

**Использование модулей, основанных на преобразовании Фурье, в архитектурах нейронных сетей для предсказания связывания транскрипционных факторов**

***Вяльцев Валерий Владимирович***

*Студент (специалист)*

Московский государственный университет имени М.В.Ломоносова, Факультет биоинженерии и биоинформатики, Москва, Россия

*E-mail: vyaltsevaleriy7@gmail.com*

Транскрипционные факторы обладают способностью связываться с ДНК и изменять уровень экспрессии генов, благодаря чему они играют важную роль в процессах онтогенеза, ответа на внеклеточные сигналы и изменение условий окружающей среды. Поэтому изменение возможности связывания различных участков ДНК соответствующими транскрипционными факторами, может привести к значительным изменениям в работе клетки, в том числе повышать риск некоторых заболеваний. В связи с этим особую важность представляет возможность предсказывать эффект различных мутаций, например, таких как SNP, инсерций и делеций, на связываемость транскрипционных факторов в данном участке ДНК.

Одним из методов анализа взаимодействия транскрипционных факторов с ДНК является ещё не опубликованный метод GHT-SELEX, принцип работы которого напоминает метод эволюции лигандов экспоненциальным обогащением HT-SELEX [2]. На основе предоставленных консорциумом Codebook и GRECO-BIT данных экспериментов геномного HT-SELEX для различных транскрипционных факторов мы хотим разработать нейросетевую архитектуру, которая, при обучении на данных эксперимента для некоторого фактора, эффективно предсказывает его связываемость с ДНК.

Нейросетевые модели, имеющие искомую архитектуру, должны отличать участки, в которых фактор может связаться с ДНК, от участков, где он не связывается, то есть должны решать задачу бинарной классификации, поэтому на основе исходных данных - позитивного класса - путем сдвига был получен негативный класс. Сперва мы проверили ряд простых архитектур, основанных на сверточных нейронных сетях. В частности, была использована широко применяемая в анализе изображений архитектура ResNet с inverted residual block, в которой присутствуют так называемые skip connections, позволяющие бороться с проблемой исчезающих градиентов [1]. Исследуемые архитектуры на разных транскрипционных факторах показали приемлемое качество, однако, с целью улучшить результаты предсказания мы рассмотрели другой вариант архитектуры.

Для дальнейшего исследования была выбрана сеть, основанная на архитектуре модели LaMa [3]. Особенностью данной архитектуры является одновременное использование небольших ядер свертки и специальных блоков, использующие преобразование Фурье, что позволяет модели использовать не только локальную информацию, но информацию о всей последовательности разом (Рис.). В результате экспериментов с адаптированной под особенности задачи архитектурой удалось получить качество предсказания выше, чем у ранее упомянутых простых сверточных сетей и ResNet для большинства исследуемых факторов.

**Источники и литература**

- 1) 1. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, “Deep Residual Learning for Image Recognition”, arXiv:1512.03385 [cs.CV], Dec. 2015
- 2) 2. Ogawa, N., & Biggin, M. D. (2011). High-Throughput SELEX Determination of DNA Sequences Bound by Transcription Factors In Vitro. Gene Regulatory Networks, 51–63. doi:10.1007/978-1-61779-292-2\_3
- 3) 3. Roman Suvorov, Elizaveta Logacheva, Anton Mashikhin, Anastasia Remizova, Arsenii Ashukha, Aleksei Silvestrov, Naejin Kong, Harshith Goka, Kiwoong Park, Victor Lempitsky, “Resolution-robust Large Mask Inpainting with Fourier Convolutions”, arXiv:2109.07161 [cs.CV], Sep. 2021

### Иллюстрации

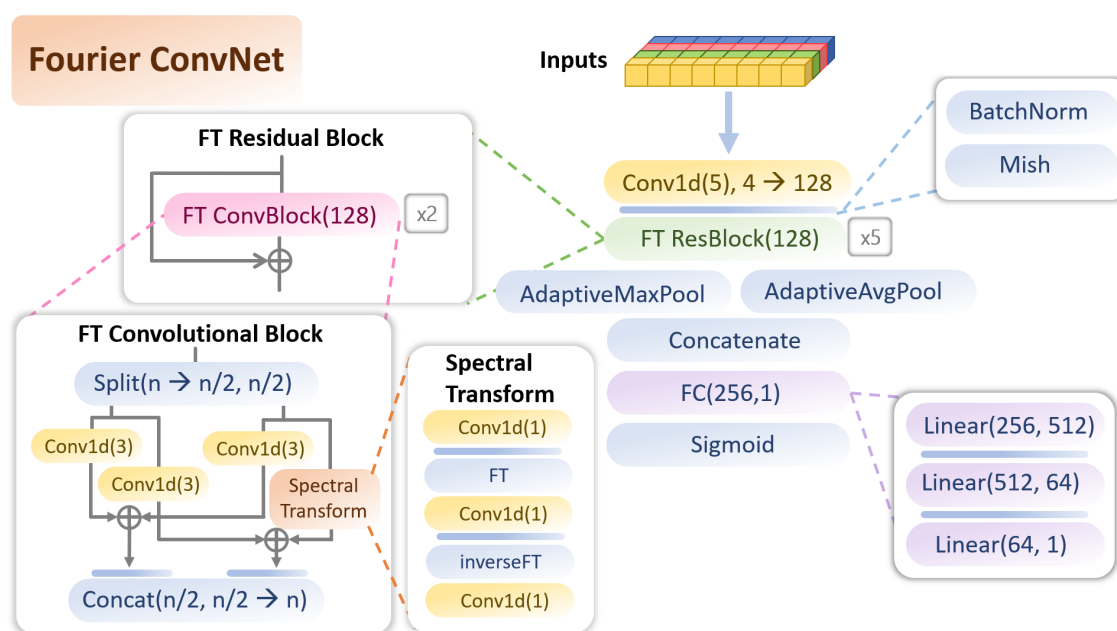


Рис. : Адаптированная архитектура модели из статьи [2].