MULTI-STAGE VISION TRANSFORMER FOR INDOOR PATHLOSS ESTIMATION

Yuuki Tachioka

Denso IT Laboratory, Tokyo, Japan

ABSTRACT

Estimation of pathloss (PL) is important. The first indoor pathloss radio map prediction challenge introduced an indoor PL estimation problem. The indoor PL estimation is different from the outdoor one. For outdoor PL estimation, semantic segmentation is crucial because it can ignore PL in building areas, but indoor PL estimation, PL in all regions must be estimated and given inputs such as reflectance and transmittance are correlated. We applied MST++ to the estimation of the PL map for dealing with correlated input data. After the estimation of direct PL estimation, we used MST++ to estimate the PL map from the input of four channels. The experiments show the effectiveness of our approach.

Index Terms— pathloss estimation, multi-stage transformer

1. INTRODUCTION

Pathloss (PL) estimation is important for estimating PL map from transmitted antenna (Tx). The first indoor pathloss radio map prediction challenge [1] is an ICASSP 2025 signal processing grand challenge, which introduced the indoor PL estimation problem. For the outdoor PL estimation, the NN tailored for Pathloss Map Prediction (PMNet) [2] achieves the SOTA. However, the indoor PL estimation is different from the outdoor one. For the estimation of outdoor PL, semantic segmentation is crucial because it can ignore PL in building areas, but for the estimation of indoor PL, PL in all regions must be estimated. For indoor PL estimation, given information such as reflectance and transmittance is correlated. It is important to deal with correlated information. In the field of hyperspectral image restoration, it is important to deal with channel-correlated input. Multi-stage spectral-wise transformer (MST++) [3] is SOTA method, which can capture channel correlation based on the attention mechanism thanks to the high performance of vision transformer [4].

We propose to apply MST++ to indoor PL estimation, after we develop direct PL level estimation. Based on this information, MST++ can estimate the PL map. Experiments on the first indoor pathloss radio map prediction challenge show that our proposed approach was effective.

2. INDOOR PATH ESTIMATION

2.1. Direct PL level estimation

The input data for this challenge are reflectance (as shown in Fig. 1) and transmittance within the building with the position of Tx, but from these inputs it is difficult to estimate PL maps. Thus, we estimate the direct PL level before estimating the PL map. Based on the reflectance information, the position of the wall within the building can be known. From the Tx position, the direct path can be drawn as shown in Fig. 2. For every transmission of the direct path through the wall, the power of the direct path is decreased. Depending on the power level, the direct PL level can be estimated as shown in Fig. 3, which indicates that the brighter color corresponds to the higher power level.

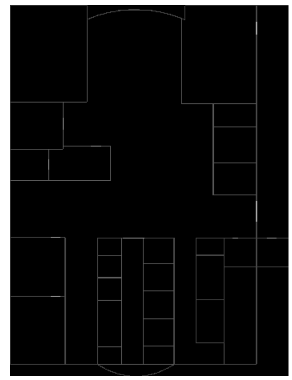

Fig. 1. The normal incidence reflectance [dB] at each point of the grid (0 for air).

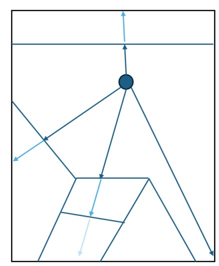

Fig. 2. The schematics of direct path estimation.

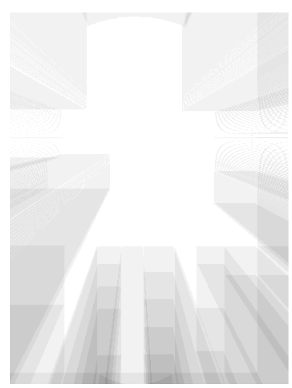

Fig. 3. Direct PL level estimation.

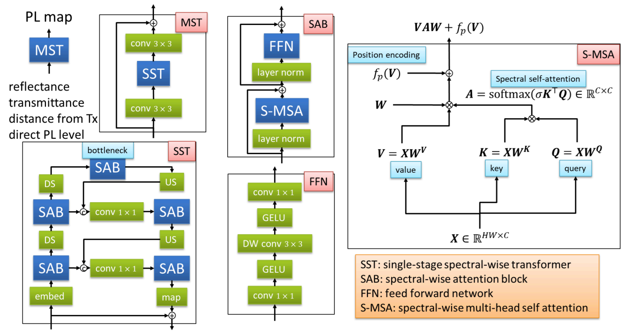

Fig. 4. Structure of multi-stage spectral-wise transformer (MST). Here, conv, DS and US are convolution, downsampling, and upsampling, respectively; matrices $\boldsymbol{W}$, $\boldsymbol{W}^V$, $\boldsymbol{W}^K$, $\boldsymbol{W}^Q$, and σ are learnable. The height, width, and channels of the input to S-MSA are H, W, and C. $\top$ is the transpose of the matrix and f_p is a position encoding function.

2.2. MST++ to pathloss estimation

We used MST++[3] to estimate the PL map from reflectance, transmittance, distance from Tx, and estimated direct PL level in the previous section, as shown in Fig. 4. This model can deal with the correlated inputs to estimate the PL map accurately.

3. EXPERIMENT

3.1. Experimental conditions

We used the first indoor pathloss radio map prediction challenge data. There were 25 different indoor geometries and PL maps were observed at three different frequencies (868 MHz, 1.8 GHz, and 3.5 GHz) with five antenna radiation patterns. The input data was composed of three channels, which were the normal incidence reflectance in dB at each point of the grid (0 for air), the normal incidence transmittance in dB at each point of the grid (0 for air), and the physical distance between Tx and each point of the grid. The output to be estimated was the PL map within the building where each point of the image denotes the PL at that point.

This challenge consisted of three tasks. Task 1 was an evaluation of the simulated data with an isotropic antenna pattern conducted at 868 MHz for the 25 buildings. For each building, 50 radio maps were generated by placing the Tx in different locations within the building. Task 2 was an evaluation of the simulated data with an isotropic antenna pattern conducted at three different frequencies (0.868, 2, and 3.5 GHz) for the 25 buildings. For each building and frequency, 50 radio maps were generated by placing the Tx in different locations within the building. Task 3 was an evaluation of the simulated data with five different antenna radiation patterns conducted at 0.868, 2 and 3.5 GHz for the 25 buildings.

We input an estimated direct PL level as shown in Fig. 3 in addition to the three channels given (in total four channels) into MST++. To train MST++, Adam optimizer was applied, and to overcome the problem of the limited amount of data, data augmentation of clop and flip was used. We divide training sets into training and validation sets as shown in Table 1.

Table 1. The number of data for training and validation sets.

task	training	validation
task1	1250	125
task2	3750	375
task3	27750	2775

Fig. 5. Estimated PL by MST++ for the validation set of task 1.

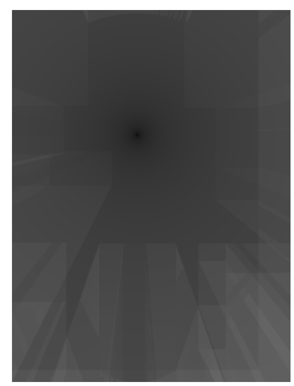

Fig. 6. Reference PL for the validation set of task 1.

Table 2. Mean square error (MSE) for validation data.

task	MSE
task 1	0.0001439
task 2	0.0004047
task 3	0.0006197

3.2. Result and discussion

Fig. 5 shows the estimated PL map within the building. The ground truth of the PL map was Fig. 6. Comparison of them shows that our proposed approach can accurately estimate the PL map. The contour of the PL map corresponds to that in Fig. 3, which indicates that the estimation of the direct PL level is important. Table 2 shows the mean square error (MSE) of the prediction for the validation data, which were not included in the training data. The predicted PL maps have a small MSE.

4. CONCLUSION

We applied MST++ to the estimation of the PL map for dealing with correlated input data. After the estimation of direct PL estimation, we used MST++ to estimate the PL map from the input of four channels. Experiments show the effectiveness of our approach.

5. REFERENCES

[1] Stefanos Bakirtzis, Cagkan Yapar, Kehai Qui, Ian Wassell, and Jie Zhang, "The first indoor pathloss radio map prediction challenge," in *To appear in Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW)*, April 2025.

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

[3] 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 *Proceedings of CVPRW*, 2022, pp. 745–755.

[4] 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*, vol. abs/2010.11929, 2020.