MGCFNN: A Neural MultiGrid Solver with Novel Fourier Neural Network for **High WaveNumber Helmholtz Equations**

Yan Xie, Minrui Lv & Chen-song Zhang

LSEC, Academy of Mathematics and Systems Science, CAS

ICLR 2025

Introduction

Helmholtz equation

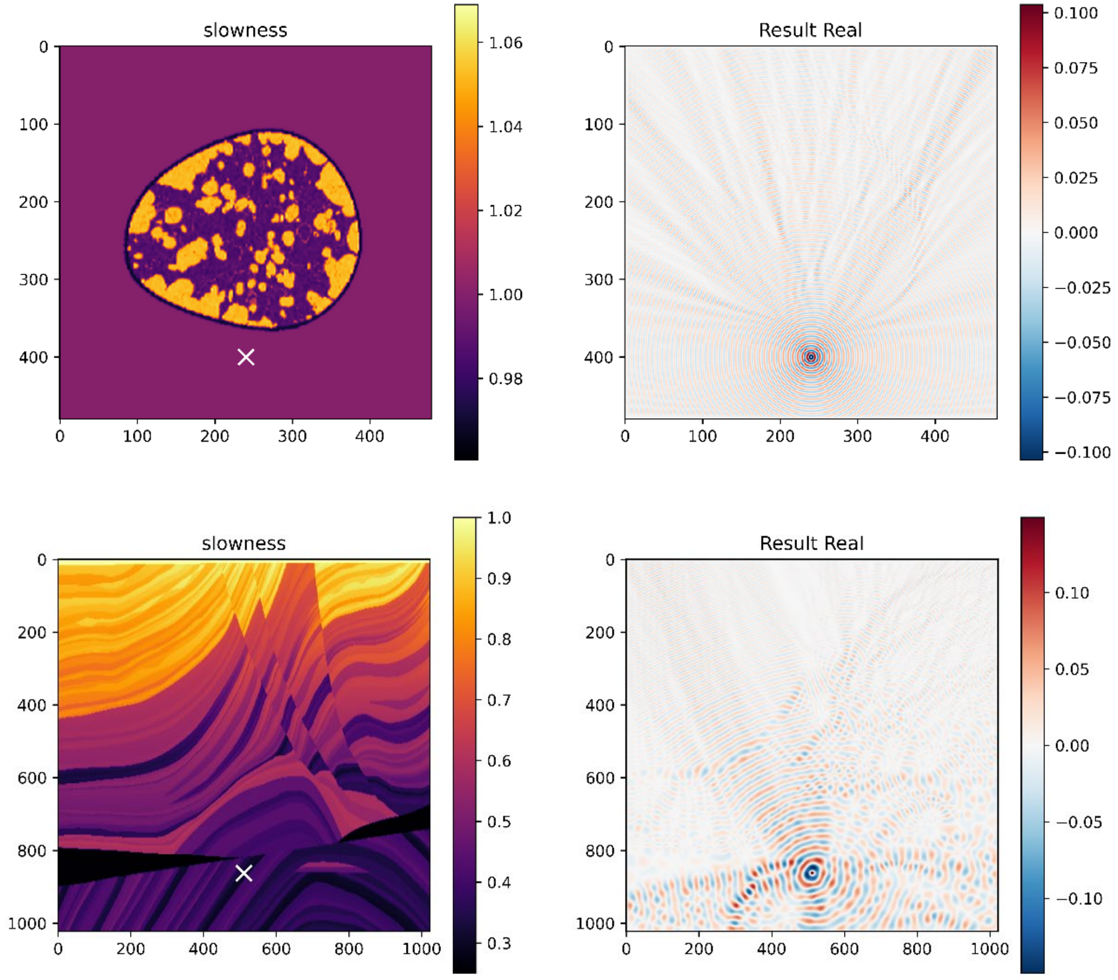
$$-\Delta u(x) - k^2(1-\gamma i)u(x) = f(x), x \in [0,1]^2$$

where $k = \omega\kappa(x) = \frac{\omega}{c(x)}$ is the wavenumber, κ/c is the slowness/speed field

Notoriously hard to solve [1]

- ☐ High wavenumber
- ☐ Heterogeneous media
- ☐ No damping

- ■ Large grid size, a thumb of rule:
  $\omega\kappa h \le \frac{2\pi}{10}$, 10 points in one wavelength
- ■ Complex propagation pattern
  reflection, refraction, diffraction…
- ■ Slow decay rate (polynomial)

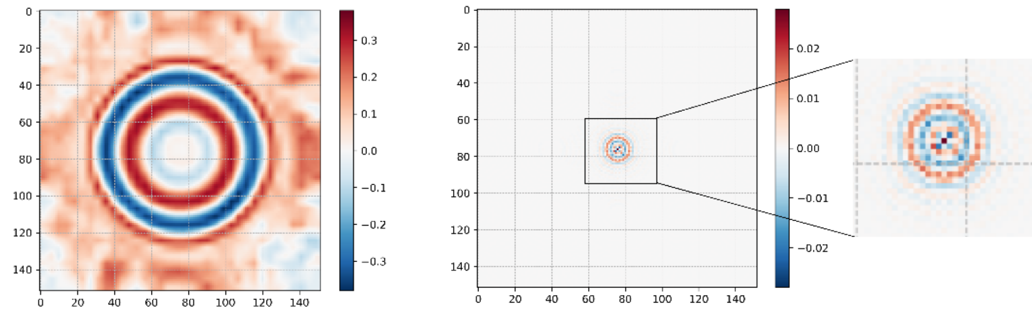


[1] Ernst O G, Gander M J. Why it is difficult to solve Helmholtz problems with classical iterative methods. *Numerical analysis of multiscale problems*, 2011: 325-363.

Fourier Neural Network (FNN)

Use Fourier transform to implement CNN with extended kernel size (see Fig.(b))

$$\mathcal{F}(f * g) = \mathcal{F}(f)\mathcal{F}(g)$$

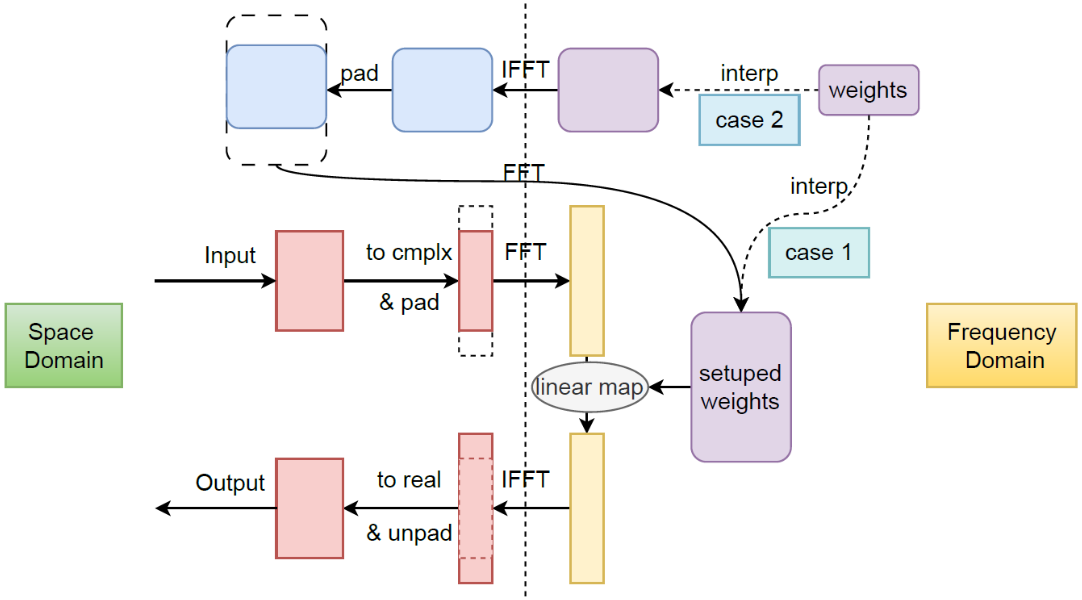


(a) FNN weights in frequency domain; (b) Equivalent kernel in space domain

Modify FNO[2]'s spectral convolution

- Use full modes
- Higher frequency with larger gird size

- Interpolate weights from small size to target size (smoothness in Fig.(a))
- Gain kernel in space domain, then pad and Fourier transform back



Notes:
- Kernel is like a damped wave
- FNN in solve phase is like Born series method[3]

[2] Li, Zongyi, et al. "Fourier Neural Operator for Parametric Partial Differential Equations." *International Conference on Learning Representations, 2021*.
[3] Osnabrugge, G., et al. A convergent Born series for solving the inhomogeneous Helmholtz equation in arbitrarily large media. *Journal of computational physics*, 322 (2016): 113-124.

Multigrid Architecture

Basically follow the multigrid architecture of our previous work MGCNN[4]

Setup phase

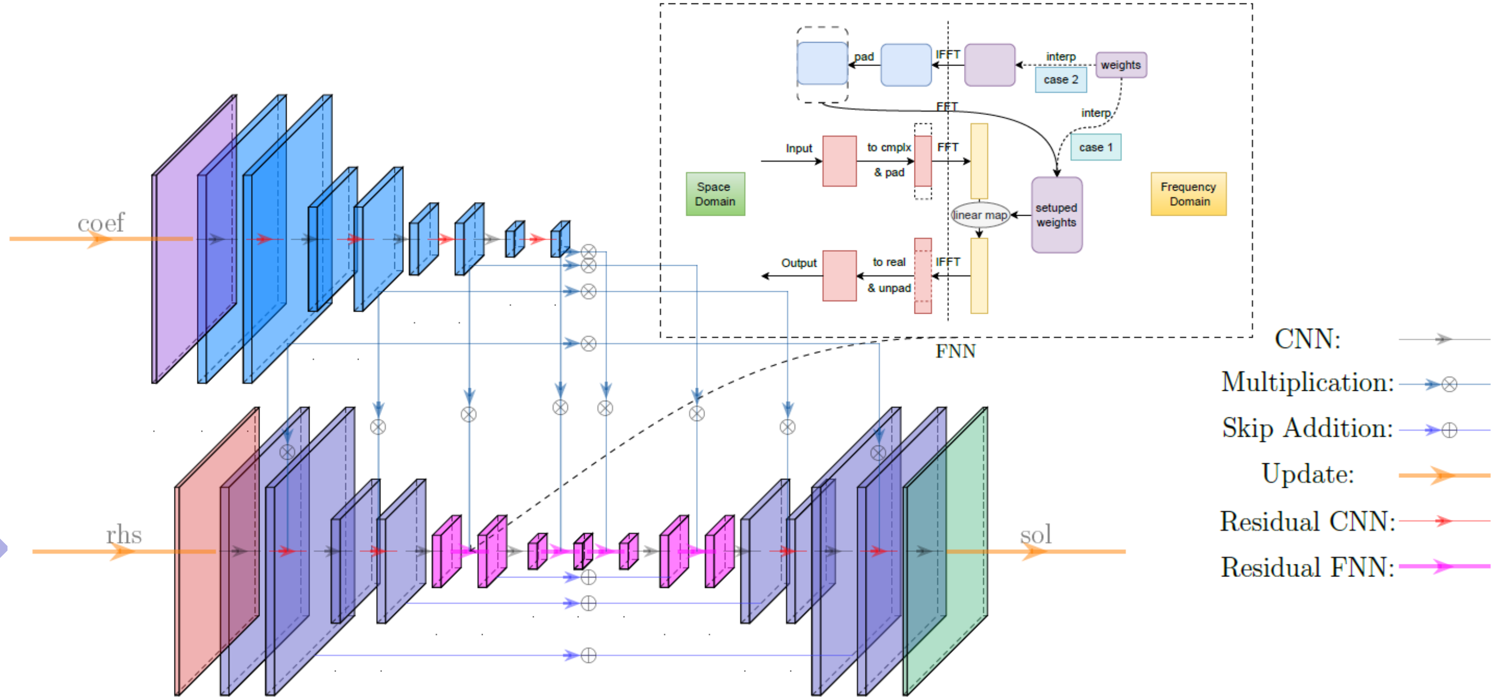
- Process coefficient nonlinearly
- Prepare $setupout_l$ for every level l

Solve phase

- Process source with $setupout_l$

$$x_l = Net(setupout_l \cdot x_l) + x_l$$

- Linear operation



Remark2:

- Multigrid levels Fixed for higher frequency in scalability test
- Double number of sweeps (layers) in coarser level

Remark1 (naming convention):

- MGCFNN: use FNN on some coarse levels of solve phase
- MGFNN: use FNN on all levels of solve phase

[4] Xie, Yan, et al. MGCNN: a learnable multigrid solver for sparse linear systems from PDEs on structured grids. *arXiv:2312.11093* (2023).

Compare with neural operator methods

- Supervised loss

$$L_{err} = \frac{1}{N}\sum_{i=1}^{N} ||u_i - sol_i||_2^2$$

- A single inference
- Separately evaluate models for

$$\kappa_{min} = 0.75, 0.5, 0.25$$

- Best respect of train error, test error and train time
- FNO, MWT and UNO have obvious early overfitting
- A single inference is not enough to reach low error

Table 1: Supervised learning on grid 256×256 with $\omega = 80\pi$ and $\kappa_{\min} = 0.75$. We record information *Where the Best Test Error* reaches.

	$\kappa_{\min} = 0.75$					
		Where Best Test Error				
Model	Train Error	Test Error	Train Error	Epoch	Train Time(s/epoch)	Parameters(MB)
MGFNN	**0.035**	**0.061**	**0.035**	120	93.2	8.9
MGCFNN	0.046	0.070	0.046	120	**67.5**	5.3
MGNO	0.063	0.079	0.063	120	110.3	4.6
FNO2D	0.085	0.561	0.496	4	103.6	46.1
MWT2D	0.119	0.527	0.475	4	147.0	26.0
U-NO	0.408	0.880	0.870	4	101.6	86.7
U-NET	0.534	0.803	0.758	31	89.9	31.0
DIL-RESNET	0.605	0.606	0.605	116	140.0	**0.6**
LSM	0.722	0.783	0.739	66	230.7	4.9

Table 2: Supervised learning on grid 256×256 with $\omega = 80\pi$ and $\kappa_{\min} = 0.50, 0.25$. We record information *Where the Best Test Error* reaches.

	$\kappa_{\min} = 0.50$				$\kappa_{\min} = 0.25$			
		Where Best Test Error				Where Best Test Error		
Model	Train Error	Test Error	Train Error	Epoch	Train Error	Test Error	Train Error	Epoch
MGFNN	**0.075**	**0.187**	0.122	25	**0.126**	**0.431**	**0.253**	18
MGCFNN	0.101	0.206	**0.104**	88	0.190	0.458	0.396	7
MGNO	0.209	0.333	0.221	81	0.361	0.633	0.614	14
FNO2D	0.122	0.749	0.718	3	0.154	0.851	0.818	3
MWT2D	0.169	0.728	0.669	4	0.186	0.845	0.821	3
U-NO	0.389	0.871	0.864	3	0.362	0.935	0.930	4
U-NET	0.634	0.774	0.716	57	0.575	0.806	0.778	26
DIL-RESNET	0.627	0.629	0.627	117	0.699	0.702	0.699	119
LSM	0.852	0.860	0.852	117	0.867	0.878	0.874	38

Iterative Solving

Table 3: Scalability comparison with other solvers.

rtol=1E-7		$\gamma = \mathbf{0.01}$ (RTX3090) Lerer et al. (2024)			$\gamma = \mathbf{0.0}$ (A100 80G) Cui et al. (2024)			$\gamma = \mathbf{0.0}$ (CPU)	
		time (s) & iters			time (s) & iters			time(s)	
ω	grid	MGCFNN	ENCODER-SOLVER	speedup	MGCFNN	WAVE-ADR-NS	speedup	CHOLMOD	speedup
80π	511×511	0.12(9)	0.65(43)	5.5	0.16(14)	15.07(28)	94.8	0.49	3.1
160π	1023×1023	0.19(11)	1.29(68)	6.8	0.30(22)	34.98(54)	116.2	8.88	29.5
320π	2047×2047	0.58(14)	3.40(85)	5.8	1.15(40)	91.63(122)	79.6	38.91	33.8
640π	4095×4095	2.77(18)	13.34(117)	4.8	8.55(83)	286.14(247)	33.5	183.61	21.5

- Unsupervised loss

$$L_{res} = \frac{1}{N}\sum_{i=1}^{N} ||rhs_i - A_i sol_i||_2^2$$

- Reach low relative residual 1E-7

Outperform other

✓ Pure AI solver: encoder-solver [5]

✓ AI-enhanced traditional solver: Wave-ADR-NS [6]

✓ Sparse direct solver: CHOLMOD [7]

[5] Bar Lerer, et al. Multigrid-augmented deep learning preconditioners for the Helmholtz equation using compact implicit layers. *SIAM Journal on Scientific Computing*, pp. S123–S144, 2024.
[6] Chen Cui, et al. A neural multigrid solver for Helmholtz equations with high wavenumber and heterogeneous media. arXiv:2404.02493, 2024.
[7] Timothy A Davis and William W Hager. Dynamic supernodes in sparse Cholesky update/downdate and triangular solves. ACM Transactions on Mathematical Software (TOMS), 35(4):1–23, 2009.

Ablation Study

MGCFNN combine the complementary features of FNN and multigrid architecture

MGCFNN is the SOTA choice

- No need using FNN at all levels
  - MGFNN only reduce few iters and need more time
- Better convergence than MGCNN
  - MGCNN rely on GMRES to reduce iters significantly

Table 4: Model architecture ablation study.

rtol=1E-7	grid size 256×256, $\omega = 80\pi$						grid size 480×480, $\omega = 150\pi$					
	standalone			GMRES			standalone			GMRES		
	MGCFNN	MGFNN	FNN	MGCFNN	MGFNN	FNN	MGCFNN	MGFNN	FNN	MGCFNN	MGFNN	FNN
iters	22	19	55	15	13	24	12	9	54	11	10	22
time(s)	0.136	0.135	0.270	**0.111**	0.116	0.151	**0.093**	0.101	0.724	0.105	0.124	0.365

Table 5: Scalability comparison between MGCFNN and MGCNN.

rtol=1E-7		standalone				GMRES			
		MGCFNN		MGCNN		MGCFNN		MGCNN	
ω	size	iters	time(s)	iters	time(s)	iters	time(s)	iters	time(s)
80π	511×511	14	0.17	35	0.27	12	0.17	21	0.20
160π	1023×1023	22	0.31	61	0.55	20	0.31	36	0.37
320π	2047×2047	41	1.18	115	2.33	35	1.13	71	1.65
640π	4095×4095	83	8.18	231	18.66	72	8.48	146	14.56

Thanks for Your Listening

xieyan2021@lsec.cc.ac.cn

https://gitee.com/xiehuohuo77/mgcfnn/