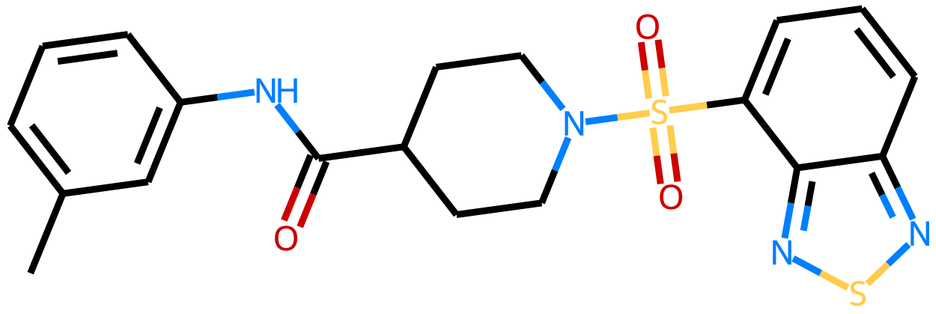
An End-to-End framework for Molecular Conformation Generation via Bilevel Programming

Minkai Xu[1,2], Wujie Wang[3], Shitong Luo[4], Chence Shi[1,2],
Yoshua Bengio[1,2], Rafael Gómez-Bombarell[1,3], Jian Tang[1,5]

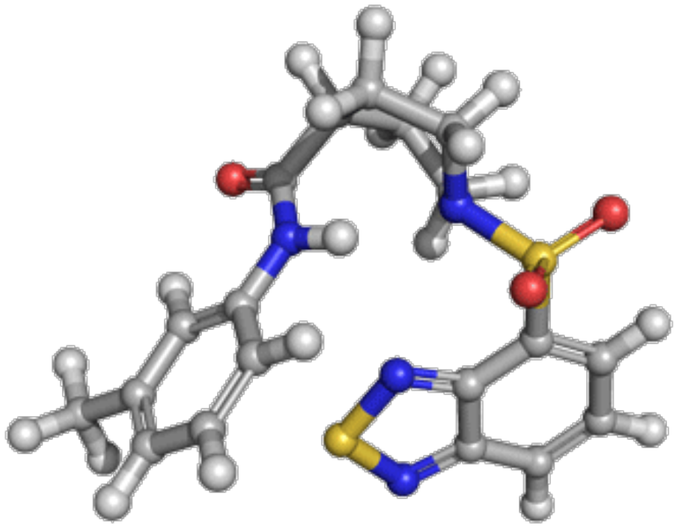
Conformation Prediction

- For real-world molecules, computing 3D structures is expensive
- We study how to predict valid and stable conformations from molecular graph
  - Molecular graph $\mathcal{G}$: 2D atom-bond graph
  - Conformation $\boldsymbol{R}$: atomic 3D coordinates



$\mathcal{G}$

$p(\boldsymbol{R}|\mathcal{G})$



$\boldsymbol{R} \propto \underbrace{\exp(-E(\boldsymbol{R})/k_B T)}_{\text{Boltzmann distribution}}$

Limitation of previous works & Motivation

- Likelihood of conformations is not rotation and translation invariant[1]. Distance based methods[2,3] learn to generate outputs (distances) as the intermediate variables of the desired object (atomic coordinates)
- This motivates us to pursue an algorithm that (C1) learns to generate conformations in an end-to-end fashion, and (C2) preserves the roto-translation equivariance of conformations.

[1]Mansimov, Elman, et al. "Molecular geometry prediction using a deep generative graph neural network." *Scientific reports* 9.1 (2019): 1-13.
[2]Simm, Gregor NC, and José Miguel Hernández-Lobato. "A generative model for molecular distance geometry." *arXiv preprint arXiv:1909.11459* (2019).
[3]Minkai Xu, Shitong Luo, Yoshua Bengio, Jian Peng, and Jian Tang. Learning neural generative dynamics for molecular conformation generation. ICLR 2021

Solution

- Learn $\mathrm{p}(\boldsymbol{R}|\mathcal{G})$ in an end-to-end manner, thus we can take the error of distance geometry into account during training!
- We formulate learning $\mathrm{p}(\boldsymbol{R}|\mathcal{G})$ as a bilevel program:
  - In bilevel program, we have two objectives F and H (outer and inner objective), and the corresponding outer and inner variables θ and w:

$$\min_{\theta} F(w_\theta) \ \text{ such that } \ w_\theta \in \arg\min_{w} H(w, \theta)$$

  - We can get w though T steps optimization

$$w_{\theta,T} = \Phi(w_{\theta,T-1}, \theta) = \Phi(\Phi(w_{\theta,T-2}, \theta), \theta)\text{o on}$$

  - Then the meta-gradient dF/dθ can be computed through the optimization path to optimize the meta-parameters θ

$$\nabla_\theta F(w_{\theta,T}) = \partial_w F(w_{\theta,T}) \nabla_\theta w_{\theta,T}$$

Solution

- We formulate learning $\mathrm{p}(\boldsymbol{R}|\mathcal{G})$ as a bilevel program:
  - Meta parameters: $p_\theta(\boldsymbol{d}|\mathcal{G})$ to generate the distance $\boldsymbol{d}_{\theta,\phi} = D_\theta(z_\phi, \mathcal{G})$
  - Inner loop: solve $\boldsymbol{R}$ from $\boldsymbol{d}$:

$$\begin{aligned}\boldsymbol{R}_{\theta,\phi} &= \arg\min_{\boldsymbol{R}} H(\boldsymbol{R}, D_\theta(z_\phi, \mathcal{G})) \\ &= \arg\min_{\boldsymbol{R}} H(\boldsymbol{R}, \boldsymbol{d}_{\theta,\phi}) \\ &= \arg\min_{\boldsymbol{R}} \Big\{ \sum_{e_{uv}\in\mathcal{E}} \big(\|\boldsymbol{r}_u - \boldsymbol{r}_v\|_2 - d_{uv}\big)^2 \Big\}\end{aligned}$$

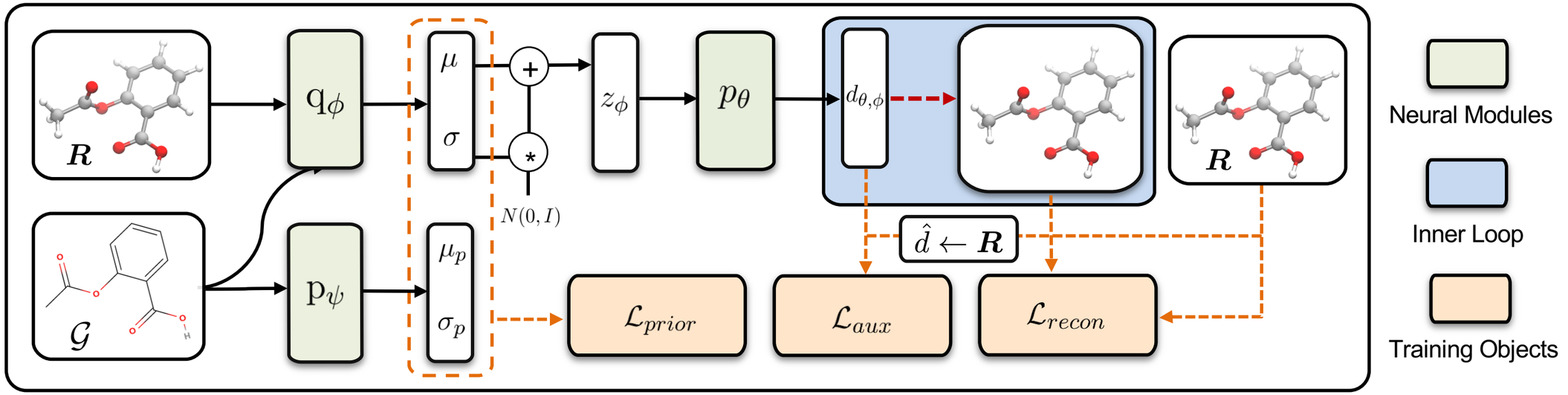
  - Outer loop: maximize the likelihood of $\mathrm{p}(\boldsymbol{R}|\mathcal{G})$ (reconstruction in VAE)
    - First align the reference conformation R*
    - Then compute the RMSD (root-mean-square deviation): $\mathrm{RMSD}(\boldsymbol{R}, \hat{\boldsymbol{R}}) = \Big(\frac{1}{n}\sum_{i=1}^{n} \|\boldsymbol{R}_i - \hat{\boldsymbol{R}}_i\|^2\Big)^{\frac{1}{2}}$

$$\begin{aligned}F(\boldsymbol{R}_{\theta,\phi}) &= \log p_\theta(\boldsymbol{R}|z, \mathcal{G}) \\ &= -\sum_{i=1}^{n}\sum_{j=1}^{3} (\boldsymbol{R}_{ij} - A(\boldsymbol{R}, \boldsymbol{R}^*)_{ij})^2\end{aligned}$$

Whole framework

- Outer objective:
  - Encoder q(z|G,R); Prior p(z|G); Decoder p(d|z,G)
  - Lrecon: first term (reconstruction loss)
  - Lprior: second term (prior regularization loss)
  - Laux: auxiliary term (an additional supervision on the distances)

$$\log P(\boldsymbol{R}|\mathcal{G}) \geq \mathbb{E}_{z\sim q_\phi(z|\boldsymbol{R},\mathcal{G})}\left[\log p_\theta(\boldsymbol{R}|z,\mathcal{G})\right] - D_{\mathrm{KL}}\left[q_\phi(z|\boldsymbol{R},\mathcal{G})\|p_\psi(z|\mathcal{G})\right]$$



Experiments

- Data Sets
  - **GEOM**: > 33 million molecular conformers by MIT group, including both small molecules in QM9 and medium-sized drug-like molecules.
- Baselines
  - **CVGAE(Mansimov et al. 2019)**: learning atom representations with GNNs and then predict the coordinates of atoms
  - **GraphDG(Simm&Hernandez-Lobato, 2020)** and **CGCF (Xu et al., 2021)**: predicting the pairwise distances between atoms with GNNs and then generate conformers based on distances
  - **RDKit**: a classical Euclidean Distance Geometry-based approach

Evaluation Metrics

- Discrepancy between two conformations: Root-Mean-Square Deviation (RMSD)

$$\mathrm{RMSD}(\boldsymbol{R}, \hat{\boldsymbol{R}}) = \left(\frac{1}{n} \sum_{i=1}^{n} \| \boldsymbol{R}_i - \hat{\boldsymbol{R}}_i \|^2 \right)^{\frac{1}{2}}$$

- Coverage (COV): the fraction of conformations in the reference set that are matched by at least one conformation in the generated conformations

$$\mathrm{COV}(\mathbb{S}_g(\mathcal{G}), \mathbb{S}_r(\mathcal{G})) = \frac{1}{|\mathbb{S}_r|} \left| \left\{ \boldsymbol{R} \in \mathbb{S}_r \middle| \, \mathrm{RMSD}(\boldsymbol{R}, \boldsymbol{R}') < \delta, \boldsymbol{R}' \in \mathbb{S}_g \right\} \right|$$

- Matching (MAT): measure the average distance of the reference conformations with their nearest neighbors in the generated conformations

$$\mathrm{MAT}(\mathbb{S}_g(\mathcal{G}), \mathbb{S}_r(\mathcal{G})) = \frac{1}{|\mathbb{S}_r|} \sum_{\boldsymbol{R}' \in \mathbb{S}_r} \min_{\boldsymbol{R} \in \mathbb{S}_g} \mathrm{RMSD}(\boldsymbol{R}, \boldsymbol{R}').$$

Results

Dataset	GEOM-QM9				GEOM-Drugs			
Metric	COV* (%)		MAT (Å)		COV* (%)		MAT (Å)	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median
CVGAE	8.52	5.62	0.7810	0.7811	0.00	0.00	2.5225	2.4680
GraphDG	55.09	56.47	0.4649	0.4298	7.76	0.00	1.9840	2.0108
CGCF	69.60	70.64	0.3915	0.3986	49.92	41.07	1.2698	1.3064
ConfVAE-	75.57	80.76	0.3873	0.3850	51.24	46.36	1.2487	1.2609
ConfVAE	**77.98**	**82.82**	**0.3778**	**0.3770**	**52.59**	**56.41**	**1.2330**	**1.2270**
RDKit	79.94	**87.20**	0.3238	0.3195	65.43	70.00	1.0962	1.0877
CVGAE + FF	63.10	60.95	0.3939	0.4297	83.08	95.21	0.9829	0.9177
GraphDG + FF	70.67	70.82	0.4168	0.3609	84.68	93.94	0.9129	0.9090
CGCF + FF	73.52	72.75	0.3131	0.3251	**92.28**	98.15	0.7740	**0.7338**
ConfVAE- + FF	77.95	79.14	0.2851	0.2817	91.48	99.21	0.7743	0.7436
ConfVAE + FF	**81.46**	**83.80**	**0.2702**	**0.2709**	91.88	**100.00**	**0.7634**	**0.7312**

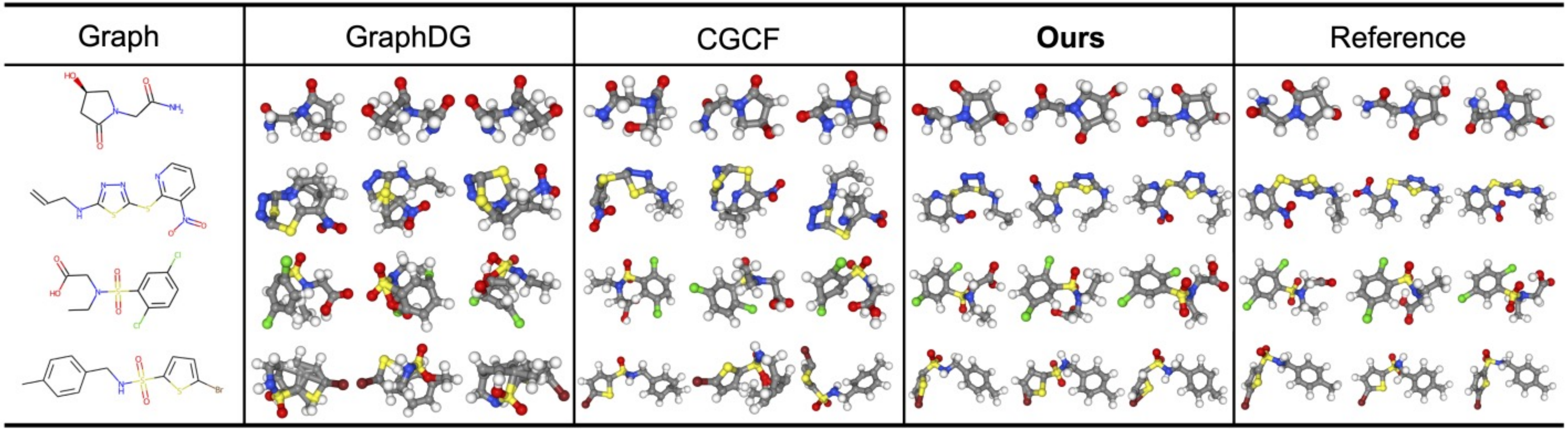
* For COV, the threshold δ is set as 0.5Å for QM9 and 1.25Å for Drugs following Xu et al. (2021).

ConfVAE-: ablation setting by removing the bilevel component

Refined by classical **Merck Molecular Force Field (MMFF)**

ConfVAE+FF: the first method that already practically useful (beat the rule-based RDKit baseline) when combined with MMFF, and achieves the state-of-the-art performance

Visualizations



Thank you!

Code is available at https://github.com/MinkaiXu/ConfVAE-ICML21
Also feel free to contact me later at xuminkai@mila.quebec