Advances in Engineering Research : Possibilities and Challenges
ISSN Print: 3079-5192
ISSN Online: 3079-5206
https://doi.org/10.63313/AERpc.9062

Architectural Choice in Inverse Design of THz Metamaterials: Local Priors vs. Global Attention under Data Scarcity

Peng Xu. Shuang Wang
College of Engineering, Tianjin University of Technology and Education, No. 1310, Dagu South Road, Hexi District, Tianjin 300222, China
Email: 550184452@qq.com

How to cite this paper: Xu, P., & Wang, S. (2025). Architectural choice in inverse design of THz metamaterials: Local priors vs. global attention under data scarcity. *Advances in Engineering Research: Possibilities and Challenges*, 3(1), 1–11. Print ISSN: 3079-5192; Online ISSN: 3079-5206. https://doi.org/10.63313/AERpc.9062
Published: 2025-12-09

Copyright © 2025 by author(s) and Erytis Publishing Limited.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/

Abstract

The inverse design of terahertz (THz) electromagnetically induced transparency (EIT) metamaterials is complicated by high-dimensional parameter spaces and complex spectral mappings. While deep learning has emerged as a potent tool for addressing these issues, the suitability of distinct neural architectures under data-constrained conditions remains underexplored. Focusing on EIT spectra characterized by long-range frequency dependencies, this study compares the inverse design performance of the Fully Connected Residual Network (FC-ResNet), which leverages local inductive biases, against the Transformer architecture, which relies on global self-attention mechanisms. We evaluated these models using a comprehensive dataset of 20,476 samples and a restricted dataset of only 500 samples. The results demonstrate that in data-rich environments, both architectures achieve exceptional accuracy ($R^2 > 0.99$), indicating that architectural differences do not constitute a bottleneck when data is abundant. However, in small-sample regimes—simulating scenarios with scarce experimental data—the FC-ResNet exhibits significantly superior generalization capabilities compared to the Transformer. Our findings suggest that while Transformers offer global modeling potential, they are prone to overfitting when data is limited; conversely, the structural priors of the FC-ResNet provide essential regularization in such contexts. This work offers empirical guidance for selecting architectures in the intelligent design of metamaterials where data availability is a limiting factor

Keywords

Terahertz metamaterials; Inverse design; Deep learning

1. Introduction

Situated between the microwave and infrared regions, the terahertz (THz) band holds significant potential for next-generation high-speed wireless communications, non-destructive testing, and biochemical sensing, owing to its unique spectral fingerprinting capabilities and transparency to non-polar materials [1-3]. However, the development of high-performance functional devices is often hindered by the weak electromagnetic response of naturally occurring materials in this frequency range. Metamaterials—artificial composite structures formed by periodically arranged sub-wavelength units—offer the degrees of freedom necessary to tailor the amplitude, phase, and polarization of electromagnetic waves [4,5]. In particular, Electromagnetically Induced Transparency (EIT) metamaterials are widely used in high-sensitivity refractive index sensing and slow-light device design. By mimicking quantum interference effects found in atomic systems, they generate an extremely narrow transparency window within a broadband absorption background [6-9].

While the physical mechanisms governing EIT metamaterials are well understood, their inverse design—deducing geometric parameters from a target spectrum—remains a classic non-convex and ill-posed problem. Since the EIT effect arises from near-field coupling between bright and dark modes, even minute perturbations in structural parameters can precipitate drastic changes in the spectral lineshape [10]. Traditional design approaches typically rely on parameter sweeps or heuristic algorithms such as Genetic Algorithms (GA) or Particle Swarm Optimization (PSO). These methods often require thousands of iterative calculations using full-wave electromagnetic simulation solvers (e.g., FDTD or FEM), resulting in prohibitive computational costs without guaranteeing convergence to a global optimum [11-13].

In recent years, Deep Learning (DL) has significantly accelerated the on-demand design of metamaterials by establishing data-driven surrogate models [14-16]. While Convolutional Neural Networks (CNN) dominate tasks involving topological pattern generation, the parameterized inverse design task addressed in this study involves discrete frequency-transmittance spectra as input. Inherently, this data constitutes a one-dimensional (1D) sequence rather than a two-dimensional image; consequently, directly applying convolutional layers designed for spatial feature extraction is not optimal. To address these data characteristics, the Fully Connected Residual Network (FC-ResNet) has been proposed and validated [17-20]. By incorporating residual learning into the Multi-Layer Perceptron (MLP), the FC-ResNet retains the efficiency of fully connected layers in processing 1D vectors while overcoming deep network degradation via shortcut connections, establishing itself as a benchmark model for spectral regression tasks.

Concurrently, in the realm of sequence data processing, the Transformer architecture—based on self-attention mechanisms—has achieved breakthrough success in natural language processing and is increasingly being applied to the physical sciences [21-23]. Unlike the FC-ResNet, which relies on fixed-weight hierarchical mapping, the Transformer dynamically calculates correlation weights between any two frequency points in a sequence. Theoretically, this makes it superior at capturing the long-range physical coupling features inherent in EIT spectra, such as the interactions between distinct resonant modes [24]. However, existing comparative studies predominantly rely on massive datasets, often overlooking the "data scarcity" dilemma ubiquitous in scientific research. In practical scenarios where high-fidelity simulation or experimental data is costly to obtain, a model's generalization capability under small-sample conditions is often more practically valuable than its asymptotic accuracy with abundant data [25].

Addressing these issues, this study systematically evaluates the suitability of the FC-ResNet and Transformer architectures for the inverse design of THz EIT metamaterials. We specifically focus on the distinct inductive biases exhibited by these models when processing 1D spectral sequences. By analyzing training dynamics and generalization loss across both large-scale and small-scale datasets, we elucidate the relationship between model architecture and data availability. Our results indicate that while the Transformer possesses a theoretical advantage in capturing global dependencies, it is highly prone to overfitting in small-sample regimes. Conversely, the FC-ResNet, leveraging its structural priors, demonstrates superior data efficiency and physical robustness when data is limited.

2. Theory and Methods

This study utilizes a classic EIT metamaterial structure as the physical validation platform, with the unit cell geometry illustrated in Figure 1. The structure consists of an aluminum metallic pattern on a silicon substrate, comprising a continuous metal wire acting as the "bright mode" and an asymmetric double U-shaped resonator (SRR) functioning as the "dark mode." Under normal incidence of y-polarized Thz waves, the bright mode is directly excited to yield a broadband response. It subsequently excites the dark mode via near-field coupling; the resulting destructive interference generates a sharp transparency window within the transmission spectrum.

To characterize the electromagnetic response, we developed a full-wave simulation model utilizing the Time Domain Solver in CST Studio Suite. In the simulation setup, a plane wave propagating along the z-axis served as the excitation source. To simulate an infinite two-dimensional periodic array and eliminate boundary reflections, we applied periodic boundary conditions ("Unit Cell") along the x- and y-directions, while assigning open boundary conditions ("Open Add Space") along the z-direction. We selected five key geometric parameters as design variables: the metal wire length L, SRR side length l, line width w, gap size g, and the separation distance s. By performing parameter sweeps within physically reasonable ranges, we generated a total of 20,476 geometry-spectrum sample pairs. The input spectral data was sampled at 501 frequency points ranging from 0.5 THz to 1.0 THz; these were integrated with frequency indices to construct a 1,002-dimensional 1D feature vector. To investigate the impact of data scale, we partitioned the data into a "large dataset," comprising all generated samples, and a "small dataset" consisting of only 500 randomly selected samples. The latter was designed to replicate authentic research scenarios characterized by a scarcity of experimental data.

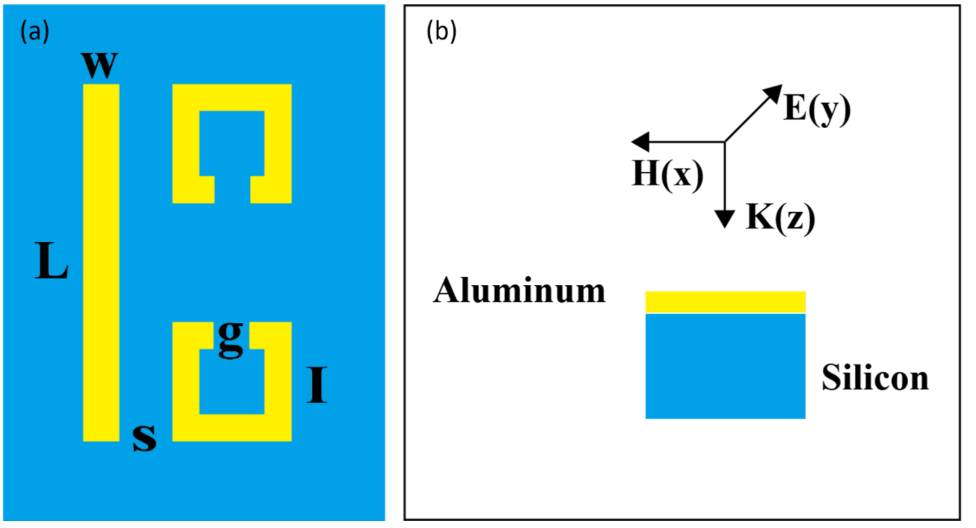

Figure 1. Schematic of the Thz EIT metamaterial and simulation setup. (a) Top view of the unit cell, consisting of a continuous metal wire (bright mode) and a pair of asymmetric split-ring resonators (dark mode). Key geometric parameters (L,s,w,g,I) are labeled. (b) 3D perspective view illustrating the excitation by a y-polarized plane wave, with periodic boundary conditions in the x-y plane and open boundary conditions in the z-direction.

To address the inherent 1D sequential nature of the spectral data, we constructed the FC-ResNet architecture, as illustrated in Figure 2 Distinct from convolutional networks typically used for image processing, this model employs a custom-designed ResidualFC block. By integrating linear layers with shortcut connections, this design effectively mitigates the gradient degradation often encountered in deep networks. To prevent overfitting during small-sample training, we incorporated a Dropout layer with a rate of 0.2 within the residual blocks. In terms of overall topology, the network adopts a "funnel-like" design: high-dimensional input spectral features traverse a stack of 14 residual blocks where the number of channels is progressively reduced, culminating in the output of the five geometric parameters via a regression head. This hierarchical process of dimensionality reduction imposes a local inductive bias, guiding the model to strip away redundant noise and extract essential physical features.

To evaluate distinct architectural characteristics, we developed a Transformer model based on self-attention mechanisms, as illustrated in Figure 3 Recognizing the long-range physical couplings among resonant modes inherent in EIT spectra, the Transformer utilizes a Multi-Head Self-Attention mechanism to dynamically capture these non-local dependencies. The core attention computation is defined by Equation (1):

$$Attention(Q,K,V) = soft\max(\frac{QK^T}{\sqrt{d_k}}) \tag{1}$$

Here, Q, K, and V denote the query, key, and value matrices, respectively, while d_k serves as the scaling factor. The data processing pipeline proceeds as follows: the input spectrum is first mapped via an embedding layer and concatenated with a learnable classification token ([CLS] token) and positional encodings. Subsequently, the sequence enters a stack comprising four encoder layers. Ultimately, the model extracts solely the output vector of the [CLS] token, mapping it to the predicted parameters via a MLP. This architectural design significantly relaxes structural constraints, endowing the model with enhanced global representation capabilities.

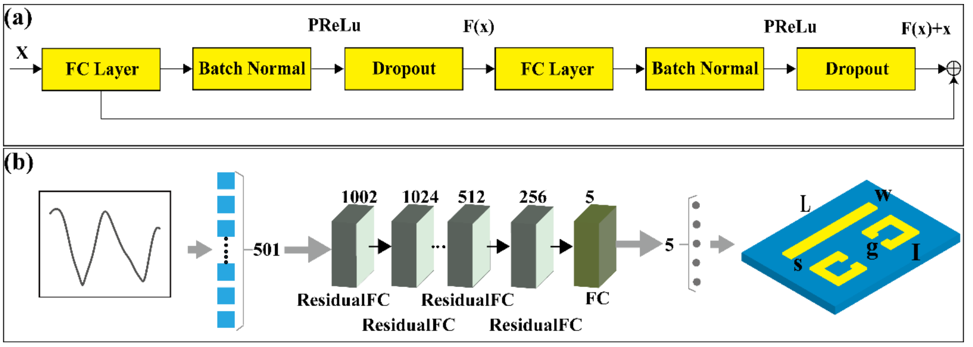

Figure 2. Architecture of the proposed FC-ResNet model for inverse design. (a) Detailed structure of the ResidualFC block, which incorporates Batch Normalization (BN), PReLU activation, and Dropout for regularization. (b) The overall network topology, featuring a "funnel-shaped" structure that progressively reduces feature dimensionality from the input spectrum to the output geometric parameters.

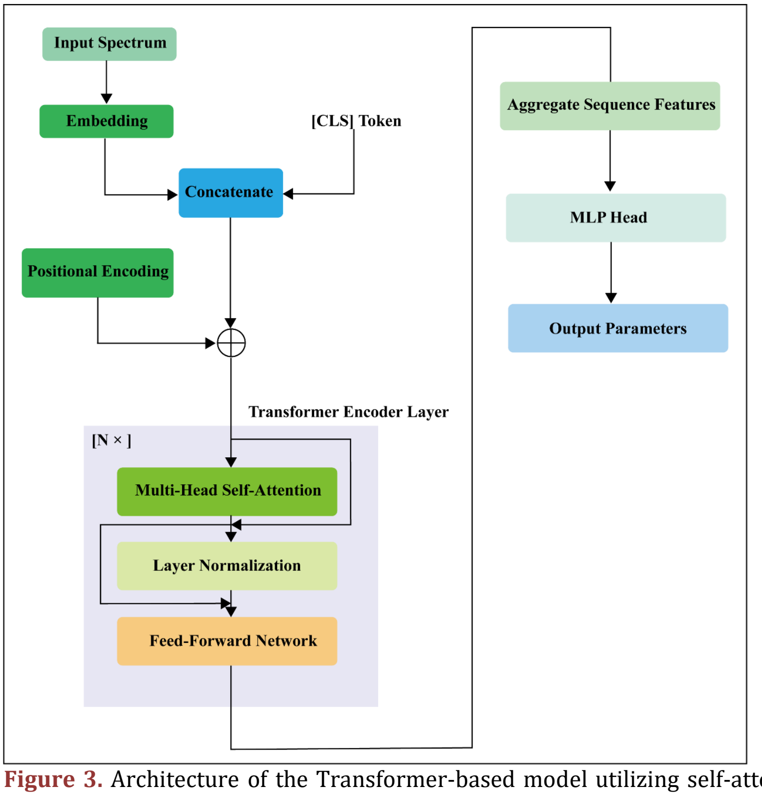

Figure 3. Architecture of the Transformer-based model utilizing self-attention mechanisms. The input spectrum is first mapped via embedding layers and combined with a learnable [CLS] token and positional encodings. The sequence is then processed by a stack of Transformer

encoder layers to capture long-range spectral dependencies, finally predicting structure parameters through an MLP head.

All models were evaluated within a unified training framework. To eliminate scale discrepancies, both input spectra and output parameters were standardized using the RobustScaler. During training, Mean Squared Error (MSE) was employed as the loss function to quantify the Euclidean distance between the network-predicted parameters and the ground truth geometric parameters. We utilized the AdamW optimizer paired with a cosine annealing learning rate scheduler; the initial learning rate was set to 5×10^{-4} for the FC-ResNet and 5×10^{-5} for the Transformer. Furthermore, to achieve closed-loop validation of the inverse design, we pre-trained a high-precision forward prediction surrogate model to rapidly verify the physical response of the structures generated by the inverse models.

3. Results and Discussion

To establish a performance benchmark for deep neural networks in the inverse design of THz metamaterials, we conducted comprehensive training and evaluation of the FC-ResNet and Transformer models using the full dataset of 20,476 samples. This phase was designed to validate the non-linear feature extraction capabilities of both architectures under data-rich conditions. Table 1 details the quantitative evaluation metrics for both models on the test set. The results indicate that given ideal data availability, both architectures achieve exceptional prediction accuracy. The FC-ResNet attained a remarkable R2 of 0.9997, with the Transformer following closely at 0.9992. Regarding the Mean Squared Error (MSE), both models converged to the order of 10^{-4}, implying that the deviation between predicted and ground-truth geometric parameters falls well below standard micro- and nanofabrication tolerances.

Table 1. Quantitative performance comparison on the large dataset.

Model Architecture	R^2 Score	MSE	MAE
FC-ResNet	0.9997	0.0003	0.0139
Transformer	0.9992	0.001250	0.019946

To visually validate the prediction results from a physical perspective, we randomly selected a sample exhibiting typical EIT characteristics from the test set. Using the trained inverse models, we predicted its geometric structure and subsequently fed these parameters into the forward surrogate model to reconstruct the spectrum. The comparative results are depicted in Figure 4. It is evident that the spectra reconstructed by both the FC-ResNet (red dashed line) and the Transformer (blue dotted line) achieve a near-perfect overlap with the target spectrum (black solid line). Specifically, key features—including the center frequency of the transparency window, the intensity of the transmission peak, and the asymmetric absorption dips flanking the peak arising from bright-dark mode coupling—are accurately reproduced. These results compellingly demonstrate that when training data sufficiently covers the high-dimensional parameter space, the choice of neural network architecture is not the primary factor limiting inverse design accuracy. Both the residual network, relying on local feature extraction, and the Transformer, based on global attention mechanisms, possess sufficient model capacity to accurately approximate the complex mappings governing electromagnetic responses.

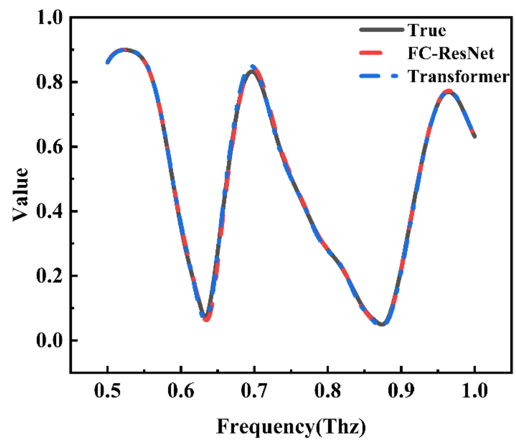

Figure 4. Comparison of inverse design performance on the large dataset. The spectra reconstructed by both FC-ResNet (red dashed line) and Transformer (blue dotted line) show excellent agreement with the target spectrum.

While the performance on the large dataset is impressive, acquiring tens of thousands of labeled data points for practical optical experiments or high-precision full-wave simulations is often prohibitively time-consuming and costly. Consequently, evaluating model performance under "small-sample" conditions holds greater practical relevance. To address this, we retained only 2.5% of the original dataset (totaling 500 samples)—split evenly into 250 for training and 250 for testing—and retrained both models from scratch. In this data-constrained scenario, the performance of the two models exhibited significant divergence. As shown in Table 2, the FC-ResNet demonstrated remarkable data efficiency; its R2 score remained high at 0.9086, while the MSE rose only slightly to 0.0544, indicating that it remains capable of accurately predicting structural parameters within a general range. In contrast, the Transformer suffered severe performance degradation. Its R2 dropped to 0.6089 and the MSE surged to 0.2874, suggesting that its predictions contain substantial random errors and fail to meet the requirements for precise design

Table 2. Quantitative performance comparison on the small dataset

Model Architecture	R^2 Score	MSE	MAE
FC-ResNet	0.9086	0.0544	0.1225
Transformer	0.6089	0.287405	0.403955

To elucidate the underlying mechanisms driving this performance divergence, we detailed the loss function trajectories during training, as depicted in Figure 5. These "learning curves" reveal starkly contrasting training dynamics between the two architectures. As observed in the Transformer curves in Figure 5(a), the training loss rapidly descends to near-zero within the initial epochs. This indicates that the model, leveraging its extensive parameter space, easily resorted to rote memorization of the training samples. However, the testing loss, following a brief decline, plateaued at a high magnitude and even exhibited oscillations in the later stages. This substantial discrepancy between training and testing errors—the "generalization gap"—confirms that the Transformer succumbed to severe overfitting in the small-sample regime. Lacking intrinsic assumptions regarding data locality (a characteristic reflecting weak inductive bias), the self-attention mechanism requires massive datasets to discern physical laws; without them, it is prone to fitting noise.

In sharp contrast, Figure 5(b) demonstrates that the test loss of the FC-ResNet closely tracks the decline of the training loss. This superior generalization is attributed to the strong inductive bias embedded within its architecture: the hierarchical structure and residual connections of the FC-ResNet implicitly encode the prior assumption that features should be abstracted hierarchically and evolve smoothly. Such structural constraints serve as potent regularization, compelling the model to disregard noise during data scarcity and focus on learning the core features governing EIT physics.

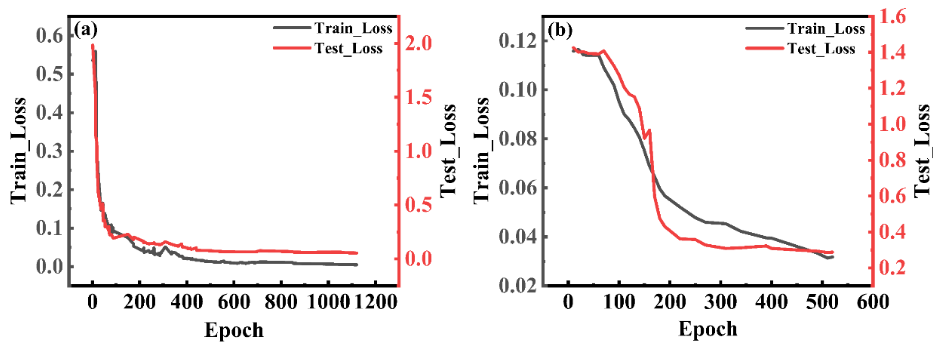

Figure 5. Analysis of training dynamics on the small dataset. (a) Loss curves for the Transformer model show a rapid decrease in training loss but a stagnant testing loss, revealing a large generalization gap and severe overfitting. (b) Loss curves for the FC-ResNet model demonstrate a consistent correlation between training and testing losses, indicating robust generalization capabilities attributed to its structural inductive bias.

For metamaterial designers, statistical metrics are merely indicative; the physical validity of the final design constitutes the ultimate criterion. To ascertain the practical utility of the FC-ResNet in small-sample scenarios, we conducted a physical closed-loop validation. We fed the geometric parameters predicted by the FC-ResNet (trained on the limited dataset) into the forward surrogate model to assess the corresponding spectral response. As illustrated in Figure 6, despite a 97.5% reduction in training data, the structure inversely designed by the FC-ResNet (red curve) reproduced the key features of the target spectrum (black curve) with remarkable accuracy. Of particular note is the transparency window—the feature most sensitive to the EIT effect—where the model's prediction exhibited negligible deviation. This finding provides a critical empirical basis for establishing data-efficient photonics design platforms: when confronting real-world scenarios with high data acquisition costs, selecting the FC-ResNet—with its appropriate inductive bias and compact structure—offers greater engineering utility than the blind pursuit of complex Transformer models.

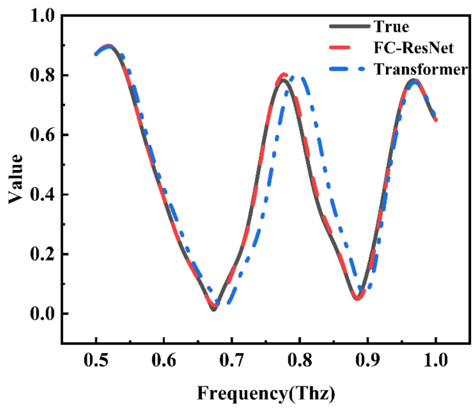

Figure 6. Closed-loop physical verification of the FC-ResNet model trained on the small dataset. Despite the scarcity of training data, the reconstructed spectrum closely matches the target EIT resonance peak and bandwidth, confirming the model's practical utility for data-limited inverse design.

4. Conclusion

Focusing on Thz EIT metamaterials characterized by complex physical coupling mechanisms, this study systematically evaluates and compares the suitability of the FC-ResNet, which relies on local inductive biases, against the Transformer architecture, grounded in global attention mechanisms, for inverse design tasks. Through comprehensive validation across both large-scale and small-scale datasets, we elucidate the intrinsic interplay among neural network architecture, data volume, and design performance.

The results indicate that under ideal conditions with abundant training data, architectural differences do not constitute a bottleneck limiting performance. Both the FC-ResNet and the Transformer are capable of establishing high-precision spectrum-structure mappings, achieving accurate reconstruction of EIT transparency windows and quality factors. However, under data-scarce conditions that more closely mirror real-world research scenarios, the two models exhibit starkly distinct learning behaviors. Lacking structural prior constraints, the Transformer demonstrates a pronounced "data hunger"; in small-sample training regimes, it is highly prone to severe overfitting, resulting in a significant deterioration of generalization capability. Conversely, by virtue of the strong inductive bias implicit in its residual connections and hierarchical structure, the FC-ResNet maintains high prediction reliability and physical robustness even in extreme scenarios where data is minimal.

In summary, this study confirms that in the inverse design of photonics, the blind pursuit of complex attention-based models is not always the optimal solution. For design tasks contingent upon high-cost experiments or high-fidelity full-wave simulations, selecting the FC-ResNet—characterized by appropriate inductive biases and a compact structure—often yields greater data efficiency and engineering utility than the Transformer. This finding not only delineates the applicability boundaries of different deep learning architectures in metamaterial design but also provides critical empirical evidence and strategic guidance for the future development of low-cost, high-efficiency platforms for intelligent Thz device design.

References

[1] Nagatsuma T, Ducournau G, Renaud C C. Advances in terahertz communications accelerated by photonics[J]. Nature Photonics, 2016, 10(6): 371-379.

[2] Mittleman D M. Twenty years of terahertz imaging[J]. Optics express, 2018, 26(8): 9417-9431.

[3] O'Hara J F, Singh R, Brener I, et al. Thin-film sensing with planar terahertz metamaterials: sensitivity and limitations[J]. Optics express, 2008, 16(3): 1786-1795.

[4] Yu N, Capasso F. Flat optics with designer metasurfaces[J]. Nature materials, 2014, 13(2): 139-150.

[5] Chen H T, Padilla W J, Zide J M O, et al. Active terahertz metamaterial devices[J]. Nature, 2006, 444(7119): 597-600.

[6] Papasimakis N, Fedotov V A, Zheludev N I, et al. Metamaterial analog of electromagnetically induced transparency[J]. Physical Review Letters, 2008, 101(25): 253903.

[7] Zhang S, Genov D A, Wang Y, et al. Plasmon-induced transparency in metamaterials[J]. Physical review letters, 2008, 101(4): 047401.

[8] Liu N, Weiss T, Mesch M, et al. Planar metamaterial analogue of electromagnetically induced transparency for plasmonic sensing[J]. Nano letters, 2010, 10(4): 1103-1107.

[9] Gu J, Singh R, Liu X, et al. Active control of electromagnetically induced transparency analogue in terahertz metamaterials[J]. Nature communications, 2012, 3(1): 1151.

[10] Molesky S, Lin Z, Piggott A Y, et al. Inverse design in nanophotonics[J]. Nature Photonics, 2018, 12(11): 659-670.

[11] Haupt R L. An introduction to genetic algorithms for electromagnetics[J]. IEEE Antennas and Propagation Magazine, 1995, 37(2): 7-15.

[12] Robinson J, Rahmat-Samii Y. Particle swarm optimization in electromagnetics[J]. IEEE transactions on antennas and propagation, 2004, 52(2): 397-407.

[13] Peurifoy J, Shen Y, Jing L, et al. Nanophotonic particle simulation and inverse design using artificial neural networks[J]. Science advances, 2018, 4(6): eaar4206.

[14] Liu D, Tan Y, Khoram E, et al. Training deep neural networks for the inverse design of nanophotonic structures[J]. Acs Photonics, 2018, 5(4): 1365-1369.

[15] Ma W, Liu Z, Kudyshev Z A, et al. Deep learning for the design of photonic structures[J]. Nature photonics, 2021, 15(2): 77-90.

[16] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[17] Asano T, Noda S. Optimization of photonic crystal nanocavities based on deep learning[J]. Optics express, 2018, 26(25): 32704-32717.

[18] Wiecha P R, Muskens O L. Deep learning meets nanophotonics: a generalized accurate predictor for near fields and far fields of arbitrary 3D nanostructures[J]. Nano letters, 2019, 20(1): 329-338.

[19] Malkiel I, Mrejen M, Nagler A, et al. Plasmonic nanostructure design and characterization via deep learning[J]. Light: Science & Applications, 2018, 7(1): 60.

[20] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.

[21] Dosovitskiy A. An image is worth 16x16 words: Transformers for image recognition at scale[J]. arXiv preprint arXiv:2010.11929, 2020.

[22] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[C]//Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers). 2019: 4171-4186.

[23] Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold[J]. nature, 2021, 596(7873): 583-589.

[24] Karniadakis G E, Kevrekidis I G, Lu L, et al. Physics-informed machine learning[J]. Nature Reviews Physics, 2021, 3(6): 422-440.

[25] Zhang Y, Ling C. A strategy to apply machine learning to small datasets in materials science[J]. Npj Computational Materials, 2018, 4(1): 25.