Wave-MambaAD: Wavelet-driven State Space Model for Multi-class Unsupervised Anomaly Detection

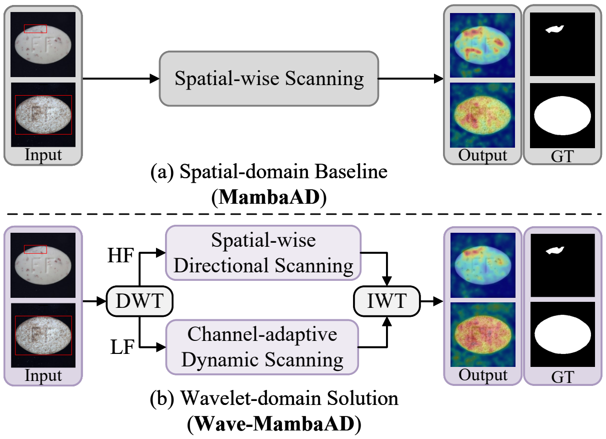
Qiao Zhang Mingwen Shao* Xinyuan Chen Xiang Lv Kai Xu
Shandong Key Laboratory of Intelligent Oil & Gas Industrial Software, Qingdao Institute of Software,
College of Computer Science and Technology, China University of Petroleum (East China)
bz24070008@s.upc.edu.cn, smw278@126.com, b24070011@s.upc.edu.cn, lvxiang1997@126.com,
s23070014@s.upc.edu.cn

Abstract

The Mamba model excels in anomaly detection through efficient long-range dependency modeling and linear complexity. However, Mamba-based anomaly detectors still face two critical challenges: (1) insufficient modeling of diverse local features leading to inaccurate detection of subtle anomalies; (2) spatial-wise scanning mechanism disrupting the spatial continuity of large-scale anomalies, resulting in incomplete localization. To address these challenges, we propose ***Wave-MambaAD****, a wavelet-driven state space model for unified subtle and large-scale anomaly detection. Firstly, to capture subtle anomalies, we design a high-frequency state space model that employs horizontal, vertical, and diagonal scanning mechanisms for processing directionally aligned high-frequency components, enabling precise anomaly detection through multidimensional feature extraction. Secondly, for comprehensive localization of large-scale anomalies, we propose a low-frequency state space model implementing channel-adaptive dynamic scanning mechanisms to maintain structural coherence in global contexts, which facilitates large-scale anomaly detection via adaptive feature integration. Finally, we develop a dynamic spatial enhancement block to improve anomalous feature representation by enhancing feature diversity through coordinated inter-channel communication and adaptive gating mechanisms. Comprehensive experiments on benchmark anomaly detection datasets show that Wave-MambaAD achieves competitive performance at lower parameters and computational costs.*

1. Introduction

Visual anomaly detection (AD) aims to identify abnormal areas in images that deviate from normal patterns. AD technology helps prevent potential risks and improve safety, and therefore has a wide range of applications in industrial defect detection [5], medical image diagnosis [44], video surveillance [7], autonomous driving [36], and agricultural spot detection [29, 41, 42], etc.


Input
Spatial-wise Scanning
Output
GT
HF
LF
DWT
Spatial-wise Directional Scanning
Channel-adaptive Dynamic Scanning
IWT
Input
Output
GT

(a) Spatial-domain Baseline (**MambaAD**)

(b) Wavelet-domain Solution (**Wave-MambaAD**)

Figure 1. **Motivation and effectiveness of our method.** Compared with the spatial-domain baseline (**MambaAD**), our approach (**Wave-MambaAD**) employs wavelet-domain modeling: directional scanning for high-frequency components to capture subtle anomalies, and channel-adaptive dynamic scanning for low-frequency components to fully localize large-scale anomalies. 'HF' and 'LF' denote the high- and low-frequency, respectively.

Previous AD methods have predominantly relied on manual labeling and extensive annotated data, a paradigm proven impractical in real-world scenarios due to the inherent scarcity of labeled anomalies. This limitation has propelled the emergence of unsupervised anomaly detection (UAD) methods, which operate without labeled data and have attracted substantial research interest. Current UAD methods are broadly classified into three categories: *augmentation-based*, *embedding-based*, and *reconstruction-based* methods. Augmentation-based methods [24, 37] address anomaly sample scarcity through

[*] Corresponding author

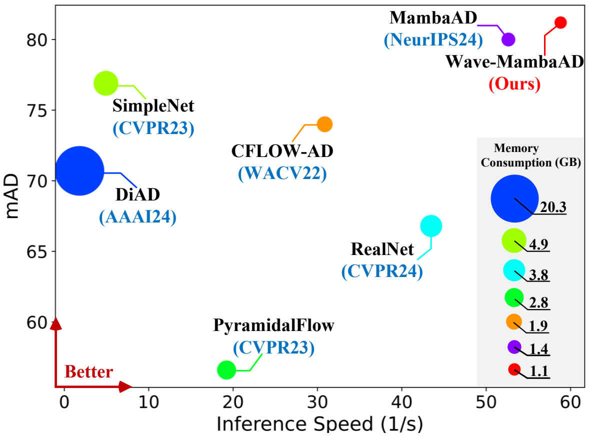


Figure 2. **Comparison with other SoTA methods regarding performance and efficiency.** Our method can achieve competitive performance and inference speed. Here, 'mAD' denotes the average performance of the three datasets, and the circle size indicates the memory consumption during inference.

data transformations and synthetic anomaly generation. However, suboptimal augmentation strategies may distort anomalous feature learning. Embedding-based methods [13, 18, 28] project normal data into low-dimensional manifolds for analysis, nevertheless, their effectiveness comes at the cost of demanding substantial computational resources. Reconstruction-based methods [6, 15, 35] have recently gained prominence owing to their robust performance and scalability. These methods model normal data distributions, identifying anomalies by quantifying reconstruction errors. For instance, RD4AD [6] establishes a single-class reconstruction framework via reverse distillation. However, it necessitated training individual models per class, leading to high computational costs.

To overcome this limitation, UniAD [35] has pioneered a unified multi-class reconstruction approach that maintains competitive performance while drastically reducing resource consumption, thereby validating the feasibility of multi-class reconstruction in cost-sensitive environments. Currently, most multi-class reconstruction methods rely on convolutional neural networks (CNNs) [6], Transformers [39], GANs [20] or Diffusion Models [15] to construct models. However, CNNs struggle with long-range dependency modeling due to their localized receptive fields, Transformers suffer from quadratic computational complexity in self-attention operations, and GANs or Diffusion Models face challenges in convergence stability and training efficiency. These limitations impede their deployment in large-scale real-time applications.

Recently, Mamba [10] has gained attention for its strong performance and linear computational efficiency. He et al. [14] proposed MambaAD, a Mamba-based anomaly detectors that introduces a Hilbert curve-based scanning strategy in the spatial domain to enhance contextual modeling, achieving competitive performance. Nevertheless, current Mamba-based anomaly detectors (e.g., MambaAD) still face two limitations. First, the limited ability to model diverse local features leads to inaccurate capture of subtle anomalies. Second, the spatial-wise scanning mechanism disrupts the spatial continuity of large-scale anomalies, leading to incomplete localization. As indicated in Figure 1 (a), MambaAD struggles to locate subtle anomalies and large-scale anomalies.

To tackle these limitations, we propose a wavelet-driven state space anomaly detector, dubbed **Wave-MambaAD**, that effectively localizes both subtle and large-scale anomalies (Figure 1(b)). Wave-MambaAD comprises a pre-trained encoder and a novel Wavelet-Mamba Decoder. Concretely, subtle anomalies manifest as weak local feature variations; however, Mamba's global modeling strength limits its sensitivity to such localized patterns. In contrast, the wavelet transform decomposes image features into directional high-frequency components (horizontal, vertical, and diagonal), which contain fine-grained texture details crucial for detecting subtle anomalies. Accordingly, we present a high-frequency state space (HFSS) model that applies directional scanning mechanisms (horizontal, vertical, and diagonal) to capture subtle anomalies from the corresponding high frequency.

Next, the low-frequency component not only reflects the global structure of image features but also captures uniform variations at large scales, both of which are crucial for detecting large-scale anomalies. Consequently, to leverage these properties, we propose a low-frequency state space (LFSS) model that employs a channel-adaptive dynamic scanning mechanism, thereby preserving global structural integrity and enhancing the detection of large-scale anomalies.

Finally, to further enhance anomaly feature representations, we designed a dynamic spatial enhancement (DSE) block that promotes inter-channel information flow and leverages adaptive gating mechanisms to enrich spatial features. Notably, HFSS, LFSS and wavelet transform are combined to form a wavelet-based dual state space (WDSS) block. WDSS and DSE are combined to form the Wavelet-Mamba Module, which is the core of the Wavelet-Mamba Decoder. We propose the Wave-MambaAD achieves competitive performance, demonstrating faster inference speeds and lower memory consumption (as illustrated in Figure 2).

Our main contributions can be summarized as:

- Wave-MambaAD combines wavelet transform and state space model to detect both subtle and large-scale anomalies via frequency-aware processing, achieving competitive performance with low complexity.
- A high-frequency state space model with directional

scanning is devised to capture subtle anomalies in multiple directions.
- A low-frequency state-space model with channel-adaptive scanning preserves global structure for detecting large-scale anomalies.
- A dynamic spatial enhancement block is proposed to enhance anomaly features through inter-channel interaction and adaptive gating.

2. Related Work

2.1. Unsupervised Anomaly Detection

Most UAD methods are broadly categorized into Augmentation-based, Embedding-based, and Reconstruction-based methods. Augmentation-based methods improve training by generating diverse anomaly data, like SimpleNet [24], which adds Gaussian noise to normal features, and DRAEM [38], which enhances training for anomalous features. However, these methods heavily rely on the design of augmentation strategies, and inappropriate augmentation may hinder the model's ability to learn true anomaly features. Meanwhile, embedding-based methods analyze normal sample embeddings to identify deviations, such as MDND [26] with Gaussian distributions, PatchCore [28] using memory banks, and CFLOW-AD [13] and PyramidFlow [18] with normalized flows. While effective, these methods require significant resources and expertise. Besides, reconstruction-based methods focus on self-trained encoders and decoders to reconstruct images, with commonly used reconstruction backbones including CNNs [27], Transformers [25], GANs [20], Diffusion Models [15], and Mamba [14]. However, generalisation of the model sometimes leads to inaccurate localization of anomalous regions.

2.2. Multi-class Anomaly Detection

Most AD methods are trained separately for each class, increasing time and memory overhead as the number of classes grows, and struggling with large intra-class diversity. Therefore, multi-class anomaly detection has become a research focus to address these challenges. For example, UniAD [35] introduced the multi-class paradigm with a unified framework, while DiAD [15] used a diffusion model with semantic guidance to ensure reconstructed image consistency. Furthermore, ViTAD [39] advanced the field with a pure Transformer architecture for better global and local anomaly detection. However, existing reconstruction-based methods—whether CNNs, Transformers, or diffusion models—have limitations: CNNs lack long-range modeling, Transformers are computationally expensive, and diffusion models are hard to optimize. Recently, MambaAD [14] addresses the above challenges by leveraging the advantages of Mamba's long-range modeling and linear complexity.

2.3. State Space Models

State space models (SSMs) [8, 10–12, 30] have gained attention for their performance in natural language processing. Although the structured state space sequence (S4) model [11] captures long-range dependencies through diagonal parameterization, it compresses all historical information, causing redundancy. Recently, Mamba [10] improves this by introducing an information selectivity mechanism, enabling efficient filtering of redundant data. Mamba's superior long-range modeling capabilities and linear computational efficiency have inspired extensive exploration of computer vision. For instance, VMamba [23] proposes a cross-scanning mechanism to address orientation sensitivity, sparking further research in fields such as medical imaging [22, 33, 34], remote sensing [3, 4, 32], and low-level vision [9, 19, 31, 43, 46]. In the field of AD, MambaAD [14] pioneered the application of Mamba to anomaly detection by designing a Hilbert-type scanning approach, achieving competitive performance.

However, MambaAD, which relies on spatial domain modeling, suffers from false or missed detections when handling multi-orientation subtle anomalies, while fixed spatial scanning disrupts the structural continuity of large-scale anomalies. Instead, our wavelet domain-based detector addresses these limitations through dual mechanisms: (1) We design the directional scanning mechanism to process high-frequency components to capture subtle anomalies in specific directions; (2) We use a channel-adaptive dynamic scanning mechanism to process low-frequency components to maintain global spatial coherence for complete localization of large-scale anomalies.

3. Method

3.1. Preliminaries

Recently, state space models such as S4 have emerged, inspired by linear time-invariant systems. In these models, a one-dimensional input function or sequence $x(t) \in \mathbb{R}$ is transformed into an output $y(t) \in \mathbb{R}$ via a hidden state $h(t) \in \mathbb{R}^N$, where the hidden state dynamics are governed by a linear ordinary differential equation (ODE):

$$\begin{cases} h'(t) = \mathbf{A}h(t) + \mathbf{B}x(t), \\ y(t) = \mathbf{C}h(t) + \mathbf{D}x(t). \end{cases} \tag{1}$$

where N denotes the state size, $\mathbf{A} \in \mathbb{R}^{N \times N}$, $\mathbf{B} \in \mathbb{R}^{N \times 1}$, $\mathbf{C} \in \mathbb{R}^{1 \times N}$, and $\mathbf{D} \in \mathbb{R}$.

Following this, the continuous-time system was embedded within deep-learning models by introducing a timescale parameter Δ, which enables the discretization of the continuous parameters $\mathbf{A}$ and $\mathbf{B}$ using the zero-order hold (ZOH) technique. This conversion yields the discrete pa-

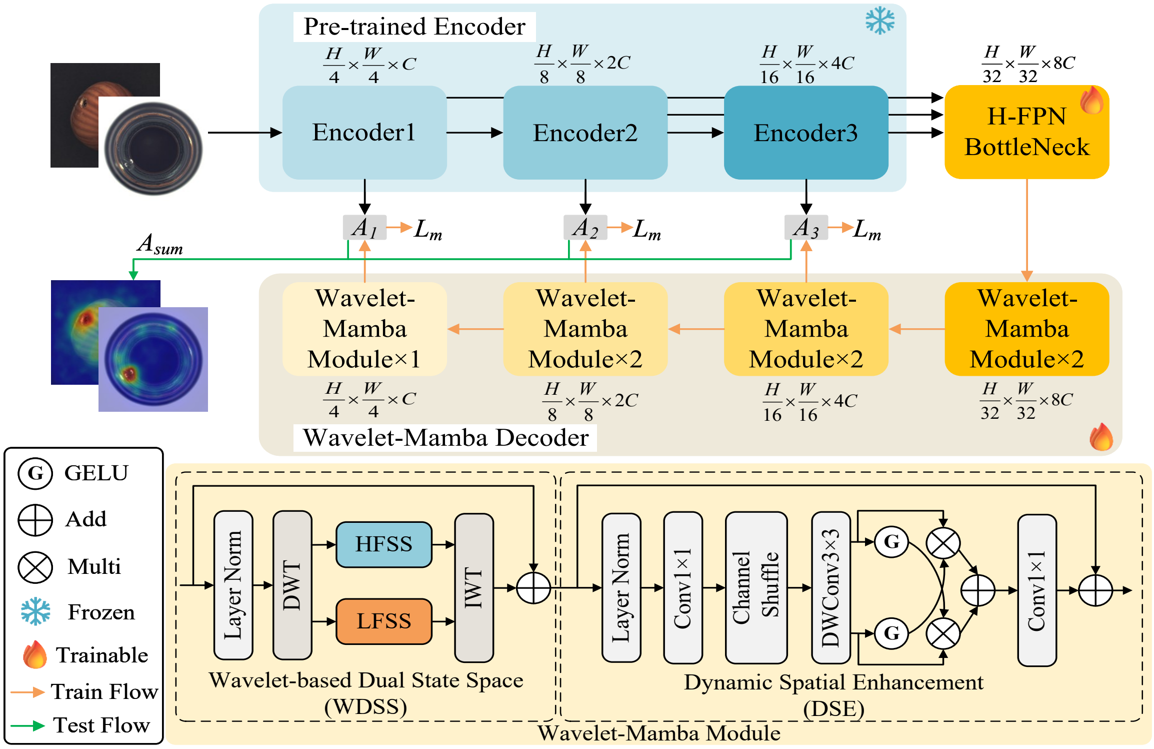


Figure 3. **Overview of the proposed Wave-MambaAD.** The framework adopts a pyramid autoencoder to reconstruct multi-scale features via the proposed Wavelet-Mamba Module. Each module comprises a wavelet-driven dual state space (WDSS) block for capturing both subtle and large-scale anomalies, as well as a dynamic spatial enhancement (DSE) block for refining the representations of anomalies. In WDSS, 'DWT' and 'IWT' refer to discrete wavelet transform and inverse wavelet transform, respectively.

rameters $\overline{\mathbf{A}}$ and $\overline{\mathbf{B}}$.

$$\begin{cases} \overline{\mathbf{A}} = exp(\mathbf{\Delta A}), \\ \overline{\mathbf{B}} = (\mathbf{\Delta A})^{-1}(exp(\mathbf{\Delta A}) - \mathbf{I})(\mathbf{\Delta B}). \end{cases} \tag{2}$$

After discretisation, Eq.1 can be represented as:

$$\begin{cases} h(t) = \overline{\mathbf{A}}h_{t-1} + \overline{\mathbf{B}}x_t, \\ y(t) = \mathbf{C}h_t + \mathbf{D}x(t). \end{cases} \tag{3}$$

The final output can be obtained directly through full convolution computation.

However, the aforementioned process parameters are static across different inputs, which restricts their adaptability. To overcome this limitation, Mamba incorporates a scanning mechanism alongside data-dependent, learnable parameters $\mathbf{\Delta}$, $\overline{\mathbf{B}}$, and $\mathbf{C}$, allowing the model to dynamically adjust its learning context in response to the specific characteristics of the input.

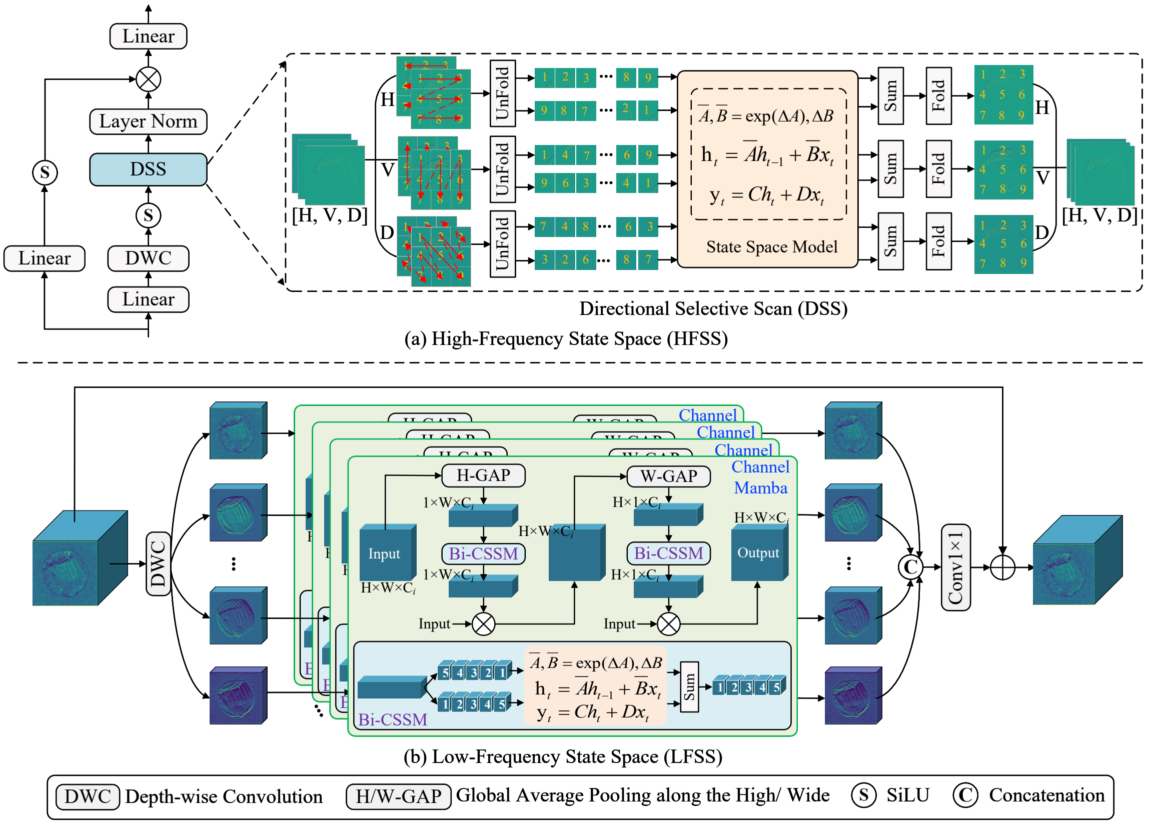
3.2. Overview

Figure 3 shows the Wave-MambaAD framework for multi-class anomaly detection. During training, a pre-trained encoder (ResNet-34 [16]) extracts multi-scale features, which are fused in the Half-FPN bottleneck [21] and then passed to a Wave-Mamba Decoder for reconstruction. The final loss function is the sum of the mean squared error (L_m) computed across feature maps at three scales. During inference, the anomaly map is generated by summing the cosine similarities of features across scales.

Within the Wavelet-Mamba Decoder, we introduce the Wavelet-Mamba Module, which consists of a wavelet-based dual state space (WDSS) block and a dynamic spatial enhancement (DSE) block. Specifically, the WDSS block comprises a high-frequency state space (HFSS) model and a low-frequency state space (LFSS) model. The HFSS model processes the high-frequency component to capture subtle anomalies, while the LFSS model processes the low-frequency component to fully localize large-scale anomalies. Finally, the DSE block is attached to the WDSS to further enhance the anomaly feature representation.

3.3. Wavelet-based Dual State Space

While Mamba-based anomaly detectors demonstrate promising performance, however, there are two limitations: (1) Inadequate modelling of diverse local features leads to inaccurate detection of subtle defects; and (2) the spatial-wise scanning tends to compromise the continuity of large-



(a) High-Frequency State Space (HFSS)

(b) Low-Frequency State Space (LFSS)

Figure 4. **Overview of the proposed HFSS and LFSS.** The HFSS employs directional selective scanning mechanisms to capture subtle anomalies in high-frequency features. LFSS utilizes channel-wise selective scanning mechanisms to preserve spatial structural integrity, effectively detecting large-scale anomalies. In (a), ‘H’, ‘V’, and ‘D’ are the horizontal, vertical, and diagonal high-frequency components.

scale anomalies causing incomplete detection. To address the above limitations, we propose a wavelet-based dual state space (WDSS) block that contains a high-frequency state space (HFSS) model and a low-frequency state space (LFSS) model. Concretely, HFSS captures subtle anomalies in the corresponding high-frequency subbands via a directional scanning mechanism (horizontal, vertical, and diagonal). Furthermore, LFSS maintains global spatial structural integrity and locates large-scale anomalies in their entirety through a channel-adaptive dynamic scanning mechanism. The detailed flow of WDSS is shown in Figure 3, and the whole process can be described as follows:

$$\begin{cases} F_h^{in}, F_l^{in} = \mathbf{DWT}(\mathbf{LN}(F_{in})), \\ F_h^{out} = \mathbf{HFSS}(F_h^{in}), \\ F_l^{out} = \mathbf{LFSS}(F_l^{in}), \\ F_{wdss} = \mathbf{IWT}(F_h^{out}, F_l^{out}) + F_{in}, \end{cases} \tag{4}$$

where $\mathbf{LN}(\cdot)$ denotes layer normalization, $\mathbf{DWT}(\cdot)$ and $\mathbf{IWT}(\cdot)$ represent Haar wavelet transform and Haar wavelet inverse transform, respectively. *The detailed formulations of* $\mathbf{DWT}(\cdot)$ *and* $\mathbf{IWT}(\cdot)$ *are provided in the supplementary materials*. $\mathbf{HFSS}(\cdot)$ and $\mathbf{LFSS}(\cdot)$ denote the high-frequency state space model and the low-frequency state space model, respectively. F_h denotes the high-frequency component, comprising horizontal, vertical, and diagonal high frequencies. F_l represents the low-frequency component. F_{wdss} is the output of WDSS.

High-Frequency State Space. The HFSS is illustrated in Figure 4 (a). The HFSS retains the architectural framework of VSSM [23], with the difference that we design a directional (horizontal, vertical, diagonal) selective scanning mechanism specifically optimized for processing direction-aligned high-frequency components, enabling precise capture of subtle anomaly signatures. The process of HFSS can be represented as:

$$\begin{cases} F_h^1 = \mathbf{LN}(\mathbf{DSS}(\mathbf{SiLU}(\mathbf{DWC}(\mathbf{Linear}(F_h^{in}))))), \\ F_h^2 = \mathbf{SiLU}(\mathbf{Linear}(F_h)), \\ F_h^{out} = \mathbf{Linear}(F_h^1 \cdot F_h^2), \end{cases} \tag{5}$$

where $\mathbf{DWC}(\cdot)$ denotes depth-wise convolution. $\mathbf{DSS}(\cdot)$

represents the directional selective scan (DSS). The DSS pipeline first serializes directional high-frequency components (H, V, D) into 1D sequences, applies corresponding directional scanning strategies, then employs discrete state space modeling to capture long-range dependencies, followed by summation-based fusion and dimensional reconstruction to recover the original 2D structure.

Low-Frequency State Space. The detailed architecture of LFSS is shown in Figure 4 (b). Specifically, we first apply depth-wise convolution to F_l^{in}, subsequently conduct channel grouping. Each partitioned group is then processed through Mamba blocks with channel-adaptive scanning mechanisms to model long-range dependencies. Finally, the processed groups are concatenated and the output feature is collectively generated via residual connections. The overall process of LFSS can be represented as:

$$\begin{cases} F_l^1, F_l^2, ..., F_l^n = \mathbf{Split}(\mathbf{DWC}(F_l^{in})), \\ F_l^1 = \mathbf{CMamba}(F_l^1), \\ F_l^2 = \mathbf{CMamba}(F_l^2), \\ ..., \\ F_l^n = \mathbf{CMamba}(F_l^n), \\ F_l^{out} = \mathbf{Conv_{1\times1}}\left(\left[F_l^1, F_l^2, ..., F_l^n\right]\right) + F_l^{in}, \end{cases} \tag{6}$$

where $\mathbf{Split}(\cdot)$, and $[\cdot]$ denote split and concatenation operations, respectively. $\mathbf{CMamba}(\cdot)$ denotes the designed Mamba block with channel-wise selective scanning (Channel Mamba). The pipeline of Channel Mamba is shown in Figure 4 (b). Specifically, we compress spatial features using the axis pooling operation and then model long-distance dependencies between channels via bidirectional channel SSM (Bi-CSSM). Axis pooling reduces computational costs while maintaining certain spatial features. The whole process can be represented as:

$$\begin{cases} F' = \mathbf{CSSM_{Bi}}(\mathbf{GAP_H}(Input)) \cdot Input, \\ Output = \mathbf{CSSM_{Bi}}(\mathbf{GAP_W}(F^{'})) \cdot Input, \end{cases} \tag{7}$$

where $\mathbf{GAP_H}(\cdot)$ and $\mathbf{GAP_W}(\cdot)$ denote the global average pooling along height and width, respectively. $\mathbf{CSSM_{Bi}}(\cdot)$ is the bidirectional channel SSM.

3.4. Dynamic Spatial Enhancement

While the WDSS architecture decouples high-frequency and low-frequency modelling in the wavelet domain through a state space model (allowing for the simultaneous detection of both subtle defects and large-scale anomalies), its frequency-centric representation exhibits inherent limitations in preserving spatial structure continuity. To bridge this representation gap, we propose a dynamic spatial enhancement (DSE) block—a lightweight yet effective component that enhances the spatial structure through coordinated inter-channel communication pathways and adaptive gating mechanisms, thereby improving anomaly feature representation. The detailed architecture of DSE is illustrated in Figure 3, and its process can be formally described by the following equation:

$$\begin{cases} F_d = \mathbf{DWConv}(\mathbf{CS}(\mathbf{Conv_{1\times1}}(\mathbf{LN}(F_{wdss})))), \\ F_d^1, F_d^2 = \mathbf{Split}(F_d), \\ F_d^1 = \mathbf{GELU}(F_d^2) \cdot F_d^1, \\ F_d^2 = \mathbf{GELU}(F_d^1) \cdot F_d^2, \\ F_{out} = \mathbf{Conv_{1\times1}}(F_d^1 + F_d^2) + F_{wdss}, \end{cases} \tag{8}$$

where $\mathbf{CS}(\cdot)$ denotes the channel shuffle operation. The DSE complements spatial structure information and further enhances the representation of anomaly features.

3.5. Anomaly Segmentation and Classification.

Anomaly segmentation provides a pixel-level anomaly score map to locate anomalies. The multi-scale anomaly maps A_i (i=1, 2, 3) from constrained features are summed to form the final anomaly map A_{sum}. For anomaly classification, we perform average pooling on A_{sum} and use its maximum value as the image-level score to determine if the image is anomalous.

4. Experiments and Analysis

4.1. Implementation Details

Our framework adopts a pre-trained ResNet-34 [16] as the feature extractor and a designed Wavelet-Mamba Decoder as the student network, where the Wavelet-Mamba Module employs a [1, 2, 2, 2] layer configuration. During both training and inference phases, input images are uniformly resized to 256^2 resolution. The optimization process utilizes the Adam optimizer with an initial learning rate of 0.005 and weight decay of 0.0001, executed 100 training epochs on an NVIDIA RTX A6000. The loss function is the mean square error (MSE) at multiple scales. During inference, the anomaly map is generated by summing cosine similarities across scales.

4.2. Datasets and Metrics

To evaluate Wave-MambaAD, we conduct experiments on three anomaly detection datasets: MVTec-AD [1], VisA [47], and MPDD [17]. Following previous work [14], we evaluate the model's performance using the following metrics: Area Under the Receiver Operating Characteristic Curve (AU-ROC), Average Precision (AP) [38], F1-score-max (F1_max) [47], and Area Under the Per-Region-Overlap (AU-PRO) [2]. Furthermore, we compute the average of all metrics (**mAD**) to measure the overall performance of the model. *A more detailed description is provided in the supplementary materials.*

Datasets	Method	Image-level			Pixel-level				mAD
		AU-ROC	AP	F1_max	AU-ROC	AP	F1_max	AU-PRO	
MVTec-AD [1]	CFLOW-AD [13]	91.6	96.7	93.4	95.7	45.9	48.6	88.3	80.0
	SimpleNet [24]	95.9	98.6	96.0	97.0	49.3	51.9	87.3	82.3
	PyramidalFlow [18]	70.2	85.5	85.5	80.0	22.3	22.0	47.5	59.0
	RealNet [45]	88.5	95.0	91.8	76.2	40.6	38.7	62.7	70.5
	DiAD [15]	97.2	99.0	96.5	96.8	52.6	55.5	90.7	84.0
	MambaAD [14]	97.8	**99.3**	97.3	97.4	55.1	57.6	**93.4**	85.4
	Wave-MambaAD	**98.2**	**99.3**	**97.4**	**97.5**	**55.9**	**58.3**	93.1	**85.6**
VisA [47]	CFLOW-AD [13]	86.5	88.8	84.9	97.7	33.9	37.2	86.8	73.7
	SimpleNet [24]	88.8	91.2	85.4	97.1	35.0	38.4	83.2	74.1
	PyramidalFlow [18]	58.2	66.3	74.4	77.0	7.2	9.6	42.8	47.9
	RealNet [45]	80.8	85.1	78.7	74.0	25.7	29.9	45.7	59.9
	DiAD [15]	86.8	88.3	85.1	96.0	26.1	33.0	75.2	70.1
	MambaAD [14]	**94.3**	94.5	89.4	98.5	39.4	**44.0**	91.0	78.3
	Wave-MambaAD	**94.4**	**94.8**	**90.1**	**98.6**	**40.1**	43.8	**91.4**	**79.0**
MPDD [17]	CFLOW-AD [13]	75.7	80.1	81.7	96.8	26.3	28.0	89.5	68.3
	SimpleNet [24]	88.4	92.0	87.9	96.5	32.0	34.6	89.0	74.3
	PyramidalFlow [18]	73.6	77.0	79.4	94.1	21.1	17.8	77.2	62.9
	RealNet [45]	85.1	90.2	88.3	83.3	36.1	39.6	68.1	70.1
	DiAD [15]	68.3	77.9	80.1	90.4	10.9	13.1	66.1	58.1
	MambaAD [14]	88.7	93.2	90.8	97.5	33.6	38.1	92.3	76.3
	Wave-MambaAD	**92.5**	**94.9**	**91.0**	**98.0**	**41.5**	**42.2**	**93.8**	**79.1**

Table 1. **Quantitative comparison with other cutting-edge multi-class anomaly detection methods.** The best results are denoted using **bold** and the second-best results are underlined.

Method	Params(M)	FLOPs(G)	Memory(MB)	Speed	mAD
CFLOW-AD	237.0	28.7	1892	30.9	74.0
SimpleNet	72.8	17.7	4946	4.9	76.9
PyramidalFlow	34.3	962.0	2836	19.2	56.6
RealNet	591.0	115.0	3794	43.5	66.8
DiAD	1331.0	451.5	20306	1.8	70.7
MambaAD	25.7	8.3	1484	52.6	80.0
Ours	**22.3**	**7.5**	**1142**	**58.8**	**81.2**

Table 2. **Comparison of efficiency with SoTA methods.** 'Memory' denotes the memory consumption during inference. The best results are shown in **bold**.

4.3. Comparison with SoTAs on AD Datasets

We compare the proposed method with SoTAs anomaly detection methods, including MambaAD [14], DiAD [15], RealNet [45], PyramidalFlow [18], SimpleNet [24], and CFLOW-AD [13]. *Notably, some of the results of the above methods are obtained from ADer [40].*

Quantitative Comparison. As shown in Table 1, our method achieves better performance on the MVTec-AD dataset across all metrics except AU-PRO, outperforming the current SoTA method (MambaAD). On the challenging VisA dataset, our method slightly underperforms MambaAD in pixel-level F1_max, yet achieves the best performance in all other metrics, with an overall improvement of 0.7 in the average metric (mAD). Notably, our method performs better on the real industrial dataset MPDD, which also proves the robustness of our method. Additionally, we present a comparison of the proposed Wave-MambaAD with other methods in terms of parameters, complexity, memory consumption, and inference speed in Table 2. Compared to other SoTA methods, our method has fewer parameters and complexity, faster inference speed, and less memory consumption. *More quantitative comparisons are in the supplementary materials.*

Qualitative Comparison. Figure 5 presents a qualitative comparison between the proposed method and other SoTA methods. Our method demonstrates superior accuracy in detecting subtle defects compared to the current SoTA method (MambaAD). Specifically, as illustrated by the examples of 'Pill', 'Pub', and 'Capsules' in Figure 5, our method achieves more precise defect localization, whereas MambaAD tends to mislocalize anomalies. Furthermore, for large-scale defects such as 'Cable' and 'Total_rust', our method provides a more complete and accurate localization of anomalous region.

4.4. Ablation Studies

Table 3 presents a series of ablation experiments conducted on the MVTec-AD dataset to validate the effectiveness of the proposed components.

Effectiveness of Wavelet-Mamba Module. As shown in Table 3 (a) and (f), replacing the Wavelet-Mamba module with HSS from MambaAD resulted in a performance decline, which confirms the effectiveness of our method and highlights the advantages of combining Mamba with wavelet transform for anomaly detection.

Effectiveness of High-Frequency State Space model. As illustrated in (b) and (f) of Table 3, removing the HFSS results in a performance decline, validating its effectiveness. Furthermore, we evaluate the importance of directional scanning in HFSS, as shown in (d) and (f) of Table 3. Replacing the directional scanning (horizontal, vertical, and diagonal) with uniform scanning in a single direction (horizontal or vertical) leads to a degradation in model performance, which further demonstrates the effectiveness of us-

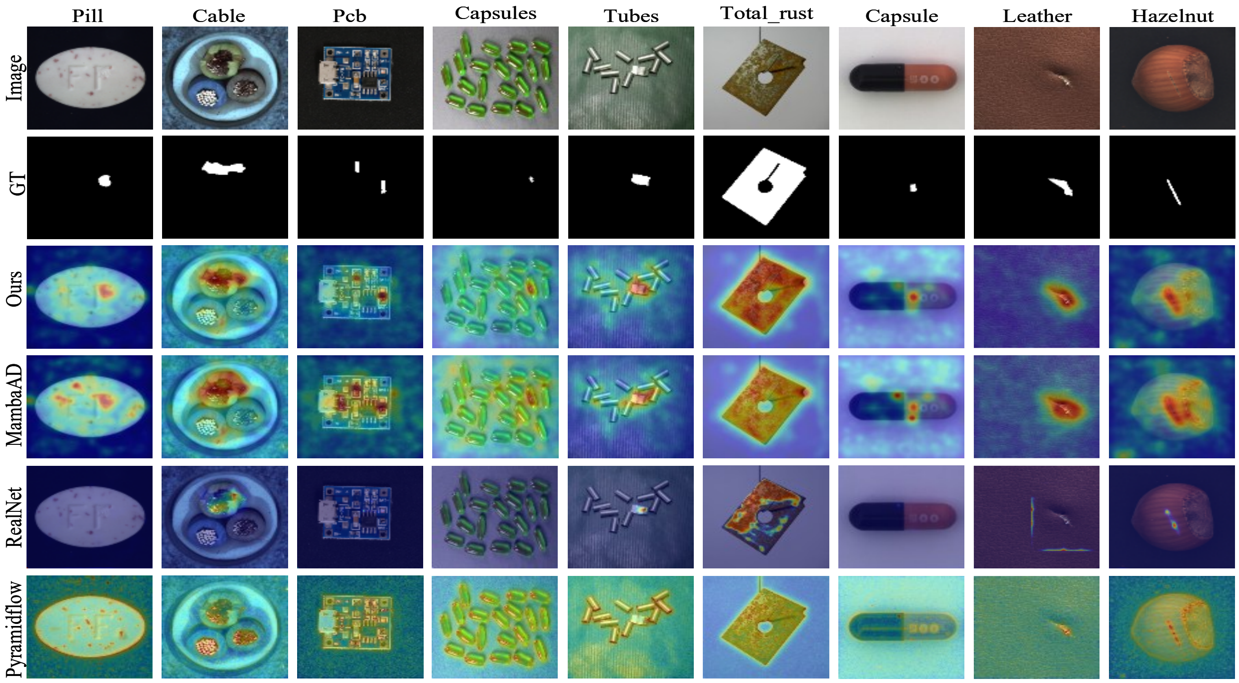


Figure 5. **Qualitative visualization of pixel-level anomaly segmentation on anomaly datasets.** Our method can accurately and completely locate anomalies.

Settings		Image-level			Pixel-level				mAD
		AU-ROC	AP	F1_max	AU-ROC	AP	F1_max	AU-PRO	
(a)	WM→HSS	97.5	98.8	96.3	96.7	52.5	56.3	91.3	84.2
(b)	w/o HFSS w/o LFSS	97.7 98.0	99.0 99.1	96.6 96.9	96.9 96.9	52.0 53.2	55.6 56.6	91.4 91.6	84.2 84.6
(c)	DSE→MLP	97.9	99.1	96.8	97.0	54.5	57.3	91.8	84.9

(d)	HFSS(Horizontal scan)
	HFSS(Vertical scan)
	HFSS(w/o Diagonal scan)
		97.9
		97.8
		98.1
			99.2
			99.1
			99.3
				97.1
				96.8
				97.4
					97.2
					97.1
					97.3
						54.8
						54.1
						55.4
							57.9
							56.8
							58.0
								92.7
								91.8
								92.8
									85.0
									85.3
									85.5

(e)	LFSS(Spatial-wise)	98.0	99.0	96.9	97.1	54.0	57.8	91.7	84.9
(f)	All	**98.2**	**99.3**	**97.4**	**97.5**	**55.9**	**58.3**	**93.1**	**85.6**

Table 3. **Ablation studies of the proposed components.** 'WM' denotes the Wavelet-Mamba Module. 'HSS' denotes the Hybrid State Space Block in MambaAD[14]. '→' is substitution. The best results are shown in **bold**.

ing a directional scanning strategy for high-frequency.

Effectiveness of Low-Frequency State Space model. By comparing (b) and (g) in Table 3, the model's overall performance decreases by 1.0 without the LFSS, confirming its effectiveness. Additionally, we replace the channel scanning in LFSS with spatial scanning, as in (e) in Table 3, and the model's performance decreases, indicating that the channel-wise scanning is more effective for large-scale anomalies.

Effectiveness of Dynamic Space Enhancement block. To validate the effectiveness of the DSE, we replace it with a multi-layer perception (MLP) network. As shown in (c) and (f) of Table 3, this results in a 0.7 drop in overall performance, highlighting the importance of supplementing spatial structure information extracted in the frequency domain.

5. Conclusion

In this work, we propose a wavelet-driven state space anomaly detector (Wave-MambaAD) to tackle the multi-class anomaly detection task. Our approach integrates three innovative components: (1) a high-frequency state space model with directional scanning to capture subtle defects; (2) a low-frequency state space model with channel-adaptive dynamic scanning to preserve global structure for accurate localization of large-scale anomalies; and (3) a dynamic spatial enhancement block that strengthens anomaly representations through inter-channel interaction and adaptive gating. Experiments show that Wave-MambaAD is a competitive anomaly detector.

Acknowledgements

This work was supported by the National Key Research and Development Program of China (Grant No. 2021YFA1000102), the National Natural Science Foundation of China (Grant Nos. 62376285 and 61673396), and Natural Science Foundation of Shandong Province, China (Grant No. ZR2022MF260).

References

[1] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Mvtec ad–a comprehensive real-world dataset for unsupervised anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9592–9600, 2019. 6, 7

[2] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4183–4192, 2020. 6

[3] Hongruixuan Chen, Jian Song, Chengxi Han, Junshi Xia, and Naoto Yokoya. Changemamba: Remote sensing change detection with spatio-temporal state space model. *arXiv preprint arXiv:2404.03425*, 2024. 3

[4] Keyan Chen, Bowen Chen, Chenyang Liu, Wenyuan Li, Zhengxia Zou, and Zhenwei Shi. Rsmamba: Remote sensing image classification with state space model. *IEEE Geoscience and Remote Sensing Letters*, 2024. 3

[5] Yajun Chen, Yuanyuan Ding, Fan Zhao, Erhu Zhang, Zhangnan Wu, and Linhao Shao. Surface defect detection methods for industrial products: A review. *Applied Sciences*, 11(16): 7657, 2021. 1

[6] Hanqiu Deng and Xingyu Li. Anomaly detection via reverse distillation from one-class embedding. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9737–9746, 2022. 2

[7] Omar Elharrouss, Noor Almaadeed, and Somaya Al-Maadeed. A review of video surveillance systems. *Journal of Visual Communication and Image Representation*, 77: 103116, 2021. 1

[8] Daniel Y Fu, Tri Dao, Khaled K Saab, Armin W Thomas, Atri Rudra, and Christopher Ré. Hungry hungry hippos: Towards language modeling with state space models. *arXiv preprint arXiv:2212.14052*, 2022. 3

[9] Hu Gao, Bowen Ma, Ying Zhang, Jingfan Yang, Jing Yang, and Depeng Dang. Learning enriched features via selective state spaces model for efficient image deblurring. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 710–718, 2024. 3

[10] Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*, 2023. 2, 3

[11] Albert Gu, Karan Goel, and Christopher Ré. Efficiently modeling long sequences with structured state spaces. *arXiv preprint arXiv:2111.00396*, 2021. 3

[12] Albert Gu, Isys Johnson, Karan Goel, Khaled Saab, Tri Dao, Atri Rudra, and Christopher Ré. Combining recurrent, convolutional, and continuous-time models with linear state space layers. *Advances in neural information processing systems*, 34:572–585, 2021. 3

[13] Denis Gudovskiy, Shun Ishizaka, and Kazuki Kozuka. Cflow-ad: Real-time unsupervised anomaly detection with localization via conditional normalizing flows. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 98–107, 2022. 2, 3, 7

[14] Haoyang He, Yuhu Bai, Jiangning Zhang, Qingdong He, Hongxu Chen, Zhenye Gan, Chengjie Wang, Xiangtai Li, Guanzhong Tian, and Lei Xie. Mambaad: Exploring state space models for multi-class unsupervised anomaly detection. *arXiv preprint arXiv:2404.06564*, 2024. 2, 3, 6, 7, 8

[15] Haoyang He, Jiangning Zhang, Hongxu Chen, Xuhai Chen, Zhishan Li, Xu Chen, Yabiao Wang, Chengjie Wang, and Lei Xie. A diffusion-based framework for multi-class anomaly detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 8472–8480, 2024. 2, 3, 7

[16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 4, 6

[17] Stepan Jezek, Martin Jonak, Radim Burget, Pavel Dvorak, and Milos Skotak. Deep learning-based defect detection of metal parts: evaluating current methods in complex conditions. In *2021 13th International congress on ultra modern telecommunications and control systems and workshops (ICUMT)*, pages 66–71. IEEE, 2021. 6, 7

[18] Jiarui Lei, Xiaobo Hu, Yue Wang, and Dong Liu. Pyramidflow: High-resolution defect contrastive localization using pyramid normalizing flow. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14143–14152, 2023. 2, 3, 7

[19] Xiaoyan Lei, Wenlong ZHang, and Weifeng Cao. Dvmsr: Distillated vision mamba for efficient super-resolution. *arXiv preprint arXiv:2405.03008*, 2024. 3

[20] Yufei Liang, Jiangning Zhang, Shiwei Zhao, Runze Wu, Yong Liu, and Shuwen Pan. Omni-frequency channel-selection representations for unsupervised anomaly detection. *IEEE Transactions on Image Processing*, 2023. 2, 3

[21] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2117–2125, 2017. 4

[22] Jiarun Liu, Hao Yang, Hong-Yu Zhou, Yan Xi, Lequan Yu, Cheng Li, Yong Liang, Guangming Shi, Yizhou Yu, Shaoting Zhang, et al. Swin-umamba: Mamba-based unet with imagenet-based pretraining. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 615–625. Springer, 2024. 3

[23] Yue Liu, Yunjie Tian, Yuzhong Zhao, Hongtian Yu, Lingxi Xie, Yaowei Wang, Qixiang Ye, and Yunfan Liu. Vmamba: Visual state space model. *arXiv preprint arXiv:2401.10166*, 2024. 3, 5

[24] Zhikang Liu, Yiming Zhou, Yuansheng Xu, and Zilei Wang. Simplenet: A simple network for image anomaly detection

and localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20402–20411, 2023. 1, 3, 7

[25] Jonathan Pirnay and Keng Chai. Inpainting transformer for anomaly detection. In *International Conference on Image Analysis and Processing*, pages 394–406. Springer, 2022. 3

[26] Oliver Rippel, Patrick Mertens, and Dorit Merhof. Modeling the distribution of normal data in pre-trained deep features for anomaly detection. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 6726–6733. IEEE, 2021. 3

[27] Nicolae-Cătălin Ristea, Neelu Madan, Radu Tudor Ionescu, Kamal Nasrollahi, Fahad Shahbaz Khan, Thomas B Moeslund, and Mubarak Shah. Self-supervised predictive convolutional attentive block for anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 13576–13586, 2022. 3

[28] Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Schölkopf, Thomas Brox, and Peter Gehler. Towards total recall in industrial anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14318–14328, 2022. 2, 3

[29] Abdullah Ali Salamai, Nouran Ajabnoor, Waleed E Khalid, Mohammed Maqsood Ali, and Abdulaziz Ali Murayr. Lesion-aware visual transformer network for paddy diseases detection in precision agriculture. *European Journal of Agronomy*, 148:126884, 2023. 1

[30] Jimmy TH Smith, Andrew Warrington, and Scott W Linderman. Simplified state space layers for sequence modeling. *arXiv preprint arXiv:2208.04933*, 2022. 3

[31] Jiangwei Weng, Zhiqiang Yan, Ying Tai, Jianjun Qian, Jian Yang, and Jun Li. Mamballie: Implicit retinex-aware low light enhancement with global-then-local state space. *arXiv preprint arXiv:2405.16105*, 2024. 3

[32] Yi Xiao, Qiangqiang Yuan, Kui Jiang, Yuzeng Chen, Qiang Zhang, and Chia-Wen Lin. Frequency-assisted mamba for remote sensing image super-resolution. *arXiv preprint arXiv:2405.04964*, 2024. 3

[33] Zhaohu Xing, Tian Ye, Yijun Yang, Guang Liu, and Lei Zhu. Segmamba: Long-range sequential modeling mamba for 3d medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 578–588. Springer, 2024. 3

[34] Zhongxing Xu, Feilong Tang, Zhe Chen, Zheng Zhou, Weishan Wu, Yuyao Yang, Yu Liang, Jiyu Jiang, Xuyue Cai, and Jionglong Su. Polyp-mamba: Polyp segmentation with visual mamba. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 510–521. Springer, 2024. 3

[35] Zhiyuan You, Lei Cui, Yujun Shen, Kai Yang, Xin Lu, Yu Zheng, and Xinyi Le. A unified model for multi-class anomaly detection. *Advances in Neural Information Processing Systems*, 35:4571–4584, 2022. 2, 3

[36] Ekim Yurtsever, Jacob Lambert, Alexander Carballo, and Kazuya Takeda. A survey of autonomous driving: Common practices and emerging technologies. *IEEE access*, 8:58443–58469, 2020. 1

[37] Vitjan Zavrtanik, Matej Kristan, and Danijel Skocaj. Draem - a discriminatively trained reconstruction embedding for surface anomaly detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 8330–8339, 2021. 1

[38] Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Draem-a discriminatively trained reconstruction embedding for surface anomaly detection. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 8330–8339, 2021. 3, 6

[39] Jiangning Zhang, Xuhai Chen, Yabiao Wang, Chengjie Wang, Yong Liu, Xiangtai Li, Ming-Hsuan Yang, and Dacheng Tao. Exploring plain vit features for multi-class unsupervised visual anomaly detection. *Available at SSRN 4866147*, . 2, 3

[40] Jiangning Zhang, Haoyang He, Zhenye Gan, Qingdong He, Yabiao Wang, Chengjie Wang, Lei Xie, Yong Liu, et al. A comprehensive library for benchmarking multi-class visual anomaly detection. . 7

[41] Qiao Zhang, Yanliang Ge, Cong Zhang, and Hongbo Bi. Tprnet: camouflaged object detection via transformer-induced progressive refinement network. *The Visual Computer*, 39 (10):4593–4607, 2023. 1

[42] Qiao Zhang, Xiaoxiao Sun, Yurui Chen, Yanliang Ge, and Hongbo Bi. Attention-induced semantic and boundary interaction network for camouflaged object detection. *Computer Vision and Image Understanding*, 233:103719, 2023. 1

[43] Qiao Zhang, Mingwen Shao, and Xinyuan Chen. Poolmamba: Pooling state space model for low-light image enhancement. *Neurocomputing*, 635:130005, 2025. 3

[44] Shaoting Zhang and Dimitris Metaxas. On the challenges and perspectives of foundation models for medical image analysis. *Medical image analysis*, 91:102996, 2024. 1

[45] Ximiao Zhang, Min Xu, and Xiuzhuang Zhou. Realnet: A feature selection network with realistic synthetic anomaly for anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16699–16708, 2024. 7

[46] Zhuoran Zheng and Chen Wu. U-shaped vision mamba for single image dehazing. *arXiv preprint arXiv:2402.04139*, 2024. 3

[47] Yang Zou, Jongheon Jeong, Latha Pemula, Dongqing Zhang, and Onkar Dabeer. Spot-the-difference self-supervised pre-training for anomaly detection and segmentation. In *European Conference on Computer Vision*, pages 392–408. Springer, 2022. 6, 7