WHEN DOES EQUIVARIANCE HELP? CANONICAL ALIGNMENT IN NEURAL FLUID SURROGATES

Anonymous authors
Paper under double-blind review

ABSTRACT

Neural surrogates can accelerate computational fluid dynamics (CFD) simulations by orders of magnitude, but practical deployment in engineering and healthcare applications requires architectures that scale to high-resolution meshes and learn effectively from limited data. Explicit equivariance offers a principled inductive bias, yet its accuracy benefits may depend on the prediction task and the distribution of anatomical orientations. We investigate this dependence across three hemodynamic benchmarks with different degrees of natural canonical alignment. To support this study, we introduce the Anchored-Branched Geometric Algebra Transformer (AB-GATr), an $E(3)$-equivariant surrogate that efficiently predicts coupled surface and volume quantities. Across these benchmarks, AB-GATr consistently outperforms the evaluated non-equivariant models, including variants trained with rotational augmentation, while achieving accuracy competitive with $E(3)$-equivariant LaB-GATr at substantially lower training cost. In comparison, rotational augmentation provides inconsistent benefits across architectures and can reduce accuracy. A controlled experiment on ShapeNet-Car shows that strong canonical alignment can favor non-equivariant models, but their accuracy generally deteriorates as training orientations broaden and can decline sharply under broader test rotations. We further investigate these patterns using extended symmetry-breaking diagnostics and probes of the predictive information associated with canonical alignment across all benchmarks. Together, these results support explicit equivariance for the evaluated hemodynamic tasks with natural orientation variation, while showing that its accuracy benefits depend on the task and orientation distribution. Our code is publicly available at [1].

1 INTRODUCTION

In-silico modelling of physical phenomena using numerical methods has long played a central role in science and engineering. In computational fluid dynamics (CFD), applications in domains such as automotive design (Martins, 2022) and medicine (Williamson et al., 2022) are primarily carried out using high-fidelity numerical solvers that approximate the Navier-Stokes (NS) equations. In medicine, rapid and reliable personalised assessment of hemodynamics, i.e., the way in which blood flows through arteries, can inform timely treatment decisions in patients suffering from pathologies such as stenoses and aneurysms (Taylor et al., 2013). While these methods provide accurate and physically grounded solutions, their computational cost and runtime remain bottlenecks in many real-world settings.

In recent years, neural surrogates - deep learning-based approximators of numerical methods - have emerged as a promising alternative, offering orders-of-magnitude acceleration by amortising simulation cost into offline training. Once trained, such models can estimate quantities of interest, such as pressure, velocity, or wall shear stress, in seconds. As a result, neural surrogates have sparked increasing interest across CFD-driven domains, including automotive aerodynamics (Wu et al., 2024; Luo et al., 2025; Zhdanov et al., 2025; Alkin et al., 2025) or hemodynamics (Brehmer et al., 2023; Suk et al., 2024c; 2025; Rygiel et al., 2023; 2025a; Griffo et al., 2026; Lannelongue et al., 2026).

Recent progress in neural surrogate modelling has been largely shaped by two research directions. The first focuses on scalability, seeking architectures that can handle high-resolution, geometrically

[1] withheld due to review process

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.

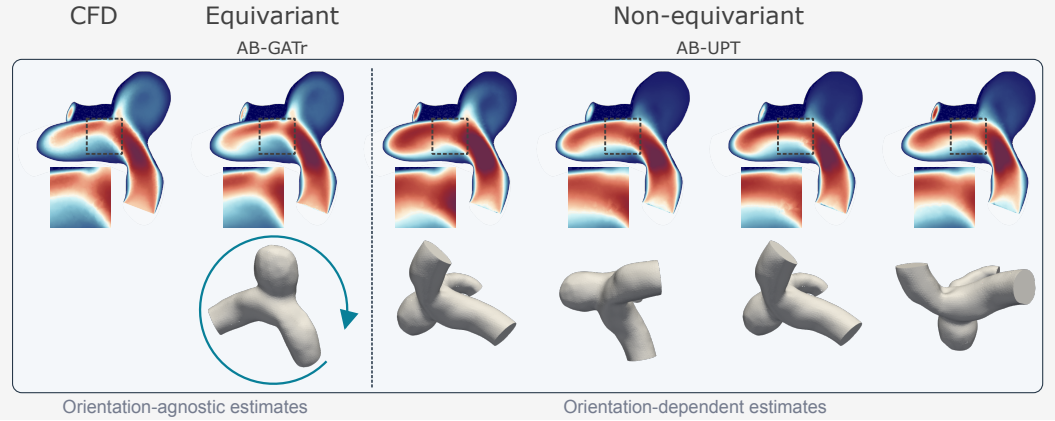

Figure 1: Vascular systems can realistically exhibit different orientation variability. Non-equivariant models treat all these orientations as different samples; hence, the estimates vary between them and may lead to degraded performance. On the other hand, equivariant models are agnostic to these orientation changes, yielding consistent and more accurate estimates.

complex surface or volumetric domains without prohibitive memory or computational cost. This has led to more computationally efficient transformer designs (Luo et al. (2025); Zhdanov et al. (2025)) and neural field formulations that decouple geometry encoding from pointwise prediction (Jaegle et al., 2021; Lu et al., 2021; Li et al., 2023; Cao et al., 2024; Alkin et al., 2024; 2025). The second direction emphasises inductive bias by incorporating physical structure - such as conservation laws (Raissi et al., 2019; Lutter et al., 2019; Suk et al., 2024a; Lannelongue et al., 2026) or symmetries (Bronstein et al., 2021) - into the model design to improve data efficiency and generalisation in domains that generally lack large training data sets. Among these, equivariant neural networks, which enforce symmetry constraints by construction, have shown strong empirical performance on a range of physics learning tasks (Brandstetter et al., 2022a;b; Brehmer et al., 2023; Suk et al., 2024b;c; Catalani et al., 2025; Kim et al., 2026; Wessels et al., 2025).

Across scientific machine learning applications, the necessity of including inductive biases in architecture design is increasingly under scrutiny, particularly as training dataset sizes increase in scale (Brehmer et al., 2024). Recent notable works in molecular design suggest that, with sufficient data, neural networks can implicitly learn physical symmetries, making explicit equivariance unnecessary (Abramson et al., 2024; Wang et al., 2024). However, in practice, this assumption does not hold uniformly across application domains and can affect model behaviour (Moskalev et al., 2023). In medical and engineering CFD problems in particular, two challenges compound. First, training samples are scarce because acquisition and simulation are costly (Alkin et al., 2025; Rygiel et al., 2025b). Second, training data can vary widely, and the distribution of geometric orientations may influence the usefulness of equivariance as an inductive bias.

We argue that the *canonical alignment* of a dataset, i.e., the extent to which geometries share similar orientations, determines the need for equivariance. For example, automotive aerodynamics benchmarks (Umetani & Bickel, 2018; Ashton et al., 2025) typically exhibit strong canonical alignment through standardised wind-tunnel configurations. In such datasets, enforcing equivariance might be counterproductive as models lose the ability to exploit canonical alignment-induced cues and therefore attain suboptimal in-distribution performance (Vadgama et al., 2025; Lawrence et al., 2026). In contrast, hemodynamic datasets naturally exhibit different degrees of alignment: Aortas retain a preferred anatomical orientation, whereas intracranial artery geometries exhibit substantially greater orientation variability. In such problems, relying solely on data to recover fundamental physical symmetries may be both inefficient and unreliable (see Fig. 1).

This work makes two contributions. First, we systematically evaluate the relation between canonical alignment and equivariance across three hemodynamic benchmarks, and assess robustness to orientation-distribution shifts. Building on Lawrence et al. (2026), we extend symmetry diagnostics with rotation-dependent response curves and integrated scores, and apply task-dependent probes to examine how canonical alignment relates to surrogate performance. Second, we introduce

Anchored-Branched Geometric Algebra Transformer (AB-GATr), an $E(3)$-equivariant neural surrogate that combines scalable anchor-attention and multi-branching (Alkin et al., 2025) with explicit $E(3)$-equivariance achieved through projective geometric algebra (Brehmer et al., 2023). AB-GATr outperforms non-equivariant models on hemodynamic tasks and achieves better or on-par performance relative to $E(3)$-equivariant LaB-GATr (Suk et al., 2024c) at substantially lower training cost.

2 Materials & Methods

We first describe the supervised CFD learning problem and distinguish solution-map equivariance from symmetry in the data distribution. We then introduce AB-GATr and the benchmarks used to investigate canonical alignment and the benefits of explicit equivariance under orientation changes.

2.1 Computational Fluid Dynamics (CFD)

CFD requires solving the incompressible Navier-Stokes (NS) equations, partial differential equations (PDEs) that govern fluid dynamics:

$$\rho\left(\frac{\partial u}{\partial t} + (u \cdot \nabla)u\right) = -\nabla p + \mu\nabla^2 u + g, \quad \nabla \cdot u = 0$$

Here, u denotes the velocity field, p the pressure field, g the external forces, and μ and ρ the fluid viscosity and density, respectively. Numerical solvers approximate the solution on a discretised domain $\Omega \subset \mathbb{R}^3$ subject to prescribed boundary conditions $\mathcal{B}$ and initial conditions. The resulting fields (velocity, pressure, etc.) are represented on the computational domain Ω and provide the targets for surrogate learning.

2.2 Equivariant neural surrogates for CFD

Neural surrogates provide a data-driven alternative to classical numerical solvers by learning to approximate the mapping from the domain and boundary conditions to the PDE solution. We view a neural surrogate as approximating the mapping $f_\theta \in \mathcal{F} : (\mathcal{X}, \mathcal{B}) \mapsto \mathcal{Y}$, where $\mathcal{X}$ denotes the input signal (e.g., Cartesian coordinates or normal vectors) defined over the spatial domain Ω, $\mathcal{B}$ represents boundary conditions, and $\mathcal{Y}$ the target solution.

Many PDEs exhibit intrinsic structure that neural surrogates can exploit through appropriate inductive biases imposed on the hypothesis class $\mathcal{F}$. These biases restrict $\mathcal{F}$ to functions consistent with known physical priors, thereby improving generalisation and data efficiency (Bronstein et al., 2021; Brehmer et al., 2024). A prominent example is equivariance and invariance to symmetry, which is particularly relevant in CFD: assuming the absence of external forces, or that such forces respect the same symmetries as the domain, the incompressible 3D NS operator is equivariant to the Euclidean group $E(3)$. Concretely, for any group action $\rho \in E(3)$, i.e., rotating, translating, or reflecting the spatial domain changes the solution in a well-defined way: $u(\rho x, t) = \rho u(x, t)$, $p(\rho x, t) = p(x, t)$. Where the velocity field transforms equivariantly as a vector field, while the pressure transforms invariantly as a scalar field. Incorporating equivariance into neural surrogate models enables them to respect the underlying physics by construction.

However, an operator's equivariance does not automatically imply equivariance of the solution map. For a neural surrogate f to be equivariant, the group action must jointly transform the input signal $\mathcal{X}$ and the boundary conditions $\mathcal{B}$: $f\big(\rho\mathcal{X}, \rho\mathcal{B}\big) = \rho f\big(\mathcal{X}, \mathcal{B}\big)$. If either the input signal or the boundary conditions violate the assumed symmetry, equivariance is broken, and the NS solution no longer transforms consistently under $E(3)$. This effect is amplified if we never explicitly supply boundary conditions to the model but require it to learn them as a function of the input signal $f(\mathcal{X}, \mathcal{B}(\mathcal{X}))$.

This distinction highlights that symmetry is a property of the entire problem formulation, not of the governing equations alone. Consequently, the effectiveness of equivariant inductive biases depends strongly on both the problem and the data's rate of canonical alignment, echoing recent findings by Lawrence et al. (2026).

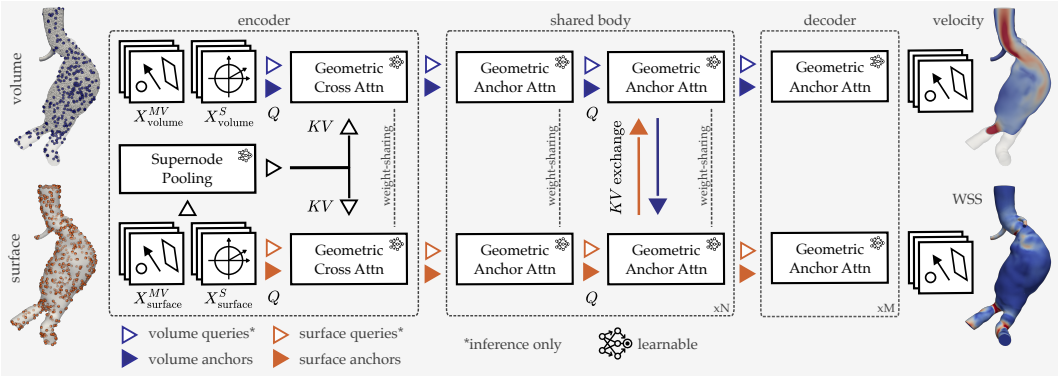

Figure 2: AB-GATr is a transformer-based $E(3)$-equivariant neural field that jointly models volumetric and surface quantities. Volume and surface nodes are embedded using geometric algebra multivectors X^{MV} alongside sine-cosine encoded scalar features X^S. Sparse surface and volume anchors are uniformly sampled and serve as keys and values for geometric anchor attention throughout the network. At inference, additional query tokens are concatenated to the query set, while keys and values remain restricted to anchors. In the shared body, branch-specific geometric anchor attention is interleaved with a KV-exchange variant, in which key and value tokens are taken from opposite branches. Branch-specific decoders are applied at the output stage, where target quantities are extracted from the resulting multivectors.

2.3 Anchored-Branched Geometric Algebra Transformer (AB-GATr)

Designing neural surrogates for real-world CFD involves simultaneously addressing two competing challenges: scalability to large, high-resolution surface and volume meshes, and a physically meaningful inductive bias. This work bridges these paradigms by proposing a scalable, equivariant architecture.

We introduce the **Anchored-Branched Geometric Algebra Transformer (AB-GATr)** (Fig. 2), an $E(3)$-equivariant neural surrogate built upon Anchored-Branched Universal Physics Transformer (AB-UPT; Alkin et al. (2025)). AB-GATr retains AB-UPT's anchor attention and multi-branching for processing large CFD meshes, while incorporating equivariance through projective geometric algebra (PGA) as introduced through GATr architecture (Brehmer et al. (2023)). The following sections describe the core components of the AB-GATr architecture.

PGA & positional encoding: Like AB-UPT, AB-GATr operates as a transformer-based neural field that processes mesh nodes as tokens. In AB-UPT, geometric information, such as Cartesian coordinates or surface normals, is encoded in an unstructured manner by concatenating node-level features into a single vector. This approach does not preserve geometric structure and is inherently non-equivariant. In contrast, AB-GATr adopts PGA to provide a structured geometric representation, associating each token with a 16-dimensional multivector X^{MV} capable of encoding geometric primitives such as points or directions (see App A.2). Operations on multivectors respect the actions of the Euclidean group $E(3)$, ensuring equivariant processing of geometric information. Similar to Brehmer et al. (2023), AB-GATr additionally supports auxiliary scalar features X^S for encoding non-geometric attributes.

Transformer architectures typically rely on positional encodings to inject spatial context and enrich the input representation. AB-UPT employs sine–cosine embeddings (Vaswani et al., 2017) and rotary positional encoding (RoPE) (Su et al., 2024), which cannot be directly applied to geometric algebra representations without breaking equivariance. To preserve $E(3)$-equivariance in AB-GATr, positional encodings are therefore restricted only to the auxiliary scalars.

Geometric anchor-attention: Self-attention is a central component of transformer architectures, modelling all pairwise interactions among the input tokens X of size N. This results in $O(N^2)$ computational complexity, which becomes prohibitively expensive for high-resolution computational meshes in real-world scenarios, making full self-attention impractical.

To address this scale challenge, Alkin et al. (2025) introduced the anchor-attention mechanism. Anchor-attention serves two primary purposes. Firstly, it reduces the complexity to $O(MN)$ by uniformly sampling a small subset of tokens $A \subset X$, with $|A| = M \ll N$, designated as *anchors*. These anchors act as the keys and values in the attention mechanisms, while all tokens act as queries. Secondly, because non-anchor tokens attend only to anchors, each non-anchor token can be processed independently, satisfying the requirements of zero-shot super-resolution (a neural field property).

We extend the idea of anchor-attention to AB-GATr by introducing *geometric anchor-attention*, an $E(3)$-equivariant analogue. Here, input tokens $X = (X^{MV}, X^S)$, are represented by a tuple of multivectors X^{MV} and corresponding auxiliary scalars X^S. Attention weights of geometric attention are a linear combination of the following three elements: the invariant inner product of the geometric algebra $\langle \cdot, \cdot \rangle$, the distance-aware inner product of nonlinearities $\phi(\cdot), \psi(\cdot)$, and the standard Euclidean inner product of the auxiliary scalars with RoPE encoding R:

$$\mathrm{W}(k, q) := \mathrm{Softmax}\left(\frac{\alpha \sum_c \langle q_c^{MV}, k_c^{MV} \rangle + \beta \sum_c \phi(q_c^{MV}) \cdot \psi(q_c^{MV}) + \gamma \sum_c \mathrm{R}(q_c^S)\mathrm{R}(k_c^S)}{\sqrt{13 n_{MV} + n_S}} \right)$$

Where n_{MV}, n_S are respectively the dimensionalities of multivectors and scalars, c label channels, and $\alpha, \beta, \gamma > 0$ head-specific learnable weights (for more details, we refer to Brehmer et al. (2023)).

Attention weights $W(\cdot, \cdot)$ can be utilised for both geometric self-attention and geometric anchor-attention in the following manner by employing equilinear projection layers $K(\cdot)$, $Q(\cdot)$, and $V(\cdot)$:

$$\begin{aligned} \text{GeometricSelfAttention}(X) &:= W(K(X), Q(X))V(X) \\ \text{GeometricAnchorAttention}(X, A) &:= W(K(A), Q(X))V(A) \end{aligned}$$

$$A = (X^{MV}[\mathcal{I}_A], X^S[\mathcal{I}_A]) \quad \mathcal{I}_A \sim \mathrm{Uniform}(\{1, 2, ..., N\}, M)$$

Supernode pooling: To inject manifold boundary into the token representations, both volume and surface tokens attend to a set of surface-pooled supernodes (see Fig. 2) The supernodes $S = A_{\text{surface}}$ correspond to the same surface indices sampled as anchors. For each supernode $a \in A_{\text{surface}}$ a local neighborhood is constructed via a radius-ball: $\mathcal{N}_r(a) := \{b \in X_{\text{surface}} \mid \|a - b\|_2 \leq r\}$, over which message-passing pooling is performed. During pooling, message signals are augmented with the relative translation vector, encoded as a *translation* primitive (App A.2), and its magnitude as an auxiliary scalar, enabling boundary-aware aggregation while preserving $E(3)$-equivariance.

Multi-branching: Following Alkin et al. (2025), we introduce a multi-branch design in which separate branches process volume and surface tokens while sharing a subset of weights and exchanging information through attention keys and values (KV exchange). AB-GATr adopts this design and extends it to the equivariant setting by jointly exchanging both multivectors X^{MV} and auxiliary scalars X^S across branches. As illustrated in Fig. 2, the *encoder* processes volume and surface tokens with shared weights, followed by a *shared body* in which standard geometric anchor-attention is interleaved with a KV exchange to enable cross-branch interaction. In the *decoder* stage, the branches become fully independent, allowing for regression of volumetric and surface quantities.

2.4 CFD BENCHMARKS

As discussed in Section 2.2, equivariance of the solution map arises only when input signal and boundary conditions jointly preserve symmetry. Consequently, equivariant inductive biases may yield different results across benchmarks with varying degrees of natural canonical alignment. This work uses the following hemodynamic benchmarks:

SynthBifCor (Suk et al., 2024b) contains $2,000$ synthetic coronary bifurcation geometries derived from anatomical atlas measurements. Steady simulations use an inflow velocity of 11.8 cm/s at Reynolds number 100. The geometries share an inlet orientation but vary in the orientation of branch configuration.

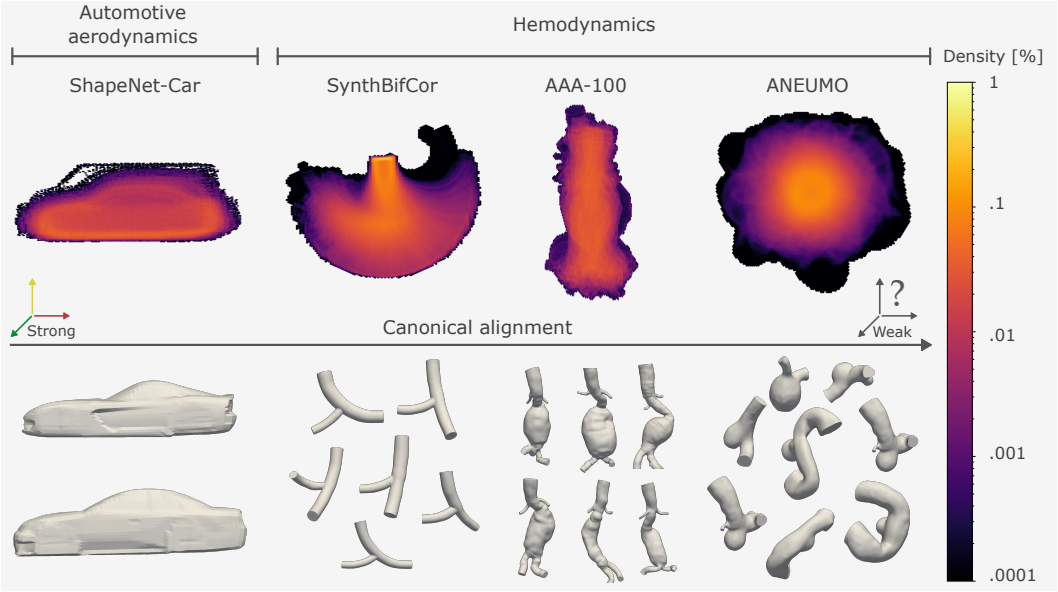

Figure 3: Canonical alignment across different benchmarks. The first row represents a histogram of point occupancy across the whole dataset. The second row shows representative geometries sampled from the dataset kept in the same orientation as the histogram.

AAA-100 (Rygiel et al. (2024; 2025a))contains 100 patient-specific abdominal aortic aneurysm (AAA) geometries reconstructed from computed tomography scans. Transient CFD simulations use a prescribed inflow waveform with a peak flow rate of 80 ml/s [2]. The geometries retain a preferred anatomical orientation with variation in tilt and axial rotation.

ANEUMO (Li et al., 2025) contains $10,660$ synthetic intracranial aneurysm geometries derived from 400 patient-specific anatomies (Juchler et al. (2022)) and simulated under steady inflow conditions with mass flow of 0.0010 kg/s [2]. The geometries exhibit broad orientation variability and an apparent lack of canonical alignment.

These hemodynamic benchmarks share the property that they represent *internal flows*, which are predominantly intrinsic to geometry, with absolute orientation carrying little or no physical meaning. In contrast, *external flows* as seen in automotive aerodynamics are defined within fixed global reference frames, which usually results in canonical alignment and explicit symmetry breaking imposed by the wind-tunnel setup. To assess the differences between these two settings, we include a fourth benchmark:

ShapeNet-Car (Umetani & Bickel, 2018) contains 3D car geometries from ShapeNet (Chang et al., 2015) with steady incompressible CFD simulations at an inflow speed of 72 km/h. All car geometries share a common canonical orientation, with a fixed inflow direction reflecting a wind-tunnel experimental setup.

These benchmarks span distinct patterns of canonical alignment and geometric variability that naturally arise in real-world CFD problems, providing a robust testbed for assessing the influence of equivariance in neural surrogate design. More details about the datasets are provided in App. B.1.

To ground our observations, we quantify canonical alignment by extending symmetry-breaking diagnostics of Lawrence et al. (2026) to examine how distinguishable the data remain from randomly rotated counterparts at original orientations and when their variability increases. Specifically, we consider an integral of symmetry response across orientation variability as a measure of data canonical alignment (App. B.2).

[2] AAA-100 and ANEUMO provide simulations under different boundary conditions. Since we are interested in the influence of geometric variability only, we use only a single set of fixed boundary conditions for both.

Table 1: We report `median` of mean relative L2 error [%] of volume velocity (u) and surface WSS (τ) over 3 runs for SynthBifCor and ANEUMO, and over 10-fold cross-validation for AAA-100. Best-performing model in bold, with an underline indicating whether its performance gain is statistically significant compared to the second-best ($p < 0.05$ in the Wilcoxon test).

	SynthBifCor				**AAA-100**				**ANEUMO**	
	original		*rotated*		*original*		*rotated*		*original*	*rotated*
	u	τ	u	τ	u	τ	u	τ	u	u
Transolver++†	1.25	8.62	103	104	71.4	81.8	123	110	48.1	48.0
AB-UPT	2.02	8.27	96.7	92.0	57.5	71.5	109	108	25.5	26.1
Transolver++$^{\dagger,\mathrm{rot}}$	2.02	8.96	2.01	8.98	92.1	66.6	88.8	65.4	20.0	19.9
AB-UPT ^{rot}	1.37	7.46	1.35	7.47	46.1	44.6	45.3	44.6	13.0	13.0
LaB-GATr†	1.21	7.46	1.21	7.46	42.0	47.3	42.0	47.3	**10.4**	**10.4**
AB-GATr	**0.98**	**6.87**	**0.98**	**6.87**	**39.0**	**38.4**	**39.0**	**38.4**	10.9	10.9

$\dagger$ - trained separately for surface and volume. rot - model trained with SO(3) data augmentation.

3 Experiments & Results

We compare AB-GATr with its non-equivariant counterpart AB-UPT, the non-equivariant Transolver++ (Luo et al., 2025), and the $E(3)$-equivariant LaB-GATr (Suk et al., 2024c), which utilises the same equivariance formulation as AB-GATr. To thoroughly benchmark the notion of equivariance, we fix AB-UPT and AB-GATr to follow the same architecture size, i.e., number of attention blocks, number of anchors/tokens, etc., with the only difference being the size of the hidden dimension and number of attention heads due to the difference between linear and geometric algebra layers (training configurations and resource requirements are reported in App. A.3).

We evaluate each model on the original test geometries and on rotated versions of the same cases. We train both non-equivariant architectures with and without $SO(3)$ rotational augmentation; augmented variants are denoted by the superscript "rot." The original test sets retain each benchmark's native orientation distribution, whereas the additional rotation evaluation tests sensitivity to a changed orientation distribution.

Hemodynamic benchmarks: Table 1 reports relative L2 errors for wall shear stress (τ) and velocity (u) estimation across hemodynamic datasets. For benchmarks with a notion of canonical alignment, namely SynthBifCor and AAA-100, non-equivariant models perform substantially worse on *rotated* datasets when trained without rotational augmentation. In contrast, for the ANEUMO benchmark, which lacks any canonical alignment, performance on oriented and rotated geometries remains similar.

Rotation augmentation can reduce the sensitivity of a model to orientation changes, but its effect on performance depends on the architecture and dataset. First, as expected, we observe that augmentation allows non-equivariant models to match performance between *original* and *rotated* test sets. However, performance gains from data augmentation are not consistent across models and benchmarks. AB-UPT's performance consistently improves across all hemodynamic benchmarks, even with original alignment. Meanwhile, Transolver++'s performance on original alignment degrades for SynthBifCor and velocity prediction in AAA-100. Nevertheless, even with augmentation, neither model can close the performance gap to AB-GATr on any benchmark and task.

LaB-GATr results suggest that the gains from explicit equivariance extend beyond specific architectural choices. Despite the architectural differences between LaB-GATr and AB-GATr, both equivariant models mostly outperform the non-equivariant baselines. Moreover, the equivariant models are the best-performing ones on every benchmark: AB-GATr leads on SynthBifCor and AAA-100, while LaB-GATr achieves a slightly lower velocity error on ANEUMO (10.4% versus 10.9%). These results identify explicit equivariance as a shared feature of the best-performing models and support symmetry preservation as an important design choice across the evaluated architectures.

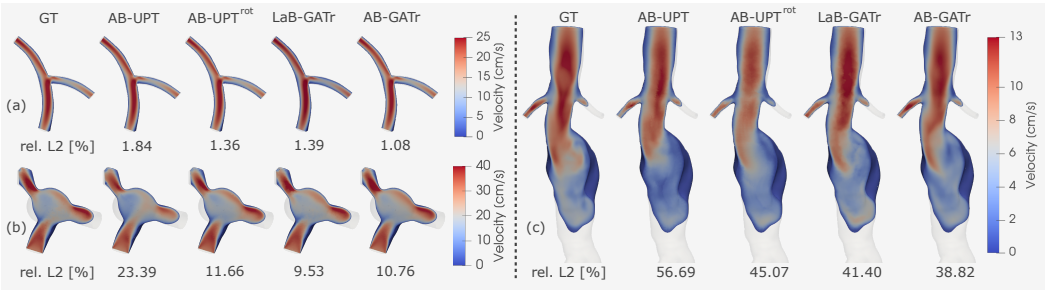

Figure 4: Qualitative comparison of neural surrogate velocity predictions across hemodynamic benchmarks: (a) SynthBifCor, (b) ANEUMO, and (c) AAA-100. In case (c), we visualise the time-averaged velocity field, since the AAA-100 benchmark consists of transient simulations.

Fig. 4 illustrates velocity field predictions for the hemodynamic benchmarks across the anchored-branched model family. Predictions on SynthBifCor (a) are largely visually indistinguishable across models, consistent with the low mean relative L2 errors ($< 2\%$). Differences are more apparent on ANEUMO (b) and AAA-100 (c), which consist of more realistic geometries and hence more complex flow patterns. Compared with LaB-GATr and AB-GATr, AB-UPT shows the weakest performance in both variants. For ANEUMO, errors become more pronounced further from the inlet, where velocity magnitudes are underestimated, and complex flow patterns are insufficiently captured. A similar behaviour is observed on AAA-100, where side branches and the aneurysm sac lack finer and high-frequency flow details present in the CFD solution. In particular, AB-GATr yields the smoothest and most coherent predictions, with the greatest improvements in the leftmost branch in example (b) and throughout the entire geometry in example (c). Additional predictions and residual maps, including Transolver++, are provided in App. B.3.

Controlled variation of canonical alignment: Because hemodynamic datasets naturally have imperfect canonical alignment, we cannot fully control its rate within a single benchmark. Hence, to thoroughly examine alignment within a fixed benchmark, we construct a controlled experiment on ShapeNet-Car. Starting from its perfect canonical orientation reflecting a wind-tunnel setup, we systematically vary the range of rotational augmentation during training and evaluation. Specifically, we train and evaluate AB-UPT using rotations sampled up to $[5°, 15°, 30°, 60°, 90°, 180°]$. Figure 5 summarises the results and reveals three trends. First, AB-UPT outperforms AB-GATr on the original canonically aligned data and under restricted rotation ranges. Second, broadening the training orientation range generally increases AB-UPT's errors across all orientations represented during training. Third, evaluation beyond that range produces severe failures of generalisation, with errors increasing sharply. AB-GATr maintains unchanged errors across the tested rotations. Together, these results indicate that problems with strong canonical alignment may benefit from a lack of equivariance, and enforcing or learning symmetry preservation can deteriorate in-distribution performance while ensuring reliable generalisation when evaluated under unseen orientations.

Diagnosing distributional symmetry breaking: As observed with ShapeNet-Car, the dataset's natural orientation can break the symmetry of the assigned problem and benefit non-equivariant models. Lawrence et al. (2026) proposed distributional symmetry breaking to diagnose this mechanism. We extend their formulation of binary classification accuracy $m(p_X)$ under random group transformation of the samples (with probability $\frac{1}{2}$) by additionally rotating all samples up to a maximum angle θ, which we denote by $m(p_X \circ r_\Theta)$. Here, $\Theta \sim \mathcal{U}_{[0,\theta]}$ and $r_\Theta \in \{R_{\theta'} \in \mathrm{SO}(3) \mid \theta' \leq \Theta\}$ where rotations $R_{\theta'}$ are described by their angle θ'. Figure 6 shows the symmetry response curve for $\theta \in [0, \pi]$. Aneumo is the only dataset without distributional symmetry breaking $m(p_X) \approx 0.5$ (i.e., random guessing), while ShapeNet-Car, SynthBifCor and AAA-100 are perfectly distinguishable $m(p_X \circ r_\theta) \approx 1$ for $\theta = 0$. Nevertheless, SynthBifCor and AAA-100 come with a weaker notion of canonical alignment; their symmetry response drops at a higher rate with increasing θ. We can quantify this phenomenon by integrating over the symmetry response curve, which provides an a priori measure of canonical alignment (App. B.2). Following the flow chart proposed by Lawrence et al. (2026), we also probe task-useful distribution symmetry breaking, which we find to correlate with our empirical observations on AB-GATr compared to AB-UPT (App. B.2).

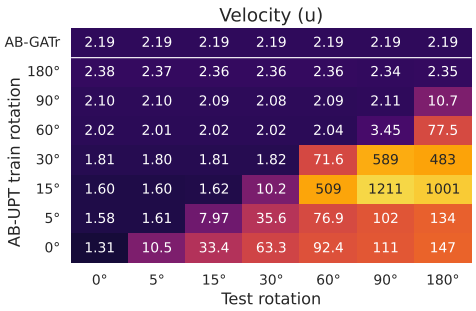

Figure 5: Median relative L2 error [%] across 5 runs on ShapeNet-Car for different training and testing rotation rates. Rotation rate $x°$ indicates that random rotations up to $x°$ were used.

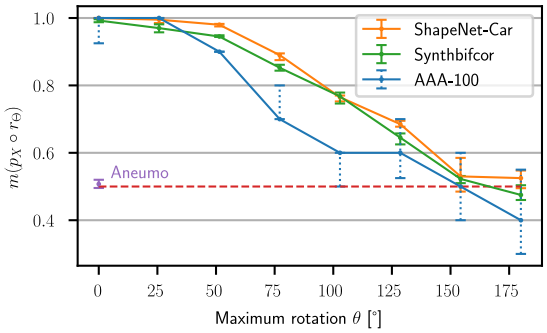

Figure 6: Symmetry response; $m(p_X \circ r_\Theta)$ measures alignment (↑). We show lower and upper quartiles over four runs (Aneumo, ShapeNet-Car, SynthBifcor) or 10-fold cross-validation (AAA-100).

4 Discussion

Dataset scale has often been cited as a key factor when assessing the benefits of equivariance (Brehmer et al., 2024). While scale matters, our results suggest it is not sufficient on its own to determine whether equivariant or non-equivariant models are preferable. In particular, the ANEUMO benchmark is an unusually large dataset (∼10k simulations), yet explicit equivariance remains beneficial. Instead, we find that the degree of canonical alignment induced by the problem formulation plays a comparably important role. Across the hemodynamic benchmarks, characterised by different orientation variation, equivariant models consistently outperform non-equivariant ones - even with rotation augmentation. Conversely, in strongly aligned settings such as ShapeNet-Car, non-equivariant models can exploit alignment-specific regularities to improve in-distribution performance, an effect also noted by Lawrence et al. (2026); Vadgama et al. (2025). The stronger symmetry response on ShapeNet-Car is consistent with AB-UPT's advantage under canonical alignment, while the results on SynthBifCor and AAA-100 demonstrate that detectable alignment does not preclude benefits from equivariance.

Limitations and future work: Our conclusions are limited to the evaluated CFD tasks, prescribed boundary-condition regimes, and model configurations. We evaluated a single notion of equivariance and architectural design, which we both chose for their computational efficiency and performance. Alternative equivariant methods such as GEM-CNN (de Haan et al., 2021), SEGNN (Brandstetter et al., 2022b) or $SE(3)$-Transformer (Fuchs et al., 2020) have been previously reported to fare worse in both performance and computational efficiency (Brehmer et al., 2023; Suk et al., 2024c). While EqGINO (Kim et al., 2026) and Platonic Transformers (Islam et al., 2026) restrict exact rotational equivariance to finite subgroups, PONITA (Bekkers et al., 2024) introduces discretisation-dependent equivariance errors. Consistent improvements in hemodynamic prediction accuracy due to explicit equivariance come at the cost of additional computational overhead. However, we found that AB-GATr substantially reduced computation cost compared to LaB-GATr (App. A.4), narrowing the gap with the non-equivariant AB-UPT model. While future work should seek to narrow this gap further, the $2-11$-fold increase in training time between AB-UPT and AB-GATr would be acceptable in most practical applications. On the other hand, other scalable architectural designs like Transolver++ (Luo et al., 2025) or Erwin (Zhdanov et al., 2025) have significantly higher computational time and memory burden at our benchmark scale, which limits adoption of equivariance in these frameworks (App. A.4). Finally, our findings are empirical and do not yet provide a principled criterion for determining when equivariance is beneficial - a key step toward guiding model design in real-world applications.

Conclusions: We introduced AB-GATr, a scalable $E(3)$-equivariant CFD surrogate, and showed that the benefits of equivariance depend on the prediction task and orientation distribution. Across the hemodynamic benchmarks studied, AB-GATr consistently outperforms the evaluated non-equivariant architectures, including variants trained with rotational augmentation.

AI USE STATEMENT

In this work, we used generative AI tools for [design or provide feedback on research methodology or experiments]. We have not used generative AI tools for [generate synthetic data sets, help develop theoretical models or conceptual frameworks, formulate mathematical claims, provide critical ingredients for proving mathematical claims, assist in the writing of proofs, implement methods, assist with translation, clean and reformat dataset, support qualitative and thematic data analysis, interpret results] and [assist with translation] are not applicable to this work.

Additionally, we used generative AI tools for [draft parts of a research paper, brainstorming, edit a research paper to improve readability, propose a title or keywords for a research paper].

We have reviewed all AI-assisted work. We checked all the polished or generated text by AI tools for potential plagiarism, factual errors and unsupported or misrepresented claims. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

Model configurations, optimisation settings, and training resource requirements are reported in App. A. App. B.1 documents benchmark statistics and data-splitting protocols, while App. B.2 provides implementation details for the symmetry-breaking diagnostics. We will publicly release the codebase, including our AB-GATr implementation and scripts for training and evaluating the models across all benchmarks.

REFERENCES

Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J. Ballard, Joshua Bambrick, Sebastian W. Bodenstein, David A. Evans, Chia-Chun Hung, Michael O'Neill, David Reiman, Kathryn Tunyasuvunakool, Zachary Wu, Akvilė Žemgulytė, Eirini Arvaniti, Charles Beattie, Ottavia Bertolli, Alex Bridgland, Alexey Cherepanov, Miles Congreve, Alexander I. Cowen-Rivers, Andrew Cowie, Michael Figurnov, Fabian B. Fuchs, Hannah Gladman, Rishub Jain, Yousuf A. Khan, Caroline M. R. Low, Kuba Perlin, Anna Potapenko, Pascal Savy, Sukhdeep Singh, Adrian Stecula, Ashok Thillaisundaram, Catherine Tong, Sergei Yakneen, Ellen D. Zhong, Michal Zielinski, Augustin Žídek, Victor Bapst, Pushmeet Kohli, Max Jaderberg, Demis Hassabis, and John M. Jumper. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630(8016):493–500, June 2024. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-024-07487-w. URL https://www.nature.com/articles/s41586-024-07487-w.

Benedikt Alkin, Andreas Fürst, Simon Lucas Schmid, Lukas Gruber, Markus Holzleitner, and Johannes Brandstetter. Universal physics transformers: A framework for efficiently scaling neural operators. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL https://openreview.net/forum?id=oUXiNX5KRm.

Benedikt Alkin, Maurits Bleeker, Richard Kurle, Tobias Kronlachner, Reinhard Sonnleitner, Matthias Dorfer, and Johannes Brandstetter. AB-UPT: Scaling neural CFD surrogates for high- fidelity automotive aerodynamics simulations via anchored- branched universal physics transformers. *Transactions on Machine Learning Research*, 2025. ISSN 2835-8856. URL https://openreview.net/forum?id=nwQ8nitlTZ.

Neil Ashton, Charles Mockett, Marian Fuchs, Louis Fliessbach, Hendrik Hetmann, Thilo Knacke, Norbert Schonwald, Vangelis Skaperdas, Grigoris Fotiadis, Astrid Walle, Burkhard Hupertz, and Danielle Maddix. Drivaerml: High-fidelity computational fluid dynamics dataset for road-car external aerodynamics, 2025. URL https://arxiv.org/abs/2408.11969.

Erik J Bekkers, Sharvaree Vadgama, Rob Hesselink, Putri A Van der Linden, and David W. Romero. Fast, expressive $\mathrm{SE}(n)$ equivariant networks through weight-sharing in position-orientation space. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=dPHLbUqGbr.

Johannes Brandstetter, Rianne van den Berg, Max Welling, and Jayesh K Gupta. Clifford neural layers for pde modeling. *arXiv preprint arXiv:2209.04934*, 2022a.

Johannes Brandstetter, Rob Hesselink, Elise van der Pol, Erik J Bekkers, and Max Welling. Geometric and physical quantities improve e(3) equivariant message passing. In *International Conference on Learning Representations*, 2022b. URL https://openreview.net/forum?id=_xwr8gOBeV1.

Johann Brehmer, Pim De Haan, Sönke Behrends, and Taco Cohen. Geometric algebra transformer. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL https://openreview.net/forum?id=M7r2CO4tJC.

Johann Brehmer, Sönke Behrends, Pim de Haan, and Taco Cohen. Does equivariance matter at scale?, 2024. URL https://arxiv.org/abs/2410.23179.

Michael M. Bronstein, Joan Bruna, Taco Cohen, and Petar Velickovic. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *CoRR*, abs/2104.13478, 2021. URL https://arxiv.org/abs/2104.13478.

Qianying Cao, Somdatta Goswami, and George Em Karniadakis. Laplace neural operator for solving differential equations. *Nat. Mach. Intell.*, 6(6):631–640, June 2024.

Giovanni Catalani, Xavier BERTRAND, Frédéric TOST, Michael BAUERHEIM, and Joseph Morlier. Geometry aware inference of steady state PDEs using equivariant neural field representations. In *NeurIPS 2025 AI for Science Workshop*, 2025. URL https://openreview.net/forum?id=Ia952enzxK.

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, 2015. URL https://arxiv.org/abs/1512.03012.

Xiangning Chen, Chen Liang, Da Huang, Esteban Real, Kaiyuan Wang, Yao Liu, Hieu Pham, Xuanyi Dong, Thang Luong, Cho-Jui Hsieh, Yifeng Lu, and Quoc V. Le. Symbolic discovery of optimization algorithms, 2023. URL https://arxiv.org/abs/2302.06675.

Pim de Haan, Maurice Weiler, Taco Cohen, and Max Welling. Gauge equivariant mesh cnns: Anisotropic convolutions on geometric graphs. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=Jnspzp-oIZE.

Fabian B. Fuchs, Daniel E. Worrall, Volker Fischer, and Max Welling. Se(3)-transformers: 3d roto-translation equivariant attention networks. In *Advances in Neural Information Processing Systems 34 (NeurIPS)*, 2020.

Bianca Griffo, Diego Gallo, David Marlevi, Marco Laudato, Girolamo Mastronuzzi, Claudio Chiastra, Alessandro Candreva, Carlos Collet, Bernard De Bruyne, Andrea Erriquez, Gianluca Campo, Simone Biscaglia, Umberto Morbiducci, and Maurizio Lodi Rizzini. Geometric deep learning-based coronary wall shear stress estimation from real-world patients. *Computers in Biology and Medicine*, 205:111583, 2026. ISSN 0010-4825. doi: https://doi.org/10.1016/j.compbiomed.2026.111583. URL https://www.sciencedirect.com/science/article/pii/S0010482526001460.

Mohammad Mohaiminul Islam, Rishabh Anand, David Wessels, Friso de Kruiff, Thijs P. Kuipers, Rex Ying, Clara I. Sánchez, Sharvaree Vadgama, Georg Bökman, and Erik J Bekkers. Platonic transformers: A solid choice for equivariance, 2026. URL https://openreview.net/forum?id=qBVIaEz6Nc.

Andrew Jaegle, Sebastian Borgeaud, Jean-Baptiste Alayrac, Carl Doersch, Catalin Ionescu, David Ding, Skanda Koppula, Andrew Brock, Evan Shelhamer, Olivier J. H'enaff, Matthew M. Botvinick, Andrew Zisserman, Oriol Vinyals, and João Carreira. Perceiver io: A general architecture for structured inputs & outputs. *ArXiv*, abs/2107.14795, 2021. URL https://api.semanticscholar.org/CorpusID:236635379.

Norman Juchler, Sabine Schilling, Philippe Bijlenga, Vartan Kurtcuoglu, and Sven Hirsch. Shape trumps size: Image-Based morphological analysis reveals that the 3D shape discriminates intracranial aneurysm disease status better than aneurysm size. *Front Neurol*, 13:809391, May 2022.

Sungwon Kim, Juho Song, Seungmin Shin, Guimok Cho, Sangkook Kim, and Chanyoung Park. EqGINO: Equivariant geometry-informed fourier neural operators for 3d PDEs. In *AI&PDE: ICLR 2026 Workshop on AI and Partial Differential Equations*, 2026. URL https://openreview.net/forum?id=rY6mQMW8nS.

Vincent Lannelongue, Paul Garnier, Pablo Jeken-Rico, Aurèle Goetz, Philippe Meliga, Yves Chau, and Elie Hachem. Physics constrained graph neural network for real time prediction of intracranial aneurysm hemodynamics. *npj Digital Medicine*, 9, February 2026. doi: 10.1038/s41746-026-02404-z.

Hannah Lawrence, Elyssa Hofgard, Vasco Portilheiro, Yuxuan Chen, Tess Smidt, and Robin Walters. To augment or not to augment? diagnosing distributional symmetry breaking. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=ZTb4YmHD9n.

Xigui Li, Yuanye Zhou, Feiyang Xiao, Xin Guo, Chen Jiang, Tan Pan, Xingmeng Zhang, Cenyu Liu, Zeyun Miao, Jianchao Ge, Xiansheng Wang, Qimeng Wang, Yichi Zhang, Wenbo Zhang, Fengping Zhu, Limei Han, Yuan Qi, Chensen Lin, and Yuan Cheng. Aneumo: A large-scale multimodal aneurysm dataset with computational fluid dynamics simulations and deep learning benchmarks, 2025. URL https://arxiv.org/abs/2505.14717.

Zijie Li, Kazem Meidani, and Amir Barati Farimani. Transformer for partial differential equations' operator learning. *Transactions on Machine Learning Research*, 2023. ISSN 2835-8856. URL https://openreview.net/forum?id=EPPqt3uERT.

Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3(3):218–229, 2021.

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 *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=AM7iAh0krx.

Michael Lutter, Christian Ritter, and Jan Peters. Deep lagrangian networks: Using physics as model prior for deep learning. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=BklHpjCqKm.

Joaquim R. R. A. Martins. Aerodynamic design optimization: Challenges and perspectives. *Computers & Fluids*, 239:105391, 2022. ISSN 0045-7930. URL https://www.sciencedirect.com/science/article/pii/S0045793022000615.

Artem Moskalev, Anna Sepliarskaia, Erik J Bekkers, and Arnold W.M. Smeulders. On genuine invariance learning without weight-tying. In Timothy Doster, Tegan Emerson, Henry Kvinge, Nina Miolane, Mathilde Papillon, Bastian Rieck, and Sophia Sanborn (eds.), *Proceedings of 2nd Annual Workshop on Topology, Algebra, and Geometry in Machine Learning (TAG-ML)*, volume 221 of *Proceedings of Machine Learning Research*, pp. 218–227. PMLR, 28 Jul 2023. URL https://proceedings.mlr.press/v221/moskalev23a.html.

M. Raissi, P. Perdikaris, and G. E. 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. ISSN 0021-9991. doi: https://doi.org/10.1016/j.jcp.2018.10.045. URL https://www.sciencedirect.com/science/article/pii/S0021999118307125.

Patryk Rygiel, Paweł Płuszka, Maciej Zięba, and Tomasz Konopczyński. CenterlinePointNet++: A New Point Cloud Based Architecture for Coronary Artery Pressure Drop and vFFR Estimation. In Hayit Greenspan, Anant Madabhushi, Parvin Mousavi, Septimiu Salcudean, James Duncan, Tanveer Syeda-Mahmood, and Russell Taylor (eds.), *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*, volume 14226, pp. 781–790. Springer Nature Switzerland, Cham, 2023. ISBN 978-3-031-43989-6 978-3-031-43990-2. doi: 10.1007/978-3-031-43990-2_73. URL https://link.springer.com/10.1007/978-3-031-43990-2_73. Series Title: Lecture Notes in Computer Science.

Patryk Rygiel, Dieuwertje Alblas, Christoph Brune, Stefan Smorenburg, Kak Khee Yeung, and Jelmer M. Wolterink. AAA-100: A curated dataset of 3d watertight abdominal aortic aneurysm models, April 2024. URL https://doi.org/10.5281/zenodo.10932957.

Patryk Rygiel, Julian Suk, Christoph Brune, Kak Khee Yeung, and Jelmer M. Wolterink. Wall shear stress estimation in abdominal aortic aneurysms: Towards generalisable neural surrogate models, 2025a. URL https://arxiv.org/abs/2507.22817.

Patryk Rygiel, Julian Suk, Kak Khee Yeung, Christoph Brune, and Jelmer M. Wolterink. Active learning for deep learning-based hemodynamic parameter estimation, 2025b. URL https://arxiv.org/abs/2503.03453.

Jianlin Su, Murtadha Ahmed, Yu Lu, Shengfeng Pan, Wen Bo, and Yunfeng Liu. RoFormer: Enhanced transformer with Rotary Position Embedding. *Neurocomputing*, 568:127063, 2024. ISSN 0925-2312. doi: https://doi.org/10.1016/j.neucom.2023.127063. URL https://www.sciencedirect.com/science/article/pii/S0925231223011864.

Julian Suk, Christoph Brune, and Jelmer M. Wolterink. SE(3) Symmetry Lets Graph Neural Networks Learn Arterial Velocity Estimation from Small Datasets. In Olivier Bernard, Patrick Clarysse, Nicolas Duchateau, Jacques Ohayon, and Magalie Viallon (eds.), *Functional Imaging and Modeling of the Heart*, volume 13958, pp. 445–454. Springer Nature Switzerland, Cham, 2023. ISBN 978-3-031-35301-7 978-3-031-35302-4. doi: 10.1007/978-3-031-35302-4_46. URL https://link.springer.com/10.1007/978-3-031-35302-4_46. Series Title: Lecture Notes in Computer Science.

Julian Suk, Dieuwertje Alblas, B A Hutten, Albert Wiegman, Christoph Brune, Pim van Ooij, and Jelmer M. Wolterink. Physics-informed graph neural networks for flow field estimation in carotid arteries. *Medical image analysis*, 110:103974, 2024a. URL https://api.semanticscholar.org/CorpusID:271865668.

Julian Suk, Pim de Haan, Phillip Lippe, Christoph Brune, and Jelmer M. Wolterink. Mesh neural networks for SE(3)-equivariant hemodynamics estimation on the artery wall. *Computers in Biology and Medicine*, 173:108328, 2024b. ISSN 0010-4825. doi: https://doi.org/10.1016/j.compbiomed.2024.108328. URL https://www.sciencedirect.com/science/article/pii/S0010482524004128.

Julian Suk, Baris Imre, and Jelmer M. Wolterink. LaB-GATr: Geometric Algebra Transformers for Large Biomedical Surface and Volume Meshes. In Marius George Linguraru, Qi Dou, Aasa Feragen, Stamatia Giannarou, Ben Glocker, Karim Lekadir, and Julia A. Schnabel (eds.), *Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*, pp. 185–195, Cham, 2024c. Springer Nature Switzerland. ISBN 978-3-031-72390-2.

Julian Suk, Guido Nannini, Patryk Rygiel, Christoph Brune, Gianluca Pontone, Alberto Redaelli, and Jelmer M. Wolterink. Deep vectorised operators for pulsatile hemodynamics estimation in coronary arteries from a steady-state prior, 2025. ISSN 0169-2607. URL https://www.sciencedirect.com/science/article/pii/S016926072500375X.

Charles A Taylor, Timothy A Fonte, and James K Min. Computational fluid dynamics applied to cardiac computed tomography for noninvasive quantification of fractional flow reserve: scientific basis. *J Am Coll Cardiol*, 61(22):2233–2241, April 2013.

Nobuyuki Umetani and Bernd Bickel. Learning three-dimensional flow for interactive aerodynamic design. *ACM Trans. Graph.*, 37(4), July 2018. ISSN 0730-0301. doi: 10.1145/3197517.3201325. URL https://doi.org/10.1145/3197517.3201325.

Sharvaree Vadgama, Mohammad Mohaiminul Islam, Domas Buracas, Christian Shewmake, Artem Moskalev, and Erik Bekkers. Probing equivariance and symmetry breaking in convolutional networks, 2025. URL https://arxiv.org/abs/2501.01999.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is All you Need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

Yuyang Wang, Ahmed A. Elhag, Navdeep Jaitly, Joshua M. Susskind, and Miguel Ángel Bautista. Swallowing the bitter pill: Simplified scalable conformer generation. In *Forty-first International Conference on Machine Learning*, 2024.

David Wessels, David M Knigge, Riccardo Valperga, Samuele Papa, Sharvaree Vadgama, Efstratios Gavves, and Erik J Bekkers. Grounding continuous representations in geometry: Equivariant neural fields. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=A4eCzSohhx.

Petra N Williamson, Paul D Docherty, Sina G Yazdi, Adib Khanafer, Natalia Kabaliuk, Mark Jermy, and Patrick H Geoghegan. Review of the development of hemodynamic modeling techniques to capture flow behavior in arteries affected by aneurysm, atherosclerosis, and stenting. *J Biomech Eng*, 144(4), April 2022.

Haixu Wu, Huakun Luo, Haowen Wang, Jianmin Wang, and Mingsheng Long. Transolver: a fast transformer solver for pdes on general geometries. ICML'24. JMLR.org, 2024.

Maksim Zhdanov, Max Welling, and Jan-Willem van de Meent. Erwin: A tree-based hierarchical transformer for large-scale physical systems. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=MrphqqwnKv.

A Implementation details

A.1 Data standarization

In all the benchmarks we study, vector-valued targets are equivariant with respect to $O(3)$. AB-UPT utilises per-axis standardisation to mean of 0 and a standard deviation of 1 for vectors (Alkin et al., 2025). However, such a standardisation technique breaks equivariance.

According to Theorem A.1, the only construction of vector standardisation that respects $O(3)$-equivariance is when the mean is 0, and the standard deviation is constant across the axes (in other words, the direction of the vectors *does not* change under standardisation). As such, to retain $O(3)$-equivariance, we standardise vector labels for equivariant models by dividing by the vector magnitude mean (i.e. divide by a single scalar). For non-equivariant, we follow per-axis standardisation to a mean of 0 and standard deviation of 1, which yielded empirically better results than magnitude standardisation only in our experiments.

Theorem A.1. *Let $f : \mathbb{R}^c \mapsto \mathbb{R}^3$ be an $O(3)$-equivariant function, i.e., $(\forall \rho \in O(3))(f(\rho x) = \rho f(x))$, where $x \in \mathbb{R}^c$. Let's consider the following function $g(x) = \sigma \odot f(x) + \mu$, which represents the standardisation inverse with mean $\mu \in \mathbb{R}^3$ and standard deviation $\sigma \in \mathbb{R}^3$. A function $g(x)$ is $O(3)$-equivariant only if $\mu = 0$ and $\sigma_1 = \sigma_2 = \sigma_3$*

Proof. A function $g(x)$ is $O(3)$-equivariant if following identity holds for every group action $\rho \in O(3)$ and every input $x \in \mathbb{R}^c$:

$$g(\rho x) = \rho g(x)$$

$$g(Ax) = \sigma \odot f(\rho x) + \mu = \sigma \odot (\rho f(x)) + \mu$$
$$\rho g(x) = \rho(\sigma \odot f(x) + \mu) = \rho(\sigma \odot f(x)) + \rho\mu$$

$$\sigma \odot (\rho f(x)) + \mu = \rho(\sigma \odot f(x)) + \rho\mu$$

First, let's set $f(x) = 0$. Then, $(\forall A \in O(3))(\mu = \rho\mu)$, which means that value of μ is invariant under $O(3)$. The only such vector is the zero vector. As such for $g(x)$ to be $O(3)$-equivariant, $\mu = 0$.

By substituting $\mu = 0$ we get that $\sigma \odot (\rho f(x)) = \rho(\sigma \odot f(x))$. By writing σ as a diagonal matrix $D = diag(\sigma_1, \sigma_2, \sigma_3)$, we get that $g(x)$ is $O(3)$-equivariant if $(\forall \rho \in O(3))(D\rho f(x) = \rho D f(x) \implies D\rho = \rho D)$. This means that D and A are commuting matrices. However, according to Schur's lemma, the only commutant of $O(3)$ group is of the form cI, where c is any scalar. As such, $\sigma_1 = \sigma_2 = \sigma_3$.

Consequently, $g(x)$ is $O(3)$-equivariant only if $\mu = 0$ and $\sigma_1 = \sigma_2 = \sigma_3$. □

A.2 Input features

For ShapeNet-Car, we follow Alkin et al. (2025) and use only Cartesian coordinates as input features for both surface and volumetric representations. Across all experiments and benchmarks, the Cartesian coordinates are scaled to fit in $[0, 1000]$ cube, following Alkin et al. (2025).

For the hemodynamic benchmarks, we adopt the input feature design proposed in Suk et al. (2024b); Rygiel et al. (2025a), which has been shown to improve model expressiveness. Surface features include Cartesian coordinates, normal vectors, and surface geodesic distances to the inlet and outlets. Volumetric features consist of Cartesian coordinates and volume geodesic distances to the vessel wall, inlet, and outlets. For the AAA-100 dataset, we additionally incorporate principal curvatures, inflow rate, and flow prior features as described in Rygiel et al. (2025a).

AB-GATr and LaB-GATr employ projective geometric algebra $\mathbf{G}(3, 0, 1)$ to embed geometric quantities as 16-dimensional multivectors, following Equation 1 and Table 2. In these models, vector-valued features (normal vectors and flow priors) are embedded as *planes*, Cartesian coordinates as *points*, and scalar features (geodesic distances, curvatures, and inflow rate) are provided as auxiliary scalars.

$$x = (x_s, \underbrace{x_0, x_1, x_2, x_3}_{\text{vectors}}, \underbrace{x_{01}, x_{02}, x_{03}, x_{12}, x_{13}, x_{23}}_{\text{bivectors}}, \underbrace{x_{012}, x_{013}, x_{023}, x_{123}}_{\text{trivectors}}, x_{0123}) \tag{1}$$

Table 2: Embedding of vectors (planes), points and scalars as 16-dimensional multivectors in PGA (Eq. equation 1), the remaining elements of x remain zero.

Geometric object / operation	PGA embedding
Scalar $s \in \mathbb{R}$	$x_s = s$
Plane with normal $\nu \in \mathbb{R}^3$, offset $\delta \in \mathbb{R}$	$(x_0, x_1, x_2, x_3) = (\delta, \nu)$
Point $\rho \in \mathbb{R}^3$	$(x_{012}, x_{013}, x_{023}, x_{123}) = (\rho, 1)$

A.3 Training details

To isolate the effect of equivariance, we fix architectural hyperparameters across all benchmarks and minimise differences unrelated to symmetry enforcement. For each benchmark (ShapeNet-Car, SynthBifCor, ANEUMO, AAA-100), we fix volume (4096, 5000, 5000, 5000) and surface (3586, 5000, 5000, 5000) anchors and tune dropout $(0\%, 0\%, 0\%, 10\%)$ and the pooling radius $(9, 9, 1, 1)$. These settings are shared across all architectures within each benchmark.

AB-UPT and AB-GATr use a single geometry-branch block, followed by 9 shared blocks alternating between standard and KV-exchange (geometric) anchor attention, and 2 separate decoder blocks for the surface and volume branches - resulting in 12 attention blocks per branch with 10 sharing weights. Due to differences between standard linear layers and geometric algebra layers, the two models employ different hidden feature dimensions. AB-UPT uses a hidden dimension of 192 with 3 attention heads, following the configuration of Alkin et al. (2025). In contrast, AB-GAT uses 16 multivector channels, 64 scalar channels, and 4 attention heads. Total number of trainable parameters is 7M and 1.6M AB-UPT and AB-GATr, respectively.

LaB-GATr utilizes 12 attention blocks and follows the same transformer block configuration as AB-GATr, with 192 hidden channels and 4 attention heads. Total number of trainable parameters is 1.3M.

Transolver++ utilizes 8 layers with hidden dimension of 192, 4 attention heads and number of slices set to 64. Total number of trainable parameters is 2M.

All models are trained with a batch size of 1, gradient clipping with a threshold of 1.0, and the LION optimiser (Chen et al. (2023)) with weight decay 5×10^{-2}. We use a linear learning-rate warmup for 5% of the training schedule, followed by cosine decay to a final learning rate of 10^{-6}. The peak learning rate is tuned separately for each model and benchmark in the range $[10^{-5}, 10^{-3}]$ (see Table 3). All models are trained in full `float32` precision, as we observed instability when using mixed `float16` precision for AB-GATr. Mean squared error is used as the training loss for both surface and volume quantities, with the final loss being a mean of both branch losses.

Table 3: Training details and resources. For non-joint models, we denote volume variant by "V" and surface variant by "S". Column "Time [h]" refers to total training time and "VRAM [GB]" to peak GPU memory usage.

Benchmark	Model	LR	Epochs	Time [h]	VRAM [GB]
SynthBifCor	AB-UPT	2e-5	1000	34	2.3
	Transolver++ (V | S)	2e-5	1000	104 | 33	24.3 | 2.7
	LaB-GATr (V | S)	5e-5	300	285 | 39	10.6 | 3.9
	AB-GATr	1e-4	300	70	9.5
AAA-100	AB-UPT	5e-5	500	1	1.9
	Transolver++ (V | S)	5e-5	500	8.3 | 2.4	71 | 4.5
	LaB-GATr (V | S)	2e-4	1000	110 | 14	20 | 4.9
	AB-GATr	2e-4	1000	11	7
ANEUMO	AB-UPT	5e-5	100	17	1.9
	Transolver++ (V)	5e-5	100	59	43
	LaB-GATr (V)	5e-5	50	226	9.7
	AB-GATr	1e-4	50	51	6.5

A.4 Scalability & resources

All the models were trained on a single NVIDIA H100/80GB, taking approximately the following resources as denoted in Table 3.

As reported in Table 3, AB-UPT offers substantially better time and memory usage than Transolver++. This fact makes it far more optimal to serve as a backbone of an equivariant network due to the usual computational overhead that comes with the introduction of equivariant layers. By adopting components of AB-UPT, AB-GATr achieves similar scaling capabilities (with equivariance overhead introduced by GATr). AB-GATr allows for the training of geometries of arbitrary size, due to its training memory complexity being dependent on the number of anchor tokens only. During inference, the memory complexity depends on the number of query tokens as well; however, a chunked inference scheme (as described in Alkin et al. (2025)) can be employed to split all query tokens into smaller subsets that fit in GPU memory. This is a substantial gain over both Transolver++ and LaB-GATr, whose training memory complexity depends on the number of all tokens. Moreover, LaB-GATr's initial subsampling employs a costly farthest-point sampling scheme that scales poorly with the number of tokens, which is reflected by the high training times as reported in Table 3.

A.5 Matching training compute across architectures

Table 3 shows that training costs differ across architectures on the same benchmark. We explored compute-matched configurations by increasing the training budgets of non-equivariant models to those of AB-GATr. In

these experiments, performance plateaued or deteriorated as overfitting increased, particularly on AAA-100. We therefore report the selected best-performing training configurations alongside their runtimes and memory requirements, allowing predictive accuracy to be assessed together with computational cost. The reported comparisons consequently use different training budgets.

B CFD BENCHMARKS

B.1 BENCHMARK DETAILS

In Table 4, we provide dataset details including their complexity, size and available target fields. In our experiments, we have left out pressure fields available for hemodynamic datasets to focus on vector-valued fields. However, the experiments and models can be naturally extended to predict pressure, as used in ShapeNet-Car.

In Fig 7 we expand upon Fig 3 to include point occupancy histogram in all Cartesian axes.

Table 4: Benchmark details. By | we separate volume (left) and surface (right) quantities. The dataset split represents `train:valid:test` dataset sizes. We list all available prediction fields per dataset and mark ones used in our benchmarks in **bold**.

Name	Type	#points (mean ± std)	Dataset split	Fields	Timepoints
SynthBifCor	hemo	150k ± 34k | 17k ± 2.3k	1600:200:200[2]	$\boldsymbol{u}, p \mid \boldsymbol{\tau}$	1
AAA-100	hemo	340k ± 115k | 30k ± 5.5k	90:0:10[1]	$\boldsymbol{u}, p \mid \boldsymbol{\tau}$	21
ANEUMO	hemo	186k ± 53k | 10k ± 2.7k	8495:1072:1093[3]	$\boldsymbol{u}, p \mid$ -	1
ShapeNet-Car	aero	29k | 3.6k	789:0:100[4]	$\boldsymbol{u} \mid p$	1

[1] 10-fold cross-validation
[2] from Suk et al. (2023)
[3] stratified according to the underlying base anatomies
[4] from Alkin et al. (2024)

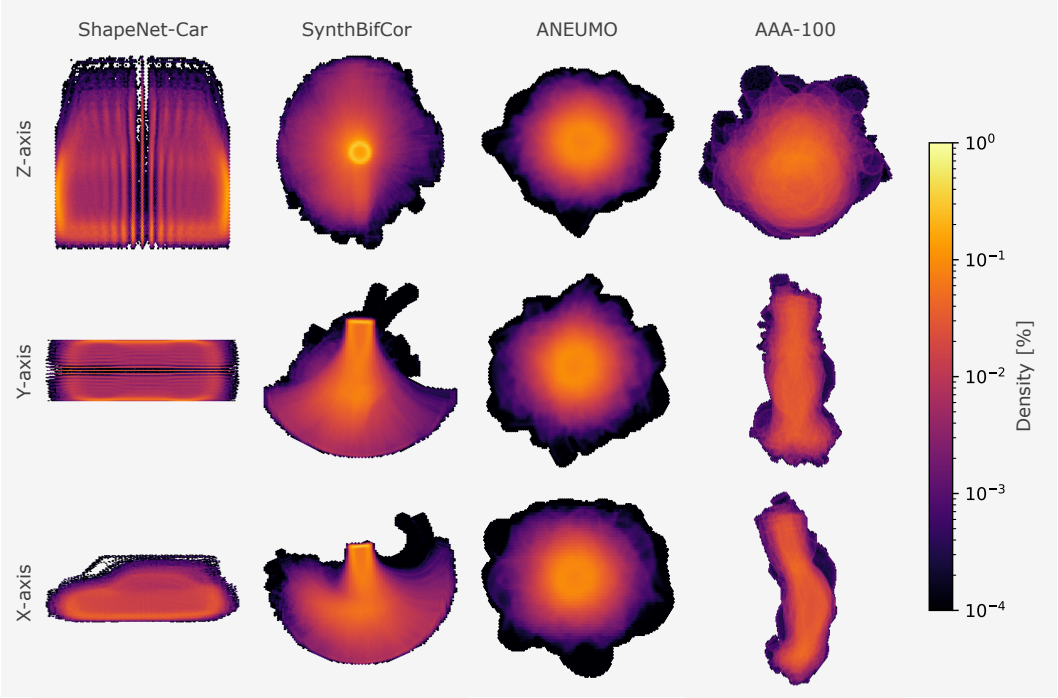

Figure 7: Canonical alignment across different benchmarks: a histogram of point occupancy across all 3 Cartesian axes.

Table 5: Symmetry response $\int m(p_X) \in [0, \pi]$ and task-useful distributional symmetry breaking $t(p_{X,Y})$ measure canonical alignment of the dataset (↑) and its influence on the task (↑), respectively. We show the mean over the test split (ShapeNet-Car, SynthBifcor) or 10-fold cross-validation (AAA-100).

	ShapeNet-Car	SynthBifCor	AAA-100
$\int m(p_X \circ r_\Theta) - \frac{\pi}{2}$	0.939	0.863	0.666
$t(p_{X,Y})$	29.23	8.52	2.35

B.2 On distributional symmetry breaking

To compute $m(p_X \circ r_\Theta)$, we trained thin classifiers (cross-attention; ShapeNet-Car, SynthBifCor and AAA-100: 108,354 parameters, points and surface normal features; Aneumo: 107,970 parameters, points features) until perceived convergence on the validation splits (ShapeNet-Car: 92,520 iterations, SynthBifCor: 44,800 iterations, Aneumo: 135,920 iterations) or training split (AAA-100: 2,967 iterations) with batch size four and Adam optimiser. These binary classifiers are trained to distinguish original samples from samples transformed under $\mathrm{SO}(3)$.

In Table 5 we report the integrated symmetry response $\int_0^\pi m(p_X \circ r_\Theta)\mathrm{d}\theta$ with $\Theta \sim \mathcal{U}_{[0,\theta]}$ and $r_\Theta \in \{R_{\theta'} \in \mathrm{SO}(3) \mid \theta' \leq \Theta\}$ where rotations $R_{\theta'}$ are described by their angle θ'. In other words, r_Θ is a random rotation up until a maximum angle θ. This provides a measure of how close the dataset is to being perfectly aligned, or on the contrary, fully symmetric. Moreover, this measure can be applied a priori, without having to train a task-specific model. We find that ShapeNet-Car and SynthBifCor are more strongly aligned relative to AAA-100, indicated by higher values of $\int m(p_X \circ r_\Theta)$.

We also report task-useful distributional symmetry breaking as introduced by Lawrence et al. (2026). We compute this from mean absolute errors on the test splits (ShapeNet-Car: p, SynthBifCor: τ) or 10-fold cross-validation (AAA-100: τ) after training AB-UPT (ca. 400,000 parameters, points and surface normal features) until perceived convergence on the validation splits (ShapeNet-Car: 180,000 iterations, SynthBifCor: 56,800 iterations) or training split (AAA-100: 10,534 iterations) with batch size four and Adam optimiser. Task-useful distributional symmetry breaking requires passing inputs through a canonicalisation function for which we chose a randomly initialised and frozen GATr geometric product layer where we break reflection equivariance by keeping a fixed reference multivector and translation equivariance by encoding the point-wise distance to the origin (after centering). This leaves only $\mathrm{SO}(3)$ equivariance intact; breaking the other symmetries was necessary for our experiments. The canonicalisation function outputs features in equivariant as well as invariant pathways. We find that ShapeNet-Car labels are strongly coupled with its canonical alignment followed by SynthBifCor and least of all, AAA-100.

We note that both presented metrics correlate with the efficacy of AB-GATr over AB-UPT (compare Table 1), i.e., equivariance. In the strongly canonically aligned dataset ShapeNet-Car, labels can be inferred from this distributional symmetry breaking which provides a powerful shortcut when test samples exhibit the same alignment as training samples. This mechanism becomes less and less powerful as alignment gets weaker or labels no longer correlate, e.g., for the ANEUMO.

B.3 Additional qualitative results

In this appendix we include additional qualitative results for SynthBifCor (Fig. 9), AAA-100 (Fig. 11) and ANEUMO (Fig 10). Each figure includes three examples that belong to roughly the first quartile (Q1), second quartile/median (Q2) and third quartile (Q3) of the relative L2 error distribution. They can be treated as easy, median and hard samples, respectively. The presented examples have been predicted by the models in their original dataset alignment. Each example is accompanied by the velocity contour plot and absolute error plot with respect to GT (CFD).

Additionally, in Fig. 8 we showcase changes in prediction quality over the anchored-branched model family with different sample orientations. Non-equivariant models yield different error rates based on the sample orientation, while equivariant AB-GATr offers the same level of predictive quality. We can also observe that AB-UPT trained with no data augmentations yields substantially different error maps across different orientations.

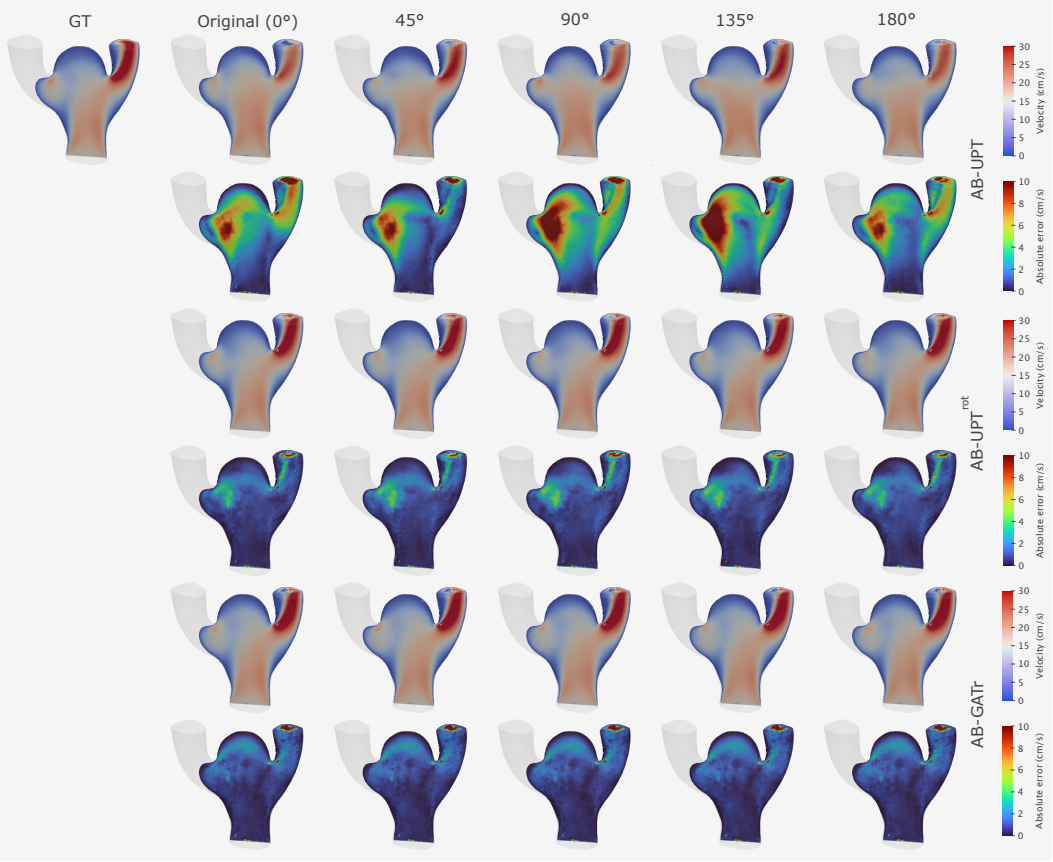

Figure 8: Qualitative comparison of neural surrogate velocity predictions single ANEUMO sample with different rotations applied ($0°$, $45°$, $90°$, $135°$, $180°$). For the sake of visualization all samples are represented in the original alignment.

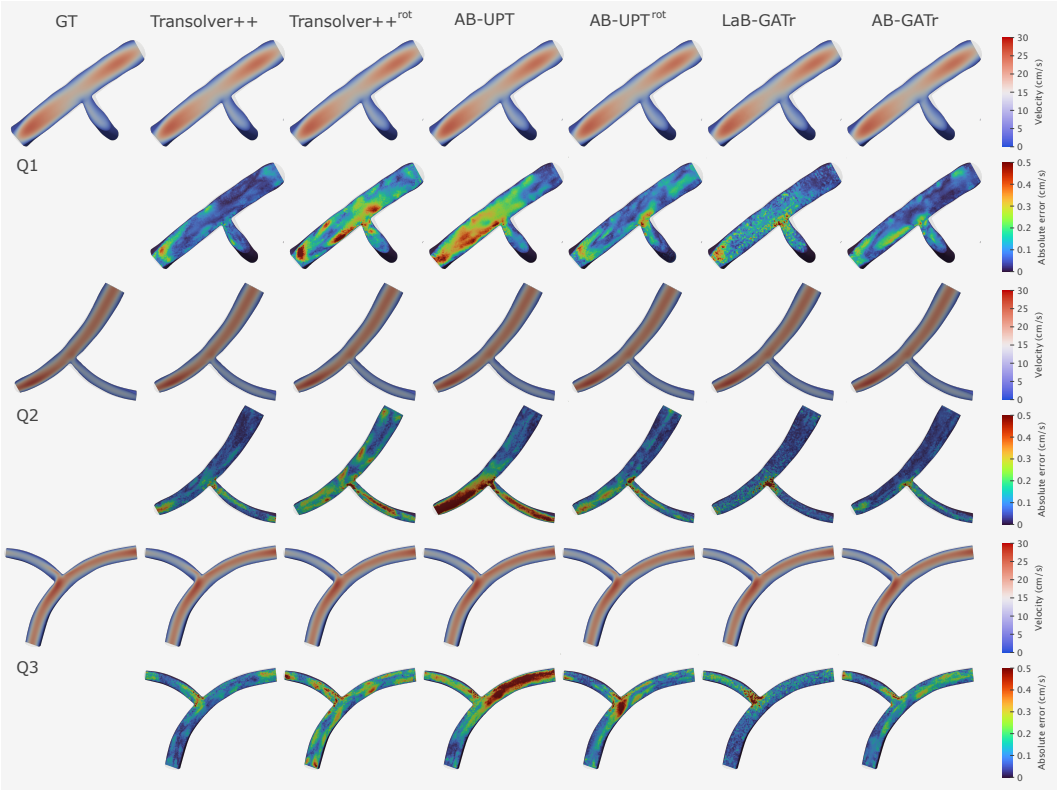

Figure 9: Extended qualitative comparison of neural surrogate velocity predictions on originally aligned SynthBifCor. Samples Q1, Q2 and Q3 correspond roughly to first, second (median) and third quartile of relative L2 error distribution.

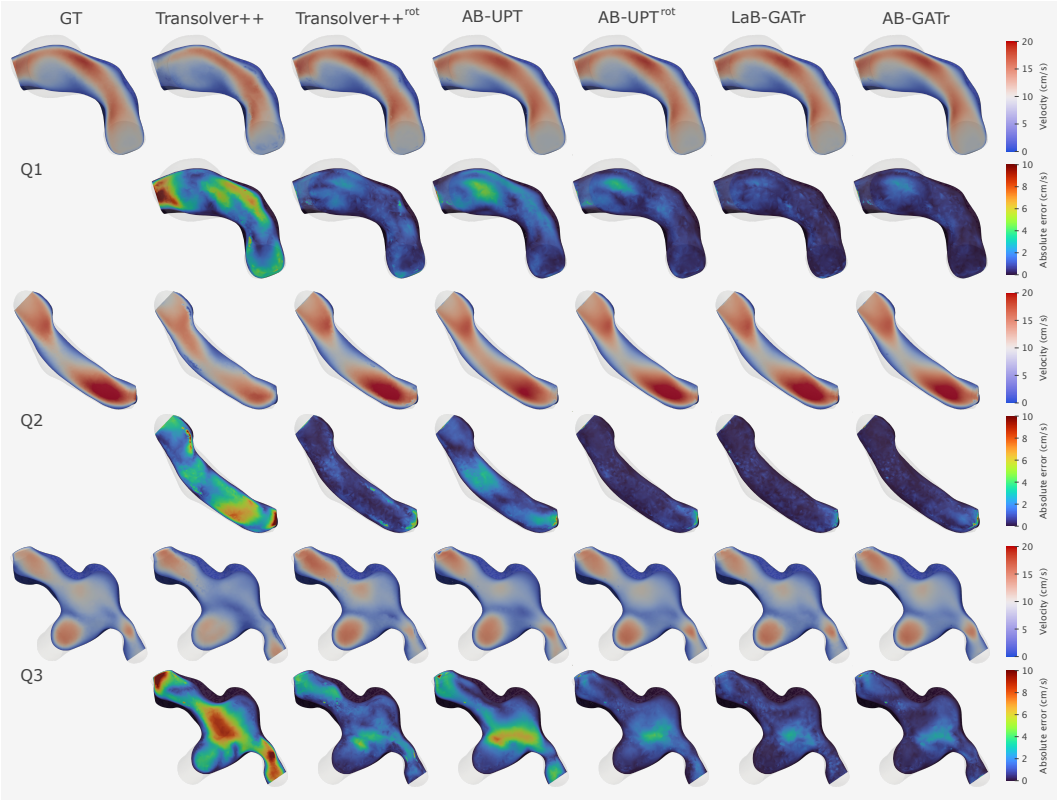

Figure 10: Extended qualitative comparison of neural surrogate velocity predictions on originally aligned ANEUMO. Samples Q1, Q2 and Q3 correspond roughly to first, second (median) and third quartile of relative L2 error distribution.

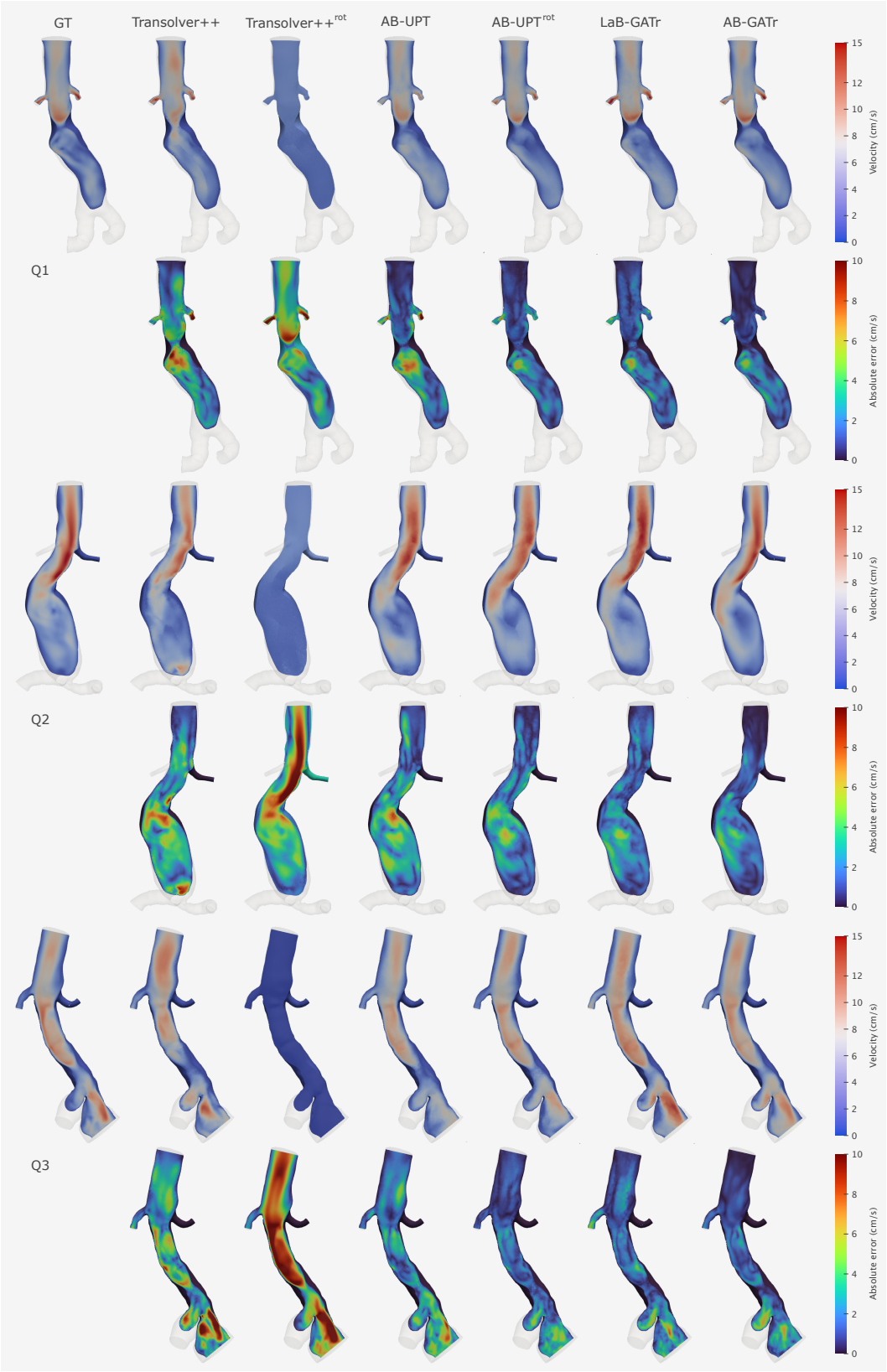

Figure 11: Extended qualitative comparison of neural surrogate velocity predictions on originally aligned AAA-100. Samples Q1, Q2 and Q3 correspond roughly to first, second (median) and third quartile of relative L2 error distribution.