MgNO: Efficient Parameterization of Linear Operators via Multigrid

Juncai He, Xinliang Liu and Jinchao Xu

Applied Mathematics and Computational Sciences (AMCS)
The King Abdullah University of Science and Technology (KAUST)

General Setting of Operator Learning

We consider pairs of functions $(\boldsymbol{u}_j, \boldsymbol{v}_j)_{j=1}^N$, where $\boldsymbol{u}_j$ is drawn from a probability measure μ and $\boldsymbol{v}_j = \mathcal{G}(\boldsymbol{u}_j)$. Given the data $(\boldsymbol{u}_j, \boldsymbol{v}_j)_{j=1}^N$, we approximate $\mathcal{G}$, by solving the network parameter set θ via an optimization problem:

$$\min_{\theta \in \Theta} \mathcal{L}(\theta) := \min_{\theta \in \Theta} \frac{1}{N} \sum_{j=1}^{N} \left[\|\widetilde{\mathcal{G}}_\theta(\boldsymbol{u}_j) - \boldsymbol{v}_j\|^2 \right]. \tag{1}$$

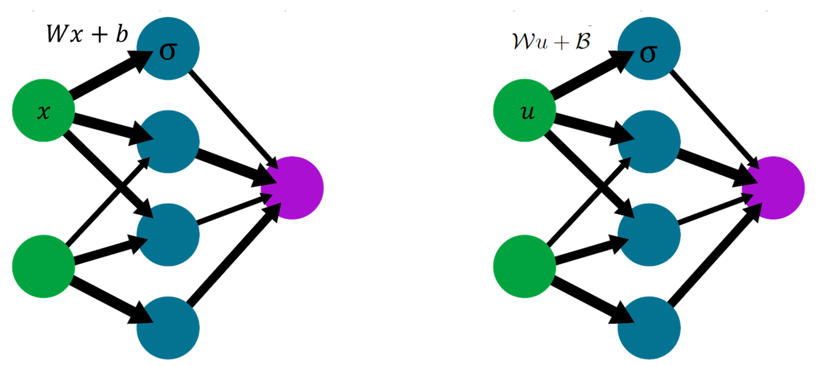
General Neural Operator

To begin, we define the following shallow neural operators with n neurons for operators from $\mathcal{X}$ to $\mathcal{Y}$ as

$$\mathcal{O}(u) = \sum_{i=1}^{n} \mathcal{A}_i \sigma(\mathcal{W}_i u + \mathcal{B}_i) \quad \forall u \in \mathcal{X} \tag{2}$$

where $\mathcal{W}_i \in \mathcal{L}(\mathcal{X}, \mathcal{Y}), \mathcal{B}_i \in \mathcal{Y}$, and $\mathcal{A}_i \in \mathcal{L}(\mathcal{Y}, \mathcal{Y})$. Here, $\mathcal{L}(\mathcal{X}, \mathcal{Y})$ denotes the set of all bounded (continuous) linear operators between $\mathcal{X}$ and $\mathcal{Y}$, and $\sigma : \mapsto \mathbb{R}$ defines the nonlinear point-wise activation.

Analogy between Neural Network and Neural Operator



$$\begin{cases} h^0 = u \in \mathcal{X} \\ h^\ell(u) = \sigma\left(\mathcal{W}^\ell h^{\ell-1}(u) + \mathcal{B}^\ell\right) \in \mathcal{Y}^{n_\ell} \quad \ell = 1 : L \\ \mathcal{O}(u) = \mathcal{W}^{L+1} h^L(u) \in \mathcal{Y} \end{cases}$$

- $\mathcal{X} = [0,1]^d, \mathcal{Y} = \mathbb{R}^k \Rightarrow \left[\mathcal{W}^\ell\right]_{ij} \in \mathbb{R} \Rightarrow$ DNNs for function approximation.
- $\mathcal{X} = [0,1]^{d \times d \times c}, \mathcal{Y} = \mathbb{R}^k \Rightarrow \left[\mathcal{W}^\ell\right]_{ij}$ is convolution $\Rightarrow$ CNNs for image tasks.
- $\mathcal{X}$ = function space, $\mathcal{Y}$ = function space $\Rightarrow \left[\mathcal{W}^\ell\right]_{ij} \in \mathcal{L}(\mathcal{Y}, \mathcal{Y}) \Rightarrow$ neural operators

Furthermore, we have the following universal approximation theorem based on this unified definition of shallow networks.

Theorem (Universal Approximation)

Let $\mathcal{X} = H^s(\Omega)$ and $\mathcal{Y} = H^{s'}(\Omega)$ for some $s, s' \geq 1$, and $\sigma \in C(\mathbb{R})$ is non-polynomial, for any continuous operator $\mathcal{O}^* : \mathcal{X} \mapsto \mathcal{Y}$, compact set $\mathcal{C} \subset \mathcal{X}$ and $\epsilon > 0$, there is n such that

$$\inf_{\mathcal{O} \in \Xi_n} \sup_{\boldsymbol{u} \in \mathcal{C}} \|\mathcal{O}^*(\boldsymbol{u}) - \mathcal{O}(\boldsymbol{u})\|_{\mathcal{Y}} \leq \epsilon, \tag{3}$$

where Ξ_n denote the shallow networks defined in equation 2 with n neurons.

Revisit $\mathcal{W}^\ell$ as Solving PDEs

Consider the elliptic PDEs given by $\mathcal{L}u(x) = f(x)$ defined over the domain $\Omega = (0,1)^2$ and subject to boundary conditions. Employing a linear finite element method (FEM) discretization with a mesh size defined as $h = \frac{1}{d}$, the discretized system can be expressed as:

$$A * u = f, \tag{4}$$

where $u, f \in \mathbb{R}^{d \times d}$.

- $A*u$ represents the standard convolutional operation for a single channel, with specific padding schemes determined by the boundary conditions;
- A, of dimensions 3×3, is dictated by the elliptic operator $\mathcal{L}$ in conjunction with the linear FEM.

Consequently, the inverse operation of $A*$ corresponds to the discrete Green's function associated with $\mathcal{L}$ under a linear FEM framework. Solving equation 4 can be precisely represented as a conventional convolution neural network.

Parametrization of $\mathcal{W}^\ell$ and MgNO

We provide a concise overview of the essential components of multigrid structure in the language of convolution as an operator mapping from f to u:

1. **Input (f) and Initialization**: Set $f^1 = f$ and initialize with $u^{1,0} = 0$.
2. **Iteration (Smoothing) Process**: The algorithm iteratively refines u based on the relation:
$$u^{\ell,i} = u^{\ell,i-1} + B^{\ell,i} * \left(f^\ell - A^\ell * u^{\ell,i-1}\right), \tag{5}$$
where $\ell = 1 : J$ and $i \leq \nu_\ell$.
3. **Hierarchical Structure via Restriction and Prolongation**: The superscript ℓ denotes the specific hierarchical level or grid within the multigrid structure. Specifically, using the residual, we restrict the input f^ℓ and the current state u^ℓ to a coarser level through convolution with a stride of 2:
$$f^{\ell+1} = R_\ell^{\ell+1} *_2 \left(f^\ell - A^\ell * u^\ell\right) \in \mathbb{R}^{d_{\ell+1} \times d_{\ell+1} \times n}, \quad u^{\ell+1,0} = 0. \tag{6}$$
Subsequently, we apply the smoothing iteration as in equation 5 to derive the correction from the coarser level. The correction is then prolonged from the coarser to the finer level using a de-convolution operation $P_{\ell+1}^\ell$ with a stride of 2 (acting as the transpose of the restriction operation).

Approximation Property

Let's represent the linear operator defined by the aforementioned V-cycle multigrid operator as $\mathcal{W}_{Mg}$. The convergence result presented in [2].

$$\|u - \mathcal{W}_{Mg}(A * u)\|_A \leq \left(1 - \frac{1}{c}\right) \|u\|_A, \tag{7}$$

demonstrates the uniform approximation capabilities of $\mathcal{W}_{Mg}$ relative to the inverse of $A*$, which corresponds to the Green's function associated with the elliptic operator $\mathcal{L}$. Here, c is a constant that is independent of the mesh size h, and $\|u\|_A^2 = (u, A * u)_{L^2(\Omega)}$ denotes the energy norm.

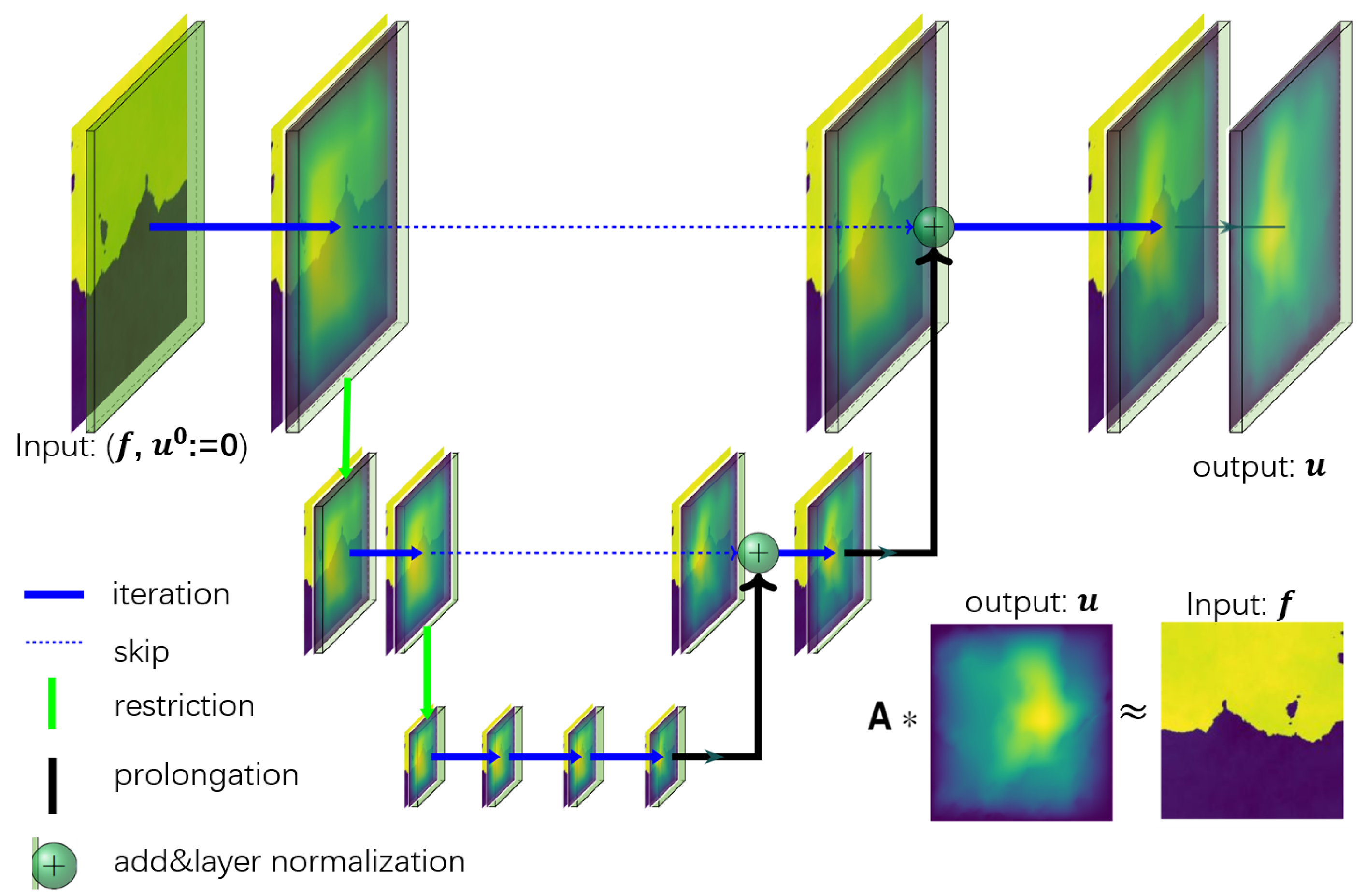


Figure 1. Overview of $\mathcal{W}_{Mg}$ using a multi-channel V-cycle multigrid framework MgNet [1].

Key features of MgNO

1. Commonly used **lifting and projecting operators** in traditional constructs **are unnecessary in MgNO**.
2. Our method establishes its superiority **not only in prediction accuracy but also in efficiency**, both in terms of **parameter count and runtime** on several PDEs, including Darcy, Helmholtz, and Navier-Stokes equations, with different boundary conditions.
3. Given the inherent ties between convolutions and multigrid methods in PDE contexts, MgNO **naturally accommodates** various boundary conditions.
4. Interesting observation: **No gap between training and test errors**, as shown in **Figure 3**.

Numerical Results

The Darcy equation writes

$$\begin{cases} -\nabla \cdot (a(x) \nabla u(x)) = f(x) & x \in D \\ u(x) = 0 & x \in \partial D \end{cases} \tag{8}$$

where the coefficient $0 < a_{\min} \leq a(x) \leq a_{\max}, \forall x \in D$, and the forcing term $f \in H^{-1}(D; \mathbb{R})$. The coefficient to solution map is $\mathcal{S} : L^\infty(D; \mathbb{R}_+) \to H_0^1(D; \mathbb{R})$, such that $u = \mathcal{S}(a)$ is the target operator.

Table 1. Performance comparison for Darcy benchmarks. Performance are measured with relative L^2 errors ($\times 10^{-2}$) and relative H^1 errors ($\times 10^{-2}$).

			Darcy smooth		Darcy rough		Darcy multiscale	
Model	time (s/iter)	params (m)	L^2	H^1	L^2	H^1	L^2	H^1
FNO2D	7.4	2.37	0.684	2.583	1.613	7.516	1.800	9.619
DilResNet	14.9	1.04	4.104	5.815	7.347	12.44	1.417	3.528
UNet	9.1	17.27	2.169	4.885	3.519	5.795	1.425	5.012
U-NO	11.4	16.39	0.492	1.276	1.023	3.784	1.187	5.380
MWT	21.7	9.80	–	–	1.138	4.107	1.021	7.245
GT	38.2	2.22	0.945	3.365	1.790	6.269	1.052	8.207
MgNO	**6.6**	**0.57**	**0.176**	**0.576**	**0.339**	**1.380**	**0.715**	**1.756**

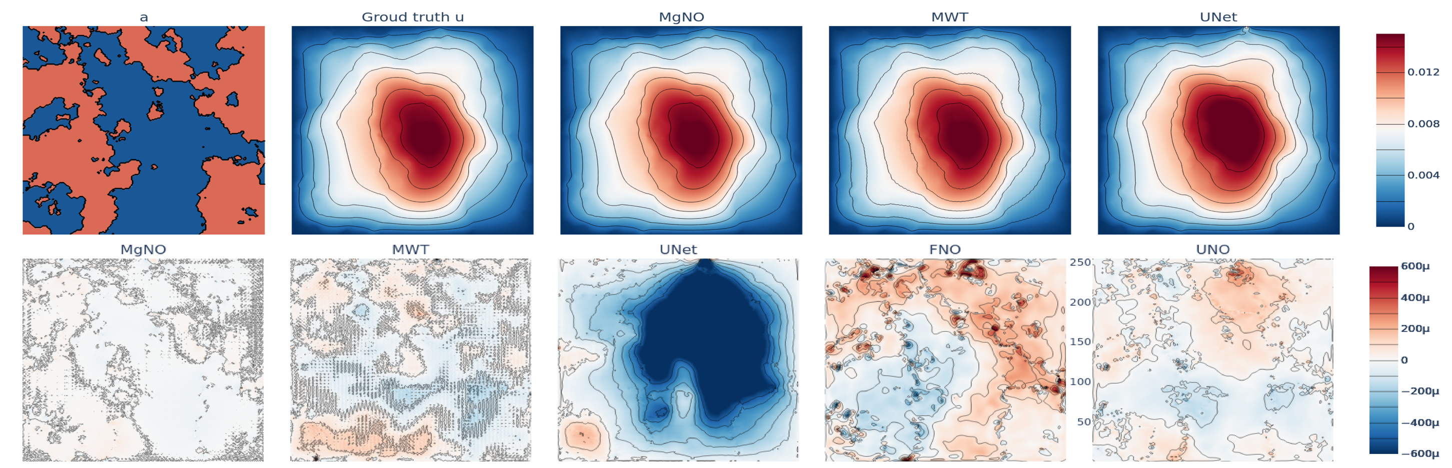


Figure 2. Qualitative comparisons on Darcy rough benchmark. Top: coefficient a, ground truth u, and predictions; bottom: the corresponding prediction error map for each model in the same color scale.

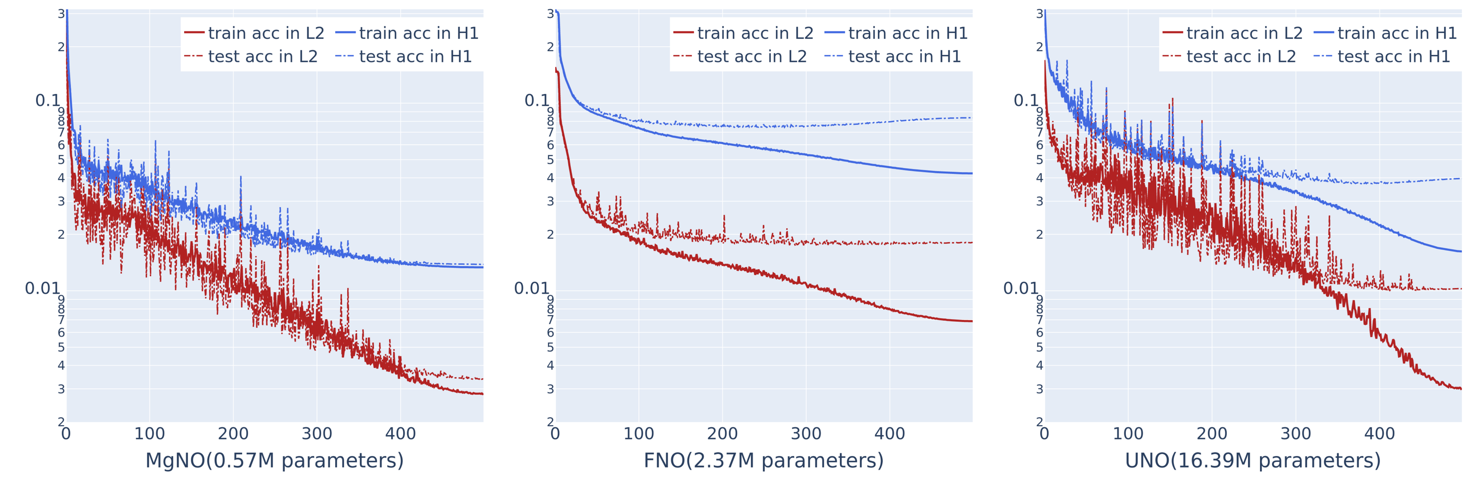


Figure 3. Comparison of generalization error between MgNO, FNO and UNO.

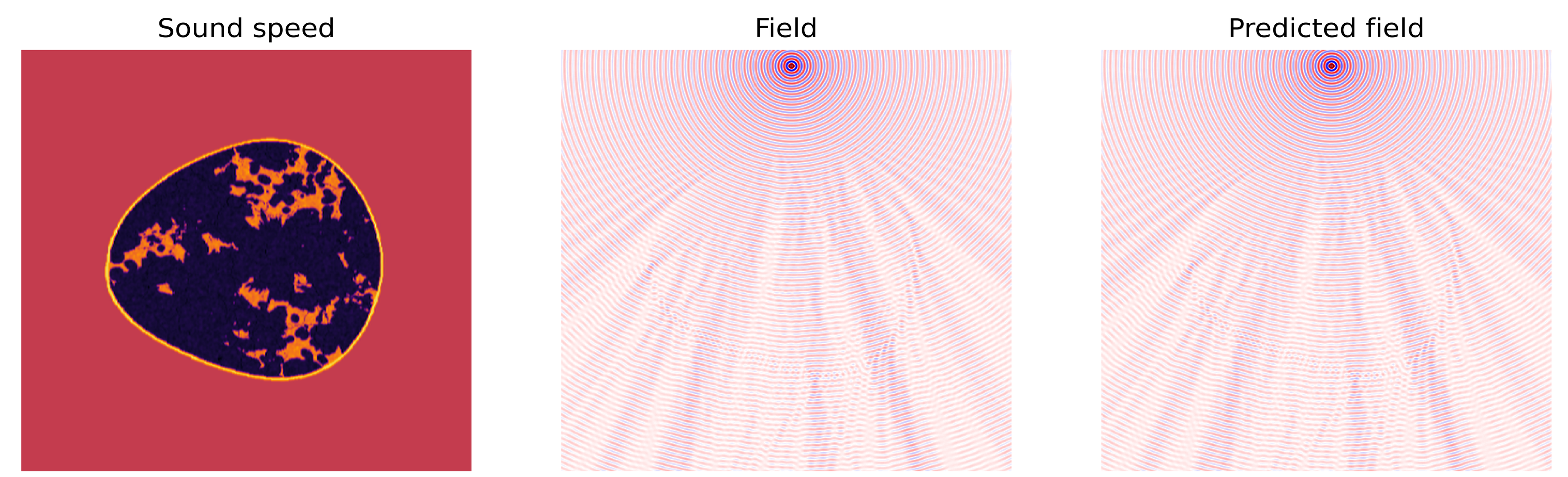


Figure 4. Wavefield Prediction for Ultrasound Breast Examination. The neural operator maps the sound speed $c(x)$ and point source f to the wave field solution u such that $-\Delta u - \frac{\omega^2}{c^2(x)} u = f$ with absorbing boundary condition. New results from this study will be detailed in an upcoming manuscript. Stay tuned for more information! Data utilized in this study is credited to Bohrium.

References

[1] Juncai He and Jinchao Xu.
Mgnet: A unified framework of multigrid and convolutional neural network.
Science china mathematics, 62:1331–1354, 2019.

[2] Jinchao Xu and Ludmil Zikatanov.
The method of alternating projections and the method of subspace corrections in hilbert space.
Journal of the American Mathematical Society, 15(3):573–597, 2002.

(juncai.he; xinliang.liu; jinchao.xu)@kaust.edu.sa