Symmetry-Preserving Conformer Ensemble Networks for Molecular Representation Learning

Yanqiao Zhu†, Yidan Shi†, Yuanzhou Chen, Fang Sun, Yizhou Sun, Wei Wang

University of California, Los Angeles   † Equal contribution

UCLA Samueli Computer Science

Motivation & Background

Why Conformer Ensembles?

- Molecules are dynamic and continuously interconvert between conformers via bond rotations and vibrations
- Many molecular properties depend on the entire distribution of thermodynamically accessible conformations
- Accurate molecular representations therefore require learning over the full conformer ensemble instead of a single geometry

Limitations of Prior Approaches

- 3D GNNs: Operate on one conformer ⇒ Ignore ensemble flexibility
- 4D-QSAR: Require rigid alignment and common scaffolds
- Multi-instance baselines: Treat conformers independently; miss cross-conformer interactions
- Structural averaging: Produces unphysical merged geometries, sensitive to conformer alignment

Problem Definition

Molecular Topology Representations

Represent molecular topology of each molecule as $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \boldsymbol{X})$

- $\mathcal{V} = \{v_i\}_{i=1}^{|\mathcal{V}|}$: Atom indices
- $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$: Chemical bonds
- $\boldsymbol{X} \in \mathbb{R}^{|\mathcal{V}| \times d_v}$: Node attributes

Conformer Ensemble Representations

For each molecule we consider n conformers $\mathcal{C} = \{\boldsymbol{C}_i\}_{i=1}^{n}$ with $\boldsymbol{C}_i \in \mathbb{R}^{|\mathcal{V}| \times 3}$ sampled from the thermodynamically accessible space

Each conformer has a Boltzmann weight (computed for ground-truth ensemble properties but withheld from the model):

$$p_i = \frac{\exp\left(-\frac{e_i}{k_B T}\right)}{\sum_j \exp\left(-\frac{e_j}{k_B T}\right)}$$

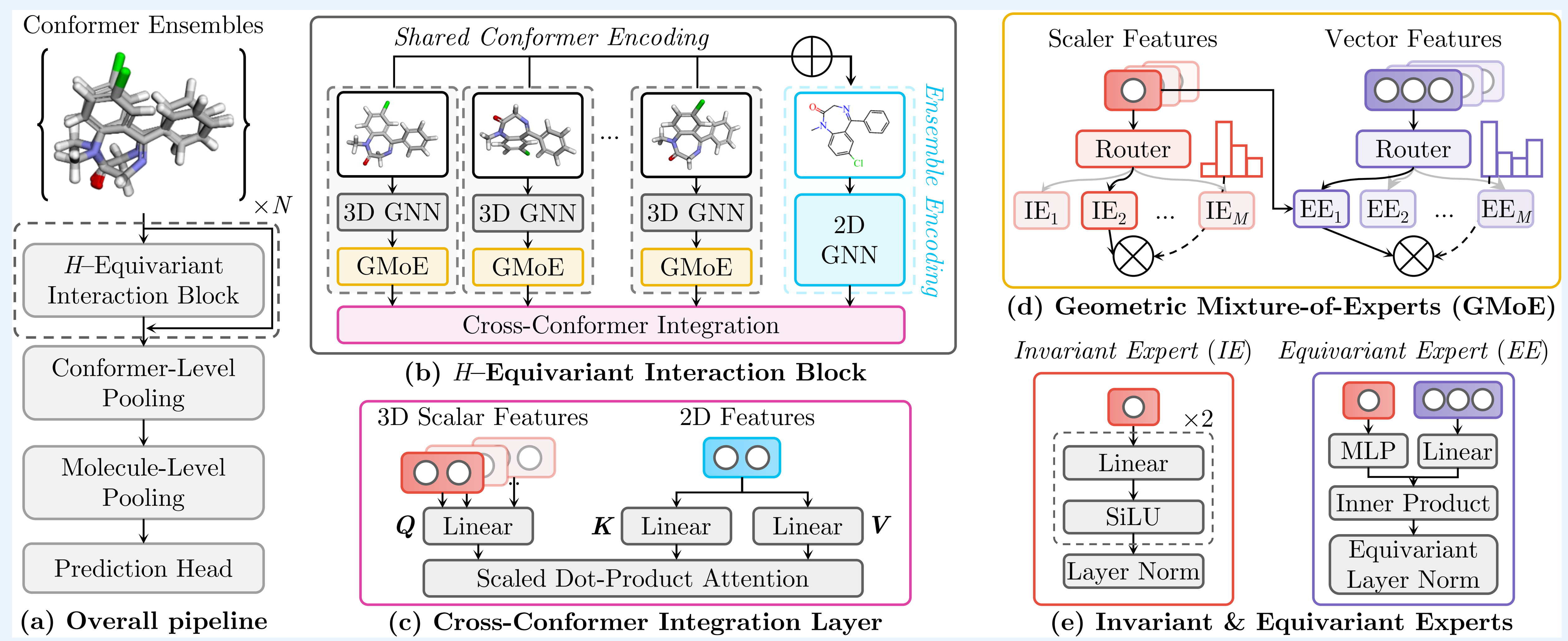
Conformer energy
Temperature
Boltzmann constant

Learning Objective

- Learn $(\mathcal{G}, \mathcal{C}) \mapsto$ Molecular properties
- Overall requirement: Functions equivariant to the product group $H = S_n \times G^n$
- S_n: Permutation symmetry over conformers
- G: Independent geometric transforms (e.g., $\mathrm{E}(3)$) per conformer

The SPiCE Architecture

TL;DR: SPiCE (Symmetry-Preserving Conformer Ensemble Networks) maintains joint equivariance to conformer permutations and geometric transformations, enabling physically meaningful molecular representations without alignment



SPiCE processes conformer ensembles through H-equivariant interaction blocks composed of: (1) shared 3D GNN encoding, (2) Geometric Mixture-of-Experts (GMoE), and (3) hierarchical ensemble encoding with cross-conformer integration

Benchmark Datasets & Tasks

Drugs-7.5K
- Diverse drug-like molecules
- Targets: Ionization Potential (IP), Electron Affinity (EA), and electronegativity (χ) with Mean Absolute Error (MAE)

Kraken
- Monodentate organophosphorus ligands
- Targets: Sterimol B_5/L and buried B_5/L descriptors for catalysis design (MAE)

CoV2
- Human-cell SARS-CoV-2 inhibition screen with extreme imbalance
- Evaluate ROC-AUC and Precision-Recall (PR)

CoV2-3CL
- 3CL protease inhibition benchmark
- Precision-critical hit discovery reported with ROC-AUC/PR

Key Components

Shared Conformer Encoding

- Shared 3D GNN layers to process conformer geometry
- Hybrid node embeddings retain scalar (type-0) and vector (type-1) channels across interaction blocks
- Operating on each conformer separately yields S_n-equivariant outputs, and the equivariant backbones ensure G^n-equivariance

Key Components

Geometric Mixture-of-Experts (GMoE)

- Weight-tied scalar (type-0) and vector (type-1) branches preserve symmetry across conformers
- Invariant experts: Two-layer MLP specialists operate on scalar channels
- Equivariant experts: Gated linear transformations process vector channels while respecting rotations
- Router design: GCN scoring for scalars and inner-product scores for vectors
- Stabilize training: Top-k selection with Gumbel-Sigmoid sampling, router z-loss, and gradual expert upcycling

Hierarchical Ensemble Encoding

- A 2D GIN injects molecular topology using graph edges and a learnable offset before passing context downstream
- Cross-attention between conformer features and the GIN summary enables selective information flow across the ensemble
- Gated updates maintain the joint symmetry while integrating conformer-level and molecule-level information

Experimental Results

Main Results

Backbone	Ensemble Strategy	Drugs-7.5K (MAE, ↓) IP	EA	χ	Kraken (MAE, ↓) B_5	L	BurB_5	BurL	CoV2 ROC (↑)	3CL ROC (↑)
▶ *1D String-based and 2D Topological Approaches*										
Fingerprint+RF [54]		0.5833	0.5277	0.3130	0.4760	0.4303	0.2758	0.1521	0.6071	0.9013
E3FP+RF [55]		0.6217	0.5774	0.3464	0.6249	0.5535	0.3692	0.1908	0.6046	0.7676
GIN [50]		0.5575	0.5116	0.2892	0.3128	0.4003	0.1719	0.1200	0.3708	0.5942
GIN-VN [56]		0.5398	0.5160	0.2937	0.3567	0.4344	0.2422	0.1741	0.4832	0.7387
GraphGPS [57]		0.5480	0.5054	0.2863	0.3450	0.4363	0.2066	0.1500	0.5601	0.8387
▶ *3D Single-Conformer Graph Neural Networks with Random Conformer Sampling [51]*										
PaiNN [37]		0.5557	0.5127	0.2924	0.3443	0.4471	0.2395	0.1673	0.2997	0.8368
ClofNet [13]		0.6316	0.6008	0.3615	0.4473	0.6369	0.3216	0.2426	0.5233	0.7562
Equiformer [34]		0.5471	0.4898	0.2887	0.2709	0.3759	0.2019	0.1526	0.4577	0.8035
ViSNet [58]		0.5393	0.4855	0.2985	0.3828	0.4495	0.2400	0.1755	0.5011	0.4774
▶ *Conformer Ensemble Approaches*										
ConfNet [60]		0.5760	0.5359	0.3057	0.4469	0.4680	0.2686	0.1657	0.5010	0.4930
ConAN-FGW [31]		0.5471	0.4945	0.2891	0.3242	0.5178	0.2026	0.1492	0.6340	0.9180
PaiNN [37]	Mean	0.5410	0.4966	0.2963	0.2877	0.3950	0.1817	0.1472	0.5722	0.8850
	DeepSets	0.5396	0.5091	0.2982	0.2225	0.3619	0.1693	0.1324	0.5802	0.6808
	Attention	0.6318	0.5985	0.3488	0.3496	0.4109	0.2123	0.1506	0.4179	0.6984
	SPiCE	0.5281	0.4929	**0.2792**	0.2178	0.3548	**0.1564**	0.1292	0.5910	0.8880
ClofNet [13]	Mean	0.5935	0.5441	0.3121	0.3986	0.5674	0.2857	0.2327	0.3900	0.7580
	DeepSets	0.5912	0.5533	0.3153	0.3314	0.5375	0.2532	0.1983	0.6208	0.7628
	Attention	0.6694	0.5949	0.3578	0.4979	0.6118	0.3353	0.2502	0.3707	0.8182
	SPiCE	0.5747	0.5283	0.3059	0.3193	0.4903	0.2477	0.1913	0.6730	**1.0000**
Equiformer [34]	Mean	0.5457	0.4932	0.2977	0.2303	0.3830	0.1680	0.1259	0.5601	0.8387
	DeepSets	0.5404	0.4888	0.2990	0.2564	0.3772	0.1782	0.1234	0.5125	0.7134
	Attention	0.5488	0.4923	0.2896	0.3187	0.4508	0.1673	0.1425	0.3882	0.7881
	SPiCE	0.5318	0.4830	0.2816	0.2241	**0.3456**	0.1611	**0.1229**	0.5650	0.8405
ViSNet [58]	Mean	0.5593	0.4927	0.2862	0.2811	0.3970	0.1874	0.1469	0.6035	0.7447
	DeepSets	**0.5280**	0.4987	0.2846	0.3104	0.4113	0.1716	0.1314	0.6626	0.4160
	Attention	0.5593	0.4988	0.2944	0.3755	0.4195	0.2384	0.1394	0.5262	0.7158
	SPiCE	0.5384	**0.4538**	0.2814	**0.2715**	0.3807	0.1657	0.1277	**0.6890**	0.7195

Observations:

- SOTA across **34/36** configurations
- Regression (Drugs-7.5K, Kraken): 3–9% MAE improvements over the best baselines
- Classification (CoV2, CoV2-3CL): High ROC-AUC and robust to severe class imbalance

Ablation Studies