# ТЕОРЕТИЧЕСИКЕ ОСНОВЫ РАЗРАБОТКИ ПРОГРАММНО-АЛГОРИТМИЧЕСКОГО КОМПЛЕКСА АНАЛИЗА ОКТ-ИЗОБРАЖЕНИЙ СЕТЧАТКИ

## Описание биотехнической системы

Биотехническая система, в которой предполагается использовать разрабатываемый программный комплекс, представлена на рисунке 1.

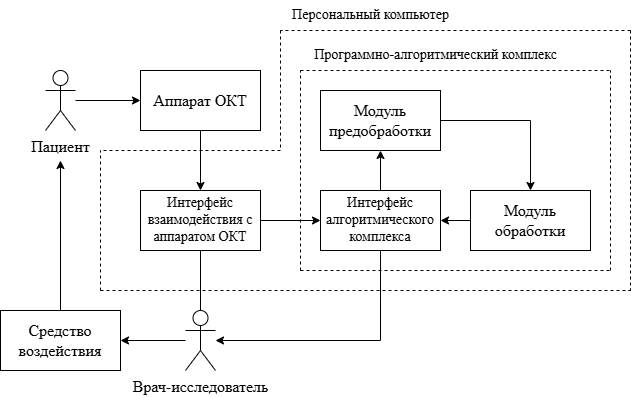


Рисунок 1 – Биотехническая система

Основными компонентами системы являются элементы, представленные далее:

1. Пациент – биологический объект, по отношению к которому применяются диагностические методы исследования для определения его состояния.
2. Аппарат ОКТ – оптический когерентный томограф, применяемый для получения изображения сетчатки глаза пациент. Аппарат ОКТ производит зондирование биологических тканей оптическим излучением ближнего инфракрасного диапазона. Отраженная и опорная световые волны создают интерференционную картину, анализ которой позволяет визуализировать морфологическое строение анализируемого биологического объекта. Результирующее изображение поступает на «Интерфейс взаимодействия с аппаратом ОКТ».
3. Интерфейс взаимодействия с аппаратом ОКТ – пользовательский интерфейс, позволяющий управлять аппаратом ОКТ: выбирать функции анализа, выбирать режим и тип съемки, выбирать методы коррекции, просматривать и анализировать полученное изображение. Далее, через интерфейс сетевого взаимодействия, полученное изображение может быть предано разрабатываемому программно-алгоритмическому комплексу для проведения аналитических преобразования, позволяющих врачу-офтальмологу получить статистическую информацию о возможных патологиях.
4. Интерфейс алгоритмического комплекса – пользовательский интерфейс, позволяющий взаимодействовать с разрабатываемым программно-алгоритмическим комплексом. Пользовательский интерфейс позволят загрузить изображения, провести анализ изображения, просмотреть статистические данные о возможных патологиях, полеченные в результате проведения анализа изображения, просмотреть статистические данные анализа каждого проанализированного изображения за данную сессию. При выборе функции анализа изображения, оно поступает на «Модуль предобработки».
5. Модуль предобработки – совокупность алгоритмических преобразований ОКТ-изображения, позволяющих выделить анализируемую область сетчатки. После проведения предварительной обработки, результирующее ОКТ-изображения поступает на «Модуль обработки».
6. Модуль обработки – модель нейронной сети, предварительно обученная на большом количестве различных ОКТ-изображений сетчатки, содержащих патологии, для проведения сегментации ОКТ-изображения сетчатки – выделения патологических участков изображения. Результатом работы модуля обработки является объект, содержащий координаты и контура выделенных областей, их классификацию и вероятность правильной классификации. Полученный объект, далее, передаётся «Модулю формирования заключения».
7. Модуль формирования заключения – совокупность алгоритмов, преобразующих объект, полученный от «Модуля обработки» в объект заключения. Модуль формирования заключения разбирает полученные вероятностные классификации патологий анализируемого изображения, сравнивает классифицированные патологии с референтными таблицами и формирует вероятностное заключение на основе вероятностей верной классификации патологий. Далее, сформированное заключение отправляется на «Интерфейс алгоритмического комплекса», что позволяет врачу-исследователю ознакомиться с результатами проведенного анализа.
8. Врач-исследователь – квалифицированный специалист в области офтальмологии, осуществляющий применение разрабатываемого программно-алгоритмического комплекса в качестве инструмента интеллектуальной поддержки процесса принятия диагностических решений. В процессе анализа полученных данных врач-исследователь интерпретирует результаты автоматизированной обработки медицинских изображений в совокупности с собственными клинико-диагностическими выводами, на основании чего формирует обоснованное заключение, отражающее интегративную оценку состояния «Пациента». При помощи «Средства воздействия» врач-исследователь обеспечивает восстановление или стабилизацию состояния «Пациента».
9. Средство воздействия – компонент, используемый врачом-исследователем для проведения различных мероприятий для стабилизации или изменения состояния «Пациента», в зависимости от произведенной интегральной оценки его состояния.

## Описание входных данных программно-алгоритмического комплекса

На рисунке 2 представлен пример ОКТ-изображения, подающегося на вход программно-алгоритмическому комплексу.

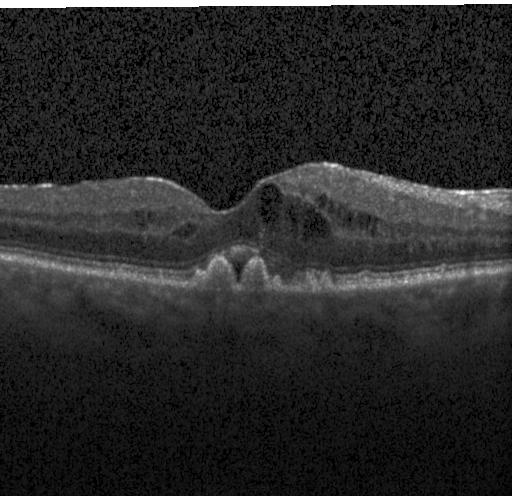


Рисунок 2 – Пример ОКТ-изображения, подающегося на вход программно-алгоритмическому комплексу

Анализируя исходное изображение, можно заметить, что оно содержит некоторую лишнюю информацию, в виде тёмных областей сверху и снизу изображения, которая может негативно повлиять на работу модуля обработки. Для того, чтобы сфокусировать внимание модуля обработки на значимой части анализируемого изображения, применятся модуль предобработки.

## Разработка модуля предобработки

На рисунке 3 представлена общая схема последовательности этапов алгоритмических преобразований, проводимых над анализируемым ОКТ-изображением, для выделения его значимой части. Каждый из этапов предобработки более подробно рассмотрен далее.



Рисунок 3 – Схема алгоритмических преобразований модуля предобработки

### Преобразование к чёрно-белому изображению

В решаемой задаче сегментации морфологических структур биологического объекта цветовая информация не представляет значимой аналитической ценности. При этом обработка цветного изображения требует больше вычислительных ресурсов и времени, чем на обработка черно-белого. Поэтому преобразование изображения в оттенки серого позволяет оптимизировать процесс обработки без потери значимой информации.

Следует отметить, что использование изображения в оттенках серого позволяет оптимизировать работу многих алгоритмов фильтрации изображений, например, алгоритмов выделения контуров и краёв, пороговой фильтрации, так как и значение интенсивности в точке.

На рисунке 4 представлено анализируемое ОКТ-изображение после прохождения первого этапа предобработки.

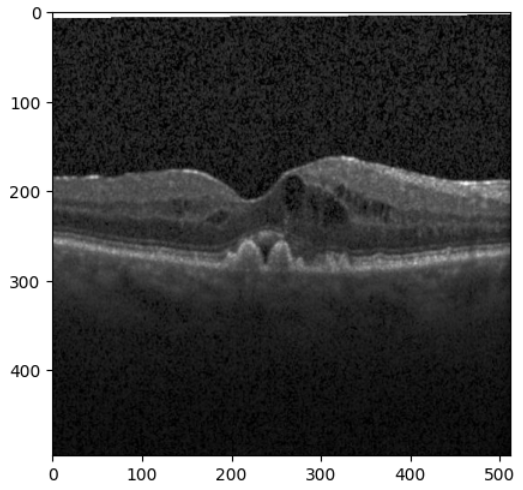


Рисунок 4 – Анализируемое ОКТ-изображение после прохождения первого этапа предобработки

Можно заметить, что рисунок 2, подаваемый на вход программно-алгоритмического комплекса, и рисунок 4, получаемый после прохождения первого этапа предобработки – идентичны. Это обусловлено тем, что рисунок 2 уже преобразован к формату в оттенках серого.

Программно-алгоритмический комплекс разработан с возможностью обработки как черно-белых, так и цветных изображений. В качестве примера на рисунке 5 представлено исходное цветное ОКТ-изображение, в котором различия в интенсивности сигнала визуализированы с помощью цветовой палитры.

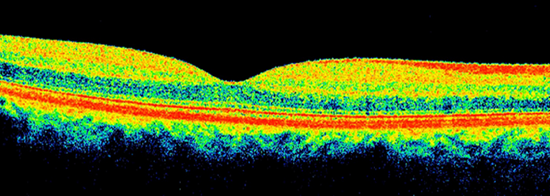


Рисунок 5 – Цветное ОКТ-изображение

### Размытие изображения

На рисунке 4 можно заметить шум типа «salt and pepper» – искажение, проявляющееся в виде чередующихся чёрных и белых точек, случайно распределённых по изображению. Наличие таких шумов существенно снижает качество визуальной информации: они могут имитировать границы объектов или искажать существующие структуры, что негативно влияет на последующий этап выделения краёв.

Для устранения данного типа искажений на первом этапе предобработки применялся медианный фильтр. Принцип его работы заключается в замене значения каждого пикселя медианой значений из его локального окружения (обычно в виде квадратного окна). Такой подход позволяет эффективно подавлять одиночные выбросы яркости – типичные для «соляного и перцового» шума – без существенного смещения краёв или структур на изображении. Медианный фильтр является одним из наиболее надёжных решений при работе с импульсными шумами.

Однако, несмотря на свою эффективность, медианный фильтр может приводить к незначительному размытию мелких деталей и снижению локальной контрастности. Чтобы компенсировать этот эффект и сохранить чёткие границы между анатомическими структурами, на следующем этапе предобработки применялся билатеральный фильтр.

Билатеральный фильтр – это метод нелинейной фильтрации, который одновременно сглаживает изображение и сохраняет резкие переходы интенсивности. Его ключевое отличие от обычных гауссовых фильтров состоит в том, что он учитывает не только пространственную близость пикселей, но и различие в их интенсивности. Это позволяет ограничить сглаживание только внутри однородных областей, сохраняя при этом границы между различными по яркости регионами. Такой подход делает билатеральную фильтрацию особенно эффективной для изображений, в которых важно одновременно удалить шум и сохранить структурные особенности.

На рисунке 6 представлено ОКТ-изображение сетчатки после прохождения второго этапа предобработки.

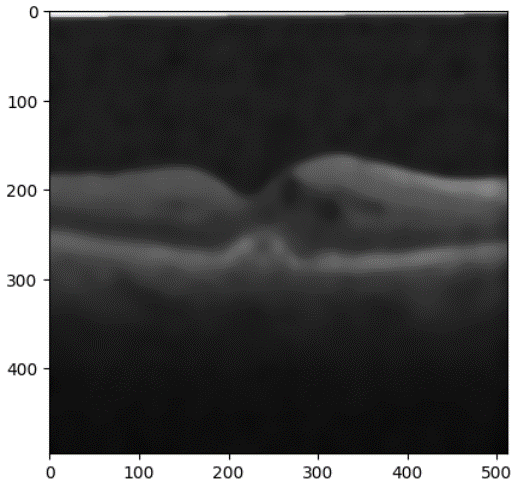


Рисунок 6 – Анализируемое ОКТ-изображение сетчатки после прохождения второго этапа предобработки

### Выделение краёв

Анализируя рисунок 6, можно заметить, что наибольшую аналитическую ценность представляют участки изображения с выраженными перепадами интенсивности, соответствующими границам анатомических структур. Для автоматического выявления таких участков на этапе предобработки применялся оператор Собеля – классический метод выделения границ на изображениях.

Оператор Собеля основан на приближённом вычислении градиента яркости изображения с помощью свёртки с двумя небольшими ядрами (обычно 3×3), чувствительными к изменениям яркости вдоль горизонтального (оси X) и вертикального (оси Y) направлений. Результат свёртки представляет собой приближённое значение производной изображения, то есть степень изменения яркости в данной точке.

В рамках данной задачи оператор Собеля был применён только по вертикальной оси (оси Y), поскольку вертикальные перепады интенсивности наиболее характерны для структур сетчатки, таких как границы между слоями. Это позволяет с высокой точностью выделить горизонтально ориентированные границы, которые имеют наибольшее значение для дальнейшего анализа и классификации.

На рисунке 7 представлено анализируемое ОКТ-изображение сетчатки после прохождения третьего этапа предобработки. на котором видно, как оператор Собеля выделяет наиболее контрастные вертикальные структуры.

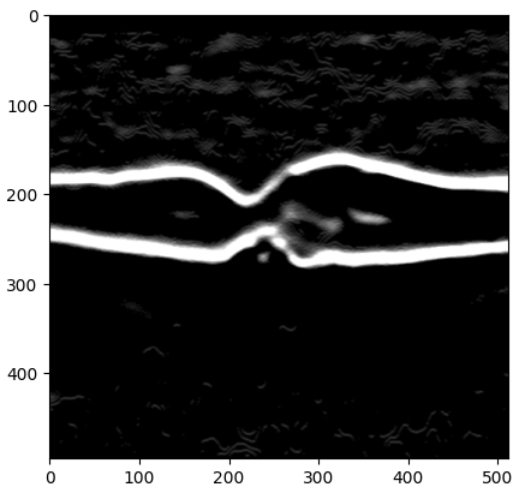


Рисунок 7 – Анализируемое ОКТ-изображение сетчатки после прохождения третьего этапа предобработки

### Пороговая фильтрация

Анализируя рисунок 7, можно заметить, что помимо значимой области были выделены некоторые шумовые артефакты, лучше всего заметные в верхних и нижних частях изображения.

Образовавшиеся артефакты могу негативно сказаться на последующих этапах обработки изображения. Поэтому, для предотвращения возможных негативных воздействий, которые могут оказать появившиеся артефакты, к изображения применяется бинарная фильтрация.

Принцип пороговой фильтрации основан на сравнении значения яркости каждого пикселя с заранее заданным порогом.

Если интенсивность пикселя превышает пороговое значение, он принимает максимальное значение (обычно белый цвет). В противном случае пиксель устанавливается в минимальное значение (обычно чёрный цвет).

На рисунке 8 представлено анализируемое ОКТ-изображения сетчатки после прохождения четвертого этапа предобработки.

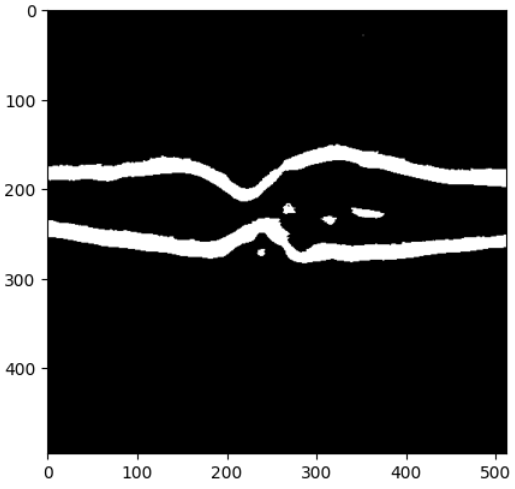


Рисунок 8 – Анализируемое ОКТ-изображения сетчатки после прохождения четвертого этапа предобработки

### Морфологические преобразования

Даже после пороговой фильтрации на изображении могут сохраняться незначительные артефакты, которые не несут аналитической ценности, но могут затруднять последующую обработку. Для устранения таких нежелательных элементов и улучшения структуры бинарного изображения применяются морфологические операции — инструменты математической морфологии, основанные на анализе формы и структуры объектов.

Морфологические преобразования выполняются с использованием специального структурного элемента (обычно небольшая матрица – например, 3×3), который последовательно перемещается по изображению и применяет логические правила к окрестности каждого пикселя.

В рамках данной работы используются две базовые операции: открытие и закрытие, применяемые последовательно.

1. Открытие – операция, состоящая из эрозии с последующим дилатированием. Она позволяет удалить мелкие шумовые объекты, не входящие в состав крупных структур, сохраняя при этом общую форму и размеры значимых областей. Эффективна для устранения мелких изолированных пикселей и сглаживания границ
2. Закрытие – операция, включающая дилатацию, за которой следует эрозия. Применяется для заполнения небольших разрывов и размыканий внутри объектов, а также для сглаживания контуров. Закрытие помогает объединить близко расположенные пиксели в единую область и устранить небольшие тёмные вкрапления в светлых сегментах.

На рисунке 9 представлено анализируемое ОКТ-изображение после прохождения пятого этапа предобработки.

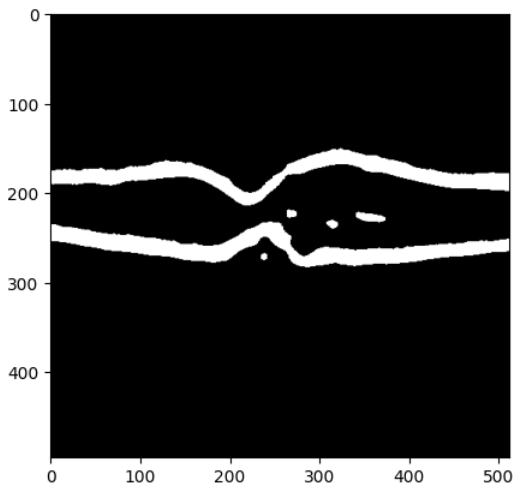


Рисунок 9 – Анализируемое ОКТ-изображение после прохождения пятого этапа предобработки

### Объединение контуров

Анализируя рисунок 9, можно заметить, что полностью избавиться от артефактов не удалось. Этап объединения контуров позволяет отдельно проанализировать каждый из выделенных контуров, провести фильтрацию по некоторому набору параметров и отобрать только конура, соответствующие анализируемой области.

Так, у каждого контура проверяется:

1. Площадь – если площадь контура менее определенной константы, контур помечается как артефактный и не учувствует в дальнейшей обработке.
2. Ширина – если ширина контура менее 1/4 от ширины всего изображения, контур помечается как артефактный и не учувствует в дальнейшей обработке.
3. Схожесть с геометрическими объектами – если в контур может быть вписана прямая линия с минимальной ошибкой по расстоянию от лини до контура, то такой помечается как артефактный и не учувствует в дальнейшей обработке.

Объект с прошедшими фильтрацию контурами передаётся на следующий этап обработки.

### Выделение фрагмента изображения

На основе контуров, выделенных на предыдущем этапе обработки, рассчитывается совокупный ограничивающий прямоугольник (boundingRect) относительно всего анализируемого изображения.

Данный прямоугольник охватывает все найденные контуры и позволяет определить ключевые геометрические параметры: общую ширину и высоту области, а также координаты правого нижнего угла по осям X и Y.

Рассчитанные характеристики используются для автоматизированного определения положения области интереса — участка сетчатки, подлежащего дальнейшему анализу. Это существенно упрощает задачу локализации и обеспечивает фокусировку обработки на информативной части изображения.

На рисунке 10 представлено ОКТ-изображение сетчатки с наложенным ограничивающим прямоугольником, определённым на данном этапе.

Координаты, полученные в результате анализа контуров, позволяют выполнить обрезку исходного изображения по границам ограничивающего прямоугольника. Полученное в результате изображение может быть использовано для обучения алгоритмов, основанных на применении нейронных сетей, где важно наличие чётко локализованной и информативной входной области.

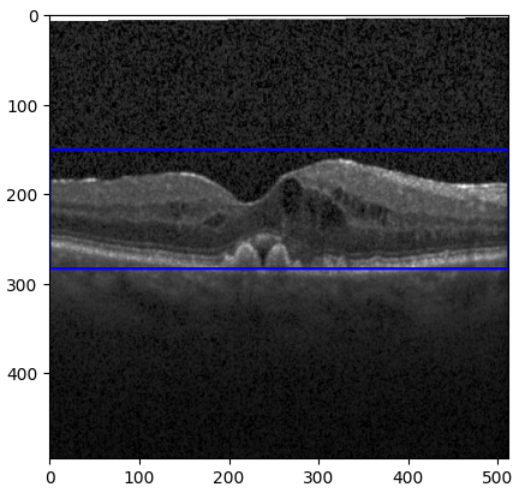


Рисунок 10 – Анализируемое ОКТ-изображение сетчатки с локализованными структурными элементами

## Составление выборки ОКТ-изображений сетчатки для обучения модели нейронной сети

Составление обучающей, тестовой и валидационной выборки выявляется одними из основных этапов разработки программного средства, включающего алгоритмы обработки изображений на основе нейронных сетей.

От качества выборки оказывает существенное влияние процесс и результаты обучения нейросетевых моделей:

* Размер выборки. От количества данных зависит эффективность и универсальность модели. Большее количество данных способствует повышения качества обобщения модели, снижает риск переобучения и позволяет модели лучше выявлять сложные закономерности в данных. Тогда как низкое количество данных снижает эффективность модели про обработки новых, ранее не встречавшихся данных.
* Сбалансированность. Выборка, содержащая равномерное распределение классов и вариантов данных, позволяет практически полностью исключить искажение обобщающей способности модели. Несбалансированная выборка приводит к тому, что модель начинать хуже распознавать редко встречающиеся классы, снижаю общую универсальность и точность системы.
* Качество аннотации. Наличие ошибок в аннотациях, например, неверная классификации или неправильное выделение области класса, ухудшает качество обучения модели. Высокое качество аннотации критично в задачах медицинского скрининга, где точность классификации имеет ключевое значение.
* Качество изображений. Высокое качество входных изображений способствует более эффективному извлечению признаков и ускоряет процесс обучения. Изображения низкого качества затрудняют обучение модели, увеличивают вероятность ошибок во время обработки и могут снизить производительность модели.

### Выбор данных для составления выборки

При выборе данных для обучающей, тестовой и валидационной выборок особое внимание уделялось общему объёму изображений, количеству размеченных патологий, а также возможностям последующего расширения выборки за счёт дополнительной аннотации существующих данных или включения новых аннотированных изображений.

В таблице 1 представлен обзор открытых выборок ОКТ-изображений, рассмотренных для использования в процессе обучения нейронной сети в соответствии с ранее сформулированными требованиями.

Таблица 1 – Характеристики рассмотренных открытых баз данных

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Название датасета | Аннотированные патологии | Добавление новых данных | Расширение существующей аннотации | Количество изображений |
| OPTIMA[] | ИК, СЖ, ОПЭ | Да | Нет | 849 |
| UMN[] | ИК, СЖ, ОПЭ | Да | Нет | 600 |
| RETOUCH[] | ИК, CЖ ОПЭ | Да | Нет | 1037 |
| KERMANY[] | — | — | — | 84495 |

Анализ проведённого обзора существующих открытых датасетов позволил установить, что наилучшим вариантом для использования является выборка RETOUCH, характеризующаяся сравнительно большим объёмом доступных изображений, а также возможностью расширения за счёт добавления новых данных.

Следует отметить, что основным затруднением при расширении данной выборки является проблема поиска изображений, сопоставимых с оригинальными по качеству сканирования, разрешению и размеру. Сохранение указанных параметров является критически важным для обеспечения однородности обучающей выборки, что напрямую влияет на качество обучения и обобщающую способность моделей сегментации.

Учитывая указанные ограничения, было принято решение о самостоятельной аннотации изображений на основе выборки KERMANY. Данный набор данных включает широкий спектр ОКТ-изображений сетчатки, преимущественно отражающих морфологические изменения сетчатки, связанные с диабетическим макулярным отёком (ДМО), влажной формой возрастной макулярной дегенерации (влажная ВМД) и наличием друз. Кроме того, в составе выборки присутствуют изображения, демонстрирующие другие патологические изменения сетчатки, в частности отслойку пигментного эпителия и отслойку стекловидного тела.

Таким образом, самостоятельная аннотация изображений из выборки KERMANY и последующее обучение нейронной сети на расширенной выборке позволит увеличить количество диагностируемых патологий и, как следствие, повысить диагностические возможности разрабатываемого программно-алгоритмического комплекса.

### Выбор сегментируемых патологий ОКТ-изображения сетчатки

В качестве исходных данных использовалась база ОКТ-изображений сетчатки Kermany [], включающая три основных диагностических категории: диабетический макулярный отёк (DME), возрастная макулярная дегенерация (AMD) и друзы.

Следует отметить, что DME и AMD являются самостоятельными офтальмологическими заболеваниями, для которых характерны специфические морфологические изменения сетчатки. В частности, наличие субретинального гиперрефлективного материала и друз является характерным проявлением возрастной макулярной дегенерации, тогда как интраретинальные кисты являются типичным признаком диабетического макулярного отёка.

Кроме того, изображения из разных диагностических классов могут содержать и другие патологические признаки, такие как субретинальная жидкость, отслойка нейроэпителия, отслойка стекловидного тела, ламеллярный макулярный разрыв, витреомакулярная тракция, а также отслойка пигментного эпителия.

Для задачи сегментации были отобраны те классы патологий, которые наиболее широко представлены в выборке. К ним относятся:

* отслойка нейроэпителия,
* отслойка стекловидного тела,
* ламеллярный макулярный разрыв,
* витреомакулярная тракция,
* отслойка пигментного эпителия,
* субретинальный гиперрефлективный материал,
* друзы,
* интраретинальные кисты.

### Аннотация ОКТ-изображений сетчатки

Сформированная выборка была обработана модулем предобработки с целью выделения значимой части изображения.

Аннотация результирующей выборки изображений производилась с помощью программного средства с открытым исходным кодом Label Studio [].

На рисунке 11 изображен интерфейс Label Studio.



Рисунок 11 – Интерфейс Label Studio

На рисунках 12, 13 представлено ОКТ-изображение сетчатки «ДО» и «ПОСЛЕ» проведения аннотации.

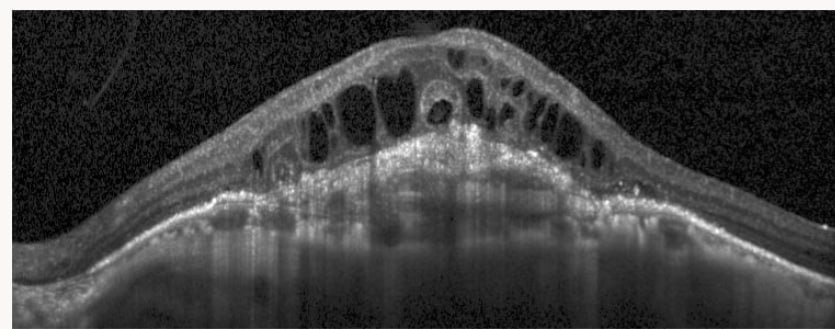


Рисунок 12 – ОКТ-изображение сетчатки «ДО» проведения аннотации

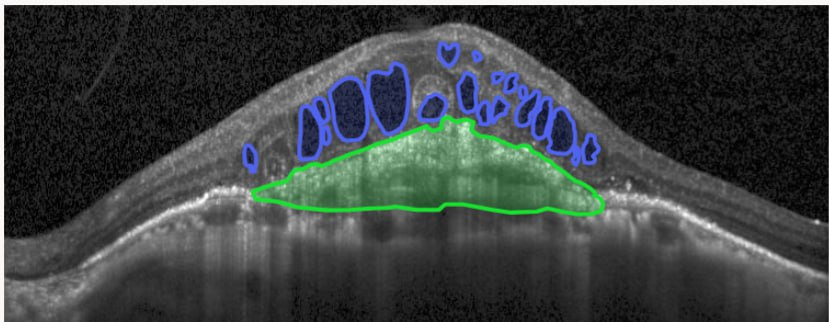


Рисунок 13 – ОКТ-изображение сетчатки «ПОСЛЕ» проведения аннотации

В таблице 2 представлена характеристика сформированной выборки.

Следует отметить, что достоверность аннотаций ОКТ-изображений сетчатки была подтверждена верификацией, проведённой офтальмологом‑специалистом.

Таблица 2 – Характеристика сформированной выборки

|  |  |  |
| --- | --- | --- |
| Тип аннотации | Общее число аннотаций | Общее количество изображений |
| retinal\_drusen (ретинальные друзы) | 1818 | 399 |
| intraretinal\_cyst (интраретинальные кисты) | 1313 |
| subretinal\_hyperreflective\_material (субретинальный гиперрефлективный материал) | 99 |
| neuroepithelium\_detachment (отслойка нейроэпителия) | 36 |
| vitreomacular\_traction (витреомакулярная тракция) | 35 |
| lamellar\_macular\_rupture (ламеллярный макулярный разрыв) | 27 |
| vitreous\_detachment (отслойка стекловидного тела) | 14 |

Сформированную выборку предполагается использовать для обучения модели нейронной сети.

## Разработка модуля обработки

Модулем обработки является модель нейронной сети, производящая сегментацию изображения. Задача модуля обработки – выделить вероятные патологии в морфологической структуре сетчатки, изображенной на обрабатываемом ОКТ-изображении.

Выбор верной архитектуры модели нейронной сети играет одну из ключевых ролей в точности и эффективности сегментации.

В целях выбора наилучшей архитектуры для решения задачи сегментации ОКТ-изображений сетчатки рассматривались несколько наиболее современных моделей нейронных сетей.

### YOLO (You Only Look Once)

YOLO – современная архитектура моделей компьютерного зрения, разработанная Джозефом Редмоном и Али Фархади в 2016.

По своей архитектуре YOLO является одноэтапным детектором, выполняющим идентификацию объектов за одни проход через сеть.

Современные версии моделей YOLO состоят из нескольких модулей:

* Backbone – сеть для первичного извлечения признаков. Backbone принимает на вход исходное изображение и извлекает из него важные признаки. Он состоит из последовательностей сверточных слоёв, нормализации и функций активации, предназначенных для выявления текстур, контуров, форм и других низко- и высокоуровневых особенностей изображения.
* Neck – это промежуточный модуль между backbone и head, задача которого – эффективно собрать и объединить признаки разного масштаба.  
  Это необходимо для успешного обнаружения объектов различных размеров: маленькие, средние и большие объекты могут быть лучше выявлены на разных уровнях детализации.
* Head – выходные слои, предсказывающие классы и координаты рамок. Head принимает признаки, подготовленные neck, и на их основе делает окончательные предсказания.

На данный момент существует 12 версий модели YOLO. В работе рассматривалось 3 новейших архитектуры:

* YOLOv9. Данная модель выделяется своим инновационным подходом к решению проблемы потери информации, присущей глубоким нейронным сетям. Благодаря интеграции Programmable Gradient Information (PGI), сеть способна эффективно сохранять и передавать градиентную информацию между слоями на различных уровнях обработки.

Механизм PGI позволяет адаптивно регулировать потоки градиентов в зависимости от сложности извлекаемых признаков, что способствует более стабильному обучению и снижению эффекта затухания градиентов.

Дополнительно в архитектуре используется модуль GLEAN (Global Local Enhancement Attention Network), который направлен на улучшение качества обработки признаков за счёт совмещения локальной и глобальной информации.

Интеграция GLEAN особенно важна для медицинских изображений, где точность выделения малых патологических областей играет критическую роль. Благодаря GLEAN модель лучше фокусируется на тонких морфологических особенностях, что способствует более точной и надежной сегментации ОКТ-изображений сетчатки.

* YOLOv11. Данная модель отличаетcя инновационным подходом к извлечению признаков и агрегации информации, что решает проблему потери важной информации, характерную для многих глубоких нейронных сетей. В архитектуре YOLOv11 использован улучшенный backbone с блоком C3k2, который эффективно извлекает признаки, обеспечивая более точное детектирование объектов. Это улучшение позволяет сети быстро и с высокой точностью обрабатывать сложные медицинские изображения, сохраняя ключевую информацию о морфологии объектов.

Кроме того, в архитектуре используется механизм пространственного внимания, реализованный через блок C2PSA, что позволяет модели фокусироваться на ключевых областях изображения. Это особенно важно для сегментации мелких патологий на медицинских изображениях, таких как ОКТ, где точность и внимание к деталям критически важны.

Завершающим компонентом является улучшенный head, включающий несколько C3k2 блоков и CBS слоёв, которые позволяют ещё более эффективно обрабатывать извлечённые признаки. Эти усовершенствования способствуют снижению вычислительных затрат, обеспечивая при этом высокую производительность и точность, что особенно важно для работы в реальном времени при анализе медицинских изображений.

* YOLOv12. Архитектура YOLOv12 продолжает традиции YOLO, внедряя несколько значительных нововведений, направленных на повышение эффективности и улучшение производительности при обработке изображений. Одним из ключевых улучшений является Area Attention Mechanism, новый подход к самовниманию, который эффективно обрабатывает большие области рецептивного поля. Этот механизм делит карты признаков на несколько равных областей (по умолчанию 4), что позволяет избежать сложных операций и существенно снизить вычислительные затраты, сохраняя при этом широкий диапазон восприятия. Это особенно важно для задач сегментации, где внимание к различным частям изображения играет важную роль.

Еще одной важной новацией является Residual Efficient Layer Aggregation Networks (R-ELAN), улучшенная версия модуля агрегации признаков, основанная на ELAN. В отличие от стандартных методов, R-ELAN использует остаточные связи на уровне блоков с масштабированием, а также переработанную методику агрегации признаков, что создаёт структуру, схожую с бутылочным горлышком. Эти изменения помогают эффективно справляться с вызовами, связанными с оптимизацией, особенно в больших моделях, ориентированных на внимание, и ускоряют процесс обучения.

Для более эффективной работы с вниманием, YOLOv12 оптимизирует стандартную архитектуру внимания. Включение FlashAttention помогает минимизировать затраты на память, а удаление позиционного кодирования упрощает и ускоряет модель. Также введены изменения в настройке отношения MLP (с 4 до 1.2 или 2), что позволяет более эффективно сбалансировать вычисления между вниманием и слоями прямого распространения. Это улучшает оптимизацию и снижает вычислительные затраты модели, особенно в контексте обработки сложных изображений, таких как медицинские.

Добавление 7x7 разделяемой свертки, известной как "position perceiver", в механизм внимания также способствует улучшению обработки позиционной информации. Это делает модель более универсальной и точной при работе с разнообразными объектами, включая сложные медицинские патологии на ОКТ-изображениях.

YOLOv12, благодаря этим усовершенствованиям, достигает более высокой точности при меньшем количестве параметров по сравнению с предыдущими моделями. Это улучшение баланса между точностью и скоростью особенно полезно для применения в реальных условиях, где важна эффективность.

### SAM2 (Segment Anything Model 2)

SAM2 (Segment Anything Model 2) — это усовершенствованная версия модели для сегментации изображений. SAM2 представляет собой одно из самых передовых решений для автоматической сегментации объектов в изображениях, предлагая высокую точность и гибкость для разных типов данных, включая медицинские изображения, спутниковые снимки и другие области.

SAM2 строится на основе предыдущих достижений в области сегментации, улучшая возможности идентификации объектов и их разделения, используя новые подходы в обучении и обработке данных. Модель включает несколько ключевых особенностей, которые обеспечивают её высокую производительность:

* Модуль внимания – использует новейшие техники самообучения, чтобы фокусироваться на наиболее важных участках изображения. Это позволяет модели точно выявлять мелкие или плохо выраженные объекты, что особенно важно для медицинских изображений, где каждый элемент может иметь значение.
* Гибкость сегментации – SAM2 обучена работать с различными типами изображений, используя подходы, которые минимизируют зависимости от конкретных данных. Это делает её универсальной для широкого спектра задач, включая автоматическую сегментацию патологий на медицинских изображениях, таких как ОКТ-изображения сетчатки.
* Эффективная обработка контекста – SAM2 использует улучшенные методы агрегации признаков для более точного учета контекста изображения. Это позволяет лучше разделять объекты, которые находятся в сложных или пересекающихся областях.
* Быстродействие – несмотря на свою сложность и высокую точность, SAM2 ориентирована на работу в реальном времени, что делает её подходящей для приложений, где важна скорость, например, в клинических системах для диагностики заболеваний.
* Zero-Shot Generalization – это важная особенность SAM2, которая позволяет модели работать с новыми типами объектов или данными, с которыми она не была явно обучена. Используя продвинутые методы обучения, модель может эффективно обрабатывать изображения, содержащие объекты, которые она не видела на обучающих данных, без необходимости дообучения или использования дополнительных меток. Это позволяет SAM2 адаптироваться к новым и разнообразным ситуациям, например, выявлять редкие заболевания на медицинских изображениях, которые не встречались в тренировочных данных.

Благодаря универсальной архитектуре, SAM2 хорошо подходит для анализа как обычных, так и специализированных изображений, включая спутниковые снимки, фотограмметрию и медицинские изображения, такие как ОКТ, МРТ и другие.

Модель SAM2 обладает преимуществами в точности сегментации, особенно в сложных и детализированных изображениях, что делает её мощным инструментом для современных задач компьютерного зрения и анализа изображений.

### DeepLabV3+

DeepLabV3+ – это одна из наиболее продвинутых моделей для семантической сегментации изображений, разработанная исследователями из Google Research. Модель улучшает предыдущие архитектуры, такие как DeepLabV3, предлагая более точные и эффективные решения для задач сегментации, включая медицинские изображения, спутниковые снимки и другие сложные сцены.

DeepLabV3+ основана на использовании новейших методов в области компьютерного зрения, включая расширенные сети глубокого обучения и методы внимательности. Вот несколько ключевых особенностей модели:

* Atrous Spatial Pyramid Pooling (ASPP) – ключевая особенность модели, которая помогает эффективно извлекать признаки на различных уровнях масштаба. ASPP использует атрозные свертки для захвата информации о контексте на разных уровнях разрешения. Это позволяет модели хорошо справляться с задачами, где объекты могут иметь различные размеры, например, в медицинских изображениях, где требуется точная сегментация тканей или патологий.
* Улучшенная декодировка – DeepLabV3+ использует декодер для восстановления пространственного разрешения после применения ASPP, что позволяет более точно локализовать границы объектов. Эта улучшенная декодировка особенно полезна для сегментации мелких объектов и деталей на изображениях, таких как области патологий на ОКТ-изображениях сетчатки.
* Гибкость архитектуры – модель DeepLabV3+ включает в себя эффективное сочетание различных сетевых компонентов, что позволяет ей работать с разнообразными типами данных. В отличие от других моделей, DeepLabV3+ может быть адаптирована для решения как простых, так и сложных задач, таких как сегментация спутниковых снимков или медицинских изображений, включая МРТ и ОКТ.
* Техника пространственного внимания – DeepLabV3+ включает в себя механизмы внимания, которые позволяют модели фокусироваться на наиболее значимых областях изображения. Это особенно важно в задачах, где критична высокая точность сегментации, например, для выделения малых патологий или особенностей на медицинских изображениях.
* Быстродействие и оптимизация – несмотря на свою сложность, DeepLabV3+ разработан с учетом эффективности в вычислениях, что делает его подходящим для реальных приложений, где важно не только качество сегментации, но и быстродействие. Это особенно актуально для клинических систем, где необходимо быстро обрабатывать изображения в реальном времени.
* Применимость в разных областях – благодаря своей универсальности, модель DeepLabV3+ эффективна в различных областях компьютерного зрения, включая автоматическую сегментацию медицинских изображений, таких как МРТ и ОКТ, а также для задач анализа спутниковых снимков, фотограмметрии и других сложных сцен.

DeepLabV3+ является мощным инструментом для задач семантической сегментации, предлагая отличное сочетание точности и эффективности. Модель прекрасно справляется с разнообразными и сложными изображениями, делая её идеальным выбором для применения в таких сферах, как медицинская диагностика, анализ спутниковых данных и другие области компьютерного зрения.

### Первые результаты

На рисунке 14 представлены первые статистические результаты обучения модели YOLOv11.

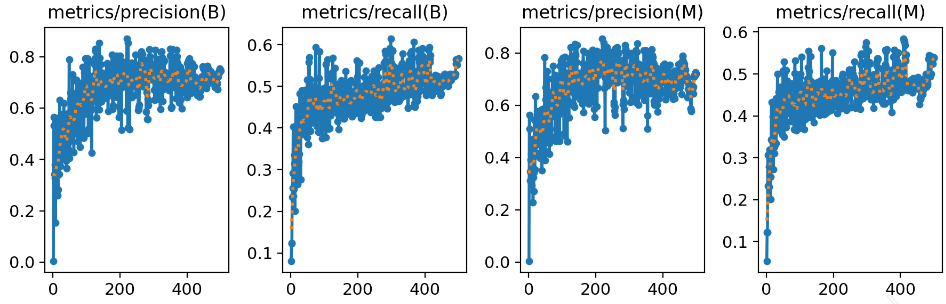


Рисунок 14 – Первые статистические результаты обучения модели YOLOv11

Анализ полученных результатов показывает, что модель способна точно определять и классифицировать патологии на изображениях, о чём свидетельствуют высокие значения показателя Precision. Это означает, что большинство выделенных моделью объектов действительно принадлежат к соответствующим классам патологий.

В то же время Recall в среднем остаётся на умеренном уровне, что указывает на то, что модель пропускает значительное количество патологий – то есть не распознаёт и не классифицирует все имеющиеся объекты на изображении.

В таблице 3 приведены метрики Precision и Recall для каждого из сегментируемых классов патологий.

Таблица 3 – Метрики Precision и Recall каждого из сегментируемых классов

|  |  |  |
| --- | --- | --- |
| Класс | Precision | Recall |
| retinal\_drusen (ретинальные друзы) | 0.83 | 0.38 |
| intraretinal\_cyst (интраретинальные кисты) | 0.87 | 0.25 |
| subretinal\_hyperreflective\_material (субретинальный гиперрефлективный материал) | 0.71 | 0.77 |

Таблица 4 – Продолжение таблицы 3

|  |  |  |
| --- | --- | --- |
| neuroepithelium\_detachment (отслойка нейроэпителия) | 0.89 | 0.2 |
| vitreomacular\_traction (витреомакулярная тракция) | 0.48 | 0.11 |
| lamellar\_macular\_rupture (ламеллярный макулярный разрыв) | 0.46 | 0.5 |
| vitreous\_detachment (отслойка стекловидного тела) | 0.86 | 0.5 |

Из анализа этих метрик можно сделать вывод, что модель в целом уверенно классифицирует обнаруженные объекты (высокий Precision), однако часто не находит все патологии на изображениях (низкий Recall). Это характерно практически для всех классов, включая наиболее заметные, такие как интраретинальные кисты и отслойка нейроэпителия.

На рисунке 15 представлен пример изображения, сегментированного с использование данной модели.

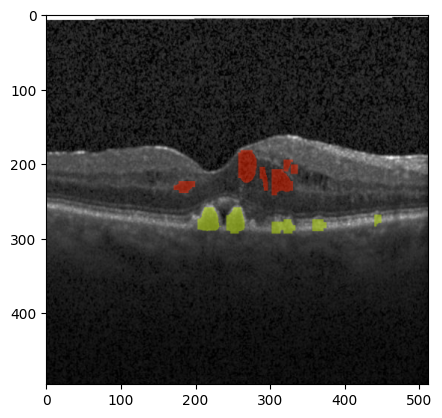


Рисунок 15 – Пример изображения, сегментированного с использованием первой обученной модели

На изображении визуализированы выделенные моделью области патологий: жёлтым цветом обозначены ретинальные друзы, красным – интраретинальные кисты. Видно, что некоторые друзы и кистозные образования остались невыделенными, что подтверждает ранее сделанные выводы: модель демонстрирует хорошую точность, но недостаточную полноту сегментации.

Возможными причинами данного поведения могут являться: несбалансированность обучающей выборки из-за чего модель начинает игнорировать классы, представленные меньшим количеством объектов; недостаточное количество обучающих эпох.

## Разработки модуля формирования заключения

Формирование заключения о возможном офтальмологическом заболевании осуществляется на основании результатов, полученных на выходе модуля обработки изображений. В результате работы модели формируются маски, соответствующие обнаруженным и классифицированным патологическим структурам на ОКТ-изображениях сетчатки.

Для интерпретации выявленных морфологических изменений проводится сопоставление классифицированных патологических классов с референтными таблицами, содержащими информацию о возможных клинических диагнозах, ассоциированных с данными структурами. Такое сопоставление позволяет выдвигать предварительные диагностические гипотезы, основанные на характере и локализации патологий.

В таблице 5 представлен пример референтной таблицы для сегментируемых классов патологий [].

Таблица 5 – Пример референтной таблицы для сегментируемых классов патологий

|  |  |
| --- | --- |
| Класс | Ассоциируемое заболевание |
| retinal\_drusen (ретинальные друзы) | Возрастная макулярная дегенерация |
| intraretinal\_cyst (интраретинальные кисты) | Диабетический макулярный отёк |
| subretinal\_hyperreflective\_material (субретинальный гиперрефлективный материал) | Хориоидальная неоваскуляризация (в т.ч. при ВМД) |
| neuroepithelium\_detachment (отслойка нейроэпителия) | Центральная серозная хориоретинопатия |
| vitreomacular\_traction (витреомакулярная тракция) | Витреомакулярный тракционный синдром |
| lamellar\_macular\_rupture (ламеллярный макулярный разрыв) | Ламеллярная макулярная дыра |
| vitreous\_detachment (отслойка стекловидного тела) | Возрастные изменения, миопия высокой степени |

Сформированное заключение вместе с сегментированным ОКТ-изображением сетчатки, далее, отправляется для отображения на клиентский интерфейс.

## Разработка интерфейса алгоритмического комплекса

На рисунке 16 представлена схема взаимодействия клиентского приложения (пользовательского интерфейса) с сервером (модулем, инкапсулирующим логику обработки изображений).

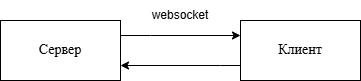


Рисунок 16 – Модель клиент-серверного взаимодействия

Клиент-серверное взаимодействия осуществляется при помощи технологии websocket. Websocket – протокол позволяющий обеспечить полнодуплексную коммуникацию между сервером и клиентом в реальном времени по единичному TCP-подключению.

Использование websocket обеспечивает возможность моментальной отправки результатов обработки ОКТ-изображения, в момент завершения обработки, клиенту, запросившему анализ соответствующего ОКТ‑изображения.

### Разработка клиентского модуля

Клиентский модуль представляет собой визуальный интерфейс для взаимодействия с серверной частью, которая выполняет основную обработку данных. Он разрабатывается с целью решения следующих задач:

* Обеспечение простого взаимодействия с моделью нейронной сети;
* Отображения результатов обработки ОКТ-изображения и отчёта, сформированного по ним;
* Хранение сессионной истории обработанных изображений и отчётов с возможность их просмотра;

Реализация клиентского модуля строится на основе таких технология, как: TypeScript и React.

TypeScript – сильно типизированный интерпретируемый язык программирования, основывающийся на JavaScript. Он повышает надёжность, читаемость и поддерживаемость кода, снижает вероятность ошибок и облегчает масштабирование проекта.

React – библиотека для разработки веб-интерфейсов. Она обеспечивает высокую производительность, компонентный подход и широкие возможности для создания динамичных, интерактивных приложений.

Интерфейс состоит из трёх основных компонентов:

* История. Столбец истории хранит ОКТ-изображений, обработанных за текущий сеанс в соответствующем порядке. При выборке одного из изображений, содержащихся в истории, на экран выведется ОКТ-изображение, отправленное она обработку, обработанное ОКТ-изображение с сегментированными патологиями, сформированное заключение. При завершении сеанса результаты обработки ОКТ-изображений, произведенные за сеанс, удаляются.
* Компонент загрузки. Компонент загрузки позволяет загрузить и просмотреть ОКТ-изображение пред отправкой на обработку. Если было выбрано неверное ОКТ-изображения, компонент позволяет перезагрузить изображение.
* Компоненты просмотра. Компонент просмотра позволяет визуализировать обработанное изображение, содержащее сегментированные объекты патологий. Также, компонент просмотра отображает текст заключения, сформированного по результатам обработки ОКТ-изображения.

На рисунке 17 представлен пользовательский интерфейс разрабатываемого модуля.

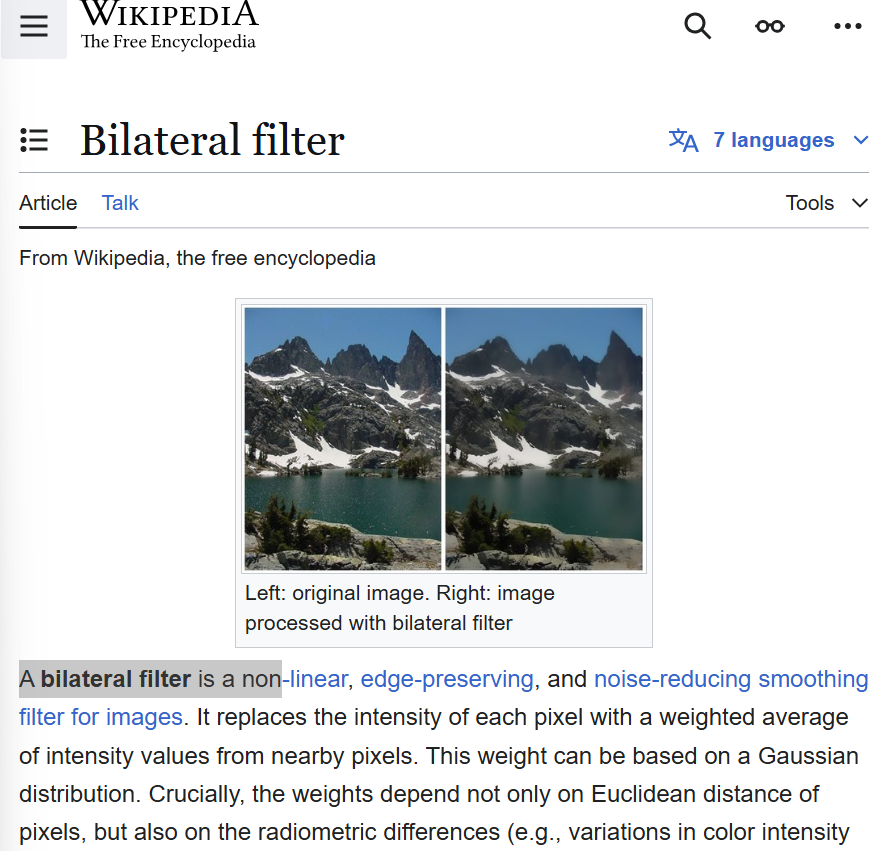


Рисунок 17 – Пользовательский интерфейс разрабатываемого модуля

### Разработки серверного модуля

Серверный модуль предоставляет веб-интерфейс для взаимодействия с модулями предобработки, анализа и формирования заключения. Он выполняет роль координатора обработки данных и служит связующим звеном между клиентским интерфейсом и внутренними вычислительными модулями.

Задачами серверного модуля являются:

* Установка последовательности взаимодействия между модулями предобработки, обработки, формирования заключения.
* Обработка множественных запросов от клиентских модулей.
* Формирование структурированного ответа на основе результатов работы внутренних модулей.

Реализация серверного модуля строится на таких технологиях, как: Python и Flask.

Python – высокоуровневый, интерпретируемый язык программирования с динамической типизацией. Он был выбран в качестве основного языка разработки благодаря своей гибкости и возможности быстрой интеграции с другими модулями, также реализованными на Python. Это позволило значительно упростить настройку взаимодействия между компонентами системы.

Flask – лёгкий и расширяемый веб-фреймворк на языке Python, предназначенный для создания веб-приложений. Его использование позволяет быстро разрабатывать и тестировать прототипы, что особенно важно на ранних этапах проектирования и в процессе итеративной разработки.

## Выводы по главе