GeomGCL: Geometric Graph Contrastive Learning for Molecular Property Prediction

Shuangli Li,[1, 2*] Jingbo Zhou,[2] Tong Xu,[1] Dejing Dou,[2] Hui Xiong[3]

[1]University of Science and Technology of China, [2]Business Intelligence Lab, Baidu Research
[3]Artificial Intelligence Thrust, The Hong Kong University of Science and Technology
lsl1997@mail.ustc.edu.cn, {zhoujingbo, doudejing}@baidu.com, tongxu@ustc.edu.cn, xionghui@ust.hk

https://github.com/agave233/GeomGCL

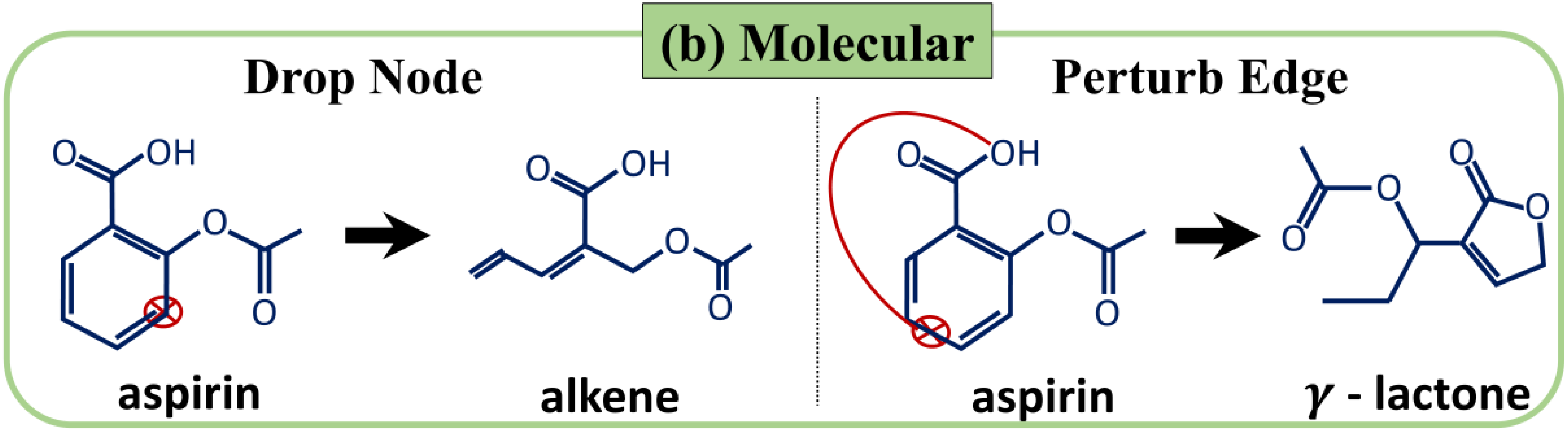
2022. 12. 22 • ChongQing

2022_AAAI

Leibniz-Institut für Sozialwissenschaften

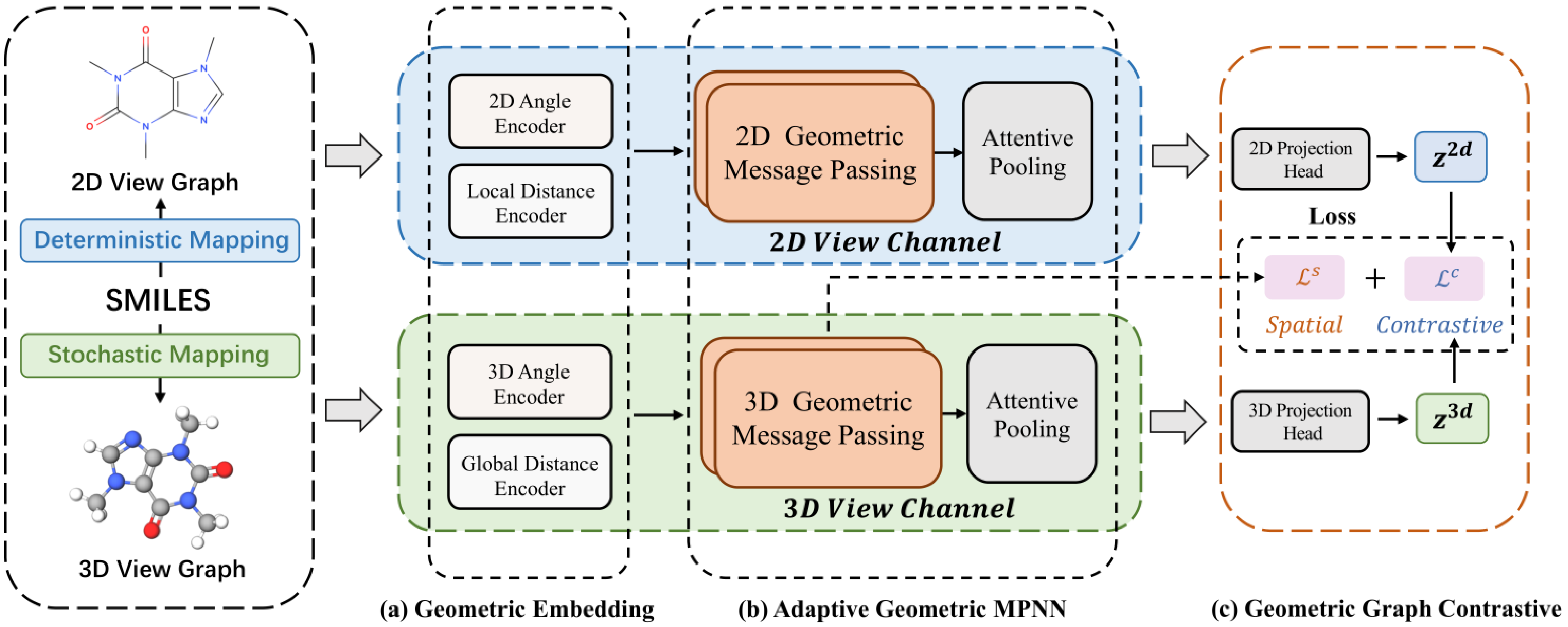
Reported by JiaWei Cheng

Introduction

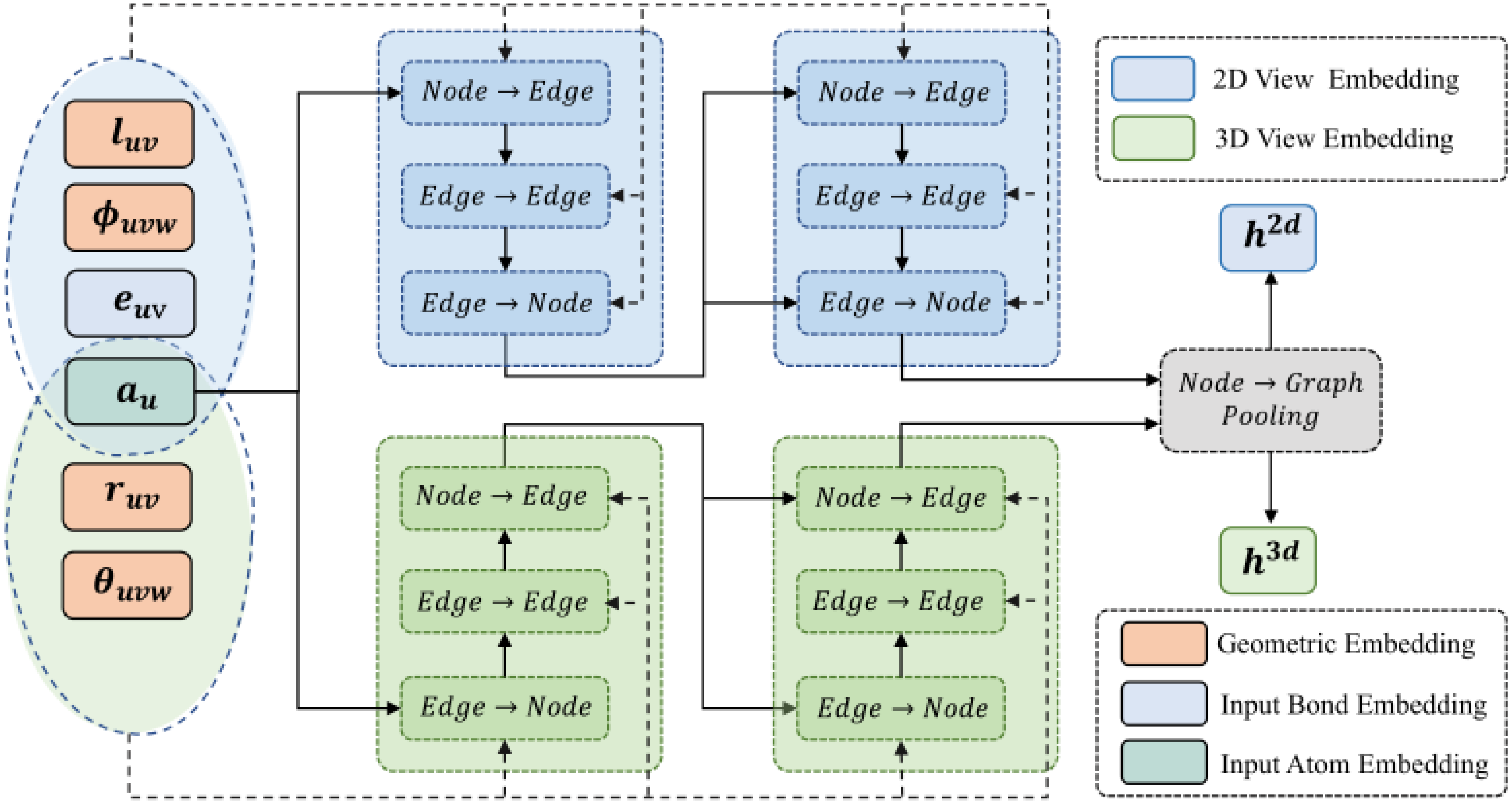


[2112.02472] Augmentation-Free Self-Supervised Learning on Graphs (arxiv.org)

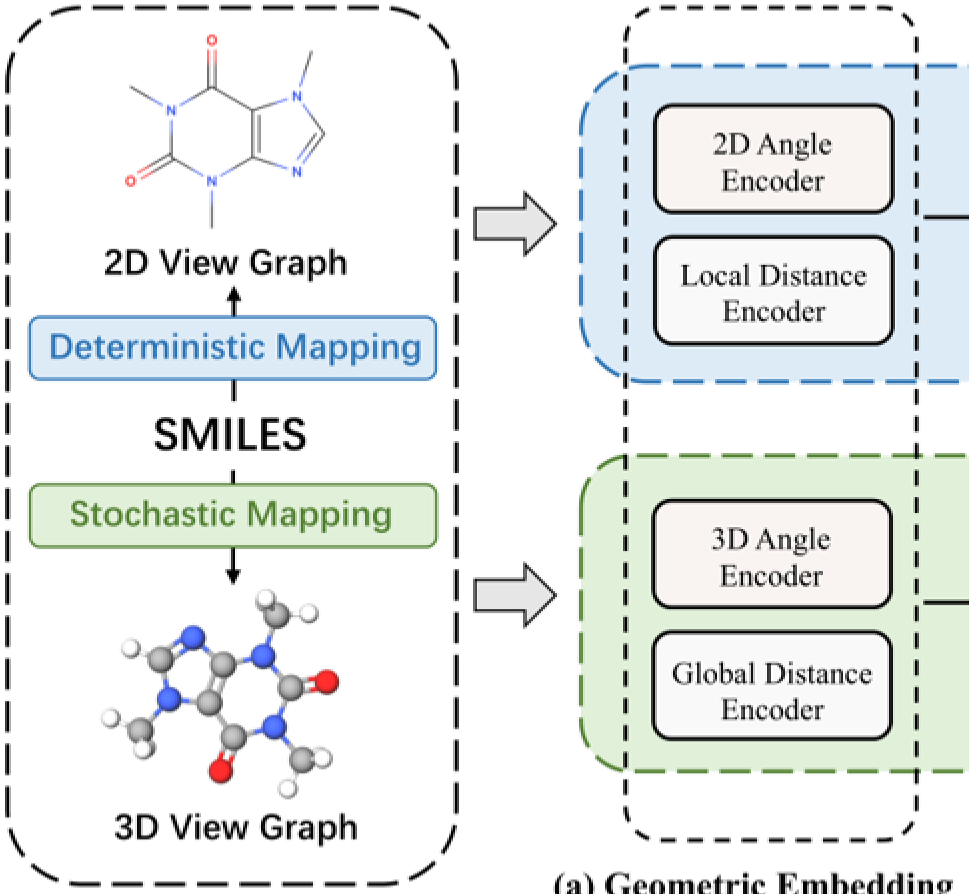
Overview


2D View Graph
Deterministic Mapping
SMILES
Stochastic Mapping
3D View Graph
2D Angle Encoder
Local Distance Encoder
3D Angle Encoder
Global Distance Encoder
2D Geometric Message Passing
Attentive Pooling
2D View Channel
3D Geometric Message Passing
Attentive Pooling
3D View Channel
2D Projection Head
z^{2d}
Loss
$\mathcal{L}^s$ + $\mathcal{L}^c$
Spatial
Contrastive
3D Projection Head
z^{3d}
(a) Geometric Embedding
(b) Adaptive Geometric MPNN
(c) Geometric Graph Contrastive

Overview


l_{uv}
ϕ_{uvw}
e_{uv}
a_u
r_{uv}
θ_{uvw}
Node → Edge
Edge → Edge
Edge → Node
Node → Edge
Edge → Edge
Edge → Node
Node → Edge
Edge → Edge
Edge → Node
Node → Edge
Edge → Edge
Edge → Node
2D View Embedding
3D View Embedding
h^{2d}
Node → Graph Pooling
h^{3d}
Geometric Embedding
Input Bond Embedding
Input Atom Embedding

Method



(a) Geometric Embedding

$$\boldsymbol{l} = RBF(l) = \overset{K}{\underset{k=1}{\frown}} \exp\big(-\beta_l(\exp(-l) - \mu_{l,k})^2\big) \quad (1)$$

$$\boldsymbol{r} = RBF(r) = \overset{K}{\underset{k=1}{\frown}} \exp\big(-\beta_r(\exp(-r) - \mu_{r,k})^2\big) \quad (2)$$

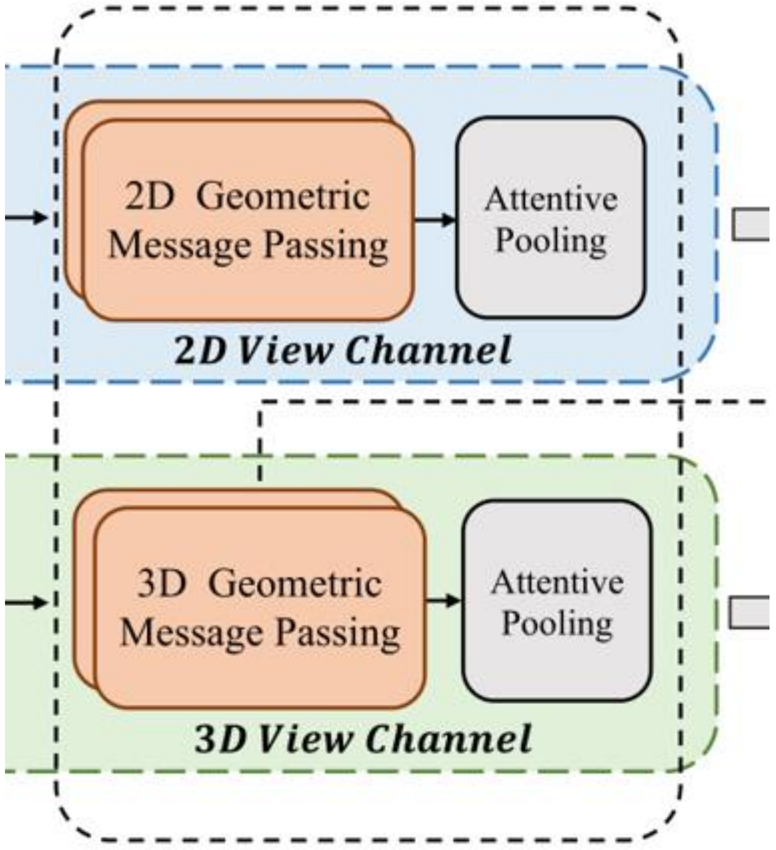
$$\boldsymbol{\phi} = RBF(\phi) = \overset{K}{\underset{k=1}{\frown}} \exp\big(-\beta_\phi(\phi - \mu_{\phi,k})^2\big) \quad (3)$$

$$\boldsymbol{\theta} = RBF(\theta) = \overset{K}{\underset{k=1}{\frown}} \exp\big(-\beta_\theta(\theta - \mu_{\theta,k})^2\big) \quad (4)$$

the K central points $\{\mu_{*,k}\}$ are uniformly selected between $\exp(-*)$ ($*$ is l or d) and 1, while $\beta_* = (\frac{2}{K}(1 - \exp(-*))^{-2}$.

For 2D and 3D angles, each $\mu_{*,k}$ is between 0 and π with $\beta_* = (\frac{2\pi}{K})^{-2}$, where $*$ denotes ϕ or θ .

Method



(b) Adaptive Geometric MPNN

Node→Edge Message Passing

$$\boldsymbol{e}_{uv}^{2d,t} = MLP(\boldsymbol{a}_u^{2d,t-1} \parallel \boldsymbol{a}_v^{2d,t-1} \parallel \boldsymbol{e}_{uv}^0) \tag{5}$$

$$\boldsymbol{e}_{uv}^{3d,t} = MLP(\boldsymbol{a}_u^{3d,t-1} \parallel \boldsymbol{a}_v^{3d,t-1} \parallel \boldsymbol{r}_{uv}) \tag{6}$$

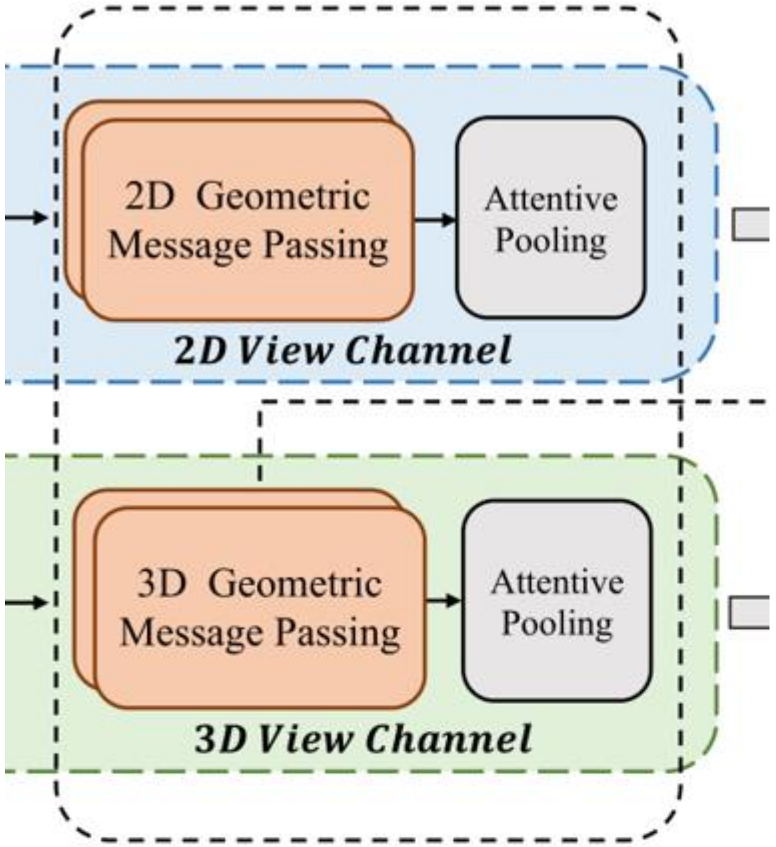
Edge→Edge Message Passing

$$\boldsymbol{e}_{uv}^{2d,t} = \sum_{e_{wu} \in \mathcal{A}(e_{uv})} W_\phi^t \boldsymbol{\phi}_{wuv} \odot (W_e^t \boldsymbol{e}_{uv}^{2d,t}) \tag{7}$$

$$\boldsymbol{e}_{uv,i}^{3d,t} = \sum_{e_{wu} \in \mathcal{A}_i(e_{uv})} W_{\theta,i}^t \boldsymbol{\theta}_{wuv} \odot (W_{e,i}^t \boldsymbol{e}_{uv}^{3d,t}) \tag{8}$$

$$\boldsymbol{e}_{uv}^{3d,t} = \Big\Vert_{i=1}^{n} \mathrm{Pool}\left(\{\boldsymbol{e}_{uv,i}^{3d,t} | 1 \leq i \leq n\}\right) \odot \boldsymbol{e}_{uv,i}^{3d,t} \tag{9}$$

Method



(b) Adaptive Geometric MPNN

Edge→Node Message Passing

$$\boldsymbol{a}_v^{2d,t} = \sum_{e_{uv} \in \mathcal{D}(a_v)} W_l^t \boldsymbol{l}_{uv} \odot (W_a^t \boldsymbol{e}_{uv}^{2d,t}) \tag{10}$$

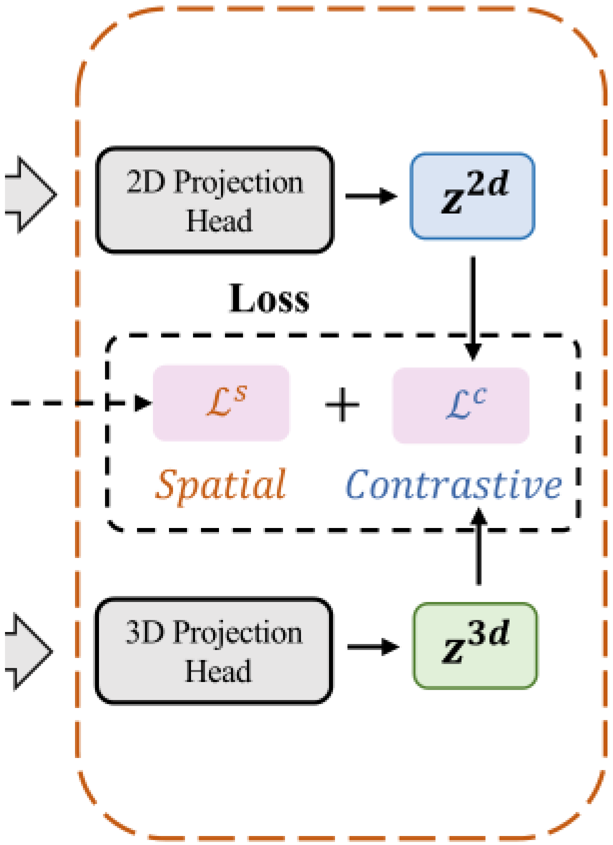
$$\boldsymbol{a}_v^{3d,t} = \Big\Vert_{i=1}^{m} \sum_{e_{uv} \in \mathcal{D}_i(a_v)} W_{r,i}^t \boldsymbol{r}_{uv} \odot (W_{a,i}^t \boldsymbol{e}_{uv}^{3d,t}) \tag{11}$$

Node→Graph Attentive Pooling

$$\boldsymbol{g}^t = \sum_{a_v \in \mathcal{V}} \mathrm{softmax}(\boldsymbol{h}^t \parallel \boldsymbol{a}_v) W_g^t \boldsymbol{a}_v \tag{12}$$

$$\boldsymbol{h}^{t+1} = \mathrm{GRU}(\boldsymbol{h}^t, \boldsymbol{g}^t),\ t = 0, 1, .., T_g \tag{13}$$

Method



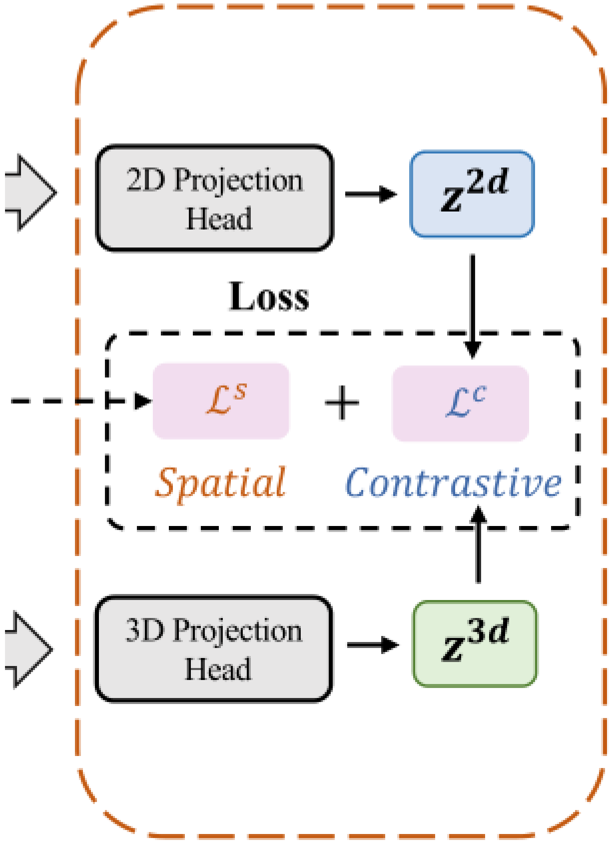
(c) Geometric Graph Contrastive

$$\boldsymbol{z}_i^{2d} = MLP(\boldsymbol{h}_i^{2d}), \quad \boldsymbol{z}_i^{3d} = MLP(\boldsymbol{h}_i^{3d}) \tag{14}$$

$$\begin{aligned} \mathcal{L}_i^c &= \mathcal{L}_i^{2d,c} + \mathcal{L}_i^{3d,c} \\ &= -\log \frac{e^{\langle \boldsymbol{z}_i^{2d}, \boldsymbol{z}_i^{3d} \rangle / \tau}}{\sum_{j=1}^{N} e^{\langle \boldsymbol{z}_i^{2d}, \boldsymbol{z}_j^{3d} \rangle / \tau}} - \log \frac{e^{\langle \boldsymbol{z}_i^{3d}, \boldsymbol{z}_i^{2d} \rangle / \tau}}{\sum_{j=1}^{N} e^{\langle \boldsymbol{z}_i^{3d}, \boldsymbol{z}_j^{2d} \rangle / \tau}} \end{aligned} \tag{15}$$

$$\mathcal{L}^s = \sum_{t=1}^{T} \sum_{i=1}^{n-1} \| W_{\theta,i+1}^t - W_{\theta,i}^t \|^2 \tag{16}$$

Method



(c) Geometric Graph Contrastive

$$\hat{y} = MLP\Big(MLP\big(f_{2d}(\mathcal{G}^{2d})\big) + MLP\big(f_{3d}(\mathcal{G}^{3d})\big)\Big) \quad (18)$$

$$\mathcal{L}_{cls} = \mathcal{L}_{ce}(\hat{y}, y) + \lambda \mathcal{L}^s \quad (19)$$

$$\mathcal{L}_{reg} = \mathcal{L}_1(\hat{y}, y) + \lambda \mathcal{L}^s \quad (20)$$

Experiment

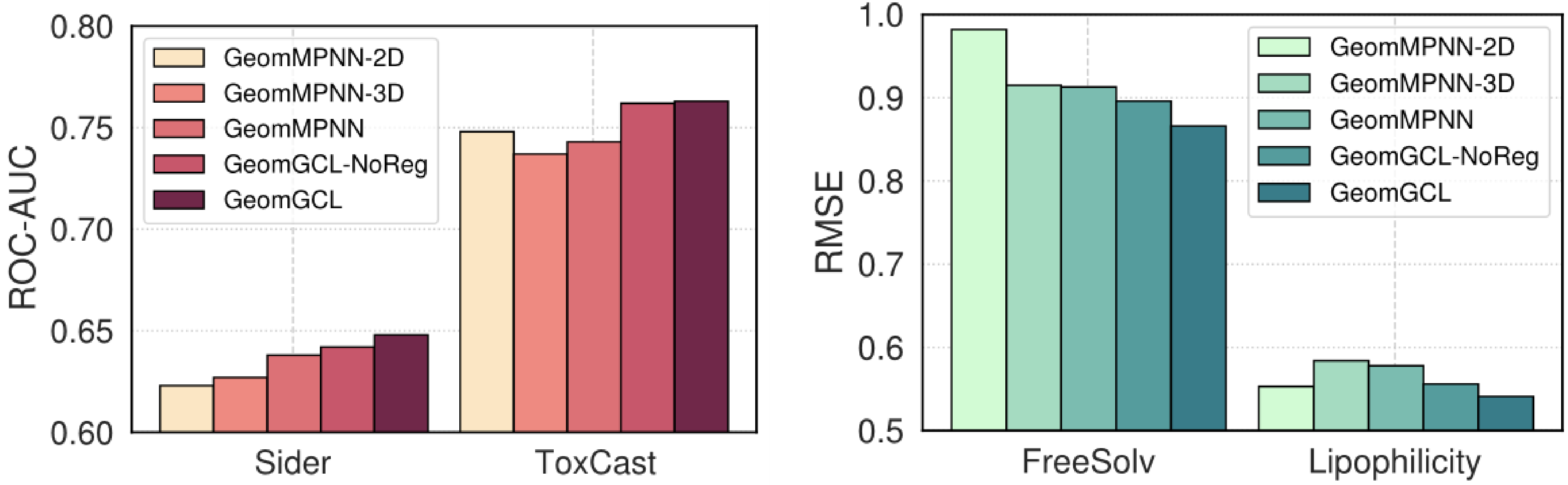
Dataset	# Tasks	Task Type	# Molecules
ClinTox	2	Classification	1484
Sider	27	Classification	1427
Tox21	12	Classification	7831
ToxCast	617	Classification	8597
ESOL	1	Regression	1128
FreeSolv	1	Regression	643
Lipophilicity	1	Regression	4200

Table 1: Statistics of seven molecular datasets.

Experiment

Model	Graph Classification (ROC-AUC) ↑					Graph Regression (RMSE) ↓			
	ClinTox	Sider	Tox21	ToxCast	Cls.Ave	ESOL	FreeSolv	Lipophilicity	Reg.Ave
AttentiveFP	0.808	0.605	0.835	0.743	0.748	0.578	1.034	0.602	0.738
DMPNN	0.886	0.637	0.848	0.743	0.779	0.647	1.092	0.591	0.777
CoMPT	0.877	0.626	0.836	0.755	0.774	0.589	1.103	0.590	0.761
SGCN	0.825	0.560	0.769	0.656	0.703	1.329	2.061	1.075	1.488
MAT	0.898	0.619	0.834	0.735	0.772	0.624	1.059	0.705	0.796
HMGNN	0.680	0.607	0.794	0.702	0.696	0.701	1.207	0.720	0.876
DimeNet	0.760	0.615	0.780	0.645	0.7000	0.633	0.978	0.614	0.742
InfoGraph	0.781	0.585	0.793	0.705	0.716	0.914	2.104	0.845	1.288
MoCL	0.739	0.629	0.824	0.718	0.727	0.934	1.478	0.742	1.051
GeomMPNN	0.900	0.638	0.838	0.743	0.780	**0.555**	0.913	0.578	0.682
GeomGCL	**0.919**	**0.648**	**0.850**	**0.763**	**0.796**	0.575	**0.866**	**0.541**	**0.661**

Experiment



(a) Results for Classification (b) Results for Regression

Figure 4: Evaluation of GeomGCL with its variants.

Experiment

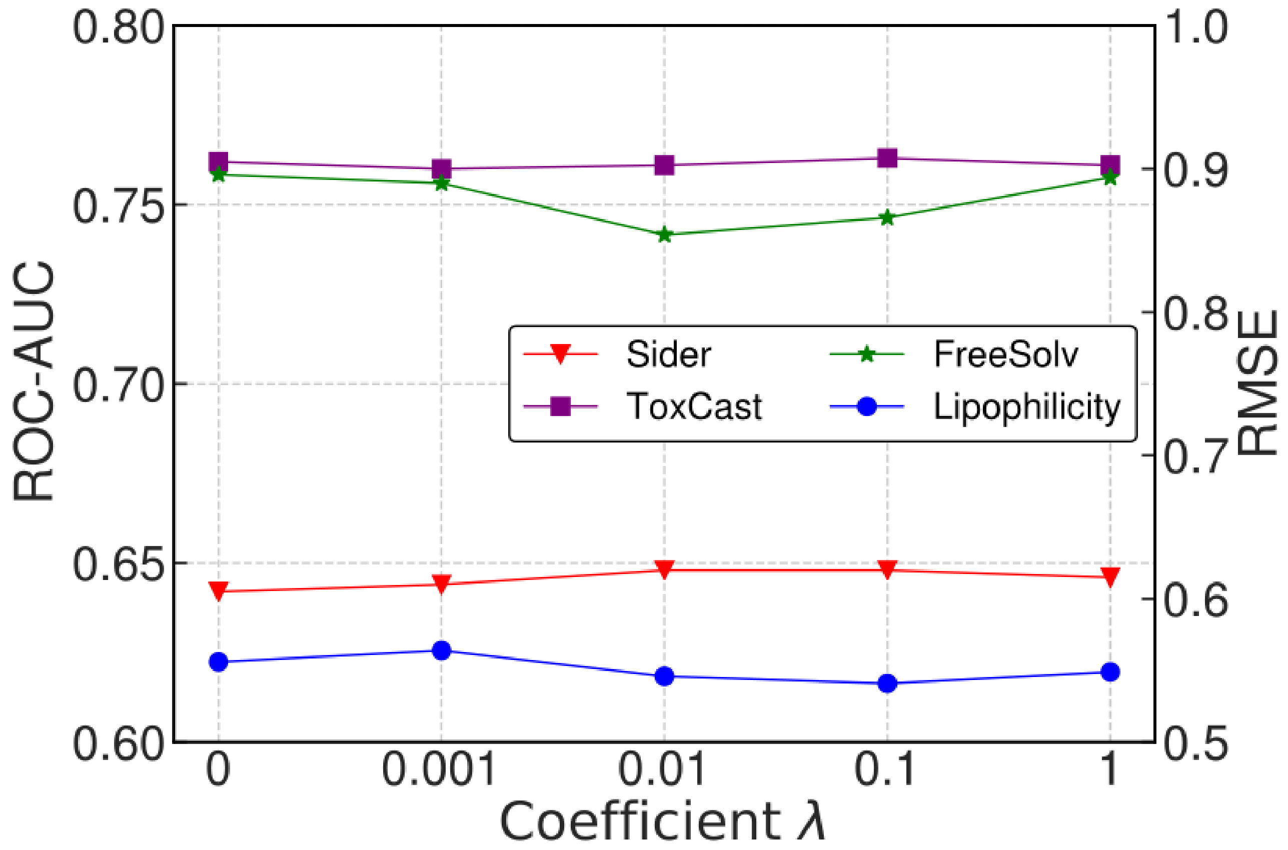


Figure 5: Analysis for the balancing parameter of $\mathcal{L}^s$

Thank you!