GDiffRetro: Retrosynthesis Prediction with Dual Graph Enhanced Molecular Representation and Diffusion Generation

Anonymous Author(s)*

ABSTRACT

Retrosynthesis prediction is a critical task in chemical research, focusing on identifying reactants capable of synthesizing a target product. Typically, the retrosynthesis prediction involves two phases: *Reaction Center Identification* and *Reactant Generation*. However, we argue that most existing methods suffer from the following two limitations: i) Existing models do not adequately capture the "face" information in molecular graphs for the reaction center identification. ii) Current approaches for the reactant generation predominantly use sequence generation in 2D space, which lacks versatility in generating reasonable distributions for completed reactive groups and overlooks the inherent 3D properties of molecules.

In this paper, we propose **GDiffRetro**, an innovative framework for retrosynthesis prediction. For the reaction center identification phase, the proposed GDiffRetro uniquely integrates the original graph with its corresponding dual graph to represent molecular structures, which helps guide the model to pay attention to the faces in the graph. For the reactant generation phase, GDiffRetro employs a conditional equivariant diffusion model in 3D to further transform the obtained synthon in the previous phase into a complete reactant. Despite operating within a semi-template framework, our experimental findings reveal that GDiffRetro outperforms contemporary state-of-the-art models in top-1 accuracy, including those reliant on predefined chemical templates. Additionally, it demonstrates performance on par with existing methods across various other evaluative metrics.

CCS CONCEPTS

• **Computing methodologies** → **Machine learning**; • **Applied computing** → Bioinformatics.

KEYWORDS

Graph Representation Learning, Conditional Diffusion, Retrosynthesis Prediction

ACM Reference Format:
Anonymous Author(s). 2018. GDiffRetro: Retrosynthesis Prediction with Dual Graph Enhanced Molecular Representation and Diffusion Generation. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation emai (Conference acronym 'XX)*. ACM, New York, NY, USA, 11 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

The retrosynthesis prediction has garnered widespread attention in the past decades due to its crucial role in drug discovery [4, 14]. The retrosynthesis task aims to find a set of reactants capable of synthesizing a given product, which is a typical one-to-many problem. Even for experienced chemists, addressing such a one-to-many problem is also extremely challenging because of the huge search space of all possible solutions.

In recent years, benefiting from the rapid advancement of deep learning, researchers have resorted to deep learning-based models to design efficient approaches for retrosynthetic prediction. Existing deep learning-based methods can be divided into three categories [30]: template-based methods, template-free methods and semi-template methods. The template-based methods rely on predefined templates (extracted from a large-scale chemical reaction database) to transform a product into reactants. For example, Dai et al. proposed the graph logic network (GLN) [5], which treats chemical knowledge of reaction templates as logical rules and utilizes graph representation learning techniques to model the conditional joint probability between rules and reactants.

Considering that template-based methods are constrained by external knowledge (predefined templates), template-free methods have been proposed, which do not require extracting reaction patterns from external databases. The template-free methods directly convert products into potential reactants. For example, the Chemformer proposed in [12] formulates the retrosynthesis prediction as a translation task, where the product SMILES and the set of reactant SMILES serve as the "source language" and the "target language", respectively. Template-free methods do not rely on predefined templates, however, they typically generate reactant SMILES by sequentially outputting individual symbols, which makes their predictions limited in diversity and interpretability.

Inspired by chemists' expertise, the semi-template framework for retrosynthesis prediction has recently been proposed to alleviate issues present in both template-based and template-free methods. Semi-template methods do not utilize reaction templates or directly transform products into reactants. Instead, they predict the final reactants through the intermediates (synthons). The semi-template framework contains two steps: first identifying the reaction center to form synthons, then completing the synthons into reactants through a generative model. For instance, G2Gs [28] first employs the Relational Graph Convolutional Network (RGCN) [23] for reaction center identification, and then generates products through the variational graph translation. More work following the semi-template framework can be found in [21, 42].

Although existing semi-template methods have achieved success in some scenarios, we argue that there are still avenues to improve. First, existing methods solely focus on the nodes within the molecular graph, neglecting the features associated with faces in the graph. The features of faces play a crucial role in the reaction

Conference acronym 'XX, June 03–05, 2018, Woodstock, NY
© 2018 Association for Computing Machinery.
ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

center identification. For example, in a benzene ring, all carbon atoms reside on one face, and the bonds connecting these carbons exhibit high stability, making them less likely to serve as reaction centers. Secondly, existing methods generate reactants based on 2D graphs, whereas real-world molecules exist in a 3D space. This inconsistency can result in the generative model ignoring the 3D structural information of molecules to some extent.

To address the above shortcomings, we propose the retrosynthesis prediction with dual graph enhanced molecular representation and diffusion generation (**GDiffRetro**). We first introduce the concept of dual graphs to the graph representation learning in the stage of reaction center identification, which aids the model in capturing information related to faces within molecular graphs. The dual graph is a way to describe a graph from the perspective of its faces. In the dual graph, each node corresponds to a face in the original graph. Our primary motivation for introducing the dual graph is to integrate the face information into the node representations, enabling the model to focus on the face information within molecular graphs, such as distinguishing whether different nodes are on the same faces. Considering the tremendous success of diffusion models in generative AI [6, 15, 45], we then employ the controllable diffusion model in the 3D space to generate final reactants. More specifically, we generate reactants conditioned on the synthons obtained from the stage of reaction center identification, and conduct the diffusion process in 3D space to preserve the reactants' inherent structural properties. The contributions of this paper are summarized below:

- To better extract information from molecular graphs, we introduce the concept of dual graphs to guide the model in focusing on the faces in the molecular. This enables the model to more precisely identify reaction centers, paving the way for the subsequent reactant generation.
- To better transform intermediates obtained from the stage of reaction center identification into reactants, we make the first attempt at applying the conditional diffusion model in the field of semi-template retrosynthesis prediction.
- Extensive experiments are conducted to evaluate the performance of the proposed method. The results show that the proposed GDiffRetro outperforms the state-of-the-art methods in top-1 accuracy, which demonstrates the effectiveness of the proposed method.

2 PRELIMINARY

2.1 Notations

Following standard convention, vectors and matrices are denoted by bold lower case letters (e.g., $\mathbf{a}$) and bold upper case letters (e.g., $\mathbf{A}$), respectively. Calligraphic letters (e.g., $\mathcal{Q}$) denote sets, and $|\cdot|$ represents the number of elements in the set (e.g., $|\mathcal{Q}|$). Superscript $(\cdot)^\top$ stands for transpose. $\|$ denotes the concatenation operation. $\mathbb{R}^{m\times n}$ is real matrix space of dimension $m \times n$. $\mathbb{E}(\cdot)$ represents the statistical expectation. $\mathcal{N}(\boldsymbol{\mu}, \mathbf{R})$ denotes a Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{R}$. $\mathcal{U}(a, b)$ denotes a uniform distribution with noise range from a to b. $\mathbf{I}$ denotes an identity matrix.

2.2 Retrosynthesis Prediction

Considering that molecules in chemistry have a natural graph structure, a chemical molecule $\mathcal{M}$ containing n atoms and q types of chemical bonds can be represented as $\mathcal{M} = \{\mathbf{A}, \mathbf{X}\}$, where $\mathbf{X} \in \mathbb{R}^{n\times d}$ is a node feature matrix (with feature dimension d) and $\mathbf{A} \in \mathbb{R}^{n\times n\times q}$ is an adjacency matrix (if there exists a bond of type k between atom i and atom j, then $A_{i,j,k} = 1$; otherwise $A_{i,j,k} = 0$). Based on the above formulation of molecules, a chemical reaction can be described as a pair of sets $(\mathcal{G}^{\mathrm{r}}, \mathcal{G}^{\mathrm{p}})$, where $\mathcal{G}^{\mathrm{r}} = \left\{\mathcal{M}_i^{\mathrm{r}}\right\}|_{i=1}^{l}$ is a set containing l reactants and $\mathcal{G}^{\mathrm{p}} = \{\mathcal{M}_j^{\mathrm{p}}\}|_{j=1}^{m}$ is a set containing m products. Following previous work, we focus only on standard single-output chemical reactions, i.e., $|\mathcal{G}^{\mathrm{p}}| = 1$. For a single-output chemical reaction $\left(\left\{\mathcal{M}_i^{\mathrm{r}}\right\}|_{i=1}^{l}, \mathcal{M}^{\mathrm{p}}\right)$, the goal of retrosynthesis prediction task is to predict the set of reactants $\left\{\mathcal{M}_i^{\mathrm{r}}\right\}|_{i=1}^{l}$ corresponding to the given product $\mathcal{M}^{\mathrm{p}}$. In this paper, we consider adopting a two-step architecture to handle the retrosynthesis prediction task. Specifically, we first conduct reaction center identification to partition the products into synthons (subgraphs of the product molecule, often not valid molecules). Then, we utilize a diffusion model to generate reactants based on the synthons obtained in the first step.

2.3 Diffusion Models

The diffusion models are widely applied across various fields [1, 7, 40] and mainly comprise the following three components [8]:

- **Forward process.** Starting from an input $\mathbf{x}_0$, the forward process aims to generate the latent variables $\{\mathbf{x}_i\}|_{i=1}^{T}$ in a Markov Chain by incrementally introducing Gaussian noise over T steps, i.e., $\mathbf{x}_0 \to \mathbf{x}_1 \to \cdots \to \mathbf{x}_T$ (the so-called diffusion steps). As $T \to \infty$, $\mathbf{x}_T$ converges to a Gaussian distribution.
- **Reverse process.** Taking $\mathbf{x}_T$ as the starting point, the goal of the inverse process is to learn a denoising process $\mathbf{x}_t \to \mathbf{x}_{t-1}$ iteratively by a network parameterized with $\boldsymbol{\theta}$.
- **Inference.** After obtaining the trained model with parameter $\boldsymbol{\theta}$, the diffusion model samples $\mathbf{x}_T$ from the $\mathcal{N}(\mathbf{0}, \mathbf{I})$ and iteratively denoises it using $p_{\boldsymbol{\theta}}(\mathbf{x}_{t-1}|\mathbf{x}_t)$ to achieve the goal of generation (e.g., $\mathbf{x}_T \to \mathbf{x}_{T-1} \to \cdots \to \mathbf{x}_0$).

3 METHODOLOGY

In this section, a two-step framework is designed for the retrosynthesis prediction task. First, we design a dual graph enhanced reaction center identification method to effectively partition the products into synthons. Considering that the obtained synthons are usually not valid molecules, we then introduce a reactant generation method based on the conditional diffusion model to transform the obtained synthons into valid chemical molecules (reactants). The overview of the proposed method is shown in Figure 1.

3.1 Dual Graph Enhanced Reaction Center Identification

The identification of reaction centers can be fundamentally viewed as a binary link prediction task. Specifically, given embeddings of two atoms in the product, the reaction center prediction model is required to output a score, which represents the probability of a reaction center existing between these two atoms. The higher the

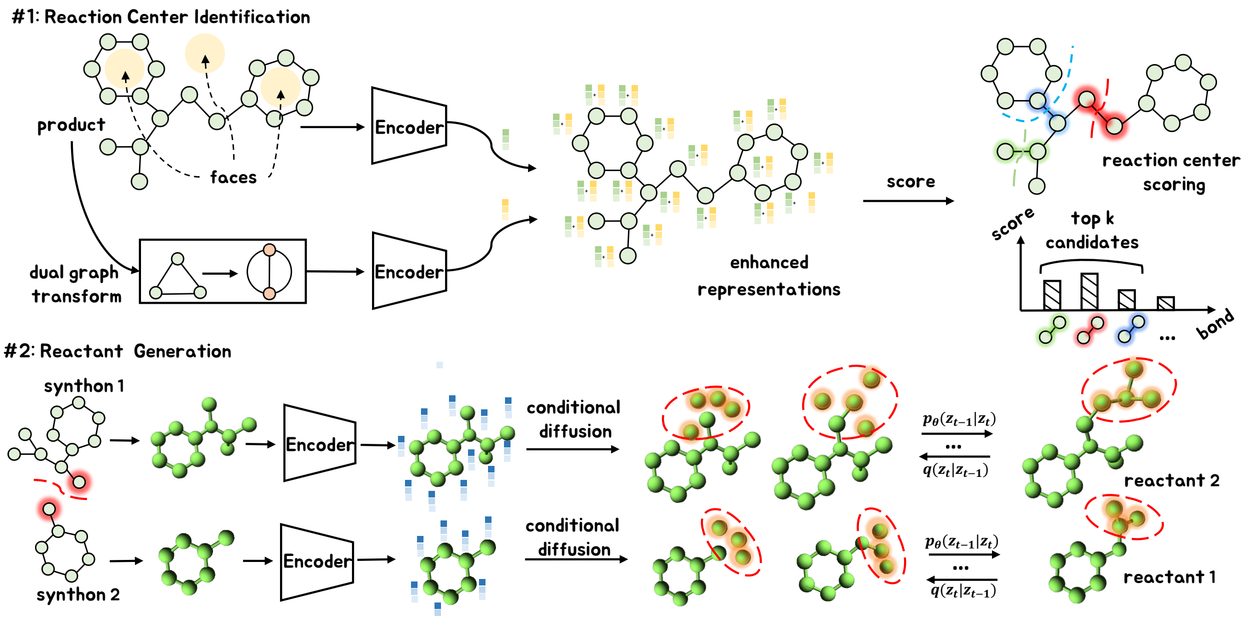

Figure 1: The demonstration of the proposed method. In the reaction center identification phase (#1), we utilize the dual graph to enhance the representations, thus leading to an improved accuracy in reaction center identification. In the reactant generation phase (#2), we employ the conditional diffusion model in 3D (conditioned on the obtained synthon) to transform synthons into reactants.

probability, the more it indicates that the product needs to break the bond between these two atoms to generate synthons.

Given a product $\mathcal{M}^{\mathrm{p}} = \{\mathrm{A}^{\mathrm{p}}, \mathrm{X}^{\mathrm{p}}\}$, we first use the GNN-based method to encode atoms in the product. Considering the heterogeneity of the molecular graph (containing different types of nodes/edges), we use the RGCN to encode atoms. Formally, the update of node i's representation can be expressed as follows:

$$\begin{cases} \mathbf{h}_i^0 = \mathrm{X}^{\mathrm{p}}\left[i,:\right], \\ \mathbf{h}_i^l = \sigma\left(\sum_{r\in\mathcal{R}}\sum_{j\in\mathcal{N}_i^r}\mathbf{W}_r^{l-1}\mathbf{h}_j^{l-1} + \mathbf{W}_0^{l-1}\mathbf{h}_i^{l-1}\right),\ l = 1,\cdots,L, \end{cases} \tag{1}$$

where $\mathcal{R}$ is the set of all edge types (chemical bonds), $\mathcal{N}_i^r$ is the set of neighbors of node i under relation r (can be obtained through $\mathrm{A}^{\mathrm{p}}[:,:,i]$), $\sigma(\cdot)$ is an activation function, $\mathbf{W}_r$ is the learnable weight matrix corresponding to the edge type r, and $\mathbf{W}_0$ is the learnable weight matrix for the self-loop edge. A RGCN with L layers can only aggregate information from nodes within L hops, while the reactivity of a reaction center may also be related to more distant nodes. Therefore, we also compute the graph-level embedding (by applying the Readout$(\cdot)$ function proposed in [37]) to introduce the influence of remote atoms, i.e.,

$$\mathbf{h}_{\mathcal{M}^{\mathrm{p}}} = \mathrm{Readout}(\mathbf{H}^L), \tag{2}$$

where $\mathbf{H}^L$ is a node embedding matrix constructed from $\mathbf{h}_i^L$.

The above representation encoding strategy is designed from the perspective of the nodes, while neglecting the faces in the molecular graph. For the reaction center identification task, the faces in the molecular graph are also very important. For example, the carbon atoms in a benzene ring are all in one face, and the bonds between these carbons are very stable, making them unlikely to become reaction centers. To make up for this shortcoming, we introduce the dual graph $\mathcal{D}^{\mathrm{p}} = \left\{\mathrm{A}_{\mathrm{d}}^{\mathrm{p}}, \mathrm{X}_{\mathrm{d}}^{\mathrm{p}}\right\}$ of $\mathcal{M}^{\mathrm{p}}$. The dual graph is constructed as follows:

- **Topological structure construction**. Given an original planar graph $\mathcal{M}^{\mathrm{p}}$, the dual graph $\mathcal{D}^{\mathrm{p}}$ is a graph that has a node for each face of $\mathcal{M}^{\mathrm{p}}$. Additionally, $\mathcal{D}^{\mathrm{p}}$ has an edge connecting two nodes if the corresponding faces in $\mathcal{M}^{\mathrm{p}}$ are separated by an edge in the $\mathcal{M}^{\mathrm{p}}$. The type of an edge in $\mathcal{D}^{\mathrm{p}}$ corresponds to the type of the edge it crosses in $\mathcal{M}^{\mathrm{p}}$. An example of dual graph construction is shown in Figure 2. In Figure 2, the five nodes in the original graph divide the entire space into three parts. Consequently, its corresponding dual graph contains three nodes, with each node representing a face. Furthermore, the original graph contains two types of edges (denoted as blue and green, respectively). Similarly, its dual graph also comprises two types of edges. Specifically, edges crossing the blue edges in the original graph belong to one type (marked as red), whereas edges crossing the green edges in the original graph belong to another type (marked as orange).
- **Node feature construction**. The feature of a node in the dual graph depends on the surface where the node is located. Formally, the feature of node i in $\mathcal{D}^{\mathrm{p}}$ is:

$$\mathrm{X}_{\mathrm{d}}^{\mathrm{p}}[i,:] = \frac{1}{|\mathcal{S}_i|}\sum_{j\in\mathcal{S}_i}\mathrm{X}^{\mathrm{p}}[j,:], \tag{3}$$

where $\mathcal{S}_i$ is the set of nodes in $\mathcal{M}^{\mathrm{p}}$ on the face where node i in $\mathcal{D}^{\mathrm{p}}$ is located.

After obtaining $\mathcal{D}^{\mathrm{p}}$, we encode the nodes in the dual graph in the same way as before (through an L-layer RGCN), i.e.,

$$\mathbf{D}^{L} = \mathrm{RGCN}(\mathcal{D}^{\mathrm{p}}), \tag{4}$$

where each row in $\mathbf{D}^{L}$ is the final embedding of a node in the dual graph $\mathcal{D}^{\mathrm{p}}$. Combining $\mathcal{M}^{\mathrm{p}}$ and $\mathcal{D}^{\mathrm{p}}$, the final embedding of node i in the original molecular graph can be expressed as:

$$\mathbf{m}_i = \mathbf{H}^{L}[i,:] \parallel \sum_{j\in\mathcal{F}_i} \mathbf{D}^{L}[j,:] \parallel \mathbf{h}_{\mathcal{M}^{\mathrm{p}}}, \tag{5}$$

where $\mathcal{F}_i$ is the set of nodes in $\mathcal{D}^{\mathrm{p}}$ on the face where node i in $\mathcal{M}^{\mathrm{p}}$ is located. In order to estimate the reactivity probability between a pair of nodes i and j, we formulate the edge embedding as follows:

$$\mathbf{e}_{ij} = \mathbf{m}_i \parallel \mathbf{m}_j \parallel \mathbf{A}^{\mathrm{p}}[i,j,:]. \tag{6}$$

Then, the reactivity score can be calculated as $s_{ij} = \mathrm{Sigmoid}\left(\phi(\mathbf{e}_{\mathrm{ij}})\right)$, where $\phi(\cdot)$ is a network for converting edge embeddings to scalar scores. For training, the dual graph enhanced reaction center identification module is optimized by maximizing the following loss function:

$$\mathcal{L}^{(1)} = -\mathbb{E}_{\mathcal{P}_{\mathrm{r}}}\left[\sum_{i}\sum_{j\neq i}\lambda Y_{ij}\log(s_{ij}) + (1-Y_{ij})\log(1-s_{ij})\right], \tag{7}$$

where $\mathcal{P}_{\mathrm{r}}$ is the set of all chemical reactions in the training data, Y_{ij} is the true label indicating whether a reaction center exists between atoms i and j, and λ is a hyper-parameter for alleviating class imbalance issues.

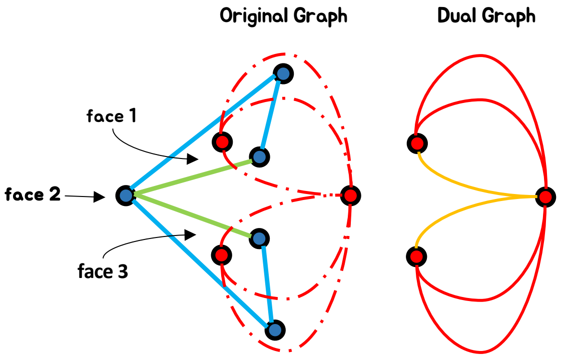

Figure 2: An example of dual graph construction. Each node in the dual graph corresponds to a face in the original graph, and the type of each edge in the dual graph depends on the type of the edge it crosses in the original graph (the different colors of edges in the graph represent different types).

Essentially, a dual graph depicts a graph from the perspective of its faces. As shown in Fig. 2, each node in the dual graph represents a face in the original graph. When we input the dual graph into the RGCN, the node representations output by the RGCN in the dual graph are actually representations of the faces in the original graph. The obtained face representations are then integrated into the node representations of the original graph through Eq. (5). In this manner, the proposed dual graph and RGCN can utilize the face information within the original graph. It is easy to see that when two nodes in the original graph exist within the same face, their corresponding face representations included in the Eq. (5) will be the same (i.e., they have the same face information), which is consistent with intuition.

After obtaining the reaction center, a product molecule can be divided into its corresponding synthons. It should be noted that the synthons may not be valid molecules. Therefore, we introduce a conditional diffusion model to transform the synthons into structurally valid reactants in the following section.

3.2 Conditional Diffusion Model-Based Reactant Generation

As shown in Figure 1, the conditional diffusion model-based reactant generation involves two crucial processes: *i)* A forward process corrupts the structure and features of a synthon by adding Gaussian noise step by step. *ii)* A reverse process learns the denoise process and outputs a reactant.

- **Forward process.** An atom $\mathbf{s}$ can be represented by a 3D coordinates $\mathbf{u}^{(\mathrm{x})} \in \mathbb{R}^3$ and d-dimensional features $\mathbf{u}^{(\mathrm{h})} \in \mathbb{R}^{\mathrm{d}}$, i.e., $\mathbf{s} = [\mathbf{u}^{(\mathrm{x})}, \mathbf{u}^{(\mathrm{h})}]$. Setting $\mathbf{z}_0 = \mathbf{s}$ as the initial state and parameterize a fixed noise process as:

$$q\left(\mathbf{z}_t|\mathbf{z}_0\right) = \mathcal{N}\left(\mathbf{z}_t|\alpha_t\mathbf{z}_0, \sigma_t^2\mathbf{I}\right), \ t = 1,\cdots,T, \tag{8}$$

where $\mathbf{z}_t$ is a latent noised representation, α_t controls the proportion of the original input to be retained, and σ_t^2 controls the intensity of added Gaussian noise. Inspired by [29, 31], we adopt a variance-preserving noise adding process, i.e., $\alpha_t = \sqrt{1-\sigma_t^2}$. In addition, a numerically stable polynomial noise schedule [9] is applied to regulate the added noise, as shown below:

$$\alpha_t = (1-2s)\left[1-(\frac{t-1}{T-1})^2\right], \ t = 1,\cdots,T, \tag{9}$$

where s is set to 10^{-5} to ensure numerical stability.

- **Reverse process.** The reverse process takes $\mathbf{x}_T$ as the starting point and attempts to learn a network with $\boldsymbol{\theta}$ as a trainable parameter for denoising, as follows:

$$p_{\boldsymbol{\theta}}(\mathbf{z}_{t-1}|\mathbf{z}_t) = \mathcal{N}(\mathbf{z}_{t-1};\boldsymbol{\mu}_{\boldsymbol{\theta}}(\mathbf{z}_t,t),\mathbf{R}_{\boldsymbol{\theta}}(\mathbf{z}_t,t)), \tag{10}$$

where $\boldsymbol{\mu}_{\boldsymbol{\theta}}(\mathbf{z}_t,t)$ and $\mathbf{R}_{\boldsymbol{\theta}}(\mathbf{z}_t,t)$ are obtained from a network parameterized by $\boldsymbol{\theta}$.

- **Optimization process.** In order to maximize the likelihood of observed input data, we optimize the variational lower bound, i.e.,

$$\begin{aligned} -\log p(\mathbf{z}_0) &= -\log\int p(\mathbf{z}_{0:T})d\mathbf{z}_{1:T} = -\log\mathbb{E}_{q(\mathbf{z}_{1:T}|\mathbf{z}_0)}\left[\frac{p(\mathbf{z}_{0:T})}{q(\mathbf{z}_{1:T}|\mathbf{z}_0)}\right] \\ &\le \textstyle\sum_{t>1}^{T}\underbrace{\mathbb{E}_{q(\mathbf{z}_t|\mathbf{z}_0)}\left[D_{\mathrm{KL}}(q(\mathbf{z}_{t-1}|\mathbf{z}_t,\mathbf{z}_0)\parallel p_{\boldsymbol{\theta}}(\mathbf{z}_{t-1}|\mathbf{z}_t))\right]}_{\text{(diffusion loss } \mathcal{L}_t\text{)}} \\ &+\underbrace{D_{\mathrm{KL}}(q(\mathbf{z}_T|\mathbf{z}_0)\parallel p(\mathbf{z}_T))}_{\text{(prior loss } \mathcal{L}_{\mathrm{p}}\text{)}} - \underbrace{\mathbb{E}_{q(\mathbf{z}_1|\mathbf{z}_0)}\left[\log p_{\boldsymbol{\theta}}(\mathbf{z}_0|\mathbf{z}_1)\right]}_{\text{(reconstruction loss } \mathcal{L}_1\text{)}}, \end{aligned} \tag{11}$$

where $D_{\mathrm{KL}}(\cdot)$ is a operation of calculating the KL divergence of two distributions.

The diffusion loss $\mathcal{L}_t$ encourages approximating the distribution $q(\mathbf{z}_{t-1}|\mathbf{z}_t,\mathbf{z}_0)$ through a distribution $p_{\boldsymbol{\theta}}(\mathbf{z}_{t-1}|\mathbf{z}_t)$ associated with a network. The closed form of $q(\mathbf{z}_{t-1}|\mathbf{z}_t, \mathbf{z}_0)$ can be expressed as:

$$q(\mathbf{z}_{t-1}|\mathbf{z}_t, \mathbf{z}_0) = \mathcal{N}(\mathbf{z}_{t-1}; \boldsymbol{\mu}_q(\mathbf{z}_t, \mathbf{z}_0, t), \sigma_q^2(t)\mathbf{I}), \tag{12}$$

where

$$\boldsymbol{\mu}_q(\mathbf{z}_t, \mathbf{z}_0, t) = \frac{\alpha_t \sigma_{t-1}^2}{\alpha_{t-1}\sigma_t^2}\mathbf{z}_t + \frac{\alpha_{t-1}^2\sigma_t^2 - \alpha_t^2\sigma_{t-1}^2}{\alpha_{t-1}\sigma_t^2}\mathbf{z}_0, \tag{13}$$

and

$$\sigma_q^2(t) = \sigma_{t-1}^2 - \frac{\alpha_t^2\sigma_{t-1}^4}{\alpha_{t-1}^2\sigma_t^2}. \tag{14}$$

Substituting Eq. (10) and Eq. (12) into the diffusion loss at the step t yields:

$$\begin{aligned}\mathcal{L}_t &= \mathbb{E}_{q(\mathbf{z}_t|\mathbf{z}_0)}\left[D_{\mathrm{KL}}(q(\mathbf{z}_{t-1}|\mathbf{z}_t, \mathbf{z}_0) \parallel p_{\boldsymbol{\theta}}(\mathbf{z}_{t-1}|\mathbf{z}_t))\right] \\ &= \mathbb{E}_{q(\mathbf{z}_t|\mathbf{z}_0)}\left[\frac{1}{2\sigma_q^2(t)}\left[\parallel \boldsymbol{\mu}_{\boldsymbol{\theta}}(\mathbf{z}_t, t) - \boldsymbol{\mu}_q(\mathbf{z}_t, \mathbf{z}_0, t) \parallel_2^2\right]\right],\end{aligned} \tag{15}$$

where $\boldsymbol{\mu}_{\boldsymbol{\theta}}(\mathbf{z}_t, t)$ can be easily expressed according to the Eq. (13):

$$\boldsymbol{\mu}_{\boldsymbol{\theta}}(\mathbf{z}_t, t) = \frac{\alpha_t\sigma_{t-1}^2}{\alpha_{t-1}\sigma_t^2}\mathbf{z}_t + \frac{\alpha_{t-1}^2\sigma_t^2 - \alpha_t^2\sigma_{t-1}^2}{\alpha_{t-1}\sigma_t^2}\hat{\mathbf{z}}_{\boldsymbol{\theta}}(\mathbf{z}_t, t), \tag{16}$$

with $\hat{\mathbf{z}}_{\boldsymbol{\theta}}(\mathbf{z}_t, t)$ as the predicted initial state $\mathbf{z}_0$ (output by a trainable network). Plugging Eq. (13) and Eq. (16) into Eq. (15) results in

$$\mathcal{L}_t = \mathbb{E}_{q(\mathbf{z}_t|\mathbf{z}_0)}\left[\frac{1}{2}\left(\frac{\alpha_{t-1}^2}{\sigma_{t-1}^2} - \frac{\alpha_t^2}{\sigma_t^2}\right) \parallel \hat{\mathbf{z}}_{\boldsymbol{\theta}}(\mathbf{z}_t, t) - \mathbf{z}_0 \parallel_2^2\right]. \tag{17}$$

Considering that $\mathbf{z}_t$ can be reparameterized as $\mathbf{z}_t = \alpha_t\mathbf{z}_0 + \sigma_t\boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, [8] suggests that using a neural network to predict $\boldsymbol{\epsilon}$ instead of $\mathbf{z}_0$ will lead to a better result, i.e., $\mathcal{L}_t$ can be simplifies to:

$$\mathcal{L}_t = \mathbb{E}_{\epsilon\sim\mathcal{N}(\mathbf{0},\mathbf{I})}\left[\frac{1}{2}\left(\frac{\alpha_{t-1}^2\sigma_t^2}{\alpha_t^2\sigma_{t-1}^2} - 1\right) \parallel \hat{\boldsymbol{\epsilon}}_{\boldsymbol{\theta}}(\mathbf{z}_t, t) - \boldsymbol{\epsilon} \parallel_2^2\right]. \tag{18}$$

According to [9], both $\mathcal{L}_{\mathrm{p}}$ and $\mathcal{L}_0$ are close to 0 (due to $\alpha_T = 0$, $\alpha_1 \approx 1$, and $\mathbf{z}_0$ is discrete). Furthermore, Ho *et al.* [8] found that removing the weight in Eq. (18) is conducive to improving sample quality. Therefore, an unweighted version of the final loss $\mathcal{L}^{(2)}$ used in the reactant generation phase can be written as:

$$\mathcal{L}^{(2)} = \mathbb{E}_{\epsilon\sim\mathcal{N}(\mathbf{0},\mathbf{I}), t\sim\mathcal{U}(1,T)}\left[\parallel \hat{\boldsymbol{\epsilon}}_{\boldsymbol{\theta}}(\mathbf{z}_t, t) - \boldsymbol{\epsilon} \parallel_2^2\right]. \tag{19}$$

• **Modeling of the $\hat{\boldsymbol{\epsilon}}_{\boldsymbol{\theta}}$.** It should be noted that during the reactant generation stage, each atom not only contains a feature vector but also includes 3D coordinates. In order to preserve equivariance of $\hat{\boldsymbol{\epsilon}}_{\boldsymbol{\theta}}$ to coordinate rotations and translations[1], we utilize the EGNN [22] to model $\hat{\boldsymbol{\epsilon}}_{\boldsymbol{\theta}}$. Define the final reactant $\mathcal{R}$ containing n atoms as $\mathcal{R} = \{v_i\}|_{i=1}^n = \{\mathcal{S}, \mathcal{Q}\}$, where $\mathcal{S} = \{v_i\}|_{i=1}^m$ is a set of m atoms in the synthon (obtained from the first stage), $\mathcal{Q} = \{v_i\}|_{i=m+1}^n$ is a set of $n - m$ atoms need to be generated[2], and the feature of atom i in the denoising time step t can be expressed as $\mathbf{z}_{i,t} = [\mathbf{z}_{i,t}^{(\mathrm{x})}, \mathbf{z}_{i,t}^{(\mathrm{h})}]$. Following the previous work [9], $\hat{\boldsymbol{\epsilon}}_{\boldsymbol{\theta}}(\mathbf{z}_{i,t}, t)$ can be written as:

$$\hat{\boldsymbol{\epsilon}}_{\boldsymbol{\theta}}(\mathbf{z}_{i,t}, t) = \left[\mathbf{e}_{i,t}^{(\mathrm{x}),L}, \mathbf{e}_{i,t}^{(\mathrm{h}),L}\right] - \left[\mathbf{z}_{i,t}^{(\mathrm{x})}, \mathbf{0}\right], \tag{20}$$

where $\left[\mathbf{e}_{i,t}^{(\mathrm{x}),L}, \mathbf{e}_{i,t}^{(\mathrm{h})}\right]$ is the embedding of atom i output by a L-layer EGNN (in the time step t), and its computation process is:

$$\begin{cases} \mathbf{m}_{ij} = \phi_e\left(\mathbf{e}_{i,t}^{(\mathrm{h}),l-1}, \mathbf{e}_{j,t}^{(\mathrm{h}),l-1}, \parallel \mathbf{d}_{ij}^{l-1} \parallel_2^2\right), \\ \mathbf{e}_{i,t}^{(\mathrm{h}),l} = \phi_h\left(\mathbf{e}_{i,t}^{(\mathrm{h}),l-1}, \sum_{j\neq i}\mathbf{m}_{ij}\right), \\ \mathbf{e}_{i,t}^{(\mathrm{x}),l} = \begin{cases} \mathbf{e}_{i,t}^{(\mathrm{x}),l-1} + \sum_{i\neq j}\dfrac{\mathbf{d}_{ij}^{l-1}\phi_r\left(\mathbf{e}_{i,t}^{(\mathrm{h}),l-1}, \mathbf{e}_{j,t}^{(\mathrm{h}),l-1}\right)}{1+ \parallel \mathbf{d}_{ij}^{l-1} \parallel_2^2}, v_i \notin \mathcal{S}, \\ \mathbf{e}_{i,t}^{(\mathrm{x}),l-1}, v_i \in \mathcal{S}, \end{cases} \end{cases} \tag{21}$$

with $\mathbf{d}_{ij}^{l-1} = \mathbf{e}_{i,t}^{(\mathrm{x}),l-1} - \mathbf{e}_{j,t}^{(\mathrm{x}),l-1}$, $\phi_e/\phi_h/\phi_r$ is a multi-layer perceptron. It can be seen that we keep the coordinates of the atoms in the $\mathcal{S}$ unchanged, so that the generation of reactants is conditioned on the synthon obtained in the first stage. In other words, the generation process is more controllable.

4 EXPERIMENTS

In this section, we assess the effectiveness of GDiffRetro through comprehensive experiments. Our experimental design is structured to address the following three research questions:

- **RQ1:** How does GDiffRetro compare with the three categories of baseline methods in terms of performance improvement across various degrees (k) of evaluation?
- **RQ2:** To what extent are the dual graph enhanced RGCN effective in *Reaction Center Prediction* task?
- **RQ3:** In contrast to existing baselines that cover the entire generation space, does GDiffRetro demonstrate superior performance with a limited number of sampling attempts?

We will answer these questions in Section 4.2, Section 4.3 and Section 4.4, respectively.

4.1 Experiment Setup

4.1.1 Dataset Setup

To assess the proposed method, we utilize the USPTO-50k dataset, a standard single-step retrosynthesis benchmark. Originally compiled by Daniel Lowe, this dataset encompasses 50k chemical reactions, categorized into 10 distinct reaction types [17]. The split of the dataset follows previous work [16, 28]. More details about the dataset can be seen in APPENDIX A.1.

4.1.2 Baselines

Existing baselines can be categorized into three groups: *Template-Based* methods, *Template-Free* methods, and *Semi-Template* methods. Specifically, for *Template-Based* baselines, we select MHNreact [25], GLN [5], LocalRetro [2], GraphRetro [30], RetroComposer [43], Dual-TB [32], and RetroExplainer [39]. For *Template-Free* methods, we select Transformer [36], SCROP [46], Retroformer [38], GTA [26], Graph2SMILES (D-GCN) [35], Transformer (*Aug.*) [33], Dual-TF [32], and Chemformer [12]. For *Semi-Template* methods, we

[1] More details about the equivariance are provided in APPENDIX A.2.

[2] The number of atoms to be generated is sometimes unknown. Following [10], the task of determining the number of atoms can be treated as a classification task. We predefine the classes according to the number of atoms, and then project the output of the EGNN into a vector that implying probabilities associated with each class. More specifically, we aim to train a network where the input is the synthon obtained from the first phase and the output is a scalar (the preset category, such as discrete numbers 1-10). The network first encodes each atom of the synthon using Eq. (21), and obtains the representation $\mathbf{g}_s$ of the synthon through mean pooling. The $\mathbf{g}_s$ is then fed into a fully connected layer to obtain the probability vector of categories. The category with the highest probability is selected as the final prediction of the number of atoms.

Table 1: Top-*k* exact match accuracy (%) on reaction dataset USPTO-50k [16]. The best result for each category is **bolded**. The result of GDiffRetro is highlighted with **red** color.

Baselines		Top-*k* accuracy %			
		$k=1$	$k=3$	$k=5$	$k=10$
Template-Based	MHNreact [25]	51.8	74.6	81.2	88.1
	GLN [5]	52.5	69.0	75.6	83.7
	LocalRetro [2]	53.4	77.5	**85.9**	**92.4**
	GraphRetro [30]	53.7	68.3	72.2	75.5
	RetroComposer [43]	54.5	77.2	83.2	87.7
	Dual-TB [32]	55.2	74.6	80.5	86.9
	RetroExplainer [39]	**57.7**	**79.2**	84.8	91.4
Template-Free	Transformer [36]	43.7	59.7	65.1	70.1
	SCROP [46]	43.7	60.0	65.2	68.7
	Transformer(*Aug.*) [33]	48.3	-	73.4	77.4
	GTA [26]	51.1	67.6	**74.8**	**81.6**
	Retroformer [38]	52.9	68.2	72.5	76.4
	Graph2SMILES (D-GCN) [35]	52.9	66.5	70.0	72.9
	Dual-TF [32]	53.6	**70.7**	74.6	77.0
	Chemformer [12]	**54.3**	-	62.3	63.0
Semi-Template	MEGAN [21]	48.1	70.7	78.4	**86.1**
	G2Gs [28]	48.9	67.6	72.5	75.5
	RetroXpert [42]	50.4	61.1	62.3	63.4
	G^2Retro [3]	51.4	72.1	78.2	83.6
	GDiffRetro (Ours)	**56.8**	**76.8**	**79.9**	81.7

select MEGAN [21], G2Gs [28], RetroXpert [42], and G^2Retro [3]. The definition of the three types is described in detail in Section 1.

4.1.3 Evaluation Metrics

In consistent with previous work [16], we employ the top-*k* exact match accuracy as our evaluation metric. To facilitate meaningful comparisons, we consider different values of k in our experimental evaluations ($k = 1, 3, 5, 10$). The accuracy is determined by comparing the canonical SMILES strings of the predicted molecules with the ground truth.

4.1.4 Implementation Details

We leverage the open-source RDKit library to construct molecular graphs based on molecular SMILES. During the "Reaction Center Identification" stage, we adopt the widely-used machine learning tool, `TorchDrug` [47], to facilitate the training and evaluation processes. To obtain the 3D conformation of a molecular graph from SMILES, we adopt the data processing method employed by DeLinker [11], which involves comparing the conformations of a SMILES representation across all possibilities and selecting the one with the lowest energy. For the generation of SMILES representations of atomic point clouds produced by the diffusion model, we rely on OpenBabel [19], a well-known open-source tool in chemical research. To obtain the top-*k* results, we sample 300 times during the inference process and select k most frequent SMILES representations as the top-*k* candidates.

4.2 Performance Comparison (RQ1)

We assess the performance of our proposed approach by evaluating the top-*k* exact match accuracy. The results are presented in Table 1. The top-*k* result of GDiffRetro is calculated by comparing the *k*-th most frequent SMILES sampled from the proposed model with the ground truth. In Table 1, it is evident that the top-1 result of GDiffRetro surpasses all existing template-free/semi-template based baselines, and most of the state-of-the-art template-based baselines. It's important to note that template-based methods rely heavily on external knowledge compared to template-free and semi-template based methods, making a direct comparison between the template-based method and template-free/semi-template based method inherently unfair. To ensure fairness, we focus on performance gains within the category. Within the "Semi- Template" category, the proposed method achieves a relative improvement of 10.5% in terms of the top-1 metric compared to the second-best method. This demonstrates that GDiffRetro can provide the most accurate retrosynthesis

Table 2: Top-k exact match accuracy of single *Reaction Center Prediction* and end-to-end *Retrosynthesis Prediction*. "*w Dual-$\mathcal{G}$*" means the *Dual-$\mathcal{G}$*raphs are constructed for molecules and considered into the representation learning. "*w/o Dual-$\mathcal{G}$*" means the vanilla RGCN without constructing the dual graphs.

a. Top-k exact match accuracy of *Reaction Center Prediction*.

Setting (*Dual-$\mathcal{G}$*)	Top-k Accuracy %				
	k=1	k=2	k=3	k=5	k=10
w/o	81.4	93.4	96.5	98.6	99.6
w	86.2	95.1	97.4	98.8	99.6

b. Top-k exact match accuracy of *Retrosynthesis Prediction*.

Setting (*Dual-$\mathcal{G}$*)	Top-k Accuracy %			
	k=1	k=3	k=5	k=10
w/o	53.1	76.1	79.7	81.4
w	56.8	76.8	79.9	81.7

prediction with just a single attempt. In real-world applications, a relatively high single-attempt success rate (i.e., top-1 accuracy) is extremely important. This is because the reactants obtained a single retrosynthetic prediction usually not commercially available. Typically, we need to recursively conduct multiple retrosynthesis predictions to obtain the final synthesis route (akin to a search tree). Obviously, under this tree-like structure, a higher top-1 accuracy rate can greatly narrow the search space, thereby enhancing efficiency and reducing resource consumption. When examining the top-3 and top-5 performances, GDiffRetro performs on par with all the template-free and semi-template baselines. Particularly, GDiffRetro outperforms all template-free and semi-template baselines in the top-3 and top-5 categories. It is worth noting that certain approaches, such as MEGAN [21], may exhibit significantly higher performance than GDiffRetro in terms of the top-10 results. We attribute this difference to the limited number of sampling iterations in GDiffRetro, which may hinder its ability to generate an ample set of SMILES candidates. Mathematically, the diversity of results generated by a diffusion model is closely related to the sampling PDF's peakiness. Moreover, in the conditional diffusion model, the peakiness depends on the mutual information between the condition and target. Considering that in our setting, the synthons (conditions) and reactants (targets) are similar, this leads to more concentrated sampling results, which in turn causes a decrease in the top 10 accuracy..

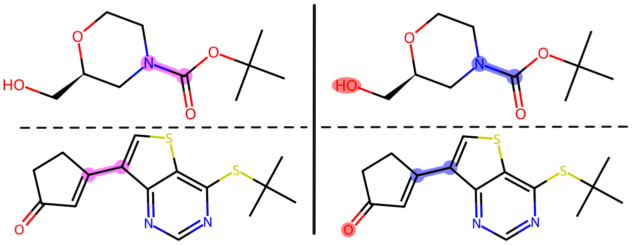

Figure 3: Illustration of reaction center on two example products predicted by GDiffRetro, with dual graphs considered (left pair) and without them (right pair). The ground truths and predictions are highlighted in blue and red, respectively. Overlapping areas, indicating agreement, appear purple. The atoms and chemical bonds are colored based on the well-known CPK coloring (carbon: black, nitrogen: blue, oxygen: red, sulphur: yellow).

4.3 Ablation Study (RQ2)

In this section, we perform ablation studies to evaluate the effectiveness of dual graph-enhanced representation learning introduced in GDiffRetro. To verify the effectiveness of the dual graphs we introduced to the RGCNs, we compare two sets of ablation experiments:

a. top-k accuracy of reaction center prediction
b. top-k accuracy of end-to-end retrosynthesis prediction

with and without dual graphs introduced in the inference process. The results are shown in Table 2a and Table 2b, respectively.

In Table 2a, it is evident that GDiffRetro achieves a significant improvement in top-1 accuracy when incorporating dual graphs. GDiffRetro with *Dual-$\mathcal{G}$* also consistently surpasses its counterpart without *Dual-$\mathcal{G}$* in top-3, top-5, and top-10 accuracy. A clear correlation between the performance of the end-to-end retrosynthesis prediction and the *Dual-$\mathcal{G}$* configuration can be observed in Table 2b, following a similar trend to Table 2a. GDiffRetro with *Dual-$\mathcal{G}$* outperforms the version without it, showing approximately a 3% increase in top-1 accuracy, and a slight improvement in top 3, top 5, and top 10 accuracy. The relatively modest improvement is attributed to the already high accuracy observed in top 3, top 5, and top 10 accuracy.

The predictions for the reaction center of two test molecules, containing 1 and 3 rings, respectively, are depicted in Figure 3. With enhanced information provided by the dual graphs, GDiffRetro with *Dual-$\mathcal{G}$* can offer more precise predictions. More examples from 10 distinct reaction classes in the USPTO-50k dataset are illustrated in APPENDIX A.3 (Figure 7).

4.4 Case Study and Visualization (RQ3)

To assess the proficiency of GDiffRetro in learning reaction templates, we visualize the end-to-end retrosynthesis prediction process for two examples from the same reaction class in Figure 4. For two products belonging to the "protections" reaction class [3], GDiffRetro accurately predicts the reaction centers. Then, it generates the reactants using two similar sets of synthons, formulating identical completed parts. These examples demonstrate that GDiffRetro's results are informed by its understanding of the reaction class, indicating its potential in capturing the underlying reaction template.

To demonstrate the validity and diversity of generated reactant candidates by GDiffRetro, we provide the top-3 results given by GDiffRetro on 3 different synthons in Figure 5. It can be seen that GDiffRetro not only always matches in a one-shot trial, but

[3] More details about the classes in the USPTO-50k dataset can be seen in APPENDIX A.1

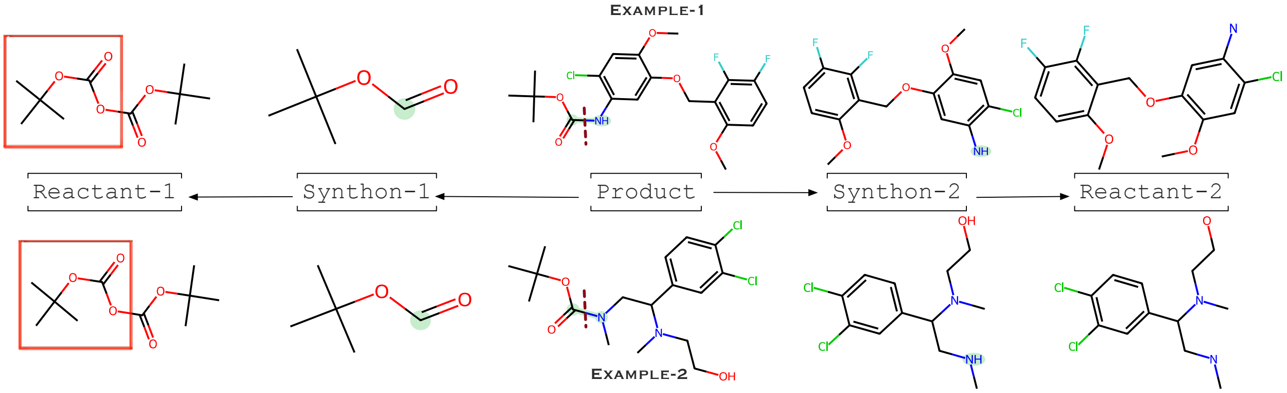

Figure 4: Depiction of the overall *Retrosynthesis Prediction* process for two examples in the "protections" reaction class. Reaction centers are highlighted on the products and synthons, while the completed parts are outlined in red on the reactants.

also gives some reasonable results in the following sub-candidates. Among them, some examples generate all reactants with the same number of completed atoms, but GDiffRetro still provides diverse meaningful candidates for reference.

The trajectory of reactant generation is provided in Figure 6. It can seen that the product is first cleaved into two synthons, and then these two synthons are step by step transformed into the final reactant through the denoising process of the diffusion model. The atoms to be generated start from a cluster of noise (with random coordinates and categories). As the denoising time step changes, reasonable atom types and their coordinates gradually emerge, ultimately leading to the accurate generation of the final reactant.

5 RELATED WORK

Retrosynthetic Prediction. Template-based methods rely on predefined templates extracted from large-scale chemical reaction databases to transform products into reactants. Some significant works in this category include GLN[5], LocalRetro [2], GraphRetro [30], and Dual-TB [32]. To overcome the constraints of external knowledge in template-based methods, template-free methods have been developed. These methods directly convert products into potential reactants without extracting reaction patterns from external databases. Key examples in this category include Chemformer[12], Transformer [36], Retroformer [38], GTA [26], Graph2SMILES (D-GCN) [35], Transformer (Aug.) [33], and Dual-TF [32]. The chemical prior, pertaining to the two stages, is deprecated in this framework. Semi-template methods have emerged to address the limitations of both template-based and template-free methods, inspired by chemists' expertise. These methods predict the final reactants through intermediates (synthons) rather than utilizing reaction templates or directly transforming products into reactants. The semi-template framework involves first identifying the reaction center to form synthons, followed by completing the synthons into reactants using a generative model. Notable works following the semi-template framework include MEGAN [21], G2Gs [28], and RetroXpert [42].

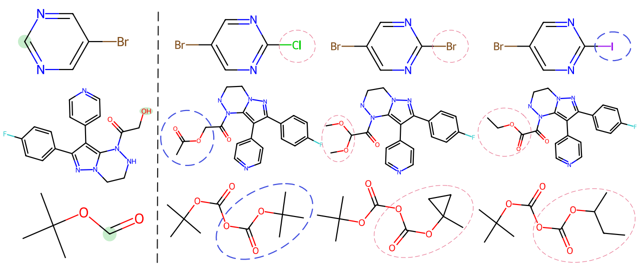

Figure 5: Top-3 reactant generation results for three distinct synthons. The leftmost column displays the original synthons with their reaction centers accentuated in green. On the right side of the dotted line, for each synthon, the 1st, 2nd, and 3rd candidates are placed in corresponding columns in order. Correctly completed parts are denoted in blue, whereas incorrect completions are marked in pink.

Molecular Generation. The problem of molecular generation is closely related to various deep generative models [20, 24]. For example, You et al. modeled the molecular generation as a sequential decision process on graphs (adding nodes or edges based on the current subgraph) and introduced the reinforcement learning for decision making [44]. However, the aforementioned methods perform molecular generation on a 2D plane, neglecting molecules' 3D properties. Several recent works introduced denoising diffusion models to 3D molecular data. Conformer generation methods GeoDiff [41] and ConfGF [27] condition the model on the adjacency matrix of the molecular graph, enabling them to compute and optimize torsion angles between atoms [13]. The equivariant diffusion model [9] generates 3D molecules from scratch, conditioned on predefined scalar properties. Another noteworthy addition is DiffLinker [10], an E(3)-equivariant 3D-conditional diffusion model for designing molecular linkers, capable of connecting an arbitrary number of molecular fragments. Other models include SMCDiff [34] for designing protein scaffolds from protein motifs, and one model for antibody design [18], which combines discrete and continuous diffusion for molecular graphs and 3D coordinates, respectively.

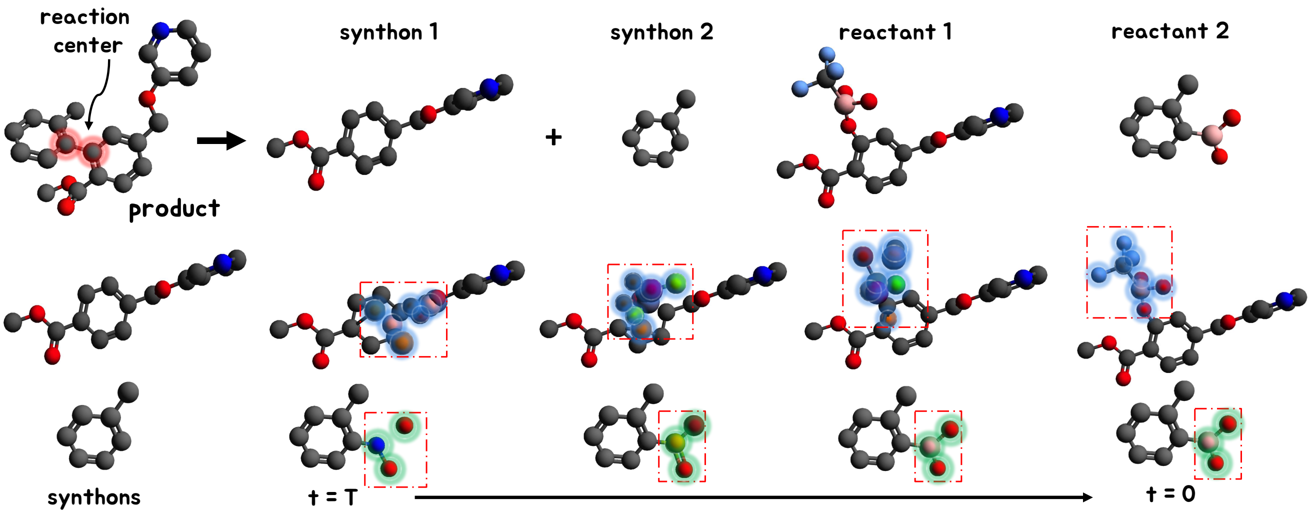

Figure 6: Trajectory of reactant generation. The identified reaction center is highlighted in **red**. The atoms undergoing changes in each time step for synthon 1 and synthon 2 are highlighted in **blue** and **green**, respectively. The overall process contains two steps. First, the target product undergoes bond cleavage at the red highlighted position to generate two synthons. Then, these two synthons are step by step transformed into the final reactants according to the denoising process, i.e., from the noisy version ($t = T$) to the final generated reactant ($t = 0$).

6 CONCLUSION

In this paper, we introduced GDiffRetro, a novel framework designed for retrosynthesis prediction. This framework notably incorporates a dual graph enhanced molecular representation for the reaction center identification, and introduces the conditional diffusion model for reactant generation in 3D space. The experimental results demonstrate that GDiffRetro not only surpasses current state-of-the-art models in top-1 accuracy, including those heavily reliant on templates, but also achieves competitive performance across top-3, top-5, and top-10 rankings. Through comprehensive ablation studies and detailed visualization, we have confirmed that the two key components proposed in GDiffRetro function independently and effectively.

REFERENCES

[1] Jacob Austin, Daniel D. Johnson, Jonathan Ho, Daniel Tarlow, and Rianne van den Berg. 2021. Structured denoising diffusion models in discrete state-spaces. In *Advances in Neural Information Processing Systems.*

[2] Shuan Chen and Yousung Jung. 2021. Deep retrosynthetic reaction prediction using local reactivity and global attention. *JACS Au* 1, 10 (2021), 1612–1620.

[3] Ziqi Chen, Oluwatosin R. Ayinde, James R. Fuchs, Huan Sun, and Xia Ning. 2023. G2Retro as a two-step graph generative models for retrosynthesis prediction. *Communications Chemistry* (2023).

[4] Connor W. Coley, Luke Rogers, William H. Green, and Klavs F. Jensen. 2017. Computer-assisted retrosynthesis based on molecular similarity. *ACS Central Science* 3, 12 (2017), 1237–1245.

[5] Hanjun Dai, Chengtao Li, Connor Coley, Bo Dai, and Le Song. 2019. Retrosynthesis prediction with conditional graph logic network. In *Advances in Neural Information Processing Systems.*

[6] Prafulla Dhariwal and Alexander Nichol. 2021. Diffusion models beat GANs on image synthesis. In *Advances in Neural Information Processing Systems.*

[7] William Harvey, Saeid Naderiparizi, Vaden Masrani, Christian Weilbach, and Frank Wood. 2022. Flexible diffusion modeling of long videos. In *Advances in Neural Information Processing Systems.*

[8] Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising Diffusion Probabilistic Models. In *arXiv:2006.11239.*

[9] Emiel Hoogeboom, Vıctor Garcia Satorras, Clément Vignac, and Max Welling. 2022. Equivariant diffusion for molecule generation in 3D. In *Proc. Int. Conf. Machine Learning.*

[10] Ilia Igashov, Hannes Stärk, Clément Vignac, Victor Garcia Satorras, Pascal Frossard, Max Welling, Michael Bronstein, and Bruno Correia. 2022. Equivariant 3D-conditional diffusion models for molecular linker design. In *arXiv:2210.05274.*

[11] Fergus Imrie, Anthony R Bradley, Mihaela van der Schaar, and Charlotte M Deane. 2020. Deep generative models for 3D linker design. *Journal of Chemical Information and Modeling* 60, 4 (2020), 1983–1995.

[12] Ross Irwin, Spyridon Dimitriadis, Jiazhen He, and Esben Jannik Bjerrum. 2022. Chemformer: a pre-trained transformer for computational chemistry. *Machine Learning: Science and Technology* 3, 1 (2022), 015022.

[13] Bowen Jing, Gabriele Corso, Jeffrey Chang, Regina Barzilay, and Tommi Jaakkola. 2022. Torsional diffusion for molecular conformer generation. In *Advances in Neural Information Processing Systems.*

[14] Simon Johansson, Amol Thakkar, Thierry Kogej, Esben Bjerrum, Samuel Genheden, Tomas Bastys, Christos Kannas, Alexander Schliep, Hongming Chen, and Ola Engkvist. 2020. AI-assisted synthesis prediction. *Drug Discovery Today: Technologies* 32 (2020), 65–72.

[15] Diederik Kingma, Tim Salimans, Ben Poole, and Jonathan Ho. 2021. Variational diffusion models. In *Advances in Neural Information Processing Systems.*

[16] Bowen Liu, Bharath Ramsundar, Prasad Kawthekar, Jade Shi, Joseph Gomes, Quang Luu Nguyen, Stephen Ho, Jack Sloane, Paul Wender, and Vijay Pande. 2017. Retrosynthetic reaction prediction using neural sequence-to-sequence models. *ACS Central Science* 3, 10 (2017), 1103–1113.

[17] Daniel Lowe. 2017. Chemical Reactions from US Patents (1976-Sep2016). https://figshare.com/articles/dataset/Chemical_reactions_from_US_patents_1976-Sep2016_/5104873

[18] Shitong Luo, Yufeng Su, Xingang Peng, Sheng Wang, Jian Peng, and Jianzhu Ma. 2022. Antigen-specific antibody design and optimization with diffusion-based generative models for protein structures. In *Advances in Neural Information Processing Systems.*

[19] Noel M O'Boyle, Michael Banck, Craig A James, Chris Morley, Tim Vandermeersch, and Geoffrey R Hutchison. 2011. Open Babel: An open chemical toolbox. *Journal of Cheminformatics* 3, 1 (2011), 1–14.

[20] Marcus Olivecrona, Thomas Blaschke, Ola Engkvist, and Hongming Chen. 2017. Molecular de-novo design through deep reinforcement learning. *Journal of Cheminformatics* 9, 1 (2017), 1–14.

[21] Mikołaj Sacha, Mikołaj Błaz, Piotr Byrski, Paweł Dabrowski-Tumanski, Mikołaj Chrominski, Rafał Loska, Paweł Włodarczyk-Pruszynski, and Stanisław Jastrzebski. 2021. Molecule edit graph attention network: modeling chemical reactions as sequences of graph edits. *Journal of Chemical Information and Modeling* 61, 7 (2021), 3273–3284.

[22] Victor Garcia Satorras, Emiel Hogeboom, Fabian Fuchs, Ingmar Posner, and Max Welling. 2021. E(n) equivariant normalizing flows. In *Advances in Neural*

Information Processing Systems.

[23] Michael Schlichtkrull, Thomas N. Kipf, Peter Bloem, Rianne van den Berg, Ivan Titov, and Max Welling. 2017. Modeling Relational Data with Graph Convolutional Networks. In *arXiv:1703.06103.*

[24] Marwin H. S. Segler, Thierry Kogej, Christian Tyrchan, and Mark P. Waller. 2018. Generating focused molecule libraries for drug discovery with recurrent neural networks. *ACS Central Science* 4, 1 (2018), 120–131.

[25] Philipp Seidl, Philipp Renz, Natalia Dyubankova, Paulo Neves, Jonas Verhoeven, Jörg K. Wegner, Marwin Segler, Sepp Hochreiter, and Günter Klambauer. 2022. Improving Few- and Zero-Shot Reaction Template Prediction Using Modern Hopfield Networks. *Journal of Chemical Information and Modeling* (2022).

[26] Seung-Woo Seo, You Young Song, June Yong Yang, Seohui Bae, Hankook Lee, Jinwoo Shin, Sung Ju Hwang, and Eunho Yang. 2021. GTA: Graph truncated attention for retrosynthesis. In *AAAI.*

[27] Chence Shi, Shitong Luo, Minkai Xu, and Jian Tang. 2021. Learning gradient fields for molecular conformation generation. In *Proc. Int. Conf. Machine Learning.*

[28] Chence Shi, Minkai Xu, Hongyu Guo, Ming Zhang, and Jian Tang. 2020. A graph to graphs framework for retrosynthesis prediction. In *Proc. Int. Conf. Machine Learning.*

[29] Jascha Sohl-Dickstein, Eric A. Weiss, Niru Maheswaranathan, and Surya Ganguli. 2015. Deep unsupervised learning using nonequilibrium thermodynamics. In *Proc. Int. Conf. Machine Learning.*

[30] Vignesh Ram Somnath, Charlotte Bunne, Connor Coley, Andreas Krause, and Regina Barzilay. 2021. Learning graph models for retrosynthesis prediction. In *Advances in Neural Information Processing Systems.*

[31] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. 2021. Score-based generative modeling through stochastic differential equations. In *Proc. Int. Conf. Learning Representations.*

[32] Ruoxi Sun, Hanjun Dai, Li Li, Steven Kearnes, and Bo Dai. 2020. Energy-based view of retrosynthesis. In *arXiv:2007.13437.*

[33] Igor V Tetko, Pavel Karpov, Ruud Van Deursen, and Guillaume Godin. 2020. State-of-the-art augmented NLP transformer models for direct and single-step retrosynthesis. *Nature Communications* 11, 1 (2020), 5575.

[34] Brian L Trippe, Jason Yim, Doug Tischer, David Baker, Tamara Broderick, Regina Barzilay, and Tommi Jaakkola. 2022. Diffusion probabilistic modeling of protein backbones in 3D for the motif-scaffolding problem. In *arXiv:2206.04119.*

[35] Zhengkai Tu and Connor W Coley. 2022. Permutation invariant graph-to-sequence model for template-free retrosynthesis and reaction prediction. *Journal of Chemical Information and Modeling* 62, 15 (2022), 3503–3513.

[36] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems.*

[37] Petar Velickovic, William Fedus, William L. Hamilton, Pietro Liò, Yoshua Bengio, and R. Devon Hjelm. 2019. Deep graph infomax. In *Proc. Int. Conf. Learning Representations.*

[38] Yue Wan, Chang-Yu Hsieh, Ben Liao, and Shengyu Zhang. 2022. Retroformer: Pushing the limits of end-to-end retrosynthesis transformer. In *Proc. Int. Conf. Machine Learning.*

[39] Yu Wang, Chao Pang, Yuzhe Wang, Junru Jin, Jingjie Zhang, Xiangxiang Zeng, Ran Su, Quan Zou, and Leyi Wei. 2023. Retrosynthesis prediction with an interpretable deep-learning framework based on molecular assembly tasks. *Nature Communications* (2023).

[40] Tian Xie, Xiang Fu, Octavian-Eugen Ganea, Regina Barzilay, and Tommi Jaakkola. 2022. Crystal diffusion variational autoencoder for periodic material generation. In *Proc. Int. Conf. Learning Representations.*

[41] Minkai Xu, Lantao Yu, Yang Song, Chence Shi, Stefano Ermon, and Jian Tang. 2022. Geodiff: A geometric diffusion model for molecular conformation generation. In *arXiv:2203.02923.*

[42] Chaochao Yan, Qianggang Ding, Peilin Zhao, Shuangjia Zheng, Jinyu Yang, Yang Yu, and Junzhou Huang. 2020. Retroxpert: Decompose retrosynthesis prediction like a chemist. In *Advances in Neural Information Processing Systems.*

[43] Chaochao Yan, Peilin Zhao, Chan Lu, Yang Yu, and Junzhou Huang. 2022. RetroComposer: Composing Templates for Template-Based Retrosynthesis Prediction. *Biomolecules* (2022).

[44] Jiaxuan You, , Bowen Liu, Rex Ying, Vijay Pande, and Jure Leskovec. 2018. Graph convolutional policy network for goal-directed molecular graph generation. In *Advances in Neural Information Processing Systems.*

[45] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. 2023. Adding conditional control to text-to-image diffusion models. In *Proc. IEEE Int. Conf. Computer Vision.*

[46] Shuangjia Zheng, Shuangjia Zheng, Jiahua Rao, Zhongyue Zhang, Jun Xu, and Yuedong Yang. 2020. Predicting retrosynthetic reactions using self-corrected transformer neural networks. *Journal of chemical information and modeling* 60, 1 (2020), 47–55.

[47] Zhaocheng Zhu, Chence Shi, Zuobai Zhang, Shengchao Liu, Minghao Xu, Xinyu Yuan, Yangtian Zhang, Junkun Chen, Huiyu Cai, Jiarui Lu, et al. 2022. Torchdrug: A powerful and flexible machine learning platform for drug discovery. In *arXiv:2202.08320.*

A APPENDIX

A.1 Dataset Details

The USPTO-50k dataset is obtained from the open soruce patent database [17], which includes approximately 50,000 chemical reactions divided into 10 classes. It should be noted that chemical reactions containing multiple products are split into multiple single-product reactions, with each reaction preserving the reactants from the original reaction. Reactions involving trivial products, such as inorganic ions and solvent molecules, are eliminated. In Section 4.4, we visualize the predictions under different classes to demonstrate the effectiveness of the proposed method. Details of the classes in the dataset are shown in Table 3. Furthermore, the dataset is divided into training, validation, and test sets in an 8:1:1 ratio.

Table 3: Information about the classes in the USPTO-50k dataset.

Class Name	#Examples
Heteroatom alkylation and arylation	15204
Acylation and related processes	11972
Deprotections	8405
C-C bond formation	5667
Reductions	4642
Functional group interconversion (FGI)	1858
Heterocycle formation	909
Oxidations	822
Protections	672
Functional group addition (FGA)	231

A.2 Equivariance

In the diffusion-based reactant generation, a node is described by its 3D coordinates $\mathbf{u}^{(\mathrm{x})}$ and a feature vector $\mathbf{u}^{(\mathrm{h})}$. Processing such features associated with 3D coordinates requires operations that respect the symmetry of the data. More specifically, we want the features $\mathbf{u}^{(\mathrm{h})}$ to be invariant to group transformations, while the positions $\mathbf{u}^{(\mathrm{x})}$ are affected by rotations and translations. Formally, an abstract encoding function $f(\cdot)$ needs to satisfy:

$$\begin{cases} \mathbf{z}^{(\mathrm{x})}, \mathbf{z}^{(\mathrm{h})} = f\left(\mathbf{u}^{(\mathrm{x})}, \mathbf{u}^{(\mathrm{h})}\right), \\ \mathbf{U}\mathbf{z}^{(\mathrm{x})} + \mathbf{t}, \mathbf{z}^{(\mathrm{h})} = f\left(\mathbf{U}\mathbf{u}^{(\mathrm{x})} + \mathbf{t}, \mathbf{u}^{(\mathrm{h})}\right), \end{cases} \tag{22}$$

where $\mathbf{U}$ is a orthogonal matrix ($\mathbf{U}\mathbf{U}^T = \mathbf{I}$) to rotate the input, $\mathbf{t}$ is a translation vector represents the input translation. It can be seen from (22) that the node features satisfy permutation invariance. Permutation invariance is common in classical graph neural networks, where nodes do not have an intrinsic order. In addition to the permutation invariance, (22) introduces an equivariant constraint for the 3D coordinates of atoms. In other words, when there is a transformation in the input coordinate attributes, such as rotation, we desire the model's output to undergo the same transformation. It is easy to verify that the encoding layer in (21) satisfies the aforementioned requirements of the perturbation invariance and equivariance constraint. More specially, the update of node features

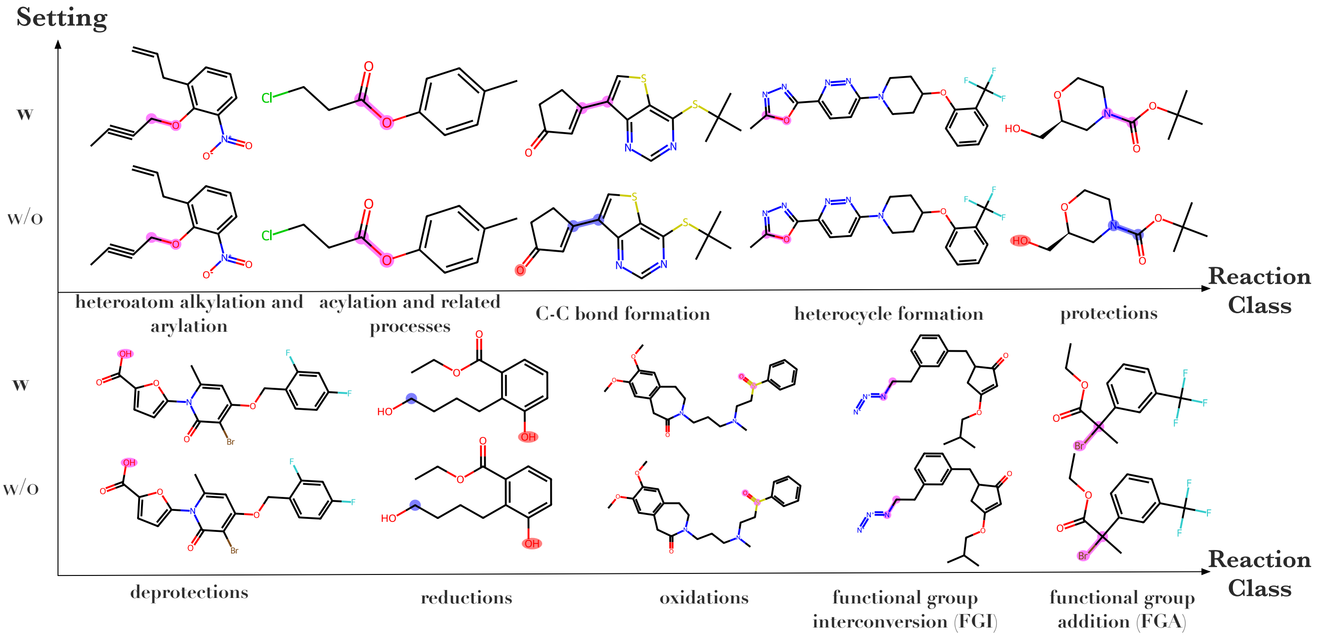
Setting
w
w/o
Reaction Class
heteroatom alkylation and arylation
acylation and related processes
C-C bond formation
heterocycle formation
protections
w
w/o
Reaction Class
deprotections
reductions
oxidations
functional group interconversion (FGI)
functional group addition (FGA)

Figure 7: Illustration of reaction center on 10 example products predicted by GDiffRetro, with dual graphs considered (*w*) and without them (*w/o*). The ground truths and predictions are highlighted in blue and red, respectively. Overlapping areas, indicating agreement, appear purple.

is determined by node features and the perturbation-invariant distance between nodes, while the coordinate update depends linearly on the difference in coordinates between the two nodes (equivariant with respect to rotation and translation).

A.3 Visualization of Reaction Center Identification on different classes

We select 10 typical examples with 10 distinctive reaction classes, as described in APPENDIX A.1, and visualize their identified reaction center with dual graph (*w*) and without dual graph (*w/o*) configuration in Figure 7. It is easy to verify that dual graph enhanced molecular representations aids in the more accurate identification of reaction centers. For example, within the class "C-C bond formation", the model with the dual graph enhanced molecular representations correctly identifies the C-C bond as the reaction center, whereas its variant (the model without dual graph enhancement) incorrectly identifies the chemical bond around the oxygen atom as the reaction center.