ENHANCING PATHLOSS ESTIMATION WITH VISION TRANSFORMERS AND DIRECT WAVE POWER INTEGRATION

Yuuki Tachioka

Denso IT Laboratory, Tokyo, Japan

ABSTRACT

Accurate pathloss (PL) estimation is essential for optimizing wireless communication networks. Traditional PL estimation methods struggle with generalization, especially in complex indoor and outdoor environments. We propose an MST++ vision transformer-based model integrated with direct wave power estimation as an auxiliary input, enabling effective line-of-sight detection and improved PL prediction. Experiments on indoor and outdoor datasets show that our approach significantly reduces estimation errors compared to existing deep learning methods, demonstrating its potential for practical PL estimation tasks.

Index Terms— Pathloss Estimation, Antenna Placement, Deep Learning, Vision Transformer, Radio Propagation

1. INTRODUCTION

Accurate pathloss (PL) estimation is essential for wireless network planning, antenna placement, and signal optimization. Traditional radio wave propagation models, based on empirical and analytical approaches, often struggle to generalize in diverse urban and indoor environments [1]. Given the increasing complexity of environments, deep learning techniques, including convolutional neural networks (CNN) [2, 3, 4] and recently transformer-based methods [5], have emerged as powerful tools to improve the accuracy of PL estimation.

Recent advances in deep learning have led to state-of-the-art (SOTA) performance in PL estimation. Pathloss map network (PMNet) [4], a CNN-based model with an encoder-decoder architecture, has demonstrated high accuracy. Although CNN-based approaches effectively capture local features, they sometimes fail to model long-range spatial correlations and line-of-sight (LOS) conditions. Meanwhile, in the field of image processing, vision transformer-based methods [6] have emerged as promising alternatives to traditional CNNs for various reconstruction tasks [7] due to their global attention mechanisms. For hyperspectral image reconstruction, multi-stage spectral-wise transformer (MST)++ [8, 9] has achieved SOTA performance [8, 10]. Given that PL estimation shares similarities with hyperspectral image reconstruction, both involving the transformation of multichannel input into a different output, we apply MST++ to PL estimation and compare its performance with PMNet.

However, existing deep learning-based PL estimation methods face challenges in capturing long-range spatial correlations and accurately determining whether a LOS path exists [11]. Distance-based auxiliary information has improved performance [5], but this method requires image rotation prior to estimation and does not fully exploit geometric information. To address these limitations, we propose an improved PL estimation framework that integrates direct wave power estimation with deep learning-based PL estimation. By incorporating estimated direct power levels as an auxiliary input, our approach improves the model's ability to capture radio propagation characteristics, leading to enhanced accuracy across diverse environments. The proposed method simplifies LOS detection and reduces estimation errors, effectively leveraging the strengths of vision transformer architectures to model long-range dependencies while maintaining computational efficiency. Through extensive evaluation on outdoor [4] and indoor [12] datasets, we demonstrate that our method achieves superior performance compared to conventional deep learning approaches.

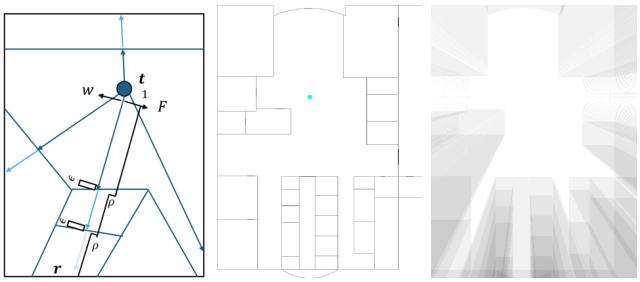
2. DIRECT POWER LEVEL ESTIMATION

This section introduces the direct power level estimation framework, which is designed to model the LOS path and the transmission loss through walls, while excluding reflections and multipath effects. The goal is to provide an interpretable estimate of the direct wave component as an auxiliary input for PL estimation. We consider two scenarios: (1) outdoor urban environments with building geometries and (2) indoor layouts with reflectance and transmittance data. For outdoor environments [4], we input a 2D map with building outlines and transmitter (Tx) positions (Fig. 1 (middle)); for indoor environments [12], reflectance and transmittance maps represent wall locations and material properties (Fig. 2 (middle)). Estimating PL maps directly from these inputs is challenging and the goal is to capture LOS and wall attenuation effects efficiently with a help of direct power level estimation.

In our proposed method, we first estimate the direct power level with a consideration of power reduction by transmission through a building or wall. The direct path from Tx is de-

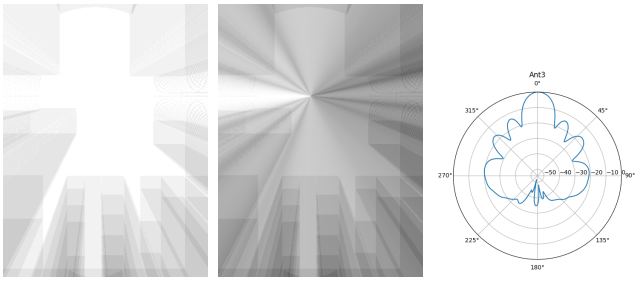
(a) Direct power level estimation (b) Geometry of buildings with Tx (c) Estimated direct power map

Fig. 1. For outdoor pathloss (PL) estimation, the schematics of direct path estimation (left), the geometry of buildings (middle), and estimated direct power map (right).



(a) Direct power level estimation (b) Normal incidence reflectance with Tx (c) Estimated direct power map

Fig. 2. For indoor PL estimation, the schematics of direct path estimation (left), the normal incidence reflectance (middle), and estimated direct map (right).



(a) Estimated direct power map when antenna is isotopic (b) Estimated direct power map when antenna radiation pattern is (c) (c) Antenna radiation pattern

Fig. 3. Direct power map estimation when antenna radiation pattern is not isotopic.

picted in Figs. 1 and 2 (left) and the estimated direct power level is visualized in Figs. 1 and 2 (right), where brighter colors indicate higher power levels and darker colors indicate lower power levels. The relative direct power level F is computed recursively with the small constant width ϵ as:

$$F(\boldsymbol{t}, \boldsymbol{r}) = F(\boldsymbol{t}, \boldsymbol{r} - \epsilon \boldsymbol{u}) - \rho w(\boldsymbol{r} - \epsilon \boldsymbol{u}, \boldsymbol{r}), \quad (1)$$

where $\boldsymbol{t} = (t_x, t_y)$ and $\boldsymbol{r} = (r_x, r_y)$ represent the Tx and receiver positions, respectively. The indicator function $w(\boldsymbol{x}_1, \boldsymbol{x}_2)$ returns 1 if a wall exists in the interval $[\boldsymbol{x}_1, \boldsymbol{x}_2)$ but not in $\boldsymbol{x}_2$, otherwise returning 0. The unit vector $\boldsymbol{u} = \frac{\boldsymbol{r}-\boldsymbol{t}}{||\boldsymbol{r}-\boldsymbol{t}||}$ points from Tx to the receiver, where $||\cdot||$ denotes the length of the vector. The reduction constant ρ accounts for the power attenuation through the walls. In Tx, $F(\boldsymbol{t}, \boldsymbol{t}) = 1$. Whether a receiver has an LOS to Tx significantly affects PL estimation, but identifying LOS can require analyzing long-range correlations, making it challenging even for deep learning-based methods. The proposed direct power map estimation simplifies LOS and the attenuated direct wave using Eq. (1).

When the antenna radiation pattern is not isotropic, it can be incorporated into direct power level estimation as:

$$F'(\boldsymbol{t}, \boldsymbol{r}) = F(\boldsymbol{t}, \boldsymbol{r}) - \nu G\left(\tan^{-1}\left(\frac{r_y - t_y}{r_x - t_x}\right) + \frac{\pi}{2}\right), \quad (2)$$

where $G(\theta)$ represents the antenna's directional gain as a function of angle θ, and ν is a deduction coefficient.

Fig. 3 illustrates the effect of the antenna radiation pattern. The left figure assumes an isotropic antenna, while the middle figure shows a power map generated using Eq. (2) based on the directional pattern in the right figure. Lower gain directions exhibit reduced power allocation.

3. PATHLOSS MAP PREDICTION

3.1. PMNet

PMNet [4] is an encoder-decoder network with skip connections [2, 13] [1], residual blocks, atrous convolutions, and an hourglass network [14], as shown in Fig. 4, designed to capture local features in PL estimation tasks.

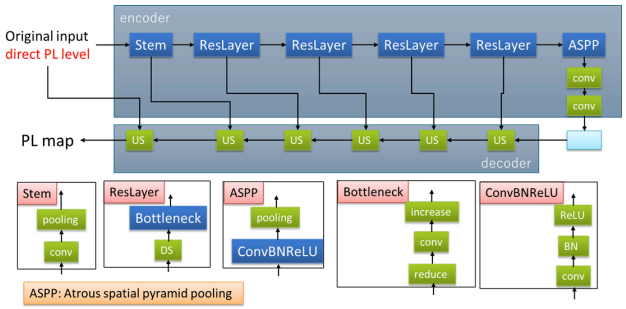


Fig. 4. Structure of PMNet, where conv, DS, US, and BN are convolution, downsampling, upsampling, batch normalization, respectively.

3.2. MST++

MST++ [8] [2] is a vision transformer-based model [6] with a sparse, coarse-to-fine structure [9], as shown in Fig. 5. Its

[1] Source codes: https://github.com/abman23/pmnet

[2] Source codes: https://github.com/caiyuanhao1998/MST-plus-plus

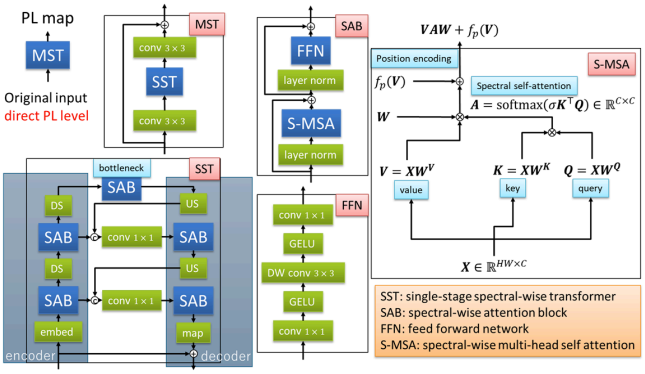


Fig. 5. Structure of multi-stage spectral-wise transformer (MST). Matrices $\boldsymbol{W}$, $\boldsymbol{W}^V$, $\boldsymbol{W}^K$, $\boldsymbol{W}^Q$, and σ are learnable. The height, width, and channels of the input to S-MSA are H, W, and C. $\top$ is the transpose of the matrix and f_p is a position encoding function.

Table 1. The specification of outdoor dataset.

task	# train	# valid	Freq.(GHz)
USC	17114	1902	2.5
UCLA	3398	378	3.0
Boston	2828	315	3.0

key innovation is spectral-wise multihead self-attention (S-MSA), which focuses on channel correlations rather than spatial dependencies, reducing computational complexity while improving spectral feature learning. MST++ follows a hierarchical U-Net-like structure called a single-stage spectral-wise transformer (SST), which consists of an encoder, bottleneck, and decoder. The encoder extracts image features at different resolutions, while the decoder reconstructs the PL map with skip connections. By stacking multiple SSTs, MST++ progressively refines image reconstruction, achieving high accuracy with a lower computational cost. MST++ can capture channel correlations using an attention mechanism, which is essential for accurate PL estimation, because geometrical location and antenna power must be integrated. In our implementation, we integrate direct wave power maps as an auxiliary input, allowing the model to efficiently utilize LOS information and reduce estimation errors.

4. EXPERIMENT

4.1. Experimental conditions

For the estimation of outdoor PL, we used simulation data [4] as shown in Table 1[3]. Ray tracing was applied to the geographical and morphological maps of the University of Southern California (USC) campus, the University of California, Los Angeles (UCLA) campus, and the Boston area. The USC and UCLA datasets represent light urban environments with mostly low-rise buildings, while the Boston dataset represents a dense metropolitan area with high-rise buildings and irregular street layouts. Each dataset varies in scale, geographical features, and environmental characteristics. PMNet [4] has demonstrated SOTA performance on these datasets, making it a strong baseline model for comparison.

[3]Available at `https://github.com/abman23/pmnet`

Table 2. The specification of indoor dataset.

task	# train	# valid	Freq.(GHz)	Tx pattern
task1	1125	125	0.868	1
task2	3375	375	0.868/1.8/3.5	1
task3	24975	2775	0.868/1.8/3.5	5

For indoor PL estimation, we used the data from the "first indoor pathloss radio map prediction challenge" [12]. The input data consist of three channels: normal incidence reflectance and transmittance (both in dB, with 0 for air) and the physical distance from Tx to each grid. The output is the PL map within the building. There are 25 building patterns. The challenge consists of three tasks. Task 1 evaluates isotropic antenna patterns, generating 50 radio maps per building. Task 2 extends this to three frequencies. Task 3 extends this to five different antenna radiation patterns. Since the input image size varies, we applied data augmentation, including cropping and flipping. Table 2 shows the specification of the task.

PMNet and MST++ were trained using the Adam optimizer with a learning rate of 0.004, which was halved every 10 epochs. The batch size was set to 16 for outdoor data and 8 for indoor data. The total number of epochs was 30. An estimated direct power level map was added to the input, resulting in three input channels for outdoor data and four for indoor data. The direct power map estimation was performed using the parameters $\rho = 0.05, \quad \nu = \frac{0.5}{\min(G)}$.

4.2. Result and discussion (Outdoor)

Table 3 shows the mean square error (MSE) with the original input, which does not include direct power level estimation. This serves as a baseline for evaluating the effectiveness of our proposed auxiliary input. PMNet's performance is highly dependent on the amount of training data, resulting in lower accuracy in the Boston dataset. In contrast, when sufficient data are available, such as in the USC dataset, PMNet achieves high accuracy. MST++ exhibits stable learning performance across datasets. The reference and estimated PL map using MST++ for the USC dataset are shown in Figs. 6 (a) and (b). The predicted PL map tends to be excessively smoothed, resulting in uniformly low PL values throughout the region.

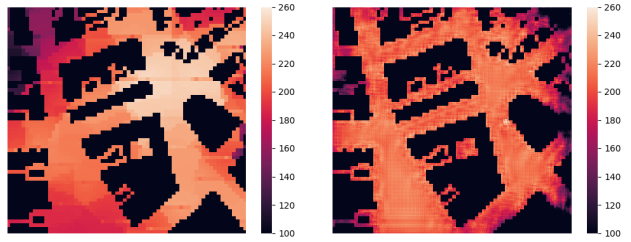
Table 4 shows the MSE when the proposed auxiliary input is introduced. Both PMNet and MST++ achieve a significant reduction in MSE. The PL maps estimated by PMNet and MST++ with the auxiliary input are shown in Figs. 6 (c)

Table 3. Mean square error (MSE) with original input.

task	PMNet	MST++
UCLA	$2.121 \cdot 10^{-3}$	$2.750 \cdot 10^{-2}$
USC	$9.750 \cdot 10^{-5}$	$6.501 \cdot 10^{-3}$
Boston	$2.538 \cdot 10^{-1}$	$1.266 \cdot 10^{-2}$

Table 4. MSE with auxiliary input.

task	PMNet	MST++
UCLA	$5.780 \cdot 10^{-5}$	$7.872 \cdot 10^{-6}$
USC	$2.489 \cdot 10^{-6}$	$1.822 \cdot 10^{-7}$
Boston	$1.059 \cdot 10^{-4}$	$1.148 \cdot 10^{-5}$



(a) Reference (b) MST++

(c) PMNet with Eq. (1) (d) MST++ with Eq. (1)

Fig. 6. Reference and estimated PL by PMNet and MST++ for the validation set of USC.

and (d), respectively. In both cases, the predicted PL maps closely match the reference. Compared to Fig. 6 (b), the auxiliary input improves the accuracy of the estimation, particularly in the plaza area in the upper left and right regions where the lower PL values are better captured. These results confirm that the proposed auxiliary input is effective in reducing MSE. Additionally, with the auxiliary input, PMNet outperforms MST++ in terms of accuracy.

4.3. Result and discussion (Indoor)

Table 5 shows the MSE for the validation data without using Eq. (1) or (2). Similarly to the outdoor case, the performance of PMNet depends on the volume of training data, while MST++ exhibits stable learning. Fig. 7 shows the reference (a) and the PL map estimated by MST++ (b). The predicted PL map is overly smooth and does not capture the detailed shapes of walls and other structures.

Table 6 presents the results when the proposed auxiliary input is used. The auxiliary input effectively reduces MSE. Figs. 7 (c) and (d) show the PL maps estimated by PMNet and MST++, respectively. PMNet still struggles to produce accurate estimations, whereas MST++ achieves high accuracy, successfully capturing room shapes in low-PL areas. These results confirm the effectiveness of the proposed method for indoor PL estimation as well. However, in the lower part of the image, the reflected waves are not well predicted, indicating a remaining challenge in handling reflections.

Table 5. MSE with original input.

task	PMNet	MST++
task1	$2.739 \cdot 10^{-2}$	$3.511 \cdot 10^{-4}$
task2	$4.812 \cdot 10^{-4}$	$6.821 \cdot 10^{-4}$
task3	$5.411 \cdot 10^{-4}$	$8.509 \cdot 10^{-4}$

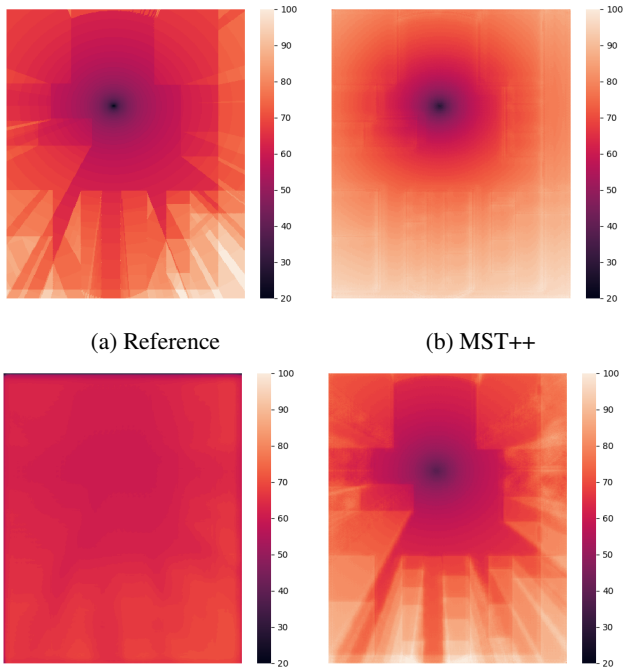
Table 6. MSE with auxiliary input.

task	PMNet	MST++
task1	$2.761 \cdot 10^{-3}$	$1.439 \cdot 10^{-4}$
task2	$4.636 \cdot 10^{-4}$	$4.047 \cdot 10^{-4}$
task3 (Eq. (1))	$6.203 \cdot 10^{-4}$	$6.197 \cdot 10^{-4}$
task3 (Eqs. (1) and (2))	$3.389 \cdot 10^{-4}$	$4.114 \cdot 10^{-4}$

Next, we compare the results for task 3, where different antenna radiation patterns are considered. Task 3 of Table 6 shows that incorporating antenna directivity (Eqs. (1) and (2)) achieves a lower MSE compared to those that do not consider directivity (Eq. (1)). Fig. 8 shows the results for the same antenna location and building structure as in Fig. 7, but with the radiation pattern of Fig. 3 (c). The reference clearly reflects the influence of the directivity of the antenna. When directivity is not considered, the gain remains constant regardless of direction. In contrast, Figs. 8 (c) and (d) show the results with directivity taken into account. It can be observed that areas with lower gain correspond to higher estimated PL values. These findings demonstrate the importance of considering antenna directivity in PL estimation.

5. CONCLUSION

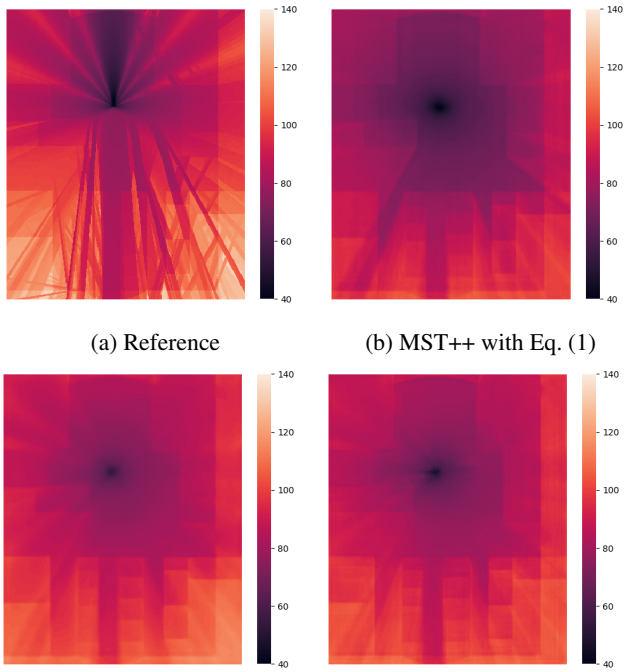
In this paper, we proposed to improve Pathloss (PL) estimation by integrating direct wave information and leveraging vision transformer architectures. Our experiments demonstrate that incorporating direct wave PL estimation significantly improves prediction accuracy in both indoor and outdoor environments. The results confirm that our method extends the applicability of DNN-based PL estimation methods and outperforms CNN-based approaches. Future work will explore the integration of low-order reflected waves as auxiliary information and evaluate the model's generalization to unseen environments.



(a) Reference (b) MST++

(c) PMNet with Eq. (1) (d) MST++ with Eq. (1)

Fig. 7. Reference and estimated PL map for the validation set of task 1.



(a) Reference (b) MST++ with Eq. (1)

(c) PMNet with Eqs. (1) and (2) (d) MST++ with Eqs. (1) and (2)

Fig. 8. Reference and estimated PL map for the validation set of task 3.

6. REFERENCES

[1] Han-Shin Jo, Chanshin Park, Eunhyoung Lee, Haing Kun Choi, and Jaedon Park, "Path loss prediction based on machine learning techniques: Principal component analysis, artificial neural network, and Gaussian process," *Sensors*, vol. 20, no. 7, p. 1927, 2020.

[2] Ron Levie, Çağkan Yapar, Gitta Kutyniok, and Giuseppe Caire, "RadioUNet: Fast radio map estimation with convolutional neural networks," *IEEE Trans on Wireless Communications*, vol. 20, no. 6, pp. 4001–4015, 2021.

[3] Rong-Terng Juang, "Explainable deep-learning-based path loss prediction from path profiles in urban environments," *Applied Sciences*, vol. 11, no. 15, pp. 6690, 2021.

[4] Ju-Hyung Lee and Andreas F. Molisch, "A scalable and generalizable pathloss map prediction," *IEEE Trans on Wireless Communications*, vol. 23, no. 11, pp. 17793–17806, 2024.

[5] Thomas Hehn, Tribhuvanesh Orekondy, Ori Shental, Arash Behboodi, Juan Bucheli, Akash Doshi, June Namgoong, Taesang Yoo, Ashwin Sampath, and Joseph Soriaga, "Transformer-based neural surrogate for link-level path loss prediction from variable-sized maps," arXiv:2310.04570, 2023.

[6] 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," arXiv: 2010.11929, 2020.

[7] Asifullah Khan, Zunaira Rauf, Anabia Sohail, Abdul Rehman Khan, Hifsa Asif, Aqsa Asif, and Umair Farooq, "A survey of the vision transformers and their CNN-transformer based variants," *Artificial Intelligence Review*, vol. 56, pp. 2917–2970, 2023.

[8] Yuanhao Cai, Jing Lin, Zudi Lin, Haoqian Wang, Yulun Zhang, Hanspeter Pfister, Radu Timofte, and Luc Van Gool, "MST++: Multi-stage spectral-wise transformer for efficient spectral reconstruction," in *CVPRW*, 2022, pp. 745–755.

[9] Yuanhao Cai, Jing Lin, Xiaowan Hu, Haoqian Wang, Xin Yuan, Yulun Zhang, Radu Timofte, and Luc Van Gool, "Coarse-to-fine sparse transformer for hyperspectral image reconstruction," in *ECCV*, 2022, pp. 686–704.

[10] Yuuki Tachioka, "Restoration of hyperspectral images from RGB images by using two-band models," in *International Conference on Computer and Communications Management*, 2024, pp. 48–54.

[11] Han-Hee Lee, Jae Lee, Jong Kwon, Jong-Hwan Hwang, and Chang Hee Hyoung, "Prediction of radio-wave propagation in a shield room: Measurement, simulation, and theoretical method," *Journal of Electromagnetic Engineering and Science*, vol. 20, pp. 45–52, 2020.

[12] Stefanos Bakirtzis, Cagkan Yapar, Kehai Qui, Ian Wassell, and Jie Zhang, "The first indoor pathloss radio map prediction challenge," in *ICASSPW*, 2025.

[13] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention*. 2015, 234–241.

[14] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *ECCV*, 2018, pp. 833–851.