

# RadioMamba: Breaking the Accuracy-Efficiency Trade-Off in Radio Map Construction via a Hybrid Mamba-UNet

Honggang Jia, *Student Member, IEEE*, Nan Cheng<sup>✉</sup>, *Senior Member, IEEE*, Xiucheng Wang<sup>✉</sup>, *Member, IEEE*, Conghao Zhou<sup>✉</sup>, *Member, IEEE*, Ruijin Sun<sup>✉</sup>, *Member, IEEE*, and Xuemin Shen<sup>✉</sup>, *Fellow, IEEE*

**Abstract**—Radio map (RM) has recently attracted much attention since it can provide real-time and accurate spatial channel information for 6G services and applications. However, current deep learning-based methods for RM construction exhibit well known accuracy-efficiency trade-off. In this paper, we introduce RadioMamba, a hybrid Mamba-UNet architecture for RM construction to address the trade-off. Generally, accurate RM construction requires modeling long-range spatial dependencies, reflecting the global nature of wave propagation physics. RadioMamba utilizes a Mamba-Convolutional block where the Mamba branch captures these global dependencies with linear complexity, while a parallel convolutional branch extracts local features. This hybrid design generates feature representations that capture both global context and local detail. Experiments show that RadioMamba achieves higher accuracy than existing methods, including diffusion models, while operating nearly 20 times faster and using only 2.9% of the model parameters. By improving both accuracy and efficiency, RadioMamba presents a viable approach for real-time intelligent optimization in next generation wireless systems.

**Index Terms**—6G wireless networks, radio map, Mamba, lightweight model, real-time optimization.

## I. INTRODUCTION

THE continuous advancement towards sixth-generation (6G) wireless networks is enabling a future with the Internet of Things (IoT), autonomous systems, and immersive cyber-physical experiences [1], [2], [3], [4]. This vision, which underpins future applications like the metaverse [5], motivates the transition from reactive network management to proactive, data-driven control [6]. A key component of this evolution is the network digital twin (NDT) [7], [8], a

high-fidelity virtual replica of the physical network environment that enables simulation, prediction, and optimization in real-time [6], [9], [10], [11]. The NDT concept is becoming a cornerstone for managing the unprecedented complexity of 6G systems, offering a paradigm to operate, manage, and optimize wireless networks with enhanced intelligence and autonomy [8].

For wireless systems, especially those involving mobile elements like AAVs [12], a high-fidelity twin must extend beyond the physical infrastructure to also model the complex radio propagation environment, a task that comes with its own set of security and reliability challenges [13]. This functionality is provided by a radio map (RM)—a detailed spatial representation of signal propagation characteristics, such as pathloss or received signal strength [14]. The precision of such maps can be further enhanced by technologies like integrated sensing and communication (ISAC) [15], [16], [17], [18]. Consequently, an accurate and continuously updated RM is essential for the NDT and a wide range of other critical 6G applications. These include dynamic spectrum management, interference coordination, user-cell association, and on-the-fly trajectory planning for autonomous aerial vehicles (AAVs) in complex air-ground networks [19], [20], [21], [22], [23]. Furthermore, high-definition maps, a specialized form of RMs, are critical for ensuring the safety and efficiency of vehicular networks by providing centimeter-level accuracy for autonomous vehicle navigation and coordination [24], [25].

However, traditional RM construction methods have several limitations. These traditional approaches are broadly categorized into two types: measurement-based and physical model-based. Measurement-based techniques rely on deploying extensive sensors or conducting drive-tests to gather signal measurements, which are then used with interpolation algorithms like Kriging to construct the RM [26], [27], [28]. However, these methods are often expensive, labor-intensive, and face challenges in scaling in complex urban environments. The quality of the resulting RM is highly dependent on the quantity and accuracy of the measurements, and these methods are challenged in constructing maps for inaccessible regions, limiting their use in applications like AAV path planning [29], [30], [31]. Alternatively, physical model-based methods using electromagnetic ray-tracing (ERT) [32], [33] can yield highly precise RMs by

Received 24 July 2025; revised 19 August 2025; accepted 26 September 2025. Date of publication 6 October 2025; date of current version 11 December 2025. This work was supported in part by the National Key Research and Development Program of China under Grant 2024YFB2907500, and in part by the National Natural Science Foundation of China under Grant 62571402. Recommended for acceptance by Dr. Celimuge Wu. (*Corresponding author: Nan Cheng.*)

Honggang Jia, Nan Cheng, Xiucheng Wang, Conghao Zhou, and Ruijin Sun are with the State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an 710071, China, and also with the School of Telecommunications Engineering, Xidian University, Xi'an 710071, China (e-mail: jiahg@stu.xidian.edu.cn; dr.nan.cheng@ieee.org; xcwang\_1@stu.xidian.edu.cn; conghao.zhou@ieee.org; sunruijin@xidian.edu.cn).

Xuemin Shen is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: sshen@uwaterloo.ca).

Digital Object Identifier 10.1109/TNSE.2025.3617102

simulating wave propagation in a 3D environmental model. Yet, they are computationally intensive. This high latency makes both traditional approaches less suitable for dynamic 6G scenarios where the environment, affected by elements like moving vehicles, changes in seconds [34], [35], [36].

To address these challenges, the community has turned to deep learning (DL), which can learn the complex, non-linear mapping from environmental features to a complete RM with rapid inference capabilities. Initial works like RadioUNet [37] and RME-GAN [38] established the viability of this approach. Most recently, diffusion models, exemplified by RadioDiff [39], have become the state-of-the-art (SOTA) in terms of raw accuracy, leveraging an iterative denoising process to generate high-fidelity maps. While their inference process is faster than traditional ray-tracing simulations, the iterative nature of diffusion models still incurs latencies that are too high for the most demanding 6G applications [40]. Scenarios such as real-time network digital twins require near-instantaneous environmental updates, a requirement that current SOTA models cannot meet. This situation creates a trade-off: a choice between the high accuracy of heavyweight generative models and the low latency of their less accurate, lightweight predecessors. Overcoming this trade-off is critical for the widespread deployment of DL-based RM solutions in future real-time systems.

This accuracy-efficiency trade-off is not merely an issue of model scaling but is deeply rooted in the foundational architectures of current deep learning models, which are often adapted from the domain of computer vision. In radio map construction, the input is typically an image-like representation of the environment [37] where long-range physical phenomena, such as reflections and diffractions, manifest as subtle, spatially distant correlations. While convolutional neural networks (CNNs) are efficient, their hierarchical nature, built upon stacked local operations, may be suboptimal for this task. The process of downsampling and repeated local feature extraction in a deep CNN can progressively weaken or even lose these critical, non-local physical cues. Consequently, architectures with an innate global receptive field, like vision Transformers (ViTs) [41], appear to be a more natural fit. They are theoretically well-equipped to capture these dependencies from the outset. However, their global self-attention mechanism comes at the cost of quadratic complexity with respect to the input size, making them too computationally intensive for the real-time, high-resolution RM construction required by 6G systems. Moreover, ViTs often require extensive pre-training on massive datasets to develop effective visual representations [42]. This creates a critical dilemma: the architectural need for global context to respect the underlying physics is fundamentally at odds with the practical need for low-latency inference and precise local detail.

Structured state space models (SSMs) [43], and particularly the Mamba architecture [44], have recently emerged as a promising approach to this problem. Mamba is specifically designed to model long-range dependencies with linear-time complexity, emulating the global context awareness of Transformers without the associated computational burden. The power of this approach has led to its swift adaptation in computer vision, with various

works developing hybrid Mamba-CNN backbones for image-based tasks [45], [46]. Furthermore, the ongoing development of even more efficient variants optimized for training speed and edge deployment underscores its practicality for real-world systems [47], [48]. This combination of capabilities, achieving the global reach of Transformers with the efficiency of CNNs, suggests Mamba as a highly suitable architectural foundation for building models that are both highly accurate and fast enough for real-time RM construction.

Based on these capabilities, we propose **RadioMamba**, a hybrid Mamba-UNet architecture designed to resolve the accuracy-efficiency trade-off. The ability of Mamba to model global context with linear complexity offers a potential solution to the limitations of previous models. We hypothesize that by augmenting a computationally efficient, CNN-based architecture with the ability to model the long-range spatial dependencies that govern the underlying physics of radio propagation, it is possible to break the established trade-off, creating a model that can achieve state-of-the-art accuracy while maintaining the high computational efficiency required for real-time systems. RadioMamba is designed to test this hypothesis by embedding a hybrid Mamba-Convolutional block within the U-Net framework, where the Mamba branch models global interactions and a parallel convolutional branch extracts local features. This high-speed, high-fidelity approach serves as a key enabler for dynamic 6G applications, such as real-time digital twin synchronization and autonomous AAV navigation. The result is a single-pass model that not only improves upon prior methods in performance and speed but also unlocks new operational capabilities. Our primary contributions are as follows.

- 1) We analyze the accuracy-efficiency trade-off in radio map construction, tracing it to the architectural limitations of current models. We identify that CNNs' local receptive fields fail to capture global wave physics, while global-context models like Transformers are computationally prohibitive for real-time use. Our analysis further establishes the unique suitability of the Mamba architecture, which models long-range dependencies with linear complexity, as a foundation to resolve this trade-off.
- 2) We propose RadioMamba, a novel hybrid architecture that integrates the strengths of Mamba and CNNs within a U-Net framework. The core of our design is a synergistic Mamba-Convolutional block that captures both global long-range dependencies and fine-grained local features, making it uniquely suited for the physics-informed task of radio map construction.
- 3) Our experiments show that RadioMamba establishes a new SOTA, outperforming existing methods like RadioDiff across all accuracy and perceptual metrics. It achieves this while being nearly 20 times faster and using only 2.9% of the model parameters, demonstrating a resolution to the accuracy-efficiency trade-off and confirming its viability for real-time 6G systems.

The remainder of this paper is organized as follows. Section II reviews related works. Section III presents the system model and problem formulation. Section IV focus on the proposed

RadioMamba architecture. Section V details the comprehensive experimental setup, results, and analyses. Finally, Section VI concludes the paper and discusses future research directions.

## II. RELATED WORK

The methodologies for constructing RMs are broadly divided into two categories: measurement-based, which relies on physical data collection and interpolation [26], [27], and sampling-free, which generates maps directly from environmental data. Our work focuses on the latter, a field that has been revolutionized by DL due to its power in solving complex generative tasks [37], [38], [39]. To provide a thorough background, this section reviews the pertinent literature across several key domains. We first survey the evolution of radio propagation modeling paradigms, then discuss the specific DL architectures for RM construction. Subsequently, we explore the broader context of AI-enabled wireless network optimization and the challenges of model efficiency, before finally delving into the state space models that form the technical foundation of our proposed method.

### A. Radio Propagation Modeling Paradigms

The prediction of radio wave propagation is a long-standing challenge in wireless communications. The approaches can be broadly categorized into model-based (or physics-based) and data-driven methods [49].

*Model-Based Approaches:* These methods rely on fundamental physical principles to predict signal strength. They can be further divided into empirical and deterministic models. Empirical models, such as the Okumura-Hata and Lee models, are derived from extensive measurement campaigns and provide statistical characterizations of path loss as a function of distance, frequency, and antenna heights [50]. While computationally simple and widely used for large-scale network planning, their accuracy is limited as they generalize environmental characteristics into broad categories and may not capture site-specific details effectively [51].

Deterministic models, most notably those based on ray tracing, offer a more precise alternative by simulating the paths of electromagnetic waves as they reflect, diffract, and scatter within a detailed 3D model of the environment [52], [53]. Ray-tracing techniques can provide highly accurate, site-specific predictions for both narrowband and wideband channel characteristics, making them a powerful tool for designing and optimizing complex radio networks, from indoor WLANs to large-scale urban deployments [54], [55]. However, the accuracy of ray tracing is critically dependent on the quality and detail of the environmental database, and the computational cost can be prohibitive, especially for large areas or when considering a high number of interactions [53].

*Data-Driven Approaches:* With the proliferation of sensor data and advancements in machine learning, data-driven modeling has emerged as a powerful paradigm [56], [57]. Instead of relying on first principles, these methods learn the complex, nonlinear mapping directly from input-output data. In the context of wireless communications, this involves

learning the relationship between environmental features and channel characteristics from measured or simulated channel data [58]. The key advantage of data-driven methods is their ability to model highly complex systems where analytical or deterministic models are infeasible. Our work falls into this category, leveraging the power of deep learning to learn the mapping from environmental layouts to complete radio maps.

### B. Deep Learning for Radio Map Construction

The evolution of DL for RM construction has seen a progression of architectures, each with distinct advantages and drawbacks [53], [59], [60], [61].

*CNN Approaches:* CNNs, particularly encoder-decoder architectures, were among the first architectures applied to RM construction. **RadioUNet** [37] was one of the first models to apply the U-Net architecture [62] for RM construction. By representing the building layout and transmitter location as input channels, RadioUNet learns a direct mapping to the final pathloss map. The skip connections in U-Net are particularly effective, allowing the decoder to combine deep, semantic features with shallow, high-resolution features, which is essential for preserving sharp details. The primary strength of RadioUNet is its computational efficiency. However, its core convolution operator has an inherently local receptive field, limiting its ability to capture the global context of the radio environment and capping its ultimate performance.

*GAN Approaches:* To generate more perceptually realistic RMs, researchers turned to GANs [63], [64]. **RME-GAN** [38] is a prime example, utilizing a conditional GAN (cGAN) framework [65] where a generator network creates RMs and a discriminator distinguishes them from ground-truth maps. Other works have also explored GANs for fast and accurate cooperative radio map estimation [66]. This adversarial process often results in maps with sharper details. GANs are also computationally efficient during inference, but their training can be unstable and requires careful tuning.

*Advanced Generative Models:* Recently, denoising diffusion probabilistic models (DDPMs) [67], [68] have established SOTA performance. **RadioDiff** [39] is the first work to adapt this paradigm to RM construction. It formulates the task as a conditional generation process, starting from pure random noise and iteratively denoises it over hundreds of timesteps to produce a high-fidelity RM. This iterative refinement allows diffusion models to achieve high accuracy. However, this improved performance comes at the cost of high computational complexity and slow inference, making them orders of magnitude slower than single-pass models and posing challenges for real-time deployment. The broader field of wireless communications has also seen a surge in the application of generative AI (GAI), for instance, to construct channel knowledge maps (CKMs) [69] or to enhance physical layer security, demonstrating the versatility of these models beyond pure generation tasks [70].

### C. AI-Enabled Wireless Network Optimization

The construction of an accurate RM is often not an end in itself, but a means to enable more intelligent network management and optimization. The application of AI, particularly reinforcement learning (RL) and its variants, to wireless communications is a burgeoning field [71].

In 6G networks, which are envisioned as a ‘network of subnetworks,’ dynamic resource management becomes exceptionally challenging due to high mobility and mutual interference. Multi-agent reinforcement learning (MARL) has been proposed to address this, enabling subnetworks to learn cooperative policies for channel selection and power control. Advanced MARL architectures, such as those incorporating graph attention networks, can effectively reason about inter-subnetwork relationships using only accessible information like RSSI, avoiding the need for hard-to-obtain global channel state information [72]. Similarly, in D2D-enabled 6G networks, federated reinforcement learning provides a framework for decentralized resource allocation that maximizes network capacity while respecting user privacy [73].

Accurate channel estimation is another critical task where AI is making inroads. In Massive MIMO systems, for instance, data-driven approaches using neural networks are being explored to estimate effective downlink channel gains without the need for downlink pilots, outperforming traditional model-based methods, especially in environments with low channel hardening [49]. This trend of AI-enabled, data-driven channel modeling, which includes diverse applications from radio map construction to modulation classification [74], is seen as a key enabler for future communication systems, promising to deliver more accurate and adaptable models than classical methods [58].

### D. Model Efficiency and Edge Deployment

A major bottleneck for deploying advanced AI models in real-world wireless systems, particularly on edge devices like AAVs and vehicles, is their computational and communication overhead. This has spurred significant research into model efficiency.

Federated Learning (FL), while designed to protect data privacy by training models locally on devices, introduces a significant communication burden due to the repeated transmission of model parameters between clients and a central server [75]. To mitigate this, various model compression techniques have been investigated. These techniques aim to reduce the size of the models being transmitted, for both upstream (client-to-server) and downstream (server-to-client) communication, by creating sparse models from dense ones without significant loss in accuracy [76]. Hierarchical FL architectures, which introduce intermediate aggregation servers, have also been proposed to reduce communication costs and better handle data heterogeneity in large-scale systems like the Industrial IoT [77].

The challenge of deploying AI extends to Digital Twins (DTs) as well, where tasks like adaptive caching are crucial for optimizing performance in heterogeneous IoT environments [78]. The construction and maintenance of high-fidelity DTs for complex systems like air-ground networks or vehicular edge networks require substantial resources. Consequently, research has focused

on intelligent task offloading and resource management. For instance, federated deep reinforcement learning has been used to optimize task offloading in DT edge networks [79], while incentive mechanisms based on game theory have been designed to encourage participation in FL for dynamic DTs in air-ground networks [22]. These efforts, spanning from general IoT to specific domains like secure vehicular networks [80], highlight the critical need for communication- and computation-efficient AI solutions, which directly motivates our work on developing a lightweight yet powerful architecture for RM construction.

### E. State Space Models and Mamba

Our work is directly inspired by recent advances in structured SSMs, which have recently emerged as a powerful and highly efficient alternative to Transformers for modeling long-range dependencies.

1) *Foundations of State Space Models:* SSMs [43], originating from classical control theory, model a system by mapping a 1-D input sequence  $\mathbf{u}(t)$  to an output  $\mathbf{y}(t)$  via a latent state vector  $\mathbf{h}(t) \in \mathbb{R}^N$ . A continuous-time SSM is defined by the linear ordinary differential equations (ODEs) as follows.

$$\frac{d\mathbf{h}(t)}{dt} = \mathbf{A}\mathbf{h}(t) + \mathbf{B}\mathbf{u}(t), \quad (1)$$

$$\mathbf{y}(t) = \mathbf{C}\mathbf{h}(t) + \mathbf{D}\mathbf{u}(t), \quad (2)$$

where  $\mathbf{A} \in \mathbb{R}^{N \times N}$  is the state matrix that governs the internal state dynamics, and  $\mathbf{B} \in \mathbb{R}^{N \times 1}$ ,  $\mathbf{C} \in \mathbb{R}^{1 \times N}$  are projection matrices that map the input to the state and the state to the output, respectively.

For use in deep learning with discrete data, such as tokens or pixels, these continuous parameters must be discretized. A fixed sampling timestep  $\Delta$  is introduced, and a discretization rule, such as the zero-order hold (ZOH), transforms the continuous parameters into their discrete counterparts,  $\bar{\mathbf{A}}$  and  $\bar{\mathbf{B}}$ , which is as follows.

$$\bar{\mathbf{A}} = \exp(\Delta\mathbf{A}), \quad (3)$$

$$\bar{\mathbf{B}} = (\Delta\mathbf{A})^{-1}(\exp(\Delta\mathbf{A}) - \mathbf{I})\Delta\mathbf{B}, \quad (4)$$

The SSM can then be computed either as a standard recurrence, which is efficient for inference, or a global convolution, which is efficient for training. It can be formulated as follows.

$$\mathbf{h}_k = \bar{\mathbf{A}}\mathbf{h}_{k-1} + \bar{\mathbf{B}}\mathbf{u}_k, \quad (5)$$

$$\mathbf{y}_k = \mathbf{C}\mathbf{h}_k, \quad (6)$$

$$\bar{\mathbf{K}}_k = \mathbf{C}\bar{\mathbf{A}}^k\bar{\mathbf{B}}, \quad (7)$$

$$\mathbf{y} = \mathbf{u} * \bar{\mathbf{K}}. \quad (8)$$

While powerful, the computational cost of these formulations, especially for long sequences, limited early SSMs. The breakthrough came with structured variants, like S4 [43], which imposed structure (e.g., diagonal plus low-rank) on the  $\mathbf{A}$  matrix. This allowed for the efficient computation of the convolutional kernel  $\bar{\mathbf{K}}$  in near-linear time, enabling SSMs to rival Transformers on long-sequence benchmarks.

2) *Mamba. Selective State Spaces and Efficiency:* **Mamba** [44] represents an advancement in this area, overcoming a critical limitation of prior SSMS: their time- and input-invariance. Models like S4 use fixed parameters that are not data-dependent, meaning they cannot adapt their behavior to the specific input content. Mamba addresses this with two key innovations:

a) *Selection mechanism:* The key advantage of Mamba is its ability to perform content-based reasoning by making the SSM parameters input-dependent. For each token  $\mathbf{u}_k$  in the input sequence, Mamba dynamically generates specific parameters  $\Delta_k$ ,  $\mathbf{B}_k$ , and  $\mathbf{C}_k$  via linear projections from the input as follows.

$$\Delta_k = \text{softplus}(\text{Linear}_\Delta(\mathbf{u}_k)), \quad (9)$$

$$\mathbf{B}_k = \text{Linear}_B(\mathbf{u}_k) \in \mathbb{R}^{N \times 1}, \quad (10)$$

$$\mathbf{C}_k = \text{Linear}_C(\mathbf{u}_k) \in \mathbb{R}^{1 \times N}, \quad (11)$$

where  $\text{Linear}_\Delta$ ,  $\text{Linear}_B$ , and  $\text{Linear}_C$  are learnable projection layers. Note that the core state matrix  $\mathbf{A} \in \mathbb{R}^{N \times N}$  remains fixed. These dynamic parameters are then used to compute time-varying discrete matrices  $\bar{\mathbf{A}}_k$  and  $\bar{\mathbf{B}}_k$  for each step as follows.

$$\bar{\mathbf{A}}_k = \exp(\Delta_k \mathbf{A}), \quad (12)$$

$$\bar{\mathbf{B}}_k = (\Delta_k \mathbf{A})^{-1}(\exp(\Delta_k \mathbf{A}) - \mathbf{I})\Delta_k \mathbf{B}_k, \quad (13)$$

This results in a selective SSM recurrence where the state update is modulated by the input at every step as follows.

$$\mathbf{h}_k = \bar{\mathbf{A}}_k \mathbf{h}_{k-1} + \bar{\mathbf{B}}_k \mathbf{u}_k, \quad (14)$$

$$\mathbf{y}_k = \mathbf{C}_k \mathbf{h}_k. \quad (15)$$

This dynamic parameterization allows the model to selectively propagate or forget information in its latent state  $\mathbf{h}_k$ . By learning to modulate  $\Delta_k$ , for instance, the model can decide whether to focus on fine-grained local information (small  $\Delta_k$ ) or long-range dependencies (large  $\Delta_k$ ). This selectivity gives Mamba the context-aware reasoning of attention mechanisms without the quadratic computational cost.

b) *Hardware-aware parallel algorithm:* While the selection mechanism makes the model powerful, it breaks the time-invariance that allowed for efficient convolutional computation (8). A naive recurrent implementation (5) would be slow to train on parallel hardware like GPUs. Mamba's second core innovation is a **hardware-aware algorithm** that leverages the associative property of its scan operation. It expresses the computation in a way that can be parallelized for training, similar to prefix sums, while still collapsing back to a highly efficient recurrent form for autoregressive inference. This dual formulation allows Mamba to be trained at high speed on modern accelerators while retaining linear-time inference.

These innovations combined give Mamba the ability to model extremely long-range dependencies with a computational complexity that is linear in sequence length. This characteristic is particularly suitable for radio map construction. The radio environment is a system where the 'state' (pathloss strength) at one location is dependent on a long 'sequence' of interactions with obstacles across the entire map. Mamba's ability to efficiently

process this global context makes it uniquely suited for this task. While originally designed for 1D sequences, recent works like Vision Mamba [45], [81] have adapted it to 2D vision tasks by scanning images along multiple directions. Our work is inspired by this principle, adapting Mamba into a lightweight and synergistic architecture tailored for the physics-based generation task of radio map construction.

### III. SYSTEM MODEL AND PROBLEM FORMULATION

In this section, we formally define the environment for sampling-free radio map construction and formulate the task as a conditional image generation problem, drawing upon the detailed setup from prior work such as RadioDiff [39] for consistency and comparability.

#### A. Environment and Radio Map Representation

We consider a geographical area discretized into a 2D grid of size  $N \times N$ , where each grid cell (pixel) represents a small square area (e.g., 1m  $\times$  1m). The environment within this grid is characterized by the presence of various obstacles that affect radio wave propagation.

*Input Features:* The model is conditioned on the physical layout of the environment, which is provided as a multi-channel tensor. This input, denoted as  $\mathbf{C}_{in} \in \mathbb{R}^{C \times N \times N}$ , is composed of several channels that describe the scene:

- *Static Obstacle Map ( $\mathbf{H}_s$ ):* A binary matrix of size  $N \times N$ . An element  $h_{i,j}^s = 1$  indicates the presence of a permanent, static obstacle like a building at location  $(i, j)$ , while  $h_{i,j}^s = 0$  signifies free space. This map provides the primary, unchanging geometry of the environment.
- *Dynamic Obstacle Map ( $\mathbf{H}_d$ ):* A binary matrix of size  $N \times N$  that captures transient objects, such as vehicles. An element  $h_{i,j}^d = 1$  indicates the presence of a dynamic obstacle. This channel allows the model to account for temporary changes in the environment, which is crucial for dynamic RM (DRM) construction. For static RM (SRM) construction, this channel may be zero-filled or omitted.
- *Transmitter Location Map ( $\mathbf{R}$ ):* A binary matrix of size  $N \times N$  that serves as a one-hot encoding of the transmitter's position. The pixel corresponding to the transmitter's coordinates  $(x_{tx}, y_{tx})$  has a value of 1, and all other pixels are 0.

These maps, representing different aspects of the physical world, are concatenated channel-wise to form the complete input tensor for the neural network. In our experiments, following the RadioMapSeer benchmark [37],  $N = 256$ .

*Output Radio Map ( $\mathbf{P}$ ):* The model's objective is to generate the radio map, which is a single-channel, real-valued matrix  $\mathbf{P} \in \mathbb{R}^{N \times N}$ . Each element  $p_{i,j}$  in this matrix represents the predicted pathloss value in decibels (dB) at the grid location  $(i, j)$ . This dense map provides a comprehensive characterization of the signal coverage throughout the area for a given transmitter location and environmental state. Pathloss is typically a large negative value, which is normalized to a grayscale range (e.g.,  $[0, 1]$ ) for model training and prediction.

## B. Problem Formulation

The task of radio map construction is to learn a deterministic function,  $\mathcal{F}_\theta$ , parameterized by a set of learnable weights  $\theta$ , that maps the multi-channel environmental input  $\mathbf{C}_{\text{in}}$  to its corresponding radio map  $\mathbf{P}$ . This function must implicitly learn the complex underlying physics of radio wave propagation, including reflection, diffraction, and scattering. The problem can be formally stated as:

*Problem 1. (Radio Map construction):* Given a dataset  $\mathcal{D} = \{(\mathbf{C}_{\text{in}}^{(i)}, \mathbf{P}^{(i)})\}_{i=1}^M$  consisting of  $M$  pairs of environmental configuration tensors and their corresponding ground-truth radio maps. The goal is to find the optimal set of parameters  $\theta^*$  for a deep neural network  $\mathcal{F}_\theta$  that minimizes a predefined loss function  $\mathcal{L}(\cdot, \cdot)$  over the dataset:

$$\theta^* = \arg \min_{\theta} \mathbb{E}_{(\mathbf{C}_{\text{in}}, \mathbf{P}) \sim \mathcal{D}} [\mathcal{L}(\mathcal{F}_\theta(\mathbf{C}_{\text{in}}), \mathbf{P})]. \quad (16)$$

where  $\hat{\mathbf{P}} = \mathcal{F}_\theta(\mathbf{C}_{\text{in}})$  is the predicted radio map.

This task is fundamentally a **complex spatial mapping problem**. The network’s objective is to learn a deterministic function that translates an input (a single transmitter point and a building layout) into a complete, continuous-valued spatial distribution, the radio map  $\hat{\mathbf{P}}$ . The pathloss value at any given location  $(i, j)$  is not merely a function of its local neighborhood; it is dependent on the **entire global geometry** of the environment. Obstacles hundreds of meters away can cast long shadows (non-line-of-sight conditions) or create complex multi-path reflections that fundamentally alter the signal strength. This non-local nature of wave physics requires that a successful model must possess a powerful mechanism for capturing long-range spatial dependencies. A model with only a limited, local receptive field is inherently limited, as it cannot ‘see’ the distant obstacles that are crucial for an accurate prediction. This physical requirement motivated our design of RadioMamba.

## IV. THE PROPOSED RADIOMAMBA

To resolve the performance-efficiency dilemma, we designed RadioMamba, an architecture that leverages the hierarchical feature extraction of U-Net while enhancing its feature representation capabilities with a hybrid block that efficiently models both global and local spatial features.

### A. Overall Architecture

The overall architecture of RadioMamba is based on the well-established encoder-decoder design of U-Net, illustrated in Fig. 2. This structure is inherently suited for image-to-image tasks due to its ability to preserve spatial information across multiple scales.

- *Input:* The model accepts a multi-channel input tensor of size  $256 \times 256$ , representing the building map, transmitter location, and an auxiliary channel for inputs such as dynamic obstacles. An initial convolution projects this input into a high-dimensional feature space.
- *Encoder:* The encoder path consists of three stages. Each stage comprises a series of our custom *MambaConvBlocks* followed by a downsampling operation using a  $2 \times 2$

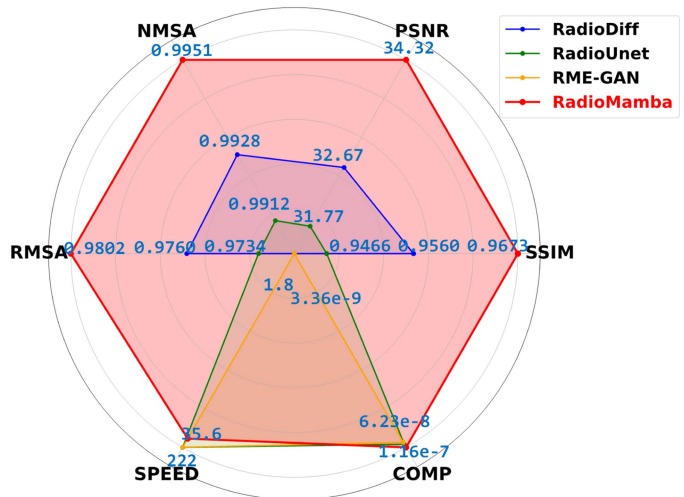


Fig. 1. A radar chart comparing four RM construction models across six metrics: SSIM, PSNR, NMSA (calculated as 1-NMSE), RMSA (calculated as 1-RMSE), SPEED (inverse of inference time), and COMP (model compactness, inverse of parameter count).

strided convolution that halves the spatial resolution (e.g.,  $256 \rightarrow 128 \rightarrow \dots \rightarrow 32$ ) while doubling the number of feature channels. This process builds a feature representation that captures increasingly abstract semantic information.

- *Bottleneck:* At the lowest resolution ( $32 \times 32$ ), the feature map passes through the final set of MambaConvBlocks, acting as the bottleneck which processes the most compressed and semantically rich feature representation.
- *Decoder:* The decoder path symmetrically mirrors the encoder. It consists of three stages that use transposed convolutions to progressively upsample the feature maps back to the original resolution. At each stage, the upsampled feature map is concatenated with the corresponding high-resolution feature map from the encoder via a skip connection. A  $1 \times 1$  convolution first fuses the concatenated features, which are then processed by a series of MambaConvBlocks to refine the representation.
- *Output:* A final  $1 \times 1$  convolutional layer projects the feature map from the last decoder stage back to the desired single output channel for the radio map, producing the final prediction.

The core innovation of our model lies not in the U-Net skeleton itself, but in the effective building blocks used at each stage of the encoder and decoder.

### B. The MambaConvBlock

Standard U-Nets use blocks of convolutional layers. To overcome their limited receptive field, we designed the **MambaConvBlock**, a hybrid module that operates with two parallel branches to process information concurrently, as shown in Fig. 3. This design allows the block to benefit from the best of both worlds: the local inductive bias of CNNs and the global receptive field of SSMs.

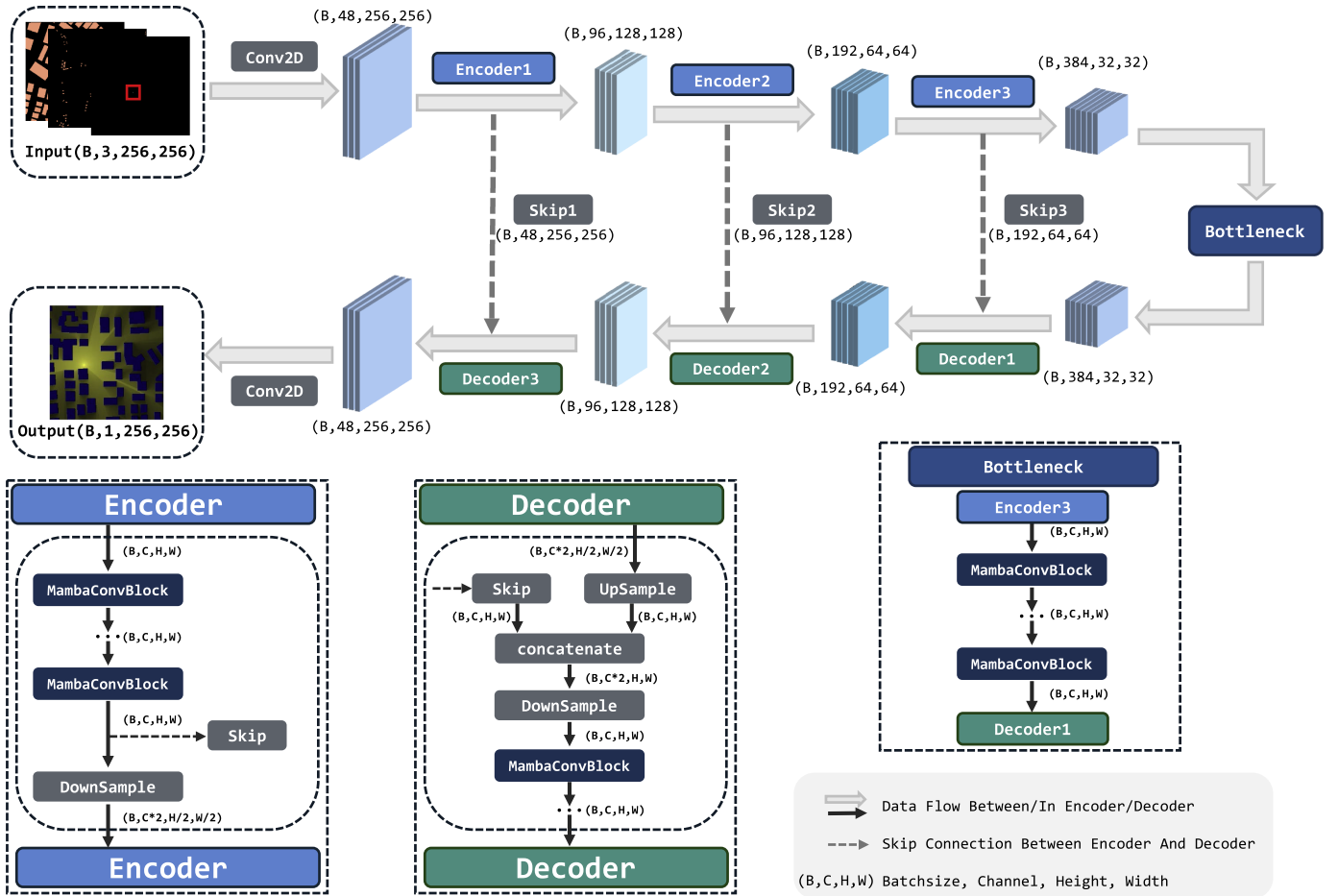


Fig. 2. The overall architecture of RadioMamba.

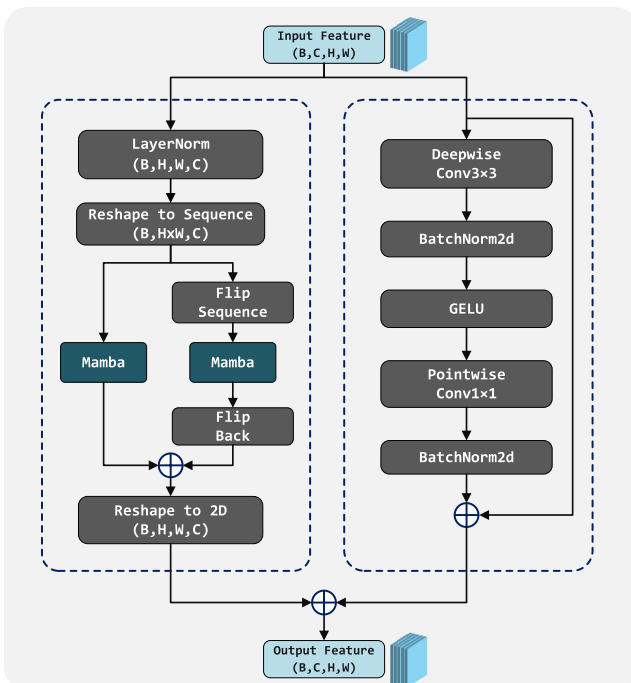


Fig. 3. Schematic of the proposed MambaConvBlock.

1) *Local Feature Branch*: This branch is engineered to extract local textures and high-frequency details, which are important for delineating the sharp, physically-correct shadow boundaries around environmental obstacles. To achieve this with maximum computational efficiency, the branch is constructed as a ResidualConvBlock which leverages the principle of depthwise separable convolutions. This modern architectural pattern factorizes a standard convolution into two distinct, more efficient operations.

Let the input tensor be  $\mathbf{X} \in \mathbb{R}^{C_{in} \times H \times W}$ . A standard 2D convolution would directly map this to an output with a computational cost, measured in multiply-accumulate operations (MACs) approximately as follows.

$$\text{Cost}_{\text{std}} = H \cdot W \cdot C_{in} \cdot C_{out} \cdot K_h \cdot K_w, \quad (17)$$

Depthwise separable convolutions reduce this cost by splitting the process.

a) *Depthwise convolution*: The first step is a depthwise convolution, which applies a single spatial filter to each input channel independently. This operation handles the spatial filtering, capturing patterns like edges, but does not mix information across channels. Its cost is as follows.

$$\text{Cost}_{\text{dw}} = H \cdot W \cdot C_{in} \cdot K_h \cdot K_w, \quad (18)$$

*b) Pointwise convolution:* The second step is a pointwise convolution (a  $1 \times 1$  convolution), which projects the intermediate features from the depthwise step into the final output channel space by linearly combining features across channels. This step is responsible for feature mixing, and its cost is as follows.

$$\text{Cost}_{\text{pw}} = H \cdot W \cdot C_{\text{in}} \cdot C_{\text{out}}, \quad (19)$$

The primary motivation for this factorization is the reduction in computational load. The ratio of the separable cost ( $\text{Cost}_{\text{dw}} + \text{Cost}_{\text{pw}}$ ) to the standard cost is given as follows.

$$\frac{\text{Cost}_{\text{dw}} + \text{Cost}_{\text{pw}}}{\text{Cost}_{\text{std}}} = \frac{1}{C_{\text{out}}} + \frac{1}{K_h K_w}, \quad (20)$$

For a typical  $3 \times 3$  kernel and a large number of output channels, this ratio approaches  $\frac{1}{9}$ , implying an almost 9-fold reduction in MACs and parameters.

The entire operation is wrapped in a residual connection. Let  $\mathcal{F}_{\text{conv}}$  represent the sequence of depthwise convolution, a GELU activation, and a pointwise convolution. The output of the ResidualConvBlock  $\mathbf{Y}_{\text{conv}}$  is as follows.

$$\mathbf{Y}_{\text{conv}} = \mathbf{X} + \mathcal{F}_{\text{conv}}(\mathbf{X}). \quad (21)$$

This residual connection is crucial for enabling stable training in a deep architecture, as it facilitates unimpeded gradient flow and ensures that fine-grained features are preserved and progressively refined.

*2) Global Context Branch:* This branch is the key to capturing long-range dependencies across the entire feature map, modeling the global nature of wave propagation. It employs our Mamba module, which we term **SS2D-Mamba**. To adapt the inherently 1D Mamba model to 2D image data, we first transform the input feature map into a sequence and then perform a bidirectional scan for comprehensive context aggregation [81].

Given an input feature tensor  $\mathbf{X} \in \mathbb{R}^{B \times C \times H \times W}$ , it is first normalized using LayerNorm for stable processing. It is then unrolled into a sequence of length  $L = H \times W$  via a standard raster-scan flattening operation, which we denote as  $\mathcal{S}$ . This yields the sequence  $\mathbf{x}_{\text{seq}} \in \mathbb{R}^{B \times L \times C}$  as follows.

$$\mathbf{x}_{\text{seq}} = \mathcal{S}(\text{LayerNorm}(\mathbf{X})), \quad (22)$$

This sequence is then processed by the core Mamba model, which we represent with the operator  $\mathcal{M}$ , in both forward and backward directions to ensure that the context for each token is non-causal and includes information from the entire sequence. Let  $\mathcal{R}$  define the sequence reversal operator, which flips a sequence along its length dimension. The forward and backward outputs are then computed as follows.

$$\mathbf{y}_{\text{fwd}} = \mathcal{M}(\mathbf{x}_{\text{seq}}), \quad (23)$$

$$\mathbf{y}_{\text{bwd}} = \mathcal{R}(\mathcal{M}(\mathcal{R}(\mathbf{x}_{\text{seq}}))). \quad (24)$$

The outputs from these two directional scans are then aggregated, typically via element-wise addition, and the resulting sequence is reshaped back into a 2D feature map  $\mathbf{X}_{\text{mamba}} \in \mathbb{R}^{B \times C \times H \times W}$ . This process allows every pixel in the feature map to efficiently integrate information from all other pixels, providing a comprehensive global context.

We acknowledge a theoretical limitation in this approach: radio wave propagation is physically isotropic, meaning it spreads uniformly in all directions, whereas the raster-scan flattening method is anisotropic, prioritizing horizontal and vertical spatial relationships. However, we hypothesize that this limitation is effectively mitigated by several factors in our architecture. First, the bidirectional scan captures dependencies in two primary directions. Second, the hierarchical structure of the U-Net, which processes features at multiple scales, helps to integrate contextual information more holistically. Finally, the parallel convolutional branch, with its strong local and isotropic inductive bias, complements the Mamba branch by reinforcing spatially-agnostic local features. This synergy allows the model to learn a robust quasi-isotropic representation that effectively models wave propagation physics. Future work could explore more advanced, isotropic scanning patterns, such as space-filling curves, to potentially align the model even more closely with the underlying physics.

*3) Branch Fusion:* The outputs from the two parallel branches are synergistically combined via simple element-wise addition. We chose this strategy for several reasons. First, it is computationally lightweight, introducing no additional parameters or latency, which aligns with our goal of maximizing efficiency. Second, addition allows the global context from the Mamba branch to act as a residual correction or modulation on the local features from the convolutional branch, a proven technique for effective feature integration. While alternative methods like channel-wise concatenation followed by a  $1 \times 1$  convolution could offer more learnable flexibility, they would increase model complexity and parameters. Our empirical results confirm that element-wise addition provides a powerful yet efficient mechanism for fusing multi-scale spatial information for this task. The outputs are as follows.

$$\mathbf{Y}_{\text{out}} = \mathbf{Y}_{\text{conv}} + \mathbf{X}_{\text{mamba}}. \quad (25)$$

This MambaConvBlock serves as the fundamental building unit throughout our RadioMamba architecture, enabling it to achieve better feature representation at every level of the hierarchy.

## V. EXPERIMENTS

### A. Experimental Setup

*1) Dataset and Preprocessing:* We conduct all experiments on the publicly available **RadioMapSeer** dataset [37], [39], a standard benchmark for this task. The dataset consists of 700 maps from several cities, including Berlin and London, sourced from OpenStreetMap, with 80 different transmitter locations simulated for each map, totaling 56,000 unique scenarios. The maps are represented as  $256 \times 256$  pixel images, where each pixel corresponds to a  $1\text{m} \times 1\text{m}$  area. Building heights are fixed at 25 meters, and transmitter/receiver heights are set to 1.5 meters. The ground truth RMs for this benchmark are generated using a high-fidelity dominant path model (DPM) simulation, which provides a physically accurate basis for evaluation. We evaluate performance on two sub-tasks:

- *SRM construction*: Only static buildings are considered as obstacles. The input tensor comprises channels for the building layout and transmitter location.
- *DRM construction*: Both static buildings and transient obstacles, which are randomly generated vehicles along roads, are considered. An additional channel for dynamic obstacles is included in the input tensor.

We adhere to the standard data split of 550 maps for training, 50 for validation, and 100 for testing, ensuring no geographical overlap between sets. Input maps (buildings, transmitters, etc.) are converted to binary images. The output pathloss maps, originally in dB, are normalized to a grayscale range of [0, 1] for model training and evaluation.

2) *Baselines*: We compare our proposed RadioMamba with several SOTA baselines that represent different architectural philosophies to provide a comprehensive evaluation:

- *RadioUNet [37]*: The influential CNN-based method. It is highly efficient but its accuracy is limited by the local receptive field of its convolutional operators.
- *RME-GAN [38]*: A representative GAN-based model. It is efficient at inference but known for training instability and potential for artifacts.
- *RadioDiff [39]*: The current SOTA in construction accuracy. This diffusion-based model generates high-fidelity maps via an iterative denoising process, but at the cost of being extremely parameter-heavy and having high inference latency.

3) *Implementation Details*: Our model was implemented using PyTorch and trained on NVIDIA A40 GPUs. We utilized the AdamW optimizer with an initial learning rate of  $9 \times 10^{-4}$  and a weight decay of  $10^{-4}$ . A cosine annealing scheduler was employed to manage the learning rate. To accelerate training and reduce memory footprint, we used 16-bit mixed-precision. The final RadioMamba model has approximately 8.6 million parameters, highlighting its lightweight nature.

4) *Loss Function*: To guide the network towards producing physically plausible and perceptually high-quality maps, we employ a composite loss function. The total loss  $\mathcal{L}_{\text{total}}$  is a weighted sum of four components as follows.

$$\mathcal{L}_{\text{total}} = w_1 \mathcal{L}_{\text{L1}} + w_2 \mathcal{L}_{\text{MSE}} + w_3 \mathcal{L}_{\text{SSIM}} + w_4 \mathcal{L}_{\text{Grad}}. \quad (26)$$

The components are as follows, where  $\hat{\mathbf{P}}$  is the predicted radio map and  $\mathbf{P}$  is the ground truth.

- *Mean absolute error ( $\mathcal{L}_{\text{L1}}$ ) and mean squared error ( $\mathcal{L}_{\text{MSE}}$ )*: These standard pixel-wise losses ensure numerical accuracy.
- *Structural similarity loss ( $\mathcal{L}_{\text{SSIM}}$ )*: Formulated as  $1 - \text{SSIM}(\hat{\mathbf{P}}, \mathbf{P})$ , this loss enhances perceptual quality by focusing on structural information.
- *Gradient loss ( $\mathcal{L}_{\text{Grad}}$ )*: To enforce sharp edges and preserve fine details, we compute the L1 loss between the gradients of the prediction and the target, obtained via a Sobel filter:  $\mathcal{L}_{\text{Grad}} = \|\nabla \hat{\mathbf{P}} - \nabla \mathbf{P}\|_1$ . This is crucial for accurately rendering diffraction patterns.

Through empirical validation, the weights were set to  $w_1 = 0.4$ ,  $w_2 = 0.1$ ,  $w_3 = 0.2$ ,  $w_4 = 0.3$  to achieve a balanced optimization.

5) *Evaluation Metrics*: We provide a comprehensive evaluation using four metrics for construction quality and three for computational efficiency.

- *Accuracy Metrics*: normalized mean square error (NMSE) and root mean square error (RMSE).
- *Perceptual Metrics*: structural similarity index measure (SSIM) and peak signal-to-noise ratio (PSNR, in dB).
- *Efficiency Metrics*: inference time (seconds per map), model parameters (millions), and GPU memory usage (MB). Lower is better for all three.

## B. Results and Analysis

In this section, we analyze RadioMamba's performance, comparing it to other methods in terms of accuracy and efficiency. Our results show high accuracy and efficiency, addressing the trade-off between the two.

1) *SRM Construction Results*: The SRM task serves as a fundamental benchmark for a model's ability to learn propagation physics. In Table II, the best-performing metrics are highlighted in bold red, while the second-best are underlined in blue.

As shown in Table II, RadioMamba achieves lower error and higher similarity scores compared to all baselines, including the RadioDiff model. The reduction in NMSE from 0.0072 (RadioDiff) to 0.0050 represents a **30.5% improvement**, while the RMSE is reduced by **17.1%**. This performance gain can be attributed to the model's hybrid architectural design. The pathloss at any point is governed by global phenomena, such as large-scale shadowing from distant building clusters. The Mamba branch, with its global receptive field, is equipped to model these long-range interactions. Simultaneously, the convolutional branch captures fine-grained, local diffraction patterns. By combining these two capabilities, RadioMamba develops a more holistic and physically consistent understanding of the radio environment than models reliant on purely local (CNNs) or purely iterative (Diffusion) mechanisms.

2) *DRM Construction Results*: The DRM task introduces additional complexity by adding transient obstacles, simulating a more realistic urban environment. The results are presented in Table II.

As expected, the performance of all models degrades in this more challenging scenario. However, RadioMamba maintains its lead, once again outperforming all competitors by a wide margin. It achieves an NMSE of 0.0063, a **30% improvement** over RadioDiff's 0.0090. This result demonstrates the model's adaptability. The model effectively leverages its global context awareness (via Mamba) to account for the large-scale changes in signal clutter introduced by dynamic obstacles, while its local feature extraction (via CNN) precisely delineates the new, sharp shadows they cast. This confirms that the benefits of our hybrid design are not limited to static cases but are equally potent in complex, changing environments.

3) *Efficiency and Complexity*: While accuracy is crucial, practical deployment in 6G systems is ultimately dictated by efficiency. Table III provides a comparison of model complexity and resource consumption, where the percentages indicate the performance improvement of our model relative to RadioDiff.

TABLE I  
COMPARISON WITH TRADITIONAL AND AI-ENABLED PARADIGMS

Model/Method	AI-enabled	Global Context	Real-time Speed	High Accuracy
[27]	✗	✓	✗	✗
[34]	✗	✓	✗	✓
[38]	✓	✗	✓	✗
[39]	✓	✗	✓	✗
[40]	✓	✓	✗	✓
<b>RadioMamba(Ours)</b>	✓	✓	✓	✓

TABLE II  
QUANTITATIVE COMPARISON FOR SRM AND DRM CONSTRUCTION

	Model	NMSE ↓	RMSE (dB) ↓	SSIM ↑	PSNR (dB) ↑
SRM	RME-GAN [39]	0.0096	0.0279	0.9431	31.35
	RadioUNet [38]	0.0088	0.0266	0.9466	31.77
	RadioDiff [40]	<u>0.0072</u>	<u>0.0240</u>	<u>0.9560</u>	<u>32.67</u>
	<b>RadioMamba(Ours)</b>	<b>0.0050</b>	<b>0.0199</b>	<b>0.9673</b>	<b>34.32</b>
DRM	RME-GAN	0.0115	0.0306	0.9276	30.42
	RadioUNet	0.0107	0.0291	0.9291	30.89
	RadioDiff	<u>0.0090</u>	<u>0.0266</u>	<u>0.9432</u>	<u>31.71</u>
	<b>RadioMamba(Ours)</b>	<b>0.0063</b>	<b>0.0221</b>	<b>0.9580</b>	<b>33.33</b>

TABLE III  
EFFICIENCY AND COMPLEXITY COMPARISON. THE PERCENTAGES INDICATE THE PERFORMANCE IMPROVEMENT OF OUR MODEL RELATIVE TO RADIODIFF.

Model	Time (s) ↓	Params (M) ↓	Memory (MB) ↓
RME-GAN	0.0045	16.04	527
RadioUNet	0.0041	13.27	525
RadioDiff	0.5535	297.74	2067
<b>RadioMamba(Ours)</b>	<b>0.0280</b>	<b>8.60</b>	<b>808</b>
	<b>↓94.94%</b>	<b>↓97.11%</b>	<b>↓60.91%</b>

RadioMamba performs inference in just **28 milliseconds**, a speed that is **nearly 20 times faster** than RadioDiff’s 553 milliseconds. This performance places it well within the tens-of-milliseconds latency budget required for demanding real-time applications, such as network control loops or AAV navigation. In contrast, RadioDiff’s half-second delay relegates it to offline analysis tasks.

Furthermore, RadioMamba achieves its SOTA accuracy with only **8.6 million parameters**, a **97.1% reduction** compared to RadioDiff’s 297.74 million. This reduction in model size, along with its lower memory footprint (808 MB vs. 2067 MB), makes RadioMamba a prime candidate for deployment on edge devices with limited computational resources. It challenges the notion that top-tier accuracy must come at the cost of massive model size and high latency.

4) *Qualitative Comparison:* Quantitative metrics are crucial, but a visual inspection of the generated maps provides deeper insight into model performance. Figs. 4 and 5 display sample outputs from all models against the ground truth.

The qualitative results are consistent with our quantitative findings. The maps generated by RadioMamba are visually more accurate, exhibiting a close resemblance to the ground truth. Observe in Fig. 4 (e.g., row 4), how RadioMamba accurately captures the subtle gradations of signal decay in open spaces and the sharp, well-defined reflection patterns behind buildings. In contrast, RME-GAN and RadioUNet’s output appears smoothed and lacks fine detail, a direct consequence of its limited receptive field. Even RadioDiff, while generally accurate, tends to slightly blur the sharpest transitions. In the challenging DRM scenarios (Fig. 5), RadioMamba’s ability to render clean, crisp shadow

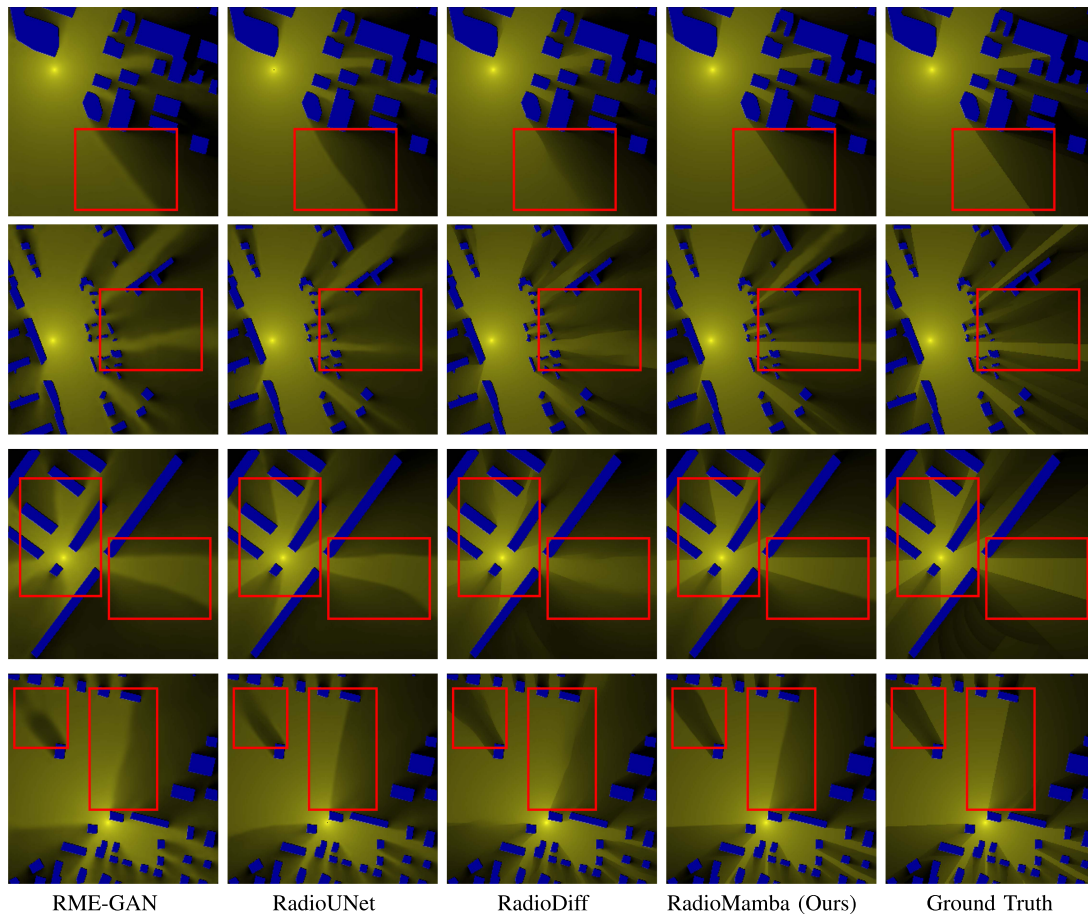


Fig. 4. Qualitative comparison for SRM construction, comparing (from left to right): RME-GAN, RadioUNet, RadioDiff, RadioMamba(Ours), and Ground Truth.

edges cast by both static and dynamic obstacles is particularly evident. This visual fidelity confirms its advanced understanding of the underlying wave propagation physics.

### C. Ablation Study

To validate our design and analyze the contributions of each component, we conduct a comprehensive series of ablation studies. Our investigation is structured into three parts: analyzing the core MambaConvBlock, justifying our choice of efficient convolutions, and evaluating the impact of the loss function components.

1) *Component Analysis of the MambaConvBlock*: The central innovation of RadioMamba is the synergistic MambaConvBlock, which combines a global Mamba branch and a local convolution branch. To demonstrate that this hybrid approach is better to using either component in isolation, we compare our full model against two variants: first, a **Conv-Only** baseline, where the Mamba branch is removed, and second, a **Mamba-Only** baseline, where the convolution branch is removed.

The qualitative results, shown in Fig. 6, are illustrative. The Conv-Only model, shown in Fig. 6(a), shows poor performance in predicting pathloss in areas far from the transmitter, producing

overly smoothed and blurry constructions. This highlights the inherent limitation of the local receptive field of convolutions for modeling global wave propagation. Conversely, the Mamba-Only model, shown in Fig. 6(b), captures the global structure but lacks fine-grained detail, resulting in blurry shadow edges and an inability to render sharp diffraction patterns. Our full RadioMamba block, shown in Fig. 6(c), combines the strengths of both, producing a map that is both globally coherent and locally precise, closely matching the ground truth shown in Fig. 6(d). This visually confirms that both branches are essential for achieving state-of-the-art results.

2) *Justification of Depthwise Separable Convolutions*: Our local feature branch utilizes depthwise separable convolutions to maximize efficiency. To quantify this design choice, we trained an alternative model using standard 2D convolutions in the MambaConvBlock. Table IV compares this variant against our proposed model.

Replacing depthwise separable convolutions with standard 2D convolutions did not yield a significant improvement in construction accuracy. The results were nearly identical, with the NMSE changing from 0.0050 to 0.0052. However, this substitution substantially increased the model's complexity, raising the parameter count from 8.6 M to 24.71 M. This finding confirms that depthwise separable convolutions offer a more effective

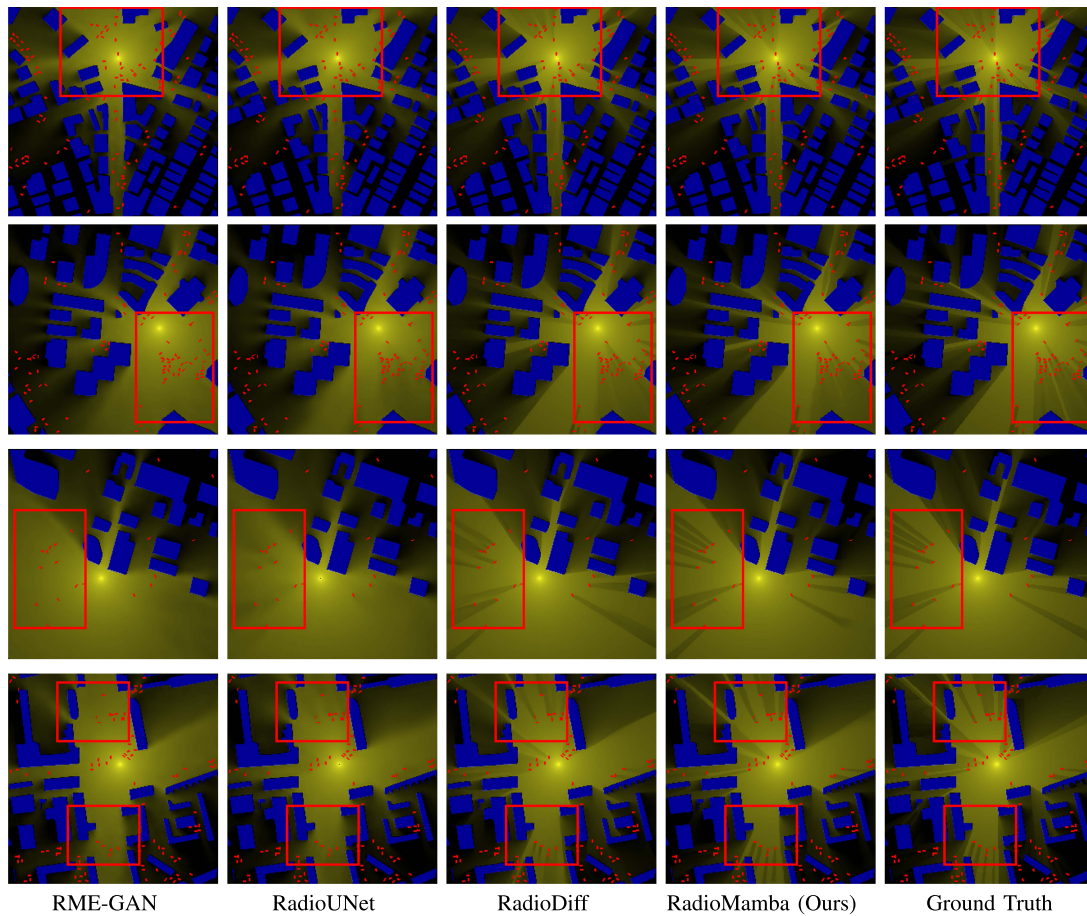


Fig. 5. Qualitative comparison for DRM construction, comparing (from left to right): RME-GAN, RadioUNet, RadioDiff, RadioMamba(Ours), and Ground Truth.

TABLE IV  
COMPARISON OF STANDARD VS. DEPTHWISE SEPARABLE CONVOLUTIONS

Conv Type	NMSE ↓	SSIM ↑	PSNR ↑	Params(M) ↓	Time(s) ↓	Memory(MB) ↓
Standard Conv	0.0052	0.9657	34.10	24.71	0.0318	902
Depthwise (Ours)	<b>0.0050</b>	<b>0.9673</b>	<b>34.32</b>	<b>8.60</b>	<b>0.0280</b>	<b>808</b>

TABLE V  
IMPACT OF LOSS FUNCTION COMPONENTS ON SRM CONSTRUCTION

Loss Configuration	NMSE ↓	RMSE (dB) ↓	SSIM ↑	PSNR (dB) ↑
$\mathcal{L}_{L1} + \mathcal{L}_{MSE}$	0.0075	0.0247	0.9541	32.39
Full Loss (Ours)	<b>0.0050</b>	<b>0.0199</b>	<b>0.9673</b>	<b>34.32</b>

balance, achieving comparable performance with a considerably smaller model size.

3) *Impact of Loss Function Components*: Finally, we analyze our composite loss function. We trained a model using only the pixel-wise  $\mathcal{L}_{L1}$  and  $\mathcal{L}_{MSE}$  losses (with weights set to 0.5 each), removing the SSIM and Gradient loss components. Table V and Fig. 7 compare this against our full loss configuration.

The results demonstrate the clear advantage of the composite loss function. The model trained with only L1+MSE loss performs worse across all evaluation metrics. Its NMSE and RMSE are higher, indicating a clear degradation in pixel-level accuracy. Furthermore, the lower SSIM and PSNR scores confirm a tangible loss in perceptual quality and structural fidelity. The visual evidence in Fig. 7 corroborates these findings; the

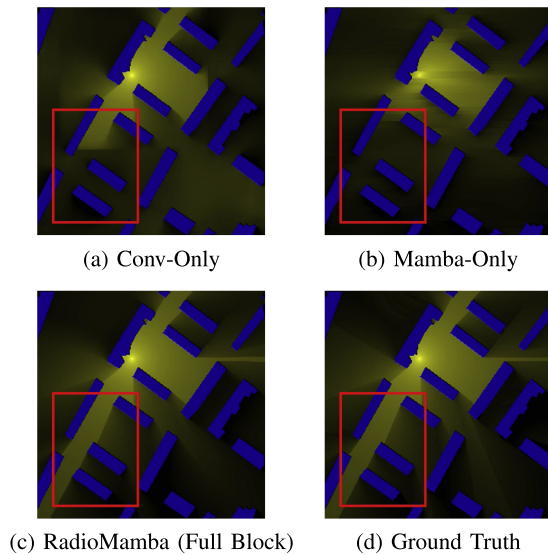


Fig. 6. Visual comparison of MambaConvBlock components.

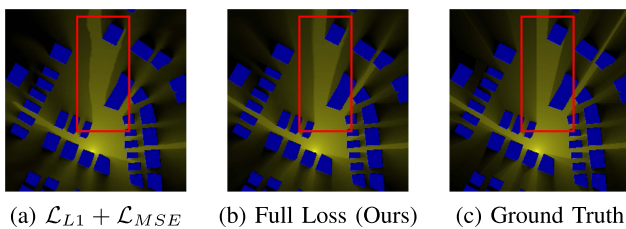


Fig. 7. Visual impact of the loss function.

map generated with only L1+MSE loss, shown in Fig. 7(a), is structurally inaccurate, with blurry and indistinct shadow edges. In contrast, our full loss function, shown in Fig. 7(b), which includes terms for structural similarity and gradients, produces a map with sharp, physically realistic boundaries that align closely with the ground truth in Fig. 7(c). This confirms that the SSIM and Gradient loss components are not merely for aesthetic improvement; they are essential for guiding the model to learn the correct high-frequency details and structural properties of the radio map, leading to a more accurate and physically plausible solution overall. With all four loss terms retained, we found the current weight ratio to be effective after multiple empirical trials. While a systematic hyperparameter search could potentially yield minor further improvements, we believe the current configuration robustly demonstrates the benefits of our composite loss.

## VI. CONCLUSION AND FUTURE WORK

In this paper, we have introduced RadioMamba, a hybrid Mamba-UNet architecture designed to address the accuracy-efficiency trade-off in radio map construction. Our model's core Mamba-Convolutional block synergistically captures both the global long-range dependencies essential for modeling wave physics and the local features that define environmental boundaries. This design leads to a more comprehensive feature representation, which translates to improved prediction accuracy.

Experiments demonstrated that RadioMamba achieves higher accuracy than previous methods, including diffusion models, while being faster and more parameter-efficient. These results suggest that an efficient architectural design aligned with the problem's underlying physics is a viable path to achieving high performance. This balance of accuracy and efficiency makes RadioMamba a suitable candidate for real-time 6G applications by providing the responsive, high-fidelity information needed for network digital twin synchronization or autonomous agent navigation. Future work will focus on extending the model to 3D scenarios, validating its generalization across diverse environments (e.g., rural), and addressing real-world deployment challenges such as bridging the domain gap between simulated and measured data for on-device applications.

## REFERENCES

- [1] S. Dang, O. Amin, B. Shihada, and M.-S. Alouini, "What should 6G be?," *Nature Electron*, vol. 3, no. 1, pp. 20–29, 2020.
- [2] F. Tao, B. Xiao, Q. Qi, J. Cheng, and P. Ji, "Digital twin modeling," *Manuf. Syst.*, vol. 64, pp. 372–389, 2022.
- [3] M. Alsabah et al., "6G wireless communications networks: A comprehensive survey," *IEEE Access*, vol. 9, pp. 148191–148243, 2021.
- [4] R. Sun et al., "A comprehensive survey of knowledge-driven deep learning for intelligent wireless network optimization in 6G," *IEEE Commun. Surv. Tuts.*, early access, doi: [10.1109/COMST.2025.3574765](https://doi.org/10.1109/COMST.2025.3574765).
- [5] L. Chang et al., "6G-enabled edge AI for metaverse: Challenges, methods, and future research directions," *Commun. Inf. Netw.*, vol. 7, no. 2, pp. 107–121, 2022.
- [6] Y. Wu, K. Zhang, and Y. Zhang, "Digital twin networks: A survey," *IEEE Internet Things J.*, vol. 8, no. 18, pp. 13789–13804, Sep. 2021.
- [7] X. Lin, L. Kundu, C. Dick, E. Obiodu, T. Mostak, and M. Flaxman, "6G digital twin networks: From theory to practice," *IEEE Commun. Mag.*, vol. 61, no. 11, pp. 72–78, Nov. 2023.
- [8] L. U. Khan, W. Saad, D. Niyato, Z. Han, and C. S. Hong, "Digital-twin-enabled 6G: Vision, architectural trends, and future directions," *IEEE Commun. Mag.*, vol. 60, no. 1, pp. 74–80, Jan. 2022.
- [9] K. B. Letaief, W. Chen, Y. Shi, J. Zhang, and Y.-J. A. Zhang, "The roadmap to 6G: AI empowered wireless networks," *IEEE Commun. Mag.*, vol. 57, no. 8, pp. 84–90, Aug. 2019.
- [10] N. Ansari, "Toward 6G and beyond: AI-driven synergies across terrestrial, non-terrestrial, and digital twin networks," *IEEE Wireless Commun.*, vol. 32, no. 3, pp. 4–5, Jun. 2025.
- [11] N. P. Kuruvatti, M. A. Habibi, S. Partani, B. Han, A. Fellan, and H. D. Schotten, "Empowering 6G communication systems with digital twin technology: A comprehensive survey," *IEEE Access*, vol. 10, pp. 112158–112186, 2022.
- [12] H. J. Hadi, Y. Cao, M. K. Khan, N. Ahmad, Y. Hu, and C. Fu, "AAV-NIDD: A dynamic dataset for cybersecurity and intrusion detection in AAV networks," *IEEE Trans. Netw. Sci. Eng.*, vol. 12, no. 4, pp. 2739–2757, Jul./Aug. 2025.
- [13] W. Wang, Y. Yang, L. U. Khan, D. Niyato, Z. Han, and M. Guizani, "Digital twin for wireless networks: Security attacks and solutions," *IEEE Wireless Commun.*, vol. 31, no. 3, pp. 278–285, Jun. 2024.
- [14] J. Morais and A. Alkhateeb, "Localization in digital twin MIMO networks: A case for massive fingerprinting," in *Proc. 2024 IEEE Int. Conf. Commun. Workshops*, 2024, pp. 276–281.
- [15] F. Liu et al., "Integrated sensing and communications: Toward dual-functional wireless networks for 6G and beyond," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 6, pp. 1728–1767, Jun. 2022.
- [16] A. Liu et al., "A survey on fundamental limits of integrated sensing and communication," *IEEE Commun. Surv. Tuts.*, vol. 24, no. 2, pp. 994–1034, Second Quarter 2022.
- [17] Z. Wei et al., "Integrated sensing and communication channel modeling: A survey," *IEEE Internet Things J.*, vol. 12, no. 12, pp. 18850–18864, Jun. 2025.
- [18] Z. Yin, N. Cheng, M. Wang, C. Li, and W. Xiang, "Conceal truth while show fake: T/f frequency multiplexing-based anti-intercepting transmission," *IEEE Trans. Inf. Forensics Secur.*, vol. 20, pp. 6884–6894, 2025.

- [19] S. Bi, J. Lyu, Z. Ding, and R. Zhang, "Engineering radio maps for wireless resource management," *IEEE Wireless Commun.*, vol. 26, no. 2, pp. 133–141, Apr. 2019.
- [20] S. Zhang and R. Zhang, "Radio map-based 3D path planning for cellular-connected AAV," *IEEE Trans. Wireless Commun.*, vol. 20, no. 3, pp. 1975–1989, Mar. 2021.
- [21] Z. Zhang and L. Dai, "Reconfigurable intelligent surfaces for 6G: Nine fundamental issues and one critical problem," *Tsinghua Sci. Technol.*, vol. 28, no. 5, pp. 929–939, 2023.
- [22] W. Sun, N. Xu, L. Wang, H. Zhang, and Y. Zhang, "Dynamic digital twin and federated learning with incentives for air-ground networks," *IEEE Trans. Netw. Sci. Eng.*, vol. 9, no. 1, pp. 321–333, Jan./Feb. 2022.
- [23] B. Zhu, E. Bederer, H. H. Nguyen, R. Barton, and J. Henry, "AAV trajectory planning in wireless sensor networks for energy consumption minimization by deep reinforcement learning," *IEEE Trans. Veh. Technol.*, vol. 70, no. 9, pp. 9540–9554, Sep. 2021.
- [24] J. Redondo, N. Aslam, J. Zhang, and Z. Yuan, "Multi-agent assessment with QoS enhancement for HD map updates in a vehicular network and multi-service environment," *IEEE Trans. Netw. Sci. Eng.*, vol. 12, no. 2, pp. 738–749, Mar./Apr. 2025.
- [25] G. Elghazaly, R. Frank, S. Harvey, and S. Safko, "High-definition maps: Comprehensive survey, challenges, and future perspectives," *IEEE Open J. Intell. Transp. Syst.*, vol. 4, pp. 527–550, 2023.
- [26] D. Mao, W. Shao, Z. Qian, H. Xue, X. Lu, and H. Wu, "Constructing accurate radio environment maps with kriging interpolation in cognitive radio networks," in *Proc. 2018 Cross Strait Quad-Regional Radio Sci. Wireless Technol. Conf.*, 2018, pp. 1–3.
- [27] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inform. Theory*, vol. 13, no. 1, pp. 21–27, Jan. 1967.
- [28] F. J. Breidt and J. D. Opsomer, "Local polynomial regression estimators in survey sampling," *Ann. Stat.*, vol. 28, pp. 1026–1053, 2000.
- [29] Q. Gong, F. Wu, D. Yang, L. Xiao, and Z. Liu, "3D radio map reconstruction and trajectory optimization for cellular-connected AAVs," *Commun. Inf. Netw.*, vol. 8, no. 4, pp. 357–368, 2023.
- [30] Y. Dong, C. He, Z. Wang, and L. Zhang, "Radio map assisted path planning for AAV anti-jamming communications," *IEEE Signal Process. Lett.*, vol. 29, pp. 607–611, 2022.
- [31] X. Liu, H. Song, and A. Liu, "Intelligent AAVs trajectory optimization from space-time for data collection in social networks," *IEEE Trans. Netw. Sci. Eng.*, vol. 8, no. 2, pp. 853–864, Second Quarter 2021.
- [32] S.-H. Oh and N.-H. Myung, "MIMO channel estimation method using ray-tracing propagation model," *Electron. Lett.*, vol. 40, no. 21, pp. 1350–1352, 2004.
- [33] B. Salski, P. Czekala, J. Cuper, P. Kopyt, H. Jeon, and W. Yang, "Electromagnetic modeling of radiowave propagation and scattering from targets in the atmosphere with a ray-tracing technique," *IEEE Trans. Antennas Propagat.*, vol. 69, no. 3, pp. 1588–1595, Mar. 2021.
- [34] P. M. Tshakwanda, S. T. Arzo, and M. Devetsikiotis, "Advancing 6G network performance: AI/ML framework for proactive management and dynamic optimal routing," *IEEE Open J. Comput. Soc.*, vol. 5, pp. 303–314, 2024.
- [35] M. Nandy and Y. D. Kumar, "Exploring the transition from 5G to 6G: Advancements in ultra-low latency, high-speed connectivity, and AI-driven network optimization," in *Proc. 2025 Int. Conf. Automat. Computation*, 2025, pp. 1234–1239.
- [36] G. Guo, F. Qamar, S. H. A. Kazmi, and M. H. ur Rehman, "Threat detection in the 6G enabled industrial IoT networks using deep learning: A review on the state-of-the-art solutions, challenges and future research directions," *Internet of Things*, vol. 33, 2025, Art. no. 101686.
- [37] R. Levie, Ç. Yapar, G. Kutyniok, and G. Caire, "RadioUNet: Fast radio map estimation with convolutional neural networks," *IEEE Trans. Wireless Commun.*, vol. 20, no. 6, pp. 4001–4015, Jun. 2021.
- [38] S. Zhang, A. Wijesinghe, and Z. Ding, "RME-GAN: A learning framework for radio map estimation based on conditional generative adversarial network," *IEEE Internet Things J.*, vol. 10, no. 20, pp. 18016–18027, Oct. 2023.
- [39] X. Wang et al., "RadioDiff: An effective generative diffusion model for sampling-free dynamic radio map construction," *IEEE Trans. Cogn. Commun. Netw.*, vol. 11, no. 2, pp. 738–750, Apr. 2025.
- [40] X. Luo, Z. Li, Z. Peng, M. Chen, and Y. Liu, "Denosing diffusion probabilistic model for radio map estimation in generative wireless networks," *IEEE Trans. Cogn. Commun. Netw.*, vol. 11, no. 2, pp. 751–763, Apr. 2025.
- [41] S. Khan, M. Naseer, M. Hayat, S. W. Zamir, F. S. Khan, and M. Shah, "Transformers in vision: A survey," *ACM Comput. Sur.*, vol. 54, no. 10s, pp. 1–41, 2022.
- [42] A. Dosovitskiy et al., "An image is worth 16 x 16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.
- [43] A. Gu, K. Goel, and C. Ré, "Efficiently modeling long sequences with structured state spaces," 2021, *arXiv:2111.00396*.
- [44] A. Gu and T. Dao, "Mamba: Linear-time sequence modeling with selective state spaces," 2023, *arXiv:2312.00752*.
- [45] X. Liu, C. Zhang, F. Huang, S. Xia, G. Wang, and L. Zhang, "Vision Mamba: A comprehensive survey and taxonomy," *IEEE Trans. Neural Netw. Learn. Syst.*, pp. 1–21, 2025, doi: [10.1109/TNNLS.2025.3610435](https://doi.org/10.1109/TNNLS.2025.3610435).
- [46] A. Hatamizadeh and J. Kautz, "MambaVision: A hybrid mamba-transformer vision backbone," in *Proc. Comput. Vis. Pattern Recognit. Conf.*, 2025, pp. 25261–25 270.
- [47] T. Dao and A. Gu, "Transformers are SSMs: Generalized models and efficient algorithms through structured state space duality," 2024, *arXiv:2405.21060*.
- [48] H. He et al., "MobileMamba: Lightweight multi-receptive visual mamba network," in *Proc. Comput. Vis. Pattern Recognit. Conf.*, 2025, pp. 4497–4507.
- [49] A. Ghazanfari, T. Van Chien, E. Björnson, and E. G. Larsson, "Model-based and data-driven approaches for downlink massive MIMO channel estimation," *IEEE Trans. Commun.*, vol. 70, no. 3, pp. 2085–2101, Mar. 2022.
- [50] M. Alim, M. Rahman, M. Hossain, and A. A. Nahid, "Analysis of large scale propagation models for mobile communications in urban area," 2010, *arXiv:1002.2187*.
- [51] I. K. Eltahir, "The impact of different radio propagation models for mobile ad hoc networks (MANET) in urban area environment," in *Proc. 2nd Int. Conf. Wireless Broadband Ultra Wideband Commun.*, 2007, pp. 30–30.
- [52] J. W. McKown and R. L. Hamilton, "Ray tracing as a design tool for radio networks," *IEEE Netw.*, vol. 5, no. 6, pp. 27–30, Nov. 1991.
- [53] Z. Yun and M. F. Iskander, "Ray tracing for radio propagation modeling: Principles and applications," *IEEE Access*, vol. 3, pp. 1089–1100, 2015.
- [54] N. Suga, Y. Maeda, and K. Sato, "Indoor radio map construction via ray tracing with RGB-D sensor-based 3D reconstruction: Concept and experiments in WLAN systems," *IEEE Access*, vol. 11, pp. 24863–24874, 2023.
- [55] Y. Corre and Y. Lohanen, "Three-dimensional urban EM wave propagation model for radio network planning and optimization over large areas," *IEEE Trans. Veh. Technol.*, vol. 58, no. 7, pp. 3112–3123, Sep. 2009.
- [56] M. K. Habib, S. A. Ayankoso, and F. Nagata, "Data-driven modeling: Concept, techniques, challenges and a case study," in *Proc. 2021 IEEE Int. Conf. Mechatron. Automat.*, 2021, pp. 1000–1007.
- [57] J. Jagannath, N. Polosky, A. Jagannath, F. Restuccia, and T. Melodia, "Machine learning for wireless communications in the Internet of Things: A comprehensive survey," *Ad Hoc Netw.*, vol. 93, 2019, Art. no. 101913.
- [58] M. Yang et al., "AI-enabled data-driven channel modeling for future communications," *IEEE Commun. Mag.*, vol. 62, no. 4, pp. 112–118, Apr. 2024.
- [59] X. Wang et al., "RadioDiff- $k^2$ : Helmholtz equation informed generative diffusion model for multi-path aware radio map construction," 2025, *arXiv:2504.15623*.
- [60] X. Wang et al., "RadioDiff-3D: A 3D  $\times$  3D radio map dataset and generative diffusion based benchmark for 6G environment-aware communication," *IEEE Trans. Netw. Sci. Eng.*, early access, Jul. 18, 2025, doi: [10.1109/TNSE.2025.3590545](https://doi.org/10.1109/TNSE.2025.3590545).
- [61] X. Wang et al., "RadioDiff-Inverse: Diffusion enhanced Bayesian inverse estimation for ISAC radio map construction," 2025, *arXiv:2504.14298*.
- [62] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Computer-Assisted Intervention*, Munich, Germany, 2015, pp. 234–241.
- [63] I. Goodfellow et al., "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, 2020.
- [64] C. Zou, F. Yang, J. Song, and Z. Han, "Generative adversarial network for wireless communication: Principle, application, and trends," *IEEE Commun. Mag.*, vol. 62, no. 5, pp. 58–64, May 2024.
- [65] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, *arXiv:1411.1784*.
- [66] Z. Zhang, G. Zhu, J. Chen, and S. Cui, "Fast and accurate cooperative radio map estimation enabled by GAN," in *Proc. 2024 IEEE Int. Conf. Commun. Workshops*, 2024, pp. 1641–1646.
- [67] J. Ho, A. Jain, and P. Abbeel, "Denosing diffusion probabilistic models," in *Proc. Int. Conf. Adv. Neural Inf. Process.*, 2020, pp. 6840–6851.
- [68] N. Van Huynh et al., "Generative AI for physical layer communications: A survey," *IEEE Trans. Cogn. Commun. Netw.*, vol. 10, no. 3, pp. 706–728, Jun. 2024.

- [69] S. Fu et al., "CKMDiff: A generative diffusion model for CKM construction via inverse problems with learned priors," 2025, *arXiv:2504.17323*.
- [70] C. Zhao et al., "Enhancing physical layer communication security through generative AI with mixture of experts," *IEEE Wireless Commun.*, vol. 32, no. 3, pp. 176–184, Jun. 2025.
- [71] S. Ergun, I. Sammour, and G. Chalhoub, "A survey on how network simulators serve reinforcement learning in wireless networks," *Comput. Netw.*, vol. 234, 2023, Art. no. 109934.
- [72] X. Du et al., "Multi-agent reinforcement learning for dynamic resource management in 6G in-X subnetworks," *IEEE Trans. Wireless Commun.*, vol. 22, no. 3, pp. 1900–1914, Mar. 2023.
- [73] Q. Guo, F. Tang, and N. Kato, "Federated reinforcement learning-based resource allocation in D2D-enabled 6G," *IEEE Netw.*, vol. 37, no. 5, pp. 89–95, Sep. 2023.
- [74] Q. Xuan et al., "AvgNet: Adaptive visibility graph neural network and its application in modulation classification," *IEEE Trans. Netw. Sci. Eng.*, vol. 9, no. 3, pp. 1516–1526, May/Jun. 2022.
- [75] M. Beitollahi and N. Lu, "Federated learning over wireless networks: Challenges and solutions," *IEEE Internet Things J.*, vol. 10, no. 16, pp. 14749–14763, Aug. 2023.
- [76] S. M. Shah and V. K. Lau, "Model compression for communication efficient federated learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 9, pp. 5937–5951, Sep. 2023.
- [77] J. Yang, W. Jiang, and L. Nie, "Hypernetworks-based hierarchical federated learning on hybrid non-IID datasets for digital twin in industrial IoT," *IEEE Trans. Netw. Sci. Eng.*, vol. 11, no. 2, pp. 1413–1423, Mar./Apr. 2024.
- [78] J. Tan, F. Tang, M. Zhao, and Y. Zhu, "Adaptive caching scheme for jointly optimizing delay and energy consumption in heterogeneous digital twin IoT," *IEEE Trans. Netw. Sci. Eng.*, vol. 10, no. 6, pp. 4020–4032, Nov./Dec. 2023.
- [79] Y. Dai, J. Zhao, J. Zhang, Y. Zhang, and T. Jiang, "Federated deep reinforcement learning for task offloading in digital twin edge networks," *IEEE Trans. Netw. Sci. Eng.*, vol. 11, no. 3, pp. 2849–2863, May/Jun. 2024.
- [80] J. Kang, Z. Xiong, D. Niyato, D. Ye, D. I. Kim, and J. Zhao, "Toward secure blockchain-enabled Internet of Vehicles: Optimizing consensus management using reputation and contract theory," *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 2906–2920, Mar. 2019.
- [81] Y. Liu et al., "VMamba: Visual state space model," in *Proc. 38th Int. Conf. Adv. Neural Inf. Process.*, 2024, pp. 103031–1030.



**Honggang Jia** (Student Member, IEEE) is currently working toward the M.S. degree with Xidian University, Xi'an, China. His research interests include intelligent networking and data-driven construction of radio maps.



**Nan Cheng** (Senior Member, IEEE) received the B.E. and M.S. degrees from the Department of Electronics and Information Engineering, Tongji University, Shanghai, China, in 2009 and 2012, respectively, and the Ph.D. degree from the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada, in 2016. From 2017 to 2019, he was a Postdoctoral Fellow with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada. He is currently a Professor with the State Key Laboratory of ISN and School of Telecommunication Engineering, Xidian University, Shaanxi, China. His research interests include B5G/6G, space-air-ground integrated network, Big Data in vehicular networks, self-driving systems, performance analysis, MAC, opportunistic communication, and application of AI for vehicular networks.



**Xiucheng Wang** (Member, IEEE) is currently working toward the Ph.D. degree with Xidian University, Shaanxi, China. His research focuses on machine learning of the wireless network.



**Conghao Zhou** (Member, IEEE) received the B.Eng. degree from Northeastern University, Shenyang, China, in 2017, the M.Sc. degree from the University of Illinois at Chicago, Chicago, IL, USA, in 2018, and the Ph.D. degree in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 2022. He is currently a Postdoctoral Fellow with the University of Waterloo. His research interests include space-air-ground-integrated networks, network slicing, and machine learning for wireless networks.



**Ruijin Sun** (Member, IEEE) received the Ph.D. degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 2019. From 2017 to 2018, she was a Visiting Student with the University of Waterloo, Waterloo, ON, Canada. From 2019 to 2021, she was a Joint Postdoctoral Fellow with Peng Cheng Laboratory, Shenzhen, China, and Tsinghua University, Beijing. She is currently a Lecturer with the School of Telecommunications Engineering and the State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an, China. Her research interests include knowledge-driven wireless resource allocation.



**Xuemin (Sherman) Shen** (Fellow, IEEE) received the Ph.D. degree in electrical engineering from Rutgers University, New Brunswick, NJ, USA, in 1990. He is currently a University Professor with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. His research interests include network resource management, wireless network security, Internet of Things, AI for networks, and vehicular networks. Dr. Shen was the recipient of the "West Lake Friendship Award" from Zhejiang Province in 2023, President's Excellence in Research from the University of Waterloo in 2022, Canadian Award for Telecommunications Research from the Canadian Society of Information Theory (CSIT) in 2021, R.A. Fessenden Award in 2019 from IEEE, Canada, Award of Merit from the Federation of Chinese Canadian Professionals (Ontario) in 2019, James Evans Avant Garde Award in 2018 from the IEEE Vehicular Technology Society, Joseph LoCicero Award in 2015 and Education Award in 2017 from the IEEE Communications Society (ComSoc), and Technical Recognition Award from Wireless Communications Technical Committee (2019) and AHSN Technical Committee (2013). He was the Editor-in-Chief of IEEE INTERNET OF THINGS JOURNAL, IEEE NETWORK, and *IET Communications*. He is also the Past President of the IEEE Communications Society. He was the Vice President for Technical & Educational Activities, Vice President for Publications, Member-at-Large on the Board of Governors, Chair of the Distinguished Lecturer Selection Committee, and Member of IEEE Fellow Selection Committee of the ComSoc. He is also a registered Professional Engineer of Ontario, Canada, an Engineering Institute of Canada Fellow, Canadian Academy of Engineering Fellow, Royal Society of Canada Fellow, Chinese Academy of Engineering Foreign Member, International Fellow of the Engineering Academy of Japan, and a Distinguished Lecturer of the IEEE Vehicular Technology Society and Communications Society.