Learning Deep Nonlocal (Integral) Operators for Heterogeneous Material Modeling

Yue Yu
Department of Mathematics, Lehigh University

Present at the Theoretical and Applied Aspects for nonlocal Models Workshop
July 21, 2022

LEHIGH UNIVERSITY

Outline

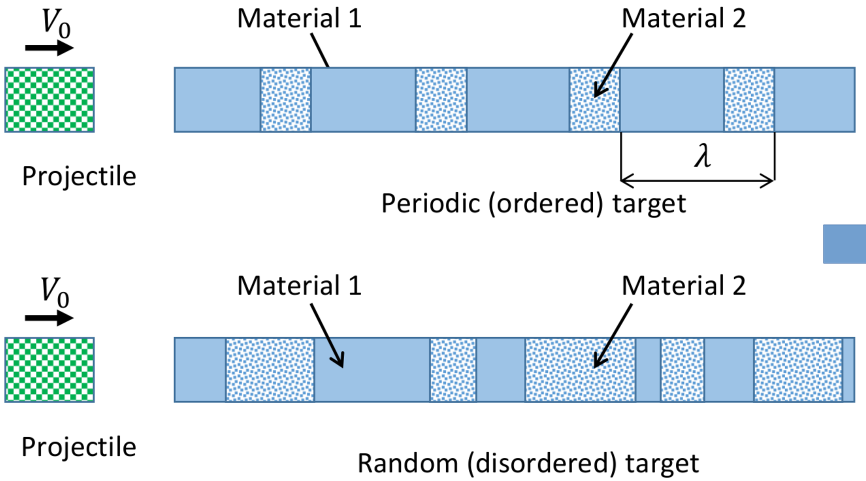
- Goal: modeling heterogeneous material behavior
- Key: Continuous and converging model via learning a nonlocal kernel
- Part I: Learning a Linear and Homogenized Model
  - ✓ To Learn: a nonlocal kernel function
- Part II: Learning a Nonlinear and Heterogeneous Model
  - ✓ To Learn: a nonlocal solution operator (kernel+NN)

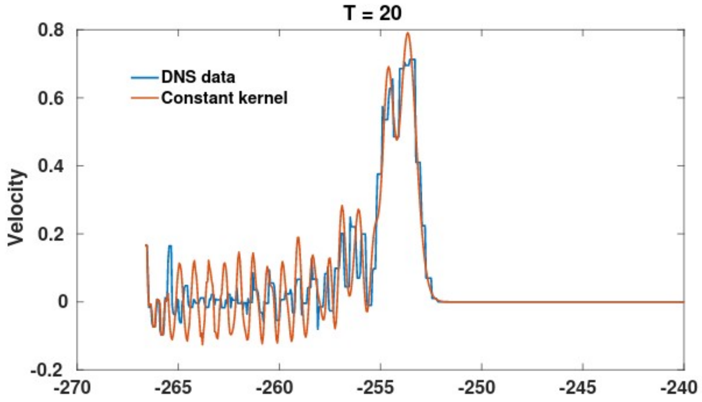
Motivation and Background

Goal: prediction and monitoring of heterogeneous material responses

- In heterogeneous materials, small-scale dynamics and interactions affect the global behavior.
- Fundamental challenges present, due to difficulties around computational scalability, variability, and data scarsity.

Exemplar problem 1: Impact on a heterogeneous bar





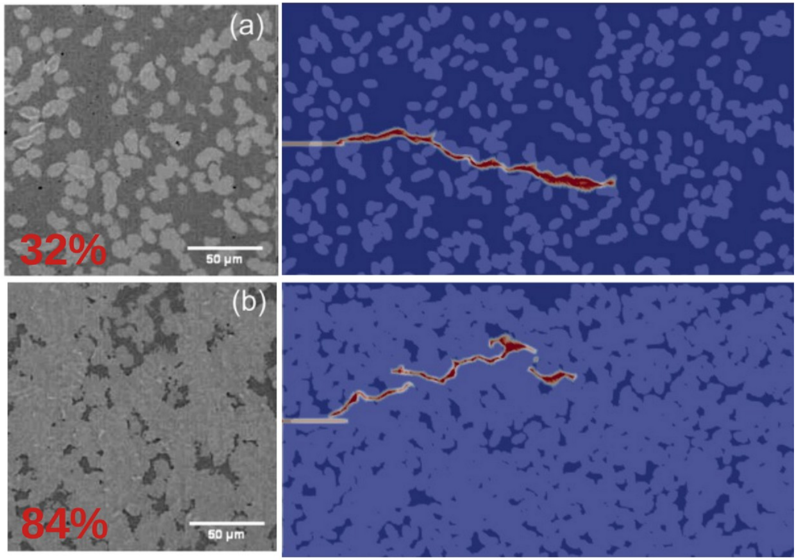
Motivation and Background

Goal: prediction and monitoring of heterogeneous material responses

- In heterogeneous materials, small-scale dynamics and interactions affect the global behavior.
- Fundamental challenges present, due to difficulties around computational scalability, variability, and data scarsity.

Exemplar problem 2: crack propagation on glass-ceramics.

Image source (iphone 12): cnet.com



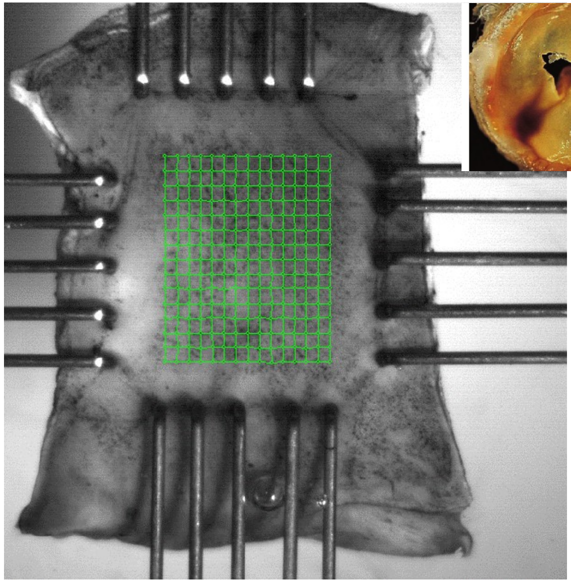
Numerical simulations using peridynamics. **Each takes 72 hours**

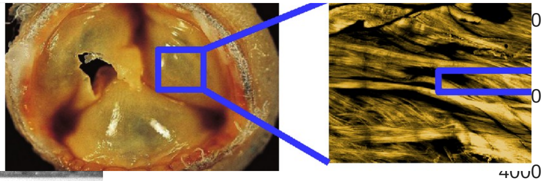
Motivation and Background

Goal: prediction and monitoring of heterogeneous material responses

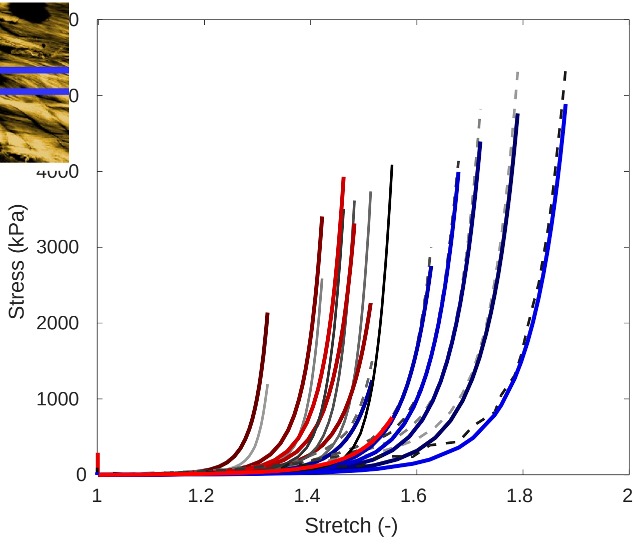
- In heterogeneous materials, small-scale dynamics and interactions affect the global behavior.
- Fundamental challenges present, due to difficulties around computational scalability, variability, and data scarsity.

Exemplar problem 3: heart valve leaflet modeling from experiment.





Mechanical Testing of heart valve leaflet



Conventional constitutive modeling: fails to capture the mechanical response (**displacement error ~ 10-30%**)

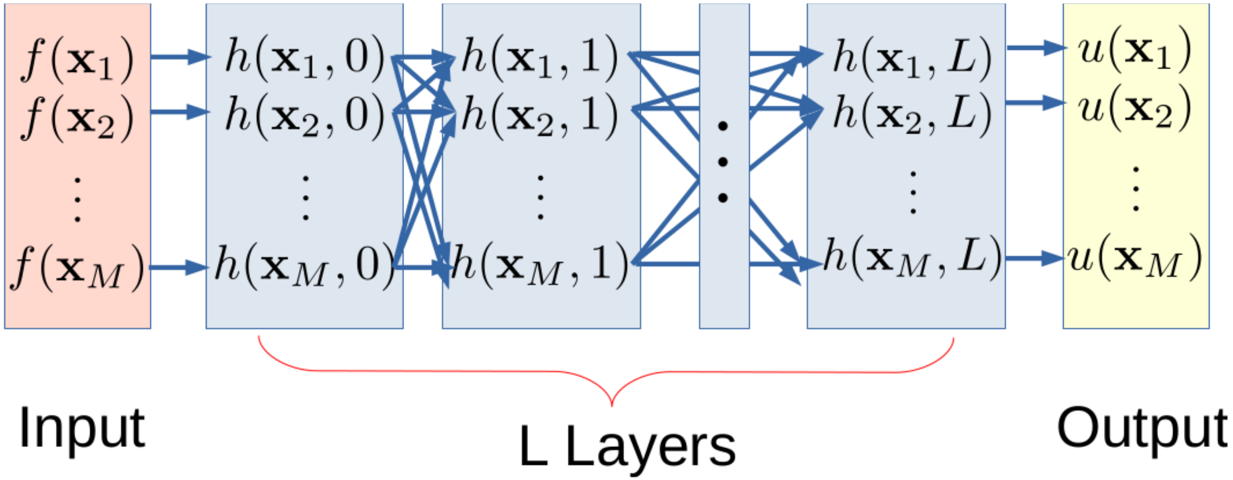
Motivation and Background

Goal: prediction and monitoring of heterogeneous material responses

- Desired properties: 1. the learnt model should be **generalizable to future prediction tasks**.
  2. the inverse problem should also be **well-posed and resolution independent, or even converging**.

$$\{b_1(\mathbf{x}_i), f_1(\mathbf{x}_i), u_1(\mathbf{x}_i)\}$$
$$\{b_2(\mathbf{x}_i), f_2(\mathbf{x}_i), u_2(\mathbf{x}_i)\}$$
$$\cdots$$
$$\{b_N(\mathbf{x}_i), f_N(\mathbf{x}_i), u_N(\mathbf{x}_i)\}$$

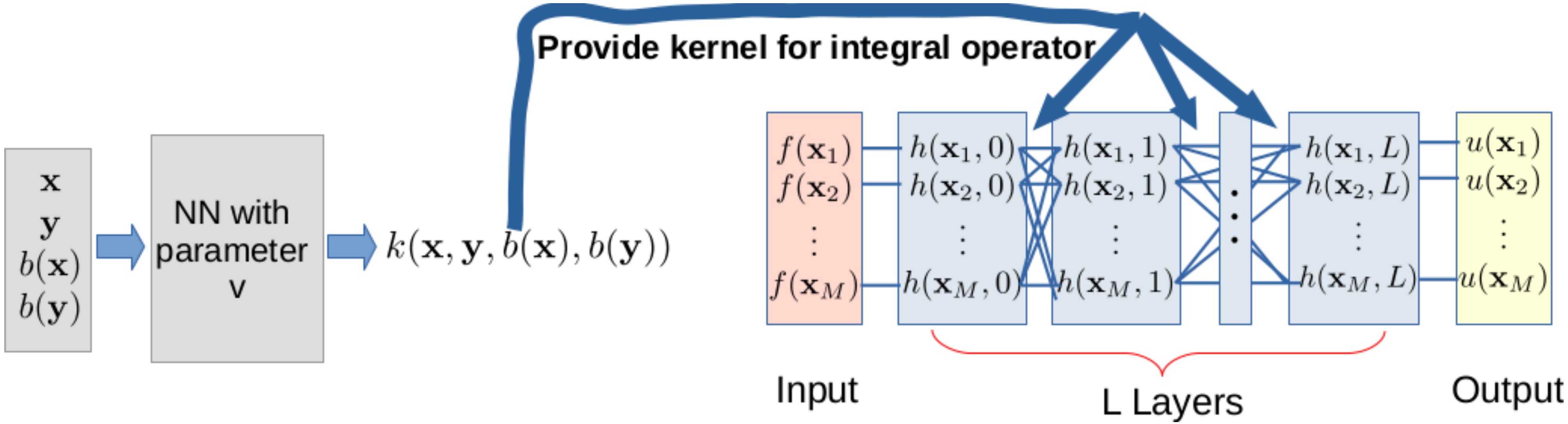
Training Samples



Motivation and Background

Goal: prediction and monitoring of heterogeneous material responses

- Desired properties: 1. the learnt model should be **generalizable to future prediction tasks**.
  2. the inverse problem should also be **well-posed and resolution independent, or even converging**.



Propose: Learning a nonlocal kernel!

Part I
Learning Kernels for Nonlocal Homogenized Models

[1] F. Lu, Q. An, Y. Yu, "Nonparametric learning of kernels in nonlocal operators". Submitted.

[2] H. You, Y. Yu, N. Trask, M. Gulian, M. D'Elia, "Data-driven learning of nonlocal physics from high-fidelity synthetic data", CMAME, 2021.

[3] H. You, Y. Yu, S. Silling, M. D'Elia, "Data-driven learning of nonlocal models: from high-fidelity simulations to constitutive laws". AAAI Spring Symposium: MLPS, 2021

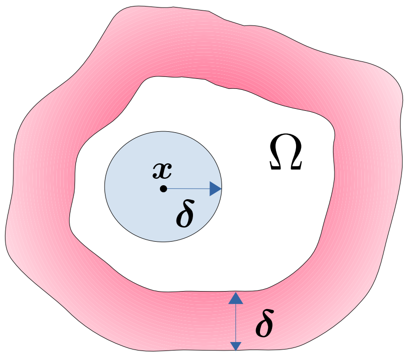
[4] H. You, Y. Yu*, S. Silling, M. D'Elia, "A data-driven peridynamic continuum model for upscaling molecular dynamics". CMAME, 2022.

[5] L. Zhang, H. You, Y. Yu*, "Meta-Learning for Metamaterials: A Provable Nonlocal Operator Regression Approach". Submitted.

What is a nonlocal (integral) model?

Basic concepts:

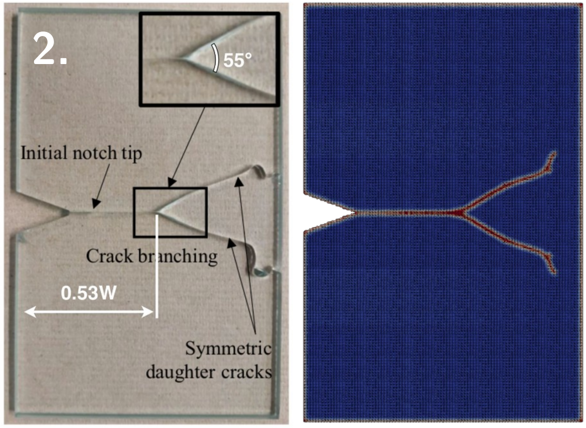
- The state of a system at any point depends on the state in a **neighborhood** of points
- Interactions can occur **at distance, without contact**
- Solutions can be irregular: non-differentiable, singular, discontinuous



Facts:

These models can capture effects that traditional PDEs **fail to capture**

1) Multiscale behavior (*nonlocal as an upscaled/homogenized model*)
2) Discontinuities such as cracks and fractures (*peridynamics)*
3) Anomalous behavior such as superdiffusion and subdiffusion *(fractional operators)*



Q. Du, B. Engquist, X. Tian, Multiscale modeling, homogenization and nonlocal effects: Mathematical and computational issues, Contemporary Mathematics 754.

Nonlocal Operator Regression (NOR)

Proposed: a 3-step recipe

- **Goal:** identify a nonlocal kernel k in $\mathcal{L}u(x) = \int_{B_\delta(x)} (u(y) - u(x)) k(x,y) dy$

$$\begin{cases} \ddot{u} = \mathcal{L}u + f & \text{in } \Omega \\ u = g & \text{on the nonlocal boundary} \end{cases}$$

1) **Collect measurements** of solution and forcing term: $\mathcal{D} = \{u_i(x), f_i(x)\}_{i=1}^N$

training set: measurements or high fidelity simulations

2) **Approximate the kernel** with a parameterization: $k(x,y) = \sum_{m=1}^{M} c_m \phi_m(x,y)$

3) **Minimize the residual** $\mathcal{E}_\lambda(k) = \frac{1}{N} \sum_{i=1}^{N} \|L_k[u_i] - f_i\|_{L^2}^2 + \lambda \mathcal{R}(k)$

subject to solvability and physical constraints.

outcome: coefficients c_m

H. You, Y. Yu, S. Silling, M. D'Elia, "Data-driven learning of nonlocal models: from high-fidelity simulations to constitutive laws". AAAI: MLPS, 2021

Nonlocal Operator Regression (NOR)

Proposed: a 3-step recipe

- **Goal:** identify a nonlocal kernel k in $\mathcal{L}_K u(x) = \int_{B_\delta(x)} (u(y) - u(x)) k(x, y; \mu) dy$

1) Collect measurements of solution and forcing term: $\mathcal{D} = \{u_i(x), f_i(x)\}_{i=1}^N$

training set: measurements or high fidelity simulations

2) Approximate the kernel with a parameterization: $k(x,y) = \sum_{m=1}^{M} c_m \phi_m(x,y)$

3) Minimize the residual $\mathcal{E}_\lambda(k) = \frac{1}{N} \sum_{i=1}^{N} \|L_k[u_i] - f_i\|_{L^2}^2 + \lambda \mathcal{R}(k)$

subject to solvability and physical constraints.

outcome: coefficients c_m

Key Algorithm Features/Contributions:

- Guarantees that the resultant surrogate model is **well-posed and physically consistent.**
- Applied through basis function design or penalization.

➡ **Genererabizable to Different Prediction Tasks**

H. You, Y. Yu, S. Silling, M. D'Elia, "Data-driven learning of nonlocal models: from high-fidelity simulations to constitutive laws". AAAI: MLPS, 2021

Nonlocal Operator Regression (NOR)

Proposed: a 3-step recipe

- **Goal:** identify a nonlocal kernel k in $\mathcal{L}_K u(x) = \int_{B_\delta(x)} (u(y) - u(x)) k(x, y; \mu) dy$

1) Collect measurements of solution and forcing term: $\mathcal{D} = \{u_i(x), f_i(x)\}_{i=1}^N$

training set: measurements or high fidelity simulations

2) Approximate the kernel with a parameterization: $k(x, y) = \sum_{m=1}^{M} c_m \phi_m(x, y)$

3) Minimize the residual $\mathcal{E}_\lambda(k) = \frac{1}{N} \sum_{i=1}^{N} \|L_k[u_i] - f_i\|_{L^2}^2 + \lambda \mathcal{R}(k)$

subject to solvability and physical constraints.

outcome: coefficients c_m

Key Algorithm Features/Contributions:

- One can selects a set of basis functions for a hypothesis space.
- **Learns the functional form of the kernel** (previous works only identify discrete parameters!).

→ **Converging Estimator (Kernel k)**

Lu, F., An, Q., & Yu, Y. (2022). Nonparametric learning of kernels in nonlocal operators. arXiv preprint arXiv:2205.11006.

Nonlocal Operator Regression (NOR)

Proposed: a 3-step recipe

- **Goal:** identify a nonlocal kernel k in $\mathcal{L}_K u(x) = \int_{B_\delta(x)} (u(y) - u(x)) k(x, y; \mu) dy$

1) Collect measurements of solution and forcing term: $\mathcal{D} = \{u_i(x), f_i(x)\}_{i=1}^N$

training set: measurements or high fidelity simulations

2) Approximate the kernel with a parameterization: $k(x,y) = \sum_{m=1}^{M} c_m \phi_m(x,y)$

3) Minimize the residual $\mathcal{E}_\lambda(k) = \frac{1}{N} \sum_{i=1}^{N} \|L_k[u_i] - f_i\|_{L^2}^2 + \lambda \mathcal{R}(k)$

subject to solvability and physical constraints.

outcome: coefficients c_m

Key Algorithm Features/Contributions:

- A regularization term is necessary, or the inverse problem becomes ill-posed as $\Delta x \to 0$.
- A **system-intrinsic data-adaptive reproducing kernel Hilbert space (SIDA-RKHS) regularization term** is developed.

→ **Identifiability and Robustness to Noise**

Lu, F., An, Q., & Yu, Y. (2022). Nonparametric learning of kernels in nonlocal operators. arXiv preprint arXiv:2205.11006.

NOR: Convergence and Robustness to Noise

- **Training set:** $\mathcal{D} = \{u_i(x), f_i(x)\}_{i=1}^N$, generated from the nonlocal equation $\mathcal{L}_K u(x) = f(x)$ where $\mathcal{L}_K$ is associated to a manufactured kernel $k_{true}(x,y) := k_{true}(|x-y|)$

- **Manufactured kernel:** $k_{true}(r) = c_{d,s}\frac{1}{r^{d+2s}}\mathbf{1}_{[0.1,6]}(x) + \frac{1}{0.1^{d+2s}}\mathbf{1}_{[0,0.1]}(x)$ where $d=1, s=0.5.$

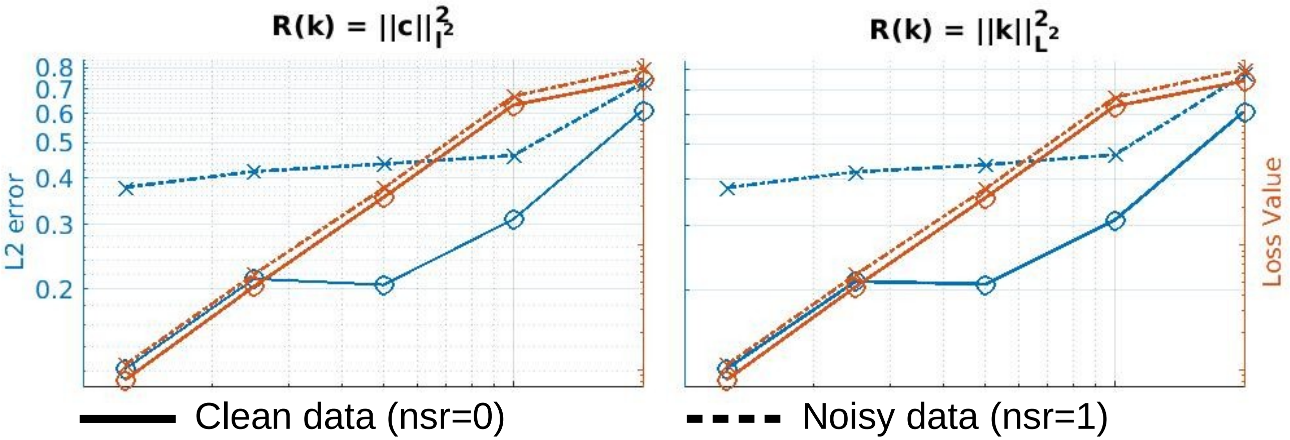
- **Optimization-based learning:** $\min_{c_m} \frac{\Delta x}{N}\sum_{i=1}^{N}\sum_{j=1}^{J} |L_k[u_i](x_j) - f_i(x_j)|^2 + \lambda\mathcal{R}(k)$

  where k is approximated by B-splines: $k(x,y) = k(|x-y|) = k(r) = \sum_{m=1}^{M} c_m\phi_m(r)$

When taking the classical Tikhonov regularization:

$$\mathcal{R}(k) = \|c\|_{l^2}^2 \text{ or } \mathcal{R}(k) = \|k\|_{L^2}^2$$

Convergence of function estimator as the data mesh-size Δx decreases from 0.2 to 0.0125:



NOR: Convergence and Robustness to Noise

- **Training set:** $\mathcal{D} = \{u_i(x), f_i(x)\}_{i=1}^N$, generated from the nonlocal equation $\mathcal{L}_K u(x) = f(x)$ where $\mathcal{L}_K$ is associated to a manufactured kernel $k_{true}(x,y) := k_{true}(|x-y|)$

Theorem (Function space of identifiability) [Lu, An, Yu, 2022]:
Consider the problem of identifying the kernel k, the function space of identifiability, in which the true kernel is the unique minimizer of the loss functional, is an RKHS (denoted by H_G) with reproducing kernel:

$$\bar{G}(r,s) = \frac{G(r,s)}{\rho'_N(r)\rho'_N(s)}, \text{ where } G(r,s) = \frac{1}{N}\sum_{i=1}^N \int_{|\eta|=1}\int_{|\xi|=1}\left[\int [u_i(x+r\xi)-u_i(x)][u_i(x+s\eta)-u_i(x)]dx\right] d\xi d\eta$$

where ρ'_N is the density of an empirical probability density $\rho_N(dr) = \frac{1}{ZN}\sum_{i=1}^N \int_\Omega\int_\Omega \delta_{|x-y|}(r)|u_i(x)-u_i(y)|dxdy$.

Theorem (Characterization of the RKHS space) [Lu, An, Yu, 2022]:
The RKHS H_G with $\bar{G}$ as reproducing kernel satisfies $H_G = \mathcal{L}_{\bar{G}}^{1/2}(L^2(\rho_N))$, where $L_{\bar{G}}$ is an integral operator defined by

$$\mathcal{L}_{\bar{G}} k(r) = \int_0^\infty k(s)\bar{G}(r,s)\rho_N(s)ds$$

The eigenvalues of $L_{\bar{G}}$ converges to zero, and its eigen-functions $\{\psi_l(r)\}$ can form a complete orthonormal basis of $L^2(\rho_N)$.

Lu, F., An, Q., & Yu, Y. (2022). Nonparametric learning of kernels in nonlocal operators. arXiv preprint arXiv:2205.11006.

NOR: Convergence and Robustness to Noise

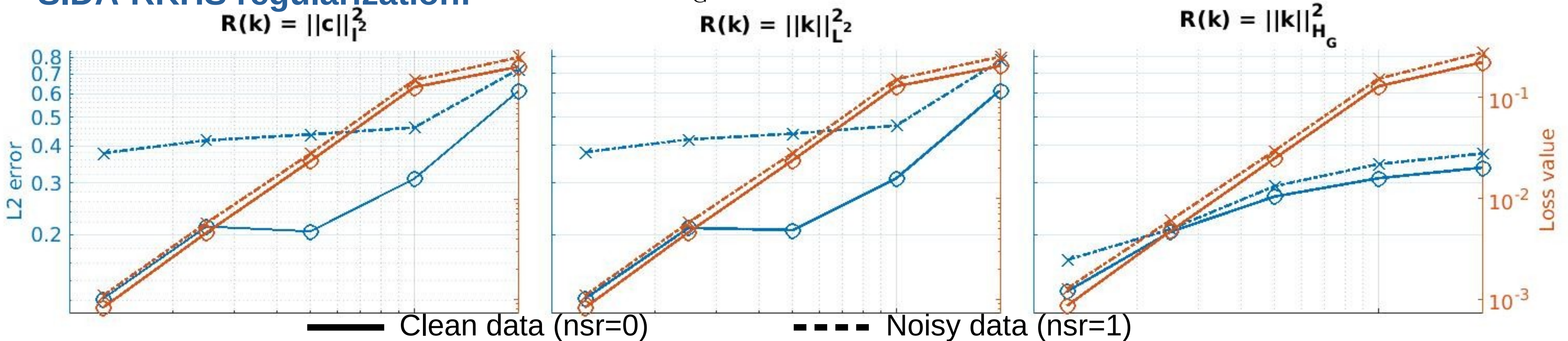
- **Training set:** $\mathcal{D} = \{u_i(x), f_i(x)\}_{i=1}^N$, generated from the nonlocal equation $\mathcal{L}_K u(x) = f(x)$ where $\mathcal{L}_K$ is associated to a manufactured kernel $k_{true}(x,y) := k_{true}(|x-y|)$

- **Manufactured kernel:** $k_{true}(r) = c_{d,s}\frac{1}{r^{d+2s}}\mathbf{1}_{[0.1,6]}(x) + \frac{1}{0.1^{d+2s}}\mathbf{1}_{[0,0.1]}(x)$ where $d=1, s=0.5.$

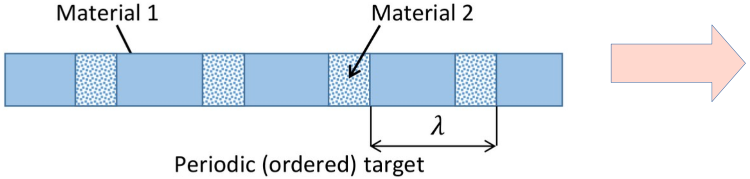
- **Optimization-based learning:** $\min_{c_m} \frac{\Delta x}{N}\sum_{i=1}^N\sum_{j=1}^J |L_k[u_i](x_j) - f_i(x_j)|^2 + \lambda\mathcal{R}(k)$

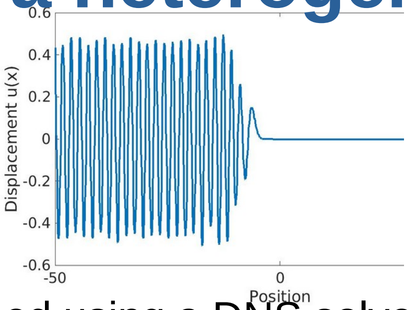
  where k is approximated by B-splines: $k(x,y) = k(|x-y|) = k(r) = \sum_{m=1}^M c_m\phi_m(r)$

- **SIDA-RKHS regularization:** $\mathcal{R}(k) = \|k\|_{H_G}^2$



NOR: Wave propagation in a heterogeneous bar



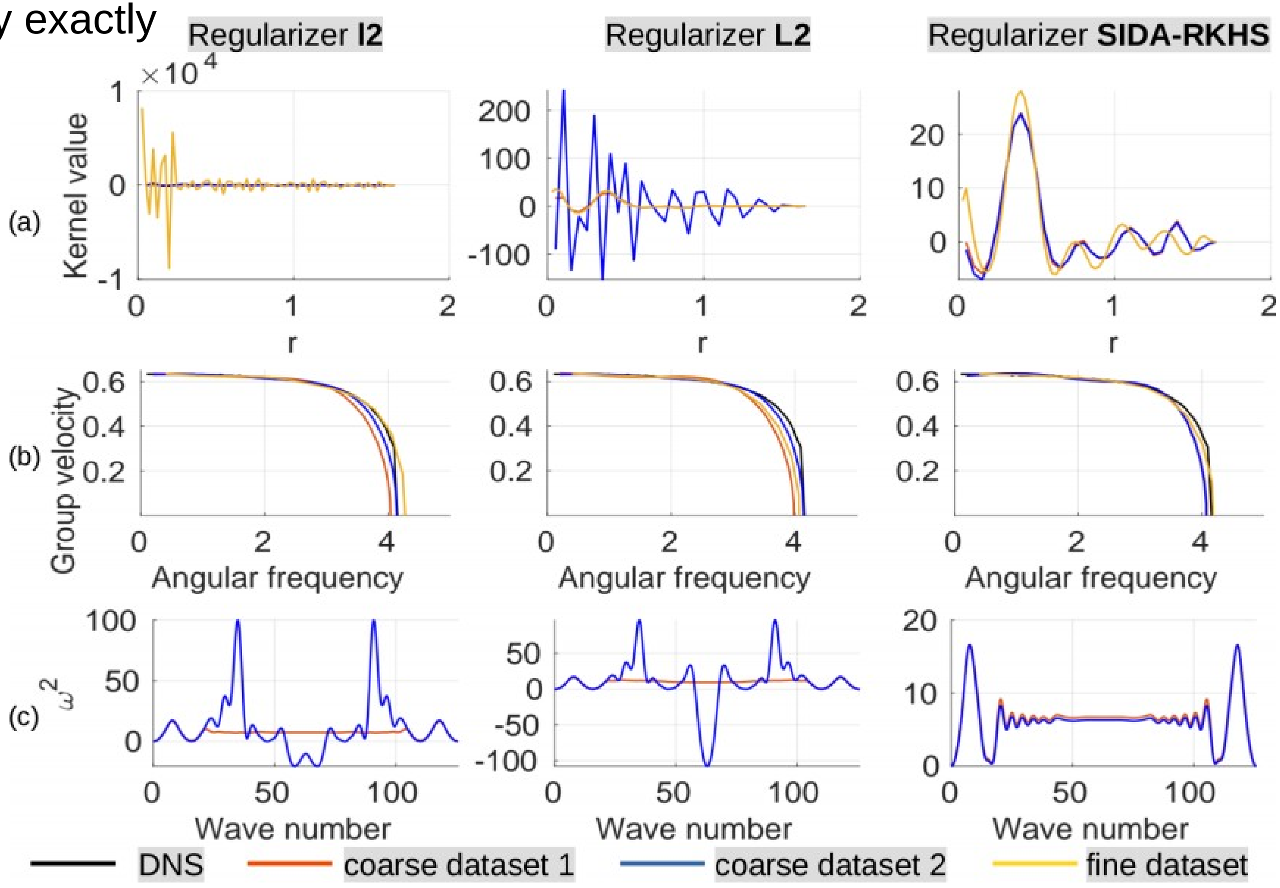


- **Training set:** oscillating source and plane wave obtained using a DNS solver that computes the velocity exactly, with t from 0 to 2.
  Oscillating source: $\Omega = [-50, 50]$, $f(x,t) = \exp^{-(\frac{2x}{5jL})^2} \exp^{-(\frac{t-0.8}{0.8})^2} \cos^2(\frac{2\pi x}{jL})$, for $j = 1, 2, \cdots, 20$.
  Plane wave 1: $\Omega = [-50, 50]$, $f(x,t) = 0$, $u(x,0) = 0$, $v(-50,t) = \cos(jt)$ for $j = 0.35, 0.7, \cdots, 3.85$.
  Plane wave 2: $\Omega = [-50, 50]$, $f(x,t) = 0$, $u(x,0) = 0$, $v(-50,t) = \sin(jt)$ for $j = 0.35, 0.7, \cdots, 3.85$.

- **Experiments:**
  Coarse data set 1: we train the estimator using ``coarse'' dataset (Δx=0.05) of oscillating source and plane wave 1.
  Coarse data set 2: we train the estimator using ``coarse'' dataset (Δx=0.05) of oscillating source and plane wave 2.
  Fine data set: we train the estimator using ``fine'' dataset (Δx=0.025) of oscillating source and plane wave 1.

NOR: Wave propagation in a heterogeneous bar

- **Training set:** oscillating source and plane wave obtained using a DNS solver that computes the velocity exactly



Kernel convergence

matching DNS indicates physical consistency

>0 indicates physical stability

NOR: Wave propagation in a heterogeneous bar

- **Training set:** oscillating source and plane wave obtained using a DNS solver that computes the velocity exactly, with t from 0 to 2.
  Oscillating source: $\Omega = [-50, 50]$, $f(x,t) = \exp^{-(\frac{2x}{5jL})^2} \exp^{-(\frac{t-0.8}{0.8})^2} \cos^2(\frac{2\pi x}{jL})$, for $j = 1, 2, \cdots, 20$.
  Plane wave 1: $\Omega = [-50, 50]$, $f(x,t) = 0$, $u(x,0) = 0$, $v(-50,t) = \cos(jt)$ for $j = 0.35, 0.7, \cdots, 3.85$.
  Plane wave 2: $\Omega = [-50, 50]$, $f(x,t) = 0$, $u(x,0) = 0$, $v(-50,t) = \sin(jt)$ for $j = 0.35, 0.7, \cdots, 3.85$.

- **Test set:** wave packet obtained using a DNS solver with a different loading and domain, from the training dataset, and with a much longer simulation time (t from 0 to 100).
  Wave packet: $\Omega = [-133.3, 133.3]$, $f(x,t) = 0$, $u(x,0) = 0$, $v(-133.3,t) = \sin(jt)\exp\left(-(t/5-3)^2\right)$, for $j = 1, 2, 3$.

The relative L2 errors of long term (T=100) displacement prediction on the test dataset:

Resolution	l2	L2	SIDA-RKHS
Coarse ($\Delta x = 0.05$)	23.5%	28.4%	**21.8%**
Fine ($\Delta x = 0.025$)	INF	23.4%	**19.2%**

Part II
Learning Integral Neural Operators for Heterogeneous Models

[1] H. You, Q. Zhang, C. Ross, C-H. Lee, Y. Yu*, “Learning Deep Implicit Fourier Neural Operators (IFNOs) with Applications to Heterogeneous Material Modeling”. CMAME, 2022.
[2] H. You, Q. Zhang, C. Ross, C-H. Lee, M-C. Hsu, Y. Yu*, “A Physics-Guided Neural Operator Learning Approach to Model Biological Tissues from Digital Image Correlation Measurements”. arXiv preprint arXiv:2204.00205.
[3] H. You, Y. Yu*, M. D’Elia, T. Gao, S. Silling, “Nonlocal Kernel Network (NKN): a stable and resolution independent deep neural network”. arXiv preprint arXiv:2201.02217
[4] S. Goswami, A. Bora, Y. Yu, G. Karniadakis*, “Physics-Informed Neural Operators". Submitted.

Neural Operator Learning

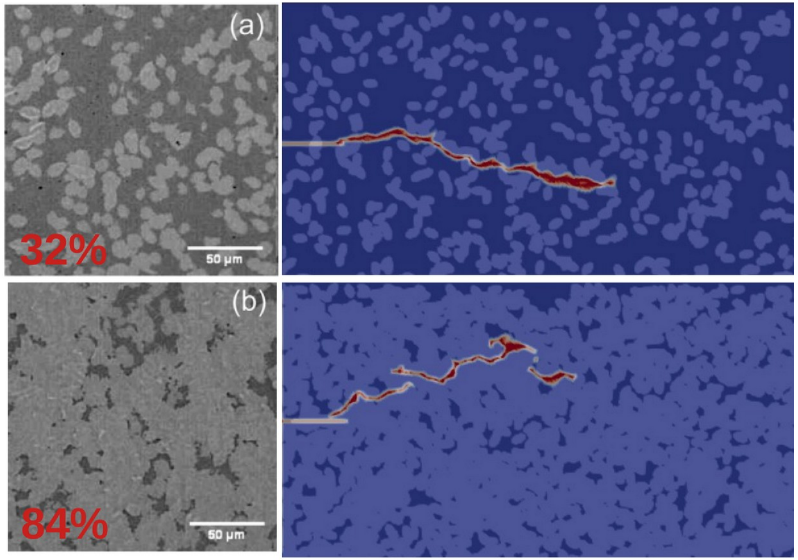
Goal: prediction and monitoring of heterogeneous material responses

- **Idea**: the material displacement and damage modeling and solving problem can be seen as to **find a solution operator:**

$$G: b(x) \rightarrow u(x)$$

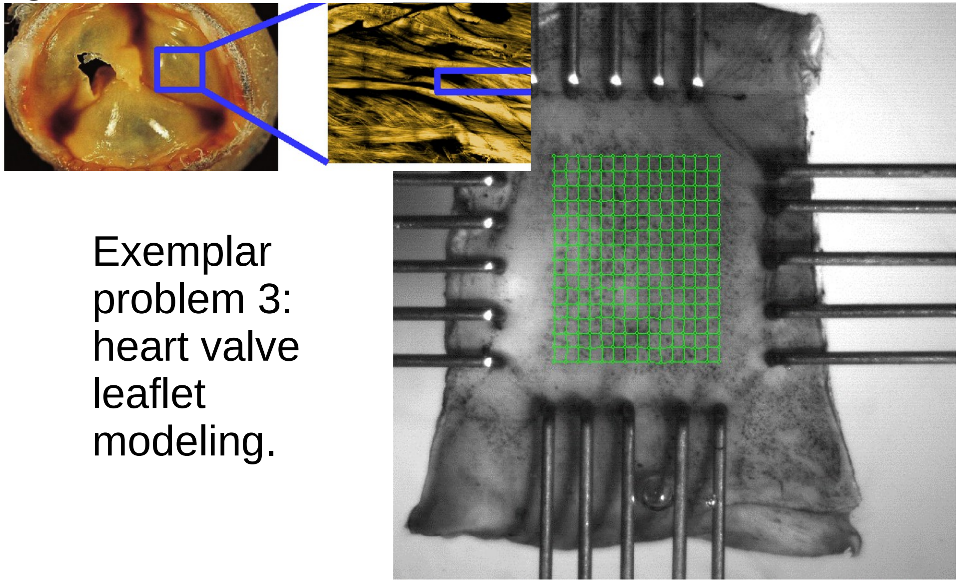
where b can be the boundary condition/external loading/initial condition/microstructure.

Exemplar problem 2: crack on glass-ceramics.



Crack propagation simulations using peridynamics.

Exemplar problem 3: heart valve leaflet modeling.



Mechanical Testing of heart valve leaflet

Neural Operator Learning

Goal: prediction and monitoring of heterogeneous material responses

- We propose to use **neural operator learning** approach, which directly learns material responses from high-fidelity simulations or experimental data.
- Assume an **unknown** governing equation

$$-\mathrm{L}_a[u](\mathbf{x}) = f(\mathbf{x}), \quad \mathbf{x} \in D,$$
$$u(\mathbf{x}) = u_{bc}(\mathbf{x}), \quad \mathbf{x} \in \partial D,$$

- Learn the operator G, such that for each $\mathbf{b}(\mathbf{x}) = [\mathbf{x}, a(\mathbf{x}), f(\mathbf{x}), u_{bc}(\mathbf{x})]$, the solution u=G(b).
- **Advantages:**

  1. Only require observed data pairs $\{(\mathbf{b}_j, \mathbf{u}_j)\}_{j=1}^N$, and hence can be applied when the underlying PDE is unknown. ➡ Exemplar problem 3

  2. For every new instance of b, requires only a forward pass of the network. ➡ Exemplar problem 2

  3. No further modification or tuning will be required for different resolutions and discretizations.

[1]L. Lu, P. Jin, G. Pang, Z. Zhang, G. E. Karniadakis, Learning nonlinear operators via deeponet based on the universal approximation theorem of operators, Nature Machine Intelligence 3 (3) (2021) 218–229.
[2]Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, A. Anandkumar, Neural operator: Graph kernel network for partial differential equations, arXiv preprint arXiv:2003.03485.
[3]Benner, P., Goyal, P., Kramer, B., Peherstorfer, B., & Willcox, K. (2020). Operator inference for non-intrusive model reduction of systems with non-polynomial nonlinear terms. Computer Methods in Applied Mechanics and Engineering, 372, 113433.

Integral Operator Learning

- Integral Kernel Networks: constructing a parametric map from b to u

$$\mathcal{G} : \mathcal{A} \times \Theta \to \mathcal{U}$$

- Consider an elliptic equation

$$\begin{aligned} -\mathrm{L}_b[u](\mathbf{x}) &= f(\mathbf{x}), \quad \mathbf{x} \in D, \\ u(\mathbf{x}) &= 0, \quad \mathbf{x} \in \partial D, \end{aligned} \quad \Longrightarrow \quad u(\mathbf{x}) = \int_D G_b(\mathbf{x}, \mathbf{y}) f(\mathbf{y}) d\mathbf{y}.$$

- Li et al[1,2] proposed to parameterize the Green's function G_b as a neural network. For an L-layer NN, the l-th layer network update is

$$\mathbf{h}(\mathbf{x}, l+1) = \sigma \left(R(l)\mathbf{h}(\mathbf{x}, l) + \int_D k(\mathbf{x}, \mathbf{y}, \mathbf{b}(\mathbf{x}), \mathbf{b}(\mathbf{y}); \mathbf{v}(l)) \mathbf{h}(\mathbf{y}, l) d\mathbf{y} + \mathbf{c}(l) \right).$$

$$k(\mathbf{x}, \mathbf{y}, \mathbf{b}(\mathbf{x}), \mathbf{b}(\mathbf{y}); \mathbf{v}(l)) := k(\mathbf{x} - \mathbf{y}; \mathbf{v}(l))$$ ⟹ FNO[1]

[1]Z. Li, N. B. Kovachki, K. Azizzadenesheli, K. Bhattacharya, A. Stuart, A. Anandkumar, et al., Fourier neural operator for parametric partial differential equations, in: International Conference on Learning Representations, 2021.

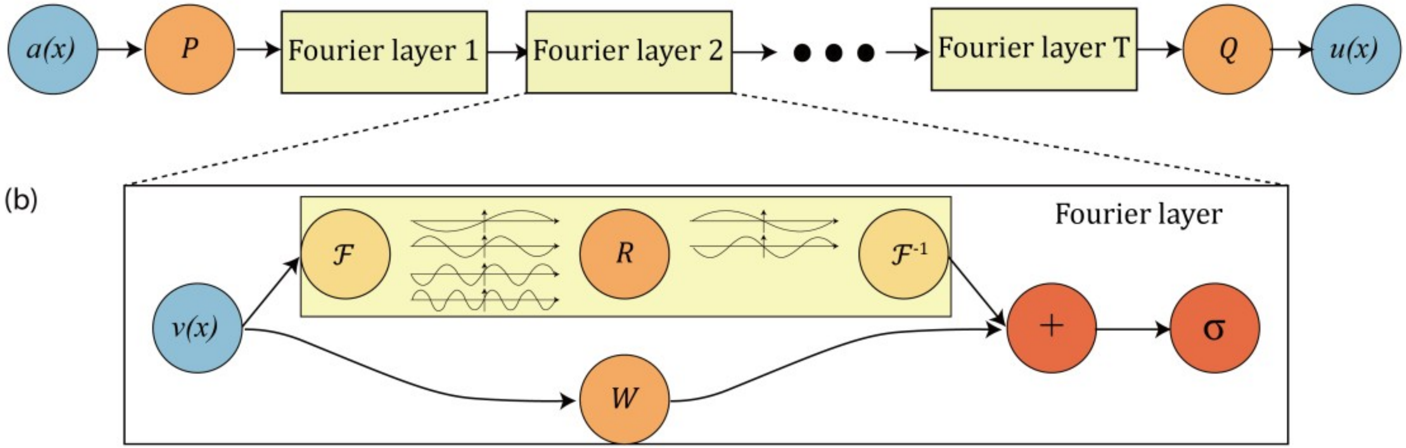
Fourier Neural Operator (FNO)

- FNO: parameterize the integral kernel directly in Fourier space, and learns the mapping between function spaces.

$$\mathbf{h}(\mathbf{x}, l+1) = \sigma\left(R(l)\mathbf{h}(\mathbf{x}, l) + \int_D k(\mathbf{x}-\mathbf{y}; \mathbf{v}(l))\mathbf{h}(\mathbf{y}, l)d\mathbf{y} + \mathbf{c}(l)\right).$$

$$\mathbf{h}(\mathbf{x}, l+1) = \sigma\left(R(l)\mathbf{h}(\mathbf{x}, l) + \mathcal{F}^{-1}(\mathcal{F}(k(\cdot; \mathbf{v}(l))) \cdot \mathcal{F}(\mathbf{h}(\cdot, l)))(\mathbf{x}) + \mathbf{c}(l)\right).$$

- Allows Fast Fourier Transform (FFT) to efficiently compute the integral.
- Generalizes well to different meshes and parameters b.



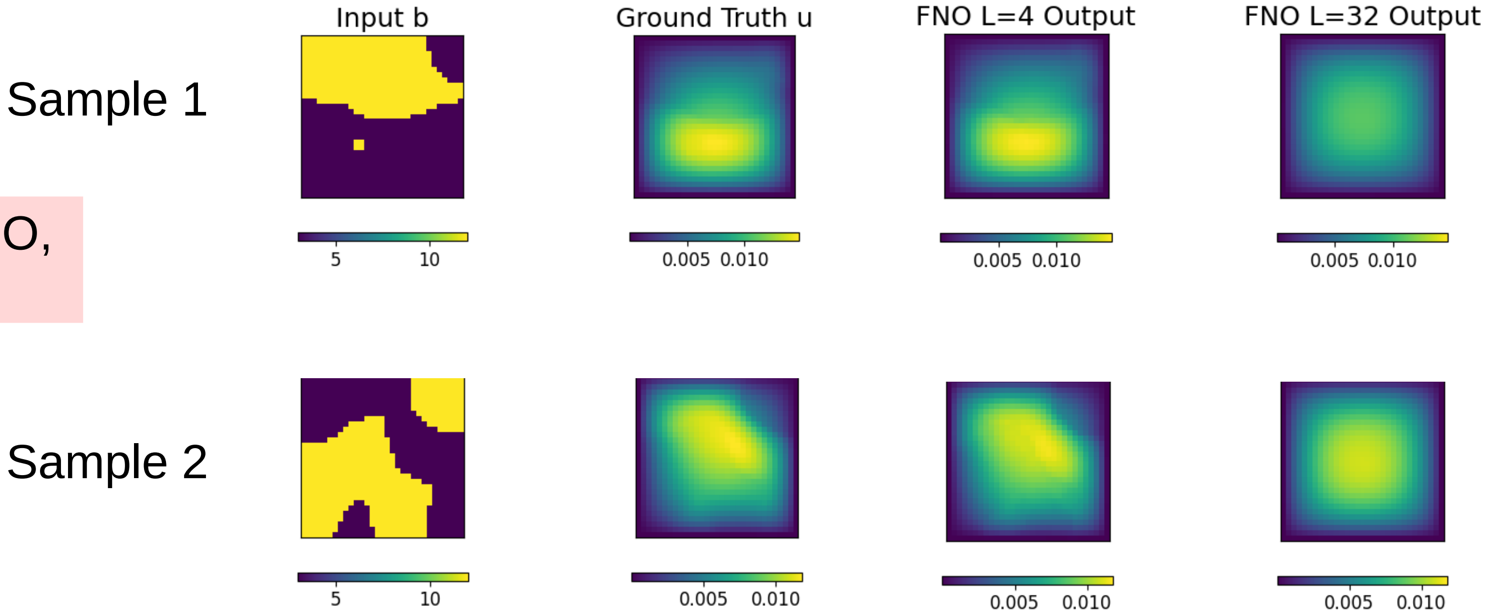
Deep FNO: a possible pitfall

- Consider a 2D Darcy's equation

$$-\nabla \cdot (a(\mathbf{x})\nabla u(\mathbf{x})) = f(\mathbf{x}) \quad \text{on} \quad D = [0,1]^2$$

use FNO to construct a mapping from a to u, with 1000 pairs of $\{(a_j(x),u_j(x))\}$.

Representative FNO, L=4&32 results

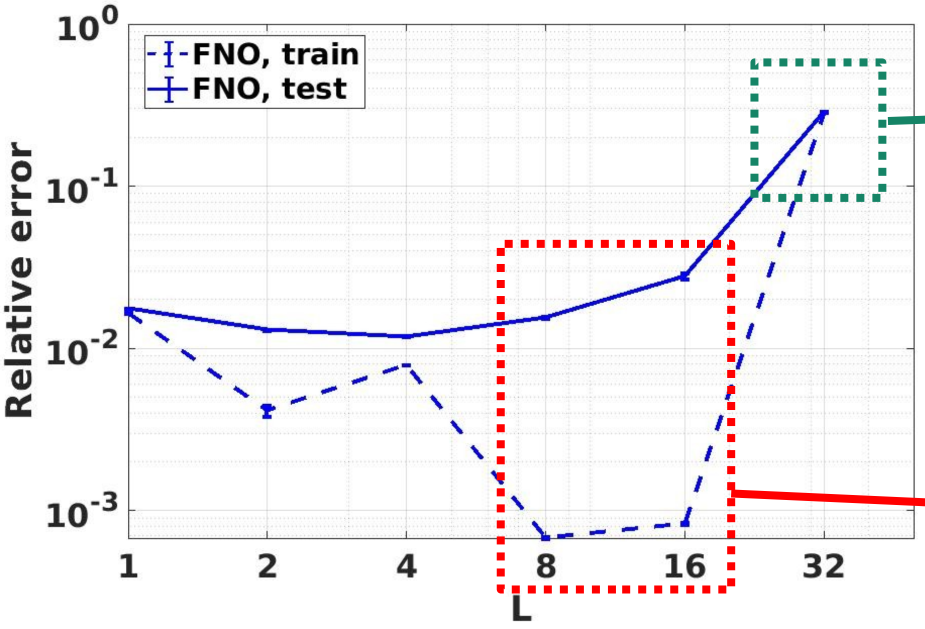


Deep FNO: a possible pitfall

- Consider a 2D Darcy's equation

$$-\nabla \cdot (a(\mathbf{x})\nabla u(\mathbf{x})) = f(\mathbf{x}) \quad \text{on} \quad D = [0,1]^2$$

use FNO to construct a mapping from a to u, with 1000 pairs of $\{(a_j(x), u_j(x))\}$.



Vanishing gradient

# of trainable parameters
FNO, L=1: 0.17M
FNO, L=32: 5.35M

Overfitting

Implicit Fourier Neural Operator (IFNO)

- Integral Kernel Networks: constructing a parametric map from b to u

$$\mathcal{G} : \mathcal{A} \times \Theta \to \mathcal{U}$$

- Consider an elliptic equation as an implicit problem:

$$\begin{aligned} -\mathrm{L}_b[u](\mathbf{x}) &= f(\mathbf{x}), \quad \mathbf{x} \in D, \\ u(\mathbf{x}) &= 0, \quad \mathbf{x} \in \partial D, \end{aligned} \quad \Longrightarrow \quad F_b(U) = 0, \text{ where } U = [\mathbf{u}(\mathbf{x}_1), \cdots, \mathbf{u}(\mathbf{x}_M)].$$

- **Idea 1**: Solve for U using the Newton-Raphson method iteratively:

$$U^{l+1} = U^l - [\nabla_U F_b(U^l)]^{-1} F_b(U^l)$$

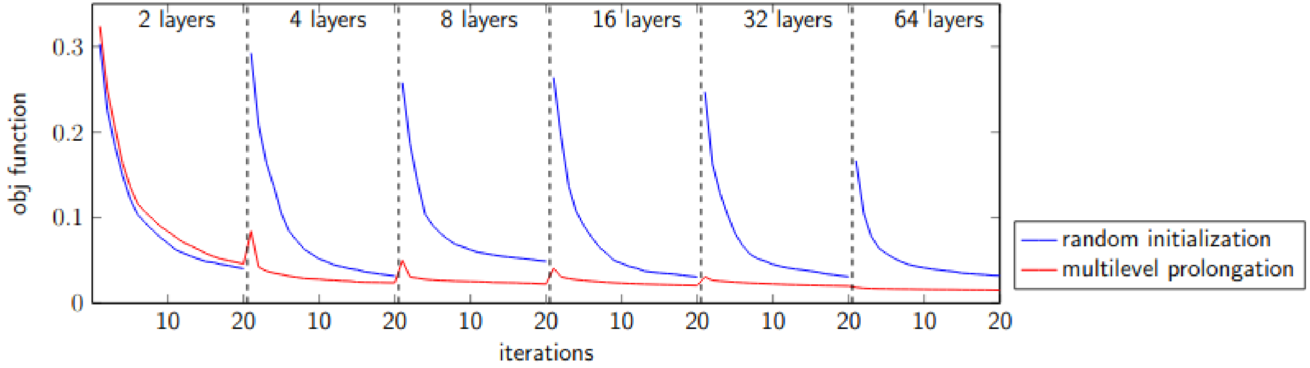
and use FNO layer to mimic the (autonomous) operator $-[\nabla_U F_b(\cdot)]^{-1} F_b(\cdot)$:

$$\mathbf{h}(\mathbf{x}, l+1) = \mathbf{h}(\mathbf{x}, l) + \sigma \left(R\mathbf{h}(\mathbf{x}, l) + \mathcal{F}^{-1}(\mathcal{F}(k(\cdot; \mathbf{v})) \cdot \mathcal{F}(\mathbf{h}(\cdot, l)))(\mathbf{x}) + \mathbf{c} \right).$$

Fung, S. W., Heaton, H., Li, Q., McKenzie, D., Osher, S., & Yin, W. (2021). JFB: Jacobian-free backpropagation for implicit networks. arXiv preprint arXiv:2103.12803.

Implicit Fourier Neural Operator (IFNO)

- **Idea 2: ResNet and Shallow-to-Deep Technique[2]**
  For a Deep NN with ResNet architecture:
  $$\mathbf{h}(l+1) = \mathbf{h}(l) + \mathcal{R}(\mathbf{h}(l); \mathbf{v}),$$
  the forward propagation can be seen as a discretization of a time-dependent nonlinear ODE:
  $$\mathbf{h}(t+\Delta t) = \mathbf{h}(t) + \Delta t\, \tilde{\mathcal{R}}(\mathbf{h}(t), \mathbf{v}).$$
- Haber et al. proposes to accelerate the training of deep networks by using the parameter v trained with depth L as the initial parameter for depth $\tilde{L} > L$:



[1]Haber, E., Ruthotto, L., Holtham, E., & Jun, S. H. (2018, April). Learning Across Scales—Multiscale Methods for Convolution Neural Networks. In Thirty-Second AAAI Conference on Artificial Intelligence.
[2]H. You, Y. Yu, M. D'Elia, T. Gao, S. Silling, "Nonlocal Kernel Network (NKN): a stable and resolution independent deep neural network". arXiv preprint arXiv:2201.02217

Implicit Fourier Neural Operator (IFNO)

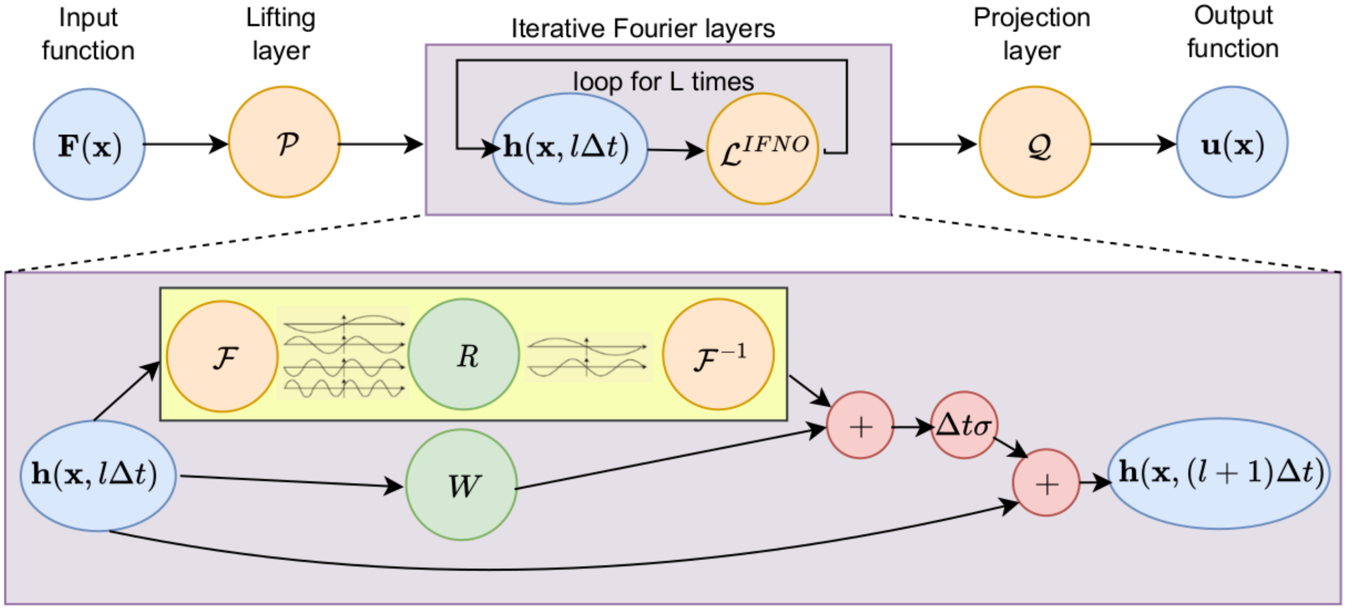
- Combining ideas 1 and 2:

FNO:

$$\mathbf{h}(\mathbf{x}, l+1) = \sigma\left(R(l)\mathbf{h}(\mathbf{x}, l) + \mathcal{F}^{-1}(\mathcal{F}(k(\cdot;\mathbf{v}(l))) \cdot \mathcal{F}(\mathbf{h}(\cdot, l)))(\mathbf{x}) + \mathbf{c}(l)\right).$$

IFNO:

$$\mathbf{h}(\mathbf{x}, t+\Delta t) = \mathbf{h}(\mathbf{x}, t) + \Delta t \sigma\left(R\mathbf{h}(\mathbf{x}, t) + \mathcal{F}^{-1}(\mathcal{F}(k(\cdot;\mathbf{v})) \cdot \mathcal{F}(\mathbf{h}(\cdot, t)))(\mathbf{x}) + \mathbf{c}\right).$$



Implicit Fourier Neural Operator (IFNO)

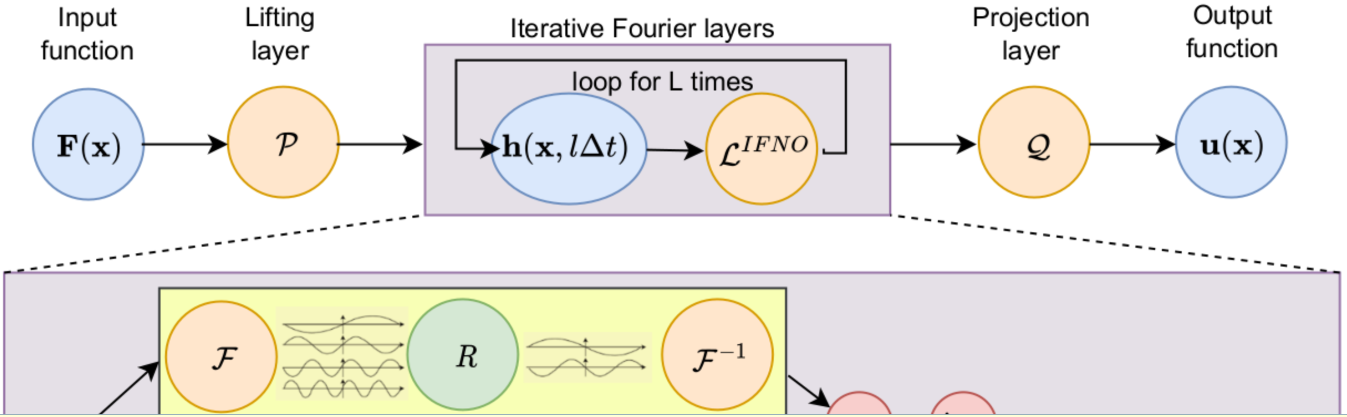
- Combining ideas 1 and 2:

FNO:

$$\mathbf{h}(\mathbf{x}, l+1) = \sigma\left(R(l)\mathbf{h}(\mathbf{x}, l) + \mathcal{F}^{-1}(\mathcal{F}(k(\cdot;\mathbf{v}(l))) \cdot \mathcal{F}(\mathbf{h}(\cdot, l)))(\mathbf{x}) + \mathbf{c}(l)\right).$$

IFNO:

$$\mathbf{h}(\mathbf{x}, t+\Delta t) = \mathbf{h}(\mathbf{x}, t) + \Delta t \sigma\left(R\mathbf{h}(\mathbf{x}, t) + \mathcal{F}^{-1}(\mathcal{F}(k(\cdot;\mathbf{v})) \cdot \mathcal{F}(\mathbf{h}(\cdot, t)))(\mathbf{x}) + \mathbf{c}\right).$$



- **Features/Contributions:**
  1) An autonomous iterative system to reduce the memory allocation and overfitting issue.
  2) The resemblance with time-dependent nonlinear ODE to allow shallow-to-deep initialization technique and resolve vanishing gradient issues.

Implicit Fourier Neural Operator (IFNO)

IFNO:

$$\mathbf{h}(\mathbf{x},0)=\mathcal{P}(\mathbf{f})(\mathbf{x}):=P\mathbf{f}(\mathbf{x})+\mathbf{p}$$

$$\mathbf{h}(\mathbf{x},t+\Delta t)=\mathbf{h}(\mathbf{x},t)+\Delta t\sigma\left(R\mathbf{h}(\mathbf{x},t)+\mathcal{F}^{-1}(\mathcal{F}(k(\cdot;\mathbf{v}))\cdot\mathcal{F}(\mathbf{h}(\cdot,t)))(\mathbf{x})+\mathbf{c}\right).$$

$$\mathbf{u}(\mathbf{x})=\mathcal{Q}(\mathbf{h}(\cdot,T))(\mathbf{x}):=Q_2\sigma(Q_1\mathbf{h}(\mathbf{x},T)+\mathbf{q}_1)+\mathbf{q}_2$$

- **Assumption (Existence of a Fixed-Point Formulation):**
  Let $\mathbf{U}=[\mathbf{u}(\mathbf{x}_1),\mathbf{u}(\mathbf{x}_2),\dots,\mathbf{u}(\mathbf{x}_M)]$ and $\mathbf{F}=[\mathbf{f}(\mathbf{x}_1),\mathbf{f}(\mathbf{x}_2),\dots,\mathbf{f}(\mathbf{x}_M)]$, there exists a fixed point formulation, $\mathbf{U}^{l+1}=\mathbf{U}^l+\mathcal{R}(\mathbf{U}^l,\mathbf{F})$ for the target problem, such that *R* is a continuous function satisfying $||\mathcal{R}(\hat{\mathbf{U}},\mathbf{F})-\mathcal{R}(\tilde{\mathbf{U}},\mathbf{F})||_{l^2(\mathbb{R}^M)}\leq m||\hat{\mathbf{U}}-\tilde{\mathbf{U}}||_{l^2(\mathbb{R}^M)}$ for any two vectors $\hat{\mathbf{U}},\tilde{\mathbf{U}}\in\mathbb{R}^M$. Moreover, for any ε>0, there exist an integer L such that $||\mathbf{U}^l-\mathbf{U}^*||_{l^2(\mathbb{R}^M)}\leq\epsilon,\ \forall l>L$ for all possible input instances **F**.

- **Theorem (Universal Approximation):**
  Let U^* be the ground-truth solution of a modeling problem that satisfies the above assumption, then for any ε>0, there exist sufficiently large layer number L>0 and feature dimension number d>0, such that one can find a parameter set $\theta_\epsilon=\{P,\mathbf{p},Q_1,Q_2,\mathbf{q}_1,\mathbf{q}_2,\mathbf{C},V\}$, with the corresponding IFNO model satisfies

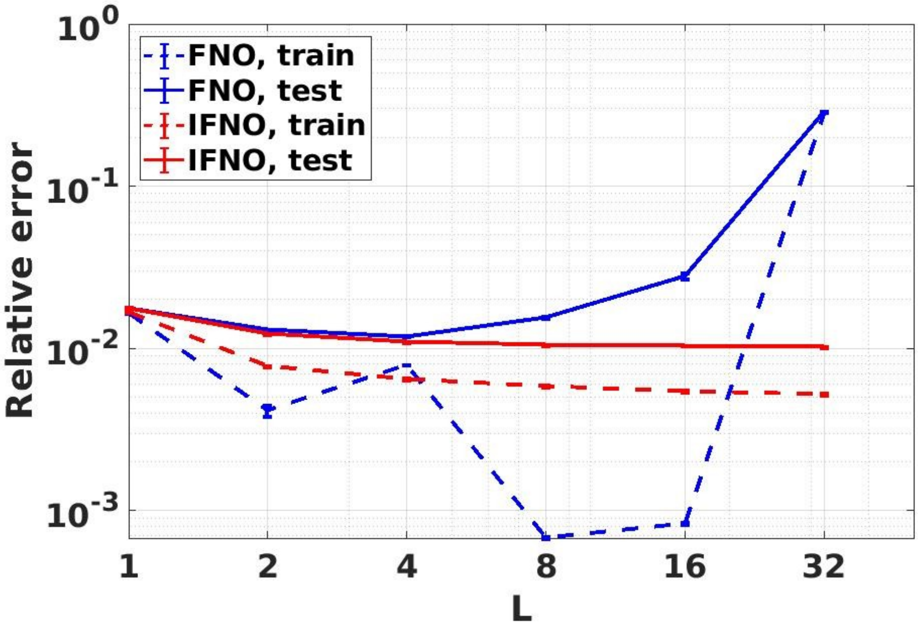
$$\left|\left|\mathcal{Q}\circ(\mathcal{L}^{IFNO})^L\circ\mathcal{P}([\mathbf{U}^0,\mathbf{F}]^{\mathrm{T}})-\mathbf{U}^*\right|\right|\leq\varepsilon,\quad\forall\mathbf{F}\in\mathbb{R}^M$$

Implicit Fourier Neural Operator (IFNO)

- Consider a 2D Darcy's equation

$$-\nabla \cdot (a(\mathbf{x})\nabla u(\mathbf{x})) = f(\mathbf{x}) \quad \text{on} \quad D = [0,1]^2$$

use FNO to construct a mapping from a to u, with 1000 pairs of $\{(a_j(x),u_j(x))\}$.



# of trainable parameters

FNO, L=1: 0.17M, 0.406 sec/epoch

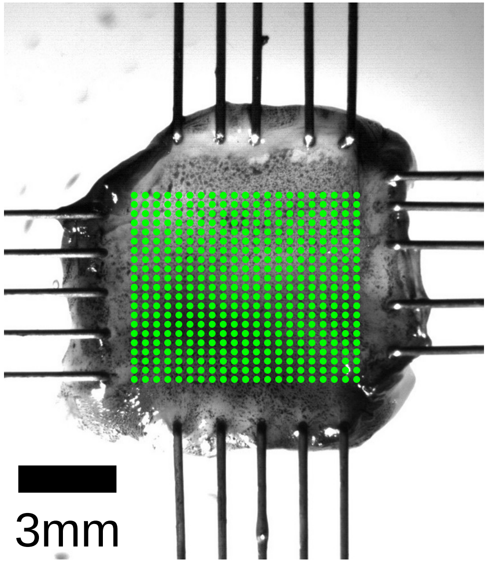
FNO, L=32: 5.35M, 5.694 sec/epoch

IFNO, L=1: 0.17M, 0.342 sec/epoch

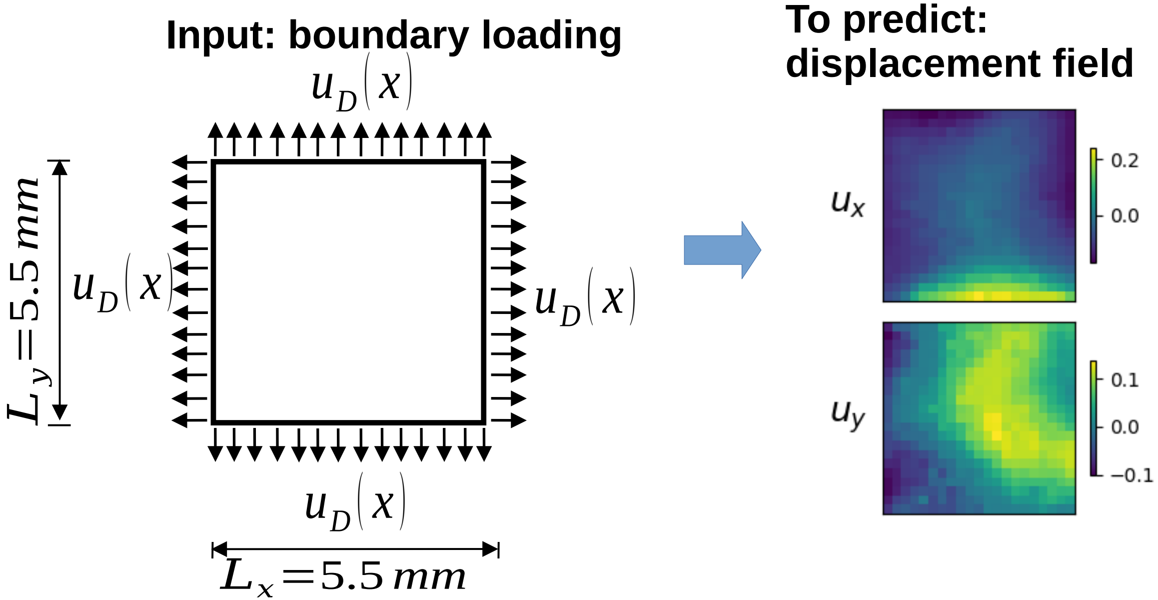
IFNO, L=32: 0.17M, 4.300 sec/epoch

Exemplar 3: Experiment Data (Heart Valve Leaflet)

- We consider the material response of heart valve leaflet, which is an anisotropic, highly heterogeneous and nonlinear material.
- 7 different Testing Protocol sets were performed, the displacement field is recorded via the DIC displacement tracking.
- For a fixed (unknown) microstructure $a(\mathbf{x})$, for each $\mathbf{b}(\mathbf{x}) = \mathbf{u}_{bc}(\mathbf{x})$, learn: $\mathcal{G}(\mathbf{b}) = \mathbf{u}(\mathbf{x})$.

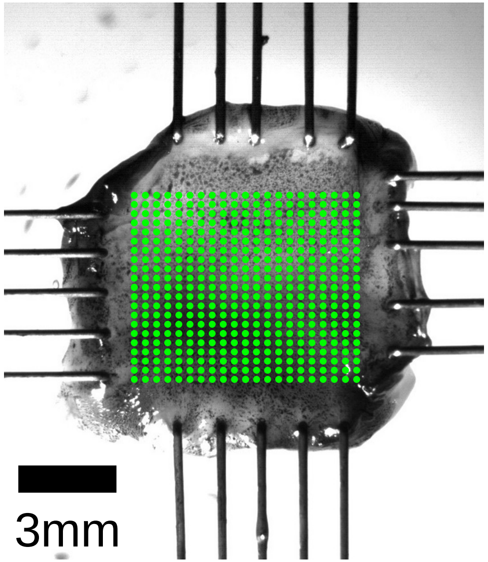


Thickness: 0.55 mm



Exemplar 3: Experiment Data (Heart Valve Leaflet)

- We consider the material response of heart valve leaflet, which is an anisotropic, highly heterogeneous and nonlinear material.
- 7 different Testing Protocol sets were performed, the displacement field is recorded via the DIC displacement tracking.
- For a fixed (unknown) microstructure $a(\mathbf{x})$, for each $\mathbf{b}(\mathbf{x}) = \mathbf{u}_{bc}(\mathbf{x})$, learn: $\mathcal{G}(\mathbf{b}) = \mathbf{u}(\mathbf{x})$.



3mm

Thickness: 0.55 mm

Experiment

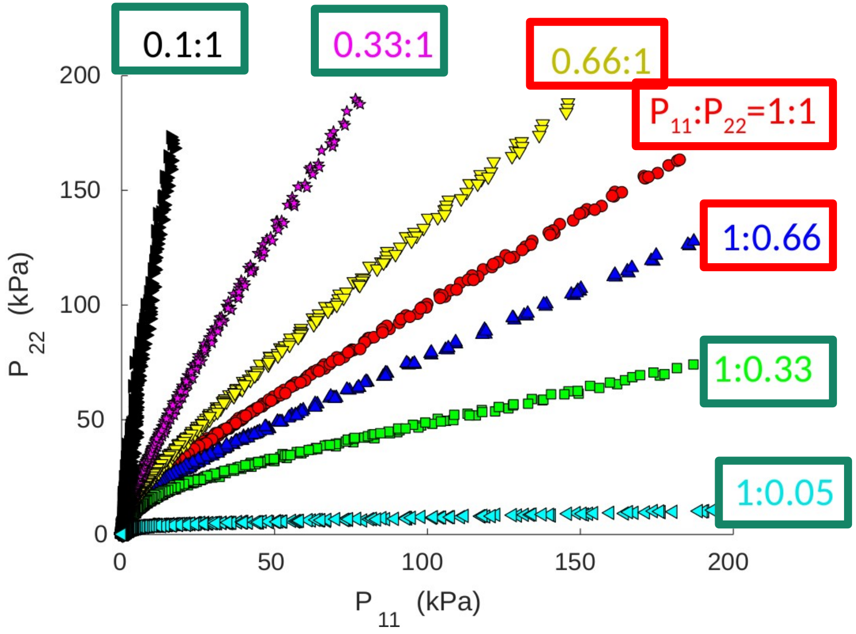
Prediction

In Distribution
IFNO training error: 1.54%
IFNO test error: **1.64%**
Fung model training error: 10.34%
Fung model test error: **10.83%**

Exemplar 3: Experiment Data (Heart Valve Leaflet)

- New challenge: generalizability when training and testing on different protocols.
- Generalizability to test samples chosen sufficiently far away from the training distribution is critical for safely deploying deep learning models in the real world.
- However, the **out-of-distribution prediction task** is generally challenging for machine learning models.

Protocol ID	Testing Protocol	Role
1	Biaxial Tension = 1:1	**Training set**
2	Biaxial Tension = 1:0.66	**Training set**
3	Biaxial Tension = 1:0.33	**Test set**
4	Biaxial Tension = 0.66:1	**Training set**
5	Biaxial Tension = 0.33:1	**Test set**
6	Constrained Uniaxial in x	**Test set**
7	Constrained Uniaxial in y	**Test set**

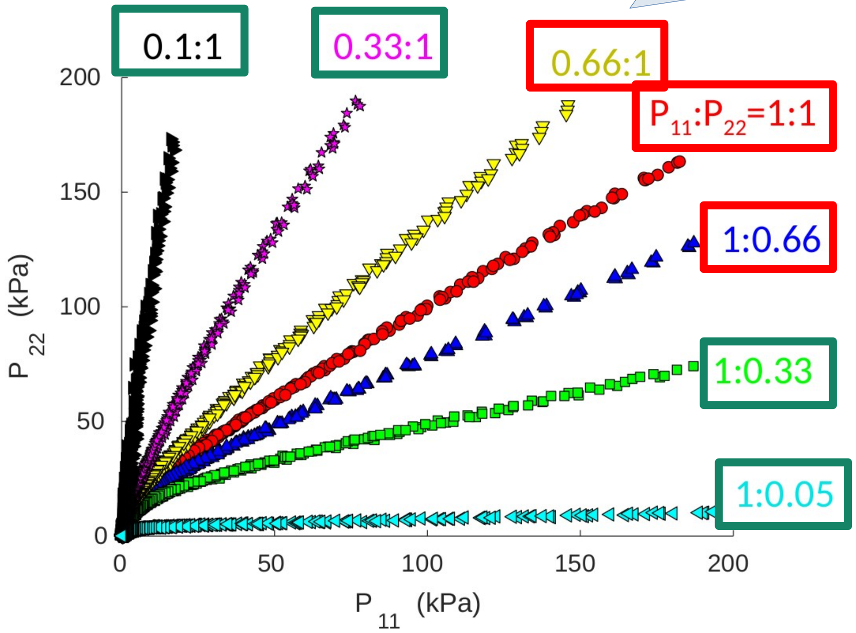


Exemplar 3: Experiment Data (Heart Valve Leaflet)

- New challenge: generalizability when training and testing on different protocols
- Generalizability to test samples chosen sufficiently far away from the training d for safely deploying deep learning models in the real world.
- However, the **out-of-distribution prediction task** is generally challenging for models.

Out of Distribution
IFNO training error: 1.53%
IFNO test error: **16.78%**
Fung model training error: 12.37%
Fung model test error: **16.80%**

Protocol ID	Testing Protocol	Role
1	Biaxial Tension = 1:1	**Training set**
2	Biaxial Tension = 1:0.66	**Training set**
3	Biaxial Tension = 1:0.33	**Test set**
4	Biaxial Tension = 0.66:1	**Training set**
5	Biaxial Tension = 0.33:1	**Test set**
6	Constrained Uniaxial in x	**Test set**
7	Constrained Uniaxial in y	**Test set**



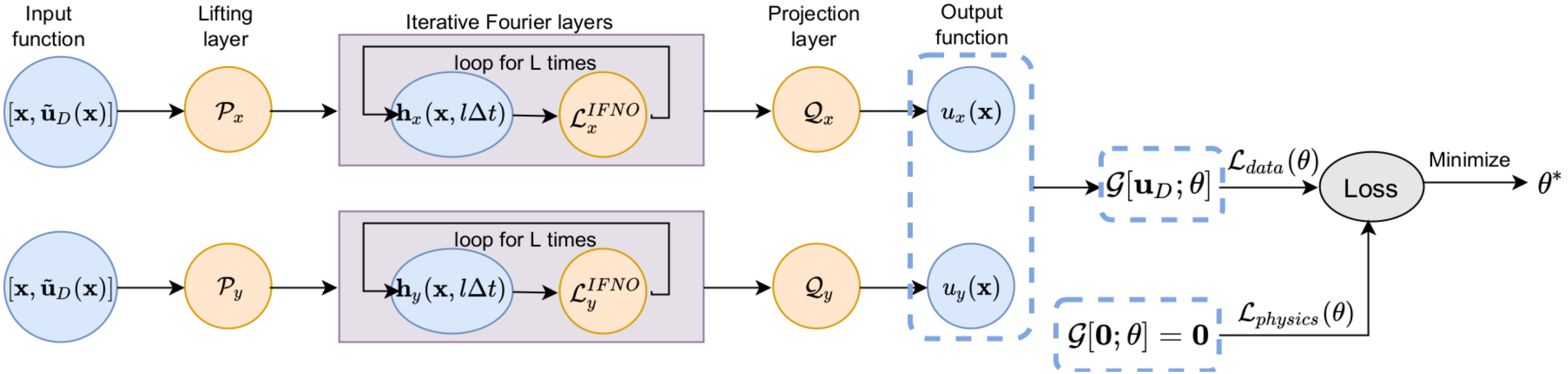
Physics-Guided IFNO

no-permanent-deformation

- **Idea 3: minimize the residual with infused physics knowledge:**

$$\min_{\theta \in \Theta} \sum_{i=1}^{N} \sum_{\mathbf{x}_j \in D} ||G((\mathbf{u}_D)_i;\theta)(\mathbf{x}_j) - \mathbf{u}_i(\mathbf{x}_j)||^2 \quad + \quad G(\mathbf{0};\theta)(\mathbf{x}) = \mathbf{0}$$

$$\Rightarrow \min_{\theta \in \Theta} \sum_{i=1}^{N} \sum_{\mathbf{x}_j \in D} ||G(\mathbf{b}_i;\theta)(\mathbf{x}_j) - \mathbf{u}_i(\mathbf{x}_j)||^2 + \lambda \sum_{\mathbf{x}_j \in D} ||G(0;\theta)(\mathbf{x}_j)||^2$$

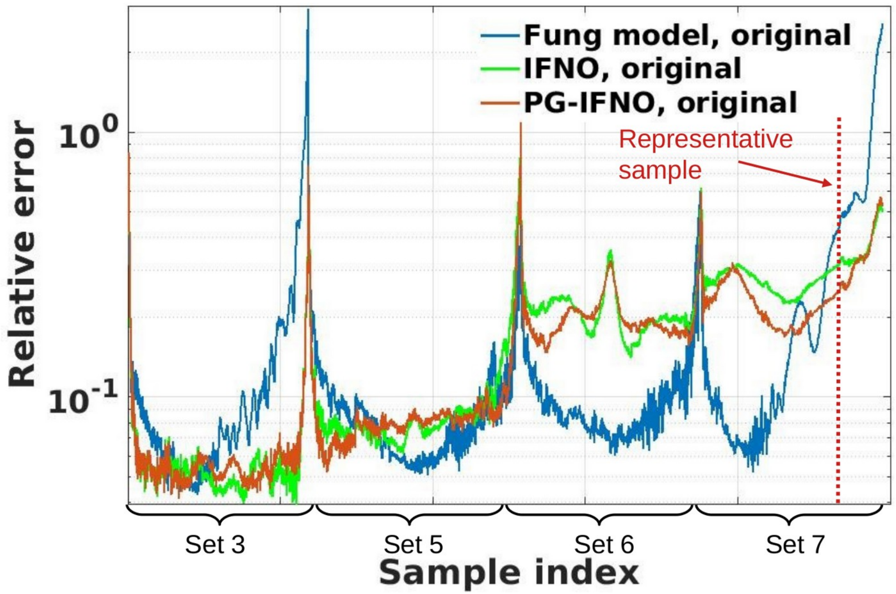


Exemplar 3: Experiment Data (Heart Valve Leaflet)

- New challenge: generalizability when training and testing on different protocols
- Generalizability to test samples chosen sufficiently far away from the training d… for safely deploying deep learning models in the real world.
- However, the **out-of-distribution prediction task** is generally challenging for ma… learning models.

Out of Distribution
IFNO test error: **16.78%**
Fung model test error: **16.80%**
PG-IFNO test error: **15.32%**

Protocol ID	Testing Protocol	Role
1	Biaxial Tension = 1:1	**Training set**
2	Biaxial Tension = 1:0.66	**Training set**
3	Biaxial Tension = 1:0.33	**Test set**
4	Biaxial Tension = 0.66:1	**Training set**
5	Biaxial Tension = 0.33:1	**Test set**
6	Constrained Uniaxial in x	**Test set**
7	Constrained Uniaxial in y	**Test set**



Conclusion

- We proposed two new nonlocal operator learning models, NORs and IFNOs, which learns continuous kernels for heterogeneous material learning tasks.
- For **homogenized model learning tasks**, the **nonlocal operator regression (NOR) model** is proposed, which learns optimal kernel functions directly from data.
- For **heterogeneous material modeling tasks**, the **implicit Fourier neural operator (IFNO) model** is proposed, which naturally embeds the material micromechanical properties and defects in the integrand.
- We employed NOR and IFNO to learn three exemplar material models directly from high-fidelity simulations/experimental measurements, and show that the learnt nonlocal operators outperform conventional constitutive models in predicting complex material responses.

Acknowledgment

- **Co-authors:**

  Huaiqian You (Ph.D. student), Quinn Zhang (undergraduate student), *Lehigh University*

  Colton Ross, Chung-Hao Lee, *University of Oklahoma*

  Fei Lu, Qingci An, *Johns Hopkins University,*

  Stewart Silling, *Sandia National Lab,* Marta D'Elia, *Meta*

- **Funding support:**

  NSF CAREER award DMS1753031

  AFOSR YIP grant FA9550-22-1-0197

- **Computational Resources:** Lehigh HPC systems

- **References:**

[1] H. You, Q. Zhang, C. Ross, C-H. Lee, Y. Yu*, "Learning Deep Implicit Fourier Neural Operators (IFNOs) with Applications to Heterogeneous Material Modeling". CMAME, 2022.

[2] H. You, Q. Zhang, C. Ross, C-H. Lee, M-C. Hsu, Y. Yu*, "A Physics-Guided Neural Operator Learning Approach to Model Biological Tissues from Digital Image Correlation Measurements". arXiv preprint arXiv:2204.00205. (code: https://github.com/fishmoon1234/IFNO-tissue)

[3] F. Lu*, Q. An, Y. Yu, "Nonparametric learning of kernels in nonlocal operators". arXiv preprint arXiv:2205.11006.