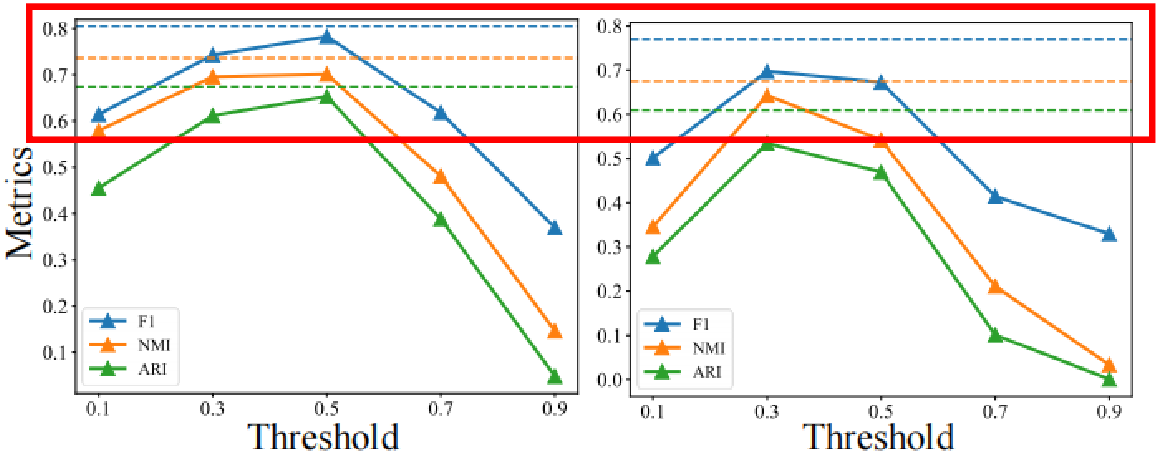


Novel anomaly class discovery in industrial scenarios generates features with smaller intra-class distances and larger inter-class distances.

Experiment

Ablation—MEBin

	Avg FPR ↓	Avg FNR ↓	NMI ↑	ARI ↑	F_1 ↑
$\epsilon = 0.1$	0.572	0.617	0.554	0.482	0.650
$\epsilon = 0.3$	0.678	0.742	0.499	0.404	0.587
$\epsilon = 0.5$	0.484	0.165	0.567	0.458	0.640
$\epsilon = 0.7$	0.269	0.247	0.495	0.395	0.623
$\epsilon = 0.9$	0.544	0.593	0.077	0.013	0.337
Otsu [33]	0.676	0.525	0.382	0.268	0.499
Ours	**0.153**	**0.035**	**0.613**	**0.526**	**0.712**

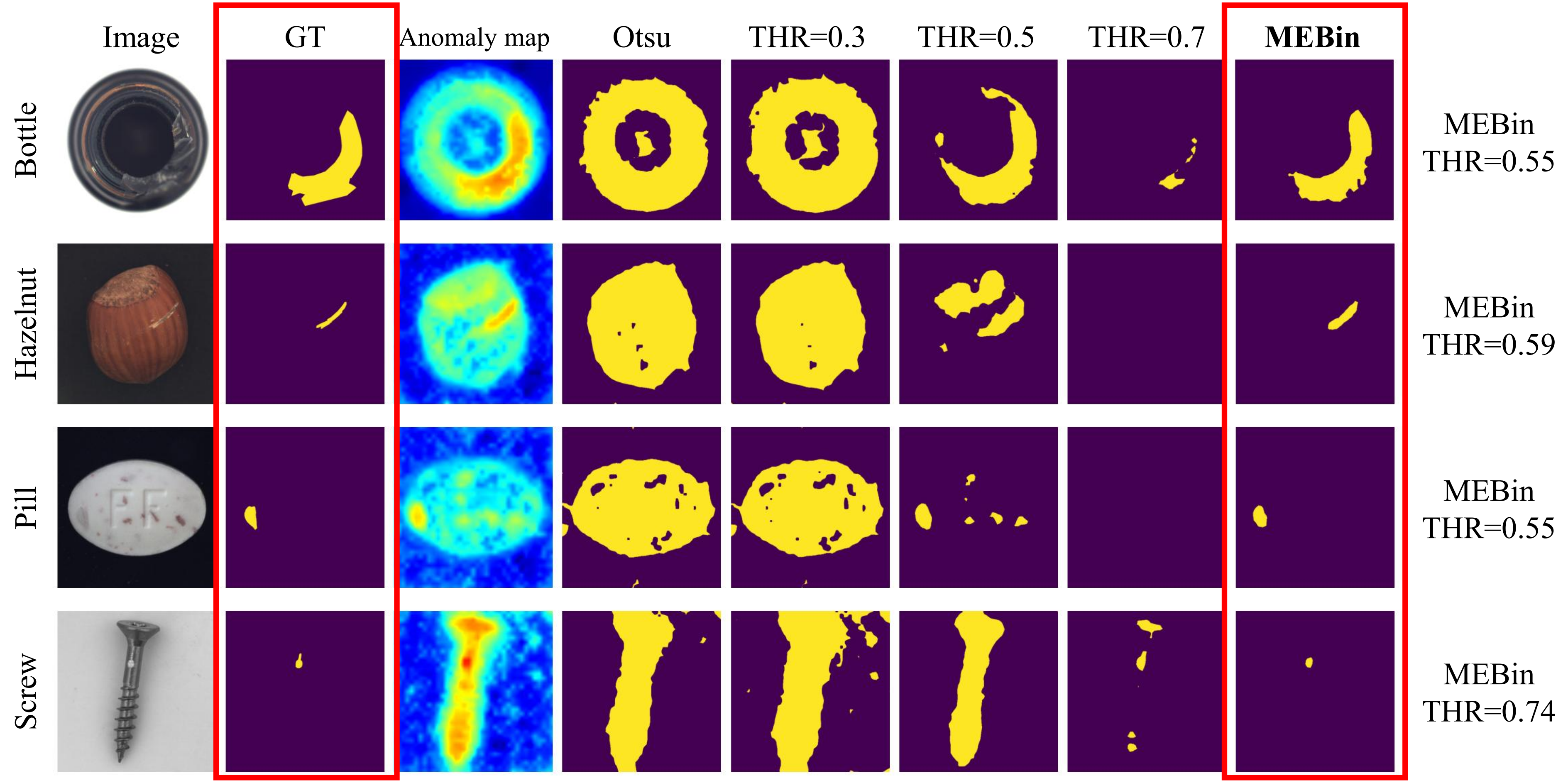


(a) CPR (b) PNI

MEBin achieves the lowest false positive and false negative rates, leading to the optimal overall classification performance of the model and demonstrating adaptability to various defect detection algorithms.

Experiment

Ablation — Qualitative results of MEBin



The binarization results of MEBin are closest to the ground-truth annotations.

Experiment

Ablation—MEBin

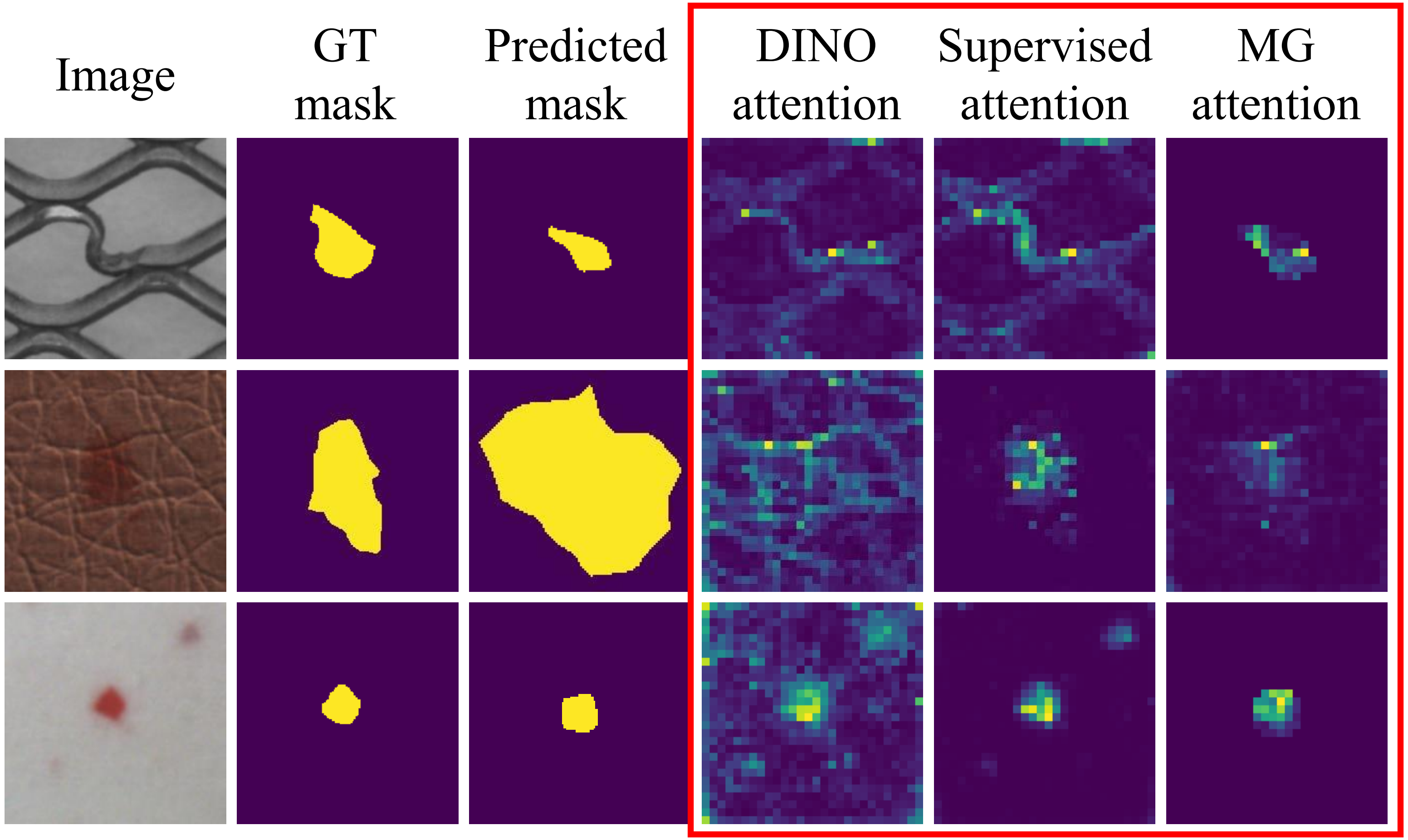
	MVTec AD			MTD		
L_m	NMI	ARI	F_1	NMI	ARI	F_1
1	0.606	0.508	0.690	0.191	0.157	0.420
3	0.608	0.511	0.694	0.209	0.173	0.436
6	**0.613**	0.519	**0.713**	0.253	0.214	0.492
9	**0.613**	**0.526**	0.712	**0.268**	**0.228**	**0.509**
12	0.609	0.521	0.712	0.249	0.213	0.492

Mask Mechanism	NMI	ARI	F_1
(a) w/o MGA	0.598	0.494	0.698
(b) all tokens	0.507	0.382	0.600
(c) patch tokens	0.563	0.467	0.686
(d) class token (Ours)	**0.613**	**0.526**	**0.712**

The model achieves the optimal classification metrics when using the mask-guided attention (MGA) approach.

Experiment

Ablation—MGViT



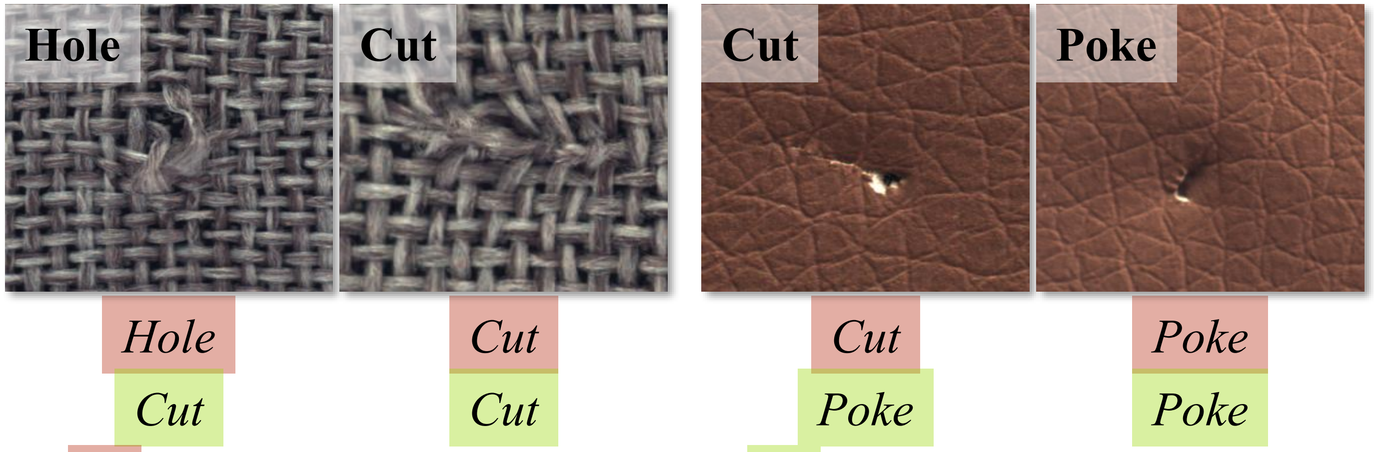
Using an unsupervised setting, it achieves defect attention effects close to those of supervised methods, saving significant human and material resources.

Experiment

Ablation—auxiliary dataset

	MVTec AD			MTD		
	NMI	ARI	F_1	NMI	ARI	F_1
w/o $\mathcal{D}^l$	0.583	0.506	0.689	0.227	0.202	0.485
w $\mathcal{D}^l$	**0.613**	**0.526**	**0.712**	**0.268**	**0.228**	**0.509**

Using auxiliary datasets can improve the overall performance of the model.



*Results **with** labled data* *Results **without** labled data*

With the aid of labeled defect classes, it can effectively distinguish **different defect classes** with **similar appearances**.

These defects are sometimes not easily distinguishable by the human eye.

Experiment

Ablation—Region Merging

Merge	MVTec AD			MTD		
	NMI	ARI	F_1	NMI	ARI	F_1
(i) Avg	0.610	0.521	0.709	0.257	0.223	0.501
(ii) Score Avg	0.600	0.513	0.703	0.228	0.208	0.485
(iii) Area Avg	**0.613**	**0.526**	**0.712**	**0.268**	**0.228**	**0.509**

Using averaging predictions of sub-images , the model achieves the best overall performance, surpassing direct averaging and anomaly score-weighted averaging.

Acknowledgments

Sincere thanks to all members of the SLOW team !

Ziming Huang

Huazhong University
of Science and Technology

zmhuang@hust.edu.cn

Xurui Li

Huazhong University
of Science and Technology

xrli_plus@hust.edu.cn

Haotian Liu

Huazhong University
of Science and Technology

htliu_master@hust.edu.cn

Feng Xue

University of Trento

feng.xue@unitn.it

Yuzhe Wang

Huazhong University
of Science and Technology

yuzw@hust.edu.cn

Yu Zhou

Huazhong University
of Science and Technology,
Wuhan JingCe Electronic Group Co., LTD

yuzhou@hust.edu.cn

Sincere thanks to Wuhan Jingce Electronic Group Co., Ltd. for its long-term funding of the research group's basic research in the field of industrial vision!

JINGCE ELECTRONIC 精测电子 | 武汉精测电子集团股份有限公司 WUHAN JINGCE ELECTRONIC GROUP CO.,LTD

華中科技大學

Huazhong University of Science and Technology

CVPR Nashville JUNE 11-15, 2025

武汉精测电子集团股份有限公司

WUHAN JINGCE ELECTRONIC GROUP CO.,LTD

Thank you for your attention!
Your corrections are sincerely appreciated!

Welcome to download and use!

Paper and Code : https://github.com/HUST-SLOW