

# RoBiS: Robust Binary Segmentation for High-Resolution Industrial Images

Xurui Li<sup>1</sup> Zhongsheng Jiang<sup>1</sup> Tingxuan Ai<sup>1</sup> Yu Zhou<sup>1,2,3,†</sup>  
{xrli\_plus, zsjiang, tingxuanai, yuzhou}@hust.edu.cn

<sup>1</sup>School of Electronic Information and Communications, Huazhong University of Science and Technology

<sup>2</sup>Hubei Key Laboratory of Smart Internet Technology, Huazhong University of Science and Technology

<sup>3</sup>Artificial Intelligence Research Institute, Wuhan JingCe Electronic Group Co., LTD

**Track I—Adapt & Detect: Robust Anomaly Detection in Real-World Applications**

## Abstract

*Robust unsupervised anomaly detection (AD) in real-world scenarios is an important task. Current methods exhibit severe performance degradation on the MVTec AD 2 benchmark due to its complex real-world challenges. To solve this problem, we propose a robust framework RoBiS, which consists of three core modules: (1) Swin-Cropping, a high-resolution image pre-processing strategy to preserve the information of small anomalies through overlapping window cropping. (2) The data augmentation of noise addition and lighting simulation is carried out on the training data to improve the robustness of AD model. We use INP-Former as our baseline, which could generate better results on the various sub-images. (3) The traditional statistical-based binarization strategy (mean+3std) is combined with our previous work, MEBin (published in CVPR2025), for joint adaptive binarization. Then, SAM is further employed to refine the segmentation results. Compared with some methods reported by the MVTec AD 2, our RoBiS achieves a **29.2% SegF1 improvement** (from 21.8% to 51.00%) on  $TEST_{priv}$  and **29.82% SegF1 gains** (from 16.7% to 46.52%) on  $TEST_{priv,mix}$ . Code is available at [github link](#).*

## 1. Introduction

Unsupervised anomaly detection addresses the critical challenge of identifying deviations from normal patterns using only defect-free training data, with applications spanning medical diagnostics [9], industrial inspection [4, 24], and video surveillance [18]. While current methods achieve near-perfect performance (AUROC > 99%) on benchmark datasets like MVTec AD [3], their real-world deployment faces fundamental limitations. Real-world scenarios introduce complex and diverse interferences, such as various lighting conditions, viewpoint changes, and sensor noise.

These challenges expose the fragility of existing approaches that lack explicit mechanisms for handling real-world condition variations, particularly in preserving detection accuracy under significant domain shifts.

The MVTec AD 2 [12] dataset introduces eight industrial inspection scenarios with 8,004 high-resolution images, establishing a rigorous benchmark for unsupervised anomaly detection. While training data maintains regular lighting conditions, the test suite comprises three distinct challenges: (1) Public testset ( $TEST_{pub}$ ) provides limited annotated samples under mixed illumination for preliminary validation, (2) Private testset ( $TEST_{priv}$ ) preserves training-domain conditions for conventional evaluation, and (3) Mixed private testset ( $TEST_{priv,mix}$ ) combines seen and unseen lighting conditions to simulate real-world uncertainties. Current methods achieving more than 99% AUROC on controlled benchmarks suffer large performance degradation under these complex conditions, exposing the urgent need for robust anomaly detection frameworks.

**Challenge 1: Robustness on varying scenarios.** Eight industrial products of the MVTec AD 2 dataset emphasize distinct technical challenges. Spatial randomness challenges emerge in categories like wall plugs and walnuts, where overlapping objects create complex occlusion patterns. Material diversity challenges occur in fabric, sheet metal and rice categories, which require texture consistency modeling. Optical complexities involve reflective surfaces and translucent materials, such as cans, vials and fruit jellies, that distort visual signatures. Illumination challenges include extreme lighting conditions—backlighting present in vials/fruit jelly, as well as dark-field lighting in metal sheets. Additionally, grayscale images, high-resolution inputs and non-uniform aspect ratios collectively simulate real-world industrial inspection conditions.

**Challenge 2: Adaptive binarization.** Current anomaly detection methods could only generate continuous anomaly maps, where the pixel value indicates the anomaly likelihood. To conduct an accurate binary discrimination for

† Corresponding Author.

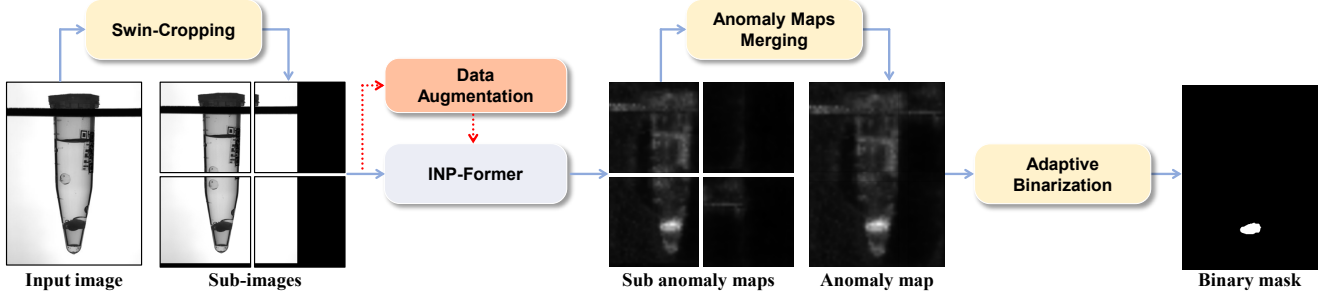


Figure 1. **The pipeline of our RoBiS.** This framework contains three important parts: (1) Swin-Cropping for dataset pre-processing (Sec. 2.1). (2) INP-Former [17] to detect anomalies (Sec. 2.2). (3) Anomaly Maps Merging to obtain the anomaly map of original image, and Adaptive Binarization to generate its corresponding binary mask (Sec. 2.3).

each pixel, it is necessary to study the adaptive binarization method. Some researches [23, 25] in the natural scenarios struggle to be used in the industrial scenarios. Otsu [19] often brings much over-detections due to the normal regions of products also have higher scores than the background. Using the average of all per-pixel anomaly scores plus three times their standard deviation as the threshold is effective, but it easily overlooks subtle anomalies with low scores.

The above challenges test the robustness and adaptability of unsupervised anomaly detection models in real-world detection scenarios. Therefore, we design the **Robust Binary Segmentation (RoBiS)** to handle these challenges and output precise binarization anomaly segmentation results.

## 2. Methodology

The pipeline of our method is illustrated in Fig. 1, which consists of three important parts: dataset pre-processing, model design and results post-processing.

### 2.1. Dataset Pre-processing

Through the statistical analysis of the MVTec AD 2 [12] dataset, we observe that all product images maintain high resolutions (ranging from 1400×1900 to 2448×2048). However, over 20% of anomalies occupy less than 281 pixels, with the minimal anomaly containing only 5 pixels. To address this challenge of small anomaly localization, we propose the sliding window-based image cropping strategy, named Swin-Cropping. This pre-processing strategy systematically crops high-resolution inputs into smaller sub-images for subsequent model training and inference.

The Swin-Cropping strategy employs a 1024×1024 window size to balance computational efficiency and detection accuracy, as excessive partitioning would lead to exponential growth in training time. To prevent boundary artifacts [15] where one anomaly might be divided by adjacent windows, we implement a 10% overlap between neighboring sub-images. This design ensures the integrity of large or elongated anomaly patterns during the segmentation process. To process peripheral regions, zero-padding is applied

to residual boundary areas, maintaining consistent input dimensions for subsequent network operations while preserving spatial relationships in marginal zones.

## 2.2. Model Design

### 2.2.1. Approach

While texture categories (e.g., fabric, rice, and sheet metal) exhibit consistent sub-image patterns, significant inter-subimage variations exist in other product categories. Traditional one-class anomaly detection (AD) methods [5, 8, 26, 28] focus on training individual model for each category, while each category in the MVTec AD 2 dataset contains multiple distinct sub-images. This architectural constraint prevents effective learning across all sub-image variations. In contrast, multi-class approaches [2, 10, 11, 17] demonstrate superior learning capacity, particularly in categories like vial and fruit jelly, which have more sub-image discrepancies. Based on the above analysis, we adopt INP-Former [17] as our baseline, which is the current state-of-the-art multi-class AD framework. Notice that our method strictly maintains the unsupervised setting and trains one model for each product category.

### 2.2.2. Architecture

Our implementation preserves the original INP-Former architecture while employing a ViT-B-14 backbone initialized with DINOv2-R [6] pre-trained weights as the encoder. To maintain the details in high-resolution, all sub-images from both training and test sets are resized to 518×518 resolution. In addition, since some anomalies may appear around image boundaries, we discard the center cropping operation to generate more reliable segmentation results.

### 2.2.3. Training

While training images are captured under regular lighting conditions, test scenarios contain significant illumination variations including underexposure, overexposure, and additional light sources. To enhance model robustness against these illumination variations, we implement an augmentation pipeline applied to 50% of training samples. For each

Object	PatchCore [20]	RD [7]	RD++ [21]	EfficientAD [1]	MSFlow [29]	SimpleNet [16]	DSR [27]	RoBiS (Ours)
Can	0.3 / 0.1	0.1 / 0.1	0.1 / 0.1	0.8 / 0.1	<b>5.0</b> / 0.1	0.6 / 0.1	0.4 / 0.1	<u>1.86</u> / <b>0.84</b>
Fabric	11.5 / 9.8	2.6 / 2.2	2.9 / 2.3	7.6 / 1.0	<u>22.0</u> / 4.1	21.6 / <u>10.2</u>	7.9 / 5.0	<b>87.46</b> / <b>73.37</b>
Fruit Jelly	8.7 / 8.2	22.5 / 22.7	26.9 / 26.7	20.8 / 18.2	<u>47.6</u> / <u>38.1</u>	25.1 / 23.0	17.9 / 17.2	<b>53.63</b> / <b>52.62</b>
Rice	3.8 / <u>4.2</u>	7.0 / 3.9	9.5 / 2.9	15.0 / 0.5	<u>19.1</u> / 1.8	11.6 / 1.0	1.5 / 1.4	<b>63.86</b> / <b>63.23</b>
Sheet Metal	1.8 / 1.1	<u>41.3</u> / <u>39.2</u>	40.9 / 37.7	9.3 / 3.8	13.0 / 7.6	14.6 / 2.8	13.9 / 14.4	<b>70.98</b> / <b>70.92</b>
Vial	2.3 / 2.2	<u>28.0</u> / <u>28.3</u>	28.2 / 22.8	30.5 / 26.5	23.3 / 6.2	<u>31.9</u> / 17.5	28.2 / 27.9	<b>48.73</b> / <b>48.83</b>
Wall Plugs	0.0 / 0.0	1.9 / 0.8	1.3 / <u>0.9</u>	4.4 / 0.3	0.1 / 0.2	1.0 / 0.3	0.4 / 0.4	<b>14.38</b> / <b>3.40</b>
Walnuts	1.2 / 1.3	41.2 / 36.7	44.1 / <u>40.5</u>	34.6 / 13.3	<u>44.5</u> / 14.3	35.2 / 14.3	17.0 / 9.6	<b>67.13</b> / <b>58.94</b>
Mean	3.7 / 3.4	18.1 / <u>16.7</u>	19.2 / <u>16.7</u>	15.4 / 8.0	<u>21.8</u> / 9.0	17.7 / 8.7	10.9 / 9.5	<b>51.00</b> / <b>46.52</b>

Table 1. Segmentation  $F_1$  score (in %) on binarized images for  $TEST_{priv}/TEST_{priv,mix}$  set on the MVTec AD 2 dataset. The best-performing result is in bold, the second-best result is underlined.

sub-image  $I_i$ , we first inject Gaussian noise with zero mean and standard deviation  $\sigma = 15$ , producing  $\bar{I}_i$ . Subsequently, we simulate illumination variations through exposure adjustment to generate the augmented image  $\hat{I}_i$  as,

$$\hat{I}_i = \text{clamp}(\bar{I}_i * 2^\lambda) \quad (1)$$

where  $\text{clamp}(\cdot)$  constrains pixel values to  $[0, 255]$ . And  $\lambda \sim \mathcal{U}(-0.2, 0.2)$  controls exposure levels, where positive values induce overexposure while negative values create underexposure effects. This dual-stage augmentation generates photometrically perturbed samples  $\hat{I}_i$  that better approximate real-world testing conditions.

Our training configuration employs the StableAdamW [22] optimizer with an initial learning rate of  $1 \times 10^{-3}$  for 200 epochs, coupled with a WarmCosineScheduler for dynamic learning rate adjustment. The scheduler incorporates 100 warmup iterations to stabilize early training phases, gradually decaying the learning rate to a final value of  $1 \times 10^{-4}$  through cosine annealing.

### 2.3. Results Post-processing

During inference, our method generates anomaly maps for each sub-image, where each pixel value indicates anomaly likelihood. To obtain final binary segmentation masks for original high-resolution images, we implement a two-stage post-processing workflow.

① **Anomaly Maps Merging.** The spatial merging of anomaly maps begins with precise coordinate alignment using positional metadata recorded during Swin-Cropping preprocessing. Each sub-image’s anomaly map is repositioned in the original high-resolution coordinate system through bilinear interpolation. For overlapping regions between adjacent sub-images (10% overlap ratio), we compute pixel-wise probabilistic averages.

② **Adaptive Binarization.** For anomaly map  $A_i$ , the traditional binarization uses the average of all per-pixel anomaly scores plus three times their standard deviation as the threshold to produce the binary mask  $\bar{M}_i$ . However, this approach may lead to missing detection for subtle anomalies, which have lower anomaly scores. To mitigate this, we employ our MEBin proposed in AnomalyNCD [13]

(CVPR 2025), which adaptively determines optimal thresholds through stable connected-component analysis, generating the mask  $\hat{M}_i$ . Then we calculate the coarse segmentation mask through the logical OR operation as,

$$M_i = \bar{M}_i \cup \hat{M}_i \quad (2)$$

To enhance segmentation precision, we develop a SAM-based refinement module (SAM-Finer) to process the coarse segmentation mask  $M_i$ . First, we extract the minimum bounding box for each abnormal region of  $M_i$ . Then each bounding box is considered as a prompt of Segment Anything (SAM) [14] to guide finer segmentation. Since the SAM decoder generates three confidence-ranked masks for each prompt, we combined them through logical OR operations to minimize false negatives for the fabric and walnuts categories. We only use the mask with the largest confidence for the other categories. Final binary mask  $\bar{M}_i$  is obtained through the spatial composition of all refined abnormal regions. The above SAM-Finer is used on all categories except rice. In addition, we use SAM-Huge architecture for  $TEST_{priv}$  and SAM-Base for  $TEST_{priv,mix}$ .

### 2.4. Dataset & Evaluation

We conduct experiments on the industrial dataset MVTec AD 2 [12]. The complete dataset pre-processing workflow and technical details of our data augmentation strategies are documented in Sec. 2.1 and Sec. 2.2.3, respectively.

We strictly adhere to the official challenge’s evaluation paradigm, employing pixel-level F1-score (SegF1) as the primary metric for assessing anomaly detection performance. Conventional AD methods only produce continuous anomaly maps. To calculate the SegF1 metric, binarizing these anomaly maps is necessary.

## 3. Results

### 3.1. Quantitative Results

As reported in Table 1, our method achieves superior segmentation performance on MVTec AD 2’s challenging test sets, attaining SegF1 scores of 51.00% and 46.52% on

Object	AucPro_0.05 (private)	ClassF1 (private)	SegF1 (private)
Can	30.28	60.93	1.86
Fabric	79.45	83.79	87.46
Fruit Jelly	74.46	87.35	53.63
Rice	62.27	72.00	63.86
Sheet Metal	75.51	87.68	70.98
Vial	76.81	84.61	48.73
Wall Plugs	62.20	75.20	14.38
Walnuts	77.05	85.42	67.13
Mean	67.25	79.62	51.00

Object	AucPro_0.05 (priv_mixed)	ClassF1 (priv_mixed)	SegF1 (priv_mixed)
Can	20.03	65.04	0.84
Fabric	79.27	83.80	73.37
Fruit Jelly	74.11	87.55	52.62
Rice	63.89	73.45	63.23
Sheet Metal	73.54	86.69	70.92
Vial	69.59	85.77	48.83
Wall Plugs	24.77	72.66	3.40
Walnuts	72.00	83.95	58.94
Mean	59.65	79.86	46.52

Table 2. More quantitative results on  $TEST_{priv}/TEST_{priv,mix}$  set.

$TEST_{priv}$  and  $TEST_{priv,mix}$  respectively. This represents significant improvements of **29.2%** and **29.82%** over state-of-the-art competitors including PatchCore [20], RD [7], RD++ [21], EfficientAD [1], MSFlow [29], SimpleNet [16] and DSR [27], as officially benchmarked in MVTec AD 2. Our comprehensive evaluation further reveals consistent advantages across multiple metrics (Table 2), such as AucPro\_0.05 and ClassF1. Notably, while the challenge ranking prioritizes SegF1, our method’s balanced performance across all metrics (ClassF1: 79.62%/79.86%, AucPro\_0.05: 67.25%/59.65%) confirms its robustness and adaptability for different industrial scenes. These metrics are all calculated by the official platform<sup>1</sup> of the challenge.

### 3.2. Qualitative Results

Since the ground truth masks of  $TEST_{priv}$  and  $TEST_{priv,mix}$  sets are unavailable, we visualize the binary mask of  $TEST_{pub}$  set output by our method in Fig. 2. Our method demonstrates effective anomaly localization across most of the MVTec AD 2 product categories. While our method successfully identifies both structural and logical anomalies, architectural constraints inherent to anomaly detection models, particularly the need for successive downsampling to achieve sufficient receptive fields. This manifests in bi-

<sup>1</sup><https://benchmark.mvtec.com/vand-leaderboard>

Swin-Cropping	Data augmentation	F1-max
✓		33.6
	✓	31.8
✓	✓	<b>35.5</b>

Table 3. Ablation of our Swin-Cropping and data augmentation operation. We report the F1-score with optimal threshold (F1-max in %) to measure the performance of anomaly maps on  $TEST_{pub}$ .

MEBin	SAM-Finer	SegF1
✓		34.4
	✓	46.5
✓	✓	<b>46.9</b>

Table 4. Ablation of MEBin and SAM-Finer. We report the SegF1 (in %) to measure the performance of binary masks on  $TEST_{pub}$ .

nary masks that are larger than ground truth annotations. Although SAM integration provides partial refinement, the 64×64 feature resolution limit of SAM’s encoder prevents more precise segmentation of tiny anomalies.

## 4. Discussion

### 4.1. Ablation study

Table 3 presents ablation studies of our Swin-Cropping and data augmentation modules on the  $TEST_{pub}$  set. Due to the direct impact on anomaly map quality, we evaluate performance using the F1-max metric, which indicates the theoretical upper bound of SegF1 under the optimal threshold. Our analysis reveals that Swin-Cropping achieves 3.7% F1-max improvement by preserving critical anomaly information in sub-images, effectively mitigating missing detections of small anomalies. The data augmentation module contributes an additional 1.9% gain through exposure variation simulation and Gaussian noise injection, demonstrating enhanced robustness against various lighting conditions.

In Table 4, we conduct the ablation studies of the MEBin and SAM-Finer modules, which directly affect the binarization performance of anomaly maps. The MEBin module achieves a 0.4% SegF1 improvement by adaptively detecting subtle anomalies overlooked by conventional “mean+3std” thresholding, reducing false negatives. SAM-Finer delivers a substantial 12.5% performance gain, primarily through internal defect filling in fabric-class anomalies where many false negatives exist. While SAM enables fine-grained refinement of slender and small anomalies, its impact on quantitative metrics remains limited.

### 4.2. Future work

While our framework demonstrates state-of-the-art performance, several promising directions emerge for ad-



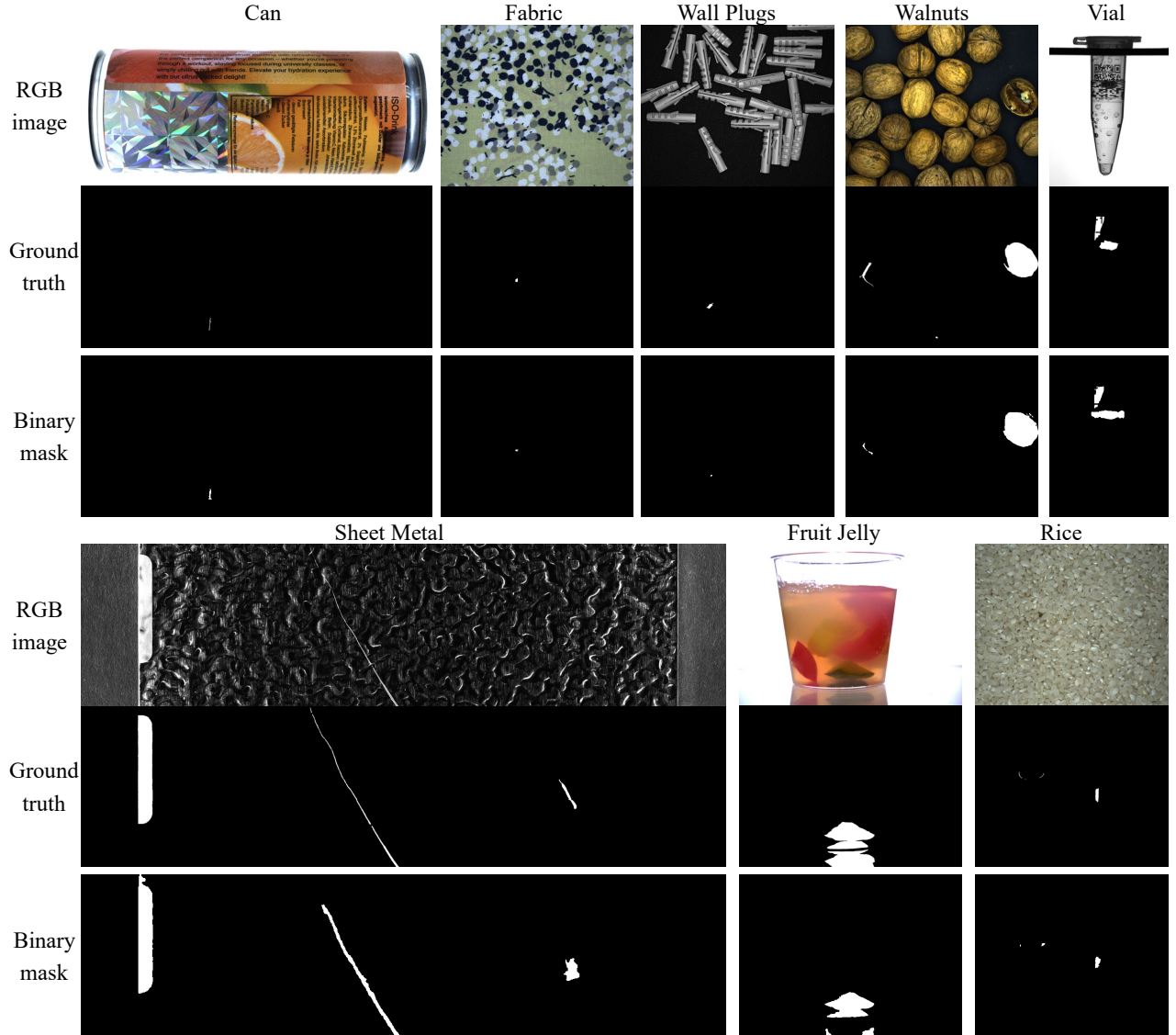


Figure 2. Visualization of anomaly segmentation results on the  $TEST_{pub}$  set of MVTec AD 2 dataset.

vancing industrial anomaly detection in real-world scenarios. Firstly, the detection of small anomalies in high-resolution industrial images remains excessive time consumption. Studying an efficient AD model for processing high-resolution images is necessary. Meanwhile, the adaptive binarization is a direction worth exploring. However, there are relatively few studies on this at present. It is promising to explore a binarization strategy that is robust to different product categories.

## 5. Conclusion

In this paper, we present our solution RoBiS for the CVPR VAND3.0 challenge in the MVTec AD 2 dataset. Three important modules are contained in our RoBiS: (1) Swin-Cropping to pre-processing high-resolution images. (2)

Data augmentation to simulate different lighting conditions and then training the AD model INP-Former. (3) Adaptive binarization and SAM-Finer to generate high-quality binary masks. Our solution achieves 51.00% and 46.52% SegF1 in the  $TEST_{priv}$  and  $TEST_{priv,mix}$ , respectively.

## 6. Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant No.62176098. The computation is completed in the HPC Platform of Huazhong University of Science and Technology.

## References

- [1] Kilian Batzner, Lars Heckler, and Rebecca König. Efficientad: Accurate visual anomaly detection at millisecond-level latencies. In *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2024. 3, 4
- [2] Farzad Beizaei, Gregory A Lodygensky, Christian Desrosiers, and Jose Dolz. Correcting deviations from normality: A reformulated diffusion model for multi-class unsupervised anomaly detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025. 2
- [3] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Mvtec ad—a comprehensive real-world dataset for unsupervised anomaly detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 1
- [4] Yunkang Cao, Xiaohao Xu, Jiangning Zhang, Yuqi Cheng, Xiaonan Huang, Guansong Pang, and Weiming Shen. A survey on visual anomaly detection: Challenge, approach, and prospect. In *arXiv preprint arXiv:2401.16402*, 2024. 1
- [5] Qiyu Chen, Huiyuan Luo, Chengkan Lv, and Zhengtao Zhang. A unified anomaly synthesis strategy with gradient ascent for industrial anomaly detection and localization. In *European Conference on Computer Vision (ECCV)*, pages 37–54. Springer, 2024. 2
- [6] Timothée Darcet, Maxime Oquab, Julien Mairal, and Piotr Bojanowski. Vision transformers need registers. In *International Conference on Learning Representations (ICLR)*, 2024. 2
- [7] Hanqiu Deng and Xingyu Li. Anomaly detection via reverse distillation from one-class embedding. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9737–9746, 2022. 3, 4
- [8] Matic Fučka, Vitjan Zavrtanik, and Danijel Skočaj. Transfusion—a transparency-based diffusion model for anomaly detection. In *European Conference on Computer Vision (ECCV)*, pages 91–108. Springer, 2024. 2
- [9] Jia Guo, Shuai Lu, Lize Jia, Weihang Zhang, and Huiqi Li. Encoder-decoder contrast for unsupervised anomaly detection in medical images. *IEEE Transactions on Medical Imaging*, 2023. 1
- [10] Jia Guo, Shuai Lu, Weihang Zhang, Fang Chen, Huiqi Li, and Hongen Liao. Dinomaly: The less is more philosophy in multi-class unsupervised anomaly detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025. 2
- [11] Haoyang He, Yuhu Bai, Jiangning Zhang, Qingdong He, Hongxu Chen, Zhenye Gan, Chengjie Wang, Xiangtai Li, Guanzhong Tian, and Lei Xie. Mambaad: Exploring state space models for multi-class unsupervised anomaly detection. *Advances in Neural Information Processing Systems (NeurIPS)*, 2024. 2
- [12] Lars Heckler-Kram, Jan-Hendrik Neudeck, Ulla Scheler, Rebecca König, and Carsten Steger. The mvtec ad 2 dataset: Advanced scenarios for unsupervised anomaly detection. *arXiv preprint arXiv:2503.21622*, 2025. 1, 2, 3
- [13] Ziming Huang, Xurui Li, Haotian Liu, Feng Xue, Yuzhe Wang, and Yu Zhou. Anomalyncd: Towards novel anomaly class discovery in industrial scenarios. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025. 3
- [14] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4015–4026, 2023. 3
- [15] Renting Liu and Jiaya Jia. Reducing boundary artifacts in image deconvolution. In *IEEE International Conference on Image Processing (ICIP)*, pages 505–508. IEEE, 2008. 2
- [16] Zhikang Liu, Yiming Zhou, Yuansheng Xu, and Zilei Wang. Simplenet: A simple network for image anomaly detection and localization. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 20402–20411, 2023. 3, 4
- [17] Wei Luo, Yunkang Cao, Haiming Yao, Xiaotian Zhang, Jianan Lou, Yuqi Cheng, Weiming Shen, and Wenyong Yu. Exploring intrinsic normal prototypes within a single image for universal anomaly detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025. 2
- [18] Amira Ben Mabrouk and Ezzeddine Zagrouba. Abnormal behavior recognition for intelligent video surveillance systems: A review. *Expert Systems with Applications*, 91:480–491, 2018. 1
- [19] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics (TSMC)*, 9(1):62–66, 1979. 2
- [20] Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Schölkopf, Thomas Brox, and Peter Gehler. Towards total recall in industrial anomaly detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 3, 4
- [21] Tran Dinh Tien, Anh Tuan Nguyen, Nguyen Hoang Tran, Ta Duc Huy, Soan Duong, Chanh D Tr Nguyen, and Steven QH Truong. Revisiting reverse distillation for anomaly detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 3, 4
- [22] Mitchell Wortsman, Tim Dettmers, Luke Zettlemoyer, Ari Morcos, Ali Farhadi, and Ludwig Schmidt. Stable and low-precision training for large-scale vision-language models. *Advances in Neural Information Processing Systems (NeurIPS)*, 36:10271–10298, 2023. 3
- [23] Weijia Wu, Yuzhong Zhao, Mike Zheng Shou, Hong Zhou, and Chunhua Shen. Diffumask: Synthesizing images with pixel-level annotations for semantic segmentation using diffusion models. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023. 2
- [24] Guoyang Xie, Jinbao Wang, Jiaqi Liu, Jiayi Lyu, Yong Liu, Chengjie Wang, Feng Zheng, and Yaochu Jin. Im-iad: Industrial image anomaly detection benchmark in manufacturing. *IEEE Transactions on Cybernetics*, pages 1–14, 2024. 1
- [25] Ruijie Xu, Chuyu Zhang, Hui Ren, and Xuming He. Dual-level adaptive self-labeling for novel class discovery in point

- cloud segmentation. In *European Conference on Computer Vision*, pages 288–305. Springer, 2024. [2](#)
- [26] Hang Yao, Ming Liu, Zhicun Yin, Zifei Yan, Xiaopeng Hong, and Wangmeng Zuo. Glad: towards better reconstruction with global and local adaptive diffusion models for unsupervised anomaly detection. In *European Conference on Computer Vision (ECCV)*, pages 1–17. Springer, 2024. [2](#)
- [27] Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Dsr—a dual subspace re-projection network for surface anomaly detection. In *European Conference on Computer Vision (ECCV)*, 2022. [3](#), [4](#)
- [28] Ximiao Zhang, Min Xu, and Xiuzhuang Zhou. Realnet: A feature selection network with realistic synthetic anomaly for anomaly detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16699–16708, 2024. [2](#)
- [29] Yixuan Zhou, Xing Xu, Jingkuan Song, Fumin Shen, and Heng Tao Shen. Msflow: Multiscale flow-based framework for unsupervised anomaly detection. *IEEE Transactions on Neural Networks and Learning Systems (TNNLS)*, 2024. [3](#), [4](#)