

* "Название": Deep Learning for reliable detection of tandem repeats in 3D protein structures

* "Задача": Описание или постановка задачи. Желательна постановка в виде задачи оптимизации (в формате argmin). Также возможна ссылка на классическую постановку задачи.

Deep learning algorithms pushed computer vision to a level of accuracy comparable or higher than a human vision. Similarly, we believe that it is possible to recognize the **symmetry of a 3D** object with a very high reliability, when the object is represented as a density map.

The optimization problem includes

- i) multiclass classification of 3D data. The output is the order of symmetry. The number of classes is ~10-20
- ii) multioutput regression of 3D data. The output is the symmetry axis (a 3-vector).

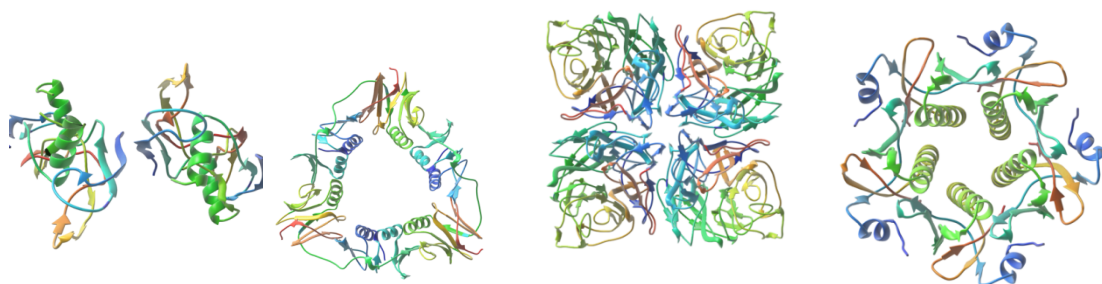
The input data are typically 24x24x24 meshes. The total amount of these meshes is of order a million.

Biological motivation : Symmetry is an important feature of protein tertiary and quaternary structures that has been associated with protein folding, function, evolution, and stability. Its emergence and ensuing prevalence has been attributed to gene duplications, fusion events, and subsequent evolutionary drift in sequence. Methods to detect these symmetries exist, either based on the structure or the sequence of the proteins, however, we believe that they can be vastly improved.

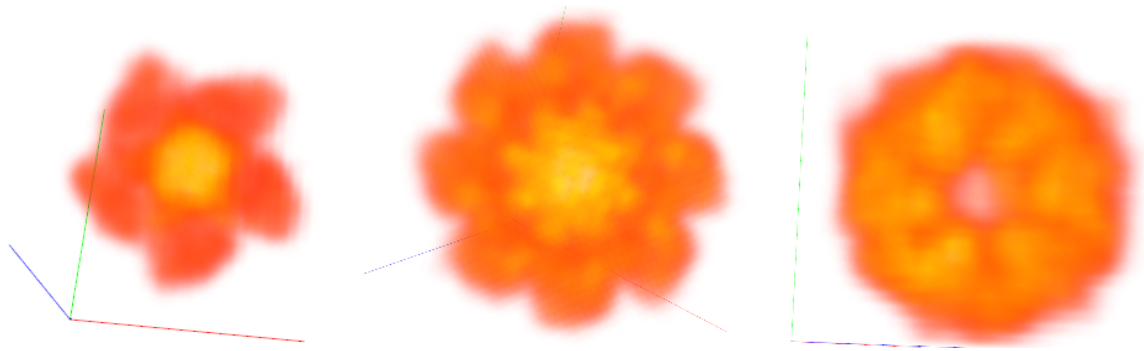
* "Данные": Краткое описание данных, используемых в вычислительном эксперименте, и ссылка на выборку.

Synthetic data are obtained by 'symmetrizing' folds from top8000 library (<http://kinemage.biochem.duke.edu/databases/top8000.php>).

These look like



Examples of symmetric protein structures



Density maps generated from other symmetric protein structures.

* "Литература": Список научных работ, дополненный 1) формулировкой решаемой задачи, 2) ссылками на новые результаты, 3) основной информацией об исследуемой проблеме.

Our previous 3D CNN:

<https://arxiv.org/abs/1801.06252>

Invariance of CNNs (and references therein):

<https://hal.inria.fr/hal-01630265/document>

<https://arxiv.org/pdf/1706.03078.pdf>

Classical methods:

[1] Myers-Turnbull, D., Bliven, S. E., Rose, P. W., Aziz, Z. K., Youkharibache, P., Bourne, P. E., & Prlić, A. (2014). Systematic detection of internal symmetry in proteins using CE-Symm. *Journal of molecular biology*, 426(11), 2255-2268.

[2] Tai, C. H., Paul, R., Kc, D., Shilling, J. D., & Lee, B. (2014). SymD webserver: a platform for detecting internally symmetric protein structures. *Nucleic acids research*, 42(W1), W296-W300.

[3] Abraham, A. L., Rocha, E. P., & Pothier, J. (2008). Swelfe: a detector of internal repeats in sequences and structures. *Bioinformatics*, 24(13), 1536-1537.

* "Базовой алгоритм": Ссылка на алгоритм, с которым проводится сравнение или на ближайшую по теме работу.

A prototype has already been created using the Tensorflow framework [4], which is capable to detect the order of cyclic structures with about 93% accuracy. The main goal of this internship is to optimize the topology of the current neural network prototype and make it rotational and translational invariant with respect to input data.

[4] <https://www.tensorflow.org/>

* "Решение": Предлагаемое решение задачи и способы проведения исследования. Способы представления и визуализации данных и проведения анализа ошибок, анализа качества алгоритма.

The network architecture needs to be modified according to the invariance properties (most importantly, rotational invariance). Please see the links below

<https://hal.inria.fr/hal-01630265/document>

<https://arxiv.org/pdf/1706.03078.pdf>

The code is written using the Tensorflow library, and the current model is trained on a single GPU (Nvidia Quadro 4000) of a desktop machine.

* "Новизна": Обоснование новизны и значимости идей (для редколлегии и рецензентов журнала).

Applications of convolutional networks to 3D data are still very challenging due to large amount of data and specific requirements to the network architecture. More specifically, the models need to be rotationally and transnationally invariant, which makes classical 2D augmentation tricks loosely applicable here. Thus, new models need to be developed for 3D data.

* "Авторы": эксперт : Sergei Grudin, консультант : Guillaume Pages