



Crowdsourced Geospatial Intelligence: Constructing 3D Urban Maps with Satellitic Radiance Fields

XIAOPENG ZHAO*, The Hong Kong Polytechnic University, Hong Kong

SHEN WANG*, The Hong Kong Polytechnic University, Hong Kong

ZHENLIN AN, The Hong Kong Polytechnic University, Hong Kong

LEI YANG†, The Hong Kong Polytechnic University, Hong Kong

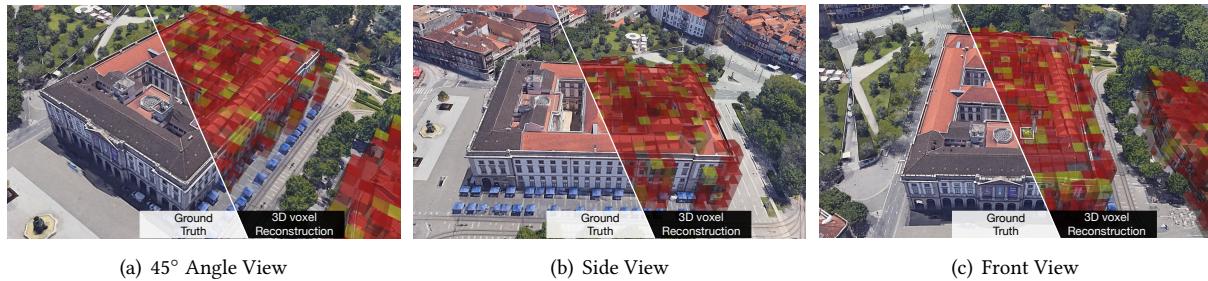


Fig. 1. Illustration of SaRF. Our framework effectively reconstructs 3D urban maps using crowdsourced GNSS data, primarily through a methodical voxelization process. In the figures, medium-sized voxels are used to reduce the overall image size.

In urban planning and research, 3D city maps are crucial for activities such as cellular network design, urban development, and climate research. Traditionally, creating these models has involved costly techniques like manual 3D mapping, interpretation of satellite or aerial images, or the use of sophisticated depth-sensing equipment. In this work, we propose a novel approach to develop 3D urban maps by examining the influence of urban structures on satellite signals, using GPS records crowdsourced from hundreds of smartphones during everyday user movements. We introduce the concept of satellitic radiance fields (SaRF), a novel neural scene representation technique designed to capture the distribution of GPS signals in urban settings. SaRF employs a sparse voxel octree framework to depict voxel-centric implicit fields, capturing physical properties like the density of each voxel. This model is progressively refined using a differentiable ray-marching process, ultimately leading to the reconstruction of 3D urban maps. Our thorough experimental evaluation, which incorporates approximately 27.4 million GPS records, reveals an average reconstruction accuracy of 83.1% in six varied urban scenes.

*Both authors contributed equally to this research.

†This is the corresponding author.

Authors' Contact Information: Xiaopeng Zhao, zhao@tagsys.org, The Hong Kong Polytechnic University, Department of Computing, Hong Kong; Shen Wang, The Hong Kong Polytechnic University, Department of Computing, Hong Kong, wangshen@tagsys.org; Zhenlin An, The Hong Kong Polytechnic University, Department of Computing, Hong Kong, an@tagsys.org; Lei Yang, The Hong Kong Polytechnic University, Department of Computing, Hong Kong, young@tagsys.org.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM 2474-9567/2024/9-ART146

<https://doi.org/10.1145/3678572>

CCS Concepts: • Human-centered computing → Ubiquitous and mobile computing; • Information systems → Geographic information systems.

Additional Key Words and Phrases: Crowdsensing, GNSS Measurement, 3D Urban Mapping

ACM Reference Format:

Xiaopeng Zhao, Shen Wang, Zhenlin An, and Lei Yang. 2024. Crowdsourced Geospatial Intelligence: Constructing 3D Urban Maps with Satellite Radiance Fields. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 8, 3, Article 146 (September 2024), 24 pages. <https://doi.org/10.1145/3678572>

1 Introduction

3D urban mapping is driven by the need for detailed and accurate representations of urban environments for many reasons, such as better urban planning [19], emergency response [27], environmental monitoring [6], vehicle navigation [17], virtual and augmented reality [10], cultural heritage preservation [45], and so on. Specifically, a comprehensive understanding of the urban landscape is vital for the widespread adoption and success of autonomous vehicles in enabling them to plan routes efficiently, avoid obstacles, and interact with other road users. By creating highly accurate digital representations of real-world environments, these maps allow users enjoying immersive virtual and augmented reality to explore and interact with urban settings in a realistic manner. 3D urban mapping also benefits drone delivery systems by enabling optimized routing, obstacle avoidance, and precise navigation, ensuring safe and efficient transportation of goods in complex urban environments.

Advanced 3D mapping technologies include several key methods: (1) LiDAR, using laser pulses to create detailed point clouds [14, 39]; (2) Photogrammetry, generating 3D data from overlapping 2D images; (3) Synthetic Aperture Radar (SAR) for high-resolution 3D mapping in challenging conditions; (4) Structure from Motion (SfM), reconstructing 3D structures from 2D drone images; and (5) SLAM, used in robotics and autonomous vehicles for real-time mapping and localization. These technologies contribute significantly to precise, high-quality mapping, enhancing our understanding of complex environments. However, they often focus on major structures and require lengthy intervals for updates, with limited access for external use. There remains a gap for cost-effective, regularly updated urban 3D modeling techniques with acceptable accuracy.

Recent studies have delved into the use of crowdsourced geospatial intelligence, leveraging GNSS data casually collected from mobile devices to identify buildings [23] and even construct 3D models of cities [26, 40, 51]. Although these methods might not match the accuracy of traditional techniques, they offer the advantages of reduced costs and broader area coverage. The underlying concept is that urban structures like buildings can block, diminish, or reflect GNSS signals. By examining the fluctuations in signal strength and the visibility of satellites, it becomes feasible to deduce the presence and bounding boxes of buildings or other urban elements. This method is economical since it does not necessitate exclusive hardware or specialized data-gathering campaigns; it instead utilizes GNSS data generated incidentally by individuals in their everyday routines. The extensive data coverage is attributed to the vast number of people using GNSS-equipped devices, which facilitates widespread data acquisition across large geographical regions and yields insights into various urban landscapes.

We expand upon the idea of crowdsourced geospatial intelligence by integrating NeRF (Neural Radiance Fields) into the process of 3D urban mapping [32]. NeRF is a pioneering deep learning approach that generates realistic 3D scenes from multiple 2D images. The model is trained to represent a scene as a continuous function, mapping 3D coordinates and viewing directions to colors and densities. Once trained, NeRF can create new views of the scene, consistent with the input images. Recognized for rendering detailed 3D scenes with few inputs, NeRF has led to significant advancements in 3D scene representation [3, 32, 34, 44, 48]. Recently, the adaptation of NeRF to work with electromagnetic data (NeRF² [59]) has opened opportunities to use crowdsourced GNSS data for fine-grained 3D building reconstruction.

In this work, we introduce the Satellite Radiance Fields (SaRF), a deep learning method for generating 3D urban maps using GPS data sourced through crowdsourcing. Similar to NeRF², SaRF compiles GPS measurements from diverse locations. It employs a sparse voxel octree structure to effectively capture voxel-based implicit fields, highlighting physical characteristics like voxel density. The model is progressively refined via a differentiable ray-marching process, resulting in detailed urban topography maps that accurately reflect variations in voxel densities. This sophisticated modeling and reconstruction approach, facilitated by SaRF, opens up new possibilities in urban planning, simulation, and analysis. Fig. 1 presents an example of voxelization applied to a library building, viewed from various angles.

Challenges. Translating the above idea into a practical system presents numerous challenges.

- The first challenge lies in the need for a comprehensive collection of GPS data from multiple angles to accurately depict a scene, a notably demanding task. To tackle this, we developed an Android app and collaborated with numerous volunteers who collected 617,286 GPS measurements across a specific scene within a year. Furthermore, we leveraged publicly available GPS data from the SenseMyCity project [37, 40], which provided extensive data from five different scenes. This dataset includes over 27.4 million GPS records, gathered by 900 unique users across a span of five years. Cumulatively, we have utilized approximately 28 million GPS records to evaluate our model, thus enriching our analysis with a diverse set of data inputs.
- Second, traditional approaches often rely on line-of-sight (LOS) attenuation models, like those in [26, 40, 51], to infer the presence of buildings along the LOS path. While straightforward, this method falls short in efficiency as it overlooks the impact of reflections from non-line-of-sight (NLOS) propagations. These reflections could be instrumental in deducing the layout of adjacent buildings and in refining the LOS propagation by distinguishing it from other NLOS paths. To address this gap, we introduce SaRF, a model designed to trace potential signal propagations from a comprehensive range of directions. Consequently, every kind of propagation captured by a GPS receiver is effectively utilized in the scene representation, ensuring a more thorough and accurate mapping process.

- Third, in our method, we adopt voxelization to depict a scene. This technique divides the scene into numerous small cubic units, or voxels, which aids in constructing the 3D urban map. Consider a case where each voxel has a volume of 10 cm³. In a scene spanning 100 m³, this equates to a total of a billion voxels, creating a substantial computational challenge in the ray marching process. To overcome this, we implement a hierarchical data structure, specifically an octree. Within this framework, each larger voxel can be subdivided into eight smaller sub-voxels during each training iteration, as required, until the smallest voxel size is achieved. This approach is known as progressive training. In the final analysis, non-air voxels whose relative density surpasses a set threshold are identified as part of a building, thereby enhancing the precision of the resultant 3D urban map.

Contributions. The key contributions of this work are outlined as follows. Firstly, to the best of our knowledge, we are the first to introduce neural radiance fields to address the challenges of a 3D urban map reconstruction using crowdsourced GPS data. Secondly, we introduce SaRF, a novel framework that transforms the training of radiance fields into a problem akin to conventional global illumination. Thirdly, we implement our proposed methodology to construct 3D maps for six diverse scenes, utilizing tens of millions of GPS records available in the public domain.

2 Related Work

In this section, we outline the relevant background and related work across four key categories:

2.1 Global Navigation Satellite System

The Global Navigation Satellite System (GNSS) is a collective term for satellite-based navigation systems providing accurate global positioning and timing. Key GNSS systems include GPS, GLONASS, Galileo, and BeiDou. Focusing

on GPS, it uses three L-band frequencies, primarily L1 at 1.57542 GHz, as L2 and L3 are less accessible. GPS signals, modulated using BPSK, contain ephemeris data and time stamps. GPS receivers calculate their locations by measuring the time delay of signals from at least four satellites, employing trilateration to determine latitude, longitude, and altitude. Various factors can impact the precision of GPS trilateration. Common sources of error include inaccuracies in satellite clocks, atmospheric delays, multipath errors, suboptimal satellite geometry, receiver noise, and selective availability. Particularly, urban buildings can significantly impact GPS accuracy and reliability in several ways: (1) **Signal Blockage**: High buildings can obstruct the line of sight between the receiver and satellites, particularly in densely built areas, leading to signal loss. (2) **Multipath Errors**: GPS signals reflecting off buildings can take longer or different paths, causing errors where the receiver miscalculates its position. (3) **Signal Attenuation**: Urban structures can attenuate, or weaken, GPS signals before they reach the receiver, which can further degrade positioning accuracy. Consequently, under optimal conditions with clear skies and no enhancements, standard GPS typically achieves a horizontal accuracy of around 5-10 meters [54].

To mitigate these potential errors, mobile platforms such as iOS and Android incorporate various advanced GPS methods. (1) **Assisted GPS**: It uses an internet connection to assist in the acquisition of satellite data [49]. By accessing information from a network server, mobile devices can reduce the time to find and lock onto satellites, especially in urban settings where GPS signals might be obstructed. (2) **Differential GPS**: The system [33] improves accuracy by using a network of fixed, ground-based reference stations to broadcast the difference between the positions indicated by the satellite systems and the known fixed positions. Mobile devices can use it to correct their own GPS data. (3) **Augmentation Systems**: Systems like WAAS (Wide Area Augmentation System) in North America and EGNOS (European Geostationary Navigation Overlay Service) in Europe [50], provide corrections and additional information to improve GNSS performance in terms of accuracy, integrity, and availability. (4) **Multi-GNSS Support**: Modern mobile devices are increasingly capable of receiving signals from multiple GNSS networks (like GLONASS, Galileo, BeiDou) simultaneously [36], increasing the number of satellites available for positioning and thereby enhancing accuracy. (5) **Integration with Other Sensors**: Modern mobile devices often integrate GPS with other sensors like accelerometers, gyroscopes, and barometers to improve location accuracy [58]. This sensor fusion approach helps in providing more accurate positioning, especially where GPS signals are weak or unavailable. The combined use of the above methods can significantly enhance GPS accuracy, often achieving cm-level precision even in urban settings.

2.2 GNSS-based 3D voxel Mapping

SaRF aims to reconstruct the 3D voxel map (i.e., occupancy map), which is a spatial representation technique used primarily in robotics and autonomous vehicle navigation. It involves dividing a three-dimensional space into discrete, uniformly sized cubes known as voxels, each of which can store various types of data about the environment. This method allows for highly detailed and dynamic modeling of physical spaces, facilitating precise object and obstacle detection critical for navigation and decision-making in autonomous systems. For example, companies like Tesla incorporate 3D voxel maps in their self-driving cars [15] to process and interpret vast amounts of sensory data. These maps are instrumental in understanding complex environments, enabling the vehicle to navigate safely by identifying and classifying objects. Key benefits of using 3D voxel maps include improved accuracy in object perception and the ability to perform terrain analysis. This technology not only enhances the safety features of autonomous vehicles but also improves their operational efficiency in diverse driving conditions.

RF signals have demonstrated the capacity to perceive the environment beyond their original applications, extending to areas such as occupancy detection [5, 25, 53], human activity recognition [55, 56], localization [11] and imaging [21]. 3D city maps, crucial for applications ranging from network planning to climate studies [4, 13, 16], have traditionally relied on expensive methods such as manual annotation and depth-sensing technologies [1, 14,

[31, 39]. However, GNSS signals, commonly used for positioning, have transformed urban mapping. They provide a globally accessible, economical option for 3D mapping, enhanced by smartphone-based passive data collection [2, 20]. GNSS-based 3D mapping, employing building signal obstruction and GPS for urban reconstruction, shows significant potential [26, 46]. The integration of crowdsourced data and probabilistic methods using GNSS SNR measurements further demonstrates the adaptability of GNSS for urban modeling [12, 23, 40, 41], mitigating the inaccuracies inherent in crowdsourced data, thereby improving the reliability and precision of the urban models generated. While GNSS mapping holds promise, its accuracy faces challenges from signal strength classification and processing, especially in complex urban settings. Unlike prior methods using SNR-based classifiers for identifying building voxels, SaRF leverages an RF model-informed neural radiance field to discern voxel attenuation properties, thereby improving accuracy.

2.3 Crowdsensing

Crowdsensing has become a popular method for conducting large-scale, diverse data across various domains. It's used to many applications from location and driving information [7, 8, 12, 24], spectrum monitoring with smartphones [35, 42], construct Wi-Fi maps [57]. However, these systems often face challenges related to the quality of data, especially when the collection process is opportunistic and lacks controlled environments [57]. In addressing these challenges, recent research has proposed various methodologies. For instance, Fang et al. explore pervasive vehicular sensing to refine urban map inference, which can potentially enhance the accuracy and reliability of location-based data [17]. Uvlens [9] integrates crowdsourced data with open government datasets to identify urban village boundaries and estimate populations, which showcases the potential of crowdsourcing in urban planning. Additionally, MultiCell develops a model for urban population dynamics using multiple cellphone networks, offering a novel approach to demographic analytics [18]. Wang et al. focus on constructing a cellular signal map, which serves to improve network coverage analysis through mobile crowdsensing [52]. SaRF employs crowdsourced GNSS data from mobile devices to construct a 3D occupancy map.

2.4 Neural Radiance Fields

NeRF has become a focal point in computer vision research, leading to diverse developments. Its effectiveness in 3D modeling and environmental reconstruction is well-documented [28, 38, 47], along with its use in scene relighting and view synthesis [30, 44]. The adaptation of NeRF for RF signals first occurred with NeRF² [59]. SaRF introduces a new application of NeRF², uniquely adapting it to work with GPS signals. Compared with NeRF², SaRF has unique contributions from the three aspects: **(1) Target and Representation:** While NeRF² primarily focuses on signal prediction within room-scale environments, SaRF is designed for reconstructing 3D urban voxel maps. To achieve this, we employ an explicit octree-based representation, enhancing the structural depiction of buildings and enabling finer reconstruction granularity. **(2) Scale and Computational Efficiency:** SaRF is designed for large-scale urban scenes, contrasting with the room-scale orientation of NeRF². To manage the computational demands of ray marching through billions of voxels, SaRF utilizes a progressive training methodology. In this process, each larger voxel is subdivided into eight smaller sub-voxels during successive training iterations, continuing until the desired voxel resolution is attained. **(3) Training Methodology:** Unlike NeRF², SaRF primarily concentrates on the attenuation properties of voxels to distinguish non-air elements, making the training of the attenuation subnetwork crucial for accurately reconstructing building structures. This focus necessitates a specialized training approach for the attenuation subnetwork. We utilize readily available 2D maps to guide this training process. Given the challenges of directly obtaining ground-truth data for attenuation coefficients, we apply a semi-supervised learning strategy. This approach leverages GPS signal samples from smartphones to jointly train the radiance and attenuation subnetworks, facilitating comprehensive learning and precise scene reconstruction.

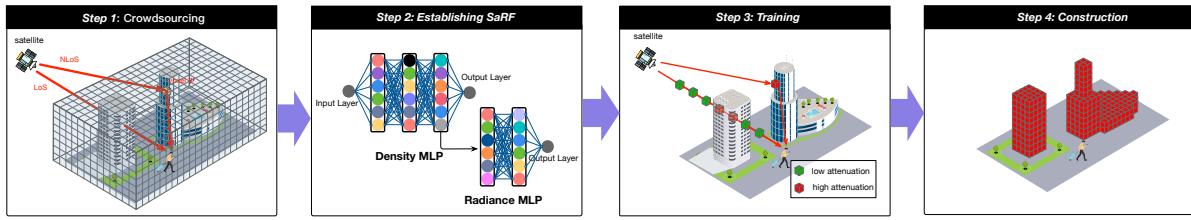


Fig. 2. Approach to Building 3D Urban Maps. This involves (1) amassing extensive raw GPS data, encompassing raw RF signals, positions of satellites, and locations of receivers; (2) creating a SaRF, depicted through two Multilayer Perceptrons (MLPs); (3) training the neural network using ray tracing within a semi-supervised learning framework; and (4) forming the 3D map by eliminating superfluous aerial voxels.

3 Overview

we focus on demonstrating our system using GPS signals, although our dataset includes various types of GNSS data. Modern mobile devices use advanced localization algorithms that integrate previously mentioned GPS techniques, typically achieving an accuracy of less than 1 meter, as commonly experienced in daily usage. Based on this, we adopt a practical assumption:

ASSUMPTION 1. The positions determined by GPS receivers are deemed relatively accurate, despite raw GPS signals being susceptible to disturbances from urban architecture, like signal blockage and multipath errors.

Furthermore, satellite positions, known accurately through ephemeris data, enable the precise establishment of both GPS transmitters (satellites) and receivers (devices) locations, despite potential interference from buildings on raw GPS signals. This observation motivates us to discover building structures using raw GPS signals.

Given that smartphones, commonly equipped with GPS receivers, are widely used, it becomes straightforward to amass substantial data via crowdsourcing. This cost-effective method of data collection encourages more frequent and highly efficient updates of 3D maps. However, our aim is not to supplant existing high-accuracy measurement methods like LiDAR and SAR but to offer an additional methodology for real-time updates. As shown in Fig. 2, the workflow of our approach contains four steps:

- **Step 1 - Data Collection:** Gather comprehensive raw GPS data, including raw RF signals, satellite positions, and receiver locations. This data forms the foundation for accurate 3D mapping, providing insights into spatial relationships.

- **Step 2 - Establishing SaRF:** Develop a SaRF model to interpret GPS data and model the urban environment's physical characteristics. SaRF serves as the basis for transforming GPS signals into a structured representation for a 3D scene.

- **Step 3 - Training:** Train the neural network through ray tracing, a technique that simulates GPS signal propagation in the environment. This semi-supervised approach enhances learning from both labeled and unlabeled data, adapting to diverse urban settings.

- **Step 4 - Reconstruction:** Process and analyze data to create a 3D map by pruning unnecessary aerial voxels. This step focuses the map on critical urban features, resulting in a realistic and functional 3D urban map.

The following subsequent will elaborate on each step.

4 Data Collection

In this section, we provide the methodology and data analysis of our data collection process utilizing crowdsourcing.

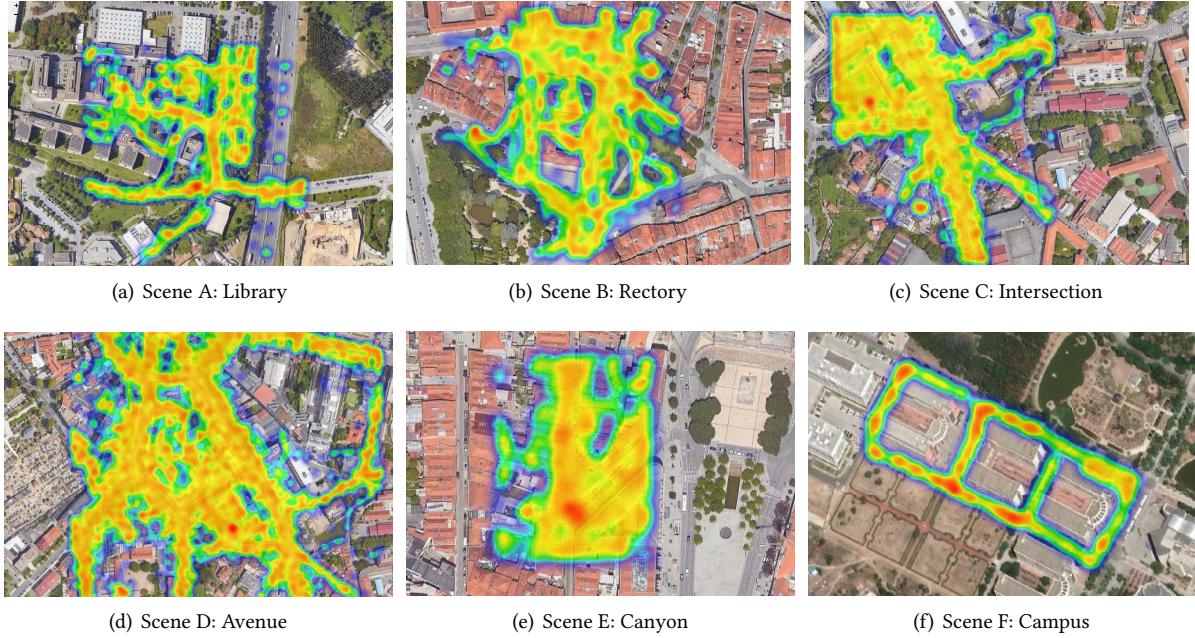


Fig. 3. Spatial distribution of GNSS data across various scenes. (a)-(b) display the GPS data from five different scenes sourced from the SenseMyCity dataset, while (f) illustrates the GPS records we gathered around the architectural complex of a university campus.

4.1 Methodology

Our team developed a crowdsensing application that autonomously collects GNSS data from users' smartphones during their regular activities. This application has successfully gathered 617,286 million geo-tagged data points from participants' daily movements. Ten volunteers participated in the data collection process, traversing the perimeters of three buildings, each approximately 23 meters in height. We use the Android API on smartphones to report the GNSS data. Each data record consists of data following:

$$[\text{Lon}_p, \text{Lat}_p, \text{Alt}_p, \text{Lon}_s, \text{Lat}_s, \text{Alt}_s, \text{SNR}, \theta, \text{PRN}, \text{Timestamp}]$$

where Lon_p , Lat_p , Alt_p , Lon_s , Lat_s , and Alt_s are the longitude, latitude, and altitude of the smartphone and satellite, respectively. SNR and θ denote the received signal-to-noise ratio and carrier phase of the GNSS signal. PRN code is used to identify the ID of the satellites. Additionally, the use of network-based location data from Wi-Fi and cellular sources by Android devices augments the location accuracy, particularly in areas with poor GPS signal reception.

Our study also leverages publicly accessible crowdsourced GPS data from the SenseMyCity [37, 40], which amassed 27.4 million data points through crowdsourcing involving 900 unique users over a period of five years. The data undergoes the pre-processing [40], leading to the absence of phase and timestamp. The SenseMyCity application employs a low-rate, energy-efficient sensing algorithm to detect user movement, enhancing the frequency of data collection specifically during periods of travel.

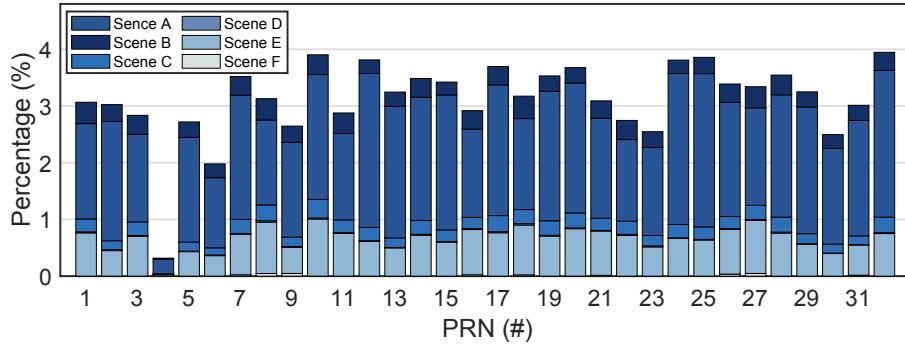


Fig. 4. Histogram of GPS records across 32 GPS satellites

In short, we utilized approximately 28 million GPS records collected through crowdsourcing efforts, of which 617,286 were collected by our team and the remainder sourced from publicly accessible datasets for large-scale evaluation. These extensive datasets fuel the next step of the establishment of SaRF.

4.2 Data Analysis

Fig. 3 illustrates the record distributions for six unique scenes, each serving as a benchmark for our experimental analysis. These scenes, characterized by their distinct layouts, are as follows: **Scene A**: Features a $30 \times 30 \times 34 \text{ m}^3$ building, surrounded by open areas and a car park, with notably dense data collection in the northern section. **Scene B**: Encompasses a larger, yet shorter building ($60 \times 90 \times 19 \text{ m}^3$) situated downtown, surrounded by pedestrian zones and a park. **Scene C**: Distinguished by a 5-way junction with buildings 24-35m in height, this scene's data predominantly originates from vehicles. **Scene D**: A busy, 100m segment of a 22 m-wide avenue, lined with buildings 16-28m tall, where data collection is largely vehicle-based. **Scene E**: An urban canyon with a narrow road flanked by buildings 15-20m tall, which likely experiences poor GPS reception. **Scene F**: Represents the architectural layout of a university campus, featuring three main buildings. Each scene presents unique architectural features, with diverse recesses and protrusions typical of urban structures.

Fig. 4 shows a histogram of the number of satellites visible in the entire dataset, with an average of 7.4 satellites visible at each GNSS location. Notably, satellite #4, the first GPS satellite launched in 1978 and now decommissioned, still appears occasionally in the data despite its retirement [43]. The distribution indicates that Scene A (the library) forms the largest part of the dataset, while Scene F (the campus) has a smaller representation. The amount of data from each scene is determined by factors like user traffic in the area and the duration of data collection.

5 Satellitic Radiance Fields

In this section, we transform the task of estimating the satellite radiance field into a problem akin to global illumination, a concept extensively explored in 3D computer graphics through radiosity. We discuss the attenuation and radiance characteristics of voxels. Our neural radiance network, inspired by Nerf², uses two MLPs to link 3D voxel coordinates with their attenuation and radiance properties, facilitating a detailed understanding of GNSS signal interactions within urban environments.

5.1 Radiosity

We segment an urban landscape into a multitude of small, continuous voxels and apply discrete geometry calculations to each. As GPS satellites emit continuous signals from the sky toward the ground, these voxels are illuminated in a manner akin to sunlight, with the satellite serving as the light source. Buildings within the scene either attenuate or reflect these RF signals, which are then captured by GPS receivers. This process resembles global illumination in computer graphics, where the intricate behavior of light in the real world is simulated to create more lifelike images.

The size of each voxel is a customizable parameter, balancing between computational performance and spatial accuracy. In terms of RF signal propagation, each voxel is characterized by specific attenuation and radiation properties.

5.1.1 Attenuation Characteristic. When a GPS signal traverses a voxel, it undergoes attenuation, which is contingent on the voxel's physical properties. For example, a voxel filled with air only causes negligible signal fading. In the standard RF model, the attenuation coefficient of the i^{th} voxel denoted by V_i is represented as a complex number, expressed by the equation:

$$h(V_i) = \Delta a(V_i) e^{j\Delta\theta(V_i)} \quad (1)$$

where $\Delta a(V_i)$ (normalized from 0 to 1) signifies the reduction in amplitude, while $\Delta\theta(V_i)$ (varying from 0 to 2π) indicates the phase shift. For ease of computation in the ray-marching algorithm, we transform this attenuation coefficient into a negative logarithmic form as below:

$$\begin{aligned} \delta(V_i) &= -\ln(h(V_i)) = -\ln(\Delta a(V_i) e^{j\Delta\theta(V_i)}) \\ &= -\ln \Delta a(V_i) - j\Delta\theta(V_i) \end{aligned} \quad (2)$$

where the real part $-\ln \Delta a(V_i)$ is non-negative since $\Delta a(V_i)$ is less than or equal to 1. Conversely, the imaginary part $-j\Delta\theta(V_i)$ is negative, reflecting the extent of phase rotation. The preference for using the negative logarithm in this context is due to two ranges being easily confined through the applications of ReLu and Sigmoid activation functions in the subsequent neural network layers.

The attenuation characteristic at a given RF frequency is influenced by physical attributes such as size, density, and the composition of the material, irrespective of whether the voxel is illuminated or not. Typically, a voxel filled with air exhibits a substantially lower attenuation coefficient compared to one made of concrete. This variance in attenuation properties plays a crucial role in accurately defining the contours of buildings in the next step.

5.1.2 Radiance Characteristic. In accordance with the Huygens–Fresnel principle, a voxel acts as a secondary source of radiance upon receiving signals from satellites directly or other voxels attributed to multipath effects. Consequently, we conceptualize each voxel V_i as an emergent RF source originating from the GPS signals transmitted by the GPS satellite O . This voxel then re-emits a new signal along the direction ω , which can be mathematically expressed as:

$$S(V_i, O, \omega) = a(V_i) e^{j\theta(V_i)} \quad (3)$$

where $\theta(V_i)$ represents the initial phase value and $a(V_i)$ the initial amplitude of the signal. Given that a voxel can potentially radiate in any direction, ω symbolizes a prospective 2D direction, denoted as $\omega = (\alpha, \beta)$. The parameters α and β , varying within $(0, 2\pi]$ and $(0, \pi/2]$ respectively, correspond to the azimuthal and elevation angles. This model allows us to account for the diverse and complex ways in which voxels may radiate RF signals, influenced by their interactions with incoming satellite signals and the surrounding environment. Understanding these radiative behaviors is crucial for accurately mapping and interpreting the RF signal propagation within urban landscapes.

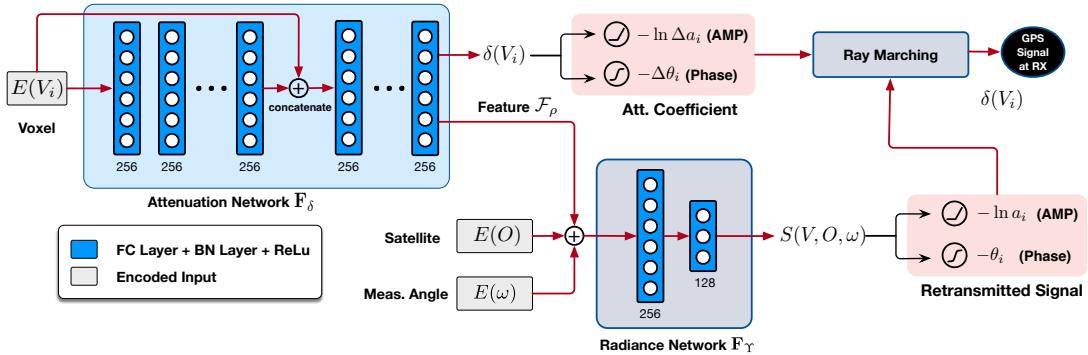


Fig. 5. Neural Network Architecture of SaRF

5.2 Neural Radiance Network

Drawing inspiration from NeRF², we employ two MLPs to model two key characteristics: the attenuation and radiance of each voxel. This approach leads us to designate the neural network as *satellitic radiance fields* (SaRF), as shown in Fig. 5. Central to SaRF is the concept of representing a scene using two distinct continuous functions. These functions, embodied by the MLPs, are tasked with linking 3D coordinates to the attenuation and radiance properties of a voxel. They process inputs such as the voxel's position V_i , the satellite's position s , and the viewing direction ω , to output the attenuation coefficient $\delta(V_i)$ and the reemitted GPS signal $S(V_i, \omega, s)$. The network is fundamentally structured into two segments: the attenuation subnetwork, which focuses on signal weakening, and the radiance subnetwork, which handles signal retransmission.

5.2.1 Attenuation Subnetwork. The first MLP, termed the *attenuation subnetwork*, is designed to correlate a voxel's position with its attenuation coefficient, which is represented by the following equation:

$$F_\delta : (V_i) \rightarrow (\delta(V_i), \mathcal{F}(V_i)) \quad (4)$$

The network processes the 3D position of a voxel and yields two outputs: the attenuation coefficient $\delta(V_i)$ and a 256-dimensional feature vector $\mathcal{F}(V_i)$. The coefficient $\delta(V_i)$ is a complex number. Its real component is modified using a ReLU activation function to guarantee that $-\ln(\Delta a(V_i)) \geq 0$ (implying $\Delta a(V_i) \leq 1$). The imaginary component, on the other hand, is adjusted with a $2\pi \times$ sigmoid function to constrain the phase shift between 0 and 2π . This feature vector $\mathcal{F}(V_i)$ is then employed as input for the subsequent radiance subnetwork.

Composed of eight fully connected layers, each with ReLU activations and 256 channels, the MLP is structured to process this data effectively. It is important to note that the attenuation characteristic is solely determined by the voxel's density and the structural composition of the scene, making it independent of the incoming RF signals. As a result, the attenuation subnetwork does not require satellite information as part of its input.

5.2.2 Radiance Subnetwork. The radiance subnetwork, represented by F_γ , is tasked with predicting the characteristics of the GPS signal that is retransmitted by a voxel. This prediction is based on the voxel's attenuation feature vector $\mathcal{F}(V_i)$, the observation direction ω , and the satellite position O (i.e., orbiter). The functionality of this subnetwork is encapsulated in the equation:

$$F_\gamma : (\mathcal{F}(V_i), O, \omega) \rightarrow (a(V_i), \theta(V_i)) \quad (5)$$

It is worth noting that the voxel's position is embedded into the feature vector. Comprising two fully connected layers equipped with ReLU activation functions, the subnetwork features 256 channels in the first layer and 128

in the second. Its output is the direction-dependent retransmitted GPS signal $a(V_i)e^{j\theta(V_i)}$, where the amplitude and phase components are respectively refined using ReLu and Sigmoid activation functions.

The radiance subnetwork plays a crucial role in modeling how each voxel interacts with the incoming GPS signals, transforming and re-emitting them based on their unique characteristics. The subnetwork's structure allows for the nuanced interpretation of signal behavior, accounting for variations in signal strength (amplitude) and the phase shift by the voxel's material properties and position relative to the satellite. By processing the amplitude with ReLU, the network ensures a non-negative output, while the Sigmoid function applied to the phase ensures it remains within a valid range. This detailed modeling is instrumental in accurately reconstructing the complex signal interactions within an urban environment, facilitating a deeper understanding of GPS signal propagation and its interaction with various urban structures.

5.3 Summary

Our approach contrasts with NeRF², which primarily targets signal prediction, making both of its subnetworks crucial. Different materials, such as concrete or wood, have distinct attenuation characteristics, enabling us to identify voxels composed of non-air materials. Once the attenuation subnetwork is proficiently trained, it becomes possible to reconstruct building structures within the scene. While the radiance subnetwork might appear superfluous for our end goal, it is in fact essential for the training process. Direct collection of ground truth data for attenuation coefficients is impractical, making it challenging to train the attenuation subnetwork in isolation. The training dataset, crowdsourced via smartphones, offers only GPS signal samples captured at various locations within the scene. Consequently, the training approach amalgamates both subnetworks, employing a semi-supervised method to train them jointly. This integrated training methodology ensures a comprehensive learning process, leveraging the strengths of both subnetworks to achieve accurate scene reconstruction.

6 Training

This section describes the self-supervised training approach for SaRF, utilizing crowdsourced GPS signal data. Training involves a divide-and-conquer strategy that predicts GPS signals using ray marching techniques, which trace and combine signals from individual voxels and satellite paths. The method adjusts for multiple satellites and all possible directions to minimize the prediction error against actual signal strengths. Additionally, 2D city maps are integrated to refine the model's urban accuracy, projecting 3D signal data onto a 2D plane to ensure alignment with real-world map features.

6.1 Divide and Conquer

Self-supervised learning is centered around devising a pretext task, a learning problem that can be resolved using only the input data. By tackling this task, the model learns to extract meaningful representations. For SaRF, the task is to accurately predict the GPS signal at a specific location. The objective of the training is to reduce the difference between the real GPS signal received by a smartphone and the model's predicted signal, all without direct supervision. This means that there are no pre-established ground-truth outputs for any of the two subnetworks involved in the process.

To leverage SaRF for predicting the GPS signal received at a specific location, we adopt a divide-and-conquer approach. As shown in Fig. 6, the GPS signal a smartphone receives can be methodically broken down. Specifically, ① the overall GPS signal is an amalgamation of signals from various satellites → ② the signal from each satellite is itself composed of signals from multiple directions → ③ the signal from a particular direction is an aggregate of signals from all voxels along that path → ④ the signal retransmitted from a specific voxel. By tracing the signal from individual voxels and cumulatively combining them, we can predict the final received signal. This technique, akin to ray marching in computer graphics, involves progressively tracing rays from the smartphone's

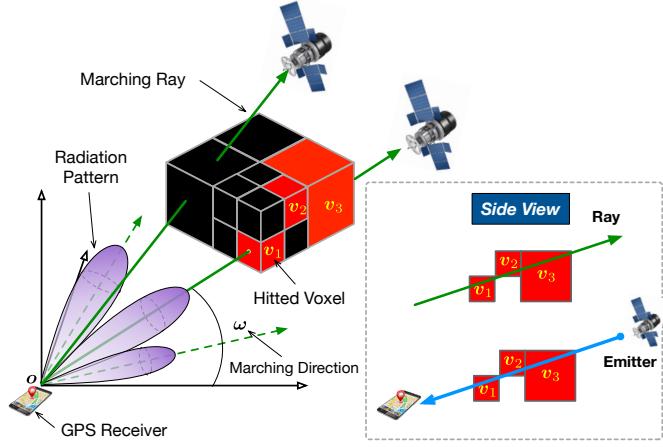


Fig. 6. Ray Marching. We progressively trace rays from the smartphone’s viewpoint into the scene to predict the GPS signal received by the receiver.

viewpoint into the scene. The subsequent sections will delve deeper into utilizing ray marching for predicting GPS signals received by smartphones.

6.2 Tracing from a Single Voxel

The RF signal propagation model can generally be depicted using the Friis transmission equation, represented as:

$$S_{RX} = H_{TX \rightarrow RX} S_{TX} \quad (6)$$

where S_{TX} and S_{RX} denote the transmitted and received signals, while $H_{TX \rightarrow RX}$ symbolizes the channel attenuation. Accordingly, the GPS signal emitted from a specific voxel V_i can be expressed as:

$$S_{vox}(P_0, V_i) = H_{V_i \rightarrow P_0} S(V_i, O, \omega) \quad (7)$$

In this scenario, $S(V_i, O, \omega)$ indicates the GPS signal re-emitted from V_i in the direction ω to the receiver located at P_0 , originating from the satellite O . If there are M voxels, denoted as $\{V_1, V_2, \dots, V_M\}$, along the path between V_i and P_0 , they collectively attenuate the GPS signal $S(V_i, O, \omega)$. The channel attenuation is thus formulated as:

$$\begin{aligned} H_{V_i \rightarrow P_0} &= H(V_1) \cdot H(V_2) \cdots \cdot H(V_M) = \prod_{m=1}^M \left(\Delta a(V_m) e^{j \Delta \theta(V_m)} \right) \\ &= \exp \left(- \sum_{m=1}^M -\ln (\Delta a(V_m) e^{j \Delta \theta(V_m)}) \right) \\ &= \exp \left(- \sum_{m=1}^M \delta(V_m) \right) \end{aligned} \quad (8)$$

where $\delta(V_m)$ represents the negative logarithmic attenuation caused by voxel V_m . The logarithmic form simplifies calculations, transforming multiplications into summations. Integrating this equation into Eqn. 7, we derive the GPS signal received at the receiver contributed by V_i as follows:

$$S_{vox}(P_0, V_i) = \exp \left(- \sum_{m=1}^M \delta(V_m) \right) \cdot S(V_i, O, \omega) \quad (9)$$

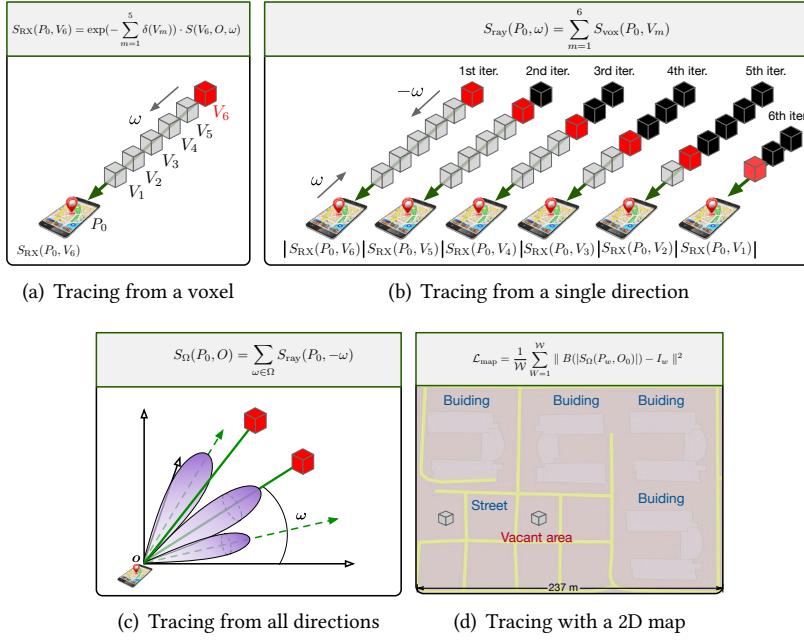


Fig. 7. Divide-and-Conquer Ray Marching Algorithm

where $\delta(V_m)$ and $S(V_i, O, \omega)$ are obtained from Eqn. 4 and Eqn. 5, respectively. Both are fitted by the two neural subnetworks.

Fig. 7(a) shows an example. The GPS signal $S(V_6, O, \omega)$, re-emitted by voxel V_6 (depicted as the red voxel), experiences attenuation while passing through voxels $V_1 - V_5$ (shown as gray voxels). The direction ω , relative to V_6 , results in an attenuation coefficient that is the sum of $\delta_1 + \delta_2 + \delta_3 + \delta_4 + \delta_5$, measured in dB units. For signal computation, this coefficient is reconverted to volts, and the received signal at the position P_0 is expressed as $S_{\text{vox}}(P_0, V_6) = \exp(-(\delta_1 + \dots + \delta_5)) \cdot S(V_6, O, \omega)$. There might be concerns about neglecting the RF signals reflected from adjacent voxels, which originally emanate from V_6 . It's crucial to recognize that in this model, each voxel is conceptualized as an RF source that captures signals from all possible directions. Therefore, reflections are treated as an integral part of the re-emission process of the neighboring voxels. This approach ensures that when tracing the signals from nearby voxels, such reflections are automatically counted in the analysis, providing a thorough representation of signal dynamics within the scene.

6.3 Tracing from a Single Direction

A given direction ω can be conceptualized as a ray originating from the position P_{RX} and extending towards ω . The points along this ray can be mathematically described as:

$$P(r, \omega) = P_0 + r\omega \quad (10)$$

where r represents the radial distance from the RX to the points on the ray, with $P(0, \omega) = P_0$ denoting the starting point. Consequently, the RX is capable of receiving signals that are retransmitted only from the voxels intersecting this particular ray.

Imagine a scenario where a ray intersects with N voxels, labeled as $\{V_1, V_2, \dots, V_N\}$, with V_1 being the closest to the receiver and V_N the farthest, also acting as the scene's boundary voxel. The signal received from direction

ω can be understood as the cumulative effect of the GPS signals re-emitted from these N voxels along the ray. This can be mathematically represented as:

$$\begin{aligned} S_{\text{ray}}(P_0, \omega) &= \sum_{n=1}^N S_{\text{vox}}(P_0, V_n) \\ &= \sum_{n=1}^N \left(\exp \left(- \sum_{m=1}^{n-1} \delta(V_m) \right) \cdot S(V_n, O, -\omega) \right) \end{aligned} \quad (11)$$

This formulation is derived from Eqn. 9. It's important to note that while the direction ω is defined with respect to the receiver, the re-emitted GPS signal from a voxel is relative to the voxel itself, effectively 180° opposite to ω . Therefore, $-\omega$ is used in the term $S(\cdot)$. An illustrative example is shown in Fig. 7(b), where we execute six iterations of voxel-based tracing. Each iteration calculates the GPS signal re-emitted from voxels V_6 to V_1 sequentially.

6.4 Tracing from all Directions

GPS receivers in smartphones typically come with directional antennas, but the orientation of the phones in the crowdsourced dataset is not known. Therefore, it's necessary to consider all possible directions, represented by Ω . As a result, the GPS signal received by a receiver from a satellite is the cumulative effect of signals from all these directions:

$$\begin{aligned} S_{\Omega}(P_0, O) &= \sum_{\omega \in \Omega} S_{\text{ray}}(P_0, \omega) = \sum_{\omega \in \Omega} \sum_{n=1}^N S_{\text{vox}}(P_0, V_n) \\ &= \sum_{\omega \in \Omega} \sum_{n=1}^N \left(\exp \left(- \sum_{m=1}^{n-1} \delta(V_m) \right) \cdot S(V_n, O, -\omega) \right) \end{aligned} \quad (12)$$

where the voxels of $\{V_1, V_2, \dots, V_N\}$ are on the direction ω . An example of this is illustrated in Fig. 7(c). Typically, Ω represents all possible directions. Yet, when the directionality of the GPS receiver is known, this range can be narrowed down significantly.

6.5 Tracing from all Satellites

For a specific detected satellite, we can collect a set of GPS signals at N distinct locations through crowdsourcing. The training process aims to minimize the following objective function:

$$\mathcal{L}_{\text{gps}} = \frac{1}{N} \sum_{n=1}^N \| |S_{\Omega}(P_n, O)| - |\tilde{S}_{\Omega}(P_n, O)| \| ^2 \quad (13)$$

where $|\tilde{S}_{\Omega}(P_n, O)|$ denotes the strength of actual raw signal received at position P_n , which is originated from the satellite O (i.e., the location of the satellite at the current time). The goal of this objective function is to reduce the strength discrepancy between the predicted GPS signal, $S_{\Omega}(P_n, O)$, as calculated by our model, and the real-world signal, $\tilde{S}_{\Omega}(P_n, O)$, gathered from the field.

A smartphone typically has the capability to simultaneously receive GPS signals from at least three satellites at any given location. This means that the voxels within a scene are illuminated by multiple satellites. Utilizing the CDMA encoding scheme, the smartphone is able to distinguish and decode each satellite's GPS signals, capturing both amplitude and phase information. Let us assume that the smartphone collects GPS signals from \mathcal{K} satellites, represented as $O_1, O_2, \dots, O_{\mathcal{K}}$. Under these circumstances, the loss function is modified as follows:

$$\mathcal{L}_{\text{gps}} = \frac{1}{\mathcal{K}N} \sum_{k=1}^{\mathcal{K}} \sum_{n=1}^N \| |S_{\Omega}(P_n, O_k)| - |\tilde{S}_{\Omega}(P_n, O_k)| \| ^2 \quad (14)$$

Assuming that the scene remains static, the objective of the training is to minimize the difference between the received and predicted GPS signals at various positions, emanating from different satellites. Typically, smartphones report the estimated amplitude of signals in dBm and their phase. For the purposes of our model, it is necessary to convert the amplitude from dBm to volts. Once converted, these amplitude values, along with the phase information, are combined to form a complex exponential number. This approach allows for a more accurate representation of the GPS signals and facilitates more precise training and prediction of signal behaviors in the model.

6.6 Tracing with a Known 2D Map

2D city maps, providing a bird's eye view of urban areas, display streets, landmarks, buildings, and other features, aiding in navigation and urban analysis. These maps are useful for various stakeholders, including tourists, residents, and planners, for city exploration and infrastructure study. An example from Google Maps, shown in Fig. 7(d), illustrates buildings in gray, roads in yellow, and vacant areas in pink, effectively portraying the physical layout. Leveraging the accessibility of 2D maps, we use them as a supportive tool in training SaRF. We simulate GPS signals as coming from above to the ground level, where the receiver is conceptually placed corresponding to a pixel on the 2D map. If the receiver is under a building, complete signal attenuation is expected; if in open space, minimal attenuation occurs. The 2D map is thus converted to a binary format, with buildings marked as zero and open areas as one. Our training focuses on minimizing a specific loss function derived from this binary map representation:

$$\begin{aligned} \mathcal{L}_{\text{map}} &= \frac{1}{W} \sum_{w=1}^W \| B(|S_\Omega(P_w, O_0)|) - I_w \|^2 \\ \text{where } B(x) &= \begin{cases} 1 & x > \xi \\ 0 & \text{otherwise} \end{cases} \end{aligned} \quad (15)$$

where P_w refers to the GPS coordinates of the w^{th} pixel in the 2D map, which spans the scene of interest. ξ represents the threshold for minimal detectable GPS signal strength. O_0 is the position of a virtual satellite, strategically placed at the center above the scene at an altitude typical of GPS satellites (20,200 km), allowing us to assume that the GPS signals are parallel with a 90° angle of incidence. For consistency, all coordinates in the earlier equations are converted to GPS-based coordinates, including longitude, latitude, and altitude. I_w is the value of the binarized pixel, set to zero for pixels within building outlines, implying no measurable GPS signal ($|S_\Omega(P_w, O_0)| \leq \xi$) underneath a building. The network is expected to recognize these as concrete structures. Conversely, when $I_w = 1$, the actual GPS signals should align with the predicted values. This process essentially projects the predicted 3D map onto a 2D plane, facilitating a comparison with the actual 2D map, particularly regarding building locations.

6.7 Summary

Finally, we put the pieces together to get the joint training loss function as follows:

$$\mathcal{L} = (1 - \lambda_1 - \lambda_2) \mathcal{L}_{\text{gps}} + \lambda_1 \mathcal{L}_{\text{map}} + \lambda_2 \cdot \zeta(|S_{\text{RX}}(p_0, s_k)|) \quad (16)$$

where ζ is a beta-distribution regularizer introduced by [28, 29], λ_1 and λ_2 are hyperparameters. This loss function aims to minimize the difference between the measured and predicted GPS signals at the smartphone's positions. Meanwhile, it obtains hints about the boundaries and shapes of the building from the 2D map. Certainly, the map loss is optional if the 2D map is not available.

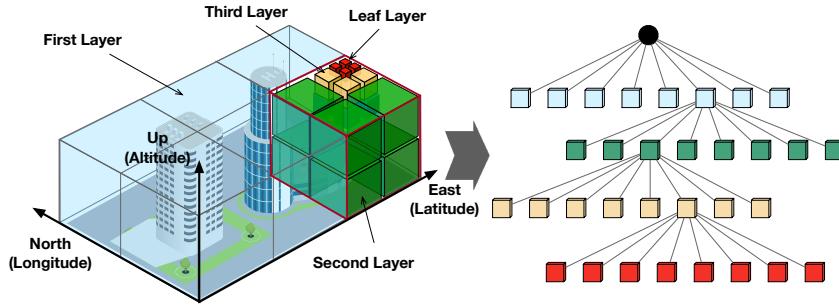


Fig. 8. Octree-based Voxelization

7 3D Map Reconstruction

In this section, we detail the process for constructing a 3D map utilizing SaRF. Materials with higher density tend to cause more significant attenuation of RF signals. This implies a direct correlation between a voxel's density and the level of attenuation it imparts on a GPS signal. We thus define the relative density of a voxel as follows:

$$\rho(V_i) = |\exp(\delta(V_i))| = \frac{1}{\Delta a(V_i)} \quad (17)$$

where $\Delta a(V_i)$ indicates the attenuation of the RF signal that traverses through the voxel V_i . The density of a voxel is inversely proportional to the amount of signal it absorbs. The relative densities of all voxels are acquired from the attenuation subnetwork. By establishing a threshold value, denoted as ρ_{air} , we classify a voxel to be a component of a building when its relative density $\rho(V_i)$ exceeds ρ_{air} . If not, the voxel is identified as air. This approach enables the reconstruction of buildings within the scene by omitting voxels identified as air, thereby effectively differentiating between solid structures and open areas. The final step involves identifying a suitable bounding box that encompasses all the non-air voxels, facilitating the creation of a comprehensive 3D map.

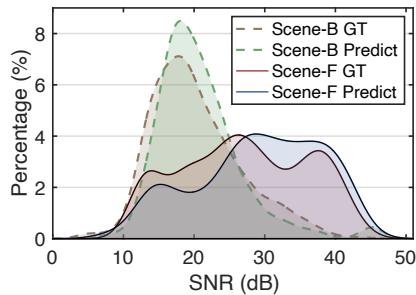
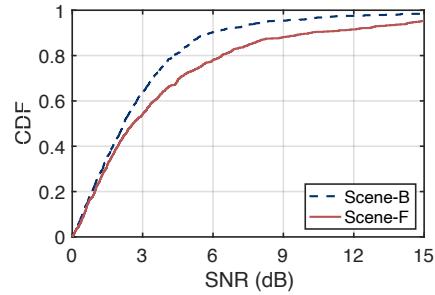
Progressive Training. Employing fine-grained uniform voxelization enhances the model's resolution and accuracy. However, the sheer number of voxels poses a challenge for computational efficiency, especially during ray marching procedures. Consider a scenario where each voxel measures 10 cm^3 . In a 100 m^3 scene, this results in a billion voxels. Consequently, a single ray marching along a direction could involve interactions with over a thousand voxels, leading to an excessively high computational burden. Moreover, many areas in a typical scene are ‘vacant,’ filled predominantly with air, contributing minimal attenuation to the ray marching process. To address this, we adopt a hierarchical data structure known as an octree for representing the 3D scene as shown in Fig. 8. The fundamental principle of an octree is to recursively subdivide space into eight smaller segments, or ‘octants.’ Each octant corresponds to a node in the tree structure. The tree is structured such that each node either has eight child nodes or none, with the root node encapsulating the entire space and each subsequent level representing a division into smaller subspaces. Importantly, each node in this structure contains numerous voxels. To build an octree in local coordinate, we transform the coordinate from LLA (Longitude, Latitude, Altitude) to ENU (East, North, Up), in which we select a reference point as the origin, east is the x-axis, north is the y-axis, and up is the z-axis.

8 Results

In this section, we evaluate the performance of the 3D map reconstruction using SaRF.

Table 1. Annotated Voxels Description

Scenes (#)	A	B	C	D	E	F
Records	17.1M	2.6M	1.8M	5.8M	121K	617K
Building	14,012	12,245	13,680	19,097	12,371	7,026
Vacant	23,428	18,811	30,864	22,999	13,069	23,598
Total	37,440	31,056	44,544	42,096	25,440	30,624

**Fig. 9. PDFs of SNR Prediction****Fig. 10. CDFs of SNR Prediction**

8.1 Implementation

In our experiment, we set $\lambda_1 = 0.01$ and $\lambda_2 = 0.001$. We use a batch size of 16 coupled with a cosine learning rate scheduler that varies between 10^{-4} and 10^{-6} . The loss is computed for each sample. We use Adam to optimize the loss function. The direction space Ω is sampled to $36 \times 9 = 324$ directions. We conducted our training over approximately 500,00 iterations, which took about 10 hours on a single NVIDIA GTX 4090. We start by creating a balanced octree with a root node size of 200 m with coarse leaf nodes of 6.25 m. Nodes can be further subdivided down to the smallest unit, a 0.39 m voxel, if its relative density exceeds $\rho_{\text{air}} = 0.5$. If not, the node undergoes self-pruning. This method not only maintains the model's high resolution where necessary but also significantly reduces computational load in areas with little to no significant content.

Ground Truth. For the ground truth 3D map of the campus, we utilized Google Earth for its high-resolution 3D building models and footprints. Buildings and open spaces were manually annotated with voxels, as shown in Table 1, which lists the number of voxels annotated. "Records" means the amount of collected GNSS data in the scenes. Voxels within buildings were labeled as "Building", while those in open spaces near buildings were marked as "Vacant". These labels, numbering over 25,000 per scene, are crucial for assessing the reconstruction accuracy of our model.

8.2 Accuracy of Satellite SNR Prediction

Evolving from NeRF², which is capable of predicting received signals at any given location, SaRF has also been endowed with the same predictive capability. A higher level of prediction accuracy signifies a more precise fit of the neural radiance fields, as the predictions emerge from ray tracing by using the two MLPs. Therefore, our initial evaluation of SaRF focuses on its accuracy in predicting the GPS SNR. During our experiments, we randomly selected 80% of the datasets from Scenes B and F for training, reserving the remaining 20% for testing purposes.

Table 2. Accuracy of Reconstruction

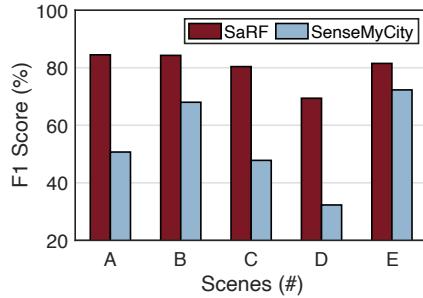
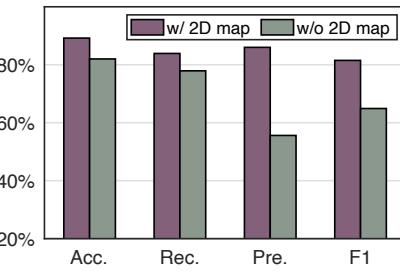
Dataset	SaRF			SenseMyCity		
	Acc. ↑	Rec. ↑	Pre. ↑	Acc. ↑	Rec. ↑	Pre. ↑
Scene A	<u>85.8%</u>	<u>77.8%</u>	92.5%	78.5%	73.2%	49.5%
Scene B	<u>85.1%</u>	<u>79.4%</u>	<u>89.8%</u>	81.5%	69.5%	76.5%
Scene C	<u>80.9%</u>	<u>92.8%</u>	70.9%	68.0%	38.0%	<u>81.0%</u>
Scene D	<u>73.7%</u>	<u>59.7%</u>	83.0%	63.3%	33.5%	<u>85.3%</u>
Scene E	<u>84.0%</u>	96.3%	<u>70.6%</u>	82.5%	93.8%	61.0%
Scene F	89.2%	83.9%	86.0%	—	—	—
Avg	83.1%	81.7%	82.1%	74.8%	61.6%	70.7%

The results of the predicted SNR, when compared with the ground truth (GT) and represented as a PDF, are depicted in Fig. 9. The prediction closely mirrors the actual SNR distribution. Notably, the SNR predominantly ranges between 10 and 30 dB for Scene B, while it spans a broader spectrum of 10 – 40 dB for Scene F, attributable to the latter's more extensive area coverage. Further quantifying the accuracy, we computed the SNR errors, defined as the absolute deviation between the predicted SNR and the ground truth. The CDF of these prediction errors is presented in Fig. 10. For Scene B, we attained a median error of 2.3 dB (with the 10th percentile at 0.4 dB and the 90th percentile at 5.9 dB), and for Scene F, the median error was 2.2 dB (with the 10th percentile at 0.6 dB and the 90th percentile at 10.3 dB). These findings affirm that SaRF possesses a high degree of accuracy in modeling GPS signal propagation, which is advantageous for reconstructing 3D maps.

8.3 Accuracy of Reconstruction

We extended our evaluation to assess building reconstruction performance using SaRF and the SenseMyCity dataset. To measure the outcomes, we employed balanced accuracy (Acc.), precision (Pre.), and recall (Rec.) as principal metrics. Accuracy is a universal metric reflecting the proportion of voxels accurately classified. This metric treats all voxels with equal weight, which can lead to skewed results in datasets with imbalances, like a predominance of "vacant" voxels over "Building" voxels. Balanced accuracy, defined as $\frac{1}{2} \left(\frac{\text{TP}}{\text{TP}+\text{FN}} + \frac{\text{TN}}{\text{TN}+\text{FP}} \right)$, addresses this by computing the mean of the correct classification rates for each class separately, thereby adjusting for any disparities in class distribution. Precision, denoted as $\frac{\text{TP}}{\text{TP}+\text{FP}}$, evaluates the correct identification rate of building voxels within the octree, with TP being true positives and FP false positives. Recall, defined as $\frac{\text{TP}}{\text{TP}+\text{FN}}$, measures the model's success in correctly identifying actual building voxels, indicative of its comprehensive detection ability. Higher values for these three metrics are preferable.

Partial Analysis. Table 2 showcases the reconstruction performance comparison between SaRF and SenseMyCity, with the latter's results sourced from [40] (Scene F is unavailable to SenseMyCity). Our findings indicate that (1) SaRF surpasses SenseMyCity with an average balanced accuracy, recall, and precision of 83.1%, 81.7%, and 82.1%, respectively, while SenseMyCity, utilizing the Random Forest classifier, achieves 74.8%, 61.6%, and 70.7%. This underscores SaRF's superior performance in precise classification and detection of building voxels, along with more effective voxel pruning within the octree framework. (2) SaRF attains peak balanced accuracy of 89.2% in Scene F, a peak recall of 96.3% in Scene E, and peak precision of 92.5% in Scene A. These scenes, associated with residential and educational environments, are characterized by fewer dynamic objects such as vehicles that could otherwise impact model precision.

**Fig. 11. Performance of F1 score****Fig. 12. Ablation Study**

Overall Analysis. For a comprehensive evaluation of SaRF’s performance, we utilize the F1 score, which represents the harmonic mean of precision and recall and accounts for both metrics concurrently. Higher F1 scores are indicative of more accurate reconstruction. Fig. 11 shows a comparison of the F1 scores between SaRF and SenseMyCity across five different scenes. The data reveals that (1) SaRF nearly attains F1 scores exceeding 80%, with specific scores of 84.5%, 84.3%, 80.4%, 69.4% and 81.5% for Scene A, B, C, D and E, respectively. On the other hand, SenseMyCity’s F1 scores are comparatively lower, at 50.7%, 68.0%, 47.8%, 32.3%, and 72.3%, respectively. SaRF outperforms SenseMyCity by 25.6%. (2) The F1 scores in Scene D are notably lower, a consequence of particularly poor recall in that scene (59.7% and 33.5%). In Scene D, which features a complex architectural layout with irregular buildings and ambiguous open spaces, numerous voxels are misclassified from open air to building, impacting the accuracy.

Summary. SaRF surpasses existing state-of-the-art methodologies due to two primary factors. Firstly, prior approaches rely solely on line-of-sight (LOS) assumptions to deduce the presence of buildings, neglecting non-line-of-sight (NLOS) signal paths. This simplification is problematic since distinguishing LOS from NLOS components in received signals is nearly impossible, often rendering the assumption inaccurate. In contrast, SaRF fully accounts for all potential signal paths using the ray marching algorithm. Secondly, SaRF enhances prediction accuracy by incorporating the physical properties of objects encountered along signal paths—something not considered in previous models. These properties are finely represented through the dual MLPs within our system.

8.4 Ablation Study

An ablation study was performed to assess the contribution of \mathcal{L}_{map} to the model’s performance using Scene F as the test case. Specifically, we evaluated the performance of SaRF with and without the integration of the 2D map in the training process, focusing on balanced accuracy, precision, recall, and F1 score. The findings revealed that without the 2D map, SaRF’s performance metrics experienced a drop: accuracy fell to 82.0%, recall to 77.9%, precision to 55.6%, and F1 score to 64.9%. These figures reflect declines of 7.2%, 6.0%, 30.4%, and 16.6%, respectively, when compared to the results achieved with the inclusion of \mathcal{L}_{map} . The most significant reductions were observed in precision and the F1 score, underscoring the 2D map’s pivotal role in curbing false positives, where non-building voxels are incorrectly identified as part of a building. Hence, the integration of \mathcal{L}_{map} into the training markedly curtails false positives, thereby bolstering the precision and overall performance of SaRF.

8.5 Impact of Altitude

Then, we assessed the balanced accuracy of SaRF across different voxel altitudes within the reconstructed 3D map. The evaluation was carried out on Scene A (Library) and Scene F (Campus), calculating the average accuracy for reconstructed voxels at different height levels. The campus scene featured building heights around 23 m, while the library scene had buildings approximately 31 m tall. We analyzed accuracy for voxel heights from 2 m

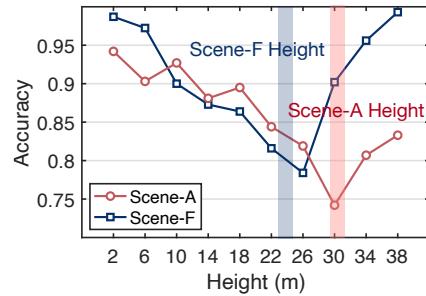


Fig. 13. Impact of Altitude

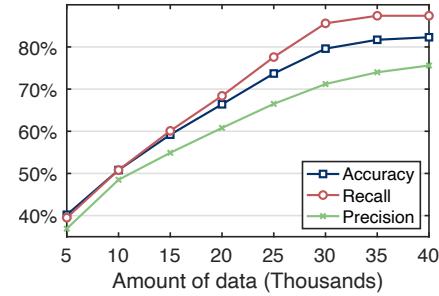


Fig. 14. Impact of Data Amount

to 38 m, in increments of 4 m, as depicted in Fig. 13. Our results indicated that accuracy fluctuated across different altitudes. Typically, accuracy peaked at ground level, then gradually diminished as height increased, up to a point close to the building's height, before rebounding. For example, accuracy was (98.1%, 94.3%) at 2m, dipped to (78.4%, 74.2%) at 23m and 30m, and then increased to (99.8%, 83.6%) at 38m. The average accuracy across all layers was 89.2% for Scene A and 85.8% for Scene F.

The pattern of varying accuracy across different voxel altitudes can be explained by the limitations of the ray tracing technique. GPS signals are predominantly captured at ground level, resulting in ray tracing predominantly occurring in upward directions from the ground. Particularly during directional tracing, voxels nearer to the ground are more frequently traced, while those at higher altitudes may be traced less often. This occurs because the separation between two distinct directional paths increases with altitude, leading to less effective tracing of higher voxels. Additionally, the standard practice of assuming voxels to be vacant by default contributes to improved accuracy at levels above the actual height of buildings, thereby enhancing the accuracy observed in the upper layers of the reconstruction.

8.6 Impact of Training Scale

Next, we assessed how varying amounts of data affect the performance of 3D building reconstruction in SaRF. For this evaluation, we selected a building in the Scene B. Fig.14 illustrates the changes in accuracy, recall, and precision as the volume of data increments from 5K to 40K in steps of 5K. Initial results with 5K data show SaRF achieving 40.2% accuracy, 36.9% recall, and 39.5% precision. As the data volume increases, these metrics progressively improve. However, beyond 30K data, the rate of performance enhancement slows down. At 40K data, SaRF reaches 82.3% accuracy, 75.6% recall, and 87.4% precision. Compared to the full dataset training results (approximately 120K), detailed in Table2, there is a performance decrease of 4.1%, 4.8%, and 2.8% in accuracy, recall, and precision, respectively. Considering the balance between data collection efforts and performance efficiency, we recommend 40K data as an optimal amount for effective building reconstruction.

8.7 Visualization

Finally, we visualize the reconstruction results in Fig. 15. We employ the City3D toolbox [22] to draw the recognized building voxels and bounding boxes on Google Earth. In the model, red and yellow represent the higher and lower-density building voxels, respectively. The visualization distinctly shows that our recognition results are more detailed and precise compared to those achieved by SenseMyCity. Regardless of the coverage or accuracy, SaRF outperforms SenseMyCity.



Fig. 15. Illustration of 3D Urban Constructions. The first row displays the ground truth for five scenes, with 3D building models sourced from Google Earth and voxels labeled manually. The second row illustrates the reconstruction outcomes from the related project, SenseMyCity. The third row presents the 3D voxels reconstructed using SaRF. The fourth row depicts the bounding boxes of the reconstructed buildings, where each box circumscribes relevant voxels.

9 Limitations and Future Works

Many aspects of our work can be improved, as summarized below:

Dynamics Interference: In SaRF, we assume that GNSS signals are affected by urban structures, enabling the construction of satellite neural radiance fields to analyze the attenuation properties of each voxel. However, GNSS signals are influenced by a variety of factors, including moving traffic, atmospheric conditions, and human gestures. Specifically, obstructions caused by the human body can lead to weaker GNSS signals. A potential solution for future work could involve utilizing a calibration algorithm to filter data exclusively related to building structures.

Degradation Near Building Heights: Our experiments indicate an accuracy degradation near building heights. This occurs because most data are collected at ground level, where human activity is prevalent. Consequently, fewer rays are traced at higher altitudes, leading to an imbalance in training samples. This limitation could be addressed by crowdsourcing GNSS data from higher floors or using data collected by UAVs.

Scene Representation: We currently use an octree-based voxel representation to model scenes. While efficient, our model is constrained to block-scale scenes due to the limited scalability of the MLP in SaRF. To adapt to larger scales, we could employ multiple sub-SaRF to integrate into a city-scale model. Furthermore, the exploration of more effective and explicit scene representations, such as 3D Gaussian spheres or point cloud-based models, could provide more accurate and scalable solutions.

10 Conclusion

We introduce SaRF, a framework designed for 3D urban mapping utilizing crowdsourced GNSS data. The key of SaRF lies in its innovative voxel-based representation of urban structures, coupled with the use of radio frequency

neural radiance fields for learning the attenuation properties of these voxels. This unique combination allows our framework to efficiently and accurately reconstruct urban building structures based on GNSS data, representing a significant advancement in the field of urban mapping.

Acknowledgments

This work is supported by the NSFC Key Program (No. 61932017), UGC/GRF (No. 15204820, 15215421), Innovation and Technology Fund (ITS/099/21), and Shenzhen Fundamental Research Program (No. JCYJ20230807140410022). We thank all the anonymous reviewers for their valuable comments and helpful suggestions.

References

- [1] Fatemeh Alidoost, Hossein Arefi, and Federico Tombari. 2019. 2D Image-To-3D Model: Knowledge-Based 3D Building Reconstruction (3DBR) Using Single Aerial Images and Convolutional Neural Networks (CNNs). *Remote Sensing* 11, 19 (2019), 2219.
- [2] Simon Banville and Frank van Diggelen. 2016. Precision GNSS for Everyone. *GPS World* 27, 11 (2016), 43–48.
- [3] Jonathan T Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P Srinivasan. 2021. Mip-NeRF: A Multiscale Representation for Anti-Aliasing Neural Radiance Fields. In *Proc. of IEEE/CVF ICCV*. 5855–5864.
- [4] Michael Bevis, Steven Businger, Thomas A Herring, Christian Rocken, Richard A Anthes, and Randolph H Ware. 1992. GPS Meteorology: Remote Sensing of Atmospheric Water Vapor Using the Global Positioning System. *Journal of Geophysical Research: Atmospheres* 97, D14 (1992), 15787–15801.
- [5] Md Fazlay Rabbi Masum Billah, Nurani Saoda, Jiechao Gao, and Bradford Campbell. 2021. BLE Can See: A Reinforcement Learning Approach for RF-Based Indoor Occupancy Detection. In *Proc. of ACM/IEEE IPSN*. 132–147.
- [6] Martin Breunig and Sisi Zlatanova. 2011. 3D Geo-Database Research: Retrospective and Future Directions. *Computers & Geosciences* 37, 7 (2011), 791–803.
- [7] Chu Cao, Zhidan Liu, Mo Li, Wenqiang Wang, and Zheng Qin. 2018. Walkway Discovery from Large Scale Crowdsensing. In *Proc. of ACM/IEEE IPSN*. IEEE, 13–24.
- [8] Dongyao Chen and Kang G Shin. 2019. TurnsMap: Enhancing Driving Safety at Intersections with Mobile Crowdsensing and Deep Learning. *Proc. of the ACM IMWUT* 3, 3 (2019), 1–22.
- [9] Longbiao Chen, Chenhui Lu, Fangxu Yuan, Zhihan Jiang, Leye Wang, Daqing Zhang, Ruixiang Luo, Xiaoliang Fan, and Cheng Wang. 2021. UVLens: Urban Village Boundary Identification and Population Estimation Leveraging Open Government Data. *Proc. of the ACM IMWUT* 5, 2 (2021), 1–26.
- [10] Arnis Cirulis and Kristaps Brīgmanis. 2013. 3D Outdoor Augmented Reality for Architecture and Urban Planning. *Procedia Computer Science* 25 (2013), 71–79.
- [11] Pablo Corbalán, Gian Pietro Picco, and Sameera Palipana. 2019. Chorus: UWB Concurrent Transmissions for GPS-like Passive Localization of Countless Targets. In *Proc. of ACM/IEEE IPSN*. 133–144.
- [12] Vladimir Coric and Marco Gruteser. 2013. Crowdsensing Maps of On-Street Parking Spaces. In *Proc. of IEEE DC OSS*. IEEE, 115–122.
- [13] Yoann Corre and Yves Lostanlen. 2009. Three-Dimensional Urban EM Wave Propagation Model for Radio Network Planning and Optimization over Large Areas. *IEEE Transactions on Vehicular Technology* 58, 7 (2009), 3112–3123.
- [14] Esteban Egea-Lopez, Jose Maria Molina-Garcia-Pardo, Martine Lienard, and Pierre Degauque. 2021. Opal: An Open Source Ray-Tracing Propagation Simulator for Electromagnetic Characterization. *Plos one* 16, 11 (2021), e0260060.
- [15] Ashok Elluswamy. 2022. CVPR 2022 Workshop on Autonomous Driving. https://www.youtube.com/watch?v=jPCV4GKX9Dw&ab_channel=WADatCVPR.
- [16] Omid Esrafilian, Rajeev Gangula, and David Gesbert. 2020. Three-Dimensional-Map-Based Trajectory Design in UAV-Aided Wireless Localization Systems. *IEEE Internet of Things Journal* 8, 12 (2020), 9894–9904.
- [17] Zhihan Fang, Guang Wang, Xiaoyang Xie, Fan Zhang, and Desheng Zhang. 2021. Urban Map Inference by Pervasive Vehicular Sensing Systems with Complementary Mobility. *Proc. of the ACM IMWUT* 5, 1 (2021), 1–24.
- [18] Zhihan Fang, Fan Zhang, Ling Yin, and Desheng Zhang. 2018. MultiCell: Urban Population Modeling Based on Multiple Cellphone Networks. *Proc. of the ACM IMWUT* 2, 3 (2018), 1–25.
- [19] Lovnesh Kumar Goyal, Rajiv Chauhan, Raman Kumar, and Hardeep Singh Rai. 2020. Use of BIM in Development of Smart Cities: A Review. In *IOP Conference Series: Materials Science and Engineering*, Vol. 955. IOP Publishing, 012010.
- [20] Paul D Groves. 2011. Shadow Matching: A New GNSS Positioning Technique for Urban Canyons. *The journal of Navigation* 64, 3 (2011), 417–430.
- [21] Junfeng Guan, Sohrab Madani, Suraj Jog, Saurabh Gupta, and Haitham Hassanieh. 2020. Through Fog High-Resolution Imaging Using Millimeter Wave Radar. In *Proc. of IEEE/CVF CVPR*.

- [22] Jin Huang, Jantien Stoter, Ravi Peters, and Liangliang Nan. 2022. City3D: Large-Scale Building Reconstruction from Airborne LiDAR Point Clouds. *Remote Sensing* 14, 9 (2022), 2254.
- [23] Andrew T Irish, Jason T Isaacs, François Quitin, Joao P Hespanha, and Upamanyu Madhow. 2014. Probabilistic 3D Mapping Based on GNSS SNR Measurements. In *Proc. of IEEE ICASSP*. IEEE, 2390–2394.
- [24] Qideng Jiang, Yongtao Ma, Kaihua Liu, and Zhi Dou. 2016. A Probabilistic Radio Map Construction Scheme for Crowdsourcing-Based Fingerprinting Localization. *IEEE Sensors Journal* 16, 10 (2016), 3764–3774.
- [25] Avinash Kalyanaraman, Elahe Soltanaghaei, and Kamin Whitehouse. 2019. Doorpler: A Radar-Based System for Real-Time, Low Power Zone Occupancy Sensing. In *Proc. of IEEE RTAS*. IEEE, 42–53.
- [26] Kihwan Kim, Jay Summet, Thad Starner, Daniel Ashbrook, Mrunal Kapade, and Irfan Essa. 2008. Localization and 3D Reconstruction of Urban Scenes using GPS. In *2008 12th IEEE international symposium on wearable computers*. IEEE, 11–14.
- [27] Thomas H Kolbe, Gerhard Gröger, and Lutz Plümer. 2008. CityGML—3D City Models and Their Potential for Emergency Response. In *Geospatial information technology for emergency response*. CRC Press, 273–290.
- [28] Lingjie Liu, Jiatao Gu, Kyaw Zaw Lin, Tat-Seng Chua, and Christian Theobalt. 2020. Neural Sparse Voxel Fields. *Advances in Neural Information Processing Systems* 33 (2020), 15651–15663.
- [29] Stephen Lombardi, Tomas Simon, Jason Saragih, Gabriel Schwartz, Andreas Lehrmann, and Yaser Sheikh. 2019. Neural Volumes: Learning Dynamic Renderable Volumes from Images. *ACM Transactions on Graphics (TOG)* 38, 4 (2019), 1–14.
- [30] Ricardo Martin-Brualla, Noha Radwan, Mehdi SM Sajjadi, Jonathan T Barron, Alexey Dosovitskiy, and Daniel Duckworth. 2021. NeRF in the Wild: Neural Radiance Fields for Unconstrained Photo Collections. In *Proc. of the IEEE/CVF CVPR*. 7210–7219.
- [31] Hiroshi Masaharu and Hiroyuki Hasegawa. 2000. Three-Dimensional City Modeling from Laser Scanner Data by Extracting Building Polygons Using Region Segmentation Method. *International Archives of Photogrammetry and Remote Sensing* 33, B3/1; PART 3 (2000), 556–562.
- [32] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. 2020. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In *Proc. of ECCV*. Springer, 405–421.
- [33] Alex Minetto, Maria Chiara Bello, and Fabio Dovis. 2022. DGNS Cooperative Positioning in Mobile Smart Devices: A Proof of Concept. *IEEE Transactions on Vehicular Technology* 71, 4 (2022), 3480–3494.
- [34] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. 2022. Instant Neural Graphics Primitives with A Multiresolution Hash Encoding. *ACM transactions on graphics (TOG)* 41, 4 (2022), 1–15.
- [35] Ana Nika, Zhijing Li, Yanzi Zhu, Yibo Zhu, Ben Y Zhao, Xia Zhou, and Haitao Zheng. 2016. Empirical Validation of Commodity Spectrum Monitoring. In *Proc. of ACM Sensys*. 96–108.
- [36] Jacek Paziewski, Marco Fortunato, Augusto Mazzoni, and Robert Odolinski. 2021. An Analysis of Multi-GNSS Observations Tracked by Recent Android Smartphones and Smartphone-Only Relative Positioning Results. *Measurement* 175 (2021), 109162.
- [37] SENSEMYCITY Project. 2019. Sense My City. <https://sensemycity.pt/>.
- [38] Konstantinos Rematas, Andrew Liu, Pratul P Srinivasan, Jonathan T Barron, Andrea Tagliasacchi, Thomas Funkhouser, and Vittorio Ferrari. 2022. Urban Radiance Fields. In *Proc. of the IEEE/CVF CVPR*. 12932–12942.
- [39] Remcom Inc. 2024. Electromagnetic Simulation Software. <https://www.remcom.com/>.
- [40] João GP Rodrigues and Ana Aguiar. 2019. Extracting 3D Maps from Crowdsourced GNSS Skyview Data. In *Proc. of ACM MobiCom*. 1–15.
- [41] João GP Rodrigues, Ana Aguiar, and Cristina Queirós. 2016. Opportunistic Mobile Crowdsensing for Gathering Mobility Information: Lessons Learned. In *Proc. of IEEE ITSC*. IEEE, 1654–1660.
- [42] Jinghao Shi, Zhangyu Guan, Chunming Qiao, Tommaso Melodia, Dimitrios Koutsonikolas, and Geoffrey Challen. 2014. Crowdsourcing Access Network Spectrum Allocation Using Smartphones. In *Proc of ACM workshop on HotNets*. 1–7.
- [43] John L. Sloan. 2018. GPS Satellite PRN 4. <https://coverclock.blogspot.com/2018/11/gps-satellite-prn-4.html>.
- [44] Pratul P Srinivasan, Boyang Deng, Xiuming Zhang, Matthew Tancik, Ben Mildenhall, and Jonathan T Barron. 2021. NeRV: Neural Reflectance and Visibility Fields for Relighting and View Synthesis. In *Proc. of the IEEE/CVF CVPR*. 7495–7504.
- [45] Deni Suwardhi, Shafarina Wahyu Trisyanti, Riantini Virtriana, Adib Aulia Syamsu, Shefi Jannati, and Rizky Syamsudin Halim. 2022. Heritage Smart City Mapping, Planning and Land Administration (Hestya). *ISPRS International Journal of Geo-Information* 11, 2 (2022), 107.
- [46] RP Swinford. 2005. Building On-the-Fly World Models for Pervasive Gaming and Other Ubicomp Applications Using GPS Availability Data. In *The IEEE International Workshop on Intelligent Environments*. IET, 133–142.
- [47] Matthew Tancik, Vincent Casser, Xinchen Yan, Sabeek Pradhan, Ben Mildenhall, Pratul P Srinivasan, Jonathan T Barron, and Henrik Kretzschmar. 2022. Block-NeRF: Scalable Large Scene Neural View Synthesis. In *Proc. of IEEE/CVF CVPR*. 8248–8258.
- [48] Matthew Tancik, Pratul Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan Barron, and Ren Ng. 2020. Fourier Features Let Networks Learn High Frequency Functions in Low Dimensional Domains. *Advances in Neural Information Processing Systems* 33 (2020), 7537–7547.
- [49] Narseo Vallina-Rodriguez, Jon Crowcroft, Alessandro Finamore, Yan Grunenberger, and Konstantina Papagiannaki. 2013. When Assistance Becomes Dependence: Characterizing the Costs and Inefficiencies of A-GPS. *ACM SIGMOBILE Mobile Computing and*

Communications Review 17, 4 (2013), 3–14.

- [50] Javier Ventura-Traveset, Carlos López de Echazarreta, Juan-Pedro Lam, and Didier Flament. 2015. An Introduction to EGNOS: The European Geostationary Navigation Overlay System. *GALILEO Positioning Technology* (2015), 323–358.
- [51] Yutaro Wada, Li-Ta Hsu, Yanlei Gu, and Shunsuke Kamijo. 2017. Optimization of 3D Building Models by GPS Measurements. *GPS solutions* 21 (2017), 65–78.
- [52] Hai Wang, Baoshen Guo, Shuai Wang, Tian He, and Desheng Zhang. 2021. CSMC: Cellular Signal Map Construction via Mobile Crowdsensing. *Proc. of the ACM IMWUT* 5, 4 (2021), 1–22.
- [53] Xuanzhi Wang, Kai Niu, Anlan Yu, Jie Xiong, Zhiyun Yao, Junzhe Wang, Wenwei Li, and Daqing Zhang. 2023. WiMeasure: Millimeter-level Object Size Measurement with Commodity WiFi Devices. *Proc. of the ACM IMWUT* 7, 2 (2023), 1–26.
- [54] Michael G Wing, Aaron Eklund, and Loren D Kellogg. 2005. Consumer-Grade Global Positioning System (GPS) Accuracy and Reliability. *Journal of forestry* 103, 4 (2005), 169–173.
- [55] Chenhao Wu, Xuan Huang, Jun Huang, and Guoliang Xing. 2023. Enabling Ubiquitous WiFi Sensing with Beamforming Reports. In *Proc. of ACM SIGCOMM*. 20–32.
- [56] Dan Wu, Daqing Zhang, Chenren Xu, Yasha Wang, and Hao Wang. 2016. WiDir: Walking direction estimation using wireless signals. In *Proc. of ACM UbiComp*. 351–362.
- [57] Xiangou Wu, Panlong Yang, Shaojie Tang, Xiao Zheng, and Yan Xiong. 2015. Privacy Preserving RSS map generation for a crowdsensing network. *IEEE Wireless Communications* 22, 4 (2015), 42–48.
- [58] Farzaneh Zangenehnejad and Yang Gao. 2021. GNSS Smartphones Positioning: Advances, Challenges, Opportunities, and Future Perspectives. *Satellite navigation* 2 (2021), 1–23.
- [59] Xiaopeng Zhao, Zhenlin An, Qingrui Pan, and Lei Yang. 2023. NeRF2: Neural Radio-Frequency Radiance Fields. In *Proc. of ACM MobiCom*. 1–15.