LGO: A LOCAL GEOMETRY BRANCH FOR AUGMENTING NEURAL OPERATORS IN 3D INDOOR AIRFLOW PREDICTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Predicting indoor airflow across changing diffuser layouts requires models to capture local supply-jet behaviour, governed by nearby flow-driving boundaries, and room-scale recirculation set by the enclosure. Neural operators efficiently model global flow fields but typically lack explicit local boundary information at each query point. We introduce the *Local–Global Operator* (LGO), a lightweight local-geometry module using query-relative boundary coordinates stratified by boundary type to preserve small but important supply and return patches. A zero-initialised projection enables controlled comparison with the original operator. We evaluate LGO on quality-controlled CFD simulations of a single room with diffuser layout as the only varying design parameter. Generalisation is tested using separation-controlled train–test splits based on a dimensionless layout distance, with the threshold set where a nearest-layout retrieval baseline loses predictive skill. Across three neural-operator architectures, LGO adds about 1% more parameters, improves full-field velocity R^2 on both splits, and remains more stable as test layouts become more separated from training. Improvements concentrate near supply jets, while no consistent gain is observed in the ASHRAE Standard 55 occupied zone, where model differences are comparable to CFD repeatability and a simple lookup baseline. These results show that query-relative boundary information improves geometric transfer where local boundary forcing is important, at the cost of additional inference time.

1 INTRODUCTION

The performance of heating, ventilation, and air-conditioning (HVAC) systems depends strongly on how conditioned air is distributed within the occupied space. Among the geometric factors governing indoor airflow, supply-diffuser placement is especially important because it shapes supply-jet paths, air distribution, short-circuiting to returns, and room-scale transport of heat and contaminants. Computational fluid dynamics (CFD) has therefore long been used to resolve indoor airflow and evaluate ventilation configurations (Goldasteh et al., 2012; Gilani et al., 2016; Sørensen & Nielsen, 2003; Rodríguez-Vázquez et al., 2020; Tian et al., 2018).

Systematic layout exploration with CFD remains computationally expensive because each configuration requires updated boundary conditions and a new converged flow solution. Learned surrogates can reduce this cost (Chen et al., 2025; Wei & Ooka, 2025), while neural operators are well suited to learning mappings from input conditions to spatial flow fields (Lu et al., 2021; Li et al., 2021; Kovachki et al., 2023). However, varying diffuser locations creates a geometric representation challenge because indoor airflow combines local supply jets, which depend strongly on nearby flow-driving boundaries, with room-scale recirculation governed by the overall geometry and boundary conditions. Neural operators based mainly on global room information and absolute query coordinates can capture the overall flow pattern but do not explicitly provide each query point with nearby boundary geometry. As a result, local jet structures and strong velocity gradients must be inferred indirectly from global information.

Existing methods address only part of this representation problem. Neural operators improve global field representations using Fourier features (Tancik et al., 2020), spectral kernels (Li et al., 2021;

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.

2023), or learned token interactions (Hao et al., 2023; Wu et al., 2024), but they do not explicitly provide each query point with the relative geometry of nearby flow-driving boundaries. Point-cloud and mesh-based methods can encode local geometry through neighbourhood aggregation (Qi et al., 2017b; Thomas et al., 2019; Zhao et al., 2021; Pfaff et al., 2021), but standard distance-based neighbourhoods may be dominated by large wall surfaces, leaving small but important inlet and outlet patches poorly represented. In indoor airflow applications, existing surrogates have also focused mainly on changes in operating conditions within fixed or limited geometries, rather than generalisation across different diffuser layouts (Bian et al., 2026). Thus, two key gaps remain: existing approaches do not explicitly combine query-relative boundary geometry with enclosure-scale context while preserving the visibility of flow-driving surfaces; and, because layouts sampled from a low-dimensional design space can be highly similar, standard evaluation protocols may not distinguish genuine transfer to unseen layouts from interpolation around near-identical training cases (Kapoor & Narayanan, 2023).

We address both gaps with the *Local–Global Operator* (LGO), which augments a neural operator with local boundary information while retaining its global room representation. We also introduce a separation-controlled protocol for evaluating transfer across diffuser layouts. Our contributions are:

1. **A general local-geometry module** that provides each query point with nearby boundary geometry in relative coordinates, stratified by boundary type so that important flow-driving surfaces remain represented at a fixed neighbour count (§4).
2. **Controlled ablation across different neural operators.** A zero-initialised projection makes the augmented model initially reproduce the base model, enabling direct comparison with and without local boundary information. The same module is tested on three distinct neural-operator architectures (§4.4).
3. **A separation-controlled evaluation protocol** based on a dimensionless layout distance, with the threshold determined where a nearest-layout retrieval baseline loses predictive skill rather than selected heuristically (§3.1).

On held-out layouts, LGO improves full-field velocity R^2 for all three neural operators under both splits and remains more stable as geometric separation from the training set increases. No consistent improvement is observed in the occupied zone; this limitation is discussed in §7 and Appendix E. Appendices A–F are provided in the supplementary material.

2 Related Work

Neural operators for geometry-conditioned fields. DeepONet represents the solution through a branch–trunk decomposition, with the branch encoding the input function and the trunk evaluating the field at query coordinates (Lu et al., 2021). Fourier Neural Operators learn global integral operators through spectral convolutions on regular grids (Li et al., 2021; Kovachki et al., 2023), while graph-based formulations extend operator learning to irregular discretisations (Li et al., 2020). Geometry-informed operators further incorporate domain geometry to handle varying shapes (Li et al., 2023), and physics-informed neural networks impose governing equations through residual terms in the training loss (Raissi et al., 2019). These methods provide strong global representations, but moving diffusers introduce a more local requirement. The flow at an interior point can depend strongly on its position relative to small supply or return patches. Increasing global capacity through Fourier features (Tancik et al., 2020), finer latent representations, or richer operator kernels does not explicitly provide this query-relative boundary information. This motivates combining global operator context with direct local geometric encoding.

Local geometric learning on point clouds and meshes. Point-cloud methods learn directly from irregular geometric data. PointNet combines per-point encoding with symmetric aggregation (Qi et al., 2017a; Zaheer et al., 2017), while PointNet++ introduces hierarchical local neighbourhoods (Qi et al., 2017b). Kernel-point convolutions (Thomas et al., 2019) and point-transformer models (Zhao et al., 2021) further capture local interactions using relative spatial information, and related equivariant architectures extend these ideas to transformations of the input geometry (Cohen & Welling, 2016). For numerical simulation, mesh-based graph networks propagate information between connected nodes (Pfaff et al., 2021), while more recent operator models combine global-,

graph- and point-level representations (Wang et al., 2025). These methods provide strong local geometric encoding, but two limitations remain for ventilated enclosures. First, distance-based neighbourhoods may be dominated by large wall surfaces, while smaller supply and return patches that directly impose the flow can be under-represented at a fixed neighbour count. Second, purely local aggregation provides limited enclosure-scale context unless information is propagated over multiple layers or combined with a global representation. This motivates a boundary-aware local representation coupled with the global context of a neural operator.

Learned surrogates for indoor airflow. Machine-learning surrogates have been increasingly used to reduce the cost of CFD-based indoor airflow prediction (Chen et al., 2025). Operator-learning models have been applied to ventilation prediction and optimisation (Bian et al., 2026), while physics-informed approaches have been explored for indoor airflow prediction and temporal extrapolation (Wei & Ooka, 2025). Most studies, however, vary operating conditions such as supply rate, vent angle, or thermal load within a fixed or restricted geometry. Generalisation across different diffuser locations remains less studied. This is important because moving a diffuser changes the spatial relation between each prediction point and the flow-driving boundary. A surrogate for layout exploration must therefore learn how airflow changes with boundary placement, rather than only interpolate among operating conditions for a fixed configuration.

Evaluating generalisation across layouts. Random train–test splits can overestimate generalisation when test samples are very similar to training cases (Kapoor & Narayanan, 2023). Other machine-learning fields address this by defining splits from domain-specific similarity measures (Wu et al., 2018) and by comparing learned models with simple baselines (Ferrari Dacrema et al., 2019). The same issue arises for diffuser layouts generated from a low-dimensional design space, where nominally different configurations may remain geometrically close. A random split can therefore confound transfer to new layouts with interpolation around nearby training cases. The key question is how much separation is sufficient. We determine this empirically using a nearest-layout retrieval baseline and set the threshold where the closest training layout no longer provides a reliable prediction.

Positioning. LGO combines global operator learning with explicit local boundary geometry. The base neural operator provides room-scale context, while the added module supplies each query point with relative information about nearby flow-driving boundaries. A zero-initialised residual projection, following established practice (Bachlechner et al., 2021; Zhang et al., 2023), makes the augmented model initially reproduce the base model and enables controlled ablation. Attention-based operators (Hao et al., 2023; Wu et al., 2024) may already capture similar local interactions through dense query–boundary attention. Our claim is therefore limited to operator architectures that do not otherwise provide explicit query-relative boundary information.

3 Problem Setting and Dataset

We seek an operator $\mathcal{G}$ carrying a room description – boundary geometry and boundary conditions – to a field $\mathbf{x} \mapsto (u, v, w, p_{\mathrm{rgh}}, T)(\mathbf{x})$ evaluable at any interior coordinate. Queries are points rather than a grid, so the model is resolution-independent and is scored on the CFD's own $\approx 5.7 \times 10^5$ cell centres per case, with no interpolation. Because 8.3% of the volume carries 81.4% of the velocity variance, any aggregate full-field metric is a near-jet metric in disguise, which is why the ASHRAE Standard 55 occupied zone (de Dear & Brager, 2002) is reported separately throughout and never inferred from a full-field score.

We generate 193 quality-controlled steady-state cases in a single room of $8.80 \times 6.10 \times 3.20$ m, carrying three supply diffusers and three return grilles, each $w = 0.600$ m square. Their positions are the only variable: envelope, furniture, flow rate and temperature, wall conditions, solver, mesh and iteration protocol are identical across cases, only the ceiling and six vent patches are rebuilt per layout, and the remaining 18 boundary patches are copied byte-identically. That is what licenses attributing a measured difference between cases to the layout rather than to the pipeline – and it is equally the principal limit on scope: one room topology, one vent count. Every case has exactly three supply and three return panels in a fixed enclosure, so what is demonstrated is generalisation across layouts, not across rooms, scales or vent counts (Appendix D). Layouts come from a scram-

bled Sobol sequence (Sobol', 1967), widened by 66 layouts over the full feasible region chosen by farthest-point search. Cases are solved with a steady buoyant RANS solver (Weller et al., 1998) using k–ε turbulence with wall functions (Launder & Spalding, 1974), under an export-time quality gate rather than a report afterwards (Appendix A).

These are not converged steady states, and that bounds one of the two regions we report. Two snapshots of the same case, at iterations 1500 and 3000, correlate at $\rho = +0.166$ across the ASHRAE band against $+0.932$ full-field: in the comfort band, changing the iteration at which the solve is stopped moves the field more than changing the layout does. Roughly half of what a model learns there is solver trajectory rather than room physics, and we carry this as a ceiling on every comfort-band number we report (Appendix C).

3.1 Separation-Controlled Splits

Layout distance is the mean Hungarian-matched panel displacement (Kuhn, 1955), supplies to supplies and returns to returns, reading directly as how far the vents moved. We report it non-dimensionally, as a fraction of the vent panel width $w = 0.600$ m,

$$\delta \;=\; \frac{1}{6}\sum_{k=1}^{6}\left\|\mathbf{c}_k - \mathbf{c}'_{\pi(k)}\right\| \Big/\; w, \tag{1}$$

where $\mathbf{c}_k$ are the six vent centres of one layout, $\mathbf{c}'$ those of the other and π the optimal within-class matching. The panel width is the right scale because it is the length over which the boundary condition itself changes: $\delta = 1$ means a held-out diffuser sits entirely off the footprint of the training diffuser matched to it, leaving face overlap $1 - \delta$ along the direction of motion. Cases closer than a threshold are joined by an edge and each whole connected component is assigned to one split, so any two cases in different splits are at least the threshold apart and none is discarded into a buffer zone.

The threshold is measured, not chosen – and what it measures is $\delta = 1$. When a nearest-layout baseline is forced to retrieve only from training layouts at least δ away, its skill degrades monotonically and crosses below that of predicting the training mean at one panel width (Appendix A). The collapse distance is $\delta = 1$, which is what a geometric reading predicts and is why the threshold needs no reference to this room's size.

We therefore build two splits, at $\delta \approx 0.50$ and $\delta \approx 0.97$ (equivalently 0.30 and 0.58 m, the names the released manifests carry), leaving about 50% and 3% of a held-out diffuser face over training ground. On the wider split a lookup table has nothing left to copy: retrieval scores zero in the comfort band and is negative on the full field, so any skill there is skill a lookup cannot reproduce. Retrieval is also strongly regional, and the near field is where it dies first – near-jet retrieval R^2 is -0.056 at $\delta \approx 0.50$ and -0.243 at $\delta \approx 0.97$, worse at both than predicting the mean. The two splits are therefore not interchangeable: $\delta \approx 0.97$ carries the generalisation claim, its held-out layouts being geometrically extreme by construction, while $\delta \approx 0.50$ is the realistic-layout check, the only held-out set drawn from layouts a designer would install. They share training layouts, so results on them are not statistically independent.

4 Method: The Local–Global Operator

The Local–Global Operator is a small module that runs beside an existing neural operator and answers one question the host cannot: what does the boundary look like from where this query is standing? For each query point it retrieves the nearest boundary points of each patch type separately – supplies, returns, solid, leak – so the surfaces that drive the flow are always among them. Each retrieved point becomes a token built from relative quantities only: the offset from the query to that point, its distance, the surface normal there, and what kind of surface it is. No absolute coordinate enters. One cross-attention layer lets the query read its own neighbour set, and the result is added into the host's per-query state through a projection initialised at zero, so at step 0 the augmented model computes exactly what the host computed. Training moves it away from zero; what it learns is a correction that the host had no channel to express. The subsections take these pieces in turn.

4.1 WHY TWO STREAMS

The argument is a measurement, not an intuition. The 8.3% of the volume within 1 m of a diffuser carries 81.4% of the velocity variance, and there the jet's direction, spreading angle and decay are properties of the source rather than of where it sits – so relative-coordinate locality is the right inductive bias. The far field inverts every one of those properties, is set by the shape of the whole enclosure, and is not recoverable from a neighbourhood at any affordable neighbour count. Hence two streams. A global stream encodes the whole boundary cloud into a room latent each query interrogates: permutation-invariant (Zaheer et al., 2017), computed once per case, and the only route by which room shape reaches a query. A local stream attends to per-query neighbour sets in relative coordinates: translation-equivariant, resolution-independent, and the only route by which a jet's offset from its source reaches a query.

4.2 STRATIFIED NEIGHBOUR GATHER

For each query $\mathbf{x}$ we retrieve the K nearest boundary points of each patch class independently – $K = 32$ supply, 16 return, 32 solid, 8 leak, so 88 per query. Stratification is load-bearing, not cosmetic: plain k-nearest-neighbour retrieval for a mid-room query returns only walls, the solid cloud outnumbering a supply panel by more than an order of magnitude, so the vents that cause the flow would be invisible. Per-type sets guarantee every query the nearest vent of each type, at whatever distance, always in relative coordinates.

It is equally important to state what stratification does not buy. A supply panel holds 2209 boundary points at about 1.3 cm spacing, so the 32 nearest supply points of any query lie on one panel; a second would need $K > 2209$. The arrangement of the other panels reaches a query only through the global stream. K is therefore a resolution parameter on the nearest surface of a type, not a coverage parameter: $K = 32$ spans 0.15 m of a 0.600 m panel face, and even $K = 512$ spans only 0.58 m and still reaches one panel of three. Cost is linear in K. The sweep of Appendix E confirms the reading: only the supply group is decisive, and K on it is flat from 8 to 128, which is what a resolution reading predicts and a coverage reading would not.

4.3 RELATIVE-ONLY ENCODING AND LOCAL ATTENTION

Each neighbour $\mathbf{p}$ becomes one token built from relative quantities only. With $\mathbf{d} = \mathbf{p} - \mathbf{x}$ the offset in metres and $\mathbf{n}$ the boundary normal at $\mathbf{p}$,

$$\mathbf{t}(\mathbf{p}, \mathbf{x}) = \mathrm{MLP}\Big(\mathbf{d},\ \log \|\mathbf{d}\|,\ \hat{\mathbf{d}},\ \mathbf{n} \cdot \hat{\mathbf{d}},\ \mathrm{RBF}_{12}(\|\mathbf{d}\|),\ \mathbf{f}_{\mathbf{p}},\ \mathbf{g}_{\mathbf{p}}\Big) \in \mathbb{R}^{64}, \tag{2}$$

where RBF_{12} is a bank of 12 Gaussians log-spaced from 0.02 m to 4.0 m, $\mathbf{f}_{\mathbf{p}}$ collects the point's normal, blinded boundary velocity, blinded boundary temperature and leak magnitude, and $\mathbf{g}_{\mathbf{p}}$ is the patch-group one-hot. The logarithm is taken of $\|\mathbf{d}\| + 10^{-3}$ m and $\hat{\mathbf{d}}$ divides by $\max(\|\mathbf{d}\|, 10^{-6})$, so a token stays finite even for a query on the boundary. Neither guard is needed for finiteness here: queries are interior cell centres and neighbours boundary face samples, two disjoint sets, so $\|\mathbf{d}\|$ is bounded below by the near-wall cell height.

No absolute coordinate enters Equation 2. This makes the local stream translation-equivariant by construction, in the strict sense that the property follows from the functional form and therefore holds at initialisation and after any amount of training. We measure the residual under a joint translation of room and query at 1.2×10^{-7}, float32 round-off, against 3.9×10^{-3} when the branch is instead given absolute coordinates. Tokens are deliberately narrow at 64 channels, for memory rather than taste: every token tensor is per (query, neighbour) pair, so at 5120 queries and 88 neighbours it carries 450k rows. Each query then attends to its own neighbour set, batched per query and never global over the cloud, through one cross-attention layer (Vaswani et al., 2017) with 8 heads at width 64. The query stream is initialised at zero, so the layer-0 query is a learned constant carrying no coordinate at all. Figure 1 shows the complete module.

4.4 ZERO-INITIALISED INJECTION, AND THE THREE HOSTS

The branch's output is projected to the host's own width d_{model} through a zero-initialised linear layer, and two consequences follow. First, d_{model} is read from the host, never configured. The

Table 1: The injection contract: the identical module enters each host at one per-query point, before that host's own shared output head. Only Host B reads no absolute coordinate end to end, so it is the only arm translation-equivariant as a whole rather than in its local stream alone.

Host	Mechanism	Injection point	Added params	Abs. coord.
A	Geom-DeepONet, rank 512	before the shared linear head	+87,040 (+1.5%)	Fourier($\mathbf{x}$)
B	Cross-attention trunk, d=256	query seed, and before the head	+54,016 (+1.0%)	none
C	Graph kernel + FNO, d=64	before the shared projection MLP	+29,248 (+0.55%)	latent grid

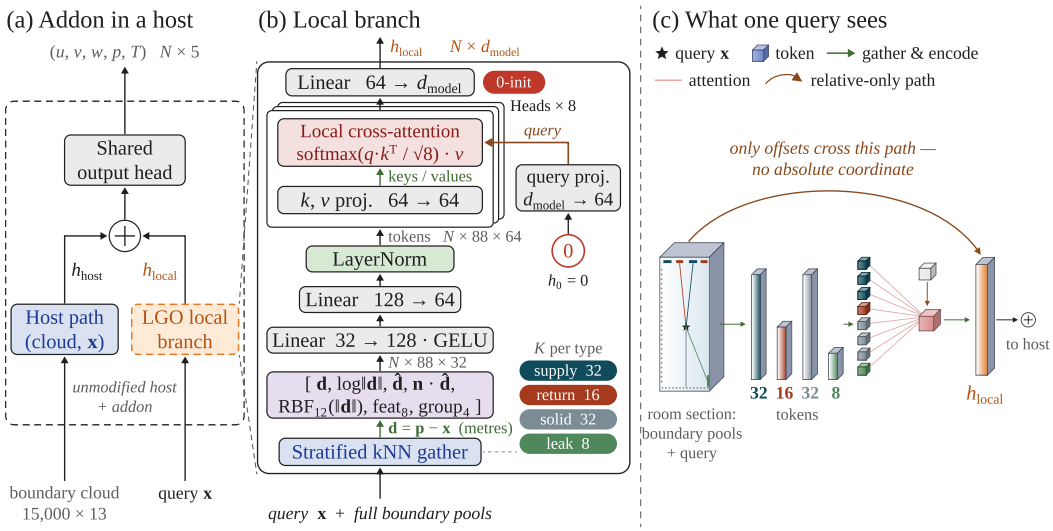

Figure 1: The LGO local branch. (a) Its output h_{local} is added to an unmodified host's per-query state before the host's own output head. (b) Stratified gather, relative-only tokens (Equation 2) and one cross-attention layer queried from a zero state, returned through a zero-initialised projection. (c) One query's view: 88 boundary points enter only as offsets.

branch therefore has no output width of its own and no prediction head of its own: this single projection is its only interface, and it emits a correction into the host's representation rather than a field. That is why the added parameter count spans 0.55 to 1.5% across the three hosts (Table 1) – the projection is $64 \times d_{\text{model}}$, so the cost is a property of the host, not of the add-on, and one arm's delta must not be quoted as "the cost of the branch".

Second, because the output projection starts at zero, $h_{\text{local}} \equiv 0$ at initialisation, so host plus branch is bit-identical to host alone before any training. The branch-less arms are therefore not matched baselines but the same model with one module removed, which makes all three ablation pairs exact rather than approximate. Unit tests assert the bit-identity rather than trusting it, and the parameter accounting is exact: $5{,}888{,}438 = 5{,}801{,}398 + 87{,}040$ for the modal host, and likewise for the other two. All seven arms sit within $\pm 6\%$ of one another in parameter count, from 5.36 M to 5.99 M, so no comparison reported here is a capacity artefact.

5 Experimental Protocol

Every arm is trained for 450 epochs, leak-free, at its own tuned learning rate, each a bracketed interior optimum of an identical grid, so both members of every ablation pair receive the same search; the objective and schedule are given in Appendix B.

Metrics are computed per case and then averaged over cases: pattern correlation ρ, amplitude ratio s, the Taylor skill score (Taylor, 2001), R^2 and dimensional errors. R^2 is reported but never used for selection: for a centred field $R^2 = \rho^2 - (\rho - s)^2 - \text{bias}^2$ is maximised at $s = \rho$, i.e. by under-

Table 2: Held-out test, evaluated once (seed 0; 21 and 25 cases). The near-jet column is where retrieval has no skill, the ASHRAE column where it has too much (Appendix E).

	$\delta \approx 0.50$ (realistic layouts)			$\delta \approx 0.97$ (generalisation)		
Arm	full v R^2	near-jet R^2	ASHRAE ρ	full v R^2	near-jet R^2	ASHRAE ρ
Host A + LGO	$+0.975$	$+0.995$	$+0.783$	$+0.926$	$+0.967$	$+0.666$
Host B + LGO	$\mathbf{+0.978}$	$+0.996$	$+0.785$	$+0.936$	$+0.974$	$+0.661$
Host C + LGO	$+0.971$	$\mathbf{+0.996}$	$+0.734$	$\mathbf{+0.942}$	$\mathbf{+0.979}$	$+0.670$
Host A (Geom-DeepONet)	$+0.641$	$+0.624$	$+0.790$	$+0.265$	$+0.241$	$+0.662$
Host B (cross-attention)	$+0.673$	$+0.667$	$\mathbf{+0.800}$	$+0.347$	$+0.338$	$+0.639$
Host C (graph kernel)	$+0.920$	$+0.942$	$+0.732$	$+0.644$	$+0.627$	$\mathbf{+0.693}$
Slice attention	$+0.967$	$+0.992$	$+0.727$	$+0.582$	$+0.584$	$+0.565$
Nearest-layout retrieval	$+0.003$	-0.079	*+0.729*	-0.081	-0.161	*+0.611*
Training mean	-0.001	-0.005	$+0.000$	-0.003	-0.010	$+0.000$

predicting whenever correlation is low, so ranking on it rewards damping. Checkpoints are selected on the Taylor score, in the form

$$S = \frac{2\,(1+\rho)}{\left(s + s^{-1}\right)^2}, \qquad S_{\text{joint}} = \tfrac{1}{2}\left(S_{\mathbf{v}} + S_T\right), \tag{3}$$

the mean of the velocity and temperature scores, maximised only at $\rho = 1$ and $s = 1$ and so not purchasable by damping; we fix the form because Taylor (2001) defines a family of such scores and the choice changes which checkpoint is selected. Two orthogonal regions are reported and must not be confused: distance bands from the nearest supply vent, which are not the occupied volume and on which no comfort claim may rest, and the ASHRAE Standard 55 occupied zone (Appendix C).

Seed noise here is arm-dependent by up to three orders of magnitude, so no single threshold exists and one was fixed in advance instead: every headline effect is reported against both its case standard error and the seed spread, checked against the larger of the two arms' spreads, and an effect smaller than that spread is null however many standard errors it clears (Appendix C). Both splits are seed-replicated over seeds $0/1/2$; paired over three seeds, significance requires $|t| > 4.30$.

Validation, test, and what each may decide. Checkpoint selection, learning-rate tuning and every ablation are decided on validation, so each seed spread we quote is a lower bound on true seed variance. To keep the headline free of that optimisation bias, a test set on each split – sharing no layout and no connected component with training or validation – was evaluated exactly once, over all arms and both baselines, after the claim had been recorded in writing. That pass is seed 0 only, which is its cost; the test sets are spent and are not reused.

Verifying the external harness. Running on someone else's benchmark tests our export, metric and training loop as much as it tests the branch, so we first tried to reproduce a published number end to end. Transolver (Wu et al., 2024) was chosen because its code and full configuration are released, where the GINO number we sit beside is a re-implementation whose code is not; landing near its published 0.0207 would have verified all three at once. It did not land: 0.1577 at the published budget, $7.6\times$ off and worse than a constant-field null. The architecture is not the cause – our implementation reads line by line against the released file, identical but for a clamp on the slice temperature – so the gap is the data contract, which Appendix F enumerates. This is load-bearing rather than embarrassing: our GINO also sits above its published number, and a second, architecturally verified model reproducing the same direction shows the offset belongs to the contract rather than to GINO. Accordingly no Transolver-based result enters our comparison.

6 Results

Seven arms on both splits – three hosts, three host-plus-branch pairs, and a slice-attention operator (Wu et al., 2024) in no pair – against both zero-parameter baselines (Appendix E).

Table 3: Three parameter-exact ablations, each row removing one module from one host. Test columns are from Table 2; validation columns are means over seeds $0/1/2$ at $\delta \approx 0.97$, paired over seeds. Near-jet R^2 moves with full-field throughout: $+0.862$, $+0.794$, $+0.317$ at $t = 14.9$, 11.6, 16.6.

		Δ full v R^2, test		$\delta \approx 0.97$ validation, 3 seeds			
Pair	added params	$\delta \approx 0.50$	$\delta \approx 0.97$	full v R^2	t	T MAE [K]	t
Host A $\rightarrow$ +LGO	87,040 (+1.5%)	+0.335	+0.661	+0.734	13.9	−0.050	−16.4
Host B $\rightarrow$ +LGO	54,016 (+1.0%)	+0.305	+0.589	+0.690	12.5	−0.059	−19.8
Host C $\rightarrow$ +LGO	29,248 (+0.55%)	+0.051	+0.298	+0.276	20.6	−0.087	−15.7

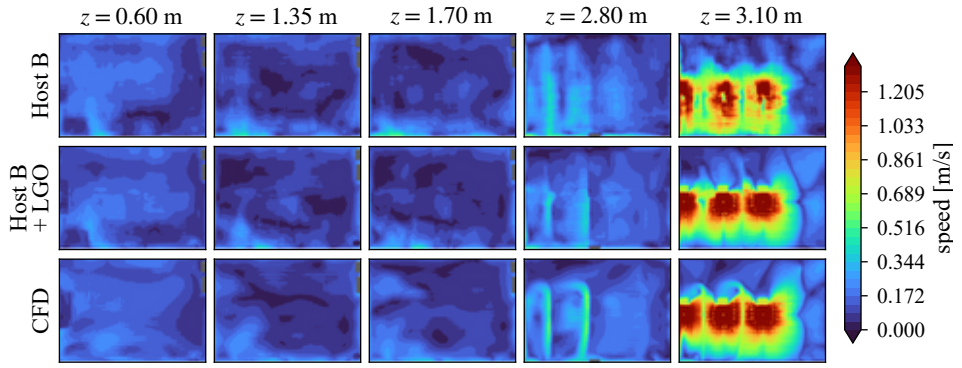

Figure 2: Host B without and with the local branch, against CFD, on the median held-out test case of the $\delta \approx 0.50$ split (layout D010, seed 0): speed at all 563,858 interior cell centres, on the CFD colour scale.

6.1 The Branch, on Three Hosts

The claim concerns the branch contribution, evaluated through three exact ablations (Table 3), rather than an architecture-level comparison. All six model–split combinations improve both full-field and near-jet R^2, with no validation case degraded on either metric. The improvement is three to fifteen times larger under the widely separated split, consistent with stronger gains under more demanding generalisation. Global temperature error also decreases for every host architecture and random seed, providing the most consistent result without requiring a noise-related qualification.

Host C is the test that could have failed, and did not. Not a trunk decoder, and already owning a partial route to local geometry, it should gain least – and does, the smallest add-on still worth $+0.298$ on test. A null here would have falsified the account rather than merely weakened it.

6.2 The Branch Makes the Score Transfer

The result visible only with a held-out set is not the size of the gain but its stability. From validation to unseen layouts at $\delta \approx 0.50$ the augmented models move by -0.002 to -0.003 full-field velocity R^2 while the branch-less hosts lose 0.070 and 0.112 – twenty-five to fifty times more. At $\delta \approx 0.97$ they improve on test where two of three hosts fall, and five of six test deltas exceed their validation counterparts, so the validation numbers are conservative. The branch does not only raise the score; it makes the score survive a change of layout. Figure 2 shows it as a field: the host alone smears the supply jets into a diffuse plume, the augmented one recovers the jet cores and their attachment to the ceiling, at $z = 2.80$ and 3.10 m.

A six-cell sweep at constant parameter count isolates where the skill comes from (Appendix E). Dropping the supply group costs 0.101 full-field velocity R^2, losing $28/28$ cases; the leak group measures as nothing; and K on supply is flat from 8 to 128, confirming §4 – K buys face resolution the model does not need, not the coverage it does.

Table 4: ShapeNet-Car, volume velocity, relative L^2 (lower is better), at the published fold 0, 200 epochs and lr 10^{-3}. A zero-parameter training-set mean scores 0.0939; the published GINO, which also receives surface normals and a joint velocity-and-pressure loss, reaches 0.0386 (Appendix F).

Arm	no distance channel	+ distance channel	Δ
Graph-kernel host	0.0899	0.0792	-11.9%
+ LGO	$\mathbf{0.0532}$	$\mathbf{0.0515}$	-3.2%

6.3 An External Benchmark, and What the Branch Adds Over an Explicit Input

The dataset above is ours, so the same ablation was run on data we did not build: ShapeNet-Car, volume velocity in relative L^2 on the graph-kernel host at the published fold, rate and budget. It also supplies a per-point distance to the nearest boundary that every published model on it consumes and our hosts do not – explicit geometric information, free to the model, and so the sharpest alternative to the branch – so we ran the full 2×2 (Table 4).

The branch cuts relative L^2 by 40.8% for $+0.55\%$ parameters, at an amplitude ratio of 100.2% – the field matched, not damped toward the mean. The columns together say more than the pair alone. The channel has real value to a host without the branch, which gains 11.9%, so it is not worthless here. It does nothing for the augmented model: -3.2% on this fold, $+1.2\%$ on the other at its tuned rate – a flipped sign, and so noise by this paper's own rule. That is a mechanism claim rather than a score: the branch has already extracted what the distance channel provides, as a relative-offset gather should, distance being derivable from the offsets it reads. Nor is it standing in for the missing input – given the channel, the branch-less host still reads 0.0792 against 0.0515. Every cell is seed 0, so the argument rests on that flipped sign rather than on the size of the -3.2%; the pair is the controlled claim and the published 0.0386 a weaker, unmatched external comparison (Appendix F).

7 Conclusion

This work shows that explicit local boundary geometry can improve neural-operator generalisation across changing diffuser layouts. The proposed Local–Global Operator (LGO) uses patch-stratified neighbourhoods to preserve flow-driving surfaces and relative offsets to represent local geometry. Across three neural-operator architectures and two separation-controlled splits, LGO improves held-out full-field velocity R^2 by $+0.051$ to $+0.661$ with only 0.55–1.5% more parameters. At $\delta \approx 0.50$, augmented models vary by at most 0.003 between validation and unseen layouts, while base models lose up to 0.112. Ablations show that the gain is driven mainly by supply-boundary information. On ShapeNet-Car, LGO reduces relative L^2 error by 40.8%, while explicit boundary-distance input adds no measurable benefit to the augmented model. Overall, LGO provides a simple way to complement global neural operators with local geometric information for geometry-dependent flow prediction.

Limitations. We claim nothing in the ASHRAE occupied zone: its self-correlation is $\rho = +0.166$ and a zero-parameter lookup scores inside the arms' own spread, so it cannot arbitrate architecture. Of three physics diagnostics only one favours the branch – near-wall fidelity, $3/3$ pairs on both splits – while wall compliance is mixed and the divergence excess is architecture-independent, shared by every arm including the baselines. Which host wins reverses between splits, the second reason we name no headline model. And the branch is cheap in parameters and expensive in time: a full field takes 11.3–13.3 s, 126 to $149\times$ faster than the 28-minute mean CFD solve, yet the branch adds ≈ 1.7 s per 100k queries and surrenders 5 to $45\times$ of its host's own speed-up, so the $+0.55\%$ parameter figure must never be quoted alone. Four limits bound the scope: one room topology, an untested vent count, no local message-passing or multiscale-operator baseline (Pfaff et al., 2021; Wang et al., 2025), and test sets spent in a single pass at one seed (Appendix D, Appendix E). The next step is the one this design was built for: vary the boundary condition rather than the boundary.

AI USE STATEMENT

We used generative AI tools for drafting and editing prose in this manuscript, and for assistance with analysis and plotting code. We did not use them to generate research ideas, to produce or modify any experimental measurement, or to write any numeric value reported here: every number in this paper is read from a logged experimental record. We have reviewed all AI-assisted work, and all claims, figures and tables were checked against the underlying result files by the authors. We take responsibility for the final content of this work.

REPRODUCIBILITY STATEMENT

Everything needed to reproduce this work is specified in the paper or in artefacts that accompany it; we note here where each part lives rather than restating it.

Dataset. Section 3 and Appendix A give the room, the layout parameterisation, the design tiers, the flow configuration and the solver settings. Layouts are generated deterministically from a fixed Sobol seed, so the design of experiments is reproducible without storing it. Two machine-readable manifests accompany the dataset: a layout manifest recording, for every designed layout, its tier, its parameters, its six vent centres, its feasibility status, its nearest other layout and that distance, and its split assignments; and a case manifest recording point counts, patch counts, the full quality-control record and a checksum of the field data. The export-time gate and its thresholds are in Appendix A, including the two rejected layouts, retained on disk as regression fixtures for the gate itself.

Splits. Section 3.1 defines the layout distance and the connected-component construction. Assignments are frozen in a manifest rather than recomputed, so a growing dataset cannot silently re-deal them.

Model and training. Section 4 specifies the local branch completely, including the per-class gather sizes and the token construction of Equation 2. Section 4.4 specifies each host, its single injection point and the exact added parameter counts (Table 1); the bit-identity of host-plus-branch with the host at initialisation is asserted by unit tests rather than assumed, which is what makes each pair an exact ablation. Architecture, training and normalisation configurations are recorded alongside checkpoint checksums. The return-velocity blinding is standard on every reported run, and the evaluator reads the flag back from the run configuration, so a blinded model cannot be scored unblinded.

Evaluation. Section 5 and Appendix C give the metric definitions, the two scoring regions and why they are not substitutes, the checkpoint-selection criterion of Equation 3, and the seed protocol with the decision rule fixed in advance. Both zero-parameter baselines are evaluated through the same metric code, on the same point sets, as the models.

Validation and test. Every design decision, hyperparameter and checkpoint in this paper was selected on validation sets. The held-out test sets were evaluated in a single pass, at one seed, after the claim was recorded in writing, and are reported in Table 2; they have not been reused, and no number in this paper was chosen with reference to them. Seed spreads are lower bounds, because checkpoints are selected on the same held-out set the seeds are scored on.

Artefacts. Source code, the training and evaluation harness, the CFD case generator with its mesh and boundary-condition dictionaries, the layout and case manifests, the frozen split assignments and the per-arm run configurations are available in an anonymised repository, `https://anonymous.4open.science/r/lgo-indoor-airflow-5252/`, which also gives the persistent identifier of the dataset. Anything in Appendix A and Appendix B that is specified by a configuration file rather than by prose – the furniture and corner-column geometry, the volume-meshing parameters, and the fitted diffuser profile – is resolvable there.

ETHICS STATEMENT

This work raises no human-subjects, privacy or personal-data concerns. The dataset is entirely synthetic: steady-state CFD solutions in a single unoccupied room with no occupants modelled, no measured building data and no information about any real person, site or organisation. No human subjects were involved at any stage and no institutional review was required.

Two points are worth stating rather than assuming. First, the application domain is ventilation design, where a surrogate that is wrong in the occupied zone could inform a decision affecting the comfort or health of building occupants. This is exactly why we decline a comfort-band claim and report the solver-noise ceiling and the lookup baseline beside every comfort-band number: the honest reading of our evidence is that the model is not yet validated for comfort assessment, and we would consider it misused if deployed as though it were. Second, the surrogate is trained against a particular RANS configuration and inherits its modelling assumptions and its unconverged numerics; it approximates that solver, not reality, and Appendix A records the gap.

REFERENCES

Thomas C. Bachlechner, Bodhisattwa Prasad Majumder, Huanru Henry Mao, Garrison W. Cottrell, and Julian McAuley. ReZero is all you need: Fast convergence at large depth. In *Proceedings of the 37th Conference on Uncertainty in Artificial Intelligence (UAI)*, volume 161 of *PMLR*, pp. 1352–1361, 2021.

Yuexin Bian, Oliver Schmidt, and Yuanyuan Shi. Operator learning for energy-efficient building ventilation control with computational fluid dynamics simulation of a real-world classroom. *Applied Energy*, 404:127035, 2026. doi: 10.1016/j.apenergy.2025.127035.

Tianle Chen, Ruibin Li, Xuesong Hu, Bingyang Zhang, Yuanyuan Liu, Liangzhu (Leon) Wang, and Naiping Gao. Machine learning as CFD surrogate models for rapid prediction of building-related physical fields: A review of methods and state-of-the-art. *Building and Environment*, 285:113667, 2025. doi: 10.1016/j.buildenv.2025.113667.

Taco Cohen and Max Welling. Group equivariant convolutional networks. In *Proceedings of the 33rd International Conference on Machine Learning (ICML)*, volume 48 of *PMLR*, pp. 2990–2999, 2016.

Richard J. de Dear and Gail S. Brager. Thermal comfort in naturally ventilated buildings: Revisions to ASHRAE Standard 55. *Energy and Buildings*, 34(6):549–561, 2002. doi: 10.1016/s0378-7788(02)00005-1.

Maurizio Ferrari Dacrema, Paolo Cremonesi, and Dietmar Jannach. Are we really making much progress? A worrying analysis of recent neural recommendation approaches. In *Proceedings of the 13th ACM Conference on Recommender Systems (RecSys)*, pp. 101–109, 2019. doi: 10.1145/3298689.3347058.

Sara Gilani, Hamid Montazeri, and Bert Blocken. CFD simulation of stratified indoor environment in displacement ventilation: Validation and sensitivity analysis. *Building and Environment*, 95: 299–313, 2016. doi: 10.1016/j.buildenv.2015.09.010.

Iman Goldasteh, Goodarz Ahmadi, and Andrea Ferro. CFD simulation of particle transport and dispersion in indoor environment by human walking. In *Proceedings of the ASME 2012 Fluids Engineering Division Summer Meeting*, volume 1, pp. 267–272, 2012. doi: 10.1115/FEDSM2012-72149.

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 (ICML)*, volume 202 of *PMLR*, pp. 12556–12569, 2023.

Sayash Kapoor and Arvind Narayanan. Leakage and the reproducibility crisis in machine-learning-based science. *Patterns*, 4(9):100804, 2023. doi: 10.1016/j.patter.2023.100804.

Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew M. Stuart, and Anima Anandkumar. Neural operator: Learning maps between function spaces with applications to PDEs. *Journal of Machine Learning Research*, 24(89):1–97, 2023.

H. W. Kuhn. The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1–2):83–97, 1955. doi: 10.1002/nav.3800020109.

B. E. Launder and D. B. Spalding. The numerical computation of turbulent flows. *Computer Methods in Applied Mechanics and Engineering*, 3(2):269–289, 1974. doi: 10.1016/0045-7825(74)90029-2.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew M. Stuart, and Anima Anandkumar. Neural operator: Graph kernel network for partial differential equations. *arXiv preprint arXiv:2003.03485*, 2020. doi: 10.48550/arxiv.2003.03485.

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

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

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. doi: 10.1038/s42256-021-00302-5.

Tobias Pfaff, Meire Fortunato, Alvaro Sanchez-Gonzalez, and Peter W. Battaglia. Learning mesh-based simulation with graph networks. In *International Conference on Learning Representations (ICLR)*, 2021.

Charles R. Qi, Hao Su, Kaichun Mo, and Leonidas J. Guibas. PointNet: Deep learning on point sets for 3D classification and segmentation. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 77–85, 2017a. doi: 10.1109/cvpr.2017.16.

Charles R. Qi, Li Yi, Hao Su, and Leonidas Guibas. PointNet++: Deep hierarchical feature learning on point sets in a metric space. In *Advances in Neural Information Processing Systems 30 (NeurIPS)*, 2017b.

Maziar Raissi, Paris Perdikaris, and George Em Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019. doi: 10.1016/j.jcp.2018.10.045.

Martín Rodríguez-Vázquez, Iván Hernández-Pérez, Jesús Xamán, Yvonne Chávez, Miguel Gijón-Rivera, and Juan M. Belman-Flores. Coupling building energy simulation and computational fluid dynamics: An overview. *Journal of Building Physics*, 44(2):137–180, 2020. doi: 10.1177/1744259120901840.

I. M. Sobol'. On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Computational Mathematics and Mathematical Physics*, 7(4):86–112, 1967. doi: 10.1016/0041-5553(67)90144-9.

D. N. Sørensen and P. V. Nielsen. Quality control of computational fluid dynamics in indoor environments. *Indoor Air*, 13(1):2–17, 2003. doi: 10.1111/j.1600-0668.2003.00170.x.

Matthew Tancik, Pratul P. Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan T. Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. In *Advances in Neural Information Processing Systems 33 (NeurIPS)*, 2020.

Karl E. Taylor. Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres*, 106(D7):7183–7192, 2001. doi: 10.1029/2000jd900719.

Hugues Thomas, Charles R. Qi, Jean-Emmanuel Deschaud, Beatriz Marcotegui, Francois Goulette, and Leonidas Guibas. KPConv: Flexible and deformable convolution for point clouds. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 6410–6419, 2019. doi: 10.1109/iccv.2019.00651.

Wei Tian, Xu Han, Wangda Zuo, and Michael D. Sohn. Building energy simulation coupled with CFD for indoor environment: A critical review and recent applications. *Energy and Buildings*, 165:184–199, 2018. doi: 10.1016/j.enbuild.2018.01.046.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems 30 (NeurIPS)*, 2017.

Qinxuan Wang, Chuang Wang, Mingyu Zhang, Jingwei Sun, Peipei Yang, Shuo Tang, and Shiming Xiang. MNO: Multiscale neural operator for 3D computational fluid dynamics. *arXiv preprint arXiv:2510.16071*, 2025. doi: 10.48550/arxiv.2510.16071.

Chenghao Wei and Ryozo Ooka. Applying a physics-informed neural network to an indoor airflow time-extrapolation prediction. *Building and Environment*, 282:113246, 2025. doi: 10.1016/j.buildenv.2025.113246.

H. G. Weller, G. Tabor, H. Jasak, and C. Fureby. A tensorial approach to computational continuum mechanics using object-oriented techniques. *Computers in Physics*, 12(6):620–631, 1998. doi: 10.1063/1.168744.

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 (ICML)*, volume 235 of *PMLR*, pp. 53681–53705, 2024.

Zhenqin Wu, Bharath Ramsundar, Evan N. Feinberg, Joseph Gomes, Caleb Geniesse, Aneesh S. Pappu, Karl Leswing, and Vijay Pande. MoleculeNet: A benchmark for molecular machine learning. *Chemical Science*, 9(2):513–530, 2018. doi: 10.1039/c7sc02664a.

Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabás Póczos, Ruslan Salakhutdinov, and Alexander J. Smola. Deep sets. In *Advances in Neural Information Processing Systems 30 (NeurIPS)*, 2017.

Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 3813–3824, 2023. doi: 10.1109/iccv51070.2023.00355.

Hengshuang Zhao, Li Jiang, Jiaya Jia, Philip Torr, and Vladlen Koltun. Point transformer. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 16239–16248, 2021. doi: 10.1109/iccv48922.2021.01595.