

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

Веселова Е.Р.¹

veselova.er@phystech.edu

¹Московский физико-технический институт (МФТИ)

В работе рассматривается задача регрессионного выделения осей симметрии трёхмерных объектов и классификационного выделения порядков найденных осей. Обе задачи решаются с помощью применения свёрточных нейросетей к синтетическому датасету, полученному размножением 3D моделей белковых структур. Относительно трансляции данных свёрточные нейросети обладают свойством устойчивости, что не выполняется для вращений. Предлагается применение трёхмерных сферических гармоник вместо классических свёрточных фильтров в CNN. Решение данной задачи позволит увеличить точность обработки и автоматического выделения свойств белков.

Ключевые слова: *CNN, сферические гармоники, ось симметрии, 3D объект*

1 Введение

Машинное обучение широко применяется в задачах современных естественных наук, в частности в задачах структурной биологии. Большая часть получаемых на практике белков обладает повторяющимися элементами структуры или симметрией, которые влияют на функции белков и позволяют исследовать их эволюцию. Нахождение симметрий и повторов является важной задачей, решённой с помощью классических методов машинного обучения в 2006 году [4], поэтому особый интерес представляет применение методов глубокого обучения и свёрточных нейросетей, которые позволяют получать на нижних уровнях сети легко интерпретируемые характерные черты изучаемых объектов.

Предсказание трёхмерной структуры белка по его аминокислотному составу [1] и детектирование тандемных повторов и внутренних симметрий с высокой точностью решается свёрточными нейросетями [2], однако существующие свёрточные нейросети не имеют возможности одинаково качественно обрабатывать входные данные при любых поворотах и сдвигах. Основная цель работы состоит в адаптации построенных нейросетей для выявления повторов и симметрий к различным преобразованиям входных данных.

Если трансляция входного объекта при обработке свёрточной нейросетью даёт пропорционально транслированную карту характеристик [3], то для вращений входного объекта подобное свойство не реализуется. Искомое свойство переноса преобразования входных данных на выходные называется эквивариантностью. Кроме того, сохраняющаяся во всех слоях нейросети эквивариантность позволяет отслеживать свойства исследуемых структур уже на нижних уровнях нейросети. 2D CNN, относительно которой данные были бы эквивариантны, была реализована заменой стандартных свёрточных фильтров на комплексные сферические гармоники, обеспечивающие вращательную эквивариантность без необходимости использовать сильную аугментацию данных [6]. Трёхмерной интерпретацией данного подхода, способной к работе с любым преобразованием из группы симметрий $SE(3)$, стала 3D Steerable CNN, в которой стандартные трёхмерные свёрточные фильтры заменяются линейной комбинацией аналитически определённого вращательного базиса [5].

*Задачу поставил: Grudinin S. Консультант: Pages G.

Более того, в силу одинакового изменения RGB матриц при исследуемых преобразованиях, построенная CNN обрабатывает и цветные изображения, в отличие от предыдущих работ. По этим причинам в статье реализуется именно метод трёхмерного вращательного базиса.

Проводимое исследование нацелено на эффективную имплементацию полученных результатов в существующей модели выделения tandemных повторов и симметрий в белках для получения идентичных результатов при любых вращениях исходных карт атомных плотностей белковых 3D моделей [2]. В качестве входных данных выступает синтетический датасет, полученный «симметризацией» белковых структур из датасета Top8000*.

2 Вывод

Литература

- [1] Derevyanko Georgy, Grudinin Sergei, Bengio Yoshua, and Lamoureux Guillaume. Deep convolutional networks for quality assessment of protein folds. *ArXiv e-prints*, January 2018.
- [2] Pagès Guillaume and Grudinin Sergei. DeepSymmetry : Using 3D convolutional networks for identification of tandem repeats and internal symmetries in protein structures. working paper or preprint, October 2018.
- [3] K. Lenc and A. Vedaldi. Understanding image representations by measuring their equivariance and equivalence. *International Journal of Computer Vision*, 2018.
- [4] Mitra, Guibas, and Pauly. Partial and approximate symmetry detection for 3D geometry. *ACMTG: ACM Transactions on Graphics*, 25, 2006.
- [5] 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.
- [6] 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>