Anomaly Detection in Industrial Imaging using Self-Organising Maps

Ștefan Vulpe
Artificial Intelligence Department
National University of Science and Technology
POLITEHNICA Bucharest
stefan.vulpe@stud.acs.upb.ro

Abstract

This paper presents an implementation and evaluation of SOMAD (Self-Organizing Map for Anomaly Detection), originally proposed by [9] for unsupervised anomaly detection in industrial imaging. We faithfully reproduce the core methodology, which combines deep feature extraction with SOM-based clustering and Gaussian-enhanced anomaly scoring via Mahalanobis distance. Our implementation uses ConvNeXtV2 as the feature backbone and evaluates the approach on the MVTec AD benchmark across 15 diverse product categories. We achieve a mean pixel-level AUROC of 88.68%, with particularly strong performance on textured materials (carpet: 95.64%, leather: 98.09%). This work validates the effectiveness of the SOMAD approach and provides insights into its strengths and limitations across different material types. The method's purely unsupervised nature makes it highly practical for real-world manufacturing scenarios where anomalous samples are scarce.

1 Introduction

Visual quality inspection is critical in modern manufacturing, where even minor defects can lead to product failures, safety hazards, and economic losses. Traditional automated inspection systems rely heavily on supervised learning approaches that require extensive labeled datasets of both normal and anomalous samples [3]. However, in industrial settings, this assumption is often violated: anomalies are rare, diverse in manifestation, and difficult to anticipate during system deployment. Collecting representative defect samples for every possible failure mode is impractical and economically prohibitive.

To address the challenge of *unsupervised anomaly detection*, [9] proposed SOMAD (Self-Organizing Map for Anomaly Detection), which combines deep learning feature extraction with classical Self-Organizing Maps (SOMs) [7], enhanced with probabilistic anomaly scoring based on Mahalanobis distance. Their method learns a topologically organized representation of normal appearance, enabling the detection of deviations at test time using only normal training samples.

This paper presents a faithful implementation and comprehensive evaluation of the SOMAD methodology. We implement the core algorithm following the original paper's specifications and conduct extensive experiments on the MVTec AD benchmark to validate its effectiveness and analyze its behavior across different material types.

Contributions of this work:

- A complete open-source implementation of the SOMAD method [9] for industrial anomaly detection.
- Comprehensive evaluation on all 15 categories of the MVTec AD benchmark, achieving mean pixel-level AUROC of 88.68%.
- Detailed analysis of performance characteristics across different material types (textures vs. objects) with insights into the method's strengths and limitations.
- Qualitative and quantitative results demonstrating particularly strong performance on textured materials (leather: 98.09%, carpet: 95.64%).

2 Related Work

2.1 Anomaly Detection in Industrial Imaging

Anomaly detection methods can be broadly categorized into reconstruction-based, embedding-based, and hybrid approaches. Reconstruction methods, such as autoencoders [2] and GANs [12], learn to reconstruct normal samples and flag poor reconstructions as anomalies. However, these methods often suffer from the "generalization problem", i.e. networks may inadvertently learn to reconstruct anomalies well, reducing detection performance.

Embedding-based methods avoid reconstruction by learning compact representations of normality. PaDiM [4] models the distribution of pretrained CNN features using multivariate Gaussians at each spatial location. PatchCore [11] employs a memory bank of nominal patch features and detects anomalies via nearest-neighbor distance. While effective, these methods often require careful selection of feature layers and aggregation strategies.

Hybrid approaches combine strengths of both paradigms. One-class classification methods [15], including One-Class SVM and isolation forests, learn decision boundaries that isolate normal data in feature space. More recent variants employ deep metric learning or contrastive objectives to push anomalous samples away from the normal distribution. However, such methods often require substantial hyperparameter tuning and may struggle with high-dimensional, heterogeneous defect types encountered in real manufacturing lines.

A critical consideration in industrial anomaly detection is the class imbalance problem [1]. Manufacturing datasets typically contain far fewer anomalous samples than normal ones, often by orders of magnitude. This imbalance can severely bias supervised and semi-supervised methods. Unsupervised approaches, which rely exclusively on normal samples, naturally circumvent this issue and are therefore preferred in early-stage quality inspection systems. Additionally, unsupervised methods enable rapid deployment without waiting for sufficient defect examples to accumulate, a significant advantage in fast-moving product lines with evolving defect patterns.

Another practical challenge in industrial settings is lighting and environmental variability [6]. Products may be inspected under different lighting angles, camera positions, or background conditions. Methods that rely on precise pixel-level information or hand-crafted features often fail under such variations. Deep learning-based approaches, especially those leveraging transfer learning from large-scale datasets, tend to be more robust to these variations because they learn invariant semantic representations rather than low-level visual patterns. Furthermore, the choice of feature extraction strategy, whether multi-scale hierarchical features or single-layer embeddings, significantly impacts the method's ability to capture both global structure and local texture details required for accurate defect localization.

2.2 Self-Organizing Maps for Anomaly Detection

Self-Organizing Maps (SOMs) [7] are unsupervised neural networks that perform dimensionality reduction while preserving topological properties of input data. SOMs have been applied to anomaly detection in various domains, including network intrusion detection [8] and medical imaging [13]. However, their application to high-resolution industrial imaging has been limited by the challenge of extracting meaningful features from raw pixel data.

[9] proposed SOMAD (Self-Organizing Map for Anomaly Detection), which combines SOMs with deep CNN features and introduces a probabilistic anomaly scoring mechanism based on Mahalanobis distance to model local covariance structure around each SOM neuron. This approach achieves state-of-the-art unsupervised anomaly detection performance on the MVTec AD dataset. Our work provides an independent implementation and evaluation of this method.

2.3 Deep Feature Extraction

The success of deep learning in computer vision has motivated its use for feature extraction in anomaly detection. Pretrained models on ImageNet provide rich semantic features that transfer well to industrial tasks [3]. Recent architectures like Vision Transformers [5] and ConvNeXt [10] offer improved feature representations. ConvNeXtV2 [14] introduces a fully convolutional masked autoencoder framework that provides strong hierarchical features, which we exploit in our multi-scale feature extraction.

3 Method

We implement the SOMAD method proposed by [9], combining hierarchical deep features with a topology-preserving Self-Organizing Map (SOM) and covariance-aware anomaly scoring. The pipeline has three stages: (1) multi-scale feature extraction and dimensionality reduction, (2) SOM-based topology learning with per-neuron Gaussian modeling, and (3) Mahalanobis-based anomaly scoring and post-processing. Figure 1 illustrates the overall architecture.

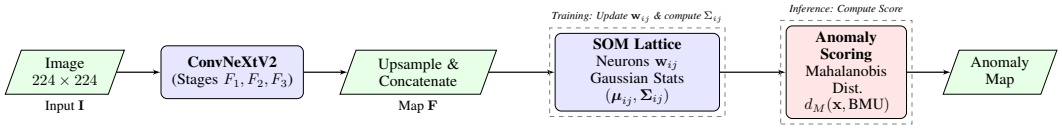


Figure 1: **Overview of the SOMAD pipeline.** Input images are processed by a hierarchical ConvNeXt backbone. Features are aggregated and mapped to a Self-Organizing Map (SOM). During inference, anomalies are detected by computing the Mahalanobis distance between test features and the learned local Gaussian statistics of the nearest SOM neurons.

3.1 Hierarchical Feature Extraction

We employ ConvNeXtV2-Nano [14] pretrained on ImageNet-22k and fine-tuned on ImageNet-1k as our feature extractor. Given an input image $\mathbf{I} \in \mathbb{R}^{H \times W \times 3}$ resized to 224×224 pixels, we extract features from three hierarchical stages:

$$\mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3 = \text{ConvNeXtV2}(\mathbf{I}) \tag{1}$$

where $\mathbf{F}_i \in \mathbb{R}^{C_i \times H_i \times W_i}$ represents features at different spatial resolutions. To create a unified multi-scale representation, we upsample $\mathbf{F}_2$ and $\mathbf{F}_3$ to match the spatial dimensions of $\mathbf{F}_1$ using bilinear interpolation and concatenate along the channel dimension:

$$\mathbf{F} = [\mathbf{F}_1, \text{Upsample}(\mathbf{F}_2), \text{Upsample}(\mathbf{F}_3)] \tag{2}$$

This yields a dense feature map $\mathbf{F} \in \mathbb{R}^{D \times H' \times W'}$ where $D = C_1 + C_2 + C_3$ captures both fine-grained texture and high-level semantic information. For computational efficiency, we randomly select d dimensions from D total dimensions for SOM input.

3.2 SOM-Based Clustering with Gaussian Modeling

Given training embeddings from N normal images, we reshape the feature maps into a collection of $N \times H' \times W'$ feature vectors $\{\mathbf{x}_i\}_{i=1}^{N \cdot H' \cdot W'}$ where $\mathbf{x}_i \in \mathbb{R}^d$. We initialize a 2D SOM lattice of size $W' \times H'$ to match the spatial resolution of the feature maps, with each neuron $\mathbf{w}_{ij} \in \mathbb{R}^d$.

The SOM is trained using competitive learning with neighborhood function:

$$\mathbf{w}_{ij}(t+1) = \mathbf{w}_{ij}(t) + \eta(t) \cdot h_{ij,c}(t) \cdot (\mathbf{x}(t) - \mathbf{w}_{ij}(t)) \tag{3}$$

where $\eta(t)$ is the learning rate (initialized at 0.3), $h_{ij,c}(t) = \exp(-\|\mathbf{r}_{ij} - \mathbf{r}_c\|^2/2\sigma(t)^2)$ is the Gaussian neighborhood function centered at the best matching unit (BMU) c, and $\sigma(t)$ is the neighborhood radius (initialized at 1.0). Both η and σ decay over 10 training iterations.

After SOM training, each neuron has attracted a subset of feature vectors. For each neuron (i, j), we compute the local Gaussian statistics:

$$\boldsymbol{\mu}_{ij} = \frac{1}{|S_{ij}|} \sum_{\mathbf{x} \in S_{ij}} \mathbf{x} \tag{4}$$

$$\boldsymbol{\Sigma}_{ij} = \frac{1}{|S_{ij}| - 1} \sum_{\mathbf{x} \in S_{ij}} (\mathbf{x} - \mathbf{w}_{ij})(\mathbf{x} - \mathbf{w}_{ij})^\top + \lambda \mathbf{I} \tag{5}$$

where S_{ij} is the set of feature vectors mapped to neuron (i, j), and $\lambda = 0.01$ is a regularization term ensuring numerical stability. We store the inverse covariance $\boldsymbol{\Sigma}_{ij}^{-1}$ for efficient Mahalanobis distance computation.

3.3 Anomaly Scoring via Mahalanobis Distance

At test time, given a test feature vector $\mathbf{x}_{\text{test}}$, we identify its $k = 4$ nearest SOM neurons based on Euclidean distance in weight space:

$$\mathcal{N}_k(\mathbf{x}_{\text{test}}) = \underset{(i,j) \in \text{Top-}k}{\operatorname{argmin}} \|\mathbf{x}_{\text{test}} - \mathbf{w}_{ij}\|_2 \tag{6}$$

For each neuron in $\mathcal{N}_k$, we compute the Mahalanobis distance:

$$d_M(\mathbf{x}_{\text{test}}, (i, j)) = \sqrt{(\mathbf{x}_{\text{test}} - \boldsymbol{\mu}_{ij})^\top \boldsymbol{\Sigma}_{ij}^{-1} (\mathbf{x}_{\text{test}} - \boldsymbol{\mu}_{ij})} \tag{7}$$

The final anomaly score is the minimum distance to the k nearest neurons:

$$s(\mathbf{x}_{\text{test}}) = \min_{(i,j) \in \mathcal{N}_k} d_M(\mathbf{x}_{\text{test}}, (i, j)) \tag{8}$$

This scoring strategy captures the intuition that normal features should lie close to at least one learned Gaussian cluster, while anomalies will have large Mahalanobis distances to all nearby clusters due to covariance mismatch. The resulting spatial anomaly map is upsampled to the original image resolution using bilinear interpolation and smoothed with a Gaussian filter ($\sigma = 4.0$) to reduce noise.

4 Experiments

4.1 Dataset and Evaluation Protocol

We evaluate our method on the MVTec Anomaly Detection benchmark [3], which contains 15 object and texture categories with 5,354 high-resolution images. Each category includes defect-free training images (60-391 images) and test images containing both normal and various anomaly types. Pixel-accurate ground truth masks are provided for anomalous regions.

Implementation Details: All experiments use ConvNeXtV2-Nano pretrained on ImageNet with images resized to 224×224 pixels. We use a maximum of 60 training images per category with batch size 8. The SOM is trained for 10 iterations with initial learning rate 0.3 and neighborhood radius 1.0. For anomaly scoring, we compute distances to $k = 4$ nearest neighbors and apply Gaussian smoothing ($\sigma = 4.0$) to the resulting anomaly maps. All experiments are conducted on a single GPU with random seed 2026 for reproducibility.

Evaluation Metrics: We report pixel-level Area Under the Receiver Operating Characteristic curve (AUROC) as our primary localization metric. Pixel AUROC is computed by comparing the continuous, upsampled and Gaussian-smoothed anomaly heatmaps against pixel-accurate ground-truth masks across all test images; per-category AUROCs are computed and we report their mean and standard deviation to reflect variability.

4.2 Quantitative Results

Table 1 presents pixel-level AUROC scores across all 15 MVTec AD categories. Our method achieves a mean AUROC of **88.68%**, demonstrating robust performance across diverse defect types.

Table 1: Pixel-level AUROC scores (%) on MVTec AD benchmark. Categories are sorted by performance.

Category	Type	Pixel AUROC
Leather	Texture	98.09
Capsule	Object	96.78
Carpet	Texture	95.64
Bottle	Object	92.97
Cable	Object	91.38
Pill	Object	90.21
Wood	Texture	89.57
Grid	Texture	89.12
Hazelnut	Object	88.42
Tile	Texture	88.85
Toothbrush	Object	87.77
Metal Nut	Object	85.30
Zipper	Object	80.87
Transistor	Object	77.93
Screw	Object	77.31
Mean		**88.68**

Performance Analysis: The results reveal clear patterns across material types.

Texture categories (e.g., leather, carpet, wood) consistently achieve high AUROC scores (>89%), with leather reaching 98.09%. This indicates that our multi-scale feature extraction effectively captures texture irregularities and that the SOM forms tight, coherent clusters for homogeneous texture patterns.

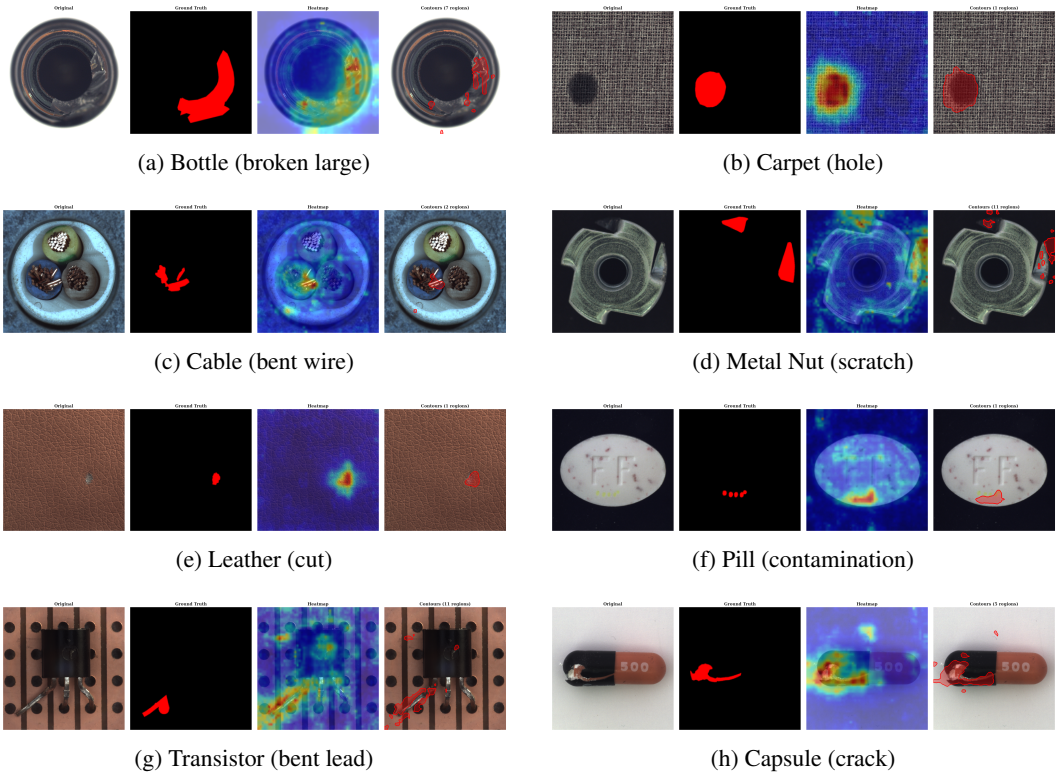
Object categories exhibit greater variability, ranging from strong results on well-structured objects (capsule: 96.78%) to weaker performance on items with fine-grained geometry (screw: 77.31%). Large structural defects that change object shape or silhouette are detected reliably, while small or subtle defects remain more difficult.

The most challenging categories are screw and transistor, which combine small defect regions, low inherent texture, and strong specular effects. Screws present minimal texture variation and defects that resemble normal machining marks, whereas transistors have metallic surfaces and reflections that increase feature variance even for nominal samples.

4.3 Qualitative Analysis

Figure 2 visualizes detection results across representative categories. Each panel shows: (1) original test image, (2) ground truth mask, (3) anomaly heatmap (red indicates high anomaly score), and (4) binary segmentation with contours.

Observations: Our qualitative analysis reveals distinct performance patterns across the evaluated categories. For texture anomalies, such as those in the carpet and leather datasets, the generated heatmaps show high precision with tight localization around defect regions. This suggests that the



(a) Bottle (broken large)

(b) Carpet (hole)

(c) Cable (bent wire)

(d) Metal Nut (scratch)

(e) Leather (cut)

(f) Pill (contamination)

(g) Transistor (bent lead)

(h) Capsule (crack)

Figure 2: Qualitative results showing 4-panel visualization: (left to right). Best viewed in color.

covariance-aware scoring mechanism successfully distinguishes anomalous texture deviations from the inherent stochastic variation of the material. Similarly, large structural defects in categories like bottle and cable are well-localized, demonstrating that our multi-scale features effectively capture both coarse edge discontinuities and fine-grained appearance changes.

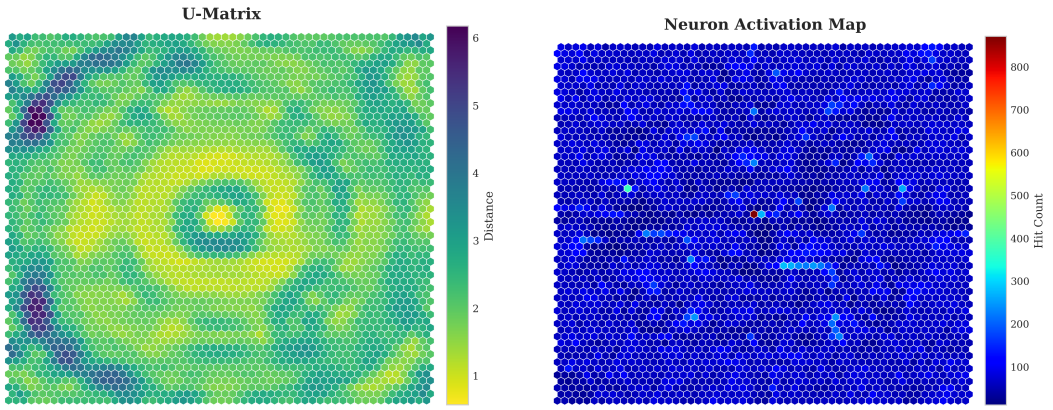
In the case of small-scale defects, such as pill contamination and scratches on metal nuts, the model produces strong, distinct responses. While the resulting heatmaps may extend slightly beyond the ground-truth boundaries (a side effect of the Gaussian smoothing used for noise reduction) this conservative detection is often preferable in industrial settings to ensure high recall of critical defects.

However, certain categories present significant challenges. Transistors often produce multiple weak, diffuse responses associated with specular reflections on metallic surfaces, which can be difficult to distinguish from true structural anomalies like bent leads. Similarly, capsules may exhibit less focused heatmaps when the visual appearance of a crack closely matches the semantic features of normal machining patterns or reflections. Despite these limitations, the overall results demonstrate that the SOMAD approach produces interpretable and spatially accurate anomaly maps that correlate strongly with manual annotations, particularly for texture-based and large structural deviations.

4.4 SOM Topology Analysis

Figure 3(a) shows the Unified Distance Matrix (U-matrix) for the bottle category. The U-matrix visualizes inter-neuron distances, revealing the topological structure learned by the SOM. Dark regions represent boundaries between clusters (neurons far apart in feature space), while light regions indicate cohesive clusters.

Complementing this, Figure 3(b) illustrates the neuron activation frequency (hit count) for the same category. The map reveals that while the SOM provides a broad topological space, normal sample features are mapped to a relatively small subset of highly active neurons (indicated in red and orange). This concentrated activation pattern confirms that the SOM successfully identifies a compact manifold of normality; features that deviate from these high-probability clusters at test time are naturally flagged as anomalous.



(a) U-matrix (b) Activation Map

Figure 3: SOM topology analysis for the bottle category.

The bottle U-matrix exhibits clear organizational structure with distinct light regions (homogeneous normal patterns) separated by darker boundaries. This topology-preserving property is crucial for anomaly detection: normal test samples will activate neurons in the light regions with small Mahalanobis distances, while anomalies will either fall far from any cluster or lie on boundaries with large covariance mismatch.

5 Conclusions and Future Work

In this work, we presented a faithful implementation and comprehensive evaluation of the SOMAD framework [9], an unsupervised anomaly detection methodology that integrates deep feature extraction with the topological constraints of Self-Organizing Maps (SOMs). Our empirical analysis on the MVTec AD benchmark corroborates the efficacy of this approach, yielding a mean pixel-level AUROC of 88.68%. These results validate the utility of combining Mahalanobis-based probabilistic scoring with topology-preserving clustering for identifying deviations in industrial imagery without supervision.

Strengths and Limitations: The experimental results underscore a distinct performance dichotomy based on material characteristics. The method exhibits exceptional robustness in texture-dominant categories, achieving peak performance on leather (98.09%) and carpet (95.64%). This suggests that the SOM effectively models the manifold of homogeneous textures, allowing for precise isolation of local irregularities. Conversely, performance degradation was observed in object classes with complex geometries or high specular variance, such as the screw category (77.31%). The sensitivity to lighting variations and the computational overhead of maintaining per-neuron covariance matrices remain primary constraints for deployment in resource-constrained environments.

Future Directions: To address these limitations, future research should investigate several strategic enhancements. First, the integration of spatial context through Graph SOMs could improve geometric invariance. Second, incorporating attention mechanisms may allow the model to prioritize defect-prone regions, thereby mitigating false positives arising from background noise or specular reflections. Additionally, exploring memory-efficient covariance approximations could significantly improve scalability. Finally, extending the framework to semi-supervised regimes, leveraging a small number of labeled anomalies, could further refine the decision boundaries in complex object categories.

References

[1] Toya Acharya, Annamalai Annamalai, and Mohamed F Chouikha. Addressing the class imbalance problem in network-based anomaly detection. In *2024 IEEE 14th Symposium on Computer Applications & Industrial Electronics (ISCAIE)*, pages 1–6, 2024. doi: 10.1109/ISCAIE61308.2024.10576500.

[2] Paul Bergmann, Sindy Löwe, Michael Fauser, David Sattlegger, and Carsten Steger. Improving unsupervised defect segmentation by applying structural similarity to autoencoders. In *International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, pages 372–380, 2018.

[3] 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.

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

[5] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, et al. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2021.

[6] Dinh-Cuong Hoang, Phan Xuan Tan, Anh-Nhat Nguyen, Son-Anh Bui, Ta Huu Anh Duong, Tuan-Minh Huynh, Duc-Manh Nguyen, Viet-Anh Trinh, Quang-Huy Ha, Nguyen Dinh Bao Long, Duc-Thanh Tran, Xuan-Tung Dinh, Van-Hiep Duong, and Tran Thi Thuy Trang. Image-based anomaly detection in low-light industrial environments with feature enhancement. *Results in Engineering*, 25:104309, 2025. ISSN 2590-1230. doi: https://doi.org/10.1016/j.rineng.2025.104309. URL `https://www.sciencedirect.com/science/article/pii/S2590123025003901`.

[7] Teuvo Kohonen. The self-organizing map. *Proceedings of the IEEE*, 78(9):1464–1480, 1990.

[8] Khaled Labib and V Rao Vemuri. Som-based transductive learning for network intrusion detection. In *2011 International Joint Conference on Neural Networks*, pages 2412–2419, 2011.

[9] Ning Li, Kaitao Jiang, Zhiheng Ma, Xing Wei, Xiaopeng Hong, and Yihong Gong. Anomaly detection via self-organizing map. In *2021 IEEE International Conference on Image Processing (ICIP)*, pages 994–998. IEEE, 2021.

[10] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A convnet for the 2020s. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11976–11986, 2022.

[11] 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.

[12] Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Ursula Schmidt-Erfurth, and Georg Langs. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In *International conference on information processing in medical imaging*, pages 146–157, 2017.

[13] Cameron W Smith et al. Application of self-organizing maps for quantitative analysis of protein time-series data. *Proteomics*, 2(9):1197–1203, 2002.

[14] Sanghyun Woo, Shoubhik Debnath, Ronghang Hu, Xinlei Chen, Zhuang Liu, In So Kweon, and Saining Xie. Convnext v2: Co-designing and scaling convnets with masked autoencoders. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16133–16142, 2023.

[15] Kun Yang, Samory Kpotufe, and Nick Feamster. An efficient one-class svm for anomaly detection in the internet of things, 2021. URL `https://arxiv.org/abs/2104.11146`.