BEYOND SOURCE SIMULATION: PRETRAINING WITH PROXY GOVERNING EQUATIONS

Anonymous authors
Paper under double-blind review

ABSTRACT

Physical simulation is fundamental to aerospace, marine, and automotive engineering design, but obtaining high-fidelity solutions for diverse geometries and operating conditions remains computationally expensive. Although neural surrogates accelerate repeated prediction, their training still relies heavily on costly simulated fields for supervision. To overcome this limitation, we propose *proxy governing equation pretraining*, a new paradigm that learns transferable physical structure directly from unlabeled 3D geometries and governing laws, without precomputed source simulation fields. Specifically, we pair each geometry with randomly sampled physical conditions and pretrain a shared predictor using proxy equation residuals and boundary constraints. We adopt the classical Oseen system as a physically rich proxy that retains transport, pressure coupling, and viscous diffusion, but its second-order derivatives are difficult to estimate accurately on irregular 3D point sets. To address this, we introduce auxiliary momentum fluxes and reformulate the system into a provably equivalent first-order mixed form that preserves the proxy dynamics while requiring only first-order derivatives. Combined with boundary-aware operators and volume–surface pressure consistency, this formulation enables effective physical supervision on irregular 3D geometries. The pretrained model is then fine-tuned with high-fidelity target labels for diverse downstream tasks. Across five benchmarks spanning aerodynamic, hydrodynamic, and structural prediction, it transfers to substantially different target physics, including fluid-to-structure transfer. Our method matches full-data scratch accuracy with up to 60% fewer labeled simulations and further improves prediction accuracy on the data-rich DrivAerNet++ benchmark. These results show that proxy governing equations provide a principled way to lift geometry into transferable physical structure, extending physical-field pretraining beyond fixed collections of simulated solutions.

1 INTRODUCTION

Physical simulation aims to predict how a system responds to changes in different geometries and operating conditions. By estimating fields such as velocity, pressure, or stress, it supports performance analysis and design in aerospace, marine, and automotive engineering. High-fidelity numerical solvers provide detailed predictions, but each geometry-condition configuration requires substantial computation. Neural simulators, including neural operators, accelerate repeated predictions by learning mappings from geometry and operating conditions to physical responses (Kovachki et al., 2023; Li et al., 2023; Wu et al., 2024). Their effectiveness, however, still depends heavily on expensive precomputed physical solutions. For example, generating the computational fluid dynamics (CFD) data in DrivAerNet++ required approximately 3×10^6 CPU-hours (Elrefaie et al., 2024b). Extending such datasets to a broader range of geometries and operating conditions can be prohibitively expensive, as each new configuration typically requires rerunning the numerical solver. The cost of obtaining these fields therefore remains a major obstacle to scaling neural simulation.

Pretraining offers a promising way to reduce this dependence by learning reusable physical knowledge before adaptation to individual physical tasks. Existing physical pretraining approaches such as Poseidon and DPOT learn from collections of simulated PDE solutions, while AdaField uses aerodynamic data for pretraining (Herde et al., 2024; Hao et al., 2024; Zou et al., 2026). Although

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

these methods improve the use of target data, obtaining their source supervision still requires numerical simulation. An alternative is to pretrain on readily available 3D geometries without requiring physical-field solutions (Chang et al., 2015; Zhang et al., 2026), as shown in Figure 1.(a). Geometric reconstruction captures static shape information, but does not directly describe how physical fields respond to changing operating conditions. GeoPT introduces synthetic dynamics to supervise the evolution of geometric features (Wu et al., 2026), extending pretraining beyond static shape reconstruction, as shown in Figure 1.(b). Nevertheless, these geometric objectives do not explicitly enforce the coupled relations among velocity, pressure, and boundaries in a predicted physical field.

The resulting gap motivates a central question: *Can unlabeled geometries support learning physical structure without precomputed source solutions?* Our key observation is that a governing equation and its boundary conditions specify relations that a predicted field should satisfy, even when the corresponding solution is unavailable. A tractable proxy equation can therefore turn each geometry into a family of condition-dependent physical tasks. Learning across these tasks offers a way to acquire reusable geometry–condition–field relations, which we refer to as physical structure.

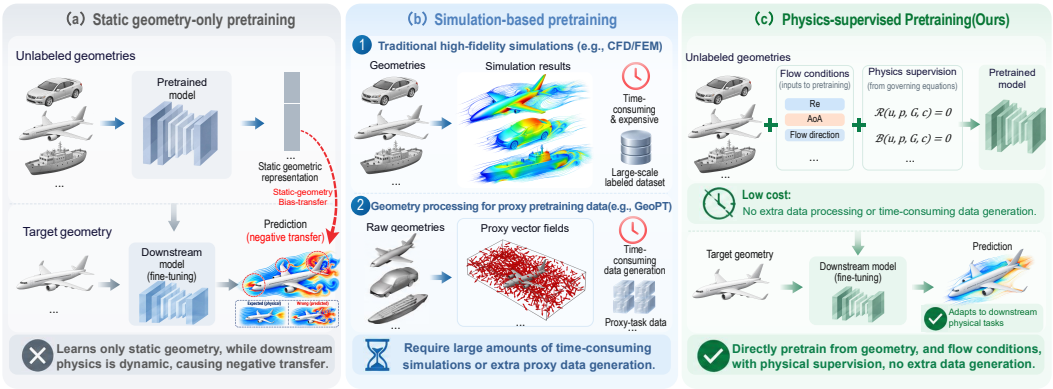

Figure 1: Comparison of three pretraining paradigms. (a) Geometry-only pretraining learns static geometric features with limited adaptation to dynamic downstream physics. (b) Simulation-based pretraining requires costly simulations or additional proxy-data generation. (c) Our method directly learns transferable physical structure from geometry, flow conditions, and governing constraints without source simulation fields.

Building on this insight, we propose *proxy governing equation pretraining*. As illustrated in Figure 1.(c), we pair source geometries with sampled physical conditions and train a shared model through equation residuals and boundary constraints without requiring source simulations. Resampling new conditions expands the pretraining task distribution without generating solution fields. After pretraining, we remove the proxy constraints and then fine-tune with a small amount of high-fidelity target labels to adapt to different downstream tasks, while the proxy equation need not strictly match the target governing physics. The proxy thus provides an inductive bias for adaptation rather than prescribing the downstream solution. Realizing this paradigm to complex 3D geometries is nontrivial because governing constraints must be evaluated on irregular volume and surface point sets. Specifically, we adopt the classical Oseen system (Galdi, 2011), a linearization of the incompressible Navier-Stokes equations that retains directional transport, pressure coupling, and viscous diffusion while omitting nonlinear disturbance self-advection. To avoid potentially inaccurate second-order derivative estimates on irregular 3D geometries, we simplify this Oseen system through an auxiliary momentum-flux tensor and reformulate it in a provably equivalent first-order mixed form. Jointly predicting velocity, pressure, and flux enables residual evaluation using only first-order derivatives while preserving the proxy dynamics. Boundary-aware moving-least-squares operators (Lancaster & Salkauskas, 1981) incorporate prescribed boundary information into local derivative estimates, while a fluid-side pressure trace connects volumetric supervision to surface prediction. Together, these components provide coupled physical supervision across the interior and boundary of each source geometry.

Our main contributions are as follows:

- We introduce *proxy governing equation pretraining*, which uses equation residuals and boundary constraints to learn from unlabeled 3D geometries without source simulation. Unlike pretraining from fixed simulated fields, the source task distribution can be continually enriched by resampling physical conditions without requiring corresponding simulated solution fields.
- We develop an Oseen-based implementation for irregular 3D domains, combining a first-order mixed formulation, boundary-aware differential operators, and volume-surface pressure consistency, enabling physical supervision across both interior and surface predictions. The pretrained model is adapted through target supervision, allowing the proxy and target governing physics to differ.
- Experiments on five downstream benchmarks demonstrate transfer across aerodynamic, hydrodynamic, and structural tasks. Our method achieves the lowest reported relative L2 error among the compared methods on all four smaller benchmarks, matches full-data training from scratch with up to 60% fewer high-fidelity labels, and further improves prediction accuracy on DrivAerNet++.

2 Related Work

Neural operators learn parametric PDE solution maps from simulation data. FNO (Li et al., 2021) uses Fourier-space parameterizations, while GINO (Li et al., 2023), GNOT (Hao et al., 2023), and Transolver (Wu et al., 2024) address irregular or complex geometries; Unisolver (Zhou et al., 2025) conditions prediction on PDE components to handle diverse equations. Beyond architecture, pretraining methods differ in where their supervision comes from. Poseidon (Herde et al., 2024) and DPOT (Hao et al., 2024) pretrain on simulated PDE fields, while AdaField (Zou et al., 2026) uses large aerodynamic datasets for pretraining; all require source physical responses. Geometry-based alternatives avoid these fields: geometry reconstruction (Zhang et al., 2026) provides shape-based supervision, whereas GeoPT (Wu et al., 2026) predicts trajectories of geometric features under synthetic dynamics. Equation-based learning offers another route. PINNs (Raissi et al., 2019) and PINO (Li et al., 2024) incorporate PDE residuals into training; FO-PINN (Gladstone et al., 2025) uses a first-order formulation, and DT-PINNs (Sharma & Shankar, 2022) evaluate derivatives through meshless discretization. PFEM (Wang et al., 2026) also pretrains without solution labels, using governing equations to produce initial guesses for finite-element solves. We instead use a proxy equation to supervise predicted physical fields across unlabeled geometry–condition pairs before adaptation to downstream physics that may differ from the pretraining system.

3 Method

3.1 Problem Setup and Framework Overview

For a downstream task t, we consider the prediction of a physical field $y^{(t)}(x; G, C_t)$ at spatial location x, conditioned on a geometry G and a physical condition C_t. The target task provides high-fidelity examples

$$\mathcal{D}_{\text{lab}}^{(t)} = \{(G, C_t, X_t, Y_t)\}, \tag{1}$$

where X_t denotes query locations and Y_t the corresponding field values. In contrast, pretraining assumes access only to a collection of source geometries $\mathcal{D}_{\text{geo}} = \{G_i\}_{i=1}^{N}$ without corresponding source solution fields. Our goal is to learn transferable geometry–condition–field relations, which we refer to as *physical structure*, and adapt them to different downstream physical tasks.

We lift each source geometry into a family of equation-constrained physical tasks. For a geometry G, spatial samples $\mathcal{X}_G = \{x_j\}_{j=1}^{N_G}$ represent its physical domain and boundaries. Sampling a proxy condition $c \sim P_c$ defines

$$\mathcal{T}(G, c) = \big(\Omega_G, \mathcal{X}_G, \mathcal{R}_c, \mathcal{B}_{G,c}\big), \tag{2}$$

where Ω_G is the geometry-dependent domain, $\mathcal{R}_c$ denotes the governing-equation residuals under condition c, and $\mathcal{B}_{G,c}$ specifies the associated boundary constraints. A shared predictor is trained across these geometry–condition tasks, using the proxy equations as supervision rather than independently solving each source PDE instance.

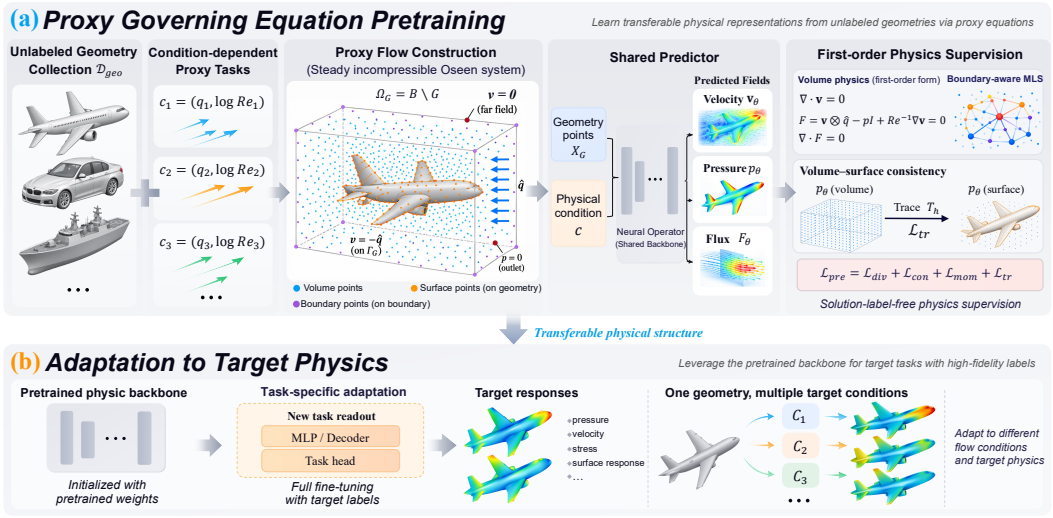

Figure 2: Overview of the two-stage framework. Source geometries are paired with sampled physical conditions to form proxy tasks supervised by equation residuals and boundary constraints. The pretrained predictor is then adapted using high-fidelity labels from a target task.

Our framework consists of two stages, as illustrated in Figure 2. During pretraining, the shared model learns from geometry–condition tasks through

$$\mathcal{L}_{\text{pre}} = \mathbb{E}_{G\sim\mathcal{D}_{\text{geo}},\, c\sim P_c}\left[\mathcal{L}_{\text{proxy}}(G, c)\right], \tag{3}$$

where $\mathcal{L}_{\text{proxy}}$ enforces the proxy governing relations, geometry-dependent boundary constraints, and volume–surface consistency defined in the following sections. During adaptation, the proxy constraints are removed, and the pretrained model is specialized to a target task using high-fidelity supervision:

$$\mathcal{L}^{(t)}_{\text{target}} = \mathbb{E}_{\mathcal{D}^{(t)}_{\text{lab}}}\left[\ell_t(\widehat{Y}_t, Y_t)\right]. \tag{4}$$

The proxy stage therefore determines the transferable physical structure learned from source geometries, while target supervision determines the task-specific high-fidelity response.

3.2 Condition-Dependent Proxy Flow Tasks

We instantiate the proxy tasks with the steady incompressible Oseen system (Galdi, 2011). This simplified system retains directional transport, pressure coupling, and viscous diffusion, but omits the disturbance self-advection term of the full Navier–Stokes equations. It is used to provide structured pretraining supervision.

For a geometry G embedded in a bounding box B, the fluid domain is $\Omega_G = B \setminus G$. Let q be a nonzero background-flow vector, $U = \|q\|_2$, and $\hat{q} = q/U$. The network receives the condition feature $c = [q_x, q_y, q_z, \log Re]$, whereas the dimensionless proxy constraints are parameterized by $(\hat{q}, Re)$. Writing the dimensional total velocity as $u^{\dim}/U = \hat{q} + v$, the disturbance velocity v and pressure p satisfy

$$\nabla \cdot v = 0, \qquad (\hat{q} \cdot \nabla)v + \nabla p - Re^{-1}\Delta v = 0 \quad \text{in } \Omega_G. \tag{5}$$

Changing $\hat{q}$ or Re produces distinct proxy tasks on the same geometry, turning a static shape into a family of condition-dependent physical tasks.

Let Γ_G denote the body surface, Γ_{far} the far-field boundary, and Γ_{out} the designated outlet. We impose

$$v = -\hat{q} \quad \text{on } \Gamma_G, \qquad v = 0 \quad \text{on } \Gamma_{\text{far}}, \qquad p = 0 \quad \text{on } \Gamma_{\text{out}}. \tag{6}$$

The body and far-field velocities are prescribed at sampled points, corresponding to no slip and undisturbed background flow, respectively. A hard pressure lifting enforces $p_\theta = 0$ on the designated outlet, while outlet velocity remains predicted. The prescribed velocities also enter local derivative estimates near the boundaries; no separate boundary-loss term is required.

3.3 PHYSICS-SUPERVISED PRETRAINING ON IRREGULAR 3D DOMAINS

Direct evaluation of the second-order diffusion term Δv is challenging on irregular 3D point sets (Gladstone et al., 2025; Sharma & Shankar, 2022). We therefore introduce an auxiliary momentum flux $F = v \otimes \hat{q} + pI - Re^{-1}\nabla v$ and write the proxy system in first-order form:

$$\nabla \cdot v = 0, \qquad F - v \otimes \hat{q} - pI + Re^{-1}\nabla v = 0, \qquad \nabla \cdot F = 0. \tag{7}$$

With row-wise tensor divergence, $\nabla \cdot F$ recovers the momentum equation in Eq. 5. The three relations jointly constrain the predicted fields $(v_\theta, p_\theta, F_\theta)$ while requiring only first-order spatial derivatives.

We evaluate the required first-order derivatives using local moving least squares (MLS) (Lancaster & Salkauskas, 1981). For a fluid-volume point x_i and coordinate direction k,

$$(D_{h,k} f)_i = \sum_{j \in \mathcal{N}_i} w_{ij}^{(k)} (f_j - f_i), \qquad k \in \{1, 2, 3\}, \tag{8}$$

where the weights depend on the local point configuration. Near geometric boundaries, these neighborhoods become one-sided and do not directly encode prescribed boundary values. We therefore augment them with boundary anchors:

$$(D_{h,k}^{\mathrm{bc}} f)_i = \sum_{j \in \mathcal{N}_i \cup \mathcal{B}_i} w_{ij}^{(k),\mathrm{bc}} (f_j - f_i), \qquad k \in \{1, 2, 3\}. \tag{9}$$

where $\mathcal{B}_i$ contains the boundary anchors associated with x_i and the weights are fitted on the augmented configuration. We use D_h^{bc} for velocity derivatives with prescribed boundary values and the standard operator for flux divergence. The resulting volumetric residuals are

$$\begin{aligned} r_{\mathrm{div}} &= D_h^{\mathrm{bc}} \cdot v_\theta, \\ r_{\mathrm{con}} &= F_\theta - v_\theta \otimes \hat{q} - p_\theta I + Re^{-1} D_h^{\mathrm{bc}} v_\theta, \\ r_{\mathrm{mom}} &= D_h \cdot F_\theta. \end{aligned} \tag{10}$$

Volumetric residuals alone do not directly constrain physical quantities predicted on the body surface. We therefore construct a fluid-side trace of the volumetric pressure at each surface sample s:

$$(T_h p_\theta)_s = \sum_{j \in \mathcal{N}_s^{\mathrm{tr}}} \tau_{sj} p_\theta(x_j), \tag{11}$$

where τ_{sj} are geometry-dependent trace weights. The surface-pressure prediction is coupled to this trace through

$$\mathcal{L}_{\mathrm{tr}}(G, c) = \frac{\sum_{s \in \mathcal{S}_G} A_s \left[p_\theta(s) - (T_h p_\theta)_s \right]^2}{\sum_{s \in \mathcal{S}_G} A_s}, \tag{12}$$

where $\mathcal{S}_G$ denotes the body-surface samples and A_s their associated area weights. This constraint allows volumetric equation supervision to propagate to surface prediction without source surface-field labels.

Combining the volumetric residuals with the pressure-trace constraint gives the proxy objective introduced in Sec. 3.1:

$$\begin{aligned} \mathcal{L}_{\mathrm{proxy}}(G, c) = \lambda_{\mathrm{div}} \mathcal{M}_G(r_{\mathrm{div}}) + \lambda_{\mathrm{con}} \mathcal{M}_G(r_{\mathrm{con}}) \\ + \lambda_{\mathrm{mom}} \mathcal{M}_G(r_{\mathrm{mom}}) + \lambda_{\mathrm{tr}} \mathcal{L}_{\mathrm{tr}}(G, c), \end{aligned} \tag{13}$$

where $\mathcal{M}_G$ denotes the mean squared residual over valid fluid-volume points. Together, these terms impose coupled physical supervision across irregular volume and surface samples without requiring source solution fields. Further discrete-operator and pressure-trace details appear in Appendix A.

3.4 IMPLEMENTATION DETAILS

We instantiate the shared predictor with Transolver (Wu et al., 2024). The pretraining model predicts 13 channels: disturbance velocity, pressure, and momentum flux. During adaptation, we transfer the pretrained model parameters, reinitialize the task-specific output layers, and fine-tune the model using high-fidelity target labels.

Pretraining uses 10,253 unlabeled ShapeNetV2 geometries (Chang et al., 2015), covering airplanes, cars, buses, and watercraft. Each geometry provides 40,000 fluid-volume points, 10,000 body-surface points, and 10,000 outer-boundary points. Instead of constructing a separate proxy dataset before training, our method directly generates equation-constrained proxy tasks from geometry by sampling physical conditions and evaluating governing-equation residuals during training. In contrast, synthetic-supervision approaches such as GeoPT require an additional offline stage to generate geometric trajectory supervision before pretraining, which takes around 3 days on 80 CPU cores (Wu et al., 2026). Our source-solution-free formulation removes this additional data-generation stage and allows physical supervision to be directly obtained from unlabeled geometries.

We optimize the model with AdamW (Loshchilov & Hutter, 2019) for 300 epochs during pretraining and 200 epochs during downstream adaptation. The learning rate follows a linear warmup schedule followed by cosine decay. Full optimization configurations and task-specific training settings are provided in Appendix B.

4 Experiments

We extensively evaluate our proxy governing equation pretraining across five downstream physics simulation tasks with diverse geometries, physical conditions, and target fields, demonstrating its transferability beyond the proxy system.

4.1 Downstream Tasks and Baselines

We evaluate our method on five downstream physics benchmarks covering aerodynamic, hydrodynamic, and structural field prediction. Aircraft (Luo et al., 2025) and NASA-CRM (Bekemeyer et al., 2025) evaluate aerodynamic field prediction on complex aircraft geometries, where the model predicts surface aerodynamic quantities under different operating conditions. DTCHull (Wu et al., 2026) evaluates hydrodynamic prediction on ship geometries, involving flow fields around different hull configurations. Car-Crash (Wu et al., 2026) evaluates structural response prediction under impact scenarios, with deformation-related physical quantities as targets. In addition, we evaluate on DrivAerNet++ (Elrefaie et al., 2024b), a large-scale automotive aerodynamics benchmark for predicting surface pressure and wall shear stress on vehicle geometries. Together, these benchmarks cover diverse geometries, physical conditions, and target fields, enabling evaluation of transfer beyond the proxy system. Dataset sizes, splits, prediction targets, and evaluation metrics are detailed in Appendix B.

For the four-task comparison, we include representative neural operator and physical pretraining methods, including GNOT (Hao et al., 2023), UPT (Alkin et al., 2024), Transolver++ (Luo et al., 2025), LinearNO (Hu et al., 2026), HNO (Pei et al., 2026), and GeoPT (Wu et al., 2026). For DrivAerNet++, we compare with specialized aerodynamic prediction methods, including RegDGCNN (Elrefaie et al., 2024a), Transolver (Wu et al., 2024), FIGConvNet (Choy et al., 2025), TripNet (Chen et al., 2026), AdaField (Zou et al., 2026), GA-Field (Zheng et al., 2026), and GeoPT (Wu et al., 2026).

4.2 Transfer Across Target Physics

We test whether pretraining with a simplified Oseen proxy benefits downstream prediction when the target geometries, fields, and physical processes differ from those of the source task. The four full-label benchmarks cover aircraft aerodynamics (Aircraft and NASA-CRM), hydrodynamics (DTCHull), and impact-induced structural response (Car-Crash). Car-Crash provides the clearest proxy–target mismatch: the model is pretrained with a flow equation but adapted to predict peak von Mises stress.

Table 1 shows that our method obtains the lowest overall relative L2 error and the highest reported field-specific R^2 on all four benchmarks. GeoPT is the runner-up in overall relative L2 on the three fluid benchmarks, and our method improves on it in each case. On Car-Crash, our relative L2 error is 0.1762, compared with 0.1846 for the strongest competing result and 0.2183 for GeoPT. The structural target differs substantially from the proxy flow task, so this result supports the usefulness of

Table 1: Full-label comparison on four downstream benchmarks. Relative L2 aggregates all predicted channels for Aircraft and DTCHull and evaluates the single target field for NASA-CRM and Car-Crash. R^2 is reported for the field indicated in each column header. Lower relative L2 and higher R^2 are better; bold and underlined values mark the best and second-best results.

METHOD	AIRCRAFT		DTCHULL		NASA-CRM		CAR-CRASH	
	Rel. L2 ↓	$R^2_{C_p}$ ↑	Rel. L2 ↓	R^2_p ↑	Rel. L2 ↓	$R^2_{C_p}$ ↑	Rel. L2 ↓	$R^2_{\sigma_{\mathrm{VM}}}$ ↑
GNOT	0.1574	0.7163	0.2248	0.8697	0.1811	0.9654	0.2835	0.8193
UPT	0.2317	0.3682	0.5011	0.4169	0.6028	0.6122	0.3915	0.6556
Transolver++	0.0787	0.9372	0.2458	0.8633	0.1069	0.9872	<u>0.1846</u>	<u>0.9226</u>
LinearNO	0.0517	0.9552	0.1666	0.9374	0.1098	0.9865	0.2109	0.8993
HNO	0.0582	0.9578	0.1650	0.9394	0.1018	0.9882	0.1879	0.9198
GeoPT	<u>0.0412</u>	<u>0.9826</u>	<u>0.1444</u>	<u>0.9512</u>	<u>0.0919</u>	<u>0.9894</u>	0.2183	0.8919
Ours	**0.0405**	**0.9879**	**0.1362**	**0.9565**	**0.0866**	**0.9910**	**0.1762**	**0.9295**

proxy-equation pretraining beyond downstream problems that closely resemble its source dynamics. Field-wise results and source-training diagnostics are provided in Appendix C.

We also inspect the pretrained predictor before adaptation to a target task. Figure 3 holds one NASA-CRM geometry fixed and displays its proxy responses under three sampled conditions. This comparison examines whether the same geometry produces distinct predicted fields as the physical conditions change.

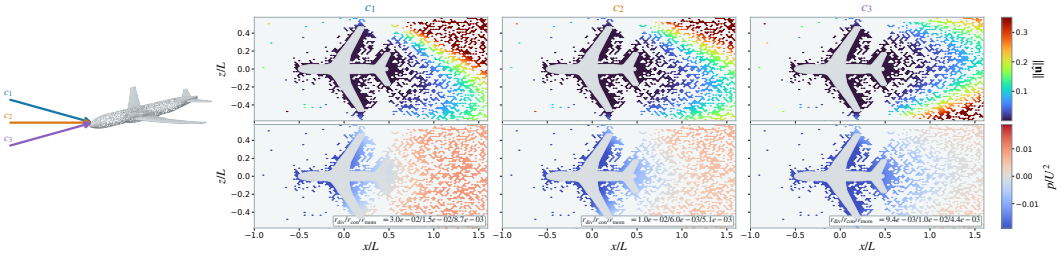

Figure 3: Condition-dependent proxy responses on the same NASA-CRM geometry before downstream fine-tuning. The proxy condition is $c = (U, \mathrm{yaw}, \mathrm{pitch}, Re)$, with $c_1 = (0.80, -20.00°, -2.00°, 10.00)$, $c_2 = (0.86, -10.26°, -1.47°, 37.63)$, and $c_3 = (1.11, 6.19°, 1.98°, 13.49)$. The upper and lower rows show two-dimensional center-plane (x–z) fields of velocity magnitude and normalized pressure p/U^2, respectively.

Across the three conditions, the lateral distribution of the high-velocity wake changes and the pressure field is redistributed, while near-body response patterns remain anchored to the same aircraft geometry. The annotated divergence, constitutive, and momentum residuals are on the order of 10^{-2}–10^{-3} at the evaluated points, consistent with the imposed discrete proxy constraints. The visualization thus shows condition-sensitive proxy responses before fine-tuning and complements the downstream transfer results. Additional visualizations are shown in Appendix D.

4.3 LABEL EFFICIENCY UNDER LIMITED SUPERVISION

We further evaluate whether proxy governing equation pretraining reduces the amount of high-fidelity supervision required for downstream adaptation. For each target task, we fine-tune the pretrained model using different fractions of the available training cases, including 20%, 40%, 60%, 80%, and 100% label budgets. The results are shown in Figure. 4.

Figure 4 compares the adaptation performance under different amounts of target supervision. Using the full-data scratch model as the reference, proxy governing equation pretraining achieves com-

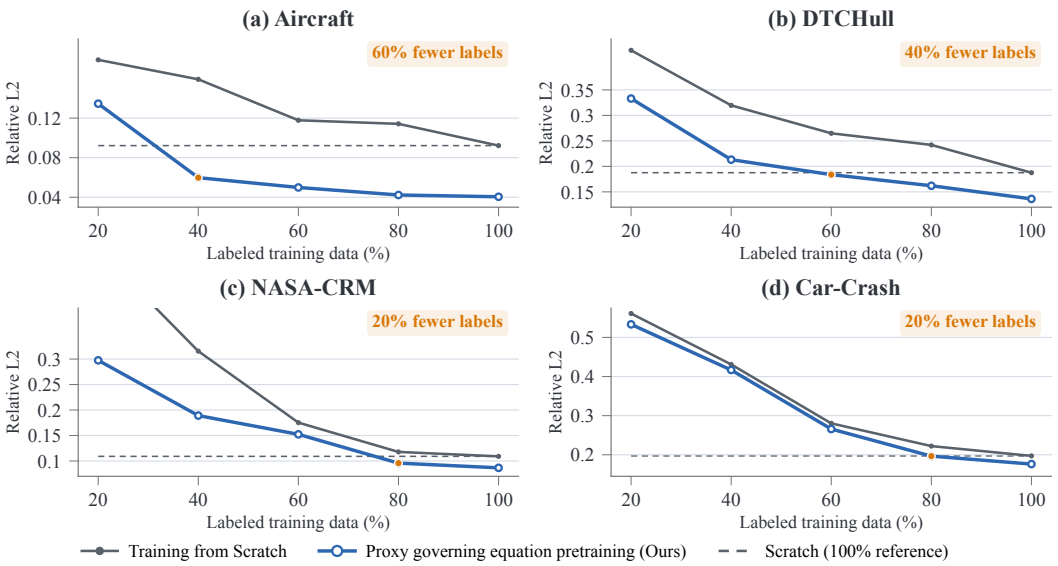

Figure 4: Label efficiency under reduced high-fidelity supervision. Curves show relative L2 error when adapting from scratch or from proxy governing equation pretraining with different fractions of target labels. The dashed line denotes the full-data scratch baseline, and markers indicate the smallest label budget required for the pretrained model to match or exceed that baseline.

parable or better performance with substantially fewer labeled cases. Specifically, the pretrained model reaches the full-data scratch relative L2 error using only 40% of the labels on Aircraft, 60% on DTCHull, and 80% on both NASA-CRM and Car-Crash. This corresponds to reductions of 60%, 40%, 20%, and 20% in required high-fidelity labels, respectively.

Table 2: Ablation of coupled proxy physics under full target supervision. Lower relative L2 errors and higher R^2 values indicate better performance.

Dataset	Field	Scratch		Partial proxy constraints		**Full proxy physics (Ours)**	
		Rel. L2↓	R^2 ↑	Rel. L2↓	R^2 ↑	Rel. L2↓	R^2 ↑
Aircraft	C_p	0.2896	0.8906	0.2152	0.9378	**0.0961**	**0.9879**
	$\mathbf{u}$	0.0632	0.9936	0.0457	0.9964	**0.0263**	**0.9987**
DTCHull	p	0.1876	0.9172	0.1572	0.9446	**0.1362**	**0.9565**
	$\mathbf{u}$	0.1727	0.9553	0.1628	0.9601	**0.1595**	**0.9618**
NASA-CRM	C_p	0.1091	0.9863	0.1153	0.9848	**0.0866**	**0.9910**
Car-Crash	$\sigma_{\mathrm{VM}}^{\max}$	0.1971	0.9119	0.1830	0.9239	**0.1762**	**0.9295**

These results demonstrate that proxy governing equation pretraining improves the utilization of high-fidelity simulation data. By learning transferable geometry–condition–field structure before adaptation, the model requires fewer labeled simulations to achieve the performance of a fully supervised model trained from scratch.

4.4 Ablation of Coupled Proxy Physics

We assess the contribution of the coupled flux–momentum constraints in proxy-equation pretraining by comparing training from scratch, a partial proxy objective, and the full objective. The partial objective jointly sets $\lambda_{\mathrm{con}} = \lambda_{\mathrm{mom}} = 0$, while retaining the incompressibility and pressure-trace terms and the same boundary treatment: $\mathcal{L}_{\mathrm{partial}} = \lambda_{\mathrm{div}}\mathcal{M}_G(r_{\mathrm{div}}) + \lambda_{\mathrm{tr}}\mathcal{L}_{\mathrm{tr}}$. We remove the two terms together because the constitutive residual couples the auxiliary flux to velocity and pressure,

whereas the flux divergence residual enforces momentum balance. This ablation therefore tests the coupled flux–momentum block as a whole.

Table 2 shows that the full objective achieves the lowest relative L2 error and highest R^2 on all six evaluated fields. The partial objective improves over scratch training on five fields, with NASA-CRM C_p as the exception. Its performance remains below the full objective on every field, consistent with the coupled flux–momentum constraints providing a stronger pretraining signal than incompressibility and pressure trace alone.

4.5 Beyond Data Scarcity: Data-Rich Transfer on DrivAerNet++

We further investigate whether proxy governing equation pretraining improves the final prediction accuracy when sufficient target supervision is available. Unlike the previous label-efficiency evaluation, this experiment uses the full training set and focuses on the achievable performance after adaptation. Table 3 compares our method with representative neural operators and pretraining approaches.

Table 3: DrivAerNet++ pressure and wall shear stress prediction. Relative errors are reported in percent; MSE and MAE use units of 10^{-2} and 10^{-1}, respectively. Lower is better.

	Pressure					**Wall Shear Stress**				
Method	Rel L2↓ (%)	Rel L1↓ (%)	MSE↓ (10^{-2})	MAE↓ (10^{-1})	MaxAE↓	Rel L2↓ (%)	Rel L1↓ (%)	MSE↓ (10^{-2})	MAE↓ (10^{-1})	MaxAE↓
RegDGCNN	27.72	26.21	8.29	1.61	10.81	36.42	36.57	13.82	3.64	11.54
Transolver	23.87	22.57	7.15	1.41	7.12	22.49	17.65	8.95	2.06	8.45
FIGConvNet	20.86	21.12	4.99	1.22	6.55	22.32	17.65	9.86	2.22	7.84
TripNet	20.05	20.93	5.14	1.25	6.35	22.07	17.18	9.52	2.15	8.14
AdaField	19.81	17.28	4.58	1.05	6.22	—	—	—	—	—
GA-Field	19.28	17.00	4.13	1.04	6.13	20.54	16.12	8.18	1.62	5.64
GeoPT	<u>18.28</u>	<u>16.18</u>	<u>3.98</u>	<u>0.99</u>	<u>6.01</u>	<u>18.86</u>	<u>14.99</u>	<u>6.95</u>	<u>1.51</u>	<u>5.15</u>
Ours	**17.91**	**15.72**	**3.77**	**0.97**	**5.88**	**18.51**	**14.67**	**6.70**	**1.48**	**5.02**

Our method achieves the best performance across all reported pressure and wall shear stress metrics. Compared with Transolver trained directly on the target data, our method reduces the relative L2 error from 23.87% to 17.91% for pressure prediction and from 22.49% to 18.51% for wall shear stress prediction. It also consistently outperforms the geometry-based pretraining method GeoPT.

These results demonstrate that the benefit of proxy governing equation pretraining is not limited to improving data efficiency. Even in a data-rich regime with abundant target supervision, the learned geometry–condition–field structure provides complementary information that further improves the attainable prediction accuracy.

5 Conclusion

High-fidelity physical fields are expensive to obtain, whereas unlabeled 3D geometries are abundant. We introduced *proxy governing equation pretraining*, which turns each geometry and its sampled conditions into a family of physics-supervised tasks without requiring source simulation fields. A first-order mixed Oseen formulation, boundary-aware differentiation, and a fluid-side pressure trace make this supervision applicable to irregular 3D domains. The proxy equation provides physical structure for a shared predictor; target labels then specialize it to the downstream task, even when the target physics differs from the proxy. Across four aerodynamic, hydrodynamic, and structural benchmarks, our method achieves the lowest relative L2 error and matches full-data scratch accuracy with 20–60% fewer high-fidelity labels. It also improves pressure and wall-shear-stress prediction on DrivAerNet++ under full target supervision. These results suggest a broader role for governing equations: beyond defining individual simulation problems, they can provide reusable supervision that turns inexpensive geometry collections into predictors adaptable to distinct physical tasks.

AI USE STATEMENT

In this work, we used generative AI tools for language polishing and for refining figure layouts and captions. We did not use generative AI to generate experimental data or to determine the scientific conclusions of this work. All AI-assisted content was carefully reviewed and revised by the authors. In particular, language edits were checked to ensure that they preserved the intended technical meaning, and AI-assisted figure layouts and captions were verified against the corresponding methods and experimental results. We take full responsibility for the final content of this work, including all text, claims, figures, and other artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

We provide detailed descriptions of the proposed proxy governing equation pretraining framework, including the proxy task construction, first-order formulation, differential operators, boundary treatment, and training objectives in Section 3. The datasets, task definitions, evaluation metrics, data splits, and preprocessing procedures used in the experiments are described in Section 4 and Appendix B. Additional implementation and optimization details, including sampling strategies, training schedules, and task-specific configurations, are provided there. We also report ablation studies and evaluations under different supervision regimes to facilitate verification of the main empirical findings.

REFERENCES

Benedikt Alkin, Andreas Fürst, Simon Schmid, Lukas Gruber, Markus Holzleitner, and Johannes Brandstetter. Universal physics transformers: A framework for efficiently scaling neural operators. In *Advances in Neural Information Processing Systems*, volume 37, 2024.

Philipp Bekemeyer, Nathan Hariharan, Andrew M. Wissink, and Jason Cornelius. Introduction of applied aerodynamics surrogate modeling benchmark cases. In *AIAA SCITECH 2025 Forum*, 2025.

Angel X. Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, Jianxiong Xiao, Li Yi, and Fisher Yu. ShapeNet: An information-rich 3D model repository. *arXiv preprint arXiv:1512.03012*, 2015.

Qian Chen, Mohamed Elrefaie, Angela Dai, and Faez Ahmed. TripNet: Learning large-scale high-fidelity three-dimensional car aerodynamics with triplane networks. *Physics of Fluids*, 38(6): 062106, 2026.

Chris Choy, Alexey Kamenev, Jean Kossaifi, Max Rietmann, Jan Kautz, and Kamyar Azizzadenesheli. Factorized implicit global convolution for automotive computational fluid dynamics prediction. *arXiv preprint arXiv:2502.04317*, 2025.

Mohamed Elrefaie, Angela Dai, and Faez Ahmed. DrivAerNet: A parametric car dataset for data-driven aerodynamic design and graph-based drag prediction. In *Proceedings of the ASME 2024 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, volume 3A, pp. V03AT03A019. American Society of Mechanical Engineers, 2024a.

Mohamed Elrefaie, Florin Morar, Angela Dai, and Faez Ahmed. DrivAerNet++: A large-scale multimodal car dataset with computational fluid dynamics simulations and deep learning benchmarks. In *Advances in Neural Information Processing Systems*, volume 37, pp. 499–536, 2024b.

Giovanni P. Galdi. *An Introduction to the Mathematical Theory of the Navier–Stokes Equations: Steady-State Problems*. Springer Monographs in Mathematics. Springer, second edition, 2011.

Rini Jasmine Gladstone, Mohammad Amin Nabian, N. Sukumar, Ankit Srivastava, and Hadi Meidani. FO-PINN: A first-order formulation for physics-informed neural networks. *Engineering Analysis with Boundary Elements*, 174:106161, 2025.

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 *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 12556–12569. PMLR, 2023.

Zhongkai Hao, Chang Su, Songming Liu, Julius Berner, Chengyang Ying, Hang Su, Anima Anandkumar, Jian Song, and Jun Zhu. DPOT: Auto-regressive denoising operator transformer for large-scale PDE pre-training. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 17616–17635. PMLR, 2024.

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

Wenjie Hu, Sidun Liu, Peng Qiao, Zhenglun Sun, and Yong Dou. Transolver is a linear transformer: Revisiting physics-attention through the lens of linear attention. *Proceedings of the AAAI Conference on Artificial Intelligence*, 40(1):408–416, 2026.

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.

P. Lancaster and K. Salkauskas. Surfaces generated by moving least squares methods. *Mathematics of Computation*, 37(155):141–158, 1981.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*, 2021.

Zongyi Li, Nikola Kovachki, Chris Choy, Boyi Li, Jean Kossaifi, Shourya Otta, Mohammad Amin Nabian, Maximilian Stadler, Christian Hundt, Kamyar Azizzadenesheli, and Anima Anandkumar. Geometry-informed neural operator for large-scale 3D PDEs. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

Zongyi Li, Hongkai Zheng, Nikola Kovachki, David Jin, Haoxuan Chen, Burigede Liu, Kamyar Azizzadenesheli, and Anima Anandkumar. Physics-informed neural operator for learning partial differential equations. *ACM / IMS Journal of Data Science*, 1(3):1–27, 2024.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019.

Huakun Luo, Haixu Wu, Hang Zhou, Lanxiang Xing, Yichen Di, Jianmin Wang, and Mingsheng Long. Transolver++: An accurate neural solver for PDEs on million-scale geometries. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 41432–41449. PMLR, 2025.

Jieyuan Pei, Zhuoxuan Li, Wei Li, Haobo Zhang, Jiawei Jiang, and Jianwei Zheng. Hyperbolic neural operator. In *Proceedings of the 43rd International Conference on Machine Learning*, volume 306 of *Proceedings of Machine Learning Research*. PMLR, 2026.

Maziar Raissi, Paris Perdikaris, and George Em Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.

Ramansh Sharma and Varun Shankar. Accelerated training of physics-informed neural networks (PINNs) using meshless discretizations. In *Advances in Neural Information Processing Systems*, volume 35, 2022.

Yizheng Wang, Zhongkai Hao, Mohammad Sadegh Eshaghi, Cosmin Anitescu, Xiaoying Zhuang, Timon Rabczuk, and Yinghua Liu. Pretrain finite element method: A pretraining and warm-start framework for PDEs via physics-informed neural operators. *Journal of the Mechanics and Physics of Solids*, 214:106682, 2026.

Haixu Wu, Huakun Luo, Haowen Wang, Jianmin Wang, and Mingsheng Long. Transolver: A fast transformer solver for PDEs on general geometries. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 53681–53705. PMLR, 2024.

Haixu Wu, Minghao Guo, Zongyi Li, Zhiyang Dou, Mingsheng Long, Kaiming He, and Wojciech Matusik. GeoPT: Scaling physics simulation via lifted geometric pre-training. In *International Conference on Machine Learning*, 2026.

Zhizhou Zhang, Youjia Wu, Kaixuan Zhang, and Yanjia Wang. From cheap geometry to expensive physics: A physics-agnostic pretraining framework for neural operators. In *International Conference on Learning Representations*, 2026.

Zhenhua Zheng, Lu Zhang, Junhong Zou, Shitong Liu, Zhen Lei, Xiangyu Zhu, and Zhiyong Liu. GA-Field: Geometry-aware vehicle aerodynamic field prediction. *arXiv preprint arXiv:2602.20609*, 2026.

Hang Zhou, Yuezhou Ma, Haixu Wu, Haowen Wang, and Mingsheng Long. Unisolver: PDE-conditional transformers towards universal neural PDE solvers. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 79061–79088. PMLR, 2025.

Junhong Zou, Wei Qiu, Zhenxu Sun, Xiaomei Zhang, Zhaoxiang Zhang, and Xiangyu Zhu. AdaField: Generalizable surface pressure modeling with physics-informed pre-training and flow-conditioned adaptation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 40(2): 1676–1684, 2026.

A Proxy Formulation and Discrete Operators

A.1 Nondimensional variables and mixed formulation

For a nonzero background velocity q, define $U = \|q\|_2$ and $\hat{q} = q/U$. With reference length L and constant density ρ,

$$x = x^{\dim}/L, \qquad v = u^{\dim}/U - \hat{q}, \qquad p = P^{\dim}/(\rho U^2). \tag{14}$$

The Oseen linearization omits disturbance self-advection $(v \cdot \nabla)v$ while retaining advection by $\hat{q}$. Writing $(\nabla v)_{ij} = \partial_j v_i$ and $(\nabla \cdot F)_i = \sum_j \partial_j F_{ij}$, the mixed relation is

$$F = v \otimes \hat{q} + pI - Re^{-1}\nabla v, \qquad \nabla \cdot F = (\hat{q} \cdot \nabla)v + \nabla p - Re^{-1}\Delta v. \tag{15}$$

Equivalence with the Oseen system. Let $\hat{q} \in \mathbb{R}^3$ be spatially constant, $Re > 0$ be constant, $v \in C^2(\Omega_G; \mathbb{R}^3)$, and $p \in C^1(\Omega_G)$. If (v, p) satisfies Eq. 5, define $F_{ij} = v_i \hat{q}_j + p\delta_{ij} - Re^{-1}\partial_j v_i$. Then the constitutive relation in Eq. 7 holds by definition, $\nabla \cdot v = 0$ is unchanged, and, componentwise,

$$\begin{aligned}(\nabla \cdot F)_i &= \sum_{j=1}^{3} \partial_j \big(v_i \hat{q}_j + p\delta_{ij} - Re^{-1}\partial_j v_i\big) \\ &= (\hat{q} \cdot \nabla)v_i + \partial_i p - Re^{-1}\Delta v_i = 0.\end{aligned} \tag{16}$$

Hence (v, p, F) satisfies the mixed system. Conversely, suppose (v, p, F) satisfies all three relations in Eq. 7 with the same regularity and constant coefficients. The constitutive relation fixes F uniquely in terms of (v, p); substituting it into $\nabla \cdot F = 0$ and using the identity above yields the Oseen momentum equation, while the first mixed relation gives incompressibility. The boundary conditions on v and p in Eq. 6 are identical in both formulations. Thus, under the stated regularity and constant-coefficient assumptions, the two systems have the same (v, p) solutions, with a unique associated flux F for each solution. The mixed form therefore preserves the continuum Oseen equations while expressing the training constraints through first-order derivatives. The source readout contains 13 values: three disturbance-velocity components, one pressure, and nine independent flux entries.

A.2 Geometry and boundary-aware first derivatives

Each source input has 40,000 volume, 10,000 body-surface, and 10,000 outer-boundary points. Geometry features encode position, wall distance, and a three-component wall/boundary direction. A four-component condition encodes background velocity and a normalized logarithmic Reynolds number. The physical Re, rather than its network-input encoding, enters the residual. Attention receives all three point sets. Each stored MLS fit uses 40 neighbors, and each fluid-side pressure trace uses 32 neighbors. The minimum valid-row fractions are 0.9 for both MLS derivatives and pressure traces.

For a scalar field f at fluid-volume center x_i, a stored MLS row is

$$(D_{h,k} f)_i = \sum_{j \in \mathcal{N}_i} w_{ij}^{(k)} (f_j - f_i). \tag{17}$$

We obtain the derivative weights from a locally scaled, regularized weighted least-squares fit. Specifically, let $h_i = \max_{j \in \mathcal{N}_i} \|x_j - x_i\|_2$, $\xi_{ij} = (x_j - x_i)/h_i$, and use the quadratic basis $\phi(\xi) = (1, \xi_1, \xi_2, \xi_3, \xi_1^2, \xi_1\xi_2, \xi_1\xi_3, \xi_2^2, \xi_2\xi_3, \xi_3^2)^\top$. If P_i stacks the basis vectors over the 40 neighbors and $W_i = \mathrm{diag}(\exp(-4\|\xi_{ij}\|_2^2) + 10^{-3})$, then

$$\mathbf{w}_i^{(k)} = h_i^{-1} W_i P_i (P_i^\top W_i P_i + \eta_i I)^{-1} e_k, \tag{18}$$

where e_k selects the kth linear coefficient and $\eta_i > 0$ is a small diagonal regularizer. These weights depend only on the local geometry and are precomputed. Thus, a quadratic local fit supplies first-derivative rows without evaluating second-order PDE derivatives. Candidate neighbors are screened for wall-side and nearest-wall consistency. A row enters the residual only if it has an admissible neighborhood and passes conditioning, basis-reproduction, noise-gain, anisotropy, and finite-value checks. For a near-boundary center, the same fit is recomputed with an anchor carrying the prescribed velocity:

$$(D_{h,k}^{\mathrm{bc}} v)_i = \sum_{j \in \mathcal{N}_i} \bar{w}_{ij}^{(k)} (v_j - v_i) + a_i^{(k)} (v_i^{\mathrm{bc}} - v_i). \tag{19}$$

The disturbance velocity is imposed as $-\hat{q}$ at the body and zero at the prescribed far-field points. Volume and outlet velocities remain predicted. Velocity gradients use the boundary-aware rows; flux divergence uses the base first-derivative rows. Residuals are evaluated at valid fluid-volume centers. The augmented derivative rows incorporate prescribed boundary values into the local volumetric constraints.

A.3 Pressure lifting and fluid-side trace

The streamwise box coordinate has an upstream $x_{\min}$ and a downstream $x_{\max}$. A pressure lifting

$$p_\theta(x) = g_{\text{out}}(x)\widetilde{p}_\theta(x), \qquad g_{\text{out}}(x) = \left(\frac{x_{\max} - x_1}{x_{\max} - x_{\min}}\right)^\alpha, \quad \alpha > 0, \tag{20}$$

vanishes at the outlet, enforcing homogeneous pressure on that boundary. The proxy pressure is defined relative to this outlet reference. We use $\alpha = 1$.

At body point s, the fluid-side trace is

$$(T_h p_\theta)_s = \sum_{j \in \mathcal{N}_s^{\text{tr}}} \tau_{sj} p_\theta(x_j). \tag{21}$$

It extrapolates from admissible nearby fluid-volume points. The trace loss compares this value with surface pressure using represented surface areas and valid trace rows, as in Eq. 12. The trace weights come from the constant coefficient of a quadratic weighted least-squares fit in the surface's local tangent–normal frame. Only neighborhoods containing 32 strictly admissible fluid-side points and passing numerical quality checks contribute to the loss; the fitted weights reproduce constants. Surface samples are drawn proportionally to triangle area, so each carries $A_s = |\Gamma_G|/|\mathcal{S}_G|$ before restriction to valid trace rows. For residual r,

$$\mathcal{M}_G(r) = \frac{1}{|\mathcal{V}_G| d_r} \sum_{i \in \mathcal{V}_G} \|r_i\|^2, \qquad d_{\text{div}} = 1, \quad d_{\text{con}} = 9, \quad d_{\text{mom}} = 3. \tag{22}$$

The vector and tensor norms are Euclidean and Frobenius, respectively. Geometry-level reductions precede batch averaging. The full objective uses $(\lambda_{\text{div}}, \lambda_{\text{con}}, \lambda_{\text{mom}}, \lambda_{\text{tr}}) = (1, 1, 1, 1)$ for the divergence, constitutive, momentum, and pressure-trace terms; the partial control sets $\lambda_{\text{con}} = \lambda_{\text{mom}} = 0$ while retaining the other constraints and boundary treatments. Together, the mixed system, boundary-aware derivatives, and fluid-side trace provide first-order physical supervision on geometry-dependent volume and surface point sets.

B Training Data and Experimental Settings

B.1 Pretraining corpus and preprocessing

Table 4 summarizes the source geometry corpus and the five downstream benchmarks. Source counts refer to distinct geometries; downstream partitions count simulation cases. The source corpus has no precomputed velocity, pressure, flux, or other solution fields.

Table 4: Pretraining and downstream datasets. A dash indicates that no separate validation partition is used.

Pretraining: ShapeNetV2 geometries without solution fields					
Category	Airplane	Car	Bus	Watercraft	Total
Geometries	4,020	3,443	925	1,865	10,253
Downstream: high-fidelity simulation cases					
Dataset	Train	Validation	Test	Prediction targets	
Aircraft	140	–	10	$C_p, \rho, \mathbf{u}, p$	
DTCHull	100	–	20	$p, \mathbf{u}$	
NASA-CRM	105	–	44	Surface C_p	
Car-Crash	100	–	30	Peak von Mises stress	
DrivAerNet++	5,817	1,148	1,154	Surface pressure and WSS	

Table 5 summarizes the preprocessing and condition generation used to construct the source inputs. No source CFD solution fields are generated.

Table 5: Source-input preprocessing and proxy-condition generation.

ITEM	CONFIGURATION
Source geometries	ShapeNetV2 airplane (02691156), car (02958343), watercraft (04530566), and bus (02924116); all 10,253 geometries, without a source split.
Fluid-region check	Triangle meshes need not be watertight. Invalid or misregistered BINVOX occupancy is rejected before identifying exterior fluid points.
Coordinates	Rotate coordinates and normals by $(x, y, z) \mapsto (z, y, -x)$; ground at $y = 0$, center horizontally, and scale the transformed x extent to $L_{\mathrm{ref}} = 1$.
Points per geometry	40,000 exterior fluid-volume points, 10,000 body-surface points with normals, and 10,000 outer-boundary points.
Volume allocation	Global / near-wall / wake: nominally 35% / 45% / 20%; difficult meshes receive more near-wall samples.
Nearest-wall features	$[x, y, z, d_{\mathrm{wall}}, r_x, r_y, r_z]$ for volume and outer-boundary points, where $\mathbf{r}$ points from the nearest wall to the query point.
Discrete templates	First-order volume MLS with 40 neighbors and fluid-side surface-pressure trace with 32 valid volume neighbors; no second-order or surface PDE derivative stencil.
Proxy conditions	100 deterministic draws per geometry: yaw $[-20^\circ, 20^\circ]$, pitch $[-5^\circ, 5^\circ]$, speed $[0.8, 1.2]$, and $Re \in [10, 300]$.

The 100 conditions yield 1,025,300 available geometry–condition tasks. During training, each geometry uses one of its stored conditions per epoch; the selected index advances by 37 modulo 100 across epochs. The network receives $c = [q_x, q_y, q_z, \log Re]$, while the proxy residual uses $\hat{q} = q/\|q\|_2$ for directional transport and prescribed velocity conditions and Re for viscous diffusion. The latter is a dimensionless proxy parameter at $L_{\mathrm{ref}} = 1$, not a full-scale aircraft or vehicle Reynolds number.

B.2 DOWNSTREAM TASKS AND TRAINING SETTINGS

Table 4 lists the benchmark partitions and prediction targets. A case denotes one simulated geometry–condition instance. Aircraft predicts $[C_p, \rho, u_x, u_y, u_z, p]$; DTCHull predicts $[p, u_x, u_y, u_z]$; NASA-CRM predicts surface C_p; and Car-Crash predicts the maximum von Mises stress over the impact trajectory. DrivAerNet++ evaluates surface pressure and wall shear stress (WSS). These target fields retain their benchmark conventions and are distinct from the nondimensional source proxy fields.

Our training settings. Table 6 reports the configurations used for our method. The downstream settings apply to the four cross-physics tasks; DrivAerNet++ is listed separately. The pretrained Transolver backbone has eight layers, hidden width 256, eight attention heads, and 32 Physics-Attention slices per layer.

Label-budget protocol. The reduced-label experiments use nested 20%, 40%, 60%, and 80% subsets of training cases. The corresponding case counts are 28/56/84/112 for Aircraft, 20/40/60/80 for DTCHull and Car-Crash, and 21/42/63/84 for NASA-CRM. The test partitions remain unchanged.

Table 6: Training configurations for our method.

ITEM	CONFIGURATION
Proxy pretraining	
Optimizer	AdamW
Base learning rate	1.0×10^{-4}
Global batch size	48 geometry–condition tasks
Epochs	300
Hardware	2 NVIDIA A800 80 GB GPUs
Training compute	103.94 GPU-hours
Four cross-physics tasks: adaptation	
Optimizer	AdamW
Base learning rate	1.2×10^{-3}
Batch size	1 case
Epochs	200
Hardware	1 NVIDIA RTX 4090 48GB GPU
DrivAerNet++: adaptation	
Optimizer	AdamW
Base learning rate	1.5×10^{-3}
Epochs	200

B.3 EVALUATION METRICS AND AGGREGATION

For a chosen field, flatten its valid entries within case c to a vector y_c of length M_c, and set $e_c = \widehat{y}_c - y_c$. We use

$$E_{2,c} = \frac{\|e_c\|_2}{\|y_c\|_2}, \qquad E_{1,c} = \frac{\|e_c\|_1}{\|y_c\|_1}, \tag{23}$$

$$\mathrm{MSE}_c = \frac{\|e_c\|_2^2}{M_c}, \qquad \mathrm{MAE}_c = \frac{\|e_c\|_1}{M_c}, \tag{24}$$

$$\mathrm{RMSE}_c = \sqrt{\mathrm{MSE}_c}, \qquad R_c^2 = 1 - \frac{\|e_c\|_2^2}{\sum_{k=1}^{M_c}(y_{ck} - \bar{y}_c)^2}. \tag{25}$$

Each metric is computed per case and then averaged across test cases, giving equal weight to geometries with different point counts. Mean RMSE and the square root of mean MSE correspond to different averaging orders. For the four-task summaries, MaxAE is the global maximum $\max_c \max_k |e_{ck}|$.

Overall and field-wise relative error. The main four-task comparison computes overall relative L2 after concatenating all six Aircraft channels or all four DTCHull channels. NASA-CRM and Car-Crash each have one output channel. Table 8 reports field-wise errors to show the contribution of individual physical quantities. The aggregate multi-field metric retains the channel units and scales of each benchmark. For DTCHull, the pressure and grouped-velocity R^2 values are 0.9565 and 0.9618, respectively.

DrivAerNet++ evaluation space. Pressure benchmarking uses the fixed transformation

$$p^\star = \frac{p - (-94.5)}{117.25}, \qquad \widehat{p}^\star = \frac{\widehat{p} - (-94.5)}{117.25}. \tag{26}$$

We apply this transformation to both decoded predictions and targets before computing the reported pressure metrics. For WSS, first form the scalar magnitudes $m_i = \|\boldsymbol{\tau}_i\|_2$ and $\widehat{m}_i = \|\widehat{\boldsymbol{\tau}}_i\|_2$, then evaluate every reported WSS metric on $\widehat{m}$ and m. The sample-averaged MaxAE convention used for the main DrivAerNet++ comparison is

$$\mathrm{MaxAE}_{\text{case mean}} = \frac{1}{C}\sum_{c=1}^{C} \max_i |\widehat{y}_{ci} - y_{ci}|, \tag{27}$$

where $y = p^\star$ for pressure and $y = m$ for WSS. This convention averages the maximum scalar error across test cases.

C Additional Quantitative Results

C.1 Source-training diagnostics

Table 7 summarizes proxy residuals and valid discrete-operator rows over 526 source geometries at selected epochs. By epoch 300, the divergence, constitutive, and momentum RMS residuals are below 0.008, while both valid-row fractions exceed 0.97.

Table 7: Source-training diagnostics. RMS residuals are averaged over 526 geometries at each displayed epoch; valid fractions measure usable discrete-operator rows. Bold marks the preferable value in each row (lower RMS or higher valid fraction, including ties).

Diagnostic	Epoch 1	Epoch 300
Divergence RMS	2.0557	**0.00708**
Constitutive RMS	0.2872	**0.00749**
Momentum RMS	0.3270	**0.00353**
Pressure-trace RMS	0.02245	**0.000115**
Valid MLS fraction	**0.9943**	**0.9943**
Valid trace fraction	**0.9782**	0.9770

C.2 Field-wise results and label efficiency

Table 8 separates the coupled outputs that are combined in the overall relative error. Velocity metrics concatenate the three vector components. The full proxy model improves all six listed field-relative errors over both controls, covering pressure, velocity, density, and surface-pressure prediction. For DTCHull, the field-wise breakdown shows gains in both pressure and grouped velocity.

Table 8: Field-wise full-label test results. Relative L2 is a fraction, not a percentage. Metrics are averaged per test case. Bold marks the best value for each field and metric.

		Scratch		w/o con+mom		Ours	
Dataset	Field	Rel. L2	R^2	Rel. L2	R^2	Rel. L2	R^2
Aircraft	C_p	0.2896	0.8906	0.2152	0.9378	**0.0961**	**0.9879**
Aircraft	ρ	0.1630	0.8832	0.1017	0.9433	**0.0566**	**0.9853**
Aircraft	$\mathbf{u}$	0.0632	0.9936	0.0457	0.9964	**0.0263**	**0.9987**
Aircraft	p	0.1363	0.9475	0.0992	0.9650	**0.0626**	**0.9889**
DTCHull	p	0.1876	0.9172	0.1572	0.9446	**0.1362**	**0.9565**
DTCHull	$\mathbf{u}$	0.1727	0.9553	0.1628	0.9601	**0.1595**	**0.9618**

Table 9 gives the data underlying Figure 4, using the overall relative error defined in Appendix B. It compares scratch and proxy-pretrained initialization at all five label budgets on the four downstream tasks. The pretrained model matches the full-data scratch error with 40% of labels on Aircraft, 60% on DTCHull, and 80% on NASA-CRM and Car-Crash.

Table 9: Label-efficiency results. Columns are fractions of the original training pool. Entries are overall relative L2 errors; bold marks the lower error at each budget.

Initialization	20%	40%	60%	80%	100%
Aircraft					
Scratch	0.1789	0.1592	0.1178	0.1142	0.0922
Ours	**0.1346**	**0.0598**	**0.0499**	**0.0423**	**0.0405**
DTCHull					
Scratch	0.4277	0.3197	0.2649	0.2421	0.1876
Ours	**0.3331**	**0.2132**	**0.1838**	**0.1620**	**0.1362**
NASA-CRM					
Scratch	0.5148	0.3155	0.1753	0.1178	0.1091
Ours	**0.2975**	**0.1891**	**0.1523**	**0.0958**	**0.0866**
Car-Crash					
Scratch	0.5611	0.4309	0.2803	0.2222	0.1971
Ours	**0.5334**	**0.4167**	**0.2658**	**0.1965**	**0.1762**

C.3 Additional error metrics

Tables 10–13 supplement the main relative-L2 comparison with relative L_1, MSE, RMSE, MAE, and global MaxAE on the test sets. Absolute errors use each benchmark's native decoded units. Across all four tasks, the full proxy model attains the lowest relative L_1, MSE, RMSE, and MAE among the methods reported in these tables. Global MaxAE measures the largest local error and can rank methods differently from the averaged metrics.

Table 10: Additional Aircraft metrics on all six outputs. Lower is better; bold marks the best value in each column.

Method	Rel. L_1	MSE	RMSE	MAE	Global MaxAE
GNOT-base	0.1193	0.2597	0.4474	0.1991	21.7613
UPT-small	0.1908	0.5436	0.6754	0.3278	24.7645
Transolver++	0.0675	0.0512	0.2125	0.1095	21.0623
LinearNO	0.0354	0.0203	0.1350	0.0559	13.9061
HNO	0.0436	0.0291	0.1566	0.0691	**12.8131**
GeoPT	0.0279	0.0152	0.1115	0.0450	19.5201
Scratch	0.0739	0.0776	0.2560	0.1213	27.5317
w/o con+mom	0.0518	0.0313	0.1691	0.0824	15.7990
Ours	**0.0255**	**0.0146**	**0.1087**	**0.0407**	13.8763

Table 11: Additional DTCHull metrics on pressure and three velocity components. Lower is better; bold marks the best value in each column.

Method	Rel. L_1	MSE	RMSE	MAE	Global MaxAE
GNOT-base	0.1217	37.0234	5.7309	1.2636	876.1799
UPT-small	0.3017	187.0740	12.9211	3.1284	1350.4707
Transolver++	0.1654	43.3988	6.3167	1.7217	877.7742
LinearNO	0.0958	19.9225	4.2822	0.9976	814.8098
HNO	0.0968	19.7824	4.2461	1.0082	887.1795
GeoPT	0.0773	14.7365	3.7019	0.8036	**622.2211**
Scratch	0.1112	24.6756	4.7988	1.1544	1147.0984
w/o con+mom	0.0887	18.0220	4.0461	0.9251	963.4878
Ours	**0.0708**	**12.9855**	**3.4858**	**0.7373**	638.5015

Table 12: Additional NASA-CRM surface C_p metrics. Lower is better; bold marks the best value in each column.

Method	Rel. L_1	MSE	RMSE	MAE	Global MaxAE
GNOT-base	0.1327	0.0034	0.0565	0.0299	2.8039
UPT-small	0.5032	0.0426	0.1935	0.1140	2.8018
Transolver++	0.0814	0.0014	0.0340	0.0185	2.7587
LinearNO	0.0861	0.0014	0.0347	0.0195	2.8266
HNO	0.0786	0.0013	0.0325	0.0179	2.7367
GeoPT	0.0658	0.0012	0.0297	0.0151	2.9000
Scratch	0.0841	0.0015	0.0347	0.0192	2.7011
w/o con+mom	0.0884	0.0016	0.0367	0.0201	**2.6340**
Ours	**0.0634**	**0.0010**	**0.0277**	**0.0145**	2.6884

Table 13: Additional Car-Crash maximum von Mises stress metrics. Lower is better; bold marks the best value in each column.

Method	Rel. L_1	MSE	RMSE	MAE	Global MaxAE
GNOT-base	0.2636	683.7354	26.0200	17.9528	636.9171
UPT-small	0.3824	1.252×10^3	35.2860	24.9957	704.0164
Transolver++	0.1667	295.3389	17.0224	11.4121	627.8551
LinearNO	0.1938	382.5221	19.4076	13.2321	634.1409
HNO	0.1701	305.8686	17.3262	11.6448	637.5948
GeoPT	0.1995	404.7718	19.9464	13.4724	624.7348
Scratch	0.1789	335.0427	18.1522	12.2328	630.4558
w/o con+mom	0.1656	289.9019	16.8685	11.3324	628.1026
Ours	**0.1580**	**268.7784**	**16.2434**	**10.8150**	**623.9977**

D Additional Visualizations

D.1 Downstream predictions and error maps

Figures 5–8 show selected test cases from the four cross-physics tasks. Aircraft and DTCHull each display two fields from the same case. Within each panel, references and predictions share the viewpoint and field color scale; signed-error maps use a common scale, with blue for underprediction and red for overprediction. Color limits are chosen for visibility, while all quantitative metrics are computed from the original test fields.

Across these selected cases, full-proxy pretraining reduces the extent of surface-pressure errors on Aircraft and NASA-CRM, while its DTCHull and Car-Crash predictions reproduce near-hull flow and front-body stress patterns. These observations complement the quantitative results in Tables 1 and 2 across distinct target physics.

DrivAerNet++ surface predictions and error atlases.

These plots use validation-selected checkpoints. Pressure is displayed in native decoded units; the normalized-pressure benchmark instead uses Eq. (26). For WSS the displayed scalar is the three-component magnitude, and its error is the difference between predicted and reference magnitudes.

The predicted pressure and WSS retain the main spatial patterns in Figure 9; the error atlases locate the remaining discrepancies across test vehicles. This complements the data-rich gains in Table 3.

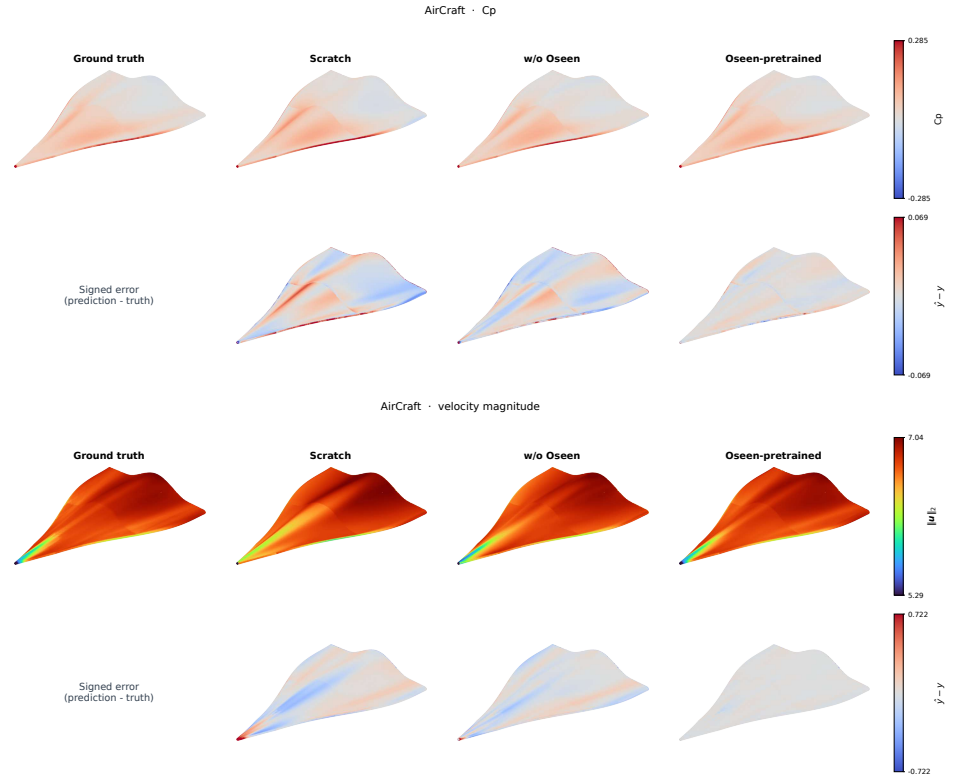

Figure 5: Aircraft: surface C_p (upper panel) and speed (lower panel) for one test geometry. In each panel the top row shows reference and predictions, and the bottom row shows signed errors. Full proxy pretraining reduces broad C_p discrepancies across the swept surface visible in the scratch and partial-objective maps. Speed is shown for readability; the quantitative velocity relative error evaluates the full vector.

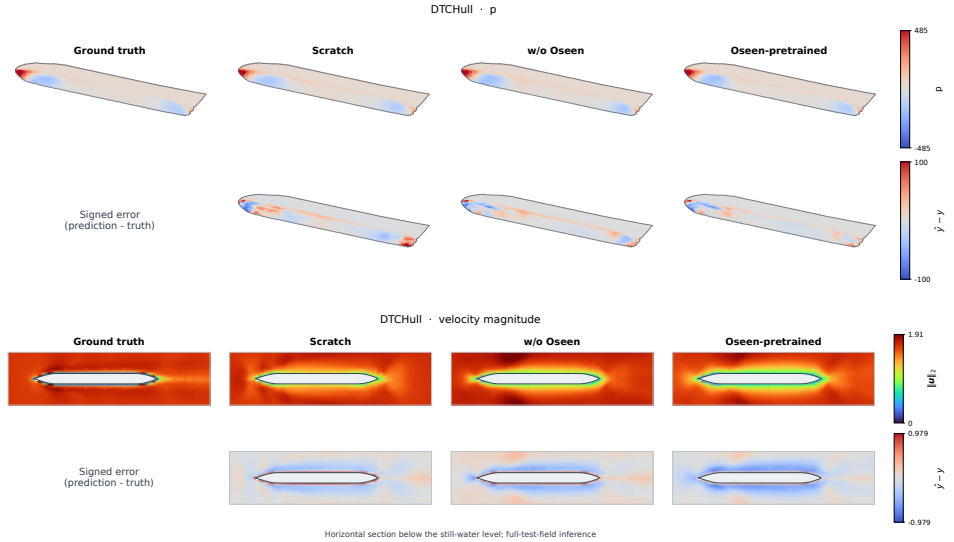

Figure 6: DTCHull: hull pressure and an underwater speed section from one test case. Hull pressure uses the original surface connectivity. The speed view intersects the volume mesh at 60% of the still-water draft measured upward from the keel. All methods receive the same 229,000-point test input. The gray region masks the solid hull, and display interpolation is used only for visualization. The shared-scale error maps show differences near the hull ends and in the near-wall flow.

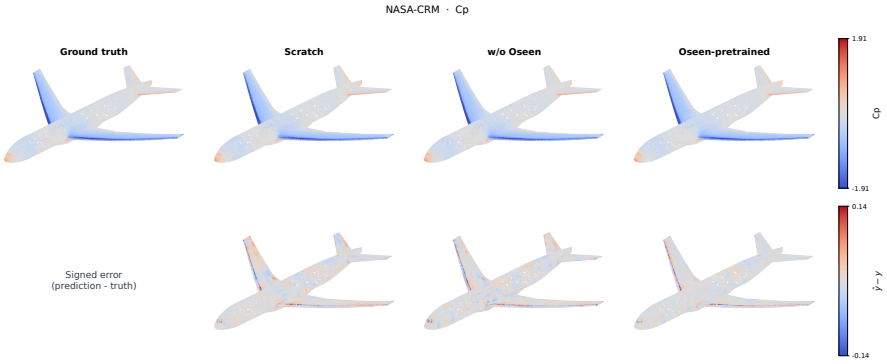

Figure 7: NASA-CRM surface C_p predictions and signed errors. The full-proxy model more closely reproduces the wing and tail pressure patterns, with shared scales exposing spatial differences among methods.

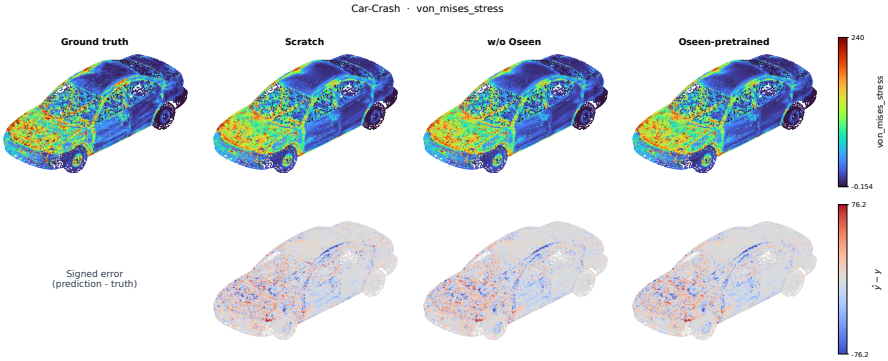

Figure 8: Car-Crash maximum von Mises stress predictions and signed errors. After target-task adaptation, the full-proxy model captures the stress concentration on the front body and structural edges. The shared-scale error maps compare local deviations across methods.

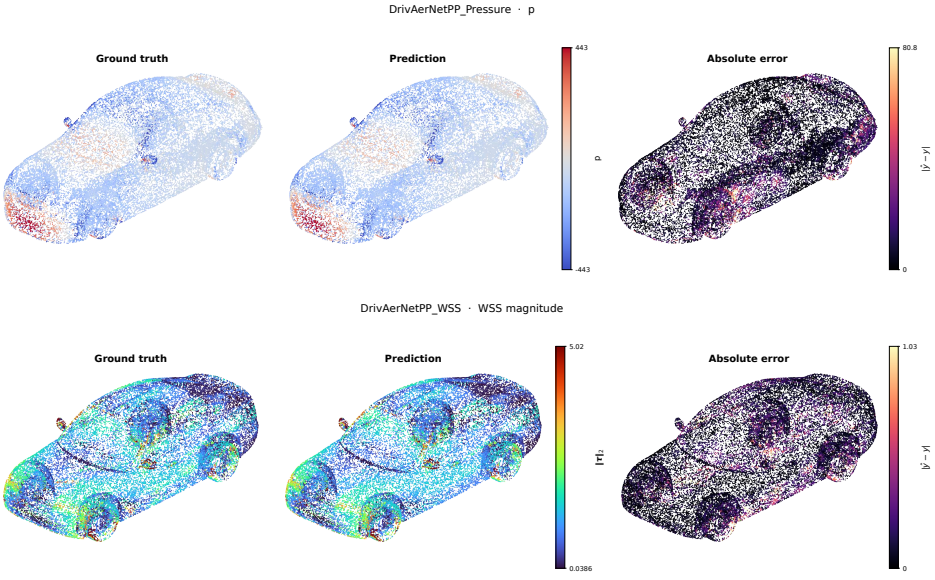

Figure 9: DrivAerNet++ surface pressure (top) and WSS magnitude (bottom). Columns show the reference, prediction, and absolute scalar error. Pressure and WSS panels use their respective test examples. The maps show predicted field structure and local scalar errors on complex vehicle surfaces.

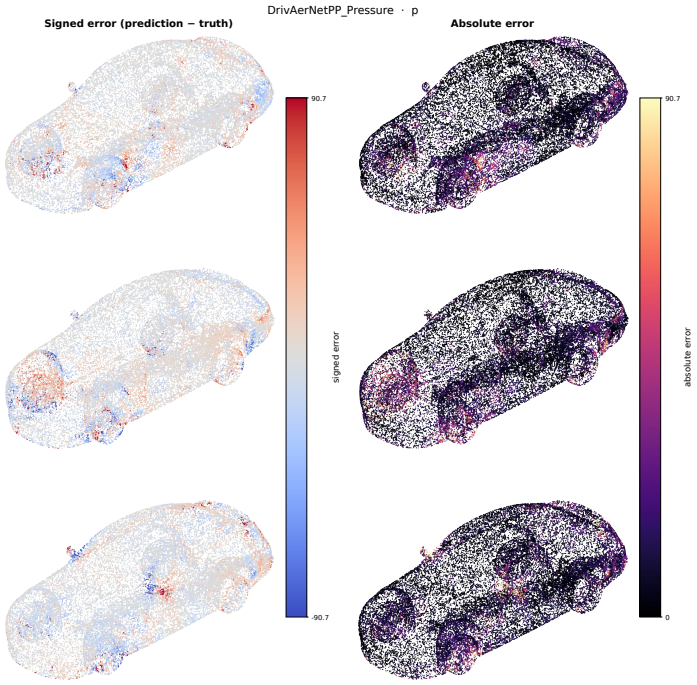

Figure 10: Pressure error atlas over three DrivAerNet++ test cases. Signed errors retain over-/underprediction information, while absolute errors emphasize localized discrepancies. All rows share the display error range. Pressure errors are shown in native decoded units; evaluation metrics use the transformation in Eq. 26.

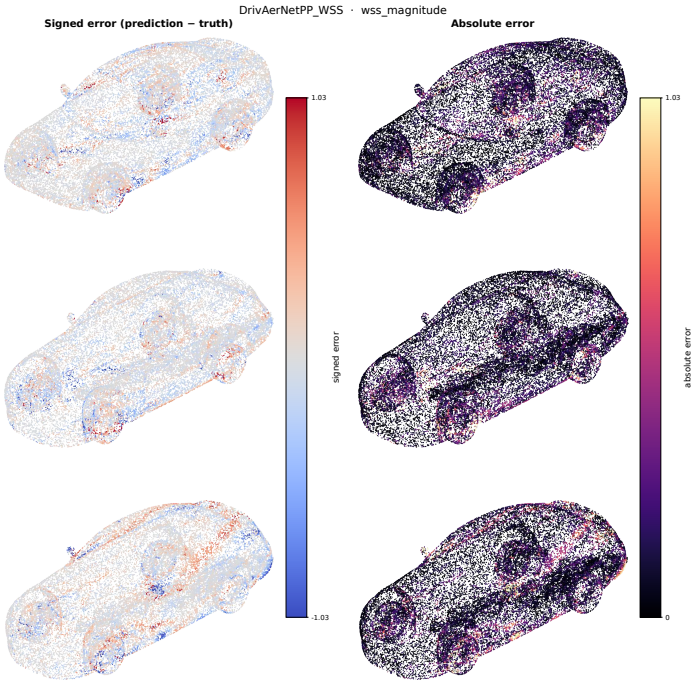

Figure 11: WSS-magnitude error atlas over three existing test examples. Errors are computed from predicted and reference vector magnitudes. A shared display scale shows spatial error patterns across surface regions and geometries.

D.2 SOURCE-TASK RESPONSES AND MODEL ATTENTION

Figure 12 extends the NASA-CRM example in the main text to Aircraft, DTCHull, and DrivAerNet++ geometries. All use the same source checkpoint before downstream adaptation and the following three physical conditions:

$$\begin{aligned} c_1 &= (0.80, -20.00^\circ, -2.00^\circ, 10.00), \\ c_2 &= (0.86, -10.26^\circ, -1.47^\circ, 37.63), \qquad c = (U, \text{yaw}, \text{pitch}, Re). \\ c_3 &= (1.11, \quad 6.19^\circ, \quad 1.98^\circ, 13.49), \end{aligned} \tag{28}$$

The displayed velocity is the magnitude of the dimensionless total field $\hat{q} + v_\theta$. In the plots, the colorbar notation p/U^2 denotes the model output $p_\theta = P^{\text{dim}}/(\rho U^2)$ defined in Eq. 14. The near-center-plane views show median values in hexagonal bins; the silhouette and empty bins are masked. Each geometry and field uses a common scale across the three conditions, with separate scales across geometries.

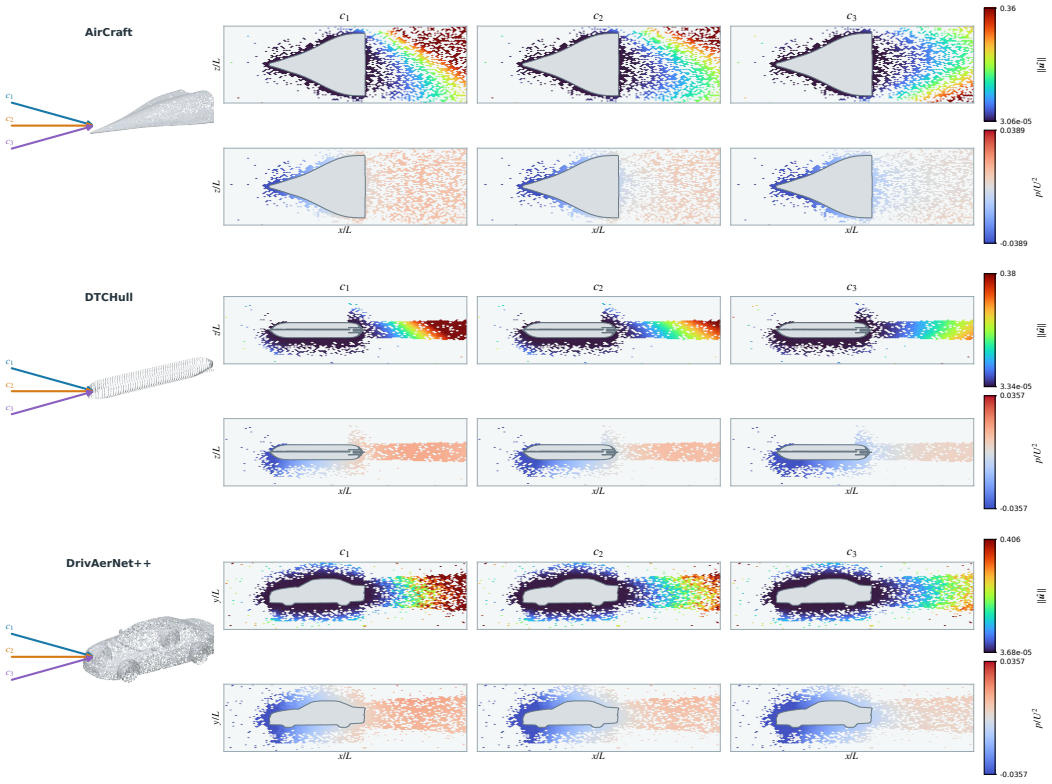

Figure 12: Additional source-task responses under the conditions in Eq. (28). Each geometry shows velocity magnitude and proxy pressure in two-dimensional near-center-plane views. Aircraft and DTCHull use the x–z view; the car uses the x–y view. The samples exhibit geometry-dependent low-velocity regions and changes in downstream and pressure patterns during proxy pretraining.

Condition-dependent proxy responses. For each fixed geometry, jointly varying the flow condition changes the spatial distribution and amplitude of proxy velocity and pressure. Near-body patterns and downstream asymmetry remain geometry-dependent. These qualitative responses complement the numerical proxy-residual measurements in Table 7. Observed before target adaptation on three geometry classes, these responses illustrate how one shared network couples geometry and conditions to proxy fields.

Condition-dependent final-layer attention.

Figures 13–16 show point-to-slice softmax assignment weights from head 0 of the pretrained model's last Physics-Attention layer. Each condition block contains the same 32 slices in row-major order, with fixed slice indices across the conditions in Eq. 28.

For slice s, let $a_s(x, c)$ be its assignment probability. Display values use the 0.5th and 99.5th percentiles pooled over surface points and all three conditions:

$$\widetilde{a}_s(x, c) = \mathrm{clip}_{[0,1]}\left(\frac{a_s(x, c) - Q_{0.005,s}}{Q_{0.995,s} - Q_{0.005,s}}\right). \tag{29}$$

Pooling the normalization range across conditions enables comparison of the same slice as c varies. The colorbar shows the resulting display scale for each slice. For example, NASA-CRM raw probabilities span approximately 0.0275–0.0348, near the uniform value $1/32$; rescaling makes the condition-dependent variation visible. Surface interpolation is used only for rendering.

These maps show condition-sensitive spatial allocation within the network, including changes around noses, wings, hull ends, and car-body surfaces. They provide a model-level view of the condition dependence also visible in the proxy-field responses.

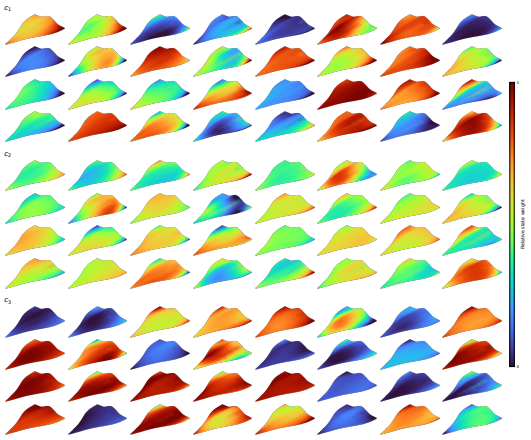

Figure 13: Final-layer surface-to-slice assignment heatmaps for Aircraft. Blocks correspond to c_1, c_2, c_3 and show all 32 slices of one fixed head. Display normalization follows Eq. (29); colors are comparable across conditions within a slice.

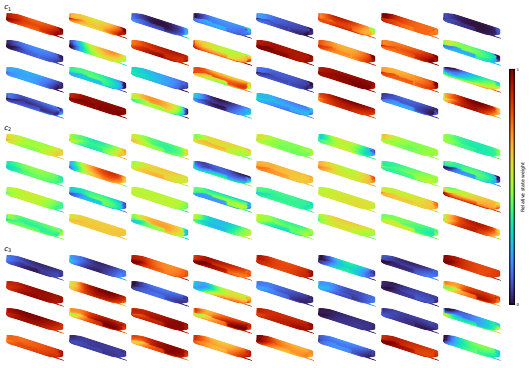

Figure 14: Final-layer surface-to-slice assignment heatmaps for DTCHull. Original hull connectivity is used for rendering. The head, slice order, and per-slice condition-pooled normalization remain fixed across blocks.

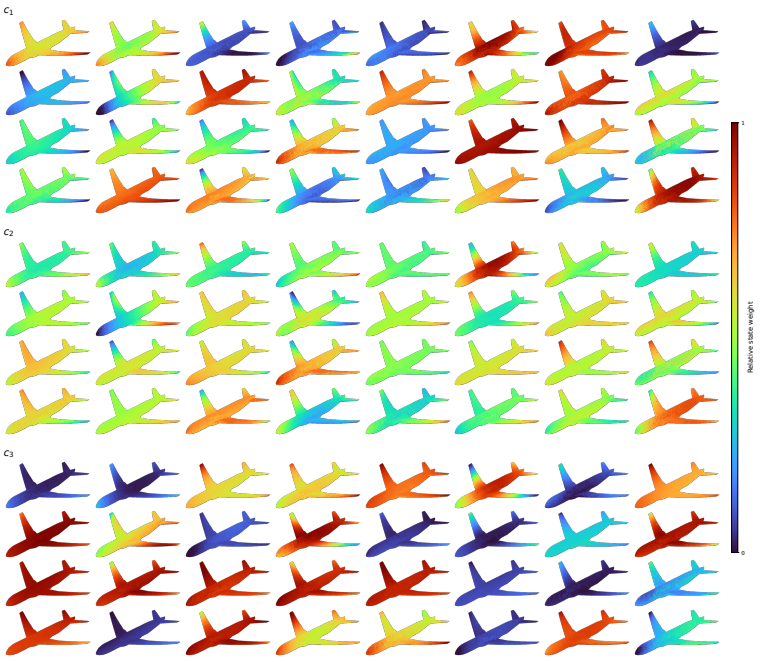

Figure 15: Final-layer surface-to-slice assignment heatmaps for NASA-CRM. Maps reveal relative spatial changes within each slice, including wing, fuselage, and tail regions. Display contrast follows the per-slice normalization in Eq. 29.

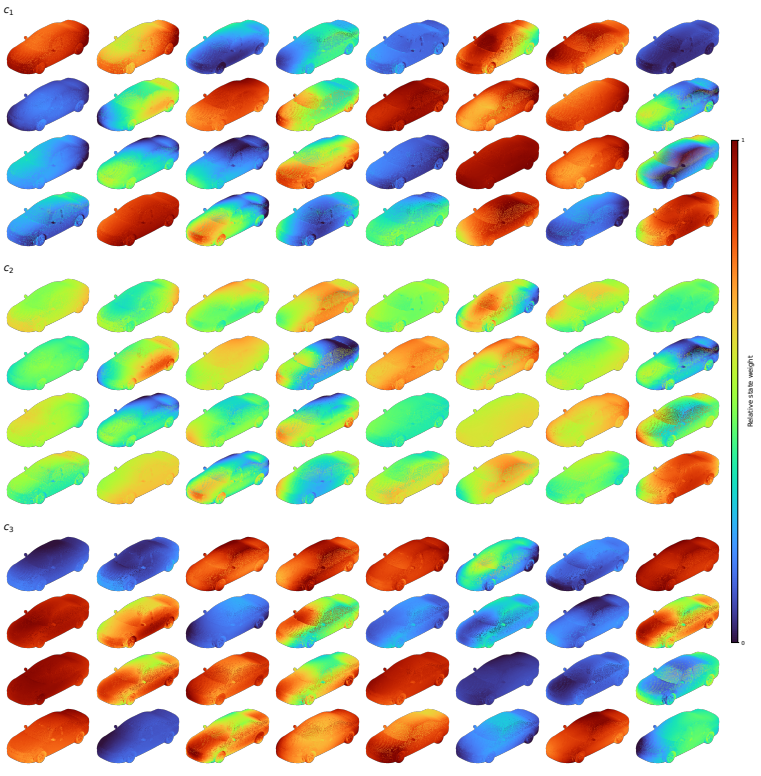

Figure 16: Final-layer surface-to-slice assignment heatmaps for DrivAerNet++. All panels use the pretrained model before fine-tuning and share the per-slice normalization across conditions.