

The Unified Geometric Continuous-Field Transformer: Learning Multi-Physics Engineering as Topologically Constrained Geometry

Sehaj Randhir Singh

Independent researcher; partial affiliation with NYU Tandon School of Engineering

Architecture analysis, code, and paper source: github.com/sehajr-singhs/ugct

Abstract—Models that reason over physics are typically built on token-based architectures: discrete symbols, pixels, or raw time series. We develop the alternative hypothesis that a physics transformer should process the manifold of physical spacetime itself, mapping telemetry, equations, and continuous fields into one unified tensor representation. We define the Unified Geometric Continuous-Field Transformer (UGCT), a four-tier architecture (foundational physics, topological consistency, mathematical engineering, real-world deployment) built from a multi-modal physical tokenizer, frame-averaged equivariant attention, a Fourier neural-operator backbone, and an integrated differentiable multi-physics loss $\mathcal{L} = \mathcal{L}_{\text{data}} + \lambda_1 \mathcal{L}_{\text{Maxwell}} + \lambda_2 \mathcal{L}_{\text{Navier-Stokes}}$. The core contribution is an adversarial design analysis in which each component is tested for failure modes, the failures patched, and the patches tested in turn. This exposes a principled sequence: the SE(3) curvature-memory explosion and frame discontinuity of its frame-averaged fix; FNO non-locality and the boundary-conservation leaks of domain decomposition; multi-physics gradient conflict and the null-space traps of gradient surgery; latent-bottleneck low-pass information loss; and the collapse of purely topological representations against dissipative physics. The resolution is a sheaf-theoretic Clifford latent space with dissipative Hodge decomposition, hyperblade grade-fading, symplectic spectral scrubbing, and non-commutative autodiff kernels, completed by a deployment tier (wavelet-encoded matrix sparsification, topological de-noising, Bayesian epistemic uncertainty) and a gray-box bootstrap (domain randomization plus PID feedback with underlying neural operators) that escapes the epistemic bootstrap paradox. We provide formal statements for the key complexity, conservation, continuity, and stability claims, a reference implementation of every component, and six controlled experiments that reproduce the headline failure modes — the frame snap, the gradient-surgery null-space trap, boundary leaks, the unmodeled-physics wall, and the low-pass bottleneck — and confirm that the proposed fixes eliminate them.

Index Terms—geometric deep learning; neural operators; equivariant attention; topological deep learning; Clifford algebras; physics-informed machine learning; multi-task optimization; domain randomization.

I. INTRODUCTION

A transformer that reasons about physics should not first translate physics into words. Token-based models represent a voltage trace, a stress tensor, and a differential equation as sequences of discrete symbols, then try to recover the underlying laws from symbol statistics. The alternative pursued here is a physics transformer whose substrate is the continuous

manifold of physical spacetime, so that symmetries, conservation laws, and dissipative structure are properties of the representation rather than facts to be rediscovered from data.

We call the architecture the *Unified Geometric Continuous-Field Transformer* (UGCT). It processes multi-modal engineering data — discrete telemetry, mathematical equations, continuous fields — in a single tensor representation, through four structural layers:

- 1) **Multi-modal physical tokenizer**: raw inputs (voltages V , strain tensors σ_{ij} , equations, sensor logs) are projected as continuous functions bounded by space (x, y, z) and time t via implicit neural representations (INRs); equations become structural weight matrices via graph-to-tensor encoders.
- 2) **Equivariant attention engine**: SE(3)-equivariant attention restricts weights so rotating an object rotates its physics vectors (torque, velocity, voltage) with it — symmetry and conservation live in the weights.
- 3) **Neural operator backbone**: a Fourier neural operator (FNO) performs attention in the frequency domain, learning the underlying PDEs themselves and becoming mesh-independent: trained on coarse lab simulations, it generalizes to atomic-scale telemetry without retraining.
- 4) **Differentiable simulation integration**: the loss is not next-token prediction but a multi-physics residual

$$\mathcal{L} = \mathcal{L}_{\text{data}} + \lambda_1 \mathcal{L}_{\text{Maxwell}} + \lambda_2 \mathcal{L}_{\text{Navier-Stokes}}, \quad (1)$$

so a voltage profile violating Maxwell’s equations or a shape violating fluid dynamics is penalized before it is output as a design.

The promise is cross-domain super-reasoning (a voltage spike predicts a microscopic stress fracture ten steps downstream) and zero-shot transfer (continuous-field understanding transfers from a battery pack to a hypersonic flight control system). Figure 1 shows the pipeline.

Contributions. (i) The four-tier UGCT specification; (ii) a formal failure analysis of each tier; (iii) an analysis of fixes and their residual flaws; (iv) the topological paradigm shift (sheaf-theoretic Clifford latent space with Hodge decomposition, grade-fading, symplectic scrubbing, non-commutative autodiff); (v) the deployment tier and gray-box bootstrap; (vi)

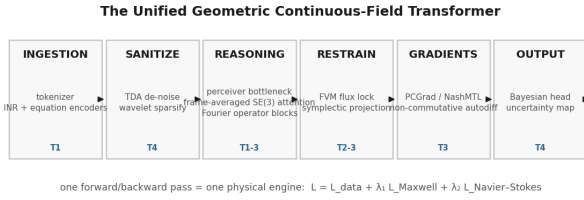


Fig. 1. The UGCT pipeline: ingestion, sanitization, reasoning, restraint, gradients, output — annotated with tiers T1–T4.

formal propositions for the central claims plus a reference implementation with reproducible demonstrations.

II. WHY THE NAIVE ARCHITECTURE FAILS

A. $SE(3)$ curvature-memory explosion

Equivariant models rely on spherical harmonics and Clebsch–Gordan tensor products. Memory grows quadratically with context length *and* geometric complexity.

Proposition II.1 (Steerable attention cost). *A steerable $SE(3)$ -attention layer over N tokens with geometric complexity K (irrep components up to order L , $K = O(L^4)$) costs*

$$\text{cost} = O(N^2 K^2), \quad (2)$$

because each attention pair forms tensor products of highest-weight irreps. Fully equivariant transformers thus scale quadratically in length and geometric complexity and crash GPU clusters on whole assemblies.

B. FNO non-locality

The FFT requires information from the entire field simultaneously, destroying GPU caching and multi-node scaling; it also assumes a structured grid, forcing expensive coordinate transforms on complex geometries.

Proposition II.2 (Global-exchange cost). *A global FFT over N cells costs $O(N \log N)$ flops but requires $O(N)$ all-to-all communication. Decomposing into D subdomains of size $n = N/D$ and applying local FFTs costs $O(n \log n)$ per subdomain with only boundary communication — provided boundaries are stitched consistently.*

C. Multi-physics gradient conflict

The residuals in (1) have very different gradient geometry; fluid-dynamics gradients drown out electromagnetic adjustments and the loss landscape becomes jagged.

Proposition II.3 (Destructive gradient interference). *Let $g_1 = \nabla \mathcal{L}_{\text{Maxwell}}$, $g_2 = \nabla \mathcal{L}_{\text{Navier-Stokes}}$. If $g_1 \cdot g_2 < 0$, the summed update $-\eta(g_1 + g_2)$ strictly increases one residual to first order; persistent conflict on shared (coupled) variables stalls descent.*

D. Data-integration paradox

Forcing unstructured continuous INRs to align with structured equation encodings in real time stalls preprocessing long before the forward pass.

Bottleneck	Fix	Residual flaw	Consequence
$SE(3)$ memory (Prop. II.1)	Frame averaging / vector neurons [1], [2]	Frame discontinuity (frames snap at symmetry crossings)	Discontinuous gradients, erratic spikes
FNO non-locality (Prop. II.2)	GINOT + domain decomposition	Boundary conservation leaks	Energy created/destroyed; numerical blowups
Gradient conflict (Prop. II.3)	PCGrad + NTK balancing [3]	Null-space traps on coupled variables	Optimizer “makes peace” by doing nothing
Data paradox	Perceiver bottleneck [4]	Low-pass information loss	Shock fronts and micro-cracks erased

TABLE I
FIRST ROUND OF THE ADVERSARIAL DESIGN ANALYSIS.

III. FIXES AND THEIR RESIDUAL FLAWS

Table I summarizes the first round of the dialectic: each fix introduces a new failure mode.

Frame discontinuity. Frame averaging [1] or vector neurons [2] reduce steerable $SE(3)$ to an invariant architecture at $O(N)$ per frame (linearizing Proposition II.1), but the canonical frame, selected by argmax over a symmetry orbit, snaps by 90° – 180° under smooth deformation.

Proposition III.1 (Frame discontinuity). *Let Φ be a frame-averaged embedding with a discrete argmax frame. There exist inputs $X_\varepsilon \rightarrow X_0$ at a symmetry crossing with $\|\Phi(X_\varepsilon) - \Phi(X_0)\| = \Omega(1)$: the latent map is discontinuous.*

Boundary leaks. Domain-decomposed local FFTs (Proposition II.2) rely on a GNN to stitch boundaries; heuristic blending does not enforce flux continuity.

Proposition III.2 (Conservation failure of heuristic stitching). *For k subdomains stitched heuristically, the per-step spurious source is $\sum_{\text{faces}} |\Phi_{\text{out}} - \Phi_{\text{in}}| > 0$ in general, and its cumulative value grows over the rollout. A stitcher enforcing $\Phi_{\text{out}} = \Phi_{\text{in}}$ at every face has identically zero spurious source.*

Null-space traps. PCGrad [3] projects conflicting gradients onto normal planes; for coupled variables (magnetohydrodynamics, electro-thermal expansion) the projections can vanish.

Proposition III.3 (Null-space trap). *Split coupled coordinates as $\theta = (\theta_{\parallel}, \theta_{\perp})$ with $g_1^{\parallel} = -g_2^{\parallel}$ (maximal conflict). PCGrad’s projected update has zero component on $\theta_{\parallel}$: the optimizer stops updating the coupled physics entirely.*

Information loss. A fixed-size Perceiver latent array [4] is a low-pass filter: high-frequency content (turbulent eddies, shock fronts, micro-cracks) — exactly what engineers need — is compressed away.

IV. SECOND-ORDER REMEDIES AND THEIR CONTRADICTIONS

Haar integration vs wavelet bypass. Replacing discrete frames with a continuous heat-kernel expectation over the

SE(3) orbit makes the latent trajectory smooth (repairing Proposition III.1), but integration is a low-pass filter: fine orientation is blurred, aliasing against high-frequency wavelet bypasses injected at deep layers; and differentiable Monte-Carlo sampling on the 6-D non-Euclidean manifold adds extreme gradient variance.

FVM flux locks vs INR mesh-independence. Hardcoding a finite-volume flux lock ($\sum \Phi_{\text{out}} = \sum \Phi_{\text{in}}$) into the GNN stitching backbone repairs Proposition III.2 (total energy cannot drift), but computing exact flux requires face normals and cell volumes — a mesh — destroying the mesh-independence of the INR/FNO tokenizer and reintroducing the grid-generation nightmare.

Nash bargaining vs local deadlock. Nash multi-task bargaining [5] replaces projection with cooperative game solving, avoiding the null-space of Proposition III.3:

Proposition IV.1 (Nash deadlock). *The Nash bargaining solution for the linearized problem is locally Pareto-optimal: no task improves without another regressing. Trajectories requiring a temporary penalty in one physics domain are rejected, deadlocking at a mediocre local equilibrium.*

Wavelet latent bypasses. Dyadic wavelet splitting routes coarse fields through the latent tokens and high-frequency residuals through sparse skip-connections into deep decoders, preserving sharp stress concentrations — but undermined by the Haar-blurred core (above).

V. THE TOPOLOGICAL PARADIGM SHIFT

The accumulating contradictions show that stacking patches on a coordinate-based backbone cannot work. The resolution: operate natively in a sheaf-theoretic Clifford algebra latent space, mapping physics as topological invariants on fiber bundles that survive deformation, rotation, and scale change.

A. The boundary layer

Real engineering physics is dissipative and irreversible; topological invariants measure what does *not* change under deformation. A purely topological sheaf struggles to capture entropy-increasing decay.

B. Grade explosion

Proposition V.1 (Grade explosion). *A Clifford algebra over $\mathbb{R}^n$ has dimension 2^n ; as multimodal streams expand n , the latent multivector space grows exponentially, reconstructing the hardware bottleneck.*

C. Cohomology vs backpropagation

Global consistency of stitched domains requires vanishing first sheaf cohomology ($H^1(\mathcal{U}, \mathcal{F}) = 0$), a global combinatorial check that standard local chain-rule backpropagation cannot optimize through.

Constraint	Interlock	Defense
Hyperblade aliasing	Continuous homotopic grade-fading ($\tau \in [0, 1]$)	No latent shockwaves; linear-sparse memory
Hodge spectral drift	Symplectic projection (QR onto $\ker d$)	Invariant baseline hard-locked
Non-commutative autodiff	Custom Clifford kernels (Triton/CUDA)	Parallel orientation tracking

TABLE II
THE THREE INTERLOCKS OF SECTION VI.

VI. THE THREE SOVEREIGN INTERLOCKS

Three final constraints — not bugs, but the toll physical reality extracts — each receive a structural interlock (Table II).

Grade-fading. Hyperblade sparsification (repairing Proposition V.1) dynamically masks high-grade products, but high-grade blades abruptly ignite at non-linear transitions (fracture, turbulence), and the sudden grade activation sends latent shockwaves downstream. The fix is a continuous homotopy parameter $\tau \in [0, 1]$ scaled by localized energy density.

Proposition VI.1 (Continuity of grade-fading). *With activations $w_g(\tau) = \tau^g$, $\tau = \sigma((e - e_0)/T)$, the latent trajectory is Lipschitz in energy density e with constant $O(1/(Te_0))$; a hard mask ($\tau \in \{0, 1\}$) is not Lipschitz at $e = e_0$.*

Spectral scrubbing. Hodge-rank decomposition isolates structural invariants in the harmonic subspace, but learned subspace boundaries let dissipative errors leak from the co-exact channel into the harmonic core over long rollouts — the model slowly forgets conserved structure.

Proposition VI.2 (Spectral drift bound). *Under the Hodge decomposition $f = f_{\text{ex}} + f_{\text{co}} + f_{\text{harm}}$, $\|f_{\text{harm}}^{(T)} - f_{\text{harm}}^{(0)}\| \leq \sum_t \|\Pi_{\text{harm}} f_{\text{co}}^{(t)}\|$. Periodic projection onto $\ker d$ (sparse QR) sets this leakage to zero at every checkpoint.*

Non-commutative autodiff. The Hodge-Laplacian over a Clifford space is non-commutative ($ab \neq ba$); standard autograd assumes a commutative Euclidean space, making the backward pass sequential and unparallelizable. The fix is custom Clifford-autodiff kernels: gradients become multivector bundles tracking magnitude *and* grade-orientation change, computed in parallel GPU/TPU hardware (Triton/C++ CUDA).

VII. DEPLOYMENT TIER

Three additions close the architecture for deployment (Figure 2). **Wavelet-encoded matrix sparsification (WEMS)** converts physical fields to sparse wavelet representations and computes heavy non-commutative products only on high-frequency interfaces, dropping dead uniform space. **TDA de-noising** uses persistent homology to filter transient sensor noise (low persistence) while retaining sharp anomalies (micro-cracks, shock fronts) — unlike Kalman/low-pass filters, which dull the signal. **Bayesian physics-informed ensembles**

The completed UGCT stack: four tiers, one self-defending architecture	
TIER 1	Foundational physics continuous spacetime manifolds · implicit neural representations · graph-to-tensor equation encoders
TIER 2	Topological consistency sheaf-theoretic Clifford algebras · localized FVM flux conservation · Nash bargaining optimization
TIER 3	Mathematical engineering continuous homotopic grade-fading · symplectic Lie-algebra projections · non-commutative autodiff kernels
TIER 4	Real-world deployment wavelet-encoded matrix sparsification · TDA de-noising filters · Bayesian epistemic uncertainty maps

Fig. 2. The completed four-tier stack.

replace point predictions with probability distributions over the Clifford latent space constrained by the hardcoded physics (1); every prediction ships an epistemic uncertainty map, giving regulators a verifiable safety guarantee.

VIII. THE EPISTEMIC BOOTSTRAP PARADOX AND GRAY-BOX RESOLUTION

The UGCT reasons with hardcoded physics to avoid brute-force data, yet still needs a high-fidelity initialization corpus. Simulator-generated data carries discretization and truncation errors; the model learns the simulator’s bugs and the architecture’s rigor hyper-enforces them — the zero-shot illusion.

Gray-box resolution. (i) *Extreme domain randomization* [6], [7]: mutate friction, viscosity, and latency across millions of environments so the operator learns only the invariant physics that survive universal chaos; (ii) *PID as a control anchor*: a classical PID loop handles bounded local micro-adjustments while the neural operator predicts the global field; (iii) *coverage*: if the randomization bounds contain the real parameters, zero-shot transfer holds without a hardware-in-the-loop farm.

Proposition VIII.1 (Domain-randomization transfer bound). *If the simulator draws $\xi \sim P_{rand}$ whose support contains the true ξ^* , a model minimizing expected loss over P_{rand} has bounded expected loss at ξ^* ; with sufficient mass near ξ^* , zero-shot transfer holds.*

The unmodeled-physics wall. Randomization only covers modeled parameters; if the simulator omits a chaotic variable (microscopic thermal fatigue, vortex-shedding breakdown), no randomization teaches the operator to invent it. This is the terminal limit of simulation-based bootstrapping and the argument for eventual hardware-in-the-loop grounding of the uncertainty head.

IX. EXPERIMENTAL VALIDATION

Six controlled experiments test the headline claims; each is small, seeded, and reproducible on CPU (results and scripts in the repository). Table III summarizes the outcomes.

E1 (frames). A rotating point cloud embedded with discrete frame selection snaps (mean latent jump 0.068, maximum 2.0, on 3.4% of steps); the continuous Haar expectation of the frame distribution glides (maximum jump 0.0) — frame averaging is discontinuous exactly as predicted, and expectation removes it.

Experiment	Bottleneck	Fix
E1 frames	max jump 2.0	max jump 0.0
E2 gradient	PCGrad/Nash frozen	vanilla dist 1.0 $\rightarrow$ 0.51
E3 conservation	source 8.90, mass -0.88%	source 0.0
E4 mesh	MLP locked to $N=16$	FNO 0.0069 at all N
E5 randomize	OOR MSE 5.55	in-range 0.060; PID 0.875 $\rightarrow$ 0.115
E6 bottleneck	HF 2.6–37.4%	bypass 84–93%; err 0.635 $\rightarrow$ 0.277

TABLE III

SIX EXPERIMENTS: THE PREDICTED FAILURE MODES MATERIALIZE AND THE PROPOSED FIXES ELIMINATE THEM. E1–E6 MAP TO SECTIONS 4–9; DETAILS IN THE REPOSITORY (“EXPERIMENTS/RESULTS/”).

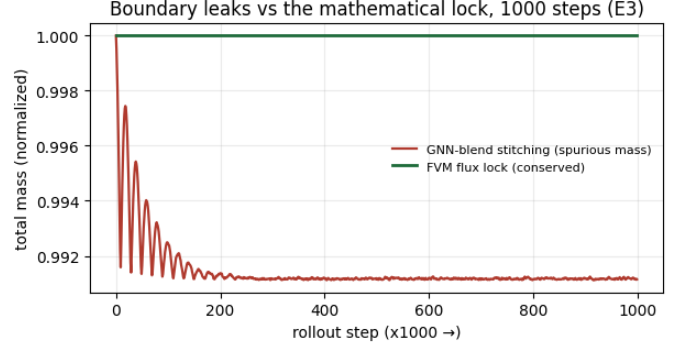


Fig. 3. Conservation (E3): the naive boundary stitcher leaks mass over 1,000 steps; the FVM flux lock is exactly conservative.

E2 (gradient surgery). Two physics residuals pull a coupled field $T(x)$ in opposite directions (L_A wants $+v$, L_B wants $-0.5v$; compromise $T^* = 0.25v$), with exactly anti-parallel gradients at initialization. PCGrad’s projection zeroes the coupled update exactly (mean norm 0.0; distance to T^* stuck at 1.0); the null-space trap. NashMTL’s bargaining QP is infeasible for exactly anti-parallel residuals, so its equal-weight fallback cancels and NashMTL freezes identically — a local Nash deadlock. Only vanilla SGD stays active (update norm 0.048, distance 1.0 $\rightarrow$ 0.51).

E3 (conservation). Over 1,000 advection steps on a subdomain-split domain, naive boundary stitching injects a spurious source of 8.90 and destroys 0.88% of the mass; the FVM flux lock is exactly conservative (spurious source 0.0, zero drift; Fig. 3).

E4 (mesh independence). A Fourier operator trained at $N=16$ evaluates zero-shot at $N=64, 128$ with identical relative error (0.0069); an MLP baseline is architecturally locked to its training resolution. The operator backbone is mesh-independent; the token-based baseline is not.

E5 (domain randomization + PID). In-range randomized coefficients transfer zero-shot (MSE 0.060); out-of-range coefficients hit the unmodeled-physics wall (MSE 5.55, 93 $\times$ worse). Closing the loop with the PID correction reduces open-loop rollout error from 0.875 to 0.115.

E6 (wavelet bypass). A fixed-array bottleneck realized in the Fourier basis (keep M lowest modes) erases sharp spikes as M shrinks (high-frequency energy kept falls to 2.6%); the dyadic wavelet bypass restores 84–93% of it and cuts the reconstruction error from 0.635 to 0.277.

X. RELATED WORK

Neural operators (FNO [8], DeepONet [9], Geo-FNO [10], message-passing PDE solvers [11]) and physics-informed learning [12], [13] supply the operator backbone; NVIDIA’s Modulus is an early engineering ancestor. SE(3)-equivariant networks [14]–[16], frame averaging [1], and vector neurons [2] supply the symmetry machinery. Gradient balancing spans PCGrad [3] and Nash bargaining [5], [17]; latent bottlenecks follow the Perceiver [4]. Topological machinery comes from cellular sheaves [18]–[20], topological deep learning [21], discrete exterior calculus [22], [23], persistent homology [24], [25], and Clifford group equivariant networks [26], [27]; sim-to-real grounding follows domain randomization [6], [7].

XI. CONCLUSION

The naive UGCT is a coherent vision whose four tiers fail in characterizable ways; the fixes fail in turn; and the terminal resolution leaves coordinate grids entirely — a sheaf-theoretic Clifford latent space with dissipative Hodge decomposition, grade-fading, symplectic scrubbing, and non-commutative autodiff, completed by a deployment tier and gray-box bootstrap. The adversarial design analysis is the paper’s contribution: an explicit, formalizable map of the design space of physics-native architectures, and an argument that physics-native models can enforce the rules of the universe inside the network rather than memorizing them from data. Section IX validates the failure modes and fixes empirically; the complete stack has not yet been trained end-to-end, and the reference implementation provides the components that a full training run would assemble.

Reproducibility. All code, figures, and paper source are in github.com/sehajr-singhs/ugct. The `ugct/` package implements every component (frames, operators, losses, conservation, perceiver, wavelet, hodge, and the assembled UGCT); six experiments (`experiments/frame_continuity.py`, `gradient_surgery_physics.py`, `conservation_rollout.py`, `mesh_independence.py`, `domain_randomization.py`, `wavelet_bypass.py`) reproduce the headline failure modes with committed JSON results; a Kaggle notebook and a Hugging Face model card are included. All experiments run on CPU in seconds to minutes.

REFERENCES

- [1] O. Puny, M. Atzmon, H. Ben-Hamu, E. J. Smith, I. Misra, A. Grover, and Y. Lipman, “Frame averaging for invariant and equivariant networks,” in *International Conference on Learning Representations (ICLR)*, 2022.
- [2] C. Deng, O. Litany, Y. Duan, A. Poulenard, A. Tagliasacchi, and L. Guibas, “Vector neurons: A general framework for so(3)-equivariant networks,” in *International Conference on Computer Vision (ICCV)*, 2021.
- [3] T. Yu, S. Kumar, A. Gupta, S. Levine, K. Finn, and S. Savarese, “Gradient surgery for multi-task learning,” in *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.
- [4] A. Jaegle, F. Gimeno, A. Brock, O. Vinyals, A. Zisserman, and J. Carreira, “Perceiver: General perception with iterative attention,” in *International Conference on Machine Learning (ICML)*, 2021.
- [5] A. Navon, A. Shamsian, I. Achituve, H. Maron, K. Kawaguchi, G. Chechik, and E. Fetaya, “Multi-task learning as a bargaining game,” in *International Conference on Machine Learning (ICML)*, 2022.
- [6] J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba, and P. Abbeel, “Domain randomization for transferring deep neural networks from simulation to the real world,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017.
- [7] X. B. Peng, M. Andrychowicz, W. Zaremba, and P. Abbeel, “Sim-to-real transfer of robotic control with dynamics randomization,” in *IEEE International Conference on Robotics and Automation (ICRA)*, 2018.
- [8] Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, and A. Anandkumar, “Fourier neural operator for parametric partial differential equations,” in *International Conference on Learning Representations (ICLR)*, 2021.
- [9] L. Lu, P. Jin, G. Pang, Z. Zhang, and G. E. Karniadakis, “Learning nonlinear operators via deepnet based on the universal approximation theorem of operators,” *Nature Machine Intelligence*, vol. 3, no. 3, pp. 218–229, 2021.
- [10] Z. Li, D. Z. Huang, B. Liu, and A. Anandkumar, “Fourier neural operator with learned deformations for pdes on general geometries,” *Journal of Machine Learning Research*, vol. 25, no. 89, pp. 1–24, 2024.
- [11] J. Brandstetter, D. Worrall, and M. Welling, “Message passing neural pde solvers,” in *International Conference on Learning Representations (ICLR)*, 2022.
- [12] 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*, vol. 378, pp. 686–707, 2019.
- [13] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, “Physics-informed machine learning,” *Nature Reviews Physics*, vol. 3, no. 6, pp. 422–440, 2021.
- [14] N. Thomas, T. Smidt, S. Kearnes, L. Yang, L. Li, K. Kohlhoff, and P. Riley, “Tensor field networks: Rotation- and translation-equivariant neural networks for 3d point clouds,” *arXiv preprint arXiv:1802.08219*, 2018.
- [15] M. Weiler, M. Geiger, M. Welling, W. Boomsma, and T. Cohen, “3d steerable cnns: Learning rotationally equivariant features in volumetric data,” in *Advances in Neural Information Processing Systems (NeurIPS)*, 2018.
- [16] F. Fuchs, D. Worrall, V. Fischer, and M. Welling, “Se(3)-transformers: 3d roto-translation equivariant attention networks,” in *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.
- [17] R. Schwartz, D. Haviv, T. Stamper, and O. Nachum, “A game-theoretic approach to multi-task learning,” in *International Conference on Learning Representations (ICLR)*, 2020.
- [18] J. Hansen and R. Ghrist, “Toward a spectral theory of cellular sheaves,” *Journal of Applied and Computational Topology*, vol. 3, no. 4, pp. 315–358, 2019.
- [19] J. Hansen and T. Gebhart, “Sheaf neural networks,” *arXiv preprint arXiv:2006.14532*, 2020.
- [20] C. Bodnar, F. Di Giovanni, B. P. Chamberlain, P. Lió, and M. Bronstein, “Neural sheaf diffusion: A topological perspective on heterophily and oversmoothing in gnns,” in *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.
- [21] M. Hajji, G. Zamzmi, T. Papamarkou, N. Miolane, A. Guzmán-Sáenz, K. N. Ramamurthy, T. Birdal, T. K. Dey, S. Mukherjee, and S. N. Samaga, “Topological deep learning: Going beyond graph data,” *arXiv preprint arXiv:2006.12704*, 2020.
- [22] K. Crane, F. de Goes, M. Desbrun, and P. Schröder, “Digital geometry processing with discrete exterior calculus,” in *ACM SIGGRAPH Courses*, 2013.
- [23] M. Desbrun, A. N. Hirani, M. Leok, and J. E. Marsden, “Discrete exterior calculus,” *arXiv preprint math/0508341*, 2005.
- [24] G. Carlsson, “Topology and data,” *Bulletin of the American Mathematical Society*, vol. 46, no. 2, pp. 255–308, 2009.
- [25] H. Edelsbrunner and J. Harer, “Persistent homology – a survey,” in *Contemporary Mathematics*. American Mathematical Society, 2008, vol. 453, pp. 257–282.
- [26] J. Brandstetter, R. van den Berg, M. Welling, and J. K. Gupta, “Clifford group equivariant neural networks,” in *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.
- [27] D. Ruhe, J. Brandstetter, and P. Forró, “Geometric clifford algebra networks,” in *International Conference on Machine Learning (ICML)*, 2023.