

**Санкт-Петербургский политехнический университет  
Петра Великого  
Институт Компьютерных наук и Технологий**

На правах рукописи

**Сергеев Даниэл Игоревич**

**Разработка автоматизированных методов компьютерного зрения для  
обработки медицинских изображений**

Направление подготовки: 09.06.01\_06 Информатика и вычислительная техника

---

*Код и наименование*

Направленность: 05.13.11 Математическое и программное обеспечение  
вычислительных машин, комплексов и компьютерных сетей

---

*Код и наименование*

**НАУЧНЫЙ ДОКЛАД**

об основных результатах научно-квалификационной работы (диссертации)

Автор работы: Сергеев Даниэл Игоревич  
Научный руководитель: к.т.н, доцент,  
Дробинцев Павел Дмитриевич

Санкт Петербург – 2021

Научно-квалификационная работа выполнена в ВШ/на кафедре Института Компьютерных наук и технологий федерального государственного автономного образовательного учреждения высшего образования «Санкт-Петербургский политехнический университет Петра Великого»

Директор ВШ/зав. кафедрой:

Научный руководитель:

Рецензент:

С научным докладом можно ознакомиться в библиотеке ФГАОУ ВО «Санкт-Петербургский политехнический университет Петра Великого» и на сайте Электронной библиотеки СПбПУ по адресу: <http://elib.spbstu.ru>

## **ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ**

### **Актуальность работы**

Проблема обработки биомедицинских изображений имеет в настоящее время большое значение для биологии и медицины. Лаборантам-патоморфологам приходится обрабатывать несколько сотен изображений препаратов в день.

Типовыми задачами для обработки препаратов являются обнаружение и подсчет клеточных структур, таких как ядра клеток либо железы, на единицу площади. Также выполняется подсчет структур, выделенных различными маркерами.

При выполнении обработки изображений вручную патоморфологи не только выполняют рутинную работу, но и допускают ошибки. Поэтому разработка методов и приложений, осуществляющих анализ медицинских изображений актуальна.

Обработка медицинских изображений может осуществляться в полуавтоматическом и автоматическом режимах. Полуавтоматический режим позволяют уменьшить число действий оператора для обработки одного препарата, однако требует вмешательства оператора при обработке каждого препарата. Автоматический режим не требует непосредственного вмешательства оператора при обработке препаратов, но лишь начальной настройки.

Особенно актуально распознавание препаратов именно в автоматическом режиме

### **Степень разработанности темы исследования.**

К задаче распознавания биомедицинских изображений применяют современные алгоритмы компьютерного зрения и машинного обучения.

Предварительная обработка изображений с использованием морфологических операторов и фильтров повышения яркости подробно описана в работах Chen S, Zhao M.

Для детектирования объектов интереса применяют различные алгоритмы, такие как алгоритм активных контуров, алгоритм водоразделов, сегментация изображений по известным классам, сегментация с предварительным обнаружением границ классов.

Заключительный этап - классификация объектов интереса. Этот этап позволяет отнести каждый из объектов интереса к одному из нескольких классов, либо ни к одному из них. На этапе классификации объекты интереса на изображениях препаратов выделяют маркерами. Это позволяет подкрасить интересующие нас ядра определенным цветом. Всего выделяют четыре типа классов: ядро, подкрашенное маркером; ядро, не подкрашенное маркером; фон (строма); группа ядер. Популярные классификаторы, такие как метод

опорных векторов, классификатор Байеса, каскад Хаара, сверточная нейронная сеть, описаны в работах Gurcan M. N

На рынке распознавания биомедицинских изображений представлены современные алгоритмы и приложения, такие как ImageJ, CellProfiler, Orbit Image Analysis, AxioVision 4.8, Видео-Тест-Морфология 5.2, CellSens, IMAGE-PRO-Premium, BioVision. Однако эти приложения способны выполнять обработку медицинских изображений лишь в полуавтоматическом режиме, либо имеют высокую стоимость и отсутствуют в открытом доступе.

Поэтому было решено провести исследование совместно с НИИ акушерства, гинекологии и репродуктологии имени Д.О. Отта и разработать и реализовать бесплатный и эффективный метод, распознающий биомедицинские изображения в полностью автоматическом режиме.

### **Цель и задачи исследования**

Целью диссертационной работы является исследование и разработка полностью автоматизированных методов компьютерного зрения для обработки медицинских изображений.

### **Научная новизна**

Применение современных алгоритмов компьютерного зрения и машинного обучения для распознавания клеточных структур на изображениях конфокальной и световой микроскопии:

1. В работе разработан метод распознавания ядер клеток на изображениях световой микроскопии;
2. В работе разработан метод распознавания ядер клеток на изображениях конфокальной микроскопии;
3. В работе разработан метод распознавания контуров желез на изображениях конфокальной микроскопии;

### **Теоретическая и практическая значимость**

Теоретическая значимость работы заключается в правильной комбинации существующих алгоритмов для решения прикладных задач по распознаванию биомедицинских изображений. Разработанный метод может быть использован для широкого спектра аналогичных задач распознавания и подсчета большого числа схожих объектов интереса на изображениях.

Практическая значимость работы заключается в том, что предложенные в работе методы распознавания биомедицинских изображений не только позволят автоматизировать труд лаборантов-патоморфологов и

исключить фактор человеческой ошибки при распознавании препаратов, но и могут быть применены как основа для автоматической установки диагноза пациента, в частности для распознавания некоторых типов онкологии.

### **Методология и методы диссертационного исследования**

Методология данного диссертационного исследования базировалась на междисциплинарном подходе информатики, биологии и медицины, с применением методов микроскопии, выделения клеточных структур с помощью маркеров, компьютерного зрения, машинного обучения, статистического анализа.

### **Положения, выносимые на защиту:**

1. Метод распознавания ядер клеток на изображениях световой микроскопии, подсчет ядер клеток, выделенных и не выделенных маркерами, распознавание ядер, перекрывающих друг друга;
2. Метод распознавания ядер клеток на изображениях конфокальной микроскопии, подсчет ядер клеток, выделенных и не выделенных маркерами, распознавание ядер, перекрывающих друг друга;
3. Метод распознавания контуров желез на изображениях конфокальной микроскопии, подсчет количества желез на изображениях, выделение внешних и внутренних контуров желез, подсчет ядер клеток внутри железы и вне нее;

### **Реализация результатов работы**

В результате работы было разработано консольное приложение, позволяющее обрабатывать изображения биомедицинских препаратов в автоматическом режиме и выдавать результаты в виде размеченных изображений и файла статистики в формате csv. В настоящее время приложение применяется лаборантами-патоморфологами в их научных работах по анализу препаратов с целью выявления заболеваний.

### **Апробация полученных результатов и публикациях.**

Данная работа была представлена на нескольких международных конференциях по биологии и информатике. Результаты были опубликованы в 3х научных работах, в том числе 2 – в индексируемых в базах SCOPUS и Web of Science. Также было сделано несколько публикаций в журналах с медицинской тематикой.

### **Личный вклад.**

Данная работа выполнялась совместно с сотрудниками НИИ Отта и студентами ИКНТ политехнического университета. Весомый вклад внесли Дробинцева А., Кукавица Н., Ценевска С., Андреев А.

## СОДЕРЖАНИЕ РАБОТЫ

*(по главам)*

**Во введении** доказывається актуальность темы исследования, описываются цель и поставленные задачи, формулируется теоретическая и практическая значимость работы и описываются краткие результаты.

**В первой главе** более подробно рассматриваются задачи, которые стоят перед алгоритмами обработки медицинских изображений. Приводится сравнение алгоритмов, работающих в автоматическом и полуавтоматическом режимах, и доказывається, что полностью автоматизированные алгоритмы более подходят для повседневных задач патоморфологов.

Приводится обзор существующих программных комплексов и перечисляются их достоинства и недостатки

Перечисляются шаги по обработке биомедицинских изображений с применением современных алгоритмов и выполняется обзор этих алгоритмов

В качестве первого шага рассматривается предварительная обработка изображений. Такая обработка применяется для создания условий, которые повышают качество детектирования объектов интереса на биомедицинских изображениях. Предварительная обработка может быть осуществлена с помощью морфологических операторов и фильтров, детекторов границ и фильтров с нормализацией яркости.

Второй шаг – детектирование объектов интереса. На этом шаге необходимо определить координаты  $X$  и  $Y$  предполагаемого центра каждого объекта интереса. В результате этапа получается набор объектов интереса, которые, вероятно, являются ядрами. Параметры алгоритма на данном шаге подбираются таким образом, чтобы число объектов интереса было избыточным. Другими словами, допускается детектирование ложных объектов, но не допускается пропуск настоящих ядер. В основном, используют следующие алгоритмы: алгоритм активных контуров, алгоритм водоразделов, сегментация изображений по известным классам, сегментация с предварительным обнаружением границ классов

Заключительный этап позволяет отнести каждый из объектов интереса к одному из целевых классов. Для задачи классификации ядер на конфокальной и флуоресцентной микроскопии выделяют характеристики следующих классов: топологические, текстурные и основанные на интенсивности. На этапе классификации классы объектов интереса определяются маркерами, использующимися для подкраски ядер и включают в себя:

- Не подкрашенное маркером ядро;

- Фон (строма);
- Ядро, подкрашенное маркером;
- Несколько ядер.

Отдельной задачей является разделение наложенных друг на друга ядер. Это может быть вызвано делением клеток, углом обзора камеры при снятии препарата, а также, расположением ядер друг над другом в толще исследуемой ткани. Для разделения слившихся ядер и точного определения их числа могут быть использованы следующие подходы, описанные в работе Y. Yorozu:

- Метод активных контуров с предварительным применением эрозии;
- Классификаторы (сверточная нейронная сеть), предварительно обученные на классы, определяющие число клеток в области заданного размера;
- Алгоритмы водоразделов;
- Алгоритмы сегментации, ориентированные на топологические особенности объектов.

**В третьей главе** описаны объекты (предмет) исследования. В качестве данных для анализа были собраны и размечены изображения препаратов клеточных структур и тканей световой и конфокальной микроскопии (см. Табл. 1).

<b>Класс собранных изображений</b>	<b>Кол-во изображений</b>
Препараты клеточных структур световой микроскопии	100
Препараты клеточных структур конфокальной микроскопии	30
Препараты тканей конфокальной микроскопии	92

Таблица 1. Собранные изображения

Материал и оборудование для съемки было предоставлено НИИ акушерства, гинекологии и репродуктологии имени Д.О. Отта. Съемка отдельных классов изображений осуществлялась с фиксированным увеличением микроскопа. Были собраны изображения препаратов различных типов тканей с различными условиями освещенности и цветами маркеров.

Для ускорения разметки препаратов было разработано приложение, обеспечивающее разметку клеточных структур на изображениях и сохранение результатов разметки в csv файлы. Разметка выполнялась специалистом патоморфологом из НИИ Отта.

**В четвертой главе** описана цель работы и поставленные задачи. Целью работы стало исследование и разработка следующих алгоритмов:

- Подсчет числа ядер клеток с экспрессией и без экспрессии исследуемого маркера на изображениях препаратов световой микроскопии;

- Подсчет числа ядер клеток с экспрессией и без экспрессии исследуемого маркера на изображениях препаратов конфокальной микроскопии;
- Выделение внутренних и внешних границ желез на изображениях препаратов конфокальной микроскопии.

Для всех алгоритмов были установлены следующие требования:

- работа без участия оператора;
- устойчивость к изменению яркости препарата;
- устойчивость к различным цветам маркеров;
- масштаб изображения является входным параметром алгоритмов.

**В пятой главе** описываются предложенные методы для решения поставленных задач по распознаванию биомедицинских изображений:

**А.** Подсчет числа ядер клеток на изображениях конфокальной микроскопии.

- Считать цветное изображение  $I$  в формате RGB, глубина 8 бит на канал;
- Привести изображение к масштабу 1.5 нанометров на пиксел
- Преобразовать  $I$  в формат HSV, записать  $V$  компоненту в переменную  $V$ ;
- Применить на изображении  $V$  метод contrast limited adaptive histogram equalization с параметрами clipLimit = 2, titleGridSize = 8;
- Выполнить на изображении  $V$  эрозию с ядром эллипса размера 3;
- Вычислить mean как среднее значение пикселей  $V$ ;
- Для каждого пиксела  $V_{ij}$ : если  $V_{ij} > \text{mean} + 20$ , присвоить  $V_{ij} = \min(V_{ij} + 100, 255)$ , если  $V_{ij} < \text{mean} - 20$ , присвоить  $V_{ij} = \max(V_{ij} - 100, 0)$ ;
- Применить к изображению  $V$  медианный фильтр с ядром размера 5;
- Выполнить пороговую бинаризацию изображения  $V$  с порогом 127, результат записать в переменную  $B$ ;
- Выполнить поиск контуров на изображении  $B$ , оставив контуры, не имеющие вложенных контуров. Результат записать в переменную contours;
- Вычислить центры масс ( $C_x$ ,  $C_y$ ) для каждого контура из contours по формулам:

$$C_x = m10 / m00$$

$$C_y = m01 / m00$$

$$m(p, q) = \text{Sum } [i=1..n] (x, y) x^p y^q \quad (1)$$

Оставить лишь те контуры, у которых

$$\text{abs}(C_x - C_y) < 1;$$

- Число полученных контуров будет суммарным числом ядер

**В.** Выделение внутренних контуров желез на изображениях конфокальной микроскопии

- Считать цветное изображение  $I$  в формате RGB, глубина 8 бит на канал;
- Привести изображение к масштабу 1.5 нанометров на пиксел
- Преобразовать  $I$  в формат HSV, записать  $V$  компоненту в переменную  $V$ ;
- Использовать на изображении  $V$  метод contrast limited adaptive histogram equalization с параметрами clipLimit = 2, titleGridSize = 8;
- Выполнить на изображении  $V$  пороговую бинаризацию с порогом 127;
- Выполнить на изображении  $V$  27 шагов эрозии с ядром эллипса размера 3
- Выполнить поиск контуров на изображении  $B$ , оставив контуры, не имеющие вложенных контуров. Результат записать в переменную contours;
- Для каждого контура contour: вычислить площадь контура и записать в contourArea, посчитать число пикселей  $V_{ij}$ , попадающих в данный контур, со значением  $V_{ij} > 15$  и записать в переменную nonZeroPixelsArea. Считать что contour является внутренней границей железы, если nonZeroPixelsArea < contourArea \* 0.4.

### С. Подсчет числа ядер клеток на изображениях световой микроскопии.

- Считать цветное изображение  $I$  в формате RGB, глубина 8 бит на канал и привести его к масштабу 1.5 нанометров на пиксел
- Преобразовать  $I$  в HSV формат, записать  $V$  компоненту в переменную  $V$  (по аналогии с алгоритмами А и В)
- Нормализовать гистограмму интенсивности  $V$  компоненты. Для этого:
- Получить вектор распределения цветов (0..255);
- Записать значения гистограммы в массив chist, где  $chist[n] = \text{Sum}(0..n) \text{ hist}[i] (\text{hist}[i])$ ;
- Вычислить  $V_{max} = chist[\text{len}(chist) - 1] * (100 - \text{thresh}) / 100$ , где thresh – пороговое значение, равное единице;
- Заменить пиксели изображения  $V$  со значением большим, чем  $V_{max}$ , на значение  $V_{max}$ ;
- Применить преобразование Лапласа на изображении  $V$  с sigma 3, 6 и 9 и сохранить результаты в L1, L2, L3 соответственно;
- Вычислить сумму изображений L1, L2, L3 по пикселям и записать в Lsum, нормализовать Lsum в диапазоне от 0 до 255;
- Выполнить бинаризацию Отсу на Lsum и записать результат в переменную B;
- Выполнить поиск контуров на изображении B, оставляя только внутренние контуры (также, как в алгоритме А). Результат записать в массив contours;
- Вычислить моменты контуров  $S_{xm}$   $S_{ym}$  по формуле (1) и оставить только те контуры, у которых  $|S_x - S_y| < 1$ ;

Коэффициенты алгоритмов были подобраны опытным путем при тестировании на анализируемых препаратах

## Результаты и их обсуждение

1. Алгоритм подсчета числа ядер клеток на изображениях конфокальной микроскопии.

Процент успешного распознавания составил 84% на исследуемых изображениях. Пример распознавания изображения представлен на Рис. 1.

Для задачи подсчета ядер процент успешного распознавания выдавался как усредненное значение для анализируемых изображений. Для каждого изображения  $I$  имелось число обнаруженных ядер `detectedCount` и число размеченных ядер `labelledCount`. Процент  $p_i$  распознавания для изображения  $I$  вычислялся по формуле (2):

$$p_i = \text{abs}(\text{detectedCount}/\text{labelledCount} - 1) * 100\% \quad (2)$$

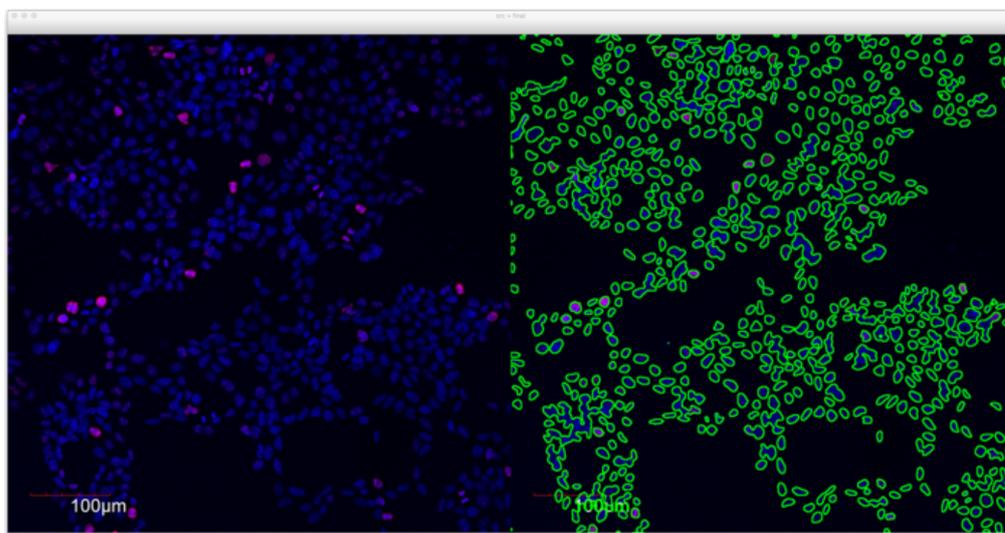


Рис. 1. Выделение ядер на препарате культуры клеток эндометрия визуализированом при помощи конфокальной микроскопии. Исходное изображение (слева), изображение с выделенными ядрами (справа)

2. Выделение внутренних контуров желез на изображениях конфокальной микроскопии.

Процент успешного распознавания составил 70% на исследуемых изображениях. Пример распознавания изображения представлен на Рис. 2.

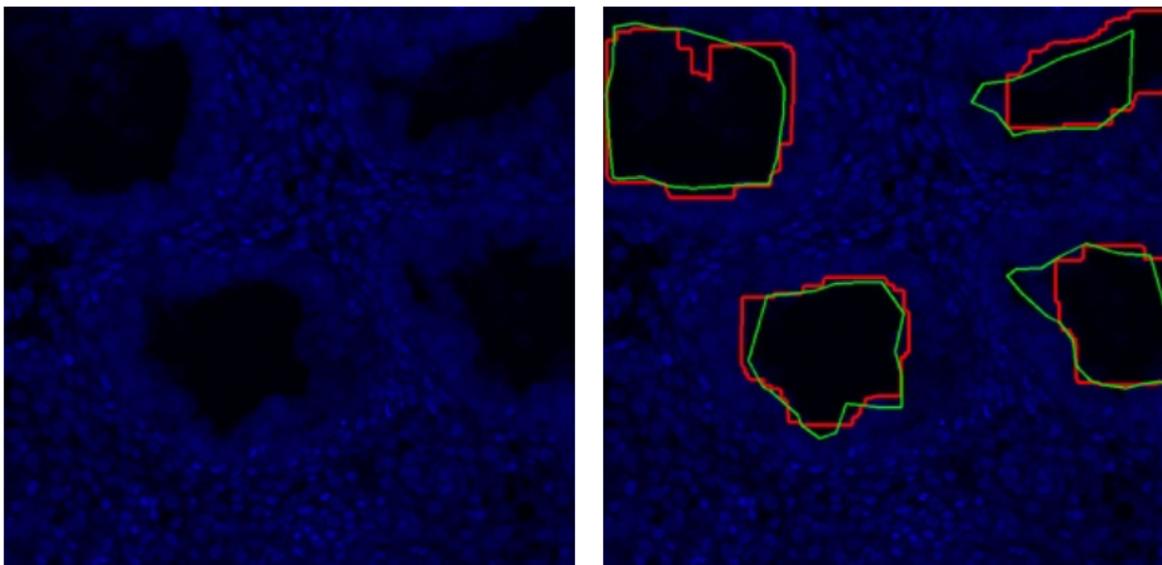


Рис. 2. Выделение ядер на гистологических срезах эндометрия визуализированных при помощи конфокальной микроскопии. Исходное изображение (слева), изображение с выделенными ядрами (справа)

Для задачи подсчета желез процент успешного распознавания выдавался как усредненное значение для анализируемых изображений. Для каждого изображения  $I$  имелись внутренние контуры желез, детектированные алгоритмом – `detectedContours` и внутренние контуры желез, размеченные патоморфологом – `labelledContours`. Вычислялась площадь пересечения размеченных и задетектированных контуров `intersectArea` и суммарная площадь размеченных контуров – `labelledArea`. Процент  $p_i$  распознавания для изображения  $I$  вычислялся по формуле:

$$p_i = \text{abs}(\text{intersectArea}/\text{labelledArea} - 1) * 100\% \quad (3)$$

3. Подсчет числа ядер клеток на изображениях световой микроскопии  
 размеченных контуров  
 Процент успешного распознавания составил 90% на исследуемых изображениях. Пример распознавания изображения представлен на Рис. 3.

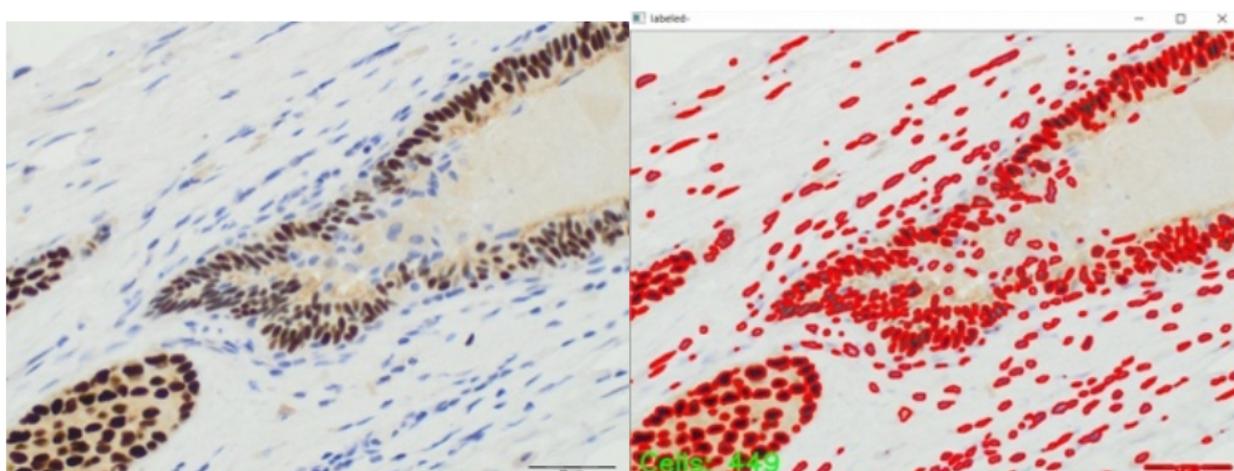


Рис. 3. Выделение ядер на гистологических срезах молочной железы визуализированных при помощи световой микроскопии. Исходное изображение (слева), изображение с выделенными ядрами (справа)

Для задачи подсчета ядер на препаратах световой микроскопии процент успешного распознавания анализировался по формуле (2) аналогично проценту распознавания для конфокальной микроскопии.

#### 4. Сравнение с существующими приложениями

Целью последнего этапа работы было сравнение полученных данных с результатами уже существующих на рынке приложений для распознавания микрофотографий.

Поскольку большая часть программного обеспечения, предназначенного для цитологических исследований либо имеет высокую стоимость, либо требует знания языков программирования, было решено сравнить производительность созданного приложения с Fiji – плагином ImageJ для оценки микроскопических изображений, так как этот плагин эффективен, популярен среди патоморфологов и не требует большого порога вхождения. Результаты сравнения предложенного алгоритма и Фиджи для задачи распознавания ядер на конфокальных изображениях представлены в таблице 2.

No	Размечено вручную	Fiji	%Fiji	Наш алгоритм	% нашего алгоритма
1 (5)	11	15	73.3	14	78.5
1 (6)	11	17	64.7	18	61.1
1 (7)	95	109	87.2	94	98.9
1 (9)	12	25	48.0	21	57.1
1 (13)	360	424	84.9	195	54.2
1 (30)	145	261	55.6	165	87.8
1 (31)	128	161	79.5	128	100
1 (38)	45	59	76.3	57	78.9
1 (41)	189	192	98.4	182	96.2
1 (42)	123	121	98.4	117	95.1
...					
Результат Fiji				76% +/- 16.1%	
Результат нашего алгоритма				83.8% +/- 11.1%	

Таблица 2. Сравнение предложенного алгоритма и плагина Fiji на базе конфокальных изображений

Для задачи подсчета ядер на препаратах световой микроскопии процент успешного распознавания был подсчитан по формуле (2) аналогично проценту распознавания для препаратов конфокальной микроскопии.

Таким образом, разработанный алгоритм превышает точность FiJi на 7,8%. Следует отметить, что для достижения максимальной точности в FiJi требуется задать пороговый параметр и предполагаемый радиус интересующего объекта должен быть введен как нижний предел, тогда как в написанном коде происходит определение размера объектов автоматически, что исключает элемент субъективности исследования и необходимость предварительной обработки фотографий. Например, для оценки изображения созданный алгоритм не требует предварительного удаления масштабной линейки с изображения. Кроме того, для выражения различных цветовых маркеров в задаче классификации ядер с использованием FiJi необходимо установить значение цветового порога для каждого типа маркера на каждой фотографии отдельно, что не только значительно увеличивает время анализа, проводимого патоморфологом, но также значительно снижает качество этого анализа, поскольку ядра со слабой экспрессией, скорее всего, не будут включены в соответствующую группу. В отличие от FiJi, предложенный нами алгоритм одинаково эффективно справляется как с задачей подсчета общего количества ядер, так и с задачей их классификации.

### **Заключение**

В заключение цель поставленной работы была выполнена, а именно были исследованы и разработаны автоматизированные методы компьютерного зрения для обработки медицинских изображений:

1. Метод распознавания ядер клеток на изображениях световой микроскопии, подсчет ядер клеток, выделенных и не выделенных маркерами, распознавание ядер, перекрывающихся друг друга;
2. Метод распознавания ядер клеток на изображениях конфокальной микроскопии, подсчет ядер клеток, выделенных и не выделенных маркерами, распознавание ядер, перекрывающихся друг друга;
3. Метод распознавания контуров желез на изображениях конфокальной микроскопии, подсчет количества желез на изображениях, выделение внешних и внутренних контуров желез, подсчет ядер клеток внутри железы и вне нее;

Также было разработано консольное приложение, реализующее предложенные алгоритмы и дающее лучшие результаты распознавания, чем у существующих аналогов.

В дальнейшем планируется продолжить исследования в области обработки микрофотографий и улучшить показатели надежности алгоритмов с помощью применения U-net сетей.

### **Список работ, опубликованных по теме научно-квалификационной работы (диссертации)**

#### **Публикации в изданиях, рецензируемых ВАК**

1. Sergeev D. I. et al. Development of automated computer vision methods for cell counting and endometrial gland detection for medical images processing //Труды института системного программирования РАН. – 2020. – Т. 32. – №. 3. – С. 119-130
2. A. Drobintseva, S Cenevska, A Andreev, D Sergeev, P Drobintsev. P318. New approach to detection and analysis of cancer-associated proteins in fluorescent digital microscopy.
3. Андреев А.Е., Ценевска С.Л., Сергеев Д.И., Кукавица Н., Дробинцева А.О., Дробинцев П.Д. КОМПЬЮТЕРНЫЙ АНАЛИЗ МИКРОФОТОГРАФИЙ НА ПРИМЕРЕ ВЫЯВЛЕНИЯ КЛЕТОЧНЫХ ЯДЕР // Биотехносфера. 2018. № 1 (55). С. 8-14.

#### **Публикации в других изданиях**

1. Кукавица Н. и др. АВТОМАТИЗАЦИЯ АНАЛИЗА ИЗОБРАЖЕНИЙ КЛЕТОЧНЫХ СТРУКТУР ОКРАШЕННЫХ ИММУНОГИСТОХИМИЧЕСКИМИ МАРКЕРАМИ, С ИСПОЛЬЗОВАНИЕМ ТЕХНОЛОГИЙ КОМПЬЮТЕРНОГО ЗРЕНИЯ //НЕДЕЛЯ НАУКИ СПбПУ. – 2019. – С. 146-148.
2. Андреев А.Е., Сергеев Д.И., Кукавица Н., Дробинцева А.О., Крылова Ю.С. ПАТОЛОГИЧЕСКАЯ АНАТОМИЯ: ТРАДИЦИИ И СОВРЕМЕННОСТЬ  
Сборник научных трудов Всероссийской конференции, посвященной 160-летию кафедры патологической анатомии ВМедА им. С.М. Кирова. – 2019. – С. 12-17.

Аспирант \_\_\_\_\_ Сергеев Даниэл Игоревич  
(подпись)