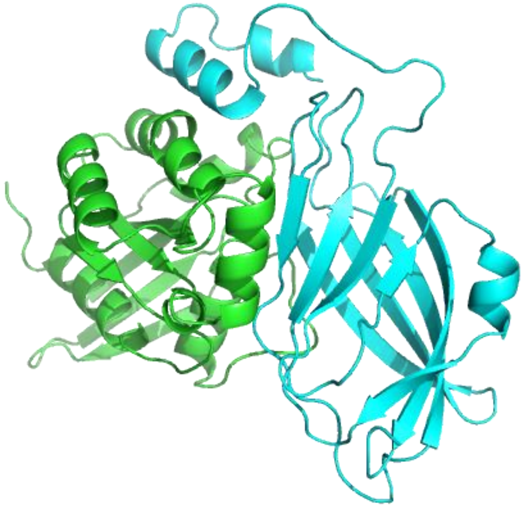
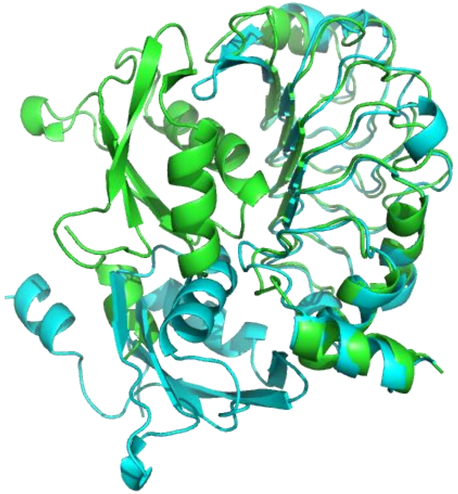
MMRefiner
A Deep Learning Guided Protein Complex Refinement Protocol

MOHIMENUL KARIM

Protein Complex





Motivation

Experimental methods are expensive and time consuming

Computational methods can generate structures with deviation from the native

Orientation of the individual monomers are not always optimal

No deep learning method to refine the near native protein complexes

Method

A deep learning method to predict the refined protein complex structure

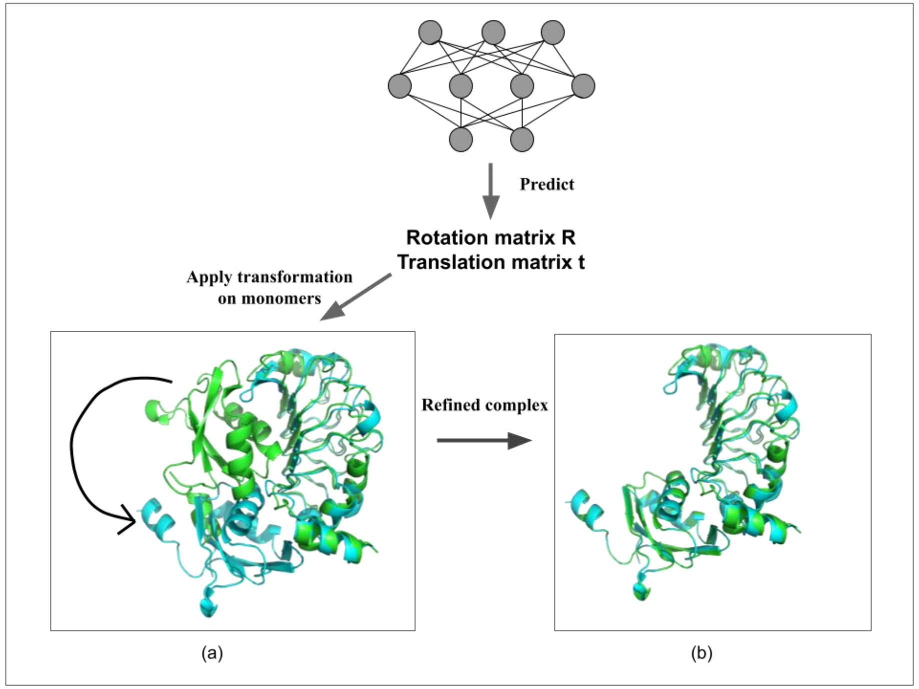
Predict the transformation matrix of individual chains

- Translation matrix
- Rotation matrix

Apply the transformation on the individual chains

Generate the refined protein complex

Considered only chain 2 in this study



Dataset

Dockground

35 complexes (deviation spanning from 1Å to 6Å)

10 decoys for each

Training: 1647 complexes

Test: 480 complexes

No overlap in training and test dataset

Features

Relative position (residue number/chain length)

Residue type (one hot encoded)

Secondary structure (three state; one hot encoded)

Solvent accessibility

Phi (both sine and cosine)

Psi (both sine and cosine)

Translation

$t_i \in t_x, t_y, t_z$

$s_{score} = \frac{1}{1+\left(\frac{ti}{d0}\right)^2}$ $where,\ t_i \in t_x, t_y, t_z\ and\ d_0 = 10$ ……………………….…. (1)

$If\ t_i < 0, then\ s_{score}\ is\ negative$

After prediction

Calculate t_i using equation 1

$If\ s_{score}\ is\ negative, then\ t_i\ is\ negative$

Rotation

Calculate the angles of rotation along x, y and z axis from the rotation matrix

Predict sine and cosine of the angles of rotation

Calculate the angles back from the prediction using the following equation

$$\theta = tan^{-1}\frac{sin\theta}{cos\theta}$$

Calculate the rotation matrix from the angles

Architecture 1

Conv2D (input channels=60, output channels=128)

ReLU

Max pool (2x2)

Conv2D (input channels=128, output channels=256)

ReLU

Max pool (2x2)

Fully connected layer (256)

Output layer (3 for translation and 6 for rotation)

Architecture 2

Conv2D (input channels=60, output channels=128)

ReLU

Max pool (2x2)

Conv2D (input channels=128, output channels=256)

ReLU

Max pool (2x2)

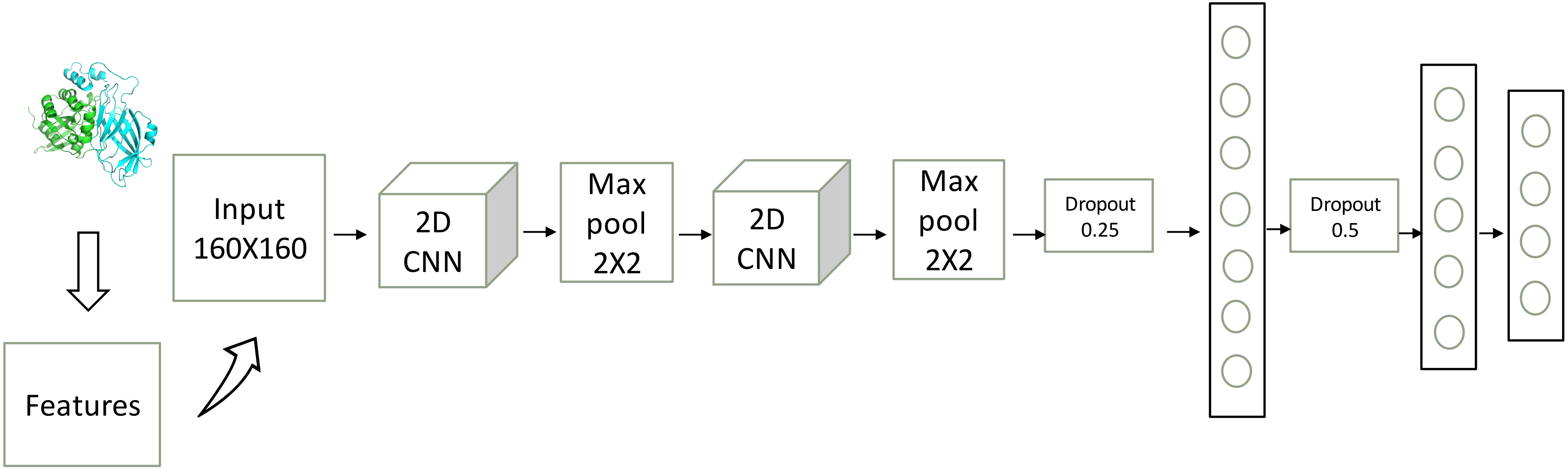
Dropout (0.25)

Fully connected layer (256)

Dropout (0.5)

Output layer (3 for translation and 6 for rotation)

Architecture 2



Input and Training

Dimension: 160x160

Each pixel (i, j)

- Feature of residue i from chain 1
- Feature of residue j from chain 2
- Concatenate the features of i and j

Training

- Learning rate: 0.001
- Optimize MSE loss

Experiments

Exp1

- Architecture 1
- Epochs: 500
- Batch size: 100

Exp2

- Architecture 2
- Epochs: 500
- Batch size: 100

Exp3

- Translation from Exp1
- Rotation from Exp2

Experiments

Exp4

- Resnet18
- Epochs: 35

Exp5

- Translation from Exp4
- Rotation (Exp4; stops training if loss<0.2 and at least 20 epochs)

Result

Experiment	Average TM score before	Average TM score after
Exp1	0.52258	0.4991
Exp2	0.52258	0.4957
Exp3	0.52258	0.4986
Exp4	0.52258	0.4965
Exp5	0.52258	0.4972

Result

Experiment	TM increased	TM decreased	TM same
Exp1	136	276	68
Exp2	107	301	72
Exp3	103	303	74
Exp4	101	300	79
Exp5	104	301	75

Result

Significant improve -> TM is increased by at least 0.04

Experiment	#Significant Improve (out of 480)
Exp1	75
Exp2	68
Exp3	70
Exp4	70
Exp5	69

Result

Experiment	Average lDDT before	Average lDDT after
Exp1	0.678	0.669
Exp2	0.678	0.66784
Exp3	0.678	0.66785
Exp4	0.678	0.668
Exp5	0.678	0.66783

Result

Experiment	lDDT increased	lDDT decreased	lDDT same
Exp1	48	386	46
Exp2	2	420	58
Exp3	0	422	58
Exp4	11	411	58
Exp5	1	422	57

Result

Significant decrease -> Decrease by at least 0.05

Experiment	#Significant Decrease (out of 480)
Exp1	22
Exp2	23
Exp3	22
Exp4	23
Exp5	23

Findings

Refinement is sensitive to the prediction of sign

Chain 2 was taken far away in many of the complex structures

Convolutional neural networks may not work

Many of the structures had the same TM score after the transformation, because only chain 1 was considered during the alignment and chain 2 was ignored (chain 2 was too far to be aligned)

A few number of structures had significant improvement and therefore, the proposed method has potential to work better

Future work

Graph Neural Network (on progress)

More decoys for training and testing

Flexible refinement to capture conformational changes

Questions