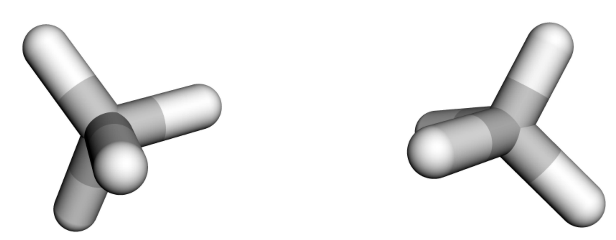
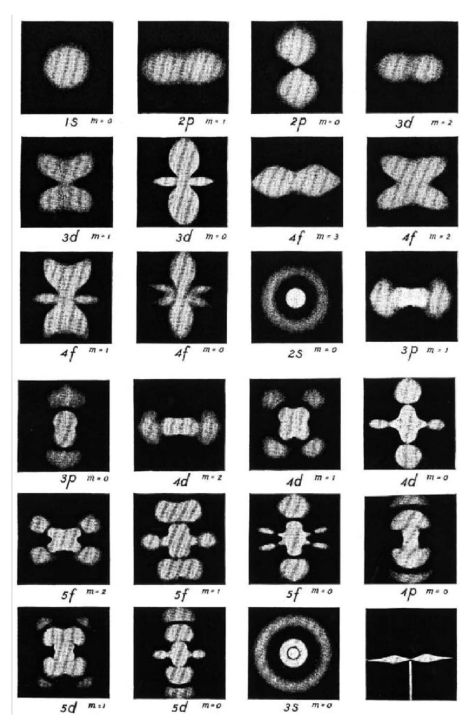
RETHINKING THE BENEFITS OF STEERABLE FEATURES IN 3D EQUIVARIANT GRAPH NEURAL NETWORKS

Shih-Hsin Wang[1], Yung-Chang Hsu[2], Justin Baker[1], Andrea L. Bertozzi[3], Jack Xin[4], Bao Wang[1,5]
[1]University of Utah [2]University of Purdue [3]University of California Los Angeles [4]University of California Irvine [5]Scientific Computing and Imaging Institute

Background

- **Science, Symmetry & Steerable Features:**
  - Features exhibit symmetries – they remain unchanged or change accordingly under group transformations, e.g., rotations, reflections, and translations.





  - Steerable features are those that transform linearly under transformations, e.g. a matrix "steering" the feature vector,

$$g \cdot \boldsymbol{v} = \boldsymbol{D}(g) \cdot \boldsymbol{v}.$$

  - For $\mathrm{SO}(3)$ and $\mathrm{O}(3)$, steerable features can be decomposed into irreducible components, known as type-L steerable features, with actions defined by Wigner D-matrices $\boldsymbol{D}^l(g)$.
- **Geometric Graph Neural Networks:**
  - To leverage symmetries, people have developed geometric GNNs to learn steerable features up to type-L:

$$\boldsymbol{f}_i^{(t+1)} = \mathrm{UPD}\left(\boldsymbol{f}_i^{(t)}, \mathrm{AGG}(\{\!\{\boldsymbol{f}_i^{(t)}, \boldsymbol{f}_j^{(t)}, \boldsymbol{x}_i - \boldsymbol{x}_j \mid j \in \mathcal{N}_i^{(k)}\}\!\})\right), \quad (1)$$

    - $\boldsymbol{f}_i^{(t)}$ denotes the steerable feature of node i at layer t,
    - $\boldsymbol{x}_i$ represents the input coordinates of node i,
    - UPD and AGG are learnable update and aggregate equivariant functions, respectively, where $\{\!\{\cdot\}\!\}$ denotes a multiset,
    - $\mathcal{N}_i^{(k)}$ denotes the k-hop neighborhood of node i, the set of nodes in $\mathcal{V}$ that are reachable from i through a path with k edges or fewer.
- **Existing Insights on Geometric GNNs' Performance:**
  - Joshi et al.[1] have shown that 1-hop invariant GNNs ($L = 0$) may underperform equivariant GNNs ($L = 1$).
  - Several studies[2345] showed that equivariant GNNs using steerable features up to type-L improve with higher L.

Contributions

- **Remaining Questions**
  - The impact of introducing multi-hop message passing aggregation (e.g. SphereNet[6] and ComENet[7]) into invariant GNNs remains unexplored.
  - Existing experiments comparing different types of steerable features often lack control over feature dimensions, making it difficult to isolate the true effect of feature type.
- **Our Answers**
  - Even introducing multi-hop message passing aggregation, invariant GNNs lack the intrinsic capability to capture geometric information between local neighborhoods, and hence, fail to obtain accurate global invariant features.
  - When preserving the feature dimension, the performance of equivariant GNNs using steerable features up to type-L may not increase as L grows.

k-hop Invariant GNNs: Missing Global Features

Theorem 1. *If $\mathcal{G}_1$ and $\mathcal{G}_2$ are two k-hop identical graphs, then any iteration of k-hop invariant GNNs will get the same output from these two graphs. That is, there is a graph isomorphism b such that $\boldsymbol{f}_i^{(t+1)}(\mathcal{G}_1) = \boldsymbol{f}_{b(i)}^{(t+1)}(\mathcal{G}_2)$ for any i, even though $\mathcal{G}_1$ and $\mathcal{G}_2$ may not be identical up to group action.*

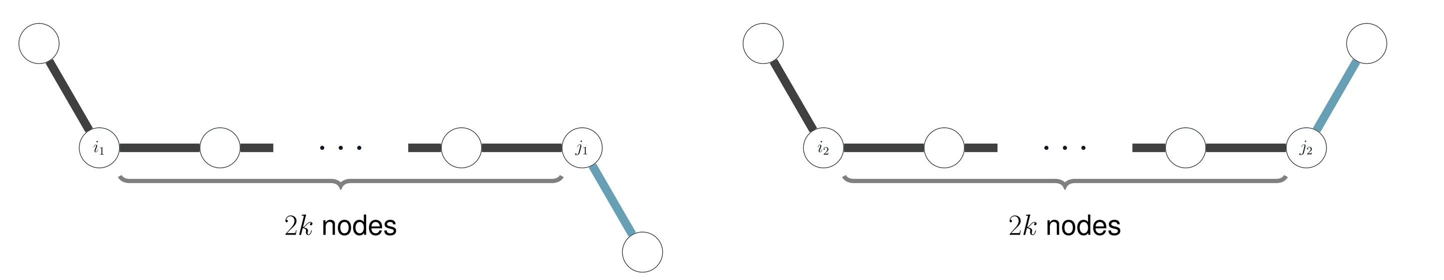


Figure 1: A pair of graphs each consisting of $2k+2$ nodes, called k-chains, introduced from Joshi et al.[1]. These graphs are nearly identical, differing only in the orientation of a single edge, marked in blue. Despite this minor distinction, these graphs remain k-hop identical.

Equivariant GNNs:

- **The importance of faithfulness:**

Theorem 2. *Consider 1-hop equivariant GNNs learning features on steerable vector space V where the aggregate function AGG learns features on steerable vector space W. Suppose V and W are faithful representations, and AGG and UPD are G-orbit injective and G-equivariant multiset functions. Then with k iterations, these equivariant GNNs learn different multisets of node features $\{\!\{\boldsymbol{f}_i^{(k)}\}\!\}$ on two k-hop distinct geometric graphs.*

- **Correspond steerable features to invariant features:** according to Winter et al.[8], we can represent any $\boldsymbol{X} \in \mathbb{R}^{3\times m}$ using a group element $g_{\boldsymbol{X}} \in G$ and a canonical representative $c(\boldsymbol{X}) \in \mathbb{R}^{3\times m}$ where we have $g_{\boldsymbol{X}} \cdot c(\boldsymbol{X}) = \boldsymbol{X}$.

Lemma 3. *Let V be a d-dimensional G-steerable vector space with the assigned group representation $\rho : G \to \mathrm{GL}(V)$. If $f : \mathbb{R}^{3\times m} \to V$ is G-equivariant, then there exists a unique G-invariant function $\lambda : \mathbb{R}^{3\times m} \to V_0^{\oplus d}$ s.t. $f(\boldsymbol{X}) = \rho(g_{\boldsymbol{X}})\lambda(\boldsymbol{X})$, where V_0 denotes the 1D trivial representation of G[9]. In particular, the following map is well-defined*

$$\{f : \mathbb{R}^{3\times m} \to V \mid f : G\text{-equivariant}\} \to \{\lambda : \mathbb{R}^{3\times m} \to V_0^{\oplus d} \mid \lambda : G\text{-invariant}\}.$$

- **Learning steerable features of the same dimension:**

Corollary 4. *Let V and W be two steerable vector spaces of dimension d. Then for any G-equivariant function $f_V : \mathbb{X}_3 \to V$, there is a G-equivariant function $f_W : \mathbb{X}_3 \to W$ such that for any $\boldsymbol{X} \in \mathbb{X}_3$, we have $f_V(\boldsymbol{X}) = \rho_V(g_{\boldsymbol{X}})\lambda(\boldsymbol{X})$ and $f_W(\boldsymbol{X}) = \rho_W(g_{\boldsymbol{X}})\lambda(\boldsymbol{X})$ for the same G-invariant function λ where ρ_V, ρ_W are the group representation on V and W, resp.*

Theorem 5. *Consider two geometric GNNs learning features on steerable vector spaces V and W of the same dimension, resp. Denote their update and aggregation functions at iteration t as $\mathrm{UPD}_V^{(t)}$, $\mathrm{UPD}_W^{(t)}$ and $\mathrm{AGG}_V^{(t)}$, $\mathrm{AGG}_W^{(t)}$. Then for any collection $\{(\mathrm{UPD}_V^{(t)}, \mathrm{AGG}_V^{(t)})\}_t$, there exists a collection $\{(\mathrm{UPD}_W^{(t)}, \mathrm{AGG}_W^{(t)})\}_t$ such that for any fully connected graph, they learn the same corresponding invariant features $\lambda_i^{(t)}$ for any iteration $t \geq 0$ on each node i.*

- **Remark:** Theorem 5 establishes the equivalence of geometric GNNs on fully connected graphs without strong assumptions on the injectivity of update and aggregate functions, holding for any representation.

Numerical Results

Layers	1	2	3	1	2	3	4
k-hop chain	$k=2$	$k=2$	$k=2$	$k=3$	$k=3$	$k=3$	$k=3$
			$L=0$				
SchNet	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0	50.1 ± 0.2	50.0 ± 0.0	50.0 ± 0.0
DimeNet++	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0
SphereNet	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0
ComENet	55.0 ± 4.5	59.0 ± 11.6	53.0 ± 6.4	54.0 ± 6.2	50.0 ± 0.0	46.5 ± 5.0	51.0 ± 2.0
EquiformerV2	71.0 ± 3.0	76.0 ± 8.0	83.0 ± 6.4	43.0 ± 9.0	67.0 ± 4.6	67.9 ± 9.0	61.0 ± 5.4
			$L=1$				
EGNN	50.0 ± 0.0	100.0 ± 0.0	95.0 ± 15.0	50.0 ± 0.0	50.0 ± 0.0	90.0 ± 20.0	100.0 ± 0.0
GVP	50.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0	50.0 ± 0.0	92.5 ± 16.0	91.5 ± 17.3	95.0 ± 15.0
ClofNet	50.0 ± 0.0	50.0 ± 0.0	100.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0
MACE	50.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0	50.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0
eSCN	64.0 ± 8.0	60.5 ± 10.0	64.3 ± 18.2	53.0 ± 4.6	63.0 ± 9.0	60.0 ± 13.4	56.0 ± 10.2
EquiformerV2	90.0 ± 0.0	95.0 ± 5.0	96.0 ± 4.9	76.0 ± 6.6	84.0 ± 6.6	92.0 ± 6.0	98.0 ± 4.0
			$L=2$				
MACE	50.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0	50.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0
eSCN	62.0 ± 7.5	61.0 ± 9.4	52.0 ± 4.0	62.0 ± 10.8	59.0 ± 9.4	56.0 ± 10.2	54.0 ± 6.6
EquiformerV2	73.0 ± 4.6	88.0 ± 4.0	86.0 ± 4.9	86.0 ± 4.9	89.0 ± 3.0	88.0 ± 4.0	83.0 ± 9.0

Table 1: Test accuracy for the k-chain dataset with different ks. Models are further distinguished by their use of type-L features. Cell shading is based on two standard deviations above or below the expected value. Unit:%.

Model	L	c	Feat. Dim.	# Param.	**Loss** ↓	**Energy MAE** [meV] ↓	**EwT** [%] ↑
eSCN	1	464	1856	11M	0.380 ± 0.006	865 ± 14	1.91 ± 0.09
eSCN	2	206	1854	10M	$\mathbf{0.369 \pm 0.006}$	$\mathbf{842 \pm 13}$	$\mathbf{1.94 \pm 0.12}$
eSCN	3	133	1862	9M	0.397 ± 0.001	904 ± 3	$1.85 \pm .12$
eSCN	4	98	1862	9M	0.408 ± 0.006	929 ± 15	1.74 ± 0.12
eSCN	5	77	1848	8M	0.409 ± 0.003	933 ± 7	$1.61 \pm .12$
eSCN	6	64	1856	8M	0.3836 ± 0.003	872 ± 6	1.91 ± 0.19
EquiformerV2	1	77	304	7M	OOM	OOM	OOM
EquiformerV2	2	34	306	9M	0.369 ± 0.009	841 ± 21	2.02 ± 0.14
EquiformerV2	3	22	306	12M	$\mathbf{0.363 \pm 0.009}$	$\mathbf{828 \pm 21}$	1.94 ± 0.08
EquiformerV2	4	16	304	15M	0.364 ± 0.005	832 ± 11	$\mathbf{2.03 \pm 0.14}$

Table 2: Validation results of the steerable model ablation study on L and c over 4-folds of the IS2RE dataset with 10k training molecules. We observe that higher type-L steerable models may not perform best. OOM denotes models that run out of memory during training.

Conclusion and Discussion

- To achieve equivalent expressiveness in invariant GNNs as in equivariant GNNs, it is essential to integrate global features that extend beyond the confines of fixed k-hop neighborhoods.
- The traditional trade-off between performance and computational cost of using steerable features in equivariant GNNs should be reevaluated. Specifically, when maintaining a constant feature dimension, the utilization of higher-type steerable features in equivariant GNNs might not ensure improved performance and could entail additional computational overhead.
- Limitation: our analysis of expressiveness focuses on the capacity of features to capture information. A broader view also considers the ability to extract features from data (universality).

Reference: 1. Joshi et al., "On the expressive power of geometric graph neural networks", PMLR, 2023 2. Batzner et al., "E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials.", Nature Communications. 3. Batatia et al., "Mace: Higher order equivariant message passing neural networks for fast and accurate force fields.", NeurIPS, 2022. 4. Liao & Smidt, "Equiformer: Equivariant graph attention transformer for 3D atomistic graphs.", ICLR, 2023. 5. Passaro & Zitnick, "Reducing SO(3) convolutions to SO(2) for efficient equivariant gnns", PMLR, 2023. 6. Liu et al.,"Spherical message passing for 3D molecular graphs", ICLR 2022 7. Wang et al.,"ComENet: Towards complete and efficient message passing for 3D molecular graphs", NeurIPS 2022 8. Winter et al.,"Unsupervised learning of group invariant and equivariant representations.", NeurIPS 2022

THE UNIVERSITY OF UTAH

www.sci.utah.edu