Structure-aware Interactive Graph Neural Networks for the Prediction of Protein-Ligand Binding Affinity

Shuangli Li[1,2†], Jingbo Zhou[2*], Tong Xu[1], Liang Huang[4,5], Fan Wang[3]
Haoyi Xiong[3], Weili Huang[3,6], Dejing Dou[2*], Hui Xiong[7*]

[1]University of Science and Technology of China, [2]Business Intelligence Lab, Baidu Research
[3]Baidu Inc., [4]Baidu Research USA, [5]Oregon State University, [6]HWL Consulting LLC, [7]Rutgers University
lsl1997@mail.ustc.edu.cn,{zhoujingbo,wangfan04,xionghaoyi,doudejing}@baidu.com
tongxu@ustc.edu.cn,{liang.huang.sh,lwlily99}@gmail.com,hxiong@rutgers.edu

ABSTRACT

Drug discovery often relies on the successful prediction of protein-ligand binding affinity. Recent advances have shown great promise in applying graph neural networks (GNNs) for better affinity prediction by learning the representations of protein-ligand complexes. However, existing solutions usually treat protein-ligand complexes as topological graph data, thus the biomolecular structural information is not fully utilized. The essential long-range interactions among atoms are also neglected in GNN models. To this end, we propose a structure-aware interactive graph neural network (SIGN) which consists of two components: polar-inspired graph attention layers (PGAL) and pairwise interactive pooling (PiPool). Specifically, PGAL iteratively performs the node-edge aggregation process to update embeddings of nodes and edges while preserving the distance and angle information among atoms. Then, PiPool is adopted to gather interactive edges with a subsequent reconstruction loss to reflect the global interactions. Exhaustive experimental study on two benchmarks verifies the superiority of SIGN.

CCS CONCEPTS

• **Computing methodologies** → **Neural networks**; • **Applied computing** → **Bioinformatics**; *Computational biology*.

KEYWORDS

Binding Affinity Prediction; Graph Neural Network;Drug Discovery

†This work was done when the first author was an intern in Baidu Research under the supervision of the second author.
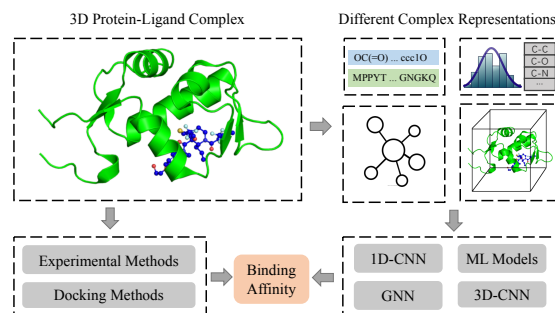*Corresponding authors.

Figure 1: A brief summary for protein-ligand binding affinity prediction. (1) Top left: An example of protein-ligand complex (PDB: 5HMI). (2) Top right: Various representations of complex. (3) Bottom left: Traditional Methods. (4) Bottom right: Machine learning and deep learning methods.

1 INTRODUCTION

The prediction of protein-ligand binding affinity has been widely considered as one of the most important tasks in computational drug discovery [16]. Here ligands are usually drug candidates including small molecules and biologics which can interact with proteins as agonists or inhibitors in the biological processes to cure diseases. The binding affinity, defined as the strength of the binding interaction between a protein and a ligand (e.g., drug), can be measured by experimental methods. However, those biological tests are laborious and time-consuming. With the computer aided simulation methods and the data-driven learning models, binding affinities can be predicted in the early stage of drug discovery. Instead of applying costly biological methods directly to screen numerous candidate molecules, the prediction of binding affinity can help to rank drug candidates and prioritize the appropriate ones for subsequent testing to accelerate the process of drug screening [21].

With the development of structural biology and protein structure prediction, especially the recent Alphafold II model [4], there are growing three-dimensional (3D) structure protein data, which enables a new paradigm for structure-based drug discovery [3, 13, 27]. It has been demonstrated that 3D structural information can effectively contribute to the drug design [18]. Indeed, since there are already many accurate and robust algorithms to find poses of protein-ligand complexes (e.g., binding site prediction methods and docking methods), it is significant to focus on the much harder task of binding affinity prediction [2]. To learn useful 3D structure from a protein-ligand complex, as illustrated in Figure 1, many efforts have been devoted to estimating more accurate binding affinity

for effective drug design. Docking methods [1, 12, 38] play an important role to predict how a specific ligand binds to the target protein with affordable computational costs. While the docking process can identify the binding pose of the protein-ligand complex with relatively high accuracy, its prediction of binding affinity is inaccurate and unreliable [2, 28] due to poor scoring functions, which limits the applicability of docking methods in drug discovery. Compared to docking calculations, traditional machine learning methods [2, 14] have improved the performance by learning the extracted features from protein-ligand complexes. However, these approaches with limited generalizability require expert knowledge and heavily rely on feature engineering.

Recently, deep learning for binding affinity prediction have become an emerging research area, which represents the complex as sequence data [31], 3D grid-like data [40] or graph data [30] to employ various neural networks. One of the key challenges of deep learning in structural biology is how to model the 3D spatial structure for better performance. To this end, most of the existing works [32, 35, 40] attempt to apply 3D convolutional neural networks (3D CNNs) by treating the complex as a 3D-grid representation. However, the cost of these models is huge, especially when considering long-range interactions. What's more, both the absence of topological information and the sensitivity to rotation in the complex have a negative effect on the prediction results.

Despite the powerful ability of graph neural networks (GNNs) to learn graph representations [22, 24, 47], there are only a few studies [23, 30] on using GNNs to predict the protein-ligand binding affinity. By contrast, many researchers have greatly developed GNN models in other fields of drug discovery [37, 45], such as predicting molecular property [17, 26, 44] and chemical reaction [7]. Nevertheless, these domain-specific models tend to lose their effectiveness when modeling the larger biomolecules, e.g., protein-ligand complexes. In general, most of the existing GNNs in drug design aim to learn the spatial structure by incorporating the distance information, which is insufficient to model the 3D structure of complex. Moreover, the fundamental long-range interactive information between proteins and ligands, which is valuable for predicting the binding affinity [19], cannot be handled under the current GNN framework.

To overcome the above limitations, we propose a novel Structure-aware Interactive Graph Neural Network (SIGN) to learn the constructed complex graph for predicting the protein-ligand binding affinity. SIGN is equipped with two designed components to correspondingly address the challenges, namely the *polar-inspired graph attention layers (PGAL)* for modeling 3D spatial structure and the *pairwise interactive pooling (PiPool)* for leveraging long-range interactions. Firstly, the key idea of PGAL is to establish a polar coordinate system for each central target and to preserve both distance and angle information of neighbors when performing the aggregation process. More specifically, we apply the node-edge interactive scheme iteratively with graph attention to integrate spatial factors for effectively learning the 3D structure of complex.

In view of the large size of the protein, it is redundant to contain the complete protein structure in the complex graph, but in this way the long-range interactive information between the protein and the ligand is also lost. To deal with this issue, PiPool, the secondary part of SIGN is designed to incorporate such global interactions into our model, which employs an atomic type-aware pooling process

on edges with introducing an auxiliary learning task to reconstruct the atomic interaction matrix. By this means, SIGN can enhance the representation learning for complexes with involving both 3D spatial structures and global interactions. To summarize, the main contributions of this paper are as follows:

- To the best of our knowledge, we are among the first to develop graph neural networks from the perspective of polar coordinates for structure-based binding affinity prediction.
- We propose a novel structure-aware interactive graph neural network (SIGN), which can capture not only 3D spatial information through polar-inspired graph attention layers (PGAL), but also global long-range interactions through pairwise interactive pooling (PiPool) in a semi-supervised manner.
- We conduct extensive experiments using two benchmark datasets to evaluate the performance of the proposed model, which demonstrates the effectiveness of our SIGN with better generalizability.

2 RELATED WORK

In this section, we first review the related literatures about predicting protein-ligand binding affinity and then detail recent advances in graph neural networks for drug discovery.

Protein-Ligand Binding Affinity Prediction. As a crucial stage in drug discovery, predicting protein-ligand binding affinity has been intensively studied for a long time [11, 34], which is of great importance for efficient and accurate drug screening. The earlier empirical-based methods [9, 38, 42] design docking and scoring functions specially to make predictions, while expert domain knowledge is required to encode internal biochemical interactions. Later on, statistical and machine learning-based methods [5] are developed to predict binding affinity based on data-driven learning, which attempt to extract protein-ligand features and use classic models for regression, such as random forest [2] and SVM [14]. These approaches are dependent on the quality of hand-crafted features and lack of generality on the larger dataset. Recently, several deep learning-based models [31] utilize 1D convolutions and pooling to capture potential patterns from raw sequence information of both ligand and protein. However, only using separate character representations fails to achieve desirable performance.

With the recent advances in predicting structures of proteins [4] and the increasing availability of 3D-structure protein-ligand data [41], there is another hot research area of studying structure-based approaches, which focus on learning from 3D-structure protein-ligand complexes to predict binding affinity. Some recent works [32, 35] represent the protein-ligand complex as 3D grid-like data and use 3D convolutions (3D-CNNs) to take advantage of spatially-local correlations. Though these approaches can learn spatial information, one limitation is that positions of proteins and ligands in different complexes are changeable, such as different angle rotations, which means the spatial structure of 3D grid-like modeling is inevitably incomplete. More recently, OnionNet [46] employs CNN models to learn the complex representation from the extracted element-specific interaction features between a protein and its ligand. However, all the above models neglect the critical topological structure information of complex. In the work [23], a protein-ligand complex is represented as a weighted graph with distance information. Then graph attention networks are applied to predicting the interactions. Nevertheless, only distance information between atoms

Table 1: Mathematical notations.

Notation	Description
$\mathcal{V}^P, \mathcal{V}^L$	The atom node sets of protein and ligand
M^P, M^L	The 3D position matrices of protein and ligand
$\mathcal{G}_I$	The complex interaction graph
a_i	The i-th atom node in $\mathcal{G}_I$
e_{ij}	The directed edge from atom a_i to atom a_j
$\mathcal{N}_e(a_i)$	The neighboring edges of atom a_i
$\mathcal{N}_e(e_{ij})$	The neighboring edges of edge e_{ij}
$\boldsymbol{a}_i, \boldsymbol{e}_{ij}$	The embedding vectors of atom a_i and edge e_{ij}
$\boldsymbol{d}_{ij}$	The spatial embedding vector between a_i and a_j

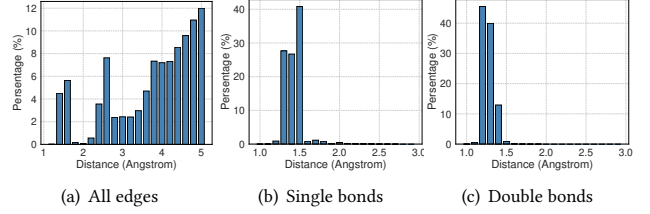(a) 3D Protein-Ligand Complex (b) Complex Interaction Graph
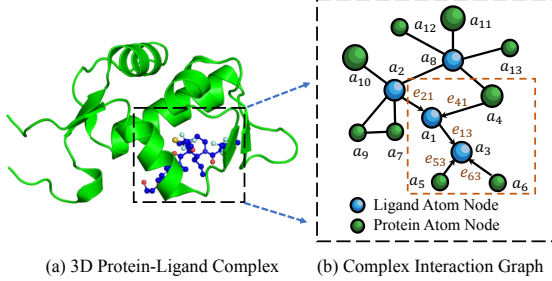
Figure 2: An illustrative example of converting the protein-ligand complex into a complex interaction graph.

is not adequate to model 3D-structure interactions. In this paper, we also focus on the structure-based prediction of protein-ligand binding affinity with incorporating abundant spatial information.

Graph Neural Networks for Drug Discovery. Inspired by the great advantage of graph neural networks (GNNs) in modeling graph data, much attention has been devoted to applying them in computational drug discovery [37], such as the prediction of molecular property [10] and protein interface [25]. Treating the molecule as a graph, GNNs can learn the graph-level representation for drug or protein by aggregating structural information. GraphDTA [30] adopts GNN models [15, 39, 43] to learn drug presentation with combining the protein representation from 1D convolutions to predict binding affinity. In attributed molecular graphs, the edges between atoms contain valuable information, such as distance or bond order. To leverage rich attributes in the molecule, edge-oriented message passing neural networks [33, 44, 48] are proposed to update both node and edge embeddings . Meanwhile, there are also some efforts to model the 3D-structure of molecule by improving GNNs with spatial information, such as distance [23, 26], angle [17], and 3D coordinate [6]. However, these models fail to consider the spatial interactions between proteins and ligands. In addition, the function of learning angle information in [17] is designed for density functional theory, which is only beneficial for predicting molecular properties rather than protein-ligand binding affinity. To overcome these limitations, we propose an interaction-aware GNN framework with integrating both distance and angle factors harmoniously.

3 PRELIMINARIES

In this section, we introduce some definitions used in our model and formulate the structure-based prediction problem for protein-ligand binding affinity. The frequently used notations in this paper are summarized in Table 1.

(a) All edges (b) Single bonds (c) Double bonds

Figure 3: The distribution of distance between atoms within 5 Å in the protein-ligand complex from PDBbind dataset.

Definition 3.1. **Complex Interaction Graph.** Given a protein-ligand complex as shown in Figure 2(a), we define the atom node sets of protein and ligand as $\mathcal{V}^P = \{a_1^P, ..., a_m^P\}$ and $\mathcal{V}^L = \{a_1^L, ..., a_n^L\}$ with the position matrix $M^P \in \mathbb{R}^{m \times 3}$ and $M^L \in \mathbb{R}^{n \times 3}$ for 3D atomic coordinates, respectively. Then we define the complex interaction graph as a directional graph $\mathcal{G}_I = <\mathcal{V}, \mathcal{E}>$, where the vertex set $\mathcal{V} \subseteq \mathcal{V}^P \cup \mathcal{V}^L$ and the edge set $\mathcal{E} = f_e(\mathcal{V}^P, \mathcal{V}^L, M^P, M^L)$ are constructed based on the spatial positions of atoms. The graph construction process is introduced in Appendix A.1.

Definition 3.2. **Edge-oriented Neighbors.** Given an atom node a_i or a directed edge e_{ij} (i.e., $a_i \rightarrow a_j$) in the complex interaction graph $\mathcal{G}_I$, the edge-oriented neighbors $\mathcal{N}_e$ of a_i or e_{ij} are defined as the sets of directed edges $\{e_{ki}, ..., e_{li}\}$ which point to the target atom a_i or the target edge e_{ij}.

Taking Figure 2(b) as an example, the edges e_{21} and e_{41} are connected to the edge e_{13} via the common node a_1, the edge-oriented neighbors of e_{13} are denoted as $\mathcal{N}_e(e_{13}) = \{e_{21}, e_{41}\}$. Similarly, the edges e_{13}, e_{53} and e_{63} point to the atom node a_3, resulting in the neighbors set $\mathcal{N}_e(a_3) = \{e_{13}, e_{53}, e_{63}\}$.

Definition 3.3. **Structure-based Protein-Ligand Binding Affinity Prediction.** Given a protein-ligand complex with 3D structure, i.e., the complex interaction graph $\mathcal{G}_I$ and the 3D position matrix M consisting of M^P and M^L, our goal is to learn a model $f : (\mathcal{G}_I | M) \rightarrow y$ to precisely predict the binding affinity y with preserving the spatial structure.

4 MODEL FRAMEWORK

In this section, we present the proposed SIGN model for protein-ligand binding affinity prediction. We first introduce the overall framework and then describe the details of each component.

4.1 Overview

To make accurate predictions for protein-ligand binding affinity, there are two challenges. Firstly, as shown in Figure 2, the complex graph has the unique spatial structure, which is different with general graph. Secondly, the long-range interactions between protein and ligand are also critical to the binding affinity [19]. However, the existing GNNs are incapable of capturing such spatial information and interactions. To overcome the limitations, we propose a novel *Structure-aware Interactive Graph Neural Network (SIGN)* to model the 3D structural complex and protein-ligand spatial interactions.

Figure 4 exhibits the architecture which takes the complex interaction graph $\mathcal{G}_I$ as input. We start with the polar-inspired graph attention layers (*PGAL*), which are composed of *node→edge* and *edge→node* interaction layers. PGAL can propagate the node's and
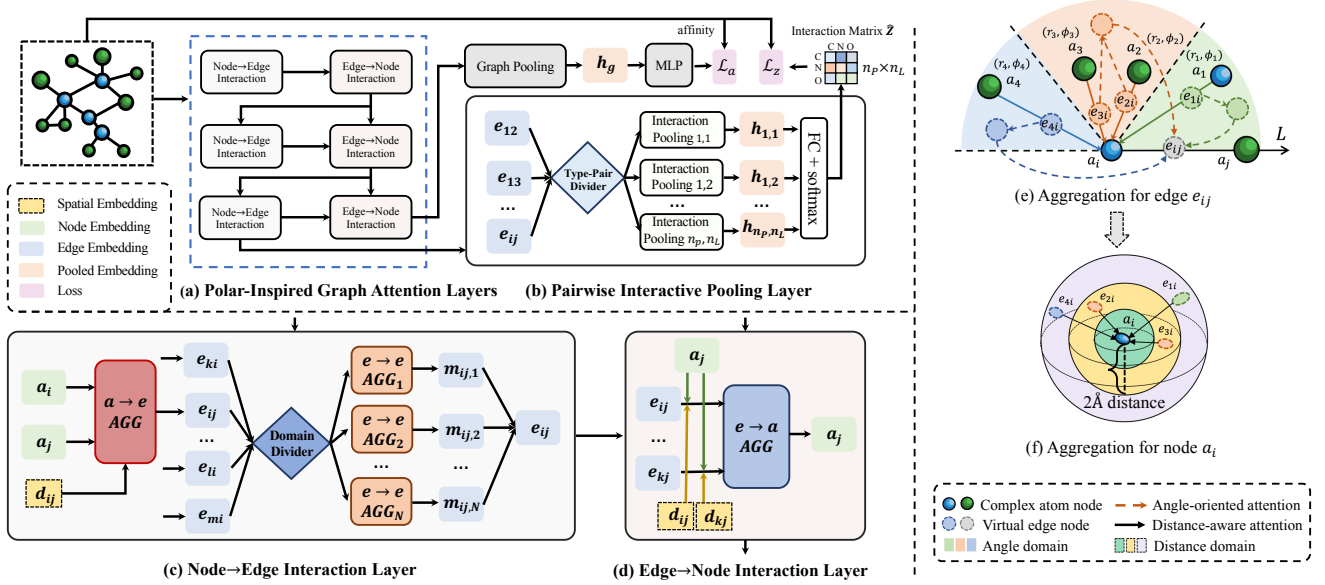
Figure 4: Illustration of the proposed SIGN framework. (a), (b): The two key components PGAL and PiPool. (c), (d): The two inner structures of component PGAL. (e), (f): The aggregation processes in *node→edge* and *edge→node* interaction layers.

edge's embeddings alternately with learning the spatial distance and angle information. The two parts of PGAL play a synergistic effect on modeling the spatial structure of the complex. After that, we apply a pairwise interactive pooling layer (*PiPool*) which performs on the edges' representations to obtain the atomic type-based interaction matrix of the complex. From a global view, PiPool aims to approximate the overall interactions between proteins and ligands to improve the prediction performance. Finally, the model is trained through multi-task learning with augmented constraints for the interaction matrix, which serves as a self-supervised task.

4.2 Polar Coordinate-Inspired Graph Attention

Standard GNNs have shown great advantages in learning topological structure of the general graph, which cannot take atom's spatial position into account in the 3-dimensional space. To model the 3D structure of a complex, an intuitive method is to provide atom's 3-dimensional coordinate in the GNN architecture [6]. However, the position information under the Cartesian coordinate system is sensitive to both translations and rotations, causing poor generalization of model when learning the complex representation. Several models, such as GNN-DTI [17] and MAT [26], manage to combine the distance information in the aggregation process, while only pairwise distance is not adequate. Different from DimeNet [17], which specially designs Bessel functions in GNN for density functional theory (DFT) approximation with limited ability to model the larger biological complex, we employ iterative *node→edge* and *edge→node* interaction layers to incorporate both distance and angle information from a spatial distribution perspective.

4.2.1 Polar-Inspired Attentive Learning Architecture. Inspired by polar coordinate which is composed of radial distance r and polar angle ϕ, we develop an interaction-based graph attention network to leverage both the distance between nodes and the angle between

edges in a collaborative framework. As illustrated in Figure 4(e), when aggregating for edge e_{ij}, we treat it as the polar axis L. Under such a definite polar coordinate system, the edge-oriented neighbors are distributed around e_{ij} with unique identifying coordinates (r, ϕ). Through the method of dividing angle domains, the spatial distribution for the complex can be taken into account by means of angle-oriented attention in the first aggregation stage for edges.

Moreover, the distance factor is also helpful for structure modeling, which reveals spatial correlations. Figure 3 shows the statistical distribution of distance between atoms. It can be seen that the distances of covalent bonds mainly range from 1 to 2 Å, while noncovalent interactions, like hydrophobic and van der Waals interactions, and hydrogen bonds, are distributed over longer distances. Atomic interactions in the complex vary from different distances, which indicate different spatial relations for atom pairs. Given the radial distance r between atoms a_i and a_j, as shown in Figure 4(f), we first map r to a bucket (i.e., a distance domain corresponding to a type of relation) and obtain the one-hot vector x_{ij}. Then we apply a dense layer transformation to get the spatial relation embedding:

$$d_{ij} = W_s x_{ij}, \quad (1)$$

where $W_s \in \mathbb{R}^{d_s \times b}$ is the transformation weight matrix and b is the number of buckets (i.e., spatial relations). To factor in these correlations, we design the distance-aware attention in the second aggregation stage for nodes. As shown in Figure 4(a), the overall attentive interaction process at l-th layer is defined as:

$$e_{ij}^{(l)} = f^{(l)}\left(\left\{(a_k^{(l-1)}, a_i^{(l-1)}), \forall e_{ki} \in \mathcal{N}_e(e_{ij})\right\}\right), \quad (2)$$

$$a_j^{(l)} = g^{(l)}\left(a_j^{(l-1)}, \left\{e_{kj}^{(l)}, \forall e_{kj} \in \mathcal{N}_e(a_j)\right\}\right), \quad (3)$$

where $e_{ij}^{(l)}$ is the edge embedding, $a_j^{(l)}$ is the node (atom) embedding, $f(\cdot)$ and $g(\cdot)$ are interaction functions of *node→edge* and

$edge \rightarrow node$ layers, $\mathcal{N}_e(e_{ij})$ and $\mathcal{N}_e(a_j)$ are the edge-oriented neighbors of edge e_{ij} and node a_j respectively.

4.2.2 Angle-oriented Node→Edge Interaction Layer.

Failing to distinguish neighbor nodes from different directions in the aggregation process is a weakness of the existing GNN models. To overcome this inadequacy, we adopt an angle-oriented graph attention layer to update the edge representations with integrating spatial angle information. Since the angle exists between the two edges, as shown in Figure 4(c), we first get the edge embedding through aggregating the node features:

$$e_{ij}^{(l)} = \sigma(W_{a \rightarrow e}^{(l)} \cdot [a_i^{(l-1)} \parallel a_j^{(l-1)} \parallel d_{ij}]), \qquad (4)$$

where $W_{a \rightarrow e}^{(l)}$ is the transformation matrix for atomic combination, the operator $\parallel$ represents concatenation, and σ is the Relu function.

After obtaining the representations $\{e_{ij}^{(l)}, e_{ki}^{(l)}, .., e_{mi}^{(l)}\}$ of edge e_{ij} and its neighbors, we further separate the neighboring edges in 3-dimensional space by applying an angle-domain divider D_A, which plays an intermediate role to assign each neighbor to the specific angle domain. For example, in Figure 4(e), there are four edge-oriented neighbors e_{1i}, e_{2i}, e_{3i} and e_{4i} around the central target edge e_{ij}. These neighboring edges are located in three different local angle domains according to the angles between edge e_{ij} and its neighbors. Given the number of angle domains N (e.g., $N = 3$ in Figure 4(e)) and the target edge e_{ij} for aggregation, D_A can map each neighbor e_{ki} to the located angle domain index:

$$Ind_{ki} = D_A(e_{ki}, e_{ij}, N) = \lceil N \cdot \frac{\phi_{kij}}{180} \rceil, \qquad (5)$$

where $\lceil \cdot \rceil$ denotes rounding operation to get the integer index, $\phi_{kij} \in [0, 180°]$ is the calculated angle between edges e_{ki} and e_{ij}. Then the subset of edge-oriented neighbors which are located in q-th angle domain can be defined as:

$$\mathcal{N}_e^q(e_{ij}) = \{e_{ki} \mid e_{ki} \in \mathcal{N}_e(e_{ij}) \wedge Ind_{ki} = q\}. \qquad (6)$$

After reorganizing the neighbors of e_{ij} through divider D_A based on the polar coordinate system, we then feed N neighbor subsets from different angle domains into N independent propagation layers to capture long-range dependencies in the complex interaction graph. Firstly, we devise the domain-specific aggregation process along edges for the q-th angle domain:

$$m_{ij,q}^{(l)} = \sum_{e_{ki} \in \mathcal{N}_e^q(e_{ij})} \alpha_{ki,q}^{(l)} \cdot e_{ki}^{(l)}, \quad 1 \le q \le N, \qquad (7)$$

where $m_{ij,q}^{(l)}$ is the q-th local aggregated edge representation at l-th layer, $\alpha_{ki,q}^{(l)}$ is the attention weight of the neighboring edge e_{ki} across the q-th angle domain. Concretely, we apply the angle-oriented attention mechanism, which first uses $attn_q^l$ function to calculate the coefficient between two edges and then adopts the softmax function for normalization:

$$attn_q^l(e_{ij}, e_{ki}) = u_{l,q}^T \cdot tanh(W_{e,q}^{(l)} \cdot [e_{ij}^{(l)} \parallel e_{ki}^{(l)}] + b_{e,q}^{(l)}), \qquad (8)$$

$$\alpha_{ki,q}^{(l)} = \frac{\exp(attn_q^l(e_{ij}, e_{ki}))}{\sum_{e_{ti} \in \mathcal{N}_e^q(e_{ij})} \exp(attn_q^l(e_{ij}, e_{ti}))}, \qquad (9)$$

where $u_{l,q}, W_{e,q}^{(l)}$ and $b_{e,q}^{(l)}$ are the learnable attention parameters of the specific q-th angle domain, and we use tanh as the nonlinear activation function.

Secondly, we combine all aggregated edge embeddings obtained from Eq. (7). To completely preserve the spatial information in different local angle domains, we concatenate the representations as the global aggregation to update the angle-aware edge embedding:

$$e_{ij}^{(l)} = [m_{ij,1}^{(l)} \parallel m_{ij,2}^{(l)} \parallel \cdots \parallel m_{ij,N}^{(l)}]. \qquad (10)$$

4.2.3 Distance-aware Edge→Node Interaction Layer.

After injecting the angle information into the edge embedding $e_{ij}^{(l)}$, we make further efforts to develop an attention-based edge→node interaction layer to incorporate another spatial factor in the polar coordinate system, that is distance. As we stated in Section 4.2.1, the distance between atoms is implicated in different meaningful correlations. Therefore, it's momentous to explore the influence of distance while learning representations for protein-ligand complexes. Specifically, since edges and nodes (atoms) have different feature spaces, we first convert the edge embedding and node embedding into the hidden representation $\tilde{e}_{ij}^{(l)}$ and $\tilde{a}_j^{(l-1)}$ in the same vector space:

$$\tilde{e}_{ij}^{(l)} = W_e^{(l)} \cdot e_{ij}^{(l)}, \qquad (11)$$

$$\tilde{a}_j^{(l)} = W_a^{(l)} \cdot a_j^{(l-1)}, \qquad (12)$$

where $W_e^{(l)}$ and $W_a^{(l)}$ are linear transformation matrices, $a_j^{(l-1)}$ is the embedding of atom a_j from $(l-1)$-th layer.

As a result of the variant distances and atomic attributes, the neighboring edges have different impacts on the target node. However, the existing GNN models cannot effectively capture the influence of the distance factor. Hence, as shown in Figure 4(d) and 4(f), we propose to extend the original GAT [39] with the distance-aware attention to fuse the distance information with the capability of discriminating multiple spatial relations among atoms:

$$w_{ij}^{(l)} = LeakyRelu(v_l^T \cdot [\tilde{e}_{ij}^{(l)} \parallel \tilde{a}_j^{(l)} \parallel W_d^{(l)} d_{ij}]), \qquad (13)$$

$$\beta_{ij}^{(l)} = \frac{\exp(w_{ij}^{(l)})}{\sum_{e_{tj} \in \mathcal{N}_e(a_j)} \exp(w_{tj}^{(l)})}, \qquad (14)$$

where v_l is the parameter of edge→node attention at l-th layer, $W_d^{(l)}$ is the trainable parameter matrix for distance transformation, the final calculated attention weight $\beta_{ij}^{(l)}$ reflects how important the edge e_{ij} is for the node a_j. Then we develop the distance-aware attention to multi-head attention version as GAT for better stability and apply the aggregation process from edge to node:

$$a_j^{(l)} = \frac{1}{C} \sum_{c=1}^{C} \sum_{e_{ij} \in \mathcal{N}_e(a_j)} \beta_{ij,c}^{(l)} \cdot \tilde{e}_{ij,c}^{(l)}, \qquad (15)$$

where C is the number of independent attention heads. Due to the angle injection for edge embedding $\tilde{e}_{ij,c}^{(l)}$ and the distance injection for attention weight $\beta_{ij,c}^{(l)}$, our proposed model can comprehensively incorporate spatial information in the complex.

After performing L polar-inspired graph attention layers, we obtain the node embedding $a_j^{(L)}$ for atom a_j and the edge embedding $e_{ij}^{(L)}$ between atoms a_i and a_j.

4.3 Pairwise Interactive Pooling Constraint

As introduced in Appendix A.1, the constructed complex graph $\mathcal{G}_I$ only contains the partial protein structure due to the limitation of graph size and needless noise. However, the long-range intermolecular interactions between protein and ligand have effects on the binding affinity [2, 19], while $\mathcal{G}_I$ cannot provide such interactive information. To capture the long-range interactions in the complex (e.g., the Carbon-Carbon co-occurrence interaction), we design an atomic type-aware pooling layer for edges between the protein and the ligand, which generates a proximity interaction matrix of atom type pair and enhances the representation learning process through the additional self-supervised training.

Specifically, we first construct the pairwise interaction matrix $Z \in \mathbb{R}^{|S_P| \times |S_L|}$ from the complete protein and its ligand, where S_P and S_L are atomic type sets of the protein and its ligand. Each element T_k in S_P or T_l in S_L represents the atomic number (e.g., 6) of a certain atom (e.g., carbon atom C). Following the previous work [2], we calculate the number of occurrences for a specific atomic type pair (T_k, T_l) (e.g., (6, 7) for <C,N> pair) within a certain distance and normalize the result to get the matrix Z:

$$n(T_k, T_l) = \sum_{a_i \in \mathcal{V}^P} \sum_{a_j \in \mathcal{V}^L} \delta(\tau(a_i), T_k)\delta(\tau(a_j), T_l)\Theta(d_\rho - d_{ij}), \tag{16}$$

$$Z_{kl} = \frac{n(T_k, T_l)}{\sum_{(a_i, a_j) \in \mathcal{V}^P \times \mathcal{V}^L} \Theta(d_\rho - d_{ij})}, \tag{17}$$

where the function $\tau(a_i)$ returns the atomic type of a_i, $\delta(\cdot, \cdot)$ is a Kronecker delta function which outputs 1 only if the type of atom is T_k (or T_l) and 0 otherwise, d_ρ is referred to as the interaction cutoff distance and a Heaviside step function Θ is adopted to count protein–ligand atomic type pairs within the distance d_ρ.

Secondly, we take the edge embeddings obtained from PGAL as input to the atomic type-aware pooling layer, which is shown in Figure 4(b). There are $|S_P| \times |S_L|$ pooling blocks for type pairs. One block to gather edge representations belonging to atomic type pair (T_k, T_l) can be formulated as:

$$h_{k,l} = \sum_{e_{ij} \in \mathcal{E}_I} \underbrace{\delta(\tau(a_i), T_k)\delta(\tau(a_j), T_l)}_{\text{Divider}} W_h e_{ij}^{(L)}, \tag{18}$$

where W_h is the shared parameter matrix for edge pooling, $\mathcal{E}_I \subset \mathcal{E}$ contains all the intermolecular edges in the complex G_I, a_i and a_j are atom nodes connected by e_{ij}, the two $\delta(\cdot, \cdot)$ functions act as a divider to pick up the corresponding edges. Then we calculate each value of the approximate interaction matrix:

$$\tilde{Z}_{kl} = \frac{\exp(q^T h_{k,l})}{\sum_{i,j} \exp(q^T h_{i,j})}, \tag{19}$$

where q is the trainable parameter. In the training stage, we use an additional proximity loss to draw the interaction matrix $\tilde{Z}$ and Z

closer:

$$\mathcal{L}_z = \sum_{\mathcal{G}_I \in \mathcal{D}} \|F(\tilde{Z}) - F(Z)\|, \tag{20}$$

where $F(\cdot)$ is the flatten operation for matrix, $\mathcal{D}$ is the training set.

4.4 Optimization Objective

At the last part, we add together node (atom) embeddings to get the complex representation and use MLP layers as the regressor to predict the protein-ligand binding affinity:

$$\hat{y} = MLP\Big(\sum_{a_i \in \mathcal{V}} a_i^{(L)}\Big). \tag{21}$$

Then the absolute error between the predicted binding affinity $\hat{y}$ and the measured ground truth y is used to calculate the loss. Thus, we adopt the L1 loss function to optimize the model:

$$\mathcal{L}_a = \sum_{\mathcal{G}_I \in \mathcal{D}} |\hat{y} - y|, \tag{22}$$

where $\mathcal{D}$ contains all the protein-ligand complexes with binding affinities. To integrate the interaction effectiveness for better complex representation learning, we further combine with the complex interaction constraint in Eq. (20) and reach the following overall objective function:

$$\mathcal{L} = \mathcal{L}_a + \lambda \mathcal{L}_z, \tag{23}$$

where λ is the balancing hyper-parameter to control the strength of interaction loss. The detailed process for training the proposed SIGN is provided in Algorithm 2.

5 EXPERIMENTS

In this section, we conduct experiments on two standard datasets to investigate the following research questions:

- **RQ1.** How does the proposed SIGN model perform compared against the state-of-the-art methods?
- **RQ2.** How does the generalizability of SIGN and competitors when trained on the larger but lower-quality dataset?
- **RQ3.** Do the spatial and interactive factors benefit the prediction?
- **RQ4.** How do the parameter settings (e.g., the cutoff distance and angle domain divisions) affect the prediction result?

5.1 Experiment Settings

5.1.1 Datasets. We evaluate all models on the following public standard datasets for protein-ligand binding affinity prediction.

PDBbind[1] is a well-known public dataset [41] in development which provides 3D binding structures of protein-ligand complexes with experimentally determined binding affinities (refer to Appendix A.2). In our experiment, we mainly use the PDBbind v2016 dataset, which is most frequently used in recent works [35, 46]. Specifically, it includes three overlapping subsets, i.e., *general*, *refined* and *core set*. The *general set* contains all 13,283 protein-ligand complexes, while the 4,057 complexes in *refined set* are selected out of the *general set* with better quality. Moreover, the *core set* with 290 complexes serves as the highest quality benchmark for testing through a careful selection process [36]. Conveniently, we call the difference between the *refined* and *core* subsets, that is 3,767 complexes, as *refined set* of PDBbind in the following.

[1]http://www.pdbbind-cn.org

Table 2: Performance comparison on *PDBbind core set* and *CSAR-HiQ set*.

Method		PDBbind core set				CSAR-HiQ set			
		RMSE ↓	MAE ↓	SD ↓	R ↑	RMSE ↓	MAE ↓	SD ↓	R ↑
ML-based Methods	LR	1.675 (0.000)	1.358 (0.000)	1.612 (0.000)	0.671 (0.000)	2.071 (0.000)	1.622 (0.000)	1.973 (0.000)	0.652 (0.000)
	SVR	1.555 (0.000)	1.264 (0.000)	1.493 (0.000)	0.727 (0.000)	1.995 (0.000)	1.553 (0.000)	1.911 (0.000)	0.679 (0.000)
	RF-Score	1.446 (0.008)	1.161 (0.007)	1.335 (0.010)	0.789(0.003)	1.947 (0.012)	1.466 (0.009)	1.796 (0.020)	0.723 (0.007)
CNN-based Methods	Pafnucy	1.585 (0.013)	1.284 (0.021)	1.563 (0.022)	0.695 (0.011)	1.939 (0.103)	1.562 (0.094)	1.885 (0.071)	0.686 (0.027)
	OnionNet	1.407 (0.034)	1.078 (0.028)	1.391 (0.038)	0.768 (0.014)	1.927 (0.071)	1.471 (0.031)	1.877 (0.097)	0.690 (0.040)
GraphDTA Methods	GCN	1.735 (0.034)	1.343 (0.037)	1.719 (0.027)	0.613 (0.016)	2.324 (0.079)	1.732 (0.065)	2.302 (0.061)	0.464 (0.047)
	GAT	1.765 (0.026)	1.354 (0.033)	1.740 (0.027)	0.601 (0.016)	2.213 (0.053)	1.651 (0.061)	2.215 (0.050)	0.524 (0.032)
	GIN	1.640 (0.044)	1.261 (0.044)	1.621 (0.036)	0.667 (0.018)	2.158 (0.074)	1.624 (0.058)	2.156 (0.088)	0.558 (0.047)
	GAT-GCN	1.562 (0.022)	1.191 (0.016)	1.558 (0.018)	0.697 (0.008)	1.980 (0.055)	1.493 (0.046)	1.969 (0.057)	0.653 (0.026)
GNN-based Methods	SGCN	1.583 (0.033)	1.250 (0.036)	1.582 (0.320)	0.686 (0.015)	1.902 (0.063)	1.472 (0.067)	1.891 (0.077)	0.686 (0.030)
	GNN-DTI	1.492 (0.025)	1.192 (0.032)	1.471 (0.051)	0.736 (0.021)	1.972 (0.061)	1.547 (0.058)	1.834 (0.090)	0.709 (0.035)
	DMPNN	1.493 (0.016)	1.188 (0.009)	1.489 (0.014)	0.729 (0.006)	1.886 (0.026)	1.488 (0.054)	1.865 (0.035)	0.697 (0.013)
	MAT	1.457 (0.037)	1.154 (0.037)	1.445 (0.033)	0.747 (0.013)	1.879 (0.065)	1.435 (0.058)	1.816 (0.083)	0.715 (0.030)
	DimeNet	1.453 (0.027)	1.138 (0.026)	1.434 (0.023)	0.752 (0.010)	1.805 (0.036)	1.338 (0.026)	1.798 (0.027)	0.723 (0.010)
	CMPNN	1.408 (0.028)	1.117 (0.031)	1.399 (0.025)	0.765 (0.009)	1.839 (0.096)	1.411 (0.064)	1.767 (0.103)	0.730 (0.052)
Ours	SIGN	**1.316 (0.031)**	**1.027 (0.025)**	**1.312 (0.035)**	**0.797 (0.012)**	**1.735 (0.031)**	**1.327 (0.040)**	**1.709 (0.044)**	**0.754 (0.014)**
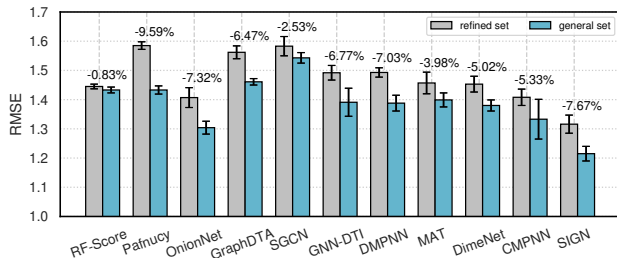
Figure 5: Performance improvements on PDBbind benchmark when training on *general set*.

CSAR-HiQ[2] is an additional benchmark dataset [8], containing two subsets with 176 and 167 protein-ligand complexes. We use this external dataset from an independent source to further evaluate the generalization ability of models.

5.1.2 Setup. Following [2], we choose the *refined set* of PDBbind as our primary training data because there is considerable overlap between the full *general set* and CSAR-HiQ dataset. We randomly split the protein-ligand complexes in *refined set* with a ratio of 9:1 for training and validation. For testing sets, we use the *core set* and CSAR-HiQ set with removing the complexes present in *refined set*.

Since the lower-quality data of *general set* can still improve the performance of models [20], we conduct the supplemental experiment on the full *general set* which is larger but of worse quality to analyze the generalizability of our model. As stated above, we can only evaluate the performance on the *core set* due to the overlapping problem of CSAR-HiQ dataset. Following [35, 46], we randomly select 1,000 complexes from *refined set* as the validating set. The remaining 11,993 complexes in *general set* are used for training.

5.1.3 Evaluation Metrics. To comprehensively evaluate the model performance, following [35, 46], we use Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Pearson's correlation coefficient (R) and the standard deviation (SD) in regression to measure the prediction error. The detail is introduced in Appendix A.3.

5.1.4 Baselines. We compare our proposed model with comparative methods including machine learning-based methods (**LR**, **SVR**,

and **RF-Score** [2]), CNN-based methods (**Pafnucy** [35] and **Onion-Net** [46]), and GNN models **GraphDTA** [30] for protein-ligand binding affinity prediction. Moreover, various state-of-the-art GNN-based models (**SGCN** [6], **GNN-DTI** [23], **DMPNN** [44], **MAT** [26], **DimeNet** [17], and **CMPNN** [33]) which also consider the spatial information for molecular modeling are compared to evaluate the performance of SIGN. The details of experiment settings and baseline descriptions are provided in Appendix A.3 and A.4.

5.2 Performance Evaluation

5.2.1 Overall Comparison (RQ1). We first compare our proposed SIGN with baseline approaches on two benchmark datasets. As shown in Table 2, the average and the standard deviation of four indicators for testing performance are reported across five random runs. In general, we can observe that SIGN achieves the best performance on two datasets, with 6.5% and 3.9% improvement of RMSE over the best baseline models on PDBbind and CSAR-HiQ datasets, respectively. We further have the following observations.

Among all baselines, GraphDTA methods show relatively poor performance due to the failure of considering the spatial structure and interactions between proteins and ligands. It indicates that simply modeling the molecular graph with protein sequence information is not capable of predicting structure-based protein-ligand binding affinity. By contrast, from the perspective of inter-action modeling, the machine learning-based methods and Onion-Net model take advantage of long-range interaction features and achieve better results. However, these data-driven approaches relying on feature engineering ignore the informative spatial structures
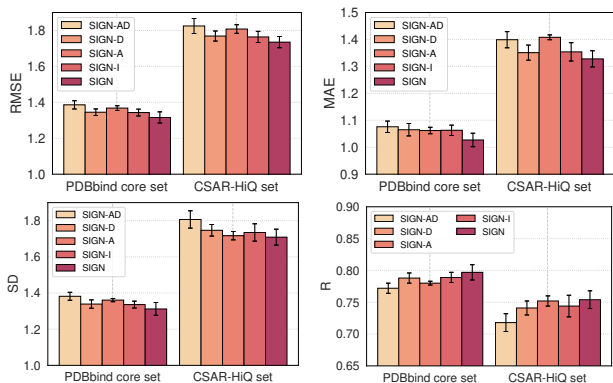
Figure 6: Contribution of spatial and interactive factors.

of complexes and have limited generalization capability on the additional CSAR-HiQ dataset. From the perspective of spatial structural modeling, we find that SGCN and GNN-DTI which incorporate position and distance information exhibit considerable improvement over the vanilla GCN and GAT. Since SGCN takes atomic position coordinates as input directly, it will be easily affected by the rotation and translation of atoms, and the 3D CNN model Pafnucy suffers from a similar issue. Thus, the prediction results are not ideal. Despite leveraging a transformer-like attention mechanism to handle the spatial structure, MAT is not better than RF-Score and OnionNet, suggesting the importance of combining spatial and interactive information. The edge-oriented model CMPNN outperforms the above methods because it enhances DMPNN with communication while propagating the distance information, which shows the significance of node-edge message passing process. Although DimeNet can learn from angle information and perform slightly better, the performance is still not ideal due to its limited ability of modeling larger biomolecules. Our proposed SIGN can not only capture more comprehensive angle-enhanced structural information instead of just distance, but also handle interactions in the complex through multi-task learning framework. Therefore, SIGN is much effective for modeling the protein-ligand complex and can accurately predict the binding affinity.

5.2.2 Generalizability Comparison (RQ2). There is increasing 3D structure-based protein-ligand data with binding affinity, whereas the amount of high-quality data in *refined set* is relatively small. Thus, the ability of utilizing more lower-quality data to improve performance shows the generalizability of model, which is another necessary measurement of performance evaluation. As introduced in Section 5.1.2, we conduct the extra experiment of generalizability on the *general set* of PDBbind dataset. As illustrated in Figure 5, we compare the proposed SIGN with major competitive baselines on two training sets. The results show that SIGN gets the lowest prediction error remarkably under both training settings. More importantly, our model improves the performance by around 8% when trained on the *general set* and it further expands the prediction advantage compared to baselines. Therefore, SIGN is proved to be more generalizable to more data in large quantity but poor quality.

5.2.3 Impact of Spatial and Interactive Factors (RQ3). To verify the effectiveness of factors that influence the final performance, we compare SIGN with its variants on the two benchmarks.
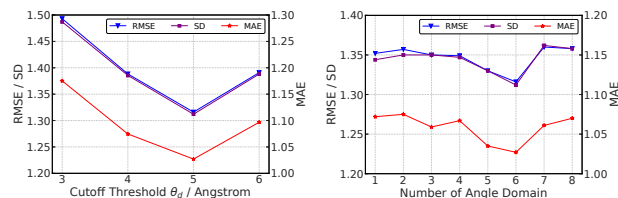

Figure 7: Parameter analysis on PDBbind *core set*.

- **SIGN-AD** uses the vanilla GAT layers for node-edge interaction without either angle or distance information.
- **SIGN-D** uses the vanilla GAT layer without distance information.
- **SIGN-A** uses the vanilla GAT layer without angle information.
- **SIGN-I** removes the interaction loss $\mathcal{L}_z$.

Figure 6 presents the comparison results on all metrics. As we can see, the proposed SIGN outperforms other variants, proving the necessity of handling the spatial and interactive information synergistically which is essential for protein-ligand binding affinity prediction. Specifically, SIGN-D and SIGN-A perform worse than SIGN since they can only capture the one-sided spatial structural information, i.e., distance or angle information in the complex. Furthermore, the prediction error of SIGN-AD is especially high among all variants. It indicates that modeling the complete spatial structure has a significant impact on performance improvement. The lack of long-range interactions in SIGN-I also leads to performance reduction, which confirms that only utilizing the spatial factors is insufficient and will lose the important interactive information.

5.2.4 Parameters Analysis (RQ4). As depicted in Figure 7, we investigate the performance variation for SIGN w.r.t several necessary hyper-parameters by varying each parameter while keeping others fixed as default settings. More results are in Appendix A.5.

Cutoff distance θ_d. We analyze the effect of cutoff distance for complex graph construction when varying θ_d from 3 to 6. With the increase of θ_d, more spatial information in the complex is available to our model and beneficial for learning complex representation better, which leads to dramatic performance improvements when $\theta_d \leq 5$ Å. After that, too long cutoff distance will introduce additional redundancies and degrade the performance.

Angle domain divisions N. To look deeper into the impact of angle information in our model, we divide the angle domains varying from 1 to 8. We can see that the model performs best when the number of angle domains is 5 or 6. Too fine-grained or coarse-grained divisions will result in performance degradation. One possible explanation is that too fine-grained divisions cannot provide distinguishable information in space while the angle domain at a too big granularity contains quite sparse atomic neighbors, both of which have an adverse effect on learning spatial information.

6 CONCLUSION

In this paper, we investigated how to improve the prediction of binding affinity between proteins and ligands. Specifically, we proposed a GNN-based model, SIGN, to learn the representations of protein-ligand complexes for better binding affinity prediction by leveraging the fine-grained structure and interaction information among atoms. Along this line, we designed the polar-inspired graph attention layers (PGAL) to integrate both distance and angle information for 3D spatial structure modeling. Also, to further improve

the prediction performance, we introduced a well-designed pooling process along with a reconstruction learning task for interaction matrix. Finally, the experimental results on two benchmarks showed the effectiveness and the generalizability of the proposed model.

REFERENCES

[1] William J Allen, Trent E Balius, Sudipto Mukherjee, Scott R Brozell, Demetri T Moustakas, P Therese Lang, David A Case, Irwin D Kuntz, and Robert C Rizzo. 2015. DOCK 6: Impact of new features and current docking performance. *Journal of computational chemistry* 36, 15 (2015), 1132–1156.

[2] Pedro J Ballester and John BO Mitchell. 2010. A machine learning approach to predicting protein–ligand binding affinity with applications to molecular docking. *Bioinformatics* 26, 9 (2010), 1169–1175.

[3] Maria Batool, Bilal Ahmad, and Sangdun Choi. 2019. A structure-based drug discovery paradigm. *International journal of molecular sciences* 20, 11 (2019), 2783.

[4] Ewen Callaway. 2020. 'It will change everything': DeepMind's AI makes gigantic leap in solving protein structures. *Nature* (2020).

[5] Lucy J Colwell. 2018. Statistical and machine learning approaches to predicting protein–ligand interactions. *Current opinion in structural biology* 49 (2018), 123–128.

[6] Tomasz Danel, Przemysław Spurek, Jacek Tabor, Marek Śmieja, Łukasz Struski, Agnieszka Słowik, and Łukasz Maziarka. 2020. Spatial graph convolutional networks. In *International Conference on Neural Information Processing*. Springer, 668–675.

[7] Kien Do, Truyen Tran, and Svetha Venkatesh. 2019. Graph transformation policy network for chemical reaction prediction. In *SIGKDD*. 750–760.

[8] James B Dunbar Jr, Richard D Smith, Chao-Yie Yang, Peter Man-Un Ung, Katrina W Lexa, Nickolay A Khazanov, Jeanne A Stuckey, Shaomeng Wang, and Heather A Carlson. 2011. CSAR benchmark exercise of 2010: selection of the protein–ligand complexes. *Journal of chemical information and modeling* 51, 9 (2011), 2036–2046.

[9] Holger Gohlke, Manfred Hendlich, and Gerhard Klebe. 2000. Knowledge-based scoring function to predict protein-ligand interactions. *Journal of molecular biology* 295, 2 (2000), 337–356.

[10] Zhongkai Hao, Chengqiang Lu, Zhenya Huang, Hao Wang, Zheyuan Hu, Qi Liu, Enhong Chen, and Cheekong Lee. 2020. ASGN: An active semi-supervised graph neural network for molecular property prediction. In *SIGKDD*. 731–752.

[11] Laurent Jacob and Jean-Philippe Vert. 2008. Protein-ligand interaction prediction: an improved chemogenomics approach. *Bioinformatics* 24, 19 (2008), 2149–2156.

[12] Ajay N Jain. 2003. Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. *Journal of medicinal chemistry* 46, 4 (2003), 499–511.

[13] Harren Jhoti and Andrew R Leach. 2007. *Structure-based drug discovery*. Vol. 1. Springer.

[14] Sarah L Kinnings, Nina Liu, Peter J Tonge, Richard M Jackson, Lei Xie, and Philip E Bourne. 2011. A machine learning-based method to improve docking scoring functions and its application to drug repurposing. *Journal of chemical information and modeling* 51, 2 (2011), 408–419.

[15] Thomas N. Kipf and Max Welling. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *ICLR*.

[16] Douglas B Kitchen, Hélène Decornez, John R Furr, and Jürgen Bajorath. 2004. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nature reviews Drug discovery* 3, 11 (2004), 935–949.

[17] Johannes Klicpera, Janek Groß, and Stephan Günnemann. 2020. Directional Message Passing for Molecular Graphs. In *ICLR*.

[18] Andrew R Leach, Brian K Shoichet, and Catherine E Peishoff. 2006. Prediction of protein- ligand interactions. Docking and scoring: successes and gaps. *Journal of medicinal chemistry* 49, 20 (2006), 5851–5855.

[19] DE Leckband, JN Israelachvili, FJ Schmitt, and W Knoll. 1992. Long-range attraction and molecular rearrangements in receptor-ligand interactions. *Science* 255, 5050 (1992), 1419–1421.

[20] Hongjian Li, Kwong-Sak Leung, Man-Hon Wong, and Pedro J Ballester. 2015. Low-quality structural and interaction data improves binding affinity prediction via random forest. *Molecules* 20, 6 (2015), 10947–10962.

[21] Qingliang Li and Salim Shah. 2017. Structure-Based Virtual Screening. *Methods in molecular biology (Clifton, N.J.)* 1558 (2017), 111–124.

[22] Shuangli Li, Jingbo Zhou, Tong Xu, Hao Liu, Xinjiang Lu, and Hui Xiong. 2020. Competitive Analysis for Points of Interest. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 1265–1274.

[23] Jaechang Lim, Seongok Ryu, Kyubyong Park, Yo Joong Choe, Jiyeon Ham, and Woo Youn Kim. 2019. Predicting drug–target interaction using a novel graph neural network with 3D structure-embedded graph representation. *Journal of chemical information and modeling* 59, 9 (2019), 3981–3988.

[24] Hao Liu, Jindong Han, Yanjie Fu, Jingbo Zhou, Xinjiang Lu, and Hui Xiong. 2021. Multi-Modal Transportation Recommendation with Unified Route Representation Learning. *Proceedings of the VLDB Endowment* 14, 3 (2021), 342–350.

[25] Yi Liu, Hao Yuan, Lei Cai, and Shuiwang Ji. 2020. Deep learning of high-order interactions for protein interface prediction. In *SIGKDD*. 679–687.

[26] Łukasz Maziarka, Tomasz Danel, Sławomir Mucha, Krzysztof Rataj, Jacek Tabor, and Stanisław Jastrzębski. 2020. Molecule Attention Transformer. *arXiv preprint arXiv:2002.08264* (2020).

[27] Xuan-Yu Meng, Hong-Xing Zhang, Mihaly Mezei, and Meng Cui. 2011. Molecular docking: a powerful approach for structure-based drug discovery. *Current computer-aided drug design* 7, 2 (2011), 146–157.

[28] N Moitessier, P Englebienne, D Lee, J Lawandi, and CR Corbeil. 2008. Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. *British journal of pharmacology* 153, S1 (2008), S7–S26.

[29] Ingo Muegge and Yvonne C Martin. 1999. A general and fast scoring function for protein- ligand interactions: a simplified potential approach. *Journal of medicinal chemistry* 42, 5 (1999), 791–804.

[30] Thin Nguyen, Hang Le, Thomas P Quinn, Tri Nguyen, Thuc Duy Le, and Svetha Venkatesh. 2020. GraphDTA: Predicting drug–target binding affinity with graph neural networks. *Bioinformatics* (10 2020). btaa921.

[31] Hakime Öztürk, Arzucan Özgür, and Elif Ozkirimli. 2018. DeepDTA: deep drug–target binding affinity prediction. *Bioinformatics* 34, 17 (2018), i821–i829.

[32] Matthew Ragoza, Joshua Hochuli, Elisa Idrobo, Jocelyn Sunseri, and David Ryan Koes. 2017. Protein–ligand scoring with convolutional neural networks. *Journal of chemical information and modeling* 57, 4 (2017), 942–957.

[33] Ying Song, Shuangjia Zheng, Zhangming Niu, Zhang-Hua Fu, Yutong Lu, and Yuedong Yang. 2020. Communicative representation learning on attributed molecular graphs. In *IJCAI*. 2831–2838.

[34] Sergio Filipe Sousa, Pedro Alexandrino Fernandes, and Maria Joao Ramos. 2006. Protein–ligand docking: current status and future challenges. *Proteins: Structure, Function, and Bioinformatics* 65, 1 (2006), 15–26.

[35] Marta M Stepniewska-Dziubinska, Piotr Zielenkiewicz, and Pawel Siedlecki. 2018. Development and evaluation of a deep learning model for protein–ligand binding affinity prediction. *Bioinformatics* 34, 21 (2018), 3666–3674.

[36] Minyi Su, Qifan Yang, Yu Du, Guoqin Feng, Zhihai Liu, Yan Li, and Renxiao Wang. 2018. Comparative assessment of scoring functions: the CASF-2016 update. *Journal of chemical information and modeling* 59, 2 (2018), 895–913.

[37] Mengying Sun, Sendong Zhao, Coryandar Gilvary, Olivier Elemento, Jiayu Zhou, and Fei Wang. 2020. Graph convolutional networks for computational drug development and discovery. *Briefings in bioinformatics* 21, 3 (2020), 919–935.

[38] Oleg Trott and Arthur J Olson. 2010. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry* 31, 2 (2010), 455–461.

[39] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2018. Graph Attention Networks. In *ICLR*.

[40] Izhar Wallach, Michael Dzamba, and Abraham Heifets. 2015. AtomNet: a deep convolutional neural network for bioactivity prediction in structure-based drug discovery. *arXiv preprint arXiv:1510.02855* (2015).

[41] Renxiao Yang, Xueliang Fang, Yipin Lu, Chao-Yie Yang, and Shaomeng Wang. 2005. The PDBbind database: methodologies and updates. *Journal of medicinal chemistry* 48, 12 (2005), 4111–4119.

[42] Renxiao Wang, Luhua Lai, and Shaomeng Wang. 2002. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *Journal of computer-aided molecular design* 16, 1 (2002), 11–26.

[43] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. 2019. How Powerful are Graph Neural Networks?. In *ICLR*.

[44] Kevin Yang, Kyle Swanson, Wengong Jin, Connor Coley, Philipp Eiden, Hua Gao, Angel Guzman-Perez, Timothy Hopper, Brian Kelley, Miriam Mathea, et al. 2019. Analyzing learned molecular representations for property prediction. *Journal of chemical information and modeling* 59, 8 (2019), 3370–3388.

[45] Chengxi Zang and Fei Wang. 2020. MoFlow: an invertible flow model for generating molecular graphs. In *SIGKDD*. 617–626.

[46] Liangzhen Zheng, Jingrong Fan, and Yuguang Mu. 2019. OnionNet: a Multiple-Layer Intermolecular-Contact-Based Convolutional Neural Network for Protein–Ligand Binding Affinity Prediction. *ACS omega* 4, 14 (2019), 15956–15965.

[47] Zhi Zheng, Chao Wang, Tong Xu, Dazhong Shen, Penggang Qin, Baoxing Huai, Tongzhu Liu, and Enhong Chen. 2021. Drug Package Recommendation via Interaction-aware Graph Induction. *arXiv preprint arXiv:2102.03577* (2021).

[48] Jingbo Zhou, Shuangli Li, Liang Huang, Haoyi Xiong, Fan Wang, Tong Xu, Hui Xiong, and Dejing Dou. 2020. Distance-aware Molecule Graph Attention Network for Drug-Target Binding Affinity Prediction. *arXiv preprint arXiv:2012.09624* (2020).

A APPENDIX

In the appendix, we first introduce the construction process of complex interaction graph. Then the details of experimental settings

and baseline descriptions are given. Finally, we show the additional results of parameter analysis on another dataset. The pseudocode of SIGN training procedure is described in Algorithm 2. The implement our model based on PaddlePaddle[3]. We train all models on 24 Intel CPUs and a Tesla V100 GPU with 32 GB memory. The code is available at: https://github.com/PaddlePaddle/PaddleHelix/tree/dev/apps/drug_target_interaction/sign.

A.1 Complex Interaction Graph Construction

As shown in Figure 2(a), there is no available intermolecular connection information between the ligand and its protein in the dataset. What's more, the local covalent bonds in the original molecular graph cannot provide adequate 3D structure information. To include non-local correlations, here we aim to construct the spatial-based complex interaction graph. Since the size of a protein is much larger than that of a ligand, it's unnecessary to build the complete protein structure in the complex graph, which might be noisy and time-consuming. Therefore, we apply a sampling-based process to construct the graph, which preserves the key structure of complex. The detail of the construction process is provided in algorithm 1. We first initialize the atom node set $\mathcal{V}$ with ligand's atom set $\mathcal{V}^L$. Then the protein's atoms which are close to the ligand from $\mathcal{V}^P$ are selected to add into the atom node set $\mathcal{V}$. Finally, we update the complex edge set $\mathcal{E}$ by adding into the edges of atom pairs whose distances are smaller than the cutoff threshold r_θ.

Algorithm 1: Complex Interaction Graph Construction.

Input : The position matrix M^P and node set $\mathcal{V}^P$ of protein
The position matrix M^L and node set $\mathcal{V}^L$ of ligand
The cutoff distance r_θ
Output: The complex interaction graph $\mathcal{G}_I = <\mathcal{V}, \mathcal{E}>$

1 Initialize $\mathcal{V} \leftarrow \mathcal{V}^L, \mathcal{E} \leftarrow \{\}$;
2 **for** *atom node pair* $(a_i, a_j) \in \mathcal{V}^L \times \mathcal{V}^P$ **do**
3   Calculate distance $d_{ij} \leftarrow |M^L(a_i) - M^P(a_j)|$;
4   **if** $d_{ij} \le r_\theta$ **then**
5   Update node set $\mathcal{V} \leftarrow \mathcal{V} \cup \{a_j\}$;
6   **end**
7 **end**
8 Combined position matrix $M \leftarrow CONCAT(M^L, M^P)$;
9 **for** *atom node pair* $(a_i, a_j) \in \mathcal{V} \times \mathcal{V}$ **do**
10   Calculate distance $d_{ij} \leftarrow |M(a_i) - M(a_j)|$;
11   **if** $d_{ij} \le r_\theta$ **then**
12   Update edge set $\mathcal{E} \leftarrow \mathcal{E} \cup \{e_{ij} = (a_i, a_j)\}$;
13   **end**
14 **end**
15 **return** $\mathcal{V}, \mathcal{E}$

A.2 Instruction of the Binding Affinity

In biology experiments, the binding affinity between protein and ligand can be determined as the value K_d (dissociation constant), K_i (inhibition constant), or IC_{50} (half inhibition concentration). In practice, the experimental binding affinity (i.e., the ground truth for the binding affinity prediction task) is expressed with the negative

---

Algorithm 2: Training Procedure for SIGN.

Input : Training set $\mathcal{D}$
Output: Trained model parameters θ

1 Randomly initialize the parameter θ of SIGN;
2 **for** *iteration = 1, 2, ...* **do**
3   **for** *each batch from training samples* **do**
4   Calculate spatial relation embeddings $\boldsymbol{d}$ using Eq. (1);
5   **for** *l = 1...L* **do**
6   Obtain edge embeddings $\boldsymbol{e}^{(l)}$ using Eq. (4)-(10);
7   Obtain node embeddings $\boldsymbol{a}^{(l)}$ using Eq. (11)-(15);
8   **end**
9   Build the interaction matrix Z using Eq. (16)-(17);
10   Estimate the interaction matrix $\tilde{Z}$ using Eq. (18)-(19);
11   Calculate the interaction loss $\mathcal{L}_z$ using Eq. (20);
12   Calculate the prediction $\hat{y}$ using Eq. (21);
13   Calculate the prediction loss $\mathcal{L}_a$ using Eq. (22);
14   Update parameters θ according to the gradient of $\mathcal{L}$;
15   **end**
16 **end**
17 **return** θ

---

logarithm pk_a of the determined value (e.g, $-logK_d$, $-logK_i$, or $-logIC_{50}$) on PDBbind and CSAR-HiQ datasets.

A.3 Experiment Details

A.3.1 Evaluation Metrics. We first detail the four metrics used in our experiment. Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Pearson correlation coefficient (R) are defined as:

$$RMSE = \sqrt{\frac{1}{|\mathcal{D}|}\sum_{i=1}^{|\mathcal{D}|}(\hat{y}_i - y_i)^2}, \; MAE = \frac{1}{|\mathcal{D}|}\sum_{i=1}^{|\mathcal{D}|}|\hat{y}_i - y_i| \quad (24)$$

$$R = \frac{\sum_{i=1}^{|\mathcal{D}|}(\hat{y}_i - \bar{\hat{y}})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{|\mathcal{D}|}(\hat{y}_i - \bar{\hat{y}})^2(y_i - \bar{y})^2}} \quad (25)$$

$\hat{y}_i$ and y_i respectively represent the predicted and experimental binding affinity of the i-th complex in dataset $\mathcal{D}$. As introduced in [35], the standard deviation (SD) is defined as follows:

$$SD = \sqrt{\frac{1}{|\mathcal{D}| - 1}\sum_{i=1}^{|\mathcal{D}|}[y_i - (a + b\hat{y}_i)]^2} \quad (26)$$

where a and b are the intercept and the slope of the regression line, respectively.

A.3.2 Input Graph and Features. For all GNN-based methods, we use the same input complex graph as introduced in Appendix A.1 for protein-ligand binding affinity prediction. For GraphDTA models, we input the protein sequence as well as the ligand molecular graph or our constructed complex graph. In this paper, we report the best result when using the complex graph as input. For 3D-CNN and GNN models, the atom features used according to [35] include atomic types, hybridization, the number of bonds with other heavy-atoms and hetero-atoms, atom properties such as aromaticity, and the partial charge. In total, 18 features are used to describe an
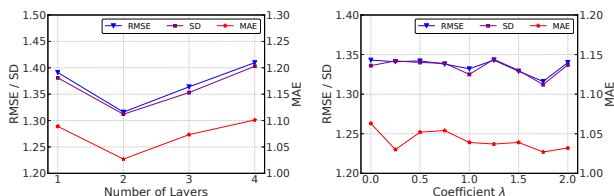
Figure 8: Parameter analysis on PDBbind *core set*.

atom. Considering the heterogeneity in the complex graph, we further extend atom features to a 36-dimension vector with zero-padding, where the 1st to 18th elements represent the features of ligand atoms and the 19th to 36th elements represent the features of protein atoms. As to edge features for edge-based GNN models, we combine the two atom features and encoded distance features between atoms as the input vector. In brief, we provide the same input complex graph and atomic features for our SIGN and all GNN-based baselines to make fair comparisons. For ML-based methods and OnionNet, they cannot take the complex graph or grid-like data as input and only receive the specific molecule-level features. We extract the input feature vectors based on the distance as described in their original papers [2, 46]. Note that these extracted features also reflect the structural information of complex in a global view.

A.3.3 Parameter Settings. For the proposed SIGN, we use Adam optimizer for model training with a learning rate of 0.001 and set the batch size as 32. The balancing hyper-parameter λ is set to 1.75 according to the performance on validation set. We construct the complex graph and interaction matrix with cutoff-threshold $\theta_d = 5$ and $\theta_\rho = 12$ as suggested in [29], respectively. The basic dimensions of node and edge embeddings are both set to 128. The number of buckets for spatial relation b is set to 4 with the splitting granularity of 1Å. For PGAL layers, we set the number of attention heads C to 4, the dropout rate to 0.2, and the number of angle domains N to 6. For PiPool layer, there are 36 pooling blocks in total, where the two atomic type sets S_P and S_L are defined as stated in [2].

For baseline models, we tune the parameters based on the default settings to get optimal performance. Specifically, the number of decision trees in RF-score is set to 100, the max-depth of trees is set to 5, the maximum number of features is set to 3 and the minimum number of samples required to split is set to 10. For the CNN-based models, we set the channels of three-layer 3D convolutions for Pafnucy as 64,128 and 256. For OnionNet, the number of input features is 3840 and there are 32, 64, and 128 filters in the three convolutional layers with the kernel size as 4. The maximum length of protein sequences is set to 1000 in GraphDTA. For GNN-based models, the number of filters in SGCN is set to 32 with the dimension as 36. We also apply the data augmentation process to ensure optimal performance. For fair comparison, the embedding dimension of other baselines is set to 128 (same as SIGN). For GNN-DTI, the initial μ and δ for distance learning in GAT layers are set to 4.0 and 1.0, respectively. For DimeNet, the number of spherical harmonics and radial basis functions are set to 4 and 3, respectively. We use two-layer interaction blocks and three-layer bilinear layers to make DimeNet work in our experiment. For DMPNN and CMPNN, the layer of edge-oriented message passing layers is set to 3 and we use MLP as the communication module in CMPNN. The weighting

coefficients for self-attention, distance, and adjacency matrices in MAT are set to 0.3, 0.3, and 0.4, respectively.

A.4 Baseline Descriptions

We compare our SIGN model with the following methods to predict the protein-ligand binding affinity:

- **ML-based methods** include linear regression (LR), support vector regression (SVR), and random forest (RF). These methods take the inter-molecular interaction features introduced in RF-Score [2] as input and predict the protein-ligand binding affinity.
- **Pafnucy** [35] is a representative 3D CNN-based model which can learn the spatial structure of protein-ligand complexes.
- **OnionNet** [46] generates two-dimensional interaction features based on rotation-free element-pair contacts in complexes and adopts CNN to learn representations for prediction.
- **GraphDTA** [30] introduces GNN models to learn the complex graph and uses CNN to learn the protein sequence. It has four variants with different GNN models: **GCN** [15], **GAT** [39], **GIN** [43] and **GAT-GCN** which combines the former two models.
- **SGCN** [6] leverages node positions based on graph convolutional network, which directly utilizes atomic coordinates.
- **GNN-DTI** [23] is a distance-aware graph attention network with considering 3D structural information to learn the intermolecular interactions for protein-ligand complexes.
- **DMPNN** [44] is an edge-based message passing neural network. It can incorporate the spatial information between atoms by applying the aggregation process for edges.
- **MAT** [26] employs a molecule-augmented attention mechanism based on transformer for graph representation learning with using the inter-atomic distances.
- **DimeNet** [17] is a recent state-of-the-art model for small molecular graph learning using directional message passing scheme. Bessel functions are employed to encode the angle and distance information in graph neural network.
- **CMPNN** [33] further develops DMPNN to build a communicative message passing scheme between nodes and edges for better molecular representation learning.

A.5 Additional Parameters Analysis

Number of PGAL layers L. As shown in Figure 8, we first present the influence of multi-hop propagation with stacking node-edge interaction layers from 1 to 4. We observe that increasing the number of layers would not always give rise to a better result. The model with one PGAL layer has limited ability to model high-order information in the complex. As a result of over-fitting, the performance of the model using more than 3 layers starts to degenerate gradually. Therefore, applying two interaction layers in SIGN is enough to capture sufficient spatial information.

Balancing coefficient λ. Moreover, we change the coefficient λ to control the trade-off between the prediction loss and interaction loss. From the results, we observe that the performance first tends to get better with incorporating more interactive information for long-range dependencies, and then begins to drop off slightly. In general, our model is stable with varying coefficients and always achieves better performance than all baseline methods.