RF-Vision: Object Characterization using Radio Frequency Propagation in Wireless Digital Twin

Sunday Amatare, Wentao Gao, Mohammad Hasibur Rahman, Aavash Kharel, Raul Shakya, Xiaojun Shang, and Debashri Roy
Department of Computer Science and Engineering, The University of Texas at Arlington
Emails: {saa3326, wxg8464, axk8168, rxs6339, mhr9808}@mavs.uta.edu, {xiaojun.shang, debashri.roy}@uta.edu

***Abstract*—In today's rapidly evolving technological landscape, accurate object characterization is crucial for a wide range of applications, from autonomous systems to smart environments and security. Simultaneously, the growing concern for privacy necessitates innovative approaches that can characterize objects without compromising sensitive visual information. In this paper, we introduce a novel approach for object characterization using Radio Frequency (RF) propagation map generation through ray-tracing within a Digital Twin (DT) framework. We outline a systematic pipeline for leveraging NVIDIA's Sionna Ray-Tracing tool to generate DT propagation maps created in Blender for indoor environments. Using these propagation maps, we propose a machine learning-based approach to facilitate object characterization. Our results demonstrate the feasibility of object characterization through strategic scene configuration using a small dataset that leverages RF maps within DTs. This paper provides valuable insights into the potential of our framework as a reliable and more efficient method for object characterization, offering a promising alternative to traditional vision-based techniques in scenarios where privacy concerns or environmental constraints limit the use of conventional imaging methods.**

***Index Terms*—Digital twin, RF propagation, Ray-tracing, Object characterization.**

I. Introduction

Accurate localization of objects within an environment is essential in fields such as robotics and automation, augmented and virtual reality, smart homes and IoT, security and surveillance, healthcare, transportation, industrial manufacturing, agriculture, and environmental monitoring. This is crucial as it enables systems to precisely identify the position of objects, leading to enhanced efficiency, safety, and responsiveness. To achieve a comprehensive understanding of these objects, we must focus not only on determining their locations but also on accurately characterizing their diverse properties. This process is referred to as *object characterization* [1]. Also, as one of the fundamental problems in computer vision, object characterization provides valuable insights for the semantic understanding of images and is connected to numerous applications, including image classification [2], human behavior analysis [3], face recognition [4], and autonomous driving [5].

Camera Image-based Object Characterization. Machine learning (ML) has emerged as a powerful tool for precise, image-based object localization and characterization, significantly enhancing the capabilities of computer vision systems. By leveraging large camera image datasets and sophisticated algorithms such as convolutional neural networks (CNN), region-based CNN (R-CNN), Faster R-CNN [6], and you only look once (YOLO) [7], [8], [9], these systems can continuously learn from examples. Overtime, this improves their accuracy, making them particularly effective in dynamic and complex environments. In addition, the advancements in GPUs and computing technology have greatly accelerated the processing and computation of large datasets for object characterization tasks [10]. This enhanced processing power enables more complex models to be trained and deployed quickly, allowing real-time analysis and improved performance in applications such as autonomous systems and image recognition.

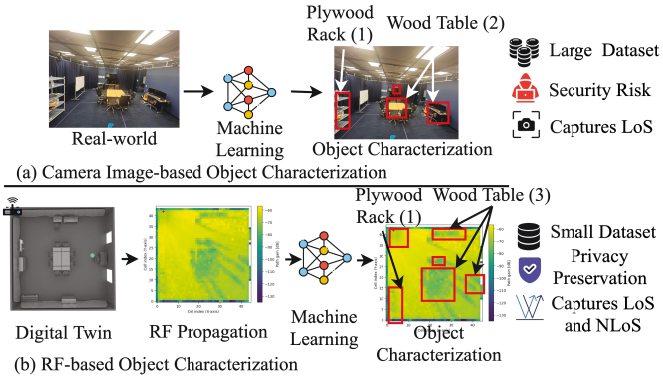


Fig. 1. Comparison of camera image-based and RF-Vision-based object characterization. We show that RF-Vision is privacy preserving and more efficient. It is shown that RF-Vision is able to detect 3 wood tables in the room while the camera image based approach is only able to detect 2 of them due to non line of sight regions.

Concerns with Image-based Object Characterization. Camera image-based datasets for object characterization present several challenges. Training ML models requires extensive, labeled datasets, which can be resource-intensive and vulnerable to security risks and adversarial attacks [11]. Furthermore, camera images often divulge private information, raising significant privacy concerns. Additionally, image-based methods are limited in capturing objects in non-line-of-sight (NLoS) conditions. These limitations necessitate alternative approaches for robust, privacy-preserving, and comprehensive object characterization.

Digital Twin (DT)-based Object Characterization. In response to previous challenges, researchers are now exploring the use of DTs to enhance object localization within virtual environments [12], [13]. Standardization bodies such as the 3rd Generation Partnership Project (3GPP) and the International Telecommunication Union Telecommunication Standardiza-

ICC 2025 - IEEE International Conference on Communications | 979-8-3315-0521-9/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICC52391.2025.11160847

Authorized licensed use limited to: University of Texas at Arlington. Downloaded on November 01,2025 at 19:06:24 UTC from IEEE Xplore. Restrictions apply.

tion Sector (ITU-T) have recognized the importance of DTs [14]. Furthermore, DT technology and RF propagation have proven to be effective techniques for selectively capturing and analyzing visual and spatial data from the surrounding environment, enabling enhanced privacy preservation [15].

Novel Contributions. Building upon these advancements in DT technology, we propose RF-Vision, an innovative framework that combines RF propagation and ML within a DT to locate and characterize objects in a scene. RF-Vision accomplishes this with minimal training data, preserves privacy, and performs effectively in detecting line of sight (LoS) objects as well as capturing overall scene details. As shown in Fig. 1, RF-Vision equips the scene with RF devices that enable ray-tracing to localize objects in real-time and generate maps for object characterization. This novel application of RF propagation for object characterization involves three key steps: (a) creating a DT to simulate real-world scenarios, (b) generating an RF propagation model within this DT, and (c) developing an ML approach that leverages the propagation characteristics of the DT to characterize objects. Specifically, various advantages of RF-Vision includes: (a) privacy preservation, (b) characterize objects in LoS regions and capture overall scene details, and (c) training on smaller dataset, as focused in Fig. 1. Formally, this paper's contributions are:

1. We propose a high-fidelity DT creation method that accurately models indoor environments by extracting real-world features with open-source Blender tool.

2. Privacy preservation through ray-tracing. We propose a methodology for precisely configuring scenes and generating propagation characteristics with transmitters placed at various positions within the environment. This approach utilizes ray-tracing (RT) on the DT of each scene, leveraging NVIDIA's Sionna RT-based software to accurately localize objects and create coverage maps for each transmitter position across all scenes.

3. Overall scene coverage. We propose an ML model that uses RT-generated coverage maps to accurately characterize LoS objects and capture overall scene details. This model effectively learns the feature characteristics of objects and the contextual information derived from these coverage maps to enable precise detection.

4. Smaller Dataset. We present a $10MB$ dataset of first-of-its-kind (to the best of our knowledge) RT-based DT indoor scenarios for object characterization. The detailed process of generating this dataset is thoroughly documented in this paper. We release our codebase and generated dataset for broader community use in [16], facilitating reproducibility and further research exploration.

II. Related Works and Motivation

Generic object detection involves locating and classifying objects within an image, marking them with rectangular bounding boxes to indicate confidence levels for each detection. The detection frameworks are typically divided into two main categories: region-based methods and classification- or regression-based methods. Girshick *et al.* [17] apply high-capacity CNNs to bottom-up region proposals to localize and segment objects. They use a simple bounding-box regression stage to improve localization performance, utilizing a training dataset of $395,918$ samples. Zhang *et al.* [18] present an enhanced Faster R-CNN model to detect healthy tomato leaves and identify four common diseases. Their approach utilizes the k-means algorithm to cluster bounding boxes, applied to a dataset of $4,178$ images, achieving a 2.71% improvement in recognition accuracy over the original Faster R-CNN model. Similarly, Feroz *et al.* [19] employ single-shot detector (SSD) and YOLO models, trained using a common objects in context (COCO) dataset containing $330,000$ images, to improve real-time object detection and recognition from webcam video, achieving an accuracy range of approximately $63-90\%$. Qin *et al.* [20] introduce an integrated framework that improves outcomes by leveraging synergistic information from multiple jointly trained CNNs, using a dataset of $53,000$ samples collected from online social networks.

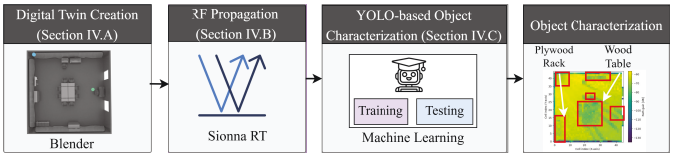


Fig. 2. The overall framework and working principle of RF-Vision.

Motivation for designing RF-Vision: All the cited work on object localization and characterization relies solely on image dataset and ML models that require large training samples, which are susceptible to security risks and adversarial attacks. Furthermore, these datasets are captured using cameras that cannot detect NLoS objects within a scene. In this paper, we propose an innovative framework, RF-Vision, for object characterization, which is designed to preserve privacy and function effectively with small sample sizes by combining RF ray-tracing with ML within a DT environment.

III. System Model and Problem Formulation

A. Problem Formulation

We denote the input scenario as S with O objects. The object characterization function f_o is defined as: $f_o : \mathrm{S} \rightarrow \{(b_i, c_i, s_i)\}_{i=1}^{N}$, where N is the number of detected objects in the scenario S (this number can be different that actual number of objects present, i.e., O), b_i represents the bounding box for the ith object, c_i is the class label for the ith object, and s_i is the confidence score for the ith object. Each bounding box b_i is represented as: $b_i = (x_i, y_i, w_i, h_i)$, where: (x_i, y_i) are the coordinates of the top-left corner of the box, w_i is the width of the box, h_i is the height of the box. The class label c_i is an integer representing the object category: $c_i \in \{1, 2, ..., K\}$, where K is the total number of object classes. The confidence score s_i is a real number between 0 and 1: $s_i \in [0, 1]$.

B. System Architecture in RF-Vision

Our framework is illustrated in Fig. 2 and is organized into three main modules as follows:

Authorized licensed use limited to: University of Texas at Arlington. Downloaded on November 01,2025 at 19:06:24 UTC from IEEE Xplore. Restrictions apply.

- **Digital Twin (DT) Creation** *(Module 1)*: We create a digital replica of each real-world scene by directly extracting its features and place RF devices at various locations within these scenes (details in Sec. IV-A).
- **Propagation Modeling** *(Module 2)*: We strategically place a transmitter in the ceiling area and a receiver at the center of each scene. We then propagate a signal to localize the objects within these scenes (details in Sec. IV-B).
- **ML-based Object Characterization** *(Module 3)*: We implement an ML model that leverages propagation characteristics and scene maps to characterize objects in each scene (details in Sec. IV-C).

IV. RF-Vision Framework

In this section, we discuss different steps and components of our proposed framework.

A. Module 1: Digital Twin (DT) Creation

In the RF-Vision, we take into account aspects related to map accuracy and characteristics of RF propagation. While incorporating several specific key metrics into RF-Vision, our baseline can be adapted in the future to accommodate different environmental configurations. We initialize the created DT as $\mathcal{E} = f(\mathrm{S}, \rho) = f(\texttt{map}, \mathrm{O}, \rho)$. Here, `map`, O, and ρ denote the imported Blender [21] map, present structures or objects for the twin $\mathcal{E}$, and number of allowed reflections for the created twin, respectively.

B. Module 2: Propagation Modeling

We use the open source Sionna RT [15] tool to generate the propagation characteristics of the created DT $\mathcal{E}$ by employing RF ray-tracing. For a given transmitter `TX`, a propagation map is a rectangular surface with arbitrary orientation subdivided into rectangular cells of size $|\mathcal{C}|$. A parameter η controls the granularity of the map. The propagation map associates with every cell $(\mathcal{C}_i, \mathcal{C}_j)$. The channel gain $\mathcal{G}_{i,j}$ for each cell of DT $\mathcal{E}$ is denoted as:

$$\mathcal{G}_{i,j} = \frac{1}{|\mathcal{C}|} \int_{(\mathcal{C}_i, \mathcal{C}_j)} |h(x,y)|^2 dx.dy,$$

where $h(s)$ is the amplitude of the path coefficients at position (x, y) within $(\mathcal{C}_i, \mathcal{C}_j)$ [22].

C. Module 3: ML-based Object Characterization

After completing propagation modeling for all digitally created scenes and generating the corresponding propagation maps, we use ML-based object localization and characterization in RF-Vision. The object characterization problem f_θ is solved by training a parameterized ML model f_θ, where $f_\theta : \mathbb{R}^{|\mathcal{C}|} \rightarrow \mathbb{R}^6$, where $|\mathcal{C}|$ represents the domain of the scenario S generated by propagation modeling and 6 represents the generated values corresponding to bounding box (4 values), class label (1 value), and confidence score (1 value), details in Sec. III-A.

For training f_θ we use a multi-task loss function $\mathcal{L}$ that combines multiple components: $\mathcal{L} = \lambda_1 \mathcal{L}_{\text{box}} + \lambda_2 \mathcal{L}_{\text{dfl}} + \lambda_3 \mathcal{L}_{\text{cls}}$, where $\mathcal{L}_{\text{box}}$ is the localization loss for bounding box regression, $\mathcal{L}_{\text{dfl}}$ is distribution focal loss for imbalanced class prediction, $\mathcal{L}_{\text{cls}}$ is the classification loss for class prediction, λ_1, λ_2, and λ_3 are weighting factors [23]. The ML model f_θ is trained over multiple epochs to minimize the multi-task loss, formally, $\hat{\theta} = \arg\min_\theta \mathcal{L}(f_\theta(I), Y)$, where θ represents the parameters of the ML model, $\mathcal{L}$ is a loss function, Y represents the ground truth bounding box annotations. Overall, we generate the object characterization function as $f_o(.) = f_{\hat{\theta}}(.)$.

During inference, for a new scenarios S_{new}, the bounding boxes are predicted as: $(\hat{b}_{\text{new}}, \hat{c}_{\text{new}}, \hat{s}_{\text{new}} = f_{\hat{\theta}}(\mathrm{S}_{new})$, where $\hat{b}_{\text{new}}, \hat{c}_{\text{new}}, \hat{s}_{\text{new}}$ are the new predicted bounding box, class label and confidence scores, respectively.

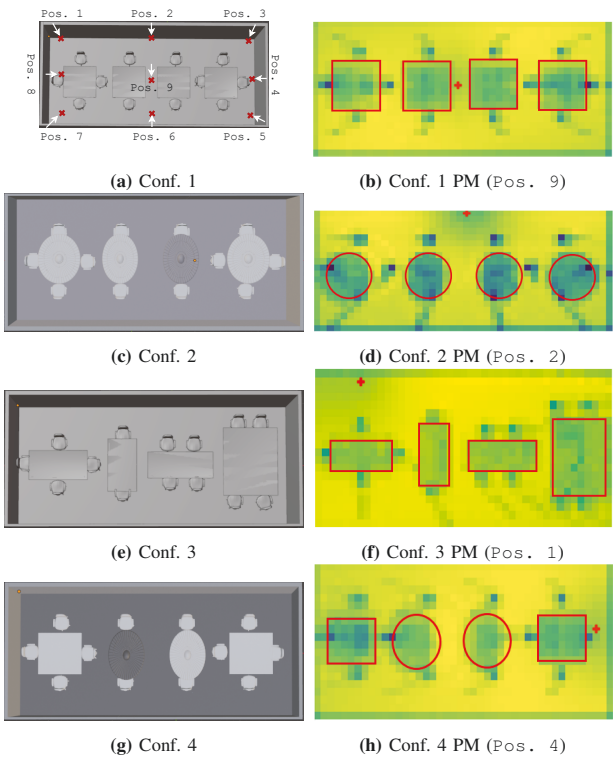
V. Experiments

A. Dataset Generation

To validate our experiments we curate a dataset featuring realistic digital twins (DTs) and their corresponding RF propagation maps of indoor scenarios with objects of various shapes and four materials. We meticulously design the DT scenes using Blender [21] and generate the corresponding propagation characteristics and maps with Sionna RT [15]. Overall, we generate 324 RF propagation maps from 36 scene configurations featuring 2 distinct floor plans and 9 different transmitter positions. For each configuration, the transmitter is placed 13 ft above the floor, i.e., at the ceiling of the room: (a) `Position 1`: top-left, (b) `Position 2`: top-middle, (c) `Position 3`: top-right, (d) `Position 4`: middle-right, (e) `Position 5`: bottom-right, (f) `Position 6`: bottom-middle, (g) `Position 7`: bottom-left, (h) `Position 8`: bottom-middle, and (i) `Position 9`: center, as marked in Fig. 3(a) and Fig. 4(a). Moreover, each configuration has four different material properties: *wood*, *metal*, *glass*, and *concrete* to represent a table object in the scene. We model these four materials as: *ITU-wood*, *ITU-metal*, *ITU-glass*, and *ITU-concrete* radio materials, respectively, in Sionna RT. Each configuration also features a unique table shape, with object positions rearranged in each subsequent setup. For example, if one configuration has tables arranged from top-left corner clockwise as wood, metal, glass, and concrete, the next configuration of the same table shape will rearrange them as metal, glass, concrete, and wood, and so forth. We generate a propagation map for each configuration.

1) Digital Twin (DT) Creation: We use Blender to create a DT of all scene configurations, digitally replicating real-world settings, including scene objects with various radio material properties, as well as the walls and ground plane of each scene. After carefully creating the scenes, we export each one from Blender in `Mitsuba 3 .xml` file format, which handles scene rendering, and import it into Sionna RT for propagation modeling. The details of DT creation is presented in our previous work [24], [25]. Snapshots of selected scene configurations of generated DTs are presented in the first columns of Figs. 3 and 4.

Authorized licensed use limited to: University of Texas at Arlington. Downloaded on November 01,2025 at 19:06:24 UTC from IEEE Xplore. Restrictions apply.



(a) Conf. 1 **(b)** Conf. 1 PM (Pos. 9)
(c) Conf. 2 **(d)** Conf. 2 PM (Pos. 2)
(e) Conf. 3 **(f)** Conf. 3 PM (Pos. 1)
(g) Conf. 4 **(h)** Conf. 4 PM (Pos. 4)

Fig. 3. A snapshot of selected scenes and their propagation maps is shown, where the first column displays the digitally created scenes and the second column shows their corresponding propagation maps (PM). Objects in the propagation maps are marked to indicate localization following propagation modeling. The red cross on the PMs indicates different transmitter positions.

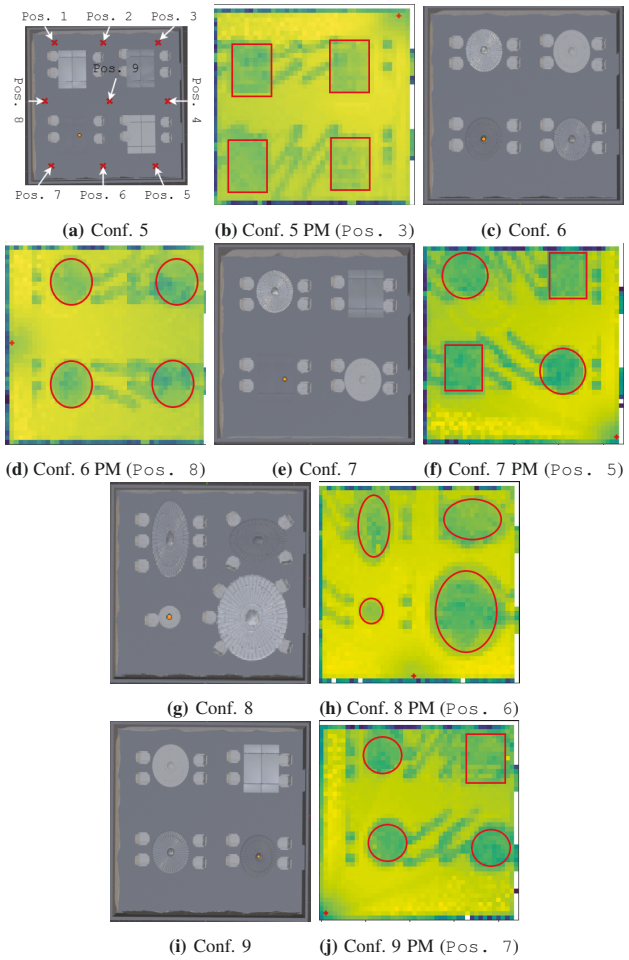
2) Propagation Modeling: Following the creation of the DT of each scene and its import into Sionna RT [15], the transmitter and receiver antennas are configured as 8×2 `tr38901` [26] with dual polarization and 1×1 *diapole* planner array, respectively. The transmitter in the scene operates at 2.4 GHz with power of 44 dBm. Using the compute path function in Sionna, we calculate compute paths to generate the propagation map, setting `max-depth = 5` and `num-samples = 5000`. Snapshots of the corresponding propagation maps with different transmitter positions are presented in the second columns of Figs. 3 and 4.

Remark 1. *The generated propagation maps capture various object properties within a scene while preserving privacy by omitting detailed visual information from camera images (see Figs. 3 and 4, validates Contribution 2)*

Generated Dataset. By combining various configurations and positions, we generate 324 propagation maps for indoor scenarios involving four distinct materials: (a) wood, (b) metal, (c) glass, and (d) concrete, resulting in a compact dataset of only 10 MB (*validates Contribution 4*).

B. YOLO-based Object Characterization

Machine Learning (ML) Model. In recent years, complex ML models such as YOLO [7], [8] and R-CNN [6] have gained significant popularity for object detection and



(a) Conf. 5 **(b)** Conf. 5 PM (Pos. 3) **(c)** Conf. 6
(d) Conf. 6 PM (Pos. 8) **(e)** Conf. 7 **(f)** Conf. 7 PM (Pos. 5)
(g) Conf. 8 **(h)** Conf. 8 PM (Pos. 6)
(i) Conf. 9 **(j)** Conf. 9 PM (Pos. 7)

Fig. 4. Selected scene configurations and their propagation maps are shown. Various transmitter's positions are shown in (a). In the second column, green-marked regions indicate weak signal strength due to the presence of an object, while yellow regions represent strong signal intensity where no object is present.

characterization in vision datasets. In RF-Vision, we leverage these state-of-the-art methods to effectively characterize various objects from propagation maps. In Figs. 3 and 4, we observe that due to the privacy-preserving nature of the data, the objects lack distinct edges. Hence, through extensive experimentation, we have selected a lightweight variant of the widely recognized YOLO model, namely *YOLO v11 nano* or *YOLO v11n*, as our candidate model [27] as small datasets generally perform better with models with smaller number of parameters and simpler architecture. We specifically fine-tune YOLO v11n with the generated dataset, producing a version with 238 fused layers, $2,582,932$ parameters, 0 gradients, and 6.3 Giga Floating Point Operations per Second (GFLOPS).

Experimental Platform and Performance Metrics. We employ Ultralytics with PyTorch, providing a robust and versatile platform. We run our experiments on *google colab* using an NVIDIA A100 GPU. We use standard performance metrics, including recall, precision, and mAP50, to evaluate the effectiveness of object characterization, where mAP50 represents the mean average precision calculated at an intersection over

Authorized licensed use limited to: University of Texas at Arlington. Downloaded on November 01,2025 at 19:06:24 UTC from IEEE Xplore. Restrictions apply.

union threshold of 0.5.

Training Parameters. In our experiments, YOLO v11n model is fine-tuned for object characterization with 100 epochs, an early stopping patience of 100, batch size of 16, and 640×640 resolution for the propagation map images. We use stochastic gradient descent optimizer, with an initial learning rate of 0.01, momentum of 0.937, and weight decay of 0.0005. Mixed precision was enabled, using 8 workers, and a fixed seed ensured reproducibility.

Data Augmentation. We also perform data augmentations including horizontal flipping (fliplr=0.5), hue, saturation, and value adjustments ($hsv_h = 0.015, hsv_s = 0.7, hsv_v = 0.4$), translation (0.1), scaling (0.5), and RandAugment [28], aiming to increase the robustness of the model. The augmentation process expanded the dataset by a factor of 2.27, resulting in a final size of approximately 27 MB, which remains relatively compact for computational efficiency.

Training. The augmented dataset is divided into training, validation, and test sets with an 80/10/10 split. We design a series of experiments, where we perform training based on the data group with respect to different transmitter locations. The details of various experiments are shown in Table I: top (`Position 2`), right (`Position 4`), bottom (`Position 6`), left (`Position 8`), center (`Position 9`), and combined. From the Table I, we observe 19.2%, 27.6%, 28.9%, 24.7%, and 89.0% for transmitter positions `2`, `4`, `6`, `8`, and `9`, respectively, on correctly detecting different materials. Based on our observations, RF-Vision demonstrates competitive performance in detecting object materials when the ray-tracing is conducted with a transmitter positioned centrally within an indoor scenario, as the varied performance across different transmitter locations leads to reduced overall performance when trained on combined data. The training plots with the transmitter in the center are shown in Fig. 5. The training trend is shown through localization loss (`box_loss` denoted as $\mathcal{L}_{\text{box}}$ in Sec. IV-C), classification loss (`cls_loss` denoted as $\mathcal{L}_{\text{cls}}$ in Sec. IV-C), and distribution focal loss (`dfl_loss` denoted as $\mathcal{L}_{\text{dfl}}$ in Sec. IV-C). The performance is captured through precision, recall, and AP50 metrics for both training and validation.

Observation 1. *We observe that RF-Vision yields competitive training performance for object characterization when the transmitter is placed in the center (`Position 9`) of a scene.*

Transmitter Pos.	Precision	Recall	mAP50	Inference Time
Top (`Pos. 2`)	19.2%	85.3%	62.2%	15.6 ms
Right (`Pos. 4`)	27.6%	94.5%	51.7%	11.2 ms
Bottom (`Pos. 6`)	28.9%	91.9%	42.7%	4.6 ms
Left (`Pos. 8`)	24.7%	80.8%	44.8%	6.2 ms
All (combined)	29.8%	83.5%	44.7%	2.8 ms
Center (`Pos. 9`)	**89.0%**	**95.5%**	**98.5%**	**2.2 ms**

Table I: Ablation study on datasets with distinctive transmitter positions.

YOLO-based Inference. We show various samples of inference while using the trained model in Fig 6. Overall, it achieves high confidence on predicting the objects shapes and materials (wood, metal, glass, and concrete) from the RF propagation maps.

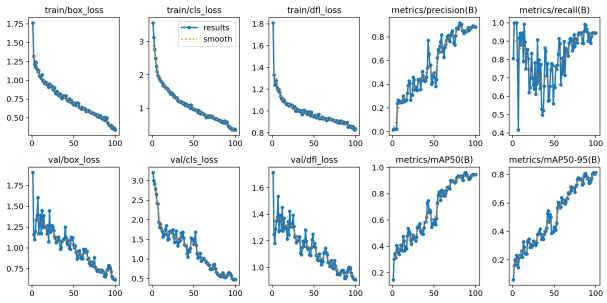


Fig. 5. Training and validation plots while trained on the dataset containing transmitter in the middle (`Pos. 9`). The `box_loss`, `cls_loss` and `dfl_loss` correspond to $\mathcal{L}_{\text{box}}$, $\mathcal{L}_{\text{cls}}$, and $\mathcal{L}_{\text{dfl}}$, defined in Sec. IV-C. The performance is captured through precision, recall, and AP50 metrics for both training and validation.

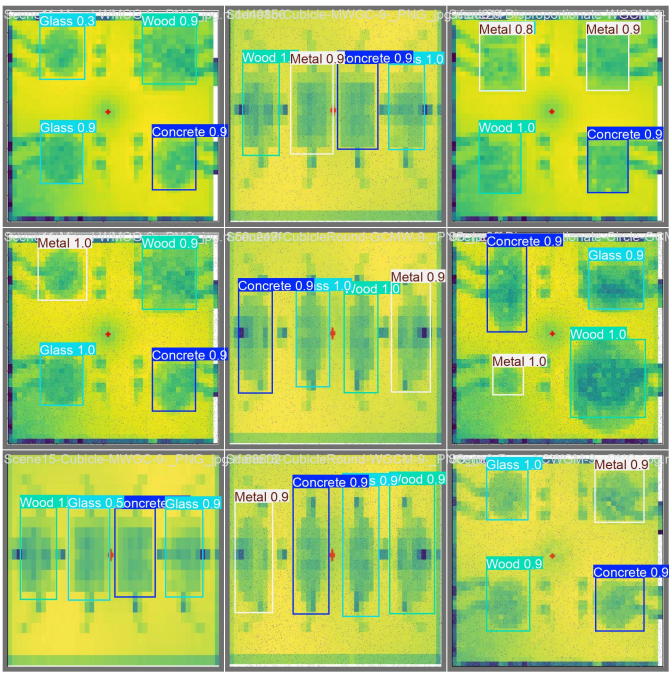


Fig. 6. Samples of inferences instances from the YOLO v11n model, which was trained on propagation maps generated with the transmitter positioned centrally (`Pos. 9`). The trained model is able to detect the shapes and material properties of the objects with high confidence.

Observation 2. *We observe that the YOLO-based detection is able to characterize objects when the transmitter is placed at the center of a scene (see Fig. 6, validates Contribution 3).*

• **Comparison with State-of-the-art.** We compare the inference time per frame with the state-of-the-art [29], [30]. As shown in Table II, RF-Vision outperforms other methods in inference time and uniquely utilizes privacy-preserving propagation maps.

• **Co-existence of RF-Vision with Existing Communication Infrastructure.** In typical indoor environments, RF

Authorized licensed use limited to: University of Texas at Arlington. Downloaded on November 01,2025 at 19:06:24 UTC from IEEE Xplore. Restrictions apply.

Table II: Comparison with state-of-the-art w.r.t. inference time.

Paper	Model	CPU/GPU	Inference Time	Exposed Visual Info.
Wang *et al.* [29]	YOLOv7	RTX 4090	5.4 ms	Yes
Zhou [30]	YOLO-NL	Ryzen 7 5700X	10.4 ms	Yes
RF-Vision	YOLOv11n	A100	2.2 ms	No

transmitters such as WiFi access points operating in the 2.4 GHz frequency band (the same band utilized in our experiments) are commonly installed on ceilings. This existing infrastructure can be used to create DTs of the environment and characterize objects within the scene. Ceiling-mounted deployments offer several advantages: they provide an elevated, bird's-eye perspective of the environment, significantly reducing the occurrence of NLoS conditions; the existing configuration can be directly utilized without additional hardware installation; and the diffraction properties inherent in RF ray-tracing [15], further minimize NLoS occurrences for objects within the scenario. This synergy between RF-Vision and existing communication infrastructure not only enhances the system's ability to characterize objects but also offers a cost-effective and non-intrusive solution for indoor environment mapping and analysis, closely mirroring practical scenarios and increasing the applicability of our findings.

VI. Conclusion

This paper presents RF-Vision, an innovative system that combines RF propagation with ML for object characterization in indoor environments. Our approach involves creating a DT of real-world scenarios, simulating propagation characteristics to generate RF maps, and using these maps to train a ML model for object characterization. Through extensive experimental validation, we demonstrate that RF-Vision achieves accurate object characterization with minimal training data generated from transmitters of relative center positions. Future work will aim to extend this system by three aspects. First, generate more high-quality data with relevant features to further fine-tune the ML model for robustness and generalization on indoor environments. Second, explore solid pipeline(s) in outdoor settings to assess its adaptability. Lastly, evaluate the privacy-preserving performance with specific quantitative metrics.

References

[1] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 9, pp. 1627–1645, 2009.

[2] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proceedings of the 22nd ACM international conference on Multimedia*, 2014, pp. 675–678.

[3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, 2012.

[4] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2d pose estimation using part affinity fields," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7291–7299.

[5] X. Chen, H. Ma, J. Wan, B. Li, and T. Xia, "Multi-view 3d object detection network for autonomous driving," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2017, pp. 1907–1915.

[6] S. Bhatlawande, S. Shilaskar, M. Agrawal, V. Ashtekar, M. Badade, S. Belote, and J. Madake, "Study of object detection with faster rcnn," in *2022 2nd International Conference on Intelligent Technologies (CONIT)*, 2022, pp. 1–6.

[7] J. Redmon, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[8] G. Jocher, "Yolov5," https://github.com/ultralytics/yolov5, 2020, accessed: 2024-11-07.

[9] R. Khanam and M. Hussain, "Yolov11: An overview of the key architectural enhancements," *arXiv preprint arXiv:2410.17725*, 2024.

[10] Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye, "Object detection in 20 years: A survey," *Proceedings of the IEEE*, vol. 111, no. 3, pp. 257–276, 2023.

[11] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," *arXiv preprint arXiv:1412.6572*, 2014.

[12] J. Morais and A. Alkhateeb, "Localization in digital twin mimo networks: A case for massive fingerprinting," *arXiv preprint arXiv:2403.09614*, 2024.

[13] S. Amatare, G. Singh, A. Kharel, and D. Roy, "Real-time localization of objects using radio frequency propagation in digital twin," *Available at SSRN 4937841*, 2024.

[14] *Digital twin Network - Requirements and architecture*, International Telecommunication Union Telecommunication Standardization Sector, recommendation ITU-T Y.3090, 2022. [Online]. Available: https://www.itu.int/rec/T-REC-Y.3090

[15] J. Hoydis, F. A. Aoudia, S. Cammerer, M. Nimier-David, N. Binder, G. Marcus, and A. Keller, "Sionna rt: Differentiable ray tracing for radio propagation modeling," *arXiv preprint arXiv:2303.11103*, 2023.

[16] https://github.com/TWIST-Lab/RF-Vision.

[17] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-based convolutional networks for accurate object detection and segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 1, pp. 142–158, 2015.

[18] Y. Zhang, C. Song, and D. Zhang, "Deep learning-based object detection improvement for tomato disease," *IEEE Access*, vol. 8, pp. 56 607–56 614, 2020.

[19] M. A. Feroz, M. Sultana, M. R. Hasan, A. Sarker, P. Chakraborty, and T. Choudhury, "Object detection and classification from a real-time video using ssd and yolo models," in *Computational Intelligence in Pattern Recognition: Proceedings of CIPR 2021*. Springer, 2022, pp. 37–47.

[20] H. Qin, J. Yan, X. Li, and X. Hu, "Joint training of cascaded cnn for face detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 3456–3465.

[21] B. O. Community, *Blender - a 3D modelling and rendering package*, Blender Foundation, Stichting Blender Foundation, Amsterdam, 2018. [Online]. Available: http://www.blender.org

[22] NVIDIA Research, "Sionna ray tracing documentation," https://nvlabs.github.io/sionna/api/rt.html, 2024, accessed: [11/05/2024].

[23] Z.-Q. Zhao, P. Zheng, S.-t. Xu, and X. Wu, "Object detection with deep learning: A review," *IEEE transactions on neural networks and learning systems*, vol. 30, no. 11, pp. 3212–3232, 2019.

[24] S. Amatare, M. Samson, and D. Roy, "Testbed design for robot navigation through differential ray tracing," in *2024 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*, 2024, pp. 173–174.

[25] S. Amatare, G. Singh, M. Samson, and D. Roy, "RagNAR: Ray-tracing based Navigation for Autonomous Robot in Unstructured Environment," in *IEEE Global Communications Conference*, December 2024.

[26] N. Inc., "Ray tracing," 2024, last accessed 6 April 2024. [Online]. Available: https://nvlabs.github.io/sionna/api/rt.html

[27] L. Brigato and L. Iocchi, "A close look at deep learning with small data," in *2020 25th International Conference on Pattern Recognition (ICPR)*, 2021, pp. 2490–2497.

[28] E. D. Cubuk, B. Zoph, J. Shlens, and Q. V. Le, "Randaugment: Practical automated data augmentation with a reduced search space," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2020, pp. 702–703.

[29] M. Wang, H. Sun, J. Shi, X. Liu, X. Cao, L. Zhang, and B. Zhang, "Q-yolo: Efficient inference for real-time object detection," in *Asian Conference on Pattern Recognition*. Springer, 2023, pp. 307–321.

[30] Y. Zhou, "A yolo-nl object detector for real-time detection," *Expert Systems with Applications*, vol. 238, p. 122256, 2024.

Authorized licensed use limited to: University of Texas at Arlington. Downloaded on November 01,2025 at 19:06:24 UTC from IEEE Xplore. Restrictions apply.