

**Сравнение классических подходов к автоматической сегментации клеток с применением алгоритмов глубокого обучения на данных иммунофлюоресценции при помощи различных метрик качества**

**Научный руководитель – Галкин Илья Николаевич**

*Дымов Даниил Антонович*

*Студент (бакалавр)*

Московский физико-технический институт, Москва, Россия

*E-mail: dymov.da@phystech.edu*

Методы секвенирования нового поколения (англ. Next Generation Sequencing – NGS) и проточной цитометрии позволили эффективно изучать клеточный состав тканей, индивидуальные особенности клеток и организацию клеточных процессов, но интерес представляет еще и пространственная организация клеток, что особенно важно для анализа опухолевого микроокружения. Мультиплексная иммунофлюоресценция (англ. Multiplexed Immunofluorescence – MxIF), в свою очередь, как раз позволяет получать данные о пространственном распределении белков в клетках ткани *in situ*. Анализ изображений, полученных при помощи иммунофлюоресценции, предоставляет возможность одновременного получения информации о расположении клеток и экспрессии различных маркеров в них, что открывает новые возможности для изучения тканей и межклеточных взаимодействий. Качество такого анализа напрямую зависит от качества сегментации, а именно от точности определения контуров клеток на слайде.

Способ иммунофлюоресцентного окрашивания гистологических препаратов, реализованный в технологии CODEX[1] от компании Akoya, позволяет получать данные до 60 маркеров на препарат и проводить последующий анализ при помощи программного обеспечения, поставляемого вместе с аппаратом. Akoya предоставляет наборы подобных данных в открытом доступе, включающие флуоресцентные изображения, маски клеточной сегментации и таблицы параметров для каждой найденной клетки. В качестве способа сегментации клеток на наборе флуоресцентных изображений, CODEX предлагает подход, сходный с реализованным в программе CellProfiler[2]. Недостатками такого подхода являются:

- Высокая чувствительность к “шуму” на изображении;
- Точность воспроизведения контуров клеток по мембранным маркерам;
- Точность количественного воспроизведения клеток в областях с высокой плотностью или для клеток с ядрами, отличающимися от округлой формы.

В поиске возможных решений описанных выше проблем сегментации, авторами были применены методы глубокого обучения, а именно автоэнкодеры с Unet-подобной архитектурой[3] и алгоритмы для нахождения клеток как отдельных объектов на изображении [4]. Названные выше подходы были протестированы на открытом датасете CODEX. Второй задачей было сравнение различных метрик качества сегментации. Чтобы сравнить маски, полученные при помощи методов глубокого обучения, с масками, полученными при помощи программного обеспечения CODEX, была использована маска клеток, сделанная вручную гистопатологом.

В задаче оценки качества авторами были выделены следующие, наиболее существенные критерии:

- Точность количественного воспроизведения клеток на слайде;
- Точность воспроизведения контуров клеток;
- Количество необнаруженных и ложно-определенных клеток в предсказании.

В соответствии с приведенными критериями в ходе литературного анализа из открытых источников были выбраны метрики (Panoptic Quality[5], Hausdorff Distance[6], Adjusted Rand Score[6]) и впоследствии имплементированы, наряду с собственными разработками.

Была доказана состоятельность введенных метрик качества сегментации, затем при помощи этих метрик было продемонстрировано, что методы глубокого обучения показывают более высокое качество воспроизведения клеточных контуров. При этом, согласно результатам, изъясном подходе семантической сегментации является необходимость постобработки клеточной маски для дополнительного разделения клеток, вследствие которой в результат может вноситься существенная ошибка пересегментации. Было доказательно показано, что наиболее высокий по качеству результат в соответствии с указанными критериями качества сегментации (т.е. наиболее близкий к разметке, сделанной человеком) продемонстрировал подход предсказания клеток как отдельных объектов.

### Источники и литература

- 1) Goltsev Y. [et al.]. Deep Profiling of Mouse Splenic Architecture with CODEX Multiplexed Imaging // Cell. 2018. V. 174. N. 4. P. 968–981
- 2) Carpenter A.E., Jones T.R., Lamprecht M.R. [et al.]. CellProfiler: image analysis software for identifying and quantifying cell phenotypes // Genome Biol. 2006. V. 7. N. 10. R100.
- 3) Ronneberger O., Fischer P., Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation // Lecture Notes in Computer Science. 2015. V. 9351. P. 234-241
- 4) K. He, G. Gkioxari, P. Dollár and R. Girshick, Mask R-CNN // IEEE International Conference on Computer Vision (ICCV) 2017. P. 2980-2988
- 5) Kirillov A., He K., Girshick R., Rother C., Dollár P. Panoptic Segmentation // IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 2019. P. 9396-9405
- 6) Beneš M, Zitová B. Performance evaluation of image segmentation algorithms on microscopic image data // J Microsc. 2015. V. 257. N. 1. P. 65-85.