Focus the Discrepancy: Intra- and Inter-Correlation Learning for Image Anomaly Detection

ICCV 2023

Xincheng Yao
Shanghai Jiao Tong University

WED-PM

Coauthors:
Ruoqi Li, Zefeng Qian, Yan Luo, Chongyang Zhang*

Preview

ICCV23 PARIS

SHANGHAI JIAO TONG UNIVERSITY

How humans recognize anomalies?

Larger patch-wise discrepancies

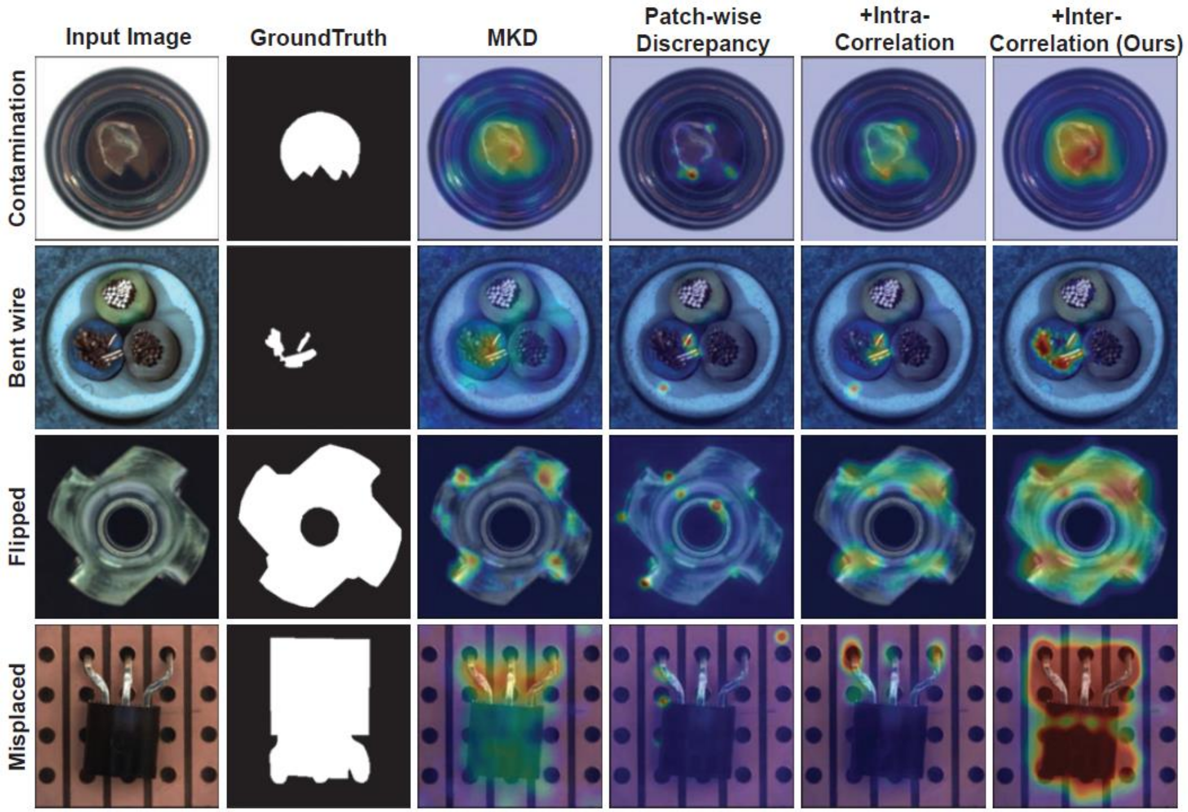
Weaker patch-to-normal-patch correlations

So, Focus the Discrepancy!

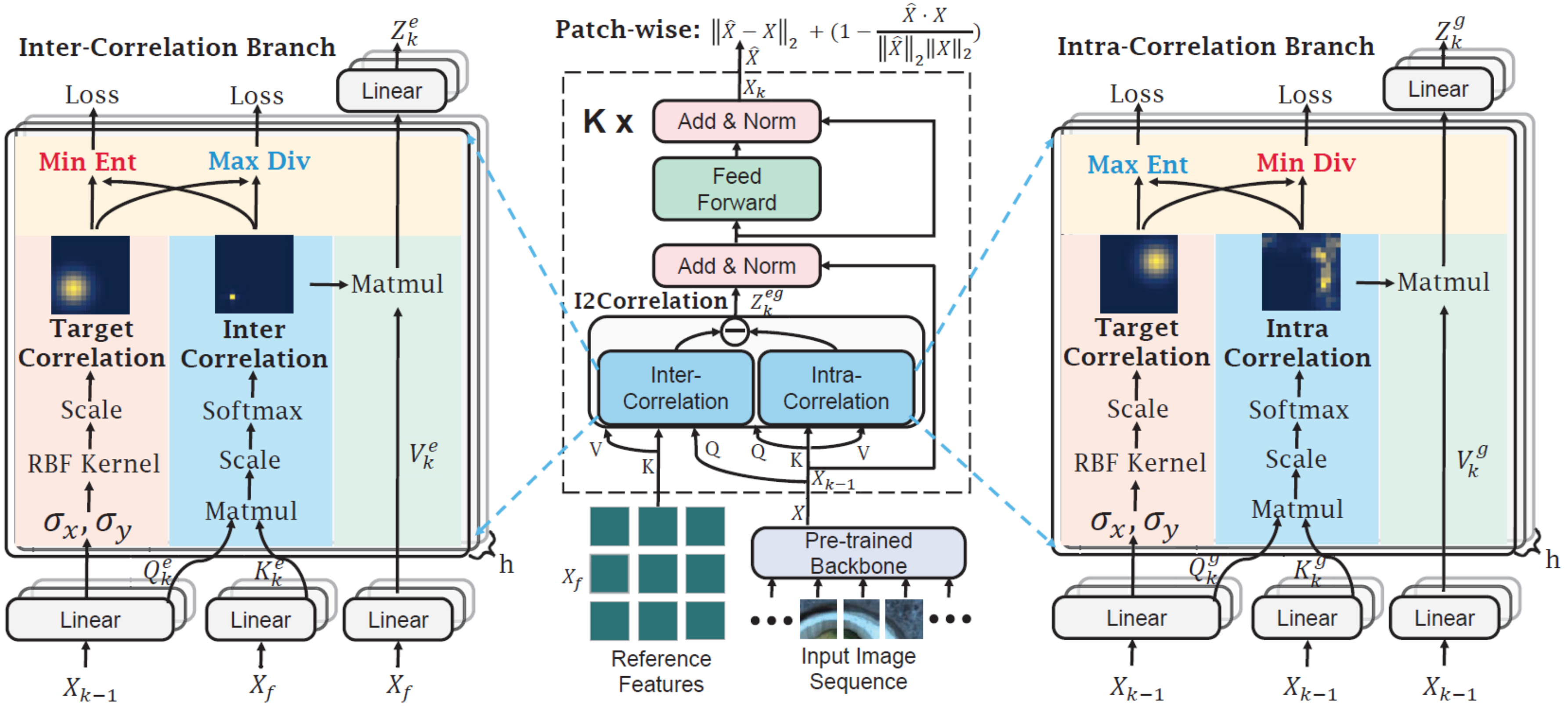
1. Patch-wise representations are different from the normal visuals

2. Different from most patches within one image

2. Deviate from our accumulated knowledge of normality



FOD: Focus the Discrepancy



Three parts: Patch-wise Discrepancy Branch, Intra-Correlation Branch, Inter-Correlation Branch.

Outline

ICCV23 PARIS

SHANGHAI JIAO TONG UNIVERSITY

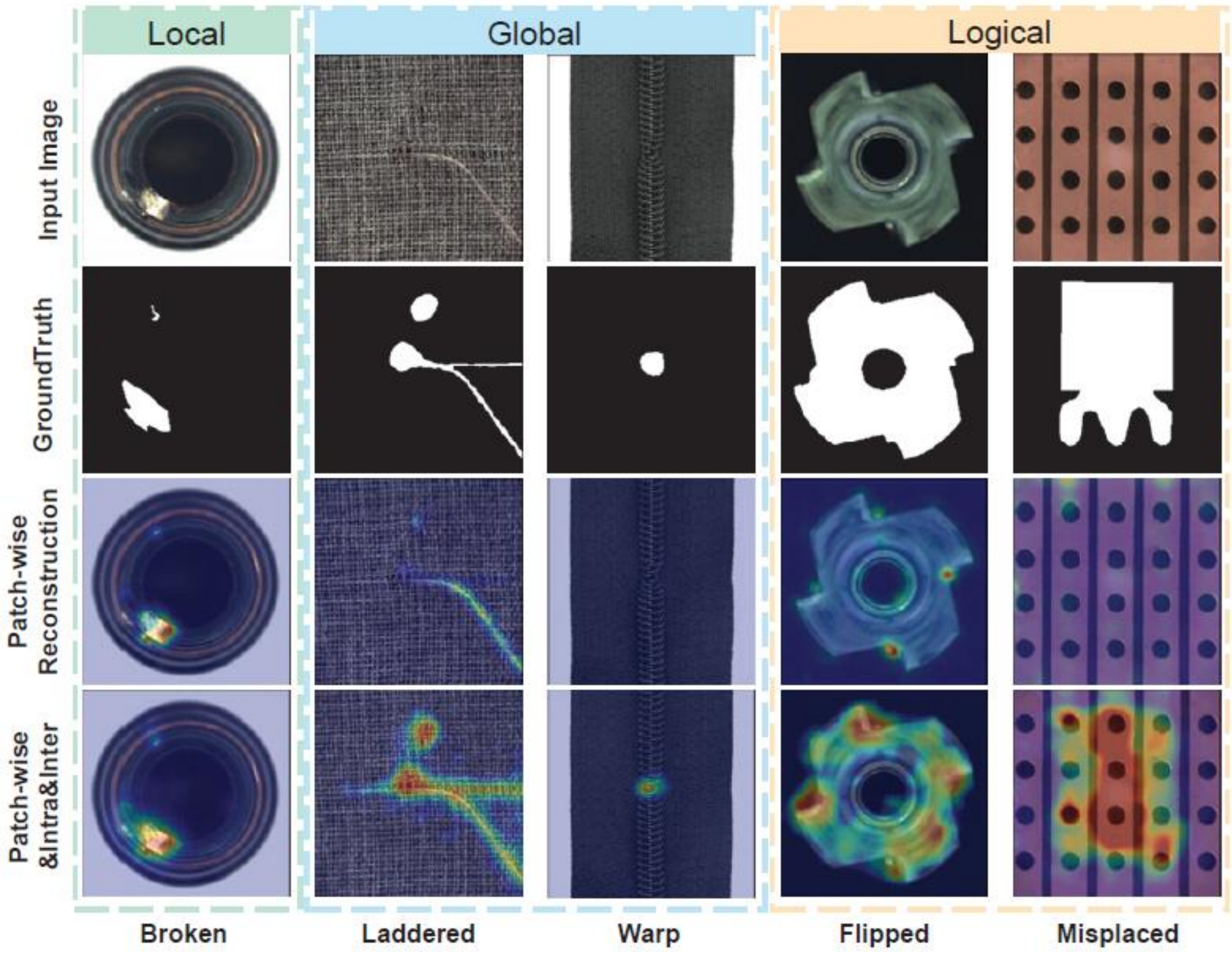
1. Motivation
2. Our Approach: Focus the Discrepancy
3. Experiments
4. Ablations
5. Conclusions

| Motivation

Our core insight: AD by sufficiently focusing the discrepancy!

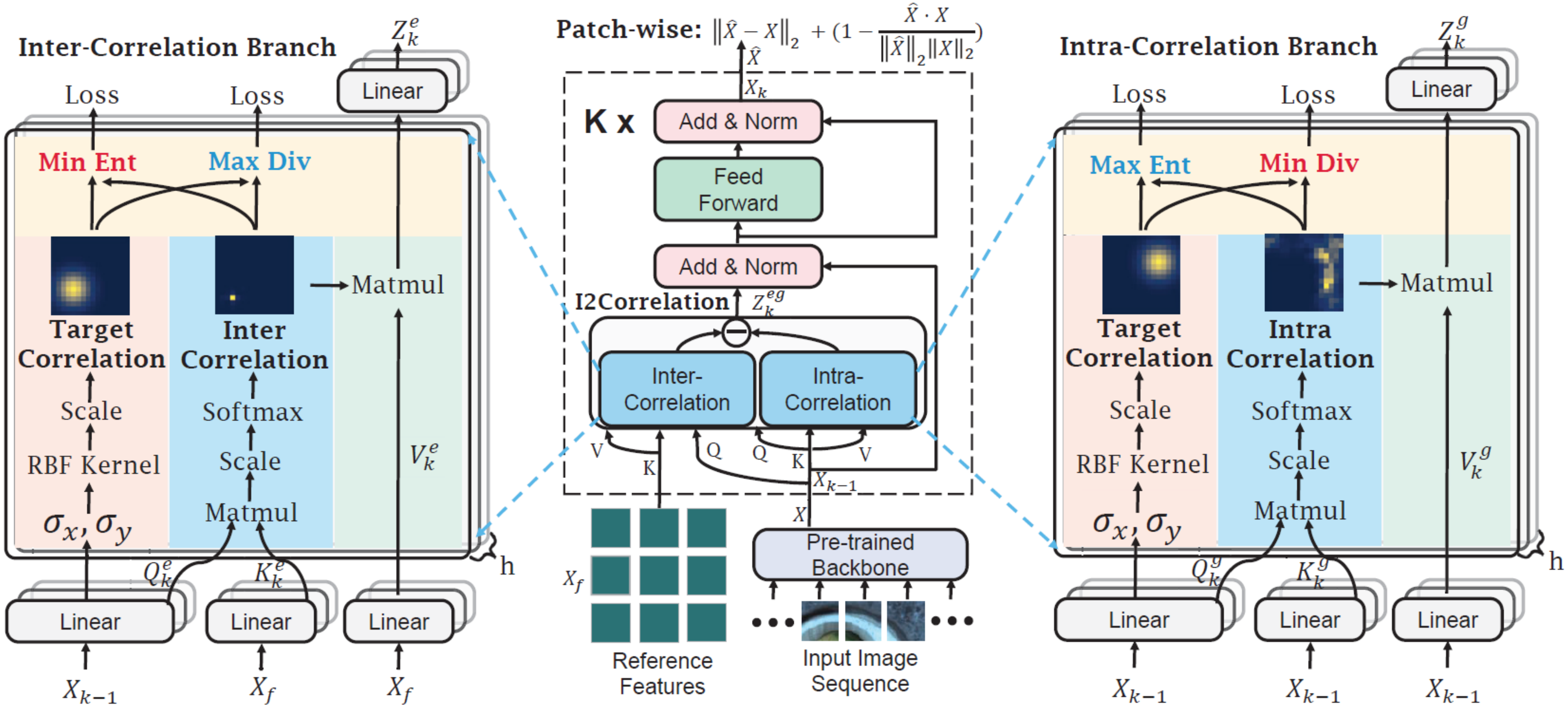
Analogy to humans:

- Patch patterns that differentiate from the normal visuals (**local**).
- Image regions that destroy textures or structures (**within image**).
- Novel appearances that deviate from accumulated knowledge of normality (**cross image**).



| Our Approach: FOD

- **Focus the Discrepancy, Model Overview:**



Three parts: Patch-wise Discrepancy Branch, Intra-Correlation Branch, Inter-Correlation Branch.

Our Approach: FOD

- **Patch-Wise Reconstruction Discrepancy**

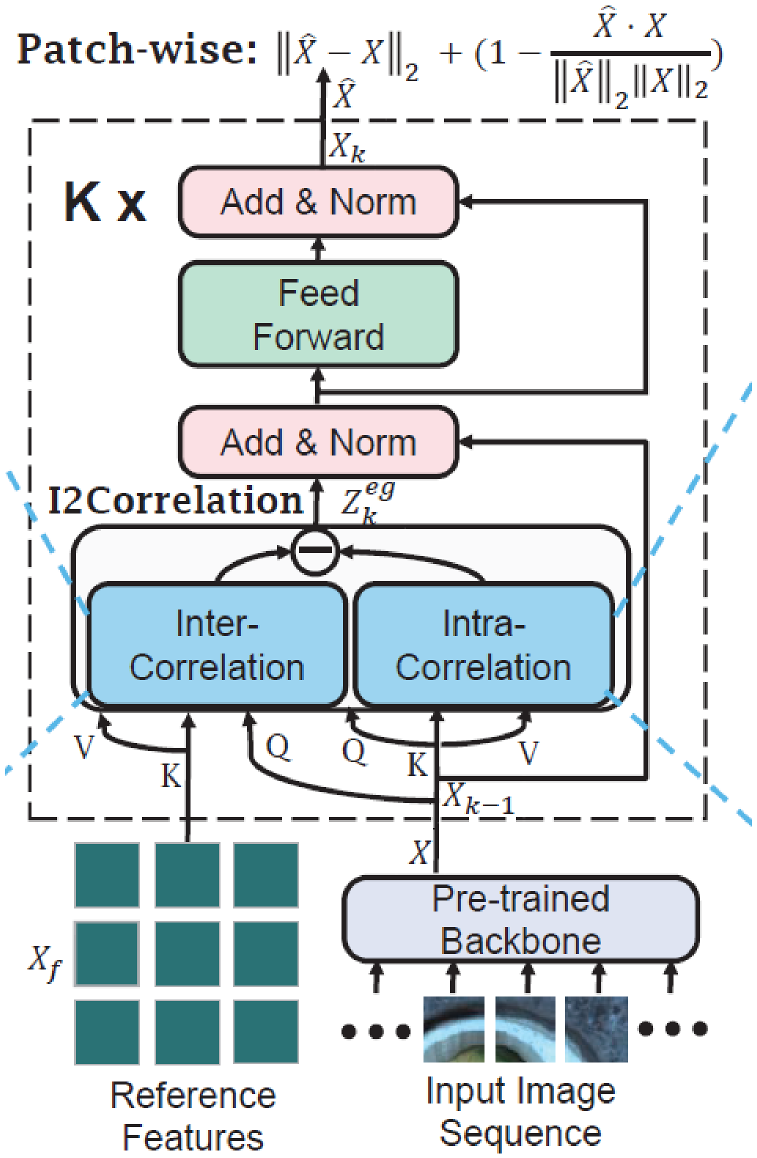
- This corresponds to the first recognition view:
  patch-wise discrepancy.

- For simplicity, we employ **feature reconstruction**.

- **Learning objective:**

$$\mathcal{L}_l = ||\hat{X} - X||_2 + \left(1 - \frac{\hat{X} \cdot X}{||\hat{X}||_2 ||X||_2}\right)$$

L2 distance + cosine distance



Our Approach: FOD

ICCV23 PARIS · SHANGHAI JIAO TONG UNIVERSITY

- **Intra-Correlation Learning Branch**

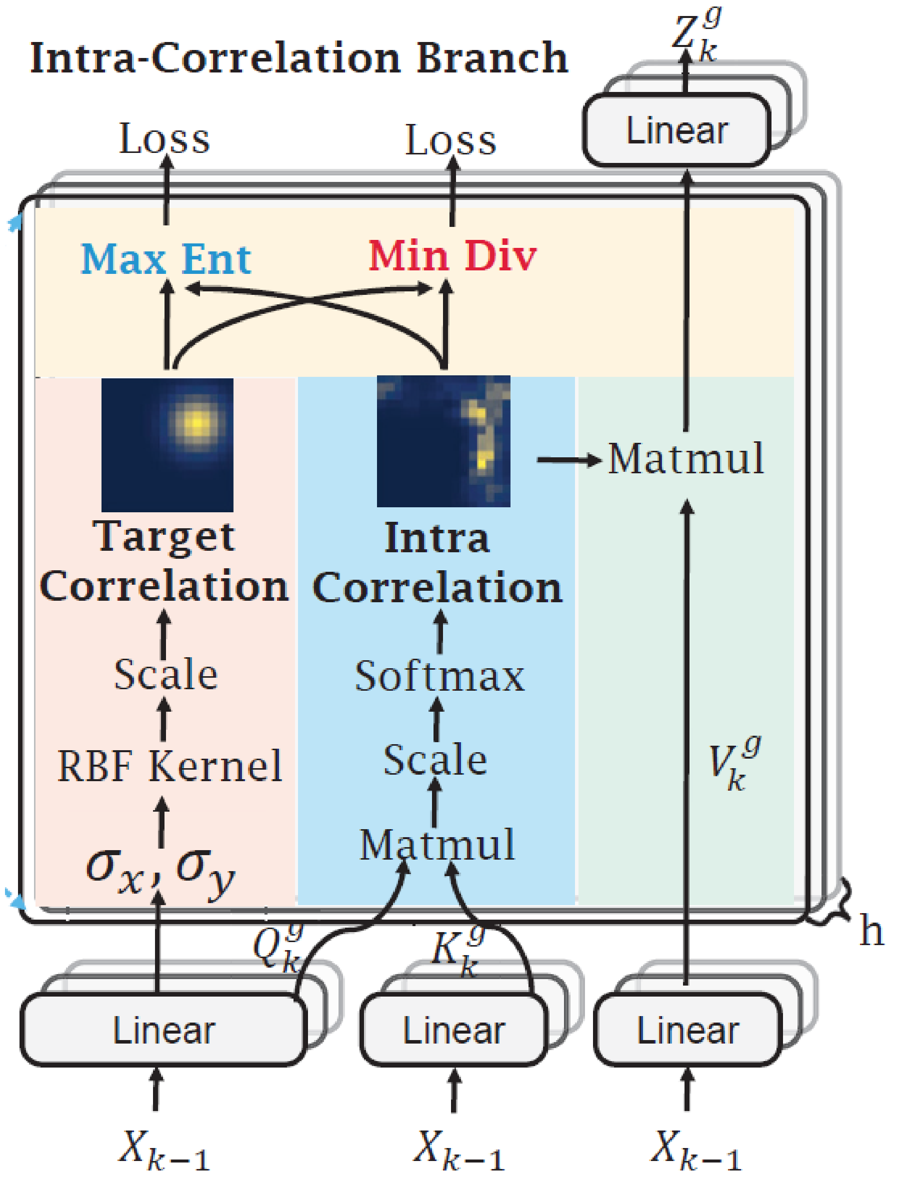
- This corresponds to the second recognition view: **intra-image discrepancy**.
- We **explicitly** take advantage of the self-attention maps as intra-correlation matrices.

- **So, how to learn?**
- We introduce RBF kernel based **Target Correlation**:

$$T_k^g = \frac{1}{2\pi\sigma_x\sigma_y}\exp\Big(-\frac{||x_{ij} - x_{i'j'}||_2^2}{2(\sigma_x^2+\sigma_y^2)}\Big)$$
$$i, i' \in \{1, \dots, H\}; j, j' \in \{1, \dots, W\}$$

- Then, measure distance by KL:

$$\mathrm{Div}(\mathcal{T}^g, \mathcal{S}^g) = \frac{1}{K}\sum_{k=1}^{K}\Big(KL(T_k^g||S_k^g) + KL(S_k^g||T_k^g)\Big)$$



Our Approach: FOD

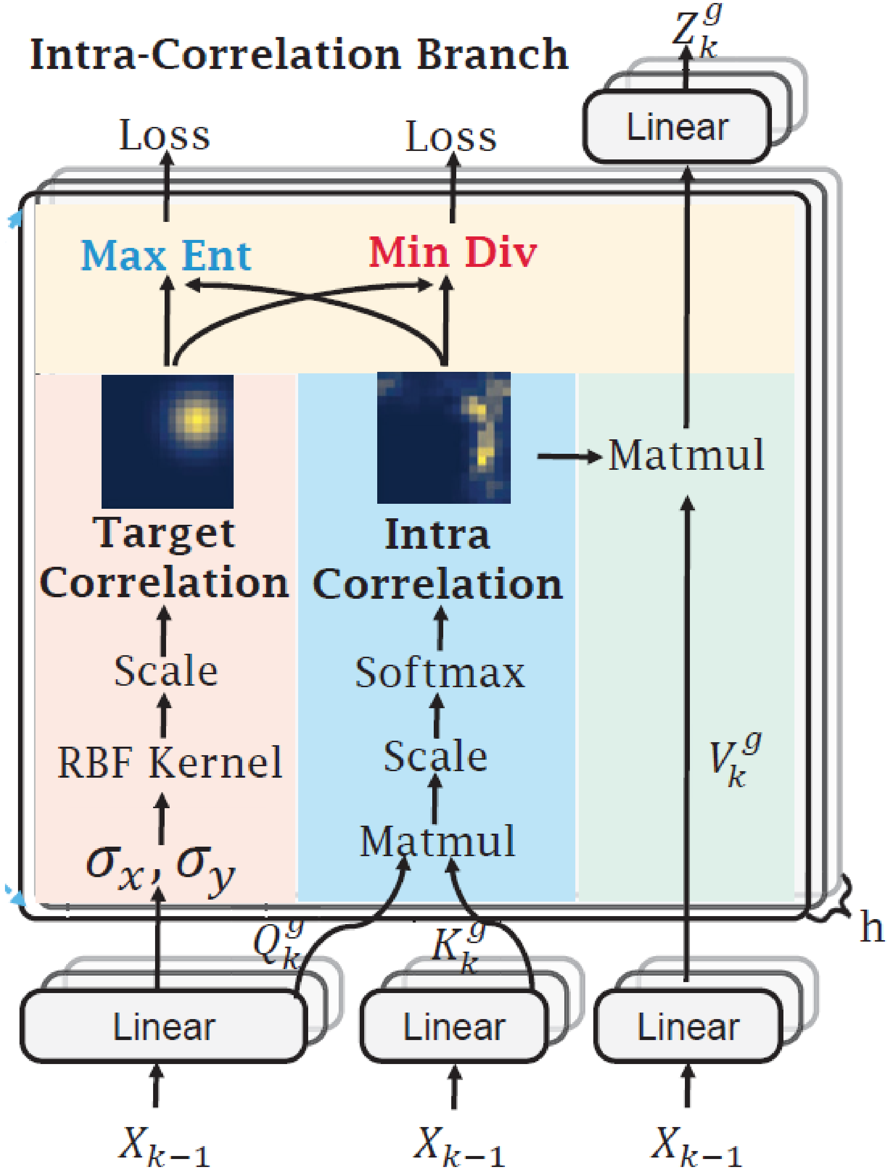
- **Entropy Constraint**

- **Why?**
- Learned correlation distributions of normal patches may also easily concentrate on the adjacent patches.

- So, we further introduce an entropy constraint item for making normal patches establish **strong associations** with **most normal patches**.

- Then, we maximize the entropy and minimize the KL:

$$\mathcal{L}_g = \lambda_1 \mathrm{Div}(\mathcal{T}^g, \mathcal{S}^g) - \lambda_2 \mathrm{Ent}(\mathcal{S}^g)$$



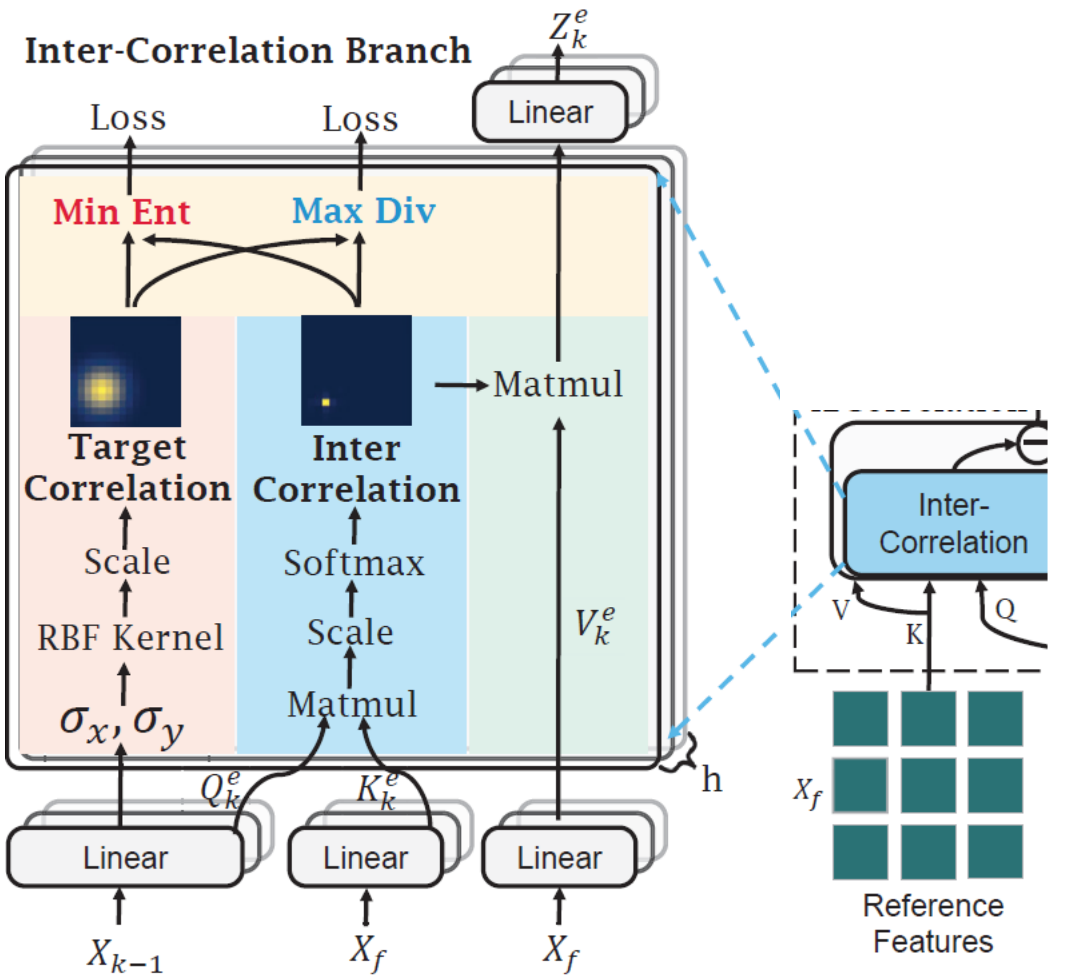
Our Approach: FOD

- **Inter-Correlation Learning Branch**

- This corresponds to the third recognition view:
  inter-image discrepancy.
- We can effectively take advantage of the known normal patterns from the normal training set by **External Reference Features**.
- Patch-wise averaged features are good empirically.

- **So, how to learn?**
- Loss function has opposite optimization direction to $\mathcal{L}_g$:

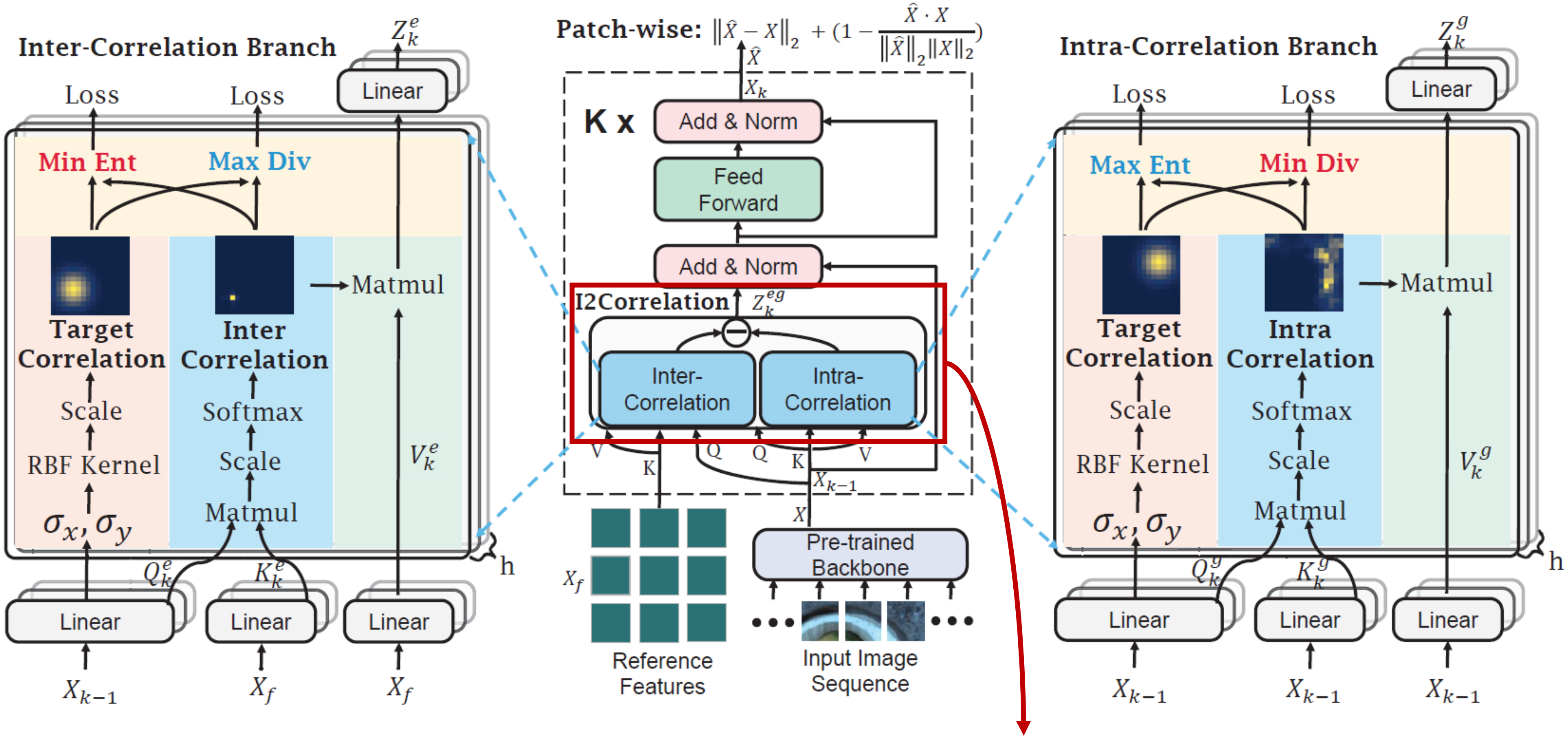
$$\mathcal{L}_e = -\lambda_1 \mathrm{Div}(\mathcal{T}^e, \mathcal{S}^e) + \lambda_2 \mathrm{Ent}(\mathcal{S}^e)$$



Our Approach: FOD

ICCV23 PARIS · SHANGHAI JIAO TONG UNIVERSITY

- **I2Correlation**



I2Correlation: the residual features are more conducive to **spotlight** the abnormal patterns.

Experiments

- **Datasets:**
  - MVTecAD: 5534 high-resolution images, 15 categories, 73 anomaly types, and 1900 abnormal regions.
  - BTAD: This dataset contains 2830 real-world images of 3 industrial products.
  - MVTec3D-RGB: This dataset contains 4147 RGB images from 10 real-world categories.
- **Metrics:**
  - Area under the curve of the receiver operating characteristic (AUROC), image-level and pixel-level.

Experiments

ICCV23 PARIS | SHANGHAI JIAO TONG UNIVERSITY

- **Comparison with SOTAs:**

Discrepancy Type	Method	Venue	Image-level AUROC	Pixel-level AUROC
Patch-wise Representation Discrepancy	STAD [4]	CVPR 2020	0.877	0.939
	PaDiM [11]	ICPR 2020	0.955	0.975
	DFR [57]	Neurocomputing 2021	/	0.950
	FCDD [25]	ICLR 2021	/	0.920
	MKD [41]	CVPR 2021	0.877	0.907
	Hou *et al.* [17]	ICCV 2021	0.895	/
	Metaformer [54]	ICCV 2021	0.958	/
	DRAEM [64]	ICCV 2021	0.980	0.973
	RDAD [12]	CVPR 2022	0.985	0.978
	SSPCAB [34]	CVPR 2022	0.989	0.972
	DSR [65]	ECCV 2022	0.982	/
	NSA [43]	ECCV 2022	0.972	0.963
	UniAD [62]	NIPS 2022	0.966	0.966
	UTRAD [8]	Neural Networks 2022	0.960	0.967
Patch-to-patch Feature Distance	PatchSVDD [60]	ACCV 2020	0.921	0.957
	DifferNet [36]	WACV 2020	0.949	/
	CFLOW [15]	WACV 2022	0.983	0.986
	CS-FLOW [37]	WACV 2022	0.987	/
	Tsai *et al.* [15]	WACV 2022	0.981	0.981
	PatchCore [35]	CVPR 2022	0.991	0.980
Others	CutPaste [21]	CVPR 2021	0.952	0.960
	Wang *et al.* [51]	CVPR 2021	/	0.91
	SPD [68]	ECCV 2022	0.946	0.946
Patch-wise&Intra&Inter	FOD (Ours)	-	0.992	0.983

- **Detailed Results (same backbone):**

Category	Image-level Anomaly Detection					
	DRAEM [64]	PatchSVDD [60]	MKD [41]	PatchCore [35]	CFLOW [15]	FOD (Ours)
Carpet	0.978	0.963	**1.000**	**1.000**	0.987	**1.000**
Grid	**1.000**	0.892	0.975	0.992	0.996	**1.000**
Leather	**1.000**	0.953	0.956	**1.000**	**1.000**	**1.000**
Tile	0.998	0.969	0.999	**1.000**	0.999	**1.000**
Wood	**0.991**	0.989	0.989	0.985	**0.991**	**0.991**
Bottle	0.993	0.976	0.989	**1.000**	**1.000**	**1.000**
Cable	0.929	0.899	0.972	0.992	0.976	**0.995**
Capsule	0.984	0.763	0.979	0.984	0.977	**1.000**
Hazelnut	**1.000**	0.912	0.997	**1.000**	**1.000**	**1.000**
Metal nut	0.989	0.941	0.972	**1.000**	0.993	**1.000**
Pill	0.981	0.791	0.971	0.954	0.968	**0.984**
Screw	0.939	0.825	0.870	0.953	0.919	**0.967**
Toothbrush	**1.000**	0.992	0.886	0.906	0.997	0.944
Transistor	0.914	0.874	0.956	0.995	0.952	**1.000**
Zipper	**1.000**	0.982	0.981	0.989	0.985	0.997
Mean	0.980	0.915	0.966	0.983	0.983	**0.992**

- In addition to pill, screw and toothbrush, our method achieves more than 99% AUROC in all other classes, others only achieve this on 9 classes.
- Our method performs much better on global and logical anomalies.

Experiments

- **More Results:**

Method	DRAEM [64]	PatchSVDD [60]	MKD [41]	PatchCore [35]	CFLOW [15]	FOD (Ours)
BTAD Dataset						
Image-level AUROC	0.922	0.924	0.935	0.934	0.948	**0.960**
Pixel-level AUROC	0.942	0.964	0.965	0.976	**0.978**	0.975
MVTec3D-RGB Dataset						
Image-level AUROC	0.757	0.743	0.688	0.839	0.851	**0.884**
Pixel-level AUROC	0.974	0.852	0.970	**0.977**	0.974	0.976

- We can outperform the best competitors on both BTAD and MVTec3D-RGB datasets.

Ablations

ICCV23 PARIS

SHANGHAI JIAO TONG UNIVERSITY

- **Ablation study results:**

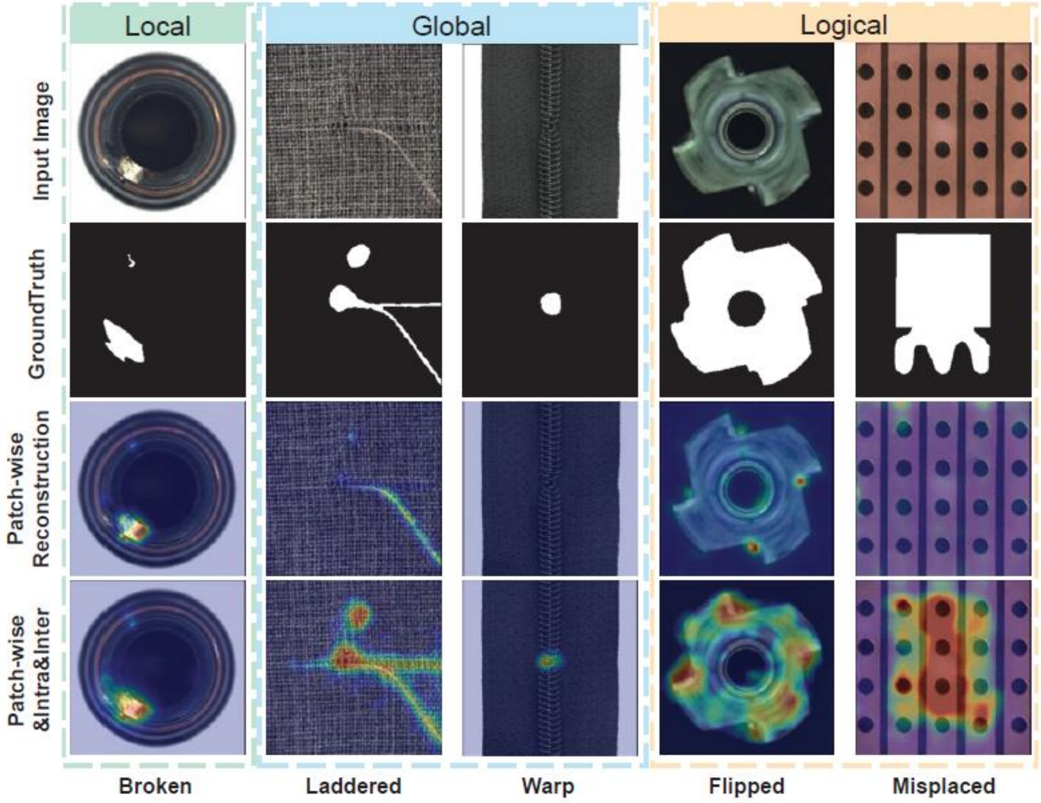
Recognition Views	Entropy Constraint	Reference Features	Anomaly Scoring	MVTecAD	BTAD	MVTec3D-RGB	Avg
Patch-wise	/	/	Rec	0.972	0.954	0.790	0.905
Intra	w/o	/	Div	0.700	0.811	0.708	0.740
	w/	/	Div	0.911	0.822	0.717	0.817
	w/	/	Rec&Div	0.974	0.952	0.818	0.915
Inter	w/	Mean	Rec&Div	0.980	0.958	0.832	0.923
	w/	Coreset	Rec&Div	0.925	0.884	0.700	0.836
Intra+Inter	w/	Mean	Div	0.896	0.922	0.814	0.877
Patch-wise+Intra +Inter (Ours)	w/	Mean	Rec&Div	**0.992**	**0.960**	**0.884**	**0.945**

- 1. The entropy constraint is quite effective and necessary in the intra- and inter-correlation branches.
- 2. The reconstruction errors and the intra- and inter-correlations can collaborate to improve detection performance.
- 3. Our FOD can surpass the pure reconstruction Transformer by 4.0% absolute improvement.

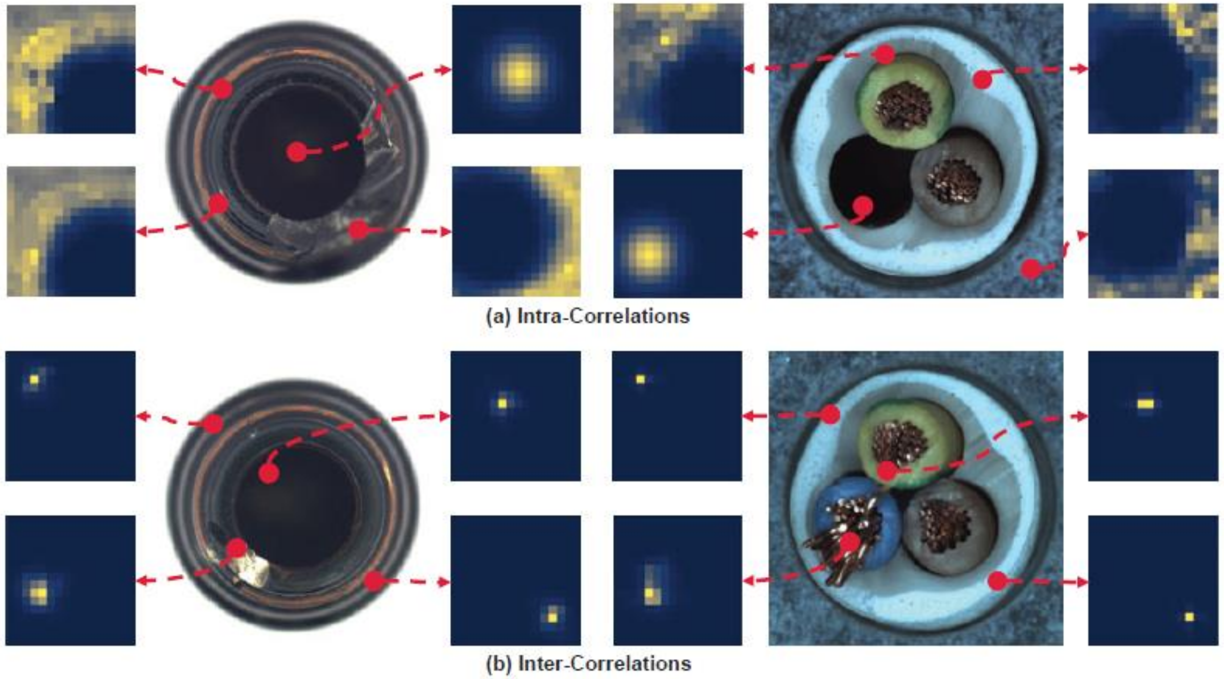
Qualitative Results

ICCV23 PARIS

SHANGHAI JIAO TONG UNIVERSITY



Detection effect of different anomalies



Visualization of intra- and inter-correlations

Conclusions

ICCV23 PARIS · SHANGHAI JIAO TONG UNIVERSITY

Anomalies are recognized through sufficient comparisons with normals.

Focus the Discrepancy!

ICCV23
PARIS

SHANGHAI JIAO TONG UNIVERSITY

Thanks!

Contact Us:
sunny_zhang@sjtu.edu.cn