

SymNet: A Multi-Task Network for Joint Radio Map Reconstruction and Transmitter Localization

Lyuzhou Ye

University of North Texas

1155 Union Circle

Denton, Texas

lyuzhouye@my.unt.edu

Thanh Dat Le

University of North Texas

1155 Union Circle

Denton, Texas

ThanhLe5@my.unt.edu

Yan Huang

University of North Texas

1155 Union Circle

Denton, Texas

Yan.Huang@unt.edu

Abstract

Accurately predicting directional radio maps is essential for wireless applications, yet prior approaches primarily focus on omnidirectional signals and typically treat transmitter localization and signal map reconstruction as separate tasks. In omnidirectional settings, predicting the maximum signal location often coincides with the transmitter position, which limits the need for explicit joint modeling. However, in directional propagation where angular effects, reflections, and building occlusions play critical roles, this assumption no longer holds. To address this gap, we propose SymNet, a unified framework that jointly predicts directional radio maps and transmitter locations from sparse signal measurements. SymNet incorporates a prediction head for transmitter localization alongside radio map reconstruction, enabling simultaneous learning of both tasks. This joint formulation leverages their complementary information and leads to consistent improvements over treating them separately. Experiments on challenging directional scenarios demonstrate that SymNet outperforms state-of-the-art baselines, achieving superior accuracy in both radio map reconstruction and transmitter localization.

1. Introduction

Wireless systems increasingly rely on directional propagation (e.g., in mmWave and beamforming-based 5G), where received signal strength is highly sensitive to angles, occlusions, and reflections. A fundamental challenge in this setting is to reconstruct accurate directional radio maps from sparse measurements and to determine the transmitter location simultaneously. Both tasks are critical for applications such as network planning, localization, and resource management, yet they are often studied in isolation.

On the one hand, recent learning-based approaches have advanced radio map prediction by leveraging convolutional or U-Net-like architectures [3, 14, 16, 32]. These models

focus on reconstructing omnidirectional signal distributions and achieve strong performance compared to classical interpolation methods, but typically assume that the transmitter position can be directly inferred from the heatmap maximum. Meanwhile, localization-focused methods such as LocNet [12], TLDL [19], and DSLoc [15] employ U-Net-style encoder-decoder networks (or HRNet in DSLoc) to directly regress transmitter heatmaps and extract positions. While effective in omnidirectional or simplified propagation scenarios, these methods do not explicitly address the broader radio map reconstruction problem.

In directional settings, the strongest signal region often deviates from the transmitter location due to angular effects and multi-path propagation. This breaks the implicit alignment between the two tasks, making radio map reconstruction and transmitter localization not only distinct but also potentially complementary. Despite this, prior work in both areas has either reconstructed maps without explicit localization or localized transmitters without modeling the full signal distribution. Our contributions are as follows:

- We identify and formulate the joint prediction problem of directional radio maps and transmitter localization, a task not explicitly studied in omnidirectional-focused studies.
- We present SymNet, a novel architecture that couples heatmap reconstruction with transmitter prediction through a shared backbone and task-specific heads.
- Through experiments in challenging directional scenarios, we show that SymNet achieves consistent gains in both tasks compared to state-of-the-art methods: it reduces localization error (meters) by an average of 34.36% relative to DSLoc across 20–100 samples, lowers **RMSE** by 4.08% compared with ViT-RefineNet, and improves **SSIM** by +4.70% on average. Moreover, it surpasses the omnidirectional SOTA DC-Net by 15% in **RMSE**, highlighting its advantage under angular-sensitive and multipath-dominated conditions.

2. Related work

2.1. Methods for Radio Map Generation

Classical approaches for constructing radio maps can be broadly divided into parametric and non-parametric methods. Parametric techniques, such as compressed sensing [23] and dictionary learning [10], assume detailed knowledge of transmitter characteristics. Non-parametric methods, such as Ordinary Kriging [2, 17] and Radial Basis Function interpolation [9], rely on hand-crafted models of signal propagation. Both categories face substantial limitations in complex urban environments, where multipath effects and severe attenuation often invalidate their underlying assumptions.

Recent work on deep learning-based radio map generation can be divided into two categories: those assuming known transmitter locations and those assuming unknown transmitter locations. When the transmitter position is known, models such as RadioUNet [14] and its extensions [13, 20, 22] learn robust propagation patterns by combining location priors with environmental data. More lightweight designs, including GAN-based models [18, 26] and Transformer-augmented autoencoders [25], have further improved efficiency and map realism.

When the transmitter position is unknown, the task becomes significantly more challenging, as models must infer radio maps solely from sparse receiver measurements without prior knowledge of the source. This setting has been the focus of much recent research.

Teganya et al. [24] first demonstrated that a U-Net could generate accurate omnidirectional radio maps using only sparse samples and a coarse environment map. Building on this idea, SkipNet [16] introduced lightweight skip connections to improve stability, while RadioUNet-S [14] explored deeper variants to increase representational capacity. Generative approaches such as GAN-CRME [33] streamlined the U-Net generator, utilizing adversarial training to enhance realism with reduced computational cost. Enes et al. [11] further advanced this line by employing twin networks to jointly predict signal power and uncertainty, improving robustness in highly variable environments.

Despite these advances, convolutional designs fundamentally struggle to capture long-range dependencies, which are critical for directional propagation. To overcome this, later methods have emphasized multi-scale context aggregation. ACT-GAN [4] and DC-Net [3] mimicked Transformer-like receptive fields through pyramid pooling and dilated convolutions, partially alleviating locality constraints, though these approximations may struggle to capture precise angular sensitivities or long-range interactions.

Most recently, ViT-RefineNet [29] introduced a hybrid design that combines a Vision Transformer backbone with U-Net-like refinement modules, effectively balancing

global dependency modeling with fine-grained detail recovery. Although achieving strong performance in the prediction of the directional radio map, this model, similar to prior approaches, it focused exclusively on heatmap reconstruction and did not address transmitter localization as a distinct and complementary task.

2.2. Methods for Transmitter Localization

Traditional transmitter localization methods, such as time of arrival (TOA), time difference of arrival (TDOA), and angle of arrival (AOA) [6, 7], require complex hardware support and are difficult to deploy at scale. RSS-based localization is easier to implement [1, 21] but is highly sensitive to non-line-of-sight propagation and typically depends on propagation models that fail in dense urban environments [8].

With the rise of deep learning, researchers have explored data-driven localization frameworks that learn propagation characteristics directly from measurements. Coordinate-based approaches predict the transmitter's (x, y) location explicitly. Zhang et al [31] pioneered the use of deep learning for transmitter localization, combining an MLP feature extractor with an HMM to directly predict transmitter coordinates. DeepTxFinder [35] extended this direction by mapping sparse RSS readings into grid cells and using a CNN-MLP design to estimate both the number and positions of transmitters, showing good performance in dense sampling scenarios. Building on this line, Wang et al. [27] proposed MT-GCNN, a multi-task gated CNN that jointly performs containment cell classification and distance regression, improving robustness under varying sensor deployments. While effective, these coordinate-based methods often face generalization challenges when sensor layouts change or sampling density is low.

More recently, heatmap-based approaches have become dominant. DeepMTL [30] reframed localization as object detection, adapting YOLOv3 to produce transmitter probability maps. TL;DL [19] predicts transmitter heatmaps with a U-Net and then thresholds/suppresses peaks to extract transmitter coordinates from the output, which is capable of localizing multiple transmitters. LocNet [12] further demonstrated that a lightweight U-Net variant can achieve competitive accuracy with orders of magnitude fewer parameters. Most recently, DSLoc [15] addressed the extreme sparsity challenge by introducing bias correction and sparse expansion preprocessing, adopting HRNet as the backbone to preserve high-resolution features, and replacing argmax with centroid regression to improve localization precision.

While these deep learning approaches have achieved promising results, they are primarily localization-focused solutions and generally do not attempt to reconstruct full radio maps and are largely developed under omnidirectional assumptions, where the transmitter location aligns with the signal maximum. However, in directional propagation, this

assumption no longer holds, and localization and radio map prediction become distinct but complementary tasks. This gap motivates our proposed SymNet, which unifies both tasks in a joint framework.

3. Preliminary

3.1. Problem Definition

Given a region of interest (ROI) that is divided into a two-dimensional grid of $H \times W$, where H and W are the height and width of the grid, respectively, with $(H, W \in \mathbb{Z}^+)$. A single transmitter, equipped with a directional antenna whose orientation (0–360 degrees) is randomly assigned in each scenario, is located at (T_x, T_y) within this grid. A subset of pixels in the area is randomly sampled. Our goal is to develop a model to predict the radio strength at each pixel in the given region as well as to predict the position of the transmitter.

3.2. Grid-based Representation

We use two matrices to represent the environment. The first matrix is the building map $\mathbf{E} = (b_{i,j})$, where

$$b_{i,j} = \begin{cases} -1, & \text{if } (i, j) \text{ is a building pixel,} \\ 1, & \text{if } (i, j) \text{ is a sampled pixel,} \\ 0, & \text{if } (i, j) \text{ is a non-building non-sampled pixel.} \end{cases}$$

The second matrix is the signal map $\mathbf{R} = (r_{i,j})$, defined as

$$r_{i,j} = \begin{cases} s_{i,j}, & \text{if } (i, j) \text{ is sampled (with reading } s_{i,j}), \\ 0, & \text{otherwise.} \end{cases}$$

4. Proposed Model Framework

Our model architecture, illustrated in Figure 1 (Top), is composed of several modules connected through a sequential data flow. We concatenate the output of each module with the original environment information at the fusion points to ensure that the model utilizes environmental information.

The data flow begins with a SkipNet with attention gates that enhances local features. Its output, after concatenation with the original environment map, is fed into two parallel paths to capture the global contextual information, one for heatmap generation and the other for transmitter location prediction. After patchification and embedding, the data passes through six self-attention encoding blocks of the Vision Transformer [5], the outputs are then fused by one cross-attention module, and added back to the original outputs. Then the two separated fused outputs are fed into two separate self-attention encoding blocks again, decoded

by two lightweight decoders, concatenated with each other and then the environment map again, and passed to two extra SkipNet modules for final output.

This design is inspired by the intuitive relationship between the two tasks: the signal heatmap provides structural cues that narrow the plausible transmitter region, while the estimated location offers global guidance that helps refine the heatmap.

4.1. Model Input

For the input of our model in Figure 1 (Left), in addition to the building map and signal map, we compute the distance from each pixel to its nearest building pixel and construct an additional channel called **Distance from Nearest Building (DNB)**, denoted as $\mathbf{D} = (d_{i,j})$, where

$$d_{i,j} = \min_{(p,q) \in \mathcal{B}} \sqrt{(i-p)^2 + (j-q)^2},$$

and $\mathcal{B} = \{(p, q) \mid b_{p,q} = -1\}$ denotes the set of all building pixels. We normalize this data as follows:

$$d'_{i,j} = \begin{cases} 1 - \frac{d_{i,j}}{d}, & 0 \leq d_{i,j} \leq d, \\ 0, & d_{i,j} > d. \end{cases}$$

denote the result as D' and include it as one of the model's input channels. This design encourages the model to focus on building edge regions, thus improving its ability to learn signal reflection patterns.

4.2. ① SkipNet with Attention Gates

SkipNet [16], based on the U-Net architecture, is an encoder-decoder network with skip connections. The encoder progressively downsamples the input to extract hierarchical features, while the decoder upsamples and receives the hierarchical features from the encoder through skip connections to reconstruct the spatial resolution for final prediction. LocNet [12] took a similar architecture and validated the effectiveness in the localization task as well. In ViT-RefineNet [29], they incorporated attention gates into the SkipNet module, reducing the RMSE of the final prediction.

4.3. ② Patchify and Embedding

This module was proposed in original ViT [5], which divides an image into $\frac{256}{P} \times \frac{256}{P}$ patches of equal size $P \times P$. After patch embedding, we add a fixed 2D positional encoding over the patch grid. Let (x, y) be the patch coordinates on a $\frac{256}{P} \times \frac{256}{P}$ grid and d the embedding dimension. The subscript $[2k]$ (or $[2k+1]$) indexes the $(2k)$ -th (or $(2k+1)$ -th) element in the d -dimensional vector. We allocate half of

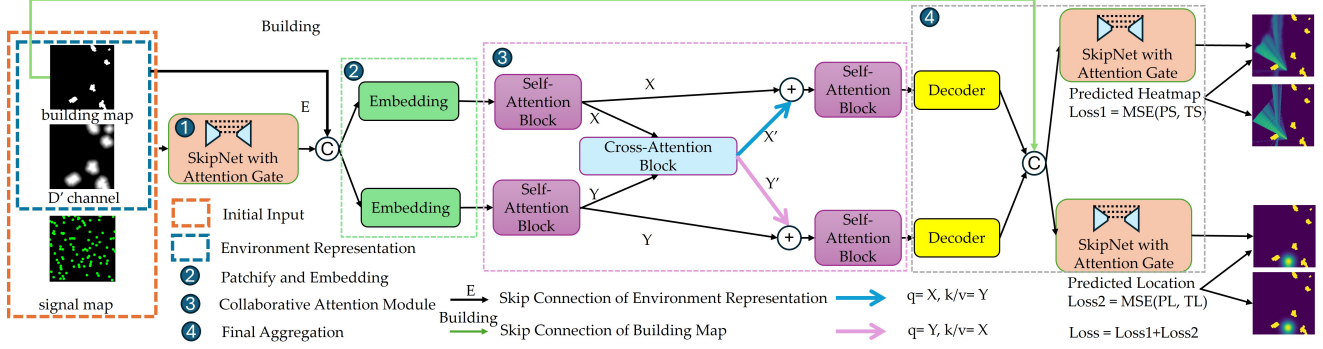


Figure 1. Proposed Framework for Radio Map Prediction. The outputs after cross-task fusion are denoted as X' and Y' . All the LayerNorm normalizations are replaced with Dynamic-Tanh. Modules at different positions do not share parameters.

the channels to x and the other half to y :

$$\begin{aligned} \text{for } k = 0, \dots, \frac{d_x}{2} - 1 : PE_x[2k] &= \sin\left(\frac{x}{10000^{2k/d_x}}\right), \\ PE_x[2k+1] &= \cos\left(\frac{x}{10000^{2k/d_x}}\right), \\ \text{for } k = 0, \dots, \frac{d_y}{2} - 1 : PE_y[2k] &= \sin\left(\frac{y}{10000^{2k/d_y}}\right), \\ PE_y[2k+1] &= \cos\left(\frac{y}{10000^{2k/d_y}}\right), \\ \text{and } PE(x, y) &= [PE_x; PE_y] \in \mathbb{R}^d, \end{aligned}$$

where $d_x = d_y = \frac{d}{2}$. (Note that $d \equiv 0 \pmod{4}$.)

4.4. ③ Collaborative Attention Module

The attention encoder blocks operate on embedded image patches. Let $X \in \mathbb{R}^{n \times d}$ denote the input sequence of embedded patches, where n is the number of patches and d is the embedding dimension. Before entering the attention layers, the sequence is normalized by a learnable dynamic-tanh function [34], defined as:

$$DT(X) = W \odot \tanh(\alpha X) + B, \quad (1)$$

where α is a learnable scalar and $\tanh(x) = \frac{e^{2x}-1}{e^{2x}+1}$ is the hyperbolic tangent function.

4.4.1. Self-Attention

The normalized patches are processed by a Multi-Head Self-Attention (MHSA) module, which models contextual dependencies within the input sequence. The standard query-key-value formulation is as follows:

$$Q = XW^Q, \quad K = XW^K, \quad V = XW^V,$$

with $W^Q, W^K, W^V \in \mathbb{R}^{d \times d_k}$. The scaled dot-product attention is:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right) V.$$

MHSA uses multiple parallel heads and concatenates their outputs:

$$\text{MultiHead}(X) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^O,$$

where each $\text{head}_i = \text{Attention}(Q_i, K_i, V_i)$ and $W^O \in \mathbb{R}^{hd_k \times d}$.

4.4.2. Cross-Attention

To enable interaction between the two prediction tasks, radio map reconstruction and transmitter localization, we introduce a task-level cross-attention module. Let X and Y denote the hidden representations for the radio map and localization heads, respectively. Cross-attention allows one task to use the features of the other as context:

$$Q_X = XW^Q, \quad K_Y = YW^K, \quad V_Y = YW^V,$$

so that the radio map head can attend to localization cues. Symmetrically, localization features attend to radio map representations:

$$Q_Y = YW^Q, \quad K_X = XW^K, \quad V_X = XW^V.$$

The attention outputs from all heads are also concatenated and linearly projected, and the final output is obtained by a residual connection followed by Dynamic-Tanh:

$$X' = DT(X + \text{MultiHead}(Q_X, K_Y, V_Y)), \quad (2)$$

$$Y' = DT(Y + \text{MultiHead}(Q_Y, K_X, V_X)). \quad (3)$$

This bi-directional interaction encourages complementary learning: the radio map task captures fine-grained propagation patterns, whereas the localization task enforces spatial consistency around the transmitter position. The output of the cross-attention module will go through another six self-attention blocks for further refinement.

4.4.3. Residual and Feed-Forward in Self/Cross Attention Blocks

Both self-attention and cross-attention outputs are followed by residual connections, Dynamic-Tanh normalization, and a feed-forward multilayer perceptron (MLP) block with nonlinear activation. Each sub-block uses an additional residual connection. Following [34], we replace traditional normalization with the dynamic-tanh regularization to improve stability and representation capacity.

Table 1. Definitions and value ranges of factors used to evaluate model performance.

Factors	Description	Possible Values
#samples (Ω)	Number of sampled points.	20, 40, 60, 80, 100
Positive Ratio (pr)	Proportion of nonzero signals.	Varies (0%–100%)
d_{mean}	Average distance from sampled pixels to the transmitter	$[0, 256\sqrt{2}]$

4.5. ④ Final Aggregation

After the collaborative attention module, following common practice in Vision Transformer–based dense prediction architectures, we employ a lightweight decoder (we denote this module as decoder in Figure 3) to restore spatial structure from the refined patch embeddings: each resulting patch embedding is linearly projected back to the pixel-level representation. These projected embeddings are then reshaped and rearranged into a $N \times N$ 2D grid to align with the original spatial layout, concatenated with the original building map, and fed to two different SkipNets for the final heatmap and localization prediction.

5. Experiment Design and Result

In this section, we conduct extensive experiments to evaluate the performance of our method. In the directional signal scenario, it is common to sample zero-signal pixels. Therefore, we evaluate the impact of the positive ratio (**pr**), defined as

$$\text{pr} = \frac{\text{\# of sampled pixels with nonzero signal strength}}{\text{\# of sampled pixels}} \times 100\%.$$

We seek answers to the following research questions, comparing with both radio map prediction methods and transmitter localization methods:

- **RQ1:** How does our model perform compared to the baseline models as the number of sampled points increases?
- **RQ2:** How does our model perform compared to the state-of-the-art models with different positive ratios.
- **RQ3:** How does our model perform compared to the state-of-the-art models over varying average distances between the sampled pixels and the transmitter?
- **RQ4:** How do the designs and modules in our model affect performance?

See Table 1 for the factors we plan to test.

5.1. Comparison Models

For the radio map generation task, we compare our proposed method with the following state-of-the-art models:

- **RadioUNet** [14]: A standard U-Net architecture characterized by its use of deep convolutional filters.
- **SkipNet** [16]: A compact U-Net variant, similar to RadioUNet, which utilizes a constant number of filters across its corresponding encoder and decoder layers.
- **GAN-CRME** [33]: A Conditional Generative Adversarial Network (c-GAN) designed specifically for lightweight and efficient generation of radio maps.



Figure 2. Visualization of the dataset

- **DC-Net** [3]: A model featuring cascaded U-Nets. It integrates Pyramid Spatial Pooling at each downsampling layer of the encoder to improve feature extraction, while the second U-Net in the cascade works to refine and correct the initial output.
- **ViT-RefineNet** [29]: A model using ViT as a backbone and SkipNet for multistep refinement, the state-of-the-art model for the directional signal dataset.

For the transmitter prediction task, we compare our proposed method with the following state-of-the-art models:

- **LocNet** [12]: A lightweight U-Net/SkipNet-style encoder–decoder that predicts transmitter heatmaps; uses *Focal Loss* to address class imbalance.
- **TL;DL** [19]: A U-Net-based image-to-image model with a large receptive field that maps sparse/noisy RSSI grids to transmitter heatmaps, followed by thresholding and local-maximum suppression to extract coordinates.
- **DSL** [15]: Addresses extreme sparsity via bias-correction and sparse-expansion preprocessing; adopts *HRNet* to preserve high-resolution features and replaces argmax with centroid regression for finer localization.

Table 2. Hyperparameter settings.

Hyperparameter	Setting
Learning Rate (SkipNets)	5e-4
Learning Rate (Other Modules)	1e-3
Batch Size	32
Number of Epochs	100
Optimizer	AdamW
Loss Function	MSE
Patch Size	8
Width	192
Number of Heads	16
Each Self-Attention Encoder’s Depth	6
MLP Ratio	4
Cross-Attention Encoder’s Depth	1
Image Size	256×256
σ (standard deviation of the Gaussian kernel)	17

5.2. Dataset and Training

We used the dataset provided in ViT-RefineNet paper [29] to train and evaluate our model. Each propagation map of the dataset is paired with multiple sampling maps to simulate different sensing conditions:

- **Training set:** Each map (27,600 in total) is paired with 10 sampling maps, totaling 276,000 training samples.
- **Validation set:** Each map (5,900 in total) is paired with 40 sampling maps, totaling 236,000 validation samples.
- **Test set:** Each map (5,900 in total) is paired with 100 sampling maps.

Table 3. Comparison across three metrics. All baselines were reproduced using their published implementations and the hyperparameters recommended in the original papers to ensure fairness. **Localization Error** and **RMSE**: lower is better; **SSIM**: higher is better. Best per row and in **bold**, and SOTA baselines are also highlighted in **bold** as reference for relative improvements. For Localization Error, SymNet shows % improvement vs. **DSLoc**; for **RMSE** and **SSIM**, SymNet shows % improvement vs. **ViT-RefineNet**.

# samples (random)	Localization Error (m)					RMSE					SSIM				
	SymNet (Δ vs DSLoc)	DC-Net	DSLoc	TL;DL	LocNet	SymNet (Δ vs ViT)	ViT-RefineNet	DC-Net	GAN-CRME	RadioUNet	SymNet (Δ vs ViT)	ViT-RefineNet	DC-Net	GAN-CRME	RadioUNet
20	27.891 $\downarrow$ 24.73%	59.705	37.056	69.779	61.316	0.082819 $\downarrow$ 0.38%	0.083132	0.088771	0.091769	0.103657	0.742213 $\uparrow$ 7.00%	0.693653	0.670583	0.617352	0.715857
40	14.142 $\downarrow$ 33.96%	45.051	21.416	64.182	39.912	0.062740 $\downarrow$ 3.82%	0.065234	0.071234	0.070210	0.086102	0.803986 $\uparrow$ 5.10%	0.764947	0.753524	0.733005	0.767339
60	9.635 $\downarrow$ 36.86%	37.503	15.259	62.899	29.779	0.054712 $\downarrow$ 4.95%	0.057563	0.063032	0.061899	0.077658	0.829305 $\uparrow$ 4.23%	0.795613	0.789412	0.773503	0.790645
80	7.538 $\downarrow$ 38.01%	33.124	12.159	63.079	24.183	0.050327 $\downarrow$ 5.47%	0.053239	0.058185	0.057349	0.072550	0.843697 $\uparrow$ 3.74%	0.813296	0.810267	0.795111	0.804493
100	6.367 $\downarrow$ 38.22%	30.144	10.306	63.688	20.643	0.047480 $\downarrow$ 5.80%	0.050402	0.054882	0.054414	0.069025	0.853431 $\uparrow$ 3.41%	0.825285	0.824221	0.809011	0.814019

In the dataset, the maximal possible signal is $S_{\max} = -4.5889$ dB, and a cutoff threshold of $S_{\min} = -120$ dB is applied, indicating that any signal strength below this threshold is treated as zero. Min–Max normalization is applied to the raw data and as follows:

$$\frac{S_{\text{true}} - S_{\min}}{S_{\max} - S_{\min}}.$$

For each propagation map, there is a corresponding antenna map, where 1 indicates transmitter pixels and 0 indicates non-transmitter pixels. We generate the ground truth heatmap for the localization head by convolving the binary antenna map with a Gaussian kernel ($\sigma = 17$) and normalizing the result to $[0, 1]$:

$$A(x, y) = \frac{(\delta * G_{\sigma})(x, y)}{\max_{x, y} (\delta * G_{\sigma})(x, y)}, G_{\sigma}(u, v) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{u^2 + v^2}{2\sigma^2}\right).$$

where $\delta(x, y)$ is the binary indicator of antenna locations, and G_{σ} is the two-dimensional Gaussian kernel.

The training settings are listed in Table 2. See Figure 2 for the visualization of the dataset.

5.3. Metric

To study the strength and shape of the signal propagation, we use **RMSE** to quantify intensity error and **SSIM** [28] to assess perceptual quality. **SSIM** is computed following the standard definition [28], using local Gaussian-weighted windows averaged over the image. Specifically,

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}, \quad \text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)},$$

where N is the number of non-building pixels, y_i is the predicted normalized signal value, and $\hat{y}_i$ is the normalized ground-truth signal value; where μ_x, μ_y are the mean values of x and y , σ_x^2, σ_y^2 are the variances, σ_{xy} is the covariance, and C_1, C_2 are small constants to stabilize the division. To measure the accuracy of transmitter localization, we use Euclidean distance from predicted location to ground truth location. We take the centroid of the prediction of the localization head as the predicted transmitter location.

5.4. Impact of Number of Sampled Points (RQ1)

See Table 3. From 20 to 100 sampled points, **SymNet** remains the best on localization/pixel-wise errors and is competitive or superior on structural similarity; moreover, its

advantages on error-sensitive metrics *increase* with sampling density.

Localization error becomes much lower. Across 20/40/60/80/100 points, SymNet reduces error by **34.36%** on average relative to DSLoc. The margin grows steadily with more samples, indicating stronger sample efficiency.

Pixel-wise error of heatmap prediction is also lower. SymNet improves over ViT-RefineNet by **4.08%** on average. The improvement is monotonic with sampling density, suggesting better performance as more information becomes available. It also improves over the omnidirectional SOTA DC-Net by 15% on average, demonstrating that our design is better suited to directional propagation scenarios where angular sensitivity and multipath effects dominate.

Structural similarity of heatmap prediction becomes higher. Compared with ViT-RefineNet, SymNet achieves average **+4.70% SSIM**.

In summary, at low sampling (20 points), **SymNet** achieves the largest *absolute* localization gain—about **10 m** vs. DSLoc, and the largest **SSIM** gains vs. ViT-RefineNet. As sampling increases, the absolute localization accuracy gap narrows, but SymNet *consistently outperforms* baselines on error metrics; **RMSE** and **SSIM** improve consistently; **RMSE** decreases more markedly, while **SSIM** increases less as the number of sampled points grows. We therefore emphasize the strong low-sample advantage and consistent superiority. *Crucially, DC-Net is an omnidirectional SOTA; when transferred to our directional setting, both localization and heatmap generation degrade substantially compared with our model* (e.g., 20 pts: 59.70 m vs. 27.89 m; 100 pts: 30.14 m vs. 6.37 m— $\sim 2.1 \times$ – $4.7 \times$ higher error), and it also trails on pixel-wise error and **SSIM** (e.g., 20 pts **RMSE** 0.0888 vs. 0.0828; 100 pts **SSIM** 0.8277 vs. 0.8534). This highlights a *task mismatch*: architectures tuned for omnidirectional propagation (which can reach sub-1 m on their native task) transfer poorly to directional settings.

5.5. Impact of Positive Ratio (RQ2)

See Table 4. Our model outperforms all the SOTAs in almost every scenario. In the localization task, the prediction error of our model reduces by 30% compared to the local-

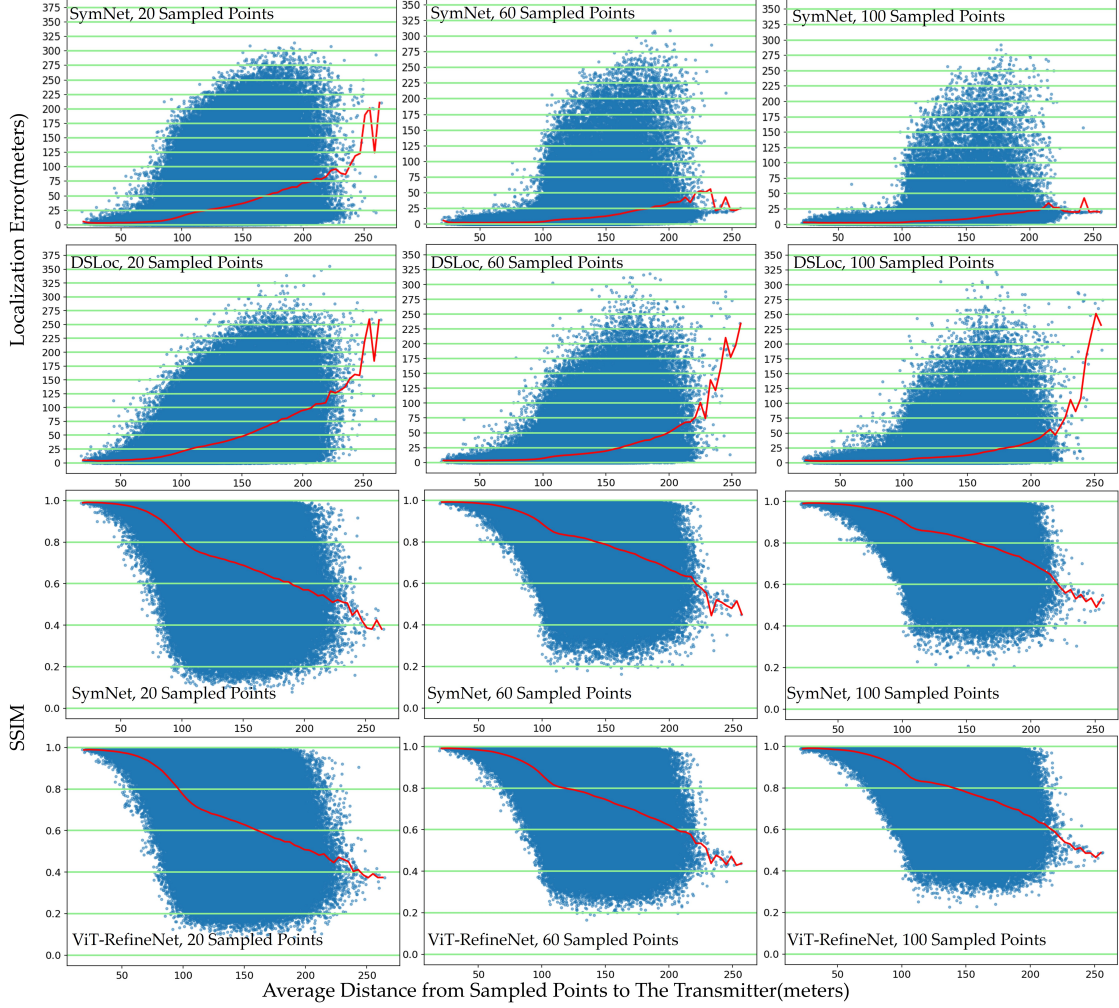


Figure 3. Localization error/ **SSIM** versus the average distance from sampled points to the transmitter (d_{mean}) under 20 and 100 sampling points. Blue dots denote individual samples, and the red line represents the binned average Localization error/SSIM.

Table 4. Performance by positive-ratio bins. Localization Error & **RMSE**: lower is better; **SSIM**: higher is better. # of samples $\in [20, 100]$. ViT-R.N. refers to ViT-RefineNet [29].

pr (%)	↑ SSIM		↓ Localization Error (m)		↓ RMSE				
	SymNet	ViT-R.N.	SymNet	DSLoc	SymNet	ViT-R.N.			
[0, 10)	0.702	↑ 8.56%	0.647	43.029	↓ 15.98%	51.207	0.099	↑ 3.46%	0.096
[10, 20)	0.775	↑ 5.89%	0.732	19.810	↓ 32.77%	29.475	0.072	↓ 2.46%	0.074
[20, 30)	0.806	↑ 5.01%	0.767	12.348	↓ 40.16%	20.632	0.062	↓ 4.89%	0.065
[30, 40)	0.823	↑ 4.49%	0.787	9.069	↓ 43.32%	15.997	0.056	↓ 5.88%	0.059
[40, 50)	0.834	↑ 4.16%	0.801	7.255	↓ 44.79%	13.135	0.052	↓ 6.46%	0.056
[50, 60)	0.843	↑ 3.90%	0.811	6.150	↓ 45.06%	11.190	0.050	↓ 6.62%	0.053
[60, 70)	0.849	↑ 3.72%	0.818	5.432	↓ 44.63%	9.810	0.048	↓ 6.72%	0.052
[70, 80)	0.854	↑ 3.55%	0.824	4.952	↓ 43.74%	8.806	0.047	↓ 6.73%	0.050
[80, 90)	0.858	↑ 3.40%	0.829	4.625	↓ 42.50%	8.039	0.046	↓ 6.66%	0.049
[90, 100]	0.861	↑ 3.27%	0.834	4.362	↓ 41.12%	7.412	0.045	↓ 6.53%	0.048

ization SOTA DSLoc when the positive ratio is greater than 10%; in the radio map reconstruction task, the **RMSE** of our model reduces by 5% on average and **SSIM** increases by 5% on average, demonstrating that our model’s robustness across different positive ratios.

5.6. Impact of Average Distance from Sampled Points to Transmitter (RQ3)

For the localization problem, as shown in Figure 3, our method exhibits reduced sensitivity to sampling distance. When all sampling points are located within 200 m on average, the averaged localization error remains within 75 m under 20 sampling points, while DSLoc exceeds 100 m under the same conditions. When 60 or 100 points are sampled, the average localization error is roughly bounded by 50 m, while DSLoc exceeds 75 m in the same conditions. This indicates that our approach maintains robustness in far-field scenarios where the SOTA method degrades significantly.

For the radio map generation task, we only show how **SSIM** changes when the average distance from sampled points to the transmitter is increasing, **RMSE** follows the same pattern. The results in Figure 3 demonstrate that SymNet is more robust when increasing the transmitter distance compared to ViT-RefineNet: with only 20 sampled points,

Table 5. Results of the ablation study across three metrics. No D' shows % improvement vs. original SymNet.

# samples	↓ RMSE					↑ SSIM					↓ Localization Error (m)				
	SymNet	No D'	Single-task	2-CAB	0-CAB	SymNet	No D'	Single-task	2-CAB	0-CAB	SymNet	No D'	Single-task	2-CAB	0-CAB
20	0.0828	0.0853 ↑ 2.97%	0.1639	0.1023	0.1646	0.742	0.763 ↑ 2.82%	0.0987	0.6535	0.0888	27.891	29.503 ↑ 5.78%	91.948	56.625	91.814
40	0.0627	0.0659 ↑ 4.96%	0.1490	0.0908	0.1496	0.804	0.814 ↑ 1.31%	0.1259	0.6774	0.1149	14.142	15.490 ↑ 9.54%	82.136	47.725	81.154
60	0.0547	0.0578 ↑ 5.60%	0.1342	0.0852	0.1352	0.829	0.836 ↑ 0.80%	0.1704	0.6834	0.1561	9.635	10.721 ↑ 11.27%	74.767	43.189	73.486
80	0.0503	0.0533 ↑ 5.99%	0.1204	0.0816	0.1215	0.844	0.848 ↑ 0.54%	0.2423	0.6852	0.2203	7.538	8.518 ↑ 13.00%	68.848	40.298	67.477
100	0.0475	0.0504 ↑ 6.22%	0.1082	0.0790	0.1094	0.853	0.857 ↑ 0.37%	0.3769	0.6855	0.3337	6.367	7.317 ↑ 14.91%	62.728	38.319	61.652

the average SSIM of SymNet remains above 0.6 until 200 m, whereas ViT-RefineNet already drops to the same level beyond 170 m. The behavior of two models in other scenarios are similar. This indicates that SymNet preserves structural similarity more effectively in far-field scenarios, highlighting its stronger generalization under sparse and distant sampling conditions.

5.7. Ablation Study (RQ4)

We performed an ablation study to assess the contribution of each design. To align with real-world conditions, we restricted the analysis to random sampling and varied the number of sampled points. We then examined the effect of the key modifications:

- **No D'**: removing the D' channel.
- **0/2 Cross-Attention Blocks(CABs)**: removing all cross-attention blocks to test whether cross-task interaction is necessary, or doubling them with separate projection matrices to test whether the two tasks should share the same projection and whether it is better to have more cross-attention blocks.
- **Single-Task Path**: removing one branch of the joint architecture and training two separate models for radio map reconstruction and localization, to examine whether joint learning is essential for convergence.

The ablation study result in Table 5 shows that removing the D' channel leads to a slight improvement in SSIM (up to +2.8%), but at the cost of significant increases in localization error and RMSE (up to +15%), indicating that D' contributes to reducing reconstruction error while preserving structural similarity. Variants altering the cross-attention design converge only to degenerate solutions with substantially worse results. In 0-CAB, the lack of complementary guidance leads to trivial convergence, yielding behavior similar to the single-task path variant. In 2-CAB, the independent projections create inconsistent feature spaces and unstable optimization.

6. Conclusion

In this work, we proposed SymNet, a cross-attention-based architecture for directional radio map reconstruction and transmitter localization. Through extensive experiments, we demonstrated that SymNet achieves superior accuracy

and robustness compared to state-of-the-art baselines such as ViT-RefineNet. Notably, SymNet remains less sensitive to the distance of sampling points: even when the average distance of sampled points exceeds 100 m, the model maintains stable localization errors and structural similarity, whereas competing methods degrade sharply.

Our ablation study confirmed the importance of the joint design. Removing the D' channel results in a marginal SSIM gain but simultaneously increases localization error and RMSE by a considerable margin, underscoring its role in balancing structural fidelity and numerical precision.

Other simplified variants, including modifications to the cross-attention block, separated single-task training, or alternative head designs, converged to ineffective solutions under the same training conditions. This indicates that designs are essential for effective learning.

Overall, these findings highlight a previously overlooked challenge in wireless sensing: in directional propagation, the strongest signal often deviates from the transmitter location, making radio map reconstruction and localization complementary rather than interchangeable tasks. SymNet preserves both fine-grained structural information and global localization accuracy under sparse and distant sampling conditions, providing a robust and scalable solution for real-world deployment. Future work will extend SymNet to multi-transmitter scenarios and investigate integration with physical propagation priors to further enhance generalization.

7. Acknowledgment

Research was partially sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-23-2-0014. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

References

- [1] P. Bahl and V.N. Padmanabhan. Radar: an in-building rf-based user location and tracking system. In *Proceedings IEEE INFOCOM 2000. Conference on Computer Communications. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies (Cat. No.00CH37064)*, pages 775–784 vol.2, 2000. 2
- [2] Gabriele Boccolini, Gustavo Hernández-Peñaloza, and Baltasar Beferull-Lozano. Wireless sensor network for spectrum cartography based on kriging interpolation. In *2012 IEEE 23rd International Symposium on Personal, Indoor and Mobile Radio Communications - (PIMRC)*, pages 1565–1570, 2012. 2
- [3] Qi Chen, Ming Huang, and JingJing Yang. Dc-net: A distant-range content interaction network for radio map construction. *ICT Express*, 10(5):1145–1150, 2024. 1, 2, 5
- [4] Qi Chen, Jingjing Yang, Ming Huang, and Qiang Zhou. Act-gan: Radio map construction based on generative adversarial networks with act blocks. *IET Communications*, 18(19): 1541–1550, 2024. 2
- [5] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021. 3
- [6] Sinan Gezici. A survey on wireless position estimation. *Wireless personal communications*, 44(3):263–282, 2008. 2
- [7] Fredrik Gustafsson and Fredrik Gunnarsson. Mobile positioning using wireless networks: possibilities and fundamental limitations based on available wireless network measurements. *IEEE Signal processing magazine*, 22(4):41–53, 2005. 2
- [8] Andreas Haeberlen, Eliot Flannery, Andrew M Ladd, Algis Rudys, Dan S Wallach, and Lydia E Kavvaki. Practical robust localization over large-scale 802.11 wireless networks. In *Proceedings of the 10th annual international conference on Mobile computing and networking*, pages 70–84, 2004. 2
- [9] Mohamed Hamid and Baltasar Beferull-Lozano. Non-parametric spectrum cartography using adaptive radial basis functions. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3599–3603, 2017. 2
- [10] Seung-Jun Kim and Georgios B. Giannakis. Cognitive radio spectrum prediction using dictionary learning. In *2013 IEEE Global Communications Conference (GLOBECOM)*, pages 3206–3211, 2013. 2
- [11] Enes Krijestorac, Samer Hanna, and Danijela Cabric. Spatial signal strength prediction using 3d maps and deep learning. In *ICC 2021-IEEE international conference on communications*, pages 1–6. IEEE, 2021. 2
- [12] Thanh Dat Le and Yan Huang. Localization through deep learning in new and low sampling rate environments. In *Advances in Knowledge Discovery and Data Mining*, pages 297–308, Singapore, 2024. Springer Nature Singapore. 1, 2, 3, 5
- [13] Ju-Hyung Lee and Andreas F. Molisch. A scalable and generalizable pathloss map prediction. *IEEE Transactions on Wireless Communications (TWC)*, 2024. 2
- [14] Ron Levie, cCaugkan Yapar, Gitta Kutyniok, and G. Caire. Radiounet: Fast radio map estimation with convolutional neural networks. *IEEE Transactions on Wireless Communications*, 20:4001–4015, 2019. 1, 2, 5
- [15] Runjie Liu, Qionggui Zhang, Yuankang Zhang, Rui Zhang, and Tao Meng. Deep learning-based transmitter localization in sparse wireless sensor networks. *Sensors*, 24(16), 2024. 1, 2, 5
- [16] William Locke, Nikita Lokhmachev, Yan Huang, and Xinrong Li. Radio map estimation with deep dual path autoencoders and skip connection learning. In *2023 IEEE 34th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, pages 1–6, 2023. 1, 2, 3, 5
- [17] Danlei Mao, Wei Shao, Zuping Qian, Hong Xue, Xin Lu, and Hongsheng Wu. Constructing accurate radio environment maps with kriging interpolation in cognitive radio networks. In *2018 Cross Strait Quad-Regional Radio Science and Wireless Technology Conference (CSQRWC)*, pages 1–3, 2018. 2
- [18] Ahmed Marey, Mustafa Bal, Hasan F. Ates, and Bahadır K. Gunturk. Pl-gan: Path loss prediction using generative adversarial networks. *IEEE Access*, 10:90474–90480, 2022. 2
- [19] Frost Mitchell, Aniqua Baset, Neal Patwari, Sneha Kumar Kasera, and Aditya Bhaskara. Deep learning-based localization in limited data regimes. In *Proceedings of the 2022 ACM Workshop on Wireless Security and Machine Learning*, page 15–20, New York, NY, USA, 2022. Association for Computing Machinery. 1, 2, 5
- [20] Ozan Ozyegen, Sanaz Mohammadjafari, Mucahit Cevik, Karim El Mokhtari, Jonathan Ethier, and Ayse Basar. An empirical study on using cnns for fast radio signal prediction. *SN Computer Science*, 3(2):131, 2022. 2
- [21] Neal Patwari, Joshua N Ash, Spyros Kyperountas, Alfred O Hero, Randolph L Moses, and Neiyer S Correal. Locating the nodes: cooperative localization in wireless sensor networks. *IEEE Signal processing magazine*, 22(4):54–69, 2005. 2
- [22] Kehai Qiu, Stefanos Bakirtzis, Hui Song, Ian Wassell, and Jie Zhang. Deep learning-based path loss prediction for outdoor wireless communication systems. In *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–2, 2023. 2
- [23] Feng Shen, Zheng Wang, Guoru Ding, Kezhi Li, and Qihui Wu. 3d compressed spectrum mapping with sampling locations optimization in spectrum-heterogeneous environment. *IEEE Transactions on Wireless Communications*, 21(1):326–338, 2022. 2
- [24] Yves Teganya and Daniel Romero. Deep completion autoencoders for radio map estimation. *IEEE Transactions on Wireless Communications*, 21(3):1710–1724, 2021. 2
- [25] Yu Tian, Shuai Yuan, Weisheng Chen, and Naijin Liu. Transformer based radio map prediction model for dense urban

- environments. In *2021 13th International Symposium on Antennas, Propagation and EM Theory (ISAPE)*, pages 1–3, 2021. [2](#)
- [26] Satya Kumar Vankayala, Swaraj Kumar, Ishaan Roy, D Thirumulanathan, Seungil Yoon, and Ignatius Samuel Kanakaraj. Radio map estimation using a generative adversarial network and related business aspects. In *2021 24th International Symposium on Wireless Personal Multimedia Communications (WPMC)*, pages 1–6, 2021. [2](#)
- [27] Wenyu Wang, Lei Zhu, Zhen Huang, Baozhu Li, Lu Yu, and Kaixin Cheng. Mt-gcnn: multi-task learning with gated convolution for multiple transmitters localization in urban scenarios. *Sensors*, 22(22):8674, 2022. [2](#)
- [28] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4): 600–612, 2004. [6](#)
- [29] Lyuzhou Ye, Thanh Dat Le, and Yan Huang. ViT-RefineNet for Directional Signal Radio Map Reconstruction from Sparse Samples. In *Proceedings of the 33rd ACM International Conference on Advances in Geographic Information Systems (SIGSPATIAL '25)*, New York, NY, USA, 2025. Association for Computing Machinery. [2](#), [3](#), [5](#), [7](#)
- [30] Caitao Zhan, Mohammad Ghaderibaneh, Pranjal Sahu, and Himanshu Gupta. Deepmtl: Deep learning based multiple transmitter localization. In *2021 IEEE 22nd International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, pages 41–50. IEEE, 2021. [2](#)
- [31] Wei Zhang, Kan Liu, Weidong Zhang, Youmei Zhang, and Jason Gu. Deep neural networks for wireless localization in indoor and outdoor environments. *Neurocomputing*, 194: 279–287, 2016. [2](#)
- [32] Zezhong Zhang, Guangxu Zhu, Juntong Chen, and Shuguang Cui. Fast and accurate cooperative radio map estimation enabled by gan. *2024 IEEE International Conference on Communications Workshops (ICC Workshops)*, pages 1641–1646, 2024. [1](#)
- [33] Yi Zheng, Ji Wang, Xingwang Li, Jiping Li, and Shouyin Liu. Cell-level rsrp estimation with the image-to-image wireless propagation model based on measured data. *IEEE Transactions on Cognitive Communications and Networking*, 9(6): 1412–1423, 2023. [2](#), [5](#)
- [34] Jiachen Zhu, Xinlei Chen, Kaiming He, Yann LeCun, and Zhuang Liu. Transformers without normalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025. [4](#)
- [35] Anatolij Zubow, Suzan Bayhan, Piotr Gawłowicz, and Falko Dressler. Deeptxfinder: Multiple transmitter localization by deep learning in crowdsourced spectrum sensing. In *2020 29th International Conference on Computer Communications and Networks (ICCCN)*, pages 1–8. IEEE, 2020. [2](#)