

ИСПОЛЬЗОВАНИЕ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ВЫДЕЛЕНИЯ КОМЕТ МОЛЕКУЛ ДНК

С. В. КЕЗИК, Т. А. КАБУЛОВ,
канд. техн. наук, доц. О. В. НЕДЗЬВЕДЬ, С. Н. ЗВОНАРЕВ,
д-р биол. наук, проф. В. В. ДЕМИДЧИК
(Белорусский государственный университет, г. Минск)

Аннотация. 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]. При помещении в электрическое поле фрагменты молекул ДНК вытягиваются по направлению поля, визуально становясь похожими на кометы (рисунок 1). Анализ изображений флуоресцентно-окрашенных ядер включает в себя выделение комет и их дифференциацию, на основе которой делаются выводы о степени поврежденности ДНК. Такой вид исследований является очень трудоемким, так как одновременно необходимо производить анализ большого количества комет [2]. Использование нейронных сетей для распознавания и анализа ДНК-комет позволяет автоматизировать этот процесс.

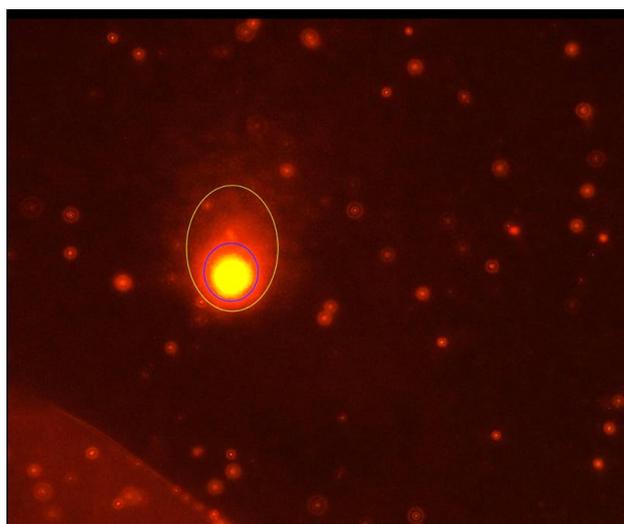


Рисунок 1. – Пример изображения ДНК кометы
(синим цветом выделена голова кометы, желтым – хвост)

Набор данных для обучения и контроля состоит из 1600 исходных изображений, из них 200 изображений для тестовой выборки, 140 – для валидации (таблица 1). Для улучшения качества исходных изображений перед их использованием в тренировке была применена серия обработок. Цель этих преобразований – снижение уровня фонового шума и увеличение резкости изображений комет. В первую очередь, цветные изображения были преобразованы в полутоновые изображения (рисунок 2). Затем была выполнена нормализация изображений, которая заключается в изменении диапазона значений пикселей для улучшения визуального восприятия и анализа. Затем было сделано вычитание среднего значения пикселей из каждого изображения (рисунок 2).

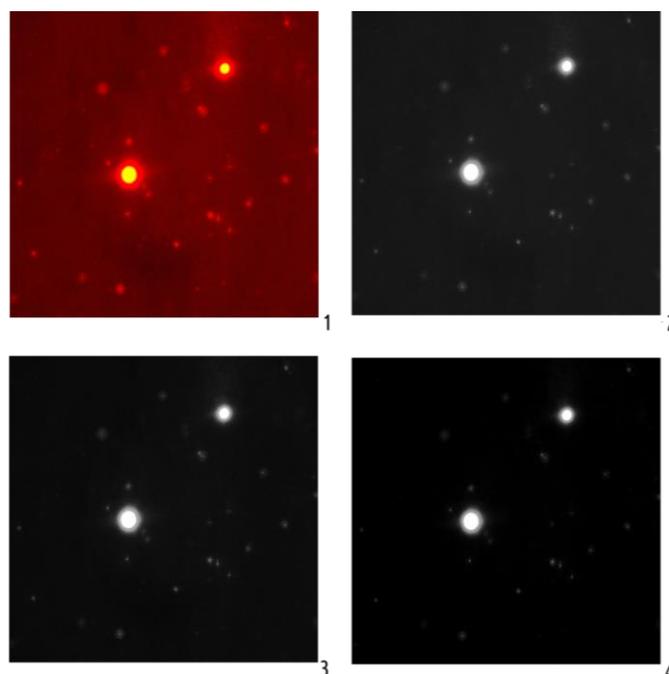


Рисунок 2. – Изображения в наборе данных: исходное изображение (1); изображение, преобразованное в полутоновое (2); изображение после нормализации (3), результат вычитания среднего значения пиксела (4)

Аугментация включала в себя ряд преобразований: добавление шума различных типов, таких как гауссовский шум и мультипликативный шум, повороты на 30, 45, 60 и 90 градусов, увеличение резкости изображений, а также изменения контраста и яркости. Эти манипуляции применялись ко всем изображениям набора данных для создания дополнительных условий обучения. Также было сгенерировано 600 «зашумленных» изображений, не содержащих кометы.

Таблица 1. – Набор данных для обучения

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

Для сегментации изображений использовалась нейронная сеть PSPNet, с энкодером resnext101_32x4d и весами ssl, доступные в библиотеке из открытых источников (https://github.com/qubvel/segmentation_models.pytorch) [3]. Использование в данной архитектуре обходных связей позволило распространить большие градиенты до исходных слоев, смягчая эффект проблемы исчезающих градиентов и улучшая точность глубоких остаточных сетей по сравнению с аналогами. В результате модель учитывает слабые градиенты на изображениях, которые свойственны образованию комет при разрыве клетки. Широкое использование пакетной нормализации (батч-нормализации) по всей структуре модели повышает стабильность сети.

Для обучения модели использовались следующие гиперпараметры: коэффициент скорости обучения 0,001, оптимизатор Adam, размер батча 8, остальные параметры были использованы по умолчанию при работе модели.

Результатом является бинарное изображение с выделенной областью кометы, построенное на основе тепловой карты в последнем слое модели.

Модель тренировалась на трех вариантах набора данных, указанных в таблице 1. Точность моделей, натренированных на четырех наборах данных, на валидационной и тестовой выборке приведены в таблице 2.

Таблица 2. – Точность работы моделей

Набор данных	Оригинал	Оригинал + шум	Оригинал + аугментация	Оригинал + аугментация + шум
Валидация	0.855	0.93	0.851	0.871
Тестовый	0.75	0.808	0.863	0.838

Исходное изображение и результат сегментации представлены на рисунке 3.

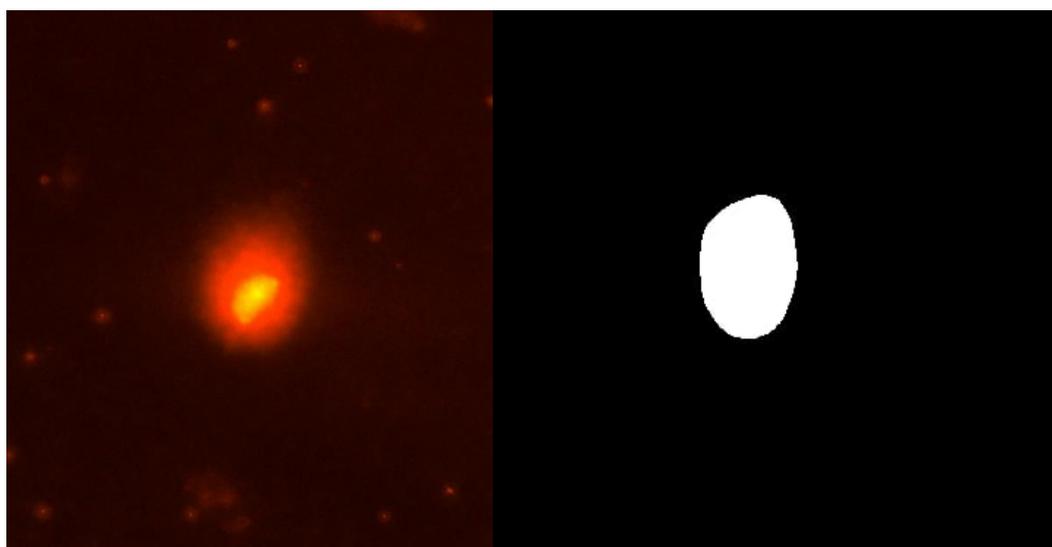


Рисунок 3. – Исходное изображение кометы и результат сегментации

По выделенной в ходе сегментации области рассчитывались следующие параметры: длину хвоста (в пикселях), площадь (в пикселях) и яркость хвоста и головы, процент повреждения ДНК и хвостовой момент, который отражает количества ДНК, перешедшей из ядра в хвост. Эти параметры позволяют характеризовать свойства комет и их текущее состояние.

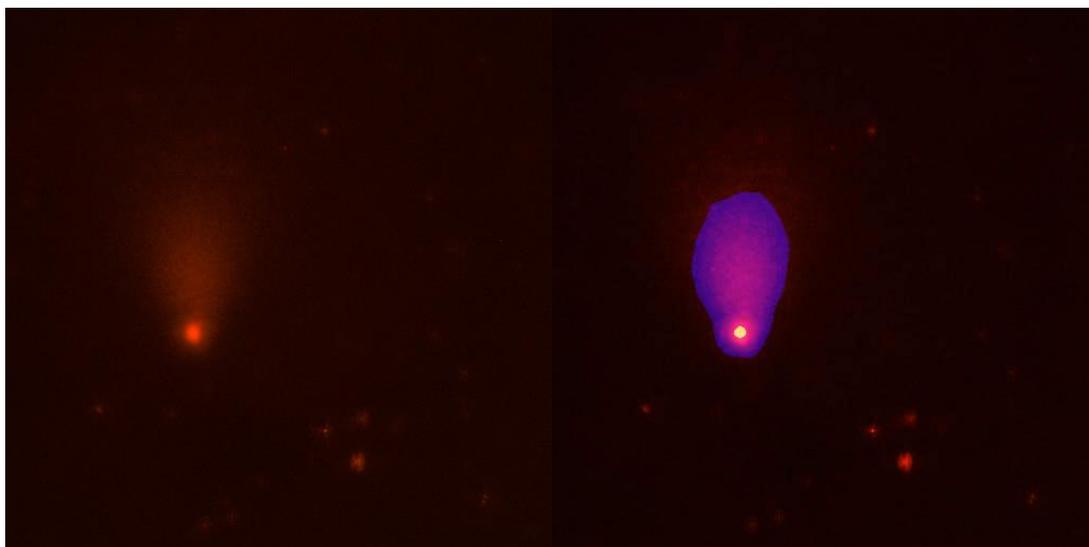


Рисунок 4. – Исходное изображение кометы и результат обработки (синим выделен хвост, желтым голова)

Таблица 3. – Пример подсчитанных параметров для рисунка 4

Параметр	Длина хвоста	Площадь хвоста	Площадь головы	Яркость хвоста	Яркость головы	Хвостовой момент	Процент повреждения ДНК
Значение	618	282465	963	5.77	111.77	615.9	99.66

Предложенная нами модель предоставляет собой инновационный подход к автоматизации анализа визуальных данных, получаемых через метод ДНК-комет. Она позволяет не только значительно упростить обработку изображений, но и открывает путь для интеграции с современными микроскопическими системами, обеспечивая автоматическую аквизицию высококачественных изображений. Кроме того, модель специально адаптирована для работы с изображениями растительных клеток, которые традиционно характеризуются повышенным уровнем шума и менее выраженными кометными хвостами.

ЛИТЕРАТУРА

1. Zvanarou, S. Salt stress triggers generation of oxygen free radicals and DNA breaks in *Physcomitrella patens* protonema/ S. Zvanarou, V. Demidchik et al. // *Environmental and Experimental Botany*. – 2020. – Vol. 180. – <https://doi.org/10.1016/j.envexpbot.2020.104236>.

2. Ubbens, J. R. Deep Plant Phenomics: A Deep Learning Platform for Complex Plant Phenotyping Tasks / J. R. Ubbens, I. Stavness. // *Frontiers in Plant Science*. – Vol. 8. – <https://doi.org/10.3389/fpls.2017.01190>.
3. Zhao, H. Pyramid Scene Parsing Network / H. Zhao et al // *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. – 2017. – P. 6230–6239.