Pseudo-Differential Integral Autoencoder Networks for Inverse PDE Operators

Chunmei Wang
Department of Mathematics
University of Florida
Supported by NSF Grant DMS-2206332
Collaborated with: Ke Chen (UMD College Park), Jasen Lai
(University of Florida starting from Fall 2024)

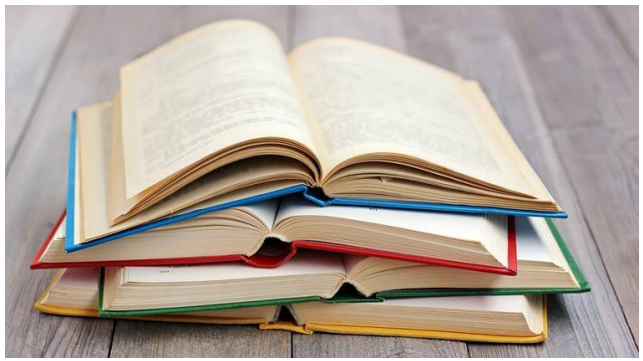
Scientific Machine Learning: Theory and Algorithms
Brin Mathematics Research Center
Department of Mathematics, University of Maryland
February 21-23, 2024

Outline of Talk

1. Inverse Problems
2. Pseudo-Differential Integral Autoencoder (pd-IAE) Algorithm
3. Numerical Results
    - Inverse scattering
    - Optical tomography
    - Electrical impedance tomography
    - Discretization Invariance
    - Robustness to noise

Operator Learning

New diagram for solutions and new opportunities for mathematics



Conventional solvers

- Years of design to solve
- Months of coding
- Accurate but may be slow

Data-driven methods

- Learning to solve from data
- Days or months of training
- Fair and fast solution

Operator Learning

Problem Statement:
Given: Data samples $\mathcal{S} = \{u_i, v_i\}_{i=1}^n$ with $v_i = \Psi(u_i)$
Goal: Learn an operator $\Psi : \mathcal{X} \to \mathcal{Y}$ from data samples $\mathcal{S}$

Broad applications:

- Solving parametric PDEs: parameter functions to PDE solutions
- Solving inverse problems: data to images
- Image processing: image to image
- Predictive data science: historical states to future states
- Learning to optimize: objective functions to optimal updating rules

Inverse Problems

Mathematical Model: Find u such that

$$\begin{cases} \mathcal{L}_a u = h, & \text{in } \Omega \subset \mathbb{R}^d \\ \mathcal{B} u = f, & \text{on } \partial\Omega \end{cases}$$

- a: unknown target function
- $\mathcal{L}_a$: differential operator defined in Ω
- h: known source function
- $\mathcal{B}$: operator defined on $\partial\Omega$
- f: boundary condition

Our Objective: Reconstruct the unknown function a based on

$$\mathcal{S}_a^n = \{(f_i, g_i), i = 1, \ldots, n\}$$

- different f_i lead to boundary measurement g_i
- f_i are predetermined and depend on the experimental design
- g_i depend on the setup of receivers

Traditional Computational Approaches

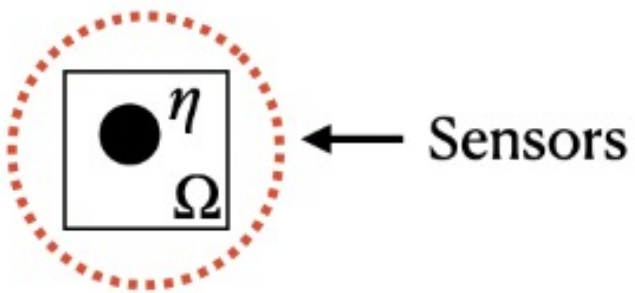
Iterative Optimization

- Require good initial guess
- Require good regularization $R(\eta)$
- Convergence issue
- Solve expensive linear systems

$$\min_{\eta} L(\eta, d) + R(\eta)$$

Direct Methods

- Require too many sensors
- Solve expensive linear systems



Operator Learning

Operator Learning: learn an operator $\psi : S_a^n \mapsto a$

- utilize a neural network (NN) ϕ_θ to parameterize ψ
- ϕ_θ is trained on data samples $\mathcal{S}_a^n$ to optimize θ
- once trained with an optimal parameter θ^*, ϕ_{θ^*} is used to estimate any target function a from the corresponding measurement data

Operator Learning

The measurement data can be represented by

$$S_a^n = [S_{ij}] \in \mathbb{R}^{n \times n}$$

- n: number of grid points on the boundary
- S_{ij}: measurement data collected at receiver j for the solution generated by the i-th source on the boundary

Target medium function can be expressed as $a = [a(x_i)] \in \mathbb{R}^M$

- x_i: the grid point
- M: number of grid points in the domain

Objective: learn a NN mapping from $\mathbb{R}^{n \times n}$ to $\mathbb{R}^M$
Challenges: Repeated and expensive training for different data structures due to varying number of grid points n on the boundary and M in the domain.
Solution: Discretization invariant operator learning

Integral Autoencoder networks

Question: How to achieve discretization invariance in operator learning?
Answer: Integral Autoencoder networks (IAEnet)
Question: What is IAEnet?
Answer: IAEnet has the computational flow:

$$\bar{f} \xrightarrow{F} a_0 \xrightarrow{\mathcal{I}_1} a_1 \xrightarrow{\mathcal{I}_2} \cdots \xrightarrow{\mathcal{I}_L} a_L \xrightarrow{G} \bar{g}$$

- F: a pre-processing NN function
- G: a post-processing NN function
- $\mathcal{I}_1, \ldots, \mathcal{I}_L$: discretization-invariant IAE blocks

IAE block

Question: What is a discretization invariant IAE block?

$$\mathcal{I}: \quad a \xrightarrow{\text{encoder}} v \xrightarrow{\text{FNN}} u \xrightarrow{\text{decoder}} b$$

- $a \in \Omega_a$ and $b \in \Omega_b$ are discretized on the same grid points $S = \{x_i\}_{i=1}^{s}$
  - s may vary for different training and testing data pairs
- v and u defined on Ω_z are discretized on grid points $S_z = \{z_j\}_{j=1}^{m}$
  - m is a fixed number

Answer: In a discretization invariant IAE block, FNN within an IAE block has fixed input and output dimensions of m.

Encoder Function and Decoder Function

Question: How to design discretization invariant IAE block?

$$\mathcal{I}: \quad a \xrightarrow{\text{encoder}} v \xrightarrow{\text{FNN}} u \xrightarrow{\text{decoder}} b$$

Answer: Nonlinear integral transform to design encoder and decoder.

Encoder:

$$v(z) = \int_{\Omega_a} \phi_1\left(a(x), x, z; \theta_{\phi_1}\right) a(x) dx\,, \quad z \in \Omega_z$$

Decoder:

$$b(x) = \int_{\Omega_z} \phi_2\left(u(z), x, z; \theta_{\phi_2}\right) u(z) dz\,, \quad x \in \Omega_b$$

Challenges: (1) computationally expensive encoders and decoders; (2) learning the kernels ϕ_1 and ϕ_2 is challenging.

Pseudo-differential Integral Autoencoder

Solution: Pseudo-differential Integral Autoencoder (pd-IAE)

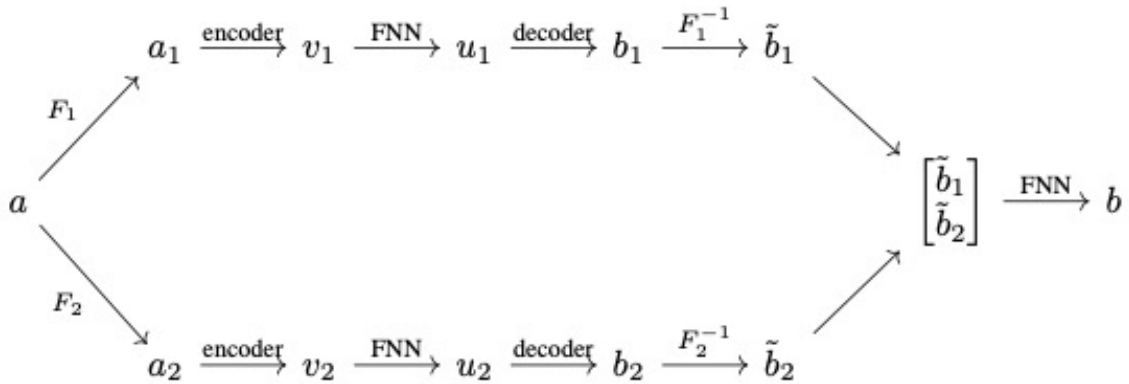
$$Pu(x) = \mathcal{F}^{-1}(\mathcal{A}(x,\cdot)\hat{u}(\cdot)) = \frac{1}{(2\pi)^d}\int e^{ix\cdot\xi}\mathcal{A}(x,\xi)\hat{u}(\xi)d\xi$$

- $\hat{u}$: Fourier transform of u
- $\mathcal{A}(x,\xi)$: a smooth function known as the symbol of P
- Example: if $\mathcal{A}(x,\xi)$ is a polynomial $p(\xi)$, the operator P corresponds to the classical differential operator $p(-i\partial_x)$.

Multi-channel pd-IAE block

Question: What is multi-channel pd-IAE block?

Answer: A multi-channel pd-IAE block has the computational flow



Question: Why multi-channel pd-IAE block?

Answer:

- additional Fourier (or Wavelet) channel proves beneficial for data featuring oscillatory or sparse characteristics
- input function $a(\xi)$ is transformed into the frequency domain

Multi-channel pd-IAE blocks

Simplified encoder function

$$v(z) = \mathcal{F}^{-1}\left(\mathcal{A}(z,\cdot)a(\cdot)\right) = \int_{\Omega_a} e^{iz\cdot\xi} A(z,\xi)a(\xi)d\xi$$

- involves only one inverse Fourier transform
- does not depend on input grid points

Use a low-rank factorization to approximate $A(z,\xi)$

$$A(z,\xi) \approx \sum_{k=1}^{K} p_k(\xi)q_k(z)\,.$$

Simplified encoder function

$$v(z) \approx \sum_{k=1}^{K} q_k(z;\theta_{q_k}) \int_{\Omega_a} e^{iz\cdot\xi} a(\xi)p_k(\xi;\theta_{p_k})d\xi.$$

Pseudo-Differential Decoder Function

Question: How to ensure that the intermediate FNN within a pd-IAE block has a fixed input size?

$$\mathcal{I}: \quad a \xrightarrow{\text{encoder}} v \xrightarrow{\text{FNN}} u \xrightarrow{\text{decoder}} b$$

Answer:

- $a(x) \in \mathbb{R}^s$ consists of function values at arbitrary grid points $S = \{x_i\}_{i=1}^s$.
- $a(x) \in \mathbb{R}^s$ is mapped to a vector $a(\xi)$ with a fixed size due to high frequency truncation.
- **Crucial for achieving discretization invariance:** Since $a(\xi)$ is in the frequency domain, we can truncate high-frequency modes in $a(\xi)$.

Pd decoder function

$$b(x) = \sum_{k=1}^{K} \tilde{p}_k(x; \theta_{\tilde{p}_k}) \mathcal{F}\left(\tilde{q}_k(\cdot; \theta_{\tilde{q}_k}) u(\cdot)\right)$$

Our Contributions

- The low-rank factorization in the integral kernel significantly simplifies the encoder and decoder functions.
- The computation complexity of encoders and decoders becomes almost linear to enhance the efficiency of pd-IAEnet.

Numerical Results

Model	Parameters	Disk space
pd-IAEnet	5,618,785	64.8 MB
IAEnet[1]	6,040,657	69.4 MB
FNO [2]	9,462,849	108 MB
DeepONet [3]	81,597,697	933 MB
ResNet [4]	5,534,593	63.5 MB

Table: This table presents the number of trainable parameters, and the disk space usage for each method.

[1] Y Ong, Z Shen, and H Yang, Integral autoencoder network for discretization-invariant learning, Journal of Machine Learning Research, 2022

[2] Z Li, N Kovachki, K Azizzadenesheli, B Liu, K Bhattacharya, A Stuart, and A Anandkumar, Fourier neural operator for parametric partial differential equations, ICML, 2021

[3] L Lu, P Jin, and G Karniadakis, Deeponet: Learning nonlinear operators for identifying differential equations based on the universal approximation theorem of operators, Nature Machine Intelligence, 2021

[4] K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition, Proceedings of the IEEE conference on computer vision and pattern recognition, 2016.

Inverse scattering

Inverse scattering problem is described by Helmholtz equation

$$\left(-\Delta - \frac{w^2}{c(x)^2}\right) u = 0\,, \quad x \in \mathbb{R}^d\,,$$

- $w \in \mathbb{R}_+$: frequency
- $c(x) > 0$: unknown inhomogeneous wave speed

Scatter function

$$\eta(x) = \frac{w^2}{c(x)^2} - \frac{w^2}{c_0(x)^2}\,, \quad x \in \Omega\,,$$

- background wave speed $c_0(x)$ is known and identical to $c(x)$ except on a bounded domain Ω within an unit ball in 3D or an unit circle in 2D

Objective: reconstruct $\eta(x)$ from measurements collected at multiple sources and receiver locations.

Point scatter distribution

Point scatter media consists of isolated points with small supports.

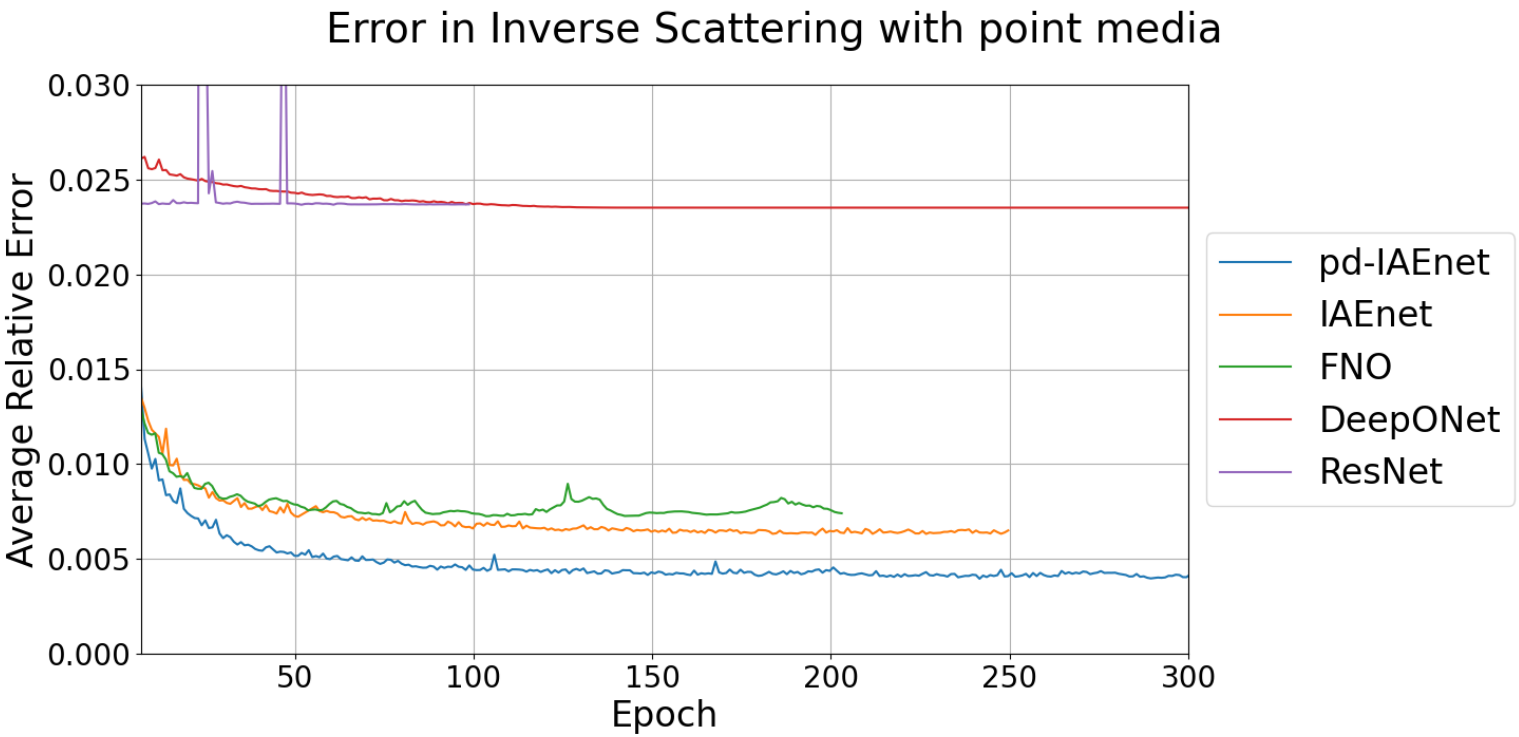
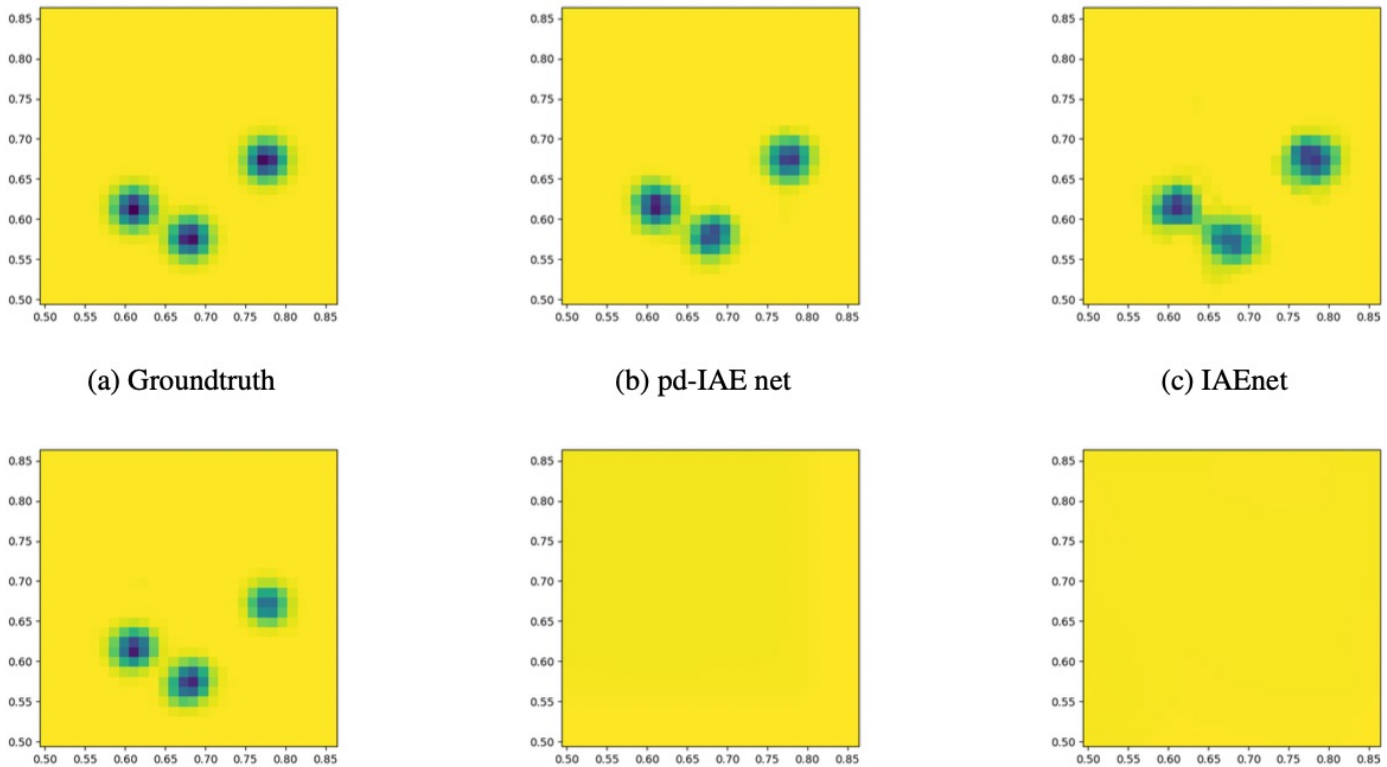


Figure: The plot illustrates the average relative errors versus the number of epochs for the inverse scattering problem with point scatter media. The average error is calculated from the following discretizations: 27×27, 41×41, 81×81, 161×161, and 241×241. Pd-IAE net consistently achieves the lowest average relative error among all the methods. DeepONet and ResNet struggle to learn the inverse operator.

Point scatter distribution



(a) Groundtruth (b) pd-IAE net (c) IAEnet

(d) FNO (e) DeepONet (f) ResNet

Figure: Reconstructed point scatterer at a discretization of 81×81 from all methods based on randomly selected testing data. pd-IAE net, IAEnet, and FNO accurately locate the positions of the point scatterers. pd-IAEnet accurately reconstructs the points with uniform magnitudes for the three depicted points. IAEnet reconstructs the bottom two points with an overlap not observed in the ground truth. FNO's upper-right point has a noticeably smaller magnitude compared to the other two points. Both DeepONet and ResNet struggle to reconstruct a meaningful scatterer, ultimately producing a near-zero solution.

Shepp-Logan scatter distribution

This media consists of indicator functions supported on ellipses with varying axis lengths, positions, and rotation angles.

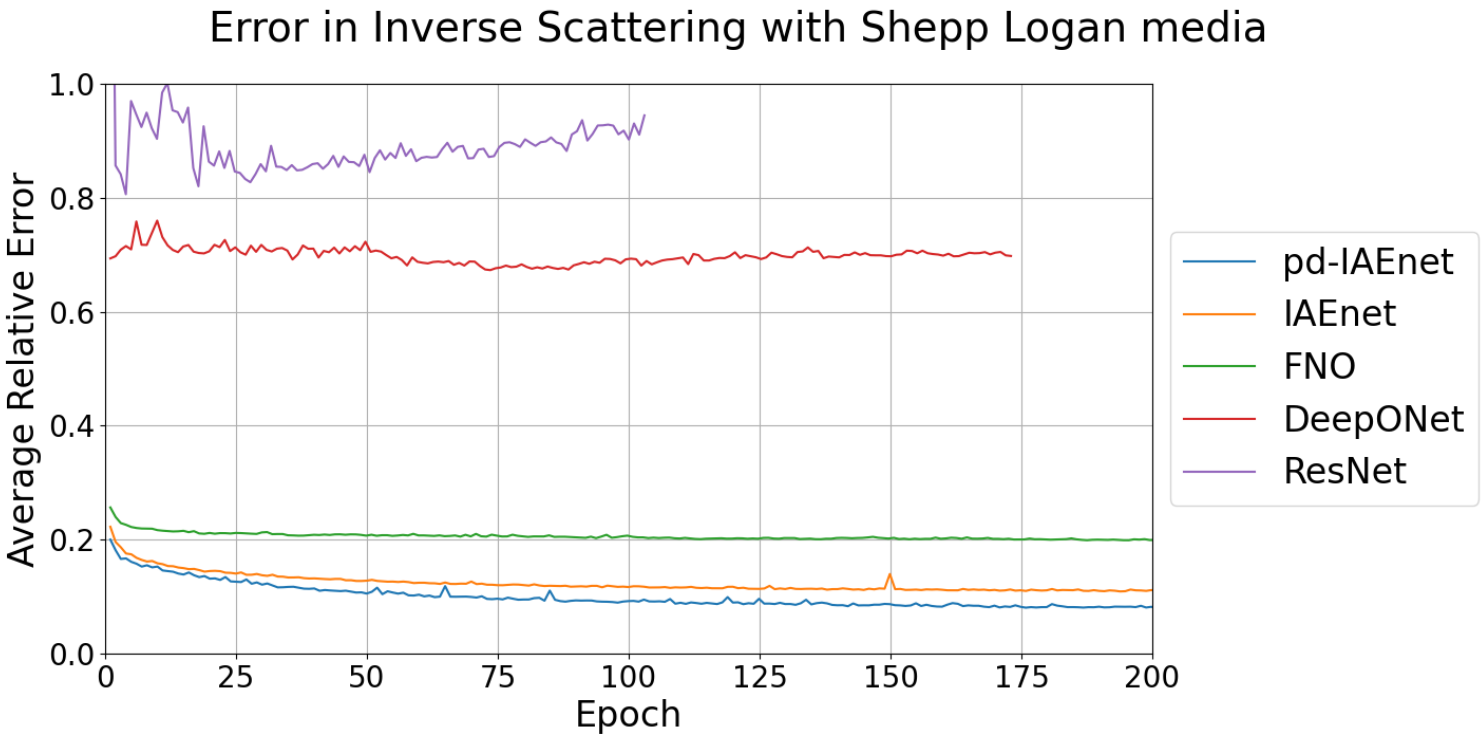
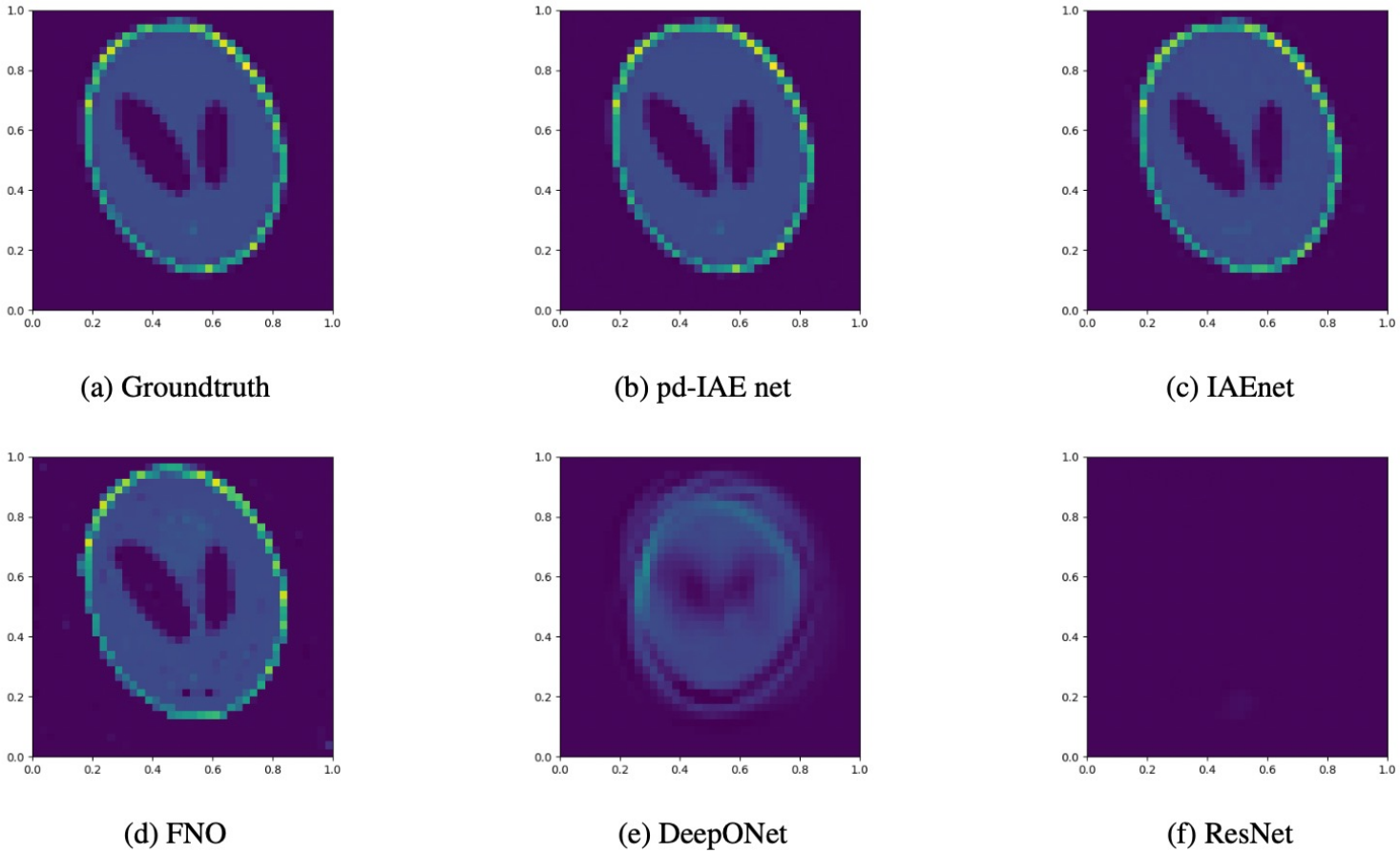


Figure: The graph illustrates the average relative errors over epochs for the inverse scattering problem with Shepp-Logan media. The average relative error is computed from errors obtained at different discretizations: 40×40, 60×60, 80×80, 100×100, and 240×240. Pd-IAEnet achieves the lowest average relative error. DeepONet and ResNet struggle at a large average relative error.

Shepp-Logan scatter distribution



(a) Groundtruth (b) pd-IAE net (c) IAEnet

(d) FNO (e) DeepONet (f) ResNet

Figure: Reconstructed media at 40×40 discretization level for the inverse scattering problem with Shepp-Logan media. Pd-IAE net provides the most accurate reconstruction of the ground truth. IAEnet introduces a minor inaccuracy below the two internal ellipses and yields a slightly different magnitude along the bottom boundary. FNO exhibits several artifacts, with the most noticeable ones just above the bottom boundary. DeepONet produces a blurry reconstruction of the boundary and the two internal ellipses. ResNet struggles and approximates a near-zero solution.

Optical tomography (OT)

OT plays a crucial role in reconstructing optical medium properties based on measurements of light transmitted and scattered through the medium with applications in biomedical imaging for tissues like brain and breast.

Mathematical Model (Radiative Transfer Equation (RTE)):

$$v \cdot \nabla \rho(x,v) = \sigma_s(x) \left(\int_{\mathbb{S}^{d-1}} u(x,v')dv' - u(x,v) \right), (x,v) \in \Omega \times \mathbb{S}^{d-1},$$

$$\rho(x,v) = f(x,v), (x,v) \in \Gamma_-,$$

- $\sigma_s(x) > 0$: scattering coefficient
- incoming boundary $\Gamma_- = \{(x,v) \in \partial\Omega \times \mathbb{S}^{d-1} \mid n_x \cdot v < 0\}$
- incoming boundary condition $f(x,v)$ represents sources of photons injected into the domain
- outgoing boundary $\Gamma_+ = \{(x,v) \in \partial\Omega \times \mathbb{S}^{d-1} \mid n_x \cdot v > 0\}$
- $g = \rho(x,v)|_{\Gamma_+}$: outgoing photon intensity

Objective: reconstruct the scattering coefficient σ_s from multiple pairs of incoming and outgoing photon intensities.

Optical tomography

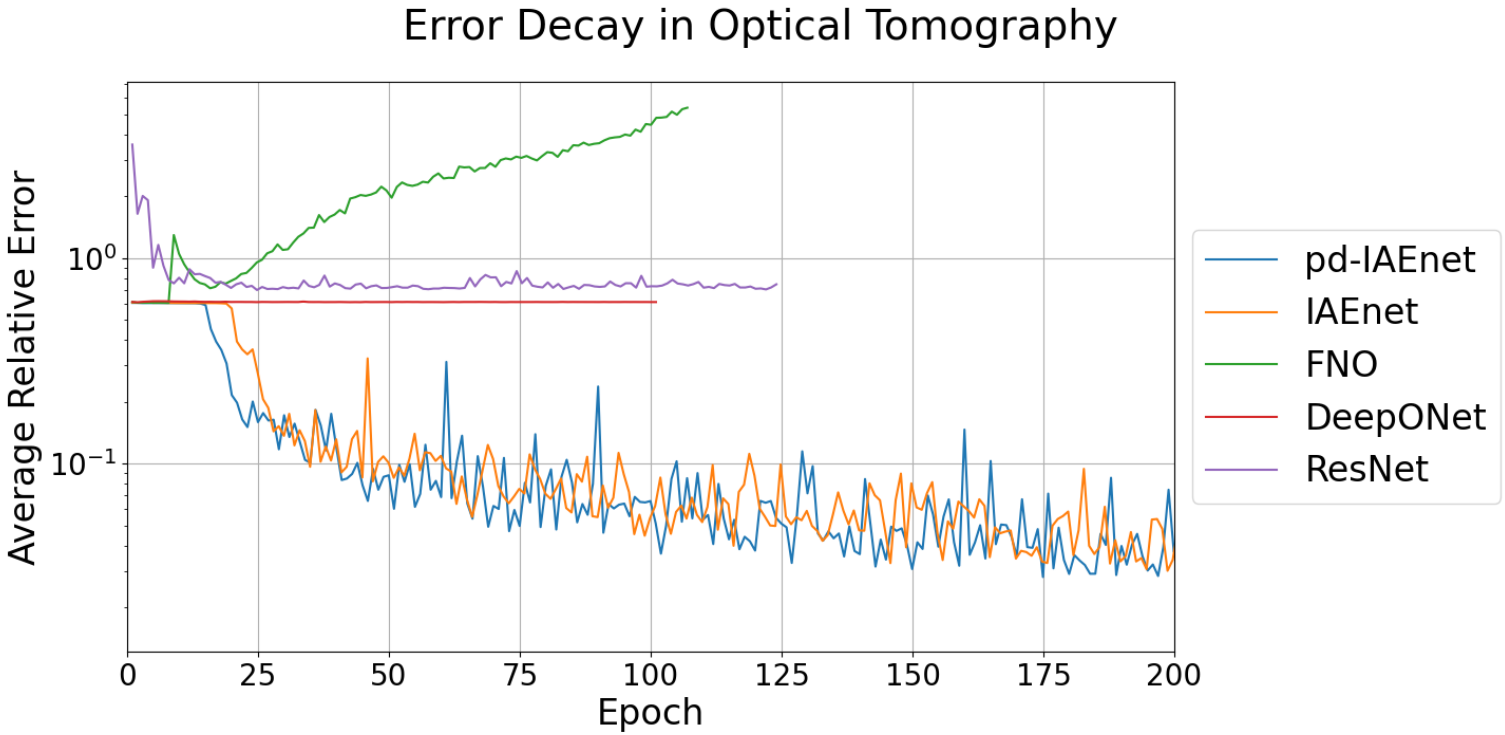
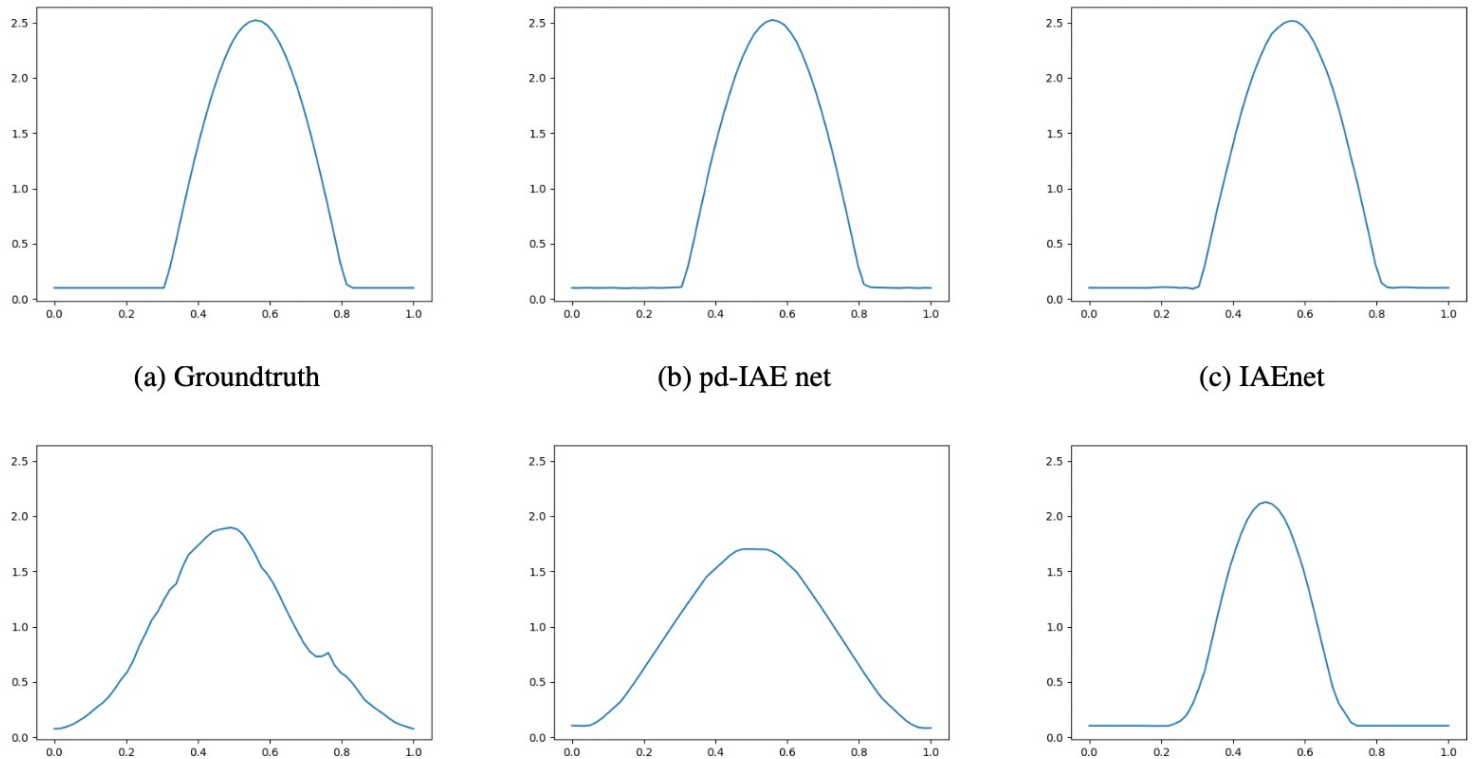


Figure: The plot illustrates the average relative errors as a function of training epochs for OT. This average error is computed from the errors at different discretizations: 30×30, 40×40, 50×50, 60×60, and 70×70. Both pd-IAEnet and IAEnet consistently achieve low errors, each below 0.1. DeepONet and ResNet stabilize at a high error level. FNO exhibits a divergent average relative error.

Optical tomography



(a) Groundtruth (b) pd-IAE net (c) IAEnet

(d) FNO (e) DeepONet (f) ResNet

Figure: Reconstruction of the sinusoidal scattering coefficient in OT with a 60×60 discretization grid. Both pd-IAE net and IAEnet yield reconstructions that closely align with the ground truth. IAEnet's reconstruction exhibits minor perturbations, particularly on the left side of the wave's base. Both FNO and DeepONet encounter difficulties in accurately capturing the location and wavelength of the sine wave. ResNet excels in identifying the location of the wave but faces challenges in accurately determining its magnitude.

Electrical impedance tomography (EIT)

EIT (Calderón problem) plays a significant role in noninvasive medical imaging, particularly in the early diagnosis of breast cancer.
Mathematical Model:

$$\begin{cases} -\text{div}\left(e^{a(x)}\nabla u(x)\right) = 0\,, & x \in \Omega\,, \\ u(x) = f(x)\,, & x \in \partial\Omega\,, \end{cases}$$

- $a(x)$: unknown medium conductivity
- f: boundary voltage

Objective: to reconstruct the function $a(x)$, given the Dirichlet-to-Neumann data pairs $S_a^n = \{(f_i, g_i) \mid i = 1, \ldots, n\}$ where current measurements are the Neumann derivatives of the solution $g = e^a \frac{\partial u}{\partial n}|_{\partial\Omega}$ on the boundary.

Electrical impedance tomography (EIT)

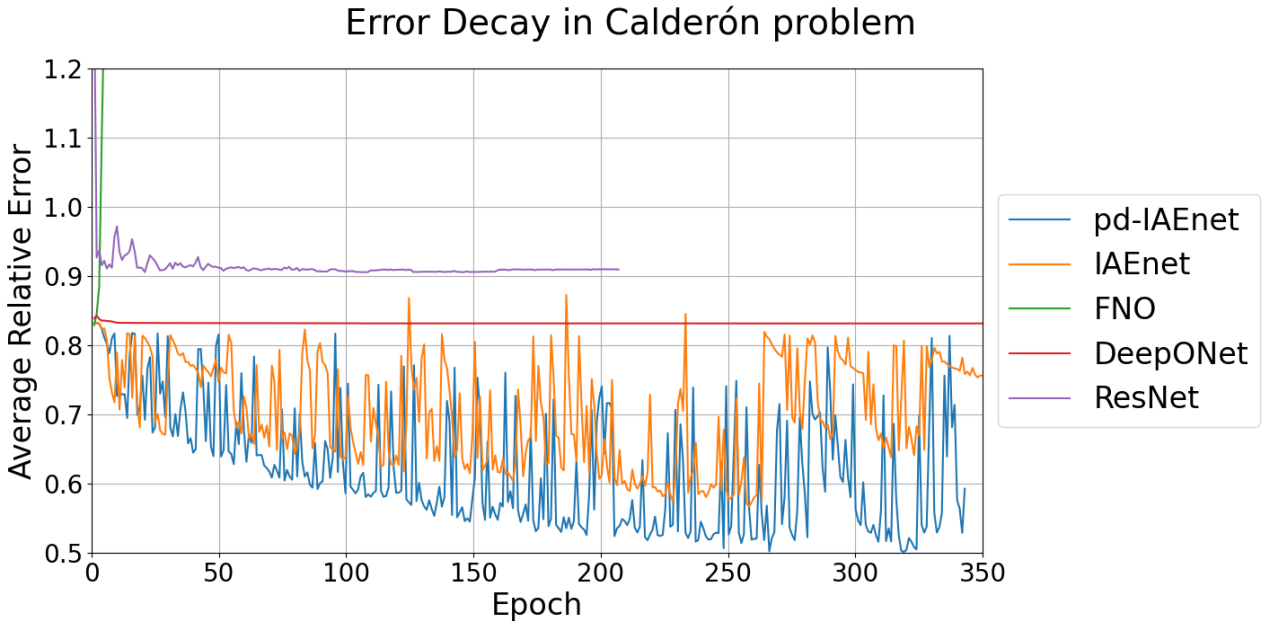
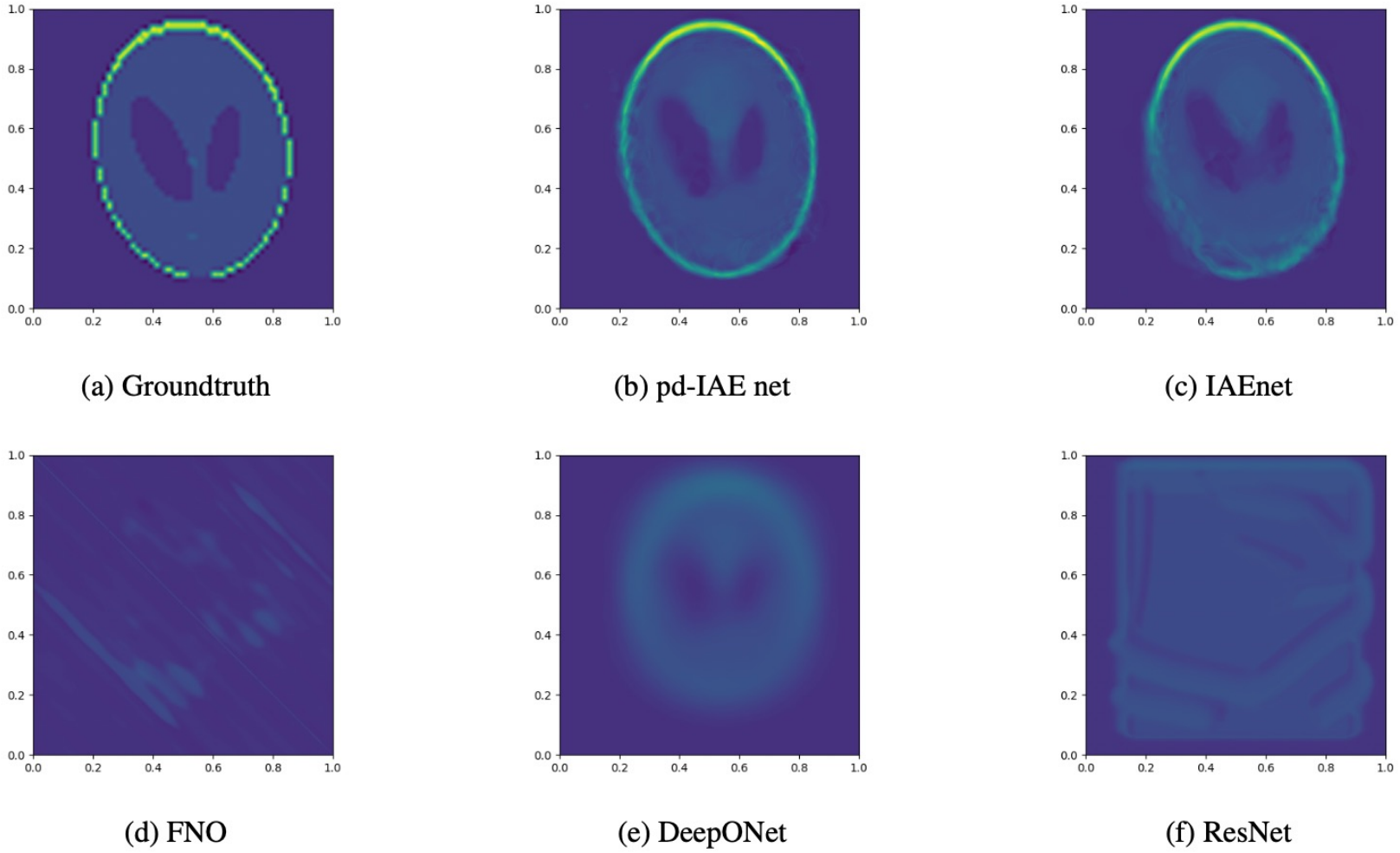


Figure: The plot presents the average relative errors plotted against the number of epochs for the Calderón problem. The relative errors are averaged across different discretizations: 42×42, 63×63, 84×84, 126×126, and 252×252. FNO exhibits a diverging trend in its average relative error, indicating instability during training. DeepONet and ResNet converge, but with persistently high relative errors. Both pd-IAE net and IAE-net consistently achieve significantly lower errors, highlighting their robust training. IAE-net reaches its smallest error of 0.56 during the training process, while pd-IAE net outperforms with the smallest error of 0.50.



(a) Groundtruth (b) pd-IAE net (c) IAEnet

(d) FNO (e) DeepONet (f) ResNet

Figure: Reconstructed images generated by all models based on randomly selected measurement data with a discretization of 252×252 for Calderón problem. Both pd-IAEnet and IAEnet consistently deliver superior reconstructions. pd-IAEnet excels in distinguishing the internal ellipses and provides a more accurate reconstruction of the boundary compared to IAEnet. DeepONet is capable of producing a somewhat blurry image of the Shepp Logan phantom. FNO and ResNet clearly indicate their inability to accurately reconstruct the Shepp Logan phantom.

Discretization Invariance

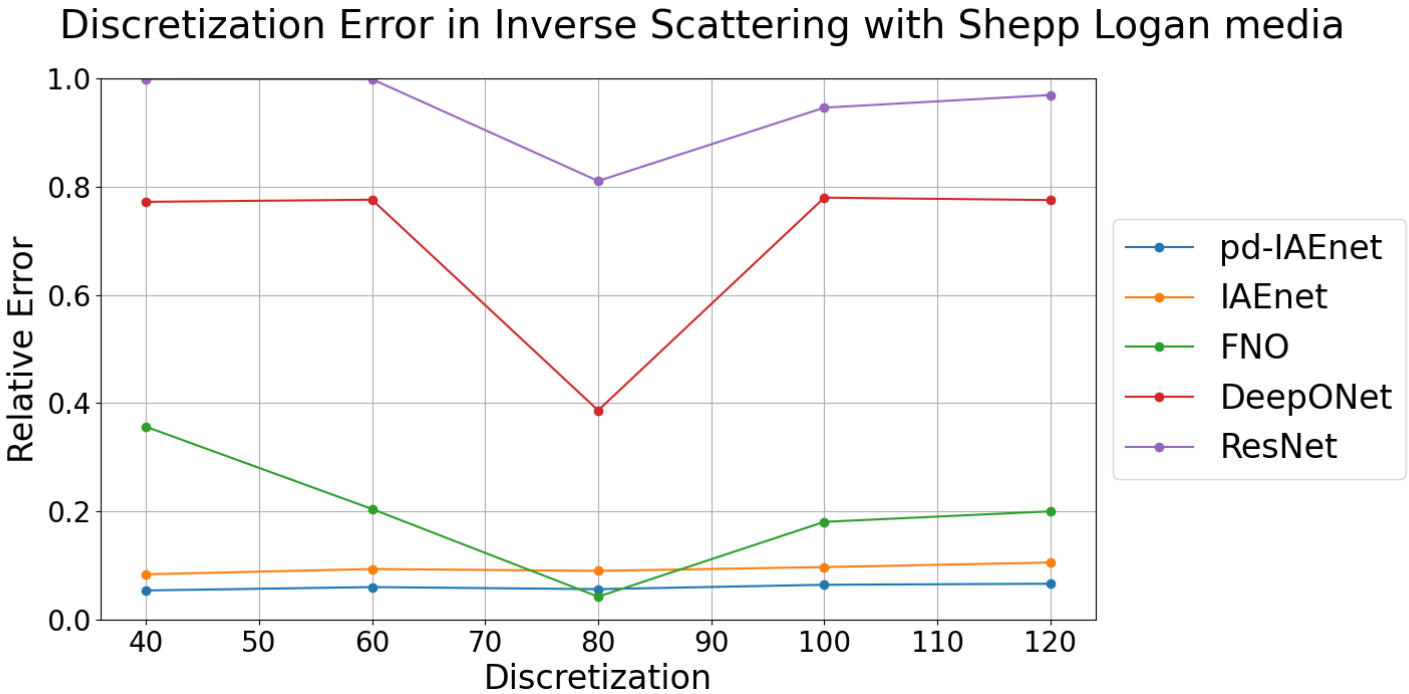


Figure: This graph provides the relative errors across various discretizations for inverse scattering with Shepp Logan media. The x-axis shows different discretization sizes: 40×40, 60×60, 80×80, 100×100, and 120×120. As the discretizations move away from the original setting of 80×80, the accuracy of FNO, DeepONet, and ResNet noticeably declines. Pd-IAE net and IAEnet consistently maintain high accuracy levels across all discretizations. Pd-IAE net stands out with the best overall accuracy and maintains near-uniform accuracy levels across all discretizations.

Discretization Invariance

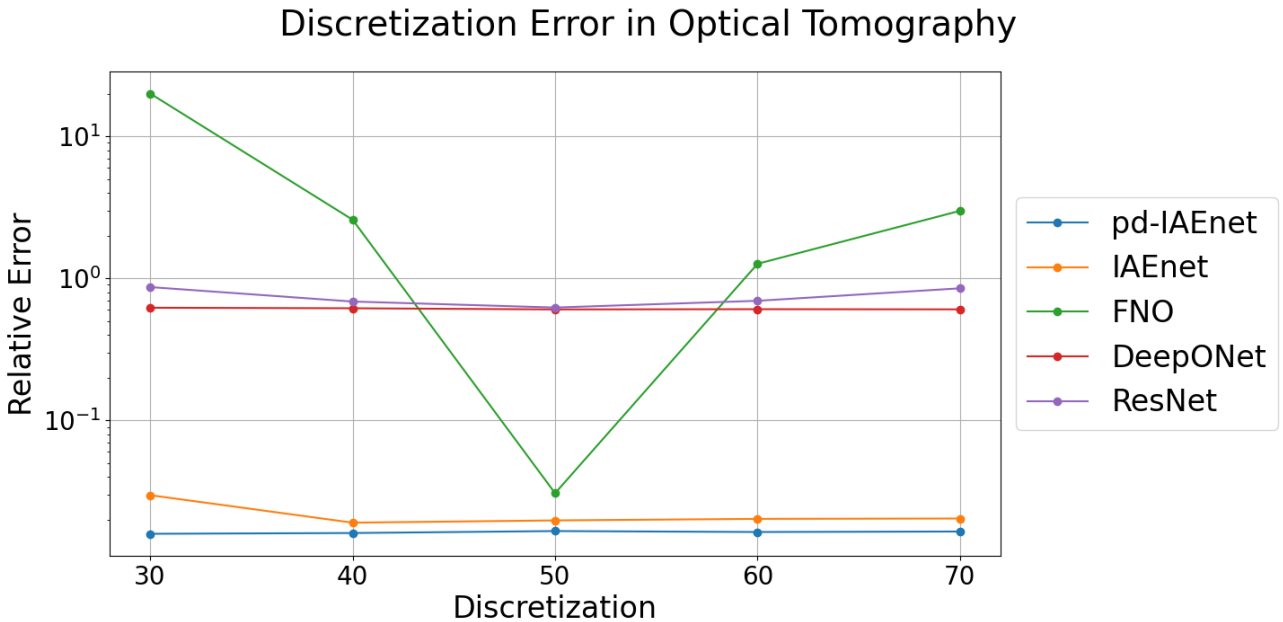


Figure: The graph offers the relative errors across a range of discretizations for OT. x-axis shows different discretization sizes: 30×30, 40×40, 50×50, 60×60, and 70×70. pd-IAE net and IAE net consistently exhibit uniform accuracy across all discretization meshes. Pd-IAE net stands out with the best overall accuracy and maintains near-uniform accuracy levels across all discretizations. FNO manages to achieve good accuracy on the 50×50 mesh but experiences a substantial relative error exceeding 100% on other meshes. Both DeepONet and ResNet display uniform errors across all discretization meshes with lower overall accuracy levels.

Robustness to noise

Model	Scattering (point media)	Scattering (Shepp-Logan media)	RTE
pd-IAEnet	0.004679	0.09102	0.01630
IAEnet	0.005768	0.09139	0.0197
FNO	0.007085	0.1969	0.6059
DeepONet	0.02351	0.6730	0.6090
ResNet	0.02372	0.8105	0.6987

Figure: The table shows the average relative errors with no noise.

Model	Scattering (point media)	Scattering (Shepp-Logan media)	RTE
pd-IAEnet	0.005187	0.1231	0.4042
IAEnet	0.007474	0.1212	0.4522
FNO	0.007357	0.2086	0.7704
DeepONet	0.02352	0.7150	0.6126
ResNet	0.02397	0.8804	0.7693

Figure: The table provides the average relative errors with noisy data. In inverse scattering problem, 1% additive noise is added to the measurement data for the point media and Shepp-Logan media. For OT, 0.25% noise is added to the measurement data.

Thank you very much for your attention!

Chunmei Wang
Department of Mathematics
University of Florida
Email: chunmei.wang@ufl.edu