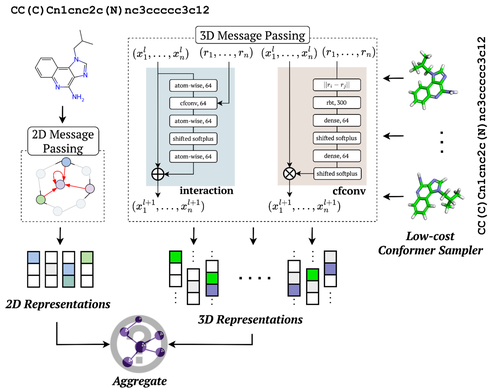
Structure-Aware E(3)-Invariant Molecular Conformer Aggregation Networks

Duy M. H. Nguyen*[1,2,3], Nina Lukashina*[1,2], Tai Nguyen[3], An T. Le[4], TrungTin Nguyen[5], Nhat Ho[6], Jan Peters[3,4,7], Daniel Sonntag[3,8], Viktor Zaverkin[9], Mathias Niepert[1,2,9]

[1]University of Stuttgart, [2]IMPRS for Intelligent Systems, [3]German Research Center for Artificial Intelligence, [4]Technical University of Darmstadt, [5]University of Queensland, [6]University of Texas at Austin, [7]Hessian.AI, [8]Oldenburg University, [9]NEC Laboratories Europe

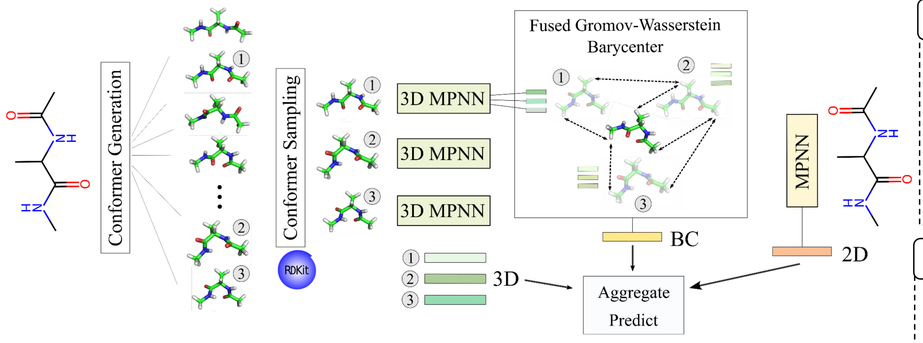
Motivations

- Prior **molecular representation learnings** consider **2D** molecular graphs or **3D** conformer structure representations in isolation.



- Invariant to E(3) transformations and permutations of conformers
- Modular and geometry-aware 2D-3D structures aggregation
- Works with conformers efficiently generated based on distance geometry
- Run on large-scale settings, be trainable with multiple GPUs

Conformer Aggregation Networks (ConAN)



2D message passing

Given a 2D graph $\mathbf{G} = (\mathbf{V}, \mathbf{E})$, output of initial node embedding $\mathbf{h}_v^{(0)}$ after $\mathbf{L}$ layers and $\mathbf{M}$ multi-heads

$$\mathbf{h}_v^{(L)} = \sigma\left(\frac{1}{M}\sum_{m=1}^{M}\sum_{u\in N(v)} \alpha_{v,u}^{(m)}\mathbf{W}^{(m)}\mathbf{h}_u^{L-1}\right)$$

3D message passing

Each 3D conformer graph $\mathbf{G}_k = (\mathbf{V}_k, \mathbf{E}_k)$ has output embedding:

$$\mathbf{h}_v^{(L)} = \sum_{u\in N(v)} \mathbf{h}_u^{(L-1)} \odot \mathbf{W}\left(\exp\left(-\gamma\left(\mathbf{e}_{uv} - \mu\right)\right)\right)$$

$\mathbf{e}_{uv}$ be interatomic distances between atoms.

FGW Barycenter Aggregation

Given $\mathbf{K}$ conformer graphs $\mathbf{G}_k = (\mathbf{H_k}, \mathbf{A_k}, \omega_\mathbf{k})$ where $\mathbf{H_k}$ be feature node matrix after 3D MPN, $\mathbf{A_k}$ be weighted adjacency matrix of pairwise atomic distance, $\omega_\mathbf{k}$ be probability mass of each atom

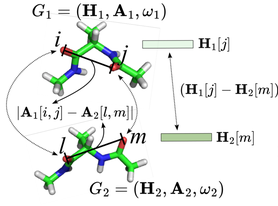
$$\overline{\mathbf{G}} := \arg\min_{\mathbf{G}} \sum_{k=1}^{K} \lambda_k \mathrm{FGW}_{p,\alpha}(\mathbf{G}, \mathbf{G_k}),$$

$$\boxed{\mathrm{FGW}_{p,\alpha}(\mathbf{G_i}, \mathbf{G_j})} := \min_{\boldsymbol{\pi}\in\boldsymbol{\Pi}(\boldsymbol{\omega}_i,\boldsymbol{\omega}_j)} \langle (1-\alpha)\mathbf{M} + \alpha\mathbf{L}(\mathbf{A}_i, \mathbf{A}_j) \otimes \boldsymbol{\pi}, \boldsymbol{\pi}\rangle$$

where $\boldsymbol{M} := (d_f(\boldsymbol{H}_i[a], \boldsymbol{H}_j[b])^p)_{n_1\times n_2} \in \mathbb{R}^{n_i\times n_j}$ is the pairwise node distance matrix,

$\mathbf{L}(\boldsymbol{A}_i, \boldsymbol{A}_j) = (|\boldsymbol{A}_i[a,b] - \boldsymbol{A}_j[c,d]|^p)_{abcd}$ the 4-tensor representing the alignment cost matrix,

$\boldsymbol{\Pi}(\boldsymbol{\omega}_i, \boldsymbol{\omega}_j) := \{\boldsymbol{\pi} \in \mathbb{R}_+^{n_i\times n_j} | \boldsymbol{\pi}\mathbf{1}_{n_j} = \boldsymbol{\omega}_i,\ \boldsymbol{\pi}\mathbf{1}_{n_i} = \boldsymbol{\omega}_j\}$ is the set of all valid couplings between two node distributions $\boldsymbol{\omega}_i$ and $\boldsymbol{\omega}_j$.



FGW distance constrains both **feature-based and structure aware**

Theorem 4.1. *Let $P \in \mathcal{P}_2(\mathcal{P}_2(\mathbf{\Omega}))$ be a probability measure on the 2-FGW space. Let $\overline{\mu}_0 \in \mathcal{P}_2(\mathbf{\Omega})$ and σ_P^2 be barycenter and variance functional of P satisfying (10) and (9), respectively. Let $\gamma, \beta > 0$ and suppose that every $\mu \in \mathrm{supp}(P)$ is the pushforward of $\overline{\mu}_0$ by the gradient of an γ-strongly convex and β smooth function $\psi_{\overline{\mu}_0\to\mu}$, i.e., $\mu = (\nabla\psi_{\overline{\mu}_0\to\mu})_{\#}\overline{\mu}_0$. If $\beta - \gamma < 1$, then $\overline{\mu}_0$ is unique and any empirical barycenter $\overline{\mu}_K$ of P satisfies*

$$\mathbb{E}\left(FGW_{2,\alpha}^2(\overline{\mu}_0, \overline{\mu}_K)\right) \le \frac{4\sigma_P^2}{(1-\beta+\gamma)^2 K}. \quad (12)$$

⇒ **Choosing number of conformers K**

ConAN accelerates FGW barycenter computation using entropic techniques, **enabling multi-GPUs training**

Algorithm 1 Entropic FGW Barycenter

Input: $\overline{\boldsymbol{\omega}}, \{G_s := (\boldsymbol{H}_s, \boldsymbol{A}_s, \boldsymbol{\omega}_s)\}_{s=1}^K, \epsilon$.
Optimizing: $\overline{G}, \{\boldsymbol{\pi}_s \in \Pi(\overline{\boldsymbol{\omega}}, \boldsymbol{\omega}_s)\}_{s=1}^K$.
repeat
  for $s = 1$ **to** K **do**
    Solve $\arg\min_{\boldsymbol{\pi}_s^{(k)}} \mathrm{FGW}_{p,\alpha}^{\epsilon}(\overline{G}^{(k)}, G_s)$ with Alg. 2.
  end for
  Update $\overline{\boldsymbol{A}}^{(k+1)} \leftarrow \frac{1}{\overline{\boldsymbol{\omega}}\overline{\boldsymbol{\omega}}^\top}\frac{1}{K}\sum_{s=1}^K \boldsymbol{\pi}_s^{(k)}\boldsymbol{A}_s\boldsymbol{\pi}_s^{(k)\top}$.
  Update $\overline{\boldsymbol{H}}^{(k+1)} \leftarrow \mathrm{diag}(1/\overline{\boldsymbol{\omega}})\frac{1}{K}\sum_{s=1}^K \boldsymbol{\pi}_s^{(k)}\boldsymbol{H}_s$
until k in *outer iterations* and *not converged*

Experimental Results

- **Molecular Property Prediction:** ConAN is compared against **13 baselines** across **4 datasets**.
- **Molecular Classification:** ConAN is compared against **7 baselines** across **2 datasets.**
- **Molecular Conformer Ensemble**: **6 baselines** using the **2nd-largest** reaction-level molecule properties **dataset**.

Model	Lipo	ESOL	FreeSolv	BACE
2D-GAT	1.387 ± 0.206	2.288 ± 0.017	8.564 ± 1.345	1.844 ± 0.33
D-MPNN	0.683 ± 0.016	1.050 ± 0.008	2.082 ± 0.082	2.253
Attentive FP	0.721 ± 0.001	0.877 ± 0.029	2.073 ± 0.183	-
PretrainGNN	0.739 ± 0.003	1.100 ± 0.006	2.764 ± 0.002	-
GROVER_large	0.823 ± 0.010	0.895 ± 0.017	2.272 ± 0.051	-
ChemBERTa-2*	0.798	0.889	-	1.363
ChemRL-GEM	0.660 ± 0.008	0.798 ± 0.029	1.877 ± 0.094	-
MolFormer	0.700 ± 0.012	0.880 ± 0.028	2.342 ± 0.052	1.047 ± 0.029
UniMol	0.603 ± 0.010	0.788 ± 0.029	1.480 ± 0.048	-
SchNet-scalar	0.704 ± 0.032	0.672 ± 0.027	1.608 ± 0.158	0.723 ± 0.1
SchNet-emb	0.589 ± 0.022	0.635 ± 0.057	1.587 ± 0.136	0.692 ± 0.028
ChemProp3D	0.602 ± 0.035	0.681 ± 0.023	2.014 ± 0.182	0.815 ± 0.17
ConfNet	1.360 ± 0.038	2.115 ± 0.484	-	1.329 ± 0.042
ConAN	0.556 ± 0.013	0.571 ± 0.019	1.496 ± 0.158	0.635 ± 0.051
ConAN-FGW	**0.487 ± 0.031**	**0.529 ± 0.022**	**1.068 ± 0.083**	**0.549 ± 0.016**

Models evaluation on regression tasks (RMSE) ↓

Method	Num Conformers	Dataset	ROC ↑	PRC ↑
SchNetFeatures	200	CoV-2 3CL	0.86	0.26
ChemProp3D	200	CoV-2 3CL	0.66	0.20
CP3D-NDU	200	CoV-2 3CL	0.901	0.413
SchNetFeatures	average neighbors	CoV-2 3CL	0.84	0.29
ChemProp3D	average neighbors	CoV-2 3CL	0.73	0.31
CP3D-NDU	average neighbors	CoV-2 3CL	0.916	**0.467**
ConfNet	{20, 40}	CoV-2 3CL	0.493	0.128
ConAN	10	CoV-2 3CL	0.881 ± 0.009	0.317 ± 0.052
ConAN-FGW	5	CoV-2 3CL	**0.918 ± 0.012**	0.423 ± 0.045
SchNetFeatures	200	CoV-2	0.63	0.037
ChemProp3D	200	CoV-2	0.53	0.032
CP3D-NDU	200	CoV-2	0.663	0.06
SchNetFeatures	average neighbors	CoV-2	0.61	0.027
ChemProp3D	average neighbors	CoV-2	0.56	**0.10**
CP3D-NDU	average neighbors	CoV-2	0.647	0.058
ConfNet	{20, 40}	CoV-2	0.501 ± 0.001	0.36 ± 0.2
ConAN	10	CoV-2	0.634 ± 0.053	0.031 ± 0.023
ConAN-FGW	10	CoV-2	**0.6735 ± 0.032**	0.036 ± 0.014

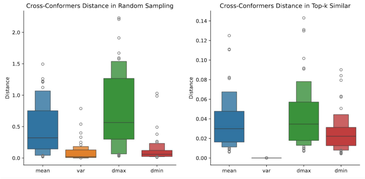
Performance on molecular classification tasks

	SchNet	DimeNet++	GemNet	PainNN	ClofNet	LEFTNet	ConAN-FGW[1]	ConAN-FGW[2]
Conf.	20	20	20	20	20	20	10	10
MAE ↓	1.9737	**1.4741**	1.6059	1.8744	2.0106	1.5276	*1.6047*	**1.4829**

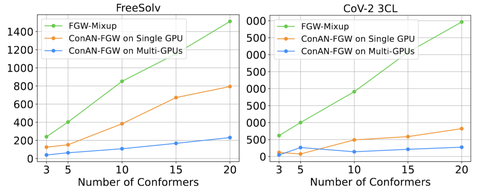
Performance of different conformer ensemble strategies on reaction molecules prediction.

Model	Lipo	ESOL	FreeSolv	BACE
ConAN (VisNet)	0.554 ± 0.448	1.025 ± 0.119	0.692 ± 0.032	0.612 ± 0.148
ConAN-FGW	0.495 ± 0.008	0.552 ± 0.052	**0.643 ± 0.015**	**0.469 ± 0.012**
ConAN (SchNet)	0.556 ± 0.013	0.571 ± 0.019	1.496 ± 0.158	0.635 ± 0.051
ConAN-FGW	**0.487 ± 0.031**	**0.529 ± 0.022**	1.068 ± 0.083	0.549 ± 0.016

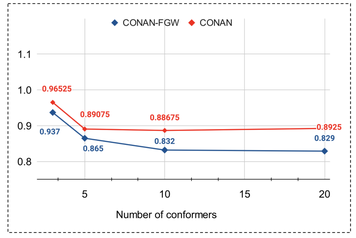
ConAN results VisNet and SchNet on regression tasks



Conformer distance distribution



Solver running time comparison.



Ablation study on the number conformers on the FGW barycenter component