CG-QLEAR: COUNTERFACTUAL GAIN GUIDED RESIDUAL CORRECTION FOR VEHICLE SURFACE PRESSURE MODELING

Anonymous authors
Paper under double-blind review

ABSTRACT

Accurate vehicle surface pressure prediction requires resolving local errors while respecting applicable geometric symmetries. Yet a large prediction error or candidate residual does not necessarily indicate a beneficial correction. We introduce CG-QLEAR, a framework that selects residual updates by their predicted counterfactual gains. A candidate generator proposes local pressure updates, and a gain estimator learns the reduction in absolute error obtained by applying each update. Correction decisions and residual values are coupled across eligible mirror pairs, while asymmetric regions remain independent. A fixed point budget controls the selected residual-correction stages before final refinement. We compare structured correction with continued backbone optimization from the same 55 epoch checkpoint. The protocol uses 5,817 training, 1,148 validation, and 1,154 test vehicles from DrivAerNet++, with 32,768 surface points per vehicle. Against 15 additional backbone epochs, CG-QLEAR reduces pooled MSE by 18.48%, pooled MAE by 3.77%, and mean per vehicle maximum absolute error by 12.52%, while using 4.45% less logged optimization and validation time after the shared checkpoint, excluding one-time setup. The same correction interface also improves all five reported metrics on a Transolver backbone, supporting counterfactual gain as a useful criterion for allocating local pressure updates.

1 INTRODUCTION

Predicting surface pressure from vehicle geometry can provide a fast approximation to computational fluid dynamics (CFD). The pressure field describes spatial effects around components such as the front end, mirrors, and underbody, so both global accuracy and local errors matter. DrivAerNet++ provides vehicle geometries and CFD fields at a scale that supports systematic evaluation of learned surrogates (Elrefaie et al., 2024). Each vehicle in our study is represented by 32,768 sampled surface points, making accurate local prediction a substantial learning problem.

Recent models have improved the geometric representation used for physical field prediction. GINO handles irregular geometry through graph and operator computation, while Transolver learns interactions on general geometries through physics attention (Li et al., 2023; Wu et al., 2024a). AdaField adapts a pressure predictor to flow conditions and provides a strong backbone for vehicle surfaces (Zou et al., 2026). These advances leave a practical question after backbone training: where should a limited amount of additional computation go? Continuing to optimize the entire predictor treats all surface regions alike, although remaining errors may be concentrated and may differ in how reliably they can be corrected.

A local correction method must solve two problems. It must propose a signed residual and decide whether applying that particular residual helps. Refinement methods have shown the value of improving an initial physical prediction (Lippe et al., 2023; Zheng et al., 2026), but selecting corrections by base error or residual magnitude can be misleading. A large candidate may overshoot the CFD target or point in the wrong direction. The useful quantity is the reduction in error caused by applying the candidate, which can be negative even where the original error is large. Selective prediction similarly distinguishes a decision's value from a model's raw confidence (Yuksekgonul et al., 2023; Mao et al., 2024).

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.

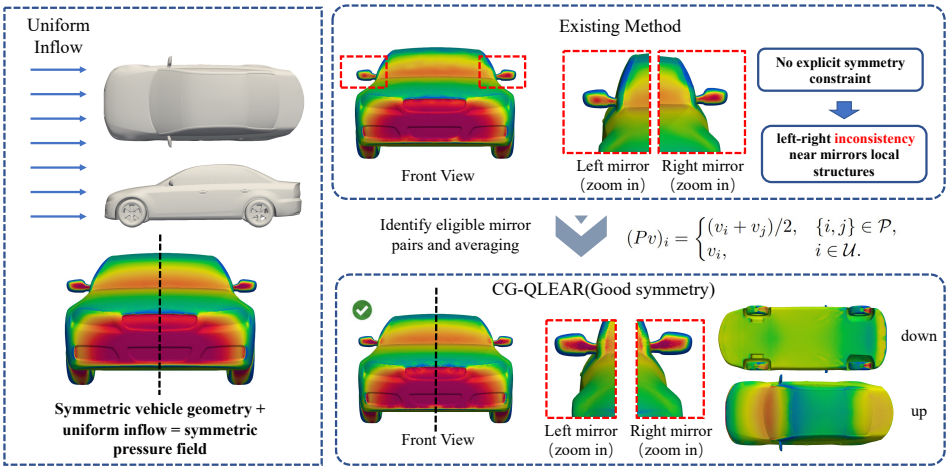


Figure 1: Motivation for partial symmetry correction. Under uniform inflow, corresponding regions should have consistent pressure. CG-QLEAR couples eligible mirror corrections while leaving asymmetric or unmatched regions independent.

We introduce CG-QLEAR, a framework that proposes local pressure updates and selects them by predicted counterfactual gain, the error reduction from applying a proposed update. A Gain Gate learns this quantity from CFD-labeled training vehicles and predicts it without pressure labels at inference. Eligible hard-exact mirror pairs share residual values and selection decisions, while soft-matched and unmatched regions remain independent, as Figure 1 illustrates (Xu et al., 2024; Vargiemezis et al., 2025). Under a fixed point budget, the method selects localized residual corrections before a bounded MLP Median Refinement stage (Xue et al., 2023; Anagnostidis et al., 2024).

Our contributions are counterfactual supervision for update selection, partial symmetry for paired correction, and a budgeted formulation with explicit point costs and a conditional selection bound. To test whether these choices use additional training resources effectively, both experimental routes start from the same 55 epoch backbone checkpoint and share the data, normalization, and evaluation protocol. On 1,154 unseen vehicles, CG-QLEAR reduces pooled MSE by 18.48% relative to continued AdaField training with less logged post-checkpoint optimization and validation time. Stage ablations and budget sweeps examine the correction mechanism, and a Transolver experiment tests its transfer to another backbone.

2 Related Work

2.1 Learning Surface Pressure Fields from Geometry

Neural operators address complex domains through attention, spectral representations, neural fields, implicit equilibrium models, and iterative refinement (Hao et al., 2023; Wu et al., 2023; Raonic et al., 2023; Serrano et al., 2023; Marwah et al., 2023; Lippe et al., 2023). For irregular geometry, GINO combines graph and operator computation, while Transolver and Transolver++ learn physics aware interactions on general geometries (Li et al., 2023; Wu et al., 2024a; Luo et al., 2025). Reusable physics models and scalable point networks further improve representation capacity (Alkin et al., 2024; Herde et al., 2024; Wu et al., 2024b; Liang et al., 2024). AdaField targets surface pressure through flow conditioned adaptation (Zou et al., 2026). These methods mainly improve representation and global field prediction. Our framework can use such models as backbones and allocates later computation to their remaining local errors.

2.2 Residual Refinement, Selective Prediction, and Adaptive Computation

Residual prediction underlies boosting and recent refinement methods for physical systems (Friedman, 2001; Lippe et al., 2023; Berman & Peherstorfer, 2023; Zheng et al., 2026). Selective prediction estimates when to defer, while adaptive computation changes sequence length or token count

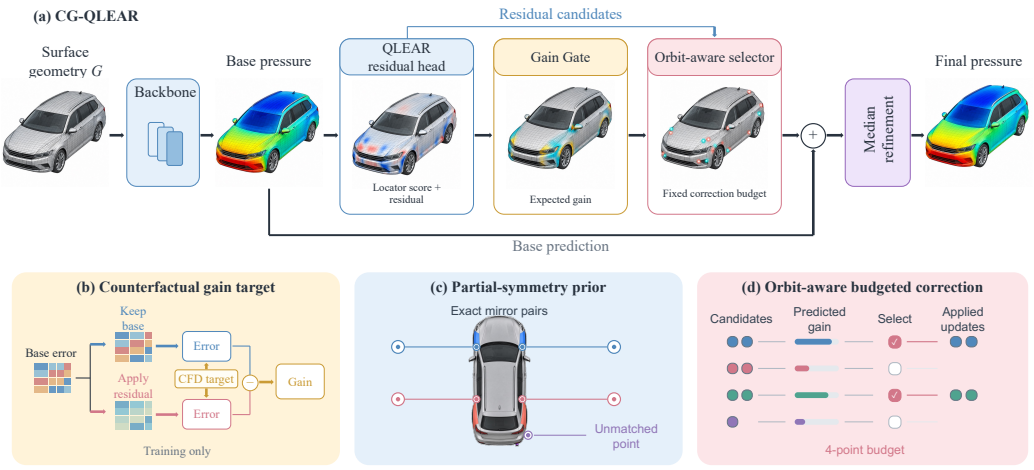


Figure 2: CG-QLEAR overview. (a) QLEAR proposes residuals and the Gain Gate scores their benefits. (b) Counterfactual gain measures error reduction after correction. (c) Eligible mirror regions share updates. (d) QLEAR first selects a fixed locator-based set; the Gain Gate adds positive-gain units under a point budget, followed by bounded MLP Median Refinement.

under limited resources (Yuksekgonul et al., 2023; Mao et al., 2024; Xue et al., 2023; Bolya et al., 2023; Chen et al., 2023; Anagnostidis et al., 2024). CG-QLEAR instead selects local pressure updates. Its gate predicts the realized utility of a fixed candidate, rather than general confidence or token importance. Freezing the candidate generator during gate training separates update quality from selection quality.

2.3 Symmetry in Physical Field Learning

Equivariant convolutions, mesh networks, and graph operators encode known geometric transformations (Passaro & Zitnick, 2023; Trang et al., 2024; Xu et al., 2024). Canonicalization, frame averaging, and latent discovery extend symmetry handling to broader settings (Kaba et al., 2023; Lin et al., 2024; Yang et al., 2024); WindMiL applies reflection equivariance to pressure coefficients (Vargiemezis et al., 2025). Standard equivariance relates outputs under transformed inputs. Our constraint acts within one vehicle: it couples selected corrections only in eligible mirror regions and preserves independent predictions elsewhere.

3 Method

3.1 Problem Setup and Structured Correction Objective

Consider a vehicle surface represented by N sampled points. The input is $X = (\xi_i)_{i=1}^N$, where ξ_i contains the coordinate, surface normal, and component identity of point i. The normalized pressure target is $p = (p_i)_{i=1}^N \in \mathbb{R}^N$. A backbone f_θ with parameters θ produces the initial field $y = f_\theta(X) \in \mathbb{R}^N$. The backbone is frozen throughout correction training. For $r \in \{1, 2\}$, define the normalized pointwise loss

$$\ell_r(v, p) = \frac{1}{N} \sum_{i=1}^{N} |v_i - p_i|^r, \qquad v \in \mathbb{R}^N. \tag{1}$$

Thus, ℓ_1 is mean absolute error and ℓ_2 is mean squared error in the shared pressure scale.

The correction stage operates on a partition $\mathcal{C}$ of the point indices. Each unit $C \in \mathcal{C}$ is either an eligible mirror pair or an independent singleton, as defined in Section 3.3. Let $c_C = |C| \in \{1, 2\}$ denote its point cost, $s_C \in \{0, 1\}$ its selection variable, and $\mathbf{1}_C \in \{0, 1\}^N$ its indicator vector. Write $s = (s_C)_{C \in \mathcal{C}}$ for the complete selection. Let $\odot$ denote elementwise multiplication, and let

$K \in \{0, \dots, N\}$ be the maximum number of corrected points. Given a symmetry compatible base field $b \in \mathbb{R}^N$ and candidate residual $\delta \in \mathbb{R}^N$, the selected field and feasible selection set are

$$q_s = b + \sum_{C \in \mathcal{C}} s_C (\mathbf{1}_C \odot \delta), \qquad \mathcal{S}_K = \left\{ s \in \{0,1\}^{|\mathcal{C}|} : \sum_{C \in \mathcal{C}} c_C s_C \le K \right\}, \tag{2}$$

Equation 2 separates three questions: which residual to propose, which units to update, and how to preserve the valid symmetry relations.

3.2 QLEAR Candidate Residual Generation

QLEAR proposes a signed update before the method decides whether to apply it. Let $\mathcal{N}(i)$ be the geometric neighborhood of point i, and let $|\mathcal{N}(i)|$ denote its size. We compute the local pressure contrast

$$h_i = y_i - \frac{1}{\max\{1, |\mathcal{N}(i)|\}} \sum_{j \in \mathcal{N}(i)} y_j. \tag{3}$$

The 13-dimensional standardized feature z_i concatenates canonicalized coordinates and normals, component identity, the backbone prediction y_i, the contrast h_i, and four mirror descriptors (normalized mirror distance, mirror-normal cosine, pair-validity, and hard-exactness). A shared encoder H_ϕ with parameters ϕ produces $u_i = H_\phi(z_i)$. A residual head R_ϕ returns the candidate $d_i = R_\phi(u_i)$, while a locator head A_ϕ returns a logit $a_i = A_\phi(u_i)$ for a large base error.

Let $e_i = p_i - y_i$ be the target residual. Let $\tau \in (0,1)$ be a fixed quantile level, and let $Q_\tau(|e|)$ denote the τ quantile of absolute residuals on the current training surface. We define the locator target as $r_i = \mathbb{I}\{|e_i| \ge Q_\tau(|e|)\}$, where $\mathbb{I}$ is the indicator function. Let $\mathcal{I}_{\mathrm{Q}}$ be the set of QLEAR training points. We use constants $\alpha \ge 0$, $\lambda_{\mathrm{Q}} \ge 0$, $\beta > 0$, and $\kappa > 0$. The smooth absolute loss is

$$\rho_\kappa(t) = \begin{cases} t^2/(2\kappa), & |t| \le \kappa, \\ |t| - \kappa/2, & |t| > \kappa, \end{cases} \tag{4}$$

and the weighted binary cross entropy is

$$\mathrm{BCE}_\beta(a, r) = -\beta r \log \sigma(a) - (1 - r) \log(1 - \sigma(a)), \qquad \sigma(a) = (1 + \exp(-a))^{-1}. \tag{5}$$

Here, α controls residual emphasis, λ_{Q} weights the locator objective, β is the positive class weight, and κ is the transition point of the smooth absolute loss. We train QLEAR with

$$\mathcal{L}_{\mathrm{Q}}(\phi) = \frac{1}{|\mathcal{I}_{\mathrm{Q}}|} \sum_{i \in \mathcal{I}_{\mathrm{Q}}} \left[(1 + \alpha r_i) \rho_\kappa(d_i - e_i) + \lambda_{\mathrm{Q}} \, \mathrm{BCE}_\beta(a_i, r_i) \right], \tag{6}$$

Freezing f_θ while minimizing Eq. 6 makes $d = (d_i)_{i=1}^N$ a fixed candidate action for the subsequent gain learning stage. We write $a = (a_i)_{i=1}^N$ for the vector of locator logits.

3.3 Partial Symmetry as an Orbit Projection

Let $\mathcal{P}$ be a set of disjoint hard-exact mirror pairs $\{i, j\}$ identified from reflected coordinates and normals. Let $\mathcal{U}$ contain all soft-matched and unmatched indices. The correction units are $\mathcal{C} = \mathcal{P} \cup \{\{i\} : i \in \mathcal{U}\}$. This construction treats only hard-exact mirror correspondences as two-point orbits and all other points as singleton orbits.

For any $v \in \mathbb{R}^N$, define the partial symmetry projection $\Pi : \mathbb{R}^N \to \mathbb{R}^N$ by

$$(\Pi v)_i = \begin{cases} \frac{1}{2}(v_i + v_j), & \text{if } \{i, j\} \in \mathcal{P}, \\ v_i, & \text{if } i \in \mathcal{U}. \end{cases} \tag{7}$$

The same paired value is assigned to both members of every pair. Hence $\Pi^2 = \Pi$. We use $b = \Pi y$, $\delta = \Pi d$, and $\bar{a} = \Pi a$ for the base pressure, candidate residual, and locator score used by the selector.

Assumption 1 (Validity of eligible pairs)**.** *For every eligible pair $\{i, j\} \in \mathcal{P}$, the target pressures satisfy $p_i = p_j$. Equivalently, $\Pi p = p$.*

Lemma 1 (Safety of a valid symmetry projection)**.** *Under Assumption 1, $\ell_r(\Pi v, p) \le \ell_r(v, p)$ for every $v \in \mathbb{R}^N$ and $r \in \{1, 2\}$.*

Lemma 1 states that pair averaging cannot increase MAE or MSE when the assumed pressure symmetry is exact. The condition $\Pi p = p$ motivates singleton treatment outside hard-exact mirror orbits. The proof is given in Appendix B.1.

Proposition 2 (Symmetry preservation under paired correction)**.** *For any $s \in \mathcal{S}_K$, the field q_s in Eq. 2 satisfies $\Pi q_s = q_s$ whenever $\Pi b = b$, $\Pi\delta = \delta$, and each selection variable s_C is shared by all points in C.*

Proposition 2 guarantees that sparse selection cannot break an eligible mirror pair. The construction applies the projection only to hard-exact pairs determined from inference-time geometry; all remaining points are handled independently. Appendix B.2 provides the proof.

3.4 Counterfactual Gain Supervision

The pressure label makes it possible to compare two actions during training: retaining the raw backbone prediction y_i and applying the orbit-synchronized candidate $\delta_i = (\Pi d)_i$. We define the realized point gain

$$g_i = \underbrace{|p_i - y_i|}_{\text{error without correction}} - \underbrace{|p_i - (y_i + \delta_i)|}_{\text{error with correction}}, \qquad i \in \{1, \dots, N\}. \tag{8}$$

A positive g_i means that the candidate reduces absolute error, while a negative g_i means that the candidate is harmful. The gain depends jointly on the current error and the proposed action. Candidate magnitude or base error alone cannot determine its sign.

The Gain Gate receives $\gamma_i = [z_i, \delta_i, |\delta_i|, \bar{a}_i]$, where brackets denote feature concatenation. Let G_ψ be a scalar regressor with parameters ψ, and let $\mathcal{I}_{\mathrm{G}}$ be the set of Gain Gate training points. Let μ_γ and σ_γ denote the feature-wise training-set mean and standard deviation. The gate objective and prediction are

$$\mathcal{L}_{\mathrm{G}}(\psi) = \frac{1}{|\mathcal{I}_{\mathrm{G}}|} \sum_{i \in \mathcal{I}_{\mathrm{G}}} \rho_\kappa(G_\psi((\gamma_i - \mu_\gamma)/\sigma_\gamma) - g_i), \qquad \widehat{g}_i = G_\psi((\gamma_i - \mu_\gamma)/\sigma_\gamma), \tag{9}$$

Pressure labels appear in g_i only during training. At inference, G_ψ uses geometry and model outputs. For each correction unit $C \in \mathcal{C}$, define its true and estimated utilities as $g_C = \sum_{i \in C} g_i$ and $\widehat{g}_C = \sum_{i \in C} \widehat{g}_i$, respectively.

3.5 Gain Guided Selection under a Fixed Budget

The selector allocates the point budget by solving the estimated utility problem

$$\widehat{s} \in \arg\max_{s \in \mathcal{S}_K} \sum_{C \in \mathcal{C}} s_C \widehat{g}_C, \qquad \mathcal{S}_K = \left\{ s : \sum_{C \in \mathcal{C}} c_C s_C \le K \right\}. \tag{10}$$

Units with nonpositive estimated utility need not be selected because the budget is an upper bound. A singleton costs one point and a mirror pair costs two points. The requested budget may therefore exceed the realized correction fraction when too few units have positive estimated gain.

The implementation uses Eq. 10 as a ranking policy rather than an exact mixed-cost solver. QLEAR first selects a fixed 1% locator-ranked base set. The Gain Gate then greedily adds positive-gain singleton or paired units, excluding the protected QLEAR set and respecting its coverage cap.

Proposition 3 (Exact utility decomposition)**.** *For fixed raw base y, orbit-synchronized candidate δ, and partition $\mathcal{C}$, define $\widetilde{q}_s = y + \sum_{C \in \mathcal{C}} s_C(\mathbf{1}_C \odot \delta)$. Every $s \in \mathcal{S}_K$ satisfies*

$$\ell_1(y, p) - \ell_1(\widetilde{q}_s, p) = \frac{1}{N} \sum_{C \in \mathcal{C}} s_C g_C. \tag{11}$$

Consequently, maximizing the true unit gains over $\mathcal{S}_K$ is equivalent to minimizing the pre-projection, pre-refinement MAE over the same feasible set.

Proposition 3 connects the training target in Eq. 8 to the pre-projection selection objective. The deployed route subsequently projects the base field and selected output on hard-exact pairs. The result holds because candidates are fixed before selection and the units form a partition. Appendix B.3 gives the proof.

Theorem 4 (Selection regret under gain estimation error). *Let $s^* \in \arg\max_{s\in\mathcal{S}_K} \sum_{C\in\mathcal{C}} s_C g_C$ be an oracle selection. Suppose $|\widehat{g}_C - g_C| \le \varepsilon_C$ for every $C \in \mathcal{C}$, where $\varepsilon_C \ge 0$. Define $c_{\min} = \min_{C\in\mathcal{C}} c_C$ and $\varepsilon_{\max} = \max_{C\in\mathcal{C}} \varepsilon_C$. Let $\eta \ge 0$ be an optimization tolerance, and suppose the implemented selection $\widehat{s}$ satisfies*

$$\max_{s\in\mathcal{S}_K} \sum_{C\in\mathcal{C}} s_C \widehat{g}_C - \sum_{C\in\mathcal{C}} \widehat{s}_C \widehat{g}_C \le \eta. \tag{12}$$

Then

$$\sum_{C\in\mathcal{C}} s^*_C g_C - \sum_{C\in\mathcal{C}} \widehat{s}_C g_C \le \sum_{C\in\mathcal{C}} (s^*_C + \widehat{s}_C)\varepsilon_C + \eta \le \frac{2K}{c_{\min}}\varepsilon_{\max} + \eta, \tag{13}$$

Theorem 4 separates two failure sources: gain prediction error and approximate budget optimization. Dividing Eq. 13 by N gives the corresponding bound on missed MAE reduction. The proof appears in Appendix B.4.

3.6 Median Refinement and Staged Optimization

Let $\mathcal{M}_\omega$ denote the final refinement map with parameters ω. In the implementation it is a bounded MLP applied to recomputed 13-dimensional features of the selected field: $r_i^{\mathrm{ref}} = 0.2\tanh(\mathcal{M}_\omega(z_i))$. Its correction is hard-projected on eligible pairs and is set to zero on points already protected by QLEAR or the Gain Gate before being added to the selected field. Thus, the point budget in Eq. 2 applies to the QLEAR and Gain Gate residual-selection stages. Median Refinement preserves their selected points, enforces equality on hard-exact pairs, and may refine the remaining unprotected surface points.

Training proceeds in stages. We first optimize the backbone parameters θ. We then freeze θ and fit the QLEAR parameters ϕ using Eq. 6. After freezing both θ and ϕ, we compute counterfactual targets from Eq. 8 and optimize the Gain Gate parameters ψ with Eq. 9. Finally, we fit ω for Median Refinement. Validation data select checkpoints without parameter updates. At inference, the method uses only X, the frozen model outputs, and the fixed pair construction; the pressure target p is unavailable and unnecessary.

4 Experiments

4.1 Experimental Setup

We evaluate on 5,817 training, 1,148 validation, and 1,154 test vehicles from DrivAerNet++Surf, with 32,768 sampled surface points per vehicle. All routes share the data split, pressure normalization, and evaluation protocol. Starting from the same AdaField-55 checkpoint, Backbone-70 continues backbone training for 15 epochs, whereas CG-QLEAR freezes the backbone and trains its correction modules. The logged post-checkpoint optimization and validation times are 4.9643 h and 4.7436 h, respectively; both exclude one-time cache and feature-statistics setup. We report MSE, MAE, MaxAE, RelL2, and RelL1 using pooled test points, and provide per-vehicle averages where relevant. Appendix A gives the full protocol, metric definitions, and timing details.

4.2 Main Results under Comparable Logged Training Time

Table 1 is the primary result: both rows use the same held-out 1,154 vehicles and the same 32,768-point evaluation. Although Backbone-70 has slightly longer logged post-checkpoint optimization and validation time, CG-QLEAR improves all five pooled metrics. The largest reduction is 18.48% in MSE, followed by 9.71% in RelL2. MAE and RelL1 improve by 3.77%, while the dataset-wide Global MaxAE decreases by 0.75%. These results establish the advantage of structured residual correction over continued AdaField optimization under similar

Table 1: Primary pooled test results on 1,154 held-out vehicles. Both routes start from the same AdaField-55 checkpoint; Backbone-70 uses 4.7% more logged post-checkpoint optimization and validation time.

Method	MSE ↓	MAE ↓	Global MaxAE ↓	RelL2 (%) ↓	RelL1 (%) ↓
Backbone-70	1637.0570	19.3388	12121.7654	26.8246	16.4841
CG-QLEAR	**1334.5349**	**18.6102**	**12031.1439**	**24.2195**	**15.8630**
Relative reduction	18.48%	3.77%	0.75%	9.71%	3.77%

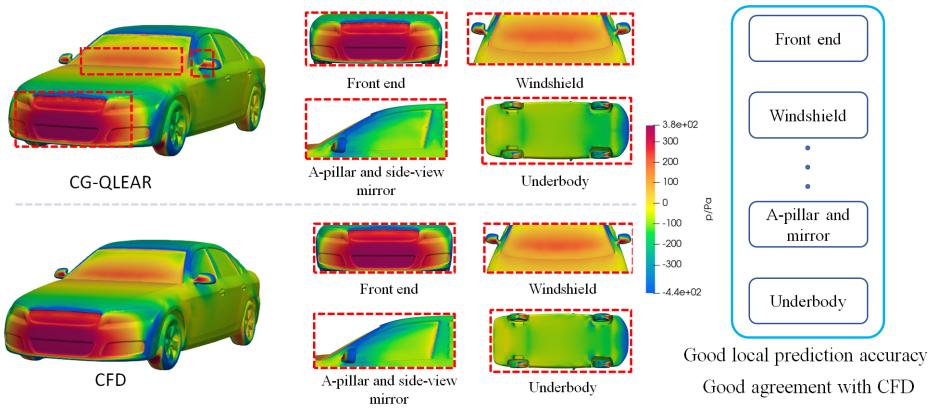


Figure 3: Qualitative pressure-field fidelity on a representative held-out vehicle. CG-QLEAR prediction (top) and CFD reference (bottom) use the same pressure color scale. Global views and enlarged front-end, windshield, A-pillar/side-mirror, and underbody regions show close agreement in both overall distribution and local pressure structures.

logged post-checkpoint optimization budgets. To complement the pooled aggregation, Table 2 reports unweighted vehicle-level averages. This view gives every vehicle equal weight and reveals whether the improvement extends to a typical vehicle rather than being dominated by a small subset of the test collection. Table 2 shows that Macro MSE, Macro MAE, Macro MaxAE, Macro RelL2, and Macro RelL1 fall by 18.48%, 3.77%, 12.52%, 9.92%, and 3.80%, respectively. The modest 0.75% change in Global MaxAE therefore coexists with a 12.52% reduction in the average vehicle's worst point. Appendix Figure 6 visualizes the macro comparison.

Table 2: Macro test results, i.e., the unweighted average of 1,154 vehicle-level metrics. Macro MaxAE averages the per-vehicle maximum error.

Method	Macro MSE ↓	Macro MAE ↓	Macro MaxAE ↓	Macro RelL2 (%) ↓	Macro RelL1 (%) ↓
Backbone-70	1637.0570	19.3388	1014.1419	26.6781	16.5525
CG-QLEAR	**1334.5349**	**18.6102**	**887.1953**	**24.0326**	**15.9239**
Relative reduction	18.48%	3.77%	12.52%	9.92%	3.80%

4.3 Qualitative Pressure-Field Fidelity

Figure 3 qualitatively compares the CG-QLEAR pressure field with the CFD reference on a representative held-out vehicle. In addition to the global surface pattern, the front end, windshield, A-pillar/side-mirror region, and underbody exhibit consistent spatial structures and pressure ranges. Together with the aggregate test metrics, this visualization demonstrates that CG-QLEAR preserves physically coherent global pressure fields while resolving local pressure structures at aerodynamically important regions.

4.4 Ablations of Gain Supervision and Refinement Stages

The staged route is evaluated from the same frozen Backbone-55 base. Table 3 shows that QLEAR provides the dominant first correction, while the Gain Gate and Median Refinement provide further consistent reductions in the norm-based metrics. Figure 4 plots the cumulative reductions relative to Backbone-55 and makes the progressive contribution of each stage explicit. The global extreme is already reduced after QLEAR and changes little afterward; Macro MaxAE gives a more representative view of the per-vehicle local-error improvement. The stage-wise analysis transparently attributes the final result to candidate residual learning, gain-guided selection, and median refinement.

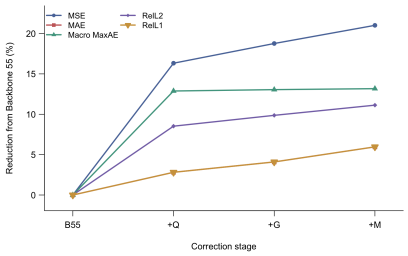


Figure 4: Cumulative error reduction relative to Backbone-55 (B55) after QLEAR (Q), Gain Gate (G), and Median Refinement (M).

Table 3: Stage ablation of the correction route from the shared Backbone-55 checkpoint.

Stage	MSE	MAE	Global MaxAE	Macro MaxAE	RelL2 (%)	RelL1 (%)
Backbone-55	1689.6215	19.7907	12138.7119	1021.7854	27.2518	16.8692
+ QLEAR 20 epochs	1413.8252	19.2340	12031.1442	890.1648	24.9286	16.3948
+ Gain Gate 10 epochs	1372.5816	18.9801	12031.1442	888.4356	24.5623	16.1784
+ Median Refinement 15 epochs	**1334.5349**	**18.6102**	**12031.1439**	**887.1953**	**24.2195**	**15.8630**

Table 4: Measured resource records for the AdaField-backbone experiment. Training is on two PPU-ZW810E devices. Logged time includes optimization and validation but excludes one-time cache and feature-statistics setup. CG-QLEAR end-to-end inference excludes disk I/O and includes all five pipeline stages; the Backbone-70 record follows the stored cached-output evaluation protocol.

Method	Total params	Added params	Logged total time	Logged time after epoch 55	Inference (s/car)	Peak memory
Backbone-70	19,957,189	0	23.3455 h	4.9643 h	0.2553	396.31 MiB
CG-QLEAR	20,014,025	+56,836 (+0.285%)	**23.1247 h**	**4.7436 h**	0.5950	388.41 MiB

4.5 Partial Symmetry and Asymmetric Regions

The preprocessing partitions points into exact mirror orbits and independent points. CG-QLEAR shares correction decisions and projected values inside exact orbits; geometrically soft matches and unmatched regions, including genuinely asymmetric structures, retain independent corrections. Figure 1 visualizes this point-wise policy and its treatment of eligible mirror regions. The resulting policy aligns the symmetry prior with regions having reliable geometric correspondence.

4.6 Correction Budgets and Computational Efficiency

We separately evaluate the inference-time selection budget with all learned weights frozen. QLEAR's base budget remains fixed at approximately 1% of surface points; we sweep the additional Gain Gate budget over $\{0, 1, 5, 10, 25, 50\}\%$ on the same 1,154 test vehicles. Table 4 records the parameter, logged training-time, inference-time, and memory profile, while Table 5 and Figure 5 report the accuracy response to available selection coverage. Since exact mirror points are selected jointly, the horizontal axis shows the measured fraction of points selected by QLEAR and the Gain Gate, rather than the requested Gate fraction. Median Refinement may also update unselected points. Table 4 shows that the correction modules add 56,836 parameters (0.285% of the Backbone-70 count), while using 4.7436 h of logged post-checkpoint optimization and validation time versus 4.9643 h for continued backbone training. The Backbone-70 parameter count includes the boundary kernel used in the pressure backbone; relative to the bare AdaField backbone alone, the same added modules correspond to 0.396%. End-to-end CG-QLEAR inference takes 0.5950 s per vehicle under the stated protocol, with a peak of 388.41 MiB per rank. The table reports each route under its recorded evaluation protocol and documents the full resource profile of the proposed pipeline.

Table 5: Post-hoc correction-budget sensitivity with fixed CG-QLEAR weights. Selected-point fractions count QLEAR and Gain Gate updates before Median Refinement. The first row includes the fixed QLEAR base budget; exact mirror points are selected jointly.

Requested Gate (%)	Selected points (%)	MSE ↓	MAE ↓	RelL2 (%) ↓
0	1.00	1348.6060	18.6662	24.3469
1	2.00	1340.9077	18.6299	24.2773
5	6.00	1334.3798	18.5712	24.2181
10	11.00	1333.5800	**18.5632**	24.2109
25	26.00	**1333.3468**	18.5732	**24.2087**
50	41.01	1334.5349	18.6102	24.2195

The budget effect is stable but not monotonic at arbitrarily large coverage. Increasing the QLEAR and Gain Gate selected-point fraction from 1.00% to 6.00% reduces pooled MSE from 1348.6060 to 1334.3798 and MAE from 18.6662 to 18.5712. MSE and RelL2 attain their lowest observed values near 26% selected points, while MAE attains its lowest value near 11%. At a requested Gate budget of 50%, the QLEAR and Gain Gate stages select 41.01% of points; the Gain Gate adds only units with positive predicted gain. MSE and MAE do not improve further. Thus, the curve supports a selective correction interpretation: increasing the available budget is useful at low coverage and eventually reaches diminishing returns.

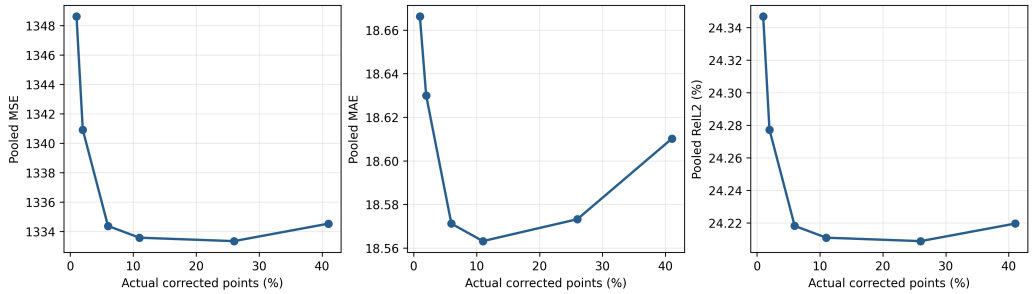


Figure 5: Post-hoc selection-budget sensitivity on the held-out test set under frozen CG-QLEAR weights. The horizontal axis, labeled "Actual corrected points" in the plot, counts points selected by QLEAR and the Gain Gate before Median Refinement; exact mirror points are selected jointly. Error falls from 1% to approximately 6–11% selected points and then saturates.

Table 6: Cross-backbone test results. Both routes start from Transolver-55. The measured continuation time is 4.877 h; the correction route targets 4.864 h of additional training.

Method	MSE ↓	MAE ↓	Global MaxAE ↓	RelL2 (%) ↓	RelL1 (%) ↓
Transolver-154	1104.9460	15.0222	12261.4578	22.0379	12.8047
Transolver + CG-QLEAR	**787.7941**	**14.0573**	**12198.6561**	**18.6083**	**11.9822**
Relative reduction	28.70%	6.42%	0.51%	15.56%	6.42%

4.7 Transfer across Prediction Backbones

To test whether the correction interface transfers beyond AdaField, we repeat the protocol with Transolver. Both routes start from the same Transolver-55 checkpoint. The continuation control trains the backbone for 99 further epochs, reaching Transolver-154. The CG-QLEAR route instead trains QLEAR, Gain Gate, and Median Refinement for 20, 10, and 15 epochs. The continuation route took 4.877 h; the correction route was targeted at 4.864 h. The available timing record does not establish the correction route's realized added training time.

All reported Transolver results use the same locked 1,154-vehicle test set and 32,768 points per vehicle. Table 6 gives the pooled primary results. The corresponding macro metrics also improve, including a 20.81% reduction in Macro MaxAE; their full values appear in Appendix Table 8. Paired vehicle-level comparisons favor CG-QLEAR on 766/1,154 vehicles for MSE (66.38%), 779/1,154 for MAE (67.50%), and 835/1,154 for MaxAE (72.36%). A 2,000-resample paired vehicle bootstrap gives positive 95% confidence intervals for the mean control-minus-CG difference: [296.34, 336.92] for MSE, [0.880, 1.048] for MAE, and [174.11, 201.85] for per-vehicle MaxAE.

The correction interface is lightweight relative to the Transolver backbone: it adds 56,836 parameters, from 3,860,033 to 3,916,869 (+1.47%). Under the same two-PPU test protocol, mean measured inference time is 0.3291 s/vehicle for Transolver+CG-QLEAR and 0.0667 s/vehicle for Transolver-154. The all-metric test improvement and paired vehicle-level evidence support cross-backbone applicability of the correction interface; a measured correction-route training time is needed to assess compute matching for this experiment.

5 Conclusion and Future Work

CG-QLEAR frames local pressure refinement as a decision about the benefit of a proposed update. Counterfactual gain supervision, paired correction on eligible regions, and a point budget give the residual selection decision an explicit target and structure. The aggregate comparison favors the complete correction pipeline over 15 additional backbone epochs across MSE, MAE, macro MaxAE, RelL2, and RelL1. The Transolver experiment further shows positive paired bootstrap intervals for MSE, MAE, and per-vehicle MaxAE. Future work can extend the candidate generator, adapt the pairing policy to broader flow conditions, and study transfer across additional backbones. It can also reduce candidate-generation cost while retaining budgeted application of the selected updates.

6 AI USE STATEMENT

Generative AI tools were used to assist with manuscript drafting, mathematical exposition, reference organization, and language editing. The authors are responsible for reviewing and verifying all AI assisted content and take full responsibility for the final manuscript.

7 ETHICS STATEMENT

The study concerns vehicle surface pressure prediction using computational fluid dynamics data and does not involve human participants or personal information. Learned pressure predictions can contain local errors and should be validated against appropriate simulations or measurements before informing engineering decisions. The symmetry constraints apply only where the geometry and flow conditions support the assumed pressure symmetry. Performance on the evaluated dataset does not establish reliability across all vehicle designs or operating conditions.

8 REPRODUCIBILITY STATEMENT

We describe CG-QLEAR in Section 3 and the experimental protocol in Section 4. Appendix A documents the data manifests, model configurations, training schedules, normalization statistics, and compute accounting used for reproduction.

REFERENCES

Benedikt Alkin, Andreas Fürst, Simon Schmid, Lukas Gruber, Markus Holzleitner, and Johannes Brandstetter. Universal physics transformers. In *Advances in Neural Information Processing Systems*, volume 37, 2024. doi: 10.52202/079017-0793. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/2cd36d327f33d47b372d4711edd08de0-Abstract-Conference.html.

Sotiris Anagnostidis, Gregor Bachmann, Imanol Schlag, and Thomas Hofmann. Navigating scaling laws: Compute optimality in adaptive model training. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 1511–1530. PMLR, 2024. URL https://proceedings.mlr.press/v235/anagnostidis24a.html.

Jules Berman and Benjamin Peherstorfer. Randomized sparse neural galerkin schemes for solving evolution equations with deep networks. In *Advances in Neural Information Processing Systems*, volume 36, 2023. doi: 10.52202/075280-0179. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/0cb310ed8121549488fea8e8c2056096-Abstract-Conference.html.

Daniel Bolya, Cheng-Yang Fu, Xiaoliang Dai, Peizhao Zhang, Christoph Feichtenhofer, and Judy Hoffman. Token merging: Your ViT but faster. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=JroZRaRw7Eu.

Mengzhao Chen, Wenqi Shao, Peng Xu, Mingbao Lin, Kaipeng Zhang, Fei Chao, Rongrong Ji, Yu Qiao, and Ping Luo. DiffRate: Differentiable compression rate for efficient vision transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 17164–17174, 2023. URL https://openaccess.thecvf.com/content/ICCV2023/html/Chen_DiffRate__Differentiable_Compression_Rate_for_Efficient_Vision_Transformers_ICCV_2023_paper.html.

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. Curran Associates, Inc., 2024. doi: 10.52202/079017-0016. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/013cf29a9e68e4411d0593040a8a1eb3-Abstract-Datasets_and_Benchmarks_Track.html.

Jerome H. Friedman. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5):1189–1232, 2001. doi: 10.1214/aos/1013203451. URL https://doi.org/10.1214/aos/1013203451.

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. URL https://proceedings.mlr.press/v202/hao23c.html.

Maximilian Herde, Bogdan Raonic, 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, 2024. doi: 10.52202/079017-2311. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/84e1b1ec17bb11c57234e96433022a9a-Abstract-Conference.html.

Sékou-Oumar Kaba, Arnab Kumar Mondal, Yan Zhang, Yoshua Bengio, and Siamak Ravanbakhsh. Equivariance with learned canonicalization functions. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 15546–15566. PMLR, 2023. URL https://proceedings.mlr.press/v202/kaba23a.html.

Zongyi Li, Nikola Kovachki, Chris Choy, Boyi Li, Jean Kossaifi, Shourya Otta, Mohammad Amin Nabian, Maximilian Stadler, Christian Hundt, Kamyar Azizzadenesheli, and Animashree Anandkumar. Geometry-informed neural operator for large-scale 3D PDEs. In *Advances in Neural Information Processing Systems*, volume 36, pp. 35836–35854. Curran Associates, Inc., 2023. doi: 10.52202/075280-1556. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/70518ea42831f02afc3a2828993935ad-Abstract-Conference.html.

Dingkang Liang, Xin Zhou, Wei Xu, Xingkui Zhu, Zhikang Zou, Xiaoqing Ye, Xiao Tan, and Xiang Bai. Pointmamba: A simple state space model for point cloud analysis. In *Advances in Neural Information Processing Systems*, volume 37, 2024. doi: 10.52202/079017-1026. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/395371f778ebd4854b88521100af30ad-Abstract-Conference.html.

Yuchao Lin, Jacob Helwig, Shurui Gui, and Shuiwang Ji. Equivariance via minimal frame averaging for more symmetries and efficiency. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 30042–30079. PMLR, 2024. URL https://proceedings.mlr.press/v235/lin24i.html.

Phillip Lippe, Bas Veeling, Paris Perdikaris, Richard Turner, and Johannes Brandstetter. PDE-Refiner: Achieving accurate long rollouts with neural PDE solvers. In *Advances in Neural Information Processing Systems*, volume 36, 2023. doi: 10.52202/075280-2946. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/d529b943af3dba734f8a7d49efcb6d09-Abstract-Conference.html.

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. URL https://proceedings.mlr.press/v267/luo25o.html.

Anqi Mao, Mehryar Mohri, and Yutao Zhong. Regression with multi-expert deferral. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 34738–34759. PMLR, 2024. URL https://proceedings.mlr.press/v235/mao24d.html.

Tanya Marwah, Ashwini Pokle, J. Zico Kolter, Zachary Lipton, Jianfeng Lu, and Andrej Risteski. Deep equilibrium based neural operators for steady-state PDEs. In *Advances in Neural Information Processing Systems*, volume 36, 2023. doi: 10.52202/075280-0691. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/32cc61322f1e2f56f989d29ccc7cfbb7-Abstract-Conference.html.

Saro Passaro and C. Lawrence Zitnick. Reducing SO(3) convolutions to SO(2) for efficient equivariant GNNs. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 27420–27438. PMLR, 2023. URL https://proceedings.mlr.press/v202/passaro23a.html.

Bogdan Raonic, Roberto Molinaro, Tim De Ryck, Tobias Rohner, Francesca Bartolucci, Rima Alaifari, Siddhartha Mishra, and Emmanuel de Bézenac. Convolutional neural operators for robust and accurate learning of PDEs. In *Advances in Neural Information Processing Systems*, volume 36, 2023. doi: 10.52202/075280-3376. URL https://proceedings.neurips.cc/paper/2023/hash/f3c1951b34f7f55ffaecada7fde6bd5a-Abstract-Conference.html.

Louis Serrano, Lise Le Boudec, Armand Kassaï Koupaï, Thomas X. Wang, Yuan Yin, Jean-Noël Vittaut, and Patrick Gallinari. Operator learning with neural fields: Tackling PDEs on general geometries. In *Advances in Neural Information Processing Systems*, volume 36, 2023. doi: 10.52202/075280-3093. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/df54302388bbc145aacaa1a54a4a5933-Abstract-Conference.html.

Thuan Anh Trang, Nhat Khang Ngo, Daniel T. Levy, Thieu Ngoc Vo, Siamak Ravanbakhsh, and Truong Son Hy. $E(3)$-equivariant mesh neural networks. In *Proceedings of the 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pp. 748–756. PMLR, 2024. URL https://proceedings.mlr.press/v238/anh-trang24a.html.

Themistoklis Vargiemezis, Charilaos Kanatsoulis, and Catherine Gorlé. WindMiL: Equivariant graph learning for wind loading prediction. *arXiv preprint arXiv:2511.01226*, 2025. doi: 10.48550/arXiv.2511.01226. URL https://arxiv.org/abs/2511.01226.

Haixu Wu, Tengge Hu, Huakun Luo, Jianmin Wang, and Mingsheng Long. Solving high-dimensional PDEs with latent spectral models. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 37417–37438. PMLR, 2023. URL https://proceedings.mlr.press/v202/wu23f.html.

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, 2024a. URL https://proceedings.mlr.press/v235/wu24r.html.

Xiaoyang Wu, Li Jiang, Peng-Shuai Wang, Zhijian Liu, Xihui Liu, Yu Qiao, Wanli Ouyang, Tong He, and Hengshuang Zhao. Point transformer V3: Simpler, faster, stronger. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4840–4851, 2024b. URL https://openaccess.thecvf.com/content/CVPR2024/html/Wu_Point_Transformer_V3_Simpler_Faster_Stronger_CVPR_2024_paper.html.

Minkai Xu, Jiaqi Han, Aaron Lou, Jean Kossaifi, Arvind Ramanathan, Kamyar Azizzadenesheli, Jure Leskovec, Stefano Ermon, and Anima Anandkumar. Equivariant graph neural operator for modeling 3D dynamics. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 55015–55032. PMLR, 2024. URL https://proceedings.mlr.press/v235/xu24j.html.

Fuzhao Xue, Valerii Likhosherstov, Anurag Arnab, Neil Houlsby, Mostafa Dehghani, and Yang You. Adaptive computation with elastic input sequence. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 38971–38988. PMLR, 2023. URL https://proceedings.mlr.press/v202/xue23e.html.

Jianke Yang, Nima Dehmamy, Robin Walters, and Rose Yu. Latent space symmetry discovery. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235

of *Proceedings of Machine Learning Research*, pp. 56047–56070. PMLR, 2024. URL https://proceedings.mlr.press/v235/yang24g.html.

Mert Yuksekgonul, Linjun Zhang, James Y. Zou, and Carlos Guestrin. Beyond confidence: Reliable models should also consider atypicality. In *Advances in Neural Information Processing Systems*, volume 36, 2023. doi: 10.52202/075280-1669. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/7900318ffaf5e9bc60250f134c6cc3c7-Abstract-Conference.html.

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. doi: 10.48550/arXiv.2602.20609. URL https://arxiv.org/abs/2602.20609.

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. doi: 10.1609/aaai.v40i2.37145. URL https://ojs.aaai.org/index.php/AAAI/article/view/37145.

A Data Audit and Reproducibility

A.1 Split Manifests, Exclusions, and Sampling

Table 7: Locked dataset partition used throughout the experiments. Counts refer to the usable aligned DrivAerNet++Surf collection supplied for this study.

Partition	Vehicles	Role
Training	5,817	Parameter fitting
Validation	1,148	Checkpoint and hyperparameter selection
Test	1,154	Locked final evaluation
Total	8,119	32,768 surface points per vehicle

The reported author-provided DrivAerNet++Surf release contains 8,119 usable vehicles, split deterministically into 5,817 training, 1,148 validation, and 1,154 locked-test vehicles. The released manifests, exclusion records, and sampling rule are retained with the experiment package. Repeated evaluation of a vehicle uses the same 32,768 sampled surface points across methods. Pair construction depends only on permitted inference inputs; pressure labels do not determine pair eligibility at inference.

A.2 Architectures, Hyperparameters, and Training Schedules

The AdaField continuation route trains for 15 epochs after the shared epoch 55 checkpoint. The correction route trains QLEAR for 20 epochs, the Gain Gate for 10 epochs, and Median Refinement for 15 epochs with the backbone frozen. The Transolver continuation route trains for 99 epochs after its shared epoch 55 checkpoint, while the Transolver correction route uses the same 20, 10, and 15 epoch stage schedule. AdaField routes are compared by logged additional optimization and validation time rather than by their total epoch counts. For Transolver, 4.864 h is the correction-route target, not a reported measurement.

QLEAR uses the 13-dimensional feature vector described in Section 3.2; the Gain Gate augments it with the proposed residual, its magnitude, and locator score, yielding 16 features. Median Refinement uses a 13-to-256-to-128-to-1 MLP with a bounded $0.2\tanh(\cdot)$ output. The QLEAR locator uses the per-vehicle 0.99 error quantile and a fixed 1% base correction budget. Feature normalization statistics are fitted on training data only; counterfactual gain targets remain in their native normalized-pressure units.

A.3 Evaluation Metrics and Aggregation

For vehicle v, let p_v be the pressure target, $\hat{p}_v$ the prediction, $e_v = \hat{p}_v - p_v$ the error, and $N = 32{,}768$ the number of sampled surface points. Let $\epsilon > 0$ prevent division by zero in the relative metrics. The vehicle-level metrics are

$$\mathrm{MSE}_v = N^{-1}\|e_v\|_2^2, \qquad \mathrm{MAE}_v = N^{-1}\|e_v\|_1, \qquad \mathrm{MaxAE}_v = \|e_v\|_\infty, \tag{14}$$

$$\mathrm{RelL2}_v = 100\frac{\|e_v\|_2}{\max(\|p_v\|_2, \epsilon)}, \quad \mathrm{RelL1}_v = 100\frac{\|e_v\|_1}{\max(\|p_v\|_1, \epsilon)}. \tag{15}$$

Pooled metrics concatenate the 1,154 test vehicles before evaluation, whereas macro metrics average the corresponding vehicle-level values. Global MaxAE is the largest absolute error over all test points; Macro MaxAE is the mean of the per-vehicle maxima. Per-vehicle records are retained for both the AdaField and Transolver comparisons. The paired Transolver bootstrap uses 2,000 vehicle-level resamples, preserving the vehicle pairing between methods.

A.4 Compute Accounting and Inference Timing

Timing uses the same two PPU-ZW810E devices as the main comparison. The end-to-end CG-QLEAR profiling pass includes geometry, backbone, QLEAR, Gain Gate, Median Refinement, and pressure output. It excludes disk I/O and uses five warm-up vehicles per rank. Peak memory is the

largest per-rank allocation. Backbone-70 latency follows the stored cached-prediction evaluation path, whereas CG-QLEAR latency measures the complete inference pipeline; the table retains these as protocol-specific resource records. Logged training time separates the shared backbone stage from all additional stages and excludes one-time cache and feature-statistics setup. The table reports exact total and added parameter counts, while elapsed wall-clock time is reported for the two-PPU runs.

B Additional Method Details

B.1 Proof of Lemma 1

Proof. Because $\mathcal{P}$ contains disjoint pairs and $\mathcal{U}$ contains the remaining singleton indices, it is sufficient to compare the loss on each unit $C \in \mathcal{C}$. A singleton is unchanged by Π, so its contribution is equal before and after projection.

Consider a pair $C = \{i, j\} \in \mathcal{P}$. The assumption $\Pi p = p$ implies $p_i = p_j = t$ for some $t \in \mathbb{R}$. Let $\bar{v} = (v_i + v_j)/2$. For $r \in \{1, 2\}$, the function $x \mapsto |x - t|^r$ is convex. Jensen's inequality therefore gives

$$|\bar{v} - t|^r \leq \frac{1}{2}|v_i - t|^r + \frac{1}{2}|v_j - t|^r. \tag{16}$$

Both projected entries equal $\bar{v}$, so multiplication by two yields

$$|(\Pi v)_i - p_i|^r + |(\Pi v)_j - p_j|^r \leq |v_i - p_i|^r + |v_j - p_j|^r. \tag{17}$$

Summing the inequality over all pairs, adding the unchanged singleton terms, and dividing by N proves $\ell_r(\Pi v, p) \leq \ell_r(v, p)$. □

B.2 Proof of Proposition 2

Proof. Fix an eligible pair $C = \{i, j\} \in \mathcal{P}$. Since $\Pi b = b$ and $\Pi \delta = \delta$, Eq. 7 implies $b_i = b_j$ and $\delta_i = \delta_j$. The selection variable is defined per unit, so both points use the same value s_C. Equation 2 gives

$$(q_s)_i = b_i + s_C \delta_i = b_j + s_C \delta_j = (q_s)_j. \tag{18}$$

Thus, q_s is constant on every eligible pair. The projection leaves all singleton entries unchanged, which proves $\Pi q_s = q_s$. □

B.3 Proof of Proposition 3

Proof. Every index i belongs to exactly one unit $C \in \mathcal{C}$. For the pre-projection field defined in Proposition 3, $(\widetilde{q}_s)_i = y_i + s_C \delta_i$. Since s_C is binary,

$$|p_i - y_i| - |p_i - (y_i + s_C \delta_i)| = s_C \left(|p_i - y_i| - |p_i - (y_i + \delta_i)|\right) = s_C g_i. \tag{19}$$

Summing over points, grouping indices by their unique units, and applying the definition $g_C = \sum_{i \in C} g_i$ gives

$$\ell_1(y, p) - \ell_1(\widetilde{q}_s, p) = \frac{1}{N} \sum_{C \in \mathcal{C}} s_C g_C. \tag{20}$$

The term $\ell_1(y, p)$ does not depend on s. Therefore, a feasible selection maximizes the true gain sum if and only if it minimizes $\ell_1(\widetilde{q}_s, p)$ over $\mathcal{S}_K$. □

B.4 Proof of Theorem 4

Proof. Define the true and estimated utilities

$$U(s) = \sum_{C \in \mathcal{C}} s_C g_C, \qquad \widehat{U}(s) = \sum_{C \in \mathcal{C}} s_C \widehat{g}_C. \tag{21}$$

The assumed optimization tolerance means $\max_{s\in\mathcal{S}_K} \widehat{U}(s) - \widehat{U}(\widehat{s}) \le \eta$. Because $s^* \in \mathcal{S}_K$, it follows that $\widehat{U}(s^*) - \widehat{U}(\widehat{s}) \le \eta$. Add and subtract the two estimated utilities to obtain

$$\begin{aligned} U(s^*) - U(\widehat{s}) &= \big[U(s^*) - \widehat{U}(s^*)\big] + \big[\widehat{U}(s^*) - \widehat{U}(\widehat{s})\big] + \big[\widehat{U}(\widehat{s}) - U(\widehat{s})\big] \\ &\le \sum_{C\in\mathcal{C}} s^*_C \varepsilon_C + \eta + \sum_{C\in\mathcal{C}} \widehat{s}_C \varepsilon_C. \end{aligned} \tag{22}$$

This proves the first inequality in Eq. 13. Every selected unit costs at least $c_{\min}$. Hence each feasible selection contains at most $K/c_{\min}$ units. Applying $\varepsilon_C \le \varepsilon_{\max}$ to both selected sets proves the second inequality. □

The bound concerns the field q_s before Median Refinement. A later map can couple points and change the realized final utility. Under equal unit costs, exact maximization of Eq. 10 selects the largest positive estimated gains until the budget is exhausted. Mixed singleton and pair costs require a cost aware solver or a stated approximation, whose worst case contribution is represented by η.

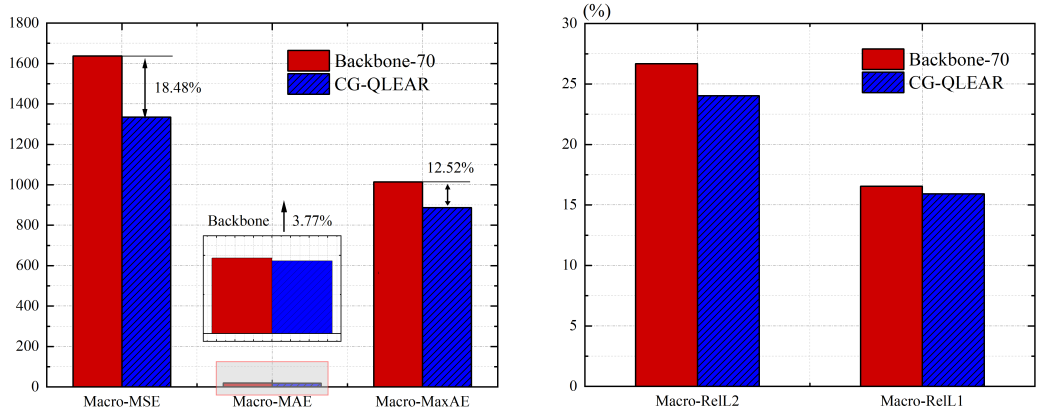
B.5 Conditions on Median Refinement

The implemented refinement map receives recomputed features of the selected field, produces a bounded residual, projects that residual on hard-exact pairs, and masks it to zero on the protected QLEAR/Gain-Gate points. It therefore preserves the earlier selected values and eligible pair equality. Because it may act on unprotected points, the correction budget describes the selected residual stages rather than the final refinement support.

C Extended Results

C.1 Macro AdaField Comparison

Figure 6 displays the vehicle-level averages reported in Table 2. Relative to Backbone-70, CG-QLEAR reduces Macro MSE by 18.48%, Macro MAE by 3.77%, and Macro MaxAE by 12.52%. Macro RelL2 and RelL1 also decrease by 9.92% and 3.80%, respectively. The larger improvement in Macro MaxAE than in Global MaxAE indicates that the average vehicle's largest local error falls even though the single largest error across the entire test set changes little.



(a) Macro MSE, MAE, and MaxAE. (b) Macro RelL2 and RelL1.

Figure 6: Macro test comparison between Backbone-70 and CG-QLEAR. Both routes start from the same AdaField-55 checkpoint; the continuation control uses 4.7% more logged post-checkpoint optimization and validation time. Lower is better.

C.2 Macro Transolver Results

Table 8 gives equal weight to each of the 1,154 test vehicles when evaluating the Transolver comparison. Adding CG-QLEAR reduces Macro MSE from 1104.9460 to 787.7941 and Macro MaxAE

from 902.3149 to 714.5313, corresponding to 28.70% and 20.81% reductions. Macro MAE, RelL2, and RelL1 also improve. These vehicle-level results agree with the pooled improvements in Table 6 and show that the gain is present under both aggregation rules.

Table 8: Macro Transolver generalization results: unweighted averages of the 1,154 vehicle-level metrics. Lower is better.

Method	Macro MSE ↓	Macro MAE ↓	Macro MaxAE ↓	Macro RelL2 (%) ↓	Macro RelL1 (%) ↓
Transolver-154	1104.9460	15.0222	902.3149	20.9710	12.9807
Transolver + CG-QLEAR	**787.7941**	**14.0573**	**714.5313**	**18.0805**	**12.1149**
Relative reduction	28.70%	6.42%	20.81%	13.78%	6.67%

C.3 Paired Bootstrap and Per Vehicle Results

The complete records provide per-vehicle metrics for the common checkpoint, continued backbone, and full correction pipeline. Paired resampling preserves the shared vehicle difficulty between methods. Confidence intervals are computed from paired records rather than reconstructed from rounded aggregate metrics.

C.4 Additional Ablations, Budget Curves, and Visualizations

The reported experiments include a stage ablation, a correction-budget curve, measured parameter and timing records, and a transfer experiment with Transolver. Together, these analyses connect the staged correction mechanism, its budgeted selection behavior, and its transfer across pressure-prediction backbones.