Using the DeepAtom framework for DPI prediction

Weekly update

Week 1 April, 2021

Overview

Data pre-processing: 3D grid creation
- i. Rasterization
- ii. Channelization

Network architecture
- i. Atom information integration block
- ii. Stacked feature extraction block
- iii. Global affinity regression block

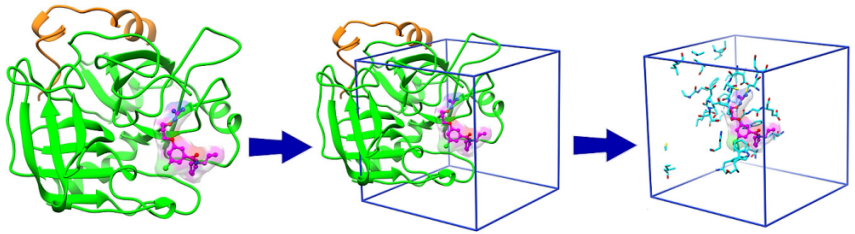
Current progress

DeepAtom Overview

A 3D-CNN based approach to predicting binding affinity.

1. Input: "mostly" just the ligand part of the co-complex, rasterized into a 3D grid box.
2. Novel voxel representation: each voxel has 24 channels, "embedding the different raw information of atoms located around the voxel".
3. Grid processed using ...

i. Rasterization (1/2)



A 3D grid of dimensions $32 \times 32 \times 32$ Å^3 with resolution 1 Å is created, centered on the ligand.

1. The greatest end-to-end distance of a ligand in PDBbind v.2016 is ~ 32 Å.
2. The vdW radius of the heavy atoms found in PDBbind ligands is ≥ 1.4 Å.

i. Rasterization (2/2)

PCMax algorithm: when converting atoms to voxels, atoms-as-voxels have a continuous effect on their neighbors.

For each voxel, its occupancy taken as the maximum of

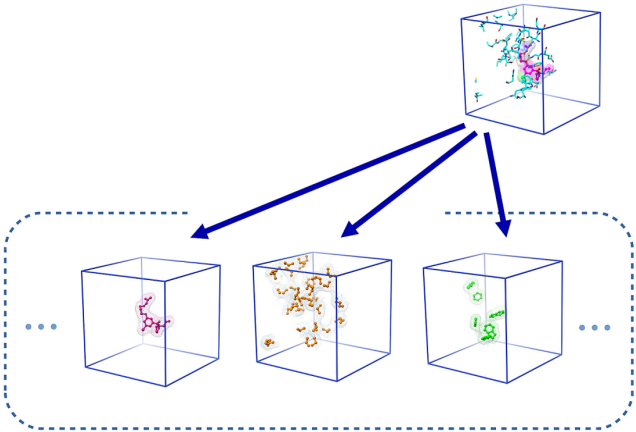
$$n(r_i) = 1 - \exp\left(-\left(\frac{r_{vdW,i}}{r_i}\right)^{12}\right)$$

over all atoms i, where $r_{vdW,i}$ is the i^{th} atom radius and r_i is the inter-voxel distance.

ii. Channelization

The input to the network is $32 \times 32 \times 32 \times 24$: 12 channels each created from the ligand and protein.

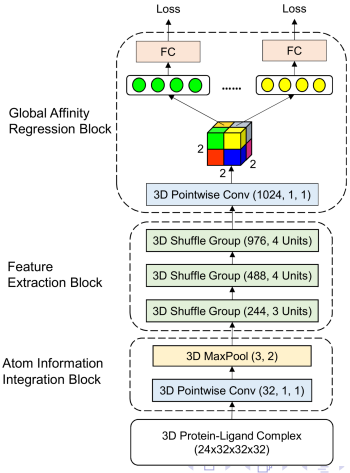
- 1 excluded volume channel.
- 11 channels of atoms involved in certain interatomic interactions.



The Three Block Overview

DeepAtom is relatively shallow, and has three blocks:

1. Atom information integration block.
2. Stacked feature extraction block.
3. Global affinity regression block.

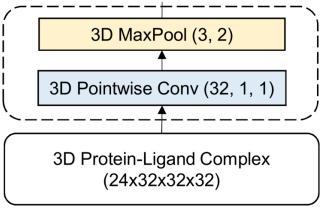


i. Atom Information Integration Block

Atomic information across the 24 channels is aggregate in a pointwise fashion, and the output is then pooled.

1. $\text{PWConv}(W, h)_{(i,j,k)} = \Sigma_m^{24} W_m \cdot h_{(i,j,k,m)}$ is used to map the 24×32^3 input to 32^3.
2. The max-pooling downsamples the tensor to 16^3.

Semantically, this block processes the input into non-linear function of the linear combination of the various channels (interaction types), parsimoniously.

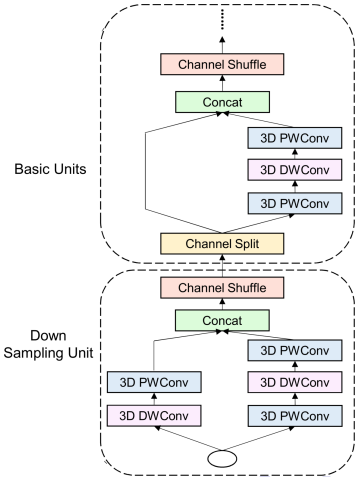


ii. Stacked Feature Extraction Block

Three 3D shuffle groups are stacked, each containing a parsimonious combination of pointwise and depthwise convolutions.

Shuffling, splitting, and then processing only certain channels encourages parsimonious while maintaining performance.

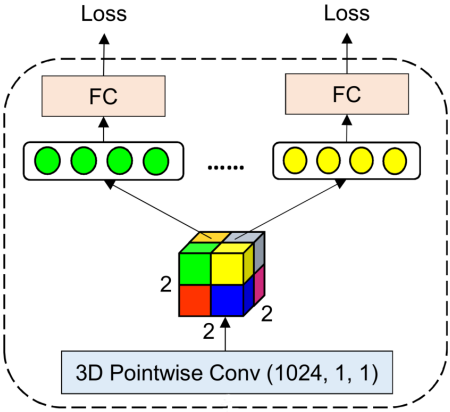
$$\text{DWConv}(W, h)_{(i,j,k)} = \Sigma_{s,t,r}^{S,T,R} W_{s,t,r} \odot h_{(i+s,j+t,k+r)}$$



iii. Global Affinity Regression Block

The output from the shuffle groups is downsampled and flattened into eight 1024-vectors.

- Each vector represents the output from a distinct channel subset, and is processed to create a BAP.
- The eight losses are used to train the network concurrently. During evaluation, the eight vectors are averaged before a single BAP is created.



Current Progress

1. Rasterize protein PDBs joined to ligand MOL2 files.
   - Done using Python's `HTMD` package.
2. Channelize 3D grids.
   - Done using an offline implementation of Arpeggio's web service.
3. Implement DeepAtom architecture.
   - Shuffle groups expected to be a particular challenge.
4. Train, tune, and validate DeepAtom on NSCC.
   - I have more "low-level" control over this DeepAtom implementation than I do of those of DrugVQA and GGNN.

For More Details...

DeepAtom:

Li, Y., Rezaei, M. A., Li, C., & Li, X. (2019)
DeepAtom: A framework FOR Protein-Ligand binding affinity prediction
IEEE/ACM Transactions on Computational Biology and Bioinformatics.

PCMax algorithm/rasterization of atoms:

Jiménez, J., Doerr, S., Martínez-Rosell, G., Rose, A. S., & De Fabritiis, G. (2017)
Deepsite: Protein-binding site predictor using 3d-convolutional neural networks
Bioinformatics 33(19), 3036-3042.

For More Details...

Arpeggio interatomic interaction finder:

Jubb, H., Higueruelo, A., Ochoa-Montaño, B., Pitt, W., Ascher, D., & Blundell, T. (2017)

Arpeggio: A web server for calculating and visualising interatomic interactions in protein structures

Journal of Molecular Biology 429(3), 365-371.

ShuffleNet:

Zhang, X., Zhou, X., Lin, M., & Sun, J. (2018)

Shufflenet: An extremely efficient convolutional neural network for mobile devices

2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition.