# A Review of Industrial Image Anomaly Detection Based on Unsupervised Deep Learning

**Zhexin Huang**

School of Business Administration, Jimei University, Xiamen, China
Email: jessiehuang_zx@163.com

**Abstract** Under industry 4.0, automated quality inspection, aiming to automatically identify surface defects on product, is vital for intelligent manufacturing. In the real-world industrial scenarios, the lack of defective samples and the high cost of manual annotation pose significant obstacles to the effective application of supervised learning methods. Unsupervised deep learning uses only normal samples to train the model, which provides a viable approach. This review surveys unsupervised deep learning-based methods for industrial image anomaly detection. First, the necessity of unsupervised algorithms and the main challenges they face were demonstrated. Existing approaches are classified into five categories: reconstruction-based, teacher–student network-based, feature memory repository-based, generative network-based, and vision–language model-based. For each category, I discuss its underlying mechanism, advantages, failure scenarios, and applicable tasks, and further provide a comparative analysis among them. Then, common benchmark datasets such as MVTec AD and evaluation metrics such as AUROC and PRO were introduced, and the influence of datasets and metrics on method evaluation was discussed. Finally, open challenges such as cross-domain generalization and real-time constraints were summarized, and future research trends were prospected from directions such as multimodal fusion and lightweight deployment.

**Keywords** Industrial Image Anomaly Detection; Unsupervised Deep Learning; Deep Learning; Anomaly Localization; Industrial Inspection

## 1. Introduction

Industry 4.0 has brought significant changes to manufacturing, driving the industry towards more intelligent and automated production. In this new environment, quality control has become a key part of smart manufacturing. Traditional manual inspection is affected by subjective fatigue and inconsistent standards, making it inefficient and unreliable. With respect to rule-based methods, they depend on handcrafted features, so their generalization ability is limited and they struggle to cope with complex and changing defect patterns. Therefore, data-driven methods, which have achieved breakthroughs in visual tasks with the rapid development of deep learning, are gradually replacing traditional manual visual checks and rule-based methods, becoming the mainstream solution in industrial inspection.

However, supervised learning depends on a large number of labelled defect samples. In industrial scenarios, defect samples are scarce, their types are diverse and unpredictable, and pixel-level annotation is extremely costly, which limits its application. Unsupervised learning, in contrast, trains models only on normal samples. It identifies anomalies by recognizing samples that deviate from the normal distribution. Since it requires no defect labels, it is cost-effective and scalable, thereby making it closer to real industrial needs.

Despite these advantages, unsupervised industrial anomaly detection still faces many challenges: defect types are highly diverse, and a single model can hardly cover all abnormal patterns; variations in lighting, viewpoint and surface texture affect model robustness; high-precision models are computationally complex and struggle to meet the real-time requirements of production lines; and achieving pixel-level localization without defect labels is especially difficult.

This paper systematically reviews unsupervised deep-learning-based methods for industrial image anomaly detection. Its contributions are as follows: first, existing methods are classified into five categories: reconstruction-based, teacher-student network-based, feature

ISSN 3106-7018 (Print) ISSN 3106-5791 (Online)

memory bank-based, generative network-based, and vision-language model-based. Note that the vision-language model-based category often involves zero-shot or few-shot settings rather than conventional unsupervised learning; however, since these methods share the core idea of training with only normal samples, they are included within the scope of this review. Their mechanisms, strengths and failure cases are compared. Second, commonly used benchmark datasets and evaluation metrics are reviewed. Third, open challenges are summarized and future directions are discussed. The rest of the paper is organized as follows: Section 2 presents the classification of methods, Section 3 introduces datasets and metrics, Section 4 discusses challenges and future work, and Section 5 concludes the paper.

## 2. Methodology

Unsupervised industrial image anomaly detection methods can be categorized into five technical routes according to their core anomaly modelling mechanism. This classification is not strictly mutually exclusive. Some methods draw on ideas from more than one category, and vision-language approaches often involve zero-shot or few-shot settings rather than conventional unsupervised learning. When that happens, we place a method in the category that best reflects its primary mechanism. The following subsections take up each category in turn, covering its basic mechanisms, advantages, main failure scenarios, and applicable industrial tasks. The section closes with a cross-category comparison of the five methods.

### 2.1 Reconstruction-based Methods

The reconstruction-based approach is one of the earliest developed paradigms for anomaly detection. The core idea that training a model to learn the underlying distribution of normal samples enables the accurate reconstruction of normal images is straightforward. During the test, the abnormal samples were outside the normal distribution range, thus resulting in significant reconstruction errors, which can serve as the fundamental basis for identifying abnormalities. Typical examples are Autoencoders (AE), Variational Autoencoders (VAE), and memory-augmented autoencoders [1-2]. Such methods are intuitive, easy to implement and have low training costs, so they are often used as baseline approaches. However, their main failure scenario is over-generalization. If anomalies are semantically similar to normal patterns, the model may still reconstruct anomalous regions fairly well, and the anomaly is then missed. Therefore, these methods may be more suitable for texture defect detection tasks with high structural regularity, such as surface inspection of fabric, leather, and wood [3].

### 2.2 Teacher-student Network-based Methods

Knowledge distillation is widely used in teacher-student anomaly detection. The basic process is relatively simple. First of all, the teacher network conducts pre-training on large data sets in order to learn general visual features. After the teacher's network is frozen, its output is used as the target of training the student network. Students are trained in general industrial data and learn to copy teachers' responses to these samples. In the reasoning stage, abnormal samples may cause differences between the outputs of two networks, which can then be used to identify and locate defects. An important advantage of this method is its high detection accuracy, especially its ability to perform pixel-level positioning. However, the method also depends to a large extent on the teacher's model. If teachers do not have enough discrimination ability, students may not be able to learn useful feature representations, which may lead to reduced detection performance. For this reason, this method is applicable to accurate component detection, including electronic components and mechanical components, of which accurate abnormal positioning is particularly important [4].

### 2.3 Feature Memory Bank-Based Methods

These methods first extract features from a large number of normal samples and store them in the feature memory bank. During the detection process, distance measurement is used to compare the features of the test sample with the stored normal prototypes. Then use the distance to determine whether the sample is abnormal. PatchCore and its variants are representative examples of this method and achieve the most advanced detection performance [5]. Another advantage is that distance measurement have relatively clear geometric meanings, which make the test results easier to interpret. However, these methods may require a lot of memory and storage, especially when the data set is large. Their performance is also affected by non-semantic changes in appearance, such as lighting or color changes, which may lead to false alarms. In practice, they are more suitable for standardized production line testing, in which the normal appearance is clearly defined and changes are relatively small [6].

### 2.4 Generative Network-Based Methods

Unlike reconstruction-based methods, which mainly rely on autoencoders, generative approaches are built on GANs or diffusion models. These models learn the distribution of normal samples and can reconstruct or generate normal images. At test time, an anomalous sample is often poorly reconstructed, or the discriminator gives an abnormal response, so the defect can be identified. Representative methods include GANomaly and diffusion-based anomaly detection [7]. Such methods offer strong generative modelling ability and suit scenarios with highly complex normal patterns. However, GAN training is unstable and prone to mode collapse, while diffusion models are computationally expensive and slow at inference, making real-time deployment difficult. They are therefore more

suitable for surface inspection with diverse normal patterns and complex defect forms, such as natural textures and cast surfaces [8].

### 2.5 Vision-Language Model-Based Methods

Visual-language model based methods are particularly useful when the production line changes rapidly and labeled data is scarce. They can exploit large pre-trained models like CLIP for cross-modal understanding, which enables zero-shot or few-shot anomaly detection. WinCLIP and AnomalyGPT are two typical examples [9-10]. Text descriptions can also specify the anomaly type, so the detection target can be defined during inference and anomaly detection with an open vocabulary is supported. Deployment is the main obstacle: these models are large in size and costly to run, which makes them difficult to use on edge devices and largely confines them to the cloud or offline analysis [11].

### 2.6 Cross-category Comparison

Several studies have compared different categories of anomaly detection methods. Cui et al. [1] compared multiple classes of unsupervised algorithms and summarized the advantages and disadvantages reported for each class. Tao et al. [2] focused on anomaly localization and gave quantitative comparisons of the performance of different localization methods. Diers and Pigorsch [12] reviewed image-based automated quality control methods and discussed the practical trade-offs involved in different approaches. Taken together, these studies suggest that different method categories have different trade-offs, rather than there being one approach that is dominant across all cases. Feature memory bank-based methods, such as PatchCore, have reported state-of-the-art detection performance on benchmark datasets [5]. Generative methods represented by GANomaly have also shown competitive detection accuracy in some settings [7]. Reconstruction-based methods are often considered relatively intuitive and easy to implement, although their detection accuracy has been reported to have certain limitations [1]. Vision-language model-based methods have shown strong zero-shot capability, while their deployment can still be affected by factors such as model size and inference cost [9][10]. Tao et al. [2] also carried out a separate comparison of localization methods, which can provide a useful reference for understanding how different method categories are positioned in terms of pixel-level localization.

Overall, existing studies indicate that different categories have their own strengths as well as different failure scenarios. The choice of method is therefore related to the specific industrial task, data conditions, and deployment constraints. A unified ranking across categories is difficult to establish, since the evaluation settings and experimental conditions are not necessarily consistent across studies.

## 3. Datasets and Evaluation Metrics

Comparisons between different methods can be affected by the choice of datasets and evaluation metrics. This section first introduces the benchmark datasets and evaluation metrics commonly used in this field, and then considers how these choices influence the evaluation of method performance.

### 3.1 Public Benchmark Datasets

MVTec AD is currently the most widely used benchmark dataset, containing 15 categories of texture and object images and serving as the standard evaluation baseline for most unsupervised anomaly detection methods [13]. MVTec AD 2 is an advanced version that introduces more challenging scenarios, such as subtle defects and complex backgrounds [14]. As for real industrial scenarios, BTAD can provide real defect images of three industrial products, making its evaluation setting closer to actual production lines [15]. MSC-AD, then, focuses on casting surface defects by multiple scenarios, making it suitable for evaluating cross-scenario generalization [16]. In addition, datasets such as KSDD (steel surface defects) and DAGM (textured surface defects) are also widely adopted in specific tasks [17-18].

The choice of dataset directly affects conclusions of evaluations of methods. Texture-intensive datasets, such as the texture categories in MVTec AD, have strong structural regularity and are better suited to evaluating reconstruction-based methods. In contrast, structural datasets, such as the object categories in MVTec AD, place higher demands on localization accuracy and are more useful for testing teacher-student and memory bank-based methods.

### 3.2 Evaluation Metrics

For image-level detection, AUROC (Area Under the Receiver Operating Characteristic curve) is commonly used to evaluate whether a model can distinguish normal images from anomalous ones. In this case, an entire image is treated as one sample. The same metric can also be used for pixel-level localization, but the unit of evaluation changes: each pixel is treated as an individual sample, and the resulting AUROC reflects how well anomalous pixels can be distinguished from normal ones.

PRO (Per-Region Overlap) is commonly used alongside pixel-level AUROC when evaluating localization. Rather than evaluating pixels separately, it considers the overlap between the predicted anomalous regions and the corresponding ground-truth regions. This makes PRO more directly related to the quality of the detected regions. F1-score, precision, and recall are also frequently included in reported results, particularly when performance at a specific threshold needs to be considered.

The choice of metrics is therefore relevant when comparing different methods. Image-level detection and pixel-level localization focus on different aspects of anomaly

ISSN 3106-7018 (Print) ISSN 3106-5791 (Online)

detection, so a single metric does not fully describe both. AUROC provides a way to evaluate the separation of normal and anomalous samples at either the image or pixel level, while PRO provides additional information about the overlap of anomalous regions. F1-score, precision, and recall can further describe the detection results under a given threshold.

## 4. Challenges and Future Directions

The method categories and evaluation metrics discussed above show that unsupervised industrial anomaly detection has a clear advantage: it depends much less on annotation. Even so, it remains far from large-scale industrial deployment. These gaps are the main challenges the field faces today, and they also point to directions for future research.

### 4.1 Open Challenges

Section 2 points to four limitations that keep recurring. Cross-domain generalization is weak across all method categories: models trained on one benchmark often fail on new products or production lines. Reconstruction-based and generative methods are hit hardest, since they depend heavily on the training distribution. Real-time deployment runs into computational cost, which is most severe for diffusion-based generative models and vision-language models. Teacher-student and memory bank methods sit in the middle, while reconstruction-based methods are the lightest. Fine-grained localization also remains difficult, especially for reconstruction-based and generative methods. Teacher-student and memory bank approaches do better, but they still need improvement. Evaluation standards, meanwhile, are inconsistent: the lack of unified protocols and the mixing of image-level and pixel-level metrics affect every category and make fair comparison difficult.

These four issues lead directly to the future directions in Section 4.2.

### 4.2 Future Research Directions

Closing these gaps will require work on several fronts. Cross-domain and few-shot learning offer a promising start: by drawing on vision-language foundation models and meta-learning strategies, it can help models adapt to new scenarios. Multimodal fusion points to another direction: combining RGB, 3D point cloud, and infrared data can offset the weaknesses of any single modality in complex defect detection. Interpretability is equally important, since industrial users need to conduct root-cause analysis and place greater trust in automated decisions. Edge deployment also calls for lightweight solutions, where model compression and acceleration are used to reduce inference costs. Evaluation itself needs attention: standardized protocols, unified benchmarks, and metrics would allow different methods to be compared under the same conditions.

## 5. Conclusion

This paper has reviewed unsupervised deep learning methods for industrial image anomaly detection. By training only on normal samples, unsupervised learning effectively eases a central problem in industrial settings: defect samples are scarce and annotation is expensive. The methods discussed here fall into five categories—reconstruction-based, teacher-student network-based, feature memory bank-based, generative network-based, and vision-language model-based—and each comes with its own applicable tasks and failure boundaries. However, it should be noted that this classification is not strictly mutually exclusive, as some methods draw on ideas from more than one category, and vision-language approaches often involve zero-shot or few-shot settings rather than conventional unsupervised learning. In addition, the experimental protocols and evaluation metrics used across different studies are not always consistent, which makes direct cross-paper comparison difficult. That said, true deployment of unsupervised anomaly detection still hinges on cross-scenario generalization, reliable pixel-level localization, keeping computational cost under control, and consistency in evaluation protocols. Further work should continue to advance cross-domain learning, multimodal fusion, and lightweight deployment, so the technology can move from benchmark testing to real industrial application.

## 6. References

[1] Cui, Y., Liu, Z., & Lian, S. (2023). A survey on unsupervised anomaly detection algorithms for industrial images. IEEE Access, 11, 55297–55315. https://doi.org/10.1109/ACCESS.2023.3282992

[2] Tao, X., Gong, X., Zhang, X., Yan, S., & Adak, C. (2022). Deep learning for unsupervised anomaly localization in industrial images: A survey. IEEE Transactions on Instrumentation and Measurement, 71, 1–21. https://doi.org/10.1109/TIM.2022.3196436

[3] Liu, J., Xie, G., Wang, J., Li, S., Wang, C., Zheng, F., & Jin, Y. (2024). Deep industrial image anomaly detection: A survey. Machine Intelligence Research, 21(1), 104–135. https://doi.org/10.1007/s11633-023-1459-z

[4] Li, Z., Yan, Y., Wang, X., Ge, Y., & Meng, L. (2025). A survey of deep learning for industrial visual anomaly detection. Artificial Intelligence Review, 58, Article 279. https://doi.org/10.1007/s10462-025-11287-7

[5] Roth, K., Pemula, L., Zepeda, J., Schölkopf, B., Brox, T., & Gehler, P. (2022). Towards total recall in industrial anomaly detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 14318–14328). IEEE.

[6] Cui, Y., Liu, Z., & Lian, S. (2025). A survey on RGB, 3D, and multimodal approaches for unsupervised industrial image anomaly detection. Information Fusion, 120, 103139. https://doi.org/10.1016/j.inffus.2025.103139

[7] Akçay, S., Atapour-Abarghouei, A., & Breckon, T. P. (2018). GANomaly: Semi-supervised anomaly detection

via adversarial training. In Proceedings of the Asian Conference on Computer Vision (ACCV) (pp. 622–637). Springer.

[8] Shukla, V., Shukla, A., S. K., S. P., & Shukla, S. (2025). A systematic survey: Role of deep learning-based image anomaly detection in industrial inspection contexts. Frontiers in Robotics and AI, 12, Article 1554196. https://doi.org/10.3389/frobt.2025.1554196

[9] Jeong, J., Zou, Y., Kim, T., Zhang, D., Ravichandran, A., & Dabeer, O. (2023). WinCLIP: Zero-/few-shot anomaly classification and segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 19606–19616). IEEE.

[10] Gu, Z., Zhu, B., Zhu, G., Chen, Y., Tang, M., & Wang, J. (2024). AnomalyGPT: Detecting industrial anomalies using large vision-language models. In Proceedings of the AAAI Conference on Artificial Intelligence, 38(3), 1932–1940.

[11] Liso, A., Cardellicchio, A., Patruno, C., Nitti, M., Ardino, P., & Stella, E. (2024). A review of deep learning-based anomaly detection strategies in Industry 4.0 focused on application fields, sensing equipment, and algorithms. IEEE Access, 12, 93911–93923. https://doi.org/10.1109/ACCESS.2024.3424242

[12] Diers, J., & Pigorsch, C. (2023). A survey of methods for automated quality control based on images. International Journal of Computer Vision, 131(10), 2553–2581. https://doi.org/10.1007/s11263-023-01822-w

[13] Bergmann, P., Fauser, M., Sattlegger, D., & Steger, C. (2019). MVTec AD — A comprehensive real-world dataset for unsupervised anomaly detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 9592–9600). IEEE.

[14] Heckler-Kram, L., Neudeck, J.-H., Scheler, U., König, R., & Steger, C. (2026). The MVTec AD 2 dataset: Advanced scenarios for unsupervised anomaly detection. International Journal of Computer Vision, 134(4). https://doi.org/10.1007/s11263-026-02743-0

[15] Mishra, P., Verk, R., Fornasier, D., Piciarelli, C., & Foresti, G. L. (2021). VT-ADL: A vision transformer network for image anomaly detection and localization. In 2021 IEEE 30th International Symposium on Industrial Electronics (ISIE) (pp. 1–6). IEEE.

[16] Zhao, Q., Wang, Y., Wang, B., Lin, J., Yan, S., Song, W., Liotta, A., Yu, J., Gao, S., & Zhang, W. (2024). MSC-AD: A multiscene unsupervised anomaly detection dataset for small defect detection of casting surface. IEEE Transactions on Industrial Informatics, 20(4, Pt. 1), 6041–6052. https://doi.org/10.1109/TII.2023.3341259

[17] Tabernik, D., Šela, S., Skvarč, J., & Skočaj, D. (2019). Segmentation-based deep-learning approach for surface-defect detection. Journal of Intelligent Manufacturing, 31, 759–776. https://doi.org/10.1007/s10845-019-01476-x

[18] Wieler, M., & Hahn, T. (2007). Weakly supervised learning for industrial optical inspection [Data set]. Robert Bosch Corporate Research. https://hci.iwr.uni-heidelberg.de/content/weakly-supervised-learning-industrial-optical-inspection