(Received 17 July 2024, Accepted 13 August 2024, Scheduled 24 August 2024)

A Lightweight Deep Learning Model for Full-Wave Nonlinear Inverse Scattering Problems

Yixin Xia and Siyuan He*

College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310058, China

ABSTRACT: Nowadays, deep learning schemes (DLSs) have gradually become one of the most important tools for solving inverse scattering problems (ISPs). Among DLSs, the dominant current scheme (DCS), which extracts physical features from the dominant components of the induced currents, has shown its successes by simplifying the learning process in solving ISPs. It has shown excellent performance in terms of efficiency and accuracy, but the increasing number of channels in DCS often requires higher computational costs and memory usage. In this paper, a lightweight deep learning model for DCS is proposed to reduce the burden of memories in the training and testing processes of network structure. And extensive tests of the model are conducted, where comparisons with results from the U-Net structure are provided. The comparison results validate its potential application in utilizing DCS under limited resource conditions.

1. INTRODUCTION

Imaging techniques based on the inverse scattering technique are widely applied in many critical fields, such as granary inspection in agriculture [1], imaging diagnostics in biomedicine [2], and exploration in seismology [3–5]. However, ISPs face two primary challenges: nonlinearity and ill-posedness [6], which to some extent restrict their further applications. To address these difficulties, alongside the rapid advancement of deep learning, utilizing deep learning solutions for ISPs [7–10] has garnered significant attention. This approach is primarily categorized into four types [11]: direct learning approach, learning-assisted objective-function approach [12–14], physics-assisted learning approach [15–17], and other approaches. Among them, physics-assisted learning approach integrates physical models and mathematical properties into the input or internal architecture of neural networks, which enhances data efficiency and generalization capabilities, thus attracts extensive attentions. The dominant current scheme (DCS) proposed on this basis has gained more application due to its superior performance in handling highly nonlinear problems and achieving precise reconstruction.

Nevertheless, in addressing practical problems, there is often a need to balance between high accuracy and low computational cost. Storing and computing DCS based on the U-net structure remains a significant challenge when only limited resources are available. In many fields, such as medical lesion imaging [18–20], real-time human body security screening [21, 22], and crop disease recognition and type identification [23–25], many efforts are made in compressing and refining the U-net model. Inspired by this, this work proposes a lightweight improvement scheme for the U-net structure used in DCS, as shown in Fig. 1.

In this work, we use $\overline{X}$ and $\overline{\overline{X}}$ to denote the vector and matrix of the discretized operator or parameter X, respectively. Furthermore, we use $\|\cdot\|_F$ and $\|\cdot\|$ to denote Frobenius norm of a matrix and Euclidian length of a vector, respectively.

2. PRINCIPLE

2.1. Theory

ISPs refer to a class of problems where the interaction between incident waves and a scattering object (the object under test) is used to reconstruct features such as the sizes, shapes, and material compositions of the scattering objects based on measurements of the secondary field (scattered field) excited by the scattering object [6]. In our consideration, under the condition of a two-dimensional transverse magnetic field in free-space background, a non-magnetic scattering object is located in the domain of interest (DOI), $D \subset R^2$, and illuminated by N_i line sources positioned at r_p^i in the z-direction, with $p = 1, 2, \cdots, N_i$.

The relationship between the electric field and the scattering object is described by the following Lippmann-Schwinger equation:

$$E^t(r) = E^i(r) + k_0^2 \int_D g\left(r, r'\right) \xi\left(r'\right) E^t\left(r'\right) dr', \quad \text{for } r \in D \tag{1}$$

where $E^t(r)$ and $E^i(r)$ respectively represent the total electric field and incident electric field. $k_0 = \omega\sqrt{\mu_0\varepsilon_0}$ is the free-space wave number, and $g(r, r')$ is the 2-D free space Green's function. Then, $\xi(r) = \varepsilon_r(r) - 1$ represents the contrast permittivity.

* Corresponding author: Siyuan He (12231036@zju.edu.cn).

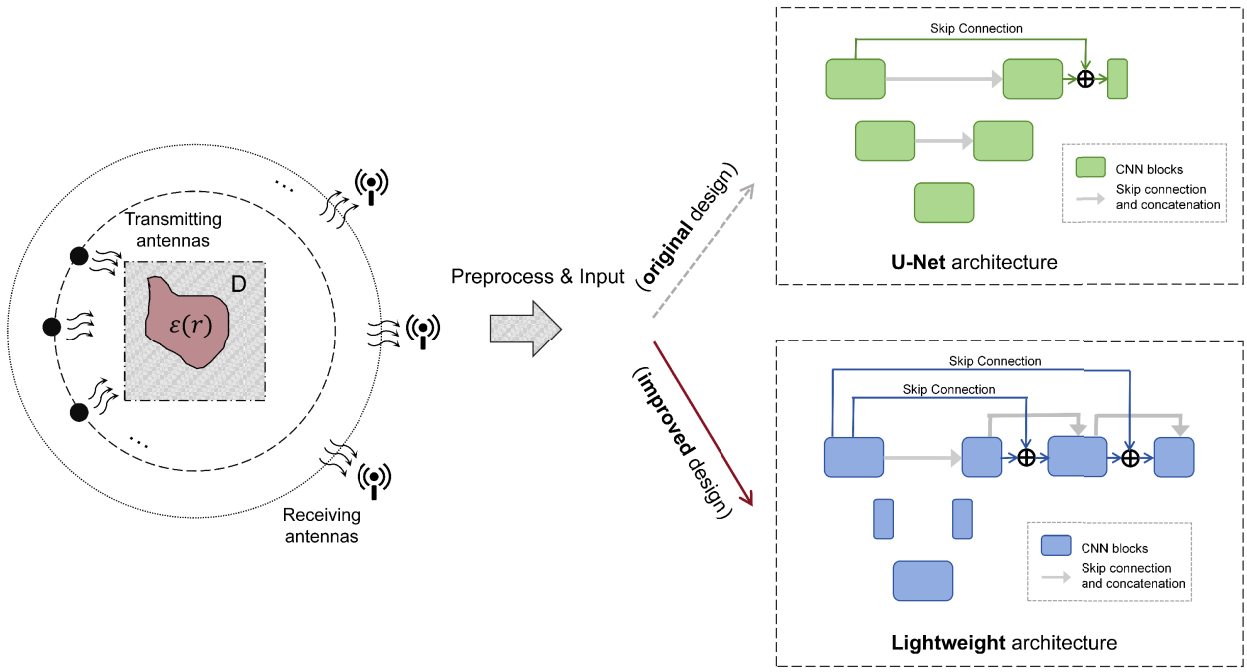

FIGURE 1. Schematic architecture of the original model and the proposed model.

The scattered electric field collected can be formulated by:

$$E^s(r) = k_0^2 \int_D g\left(r, r'\right) \xi\left(r'\right) E^t\left(r'\right) dr', \quad \text{for } r \in S \quad (2)$$

where $E^s(r)$ is the scattered field measured on the surface. Therefore, the inverse problem is to reconstruct the permittivity $\varepsilon_r(r)\,(r \in D)$ from the measured scattered field.

The starting point of the dominant current scheme is to obtain the solution of the dominant part of the unknowns and use this solution as an initial value to simplify the problem.

$$\overline{I}^d = \overline{I}^+ + \overline{I}^l \quad (3)$$

$$\overline{\xi}_p^d(n) = \frac{\overline{I}_p^d(n) \cdot \left[\overline{E}_p^{t,b}(n)\right]^*}{\left\|\overline{E}_p^{t,b}(n)\right\|^2} \quad (4)$$

where $\overline{I}^l$ and $\overline{I}^+$ are the low-frequency induced current and induced current, respectively. Then $\overline{\xi}_p^d(n)$ represents the nth element of the contrast, and $\overline{E}_p^{t,b}$ is the updated total electrical field of DCS for the pth incidence.

Specifically, through eigenvalue decomposition, the dominant part of the current which is computed from Eq. (3) generated by each incident is identified. Based on this dominant current, an approximate contrast of relative permittivity is derived from Eq. (4), which serves as input for subsequent CNN networks. The advantage of this approach lies in significantly reducing computational costs and excluding the high-frequency components of induced currents that are susceptible to input noise contamination, thereby concentrating most of the information from natural images in the low spatial frequency band [26].

2.2. Network Architecture

The proposed U-net convolutional neural network structure [26], as shown in Fig. 2, consists of a contraction path (left side) and an expansion path (right side). The advantage of this architecture for ISPs lies firstly in the incorporation of skip connections within the neural network, allowing information to flow directly from the input to the output layers. This reduces the path length over which information needs to propagate in the network and effectively mitigates the issues of vanishing or exploding gradients during deep neural network training [27, 28]. Secondly, as downsampling proceeds, the receptive field gradually enlarges, expanding the area perceivable per unit area [29, 30]. This enhances the perception of low-frequency information in images and facilitates the capture of large-scale contour information. Simultaneously, during the upsampling recovery process, information from the downsampling stages is integrated, enabling the network to effectively preserve all aspects of the image.

Therefore, in our lightweight design process, we retain skip connections and upsampling-downsampling operations. The specific structure is illustrated in Fig. 3.

The left contraction path consists of repeated applications of 3×3 convolutions and Rectified Linear Units (ReLU), followed by 2×2 max-pooling operations. In the right expansion path, 3×3 upsampling replaces max-pooling, along with a corresponding connection (feature fusion) from the contraction to expansion paths and two residual connections. Additionally, two skip connections are added in the expansion path.

In the lightweight structure, more residual connections are employed to maintain network simplicity while ensuring effective information transmission and learning, thereby improving network performance. Secondly, this structure integrates

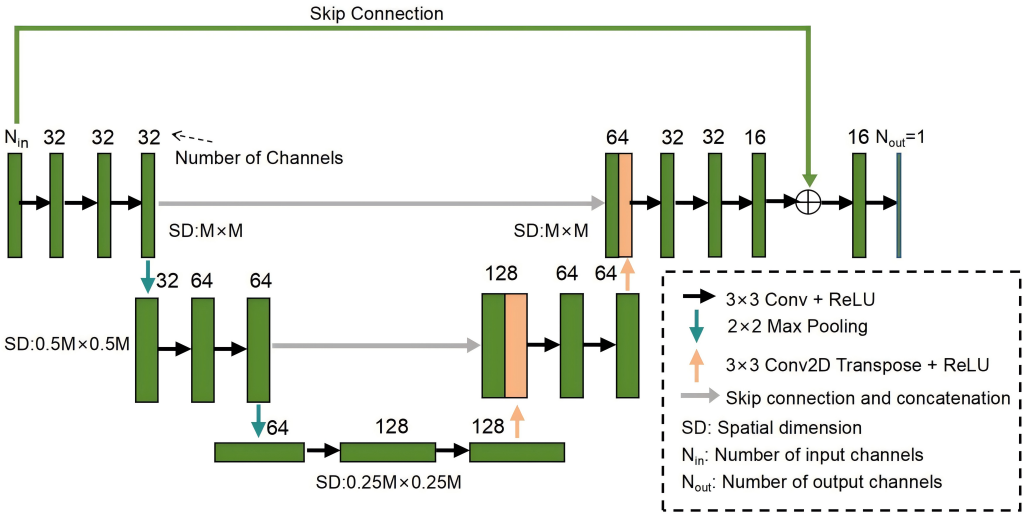

FIGURE 2. U-Net architecture for the DCS [26].

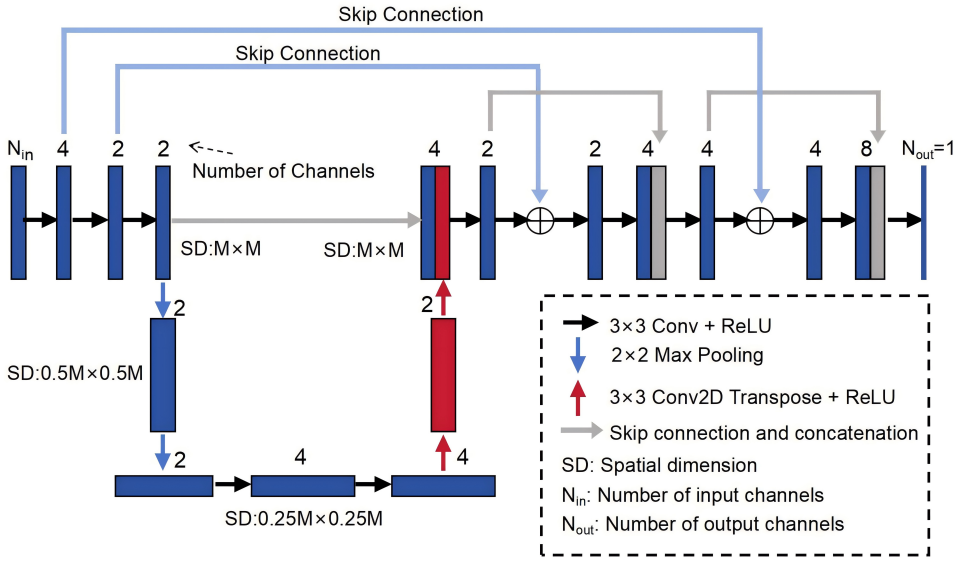

FIGURE 3. Lightweight architecture for the DCS.

features before and after residual connections, obtaining more comprehensive and integrated feature representations that aid in optimizing the model's training process and enhancing its performance. Thirdly, this design significantly reduces the required number of channels during training, decreases network complexity, and requires fewer computational resources, making it suitable for operation in resource-constrained environments.

3. RESULTS

3.1. Environment and Numerical Settings

For fair comparison, all experiments were conducted on the same personal computer with a CPU of 2.5-GHz Intel Core i5 Processor and 16-GB RAM and a GPU of NVIDIA GeForce MX450. All networks and operations in this work were implemented using TensorFlow 2.9.0. The hyperparameter configurations for this work are shown in Table 1.

TABLE 1. Hyperparameter configurations.

Parameter	Setting
Image resolution(H×W)	64×64
Input/Output channels(C)	16 / 1
Kernel size (K_1×K_2)	3×3

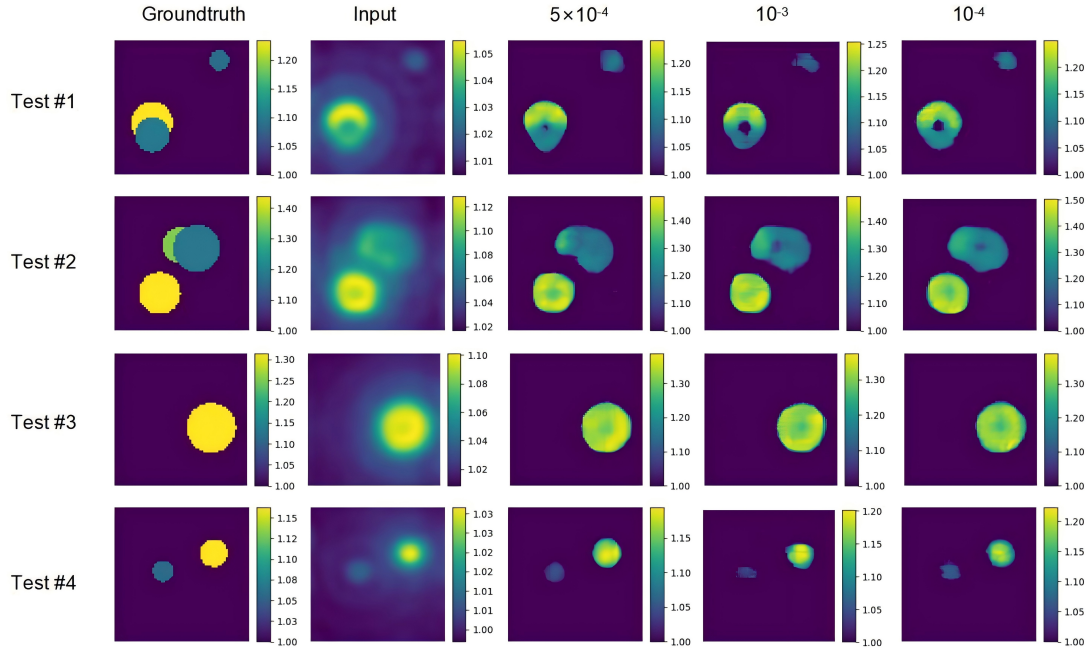

FIGURE 4. Test results of lightweight training models under learning rates of 5×10^{-4}, 10^{-3}, 10^{-4}.

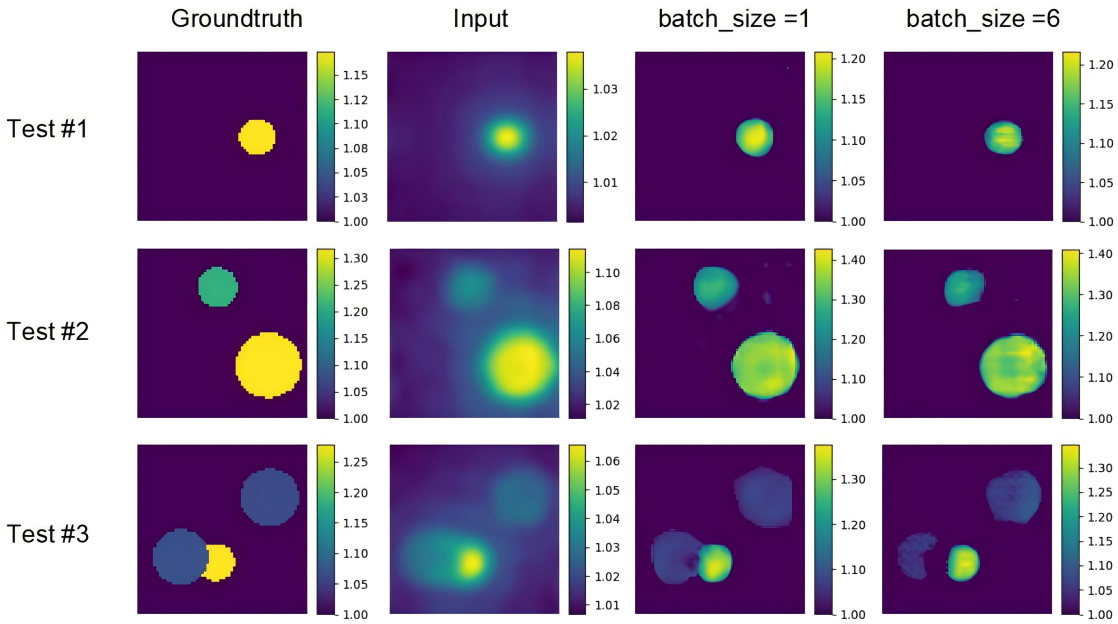

FIGURE 5. Test results of lightweight training models under batch size of 1, 6.

3.2. Model Tests

In the example, the relative permittivity is set between 1 and 1.5. The lightweight network and U-net are first trained with different hyperparameters (learning rate, batch size), followed by testing on 50 different profiles using the trained networks. To quantitatively evaluate performance, the relative errors for these 50 tests are further calculated. Finally, based on the results, optimal hyperparameters suitable for the lightweight network are selected, and the performance of both the lightweight network and U-net are comprehensively evaluated.

First of all, training comparisons were conducted with learning rates of 10^{-3}, 5×10^{-4}, and 10^{-4}, and a batch size of 1. From Fig. 4, it can be observed that the network performs best with a learning rate of 5×10^{-4}. A learning rate of 10^{-4} can only reconstruct simpler profiles, such as single or two non-overlapping cylinders. For more complex profiles, a learning rate of 10^{-4} can only show some relative positions, while

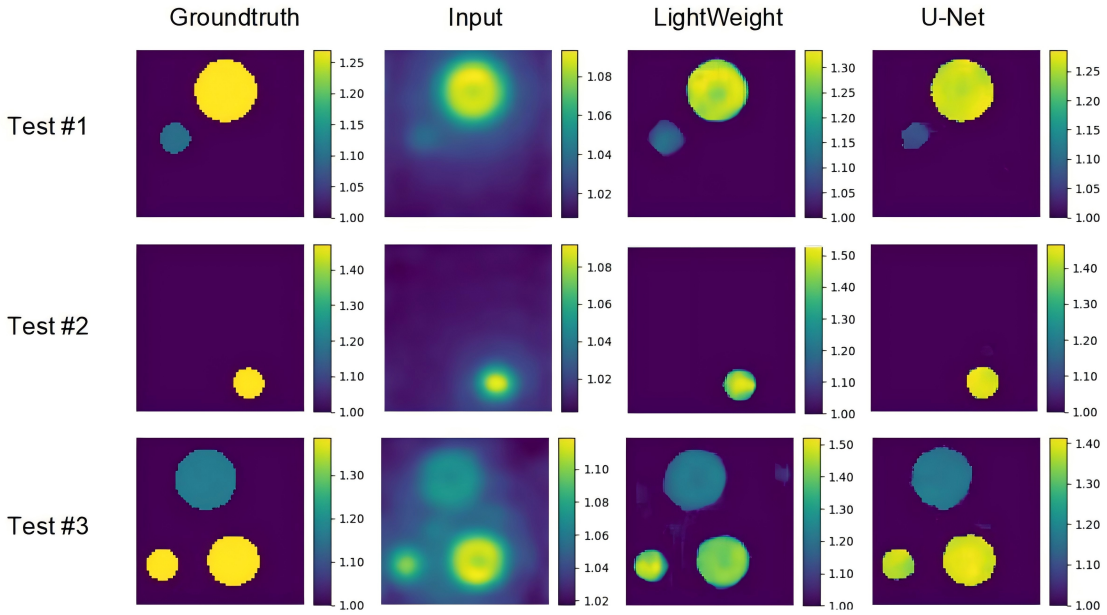

FIGURE 6. Test results of lightweight and U-Net training models under batch size of 1 and learning rate of 5×10^{-4}.

TABLE 2. MSEs and trainable parameters for lightweight and U-Net under different learning rates and batch sizes.

Model	batch_size = 1 lr = 5×10^{-4}	batch_size = 1 lr = 10^{-3}	batch_size = 1 lr = 10^{-4}	batch_size = 6 lr = 5×10^{-4}	Trainable params
LightWeight	1.81%	1.94%	1.99%	2.14%	1195
U-Net	1.72%	1.74%	1.84%	1.79%	480209

shapes reconstructed with 10^{-3} are not as complete as those reconstructed with 5×10^{-4}.

Secondly, comparisons were made by training with a learning rate of 5×10^{-4}, using batch sizes of 1 and 6. From Fig. 5, it can be seen that performance is superior when the batch size is 1. With a batch size of 1, reconstructed shapes exhibit clearer contours and edges. However, with a batch size of 6, the reconstructed images show some degree of degradation, only able to display the relative positions and basic outlines of the shapes.

Thirdly, comparisons were made by training the lightweight model and U-Net model with a learning rate of 5×10^{-4} and a batch size of 1. From Fig. 6, it can be observed that the results of the lightweight model already shows complex profiles relatively clearly, with minor deficiencies mainly in the sharpness of edges and smoothness of colors within the profiles.

$$\text{MRE} = \frac{1}{N}\sum_{i=1}^{N}\left\|\bar{\bar{\varepsilon}}_r^{\,t} - \bar{\bar{\varepsilon}}_r^{\,r}\right\|_F \Big/ \left\|\bar{\bar{\varepsilon}}_r^{\,t}\right\|_F \tag{5}$$

where $\bar{\bar{\varepsilon}}_r^{\,t}$ and $\bar{\bar{\varepsilon}}_r^{\,r}$ respectively represent the real relative permittivity and the reconstructed relative permittivity of the profile. N is the number of tests.

Lastly, we quantitatively evaluated each training result using the Mean Relative Error (MRE) as shown in Eq. (5) and calculated the number of trainable parameters for both models.

In Table 2, it can be seen that for ISP data, the lightweight model performs best when hyperparameters such as learning rate of 5×10^{-4} and batch size of 1 are chosen. Although its ability to reconstruct images is slightly inferior to the original U-Net model, it reduces trainable parameters by 99.75% with only a 0.09% increase in MRE, which significantly lowers the computational resources required. This demonstrates that lightweight models have great potential for applications where high precision is not critical, and resources are limited.

4. CONCLUSION

In this paper, we propose a simple yet effective lightweight DCS approach to address the inverse scattering problem. Unlike traditional U-Net structures, we reduce the number of channels and introduce residual connections and feature fusion between contraction-expansion paths and internal expansion paths. This approach achieves significant compression of the overall structure while maintaining basic reconstruction accuracy, yielding satisfactory results.

In practical applications, the advantage of lightweight models in reducing model size effectively decreases the storage space required for handling large-scale datasets and complex data model matching. This makes it easier to deploy models on embedded and mobile devices. Simultaneously, when integrating technologies for solving nonlinear inverse scattering problems with other systems, lightweight structures reduce the consumption of computational resources, thereby enhancing computational efficiency, and maintaining high performance in faster training times. Moreover, simplified network architectures and fewer channels make lightweight models easier to maintain, which is crucial for reliability and maintainability in practical engineering applications.

Furthermore, there is potential for further research to minimize performance degradation in lightweight models, opening up more possibilities for practical applications.

REFERENCES

[1] Khoshdel, V., M. Asefi, A. Ashraf, and J. LoVetri, "A multi-branch deep convolutional fusion architecture for 3D microwave inverse scattering: Stored grain application," *Neural Computing and Applications*, Vol. 33, 13 467–13 479, 2021.

[2] Quarteroni, A., L. Formaggia, and A. Veneziani, *Complex Systems in Biomedicine*, Springer, Milan, Italy, 2006.

[3] Mao, W., W. Li, and W. Ouyang, "Review of seismic inverse scattering migration and inversion," *Reviews of Geophysics and Planetary Physics*, Vol. 52, No. 1, 27–44, 2021.

[4] Ding, K., S. G. Song, and Z. Q. Xie, "Development and future application of inverse scattering theory," *Progress in Geophysics (in Chinese)*, Vol. 20, No. 3, 661–666, 2005.

[5] Fang, X. Z., F. L. Niu, and D. Wu, "Least-squares reverse-time migration enhanced with the inverse scattering imaging condition," *Chinese Journal of Geophysics*, Vol. 61, No. 9, 3770–3782, 2018.

[6] Chen, X., *Computational Methods for Electromagnetic Inverse Scattering*, John Wiley & Sons, Singapore, 2018.

[7] Zhou, H., X. Huang, and Y. Wang, "Nonlinear inverse scattering imaging method based on iterative multi-scale network," *Chinese Journal of Radio Science*, Vol. 37, No. 6, 1019–1024, 2022.

[8] Li, L., L. G. Wang, F. L. Teixeira, C. Liu, A. Nehorai, and T. J. Cui, "DeepNIS: Deep neural network for nonlinear electromagnetic inverse scattering," *IEEE Transactions on Antennas and Propagation*, Vol. 67, No. 3, 1819–1825, 2018.

[9] Sun, Y., Z. Xia, and U. S. Kamilov, "Efficient and accurate inversion of multiple scattering with deep learning," *Optics Express*, Vol. 26, No. 11, 14 678–14 688, 2018.

[10] Massa, A., D. Marcantonio, X. Chen, M. Li, and M. Salucci, "DNNs as applied to electromagnetics, antennas, and propagation — A review," *IEEE Antennas and Wireless Propagation Letters*, Vol. 18, No. 11, 2225–2229, 2019.

[11] Chen, X., Z. Wei, M. Li, and P. Rocca, "A review of deep learning approaches for inverse scattering problems (invited review)," *Electromagnetic Waves*, Vol. 167, 67–81, 2020.

[12] Sanghvi, Y., Y. Kalepu, and U. K. Khankhoje, "Embedding deep learning in inverse scattering problems," *IEEE Transactions on Computational Imaging*, Vol. 6, 46–56, 2019.

[13] Chen, G., P. Shah, J. Stang, and M. Moghaddam, "Learning-assisted multimodality dielectric imaging," *IEEE Transactions on Antennas and Propagation*, Vol. 68, No. 3, 2356–2369, 2019.

[14] Guo, R., X. Song, M. Li, F. Yang, S. Xu, and A. Abubakar, "Supervised descent learning technique for 2-D microwave imaging," *IEEE Transactions on Antennas and Propagation*, Vol. 67, No. 5, 3550–3554, 2019.

[15] Wang, Y., Z. Zong, S. He, R. Song, and Z. Wei, "Push the generalization limitation of learning approaches by multi-domain weight-sharing for full-wave inverse scattering," *IEEE Trans. on Geoscience and Remote Sensing*, Vol. 61, 2003814, 2023.

[16] Guo, R., T. Huang, M. Li, H. Zhang, and Y. C. Eldar, "Physics-embedded machine learning for electromagnetic data imaging: Examining three types of data-driven imaging methods," *IEEE Signal Processing Magazine*, Vol. 40, No. 2, 18–31, 2023.

[17] Liu, Y., H. Zhao, R. Song, X. Chen, C. Li, and X. Chen, "SOM-Net: Unrolling the subspace-based optimization for solving full-wave inverse scattering problems," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 60, 1–15, 2022.

[18] Liu, Y., Y. Zhang, L. Ouyang, and T. Shi, "Design of lightweight fundus lesion segmentation algorithm based on improved U-Net," *Electric Measurement Technology*, Vol. 47, No. 3, 127–134, 2024.

[19] Pang, B., L. Chen, Q. Tao, E. Wang, and Y. Yu, "GA-UNet: A lightweight ghost and attention U-Net for medical image segmentation," *Journal of Imaging Informatics in Medicine*, Vol. 37, 1874–1888, 2024.

[20] Deng, Y., Y. Hou, J. Yan, and D. Zeng, "ELU-Net: An efficient and lightweight U-Net for medical image segmentation," *IEEE Access*, Vol. 10, 35 932–35 941, 2022.

[21] Yang, H., D. Zhang, S. Qin, T. Cui, and J. Miao, "Real-time detection of concealed threats with passive millimeter wave and visible images via deep neural networks," *Sensors*, Vol. 21, No. 24, 8456, 2021.

[22] Li, L. and S. Qin, "Real-time detection of hiding contraband in human body during the security check based on lightweight U-Net with deep learning," *Journal of Electronics & Information Technology*, Vol. 44, No. 10, 3435–3446, 2022.

[23] Zhou, Q., L. Ma, L. Cao, and H. Yu, "Identification of tomato leaf diseases based on improved lightweight convolutional neural networks MobileNetV3," *Smart Agriculture*, Vol. 4, No. 1, 47, 2022.

[24] Shi, H., D. Shi, S. Wang, W. Li, H. Wen, and H. Deng, "Crop plant automatic detecting based on in-field images by lightweight DFU-Net model," *Computers and Electronics in Agriculture*, Vol. 217, 108649, 2024.

[25] Hu, L. Y., T. Zhou, W. Xu, Z. M. Wang, and Y. K. Pei, "An improved SqueezeNet lightweight model for tomato disease recognition," *Journal of Zhengzhou University (Natural Science Edition)*, Vol. 54, No. 4, 71–77, 2022.

[26] Wei, Z. and X. Chen, "Deep-learning schemes for full-wave nonlinear inverse scattering problems," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 57, No. 4, 1849–1860, 2018.

[27] He, K., X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778, Las Vegas, NV, USA, 2016.

[28] Jin, K. H., M. T. McCann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," *IEEE Transactions on Image Processing*, Vol. 26, No. 9, 4509–4522, 2017.

[29] Xie, J. and K. Zhang, "XR-MSF-Unet: Automatic segmentation model for COVID-19 lung CT images," *Journal of Frontiers of Computer Science & Technology*, Vol. 16, No. 8, 1850, 2022.

[30] Johnson, J., A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Computer Vision — ECCV 2016: 14th European Conference*, 694–711, Amsterdam, The Netherlands, 2016.