

# Unsupervised Learning for Industrial Defect Detection: A Case Study on Shearographic Data <sup>★</sup>

Jessica Plassmann<sup>1,2</sup> , Nicolas Schuler<sup>1,3</sup> , Georg von Freymann<sup>2,4</sup> , and Michael Schuth<sup>1</sup> 

<sup>1</sup> Trier, University of Applied Science, 54293 Trier, Germany  
{plassmaj, schulern, schuth}@hochschule-trier.de

<sup>2</sup> Department of Physics and research center OPTIMAS, RPTU University  
Kaiserslautern-Landau, 67663 Kaiserslautern, Germany,  
georg.freymann@rptu.de

<sup>3</sup> University of Luxembourg, 4365 Esch-Belval Esch-sur-Alzette, Luxembourg

<sup>4</sup> Fraunhofer Institute for Industrial Mathematics ITWM, 67663 Kaiserslautern,  
Germany

**Abstract.** Shearography is a non-destructive testing method for detecting subsurface defects, offering high sensitivity and full-field inspection capabilities. However, its industrial adoption remains limited due to the need for expert interpretation. To reduce reliance on labeled data and manual evaluation, this study explores unsupervised learning methods for automated anomaly detection in shearographic images. Three architectures are evaluated: a fully connected autoencoder, a convolutional autoencoder, and a student-teacher feature matching model. All models are trained solely on defect-free data. A controlled dataset was developed using a custom specimen with reproducible defect patterns, enabling systematic acquisition of shearographic measurements under both ideal and realistic deformation conditions. Two training subsets were defined: one containing only undistorted, defect-free samples, and one additionally including globally deformed, yet defect-free, data. The latter simulates practical inspection conditions by incorporating deformation-induced fringe patterns that may obscure localized anomalies. The models are evaluated in terms of binary classification and, for the student-teacher model, spatial defect localization. Results show that the student-teacher approach achieves superior classification robustness and enables precise localization. Compared to the autoencoder-based models, it demonstrates improved separability of feature representations, as visualized through t-SNE embeddings. Additionally, a YOLOv8 model trained on labeled defect data serves as a reference to benchmark localization quality. This study underscores the potential of unsupervised deep learning for scalable, label-efficient shearographic inspection in industrial environments.

---

<sup>★</sup> This preprint has not undergone peer review or any post-submission improvements or corrections. The Version of Record of this contribution is published in: Artificial Intelligence XLII. Lecture Notes in Computer Science. Springer Cham, 2026. DOI: [https://doi.org/10.1007/978-3-032-11442-6\\_22](https://doi.org/10.1007/978-3-032-11442-6_22)

**Keywords:** Shearography · Non-destructive testing (NDT) · Anomaly detection · Unsupervised learning.

## 1 Introduction

Electronic Speckle-Pattern Shearing Interferometry (ESPSI), or shearography, is a well-established optical method for non-destructive testing (NDT), particularly effective for detecting subsurface defects in composites and other industrial components [27,26]. Despite its high sensitivity to small surface deformations and rapid inspection capabilities, shearography is underused in serial applications due to the lack of fully automated evaluation processes [18].

A critical challenge in automating shearographic inspection is the reliance on supervised learning approaches, such as convolutional neural networks (CNNs) and object detection models like YOLO[20,11]. These achieve their high performance primarily through large, extensive labeled datasets, such as ImageNet [5]. Creating such datasets demands extensive manual effort and cost, and issues such as data scarcity, privacy, and labeling constraints further hinder dataset generation. Moreover, many companies possess vast unlabeled data unusable by supervised methods. [23]

Studies like Li (2022) [13] have explored the use of synthetic data for generating datasets with sufficient and reproducible defects. However, generating synthetic data with realistic defect characteristics and variations remains challenging. Overly diverse synthetic datasets may lead to overfitting, particularly in large-scale applications [13]. This highlights the need to tailor datasets to specific components and testing conditions to ensure reliable defect detection.

Given these challenges, unsupervised learning methods offer a promising alternative. Instead of relying on labeled defect data, these approaches learn the characteristics of defect-free components and identify deviations from this baseline [16]. This method significantly reduces dependency on manually curated defect annotations. This paper explores the application of unsupervised learning for automating shearographic defect detection, with a particular focus on autoencoders [3] and student-teacher models [31]. These approaches are evaluated in terms of their effectiveness, robustness, and practical feasibility [30,22]. The comparison aims to determine the most suitable method for enabling reliable and automated shearographic inspection in industrial settings.

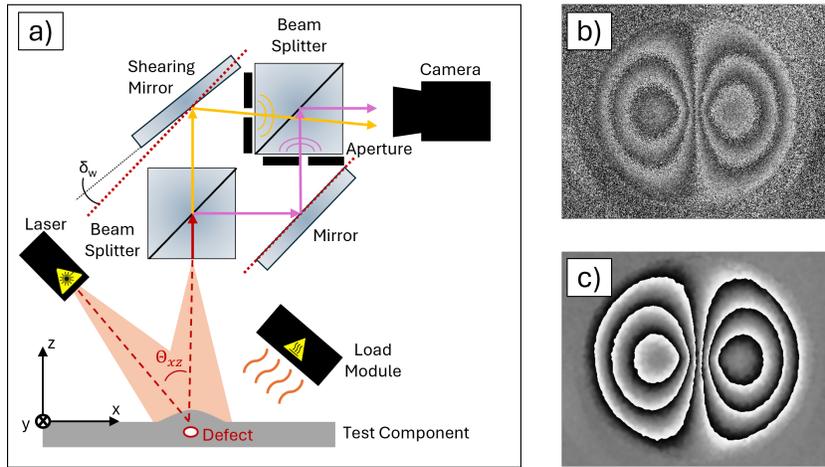
## 2 Domain Background: Shearography

Shearography is a non-destructive optical testing method for detecting subsurface defects in materials. The method measures deformation changes between two loading states with submicrometer sensitivity, enabling detection of minute surface displacement gradients. The component surface, which appears optically rough at the laser wavelength scale, is illuminated with coherent light, producing a granular speckle pattern due to random interference. This speckle pattern

carries phase information and propagates through an interferometric setup to the camera. [24]

In the setup, the backscattered wavefront is laterally shifted (sheared) by tilting one beam path, causing interference between neighboring surface points. Each image pixel thus results from interference between spatially offset points on the object surface, enabling measurement of deformation gradients. [24, p. 554]

Phase information necessary for quantitative evaluation is extracted using phase-shifting techniques. Temporal phase shifting relies on sequential image acquisition per loading state and thus requires static loading during the capture sequence. In contrast, spatial phase shifting enables single-frame evaluation, allowing measurements under dynamic loading conditions [18]. Figure 1a illustrates the setup based on a Mach-Zehnder interferometer using spatial phase shifting and thermal excitation.

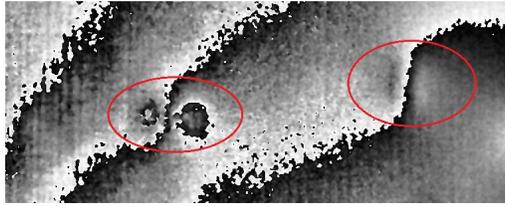


**Fig. 1.** a) Shearography setup using a Mach-Zehnder interferometer with spatial phase shifting and thermal excitation, own illustration based on [18, p. 97]. b) Unfiltered and c) corresponding filtered phase image of a central out-of-plane deformation at a circular defect. [24, p. 555]

Measurements are based on comparing two states of the same object: a reference state and a state under external load. This load, typically thermal or mechanical, induces a deformation gradient that is captured as a differential phase map, known as a shearogram [24, pp. 314–327]. These maps visualize minute changes in surface deformation, enabling the detection of subsurface flaws through local anomalies in the phase pattern. However, environmental noise and speckle decorrelation can affect phase quality. Spatial or frequency-domain filtering is therefore applied to enhance relevant structures and suppress artifacts (Figure 1b–c) [24]. Although the technique originated in the 1970s, only recent

advances in coherent light sources and computational processing have made its integration into industrial workflows viable [18].

Despite its potential, shearography remains underutilized due to the difficulty of interpreting the resulting phase images. Phase fringes indicate deformation, but not necessarily defects. Accurate assessment requires prior knowledge of the expected deformation response under load. Subtle or localized anomalies caused by internal flaws can be obscured by various effects such as global deformation, measurement noise, or speckle decorrelation. As illustrated in Figure 2, even clearly defective regions may remain difficult to identify visually [24].



**Fig. 2.** Filtered phase image with defect locations marked in red. Detection is challenging, especially when defect geometry is unknown and global deformations cause overlapping phase fringes. [24, p. 556]

As in many machine learning contexts, data availability is a key challenge. Industrial datasets are typically limited and imbalanced, with significantly more conforming parts than defective ones. Furthermore, defective samples with known flaws are rarely accessible due to confidentiality concerns. Consequently, this work explores unsupervised approaches trained exclusively on defect-free data to enable anomaly detection without labeled defects.

### 3 Methodology

The following section presents the architectural foundations of the implemented models, each designed to detect anomalies in shearographic image data. Two distinct approaches are introduced: an autoencoder-based reconstruction method and a Student-Teacher Feature Matching framework.

#### 3.1 Architecture

**Autoencoder** A widely used model for unsupervised anomaly detection is the autoencoder. It reconstructs input data by first encoding it into a lower-dimensional latent space and then decoding it back to its original form. For anomaly detection, the autoencoder is trained exclusively on defect-free samples, allowing it to learn the underlying distribution of normal data. When applied

to defective samples, reconstruction fails due to deviations from the learned distribution, leading to a higher reconstruction error, which serves as an anomaly indicator. [3]

Autoencoders offer flexibility in their architecture. A simple feed-forward autoencoder reduces input dimensions through multiple layers, while convolutional autoencoders use CNN-based encoders for feature extraction. The latter is particularly effective for image data, as it preserves spatial features. [3]

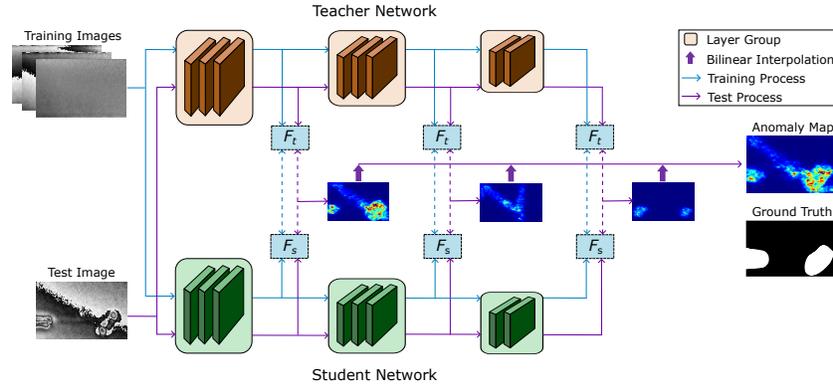
Two autoencoder architectures were implemented. The first is a feed-forward autoencoder without convolutional layers, referred to as AE. The input image is resized to  $96 \times 50$  pixels from an original resolution of  $1920 \times 1050$  pixels and then flattened into a vector of size 4800. The encoder consists of four fully connected layers with output dimensions of 256, 128, 64, and 10, respectively. The latent representation of the data is thus a vector in  $\mathbf{R}^{10}$ . Each layer is followed by a dropout layer and a ReLU activation function to improve generalization and prevent overfitting.

The second implementation, referred to as ConvAE, is based on [9,8], an autoencoder originally designed for time-series anomaly detection, adapted for two-dimensional data. The CNN-based encoder consists of four convolutional layers with 96, 128, and 256 feature maps using  $3 \times 3$  convolutional kernels. Each layer is followed by batch normalization and a LeakyReLU activation function. The decoder mirrors the encoder structure to reconstruct the input data from the latent representation.

**Student Teacher Model** In this work, the Student-Teacher Feature Pyramid Matching (STFPM) approach [29,30] was selected due to its high computational efficiency, widespread adoption, and competitive accuracy [14,7]. The model has been successfully applied and adapted across various domains, making it a robust choice for detecting defects in shearographic inspections [25,28,33,32].

The STFPM architecture, shown in Figure 3, consists of two neural networks with identical structures, referred to as the Student and the Teacher. Unlike siamese networks, these architectures do not share trainable weights. The Teacher network is a pre-trained model that remains frozen during training, serving as a reference by providing feature representations of defect-free inputs. The Student network is trained exclusively on non-defective samples and learns to approximate the feature representations extracted by the Teacher. The training objective is to minimize the discrepancy between the feature outputs of both networks, with the loss function computed as the cumulative difference between the Student’s and Teacher’s activations across multiple layers. [31]

Once trained, the model produces a heatmap that visualizes the pixel-wise differences between the feature representations of the Student and the Teacher. Since the model is trained solely on non-defective data, it learns to reproduce normal feature distributions with minimal error. When applied to defective samples, however, the reconstruction error increases, particularly in the regions where anomalies are present. This characteristic enables STFPM not only to classify defective and non-defective samples but also to localize defects at a pixel level.



**Fig. 3.** Architecture of the Student-Teacher Feature Pyramid Matching model, adapted from [31].

For the implementation in this work, the official STFPM model [29] was adopted and integrated into the existing class structures. Both the Student and Teacher networks use a ResNet18 architecture, with the pre-trained weights for the Teacher being sourced from the TorchVision model zoo [1]. In addition to the core implementation of STFPM, the model was embedded into multiple classes designed to facilitate visualization and analysis of the generated heatmaps. These include peak detection mechanisms to assist in the localization of potential defects within the input images.

Two variants of the student-teacher model were implemented: STFPM (Peaks) and STFPM (Means). Both share identical model weights; the only difference lies in the scoring strategy. STFPM (Peaks) uses the highest anomaly score from the predicted heatmap, whereas STFPM (Means) computes the average score over the input region.

### 3.2 Dataset Preparation and training Setup

A custom specimen was designed to simulate industrial conditions while ensuring reproducibility. A 500 mm × 50 mm acrylic glass strip was coated with a 1.2 mm thick, 45 mm wide rubber layer. Thin inserts of heat-shielding foil were embedded between the materials to simulate subsurface defects. These air-filled inclusions remain invisible from the surface but are verifiable from the transparent backside, providing labeled ground truth.

The specimen’s dimensions match the field of view of the shearography optics and the travel range of a UR5e robot arm, allowing automated scanning in 10 mm increments. Each frame contains at most two defects, enabling isolated evaluation. Data were acquired under two conditions: minimal global deformation through rigid fixation, and varied thermal excitation to induce diverse global deformation patterns. This allowed for capturing both localized defect responses and representative background variation.

The complete dataset comprises 10,498 shearography recordings acquired under controlled laboratory conditions. These recordings are divided into three categories. A total of 4,311 images contain deliberately introduced subsurface defects. Another 2,537 recordings represent defect-free specimens measured under fixed boundary conditions, which suppress global deformation. The remaining 3,650 images were acquired from defect-free specimens without fixation, leading to varying global deformation patterns and corresponding fringe structures.

The defective samples were manually annotated by a domain expert. Each annotated instance consists of a set of four temporally offset images capturing the same region under identical loading conditions, accompanied by bounding boxes marking the observed defect indications. The defect-free recordings serve two complementary roles: the measurements under fixed boundary conditions represent an idealized baseline free of global deformation, while the unfixed specimens introduce realistic variability in the form of fringe patterns caused by global deformation, thereby increasing the complexity of distinguishing local defect-related features from non-defect-induced surface behavior.

Based on the available dataset, two subsets were defined to address different evaluation scenarios in anomaly detection. The first, referred to as Subset A, includes only defect-free images acquired under fixed conditions, in which no global deformation patterns are present. This configuration yields a controlled baseline, consisting of 2,020 images for training, 254 for validation, and 254 for testing. Subset B includes all 6,187 defect-free recordings, incorporating both undistorted and globally deformed samples. In this case, the training set comprises 4,582 images, with 584 used for validation and 697 for testing. In both variants, defective samples are added to the validation and test sets to evaluate binary classification performance under realistic conditions. Specifically, 714 annotated defect samples are included in the test set and 391 in the validation set. These defective samples are excluded from anomaly detection training and reserved for model evaluation.

To quantitatively assess the localization performance of the proposed STFPM-based anomaly detection, a dedicated subset of the annotated defect data was used for training a YOLOv8 detection model [12]. The same 714 and 391 samples that were already included in the anomaly detection test and validation sets were reused to ensure consistency across evaluations. The remaining 3,523 annotated defective samples were reserved exclusively for training the YOLOv8 model, ensuring an unbiased benchmark.

Additionally, to evaluate the separability of the learned feature representations and the underlying structure of the data, t-distributed stochastic neighbor embedding (t-SNE) [2] was applied.

All experiments were conducted on a Windows 11 desktop PC equipped with a 13th Gen Intel(R) Core(TM) i7-13700 processor, 128 GB RAM, and an NVIDIA RTX 4000 GPU with 16 GB VRAM. All models were implemented in Python 3.11 using the PyTorch framework [17], with CUDA [15] support for accelerated training. Code and dataset will be made available at [19].

### 3.3 Model Comparison and Evaluation Metrics

The evaluation of the proposed method addresses two key aspects: binary classification between defect-free and defective samples, and the quantitative assessment of localization performance.

Classification performance was assessed using the receiver operating characteristic (ROC) curve [10] and the precision-recall (PR) curve [4]. The area under the ROC curve (AUC) quantifies the trade-off between true positive rate and false positive rate, with an AUC of 0.5 indicating random performance for binary classification. The PR curve illustrates the relationship between precision and recall, with the average precision (AP) summarizing precision across all recall levels.

To quantify localization accuracy, the STFPM and YOLO models were evaluated using common metrics from object detection [21]. These include the intersection over union (IoU), the mean average precision (mAP), mAP at fixed IoU thresholds of 0.5 and 0.75 (mAP@0.5, mAP@0.75), and the mean average recall (mAR) at one and ten predictions per image (mAR@1, mAR@10). The IoU between a predicted region  $B$  and a ground truth region  $A$  is defined as the ratio of their intersection over their union. In cases with multiple predictions per ground truth region, only the prediction with the highest IoU is considered.

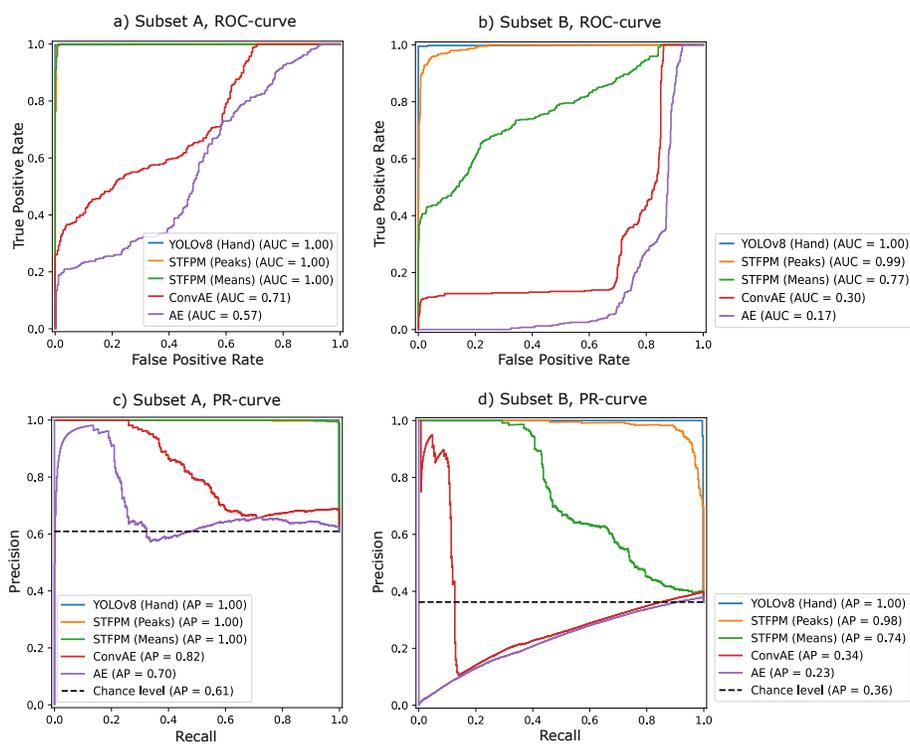
Since only a single defect class is evaluated, the mean average precision equals the average precision ( $mAP = AP$ ). The fixed IoU thresholds define the criteria for correct matches, while the recall metrics quantify the fraction of correctly detected regions under limited prediction counts. All localization metrics were computed using the mean average precision implementation from the TorchMetrics library [6]. As the Student-Teacher model does not produce conventional classification scores, the maximum anomaly score within each predicted region was used as a confidence proxy for ranking and evaluation.

## 4 Results

### 4.1 Binary Classification

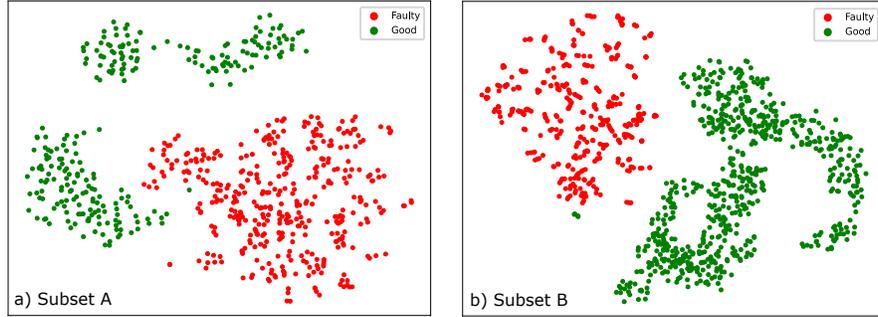
The corresponding ROC and PR curves for both subsets are shown in Figure 4, illustrating the performance differences across all models. For Subset A, the STFPM-based methods achieve perfect performance, with AUC and AP values of 1.0, matching the supervised YOLOv8 baseline. The ConvAE model achieves an AUC of 0.71 and an AP of 0.82, while the simple Autoencoder performs notably worse (AUC of 0.57, AP of 0.70). Given a chance level of 0.61 for AP in this subset, only STFPM and ConvAE yield results above random performance.

In Subset B, the supervised YOLOv8 again achieves perfect scores (AUC and AP of 1.0). STFPM Peaks maintains strong performance with an AUC of 0.99 and an AP of 0.98, while STFPM Means shows reduced but still competitive results, with an AUC of 0.77 and an AP of 0.74. In contrast, ConvAE and AE exhibit significantly lower performance in this more challenging setting, achieving AUC values of 0.30 and 0.17, and AP values of 0.34 and 0.23, respectively. The chance level for AP in Subset B is 0.36.



**Fig. 4.** ROC and PR curves for the binary classification task on Subset A and Subset B. Each plot compares unsupervised models (STFPM Peaks/Means, ConvAE, AE) against the supervised YOLOv8 baseline. (a) ROC Subset A, (b) ROC Subset B, (c) PR Subset A, (d) PR Subset B.

## 4.2 Dataset



**Fig. 5.** t-SNE embeddings illustrating the feature separability of defect-free and defective samples for (a) Subset A, trained exclusively on undistorted data, and (b) Subset B, which includes globally deformed defect-free samples.

The t-SNE visualization in Figure 5 illustrates the learned feature space for both training subsets. Defect-free samples are shown in green, while defective samples are marked in red. In both subsets, a clear separation between classes is observed, indicating that the learned representations effectively capture discriminative characteristics.

In Subset A, which consists solely of undistorted defect-free samples, this separation is particularly pronounced, with only one defect-free sample located near the defective cluster. Furthermore, the defect-free samples in Subset A exhibit a distinct internal structure, forming at least two well-separated clusters. Subset B, containing globally deformed but non-defective samples, also demonstrates good overall class separability. However, two defect-free samples lie closer to the defective cluster boundary. Unlike Subset A, the defect-free samples in Subset B form a more contiguous cluster but contain noticeable low-density regions, suggesting non-uniform distribution within the normal class. These findings indicate that both subsets enable clear class separation while revealing meaningful intra-class variability within defect-free samples.

## 4.3 Localization

Table 1 presents the localization results for the STFPM method, evaluated on Subset A and Subset B. For this experiment, only the Peaks variant was used. The predicted anomaly maps were smoothed and then binarized using fixed thresholds, which were optimized separately for each subset based on validation data. A threshold of 0.1 was selected for Subset A, and 0.001 for Subset B. The results are compared to those of a supervised YOLOv8 model trained with pixel-wise defect annotations.

**Table 1.** Localization performance comparison of STFPM and YOLOv8.

| Model            | IoU    | mAP    | mAP@50 | mAP@75 | mAR@1  | mAR@10 |
|------------------|--------|--------|--------|--------|--------|--------|
| YOLOv8           | 0.8695 | 0.7435 | 0.9901 | 0.9195 | 0.5817 | 0.7884 |
| STFPM (Subset A) | 0.6965 | 0.3389 | 0.7421 | 0.1727 | 0.3209 | 0.4782 |
| STFPM (Subset B) | 0.7113 | 0.3852 | 0.7832 | 0.2611 | 0.3703 | 0.5127 |

## 5 Discussion

### 5.1 Binary Classification

The results indicate a clear divergence in performance between the STFPM-based methods and the autoencoder-based models across both subsets. For Subset A, representing near-ideal conditions, the STFPM Peaks and Means approaches achieve perfect classification performance, matching the supervised YOLOv8 baseline with AUC and AP scores of 1.0. This demonstrates that the STFPM methods effectively capture the discriminative features necessary to distinguish defective from defect-free samples under these conditions.

The STFPM Means method is conceptually similar to the autoencoder models, as it computes the average reconstruction error over the heatmap, analogous to the aggregation of the autoencoder’s reconstruction error. Nevertheless, STFPM Means consistently achieves higher performance than both the convolutional autoencoder and the simple autoencoder, which exhibit moderate and lower performance, respectively (ConvAE AUC 0.71, AP 0.82; AE AUC 0.57, AP 0.70). Furthermore, STFPM Peaks, which detects localized maximum errors instead of averages, attains the best performance among the unsupervised approaches, underscoring the benefit of capturing localized anomalies over global averages.

This pattern becomes even more pronounced on Subset B, which introduces global deformations and thus greater complexity. While YOLOv8 and STFPM Peaks maintain near perfect scores (AUC and AP of 1.0 and 0.99 / 0.98 respectively), STFPM Means experiences a reduction in performance (AUC 0.77 and AP 0.74) but still notably outperforms both autoencoder models, which perform below or near chance level (ConvAE AUC 0.30 and AP 0.34; AE AUC 0.17 and AP 0.23). These results further confirm the limitations of the autoencoder architectures in this setting.

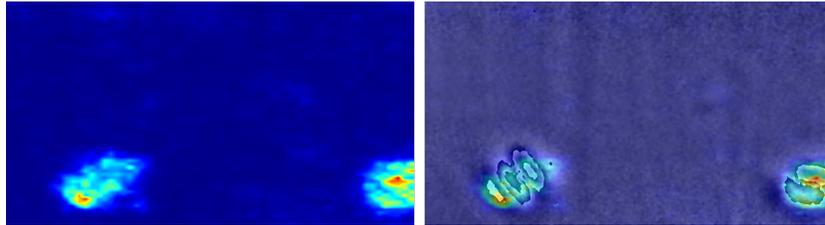
Overall, these findings suggest that the relatively poor performance of the autoencoder models is likely attributable to architectural or training limitations rather than an inherent lack of class separability in the data. This is supported by the t-SNE visualization in Figure 5, which reveals well-defined clusters corresponding to the two classes. The moderate success of the STFPM Means method further indicates that the data contain discriminative features accessible to appropriate models. To fully exploit this intrinsic separability, future work should investigate alternative model architectures or enhanced preprocessing techniques. In contrast, the consistently strong performance of the STFPM

Peaks method underscores its robustness and effectiveness for anomaly detection, even under challenging conditions with significant global deformations.

## 5.2 Localization

Despite being an unsupervised approach, STFPM demonstrates promising localization capabilities across both datasets. Since localization was evaluated exclusively on defective samples, the YOLO model is unaffected by the subset definitions and can be compared consistently across both. The localization performance depends strongly on the chosen thresholds. STFPM tends to generate multiple smaller bounding boxes for larger defect areas, which may be detected as separate anomaly regions instead of one unified defect. This fragmentation naturally affects metrics such as mean Average Precision and mean Average Recall, since these metrics penalize excessive or overlapping detections.

Despite the unsupervised nature of the method and the variability in defect size and appearance, overall detection quality remains high. Figure 6 illustrates this with a representative sample: the left image shows the anomaly heatmap generated by STFPM, while the right image presents the corresponding sample overlaid with the heatmap. The visualization highlights the model’s ability to localize defects effectively, although spatial accuracy may occasionally be limited by thresholding or peak fragmentation. For illustration, Figure 6 shows an example image with a defect, the corresponding anomaly heatmap generated by STFPM, and the detected peak localization. This visualization highlights how STFPM successfully identifies the defect area despite some spatial imprecision.



**Fig. 6.** Exemplaric defect localization using STFPM. Left: predicted anomaly heatmap. Right: same heatmap overlaid on the original shearography image. High anomaly scores (red) correspond to localized defect regions.

## 6 Conclusions

This study demonstrates that STFPM, especially its Peaks variant, is a robust unsupervised approach for defect detection and localization in shearographic inspection of industrial components. It reliably distinguishes between defect-free

and defective samples, even under complex conditions with global deformations, closely matching the performance of supervised YOLOv8 models that require extensive labeled data. In contrast, autoencoder-based models show limited effectiveness in this domain, likely due to architectural constraints rather than inherent data challenges, as confirmed by the clear class separability observed in feature space visualizations.

Importantly, STFPM’s unsupervised nature addresses a key industrial challenge: enabling automated evaluation of shearographic data without costly, time-consuming defect annotations. While localization precision can be further improved, the promising results underline STFPM’s potential for scalable, automated non-destructive testing in real-world industrial applications.

Future work should focus on refining localization accuracy and adapting models to varying defect types and materials to support broader deployment of automated shearographic inspection.

**Acknowledgments.** This research was funded by the Ministry of Science and Health Rhineland-Palatinate as part of the Young Researchers Fund of the Research Initiative (Funding Period 2025) and by the Federal Ministry for Economic Affairs and Climate Protection (BMWK) through the ‘Central Innovation Program for SMEs (ZIM)’, funding code KK5060010SY3.

**Disclosure of Interests.** The authors declare no conflict of interest.

## References

1. Pretrained models for pytorch. <https://modelzoo.co/model/pretrained-modelspytorch> (2021)
2. Cai, T.T., Ma, R.: Theoretical foundations of t-sne for visualizing high-dimensional clustered data (2022), <https://arxiv.org/abs/2105.07536>
3. Chen, S., Guo, W.: Auto-encoders in deep learning—a review with new perspectives. *Mathematics: Mathematical Methods and Applications for Artificial Intelligence and Computer Vision* **11**(18) (April 2023). <https://doi.org/10.3390/math11081777>
4. Cook, J., Ramadas, V.: When to consult precision-recall curves. *The Stata Journal* **20**(1), 131–148 (2020). <https://doi.org/10.1177/1536867X20909693>
5. Deng, J., et al.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition (2009). <https://doi.org/10.1109/CVPR.2009.5206848>
6. Detlefsen, N.S., Borovec, J., Schock, J., Jha, A.H., Koker, T., Liello, L.D., Stancel, D., Quan, C., Grechkin, M., Falcon, W.: Torchmetrics - measuring reproducibility in pytorch. *Journal of Open Source Software* **7**(70), 4101 (2022). <https://doi.org/10.21105/joss.04101>
7. Dutta, P., et al.: Exploring deep learning-based unsupervised image anomaly detection and localization methods for industrial quality assurance. In: 2024 1st International Conference on Cognitive, Green and Ubiquitous Computing (IC-CGU) (2024). <https://doi.org/10.1109/IC-CGU58078.2024.10530676>

8. Gong, D., et al.: Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection. *Computer Science: Computer Vision and Pattern Recognition* (August 2019). <https://doi.org/10.48550/arXiv.1904.02639>, <https://arxiv.org/abs/1904.02639>
9. Gong, D., et al.: STFPM - Official Implementation. <https://github.com/donggong1/memae-anomaly-detection> (2019)
10. Hoo, Z.H., et al.: What is an roc curve? *Emergency Medicine Journal* **34**(6), 357–359 (2017). <https://doi.org/10.1136/emermed-2017-206735>
11. Hussain, M.: YOLO-v1 to YOLO-v8, the Rise of YOLO and Its Complementary Nature toward Digital Manufacturing and Industrial Defect Detection. *Machines: State-of-the-Art in Digital Manufacturing Systems* **11**(7), 677 (2023). <https://doi.org/10.3390/machines11070677>
12. Jocher, G., et al.: Ultralytics yolov8 (2023), <https://github.com/ultralytics/ultralytics>
13. Li, W., et al.: Simulation dataset preparation and hybrid training for deep learning in defect detection using digital shearography. *Applied Sciences* **12**(14) (July 2022). <https://doi.org/10.3390/app12146931>
14. Mylläri, J., Nurminen, J.K.: Discrepancy scaling for fast unsupervised anomaly localization. In: 2023 IEEE 47th Annual Computers, Software, and Applications Conference (COMPSAC) (2023). <https://doi.org/10.1109/COMPSAC57700.2023.00042>
15. NVIDIA, Vingelmann, P., Fitzek, F.H.: Cuda, release: 10.2.89 (2020), <https://developer.nvidia.com/cuda-toolkit>
16. Pang, G., et al.: Deep learning for anomaly detection: A review. *ACM Computing Surveys (CSUR)* **54**(2), 1 – 38 (March 2021). <https://doi.org/10.1145/3439950>
17. Paszke, A., et al.: Pytorch: An imperative style, high-performance deep learning library. In: *Advances in Neural Information Processing Systems*. vol. 32 (2019)
18. Petry, C.M.: Weiterentwicklung der Shearografie mit räumlichem Phasenschieben als zerstörungsfreies Prüfverfahren für die automatisierte Serienüberwachung. Ph.D. thesis, Universität des Saarlandes (2021)
19. Plassmann, J., et al.: Unsupervised shearography dataset. <https://github.com/JessicaPlassmann/Unsupervised-Shearography> (2025)
20. Redmon, J., et al.: You only look once: Unified, real-time object detection. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 779–788 (2016). <https://doi.org/10.1109/CVPR.2016.91>
21. Rezatofghi, H., et al.: Generalized intersection over union: A metric and a loss for bounding box regression (2019), <https://arxiv.org/abs/1902.09630>
22. Rudolph, M., et al.: Asymmetric student-teacher networks for industrial anomaly detection. *Computer Science: Machine Learning* (October 2022). <https://doi.org/10.48550/arXiv.2210.07829>, <https://arxiv.org/abs/2210.07829>
23. Schmarje, L., et al.: A survey on semi-, self- and unsupervised learning for image classification. *IEEE Access* **9**, 82146 –82168 (May 2021). <https://doi.org/10.1109/ACCESS.2021.3084358>
24. Schuth, M., Buerakov, W.: *Handbuch Optische Messtechnik: Praktische Anwendungen für Entwicklung, Versuch, Fertigung und Qualitätssicherung*. Hanser, Munich, 1 edn. (2017)
25. Schwarz, J., et al.: A patch-based student-teacher pyramid matching approach to anomaly detection in 3d magnetic resonance imaging. In: *Medical Imaging with Deep Learning* (2024), <https://openreview.net/forum?id=vh01Nd5PC1>

26. Tao, N., et al.: FEM-assisted shearography with spatially modulated heating for non-destructive testing of thick composites with deep defects. *Composite Structures* **297** (October 2022). <https://doi.org/10.1016/j.compstruct.2022.115980>
27. Tao, N., et al.: Shearography non-destructive testing of thick gfrp laminates: Numerical and experimental study on defect detection with thermal loading. *Composite Structures* **282** (February 2022). <https://doi.org/10.1016/j.compstruct.2021.115008>
28. Valjakka, J., et al.: Anomaly localization in audio via feature pyramid matching. In: 2023 IEEE 47th Annual Computers, Software, and Applications Conference (COMPSAC) (2023). <https://doi.org/10.1109/COMPSAC57700.2023.00044>
29. Wang, G., et al.: Stfpm - official implementation. <https://github.com/gdwang08/STFPM> (2021)
30. Wang, G., et al.: Student-teacher feature pyramid matching for anomaly detection. *Computer Science: Computer Vision and Pattern Recognition* (March 2021). <https://doi.org/10.48550/arXiv.2103.04257>, <https://arxiv.org/abs/2103.04257>
31. Wang, L., Yoon, K.J.: Knowledge distillation and student-teacher learning for visual intelligence: A review and new outlooks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **44**(6), 3048 – 3068 (2021). <https://doi.org/10.1109/TPAMI.2021.3055564>
32. Yamada, S., Hotta, K.: Reconstruction student with attention for student-teacher pyramid matching. *Computer Science: Computer Vision and Pattern Recognition* (November 2021). <https://doi.org/10.48550/arXiv.2111.15376>, <https://arxiv.org/abs/2111.15376>
33. Zhu, J., et al.: Asymmetric teacher–student feature pyramid matching for industrial anomaly detection. *IEEE Transactions on Instrumentation and Measurement* **73** (December 2023). <https://doi.org/10.1109/TIM.2023.3338681>