Functional Attention: From Pairwise Affinities to Functional Correspondences

Jiefang Xiao[1,2] Maolin Gao[1,2,*] Simon Weber[3,*] Guandao Yang[4] Daniel Cremers[1,2]
[1] TU Munich [2] MCML [3] University of Oxford [4] ECE, UT Austin * Corresponding Author

ICML International Conference On Machine Learning

Github

Why this matters

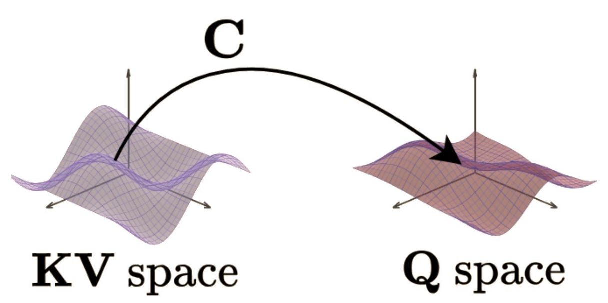
Beyond Tokens

Standard attention builds an explicit n×n affinity matrix between points.

Key Limitations Today

- Neurall operators limited to regular domain
- Attention directly applied on > 10k tokens

New View



We reinterpret softmax attention as **functional correspondence** between two spaces.

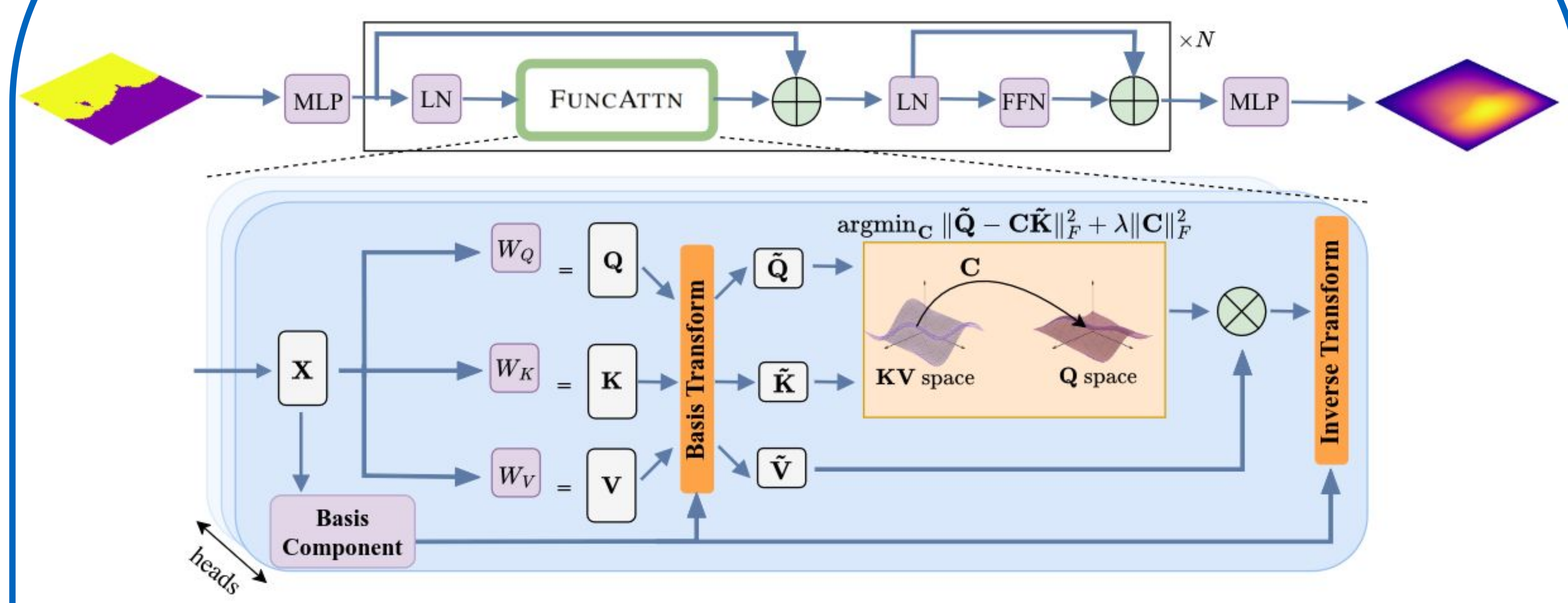
Our Motivation

Inspired by Functional map[1]: under suitable bases, correspondence between two function spaces reduces to a compact linear operator.

17.56% improvement on five standard benchmark

Method ($\times 10^{-2}$)	Elasticity	Airfoil	Darcy	Pipe	Navier–Stokes	Plasticity
2nd Best	0.64 (Transolver)	0.53 (Transolver)	0.57 (Transolver)	**0.25** (LNO)	8.45 (LNO)	0.13 (Transolver)
Ours	**0.50** 21.9%	**0.43** 18.9%	**0.42** 26.3%	0.29 –	**8.00** 5.3%	**0.11** 15.4%

Methodology

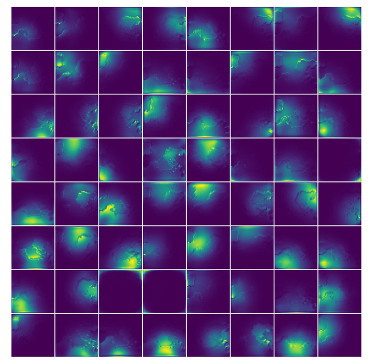


1. Data-dependent learned basis

$\mathcal{B} = \mathrm{Softmax}(\mathrm{Linear}(\mathbf{X})) \in \mathbb{R}^{n \times k}$

soft assignment from mesh points to learned bases[2].

As a generalization of the P_0 Finite element



Darcy flow basis visualization

Basis transformation

2. Solve functional correspondence

Regularized least-squares in latent space:

$$\min_{\mathbf{C}} \|\tilde{Q} - \mathbf{C}\,\tilde{K}\|_F^2 + \lambda\|\mathbf{C}\|_F^2$$

Closed-form solution:

$$\mathbf{C}^* = \tilde{Q}\,\tilde{K}^\top\left(\tilde{K}\,\tilde{K}^\top + \lambda\mathbf{I}_k\right)^{-1}$$

Vanilla attention[3]

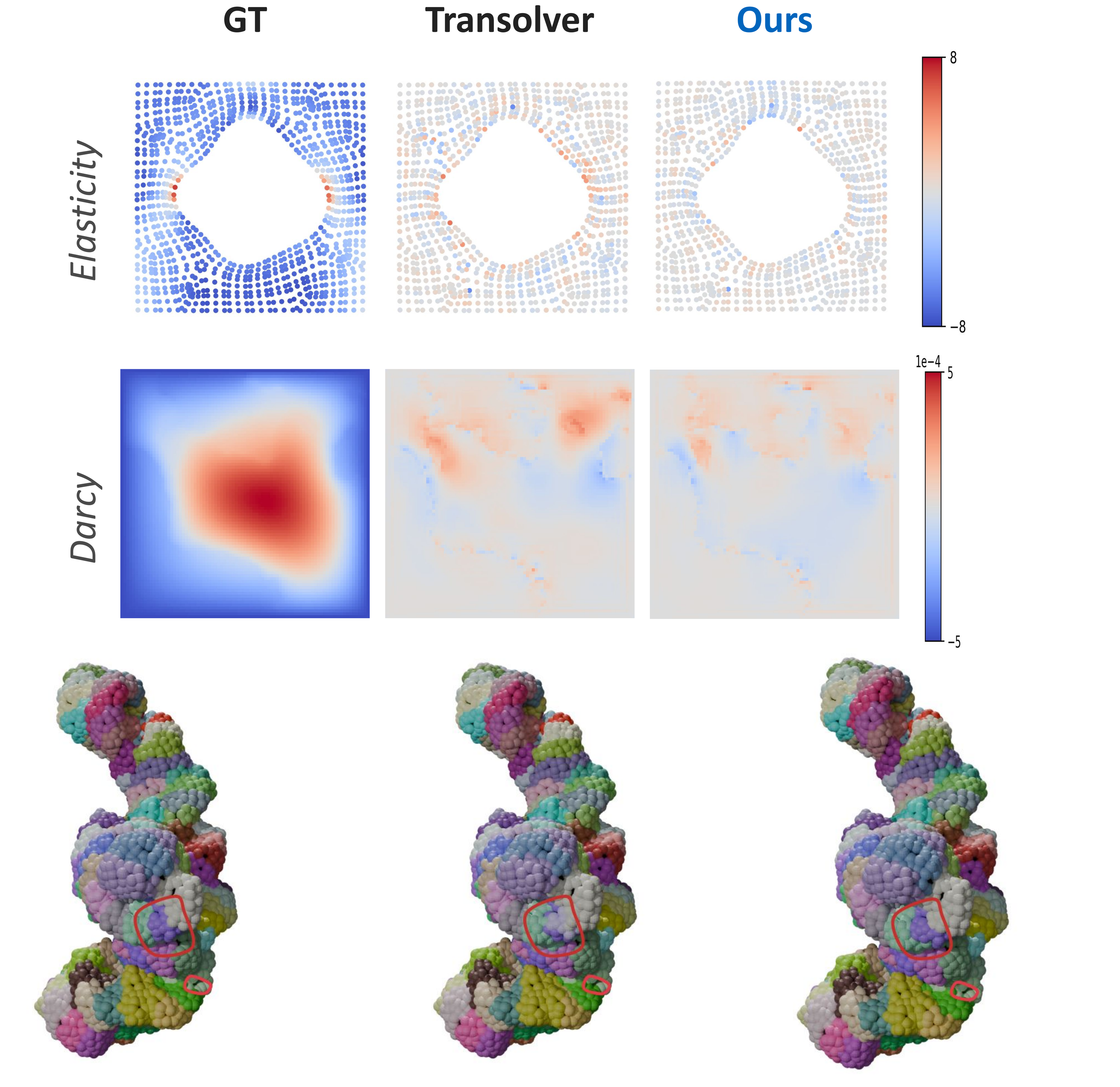
$$O = \mathrm{Softmax}\left(\frac{QK^\top}{\sqrt{d}}\right)V$$

Functional attention

$$O = B_Q C^* B_K^\top V$$

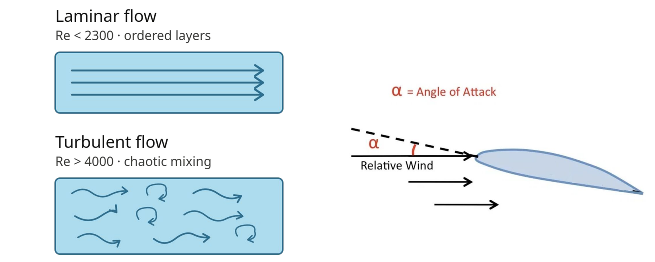
Results at a glance

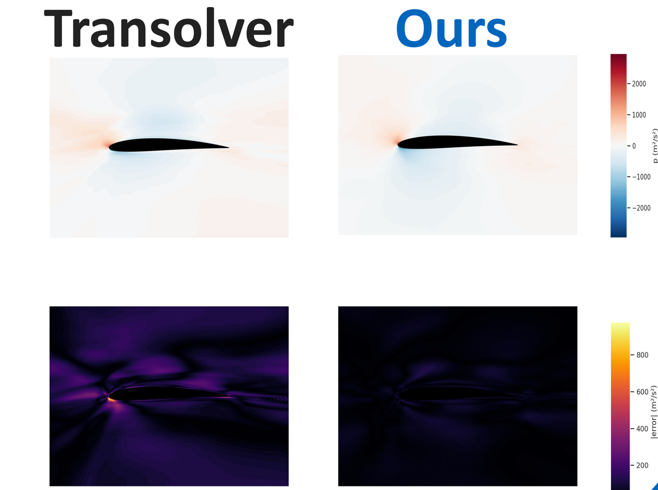
Qualitative Results



OOD Generalizability on Unseen conditions

Models	OOD Reynolds C_L (↓)	OOD Reynolds ρ_L (↑)	OOD Angles C_L (↓)	OOD Angles ρ_L (↑)
Simple MLP	62.1	95.8	41.3	95.7
GraphSAGE (2017)	43.3	97.1	25.4	98.9
PointNet (2017)	38.4	98.1	44.3	97.8
Graph U-Net (2019)	46.6	96.5	37.6	98.2
MeshGraphNet (2020)	177.2	76.3	65.3	89.3
GNO (2023a)	44.1	98.8	30.4	98.8
Galerkin (2021)	46.2	98.3	38.1	98.2
GNOT (2023)	32.7	98.7	35.0	98.7
GINO (2023b)	41.8	96.5	25.8	99.2
Transolver (2024)	32.2	98.7	22.8	99.0
Ours	**23.4**	**99.4**	**13.3**	**99.7**





References:
[1] Ovsjanikov et al. Functional Maps: A Flexible Representation of Maps Between Shapes. ACM TOG (SIGGRAPH) 2012.
[2] Wu et al. Transolver: A Fast Transformer Solver for PDEs on General Geometries. ICML 2024.
[3] Vaswani et al. Attention Is All You Need. NeurIPS 2017.