

Review

Not peer-reviewed version

---

# Mamba for Remote Sensing: Architectures, Hybrid Paradigms, and Future Directions

---

[Zefeng Li](#), [Long Zhao](#), Yihang Lu, Ma Yue, [Guoqing Li](#)\*

Posted Date: 2 December 2025

doi: 10.20944/preprints202512.0078.v1

Keywords: visual state-space models; mamba; remote sensing; earth observation; hyperspectral imaging; multimodal data fusion; image restoration; foundation models



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Review

# Mamba for Remote Sensing: Architectures, Hybrid Paradigms, and Future Directions

Zefeng Li <sup>1,2</sup>, Zhao Long <sup>1</sup>, Yihang Lu <sup>1,2</sup>, Ma Yue <sup>1</sup> and Guoqing Li <sup>1,\*</sup>

<sup>1</sup> Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China

<sup>2</sup> School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100049, China

\* Correspondence: ligq@radi.ac.cn

## Highlights

### What are the main findings?

- The survey shows that Mamba-style visual state-space models benefit remote sensing primarily when sequences are genuinely long and well-structured; scan design and hybridisation matter more than simply swapping the backbone.
- A unified taxonomy is provided across hyperspectral analysis, multimodal fusion, segmentation, detection, restoration, and emerging RS foundation models, together with critical evidence on when Mamba improves accuracy, when it mainly saves resources, and when classical CNN or Transformer designs remain sufficient.

### What are the implications of the main findings?

- The survey shows that Mamba-style visual state-space models benefit remote sensing primarily when sequences are genuinely long and well-structured; scan design and hybridisation matter more than simply swapping the backbone.
- A unified taxonomy is provided across hyperspectral analysis, multimodal fusion, segmentation, detection, restoration, and emerging RS foundation models, together with critical evidence on when Mamba improves accuracy, when it mainly saves resources, and when classical CNN or Transformer designs remain sufficient.

## Abstract

Modern Earth observation combines high spatial resolution, wide swath, and dense temporal sampling, producing image grids and sequences far beyond the regime of standard vision benchmarks. Convolutional networks remain strong baselines but struggle to aggregate kilometre-scale context and long temporal dependencies without heavy tiling and downsampling, while Transformers incur quadratic costs in token count and often rely on aggressive patching or windowing. Recently proposed visual state-space models, typified by Mamba, offer linear-time sequence processing with selective recurrence and have therefore attracted rapid interest in remote sensing. This survey analyses how far that promise is realised in practice. We first review the theoretical substrates of state-space models and the role of scanning and serialization when mapping two- and three-dimensional EO data onto one-dimensional sequences. A taxonomy of scan paths and architectural hybrids is then developed, covering centre-focused and geometry-aware trajectories, CNN- and Transformer-Mamba backbones, and multimodal designs for hyperspectral, multisource fusion, segmentation, detection, restoration, and domain-specific scientific applications. Building on this evidence, we delineate the task regimes in which Mamba is empirically warranted—very long sequences, large tiles, or complex degradations—and those in which simpler operators or conventional attention remain competitive. Finally, we discuss green computing, numerical stability, and reproducibility, and outline directions for physics-informed state-space models and remote-sensing-specific foundation architectures. Overall, the survey argues that Mamba should be used as a targeted, scan-aware component in EO pipelines rather than a drop-in replacement for existing

backbones, and aims to provide concrete design principles for future remote sensing research and operational practice.

**Keywords:** visual state-space models; mamba; remote sensing; earth observation; hyperspectral imaging; multimodal data fusion; image restoration; foundation models

---

## 1. Introduction

Modern optical and SAR constellations now image the Earth at metre- and sub-metre-scale spatial resolutions, tens to hundreds of spectral bands, and revisit periods of hours to days. Sentinel-2, for example, provides 10 m multispectral imagery with 13 spectral bands and a global revisit frequency of about five days, while commercial missions deliver sub-metre panchromatic and 3–5 m multispectral data. Taken together, this combination of high spatial resolution, high spectral dimensionality, and high temporal frequency yields archives in which individual scenes can reach gigapixel scale and time series can span thousands of acquisitions. Models must capture long-range dependencies across space, spectrum, and time without losing small structures or exceeding realistic memory and energy budgets.

Deep neural networks have replaced hand-crafted descriptors by learning hierarchical features directly from data [1,2]. In remote sensing, convolutional neural networks (CNNs) became the standard backbone for scene classification, semantic segmentation, and object extraction [3,4]. Multi-scale encoders, dilated convolutions, and spectral–spatial fusion networks mitigate small objects, class imbalance, and label noise [5,6]. Yet even sophisticated CNN designs approximate long-range interactions only indirectly through deep stacking or large kernels, so the effective receptive field grows slowly and dependencies across full  $4k \times 4k$  tiles or long image time series remain difficult to model under realistic GPU constraints.

Transformer-based architectures address this limitation by replacing purely local processing with global self-attention over token sequences [7,8]. When adapted to remote sensing, ViT- and Swin-style backbones improve performance on land-cover mapping, change detection, and multi-temporal analysis [9]. However, the quadratic cost of self-attention in the sequence length  $L$  is difficult to reconcile with the combined demands of high spatial resolution, large scene extent, and dense temporal sampling. For a  $2048 \times 2048$  image divided into  $16 \times 16$  patches,  $L = 16384$ , so a single attention map contains  $L^2 \approx 2.7 \times 10^8$  entries; storing this map alone in 32-bit floating point requires roughly 1.1 GB per layer. Practical models therefore rely on patch cropping, windowed attention, or aggressive downsampling [10,11]. These remedies improve tractability but fracture global context and can suppress precisely those small structures—river channels, roads, building outlines—that are most relevant in Earth observation.

Structured state-space models (SSMs) offer a different route to long-range modeling. S4 showed that certain continuous-time linear dynamical systems can be implemented as convolutional kernels that model very long sequences with linear-time complexity and good numerical stability [12]. Mamba extends this line by introducing selective state-space models whose parameters depend on the current token [13–15]. Instead of fixed transition and input matrices, Mamba modulates them as functions of the feature vector, implementing content-aware scans that amplify informative regions and suppress background, while retaining  $O(L)$  time and memory complexity via parallel scan algorithms.

For Earth observation data, these properties matter at scale. Flattening a  $4096 \times 4096$  tile into a sequence yields  $L$  on the order of  $10^7$ , and even patch-wise tokenization produces sequences far longer than in typical natural-image benchmarks. Long spectral vectors and dense temporal stacks further increase  $L$ . Mamba-style architectures provide a way to process such sequences with linear complexity while conditioning state updates on scene content, an appealing feature for anisotropic geophysical structures and heterogeneous urban layouts. At the same time, serializing two- and three-dimensional EO data into one-dimensional sequences introduces design questions without

direct analogues in text: scan paths, multi-directional traversal, and hybridization with convolutions or attention all shape which spatial, spectral, and temporal neighborhoods are effectively modeled.

Bao et al. recently presented the first dedicated survey of Vision Mamba techniques in remote sensing, organising roughly 120 studies by backbone design, scan strategy, and application task [16]. Complementary surveys of visual Mamba architectures in generic computer vision [17–20] and the broader Mamba-360 review of state-space models, which categorises foundational SSMs into gating, structural, and recurrent paradigms across modalities, further situate Mamba within the wider SSM landscape [21]. Building on these efforts, this review adopts an EO-centric perspective and deliberately narrows its focus to questions that are specific to, or particularly critical in, Earth observation.

The remainder of this review is structured as follows. **Section 2** revisits continuous- and discrete-time SSM formulations, links them to selective recurrence in visual Mamba backbones, and examines how different scanning mechanisms serialize two- and three-dimensional EO data. **Section 3** turns to spectral analysis and multi-source fusion, asking in which regimes Mamba-based models empirically improve over strong CNN and Transformer baselines, and where they mostly replicate existing behaviour with added complexity. **Section 4** then examines high-resolution visual perception—semantic segmentation, object detection, and change detection—from the standpoint of scan design and hybrid backbones rather than as a catalogue of architectures. **Section 5** covers restoration and generative applications, including super-resolution, pan-sharpening, and spatiotemporal fusion, and highlights where long-range propagation is demonstrably useful and where classical models remain adequate. **Section 6** synthesizes cross-cutting issues such as stability, numerical robustness, physical consistency, efficiency, and the emerging landscape of Mamba-based foundation models, and identifies open questions that, in our view, must be addressed before Mamba can be treated as a default choice in EO pipelines. **Section 7** summarises when Mamba is warranted in Earth observation, which architectural choices have practical impact, and which research questions remain most pressing.

## 2. Theoretical Foundations and Architectural Evolution

### 2.1. From Linear State-Space Models to Visual Mamba

State-space models (SSMs) describe linear dynamical systems via a latent state coupled to the input–output relation and admit standard tools for stability, controllability, and frequency analysis. Recent SSM-based neural sequence models implement this formalism as parameterised recurrent updates in discrete time. These architectures are designed to propagate information with linear complexity in sequence length and can complement self-attention in regimes with long contexts or tight memory budgets. Here we focus on the linear time-invariant (LTI) formulation and its discretization, and relate selective state-space updates to the design of visual Mamba backbones.

#### 2.1.1. LTI State-Space Systems, Discretization, and Selective Recurrence

A continuous-time LTI state-space model with input  $u(t)$ , hidden state  $x(t) \in \mathbb{R}^d$ , and output  $y(t)$  is written as

$$\dot{x}(t) = Ax(t) + Bu(t), \quad y(t) = Cx(t),$$

where  $A \in \mathbb{R}^{d \times d}$  governs autonomous dynamics,  $B \in \mathbb{R}^{d \times 1}$  injects the input, and  $C \in \mathbb{R}^{1 \times d}$  maps the state to the output. For stable  $A$ , the system defines an impulse response and hence a convolution kernel  $k(\tau)$  that encodes long-range dependencies. Structured SSMs such as S4 parameterize  $A$  so that the corresponding kernel can be evaluated efficiently and remains numerically stable over very long sequences [22–29].

To use an SSM inside a neural network, the continuous system is discretized with step size  $\Delta$  (for instance under a zero-order hold on the input). The discrete-time dynamics between indices  $k - 1$  and  $k$  read

$$x_k = \bar{A}x_{k-1} + \bar{B}u_k, \quad y_k = Cx_k,$$

with discrete operators  $\bar{A}$  and  $\bar{B}$  obtained from  $A$ ,  $B$ , and  $\Delta$  by matrix exponentials or equivalent transforms. Unrolling this recursion yields a one-dimensional convolution  $y = k * u$  whose kernel is determined by  $(\bar{A}, \bar{B}, C)$ , so SSM layers can be implemented either as recurrent updates or as convolutions in the time or frequency domain. In both views, the cost of processing a sequence of length  $L$  scales linearly with  $L$ , in contrast to the quadratic cost of full self-attention.

Classical SSM layers keep  $(A, B, C)$  fixed across the sequence and learn them as global parameters. Mamba instead introduces *selective* SSMs in which the effective dynamics depend on the current token. Given an input feature  $x_t \in \mathbb{R}^d$  at position  $t$ , the model computes token-dependent parameters

$$B_t = \text{Linear}_B(x_t), \quad C_t = \text{Linear}_C(x_t), \quad \Delta_t = \text{softplus}(\text{Linear}_\Delta(x_t)),$$

and uses them to construct discrete operators  $\bar{A}_t$  and  $\bar{B}_t$ . The hidden state is updated as

$$h_t = \bar{A}_t h_{t-1} + \bar{B}_t x_t, \quad y_t = C_t h_t,$$

where  $h_t$  denotes the recurrent state. This selective recurrence enables the system to amplify informative tokens and suppress background responses, while still enjoying  $O(L)$  time and memory complexity via highly optimized parallel scan implementations. For long EO sequences and large image tiles, this content-dependent yet linear-time mechanism is the key distinction from attention-based models.

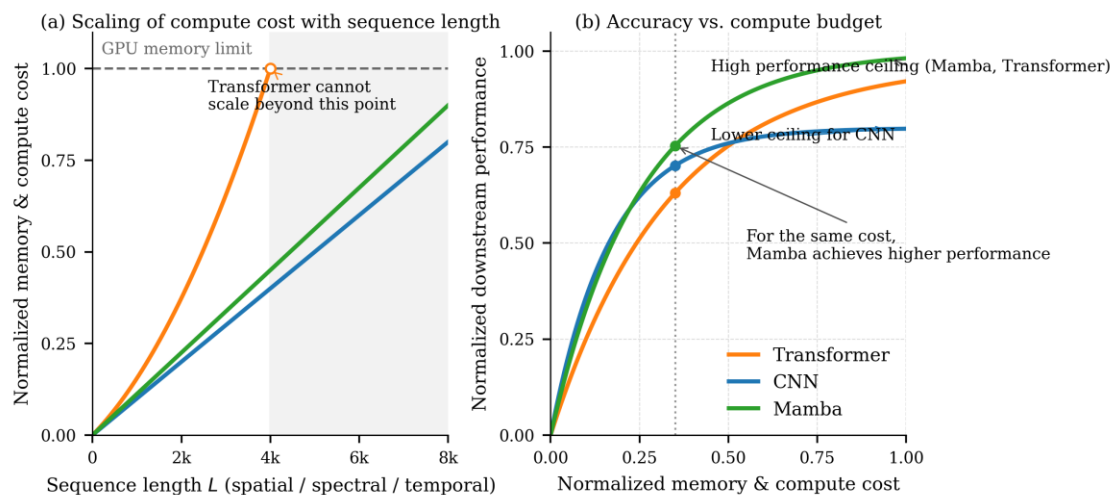
### 2.1.2. Topological Mismatch Between 1D Priors and 2D EO Data

The formulation above assumes a one-dimensional sequence with a natural ordering, as in time series, audio, or text. In those settings, neighbouring tokens in index space are also neighbours in the underlying domain, and long-range dependencies correspond to genuinely large temporal or spatial gaps. Satellite and aerial images, by contrast, lie on a two-dimensional grid, and any flattening into a one-dimensional sequence inevitably distorts neighbourhood relations. This extends naturally to higher-dimensional tensors when spectral and temporal axes are included.

A naive raster scan that visits pixels row by row or column by column preserves locality along the fast axis but breaks it along the slow axis: adjacent tokens in the sequence may correspond to distant pixels, while diagonally adjacent pixels can be separated by many steps. As the spatial extent  $H \times W$  grows, the mismatch between sequence distance and Euclidean distance increases, slowing down information propagation and inducing anisotropic artefacts in dense prediction maps on large images [30]. These effects are especially relevant in EO, where rivers, roads, coastlines, and urban blocks exhibit strong directional structure that should be propagated coherently.

Visual SSM architectures mitigate this mismatch by redesigning both the scan trajectory and the recurrent update. Some architectures abandon strict causality and run both forward and backward passes along the same path. Others introduce cross-shaped or diagonal scans so that each pixel can exchange information with neighbours along rows, columns, and diagonals within a small number of steps. A third line of work treats the image as a continuous 2D trajectory and designs serpentine or space-filling curves that shorten the average path between semantically related regions [31–33]. Together, these strategies narrow the gap between one-dimensional recurrence and two-dimensional geometry and provide more isotropic context for dense prediction and restoration.

In EO applications, the scan is therefore more than a numerical detail; it is a modelling assumption. Paths that align with expected physical anisotropies—for example along flight direction, river networks, or road grids—favour information flow along those structures, whereas near-isotropic scans are preferable when no dominant direction is known a priori. Many Mamba-based EO models implicitly encode such preferences through their scanning choices, even when these are presented as purely architectural variants.



**Figure 1.** Theoretical compute scaling and accuracy–cost trade-offs of CNN, Transformer, and Mamba long-range operators.

### 2.1.3. Visual backbones and hybridization

Once SSM layers had become competitive with attention in language modelling, they were rapidly imported into vision backbones. One class of designs replaces convolutional or Transformer blocks by visual SSM blocks at selected stages of the hierarchy, accumulating long-range context on high-resolution feature maps with linear-time complexity [34]. In parallel, lightweight variants adapt the state dimension, projection layers, and gating mechanisms so that SSM layers remain stable and efficient when stacked deeply on large images [35–37].

Pure SSM backbones are attractive from an efficiency standpoint but are not always optimal for EO tasks that demand both precise local texture modelling and long-range reasoning. Hybrid architectures therefore combine Mamba-style SSM blocks with convolutions or attention. A common pattern uses CNN layers in early stages to capture local edges, fine structures, and radiometric statistics, while SSM layers appear at mid and late stages to propagate information across large spatial extents or long temporal windows. Some models retain a small number of Transformer layers at the coarsest scales to explicitly model region-to-region and object-to-object relations on top of SSM features [34,35].

From a remote sensing perspective, this evolution turns visual SSMs into flexible building blocks rather than monolithic replacements. The same Mamba layer can augment a U-Net-like encoder–decoder, a Swin-style hierarchical Transformer, or a multi-branch fusion network, depending on spatial resolution, modality mix, and resource envelope [35–37]. In the remainder of this review, the interplay between structured linear dynamics, selective recurrence, and hybridization with convolutions or attention is a recurring theme.

## 2.2. Scanning Mechanisms in Remote Sensing

In visual SSMs, the scan specifies the order in which pixels, spectral bands, or time steps are visited and thus which positions become neighbours along the one-dimensional sequence processed by the model. For EO imagery this ordering is not imposed by the sensor but chosen in modelling: it determines which locations can exchange information within a few recurrent updates and how closely the induced one-dimensional prior respects the underlying two- or three-dimensional geometry. In this review we consider scan schemes on images of size  $H \times W$  and characterise them by two quantities: the effective path length, defined as the maximum number of recurrent updates required for information to propagate between two pixels, and the computational complexity as a function of the number of pixels  $HW$ . Continuity-aware scanning has also been explored outside EO, for example in DiT-style Mamba diffusion models such as ZigMa, where zigzag paths enforce spatial continuity between neighbouring patches on the image grid [38].

### 2.2.1. Directional and Multi-Directional Scans

The simplest image serialization is a row-wise or column-wise raster scan. Tokens that are adjacent along the fast axis remain close in the sequence, but neighbours along the slow axis can be separated by  $O(H)$  or  $O(W)$  steps, so information travels slowly across large tiles. Bidirectional variants alleviate this bias by running the recurrence forward and backward along the same path, allowing past and future tokens to influence each position. Bi-MambaHSI applies such bidirectional scans along spatial and spectral axes of hyperspectral cubes so that each pixel aggregates context beyond what standard convolutions provide [39].

Omnidirectional schemes increase directional coverage. RS-Mamba designs an omnidirectional selective scan in which states are updated along horizontal, vertical, and diagonal paths, approximating isotropic receptive fields while keeping overall cost linear in the number of pixels [40]. Cross-scan mechanisms such as those in VMamba restrict the number of directions but route information along a set of criss-cross paths; the diameter of the induced token graph then shrinks to  $O(\sqrt{HW})$ , which is advantageous for dense prediction on very large scenes [32].

Spiral and centre-focused scans encode yet another prior. SpiralMamba starts from a central pixel and follows an outward spiral so that the prediction region and its immediate neighbourhood appear in a contiguous subsequence [41]. Related schemes place the target pixel near the sequence centre and use tokens on both sides as spatial context. These designs are well suited to tasks where labels are attached to specific pixels or small patches, because the most relevant neighbourhood is processed as a compact segment while long-range dependencies are propagated along the remainder of the sequence.

In all cases, scan design implicitly states where relevant context is expected—along a dominant axis, across several orientations, or around a centre—and the SSM inherits this inductive bias through its recurrence.

### 2.2.2. Data-Adaptive and Geometry-Aware Scans

In heterogeneous scenes, hand-crafted raster or cross patterns can be misaligned with object boundaries or acquisition geometry. Several recent models therefore learn traversal paths or receptive regions from data so that the scan better follows local structure. QuadMamba uses a quadtree partitioning guided by learned locality scores; tokens are grouped into blocks with strong internal correlation, and an omnidirectional window-shift strategy moves information between blocks while preserving spatial coherence [43]. FractalMamba++ serializes two-dimensional patches along Hilbert-like fractal curves, which preserve locality across scales and adapt to varying input resolutions without redesigning the scan [43]. DAMamba couples the scan order with learned masks that concentrate updates on roads, building edges, and other salient structures, whereas MDA-RSM reweights multi-directional paths to reflect dominant building orientations and symmetries in urban layouts [44,45].

Task geometry also motivates specialized scans. For road and building extraction, traversals that approximate centre lines or follow estimated curvatures allow elongated man-made structures to be covered by short token paths. In change detection, cross-temporal scans traverse bi-temporal feature pairs at multiple scales, highlighting subtle structural differences between acquisitions while retaining linear complexity in the number of tokens. AtrousMamba adopts an atrous-window strategy with adjustable dilation rates: by increasing dilation, the receptive field grows without adding heavy attention blocks, which is beneficial when large-scale context and fine details must be captured simultaneously [46]. These geometry-aware schemes reduce the discrepancy between Euclidean layout and one-dimensional ordering.

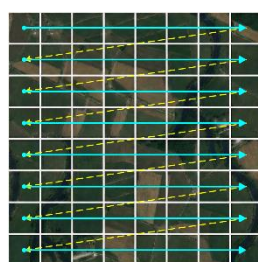
### 2.2.3. Transform-Domain and Irregular-Geometry Serialization

Scanning is not limited to regular grids. In several restoration models, SSM backbones are combined with Fourier or wavelet transforms: spatial tokens are serialized and processed by Mamba

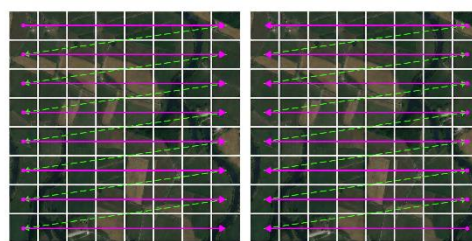
layers, while frequency-domain modules refine high-frequency details and impose priors on aliasing and noise [47,48]. FaRMamba makes this interaction explicit by restoring attenuated high-frequency components via multi-scale frequency blocks conditioned on Mamba features [49]. For LiDAR and other irregular point sets, token sequences follow acquisition trajectories or learned neighbourhood graphs instead of raster order, aligning better with local geometry in geometric-semantic fusion networks [50]. In satellite image time series, architectures such as SITSMamba process spatial features with CNN encoders and then apply Mamba along the temporal axis to model crop phenology and related dynamics under multi-task objectives [51]. In these cases, “scan” refers as much to traversal in time, spectrum, or graph space as to traversal across pixel grids.

**Table 1.** Typical scanning mechanisms in visual state-space models.

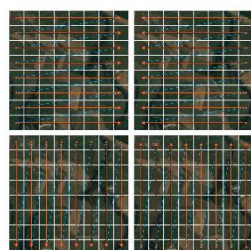
Scan type	Path length	Directions	Complexity	Example
Raster	$O(HW)$	1	Low	Vim [31]
Bidirectional	$O(HW)$	2	Low	Bi-MambaHSI [39]
Cross-scan	$O(\sqrt{HW})$	4	Medium	VMamba [32]
Spiral	$O(HW)$	1	Low	SpiralMamba [41]
Omnidirectional	$O(\sqrt{HW})$	6–8	Higher	RS-Mamba [40]
Adaptive	Learned	Dynamic	Variable	DAMamba [44]



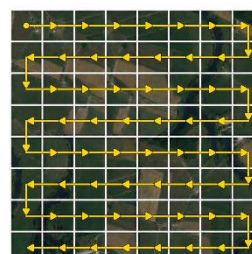
(a)



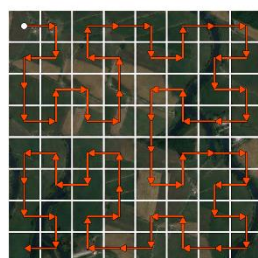
(b)



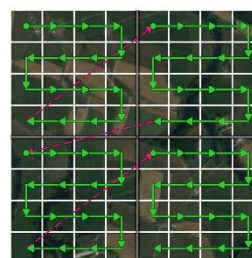
(c)



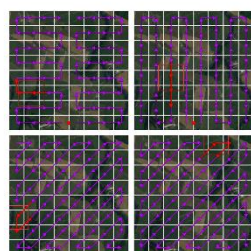
(d)



(e)



(f)



(g)



(h)

**Figure 2.** (a) raster scan (baseline), (b) bidirectional scan, (c) four-way selective scan with four orthogonal directions, (d) zigzag scan forming 8 continuous serpentine directions obtained from horizontal and vertical zigzags and their flips, (e) Hilbert-curve/fractal scan that preserves strong spatial locality across the grid, (f) window-based continuous scan with locally connected U-shaped trajectories, (g) omnidirectional spectral scan (OS-Scan) combining row-wise, column-wise and  $\pm$ diagonal trajectories, and (h) dynamic adaptive scan (DAS) whose path follows a data-driven importance map and largely skips low-importance regions.

#### 2.2.4. Empirical Guidelines and Design Trade-Offs

Comparative ablations show that scan patterns mainly differ in how they balance local continuity, directional coverage, and computational cost. In published studies, performance gaps are often modest and depend on the dataset, backbone, and scene structure. Zhu et al. evaluated one-directional, bidirectional, cross, and omnidirectional patterns across several backbones and datasets and observed that simple raster-like or bidirectional scans already provide strong baselines, while richer directional coverage yields consistent gains mainly for large scenes, pronounced anisotropy, or relatively shallow networks [52]. For EO applications, three pragmatic considerations are particularly useful:

- **Scene geometry and directional structure.** Long, thin, or strongly oriented structures (rivers, roads, building blocks) benefit from cross or omnidirectional scans that shorten paths along their main axes.
- **Spectral-spatial structure and coupling.** Hyperspectral cubes and multi-source stacks call for scans that traverse spatial and spectral axes jointly or explicitly interleave them, rather than treating each band independently.
- **Sequence length and computational budget.** Raster and bidirectional scans incur the smallest overhead and are suitable for very long sequences; omnidirectional and adaptive scans increase constant factors through additional branches or routing modules, even though asymptotic complexity in  $HW$  remains linear.

In Mamba-based EO models, the scan should therefore be treated as a task-dependent design choice tied to sensing geometry and resource constraints. Selecting an appropriate scan is often as influential as choosing the backbone itself, because it determines which neighbourhoods the SSM connects within a few recurrent steps and which structures the model can represent efficiently.

### 2.3. Architectural Hybrids and Design Patterns in EO

In current remote-sensing systems, Mamba is rarely used as a stand-alone backbone. Visual SSM blocks are instead inserted into CNN or Transformer pipelines, or attached as temporal and fusion modules around existing encoders. Architectures differ primarily in where Mamba is placed and how it interacts with local operators, not in the precise variant of the state-space layer. This subsection groups common design patterns and highlights their implications for dense prediction, multimodal fusion, and time-series analysis.

#### 2.3.1. CNN–Mamba Hybrids for Dense Prediction

A first family of designs attaches Mamba blocks to convolutional encoders or decoders for dense prediction. Typical U-Net-style hybrids retain a CNN encoder to capture local textures and object boundaries and replace bottleneck or decoder stages by Mamba layers so that large-scale context is propagated with linear complexity [41,53–56]. For very-high-resolution semantic segmentation and change detection, such CNN–Mamba U-Nets have been adapted to ultra-large tiles by combining shallow convolutional stems with deep Mamba decoders that aggregate global information while preserving fine details through skip connections [41,57].

A second pattern runs CNN and Mamba branches in parallel. The CNN stream emphasizes high-frequency details and radiometric statistics, whereas a Mamba stream handles long-range dependencies; their outputs are fused by attention or gating modules that weight features according

to local signal characteristics [53–56]. This division of labour is particularly effective for road networks, river systems, and settlement structures, where sharp boundaries and global connectivity are simultaneously important.

Across these variants, a consistent principle emerges. Convolutional layers are typically placed close to the input, where an inductive bias for local stationarity and edge detection is beneficial, while Mamba blocks appear at deeper or lower-resolution stages, where tokens represent larger spatial regions and cross-region context becomes crucial [41,53–57]. In practice, the strongest CNN–Mamba hybrids are not those that replace all convolutions, but those that allocate Mamba capacity specifically to stages with large token counts and long-range interactions.

### 2.3.2. Transformer–Mamba Architectures for Efficiency and Attention

A second family of hybrids couples Mamba with Transformer layers or attention modules. One motivation is to retain the relational modelling power of self-attention at coarse scales while reducing quadratic cost at high resolutions. Vision backbones that assign Mamba blocks to early, high-resolution stages and reserve a few attention layers for coarse feature maps follow this strategy: they keep explicit attention where fine-grained relational reasoning is most useful, while SSM layers handle long-range propagation over dense token grids [34,58].

In hyperspectral analysis, several architectures replace parts of the Transformer encoder by Mamba blocks to handle long spectral–spatial sequences more efficiently [59,60]. Some designs use Transformer layers within each scale to model rich interactions among tokens, while Mamba experts handle cross-scale propagation in a mixture-of-experts fashion, allocating capacity according to local spectral or spatial patterns [58]. Others employ Mamba as a spectral or temporal branch in parallel to spatial Transformers, so that spectral signatures and spatial layouts are captured with different inductive biases and then fused at intermediate layers [59–61].

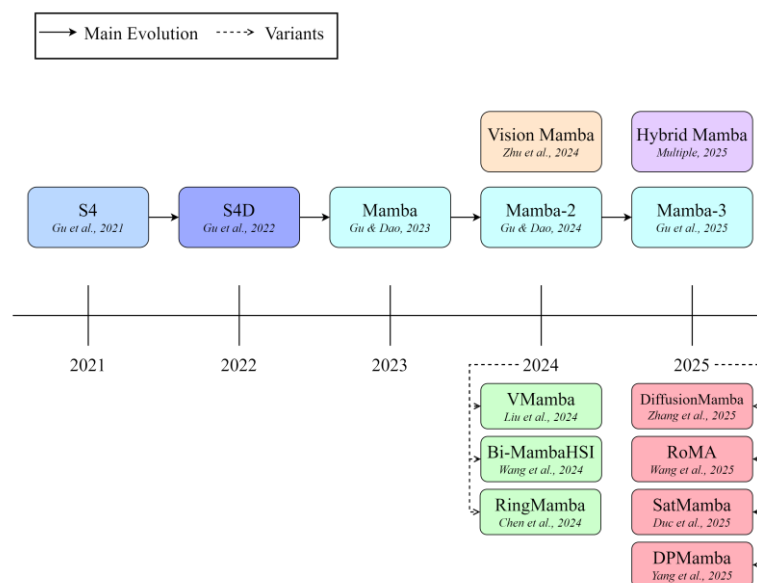
For multimodal fusion between optical and SAR images, hybrid backbones often combine CNN or Transformer encoders with Mamba modules that operate on modality-agnostic representations [61]. In these systems, attention is used sparingly to align features across modalities or regions, whereas Mamba layers act as efficient carriers of context across the full scene. Overall, Transformer–Mamba hybrids tend to favour architectures in which attention is concentrated at a few strategic locations, and Mamba provides the long-range backbone.

### 2.3.3. Multimodal and Temporal Pipelines

A third pattern embeds Mamba inside multimodal or temporal pipelines rather than in purely spatial backbones. For satellite image time series, a common choice is a CNN-based spatial encoder followed by a Mamba temporal encoder that processes sequences of spatial features under multi-task objectives such as crop classification and phenology monitoring [51,62]. This arrangement leverages CNNs for local land-cover geometry and uses linear-time recurrence to handle long acquisition histories.

In multimodal fusion, Mamba layers often appear as cross-modal or cross-scale fusion blocks. For example, some networks adopt separate CNN encoders for hyperspectral and LiDAR or for optical and SAR images, and then use Mamba to aggregate features across modalities and resolutions before prediction [61,63,64]. Other designs place Mamba modules at intermediate stages of encoder–decoder networks to propagate information between spatial scales or acquisition times, improving robustness to missing bands and acquisition gaps [62,65,66].

Across these hybrids, one design rule is consistent with the discussion in Section 2.2: convolutions and local attention are best used where spatial resolution is high and local details are critical, while Mamba layers are most effective at stages where tokens summarize larger receptive fields or encode long temporal windows. Treating Mamba as a flexible building block—rather than a wholesale replacement for CNNs or Transformers—has so far produced the most competitive EO architectures.



**Figure 3.** Conceptual evolution of Mamba-style state-space models and their vision and Earth observation variants.

### 3. Spectral Analysis

Spectral analysis tasks exploit the rich information contained in hyperspectral and multispectral measurements, often in combination with auxiliary modalities such as LiDAR or SAR. Mamba-based state-space models have been introduced here not as drop-in replacements for CNNs or Transformers, but as linear-time backbones that can follow tailored scan paths, integrate prior knowledge, and remain deployable under real-world resource constraints. This section first discusses hyperspectral image (HSI) classification, then multi-source fusion, and finally unmixing, target detection, and anomaly detection.

#### 3.1. Hyperspectral Image Classification

HSI classification assigns land-cover or material labels to pixels from data cubes with hundreds of contiguous bands, where subtle spectral variations and fine spatial structures jointly determine class boundaries. Early deep-learning approaches relied mainly on CNN backbones and shallow fusion of spectral and spatial cues, which limited their ability to model long-range spectral dependencies and cross-scene variability [5,67,68]. Subsequent Transformer-based HSI classifiers improved global receptive fields but inherited quadratic attention costs and often required aggressive patching or windowing, which is difficult to reconcile with very long spectral-spatial sequences [69–73]. Visual SSMs such as Mamba offer a different compromise by treating spectra, spatial neighbourhoods, or mixed spectral-spatial tokens as 1-D sequences with linear-time recurrence, but their real advantage depends critically on how sequences are constructed and how Mamba is embedded into larger architectures [74–76].

##### 3.1.1. Serialization and Selective Scanning

The first design axis is the **scan path** used to serialize spectral-spatial cubes. Naïve raster scans over  $H \times W \times B$  grids generate very long sequences and mix foreground and background in a way that slows information propagation and weakens inductive bias. A series of Mamba-based HSI classifiers therefore designs the scan itself as an explicit modelling choice rather than an implementation detail.

Centre-focused trajectories reorganise small windows so that spectrally and spatially informative samples around the prediction pixel form a compact contiguous subsequence. Designs such as spiral or centre-path scanning place the central region near the beginning or middle of the sequence and update the state from both directions, which improves robustness to label noise and

mixed pixels on benchmarks such as Indian Pines and Pavia University [40,77,78]. Other works extend this idea to 3D spectral–spatial paths, where sequences interleave bands and spatial neighbours so that Mamba propagates information jointly across wavelength and space while retaining linear complexity [69,79–82].

Beyond regular grids, several networks build sequences over superpixels or graphs. Superpixel-based and graph-based Mamba variants unfold tokens along region boundaries or k-NN graphs, which shortens effective sequence length and sharpens boundary representations in heterogeneous landscapes [83–89]. These structured scans consistently report moderate but reliable gains over raster or cross-scan baselines on standard HSI benchmarks while keeping the recurrent kernel simple [90–94]. Together, these results support a clear conclusion: for HSI classification, scan design is not a minor engineering detail but a primary source of performance, and centre-focused, 3D, and region-adaptive paths should be chosen deliberately according to acquisition geometry and label structure [95–99].

### 3.1.2. Hybrid CNN–Mamba and Transformer–Mamba Architectures

A second design axis is how Mamba is integrated with convolutional and attention modules. Most competitive HSI classifiers now adopt **hybrid backbones** where convolutions extract local edges and textures, while Mamba layers capture long-range spectral–spatial dependencies. Representative CNN–Mamba architectures use simple convolutional encoders followed by Mamba blocks or interleaved CNN–Mamba stages to extend context without sacrificing translation equivariance or efficient downsampling [71,72,74,75,100]. In parallel, Transformer–Mamba hybrids retain attention in selected spectral or channel dimensions and delegate long-range spatial or cross-band modelling to SSM layers, thereby reducing FLOPs and memory relative to pure Transformers while maintaining accuracy on Houston, Pavia, and other public benchmarks [70,76,80–82,101–104].

These results suggest a pragmatic view: in HSI classification, Mamba is most effective as a **complement** to convolutions or attention rather than a wholesale replacement. In particular, hybrid designs are attractive when high-resolution maps, long spectral profiles, and limited labels co-exist, while small-patch settings with abundant annotations may still favour lighter CNNs.

### 3.1.3. Frequency- and Morphology-Enhanced Modelling

A third line of work enriches Mamba with frequency-domain and morphological operators that encode prior knowledge about spectral smoothness and object shape. Wavelet- and Fourier-based architectures decompose HSI cubes into multi-scale frequency components before or within Mamba blocks, allowing the network to treat low-frequency background and high-frequency structures differently and to better reconstruct textured classes such as urban materials or vegetation mosaics [105–107]. Morphology-aware designs integrate dilation, erosion, and related operators into token generation, so that shape information is already embedded when sequences are fed into the SSM [87,108,109]. These frequency- and morphology-enhanced models typically provide incremental but consistent improvements over plain CNN–Mamba baselines, especially on noisy scenes or when object boundaries are poorly aligned with pixel grids [92–94,107].

### 3.1.4. Efficient, Few-Shot, and Transferable Learning

In practice, hyperspectral missions are usually constrained by scarce labels, domain shift between sensors, and tight memory and energy budgets on operational platforms. Under these constraints, recent work has begun to explore self-supervised, few-shot, and lightweight Mamba variants. Self-supervised methods embed composite-scanning Mamba blocks inside masked-reconstruction or contrastive frameworks, using large unlabeled HSI archives to pretrain representations that transfer well to downstream classification on Houston, Pavia, and WHU-Hi datasets [92–94,110,111]. Few-shot extensions introduce metric-learning heads, dynamic token

augmentation, or mixture-of-experts routing and consistently report higher accuracy than supervised Mamba baselines under five-shot settings [78,95,111].

On the efficiency side, several architectures apply depth and width reductions, re-parameterisation, or structured pruning to Mamba backbones, often guided by insights from MambaOut-style analyses originally developed for classification networks [76,89,97,99,100]. SpectralMamba-style families further compress spectra by sharing parameters across bands or compressing spectral channels, reducing FLOPs and parameters by factors of three to ten while maintaining comparable accuracy on datasets such as Houston2013 [98,101–105]. These results indicate that Mamba can be made competitive for onboard or embedded HSI processing when architectures are explicitly tuned for parameter and energy efficiency.

### 3.1.5. Summary

Across these strands, Mamba-based HSI classifiers have evolved from generic visual SSM backbones into task-aware architectures whose performance hinges on scan design, hybridisation with convolutions and attention, and efficiency-oriented training. First, careful design of centre-focused, 3D spectral-spatial, and region-adaptive scans is the main mechanism by which Mamba outperforms raster baselines on real HSI benchmarks, and there is little benefit in using state-space models with naïve sequence orderings. Second, hybrid CNN- and Transformer-Mamba backbones typically provide a better accuracy-efficiency trade-off than either pure CNNs or pure Transformers, especially for large scenes with limited labels. Third, frequency- and morphology-enhanced modules and self-supervised, few-shot, or lightweight variants add robustness in noisy or resource-constrained regimes but also increase design complexity and thus should be reserved for settings where their benefits have been empirically demonstrated.

## 3.2. Multi-Source Fusion

Many remote-sensing applications combine hyperspectral, multispectral, LiDAR, DSM, and SAR data rather than relying on a single sensor. Differences in imaging physics, spatial resolution, and coverage make fusion non-trivial, especially at large scale where long-range dependencies and misregistration must be handled under strict computational budgets. Transformer-based fusion networks alleviate some of these difficulties but their quadratic-complexity attention struggles at very high resolution or for long sequences.

Visual state-space models provide an alternative backbone with linear-time sequence modelling and flexible scan strategies. In multi-source settings, Mamba is used not only to propagate long-range context but also to encode cross-modal couplings directly within the state update. Current work can be grouped into three paradigms: cross-state interaction in heterogeneous classification, hybrid backbones with geometric or semantic priors, and frequency-aware or lightweight fusion, extended towards generative and reconstruction tasks.

### 3.2.1. Heterogeneous Modality Classification (HSI + LiDAR/DSM)

Joint classification of HSI with LiDAR or DSM requires exploiting both spectral signatures and height or geometric cues. A first paradigm replaces late feature concatenation with **cross-state interaction**: hidden states from different modalities interact during scanning so that cross-modal correlations are encoded in the recurrence.

CSFMamba implements a Cross-State Fusion module that couples convolutional feature extraction with Mamba-based global context modelling for HSI-LiDAR fusion [112]. On MUUFL and Houston2018, it achieves overall accuracies several percentage points higher than CoupledCNN late-fusion baselines, and ablations confirm that removing the cross-state module leads to substantial performance drops. S<sup>2</sup>CrossMamba extends this idea with an inverted-bottleneck Cross-Mamba design that updates multimodal states dynamically, reaching overall accuracies around 96% on MUUFL and clearly outperforming Transformer-based fusion backbones on MUUFL and Augsburg

[113]. MSFMamba similarly combines multi-scale spatial and spectral Mamba blocks with dedicated fusion modules for HSI and LiDAR or SAR, while CMFNet, TBi-Mamba, Mb-CMIFSD, and M2FMNet refine these cross-modal interactions through redundancy-aware fusion, triple bidirectional scanning, prototype-constrained self-distillation, or elevation-enhanced Mamba blocks [114–120].

A second paradigm overlays geometric or semantic priors onto Mamba backbones. DAHGMN couples graph convolutional networks with Mamba through hybrid GCN–Mamba blocks and dual-feature attention, where the GCN captures local geometric relationships from LiDAR and Mamba supplies long-range context [121]. Removing either branch degrades performance, indicating that local graph structure and global state-space dynamics are complementary. Other fusion networks integrate CLIP-guided semantics, tri-branch encoders, or edge-aware priors with Mamba to emphasise semantic structure and contours in complex urban scenes [122–124].

A third line focuses on frequency-aware and lightweight designs. LW-FVMamba combines skip-scanning Mamba backbones with frequency-domain channel learners to align multimodal features in the spectral frequency domain [125]. On Houston, it uses roughly 0.12 million parameters and around 40 million FLOPs—significantly lower than ExViT, NCGLF<sup>2</sup>, or standard VMamba—while slightly improving overall accuracy. TFFNet integrates fuzzy logic with Fourier and wavelet transform fusion to handle both uncertainty and spectral–spatial details in misregistered HSI–LiDAR pairs [126].

These three paradigms—cross-state interaction, GCN- or semantics-augmented hybrids, and frequency-aware lightweight designs—form a compact taxonomy for HSI–LiDAR/DSM classification.

### 3.2.2. Generative Fusion and Reconstruction

Beyond classification, Mamba backbones are also used for generative fusion and reconstruction, including HSI–MSI fusion, pan-sharpening, and spatiotemporal super-resolution. Here the goal is to reconstruct high-quality images from complementary observations while preserving spectral fidelity and fine spatial detail under ill-posed conditions and imperfect registration.

HSI–MSI fusion networks such as FusionMamba and SSCM extend the standard Mamba block into dual-input or cross-Mamba variants that jointly process HSI and MSI streams in the state update [127,128]. Long-range spatial–spectral dependencies are modelled by Mamba layers, while Laplacian or wavelet modules emphasise high-frequency texture. SSRFN decouples spectral correction from spatial enhancement: a CNN-based spectral module first compensates upsampling errors, and a Mamba branch then injects global context into the corrected features [129]. S<sup>2</sup>CMamba tackles pan-sharpening with dual-branch priors that jointly enforce spatial sharpness and spectral fidelity [130]. MCIFNet instead uses a Mamba backbone to generate latent codes and an implicit decoder that maps coordinates and codes to pixel values, allowing reconstruction at arbitrary resolutions [131].

Registration is sometimes integrated into the fusion process itself. PRFCoAM alternates between a modal-unified local-aware registration module and an interactive attention–Mamba fusion module, mitigating error accumulation that typically occurs in two-stage pipelines [132]. SINet couples Mamba with a multiscale invertible neural network based on Haar wavelets and regularises forward and inverse transforms to limit information loss during fusion [133]. For spatiotemporal fusion, MambaSTFM and STFMamba apply visual state-space encoders to long sequences and use task-specific decoders with expert modules for spatial alignment or temporal prediction, so that dense time series can be reconstructed with linear-time encoders and modest decoder overhead [62,134].

### 3.2.3. Summary

Multi-source fusion showcases how Mamba can be specialized for remote sensing. Cross-state fusion replaces patch-wise concatenation with recurrent interaction across modalities; hybrid GCN–Mamba or semantics-guided designs inject geometric and task priors; and frequency-aware, skip-scanning variants demonstrate that high accuracy is compatible with strict parameter and FLOPs

budgets. In generative and reconstruction tasks, dual-input Mamba blocks, registration–fusion coupling, and invertible or implicit decoders adapt state-space modelling to ill-posed inverse problems.

### 3.3. Hyperspectral Unmixing, Target and Anomaly Detection

Hyperspectral imagery provides dense spectral sampling, yet individual pixels typically contain mixtures of several materials embedded in structured backgrounds. Unmixing, target detection, and anomaly detection therefore need to exploit spectral correlations together with spatial context and background statistics. Classical linear-mixing and subspace models are appealing for their physical interpretability but become inaccurate in the presence of nonlinear interactions and complex clutter, while attention-based deep networks improve flexibility at the cost of substantial computation on full images. Recent work introduces Mamba-style state-space models that serialise spectra, spatial neighbourhoods, or pixel trajectories and use tailored scan schemes to balance global-context modelling with computational efficiency.

#### 3.3.1. Hyperspectral Unmixing

Hyperspectral unmixing estimates endmember spectra and their abundances from mixed pixels, a problem that becomes increasingly ill-posed in the presence of noise, nonlinear mixing, and limited supervision. Classical geometrical and statistical approaches provide physically grounded solutions but face difficulties with nonlinear effects and large scenes [135]. Mamba-based networks address these issues by combining local structure, long-range spectral–spatial context, and recurrent state updates.

MBUNet is a representative dual-stream design in which spatial and spectral features are extracted in parallel [136]. Convolutional layers capture local spatial patterns, while a bidirectional Mamba module aggregates global information along spectral–spatial dimensions. On synthetic benchmarks such as Samson, Jasper Ridge, and Urban, MBUNet substantially reduces mean spectral angle distance and root mean square error compared with both a Transformer baseline (DeepTrans) and a pure Mamba model (UNMamba), indicating that joint use of convolutions and bidirectional Mamba scanning is crucial for accurate abundance estimation.

Progressive sequence models such as ProMU treat unmixing as a sequence prediction problem over pixels or regions [137]. Stage-aware Mamba modules and progressive context selection refine abundances step by step. On Urban, ProMU reaches abundance RMSE comparable to image-level Mamba baselines while requiring roughly an order of magnitude fewer FLOPs than pixel-level Transformer models. Similar ideas appear in Mamba-SSFN and DGMNet, where Mamba branches are coupled with multi-scale convolutions or graph convolution to capture non-Euclidean spatial relationships and to improve scalability on large scenes [138–140].

#### 3.3.2. Hyperspectral Target Detection

In hyperspectral data, target detection aims to identify pixels belonging to specified materials or objects within complex, structured backgrounds. Effective detectors must remain sensitive to small targets while being robust to background variability and spectral perturbations.

HTMNet adopts a two-branch hybrid, with a Transformer stream for global multi-scale features and a LocalMamba stream with circular scanning that gathers local context around potential targets [141]. A feature interaction fusion module combines the outputs so that both global background structure and fine-scale neighbourhood cues influence detection decisions. Across benchmarks such as San Diego I and II, Abu-airport-2, and low-contrast Salinas scenes, HTMNet achieves area-under-curve values effectively at saturation and slightly higher than both a pure Mamba detector (HTD-Mamba) and a Transformer-based baseline (TSTTD), showing that coupling local Mamba recurrences with Transformer-scale context is beneficial in cluttered backgrounds.

HTD-Mamba approaches target detection from a self-supervised perspective [142]. A pyramid Mamba backbone and spatial-encoded spectral enhancement modules generate multiple spectral views for contrastive training, encouraging representations that are stable under spectral variations and effective at modelling background structure. Experiments indicate that such pretraining improves robustness in low-signal or few-label regimes, complementing hybrid architectures like HTMNet.

### 3.3.3. Hyperspectral Anomaly Detection

In hyperspectral imagery, anomaly detection is concerned with pixels whose spectra deviate from an estimated background model, typically in the absence of an explicit target signature. Performance depends critically on how accurately the background is modelled and reconstructed.

DPMN introduces a deep-prior Mamba network that uses a bidirectional Mamba-based abundance generation module to obtain background representations, coupled with a learnable background dictionary that partitions the background into several subspaces [143]. A regularization term combining total-variation and low-rank constraints enforces spatial smoothness and compactness, making it easier to separate anomalies from structured clutter. MMR-HAD adopts a reconstruction-based strategy with a multiscale Mamba reconstruction network, random masking to reduce the influence of anomalies on background estimation, dilated-attention enhancement, and dynamic feature fusion [144]. On standard anomaly benchmarks, both methods report improved detection accuracy, particularly when anomalies are subtle or densely distributed, compared with RX-type and CNN-based approaches.

### 3.3.4. Summary

Across unmixing, target detection, and anomaly detection, Mamba is rarely used in isolation. Dual-stream unmixing networks rely on Mamba to propagate information along spectral and spatial dimensions while preserving explicit endmember modelling; progressive sequence models trade a small loss in accuracy for substantial reductions in computational cost. Hybrid target detectors combine Transformer-scale global context with local Mamba scans and benefit from self-supervised pretraining, while anomaly detectors use Mamba-based reconstructions as flexible background models combined with dictionaries, masking strategies, and regularizers.

## 4. General Visual Perception

High-resolution semantic segmentation, object and change detection, and scene classification underpin many operational Earth observation products. Models must process gigapixel scenes, capture long-range spatial dependencies, and scale across archives and sensors. Mamba backbones provide linear-complexity context modeling and are increasingly inserted into CNN or hybrid networks as drop-in replacements for attention or as dedicated long-range branches.

### 4.1. Semantic Segmentation

Semantic segmentation assigns a land-cover class to every pixel and is therefore a stringent test for models that must combine fine boundaries with kilometre-scale context. Classical CNN and encoder-decoder architectures, including U-Net-style and pyramid pooling variants, have built strong baselines for EO mapping but struggle to aggregate information over very large tiles without resorting to aggressive downsampling or tiling [41,54,145–150]. Transformer-based segmentors extend the receptive field but are often memory-bound on high-resolution aerial and satellite images, which limits the spatial extent or batch size that can be processed in practice [151,152]. Mamba-based segmentation models attempt to retain CNN-like efficiency while adding linear-complexity propagation of long-range cues, and recent work has converged on a small number of design patterns rather than isolated architectures [57,153,154].

#### 4.1.1. Global–Local and Multiscale Architectures

Most Mamba segmentation networks adopt **global–local hybrids** in which convolutions handle local texture and boundary details, while Mamba branches transport information across downsampled feature maps. Samba and MF-Mamba, for example, attach Mamba encoders to CNN feature pyramids so that the SSM state evolves over multi-scale semantic maps rather than raw pixels, leading to improved mIoU on Potsdam, Vaihingen, and LoveDA with modest parameter overhead relative to their CNN baselines [57,153,155]. PPMamba wraps pyramid pooling modules with Mamba blocks and uses global–local state updates to refine predictions for large buildings and roads without oversmoothing small structures [154,156,157]. FMLSNet and related designs extend this idea by coupling ResNet-style encoders with lightweight Mamba layers, focusing on long-range refinement of large objects while leaving edge sharpening to convolutional decoders [158–161].

Other works target **data efficiency and adaptation**. LMVMamba and related models insert lightweight Mamba branches into multi-scale CNN encoders and share parameters across levels so that multi-scale features can be projected into a common linear dimension, which eases training on small labelled sets [147,148,162–164]. Multi-scale feature aggregation combined with state-space propagation has been shown to reduce fragmentation in large objects and to stabilise predictions under distribution shifts between cities or acquisition conditions [152,165–168].

Taken together, these results indicate that Mamba primarily serves as a global context carrier sitting on top of otherwise conventional segmentation stacks. Pure SSM encoders without convolutions remain rare and, on current benchmarks, offer limited evidence of clear benefits over carefully tuned CNNs and CNN–Transformer hybrids.

#### 4.1.2. Spectral–Channel, Multimodal, and Generative Designs

A second line of research exploits Mamba beyond purely spatial modelling by acting along spectral channels, across modalities, or inside generative decoders. Spectral–channel networks such as CPSSNet treat channels as ordered sequences and insert Mamba along the channel dimension, which improves discrimination of classes whose signatures differ mainly in subtle spectral patterns rather than geometry [158,169,170]. Multimodal designs push this idea further. MGF-GCN combines a graph encoder for DSM or LiDAR structure with a Mamba branch for optical imagery, using cross-modality fusion modules to align geometric and radiometric context for urban mapping [171]. MoViM integrates Vision Mamba into paired SAR–optical streams, showing that state-space branches can propagate shared context while leaving modality-specific artefacts to CNN or Transformer sub-networks [172].

Beyond discriminative models, DiffMamba couples CNN–Transformer encoders with diffusion decoders regularised by Mamba-style sequence propagation [173]. In these architectures, Mamba primarily stabilises long-range dependencies inside the generative head and improves the realism of predicted segmentations under heavy clutter or class imbalance, rather than replacing spatial convolutions.

Overall, segmentation results to date support a restrained but positive assessment of Mamba in EO. Global–local hybrids clearly help when tiles are large, classes are highly imbalanced, and infrastructure patterns span long distances, while spectral–channel and multimodal variants extend these gains to multi-band and multi-sensor settings [174–176]. At the same time, well-designed CNN or CNN–Transformer segmentors remain strong baselines, and the added complexity of Mamba branches is most defensible when long-range context or cross-modal coupling is demonstrably important.

#### 4.2. Object Detection

Object detection in remote sensing covers oriented ships and vehicles, multi-scale buildings, and very small targets for traffic or aviation monitoring. CNN and Transformer detectors are strong baselines on public benchmarks, but they still trade off ultra-high-resolution inputs, small objects,

and memory or latency constraints, so recent Mamba-based designs recast detection as sequence modelling, using SSM branches along scan paths aligned with object geometry to propagate long-range context at roughly linear cost.

#### 4.2.1. Oriented and Multimodal Detection

Multimodal detectors for RGB-IR UAV imagery explicitly account for modality-dependent disparities and spatial offsets. Several networks adopt dual branches with mask-guided regularisation and offset-guided fusion so that cross-modal features remain stable under misregistration [177–179]. Hybrid CNN-Mamba backbones, as in RemoteDet-Mamba, further encode cross-sensor context and background statistics, improving robustness in cluttered scenes [180]. For hyperspectral data, edge-preserving dimensionality reduction combined with visual Mamba enhances spatial-spectral representations and improves small-object separability [181].

A second line of work inserts Mamba blocks directly into detection pyramids. SSMNet augments the feature pyramid with state-space modules that aggregate information consistently across scales [182]. For small objects in UAV imagery, MV-YOLO introduces hierarchical feature modulation, while YOLOv5\_mamba couples bidirectional dense feedback with adaptive gate fusion to refine small-object representations in cluttered scenes [183,184]. Programmable gradients within SSMs have also been exploited in Soar to sharpen small-body detection under scarce or imbalanced data [185].

For oriented detection, OriMamba builds a hybrid Mamba pyramid with a dynamic double head that decouples classification and regression, whereas MambaRetinaNet combines multi-scale convolutions with Mamba blocks to balance global context and local detail [186,187]. Multi-directional scanning strategies further improve infrared object detection by integrating features along several orientations to suppress structured clutter [188]. In SAR ship detection, domain-adaptive state-space modules within a mean-teacher framework support unsupervised cross-domain transfer, complemented by large-strip convolutions and multi-granularity Mamba blocks that capture the elongated context of high-aspect-ratio targets [189,190]. Rotation-invariant backbones such as M-ReDet refine fine-grained features in dense ship clusters and other highly anisotropic scenes [191]. Beyond bounding boxes, context-aware state-space models have been extended to multi-category counting by scanning local neighbourhoods during inference and to single-stream object tracking that maintains localisation in cluttered or forested environments [192,193].

#### 4.2.2. Infrared Small-Target Detection

ISTD requires distinguishing faint, often sub-pixel targets from structured backgrounds. Several U-shaped architectures combine CNN encoders with Mamba blocks so that local detail is preserved while long-range context regularises background clutter. EAMNet introduces an adaptive filter module before Mamba encoding to enhance target visibility [194]. HMCNet and SBMambaNet insert spatial-bidirectional Mamba blocks into hybrid CNN-Mamba encoders to improve suppression of structured background clutter [195,196]. SMILE applies a perspective transform to sparsify the background and uses spiral spectral scanning to learn coupled spatial-spectral features [197], whereas MiM-ISTD introduces a “Mamba-in-Mamba” encoder with nested recurrences across spatial scales [198]. Together, these designs treat Mamba mainly as an efficient background modeller that normalises structured clutter and highlights salient responses.

#### 4.2.3. Salient Object Detection

In optical remote sensing images, salient object detection (SOD) aims to localise the most prominent geospatial targets in complex scenes so as to support subsequent analysis and decision-making [199]. Topology-aware hierarchical Mamba networks impose structural constraints that suppress spurious saliency responses [200]. TSFANet aligns multi-scale features in a Transformer-Mamba hybrid to maintain semantic consistency [201], whereas LEMNet uses edge cues in a

lightweight Mamba backbone to mitigate the lack of dense pixel annotations under weak supervision [202].

### 4.3. Change Detection

Change detection estimates land-cover transitions between multi-temporal images while suppressing pseudo-changes caused by illumination, sensor differences, or registration errors. High-resolution urban and peri-urban scenes add further complexity through small building footprints, thin roads, and non-rigid deformations. A recent review indicates that change detection is moving towards foundation models and efficient long-sequence encoders [203].

#### 4.3.1. Spatiotemporal Interaction Backbones

Spatiotemporal interaction backbones place Mamba at the core of bi-temporal feature fusion. ChangeMamba uses shared Siamese encoders followed by a visual Mamba block that scans concatenated pre- and post-event features, allowing each location to integrate cross-time context at linear cost [204]. CD-STMamba extends this idea with a Spatio-Temporal Interaction Module that encodes multi-dimensional correlations during both encoding and decoding [205]. CD-Lamba introduces a Cross-Temporal Locally Adaptive State-Space Scan (CT-LASS) that is designed to enhance the locality perception of the scanning strategy while maintaining global spatio-temporal context in bi-temporal features [206].

Methods such as 2DMCG, KAMamba, ST-Mamba, and SPRMamba explicitly target feature alignment and temporal reasoning on top of the core backbone. 2DMCG, for instance, couples a 2D Mamba encoder with change-flow guidance in the decoder to align bi-temporal features and mitigate spatial misregistration during fusion [207]. KAMamba targets long MODIS time series by combining a knowledge-aware transition-matrix loss with sparse deformable Mamba modules to model land-cover dynamics [208]. ST-Mamba introduces a Spatio-Temporal Synergistic Module that maps bi-temporal features into a shared latent space before Mamba propagation, thereby improving background consistency [209]. SPRMamba balances salient and non-salient changes via a saliency-proportion reconciler and squeezed-window scanning [210]. SMNet and LBCDMamba modify the scan pattern and pair Mamba blocks with modules such as RWKV or multi-branch patch attention, with the aim of improving long-range interaction modelling while keeping explicit pathways for local detail [211,212].

#### 4.3.2. Hybrid Convolution-Mamba Architectures

Hybrid architectures retain convolutional blocks for local structure and insert Mamba modules as long-range aggregators. CDMamba interleaves convolutional and Mamba branches via Scaled Residual ConvMamba blocks so that local texture and edges are refined while change cues propagate over larger areas [213]. CWmamba fuses a CNN-based base feature extractor with Mamba to jointly exploit local detail and global context, and Hybrid-MambaCD uses an iterative global-local feature fusion mechanism to merge CNN and Mamba features across scales [214,215]. ConMamba pushes this principle further by building a high-capacity hybrid encoder that deepens the interaction between convolutional and state-space features [216].

Multiscale aggregation is handled explicitly in SPMNet, which adopts a Siamese pyramid Mamba network with hybrid fusion of high- and low-channel semantic features, and in LCCDMamba, whose multiscale information spatio-temporal fusion module aggregates difference information for land-cover change detection [217,218]. MF-VMamba combines a VMamba-based encoder with a multilevel attention decoder to interactively fuse global and local representations, whereas VMMCD targets efficiency with a lightweight design and a feature-guiding fusion module that removes redundancy while preserving accuracy [219,220]. Residual wavelet transforms have been integrated with Mamba to refine fine-grained structural changes and suppress noise [221,222]. Attention-Mamba combinations such as Mamba-MSCCA-Net and AM-CD further enhance

hierarchical feature representation, and TTMGNet exploits a tree-topology Mamba to guide hierarchical incremental aggregation [223–225]. For unsupervised scenarios, RVMamba couples visual Mamba with posterior-probability-space analysis to detect changes without labelled pairs [226].

To manage multi-scale features more efficiently, a pyramid sequential processing strategy serialises multi-scale tokens into a long sequence and fuses them through Mamba updates [227]. Generative approaches such as IMDCD combine Swin-Mamba encoders with diffusion models, using iterative denoising to refine change maps and reduce artefacts [228]. Collectively, these results suggest that Mamba is most effective when it complements rather than replaces convolution, specialising in coherent long-range aggregation while CNN blocks handle precise localisation.

#### 4.3.3. Alignment-Aware Designs

Geometric misalignment between bi-temporal images is a major source of false alarms. DC-Mamba adopts an “align-then-enhance” strategy: bi-temporal deformable alignment first corrects spatial offsets at the feature level, after which Mamba layers refine change cues [229]. MSA (Mamba Semantic Alignment) instead operates at the semantic level, using a semantic-offset correction block to adjust deeper responses [230]. Building on vision foundation models, SAM-Mamba and SAM2-CD adapt SAM2 encoders to change detection by combining activation-selection gates or Mamba decoders that suppress task-irrelevant variations while sharpening change boundaries [231,232]. These studies underline that state-space dynamics perform best when applied to feature fields that already respect the imaging geometry [229,230,233].

#### 4.3.4. Hyperspectral and Challenging Scenarios

Mamba’s linear complexity is particularly attractive for data-intensive modalities such as hyperspectral imaging and for adverse conditions such as low-light scenes. GDAMamba captures global contextual differences at the image level and enhances temporal spectral discrepancies for hyperspectral change detection [234]. WDP-Mamba introduces a wavelet-augmented dual-branch design with adaptive positional embeddings to better preserve spatial–spectral topology [235]. SFMS couples a tri-plane gated Mamba with SAM-guided priors to stabilise learning for rare classes in hyperspectral change detection [236]. For low-light optical imagery, Mamba-LCD introduces illumination-aware state transitions that amplify weak signals in dark urban regions [237].

#### 4.3.5. Summary

Current Mamba-based change detectors span pure spatiotemporal backbones, hybrid CNN–Mamba networks, and designs that explicitly model cross-time alignment. Across these models, Mamba modules propagate long-range bi-temporal context at roughly linear complexity, while convolutional components refine high-frequency details, precise boundaries, and low-level registration [204,206,214,215,227]. Dedicated alignment blocks—either deformable or foundation-model-based—supply the geometric consistency that state-space dynamics alone do not guarantee [229,230,233].

**Table 2.** Representative Mamba-based change detection methods and performance on WHU-CD dataset (Optical).

Method	Core Idea	F1 (%)
SPMNet [217]	Siamese pyramid Mamba with hybrid fusion	91.80
VMMCD [220]	Lightweight design & feature-guiding fusion	92.52
SAM2-CD [232]	SAM2 adapter + activation selection gate	92.56
GSSR-Net [222]	Geo-spatial structural refinement & wavelet	93.07
Mamba-LCD [237]	Illumination-aware state transitions	93.60
SMNet [211]	Semantic-guided Mamba with RWKV integration	93.95

ChangeMamba [204]	Spatiotemporal interaction on concatenated features	94.19
SAM-Mamba [231]	SAM2 encoder + Mamba decoder (two-stage)	94.83
2DMCG [207]	Change-flow guidance	95.07
<b>CD-STMamba [205]</b>	<b>Spatio-temporal interaction module (STIM)</b>	<b>95.45</b>

#### 4.4. Scene Classification

Scene classification assigns semantic labels such as residential, industrial, or farmland to image patches and therefore requires models that capture global layout, multi-scale structure, and label co-occurrence. In this setting, Mamba backbones are mainly used as linear-complexity substitutes for attention, often combined with convolutional modules.

For single-label classification, recent work focuses on how 1D state updates can approximate non-causal 2D structure. RSMamba uses a dynamic multi-path scanning mechanism that mixes forward, reverse, and random traversals and attains F1-scores around 95% on UCM and RESISC45 with fewer parameters than ViT-Base or Swin-Tiny [238]. HC-Mamba couples a local content extraction module with cross-activation between convolutional features and Mamba states, while G-VMamba adds a contour enhancement branch to preserve luminance gradients [239,240]. To handle scale variation and limited labels, MPFASS-Net introduces progressive feature aggregation with orthogonal clustering self-supervision, ECP-Mamba applies multiscale contrastive learning to PolSAR data, and HSS-KAMNet hybridises spectral-spatial Kolmogorov-Arnold networks with dual Mamba branches for fine-grained land-cover identification [241-243].

For multi-label scene classification, the main challenge is modeling dependencies between co-occurring categories. MLMamba combines a pyramid Mamba encoder with a feature-guided semantic modeling module that refines class-wise embeddings and their relations, achieving competitive mean average precision on UCM-ML and AID-ML with substantially reduced FLOPs and parameter counts compared with Transformer-based baselines [244]. Overall, current Mamba-based scene classifiers fall into two patterns: multi-path or cross-activation scans for single-label scenes, and pyramid Mamba encoders coupled with semantic-relation modeling for multi-label settings.

## 5. Restoration, Generation, and Domain-Specific Applications

Visual state-space models are currently most mature in image restoration, where long-range dependencies and flexible scanning address the limits of local filters and quadratic-cost attention. Remote-sensing studies then extend these backbones to multimodal generation, compression, security, and scientific EO applications. This chapter therefore focuses on design patterns and where Mamba genuinely shifts the accuracy-efficiency-robustness trade space, rather than enumerating every variant.

### 5.1. Image Restoration and Geometric Reconstruction

Image restoration is a standard but critical stage in EO processing chains. Super-resolution, dehazing, denoising, and geometric reconstruction directly influence radiometric consistency, change-detection reliability, and the robustness of downstream products [245,246]. These problems combine local operators dictated by sensor physics and geometry with long-range correlations introduced by illumination, atmosphere, and acquisition layout. In this setting, Mamba-based models are appropriate only when long-sequence propagation plays a central role in the degradation process; otherwise, the additional architectural complexity is unlikely to provide clear benefits over well-engineered CNN- or Transformer-based restorers [247,248].

### 5.1.1. Super-Resolution

Remote-sensing super-resolution (SR) must sharpen man-made structures and edges at large scale without introducing spectral artefacts or aliasing [47,249]. Existing Mamba-based SR methods follow two broad strategies.

The first emphasises lightweight and hybrid architectures. Rep-Mamba, for instance, couples cross-scale state propagation with re-parameterised convolutions so that SSM layers carry context across large receptive fields while convolutions refine local details [250]. Other works keep a conventional CNN backbone and plug Mamba blocks only into deeper layers or skip connections, which reduces the marginal cost of adopting Mamba and simplifies deployment on existing SR pipelines [251–253]. These designs generally deliver modest but consistent PSNR / SSIM gains over pure CNN baselines on optical and infrared SR benchmarks, while reducing FLOPs compared with attention-heavy Transformers.

The second line develops frequency- and physics-aware SR models. Some architectures operate in wavelet or Fourier domains, where Mamba propagates information along multi-scale frequency coefficients rather than pixels, improving reconstruction of repetitive textures and fine structures [48,254,255]. Others embed priors on degradation operators and spectral response. Spectral super-resolution models employ Mamba to couple high-resolution multispectral bands with low-resolution hyperspectral measurements, decomposing the mapping into a physically constrained range space and a learned null space that absorbs residual correlations [256]. Here, the benefit of Mamba is clearest when spectral and spatial correlations interact over long ranges; on simple bicubic upsampling baselines with limited aliasing, CNNs often remain strong competitors.

Overall, SR studies suggest that Mamba is most valuable when SR is part of a broader multi-degradation or cross-sensor pipeline and when token sequences reflect physically meaningful propagation paths, not just flattened patches [47,48,249–263].

### 5.1.2. Atmospheric and Weather Restoration

Remote sensing imagery is frequently degraded by haze, clouds, rain streaks, and complex atmospheric scattering, so atmospheric and weather restoration methods seek to recover clear-sky surface reflectance from such observations [264,265]. In these settings, long-range dependencies arise naturally: scattering patterns and cloud fields evolve smoothly over space and time, and restoration must respect radiometric consistency across large regions.

Mamba-based dehazing and deraining networks typically adopt hybrid encoders: convolutional stages extract local gradients and edges, while Mamba branches propagate information along scanlines or multi-scale windows to model extended haze layers and cloud structures [266–268]. Some designs move to transform or polarimetric domains, where Mamba captures correlations among frequency components or scattering channels that are difficult to handle with local filters alone [269,270]. Compared with pure CNN baselines, these hybrids often improve structural similarity and colour fidelity under heavy haze, but the gains shrink on mild degradations where simple residual networks already perform well.

Weather-centric models apply similar ideas to rain, snow, and mixed artefacts, sometimes using recurrent or temporal SSM branches when short time series are available [271,272]. Additional illumination-aware and direction-adaptive dehazing and deraining variants follow the same pattern, combining scan-aware Mamba modules with conventional encoders to regularise large-scale shadow and cloud structures [257,258,260–263]. Here, Mamba's linear-time recurrence allows the same kernel to process sequences of varying length without changing architecture, which is appealing for operational systems that must handle irregular revisit intervals.

### 5.1.3. Denoising and Generalised Restoration

Hyperspectral denoising is a stringent test for restoration models: Gaussian and shot noise, striping, dead lines, and mixed artefacts all appear across hundreds of correlated bands [248,249,273].

In this context, Mamba provides a flexible way to couple spectral and spatial dependencies without quadratic attention.

Several works focus on specialised denoisers. Stripe-adapted and omni-selective scanning variants, for example, rearrange tokens along degradation-aware windows or channel groups so that Mamba updates align with noise patterns [274,275]. Cube-selective and continuous-scanning designs reorder voxels so that Mamba updates follow natural spectral–spatial orderings instead of arbitrary raster scans, reducing the effective sequence length while preserving physical neighbourhoods [276–278]. LaMamba combines linear attention with a bidirectional state-space layer and spectral attention, using Mamba to propagate information along long spectral paths while attention modules focus on local band interactions [278]. These models tend to outperform pure CNN and Transformer baselines on heavily degraded HSI benchmarks, particularly when noise patterns vary with wavelength or across sensors [276–278].

Beyond specialised denoisers, several architectures tackle **generalised restoration** where multiple degradations (blur, noise, haze, compression artefacts) are handled within a single network [279]. Spatial-frequency hybrids use CNN encoders to capture local textures, Fourier blocks to separate low- and high-frequency components, and Mamba layers to connect these representations across space. The main empirical conclusion is that Mamba helps when degradations introduce long-range correlations—such as global illumination shifts or structured striping—but brings limited extra value for purely local noise where classical CNN priors already suffice [247,248,276–279].

#### 5.1.4. Geometric Reconstruction: Stereo and Stitching

Geometric reconstruction tasks—stereo disparity estimation, DEM refinement, parallax-based stitching, and related problems—map image content to geometry and camera motion rather than to radiance [47,246]. They rely heavily on epipolar constraints, multi-view geometry, and sensor models, and thus offer a different test bed for Mamba.

Current work mostly uses Mamba as an auxiliary module inside otherwise geometry-aware systems. MEMF-Net and RT-UDRSIS integrate Mamba branches into multi-scale stereo or stitching pipelines, where SSM layers aggregate long-range matching cues while conventional cost volumes and warping handle geometric consistency [33,267,268,280–282]. Empirically, these hybrids can reduce artefacts in weakly textured or repetitive regions and improve robustness to radiometric differences between views. However, they do not replace explicit geometric components such as disparity regularisation or bundle adjustment; instead, they act as learnable priors that refine correspondences and fill gaps.

So far, there is little evidence that replacing entire stereo or SfM pipelines with SSM-only architectures would be beneficial. On the contrary, the strongest results come from **tight coupling between geometry-aware modules and lightweight Mamba branches**, suggesting that the role of SSMs in geometric reconstruction is to complement, not substitute, established model-based methods [33,267,268,282].

### 5.2. Vision–Language and Generation

Vision–language models for EO must link long visual sequences to text or symbolic outputs under strict memory and latency constraints. Gigapixel scenes and multi-temporal stacks inflate the number of visual tokens far beyond natural-image VLMs, and self-attention in the visual encoder becomes the main bottleneck. Mamba-style visual state-space backbones provide linear-time sequence processing and task-specific scanning, making them attractive as visual encoders for captioning, alignment, and generative pipelines in EO VLMs [284].

#### 5.2.1. Multimodal Alignment and Captioning

A recent taxonomy decomposes multimodal alignment into data-level, feature-level, and output-level schemes, with the visual encoder carrying most of the burden in remote sensing [285].

To cope with long sequences, Mamba-based captioning models modify how feature maps are serialized so that one-dimensional sequences reflect spatial and temporal structure rather than raw raster order.

RSIC-GMamba uses a Mamba backbone whose scan paths over feature maps are optimised by genetic crossover and mutation, reordering spatial tokens before state updates; semantically related regions are brought closer along the sequence and recurrence integrates context along coherent land-cover patterns, yielding consistent CIDEr and BLEU-4 gains over ViT encoders such as MG-Transformer at similar model scales [286]. RSCaMa extends to change captioning with temporal and spatial Mamba branches that serialize pre- and post-event images so that the state jointly encodes within-image structure and cross-time differences, improving the fidelity of descriptions of building construction or land-cover conversion [287,288]. DynamicVis adopts selective state-space modules for a high-resolution visual foundation model, concentrating computation on informative regions in 2048×2048 scenes and reducing latency and memory by more than 90 % relative to ViT baselines while retaining competitive captioning and retrieval performance in multimodal frameworks [289].

### 5.2.2. Generative Reconstruction, Compression, and Security

Mamba backbones also appear in generative pipelines where global structure affects rate-distortion trade-offs and robustness. Pan-Mamba addresses pan-sharpening with a backbone that jointly processes high-resolution panchromatic and lower-resolution multispectral inputs, letting state updates propagate structural cues from the panchromatic channel while preserving spectral consistency from multispectral bands and improving edge sharpness and spectral fidelity in urban scenes [290]. VMIC targets learned compression and replaces CNN hyperpriors with cross-selective scan blocks based on visual SSMs; bidirectional two-dimensional scans expose long-range spatial and channel correlations to the entropy model and yield BD-rate reductions of about 4–10 % over the VTM reference codec without a comparable complexity increase [291]. Dimba combines Transformer and Mamba layers for text-to-image diffusion, trading some attention overhead for higher throughput and lower memory usage during sampling [292].

In security, Mamba is often a teacher rather than a target. DMFAA distils features from a Mamba backbone into adversarial perturbations that are then applied to diverse student models, including CNNs and Transformers [293]. On scene-classification benchmarks such as AID, UCM, and NWPU, DMFAA achieves attack success rates 3–13 % higher than state-of-the-art black-box attacks under the same perturbation budgets, suggesting that long-range dependencies captured by Mamba provide additional degrees of freedom for transferable attacks and constitute demanding stress tests for RS recognition systems [293].

### 5.3. Domain-Specific Scientific Applications

Beyond generic benchmarks, Mamba-based architectures now appear in scientific EO applications where geometry, dynamics, and multi-sensor constraints dominate classic vision metrics. Across agriculture, disaster response, marine monitoring, and meteorology or infrastructure, CNN and attention branches typically handle local texture and sensor physics, while Mamba modules propagate information along long spatial or temporal trajectories with scan paths aligned to crop phenology, flood evolution, road networks, or ocean currents.

#### 5.3.1. Agriculture and Forestry

In precision agriculture and forestry, crop and canopy dynamics evolve over long temporal sequences, which pushes models to exploit extended time series. SITSMamba combines a CNN spatial encoder with a Mamba temporal encoder to model multi-year crop phenology as long sequences with position-weighted reconstruction [51]. STSMamba further addresses temporal-spectral coupling in MODIS time series through sparse deformable token sequences [294]. MSTFNet fuses improved Mamba modules with dual Swin-Transformers for hyperspectral precision

agriculture [295]. At object level, YOLO11-Mamba adds Efficient Mamba Attention to YOLO11 for maize emergence from UAV imagery [296]. CMRNet uses a hybrid CNN–Mamba backbone to enhance semantic features for rapeseed counting and localisation [297], and Succulent-YOLO integrates Mamba-based SR with CLIP-enhanced detection for succulent farmland monitoring [298]. RSVMamba for tree-species classification and EGCM-UNet for parcel boundary delineation follow the same template: CNN branches describe plant morphology and edges, whereas Mamba branches preserve parcel-level continuity and context [299,300].

### 5.3.2. Disaster Assessment and Emergency Response

Operational disaster assessment and emergency response are constrained by limited latency and resources, which places efficiency and robustness at the centre of model design. EMA-YOLOv9 augments YOLOv9 with Efficient Mamba Attention for real-time fire detection [301]. Flood-DamageSense adopts a multimodal design for flood assessment, fusing SAR/InSAR, optical imagery, and risk layers via a Feature Fusion State-Space module that projects them into a shared representation [302,303]. For geological hazards, SegMamba2D uses a lightweight encoder–decoder with Mamba modules to balance global context with local features for landslide mapping [304]. Mamba-MDRNet integrates pre-trained large language models with Mamba mechanisms to select reliable modalities for natural-disaster scene recognition [305]. LinU-Mamba for wildfire spread and C2Mamba for building change similarly use Mamba blocks to propagate long temporal or spatial links, while CNN and attention components manage local detail and cross-sensor alignment [306,307].

### 5.3.3. Marine Environment and Water Resources

Marine and inland-water applications are naturally suited to state-space modeling because structures are elongated, boundaries diffuse, and labels sparse. OSDMamba handles SAR-based oil-spill detection with an asymmetric decoder integrating ConvSSM components and deep supervision [308]. Algae-Mamba incorporates Kolmogorov–Arnold networks into a visual state space for algae extraction [309], and a synergistic fusion framework uses Mamba-based coral reef habitat classification to refine satellite-derived bathymetry [310]. OWTDNet detects offshore wind turbines with a dual-branch CNN–Mamba architecture in which a lightweight CNN branch captures turbine signatures and a Mamba branch encodes large-scale ocean context before alignment [311]. STD Mamba models sea-surface-temperature time series with temporal convolutions and bidirectional Mamba2 modules, while MMamba uses mutual-information-based feature selection and a Mamba reconstruction module for wind-speed gap filling [312,313].

### 5.3.4. Meteorology and Infrastructure

Road networks, meteorological fields, and environmental variables share elongated structures, smooth spatial patterns, and strong links to terrain. TrMamba steers its scan along predicted road directions in high-resolution imagery and applies selective Mamba updates along these paths to delineate road networks [314]. FDMamba employs a frequency-driven dual-branch structure to capture fine-grained edge details for road extraction [315], and other hybrids explicitly integrate multi-task learning to refine road-network topology [316]. Mamba-UNet applies selective state-space modeling to precipitation nowcasting via a dual-branch fusion module with multiscale spatiotemporal attention [317]. MCPNet uses an asymmetric Mamba–CNN collaborative architecture to balance memory usage and global modeling for large-scene segmentation [318]. Mambads for terrain-aware downscaling, BS-Mamba for black-soil degradation, and kMetha-Mamba for methane plume segmentation share a common design: CNN components encode local structure and process-related features, and Mamba components propagate information along spatial networks or temporal trajectories [319–321].

**Table 3.** Representative Mamba-based models and data/sequence regimes.

Domain	Typical data regime	Main sequence design	Dominant architecture pattern	Example methods
Hyperspectral classification / unmixing	long spectra, medium tiles	per-pixel spectral / spectral-spatial	dual-stream CNN-Mamba, bidirectional SSM	BiMambaHSI, SpiralMamba, HSS-KAMNet
VHR optical / SAR segmentation	4k-8k tiles	spatial raster / directional scans	CNN backbone + Mamba blocks, pyramid Mamba	PyramidMamba, RS-Mamba, LGMamba
Object detection / BEV segmentation	multi-scale, oriented targets	spatial-directional	dual-branch CNN-Mamba, BEV-Mamba	DMM, RemoteDet-Mamba, RSBEV-Mamba
Change detection / multi-temporal	bi-temporal or short sequences	temporal / spatio-temporal sequences	Siamese / parallel Mamba, atrous scans	AtrousMamba, Mamba-LCD, EdgePVM
Image restoration / fusion (SR, dehaze, clouds)	very high-res patches	spatial / spectral-spatial sequences	hybrid CNN-Mamba, frequency-aware Mamba	Frequency-assisted Mamba, RSDehamba, Pan-Mamba
Captioning, VLMs & foundation models	image-text pairs, multimodal corpora	multimodal token sequences	vision backbone + Mamba, Mamba-VLM	RSIC-GMamba, RSCaMa, DynamicVis
Scientific / geophysical EO	gridded fields, time series	spatial / temporal trajectories	Mamba + Conv, neural-operator-style SSM	STSMamba, Mamba-UNet, MMamba, Algae-Mamba

## 6. Advanced Frontiers & Future Directions

This section concentrates on questions that remain open before Mamba-style state-space models can be regarded as mature tools for Earth observation. Rather than listing generic “future work”, we focus on five cross-cutting themes: (i) theoretical validity and task regimes, (ii) hardware-aware deployment, (iii) physics-informed designs, (iv) remote-sensing foundation models and scaling behaviour, and (v) green computing, efficiency, and reproducibility.

### 6.1. Theoretical Substrates and Task Validity

We first revisit the theoretical substrates of visual state-space models and ask in which EO regimes their extra recurrence is justified.

#### 6.1.1. Task Regimes and the Limits of SSMs

The central question is not whether state-space models “work” on benchmarks but for which EO problems their extra recurrence is worth the cost. MambaOut addresses this directly by comparing Mamba-style backbones with gated CNNs and defining a characteristic sequence length  $\tau$  that scales with channel dimension [322]. On short sequences far below this length, such as ImageNet classification with  $224 \times 224$  inputs, gated CNNs match or slightly exceed VMamba accuracy while using fewer parameters and FLOPs [322]. These results, together with S4 studies showing that structured state-space layers are most advantageous on sequences of thousands of steps [12], suggest that SSM blocks are often unnecessary for short-range vision tasks.

EO workloads, however, range from small tiles to gigapixel images and multi-year time series. Patch-wise scene classification on  $256 \times 256$  crops or detectors with narrow receptive fields rarely reach token counts where SSMs help, and robust CNNs or light CNN–Transformer hybrids are usually easier to optimise and deploy. By contrast, dense prediction on very large scenes, long satellite image time series, or multi-sensor fusion with long correlation lengths resemble the long-range settings where structured SSMs and recent visual variants such as Spatial-Mamba have demonstrated clear benefits [12,323]. For such regimes, linear-time state updates and scan-aware design are more likely to translate into real efficiency and accuracy gains.

### 6.1.2. Structured SSMs, Mamba-3, and the Linear-Attention Frontier

Beyond the basic single-input SSM formulation, recent variants introduce multi-dimensional, structured, and hardware-aware parameterisations that fit visual and EO workloads more directly. Sparse Mamba imposes controllability, observability, and stability constraints on the state-transition matrix and promotes sparsity, reducing parameter counts and training time without loss of language-model perplexity [324]. Mamba-3 introduces multi-input–multi-output updates and richer recurrence patterns with explicit accuracy–efficiency trade-offs under fixed inference budgets [15]. Visual State Space Duality (VSSD) adapts causal SSD mechanisms to non-causal vision data by discarding state–token interaction magnitudes, enabling bidirectional context while retaining efficient scan implementations [325]. These developments suggest that operator structure, not just asymptotic complexity, is central for performance and stability.

At the same time, analyses of linear-attention Transformers and architectures such as RoMA highlight that structured SSMs and linear-attention ViTs occupy a shared design space: both achieve linear memory scaling, and differences arise mainly from kernel parameterisation, implicit priors, and hardware friendliness. For EO, the relevant question is therefore not “SSM or Transformer?”, but which linear-time operator best matches the sensing geometry, dynamics, and compute budget of a given task. On GPUs with highly optimised FlashAttention kernels, a well-implemented linear attention block may be more efficient than a naive Mamba layer, whereas on memory-limited edge devices the fixed-size state update can be preferable. Systematic comparisons under controlled budgets are still rare, especially for remote-sensing workloads, and represent an immediate research gap.

## 6.2. Hardware-Aware Deployment

We next consider how Mamba-based models behave under the size–weight–power constraints of EO platforms.

### 6.2.1. On-Board Inference from UAVs to Constellations

RTMamba demonstrates that visual Mamba backbones can be tailored to edge UAV platforms by replacing self-attention with state-space blocks in a semantic segmentation encoder and discarding redundant high-resolution features before expensive processing [326]. The resulting models run at real-time frame rates on embedded GPUs or NPUs with accuracy comparable to heavier Transformer baselines, and similar state-space encoders have been applied to event-based vision for space situational awareness, where asynchronous streams must be processed with low latency [326,327].

At orbital altitudes, EdgePVM adopts a parallel Siamese Vision Mamba architecture for on-board change detection in serverless satellite-edge constellations [328]. Multi-scale encoders with Mamba backbones and lightweight fusion modules produce change maps directly on board, halving complexity relative to Transformer baselines and enabling real-time inference under satellite power and memory budgets. Only compact change products need to be downlinked, rather than full-resolution imagery. These examples align with broader trends in 6G and edge intelligence and suggest that claims of scalability for EO architectures should increasingly be supported by evidence of on-board or near-sensor deployment [329].

### 6.3. Physics-Informed State-Space Models

Most Mamba-based EO models are currently trained as generic sequence learners, with only loose links to the partial differential equations that govern geophysical processes. The state-space formalism, however, sits naturally alongside recent advances in scientific machine learning, where physics-informed neural networks and neural-operator architectures seek to embed differential constraints directly into the learning process [330–332]. These developments suggest that Mamba-style SSMs need not remain purely data-driven encoders, but can be used as learnable discretisations of physical dynamics. In what follows, we discuss two complementary directions: viewing Mamba-style SSMs as neural operators for geophysical dynamics, and coupling them with classical numerical solvers through residual or subgrid parameterisations.

#### 6.3.1. SSMs as Neural Operators for Geophysical Dynamics

Recent work at the interface between dynamical-systems modelling and machine learning shows that Mamba-like architectures can function as neural operators for nonlinear dynamical systems, achieving competitive or superior long-horizon stability and extrapolation on chaotic benchmarks relative to Transformer and classical neural-operator baselines [331–334]. In meteorology and hydrology, spatial–temporal Mamba variants such as Mamba-ND, MetMamba, and RiverMamba learn mappings from multivariate reanalysis or forecast fields to future atmospheric states, regional forecasts, or global river discharge, while retaining linear-time complexity in sequence length and demonstrating gains over attention-based and physics-only reference models [335–337]. These studies treat the SSM update as a learnable time-stepping scheme whose parameters capture underlying dynamics, rather than as a generic sequence model.

Analogous opportunities exist in EO. Applications such as ocean-colour retrieval, SAR/InSAR time-series analysis, and radiative-transfer inversion could, in principle, benefit from state matrices whose structure reflects diffusion, advection, elasticity, or energy-balance constraints instead of being left fully unconstrained. At present, however, most EO SSMs still learn generic state matrices. Replacing these parameterisations with designs that encode conservation laws or stability criteria, and evaluating their behaviour under distribution shift, is a concrete and overdue research direction.

#### 6.3.2. Coupling SSMs with Classical Solvers

A complementary direction is to couple SSMs with numerical solvers rather than learning full dynamics from scratch. In such hybrids, Mamba blocks would not replace established PDE solvers but would learn residual tendencies—for instance, subgrid terms for coarse-resolution climate simulations or corrections to flood and fire propagation models—using time-marching structures that align naturally with the state-space update. Hybrid strategies of this kind have already proved effective with other network families, from physics-informed neural networks and deep-learning-based subgrid parameterisations to neural general circulation models and machine-learning corrections of analytical or reduced-order models [330,338–341]. Recent work on next-generation Earth system models argues that reliable weather and climate prediction will increasingly rely on such AI–physics hybrids rather than purely data-driven surrogates [342].

What is missing, specifically for state-space architectures, is a systematic framework for choosing state dimensions that align with physical modes, regularising SSM parameters according to conservation laws or stability criteria, and evaluating long-horizon stability and energy or mass conservation relative to established numerical baselines. Progress along these lines will determine whether physics-informed SSMs and Mamba-style operators remain proof-of-concept demonstrations or become components of operational geoscience models.

### 6.4. Remote-Sensing Foundation Models and State-Space Multimodal Alignment

Remote-sensing foundation models (RSFMs) amortise large-scale pretraining across tasks and sensors, but most current designs remain ViT based and inherit the quadratic complexity of attention.

Mamba-based alternatives aim to retain representational power while relaxing quadratic scaling once long time series, multispectral stacks, or large tiles are used in pretraining, and recent reviews summarise the rapidly growing literature [343,344].

#### 6.4.1. Transformer-Based Remote-Sensing Foundation Models

Early RSFMs adapt masked autoencoding and continual pretraining from natural images to satellite data with ViT encoders. SatMAE extends MAE to temporal and multispectral Sentinel-2 imagery by jointly masking spatial patches, spectral bands, and time steps, and trains large ViT backbones to reconstruct the missing content, yielding strong downstream performance on classification and segmentation [345]. However, attention remains the dominant cost and in practice constrains tile size and temporal depth.

Scale-MAE introduces a scale-aware MAE in which multi-resolution patches are tokenised and jointly masked so that the encoder must infer cross-scale relations, improving scale invariance but leaving attention complexity unchanged [346]. GFM (Geospatial Foundation Model) adopts continual pretraining: a ViT backbone is first trained on ImageNet-22k and then adapted to large geospatial corpora such as GeoPile using teacher–student distillation and progressive fine-tuning, reducing effective pretraining time compared with SatMAE-style training while retaining the same quadratic profile in sequence length [347]. Collectively, these Transformer-based RSFMs demonstrate that MAE-style objectives transfer well to EO but leave scaling to longer sequences and larger scenes largely unresolved.

#### 6.4.2. Mamba-Based RSFMs and High-Resolution Backbones

Mamba-based RSFMs replace attention with state-space layers while keeping objectives similar to SatMAE or GFM. SatMamba swaps the ViT encoder for a visual SSM backbone within the MAE framework and reports comparable or slightly better performance than ViT-MAE on fMoW, with faster convergence of reconstruction loss but also larger parameter counts and depth, which complicates strict same-budget comparisons [346,348]. DynamicVis targets 2048×2048 optical tiles and combines selective state-space modules with dynamic region perception so that computation is concentrated on informative regions; on such inputs it achieves order-of-magnitude reductions in latency and memory relative to ViT backbones at similar accuracy [289]. RoMA introduces rotation-aware self-supervised pretraining to obtain orientation-robust representations, an important property for nadir and off-nadir views [349]. RingMamba moves toward multi-sensor pretraining by coupling optical and SAR streams through scan-and-scan-couple modules and a mixture of generative and contrastive losses [350]. Geo-Mamba integrates dynamic, static, and categorical geophysical factors into a unified state-space framework, often combined with Kolmogorov–Arnold networks for regression heads, thus linking RSFMs and geophysical forecasting [351]. VMIC uses cross-selective scan Mamba modules as priors in learned image compression and achieves BD-rate reductions of roughly 4–10 % over the VTM standard on remote-sensing imagery, indicating that SSM-based priors can improve rate–distortion performance in bandwidth-limited settings [291]. Across these models, parameter counts typically range from roughly 80 M to 350 M [348–350].

#### 6.4.3. Scaling Laws and Fair Evaluation

Despite this activity, evidence on scaling behaviour in Mamba-based RSFMs remains fragmentary. Most published comparisons differ simultaneously in pretraining datasets, augmentation schemes, and compute budgets, making it difficult to attribute gains to architectural choices. SatMamba, for instance, borrows the ≈800-epoch schedule of SatMAE on fMoW but uses a deeper and larger encoder [346,348], while RingMamba and related models vary backbone width and multi-sensor composition [350]. A critical next step is to design controlled studies in which ViT-MAE and Mamba-MAE encoders are trained under matched token counts, parameter budgets, and optimisation settings and evaluated on the same suite of downstream tasks. Such benchmarks should

explicitly report not only task accuracy but also wall-clock time, GPU hours, memory footprint, and performance as a function of context length; without these ingredients, claims about the superiority of one family over the other remain anecdotal.

#### 6.4.4. Multimodal Alignment via State Modulation

Mamba-based RSFMs also invite a different view of multimodal fusion. Whereas ViT architectures typically rely on explicit cross-attention between modalities, many state-space models coordinate modalities through interactions in the hidden state. M3amba couples hyperspectral imagery and LiDAR via a Cross-SS2D module in which the hidden state of one modality modulates the recurrent dynamics of the other, enabling fusion with linear complexity [122]. MFMamba adopts a dual-branch encoder for optical and DSM inputs and introduces an auxiliary Mamba pathway that conditions the optical branch through state modulation [55]. AVS-Mamba extends this design philosophy to audio–visual segmentation with selective temporal and cross-modal scanning [352]. These examples suggest that cross-modal fusion in SSMs can be implemented as state modulation rather than token-to-token attention, which is particularly attractive for EO scenarios with heterogeneous spatial resolution, revisit frequency, and noise characteristics, and is consistent with recent findings on the need for robust cross-modal interaction mechanisms [353].

#### 6.4.5. Summary

Transformer-based remote-sensing foundation models (RSFMs) underpin most current systems and set strong baselines on public benchmarks, but their quadratic attention limits the context length and image size that can be processed efficiently. Mamba-based RSFMs use similar parameter budgets yet achieve linear complexity in sequence length, so they scale more naturally to long sequences and high resolutions, while state modulation offers a convenient mechanism for multimodal fusion. The main open questions concern how these benefits materialise under matched training budgets and whether state-space designs continue to scale favourably at the billion-parameter level.

**Table 4.** Representative Transformer- and Mamba-based remote-sensing foundation models.

Model	Backbone	Modalities	Objective
SatMAE [345]	ViT	MSI + time	Spatio-temporal MAE
Scale-MAE [346]	ViT	Multiscale RGB/MSI	Scale-aware MAE
GFM [347]	ViT	RGB + ancillary	Continual pretraining
SatMamba [348]	Mamba	MSI + time	MAE with SSM encoder
DynamicVis [289]	Mamba	HR optical	Self-/supervised
RoMA [349]	Mamba	Optical	Rotation-aware SSL
RingMamba [350]	Mamba	Optical + SAR	Generative + contrastive
Geo-Mamba [351]	Mamba + KAN	Geophysical predictors	Spatiotemporal prediction
VMIC [291]	Mamba	Latent RS features	Rate–distortion

#### 6.5. Green Computing, Efficiency, and Reproducibility

In most EO papers, Mamba-based architectures are benchmarked by accuracy, FLOPs, and nominal memory use. For operational systems, three additional questions are central: when Mamba is actually more efficient than alternatives, how energy and carbon footprint are measured, and how reliable current implementations are from a numerical and software-engineering standpoint. MambaOut already shows that when the sequence length  $L$  is far below the characteristic threshold  $\tau = 6D$ , gated CNNs can match or slightly exceed VMamba-S accuracy on ImageNet while using less compute [322]. Datacentre GPUs such as A100 or H100 and optimised attention kernels like FlashAttention further reduce the practical penalty of quadratic attention and offer low-precision formats with high performance per watt [354–356]. In these regimes, a tuned Transformer can equal

or surpass current Mamba implementations in wall-clock time, so claims of “efficiency” require careful qualification.

Numerical behaviour is a second concern. Sparse Mamba demonstrates that diagonal state-transition matrices are not automatically stable and that unconstrained eigenvalues may cause divergence or gradient pathologies [324], while analyses of Mamba-3 suggest that hardware-driven simplifications of the state parametrisation can weaken long-range memory [15]. Mixed-precision training with FP16 or BF16 and dynamic loss scaling reduces memory and cost on tensor-core GPUs, but its interaction with long-horizon SSM dynamics in EO has barely been explored [357]. Papers introducing new Mamba variants for restoration or time-series analysis should therefore report not only accuracy but also failure modes, divergence cases, and the effect of precision and sequence length on training and inference.

Energy and carbon accounting introduce additional uncertainty. Work on sustainable machine learning has proposed CO<sub>2</sub> estimators that combine runtime, hardware, and regional electricity mix [358–360], yet Bouza et al. show that software-based tools may deviate from hardware wattmeters by up to several hundred percent for CPU and system-level power [361]. Energy or CO<sub>2</sub> numbers for Mamba-based RS models should therefore be interpreted as method-dependent estimates rather than precise measurements. To make “green” claims credible, future work should justify SSM use by sequence length and task regime, report efficiency with explicit hardware and measurement protocols, and release code, configurations, and checkpoints so that accuracy–efficiency trade-offs are independently verifiable.

## 7. Conclusions

Mamba-based visual state-space models introduce linear-time, content-dependent recurrence over long visual, spectral, and temporal sequences. For EO, this mechanism is not a drop-in replacement for existing backbones but a structured operator whose impact depends on how data are serialised, how it is coupled with convolutions and attention, and how physical priors and efficiency constraints are encoded in the architecture.

Three conclusions emerge.

- **The natural regime for Mamba in EO is long-context, high-throughput modelling.** Across hyperspectral analysis, multi-source fusion, dense perception, and restoration, the most convincing gains appear when models must maintain global context over very large tiles, long image time series, or high-dimensional spectral–spatial sequences under realistic memory budgets. In these regimes, linear-time recurrence allows global dependencies to be modelled without the quadratic overhead of full attention, provided that scanning reflects sensing geometry rather than treating EO data as arbitrary one-dimensional tokens. By contrast, for patch-wise classification, shallow detectors, and other short-sequence tasks, current evidence—including MambaOut, which reports competitive or superior performance of gated CNNs at modest token lengths while using less compute [322]—indicates that Mamba is not automatically the most economical choice. Treating state-space layers as a universal upgrade is therefore difficult to justify empirically; their use should be argued from sequence length, context requirements, and deployment constraints.
- **Current EO practice favours scan-aware hybrid architectures over purely SSM backbones.** The strongest systems rarely rely on Mamba alone. CNN–Mamba and Transformer–Mamba hybrids exploit complementary inductive biases: convolutions and local attention handle edges, textures, registration errors, and sensor noise at high spatial resolutions, while Mamba branches propagate information along carefully designed spatial, spectral, temporal, or multimodal trajectories. This division of labour underpins RS-Mamba, RS3Mamba, SITSMamba, CSFMamba, VmambaIR, and many domain-specific networks for HSI classification, HSI–LiDAR fusion, very-high-resolution segmentation, and spatiotemporal reconstruction. In these models, scan design is not an implementation detail but an explicit modelling choice: centre-focused spirals, cross and omnidirectional scans, graph- or superpixel-guided traversals, and transform-

domain paths all encode assumptions about where relevant context lies and how it should be propagated. In practice, robust designs treat Mamba as a flexible context engine inserted at stages where tokens already summarise larger receptive fields or long time windows, rather than as a wholesale replacement for convolutions or attention.

- **Physics-aware SSMs and Mamba-based foundation models define the next phase, but they must be held to higher standards of stability, efficiency, and reproducibility.** Because state-space layers are rooted in dynamical systems, they are natural candidates for physics-informed EO modelling, including long-horizon forecasting, spatiotemporal downscaling, and inverse problems where conservation laws, radiative transfer, or motion models provide strong priors. At the same time, the community is moving toward large, multimodal foundation models in which Mamba-based encoders are combined with contrastive, generative, or instruction-tuned objectives on global EO archives. Works such as SatMAE, Scale-MAE, GFM, SatMamba, RoMA, and RingMamba indicate that Mamba can act as the visual backbone of such systems when attention becomes prohibitively expensive, while still enabling alignment across sensors and tasks. The challenge now is not simply to scale these models, but to do so with explicit analyses of numerical stability, calibration, and energy use, and with open checkpoints and code so that accuracy–efficiency trade-offs can be independently verified.

On this basis, it would be premature to claim that Mamba will become the dominant backbone for EO. The available results support a more nuanced position. Mamba-style SSMs are already indispensable in certain high-context regimes, but they coexist with—and often rely on—carefully engineered convolutional and attention modules. More broadly, state-space thinking has begun to reshape how the community formulates EO problems: sequences are defined along physically meaningful scan paths; long-range couplings are treated as design primitives rather than afterthoughts; and hybrid backbones are evaluated not only by benchmark accuracy, but also by whether they respect sensing geometry and deployment budgets.

Looking forward, progress will depend as much on disciplined experimental practice as on new architectures. Systematic benchmarks that vary sequence length, scan strategy, and hardware platform are needed to delineate when SSMs are genuinely advantageous. Physics-informed SSMs should be evaluated against strong baselines in data assimilation, downscaling, and hazard monitoring, not only on generic vision datasets. Foundation models that build on Mamba must report scaling behaviour, robustness across regions and sensors, and carbon and energy footprints alongside task performance. If these conditions are met, Mamba-style architectures are well positioned to become core components of EO pipelines—not because they are fashionable, but because their structured recurrence, when used judiciously, addresses concrete limitations of CNNs and Transformers under the combined pressures of high spatial resolution, high spectral dimensionality, and dense temporal sampling.

**Author Contributions:** Zefeng Li: Conceptualization, Investigation (literature search and analysis), Writing – original draft, Visualization. Zhao Long: Validation, Writing – review & editing. Yihang Lu: Validation, Visualization, Writing – review & editing. Ma Yue: Validation, Writing – review & editing. Guoqing Li: Conceptualization, Supervision, Project administration, Funding acquisition, Writing – review & editing. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Earth Observation Data Center, grant number 2024YFB3908404-03. The APC was funded by the same grant.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.
2. Zhu, X. X., Tuia, D., Mou, L., Xia, G. S., Zhang, L., Xu, F., & Fraundorfer, F. (2017). Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4), 8–36.
3. Li, Y., Zhang, H., Xue, X., Jiang, Y., & Shen, Q. (2018). Deep learning for remote sensing image classification: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(6), e1264.
4. Zhang, L., Zhang, L., & Du, B. (2016). Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geoscience and Remote Sensing Magazine*, 4(2), 22–40.
5. Chen, Y., Lin, Z., Zhao, X., Wang, G., & Gu, Y. (2016). Deep learning-based classification of hyperspectral data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(2), 583–594.
6. Kampffmeyer, M., Salberg, A. B., & Jenssen, R. (2016). Semantic segmentation of small objects and modeling of uncertainty in urban remote sensing images using deep convolutional neural networks. *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 680–688.
7. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.
8. Dosovitskiy, A. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.
9. Paheding, S., Saleem, A., Siddiqui, M. F. H., Rawashdeh, N., Essa, A., & Reyes, A. A. (2024). Advancing horizons in remote sensing: a comprehensive survey of deep learning models and applications in image classification and beyond. *Neural Computing and Applications*, 36(27), 16727-16767.
10. Aleissae, A. A., Kumar, A., Anwer, R. M., Khan, S., Cholakkal, H., Xia, G. S., & Khan, F. S. (2023). Transformers in remote sensing: A survey. *Remote Sensing*, 15(7), 1860.
11. Fichtl, A. M., Bohn, J., Kelber, J., Mosca, E., & Groh, G. (2025). The End of Transformers? On Challenging Attention and the Rise of Sub-Quadratic Architectures. *arXiv preprint arXiv:2510.05364*.
12. Gu, A., Goel, K., & Ré, C. (2022). Efficiently modeling long sequences with structured state spaces. *arXiv preprint arXiv:2111.00396*.
13. Gu, A., & Dao, T. (2024, May). Mamba: Linear-time sequence modeling with selective state spaces. In *First conference on language modeling*.
14. Gu, A., & Dao, T. (2024, May). Mamba: Linear-time sequence modeling with selective state spaces. In *First conference on language modeling*.
15. Anonymous Authors. (2025). *Mamba-3: Improved Sequence Modeling Using State-Space Systems*. *arXiv preprint arXiv:2502.01345*.
16. Bao, M., Lyu, S., Xu, Z., Zhou, H., Ren, J., Xiang, S., ... & Cheng, G. (2025). Vision Mamba in Remote Sensing: A Comprehensive Survey of Techniques, Applications and Outlook. *arXiv preprint arXiv:2505.00630*.
17. Xu, R., Yang, S., Wang, Y., Cai, Y., Du, B., & Chen, H. (2024). Visual mamba: A survey and new outlooks. *arXiv preprint arXiv:2404.18861*.
18. Rahman, M. M., Tutul, A. A., Nath, A., Laishram, L., Jung, S. K., & Hammond, T. (2024). Mamba in vision: A comprehensive survey of techniques and applications. *arXiv preprint arXiv:2410.03105*.
19. Liu, X., Zhang, C., Huang, F., Xia, S., Wang, G., & Zhang, L. (2025). Vision mamba: A comprehensive survey and taxonomy. *IEEE Transactions on Neural Networks and Learning Systems*.
20. Zhang, H., Zhu, Y., Wang, D., Zhang, L., Chen, T., Wang, Z., & Ye, Z. (2024). A survey on visual mamba. *Applied Sciences*, 14(13), 5683.
21. Patro, B. N., & Agneeswaran, V. S. (2025). Mamba-360: Survey of state space models as transformer alternative for long sequence modelling: Methods, applications, and challenges. *Engineering Applications of Artificial Intelligence*, 159, 111279.
22. Gupta, A., Gu, A., & Berant, J. (2022). Diagonal state spaces are as effective as structured state spaces. *Advances in neural information processing systems*, 35, 22982-22994.
23. Gu, A., Goel, K., Gupta, A., & Ré, C. (2022). On the parameterization and initialization of diagonal state space models. *Advances in Neural Information Processing Systems*, 35, 35971-35983.

24. Smith, J. T., Warrington, A., & Linderman, S. W. (2023, January). Simplified State Space Layers for Sequence Modeling. In *ICLR*.
25. Hasani, R., Lechner, M., Wang, T. H., Chahine, M., Amini, A., & Rus, D. (2022). Liquid structural state-space models. *arXiv preprint arXiv:2209.12951*.
26. Ma, X., Zhou, C., Kong, X., He, J., Gui, L., Neubig, G., ... & Zettlemoyer, L. (2022). Mega: Moving average equipped gated attention. *arXiv preprint arXiv:2209.10655*.
27. Li, Y., Cai, T., Zhang, Y., Chen, D., & Dey, D. (2022). What makes convolutional models great on long sequence modeling?. *arXiv preprint arXiv:2210.09298*.
28. Orvieto, A., Smith, S. L., Gu, A., Fernando, A., Gulcehre, C., Pascanu, R., & De, S. (2023, July). Resurrecting recurrent neural networks for long sequences. In *International Conference on Machine Learning* (pp. 26670-26698). PMLR.
29. Poli, M., Massaroli, S., Nguyen, E., Fu, D. Y., Dao, T., Baccus, S., ... & Ré, C. (2023, July). Hyena hierarchy: Towards larger convolutional language models. In *International Conference on Machine Learning* (pp. 28043-28078). PMLR.
30. Yang, C., Chen, Z., Espinosa, M., Ericsson, L., Wang, Z., Liu, J., & Crowley, E. J. (2024). Plainmamba: Improving non-hierarchical mamba in visual recognition. *arXiv preprint arXiv:2403.17695*.
31. Zhu, L., Liao, B., Zhang, Q., Wang, X., Liu, W., & Wang, X. (2024). Vision mamba: Efficient visual representation learning with bidirectional state space model. *arXiv preprint arXiv:2401.09417*.
32. Liu, Y., Tian, Y., Zhao, Y., Yu, H., Xie, L., Wang, Y., ... & Liu, Y. (2024). Vmamba: Visual state space model. *Advances in neural information processing systems*, 37, 103031-103063.
33. Shi, Y., Xia, B., Jin, X., Wang, X., Zhao, T., Xia, X., ... & Yang, W. (2025). Vmambair: Visual state space model for image restoration. *IEEE Transactions on Circuits and Systems for Video Technology*.
34. Hatamizadeh, A., & Kautz, J. (2025). Mambavision: A hybrid mamba-transformer vision backbone. In *Proceedings of the Computer Vision and Pattern Recognition Conference* (pp. 25261-25270).
35. Wang, F., Wang, J., Ren, S., Wei, G., Mei, J., Shao, W., ... & Xie, C. (2025). Mamba-Reg: Vision Mamba Also Needs Registers. In *Proceedings of the Computer Vision and Pattern Recognition Conference* (pp. 14944-14953).
36. Behrouz, A., Santacatterina, M., & Zabih, R. (2024). Mambamixer: Efficient selective state space models with dual token and channel selection. *arXiv preprint arXiv:2403.19888*.
37. Patro, B. N., & Agneeswaran, V. S. (2024). Simba: Simplified mamba-based architecture for vision and multivariate time series. *arXiv preprint arXiv:2403.15360*.
38. Hu, V. T., Baumann, S. A., Gui, M., Grebenkova, O., Ma, P., Fischer, J., & Ommer, B. (2024, September). Zigma: A dit-style zigzag mamba diffusion model. In *European conference on computer vision* (pp. 148-166). Cham: Springer Nature Switzerland.
39. Mao, J., Ma, H., & Liang, Y. (2025). BiMambaHSI: Bidirectional Spectral-Spatial State Space Model for Hyperspectral Image Classification. *Remote Sensing*.
40. Tang, X., Yao, Y., Ma, J., Zhang, X., Yang, Y., Wang, B., & Jiao, L. (2025). SpiralMamba: Spatial-Spectral Complementary Mamba with Spatial Spiral Scan for Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*.
41. Zhao, S., Chen, H., Zhang, X., Xiao, P., Bai, L., & Ouyang, W. (2024). Rs-mamba for large remote sensing image dense prediction. *IEEE Transactions on Geoscience and Remote Sensing*.
42. Li, T., Li, C., Lyu, J., Pei, H., Zhang, B., Jin, T., & Ji, R. (2025). DAMamba: Vision State Space Model with Dynamic Adaptive Scan. *arXiv preprint arXiv:2502.12627*.
43. Xie, F., Zhang, W., Wang, Z., & Ma, C. (2024). Quadmamba: Learning quadtree-based selective scan for visual state space model. *Advances in Neural Information Processing Systems*, 37, 117682-117707.
44. Li, B., Xiao, H., & Tang, L. (2025). Scaling Vision Mamba Across Resolutions via Fractal Traversal. *arXiv preprint arXiv:2505.14062*.
45. Zhao, M., Zhang, C., Yue, P., Cai, C., & Ye, F. (2025). MDA-RSM: multi-directional adaptive remote sensing mamba for building extraction. *GIScience & Remote Sensing*, 62(1), 2568776.
46. Wang, T., Bai, T., Xu, C., Liu, B., Zhang, E., Huang, J., & Zhang, H. (2025). AtrousMamaba: An Atrous-Window Scanning Visual State Space Model for Remote Sensing Change Detection. *arXiv preprint arXiv:2507.16172*.

47. Xiao, Y., Yuan, Q., Jiang, K., Chen, Y., Zhang, Q., & Lin, C. W. (2024). Frequency-assisted mamba for remote sensing image super-resolution. *IEEE Transactions on Multimedia*.
48. Zhang, Z., Hu, Z., Cao, B., Li, P., Su, Q., Dong, Z., & Wang, T. (2025). Wiener filter-based Mamba for Remote Sensing Image Super-Resolution with Novel Degradation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
49. Rong, Z., Zhao, Z., Wang, Z., & Ma, L. (2025). FaRMamba: Frequency-based learning and Reconstruction aided Mamba for Medical Segmentation. *arXiv preprint arXiv:2507.20056*.
50. Lu, D., Gao, K., Li, J., Zhang, D., & Xu, L. (2025). Exploring Token Serialization for Mamba-Based LiDAR Point Cloud Segmentation. *IEEE Transactions on Geoscience and Remote Sensing*.
51. Qin, X., Su, X., & Zhang, L. (2024). SITSMamba for crop classification based on satellite image time series. *arXiv preprint arXiv:2409.09673*.
52. Zhu, Q., Fang, Y., Cai, Y., Chen, C., & Fan, L. (2024). Rethinking scanning strategies with vision mamba in semantic segmentation of remote sensing imagery: An experimental study. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
53. Wang, Z., Zheng, J. Q., Zhang, Y., Cui, G., & Li, L. (2024). Mamba-unet: Unet-like pure visual mamba for medical image segmentation. *arXiv preprint arXiv:2402.05079*.
54. Ma, X., Zhang, X., & Pun, M. O. (2024). Rs3mamba: Visual state space model for remote sensing image semantic segmentation. *IEEE Geoscience and Remote Sensing Letters*, 21, 1-5.
55. Wang, Y., Cao, L., & Deng, H. (2024). MFMamba: A mamba-based multi-modal fusion network for semantic segmentation of remote sensing images. *Sensors (Basel, Switzerland)*, 24(22), 7266.
56. Liu, M., Dan, J., Lu, Z., Yu, Y., Li, Y., & Li, X. (2024). CM-UNet: Hybrid CNN-Mamba UNet for remote sensing image semantic segmentation. *arXiv preprint arXiv:2405.10530*.
57. Xiao, P., Dong, Y., Zhao, J., Peng, T., Geiß, C., Zhong, Y., & Taubenböck, H. (2025). MF-Mamba: Multi-Scale Convolution and Mamba Fusion Model for Semantic Segmentation of Remote Sensing Imagery. *IEEE Transactions on Geoscience and Remote Sensing*.
58. He, Y., Tu, B., Liu, B., Li, J., & Plaza, A. (2025). HSI-MFormer: Integrating Mamba and Transformer Experts for Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*.
59. Chen, X., Hu, W., Dong, X., Lin, S., Chen, Z., Cao, M., ... & Liang, X. (2025). Transmamba: Fast universal architecture adaption from transformers to mamba. *arXiv preprint arXiv:2502.15130*.
60. Li, Y., Xie, R., Yang, Z., Sun, X., Li, S., Han, W., ... & Jiang, J. (2025). Transmamba: Flexibly switching between transformer and mamba. *arXiv preprint arXiv:2503.24067*.
61. Li, J., Liu, Z., Liu, S., & Wang, H. (2025). MBSSNet: A Mamba-Based Joint Semantic Segmentation Network for Optical and SAR Images. *IEEE Geoscience and Remote Sensing Letters*.
62. Zhang, Q., Zhang, X., Quan, C., Zhao, T., Huo, W., & Huang, Y. (2025). Mamba-STFM: A Mamba-Based Spatiotemporal Fusion Method for Remote Sensing Images. *Remote Sensing*, 17(13), 2135.
63. Li, Z., Wu, J., Zhang, Y., & Yan, Y. (2025). MHCmamba: Multiscale Hybrid Convolution Mamba Network for Hyperspectral and LiDAR Image Classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
64. Wang, W., Yu, P., Li, M., Zhong, X., He, Y., Su, H., & Zhou, Y. (2025). Tdfnet: Twice decoding v-mamba-cnn fusion features for building extraction. *Geo-spatial Information Science*, 1-20.
65. Zhao, Z., & He, P. (2024). Yolo-mamba: object detection method for infrared aerial images. *Signal, Image and Video Processing*, 18(12), 8793-8803.
66. Huang, L., Tan, J., & Chen, Z. (2024). Mamba-UAV-SegNet: A Multi-Scale Adaptive Feature Fusion Network for Real-Time Semantic Segmentation of UAV Aerial Imagery. *Drones*, 8(11), 671.
67. Zhang, G., Zhang, Z., Deng, J., Bian, L., & Yang, C. (2024). S<sup>2</sup>CrossMamba: Spatial-Spectral Cross-Mamba for Multimodal Remote Sensing Image Classification. *IEEE Geoscience and Remote Sensing Letters*.
68. Luo, L., Zhang, Y., Xu, Y., Yue, T., & Wang, Y. (2025). A VMamba-based Spatial-Spectral Fusion Network for Remote Sensing Image Classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
69. He, Y., Tu, B., Jiang, P., Liu, B., Li, J., & Plaza, A. (2025). Classification of Multisource Remote Sensing Data Using Slice Mamba. *IEEE Transactions on Geoscience and Remote Sensing*.

70. Liu, C., Wang, F., Jia, Q., Liu, L., & Zhang, T. (2025). AMamNet: Attention-Enhanced Mamba Network for Hyperspectral Remote Sensing Image Classification. *Atmosphere*, 16(5), 541.
71. Yang, X., Yang, J., Li, L., Xue, S., Shi, H., Tang, H., & Huang, X. (2025). HG-Mamba: A Hybrid Geometry-Aware Bidirectional Mamba Network for Hyperspectral Image Classification. *Remote Sensing*, 17(13), 2234.
72. Yang, X., Li, L., Xue, S., Li, S., Yang, W., Tang, H., & Huang, X. (2025). MRFP-Mamba: Multi-Receptive Field Parallel Mamba for Hyperspectral Image Classification. *Remote Sensing*, 17(13), 2208.
73. He, Y., Tu, B., Liu, B., Li, J., & Plaza, A. (2024). 3DSS-Mamba: 3D-spectral-spatial mamba for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*.
74. Li, G., & Ye, M. (2025). MVNet: Hyperspectral Remote Sensing Image Classification Based on Hybrid Mamba-Transformer Vision Backbone Architecture. *arXiv preprint arXiv:2507.04409*.
75. Sheng, J., Zhou, J., Wang, J., Ye, P., & Fan, J. (2024). Dualmamba: A lightweight spectral-spatial mamba-convolution network for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*.
76. Yao, J., Hong, D., Li, C., & Chanussot, J. (2024). Spectralmamba: Efficient mamba for hyperspectral image classification. *arXiv preprint arXiv:2404.08489*.
77. Zhang, T., Xuan, C., Cheng, F., Tang, Z., Gao, X., & Song, Y. (2025). CenterMamba: Enhancing Semantic Representation with Center-Scan Mamba Network for Hyperspectral Image Classification. *Expert Systems with Applications*, 127985.
78. Bai, Y., Wu, H., Zhang, L., & Guo, H. (2025). Lightweight Mamba Model Based on Spiral Scanning Mechanism for Hyperspectral Image Classification. *IEEE Geoscience and Remote Sensing Letters*.
79. Wang, G., Zhang, X., Peng, Z., Zhang, T., & Jiao, L. (2025). S 2 mamba: A spatial-spectral state space model for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*.
80. Zhang, H., Liu, H., Shi, Z., Mao, S., & Chen, N. (2025). ConvMamba: Combining Mamba with CNN for hyperspectral image classification. *Neurocomputing*, 131016.
81. Huang, L., Chen, Y., & He, X. (2024). Spectral-spatial mamba for hyperspectral image classification. *arXiv preprint arXiv:2404.18401*.
82. Ahmad, M., Butt, M. H. F., Usama, M., Altuwaijri, H. A., Mazzara, M., Distefano, S., & Khan, A. M. (2025). Multi-head spatial-spectral mamba for hyperspectral image classification. *Remote Sensing Letters*, 16(4), 339-353.
83. He, Y., Tu, B., Jiang, P., Liu, B., Li, J., & Plaza, A. (2024). IGroupSS-Mamba: Interval group spatial-spectral mamba for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*.
84. Lu, S., Zhang, M., Huo, Y., Wang, C., Wang, J., & Gao, C. (2024). SSUM: Spatial-spectral unified Mamba for hyperspectral image classification. *Remote Sensing*, 16(24), 4653.
85. Duan, Y., Yu, L., Chen, J., Zeng, Z., Li, J., & Plaza, A. (2025). A New Multiscale Superpixel Mamba for Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*.
86. Song, Q., Tu, B., He, Y., Liu, B., Li, J., & Plaza, A. (2025). Superpixel-Integrated Dual-Stage Mamba for Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*.
87. Yang, A., Li, M., Ding, Y., Fang, L., Cai, Y., & He, Y. (2024). Graphmamba: An efficient graph structure learning vision mamba for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*.
88. Wang, Y., Liu, L., Xiao, J., Yu, D., Tao, Y., & Zhang, W. (2025). MambaHSI+: Multidirectional State Propagation for Efficient Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*.
89. Wang, Y., Liu, L., Xiao, J., Yu, D., Tao, Y., & Zhang, W. (2025). MambaHSI+: Multidirectional State Propagation for Efficient Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*.
90. Ming, R., Chen, N., Peng, J., Sun, W., & Ye, Z. (2025). Semantic Tokenization-Based Mamba for Hyperspectral Image Classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.

91. Zhao, F., Zhang, Z., Huang, L., Hai, Y., Fu, Z., & Tang, B. H. (2025). MHS-Mamba: A Multi-Hierarchical Semantic Model for UAV Hyperspectral Image Classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
92. Du, A., Zhao, G., Cao, M., Wang, Y., Dong, A., Lv, G., ... & Dong, X. (2025). Cross-domain hyperspectral image classification via mamba-CNN and knowledge distillation. *IEEE Transactions on Geoscience and Remote Sensing*.
93. Huang, X., Zhang, Y., Luo, F., & Dong, Y. (2024). Dynamic token augmentation mamba for cross-scene classification of hyperspectral image. *IEEE Transactions on Geoscience and Remote Sensing*.
94. Xu, Y., Wang, D., Jiao, H., Zhang, L., & Zhang, L. (2025). MambaMoE: Mixture-of-Spectral-Spatial-Experts State Space Model for Hyperspectral Image Classification. *arXiv preprint arXiv:2504.20509*.
95. Ahmad, M., Butt, M. H. F., Usama, M., Mazzara, M., Distefano, S., Khan, A. M., & Hong, D. (2025). Hybrid State-Space and GRU-based Graph Tokenization Mamba for Hyperspectral Image Classification. *arXiv preprint arXiv:2502.06427*.
96. Wang, H., Zhuang, P., Zhang, X., & Li, J. (2025). DBMGNet: A Dual-Branch Mamba-GCN Network for Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*.
97. Liao, J., & Wang, L. (2025). HyperspectralMamba: A Novel State Space Model Architecture for Hyperspectral Image Classification. *Remote Sensing*, 17(15), 2577.
98. Sun, M., Zhang, J., He, X., & Zhong, Y. (2024). Bidirectional mamba with dual-branch feature extraction for hyperspectral image classification. *Sensors*, 24(21), 6899.
99. Liu, Y., Zhang, Y., Guo, Y., & Li, Y. (2024). Lightweight spatial-spectral shift module with multi-head MambaOut for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
100. Sun, M., Wang, L., Jiang, S., Cheng, S., & Tang, L. (2025). HyperSMamba: A Lightweight Mamba for Efficient Hyperspectral Image Classification. *Remote Sensing*, 17(12), 2008.
101. Liang, L., Zhang, J., Duan, P., Kang, X., Wu, T. X., Li, J., & Plaza, A. (2025). LKMA: Learnable Kernel and Mamba with Spatial-Spectral Attention Fusion for Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*.
102. Arya, R. K., Jain, S., Chattopadhyay, P., & Srivastava, R. (2025). HSIRMamba: An effective feature learning for hyperspectral image classification using residual Mamba. *Image and Vision Computing*, 154, 105387.
103. Paoletti, M. E., Wu, Z., Zheng, P., Hong, D., & Haut, J. M. (2025). DenseMixerMamba: Residual Mixing for Spectral-Spatial Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*.
104. Wang, C., Huang, J., Lv, M., Du, H., Wu, Y., & Qin, R. (2024). A local enhanced mamba network for hyperspectral image classification. *International Journal of Applied Earth Observation and Geoinformation*, 133, 104092.
105. Zhang, J., Sun, M., & Chang, S. (2025). Spatial and Spectral Structure-Aware Mamba Network for Hyperspectral Image Classification. *Remote Sensing*, 17(14), 2489.
106. Ahmad, M., Usama, M., Mazzara, M., & Distefano, S. (2024). Wavemamba: Spatial-spectral wavelet mamba for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*.
107. Ahmad, M., Butt, M. H. F., Khan, A. M., Mazzara, M., Distefano, S., Usama, M., ... & Hong, D. (2025). Spatial-spectral morphological mamba for hyperspectral image classification. *Neurocomputing*, 636, 129995.
108. Zhang, H., Xu, X., Li, S., & Plaza, A. (2025). Wavelet Decomposition-Based Spectral-Spatial Mamba Network For Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*.
109. Zhuang, P., Zhang, X., Wang, H., Zhang, T., Liu, L., & Li, J. (2025). Fahm: Frequency-aware hierarchical mamba for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
110. Zhu, M., Wang, H., Meng, Y., Xu, S., Lin, Y., Shan, Z., & Ma, Z. (2025). Self-Supervised Mamba for Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*.
111. Ding, H., Liu, J., Wang, Z., Peng, Y., & Li, H. (2025). Mamba-Driven Multi-Scale Spatial-Spectral Fusion Network for Few-Shot Hyperspectral Image Classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.

112. Wang, Q., Jiang, X., & Xu, G. (2025). CSFMamba: Cross State Fusion Mamba Operator for Multimodal Remote Sensing Image Classification. *arXiv preprint arXiv:2509.00677*.
113. Zhang, G., Zhang, Z., Deng, J., Bian, L., & Yang, C. (2024). S2CrossMamba: Spatial-Spectral Cross-Mamba for Multimodal Remote Sensing Image Classification. *IEEE Geoscience and Remote Sensing Letters*.
114. Xing, Y., Jia, Y., Gao, S., Hu, J., & Huang, R. (2025). Frequency-enhanced mamba for remote sensing change detection. *IEEE Geoscience and Remote Sensing Letters*.
115. Gao, F., Jin, X., Zhou, X., Dong, J., & Du, Q. (2025). MSFMamba: Multi-scale feature fusion state space model for multi-source remote sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*.
116. Li, S., & Huang, S. (2024). AFA-Mamba: adaptive feature alignment with global-local mamba for hyperspectral and LiDAR data classification. *Remote Sensing*, 16(21), 4050.
117. Pan, H., Zhao, R., Ge, H., Liu, M., & Zhang, Q. (2025). Multi-Modal Fusion Mamba Network for Joint Land Cover Classification Using Hyperspectral and LiDAR Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
118. Li, D., Li, B., & Liu, Y. (2025). Mamba Cross-Modal Information Fusion Self-Distillation Model for Joint Classification of LiDAR and Hyperspectral Data. *IEEE Transactions on Geoscience and Remote Sensing*.
119. Shi, C., Zhu, F., Shi, K., Wang, L., & Pan, H. (2025). TBi-Mamba: Rethinking Joint Classification of Hyperspectral and LiDAR Data With Bidirectional Mamba. *IEEE Transactions on Geoscience and Remote Sensing*.
120. Li, Z., Wu, J., Zhang, Y., & Yan, Y. (2025). CMFNet: Cross Mamba Fusion Network for Hyperspectral and LiDAR Data Classification. *IEEE Transactions on Geoscience and Remote Sensing*.
121. Xie, Z., Lv, L., Gao, H., Xu, S., & Xie, H. (2025). Dual-Feature Attention Hybrid GCN Mamba Network for Joint Hyperspectral and LiDAR Classification. *IEEE Transactions on Geoscience and Remote Sensing*.
122. Cao, M., Xie, W., Zhang, X., Zhang, J., Jiang, K., Lei, J., & Li, Y. (2025). M 3 amba: CLIP-driven Mamba Model for Multi-modal Remote Sensing Classification. *IEEE Transactions on Circuits and Systems for Video Technology*.
123. Ye, F., Tan, S., Huang, W., Xu, X., & Jiang, S. (2025). MambaTriNet: A Mamba based Tri-backbone multimodal remote sensing image semantic segmentation model. *IEEE Geoscience and Remote Sensing Letters*.
124. Liao, D., Wang, Q., Lai, T., & Huang, H. (2024). Joint classification of hyperspectral and lidar data base on mamba. *IEEE Transactions on Geoscience and Remote Sensing*.
125. He, X., Han, X., Chen, Y., & Huang, L. (2025). A light-weighted fusion vision mamba for multimodal remote sensing data classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
126. Yue, Z., Xu, J., Yan, Y., & Su, M. (2025). TFFNet: Transform Fusion Fuzzy Network for Multimodal Remote Sensing Classification. *IEEE Geoscience and Remote Sensing Letters*.
127. Peng, S., Zhu, X., Deng, H., Deng, L. J., & Lei, Z. (2024). Fusionmamba: Efficient remote sensing image fusion with state space model. *IEEE Transactions on Geoscience and Remote Sensing*.
128. Wu, H., Sun, Z., Qi, J., Zhan, T., Xu, Y., & Wei, Z. (2025). Spatial-Spectral Cross Mamba Network for Hyperspectral and Multispectral Image Fusion. *IEEE Transactions on Geoscience and Remote Sensing*.
129. Zhao, G., Wu, H., Luo, D., Ou, X., & Zhang, Y. (2024). Spatial spectral interaction super-resolution cnn-mamba network for fusion of satellite hyperspectral and multispectral image. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
130. Zhang, Y., Song, Y., Duan, Q., Yu, N., Li, B., & Gao, X. (2025). S 2 CMamba: A Mamba-based Pan-sharpening Model Incorporating Spatial and Spectral Consistency. *IEEE Transactions on Geoscience and Remote Sensing*.
131. Zhu, C., Deng, S., Song, X., Li, Y., & Wang, Q. (2025). Mamba collaborative implicit neural representation for hyperspectral and multispectral remote sensing image fusion. *IEEE Transactions on Geoscience and Remote Sensing*.
132. Li, Z., Wen, Y., Xiao, S., Qu, J., Li, N., & Dong, W. (2025). A Progressive Registration-Fusion Co-Optimization A-Mamba Network: Towards Deep Unregistered Hyperspectral and Multispectral Fusion. *IEEE Transactions on Geoscience and Remote Sensing*.
133. Xiao, L., Guo, S., Mo, F., Song, Q., Yang, Y., Liu, Y., ... & Dian, R. (2025). Spatial Invertible Network With Mamba-Convolution for Hyperspectral Image Fusion. *IEEE Journal of Selected Topics in Signal Processing*.

134. Zhao, M., Jiang, X., & Huang, B. (2025). STFMamba: Spatiotemporal satellite image fusion network based on visual state space model. *ISPRS Journal of Photogrammetry and Remote Sensing*, 228, 288-304.
135. Bioucas-Dias, J. M., Plaza, A., Dobigeon, N., Parente, M., Du, Q., Gader, P., & Chanussot, J. (2012). Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *IEEE journal of selected topics in applied earth observations and remote sensing*, 5(2), 354-379.
136. Zhang, M., Xie, H., Yang, M., Jiao, Q., Xu, L., & Tan, X. (2025). Mamba-Enhanced Spatial-Spectral Feature Learning for Hyperspectral Unmixing. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
137. Liu, Y., Liu, S., & Wang, H. (2025). Efficient Progressive Mamba Model for Hyperspectral Sequence Unmixing. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
138. Chen, D., Zhang, J., & Li, J. (2025). UNMamba: Cascaded Spatial-Spectral Mamba for Blind Hyperspectral Unmixing. *IEEE Geoscience and Remote Sensing Letters*.
139. Gan, Y., Wei, J., & Xu, M. (2025). Mamba-based spatial-spectral fusion network for hyperspectral unmixing. *Journal of King Saud University Computer and Information Sciences*, 37(3), 32.
140. Qu, K., Wang, H., Ding, M., Luo, X., & Bao, W. (2025). DGMNet: Hyperspectral Unmixing Dual-Branch Network Integrating Adaptive Hop-Aware GCN and Neighborhood Offset Mamba. *Remote Sensing*, 17(14), 2517.
141. Zheng, X., Kuang, Y., Huo, Y., Zhu, W., Zhang, M., & Wang, H. (2025). HTMNet: Hybrid Transformer-Mamba Network for Hyperspectral Target Detection. *Remote Sensing*, 17(17), 3015.
142. Shen, D., Zhu, X., Tian, J., Liu, J., Du, Z., Wang, H., & Ma, X. (2025). HTD-Mamba: Efficient Hyperspectral Target Detection with Pyramid State Space Model. *IEEE Transactions on Geoscience and Remote Sensing*.
143. Li, L., & Wang, B. (2025). DPMN: Deep Prior Mamba Network for Hyperspectral Anomaly Detection. *IEEE Transactions on Geoscience and Remote Sensing*.
144. Fu, X., Zhang, T., Cheng, J., & Jia, S. (2025). MMR-HAD: Multi-scale Mamba Reconstruction Network for Hyperspectral Anomaly Detection. *IEEE Transactions on Geoscience and Remote Sensing*.
145. Li, F., Wang, X., Wang, H., Karimian, H., Shi, J., & Zha, G. (2025). LMVMamba: A Hybrid U-Shape Mamba for Remote Sensing Segmentation with Adaptation Fine-Tuning. *Remote Sensing*, 17(19), 3367.
146. Cao, Y., Liu, C., Wu, Z., Zhang, L., & Yang, L. (2025). Remote sensing image segmentation using vision mamba and multi-scale multi-frequency feature fusion. *Remote Sensing*, 17(8), 1390.
147. Zhu, E., Chen, Z., Wang, D., Shi, H., Liu, X., & Wang, L. (2024). Unetmamba: An efficient unet-like mamba for semantic segmentation of high-resolution remote sensing images. *IEEE Geoscience and Remote Sensing Letters*.
148. Du, W. L., Gu, Y., Zhao, J., Zhu, H., Yao, R., & Zhou, Y. (2024). A mamba-diffusion framework for multimodal remote sensing image semantic segmentation. *IEEE Geoscience and Remote Sensing Letters*.
149. Zhou, W., Yang, P., & Liu, Y. (2025). HLMamba: Hybrid Lightweight Mamba-Based Fusion Network for Dense Prediction of Remote Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing*.
150. Sun, H., Liu, J., Yang, J., & Wu, Z. (2025). HMAFNet: Hybrid Mamba-Attention Fusion Network for Remote Sensing Image Semantic Segmentation. *IEEE Geoscience and Remote Sensing Letters*.
151. Zheng, K., Yu, M., Liu, Z., Bao, S., Pan, Z., Song, Y., ... & Xie, Z. (2025). FREQUENCY AND PROMPT LEARNING COOPERATION ENHANCED MAMBA FOR REMOTE SENSING SEMANTIC SEGMENTATION. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
152. Huang, P., Zhang, K., Ma, M., Mei, S., & Wang, J. (2025). Semantic-Geometric Consistency-enforcing with Mamba-augmented Network for Remote Sensing Image Segmentation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
153. Zhu, Q., Cai, Y., Fang, Y., Yang, Y., Chen, C., Fan, L., & Nguyen, A. (2024). Samba: Semantic segmentation of remotely sensed images with state space model. *Heliyon*, 10(19).
154. Mu, J., Zhou, S., & Sun, X. (2024). PPMamba: Enhancing Semantic Segmentation in Remote Sensing Imagery by SS2D. *IEEE Geoscience and Remote Sensing Letters*.
155. Li, M., Xing, Z., Wang, H., Jiang, H., & Xie, Q. (2025). SF-Mamba: A Semantic-flow Foreground-aware Mamba for Semantic Segmentation of Remote Sensing Images. *IEEE MultiMedia*.

156. Fang, X., Liu, Z., Xie, S. A., & Ge, Y. (2025). Semantic Segmentation of High-Resolution Remote Sensing Images Based on RS3Mamba: An Investigation of the Extraction Algorithm for Rural Compound Utilization Status. *Remote Sensing*, 17(20), 3443.
157. Wen, R., Yuan, Y., Xu, X., Yin, S., Chen, Z., Zeng, H., & Wang, Z. (2025). MambaSegNet: A Fast and Accurate High-Resolution Remote Sensing Imagery Ship Segmentation Network. *Remote Sensing*, 17(19), 3328.
158. Yan, L., Feng, Q., Wang, J., Cao, J., Feng, X., & Tang, X. (2025). A multilevel multimodal hybrid mamba-large strip convolution network for remote sensing semantic segmentation. *Remote Sensing*, 17(15), 2696.
159. Qiu, J., Chang, W., Ren, W., Hou, S., & Yang, R. (2025). MMFNet: A Mamba-Based Multimodal Fusion Network for Remote Sensing Image Semantic Segmentation. *Sensors*, 25(19), 6225.
160. Li, H., Pan, H., Liu, X., Ren, J., Du, Z., & Cao, J. (2025). GLVMamba: A Global-Local Visual State Space Model for Remote Sensing Image Segmentation. *IEEE transactions on geoscience and remote sensing*.
161. Hu, Y., Ma, X., Sui, J., & Pun, M. O. (2024). Ppmamba: A pyramid pooling local auxiliary ssm-based model for remote sensing image semantic segmentation. *arXiv preprint arXiv:2409.06309*.
162. Zhang, Q., Geng, G., Zhou, P., Liu, Q., Wang, Y., & Li, K. (2024). Link aggregation for skip connection-mamba: Remote sensing image segmentation network based on link aggregation mamba. *Remote Sensing*, 16(19), 3622.
163. Ma, C., & Wang, Z. (2024). Semi-Mamba-UNet: Pixel-level contrastive and cross-supervised visual Mamba-based UNet for semi-supervised medical image segmentation. *Knowledge-Based Systems*, 300, 112203.
164. Zhu, Q., Li, H., He, L., & Fan, L. (2025). SwinMamba: A hybrid local-global mamba framework for enhancing semantic segmentation of remotely sensed images. *arXiv preprint arXiv:2509.20918*.
165. Wang, L., Li, D., Dong, S., Meng, X., Zhang, X., & Hong, D. (2024). PyramidMamba: Rethinking pyramid feature fusion with selective space state model for semantic segmentation of remote sensing imagery. *arXiv preprint arXiv:2406.10828*.
166. Chen, H., Luo, H., & Wang, C. (2025). AfaMamba: Adaptive Feature Aggregation With Visual State Space Model for Remote Sensing Images Semantic Segmentation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
167. Lin, B., Zou, Z., & Shi, Z. (2025). RSBEV-Mamba: 3D BEV Sequence Modeling for Multi-View Remote Sensing Scene Segmentation. *IEEE Transactions on Geoscience and Remote Sensing*.
168. Li, L., Yi, J., Fan, H., & Lin, H. (2025). A Lightweight Semantic Segmentation Network Based on Self-attention Mechanism and State Space Model for Efficient Urban Scene Segmentation. *IEEE Transactions on Geoscience and Remote Sensing*.
169. Shen, Y., Xiao, L., Chen, J., Du, Q., & Ye, Q. (2025). Learning Cross-task Features with Mamba for Remote Sensing Image Multi-task Prediction. *IEEE Transactions on Geoscience and Remote Sensing*.
170. Yang, Y., Yuan, G., & Li, J. (2025). Dual-Branch Network for Spatial-Channel Stream Modeling Based on the State Space Model for Remote Sensing Image Segmentation. *IEEE Transactions on Geoscience and Remote Sensing*.
171. Zhao, Y., Qiu, L., Yang, Z., Chen, Y., & Zhang, Y. (2025). MGF-GCN: Multimodal interaction Mamba-aided graph convolutional fusion network for semantic segmentation of remote sensing images. *Information Fusion*, 122, 103150.
172. Du, W. L., Tang, S., Zhao, J., Yao, R., & Zhou, Y. (2025). MoViM: A Hybrid CNN Vision Mamba Network for Lightweight Semantic Segmentation of Multimodal Remote Sensing Images. *IEEE Geoscience and Remote Sensing Letters*.
173. Wang, Z., Xu, N., You, Z., & Zhang, S. (2025). DiffMamba: semantic diffusion guided feature modeling network for semantic segmentation of remote sensing images. *GIScience & Remote Sensing*, 62(1), 2484829.
174. Wang, Z., Yi, J., Chen, A., Chen, L., Lin, H., & Xu, K. (2025). Accurate semantic segmentation of very high-resolution remote sensing images considering feature state sequences: From benchmark datasets to urban applications. *ISPRS Journal of Photogrammetry and Remote Sensing*, 220, 824-840.
175. Chai, X., Zhang, W., Li, Z., Zhang, N., & Chai, X. (2025). AECA-FBMamba: A Framework with Adaptive Environment Channel Alignment and Mamba Bridging Semantics and Details. *Remote Sensing*, 17(11), 1935.
176. Li, D., Zhao, J., Chang, C., Chen, Z., & Du, J. (2024). LGMamba: Large-Scale ALS Point Cloud Semantic Segmentation with Local and Global State Space Model. *IEEE Geoscience and Remote Sensing Letters*.

177. Zhou, M., Li, T., Qiao, C., Xie, D., Wang, G., Ruan, N., ... & Shen, H. T. (2025). Dmm: Disparity-guided multispectral mamba for oriented object detection in remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*.
178. Wang, S., Wang, C., Shi, C., Liu, Y., & Lu, M. (2024). Mask-guided mamba fusion for drone-based visible-infrared vehicle detection. *IEEE Transactions on Geoscience and Remote Sensing*.
179. Liu, C., Ma, X., Yang, X., Zhang, Y., & Dong, Y. (2025). COMO: Cross-mamba interaction and offset-guided fusion for multimodal object detection. *Information Fusion*, 103414.
180. Ren, K., Wu, X., Xu, L., & Wang, L. (2024). Remotedet-mamba: A hybrid mamba-cnn network for multimodal object detection in remote sensing images. *arXiv preprint arXiv:2410.13532*.
181. Li, W., Yuan, F., Zhang, H., Lv, Z., & Wu, B. (2024). Hyperspectral object detection based on spatial-spectral fusion and visual mamba. *Remote Sensing*, 16(23), 4482.
182. Rong, Q., Jing, H., & Zhang, M. (2025). Scale Sensitivity Mamba Network for Object Detection in Remote Sensing Images. *IEEE Sensors Journal*.
183. Wu, S., Lu, X., & Guo, C. (2024). YOLOv5\_mamba: unmanned aerial vehicle object detection based on bidirectional dense feedback network and adaptive gate feature fusion. *Scientific Reports*, 14(1), 22396.
184. Wu, S., Lu, X., Guo, C., & Guo, H. (2025). MV-YOLO: An Efficient Small Object Detection Framework Based on Mamba. *IEEE Transactions on Geoscience and Remote Sensing*.
185. Verma, T., Singh, J., Bhartari, Y., Jarwal, R., Singh, S., & Singh, S. (2024). Soar: Advancements in small body object detection for aerial imagery using state space models and programmable gradients. *arXiv preprint arXiv:2405.01699*.
186. Xiao, Z., Li, Z., Cao, J., Liu, X., Kong, Y., & Du, Z. (2025). OriMamba: Remote sensing oriented object detection with state space models. *International Journal of Applied Earth Observation and Geoinformation*, 143, 104731.
187. Chen, J., Wei, J., Wu, G., Yang, J., Shang, J., Guo, H., ... & Zhu, S. (2025). MambaRetinaNet: Improving remote sensing object detection by fusing Mamba and multi-scale convolution. *Applied Computing and Geosciences*, 100305.
188. Tian, B., Lu, Z., Zhang, C., Li, H., & Yu, P. (2025). MSMD-YOLO: Multi-scale and multi-directional Mamba scanning infrared image object detection based on YOLO. *Infrared Physics & Technology*, 106011.
189. Yan, L., He, Z., Zhang, Z., & Xie, G. (2025). LS-MambaNet: Integrating Large Strip Convolution and Mamba Network for Remote Sensing Object Detection. *Remote Sensing*, 17(10), 1721.
190. Tu, H., Wang, W., Guo, Y., & Chen, S. (2025). Mamba-UDA: Mamba Unsupervised Domain Adaptation for SAR Ship Detection. *IEEE Geoscience and Remote Sensing Letters*.
191. Liu, X., Feng, C., Zi, S., Qin, Z., & Guan, Q. (2025). M-ReDet: A mamba-based method for remote sensing ship object detection and fine-grained recognition. *PLoS One*, 20(8), e0330485.
192. Liu, P., Lei, S., & Li, H. C. (2025). Mamba-MOC: A Multicategory Remote Object Counting via State Space Model. *arXiv preprint arXiv:2501.06697*.
193. Wang, Q., Zhou, L., Jin, P., Qu, X., Zhong, H., Song, H., & Shen, T. (2024). TrackingMamba: Visual state space model for object tracking. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
194. Jiang, J., Liao, S., Yang, X., & Shen, K. (2025). EAMNet: Efficient Adaptive Mamba Network for Infrared Small Target Detection. *IEEE Transactions on Geoscience and Remote Sensing*.
195. Li, B., Rao, P., Su, Y., & Chen, X. (2025). HMCNet: A Hybrid Mamba-CNN UNet for Infrared Small Target Detection. *Remote Sensing*, 17(3), 452.
196. Yu, Z., Zhang, Z., Tian, H., Zhou, Q., & Zhang, H. (2025). SBMambaNet: Spatial-BiDirectional Mamba Network for infrared small target detection. *Infrared Physics & Technology*, 105928.
197. Ge, Y., Liang, T., Ren, J., Chen, J., & Bi, H. (2025). Enhanced salient object detection in remote sensing images via dual-stream semantic interactive network. *The Visual Computer*, 41(7), 5153-5169.
198. Yang, W., Yi, Z., Huang, A., Wang, Y., Yao, Y., & Li, Y. (2025). Topology-Aware Hierarchical Mamba for Salient Object Detection in Remote Sensing Imagery. *IEEE Transactions on Geoscience and Remote Sensing*.
199. Li, J., Wang, Z., Xu, N., & Zhang, C. (2025). TSFANet: Trans-Mamba Hybrid Network with Semantic Feature Alignment for Remote Sensing Salient Object Detection. *Remote Sensing*, 17(11), 1902.

200. Xing, G., Wang, M., Wang, F., Sun, F., & Li, H. (2025). Lightweight edge-aware mamba-fusion network for weakly supervised salient object detection in optical remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*.
201. Li, Y., Wang, L., & Chen, S. (2025). SMILE: Spatial-Spectral Mamba Interactive Learning for Infrared Small Target Detection. *IEEE Transactions on Geoscience and Remote Sensing*.
202. Chen, T., Ye, Z., Tan, Z., Gong, T., Wu, Y., Chu, Q., ... & Ye, J. (2024). Mim-istd: Mamba-in-mamba for efficient infrared small target detection. *IEEE Transactions on Geoscience and Remote Sensing*.
203. Yu, C., Yang, H., Ma, L., Yang, J., Jin, Y., Zhang, W., ... & Zhao, Q. (2025). Deep Learning-Based Change Detection in Remote Sensing: A Comprehensive Review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
204. Chen, H., Song, J., Han, C., Xia, J., & Yokoya, N. (2024). ChangeMamba: Remote sensing change detection with spatiotemporal state space model. *IEEE Transactions on Geoscience and Remote Sensing*, 62, 1-20.
205. Liu, S., Wang, S., Zhang, W., Zhang, T., Xu, M., Yasir, M., & Wei, S. (2025). CD-STMamba: Towards Remote Sensing Image Change Detection With Spatio-Temporal Interaction Mamba Model. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
206. Wu, Z., Ma, X., Lian, R., Zheng, K., Ma, M., Zhang, W., & Song, S. (2025). CD-lamba: Boosting remote sensing change detection via a cross-temporal locally adaptive state space model. *arXiv preprint arXiv:2501.15455*.
207. Kaung, J., & Ge, H. (2025). 2DMCG: 2DMambawith Change Flow Guidance for Change Detection in Remote Sensing. *arXiv preprint arXiv:2503.00521*.
208. Xu, Z., Zhu, Y., Dewis, Z., Heffring, M., Alkayid, M., Taleghanidoozdoozan, S., & Xu, L. L. (2025). Knowledge-Aware Mamba for Joint Change Detection and Classification from MODIS Times Series. *arXiv preprint arXiv:2510.09679*.
209. Zhao, J., Xie, J., Zhou, Y., Du, W. L., Yao, R., & El Saddik, A. (2025). ST-Mamba: Spatio-Temporal Synergistic Model for Remote Sensing Change Detection. *IEEE Transactions on Geoscience and Remote Sensing*.
210. Zhou, S., Xu, C., Fan, G., Li, J., Hua, Z., & Zhou, J. (2025). Sprmamba: A mamba-based saliency proportion reconciliatory network with squeezed windows for remote sensing change detection. *IEEE Transactions on Geoscience and Remote Sensing*.
211. Xu, G., Liu, Y., Deng, L., Wang, X., & Zhu, H. (2025). Smnet: A semantic guided mamba network for remote sensing change detection. *IEEE Transactions on Aerospace and Electronic Systems*.
212. Wang, L., Sun, Q., Pei, J., Khan, M. A., Al Dabel, M. M., Al-Otaibi, Y. D., & Bashir, A. K. (2025). Bi-Temporal Remote Sensing Change Detection with State Space Models. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
213. Zhang, H., Chen, K., Liu, C., Chen, H., Zou, Z., & Shi, Z. (2025). CDMamba: Incorporating local clues into mamba for remote sensing image binary change detection. *IEEE Transactions on Geoscience and Remote Sensing*.
214. Liu, Y., Cheng, G., Sun, Q., Tian, C., & Wang, L. (2025). CWmamba: Leveraging CNN-Mamba fusion for enhanced change detection in remote sensing images. *IEEE Geoscience and Remote Sensing Letters*.
215. Feng, Y., Zhuo, L., Zhang, H., & Li, J. (2025). Hybrid-MambaCD: Hybrid Mamba-CNN Network for Remote Sensing Image Change Detection With Region-Channel Attention Mechanism and Iterative Global-Local Feature Fusion. *IEEE Transactions on Geoscience and Remote Sensing*.
216. Dong, Z., Yuan, G., Hua, Z., & Li, J. (2024). ConMamba: CNN and SSM high-performance hybrid network for remote sensing change detection. *IEEE Transactions on Geoscience and Remote Sensing*.
217. Wang, J., Song, J., Zhang, H., Zhang, Z., Ji, Y., Zhang, W., ... & Wang, X. (2025). SPMNet: A Siamese Pyramid Mamba Network for Very-High-Resolution Remote Sensing Change Detection. *IEEE Transactions on Geoscience and Remote Sensing*.
218. Huang, J., Yuan, X., Lam, C. T., Wang, Y., & Xia, M. (2025). LCCDMamba: Visual state space model for land cover change detection of VHR remote sensing images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
219. Zhang, Z., Fan, X., Wang, X., Qin, Y., & Xia, J. (2024). A novel remote sensing image change detection approach based on multi-level state space model. *IEEE Transactions on Geoscience and Remote Sensing*.

220. Chen, Z., Chen, H., Leng, J., Zhang, X., Gao, Q., & Dong, W. (2025). VMMCD: VMamba-Based Multi-Scale Feature Guiding Fusion Network for Remote Sensing Change Detection. *Remote Sensing*, 17(11), 1840.
221. Wang, S., Cheng, D., Yuan, G., & Li, J. (2025). RDSF-Net: Residual Wavelet Mamba-Based Differential Completion and Spatio-Frequency Extraction Remote Sensing Change Detection Network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
222. Wang, S., Yuan, G., & Li, J. (2025). GSSR-Net: Geo-Spatial Structural Refinement Network for Remote Sensing Change Detection. *IEEE Transactions on Geoscience and Remote Sensing*.
223. Song, Z., Wu, Y., & Huang, S. (2025). Mamba-MSCCA-Net: Efficient change detection for remote sensing images. *Displays*, 103097.
224. Guo, Y., Xu, Y., Tang, G., Yu, Z., & Tang, Q. (2025). AM-CD: Joint Attention and Mamba for Remote Sensing Image Change Detection. *Neurocomputing*, 130607.
225. Wang, H., Ye, Z., Xu, C., Mei, L., Lei, C., & Wang, D. (2024). TTMGNet: Tree Topology Mamba-Guided Network Collaborative Hierarchical Incremental Aggregation for Change Detection. *Remote Sensing*, 16(21), 4068.
226. Song, J., Yang, S., Li, Y., & Li, X. (2024). An Unsupervised Remote Sensing Image Change Detection Method Based on RVMamba and Posterior Probability Space Change Vector. *Remote Sensing*, 16(24), 4656.
227. Ma, J., Li, B., Li, H., Meng, S., Lu, R., & Mei, S. (2025). Remote Sensing Change Detection by Pyramid Sequential Processing With Mamba. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
228. Liu, F., Wen, Y., Sun, J., Zhu, P., Mao, L., Niu, G., & Li, J. (2024). Iterative Mamba Diffusion Change-Detection Model for Remote Sensing. *Remote Sensing*, 16(19), 3651.
229. Sun, M., & Guo, F. (2025). DC-Mamba: Bi-temporal deformable alignment and scale-sparse enhancement for remote sensing change detection. *arXiv preprint arXiv:2509.15563*.
230. Huang, Z., Duan, P., Yuan, G., & Li, J. (2025). MSA: Mamba Semantic Alignment Networks for Remote Sensing Change Detection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
231. Li, Y., Liu, W., Li, E., Zhang, L., & Li, X. (2025). Sam-mamba: a two-stage change detection network combining the adapting segment anything and mamba models. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
232. Qin, Y., Wang, C., Fan, Y., & Pan, C. (2025). SAM2-CD: Remote Sensing Image Change Detection With SAM2. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
233. Zhang, J., Chen, R., Liu, F., Liu, H., Zheng, B., & Hu, C. (2024). DC-Mamba: A novel network for enhanced remote sensing change detection in difficult cases. *Remote Sensing*, 16(22), 4186.
234. Chen, D., Liang, X., Wang, L., Guo, Q., & Zhang, J. (2025). Global Difference-Aware Mamba for Hyperspectral Image Change Detection. *IEEE Transactions on Geoscience and Remote Sensing*.
235. Ding, C., Hao, X., Zheng, S., Dong, Y., Hua, W., Wei, W., ... & Zhang, Y. (2025). A Wavelet-Augmented Dual-Branch Position-Embedding Mamba Network for Hyperspectral Image Change Detection. *IEEE Transactions on Geoscience and Remote Sensing*.
236. Zhan, T., Qi, J., Zhang, J., Yu, X., Du, Q., & Wu, Z. (2025). Spatial-Spectral Feature-Enhanced Mamba and SAM-Guided Hyperspectral Multi-class Change Detection. *IEEE Transactions on Geoscience and Remote Sensing*.
237. Fu, Y., Wu, Z., Zheng, Z., Zhu, Q., Gu, Y., & Kwan, M. P. (2025). Mamba-LCD: Robust Urban Change Detection in Low-Light Remote Sensing Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
238. Chen, K., Chen, B., Liu, C., Li, W., Zou, Z., & Shi, Z. (2024). Rsmamba: Remote sensing image classification with state space model. *IEEE Geoscience and Remote Sensing Letters*, 21, 1-5.
239. Yang, M., & Chen, L. (2025). HC-Mamba: Remote Sensing Image Classification via Hybrid Cross-activation State Space Model. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
240. Yan, L., Zhang, X., Wang, K., & Zhang, D. (2024). Contour-enhanced visual state-space model for remote sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*.

241. Li, D., Liu, R., & Liu, Y. (2025). MPFASS-Net: A Mamba Progressive Feature Aggregation Network With Self-Supervised for Remote Sensing Image Scene Classification. *IEEE Transactions on Geoscience and Remote Sensing*.
242. Roy, S., Sar, A., Kaushish, A., Choudhury, T., Um, J. S., Israr, M., ... & Abhraham, A. (2025). HSS-KAMNet: A Hybrid Spectral-Spatial Kolmogorov-Arnold Mamba Network for Residential Land Cover Identification on RS Imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
243. Kuang, Z., Bi, H., Li, F., & Xu, C. (2025). ECP-Mamba: An Efficient Multi-scale Self-supervised Contrastive Learning Method with State Space Model for PolSAR Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*.
244. Du, R., Tang, X., Ma, J., Zhang, X., & Jiao, L. (2025). MLMamba: A Mamba-based Efficient Network for Multi-label Remote Sensing Scene Classification. *IEEE Transactions on Circuits and Systems for Video Technology*.
245. Jiang, K., Yang, M., Xiao, Y., Wu, J., Wang, G., Feng, X., & Jiang, J. (2025). Rep-Mamba: Re-Parameterization in Vision Mamba for Lightweight Remote Sensing Image Super-Resolution. *IEEE Transactions on Geoscience and Remote Sensing*.
246. Liu, Y., Zhang, R., Fu, W., Chen, J., & Dai, A. (2025). CM2-Net: A Hybrid CNN-Mamba2 Net for 3D Electromagnetic Tomography image reconstruction. *IEEE Sensors Journal*.
247. Teng, Y., Wu, Y., Shi, H., Ning, X., Dai, G., Wang, Y., ... & Liu, X. (2024). Dim: Diffusion mamba for efficient high-resolution image synthesis. *arXiv preprint arXiv:2405.14224*.
248. Zhou, H., Wu, X., Chen, H., Chen, X., & He, X. (2024). Rsdehamba: Lightweight vision mamba for remote sensing satellite image dehazing. *arXiv preprint arXiv:2405.10030*.
249. Chi, K., Guo, S., Chu, J., Li, Q., & Wang, Q. (2025). Rsmamba: biologically plausible retinex-based mamba for remote sensing shadow removal. *IEEE Transactions on Geoscience and Remote Sensing*.
250. Dong, J., Yin, H., Li, H., Li, W., Zhang, Y., Khan, S., & Khan, F. S. (2024). Dual hyperspectral mamba for efficient spectral compressive imaging. *arXiv preprint arXiv:2406.00449*.
251. Zhang, C., Wang, F., Zhang, X., Wang, M., Wu, X., & Dang, S. (2024). Mamba-CR: A state-space model for remote sensing image cloud removal. *IEEE Transactions on Geoscience and Remote Sensing*.
252. Liu, J., Pan, B., & Shi, Z. (2025). CR-Famba: A frequency-domain assisted mamba for thin cloud removal in optical remote sensing imagery. *IEEE Transactions on Multimedia*.
253. Wu, T., Zhao, R., Lv, M., Jia, Z., Li, L., Liu, M., ... & Vivone, G. (2025). Efficient Mamba-Attention Network for Remote Sensing Image Super-Resolution. *IEEE Transactions on Geoscience and Remote Sensing*.
254. Liu, W., Luo, B., Liu, J., Nie, H., & Su, X. (2025). FEMNet: A Feature-Enriched Mamba Network for Cloud Detection in Remote Sensing Imagery. *Remote Sensing*, 17(15), 2639.
255. Huang, Y., Miyazaki, T., Liu, X., & Omachi, S. (2025). Irsrmamba: Infrared image super-resolution via mamba-based wavelet transform feature modulation model. *IEEE Transactions on Geoscience and Remote Sensing*.
256. Weng, M., Liu, J., Yang, J., Wu, Z., & Xiao, L. (2025). Range-Null Space Decomposition with Frequency-Oriented Mamba for Spectral Super-Resolution. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
257. Meng, S., Gong, W., Li, S., Song, G., Yang, J., & Ding, Y. (2025). CDWMamba: Cloud Detection with Wavelet-Enhanced Mamba for Optical Satellite Imagery. *Remote Sensing*, 17(11), 1874.
258. Li, M., Xiong, C., Gao, Z., & Ma, J. (2025). HAM: Hierarchical Attention Mamba with Spatial-Frequency Fusion for Remote Sensing Image Super-Resolution. *IEEE Transactions on Geoscience and Remote Sensing*.
259. Wang, Y., Li, Y., Yang, X., Jiang, R., & Zhang, L. (2025). HDAMNet: Hierarchical Dilated Adaptive Mamba Network for Accurate Cloud Detection in Satellite Imagery. *Remote Sensing*, 17(17), 2992.
260. Zhi, R., Fan, X., & Shi, J. (2024). MambaFormerSR: A lightweight model for remote-sensing image super-resolution. *IEEE Geoscience and Remote Sensing Letters*.
261. Xue, T., Zhao, J., Li, J., Chen, C., & Zhan, K. (2025). CD-Mamba: cloud detection with long-range spatial dependency modeling. *Journal of Applied Remote Sensing*, 19(3), 038507-038507.
262. Xu, Y., Wang, H., Zhou, F., Luo, C., Sun, X., Rahardja, S., & Ren, P. (2025). MambaHSISR: Mamba hyperspectral image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing*.

263. Zhu, Q., Zhang, G., Zou, X., Wang, X., Huang, J., & Li, X. (2024). ConvMambaSR: Leveraging state-space models and CNNs in a dual-branch architecture for remote sensing imagery super-resolution. *Remote Sensing*, 16(17), 3254.
264. Chu, J., Chi, K., & Wang, Q. (2025). RMMamba: Randomized Mamba for Remote Sensing Shadow Removal. *IEEE Transactions on Geoscience and Remote Sensing*.
265. Sui, T., Xiang, G., Chen, F., Li, Y., Tao, X., Zhou, J., ... & Qiu, Z. (2025). U-Shaped Dual Attention Vision Mamba Network for Satellite Remote Sensing Single-Image Dehazing. *Remote Sensing*, 17(6), 1055.
266. Zhao, Z., Gao, Q., Yan, J., Li, C., & Tang, J. (2025). HSFMamba: Hierarchical selective fusion Mamba network for optics-guided joint super-resolution and denoising of noise-corrupted SAR images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
267. Duan, P., Luo, Y., Kang, X., & Li, S. (2025). LaMamba: Linear Attention Mamba for Hyperspectral Image Denoising. *IEEE Transactions on Geoscience and Remote Sensing*.
268. Xie, Z., Miao, G., & Chang, H. (2025). MTSR: Mamba-Transformer Super-Resolution Model for Hyperspectral Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
269. Xin, X., Deng, Y., Huang, W., Wu, Y., Fang, J., & Wang, J. (2025). Multi-Pattern Scanning Mamba for Cloud Removal. *Remote Sensing*, 17(21), 3593.
270. Li, C., Pan, Z., & Hong, D. (2025). Dynamic State-Control Modeling for Generalized Remote Sensing Image Super-Resolution. In *Proceedings of the Computer Vision and Pattern Recognition Conference* (pp. 3076-3084).
271. Si, P., Jia, M., Wang, H., Wang, J., Sun, L., & Fu, Z. (2025). DC-Mamba: A Degradation-Aware Cross-Modality Framework for Blind Super-Resolution of Thermal UAV Images. *IEEE Transactions on Geoscience and Remote Sensing*.
272. Wu, S., He, X., & Chen, X. (2025). Weamba: Weather-Degraded Remote Sensing Image Restoration with Multi-Router State Space Model. *Remote Sensing*, 17(3), 458.
273. Deng, N., Han, J., Ding, H., Liu, D., Zhang, Z., Song, W., & Tong, X. (2025). OSSMDNet: An Omni-Selective Scanning Mechanism for a Remote Sensing Image Denoising Network Based on the State-Space Model. *Remote Sensing*, 17(16), 2759.
274. Zhu, Z., Chen, Y., Zhang, S., Luo, G., & Zeng, J. (2025). Mamba-Based Unet for Hyperspectral Image Denoising. *IEEE Signal Processing Letters*.
275. Chen, C., Li, J., Liu, X., Yuan, Q., & Zhang, L. (2025). Bidirectional-Aware Network Combining Transformer and Mamba for Hyperspectral Image Denoising. *IEEE Transactions on Geoscience and Remote Sensing*.
276. Fu, H., Sun, G., Li, Y., Ren, J., Zhang, A., Jing, C., & Ghamisi, P. (2024). HDMba: hyperspectral remote sensing imagery dehazing with state space model. *arXiv preprint arXiv:2406.05700*.
277. Shao, M., Tan, X., Shang, K., Liu, T., & Cao, X. (2025). A Hybrid Model of State Space Model and Attention for Hyperspectral Image Denoising. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
278. Liu, Y., Xiao, J., Song, X., Guo, Y., Jiang, P., Yang, H., & Wang, F. (2024). HSIDMamba: Exploring bidirectional state-space models for hyperspectral denoising. *arXiv preprint arXiv:2404.09697*.
279. Luan, X., Fan, H., Wang, Q., Yang, N., Liu, S., Li, X., & Tang, Y. (2025). FMambaIR: A hybrid state space model and frequency domain for image restoration. *IEEE Transactions on Geoscience and Remote Sensing*.
280. Qiu, L., Xie, F., Liu, C., Che, X., & Shi, Z. (2025). Radiation-Tolerant Unsupervised Deep Image Stitching for Remote Sensing. *IEEE Transactions on Geoscience and Remote Sensing*.
281. Yang, M., Jiang, S., Jiang, W., & Li, Q. (2025). Mamba-based Feature Extraction and Multi-Frequency Information Fusion for Stereo Matching of High-Resolution Satellite Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
282. Li, B., Zhao, H., Wang, W., Hu, P., Gou, Y., & Peng, X. (2025). Mair: A locality-and continuity-preserving mamba for image restoration. In *Proceedings of the Computer Vision and Pattern Recognition Conference* (pp. 7491-7501).
283. Fu, G., Xiong, F., Lu, J., & Zhou, J. (2024). SSUMamba: Spatial-spectral selective state space model for hyperspectral image denoising. *IEEE Transactions on Geoscience and Remote Sensing*.

284. Patnaik, N., Nayak, N., Agrawal, H. B., Khamaru, M. C., Bal, G., Panda, S. S., ... & Vadlamani, K. (2025). Small Vision-Language Models: A Survey on Compact Architectures and Techniques. *arXiv preprint arXiv:2503.10665*.
285. Li, S., & Tang, H. (2024). Multimodal alignment and fusion: A survey. *arXiv preprint arXiv:2411.17040*.
286. Meng, L., Wang, J., Huang, Y., & Xiao, L. (2025). RSIC-GMamba: A State Space Model with Genetic Operations for Remote Sensing Image Captioning. *IEEE Transactions on Geoscience and Remote Sensing*.
287. Liu, C., Chen, K., Chen, B., Zhang, H., Zou, Z., & Shi, Z. (2024). Rscama: Remote sensing image change captioning with state space model. *IEEE Geoscience and Remote Sensing Letters*, 21, 1-5.
288. Liu, C., Zhang, J., Chen, K., Wang, M., Zou, Z., & Shi, Z. (2025). Remote Sensing Spatiotemporal Vision-Language Models: A comprehensive survey. *IEEE Geoscience and Remote Sensing Magazine*.
289. Chen, K., Liu, C., Chen, B., Li, W., Zou, Z., & Shi, Z. (2025). Dynamicvis: An efficient and general visual foundation model for remote sensing image understanding. *arXiv preprint arXiv:2503.16426*.
290. He, X., Cao, K., Zhang, J., Yan, K., Wang, Y., Li, R., ... & Zhou, M. (2025). Pan-mamba: Effective pan-sharpening with state space model. *Information Fusion*, 115, 102779.
291. Wang, Y., Liang, F., Wang, S., Chen, H., Cao, Q., Fu, H., & Chen, Z. (2025). Towards an Efficient Remote Sensing Image Compression Network with Visual State Space Model. *Remote Sensing*, 17(3).
292. Fei, Z., Fan, M., Yu, C., Li, D., Zhang, Y., & Huang, J. (2024). Dimba: Transformer-mamba diffusion models. *arXiv preprint arXiv:2406.01159*.
293. Peng, X., Zhou, J., & Wu, X. (2025). Distillation-Based Cross-Model Transferable Adversarial Attack for Remote Sensing Image Classification. *Remote Sensing*, 17(10), 1700.
294. Dewis, Z., Xu, Z., Zhu, Y., Alkayid, M., Heffring, M., & Xu, L. L. (2025). Spatial-Temporal-Spectral Mamba with Sparse Deformable Token Sequence for Enhanced MODIS Time Series Classification. *arXiv preprint arXiv:2508.02839*.
295. Li, D., & Bhatti, U. A. (2024). MSTFNet: A Mamba and dual Swin-Transformer fusion network for remote sensing image classification for precision agriculture land processing. *Available at SSRN 5033170*.
296. Zhao, M., Wang, D., Zhang, G., Cao, W., Xu, S., Li, Z., & Liu, X. (2025). Evaluating Maize Emergence Quality with Multi-task YOLO11-Mamba and UAV-RGB Remote Sensing. *Smart Agricultural Technology*, 101351.
297. Li, J., Yang, C., Zhu, C., Qin, T., Tu, J., Wang, B., ... & Qiao, J. (2025). CMRNet: An Automatic Rapeseed Counting and Localization Method Based on the CNN-Mamba Hybrid Model. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
298. Li, H., Zhao, F., Xue, F., Wang, J., Liu, Y., Chen, Y., ... & Kobayashi, H. H. (2025). Succulent-YOLO: Smart UAV-Assisted Succulent Farmland Monitoring with CLIP-Based YOLOv10 and Mamba Computer Vision. *Remote Sensing*, 17(13), 2219.
299. Zhang, X., Gu, J., Azam, B., Zhang, W., Lin, M., Li, C., ... & Akhtar, N. (2025). RSVMamba for Tree Species Classification Using UAV RGB Remote Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing*.
300. Zheng, J., Fu, Y., Chen, X., Zhao, R., Lu, J., Zhao, H., & Chen, Q. (2025). EGCM-UNet: Edge Guided Hybrid CNN-Mamba UNet for farmland remote sensing image semantic segmentation. *Geocarto International*, 40(1), 2440407.
301. Li, Y., Wang, Y., Shao, X., & Zheng, A. (2025). An efficient fire detection algorithm based on Mamba space state linear attention. *Scientific Reports*, 15(1), 11289.
302. Ho, Y. H., & Mostafavi, A. (2025). Multimodal Mamba with multitask learning for building flood damage assessment using synthetic aperture radar remote sensing imagery. *Computer-Aided Civil and Infrastructure Engineering*.
303. Ho, Y. H., & Mostafavi, A. (2025). Flood-DamageSense: Multimodal Mamba with Multitask Learning for Building Flood Damage Assessment using SAR Remote Sensing Imagery. *arXiv preprint arXiv:2506.06667*.
304. Tang, X., Lu, Z., Fan, X., Yan, X., Yuan, X., Li, D., ... & Catani, F. (2025). Mamba for landslide detection: A lightweight model for mapping landslides with very high-resolution images. *IEEE Transactions on Geoscience and Remote Sensing*.
305. Shao, Y., & Xu, L. (2025). Multimodal Natural Disaster Scene Recognition with Integrated Large Model and Mamba. *Applied Sciences*, 15(3), 1149.

306. Andrianarivony, H. S., & Akhloufi, M. A. (2025). LinU-Mamba: Visual Mamba U-Net with Linear Attention to Predict Wildfire Spread. *Remote Sensing*, 17(15), 2715.
307. Li, W., Ma, G., Zhang, H., Chen, P., Wang, D., & Chen, R. (2025). Multi-scenario building change detection in remote sensing images using CNN-Mamba hybrid network and consistency enhancement learning. *Expert Systems with Applications*, 129843.
308. Chen, S., Wang, F., Ren, P., Luo, C., & Fu, Z. (2025). OSDMamba: Enhancing Oil Spill Detection from Remote Sensing Images Using Selective State Space Model. *arXiv preprint arXiv:2506.18006*.
309. Zhang, Y., Wang, S., Chen, Y., Wei, S., Xu, M., & Liu, S. (2025). Algae-Mamba: A Spatially Variable Mamba for Algae Extraction from Remote Sensing Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
310. Zhang, X., Ma, Y., Zhang, F., Li, Z., & Zhang, J. (2025). Multi-Model Synergistic Satellite-Derived Bathymetry Fusion Approach Based on Mamba Coral Reef Habitat Classification. *Remote Sensing*, 17(13), 2134.
311. Sha, P., Lu, S., Xu, Z., Yu, J., Li, L., Zou, Y., & Zhao, L. (2025). OWTNet: A Novel CNN-Mamba Fusion Network for Offshore Wind Turbine Detection in High-Resolution Remote Sensing Images. *Journal of Marine Science and Engineering*, 13(11), 2124.
312. Jiang, X., Wang, S., Li, W., Yang, H., Guan, J., Zhang, Y., & Zhou, S. (2025). STDMamba: Spatio-Temporal Decomposition Mamba for Long-Term Fine-Grained SST Prediction. *IEEE Transactions on Geoscience and Remote Sensing*.
313. Shi, X., Ni, W., Duan, B., Su, Q., Liu, L., & Ren, K. (2025). MMamba: An Efficient Multimodal Framework for Real-Time Ocean Surface Wind Speed Inpainting Using Mutual Information and Attention-Mamba-2. *Remote Sensing*, 17(17), 3091.
314. Sun, Y., Song, J., Cai, Z., & Xiao, L. (2025). Tracking Mamba for Road Extraction From Satellite Imagery. *IEEE Geoscience and Remote Sensing Letters*.
315. Wang, Z., Yuan, S., Li, R., Xu, N., You, Z., & Huang, D. (2025). FDMamba: Frequency-Driven Dual-Branch Mamba Network for Road Extraction From Remote Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing*.
316. Li, B., Shen, C., Gu, S., Zhao, Y., & Xiao, F. (2025). Explicitly Integrated Multi-Task Learning in a Hybrid Network for Remote Sensing Road Extraction. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
317. Zhao, S., Wang, F., Huang, X., Yang, X., Jiang, N., Peng, J., & Ban, Y. (2025). Mamba-UNet: Dual-Branch Mamba Fusion U-Net With Multiscale Spatio-Temporal Attention for Precipitation Nowcasting. *IEEE Transactions on Industrial Informatics*.
318. Zhang, J., Chen, M., Zhao, Y., Shan, L., Li, C., Hu, H., ... & Xu, B. (2025). Asymmetric Mamba-CNN Collaborative Architecture for Large-Size Remote Sensing Image Semantic Segmentation. *IEEE Transactions on Geoscience and Remote Sensing*.
319. Liu, Z., Chen, H., Bai, L., Li, W., Ouyang, W., Zou, Z., & Shi, Z. (2024). Mambads: Near-surface meteorological field downscaling with topography constrained selective state space modeling. *IEEE Transactions on Geoscience and Remote Sensing*.
320. Ma, X., Lv, Z., Ma, C., Zhang, T., Xin, Y., & Zhan, K. (2025). BS-Mamba for black-soil area detection on the Qinghai-Tibetan plateau. *Journal of Applied Remote Sensing*, 19(2), 028502-028502.
321. Liu, Y., Shi, H., Cao, K., Wu, S., Ye, H., Wang, X., ... & Xiong, W. (2025). kMetha-Mamba: K-means clustering mamba for methane plumes segmentation. *International Journal of Applied Earth Observation and Geoinformation*, 142, 104664.
322. Yu, W., & Wang, X. (2025). Mambaout: Do we really need mamba for vision?. In *Proceedings of the Computer Vision and Pattern Recognition Conference* (pp. 4484-4496).
323. Xiao, C., Li, M., Zhang, Z., Meng, D., & Zhang, L. (2024). Spatial-mamba: Effective visual state space models via structure-aware state fusion. *arXiv preprint arXiv:2410.15091*.
324. Hamdan, E., Pan, H., & Cetin, A. E. (2024). Sparse Mamba: Introducing Controllability, Observability, And Stability To Structural State Space Models. *arXiv preprint arXiv:2409.00563*.

325. Shi, Y., Li, M., Dong, M., & Xu, C. (2025). Vssd: Vision mamba with non-causal state space duality. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 10819-10829).
326. Ding, H., Xia, B., Liu, W., Zhang, Z., Zhang, J., Wang, X., & Xu, S. (2024). A novel mamba architecture with a semantic transformer for efficient real-time remote sensing semantic segmentation. *Remote Sensing*, 16(14), 2620.
327. Díaz, A. H., Davidson, R., Eckersley, S., Bridges, C. P., & Hadfield, S. J. E-mamba: Using state-space-models for direct event processing in space situational awareness. In *Proceedings of SPAICE2024: The First Joint European Space Agency/IAA Conference on AI in and for Space* (pp. 509-514).
328. Sedeh, M. A., & Sharifian, S. (2025). EdgePVM: A serverless satellite edge computing constellation for changes detection using onboard Parallel siamese Vision MAMBA. *Future Generation Computer Systems*, 107985.
329. Jiang, F., Pan, C., Dong, L., Wang, K., Debbah, M., Niyato, D., & Han, Z. (2025). A comprehensive survey of large ai models for future communications: Foundations, applications and challenges. *arXiv preprint arXiv:2505.03556*.
330. Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378, 686-707.
331. Hu, Z., Daryakenari, N. A., Shen, Q., Kawaguchi, K., & Karniadakis, G. E. (2024). State-space models are accurate and efficient neural operators for dynamical systems. *arXiv preprint arXiv:2409.03231*.
332. Cheng, C. W., Huang, J., Zhang, Y., Yang, G., Schönlieb, C. B., & Aviles-Rivero, A. I. (2024). Mamba neural operator: Who wins? transformers vs. state-space models for pdes. *arXiv preprint arXiv:2410.02113*.
333. Hu, Z., Daryakenari, N. A., Shen, Q., Kawaguchi, K., & Karniadakis, G. E. (2024). State-space models are accurate and efficient neural operators for dynamical systems. *arXiv preprint arXiv:2409.03231*.
334. Liu, C., Zhao, B., Ding, J., Wang, H., & Li, Y. (2025). Mamba Integrated with Physics Principles Masters Long-term Chaotic System Forecasting. *arXiv preprint arXiv:2505.23863*.
335. Li, S., Singh, H., & Grover, A. (2024, September). Mamba-nd: Selective state space modeling for multi-dimensional data. In *European Conference on Computer Vision* (pp. 75-92). Cham: Springer Nature Switzerland.
336. Qin, H., Chen, Y., Jiang, Q., Sun, P., Ye, X., & Lin, C. (2024). Metmamba: Regional weather forecasting with spatial-temporal mamba model. *arXiv preprint arXiv:2408.06400*.
337. Eddin, M. H. S., Zhang, Y., Kollet, S., & Gall, J. (2025, May). RiverMamba: A State Space Model for Global River Discharge and Flood Forecasting. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*.
338. Rasp, S., Pritchard, M. S., & Gentine, P. (2018). Deep learning to represent subgrid processes in climate models. *Proceedings of the national academy of sciences*, 115(39), 9684-9689.
339. Yuval, J., & O'Gorman, P. A. (2020). Stable machine-learning parameterization of subgrid processes for climate modeling at a range of resolutions. *Nature communications*, 11(1), 3295.
340. Kochkov, D., Yuval, J., Langmore, I., Norgaard, P., Smith, J., Mooers, G., ... & Hoyer, S. (2024). Neural general circulation models for weather and climate. *Nature*, 632(8027), 1060-1066.
341. Bock, F. E., Keller, S., Huber, N., & Klusemann, B. (2021). Hybrid modelling by machine learning corrections of analytical model predictions towards high-fidelity simulation solutions. *Materials*, 14(8), 1883.
342. Beucler, T., Koch, E., Kotlarski, S., Leutwyler, D., Michel, A., & Koh, J. (2023). Next-generation earth system models: Towards reliable hybrid models for weather and climate applications. *arXiv preprint arXiv:2311.13691*.
343. Huo, C., Chen, K., Zhang, S., Wang, Z., Yan, H., Shen, J., ... & Wang, Z. (2025). When Remote Sensing Meets Foundation Model: A Survey and Beyond. *remote sensing*, 17(2).
344. Xiao, A., Xuan, W., Wang, J., Huang, J., Tao, D., Lu, S., & Yokoya, N. (2025). Foundation models for remote sensing and earth observation: A survey. *IEEE Geoscience and Remote Sensing Magazine*.
345. Cong, Y., Khanna, S., Meng, C., Liu, P., Rozi, E., He, Y., ... & Ermon, S. (2022). Satmae: Pre-training transformers for temporal and multi-spectral satellite imagery. *Advances in Neural Information Processing Systems*, 35, 197-211.

346. Reed, C. J., Gupta, R., Li, S., Brockman, S., Funk, C., Clipp, B., ... & Darrell, T. (2023). Scale-mae: A scale-aware masked autoencoder for multiscale geospatial representation learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 4088-4099).
347. Mendieta, M., Han, B., Shi, X., Zhu, Y., Chen, C., & Li, M. (2023). GFM: Building geospatial foundation models via continual pretraining. *arXiv preprint arXiv:2302.04476*, 3.
348. Duc, C. M., & Fukui, H. (2025). SatMamba: Development of Foundation Models for Remote Sensing Imagery Using State Space Models. *arXiv preprint arXiv:2502.00435*.
349. Wang, F., Wang, Y., Chen, M., Zhao, H., Sun, Y., Wang, S., ... & Zhang, J. (2025). Roma: Scaling up mamba-based foundation models for remote sensing. *arXiv preprint arXiv:2503.10392*.
350. Wang, P., Chang, H., Hu, H., Li, X., Liu, X., Liu, Y., ... & Sun, X. (2025). RingMamba: Remote Sensing Multi-sensor Pre-training with Visual State Space Model. *IEEE Transactions on Geoscience and Remote Sensing*.
351. Shi, Z., Zhao, C., Wang, K., Kong, X., & Zhu, J. (2025). Geo-Mamba: A data-driven Mamba framework for spatiotemporal modeling with multi-source geographic factor integration. *International Journal of Applied Earth Observation and Geoinformation*, 144, 104854.
352. Gong, S., Zhuge, Y., Zhang, L., Wang, Y., Zhang, P., Wang, L., & Lu, H. (2025). Avs-mamba: Exploring temporal and multi-modal mamba for audio-visual segmentation. *IEEE Transactions on Multimedia*.
353. Zhou, Y., et al. (2025). Advances on multimodal remote sensing foundation models for Earth observation downstream tasks: A survey. *Remote Sensing*, 17(3532), 1–35.
354. Dao, T., Fu, D., Ermon, S., Rudra, A., & Ré, C. (2022). Flashattention: Fast and memory-efficient exact attention with io-awareness. *Advances in neural information processing systems*, 35, 16344-16359.
355. Choquette, J. (2023). Nvidia hopper h100 gpu: Scaling performance. *IEEE Micro*, 43(3), 9-17.
356. Dao, T. (2023). Flashattention-2: Faster attention with better parallelism and work partitioning. *arXiv preprint arXiv:2307.08691*.
357. Micikevicius, P., Narang, S., Alben, J., Diamos, G., Elsen, E., Garcia, D., ... & Wu, H. (2017). Mixed precision training. *arXiv preprint arXiv:1710.03740*.
358. Strubell, E., Ganesh, A., & McCallum, A. (2019, July). Energy and policy considerations for deep learning in NLP. In *Proceedings of the 57th annual meeting of the association for computational linguistics* (pp. 3645-3650).
359. Lacoste, A., Luccioni, A., Schmidt, V., & Dandres, T. (2019). Quantifying the carbon emissions of machine learning. *arXiv preprint arXiv:1910.09700*.
360. Lannelongue, L., Grealey, J., & Inouye, M. (2021). Green algorithms: quantifying the carbon footprint of computation. *Advanced science*, 8(12), 2100707.
361. Bouza, L., Bugeau, A., & Lannelongue, L. (2023). How to estimate carbon footprint when training deep learning models? A guide and review. *Environmental Research Communications*, 5(11), 115014.
362. Li, G., Chen, B., Zhao, C., Zhang, L., & Zhang, J. (2025). Osmamba: Omnidirectional spectral mamba with dual-domain prior generator for exposure correction. In *Proceedings of the Computer Vision and Pattern Recognition Conference* (pp. 7480-7490).

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.