A Graph Neural Network-Based Radio Map Construction With Uncertainty Prediction

Fatema Islam Tania Seung-Jun Kim

Department of Computer Science and Electrical Engineering
University of Maryland, Baltimore County, Baltimore, MD 21250
E-mail:{fatemat1,sjkim}@umbc.edu

***Abstract*—A graph neural network (GNN)-based method is proposed for radio map construction and uncertainty prediction in urban environment. Based on sparse measurements, the received signal strength (RSS) values are interpolated over the entire area. A graph is constructed using the city map and base station (BS) locations to capture the spatial correlation pattern of the RSS. Our proposed graph construction rule allows faithful reconstruction of the sharp transitions in the RSS values across line-of-sight (LOS) and non-line-of-sight (NLOS) regions. The uncertainty levels associated with the RSS estimates are jointly predicted using the negative log-likelihood (NLL) training cost. Numerical tests show that the proposed algorithm achieves up to a 30% reduction in root mean square error (RMSE) compared with an existing GNN-based approach, and significantly outperforms a convolutional neural network (CNN)-based method with a comparable parameter count. Furthermore, the uncertainty estimates are readily interpretable, typically assigning higher uncertainty in areas with larger prediction errors or fewer available samples.**

I. Introduction

In wireless communication systems, accurate prediction of the received signal strength (RSS) in the deployment area is important for various network optimization tasks such as power/interference control, spectrum sharing, multihop routing, and path planning of unmanned aerial vehicles (UAVs). It is also essential for long-term network planning for better coverage and quality of service.

However, the radio propagation characteristics are complex, especially in urban environments. The non-line-of-sight (NLOS) condition caused by obstacles can lead to significant variations in the RSS even in nearby locations. This creates highly non-smooth spatial patterns, making the radio map construction problem challenging, being sensitive to measurement density, while also causing high prediction uncertainty. Properly capturing the spatial correlation pattern of the RSS map can thus significantly enhance the map construction performance. Furthermore, providing not only the predictions but also the associated uncertainty is instrumental in guiding the subsequent decision-making process [1].

Recent deep learning approaches, such as convolutional neural network (CNN) [2], transformer [3], and graph neural network (GNN) [4], have shown promise in capturing spatial relationships. In [5], a lightweight three-layer CNN is designed that can effectively interpolate radio maps from sparse measurements by learning both global path-loss and local shadowing patterns. Another CNN based approach is proposed in [6], where each radio map is treated as an image and a three-layer CNN is trained to perform spatial interpolation from sparse measurements. While these models reduce reconstruction error, they rely solely on sparsely sampled measurement data and do not incorporate information on the propagation environmental structure or the LOS/NLOS propagation conditions. A deep neural network-based method in [1] used a 3D map of the urban area with measurement data to predict both the RSS and the associated uncertainty. Another study in [7] deployed a two-stage first-predict-then-correct (FPTC) framework, where an environment-driven generative adversarial network (GAN) first produces a coarse map and a second GAN corrects it using sparse measurements. These works demonstrate that combining environmental priors with limited on-site samples can significantly boost reconstruction accuracy.

GNN-based methods can naturally leverage spatial dependency as priors. A graph attention network (GAT) was employed in [8] for multiband radio map reconstruction, where the transmitter position information was utilized for graph construction. In [4], an undirected graph was constructed from the city map and a GNN was employed to predict the RSS values from sparse RSS measurements, achieving higher accuracy and robustness than traditional interpolation and CNN-based methods. A 3D city map reconstruction problem was tackled in [9] using radio measurements and the LOS/NLOS states of ground users. The last two works also highlight that the LOS/NLOS statuses carry useful geometric and propagation information that can significantly enhance RF propagation environment modeling.

Building on these observations, our contribution is to markedly improve the GNN-based RSS map construction in two key aspects. First, an improved graph construction rule is proposed so that sharp transition edges in the RSS map can be better predicted. Secondly, the RSS values and their uncertainties are predicted jointly by training the model using a negative log-likelihood (NLL) cost function [10]. The numerical tests show that the proposed method significantly improves RSS prediction, readily capturing sharp transitions near LOS/NLOS borders, while providing reliable and intuitive uncertainty estimates.

The rest of the paper is organized as follows. The system model is described in Sec. II. The proposed method, including graph construction and NLL-based training, is delineated in Sec. III. The results from numerical tests are presented in Sec. IV. The conclusion is provided in Sec. V.

This work was supported by NSF grant 2242412.

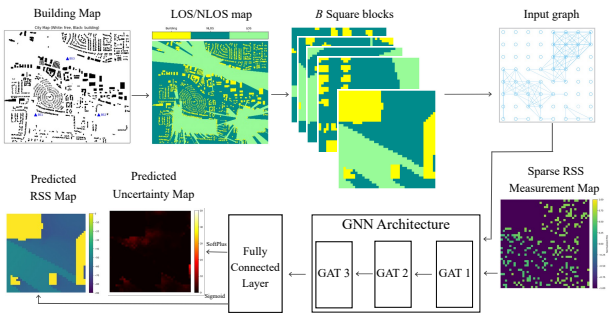


Fig. 1: Diagram of the proposed method.

II. System Model and Problem Formulation

The received signal power P_R in dBm at a location distance d away from a base station (BS) can be expressed as

$$P_R(d) = P_T + G - PL(d) + \xi \tag{1}$$

where P_T is the transmission power in dBm, $PL(d)$ is the path loss, G represents the combined antenna gain of the transmitter and the receiver, and ξ denotes the shadow fading. The path loss PL can be modeled as a function of distance as

$$PL(d) = K + 10\gamma \log_{10} d \tag{2}$$

where K is a path loss constant, and γ the path loss exponent. To accurately predict $P_R(d)$, the LOS/NLOS condition of the link is important, as it strongly affects the shadowing ξ.

Consider a rectangular-shaped urban area of interest. The area has $N_x \times N_y$ grid points with the uniform grid interval d_{grid}. The area is subdivided into B square blocks of grid size $n \times n$. An $n \times n$ sparse measurement matrix is formed for each block with RSS measurements at the α fraction of the n^2 grid points, randomly selected from the outdoor grid locations. Then, the measurement matrix looks like

$$\tilde{\mathbf{R}} = \begin{bmatrix} \tilde{r}_{1,1} & \tilde{r}_{1,2} & \cdots & \cdot \\ \tilde{r}_{2,1} & \cdot & \cdots & \tilde{r}_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ \cdot & \tilde{r}_{n,2} & \cdots & \tilde{r}_{n,n} \end{bmatrix} \tag{3}$$

where $\tilde{r}_{i,j}$ denotes the RSS measurement at the (i,j) grid location and $\cdot$ represents the lack of a measurement at the grid point.

Our goal is to learn an interpolator f_θ that can map the sparse RSS observations $\tilde{\mathbf{R}}$ to the predicted RSS values and their uncertainties at all grid points. For training of f_θ, it is assumed that a data set of B_{tr} blocks with dense (i.e., n^2) measurements is available. It is also assumed that one can utilize the known BS location and a building map of the deployment area.

III. Proposed Method

Our approach is to train a GAT based on a graph that is constructed based on the LOS/NLOS conditions of the grid locations, which can be inferred from the building map. The network is then trained using a NLL cost function to jointly predict the RSS values and their uncertainties. The overall design of the proposed method is illustrated in Fig. 1.

A. Graph Construction

For each block with n^2 grid points, an undirected weighted graph $G = (\mathcal{V}, \mathcal{E}, \mathcal{W})$ is constructed, where each grid point corresponds to a node in $\mathcal{V}$, and the construction of the edge set $\mathcal{E}$ and the edge weights $\mathcal{W}$ will be explained shortly. The node attributes consist of the normalized RSS measurements given for node $v_k \in \mathcal{V}$ as

$$r_k = \begin{cases} \dfrac{\tilde{r}_k - \tilde{r}_{\min}}{\tilde{r}_{\max} - \tilde{r}_{\min}}, & \text{if } v_k \text{ has measurement } \tilde{r}_k \\ -1, & \text{otherwise} \end{cases} \tag{4}$$

where $\tilde{r}_{\min}$ and $\tilde{r}_{\max}$ are the minimum and the maximum RSS values, respectively. Note that $r_k \in [0,1]$ if a measurement is available and $r_k = -1$ if no measurement is available.

The edge set $\mathcal{E}$ can encode the spatial dependency of the RSS values between pairs of nodes. This can be inferred from the building map of the area, which reveal the locations of the obstacles in the propagation paths. In [4], an edge construction rule is proposed, given by

$$(v_l, v_k) \in \mathcal{E} \text{ if } \begin{cases} 1)\ d_{l,k} \leqslant d_{\text{th}}; \text{and} \\ 2)\ \text{no obstacles between } v_l \text{ and } v_k \end{cases} \tag{5}$$

where $d_{l,k}$ represents the Euclidean distance between grid points v_l and v_k. That is, nodes v_l and v_k are connected only if their distance is closer than a threshold d_{th} and the two nodes have a LOS path between them according to the building map. This rule reflects the intuition that if two nodes are far away or there are obstacles between them, their RSS values would tend to decorrelate.

However, this edge rule does not directly take into account the LOS or NLOS propagation states of each grid location relative to the serving BS. Therefore, sharp transitions in the RSS values across the LOS/NLOS boundaries are not well captured in the constructed edges. To mitigate this, we first classify each grid point as either in the LOS area of the BS or in the NLOS area. If there is no building obstructing the direct path from the BS to the grid point, the point is considered in the LOS condition. Otherwise, it is a NLOS location. If there are multiple BSs, one that is the nearest to the grid point is considered. Thus, our proposed edge construction rule is

$$(v_l, v_k) \in \mathcal{E} \text{ if } \begin{cases} 1)\ d_{l,k} \leqslant d_{\text{th}}; \\ 2)\ \text{no obstacles between } v_l \text{ and } v_k; \text{and} \\ 3)\ v_l, v_k \text{ have common LOS/NLOS states.} \end{cases} \tag{6}$$

In the experimental results in Sec. IV, this new edge rule is seen to significantly improve the prediction performance, compared to the old edge rule (5).

The correlation structure associated with the distance $d_{l,k}$ can be further captured through the edge weights of the graph. A Gaussian kernel is adopted to compute the edge weights. If $(v_l, v_k) \in \mathcal{E}$, the corresponding weight $w_{l,k}$ is given by

$$w_{l,k} = \exp\left(-\frac{d_{l,k}^2}{2\sigma_d^2}\right) \tag{7}$$

where σ_d^2 is a scale parameter that controls the rate at which the edge weights decay with distance.

B. Graph Attention Network

The graph attention network v2 (GATv2) extends the original GAT framework by introducing a dynamic attention mechanism [11]. Let $\mathcal{N}(v)$ denote the set of neighbors of node $v \in \mathcal{V}$ including v itself. Also, the feature vector for node v_l is denoted as $\mathbf{h}_l$. Suppose that H attention heads are employed. For each edge connecting node v_l to its neighbor $v_k \in \mathcal{N}(v_l)$, the attention score of the h-th attention head is computed as

$$e_h(v_l, v_k) = \mathbf{a}_h^T \,\text{LeakyReLU}\left([\mathbf{W}_h \mathbf{h}_l \| \mathbf{W}'_h \mathbf{h}_k \| \mathbf{w}_h w_{l,k}]\right) \tag{8}$$

where $\cdot^T$ denotes transposition, and $\mathbf{W}_h$, $\mathbf{W}'_h$, $\mathbf{w}_h$, and $\mathbf{a}_h$ are learnable weight matrices/vectors. The normalized attention coefficient for $v_k \in \mathcal{N}(v_l)$ is computed using a softmax operation over v_l's neighbors $\mathcal{N}(v_l)$ as

$$\alpha_{l,k}^{(h)} = \frac{\exp(e_h(v_l, v_k))}{\sum\limits_{v_{k'} \in \mathcal{N}(v_l)} \exp(e_h(v_l, v_{k'}))}. \tag{9}$$

Combining H attention heads through averaging, the output feature vector $\mathbf{h}'_l$ for node v_l is given by

$$\mathbf{h}'_l = \sigma\left(\frac{1}{H}\sum_{h=1}^{H} \sum_{v_k \in \mathcal{N}(v_l)} \alpha_{l,k}^{(h)} \mathbf{W}'_h \mathbf{h}_k\right) \tag{10}$$

where $\sigma(\cdot)$ denotes a nonlinear activation function.

The operations in (8)–(10) are repeated over L layers. In the first layer, the input features are the normalized RSS measurements in (4). In the subsequent layers, the outputs from the previous layer are used as the input features.

C. NLL-based Training

The output from the GATv2 is processed using a lightweight fully-connected linear layer to produce two outputs. One is for predicting the RSS values, and the other for the associated uncertainty. The prediction is done only for the grid points that reside outside any buildings. Thus, for a non-building grid point v_k, the RSS prediction $\hat{r}_k$ and the associated variance σ_k^2 are generated. The positivity of the variance is enforced by applying a softplus nonlinearity to the raw output s_k, along with a small constant ϵ for numerical stability.

$$\sigma_k^2 = \log\left(1 + e^{s_k}\right) + \epsilon. \tag{11}$$

To train the overall architecture, recall that the availability of dense measurements is assumed in B_{tr} training blocks. The training is done using a Gaussian NLL loss function, which is defined for a given block as

$$\mathcal{L} = \frac{1}{N} \sum_{\text{non-building } v_k} \left[\frac{(\hat{r}_k - r_k)^2}{2\,\sigma_k^2} \;+\; \frac{1}{2}\,\log \sigma_k^2\right] \tag{12}$$

where N is the total number of non-building nodes in the block.

In practice, to improve the stability of the NLL-based training, a couple of tricks can be employed. First, the model is pretrained only for RSS prediction using a mean square error (MSE) loss function. After the RSS branch becomes reasonably stabilized, the variance branch can be activated and the whole model is fine-tuned with the NLL loss. In addition, we found that the errors associated with the nodes with very few neighbors tend to dominate the overall performance metric. To mitigate this, disparate weighting can be applied for the terms in (12). For instance, the terms for the nodes whose degrees are below a threshold θ are weighted higher than others. This strategy encourages to put more effort on improving the accuracy of the sparsely connected nodes, which also enhances the overall training stability.

Hyperparameter	Value
GNN layers	3
No. of attention heads per layer	(4, 4, 4)
Hidden channels per layer	(64, 64, 40)
Activation function	ReLU
Batch size	8
Training epochs	200
Optimizer	AdamW
Learning rate	0.002
Learning rate scheduler	ReduceLROnPlateau

TABLE I: Model and training hyperparameters.

IV. Numerical Tests

A. Setup

The BART-Lab Radiomap Dataset[1] is used to evaluate the proposed approach. The dataset provides ground truth RSS maps for 2,000 dense urban areas in 5 frequency bands with size around 500×500, generated using the Altair Feko simulator. In this study, we used the maps for the 1750 MHz band for three urban regions located in Davis, California; Weddington, North Carolina; and Kalamazoo, Michigan. Each area contains three BSs and the region of interests are uniformly discretized into grids of 328×369, 492×533, and 451×615 respectively. The transmission power is 46 dBm. The area is divided into $B = 393$ blocks each with grid size 41×41. Among these, 273 blocks are used for training, 60 for validation, and the remaining 60 blocks for testing.

The hyperparameters for the neural network model and training are listed in Table I. The value of d_{th} is set to $3\sqrt{2}$, σ_d^2 to 1 and θ to 10. In a given block, out of the $n^2 = 41^2$ grid points, a fraction α located in the outdoor areas was randomly selected and assigned RSS measurements. The remaining ones were assumed to have no measurements. In the experiments, α was varied in the range of $[5, 30]\%$.

To evaluate the prediction performance, two test metrics are used, namely, the root mean squared error (RMSE) and the average NLL. The RMSE is computed as

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{\text{non-building } v_k} \left(r_k - \hat{r}_k\right)^2} \tag{13}$$

and the average NLL is calculated using (12). The metrics are averaged over all 60 test blocks.

To comparatively assess the performance of the proposed method, a number of baseline methods are considered. First, the proposed method is trained with the MSE loss with the uncertainty prediction branch deactivated. Secondly, the same neural network model as the proposed method is employed but the graph is constructed using the conventional edge rule in (5). This model is trained using either the MSE loss (for RSS prediction only) or the NLL loss (for both RSS and uncertainty prediction).

[1] https://github.com/BRATLab-UCD/Radiomap-Data

Sampling Rate	New +MSE	New +NLL	Old +MSE	Old +NLL	SICNN +MSE	SICNN +NLL
5%	1.4146	1.4129	2.0222	2.0166	2.4366	2.0665
10%	1.1702	1.1988	1.6743	1.7211	1.8806	1.8071
15%	1.0521	1.0431	1.5523	1.5207	1.8970	1.7101
20%	1.0137	0.9699	1.3754	1.3932	1.7024	1.5793
25%	0.9631	0.9108	1.3313	1.3159	1.5895	1.5171
30%	0.9126	0.8621	1.2078	1.1958	1.6274	1.4480

TABLE II: RMSE comparison across sampling rates.

Sampling Rate	New+NLL	Old+NLL	SICNN+NLL
5%	0.6487	0.9040	1.0823
10%	-0.0032	0.4459	0.9917
15%	-0.2481	0.1764	0.9302
20%	-0.3741	0.0415	0.8499
25%	-0.3994	-0.0322	0.8313
30%	-0.4678	-0.1929	0.7302

TABLE III: Average NLL comparison across sampling rates.

Finally, to compare with a CNN-based method, we also implemented the SICNN model in [5]. The model accepts two input channels. One channel is for the sparse RSS measurement map and the other for the building map. For a fair comparison, we overlaid the LOS/NLOS status map onto the building map, using ternary values (0 for a building, 0.5 for NLOS, and 1 for LOS). Furthermore, the number of convolutional filters is slightly reduced from that of the original SICNN so that the total neural network parameter count is comparable to that of our GNN-based architecture. Again, we test the SICNN models trained with the MSE loss and the NLL loss.

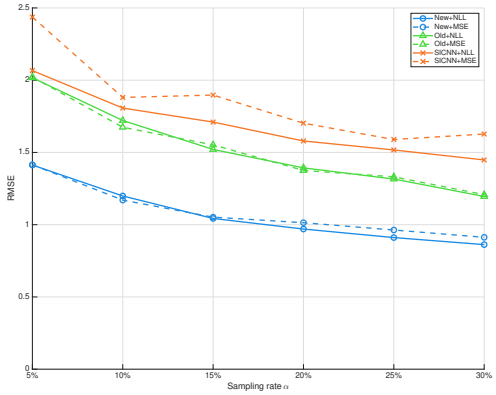
B. Results

Tables II and III summarize the RMSE and the NLL performance metrics of different methods at various sampling rates α. "New" refers to the new edge rule and "Old" means the old edge rule. "MSE" and "NLL" indicate that the training was done using the MSE loss and the NLL loss, respectively. The results are also plotted in Fig. 2. Fig. 2(a) depicts the RMSE performances of various methods, where the solid curves correspond to the results of NLL-based training and the dashed ones are from MSE-based training. Fig. 2(b) shows the NLL curves, all from NLL-based training.

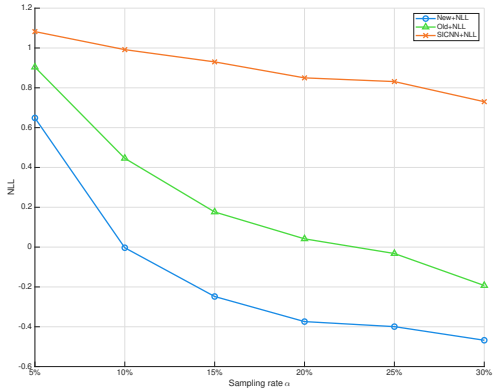
It is first seen from Table II and Fig. 2(a) that NLL-based training yields the RMSE performance that is at least comparable to that from MSE-based training. For SICNN, NLL-based training significantly reduces RMSE throughout different sampling rates. For the GNN models, NLL consistently achieves RMSE performance on par with MSE, and occasionally provides modest improvements.

Again from Table II and Fig. 2(a), the GNN model with the proposed edge rule achieves the best overall RMSE performance among all models. At the 5% sampling rate, for instance, "New+MSE" and "New+NLL" obtain the RMSE values around 1.41, outperforming the GNN with the old edge rule, which achieves around 2.02. In fact, the proposed edge rule consistently outperforms the conventional one, yielding an average 29% reduction in RMSE. Also, it is worth noting that the GNN-based models, regardless of the edge rule used, significantly outperform the CNN-based model.

Table III and Fig. 2(b) show that the proposed method produces well-calibrated uncertainty estimates in terms of the



(a) RMSE performance



(b) Average NLL performance

Fig. 2: Performance comparison.

average NLL metric. The "New+NLL" configuration consistently achieves the lowest NLL values across all sampling rates, which indicates accurate RSS predictions coupled with well-calibrated uncertainty estimates. In contrast, "Old+NLL" yields much higher NLL values, with "SICNN+NLL" exhibiting even poorer performance.

To visually examine the quality of the constructed RSS maps, the predicted maps for an example test block are shown in Figs. 3 and 4 for $\alpha = 20\%$. The ground truth RSS map is shown in Fig. 3(a) and the sparse measurement locations are depicted in Fig. 3(b). Also shown in Fig. 3 are the predicted RSS maps from different methods trained using the MSE loss. It can be seen that the proposed method with the new edge rule very closely reconstructs the ground truth RSS map. In particular, the sharp transitions at the boundary of the LOS/NLOS areas are extremely well captured. This highlights the benefit of the proposed edge construction rule, which eliminates the edges across the LOS/NLOS boundaries. On the other hand, the GNN with the old edge rule shows a blurry boundary. Although SICNN baseline uses the LOS/NLOS status map as an input, it still produces somewhat blurred LOS/NLOS borders.

In Fig. 4, the maps from NLL-based training are presented. Three maps are shown for each method. The leftmost map corresponds to the predicted RSS map, the middle one the squared error between the predicted and the ground truth maps, and the rightmost map depicts the map of uncertainty predictions. The proposed method produces an uncertainty map

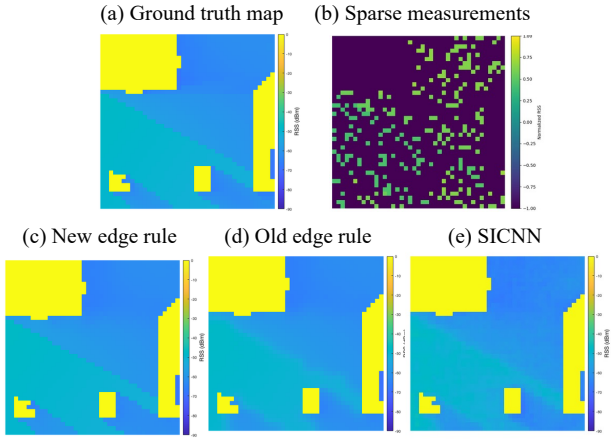
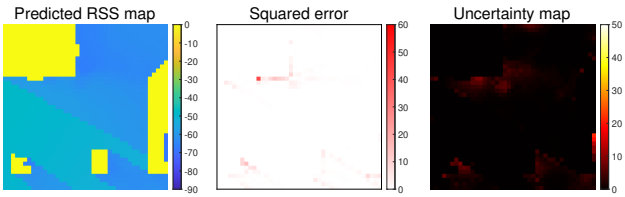
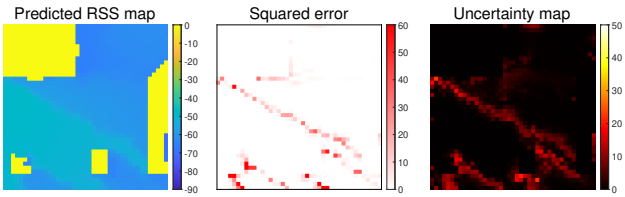


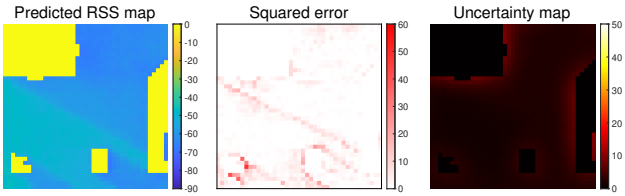
Fig. 3: Ground truth and predicted RSS maps. The methods were trained using the MSE loss.



(a) GNN with the new edge rule (Proposed)



(b) GNN with the old edge rule



(c) SICNN

Fig. 4: Predicted RSS map, square error map, and uncertainty map. The methods were trained using the NLL loss.

with low uncertainty values in the readily predictable region with smooth changes, and high uncertainty in locations like behind buildings and regions with very few measurements (cf. Fig. 3(b)). Also, the high uncertainty predictions align well with high squared errors, indicating the model's uncertainty estimates are well calibrated.

On the contrary, the GNN with the old edge rule assigns high uncertainty along the LOS/NLOS borders, as these regions are susceptible to large errors for the method. For SICNN, the squared error map reveals large residuals along building boundaries as well as the LOS/NLOS borders. The predicted uncertainty also rises in these areas, indicating that SICNN struggles in using the structural cues effectively. This is expected because SICNN performs convolution over a regular grid without explicitly accounting for the LOS/NLOS statuses. As a result, uncertainty tends to spread broadly across map edges and building borders.

V. Conclusion

In this work, a GNN-based radio map estimation method for urban environments has been proposed. Our method can utilize LOS/NLOS status information associated with each grid point to accurately capture sharp transitions in the RSS values, thereby improving the overall accuracy of radio map construction. In addition, the model jointly predicts uncertainty levels associated with the RSS estimates using NLL training cost. Numerical results indicate that the proposed method outperforms baseline models in terms of map construction accuracy and the quality of its well-calibrated, interpretable uncertainty estimates.

References

[1] E. Krijestorac, S. Hanna, and D. Cabric, "Spatial signal strength prediction using 3D maps and deep learning," in *Proc. of the IEEE Int. Conf. Commun. (ICC)*, Jun. 2021, pp. 1–6.

[2] R. Levie, C. Yapar, G. Kutyniok, and G. Caire, "RadioUNet: Fast radio map estimation with convolutional neural networks," *IEEE Trans. Wireless Commun.*, vol. 20, no. 6, pp. 4001–4015, Jun. 2021.

[3] H. Yu, Z. Hou, Y. Gu, P. Cheng, W. Ouyang, Y. Li, and B. Vucetic, "Distributed signal strength prediction using satellite map empowered by deep vision transformer," in *Proc. of the IEEE Globecom Workshops*, Madrid, Spain, Dec. 2021.

[4] G. Chen, Y. Liu, T. Zhang, J. Zhang, X. Guo, and J. Yang, "A graph neural network based radio map construction method for urban environment," *IEEE Commun. Lett.*, vol. 27, no. 5, pp. 1327–1331, May 2023.

[5] R. Hashimoto and K. Suto, "SICNN: Spatial interpolation with convolutional neural networks for radio environment mapping," in *Proc. of the IEEE Int. Conf. Artif. Intell. Info. Commun. (ICAIIC)*, 2020, pp. 167–172.

[6] Q. Niu, Y. Nie, S. He, N. Liu, and X. Luo, "RecNet: A convolutional network for efficient radiomap reconstruction," in *Proc. of the IEEE Int. Conf. Commun. (ICC)*, 2018, pp. 1–7.

[7] Y. Wang, S. Sun, N. Liu, L. Xu, and L. Wang, "Two-stage radio map construction with real environments and sparse measurements," *IEEE Wireless Commun. Lett.*, vol. 14, no. 4, pp. 969–973, 2025.

[8] X. Li, S. Zhang, H. Li, X. Li, L. Xu, H. Xu, H. Mei, G. Zhu, N. Qi, and M. Xiao, "RadioGAT: A joint model-based and data-driven framework for multi-band radiomap reconstruction via graph attention networks," *IEEE Trans. Wireless Commun.*, vol. 23, no. 11, pp. 17 777–17 792, Nov. 2024.

[9] O. Esrafilian and D. Gesbert, "3D city map reconstruction from UAV-based radio measurements," in *Proc. of the IEEE Global Commun. Conf. (GLOBECOM)*, Singapore, Dec. 2017, pp. 1–6.

[10] B. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and scalable predictive uncertainty estimation using deep ensembles," in *Proc. the 31st Int. Conf. Neural Info. Processing Syst. (NeurIPS)*, Long Beach, CA, Dec. 2017, pp. 6405–6416.

[11] S. Brody, U. Alon, and E. Yahav, "How attentive are graph attention networks?" in *Proc. of the Int. Conf. Learn. Represent (ICLR)*, Virtual, Apr. 2022. [Online]. Available: https://openreview.net/forum?id=FjRrYySKwB