

ОБРАБОТКА ИЗОБРАЖЕНИЙ С БИОЛОГИЧЕСКИМИ КЛЕТКАМИ

Н. П. Жемойтяк¹⁾, О. А. Лаврова²⁾

¹⁾ Белорусский государственный университет,
пр. Независимости, 4, 220030, г. Минск, Беларусь, natalliazhamaitsiak@gmail.com

²⁾ Белорусский государственный университет,
пр. Независимости, 4, 220030, г. Минск, Беларусь, lavrovaao@bsu.by
Научный руководитель: О. А. Лаврова, кандидат физико-математических наук

Целью данного исследования является обработка изображений для выделения биологических клеток и последующего анализа их геометрических характеристик. В работе представлены способы сегментации контуров клеток с помощью статистических методов библиотеки OpenCV, а также с глубокими сверточными сетями. В качестве геометрических характеристик для классификации клеток на больные и здоровые рассматриваются площадь клетки и ее выпуклость.

Ключевые слова: Компьютерное зрение; классификация; сегментация; биологические клетки.

Изображения биологических клеток являются важным инструментом для понимания клеточной биологии, диагностики заболеваний и разработки методов лечения. Однако, анализ этих изображений представляет собой сложную задачу, требующую передовых методов обработки изображений и анализа данных. С развитием медицинской технологии объем и разнообразие доступных изображений клеток значительно увеличился, что создает потребность в эффективных методах автоматизированного анализа и интерпретации этих данных. В данной статье описан подход к анализу изображений с биологическими клетками с помощью сегментации клеток и вычисления их геометрических характеристик.



Рис. 1. Пример изображения с биологическими клетками [1]

Сегментация клеток

Сегментация изображений представляет собой процесс разделения изображения на отдельные части или объекты. В контексте биологических клеток, сегментация позволяет выделить каждую клетку на изображении, что является важным шагом для дальнейшего анализа и классификации.

Один из подходов к сегментации клеток включает использование библиотеки компьютерного зрения OpenCV. В этом методе обычно применяются классические алгоритмы обработки

изображений, такие как фильтры, пороговая обработка и операции морфологического преобразования. Например, для сегментации клеток можно использовать адаптивную пороговую обработку для выделения контуров клеток, а затем применить морфологические операции для удаления шума и объединения близко расположенных контуров. На рис. 2 видно, что подход OpenCV не самый подходящий, потому что выделяет шум и не разграничивает клетки.

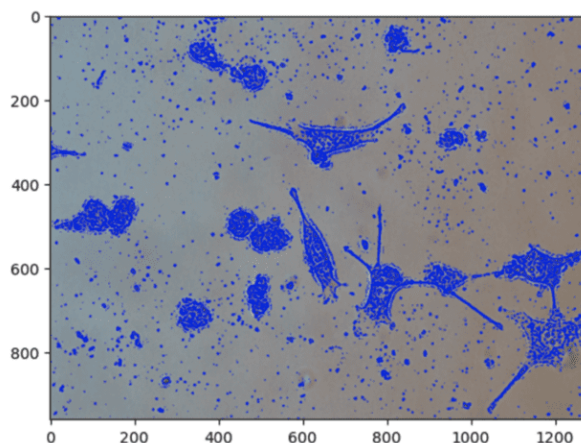


Рис. 2. Результат сегментации с помощью библиотеки OpenCV[4]

Другой подход к сегментации клеток основан на использовании нейронных сетей. Современные глубокие нейронные сети, такие как сверточные нейронные сети (CNN), демонстрируют высокую эффективность в сегментации изображений. В этом подходе сначала происходит обучение нейронной сети на большом наборе размеченных изображений, где каждая клетка на изображении имеет свой сегментированный контур. Затем обученная нейронная сеть может быть применена к новым изображениям для автоматической сегментации клеток.

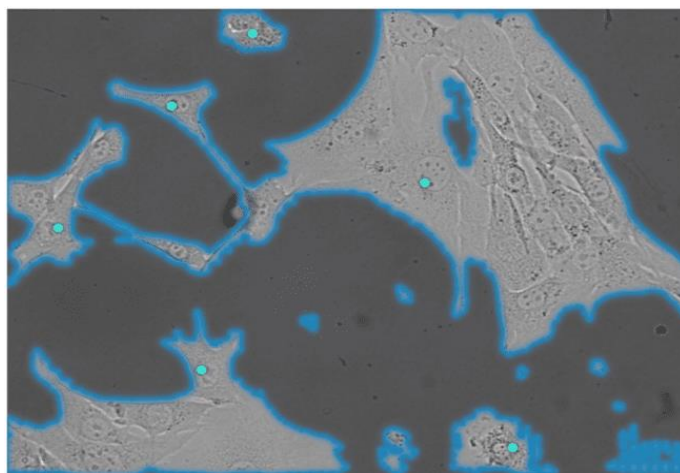


Рис. 3. Результат сегментации с помощью модели Segment Anything(SAM)[5]

Классификация клеток методом анализа их геометрических характеристик

Задача классификации заключается в присвоении метки целевому объекту. В данном случае целевым объектом является биологическая клетка, которая может быть больной или здоровой, см. рис. 3.

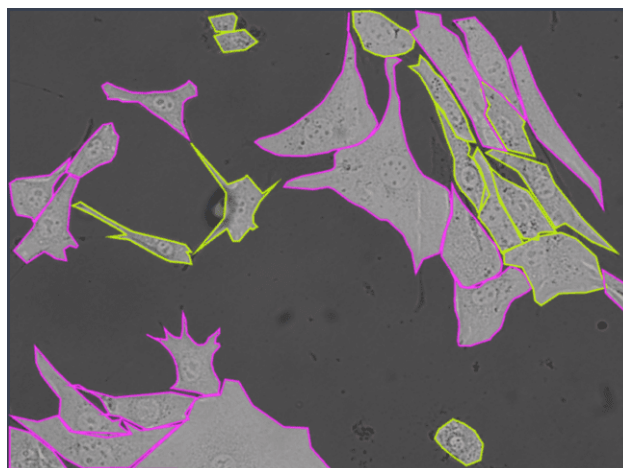


Рис. 4. Больные клетки обозначены желтым контуром, здоровые – розовым.

На рис. 4 видно, что клетки различаются по геометрическим характеристикам. Некоторые из них больше остальных по площади, некоторые же более выпуклые по форме. Исследуем несколько базовых геометрических характеристик и посмотрим на распределение значений характеристик по двум классам больных и здоровых клеток. Клетки каждого класса были выделены на изображениях вручную на основании экспертных знаний биологов [1].

Вычисление площади клетки

Расчет площади клетки осуществляется с помощью построения триангуляции многоугольной области клетки с последующим суммированием площадей всех треугольников сетки. В статье [2] было описано, как использовать библиотеку MeshPy на Python [3] для разбиения трехмерных областей на тетраэдры. Для двумерных областей будем использовать похожий подход. Пример триангуляции области клетки представлен на рис. 5.

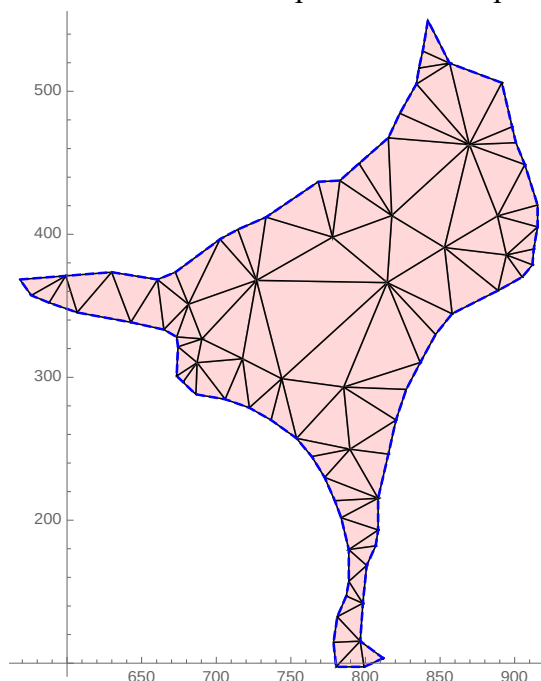


Рис. 5. Триангуляция области клетки

Классификация клеток по площади

Первая геометрическая характеристика, которую рассмотрим для классификации клеток на больные и здоровые, это площадь клетки. Вычислим площадь всех клеток и построим гистограмму распределения больных и здоровых клеток, см. рис. 6. По распределению площади на рис. 6 видно, что больные клетки, в основном, меньше по площади, чем здоровые клетки. При этом есть диапазоны совпадения площади больных и здоровых клеток. Можно сделать вывод, что площадь, как одна из геометрических характеристик клетки, не является достаточным критерием для разделения клеток на два класса.

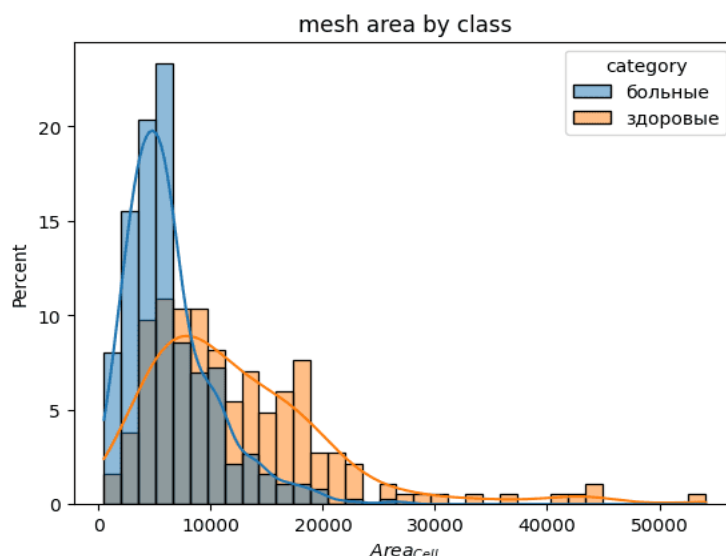


Рис. 6. Гистограмма распределения площади больных и здоровых клеток

Классификация клеток по выпуклости

Еще одна геометрическая характеристика для классификации клеток на больные и здоровые — это оценка выпуклости формы клетки.

Введем оценку выпуклости по следующей формуле:

$$\text{convexity} = \frac{\text{Area}_{\text{Cell}}}{\text{Area}_{\text{ConvexHull}}}$$

где $\text{Area}_{\text{Cell}}$ — это площадь клетки, а $\text{Area}_{\text{ConvexHull}}$ — площадь выпуклой оболочки для области клетки. Если клетка представлена выпуклой областью, то ее площадь будет совпадать с площадью выпуклой оболочки и оценка выпуклости будет равна 1. В противном случае оценка выпуклости принимает значения из промежутка (0,1). Для вычисления площади выпуклой оболочки воспользуемся функцией `ConvexHull` из библиотеки `opencv` на Python [4]. Эта функция возвращает список вершин, которые задают оболочку. Очевидно, что выпуклая оболочка является многоугольником и поэтому ее площадь можно посчитать с помощью триангуляции области.

Гистограмма на рис. 7 построена с использованием данных о площади клеток из рис. 6. По гистограмме на рис. 7 видно, что большое количество больных клеток является выпуклым, тогда как здоровые клетки, в основном, не являются выпуклыми. Однако такое различие в распределении больных и здоровых клеток не является достаточным для четкого разделения всех клеток на два класса.

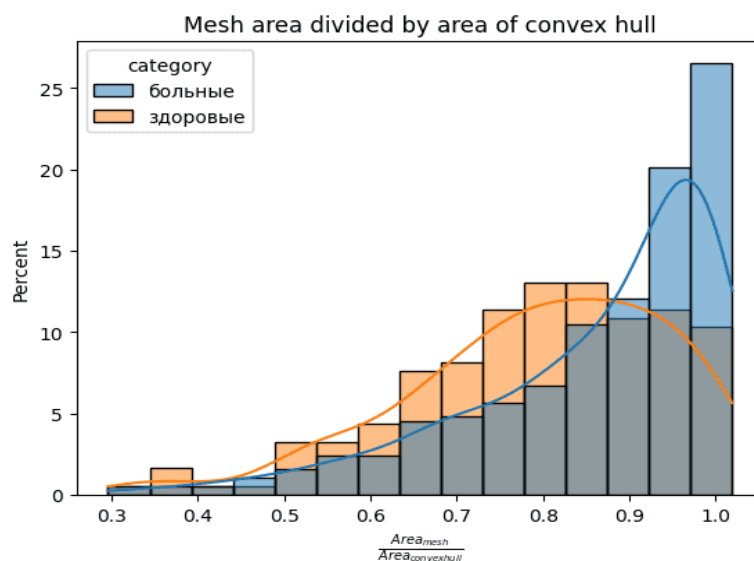


Рис. 7. Гистограмма распределения выпуклости больных и здоровых клеток

Библиографические ссылки

1. Mathematical prediction of polydopamine-coated silica-embedded SPIONs biocompatibility / V. Goranov [et al.] // Pisen: EUICЗBCB, 2023. 4 p. (принято к опубликованию).
2. Жемойтяк Н. Математическое моделирования процессов цитотоксичности клеток / Дипломная работа, Минск, 2023.
3. MeshPy Documentation [Электронный ресурс]. URL: <https://documen.tician.de/meshpy/> (дата обращения: 05.12.2023).
4. OpenCV Documentation [Электронный ресурс]. URL: <https://docs.opencv.org/4.x/index.html/> (дата обращения: 05.12.2023).
5. Segment Anything / A Kirillov [et al.] // New York: Meta AI Research, 2023. 30 p.