

3D Chromosome Structure Reconstruction Using Graph Convolutional Neural Networks

Van Hovenga¹, Jugal Kalita², and Oluwatosin Oluwadare^{2,*} *corresponding author

¹. Mathematics Department, University of Colorado, Colorado Springs, 80918

². Computer Science Department, University of Colorado, Colorado Springs, 80918

Abstract

Chromosome conformation capture (3C) is a method of measuring chromosome topology in terms of loci interaction¹. The Hi-C method is a derivative of 3C that allows for genome-wide quantification of chromosome interaction^{2,3}. From such interaction data, it is possible to infer the three-dimensional (3D) structure of the underlying chromosome. In this work, we developed a novel method, HiC-GNN⁴, for predicting the three-dimensional structures of chromosomes from Hi-C data (Figure 1). HiC-GNN is unique from other methods for chromosome structure prediction in that the models learned by HiC-GNN can be generalized to unseen data.

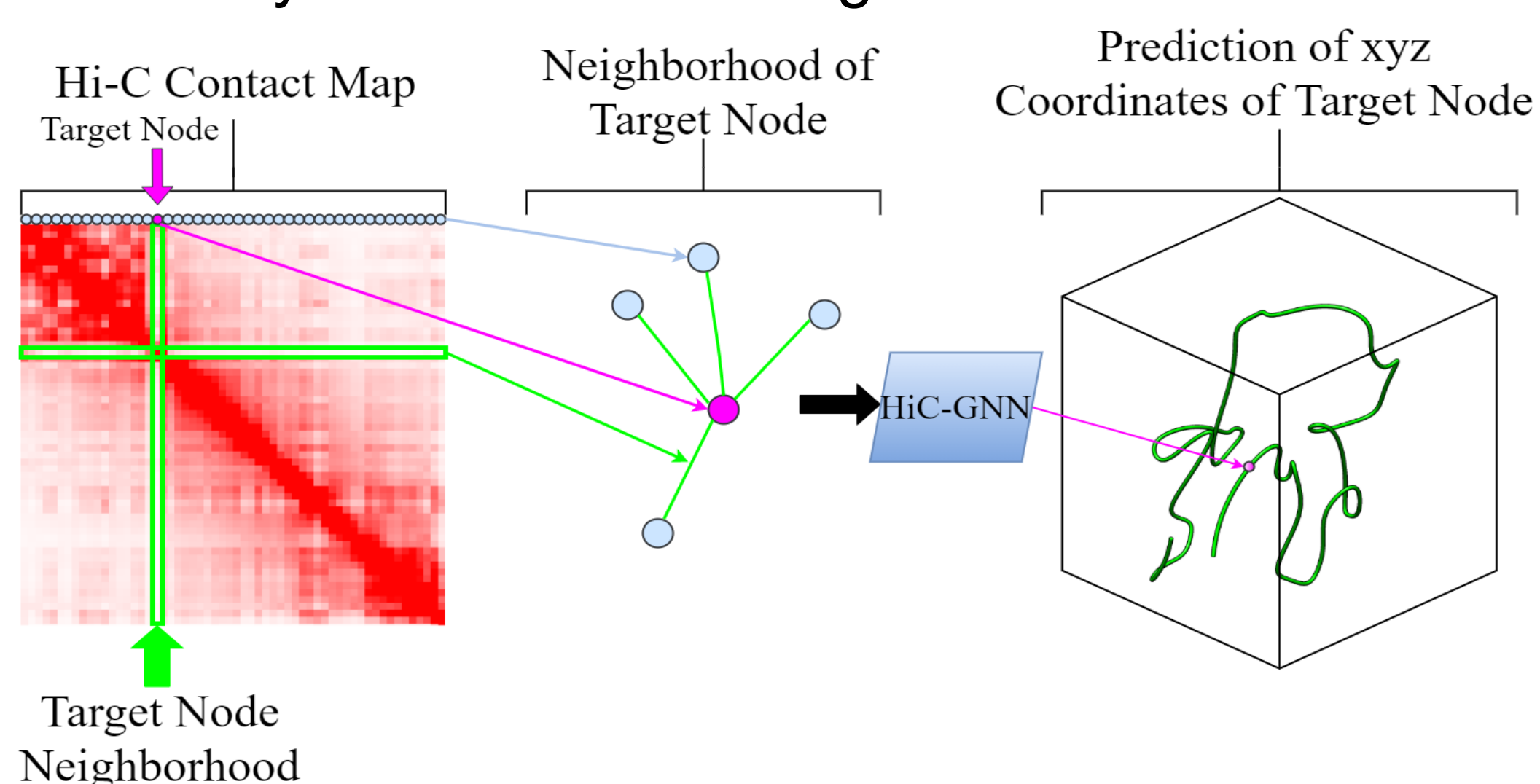


Figure 1. HiC-GNN 3D Chromosomal Structure Prediction Pipeline From Hi-C Data.

Contributions

- ✓ Generalization across resolutions: A model trained on the Hi-C map of a fixed chromosome at one resolution can be used to accurately predict the structure of the same fixed chromosome using a different Hi-C map resolution as the input. This allows us to train a model on low resolution data and make predictions for high-resolution data.
- ✓ The generalizability of our method could also reduce the computational requirements of generating structures on high resolution data via pre-training on low resolution data.

General Pipeline of HiC-GNN

From the raw input Hi-C contact map, we calculate “wish” or expected distances, we generate node embeddings, and we normalized the contact map. The node feature vectors and the normalized Hi-C map are then used as inputs to the graph convolutional neural network (GCNN). The GCNN is optimized by minimizing the MSE of the pairwise distances of the output structure to the wish distances. Here, N refers to the number of loci and E refers to the size of the embeddings. Figure 2 shows a graphical illustration of the pipeline.

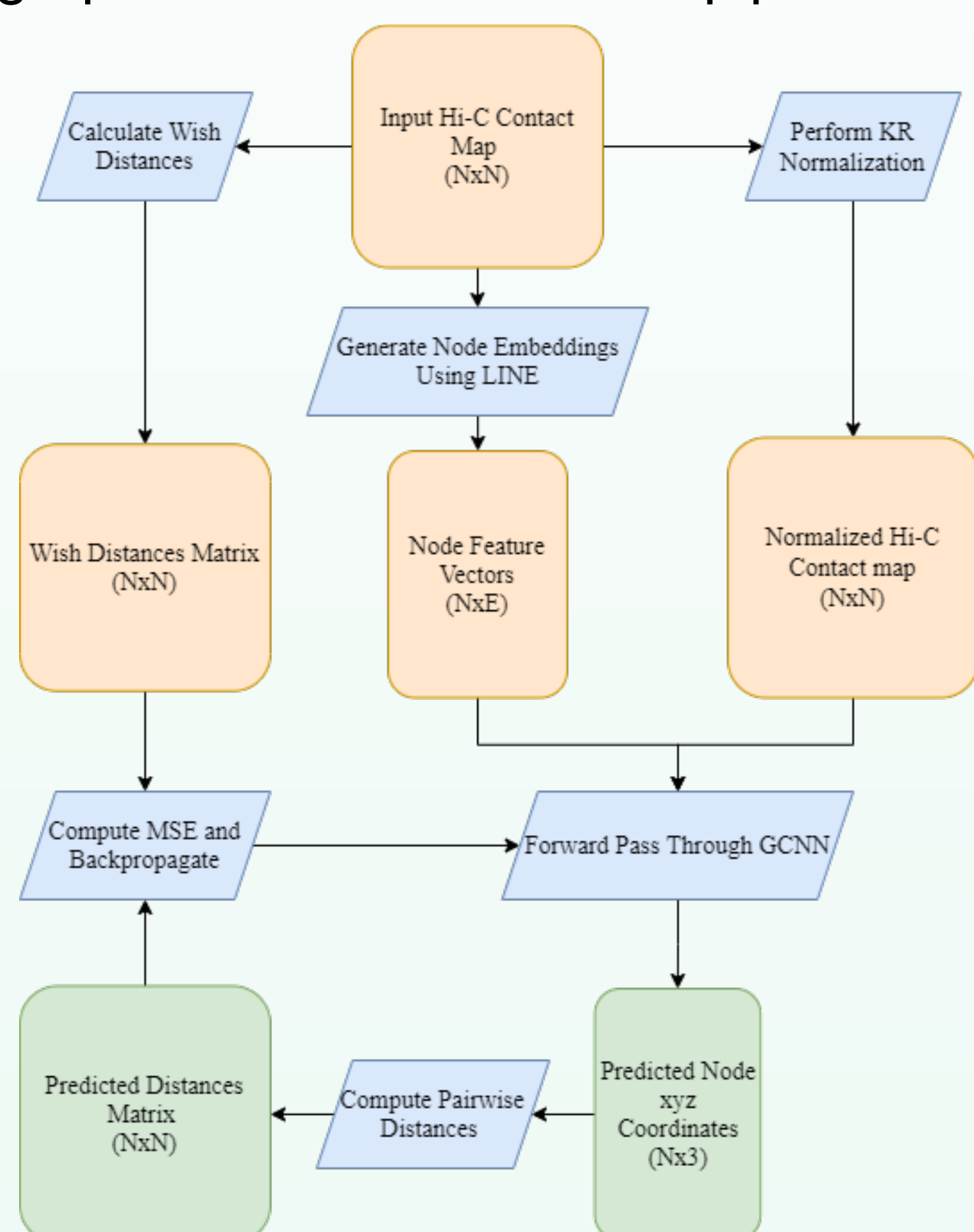


Figure 2. The pipeline for the entire HiC-GNN algorithm.

Results: HiC-GNN vs. State-of-the-Art algorithms

The Figure 3 shows a comparison of HiC-GNN with the other methods on the 1mb (A), 500kb (B), and 250kb (C) GM12878 cell line data using the distance Spearman Correlation Coefficient (dSCC) [on the Left] and distance root mean squared error (dRMSD) [on the Right]. The human GM12878 cell line Hi-C dataset from Rao et al.³ consists of the Hi-C maps of 23 chromosomes generated from the MboI restriction enzyme at 1mb, 500kb, and 250kb resolutions. HiC-GNN is either on-par or outperforms the other methods on the majority of the chromosomes. See some generated 3D structures in Figure 4.

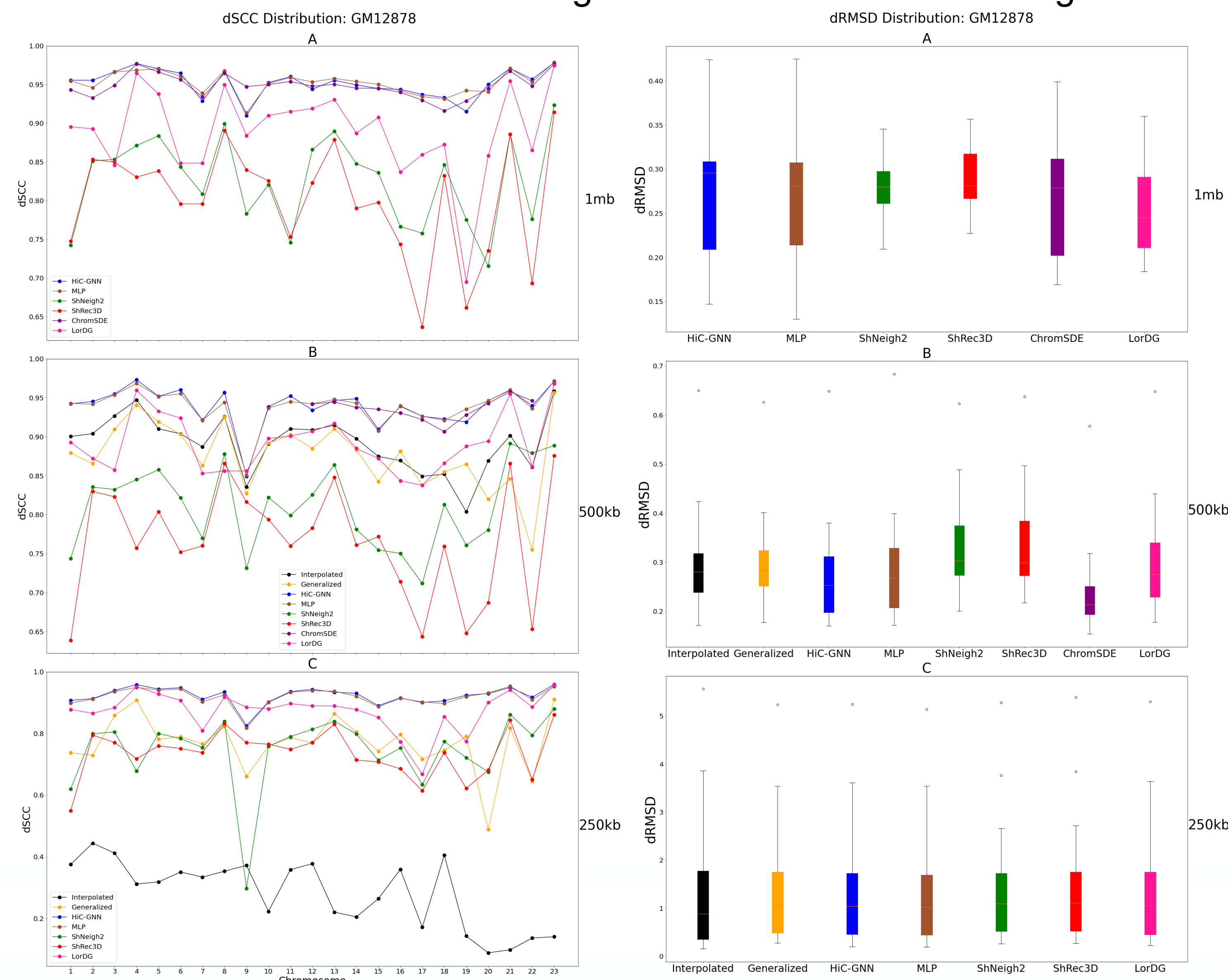


Figure 3. Comparison with existing state-of-the-art 3D chromosome structure reconstruction algorithms.

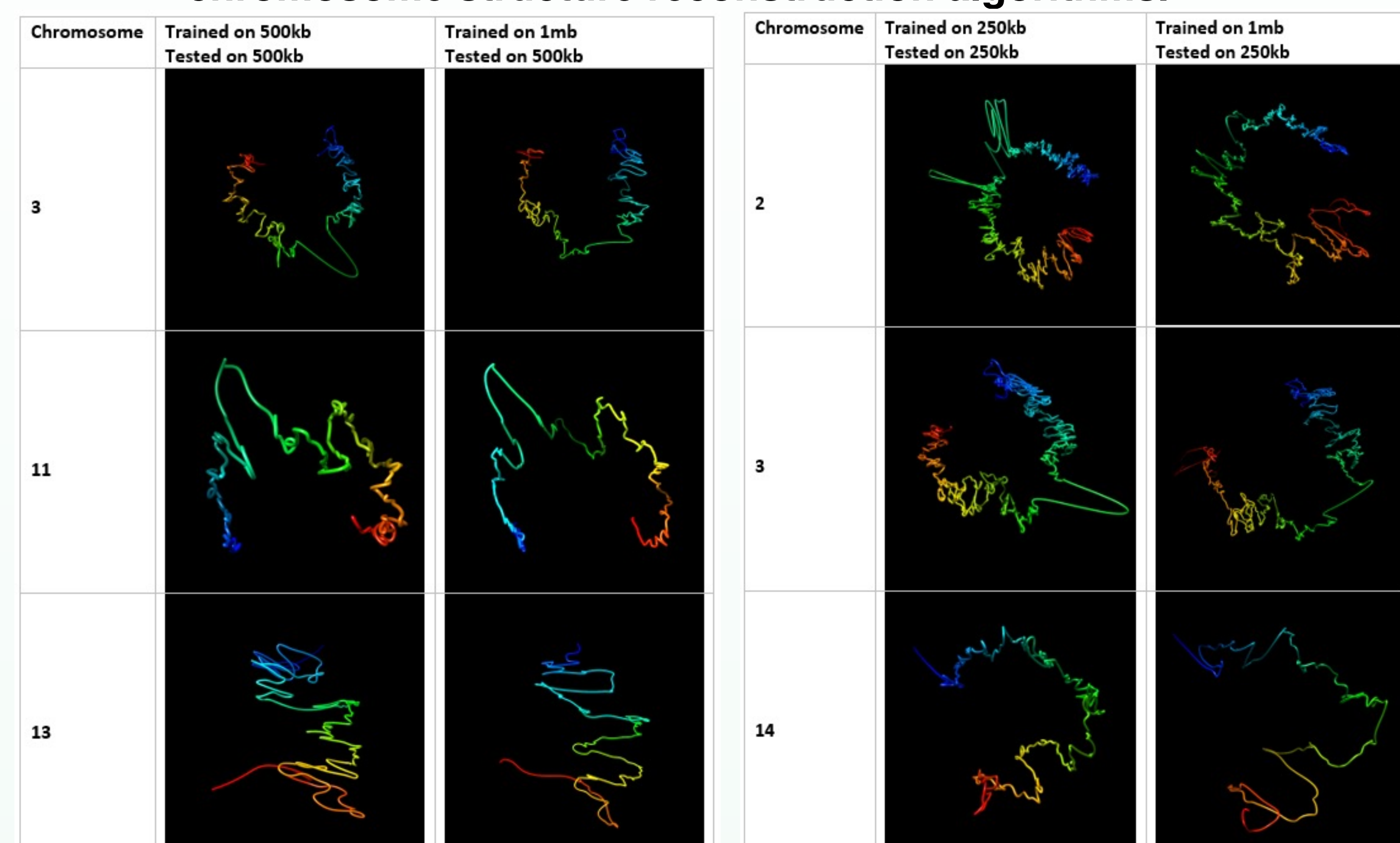


Figure 4. 3D chromosome structures generated by HiC-GNN.

ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation (NSF) CRII award (grant no: 2153205) and start-up funding from UCCS to Oluwatosin Oluwadare.

REFERENCES:

1. Dekker J, Marti-Renom MA, Mirny LA. 2013. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nat Rev Genet* 14: 390–403.
2. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragozcy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, et al. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326: 289–93.
3. Rao SS, Huntley MH, Dur NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, et al. 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159: 1665–80.
4. Hovenga, V., Oluwadare, O. and Kalita, J., 2021. HiC-GNN: A Generalizable Model for 3D Chromosome Reconstruction Using Graph Convolutional Neural Networks. *bioRxiv*.