Molecular docking is an important tool for computational drug discovery that aims to predict the binding pose of a ligand (drug) to a target protein. Identifying a correctly oriented pose requires a scoring function that has a global optimum close to the experimentally observed pose. Additionally, it should be differentiable with respect to atomic positions so that it can be used for gradient-based pose optimization.

We describe a differentiable grid-based convolutional neural network scoring function that can be explored as a global optimization problem in an end-to-end GPU-optimized molecular docking workflow. We show that convolutional neural networks trained on experimental data can successfully identify correct binding modes and meaningful rank and score compounds. We visualize these learned functions and explore their application in an end-to-end GPU-optimized molecular docking pipeline.

### Background

Protein-ligand scoring provides a metric of binding strength between small molecules and target proteins and is critical in identifying ligands for a desired protein. An ideal scoring function would correctly identify accurate ligand poses and predict the binding affinity of the ligand for the protein.

One approach for scoring is to use machine learning. Traditionally, this has required manually selecting molecular features, such as pairwise interactions and counts of typical chemical interaction patterns, that are used to train a model. However, other, non-parametric, machine learning models can learn the most important features directly from low-level representations of the data.

Neural networks are a supervised machine learning algorithm inspired by the nervous system. A basic network consists of an input layer, one or more hidden layers, and an output layer of interconnected nodes. Each hidden node computes a feature that is a non-linear function of the weighted input it receives from the nodes of the previous layer. A neural network with a finite number of nodes can approximate any continuous function to within a given error over a bounded input domain.

Data augmentation is performed by juxtaposing test structures with translations (±6Å) to protein–ligand complex structures. This reduces overfitting and compensates for the coordinate-frame dependency of a 3D grid representation.

### Data Representation

24x24x24x4 grid at 0.5Å resolution 14 ligand and 14 receptor atom types Continuous Gaussian density CNN optimized grid generation

### Datasets

GNN

SMINA docked and minimized poses are used for training.

### Redocked

POBind refined set

4053 complexes

52,166 ligand poses

Affinity data for all ligands

### Cross-Docked

Structures from Pocketome

2923 distinct pockets

27,142 receptor structures

4,138,117 non-redundant ligand poses

### Pose Optimization

Loss gradients of a trained model can be backpropagated onto the input grid. They indicate how the input can be changed to increase its score. These gradients can be further propagated onto atom centers as a vector quantity that can then be interpreted as a force in a pose optimization algorithm.

### Hyperparameter Optimization

50 parameters for training and the network topology were explored both automatically and manually.

### Models

#### Default2017 baseline model for optimization

#### Default2018 Optimized model with network-in-network architecture and fast evaluation

#### HilRes Affinity High resolution affinity model

### Cross Validation Performance

<table>
<thead>
<tr>
<th>CNN Atomic Potentials</th>
<th>Aliphatic/Carbon</th>
<th>Oxygen/Derivatives/Acceptors</th>
</tr>
</thead>
</table>

### ACKNOWLEDGMENTS

This research was supported by R01GM118349 from the National Institutes of General Medical Sciences. Additional contributions from aagrant.org, Google Cloud, NVIDIA Corporation, and the University of Pittsburgh Center for Simulation and Modeling.