



Московский государственный университет имени М. В. Ломоносова

Факультет вычислительной математики и кибернетики

Кафедра математических методов прогнозирования

Севастопольский Артем Михайлович

Методы распознавания глаукомы на основе нейронных сетей

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

Научный руководитель:

д.ф-м.н., профессор

А. Г. Дьяконов

Москва, 2017

Содержание

1	Введение	3
1.1	Определения и обозначения	6
1.2	Обзор литературы	7
2	Архитектура построенных моделей	8
3	Вычислительные эксперименты	12
3.1	Доступные базы данных	12
3.2	Условия эксперимента	12
3.3	Результаты экспериментов	13
3.4	Обсуждение и выводы	17
4	Заключение	19
	Список литературы	21

Аннотация

Глаукома — это заболевание глаза, которое характеризуется существенным поражением зрительного нерва. В 2010 г. число больных глаукомой оценивалось в 60 млн. человек по всему миру, и она является второй по численности причиной потери зрения в мире. Данная болезнь может привести к необратимым разрушениям зрительного нерва глаза вплоть до слепоты, поэтому диагностика глаукомы на ранних стадиях чрезвычайно важна. Визуальное исследование состояния зрительного нерва, которое включает в себя подсчёт отношения высот зрительного нерва и его экскавации (cup-to-disk ratio, CDR), является самым надёжным методом первичной диагностики болезни. Оценка CDR требует сегментации изображения зрительного нерва и его экскавации на снимках глазного дна может быть осуществлена современными алгоритмами компьютерного зрения. В данной выпускной квалификационной работе предлагается универсальный подход к автоматической сегментации данных органов, основанный на глубоких полносвёрточных нейронных сетях. Приведены эксперименты со сравнением предлагаемого метода на публично доступных базах данных. Для задач сегментации изображения зрительного нерва и его экскавации метод достигает качества, сравнимого с наилучшими известными методами других авторов, и превосходит их по времени получения ответов для новых изображений.

1 Введение

Глаукома — это заболевание глаза, которое характеризуется существенным поражением зрительного нерва [1]. В 2010 г. число больных глаукомой оценивалось в 60 млн. человек по всему миру, и она является второй по численности причиной потери зрения в мире [2, 3]. Согласно подсчётам [3], к 2020 г. ожидается увеличение числа людей, страдающих глаукомой, до 80 млн. человек. Данная болезнь может привести к необратимым разрушениям зрительного нерва глаза вплоть до слепоты, поэтому диагностика глаукомы на ранних стадиях чрезвычайно важна.

В большинстве случаев глаукома развивается в результате того, что внутриглазное давление постоянно или периодически превышает толерантный (индивидуально переносимый) уровень. В результате развивается характерное поражение головки зрительного нерва и ганглионарных клеток сетчатки; возникают характерные для глаукомы нарушения зрительных функций, такие как уменьшение поля зрения [1].

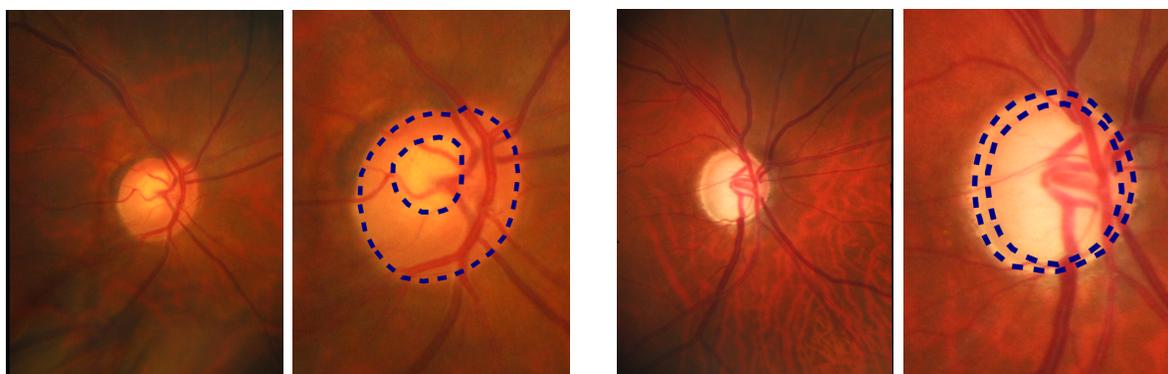
Для диагностики глаукомы в медицине широко используют следующие тесты:

- тонометрия — измерение внутриглазного давления
- офтальмоскопия или фотографирование глазного дна — визуальная оценка состояния глазного дна
- периметрия — измерение поля зрения; используется, как правило, для отслеживания изменений при ранее диагностированной глаукоме
- гониоскопия — оценка правильности строения угла, через который осуществляется отток жидкости из глаза.

Повышенное внутриглазное давление часто приводит не к глаукоме, а к другим болезням (например, окулярной гипертензии), а глаукома не всегда сопровождается повышенным внутриглазным давлением [4]. Периметрия позволяет оценить не саму патологию, а последствия от неё — насколько было уменьшено поле зрения из-за глаукомы [2]. Офтальмоскопия, т.е. анализ глазного дна, является наиболее надёжным общим тестом для диагностики глаукомы, т.к. обнаруживает непосредственное разрушение зрительного нерва при глаукоме вне зависимости от того, чем оно было вызвано [2, 5]. Другими словами, визуальная оценка состояния зрительного нерва

позволяет обнаружить проявления глаукомы, а повышенное внутриглазное давление, наследственность и другие признаки являются лишь факторами риска [4, 5].

Анализ глазного дна требует от врача локализации областей зрительного нерва и экскавации зрительного нерва (его внутреннего углубления, относительно ярко выраженного на снимке) и выделения их точных границ. Наличие проявлений глаукомы может быть определено по соотношению размеров зрительного нерва и его экскавации на полученном снимке (см. Рис. 1). Одним из часто используемых на практике показателей является cup-to-disk ratio (CDR) [2] — величина, равная отношению между высотами экскавации зрительного нерва и самого зрительного нерва. CDR является одним из самых репрезентативных признаков глаукомы, и в медицинской практике применяется следующее правило: следует подозревать глаукому, если $CDR \geq 0.65$ [6]. Сегментация изображения зрительного нерва и его экскавации и определение CDR — задачи, требующие большое время для выполнения и выполняемые на сегодняшний день только высоко квалифицированными профессионалами. Согласно исследованию, результаты которого приведены в работе [7], полная сегментация изображения зрительного нерва и его экскавации требует в среднем 8 минут на снимок.



(a) Здоровый глаз ($CDR \approx 0.45$)

(b) Глаз с подозрением на глаукому
($CDR \approx 0.90$)

Рис. 1: Пример снимка здорового глаза и глаза с патологией из базы данных RIM-ONE v.3 [8]. Для каждого из случаев приведено два изображения: слева — центральная часть глазного дна, справа — увеличенная область вокруг зрительного нерва с разметкой зрительного нерва (внешняя пунктирная линия) и его экскавации (внутренняя пунктирная линия). Можно наблюдать, что CDR значительно выше для глаза с подозрением на глаукому.

В данной работе рассматривается задача автоматической сегментации изображения зрительного нерва и экскавации зрительного нерва по фотоснимку глазного дна с целью принятия решения о возможности развития глаукомы. Данная задача предполагает, что от вычислительной системы требуется построить разметку рассматриваемых органов глаза, т.е. для каждого пикселя снимка определить, принадлежит он сегментируемому органу или нет. Решения для автоматической диагностики глаукомы позволяют во много раз снизить время обработки снимка, что может быть полезно в различных ситуациях, таких как массовый скрининг и оказание медицинской помощи в странах с существенной нехваткой высоко квалифицированных специалистов [9, 10].

Основные требования к системе автоматической сегментации органов глаза — это качество распознавания и быстрая обработка снимка. Для того, чтобы программа могла бы использоваться в качестве системы принятия решений или, по крайней мере, обработчика снимков глаза, ошибки сегментации должны быть минимальными. Время обработки снимка также является важным фактором, особенно в ситуациях, когда требуется анализ большого количества снимков. Если время обработки достаточно мало, это позволяет создать программные решения для анализа глазного дна в реальном времени. Время настройки модели распознавания также может быть важным параметром, к примеру, в случаях, когда требуется многократно обучать модель на новых данных.

Задача сегментации изображения зрительного нерва изучается во многих работах. Работы [11, 12] описывают методы, основанные на анализе отличительных характеристик зрительного нерва на снимке, таких как яркость и эллиптическая форма. В [13, 14] обучаются фильтры свёртки с помощью алгоритма Gentle Adaboost [15]. В [7] сначала удаляются сосуды, а затем изображение разбивается на маленькие части, и к каждой из них применяется нейронная сеть с несколькими слоями. В [16] используются глубокие свёрточные нейронные сети с техникой transfer learning. Сегментация изображения экскавации зрительного нерва освещается в меньшем числе работ ([13, 14]) и представляет собой более сложную задачу, т.к. граница экскавации менее заметна.

За счёт применения полносвёрточных нейронных сетей [17, 18] и их последующей модификации, автору удалось построить метод сегментации изображения органов глаза для диагностики глаукомы, который достигает сравнимого с наилучшими существующими методами качества распознавания (а в некоторых случаях — лучшего) и при этом обладает наибольшей скоростью обработки новых снимков и очень низкими требованиями к памяти. Таким образом, удалось построить универсальный метод для обеих рассматриваемых задач, объединяющий достоинства, присущие различным известным методам. Для наилучшего из построенных решений была подготовлена открыто доступная программная реализация, позволяющая обучать модели и использовать готовые модели для получения ответов на новых изображениях [19]. Это облегчает использование построенных методов на практике.

Далее текст работы организован следующим образом. В разделе 2 описываются предлагаемые модели для задачи сегментации, раздел 3 содержит описание постановки и результаты вычислительных экспериментов, в разделе 4 подводится итог выполненной работы и возможные направления дальнейшего исследования.

1.1 Определения и обозначения

Пусть $X \in \mathbb{R}^{H \times W \times 3}$, $x_{ij} \in [0, 1]$, $i = \overline{1, n}$, $j = \overline{1, m}$ — входное цветное изображение, закодированное в цветовом пространстве с 3 каналами (например, RGB). $Y \in \mathbb{R}^{H \times W}$, $y_{ij} \in \{0, 1\}$, $i = \overline{1, n}$, $j = \overline{1, m}$ — выходное чёрно-белое изображение, представляющее собой карту сегментации изображения: $y_{ij} = 0$, если пиксель в позиции (i, j) на снимке принадлежит фону, $y_{ij} = 1$ — если пиксель в позиции (i, j) на снимке принадлежит детектируемому объекту. Задача сегментации изображений состоит в построении функции $f : Y = f(X)$, переводящей произвольное входное изображение в карту сегментации. В рамках подхода машинного обучения с учителем, задача сегментации изображений решается за счет выбора семейства функций $f_w(X, w)$, $w \in W$, и настройки неизвестных параметров w на основе определённого количества обучающих данных $\mathbf{X} = \{X_i\}_{i=1}^N$, $\mathbf{Y} = \{Y_i\}_{i=1}^N$, где X_i — i -ое входное изображение, Y_i — соответствующая i -ому входному изображению карта сегментации. В литературе, как правило, рассматривается более широкая постановка задачи, когда $y_{ij} \in \{0, 1, \dots, K\}$, $K \in \mathbb{N}$ (см., к примеру, [17]).

В данной работе рассматривается две задачи сегментации: задача сегментации изображения, представляющего собой снимок глазного дна человека, на зрительный нерв и фон, и на экскавацию зрительного нерва и фон.

1.2 Обзор литературы

В данном разделе проводится обзор различных методов сегментации изображения зрительного нерва и его экскавации, которые были оценены их авторами на публично доступных базах данных с предоставленной корректной разметкой.

Для решения задачи сегментации изображения зрительного нерва, авторы [16] используют глубокую нейронную сеть из семейства полносвёрточных нейронных сетей [17]. Данное семейство моделей объединяет глубокие нейронные сети, принимающие полное изображение на входе и возвращающие полную карту сегментации на выходе. Используемая нейронная сеть основана на базовой сети VGG-16 [20] и при обучении использует технику transfer learning, заключающуюся в переносе весов первых слоёв из модели, обученной на базе данных с намного большим числом изображений. Обученная модель показывает более высокие результаты (с точки зрения функции качества Dice — см. раздел 3.2), чем дополнительная разметка второго эксперта, и авторы [16] делают вывод, что модель способна распознавать границу зрительного нерва лучше, чем человек.

Для решения задачи сегментации экскавации зрительного нерва, авторы [13] используют двухуровневый набор свёрточных фильтров, обрабатывающих изображения в нескольких разрешениях. В качестве предобработки используется энтропийная фильтрация [21] в цветовом пространстве $L^*a^*b^*$ для извлечения наиболее значимых окрестностей точек изображения. Затем к ним применяется нормализация контраста. Обучение фильтров проходит как независимое обучение линейных регрессоров с помощью алгоритма Gentle Adaboost [15]. После прохождения через свёрточные фильтры изображение подвергается процедуре, основанной на построении разрезов графов [22]. Качество метода было оценено на базе DRISHTI-GS [23, 24], и он превзошёл другие существующие методы по метрикам IOU и Dice (см. раздел 3.2). Однако важно заметить, что данный метод требует обрезки изображений по области их зрительного нерва (экскавации) перед сегментацией изображения зрительного нерва

(экскавации). Это делает метод неприменимым к новым снимкам глазного дна, т.к. требует, чтобы обрамляющий прямоугольник зрительного нерва и экскавации был заранее известен.

Работа [14] предлагает усовершенствование метода [13], заключающееся преимущественно в изменении процедуры обучения свёрточных фильтров. Также представлена оценка метода на базах DRISHTI-GS и RIM-ONE v.3 [8]. В отличие от метода [13], метод [14] не требует, чтобы для сегментации экскавации изображение было заранее обрезано по её области.

Метод из работы [16] имеет несколько недостатков. Он использует глубокую нейронную сеть, которая требует большое время для обучения; занимает большой объём видеопамяти; файл с параметрами нейронной сети также занимает большой объём. Кроме того, не удалось успешно воспроизвести результаты эксперимента, приводящего к заявленным результатам. Методы из работ [13, 14] достаточно сложны для программирования и воспроизведения результатов. Будучи подготовленными для выполнения на центральном процессоре, они также обладают большим временем получения ответов для новых изображений. Как было ранее упомянуто, метод [13] требует, чтобы изображения были заранее обрезаны по области того органа, изображение которого сегментируется, что представляет ещё один недостаток данного метода. Многие методы, не указанные в данном разделе, такие как [7, 11, 12], были в большинстве своём либо оценены на базах данных, которые в данный момент не доступны для публичного использования, либо оценены с помощью метрик, зависящих от пропорций между классами. Эти особенности делают сравнение с такими методами достаточно сложным. Обзор наилучших подходов на базах данных, которые в данный момент не доступны публично, представлен в работе [2].

2 Архитектура построенных моделей

В данном разделе описывается решение, показавшее в целом наилучшие результаты среди всех рассмотренных модификаций. Другие модификации описаны в разделе 3.

Главная часть предлагаемого решения основана на техниках глубинного обучения, а именно свёрточных нейронных сетях, которые проявили себя как одни из наилучших на сегодняшний день решений практически во всех задачах компьютерного зрения, в том числе классификации и сегментации изображений. Другое преимущество свёрточных нейронных сетей состоит в том, что одна и та же сеть может использоваться для распознавания различных объектов.

Для обеих рассматриваемых задач в качестве предобработки применяется Contrast Limited Adaptive Histogram Equalization (CLAHE, [25]). Это позволяет выровнять контраст изображения за счет изменения контраста его регионов и интерполяции результата между регионами. Для сегментации изображения экскавации, снимки сначала обрезаются по области зрительного нерва (с отступом с каждой стороны), которая может быть получена после применения алгоритма для сегментации изображения зрительного нерва.

После предобработки к изображению применяется свёрточная нейронная сеть, за основу которой взята модель архитектуры U-Net [18], которая показала очень высокие результаты на многих соревнованиях по сегментации биомедицинских изображений. U-Net была изначально построена для работы с обучающими выборками малого объёма. Это было мотивирующим фактором при выборе базовой архитектуры модели, т.к. выборки, используемые в рассматриваемых задачах, достаточно малы. U-Net является полносвёрточной нейронной сетью [17], т. е. принимает на вход изображение целиком и возвращает его полную карту сегментации.

Архитектура, предлагаемая в данной работе, изображена на Рис. 2. Как и U-Net в оригинальной статье [18], она состоит из сжимающего пути (пути, вдоль которого разрешение изображения уменьшается) и расширяющего пути (пути, вдоль которого разрешение изображения увеличивается). Сжимающий путь во многом структурно повторяет типичную архитектуру свёрточной части глубокой нейронной сети для классификации изображений, например, VGG-16 [20]. На расширяющем пути соединяется информация от слоёв того же разрешения с сжимающего пути и от слоёв меньшего разрешения с расширяющего пути. Сначала входное изображение пропускается через свёрточный слой с фильтрами размера 3 x 3 пикселя. Далее применяется Dropout-регуляризация [26] и функция активации ReLu ($f(x) = \max(0, x)$).

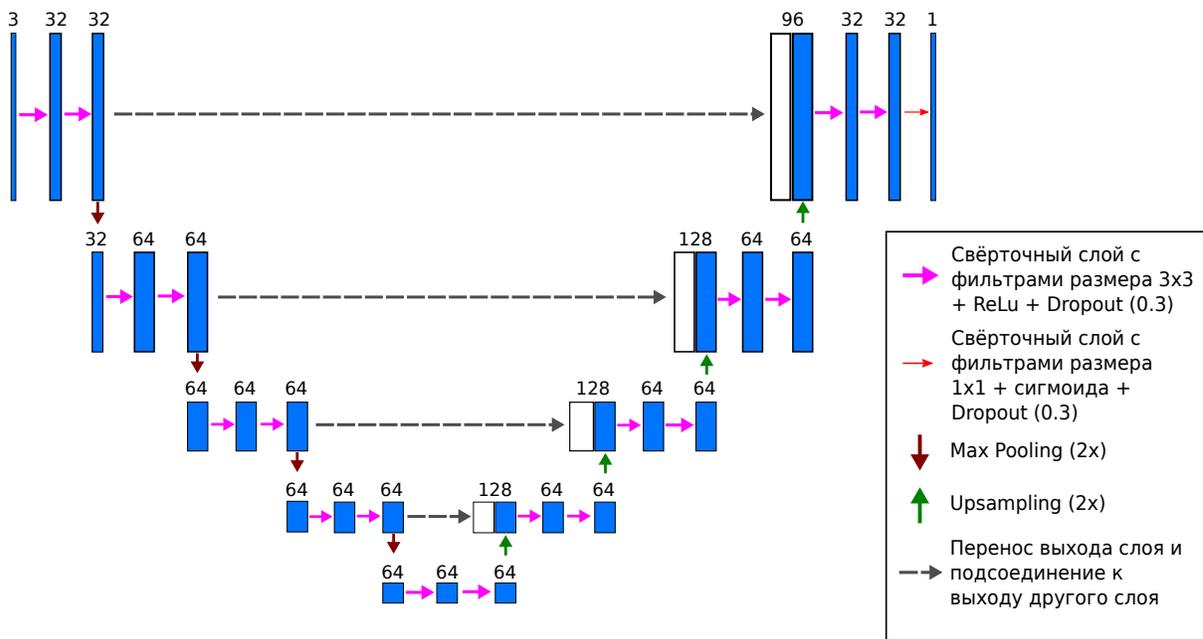


Рис. 2: Архитектура нейронной сети предлагаемого метода. Каждый синий прямоугольник отражает изображение, получающееся после применения очередной операции нейронной сети. Над каждым прямоугольником подписано число каналов соответствующего изображения.

Та же последовательность операций повторяется, затем применяется операция Max Pooling, уменьшающая разрешение изображения вдвое по каждой стороне [27]. Изображение пропускается через указанную последовательность слоёв еще несколько раз, пока разрешение не достигнет минимального. На расширяющем пути применяется та же последовательность свёрточных фильтров, перемежающаяся операциями Upsampling, повышающими разрешение изображения вдвое по каждой стороне за счет размножения каждого из пикселей в четыре пикселя (в два по каждой стороне).

В отличие от работы [18], в которой была впервые предложена нейронная сеть U-Net, представленная модификация имеет меньше фильтров в свёрточных слоях, и их число не увеличивается с каждым уровнем разрешения. Эксперименты показали, что такие изменения не приводят к ухудшению качества распознавания, но делают архитектуру модели более легковесной, что позволяет снизить требования к памяти, время обучения и обработки новых снимков. Кроме того, в предлагаемой модели используются операции Upsampling вместо обратных свёрточных слоёв для повышения разрешения изображения, применяется Dropout-регуляризация.

В качестве функции потерь для нейронной сети использовалась следующая функция $l(A, B)$:

$$l(A, B) = -\log d(A, B), \text{ где } d(A, B) = \frac{2 \sum_{i,j} a_{ij} b_{ij}}{\sum_{i,j} a_{ij}^2 + \sum_{i,j} b_{ij}^2},$$

где $A = (a_{ij})_{i=1}^H \times_{j=1}^W$ — построенная алгоритмом карта сегментации (возможно, содержащая вещественные значения), $B = (b_{ij})_{i=1}^H \times_{j=1}^W$ — корректная карта сегментации (с бинарными значениями). $d(A, B)$ является расширением функции качества Dice ($Dice(A, B) = \frac{2|A \cap B|}{|A| + |B|}$) для карт сегментации с вещественными значениями: если A и B содержат только бинарные значения, $d(A, B) = Dice(A, B)$, но $d(A, B)$ также поддерживает случай, когда $a_{ij} \in (0, 1)$, $b_{ij} \in (0, 1)$. Такое расширение позволяет вычислять градиент функции потерь и делает возможным обучение нейронной сети. Основное преимущество функции d как расширения Dice перед кросс-энтропийной функцией потерь, предлагаемой в работе [18], в том, что функция Dice не зависит от соотношения между площадью объекта и фона. Эксперименты показали, что с кросс-энтропийной функцией потерь не удаётся добиться сходимости процесса обучения для рассматриваемых задач.

В качестве метода оптимизации для обучения модели использовался стохастический градиентный спуск с моментом инерции (SGD with momentum, [28]). В течение всего процесса обучения использовались техники аугментации данных, т.е. дополнения обучающей выборки искусственными снимками, полученными из исходных случайными поворотами, увеличениями, уменьшениями, сдвигами и зеркальными отражениями.

Описанный метод не требует предварительной обрезки снимка по области того органа, сегментацию которого требуется получить. Детальное сравнение описанного метода с другими его модификациями и известными методами представлено в разделе 3.

3 Вычислительные эксперименты

3.1 Доступные базы данных

Для обучения и сравнения как построенных, так и известных методов используются базы данных, размещённые в сети Интернет и публично доступные для исследования.

- DRIONS-DB [29] — 110 цветных снимков без обрезки с предоставленной корректной разметкой изображения зрительного нерва,
- RIM-ONE v.3 [8] — 159 цветных снимков, обрезанных, так что диаметр зрительного нерва занимает в среднем приблизительно пятую часть длины изображения, с предоставленной корректной разметкой изображения зрительного нерва и его экскавации,
- DRISHTI-GS [23, 24] — 50 цветных снимков без обрезки с предоставленной корректной разметкой изображения зрительного нерва и его экскавации.

Для базы данных RIM-ONE версия № 3 является актуальной и содержит наиболее качественные снимки.

3.2 Условия эксперимента

В данном разделе представлено сравнение исследованных методов между собой и с известными методами других авторов.

Для оценки качества используются функции качества Intersection-over-Union score (IOU): $\frac{|A \cap B|}{|A \cup B|} \in [0, 1]$ и Dice score: $\frac{2|A \cap B|}{|A| + |B|} \in [0, 1]$, где $A = (a_{ij})_{i=1}^H \times_{j=1}^W$ — разметка, выданная алгоритмом (содержит, вообще говоря, вещественные числа, отражающие вероятность принадлежности каждого пикселя сегментируемому объекту), и $B = (b_{ij})_{i=1}^H \times_{j=1}^W$ — правильная разметка (состоящая из бинарных величин). Данные метрики качества не зависят от масштаба изображения и масштаба объекта на изображении.

Для основной модификации метода, описанной в разделе 2, значение параметра learning rate (темп обучения) было выставлено в 10^{-3} для сегментации зрительного

нерва и в $3 \cdot 10^{-4}$ для сегментации экскавации, значение параметра momentum — в 0.95. Размер mini-batch в методе оптимизации SGD был выбран равным 1, чтобы минимизировать объём используемой видеопамяти. В качестве разрешения входных изображений было выбрано 256 x 256 для сегментации изображения зрительного нерва и 512 x 512 для экскавации (до кадрирования; область интереса затем масштабировалась до разрешения 128 x 128 билинейной интерполяцией).

3.3 Результаты экспериментов

	RIM-ONE v.3	
	IOU	Dice
U-Net из оригинальной статьи с слоями Upsampling (число фильтров в сжимающем пути: 64 x 2, 128 x 2, 256 x 2, 512 x 2, 512 x 2)	0.84	0.93
U-Net из оригинальной статьи с слоями Upsampling + CLAHE	0.87	0.95
U-Net из оригинальной статьи с слоями Upsampling + CLAHE + обучение в режиме transfer learning на основе VGG-16 [20]	0.85	0.94
U-Net из оригинальной статьи с слоями Deconvolution [27] + CLAHE	0.85	0.94
U-Net с слоями Upsampling + CLAHE; число фильтров в сжимающем пути: $64 \times 4, 64 \times 2, 128 \times 2, 128 \times 2, 128 \times 2$	0.85	0.93
U-Net с слоями Upsampling + CLAHE; число фильтров в сжимающем пути: $32 \times 2, 64 \times 2, 128 \times 2, 128 \times 2, 128 \times 2$	0.88	0.96
Модификация, описанная в разделе 2 (число фильтров в сжимающем пути: $32 \times 2, 64 \times 2, 64 \times 2, 64 \times 2, 64 \times 2$)	0.88	0.96
U-Net с слоями Upsampling + CLAHE; число фильтров в сжимающем пути: $32 \times 2, 32 \times 2, 32 \times 2, 32 \times 2, 32 \times 2$	0.86	0.95
Модификация, описанная в разделе 2 + обучаемый фильтр в конце сети	0.83	0.90

Таблица 1: Сравнение исследованных архитектур нейронных сетей и методов предобработки для сегментации изображения **зрительного нерва**.

В таблицах 1 и 2 приведены результаты сравнения исследованных методов для сегментации изображения зрительного нерва. Таблица 1 отражает сравнение качества работы различных нейронных сетей и методов пред-обработки, Таблица 2 — сравнение качества работы различных методов пост-обработки изображений. Для каждой из модификации в таблице кратко указаны её отличительные параметры.

	RIM-ONE v.3	
	IOU	Dice
Модификация, описанная в разделе 2	0.881	0.943
Модификация, описанная в разделе 2 + оставление 1 связной компоненты	0.881	0.943
Модификация, описанная в разделе 2 + оставление 1 связной компоненты + замена на подобранный эллипс	0.882	0.944
Модификация, описанная в разделе 2 + оставление 1 связной компоненты + натягивание выпуклой оболочки	0.880	0.943
Модификация, описанная в разделе 2 + оставление 1 связной компоненты + метод главных компонент	0.830	0.912
Модификация, описанная в разделе 2 + mean filter	0.886	0.946

Таблица 2: Сравнение методов пост-обработки для сегментации изображения **зрительного нерва**.

	RIM-ONE v.3	
	IOU	Dice
U-Net из оригинальной статьи с слоями Upsampling	0.70	0.82
U-Net из оригинальной статьи с слоями Upsampling + CLAHE	0.72	0.82
U-Net из оригинальной статьи с слоями Upsampling + CLAHE + предварительная обрезка по области ЗН	0.71	0.83
Модификация, описанная в разделе 2	0.62	0.79
Модификация, описанная в разделе 2 + предварительная обрезка по области ЗН	0.72	0.84
U-Net с слоями Upsampling + CLAHE; число фильтров в сжимающем пути: $32 \times 2, 32 \times 2, 32 \times 2, 32 \times 2, 32 \times 2$	0.68	0.85
U-Net с слоями Upsampling + CLAHE; число фильтров в сжимающем пути: $32 \times 2, 32 \times 2, 32 \times 2, 32 \times 2, 32 \times 2$ + предварительная обрезка по области ЗН	0.70	0.82

Таблица 3: Сравнение исследованных архитектур нейронных сетей и методов предобработки для сегментации изображения **экскавации зрительного нерва**.

Перебор вариантов построен так, чтобы показать, как ключевые параметры архитектуры модели влияют на результат. В первых строках Таблицы 1 отражается эффект от применения CLAHE, затем перебираются способы повышения разрешения в расширяющем пути: Upsampling (повышение разрешения в два раза по каждой стороне

	DRIONS-DB		RIM-ONE v.3		Время обучения на RIM-ONE v.3	Время сегментации
	IOU	Dice	IOU	Dice		
Предлагаемый подход	0.89 ± ±0.01	0.94 ± ±0.01	0.89 ± ±0.01	0.95 ± ±0.01	26 · 382 = 9932 с	0.1 с
DRIU [16]	0.88	0.97	0.89	0.96	56 · 200 = 11200 с	0.13 с
Zilly et al. [14]	—	—	0.89	0.94	3296 с	5.3 с

Таблица 4: Сравнение наилучшего из построенных методов с известными методами других авторов для сегментации изображения **зрительного нерва**. Для оценок качества наилучшего из методов также приведено стандартное отклонение качества. «—» показывает, что результат не опубликован. Время обучения нейросетевых моделей указано как произведение времени одной эпохи и среднего числа эпох. Под временем сегментации понимается время получения разметки для нового изображения.

	Число параметров	Объём видеопамати для обучения	Объём файла с параметрами сети
Предлагаемый подход	6,6 · 10⁵	433 МБ	5 МБ
DRIU [16]	1,5 · 10 ⁷	860 МБ	61 МБ

Таблица 5: Сравнение методов, основанный на свёрточных нейронных сетях, для сегментации изображения **зрительного нерва** по характеристикам, связанным с объёмом требуемой памяти. Характеристики указаны для библиотеки Keras в версии 2.0.2 с базовым вычислителем Theano [31].

	DRISHTI-GS		RIM-ONE v.3		Время обучения на RIM-ONE v.3	Время сегментации
	IOU	Dice	IOU	Dice		
Предлагаемый подход	0.75 ± ±0.04	0.85 ± ±0.03	0.69 ± ±0.04	0.82 ± ±0.03	12 · 311 = 3732 с	0.06 с
Zilly et al. [14]	0.85	0.87	0.80	0.82	3296 с	5.3 с
BCF [13]	0.86	0.83	—	—	—	—

Таблица 6: Сравнение наилучшего из построенных методов с известными методами других авторов для сегментации изображения **экскавации зрительного нерва**. Для оценок качества наилучшего из методов также приведено стандартное отклонение качества. «—» показывает, что результат не опубликован. Время обучения нейросетевых моделей показано как произведение времени одной эпохи и среднего числа эпох. Под временем сегментации понимается время получения разметки для нового изображения.

за счет подстановки 4 пикселей на место каждого из исходных) и Deconvolution [27], или обратные свёрточные слои (применение обучаемого свёрточного фильтра не к

входному изображению, а к выходному, с разрешением вдвое больше по каждой стороне). Третья строка таблицы описывает архитектуру, обучаемую в режиме transfer learning: все слои сжимающего пути сети структурно совпадают с слоями сети VGG-16 [20], а их веса перед обучением копируются из такой сети, заранее обученной на крупном наборе данных ImageNet [30], чтобы задействовать результаты обучения на значительно большем наборе данных. Дальнейшие строки описывают облегчения архитектуры по сравнению с описанной в оригинальной статье [18]; при этом предполагается, что последовательность числа фильтров в расширяющем пути симметрична указанной последовательности числа фильтров в сжимающем пути. Модификация с обучаемым фильтром в конце сети (одним свёрточным слоем) была введена для попытки настроить оптимальный метод пост-обработки.

В Таблице 2 приведено сравнение методов возможной пост-обработки изображений, получающихся в результате применения нейронной сети (в модификации, описанной в разделе 2). Под заменой на подобранный эллипс понимается замена выходной разметки разметкой, которая лучше всего аппроксимирует контур получающейся фигуры в терминах метода наименьших квадратов. Под методом главных компонент (PCA) понимается его применение к векторам, получающимся при вытягивании разметки в вектор. PCA обучается на разметках обучающей выборки. Под mean filter понимается применение усредняющего фильтра к выходной разметке для её размытия, а затем пороговое отсечение по порогу 0.5. Радиус фильтра был подобран равным 7.

Аналогично, в Таблице 3 приведены результаты сравнения методов для сегментации изображения экскавации. Все методы сравниваются с использованием базы RIM-ONE v.3 с фиксированным разбиением на обучающую и контрольную выборку (для нейронных сетей для сегментации изображения зрительного нерва — в соотношении 99 / 60, для пост-обработки сегментации изображения зрительного нерва и для экскавации — 80 / 20). На Рис. 3 приведены изображения, отражающие лучший и худший случай применения метода к базе данных RIM-ONE v.3 для обеих из рассматриваемых задач.

Результаты сравнения наилучшего из исследованных методов с другими известными методами приведены в Таблицах 4, 5 и 6. Оценки качества построенного метода

получены по кросс-валидации с 5 фолдами. Для других методов указаны оценки качества, приведённые в соответствующих работах.

Предложенные методы были реализованы для вычислений на графическом процессоре (GPU). В качестве среды программирования использовался язык программирования Python 2.7, библиотека Keras для обучения нейронных сетей с базовым вычислителем Theano [31], библиотека Scikit-Image для реализации метода CLANE, библиотека OpenCV для реализации подбора эллипса и поиска выпуклой оболочки, библиотека Scikit-learn для реализации метода главных компонент (PCA). Оценки времени работы приведены для машины, совпадающей по характеристикам с машиной типа g2.2xlarge, предоставляемой Amazon Web Services [32] (видеокарта NVIDIA GRID (Kepler GK104), центральный процессор Intel Xeon E5-2670). Оценки времени работы для метода Zilly et al. [14] приведены для четырехъядерного центрального процессора мощностью 2.66 GHz, как указано в работе авторов метода. Такое сравнение времени можно считать корректным, т. к. данные методы построены для вычисления на центральном процессоре, а остальные, включая предложенный, — на графическом. Можно ожидать, что оценки времени работы метода BCF [13] близки к соответствующим оценкам для Zilly et al., т.к. эти методы очень схожи.

3.4 Обсуждение и выводы

На выбор модификации, описанной в разделе 2, повлияли следующие выводы о результатах экспериментов, отраженные в таблицах предыдущего раздела:

1. Предобработка методом CLANE улучшает результат.
2. Повышение разрешения методом Upsampling вместо Deconvolution не ухудшает результат, но уменьшает время обучения и обработки новых изображений.
3. Техника transfer learning не улучшает результат для задачи сегментации изображения зрительного нерва.
4. Число фильтров в свёрточных слоях нейронной сети уменьшено до возможного минимума без ухудшения качества.

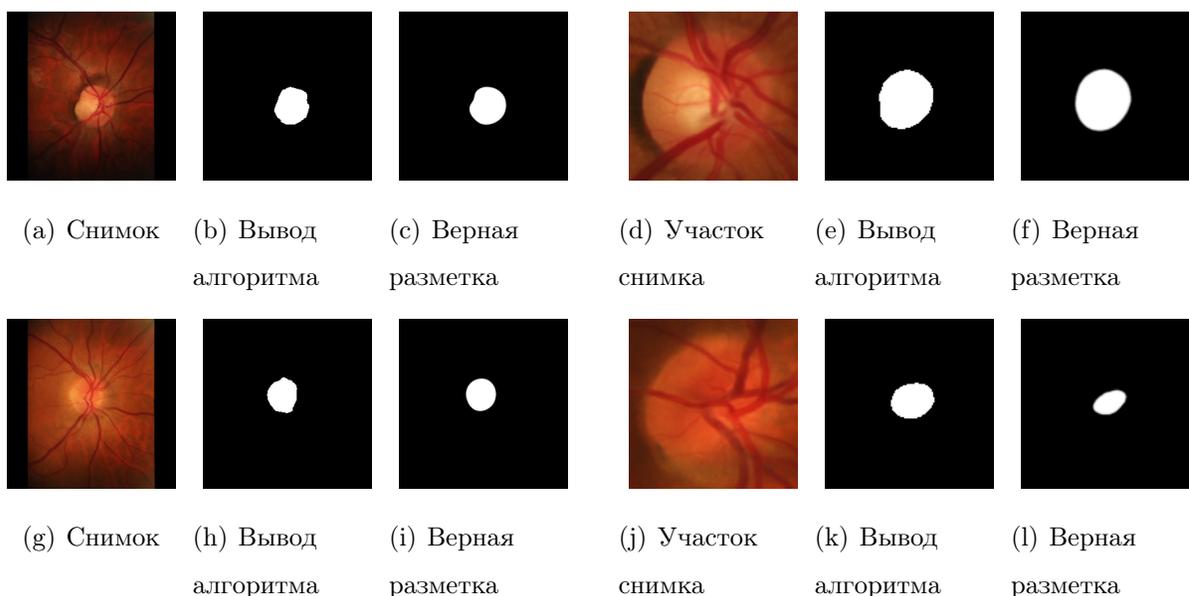


Рис. 3: Сравнение вывода алгоритма и корректного результата на базе RIM-ONE v.3 для зрительного нерва (a)-(c), (g)-(i) и экскавации (d)-(f), (j)-(l). На (d)-(f), (j)-(l) изображение увеличено до области зрительного нерва.

Для зрительного нерва: (a)-(c): лучший случай (IOU = 0.93, Dice = 0.97),

(g)-(i): худший случай (IOU = 0.80, Dice = 0.90);

для экскавации зрительного нерва: (d)-(f): лучший случай (IOU = 0.93, Dice = 0.97),

(j)-(l): худший случай (IOU = 0.46, Dice = 0.64).

5. Среди методов пост-обработки только применение mean filter существенно повышает качество, но этот эффект не меняет результат при подсчёте оценки качества по кросс-валидации.
6. Для задачи сегментации изображения экскавации зрительного нерва, предварительная обрезка по области зрительного нерва либо не ухудшает, либо улучшает результат.

Таблицы 4, 6 показывают, что построенное решение превосходит существующие методы по качеству в некоторых ситуациях; в остальных ситуациях демонстрирует сравнимые результаты, а также превосходит существующие методы по скорости получения разметки для новых изображений. Кроме того, Таблица 5 показывает, что предлагаемый метод превосходит наилучший известный нейросетевой метод по расходу видеопамати (что важно в силу большой стоимости памяти GPU) и по расходу памяти для сохранения параметров модели. Вместе с малым временем получения ре-

зультатов для новых изображений позволяет снизить требования к вычислительной системе, осуществляющей решение рассматриваемых задач.

4 Заключение

В рамках проведённого исследования был построен метод автоматической сегментации изображения зрительного нерва и экскавации зрительного нерва на снимках глазного дна для диагностики глаукомы. Представленное решение основывается на техниках глубинного обучения, а именно на полносвёрточных нейронных сетях. Построенный метод обладает следующими преимуществами:

1. Превосходит существующие методы по качеству в некоторых ситуациях; в остальных ситуациях демонстрирует сравнимые результаты.
2. Превосходит существующие методы по скорости получения разметки для новых изображений.
3. Методу требуется очень малый объём памяти для сохранения параметров модели (5 МБ) и малый объём видеопамати при обучении (433 МБ).
4. Метод универсален и демонстрирует высокое качество сегментации сразу для обеих рассматриваемых задач.
5. Метод прост как по своей структуре, так и для программирования с использованием современных библиотек глубинного обучения.
6. Для метода подготовлена и выложена в открытый доступ программная реализация [19], позволяющая, в отличие от реализаций методов других авторов, обучать модель на новых данных.

Достоинства метода позволяют снизить требования к вычислительной системе, осуществляющей автоматическую диагностику глаукомы, без потери качества распознавания, и тем самым сделать решения для диагностики более доступными для использования.

Количественная и визуальная оценка качества работы методов показывает, что изображение зрительного нерва сегментируется с очень высокой точностью, однако

для экскавации зрительного нерва достигается меньшая точность. Данного результата следовало ожидать, т. к. граница экскавации более трудноразличима. Автор полагает, что для задачи сегментации изображения экскавации зрительного нерва есть пространство для улучшения. Предполагаемые направления дальнейших исследований:

1. Собрать бóльшую базу данных снимков и использовать методы сегментации изображений, построенные для работы с большими объёмами данных.
2. Улучшить методы пост-обработки для уточнения выходной разметки.

По результатам исследований была подготовлена и принята к публикации научная работа [33]. Для программирования и исследования моделей успешно использовался опыт, полученный в ходе преддипломной практики.

Список литературы

- [1] Волков В.В. и др. Глазные болезни. Основы офтальмологии. — 2012.
- [2] Optic disc and optic cup segmentation methodologies for glaucoma image detection: a survey / Ahmed Almazroa, Ritambhar Burman, Kaamran Raahemifar, Vasudevan Lakshminarayanan // *Journal of ophthalmology*. — 2015. — Vol. 2015.
- [3] Quigley Harry A, Broman Aimee T. The number of people with glaucoma worldwide in 2010 and 2020 // *British journal of ophthalmology*. — 2006. — Vol. 90, no. 3. — Pp. 262–267.
- [4] Егоров Е.А., Алябьева Ж.Ю. Глаукома с нормальным давлением: патогенез, особенности клиники и лечения // Материалы конференции “Глаукома. Итоги и перспективы на рубеже тысячелетий”: М. — 1999.
- [5] *International Council of Ophthalmology*. ICO guidelines for glaucoma eye care. — 2016. — URL: http://www.icoph.org/enhancing_eyecare/glaucoma.html.
- [6] Glaucoma detection using novel optic disc localization, hybrid feature set and classification techniques / M Usman Akram, Anam Tariq, Shehzad Khalid et al. // *Australasian physical & engineering sciences in medicine*. — 2015. — Vol. 38, no. 4. — Pp. 643–655.
- [7] Integrated optic disc and cup segmentation with deep learning / Gilbert Lim, Yuan Cheng, Wynne Hsu, Mong Li Lee // 2015 IEEE 27th International Conference on Tools with Artificial Intelligence (ICTAI) / IEEE. — 2015. — Pp. 162–169.
- [8] RIM-ONE: An open retinal image database for optic nerve evaluation / F Fumero, Silvia Alayón, JL Sanchez et al. // 2011 24th International Symposium on Computer-Based Medical Systems (CBMS) / IEEE. — 2011. — Pp. 1–6.
- [9] Development and validation of a smartphone-based visual acuity test (peek acuity) for clinical practice and community-based fieldwork / Andrew Bastawrous, Hillary K Rono, Iain AT Livingstone et al. // *JAMA ophthalmology*. — 2015. — Vol. 133, no. 8. — Pp. 930–937.

- [10] Acceptability, Usability, and Views on Deployment of Peek, a Mobile Phone mHealth Intervention for Eye Care in Kenya: Qualitative Study / Vaishali Lodhia, Sarah Karanja, Shelley Lees, Andrew Bastawrous // *JMIR mHealth and uHealth*. — 2016. — Vol. 4, no. 2.
- [11] *Li Huiqi, Chutatape Opas*. Automated feature extraction in color retinal images by a model based approach // *IEEE Transactions on biomedical engineering*. — 2004. — Vol. 51, no. 2. — Pp. 246–254.
- [12] *Jose Jyothis, Kuruvilla Jinsa*. Detection of Red Lesions and Hard Exudates in Color Fundus Images // *International Journal of Engineering and Computer Science*. — 2014. — Vol. 3, no. 10. — Pp. 8583–8588.
- [13] *Zilly Julian G, Buhmann Joachim M, Mahapatra Dwarikanath*. Boosting convolutional filters with entropy sampling for optic cup and disc image segmentation from fundus images // *International Workshop on Machine Learning in Medical Imaging / Springer*. — 2015. — Pp. 136–143.
- [14] *Zilly Julian, Buhmann Joachim M, Mahapatra Dwarikanath*. Glaucoma detection using entropy sampling and ensemble learning for automatic optic cup and disc segmentation // *Computerized Medical Imaging and Graphics*. — 2017. — Vol. 55. — Pp. 28–41.
- [15] *Doğan Hatice, Akay Olcay*. Using AdaBoost classifiers in a hierarchical framework for classifying surface images of marble slabs // *Expert Systems with Applications*. — 2010. — Vol. 37, no. 12. — Pp. 8814–8821.
- [16] Deep retinal image understanding / Kevis-Kokitsi Maninis, Jordi Pont-Tuset, Pablo Arbeláez, Luc Van Gool // *International Conference on Medical Image Computing and Computer-Assisted Intervention / Springer*. — 2016. — Pp. 140–148.
- [17] *Long Jonathan, Shelhamer Evan, Darrell Trevor*. Fully convolutional networks for semantic segmentation // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. — 2015. — Pp. 3431–3440.

- [18] *Ronneberger Olaf, Fischer Philipp, Brox Thomas*. U-net: Convolutional networks for biomedical image segmentation // International Conference on Medical Image Computing and Computer-Assisted Intervention / Springer. — 2015. — Pp. 234–241.
- [19] *Sevastopolsky Artem*. seva100/optic-nerve-cnn: First version of the software. — 2017. — URL: <https://doi.org/10.5281/zenodo.546338>.
- [20] *Simonyan Karen, Zisserman Andrew*. Very deep convolutional networks for large-scale image recognition // *arXiv preprint arXiv:1409.1556*. — 2014.
- [21] *Gonzalez R.C., Woods R.E., Eddins S.L*. Digital Image Processing Using MATLAB. — Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2003.
- [22] *Salah Mohamed Ben, Mitiche Amar, Ayed Ismail Ben*. Multiregion image segmentation by parametric kernel graph cuts // *IEEE Transactions on Image Processing*. — 2011. — Vol. 20, no. 2. — Pp. 545–557.
- [23] A comprehensive retinal image dataset for the assessment of glaucoma from the optic nerve head analysis / Jayanthi Sivaswamy, S.R. Krishnadas, Arunava Chakravarty et al. // *JSM Biomedical Imaging Data Papers*. — 2015. — Vol. 2, no. 1.
- [24] Drishti-gs: Retinal image dataset for optic nerve head (onh) segmentation / Jayanthi Sivaswamy, S.R. Krishnadas, Gopal Datt Joshi et al. // 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI) / IEEE. — 2014. — Pp. 53–56.
- [25] *Szeliski Richard*. Computer vision: algorithms and applications. — Springer Science & Business Media, 2010.
- [26] Improving neural networks by preventing co-adaptation of feature detectors / Geoffrey E Hinton, Nitish Srivastava, Alex Krizhevsky et al. // *arXiv preprint arXiv:1207.0580*. — 2012.
- [27] *Dumoulin Vincent, Visin Francesco*. A guide to convolution arithmetic for deep learning // *arXiv preprint arXiv:1603.07285*. — 2016.

- [28] On the importance of initialization and momentum in deep learning. / Ilya Sutskever, James Martens, George E Dahl, Geoffrey E Hinton // *ICML (3)*. — 2013. — Vol. 28. — Pp. 1139–1147.
- [29] Identification of the optic nerve head with genetic algorithms / Enrique J Carmona, Mariano Rincón, Julián García-Feijoó, José M Martínez-de-la Casa // *Artificial Intelligence in Medicine*. — 2008. — Vol. 43, no. 3. — Pp. 243–259.
- [30] Imagenet large scale visual recognition challenge / Olga Russakovsky, Jia Deng, Hao Su et al. // *International Journal of Computer Vision*. — 2015. — Vol. 115, no. 3. — Pp. 211–252.
- [31] *Theano Development Team*. Theano: A Python framework for fast computation of mathematical expressions // *arXiv e-prints*. — 2016. — Vol. abs/1605.02688. — URL: <http://arxiv.org/abs/1605.02688>.
- [32] Amazon Web Services. — <https://aws.amazon.com>.
- [33] *Sevastopolsky Artem*. Optic Disc and Cup Segmentation Methods for Glaucoma Detection with Modification of U-Net Convolutional Neural Network // *arXiv preprint arXiv:1704.00979*. — 2017.