

Сравнительная оценка методов глубокого обучения и традиционных методов для сшивки изображений сетчатки на основе мобильных устройств

Али Маея

*Санкт-Петербургский государственный
электротехнический университет
«ЛЭТИ» им. В.И. Ульянова (Ленина)*

Alimayya1357@gmail.com

Зейн Муса

*Московский государственный технологический
университет «СТАНКИН»*

zeinmoussa1717@gmail.com

Аннотация. В данной статье представлено сравнительное исследование традиционных методов сшивания изображений сетчатки на основе признаков, областей и глубокого обучения с акцентом на фундус-изображениях низкого разрешения, полученных с помощью смартфонов. Традиционные методы, такие как регистрация на основе SURF и MODE и согласование шаблонов на основе взаимной информации, сравнивались с глубокими сверточными нейронными сетями, специально разработанными для оценки гомографии. Эксперименты на множестве наборов данных из государственных и полуклинических клиник показали, что хотя традиционные подходы по-прежнему надежны в сценариях контролируемой визуализации, подходы глубокого обучения всегда лучше в отношении точности сшивания, устойчивости к шумам и способности адаптироваться к разнообразным сценариям визуализации. В частности, бережливые глубокие модели обеспечивают превосходный PSNR и качество структуры, что подтверждает их пригодность для решения задач мобильной телеофтальмологии в режиме реального времени. Эти результаты свидетельствуют о большом потенциале глубокого обучения для улучшения портативной глазной диагностики и открывают новые возможности для будущих разработок в области реконструкции панорам в реальном времени и гибридной архитектуры регистрации.

Ключевые слова: анализ здоровья сетчатки глаза, сшивание изображений сетчатки, фундус-изображение со смартфона, глубокое обучение, оценка гомографии, мобильная телеофтальмология, регистрация по признакам, согласование шаблонов, обработка изображений с низким разрешением

I. ВВЕДЕНИЕ

Расстройства зрения и глазные болезни являются важной проблемой здравоохранения во всем мире. По данным Всемирной организации здравоохранения (ВОЗ), более 2,2 миллиарда человек во всем мире страдают от нарушений зрения или слепоты, и по крайней мере один миллиард из них можно предотвратить или вылечить с помощью раннего вмешательства [1]. Традиционные методы исследования сетчатки, такие как прямая фундоскопия, хотя и эффективны с клинической точки зрения, но ограничены своей доступностью, дешевой и требованием специальных навыков, особенно в

отдаленных или недостаточно обслуживаемых сообществах [2].

Недавнее распространение и внедрение устройств для визуализации сетчатки на базе смартфонов, таких как небольшие и доступные устройства, например, приставка D-EYE, произвело революцию в телеофтальмологии, обеспечив портативные, неинвазивные и масштабируемые решения для скрининга глазного здоровья [3, 4]. Однако такие системы обычно страдают от таких ограничений, как низкое пространственное разрешение, небольшие поля зрения и отсутствие стабильности в условиях освещенности. Эти факторы могут значительно ухудшить качество и диагностическую достоверность полученных фундус-изображений [3–6].

Одним из решений, позволяющих избавиться от этих ограничений, является автоматическое сшивание нескольких частично перекрывающихся изображений сетчатки в полную панораму. Правильно сшитое фундус-изображение позволяет получить более широкий обзор сетчатки, что улучшает выявление таких патологий сетчатки, как диабетическая ретинопатия, глаукома и макулярная дегенерация. Однако мозаика изображений на основе камер смартфонов имеет свои уникальные проблемы: слабая видимость объектов, деградация при движении, неоднородная яркость и малотекстурные области делают традиционные методы регистрации, основанные на ручном сопоставлении ключевых точек (например, SIFT, SURF), довольно сложными [7].

Разработки в области глубокого обучения и оценки гомографии позволили создать мощные альтернативы, обеспечивающие точные геометрические преобразования между низкокачественными фундус-изображениями без явного определения особенностей. Глубокие сверточные нейронные сети (CNN), в частности сямские архитектуры в сочетании с регрессионными головками для оценки смещения углов, продемонстрировали превосходную производительность в реальных мобильных приложениях для визуализации сетчатки [8].

Несмотря на эти технологические достижения, беспристрастное и глубокое сравнение традиционных подходов на основе признаков, подходов оптимизации на основе областей и подходов на основе глубокого

обучения для изображений сетчатки, генерируемых смартфоном, еще не изучено. Сравнительный анализ необходим для того, чтобы предложить рекомендации по выбору метода с учетом специфических потребностей приложения, таких как вычислительная мощность, требования к точности и целевые параметры развертывания.

В этой статье мы проводим сравнительную оценку некоторых методов сшивания изображений сетчатки, включая традиционный метод на основе SURF и MODE (ASRPS) [7], метод глубокого обучения, дополненный фильтром размытия фундуса [8], метод на основе регионов, использующий сокращение размерности и взаимную информацию (RetinaMatch) [9], и легкую сеть оценки гомографии на основе MobileNetV2 [10]. Мы оцениваем эти методы на данных, полученных с помощью смартфона, и клинических данных, используя показатели PSNR, SSIM и RMSE, а также качественный анализ. Наша работа показывает компромиссы между вычислительной сложностью, качеством швов и надежностью, предоставляя практическим мобильным офтальмологическим реализациям рекомендации.

II. МАТЕРИАЛ И МЕТОДЫ

A. Наборы данных изображений сетчатки глаза

Набор данных D-EYE состоит из видеок кадров низкого разрешения, которые были получены с помощью фундус-камеры смартфона D-EYE. Благодаря своей низкой стоимости и портативности система D-EYE подходит для применения в пунктах оказания медицинской помощи, однако ее изображениям свойственны сильный шум, малый угол обзора и неравномерное освещение. Этот набор данных был использован специально для оценки методов сшивания на низкокачественных изображениях, полученных с помощью мобильных устройств [8].

Набор данных STARE состоит из изображений сетчатки с сегментацией сосудов и аннотациями патологий. Изначально он использовался для оценки алгоритма RetinaMatch для сопоставления на основе шаблонов с учетом условий малого поля зрения [9].

APTOS 2019 Blindness Detection Dataset, Asia Pacific Tele-Ophthalmology Society предоставляет этот набор данных, состоящий из 3 662 учебных и 1 928 тестовых фундус-изображений с метками степени тяжести диабетической ретинопатии. Это изображения высокого разрешения, полученные в относительно контролируемых клинических условиях. Этот набор данных использовался в основном для моделирования базовых условий сшивания, при которых качество изображения сохраняется [10, 11].

FIRE (Fundus Image Registration Dataset) состоит из 134 пар изображений сетчатки с аннотациями истинности для задач регистрации изображений. Изображения включают вариации поля зрения, морфологию сосудов и патологические паттерны. FIRE является стандартным эталонным набором данных для алгоритмов регистрации и обеспечивает хорошую базу для традиционных и основанных на глубоком обучении методов сшивания [8, 10, 12].

Банк изображений сетчатки (RIB), разработанный Американским обществом специалистов по сетчатке,

представляет собой базу данных изображений фундуса смартфона в естественных условиях с высоким качеством изображения, уровнем шума и вариабельностью видимости анатомии. Изображения похожи на реальные телеофтальмологические случаи и использовались для проверки устойчивости алгоритма к некачественным входным данным [10, 13].

B. Подготовленность и подготовка данных

В каждом методе данные обрабатывались и подготавливались подходящим образом для метода сшивания, предложенного в соответствующей статье, чтобы обеспечить последовательную оценку всех методов, все изображения сетчатки были изменены до размера 128×128 или 274×274 пикселей, в зависимости от требований модели. Методы глубокого обучения, ориентированные на исходные данные низкого разрешения, такие как методы, представленные в [9, 10], генерировали перекрывающиеся фрагменты изображений для имитации условий съемки смартфоном. В отличие от них, традиционные методы, основанные на признаках [7,9], работали на клинических изображениях более высокого качества, где богатая структура сосудов способствовала более надежному сопоставлению признаков. Это различие отражает большую зависимость традиционных методов от качества изображения.

Для повышения обобщенности модели применялось расширение изображения, включая геометрические преобразования, фотометрические корректировки и введение шума. В частности, в [8] введен фильтр Fundus Blur Filter для имитации неравномерного освещения, характерного для фундус-изображений смартфонов, что повышает устойчивость модели. Матрицы гомографии, соответствующие истине, генерировались путем ручной аннотации или синтетического преобразования с использованием метода прямого линейного преобразования (DLT)[8,10]. Все изображения были нормализованы к нулю.

III. МЕТОД

A. A. Традиционный подход, основанный на характеристиках: ASRPS (Automatic Sequential Retinal Picture Stitching)

Алгоритм ASRPS, представленный в [7], использует типичный конвейер регистрации изображений на основе признаков, оптимизированный для многопоследовательного мозаичного сканирования сетчатки. Детектор SURF (Speeded-Up Robust Features) используется для получения точек признаков из зеленого канала предварительно обработанных изображений с повышенным контрастом («сосуд-фон»). Сопоставление выполняется с использованием адаптивной стратегии поиска, разбиения изображений на субрегионы и поиска соответствий с геометрическими ограничениями. Для устранения несопадающих точек и повышения устойчивости к выбросам используется метод MODE (Mode of Offset Distribution Estimation). Слияние изображений осуществляется по правилу максимального значения для плавного перехода интенсивности.

B. Сопоставление шаблонов по площади: RetinaMatch

RetinaMatch [9] решает проблему регистрации изображений сетчатки на основе шаблонов, сочетая уменьшение размерности и оптимизацию взаимной информации (MI). Сначала метод использует анализ

главных компонент (PCA) для грубой локализации шаблонов с малым полем зрения на больших эталонных фундус-изображениях. После локализации взаимная информация между деформированным шаблоном и целевым изображением оптимизируется для улучшения регистрации, что избавляет от локальных оптимумов, которыми страдают большинство методов, основанных на области. RetinaMatch особенно эффективен при регистрации небольших, зашумленных или деградированных участков, где традиционное извлечение признаков не дает результатов.

С. Оценка гомографии на основе глубокого обучения: Фундус-размытие CNN

Сантос и др. [9] предлагают методику глубокого обучения с прямой регрессией параметров гомографии по парам фундус-изображений с низким разрешением. После этого следует этап извлечения признаков с помощью сиамской конволюционной нейронной сети и четырех отдельных ветвей CNN с единственной целью - предсказать смещение каждой из угловых точек. Одним из нововведений является применение фильтра Fundus Blur Filter (FBF) во время обучения, который добавляет переменное освещение и ухудшение периферии для повышения устойчивости модели к изображениям фундуса, полученным с помощью смартфона. Матрицы гомографии рассчитываются по оцененным смещениям и используются для деформации и сшивания входных изображений.

Д. Легкая оценка гомографии на основе искусственного интеллекта (наш метод)

В нашей предыдущей работе [10] мы представили легкую, оптимизированную для мобильных устройств глубокую модель для сшивания изображений сетчатки. В модели используется сиамская сеть MobileNetV2 для эффективного извлечения признаков с помощью разделительных сверток по глубине для минимизации вычислительной сложности. Затем эти карты признаков сравниваются с помощью корреляционного слоя, основанного на расстоянии L2, и используется один регрессор CNN, который предсказывает одновременно все четыре смещения угловых точек. Затем матрицы гомографии оцениваются с помощью алгоритма DLT с учетом этих предсказаний. Данная архитектура имеет низкую вычислительную нагрузку, что позволяет работать с ней в реальном времени на мобильных системах.

IV. РЕЗУЛЬТАТЫ

Каждый алгоритм подвергался количественной и качественной оценке на различных коллекциях изображений сетчатки. Для сравнения были приняты три основных показателя:

- Пиковое отношение сигнал/шум (PSNR): измеряет точность восстановления изображения; предпочтительны более высокие значения.
- Индекс структурного сходства (SSIM): Оценивает целостность структурных данных; желательны значения, более близкие к 1.
- Среднеквадратичная ошибка (RMSE): Оценивает ошибку геометрического выравнивания; более низкие показатели означают более качественную сшивку.

А. Количественная оценка

Количественная оценка в данном исследовании означает объективное измерение качества сшивки с помощью установленных числовых показателей, включая пиковое отношение сигнал/шум (PSNR), индекс структурного сходства (SSIM) и среднеквадратичную ошибку (RMSE). Эти метрики обеспечивают стандартный способ оценки достоверности изображения, сохранения структуры и точности геометрического выравнивания для различных методов. В таблице 1 приведены сравнительные результаты, полученные на различных наборах данных сетчатки, которые подчеркивают различия в производительности между традиционными подходами и подходами, основанными на глубоком обучении.

ТАБЛИЦА 1. КОЛИЧЕСТВЕННОЕ СРАВНЕНИЕ ЭФФЕКТИВНОСТИ МЕТОДОВ СШИВАНИЯ СЕТЧАТКИ

Метод	Набор данных	PSNR (dB)	SSIM	RMSE
ASRPS (основанный на характеристиках)	APTOS / FIRE	Не указано в явном виде	Высокая визуальная согласованность	-
RetinaMatch (на основе области)	STARE / D-EYE	N/A (внимание к ошибкам)	Успешность > 94%	Средняя ошибка < 2 пикселей
Фундус-размытие CNN (глубокое обучение)	FIRE / D-EYE	25.46	0.94	25.66
Легкий CNN, управляемый искусственным интеллектом (наш)	APTOS / FIRE / Retina Bank	37.29	0.88	3.54

В. Качественные результаты

Качественная оценка показывает, что ASRPS достигает точного выравнивания на высококачественных клинических изображениях, но демонстрирует заметные несоответствия при применении к наборам данных с низким разрешением или шумом. RetinaMatch успешно выполняет локальное сопоставление шаблонов для небольших участков, сохраняя согласованность выравнивания в ограниченном поле зрения. Результаты работы Fundus Blur CNN демонстрируют сохранение структуры и высокую визуальную когерентность в условиях ухудшения качества изображения, несмотря на незначительные эффекты сглаживания восприятия. Управляемая искусственным интеллектом облегченная модель CNN создает сшитые изображения с высоким качеством восприятия, сохраняя резкость и целостность даже в сценариях с низким уровнем текстур и шумов.

С. Обсуждение

Количественные и качественные оценки выявили очевидные различия между рассмотренными методами сшивания. Традиционные подходы на основе признаков, например ASRP [7], эффективно работали с изображениями высокого разрешения, но не оправдали ожиданий при работе с зашумленными или низкотекстурными изображениями, полученными со смартфона. RetinaMatch [9] обеспечивает быстрое

локальное сопоставление шаблонов, но не масштабируется на полные мозаики сетчатки.

Методы, основанные на глубоком обучении, неизменно оказывались лучше традиционных. Надежное сохранение структуры было достигнуто с помощью Fundus Blur CNN [8] с SSIM 0,94, хотя и при низких показателях PSNR. Наша CNN, управляемая легким ИИ [10], показала самый высокий PSNR (37,29 дБ) при SSIM в приемлемом диапазоне (0,88), создавая визуально целостные сшитые изображения даже в сложных условиях.

Эти результаты подтверждают, что глубокое обучение предлагает эффективный и масштабируемый метод для срачивания низкокачественных изображений сетчатки. Тем не менее, потенциальные улучшения могут быть сделаны в направлении обобщения модели, оптимизации для работы в режиме реального времени и изучения гибридных стратегий, сочетающих глубокое обучение с традиционными методами для повышения клинической надежности.

В. ЗАКЛЮЧЕНИЕ

В данном исследовании проведена сравнительная оценка традиционных методов мозаики изображений сетчатки, основанных на признаках, областях и глубоком обучении, с особым упором на фундус-изображения низкого разрешения, полученные с помощью смартфона. Результаты ясно показывают, что методы глубокого обучения лучше традиционных подходов, особенно в плохих условиях с шумом, низким контрастом и ограниченным полем зрения.

Несмотря на то, что традиционные методы, такие как ASRPS [7], и методы на основе областей, такие как RetinaMatch [9], остаются компетентными для структурированных высококачественных данных, их производительность сильно снижается при получении низкокачественных входных данных. Глубокие конволюционные сети [8][10], однако, демонстрируют улучшенную точность сшивания, устойчивость к изменчивости и общую обобщенность для широкого набора параметров получения данных, и поэтому представляют собой технологию выбора в мобильных и телеофтальмологических приложениях.

Это достижение также открывает потенциальные возможности для будущих исследований. Методы, основанные на характеристиках, по-прежнему могут иметь преимущества при работе с высокоструктурированными данными, а методы, основанные на тонких областях, могут быть

удовлетворительными, если обеспечивается непрерывность освещения. Однако модели на основе глубокого обучения, особенно модели оценки тонкой гомографии, имеют наибольшие перспективы для применения в мобильных устройствах. Будущая работа может быть сосредоточена на архитектурах слияния, объединяющих обучение признакам и минимизацию взаимной информации, оптимизации моделей для выполнения в реальном времени на смартфонах и более эмпирическом тестировании в противоречивых клинических условиях.

Наконец, сшивание изображений на основе глубокого обучения – важная веха на пути к доступности, точности и масштабируемости диагностики сетчатки с помощью мобильных технологий.

СПИСОК ЛИТЕРАТУРЫ

- [1] “Why use telehealth? | Telehealth.HHS.gov [Electronic resource]. URL: <https://telehealth.hhs.gov/patients/why-use-telehealth> (accessed: 01.03.2024).
- [2] Gajarawala S.N., Pelkowski J.N. Telehealth Benefits and Barriers // *J Nurse Pract.* 2021. Vol. 17, № 2. P. 218–221.
- [3] Neto A., Camara J., Cunha A. Evaluations of Deep Learning Approaches for Glaucoma Screening Using Retinal Images from Mobile Device // *Sensors (Basel).* 2022. Vol. 22, № 4. P. 1449.
- [4] Shanmugam M.P. et al. Fundus imaging with a mobile phone: a review of techniques // *Indian J Ophthalmol.* 2014. Vol. 62, № 9. P. 960–962.
- [5] Eye care, vision impairment and blindness [Electronic resource]. URL: https://www.who.int/health-topics/blindness-and-vision-loss#tab=tab_1 (accessed: 05.03.2024).
- [6] A foundation model for generalizable disease detection from retinal images | *Nature* [Electronic resource]. URL: <https://www.nature.com/articles/s41586-023-06555-x> (accessed: 10.03.2024).
- [7] Yang H. et al. An Automatic Algorithm for Stitching Multi-Sequence Retinal Images: SSRN Scholarly Paper 4188695. Rochester, NY: Social Science Research Network, 2022.
- [8] Santos L. et al. Image Stitching of Low-Resolution Retinography Using Fundus Blur Filter and Homography Convolutional Neural Network: 10 // *Information. Multidisciplinary Digital Publishing Institute*, 2024. Vol. 15, № 10. P. 652.
- [9] Gong C. et al. RetinaMatch: Efficient Template Matching of Retina Images for Teleophthalmology: arXiv:1811.11874. arXiv, 2018.
- [10] [A. S. Mayya. (2025). AI-Driven Stitching of Retinal Smartphone-Based Images for Automated Fundoscopy. SCM2025 Conference.
- [11] APTOS 2019 Blindness Detection | Kaggle [Electronic resource]. URL: <https://www.kaggle.com/competitions/aptos2019-blindness-detection> (accessed: 09.03.2025).
- [12] FIRE: Fundus Image Registration Dataset [Electronic resource]. URL: <https://projects.ics.forth.gr/cvrl/fire/> (accessed: 09.03.2025).
- [13] Home-Retina Image Bank [Electronic resource]. URL: <https://imagebank.asrs.org/> (accessed: 09.03.2025).