Transformer based Radio Map Prediction Model for Dense Urban Environments

Yu Tian[1], Shuai Yuan[1], Weisheng Chen[2], and Naijin Liu[1]
[1]Qian Xuesen Laboratory of Space Technology, China Academy of Space Technology, Beijing, China
[2]School of Aerospace Science and Technology, Xidian University, Xi'an, China

Abstract*-Radio map prediction (RMP) is one of the key technologies to improve spectrum efficiency. In this paper, a novel deep learning model termed as RadioTrans is proposed for RMP task. Specifically, Transformer modules are used to capture the long-range spatial relationship in radio wave propagation. Furthermore, a Grid Anchor technique is proposed to better represent the relative position of the radiation source, destination and environment. The effectiveness of proposed method is verified on an urban radio wave propagation dataset. Compared with state-of-the-art deep learning RMP model, RadioTrans improve the prediction accuracy by 27.3%. Compared with the well-known ray-tracing based method, the prediction speed is increased by 4 orders of magnitude. Code is released at git@github.com:QXSLAB/RadioTrans_Official.git.

I. INTRODUCTION

With the advent of Fifth Generation (5G) wireless communication and massive growth of wireless devices, radio spectrum resources are becoming increasingly scarce [1]. Understanding the propagation behavior of radio waves is very important for making full use of limited spectrum resources [2]. Radio Map Prediction (RMP) aims at modeling how radio wave propagates in physical space and predicting radio power for every position in a geographic area, given radiation conditions of transmitter and environment geometry.

Fast and accurate radio map prediction is very challenging, since radio wave would reflect, refract, diffract and scatter complicatedly in environment after originated radiated from the transmitter source. Ray Tracing (RT) method is widely used in mobile communication environments to predict the propagation characteristics of radio waves [3]. However, RT-based radio map prediction suffers slow prediction speed and huge computational complexity.

Recently, Convolution Neural Networks (CNNs) have been reported for fast RMP task [4][5]. Instead of predicting each pixel output independently, CNNs extract common features of for all pixel outputs leveraging the hierarchical structure and saves lots of redundant computation. On this basis, fast prediction can be achieved.

However, the prediction accuracy of CNN-based RMP model may be limited. The reason is as follows. In radio map prediction, the source and destination may be far apart in space. At the same time, due to the presence of reflection and scattering, the radio wave may spread to a long distance and then come back. As a result, global information of the propagation environment is required to predict the radio power of a single pixel in radio map. However, convolution operation in CNN model cannot effectively model such global relationship due to its localized receptive field.

In this paper, a RadioTrans model is proposed for fast and accurate radio map prediction. In particular, Transformer is used to solve the long-range relationship modeling problem in radio wave propagation. Unlike CNN-based methods, Transformer is powerful at modeling global contexts, which has been widely witnessed in the field of machine translation and natural language processing (NLP) [6]. We specially designed a Transformer-based network architecture for RMP task. As far as we know, this is the first application of Transformer in the field of radio wave propagation. We show empirically that proposed method achieves superior results on zero-shot radio map prediction task.

II. BACKGROUND

CNN-based models are limited to characterize the relationship between positions in its receptive field. Unlike CNN, Transformer calculates the pairwise relationship between all positions. The input of Transformer is an embedding of token sequence $\mathbf{X}=[\boldsymbol{x}_1,...,\boldsymbol{x}_k,...,\boldsymbol{x}_L]$, where $\mathbf{X}\in\mathbb{R}^{d_{\text{token}}\times L}$, $\boldsymbol{x}_k$ represents the embedding of the k th input token, L represents the length of sequence, and d_{token} represents embedding dimension. The output of Transformer is also a sequence of embeddings $\mathbf{Z}\in\mathbb{R}^{d_{\text{token}}\times L}$. Transformer aims to transform the input sequence to output sequence.

In Transformer, the long-range relationship learning is mainly achieved via self-attention. In self-attention, sequence $\mathbf{X}$ is linearly projected to Query matrix $\mathbf{Q}\in\mathbb{R}^{d_{\text{token}}\times L}$, Key matrix $\mathbf{K}\in\mathbb{R}^{d_{\text{token}}\times L}$, and Value matrix $\mathbf{V}\in\mathbb{R}^{d_{\text{token}}\times L}$. The multiplication between $\mathbf{Q}$ and $\mathbf{K}^T$ characterizes pairwise matching degrees between output tokens and input tokens. On this basis, the representations of output tokens can be written in following matrix form,

$$\mathbf{Z}=\text{softmax}\left(\frac{\mathbf{Q}\cdot\mathbf{K}^T}{\sqrt{d_{\text{token}}}}\right)\cdot\mathbf{V}$$

where the superscript ^{T} is used to indicate matrix transpose.

III. ARCHITECTURE OF RADIOTRANS MODEL

The architecture of proposed RadioTrans is illustrated in Figure 1 (a). The proposed model consists of an encoder and a decoder. The role of the encoder is to extract environment features with different scales. The purpose of decoder is to

2021 13th International Symposium on Antennas, Propagation and EM Theory (ISAPE) | 978-1-6654-1812-6/21/$31.00 ©2021 IEEE | DOI: 10.1109/ISAPE54070.2021.9753644

interpret the radio propagation process and predict received power for all spatial positions.

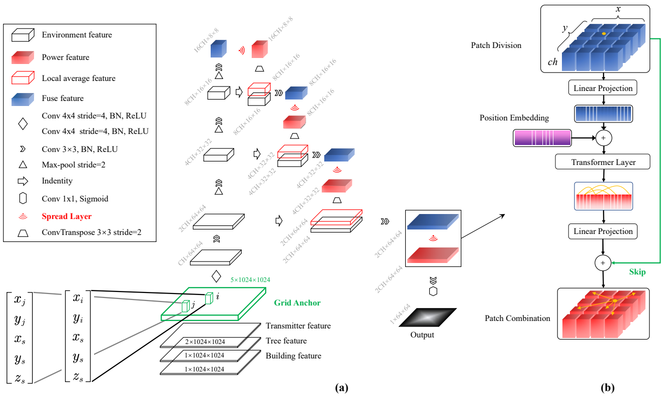

Figure 1. (a) Architecture of RadioTrans. (b) Structure of spread layer.

In order to characterize the details of radio propagation, the spread layer is proposed based on Transformer. The structure of spread layer is illustrated in Figure 1 (b). In the spread layer, the input Fuse feature is firstly divided equally in spatial dimensions. Then a linear layer is used to project divided Fuse features to patch embeddings. Resulted patch embeddings are further fed into a Transformer in order to model the long-range spatial relationship between source, destination, and environment obstacles. Finally, patch embeddings are converted to Power feature through linear projection and patch combination. A skip connection is added between the input and output, which is more conducive to gradient back propagation.

In addition, the Grid Anchor (GA) technique is proposed to more accurately describe the relative position between the source, destination and environmental obstacles. As shown in Figure 1 (a), GA incorporates environment coordinates into the network, where (x_s, y_s, z_s) and (x_i, y_i) respectively represent coordinates for transmitter and all other spatial positions.

Traditionally, the Position Embedding (PE) technique is adopted in Transformer to describe the relative position between tokens. However, the PE technique may be not sufficient for radio map prediction task. In PE, each patch is allocated a random embedding vector to distinguish different positions, and these embedding vectors are learnable during training. This method is harmful to radio propagation deduction tasks, since the spatial relative position between the source, destination and environmental obstacles can be damaged during training. Leveraging GA, spatial relative position remains fixed during training, and network can more easily model the radio-environment interaction.

IV. URBAN RADIO MAP DATASET

In order to verify the effectiveness of our method, we developed a radio map dataset for dense urban scenes. In the dataset, each sample considers a different environment geometry, a random radiation position, and a random radiation frequency. This dataset puts forward a challenging zero-shot radio map prediction task, which requires the model make accurate prediction in environments it has never met.

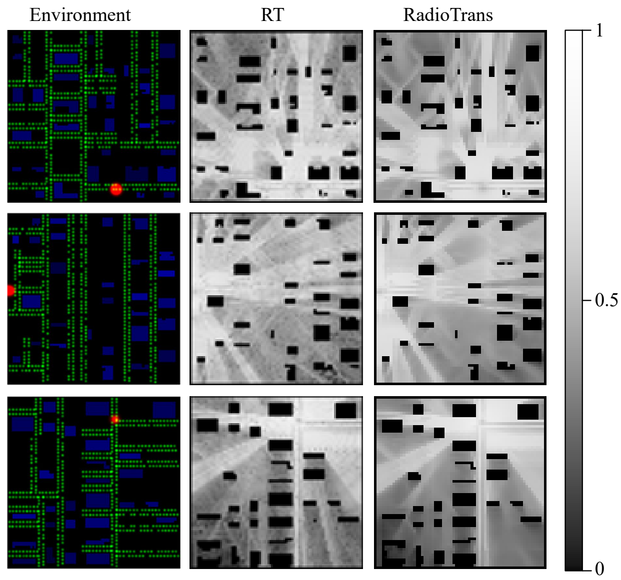

Figure 2. Environment, ground truth and prediction result

Taking the layout requirements of building, road and vegetation in typical urban scenes as constraints, environment geometries are randomly generated. The omnidirectional dipole antennas are used for radiation, and the radiation frequency ranges from 5.735 GHz to 5.825 GHz. Given radiation conditions and environment geometries, radio maps are constructed by Remcom Wireless InSite software, which has been widely used in industry and academia for wireless link design, antenna coverage optimization, and millimeter wave channel characterization [7]. The dataset construction lasted about a month, and a total of 453825 valid samples were obtained. The samples are visualized in Figure 2.

V. EXPERIMENT

A. Setup

The RadioTrans is implemented in Pytorch. Details of the model are shown in Figure 1 (a), in which the channel expansion factor CH is set to be 64. The resolution of input and output is set as 1024×1024 and 64×64. In the building feature map and tree feature map, pixel value respectively represents the height of building and tree. In the transmitter feature map, the first channel indicates the position of the transmitter in one-hot manner, where the nonzero pixel value represents the height of the transmitter. In the second channel, all pixels are filled with radiation frequency. Input features and ground truth are normalized to [0, 1]. For all spread layers, the input Fuse feature is divided into 8 equal parts along the x dimension and y dimension, forming 64 patches. In Transformer Layer, the embedding dimension d_{token} is set as 512. Unet [4] is chosen as the baseline model. As far as we know, this is the state of the art (SOTA) deep learning model for RMP task. Dataset is divided into training set and validation set according to the ratio of 99:1. All models were trained with a batch size of 64 and using the Adam optimization algorithm with learning rate of 0.0001 for 300,000 iterations. Mean absolute error (MAE) between model output and normalized ray-tracing result is used as the loss function.

B. Results

Prediction results are shown in Figure 2. The first column shows the propagation environment. Colors red, green and blue are used to represent buildings, trees and radiation source. The depth of blue, depth of green and size of red point respectively indicate the height of building, tree and radiation source. The second column shows the ground truth of radio maps, which are obtained by RT method. We use brightness to indicate the level of received power. The last column shows the prediction results given by RadioTrans model. The prediction errors of RadioTrans model and Unet model are compared in Figure 3. As can be seen, Unet model failed to predict radio power in certain regions. In contrast, the RadioTrans model can make accurate predictions over the entire environment.

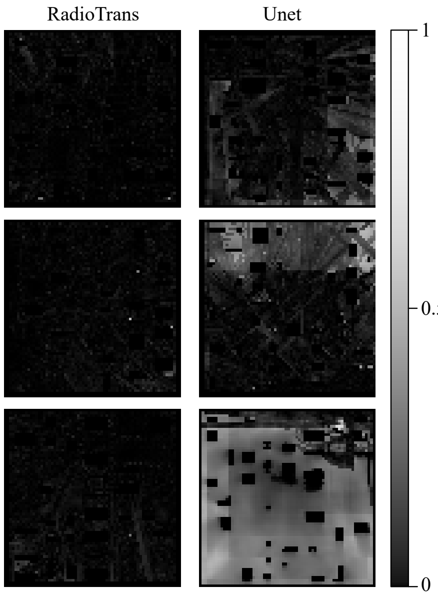

Figure 3. Prediction error of RadioTrans and Unet model

The MAE during training is shown in Figure 4. It can be seen that RadioTrans converges faster and better than Unet. Besides, RadioTrans has a lower loss on validation set, which means it has stronger generalization ability and can better model the interaction between radio waves and the environment. In addition, it can be seen from the figure that GA and skip connection in spread layer help improve the generalization ability.

The MAE on validation set are 0.0293 for RadioTrans and 0.0403 for Unet. The prediction accuracy is improved by 27.3%. The received power prediction error of RadioTrans model is 5.3dB. The relationship between MAE and received power prediction error is $error = MAE \times 180$ dB, which is determined by received power variation range [-250 dB, -70 dB]. It is worth pointing out that removing PE gives similar result as RadioTrans. This indicates that PE technique may not be necessary for radio map prediction, although it is required for NLP and computer vision tasks.

Attributed to high reusability of low-level features and high parallelism of model architecture, the proposed RadioTrans significantly improves the speed of radio map prediction. For urban scenes considered here, RT method takes 66.1 seconds to predict a radio map with a resolution of 64×64. The result is obtained by commercial software Wireless Insite. In contrast, the average inference time of RadioTrans is 5.4e-3 seconds. The prediction speed is increased by 4 orders of magnitude.

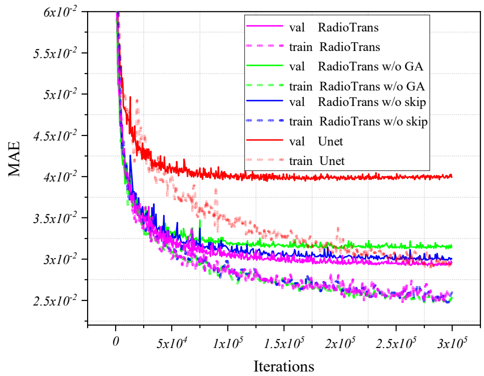

Figure 4. Loss of different models during training

VI. Conclusion

In this paper, RadioTrans deep learning model is proposed for fast and accurate radio map prediction, in which Transformer based spread layers are used for modeling the long-range dependency of source, destination and environment. In order to better anchor the topological relationship of transmitter source, destination and environment, a novel Grid Anchor technique is proposed. Experiment results show that our method can significantly improve the accuracy and speed of radio map prediction.

References

[1] A. Boulogeorgos, and G. Karagiannidis, "Low-cost cognitive radios against spectrum scarcity," arXiv preprint arXiv:1803.10536, 2018.

[2] S. Sun, T. Rappaport, M. Shafi, et al., "Propagation models and performance evaluation for 5G millimeter-wave bands," IEEE Trans. Veh. Technol., vol. 67(9), 2018, pp.8422-8439.

[3] Z. Yun, M. Iskander, "Radio propagation modeling and simulation using ray tracing," The world of applied electromagnetics. Springer, Cham, 2018, pp.275-299.

[4] R. Levie, C. Yapar, G. Kutyniok, et al., "RadioUNet: Fast radio map estimation with convolutional neural networks," IEEE Trans. Wirel.Commun., vol. 20(6), 2021, pp.4001-4015

[5] O. Ronneberger, P. Fischer, T. Brox, "U-net: Convolutional networks for biomedical image segmentation," International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015, pp.234-241.

[6] A.Vaswani, N. Shazeer, N. Parmar, et al., "Attention is all you need," arXiv preprint arXiv:1706.03762, 2017.

[7] A. Alkhateeb, S. Alex, P. Varkey, et al., "Deep learning coordinated beamforming for highly-mobile millimeter wave systems," IEEE Access, vol. 6, 2018, pp.37328-37348.