Large Language Model-Inspired Transformer Framework for Antenna Design

Chang Ge* and Qiang Chen
Department of Communications Engineering, Tohoku University, Japan
*Email: ge.chang.e5@tohoku.ac.jp

***Abstract* - Antenna design, which aims to infer geometric parameters from desired electromagnetic performance, is inherently a nonlinear inverse problem. An efficient solution, however, is essential in practical engineering. With the recent breakthroughs in large language models (LLMs) in AI, this paper explores the possibility of formulating the antenna design task as a "language translation" problem. We propose an Antenna-Transformer framework that treats antenna directivity patterns as the "source language" and geometric parameters as the "target language," enabling the AI model to learn an implicit mapping between them in a high-dimensional space. Validation through full-wave electromagnetic simulations on a dataset of two-element Yagi-Uda antennas demonstrates that the trained AI model achieves prediction errors within ±0.15 dB for forward gain and ±0.25 dB for backward gain, showing promising accuracy and reliability. This is the first study to propose a Transformer-based architecture for antenna design, offering a new perspective for solving electromagnetic inverse problems.**

***Keywords* — Antenna Inverse Design, Transformer architecture, Large Language Model, Yagi-Uda antennas.**

I. INTRODUCTION

In recent years, large language models (LLMs), such as ChatGPT and DeepSeek, have gained significant attention. These models are all powered by the Transformer architecture [1], which has been regarded as a foundational architecture of modern AI due to its strong global modeling capabilities, high parallelism, and scalability.

Transformer architecture excels at natural language processing (NLP), such as language translation task. In the electromagnetic domain, Transformer-based models have also attracted growing interest, being applied to radar propagation loss prediction [2], inverse scattering [3], and ISAR imaging [4]. These problems can naturally be framed as sequence-to-sequence modeling tasks, making Transformers a well-suited choice for their solution.

In contrast, extending Transformer architecture to antenna design poses greater challenges. Antenna design is a static hardware task that maps the antenna characteristics —mostly directivity patterns—to the antenna's geometric parameters. This process requires the model to learn a precise, continuous, bidirectional relationship between spatial field distributions and electric current distributions. Unlike natural-language tasks—where inputs and outputs are discrete, symbolic sequences—both antenna characteristics and parameters must first be converted into sequential representations compatible with the Transformer architecture.

While some efforts have explored alternative AI methods for antenna inverse design, their performance remains limited compared to Transformer-based architectures. For instance, [5] employs a multilayer perceptron (MLP), but the simple structure of MLPs lacks the capacity to capture global features and is less effective for large-scale inputs. [6] adopts Diffusion model that treats antenna structures as images and generates designs in a pixel-wise fashion. However, such image generative AI models often produce "approximately plausible" results, which fall short in terms of controllability, precision, and computational efficiency—key requirements in antenna engineering. Therefore, in this paper we focus on Transformer-based antenna inverse design. Transformer models support direct regression of continuous parameters, making them better suited for high-precision, controllable design tasks.

Motivated by these observations, we propose the Antenna-Transformer (ANTR) architecture, which adapts the original Transformer from NLP tasks to antenna inverse design. Drawing inspiration from language translation, the proposed framework treats the antenna directivity pattern as a "source language" and the geometric parameters as a "target language", enabling the model to learn a one-to-one mapping between them in a high-dimensional embedding space. This approach enables efficient and controllable prediction of antenna parameters and offers a new perspective for solving electromagnetic inverse problems.

II. MODEL DESIGN

The proposed ANTR builds on the classical Transformer encoder–decoder architecture. During training process in language translation tasks, the Transformer takes a source sequence (such as Germany) as encoder input and a target sequence (such as English) as decoder input, enabling sequence-to-sequence mapping through layers of self-attention and cross-attention. Analogously, for antenna inverse design task, we treat directivity pattern as the source sequence for the encoder and the geometric parameters as the target sequence for the decoder, aiming to map performance requirements to physical design.

In the encoder, following Vision Transformer (ViT) [7], we flatten the three-dimensional directivity pattern into a two-dimensional patch sequence. Positional information is then added to each patch to preserve angular information and enable efficient parallel processing. The encoder applies self-attention layers to extract global features of the directivity pattern.

In the decoder, the standard Transformer relies on autoregressive input of the target sequence, which is not

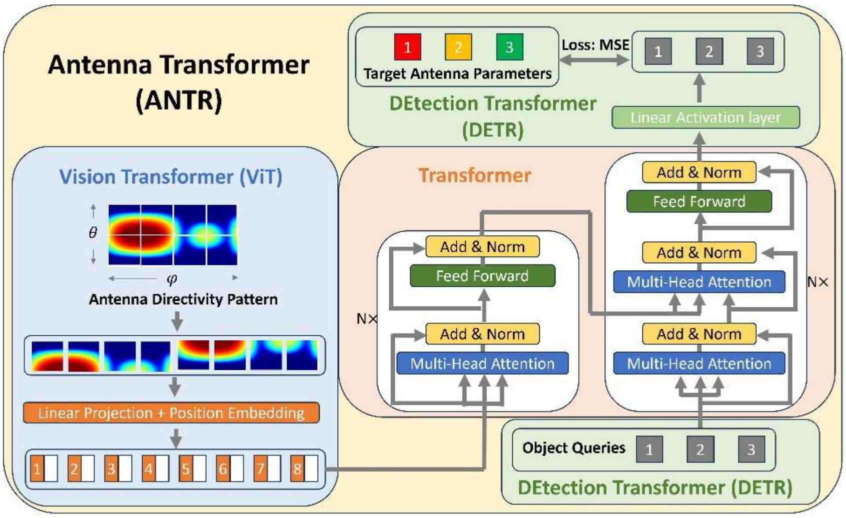

Fig. 1. The architecture of Antenna-Transformer (ANTR).

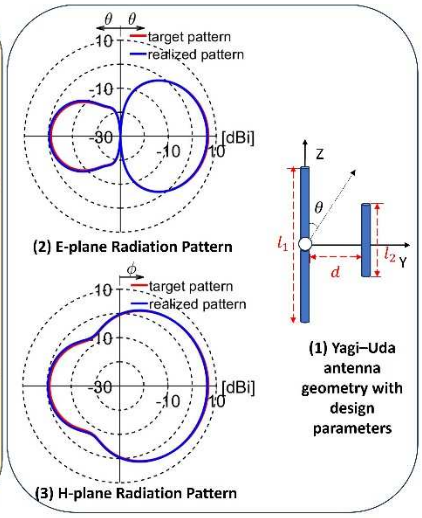

Fig.2. Comparison of predicted and target radiation patterns of the Yagi–Uda antenna.

suitable for antenna parameters, as these values do not possess a natural ordering. To address this, we adopt the strategy from Detection Transformer (DETR) [8], replacing explicit parameter sequences with learnable object queries, allowing for parallel decoding. Each query interacts with the encoder output via cross-attention, capturing relevant global context to predict each geometric parameter.

Furthermore, since antenna parameter prediction is a regression task rather than classification, we replace the softmax classification head with a linear output layer to predict continuous values. The model is trained by minimizing the mean squared error (MSE) between estimated and ground-truth antenna parameters. The overall model architecture is illustrated in Fig. 1.

III. PERFORMANCE EVALUATION

To evaluate the feasibility and accuracy of the proposed ANTR framework for antenna inverse design, we conducted experiments on a simplified two-element Yagi–Uda antenna with parameters l_1,d and l_2 (Fig. 2-1). A dataset of 339, 521 directivity pattern–parameter pairs was generated via Method of Moments (MoM) simulations. We split this dataset into 80% for training and 20% for evaluation: the training set was used to fit ANTR, and the evaluation set to assess the trained ANTR model's prediction performance.

During inference, the desired directivity pattern was input to ANTR, which output predicted geometric parameters l_1', d', l_2'. Rather than comparing these parameters directly—since aggregating them into a single error metric would obscure their individual physical meanings—we reinjected the predicted l_1', d', l_2' into MoM simulations and quantified accuracy by the difference between the realized directivity and the target directivity. Our model achieved forward-directivity errors within ± 0.15 dB and backward-directivity errors within ± 0.25 dB. Fig. 2 also presented a representative case (l_1= 0.50λ, d = 0.20λ, l_2= 0.48λ) comparing the realized and desired directivity patterns. The training process was performed on an NVIDIA RTX 4090 GPU and required approximately two hours.

IV. CONCLUSION

In this work, we proposed Antenna-Transformer (ANTR), a novel framework that adapts the Transformer architecture from LLM to antenna inverse design. By drawing an analogy to language translation, ANTR learns a high-dimensional mapping from directivity patterns to geometric parameters. Preliminary results on two-element Yagi-Uda antennas demonstrate promising prediction accuracy. This study opens new directions for integrating AI and physics-driven modeling.

ACKNOWLEDGMENT

This work was partially supported by the supporting program from Center for Diversity, Equity and inclusion (TUMUG) at Tohoku University.

REFERENCES

[1] A. Vaswani et al., "Attention Is All You Need," https://doi.org/10.48550/arXiv.1706.03762

[2] H. -J. Ji et al., "Explainable Deep Learning-Based Prediction of Radar Wave Propagation Loss in Evaporation Ducts and Exploration of Driving Factors," in IEEE Antennas and Wireless Propagation Letters, doi: 10.1109/LAWP.2025.3565556.

[3] N. Du et al., "Inhomogeneous Media Inverse Scattering Problem Assisted by Swin Transformer Network," in IEEE Transactions on Microwave Theory and Techniques, vol. 72, no. 12, pp. 6809-6820, Dec. 2024, doi: 10.1109/TMTT.2024.3412113

[4] H. Wang et al., "High-Resolution ISAR Imaging Method for Maneuvering Targets Based on a Hybrid Transformer," in IEEE Transactions on Antennas and Propagation, vol. 71, no. 10, pp. 8250-8265, Oct. 2023, doi: 10.1109/TAP.2023.3306386.

[5] M. K. Shereen et al., "Towards Bi-Directional Deep Learning Approach in Patch Antenna Design and Optimization," 2025 IEEE International Workshop on Antenna Technology (iWAT), Cocoa Beach, FL, USA, 2025, pp. 1-4, doi: 10.1109/iWAT64079.2025.10931210.

[6] Q. Pang et al., "Inverse Design Method of Antenna Based on Generative Artificial Intelligence," in IEEE Antennas and Wireless Propagation Letters, doi: 10.1109/LAWP.2025.3562934.

[7] A. Dosovitskiy et al., "An Image Is Worth 16×16 Words: Transformers for Image Recognition at Scale," https://doi.org/10.48550/arXiv.2010.11929

[8] N. Carion et al., "End-to-End Object Detection with Transformers," https://doi.org/10.48550/arXiv.2005.12872