

Deep Learning for reliable detection of tandem repeats in 3D protein structures.*

Veselova Eugenia¹

veselova.er@phystech.edu

¹Moscow Institute of Physics and Technology (MIPT)

Understanding of protein functions demands feasibility to robustly detect symmetry and structural repetitions in protein structure. This article presents an improved convolutional neural network (CNN) for repetitions and symmetries detection, that is equivariant to rigid input motions — each layer for every $SE(3)$ transformation of 3D input returns output, transformed with the same motion. Equivariance is provided by replacing regular CNN filters with linear combinations of a complete steerable filter basis, consisting of spherical harmonics. This replacement is expected to improve accuracy of structural repetitions detection.

Keywords: *CNN, spherical harmonics, structural repetitions, symmetry axes, 3D objects.*

1 Introduction

Currently deep learning is widely applied in the natural sciences, particularly in the structural biology tasks. Many existing proteins have repeating structural elements or symmetries, which affect proteins properties. Detecting such repetitions and axes of symmetry was solved through conventional machine learning methods [4], hence we are interested in applying automated deep learning approaches, e.g. convolutional neural networks, which contain easily interpreted objects features at their deepest level.

Prediction of three-dimensional protein structure from its amino-acid chain [1] and structural repetitions and internal symmetries detection are efficiently implemented by virtue of convolutional neural networks [6]. However, existing CNNs are not stable and equivariant to rigid input data motions. Therefore we are aimed at adapting already built networks for repetitions and symmetries detection to input transformations.

While translation of the input object after CNN processing returns proportionally translated features map [3], rotation of the input object does not have the same property. Furthermore, in case of persisting equivariance at each level, specific object features may be noticed even at the deepest levels. Equivariant 2D CNN was implemented with replacing CNN regular filters with linear combination of *circular harmonics*, providing rotational equivariance without need in strong data augmentation [8].

Three-dimensional interpretation of such approach are *spherical harmonics*.

Трёхмерной интерпретацией данного подхода являются сферические гармоники. Первоначально идея сферических гармоник была развита в моделировании для эффективного представления и определения степени схожести поверхностей трёхмерных объектов [2, 5]. Далее идея была применена к анализу трёхмерных признаков карт с помощью CNN. При замене стандартных трёхмерных свёрточных фильтров на линейную комбинацию аналитически определённого вращательного базиса из сферических гармоник CNN становится эквивариантна относительно любого преобразования из группы симметрий $SE(3)$ [7].

Любое движение $g \in SE(3)$ представимо как комбинация вращения $r \in SO(3)$ и трансляции $t \in \mathbb{R}^3$. При рассмотрении одного уровня свёрточной нейросети с K трёхмерных признаков карт, соответствие между входом и выходом слоя может быть записано как

*Task stated by: Grudinin S. Consultant: Pages G.

$f : \mathbb{R}^3 \rightarrow \mathbb{R}^K$. Оператор трансляции выходного векторного поля легко описывается как $t : (x - t) \mapsto x$. Вращение описывается более сложным образом, так как при повороте всей каждый вектор меняет свою позицию и поворачивается с помощью матрицы $\rho(r)$. Поэтому оператор вращения $\pi(r)$ определяется как $[\pi(r)f](x) := \rho(r)f(r^{-1}x)$, где $r^{-1}x$ описывает перемещение векторов на новые позиции. Таким образом, $g = tr$ представимо как $[\pi(tr)f](x) := \rho(r)f(r^{-1}(x - t))$. Обобщая полученные выкладки на CNN, выражение свёрточного фильтра между n и $n + 1$ слоями нейросети через базис в пространстве эквивариантных преобразований между пространствами признаков \mathcal{F}_n и \mathcal{F}_{n+1} позволяет гарантировать, что любое преобразование входа слоя будет давать такое же преобразование выхода, т.е. обеспечивать эквивариантность. Кроме того, в силу одинакового изменения всех трёх RGB матриц изображения при рассматриваемых преобразованиях полученное представление может быть перенесено и на цветные изображения.

Именно поэтому в качестве решения поставленной задачи в статье предложена эффективная имплементация сферических гармоник в существующую CNN модель выделения тандемных повторов и симметрий в белках для получения идентичных результатов при любых вращениях исходных карт атомных плотностей белковых 3D моделей [6]. В качестве входных данных выступает синтетический датасет, полученный «симметризацией» белковых структур датасета Top8000*, состоящий из карт плотностей размеров $24 \times 24 \times 24$.

2 Вывод

Литература

- [1] Georgy Derevyanko, Sergei Grudinin, Yoshua Bengio, and Guillaume Lamoureux. Deep convolutional networks for quality assessment of protein folds. *ArXiv e-prints*, 2018.
- [2] Michael Kazhdan and Thomas A. Funkhouser. Harmonic 3d shape matching. In *SIGGRAPH Abstracts and Applications*, page 191. ACM, 2002.
- [3] Karel Lenc and Andrea Vedaldi. Understanding Image Representations by Measuring Their Equivariance and Equivalence. *International Journal of Computer Vision*, 2018.
- [4] Niloy J. Mitra, Leonidas Guibas, and Mark Pauly. Partial and Approximate Symmetry Detection for 3D Geometry. *ACMTG: ACM Transactions on Graphics*, 25, 2006.
- [5] Mohamed-Hamed Mousa, Raphaëlle Chaine, Samir Akkouché, and Eric Galin. Toward an efficient triangle-based spherical harmonics representation of 3d objects. *Computer Aided Geometric Design*, 25(8):561–575, 2008.
- [6] Guillaume Pagès and Sergei Grudinin. DeepSymmetry: Using 3D convolutional networks for identification of tandem repeats and internal symmetries in protein structures. working paper or preprint, 2018.
- [7] Maurice Weiler, Mario Geiger, Max Welling, Wouter Boomsma, and Taco Cohen. 3D Steerable CNNs: Learning Rotationally Equivariant Features in Volumetric Data. *CoRR*, abs/1807.02547, 2018.
- [8] Daniel E. Worrall, Stephan J. Garbin, Daniyar Turmukhambetov, and Gabriel J. Brostow. Harmonic Networks: Deep Translation and Rotation Equivariance. In *CVPR*, pages 7168–7177. IEEE Computer Society, 2017.

*<http://kinemage.biochem.duke.edu/databases/top8000.php>