

Research paper

Training-free thermographic anomaly detection using random convolution kernels

Shaoyang Peng ^{*}, Haoxuan Deng^{*}, Sri Addepalli , Maryam Farsi

Cranfield university, Faculty of Engineering and Applied Sciences, Bedford, MK43 0AL, United Kingdom

ARTICLE INFO

Keywords:

Thermographic inspection
Anomaly detection
Zero-shot learning
Random convolution kernels
Non-destructive testing (NDT)
Composite materials
Computational efficiency

ABSTRACT

Thermographic inspection has been widely used for non-destructive evaluation of composite structures, particularly for detecting impact-induced defects in carbon fibre reinforced polymer (CFRP) materials. However, most existing approaches rely on supervised learning models that require large amounts of labeled data and significant computational resources, limiting their applicability in practical industrial environments.

In this work, a lightweight and training-free framework based on random convolution features is proposed for thermographic inspection. The method eliminates the need for model training and labeled datasets by employing randomly generated convolutional kernels to extract discriminative thermal patterns, followed by simple statistical decision mechanisms for both defect classification and pixel-wise segmentation. This enables rapid deployment and efficient processing without compromising interpretability.

Experimental results on CFRP impact damage data demonstrate that the proposed approach achieves competitive performance compared to both traditional feature-based methods and deep learning models. In classification tasks, the method consistently outperforms conventional handcrafted features, while in segmentation tasks it achieves a pixel accuracy of 99.5%, comparable to the best-performing baseline. Importantly, the proposed framework significantly reduces computational cost, with negligible model size and low memory consumption compared to data-driven approaches.

Overall, the results indicate that the proposed method provides a favorable balance between accuracy and efficiency, making it particularly suitable for real-time and resource-constrained thermographic inspection scenarios.

1. Introduction

Infrared thermography provides a non-contact method for analysing heat diffusion and detecting subsurface anomalies in materials. Thermographic inspection can be broadly classified into passive and active approaches. In active thermography, external excitation sources such as flash lamps or laser pulses induce transient thermal responses governed by heat diffusion processes. This technique has proven particularly effective for evaluating carbon-fibre-reinforced polymer (CFRP) structures, where defects such as delamination, disbonds, and impact damage can be detected through analysis of surface temperature evolution [1–4]. From a physical perspective, recent studies have revisited thermal diffusion as an energy transport process, providing deeper insights into depth-time scaling behaviour in thermographic signals [5].

With the increasing adoption of robotic inspection systems, thermographic data acquisition can now be performed automatically in industrial environments. Recent developments in cognitive sensing and feedback-based inspection further highlight the growing demand for adaptive and automated thermographic analysis pipelines [6]. However, the interpretation of thermographic data still often relies on expert operators analysing large volumes of thermal images using specialised software, which introduces significant labour and computational costs [7]. Consequently, improving the efficiency and automation of thermographic data analysis remains an important research challenge for industrial NDT applications.

In recent years, machine learning and computer vision techniques have been increasingly applied to thermographic data processing in order to improve detection accuracy and inspection efficiency [8–10]. In particular, supervised deep learning models, such as convolutional

^{*} Corresponding authors.

E-mail addresses: slysp1@hotmail.com (S. Peng), Haoxuan.Deng@cranfield.ac.uk (H. Deng), p.addepalli@outlook.com (S. Addepalli), Maryam.Farsi@cranfield.ac.uk (M. Farsi).

<https://doi.org/10.1016/j.rineng.2026.111123>

Received 30 March 2026; Received in revised form 14 May 2026; Accepted 17 May 2026

Available online 20 May 2026

2590-1230/© 2026 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

neural networks (CNNs), have demonstrated strong performance in defect detection and segmentation tasks across both medical and industrial thermography applications [11–13]. However, these approaches typically require large labelled datasets for training.

To address this limitation, recent research has investigated unsupervised and data-driven strategies. A recent survey [14] provides a systematic overview of machine learning techniques in active infrared thermography, emphasizing the rising adoption of clustering, dimensionality reduction, and autoencoder-based methods in scenarios with scarce annotations. Furthermore, spatio-temporal deep learning architectures have been proposed to exploit the temporal dynamics of thermal responses. For instance, a three-dimensional convolutional autoencoder has been presented to jointly learn spatial and temporal representations from thermographic sequences [15]. More recently, physics-informed deep learning approaches have emerged, embedding physical priors into data-driven models to enhance robustness under diverse inspection conditions [16].

Despite their effectiveness, these approaches generally require substantial training data and computational resources, which limits their applicability in practical industrial scenarios. Furthermore, thermographic inspection conditions often exhibit low signal-to-noise ratios (SNR), where measurement noise, non-uniform heating, and environmental disturbances may lead to unreliable anomaly responses [17].

To address these challenges, recent research has explored unsupervised and zero-shot anomaly detection approaches that eliminate the need for labelled datasets. For example, Aota et al. [18] proposed a zero-shot framework that assumes defect-free thermographic images exhibit homogeneous background textures, enabling anomalies to be detected as deviations from this distribution. However, existing zero-shot approaches typically rely on deep pre-trained networks for feature extraction, which introduces additional computational overhead.

In this context, there is a need for computationally efficient and training-free methodologies that can extract informative features while maintaining robustness under practical inspection conditions. The objective of this work is therefore to develop a lightweight and generalisable framework for thermographic anomaly detection that supports automated NDT inspection workflows.

In this context, there is a need for computationally efficient methodologies capable of extracting informative features from thermographic data while maintaining robustness under practical inspection conditions. The objective of this work is therefore to develop a lightweight and generalisable framework for thermographic anomaly detection that can support automated NDT inspection workflows.

The main contributions of this work are threefold:

- The introduction of a training-free Random Convolution Kernels (RCKs) based feature extraction algorithm called Racon for thermographic anomaly detection;
- A multi-scale Gaussian pooling mechanism tailored to thermographic characteristics
- A comprehensive evaluation demonstrating favourable trade-offs between detection accuracy and computational efficiency.

Beyond thermographic applications, the proposed methodology is designed to be adaptable to other inspection modalities and datasets, providing a flexible foundation for future research in multimodal defect characterisation and automated condition assessment.

2. Related work

2.1. Machine learning for thermographic inspection

To improve the reliability and efficiency of thermographic inspection, machine learning approaches have increasingly been applied to thermographic data analysis. Traditional machine learning and pattern-recognition strategies have been used to classify thermographic signals and identify defective regions, while more recent studies have extended

these workflows using deep learning models [19,20]. More recently, convolutional neural networks (CNNs) and related deep learning architectures have demonstrated strong performance in thermographic defect detection and segmentation tasks across industrial applications [21,22].

However, these supervised approaches rely heavily on labelled datasets. In practical inspection scenarios, acquiring labelled thermographic datasets can be difficult and costly because thermal data collection requires specialised equipment and controlled experimental conditions, while reliable ground-truth annotation often depends on expert interpretation or destructive validation [23]. Consequently, the availability of large and diverse thermographic datasets remains limited, which restricts the generalisation capability of supervised models.

To alleviate this limitation, recent studies have similarly explored lightweight and data-efficient strategies for thermographic defect detection and industrial anomaly analysis. For example, Xue et al. [24] proposed a memory-linked few-shot learning framework for thermographic non-destructive evaluation, where memory representations are constructed to guide segmentation under limited supervision. Likewise, Shi et al. [25] investigated memory-bank-based industrial anomaly detection strategies that improve generalisation across multiple inspection scenarios through feature inheritance and memory optimisation. In parallel, Hu et al. [26] introduced a lightweight spatial-temporal multi-feature fusion network that combines temporal thermal evolution and spatial representations to enhance defect localisation capability. However, most of these approaches still require either pretraining, support samples, normal reference data, memory-bank construction, or optimisation of learned feature representations.

2.2. Zero-shot anomaly detection

To overcome the limitations of supervised learning, unsupervised and zero-shot anomaly detection methods have recently attracted increasing attention. These approaches aim to identify abnormal regions by detecting deviations from dominant background patterns without requiring labelled training data.

Early approaches relied on statistical assumptions regarding image homogeneity. Classical techniques such as the Sauvola method provide a lightweight baseline for anomaly detection by identifying deviations in local intensity statistics [27]. However, such methods are often sensitive to noise and background variations, which are commonly observed in thermographic data.

More recently, Aota et al. [18] proposed a zero-shot anomaly detection framework that assumes defect-free images exhibit homogeneous background textures. In this approach, a pre-trained deep neural network is used to extract mid-level feature maps from the test image. These features are divided into local patches, and anomaly scores are computed by measuring the similarity between patches using a k-nearest neighbour strategy, as illustrated in Fig. 1. Regions that deviate from the dominant background distribution are identified as anomalous.

Although such approaches achieve promising results, they typically rely on large pre-trained backbone networks such as WideResNet50 [28]. Similarly, recent self-supervised or vision-language models, including DINO and CLIP, have demonstrated strong representation capability for visual anomaly detection tasks [29,30]. However, the use of large backbone networks introduces significant computational and memory requirements, which may limit their applicability in real-world inspection systems or embedded devices. Unlike physics-based thermographic processing methods such as Thermographic Signal Reconstruction (TSR) and Principal Component Thermography (PCT), which aim to enhance signal quality through temporal or statistical analysis, the proposed RCKs framework is designed as a feature extraction and anomaly detection stage. In practice, TSR or PCT can be applied as a preprocessing step to improve signal-to-noise ratio, followed by RACON for efficient defect localisation. Therefore, it is not intended to replace these established techniques, but rather to complement them within a unified thermographic inspection pipeline.

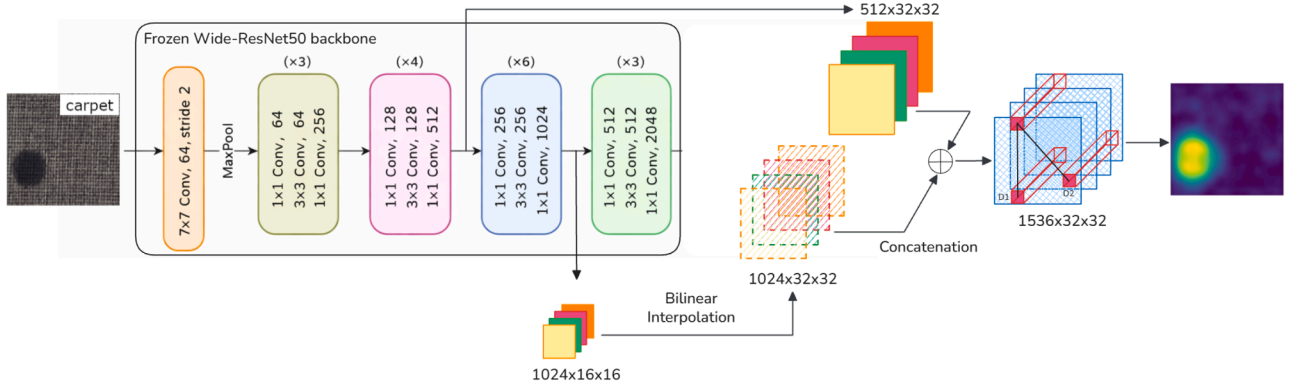


Fig. 1. The diagram of the zero-shot anomaly detection[18].

3. Random convolution kernels for feature extraction

The success of deep neural networks (DNNs) largely relies on their use of stacked layers to extract features in a hierarchical manner. While this approach is highly effective in capturing a wide range of patterns, it typically requires significant computational resources. In particular, deep models often involve high memory usage and a large number of floating-point operations, which can be limiting in real-time or resource-constrained environments.

To address this challenge, instead of relying on pre-trained deep neural networks for feature extraction, which is commonly used in many existing works, we propose a simpler and more efficient method. Our approach utilises RCKs to directly extract features from raw thermographic data. Unlike physics-based methods such as TSR and PCT, which focus on signal enhancement, RACON is designed as a feature extraction and anomaly detection stage. In practice, TSR or PCT can be used as preprocessing, followed by RACON for efficient defect localisation, making the two approaches complementary rather than competing.

3.1. Motivation

Random convolutional kernels (RCKs) have been successfully applied in 1-dimensional time series classification. An example is the ROCKET (Random Convolutional Kernels Transformation) method, which highlights the power of using a large number of randomly generated 1D convolutional filters for efficient feature extraction [31]. In ROCKET, these kernels are applied to the input time series through convolution, producing a set of feature maps. To convert these feature maps into a compact and informative feature vector, simple statistical operations are used, including taking the maximum value and computing the proportion of positive values in each feature map. This vector is then passed to a linear classifier for final prediction. In this way, the RCKs-based method allows for extremely fast training and competitive accuracy without requiring kernel learning or deep architectures.

Experiments on the open benchmark dataset demonstrate that ROCKET delivers strong classification performance, even without any parameter optimisation during the feature extraction stage. In many cases, it achieves results comparable to, or even surpasses, state-of-the-art deep learning models and ensemble methods. Notably, this performance is achieved with a significantly lower computational cost and much shorter training time, making it an efficient alternative for time series classification tasks.

One explanation behind the strong performance of ROCKET lies in the functional role of its basic unit. As suggested by Costas et al.(2011) [32], such a combination of random convolution kernels, square non-linear activation functions, and pooling operations can act as a frequency selector. Their mathematical analysis suggests that the combination of convolution and square-sum pooling acts as an effective frequency selector. The non-linearity transforms phase-dependent re-

sponses into energy-based representations, while pooling aggregates these responses to emphasise dominant spectral components. By performing gradient ascent on the output of this unit, the most responsive frequency patterns can be explicitly revealed, providing a clear interpretation of the implicit frequency filtering behaviour induced by this architecture.

Beyond the square activation, we observe that several commonly used non-linear activation functions, such as Sigmoid and ReLU, exhibit a similar phenomenon. These activations functions can emphasise distinct structural patterns, or motifs, within the input data. Despite their differences, these activations demonstrate comparable pattern-selective behaviour, enabling the model to capture a diverse set of meaningful and discriminative structures.

Importantly, all operations are carried out within a single shallow layer, resulting in a highly compact and computationally efficient model compared to deep neural networks. This lightweight yet effective design constitutes the core motivation behind the feature extraction architecture proposed in this study.

3.2. Group convolution for multi-channel image processing

The original ROCKET method is tailored for single-channel 1D time series data, focusing solely on one-dimensional signal processing. Similarly, the work by Costa et al.(2011) [32] explores the use of RCKs only in the context of single-channel greyscale images. However, in real-world scenarios, thermographic image data is commonly stored in RGB format (e.g., as PNG files), consisting of three colour channels. To address this, we extended the RCKs framework to support multi-channel inputs and introduced group convolutions to manage computational complexity. Instead of generating full three-channel 2D convolution kernels with shape $(H \times W \times 3)$, where H and W denote the height and width of the kernel, we generate a set of single-channel 2D kernels of shape $H \times W$ and divided them equally into three groups. These kernels are then divided equally into three groups, each assigned to one of the RGB channels and applied independently using group convolution. As a result, the number of kernels per channel becomes $C/3$, where C is the total number of generated kernels. Although this approach does not explicitly model cross-channel interactions as standard multi-channel convolutions do, it still delivers competitive prediction performance. Our experiments show that it compares favourably with many state-of-the-art methods while maintaining a lightweight and efficient computational profile.

3.3. Multi-scale extension of the RCKs

In this study, we further extended the use of RCKs by incorporating multi-scale convolution operations to extract features at different spatial resolutions. This design is motivated by scale-space theory, which

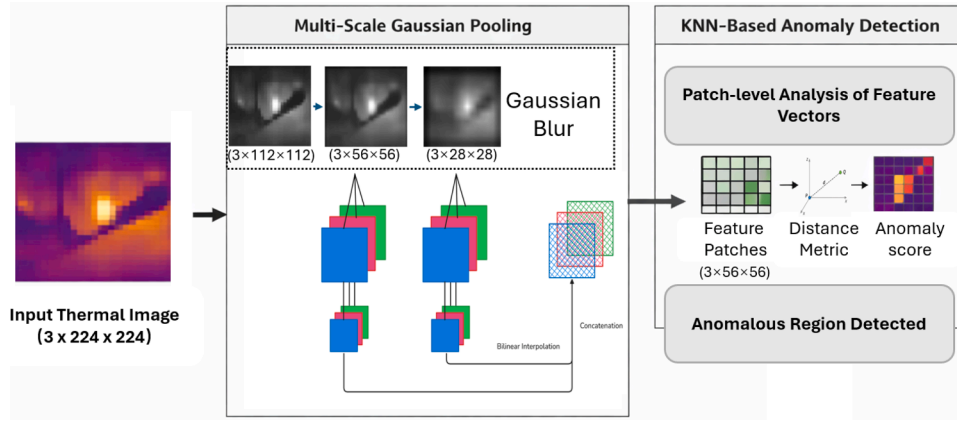


Fig. 2. The RACON pipeline consisting of multi-scale Gaussian pooling and KNN-based anomaly detection.

suggests that real-world objects consist of structures that appear at multiple scales. By using a scale-space representation, it becomes possible to identify and analyse features of varying sizes, improving the robustness of feature extraction, particularly in the presence of noise or fine-grained variations.

To implement this concept, we applied a widely used approach based on Gaussian smoothing. The Gaussian function serves as the smoothing kernel due to its well-established mathematical properties, such as scale invariance and smoothness. By convolving the input image with kernels of different scales, we can effectively capture structural patterns at multiple resolutions, further enriching the feature set extracted by the random convolutional architecture.

3.4. Overall architecture

The overall architecture of the proposed method is illustrated in Fig. 2. First, we apply a Gaussian blur before random convolution to suppress local pixel noise and amplify semantically-relevant thermal patterns, similar to the temporal smoothing benefits observed in the ROCKET framework [31]. After that, a collection of convolution kernels with random weights in the shape of $H \times W \times C$ is generated and applied to each channel independently using group convolution. This design reduces computational complexity while preserving channel-wise information.

Following the convolution and non-linear activation function, average pooling is applied to further reduce the spatial dimensions of the feature maps. This pooling step also serves as a means of aggregating local information into patch-level representations, as demonstrated in [18]. Subsequently, all feature maps from different resolution levels are upsampled to a unified spatial size using bilinear interpolation. Finally, the upsampled feature maps are concatenated along the channel dimension to form the final representation, which is then passed through the rest of the pipeline, as illustrated in Fig. 1.

4. Experiment

4.1. Dataset

The dataset used in this study was introduced by Garcia Vargas (2025) [33]. It consists of a total of 1045 RGB images in PNG format, including both normal and defective samples. The inspected specimen is a unidirectional carbon/PEEK laminate (100×100 mm) with a fiber volume fraction of 61% and a stacking sequence [02/902]. Artificial delamination was introduced by embedding Kapton tape inserts of varying sizes (2×2 mm, 3×3 mm, and 4×4 mm) at different depths (0.13 mm, 0.26 mm and 0.39 mm) prior to the moulding process and providing a controlled set of defect conditions for evaluation. Thermal

data were collected using a mid-wave infrared (MWIR) camera operating at a spatial resolution of 640×512 pixels and a frame rate of 55 Hz. Each PT excitation produced a sequence of thermograms that captured the transient heat diffusion process on the laminate surface. For annotation, binary masks are provided to indicate defect regions: pixels labelled as 0 represent defect-free areas, while pixels labelled as 1 correspond to defective regions.

Fig. 3 presents several representative examples of both normal and defective samples together with their corresponding binary masks. Although the embedded defects are nominally square in geometry and vary in size and number across different samples, their visibility depends strongly on the heat diffusion process and the resulting thermal contrast. In many frames, defects are not clearly distinguishable to the human eye because the temperature differences between defective and sound regions can be subtle, particularly as thermal gradients decay over time. Additionally, measurement noise, non-uniform heating, and the inherently smooth thermal response of composite materials may lead to blurred boundaries and spatially varying intensity patterns [34]. These factors collectively reduce visual separability and introduce significant challenges for defect identification, thereby providing a realistic and physically meaningful test scenario for evaluating the robustness of anomaly detection algorithms [35].

The entire dataset is divided into training, validation, and testing sets with a split ratio of 70%, 15%, and 15%, respectively, according to the paper [33].

4.2. Metrics

In anomaly detection and localisation tasks, especially in the context of image data, performance evaluation plays a critical role in assessing the reliability and robustness of the model. Among commonly used evaluation metrics, the Area Under the Receiver Operating Characteristic Curve (AUROC) is widely adopted due to its threshold-independent nature and ability to capture the overall discriminative performance of a model. [36–38] There are generally two levels of AUROC applied in image anomaly detection:

- **Image-Level AUROC (I-AUROC)** evaluates whether an entire image is classified as normal or anomalous. It is computed by comparing the predicted anomaly score (often a single value per image) against the ground truth label for the image (normal or defected).
- **Pixel-Level AUROC (P-AUROC)** assesses the model's ability to localise anomalies at the pixel level. Here, an anomaly score is assigned to each pixel, and the AUROC is computed by comparing these scores to the binary ground truth mask indicating defect regions.

The commonly used Intersection over Union (IoU) metric is not applied in this case due to several limitations in anomaly detection and

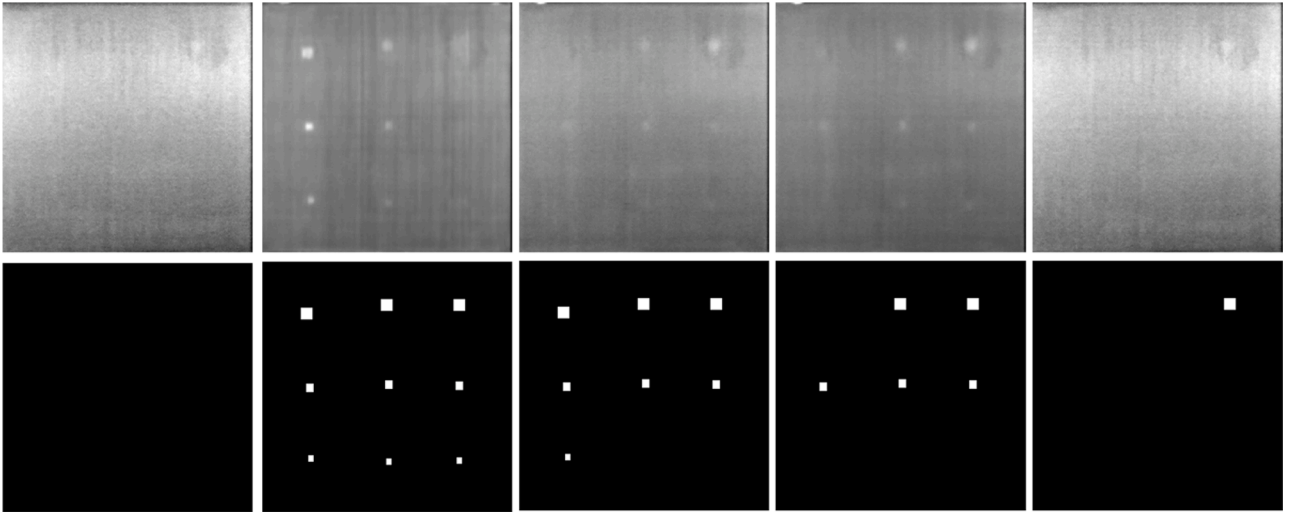


Fig. 3. Examples of original (top row) and annotated images (bottom row) [33].

localisation. IoU requires a fixed threshold to binarise the anomaly map, making performance highly sensitive to the threshold choice. It is also unreliable for detecting small or sparse anomalies, where even accurate predictions can result in low IoU scores. Moreover, IoU provides a binary evaluation, ignoring the confidence of predictions. In contrast, pixel-level AUROC offers a threshold-independent assessment, captures the model's ability to rank anomalous pixels above normal ones, and remains informative even under uncertainty or when defects are subtle.

In addition to using I/P-AUROC metrics to evaluate the detection accuracy of each model, we also assess their computational efficiency through a set of runtime metrics. These runtime metrics provide insight into the resource requirements and practical feasibility of deploying the models in real-world scenarios, especially on resource-constrained devices.

- **Inference Latency:** This refers to the time it takes for the model to process a single input image and produce a prediction. Lower latency is critical for real-time or low-latency applications, such as industrial anomaly detection systems.
- **Model File Size:** This is the storage size of the trained model on disk. A smaller model size is preferable for deployment in environments with limited memory or storage capacity, such as embedded systems or edge devices.
- **Peak Runtime Memory (RAM) Consumption:** This captures the highest amount of memory used during inference, including both model weights and intermediate computations. It provides a realistic view of the memory footprint required to run the model in practice.

By combining both detection accuracy and runtime performance metrics, we offer a comprehensive evaluation of each model in terms of predictive quality with computational efficiency.

4.3. Baselines

4.3.1. Fully supervised learning

To evaluate the effectiveness and efficiency of our proposed RCKs-based method, we compare it against several widely used baseline approaches on the thermographic dataset. These baselines are grouped into three main categories namely, fully supervised learning, few-shot unsupervised learning, and zero-shot unsupervised learning, based on the amount of labelled data and prior knowledge required during training.

For the fully supervised learning baseline, we adopt the widely used U-Net architecture, which has become a standard model for pixel-wise image segmentation tasks, including anomaly detection and localisation.

U-Net follows an encoder-decoder structure, where the encoder progressively reduces the spatial resolution of the input image to capture high-level semantic features, while the decoder upsamples these features to produce an output mask that aligns with the original image dimensions.

In our implementation, the input is a 3-channel RGB image of shape $224 \times 224 \times 3$. The encoder path consists of five downsampling blocks, each halving the spatial dimensions and doubling the number of channels, ultimately expanding the feature depth to 1024 channels. The decoder then includes five upsampling blocks, which gradually recover the spatial resolution and reduce the channel dimension back to 3, matching the original input format. This setup enables the network to learn rich hierarchical features and accurately reconstruct pixel-level defect predictions.

To optimise the training process, we use a composite loss function, denoted by $\mathcal{L}$ that balances Binary Cross-Entropy (BCE) Loss and Dice Loss [39]:

$$\mathcal{L} = \alpha \cdot BCELoss + (1 - \alpha) \cdot DiceLoss$$

Specifically, the $BCELoss$ measures the pixel-wise classification error between the predicted mask and the ground truth binary mask, and the $DiceLoss$ is a region-based metric that emphasises overlap between the predicted and actual defect regions, which is particularly useful in handling class imbalance. A weighting hyperparameter α is introduced to flexibly balance the contributions of these two loss terms.

4.3.2. Few-shot unsupervised learning

For the unsupervised few-shot learning baseline, we adopt the well-established PatchCore [40]. This approach can be used to perform anomaly detection with a very limited number of normal samples. In our setup, we randomly select normal samples from the training set to serve as the reference set. During the feature extraction stage, we follow the strategy proposed in the original PatchCore paper [40]. Specifically, we use a WideResNet-50 model pre-trained on the ImageNet-1k dataset to extract mid-level features. Features are taken from the output of the second and third convolutional blocks, which are known to capture meaningful texture and structural information useful for detecting anomalies. The extracted feature maps from these two blocks are resized to a uniform spatial resolution and concatenated along the channel dimension, resulting in a combined feature map of shape $1,536 \times 28 \times 28$ for each input image of shape $3 \times 224 \times 224$. This high-dimensional feature representation encodes both local and contextual information from the image. Next, we construct a memory bank using the features of the four normal reference samples using a core-set selection algorithm proposed in the original paper, which selects a representative subset of

Table 1

The I-AUROC(%) / P-AUROC(%) of different methods for the thermographic anomaly detection.

Fully Supervised Learning	4-shot Unsupervised Learning	Zero-shot Unsupervised Learning		
U-Net	PatchCore	Sauvola's method	Zvsm(Resnet)	Ours
94.07/98.89	83.35 $\pm$ 0.026/85.56 $\pm$ 0.024	85.04/92.65	94.04/87.83	94.0 $\pm$ 0.011/92.19 $\pm$ 0.002

features from the full memory to reduce redundancy and improve inference speed.

4.3.3. Zero-shot unsupervised learning

For the zero-shot learning baseline, we adopt the method proposed by Aota et al.(2023) [18], following the original configuration precisely. This includes using the same backbone network and specifying the same layer for feature extraction. Unlike supervised or few-shot approaches, zero-shot methods do not require any training or adaptation on the target dataset, making them particularly appealing for scenarios with limited or no labelled data. In this method, the input image is first rescaled to a resolution of 340×340 pixels. Feature extraction is then performed using the second convolutional block of a WideResNet-50 model pre-trained on ImageNet-1k. The resulting feature representation has a shape of $512 \times 40 \times 40$, capturing rich mid-level information across the entire image. The KNN then computes the distance between local features on the test image. Areas that significantly deviate from this distribution are flagged as anomalous. After that, the resulting score map is upsampled to match the original image resolution. This is done using Gaussian smoothing, which not only increases the resolution but also smooths the score map, reducing noise and improving the spatial coherence of detected regions. Finally, a centre crop of size 256×256 is applied to the upsampled image. This step helps to mitigate false positive detections near the image borders, which are a common issue in zero-shot models due to padding and boundary effects in convolutional operations.

4.4. Setup

For the U-Net model, we set the balancing coefficient $\alpha = 0.5$ to give equal weight to both components of the BCE loss and Dice Loss. The model is trained using the Adam optimiser, a widely adopted stochastic gradient-based algorithm known for its effectiveness in deep learning tasks. We use an initial learning rate of 1×10^{-4} and train the model for 60 epochs using the training split, which includes both input images and their corresponding binary defect masks. After training, we evaluate the U-Net's performance on the test split using the metrics described earlier, including I/P-AUROC and runtime measurements.

For the PatchCore baseline, we randomly select four normal samples from the training set to construct a reference memory bank, following the original procedure outlined in the PatchCore paper. To ensure statistical robustness and reduce sampling bias, this selection process is repeated using five independent random seeds, and the reported results are averaged across these runs. Once the memory bank is established, the model is evaluated on the test set. Anomalies are identified by computing the distance between local features from the test image and those stored in the memory bank. Larger deviations from the normal feature distribution indicate a higher likelihood of an anomaly in a given region.

As for Aota et al.'s method and our proposed approach, both are zero-shot unsupervised methods that do not require labelled training data. To prevent any potential data leakage, we apply both methods directly on the test set without using any training samples or annotations.

In our method, as illustrated in Fig. 2, the input image is first processed through consecutive Gaussian blurring and pooling operations to construct multi-scale representations. Given an input image of size $(3 \times 224 \times 224)$, this procedure generates feature maps at three spatial resolutions: $(C \times 64 \times 64)$, $(C \times 32 \times 32)$, and $(C \times 16 \times 16)$, where C denotes the number of random convolution kernels applied per channel.

At each resolution, random convolution kernels are applied in a channel-wise grouped manner, followed by a sigmoid non-linear activation to normalise the responses into the range $[0,1]$. Subsequently, an average pooling operation with a 2×2 windows and a stride of 2 is employed to further aggregate local information. To enable consistent feature aggregation across scales, all resulting feature maps are resized to a unified resolution of $(C \times 32 \times 32)$, forming a common representation for anomaly detection. Finally, the same KNN-based algorithm proposed by Aota et al.(2023) [18] is applied, treating these unified feature maps as patch-level descriptors.

A key hyperparameter in both methods is the *top-k* value, which defines how many nearest neighbours are considered when calculating the anomaly score for each patch. In our experiments, we set $k = 10$, ensuring a consistent and fair comparison across both approaches. To ensure statistical robustness, all experiments are repeated using five independent random seeds, and the reported results correspond to the average performance across these runs.

In addition, the classical Sauvola's method [27] is included as a simple baseline for comparison. As a local adaptive thresholding algorithm, Sauvola computes pixel-wise thresholds based on local statistics and is capable of detecting intensity deviations without requiring any learning or feature extraction.

All models are evaluated on the same hardware platform, which is equipped with an Intel® Core™i5-12600K CPU (10 cores), 32 GB of RAM, and an NVIDIA RTX 4090 GPU with 24 GB of VRAM.

4.5. Experiment results

Table 1 summarises the I-AUROC and P-AUROC performance of different approaches for thermographic anomaly detection. These indicators can accurately reflect the actual performance of an image segmentation algorithm. The fully supervised U-Net model achieves the highest pixel-wise AUROC (98.89%) and a strong image-level AUROC (94.07%), reflecting the benefit of abundant labelled data and end-to-end optimisation. Interestingly, our proposed method (labeled "Racon" in our results) attains comparable image-level performance (94.0%) to both U-Net and the zero-shot deep feature method of Aota et al.(labeled "ZvSM" in our results), while maintaining a competitive pixel-level AUROC (92.19%). This demonstrates that, despite its simplicity and lack of training, the RCK approach can reach performance levels close to fully supervised deep networks. PatchCore, as a 4-shot method, performs notably worse on both metrics, likely due to the limited number of support samples. The classical Sauvola's method threshold method also achieves reasonable pixel-level scores (92.65%) but underperforms on image-level AUROC, indicating lower robustness.

It is important to note that I-AUROC and P-AUROC capture different aspects of anomaly detection performance. I-AUROC reflects the model's ability to distinguish between defective and non-defective samples at a global level, while P-AUROC evaluates the accuracy of spatial localisation. In thermographic inspection, these two aspects are not always aligned. Methods that produce smooth anomaly maps may achieve higher I-AUROC due to consistent global responses, but may lack precise localisation, leading to lower P-AUROC. Conversely, methods that emphasise local gradients may achieve better localisation at the cost of global consistency.

The proposed RACON framework achieves a balanced performance across both metrics, indicating its capability to capture both global anomaly presence and local defect structure.

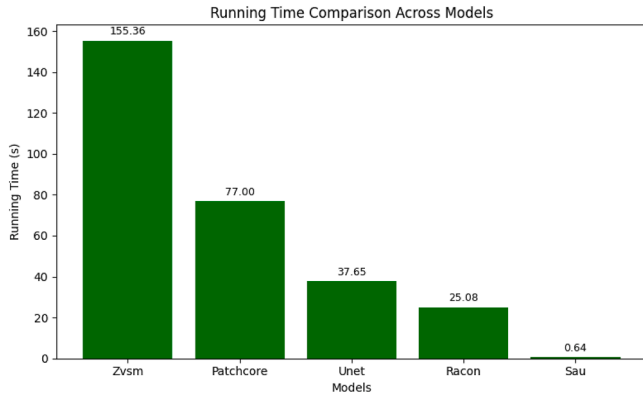


Fig. 4. The comparison of inference latency of each model.

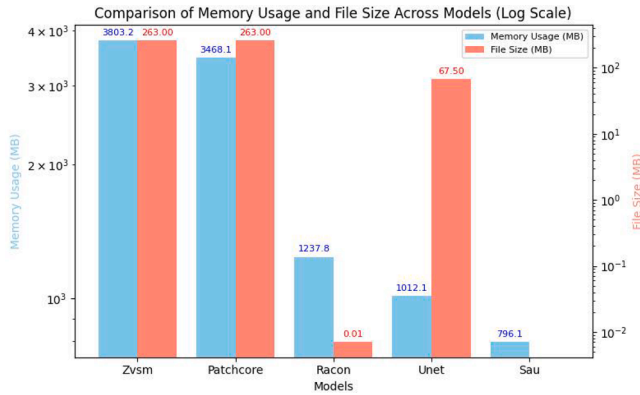


Fig. 5. The comparison of model storage and runtime memory consumption.

Fig. 4 compares the inference latency of each evaluated model under identical input dimensions and hardware conditions. As shown, traditional deep learning-based models such as U-Net and PatchCore exhibit significantly higher latency, with average inference times exceeding 150ms per image. In contrast, our proposed RCK-based method achieves inference within approximately 30ms, representing a $5\times$ reduction in latency compared to U-Net and a $4.5\times$ speed-up relative to PatchCore. This efficiency is primarily attributed to the training-free nature of the RCK pipeline, absence of learned weights, and the shallow convolutional structure that eliminates the need for backpropagation or feature aggregation layers.

Notably, while Sauvola's method demonstrates the fastest inference, its performance in anomaly localisation and robustness is substantially lower than that of RCK.

Fig. 5 presents a comparative analysis of model storage requirements and runtime memory consumption across five anomaly detection approaches: a fully supervised U-Net, the few-shot method PatchCore, the classical Sauvola thresholding baseline, the zero-shot deep feature model proposed by Aota et al.(2023), and the proposed RCK-based random convolution model. For brevity, these methods are denoted as ZVSM, PatchCore, RACON, U-Net, and SAU, respectively. The blue bars correspond to the peak runtime memory usage measured in megabytes (MB), while the red bars represent the model file sizes, each referenced to its respective y-axis. Both axes are displayed on logarithmic scales to enable clear comparison among models whose resource demands span multiple orders of magnitude.

As shown in the figure, ZVSM and PatchCore, both built upon the WideResNet50 backbone, exhibit the highest runtime memory consumption, requiring approximately 3.8 and 3.5 GB, respectively, and also incur substantial storage overheads of around 263 MB. In comparison, U-Net demonstrates a moderate runtime memory footprint of roughly 1.0

GB, yet still maintains a relatively large model size of 67.5 MB, reflecting the parameter-intensive nature of deep convolutional architectures.

By contrast, RACON achieves a dramatically reduced storage footprint, with a model size of only 0.01 MB, while its runtime memory usage remains at approximately 1.24 GB, illustrating the decoupling between parameter count and intermediate feature storage. The proposed SAU model attains the most compact and efficient profile overall, combining the lowest memory consumption (approximately 796 MB) with a negligible model file size, thereby clearly demonstrating the advantages of its lightweight design.

Overall, this comparison provides quantitative evidence of the efficiency characteristics of the proposed shallow, random-feature-based framework. RACON requires only 0.01 MB of model storage, representing more than a 99.99% reduction compared with ZVSM and PatchCore (263.00 MB). In terms of runtime memory consumption, RACON uses 1237.8 MB, which is approximately 67.5% lower than ZVSM (3803.2 MB) and 64.3% lower than PatchCore (3468.1 MB). Its memory usage is within the same order of magnitude as U-Net (1012.1 MB), although slightly higher by approximately 22.3%. These results indicate that RACON does not minimise runtime memory across all methods, but it provides a substantial reduction in model storage while maintaining a training-free inference pipeline, making it attractive for deployment in resource-constrained thermographic inspection scenarios.

5. Discussion

5.1. Why the RACON works

Fig. 6 presents a qualitative comparison of anomaly score maps produced by different methods on a representative thermographic sample. The original thermogram exhibits low contrast and spatially smooth temperature variations, which are characteristic of active thermography due to heat diffusion and subsurface attenuation. These properties make defect boundaries inherently ambiguous and challenging to localise precisely.

Clear differences can be observed in the spatial behaviour of the anomaly maps generated by each method. The fully supervised U-Net produces sharply defined and well-localised segmentation results, closely matching the ground truth mask. This is expected, as the model is trained with pixel-wise supervision, enabling accurate boundary reconstruction. However, such performance depends heavily on labelled data and may be sensitive to variations in acquisition conditions.

In contrast, ZVSM produces smooth and spatially diffused anomaly responses. While the main defect region is correctly identified, the activation extends beyond the true boundary. This behaviour reflects the nature of deep feature embeddings, which emphasise global structural consistency over fine-grained spatial precision. As a result, ZVSM typically achieves strong image-level discrimination (high I-AUROC) but weaker localisation accuracy (lower P-AUROC). PatchCore exhibits a similar trend, generating coarse anomaly maps due to its reliance on patch-level feature similarity, which captures global deviations but does not enforce pixel-level sharpness.

The classical Sauvola method produces highly fragmented and noisy anomaly maps. Its sensitivity to local intensity variations leads to scattered responses across the background, particularly in regions affected by thermal noise or surface artefacts. While this behaviour may capture fine local fluctuations, it lacks global consistency, resulting in unstable detection performance.

The proposed RACON framework achieves a balanced behaviour between these extremes. The anomaly map demonstrates clear localisation of the defect region while maintaining effective background suppression. This can be attributed to the multi-scale random convolutional feature extraction, which captures both local thermal gradients and broader structural deviations. Unlike deep feature-based methods, RACON preserves spatial consistency without excessive smoothing, and

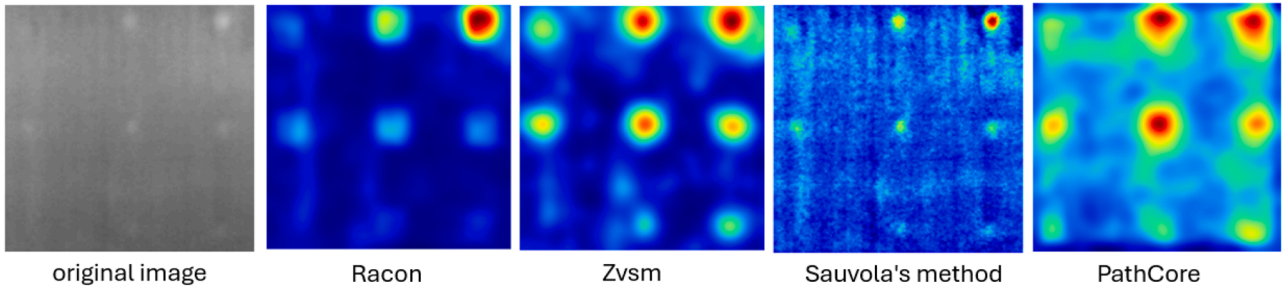


Fig. 6. Visual comparison of anomaly score maps produced by different methods on a representative thermographic sample.

unlike intensity-based approaches, it is less sensitive to noise due to its operation in a feature space rather than directly on raw intensities.

These qualitative observations are consistent with the quantitative results reported in Table 1. Methods that produce smooth anomaly maps (e.g., ZVSM, PatchCore) tend to achieve higher I-AUROC but lower P-AUROC, whereas methods with strong local sensitivity (e.g., Sauvola) may exhibit improved localisation in certain regions but suffer from poor global discrimination. RACON demonstrates a favourable trade-off, achieving competitive performance across both metrics.

This behaviour reflects a key property of thermographic data: defect signatures often manifest as subtle, low-frequency spatial anomalies rather than complex semantic structures [41]. Consequently, effective detection does not necessarily require deep hierarchical representations, but rather the ability to identify consistent spatial deviations from the background. This finding also aligns with recent progress in lightweight and data-efficient thermographic anomaly detection [24–26], where the focus is increasingly on reducing reliance on fully supervised large scale training while maintaining strong defect representation capabilities. Nonetheless, the majority of current lightweight or memory-based methods still depend on pre-trained feature extractors, memory bank construction, support-sample selection, or specialised feature optimisation schemes. By contrast, the proposed framework adopts a fundamentally different paradigm that leverages stochastic feature projection and within-image statistical comparison, thus obviating the need for feature pre-training, memory-bank optimisation, or support-image construction. By projecting the input into a high-dimensional random feature space, RACON enhances such deviations while suppressing noise-induced fluctuations. In practice, local disturbances such as sensor noise or isolated hot pixels may alter intensity values, but they rarely produce consistent responses across multiple kernels and scales. This consistency-driven representation improves robustness and leads to anomaly maps that are more aligned with the underlying defect structure compared with direct intensity-based methods such as Sauvola [27].

5.2. Ablation analysis and discussion

To systematically evaluate the design choices of the proposed RACON framework, an ablation study is conducted focusing on two key components: (i) the size of the random kernel bank, (ii) the selection of feature extraction layers. This analysis aims to identify how each component contributes to detection performance and to characterise the trade-off between accuracy and efficiency.

5.2.1. Effect of random kernel bank size

The number of random convolutional kernels is varied from 9 up to 2049, with the feature extraction layer fixed to Layer 2. As illustrated in Fig. 7, enlarging the kernel bank consistently boosts pixel-level performance, increasing P-AUROC from 0.863 to 0.953. This demonstrates that a larger kernel set enriches feature diversity and enhances defect localisation.

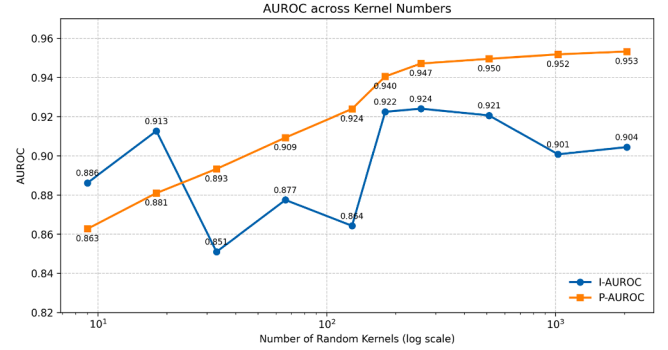


Fig. 7. Effect of kernel bank size on detection performance.

In contrast, image-level performance follows a non-monotonic pattern. Moderate increases in kernel count improve I-AUROC, but an excessive number of kernels introduces redundancy, causing slight fluctuations in performance. This implies that global discrimination benefits from kernel diversity only up to a certain threshold, after which additional kernels provide limited incremental information.

From a computational standpoint, Fig. 8a and b indicate that both runtime and memory usage grow steadily with kernel size. Average inference time rises from 0.021 s to 1.958 s per image, while peak memory usage increases from roughly 1901 MB to 2277 MB. Importantly, performance improvements saturate beyond approximately 258–513 kernels, suggesting that a moderate kernel count offers a favourable trade-off between accuracy and efficiency.

5.2.2. Effect of layer configuration

To examine the effect of hierarchical feature representations, we evaluate multiple layer configurations, including single-layer (L1, L2, L3), dual-layer (L1,2, L1,3, L2,3), and the full multi-layer setup (L1,2,3).

In the proposed framework, hierarchical features are constructed through successive Gaussian pooling operations, which progressively reduce spatial resolution while increasing the level of abstraction. Specifically, L1, L2, and L3 correspond to feature maps of size 112×112 , 56×56 , and 28×28 , respectively. The performance comparison is shown in Fig. 9a. Single-layer results reveal clear differences across representation scales. L1 and L2 achieve strong performance, while L3 performs poorly at the image level (I-AUROC = 0.114). This indicates that high-level, low-resolution features alone are insufficient for reliable global anomaly detection. The degradation can be attributed to the loss of fine-grained thermal information during repeated pooling, which is critical for identifying subtle temperature variations in thermographic data [41].

From a representational perspective, L1 preserves high spatial resolution and captures detailed local temperature gradients, which are essential for detecting small or weak defects. L2 provides a balanced representation, retaining sufficient spatial detail while incorporating moderate abstraction, leading to strong and stable performance. In contrast,

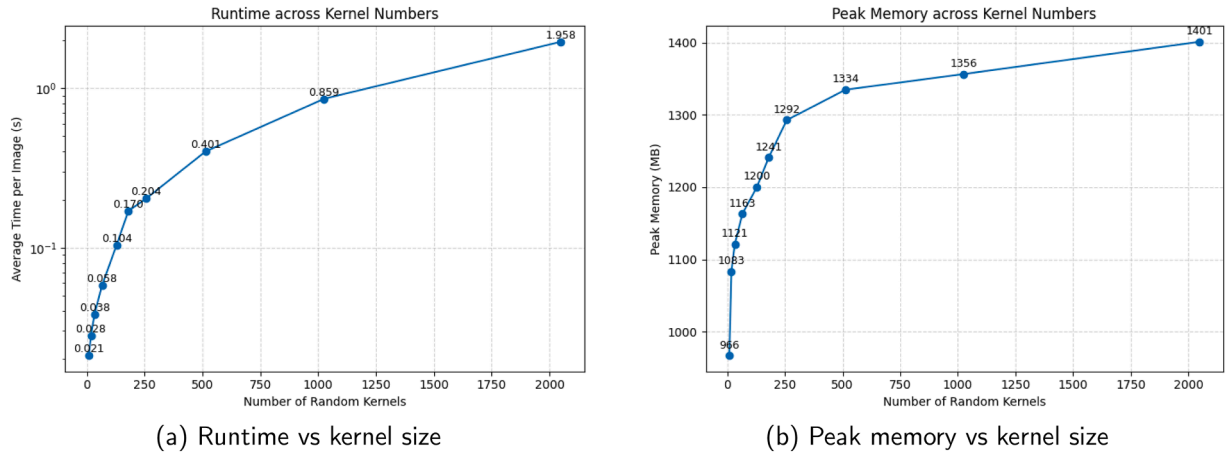


Fig. 8. Computational cost analysis with respect to the number of random kernels: (a) runtime per image (log scale), and (b) peak memory usage.

L3 focuses on coarse, highly abstracted patterns, where local thermal variations are significantly smoothed out. While such representations are effective for semantic recognition tasks, they are less suitable for thermographic anomaly detection, where defects manifest as subtle local deviations rather than high-level structures.

Multi-layer configurations consistently improve performance by combining complementary information across scales. The full multi-layer configuration L1,2,3 achieves the best overall results (I-AUROC = 0.948, P-AUROC = 0.922), demonstrating the benefit of integrating both fine and coarse representations. Similarly, L12 (I-AUROC = 0.944, P-AUROC = 0.926) and L13 (I-AUROC = 0.921, P-AUROC = 0.920) provide competitive performance, confirming the effectiveness of multi-scale feature fusion. In contrast, L2,3 shows a pronounced decline in image-level performance (I-AUROC = 0.243) while maintaining a relatively high P-AUROC, reinforcing that the lack of low-level features markedly reduces overall discriminative power.

Importantly, considering both performance and computational efficiency, L2 represents the most favourable configuration. While multi-layer setups achieve slightly higher peak accuracy, they introduce additional computational overhead due to feature concatenation and multi-scale processing. In comparison, L2 achieves near-optimal performance with significantly lower complexity, making it a more practical choice for real-world deployment scenarios where efficiency is critical. These findings highlight that effective thermographic anomaly detection requires a careful balance between spatial resolution and feature abstraction. Low-level features are essential for capturing subtle temperature variations, while higher-level features provide complementary contextual information for localisation. The integration of multi-scale representations therefore plays a critical role in achieving robust performance.

5.3. Robustness to thermal and acquisition-induced disturbances

A key challenge in thermographic anomaly detection is reliably separating genuine subsurface defects from artefacts created by acquisition conditions, including non-uniform heating, emissivity changes, reflections, and environmental disturbances. These acquisition effects frequently appear as local hotspot-like regions or structured background patterns, which can substantially modify the observed intensity distribution without corresponding to real defects.

The dataset employed in this work naturally contains thermal background noise introduced during the acquisition process. However, such noise is uncontrolled and cannot be systematically adjusted for sensitivity analysis. To enable a more rigorous robustness assessment, we introduce controlled, optical-style perturbations by superimposing 5 weak, localised hotspots with varying spatial configurations.

Table 2

Performance under optical perturbation and corresponding degradation (%) under 5-seed test.

Method	I-AUROC (%)	Drop (%)	P-AUROC (%)	Drop (%)
RACON	87.15 ± 8.55	7.29	63.57 ± 9.05	31.03
ZVSM	76.24 ± 0.67	18.93	57.57 ± 6.31	34.48
PatchCore	75.10 ± 6.25	9.90	76.18 ± 6.20	10.97
Sauvola	50.00 ± 0.00	41.23	87.23 ± 1.30	5.85
U-Net	49.33 ± 5.92	47.60	51.60 ± 0.33	47.78

Specifically, the perturbations are generated by adding a set of spatially distributed Gaussian-like hotspot patterns with randomly sampled positions, amplitudes, and spatial extents. These parameters are designed to mimic realistic acquisition artefacts commonly observed in thermographic imaging, including non-uniform excitation, sensor drift, and reflection-like responses. By varying the number, intensity, and spread of these hotspots, the perturbation process produces diverse disturbance patterns that challenge both global discrimination and local localisation capabilities. As illustrated in Fig. 10, the generated perturbations visually resemble realistic acquisition artefacts while remaining spatially controlled. To ensure statistical robustness and reduce sensitivity to specific perturbation instances, the perturbation process is repeated using three independent random seeds. All reported results correspond to the mean performance across these runs, with the standard deviation used to quantify variability.

This perturbation model operates purely in the image domain and does not attempt to fully simulate the underlying radiometric imaging physics. Instead, it provides a controlled and visually plausible approximation of typical acquisition disturbances, enabling systematic evaluation of model robustness under distribution shift.

As illustrated in Table 2 and Fig. 11, the resulting performance degradation trends remain consistent across different perturbation levels and evaluation metrics. Comparing the baseline results Table 1 with the perturbed results reveals pronounced differences in robustness between methods. The fully supervised U-Net exhibits the most severe degradation. While it achieves the highest performance under unperturbed conditions (94.07% I-AUROC and 98.89% P-AUROC), its image-level AUROC drops dramatically to 0.4933 ± 0.0592 under optical perturbation, with the pixel-level AUROC also decreasing to 0.5160 ± 0.0033 . This substantial decline highlights a strong sensitivity to distribution shift. As U-Net learns a direct mapping from image appearance to defect labels based on the training distribution, the introduction of hotspot-like artefacts alters the input statistics in ways not encountered during training. Consequently, the learned representations fail to generalise, leading to unreliable global anomaly detection.

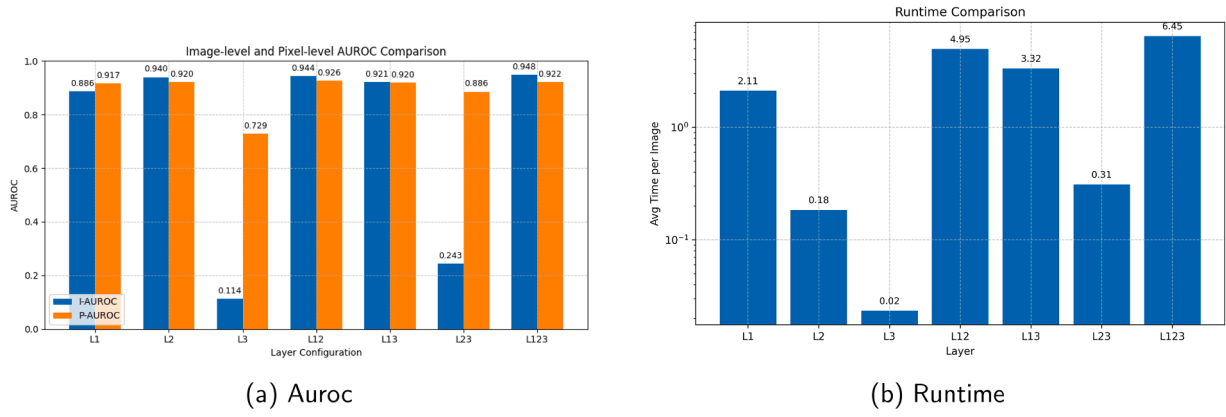


Fig. 9. Comparison of image-level and pixel-level AUROC across different layer configurations.

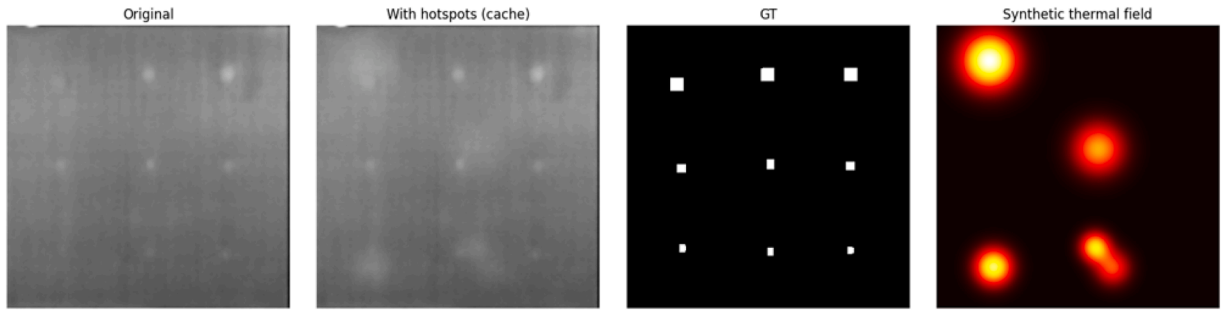


Fig. 10. Generation of optical-like perturbations. From left to right: original thermogram, perturbed image with synthetic hotspots, ground truth mask, and the corresponding synthetic thermal field used for perturbation modelling.

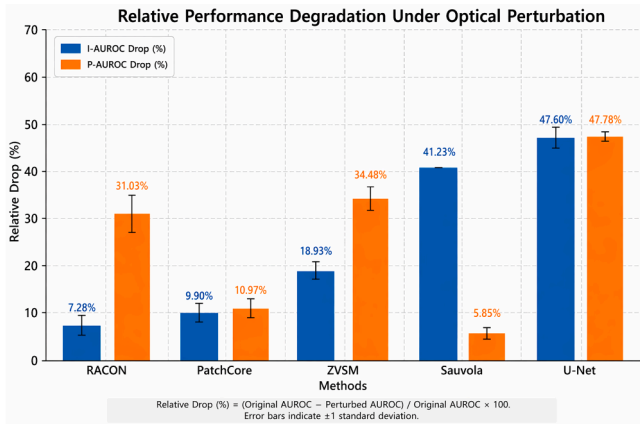


Fig. 11. Relative AUROC Degradation Under Optical Perturbations.

In contrast, training-free approaches such as RACON and ZVSM demonstrate significantly improved robustness. RACON maintains a high image-level AUROC of 0.8715 ± 0.0855 , while ZVSM achieves 0.7624 ± 0.0067 , indicating comparatively stable global discrimination. These methods do not rely on learned data distributions but instead capture intrinsic structural irregularities, making them inherently less susceptible to distribution shift. PatchCore shows moderate degradation, with I-AUROC decreasing to 0.7510 ± 0.0625 . Although it leverages pre-trained features, its reliance on a normality reference makes it partially sensitive to perturbation-induced distribution changes.

From a pixel-level perspective, a different behaviour emerges. The Sauvola method achieves a high P-AUROC of 0.8723 ± 0.0130 , primarily due to its strong sensitivity to local intensity variations. In defect-

tive images, defect-induced contrast dominates local statistics, and moderate perturbations have limited impact on detection. However, this behaviour is highly asymmetric. In defect-free images, the introduced hotspots are indistinguishable from true defects under local thresholding, leading to widespread false positives and a severe drop in I-AUROC (0.5000 ± 0.0000). PatchCore demonstrates strong overall robustness, achieving a relatively high P-AUROC (0.7618 ± 0.0620) and a moderate I-AUROC (0.7510 ± 0.0625) under perturbation, resulting in one of the smallest combined degradations across both metrics.

This behaviour stems from its modelling of the normal data distribution in feature space. By constructing a memory bank of normal samples, PatchCore identifies anomalies as deviations from this learned manifold. Under perturbation, both defects and hotspot-like artefacts deviate from the normal distribution, leading to elevated anomaly scores and improved pixel-level separability, while maintaining relatively stable global discrimination.

The relatively low variance further indicates consistent performance across different perturbation instances, suggesting reduced sensitivity to specific disturbance patterns. However, this robustness comes at the cost of reduced specificity, as the model responds similarly to true defects and perturbation-induced artefacts, limiting its ability to precisely distinguish physically meaningful anomalies from spurious responses.

RACON and ZVSM, in contrast, achieve lower P-AUROC values (0.6357 ± 0.0905 and 0.5757 ± 0.0631 , respectively), indicating a limitation in fine-grained localisation. These methods emphasise structural consistency in feature space and produce smoother anomaly maps, which suppresses noise-induced artefacts but leads to less sharply defined boundaries. This limitation is further influenced by the physical nature of thermographic data, where heat diffusion and subsurface attenuation inherently blur defect boundaries.

These observations are further supported by the qualitative results shown in Fig. 12. It can be observed that intensity-based and

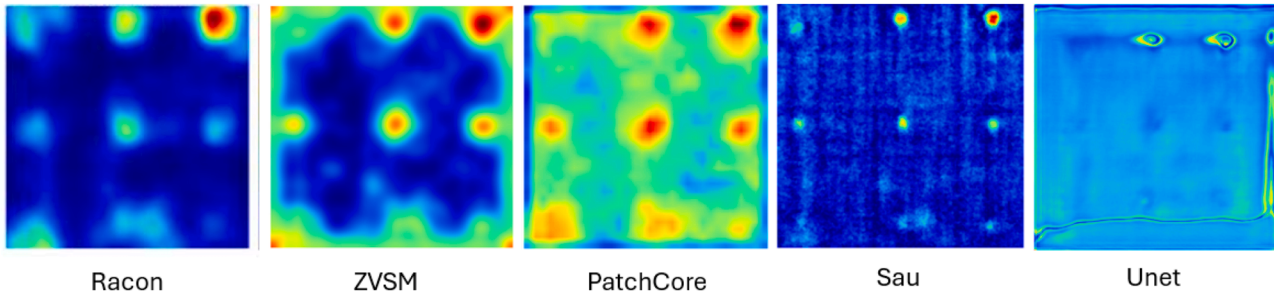


Fig. 12. Qualitative comparison of anomaly score maps under optical perturbation.

normality-deviation methods (e.g., Sauvola and PatchCore) respond strongly to hotspot artefacts, whereas RACON produces smoother and more globally consistent responses, focusing on structural deviations rather than local intensity fluctuations.

Overall, these results highlight a critical distinction between global discrimination and local sensitivity. A high P-AUROC under perturbation does not necessarily indicate accurate defect localisation but may instead reflect sensitivity to local intensity variations or deviations from normal behaviour. In contrast, I-AUROC provides a more reliable indicator of robustness under distribution shift.

Importantly, the introduced optical-like perturbations pose a significant challenge to existing methods. Fully supervised models are particularly vulnerable, as their performance strongly depends on the training data distribution and degrades severely under unseen perturbations. Unsupervised and zero-shot approaches also face difficulties, as perturbation-induced artefacts can distort feature representations or be misinterpreted as anomalies, leading to unstable responses.

5.4. Limitations

It must be acknowledged that our evaluation is limited to single thermographic data with relatively uniform defect types. The experimental dataset primarily contains one class of material (carbon-fiber composites) and defect patterns (e.g., delamination) that are fairly consistent in shape and thermal signature. Although this allowed us to demonstrate the effectiveness of the method under controlled conditions, it does not cover the full diversity of potential anomalies in real-world scenarios. In more complex settings, defects might take varied forms (cracks, irregular shape damage, multi-scale textures) or occur on backgrounds that are less uniform than the ones tested here. And all inputs in this study were from a single imaging modality (infrared thermography); many practical inspection systems may fuse information from multiple modalities (for example, combining thermal and visible-light data) or deal with environments where the “normal” background itself has significant texture or clutter. We therefore caution that further research is needed to assess the generalisability of the RCK approach. Future studies should evaluate its performance on broader datasets that feature various defects of different shapes and sizes, as well as explore its integration with other sensor data. Such investigations will help determine if any modifications or additional training signals are required to maintain high accuracy when moving beyond the relatively simple, homogeneous conditions addressed in the current work.

Finally, these findings prompt reflection on the broader landscape of anomaly detection techniques, suggesting that the choice of backbone complexity should be aligned with the variability of the target domain. In highly complex or high-variability domains (for instance, natural images or unconstrained scenes with many object types), advanced deep networks with extensive learned parameters are often indispensable to capture the rich range of patterns that define “normal” data. However, for structured, low-variability domains like the thermographic inspections studied here, we find that even classic non-parametric approaches can be surprisingly effective. Simpler methods, such as adaptive thresh-

old (e.g., the Sauvola method), already leverage the uniform background assumption and can detect obvious deviations with minimal computation, albeit usually with lower accuracy or specificity. Our RCK-based approach can be viewed as a hybrid that bridges these two extremes. It preserves the lightweight, training-free nature of traditional techniques while incorporating the representational power of convolutional feature extraction to improve detection performance. This middle-ground strategy demonstrates that one need not always resort to a massive deep model to achieve reliable results—especially if the problem domain inherently constrains the complexity of anomalies. The success of the RCK method thus encourages further work on efficient yet generalizable anomaly detectors. In future research, one could imagine extending this concept by combining random feature extractors with modest learning components or fine-tuning to handle moderately more complex scenarios without sacrificing efficiency. In summary, our study highlights that the most advanced backbone is not always necessary and that intelligently designed lightweight models can suffice (and even excel) in specialised domains, pointing the way toward anomaly detection solutions that are resource-efficient and broadly applicable.

According to the experimental results, our methods show a superior performance in runtime metrics and are much lighter than methods built upon deep neural networks, while maintaining a comparable detection performance. In contrast to extensive training and heavy hierarchical architecture, our method can deliver a significant computational resource saving up to 100x memory reduction and with 3 times faster inference.

6. Conclusion

In this work, we have presented a lightweight and efficient framework for zero-shot anomaly detection in thermographic imagery, leveraging Random Convolution Kernels (RCKs) as a computationally frugal alternative to deep pre-trained backbones. Motivated by the data scarcity and high annotation costs inherent in thermographic non-destructive testing, our approach eliminates the need for any training data by constructing a statistical reference model directly from the test image itself. By extending the RCK concept to multi-channel inputs via group convolutions and incorporating a multi-scale feature extraction strategy grounded in scale-space theory, our method effectively captures the subtle thermal deviations indicative of defects against a typically homogeneous background.

Experimental results on a dataset of CFRP laminates demonstrate that the proposed RCK-based model achieves detection performance comparable to fully supervised methods like U-Net and significantly outperforms other unsupervised baselines at the image level, while maintaining a competitive pixel-level AUROC(94.00%). Crucially, this high accuracy is attained with a dramatic reduction in computational overhead: our model exhibits orders of magnitude lower memory consumption, model size, alongside a threefold improvement in inference latency compared to zero-shot methods relying on heavy networks such as WideResNet-50.

These findings underscore a key insight: the relative simplicity and homogeneity of thermographic data often render complex, learned

feature extractors unnecessary. Our RCK framework successfully bridges the gap between the efficiency of classical thresholding techniques and the representational power of deep learning, offering a practical, out-of-the-box solution for resource-constrained industrial inspection systems.

Although our evaluation has been limited to a particular type of defect in composite materials, the proposed framework suggests several promising avenues for future investigation.

First, we plan a systematic assessment on diverse thermographic datasets and inspection modalities to more thoroughly evaluate the method's generalisation capability.

Second, we will examine how RACON can be combined with established thermographic preprocessing approaches to study potential synergies between analytical signal reconstruction and stochastic feature projection.

Finally, extending the framework to address more challenging conditions, such as heterogeneous backgrounds and multi-modal sensing, is a crucial step toward deployment in real-world industrial environments.

CRedit authorship contribution statement

Shaoyang Peng: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization; **Haoxuan Deng:** Writing – original draft, Methodology, Formal analysis, Conceptualization; **Sri Addepalli:** Writing – review & editing, Methodology; **Maryam Farsi:** Writing – review & editing.

Data Availability Statement

The raw data used for this study can be found at <https://doi.org/10.57996/cran.ceres-2828>

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

This work was funded by the EPSRC platform grant (grant number UKRI1890 Gtr).

Acknowledgment

This work was undertaken as part of a Ph.D. programme. Acknowledgments are extended to the Centre for Digital Engineering and Manufacturing, Cranfield University.

References

- [1] A.A. Sarawade, N.N. Charniya, Infrared thermography and its applications: a review, in: 2018 3rd International Conference on Communication and Electronics Systems (ICCES), 2018, pp. 280–285. <https://doi.org/10.1109/CESYS.2018.8723875>
- [2] R.H. Bossi, G.E. Georgeson, 16 - Nondestructive Testing of aerospace composites, in: P. Irving, C. Soutis (Eds.), Polymer Composites in the Aerospace Industry (Second Edition), Woodhead Publishing Series in Composites Science and Engineering, Woodhead Publishing, 2020, pp. 461–489. <https://doi.org/10.1016/B978-0-08-102679-3.00016-2>
- [3] R. Yang, Y. He, Optically and non-optically excited thermography for composites: a review, *Infrared Phys. Technol.* 75 (2016) 26–50. <https://doi.org/10.1016/j.infrared.2015.12.026>
- [4] R. Yang, Y. He, Eddy current pulsed phase thermography considering volumetric induction heating for delamination evaluation in carbon fiber reinforced polymers, *Appl. Phys. Lett.* 106 (23) (2015) 234103. <https://doi.org/10.1063/1.4922524>
- [5] Y. Kang, B. Gao, W.L. Woo, R. Chen, W. Peng, Thermal diffusion recast as energy-density transport: irrotational velocity field, boundary conservation, and depth-time scaling, *Int. J. Heat Mass Transf.* 266 (2026) 128824. <https://doi.org/10.1016/j.ijheatmasstransfer.2026.128824>
- [6] Y. Kang, L. Liu, B. Gao, J. Li, Y. Zeng, W. Lok Woo, Automated thermography cognitive sensing-Feedback inspection for large irregular sample, *IEEE Trans. Instrum. Meas.* 73 (2024) 1–10. <https://doi.org/10.1109/TIM.2024.3481540>
- [7] R.H. Bossi, *ASNT Industry Handbook: Aerospace NDT*, American Society for Nondestructive Testing, 2014.
- [8] Y. He, B. Deng, H. Wang, L. Cheng, K. Zhou, S. Cai, F. Ciampa, Infrared machine vision and infrared thermography with deep learning: a review, *Infrared Phys. Technol.* 116 (2021) 103754. <https://doi.org/10.1016/j.infrared.2021.103754>
- [9] A. Waqas, M.T. Araj, Machine learning-aided thermography for autonomous heat loss detection in buildings, *Energy Convers. Manag.* 304 (2024) 118243. <https://doi.org/10.1016/j.enconman.2024.118243>
- [10] F. Ciampa, P. Mahmoodi, F. Pinto, M. Meo, Recent advances in active infrared thermography for non-destructive testing of aerospace components, *Sensors* 18 (2) (2018). <https://www.mdpi.com/1424-8220/18/2/609>
- [11] A. Hossam, H.M. Harb, H.M. Abd El Kader, Automatic image segmentation method for breast cancer analysis using thermography, *JES. Journal of Engineering Sciences* 46 (1) (2018) 12–32.
- [12] G.K. Balakrishnan, C.T. Yaw, S.P. Koh, T. Abedin, A.A. Raj, S.K. Tiong, C.P. Chen, A review of infrared thermography for condition-based monitoring in electrical energy: applications and recommendations, *Energies* 15 (16) (2022). <https://doi.org/10.3390/en15166000>
- [13] A. Alshehri, D. AlSaeed, Breast cancer diagnosis in thermography using pre-Trained VGG16 with deep attention mechanisms, *Symmetry* 15 (3) (2023). <https://doi.org/10.3390/sym15030582>
- [14] Y. Liu, Y. Yao, F. Wang, S. Sfarra, K. Liu, Review of unsupervised machine learning methods in active infrared thermography for defect detection and analysis, *Quant. InfraRed Thermogr. J.* 0 (0) (2025) 1–28. <https://doi.org/10.1080/17686733.2025.2540662>
- [15] F. Wang, Z. Jiang, J. Jiang, Y. Yao, Y. Liu, Spatio-temporal feature extraction for infrared thermal imaging by three-dimensional convolutional autoencoder, *Quant. InfraRed Thermogr. J.* 0 (0) (2025) 1–14. <https://doi.org/10.1080/17686733.2025.2592191>
- [16] M.U. Jibril, B. Gao, W.L. Woo, G. Tian, N.A. Khan, R.S. Zakariyya, A.J. Muhammad, Physics-aware state space network with uncertainty quantification for automated defect detection in infrared NDT thermography, *NDT Int.* 158 (2026) 103586. <https://doi.org/10.1016/j.ndteint.2025.103586>
- [17] V.P. Vavilov, D.D. Burleigh, Review of pulsed thermal NDT: physical principles, theory and data processing, *NDT Int.* 73 (2015) 28–52. <https://doi.org/10.1016/j.ndteint.2015.03.003>
- [18] T. Aota, L.T.T. Tong, T. Okatani, Zero-Shot versus many-Shot: unsupervised texture anomaly detection, in: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2023, pp. 5564–5572.
- [19] Y. He, B. Deng, H. Wang, L. Cheng, K. Zhou, S. Cai, F. Ciampa, Infrared machine vision and infrared thermography with deep learning: a review, *Infrared Phys. Technol.* 116 (2021) 103754. <https://doi.org/10.1016/j.infrared.2021.103754>
- [20] N. Saeed, N. King, Z. Said, M.A. Omar, Automatic defects detection in CFRP thermograms, using convolutional neural networks and transfer learning, *Infrared Phys. Technol.* 102 (2019) 103048. <https://doi.org/10.1016/j.infrared.2019.103048>
- [21] Q. Fang, C. Ibarra-Castaneda, I. Garrido, Y. Duan, X. Maldague, Automatic detection and identification of defects by deep learning algorithms from pulsed thermography data, *Sensors* 23 (9) (2023). <https://doi.org/10.3390/s23094444>
- [22] Z. Tong, L. Cheng, S. Xie, M. Kersemans, A flexible deep learning framework for thermographic inspection of composites, *NDT Int.* 139 (2023) 102926. <https://doi.org/10.1016/j.ndteint.2023.102926>
- [23] S. Peng, S. Addepalli, M. Farsi, Machine learning in thermography non-Destructive testing: a systematic review, *Appl. Sci.* 15 (17) (2025). <https://doi.org/10.3390/app15179624>
- [24] J. Xue, B. Gao, G. Liu, Y. Zhang, W.L. Woo, Y. Yang, Y. Yu, Memory linked knowledge domain transfer few-shot learning for thermography nondestructive evaluation system, *Infrared Phys. Technol.* 132 (2023) 104718. <https://doi.org/10.1016/j.infrared.2023.104718>
- [25] Y. Shi, B. Gao, G. Yang, H. Li, W. Lok Woo, Industrial anomaly detection system: a multiclass algorithm leveraging feature information and memory bank, *IEEE Trans. Instrum. Meas.* 74 (2025) 1–9. <https://doi.org/10.1109/TIM.2025.3551852>
- [26] B. Hu, B. Gao, W.L. Woo, L. Ruan, J. Jin, Y. Yang, Y. Yu, A lightweight spatial and temporal multi-Feature fusion network for defect detection, *IEEE Trans. Image Process.* 30 (2021) 472–486. <https://doi.org/10.1109/TIP.2020.3036770>
- [27] J. Sauvola, M. Pietikäinen, Adaptive document image binarization, *Pattern Recognit.* 33 (2) (2000) 225–236. [https://doi.org/10.1016/S0031-3203\(99\)00055-2](https://doi.org/10.1016/S0031-3203(99)00055-2)
- [28] S. Zagoruyko, N. Komodakis, Wide Residual Networks, 2017. <https://arxiv.org/abs/1605.07146>
- [29] J. Jeong, Y. Zou, T. Kim, D. Zhang, A. Ravichandran, O. Dabeer, WinCLIP: zero-/few-Shot anomaly classification and segmentation, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 19606–19616.
- [30] Y. Cao, X. Xu, C. Sun, Y. Cheng, Z. Du, L. Gao, W. Shen, Segment Any Anomaly without Training via Hybrid Prompt Regularization, 5, 2023. <https://doi.org/10.48550/arXiv.2305.10724>
- [31] A. Dempster, F. Petitjean, G.I. Webb, ROCKET: Exceptionally fast and accurate time series classification using random convolutional kernels, *Data Min. Knowl. Discov.* 34 (5) (2020) 1454–1495. <https://doi.org/10.1007/s10618-020-00701-z>
- [32] A. Coates, A. Ng, H. Lee, An analysis of single-Layer networks in unsupervised feature learning, *J. Mach. Learn. Res. - Proc. Track* 15 (2011) 215–223.
- [33] I. Garcia Vargas, H. Fernandes, Thermal Inspection Dataset for Defect Segmentation in CFRP Laminates, 2025. <https://doi.org/10.17632/jrsb4b9yy5.1>

- [34] X.P.V. Maldague, *Theory and practice of infrared technology for nondestructive testing*, 2001. <https://api.semanticscholar.org/CorpusID:108423453>.
- [35] C. Ibarra-Castaneda, X. Maldague, Pulsed phase thermography reviewed, *Quant. InfraRed Thermogr. J.* 1 (1) (2004) 47–70. <https://doi.org/10.3166/qirt.1.47-70>
- [36] T. Fawcett, An introduction to ROC analysis, *Pattern Recognit. Lett.* 27 (8) (2006) 861–874. ROC Analysis in Pattern Recognition, <https://doi.org/10.1016/j.patrec.2005.10.010>
- [37] J.A. Hanley, The meaning and use of the area under a receiver operating characteristic (ROC) curve, *Radiology* 143 (1982) 29–36. <https://doi.org/10.1148/radiology.143.1.7063747>
- [38] P. Bergmann, M. Fauser, D. Sattlegger, C. Steger, MVTec AD - A comprehensive real-World dataset for unsupervised anomaly detection, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 9584–9592. <https://doi.org/10.1109/CVPR.2019.00982>
- [39] S.A. Taghanaki, Y. Zheng, S. Kevin Zhou, B. Georgescu, P. Sharma, D. Xu, D. Comaniciu, G. Hamarneh, Combo loss: handling input and output imbalance in multi-organ segmentation, *Comput. Med. Imaging Graph.* 75 (2019) 24–33. <https://doi.org/10.1016/j.compmedimag.2019.04.005>
- [40] K. Roth, L. Pemula, J. Zepeda, B. Schölkopf, T. Brox, P.V. Gehler, Towards total recall in industrial anomaly detection, *CoRR abs/2106.08265* (2021). <https://arxiv.org/abs/2106.08265>
- [41] J. Erazo-Aux, H. Loaiza-Correa, A.D. Restrepo-Girón, C. Ibarra-Castaneda, X. Maldague, Defect detection and depth estimation in composite materials for pulsed thermography images by nonuniform heating correction and oriented gradient information, *Materials* 16 (8) (2023). <https://doi.org/10.3390/ma16082998>



DSPACE

<https://dspace.org/>

Training-free thermographic anomaly detection using random convolution kernels

Peng, Shaoyang; Deng, Haoxuan; Addepalli, Sri; Farsi, Maryam

2026-06

Elsevier

<https://doi.org/10.1016/j.rineng.2026.111123>

Peng S, Deng H, Addepalli S, Farsi M. (2026) Training-free thermographic anomaly detection using random convolution kernels. Results in Engineering, Volume 30, June 2026, Article number 111123