International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 08 Issue 08 August 2025
DOI: 10.47191/ijcsrr/V8-i8-20, Impact Factor: 8.048
IJCSRR @ 2025
www.ijcsrr.org

Using a Small Dataset in Learning for Object Anomaly Detection

Lappoon R. Tang
Wildfaces Technology Ltd.

ABSTRACT: This paper introduces a novel approach to object anomaly detection using an ordered ensemble method with Procrustes distance, emphasizing data efficiency with small training sets. Unlike traditional deep autoencoder methods, which rely on pixel-wise reconstruction and require large datasets (e.g., 200 images per category in the MVTec AD dataset), our method leverages Procrustes distance to measure structural disparities between object feature shapes after translation, rotation, and scaling. By computing minimum Procrustes distances from a small set of 30 normal images per category, we derive robust thresholds for classifying objects as normal or anomalous. Evaluated on five MVTec AD categories (metal nut, cable, bottle, hazelnut, transistor), our approach achieves superior accuracy (e.g., 100% for metal nut and cable) compared to deep autoencoders, demonstrating robustness across rigid and deformable objects. This data-efficient method offers significant advantages for industrial inspection, where acquiring large defect-free datasets is challenging.

KEYWORDS: Object anomaly detection, Procrustes distance, Small sample size

1. INTRODUCTION

Procrustes analysis is a form of statistical shape analysis used to analyze the distribution of a set of shapes (Gower & Dijksterhuis, 2004). In order to compare the shapes of two or more objects, the objects must first be optimally "superimposed". Procrustes superimposition (PS) is performed by optimally moving, rotating and evenly scaling the objects. In other words, both the placement in space and the size of the objects are freely adjusted. The goal is to achieve similar placement and size by minimizing a measure of shape difference, the *Procrustes distance* between the objects. The Procrustes distance between two shapes (each defined as a set of points in a 2D plane) is a Euclidean-like distance that computes the square root of the sum of squared differences between each pair of points.

In most applications of data analysis, especially in cluster analysis, the Euclidean distance is the choice of measure of disparity between two vectors. However, if we treat a feature vector as a "shape" (i.e. a set of 2D points), we can measure the disparity between two feature vectors by using the Procrustes distance. The intuition is that two feature vectors of two normal objects (i.e. objects without defects) would look more similar after adjustment due to translation, rotation, and scaling (i.e. after Procrustes superimposition) than simply their Euclidean disparity. However, since the feature vector of an abnormal object should be intrinsically different than that of a normal object due to the presence of defects in the object, the Procrustes distance between a normal object and an abnormal object would tend to be greater. Therefore, the Procrustes distance would be more reflective of the "true" distance between a pair of objects (normal vs. abnormal). Our assumption is that the use of Procrustes distance would facilitate the derivation of a threshold for anomaly detection systems – i.e. the threshold derived from Procrustes distance will be more robust than that derived from Euclidean distance.

Object anomaly detection concerns the detection of objects with unusual defects (e.g. a developing crack) given normal objects without unusual defects (Kamoona *et. al.*, 2021). Object anomaly detection is arguably an important problem because of its potential application in predictive maintenance. More precisely, images of normal objects are used in learning a model for detecting the presence of defects in images of abnormal objects (e.g. a defective machine part). A currently trending approach to anomaly detection is using deep auto-encoders (Borghesi *et. al.*, 2019). In such an approach, feature vectors for normal objects are computed and used in learning a deep auto-encoder. The intuition is that the feature vector of an image of a defective object will be poorly re-constructed by the deep auto-encoder (i.e. reconstruction errors will be high). Thus, one can detect that an image is containing an abnormal object if the reconstruction error for the image is greater than a certain threshold. An alternative approach to object anomaly detection is using a Gaussian Mixture model (Bahrololum & Khaleghi, 2008). In this case, the feature vectors for normal objects are used for learning a Gaussian Mixture model (GMM). Outliers are feature vectors that are not close to any of the clusters – they don't belong

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 08 Issue 08 August 2025
DOI: 10.47191/ijcsrr/V8-i8-20, Impact Factor: 8.048
IJCSRR @ 2025
www.ijcsrr.org

to any of the clusters. Yet, other approaches to anomaly detection involve the use of one-class support vector machine (Zineb *et. al.*, 2012), or local outlier factor (Breunig *et. al.*, 2000).

Our approach to object anomaly detection involves computing the Procrustes distances of a selected set of normal objects to a dataset of normal objects and deriving a threshold according to the average of all minimum Procrustes distances of the selected normal objects with respect to the dataset of normal objects (i.e. the minimum Procrustes distance of each selected normal object with respect to the dataset of normal objects is computed). Given the derived threshold, an object is abnormal if its minimum Procrustes distance with respect to the dataset of objects is above the threshold.

2. RELATED WORK

A related work on object anomaly detection treats visual defect detection as a problem of anomaly detection. It provides an evaluation of different point pattern feature detectors and descriptors for defect detection within the random finite set (RFS) framework (Mahle, 2007), focusing on the manufacturing industry. More precisely, point pattern features are like those in scale-invariant feature transform (SIFT) features (Lowe, 1999), and each measured set of point features is treated as a RFS. Due to the lack of access to the defected samples, unsupervised anomaly detection is a preferred option for defect detection. In this approach, only the normal samples (i.e., defect-free samples) are used in the training phase. Similarly, the RFS-based defect detection only uses the normal samples during training to maximize RFS set density. The work emphasizes the significance of automated visual inspection in manufacturing, highlighting the limitations of manual inspection and the potential impact on product quality and production costs. Deep learning methods, such as convolutional neural networks (CNNs), for defect detection and classification in various applications can also be used.

Various handcrafted and deep learning-based point pattern feature extraction methods, such as SIFT, Harris-Laplace point detector, LF-net, D2-net, and r2d2 are covered. Experimental results, based on the MVTec AD dataset (Bergmann *et. al.*, 2021), demonstrate the performance of the proposed RFS-based defect detection framework using different feature extraction methods, showcasing the effectiveness of the RFS framework in defect detection. The results show that RFS-based defect detection with SIFT demonstrates promising performance, outperforming or performing similarly to state-of-the-art methods.

3. ALGORITHMIC APPROACH

3.1 Procrustes Analysis

We just consider objects made up from a finite number k of points in n dimensions. Often, these points are selected on the continuous surface of complex objects, such as a human bone, and in this case they are called landmark points. The shape of an object can be considered as a member of an equivalence class formed by removing the translational, rotational and uniform scaling components.

Translation

For example, translational components can be removed from an object by translating the object so that the mean of all the object's points (i.e. its centroid) lies at the origin.

Mathematically: take k points in two dimensions, for example, $\big((x_1, y_1), (x_2, y_2), \ldots, (x_k, y_k)\big)$.

The mean of these points is $(\bar{x}, \bar{y})$ where

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_k}{k}$$

$$\bar{y} = \frac{y_1 + y_2 + \cdots + y_k}{k}$$

Now, translate these points so that their mean is translated to the origin $(x_1 - \bar{x},\ y_1 - \bar{y}, \ldots)$.

Uniform scaling

Likewise, the scale component can be removed by scaling the object so that the root mean square distance (RMSD) from the points to the translated origin is 1. This RMSD is a statistical measure of the object's scale or size:

$$s = \sqrt{\frac{(x_1 - \bar{x})^2 + (y_1 - \bar{y})^2 + \cdots}{k}}$$

The scale becomes 1 when the point coordinates are divided by the object's initial scale:

$$\left(\frac{x_1 - \bar{x}}{s}, \frac{y_1 - \bar{y}}{s}\right).$$

Rotation

Removing the rotational component is more complex, as a standard reference orientation is not always available. Consider two objects composed of the same number of points with scale and translation removed. Let the points of these be $\big((x_1, y_1), \ldots\big), \big((w_1, z_1), \ldots\big)$. One of these objects can be used to provide a reference orientation. Fix the reference object and rotate the other around the origin, until you find an optimum angle of rotation θ such that the sum of the squared distances (SSD) between the corresponding points is minimized.

A rotation by angle θ gives

$$(u_1, v_1) = (\cos\theta\, w_1 - \sin\theta z_1, \sin\theta\, w_1 + \cos\theta\, z_1).$$

Shape comparison

The difference between the shape of two objects can be evaluated only after "superimposing" the two objects by translating, scaling and optimally rotating them as explained above. The square root of the above mentioned SSD between corresponding points can be used as a statistical measure of this difference in shape:

$d = \sqrt{(u_1 - x_1)^2 + (v_1 - y_1)^2 + \cdots}.$

This measure is often called **Procrustes distance**.

3.2 Learning a Model for Anomaly Detection

To learn a model for object anomaly detection, we use a set of objects without unusual defects (i.e., normal objects) as a test set and compute the Procrustes distances between these objects and the training set of normal objects. More precisely, for each object o_i in the normal *test* set, we find the minimum distance d_i of the Procrustes distances between the object o_i and the objects in the normal training set. Then, we collect the set of minimum Procrustes distances for all the objects in the test set $\{d_i\}$.

Given the set of minimum Procrustes distances $\{d_i\}$, we compute the threshold α as follows:

$$\alpha = average(\{d_i\}).$$

Given the threshold α, we decide whether a novel object is normal versus abnormal as follows. We compute the Procrustes distances of the novel object with respect to *all* the normal objects in the training set. If the minimum Procrustes distance is greater than the threshold α, the novel object is classified as abnormal (i.e., with unusual defects). Otherwise, it is classified as normal.

We employ an *ordered ensemble* approach for image anomaly detection, integrating multiple feature extraction methods to classify images as normal or anomalous. It extracts feature shapes from training images using three methods (e.g., texture, grayscale color histogram, Discrete Cosine Transform (DCT)), computing Procrustes disparity thresholds (90th percentile) against "good" reference images to establish normality criteria.

The classification process follows a hierarchical structure: 1) *Shape Similarity Check*: For a test image, compute the Procrustes distance for each feature set (texture, histogram, DCT) against the training set. If any distance exceeds the corresponding 90th percentile threshold, the image is flagged as anomalous, 2) *Feature Sum Verification*: If all shape checks pass, compute the feature sums for the test image and verify if they fall within the acceptable ranges (10th to 90th percentiles). If any sum is outside its range, the image is classified as anomalous; otherwise, it is normal.

This sequential, multi-layered approach enhances robustness. The function tests "good" images (expected normal) and general test images (potentially anomalous), logging classifications as correct or incorrect based on expected outcomes, ensuring reliable anomaly detection through diverse feature integration.

4. EXPERIMENTAL EVALUATION

To experimentally evaluate the framework of anomaly detection that uses an ordered ensemble employing Procrustes distance measure, we use the MVTec anomaly detection dataset (MVTec AD). The dataset has more than ten different categories of objects such as zipper, transistor, metal nut, toothbrush, and etc (Bergmann *et. al.*, 2021). Each category of objects has around 200 normal images for training and 30 to 40 abnormal images for testing. *However, our approach for machine learning involves only a small dataset for learning a model but still achieves relatively good performance. More precisely, for each object category, we only used*

30 normal images for training. Please notice that most current approaches reported in the literature use all training data for learning a model. For simplicity, we selected five categories of objects for experimentation, including 1) metal nut, 2) cable, 3) bottle, 4) hazelnut, and 5) transistor. Images are resized to a resolution of 700x700 before anomaly detection is carried out.

For experimental comparison, we compared our approach with the deep autoencoders approach in (Bergmann *et. al.*, 2021), which employs a convolutional autoencoder, consisting of an encoder and decoder with convolutional layers. More precisely, the encoder compresses input images (high-resolution images from MVTec AD) into a low-dimensional latent representation. The decoder reconstructs the input from this latent space, aiming to produce an output as close as possible to the defect-free training images.

Table 1. Experimental Comparison between Ordered Ensemble vs. Deep Autoencoders

Object\Approach	**Ordered Ensemble (ours)**	**Deep Autoencoders**
Metal Nut	**100%**	89%
Cable	**100%**	82%
Bottle	**96.66%**	94%
Hazelnut	**90%**	87%
Transistor	**85%**	80%

For each test image (e.g. an image of a metal nut object with a defect), we apply the learned ordered ensemble to the image to classify whether the image is “normal” (i.e. without defect) or “abnormal” (i.e. with a defect). Accuracy is computed as the percentage of correctly classified images for an object category (e.g. transistor).

The experimental results (outlined in Table 1) demonstrate the superior performance of the ordered ensemble approach compared to the deep autoencoders method across all five object categories. For the metal nut category, the ordered ensemble achieved a perfect accuracy of 100%, significantly outperforming the deep autoencoders’ 89%. This suggests that the Procrustes distance measure effectively captures the structural consistency of rigid objects like metal nuts, even with limited training data. Similarly, for the cable category, the ordered ensemble attained 100% accuracy, compared to 82% for deep autoencoders, highlighting its robustness in handling deformable objects where global context is critical. In the bottle category, the ordered ensemble scored 96.66%, slightly better than the 94% of deep autoencoders, reflecting its ability to detect subtle defects like contamination on rigid structures. For hazelnut, the ordered ensemble achieved 90% accuracy, surpassing the 87% of deep autoencoders, despite challenges posed by natural variations and random rotations. Finally, for the transistor category, the ordered ensemble recorded 85% accuracy, improving upon the 80% of deep autoencoders, though both methods struggled with reflective surfaces and structural defects like missing legs. The ordered ensemble’s consistent outperformance can be attributed to its use of Procrustes distance to align and compare image features, which better captures anomalies in diverse object types compared to the reconstruction-based approach of autoencoders, which often produces blurry outputs for complex cases. Notably, our method’s ability to achieve these results with only 30 training images per category—compared to the full dataset used by deep autoencoders—underscores its efficiency and generalization capability. These findings validate the ordered ensemble approach as a robust and data-efficient solution for unsupervised anomaly detection on the MVTec AD dataset, particularly for challenging industrial inspection scenarios.

5. DISCUSSION

The deep autoencoder approach for anomaly detection on the MVTec AD dataset, as described in (Bergmann *et al.*, 2021), relies on a convolutional autoencoder that learns to reconstruct defect-free images by encoding them into a low-dimensional latent space and decoding them back to the original resolution. Anomalies are detected by measuring reconstruction errors, with high errors indicating defects. This method leverages unsupervised learning, requiring only normal images for training, and is effective for rigid objects like bottles, where it achieves high accuracy (94%). However, its major characteristics include dependence on pixel-wise reconstruction fidelity, which can falter when modeling complex or deformable objects (e.g., cables, 82% accuracy) due to blurry reconstructions. The autoencoder struggles with global context, often producing high anomaly scores in non-defective regions, such as reflective surfaces in transistors (80% accuracy) or natural variations in hazelnuts (87% accuracy).

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 08 Issue 08 August 2025
DOI: 10.47191/ijcsrr/V8-i8-20, Impact Factor: 8.048
IJCSRR @ 2025
www.ijcsrr.org

The strengths of the deep autoencoder approach lie in its simplicity and ability to model normality without requiring labeled anomalies, making it suitable for industrial settings with abundant defect-free data. However, its weaknesses are pronounced: it requires a large number of training samples (typically 200 per category in MVTec AD) to capture the full variability of normal images, leading to high sample complexity. Suboptimal performance occurs with fewer samples, as the model fails to generalize across variations like rotations or textures. Additionally, the approach struggles with fine-grained defects (e.g., small scratches on metal nuts) and deformable objects, as the encoder-decoder architecture prioritizes low-level features over structural consistency, limiting its robustness.

In contrast, the ordered ensemble approach, utilizing Procrustes distance, demonstrates superior performance across all tested categories (e.g., 100% for metal nut and cable, 85% for transistor). Its strength lies in its ability to align and compare image features using a distance measure that captures structural consistency, making it effective for both rigid (e.g., bottle, 96.66%) and deformable objects (e.g., cable). The approach's low sample complexity—achieving high accuracy with only 30 training images per category—stems from its focus on geometric and structural alignment rather than pixel-wise reconstruction. By modeling relative feature relationships, it generalizes well with limited data, reducing overfitting risks. However, its weaknesses include potential sensitivity to extreme variations in object appearance (e.g., transistors' reflective surfaces) and computational complexity in aligning features for large datasets.

The ordered ensemble's outperformance over deep autoencoders is intrinsically tied to its ability to prioritize structural and geometric relationships over pixel-level fidelity. While autoencoders rely on reconstructing entire images, often missing subtle or global anomalies, the Procrustes distance aligns features to detect deviations in object structure, excelling in categories with complex defects (e.g., missing cable parts). Its data efficiency further enhances its applicability in real-world industrial inspection, where acquiring large training sets is costly. These results highlight the ordered ensemble's robustness and adaptability, making it a promising approach for unsupervised anomaly detection on the MVTec AD dataset.

6. CONCLUSION

The ordered ensemble approach, utilizing Procrustes distance, outperforms the deep autoencoder method for unsupervised anomaly detection on the MVTec AD dataset due to its fundamental focus on structural and geometric relationships rather than pixel-wise reconstruction. This structural alignment enables the ordered ensemble to excel across diverse object categories, achieving perfect accuracy (100%) for metal nuts and cables, and high accuracy for bottles (96.66%), hazelnuts (90%), and transistors (85%), compared to the deep autoencoder's lower scores (e.g., 89%, 82%, 94%, 87%, and 80%, respectively). While autoencoders rely on reconstructing entire images, they often produce blurry outputs, missing subtle defects (e.g., small scratches) or misclassifying non-defective regions due to poor global context modeling, particularly in deformable objects like cables or reflective surfaces like transistors. In contrast, the Procrustes distance measure aligns image features to detect deviations in object structure, effectively capturing both local and global anomalies, such as missing cable parts or structural inconsistencies in metal nuts.

The ordered ensemble's low sample complexity, requiring only 30 training images per category compared to the 200 typically used by autoencoders, stems from its emphasis on relative feature relationships rather than exhaustive pixel-level modeling. By focusing on geometric alignment, it generalizes effectively from minimal data, avoiding overfitting and capturing essential structural patterns with fewer examples. This data efficiency is critical for industrial inspection, where acquiring large defect-free datasets is costly. Despite its computational complexity and sensitivity to extreme appearance variations, the ordered ensemble's robustness and adaptability make it a superior choice for real-world applications. These results underscore its potential as a data-efficient, high-performance solution for unsupervised anomaly detection, offering significant advantages over reconstruction-based methods like deep autoencoders in challenging scenarios.

Author contributions statement: All of the work was solely done by the author himself.

Disclosure statement: The author(s) declare that there are no competing interests associated with the work presented in this article.

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. The authors declare that no sponsors were involved in the preparation of this article.

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 08 Issue 08 August 2025
DOI: 10.47191/ijcsrr/V8-i8-20, Impact Factor: 8.048
IJCSRR @ 2025
www.ijcsrr.org

REFERENCES

1. Bahrololum, M., & Khaleghi, M. (2008). Anomaly Intrusion Detection System Using Gaussian Mixture Model. In Proceedings of the Third International Conference on Convergence and Hybrid Information Technology.
2. Bergmann, P., Batzner, K., Fauser, M., Sattlegger, D., & Steger, C. (2021). The MVTec Anomaly Detection Dataset: A Comprehensive Real-World Dataset for Unsupervised Anomaly Detection. *International Journal of Computer Vision*, 129(4), 1040-1059.
3. Bergmann, P., Fauser, M., Sattlegger, D., Steger, C. (2021). Unsupervised Anomaly Detection and Localization with Deep Autoencoders on the MVTec AD Dataset. *International Journal of Computer Vision*, 129(4), 1018-1039.
4. Borghesi, A., Bartolini, A., Lombardi, M., Milano, M., & Benini, L. (2019). Anomaly Detection Using Autoencoders in High Performance Computing Systems. In Proceedings of the Thirty-First AAAI Conference on Innovative Applications of Artificial Intelligence.
5. Breunig, M. M., Kriegel, H.-P., Ng, R. T., & Sander, J. (2000). LOF: Identifying Density-based Local Outliers. In Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data.
6. Dusmanu, M., Rocco, I., Pajdla, T., Pollefeys, M., Sivic, J., Torii, A., & Sattler, T. (2019). D2-net: A Trainable CNN for Joint Description and Detection of Local Features. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.
7. Gower, J.C., & Dijksterhuis, G.B. (2004). Procrustes Problems. Oxford University Press.
8. Kamoona, A. M., Gostar, A. K., Bab-Hadiashar, A., & Hoseinnezhad, R. (2021). Evaluation of Point Pattern Features for Anomaly Detection of Defect within Random Finite Set Framework. Accessed from: *arXiv:2102.01882.*
9. Lowe, D. G. (1999). Object Recognition from Local Scale-invariant Features. In Proceedings of the International Conference on Computer Vision.
10. Mahle, R. P. (2007). *Statistical multisource-multitarget information fusion*. Artech House Norwood, MA.
11. Zineb, N., Honeine, P., & Richard, C. (2012). On Simple One-class Classification Methods. In Proceedings of the IEEE International Symposium on Information Theory.

Cite this Article: Tang, L.R. (2025). Using a Small Dataset in Learning for Object Anomaly Detection. International Journal of Current Science Research and Review, 8(8), pp. 4128-4133. DOI: https://doi.org/10.47191/ijcsrr/V8-i8-20