Adapting Generic Features to A Specific Task: A Large Discrepancy Knowledge Distillation for Image Anomaly Detection

Chenkai Zhang, Tianqi Du, Yueming Wang*

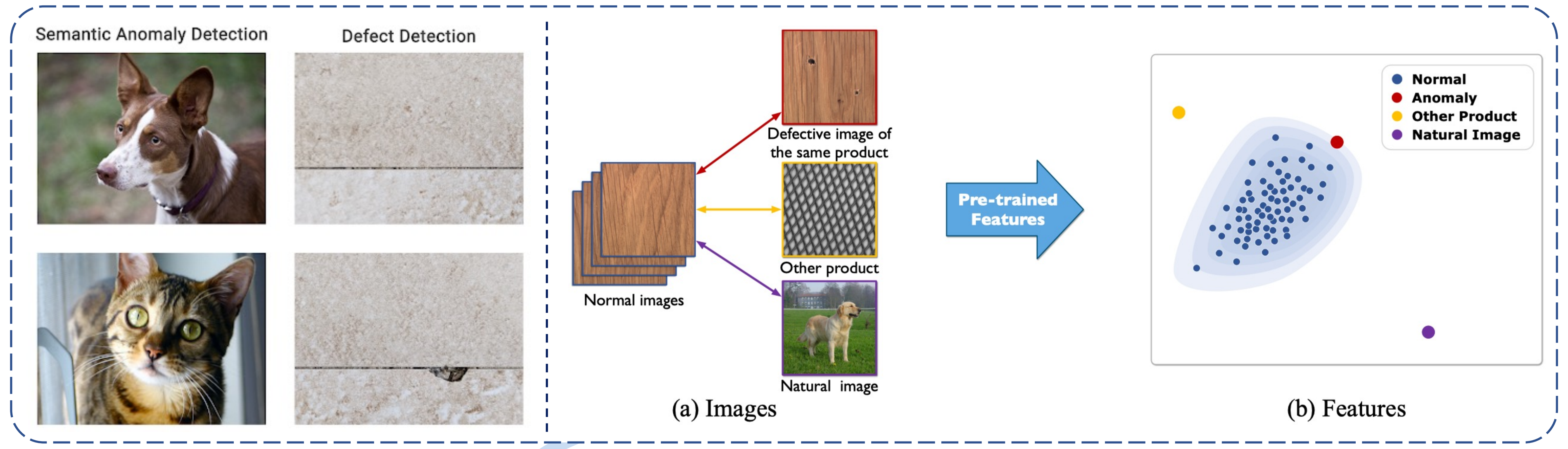
Zhejiang University

ZHEJIANG UNIVERSITY

INTRODUCTION

Motivation: The features pre-trained on natural images only have limited ability to detect anomalies in industrial images.

- **Industrial image are different:** anomalous and normal images share the same semantics and defects occur only in small regions.
- **Generic features are less capable:** the features pre-trained on natural images cannot produce discriminative features for industrial anomalies.



How to exploit generic features to a specific detection task ?

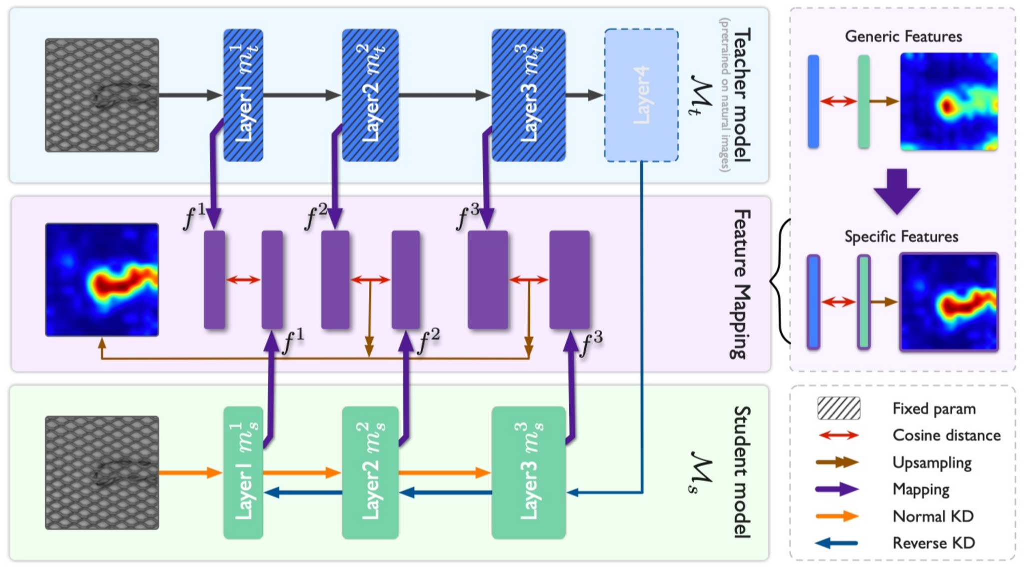
METHOD

Our idea

✓ General features are adopted to the detection task by Feature Mapping (FM)
✓ The separability of anomalies will be increased by optimizing Angular Margin Loss (AML)

Feature Mapping

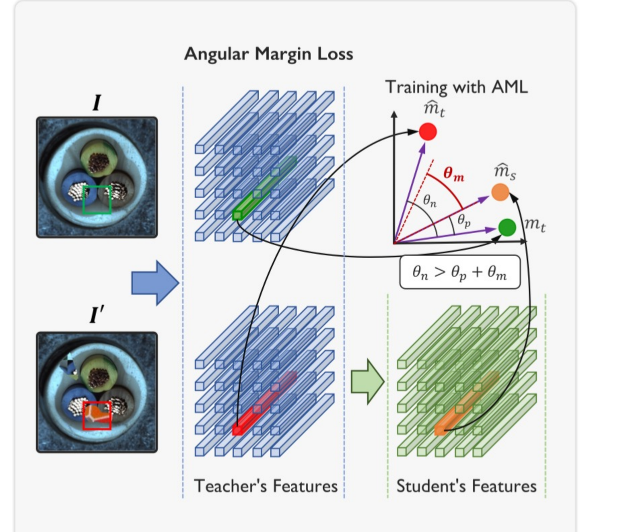
- Keeping the original features untouched
- Adjusting features by a tiny number of parameters
- Maximizing the feature discrepancies between the T-S models



Adapting the pre-trained models to the anomaly detection task.

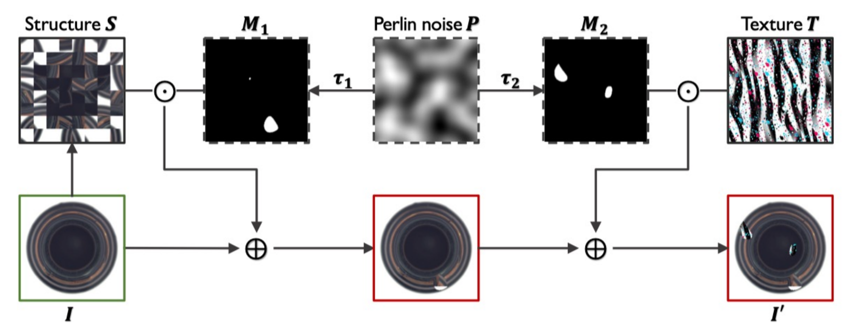
Angular Margin Loss

- Modeling feature discrepancy as angle based on the cosine similarity
- The feature discrepancies are required to meet the angular margin inequation
- Leveraging simply synthesized anomaly for contrastive learning



Synthesis Process

- Structure obtained from shuffled patches
- Texture obtained from the DTD dataset
- Fusing the structure and texture by Perlin noises



EXPERIMENTS AND RESULTS

Results

Experiment Setup

- **Datasets:**
  - MVTec dataset
  - ZJU-Leaper dataset
- **Evaluation:**
  - Image-level AUC of ROC
  - Pixel-level AUC of ROC

Performance Comparison

Table 1: The performance comparison on MVTec dataset and ZJU-Leaper Dataset.

Category	Method	MVTec		ZJU-Leaper	
		Image AUC	Pixel AUC	Image AUC	Pixel AUC
Feature Space	PuzzleAE [19]	71.1	80.7	69.1	68.2
	FCDD [15]	86.6	92.5	58.0	61.6
	SPADE [3]	85.5	96.0	83.3	88.8
	PaDiM [4]	90.3	96.1	84.8	86.3
Symmetric KD	MRKD [18]	87.7	90.7	86.9	82.3
	NKD	94.7	96.6	84.9	92.7
	NKD+FMAM	**96.7**	**96.9**	**88.6**	**93.6**
Asymmetric KD	RKD [5]	96.1	97.1	89.8	93.8
	RKD+FMAM	**98.2**	**97.3**	**91.9**	**94.7**

- FMAM achieves superior performance than many SOTA methods
- FMAM achieves better performance than original KD methods
- Improvements can be obtained on different datasets

Analysis

Analysis of Feature Distances

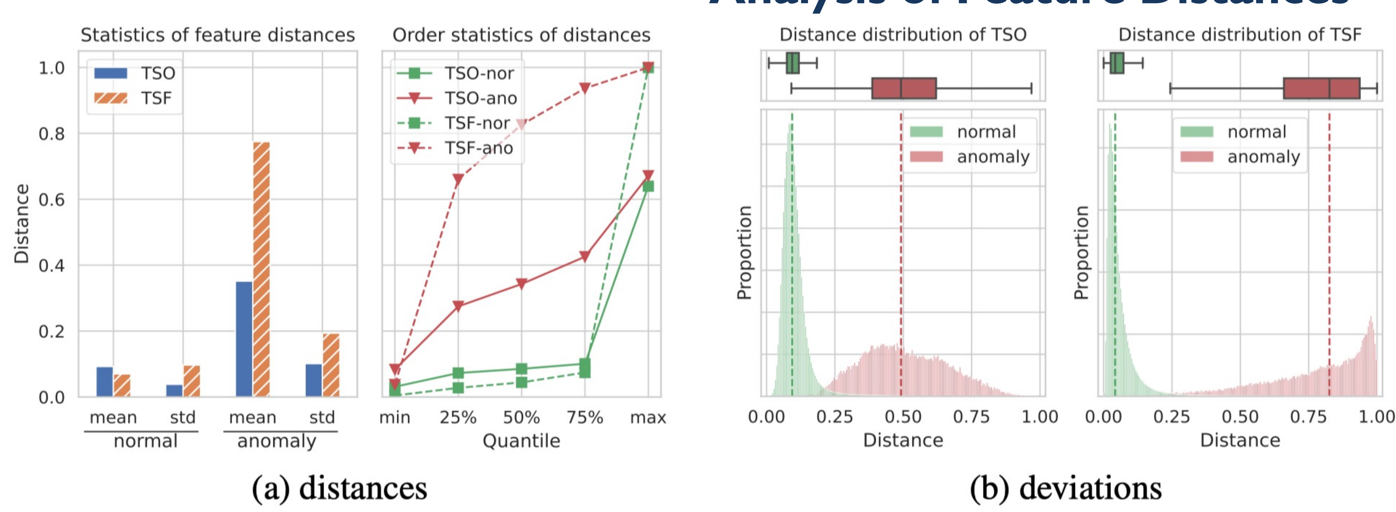


Figure 5: Statistics of feature distances of distance deviations

Statistical Analysis

- FMAM can effectively increase bias of anomalous features while keep low variance of normal features

Distribution Analysis

- FMAM can significantly increase the separability between normal and anomalous samples

Ablation Studies

Table 3: Ablation of FMAM

Model	Image AUC	Pixel AUC
RKD	96.1	97.1
RKD+FM	96.8	97.2
RKD+AML	96.9	97.2
RKD+FMAM	98.2	97.3
RKD+FM(L1ML)	93.9	96.6
RKD+FM(L2ML)	93.0	90.5

Ablation of FMAM

- Simply E2E training benefits little
- FMAM yields better performance
- Non-angular margin perform poorly

Ablation of backbones

- FMAM benefits different backbones
- Larger backbone achieves better performance

Table 4: Ablation of other backbones

Model	Image AUC	Pixel AUC
RKD(res34)	98.3	97.2
RKD(res34)+FMAM	98.3	97.4
RKD(res50)	98.5	97.6
RKD(res50)+FMAM	98.8	97.9
RKD(wres50)	98.5	97.7
RKD(wres50)+FMAM	99.1	98.1

Visualization

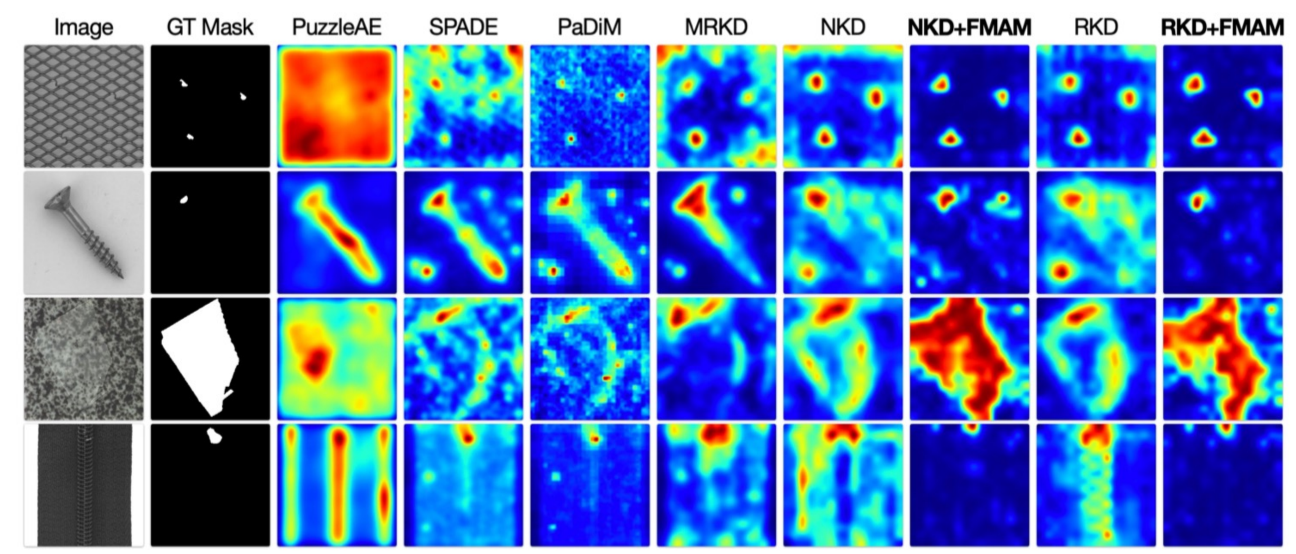


Figure 4: Visual comparison with other SOTA detection methods on the MVTec dataset.

Results of Localization

- FMAM can produce clearer and more precise localization results
- FMAM can reduce false alarms of on the heatmaps
- FMAM can better "fill" the whole anomalous regions

Conclusion

✓ Feature Mapping adapts features pre-trained on natural images for the image anomaly detection task.
✓ Training with Angular Margin Loss further increases feature discriminability.
✓ The superior performance successfully demonstrate the effectiveness of the proposed method.