# *D3PicoNet*: Enabling Fast and Accurate Indoor D-Band Millimeter-Wave Picocell Deployment

Hem Regmi; Sanjib Sur
*Department of Computer Science and Engineering, University of South Carolina, Columbia, SC, USA*
hregmi@email.sc.edu; sur@cse.sc.edu

***Abstract*—We propose *D3PicoNet*, which allows network deployers to quickly complete realistic indoor site surveys at D-band (mmWave) frequency. *D3PicoNet* models the mmWave reflection profile of a given environment, considering the primary reflecting objects. It then utilizes this model to identify places that optimize the efficiency of the reflectors. *D3PicoNet* understands an environment and deploys D-band picocells at such locations that picocells provide coverage with Non-Line-of-Sight (NLoS) paths when Line-of-Sight (LoS) is obstructed. The core module of *D3PicoNet* is a deep learning network that learns the relationship between the visual depth images to the mmWave signal reflection profiles and can accurately predict signal reflection profiles at any unobserved location, which allows *D3PicoNet* to find the best deployment locations maximizing the coverage and data rate with a minimum number of picocells in an environment. We implement and evaluate *D3PicoNet* on two buildings with multiple indoor environments. *D3PicoNet* can adapt to new environments, allowing it to be used in other indoor environments with minimal adjustments.**

***Index Terms*—Millimeter-Wave; NextG Networks; Convolutional Neural Network; Transfer Learning.**

## I. Introduction

Next-generation networks with millimeter-wave (mmWave) technology are set to revolutionize the existing wireless infrastructure, with data rates of 10s of Gigabits per second (Gbps) and milliseconds (ms) latency, enabling new applications in multiple sectors [1]. The cost of obtaining hardware at mmWave of a very high frequency, especially the D-Band (110 GHz to 170 GHz) being reasonable, makes this the ideal period to look into the difficulties and potentials of setting up networks in those frequency ranges. This will enable future network architectures beyond existing 5G, where an extremely high density of low-cost, short-range access points, called "*picocells*," can be deployed on indoor and outdoor structures, and they can operate on hundreds of frequency bands. Besides, the mmWave frequency above 100 GHz allows wider contiguous bandwidth and makes the antenna size extremely small, facilitating a higher data rate and lower latency.

Picocell communicates with the clients using very narrow beams, limiting the signals to Line-Of-Sight (LoS) paths and allowing only a few reflections in Non-LOS (NLoS) paths. Picocells often rely on NLoS paths to establish a link to the client because LoS links are vulnerable to obstructions and blockages due to narrow beam width. Moreover, not all parts of the environment can provide a strong NLoS path because of the specularity and weak reflectivity, so the D-band networks are prone to frequent outages [3][6]. We can install additional strong reflectors at required locations to facilitate more NLoS paths, although this option may not be feasible in some environments. Alternatively, we can place picocells where they can find a maximum number of NLoS paths and avoid link outages during LoS path obstructions, but finding such locations is a hard-problem. Thorough, manual site surveys collecting Signal Reflection Profiles (SRPs) at every environment location can help network deployers achieve this; however, this process is both time-consuming and costly. Wireless propagation simulations, such as Ray-tracing, to estimate SRPs may be an alternative to complete site surveys, but they are error-prone due to the unavailability of high-frequency mmWave datasets, such as D-band.

We propose *D3PicoNet* to overcome these challenges and enable robust deployment of D-band mmWave picocells for indoor environments. Inspired by prior works, such as [2], [3], that propose a picocell deployment model for lower mmWave frequencies (24 GHz), we design *D3PicoNet* specifically for the deployment of 122 GHz D-band picocells. A key idea in the model is that if two parts of the environment look similar, they will likely produce similar NLoS reflections (accounting for the distance and incident angle of the transmitted signals). By measuring NLoS reflection from one part of the environment, it is feasible to predict the reflection from the other part. Thus, the network deployers can avoid time-consuming thorough site surveys and collect a few random samples for reflection profile prediction, allowing optimal deployment. *D3PicoNet* builds on this intuition and designs a customized deep-learning framework to predict the SRP at unobserved locations within the environment. Unlike the existing works at 24 GHz, SRP measurements at 122 GHz are very noisy, with high attenuation, scattering, and diffraction, making it challenging to predict SRP at unobserved locations. *D3PicoNet* analyzes the noisy data to determine errors in ground-truth data collection and incorporates a parametric Gaussian Noise Layer into the network to compensate for the error. Adding noise layers makes the network robust and tolerant of errors.

Due to the unavailability of any open-source dataset at 122 GHz mmWave frequency, we design a custom data collection hardware by integrating an ASUS Zenfone AR smartphone with a 122 GHz mmWave transceiver [4], [5] to collect the mmWave SRPs, visual Point Cloud Data (PCD), and poses of the device. In total, we have collected over 45 GB of the dataset with $\sim$4.2 million data samples for 4 months. We have

used only ∼420K samples for training, and the rest of ∼3.8M pairs are used for testing and benchmarking all our design components. Our evaluation of a base SRP prediction model on a building shows an average SRP prediction error of 3.0 dB. But the error reaches more than 8 dB when the model is tested in different environments. However, a semantic-aware global feature augmentation to the base model can reduce the SRP prediction error to 2.2 dB with little fine-tuning (∼4 mins). *D3PicoNet* places the D-band picocells with more than 96% accuracy (an *Optimal* scheme is 100% accurate) to support data throughput up to 35 Gbps, which is more than 1.3× improvement over *Random*.

In summary: (1) We propose an SRP prediction model at the D-band mmWave frequency based on visual data and deep learning. (2) We design and evaluate picocell deployment methods for 122 GHz and present challenges and opportunities for deploying lower and higher mmWave frequencies together.

## II. Background and Motivation

Recall that the link between *picocell* and the client can be established either via LoS path or NLoS path. Even though LoS path provides a higher Signal to Noise Ratio (SNR), indoor environments are most likely to starve from such an advantage [6], [7] because of the directionality and blockage. We must consider NLoS paths along with LoS paths to increase network reliability. Surrounding strong reflectors such as metal, glass, wooden structures, *etc.*, can reflect mmWave signals, and picocell can use NLoS path by steering its beam towards one of them. Figure 1(a) shows three possible link statuses between picocell and client for an environment. The environment has few strong reflectors and comprises areas near the building elevator. In *Case 2*, the client and the picocell have LoS paths because there is no blockage. Figure 1(c) shows an example of SRP with LoS path where we get a strong signal strength of -55 dB at 1.8 m. However, in *Case 1*, the client is not in direct sight because of the wall, but the picocell can establish a link via a strong reflector on the other side of the wall. Figure 1(b) shows an example SRP with NLoS path, where picocell receives the signal strength of -69 dB (less than LoS case but enough to establish a link) with a strong reflector at 4.1 m. Unfortunately, in *Case 3*, picocell cannot find any possible NLoS paths to establish a link to the client. Figure 1(d) shows the example SRP if there are no reflecting objects under its vicinity. Thus, *Case 3* highlights the necessity of deploying picocell at the correct location based on the location of strong reflectors. To use the maximum possible NLoS paths, we should know strong reflectors and SRPs at every environment location. But measuring SRPs has the following challenges: (1) Using the Ray-tracing method [8] to estimate SRPs is error-prone due to the unavailability of path loss models at the high frequency and lack of open-source data to model them, and (2) Manual thorough site surveys to measure SRPs at all possible location is time-consuming and costly.

## III. *D3PicoNet* Design

### A. System Overview

In *D3PicoNet*, we use the learning-assisted SRP prediction to allow network deployers to deploy picocells to maximize the use of NLoS paths in the absence of LoS path. *D3PicoNet* focuses its design and deployment at D-band frequency because it offers wider, contiguous bandwidth. Deployments utilizing NLoS paths depend on strong reflectors' availability and location, and their role is even more vital at a higher frequency due to high path loss and narrow coverage. One could measure SRPs at every corner of the environment to get the correct location of strong reflectors. However, *D3PicoNet* estimates SRPs at every location with learning and a data-driven approach to eliminate costly manual site surveys. A network deployer can quickly collect visual information using an AR device, such as Google Tango [4], to create the visual PCD and collect the SRPs using a co-located mmWave transceiver [5]. Deployer can then use the *D3PicoNet*'s Deep Convolutional Neural Network (DCNN) model to map visual PCD into SRPs. After mapping, the deployer can predict SRPs at unobserved locations using visual PCD only.

Figure 2 shows the overview of the *D3PicoNet* system. *D3PicoNet* first collects the visual measurements of the environment as PCD and synchronizes with SRPs collected via co-located mmWave transceiver at random locations. *D3PicoNet* uses convolution layers to extract features from visual depth images and map those features into SRPs. The SRP prediction network then learns from thousands of samples from the environment and can predict SRPs at locations not used in training. Finally, predicted SRPs help the deployer find the best picocell locations based on the data rate requirement of the client.

### B. Data Correlation

*D3PicoNet*'s main idea is to use the visual depth images to predict SRPs at unobserved locations to reduce site surveys. We explore and analyze the correlation between visual depth images, SRPs, and visual depth images to SRPs before finding a suitable model to predict SRPs from depth images.

**Visual-to-Visual & SRP-to-SRP:** First, we explore the visual-visual and SRP-SRP similarity between environments with similar reflecting objects in their surroundings. We use the t-SNE [9] on 16K samples to extract two-dimensional features, *Feature 1* and *Feature 2*. Figures 4[a–b] show that Env A.3 and Env A.7 (marked with green arrows) map close because of their similar surrounding objects, while Env A.3 and Env A.6 (marked with red arrows) are separate due to dissimilarity. Results demonstrate that we can predict SRPs from visual depth images because visual depth images carry enough information about strong reflecting objects.

**Visual-to-SRP:** However, implementing the learning-based system to predict SRPs, there should be a strong correlation between visual depth images and SRPs. Our hypothesis is that *visually similar images produce similar mmWave signal*

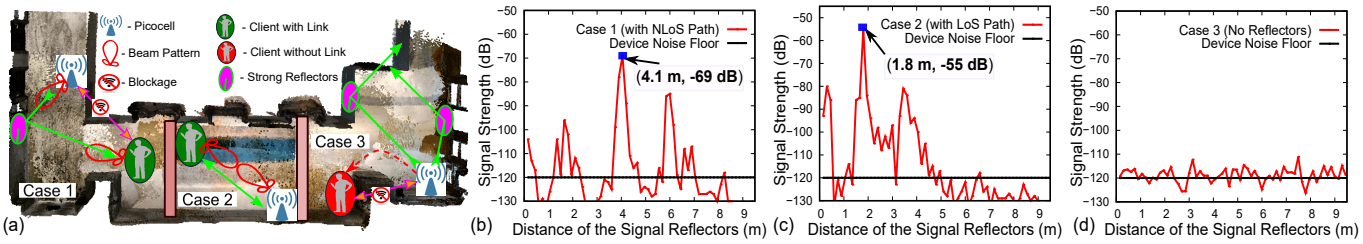


Fig. 1: (a) Illustration of picocell to the client link via NLoS path (Case 1) and LoS path (Case 2); and no connectivity (Case 3). (b–d) Examples of mmWave SRP with device noise floor for each case.

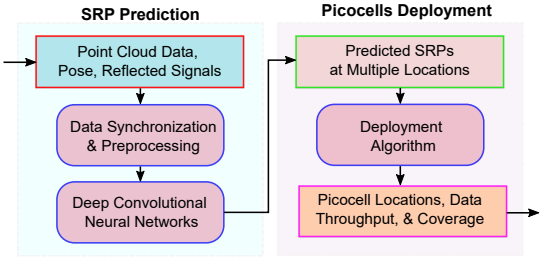


Fig. 2: System overview of *D3PicoNet*.

*reflections*. To this end, we use 320K visual depth image-SRP pairs from buildings commonly used for office spaces to find the correlation between them. We use the Structural Similarity Index Measure (SSIM) to find the similarity between the visual depth images and Mean Squared Error (MSE) between SRPs to compute the similarity between SRPs. Figure 3 shows the scatter plot of the relationship between the similarity of SRPs and the similarity of visual depth images. The non-linear relationship between data requires a complex, non-linear model.

### C. Signal Reflection Profile Prediction

*1) Deep Learning Based SRP Prediction:* Deep learning-based systems can approximate a solution from a random guess [10]. *D3PicoNet*'s learning network is similar to the approximation methods using deep learning networks. To this end, *D3PicoNet* uses convolution layers to extract the features from the visual depth image and predict the SRP measurements at the output layer. *D3PicoNet* includes the mmWave device pose on its input because SRPs are dependent on the number and location of reflectors that are reflecting. Finally, we use MSE and MAE loss combination to train the DCNN network.

**Error on Ground-Truth:** Measured SRPs have a resolution of 1 dB and may have an error on ground-truth. Even though the error is small, using it helps the DCNN model to converge faster [11]. To analyze our hypothesis that *SRP measurements are more erroneous at a higher frequency*, we collect ∼21K SRPs at both 122 GHz and 24 GHz by steering the mmWave beam towards a hard metallic reflector. Figure 6 shows that the median error is 0.77 dB at 122 GHz, ∼2× that of 24 GHz. *D3PicoNet* compensates for error on ground-truth at 122 GHz by using Gaussian Noise Layer [12]–[14].

**Data Preprocessing**: Removing redundant information and transforming it into the correct format is essential in *D3PicoNet*. Moreover, due to the unavailability of single hardware to collect both visual images and SRPs, data synchronization is also a necessary step before feeding input-output pairs into the network. We use the AR device to collect visual Point Cloud Data (PCD) co-located with the mmWave device to gather SRPs. We then use the transceiver pose, which has the measurement location and orientation from the Global PCD (GPCD), to obtain Local PCD (LPCD) (see Figure 5) with mmWave beampattern [15], [16] to remove non-reflecting points that are outside field-of-view (FoV) ($\theta$ = 60°). To get the depth images, we project the 3D points of LPCD on a 2D plane, assign the distance as depth values, and ignore reflections beyond ∼10 m because the signal strength is low. We take the inverse of depth values in Inverse Depth Image (IDI) since signal strength is inversely related to depth. Also, we mask the IDI with the beam pattern of the mmWave transceiver to encode antenna gain on DCNN input, obtain Mask Inverse Depth Image (MIDI), and pair with transceiver pose and SRPs. Finally, we follow the data preprocessing and filtering steps for each environment to generate the thousands of MIDI, transceiver pose, and SRP pairs, which will be used for training and validation of the DCNN network.

**Base DCNN Model**: In the base DCNN model, we use the depthwise and pointwise convolution layers from *MobileNetV2* [17] to make the model fast and less memory-consuming, so mobile devices can deploy the model. Convolution layers extract the 1D abstract features from the MIDI after successive convolution operations. The hierarchical nature of the convolution process allows the convolution layers to capture the relationship between local and global points in MIDI. We use the transceiver pose as input in the second last layer of the fully connected dense layers to provide the location and orientation to the network because it determines the dominant reflectors that are contributing to SRP. The output layer generates the 1×64 vector, which corresponds to the SRP at that pose of the environment. The base DCNN model uses the MSE loss function to train the network and Adam optimizer to update its parameters. The Base DCNN model accurately predicts new SRPs in the same environment but fails to incorporate transfer-learning to predict SRPs in different similar-looking environments. However, *D3PicoNet*'s aims to use one trained model across multiple environments. Since the base DCNN

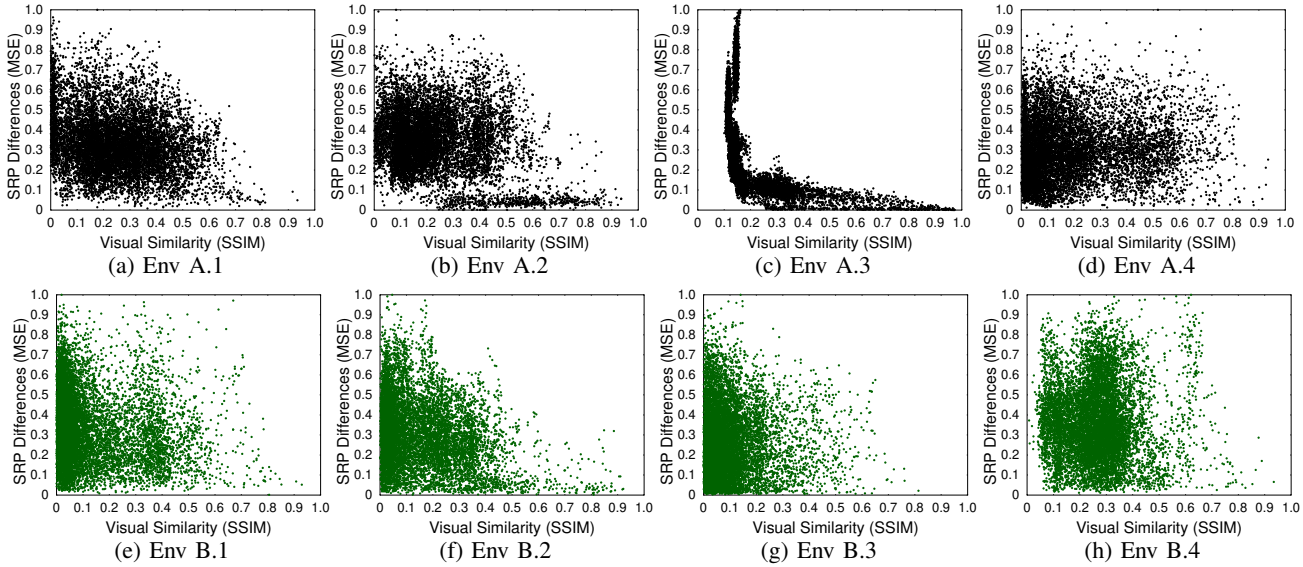


Fig. 3: Non-linear relationship between the SRPs and visual depth images for four environments of two buildings.

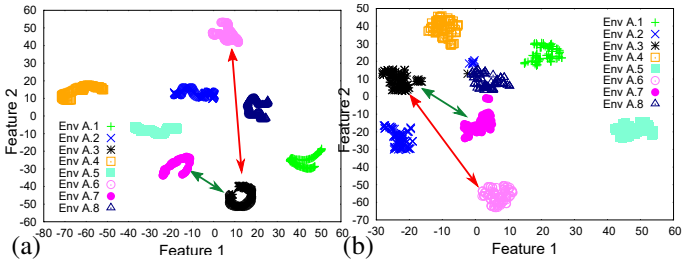


Fig. 4: The t-SNE plot of diverse environments of one building for (a) Visual - Visual depth images. (b) SRPs - SRPs.

model lacks the semantic labels required to identify the environment, they perform poorly across them. Next, we explain *D3PicoNet* semantic-aware DCNN to support transfer learning between environments.

*2) Semantic-Aware Network:* This network aims to make the trained DCNN generalizable over diverse environments. For example, if we train a network with Env A.3 (first floor - area near elevator) data samples, it should be useful for Env A.7 (second floor - area near elevator) since they have similar surrounding objects and properties. The base DCNN model couldn't achieve this because it lacks semantic features on its input. In a semantic-aware network (Figure 8), we provide *Global Features* to the base model's *Local Features* to help the DCNN better understand the environment. *D3PicoNet* first applies semantic segmentation [18] on its GPCD to obtain semantic labels for each point of GPCD. We assign a semantic label to the PCD, which could be one of 13 different classes, such as floor, ceiling, wall, clutter, *etc. Second*, we use a transceiver pose to get LPCD from semantically segmented GPCD. *Third*, we extract *Local Features* of size 1×64 with *MobileNetV2* convolution layers on MIDI (obtained from LPCD). *Fourth*, we use LPCD to construct *Semantic Features*, which is 1×22 one-dimensional vector [X, Y, Z, R, G, B, $N_X$, $N_Y$, $N_Z$, 1×13 one-hot coded vector of semantic label], where X, Y, and Z are the 3D location values of the point, R, G, and B are the color values of the point, and [$N_X$, $N_Y$, $N_Z$] are surface normals of X, Y, Z axes, respectively, to describe the surrounding objects. *D3PicoNet* extracts the *Global Features* 1×32 with PointNet [19] on *Semantic Features. Finally*, we concatenate the *Local Features* and *Global Features* and pass them through two more fully connected layers with Relu [20] activation to predict SRP of size 1×64 at the output layer. To compensate for the hardware measurement error of the 122 GHz mmWave device, we add the Gaussian Noise Layer with ($\mu$, $\sigma$) = (0, 0.77) before the output layer. Furthermore, SRPs are highly sparse [21], and most signals are noise floors. Thus, to reduce the load of the network and help in fast convergence, we don't detect those SRP values below a certain threshold $\gamma$. We accomplish this by using the mask M in the network's loss function, which is set to 1 if the SRP value falls above $75^{th}$ percentile, otherwise 0.

**Network Loss Functions**: Data samples are noisy at 122 GHz, and using Mean Squared Error (MSE) loss is insufficient for optimal network convergence. We build a custom loss function with a combination of MSE and Mean Absolute Error (MAE) as $L_C = \lambda_{MSE} \times L_{MSE} + \lambda_{MAE} \times L_{MAE}$, where $\lambda_{MSE}$ and $\lambda_{MAE}$ are hyper-parameters that need to be optimized [22]–[25] to control the direction of network convergence. Their value ranges from 0 to 1.

### *D. SRP Prediction-aided Picocell Deployment*

Accurate prediction of SRPs with a semantic-aware network allows us to reuse the pre-trained DCNN models from the different environments into the current environment for SRP prediction at unobserved locations. Predicted SRPs carry information about the number and location of strong reflectors. To increase network reliability, we have to use the maximum number of NLoS paths in the absence of LoS path due to frequent indoor blockage. However, finding the picocell

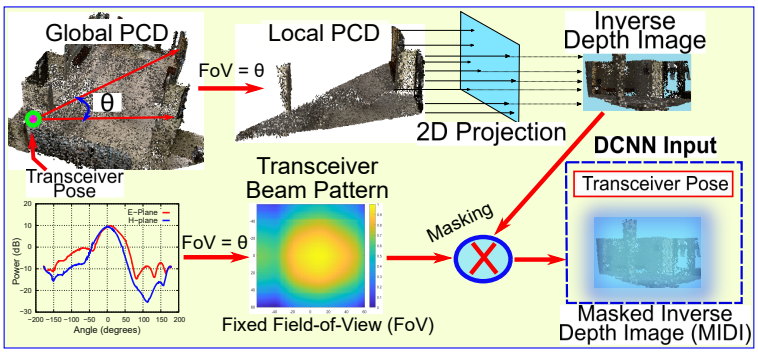


Fig. 5: *D3PicoNet*'s data preprocessing steps to generate MIDI, SRP, and transceiver pose.

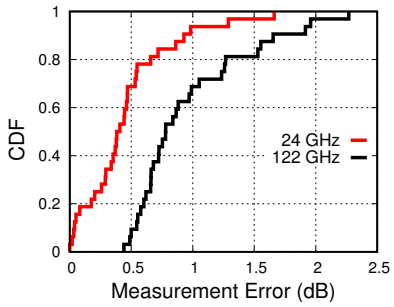


Fig. 6: Measurement error at 24 GHz and 122 GHz with a metallic reflector located at distances from 1.0 m to 3.5 m.

locations based on the SRPs of the environment to utilize the maximum available NLoS paths is complex and requires an efficient algorithm to solve it. Next, we describe the SRP emulation process and picocell deployment methods to maximize NLoS path's utilization.

**Emulating SRPs:** *D3PicoNet*'s SRP prediction uses the co-located transmitter and receiver, which is different from the actual picocell deployment scenario where the transmitter (picocell) and receiver (client) may be located at different locations. But *D3PicoNet*'s SRP prediction model provides errors in SRP estimation for the given environment based on actual measurements and can be used to correct the SRP estimation from simulation-based techniques. So, *D3PicoNet* first uses the Ray-tracing method to simulate SRPs at the measured locations (20K+ samples) to calculate the difference between measured and simulated SRPs. *Recall that we place the picocell and the client at the exact location since our measurement device is the transceiver.* We observe a constant offset (called "*SRP Offset*") between the simulated and measured SRPs however, the nature of the signal profile is similar. The offset between simulated and measured SRPs is due to the lack of certain factors in the simulation, such as noise from the environment, hardware, and surrounding interference [26], [27]. Now, *D3PicoNet* virtually places the picocells and clients at different environment locations, uses the Ray-tracing method to estimate the simulated SRPs, and uses the median of *SRP offsets* to set the correct amplitude of simulated SRPs.

**Deployment Process:** We consider emulated SRPs from each picocell to find their effective locations at 122 GHz.

► **Homogeneous Deployment:** Homogeneous deployment refers to deploying only single frequency (122 GHz) picocells across the environment. *D3PicoNet* aims to place picocells to maximize the number of clients with specific data throughput, referred to as the reduction of *Link-Outage Probability*. *D3PicoNet* compares its deployment strategy with three other methods: (1) "Random," where picocells are placed randomly at any location. (2) "Common-Sense" *a.k.a.* corner deployment, where we deploy picocells at equal distances. (3) "Optimal" refers to deploying picocell without any error in SRP prediction.

► **Heterogeneous Deployment:** It refers to deploying different frequency picocells in an environment to cover the maximum area with a minimum number of picocells to support desired data throughput. Recall that deployment at a low-frequency range may provide broader area coverage with a picocell, but due to its limited bandwidth, it may need help to support data-hungry applications. To ensure wider coverage and high data throughput support, we can deploy 24 GHz and 122 GHz picocells to work in conjunction. To this end, *D3PicoNet* first emulates SRPs at 24 GHz and 122 GHz using Ray-tracing method and then uses these SRPs to find 24 GHz and 122 GHz picocell locations that help to reduce *Link-Outage Probability* of the client in an environment for a specific data throughput support.

## IV. Implementation

**Hardware Platform:** We implement and evaluate *D3PicoNet* with real measurements from our custom-made platform with a 122 GHz mmWave transceiver [5] and a Google Tango device, ASUS Zenfone AR [4] (see Figure 9[b]). The mmWave transceiver has 4 transmit and 4 receive phased-array antenna, each placed with 2×2 format, capable of different beam patterns. The transceiver can collect SRPs in real-time, with a sampling rate of ∼67 milliseconds (ms) per frame, and is connected to a host laptop via a USB cable for storing the data. The transceiver operates on a 1 GHz bandwidth at the center carrier frequency of 122 GHz. We use the following parameters for SRP measurement: Start frequency, 119.6 GHz; frequency ramp slope, 1.43 MHz/$\mu$S; number of complex ADC samples, 512; ADC sampling rate, 487 Msps; sweep duration, 0.70 ms; pulse repetition rate, 15 Hz; and maximum receive

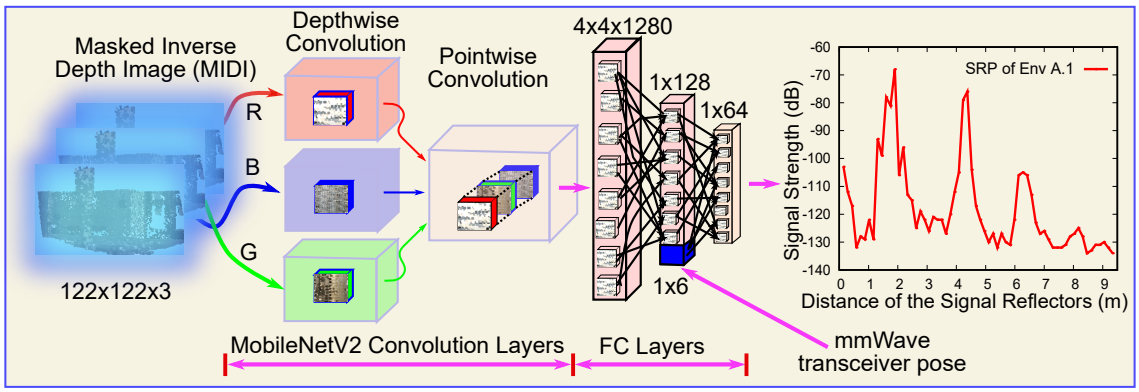


Fig. 7: Deep Convolutional Neural Network (DCNN) base model to learn the association between visual depth images and SRPs.

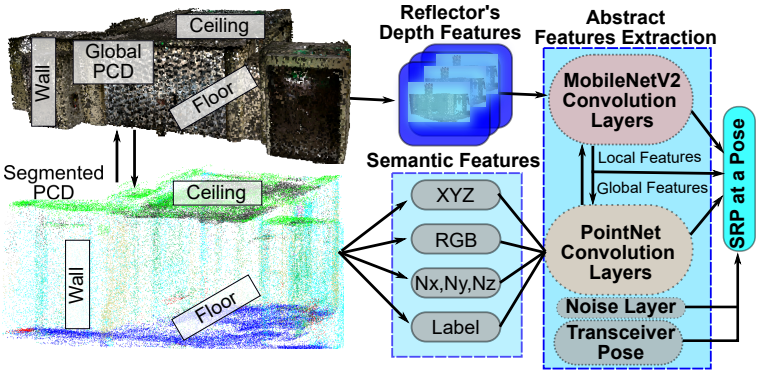


Fig. 8: Semantic-Aware DCNN network with Gaussian Noise Layer to learn a better mapping between visual PCD and SRPs.

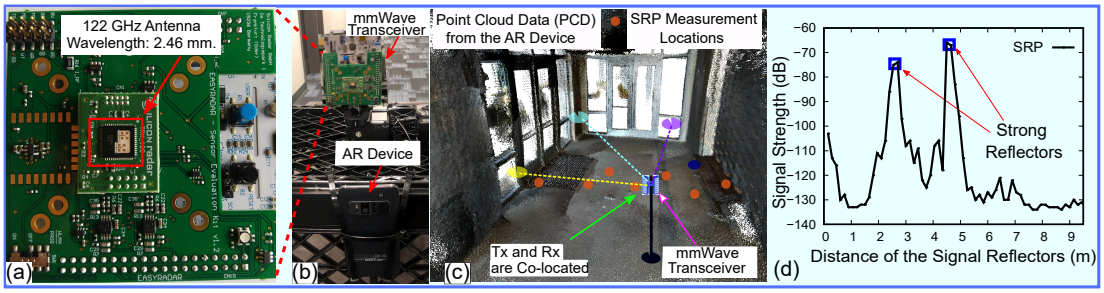


Fig. 9: (a) Zoom-in view of 122 GHz mmWave transceiver. (b) 122 GHz mmWave transceiver and AR device setup. (c) An instance of PCD with highly reflective objects and the mmWave transceiver. (d) SRP (dB) at a pose (blue circle of Figure 9[c]) with strong reflectors at ∼2.6 m and 4.6 m.

antenna gain, 56 dBi. The measured SRP at a particular pose is a 256 element vector corresponding to the number of ADC samples. Since the transceiver operates at 1 GHz bandwidth, each SRP element has a resolution of ∼0.1499 m [28]; hence, the device can gather reflections up to ∼38.40 m. However, the reflection strength at such a long distance is below the noise floor, and the picocells are supposed to operate within 10s of m in indoor settings; so, we limit the maximum range to ∼10 m, which corresponds to the first 64 elements. We keep the distance between the transceiver and AR device fixed during all the data collection and use the known distance offset to calibrate the transceiver's pose.

**Real Data Collection:** Due to a lack of tight synchronization between AR device and the mmWave transceiver, we post-process the collected dataset in software. A Matlab program running the host PC first starts the AR device to collect the visual data using the RTAB-SLAM app [29] and waits for scene stabilization while the transceiver starts recording the reflected signals. After the scene stabilizes, we move the setup around to construct the visual map of the environment and gather SRPs from various poses. A single scan of ∼4 mins generates a PCD of a typical indoor environment and can gather reflected signals from ∼3700 unique transceiver poses. We then apply 1D FFT on the reflected signals to obtain the SRPs. We post-process the pose and SRPs to match the SRP sampling rate. We identify timestamps by correlating consecutive SRPs to find movement and match with AR device's timestamp by observing self-pose change. Since AR device and transceiver are co-located, we can calibrate their

local timestamps and get synchronized SRPs and poses.

Figures 9(c–d) show an example environment and the measured SRP from one of the locations (marked in a blue circle). There are two strong reflectors at ∼2.6 m and 4.6 m from this location, corresponding to two strong peaks in the SRP. We collect the PCDs, poses, and SRPs from 16 different environments [2] to train and evaluate *D3PicoNet*. Although we use only a few minutes of scans to the train network, we scan environments thoroughly up to ∼20 hrs at multiple days to evaluate the temporal behavior and compare *D3PicoNet* with complete site surveys. In total, we have collected and analyzed over 45 GB of the dataset with ∼4.2 million data samples. We have used only ∼420K samples for training, and the rest of ∼3.8M pairs are used for testing and benchmarking all our design components. *Such data diversity and scale allow us to evaluate the performance of D3PicoNet and understand its generalizability across multiple environments*.

**Neural Network Training**: We train and evaluate the SRP prediction model with a combination of MSE and MAE loss. We follow a similar training process, number of epochs, and optimizer of prior implementation [2] and find that "Adam" optimizer [30] with a learning rate of 0.002 works best for both base and semantic-aware models. Finally, all the DCNN models are designed and implemented with Python programming language and PyTorch [31] packages on Nvidia's GPU (RTX A6000) [32]. All our networks require ∼4-6 hours to train. However, uploading the data to the Cloud TPUs [33] can further reduce the training time.

## V. Performance Evaluation

We evaluate *D3PicoNet* with two standard metrics: *Absolute Error* and *Link Outage Probability*.

► **Absolute Error**: The absolute difference between the actual and predicted values of SRPs, measured in dB.

► **Link-Outage Probability**: The probability of a client without an established link to the picocell (0% to 100%).

**Evaluation Summary**: (1) *D3PicoNet*'s base model predicts SRPs with a median error of 3.0 dB, but the error increases to 8.0 dB when the model is tested in similar-looking environments because it lacks generalizability. Semantic labels and surface normal of PCD as *"Global Features"* reduce SRP prediction error to 6 dB and drop to 2.2 dB with just ∼4 mins. of additional model fine-tuning. (2) *D3PicoNet* finds the picocell locations with more than 96% accuracy (compared to *Optimal*) and can provide coverage to nearly 100% of regions with data throughput up to ∼35 Gbps.

### A. SRP Prediction

**Base DCNN Model:** We evaluate the base DCNN model on environments Env A.1, Env A.2, Env A.3, and Env A.4, which contains regions for office spaces and areas near the elevator, with 20K data samples from each environment. Out of 20K, we use 18K for training and 2K for testing. First, we preprocess the data to generate each sample's MIDI, Pose, and SRP. We then use training samples to train the base model

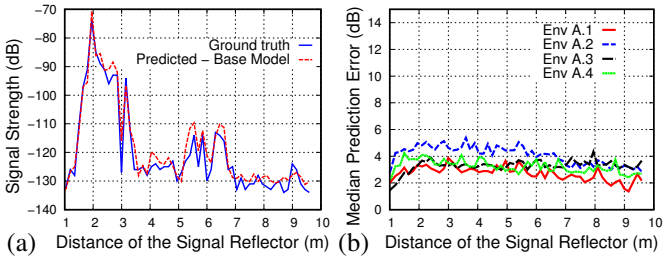


Fig. 10: (a) An illustration of SRP prediction utilizing a basic DCNN model in a single location in Environment A.1. (b) Relationship between the distance of the reflector and SRP prediction error.

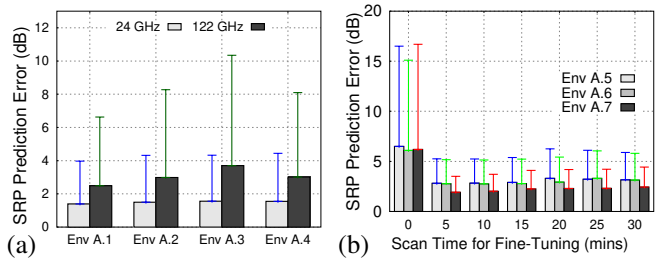


Fig. 11: (a) SRP prediction errors across four different environments. The bar is the median error, and the errorbar is the $90^{th}$ percentile error. (b) Effect on SRP prediction error with different scan times for fine-tuning.

for each environment until the models converge. We take the L1-norm between the predicted SRPs and actual SRPs of 2K test samples to determine the *Absolute Error* of the SRP prediction after training. Figure 10(a) shows an example of SRP of Env A.1 and its predicted values. SRP prediction is accurate irrespective of reflector distances (see Figure 10[b]). Finally, Figure 11(a) shows that the SRP prediction error is 3.0 dB on average across Envs A.1 to A.4, and the prediction error doesn't exceed 11 dB. Across all environments, we observe that the SRP prediction error for the base DCNN model is slightly higher for 122 GHz than 24 GHz. Higher error is due to the increased sensitivity of SRPs with their operating frequency [34]. So, the base model performs exceptionally well within the environment, but the error is more than 8 dB when tested in different environments, which is unacceptable in practice (see Figure 12[a], red curve).

**Semantic-Aware Model:** The base model performs poorly across different environments because it needs the *Semantic Features* to enable transfer learning. In *D3PicoNet*, we first correct the surface normals with [35] and apply semantic segmentation on GPCDs [18] for all data samples of building A & B. Since two buildings have multiple environments within themselves, we intend to develop a model for each building and use it across different environments within the building. GPCDs of building A & B are semantically segmented and pre-processed to generate *Semantic Features*, *MIDI*, *SRP,* and *Pose*. We use Envs A.1 to A.4 and Envs B.1 to B.4 to train models for building A & B, respectively. We use the rest of the Envs of both buildings for testing. Figure 12(a) shows the result of the semantic-aware model for building A. The

semantic-aware model can reduce SRP prediction error from 8 dB of the base model to 6.5 dB by including the *Semantic Features*. The SRP prediction error is further reduced to 6.0 dB by ignoring the noise floor of the SRP and reaches 2.2 dB by fine-tuning with ∼4 mins. of samples. We observe the consistent performance of the semantic-aware model across building A and B. Figure 12(b) shows that we can achieve SRP prediction errors of 8.0 dB and 6.75 dB without fine-tuning; 2.2 dB and 1.5 dB with limited (4 mins) fine-tuning for two buildings.

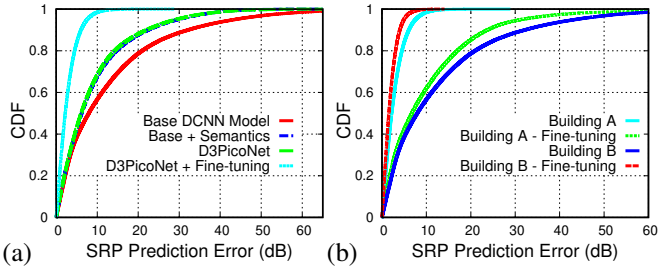


Fig. 12: (a) SRP prediction error for transfer-learning with Base Model, Base Model + Semantic Label, *D3PicoNet*, *D3PicoNet* + Fine-tuning. (b) SRP prediction error with transfer-learning with Base Model and *D3PicoNet* + Fine-tuning for buildings A and B.

**Scanning Time Requirement for Fine-Tuning:** Even though *D3PicoNet* performs sub-optimally with the inclusion of *Global Features*, we can achieve near-optimal performance with limited fine-tuning of a previously trained model with current environment samples. Introducing more samples from the current environment into the model reduces the prediction error on the rest of the test samples. However, the number of data samples necessary to achieve the near-optimal performance depends on how well the previously trained models have learned the *Global Features*. In *D3PicoNet*, we use building A to analyze the scanning time requirement. We use the previously trained model for building A (with Envs A.1 to A.4) as the starting configuration of the semantic-aware model and gradually feed the data samples from testing environments *i.e.* Envs A.5 to A.7, to fine-tune. We then use the remaining samples (not used in fine-tuning) to predict the SRPs and compute the SRP prediction error. Figure 11(b) shows the SRP prediction error to 6.0 dB from 8.0 dB (with no fine-tuning), and the error drops significantly and drops to 2.2 dB with just 5 mins. of data samples. The prediction error remains flat at 2.0 dB with additional fine-tuning showing that *D3PicoNet* achieves near-optimal results with only 5 mins. of samples. The results show that *D3PicoNet* is robust and can adapt well across multiple environments without a deployer spending time collecting more samples.

**Temporal Performance of SRP Prediction:** Indoor environments such as a hallway, classrooms, meeting offices, *etc.* may be subject to change in their surrounding structures, such as the reorientation of desks and benches and the addition or removal of wall paintings. Since *D3PicoNet* relies on the previously randomly measured SRPs to train and build the

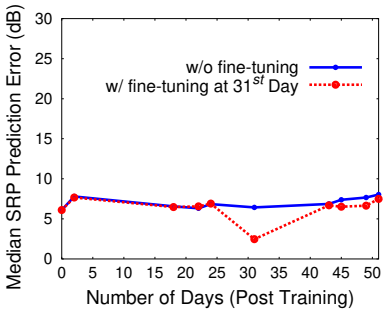


Fig. 13: Temporal performance on Building A; without fine-tuning & with fine-tuning on $31^{st}$ day.

model, these changes may affect the model performance. In *D3PicoNet*, we would like to explore how these changes affect the accuracy of SRP prediction. We use building A for temporal performance analysis, collecting more samples on various days and evaluating them on a previously trained model. Figure 13 shows that the median SRP prediction stays ∼6.5 dB on average over 51 days. The average median error reduces with fine-tuning of the model around $31^{st}$ day. The result shows that the trained model stays applicable with periodic fine-tuning.

### *B. Picocell Deployment*

**Homogeneous Deployment:** We evaluate 122 GHz homogeneous deployment on Env A.1 and Env A.2 because they represent the diverse set of environments with different structures of building A. We first emulate SRPs in these two environments using the Ray-tracing method (process described in section III-D). Emulated SRPs indicate 100% accurate SRPs at all unobserved locations and correspond to *Optimal* deployment. We use the mean, and standard deviation of the SRP prediction error from measurements to re-estimate SRPs at all locations since *D3PicoNet*'s SRP prediction is not 100% accurate. *We hypothesize that since we can predict SRP with a learning-based model, we can then predict signal strength from picocell to reflector and reflector to the client.*

Figure 15(a) shows the location of picocells obtained with *Common-Sense*, *Random*, *D3PicoNet*, and *Optimal* to reduce *Link-Outage Probability*. We see that *D3PicoNet*'s locations are overlapped more than 96% with *Optimal* (62 out of 64), showing that *D3PicoNet* achieves near-optimal deployment results. Figure 15(b) shows the 2D top-view layout of Env A.2, and we observe that most picocells are located near the lower center to maximize the use of metallic reflectors near the building entrance. Figure 16(b) shows that *Optimal* and *D3PicoNet* ensure that *Link-Outage Probability* remains below 15% across all samples indicating wide coverage. Based on the SRPs, we estimate the data throughput at the client location using the available bandwidth ∼7 GHz and noise floor (-120 dB for 122 GHz). Figure 16(a) shows that *D3PicoNet* and *Optimal* require an equal number of picocells to support data throughput of 27 Gbps. Also, *D3PicoNet* always selects picocell locations better than *Random* (red and black curves of Figure 16[a]). This result shows that *D3PicoNet* deploys

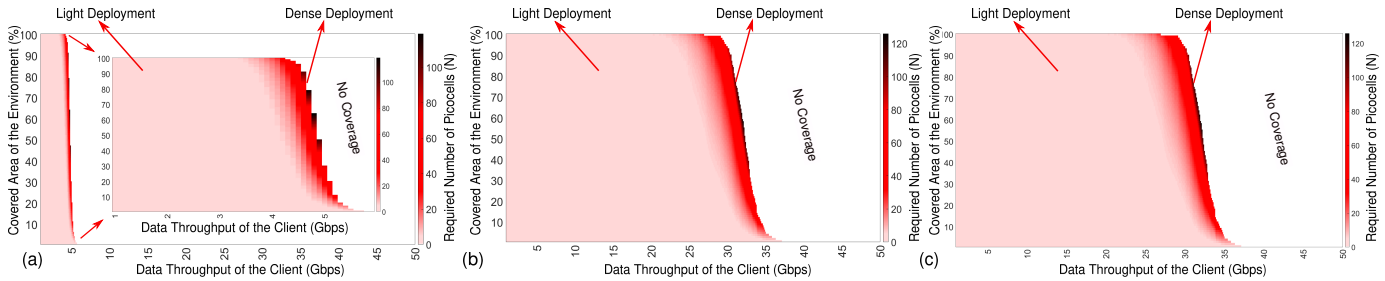


Fig. 14: Heatmap of the required number of picocells for Env A.2 for different values of the coverage of the environment and client's data throughput: (a) Homogeneous deployment for 24 GHz picocells and its zoom-in view. (b) Homogeneous deployment for 122 GHz picocells. (c) Heterogeneous deployment of deploying 24 GHz and 122 GHz picocells together.

picocells with more than 96% accuracy.

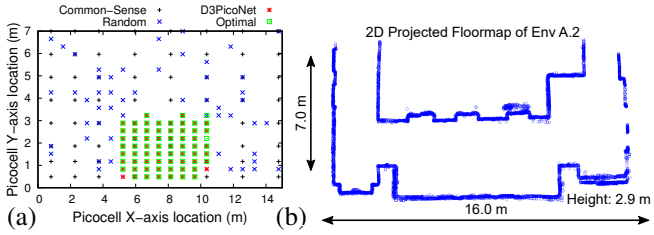


Fig. 15: (a) Picocell locations predicted by Optimal, *D3PicoNet*, Random, and Common-Sense to minimize *Link-Outage Probability*. (b) Top view of the environment near the office elevator (Env A.2).

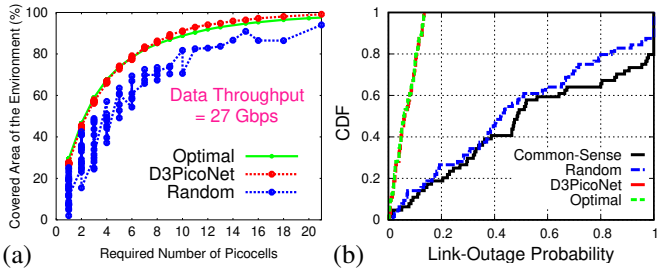


Fig. 16: (a) Number of picocells required for different area coverages of Env A.1 with Optimal, *D3PicoNet*, and Random to support 27 Gbps data rate. (b) *Link-Outage Proability* of the client with Optimal, *D3PicoNet*, Random, and Common-Sense.

**Heterogeneous Deployment**: We use Env A.2 for heterogeneous deployment. Based on the practical bandwidth range, we use the bandwidth of 1 GHz and 7 GHz for 24 GHz and 122 GHz, respectively, to calculate data rates based on their SRPs. We simulate *D3PicoNet* to find the number of picocells needed to support different data throughput and environment coverage in three different settings: (a) with 24 GHz, (b) with 122 GHz, and (c) with 24 GHz and 122 GHz together. Figures 14(a–c) show a heatmap of the picocells needed to cover a given environment region with specific data throughput support. We find that 24 GHz can only support up to ∼5 Gbps on average, while a higher number of picocells (*Dense Deployment*) are needed to support a wide area of the environment. However, D-band (122 GHz) can support up to ∼35 Gbps with a smaller number of picocells (*Light Deployment*). Deploying 24 GHz and 122 GHz together reveals that most of the coverage is dominated by 122 GHz picocells, thus requiring more 122 GHz picocells to support higher data throughput. However, a more realistic approach requires a detailed analysis of picocell cost at 122 GHz and 24 GHz, the data rate requirement of clients in a heterogeneous environment.

## VI. Related Work

**Signal Reflection Prediction at D-Band:** Signal attenuation increases with the operating frequency of the mmWave and requires a careful analysis of the environment before their prediction. Simulation-based methods such as the Ray-tracing propagation model [36] are accurate at low-frequency deployments and mostly applicable for outdoor settings. Matlab's ray tracing models for diverse environments are currently limited to 100 GHz, and there are extremely limited datasets about the signal reflection profile model at D-band. [21] used the simulation approach after finding the strong reflectors explicitly and achieved the prediction error of 2.8 dB. But their approach requires modeling each environment separately and needs to be more generalizable. Various commercial products aim to help outdoor deployment by implementing signal repeaters and exploiting channel multipath [37], [38], but they are primarily focused on outdoor settings and have minimal research on them. Besides, the specularity of mmWave [39] and weak reflectivity [40] of the indoor reflectors at higher frequency demands extensive surveys to know the SRP of the environment locations. In *D3PicoNet*, we propose the vision-assisted, semantically-corrected, and deep learning-based approach to predict SRPs across the entire environment based on a few random observations.

**SRP-aided Picocell Deployment:** Accurate signal reflection profile prediction allows multiple applications [41]–[45]. All these applications rely on the network to choose the correct coding scheme [46] to transmit data with higher data throughput. Multiple commercial tools [37], [47] include the enhanced vegetation segmentation on Ray-tracing modeling to effectively deploy small 5G cells outdoors to support the maximum number of clients. Prior work has explored the optimal deployment at 24 GHz [2] to help maximize the number of clients and provide better fairness. Furthermore, the heterogeneous deployment of 24 GHz and 122 GHz picocells together allows the picocell to support the nearby client based on their data throughput requirement. In *D3PicoNet*, we find the trade-off between the cost, coverage, and data throughput

to explore the possibility and challenges of deploying multi-frequency picocells to support various applications.

## VII. Conclusion

This work presents a deep-learning approach to reduce the manual site survey requirements for effective deployment of picocells at D-band mmWave frequency. *D3PicoNet* only uses a few randomly collected samples to learn and predict the SRPs at all remaining unobserved locations. *D3PicoNet* leverages the SRP prediction and *D3PicoNet*'s deployment algorithm to find the number of picocells and their locations to provide the required level of coverage across the environment with specific data throughput guaranteed. *D3PicoNet* also generalizes to diverse environments making it suitable for any indoor deployment with a limited model update to enable reliable next-generation connectivity.

## VIII. Acknowledgments

We sincerely thank the reviewers for their comments. This work is partially supported by the NSF under grants CAREER-2144505, CNS-1910853, and MRI-2018966.

## References

[1] L. Tan, et al., "Speech Emotion Recognition Enhanced Traffic Efficiency Solution for Autonomous Vehicles in a 5G-Enabled Space–Air–Ground Integrated Intelligent Transportation System," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 3, 2021.

[2] H. Regmi, et al., "Argus: Predictable Millimeter-Wave Picocells with Vision and Learning Augmentation," *Proc. of ACM SIGMETRICS*, 2022.

[3] T.D. Hooks, et al., "Poster: VisualMM: Visual Data and Learning Aided 5G Picocell Placement," in *ACM HotMobile*, 2021.

[4] Gsmarena, "Asus Zenfone AR ZS571KL," 2022. [Online]. Available: https://www.gsmarena.com/asus_zenfone_ar_zs571kl-8502.php

[5] Silicon Radar GmbH, "120 GHz Products," 2022. [Online]. Available: https://siliconradar.com/products/\#120ghz-radar-chips

[6] V. Raghavan, et al., "Statistical blockage modeling and robustness of beamforming in millimeter-wave systems," *IEEE T-MTT*, vol. 67, no. 7, 2019.

[7] C. Slezak, et al., "Empirical effects of dynamic human-body blockage in 60 GHz communications," *IEEE Communications Magazine*, vol. 56, no. 12, 2018.

[8] C.F. Yang, et al., "A Ray-Tracing Method for Modeling Indoor Wave Propagation and Penetration," *IEEE transactions on Antennas and Propagation*, vol. 46, no. 6, 1998.

[9] V. Maaten, et al., "Visualizing data using t-SNE." *Journal of machine learning research*, vol. 9, no. 11, 2008.

[10] L. Stéphane, et al., "A comprehensive analysis of deep regression," *IEEE transactions on pattern analysis and machine intelligence*, vol. 42, no. 9, 2019.

[11] J. Wang, et al., "Fair classification with group-dependent label noise," in *Proc. ACM on FAccT'21*, 2021.

[12] A. Camuto, et al., "Explicit regularisation in gaussian noise injections," *NIPS*, vol. 33, 2020.

[13] L. Chen, et al., "Generalized Correntropy based deep learning in presence of non-Gaussian noises," *Neurocomputing*, vol. 278, 2018.

[14] A. Mehmet, et al., "SMOTE and gaussian noise based sensor data augmentation," in *2019 4th International Conference on Computer Science and Engineering (UBMK)*, 2019.

[15] K. Sarabandi, et al., "Compact Beam Scanning 240 GHz Radar for Navigation and Collision Avoidance," in *Micro-and Nanotechnology Sensors, Systems, and Applications III*, 2011.

[16] M. Mosalanejad, et al., "Multilayer Compact Grid Antenna Array for 79 GHz Automotive Radar Applications," *IEEE Antennas and Wireless Propagation Letters*, vol. 17, no. 9, 2018.

[17] M. Sandler, et al., "MobileNetV2: Inverted residuals and linear bottlenecks," in *IEEE/CVF CVPR*, 2018.

[18] I. Armeni, et al., "3D Semantic Parsing of Large-Scale Indoor Spaces," in *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, 2016.

[19] C.R. Qi, et al., "PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation," in *IEEE-CVPR*, July 2017.

[20] K. Eckle, et al., "A comparison of deep networks with ReLU activation function and linear spline-type methods," *Neural Networks*, vol. 110, 2019.

[21] T. Wei et al., "Facilitating Robust 60 GHz Network Deployment By Sensing Ambient Reflectors," in *14th USENIX Symposium on NSDI*, 2017.

[22] M. Claesen, et al., "Hyperparameter search in machine learning," *arXiv preprint arXiv:1502.02127*, 2015.

[23] J. Bergstra, et al., "Random search for hyper-parameter optimization." *Journal of machine learning research*, vol. 13, no. 2, 2012.

[24] M. Claesen, et al., "Easy hyperparameter search using optunity," *arXiv preprint arXiv:1412.1114*, 2014.

[25] P. Balaprakash, et al., "DeepHyper: Asynchronous hyperparameter search for deep neural networks," in *2018 IEEE HiPC*, 2018.

[26] T. Leila, et al., "Device-to-device mmWave communication in the presence of interference and hardware distortion noises," *IEEE Communications Letters*, vol. 23, no. 9, 2019.

[27] C. Jiang, et al., "mmVib: micrometer-level vibration measurement with mmwave radar," in *ACM MobiCom*, 2020.

[28] D.M. Sheen, et al., "Three-dimensional millimeter-wave imaging for concealed weapon detection," *IEEE Transactions on microwave theory and techniques*, vol. 49, no. 9, 2001.

[29] Open-Source, "Real-Time Appearance-Based Mapping," 2022. [Online]. Available: http://introlab.github.io/rtabmap/

[30] "Optimizers," https://keras.io/api/optimizers/, 2022.

[31] Open-Source, "PyTorch," 2022. [Online]. Available: https://pytorch.org/

[32] NVIDIA, "GEFORCE," 2022. [Online]. Available: https://www.nvidia.com/en-us/geforce/

[33] Google, "Cloud TPU," 2022. [Online]. Available: https://cloud.google.com/tpu

[34] T. S. Rappaport, et al., "Wireless Communications and Applications Above 100 GHz: Opportunities and Challenges for 6G and Beyond," *IEEE Access*, vol. 7, 2019.

[35] Open-Source, "Open3d estimate normals," 2022. [Online]. Available: http://www.open3d.org/docs/0.7.0/python_api/open3d.geometry.estimate_normals.html

[36] MathWorks, "raytrace: Display or compute RF propagation rays," 2022. [Online]. Available: https://www.mathworks.com/help/comm/ref/txsite.raytrace.html

[37] movandi, "Unlimited Expansion," 2022. [Online]. Available: https://movandi.com/applications/

[38] Qualcomm, "See what 5G is doing for you." 2022. [Online]. Available: https://www.qualcomm.com/products/technology/5g

[39] H. Regmi, et al., "SquiggleMilli: Approximating SAR Imaging on Mobile Millimeter-Wave Devices," *Proc. of ACM IMWUT*, vol. 5, no. 3, 2021.

[40] S. Kim et al., "D-band channel measurements and characterization for indoor applications," *IEEE Transactions on Antennas and Propagation*, vol. 63, no. 7, 2015.

[41] T. Wei,et al., "Pose Information Assisted 60 GHz Networks: Towards Seamless Coverage and Mobility Support," in *Proc. of ACM MobiCom*, 2017.

[42] Y. Tsukamoto, et al., "Feedback Control for Adaptive Function Placement in Uncertain Traffic Changes on an Advanced 5G System," in *IEEE-CCNC*, 2021.

[43] A.P. Mathew, et al., "An uniform clustering based coverage and cost effective placement of serving nodes for 5G," in *ICEEIMT*, 2017.

[44] S.K. Khan, et al., "UAV-aided 5G Network in Suburban, Urban, Dense Urban, and High-rise Urban Environments," in *2020 IEEE 19th International Symposium on NCA*, 2020.

[45] S. Bartoletti, et al., "5G localization and context-awareness," *University of Bologna, University of Ferrara*, 2018.

[46] X. Duan, et al., "SDN Enabled 5G-VANET: Adaptive Vehicle Clustering and Beamformed Transmission for Aggregated Traffic," *IEEE Communications Magazine*, vol. 55, no. 7, 2017.

[47] Siradel, "Volcano 5G: Siradel announces enhanced version of its Volcano propagation model!" 2022. [Online]. Available: https://www.siradel.com/