

Bridging the Geometry Mismatch: Frequency-Aware Anisotropic Serialization for Thin-Structure SSMs

Jin Bai^{1,2}, Huiyao Zhang^{1,2}, Qi Wen¹, Ningyang Li¹, Shengyang Li¹, Atta ur Rahman³, and Xiaolin Tian⁴

¹ Technology and Engineering Center for Space Utilization, Chinese Academy of Sciences, Beijing, China

² University of Chinese Academy of Sciences, Beijing, China

³ Geography and Geomatics, University of Peshawar, Peshawar, Pakistan

⁴ Faculty of Innovation Engineering, Macau University of Science and Technology, Macau, China

`baijin25@mailsucas.ac.cn, wenqi@csu.ac.cn`

Abstract. The segmentation of thin linear structures is inherently topology-critical, where minor local errors can sever long-range connectivity. While recent State-Space Models (SSMs) offer efficient long-range modeling, their isotropic serialization (e.g., raster scanning) creates a geometry mismatch for anisotropic targets, causing state propagation across rather than along the structure trajectories. To address this, we propose FGOS-Net, a framework based on frequency-geometric disentanglement. We first decompose features into a stable topology carrier and directional high-frequency bands, leveraging the latter to explicitly correct spatial misalignments induced by downsampling. Building on this calibrated topology, we introduce frequency-aligned scanning that elevates serialization to a geometry-conditioned decision, preserving direction-consistent traces. Coupled with an active probing strategy to selectively inject high-frequency details and suppress texture ambiguity, FGOS-Net consistently outperforms strong baselines across four challenging benchmarks. Notably, it achieves 91.3% mIoU and 97.1% cDice on DeepCrack while running at 80 FPS with only 7.87 GFLOPs.

Keywords: Thin Structure Segmentation · State-Space Models · Frequency-Geometric Disentanglement · Anisotropic Serialization

1 Introduction

Thin-structure segmentation presents a persistent challenge in computer vision due to the extreme anisotropy and topological sensitivity of targets such as cracks, vessels, and roads [16, 21, 24, 45].

Among these, cracks represent a particularly challenging case due to highly variable widths, severe clutter, and low signal-to-noise ratios, which significantly

 Corresponding author.

hinder long-range connectivity preservation. While this mismatch affects thin structures broadly, we focus on crack segmentation as an extreme testbed to rigorously isolate serialization failures from confounding factors.

State-Space Models (SSMs) have emerged as an efficient paradigm for long-range dependency modeling [6,7]. However, most SSM-based segmentation frameworks rely on isotropic serialization (e.g., raster scanning) to flatten 2D feature maps [25]. This assumption breaks for anisotropic linear geometry, creating a serialization-induced geometry mismatch: the recurrent state is forced to traverse orthogonally to the structure’s dominant axis, fragmenting what should be a single continuous trace into disjoint segments. While recent lightweight SSMs like SCSegamba [21] minimize parameters, their complex scan patterns limit throughput. We instead prioritize FLOPs and FPS as the true deployment bottleneck.

To bridge this gap, we propose FGOS-Net, a lightweight framework that aligns 1D state-space modeling with 2D anisotropic geometry through frequency–geometric disentanglement. Our goal is not to claim that Haar decomposition or Hilbert indexing is novel in isolation. Instead, we use frequency decomposition as a geometry-revealing interface: the low-frequency component provides a stable topology carrier, while directional high-frequency bands provide local structural cues for alignment and detail validation. Building on this decomposition, Frequency-Aligned Scanning (FA-Scan) assigns deterministic sub-band-aligned traversal trajectories, turning serialization from a fixed raster-like operation into a geometry-conditioned modeling decision. Complementarily, Active Spectral-Geometric Probing (ASGP) filters high-frequency details through topology-conditioned probe evolution, reducing texture leakage before detail injection. These components are integrated into a hierarchical encoder with a parameter-efficient bottleneck and a parallel multi-scale decoder.

Our contributions are summarized as follows:

- We identify serialization-induced geometry mismatch as a key bottleneck for applying SSMs to topology-critical thin-structure segmentation, where fixed scan orders can disrupt structure-consistent long-range propagation.
- We propose a frequency–geometric coupling framework that uses wavelet sub-bands to guide local topology alignment, geometry-aligned SSM serialization, and topology-conditioned high-frequency detail filtering.
- We provide controlled diagnostics showing that the gains mainly come from structure-aligned serialization and active topology-conditioned gating, rather than from branch diversity, parameter scaling, or generic scan variants.
- Extensive experiments on crack benchmarks and cross-domain thin-structure datasets demonstrate strong boundary integrity, connectivity preservation, and accuracy–efficiency trade-off.

2 Related Work

Anisotropic Linear Structure Segmentation. Extracting thin linear structures (*e.g.* cracks, vessels, roads) is inherently topology-dominated: even minor

pixel errors can sever long-range connectivity [28, 37]. Existing topological losses or skeleton-based constraints [37, 46] serve as supervisory regularizers but do not alter the feature modeling itself. Recent self-similarity approaches [13] leverage fractal statistics to capture structural recurrence, yet remain agnostic to serialization order. Architecturally, most CNN- and Transformer-based models operate with isotropic kernels or global attention, leaving them vulnerable to (i) connectivity discontinuity and (ii) texture leakage [1, 11, 29, 40, 43, 49].

State-Space Models and Serialization Mismatch. Visual SSMs achieve linear-complexity long-range modeling via selective scan [3, 6, 8, 14, 25, 36, 42, 44, 47], yet their serialization of 2D maps into 1D sequences is often data-agnostic [18, 19, 22, 23, 25, 26, 35]. For thin structures, such serialization can map adjacent points along a curvilinear target to distant 1D positions, causing connectivity breaks when the scan cuts across the dominant axis. Recent locality-preserving scans and frequency-aware SSM variants show traversal order and spectral decomposition are useful design axes. We therefore do not treat Hilbert indexing or wavelet decomposition as standalone novelties. Instead, our focus is the task-specific coupling between frequency-derived structural components and SSM serialization: directional sub-bands are used to choose traversal paths, while the low-frequency carrier anchors topology-conditioned detail filtering.

Frequency-Aware Learning and Gated Injection. Frequency decomposition (*e.g.*, Wavelet or Fourier) disentangles low-frequency semantics from high-frequency boundary cues [17], yet high-frequency bands encode both meaningful edges and spurious textures. Most wavelet-based networks adopt passive fusion (*e.g.*, concatenation), indiscriminately injecting all high-frequency content [20, 31, 38]. Our approach instead redefines the low-frequency component as a stable topology carrier that actively regulates detail injection via global topological agreement.

Efficient Architecture Design. Lightweight designs [34] sacrifice high-frequency detail through aggressive downsampling, particularly harmful for thin structures [9, 32]. Our framework combines stage-adaptive LightGate bottleneck (LGB) with parallel multi-scale fusion (GFA), preserving boundary detail without additional latency.

3 Method

3.1 Overall Architecture

Problem Formulation. Standard SSMs serialize 2D features via isotropic strategies (*e.g.*, raster), implicitly assuming spatial adjacency is preserved under a fixed 1D traversal [5, 42]. For anisotropic thin structures, this breaks down: adjacent points along the structure may map to distant 1D positions.

Core Principle: Split, Align, and Model. To correct this modeling-level geometry mismatch, FGOS-Net adopts a frequency-geometric disentanglement paradigm that decouples topology modeling from noisy boundary details. Each FGOS stage (Fig. 1, right) processes features X through four steps:

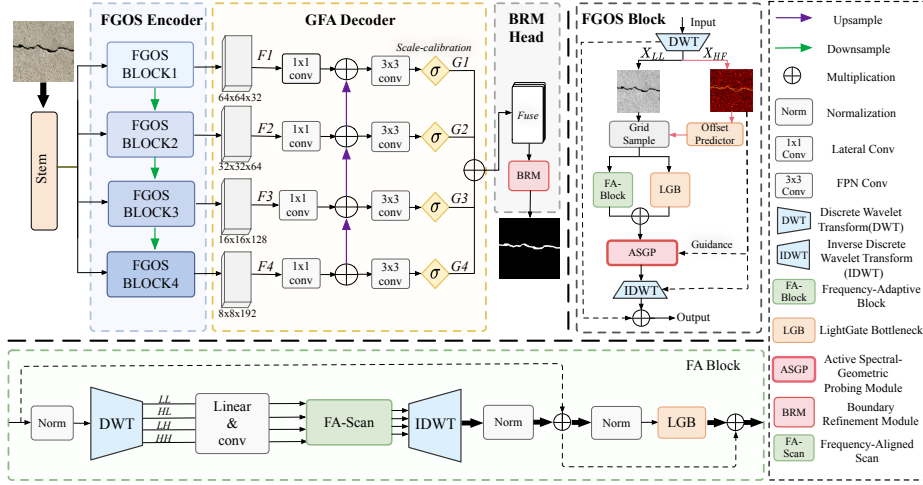


Fig. 1: Overview of FGOS-Net. Each FGOS block performs: (1) DWT-based frequency disentanglement into topology carrier X_{LL} and directional details X_{HF} ; (2) detail-driven alignment of X_{LL} ; (3) geometry-aligned FA-Scan for topology modeling and ASGP for topology-conditioned detail gating; (4) IDWT-based reconstruction. A parallel GFA decoder and BRM head fuse multi-scale features and refine boundaries.

Disentangle (DWT). We employ an invertible Haar DWT to physically separate the input into a low-frequency topology carrier (X_{LL}) and directional high-frequency bands:

$$\{X_{LL}, X_{LH}, X_{HL}, X_{HH}\} = \text{DWT}(X), \quad X_{HF} = \{X_{LH}, X_{HL}, X_{HH}\}. \quad (1)$$

We adopt the Haar wavelet for three practical reasons: (i) its 2×2 support provides a compact directional decomposition in which LH and HL respond strongly to horizontal and vertical transitions, respectively; (ii) the transform is exactly invertible, so the DWT/IDWT pair itself does not introduce approximation error; and (iii) its arithmetic cost is small enough for real-time inference. Importantly, DWT is applied to multi-channel feature maps rather than raw pixels: structural cues have already been distributed across channels by the convolutional stem, and IDWT restores the full spatial resolution within each stage.

Align (Geometric Prior). Crucially, we avoid directly fusing X_{HF} into the topology stream. Instead, we aggregate the directional bands to predict a light-weight deformation field $\Delta \mathbf{p}$, which spatially aligns the topology carrier X_{LL} to the standard coordinate grid $\mathbf{p}$ via **GridSample**:

$$\Delta \mathbf{p} = \mathcal{P}(\text{Concat}[X_{HF}]), \quad \tilde{X}_{LL} = \text{GridSample}(X_{LL}, \mathbf{p} + \Delta \mathbf{p}). \quad (2)$$

where $\mathcal{P}$ denotes a lightweight two-layer 3×3 convolutional projection predicting a 2-channel offset field. Here, misregistration refers to feature-level displacement caused by strided downsampling, not approximation error of Haar DWT/IDWT, which is exact and invertible. The HF bands provide local positional cues for correcting the LL topology carrier without directly injecting texture noise.

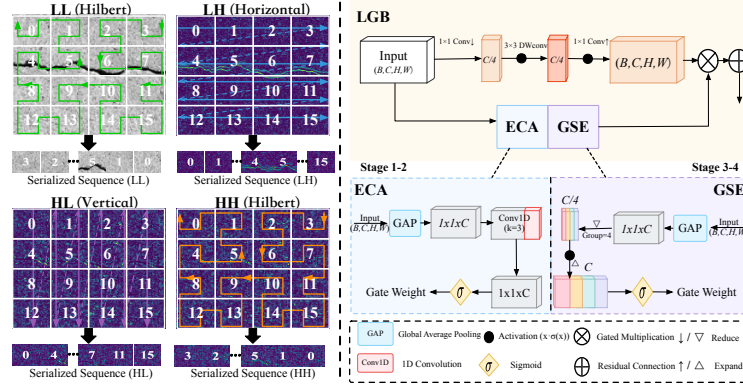


Fig. 2: FA-Block components. Left: FA-Scan assigns deterministic, sub-band-aligned serialization trajectories (Horizontal/Vertical for directional bands, Hilbert for isotropic bands). Right: The LightGate Bottleneck (LGB) employs a symmetric bottleneck ($C \rightarrow C/4 \rightarrow C$) modulated by stage-adaptive gating (ECA [41] / GSE [12]), reducing parameters by $\approx 94\%$ vs. standard FFNs. ECA and GSE denote Efficient Channel Attention and Grouped Squeeze-and-Excitation, respectively.

Model & Gate. The streams are then processed in parallel:

Topology Stream ($\tilde{X}_{LL}$): Processed by the FA-Block (Sec. 3.2), which assigns sub-band-aligned serialization trajectories (*e.g.*, Horizontal/Vertical/Hilbert) to preserve geometric continuity during SSM propagation.

Detail Stream (X_{HF}): Processed by ASGP (Sec. 3.3), which actively evolves probes on $\tilde{X}_{LL}$ to generate a spatial gate M , ensuring that only structure-consistent high-frequency details are preserved.

Merge (IDWT). Finally, the refined topology and gated details are recombined via Inverse DWT to restore the full-spectrum feature map for the next stage.

3.2 The FA-Block: Geometry-Aligned Modeling

FA-Scan: Sub-band-Aligned Serialization. To obtain geometry-aware trajectories without introducing global high-frequency noise, we perform a local Haar DWT exclusively on the aligned topology carrier $\tilde{X}_{LL}$, yielding four internal sub-bands: $\{\tilde{X}_{LL}^{ll}, \tilde{X}_{LL}^{lh}, \tilde{X}_{LL}^{hl}, \tilde{X}_{LL}^{hh}\} = \text{DWT}(\tilde{X}_{LL})$. This transforms the serialization from isotropic raster scanning into a sub-band conditioned task. Let $\tilde{X}_{LL}^{sub}$ denote these local features. We assign deterministic trajectories $\mathcal{S}_{sub}$ matching the dominant geometry of each band:

$$\mathcal{S}_{lh} = \mathcal{S}_{\text{Horz}}, \quad \mathcal{S}_{hl} = \mathcal{S}_{\text{Vert}}, \quad \mathcal{S}_{ll} = \mathcal{S}_{hh} = \mathcal{S}_{\text{Hilbert}}. \quad (3)$$

Here, $\mathcal{S}_{\text{Hilbert}}$ denotes the Hilbert space-filling curve [10]. Since IDWT reconstructs the signal by summing orthogonal components, structures at arbitrary orientations (*e.g.*, 45° cracks) are recovered through the complementary scanned sequences.

To maximize parameter efficiency, the serialized sequences are processed by a shared SSM operator Ψ (*e.g.*, VSS Block [25]), differing only in traversal order:

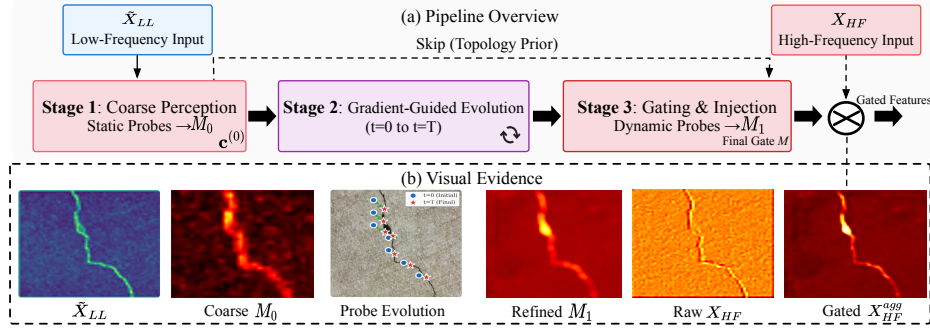


Fig. 3: Active Spectral-Geometric Probing (ASGP). The module resolves high-frequency ambiguity via: (i) Coarse Perception: Initializing probes on the topology carrier $\tilde{X}_{LL}$; (ii) Gradient-Guided Evolution: Iteratively refining probe positions ($t = 0 \rightarrow T$) via differentiable gradient ascent on the potential field M_0 ; (iii) Gating & Injection: Generating a topology-validated mask M to transform raw high-frequency details X_{HF} into gated details $X_{HF}^{agg} = X_{HF} \odot M$.

$$\hat{X}_{LL}^{sub} = \mathcal{S}_{sub}^{-1} \left(\Psi \left(\mathcal{S}_{sub}(\tilde{X}_{LL}^{sub}) \right) \right), \quad Y_{scan} = \text{IDWT}(\{\hat{X}_{LL}^{sub}\}). \quad (4)$$

Intuition: $\mathcal{S}_{\text{Horz}}$ ensures horizontal structures in $\tilde{X}_{LL}^{lh}$ remain contiguous in 1D. For isotropic bands ($\tilde{X}_{LL}^{ll}, \tilde{X}_{LL}^{hh}$), the Hilbert curve minimizes linearization distance, preserving 2D locality better than raster scans. Since $\mathcal{S}$ is a fixed indexing operation, FA-Scan adds no learnable parameters and uses precomputed Hilbert indexing.

LGB: Parameter-Efficient Selective Mixing. After FA-Scan, we use a lightweight bottleneck for efficient channel mixing (Fig. 2, Right). This module is not intended as the main conceptual contribution; it serves to keep the encoder compact after geometry-aligned serialization. It consists of a symmetric bottleneck path $\mathcal{F}(Y)$ ($C \rightarrow C/4 \rightarrow C$) and a stage-adaptive channel gate $\mathcal{G}_s(Y)$:

$$\mathcal{G}_s(Y) = \sigma \left(\begin{cases} \text{Conv1D}_k(\text{GAP}(Y)), & s \in \{1, 2\} \quad (\text{ECA}), \\ \mathbf{W}_2(\text{ReLU}(\mathbf{W}_1(\text{GAP}(Y))))), & s \in \{3, 4\} \quad (\text{grouped SE}), \end{cases} \right). \quad (5)$$

The output is $\text{LGB}(Y) = Y + \mathcal{G}_s(Y) \odot \mathcal{F}(Y)$, where $\mathcal{F}$ uses a 3×3 depthwise convolution between two pointwise projections. This design reduces parameters while preserving the accuracy–efficiency profile of the full model.

3.3 Active Spectral-Geometric Probing

Motivation: Topology-Conditioned Gating. Wavelet high-frequency bands (X_{HF}) contain both structural boundaries and background texture gradients. Naively injecting them into the semantic stream corrupts topology reasoning with noise. ASGP addresses this by formulating detail injection as a conditional gating problem: we actively localize structure-supporting regions on the stable topology carrier $\tilde{X}_{LL}$ to filter X_{HF} . Unlike passive attention, we employ an

Algorithm 1 Gradient-Guided Probe Evolution

Input: Coarse map M_0 , Feature $\tilde{X}_{LL}$, Init coords $\mathcal{C}^{(0)}$, Steps T
Hyperparams: Radius $\delta = 0.15$, $\lambda_1=0.1$, $\lambda_2=0.05$, $\epsilon = 1e^{-5}$

- 1: **for** $t = 0$ **to** $T - 1$ **do**
- 2: $\mathbf{f}_i^{(t)} \leftarrow \text{GridSample}(\tilde{X}_{LL}, \mathcal{C}^{(t)})$ ▷ Feature Extraction
- 3: $\Delta \mathbf{c}_{\text{sem}} \leftarrow \text{MLP}(\mathbf{f}_i^{(t)})$ ▷ Semantic Offset
- 4: $\mathbf{g}_i \leftarrow \nabla_{\mathcal{C}} \text{GridSample}(M_0, \mathcal{C}^{(t)})$ ▷ Differentiable Interpolation
- 5: Compute pairwise distances $d_{ij} = \|\mathbf{c}_i - \mathbf{c}_j\|_2$
- 6: $\mathcal{F}_{div}^{(t)} \leftarrow \sum_{j \neq i} \frac{\mathbf{c}_i - \mathbf{c}_j}{d_{ij} + \epsilon} \cdot \text{ReLU}\left(1 - \frac{d_{ij}}{\delta}\right)$ ▷ Diversity Constraint
- 7: $\mathcal{C}^{(t+1)} \leftarrow \text{Clamp}\left(\mathcal{C}^{(t)} + \Delta \mathbf{c}_{\text{sem}} + \lambda_1 \mathbf{g}_i + \lambda_2 \mathcal{F}_{div}^{(t)}, -1, 1\right)$
- 8: **end for**
- 9: **Return** refined coordinates $\mathcal{C}^{(T)}$

iterative optimization process to validate high-frequency evidence against global connectivity. Unlike deformable attention [2, 48], which learns offsets purely from data, ASGP guides probe migration via an explicit geometric potential field (M_0) and enforces spatial diversity through a physics-inspired repulsion constraint—properties absent in standard deformable mechanisms.

Stage 1: Coarse Perception. We initialize N static learnable probes $\{\mathbf{c}_i^{(0)}\}$ to query $\tilde{X}_{LL}$ with linear complexity $\mathcal{O}(NHW)$. These probes generate attention maps aggregated into a coarse potential field $M_0 = \sigma(\frac{1}{N} \sum_{i=1}^N \text{Attn}(\mathbf{q}_i, \tilde{X}_{LL}))$. High responses in M_0 indicate probable structural regions, serving as the initialization landscape for probe refinement. With $N=64 \ll H \times W$, the iterative evolution incurs negligible overhead versus dense attention.

Stage 2: Gradient-Guided Probe Evolution. We perform a sparse-to-dense evolution over T steps. The update rule combines semantic offsets with gradient guidance derived from the first-order topology map M_0 (via bilinear interpolation):

$$\mathbf{c}_i^{(t+1)} = \mathcal{P}_{\Omega}\left(\mathbf{c}_i^{(t)} + \Delta \mathbf{c}_{\text{sem}}^{(t)} + \lambda_1 \nabla_{\mathcal{C}} M_0(\mathbf{c}_i^{(t)}) + \lambda_2 \mathcal{F}_{div,i}^{(t)}\right). \quad (6)$$

Here, $\mathcal{P}_{\Omega}$ projects coordinates onto $[-1, 1]$. The update is driven by three terms:

1) Semantic Offset. A learnable shift $\Delta \mathbf{c}_{\text{sem}}$ predicted via MLP from local features $\mathbf{f}_i^{(t)}$, allowing data-driven adjustments. **2) Gradient Ascent.** We treat M_0 as a continuous potential field. The term $\nabla_{\mathcal{C}} M_0(\mathbf{c}_i)$ drives probes to "climb" the probability peaks of the topological structure. **3) Diversity Constraint.** Pure gradient ascent inherently drives probes towards a few global maxima (e.g., the widest crack segments), leading to spatial degeneracy. To counteract this, we introduce a truncated repulsion force $\mathcal{F}_{div} = \sum_{j \neq i} \frac{\mathbf{c}_i - \mathbf{c}_j}{\|\mathbf{c}_i - \mathbf{c}_j\|_2 + \epsilon} \cdot \max(0, 1 - \frac{\|\mathbf{c}_i - \mathbf{c}_j\|_2}{\delta})$. This truncated repulsion enforces a minimum separation δ , dispersing probes along curvilinear ridges of M_0 rather than clustering at high-contrast centroids. Since all probe coordinates operate in the normalized $[-1, 1]$ space, $\delta=0.15$ represents a fixed fraction of the feature map extent and is inherently resolution-invariant, requiring no manual adjustment across input scales or structure widths.

Stage 3: Gating and Injection. After T steps, the refined probe states are projected back to pixel space to form the fine-grained mask M_1 . The final spatial gate $M = \sigma(wM_1 + (1 - w)M_0)$ (with $w=0.5$) transforms raw high-frequency details into topology-validated details via $X_{HF}^{agg} = X_{HF} \odot M$, injecting high-frequency evidence only where it agrees with the coarse topology carrier.

3.4 Reconstruction and Refinement

Motivation: semantic–spatial mismatch across pyramid levels. Deep features are semantically strong but spatially coarse, while shallow features are spatially precise yet noisy. A sequential top-down decoder can accumulate smoothing errors. We instead use a parallel multi-scale fusion with explicit scale calibration, followed by a lightweight boundary refiner.

Gated Feature Aggregation (GFA). Given encoder features $\{F_i\}_{i=1}^4$, we project them to a shared embedding and upsample them to the spatial size of F_1 :

$$\tilde{F}_i = \mathcal{U}_{bilinear}(\phi_i(F_i), \text{size}(F_1)). \quad (7)$$

Unlike sequential FPN-style fusion, which may accumulate smoothing errors across levels, we calibrate each scale independently with a lightweight channel-wise scale-calibration gate. Let $\mathbf{W}_1$ and $\mathbf{W}_2$ be channel-reduction and channel-expansion layers. The gating vector $s_i \in \mathbb{R}^{1 \times C \times 1 \times 1}$ is computed as:

$$s_i = \sigma\left(\mathbf{W}_2 \cdot \text{ReLU}\left(\mathbf{W}_1 \cdot \text{AvgPool}(\tilde{F}_i)\right)\right), \quad G_i = s_i \odot \tilde{F}_i, \quad (8)$$

and the calibrated features are fused in one shot:

$$F_{\text{fused}} = \text{Conv}_{3 \times 3} \left(\sum_{i=1}^4 G_i \right). \quad (9)$$

Boundary Refinement Module (BRM). To counteract boundary smoothing from upsampling, BRM adopts a dual-branch residual design: a context branch (depthwise-separable convolutions) improves semantic consistency, and an edge branch (pure depthwise convolution) preserves high-frequency gradients. The two outputs are concatenated, projected to the original channel dimension, and added back to F_{fused} to sharpen boundaries without disrupting global topology.

4 Experiments

4.1 Experimental Setup

Datasets. We evaluate on four benchmarks: DeepCrack [24] (537 images) and Crack500 [45] (3,368 images) featuring low-contrast topology; CrackMap [16] (120 images) dominated by texture noise (*e.g.*, water or oil stains); and TUT [21] (1,408 images) spanning eight materially diverse environments—bitumen, cement, bricks, plastic runways, tiles, metal surfaces, generator blades, and underground pipelines—that introduce significant cross-scenario variation in texture,

Table 1: Quantitative comparison on four crack benchmarks. We report region metrics (mIoU/F1/Precision/Recall), boundary metric (ODS), and the connectivity-aware cDice [37]. Best results are highlighted in **bold**, and second-best are underlined.

Method	DeepCrack (%)						Crack500 (%)					
	mIoU	F1	Prec	Recall	ODS	cDice	mIoU	F1	Prec	Recall	ODS	cDice
RIND [30]	81.27	84.40	82.22	86.64	90.72	95.69	78.24	72.98	70.85	79.41	72.98	80.93
CT-CrackSeg [39]	84.25	88.53	90.55	88.77	91.63	<u>96.46</u>	77.81	74.06	75.65	72.53	74.38	<u>82.98</u>
SimCrack [15]	83.50	90.79	87.59	90.22	91.55	<u>96.28</u>	<u>78.76</u>	74.35	74.55	78.25	74.42	<u>82.03</u>
Swin-UMamba [22]	83.52	83.72	<u>91.02</u>	90.12	<u>91.74</u>	95.91	77.52	72.86	74.05	76.37	73.02	80.53
VMamba [25]	73.97	80.65	75.06	89.67	80.67	92.18	77.35	72.66	73.38	76.39	72.87	79.83
VM-UNet [33]	80.19	88.51	87.74	90.83	88.54	94.36	76.83	71.58	71.05	77.22	71.60	79.16
PlainMamba [44]	73.47	82.65	88.26	89.17	82.89	91.14	70.94	62.74	66.74	65.34	63.36	66.48
SCSegamba [21]	<u>90.71</u>	<u>91.20</u>	90.39	91.00	91.70	94.48	78.75	<u>77.82</u>	<u>76.27</u>	<u>79.44</u>	<u>77.82</u>	79.50
FFM [13]	88.05	85.87	81.22	95.93	89.84	86.61	78.75	74.50	74.30	78.61	74.54	82.03
GLCP [46]	89.83	91.07	89.27	89.15	91.52	96.36	77.28	74.50	72.46	76.94	74.51	82.66
FGOS-Net (Ours)	91.29	91.43	91.36	<u>91.51</u>	92.46	97.12	79.15	78.29	78.34	79.87	78.32	83.85

Method	CrackMap (%)						TUT (%)					
	mIoU	F1	Prec	Recall	ODS	cDice	mIoU	F1	Prec	Recall	ODS	cDice
RIND [30]	76.44	65.25	58.95	86.42	75.04	80.52	80.52	80.50	76.51	86.48	80.74	91.08
CT-CrackSeg [39]	<u>79.93</u>	73.36	60.09	85.74	<u>79.21</u>	90.80	81.53	82.07	80.61	84.91	82.07	<u>91.31</u>
SimCrack [15]	79.78	<u>75.11</u>	62.80	85.15	77.89	<u>91.80</u>	81.90	81.61	78.31	<u>86.79</u>	81.69	90.66
Swin-UMamba [22]	77.59	73.53	62.47	81.01	76.44	86.65	80.32	78.88	77.93	81.82	78.89	87.57
VMamba [25]	77.96	65.07	50.70	82.64	67.16	69.28	79.45	77.82	75.72	82.11	77.85	87.68
VM-UNet [33]	75.66	66.41	55.41	86.96	68.10	74.60	78.14	76.53	74.76	80.35	76.54	86.24
PlainMamba [44]	75.08	66.14	61.97	74.15	66.19	74.60	68.91	59.56	57.76	67.62	59.58	66.47
SCSegamba [21]	78.51	75.01	75.85	76.53	75.39	59.87	83.60	<u>82.21</u>	<u>81.75</u>	82.68	<u>82.21</u>	84.99
FFM [13]	76.07	70.12	57.21	93.77	75.18	89.87	<u>84.62</u>	81.88	80.43	84.70	81.88	91.09
GLCP [46]	70.30	69.80	76.33	78.35	70.34	68.37	80.74	76.12	69.46	85.41	77.34	87.52
FGOS-Net (Ours)	80.75	77.82	<u>76.27</u>	<u>87.94</u>	79.82	92.48	85.73	83.95	82.44	86.99	82.51	91.98

lighting, and crack morphology, effectively serving as a multi-domain benchmark within the thin-structure category.

Implementation & Metrics. Models follow official splits and are trained in PyTorch on a single NVIDIA RTX 3090 for 100 epochs with AdamW ($lr=1\times 10^{-4}$, cosine decay) and $\mathcal{L}_{BCE}+\mathcal{L}_{Dice}$. Inputs are resized to 256×256 with random flip, rotate ($\pm 15^\circ$), and photometric distortion. ASGP uses $T=3$, $N=64$. We report mIoU, F1, Precision, Recall, boundary-aware ODS, and connectivity-aware cDice [37]. FPS is measured at 256×256 with the full inference path including ASGP iterations. We further provide implementation details, runtime breakdown, and multi-seed robustness checks in the supplementary material. Code will be released at <https://github.com/kbaijin/FGOS-Net>.

4.2 Comparison with State-of-the-Art

FGOS-Net consistently outperforms recent CNN, SSM, and topology-specialized baselines across all benchmarks (Tab. 1).

Mitigating Connectivity Breaks (DeepCrack & Crack500). FGOS-Net ranks 1st in ODS and cDice on both datasets, reflecting strong boundary and topological integrity. Notably, GLCP [46]—a recent topology-aware method tar-

Table 2: Efficiency analysis. All metrics are measured on a single NVIDIA RTX 3090 with input size 256×256 and batch size 1, averaged over 100 runs. FGOS-Net achieves a competitive accuracy–efficiency profile.

Method	Year	Params↓	FLOPs↓	Size↓	FPS↑
RIND [30]	2021	30.63M	247.6G	117.06MB	28.46
CT-CrackSeg [39]	2023	22.88M	78.94G	87.37MB	66.0
SimCrack [15]	2023	29.58M	286.62G	225MB	40.81
Swin-UMamba [22]	2024	63.38M	104.1G	241.77MB	43.1
VMamba [25]	2024	19.50M	395.24G	65.54MB	28.8
VM-UNet [33]	2024	27.25M	29.42G	105.92MB	27.9
PlainMamba [44]	2024	16.72M	73.36G	96MB	90.6
FFM [13]	2024	55.43M	260.07G	211MB	83.5
SCSegamba [21]	2025	3.05M	18.16G	37MB	17.9
GLCP [46]	2025	46.30M	33.72G	177MB	202.6
FGOS-Net (Ours)	-	6.26M	7.87G	23.92MB	80.2

getting connectivity preservation via joint skeleton and local-discontinuity learning— achieves strong cIDice on DeepCrack (96.36%) but drops to 68.37% on CrackMap, suggesting sensitivity to background clutter. In contrast, FGOS-Net achieves the highest cIDice across all datasets, indicating robust topological preservation under geometry-aligned serialization. FFM [13] attains the highest Recall on DeepCrack (95.93%) but with substantially lower Precision (81.22%) and cIDice (86.61%), suggesting over-sensitivity to weak gradients and texture leakage.

Suppressing Texture Leakage (CrackMap & TUT). On CrackMap, many methods exhibit “high Recall, low Precision” (Precision ≈ 50 –57%), a signature of texture leakage. FFM exemplifies this pattern with 93.77% Recall but only 57.21% Precision. FGOS-Net achieves the highest mIoU (80.75%) and cIDice (92.48%) while maintaining balanced Precision–Recall. On TUT, spanning eight materially diverse environments, FGOS-Net achieves the highest mIoU (85.73%) and cIDice (91.98%), surpassing FFM (84.62%) and GLCP (80.74%), confirming that frequency-geometric disentanglement generalizes across diverse texture statistics.

Summary. FA-Scan alleviates connectivity breaks in low-contrast topology; ASGP suppresses texture leakage in cluttered environments. The combination outperforms not only general SSM backbones but also methods designed for topological preservation [46] or self-similar structure encoding [13].

4.3 Model Efficiency

FGOS-Net achieves a strong accuracy–efficiency profile (Tab. 2). Compared to Swin-UMamba, it reduces FLOPs by $>90\%$ and nearly doubles the throughput (80.2 FPS). SCSegamba is more compact (3.05M) but only reaches 17.9 FPS; FGOS-Net delivers $4.5\times$ the speed at 56% fewer FLOPs. Notably, GLCP [46] achieves the highest raw FPS (202.6) through optimized operators, yet requires $7.4\times$ more parameters and $4.3\times$ more FLOPs than FGOS-Net while lagging in accuracy by up to 10.5% mIoU (CrackMap). FFM [13] demands $33\times$ more FLOPs (260G). Crucially, FA-Scan adds no parameters because it only changes indexing order, and ASGP remains lightweight due to sparse probing ($N=64$), adding only a small inference overhead.

Table 3: System Diagnostics on DeepCrack. We conduct controlled ablations to verify: (A) The impact of stage-adaptive gating policies in LGB; (B) The sensitivity of ASGP to evolution steps (T) and probe count (N); (C) The geometric sensitivity of different scan trajectories for specific frequency bands; and (D) The stepwise contribution of each module, explicitly isolating the gains from the FA-Scan encoder versus the GFA/BRM decoder.

A. LGB Gating Policy						B. ASGP Hyper-parameters						
Policy	mIoU	F1	P	R	ODS	Setting	mIoU	F1	P	R	ODS	FPS↑
None	87.81	86.77	86.20	91.69	87.08	$T=3, N=64$	91.29	91.43	91.36	91.51	92.46	80.2
EEEE	88.54	87.62	86.75	92.06	88.42	$T=3, N=32$	90.80	90.20	89.52	91.84	91.56	81.5
GGGG	88.76	87.89	87.16	92.18	88.80	$T=1, N=64$	88.93	88.77	85.93	92.91	89.26	82.1
GEGE	88.84	87.99	87.67	91.80	88.85	$T=7, N=64$	88.30	87.39	86.77	91.34	88.66	76.4
GEE	88.86	88.00	88.58	91.91	88.93	$T=3, N=128$	87.78	86.55	85.82	90.21	86.72	79.8
C. Scan Trajectory Ablation (FA-Scan)												
Row	HH/LL Path	LH/HL Path	mIoU	F1	P	R	ODS					
D1	Hilbert	LH→H, HL→V	90.46	89.63	89.46	89.45	89.73					
D2	Raster	LH→H, HL→V	89.12	88.34	88.22	88.06	88.51					
D3	Z-order	LH→H, HL→V	90.05	89.18	89.02	88.90	89.10					
D4	Hilbert	LH→V, HL→H	88.74	87.90	87.68	87.83	87.96					
D5	Hilbert	Snake (Bi-dir)	89.82	88.95	88.80	88.62	88.93					
D6	Hilbert	Hilbert (All)	89.46	88.60	88.42	88.33	88.66					
D. Mechanism Dissection												
Row	Variant					mIoU	F1	P	R	ODS	Params↓	FLOPs↓
1	Standard Mamba Encoder					85.98	86.11	86.70	86.87	86.22	2.20M	7.81G
2	VMamba Encoder + GFA/BRM					87.56	88.12	87.45	89.30	88.65	21.50M	12.40G
3	Baseline (DWT + Raster)					86.44	86.97	87.09	87.26	86.97	6.09M	7.80G
4	DWT + Cross-Scan (4-dir)					88.15	87.62	87.90	88.48	87.95	6.10M	8.85G
5	Static Gate (M_0 Only)					88.67	87.12	88.29	89.13	88.08	6.26M	7.87G
6	ASGP ($T=3$, w/o FA-Scan)					89.83	88.93	88.51	89.36	89.63	6.22M	7.87G
7	FA-Block (FA-Scan + LGB)					90.46	89.63	89.46	89.45	89.73	6.09M	7.80G
8	FGOS-Net (w/o Align)					90.75	90.62	90.31	90.98	91.35	6.25M	7.86G
9	FGOS-Net (Full)					91.29	91.43	91.36	91.51	92.46	6.26M	7.87G

4.4 Ablation Studies and Diagnostics

To verify that the gains come from frequency–geometry alignment and active topology-conditioned gating rather than capacity scaling or branch diversity, we ablate key components on DeepCrack.

Feature Response Analysis. Fig. 4 visualizes the 1D feature response along a crack. Raster scanning shows signal dropouts when the traversal cuts across the structure, whereas FA-Scan maintains a more continuous high-amplitude response, supporting the intended role of geometry-aligned serialization.

H1: Disentangling Architecture vs. Serialization (Tab. 3C, D). A central question is whether the improvement comes from the DWT multi-branch design, generic scan diversity, or the proposed geometry-aligned serialization. We answer this through a controlled causal chain with an explicit scan-diversity baseline.

1) DWT decomposition alone is insufficient. Introducing DWT without scan alignment (Row 1 $\rightarrow$ Row 3) yields a negligible +0.46% mIoU, confirming that the multi-branch frequency architecture is not the performance driver.

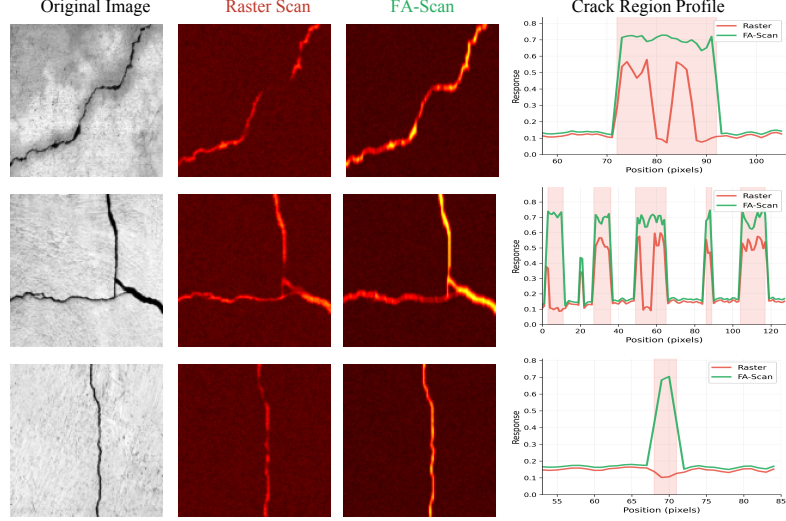


Fig. 4: Feature Response Analysis: Raster vs. FA-Scan. Feature maps are extracted from the Stage-2 encoder output (before IDWT), and profiles are sampled along the crack structure (highlighted in pink). The centerline is derived from the ground-truth skeleton for visualization purposes. Raster Scan: Features exhibit sharp signal drops when the scan path cuts across the crack, leading to fragmentation. FA-Scan: By aligning the scan trajectory with sub-band orientation, our method maintains a continuous, high-amplitude response (measured as the ℓ_2 -normalized mean channel activation), verifying the preservation of connectivity.

2) Scan diversity helps, but alignment helps far more. Replacing single-direction Raster with VMamba-style Cross-Scan (Row 3→Row 4), which applies four-directional raster to each DWT sub-band using the same shared SSM, improves mIoU by +1.71% but increases FLOPs to 8.85G. In contrast, FA-Scan (Row 3→Row 7) achieves a much larger +4.02% gain at lower FLOPs (7.80G). This indicates that frequency–geometry alignment contributes substantially more than directionality alone, while being computationally cheaper.

3) Alignment, not diversity, is the core driver. Tab. 3C provides fine-grained confirmation: deliberately mismatching sub-band assignments (D4) drops mIoU by −1.72% despite using the same number of scan paths; uniformly applying Hilbert (D6, 89.46%) or Snake scanning (D5, 89.82%) also underperforms the aligned FA-Scan (D1, 90.46%). Thus, the gain is better attributed to frequency–geometry alignment than to traversal diversity alone.

H2: Static vs. Active Probing (Tab. 3B, D). We decompose ASGP into static topology-conditioned filtering and active probe evolution. The transition from no gating (Row 3, 86.44%) to static gating (Row 5, 88.67%) brings +2.23% mIoU, showing that topology-guided suppression of high-frequency noise is already beneficial. Adding gradient-guided evolution (Row 5→Row 6) further improves mIoU by +1.16% and ODS by +1.55%, indicating that active probe relocation improves gate quality. However, excessive steps ($T=7$) may over-

Table 4: Cross-domain generalization Best results are in **bold**.

Method	CHASEDB1 [%]		Mass. Roads [%]	
	mIoU	clDice	mIoU	clDice
SimCrack [15]	75.82	75.24	72.63	71.94
VMamba [25]	73.00	76.01	71.51	71.20
SCSegamba [21]	78.17	77.30	75.41	73.40
FGOS-Net	80.45	83.35	79.83	80.86

concentrate responses and degrade performance. The diversity term is used to discourage probe concentration around a few high-response regions and to encourage broader spatial coverage along curvilinear support regions.

H3: Gating Policy (Tab. 3A). LGB is used mainly as a parameter-efficient mixing block rather than the core conceptual contribution. The hybrid EEGG policy uses ECA in shallow layers to preserve weak details and grouped SE in deeper layers to suppress false positives, yielding the best 91.29% mIoU trade-off.

H4: Encoder vs. Decoder Contribution (Tab. 3D). A strong baseline of VMamba Encoder + GFA/BRM (Row 2) lags behind the full FGOS-Net (Row 9) by 3.73% mIoU and 3.81% ODS, confirming that encoder-side geometry alignment provides structural cues the decoder alone cannot recover.

H5: Necessity of Local Spatial Alignment (Tab. 3D). Removing the detail-driven `GridSample` alignment (Row 8) forces the network to process an uncalibrated topology carrier. This drops ODS by 1.11% (92.46%→91.35%) and Precision by 1.05%. Since Haar DWT/IDWT is exact, this degradation should not be attributed to wavelet reconstruction error. Instead, it reflects feature-level misregistration introduced by downsampling and local aggregation. The Align module uses directional high-frequency cues to provide localized correction before topology modeling.

4.5 Qualitative Results

Qualitative analysis. As shown in Fig. 5, isotropic serialization-based baselines often produce fragmented predictions in low-contrast regions and false positives under texture clutter. In contrast, FGOS-Net maintains continuous crack traces and suppresses non-structural high-frequency artifacts, consistent with the intended roles of FA-Scan and ASGP.

4.6 Cross-Domain Generalization.

To verify that FGOS-Net generalizes beyond crack detection, we directly apply the same architecture—using identical hyperparameters and 256×256 inputs—to two distinct thin-structure domains: retinal vessel segmentation (CHASEDB1 [4], 28 images) and aerial road extraction (Massachusetts Roads [27], 1171 images).

As shown in Tab. 4, FGOS-Net improves mIoU/clDice over SCSegamba by +2.28/+6.05 on CHASEDB1 and +4.42/+7.46 on Massachusetts Roads, suggesting that geometry-aligned serialization and topology-conditioned gating remain effective beyond crack segmentation.

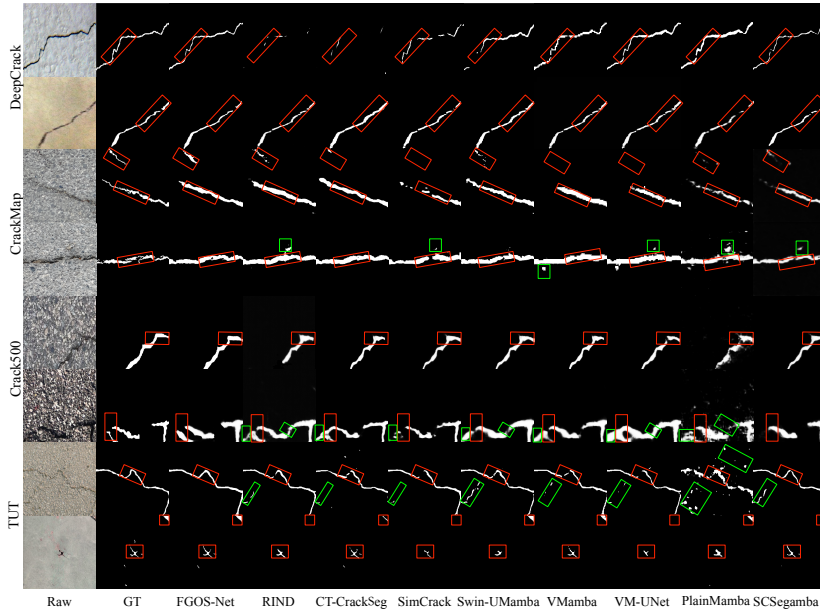


Fig. 5: Qualitative comparison on challenging scenarios. We visualize predictions from representative CNN/SSM baselines against FGOS-Net (Ours). All methods are visualized on the same test samples. Red Boxes (Topology): In low-contrast regions, baselines suffer from connectivity breaks, creating fragmented masks. Ours maintains continuous traces. Green Boxes (Texture): Under heavy clutter (*e.g.*, water stains, oil spots), baselines yield false positives (texture leakage). Ours successfully suppresses these artifacts via topology-gated injection.

5 Conclusion

We studied a geometry mismatch in applying State-Space Models to thin-structure segmentation: fixed 2D-to-1D serialization can interrupt structure-consistent propagation and aggravate texture leakage. To address this issue, we proposed FGOS-Net, a frequency-geometric framework that separates coarse topology from directional details via DWT and couples them through geometry-aligned FA-Scan and topology-conditioned ASGP. Experiments on four crack benchmarks and two cross-domain datasets show that FGOS-Net improves boundary integrity and connectivity preservation while maintaining a competitive accuracy-efficiency trade-off. The current scan assignment is a bounded inductive bias for topology-critical thin structures rather than a universal routing rule. Future work will explore more adaptive geometry-conditioned serialization while preserving the efficiency of deterministic scan paths.

Acknowledgements

This work was supported by the National Key Research and Development Program of China under Grant 2023YFB3906102.

References

1. Cao, X., Zhang, K., Jiao, L.: Csanet: Cross-scale axial attention network for road segmentation. *Remote Sensing* **15**(1), 3 (2022)
2. Dai, J., Qi, H., Xiong, Y., Li, Y., Zhang, G., Hu, H., Wei, Y.: Deformable convolutional networks. In: *Proceedings of the IEEE international conference on computer vision*. pp. 764–773 (2017)
3. Dao, T., Gu, A.: Transformers are ssms: Generalized models and efficient algorithms through structured state space duality. *arXiv preprint arXiv:2405.21060* (2024)
4. Fraz, M.M., Remagnino, P., Hoppe, A., Uyyanonvara, B., Rudnicka, A.R., Owen, C.G., Barman, S.A.: An ensemble classification-based approach applied to retinal blood vessel segmentation. *IEEE transactions on biomedical engineering* **59**(9), 2538–2548 (2012)
5. Fu, Y., Lou, M., Yu, Y.: Segman: Omni-scale context modeling with state space models and local attention for semantic segmentation. In: *Proceedings of the Computer Vision and Pattern Recognition Conference*. pp. 19077–19087 (2025)
6. Gu, A., Dao, T.: Mamba: Linear-time sequence modeling with selective state spaces. In: *First conference on language modeling* (2024)
7. Gu, A., Goel, K., Ré, C.: Efficiently modeling long sequences with structured state spaces. *arXiv preprint arXiv:2111.00396* (2021)
8. Han, D., Wang, Z., Xia, Z., Han, Y., Pu, Y., Ge, C., Song, J., Song, S., Zheng, B., Huang, G.: Demystify mamba in vision: A linear attention perspective. *Advances in neural information processing systems* **37**, 127181–127203 (2024)
9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 770–778 (2016)
10. Hilbert, D.: Über die stetige abbildung einer linie auf ein flächenstück. In: *Dritter Band: Analysis· Grundlagen der Mathematik· Physik Verschiedenes: Nebst Einer Lebensgeschichte*, pp. 1–2. Springer (1935)
11. Hou, Q., Zhang, L., Cheng, M.M., Feng, J.: Strip pooling: Rethinking spatial pooling for scene parsing. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 4003–4012 (2020)
12. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 7132–7141 (2018)
13. Huang, J., Zhou, Y., Luo, Y., Liu, G., Guo, H., Yang, G.: Representing topological self-similarity using fractal feature maps for accurate segmentation of tubular structures. In: *European Conference on Computer Vision*. pp. 143–160. Springer (2024)
14. Huang, T., Pei, X., You, S., Wang, F., Qian, C., Xu, C.: Localmamba: Visual state space model with windowed selective scan. In: *European conference on computer vision*. pp. 12–22. Springer (2024)
15. Jaziri, A., Mundt, M., Fernandez, A., Ramesh, V.: Designing a hybrid neural system to learn real-world crack segmentation from fractal-based simulation. In: *Proceedings of the IEEE/CVF winter conference on applications of computer vision*. pp. 8636–8646 (2024)
16. Katsamenis, I., Protopapadakis, E., Bakalos, N., Varvarigos, A., Doulamis, A., Doulamis, N., Voulodimos, A.: A few-shot attention recurrent residual u-net for crack segmentation. In: *International Symposium on Visual Computing*. pp. 199–209. Springer (2023)

17. Li, Q., Shen, L.: Wavesnet: Wavelet integrated deep networks for image segmentation. In: Chinese Conference on Pattern Recognition and Computer Vision (PRCV). pp. 325–337. Springer (2022)
18. Li, S., Singh, H., Grover, A.: Mamba-nd: Selective state space modeling for multi-dimensional data. In: European Conference on Computer Vision. pp. 75–92. Springer (2024)
19. Li, T., Li, C., Lyu, J., Pei, H., Zhang, B., Jin, T., Ji, R.: Damamba: Vision state space model with dynamic adaptive scan. arXiv preprint arXiv:2502.12627 (2025)
20. Li, X., Zhao, H., Han, L., Tong, Y., Tan, S., Yang, K.: Gated fully fusion for semantic segmentation. In: Proceedings of the AAAI conference on artificial intelligence. vol. 34, pp. 11418–11425 (2020)
21. Liu, H., Jia, C., Shi, F., Cheng, X., Chen, S.: Scsegamba: Lightweight structure-aware vision mamba for crack segmentation in structures. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 29406–29416 (2025)
22. Liu, J., Yang, H., Zhou, H.Y., Xi, Y., Yu, L., Li, C., Liang, Y., Shi, G., Yu, Y., Zhang, S., et al.: Swin-umamba: Mamba-based unet with imagenet-based pre-training. In: International conference on medical image computing and computer-assisted intervention. pp. 615–625. Springer (2024)
23. Liu, L., Zhang, M., Yin, J., Liu, T., Ji, W., Piao, Y., Lu, H.: Defmamba: Deformable visual state space model. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 8838–8847 (2025)
24. Liu, Y., Yao, J., Lu, X., Xie, R., Li, L.: Deepcrack: A deep hierarchical feature learning architecture for crack segmentation. *Neurocomputing* **338**, 139–153 (2019)
25. Liu, Y., Tian, Y., Zhao, Y., Yu, H., Xie, L., Wang, Y., Ye, Q., Jiao, J., Liu, Y.: Vmamba: Visual state space model. *Advances in neural information processing systems* **37**, 103031–103063 (2024)
26. Ma, J., Li, F., Wang, B.: U-mamba: Enhancing long-range dependency for biomedical image segmentation. arXiv preprint arXiv:2401.04722 (2024)
27. Mnih, V.: Machine learning for aerial image labeling. In: University of Toronto (Canada) (2013)
28. Mosinska, A., Marquez-Neila, P., Koziński, M., Fua, P.: Beyond the pixel-wise loss for topology-aware delineation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3136–3145 (2018)
29. Mou, L., Zhao, Y., Fu, H., Liu, Y., Cheng, J., Zheng, Y., Su, P., Yang, J., Chen, L., Frangi, A.F., et al.: Cs2-net: Deep learning segmentation of curvilinear structures in medical imaging. *Medical image analysis* **67**, 101874 (2021)
30. Pu, M., Huang, Y., Guan, Q., Ling, H.: Rindnet: Edge detection for discontinuity in reflectance, illumination, normal and depth. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 6879–6888 (2021)
31. Qin, Z., Zhang, P., Wu, F., Li, X.: Fcanet: Frequency channel attention networks. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 783–792 (2021)
32. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
33. Ruan, J., Li, J., Xiang, S.: Vm-unet: Vision mamba unet for medical image segmentation. *ACM Transactions on Multimedia Computing, Communications and Applications* (2024)
34. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C.: Mobilenetv2: Inverted residuals and linear bottlenecks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4510–4520 (2018)

35. Shaker, A., Wasim, S.T., Khan, S., Gall, J., Khan, F.S.: Groupmamba: Efficient group-based visual state space model. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 14912–14922 (2025)
36. Shi, Y., Li, M., Dong, M., Xu, C.: Vssd: Vision mamba with non-causal state space duality. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 10819–10829 (2025)
37. Shit, S., Paetzold, J.C., Sekuboyina, A., Ezhov, I., Unger, A., Zhylka, A., Pluim, J.P., Bauer, U., Menze, B.H.: eldice-a novel topology-preserving loss function for tubular structure segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 16560–16569 (2021)
38. Takikawa, T., Acuna, D., Jampani, V., Fidler, S.: Gated-scnn: Gated shape cnns for semantic segmentation. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 5229–5238 (2019)
39. Tao, H., Liu, B., Cui, J., Zhang, H.: A convolutional-transformer network for crack segmentation with boundary awareness. In: 2023 IEEE international conference on image processing (ICIP). pp. 86–90. IEEE (2023)
40. Wang, H., Chen, Y., Chen, W., Xu, H., Zhao, H., Sheng, B., Fu, H., Yang, G., Zhu, L.: Serp-mamba: Advancing high-resolution retinal vessel segmentation with selective state-space model. IEEE Transactions on Medical Imaging (2025)
41. Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., Hu, Q.: Eca-net: Efficient channel attention for deep convolutional neural networks. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11534–11542 (2020)
42. Xiao, C., Li, M., Zhang, Z., Meng, D., Zhang, L.: Spatial-mamba: Effective visual state space models via structure-aware state fusion. arXiv preprint arXiv:2410.15091 (2024)
43. Xu, Q., Chen, Y., Li, Y., Liu, Z., Lou, Z., Zhang, Y., Zheng, H., He, X.: Mambavesselnet++: A hybrid cnn-mamba architecture for medical image segmentation. ACM Transactions on Multimedia Computing, Communications and Applications (2025)
44. Yang, C., Chen, Z., Espinosa, M., Ericsson, L., Wang, Z., Liu, J., Crowley, E.J.: Plainmamba: Improving non-hierarchical mamba in visual recognition. arXiv preprint arXiv:2403.17695 (2024)
45. Yang, F., Zhang, L., Yu, S., Prokhorov, D., Mei, X., Ling, H.: Feature pyramid and hierarchical boosting network for pavement crack detection. IEEE transactions on intelligent transportation systems **21**(4), 1525–1535 (2019)
46. Zhou, F., Gao, Z., Zhao, H., Xie, J., Meng, Y., Zhao, Y., Lip, G.Y.H., Zheng, Y.: GLCP: Global-to-Local Connectivity Preservation for Tubular Structure Segmentation. In: proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2025. vol. LNCS 15975, pp. 237 – 247. Springer Nature Switzerland (September 2025)
47. Zhu, L., Liao, B., Zhang, Q., Wang, X., Liu, W., Wang, X.: Vision mamba: Efficient visual representation learning with bidirectional state space model. arXiv preprint arXiv:2401.09417 (2024)
48. Zhu, X., Su, W., Lu, L., Li, B., Wang, X., Dai, J.: Deformable detr: Deformable transformers for end-to-end object detection. arXiv preprint arXiv:2010.04159 (2020)
49. Zuo, X., Sheng, Y., Shen, J., Shan, Y.: Topology-aware mamba for crack segmentation in structures. Automation in Construction **168**, 105845 (2024)