LASER- AND SHOCK-INDUCED DROPLET DYNAMICS: A MACHINE LEARNING BENCHMARK FOR COMPLEX MULTIPHASE FLOWS

Anonymous authors
Paper under double-blind review

ABSTRACT

Compressible multiphase flow is central to numerous engineering applications, characterized by complex wave dynamics and challenging shock-interface interactions. Despite their importance, they remain significantly missing from existing benchmarks in the Scientific Machine Learning (SciML) community, limiting progress on generalization to impactful real-world scenarios. To address this issue, we introduce two exemplary datasets from this class, Laser-Induced Droplet Explosion (LIDE) and Shock-Induced Droplet Aero-breakup (SIDA), providing researchers with valuable references to establish reliable baselines and push boundaries of SciML. Due to the high computational cost of simulating these processes with full fidelity, we explore data-driven surrogate models designed to efficiently approximate the underlying physics at reduced cost. We benchmark these datasets on diverse architectures–UNet, Fourier Neural Operator (FNO), Vision Transformer (ViT), Scalable Operator Transformer (ScOT), and Residual Network (ResNet)–trained autoregressively and compared across varying parameter counts. A comprehensive set of ablations is carried out to analyze the performance of the models. We identify key scenarios, such as incorporating temporal sequence information and conditioning, that enable the models to accurately capture the rich and nonlinear physics embedded in the datasets. Code and datasets will be made available upon request.

1 INTRODUCTION

Modern technical applications of fluid dynamics exhibit a plethora of flow scenarios involving compressible and multiphase flows, which are characterized by discontinuities across shockwaves and phase boundaries. Gaining insights into the underlying physics of compressible flows is a cornerstone in many real-world systems. These include a wide range of scientific fields, spanning from astrophysics to engineering applications such as coating, fuel injection, biomedical treatment (Chaussy & Schmiedt, 1984), analysis of cavitation phenomena (Maeda et al., 2015), and nanoparticle synthesis (Riahi et al., 2023). Traditionally, domain experts have analyzed these phenomena through simulations and experiments. The downside of these methods is that they demand highly specialized facilities and substantial computational power.

Recent advancements in deep learning algorithms and data-driven modeling (Cai et al., 2021), (Ho et al., 2020), (Lipman et al., 2022), (Kovachki et al., 2023), (Vaswani et al., 2017)), coupled with the rapid growth of modern high-performance computing infrastructures, have accelerated discoveries in Scientific Machine Learning (SciML), enabling robust and reliable surrogate models. However, training these models requires large, multifaceted datasets that capture and correlate spatiotemporal information.

To the best of our knowledge, while datasets exist for either compressible single-phase flows (Takamoto et al., 2022), (Herde et al., 2024) or incompressible multiphase flows (Shadkhah et al., 2025), (Hassan et al., 2023), there is an absence of labeled datasets that capture the complexity of both simultaneously. We address this scarcity by providing two high-fidelity datasets pertaining to liquid droplet dynamics, called **Laser-Induced Droplet Explosion (LIDE)** and **Shock-Induced Droplet Aero-breakup (SIDA)**. This novel set of datasets involves intricate interactions of shocks

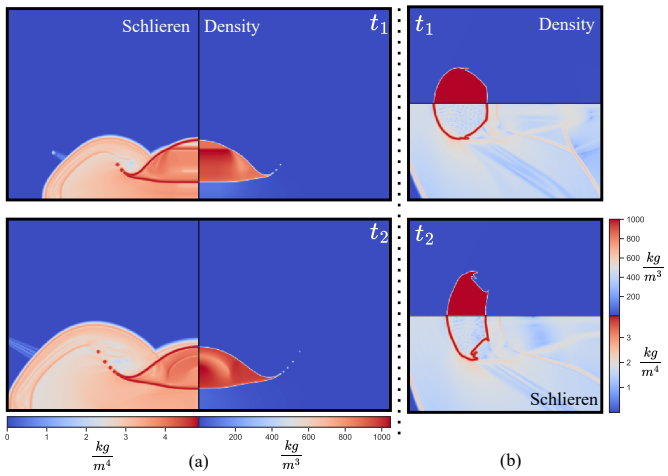

Figure 1: Two time snapshots at $t_1 = 30[s]$ and $t_2 = 50[s]$ of the Density and Schlieren field of the (a) LIDE and (b) SIDA dataset.

with interfaces-Richtmyer-Meshkov, Rayleigh-Taylor, and Kelvin-Helmholtz instabilities. It further captures the evolution of multiscale vortical structures and wave dynamics. Therefore, it requires profound domain expertise and computational resources, and our contribution lies in introducing this valuable dataset, which paves the way for advancing research in the community. An illustration of two field variables of each dataset is depicted in Figure 1. In LIDE (Paula et al., 2019), an initial high-pressure laser cavity is generated in a micro-droplet. Initiated shock-interface interactions lead to droplet breakup and cavitation events. In SIDA (Kaiser et al., 2020), a shock wave hits a droplet and initiates aero-breakup, where triggered interfacial instabilities generate small liquid fragments through different scenarios.

We propose a many-to-many training strategy (Shadkhah et al., 2025) to benchmark our datasets on a variety of neural architectures, ranging from convolution and spectral models to attention-based approaches. Specifically, we consider UNet, Residual Network (ResNet), Fourier Neural Operator (FNO), Vision Transformer (ViT), alongside Scalable Operator Transformer (ScOT). Furthermore, we identify key parameters and fields with the goal of designing an extensive set of ablations to experiment with the generalization capabilities of the models. Although training these models is computationally intensive, once trained, these models are substantially faster when used as a forward simulator. The key contributions of this work are:

- **Datasets for Complex Flow Physics.** A new high-fidelity dataset for complex flow physics involving droplet dynamics and shock-interface interactions is generated and presented.
- **Dataset Validation.** Dataset fidelity is assessed and confirmed by high-resolution simulations and independent experiments.
- **Benchmarking.** A comprehensive set of experiments is performed through side-by-side comparison with different models to gain insights into generalization capabilities.

2 Related Work

Existing benchmarks differ in scope and physical coverage. Among them, PDEBench (Takamoto et al., 2022) offers a wide variety of datasets, including single-phase compressible Navier–Stokes problems, BubbleML (Hassan et al., 2023) and MPF-Bench (Shadkhah et al., 2025) extend to multiphase problems and contribute an impressive collection of bubble and droplet datasets; however, both are limited to incompressible physics. It is noteworthy that Poseidon (Herde et al., 2024) provides an extensive set of datasets to train foundation models, although it considers only single-phase problems. However, there is no benchmark combining both compressible and multiphase physics in the same setting. Our work addresses this gap by integrating these two characteristics and further incorporates Symmetry, Dirichlet, and Neumann boundary conditions, thereby broadening the

diversity of physical scenarios available for SciML research. A summary of the aforementioned references is presented in Table 1.

Table 1: Summary of related datasets.

Name	Dimensions	Compressible	Multiphase
PDE Bench (Takamoto et al., 2022)	2	✓	✗
Poseidon (Herde et al., 2024)	2	✓	✗
BubbleML (Hassan et al., 2023)	2, 3	✗	✓
MPF-Bench (Shadkhah et al., 2025)	2, 3	✗	✓
Current study	2	✓	✓

3 DATASETS

We focus on the class of compressible multiphase problems in this paper. Breakup of liquid droplets is a significant example in this class, which can be induced by laser irradiation (LIDE) or a shock (SIDA). These two transient problems are investigated intensely through experiments and numerical simulations. The Robust Discrete Equation Method for Interface Capturing (RDEMIC) (Paula et al., 2023) is used to generate targets through solving the two-dimensional (2D) axisymmetric compressible Euler equations. Adopting an axisymmetric setup reduces computational cost compared to the full three-dimensional treatment. The set of equations, without dissipative terms in vector notation, reads

$$\partial_t \mathbf{U}_l + \nabla \cdot \mathbf{F}_l = \mathbf{B}_l \cdot \nabla \alpha_l + \mathbf{S}_l, \tag{1}$$

where subscript l denotes the index of the phase, $\mathbf{U}_l$ is the vector of conserved quantities, $\mathbf{F}_l$ is the flux tensor, $\mathbf{B}_l$ is the interaction tensor, and $\mathbf{S}_l$ is a source term to account for cylindrical symmetry,

$$\mathbf{U}_l = \begin{bmatrix} \alpha_l \\ \alpha_l \rho_l \\ \alpha_l \rho_l \mathbf{u}_l \\ \alpha_l E_l \end{bmatrix}, \quad \mathbf{F}_l = \begin{bmatrix} 0 \\ \alpha_l \rho_l \mathbf{u}_l^T \\ \alpha_l \rho_l \mathbf{u}_l \otimes \mathbf{u}_l + \alpha_l p_l \mathbf{I} \\ \alpha_l (E_l + p_l) \mathbf{u}_l^T \end{bmatrix}, \quad \mathbf{B}_l = \begin{bmatrix} -\mathbf{u}_{\text{int}}^T \\ 0 \\ p_{\text{int},l} \mathbf{I} \\ p_{\text{int},l} \mathbf{u}_{\text{int}}^T \end{bmatrix}, \quad \mathbf{S}_l = -\frac{\alpha_l u_{r,l}}{r} \begin{bmatrix} 0 \\ \rho_l \\ \rho_l \mathbf{u}_l \\ E_l + p_l \end{bmatrix}.$$

Above, α_l, ρ_l, $\mathbf{u}_l$, p_l, $u_{r,l}$, and E_l imply the volume fraction, mass density, velocity vector, pressure, velocity component in the radial direction, and total energy of phase l, respectively. Interface velocity vector and pressure are indicated by $\mathbf{u}_{\text{int}}$ and $p_{\text{int},l}$, respectively; without considering the surface tension, $p_{\text{int},l}$ is the same for all phases; r denotes the distance from the symmetry axis and $\mathbf{I}$ is the identity tensor. This method is implemented and validated extensively through the Finite Volume solver, ALPACA (Hoppe et al., 2022). In cylindrical coordinate configuration, the domain revolves around the z-axis (south, as shown in Figure 2), resulting in an axisymmetric problem. In the following sections, a brief overview of each dataset is given. For more details, refer to Appendix A.

3.1 LASER-INDUCED DROPLET EXPLOSION (LIDE)

Experimental investigations of LIDE provide a valuable insight into pure liquid states and pressure-sensitive molecular dynamics in solutions (Stan et al., 2016a). When a laser pulse hits the transparent liquid droplet, energy is deposited within nanoseconds, forming a high-pressure filament along the laser trajectory. This induces shock and expansion waves, which are reflected and subsequently generate negative pressure waves inside the droplet. Consequently, the droplet undergoes deformation and eventually ruptures if the tension is strong enough. Notably, the negative pressure at rupture is related to the tensile strength that the liquid can sustain during decompression (Stan et al., 2016b).

This problem is also numerically addressed in literature (Paula et al., 2019). Taking advantage of the symmetries, a droplet with radius R_0 is located in the bottom left corner of a square domain with length $3R_0$, as shown in Figure 2a. The filament, heated by the laser beam along the centerline, is also illustrated. The boundary conditions (BC) are Symmetry (west) and Zero-gradient (east and north). The latter refers to a special case of Neumann BC, where the normal derivative of the field variable at the boundary is set to zero. To explore the dynamics of the explosion, we vary the values for filament pressure, ambient pressure, laser half-width, and the droplet radii along perpendicular axes, which distinguishes spherical from ellipsoidal geometries. The aforementioned parameters are subsequently used as conditioning parameters during training. More details on the initial condition values and the validation of the dataset are described in Appendix A.1.

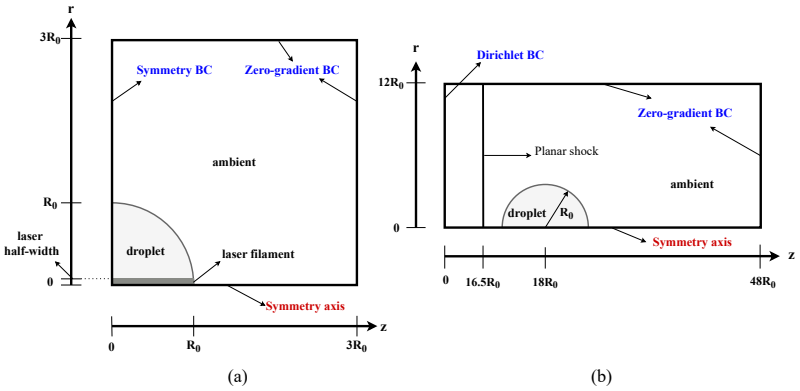

Figure 2: Initial setup for (a)LIDE and (b)SIDA.

3.2 Shock-Induced Droplet Aero-breakup (SIDA)

The droplet aero-breakup, which is caused by the sudden exposure of liquid droplets to external flow, is relevant in practical applications of fuel injection and shock-tube flow (Liang et al., 2020). The resulting shock–droplet interaction involves the evolution of reflected, transmitted, and diffracted waves, along with droplet displacement, deformation, and the development of surface instabilities. This high-speed phenomenon requires high spatiotemporal resolution to be accurately captured. Surface tension has a strong impact on the droplet breakup mode, which is characterized by the Weber number (Hinze, 1955). This non-dimensional parameter accounts for the relative dominance of aerodynamic force over surface tension. Furthermore, the external flow regime, from subsonic to supersonic, is governed by the Mach number (Kaiser et al., 2020).

Initially, we simulate the SIDA dataset in a domain of size $[48R_0, 12R_0]$, which is shown in Figure 2b. This large domain is essential to avoid undesirable boundary effects regarding wave dynamics. However, a fixed subdomain with size $[6R_0, 3R_0]$ around the droplet is saved and later used in training. This subdomain is chosen such that in the initial timestep, the shock wave is located at the west end.

Boundary conditions include Dirichlet (west) and Zero-gradient (east and north). This dataset is generated with various combinations of Mach and Weber numbers, which are later utilized as conditioning parameters in model training (Meng & Colonius, 2018), (Winter et al., 2019). More details on the initial condition values and the validation of the dataset are described in Appendix A.2.

3.3 MetaData

Each dataset[1] includes 128 trajectories, and the splitting for training/validation/inference is 86/10/32. In total, 6 fields are made available for each dataset, where density, pressure, X-velocity, Y-velocity, and schlieren are common in both datasets. The remaining channel is

[1]The uploaded supplementary material as a .zip file includes metadata.json files for each LIDE and SIDA dataset. Also, sample video files are provided for visualization.

the total energy for LIDE, and vorticity for SIDA. The spatiotemporal parameters used in the numerical solver are presented in Table 2. The datasets are stored as HDF5 files, with sizes of 75 GB and 12 GB for LIDE and SIDA, respectively, and the shapes for both are **[num_of_trajectories][num_of_timesteps][fields][X-resolution][Y-resolution]**. Each trajectory in the dataset file is assigned a unique group name based on its corresponding conditioning parameters.

Table 2: Metadata for LIDE and SIDA datasets

Dataset	**Resolution** [X, Y]	**End time [s]**	**CFL**[2]	Δt_{save} **[s]**	Δt_{solver} **[s]**[3]	Δx **[m]**
LIDE	[256, 256]	20×10^{-9}	0.35	1.00×10^{-10}	6.80×10^{-12}	1.25×10^{-7}
SIDA	[256, 128]	15×10^{-6}	0.50	0.25×10^{-6}	1.95×10^{-9}	1.17×10^{-5}

4 EXPERIMENTS

4.1 DESIGN OF EXPERIMENTS

This section outlines the Design of Experiments (DOE). Each experiment is assigned a unique tag for easier identification and comparison. We use **'P'** for Pressure, **'D'** for Density, **'U'** for X-Velocity, **'V'** for Y-Velocity, **'E'** for Energy, **'S'** for Schlieren, and **'Vo'** for Vorticity. For example, an experiment with a tag 'PDUV[ES]_T_(3,2)' implies the input channels are Pressure (P), Density (D), X-Velocity (U), Y-Velocity (V), Energy (E), and Schlieren (S). '[ES]' shows that Energy and Schlieren are counted as conditioning fields and are not predicted in the output. Furthermore, 'T' indicates that the conditioning parameters are included in the experiment. Finally, '(3,2)' corresponds to 3 consecutive inputs and 2 consecutive predicted frames. The complete DOE table is provided in the Appendix B.1.

4.2 BASELINE MODELS

We investigate the performance of the datasets on a variety of neural architecture baselines, The models under consideration are: UNet (Ronneberger et al., 2015), ResNet (He et al., 2016), FNO (Li et al., 2020), ViT (Dosovitskiy et al., 2020), and ScOT (Herde et al., 2024). Each model was trained from scratch on two parameter categories, i.e., 1M and 50M. However, ResNet is trained only with 1M model parameter count. For more details on model hyperparameters, refer to Appendix B.2.

4.3 INVESTIGATION SCENARIOS

We analyze our results by categorizing the experiments into three distinct scenarios. Each scenario addresses a certain learning problem, and experiments are grouped by altering only the learning parameter while holding all other parameters fixed. We denote the grouped experiments by 'G' in all plots in section 5 (Results). The following subsections give a brief overview of these learning problems.

4.3.1 TEMPORAL CONTEXT

Historic information, provided through additional temporal inputs (frames), has proved its efficacy (Hassan et al., 2023), (Shadkhah et al., 2025). In some experiments, to facilitate the understanding of the patterns, we incorporate multiple frames into the model. This provision is effective in learning transient trajectories. For both datasets, we experiment with either 1 input or including a sequence of 3 historic inputs. We also define a stride parameter during dataloading, which skips a fixed number of timesteps. In the LIDE and SIDA datasets, strides of 10 and 5 timesteps are employed, respectively.

[2]CFL refers to Courant-Friedrichs-Lewy criterion.

[3]This is the average solver timestep among all trajectories.

4.3.2 CONDITIONING PARAMETERS

In many fluid dynamic problems, the physics are fundamentally characterized by non-dimensional and domain parameters, which influence the system's evolution. These provide crucial information as they dictate the governing dynamics, leading to distinct flow regimes. Conditioning the model with such parameters improves generalization (Kohl et al., 2023), (Peebles & Xie, 2023). The conditioning parameters for the LIDE and SIDA datasets are mentioned in section 3 (Datasets). These are injected into the models through the normalization layers (Herde et al., 2024). More details on the implementation are provided in Appendix B.3.

4.3.3 CONDITIONING FIELDS

In this experimental scenario, additional channels are appended to the inputs before passing them to the model. These extra channels are called conditioning fields, which are derived quantities from existing inputs. For the LIDE dataset, we incorporate energy and schlieren as the conditioning fields, whereas for the SIDA dataset, vorticity and schlieren are used. We aim to test the hypothesis that this type of conditioning guides the model towards generalization.

4.4 TRAINING AUTOREGRESSIVE MODELS

In this work, we use a many-to-many training style to train each of our baselines, $\mathcal{M}_\theta$. The dataset is a discrete spatiotemporal system, containing c channels. For a particular trajectory, the mapping is given by $\mathbf{X}_t : \Omega \times [0, T] \to \mathbb{R}^c$, where $\Omega \subset \mathbb{R}^2$ and T represents the last timestep of the trajectory.

During training, we split each of the training trajectories into M windows. The length of each window is determined by the number of input and output sequences, denoted by l_1 and l_2, respectively, and s denotes the stride, which are all hyperparameters of the temporal context study as mentioned in section 4.3.1.

The input sequence of the m^{th} window is given by $\mathbf{X}_m = [\mathbf{X}_m, \dots \mathbf{X}_{m+(l_1 \times s)]} \in \mathbb{R}^{l_1 \times c}$ and the corresponding target is $\mathbf{Y}_m = [\mathbf{X}_{m+((l_1+1)\times s)]} \dots \mathbf{X}_{m+((l_1+l_2)\times s)}] \in \mathbb{R}^{l_2 \times c}$. The training loss reads:

$$\text{MSE} := \frac{1}{M} \sum_{m=1}^{M} \left\| \mathcal{M}_\theta \left(\mathbf{X}_m \right) - \mathbf{Y}_m \right\|^2, \tag{2}$$

After each training epoch, the validation loss is computed by rolling out the model autoregressively for 5 steps and then computing the Root Mean Square Error (RMSE).

4.5 INFERENCE METRICS

During inference, we start from the initial condition of each trajectory and rollout the model in an autoregressive fashion to reach the final frame. The predictions across trajectories are accumulated into a tensor, and the Mean Average Error (MAE) and Root Mean Square Error (RMSE) metrics (Refer to Appendix B.5) are obtained by comparing the predictions against the target. These metrics, henceforth, are referred to as error-type 1.

We define an error-type 2 by starting again from the initial frame and performing rollout until the end of the sequence. We compute the per-frame RMSE error, yielding a tensor of shape (N, R, T, C, spatial-dims), where N is the number of trajectories, R is the number of rollout steps, T is the number of output timesteps, C is the number of output channels, and spatial-dims is the resolution of the dataset. The error aggregation is performed in four stages:

1. In each trajectory, we first average the error over the temporal, channel, and spatial dimensions, resulting in an overall tensor with shape (N, R).
2. We compute the cumulative summation along the rollout dimension, retaining the shape (N, R).
3. We compute mean and standard deviation across trajectories (N), which results in a tensor of shape (R).

4. Finally, we reduce across the rollout dimension to obtain the overall mean and standard deviation. We denote this metric as error-type 2. The reported metrics in the plots of section 5 (Results) are error-type 2.

5 RESULTS

5.1 EFFECT OF SEQUENCE INFORMATION

Within the many-to-many autoregressive training framework, we evaluate three configurations of sequence information: (1,1), (3,1), and (3,2), corresponding to one input–one prediction, three inputs–one prediction, and three inputs–two predictions, respectively. For both the LIDE and the SIDA datasets, we observe a consistent performance improvement across all models trained with three historic timesteps, with the single prediction models having a slight metric advantage over the two consecutive predictions. A further gain in accuracy is obtained upon increasing the parameter count, with UNet performing the best. These results for the SIDA are depicted in Figure 3. The results for the LIDE are shown in Figure 12 in the Appendix C.

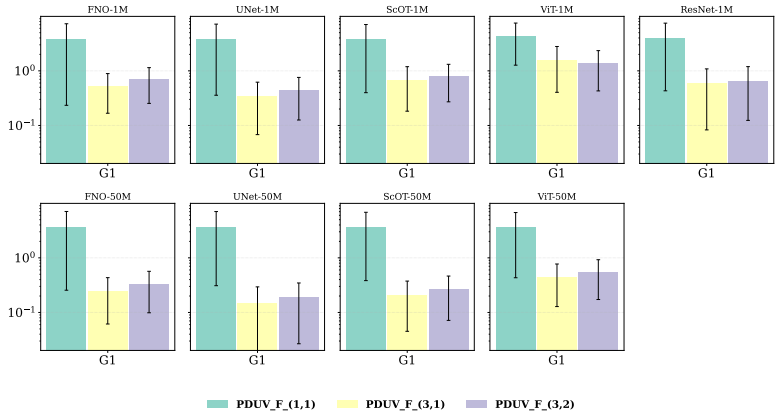

Figure 3: Effect of sequence information for SIDA dataset. Error-type 2 is presented.

5.2 EFFECT OF CONDITIONING PARAMETERS

We conduct several studies to assess if including conditioning parameters has a pronounced influence on the inference metrics. For the LIDE dataset, we observe that the effect of embedding these parameters into the baselines is evident, whereas metrics deteriorate for the SIDA dataset, as illustrated in Figure 4 and 5, respectively. It is worth noting that the characteristics of the conditioning parameters in the SIDA dataset are different from those of the LIDE dataset. In the former, the parameters are geometry-based, and for the latter, these are flow properties.

5.3 EFFECT OF CONDITIONING FIELDS

Considering the selected conditioning fields for each dataset according to the section 4.3.3, we conclude from Figure 6 that across all models and parameter counts, incorporating these fields degrades the predictions, resulting in increased errors during inference. The same results are illustrated by Figure 13 in Appendix C for the LIDE dataset.

5.4 BASELINE MODEL PERFORMANCE STUDY

We investigate the MAE and RMSE of type 1 (section 4.5) in baseline models on an identical experiment for each dataset. As a sample experiment, we present Table 3, which shows that a higher parameter count improved the prediction accuracy across all models. UNet consistently achieves superior performance compared to all the other baselines in both the 1M and 50M categories. Remaining tables are available in Appendix E.

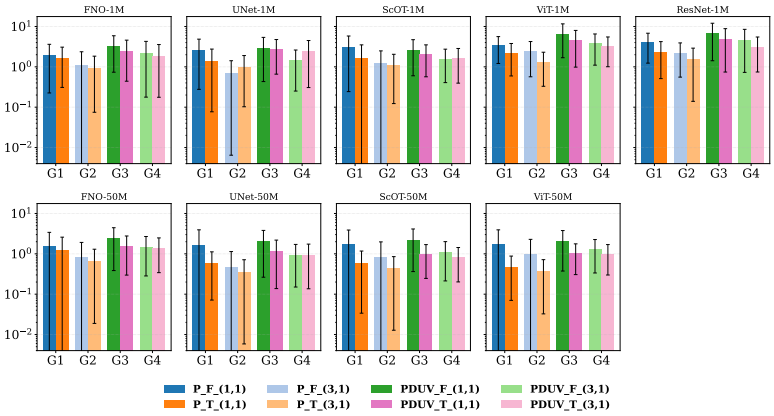

Figure 4: Effect of conditioning parameters for LIDE dataset. Error-type 2 is presented.

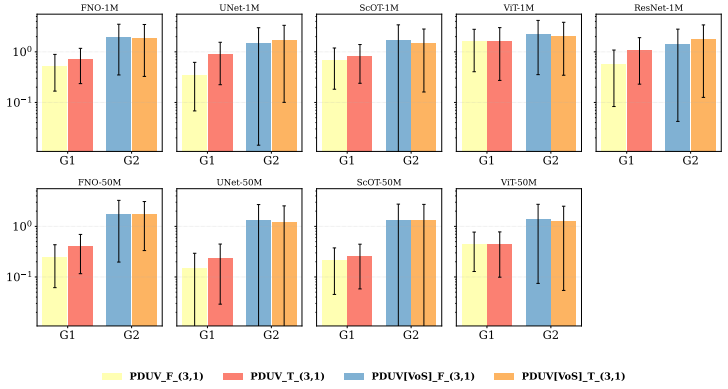

Figure 5: Effect of conditioning parameters for SIDA dataset. Error-type 2 is presented.

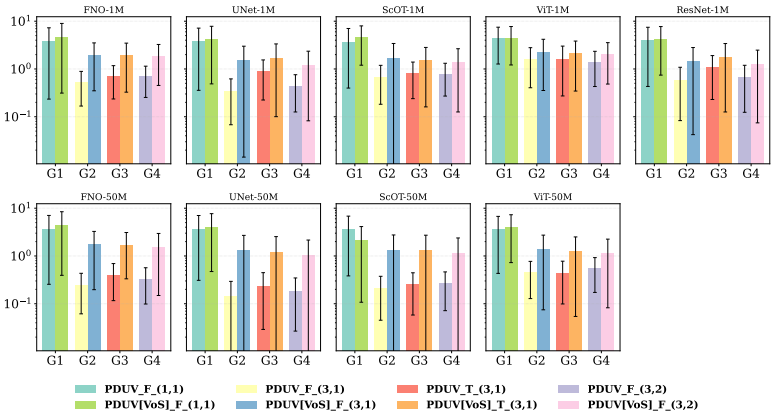

Figure 6: Effect of conditioning fields for SIDA dataset. Error-type 2 is presented.

5.5 Comparison between error types

We compare the two error types, defined in section 4.5, to correlate the metrics with the predicted rollout. It is worth emphasizing that from our ablations, error-type 2 demonstrates better coherence

Table 3: Error-type 1 for experiment PDUV_F_(3,1) for the LIDE dataset across all models.

MODEL	1M		50M	
	MAE	RMSE	MAE	RMSE
UNet	0.058351	0.201772	0.029650	0.130409
FNO	0.114536	0.359660	0.065122	0.239348
ViT	0.184868	0.493885	0.047491	0.169020
ScOT	0.068976	0.210712	0.039000	0.147999
ResNet	0.314502	0.683159		

with predicted rollouts in some cases. For example, as shown in Table 4, UNet-50M achieves higher accuracy according to error-type 2 compared to ViT-50M; UNet captures the droplet interface more precisely, indicating better performance as a surrogate relative to ViT. The corresponding plot is available in Figure 24 in Appendix D. In contrast, error-type 1 suggests that ViT predicts better. This discrepancy highlights the importance of selecting an error metric that aligns with the qualitative behavior observed in rollout plots (Luo et al., 2023). Example rollout prediction plots, during inference, for the LIDE and the SIDA datasets are shown in Appendix D.

Table 4: Error-type 1 and 2 for the experiment PDUV_T_(3,1) for the LIDE dataset across models with 50M parameters.

MODEL	Error type 1	Error type 2
UNet	0.132766	0.938997
FNO	0.185669	1.411901
ViT	0.121964	0.997704
ScOT	0.104845	0.818168

6 Conclusion

This study presents two novel datasets in the domain of compressible multiphase fluid dynamics. We benchmarked five baseline models on these datasets with varying parameter count. Our study scenarios explore the influence of historic information, conditioning parameters and fields. Inference results of trained baseline models on both the LIDE and the SIDA datasets showed superior prediction accuracy upon incorporating additional temporal context. Subsequently, introducing additional channels as conditioning fields to the input degraded the prediction accuracy during inference on both datasets. Furthermore, injecting conditional parameters into the baselines yielded bifurcating results for the datasets. Despite poor performance on the SIDA dataset, models show better accuracy on the LIDE dataset. Finally, we examined the interpretation of two error types and their correlation with the rollout plots, which illustrates the importance of selecting a suitable error metric in choosing an appropriate surrogate. In conclusion, it is essential to highlight that representing the complex physics and patterns through the current datasets by surrogates still poses a challenge. This observation motivates the integration of such datasets in the SciML community to further the development of data-driven surrogates.

Limitations and Future works. Inclusion of a broader range of models, additional error types, and analyzing different combinations of conditioning fields and parameters are future directions. Advancing toward more effective conditioning algorithms is also an important investigation. Ultimately, these developments will enable rapid and efficient exploration of parameter spaces that govern complex multiphase flow phenomena.

7 REPRODUCIBILITY STATEMENT

We introduced two datasets in this paper, which are reproducible based on our description in the main text (section 3) and supplements in the Appendix (A). These explanations include the referenced Finite Volume solver, numerical setup, and initial conditions. In addition, for reproducing model evaluations, we provide trained model weights and the code that has the complete set of instructions upon request.

REFERENCES

John David Anderson. Modern compressible flow: with historical perspective. *(No Title)*, 1990.

Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E Hinton. Layer normalization. *arXiv preprint arXiv:1607.06450*, 2016.

Shengze Cai, Zhiping Mao, Zhicheng Wang, Minglang Yin, and George Em Karniadakis. Physics-informed neural networks (pinns) for fluid mechanics: A review. *Acta Mechanica Sinica*, 37(12): 1727–1738, 2021.

Christian Chaussy and Egbert Schmiedt. Extracorporeal shock wave lithotripsy (eswl) for kidney stones. an alternative to surgery? *Urologic radiology*, 6(1):80–87, 1984.

PhysicsNeMo Contributors. Nvidia physicsnemo: An open-source framework for physics-based deep learning in science and engineering. `https://github.com/NVIDIA/physicsnemo`, February 2023. Software release.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

Sigal Gottlieb and Chi-Wang Shu. Total variation diminishing runge-kutta schemes. *Mathematics of computation*, 67(221):73–85, 1998.

Jayesh K Gupta and Johannes Brandstetter. Towards multi-spatiotemporal-scale generalized pde modeling. *arXiv preprint arXiv:2209.15616*, 2022.

Sheikh Md Shakeel Hassan, Arthur Feeney, Akash Dhruv, Jihoon Kim, Youngjoon Suh, Jaiyoung Ryu, Yoonjin Won, and Aparna Chandramowlishwaran. Bubbleml: A multi-physics dataset and benchmarks for machine learning. *arXiv preprint arXiv:2307.14623*, 2023.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

Maximilian Herde, Bogdan Raonic, Tobias Rohner, Roger Käppeli, Roberto Molinaro, Emmanuel de Bézenac, and Siddhartha Mishra. Poseidon: Efficient foundation models for pdes. *Advances in Neural Information Processing Systems*, 37:72525–72624, 2024.

Julius O Hinze. Fundamentals of the hydrodynamic mechanism of splitting in dispersion processes. *AIChE journal*, 1(3):289–295, 1955.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Nils Hoppe, Josef M Winter, Stefan Adami, and Nikolaus A Adams. Alpaca-a level-set based sharp-interface multiresolution solver for conservation laws. *Computer Physics Communications*, 272: 108246, 2022.

Guang-Shan Jiang and Chi-Wang Shu. Efficient implementation of weighted eno schemes. *Journal of computational physics*, 126(1):202–228, 1996.

JWJ Kaiser, JM Winter, S Adami, and NA Adams. Investigation of interface deformation dynamics during high-weber number cylindrical droplet breakup. *International Journal of Multiphase Flow*, 132:103409, 2020.

Georg Kohl, Li-Wei Chen, and Nils Thuerey. Benchmarking autoregressive conditional diffusion models for turbulent flow simulation. *arXiv preprint arXiv:2309.01745*, 2023.

Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Learning maps between function spaces with applications to pdes. *Journal of Machine Learning Research*, 24(89):1–97, 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*, 2020.

Yu Liang, Yazhong Jiang, Chih-Yung Wen, and Yao Liu. Interaction of a planar shock wave and a water droplet embedded with a vapour cavity. *Journal of Fluid Mechanics*, 885:R6, 2020.

Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747*, 2022.

Ze Liu, Han Hu, Yutong Lin, Zhuliang Yao, Zhenda Xie, Yixuan Wei, Jia Ning, Yue Cao, Zheng Zhang, Li Dong, et al. Swin transformer v2: Scaling up capacity and resolution. 2022 ieee. In *CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11999–12009, 2021.

Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A convnet for the 2020s. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 11976–11986, 2022.

Yining Luo, Yingfa Chen, and Zhen Zhang. Cfdbench: A large-scale benchmark for machine learning methods in fluid dynamics. *arXiv preprint arXiv:2310.05963*, 2023.

Kazuki Maeda, Wayne Kreider, Adam Maxwell, Bryan Cunitz, Tim Colonius, and Michael Bailey. Modeling and experimental analysis of acoustic cavitation bubbles for burst wave lithotripsy. In *Journal of Physics: Conference Series*, volume 656, pp. 012027. IOP Publishing, 2015.

Jomela C Meng and Tim Colonius. Numerical simulation of the aerobreakup of a water droplet. *Journal of Fluid Mechanics*, 835:1108–1135, 2018.

Thomas Paula, Stefan Adami, and Nikolaus A Adams. Analysis of the early stages of liquid-water-drop explosion by numerical simulation. *Physical Review Fluids*, 4(4):044003, 2019.

Thomas Paula, Stefan Adami, and Nikolaus A Adams. A robust high-resolution discrete-equations method for compressible multi-phase flow with accurate interface capturing. *Journal of Computational Physics*, 491:112371, 2023.

William Peebles and Saining Xie. Scalable diffusion models with transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 4195–4205, 2023.

Farbod Riahi, Alexander Bußmann, Carlos Doñate-Buendia, Stefan Adami, Nicolaus A Adams, Stephan Barcikowski, and Bilal Gökce. Characterizing bubble interaction effects in synchronous-double-pulse laser ablation for enhanced nanoparticle synthesis. *Photonics Research*, 11(12): 2054–2071, 2023.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241. Springer, 2015.

Mehdi Shadkhah, Ronak Tali, Ali Rabeh, Cheng-Hau Yang, Ethan Herron, Abhisek Upadhyaya, Adarsh Krishnamurthy, Chinmay Hegde, Aditya Balu, and Baskar Ganapathysubramanian. Mpf-bench: A large scale dataset for sciml of multi-phase-flows: Droplet and bubble dynamics. *arXiv preprint arXiv:2502.07080*, 2025.

Shubham Sharma, Awanish Pratap Singh, S Srinivas Rao, Aloke Kumar, and Saptarshi Basu. Shock induced aerobreakup of a droplet. *Journal of Fluid Mechanics*, 929:A27, 2021.

Claudiu A Stan, Despina Milathianaki, Hartawan Laksmono, Raymond G Sierra, Trevor A McQueen, Marc Messerschmidt, Garth J Williams, Jason E Koglin, Thomas J Lane, Matt J Hayes, et al. Liquid explosions induced by x-ray laser pulses. *Nature Physics*, 12(10):966–971, 2016a.

Claudiu A Stan, Philip R Willmott, Howard A Stone, Jason E Koglin, Mengning Liang, Andrew L Aquila, Joseph S Robinson, Karl L Gumerlock, Gabriel Blaj, Raymond G Sierra, et al. Negative pressures and spallation in water drops subjected to nanosecond shock waves. *The journal of physical chemistry letters*, 7(11):2055–2062, 2016b.

Makoto Takamoto, Timothy Praditia, Raphael Leiteritz, Daniel MacKinlay, Francesco Alesiani, Dirk Pflüger, and Mathias Niepert. Pdebench: An extensive benchmark for scientific machine learning. *Advances in Neural Information Processing Systems*, 35:1596–1611, 2022.

TG Theofanous and GJ Li. On the physics of aerobreakup. *Physics of fluids*, 20(5), 2008.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Josef Winter, Jakob Kaiser, Stefan Adami, and Nikolaus Adams. Numerical investigation of 3d drop-breakup mechanisms using a sharp interface level-set method. In *11th International Symposium on Turbulence and Shear Flow Phenomena, TSFP 2019*, 2019.

A DATASETS

As mentioned in the main text, we solve the compressible multiphase Euler equations (Equation 1) with the RDEMIC, which captures the interface as a diffuse zone on a Cartesian grid. This method combines the solutions of pairwise Riemann problems to obtain the finite-volume flux. By a modified partitioning of the Riemann solutions and a specific combination of fluxes and non-conservative terms, the method is made practically applicable for high-resolution shock-interface problems (Paula et al., 2023). We use this method in ALPACA (Hoppe et al., 2022), which is a well-suited environment for compressible single-phase simulations and other multi-phase methods, although originally developed as a level-set-based sharp-interface solver. Its standout features include a wide variety of Riemann solvers, high-resolution reconstruction schemes, and a state-of-the-art multiresolution algorithm for high computational efficiency. In both datasets in this study, the cell face fluxes are reconstructed with the fifth-order WENO scheme (Jiang & Shu, 1996). Furthermore, a third-order Runge-Kutta Total Variation Diminishing scheme is applied for time discretization (Gottlieb & Shu, 1998).

To close the governing equation (Equation 1), an Equation of State (EOS) is used, which relates pressure to density and internal energy. We adopt the stiffened-gas EOS to generate both datasets, which reads

$$p(\rho, e) = (\gamma - 1)\rho e - \gamma p_{\text{stiff}} \quad \Longleftrightarrow \quad e(\rho, p) = \frac{p + \gamma p_{\text{stiff}}}{(\gamma - 1)\rho}, \tag{3}$$

with p being the pressure of the fluid, ρ the mass density, e the internal energy, γ the model constant. In addition, p_{stiff} accounts for a pre-compression of the fluid. To degenerate the aforementioned equation to an ideal-gas EOS for air, we adopt $\gamma = 1.4$ and $p_{\text{stiff}} = 0$. The total energy density, $E[\frac{J}{kg}]$, is obtained by considering internal energy from Equation 3 and kinetic energy, as shown in Equation 4:

$$E = \rho e + 1/2\rho(u_r^2 + u_z^2) \tag{4}$$

Here, u_r and u_z are the velocity components in the r and z directions, respectively. Schlieren $[\frac{kg}{m^4}]$ is computed in the solver by Equation 5:

$$\text{schlieren} = \nabla\rho \tag{5}$$

Additionally, vorticity $[s^{-1}]$ is defined in Equation 6:

$$\text{vorticity} = \nabla \times \mathbf{u} \tag{6}$$

A.1 THE LIDE DATASET

To simulate this problem, careful considerations must be taken into account. The filament along the centerline, which is heated by a laser in a very short time, is pre-initialized with vapor instead of liquid water. However, it is important to note that the density of the vapor in this zone remains equal to that of liquid water, since the laser energy heats the liquid rapidly. Considering that different laser pulse energies result in different pressures in the filament (p_{filament}), we cover a range from 10^8 to 10^{10} [Pa] in our dataset. Alongside the high-pressure, the ambient pressure (p_{ambient}) varies between 10^5 and 10^6 [Pa]. In addition, the laser half-width changes in the range of 2×10^{-7} to 1.5×10^{-6} [m]. The droplet radius along the r and z axes varies from 1×10^{-5} to 1.6×10^{-5} [m]. A summary of initial condition values is presented in Table 5.

Validation. We compare the evolution of the droplet diameter in the radial direction to validate our dataset against experiments (Stan et al., 2016b). According to experimental observations, the droplet starts to expand upon the arrival of the radial shock wave, which is induced by high pressure in the filament. Due to the wave interactions, a decrease in the expansion rate is observed, which is again followed by an increase. This trend is depicted in Figure 7 and is in good agreement with experiments.

Table 5: Initial conditions for the LIDE dataset

Phase-l	ρ_l [kg m^{-3}]	$\{u_{r,l}, u_{z,l}\}$ [m s^{-1}]	p_l [Pa]	
1 (Ambient air)	0.74	0.0, 0.0	$p_{ambient}$	
2 (Liquid droplet)	998.2	0.0, 0.0	$p_{ambient}$	$z >$ laser half-width
	998.2	0.0, 0.0	$p_{filament}$	$z <$ laser half-width

In this problem, it is crucial to analyze and understand the wave interactions inside the droplet. After rapid energy deposition along the centerline, the main shock spreads radially, approaching the droplet surface. The corresponding reflection results in a curved negative-pressure wave, which increases the tension. Shortly after, this wave collapses toward the z-axis and impacts the motion of the droplet's surface (Paula et al., 2019). These phenomena are depicted step-by-step in Figure 8.

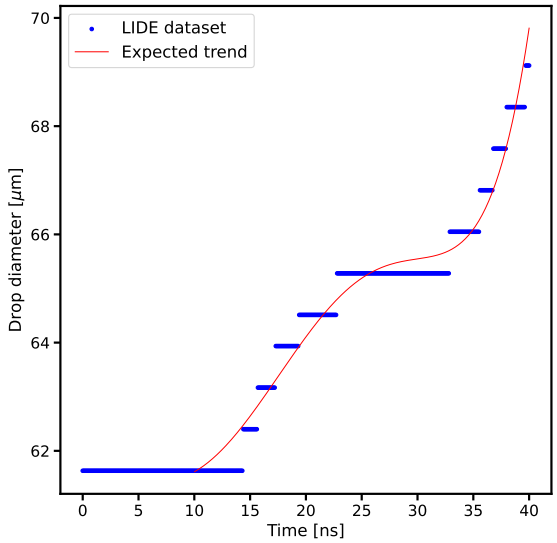

Figure 7: Validation of the LIDE data compared to the expected trend (Stan et al., 2016b).

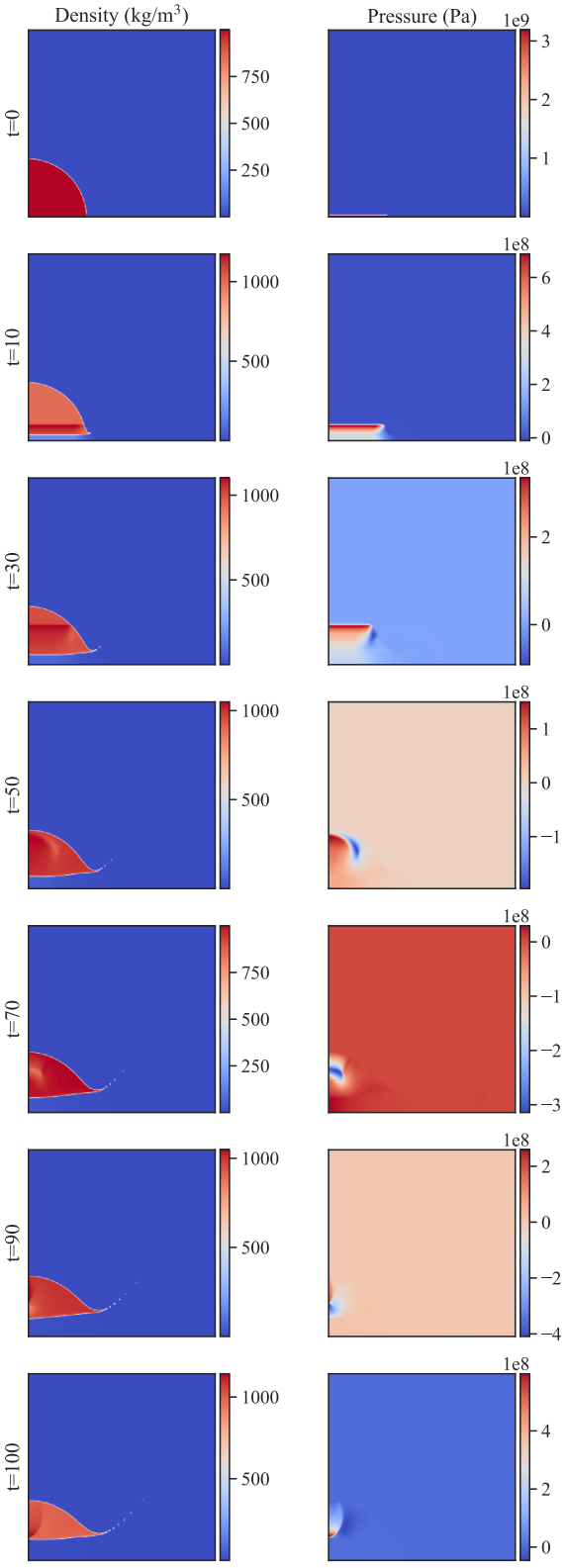

Figure 8: Visualization of droplet's motion and deformation in the LIDE dataset.

A.2 THE SIDA DATASET

To get a better understanding of this problem, both wave dynamics and droplet breakup modes are studied extensively in the literature (Sharma et al., 2021), (Theofanous & Li, 2008). Breakup modes are characterized by the Weber number (Hinze, 1955), which is defined as follows:

$$We = \frac{\rho_2 u_2^2 d}{\sigma} \tag{7}$$

In this definition, ρ_2 and u_2 refer to post-shock density and velocity of the external flow, respectively. Additionally, d is the droplet diameter and σ denotes the surface tension coefficient. For droplet aero-breakup, two major breakup modes are introduced: Rayleigh-Taylor Piercing (RTP) and shear-induced entrainment (SIE). RTP is the main instability mode for small Weber numbers (starting at $We \approx 28$), and SIE is the terminal instability mode for increasing Weber numbers ($We > 10^3$) (Theofanous & Li, 2008). For this study, we cover the Weber number in the range [530, 40000], which corresponds to the transition regions from RTP to SIE and also the SIE region itself.

After the shock impact, the post-shock flow plays a significant role in droplet deformation and breakup. The post-shock flow regime is identified by the Mach number, which is a non-dimensional parameter that relates flow velocity to the speed of sound. We compute the post-shock flow properties using the normal shock relation. These relations are given by (Anderson, 1990):

$$u_s = M_s \cdot c_1 \tag{8}$$

$$u_{1,\text{rel}} = -u_s \tag{9}$$

$$u_1 = u_{1,\text{rel}} + u_s \tag{10}$$

$$T_2 = T_1 \left(1 + \frac{2\gamma\left(M_s^2 - 1\right)}{\gamma + 1}\right)\left(\frac{2 + (\gamma - 1)M_s^2}{(\gamma + 1)M_s^2}\right) \tag{11}$$

$$c_2 = \sqrt{\gamma \cdot R_1 \cdot T_2} \tag{12}$$

$$M_{f2,\text{rel}} = \sqrt{\frac{1 + \frac{\gamma-1}{2}M_s^2}{\gamma M_s^2 - \frac{\gamma-1}{2}}} \tag{13}$$

$$u_{2,\text{rel}} = M_{f2,\text{rel}} \cdot c_2 \tag{14}$$

$$u_2 = u_s - u_{2,\text{rel}} \tag{15}$$

$$\rho_2 = \rho_1 \cdot \frac{(\gamma + 1)M_s^2}{2 + (\gamma - 1)M_s^2} \tag{16}$$

$$p_2 = p_1 \left(1 + \frac{2\gamma\left(M_s^2 - 1\right)}{\gamma + 1}\right) \tag{17}$$

We use M_s for the shock and M_f for the post-shock flow Mach number. The flow states before and after the shock wave are referred to with subscripts 1 and 2, respectively. Furthermore, T is the temperature, c is the speed of sound, $\gamma = \frac{c_p}{c_v}$ is the ratio of specific heat, and R is the specific gas constant. We consider shock Mach numbers spanning from 1.2 to 3.5. Then, based on the selected shock Mach number, we calculate ρ_2, u_2, and p_2 for the west Dirichlet boundary condition. Next, the surface tension coefficient is computed from the Weber number. A summary of initial condition values is presented in Table 6. It should be noted that the value $16.5R_0$ in the table, shows the location of the shock wave in the initial setup (refer to Figure 2).

Validation. We compare the SIDA dataset against numerical studies. Since we employ an axisymmetric setup in our simulation, a full three-dimensional study is referenced for validation (Winter

Table 6: Initial conditions for the SIDA dataset

l	ρ_l [kg m^{-3}]	$\{u_{r,l}, u_{z,l}\}$ [m s^{-1}]	p_l [Pa] $\mathbf{p_l}$ [Pa]	
1 (Ambient air)	ρ_2	0.0, u_2	p_2	$z < 16.5R_0$
	1.2	0.0, 0.0	101325.0	$z > 16.5R_0$
2 (Liquid droplet)	998.2	0.0, 0.0	101325.0	

et al., 2019), (Meng & Colonius, 2018). For this purpose, the non-dimensional time (t^*) and displacement of the center of mass (COM) in the droplet (Δz^*) are defined as

$$t^* = t\,\frac{u_2}{d}\sqrt{\frac{\rho_2}{\rho_{drop}}}, \tag{18}$$

and

$$\Delta z^* = \frac{z}{d}, \tag{19}$$

where t is the saved timestep, and ρ_{drop} is density of the liquid droplet. Upon shock and post-shock flow impact, the droplet COM accelerates. This trend is clearly observable in our dataset, which aligns with results from the literature. In Figure 10, the flattening of the droplet surface and the hat-shaped deformation are shown. Noteworthy, the perturbations on the surface of the droplet are related to shear-induced instabilities (Sharma et al., 2021).

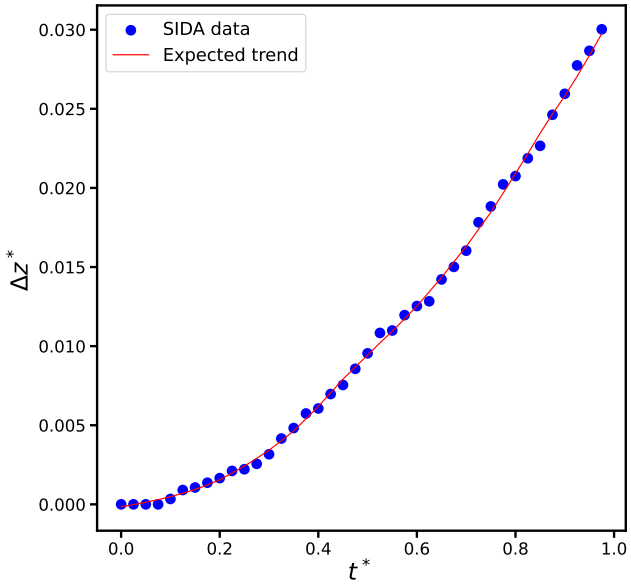

Figure 9: Validation of SIDA data against numerical studies (Winter et al., 2019),(Meng & Colonius, 2018).

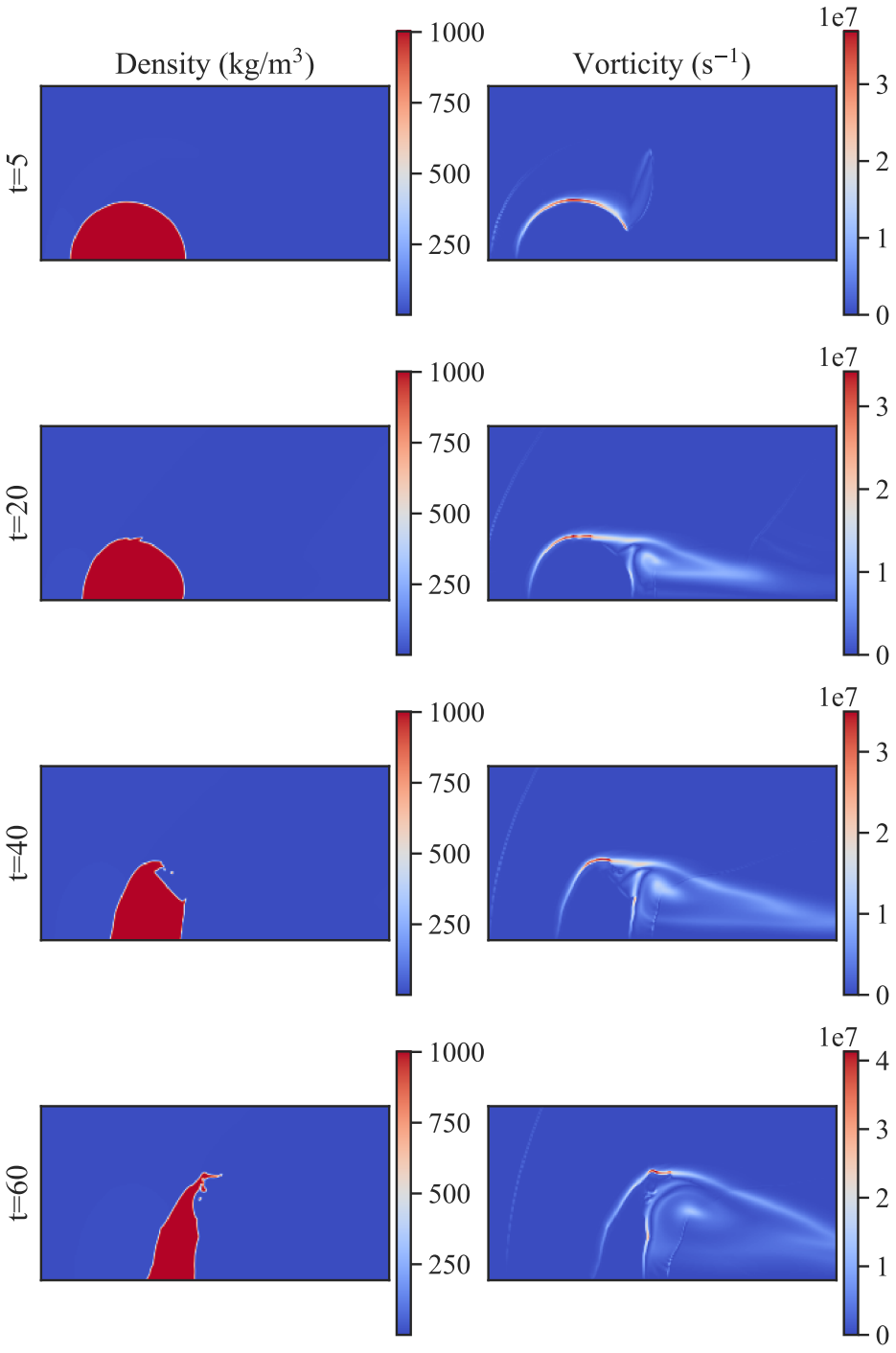

Figure 10: Visualization of droplet's motion and deformation in the SIDA dataset.

B EXPERIMENT DETAILS

B.1 DESIGN OF EXPERIMENTS

The complete set of experiments for the SIDA and the LIDE datasets is shown in Tables 7 and 8, respectively. We experiment with a variety of input, conditioning, and output channels, along with the combinations of sequence info and conditioning parameters.

Table 7: SIDA (PDUVVoS) experiments with Tag identifiers.

# Expt	Tag	Input Channels	Output Channels	Cond[4]	Seq Info [5]
1	PDUV_F_(1,1)	Pressure, Density X-velocity, Y-Velocity	Pressure, Density X-Velocity, Y-Velocity	F	1, 1, 5
2	PDUV_F_(3,1)	Pressure, Density X-velocity, Y-velocity	Pressure, Density Velocity_x, Velocity_y	F	3, 1, 5
3	PDUV_T_(3,1)	Pressure, Density X-velocity, Y-velocity	Pressure, Density X-velocity, Y-velocity	T	3, 1, 5
4	PDUV[VoS]_F_(1,1)	Pressure, Density X-velocity, Y-velocity [Vorticity, Schlieren]	Pressure, Density X-velocity, Y-velocity	F	1, 1, 5
5	PDUV[VoS]_F_(3,1)	Pressure, Density X-velocity, Y-velocity [Vorticity, Schlieren]	Pressure, Density X-velocity, Y-velocity	F	3, 1, 5
6	PDUV[VoS]_T_(3,1)	Pressure, Density X-velocity, Y-velocity [Vorticity, Schlieren]	Pressure, Density X-velocity, Y-velocity	T	3, 1, 5
7	PDUV_F_(3,2)	Pressure, Density X-velocity, Y-velocity	Pressure, Density X-velocity, Y-velocity	F	3, 2, 5
8	PDUV[VoS]_F_(3,2)	Pressure, Density X-velocity, Y-velocity [Vorticity, Schlieren]	Pressure, Density X-velocity, Y-velocity	F	3, 2, 5

[4]refers to the boolean flag indicating whether conditioning parameters are injected into the normalization layer.

[5]refers to the sequence information: [number of historic inputs, number of bundled predictions, stride between timesteps].

Table 8: LIDE (PDUVES) experiments with Tag identifiers.

# Expt	Tag	Input Channels	Output Channels	Cond	Seq Info
1	P_F_(1,1)	Pressure	Pressure	F	1, 1, 10
2	P_F_(3,1)	Pressure	Pressure	F	3, 1, 10
3	P_T_(1,1)	Pressure	Pressure	T	1, 1, 10
4	P_T_(3,1)	Pressure	Pressure	T	3, 1, 10
5	PDUV_F_(1,1)	Pressure, Density X-velocity, Y-velocity	Pressure, Density X-velocity, Y-velocity	F	1, 1, 10
6	PDUV_F_(3,1)	Pressure, Density X-velocity, Y-velocity	Pressure, Density X-velocity, Y-velocity	F	3, 1, 10
7	P[ES]_F_(1,1)	Pressure, [Energy, Schlieren]	Pressure	F	1, 1, 10
8	P[ES]_F_(3,1)	Pressure [Energy, Schlieren]	Pressure	F	3, 1, 10
9	PDUV[ES]_F_(1,1)	Pressure, Density X-velocity, Y-velocity [Energy, Schlieren]	Pressure, Density X-velocity, Y-velocity	F	1, 1, 10
10	PDUV[ES]_F_(3,1)	Pressure, Density X-velocity, Y-velocity [Energy, Schlieren]	Pressure, Density X-velocity, Y-velocity	F	3, 1, 10
11	P_F_(3,2)	Pressure	Pressure	F	3, 2, 10
12	PDUV_F_(3,2)	Pressure, Density X-velocity, Y-velocity	Pressure, Density X-velocity, Y-velocity	F	3, 2, 10
13	PDUV_T_(1,1)	Pressure, Density X-velocity, Y-velocity	Pressure, Density X-velocity, Y-velocity	T	1, 1, 10
14	PDUV_T_(3,1)	Pressure, Density X-velocity, Y-velocity	Pressure, Density X-velocity, Y-velocity	T	3, 1, 10

B.2 Baseline Model Details

In this section, we provide a brief overview of all the models used as baselines. In all the models described in this section, the LayerNorm (Ba et al., 2016) is used as the default choice of normalization layer, and the normalized grid X- and Y-coordinates are appended as additional channels with the input channels.

1. **UNet:** We implement the UNet variant as described in Gupta & Brandstetter (2022). UNets follow a structure that first performs spatial downsampling and then spatial upsampling, with each block composed of multiple convolutional layers. A distinctive feature of UNet is the inclusion of skip connections that link activations from the downsampling path to their corresponding upsampling layers. Table 9 shows the hyperparameters chosen for the two model parameter categories. The number of latent channels corresponds to the feature dimension produced after the first convolutional layer. Along the downsampling path, the base latent channel dimension is adjusted according to a channel multiplier list, with each element specifying the factor used to increase the number of channels at successive levels of the model.

Table 9: UNet hyperparameters.

Hyperparameters	**1M**	**50M**
Latent channels	28	48
Channel Multiplier	[1,4]	[1,2,2,4]
Activation	GELU	GELU

2. **Residual Network (ResNet):** The baseline ResNet is implemented as described in Gupta & Brandstetter (2022), where no up- or down-projection techniques have been used. The input channels are projected to the latent channels by a convolutional layer and subsequently passed through four ResNet blocks. Each block consists of two 3x3 convolutional layers, each followed by an activation function and a norm layer. The convolutional layers employ a stride and padding of 1, preserving the spatial resolution of the feature maps. The final output is then obtained by adding the original input to the convolutional output. Refer to Table 10 for the hyperparameters.

Table 10: ResNet hyperparameters.

Hyperparameters	**1M**
Latent channels	112
# residual blocks	[1, 1, 1, 1]
Activation	GELU

3. **Fourier Neural Operator (FNO):** The FNO is designed to approximate mappings between function spaces by performing computations directly in the Fourier domain. Its architecture can be divided into three main components: a lifting network, a sequence of Fourier layers, and a decoder network. We adopt the implementation described in Contributors (2023) and use the hyperparameters as shown in Table 11 for our experiments.

   The lifting network first maps the input channels into a higher-dimensional latent space using pointwise convolutions. The dimension of this latent space is described by the latent channels. The core of the model is composed of Fourier layers that have spectral convolution with a point-wise linear convolution layer acting as a skip connection. The activation is applied to the summation of the spectral convolutions and this convolutional skip layer. In each spectral convolution, the input is transformed into the Fourier domain using Fast Fourier Transform (FFT), where a specified number of modes are retained and updated with

learned complex weights, and the result is projected back to the spatial domain through the decoder network.

Table 11: FNO hyperparameters.

Hyperparameters	1M	50M
Latent channels	16	32
FNO Layers	4	6
Modes	16	45
Padding	8	8
Padding Type	constant	constant
Activation in Fourier Layers	GELU	GELU
Decoder layers	2	2
Decoder layers size	128	256
Decoder activation	SiLU	SiLU

4. **Vision Transformer (ViT):** A modified ViT (Dosovitskiy et al., 2020) architecture was adopted. The implementation follows the general ViT paradigm, splitting the image into square patches of size 8, embedding and passing them through a transformer encoder, and reconstructing the spatial output from the resulting latent representations with the additional capability to handle non-square inputs. The ViT model consists of a patch-based embedding, an encoder, and a decoder. Passing the input through the embedding-encoder-decoder pipeline results in a reconstruction of the original input shape. The Embedding divides the image into non-overlapping patches, embeds them via a linear projection, and adds positional encodings. For each patch, this results in a sequence of token vectors, each with dimensions specified by the latent channels. The transformer encoder processes this sequence using standard Multi-Head Self-Attention (MHSA) and feedforward layers, with the hidden size denoted by the intermediate size variable. The number of the hidden layers determines the number of the encoder layers. The number of MHSA in each layer is specified by the number of attention heads. This attention stage allows global spatial interactions across the patch grid, enabling the model to learn long-range dependencies. Table 12 shows the hyperparameters for the two learnable parameter categories.

Table 12: ViT hyperparameters.

Hyperparameters	1M	50M
Latent channels	128	504
Patch size	8	8
# hidden layers	2	12
# attention heads	4	14
intermediate size	512	1024
Activation	GELU	GELU

5. **Scalable Operator Transformer (ScOT):** The ScOT model is based on the Poseidon framework (Herde et al., 2024). At its core, ScOT adopts a hierarchical transformer architecture inspired by vision transformers with a window-based approach. The input is partitioned into a uniform grid of non-overlapping patches. We implement an additional capability to process non-square inputs. Each patch undergoes an averaging operation using a shared spatial weight matrix, followed by a linear projection into a latent embedding space, whose size is described by the latent channels. This procedure produces a piecewise-constant latent function representation over the domain, which serves as the input to the

transformer backbone. The motivation for this patch-based embedding is to reduce the computational complexity associated with global attention while preserving essential local information about the input field.

Once embedded, the representation is processed through a series of hierarchical SwinV2 Transformer blocks (Liu et al., 2021), arranged in multiple stages that progressively downsample and subsequently upsample the latent feature maps, forming a UNet-like architecture. The number of blocks per stage is defined by the variable 'depths' in Table 13. Each stage applies windowed MHSA, where attention computations are restricted to local windows rather than the entire spatial domain, significantly reducing the quadratic cost of global attention. The number of parallel MHSA per stage is determined by the number of attention heads. To ensure information exchange across windows and avoid locality bias, the attention windows are shifted between consecutive layers, enabling effective global context modeling over multiple layers.

The hierarchical design incorporates patch merging operations during the encoder phase to reduce spatial resolution and increase the feature dimension, thereby allowing deeper layers to capture global structures. Conversely, the decoder phase employs patch expansion to restore resolution, and skip connections in the form of ConvNext blocks (Liu et al., 2022), bridging the corresponding encoder and decoder stages. The number of blocks per stage in the ConvNext blocks is specified by the hyperparameter 'skip-connections'.

Table 13: ScOT hyperparameters.

Hyperparameters	**1M**	**50M**
Latent channels	27	150
Patch size	4	4
Depths	[3,3,3]	[4,4,4]
# attention heads	[3,6,12]	[6,12,24]
Skip connections	[2,2,0]	[3,3,0]
Window size	16	16
MLP ratio	2.0	4.0
Activation	GELU	GELU

B.3 Conditioning

In this section, we describe the formulation of the strategy used to integrate conditioning parameters into the model (Herde et al., 2024). For an input $x \in \mathbb{R}^d$, and k being the conditioning parameters, the conditional layer norm formulation is given by Equation 20. Figure 11 illustrates this injection of conditioning parameters into the layer norm. Here γ and β are simple Multilayer Perceptrons (MLPs).

$$\begin{aligned} LayerNorm_{\gamma(k),\beta(k)}(\mathbf{x}) &= \gamma(k) \odot \frac{\mathbf{x} - \mu_{\mathbf{x}}(x)}{\sqrt{\sigma^2_{\mathbf{x}}(x) + \epsilon}} + \beta(k), \\ \mu_{\mathbf{x}}(x) = \frac{1}{d}\sum_{j=1}^{d} x_j, &\quad \sigma^2_{\mathbf{x}}(x) = \frac{1}{d}\sum_{j=1}^{d} \big(x_j - \mu_{\mathbf{x}}(x)\big)^2. \end{aligned} \tag{20}$$

B.4 Training Hyperparameters

Table 14 shows the training hyperparameters that are common for all the models. Each model has its own specific hyperparameters, which are described in Appendix B.2. All models were trained on NVIDIA RTX A6000 48GB GPU with bf16 mixed-precision, except for the FNO, which was trained on fp32.

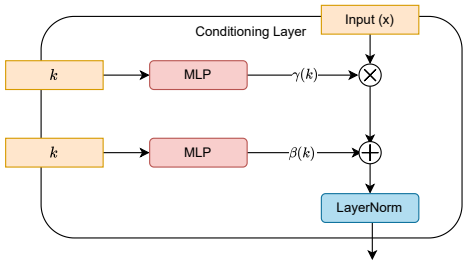

Figure 11: Conditional LayerNorm.

Table 14: Training hyperparameters.

Hyperparameter	**Value**
Number of Epochs	128
Batch Size	32
Optimizer	AdamW
Weight Decay	0.000001
Learning Rate(LR)	0.00005
LR Scheduler	Cosine
Warmup Ratio	0.0
Mix-precision	bf16 (except FNO: fp32)

B.5 Error Metrics

In this study, we employ two commonly used error measures: the Root Mean Square Error (RMSE) and the Mean Absolute Error (MAE).

The RMSE (Equation 21) measures the square root of the mean squared difference between predictions and ground-truth values, penalizing larger errors more strongly.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2} \tag{21}$$

The MAE (Equation 22) measures the average magnitude of the absolute prediction errors.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - \hat{Y}_i\right| \tag{22}$$

where Y_i denotes the ground-truth values, $\hat{Y}_i$ the corresponding model predictions, and n the total number of samples.

C RESULTS

The remaining plots for the dataset LIDE are depicted in this section. The corresponding result for the analysis of including more historic inputs is given by Figure 12.

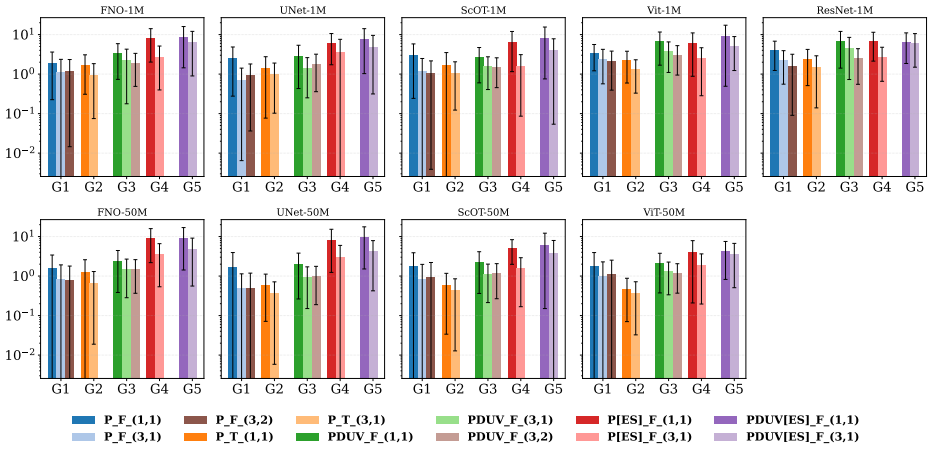

Figure 12: Effect of sequence information for the LIDE dataset. Error-type 2 is presented.

Moreover, the effect of the implementation of conditioning fields is shown in Figure 13.

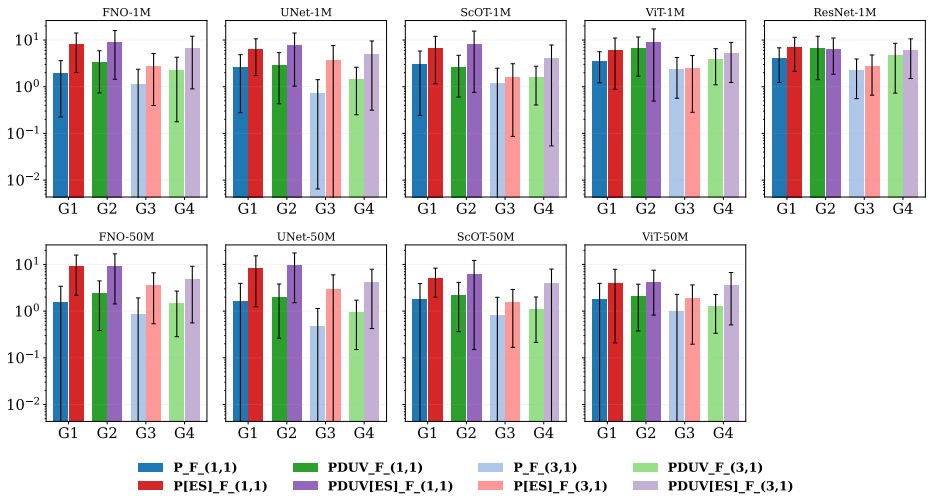

Figure 13: Effect of conditioning fields for the SIDA dataset. Error-type 2 is presented.

D INFERNCE ROLLOUT PLOTS

D.1 ROLLOUT PREDICTIONS FROM INITIAL CONDITIONS FOR THE LIDE DATASET

In the following, we present rollout predictions for various models—each with 50M parameters, except for ResNet, which has only 1M parameter count. The trajectory shown in the Figures 14, 15, 16, 17, and 18. The trajectory corresponds to the following simulation parameters: filament pressure 9.3886×10^9 [Pa], ambient pressure 1.0382×10^5 [Pa], laser half-width 1.1727×10^{-6} [m], and droplet radii 1.5966×10^{-5} [m] and 1.2139×10^{-5} [m] along z- and r-axis, respectively.

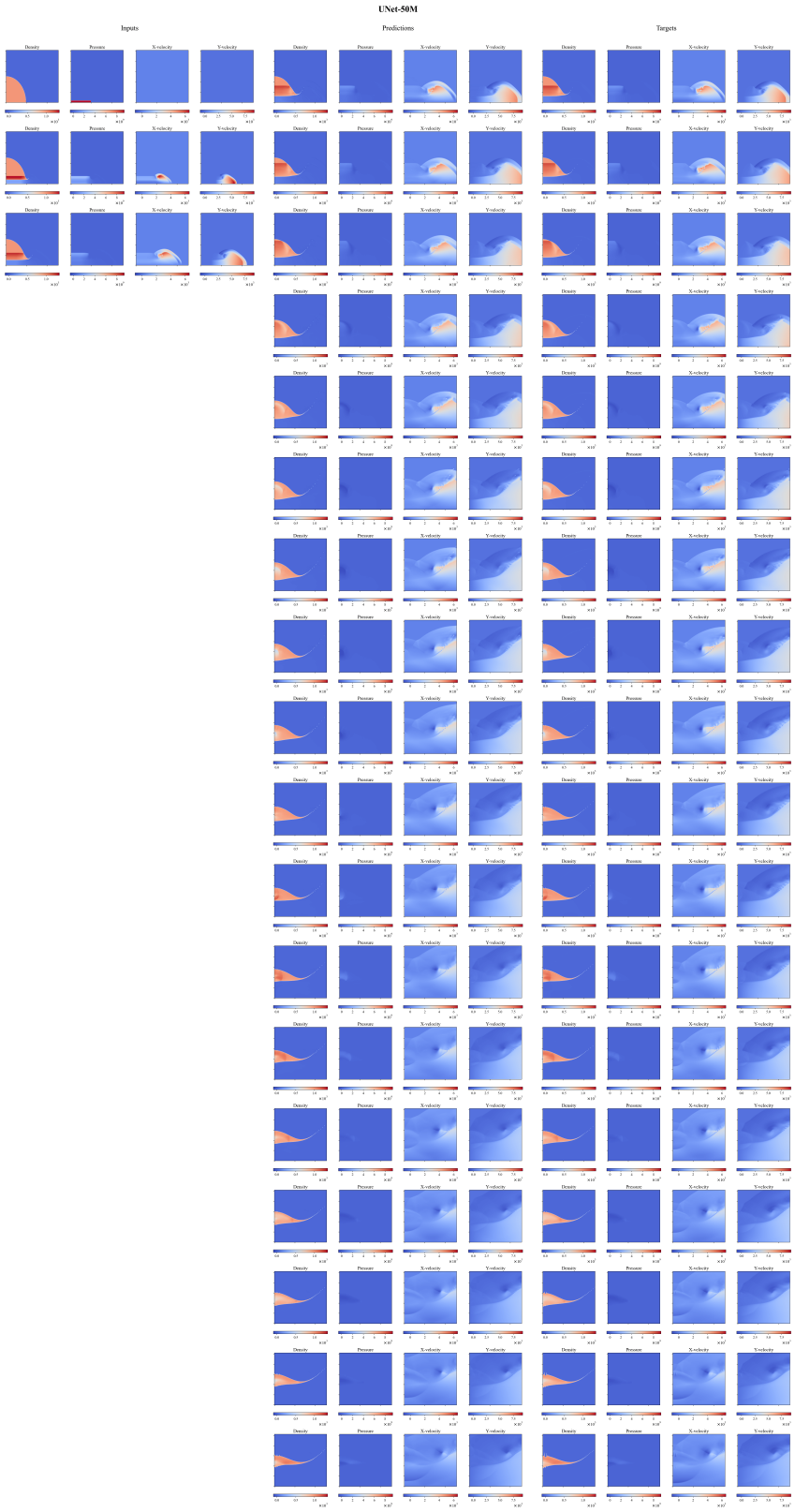

Figure 14: Rollout predictions for the LIDE-Experiment PDUV_F_(3,1) with UNet-50M.

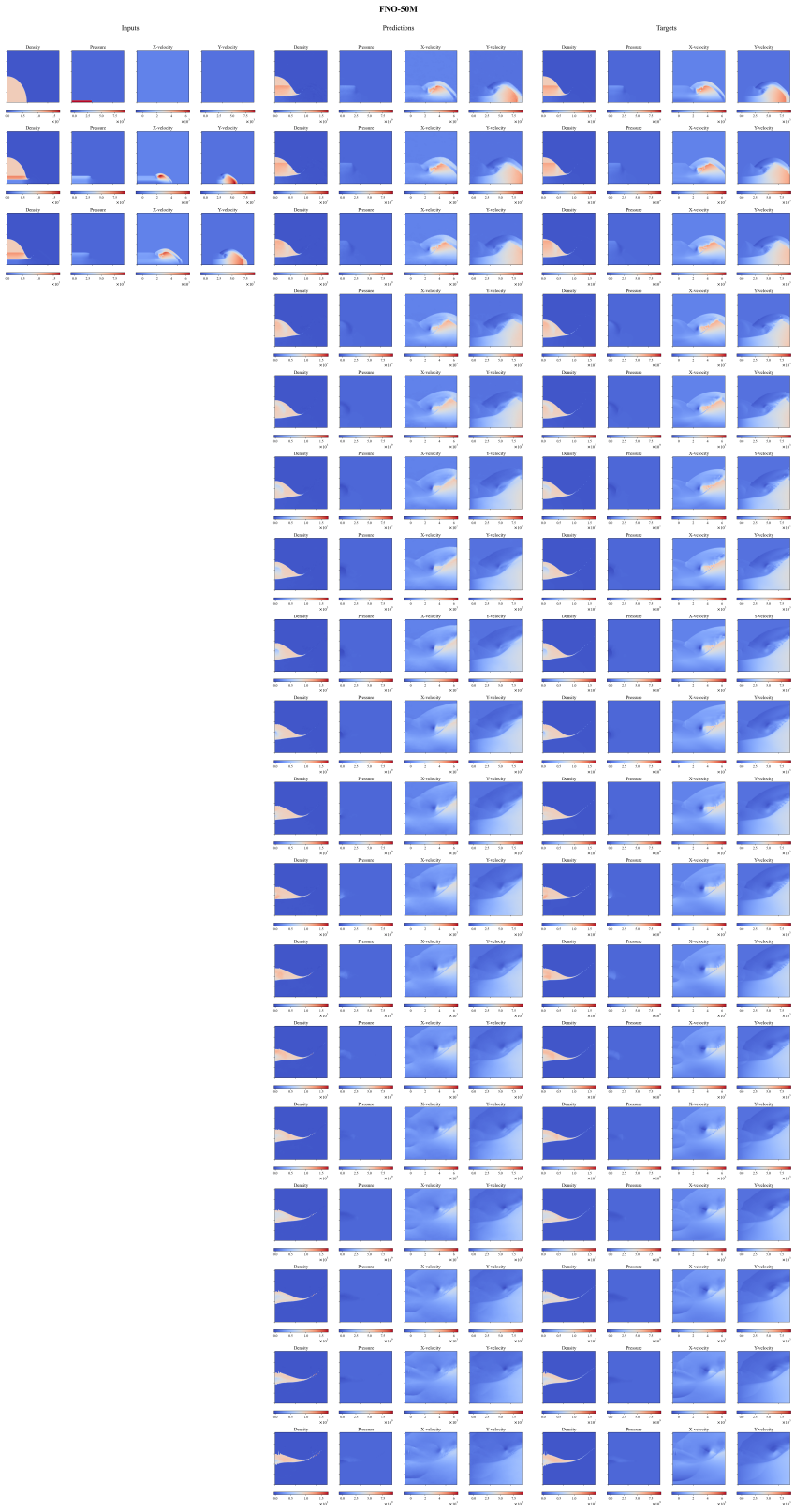

Figure 15: Rollout predictions for the LIDE-Experiment PDUV_F_(3,1) with FNO-50M.

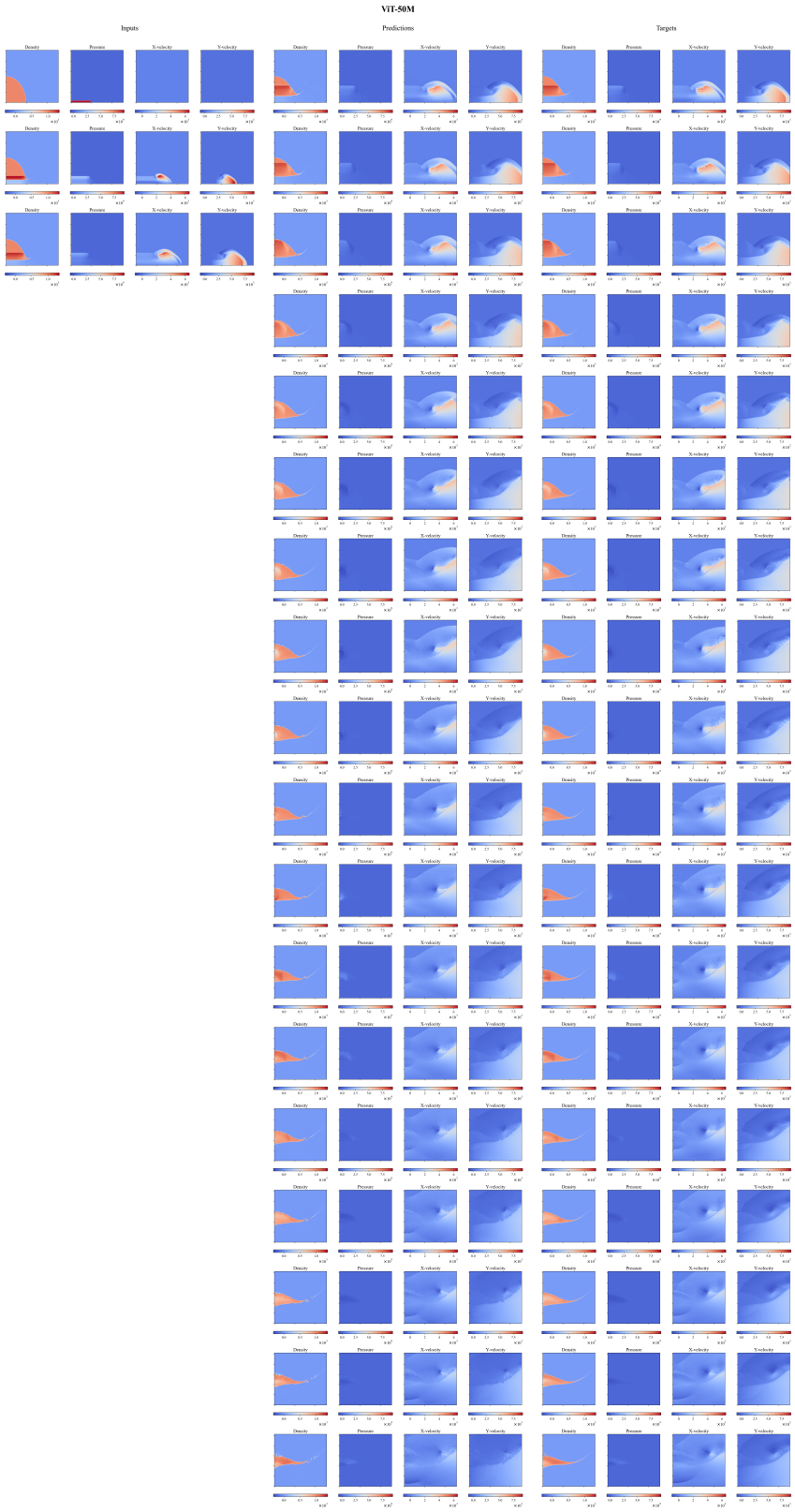

Figure 16: Rollout predictions for the LIDE-Experiment PDUV_F_(3,1) with ViT-50M.

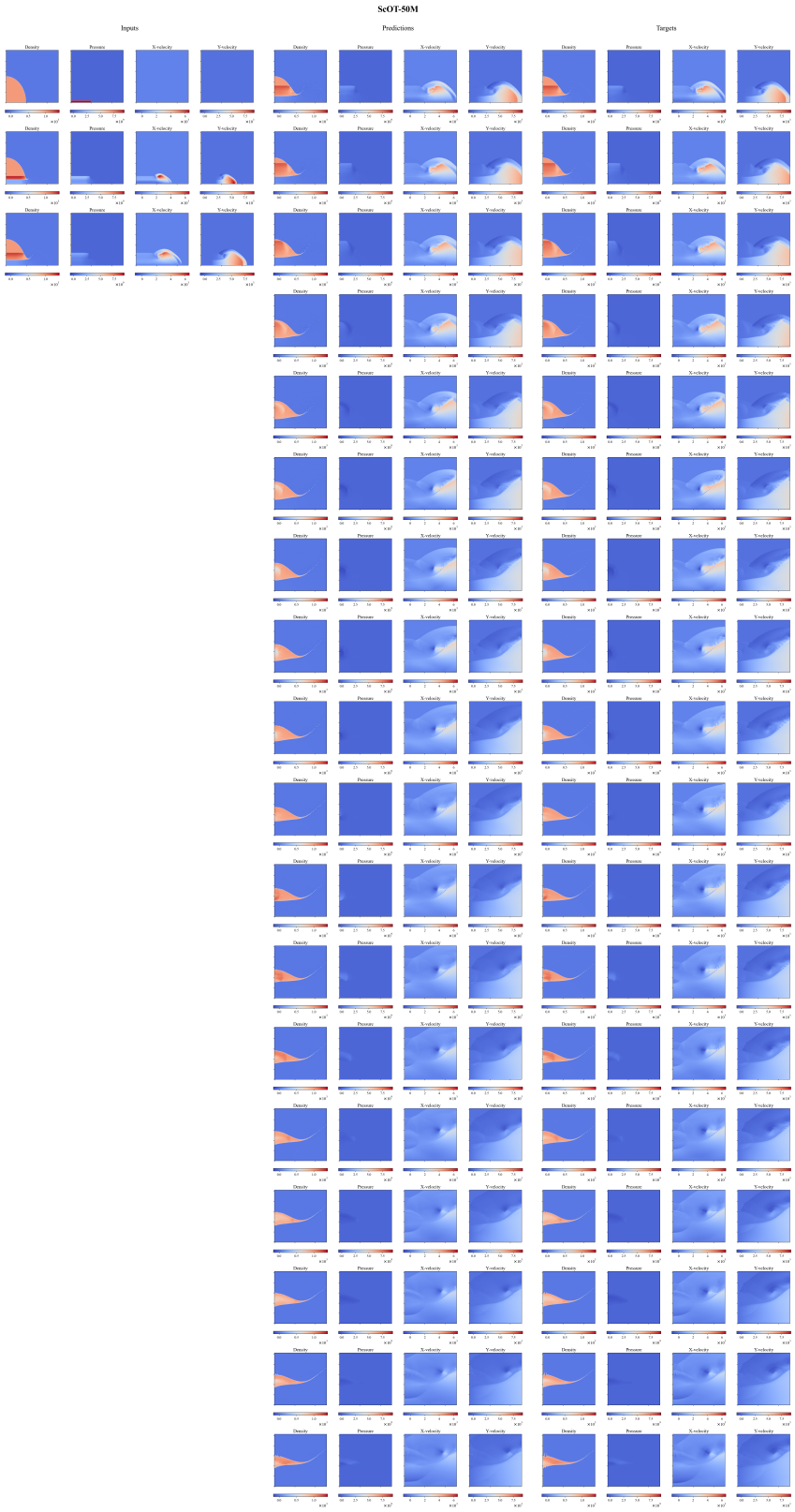

Figure 17: Rollout predictions for the LIDE-Experiment PDUV_F_(3,1) with ScOT-50M.

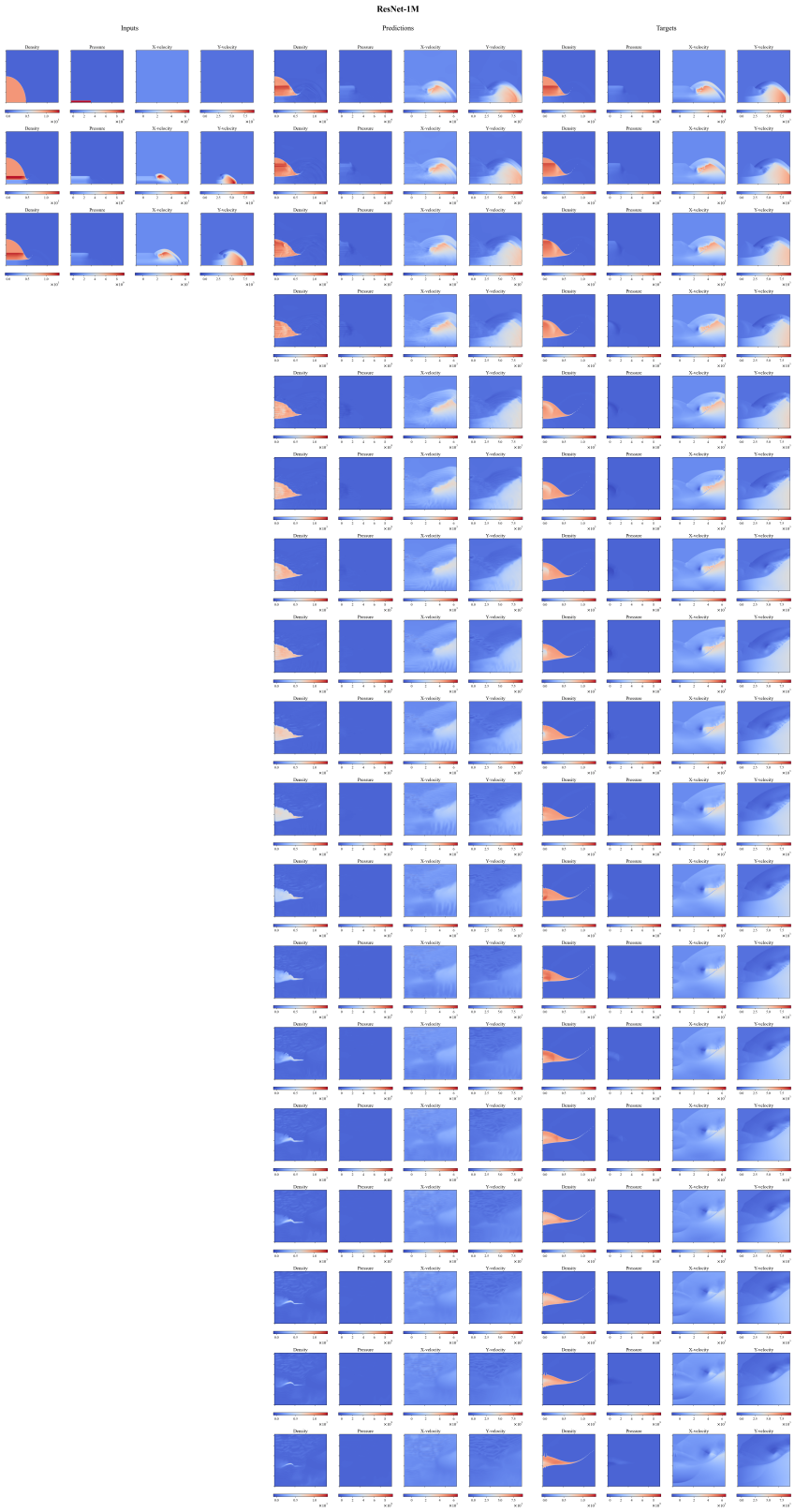

Figure 18: Rollout predictions for the LIDE-Experiment PDUV_F_(3,1) with ResNet-1M.

D.2 ROLLOUT PREDICTIONS FROM INITIAL CONDITIONS FOR THE SIDA DATASET

Here, we present rollout predictions for various models—each with 50M parameters, except for ResNet, which has only 1M parameter count. The trajectory shown in the Figures 19, 20, 21, 22, and 23 corresponds to the following simulation parameters: The shock Mach number 3.26, the flow Mach number 1.42, and the Weber number 13820.

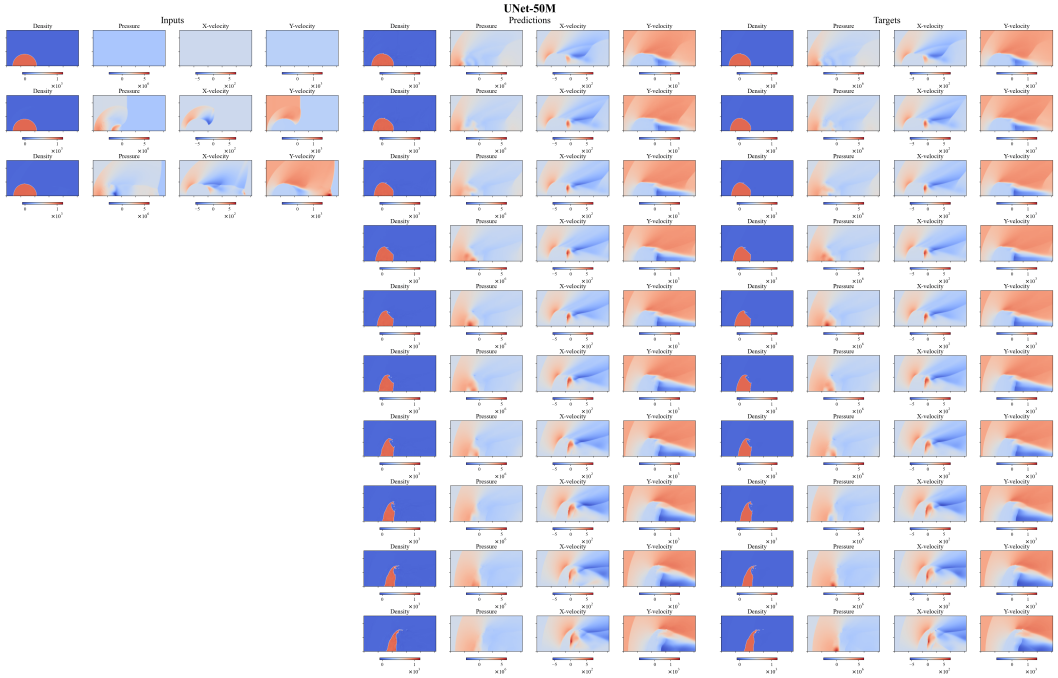

Figure 19: Rollout predictions for the SIDA-Experiment PDUV_F_(3,1) with UNet-50M.

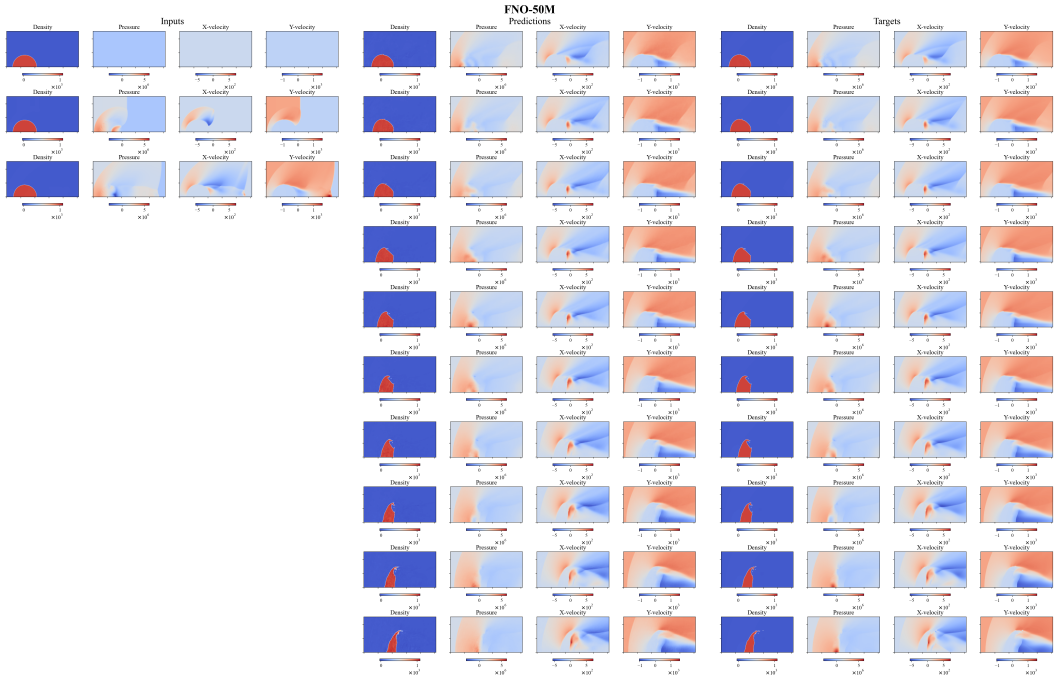

Figure 20: Rollout predictions for the SIDA-Experiment PDUV_F_(3,1) with FNO-50M.

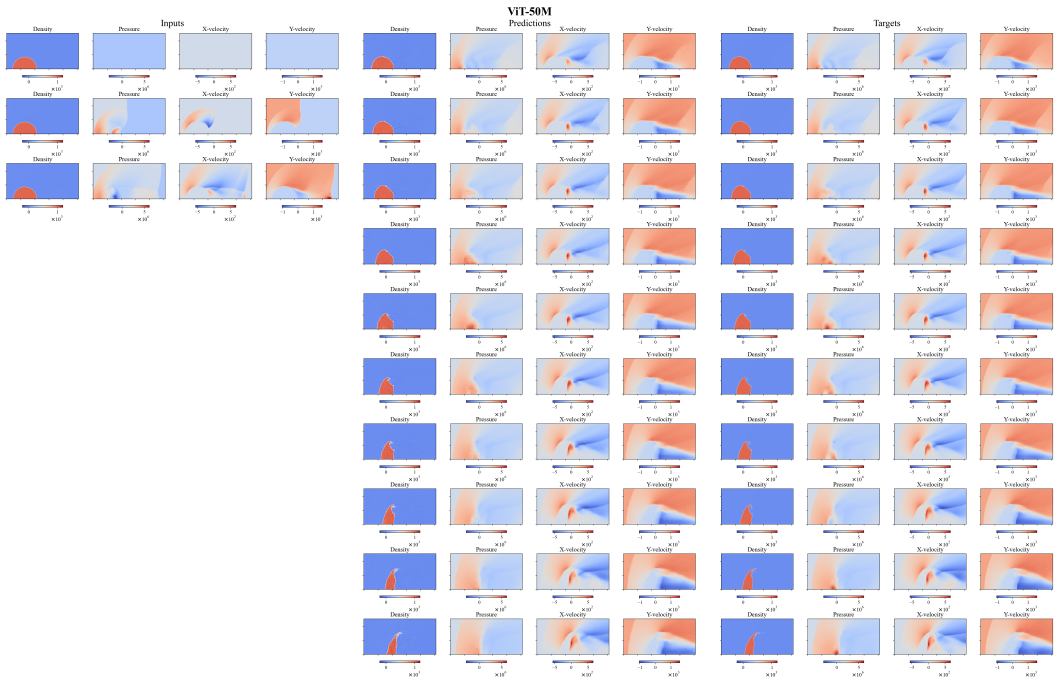

Figure 21: Rollout predictions for the SIDA-Experiment PDUV_F_(3,1) with ViT-50M.

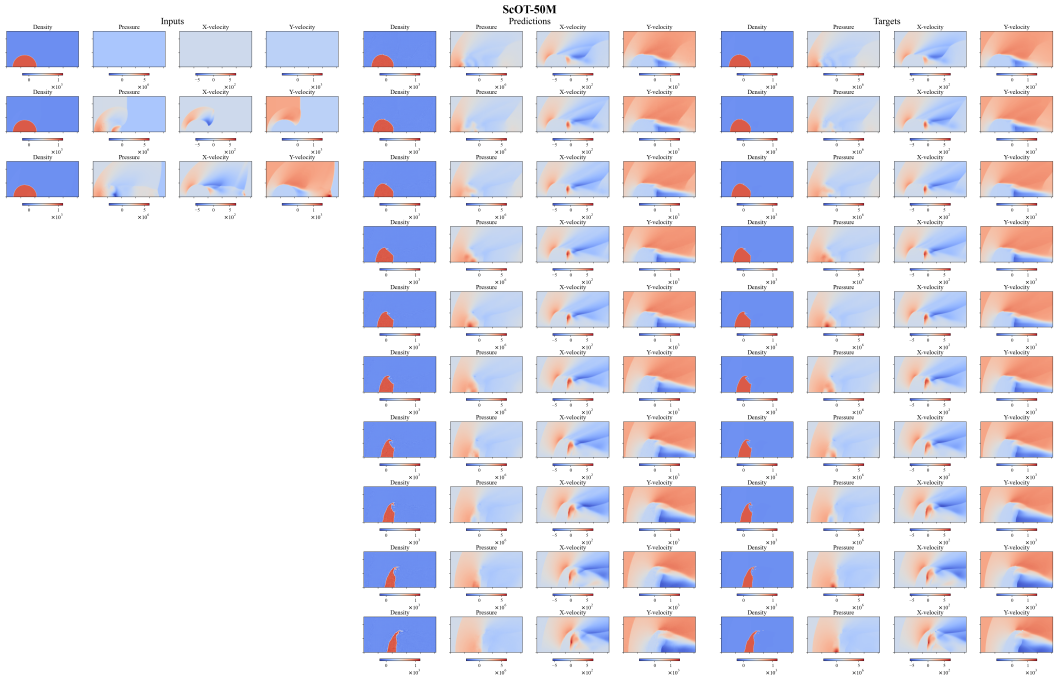

Figure 22: Rollout predictions for the SIDA-Experiment PDUV_F_(3,1) with ScOT-50M.

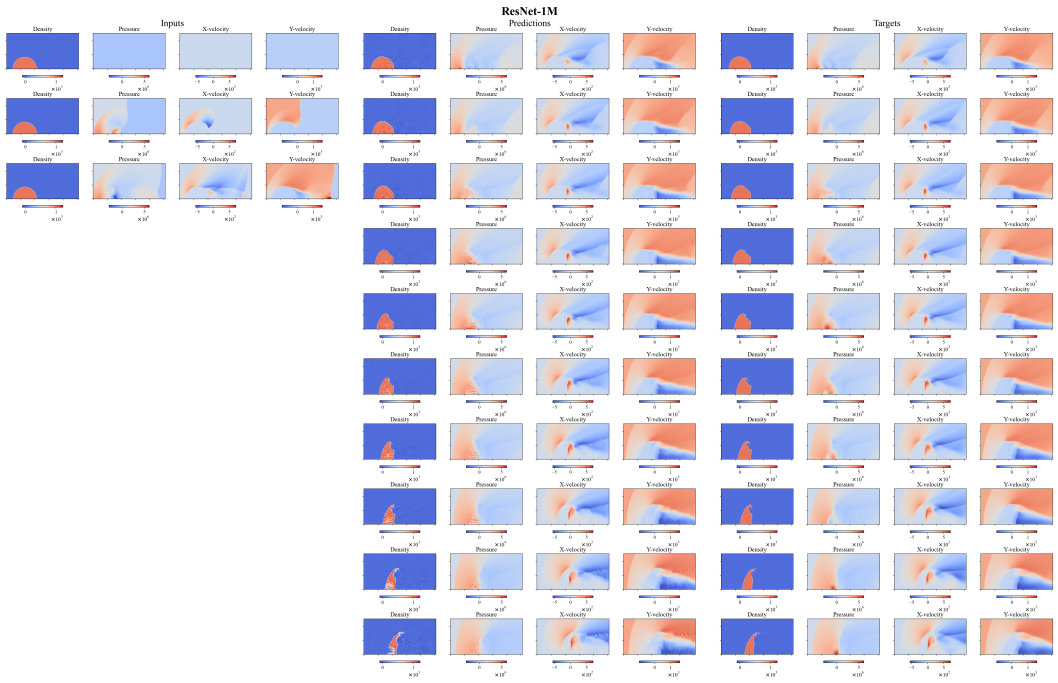

Figure 23: Rollout predictions for the SIDA-Experiment PDUV_F_(3,1) with ResNet-1M.

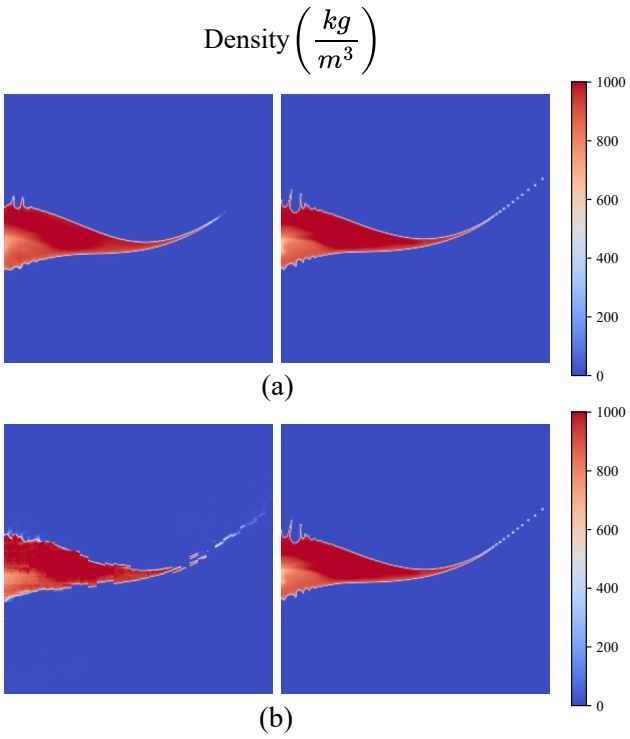

Figure 24: Comparison between UNet 50-M (a) and ViT 50-M (b) with target (for both at right) at the last rollout step for experiment PDUV_T_(3,1).

E TABLES

E.1 ERROR-TYPE 1 METRICS FOR THE LIDE DATASET

In this section, we showcase the MAE and RMSE of type 1 for all of the experiments for the LIDE dataset.

Table 15: Error-type 1 for experiment P_F_(1,1) for the LIDE dataset across all models.

MODEL	1M		50M	
	MAE	RMSE	MAE	RMSE
UNet	0.050158	0.320097	0.029947	0.263856
FNO	0.038743	0.236019	0.034457	0.228853
ViT	0.064188	0.330116	0.035981	0.265581
ScOT	0.054083	0.350523	0.030053	0.252165
ResNet	0.094531	0.440517		

Table 16: Error-type 1 for experiment P_F_(3,1) for the LIDE dataset across all models.

MODEL	1M		50M	
	MAE	RMSE	MAE	RMSE
UNet	0.018395	0.127369	0.010270	0.103838
FNO	0.026565	0.197309	0.018784	0.158807
ViT	0.051780	0.277934	0.022290	0.173438
ScOT	0.029390	0.209213	0.018250	0.159712
ResNet	0.058485	0.285421		

Table 17: Error-type 1 for experiment P_T_(1,1) for the LIDE dataset across all models.

MODEL	1M		50M	
	MAE	RMSE	MAE	RMSE
UNet	0.037299	0.207032	0.014883	0.088464
FNO	0.037581	0.219160	0.027419	0.175971
ViT	0.044926	0.251087	0.010941	0.072258
ScOT	0.074582	0.283233	0.018686	0.094912
ResNet	0.053545	0.272432		

Table 18: Error-type 1 for experiment P_T_(3,1) for the LIDE dataset across all models.

MODEL	1M		50M	
	MAE	RMSE	MAE	RMSE
UNet	0.034113	0.156079	0.009274	0.063560
FNO	0.023229	0.142621	0.016207	0.103424
ViT	0.032176	0.172894	0.009436	0.061425
ScOT	0.052968	0.165612	0.011976	0.072571
ResNet	0.040644	0.231908		

Table 19: Error-type 1 for experiment PDUV_F_(1,1) for the LIDE dataset across all models.

MODEL	1M		50M	
	MAE	RMSE	MAE	RMSE
UNet	0.109131	0.368182	0.063505	0.257702
FNO	0.170194	0.476427	0.095593	0.334874
ViT	0.276362	0.732310	0.069991	0.253188
ScOT	0.097253	0.320186	0.070541	0.270568
ResNet	0.423214	0.826303		

Table 20: Error-type 1 for experiment PDUV_F_(3,1) for the LIDE dataset across all models.

MODEL	1M		50M	
	MAE	RMSE	MAE	RMSE
UNet	0.058351	0.201772	0.029650	0.130409
FNO	0.114536	0.359660	0.065122	0.239348
ViT	0.184868	0.493885	0.047491	0.169020
ScOT	0.068976	0.210712	0.039000	0.147999
ResNet	0.314502	0.683159		

Table 21: Error-type 1 for experiment P[ES]_F_(1,1) for the LIDE dataset across all models.

MODEL	1M		50M	
	MAE	RMSE	MAE	RMSE
UNet	0.183736	0.668131	0.247058	0.972196
FNO	0.271891	0.866009	0.271549	0.986093
ViT	0.193752	0.710904	0.094765	0.480033
ScOT	0.167247	0.723314	0.110193	0.490153
ResNet	0.221558	0.687037		

Table 22: Error-type 1 for experiment P[ES]_F_(3,1) for the LIDE dataset across all models.

MODEL	1M		50M	
	MAE	RMSE	MAE	RMSE
UNet	0.112294	0.538067	0.094715	0.430769
FNO	0.098235	0.360389	0.122125	0.469131
ViT	0.077738	0.335093	0.061788	0.263886
ScOT	0.052584	0.238377	0.057491	0.223806
ResNet	0.078854	0.327747		

Table 23: Error-type 1 for experiment PDUV[ES]_F_(1,1) for the LIDE dataset across all models.

MODEL	1M		50M	
	MAE	RMSE	MAE	RMSE
UNet	0.428511	0.925083	0.520709	1.213596
FNO	0.445202	1.025386	0.415432	1.062686
ViT	0.650342	1.550977	0.178182	0.510104
ScOT	0.491391	1.162120	0.362049	1.046217
ResNet	0.345443	0.711238		

Table 24: Error-type 1 for experiment PDUV[ES]_F_(3,1) for the LIDE dataset across all models.

MODEL	1M		50M	
	MAE	RMSE	MAE	RMSE
UNet	0.326256	0.784144	0.210111	0.636456
FNO	0.352289	0.878331	0.271577	0.699626
ViT	0.261794	0.678912	0.189450	0.541396
ScOT	0.258918	0.740610	0.238986	0.754031
ResNet	0.374393	0.763712		

Table 25: Error-type 1 for experiment P_F_(3,2) for the LIDE dataset across all models.

MODEL	1M		50M	
	MAE	RMSE	MAE	RMSE
UNet	0.020932	0.150926	0.011344	0.107390
FNO	0.028200	0.185887	0.018559	0.143982
ViT	0.048801	0.258494	0.024832	0.190005
ScOT	0.026681	0.173740	0.021259	0.176111
ResNet	0.038422	0.232738		

Table 26: Error-type 1 for experiment PDUV_F_(3,2) for the LIDE dataset across all models.

MODEL	1M		50M	
	MAE	RMSE	MAE	RMSE
UNet	0.076998	0.245355	0.032054	0.131819
FNO	0.083239	0.250742	0.054791	0.189987
ViT	0.139224	0.378060	0.045292	0.149407
ScOT	0.063067	0.187791	0.040758	0.150562
ResNet	0.125493	0.324338		

Table 27: Error-type 1 for experiment PDUV_T_(1,1) for the LIDE dataset across all models.

MODEL	1M		50M	
	MAE	RMSE	MAE	RMSE
UNet	0.115898	0.324705	0.038253	0.148635
FNO	0.107092	0.362523	0.059970	0.210868
ViT	0.161391	0.521502	0.030712	0.114747
ScOT	0.087875	0.234057	0.031315	0.111389
ResNet	0.218120	0.561651		

Table 28: Error-type 1 for experiment PDUV_T_(3,1) for the LIDE dataset across all models.

MODEL	1M		50M	
	MAE	RMSE	MAE	RMSE
UNet	0.125620	0.351120	0.035482	0.132766
FNO	0.091418	0.317205	0.055380	0.185669
ViT	0.127357	0.388384	0.032883	0.121964
ScOT	0.080846	0.212856	0.030317	0.104845
ResNet	0.173521	0.413631		

E.2 Error-type 1 metrics for the SIDA Dataset

In this section, we showcase the MAE and RMSE of type 1 for all of the experiments for the SIDA dataset.

Table 29: Error-type 1 for experiment PDUV_F_(1,1) for the SIDA dataset across all models.

MODEL	1M		50M	
	MAE	RMSE	MAE	RMSE
UNet	0.389026	0.737015	0.376080	0.723840
FNO	0.388798	0.742821	0.374010	0.724714
ViT	0.448837	0.765849	0.366726	0.693767
ScOT	0.386842	0.730874	0.368502	0.700724
ResNet	0.408657	0.770825		

Table 30: Error-type 1 for experiment PDUV_F_(3,1) for the SIDA dataset across all models.

MODEL	1M		50M	
	MAE	RMSE	MAE	RMSE
UNet	0.034015	0.093065	0.011980	0.046474
FNO	0.052038	0.117692	0.022379	0.063122
ViT	0.169615	0.374392	0.041611	0.099006
ScOT	0.072025	0.161001	0.022156	0.052946
ResNet	0.051773	0.148635		

Table 31: Error-type 1 for experiment PDUV_T_(3,1) for the SIDA dataset across all models.

MODEL	1M		50M	
	MAE	RMSE	MAE	RMSE
UNet	0.085495	0.200467	0.023461	0.068501
FNO	0.064166	0.146237	0.035258	0.094010
ViT	0.195112	0.452964	0.040185	0.101045
ScOT	0.080007	0.184152	0.026292	0.062866
ResNet	0.099948	0.246592		

Table 32: Error-type 1 for experiment PDUV[VoS]_F_(1,1) for the SIDA dataset across all models.

MODEL	1M		50M	
	MAE	RMSE	MAE	RMSE
UNet	0.441329	0.816076	0.432100	0.814077
FNO	0.541915	0.963989	0.483547	0.881534
ViT	0.483162	0.840182	0.427463	0.781107
ScOT	0.528077	0.874120	0.222295	0.522612
ResNet	0.451940	0.815882		

Table 33: Error-type 1 for experiment PDUV[VoS]_F_(3,1) for the SIDA dataset across all models.

MODEL	1M		50M	
	MAE	RMSE	MAE	RMSE
UNet	0.168696	0.478806	0.143648	0.442353
FNO	0.250844	0.506633	0.248798	0.511349
ViT	0.272310	0.605585	0.169848	0.442154
ScOT	0.226147	0.553132	0.172184	0.451849
ResNet	0.153185	0.430725		

Table 34: Error-type 1 for experiment PDUV[VoS]_T_(3,1) for the SIDA dataset across all models.

MODEL	1M		50M	
	MAE	RMSE	MAE	RMSE
UNet	0.191591	0.513835	0.143940	0.423987
FNO	0.241415	0.506024	0.221999	0.449344
ViT	0.231530	0.563916	0.117042	0.387439
ScOT	0.179452	0.432399	0.155808	0.455810
ResNet	0.190392	0.499268		

Table 35: Error-type 1 for experiment PDUV_F_(3,2) for the SIDA dataset across all models.

MODEL	1M		50M	
	MAE	RMSE	MAE	RMSE
UNet	0.044712	0.103980	0.017345	0.050488
FNO	0.068911	0.145589	0.032252	0.080259
ViT	0.148939	0.302631	0.051340	0.117693
ScOT	0.080692	0.166986	0.027520	0.062177
ResNet	0.057433	0.160196		

Table 36: Error-type 1 for experiment PDUV[VoS]_F_(3,2) for the SIDA dataset across all models.

MODEL	1M		50M	
	MAE	RMSE	MAE	RMSE
UNet	0.126315	0.361863	0.112709	0.372599
FNO	0.231908	0.456978	0.207058	0.467535
ViT	0.243364	0.495933	0.129287	0.352591
ScOT	0.167449	0.413125	0.144851	0.402093
ResNet	0.139206	0.384204		

F LARGE LANGUAGE MODEL (LLM) USAGE

Large Language Models (LLMs) were utilized to polish the writing and find suitable words in some scenarios.