

VoroCNN: Deep Convolutional Neural Network Built on 3D Voronoi Tessellation of Protein Structures

Ilia Igashov^{1,2} Kliment Olechnovič³ Maria Kadukova^{1,2} Česlovas Venclovas³ Sergei Grudinin¹

Introduction

Effective use of evolutionary information has recently led to tremendous progress in computational prediction of three-dimensional (3D) structures of proteins and their complexes. Despite the progress, the accuracy of predicted structures tends to vary considerably from case to case. Since the utility of computational models depends on their accuracy, reliable estimates of deviation between predicted and native structures are of utmost importance.

Method

For the first time, we present a deep convolutional neural network (CNN) constructed on a Voronoi tessellation of 3D molecular structures.

Protein Graph. Proteins fold into specific three-dimensional (3D) structures as a result of interatomic interactions. Protein atoms interact among themselves and with the solvent, and these interactions rapidly decay with the distance. A rigorous way to define interatomic interactions is to construct a Voronoi tessellation of the protein atoms and relate every Voronoi cell to an atom and every Voronoi cell face to an interatomic contact. However, if a pair of contacting atoms is located near the surface of a protein structure, the corresponding Voronoi face may extend far away from the atoms. This problem can be circumvented by constraining the Voronoi cells of the atoms inside the boundaries defined by the solvent-accessible surface, enabling calculation of the areas for every atom-atom and atom-solvent contact. Such a solution has been implemented in Voronota (Olechnovič & Venclovas, 2014), a software package specifically optimized to construct rapid tessellations for molecular structures when the radii of balls

¹Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France ²Moscow Institute of Physics and Technology, 141701 Dolgoprudniy, Russia ³Institute of Biotechnology Life Sciences Center Vilnius University, Sauletekio 7, Vilnius, LT 10257, Lithuania. Correspondence to: Česlovas Venclovas <ceslovas.venclovas@bti.vu.lt>, Sergei Grudinin <sergei.grudinin@univ-grenoble-alpes.fr>.

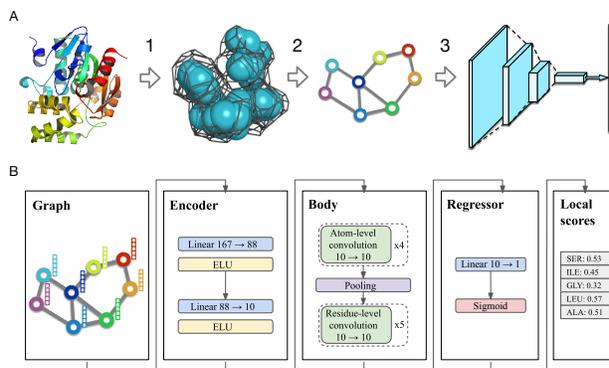


Figure 1. (A) Schematic representation of the VoroCNN quality assessment method. Firstly, a Voronoi tessellation of a 3D-model is computed with Voronota. Then, based on Voronoi 3D-tessellation, a graph is built. Finally, a graph neural network predicts local CAD-scores of all residues in the initial model. (B) Architecture of the VoroCNN network. The graph neural network consists of three blocks: Encoder, Body and Regressor. Encoder takes a graph as input and reduces node feature vectors with linear layers and ELUs. Body contains two sequences of convolutional layers and the pooling layer between them. Regressor consists of a linear layer followed by a sigmoid function and predicts local CAD-scores. A rounded colored rectangle represents a layer, dashed rounded rectangle means that it contains several identical layers, dimensions of input and output of each trainable layer are represented with the notation “x → y”. The last column illustrates the output of the network – local CAD-score predictions.

(atoms) are not very different from each other.

Graph Neural Network. Despite the irregular data domain, our data representation allows us to efficiently introduce both convolution and pooling operations and train the network in an end-to-end fashion without precomputed descriptors. Convolution operation in our case is a standard message-passing mechanism. Pooling operation aggregates the information within amino acids and downsamples the initial graph from the atom-level representation to the residue-level representation. The resultant model, VoroCNN, predicts local qualities of 3D protein folds.

Results

In this work, we report the performance of the model trained and validated on the data from CASP[8-11]. We assessed the quality of global scoring on the data from CASP12 and CASP13, and the quality of local scoring on the data from CASP13. In both cases, the prediction results are competitive to the state of the art and superior to the previous 3D CNN architectures built for the same task. We also discuss practical applications of VoroCNN, for example, in recognition of protein binding interfaces.

In the past blind protein structure prediction challenge CASP14, VoroCNN was ranked as the second protein model quality assessment method according to several evaluation metrics among more than 70 methods.

The code is available at <https://team.inria.fr/nano-d/software/vorocnn/>.

References

- Igashov, I., Olechnovič, K., Kadukova, M., Venclovas, Č., and Grudin, S. VoroCNN: Deep convolutional neural network built on 3D Voronoi tessellation of protein structures. *Bioinformatics*, 02 2021. ISSN 1367-4803. doi: 10.1093/bioinformatics/btab118. URL <https://doi.org/10.1093/bioinformatics/btab118>. btab118.
- Olechnovič, K. and Venclovas, Č. Voronota: A fast and reliable tool for computing the vertices of the Voronoi diagram of atomic balls. *Journal of computational chemistry*, 35(8):672–681, 2014.