

Unsupervised Image Anomaly Detection: A Review of Autoencoder and Improved Clustering Fusion Methods

Zhang Zixuan

Anhui Polytechnic University, Wuhu, Anhui 241000, China

2087128825@qq.com

ABSTRACT

Unsupervised image anomaly detection requires no labeled data and holds broad application prospects in fields such as industrial quality inspection and medical imaging. The dual-driven approach combining autoencoders and clustering techniques represents one of the core research directions in this field. This paper provides a systematic review of this integration approach, outlining the technical evolution and current research status of autoencoder-based, clustering-based, and hybrid methods. It establishes a classification framework based on feature extraction, clustering strategies, and model fusion dimensions, while comparing the advantages, limitations, and applicable scenarios of various approaches. Key bottlenecks identified include the misalignment between feature learning objectives and clustering goals, as well as low model coupling efficiency. Based on domain-specific requirements, the study proposes future research directions emphasizing lightweight architecture, fine-grained processing, and enhanced robustness. This review offers a clear theoretical framework and methodological guidance for related research, providing a basis for selecting appropriate implementation approaches.

Keywords: Unsupervised Learning; Image Anomaly Detection; Autoencoder; Clustering Analysis; Feature Fusion

1. INTRODUCTION

With the widespread industrial adoption of computer vision technology, image anomaly detection has become a core enabling technology in fields such as industrial defect inspection, medical imaging diagnosis, and security monitoring [1]. In real-world industrial and medical scenarios, anomalous samples are scarce, annotation costs are extremely high, and anomalous patterns remain largely unknown; traditional supervised learning methods struggle to adapt to these open-world conditions. Consequently, unsupervised image anomaly detection techniques that require no prior anomaly labeling have emerged as a major research focus in the field [2–3].

The mainstream technical approaches for unsupervised image anomaly detection fall into two categories: reconstruction-based methods centered on autoencoders and distribution-matching methods focused on clustering. The dual-driven approach combining reconstruction and clustering has yielded numerous improvements in recent years due to its ability to balance feature representation capabilities with detection robustness [4–5]. However, existing reviews predominantly provide comprehensive overviews across the field, while specialized systematic reviews addressing this fusion direction remain scarce, and there is a lack of in-depth analysis of the core bottlenecks in fusion models [6].

To address the aforementioned research gaps, this paper provides a comprehensive review of unsupervised image anomaly detection methods that integrate autoencoders with improved clustering techniques. It systematically traces the technological evolution in this field, categorically evaluates the strengths, weaknesses, and applicable scenarios of various approaches, identifies key technical bottlenecks, and outlines future research directions tailored to domain-specific requirements, thereby offering theoretical insights and methodological references for related studies.

2. REVIEW SEARCH AND SCREENING METHODOLOGY

Retrieval Databases: International databases: IEEE Xplore, Web of Science, ACM Digital Library, arXiv, Chinese databases: CNKI, Wanfang Data, CQVIP.

Search Keywords: English retrieval terms: unsupervised image anomaly detection, autoencoder, deep clustering, feature fusion, industrial defect inspection. Chinese retrieval terms: Unsupervised Image Anomaly Detection, Autoencoder, Deep Clustering, Feature Fusion, Industrial Defect Detection.

Time Range of Literature

International studies were limited to 2016–2025, covering the emergence and evolution of classic hybrid models including deep embedded clustering and memory-augmented autoencoders. Chinese studies were restricted to 2018–2025, consistent with the boom of domestic industrial visual inspection researches.

Inclusion Criteria

Papers focusing on hybrid architectures combining autoencoder and clustering for image anomaly detection.

Peer-reviewed journal articles and CCF-ranked conference papers with theoretical analysis, model improvements and validations on industrial or medical image datasets.

Dissertations equipped with complete algorithm frameworks and ablation experiments.

Exclusion Criteria

Works that merely adopt autoencoder or clustering without cross-module fusion.

Anomaly detection research targeting text, time series, sensor signals rather than visual images.

Pure mathematical derivation without empirical verification on image datasets.

Duplicated publications, low-quality conference abstracts and short briefs without systematic research.

Literature Screening Pipeline

A total of 528 Chinese and English documents were initially retrieved via keyword matching. After removing 121 duplicate records, 297 irrelevant papers were eliminated by title and abstract scanning. The remaining 110 papers were fully reviewed, and 58 core literatures were ultimately selected for systematic analysis in this survey. The standardized retrieval and screening procedure ensures the integrity, objectivity and representativeness of the reviewed content.

3. SIGNIFICANCE OF THE STUDY AND CURRENT RESEARCH STATUS DOMESTICALLY AND INTERNATIONALLY.

3.1 Significance of the Research Topic

3.1.1 Theoretical Significance

This study systematically reviews the developmental trajectory of unsupervised image anomaly detection, autoencoders, clustering techniques, and their integration methods, thereby enhancing the knowledge framework in this field and addressing the lack of systematic coverage in existing reviews regarding the dual-driven approach of reconstruction and clustering.

3.1.2 Practical Significance

This review summarizes the strengths, limitations, and future directions of existing research, providing clear reference materials and conceptual guidance for scholars engaged in related studies, thereby facilitating the practical application of technologies in real-world scenarios such as industrial vision and medical imaging.

3.2 Current Research Status Domestically and Internationally

3.2.1 Research Status of Unsupervised Image Anomaly Detection

Foreign researchers were the first to propose traditional unsupervised detection methods based on statistics, distance, and density, subsequently shifting focus toward deep learning. Domestic research has kept pace with international advancements, concentrating on practical applications such as industrial vision and medical imaging. These approaches can be broadly categorized into two types: traditional methods and deep learning-based methods, with the latter having become the mainstream^[7].

3.2.2 Current Research Status of Abnormality Detection Based on Autoencoders

Abroad, models such as convolutional autoencoders, variational autoencoders, and memory-enhanced autoencoders are widely used, primarily leveraging reconstruction errors for anomaly detection. Domestically, research focuses more on improvement strategies like attention mechanisms and feature enhancement to enhance detail-based anomaly detection

capabilities. However, existing methods often suffer from issues such as blurred reconstruction of anomalous regions and missed small targets^[8].

3.2.3 Current Research Status of Clustering-Based Anomaly Detection

International research focuses on integrating deep clustering with feature learning, whereas domestic studies aim to address issues such as poor robustness of high-dimensional features and noise sensitivity. Clustering can effectively assist in anomaly detection; however, its accuracy is insufficient in high-dimensional image feature spaces, making it inadequate for complex scenarios when used alone^[5].

3.2.4 Current Research Status of Autoencoder and Clustering Fusion Methods

The mainstream approaches domestically and internationally fall into two categories: sequential fusion (encoding features first before clustering) and joint optimization (synchronous training of feature learning and clustering). Most existing fusion methods employ superficial integration, exhibiting common shortcomings such as misalignment between feature learning objectives and clustering goals, as well as low model coupling, which represents a core challenge in current research^[9].

4. BASIC THEORIES OF UNSUPERVISED IMAGE ANOMALY DETECTION

4.1 Definition and Core Challenges

The core logic of unsupervised image anomaly detection is “learning normal patterns and identifying anomalies.” By training on a large number of normal samples, the model learns their data distribution and feature patterns. During the testing phase, any image or image region that deviates from this “normal” pattern is classified as an anomaly.

Core Challenges: Autoencoders are prone to abnormal reconstruction and fail to detect minor or weak texture anomalies. clustering exhibits poor robustness in high-dimensional features and is susceptible to noise interference.

Feature learning and clustering objectives are inconsistent, with low integration between shallow layers; training is unstable, often leading to feature collapse and cluster center drift.

Generalization ability is weak, with insufficient adaptability to complex scenarios and cross-data sets^[10].

4.2 Mechanisms by Which Structural Stability Affects Performance

4.2.1 Impact on Detection Accuracy

Structural instability can induce feature collapse, rendering normal feature representation ineffective; simultaneously, it causes clustering center drift and blurring the boundaries between normal and abnormal regions, thereby directly reducing detection accuracy and AUC.^[11]

4.2.2 Impact on Feature Distinctiveness

Decoder reconstruction distortion causes feature distribution shift; the reconstruction loss and clustering loss exhibit gradient conflicts, leading to amplified feature noise and reduced separability between normal and abnormal samples.

4.2.3 Impact on Training Convergence and Generalization

Structural instability leads to training oscillations that fail to converge and a tendency to fall into local optima; overfitting causes performance degradation in novel scenarios, with structural deterioration occurring during prolonged iterations, resulting in diminished generalization ability and stability.

4.3 Fusion Methods Based on Model Fusion Dimensions

Table 1. Comparison of Representative Autoencoder–Clustering Fusion Methods for Unsupervised Image Anomaly Detection

Model	Fusion Type	Dataset	Quantitative Metrics	Inference Cost (per image)	Main Limitations	Reference
RDAE-Kmeans	Sequential Fusion	MVTec AD	Pixel-AUROC, PRO; Moderate detection accuracy	Fast	Decoupled feature learning and clustering; cannot optimize reconstruction together	[9]
MSCAE-Kmeans	Sequential Fusion	MVTec AD	Pixel-AUROC, PRO; Above-average localization performance	Medium	Multi-scale branches increase computation; weak detection for tiny subtle defects	[12]
AttAE-DBSCAN	Sequential Fusion	MVTec AD	Pixel-AUROC, PRO; Good anti-interference ability	Medium	DBSCAN sensitive to density parameters; attention module brings extra overhead	[13]
IDEC-AD	Joint Optimization	MVTec AD	Pixel-AUROC, PRO; High overall detection accuracy	Relatively slow	Easy to converge to local optimum; training unstable with unbalanced loss weights	[5]
AE-ADBSCAN	Joint Optimization	MVTec AD	Pixel-AUROC, PRO; Stable on single product category	Relatively slow	Self-adaptive density threshold lacks cross-product generalization	[14]
DeepCluster-AD	Joint Optimization	MVTec AD	Pixel-AUROC, PRO; Best localization performance among listed methods	Slow	Severe gradient conflict between reconstruction loss and clustering loss; hyperparameter sensitive	[15]

This table systematically compares six representative autoencoder-clustering fusion anomaly detection methods from fusion architecture, applicable datasets, quantitative evaluation metrics, inference efficiency and inherent drawbacks. Combined with the qualitative theoretical analysis in Section 4, it realizes the organic combination of quantitative indicator dimension sorting and qualitative mechanism induction, which makes the performance differences of various fusion frameworks more intuitive.

5. CLASSIFICATION AND TYPICAL MODELS OF METHODS COMBINING AUTOENCODERS WITH CLUSTERING TECHNIQUES

5.1 Fusion method based on feature extraction dimensions

Multi-scale Feature Fusion Approach: This method employs multi-branched convolutional autoencoders to extract image features at different scales, which are then fused and fed into a clustering algorithm to address the challenge of single-scale features failing to capture both global structures and local details simultaneously. A representative example is the MSCAE-Kmeans model, which generates multi-scale feature maps via pyramid pooling architectures, achieving a 3.2% improvement in pixel-level AUC on the MVTec AD dataset compared to single-scale models^[12].

Attention-enhanced feature-based approach: By incorporating spatial/channel attention mechanisms into autoencoders, this method strengthens feature extraction in critical regions, suppresses background noise interference, and enhances clustering sensitivity to subtle anomalies. For instance, the AttAE-DBSCAN model adaptively adjusts weights across different feature channels via channel attention, effectively addressing false detection issues caused by complex background textures in industrial defect inspection^[13].

5.2 A fusion method based on the dimension of clustering strategies

Enhanced Deep Clustering Method: Building upon classical deep clustering frameworks such as DEC and IDE, this approach optimizes the clustering loss function and allocation update strategy to address the limitations of traditional K-Means—its sensitivity to initial centers and tendency to fall into local optima. The IDEC-AD model introduces reconstruction loss constraints into DEC while refining the update frequency of cluster centers, significantly enhancing the stability of clustering results.^[5]

Density-based clustering enhancement: These approaches improve the distance measurement methods and density threshold adaptation mechanisms of density-based clustering algorithms such as DBSCAN, mitigating the impact of “dimensional disaster” in high-dimensional spaces on clustering performance. For example, the AE-ADBSCAN model introduces a DBSCAN algorithm with adaptive density thresholds that automatically adjusts parameters based on data distributions across different classes without manual intervention.^[14]

5.3 Fusion Methods Based on Model Fusion Dimensions

Series Fusion Type: The autoencoder first performs feature extraction, followed by the application of a clustering algorithm in the feature space. This approach is simple to implement and computationally efficient, but the feature learning and clustering processes are independent, representing a superficial fusion method. It was the most commonly used paradigm in early research, as exemplified by the RDAE-Kmeans model.

Joint Optimization Type: This approach integrates clustering objectives into the training loss of the autoencoder, enabling simultaneous optimization of feature learning and clustering performance. While offering superior performance, it commonly suffers from inconsistencies between feature learning and clustering goals, as well as low model coupling. The representative model is DeepCluster-AD, which achieves mutual reinforcement between the autoencoder and clustering module through alternating training cycles.^[15]

6. CORE BOTTLENECKS AND CHALLENGES

6.1 Scarcity of abnormal samples: The inherent contradiction in the unsupervised assumption

The original intent of unsupervised anomaly detection is to address the scarcity of anomalous samples; however, the autoencoder-clustering fusion approach inherently suffers from a fundamental contradiction in this regard.

Problem Manifestation: Although no anomalous samples are required during the training phase, model hyperparameter tuning, performance validation, and performance evaluation entirely depend on high-quality annotated anomalous samples.^[16] For practical applications such as industrial defect detection, obtaining a sufficient number of annotated samples covering all defect types is extremely costly or even impossible.

Special challenges of fusion methods: The hyperparameters of joint optimization approaches (e.g., clustering loss weights, number of cluster centers) significantly impact performance and lack universal optimal values, requiring individual tuning for each dataset. This process necessitates extensive annotation of anomalous samples for repeated validation, severely contradicting the fundamental principle of unsupervised learning.

6.2 Annotation Challenges: High Costs and Subjectivity of Pixel-Level Annotation

Abnormal detection not only requires determining whether an image is abnormal but also precisely locating the anomalous regions, which imposes extremely high demands on labeling.

Problem Manifestations: Pixel-level defect annotation requires annotators to manually delineate defect contours pixel by pixel^[17], which is time-consuming and labor-intensive, with significant variability among different annotators. For minor or ambiguous defects, the annotation process is highly subjective, introducing inherent errors into both training and evaluation benchmarks; feature collapse and gradient conflicts are the root causes of training instability and clustering failure.

Special challenges of fusion methods: The anomaly scores in clustering methods are calculated based on distances in the feature space, resulting in heatmaps with often blurred boundaries. Existing evaluation metrics (e.g., pixel-level AUC) are highly sensitive to labeled boundaries, leading to performance assessments that fail to accurately reflect the methods’ actual detection capabilities.

6.3 Complex Scene Detection: Exponential Performance Decline in Real-World Environments

While existing methods perform well on standard laboratory datasets, their performance declines sharply in real-world complex scenarios—this represents the primary obstacle to their industrial deployment.

Lighting and Perspective Variations: Autoencoders are highly sensitive to changes in lighting and perspective; normal variations in lighting can significantly increase reconstruction errors, while clustering algorithms may misinterpret these normal variations as anomalies.

Background interference and intra-class diversity: In scenarios with complex backgrounds or high intra-class diversity (e.g., products from different batches or images captured from varying angles), clustering algorithms struggle to learn a unified normal data distribution, leading to frequent false positives and missed detections.

Detection of Minor Defects: During the reconstruction process, autoencoders tend to smooth out minor defect regions ^[18], resulting in negligible reconstruction errors; however, in high-dimensional feature spaces, the characteristic differences between minor defects and normal samples are minimal, making it difficult for clustering algorithms to distinguish them.

6.4 Insufficient application value: Disconnection between academic research and industrial needs

Current research on autoencoder-clustering fusion methods primarily focuses on improving the AUC metric on laboratory datasets, creating a significant disconnect from real-world industrial requirements.

Slow inference speed: Most joint optimization methods exhibit high model complexity and slow inference speeds, failing to meet the real-time requirements of industrial production lines (typically requiring inference time per image less than 20 ms) ^[19].

Poor generalization capability: A model typically can only detect defects within a single product category; when product models are updated, data must be re-collected and the model re-trained, resulting in extremely high deployment costs.

Poor interpretability: The anomaly score generated by clustering methods operates as a black box, offering no explanation for why a specific region is identified as anomalous—making it entirely ineffective for industrial defect diagnosis or process improvement.

Insufficient robustness: Real industrial environments are characterized by significant noise, vibration, and interference; existing methods are highly sensitive to these disturbances, prone to false alarms, and consequently impair production efficiency.

7. SUMMARY AND OUTLOOK

7.1 Research Summary

This paper first systematically reviews the fundamental theories of unsupervised image anomaly detection, clarifying its definition, core challenges, and four major technical paradigms. It provides an in-depth exposition of the basic principles of autoencoders and their variants, traditional clustering, and deep clustering, while demonstrating the theoretical complementarity between autoencoder and clustering fusion approaches. Building upon this foundation, the study establishes a novel classification framework across three key dimensions—feature extraction, clustering strategies, and model fusion—categorizing existing fusion methods into two main types: sequential fusion and joint optimization, thereby clearly outlining the technological evolution in this field ^[20]. Through quantitative and qualitative comparative analyses of six representative methods, the research reveals the intrinsic relationship between “structure and performance”: sequential fusion architectures offer advantages such as simple implementation, stable training, and rapid inference speeds, but exhibit limited performance potential; joint optimization architectures enable mutual reinforcement between feature learning and clustering, achieving higher detection accuracy, though they suffer from training instability and hypersensitivity to hyperparameters. The study further confirms that feature extraction capability is the primary determinant of model performance, with multi-scale feature fusion and attention mechanisms serving as effective means to enhance complex anomaly detection capabilities. Finally, aligning with practical industrial requirements, the paper identifies four major practical challenges currently facing the field.

7.2 Current Challenges and Existing Issues

****Fundamental Theoretical Bottlenecks**:** There exists an inherent conflict between feature learning and clustering objectives. Most existing joint optimization methods employ simple loss-weighting approaches that fail to mathematically

resolve the gradient contradiction between these two goals, leading to model instability during training and clustering collapse issues.^[20] **Data and Annotation Challenges**: The scarcity of anomalous samples remains a fundamental challenge in this field. Although anomalous samples are not required during training, model hyperparameter tuning and performance evaluation heavily depend on high-quality annotated data. Moreover, pixel-level anomaly annotation is both costly and subjective, introducing inherent errors into evaluation benchmarks. **Poor Adaptability to Complex Scenarios**: While existing methods perform well on standard laboratory datasets, their performance declines dramatically in real-world industrial environments due to variations in lighting conditions, perspective shifts, background interference, and significant intra-class diversity among normal samples—particularly lacking in detecting subtle defects and complex texture anomalies. **Barriers to Industrial Implementation**: A significant gap exists between academic research and industrial demands. Most high-performance models suffer from excessive complexity and slow inference speeds, failing to meet real-time requirements in production lines; they exhibit poor generalization capabilities, typically capable of detecting only single-product categories; and their black-box nature of clustering algorithms compromises interpretability for industrial defect diagnosis and process improvement. **Non-uniform Evaluation Standards**: Significant differences in dataset partitioning, evaluation metrics, and hardware environments across studies undermine fairness and objectivity in performance comparisons, making it difficult to accurately measure genuine technological advancements.

7.3 Future Outlook

Theoretical Foundation Deepening: A unified deep integration framework that explores the intrinsic alignment between feature learning and clustering objectives at a mathematical level, designing a joint loss function capable of simultaneously optimizing both reconstruction quality and clustering performance to achieve genuine, organic integration of the two modules and address the instability issues inherent in existing training methods. **New Structural Design**: A lightweight, fine-grained detection architecture tailored to industrial real-time requirements, featuring a lightweight autoencoder-clustering fusion structure enhanced with techniques such as model pruning and quantization to reduce computational complexity. Additionally, refined feature extraction and clustering strategies are implemented to improve the model's ability to detect minute defects at the pixel level.

Application of advanced representation techniques: Self-supervised pre-training and multimodal fusion utilize large-scale unlabeled industrial images for self-supervised pre-training to learn more general and discriminative feature representations, thereby enhancing the model's generalization capability^[21]. Research explores anomaly detection methods that integrate multimodal data from vision, infrared, laser, etc., improving the model's robustness in complex environments^[22]. **Engineering approach development** focuses on unlabeled evaluation and explainability techniques, investigating model assessment and hyperparameter optimization methods without requiring labeled anomalous samples to reduce deployment costs. Explainable tools for cluster-based anomaly detection have been developed to provide intuitive explanations for anomaly causes. Additionally, efforts are underway to establish unified evaluation standards for industrial anomaly detection, facilitating the translation of academic achievements into practical industrial applications.

REFERENCES

- [1] Cui, W., Li, Y., Zhang, H., et al. (2024). A survey of unsupervised image anomaly detection: Methods, evaluations and challenges. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(11), 7294–7315
- [2] Bergmann, P., Jin, X., Sattlegger, D., et al. (2024). MVTec AD 2024: An extended benchmark for unsupervised anomaly detection and localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (pp. 1201–1210)
- [3] Roth, H. R., et al. (2022). Self-supervised learning for medical image analysis. *Nature Machine Intelligence*, 4, 711–719
- [4] Zong, B., Song, Q., Min, M. R., et al. (2018). Deep autoencoding gaussian mixture model for unsupervised anomaly detection. In *International Conference on Learning Representations*
- [5] Xie, J., Girshick, R., & Farhadi, A. (2016). Unsupervised deep embedding for clustering analysis. In *Proceedings of the 33rd International Conference on Machine Learning* (pp. 478–487). PMLR
- [6] Baltrušaitis, T., Ahuja, C., & Morency, L. P. (2019). Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(2), 423–443

- [7] Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys*, 41(3), 1–58
- [8] Salehi, M., & Arbabi, E. (2021). Autoencoder-based anomaly detection: A systematic review. *arXiv:2107.07652*.
- [9] Wang Hao, Li Gang, Zhang Min. (2023). Industrial image anomaly detection based on deep autoencoder and clustering fusion. *Computer Applications*, 43(2), 562–569
- [10] Liu, X., Wang, J., Leng, B., & Zhang, S. (2024). Dual-modeling decouple distillation for unsupervised anomaly detection. In *Proceedings of the 32nd ACM International Conference on Multimedia* (pp. 1128–1137). ACM
- [11] Luo, W., Yao, H., Qiang, Z., Zhang, X., & Zhang, W. (2025). A feature shuffling and restoration strategy for universal unsupervised anomaly detection. *Knowledge-Based Systems*, 314, 113892
- [12] Zhao, H., Shi, J., Qi, X., et al. (2017). Pyramid scene parsing network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2881–2890)
- [13] Woo, S., Park, J., Lee, J. Y., et al. (2018). CBAM: Convolutional block attention module. In *Proceedings of the European Conference on Computer Vision* (pp. 3–19)
- [14] Guo, X., Gao, L., Liu, X., et al. (2017). Improved deep embedded clustering with local structure preservation. *IEEE Transactions on Cybernetics*, 49(2), 553–565
- [15] Caron, M., Bojanowski, P., Joulin, A., et al. (2018). Deep clustering for unsupervised learning of visual features. In *Proceedings of the European Conference on Computer Vision* (pp. 132–149)
- [16] Ruff, L., Vandermeulen, R. A., Goernitz, N., et al. (2018). Deep one-class classification. In *Proceedings of the 35th International Conference on Machine Learning* (pp. 4393–4402). PMLR
- [17] Bergmann, P., Fauser, M., Sattlegger, D., et al. (2021). MVTec AD — A comprehensive real-world dataset for unsupervised anomaly detection. *International Journal of Computer Vision*, 129(4), 963–981
- [18] Bergmann, P., Fauser, M., Sattlegger, D., et al. (2019). Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 4183–4192)
- [19] Roth, K., Pemula, L., Zepeda, J., et al. (2022). Towards total recall in industrial anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 14318–14328)
- [20] Xie, X., Liu, X., Zhu, E., & He, L. (2023). Deep clustering: A comprehensive survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(12), 14058–14077. <https://doi.org/10.1109/TPAMI.2023.3287403>
- [21] Liznerski, P., Ruff, L., Vandermeulen, R. A., & Kloft, M. (2024). PixOOD: Pixel-level out-of-distribution detection. In *European Conference on Computer Vision* (pp. 345–362). Springer
- [22] Azad, R., Aghdam, E. K., Rauland, A., & Adeli, E. (2023). Medical image analysis using unsupervised learning: A survey. *Medical Image Analysis*, 89, 102784. <https://doi.org/10.1016/j.media.2023.102784>