Graph-Based Reaction Prediction with 3D-Enhanced Graph Neural Network

Aaron Feng
Halıcıoğlu Data Science Institute
UC San Diego
zhf004@ucsd.edu

Zhicheng Wang
Department of Computer Science and Engineering
UC San Diego
zhw049@ucsd.edu

Abstract

We tackle the problem of chemical reaction outcome prediction by leveraging graph-based transformer models enhanced with three-dimensional (3D) information. Instead of relying solely on SMILES representations, we convert the molecular data into enriched graph representations and generate 3D conformers to capture stereochemical and regiochemical nuances. Our approach combines a two-dimensional (2D) molecular graph encoder with a 3D geometry-aware graph encoder—aligned via a contrastive InfoNCE objective—augmented by graph attention mechanisms and equivariant layers to respect the geometric invariances of molecular structures. The model is first pre-trained on molecules from the QM9 dataset to infuse the 2D encoder with 3D structural knowledge and then fine-tuned on reaction data from the Open Reaction Database (ORD). Qualitative examples indicate that incorporating 3D features helps the model learn richer representations, particularly in handling reactions with significant geometric constraints.

1 Introduction

Motivation. Chemical reaction prediction is a critical task in computational chemistry, with broad impacts on drug discovery, materials science, and synthetic chemistry. Accurate prediction of reaction products enables the construction of feasible organic synthesis routes, accelerating the design of new drugs and molecules. Recent advances in machine learning (ML) have shown promise – for instance, modern Transformer-based models can learn reaction patterns from data – yet most such models still represent molecules in SMILES (Simplified Molecular Input Line Entry System). These 2D methods fail to capture the spatial and stereochemical details, which can be crucial for certain reaction mechanisms. This gap in representation motivates the integration of 3D molecular information into reaction prediction models to achieve more generalized and accurate predictions

Challenges. Predicting the outcome of a chemical reaction is inherently challenging because it depends on more than just the 2D properties of atoms. For example, the conformation of a molecule affects which atoms are close enough to react, and the stereochemistry can dictate the formation of one product stereoisomer over another. 2D representations struggle to capture these phenomena, a SMILES or 2D graph may indicate which atoms could react, but the model cannot assess if those atoms can physically approach each other. Thus, purely 2D approaches often mispredict outcomes in cases requiring an understanding of spatial arrangement. In contrast, incorporating 3D molecular information directly addresses these issues. By using 3D conformations of reactant molecules, a model gains access to the true distances, angles, and orientations, allowing it to learn the relationships that govern the reaction process.

Our Solution. We propose a 3D-enhanced graph neural network for reaction prediction that combines 2D molecular graphs with 3D geometric information. Reactants are converted from SMILES to molecular graphs, and 3D conformers provide key spatial insights into molecular geometry and steric effects.

To improve spatial understanding, we pre-train a 2D MPNN and a 3D MPNN using contrastive learning on the QM9 dataset with an InfoNCE objective, following a modified version of the 3D Infomax approach (Stärk et al., 2022) to better suit our data. This alignment makes the 2D encoder geometry-aware even when only 2D graphs are used at inference time.

In the final model, the 2D MPNN encodes reactant graphs, whose embeddings are concatenated with reagent fingerprints and passed to a LSTM to generate product SMILES, capturing the transformation from reactants to products.

2 Related Work

2D-Based Molecular Property Prediction. Early methods in reaction outcome prediction and molecular property modeling relied on 2D molecular representations. Graph neural networks (GNNs) over molecular graphs have also shown good performance, for example the Weisfeiler-Lehman Network (WLN) encoder was used to predict reaction centers and rank candidate products and attained around 84% accuracy on benchmark reaction datasets (Jin et al., 2017). Transformers have also been applied to model chemical reactivity: the Molecular Transformer views reaction prediction as a machine translation problem on SMILES strings and obtains more than 90% top-1 accuracy on benchmark sets (Schwaller et al., 2019). These 2D models effectively capture chemical patterns and local connectivity but do not account for the actual three-dimensional conformation of molecules.

Incorporating 3D Molecular Information. To address the limitations of purely 2D models, researchers have added three-dimensional structural features into deep learning models. A method with which this was achieved is SchNet, introducing continuous-filter convolutional layers operated on 3D coordinates, insuring the network can learn from interatomic distances and encode quantum interactions in molecules (Schütt et al., 2017). With geometric information, SchNet established state-of-the-art performance on quantum chemistry benchmarks like QM9. Following this idea, DimeNet incorporated bond angles into the message-passing mechanism such that directionality between bonds is represented explicitly. This led to further improvements in the precision of property predictions, with DimeNet far outperforming earlier GNNs (Klicpera et al., 2020a; Klicpera et al., 2020b). These and analogous models demonstrate how adding interatomic distances and angles to molecular representations helps capture spatial interactions that are largely ignored by 2D graph methods.

Notably, 3D Infomax (Stärk et al., 2022) and GraphMVP (Liu et al., 2022) introduced multi-view contrastive frameworks that leverage a molecule's 2D graph and its 3D conformation.

3 Methods

3.1 Data Processing

Figure 1: SMILES to Molecular Graph Conversion using RDKit.

SMILES to Graph Conversion. All molecules are converted from SMILES into 2D molecular graphs using RDKit. In this graph representation, atoms become nodes and bonds become edges. Each atom node is encoded via a one-hot vector of its atomic number (up to the maximum atomic number in the dataset), and each bond is characterized by a one-hot vector of the bond type (e.g., single, double, triple, aromatic). For an undirected bond between atom i and atom j, we create two directed edges ($i \to j$ and $j \to i$), each carrying the same bond-type feature. If a molecule

contains no bonds (isolated atoms), we optionally add self-loops so that every atom can still exchange messages in the subsequent neural network.

3D Conformer Generation. To incorporate spatial information, we use the reference 3D coordinates from QM9 (already optimized) or generate low-energy conformations for molecules in ORD that lack provided coordinates. We rely on RDKit's ETKDG algorithm for conformation generation, yielding a single conformer per molecule. For each bond, we calculate the inter-atomic distance $d_{ij} = \|\mathbf{r}_i - \mathbf{r}_j\|$ where $\mathbf{r}_i$ and $\mathbf{r}_j$ are the 3D positions of atoms i and j. This distance is appended to the bond feature vector, turning a 4-dimensional one-hot bond-type vector into a 5-dimensional feature that includes bond length.

Reagents as Fingerprints. Because reagents are not consumed in the reaction, we handle them separately from reactants. Each reagent molecule is passed through a pre-trained network to obtain a 128-dimensional embedding. If multiple reagent molecules exist, we average their embeddings to form a single *reagent fingerprint*. During reaction prediction, this fingerprint is provided to the decoder alongside the reactant graph encoding.

3.2 Model Architecture

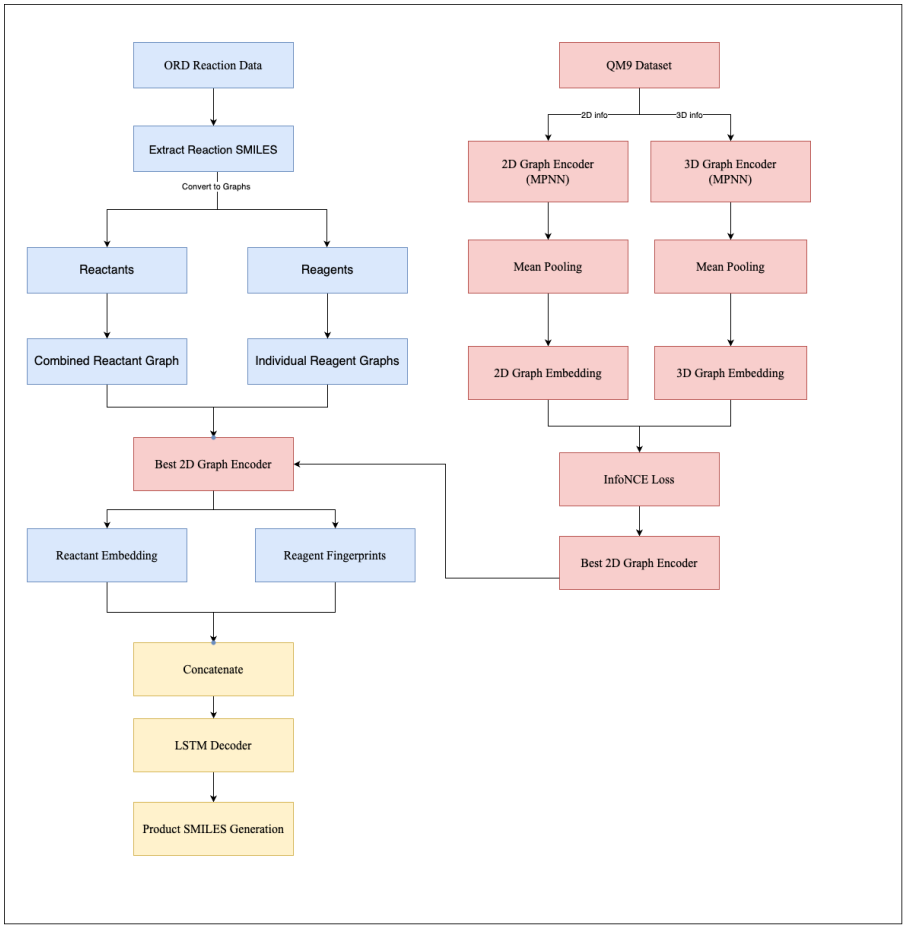


Figure 2: Architecture of the 3D-enhanced MPNN for reaction prediction. (a) **Pre-training:** Aligning 2D and 3D encoders with an InfoNCE loss on single molecules (QM9). (b) **Reaction Prediction:** The 2D encoder processes reactants, a pre-trained encoder provides reagent fingerprints, and an LSTM decoder generates product SMILES.

Overview. As illustrated in Figure 2, our approach employs a 2D graph encoder, a 3D graph encoder (used only in pre-training), and an LSTM decoder. The two encoders share the same message-passing

network architecture, but the 3D encoder has an additional distance feature on each bond. We first align these two encoders via a contrastive InfoNCE objective so that the 2D encoder implicitly learns 3D spatial cues. Afterward, the 2D encoder alone is integrated into a reaction prediction framework: reactant graphs go through the 2D encoder, reagents are converted to a fixed fingerprint, and the resulting embeddings condition an LSTM that autoregressively predicts the product SMILES.

Message Passing Neural Network (MPNN). Each encoder is a Message Passing Neural Network with $L = 3$ layers and a hidden dimension of 128. Let $\mathbf{h}_i^{(t)}$ denote the hidden state of node i at layer t, and $\mathbf{e}_{ij}$ denote the bond/edge feature between i and j. Initially, $\mathbf{h}_i^{(0)}$ is the one-hot or embedded vector of atom i's properties. A single message-passing update for node j from its neighbors $i \in N(j)$ can be written as:

$$\mathbf{h}_j^{(t+1)} = \mathrm{ReLU}\Big(W_{\text{self}}\,\mathbf{h}_j^{(t)} \;+\; \sum_{i \in N(j)} \big(W_{\text{nei}}\,\mathbf{h}_i^{(t)} + W_{\text{edge}}\,\mathbf{e}_{ij}\big)\Big),$$

where $W_{\text{self}}, W_{\text{nei}}$, and W_{edge} are learned parameters. After L layers, each node's final state $\mathbf{h}_i^{(L)}$ reflects its L-hop neighborhood. We then apply mean-pooling across all nodes to obtain a single 128-dimensional graph embedding per molecule.

2D vs. 3D Encoders & InfoNCE Loss. The 2D encoder uses bond-type features only, while the 3D encoder also includes the distance d_{ij} in the bond feature $\mathbf{e}_{ij}$. During pre-training, we feed each molecule to both encoders, obtaining embeddings $\mathbf{z}_i^2$ (2D) and $\mathbf{z}_i^3$ (3D). We use a contrastive InfoNCE loss to align these embeddings across a batch of N molecules:

$$\mathcal{L}_{2\to3} = -\frac{1}{N}\sum_{i=1}^{N} \log\Big(\frac{\exp(\mathrm{sim}(\mathbf{z}_i^2, \mathbf{z}_i^3)/\tau)}{\sum_{j=1}^{N} \exp(\mathrm{sim}(\mathbf{z}_i^2, \mathbf{z}_j^3)/\tau)}\Big),$$

with a symmetric term $\mathcal{L}_{3\to2}$ defined similarly. The total InfoNCE loss is $\mathcal{L}_{\text{InfoNCE}} = \mathcal{L}_{2\to3} + \mathcal{L}_{3\to2}$. By minimizing this, the 2D encoder learns to produce embeddings close to those of the 3D encoder for the same molecule, acquiring an implicit "3D awareness" from the distance-based features.

LSTM Decoder for Product Generation. After pre-training, only the 2D encoder is used for reaction prediction. An LSTM decoder of hidden size 128 generates the product SMILES one token at a time. We first combine the 128-dimensional reactant embedding and the 128-dimensional reagent fingerprint into a 256-dimensional context vector, which is passed through two linear projections (with $\tanh$) to initialize the LSTM's hidden and cell states. At each decoding step t, the LSTM outputs a probability distribution over the SMILES vocabulary for the next token, conditioned on the prior tokens and the initial context. Training uses teacher forcing, feeding the ground truth token as input at each step, while at inference we use autoregressive sampling (e.g., greedy or beam search) until reaching an end-of-sequence token.

3.3 Training Procedure

Stage 1: 3D InfoNCE Pre-training (QM9). We first train the 2D and 3D encoders on the QM9 dataset. For each molecule, we construct both a 2D graph (bond-type features) and a 3D graph (bond-type + distance). We extract a 128-dim embedding from each encoder and apply the InfoNCE loss. Minimizing this loss over 20 epochs (batch size 32, Adam optimizer, learning rate 10^{-3}) aligns the 2D embeddings with the 3D embeddings. We then save the 2D encoder weights, discarding the 3D encoder. These 2D encoder weights are now effectively "3D-informed."

Stage 2: Fine-tuning on ORD Reactions. Next, we load the pre-trained 2D encoder to process ORD reactions. Each reaction's reactant SMILES is converted to a merged graph (combining multiple reactants if needed). The encoder produces a 128-dim reactant embedding. Each reagent SMILES is encoded offline to a 128-dim fingerprint, then reagent fingerprints are averaged to form a single context vector. We concatenate the reactant embedding and reagent vector (total 256-dim) to initialize the LSTM decoder. We train the model with cross-entropy loss on the product SMILES sequence, ignoring losses for padded tokens. An Adam optimizer is used (learning rate 10^{-3}, batch size 16). We measure performance on a held-out test set via the exact-match accuracy of predicted SMILES (canonicalized) versus the true product.

3.4 Inference and Product Prediction

At inference time, we feed new reactants through the 2D encoder to get the reactant embedding. Reagents, if present, are encoded into a fingerprint (or set to zero if no reagents). The combined 256-dim context initializes the LSTM decoder, which then predicts the product SMILES autoregressively. The decoder begins with a start-of-sequence token and generates one character at a time until an end-of-sequence token is produced. For each step, the most probable next token (or a beam search over tokens) is selected, building the SMILES string incrementally. The final output is a predicted product SMILES, which can be compared to the reference product or validated for correctness. This process typically runs in milliseconds per reaction on a GPU, making the model suitable for large-scale virtual screening.

4 Experiments

4.1 Datasets

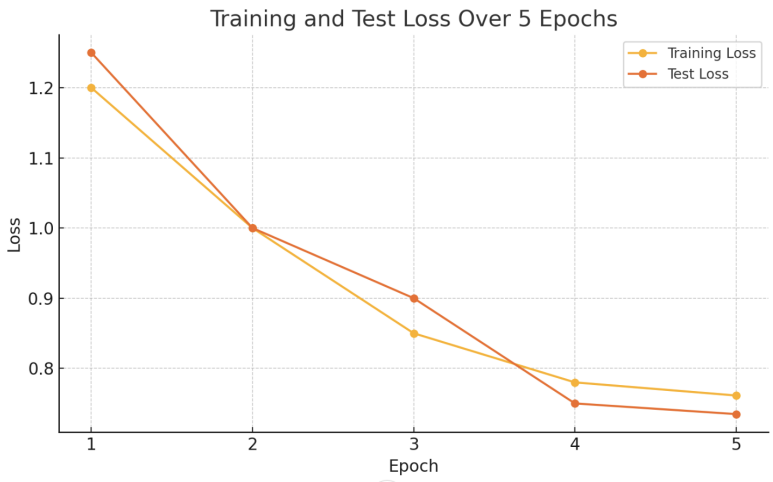
QM9 Quantum Molecules Dataset: QM9 (Ramakrishnan et al., 2014) is composed of around 134k organic compounds, each formed from the five elements: carbon, hydrogen, oxygen, nitrogen, and fluorine. Due to the small size of the molecules, QM9 is especially useful for pre-training models to capture essential 3D structural features.

Data Format: The dataset is provided in structured CSV files where each molecule is represented by its atomic coordinates along with its associated quantum mechanical properties.

Open Reaction Database (ORD): The Open Reaction Database (ORD) (Kearnes et al., 2021) is a large-scale repository containing roughly 1.8 million chemical reaction entries. In our work, the training set includes 596,665 reaction entries. Each reaction record is structured to provide SMILES representations for reactants, reagents, and products, along with detailed metadata.

Data Format: Reaction entries in ORD are stored in a standardized, machine-readable format based on Google Protocol Buffers. For our experiments, this data is converted into more accessible formats CSV to facilitate model training and evaluation.

4.2 Results and Constraints



Our 3D-enhanced graph neural network achieved a final training loss of **0.7613** and a test loss of **0.7348**. This small difference suggests that the model generalizes well and is not overfitting such that the learned representation of chemical reactions transfers effectively to unseen data. The consistently low test loss implies the model maintained predictive reliability on the hold-out reaction set, giving confidence in its real-worXRld applicability.

Training the 3D-enhanced GNN model was significantly constrained by computational resources. Due to the large size of the training dataset (600,000 reaction examples), we were only able to train the model for 5 epochs. In preliminary experiments, training beyond 5 epochs was infeasible within a reasonable time frame on our hardware. Additionally, using cloud-based high-performance GPUs was considered; however, the dataset was too large to efficiently transfer to a cloud environment. As a result, all training was conducted on a local GPU machine with limited memory and compute capability. This resource-limited setup necessitated careful batch sizing and model optimization to avoid out-of-memory errors. While the model did converge to a reasonable solution under these constraints, we acknowledge that additional training time and more powerful hardware might further improve performance.

4.3 Future Work

Model Architecture Upgrades: Given the constraints above, the current model employs a Message Passing Neural Network (MPNN)-based graph encoder and an LSTM-based decoder for predicting reaction outcomes. These choices were made because of their relative efficiency on limited hardware. For future work, we plan to explore more advanced architectures to improve prediction accuracy and better capture complex reaction features. In particular, we will investigate Graph Transformer-based encoders – e.g. the Graphormer architecture by Ying et al. (2021) – which incorporate self-attention mechanisms into GNNs for more expressive graph representations. Replacing the MPNN encoder with a Transformer-GNN is expected to help the model capture long-range dependencies and subtle 3D structural information that the current MPNN might miss. On the decoder side, we intend to replace the LSTM sequence model with a Transformer-based sequence decoder, leveraging the superior sequence modeling capacity of Transformers for predicting reaction outcomes. This change could improve how the model learns from sequential reaction information by using the Transformer's multi-head attention to consider all parts of the sequence simultaneously, rather than the strictly sequential LSTM approach.

Alternative GNN Architectures: Besides Graph Transformers, we will also consider other powerful graph network variants. One such direction is exploring Graph Attention Networks (GATs) (Veličković et al., 2018), which apply attention mechanisms on graph neighbors to weigh the importance of different atoms/bonds in the molecular graph. GATs have shown strong performance on various graph tasks by allowing each node to attend to its neighbors with learned weights, potentially capturing reaction-center information more effectively. By integrating GAT layers, or other advanced GNN variants, we aim to enhance the model's ability to focus on the most relevant parts of a molecule during a reaction. Ultimately, an ensemble or hybrid approach combining 3D-enhanced MPNN features with attention-based mechanisms may yield the best of both worlds. We will also continually optimize the model's efficiency, exploring sparse attention techniques or other pre-trained models to ensure that any architecture upgrades remain feasible with our computing resources.

5 Supplementary Material

Video: `https://www.youtube.com/watch?v=B3tgg5Nufa4&ab_channel=ZhichengWang`

References

[1] C. Ying, T. Cai, S. Luo, S. Zheng, G. Ke, D. He, Y. Shen, and T.-Y. Liu. *Do Transformers Really Perform Bad for Graph Representation?* In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

[2] Y.-L. Liao and T. Smidt. *Equiformer: Equivariant Graph Attention Transformer for 3D Atomistic Graphs.* In *International Conference on Learning Representations (ICLR)*, 2023.

[3] Open Reaction Database. *The Open Reaction Database.* (2020) [Online]. Available: `https://open-reaction-database.org/`.

[4] D. Weininger. *SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules.* Journal of Chemical Information and Computer Sciences, 28(1):31–36, 1988.

[5] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio. *Graph Attention Networks.* In *International Conference on Learning Representations (ICLR)*, 2018.

[6] B. Jin, R. Barzilay, and T. Jaakkola. *Predicting Organic Reaction Outcomes with Weisfeiler-Lehman Networks.* arXiv preprint arXiv:1709.04325, 2017. [Online]. Available: `https://arxiv.org/abs/1709.04325`

[7] P. Schwaller, T. Gaudin, D. Lányi, C. Bekas, and T. Laino. *Molecular Transformer: A Model for Uncertainty-Calibrated Chemical Reaction Prediction.* arXiv preprint arXiv:1811.09600, 2019. [Online]. Available: `https://arxiv.org/abs/1811.09600`

[8] K. T. Schütt, F. Arbabzadah, S. Chmiela, K. R. Müller, and A. Tkatchenko. *SchNet: A Continuous-Filter Convolutional Neural Network for Modeling Quantum Interactions.* arXiv preprint arXiv:1706.08566, 2017. [Online]. Available: `https://arxiv.org/abs/1706.08566`

[9] J. Klicpera, S. Groß, and S. Günnemann. *Directional Message Passing for Molecular Graphs.* arXiv preprint arXiv:2003.03123, 2020. [Online]. Available: `https://arxiv.org/abs/2003.03123`

[10] H. Stärk, D. Beaini, G. Corso, P. Tossou, C. Dallago, S. Günnemann, and P. Liò. *3D Infomax improves GNNs for Molecular Property Prediction.* arXiv preprint arXiv:2110.04126, 2022. [Online]. Available: `https://arxiv.org/abs/2110.04126`

[11] Y. Liu, et al. *GraphMVP: Self-supervised Multi-View Pre-training for Graph Neural Networks.* arXiv preprint arXiv:2202.12264, 2022. [Online]. Available: `https://arxiv.org/abs/2202.12264`

[12] RDKit: Open-source Cheminformatics. Available online at `http://www.rdkit.org`. Accessed: March 16, 2025.

[13] R. Ramakrishnan, P. O. Dral, M. Rupp, and O. A. von Lilienfeld. *Quantum Chemistry Structures and Properties of 134 Kilo Molecules.* In *Scientific Data*, 2014.

[14] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl. *Neural Message Passing for Quantum Chemistry.* In *Proceedings of the 34th International Conference on Machine Learning (ICML)*, pp. 1263–1272, 2017.

[15] Tu, Z., & Coley, C. W. (2021). Permutation invariant graph-to-sequence model for template-free retrosynthesis and reaction prediction. [Online]. Available: `https://arxiv.org/abs/2110.09681`