This supplementary material provides additional details and results, including the use of LLMs (Sec. A), physical model of Electromagnetic Inverse Scattering Problems (EISP) (Sec. B), the experimental setup (Sec. C), and extended experimental results (Sec. D).

A. The Use of Large Language Models (LLMs)

We use GPT-4 solely for the purpose of polishing our language of manuscript. This includes improving grammatical accuracy and sentence fluency. LLMs play no significant role in research ideation, methodology design, and experiment execution of this paper.

B. Physical Model of EISP

The research subject of EISP is scatterers. We can describe scatterers with their relative permittivity $\epsilon_r(\mathbf{x})$. The relative permittivity is a physical quantity determined by the material, and it represents the ability to interact with the electromagnetic field of the material. The relative permittivity of vacuum is 1 (and the relative permittivity of air is almost 1), while the relative permittivity of scatterers is over 1. The relative permittivity of metal material is positive infinity, so metal can shield against electromagnetic waves. Thus, scatterers are not composed of metal.

In EISP, the scatterer is placed within the region of interest $\mathcal{D}$. The transmitters are placed around $\mathcal{D}$, emitting incident electromagnetic field E^{i}. The incident field then interacts with the scatterer, exciting the induced current J. The induced current can act as secondary radiation sources, emitting scattered field E^{s}. In fact, for a point in the scatterer, it cannot distinguish the incident field and the scattered field at this point. So, it interacts with the sum of the incident field and the scattered field, aka the total field E^{t}. The total field can be described by Lippmann-Schwinger equation [8] as follows:

$$E^{\mathrm{t}}(\mathbf{x}) = E^{\mathrm{i}}(\mathbf{x}) + k_0^2 \int_{\mathcal{D}} g(\mathbf{x}, \mathbf{x}') J(\mathbf{x}') d\mathbf{x}', \mathbf{x} \in \mathcal{D}, \quad (4)$$

where $\mathbf{x}$ and $\mathbf{x}'$ are the spatial coordinates. k_0 is the wavelength of the electromagnetic wave determined by the frequency. $g(\mathbf{x}, \mathbf{x}')$ is the free space Green's function, which represents the impact of the induced current J at the point $\mathbf{x}'$ to the total field at the point $\mathbf{x}$. $\mathbf{x} \in \mathcal{D}$ indicates that the total field is with the region of interest $\mathcal{D}$. The relationship between the induced current J and the total field E^{t} can be expressed as follows:

$$J(\mathbf{x}) = \xi(\mathbf{x}) E^{\mathrm{t}}(\mathbf{x}), \quad (5)$$

where $\xi(\mathbf{x}) = \epsilon_r(\mathbf{x}) - 1$.

The induced current generates scattered field, and we can measure the scattered field with receivers around the region of interest $\mathcal{D}$. The scattered field can be expressed as follows:

$$E^{\mathrm{s}} = k_0^2 \int_{\mathcal{D}} g(\mathbf{x}, \mathbf{x}') J(\mathbf{x}') d\mathbf{x}', \mathbf{x} \in S, \quad (6)$$

where $\mathbf{x} \in S$ indicates that the scattered field is measured by the receivers at surface S around $\mathcal{D}$.

Since digital analysis only applies to discrete variables [28, 40], we discretize equations Eq. (4), Eq. (5), and Eq. (6). The region of interest $\mathcal{D}$ is discretized into $M \times M$ square subunits, and we use the method of moment [33] to obtain the discrete scattered field $\mathbf{E}^{\mathrm{s}}$[2]. The discrete version of Eq. (4) is as follows:

$$\mathbf{E}^{\mathrm{t}} = \mathbf{E}^{\mathrm{i}} + \mathbf{G}^{\mathrm{d}} \cdot \mathbf{J}, \quad (7)$$

where $\mathbf{G}^{\mathrm{d}}$ is the discrete free space Green's function from points in the region of interest to points in the region of interest, which is a matrix of the shape $M^2 \times M^2$. The discrete version of Eq. (5) is as follows:

$$\mathbf{J} = \mathrm{Diag}(\boldsymbol{\xi}) \cdot \mathbf{E}^{\mathrm{t}}. \quad (8)$$

And the discrete version of Eq. (6) is as follows:

$$\mathbf{E}^{\mathrm{s}} = \mathbf{G}^{\mathrm{s}} \cdot \mathbf{J}, \quad (9)$$

where $\mathbf{G}^{\mathrm{s}}$ is the discrete free space Green's function from points in the region of interest to the locations of receivers, which is a matrix of the shape $N_r \times M^2$.

C. Details of Experimental Setup

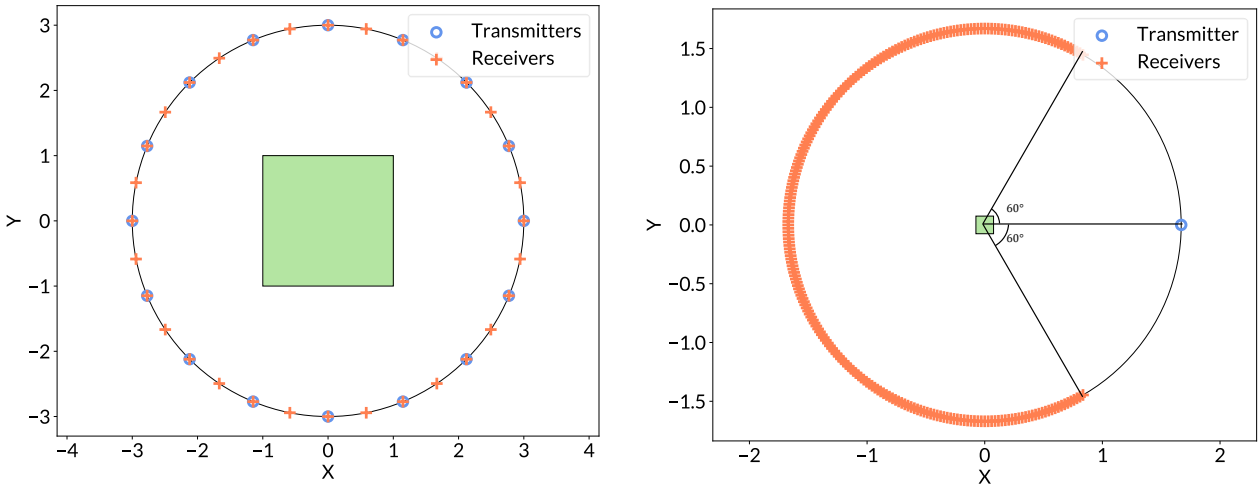
C.1. Datasets

The datasets utilized in this work are described in detail below. To enhance model robustness and training efficiency, datasets that share identical measurement settings are pooled to form consolidated training sets. We strictly follow the established rules in EISP [26, 39, 45] to set the dataset. We train and test our model on standard benchmarks used for EISP.

1) Circular [27] is synthetically generated comprising images of cylinders with random relative radius, number, location, and permittivity between 1 and 1.5. $10k$ images are generated for training purposes, and $1.2k$ images for testing.

2) MNIST [12] contains grayscale images of handwritten digits. Similar to previous settings [45, 57], we use them to synthesize scatterers with relative permittivity values between 2 and 2.5 according to their corresponding pixel values. The entire MNIST training set containing $60k$ images is used for training purposes, while $1.2k$ images from the MNIST test set are randomly selected for testing.

[2] We use the **bold letters** to represent discrete variables.



(a) (b)

Figure 10. **Positions of transmitters and receivers for 2D data.** (a) For Circular dataset and MNIST dataset, we set up $N = 16$ transmitters and $N_r = 32$ receivers. All transmitters and receivers are equally placed. For single-transmitter settings, the transmitter is positioned at the maximum x-coordinate at (3, 0). (b) For IF dataset, we set up $N = 8$ or 18 transmitters and $N_r = 241$ receivers. The transmitters are equally placed (not shown in the figure), and the locations of receivers are determined by the transmitter. For all datasets, the green square represents the region of interest $\mathcal{D}$.

Following previous work [7, 57] to generate the above two synthetic datasets, we set operating frequency $f = 400$ MHz. The region of interest is a square with the size of $2m \times 2m$. We use 16 transmitters and 32 receivers equally placed on a circle with a radius of $3m$ ($N_r = 32$ and $N = 16$). The schematic diagram of the locations of the transmitters and receivers around the region of interest is shown in Fig. 10a. The data are generated numerically using the method of moments [33] with a 224×224 grid mesh to avoid inverse crime [9]. To simulate the noise in actual measurement, we add a 5% level of noise to the scattered field $\mathbf{E}^s$ for regular setting.

3) Real-world IF dataset [15] contains three different dielectric scenarios, namely FoamDielExt, FoamDielInt, and FoamTwinDiel. $N = 8$ for FoamDielExt and FoamDielInt, $N = 18$ for FoamTwinDiel, and $N_r = 241$ for all the cases. The region of interest is a square with the size of $0.15m \times 0.15m$. Transmitters and receivers are placed on a circle with a radius of $1.67m$. The transmitters are placed equally, and the locations of receivers vary for each transmitter. There is no receiver at any position closer than $60°$ from the transmitter, and 241 receivers are placed from $+60°$ to $+300°$ with a step of $1°$ from the location of the transmitter. The schematic diagram of the locations of the transmitters and receivers around the region of interest is shown in Fig. 10b. In the real measurement, there is only one transmitter at a fixed location, and the scatterer rotates to simulate the transmitter to be in different directions. There is a movable receiver sequentially measures the scattered field at 241 different locations. After the measurement, the scatterer rotates by a certain angle for next measurement. The angle is $45°$ for FoamDielExt and FoamDielInt because $N_r = 8$, and $20°$ for FoamTwinDiel because $N_r = 18$. As for operating frequency, all cases are measured under many different frequencies, and we take the frequency $f = 5$ GHz. We evaluate these three scenarios for testing, and use the same settings to synthetically generate $10k$ images of cylinders with random number and location for training purposes.

4) Synthetic 3D MNIST [11] contains 3D images of handwritten digits. 5) Synthetic 3D ShapeNet [46] contains 3D images of various shapes. We use these two datasets to synthesize scatterers with relative permittivity value of 2. $N_r = 160$ and $N = 40$, and the operating frequency $f = 400$ MHz. The region of interest is a cube with the size of $2m \times 2m \times 2m$. The transmitters and receivers are placed at a sphere with the radius of $3m$. For the positions of transmitters, the azimuthal angle ranges from $0°$ to $315°$ with the step of $45°$, and the polar angle ranges from $30°$ to $150°$ with the step of $30°$. For the positions of receivers, the azimuthal angle ranges from $0°$ to $348.75°$ with the step of $11.25°$, and the polar angle ranges from $30°$ to $150°$ with the step of $30°$. The schematic diagram of the locations of the transmitters and receivers around the region of interest is shown in Fig. 11. The entire 3D MNIST dataset is used, including $5k$ images for training purposes and $1k$ images for testing. For 3D ShapeNet, we take $5k$ images from 5 different categories for training purposes and 500 images for testing.

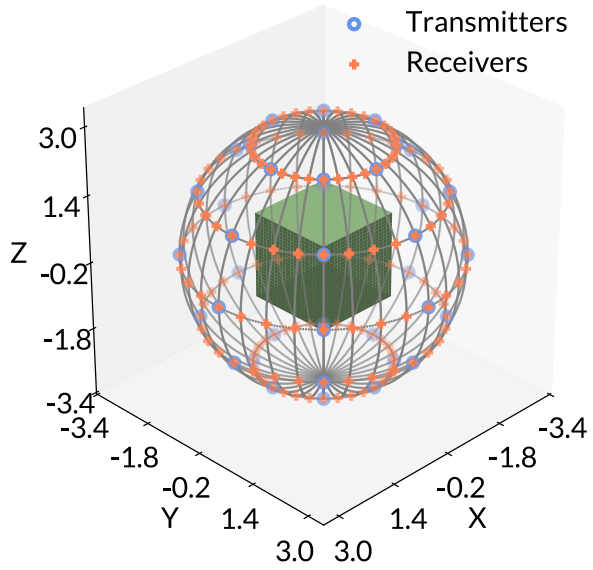


Figure 11. **Positions of transmitters and receivers for 3D MNIST dataset.** There are $N = 40$ transmitters and $N_r = 160$ receivers. The green cube represents the region of interest $\mathcal{D}$.

C.2. Metrics

Following previous work [26], we evaluate the quantitative performance of our method using PSNR [43], SSIM [44], Relative Root-Mean-Square Error (MSE) [39] and IoU. For PSNR, SSIM and IoU, a higher value indicates better performance. For MSE, a lower value indicates better performance. MSE is a metric widely used in EISP, which is defined as follows:

$$\text{MSE} = \left(\frac{1}{M \times M} \sum_{m=1}^{M} \sum_{n=1}^{M} \left|\frac{\hat{\boldsymbol{\varepsilon}}_r(m,n) - \boldsymbol{\varepsilon}_r(m,n)}{\boldsymbol{\varepsilon}_r(m,n)}\right|^2\right)^{\frac{1}{2}}, \tag{10}$$

where $\boldsymbol{\varepsilon}_r(m,n)$ and $\hat{\boldsymbol{\varepsilon}}_r(m,n)$ are the Ground Truth (GT) and predicted discrete relative permittivity of the unknown scatterers at location (m,n), respectively, and $M \times M$ is the total number of subunits over the Region of Interest (ROI) $\mathcal{D}$.

C.3. Implementation Details

We implement our method using PyTorch. Our MLP consists of 8 layers, each with 512 channels. ReLU activation is used between layers to ensure nonlinear expressiveness. Apply positional encoding to $\mathbf{x}$ before inputting it into MLP.

During training, we discretize the region of interest $\mathcal{D}$ into a 64×64 grid and optimize the model using the Adam optimizer with default values $\beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 10^{-8}$, and an initial learning rate of 1×10^{-3}, which remains constant throughout the entire training process. Positional encoding uses frequency $\Omega = 10$, and training runs for 200 iterations on an NVIDIA V100 GPU.

D. Additional Experimental Results

This section provides additional experimental results, including statistical validation (Sec. D.1), extended qualitative comparisons (Sec. D.2) and additional ablation studies on loss functions (Sec. D.3). The qualitative comparisons include more visual comparisons with baseline methods, while the ablations further analyze noise robustness, training data size, and the impact of different training losses.

D.1. Experimental Statistical Significance

Given the inherent stochasticity of the additive noise in our experiments, all quantitative results reported in the main manuscript represent averages over 3 independent trials. The noise samples were independently drawn from a $\mathcal{N}(0,1)$ distribution and subsequently scaled according to both the signal amplitude and the predefined noise ratios (5% and 30%). We report the mean and standard deviation (std) in Table 4, which shows that the experimental variations introduced less than 1% error, confirming the statistical significance of our results.

D.2. Additional Qualitative Results

Qualitative Comparison. We present more qualitative comparison on Circular dataset [27] and MNIST dataset [12] under settings with different transmitter numbers N and noise levels, as shown in Fig. 12 to Fig. 17. Our method achieves comparable or superior performance to SOTA methods in most cases under multiple-transmitter settings, as Fig. 12, Fig. 13, Fig. 16, and Fig. 17 indicate. As shown in Fig. 13 and Fig. 17, PGAN [39] introduces noticeable back-

Table 4. **Statistical results of quantitative metrics.** For Circular and MNIST datasets, we report the mean and standard deviation (std) over 3 independent trials under two noise levels: 5% and 30%.

	MNIST (5%)			MNIST (30%)			Circular (5%)			Circular (30%)		
	MSE ↓	SSIM ↑	PSNR ↑	MSE ↓	SSIM ↑	PSNR ↑	MSE ↓	SSIM ↑	PSNR ↑	MSE ↓	SSIM ↑	PSNR ↑
Number of Transmitters: $N = 16$												
mean	0.039	0.978	32.11	0.050	0.966	29.91	0.020	0.965	36.92	0.024	0.954	35.19
std	2.8×10^{-5}	5.2×10^{-5}	6.2×10^{-3}	2.1×10^{-4}	2.4×10^{-4}	1.7×10^{-2}	1.6×10^{-5}	4.8×10^{-5}	7.6×10^{-3}	7.1×10^{-5}	3.3×10^{-4}	1.4×10^{-2}
Number of Transmitters: $N = 1$												
mean	0.085	0.921	26.09	0.127	0.862	22.56	0.031	0.931	33.18	0.038	0.914	31.38
std	2.1×10^{-4}	3.1×10^{-4}	2.0×10^{-2}	8.1×10^{-5}	7.2×10^{-4}	2.2×10^{-2}	5.8×10^{-5}	1.3×10^{-4}	1.2×10^{-2}	2.6×10^{-4}	5.2×10^{-4}	3.3×10^{-2}

ground artifacts due to the lack of consideration of physics. And Fig. 12 shows that Img-Interiors occasionally fails to converge due to local optima.

For single-transmitter setting ($N = 1$), our method significantly outperforms all previous approaches across all datasets, as shown in Fig. 14 and Fig. 15. Conventional methods such as BP[4], Gs SOM [6], and 2-fold SOM [6] produce only blurry reconstructions. Deep learning-based methods such as BPS [7], Physics-Net [26], and PGAN [39] tend to "hallucinate" the digit, leading to wrong reconstruction on the MNIST dataset. Img-Interiors [27] produces results with structural errors that deviate significantly from the true morphology. In contrast, our method can still produce reasonably accurate reconstructions of the relative permittivity under such a challenging condition. This enables practical applications with fewer transmitters, significantly reducing deployment costs while preserving reconstruction quality.

3D Reconstruction. We present more qualitative comparison on 3D MNIST [11] dataset under the single-transmitter setting, as shown in Fig. 18. Our method successfully approximates permittivity reconstruction even in this difficult setting, whereas Img-Interiors [27] fails. Similar to the 2D scenario, Img-Interiors suffers from degeneration and can only produce fixed patterns, failing to capture the shape of scatterers.

Additionally, we evaluate our method on the more complex 3D ShapeNet [46] dataset to demonstrate its generalization capability. As shown in Fig. 19, our approach can successfully reconstruct various complex structures including airplanes, cars and tables under the same single-transmitter configuration, maintaining accurate permittivity distribution recovery. This demonstrates the strong potential of our method for practical real-world applications.

Noise Robustness. We present qualitative results across multiple noise levels (ranging from 5% to 30%) on the MNIST [12] dataset, providing visual support for the noise robustness analysis. As shown in Fig. 20, our method maintains high visual fidelity at low noise levels. More importantly, the method maintains stable reconstructions and preserves the essential structure of the scatterer across the entire tested noise range (5% to 30%), demonstrating a smooth and gradual degradation in quality that is fully consistent with the quantitative results.

Ablation on Training Data Size. We present qualitative results across different training data scales (ranging from 25% to 100%) on the MNIST [12] dataset, offering visual insights into the impact of data volume on reconstruction performance. As shown in Fig. 21 and Fig. 22 , our model maintains satisfactory reconstruction integrity even with limited training data.

D.3. Additional Ablation Results

We experiment with TV loss [34] for smoothness and Perceptual loss [20] for structure in addition to MSE on single-transmitter setting. Tab. 5 reports the quantitative results on the MNIST dataset under different noise levels. Overall, the auxiliary losses provide mixed improvements, while MSE alone already provides strong and stable performance, indicating that our model is relatively robust to the choice of loss function.

Table 5. **Ablation of loss functions.** We compare MSE with auxiliary TV loss [34] and perceptual loss [20] under the single-transmitter setting. Results are reported on the MNIST dataset under two noise levels: 5% and 30%.

Loss	MNIST (5%)			MNIST (30%)		
	MSE ↓	SSIM ↑	PSNR ↑	MSE ↓	SSIM ↑	PSNR ↑
Ours (MSE)	0.085	0.921	26.09	0.127	0.862	22.56
Ours (w/ TV)	0.084	0.917	26.00	0.124	0.859	22.74
Ours (w/ Percpt.)	0.084	0.923	26.20	0.130	0.861	22.49

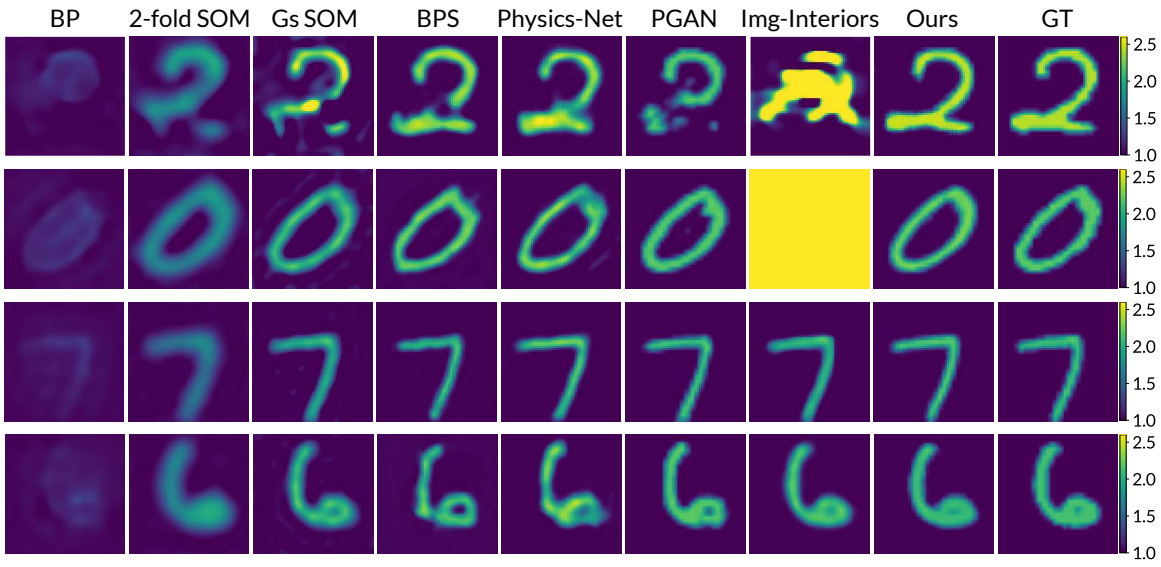


Figure 12. **Qualitative comparison under the multiple-transmitter setting on MNIST dataset.** The results are obtained with $N = 16$ transmitters and a noise level of 5%. Colors represent the values of the relative permittivity.

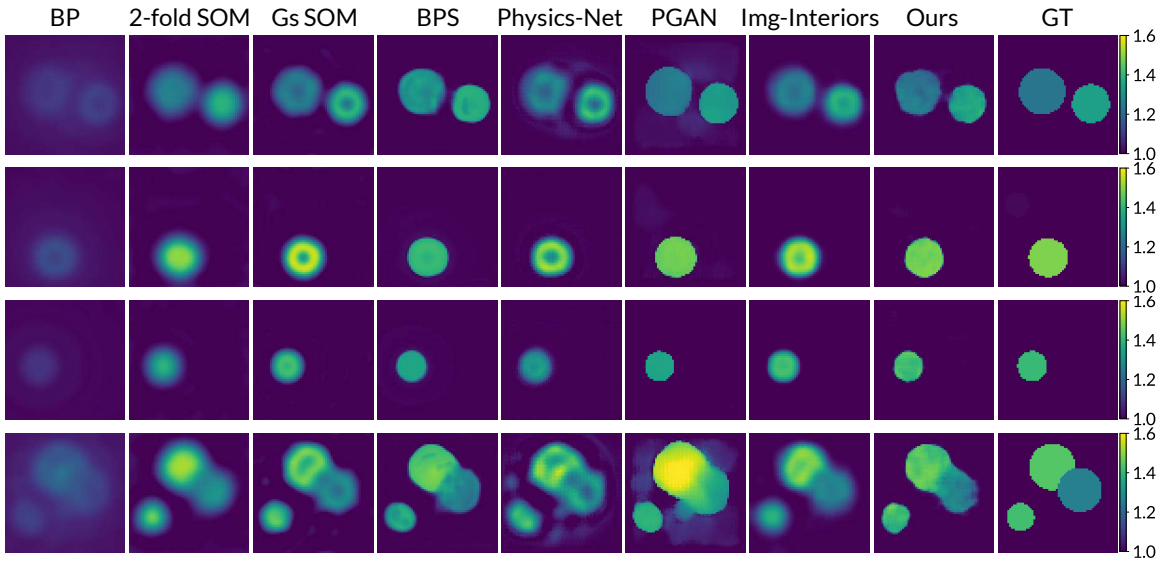


Figure 13. **Qualitative comparison under the multiple-transmitter setting on Circular dataset.** The results are obtained with $N = 16$ transmitters and a noise level of 5%. Colors represent the values of the relative permittivity.

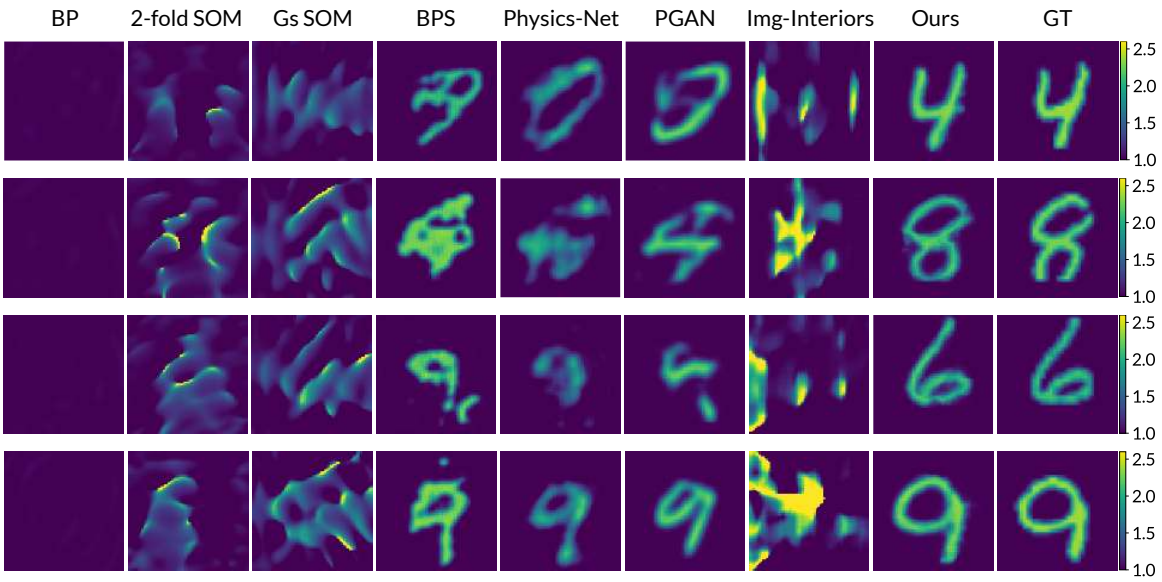


Figure 14. **Qualitative comparison under the single-transmitter setting on MNIST dataset.** The results are obtained with $N = 1$ transmitter and a noise level of 5%. Colors represent the values of the relative permittivity.

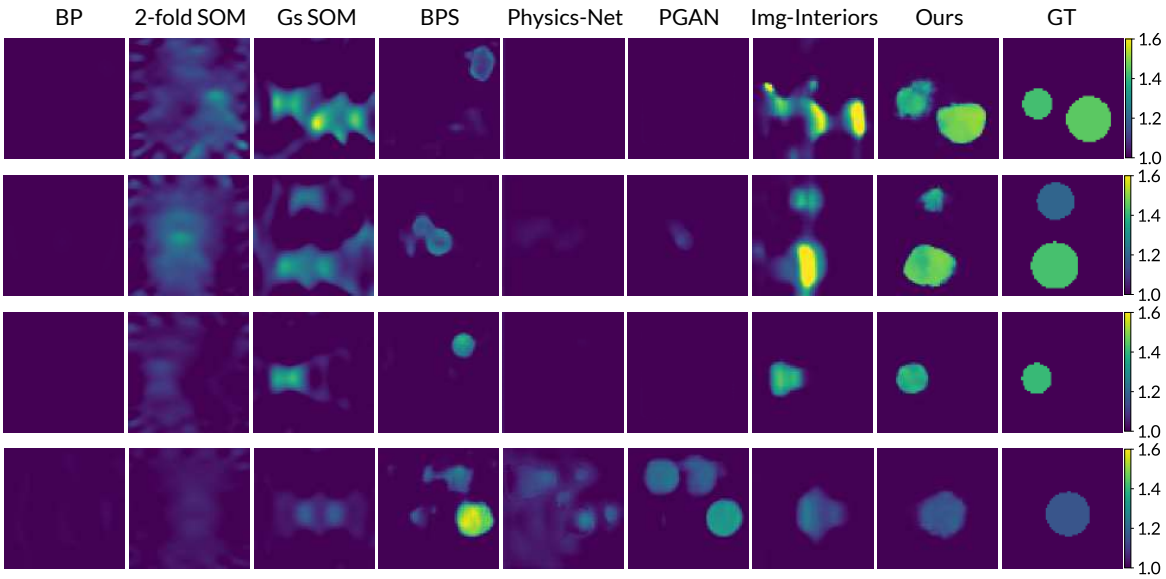


Figure 15. **Qualitative comparison under the single-transmitter setting on Circular dataset.** The results are obtained with $N = 1$ transmitter and a noise level of 5%. Colors represent the values of the relative permittivity.

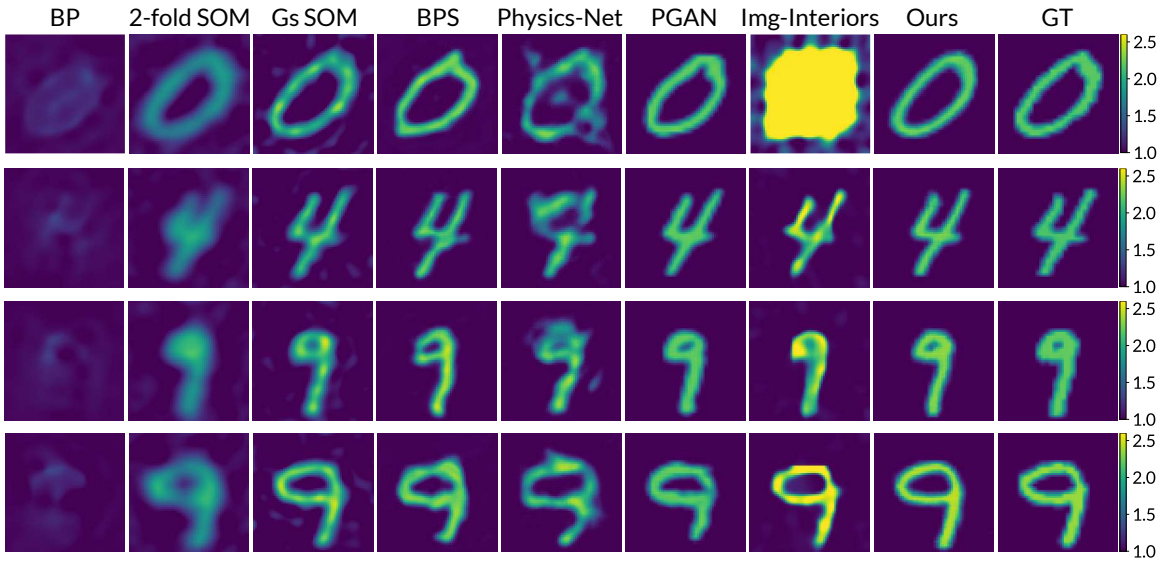


Figure 16. **Qualitative comparison under high noise setting on MNIST dataset.** The results are obtained with $N = 16$ transmitters and a noise level of 30%. Colors represent the values of the relative permittivity.

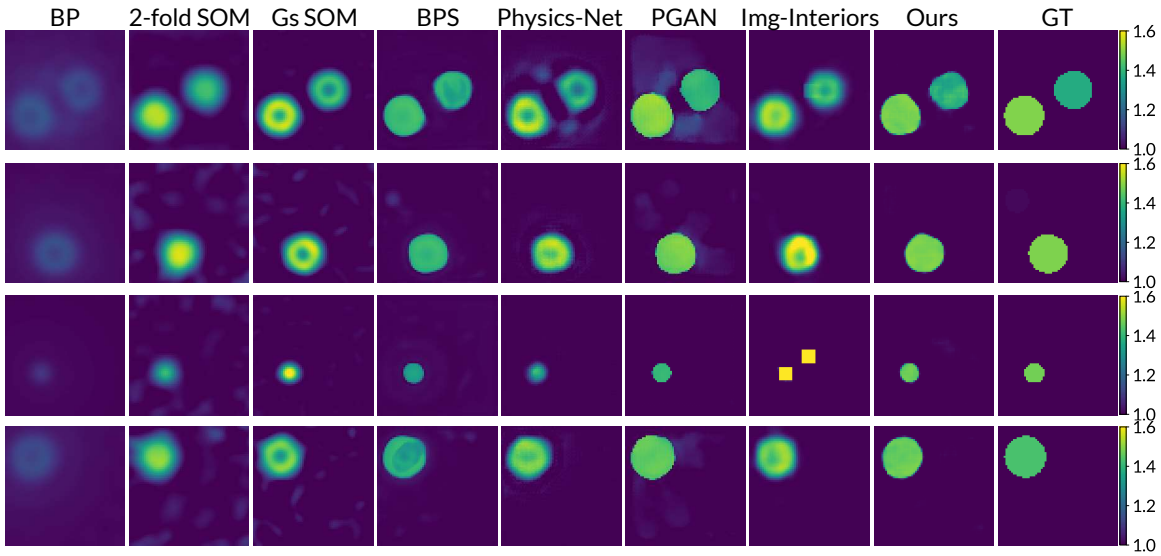


Figure 17. **Qualitative comparison under high noise setting on Circular dataset.** The results are obtained with $N = 16$ transmitters and a noise level of 30%. Colors represent the values of the relative permittivity.

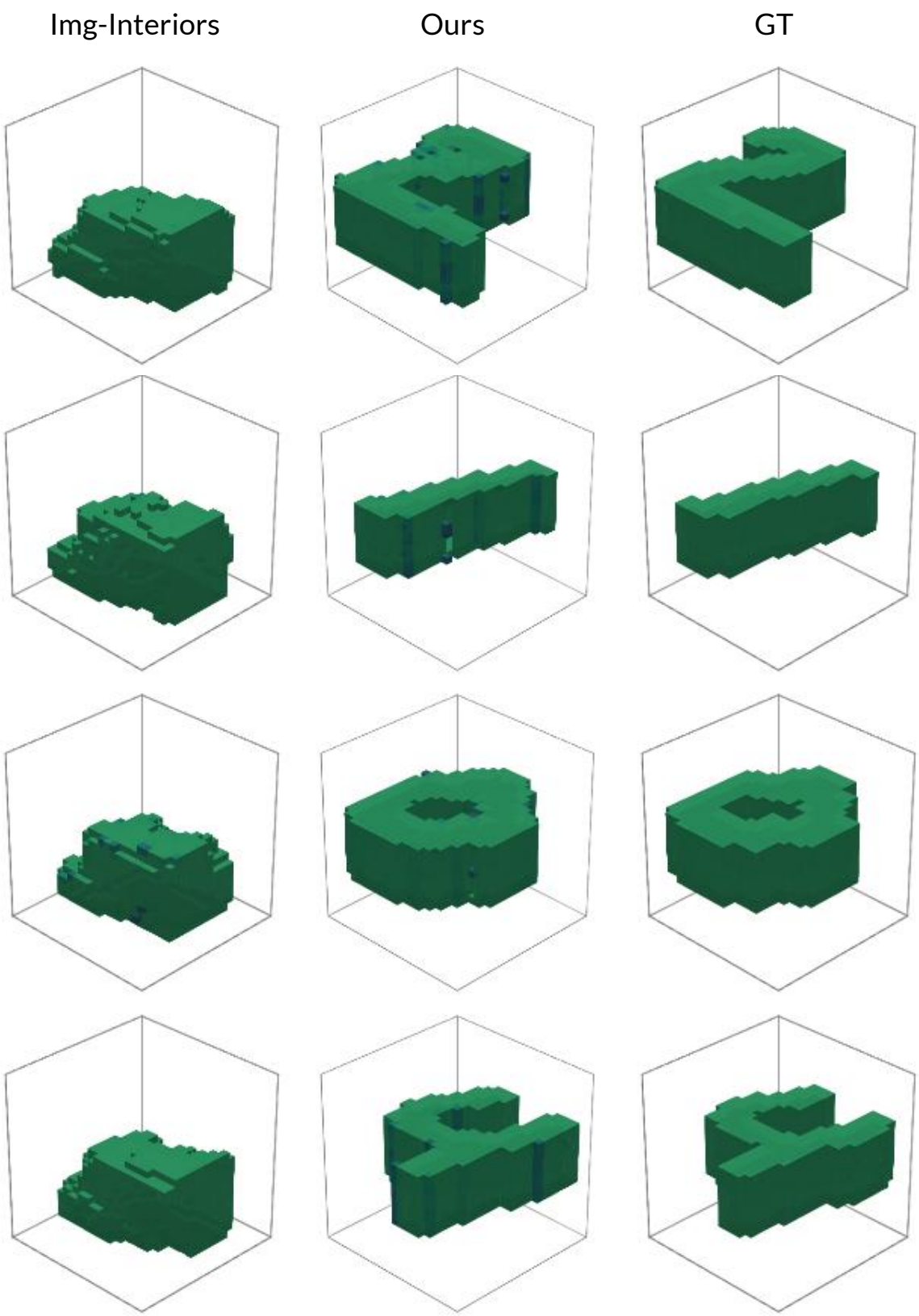


Figure 18. **Qualitative comparison under the single-transmitter setting for 3D reconstruction on 3D MNIST dataset.** The results are obtained with single transmitter ($N = 1$). The voxel colors represent the values of the relative permittivity.

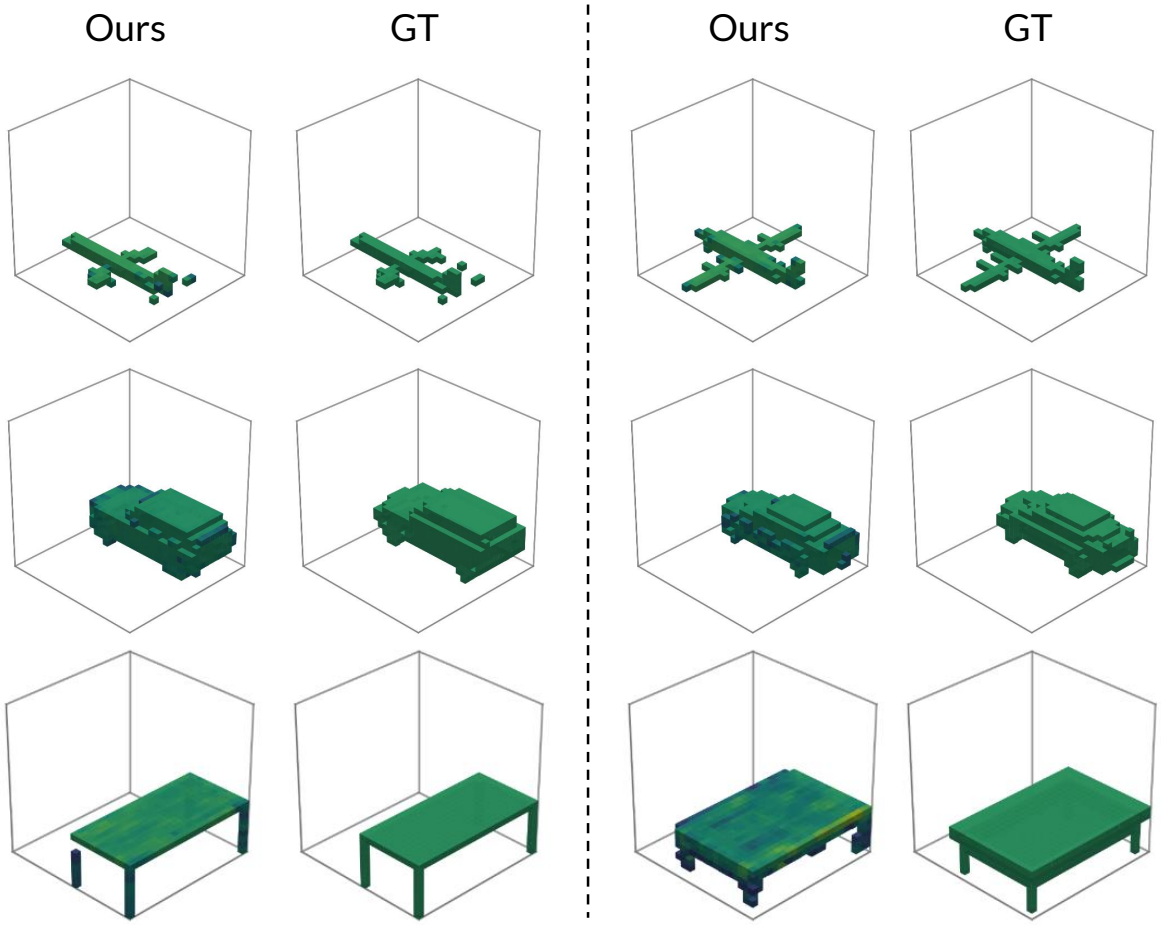


Figure 19. **Qualitative comparison under the single-transmitter setting for 3D reconstruction on 3D ShapeNet dataset.** The results are obtained with single transmitter ($N = 1$). The voxel colors represent the values of the relative permittivity.

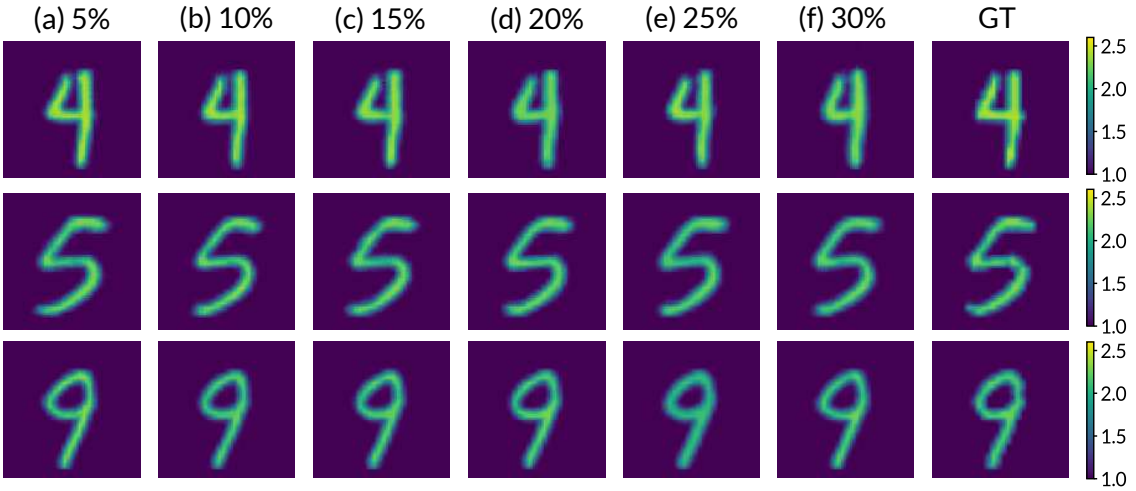


Figure 20. **Qualitative results of our noise robustness on the MNIST dataset.** The results are obtained with $N = 16$ transmitter. Colors represent the values of the relative permittivity.

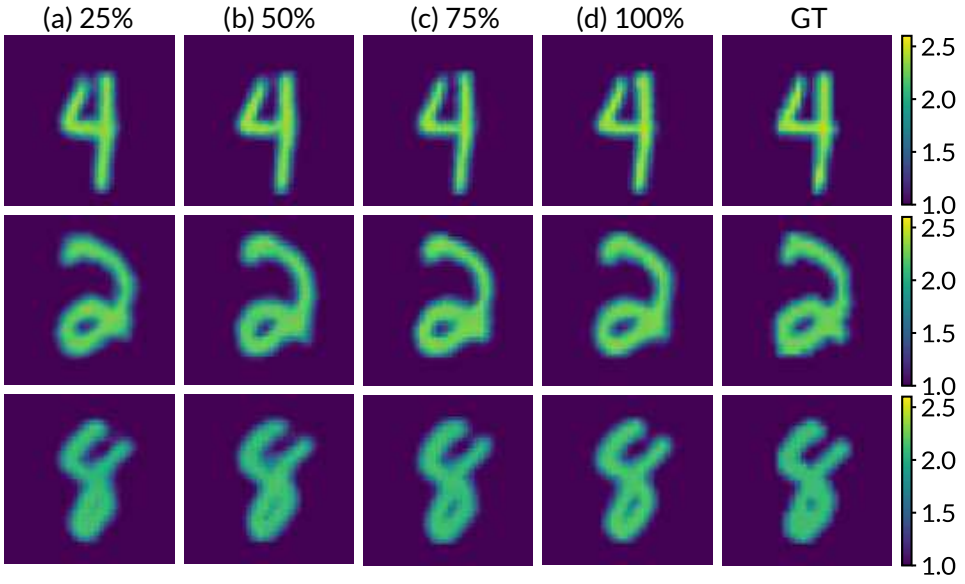


Figure 21. **Qualitative results of training data size ablation on the MNIST dataset.** The results are obtained with $N = 16$ transmitters and a noise level of 5%.

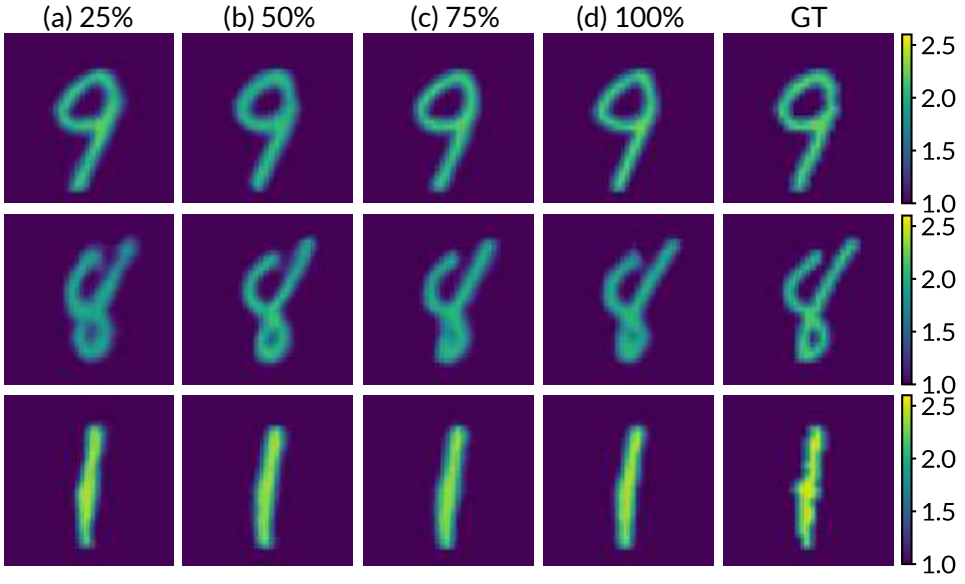


Figure 22. **Qualitative results of training data size ablation on the MNIST dataset.** The results are obtained with $N = 16$ transmitters and a noise level of 30%.