華中科技大學
Huazhong University of Science and Technology

MuSc: Zero-Shot Industrial Anomaly Classification and Segmentation with Mutual Scoring of the Unlabeled Images

Authors: Xurui Li*, Ziming Huang*, Feng Xue, Yu Zhou†

Speaker: Xurui Li

* Contributed Equally
† Corresponding Authors

Characteristics of industrial anomalies

Structural anomalies

e.g. contamination, crack

Break the regular structure

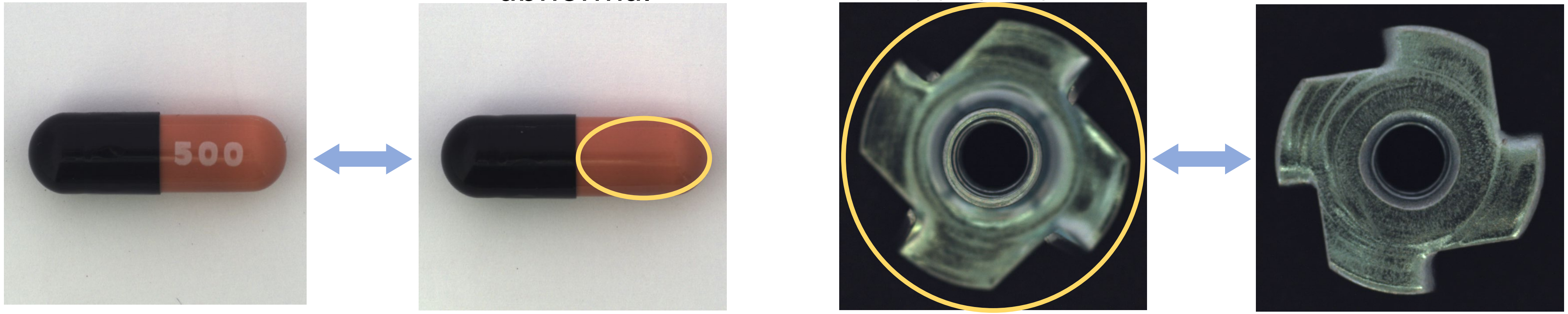
Current methods are based on the text-to-image of CLIP

Characteristics of industrial anomalies

Logical anomalies

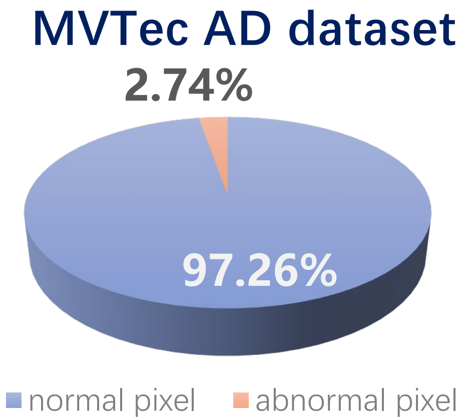
e.g. missing, flip

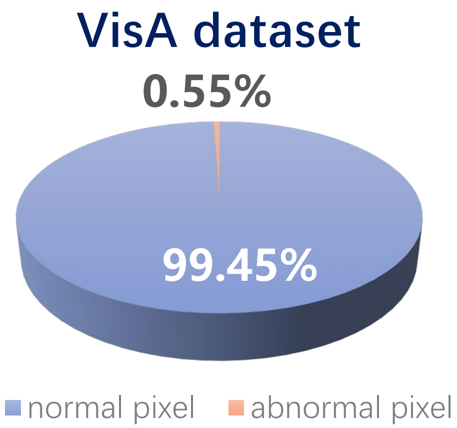
normal | abnormal | abnormal | normal



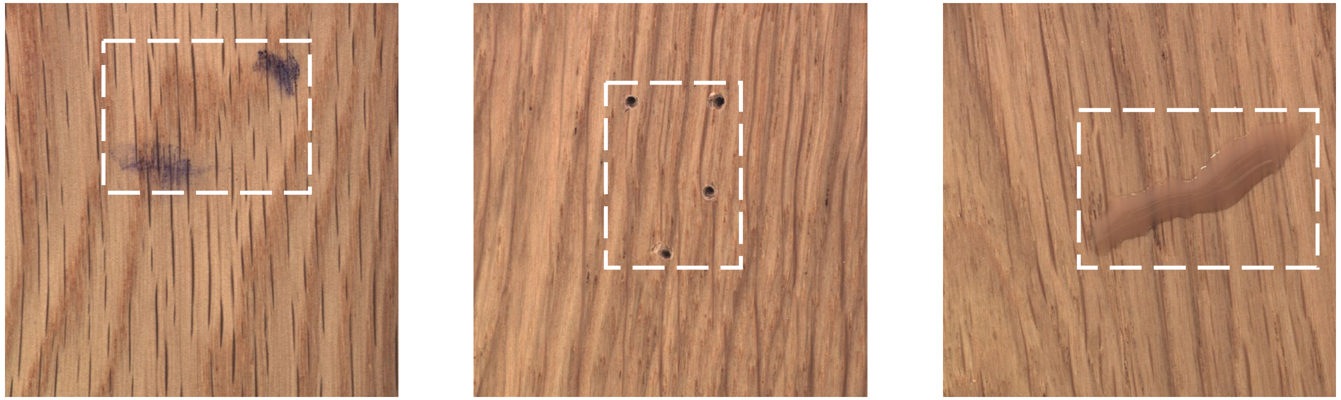
Need to be compared with normal

Motivation





Enrich normal prior information

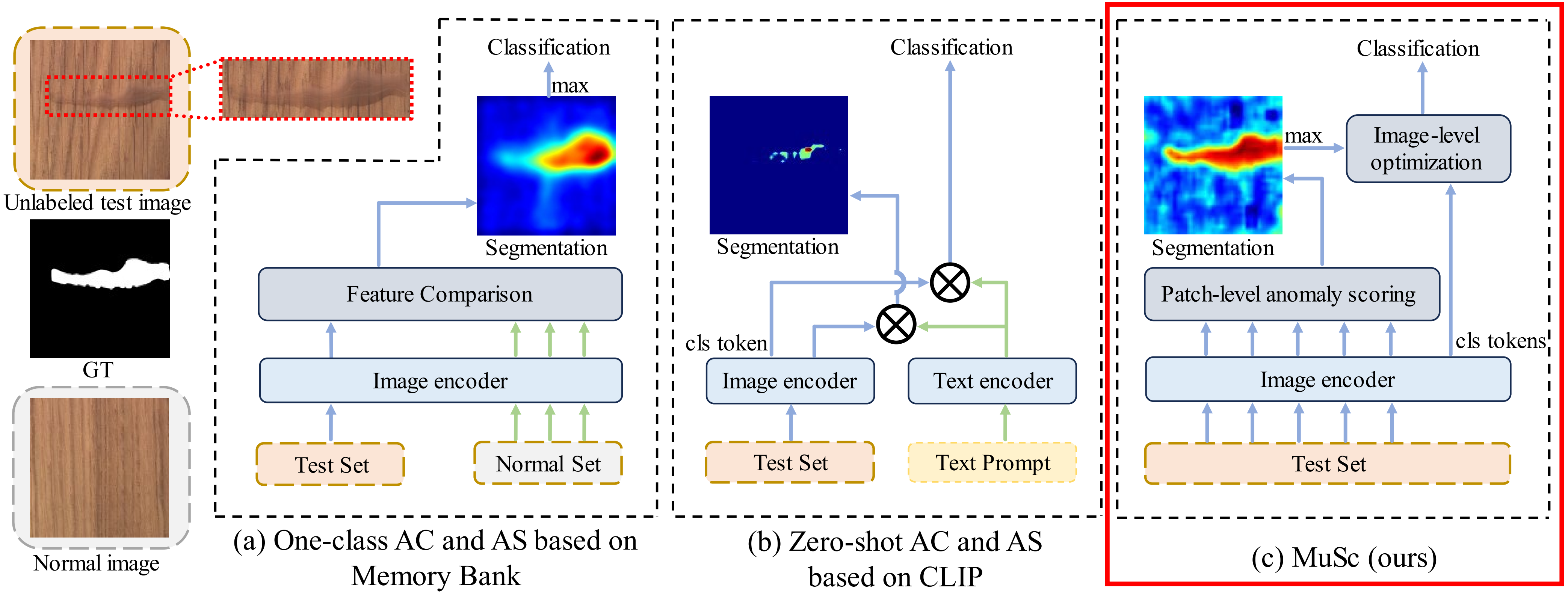


Discriminative characteristic

Normal regions could find similar regions in other unlabeled images

Abnormal regions only have a few similar regions

Mutual Scoring (MuSc)

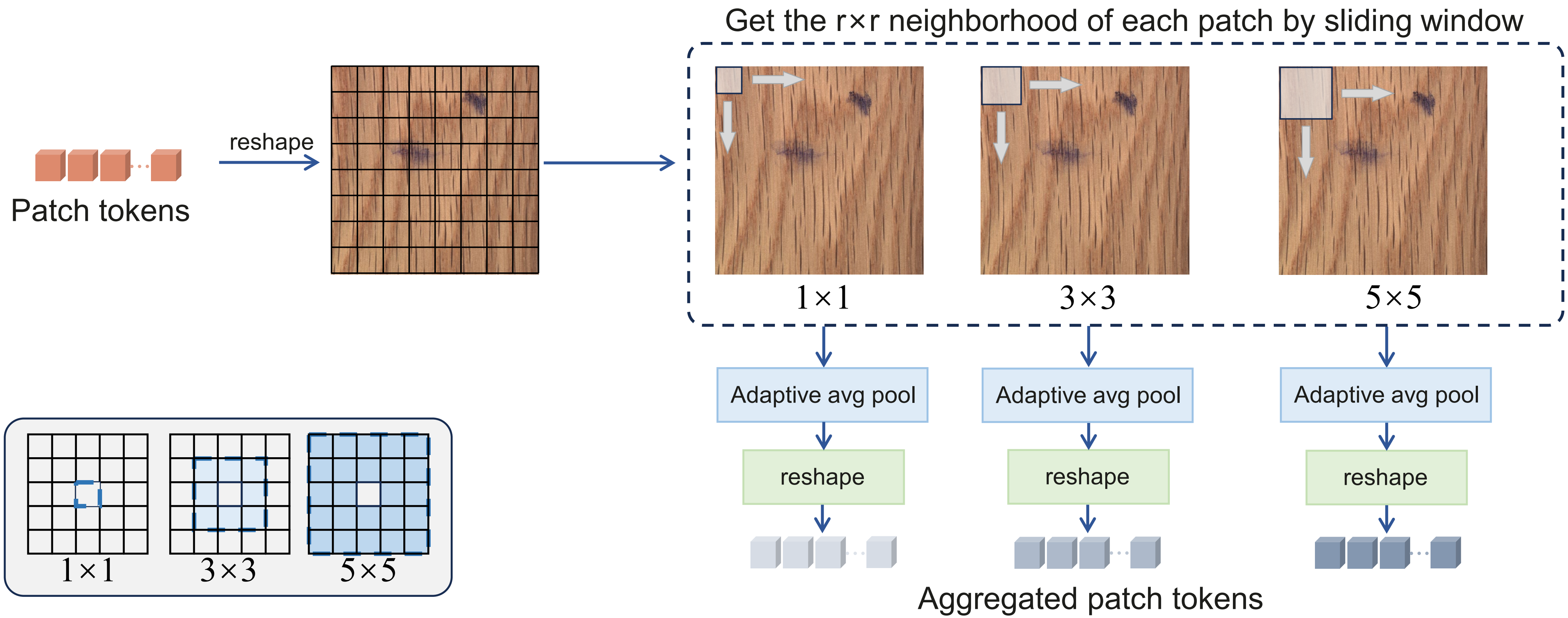


No training **No text prompt**

No additional normal reference images

Method--LNAMD

1. Local Neighborhood Aggregation with Multiple Degrees (LNAMD)

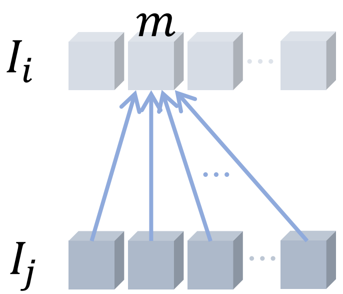


Method--MSM

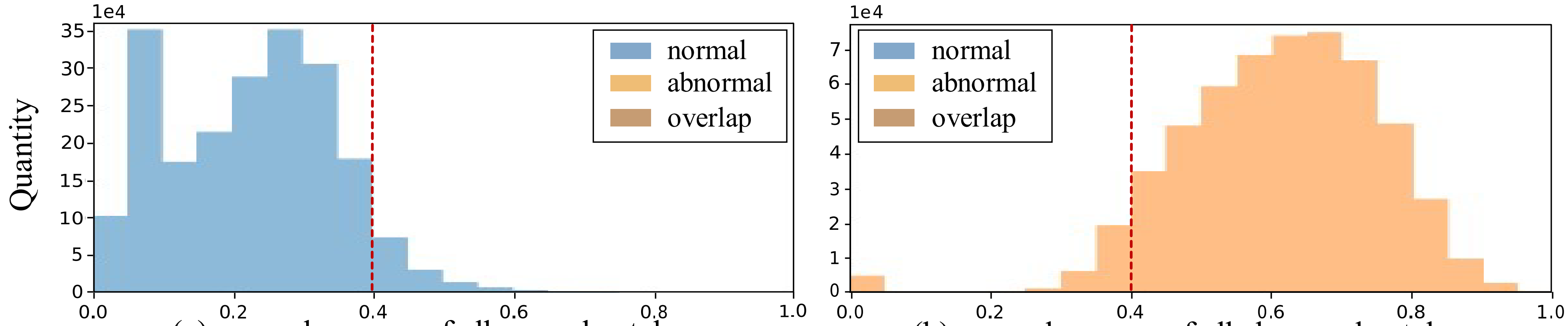
2. Mutual Scoring Mechanism (MSM)

Aggregated patch tokens： $\hat{p}_{i,l}^{m,r}$

Image I_j assigns an anomaly score for each patch token of I_i



$$a_{i,l}^{m,r}(I_j) = \min_{n} \|\hat{p}_{i,l}^{m,r} - \hat{p}_{j,l}^{n,r}\|_2$$



(a) anomaly scores of all normal patches

(b) anomaly scores of all abnormal patches

Method--RsCIN

3. Classification Re-Scoring with Constrained Image-level Neighborhood

ViT extracts the image-level features of image I_{i} as F_i
Similarity matrix W, $W_{i,j} = F_i \cdot F_j$

Multi-window Mask Operation (MMO) is used to constrain the re-scoring images

$$M_k(i,j) = \begin{cases} 1, & I_j \in N_k(I_i) \\ 0, & I_j \notin N_k(I_i) \end{cases}$$

$$\bar{M} = \{M_{k_1}, M_{k_2} \dots, M_{k_K}\},\ k \in \{k_1, k_2, \dots, k_K\},\ \ K \text{ is the number of window masks}$$

Re-Scoring

$$\hat{C} = \Big(\sum_{M_k \in \bar{M}} (D^{-1}(M_k \odot W)C) + C\Big)/(K+1)$$

$$D(i,i) = \sum_{k=1}^{n} M_k \odot W(i,k)$$

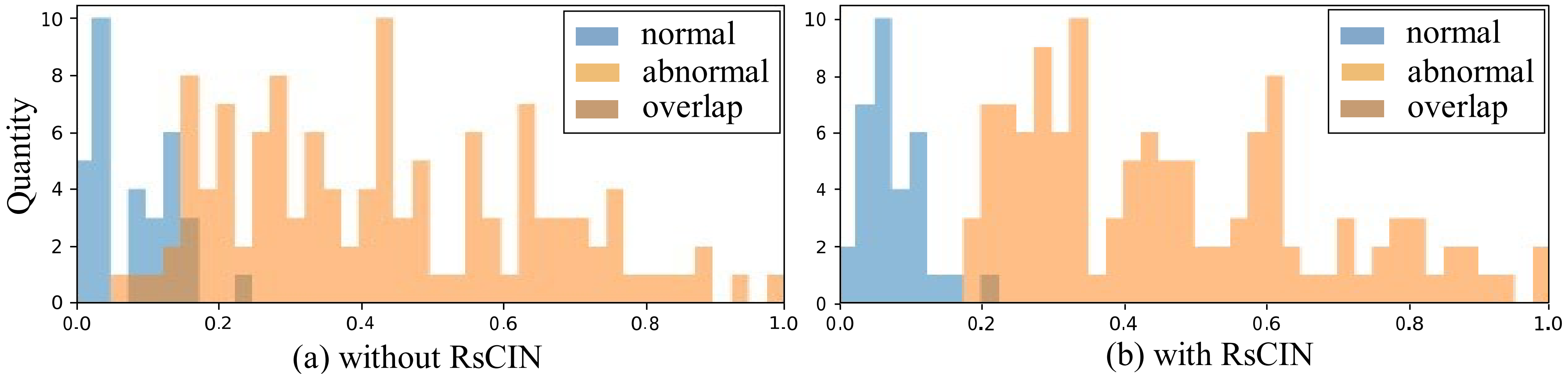
Method--RsCIN

3. Classification Re-Scoring with Constrained Image-level Neighborhood

$$\hat{c}_i = \frac{c_i}{3} + \boxed{\frac{1}{3}\sum_{j=1}^{k_2} \overline{w}_{i,j}\overline{c}_j}$$

Weighted average

$$\overline{w}_{i,j} = \begin{cases} \hat{w}_{i,j}^{k_1} + \hat{w}_{i,j}^{k_2}, & \text{if } 0 < j \leq k_1 \\ \hat{w}_{i,j}^{k_2}, & \text{if } k_1 < j \leq k_2 \end{cases}$$

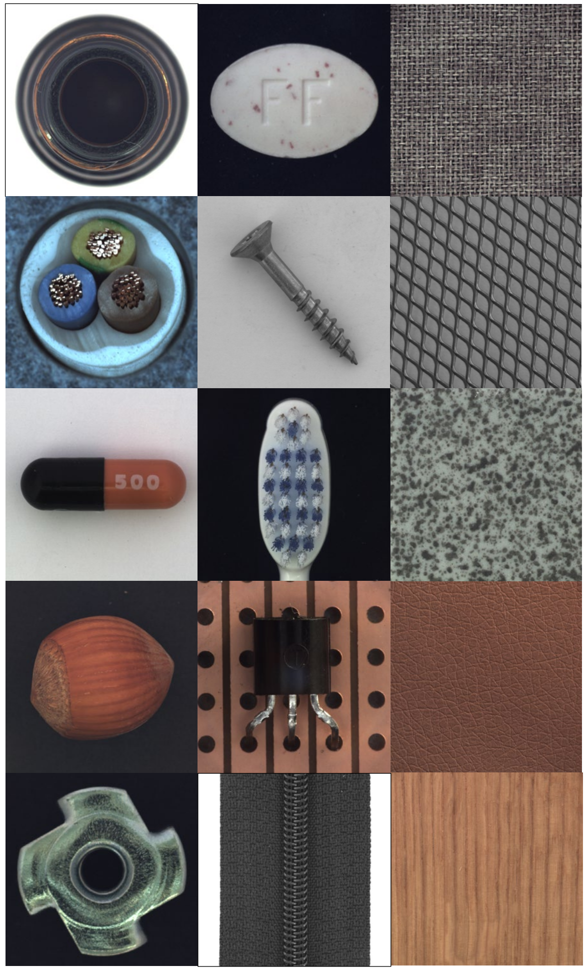


Less overlap between the scores of normal images and those of abnormal scores

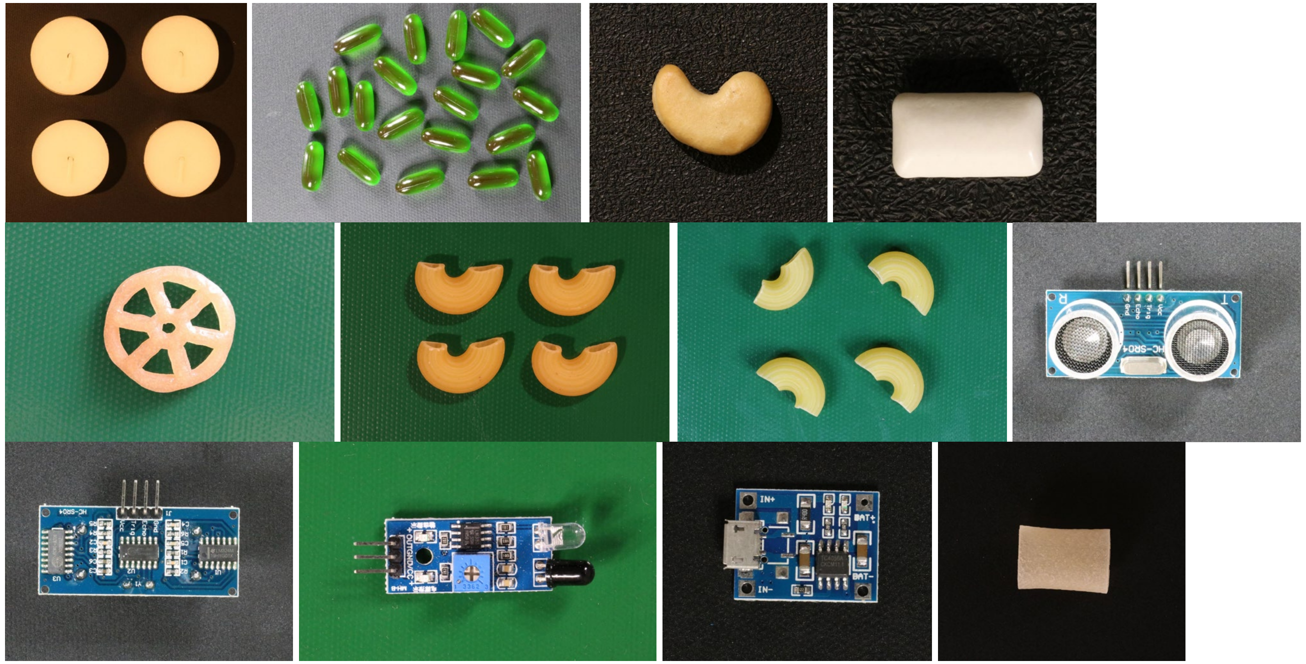
Experiments

1. Datasets

MVTec AD



VisA



Experiments

2. Quantitative results

Comparison with zero-shot methods

Dataset	Method	Setting	AUROC-cls	F1-max-cls	AP-cls	AUROC-segm	F1-max-segm	AP-segm	PRO-segm
MVTec AD	WinCLIP	0-shot	91.8	92.9	96.5	85.1	31.7	-	64.6
	APRIL-GAN	0-shot	86.1	90.4	93.5	87.6	43.3	40.8	44.0
	ACR	0-shot	85.8	91.3	92.9	92.5	44.2	38.9	72.7
	MuSc (ours)	0-shot	**97.8(+6.0)**	**97.5(+4.6)**	**99.1(+2.6)**	**97.3(+4.8)**	**62.6(+18.4)**	**62.7(+21.9)**	**93.8(+21.1)**
VisA	WinCLIP	0-shot	78.1	79.0	81.2	79.6	14.8	-	56.8
	APRIL-GAN	0-shot	78.0	78.7	81.4	94.2	32.3	25.7	86.8
	MuSc (ours)	0-shot	**92.8(+14.7)**	**89.5(+10.5)**	**93.5(+12.1)**	**98.8(+4.6)**	**48.8(+16.5)**	**45.1(+19.4)**	**92.7(+5.9)**

MVTec AD **MuSc abtains 21.1% PRO gains and 21.9% seg-AP gains**

VisA **MuSc abtains 19.4% seg-AP gains and 14.7% cls-AUROC gains**

Experiments

2. Quantitative results

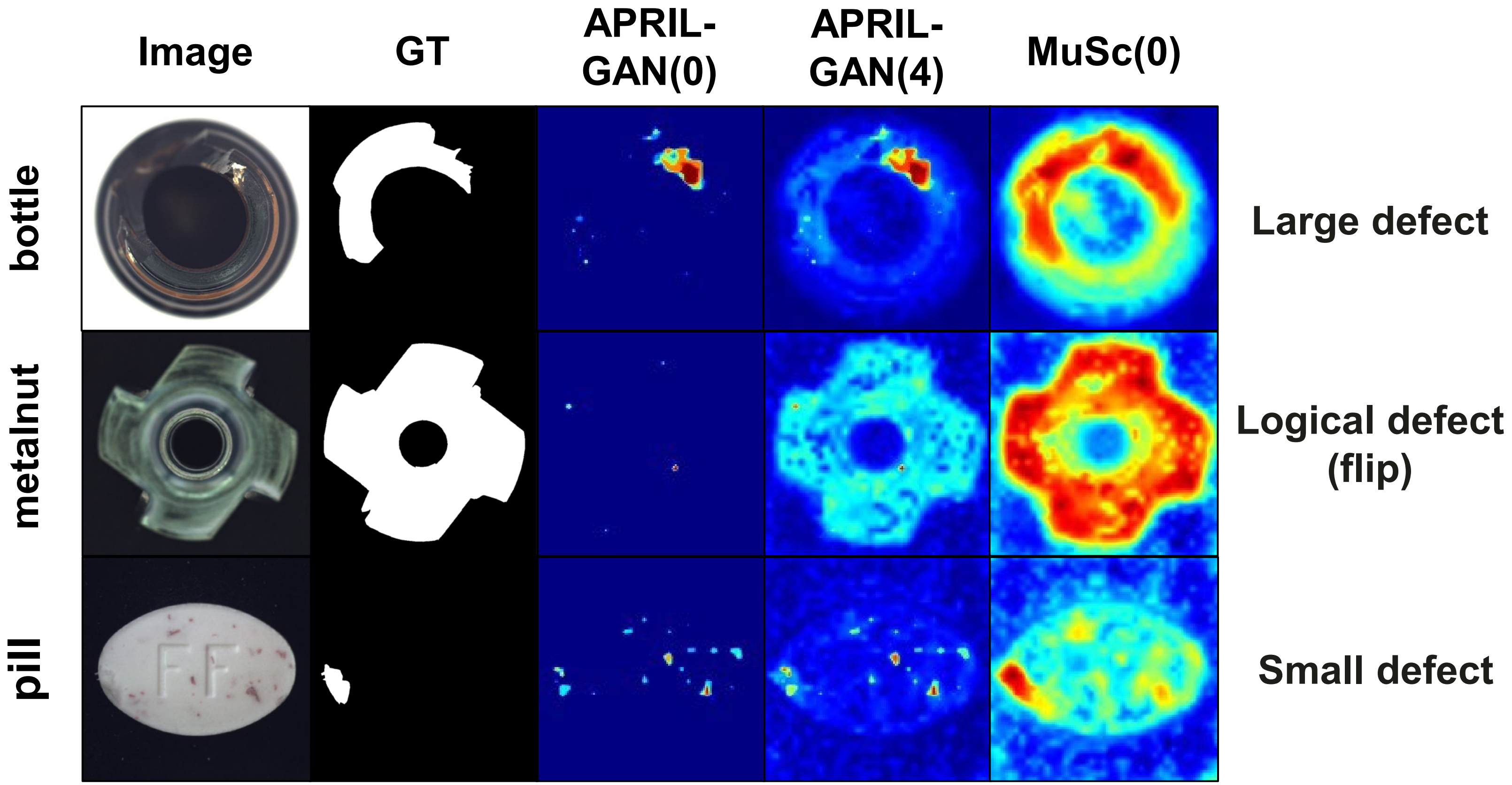
Comparison with few-shot methods

Dataset	Method	Setting	AUROC-cls	F1-max-cls	AP-cls	AUROC-segm	F1-max-segm	AP-segm	PRO-segm
MVTec AD	RegAD	4-shot	89.1	92.4	94.9	96.2	51.7	48.3	88.0
	PatchCore	4-shot	88.8±2.6	92.6±1.6	94.5±1.5	94.3±0.5	55.0±1.9	-	84.3±1.6
	WinCLIP	4-shot	95.2±1.3	94.7±0.8	97.3±0.6	96.2±0.3	59.5±1.8	-	89.0±0.8
	APRIL-GAN	4-shot	92.8±0.2	92.8±0.1	96.3±0.1	95.9±0.0	56.9±0.1	54.5±0.2	91.8±0.1
	GraphCore	4-shot	92.9	-	-	**97.4**	-	-	-
	MuSc (ours)	0-shot	**97.8**	**97.5**	**99.1**	97.3	**62.6**	**62.7**	**93.8**
VisA	PatchCore	4-shot	85.3±2.1	84.3±1.3	87.5±2.1	96.8±0.3	43.9±3.1	-	84.9±1.4
	WinCLIP	4-shot	87.3±1.8	84.2±1.6	88.8±1.8	97.2±0.2	47.0±3.0	-	87.6±0.9
	APRIL-GAN	4-shot	92.6±0.4	88.4±0.5	**94.5±0.3**	96.2±0.0	40.0±0.4	32.2±0.1	90.2±0.1
	MuSc (ours)	0-shot	**92.8**	**89.5**	93.5	**98.8**	**48.8**	**45.1**	**92.7**

MuSc outperforms most of the few-shot approaches

Experiments

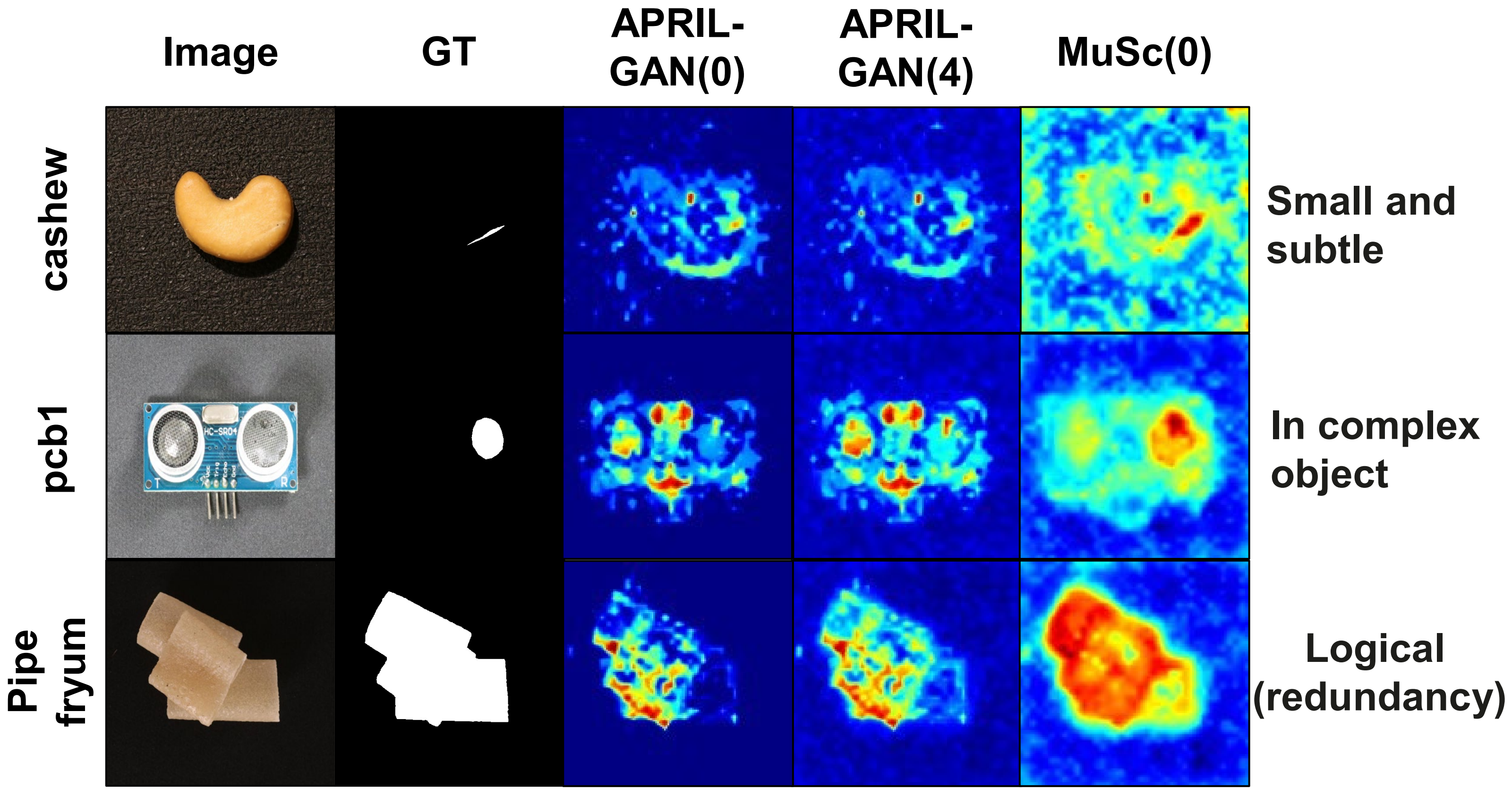
3. Qualitative result



Detect various types of defects

Experiments

3. Qualitative result



Detect various types of defects

Experiments

RsCIN

Unsupervised methods
- Based on ST, e.g. STPM
- Based on memory, e.g. SPADE and PatchCore
- Based on reconstruction, e.g. DRAEM and DSR

Few-shot methods
- Based on ST, e.g. RegAD
- Based on CLIP, e.g. APRIL-GAN

Zero-shot methods APRIL-GAN

Method	RsCIN	AUROC	F1-max	AP	Method	RsCIN	AUROC	F1-max	AP
SPADE	w/o	85.4	90.1	93.6	PatchCore	w/o	99.0	98.4	99.7
	w	**87.0**	**91.4**	**94.3**		w	**99.1**	**98.4**	**99.7**
DRAEM	w/o	98.0	97.0	99.0	DSR	w/o	98.2	96.6	99.1
	w	97.9	**97.0**	**99.1**		w	**98.2**	**96.8**	**99.3**
STPM	w/o	94.9	95.8	98.2	RegAD(2-shot)	w/o	84.8	90.7	92.5
	w	**95.6**	**96.5**	**98.5**		w	**86.2**	**91.6**	**93.1**
APRIL-GAN(0-shot)	w/o	86.1	90.4	93.5	RegAD(4-shot)	w/o	89.1	92.4	94.9
	w	**86.1**	**90.8**	**93.7**		w	**91.0**	**93.5**	**95.8**
APRIL-GAN(4-shot)	w/o	92.8	92.8	96.3	RegAD(8-shot)	w/o	91.2	92.9	95.7
	w	**93.4**	**93.1**	**96.8**		w	**92.1**	**94.0**	**96.0**
APRIL-GAN* (0-shot)	w/o	78.0	78.7	81.4	APRIL-GAN* (4-shot)	w/o	92.6	88.4	94.5
	w	**78.7**	**80.1**	**82.0**		w	**94.5**	**90.5**	**95.8**

RsCIN can effectively improve the classification results of other existing methods

華中科技大學
Huazhong University of Science and Technology

Code

https://github.com/xrli-U/MuSc

Thank you !