FACET: A Fragment-Aware Conformer Ensemble Transformer

Duy M. H. Nguyen[1,2,3], Trung Q. Nguyen[3], Ha T. H. Le[3], Mai Thanh Nhat Truong[3] , TrungTin Nguyen[4,5], Nhat Ho[6], Khoa D Doan[7]

Duy Duong-Tran[8], Li Shen[8] , Daniel Sonntag[3,9], James Zou[10], Mathias Niepert[1, 2], Hyojin Kim[11], Jonathan E Allen[11]

[1]Max Planck Research School for Intelligent Systems (IMPRS-IS) [2] University of Stuttgart [3]German Research Center for Artificial Intelligence (DFKI)

[4]ARC Centre of Excellence for the Mathematical Analysis of Cellular Systems [5]School of Mathematical Sciences, Queensland University of Technology

[6]University of Texas at Austin [7] VinUniversity [8]University of Pennsylvania [9]Oldenburg University [10] Stanford University [11] Lawrence Livermore National Labs

Correspondence to: Trung Q. Nguyen <trung.nguyen@dfki.de>, Duy M. H. Nguyen <hong01@dfki.de>

Motivation and Challenges

Motivation

- 2D graphs lack geometry
- single 3D conformer misses flexibility
- ensemble captures thermodynamic reality

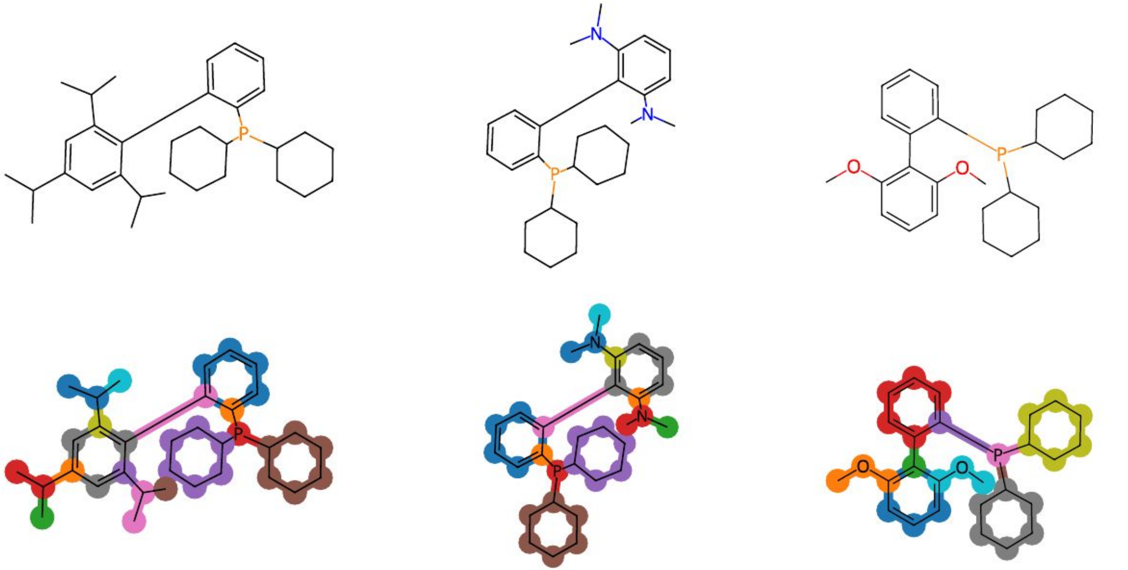
Challenges

- naive aggregation ignores geometry
- FGW alignment is accurate but expensive
- poor scalability to large datasets

Our Contributions

- replace costly FGW alignment with a trainable Graph Transformer
- introduce fragment-level structural priors
- achieve over 6x speedup compared to FGW baselines and SOTA results

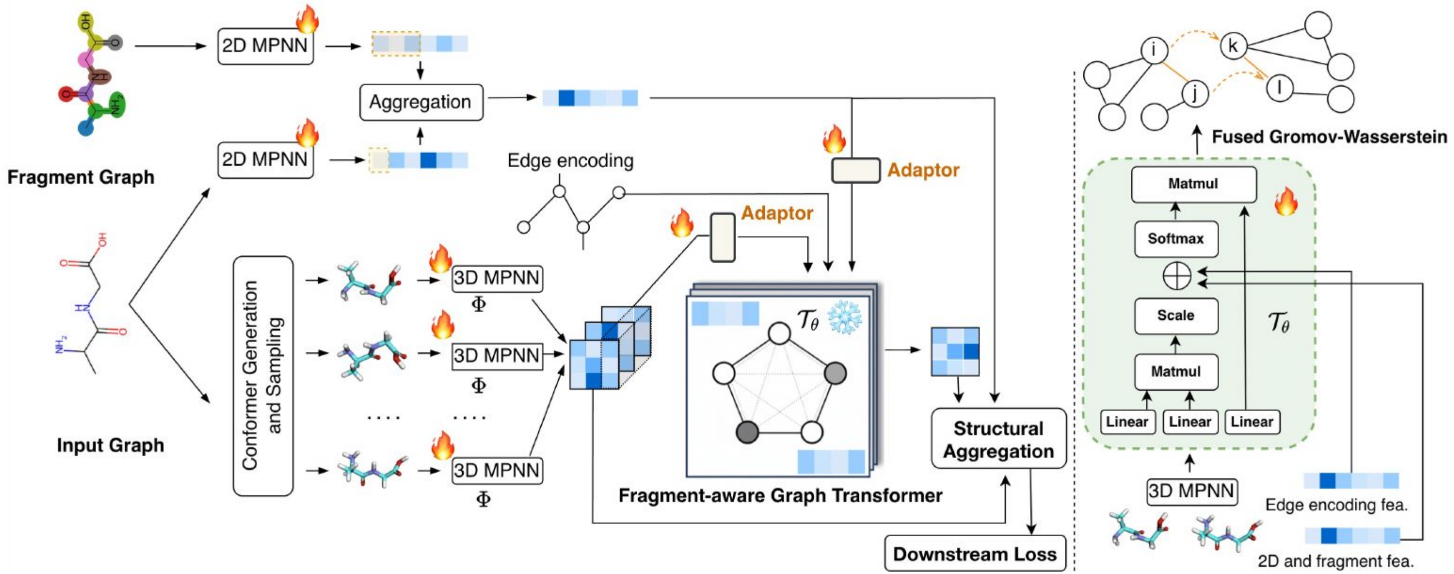
Fragment-Enhanced 2D Representation



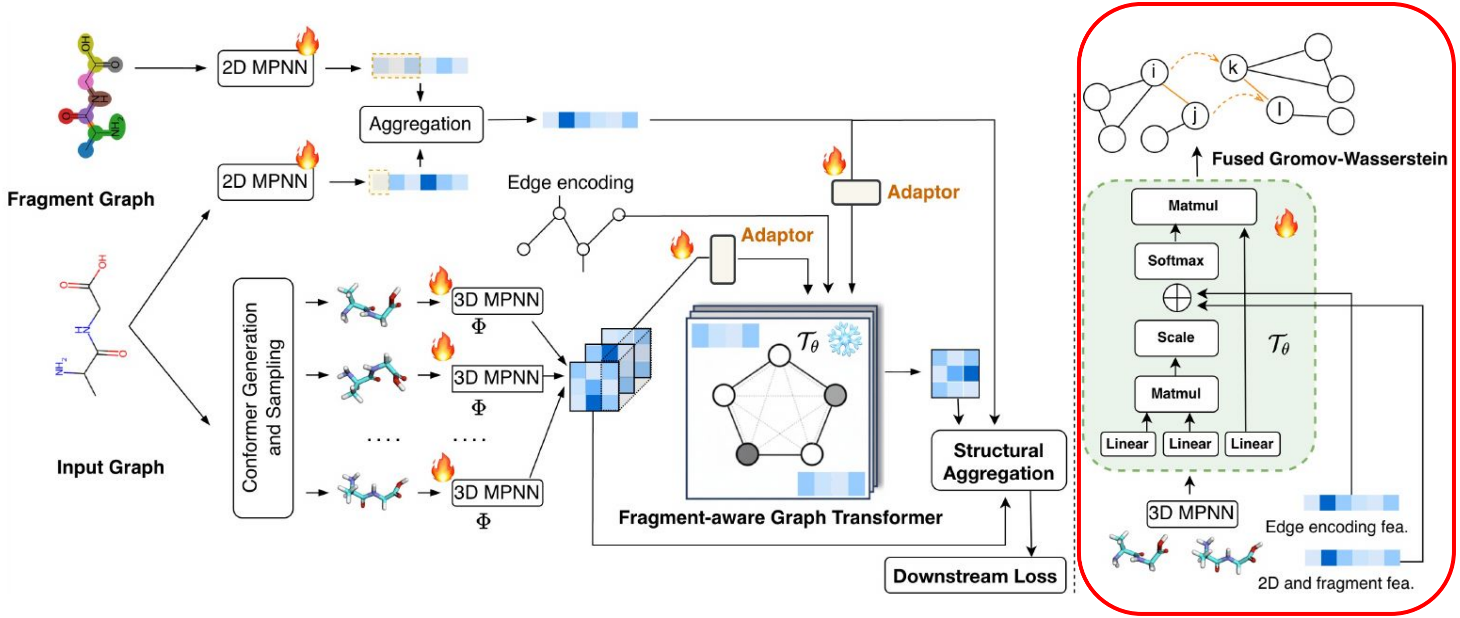
(a) Sample 1 (b) Sample 2 (c) Sample 3

Figure 6: RingsPaths decomposition on three samples of the **Kraken** dataset. Top: 2D molecules; bottom: corresponding RingsPaths decomposition results.

FACET Architecture Overview



Approximate FGW with a Graph Transformer



FGW Approximation Quality

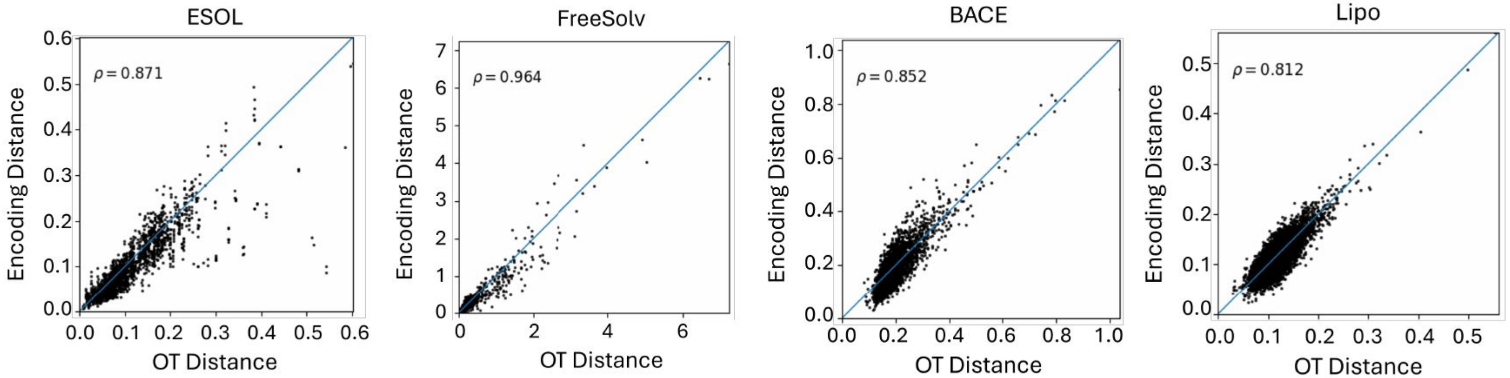
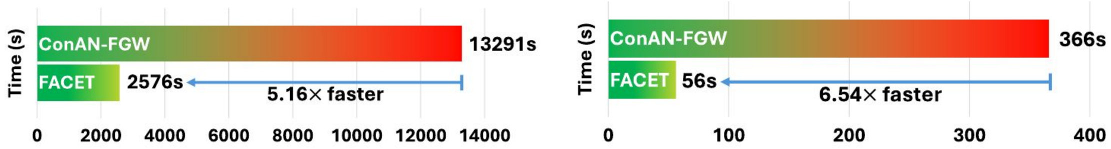


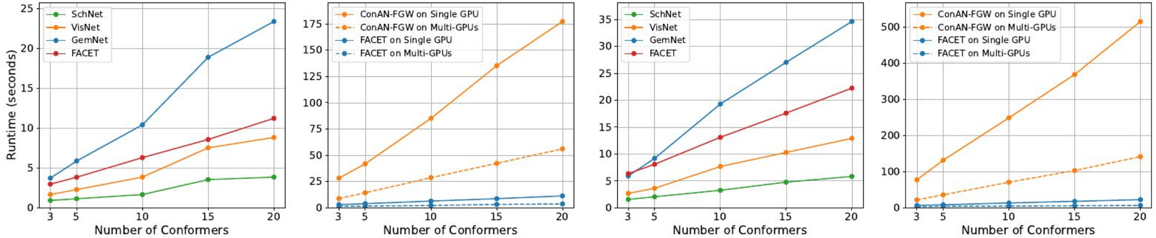
Figure 2: Correlations between FGW distance and trained GraphTransformer on four datasets in **MoleculeNet** benchmark. For each test molecule, we compute pairwise FGW distances between conformers and compare them with Euclidean distances between their Graph Transformer embeddings. The correlation ρ is reported, with the reference line $y = x$ shown in blue.

Scalability Results



(a) Drugs-75K (b) Kraken

Figure 3: Comparison of the **one-epoch training time** of CONAN-FGW (Nguyen et al., 2024b) and the proposed FACET on the Drugs-75K and Kraken datasets from the **MARCEL** benchmark.



(a) BACE (b) Lipo

Figure 4: **Inference running time comparison** between **FACET** and other **GNN-based methods** on two datasets, BACE (a) and Lipo (b). Results are shown for both single-GPU and 4-GPU (multi-GPU) configurations. Reported runtimes represent the total time required to extract structural embeddings for all molecules in the test set of each dataset.

SOTA Results — MoleculeNet

Table 3: Comparison of molecular property regression performance on the **MoleculeNet** benchmark (MSE ↓). The results of competing methods are adapted from Nguyen et al. (2024b). FACET uses a SchNet backbone.

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	0.534 ± 0.022	0.923 ± 0.045	4.213 ± 0.068	0.723 ± 0.021
Attentive FP	0.520 ± 0.001	0.771 ± 0.026	4.197 ± 0.193	-
PretrainGNN	0.545 ± 0.003	1.210 ± 0.005	6.392 ± 0.003	-
GROVER_large	0.676 ± 0.012	0.798 ± 0.018	5.162 ± 0.047	-
ChemBERTa-2*	0.639 ± 0.006	0.795 ± 0.033	-	1.858 ± 0.029
ChemRL-GEM	0.486 ± 0.008	0.706 ± 0.061	3.924 ± 0.436	-
MolFormer	0.492 ± 0.012	0.766 ± 0.026	5.485 ± 0.045	1.091 ± 0.021
ConfNet	1.360 ± 0.038	2.115 ± 0.484	-	1.329 ± 0.042
UniMol	**0.374** ± **0.012**	0.741 ± 0.014	2.867 ± 0.186	-
SchNet-scalar	0.704 ± 0.032	0.672 ± 0.027	1.608 ± 0.158	0.723 ± 0.100
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 ± 0.170
CONAN	0.556 ± 0.013	0.571 ± 0.019	1.496 ± 0.158	0.635 ± 0.051
CONAN-FGW	<u>0.422 ± 0.016</u>	<u>0.529 ± 0.022</u>	<u>1.068 ± 0.083</u>	<u>0.549 ± 0.016</u>
FACET	0.424 ± 0.009	**0.516** ± **0.044**	**0.967** ± **0.082**	**0.495** ± **0.034**

SOTA Results — MARCEL

Table 4: Comparison of molecular property regression performance on the **MARCEL** benchmark (MAE ↓). The results of competing methods are adapted from Zhu et al. (2024a).

Category	Model	Drugs-75K			Kraken			
		IP	**EA**	χ	$\mathbf{B_5}$	**L**	$\mathbf{BurB_5}$	**BurL**
2D models	GIN	0.4354	0.4169	0.2260	0.3128	0.4003	0.1719	0.1200
	GIN+VN	0.4361	0.4169	0.2267	0.3567	0.4344	0.2422	0.1741
	ChemProp	0.4595	0.4417	0.2441	0.4850	0.5452	0.3002	0.1948
	GraphGPS	0.4351	0.4085	0.2212	0.3450	0.4363	0.2066	0.1500
3D models	SchNet	0.4394	0.4207	0.2243	0.3293	0.5458	0.2295	0.1861
	DimeNet++	0.4441	0.4233	0.2436	0.3510	0.4174	0.2097	0.1526
	GemNet	0.4069	0.3922	**0.1970**	0.2789	0.3754	0.1782	0.1635
	PaiNN	0.4505	0.4495	0.2324	0.3443	0.4471	0.2395	0.1673
	ClofNet	0.4393	0.4251	0.2378	0.4873	0.6417	0.2884	0.2529
	LEFTNet	0.4174	0.3964	0.2083	0.3072	0.4493	0.2176	0.1486
Ensemble Strategy with DeepSets	SchNet	0.4452	0.4232	0.2243	0.2704	0.4322	0.2024	0.1443
	DimeNet++	0.4126	0.3944	0.2267	0.2630	0.3468	0.1783	0.1185
	GemNet	0.4066	0.3910	0.2027	0.2313	**0.3386**	0.1589	**0.0947**
	PaiNN	0.4466	0.4269	0.2294	**0.2225**	0.3619	0.1693	0.1324
	ClofNet	0.4280	0.4033	0.2199	0.3228	0.4485	0.2178	0.1548
	LEFTNet	0.4149	0.3953	0.2069	0.2644	0.3643	0.2017	0.1386
FACET	SchNet	0.4235	0.3971	0.2155	0.2508	0.3982	0.1803	0.1245
	GemNet	**0.3891**	**0.3852**	**0.1970**	**0.2225**	0.3402	**0.1503**	0.0952

Thank you

Codes are available at https://github.com/duyhominhnguyen/FACET

Affiliations

Deutsches Forschungszentrum für Künstliche Intelligenz German Research Center for Artificial Intelligence
Universität Stuttgart
imprs-is
Carl von Ossietzky Universität Oldenburg
Lawrence Livermore National Laboratory
Penn University of Pennsylvania
MACSYS
TEXAS The University of Texas at Austin
QUT Queensland University of Technology
VinUniversity