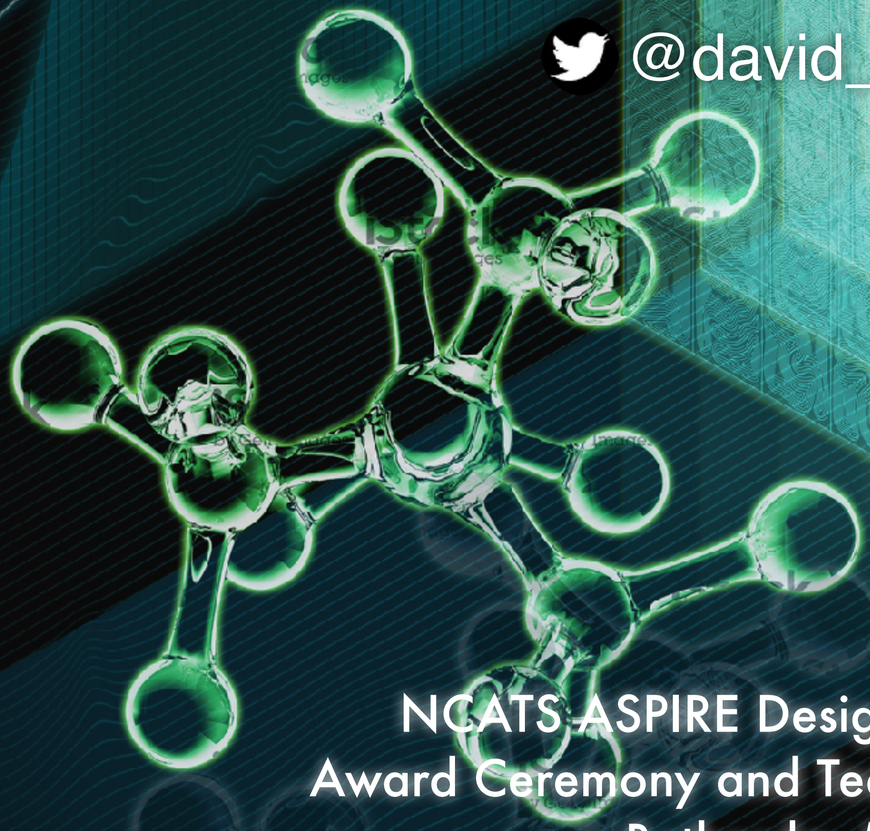
Deep and Generative Structure-Based Models for Drug Discovery

David Koes

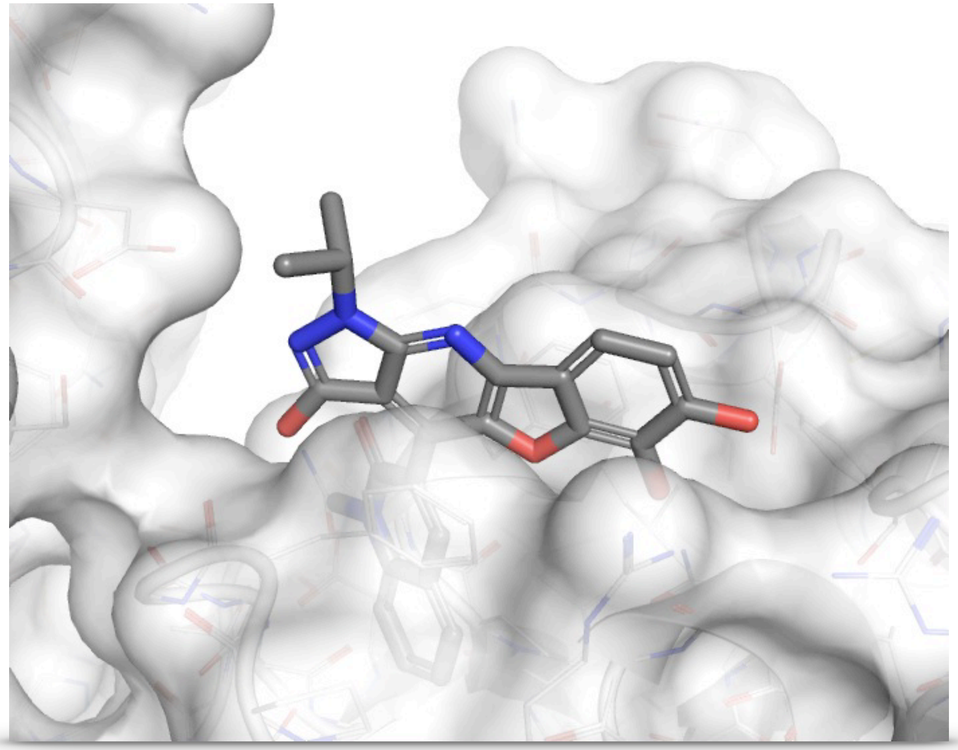
@david_koes



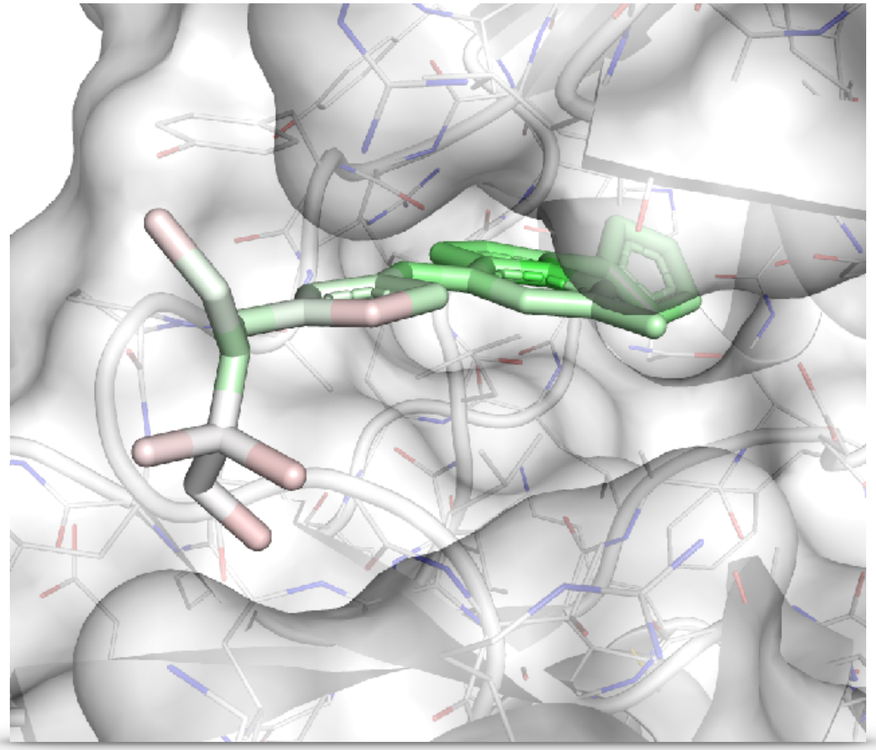
NCATS ASPIRE Design Challenges
Award Ceremony and Team-Building Event
Bethesda, MD
October 28, 2019

Structure Based Drug Design

Pose Prediction → Binding Discrimination → Affinity Prediction

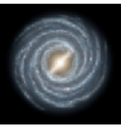


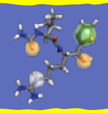
Virtual Screening



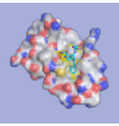
Lead Optimization

Drug Discovery Funnel

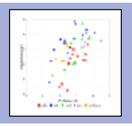
Purchasable  Accessible

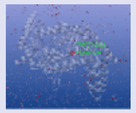


Matching

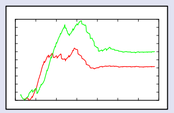


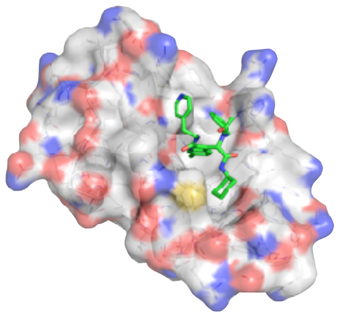
Scoring

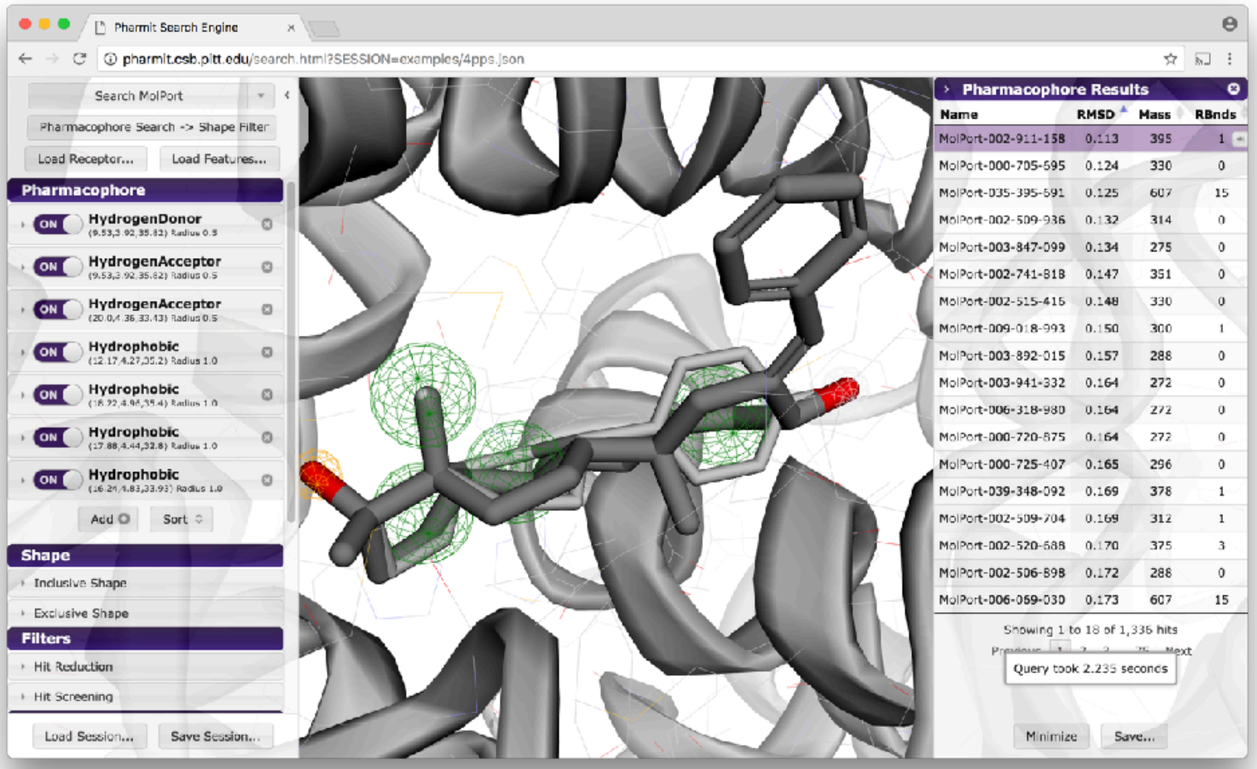




Dynamics



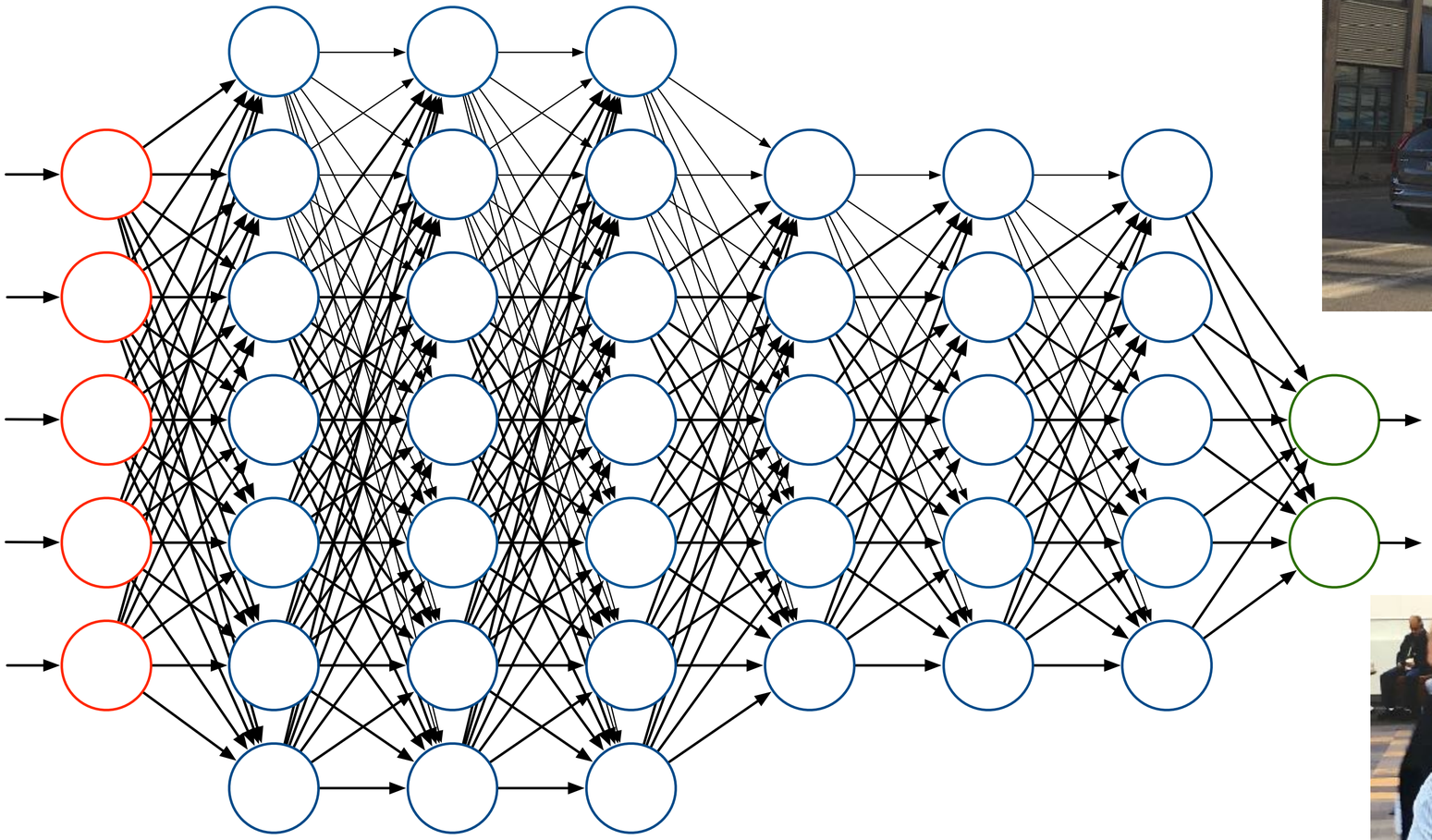




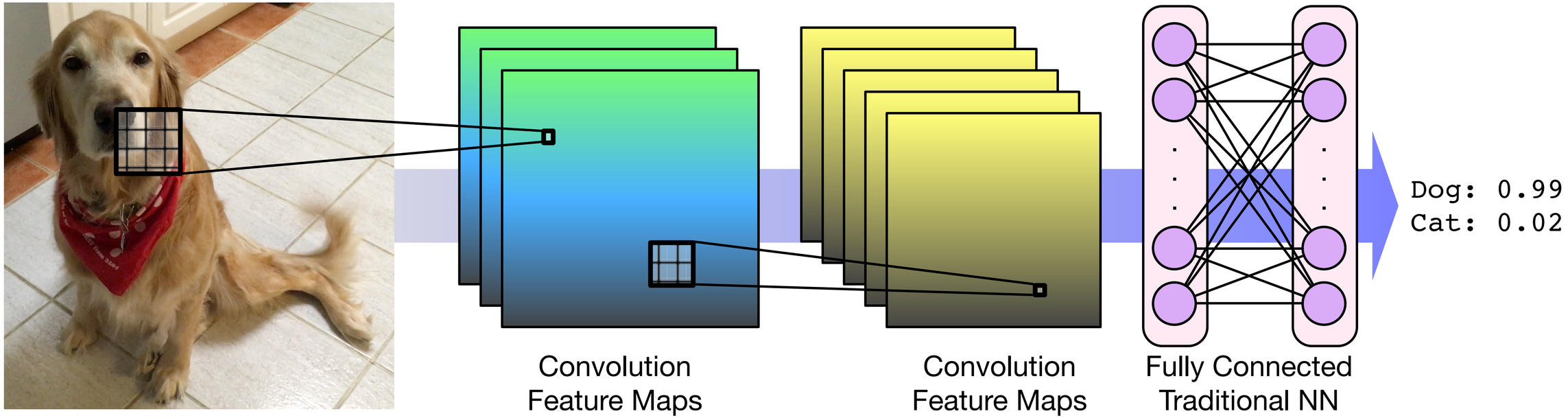
http://pharmit.csb.pitt.edu

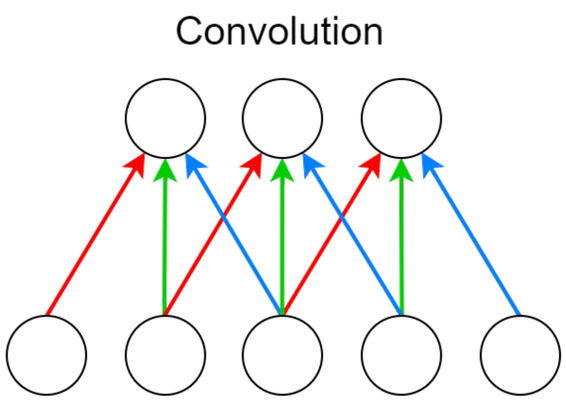
Aim 1: Develop and deploy deep 3D convolutional neural network protein-ligand scoring functions tailored for pain and opioid abuse molecular targets.

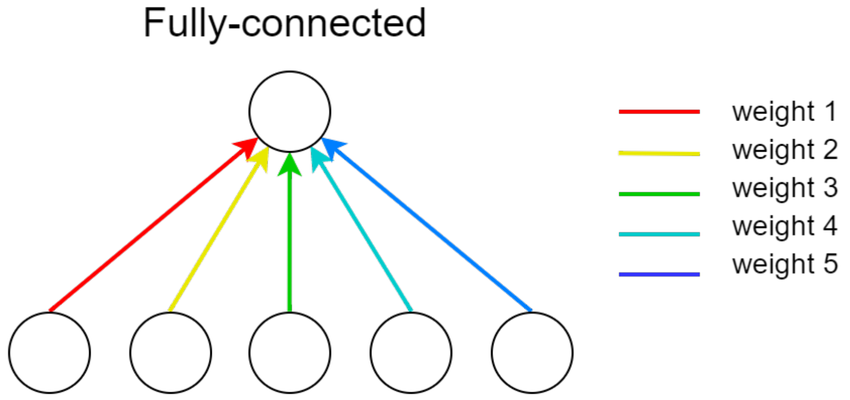
Deep Learning



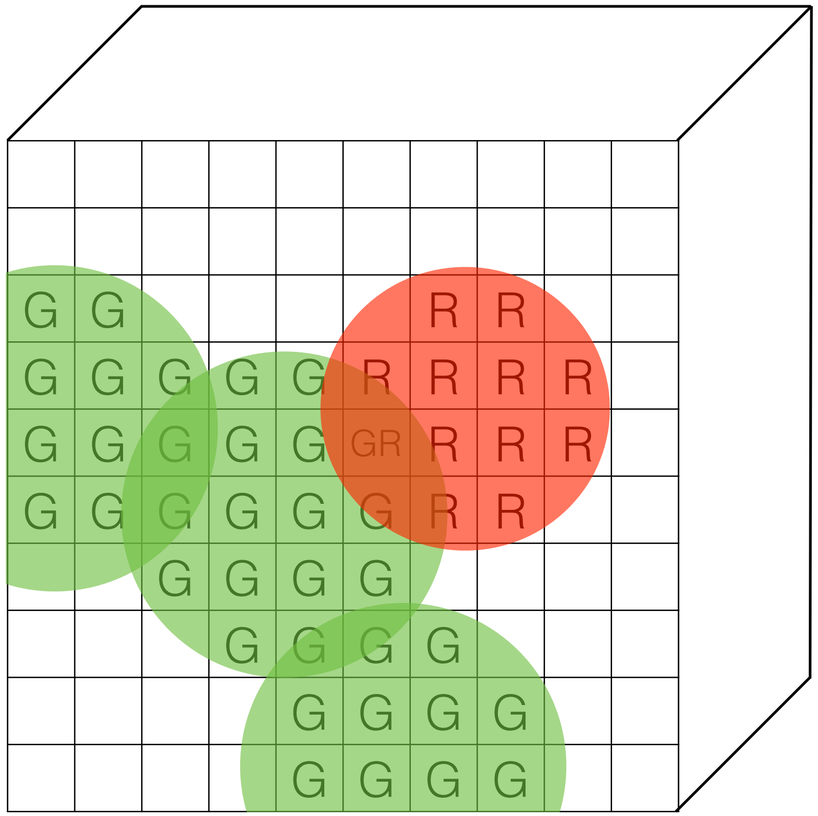
Convolutional Neural Networks







Protein-Ligand Representation

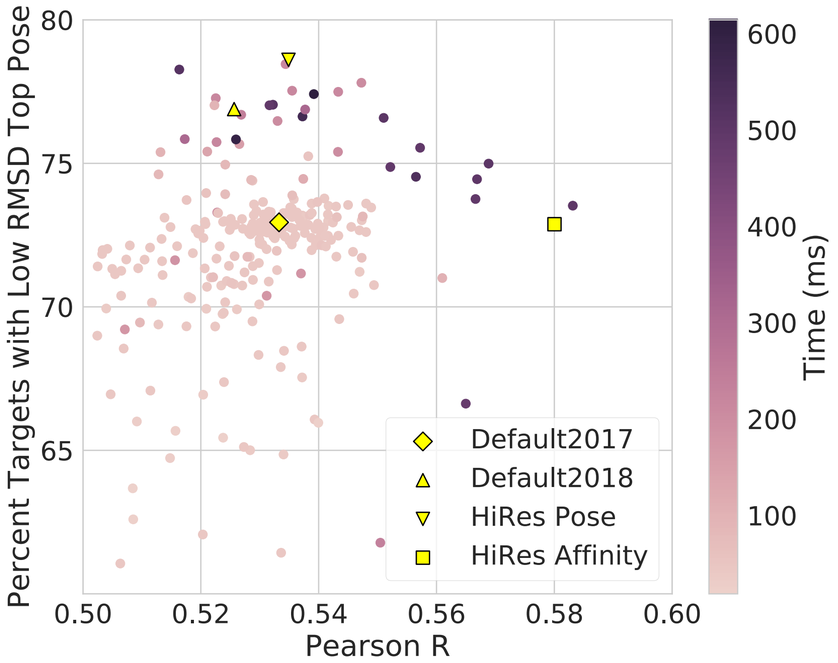


(R,G,B) pixel →

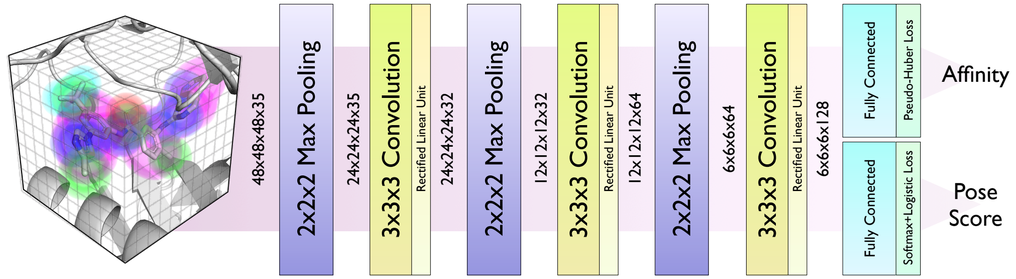
(Carbon, Nitrogen, Oxygen,…) **voxel**

The only parameters for this representation are the choice of **grid resolution**, **atom density**, and **atom types**.

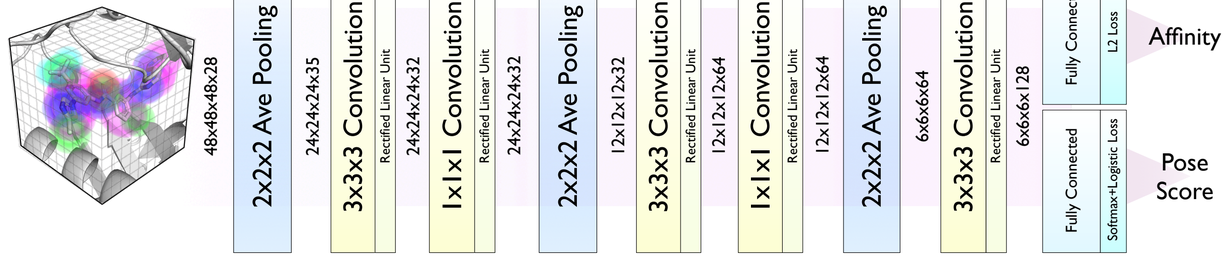
Optimized Models



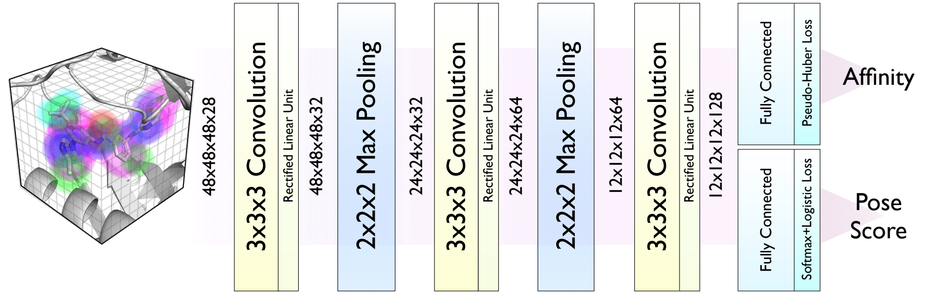
Default2017



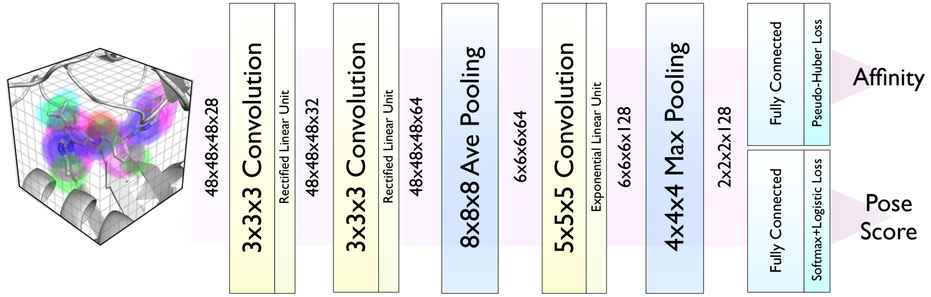
Default2018



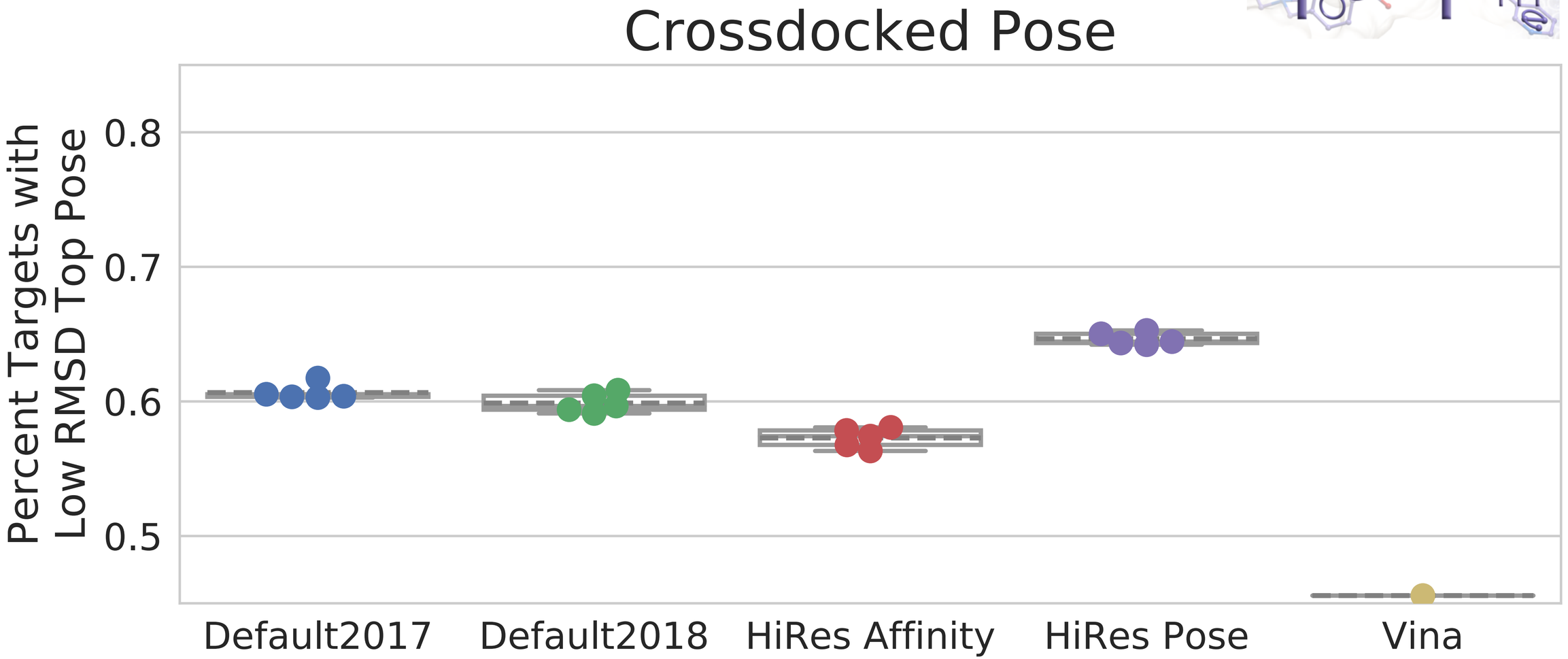
HiRes Pose



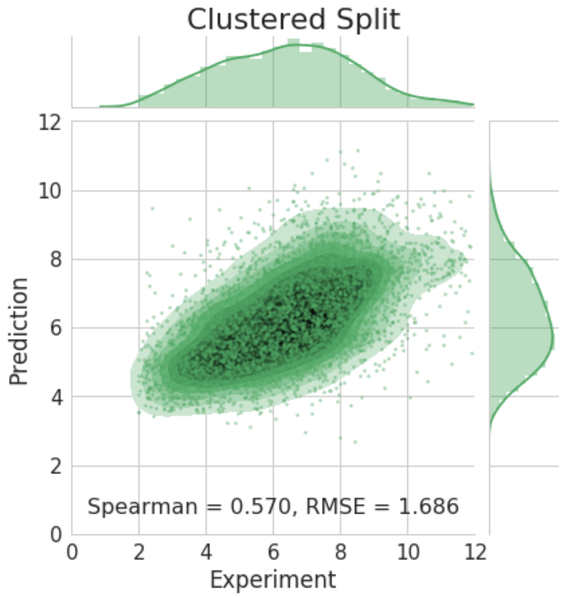
HiRes Affinity

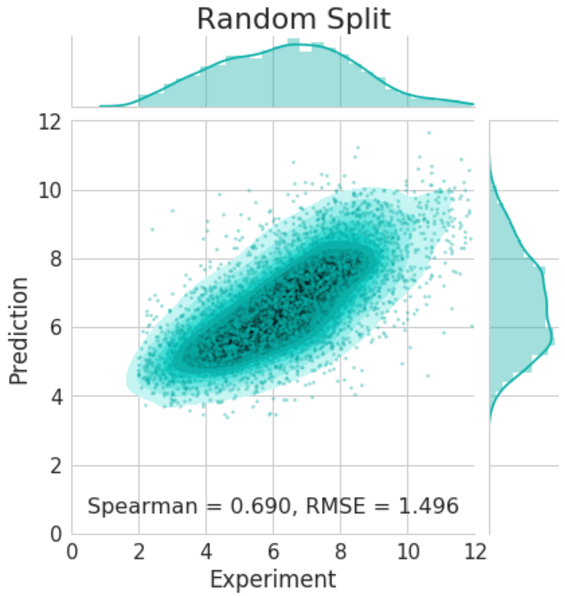


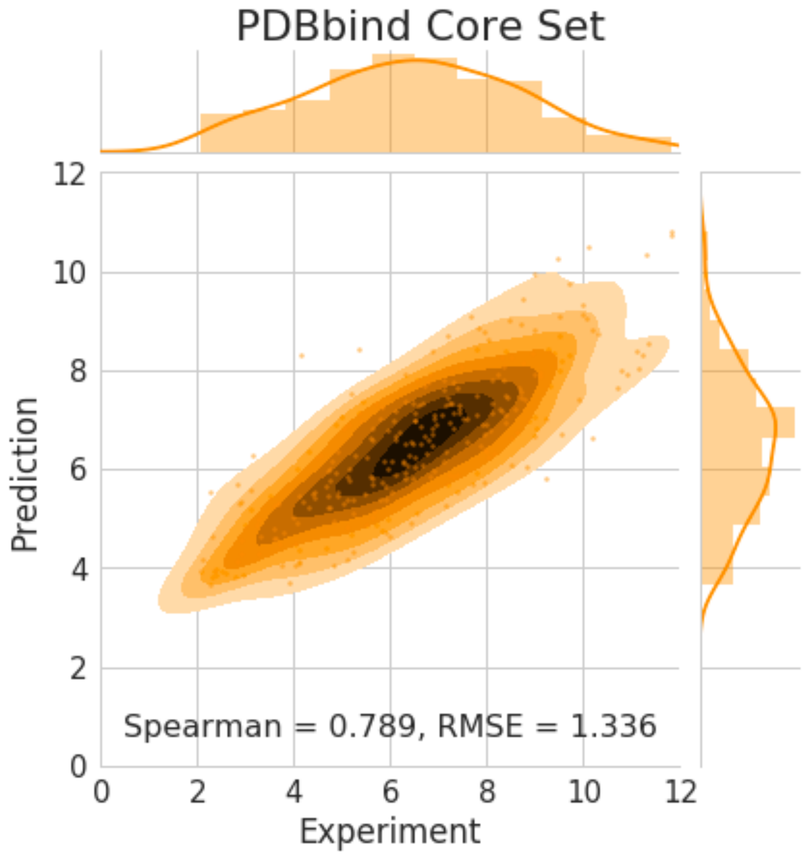
Pose Results

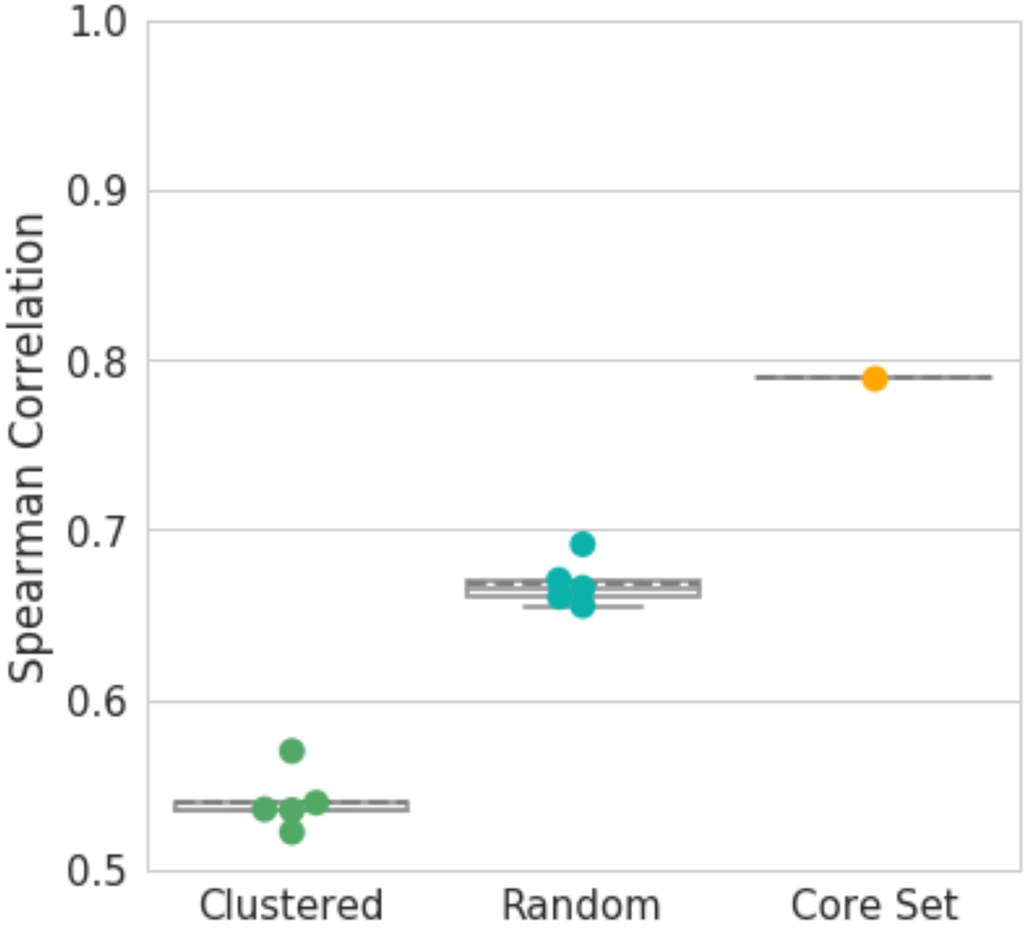


Affinity Results





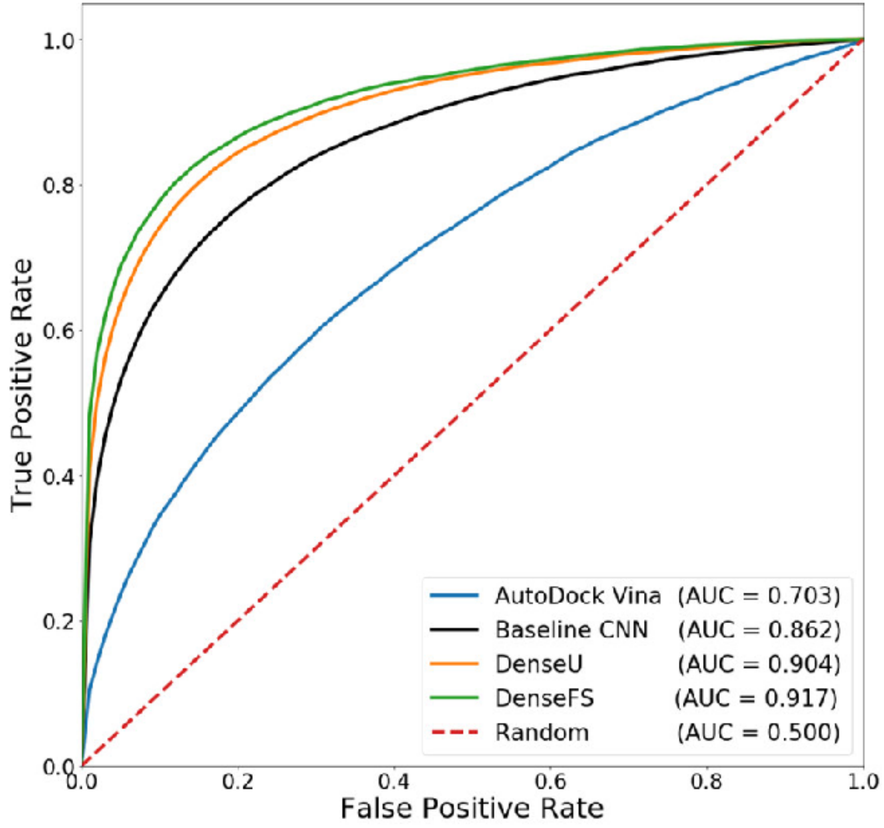


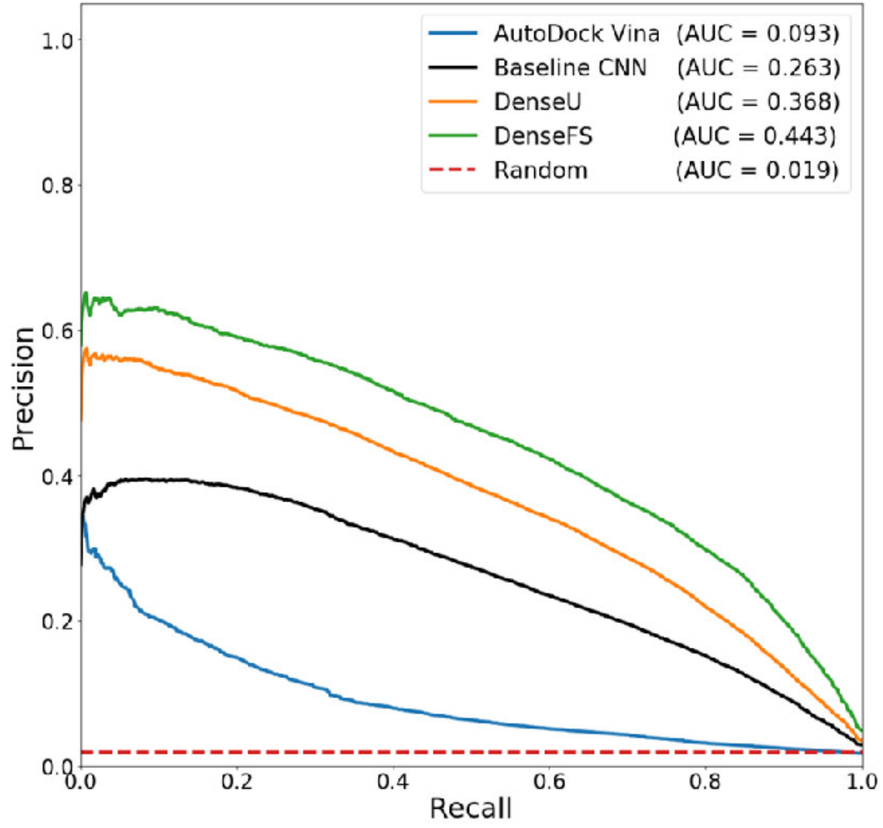


Virtual Screening

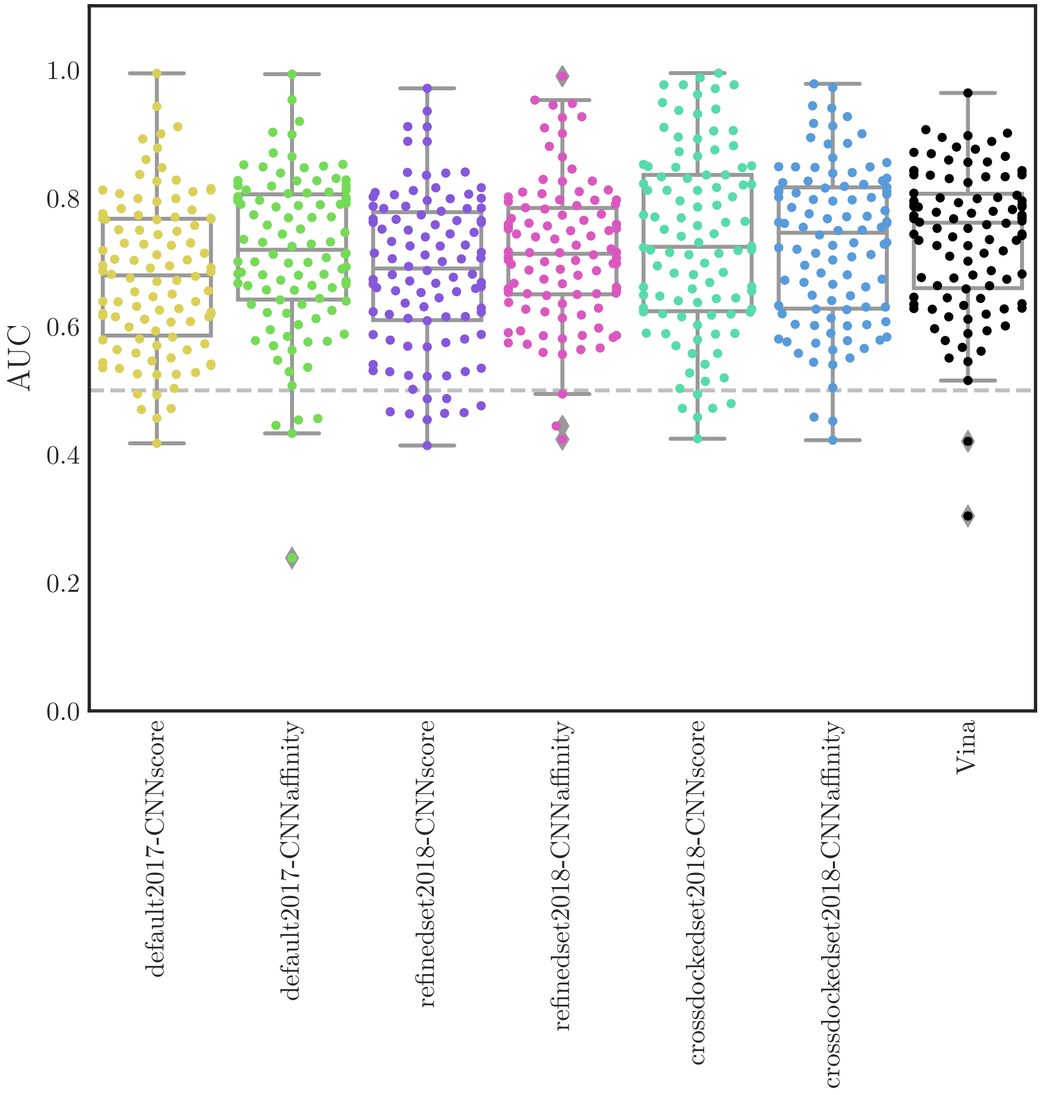
Protein Family-Specific Models Using Deep Neural Networks and Transfer Learning Improve Virtual Screening and Highlight the Need for More Data

Fergus Imrie[†], Anthony R. Bradley[‡¶§], Mihaela van der Schaar[†⊥], and Charlotte M. Deane*[†]
[†] Oxford Protein Informatics Group, Department of Statistics, University of Oxford, Oxford OX1 3LB, U.K.
[‡] Structural Genomics Consortium, University of Oxford, Oxford OX3 7DQ, U.K.
[¶] Department of Chemistry, University of Oxford, Oxford OX1 3TA, U.K.
[§] Diamond Light Source Ltd., Didcot OX11 0DE, U.K.
[⊥] Department of Engineering, University of Oxford, Oxford OX1 3PJ, U.K.
– Alan Turing Institute, London NW1 2DB, U.K.





Virtual Screening



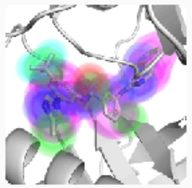
In Need of Bias Control: Evaluating Chemical Data for Machine Learning in Structure-Based Virtual Screening

Jochen Sieg, Florian Flachsenberg, and Matthias Rarey*
Universität Hamburg, ZBH - Center for Bioinformatics, Research Group for Computational Molecular Design, Bundesstraße 43, 20146 Hamburg, Germany

Hidden Bias in the DUD-E Dataset Leads to Misleading Performance of Deep Learning in Structure-Based Virtual Screening

Preprint submitted on 24.03.2019, 15:39 and posted on 25.03.2019, 12:58 by Lieyang Chen, Anthony Cruz, Steven Ramsey, Callum J. Dickson, José S. Duca, Viktor Hornak, David R. Koes, Tom Kurtzman

Search or jump to... / Pull requests Issues Marketplace Explore



gnina

Repositories 4 People 9 Teams 1 Projects 0 Settings

Find a repository... Type: All Language: All

gnina

gnina is not smina/vina

cheminformatics computational-chemistry drug-discovery

C++ ★ 35 26 Updated 11 days ago

scripts

Jupyter Notebook ★ 1 13 BSD-3-Clause Updated on Sep 17, 2018

models

Trained caffe models

★ 6 9 Updated on Aug 23, 2018

Protein–Ligand Scoring with Convolutional Neural Networks

Matthew Ragoza†‡, Joshua Hochuli‡¶, Elisa Idrobo§, Jocelyn Sunseri∥, and David Ryan Koes*∥
†Department of Neuroscience, ‡Department of Computer Science, ¶Department of Biological Sciences, and ∥Department of Computational and Systems Biology, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, United States
§ Department of Computer Science, The College of New Jersey, Ewing, New Jersey 08628, United States

J. Chem. Inf. Model., **2017**, *57* (4), pp 942–957
DOI: 10.1021/acs.jcim.6b00740
Publication Date (Web): April 3, 2017
Copyright © 2017 American Chemical Society

Cite this: *J. Chem. Inf. Model.* 2017, 57, 4, 942-957
RIS Citation GO

arXiv.org > stat > arXiv:1710.07400

Statistics > Machine Learning

Ligand Pose Optimization with Atomic Grid-Based Convolutional Neural Networks

Matthew Ragoza, Lillian Turner, David Ryan Koes

(Submitted on 20 Oct 2017)

Journal of Molecular Graphics and Modelling
Volume 84, September 2018, Pages 96-108

Visualizing convolutional neural network protein-ligand scoring

Joshua Hochuli, Alec Helbling, Tamar Skaist, Matthew Ragoza, David Ryan Koes

Journal of Computer-Aided Molecular Design
... pp 1–15 | Cite as

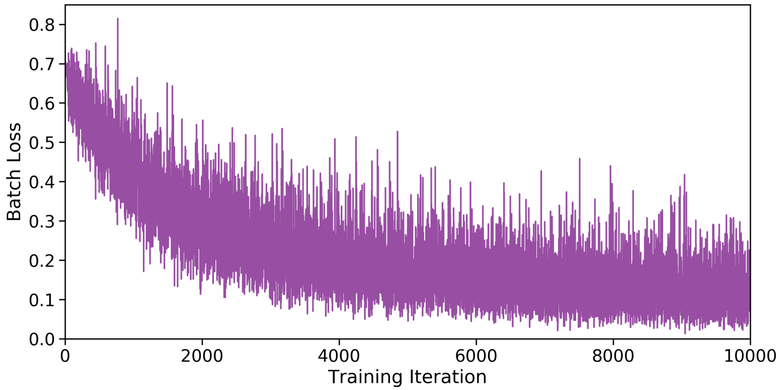
Convolutional neural network scoring and minimization in the D3R 2017 community challenge

Authors Authors and affiliations

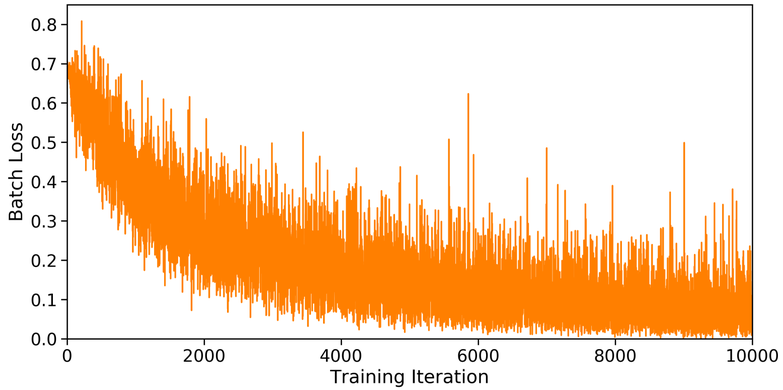
Jocelyn Sunseri, Jonathan E. King, Paul G. Francoeur, David Ryan Koes

libmolgrid

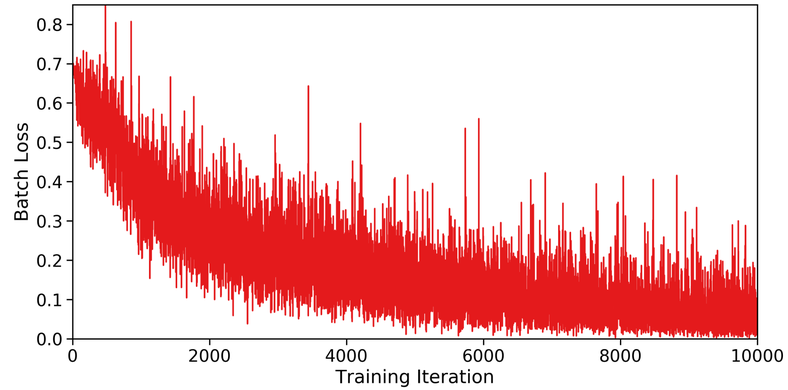
Caffe Training



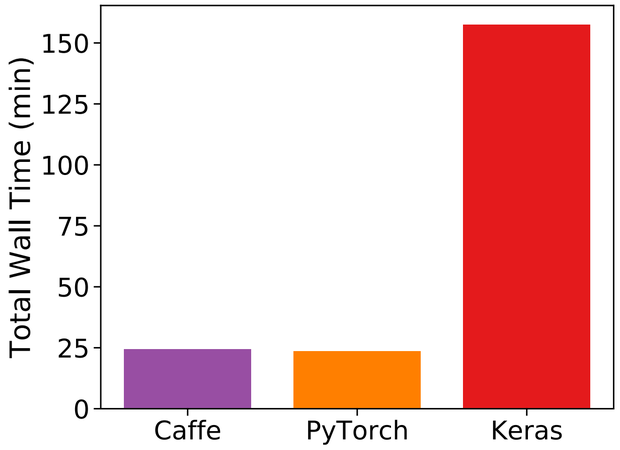
PyTorch Training



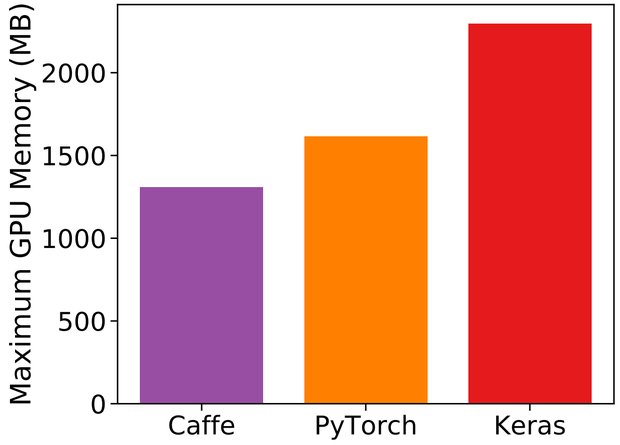
Keras Training



GPU Performance



GPU Memory Utilization

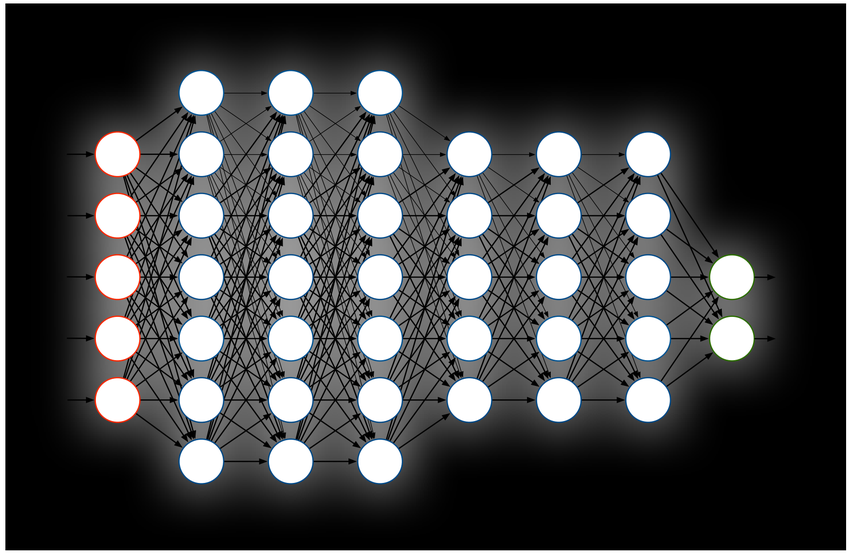



```
e = molgrid.ExampleProvider(balanced=True,shuffle=True)
e.populate('examples.txt')

gmaker = molgrid.GridMaker()

batch = e.next_batch(batch_size)
gmaker.forward(batch, input_tensor,
               random_translation=0, random_rotation=True)
```

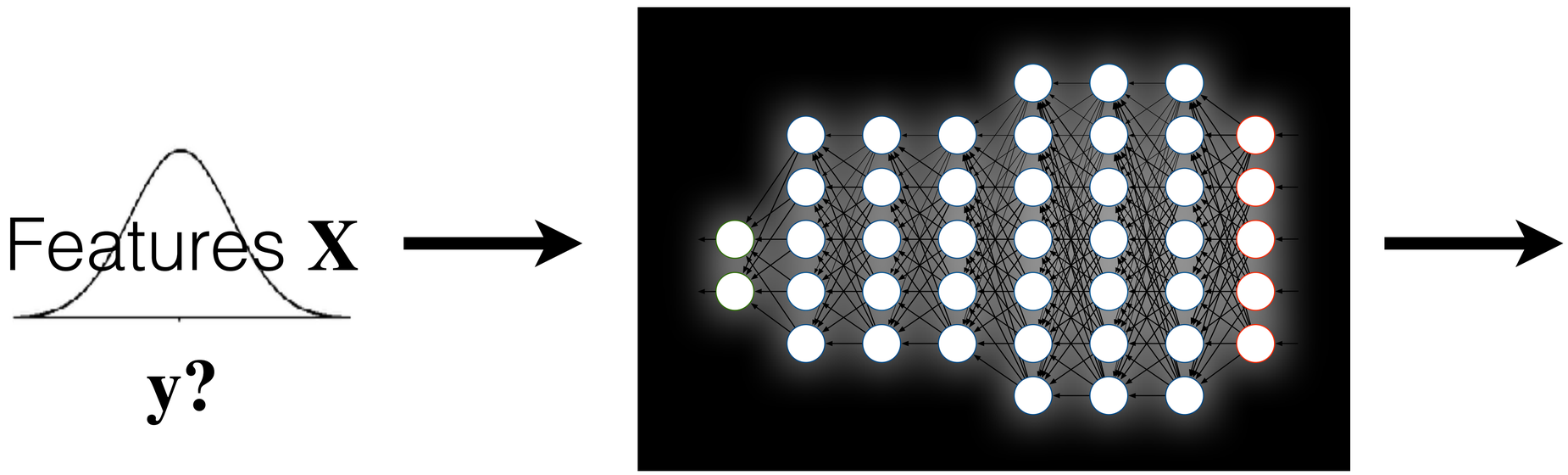
github.com/gnina/libmolgrid

Aim 2: Develop deep generative models for the rapid exploration of synthetically accessible chemical space surrounding pain and opioid abuse lead compounds.

Discriminative Model

Features $\mathbf{x}$  Prediction $\mathbf{y}$

Generative Model



Generative Adversarial Networks

2014

2015

2016

2017

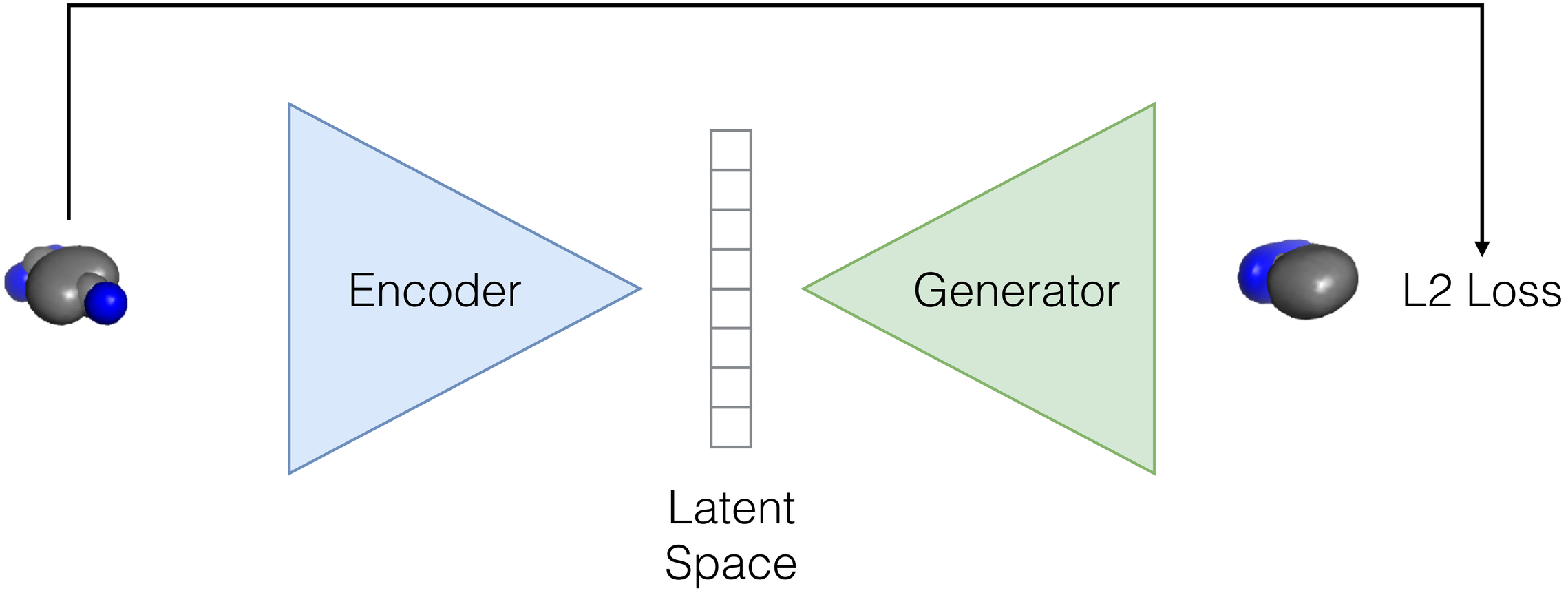
2018

Ian Goodfellow @goodfellow_ian · 2h

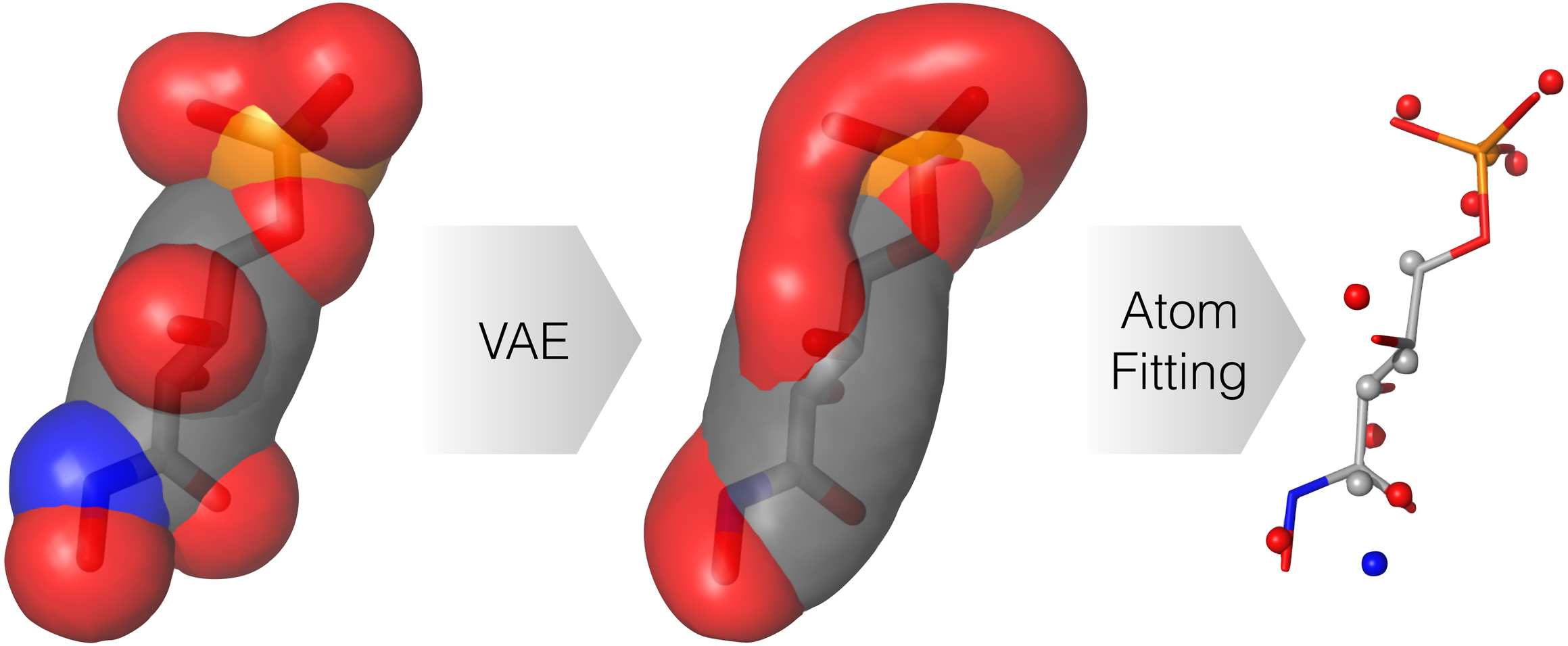
4.5 years of GAN progress on face generation. arxiv.org/abs/1406.2661 arxiv.org/abs/1511.06434 arxiv.org/abs/1606.07536 arxiv.org/abs/1710.10196 arxiv.org/abs/1812.04948

https://thispersondoesnotexist.com

Autoencoding

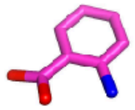
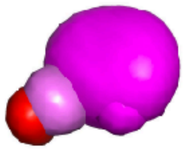
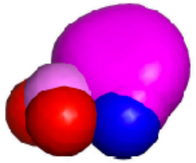
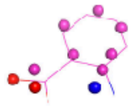
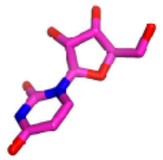
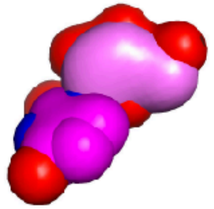
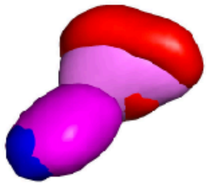
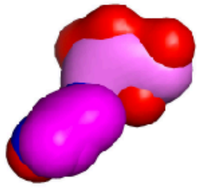
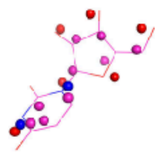
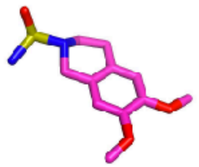
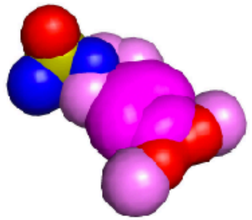
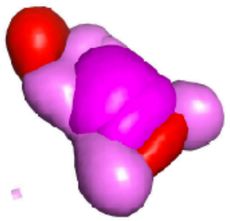
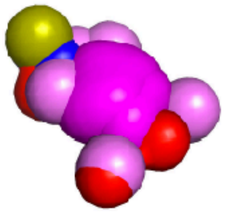
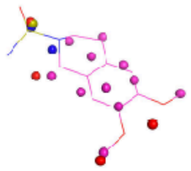
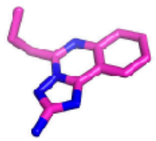
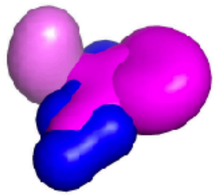
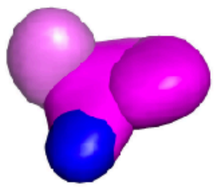
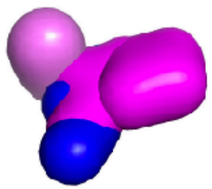
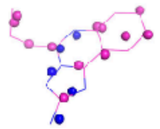


Variational Autoencoding Examples

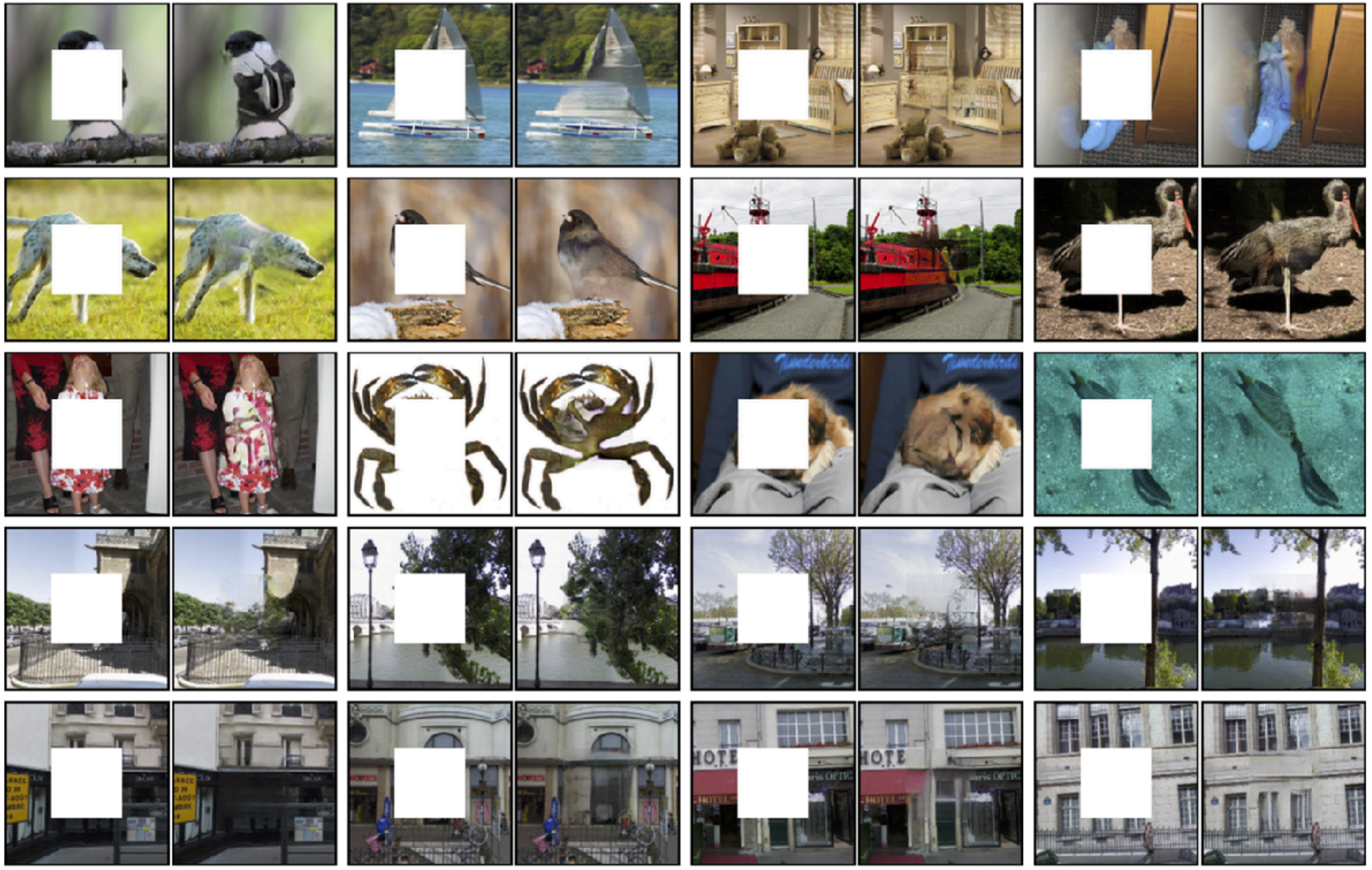


2BES

Variational Autoencoding Examples

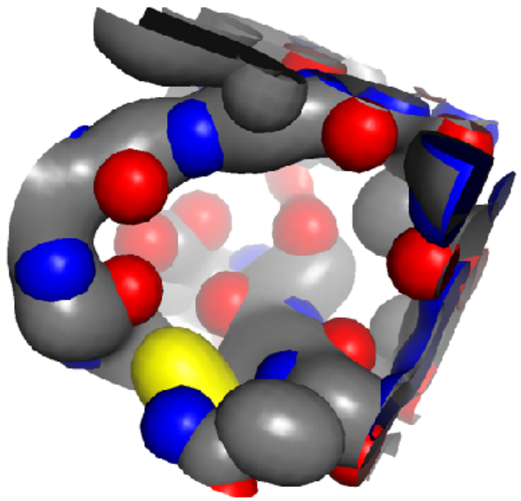
PDB	True structure	True density	Gen. density	Fit density	Fit structure	Gen. L2 distance	Fit L2 distance	Fit RMSD
3h78						9.4053	8.3141	0.6160
4jx9						13.8545	9.7198	0.8820
3igp						14.8525	12.5245	1.2066
4cwf						11.4730	9.0564	0.6725

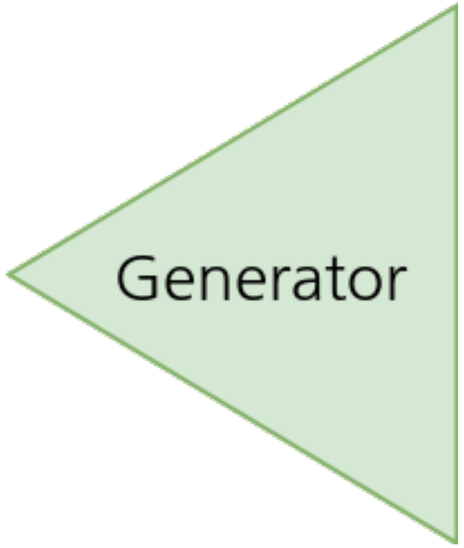
Context Encoding

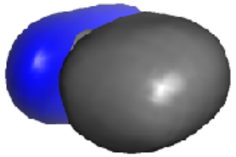


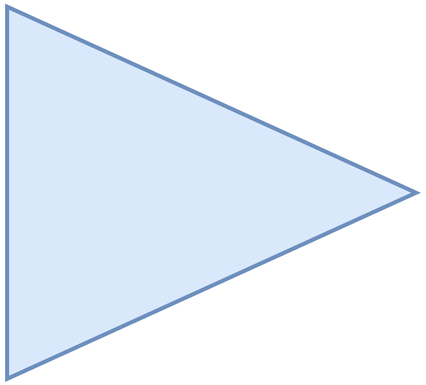
http://people.eecs.berkeley.edu/~pathak/context_encoder/

Context Encoding







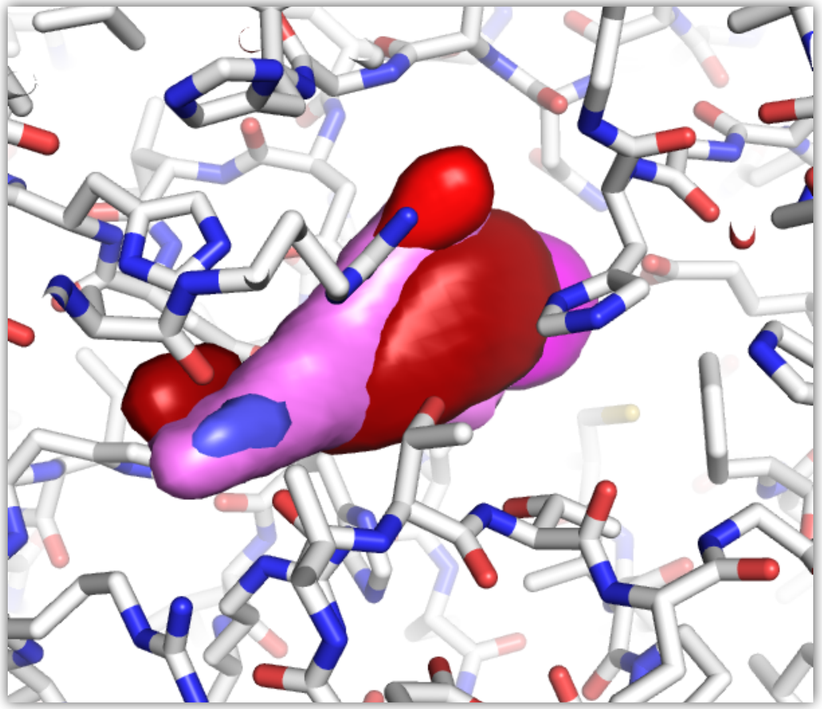


receptor grid

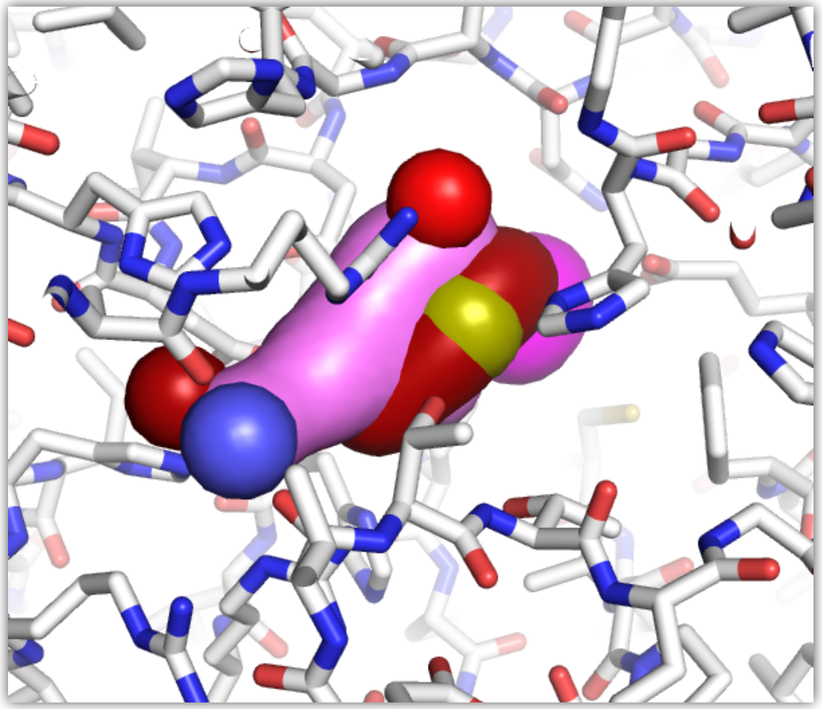
generated ligand grid

Context Encoding with Fully Convolutional Network

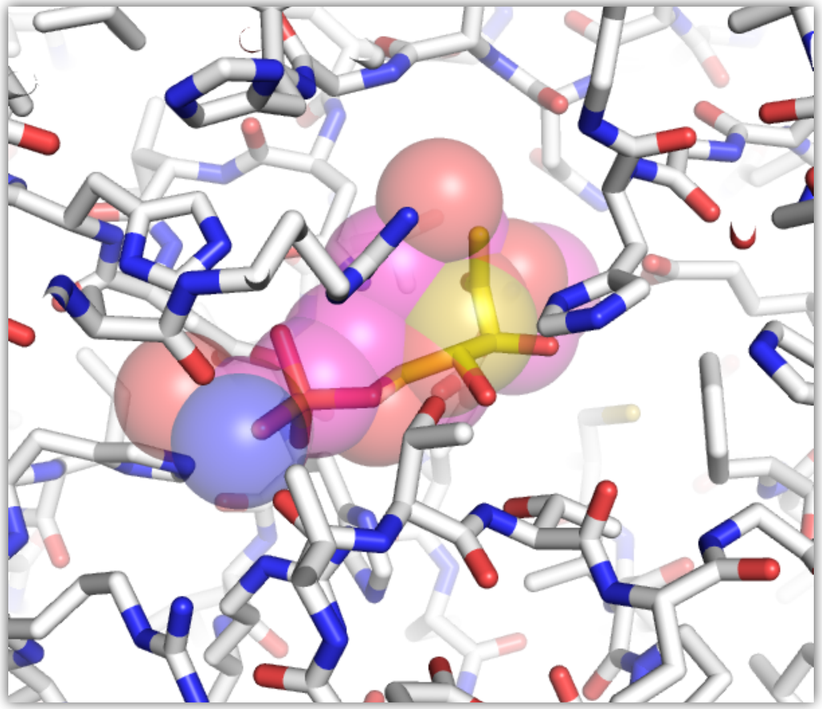
1m5w



Generated



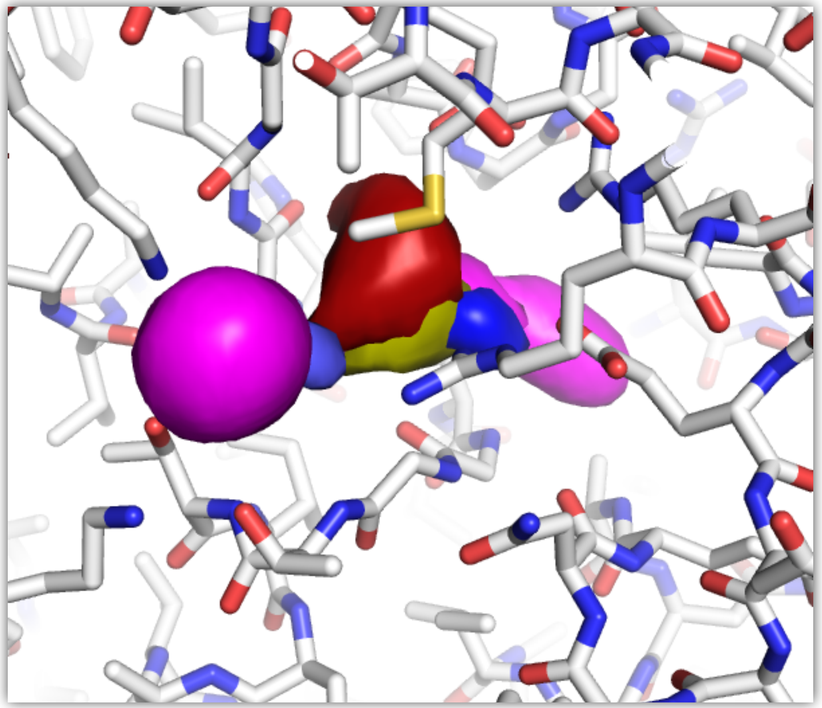
Fit Densities



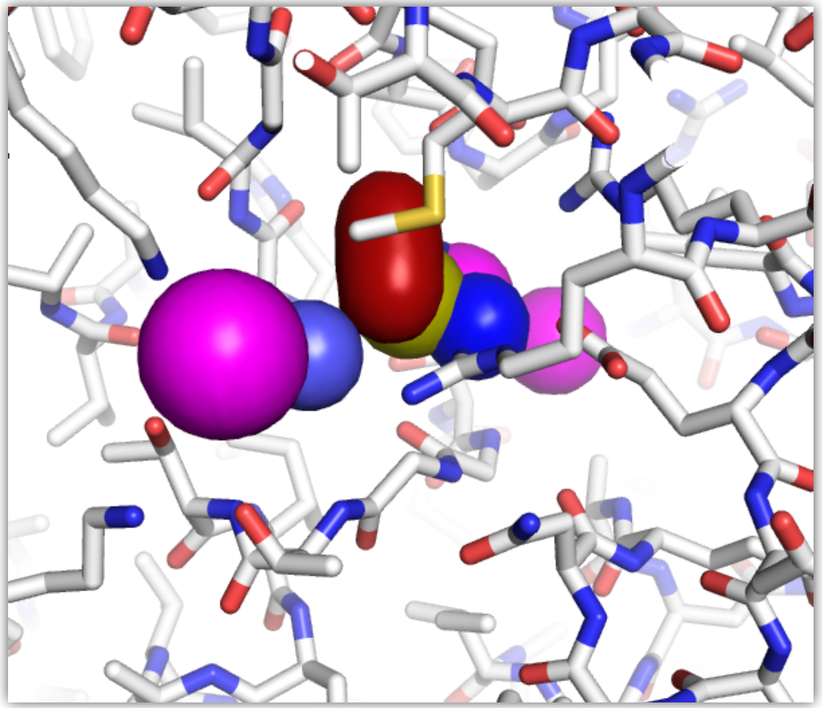
Fit Atoms

Context Encoding with Fully Convolutional Network

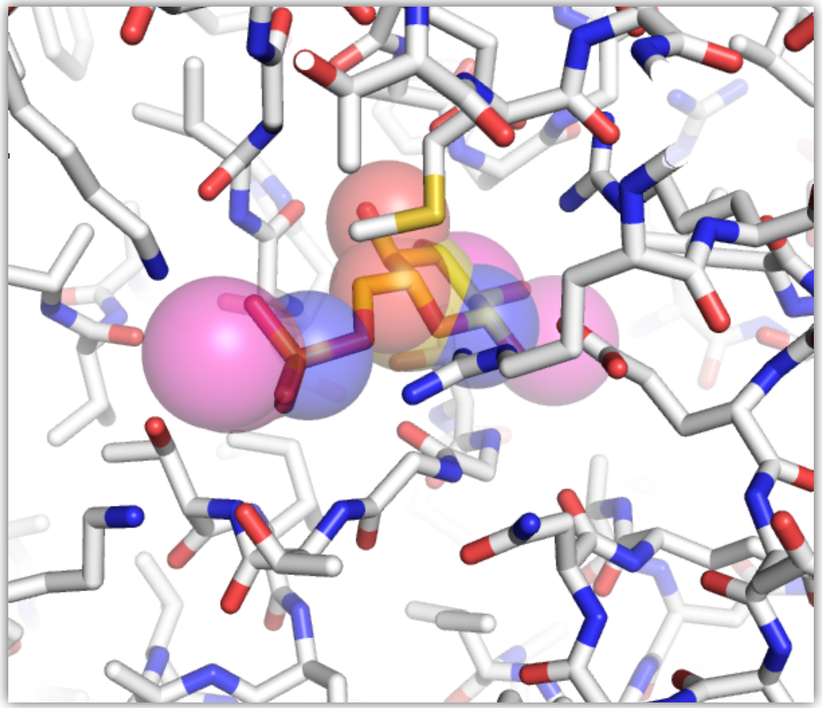
3bxg



Generated



Fit Densities

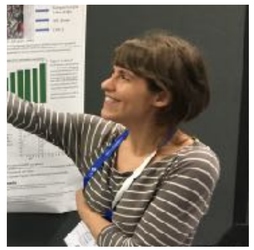


Fit Atoms

NCATS Goals

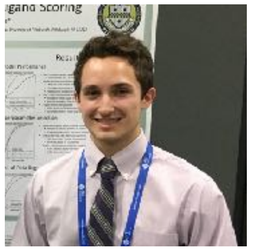
- Deploy opioid abuse structural targets as examples in Pharmit
- Construct opioid abuse focused benchmark datasets in Pharmit
- Develop target-specific CNN scoring function
- Develop 3D generative models and apply on opioid abuse targets

Acknowledgements



Jocelyn Sunseri

Paul Francoeur



Matt Ragoza

Jonathan King

Department of Computational and Systems Biology

NIH National Center for Advancing Translational Sciences

National Institute of General Medical Sciences
R01GM108340

NVIDIA

Google Cloud

RELAY THERAPEUTICS

github.com/gnina

http://bits.csb.pitt.edu

@david_koes