Accurate Anomaly Localization in Challenging Industrial Settings via a Hybrid Detection Framework

Xingao Wang Shuying Xia Zhaohong Liao Mengjie Xie Handa Wang Zhi Gao
Wuhan University
{xingaowang, , gaozhinus}@whu.edu.cn

Abstract

In this project, we present a hybrid anomaly detection framework for the CVPR Workshop 2025 Visual Anomaly and Novelty Detection Challenge, focusing on unsupervised defect detection in complex industrial environments. Our method integrates multi-scale feature embedding and prototype-based reconstruction techniques to effectively handle a wide variety of real-world anomalies. To enhance the performance of the model, we apply a range of data augmentations to normal samples. In addition, we adopt a feature prototype alignment module that strengthens anomaly localization by refining reconstruction residuals based on spatial context. Experimental results on the MVTec AD 2 dataset demonstrate that our method performs competitively across multiple objects. The model achieves promising segmentation performance with high pixel-level F1 scores, particularly in scenarios involving transparency, clutter, and variable lighting.

1. Introduction

1.1. Background

Industrial anomaly detection plays a critical role in maintaining operational efficiency, safety, and quality control in manufacturing and other industrial sectors. By identifying defects early, anomaly detection systems prevent costly failures, improve product quality, and increase production reliability. With the increasing complexity of industrial systems, automated anomaly detection is essential to handle data that traditional manual inspection cannot efficiently process. Unlike traditional supervised learning problems, anomaly detection is intrinsically challenging due to its unsupervised or one-class learning nature, where models are trained exclusively on normal samples and lack prior knowledge of actual defect types.

Current commonly used anomaly detection datasets, such as MVTec AD [1] and VisA [27], are primarily used for research and evaluation in industrial visual anomaly detection tasks. However, they often fail to simulate the full complexity of real-world industrial environments, lacking important factors like variations in lighting, backgrounds, and defect types. To address these limitations, the newly released MVTec AD 2 [9] dataset takes a significant step forward in anomaly detection research. It introduces eight challenging industrial scenarios with high-resolution images captured under diverse lighting conditions. This dataset includes previously unstudied challenges, such as transparent and overlapping objects, dark-field and backlight illumination, and objects with high variance in normal data, as well as extremely small defects.

1.2. Challenge Description

There are several challenges in unsupervised anomaly detection using the MVTec AD 2 [9]. First, the dataset contains defects that are subtle, small, and located in areas that may not be easily detected using traditional methods. Additionally, the inclusion of transparent and reflective objects complicates anomaly detection further, as these materials often create distortions that can mask defects. Moreover, the dataset is designed with realistic settings in mind, featuring variations in illumination across different scenes. Since each scene is captured under at least four different lighting conditions, anomaly detection methods must be robust to lighting shifts which is a commonly encountered challenge in industrial inspection environments.

2. Related work

2.1. Unsupervised anomaly detection(UAD)

Unsupervised anomaly detection uses only normal images for training to effectively identify deviations from normal patterns. Most UAD approaches adhere to three main paradigms: feature embedding, reconstruction-based and teacher-student methods. Feature embedding methods [4, 5, 7, 17, 24] extract patch-level features from normal images to model the distribution of normality, and detect anomalies by measuring deviations between test samples and the established normal distribution. Reconstruction-

based methods [8, 10, 11, 15, 16, 21–23, 25, 26] train models to reconstruct input images so that the output closely resemble normal patterns, identifying anomalies by comparing the input and reconstructed outputs. Teacher-student methods [2, 3, 6, 13, 18–20] use a well-trained teacher network to guide a student network trained only on normal data, with anomalies detected by analyzing the response differences between student and teacher networks.

2.2. UAD via prototype learning

Prototype-based unsupervised anomaly detection identifies anomalies by comparing test features with representative prototypes derived from normal data. Early methods such as PatchCore [17] construct a memory bank of patch features and compute anomaly scores based on the distance between test patches and their nearest normal prototypes. CPR [12] improves this by introducing a two-stage cascade retrieval process that first selects the most similar training images and then performs patch-level matching, achieving higher efficiency and accuracy. More recently, INP-Former [14] advances this paradigm by dynamically extracting Intrinsic Normal Prototypes (INPs) from the test image itself, enabling better alignment with test-time distributions and enhancing robustness under real-world variability.

3. Methodology

3.1. Model design

3.1.1. Approach

As illustrated in Figure 1, our proposed framework consists of three main stages: data preprocessing, anomaly detection, and postprocessing.

In the data preprocessing stage, we adopt two key strategies to improve training robustness. First, we generate pseudo anomalies by pasting typical industrial defects (e.g., cracks, scratches) onto normal images, simulating realistic abnormal patterns. Second, to handle resolution variability in MVTec AD 2 [9], we divide large or non-square images into smaller patches at multiple pyramid scales, enhancing the model's adaptability across object sizes.

To address the challenge of robust anomaly detection under real-world conditions presented by the MVTec AD 2 [9] dataset, we adopted a hybrid strategy that combines two state-of-the-art methods: INP-Former and CPR to leverage the strengths of both architectures in different objects to maximize accuracy and robustness.

INP-Former is selected for the majority of object types due to its strong generalization under distribution shifts, such as changes in lighting, viewpoint, or texture. Unlike conventional prototype-based methods that rely on precomputed features from training data, INP-Former dynamically extracts INPs from each test image. This image-specific prototype extraction enables better alignment with local normal patterns, leading to more precise anomaly localization—especially in categories with diverse appearances or significant intra-class variation.

For simpler and more uniform object types like fruit jelly and vial, we adopt CPR, which performs class-specific prototype reconstruction using features retrieved from normal training images. This approach is well-suited for scenarios where a single foreground object dominates the scene and where spatial consistency aids accurate retrieval. While CPR may be memory-intensive for visually complex categories, applying it selectively to these two simpler object classes achieves stable performance without compromising system efficiency. This targeted assignment allows our hybrid framework to flexibly adapt across the wide variety of object types found in MVTec AD 2 [9].

In the postprocessing stage, we address an observed limitation in the raw anomaly maps—specifically, the tendency to emphasize anomaly boundaries while neglecting the full interior of anomalous regions. This issue is especially evident in fabric categories, where added materials or structures may be only partially detected. To mitigate this, we apply morphological closing and hole filling operations to the predicted anomaly maps, effectively completing large or hollow regions and improving the accuracy of final segmentation outputs.

In the evaluation phase, determining appropriate thresholds for converting anomaly maps into binary segmentation masks is crucial. A commonly adopted strategy in related literature is to set the threshold as the mean plus three times the standard deviation of the anomaly score. Through empirical observations, we noticed that the results does not strictly follow the statistical assumption of a $\mu + 3\sigma$ threshold. Moreover, the optimal threshold varied not only across different methods, but also among different objects, due to variations in anomaly distributions. Therefore, we adopted an empirical approach: by conducting multiple trials, we manually selected thresholds that achieved the best segmentation performance in terms of the F1 score and pixel-level accuracy. This method allows us to adapt flexibly to different scenarios and contributed to the final performance gains.

3.1.2. Architecture

The system consists of two parallel branches, each tailored to a specific group of objects.

The INP-Former branch adopts a transformer-based architecture composed of a frozen ViT-L/14 encoder pretrained with DINOv2, an INP Extractor, a Bottleneck, and an INP-Guided Decoder. The encoder extracts multi-scale patch-level features from the input image. The INP Extractor dynamically extracts intrinsic normal prototypes from a single image, which the INP-Guided Decoder leverages to effectively suppress anomalous features. The original loss functions are retained, including the INP Coherence Loss for maintaining semantic consistency among the pro-

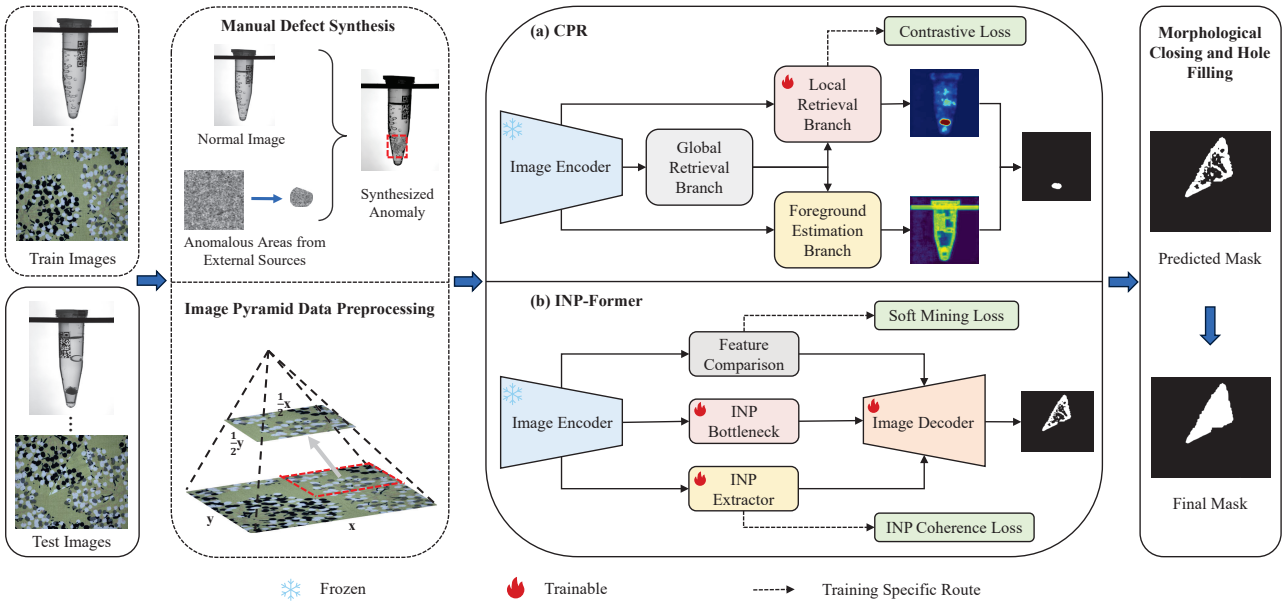

Figure 1. Our proposed framework, containing data preprocessing, anomaly detection and postprocessing.

totypes and the Soft Mining Loss to focus learning on hard-to-reconstruct normal regions. Additionally, a multi-scale bottleneck fusion is used to enhance robustness to spatial and scale variations.

The CPR branch follows the dual-path design proposed in the original method, consisting of the DenseNet201 backbone, the global retrieval module, the local retrieval module and the Foreground Estimation Branch. While the global retrieval remains unchanged, we introduce a key modification in the training of the Local Retrieval Branch. Instead of using the original learned transformation module to synthesize anomalies, we adopt a more controllable cut-paste strategy. Specifically, common industrial defects such as cracks, spots, and scratches are extracted and pasted onto clean images from the MVTec AD 2 [9] dataset to create pseudo-anomalous samples (Details see 3.2.1).

3.1.3. Training

In the training phase, each object category is treated independently, with its own model trained using either INP-Former or CPR based on its visual characteristics. No real abnormal images are used during training, in accordance with the challenge's unsupervised anomaly detection requirement. All experiments were conducted on a workstation equipped with a single NVIDIA RTX 4090 GPU.

For the INP-Former models, we use a batch size of 16 and train for 10 epochs using the StableAdamW optimizer, with a learning rate of 1e-3 and weight decay of 1e-4. The number of extracted INPs is set to six. Input images are resized to 448×448, and standard data augmentations such as horizontal flipping, color jittering, and random scaling are applied to improve robustness against real-world variability. INP-Former is optimized using a combination of INP Coherence Loss and Soft Mining Loss to ensure prototype consistency and enhance sensitivity to subtle anomalies.

For the CPR models, we follow a different configuration. A batch size of 8 is used, and training is conducted for 1300 epochs on the vial category and 2000 epochs on fruit jelly. The optimizer is AdamW with a learning rate of 0.001 and weight decay of 0.01. Input images are resized to 320×320. CPR is trained with its original combination of reconstruction and contrastive prototype losses to enable stable and discriminative learning.

3.2. Dataset and Evaluation

3.2.1. Data Utilization

To help our model effectively learn normal prototypes across different scales, we carefully preprocess the dataset. For certain objects, like fruit jelly, vial, and wallplugs, we keep the original image size. However, for objects such as fabric, can, rice, sheet metal, and walnuts, we create an image pyramid by dividing the original resolution image into multiple sub-images. This resizing approach is based on the observation that smaller, more evenly distributed objects, like rice and walnuts, benefit from higher resolution to capture fine anomalies. Cutting these objects at a higher resolution does not result in detail loss; instead, it helps magnify and capture subtle details. On the other hand, objects like jelly or vials, which are typically located at the center of the image, could lose critical information if cut, leading to a decrease in performance.

In addition to resolution adaptation, we perform synthetic anomaly generation for certain objects such as vial

and fruit jelly. We observe that many industrial defects share similar visual characteristics such as cracks, spots, and scratches. Therefore, we manually extract anomaly regions and paste them onto clean images from the MVTec AD 2 [9] dataset to create pseudo anomalous samples. This helps enhance the model's robustness and its ability to generalize to realistic anomalies, especially in objects where real anomaly samples were limited.

3.2.2. Evaluation Criteria

To evaluate the performance of our anomaly detection model, we used the official evaluation protocol. The primary metric is the F_1 score, which is the harmonic mean of precision and recall at the pixel level across different testing environments and objects in the MVTec AD 2 dataset.

The F_1 score is calculated as:

$$F_1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{1}$$

where Precision is the ratio of true positive (TP) pixels to the total predicted positive pixels, and Recall is the ratio of TP pixels to the total actual positive pixels.

In addition to the F_1 score, we also use the Area Under the Per-Region-Overlap Curve (AU-PRO) to assess the localization quality. This metric evaluates how well the model localizes anomalies within regions across the test set.

The per-region-overlap (PRO) at a given threshold is defined as:

$$\text{PRO} = \frac{1}{k} \sum_{i=1}^{|D_{\text{test}}|} \sum_{j=1}^{k_i} \frac{|P_{\text{ano}} \cap C_{i,j}|}{|C_{i,j}|} \tag{2}$$

where $C_{i,j}$ is the j-th ground truth region in image i, P_{ano} is the predicted anomalous region, and k is the total number of ground truth regions.We use an official threshold of 0.05 for the integration limit of the PRO curve ($AucPRO_{0.05}$), as recommended by the challenge organizers.

4. Results

4.1. Performance Metrics

The test data is divided into three parts: a publicly available example test set ($TEST_{pub}$) and two private test sets ($TEST_{priv}$ and $TEST_{priv,mix}$). $TEST_{priv}$ contains images with the same lighting conditions as the training set, while $TEST_{priv,mix}$ includes images from the same scenes as $TEST_{priv}$, but with randomly selected lighting conditions, encompassing both seen and unseen lighting, as well as both normal and anomalous data.

The results shown in Table 1 indicate that our approach performs robustly across various objects, with consistent performance even under changes in lighting. Notably, sheet metal achieved the highest score of 85.19% for $TEST_{priv}$. Objects such as fabric, fruit jelly, rice, sheet metal, vial and walnuts show high $AucPRO_{0.05}$ scores, consistently above 60%.

Overall, the $TEST_{priv}$ conditions yield slightly higher $AucPRO_{0.05}$ scores than the $TEST_{priv,mix}$ conditions, suggesting that lighting variations have a noticeable but manageable impact on the model's performance. Despite the challenges posed by lighting changes, our method demonstrates strong robustness.

Table 1. Segmentation $AucPRO_{0.05}$ Score (in %) of the Proposed Method on Binarized Images for $TEST_{priv}$ / $TEST_{priv,mix}$ Set

Object	$TEST_{priv}$	$TEST_{priv,mix}$
Can	46.55	40.67
Fabric	65.94	69.19
Fruit Jelly	62.45	62.38
Rice	65.25	66.78
Sheet Metal	86	85.19
Vial	64.44	60.2
Wall Plugs	26.63	22.04
Walnuts	80.47	76.57
Mean	62.22	60.38

4.2. Comparison

As shown in Table 2, our approach outperforms many others across multiple objects. For instance, under the private set, our method outperforms PatchCore, RD, RD++, and other baseline methods by a significant margin in objects like fabric and walnuts. Moreover, our method maintains a consistent advantage across the majority of objects, particularly in handling variations in lighting. The overall average of 49.01% for the $TEST_{priv}$ set and 46.26% for the $TEST_{priv,mix}$ set shows our method's robustness, significantly outperforming the other baseline methods.

5. Discussion

5.1. Challenges and Solutions

We encountered several challenges during the development of our approach. One prominent issue was the high diversity of anomaly types. Some anomalies in the dataset appeared subtle, with complex textures and low contrast, while others were extremely small and difficult to detect using traditional methods. To address this issue, we applied a series of data pre-processing strategies tailored to enhance anomaly visibility (see Section 3.2.1). Meanwhile, for objects like fruit jelly and vial, we adopted the CPR method(see Section 3.1.1), which performs well in images featuring simple object structures dominated by a single foreground object. In addition to the diversity of anomalies, the unsupervised

Table 2. Comparison of Segmentation F_1 Score (in%) Across Different Methods on Binarized Images for $TEST_{priv}$ / $TEST_{priv,mix}$ Set

Object	Methods							
	Ours	PatchCore	RD	RD++	EfficientAD	MSFlow	SimpleNet	DSR
Can	**17.03 / 9.21**	0.3 / 0.1	0.1 / 0.1	0.1 / 0.1	0.8 / 0.1	5.0 / 0.1	0.6 / 0.1	0.4 / 0.1
Fabric	**84.95 / 81.76**	11.5 / 9.8	2.6 / 2.2	2.9 / 2.3	7.6 / 1.0	22.0 / 4.1	21.6 / 10.2	7.9 / 5.0
Fruit Jelly	**35.66 / 36.08**	8.7 / 8.2	22.5 / 22.7	26.9 / 26.7	20.8 / 18.2	47.6 / 38.1	25.1 / 23.0	17.9 / 17.2
Rice	**66.34 / 66.93**	3.8 / 4.2	7.0 / 3.9	9.5 / 2.9	15.0 / 0.5	19.1 / 1.8	11.6 / 1.0	1.5 / 1.4
Sheet Metal	**69.69 / 69.74**	1.8 / 1.1	41.3 / 39.2	40.9 / 37.7	9.3 / 3.8	13.0 / 7.6	14.6 / 2.8	13.9 / 14.4
Vial	**39.5 / 37.65**	2.3 / 2.2	28.0 / 28.3	28.2 / 22.8	30.5 / 26.5	23.3 / 6.2	31.9 / 17.5	28.2 / 27.9
Wall Plugs	**11.99 / 6.24**	0.0 / 0.0	1.9 / 0.8	1.3 / 0.9	4.4 / 0.3	0.1 / 0.2	1.0 / 0.3	0.4 / 0.4
Walnuts	**66.9 / 62.45**	1.2 / 1.3	41.2 / 36.7	44.1 / 40.5	34.6 / 13.3	44.5 / 14.3	35.2 / 14.3	17.0 / 9.6
Mean	**49.01 / 46.26**	3.7 / 3.4	18.1 / 16.7	19.2 / 16.7	15.4 / 8.0	21.8 / 9.0	17.7 / 8.7	10.9 / 9.5

nature of training posed another difficulty. Since the model was trained solely on normal data, ensuring robust detection of unseen anomalies required careful tuning of residual thresholds and the design of effective feature alignment mechanisms. Moreover, variations in illumination introduced further complexity. These changes can greatly affect the appearance of both normal and anomalous samples, demanding that the model be robust to such real-world variability. Therefore, we adopted the INP-Former architecture, which excels in capturing complex feature representations(see Section 3.1.1).

5.2. Model Robustness and Adaptability

Our model is designed with robustness and adaptability in mind, particularly for real-world industrial environments where variations in lighting, object placement, and background textures are common. By incorporating multi-scale feature learning and prototype-based reconstruction, our method maintains consistent performance across different conditions.

5.3. Future Work

While our method demonstrates solid results in the MVTec AD 2 dataset, there is a need for models that can adapt to unseen anomalies in more diverse industrial settings. Future work could focus on the development of few-shot learning approaches, which would enable the model to better handle rare or previously unseen defect types with limited training data. The robustness of our model under real-world conditions, such as varying lighting and environmental changes, could be further enhanced. Incorporating more advanced data augmentation techniques, like adversarial training or domain adaptation, might allow the model to generalize better across different industrial scenarios.

6. Conclusion

In conclusion, we propose a hybrid anomaly detection framework based on INP-Former and CPR, designed for robust performance under industrial inspection scenarios. Our method is trained using a strict one-class paradigm and leverages pyramid-based preprocessing, dynamic INP prototype learning, and fine-grained postprocessing techniques. This demonstrates strong image and pixel-level anomaly detection performance on the MVTec AD 2 dataset across diverse objects, withstanding lighting variations. These architectural choices allow the model to extract precise normal features and isolate abnormalities with high confidence. The postprocessing pipeline further reduces noise and ensures segmentation completeness, helping the model remain stable under real-world data fluctuations. Our results confirm the scalability and generalization capacity of our approach, marking a promising step toward deploying anomaly detection in practical manufacturing environments.

References

[1] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Mvtec ad–a comprehensive real-world dataset for unsupervised anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9592–9600, 2019. 1

[2] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4183–4192, 2020. 2

[3] Yunkang Cao, Qian Wan, Weiming Shen, and Liang Gao. Informative knowledge distillation for image anomaly segmentation. *Knowledge-Based Systems*, 248:108846, 2022. 2

[4] Niv Cohen and Yedid Hoshen. Sub-image anomaly detec-

tion with deep pyramid correspondences. *arXiv preprint, arXiv:2005.02357*, 2020. 1

[5] Thomas Defard, Aleksandr Setkov, Angelique Loesch, and Romaric Audigier. Padim: a patch distribution modeling framework for anomaly detection and localization. In *Proceedings of the International Conference on Pattern Recognition*, pages 475–489. Springer, 2021. 1

[6] Hanqiu Deng and Xingyu Li. Anomaly detection via reverse distillation from one-class embedding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9727–9736, 2022. 2

[7] Denis Gudovskiy, Shun Ishizaka, and Kazuki Kozuka. Cflow-ad: Real-time unsupervised anomaly detection with localization via conditional normalizing flows. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 98–107, 2022. 1

[8] Hewei Guo, Liping Ren, Jingjing Fu, Yuwang Wang, Zhizheng Zhang, Cuiling Lan, Haoqian Wang, and Xinwen Hou. Template-guided hierarchical feature restoration for anomaly detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6424–6435, 2023. 2

[9] 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, 4

[10] Jinlei Hou, Yingying Zhang, Qiaoyong Zhong, Di Xie, Shiliang Pu, and Hong Zhou. Divide-and-assemble: Learning block-wise memory for unsupervised anomaly detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8791–8800, 2021. 2

[11] Teng Hu, Jiangning Zhang, Ran Yi, Yuzhen Du, Xu Chen, Liang Liu, Yabiao Wang, and Chengjie Wang. Anomalydiffusion: Few-shot anomaly image generation with diffusion model. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 8526–8534, 2024. 2

[12] Hanxi Li, Jianfei Hu, Bo Li, Hao Chen, Yongbin Zheng, and Chunhua Shen. Target before shooting: Accurate anomaly detection and localization under one millisecond via cascade patch retrieval. *IEEE Transactions on Image Processing*, 33: 5606–5621, 2024. 2

[13] Yufei Liang, Jiangning Zhang, Shiwei Zhao, Runze Wu, Yong Liu, and Shuwen Pan. Omni-frequency channel-selection representations for unsupervised anomaly detection. *IEEE Transactions on Image Processing*, 32:4327–4340, 2023. 2

[14] 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. arXiv preprint arXiv:2503.02424, 2025. 2

[15] Hyunjong Park, Jongyoun Noh, and Bumsub Ham. Learning memory-guided normality for anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14360–14369, 2020. 2

[16] Jonathan Pirnay and Keng Chai. Inpainting transformer for anomaly detection. In *International Conference on Image Analysis and Processing*, pages 394–406. Springer, 2022. 2

[17] Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Schölkopf, Thomas Brox, and Peter Gehler. Towards total recall in industrial anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14318–14328, 2022. 1, 2

[18] Mohammadreza Salehi, Niousha Sadjadi, Soroosh Baselizadeh, Mohammad H. Rohban, and Hamid R. Rabiee. Multiresolution knowledge distillation for anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14897–14907, 2021. 2

[19] Xian Tao, Dapeng Zhang, Wenzhi Ma, Zhanxin Hou, Zhenfeng Lu, and Chandranath Adak. Unsupervised anomaly detection for surface defects with dual-siamese network. *IEEE Transactions on Industrial Informatics*, 18(11):7707–7717, 2022.

[20] Tran Dinh Tien, Anh Tuan Nguyen, Nguyen Hoang Tran, Ta Duc Huy, Soan T.M. Duong, Chanh D. Tr. Nguyen, and Steven Q. H. Truong. Revisiting reverse distillation for anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 24511–24520, 2023. 2

[21] Julian Wyatt, Adam Leach, Sebastian M. Schmon, and Chris G. Willcocks. Anoddpm: Anomaly detection with denoising diffusion probabilistic models using simplex noise. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 649–655, 2022. 2

[22] Xudong Yan, Huaidong Zhang, Xuemiao Xu, Xiaowei Hu, and Pheng-Ann Heng. Learning semantic context from normal samples for unsupervised anomaly detection. In *Proceedings of the AAAI conference on artificial intelligence*, pages 3110–3118, 2021.

[23] Xincheng Yao, Ruoqi Li, Zefeng Qian, Yan Luo, and Chongyang Zhang. Focus the discrepancy: Intra- and inter-correlation learning for image anomaly detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6803–6813, 2023. 2

[24] Jihun Yi and Sungroh Yoon. Patch svdd: Patch-level svdd for anomaly detection and segmentation. In *Proceedings of the Asian Conference on Computer Vision*, pages 375–390, 2020. 1

[25] Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Draem - a discriminatively trained reconstruction embedding for surface anomaly detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8330–8339, 2021. 2

[26] Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Reconstruction by inpainting for visual anomaly detection. *Pattern Recognition*, 112:107706, 2021. 2

[27] Yang Zou, Jongheon Jeong, Latha Pemula, Dongqing Zhang, and Onkar Dabeer. Spot-the-difference self-supervised pretraining for anomaly detection and segmentation. In *European Conference on Computer Vision*, pages 392–408. Springer, 2022. 1