

# Управляемое искусственным интеллектом сшивание изображений сетчатки со смартфона для автоматизированной фундоскопии

Али Султан Маея

Санкт-Петербургский государственный электротехнический университет  
«ЛЭТИ» им. В.И. Ульянова (Ленина)  
Alimayya1357@gmail.com

**Аннотация.** В этой работе облегченная сверточная нейронная сеть (CNN) и оценка гомографии используются для сопоставления изображений сетчатки с низким разрешением, полученных с помощью смартфона. Для выделения признаков используется свисская сеть MobileNetV2, а упрощенная регрессионная модель на основе CNN используется для генерации матрицы гомографии, которая позволяет точно выровнять изображение, предсказывая смещения угловых точек. Модель достигла замечательных результатов при обучении с использованием наборов данных APTOS, FIRE и Retina Bank: PSNR составил 37,29 дБ, SSIM – 0,88 и RMSE – 3,54. Эти результаты обеспечивают вычислительную эффективность, необходимую для мобильных приложений в режиме реального времени, и при этом превосходят конкурирующие методы. Предлагаемый метод гарантирует высокоточную регистрацию изображений, обеспечивая более надежное автоматическое сканирование сетчатки, что позволяет эффективно выявлять такие заболевания, как глаукома и диабетическая ретинопатия.

**Ключевые слова:** анализ состояния сетчатки, автоматизированная фундоскопия, сшивание изображений сетчатки, оценка гомографии, глубокое обучение, Siamese MobileNetV2, CNN, мобильные устройства, приложения реального времени

## I. ВВЕДЕНИЕ

Традиционная фундоскопия имеет свои недостатки, особенно в таких областях специализации, как офтальмология. По данным Всемирной организации здравоохранения, более 2,2 млрд человек во всем мире страдают от нарушений зрения или слепоты, и по меньшей мере 1 млрд из них можно избежать или излечить путем раннего выявления и лечения [1]. Кроме того, для проведения стандартных офтальмологических обследований требуется физическое присутствие пациента на приеме у офтальмолога, что может быть затруднительно для многих людей. Люди, живущие в отдаленных районах и имеющие ограниченный доступ к специалистам (особенно в развивающихся странах) или имеющие ограниченную мобильность. С другой стороны, носимые биотехнологические устройства открывают новые возможности телемедицины, позволяя дистанционно оценивать состояние здоровья пациента. Одним из наиболее потенциальных применений является создание носимых устройств для анализа состояния глазного дна.[2].

Фундоскопия с помощью смартфона – это шаг вперед в диагностике проблем со зрением, поскольку камеры смартфонов используются для исследования глазного дна, включая сетчатку. К недостаткам этого подхода можно отнести низкое качество изображения, что увеличивает вероятность диагностических ошибок, и требования к квалифицированным специалистам [3].

Достижения в области фундоскопии с помощью смартфонов сделали скрининг сетчатки более доступным для диагностики таких заболеваний, как диабетическая ретинопатия и глаукома. Однако такие проблемы, как низкое разрешение изображения и узкое поле зрения, ограничивают надежность фундоскопии с помощью смартфонов. Автоматическое сшивание изображений сетчатки может преодолеть эти ограничения, обеспечивая полное панорамное изображение сетчатки. Однако традиционные методы сшивания часто не позволяют сохранить анатомические особенности, а также различия в освещении и ориентации. Кроме того, традиционные методы объединения изображений, основанные на сопоставлении ключевых точек или характеристик изображения, такие как SIFT и SURF или ORB, могут демонстрировать надежность и стабильность, но это зависит от многих факторов, как упоминалось выше, и они также имеют ограничения при работе с изображениями низкого разрешения с недостаточно четкими характеристиками [4].

Использование методов глубокого обучения, в частности сверточных нейронных сетей (CNN), значительно повысило точность извлечения признаков и выравнивания [5]. Цель этой работы – создать подход к оценке гомографии на основе CNN для сшивания изображений сетчатки, который можно обрабатывать в режиме реального времени на мобильных устройствах. В методе используются новейшие методы глубокого обучения, позволяющие получить точное, эффективное и масштабируемое решение для улучшения офтальмологической диагностики с помощью смартфона.

## II. МАТЕРИАЛЫ И МЕТОДЫ

### A. Ретинальные Наборы Данных

Это исследование включало три ключевых набора данных: APTOS, FIRE и Retina Bank. Эти наборы данных содержат различные изображения сетчатки, как

нормальной, так и при заболеваниях, с различным разрешением и условиями визуализации.

Азиатско-Тихоокеанское общество телеофтальмологов (APTOS) провело конкурс по выявлению слепоты APTOS 2019, в рамках которого была представлена коллекция изображений глазного дна сетчатки с аннотациями на диабетическую ретинопатию (ДР). Этот набор данных содержит 3662 фотографии для обучения и 1928 изображений для тестирования, все в высоком разрешении. Врачи оценивали каждое изображение по шкале от 0 до 4, которая отражает степень тяжести заболевания. Набор данных предназначен для создания методов машинного обучения для выявления аномалий сетчатки [6].

Набор данных для регистрации изображений глазного дна (FIRE) – это специальная коллекция снимков сетчатки, предназначенная для оценки задач выравнивания изображений, с особым акцентом на методы сшивания изображений и регистрации. Он состоит из 129 снимков сетчатки с высоким разрешением, составляющих 134 пары изображений, каждая из которых помечена точными данными, что помогает надежно оценить алгоритмы регистрации [7].

Банк изображений сетчатки, созданный по инициативе Американского общества специалистов по сетчатке, предоставляет обширную коллекцию изображений сетчатки, в том числе полученных с помощью фундоскопии со смартфона. Эти изображения имитируют реальные ситуации, что может привести к ухудшению качества изображения по сравнению с обычными клиническими камерами, снимающими глазное дно. Библиотека является отличным ресурсом для врачей и ученых, желающих изучать заболевания сетчатки при различных условиях визуализации [8].

#### *А. Предварительная обработка и подготовка данных*

Чтобы подготовить набор данных для обучения нейронной сети объединению изображений, синтезированный набор данных был использован для создания набора данных для обучения нейронной сети объединению изображений, с особым акцентом на изображения сетчатки с низким разрешением. Набор данных был создан с использованием APTOS, FIRE и ручную собранных фотографий сетчатки со смартфона из Retina Image Bank.

Чтобы имитировать пониженное качество, часто встречающееся при визуализации сетчатки с помощью смартфонов, все фотографии были масштабированы до  $128 \times 128$ . Пары изображений с перекрывающимися частями были случайным образом обрезаны и использованы для вычисления гомографии. Для повышения стабильности модели и имитации реальных aberrаций был использован ряд преобразований, включая переключение, регулировку яркости и контрастности, а также увеличение шума. Эти изменения были направлены на имитацию артефактов, которые обычно встречаются на изображениях сетчатки низкого качества.

Для упрощения оценки соответствия были созданы соответствующие метки путем ручного сопоставления угловых точек связанных пар изображений. Для каждой

пары были рассчитаны векторы смещения, которые указывают на смещение ключевых точек между исходным и преобразованным изображениями. Эти метки затем использовались для обучения модели прогнозированию смещений угловых точек, которые затем использовались для создания матрицы гомографии для точного выравнивания изображения.

Такой методологический подход позволил создать надежный синтетический набор данных, включающий систематически изменяемые и трансформируемые изображения, которые точно имитируют реальные условия получения изображений при проведении фундоскопии с помощью смартфона.

### III. МОДЕЛИ И МЕТОД

В этом разделе мы представляем предлагаемый метод сшивания изображений сетчатки, основанный на структурированном конвейере, состоящем из четырех ключевых этапов: выделение признаков, корреляция признаков, оценка гомографии и сшивание изображений.

#### *А. Извлечение объектов*

Для извлечения отличительных признаков из пар изображений сетчатки глаза используется сеть Siamese Mobile-Net V2, благодаря ее эффективности на мобильных устройствах. В Mobile-Net V2 используются разделяемые по глубине свертки, которые значительно сокращают количество параметров и сложность вычислений, сохраняя при этом высокую точность. Подсети сиамской сети извлекают карты объектов, которые затем сравниваются с использованием метрик расстояния L2 для определения сходства изображений. Это позволяет модели оценивать геометрические соотношения между изображениями, что необходимо для будущей оценки гомографии.

#### *В. Корреляция признаков*

После выделения признаков для сравнения извлеченных карт объектов используется слой корреляции признаков. На этом этапе выравниваются структурно сопоставимые области между перекрывающимися изображениями сетчатки, что повышает надежность метода оценки гомографии.

#### *С. Оценка гомографии*

В контексте сшивания изображений на сетчатке глаза оценка гомографии предполагает оценку смещения четырех угловых точек одного изображения относительно другого. В отличие от подхода, описанного в [9], в котором используются четыре полусимметричные подсети CNN, каждая из которых отвечает за прогнозирование смещения для одной угловой точки, в этом исследовании рекомендуется использовать упрощенную модель CNN для прогнозирования всех четырех смещений одновременно. Этот метод значительно сокращает вычислительные затраты, что делает его идеальным для мобильных приложений с ограниченной вычислительной мощностью.

На основе извлеченных объектов модель прогнозирует смещения угловых точек, которые затем используются для создания матрицы гомографии с помощью метода прямого линейного преобразования

(DLT) [10]. Это обеспечивает точное выравнивание изображения за счет точной регистрации угловых точек.

#### D. Деформация изображения и сшивание

Как только модель предсказывает смещение угловых точек, результаты используются для вычисления матрицы гомографии, которая выравнивает второе изображение относительно первого. Метод включает в себя проекционное преобразование второго изображения, чтобы убедиться, что его угловые точки совпадают с угловыми точками эталонного изображения.

Каждое изображение размером  $128 \times 128$  пикселей помещается в черную рамку размером  $274 \times 274$  пикселя для обеспечения плавного сшивания. Затем используется матрица обратной гомографии для проецирования пикселей со второго изображения на новые места. Эта модификация обеспечивает точное выравнивание при сохранении анатомических особенностей.

Коэффициенты нормализации используются для правильного масштабирования координат, преобразуя однородные координаты обратно в декартовы. Математическое преобразование представлено в уравнении (1):

$$(x', y', w')^T = H * (x, y, w)^T$$

- $(x, y)$  — это пиксельные координаты второго изображения перед преобразованием.
- $(x', y')$  — являются преобразованными координатами пикселей после применения матрицы гомографии  $H$
- $w$  и  $w'$  — это нормализующие факторы, обеспечивающие правильное масштабирование в проективном пространстве.

Наконец, для объединения совпадающих изображений используется алгоритм сшивания, в результате чего получается бесшовное панорамное изображение сетчатки с более широким обзором сетчатки.

На рис. 1 показан весь рабочий процесс предлагаемой модели сшивания изображений сетчатки, который включает этапы от выделения признаков до окончательного выравнивания изображения. На этом рисунке показан процесс глубокого обучения, подчеркивающий важность каждого этапа обработки для достижения высокой точности сшивания изображений на сетчатке.

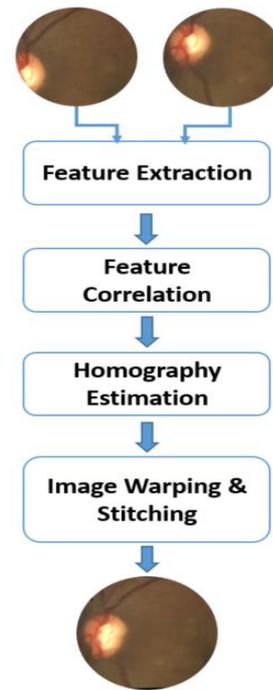


Рис. 1. Схема технологического процесса предлагаемого способа

На рис. 1 показана блок-схема предлагаемой модели, которая описывает последовательные процессы, задействованные в процессе сшивания изображений. Метод начинается с выделения объектов с использованием сети Siamese Mobile Net V2, за которой следует корреляция объектов, при которой сравниваются карты извлеченных объектов. Следующим шагом является оценка гомографии, которая включает в себя упрощенную модель CNN, предсказывающую смещение для каждой из четырех угловых точек. Эти предсказания используются для создания матрицы гомографии, которая помогает в процессе деформации изображения и сшивания, в результате чего получается правильно выровненное и безупречно сшитое изображение сетчатки.

#### IV. РЕЗУЛЬТАТЫ

А. Производительность предложенного метода оценивалась с использованием стандартных метрик оценки:

- Peak Signal-to-Noise Ratio (PSNR): измеряет качество реконструкции изображения (чем выше, тем лучше).
- Structural Similarity Index (SSIM): оценивает сохранение структуры (чем ближе к 1, тем лучше).
- Root Mean Square Error (RMSE): количественно определяет точность выравнивания (чем ниже, тем лучше).

ТАБЛИЦА I. РЕЗУЛЬТАТЫ КОЛИЧЕСТВЕННОЙ ОЦЕНКИ ПРОИЗВОДИТЕЛЬНОСТИ НА ОТКРЫТЫХ НАБОРАХ ДАННЫХ

Набор данных	RMSE	PSNR (dB)	SSIM
APTOS, FIRE	3.54	37.29	0.88
Retina Bank (Smartphone-based)	3.78	36.70	0.87

В. Результаты, представленные в табл. 1, показывают, что предложенный метод является высокоэффективным для обработки изображений сетчатки на наборах данных APTOS, FIRE и Retina Bank. Метрики оценки (RMSE, PSNR и SSIM) демонстрируют высокое качество реконструкции и близкое сходство с эталонными данными. Метод достиг RMSE 3.54 и 3.78 для наборов данных APTOS, FIRE и Retina Bank (на основе смартфона) соответственно, а также PSNR 37.29 дБ и 36.70 дБ и SSIM 0.88 и 0.87. Значения SSIM немного варьируются (среднее: 0.88, минимум: 0.74, максимум: 0.97), что можно объяснить различиями в условиях освещения и качестве изображений среди наборов данных. Все показатели имеют минимальные стандартные отклонения, что подтверждает стабильность и надежность предложенного подхода.

#### V. ЗАКЛЮЧЕНИЕ

Предложенный подход к сшивке изображений сетчатки, основанный на оценке гомографии и легковесной архитектуре CNN, демонстрирует высокую эффективность при обработке изображений сетчатки, снятых на мобильных устройствах. Метод значительно улучшает показатели RMSE и PSNR по сравнению с предыдущими подходами, обеспечивая при этом высокое качество реконструкции. Хотя значения SSIM могут быть несколько ниже в некоторых случаях, предложенная модель работает эффективно с минимальными вычислительными затратами, что делает её идеальной для мобильных приложений.

Результаты показывают, что модель способна генерировать бесшовные панорамные изображения сетчатки, преодолевая значительные ограничения, такие как узкое поле зрения и низкое разрешение при использовании смартфонов для фундоскопии. Более

того, вычислительная эффективность метода делает его подходящим для реализации в режиме реального времени при проведении офтальмодиагностики.

В будущих исследованиях следует уделить особое внимание сохранению структурных деталей при минимизации шумовых артефактов. Кроме того, внедрение адаптивных методов обучения и оптимизация модели для более быстрого вывода на мобильных устройствах могут повысить её практическую применимость. Расширение обучающего набора данных за счёт более разнообразных изображений сетчатки и тестирование метода в различных условиях съёмки могут способствовать повышению его устойчивости и обобщающей способности.

#### СПИСОК ЛИТЕРАТУРЫ

- [1] Eye care, vision impairment and blindness [Electronic resource]. URL: <https://www.who.int/health-topics/blindness-and-vision-loss> (accessed: 09.03.2025).
- [2] A foundation model for generalizable disease detection from retinal images | Nature [Electronic resource]. URL: <https://www.nature.com/articles/s41586-023-06555-x> (accessed: 09.03.2025).
- [3] Corr R.H. Fundoscopy in the smartphone age: current ophthalmoscopy methods in neurology // Arq Neuropsiquiatr. 2023. Vol. 81, № 5. P. 502–509.
- [4] Zhang Z. Research on Different Feature Matching Algorithms for Panoramic Image Stitching. Atlantis Press, 2024. P. 730–743.
- [5] A view-free image stitching network based on global homography - ScienceDirect [Electronic resource]. URL: <https://www.sciencedirect.com/science/article/abs/pii/S1047320320301784> (accessed: 09.03.2025).
- [6] APTOS 2019 Blindness Detection | Kaggle [Electronic resource]. URL: <https://www.kaggle.com/competitions/aptos2019-blindness-detection> (accessed: 09.03.2025).
- [7] FIRE: Fundus Image Registration Dataset [Electronic resource]. URL: <https://projects.ics.forth.gr/cvrl/fire/> (accessed: 09.03.2025).
- [8] Home - Retina Image Bank [Electronic resource]. URL: <https://imagebank.asrs.org/> (accessed: 09.03.2025).
- [9] Image Stitching of Low-Resolution Retinography Using Fundus Blur Filter and Homography Convolutional Neural Network [Electronic resource]. URL: <https://www.mdpi.com/2078-2489/15/10/652> (accessed: 09.03.2025).
- [10] Hartley R., Zisserman A. Multiple View Geometry in Computer Vision. 2nd ed. Cambridge: Cambridge University Press, 2004. ISBN 9780511811685. DOI: <https://doi.org/10.1017/CBO9780511811685>.