

СОДЕРЖАНИЕ

1 ВВЕДЕНИЕ	5
2 ПОСТАНОВКА ЗАДАЧИ	8
3 КОМАНДА РАЗРАБОТКИ	9
4 АЛЬТЕРНАТИВНЫЕ РЕШЕНИЯ	10
4.0.1 DeepFaceLab	10
4.0.2 FSGAN	11
4.0.3 Reface	11
4.0.4 Другие мобильные приложения	12
5 ОПИСАНИЕ МЕТОДОВ РЕШЕНИЯ	13
5.1 Нейронная сеть	13
5.2 Общая архитектура	15
5.3 Инфраструктура	15
5.4 Хранение данных	17
5.5 Предобработка данных	19
5.6 Очередь задач	21
5.7 Необходимая постобработка	21
6 РЕЗУЛЬТАТЫ	22
6.1 Анализ полученных результатов	22

0 КРАТКОЕ ОПИСАНИЕ РАБОТЫ

В данной работе рассматривается процесс разработки сервиса для автоматической замены лиц на видео. Рассматривается разработка клиентской и серверной частей проекта, внутренних сервисов, создание инфраструктуры на основе услуг облачного провайдера и запуск на ней разработанного приложения. Так же описаны проблемы связанные непосредственно с инструментами для замены лиц.

1 ВВЕДЕНИЕ

Большая проблема при производстве видеоконтента в рекламе и других областях заключается в привязке к конкретному человеку. Используя конкретного человека, а особенно знаменитость, например, в рекламе, компании имеют дело с постоянными рисками и неудобствами. Это увеличивающиеся расходы (человек становится более известным и "растет в цене"), сложности в планировании съемок, опоздания и подобные проблемы.

1.0 3D-графика

Одним из альтернативных вариантов решения являются использование 3d-графики. В этом случае настоящие съемки практически полностью заменяются на монтаж видео с использованием 3d-моделей и дальнейшей озвучки. Для этого требуется ручной труд дизайнеров, особое техническое и программное обеспечение. В целом, данный подход дает огромные возможности, которые активно используются, например в кино, позволяя создавать контент не только с людьми, но и с любыми животными или персонажами, снимать сцены, которые невозможно снять с людьми. В то же время данный подход имеет множество недостатков, не позволяющих его использовать для производства простого и массового контента. Во-первых - трудоемкость и большие затраты во времени. Процесс создания компьютерной графики может занимать недели и месяцы и требует постоянной работы команды дизайнеров. Минимальные правки, например, одежды, снова требуют очень сложного процесса моделирования в 3D.

1.0 Живые съемки

Обычным подходом являются полностью живые съемки с живыми актерами и небольшим монтажем после. Это достаточно дешево, если не привлекать к съемкам знаменитостей, и не слишком трудоемко. Минусы данно-

го подхода следуют из привязанности к конкретному актеру. Это постоянно возвращающая стоимость в случае со знаменитостями, человеческий фактор (опоздания, личные проблемы, неподобающий внешний вид), необходимость в макияже, гриме, одежде. Так же живые съемки не позволяют создавать эффекты, возможные при использовании полноценной 3d-графики.

Несмотря на недостатки, данный подход наиболее широко используется в производстве видеоконтента. Большая часть рекламы, фильмов, видеороликов снято с живыми людьми.

1.0 Автоматическая замена лиц

С последними исследованиями в области нейронных сетей оказалось, что проблему можно решить с помощью комбинированного подхода. Современные модели нейронных сетей позволяют накладывать лицо на актера на видео автоматически, без ручного труда дизайнеров и специалистов по видеомонтажу.

Такие модели позволяют снимать видео на обычную камеру с любым актером, что гораздо быстрее и дешевле как видеомонтажа с 3d-графикой, так и полноценных съемок со знаменитостями.

Ранние решения по замене лиц требовали специального обучения модели для конкретной пары актера и лица, а так же обширный набор данных, включающий съемки головы с разных ракурсов и с разными выражениями лица. Более того, требовалась ручная доработка результата, чтобы считать его удовлетворительным. Самые последние решения дают возможность замены лица по одному снимку и обеспечивают получение результата, практически не требующего доработок.

Хотя решений в этой области достаточно много, все они имеют свои недостатки. Инструменты, не представленные в виде готового к использованию сервиса, сложны в применении - требуют мощностей и навыков для разворачивания всего необходимого. Существующие сервисы в основном ориентированы на массовых пользователей, а не бизнес, из-за чего имеют недостаточное качество и множество ограничений.

В данной работе рассматривается разработка альтернативного сервиса, использующего новейшие исследования в области замены лиц и предлагающего замену лиц как сервис, ориентированный на бизнес-клиентов.

2 ПОСТАНОВКА ЗАДАЧИ

Замена лица - это задача переноса лица с исходного на целевое изображение или видео так, чтобы оно аккуратно заменяло внешность лица на целевом изображении и получался реалистичный результат[2, стр. 1].

Главной задачей данной работы является разработка информационной системы, основная функция которой - автоматическая замена лиц на видео. Разработанная ИС должна быть доступна в виде публичного интернет-сервиса, которым может воспользоваться любой человек. Основной целевой аудиторией сервиса являются бизнес-клиенты, действующие в сферах маркетинга и производства видеоконтента.

Требования к разрабатываемой ИС:

- Одновременная работа множества пользователей
- Хорошее качество замены лица
- Пользовательский интерфейс не требующий отдельного обучения
- Замена лица на видео по одному изображению, без необходимости в сборе большого датасета

3 КОМАНДА РАЗРАБОТКИ

Создать полноценный, готовый продукт такого масштаба в одиночку очень сложно. Над данным решением работала команда из нескольких человек в течение года. Команда состоит из менеджера продукта, дизайнера, двух ML-разработчиков и меня в роли разработчика клиентской и серверной части.

Часть продукта, отвечающая непосредственно за обработку видео (получает на вход видео и изображение, на выходе отдает новое видео), разрабатывалась ML-разработчиками.

Я занимался разработкой клиентского веб-приложения, серверной части, некоторых вспомогательных сервисов и оберткой вокруг нейронной сети, позволяющей встроить ее в систему.

4 АЛЬТЕРНАТИВНЫЕ РЕШЕНИЯ

4.0 Инструменты

4.0.1 DeepFaceLab

Одно из самых популярных решений в области замены лица - DeepFaceLab[4]. Оно существует с 2018 года и до сих пор остается передовой и активно развивающейся технологией.

Плюсами являются очень высокое качество наложения и хорошее сохранение похожести лица, большое количество возможностей для подстройки для конкретной пары лиц и видео, высокое разрешение лица.

Несмотря на преимущества, у DFL есть несколько серьезных недостатков, значительно усложняющих его использование. Для работы DFL требует дообучения модели для конкретной пары лиц на достаточно большом наборе данных, включающем изображения со всех ракурсов как для человека на видео, так и для накладываемого персонажа. Так же дообучение занимает значительное время и требует больших мощностей. DFL работает по принципу маски, накладывая одно лицо на другое, что делает его очень чувствительным к качеству датасетов. Так же это создает проблему разницы в освещении и цвете между видео и датасетом маски.

Упомянутые выше недостатки проявляются для потенциального пользователя в следующих минусах:

- Необходимость в сборе большого датасета для актера и накладываемого персонажа
- После сбора датасета необходимо дообучать модель в течение нескольких недель
- Датасет должен быть очень качественным - иметь изображения с разнообразных углов и с разным освещением

4.0.2 FSGAN

FSGAN[2][3] - более современная технология, появившаяся в 2019 году.

Face reenactment (перенос мимики, puppeteering) - использование движений и мимики лица в ведущем видео для управления мимикой и движением лица, представленном на другом видео или изображении.

Алгоритм работы FSGAN состоит из следующих шагов:

- Сегментация - выделение лица на исходном и целевом изображении
- Face reenactment - воссоздание мимики и положения лица на целевом видео с исходным лицом
- Смешивание лиц (face blending) - наложение получившегося лица на целевое изображение с сохранением цвета кожи и освещения

Решение, предлагаемое FSGAN не требует дообучения для конкретных пар лиц и предлагает замену лица на основе лишь одного изображения исходного лица, против большого датасета для DeepFaceLab.

Тем не менее, FSGAN имеет довольно низкое качество наложения, так выглядит пример его работы:



Рисунок 4.1 – Примеры работы замены лица и переноса мимики в FSGAN

4.0 Готовые сервисы

4.0.3 Reface

Популярным сервисом для замены лица является Reface. Это приложение для смартфонов, позволяющее наложить свою фотографию на один из видеороликов, предоставленных Reface. Приложение ориентировано на сегмент

b2c, т.е. массового пользователя. Reface по очевидным причинам не подходит для создания видеоконтента для бизнеса. Во-первых, лица можно накладывать только на ролики из существующего набора, который содержит отрывки из популярных фильмов и сериалов. Во-вторых, выходное качество видео очень низкое - это касается как качества наложения, так и выходного разрешения, частоты кадров и других параметров видео. Сервис ориентирован на то, что роликами просто будут делиться в соцсетях и мессенджерах, а не на серьезный видеопродакшн.

Минусы Reface исходят из его позиционирования на другой сегмент, и таким образом он не является хорошей альтернативой в сегменте b2b.

4.0.4 Другие мобильные приложения

Помимо Reface похожие функции появляются во многих других мобильных приложениях. Все они ориентированы на сегмент b2c, имеют низкое качество выходного видео и множество ограничений. Большая часть таких приложений накладывает лица только на фотографии, но не видео.

5 ОПИСАНИЕ МЕТОДОВ РЕШЕНИЯ

5.1 Нейронная сеть

В процессе разработки сервиса наша команда разработала собственное решение для замены лиц на базе SimSwap[1]. SimSwap[1] появился во второй половине 2021 года и позиционируется как ”эффективный фреймворк для высокоточной замены лиц”.



Рисунок 5.1 – Результаты наложения лица в исходном SimSwap

Данное решение оказалось наиболее подходящим для наших целей. SimSwap позволяет очень качественно заменять лица, используя одну предобученную модель и всего лишь одно изображение целевого лица. При этом вместо наложения маски SimSwap использует т.н. инъекцию лицевых атрибутов. Т.е. вместо наложения одного изображения на другое модель пытается изменить лицо на изображении так, чтобы оно было похоже на другое лицо. Это позволяет избежать проблем связанных с разностью в освещении и цветовой гамме изображений. При этом модель изначально работала достаточно быстро - 1 минута видео обрабатывалась около 10-15 минут.

5.2 Общая архитектура

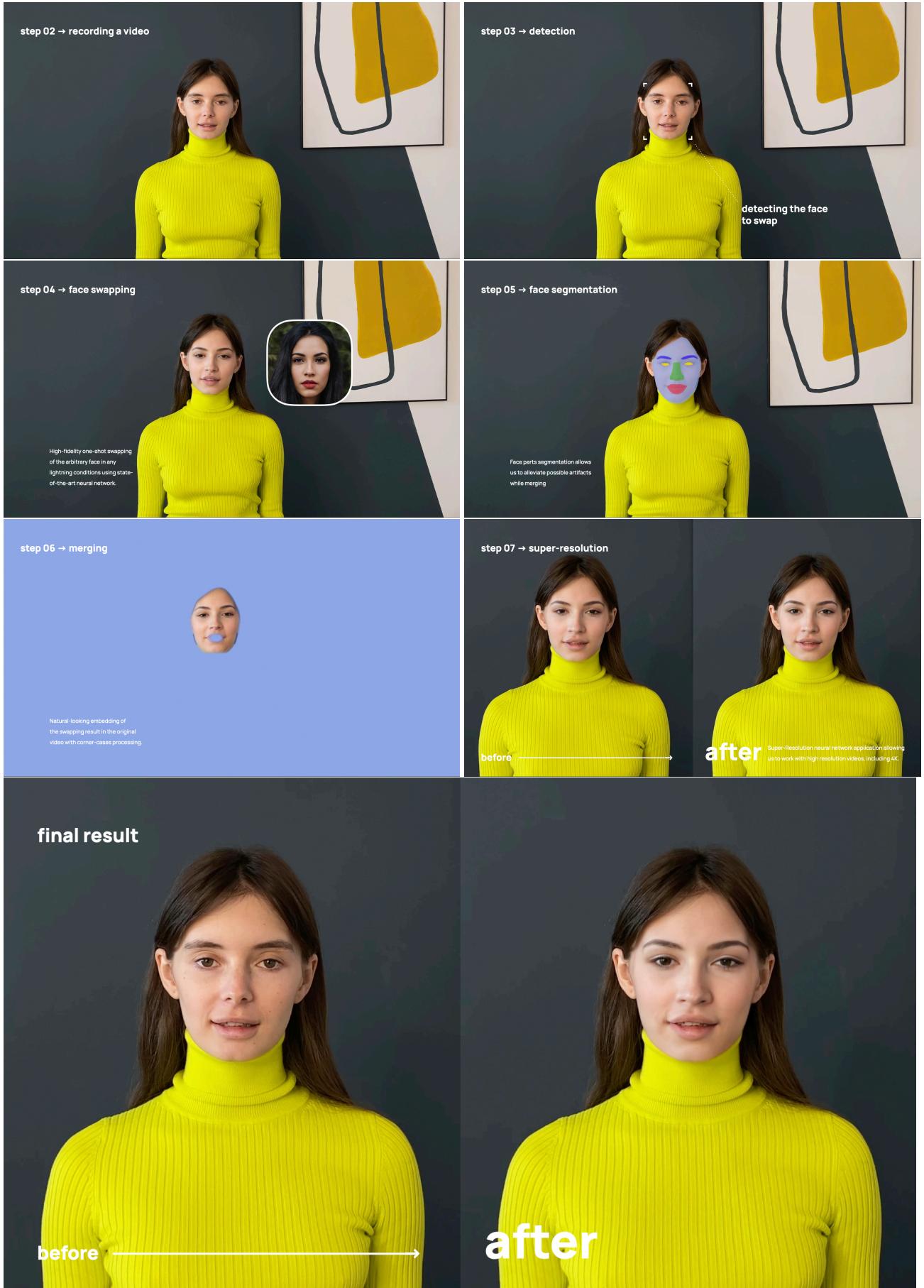


Рисунок 5.2 – Последовательность обработки видео

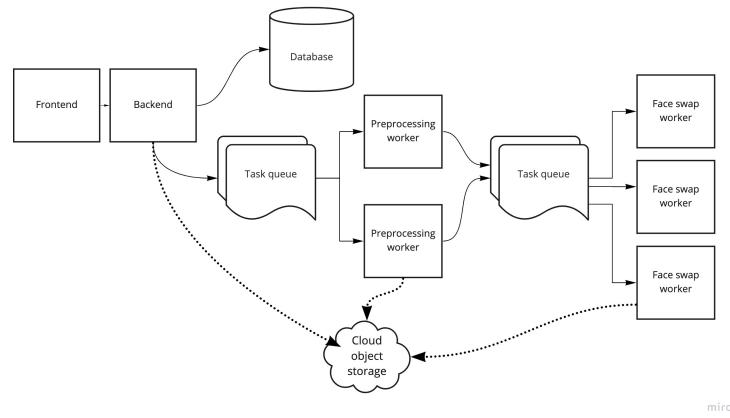


Рисунок 5.3 – Схема архитектуры сервиса

Для разработки внешней части приложения выбрана клиент-серверная архитектура.

Выполнение вычислительно сложных операций, необходимых для решения задачи, возможно на клиентских устройствах в очень ограниченном объеме и несет множество проблем, таких как необходимость держать устройство включенным, потеря

Такое решение позволяет производить вычислительно сложные операции на серверах, так же сохраняя там приватную информацию о системе.

5.3 Инфраструктура

В любой системе центральным вопросом инфраструктуры является расположение серверов. Под разработкой обычно подразумевается создание программного продукта, но когда дело касается разворачивания приложения в публичный доступ, возникает вопрос - покупать и располагать физические сервера у себя в компании или data-центре или воспользоваться облачными провайдерами и получать необходимые мощности по подписке.

Рассмотрим некоторые особенности обоих вариантов.

Собственные сервера приобретаются один раз и могут стоить достаточно дешево, в сравнении с ежемесячными счетами от облачного провайдера. На

первый взгляд из постоянных расходов они требуют лишь электричество и интернет. Но в деталях ситуация оказывается гораздо сложнее.

Собственные сервера требуют настройки и постоянного обслуживания, пример - обновления ПО. Для того, чтобы обеспечить непрерывную работы потребуются резервные интернет-каналы, запасной источник питания (например, ИБП). Так же сервера отличаются от обычных компьютеров тем - они ориентированы на постоянную работу и имеют свои особенности. Например, шум и необходимость серьезного охлаждения. Обычно это проявляются в том, что для содержания своих серверов выделяют отдельную комнату - серверную.

Помимо собственно поддержки самого аппаратного обеспечения и операционных систем серверов, собственные сервера лишают возможности использования дополнительных услуг облачных провайдеров - управляемые (managed) сервисы, автоматическая репликация, в том числе геораспределенная, управление доступом и многое другое.

Если учитывать своюимость обслуживания, собственные сервера могут оказаться гораздо дороже облачных услуг.

Облачные провайдеры, напротив, предоставляют максимально удобные, готовые и самоподдерживаемые сервисы, особенно при использовании не чистых виртуальных машин, а готовых сервисов, таких как управляемые базы данных, управляемый Kubernetes, файловое хранилище. Облачные провайдеры предоставляют автоматическое резервное копирование и обновление. Серьезным минусом использования "облаков" является так называемый vendor-lock - завязка на конкретного провайдера и невозможность (или очень высокая стоимость) переноса инфраструктуры от него. Это случается, если на проекте активно используются уникальные услуги, не предоставляемые или сильно отличающиеся у других провайдеров.

Для данного проекта было выбрано использования облачного провайдера Azure для размещения инфраструктуры. Это сервис от компании Microsoft. Он содержит огромное количество готовых сервисов и удобный веб-интерфейс для управления инфраструктурой.

Как целевая платформа для разработки приложений был выбран Kubernetes. Он имеет ряд очень существенных преимуществ по сравнению с работой с обычными серверами или виртуальными машинами.

- Позволяет автоматически разворачивать сервера и подстраивать их количество под требования системы
- Универсален для всех облачных провайдеров и позволяет легко перенести свое приложение, в том числе на собственную инфраструктуру
- Предоставляет стандартизированный способ разворачивания приложений и сервисов
- Имеет огромное сообщество, множество open-source решений имеют готовые конфигурационные файлы для Kubernetes
- Позволяет легко открывать доступ из вне к приложениям, при этом сохраняя их по умолчанию приватными
- Содержит инструменты для мониторинга и отладки приложений

5.4 Хранение данных

Для хранения пользователей, истории запросов, данных о состоянии обработки запросов, информации о подписках и многое другого была выбрана система управления базами данных PostgreSQL.

От других подобных решений она отличается полной бесплатностью и открытым исходным кодом, огромным сообществом, множеством возможностей и дополнений, работой "из коробки" с практически любыми языками и библиотеками. PostgreSQL - стандарт де-факто среди SQL баз данных для веб-приложений.

Так же исходя из специфики работы с мультимедиа в решении требуется файловое хранилище, позволяющее хранить промежуточные и финальные результаты обработки. Объектом хранения здесь являются бинарные файлы, представляющие собой фрагменты видео, целые видеофайлы, а так же фотографии лиц.

Рассмотрим различные опции:

- База данных
- Файловая система
- Оперативное хранилище, например Redis
- Облачное хранилище (S3)

Современные СУБД и частности PostgreSQL позволяют хранить бинарные данные в обычных SQL таблицах. В PostgreSQL для этого существуют 2 опции - **Large Objects** и тип данных `bytea`.

К сожалению, эти опции не очень хорошо подходят для нашего случая.

Large Objects являются нестандартной возможностью и плохо поддерживаются различными клиентами. Так же они требуют явного удаления файла, потому что сами данные хранятся отдельно от таблицы, которой принадлежат.

`bytea` плохо подходит для больших файлов, т.к. не поддерживает стриминг, а значит при работе с видеофайлами требуется много оперативной памяти. В нашем случае файлы вполне могут иметь размер в несколько гигабайт и даже больше, а значит требования к серверу базы данных (как и к клиентскому приложению) были бы огромны.

Более того - обе опции очень малопродуктивны. Это очевидно, если исходить из того, что SQL базы данных совсем не ориентированы на хранение больших бинарных блоков. Производительность запросов в этом случае будет оставлять желать лучшего.

Файловая система, что неудивительно, отлично подходит для хранения файлов. Но в нашем случае имеет ряд критических недостатков. Во-первых, к диску может одновременно иметь доступ только одна машина (виртуальная или физическая), а значит провоцирует монолитную архитектуру сервиса, что невозможно в нашем случае. Во-вторых, файловая система не позволяет автоматически переносить данные на более медленные и дешевые СХД. В-третьих, нужно самому организовывать систему резервного копирования, чтобы не потерять данные.

Третий вариант - key-value системы хранения данных, подобные Redis. Они хранят данные в памяти, иногда сгружая в постоянное хранилище. Для данного проекта они так же имеют ряд серьезных ограничений. Такие системы

не гарантируют надежность хранения данных, имеют плохую устойчивость к перезагрузкам. Они ориентированы на хранение данных в памяти, что плохо подходит для больших файлов.

Облачные файловые хранилища оказались лучшей опцией в нашем случае. Хранение большого объема файлов в нем достаточно дешево, они гарантируют надежность хранения данных и их репликацию по необходимости. Обращение к файлам являются достаточно быстрыми по скорости, особенно учитывая колокацию с серверами приложений на мощностях облачного провайдера.

Таким образом, для хранения всех медиафайлов в процессе работы проекта было выбрано облачное хранилище, а конкретно Azure Blob Storage. Решения у различных облачных провайдеров в этой области очень похожи. Для проекта самым рациональным является выбор того сервиса, который относится к используемому облачному провайдеру.

5.5 Предобработка данных

Перед наложением лица видео требует дополнительной обработки исходя из требований к сервису.

Во-первых, пользователям предлагаются разные тарифные планы, при которых обработка видео требует различного количества ресурсов. Например, для пользователей бесплатной (пробной) версии доступна обработка видео разрешением до FullHD (1920x1080) и до 30 кадров в секунду. Такие ограничения позволяют сэкономить на времени обработки, т.к. нейронные сети быстрее работают с меньшим объемом данных. Назовем этот шаг ограничением качества видео.

Во-вторых, в требованиях к сервису мы определили необходимость ускорения обработки за счет распределенной обработки на нескольких серверах одновременно. Так, например, при наличии 10 свободных серверов, мы можем обработать пользовательское видео в 10 раз быстрее.

Самый эффективный способ распределить обработку видео на несколько потоков - поделить его на сегменты примерно одинаковой длины и запустить в очередь как обработку нескольких видео. Так сервера-обработчики смогут максимально быстро разобрать и выполнить все задачи по обработке.

В текущей версии ограничение качества не происходит на этапе пре-добработки данных, а вместо этого выполняется на этапе самой обработки, непосредственно перед передачей обрабатываемого сегмента в нейронную сеть.

Второй же шаг необходим. Задача состоит в том, чтобы поделить видео на сегменты примерно равной продолжительности, не потратив на это много времени. При этом после обработки эти сегменты должно быть возможно бесшовно склеить ровно так, как было в исходном видео. Еще одна задача этого этапа - максимально сохранить качество видео. Если ограничения на качество видео не накладываются, мы хотим чтобы пользователь получил результат максимально близкий к исходному видео по качеству. Так как мы ориентируемся на бизнес-пользователей, для которых это очень критично.

Так же разделение на сегменты является важной возможностью сервиса, сильно увеличивающей возможности для контроля обработки. Например, в случае неожиданной ошибки прогресс сможет быть сохранен с точностью до длины кусочка. Так же появляется возможность приоритезации обработок разных клиентов "на ходу". Во время долгой обработки бесплатного клиента, может быть загружено видео более приоритетного клиента. Тогда в случае коротких сегментов у нас будет возможность максимально быстро обработать приоритетное видео и затем вернуться к первому клиенту.

FFMpeg - утилита - швейцарский нож в мире обработки и перекодирования видео. Она содержит множество настроек и инструментов для практически любых задач в этой области. В частности, в ней есть необходимая нам возможность разделения видео на сегменты. Нужная нам опция называется /textttsegment muxer.

Для использования этой возможности нам необходимо включить в ffmpeg режим вывода segment опцией f segment и задать целевое время сегмента с

помощью `-segment_time <количество_секунд>`. Мы используем 5 секунд на сегмент. Это хорошо подходит под среднюю длину видео, загружаемых нашими клиентами, которая не превышает 10-20 секунд.

5.6 Очередь задач

Для распределения задач обработки сегментов между воркерами используется распределенная очередь задач. Это позволяет подавать задачи на обработку максимально эффективно и в случае ошибки (например, отключение воркера) возвращать задачу в очередь и в последствии повторять обработку.

Очередь построена на базе брокера сообщений RabbitMQ[6] и фреймворка Celery[5]. Celery предоставляет протокол и интерфейс на языке Python, работающий поверх различных брокеров сообщений, позволяющий запускать задачи, распределенные по разным серверам, так же, как обычные функции в языке Python.

5.7 Необходимая постобработка

6 РЕЗУЛЬТАТЫ

6.1 Анализ полученных результатов

6.1.0 Оптимальная нагрузка

Сервис работает оптимально с точки зрения нагрузки и использования ресурсов, равномерно и максимально эффективно распределяя вычисления между обработчиками.

6.1.0 Перекодирование

Видео перекодируется несколько раз в ходе обработки, что увеличивает потери в качестве, происходит потеря информации из исходного видео. В дальнейшем эти потери хочется минимизировать. Так же это замедляет обработку из-за дополнительных операций. Отказ от множественных перекодирований связан с очень высокой нагрузкой на ИО. Огромные объёмы промежуточных данных придётся хранить и передавать по сети, записывать на диски. Это может создать большие ограничения на длину видеофайлов. Идеального решения данной задачи не существует. Каждый из подходов имеет свои минусы.

6.1.0 Параллелизм

Параллельная обработка на нескольких серверах позволяет максимально быстро и оптимально с точки зрения нагрузки выполнить определенный объем работы. Но с точки зрения множества пользователей это не всегда хорошо. Отработка достаточно длинного файла занимает работой сразу все сервера-обработчики. Если в этот момент на обработку поступает видео другого пользователя, особенно если оно очень короткое, происходит затор. Пользователь с длинным видео готов подождать, но будет сильно задерживать пользователя,

например, с картинкой. По сути параллельная обработка заставляет использовать всю пропускную способность сервиса сразу, не оставляя места для параллельной работы с разными клиентами. Так параллельная с точки зрения нагрузки обработка становится последовательной с точки зрения множества пользователей.

Эта проблема решается определенными настройками, позволяющими балансировать между двумя подходами. Например, параллельную обработку для одного пользователя можно ограничить двумя или тремя потоками. Мы потеряем в максимальной скорости, но при этом оставим место другим обработкам в очереди, даже при обработке очень длинного видеофайла. Алгоритм для работы таких ограничений пока не реализован.

6.1.0 Наложение

Сервис на данный момент имеет проблемы с качеством наложения лиц. Он плохо работает с лицами, находящимися вблизи, т.к. в этом случае сильно увеличивается их площадь и результат сильно теряет в качестве. Так же в связи с используемым алгоритмом плохо переносится похожесть лица (*identity*). Наложенное лицо не совсем совпадает с целевым, оно скорее является чем-то средним, полученным из исходного и целевого лица.

6 СПИСОК ЛИТЕРАТУРЫ

1. Renwang Chen, Xuanhong Chen, Bingbing Ni, and Yanhao Ge. Simswap: An efficient framework for high fidelity face swapping. In *MM '20: The 28th ACM International Conference on Multimedia*, 2020.
2. Yuval Nirkin, Yosi Keller, and Tal Hassner. FSGAN: Subject agnostic face swapping and reenactment. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 7184–7193, 2019.
3. Yuval Nirkin, Yosi Keller, and Tal Hassner. FSGANv2: Improved subject agnostic face swapping and reenactment. In *hello*. IEEE, 2022.
4. Ivan Perov, Daiheng Gao, Nikolay Chervonyi, Kunlin Liu, Sugasa Marangonda, Chris Umé, Mr. Dpfks, Carl Shift Facenheim, Luis RP, Jian Jiang, Sheng Zhang, Pingyu Wu, Bo Zhou, and Weiming Zhang. Deepfacelab: Integrated, flexible and extensible face-swapping framework, 2020. URL: <https://arxiv.org/abs/2005.05535>, doi:10.48550/ARXIV.2005.05535.
5. Ask Solem and contributors. Celery, 2009-2021. URL: <https://docs.celeryq.dev/en/stable/index.html>.
6. Inc. VMware. Rabbitmq documentation, 2007-2022. URL: <https://www.rabbitmq.com>.