
Benchmarking of neural operators for rigid-body fluid-structure interaction

Weiheng Zhong

University of Illinois at Urbana-Champaign
Champaign, Illinois
weiheng4@illinois.edu

Hadi Meidani

University of Illinois at Urbana-Champaign
Champaign, Illinois
meidani@illinois.edu

Abstract

Simulating rigid-body fluid–structure interaction (FSI) with both accuracy and efficiency remains a persistent challenge in computational science and engineering. While neural operators have recently shown great promise as alternatives to traditional numerical solvers in computational fluid dynamics, their potential in FSI problems—especially those involving freely moving rigid bodies—has not been thoroughly investigated. To address this gap, we present two benchmark datasets tailored for FSI scenarios where rigid-body motion acts as a control signal shaping the surrounding fluid flow. We further conduct a systematic evaluation of state-of-the-art neural operator architectures and explore three strategies for coupling structural motion with fluid initial conditions. A subset of the dataset is currently stored in Harvard Dataverse and the codes are public in Github.

1 Introduction

Simulation of rigid-body fluid-structure interaction (FSI) Kolahdouz et al. [2021] involves explaining how a rigid body’s motion influences the surrounding fluid flow. This simulation is critical to a wide range of applications in aerospace and mechanical systems, as well as civil infrastructure systems Ardekani [2019]. Conventional numerical methods, such as the finite element method (FEM) Bathe [2007] and finite volume method (FVM) Eymard et al. [2000], provide high-fidelity solutions but are computationally expensive, especially for high-dimensional, time-dependent problems. Particle-based methods Negishi et al. [2024] have also been employed to accelerate fluid simulations, although at the cost of some accuracy.

Recently, neural operators Li et al. [2020a], Lu et al. [2021] have emerged as a promising solution approach for computational fluid dynamics (CFD), addressing the computational inefficiencies of conventional solvers. Unlike standard machine learning models, neural operators learn mappings between functional spaces, enabling rapid solution of partial differential equation (PDE) across varying input geometries, initial and boundary conditions, and fluid properties. Architectures such as the Deep Operator Network (DeepONet) Lu et al. [2021], Fourier Neural Operator (FNO) Li et al. [2020a], Wavelet Neural Operator (WNO) Tripura and Chakraborty [2022], Graph Neural Operator (GNO) Li et al. [2020b], and General Neural Operator Transformer (GNOT) Hao et al. [2023] have been successfully used in applications such as turbulence modeling and multiphase flow. However, their application to FSI problems remains largely unexplored.

Existing studies on neural operator-based FSI have primarily focused on elastically mounted structures constrained by springs and influenced by vortex-induced forces. For instance, Gupta and Jaiman [2022] proposes a hybrid partitioned deep learning framework for fluid dynamics; Fan and Wang [2024] applies convolutional neural networks; and Gao and Jaiman [2024] employs graph neural networks for similar interactions. Rahman et al. [2024] investigates the simultaneous prediction of both fluid dynamics and mounted solid stress, and Xiao et al. [2024] leverages Fourier Neural

Operators (FNO) to predict vesicle dynamics driven solely by fluid pressure without elastic constraints. In all these studies, the structural motion is largely dictated by the surrounding fluid, with limited exploration of freely moving rigid bodies. Lyua et al. [2025] addresses multi-object fluid-structure interactions but remains limited to cases with small structural displacements.

Despite these advancements, neural operators remain underutilized for FSI scenarios involving unmounted rigid bodies. Most existing architectures are designed for either fluid dynamics or solid mechanics in isolation, without explicitly modeling their coupling in FSI. Furthermore, specialized machine learning approaches for FSI predominantly focus on constrained structures, and the lack of standardized datasets has impeded systematic evaluation and benchmarking of neural operator models in this domain. To address these research gaps, this paper makes the following key contributions:

- We introduce two new datasets specifically designed for FSI problems involving freely moving rigid bodies, where the rigid-body motion serves as the control signal that influences the surrounding fluid flow.
- We systematically benchmark state-of-the-art neural operator architectures using these datasets, incorporating three proposed mechanisms for coupling structural motion and fluid initial conditions.

2 Data generation

Fluid-Structure Interaction (FSI) describes the complex coupling between fluid flow and the motion or deformation of a solid structure. The governing equations of fluid dynamics in FSI are the Navier-Stokes equations, coupled through boundary conditions at the interface with the structure:

$$\begin{aligned} \frac{\partial \mathbf{u}(\mathbf{x}, t)}{\partial t} + \mathbf{u}(\mathbf{x}, t) \cdot \nabla \mathbf{u}(\mathbf{x}, t) &= -\nabla p(\mathbf{x}, t) + \frac{1}{\text{Re}} \nabla^2 \mathbf{u}(\mathbf{x}, t) + \mathbf{f}(\mathbf{x}, t), \\ \nabla \cdot \mathbf{u}(\mathbf{x}, t) &= 0, \end{aligned} \quad (1)$$

where $\mathbf{x} \in \Omega$ and $t \in [0, T]$, $\mathbf{u}(\mathbf{x}, t)$ is the velocity field, $p(\mathbf{x}, t)$ is the pressure, Re is the Reynolds number, and $\mathbf{f}(\mathbf{x}, t)$ represents external forces such as gravity. Equation (1) enforces momentum and mass conservation. At the fluid-structure interface, denoted as Γ , the no-slip boundary condition ensures that the fluid velocity at the interface matches the velocity of the solid:

$$\mathbf{u}(\mathbf{x}, t)|_{\mathbf{x} \in \Gamma} = \frac{\partial \mathbf{D}(\mathbf{x}, t)}{\partial t}|_{\mathbf{x} \in \Gamma}, \quad (2)$$

where $\mathbf{D}(\mathbf{x}, t)$ represents the structure displacement. We employed the Aquarium physics solver Lee et al. [2023] for data generation, which implements the Immersed Boundary Method and provides improved stability when handling large fluid–structure interactions.displacements.

In this problem, the rigid body movement serves as a control signal, $\mathbf{c}(t)$, for the dynamic system. The initial fluid velocity and pressure fields are denoted as $\mathbf{u}_0(\mathbf{x})$ and $p_0(\mathbf{x})$, respectively. Our goal is to predict the fluid state at the final time step T , represented by the velocity field \mathbf{u} and pressure field p , using a deep learning model \mathcal{M} parameterized by θ , which approximates the mapping:

$$[\mathbf{u}(\mathbf{x}, T), p(\mathbf{x}, T)] = \mathcal{M}_\theta(\mathbf{u}_0(\mathbf{x}), p_0(\mathbf{x}), \mathbf{c}(t)), \quad \mathbf{x} \in \Omega, \quad t \in [0, T]. \quad (3)$$

For a rigid body in 2D, motion is described by displacement in the x - and y -directions and rotation. To capture finer details, we represent it through the velocities of boundary points. Since the fluid domain is discretized on a regular grid, the boundary is projected onto this grid: each boundary point maps to its nearest grid point, which inherits its velocity components in x and y . Repeating this yields a control signal map $\mathbf{C} \in \mathbb{R}^{H \times W \times M \times 2}$, where H and W are grid dimensions and M the number of sampled signals over time T . Only N pixels contain nonzero values. Solving the governing FSI PDE 1 then produces velocity maps $\mathbf{U} \in \mathbb{R}^{H \times W \times 2}$ and pressure maps $P \in \mathbb{R}^{H \times W}$, representing the fluid state. Figure 1 illustrates the input–output representation.

3 Experiment results

Since all data is represented on a grid, most commonly used neural operator architectures can process this dataset, for example, FNO Li et al. [2020a], WNO Tripura and Chakraborty [2022], U-NO

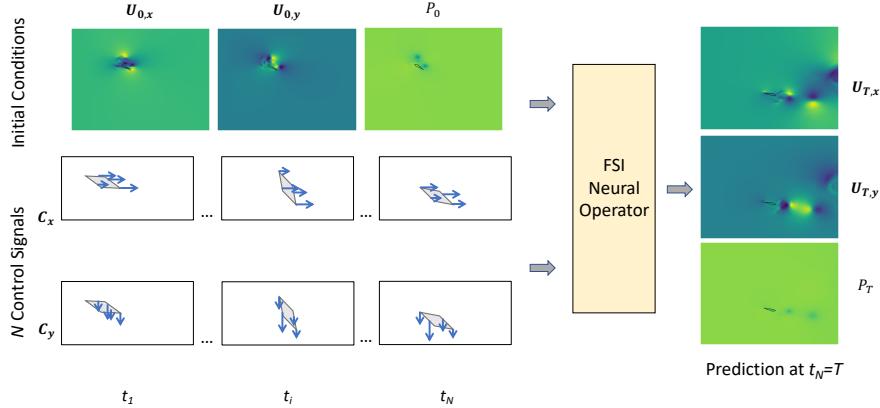


Figure 1: An example from our dataset illustrating the model input and output. The input to the model consists of three-channel feature maps representing the initial conditions ($U_{0,x}, U_{0,y}, P_0$), along with a sequence of N two-channel control signal feature maps (c_x, c_y) at different time steps. The FSI Neural Operator processes these inputs and predicts the final three-channel feature maps ($U_{T,x}, U_{T,y}, P_T$) at the final time step $t_N = T$.

Rahman et al. [2023], and Vision Transformer Han et al. [2022]. However, the key challenge lies in effectively handling the sequential nature of control signals. A simple approach is to incorporate control signals as additional input features alongside the initial conditions. In this work, we propose three strategies for integrating these two types of information. As illustrated in Figure 2, we explore the following approaches:

- **Direct concatenation:** All relevant information, including control signals and initial conditions, is concatenated and fed directly into the neural operator model, allowing it to capture long-range dependencies.
- **Temporal Pooling:** This method applies temporal average pooling to the control signals and integrates the output with the initial condition before feeding it into the model, which can reduce computational complexity and handle an arbitrary number of control signals.
- **Sequential:** The control signals at each time step are iteratively integrated with the hidden embedding of the fluid state at the corresponding time step. Here, the neural operator serves as a one-step forward mapping.

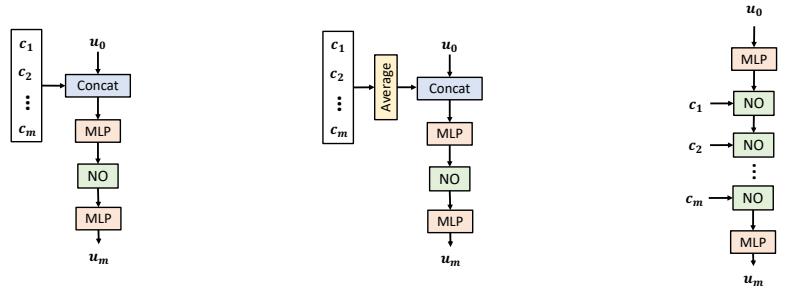


Figure 2: Illustration of three different strategies for coupling control signals with initial conditions to predict future fluid states.

For all three backends, the inputs are first projected by a linear layer, while the boundary conditions are embedded through another linear layer. These representations are concatenated, compressed by a linear transformation, and finally mapped to the output space through a linear prediction layer. The U-Net backend employs a two-level encoder–decoder architecture with a constant channel width of

64: the encoder applies a convolutional block followed by downsampling and another convolutional block, while the decoder applies convolutional blocks with skip connections, progressively merging features from the encoder. Upsampling is performed twice with transposed convolutions, using Tanh activations and no normalization layers. The FNO backend adopts a single 2D Fourier layer per step with 64 input and output channels and spectral resolution in both spatial directions. In the iterative variant, this layer is applied repeatedly at each time or boundary condition step, whereas in the concatenation/pooling variant it is applied once after compression. The ViT backend employs a Vision Transformer with a patch size of 16, an embedding dimension of 768, six encoder layers, eight attention heads, and an MLP dimension four times the embedding size. The transformer output is passed through a linear layer, reshaped into a feature map, and bilinearly upsampled back to the spatial resolution of the domain.

We evaluated each model under multiple coupling schemes to determine their effectiveness in learning rigid-body fluid-structure interaction (FSI). Table 1 shows that Vision Transformer (VT) consistently yields higher error compared to FNO and U-Net, suggesting that self-attention alone may struggle to capture the spatial-temporal complexities of FSI. By contrast, FNO and U-Net achieve noticeably lower error, highlighting their superior ability to learn fluid dynamics in the presence of rigid-body motion. Additionally, the direct concatenation approach emerges as the most effective strategy for incorporating initial conditions and control signals, outperforming iterative coupling and temporal pooling across all architectures. The iterative approach appears to accumulate prediction error over successive time steps, while simple pooling compresses essential temporal details. To better illustrate

Table 1: Mean absolute error of the baseline model predictions.

Model	Variant	Periodic Oscillation			Random Movement		
		u	v	p	u	v	p
VT	Iterative	0.066	0.086	3.91	0.198	0.182	46.1
	Pooling	0.062	0.079	3.37	0.178	0.146	44.1
	Concat	0.060	0.069	2.94	0.173	0.142	41.1
FNO	Iterative	0.072	0.099	3.57	0.168	0.161	40.7
	Pooling	0.036	0.037	2.82	0.113	0.118	24.1
	Concat	0.035	0.036	2.84	0.110	0.112	23.8
WNO	Iterative	0.069	0.091	3.32	0.196	0.152	44.0
	Pooling	0.034	0.035	2.76	0.140	0.111	20.0
	Concat	0.032	0.033	2.48	0.112	0.096	18.9
U-Net	Iterative	0.066	0.082	3.07	0.191	0.148	43.3
	Pooling	0.039	0.043	2.71	0.132	0.105	19.1
	Concat	0.030	0.028	2.19	0.105	0.090	17.8

the performance of different models, Figure 3 presents the predicted fluid velocity of U-Net model, WNO model, and FNO model with direct concatenation. We can observe that U-Net can provide a velocity map with less noise. Figure 4 compares the predicted drag force with the exact drag force for two types of rigid-body movement using U-Net.

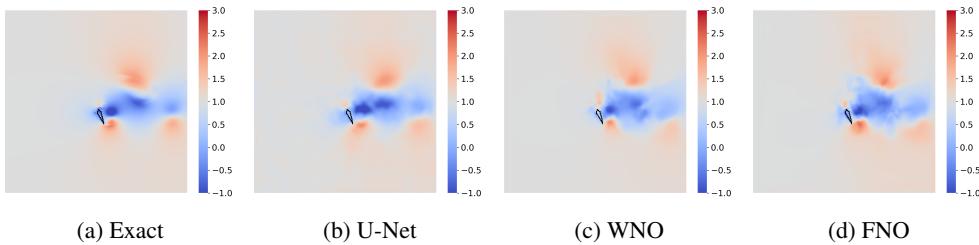


Figure 3: Comparison between the exact and predicted U_x fields by U-Net and FNO.

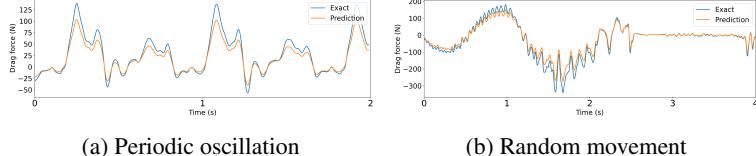


Figure 4: Drag force computations based on the predicted fluid pressure fields. Two representative cases are shown: (a) periodic oscillation and (b) random movement. The predicted drag force closely follows the ground truth over time.

4 Conclusion

In this study, we introduced two novel datasets for fluid–rigid body interaction (FSI) involving freely moving rigid bodies that drive surrounding fluid flow as control signals, and benchmarked state-of-the-art neural operators on these tasks. Our results show that concatenating initial conditions and control signals yields the most accurate predictions, while iterative coupling and purely attention-based models like the Vision Transformer are less effective compared to architectures based on U-Net and FNO. Nonetheless, the framework exhibits limitations, loss of accuracy over long horizons and inability to query fluid states at arbitrary time stamps.

An interesting direction for future work is to incorporate attention-based architectures as control signal encoders, which may provide more flexible and context-aware representations than the current design. Another promising extension is the use of Neural ODEs, which could enable continuous-time modeling and more stable long-horizon predictions. At the same time, our reliance on grid-based simulation may limit accuracy, particularly in capturing fine geometric or boundary effects. Potential improvements include adopting more adaptive discretization strategies, integrating mesh- or point-based solvers, and incorporating uncertainty quantification to assess reliability. Together, these extensions could enhance both the expressiveness of the model and the fidelity of the simulations.

References

- Hamid Alemi Ardekani. A variational principle for three-dimensional interactions between water waves and a floating rigid body with interior fluid motion. *Journal of Fluid Mechanics*, 866: 630–659, 2019.
- Klaus-Jürgen Bathe. Finite element method. *Wiley encyclopedia of computer science and engineering*, pages 1–12, 2007.
- Robert Eymard, Thierry Gallouët, and Raphaèle Herbin. Finite volume methods. *Handbook of numerical analysis*, 7:713–1018, 2000.
- Xiantao Fan and Jian-Xun Wang. Differentiable hybrid neural modeling for fluid-structure interaction. *Journal of Computational Physics*, 496:112584, 2024.
- Rui Gao and Rajeev K Jaiman. Predicting fluid–structure interaction with graph neural networks. *Physics of Fluids*, 36(1), 2024.
- Rachit Gupta and Rajeev Jaiman. A hybrid partitioned deep learning methodology for moving interface and fluid–structure interaction. *Computers & Fluids*, 233:105239, 2022.
- Kai Han, Yunhe Wang, Hanting Chen, Xinghao Chen, Jianyuan Guo, Zhenhua Liu, Yehui Tang, An Xiao, Chunjing Xu, Yixing Xu, et al. A survey on vision transformer. *IEEE transactions on pattern analysis and machine intelligence*, 45(1):87–110, 2022.
- Zhongkai Hao, Zhengyi Wang, Hang Su, Chengyang Ying, Yinpeng Dong, Songming Liu, Ze Cheng, Jian Song, and Jun Zhu. Gnot: A general neural operator transformer for operator learning. In *International Conference on Machine Learning*, pages 12556–12569. PMLR, 2023.
- Ebrahim M Kolahdouz, Amneet Pal Singh Bhalla, Lawrence N Scotten, Brent A Craven, and Boyce E Griffith. A sharp interface lagrangian-eulerian method for rigid-body fluid-structure interaction. *Journal of computational physics*, 443:110442, 2021.

- Jeong Hun Lee, Mike Y Michelis, Robert Katschmann, and Zachary Manchester. Aquarium: A fully differentiable fluid-structure interaction solver for robotics applications. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 11272–11279. IEEE, 2023.
- Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*, 2020a.
- Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Graph kernel network for partial differential equations. *arXiv preprint arXiv:2003.03485*, 2020b.
- Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via deeponet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3(3):218–229, March 2021. ISSN 2522-5839. doi: 10.1038/s42256-021-00302-5. URL <http://dx.doi.org/10.1038/s42256-021-00302-5>.
- Yanfang Lyua, Yunyang Zhang, Zhiqiang Gongb, Xiao Kangd, Wen Yao, and Yongmao Pei. A novel hybrid neural network of fluid-structure interaction prediction for two cylinders in tandem arrangement. *arXiv preprint arXiv:2504.14971*, 2025.
- Hideyo Negishi, Masahiro Kondo, Hidenao Takahashi, Hiroaki Amakawa, Shingo Obara, and Ryoichi Kurose. Fluid–rigid body coupling simulations with the passively moving solid model based on a physically consistent particle method. *Physics of Fluids*, 36(3), 2024.
- Md Ashiqur Rahman, Zachary E. Ross, and Kamyar Azizzadenesheli. U-no: U-shaped neural operators, 2023. URL <https://arxiv.org/abs/2204.11127>.
- Md Ashiqur Rahman, Robert Joseph George, Mogab Elleithy, Daniel Leibovici, Zongyi Li, Boris Bonev, Colin White, Julius Berner, Raymond A Yeh, Jean Kossaifi, et al. Pretraining codomain attention neural operators for solving multiphysics pdes. *Advances in Neural Information Processing Systems*, 37:104035–104064, 2024.
- Tapas Tripura and Souvik Chakraborty. Wavelet neural operator: a neural operator for parametric partial differential equations. *arXiv preprint arXiv:2205.02191*, 2022.
- Wang Xiao, Ting Gao, Kai Liu, Jinqiao Duan, and Meng Zhao. Fourier neural operator based fluid–structure interaction for predicting the vesicle dynamics. *Physica D: Nonlinear Phenomena*, 463:134145, 2024.