
Physics-aware Graph Neural Network for Accurate RNA 3D Structure Prediction

Shuo Zhang^{1,2}, Yang Liu², Lei Xie^{1,2,3}

¹Ph.D. Program in Computer Science, Graduate Center, City University of New York

²Department of Computer Science, Hunter College, City University of New York

³Helen & Robert Appel Alzheimer’s Disease Research Institute,

Feil Family Brain & Mind Research Institute,

Weill Cornell Medicine, Cornell University

szhang4@gradcenter.cuny.edu, thomasliuy@gmail.com, lxie@iscb.org

Abstract

Biological functions of RNAs are determined by their three-dimensional (3D) structures. Thus, given the limited number of experimentally determined RNA structures, the prediction of RNA structures will facilitate elucidating RNA functions and RNA-targeted drug discovery, but remains a challenging task. In this work, we propose a Graph Neural Network (GNN)-based scoring function trained only with the atomic types and coordinates on limited solved RNA 3D structures for distinguishing accurate structural models. The proposed Physics-aware Multiplex Graph Neural Network (PaxNet) separately models the local and non-local interactions inspired by molecular mechanics. Furthermore, PaxNet contains an attention-based fusion module that learns the individual contribution of each interaction type for the final prediction. We rigorously evaluate the performance of PaxNet on two benchmarks and compare it with several state-of-the-art baselines. The results show that PaxNet significantly outperforms all the baselines overall, and demonstrate the potential of PaxNet for improving the 3D structure modeling of RNA and other macromolecules.

1 Introduction

RNAs play an essential role in many cellular processes in organisms, such as mediating gene translation, regulating gene expression, catalyzing biochemical reactions, and transferring cellular information [1–3]. These cellular functions are determined by their unique three-dimensional (3D) structures. However, due to the high cost of experimental methods, the high-resolution RNA 3D structures deposited in the Protein Data Bank (PDB) are still very scarce [4]. Thus, there is an urgent need to develop computational RNA 3D structure prediction methods for elucidating RNA functions and promoting RNA-targeted drug discovery [5].

At the current state, most existing computational methods for the prediction of RNA 3D structures are non-deep learning methods, and are based on physics or knowledge-guided approaches like coarse-grained representations [6–8], fragment-assembling [9–11], statistical potentials [12–15], etc. Only a limited number of deep learning approaches have been proposed to be scoring functions for evaluating RNA 3D structural models [16, 17]. Specifically, ARES [17] is a Graph Neural Network (GNN) that outperforms the previous state-of-the-art methods using only a small number of RNA structures for training without any assumptions about structural characteristics. This has demonstrated the power of deep learning on tackling the challenge of RNA 3D structure prediction.

In this work, we aim at developing a new deep learning-based approach to predict RNA 3D structures. Inspired by the physical principles in molecular mechanics methods [18] that separately considers

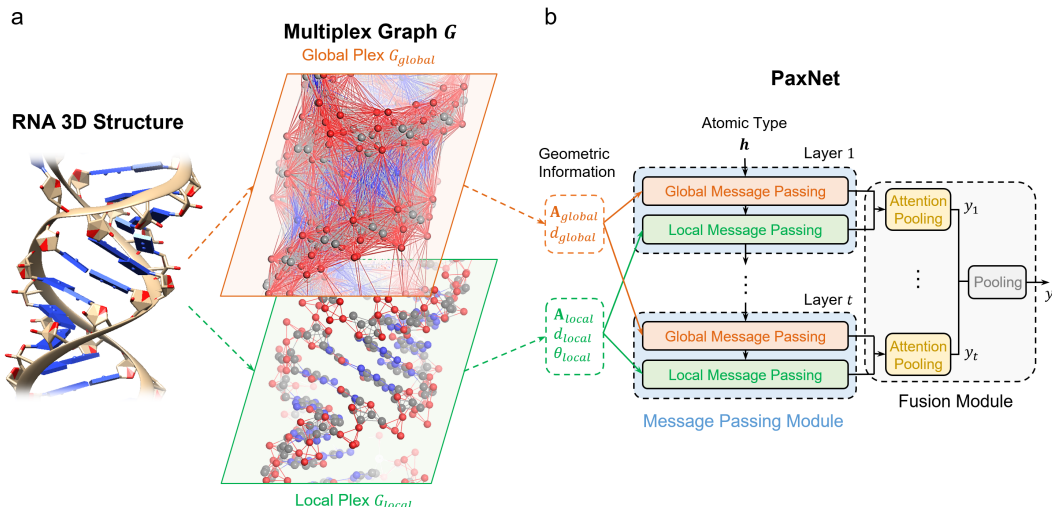


Figure 1: **Illustration of PaxNet pipeline.** (a) Construction of multiplex graph $G = \{G_{global}, G_{local}\}$ based on a given RNA 3D structure. (b) Overall architecture of PaxNet, which consists of message passing modules and a fusion module.

the local and non-local interactions when computing molecular energy, we propose Physics-aware Multiplex Graph Neural Network (PaxNet), which is a GNN based on message passing scheme [19]. PaxNet represents each RNA 3D structure using a two-plex (or layer) multiplex graph, where one plex only contains local interactions, and the other plex further contains non-local interactions. PaxNet takes the multiplex graphs as input and uses different message passing modules to incorporate the geometric information in different kinds of interactions. Furthermore, with a fusion module, the contribution of each kind of interaction can be learned and fused to compute the final prediction. Similar to ARES, PaxNet only uses atomic types and coordinates as input without having any assumptions about other structural and physiochemical characteristics.

To verify the effectiveness of PaxNet, we train and evaluate PaxNet on the datasets used in [17] for the RNA 3D structure prediction. Our PaxNet outperforms previous state-of-the-art methods on all benchmarks in general. The results suggest that PaxNet provides a powerful method for the representation learning of RNA 3D structures. Our main contributions are as follows:

- We propose a novel GNN, Physics-aware Multiplex Graph Neural Network (PaxNet) for the prediction of RNA 3D structures. Inspired by molecular mechanics, PaxNet models the local and non-local interactions differently by building multiplex graphs to represent RNA structures and designing diverse message passing modules.
- Comprehensive experiments for RNA 3D structure prediction are conducted to demonstrate that our proposed PaxNet outperforms the state-of-the-art methods.

2 Method

The overview of Physics-aware Multiplex Graph Neural Network (PaxNet) is shown in Figure 1. Given a RNA 3D structure, a corresponding multiplex graph G (Figure 1a) is built, where each plex of G contains a different group of atomic interactions. Then PaxNet (Figure 1b) uses different message passing modules to update the atomic type embeddings h with the geometric information in different plex of G accordingly. To learn the contributions from each plex, a fusion module based on attention mechanism is used for the final prediction. Further details of PaxNet, which are not covered in this section, can be found in Appendix A.2.

2.1 Multiplex Graph Representation of RNA 3D Structure

Each RNA 3D structure consists of atoms that are associated with 3D coordinates and atomic numbers. Based on solely this information, we can define different pairwise interactions between atoms, i.e. according to different cutoff values of the pairwise distances. Then a multiplex graph

can be constructed to represent the RNA: For each kind of predefined interaction, we use a plex to contain all atoms as nodes and the interactions as edges. The resulting plexes that share the same group of atoms naturally form a multiplex graph.

In particular, we are inspired by molecular mechanics [18], in which the molecular energy E is modeled with a separate consideration of local and non-local interactions: $E = E_{\text{local}} + E_{\text{non-local}}$. Motivated by this, we also decouple the modeling of these two kinds of interactions in PaxNet. For local interactions, we can define them by using either chemical bonds or finding the neighbors of each node within a relatively small cutoff distance depending on the given task. For global interactions, we define them by finding the neighbors of each node within a relatively large cutoff distance. As a result, we construct a two-plex multiplex molecular graph $G = \{G_{\text{global}}, G_{\text{local}}\}$ as shown in Figure 1a.

2.2 Physics-aware Message Passing Modules

To update the atomic (or node) embeddings, two message passing modules are used in PaxNet, which are *Global Message Passing* and *Local Message Passing* for G_{global} and G_{local} in G , respectively. The design of these message passing modules is inspired by physical principles. In molecular mechanics when modeling molecular energy $E = E_{\text{local}} + E_{\text{non-local}}$, the term for local interactions $E_{\text{local}} = E_{\text{bond}} + E_{\text{angle}} + E_{\text{dihedral}}$ includes E_{bond} that depends on bond lengths, E_{angle} on bond angles, and E_{dihedral} on dihedral angles. The term for non-local interactions $E_{\text{non-local}} = E_{\text{electro}} + E_{\text{vdW}}$ includes electrostatic and van der Waals interactions which depend on interatomic distances. So that for the geometric information in molecular mechanics, the local interactions need pairwise distances and angles, while the non-local interactions only need pairwise distances. The message passing modules in PaxNet also use geometric information in this way when modeling these interactions.

As shown in Figure 1b, the *Local Message Passing* which updates the node embeddings based on local interactions will incorporate the related adjacency matrix $\mathbf{A}_{\text{local}}$, pairwise distances d_{local} and angles θ_{local} . The *Global Message Passing* which updates the node embeddings based on global interactions (both non-local and local interactions) will contain only the related adjacency matrix $\mathbf{A}_{\text{global}}$ and pairwise distances d_{global} . The node embeddings learned by each message passing module are passed to the next layer or to the fusion module for final prediction.

2.3 Fusion Module

To combine the node embeddings from message passing modules in every hidden layer for final prediction, we design a fusion module as shown in Figure 1b. The fusion is a two-step pooling process. In the first step, we use an attention mechanism [20] for each hidden layer t of PaxNet to get the corresponding prediction y_t using the node embeddings computed by the two message passing modules in hidden layer t . In each attention operation, attention weights for each group of node embeddings are learned and used to compute the corresponding prediction y_t using a weighted summation. In the second step, the predictions from all hidden layers are averaged together to compute the final predicted result y .

3 Experiments

Task. In our experiments, we focus on the prediction of 3D RNA structures. Following the previous works [14, 21, 17], we refer it to the task of identifying accurate structural models of RNA from less accurate ones: Given a group of candidate 3D structural models generated based on an RNA sequence, a desired scoring function can distinguish accurate structural models among all candidates. For PaxNet, only a limited number of RNA structures are used for learning, and no assumptions about structural characteristics are incorporated to perform the identification. In practice, PaxNet predicts the root mean square deviation (RMSD) from the unknown true structure for each structural model. A lower RMSD would suggest a more accurate structural model predicted by PaxNet.

Datasets. To access the performance of PaxNet on predicting 3D RNA structures, we use the same datasets as those used in [17], which include a dataset for training and two benchmarks for evaluation. The **training dataset** contains 18 relatively older and smaller RNA molecules determined experimentally [22]. Each RNA is used to generate 1000 structural models via the Rosetta FARFAR2 sampling method [21]. The two benchmarks for evaluation contain relatively newer and larger RNAs. In detail, **benchmark 1** consists of the first 21 RNAs in the RNA-Puzzles structure prediction

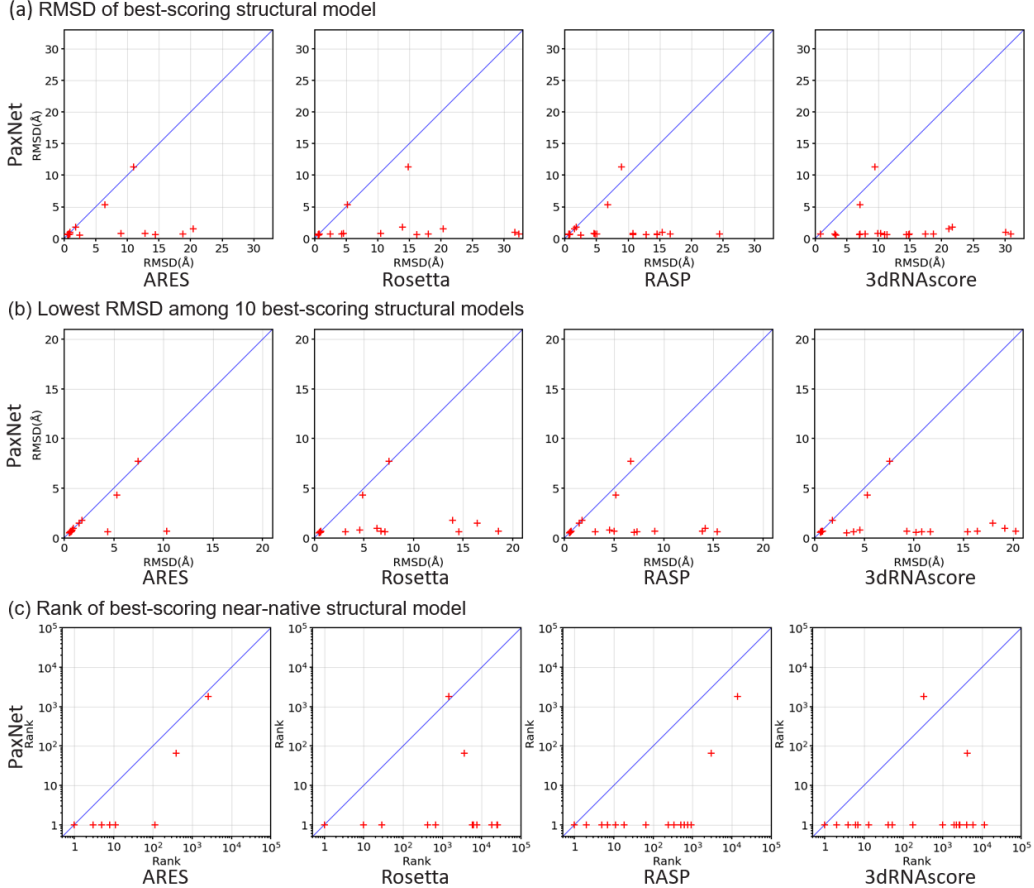


Figure 2: **Performance comparison on benchmark 1.** Given a group of candidate structural models for each RNA in benchmark 1, we rank the models using PaxNet and the other four leading scoring functions for comparison. Each cross in the figures corresponds to one RNA. (a) The best-scoring structural model of each RNA predicted by the scoring functions is compared. PaxNet in general identifies more accurate models (with lower RMSDs from the native structure) than those decided by the other scoring functions. (b) Comparison of the 10 best-scoring structural models. The identifications of PaxNet contain accurate models more frequently than those from other scoring functions. (c) The rank of the best-scoring near-native structural model for each RNA is used for comparison. PaxNet usually performs better than the other scoring functions by having a lower rank.

challenge [23]. Each RNA is used to generate at least 1500 structural models using FARFAR2, where 1% of the models are near native (i.e., within a 2\AA RMSD of the experimentally determined native structure). To simulate a difficult modeling scenario, **benchmark 2** contains no near-native models, and includes 16 RNAs that are all substantially different from any of those used in the aforementioned datasets¹. The source of the datasets can be found in [17]. The statistics of the datasets are listed in Appendix A.3.1.

Experimental Settings. We use the same data splits as those used in [17] to be the training set and validation set. For the training process, PaxNet is optimized to minimize the difference between the output value and the ground-truth RMSD of each structural model from the corresponding structure on the training set. An early-stopping strategy is adopted to decide the best epoch based on the validation loss. For the benchmark 1, we compare PaxNet with four state-of-the-art baselines: **ARES** [17], **Rosetta (2020 version)** [21], **RASP** [13], and **3dRNAscore** [14]. For benchmark 2, besides the aforementioned baselines, we additionally include **SimRNA** [6], **Rosetta (2010 version)**, and **Rosetta (2007 version)**. More details of the implementations can be found in Appendix A.3.2.

¹We use the updated version of benchmark 2 provided by the authors of [17], which leads to corrected and different results as those shown in their original paper.

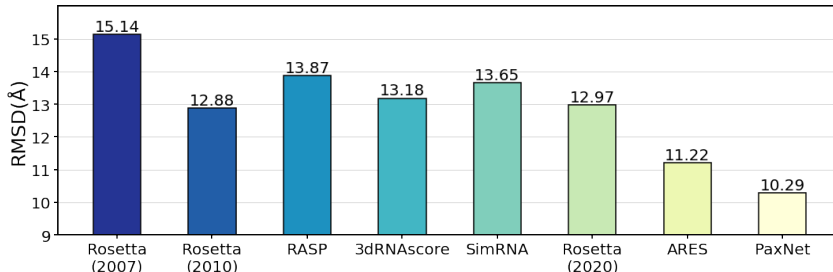


Figure 3: Performance comparison on benchmark 2.

Table 1: Comparison of space and time complexity.

Model	# of Parameters	Memory (GB)	Inference Time (s)
ARES	18073	13.47	2.14
PaxNet	12530	7.83	0.57

4 Results

Benchmark 1. On benchmark 1, PaxNet significantly outperforms all other four scoring functions as shown in Figure 2. When comparing the best-scoring structural model of each RNA (Figure 2a), the probability of the model to be near-native ($<2\text{\AA}$ RMSD from the native structure) is 90% when using PaxNet, compared with 62, 43, 33, and 5% for ARES, Rosetta, RASP, and 3dRNAscore, respectively. As for the 10 best-scoring structural models of each RNA (Figure 2b), the probability of the models to include at least one near-native model is 90% when using PaxNet, compared with 81, 48, 48, and 33% for ARES, Rosetta, RASP, and 3dRNAscore, respectively. When comparing the rank of the best-scoring near-native structural model of each RNA (Figure 2c), the geometric mean of the ranks across all RNAs is 1.7 for PaxNet, compared with 3.6, 73.0, 26.4, and 127.7 for ARES, Rosetta, RASP, and 3dRNAscore, respectively. The lower mean rank of PaxNet indicates that less effort is needed to go down the ranked list of PaxNet to include one near-native structural model. A more detailed analysis of the near-native ranking task can be found in Appendix A.4.1. We also conduct an ablation study to demonstrate the effectiveness of the global and local message passing as well as fusion modules in PaxNet. The details of the results are shown in Appendix A.4.3.

Benchmark 2. For each RNA in benchmark 2, we identify the best-scoring structural model scored by each scoring function. For each scoring function, we show the median across RNAs. As shown in Figure 3, PaxNet performs better than all scoring functions by having the lowest median RMSD across RNAs. We find a similar trend when considering the 10 best-scoring structural models for each RNA, which is shown in Appendix A.4.2.

Efficiency Evaluation. When implementing computational methods for RNAs, a common bottleneck comes from the relatively high complexity caused by the number of involved atoms. Thus we evaluate the efficiency of PaxNet by comparing it with the strongest baseline, ARES. When solving the same given task, we find PaxNet is much more efficient than ARES regarding both space and time complexity as shown in Table 1. Details of the settings can be found in Appendix A.3.2.

5 Conclusion

In this work, we focus on the task for RNA 3D structure prediction and propose a GNN-based method, PaxNet, to tackle the task from a deep learning perspective. Different from the existing deep learning-based methods in this field, PaxNet is inspired by the physical principles in molecular mechanics to decouple the modeling of local and non-local interactions in RNA 3D structures. With only atomic types and 3D coordinates and without any assumptions about structural features, PaxNet can be trained on limited RNA 3D structures to make more accurate predictions than the previous state-of-the-art models. The limitation of PaxNet is that we still need an extra sampling method to generate candidate structural models. It would be interesting to extend PaxNet to perform *de novo* RNA 3D structure prediction without a separate sampling process. PaxNet can be also possibly applied to other tasks related to macromolecules due to its general framework.

References

- [1] Run-Wen Yao, Yang Wang, and Ling-Ling Chen. Cellular functions of long noncoding rnas. *Nature cell biology*, 21(5):542–551, 2019.
- [2] Elizabeth A Dethoff, Jeetender Chugh, Anthony M Mustoe, and Hashim M Al-Hashimi. Functional complexity and regulation through rna dynamics. *Nature*, 482(7385):322–330, 2012.
- [3] Katherine Deigan Warner, Christine E Hajdin, and Kevin M Weeks. Principles for targeting rna with drug-like small molecules. *Nature reviews Drug discovery*, 17(8):547–558, 2018.
- [4] Stephen K Burley, Charmi Bhikadiya, Chunxiao Bi, Sebastian Bittrich, Li Chen, Gregg V Crichlow, Cole H Christie, Kenneth Dalenberg, Luigi Di Costanzo, Jose M Duarte, et al. Rcsb protein data bank: powerful new tools for exploring 3d structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic acids research*, 49(D1):D437–D451, 2021.
- [5] Jayesh A Kulkarni, Dominik Witzigmann, Sarah B Thomson, Sam Chen, Blair R Leavitt, Pieter R Cullis, and Roy van der Meel. The current landscape of nucleic acid therapeutics. *Nature Nanotechnology*, 16(6):630–643, 2021.
- [6] Michal J Boniecki, Grzegorz Lach, Wayne K Dawson, Konrad Tomala, Pawel Lukasz, Tomasz Soltysinski, Kristian M Rother, and Janusz M Bujnicki. Simrna: a coarse-grained method for rna folding simulations and 3d structure prediction. *Nucleic acids research*, 44(7):e63–e63, 2016.
- [7] Magdalena A Jonikas, Randall J Radmer, Alain Laederach, Rhiju Das, Samuel Pearlman, Daniel Herschlag, and Russ B Altman. Coarse-grained modeling of large rna molecules with knowledge-based potentials and structural filters. *Rna*, 15(2):189–199, 2009.
- [8] Dong Zhang and Shi-Jie Chen. Isrna: An iterative simulated reference state approach to modeling correlated interactions in rna folding. *Journal of chemical theory and computation*, 14(4):2230–2239, 2018.
- [9] Song Cao and Shi-Jie Chen. Physics-based de novo prediction of rna 3d structures. *The journal of physical chemistry B*, 115(14):4216–4226, 2011.
- [10] Mariusz Popena, Marta Szachniuk, Maciej Antczak, Katarzyna J Purzycka, Piotr Lukasiak, Natalia Bartol, Jacek Blazewicz, and Ryszard W Adamiak. Automated 3d structure composition for large rnas. *Nucleic acids research*, 40(14):e112–e112, 2012.
- [11] Yunjie Zhao, Yangyu Huang, Zhou Gong, Yanjie Wang, Jianfen Man, and Yi Xiao. Automated and fast building of three-dimensional rna structures. *Scientific reports*, 2(1):1–6, 2012.
- [12] Julie Bernauer, Xuhui Huang, Adelene YL Sim, and Michael Levitt. Fully differentiable coarse-grained and all-atom knowledge-based potentials for rna structure evaluation. *Rna*, 17(6):1066–1075, 2011.
- [13] Emidio Capriotti, Tomas Norambuena, Marc A Marti-Renom, and Francisco Melo. All-atom knowledge-based potential for rna structure prediction and assessment. *Bioinformatics*, 27(8):1086–1093, 2011.
- [14] Jian Wang, Yunjie Zhao, Chunyan Zhu, and Yi Xiao. 3drnascore: a distance and torsion angle dependent evaluation function of 3d rna structures. *Nucleic acids research*, 43(10):e63–e63, 2015.
- [15] Tongchuan Zhang, Guodong Hu, Yuedong Yang, Jihua Wang, and Yaoqi Zhou. All-atom knowledge-based potential for rna structure discrimination based on the distance-scaled finite ideal-gas reference state. *Journal of Computational Biology*, 27(6):856–867, 2020.
- [16] Jun Li, Wei Zhu, Jun Wang, Wenfei Li, Sheng Gong, Jian Zhang, and Wei Wang. Rna3dcnn: Local and global quality assessments of rna 3d structures using 3d deep convolutional neural networks. *PLoS computational biology*, 14(11):e1006514, 2018.

- [17] Raphael JL Townshend, Stephan Eismann, Andrew M Watkins, Ramya Rangan, Maria Karelina, Rhiju Das, and Ron O Dror. Geometric deep learning of rna structure. *Science*, 373(6558):1047–1051, 2021.
- [18] Tamar Schlick. *Molecular modeling and simulation: an interdisciplinary guide*, volume 21. Springer Science & Business Media, 2010.
- [19] Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural message passing for quantum chemistry. In *International Conference on Machine Learning*, pages 1263–1272. PMLR, 2017.
- [20] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *International Conference on Learning Representations*, 2018.
- [21] Andrew Martin Watkins, Ramya Rangan, and Rhiju Das. Farfar2: improved de novo rosetta prediction of complex global rna folds. *Structure*, 28(8):963–976, 2020.
- [22] Rhiju Das and David Baker. Automated de novo prediction of native-like rna tertiary structures. *Proceedings of the National Academy of Sciences*, 104(37):14664–14669, 2007.
- [23] Zhichao Miao, Ryszard W Adamiak, Maciej Antczak, Michał J Boniecki, Janusz Bujnicki, Shi-Jie Chen, Clarence Yu Cheng, Yi Cheng, Fang-Chieh Chou, Rhiju Das, et al. Rna-puzzles round iv: 3d structure predictions of four ribozymes and two aptamers. *RNA*, 26(8):982–995, 2020.
- [24] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and S Yu Philip. A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 2020.

A Appendix

A.1 Notations and Definitions

Notations. Let $G = (V, E)$ be a graph with $N = |V|$ nodes and $M = |E|$ edges. The nearest neighbors of node i are defined as $\mathcal{N}(i) = \{j | d(i, j) = 1\}$, where $d(i, j)$ is the shortest distance between node i and j . In later formulations, we will use \mathbf{h}_i as the embedding of node i , \mathbf{e}_{ji} as the edge embedding between node i and j , \mathbf{m}_{ji} as the message being sent from node j to node i in the message passing scheme, F as the hidden dimension in our model, MLP as multi-layer perceptron, \parallel as concatenation operation, \odot as element wise production and \mathbf{W} as weight matrix.

Here we give the definition of a multiplex molecular graph, which is the format of our RNA representation, as follows:

Definition 1. Multiplex Molecular Graph. We denote a molecular structure as an $(L + 1)$ -tuple $G = (V, E^1, \dots, E^L)$ where V is the set of nodes (atoms) and for each $l \in \{1, 2, \dots, L\}$, E^l is the set of edges (molecular interactions) in type l that between pairs of nodes (atoms) in V . By defining the graph $G^l = (V, E^l)$ which is also called a plex or a layer, the multiplex molecular graph can be seen as the set of graphs $G = \{G^1, G^2, \dots, G^L\}$.

Next we introduce the message passing scheme [19] which is a framework widely used in spatial-based GNNs [24] and is the basis of our PaxNet:

Definition 2. Message Passing Scheme. Given a graph G , the node feature of each node i is \mathbf{x}_i , and the edge feature for each node pair j and i is \mathbf{e}_{ji} . The message passing scheme iteratively updates message \mathbf{m}_{ji} and node embedding \mathbf{h}_i for each node i using the following functions:

$$\mathbf{m}_{ji}^t = f_m(\mathbf{h}_i^{t-1}, \mathbf{h}_j^{t-1}, \mathbf{e}_{ji}), \quad (1)$$

$$\mathbf{h}_i^t = f_u(\mathbf{h}_i^{t-1}, \sum_{j \in \mathcal{N}(i)} \mathbf{m}_{ji}^t), \quad (2)$$

where superscript t denotes the t -step iteration, f_m and f_u are learnable functions. For each node i , \mathbf{x}_i is the input node embedding \mathbf{h}_i^0 .

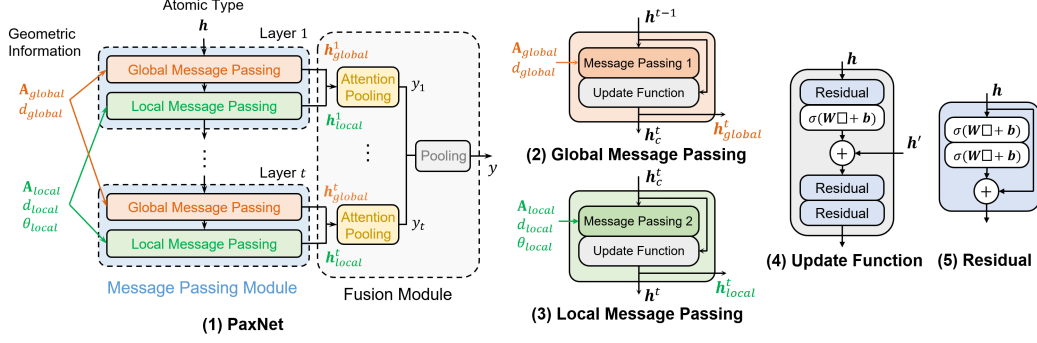


Figure A.4: **Illustration of the architecture of PaxNet.** (1) Overall architecture of PaxNet. (2) Architecture of Global Message Passing. (3) Architecture of Local Message Passing. (4) Update function in message passing. (5) Residual block.

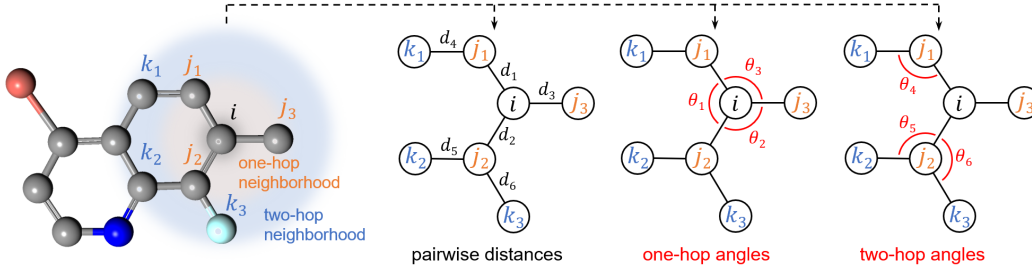


Figure A.5: **An example of the geometric information in G .** By defining the one-hop neighbors $\{j\}$ and two-hop neighbors $\{k\}$ of node i , we can define the pairwise distances d and the related angles θ including one-hop angles and two-hop angles.

A.2 Architecture of PaxNet

The architecture of PaxNet is shown in Figure A.4. The detailed description of the modules in PaxNet will be covered in this section.

Global Message Passing. In this module, we update the node embeddings in global plex G_{global} by capturing the pairwise distances d_{global} based on the message passing in Definition 2. Each message passing operation is:

$$\mathbf{m}_{ji}^{t-1} = \text{MLP}_m([\mathbf{h}_j^{t-1} \parallel \mathbf{h}_i^{t-1} \parallel \mathbf{e}_{ji}]), \quad (3)$$

$$\mathbf{h}_i^t = \mathbf{h}_i^{t-1} + \sum_{j \in \mathcal{N}(i)} \mathbf{m}_{ji}^{t-1} \odot \phi_d(\mathbf{e}_{ji}), \quad (4)$$

where $i, j \in G_{global}$ are connected nodes that define a message embedding, ϕ_d is a learnable function for pairwise distance. The edge embedding \mathbf{e}_{ji} encodes the corresponding pairwise distance information.

After the message passing, an update function containing multiple residual blocks is used to get the node embeddings for the next layer as well as \mathbf{h}_{global} to be fed into the fusion module. Illustrations of these operations are shown in Figure A.4.

Local Message Passing. For the updates of node embeddings in the local plex G_{local} , we incorporate both pairwise distances d_{local} and angles θ_{local} . When updating the embedding of node i , we consider the one-hop neighbors $\{j\}$ and the two-hop neighbors $\{k\}$ of i . Specifically for the angles related to those nodes, we show an example in Figure A.5 and cluster them depending on the edges around i : (a) The *one-hop angles* are angles between the one-hop edges (θ_1, θ_2 and θ_3 in Figure A.5). (b) The *two-hop angles* are angles between the one-hop edges and two-hop edges (θ_4, θ_5 and θ_6 in Figure A.5).

Table A.2: **Dataset statistics.**

Dataset	# of RNAs	Avg. # of Non-hydrogen Atoms	# of Structural Models
Training Set	18	865	18000
Benchmark 1	21	2172	428195
Benchmark 2	16	1681	76442

To perform message passing, we use the same way as Equation (3) to compute the message embeddings \mathbf{m} . The message passing operation in the t -th iteration is:

$$\begin{aligned} \mathbf{m}_{ji}'^{t-1} &= \mathbf{m}_{ji}^{t-1} + \sum_{j' \in \mathcal{N}(i) \setminus \{j\}} \mathbf{m}_{ji'}^{t-1} \odot \phi_d(\mathbf{e}_{j'i}) \odot \phi_\theta(\theta_{j'i,ji}) \\ &\quad + \sum_{k \in \mathcal{N}(j) \setminus \{i\}} \mathbf{m}_{kj}^{t-1} \odot \phi_d(\mathbf{e}_{kj}) \odot \phi_\theta(\theta_{kj,ji}), \end{aligned} \quad (5)$$

$$\mathbf{h}_i^t = \mathbf{h}_i^{t-1} + \sum_{j \in \mathcal{N}(i)} \mathbf{m}_{ji}'^{t-1} \odot \phi_d(\mathbf{e}_{ji}), \quad (6)$$

where $i, j, k \in G_{local}$, ϕ_d is a learnable function for pairwise distance, ϕ_α is a learnable function for angles. \mathbf{e}_{ji} encodes the corresponding pairwise distance information. $\theta_{kj,ji}$ encodes the angle $\theta_{kj,ji} = \angle kji$ accordingly. In Equation (5), we use two summation terms to separately encode the one-hop and two-hop angles with the associated pairwise distances to update \mathbf{m}_{ji} .

As shown in Figure A.4, we design the remaining functions in Local Message Passing similarly to those in Global Message Passing to get \mathbf{h}_{local} to be fused and the input for the next iteration.

Communication between Message Passings. To address the cross-layer relations between G_{global} and G_{local} , we let the information in those plexes communicate with each other as depicted in Figure A.4. We first perform Global Message Passing on G_{global} in each iteration. Then the updated node embeddings are transferred to G_{local} for Local Message Passing. Finally, the further updated node embeddings are passed back to G_{global} for the next iteration.

Fusion Module. As shown in Figure A.4, to fuse the node embeddings for a final prediction, we design Fusion Module with a two-step pooling:

In the first step, we use attention mechanism for each hidden layer t of PaxNet to get the corresponding node-level prediction y_t using the node embeddings \mathbf{h}_{global}^t and \mathbf{h}_{local}^t from hidden layer t . We first compute the attention weight $\alpha_{m,i}$ that measures the contribution of $\mathbf{h}_{m,i}$, which belongs to node i on plex m in G :

$$\alpha_{m,i}^t = \frac{\exp(\text{LeakyReLU}(\mathbf{W}_m^t \mathbf{h}_{m,i}^t))}{\sum_m \exp(\text{LeakyReLU}(\mathbf{W}_m^t \mathbf{h}_{m,i}^t))}, \quad (7)$$

where $m \in \{global, local\}$, and $\mathbf{W}_m^t \in \mathbb{R}^{1 \times F}$ is a learnable weight matrix different for each hidden layer t and plex m . With $\alpha_{m,i}^t$, we then compute the node-level prediction $y_{i,t}$ of node i in hidden layer t using weighted summation:

$$y_{i,t} = \sum_m \alpha_{m,i}^t (\mathbf{W}_{out_m}^t \mathbf{h}_{m,i}^t), \quad (8)$$

where $\mathbf{W}_{out_m}^t \in \mathbb{R}^{1 \times F}$ is a learnable weight matrix different for each hidden layer t and plex m .

In the second step, the node-level predictions $y_{i,t}$ are used to compute the final prediction y :

$$y = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T y_{i,t}. \quad (9)$$

where N is the total number of nodes, T is the number of hidden layers in PaxNet.

A.3 Experimental Settings

A.3.1 Dataset Statistics

The statistics of the datasets used in this paper are summarized in Table A.2.

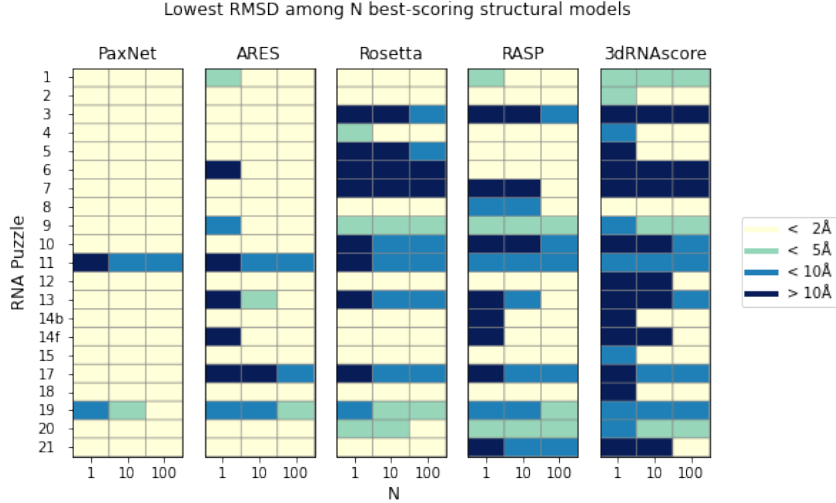


Figure A.6: **Detailed analysis of near-native ranking task on benchmark 1.** The results of the lowest RMSD among N best-scoring structural models of each RNA predicted by each scoring function are compared.

A.3.2 Implementation Details

When building multiplex graphs for RNA structures, we use cutoff distance $d_l = 2.6\text{\AA}$ for the local interactions in G_{local} and $d_g = 20\text{\AA}$ for the global interactions in G_{global} . Following [17], we only use the carbon, nitrogen, and oxygen atoms in RNA structures. For the input features, we use atomic numbers Z as node features and pairwise distances between atoms as edge features computed from atomic positions. We represent Z with randomly initialized, trainable embeddings and expand pairwise distances and angles with basis functions to reduce correlations. For the basis functions e_{RBF} and a_{SBF} , we use $N_{SHBF}=7$, $N_{SRBF}=6$ and $N_{RBF}=16$. In our message passing operations, we define $\phi_d(e) = W_e e$ and $\phi_\alpha(\alpha) = \text{MLP}_\alpha(\alpha)$, where W_e is a weight matrix, MLP_α is a multi-layer perceptrons (MLP). For the MLPs used in our model, they all have 2 layers to take advantage of the approximation capability of MLP. For all activation functions, we use the self-gated Swish activation function. PaxNet is optimized by minimizing the smooth L1 loss between the predicted value and the ground truth using the Adam optimizer. For the best-performed PaxNet, we use a batch size=8, learning rate=1e-4, number of hidden layer=1, number of hidden dimension=16. PyTorch is used to implement PaxNet. All of the experiments are done on an NVIDIA Tesla V100 GPU (32 GB).

For efficiency evaluation, we use PaxNet and ARES to predict for the structural models of RNA in puzzle 5 of RNA-Puzzles challenge. The RNA being predicted has 6034 non-hydrogen atoms. The model settings of PaxNet and ARES are the same as those used for reproducing the experiments on the benchmarks. We use batch size=8 when performing the predictions.

A.4 Additional Results

A.4.1 Detailed Analysis of Near-native Ranking Task on Benchmark 1

For each RNA in benchmark 1, we rank the structural models using PaxNet and four baseline scoring functions. For each scoring function, we select the $N \in \{1, 10, 100\}$ best-scoring structural models for each RNA. For each RNA, scoring function, and N, we show the lowest RMSD across structural models in Figure A.6). The RMSD results are quantized to determine if each RMSD is below 2\AA , between 2\AA and 5\AA , between 5\AA and 10\AA , or above 10\AA . From the results, we find that for each RMSD threshold (2\AA , 5\AA , or 10\AA) and for each N, the number of RNAs with at least one selected model that has RMSD below the threshold is greater when using PaxNet than when using any of the other four baseline scoring functions.

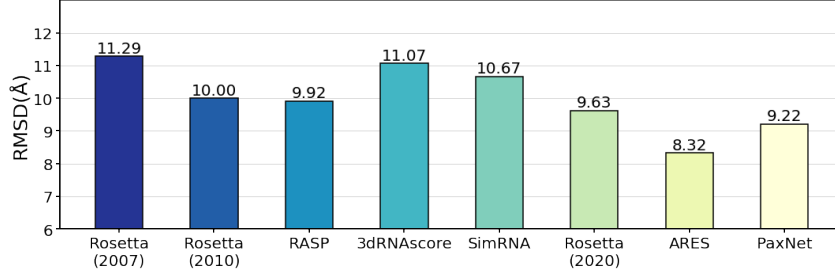


Figure A.7: **Additional Results on Benchmark 2.** For each RNA and for each of the eight scoring functions, we identify the minimum RMSD across the 10 best-scoring structural models. For each scoring function, we show the median across RNAs.

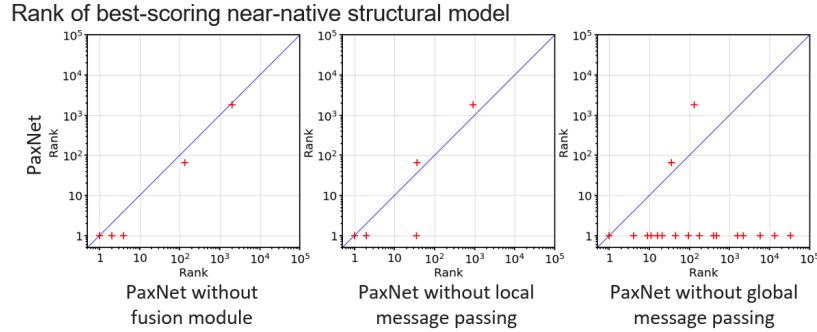


Figure A.8: **Results of ablation study on benchmark 1.** The rank of the best-scoring near-native structural model for each RNA is used for comparison.

A.4.2 Additional Results on Benchmark 2

We score all structural models for each RNA and for each of the eight scoring functions, and then identify the minimum RMSD across the 10 best-scoring ones. For each scoring function, we use the median across RNAs for comparison. As shown in Figure A.7, PaxNet outperforms all scoring functions except ARES.

A.4.3 Ablation Study

To evaluate the effectiveness of the modules in our PaxNet, we conduct ablation study on benchmark 1 by designing three PaxNet variants: PaxNet without the fusion module (replace the fusion module with average pooling), PaxNet without the local message passing module, and PaxNet without the global message passing module. From the results shown in Figure A.8, we find the original PaxNet overall performs better than all ablated variants: The geometric mean of the ranks across all RNAs is 1.7 for PaxNet, compared with 2.1, 2.0, and 79.9 for PaxNet without fusion module, PaxNet without local message passing module, and PaxNet without global message passing module, respectively. In particular, PaxNet without global message passing module performs the worst among all ablated variants, which implies the importance of non-local interactions in RNA structures.