

ИСПОЛЬЗОВАНИЕ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ВЫДЕЛЕНИЯ КОМЕТ МОЛЕКУЛ ДНК

С. В. Кезик ¹⁾, Т. А. Кабулов ²⁾

¹⁾ Белорусский государственный университет, пр. Независимости, 4, 2203030, г. Минск, Беларусь, email: sergeyk8008@gmail.com

²⁾ Белорусский государственный университет, пр. Независимости, 4, 220030, г. Минск, Беларусь

Научный руководитель — О.В. Недзьведь, кандидат технических наук, доцент

В данной статье представлены результаты автоматизации анализа изображений, получаемых при помощи техники ДНК-комет.

Ключевые слова: компьютерное зрение; глубокое обучение; семантическая сегментация; сверточная нейронная сеть; метод ДНК-комет.

Стрессовые воздействия различной природы могут приводить к повреждению генетической информации, хранящейся в ядре клетки. Одним из методов исследования этих повреждений является метод ДНК-комет, который представляет собой метод регистрации повреждения ДНК на уровне отдельных клеток [1-3]. Данный процесс является очень трудоемким, так как одновременно необходимо производить анализ сотен комет. В этой связи автоматизация анализа ДНК-комет представляется актуальной и важной проблемой.

Исходный набор данных состоял из 1260 изображений. Типичное изображение из набора представлено на рис. 1. Изображение включает многочисленные шумы, остатки препаратов.

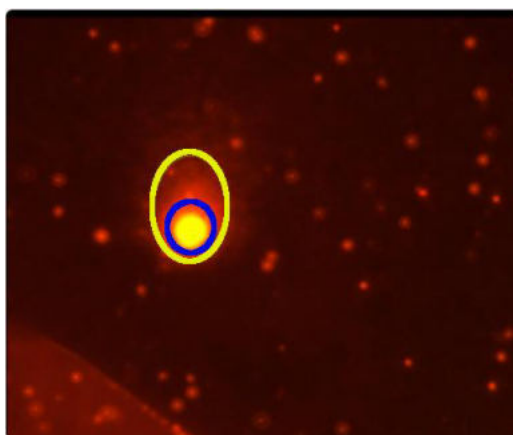


Рис. 1. Пример изображения ДНК кометы: синим цветом выделена голова кометы, желтым – хвост

Для этого к изображениям применяется серия преобразований (рис. 2). Для начало каждое изображение переводится из трехканального RGB формата в одноканальный формат оттенков серого (grayscale). Далее все изображения нормализуются с целью первичного уменьшения шума, после чего формируется среднее по изображению, которое вычитается из исходного изображения. Это позволяет существенно снизить уровень шума и повысить четкость изображений комет.

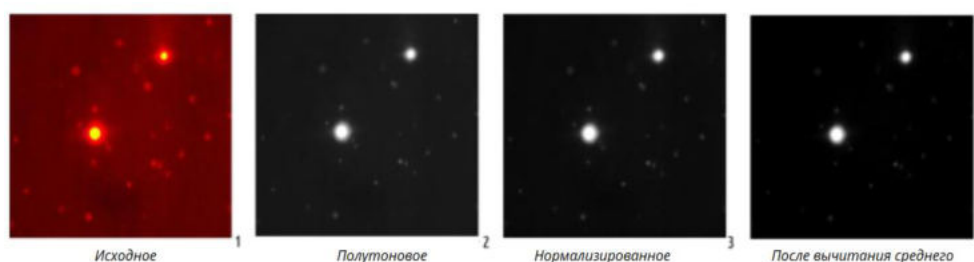


Рис. 2. Этапы предобработки изображений

Затем все изображения аннотируются вручную для максимизации качества сегментации. После аннотирования размер всех изображений уменьшается до 512*512 пикселей. На завершающем этапе набор данных разбивается на тренировочную, валидационную и тестовую выборки в соотношении 80%-10%-10%.

Для повышения точности модели и улучшения вариативности данных к тренировочному набору данных применяются различные методы аугментации изображений: шумы разных типов, включая гауссовский и мультипликативный, повороты на 30, 45, 60 и 90 градусов. Кроме того, выполняется увеличение резкости, а также корректировка контраста и яркости. Для уменьшения количества ложноположительных срабатываний модели сгенерировано 600 “шумных” изображений, не содержащих кометы. Итоговые наборы данных представлены в табл.1.

Таблица 1

Наборы данных, используемые в работе

	Оригинал	Оригинал, шум	Оригинал, шум, аугментация
Количество изображений	1260	1260+600	1260+600+600

Для тренировки использовалась модель PSPnet (Pyramid Scene Parsing Network), так как она показала себя лучше всего по сравнению с другими моделями (Feature Pyramid Network, U-net). Тренировка проходила со

следующими гиперпараметрами: количество эпох — 50, коэффициент обучения — 0.0001, размер батча 8, оптимизатор Adam. Для оценки качества работы нейронной сети использовалась метрика Дайса (Dice score).

В результате обучения модели были получены результаты, приведенные в табл. 2 в виде значений метрики для соответствующих выборок.

Таблица 2

Результаты тестирования

Набор данных	Оригинал	Оригинал + шум	Оригинал + аугментация	Оригинал + аугментация + шум
Валидация	0,855	0,93	0,871	0,851
Тестовый	0,75	0,808	0,838	0,863

Итоговый результат виден на рис. 3, где синим выделен хвост, а желтым ядро кометы.

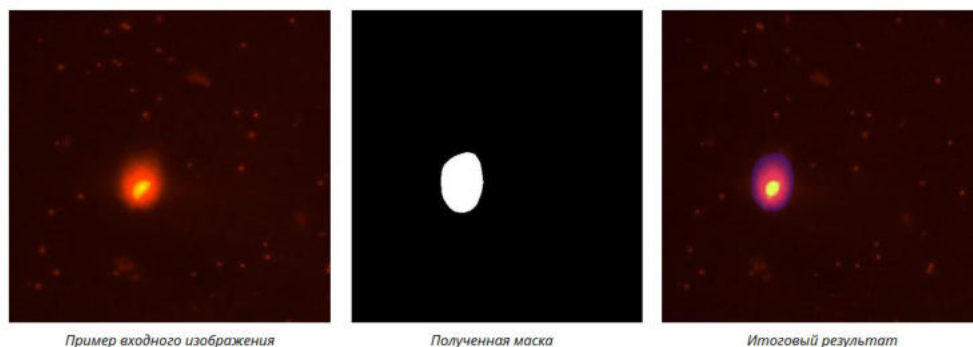


Рис. 3. Результат полной работы программы: синим выделен хвост, а желтым ядро кометы

После получения маски на выходе из модели с неё удаляются все элементы, размер которых меньше 200 пикселей, на исходном изображении выделяется область, соответствующая маске и для неё применяется градиентный фильтр для выделения ядра.

После получение маски головы и хвоста становится возможен автоматический анализ комет. Для этого для каждого изображения происходит расчет характеристик, таких как яркость и размер головы, яркость и размер хвоста, а также процент ДНК в хвосте по формуле.

$$\text{ДНК}_{\text{хвост}} = \frac{\text{Площадь}_{\text{хвост}}}{\text{Площадь}_{\text{комета}}}$$

Площадь кометы и хвоста здесь рассчитывается, как число пикселей в соответствующей области. Увеличение площади хвоста приводит к общему увеличению значения формулы, что отражает влияние того или иного стрессового фактора на повреждение ДНК.

Таким образом, предложенная нами модель позволяет не только значительно упростить обработку изображений, но и открывает путь для интеграции с современными микроскопическими системами, обеспечивая автоматический анализ изображений.

Библиографические ссылки

1. *Liao W., McNutt M., Zhu W.-G.* The comet assay: A sensitive method for detecting DNA damage in individual cells // *Methods*. 2009. Vol. 48. P. 46-53.
2. *Voulodimos A., Doulamis N.* Deep Learning for Computer Vision: A Brief Review // *Computational Intelligence and Neuroscience*. 2018. Vol.18. P.129-135.
3. *Ubbens J. R., Stavness I.* Deep Plant Phenomics: A Deep Learning Platform for Complex Plant Phenotyping Tasks // *Frontiers in Plant Science*. 2017. Vol. 8. P. 96-105.