RFCanvas: Modeling RF Channel by Fusing Visual Priors and Few-shot RF Measurements

Xingyu Chen
University of California San Diego
xic063@ucsd.edu

Zihao Feng
University of California San Diego
zif004@ucsd.edu

Ke Sun
University of California San Diego
University of Michigan Ann Arbor
kesun@ucsd.edu

Kun Qian
University of Virginia
kunqian@virginia.edu

Xinyu Zhang
University of California San Diego
xyzhang@ucsd.edu

ABSTRACT

Accurate and responsive simulation of radio frequency (RF) signal propagation is crucial for designing wireless systems operating in dynamic environments. Conventional ray tracing approaches struggle to accurately model the intricate geometries and material properties of objects that impact propagation. Recently proposed neural scene representations can learn such intricacies from RF data, but they treat the entire scene as implicit neural networks, necessitating retraining with a massive amount of RF data upon any environmental changes. In this paper, we propose RFCanvas, which fuses visual priors and RF measurements to achieve high accuracy for realistic scenes and be responsive to environmental changes. To ensure compatibility between visual priors and RF measurements, we introduce RFCanvas scene representations that model shapes and materials of substantial objects with tensorial fields and signed distance fields. We further extract motion information from visual priors to adapt RFCanvas scene representations to scene dynamics. RFCanvas is built upon an end-to-end optimization framework with differentiable RF simulation. Extensive evaluations across real-world wireless communication and sensing environments demonstrate RFCanvas's superiority over both existing methods.

CCS CONCEPTS

• **Human-centered computing** → **Ubiquitous and mobile computing**; • **Computer systems organization** → **Embedded and cyber-physical systems**.

KEYWORDS

RF Simulation, Channel Estimation, Ray Tracing, Differentiable Computing

ACM Reference Format:

Xingyu Chen, Zihao Feng, Ke Sun, Kun Qian, and Xinyu Zhang. 2024. RFCanvas: Modeling RF Channel by Fusing Visual Priors and Few-shot RF Measurements. In *ACM Conference on Embedded Networked Sensor Systems (SenSys '24), November 4–7, 2024, Hangzhou, China.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3666025.3699351

SenSys '24, November 4–7, 2024, Hangzhou, China

ACM ISBN 979-8-4007-0697-4/24/11…$15.00
https://doi.org/10.1145/3666025.3699351

1 INTRODUCTION

Simulations at the radio frequency (RF) bands have been driving the design and deployment of diverse wireless systems. Precise characterization of the RF propagation environment is pivotal for optimizing the deployment of wireless network infrastructures [43, 49] and planning the spatial protection zones for spectrum sharing [44]. Explicit modeling of the intricate multipath effects arising from signal interactions with the environment is also critical for enhancing the accuracy of wireless sensing applications, such as object detection, localization, and imaging [15, 46, 50, 56]. In addition, RF simulations are indispensable for designing and optimizing emerging electromagnetic (EM) structures [21, 27], such as reconfigurable intelligent surfaces that can tailor the RF propagation towards programmable wireless environments. Despite recent advancements in simulation accuracy [26, 57], the long simulation time and slow response to environmental changes limit the use of RF simulation in practice, especially for wireless systems with stringent downtime. Achieving rapid updates of RF simulation with sufficient accuracy has been a long-standing challenge, hindered by the propagation environment's inherent complexities and dynamics.

Table 1: Comparison of RF Simulation Approaches

RF Simulation Approach	Visual Priors	RF Measurements	Accuracy	Dynamic Scenes
Ray Tracing	Complete 3D Meshes	None	Low	✓
Neural based	None	Dense	High	×
RFCanvas	**Monocular Images**	**Sparse**	**High**	✓

Ray tracing algorithms have been widely adopted by the wireless industry [35, 38, 49, 51] for RF simulation. These algorithms model EM waves as numerous "rays" emanating from a transmitter, interacting with objects in the environment (through scattering, penetration, attenuation, etc.), and ultimately being captured by a receiver. RF ray tracing mainly relies on visual priors to reconstruct 3D models of the environment as inputs, which can be explicitly modified to account for environmental changes. However, conventional ray tracing methods suffer from the notorious sim-to-real gap due to the imperfect representation of objects' 3D structures and material properties. The visual models tend to overlook minute surface features that significantly impact RF propagation [40], especially

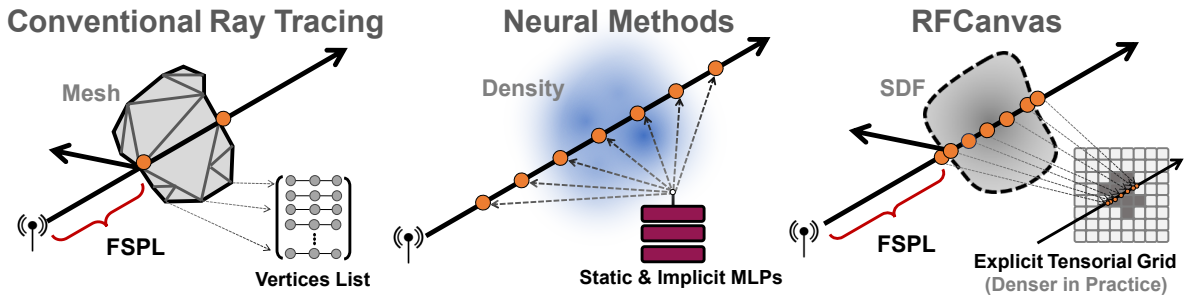

Figure 1: RFCanvas Comparison with Existing Methods.

in a multipath-rich environment. Moreover, visual 3D models fail to capture real-world objects' internal structures and nonuniform materials. The inherent vision-to-RF gap presents additional challenges, particularly in non-line-of-sight (NLOS) scenarios where cameras cannot detect objects that may influence RF propagation. This limitation can lead to inaccurate 3D structure generation and, consequently, unreliable RF predictions based on these incomplete models.

Recent data-driven approaches, such as NeRF^2 [57], have proposed using implicit neural representations to model the RF propagation within a scene. While these neural approaches demonstrate the capacity to learn intricate scene details and RF-object interactions from data, they treat the substance and radiance of the entire scene as implicit black boxes [28, 48], thereby sacrificing the flexibility inherent in conventional ray tracing methods. Additionally, these implicit neural models suffer from data inefficiency, requiring about 200 channel samples per square foot of environment. Adapting to environmental changes necessitates recollecting training samples and retraining the entire neural model—an often impractical burden.

In this paper, we introduce RFCanvas, a novel RF simulation framework that is accurate for realistic scenes and responsive to environmental changes. The core novelty of RFCanvas lies in its approaches of rapid generation and update of high-fidelity scene representations by fusing visual priors (such as monocular images) with RF channel samples of the scene, as shown in Table 1. RFCanvas leveraging the rich information of visual data to overcome data scarcity problems. Initially, RFCanvas utilizes the visual priors to reconstruct coarse yet fairly accurate scene representations. These representations are then further refined using a significantly reduced set of RF channel samples, in stark contrast to the existing scene-wise neural representation techniques [25, 57]. When scene changes, RFCanvas can swiftly adapt to the new scene by incorporating updated visual priors from camera inputs.

Realizing the full potential of RFCanvas entails three non-trivial technical challenges:

How to design a better RF representation for visual prior integration? To facilitate the integration of visual priors and enhance the RF sample efficiency, we adopt tensorial fields [5] in RFCanvas to model scenes. Specifically, the RF fields of scenes are explicitly represented by tensors, each of whose elements defines the attenuation and radiance coefficients at the corresponding physical location. The use of tensorial fields enables visual prior integration and improves RF data efficiency in two ways. First, unlike neural representations, tensorial representations decouple the RF fields at different locations. This decoupling not only enables partial editing with fewer RF measurements but also allows for the seamless integration of visual priors at specific spatial locations. Second, tensorial fields can be factorized into low-rank vectors and matrices via vector–matrix decomposition [5], which reduces the number of parameters and hence the RF measurements required for optimization.

How to generate scene representations from visual priors? An RF scene usually has a few objects sparsely distributed, with the rest of the space completely empty. While the interactions between objects and RF signals are complicated, RF signals in empty space can be accurately modeled via simple free-space propagation. Hence, RFCanvas utilizes visual priors to identify objects in the scene for rapid initialization and optimization of tensorial representations. Specifically, we use the depth information from camera images to determine the boundaries of objects. These boundaries are represented using signed distance fields (SDF) [32] in RFCanvas. An SDF specifies the distance from any point in space to the nearest surface of a 3D object. With the SDF, RFCanvas can localize interactions by tracking the distance between the front of a marching ray and the object's surface, known as sphere tracing [12]. The radiance tensors are then initialized from SDF using inverse spherical harmonics [8]. We further use camera images to obtain a coarse estimation of the objects' material composition, which helps initialize the scene's attenuation tensors. Thanks to the visual priors, the initial scene representations help significantly reduce the need for RF data and the optimization cost.

How to adapt scene representations to environmental dynamics using visual priors? A realistic scene often comprises dynamic objects. Adapting to a scene involves the detection and manipulation of changing regions. Unlike neural representation, which requires recollecting the real RF measurements and retraining the neural model, RFCanvas detects environmental changes from visual priors and directly modifies the explicit tensorial fields. Specifically, we model environmental changes with object motion vectors and develop a procedure to predict the corresponding change of dominant multipath reflection from these objects. The radiance coefficients in the dynamic regions are then recomputed via the inversion of spherical harmonics. With additional RF measurements collected, RFCanvas can further refine the tensorial fields of the scene. This design allows the rapid response of RFCanvas to environmental dynamics.

RFCanvas facilitates flexible editing of an RF simulation scene, such as modification of objects based on visual priors and configuration of radio hardware. This flexibility is highly desirable for designing and verifying wireless communication and sensing systems. Moreover, RFCanvas paves the way towards a modularized primitive library, analogous to the well-established practice of curating libraries of 3D mesh models in computer graphics. Such a library would enable large-scale, modularized RF simulation through pre-trained object representations, primed for plug-and-play integration into the design process of wireless systems.

We have implemented RFCanvas and conducted a comprehensive evaluation using various wireless testbeds, including WiFi/WiGig radios. Our microbenchmark experiments demonstrate that the RF signals generated by RFCanvas can accurately approximate the ground truth, achieving approximately 4 dB higher accuracy in RF-object characterization compared to conventional visual model based ray tracing. With a sparse dataset of merely 0.5 samples/sq ft of training data, RFCanvas surpasses the state-of-the-art (SOTA) $NeRF^2$ by around 1.5 dB in COTS WiFi channel prediction, and even higher for mmWave. In addition, RFCanvas exhibits robust adaptability to changes in a scene with a small set of 5 to 20 additional data samples.

Our contributions can be summarized as follows:

(i) We propose RFCanvas, a novel RF simulation framework that achieves comparable accuracy to neural-based methods while preserving the rapid responsiveness of conventional ray tracing.

(ii) We design a novel explicit representation capable of capturing objects' detailed structures and material properties. We develop an initialization procedure leveraging visual priors, significantly reducing the requirement for extensive RF measurements, and a few-shot fine-tuning scheme to efficiently refine the scene representations with limited RF measurements.

(iii) We conduct comprehensive evaluation of RFCanvas across diverse wireless environments. Our experiments demonstrate the effectiveness of RFCanvas and highlight its potential to foster new applications of RF simulation.

2 PRELIMINARIES AND RELATED WORKS

RF propagation and EM field simulation. Many wireless communication and sensing designs necessitate simulating how the RF signals from a source (transmitter) are distributed over space and interact with the environment. Conventional full-wave simulation approaches, such as Finite Element Method (FEM) [17] and Finite-Difference Time-Domain (FDTD) [42], can generate highly accurate simulation results. However, their substantial computational complexities limit their usage to small-scale RF device simulations, such as antenna design. Ray tracing approaches [6, 23, 35, 49, 51] enable efficient large-scale simulations by approximating electromagnetic waves as geometrical optics. A typical ray tracing simulation consists of transmitters (Tx), receivers (Rx), and mesh representation of objects in the scene. A mesh consists of numerous small polygons that delineate the contour of the object. The simulation starts by emitting rays from each Tx in all directions, with the ray density determined by predefined resolutions. Then the simulator detects the intersections between the rays and object meshes. At each intersection, the simulator replicates rays according to the number of interactions, redirects the rays based on the interaction types (*e.g.*, reflection, refraction, or diffraction). Additionally, it calculates the *directional radiance*, representing the power radiated by the ray along a specific direction after the interaction, which depends on the object's material properties (*e.g.*, permittivity and conductivity). The propagation of a ray terminates when it arrives at the Rx or reaches a predefined maximum number of interactions. Finally, all rays received by the Rx are coherently combined to obtain the received signal, with the relative phases of different rays calculated based on their respective path lengths.

Ray tracing requires accurate models of objects' geometries and material properties, which can hardly be generated using common sensors. Even for reflective objects with simple geometries, visual 3D models of the objects cannot easily capture the fine-grained shape details, which may look trivial visually but may interact significantly with RF signals. Recent work [49, 51] employed Lidar or RF transceivers themselves to create a 3D view of the environment as ray tracing input. However, these approaches still fall short of granularity and cannot model sophisticated multipath channels.

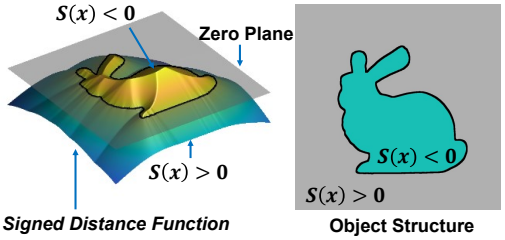

Figure 2: The geometry boundary is defined via the zero-level set of SDFs.

Neural representations for RF simulation. Neural scene representations have emerged as a powerful paradigm in computer graphics, gaining widespread adoption following the breakthrough of Neural Radiance Fields (NeRF) [28]. NeRF employs neural networks to implicitly model two key properties of a 3D scene: the volumetric density at any spatial location, and the radiance (color and brightness) emitted along any viewing direction from that location. By learning these view-dependent and location-dependent radiance mappings from a set of input images, NeRF can effectively render photorealistic images from unseen viewpoints, without requiring explicit 3D geometry or surface models.

However, when NeRF-rendered scenes are used in RF simulation, as in $NeRF^2$ [57], the entire scene has to be modeled as a monolithic blackbox, with no clearly defined boundaries of physical objects and interactions between physical objects and RF signals. This results in limited flexibility and high sensitivity to common environmental changes. To demonstrate this issue, we train a $NeRF^2$ model for the room. The result shows that when the scene remains unchanged, $NeRF^2$ is more accurate than conventional ray tracing. However, after moving some objects in the scene and recollecting RF measurements, the RSS error of $NeRF^2$ increases significantly by 4.58 dB. Moreover, existing methods, such as WiNeRT [30], partially integrate neural representations with conventional ray tracing by using neural networks to model materials' reflection parameters, which are then applied in ray tracing. However, these methods still rely on conventional 3D mesh models, which struggle to capture the fine-grained structural details of real objects.

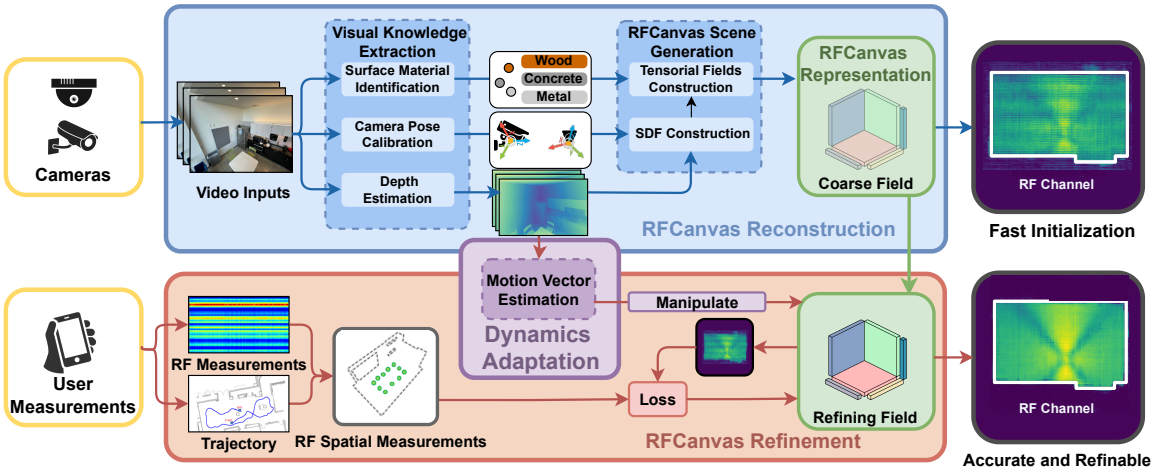

Figure 3: RFCanvas System Workflow.

Visual scene representation. To mitigate the shortcomings of mesh models, which capture fine details as well as internal and implicit structures, and the black-box nature of NeRF models, alternative representations have been widely studied.

Signed Distance Functions (SDF) [11, 13, 31, 32], offer an implicit surface representation, defining a field function $f(x, y, z)$ that returns the signed distance to the nearest surface point for any given 3D coordinate. The surface itself is represented by the zero-level set of this function, as shown in Fig 2. SDFs provide a compact representation suitable for ray tracing techniques and enable efficient collision detection. They also facilitate smooth blending between objects and support analytical operations like unions and intersections. Recent advancements in neural SDFs have shown promise in representing complex scenes with high fidelity. However, although SDF is efficient at capturing fine-grained geometric surfaces, it cannot capture the internal and multilayer structures common in RF simulation.

3D Gaussian Splatting (3DGS) [18] has emerged as a powerful technique for representing and rendering complex scenes. This method represents a scene as a set of 3D Gaussian primitives, each defined by parameters such as position, covariance matrix, and appearance attributes. A key feature of 3DGS is its use of spherical harmonics [29] to capture view-dependent appearance, allowing for efficient modeling of complex lighting interactions and material properties. 3DGS cannot explicitly represent object boundaries that are necessary for accurate multipath simulation; it can only model the Gaussian distribution of probabilities. RFCanvas adopts the idea of the spherical harmonic function to capture the sensitive view-dependent radiance attributes.

Tensorial Fields discretize 3D space into a regular grid of volumetric elements (voxels), each storing information such as occupancy or color. This representation allows for efficient spatial queries and is particularly suited for volumetric data RF fields. TensoRF [5] representing scenes using a compact factorized tensor representation. It decomposes 3D space into vector components along each dimension, enabling efficient modeling of view-dependent effects and fine geometric details. This factorized approach allows for faster training and rendering compared to traditional neural radiance fields while maintaining fidelity. However, the application of tensorial fields in RF simulation, as well as their dynamic adaptation and initialization, remains unexplored.

Summary: RFCanvas leverages the strengths of both neural-based and conventional RF simulation approaches, as illustrated in Fig. 1. It employs explicit tensorial grids of radiance and attenuation fields and SDFs from visual priors to capture the intricacies of RF signal propagation, similar to neural-based methods while maintaining the interpretability and adaptability characteristic of conventional techniques.

3 SYSTEM DESIGN

RFCanvas is a novel data-driven RF simulation system that leverages visual priors to reconstruct scenes and adapt to scene dynamics efficiently. The end-to-end workflow is shown in Fig. 3. The design components can be summarized into three parts:

(i) **RFCanvas representation** is a novel explicit scene representation for RF simulation based on tensorial fields and boundary-aware signed distance functions (SDFs).

(ii) **RFCanvas reconstruction** module initializes the RFCanvas scene from visual priors by inferring depth and material information, with no need for real RF samples.

(iii) **Dynamics adaptation** module adapts to scene changes by monitoring motion vectors derived from visual priors.

RFCanvas's workflow begins with its novel scene representation, combining tensorial fields and SDFs for an explicit, editable RF scene model. The reconstruction module then initializes this scene using visual priors from cameras, extracting depth and material information to set up SDFs and attenuation tensors, while using multipath tracing to initialize radiance tensors. As the environment changes, the dynamics adaptation module detects and classifies changes by estimating motion vectors from camera frames. It then updates the scene representation accordingly, using ray sample redirection for moving objects or partial reinitialization for new additions.

3.1 RFCanvas Scene Representation

RFCanvas scene represents a collection of objects to be simulated to estimate RF fields. To achieve rapid scene reconstruction and

adaptation from visual priors, the scene representation must meet two requirements: **1) Optimizability**: The scene representation needs to be differentiable to support efficient gradient-based optimization. **2) Editability**: The scene representation needed to be partially editable to integrate prior knowledge of object geometry, surface material types, and dynamic changes to corresponding regions of the scene. Mesh representations used in conventional ray tracing methods fail to meet the first requirement. Despite numerous studies on the differentiability of meshes in rendering, this does not transfer well to RF simulation. Additionally, mesh representations have low precision, as they can only model surface reflections and lack the ability to characterize complex scattering inside objects. In contrast, trainable neural representations, such as NeRF[2] and NeWRF, fulfill the first requirement. However, they are purely based on implicit neural networks, where assigning and updating material properties and geometrical boundaries to specific regions is infeasible. Further, they require complete retraining with massive RF measurements for updating.

Explicit and Editable Tensorial RF Field. Inspired by recent advances in computer graphics, we employ tensorial fields [5] to represent the geometric structures in RF simulation. Unlike MLPs in NeRF that implicitly map locations to RF radiance, we represent RF radiance fields as an explicit voxel grid of features. Given a 3D tensor $\mathcal{T} \in \mathbb{R}^{I\times J\times K}$ for a 3D cuboid space, each element $\mathcal{T}_{ijk}$ explicitly stores the directional radiance coefficients and attenuation coefficient of the corresponding voxel. Voxel-based data allows for high interpretability and editability, enabling users to manipulate individual voxel values similarly to editing image-like data. This includes operations such as cropping, deleting, and adding voxels, facilitating precise modifications.

Different from conventional voxel-based representations that have high memory costs and require long reconstruction times, tensorial representations use Vector-Matrix Decomposition (VM) [5] to factorize the high-dimensional tensors into multiple compact low-dimensional components as shown in Fig. 4. This can be expressed as:

$$\mathcal{T} = \sum_{r=1}^{R} \mathbf{v}_r^X \circ \mathbf{M}_r^{Y,Z} + \mathbf{v}_r^Y \circ \mathbf{M}_r^{X,Z} + \mathbf{v}_r^Z \circ \mathbf{M}_r^{X,Y}, \tag{1}$$

where $\mathbf{M}_r^{Y,Z} \in \mathbb{R}^{Y\times Z}, \mathbf{M}_r^{X,Z} \in \mathbb{R}^{X\times Z}, \mathbf{M}_r^{X,Y} \in \mathbb{R}^{X\times Y}$ are matrix factors. $\mathbf{v}_r^X, \mathbf{v}_r^Y, \mathbf{v}_r^Z$ are vector factors. R is the resolution of the cubic scene tensor.

Therefore, the query the tensor element is:

$$\mathcal{T}_{ijk} = \sum_{r=1}^{R} \sum_{m} A_{r,ijk}^{m}, m \in XYZ, \tag{2}$$

$$\mathcal{A}_{r,ijk}^{X} = \mathbf{v}_{r,i}^{X}\mathbf{M}_{r,jk}^{YZ}; \mathcal{A}_{r,ijk}^{Y} = \mathbf{v}_{r,j}^{Y}\mathbf{M}_{r,ik}^{XZ}; \mathcal{A}_{r,ijk}^{Z} = \mathbf{v}_{r,k}^{Z}\mathbf{M}_{r,ij}^{XY}. \tag{3}$$

Explicit Boundaries of Objects. Existing neural representations do not explicitly define object boundaries but point-wise samples of attenuation and radiance fields. Such dense representations waste model representability and computing resources, due to the imbalanced complexity of RF propagation inside and outside of objects. Specifically, in empty space, which usually occupies a large portion of a scene, RF signals simply follow free-space propagation. In contrast, complex RF interactions only happen at the boundary of and inside the substantial objects. However, these objects are usually just sparsely distributed in a scene.

To integrate the information on object boundaries in the scene representations, we model the object boundaries with SDF. Given a group of geometry Ω and its boundary $\partial\Omega$, each element in the SDF S^{Ω} stores the signed distance to the nearest surface:

$$S_p^{\Omega} = s(p, \Omega) \cdot d(p, \partial\Omega), \tag{4}$$

where $d(p, \partial\Omega) = \inf_{q\in\partial\Omega} d(p, q)$ denotes the minimal distance between the point p and any point q on the object's surface $\partial\Omega$. In addition, the sign $s(p, \Omega)$ indicates the *containing relation* between the point p and the object Ω, where a positive sign means that the point p is outside the object, *i.e.*, $p \notin \Omega$ while a negative sign means $p \in \Omega$.

An SDF can serve as a mask of tensorial fields, where only the tensorial fields with $S_p^{\Omega} <= 0$, *i.e.*, the boundary and inside region of the object, are sampled. The SDF is first initialized from visual priors and then optimized progressively along with tensorial fields, detailed in Section 3.2.3.

Trilinear Interpolation. RFCanvas represents the scene via tensorial fields. Although the tensor elements are discrete, they can represent continuous surfaces and geometry through trilinear interpolation.

For each sampling position, we calculate the weighted average of the 8 voxels surrounding the sampling position. The weights are determined by the distances between the voxel centers and the sampling position. However, evaluating 8 voxels for each sample can be costly. Utilizing VM decomposition, trilinear interpolation can be decomposed into a linear decomposition on vectors and bilinear interpolation on matrices, which further reduces the computational cost.

Simulation with RFCanvas scene representations. The propagation of an RF signal S from a transmitter (TX) to a receiver (RX) is characterized by the following equation:

$$S_{\mathrm{RX}} = a \cdot S_{\mathrm{TX}} \cdot e^{-j2\pi f_c \tau} \tag{5}$$

where S_{RX} denotes the received signal, and S_{TX} represents the transmitted signal. The complex factor $a = |a|e^{j\theta}$ accounts for both amplitude attenuation $|a|$ and phase shift θ introduced by the channel. The term $e^{-j2\pi f_c \tau}$ describes the additional phase shift due to the propagation delay, where f_c is the carrier frequency and τ is the time delay. This formulation effectively captures the combined effects of attenuation and phase alteration as the signal traverses the distance from TX to RX.

We use differentiable volumetric ray marching to simulate the RF signal propagation. The 3D position along a ray from receiver location p_0 at direction ω is defined as $P(t, \omega) = p_0 + \omega t.$, where t is the distance from the transmitter. The total attenuation between the point $P(t, \omega)$ and the receiver P_{RX} is

$$\begin{aligned} A_{P(t,\omega)\to P_{RX}} &= \exp\left(\int_0^t \left[\mathrm{FSPL}(\tilde{t}) \cdot (1 - f(\tilde{t})) + \mathcal{T}^{\alpha}_{P(\tilde{t},\omega)} \cdot f(\tilde{t})\right] d\tilde{t}\right) \\ &\approx \exp\left(\sum_{i=1}^{N} \left[\mathrm{FSPL}(t_i) \cdot (1 - f(t_i)) + \mathcal{T}^{\alpha}_{P(\tilde{t},\omega)} \cdot f(t_i)\right] \Delta t\right), \end{aligned} \tag{6}$$

where $\mathrm{FSPL}(t)$ is the free space path loss at distance t, $\mathcal{T}_{\alpha}(t, \omega)$ is the tensorial field sample for attenuation along the ray at distance

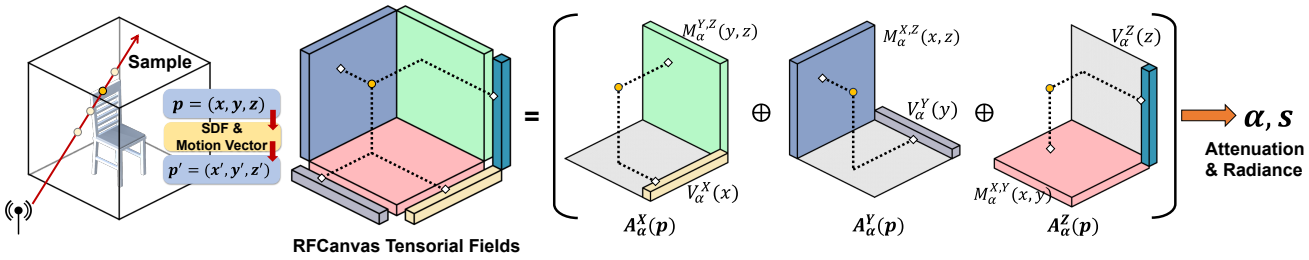

Figure 4: RFCanvas Tensorial RF Fields.

t, $f(t)$ is a binary function based on SDF sample that determines whether the signal is in a medium (1) or in free space (0) at position t, Δt is the sampling step size, and N is the number of sampling points along the ray. The resulting received signal $R(\omega)$ of the ray is computed as:

$$R(\omega) = \int_0^R \underbrace{A_{P(t,\omega)\to P_{RX}}}_{\text{Attenuation}} \cdot \underbrace{\mathcal{T}^s_{P(t,\omega)} \cdot Y_4^m(-\omega_\theta, -\omega_\phi)}_{\text{Radiance}}\, dt \tag{7}$$

Spherical Harmonics $Y_4^m(\cdot)$ are naturally adept at representing directional data. Their completeness and rotational invariance enable them to accurately capture the sensitivity of mmWave signals to minute directional changes while providing the ability to analyze signals across different frequency scales. The Spherical Harmonic function is defined as:

$$Y_4^m(\theta,\phi) = \sqrt{\frac{(2l+1)}{4\pi}\frac{(l-m)!}{(l+m)!}} P_l^m(\cos\theta) e^{im\phi} \tag{8}$$

where Y_4^m represents the spherical harmonic function, θ and ϕ are the spherical coordinates, P_l^m is the associated Legendre polynomial, and l and m are the degree and order of the harmonic, respectively.

$\mathcal{T}^s$ stores the spherical harmonics coefficients of the radiances. To query the radiance of a voxel from a given direction, one multiplies the coefficients from the tensor with the spherical harmonics for the querying direction. In other words, the surface reflection pattern is modeled by spherical harmonics along with its coefficients. Since spherical harmonics are used in a linear system, we find that they can be inverted for a given reflection pattern to predict the coefficients. This is particularly helpful for initializing the radiance tensor when the voxel reflection patterns are recalculated via multipath tracing (detailed in Section 3.2.2).

3.2 RFCanvas Scene Reconstruction

RFCanvas initializes the scene using visual priors to reduce the required RF measurements. Cameras can provide rich knowledge about the environment in terms of geometry, distance, and materials. RFCanvas does not require precise visual scanning of each object by moving the camera in the scene and synthesizing images from multiple views. Instead, RFCanvas only relies on a few static visual sensors (*e.g.*, from several distributed surveillance cameras and IoT cameras) to reconstruct partial scene representations, which are then progressively refined. This makes RFCanvas more adaptable and efficient.

3.2.1 Visual Knowledge Extraction from Cameras. RFCanvas extracts two types of information from camera images, *depth* and *material* of objects' surfaces, and uses them as visual priors to *initialize* the RF scene. By default, neither type of information is provided by common monocular cameras. Fortunately, recent advances in for computer vision [1, 52] have enabled the extraction of them from camera images. With the depth information, *i.e.*, the distance of each pixel to the camera, RFCanvas can estimate the shapes and locations of objects in the scene, which are used to initialize SDF [47]. Meanwhile, objects' materials can be classified by analyzing visual cues such as texture, color, and reflectance recorded in images [14, 41]. This information is then used to initialize the attenuation tensors of RFCanvas scene representations. Despite not being perfect, the initialization with visual priors helps reduce the search space of scene parameters, leading to more accurate simulations and faster optimization in the refinement process.

In practice, given multiple cameras, RFCanvas needs their global poses to fuse their data to reconstruct the scene. We use an automatic progress to determine the pose based on shared spatial features [22]. The calibration for the poses of static cameras only needs to be conducted once.

3.2.2 RF Field Reconstruction from Visual Knowledge. With the visual knowledge, RFCanvas initializes three elements of the scene representations, *i.e.*, SDF, attenuation tensor, and radiance tensor.

To initialize SDF, we first transform the depth image into 3D space using a calibrated camera pose and the camera's intrinsic matrix. The point clouds are then mapped to voxels. The SDF is calculated by checking each voxel within the space to determine the distance to the nearest surface, completing the SDF initialization. While occluded regions may exist, they are acceptable as the SDF will be further refined with RF measurements.

The attenuation tensor of RFCanvas tensorial fields is physically related to the properties of surface materials. With the material types extracted from camera images, RFCanvas can query a material database to obtain the materials' average dielectric constants and conductivity coefficients from literature [20, 39]. For unrecognized materials, RFCanvas uses the material properties of neighboring voxels for initialization.

The radiance tensor captures the multipath propagation of RF signals. Each voxel is considered as an effective emitter that retransmits the signal it receives in desired directions. Unlike neural-based methods where radiance fields are randomly initialized, RFCanvas uses a physically-based initialization through multipath tracing. As

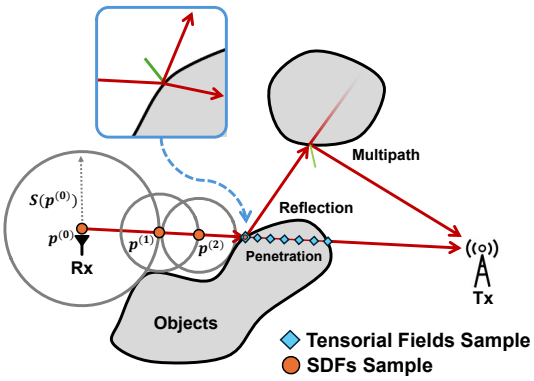

Figure 5: RFCanvas Multipath Exploration for Radiance initialization and adaptation.

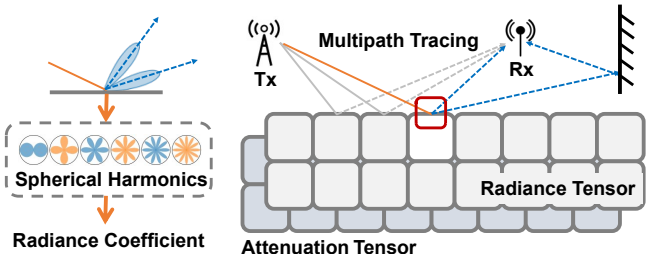

Figure 6: RFCanvas initialize the radiance tensor using reflection patterns obtained from multipath tracing by inverting spherical harmonics.

shown in Fig. 5, RFCanvas explicitly calculates the dominant multipath components with less than three reflections. Beyond three reflections, RF signals become very weak and thus negligible. RFCanvas uses a modified ray tracing algorithm to calculate all paths received and reflected by each voxel, creating a reflection pattern. Specifically, view-dependent reflection is calculated by multiplying the radiance coefficient by the spherical harmonics. RFCanvas inversely calculates and assigns the radiance coefficient by dividing the path-traced reflection pattern by the spherical harmonics, as illustrated in Fig 6.

Although visual priors only provide a rough initialization, they provide a fairly good initialization of the scene. Compared with random initialization, the initialization with visual priors is much closer to the realistic physical scene, making RFCanvas easier to refine progressively using few-shot RF measurements.

3.2.3 Refinement with real RF measurements. The scene representations created from visual priors still have sim-to-real gaps due to visual inaccuracy and occlusions. Therefore, RFCanvas further finetunes the scene representations using few-shot RF measurements from real users. Compared to conventional ray tracing with 3D meshes, RFCanvas is end-to-end trainable, thanks to the differentiable components, including SDF, attenuation, and radiance tensors. Additionally, unlike neural-based methods, RFCanvas 's explicit representation and fairly accurate initialization significantly reduce the amount of RF measurements required for training. With the rapid optimization process, RFCanvas can be progressively updated on the fly, making it more practical for real-world applications.

Refinement procedure. RFCanvas exploits RF measurements from mobile users in the scene. Nowadays, mobile devices are commonly equipped with communication modules and IMUs. We assume users with mobile devices move around freely in the scene. They are the source of major environmental changes in the scene, but simultaneously provide RF measurements, such as RSS and CSI, and their trajectories can be calculated from onboard IMUs. The combination of RF measurements and trajectories is referred to as "RF Spatial Measurements." The paired RF Spatial Measurements can then be used to evaluate the real-time accuracy of RFCanvas, and guide the subsequent optimization of RFCanvas scene representations through backpropagation.

Boundary confidence-guided optimization. To capture the real structure and material properties of the objects and minimize the sim-to-real discrepancy, RFCanvas needs to keep optimizing the attenuation tensor $\mathcal{T}a$, radiance tensor $\mathcal{T}^s$, and SDF S^Ω, *i.e.*,

$$(\mathcal{T}^a, \mathcal{T}^s, S^\Omega)^* = \underset{\mathcal{T}^a, \mathcal{T}^s, S^\Omega}{\arg min} \sum_{i=0}^{M} \mathcal{L}(\hat{s}_i, s_i), \tag{9}$$

where $\hat{s}$ are RF measurements in the scene and M is the number of measurements. $\mathcal{L}$ is the loss that quantifies the discrepancy between real measurements $\hat{s}$ and RFCanvas's RF simulation output s.

Joint optimization of SDF and tensorial fields may lead to suboptimal solutions due to their complexity and non-convexity. To address this issue, we propose a boundary confidence-guided optimization strategy. In surface regions covered by visual priors, we reduce the granularity (i.e., resolution of SDF gradients) and learning rates because these regions have higher boundary confidence. For occluded regions with lower confidence, we enable joint optimization with tensorial fields.

It is expected that a few shots of RF measurements, *i.e.*, a small M, are sufficient for scene refinement since RFCanvas obviates the need to capture expansive empty regions and only focuses on objects with the fairly accurate initialization based on visual priors.

3.3 Dynamics Adaptation

Rapidly adapting to dynamic environments is crucial for responsive RF simulation. Existing neural-based methods [53, 57] treat the entire scene as an implicit model and hence cannot adapt to environmental changes easily. While some recent works [16, 33] generate dynamic neural radiance fields, they are optimized from scratch using a complete sequence of images, but cannot adapt to any new images. In contrast, thanks to the editability of RFCanvas scene representations, RFCanvas supports rapid adaptation to environmental changes.

First, RFCanvas detects and models environmental changes by exploiting visual priors from cameras. To model the changes, RFCanvas estimates the motion vector in image coordinates by matching the spatial features across images, based on the Horn-Schunck Method[4]. After obtaining the screen space motion vector along with depth information, RFCanvas transforms the motion vector to 3D space using the camera's intrinsic matrix K, the 3D position of a pixel (u, v) with depth z can be computed as $p = z \cdot K^{-1}(u, v, 1)^\top$. Applying this transformation to consecutive frames yields the 3D

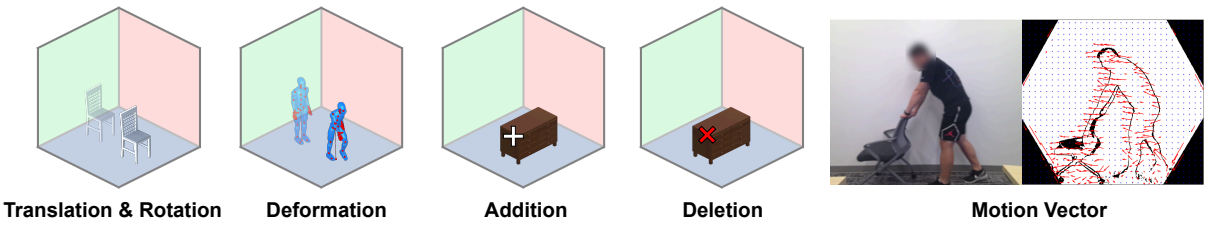

Figure 7: Motion Vectors and Scene Dynamics Adaptations.

positions p_t and p_{t+1}, from which the 3D motion vector is calculated.

After obtaining the motion vectors, we classify the scene changes into four categories, as illustrated in Fig. 7. Each category is handled differently:

(i) *Translation & rotation.* The translation $\mathbf{T} \in \mathbb{R}^3$ and rotation $\mathbf{R} \in \mathbb{R}^{3\times3}$ of dynamic voxel groups are determined either by clustering the motion vectors directly or by using vision-based 3D bounding box systems. We employ ray sample redirection to sample the tensorial field before transformation. Consider an original tensorial field $\mathcal{T}^{t0}$ and a transformed tensorial field $\mathcal{T}^{t1}$. For each query position and direction $P(t,\omega)$ on $\mathcal{T}^{t1}$, we can redirect the sample by the motion's translation and rotation. This is represented as $P'(t,\omega') = p_0'' + \omega' t$, where $p_0'' = \mathbf{R}^T(p_0 - \mathbf{T})$ and $\omega' = \mathbf{R}^T\omega$. Therefore, sampling $P(t,\omega)$ on $\mathcal{T}^{t1}$ is equivalent to sampling $P'(t,\omega')$ on $\mathcal{T}^{t0}$.

(ii) *Deformation.* Deformation involves changes in the shape of the object rather than just its position or orientation. We model deformation as a combination of small translations and rotations applied to individual parts of the object. This can be mathematically represented by a set of local transformations, each described by a translation vector $\mathbf{T_i}$ and a rotation matrix $\mathbf{R_i}$. These local transformations are applied to each voxel in the object, effectively capturing the deformation process. The cumulative effect of these local transformations results in a complex deformation that can be analyzed and visualized in the tensorial field.

(iii) *Addition.* Addition refers to the appearance of new objects or regions within the scene. This is determined by comparing the current voxel frame with the subsequent voxel frame to identify regions that were previously unseen. Since new additions do not have corresponding motion vectors from the previous frame, we identify these regions by detecting spatial features that appear in the subsequent frame but are absent in the previous one.

The manipulation of the tensorial field for addition is similar to the partial initialization from visual priors, as discussed in Section 3.2. Specifically, the SDF and the appearance region of the new voxels are recalculated. The attenuation fields are filled with the reflection coefficients obtained from material identification. The radiance field is initialized using multipath tracing through our custom-designed Ray Tracing simulator, which is specifically optimized for calculating multipath effects.

(d) *Deletion.* Deletion refers to the disappearance of objects or regions from the scene. Similar to addition, deletion is determined by comparing the current voxel frame with the subsequent voxel frame. Regions that were present in the previous frame but are missing in the subsequent frame are identified as deletions. This involves detecting spatial features that vanish and clustering the disappearing voxels to recognize coherent objects or regions that have been removed. By analyzing these deletions, we can understand changes in the scene's structure over time.

The manipulation of the tensorial field for deletion involves recalculating the SDF for the disappearing regions. The attenuation fields and radiance field are reset to zero to reflect the removal of the objects or regions.

4 IMPLEMENTATION

4.1 Software

We implement the RFCanvas framework using PyTorch. However, since PyTorch is designed for data parallelism and model parallelism, it may encounter performance bottlenecks when dealing with complex operations like ray tracing. Hence, we use custom CUDA kernels based on TinyCudaNN and the Slang Shader Language [2] for GPU computation to improve efficiency. The Adam optimizer is used for gradient descent. The training of our system is conducted on a server equipped with two NVIDIA RTX A6000 GPUs.

We train the RFCanvas scene using the Adam optimizer, starting with a learning rate of 1×10^{-3} that decays to 1×10^{-5} over 10,000 iterations. Both the tensorial and SDF fields are trained using a multi-resolution approach. Initially, the resolution for both fields is set to 128. During training, we upsample the vectors and matrices linearly and bilinearly at iterations 2000, 3000, 4000, 5500, and 7000. The upsampling process involves interpolating the number of voxels between the initial resolution and the final resolution in logarithmic space. The final resolution is determined by the size of the scene's bounding box divided by half the wavelength.

For interpreting visual priors, we employ the pre-trained DepthAnythingV2 [52] model to extract depth information. For material identification, we use a modified ViT-based semantic segmentation model [55] trained on Apple's Dense Material Segmentation Dataset [45]. The dielectric constants of the corresponding material are referenced from the Wireless InSite database [37]. The scene dynamics are estimated by monitoring the 3D bounding box of the objects. The human body dynamics is specifically modeled via SMPL model [19, 24].

4.2 Evaluation Setup

Mobile Platform. As shown in Fig. 9, we employ Turtlebot4 to collect ground truth RF scene data. The robot is equipped with integrated Light Detection and Ranging (LiDAR) sensor and a Simultaneous Localization and Mapping (SLAM) system. The robot can navigate the target space by following a predefined trajectory

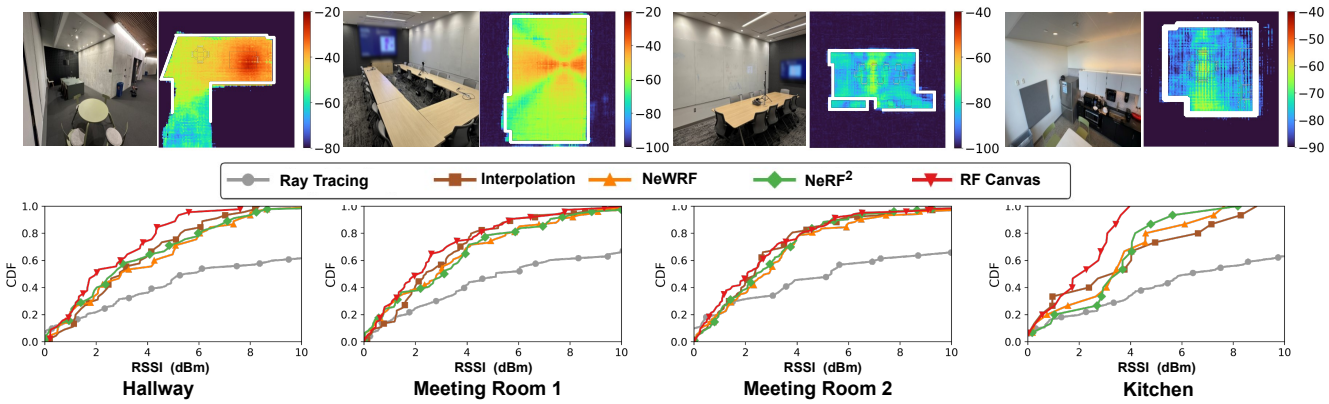

Figure 8: RF Channel Estimation for four indoor scenarios. RFCanvas outperforms all baselines in accuracy.

with designated waypoints, ensuring precise positioning within the floor plan. Furthermore, the adjustable height of wireless devices mounted on the Turtlebot4 facilitates more accurate data collection. Leveraging this platform, we affixed a smartphone and a mmWave router specifically configured for 2.4/5 GHz WiFi and 60 GHz WiGig Received Signal Strength Indicator (RSSI) acquisition.

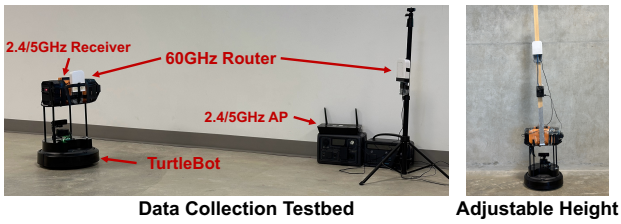

Figure 9: RFCanvas Wireless Channel Collection Setup.

Wireless Devices. For our Wi-Fi setup, we deploy an ASUS RT-AC86U router supporting 802.11ac as our access point (AP), and an iPhone 14 Pro as the client device, functioning as the transmitter and receiver, respectively. The RT-AC86U operates in concurrent dual-band mode (2.4 GHz and 5 GHz). On the client side, we utilize the "Wi-Fi Scan" feature of the AirPort Utility application to record real-time RSSI values for both frequency bands. For the WiGig setup, we employ two 802.11ad-compliant MikroTik wAP 60G×3 routers [3], with one functioning as the transmitter and the other as the receiver. To overcome the limitations of the proprietary operating system, we opt for the open-source OpenWrt operating system in combination with the Mikrotik Researcher Tools [10], a specialized OpenWrt system designed for research purposes, to facilitate efficient RSSI collection. Both setups are illustrated in Fig. 9.

Baselines. We use the following classical and state-of-the-art methods for baseline comparison:

(i) *Ray tracing.* We utilize *AutoMS* [27] for the ray tracing baseline as its implementation supports optimizing surface material reflectance with few-shot RF measurements, which aligns with RFCanvas. The 3D mesh required by ray tracing is scaned via Polycam [34].

(ii) $NeRF^2$. We use $NeRF^2$ [57] as one of the neural methods baselines. $NeRF^2$ represents the state-of-the-art (SOTA) in neural channel prediction. Our implementation for $NeRF^2$ adheres to the default configuration provided in the publicly available code archive. We train the models for 30,000 epochs using the cosine annealing scheduler to avoid overfitting. The reproduced results demonstrate a performance of 2.90 dB on BLE RSSI, 22.24 dB on WiFi SNR, and 0.82 SSIM for the RFID spectrum, which aligns with the original paper.

(iii) *NeWRF.* NeWRF [25] is used as the second neural method baseline, as it is an improved version of $NeRF^2$ tailored for stationary Tx, consistent with our experimental setup.

(iv) *Linear Interpolation.* We employ linear interpolation from the *scipy* package as a pure signal processing baseline.

Vision Prior Granularity. We use two metrics to qualify the granularity of vision priors required by RFCanvas.

i). Camera Coverage Quantification. As different cameras in the scene come in different models and usually with different resolutions and fields of view (FoV), we quantify and unify the contribution of camera sensors by employing the Voxelization method (when 3D ground truth is available), which divides the 3D space into voxels and calculates the ratio of voxels observed by cameras to the total number of voxels in the scene.

ii). Scene Dynamics Quantification. The measurements of scene changes involve the number of voxels (volumes) of changes Δv and the distance (*e.g.*, displacement) of changes Δd. We quantify the measurements as a unified metric $\Delta c = \Delta v \times \Delta d$.

5 EVALUATION

5.1 System Performance Evaluation

5.1.1 RF Channel Modeling Accuracy. As illustrated in Fig. 8, we evaluate RFCanvas in four real-world indoor scenes comprising multiple types of objects and materials. For each scene, we collected 2.4 GHz, 5 GHz WiFi, and 60 GHz WiGig RSSI using the mobile platform. For *meeting rooms*, two images are captured from ceiling-mounted cameras angled downward. For *hallways* and *kitchens*, two panoramic images are taken from the center of each region. A total of roughly 200 RF samples are collected for each scene under each frequency band, of which 80% of the samples are used for

reconstructing RFCanvas and the baselines scene. The CDFs for each room are shown in Fig. 8. The median errors of RFCanvas and baselines across different frequency bands are shown in Fig. 13. RFCanvas outperforms all baselines in all scenarios and frequency bands, with an overall median error of 2.2 dB. The results confirm that RFCanvas is capable of modeling indoor RF channels accurately using few-shot measurements and visual priors.

5.1.2 Impact of RF Samples Density. We evaluate RFCanvas and baselines in optimization performance with respect to RF sampling density. The results are shown in Fig. 11. Ray tracing precisely predicts the channel structure; however, it only achieves a 13 dB error without real RF measurements and converges to 7 dB when refining solely the surface material, due to its inability to optimize internal and complex propagation effects. The neural methods reached a 3.5 dB error in overall numerical accuracy; however, they struggle to capture the channel structure adequately because they do not represent and define the boundaries of objects. We further augment the training data for neural methods with RFCanvas simulated results. However, they remain unable to capture the scene structure even at high sampling density. Surprisingly, linear interpolation demonstrates better structural characterization of the scene compared to neural methods, achieving median errors of 3.75 dB. This is because most of the propagation simply follows the FSPL, which is linear in logarithmic space. We found linear interpolation fails in regions heavily affected by occlusion, multipath, or a lack of sufficient training samples. In comparison, RFCanvas can characterize the scene with extremely sparse samples (i.e., 0.05 sp/ft^2) and achieves high numerical accuracy with a 1.92 dB median error due to the initialization from visual priors while maintaining superior structural prediction due to the clear definitions of object boundaries.

Figure 10: The camera setup for obtaining visual priors.

5.1.3 Impact of Scene Dynamics. We evaluate RFCanvas's performance in adapting to scene dynamics. RFCanvas is initialized on the original scene and subsequently tested in three scenarios, each incorporating varying degrees of transition, rotation, deformation, addition, and deletion. The extent of dynamic changes increases across the three cases as shown in Fig. 12. The results are shown in Fig. 14. RFCanvas accurately predicted scenes (D1) and (D2) with a median error of 2.2 and 2.2 dB, respectively, compared to the baseline (without motion vector) error of 3.5 dB. In scene (D3), which exhibits the most significant scene changes, RFCanvas achieves a median error of 3.2 dB while the baseline demonstrates a 5.7 dB median error. We further refine RFCanvas using five additional real-world measurements around the changed object. The refinement reduces the error to a mere 1.9 dB.

5.2 Generalization Across Different Visual Priors

5.2.1 Impact of Camera Coverage and NLOS Scenarios. We evaluate RFCanvas's performance across varying levels of camera coverage, including scenarios with NLOS conditions due to occlusions. The setup, as shown in Fig. 10, consists of two cameras with 100-degree fields of view (FOV) and one panoramic camera at the room's center. The calculated camera coverage for a single corner camera, two cameras, and the panoramic camera are approximately 30%, 60%, and 90% of the total area, respectively. We also include a scenario with 0% scene coverage (i.e., no cameras) for comparison. The results are presented in Fig. 15. RFCanvas demonstrates robust performance even in NLOS scenarios, achieving a median error of 3.2 dB. In regions with camera coverage, the visual priors significantly enhance optimization and convergence. In occluded regions, RFCanvas requires additional samples and processing time to reduce uncertainty; however, it ultimately achieves satisfactory results with a median error of 2.0 dB.

5.2.2 Visual Prior Sensitivity Analysis. While SOTA computer vision techniques can effectively extract visual information, the precision of object localization and boundary detection may be subject to errors. To evaluate the robustness and sensitivity of RFCanvas in the presence of visual prior inaccuracies, our experiment involved systematically introducing random offsets to object boundaries within the hallway scene. As shown in Fig. 16, RFCanvas demonstrates resilience to moderate levels of visual prior errors, maintaining satisfactory performance with a median error of 2.1 to 3.5 dB when the geometric discrepancies remain below 10 cm. Notably, the system's utilization of the real RF measurements facilitates automatic geometrical error calibration.

5.3 System Consumption

RFCanvas employs vector-matrix multiplication with a computational complexity of $O(N^2)$, where N is the tensorial fields' resolution. This method reduces space complexity from $O(N^3)$ to $O(N^2)$. Neural-based methods using MLPs have a comparable computational complexity of $O(D \times W^2)$, with D as the number of layers and W as the layer dimension. Conventional ray tracing has a computational complexity of $O(U \times V^2)$, optimized to $O(U \times V \log V)$ with accelerated structures, where V is the number of scene triangles and U is the number of rays. Notably, $V > U \gg N \approx W$. In practice, RFCanvas is highly efficient, predicting RF channels in sub-millisecond time and supporting parallelization. Unlike ray tracing, its performance is independent of scene complexity. The simplicity of the vector-matrix operations allows deployment on edge devices without GPU acceleration, making RFCanvas an efficient solution for RF channel prediction across various computing environments.

RFCanvas's scene representation exhibits remarkable efficiency in training. Training a room-sized scene is accomplished in approximately five minutes using a single RTX A6000 GPU. When adapting to scene dynamics, operations such as translation, rotation, and deletion are performed instantaneously. However, for additions and significant scene alterations, such as furniture rearrangements, the process necessitates initiating multipath tracing, which requires approximately 5 seconds.

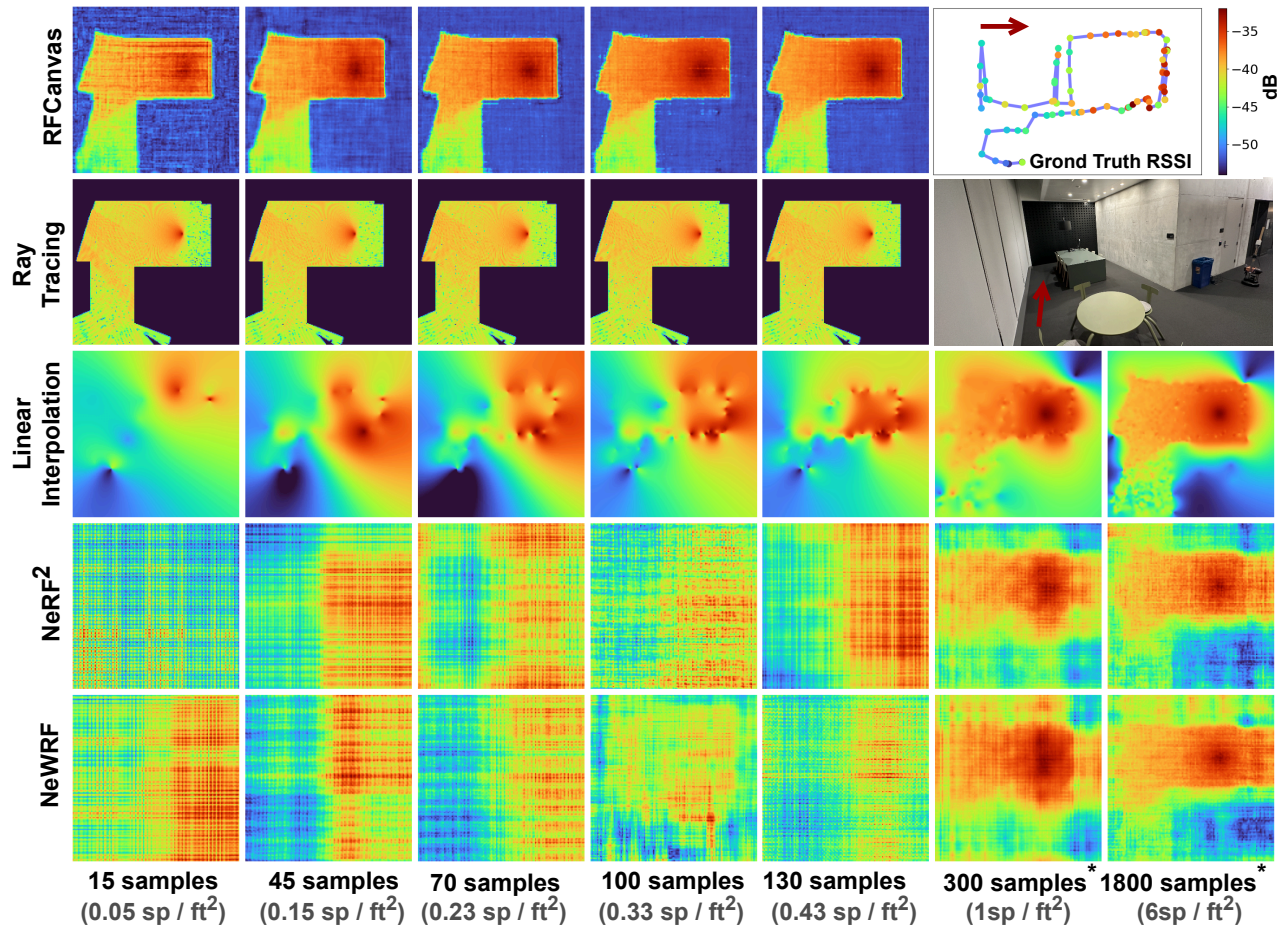

Figure 11: RF channel estimation performance under varying sampling densities. RFCanvas outperforms neural-based methods and ray tracing in both structural and numerical accuracy, especially with sparse samples.

6 CASE STUDY: METASURFACE REDEPLOYMENT

Metasurfaces (MS) have emerged as promising techniques to manipulate RF channels and coverage with fine granularity [7, 21, 27, 36]. Passive metasurfaces are typically optimized through simulations of static environments and fabricated via 3D printing, hot stamping, *etc.* Consequently, they are unable to adapt to scene changes such as furniture rearrangement. Although fabrication costs of new metasurfaces are trending downward, adapting to scene changes inevitably incurs additional expenses.

RFCanvas enables the redeployment of existing metasurfaces to new locations, facilitating adaptation to dynamic scenes without the need for new metasurfaces. We demonstrate this through a case study in which we enable RFCanvas to support metasurface (MS) optimization. This is achieved by representing the MS as a thin quad using SDF with an assigned phase texture. When a ray hits and reflects off the MS, its phase is shifted according to the texture's phase profile, while penetrations are discarded for the MS. We set the experiment in a typical indoor environment as illustrated in Fig. 17. The 60 GHz WiGig router is positioned at the end of the hallway, and the user's receiver is placed on the table. The direct path between the transmitter and the receiver is blocked by the wall. We initially optimize the MS in the scene using conventional methods [27] and fabricate an MS using 3D printing [36]. Then we alter the scene by introducing new obstacles. When a new object is introduced, obstructing the path between the target region and the metasurface, the RSSI decreases by 7 dB. We optimize the new deployment location and orientation of the metasurface by extending existing optimization methods [27] with RFCanvas simulation while maintaining the original metasurface pattern. The optimized new deployment location and rotation yield a 5 dB signal gain.

7 DISCUSSION AND FUTURE WORKS

Real-time channel modeling. RFCanvas is highly efficient for inference and training, taking under one millisecond for RF channel inference and about 20 ms per training iteration. The current implementation achieves a 5 Hz update rate for dynamic scenes, with the main bottleneck being the camera-based depth and motion vector estimation. As a framework for RF simulation, RFCanvas is poised to achieve real-time RF scene adaptation with advancements in computer vision, broadening its applicability to various real-world scenarios.

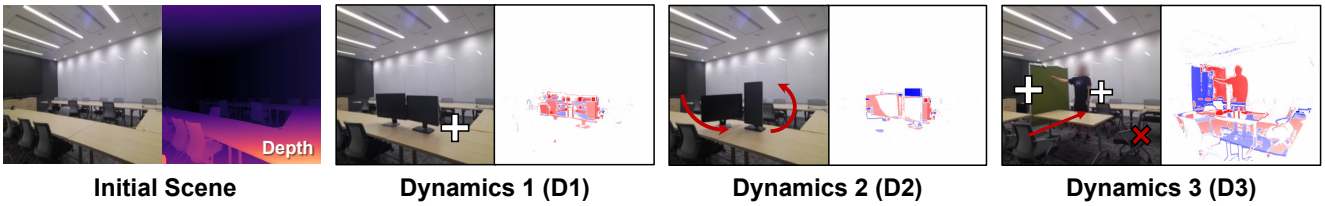

Figure 12: Motion Vectors and Scene Dynamics Adaptations

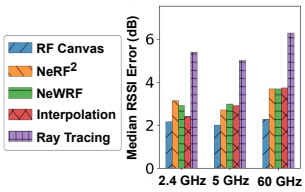

Figure 13: Performance Under Different Frequency Band

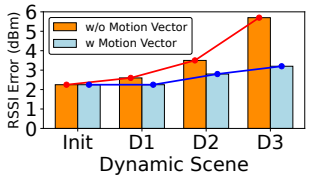

Figure 14: Impact of Scene Dynamics

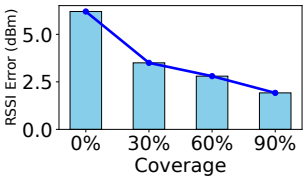

Figure 15: Impact of Camera Coverage

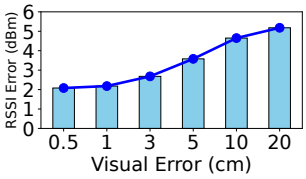

Figure 16: Visual Prior Sensitivity Analysis

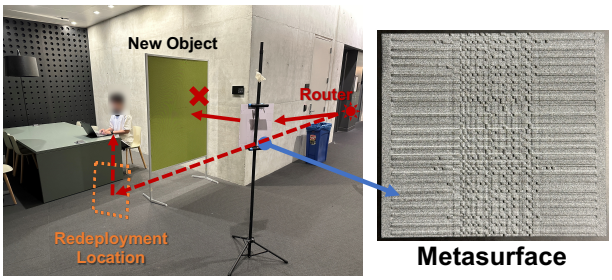

Figure 17: Experiment setup for metasurface redeployment case study.

Complex tasks and dynamic transmitters. Beyond RSSI prediction, RFCanvas supports various complex tasks such as CSI prediction, RFID spectrum prediction, and dynamic transmitters. Following existing methods, high-dimensional inputs and outputs, such as Tx location, CSI subcarriers, and angle spectrum, are managed by combining them with additional MLPs. RFCanvas demonstrated performance comparable to NeRF^2 (24 dB SNR for CSI and 0.8 SSIM for RF spectrum prediction) on its public dataset.

Outdoor environment. RFCanvas can be applied to large outdoor environments, where visual priors can be obtained from widely available street view or satellite images. In these environments, objects such as buildings and terrains are often sparsely distributed across vast areas. As a result, tensorial fields can be optimized using hierarchical acceleration structures like octrees [9] or by partitioning the space into chunks for improved efficiency. Furthermore, RFCanvas can be integrated with existing neural methods to implicitly characterize objects with highly complex shapes, such as vegetation, which are commonly found in outdoor settings.

Visual and RF disparities. RFCanvas is evaluated on common indoor materials such as concrete, wood, plastic, and metal and under the assumption that the visual and RF properties are coherent. It may suffer incorrect initialization when encountering visually disrupting material such as glass (transparent) and highly specular reflection such as mirror and unseen material from the material library. A potential solution is to incorporate a more advanced graphical 3D estimation model for initialization [54] and increase the weight on material optimized from RF signal rather than solely relying on visual material identification [30].

Beyond novel view synthesis for RF. Recent advancements, such as 3DGS [18], can also be used to achieve similar tasks as RFCanvas, such as creating a visual 3DGS from cameras and then optimizing additional spherical harmonics coefficients for RF. However, we believe it is important to go beyond simply adopting Novel View Synthesis methods from computer graphics, where scene representations are specifically designed for only generating new sensor views. We believe that to build a true digital twin for RF, the scene geometries and materials must be explicitly defined and physically grounded, rather than relying solely on neural and probabilistic models. RFCanvas represents a step toward explicit scene representations for next-generation RF digital twin simulation.

8 CONCLUSION

We have presented RFCanvas, a novel system for generating neural-free and explicit RF representations, ensuring accurate and flexible RF simulations. RFCanvas surpasses conventional ray tracing and neural scene models by utilizing visual priors and refining with few-shot RF samples. Featuring RFCanvas Scene based on tensorial fields, it offers editability and precision in modeling complex geometries and materials. Its adaptability to scene dynamics via motion vectors and its end-to-end optimization framework with differentiable simulation provide notable advantages. We envision RFCanvas as a powerful tool for wireless applications like network planning and 3D reconstruction, with further exploration left for future work.

ACKNOWLEDGMENTS

We appreciate the insightful comments and feedback from the anonymous reviewers and shepherd. This work is partially supported by the NSF under Grants CNS-2128588, CNS-2312715, and CNS-2408393.

REFERENCES

[1] Ibraheem Alhashim and Peter Wonka. 2018. High quality monocular depth estimation via transfer learning. *arXiv preprint arXiv:1812.11941* (2018).

[2] Sai Praveen Bangaru, Lifan Wu, Tzu-Mao Li, Jacob Munkberg, Gilbert Bernstein, Jonathan Ragan-Kelley, Frédo Durand, Aaron Lefohn, and Yong He. 2023. Slang. d: Fast, modular and differentiable shader programming. *ACM Transactions on Graphics (TOG)* 42, 6 (2023), 1–28.

[3] Alejandro Blanco, Pablo Jiménez Mateo, Francesco Gringoli, and Joerg Widmer. 2022. Augmenting mmWave localization accuracy through sub-6 GHz on off-the-shelf devices. In *Proceedings of the ACM MobiSys*.

[4] Andrés Bruhn, Joachim Weickert, and Christoph Schnörr. 2005. Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods. *International journal of computer vision* 61 (2005), 211–231.

[5] Anpei Chen, Zexiang Xu, Andreas Geiger, Jingyi Yu, and Hao Su. 2022. Tensorf: Tensorial radiance fields. In *European conference on computer vision*. Springer, 333–350.

[6] Xingyu Chen and Xinyu Zhang. 2023. RF Genesis: Zero-shot generalization of mmwave sensing through simulation-based data synthesis and generative diffusion models. In *Proceedings of the ACM SenSys*.

[7] Yongjian Fu, Yongzhao Zhang, Yu Lu, Lili Qiu, Yi-Chao Chen, Yezhou Wang, Mei Wang, Yijie Li, Ju Ren, and Yaoxue Zhang. 2024. Adaptive Metasurface-Based Acoustic Imaging using Joint Optimization. In *Proceedings of the 22nd Annual International Conference on Mobile Systems, Applications and Services*. 492–504.

[8] Robin Green. 2003. Spherical harmonic lighting: The gritty details. In *Archives of the game developers conference*, Vol. 56. 4.

[9] Sören Grimm, Stefan Bruckner, Armin Kanitsar, and Eduard Groller. 2004. Memory efficient acceleration structures and techniques for CPU-based volume raycasting of large data. In *2004 IEEE Symposium on Volume Visualization and Graphics*. IEEE, 1–8.

[10] IMDEA Networks Wireless Networking Group. 2024. Researcher version of Openwrt for wAP 60G/LHGG-60ad/wAP 60Gx3 AP. https://github.com/IMDEANetworksWNG/Mikrotik-researcher-tools

[11] Gaël Guennebaud and Markus Gross. 2007. Algebraic point set surfaces. *ACM Transactions on Graphics (TOG)* 26, 3 (2007), 23.

[12] John C Hart. 1996. Sphere tracing: A geometric method for the antialiased ray tracing of implicit surfaces. *The Visual Computer* 12, 10 (1996), 527–545.

[13] John C Hart. 1996. Sphere tracing: A geometric method for the antialiased ray tracing of implicit surfaces. *The Visual Computer* 12, 10 (1996), 527–545.

[14] Eric Hayman, Barbara Caputo, Mario Fritz, and Jan-Olof Eklundh. 2004. On the significance of real-world conditions for material classification. In *Computer Vision-ECCV 2004: 8th European Conference on Computer Vision, Prague, Czech Republic, May 11-14, 2004. Proceedings, Part IV 8*. Springer, 253–266.

[15] Jingzhi Hu, Tianyue Zheng, Zhe Chen, Hongbo Wang, and Jun Luo. 2023. Muse-fi: Contactless muti-person sensing exploiting near-field wi-fi channel variation. In *Proceedings of the ACM MobiCom*.

[16] Hankyu Jang and Daeyoung Kim. 2022. D-tensorf: Tensorial radiance fields for dynamic scenes. *arXiv preprint arXiv:2212.02375* (2022).

[17] Jian-Ming Jin. 2015. *The finite element method in electromagnetics*. John Wiley & Sons.

[18] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 2023. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Transactions on Graphics* 42, 4 (July 2023).

[19] Muhammed Kocabas, Nikos Athanasiou, and Michael J Black. 2020. Vibe: Video inference for human body pose and shape estimation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 5253–5263.

[20] B. Langen, G. Lober, and W. Herzig. 1994. Reflection and transmission behaviour of building materials at 60 GHz. In *5th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, Wireless Networks - Catching the Mobile Future.*, Vol. 2. 505–509 vol.2. https://doi.org/10.1109/WNCMF.1994.529141

[21] Xinyi Li, Gaoteng Zhao, Ling Chen, Xinyu Zhang, and Ju Ren. 2024. RFMagus: Programming the Radio Environment With Networked Metasurfaces. In *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*. 16–30.

[22] Tsung-Yi Lin, Peter Raymond Florence, Yen-Chen Lin, and Jonathan Tilton Barron. 2023. Inverting Neural Radiance Fields for Pose Estimation. US Patent App. 18/011,601.

[23] Hao Ling, R-C Chou, and S-W Lee. 1989. Shooting and bouncing rays: Calculating the RCS of an arbitrarily shaped cavity. *IEEE Transactions on Antennas and Propagation* 37, 2 (1989), 194–205.

[24] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. 2015. SMPL: A Skinned Multi-Person Linear Model. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)* 34, 6 (Oct. 2015), 248:1–248:16.

[25] Haofan Lu, Christopher Vattheuer, Baharan Mirzasoleiman, and Omid Abari. 2024. A Deep Learning Framework for Wireless Radiation Field Reconstruction and Channel Prediction. arXiv:2403.03241

[26] Luxcarta. 2024. Mapping your world in 3D. https://www.luxcarta.com/.

[27] Ruichun Ma, Shicheng Zheng, Hao Pan, Lili Qiu, Xingyu Chen, Liangyu Liu, Yihong Liu, Wenjun Hu, and Ju Ren. 2024. AutoMS: Automated Service of mmWave Coverage Optimization using Low-cost Metasurfaces. In *Proceedings of the ACM MobiCom*.

[28] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. *Commun. ACM* 65, 1 (2021), 99–106.

[29] Claus Müller. 2006. *Spherical harmonics*. Vol. 17. Springer.

[30] Tribhuvanesh Orekondy, Pratik Kumar, Shreya Kadambi, Hao Ye, Joseph Soriaga, and Arash Behboodi. 2022. Winert: Towards neural ray tracing for wireless channel modelling and differentiable simulations. In *Proceedings of the ICLR*.

[31] Stanley Osher and James A Sethian. 1988. Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations. *Journal of computational physics* 79, 1 (1988), 12–49.

[32] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. 2019. Deepsdf: Learning continuous signed distance functions for shape representation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 165–174.

[33] Keunhong Park, Utkarsh Sinha, Jonathan T Barron, Sofien Bouaziz, Dan B Goldman, Steven M Seitz, and Ricardo Martin-Brualla. 2021. Nerfies: Deformable neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 5865–5874.

[34] polycam. 2023. Polycam - LiDAR & 3D Scanner for iPhone & Android. https://poly.cam/.

[35] Telecom Infra Project. 2023. Open AFC. https://telecominfraproject.com/open-afc/.

[36] Kun Qian, Lulu Yao, Xinyu Zhang, and Tse Nga Ng. 2022. MilliMirror: 3D printed reflecting surface for millimeter-wave coverage expansion. In *Proceedings of the 28th Annual International Conference on Mobile Computing And Networking*. 15–28.

[37] Remcom. 2024. *Wireless InSite*. Remcom Inc., State College, PA. https://www.remcom.com/wireless-insite-em-propagation-software Version X.X (replace with the version you used).

[38] Remcom. 2024. Wireless Insite 3D wireless propagation software. https://www.remcom.com/wireless-insite-propagation-software.

[39] K. Sato, T. Manabe, T. Ihara, H. Saito, S. Ito, T. Tanaka, K. Sugai, N. Ohmi, Y. Murakami, M. Shibayama, Y. Konishi, and T. Kimura. 1997. Measurements of reflection and transmission characteristics of interior structures of office building in the 60-GHz band. *IEEE Transactions on Antennas and Propagation* 45, 12 (1997), 1783–1792. https://doi.org/10.1109/8.650196

[40] Dario Seyb, Eugene d'Eon, Benedikt Bitterli, and Wojciech Jarosz. 2024. From microfacets to participating media: A unified theory of light transport with stochastic geometry. *ACM Transactions on Graphics (Proceedings of SIGGRAPH)* 43, 4 (July 2024). https://doi.org/10.1145/3658121

[41] Shuochen Su, Felix Heide, Robin Swanson, Jonathan Klein, Clara Callenberg, Matthias Hullin, and Wolfgang Heidrich. 2016. Material classification using raw time-of-flight measurements. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3503–3511.

[42] Allen Taflove, Susan C Hagness, and Melinda Piket-May. 2005. Computational electromagnetics: the finite-difference time-domain method. *The Electrical Engineering Handbook* 3, 629-670 (2005), 15.

[43] Blare Tech. 2024. The challenge of 5G network planning.

[44] Paolo Testolina, Michele Polese, and Tommaso Melodia. 2024. Sharing Spectrum and Services in the 7-24 GHz Upper Midband. *arXiv preprint arXiv:2402.08649* (2024).

[45] Paul Upchurch and Ransen Niu. 2022. A dense material segmentation dataset for indoor and outdoor scene parsing. In *European conference on computer vision*. Springer, 450–466.

[46] Stavros Vakalis, Liang Gong, and Jeffrey A Nanzer. 2019. Imaging with wifi. *IEEE Access* 7 (2019), 28616–28624.

[47] Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. 2021. Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. *arXiv preprint arXiv:2106.10689* (2021).

[48] Xin Wang, Shu Hu, Heng Fan, Hongtu Zhu, and Xin Li. 2024. Neural Radiance Fields in Medical Imaging: Challenges and Next Steps. *arXiv preprint arXiv:2402.17797* (2024).

[49] Teng Wei, Anfu Zhou, and Xinyu Zhang. 2017. Facilitating Robust 60 GHz Network Deployment By Sensing Ambient Reflectors. In *USENIX Symposium on Networked Systems Design and Implementation (NSDI)*.

[50] Timothy Woodford, Xinyu Zhang, Eugene Chai, and Karthikeyan Sundaresan. 2022. Mosaic: Leveraging diverse reflector geometries for omnidirectional around-corner automotive radar. In *Proceedings of the 20th Annual International Conference on Mobile Systems, Applications and Services (MobiSys)*.

[51] Timothy Woodford, Xinyu Zhang, Eugene Chai, Karthikeyan Sundaresan, and Amir Khojastepour. 2021. Spacebeam: Lidar-driven one-shot mmwave beam management. In *Proceedings of the ACM Annual International Conference on Mobile Systems, Applications and Services (MobiSys)*.

[52] Lihe Yang, Bingyi Kang, Zilong Huang, Zhen Zhao, Xiaogang Xu, Jiashi Feng, and Hengshuang Zhao. 2024. Depth Anything V2. *arXiv:2406.09414* (2024).
[53] Yu-Jie Yuan, Yang-Tian Sun, Yu-Kun Lai, Yuewen Ma, Rongfei Jia, and Lin Gao. 2022. Nerf-editing: geometry editing of neural radiance fields. In *Proceedings of the IEEE/CVF CVPR*.
[54] Junyi Zeng, Chong Bao, Rui Chen, Zilong Dong, Guofeng Zhang, Hujun Bao, and Zhaopeng Cui. 2023. Mirror-NeRF: Learning Neural Radiance Fields for Mirrors with Whitted-Style Ray Tracing. In *Proceedings of the 31st ACM International Conference on Multimedia*. 4606–4615.
[55] Bowen Zhang, Zhi Tian, Quan Tang, Xiangxiang Chu, Xiaolin Wei, Chunhua Shen, et al. 2022. Segvit: Semantic segmentation with plain vision transformers. *Advances in Neural Information Processing Systems* 35 (2022), 4971–4982.
[56] Hanying Zhao, Mingtao Huang, and Yuan Shen. 2022. High-accuracy localization in multipath environments via spatio-temporal feature tensorization. *IEEE Transactions on Wireless Communications* 21, 12 (2022), 10576–10591.
[57] Xiaopeng Zhao, Zhenlin An, Qingrui Pan, and Lei Yang. 2023. NeRF2: Neural Radio-Frequency Radiance Fields. In *Proceedings of the ACM MobiCom*.