МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САМАРСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ ИМЕНИ АКАДЕМИКА С.П. КОРОЛЕВА»
(САМАРСКИЙ УНИВЕРСИТЕТ)

На правах рукописи

Козлова Юлия Ханифовна

МЕТОД СОЗДАНИЯ ПАРАМЕТРИЗОВАННОГО АВАТАРА ГОЛОВЫ ЧЕЛОВЕКА НА ОСНОВЕ НЕЙРОСЕТЕВОЙ МОДЕЛИ РЕНДЕРИНГА

2.3.8 – Информатика и информационные процессы

Диссертация на соискание ученой степени
кандидата технических наук

Научный руководитель:
Мясников Владислав
Валерьевич,
доктор физико-математических
наук, профессор

Самара – 2024

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ

Актуальность темы исследования

Задача создания цифровых аватаров людей становится более актуальной в последние годы в связи со стремительным развитием технологий виртуальной, смешанной и дополненной реальности [1]. В областях киноиндустрии и игровой индустрии также существует высокий спрос на такие технологии. При этом для создания реалистичных аватаров требуется дорогостоящее оборудование, которое включает в себя специализированное освещение [2], [3] и многокамерную установку [4], [5], а также кропотливый ручной труд специалистов, которые занимаются постобработкой полученных данных и проектированием примитивов для анимации, интегрируемых в существующие рендереры. Методы создания цифровых аватаров позволяют решать задачу телеприсутствия, что было особенно актуально во время пандемии COVID-19, которая в результате привела к частичному переходу на дистанционный формат работы, обучения, участия в различных мероприятиях, например, в научных конференциях. В работе [6] авторы показали, что интеграция технологии цифрового аватара в видеоконференции позволяет при снижении пропускной способности канала сохранять исходное качество изображения, так как не требуется выполнять передачу и сжатие всего изображения. Вместо этого на приемную сторону передается низкоразмерная закодированная информация о ключевых точках лица и положении головы, которая используется для синтеза реалистичной анимации. Ещё одним примером использования цифровых аватаров являются звонки с эффектом присутствия с помощью гарнитуры смешанной реальности. Так, в июне 2023 года компания Apple представила гарнитуру смешанной реальности «Apple Vision Pro» [7], которая с помощью специализированного приложения «Persona» [8] позволяет создать персонализированного аватара. Процесс создания аватара включает в себя процедуру сканирования головы камерами устройства в соответствии с заданными инструкциями. Конечное назначение

представленной технологии – звонки с имитацией присутствия в едином физическом пространстве, где за счет информации с датчиков и камер гарнитур производится анимация аватаров.

Здесь и далее под **аватаром головы человека** понимается цифровое представление поверхности головы (форматами представления могут выступать полигональная сетка, облако точек, нейронное неявное представление, изображение и т. п.), полученное на основе некоторых данных (трехмерное представление головы, полученное в результате сканирования; изображение; набор изображений; видеопоследовательность и т. п.), которое может быть использовано для передачи и воспроизведения изображения лица/головы без потери идентичности. Под параметризованным аватаром будет пониматься аватар, для которого возможно выполнить синтез изображений, при котором мимика и поза головы будут управляться значениями некоторых параметров.

Таким образом, задача создания параметризованного аватара головы человека заключается в разработке метода, который принимает на вход некоторую информацию, описывающую внешний вид головы человека, а на выходе формирует реалистичное представление поверхности головы, которое может быть модифицировано в зависимости от ожидаемого выражения лица и/или положения головы, а также использовано для воспроизведения/реконструкции изображения/видеопоследовательности головы человека.

Для объективного анализа актуальности выбранной темы исследования была проведена агрегация статей с упоминанием ключевых слов «Head avatar», «NeRF», «Neural rendering и NeRF» по годам, начиная с 2015 года. В рассмотрении участвовали работы, размещенные в электронном архиве с открытым доступом для научных статей и рукописей arXiv.org [9]. На рисунке 1 представлен результат анализа. Исходя из представленных результатов можно заметить нелинейный рост количества статей, посвященных как задаче создания аватаров головы человека, так и выбранному в рамках

диссертационного исследования способу пространственного представления – Neural Radiance Fields (NeRF, подробнее в подразделе 3.1.1), в частности его модификации, где для рендеринга итогового изображения применяется не алгоритм объемометрического рендеринга (англ. volume rendering), а двумерная свёрточная нейронная сеть, значительно ускоряющая эту процедуру.

Видно, что направление исследований, посвященное методам создания аватара головы человека, начало набирать популярность около пяти лет назад, однако значимых работ близких к теме диссертационного исследования от авторов из Российской Федерации или стран СНГ обнаружить не удалось. Большое количество статей с высокими показателями цитируемости опубликовано авторами из Германии (Университет Макса Планка [10]), США (Университет Стэнфорд [11], Университет Беркли [12], Университет Карнеги [13]), Швейцарии (Швейцарская высшая техническая школа Цюриха [14]), Великобритании (Оксфордский университет [15]), Китая (Шанхайский университет [16]) и других государств.

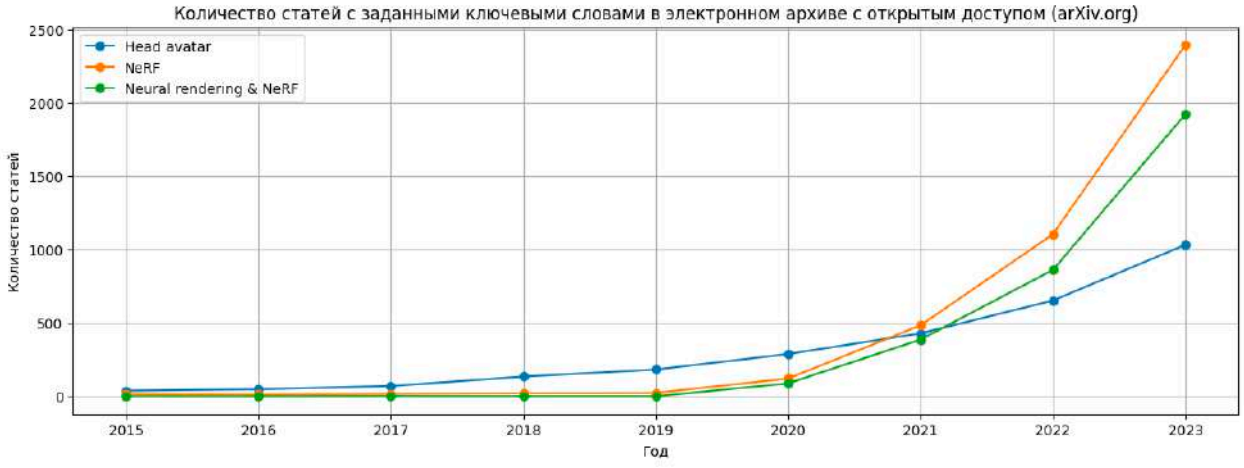


Рисунок 1 – Количество статей с заданными ключевыми словами по годам

При анализе существующих решений по теме диссертационного исследования (см. раздел 1) были выявлены следующие недостатки:

1. Низкая доступность в связи с использованием сложных многокамерных установок (в некоторых случаях также требуется специализированное освещение).
2. Низкая степень схожести между субъектом и синтезируемыми изображениями.
3. Высокая вычислительная сложность на этапах обучения и/или синтеза новых изображений.

Перечисленные недостатки присущи рассмотренным методам в разной степени, например, некоторые методы обладают высокой доступностью и низкой вычислительной сложностью, но не способны синтезировать изображения с высокой степенью схожести, другие методы, напротив, способны синтезировать высококачественные фотореалистичные изображения, но требуют большого количества вычислений и/или сложных многокамерных систем.

Таким образом, ключевым направлением диссертационного исследования является разработка метода, который позволит устранить указанные недостатки. А именно, разрабатываемый метод создания аватара должен удовлетворять следующим условиям:

1. Доступность, под которой подразумевается возможность использования одной видеопоследовательности.
2. Высокая степень схожести между субъектом и синтезируемыми изображениями.
3. Низкая вычислительная сложность по сравнению с существующими решениями.

Учитывая все изложенные выше тезисы, как тема диссертационной работы в целом, так и отдельные выбранные направления исследований в частности являются безусловно актуальными.

Цели и задачи исследования

Целью диссертационного исследования является разработка методов и алгоритмов создания параметризованного аватара головы человека,

обеспечивающих при относительно низких вычислительных затратах процесса создания повышенное качество синтезируемых изображений-проекций.

Для достижения поставленной цели в диссертации решаются следующие задачи:

1. Разработка алгоритма оценки параметров модели головы FLAME с использованием RGBD изображения.
2. Разработка и исследование параметрической модели головы человека на основе нейросетевой модели представления поверхности CNeRF, двумерного нейронного рендеринга и синтетического набора данных, генерируемого в реальном времени.
3. Разработка и исследование метода создания параметризованного аватара головы человека на основе разработанной параметрической модели головы человека.

Поставленные задачи определяют структуру работы и содержание ее разделов.

Методология и методы исследования

В диссертационной работе используются методы статистического анализа, компьютерной графики, компьютерного зрения и методы машинного обучения.

Научная новизна работы

1. Предложен алгоритм оценки параметров параметрической модели FLAME с использованием RGBD изображения.
2. Разработана параметрическая модель головы человека на основе нейросетевой модели представления поверхности CNeRF, двумерного нейронного рендеринга и синтетического набора данных, генерируемого в реальном времени.
3. Предложен метод создания параметризованного аватара головы человека на основе разработанной параметрической модели головы человека.

4. Предложен способ аугментации (расширения) реального набора данных, содержащего кадры видеопоследовательности с изображением головы человека, с использованием интерполяции промежуточных кадров.

Практическая значимость работы

Разработанные методы и алгоритмы могут быть использованы в составе систем телеприсутствия; виртуальной, дополненной, смешанной реальности; систем видеоконференцсвязи; систем создания видеоконтента и прочих системах, требующих синтеза визуально реалистичных трехмерных моделей лиц пользователей.

Положения, выносимые на защиту

На защиту выносятся:

1. Алгоритм оценки параметров модели FLAME с использованием RGBD изображения, позволяющий достичь высокой точности трехмерной реконструкции.
2. Параметрическая модель головы человека на основе нейросетевой модели представления поверхности CNeRF, архитектуры сети двумерного нейронного рендеринга с блоком повышения пространственной дискретизации, позволяющим ускорить сходимость метода создания аватара, и синтетического набора данных, генерируемого в реальном времени.
3. Метод создания параметризованного аватара головы человека на основе разработанной параметрической модели головы человека, позволяющий достичь высокой скорости создания аватара конкретного человека, а также обеспечить высокую скорость синтеза изображений-проекций аватара при высоком значении показателя качества.
4. Способ аугментации (расширения) реального набора данных, содержащего кадры видеопоследовательности с изображением головы человека, с использованием интерполяции промежуточных

кадров, который позволяет повысить качество синтезируемых изображений-проекций аватара для коротких видеопоследовательностей.

Соответствие специальности

Диссертация соответствует паспорту научной специальности 2.3.8 – «Информатика и информационные процессы» и охватывает следующие области исследования, входящие в эту специальность:

1. Разработка компьютерных методов и моделей описания, оценки и оптимизации информационных процессов и ресурсов, а также средств анализа и выявления закономерностей на основе обмена информацией пользователями и возможностей используемого программно-аппаратного обеспечения.
13. Разработка и применение методов распознавания образов, кластерного анализа, нейросетевых и нечетких технологий, решающих правил, мягких вычислений при анализе разнородной информации в базах данных.

Степень достоверности и апробация результатов

Основные результаты диссертации были представлены на 3 научных конференциях:

1. Международной конференции «Информационные технологии и нанотехнологии» (ИТНТ, Самара, Россия) - 2022 год;
2. Международной конференции «Информационные технологии и нанотехнологии» (ИТНТ, Самара, Россия) - 2023 год;
3. Международной конференции «Информационные технологии и нанотехнологии» (ИТНТ, Самара, Россия) - 2024 год.

По теме диссертации опубликовано восемь работ [17*–24*] (здесь и далее работы автора диссертации обозначаются символом «*»). Из них одна работа опубликована в изданиях, рекомендуемых ВАК, три работы опубликованы в рецензируемых изданиях, входящих в международные

реферативные базы данных Scopus и/или WebOfScience. Пять работ выполнено без соавторов.

Результаты диссертационной работы:

1. Использованы в АО «Самара-Информспутник» при выполнении хоздоговорных работ № 4/2021 от 22.09.2021 (2021–2023), № 7/2021 от 08.11.2021 (2021–2023).
2. Использованы в ФГУП «ГосНИИПП» в рамках составных частей научно-исследовательских работ по договорам №4/2021 от 22.09.2021 года и №7/2021 от 08.11.2021 года.
3. Использованы в рамках договора №69/12/2023 от 25.12.2023, в рамках гранта от Фонда Содействия Инновациям № 358ГС1ЦТС10-D5/91117 от 18.12.2023 в ООО «Давтех».
4. Использованы в учебном процессе в ФГАОУ ВО «Самарский национальный исследовательский университет имени академика С. П. Королева» в курсе лекций по дисциплине «Безопасность мультимедиа».

Диссертация состоит из четырех разделов, заключения, списка использованных источников из 102 наименований; изложена на 209 страницах машинописного текста, содержит 59 рисунков, 2 таблицы, 1 приложение.

Краткое содержание диссертации

В первом разделе диссертационного исследования дана оценка современного состояния задачи создания цифрового аватара головы человека. Приведен обзор большого количества работ, представленных на различных научных конференциях высокого класса и опубликованных в журналах с высоким рангом. Производится систематизация этих работ и разработка системы классификации методов создания аватара головы человека.

Во втором разделе диссертационного исследования представлен подробный обзор параметрической модели головы FLAME и описан классический алгоритм оценки параметров модели по RGB изображению, который применяется во многих современных работах по созданию цифрового

аватара головы на этапе предобработки. Также представлен разработанный алгоритм оценки параметров параметрической модели FLAME по RGBD изображению. Параметрическая модель головы FLAME необходима для создания «грубого» описания поверхности головы человека посредством векторов. Такое представление необходимо как один из компонентов входных данных разработанного метода создания аватара головы.

В третьем разделе диссертационного исследования приведено описание разработанной параметрической модели головы человека на основе нейросетевой модели представления поверхности CNeRF (Conditional Neural Radiance Fields) и двумерного нейронного рендеринга. Описан модуль генерации синтетических данных в реальном времени. Представлены порядок проведения и результаты экспериментальных исследований.

В четвертом разделе диссертационного исследования приведено описание разработанного метода создания параметризованного аватара головы человека. Описываются порядок проведения и результаты экспериментальных исследований метода. В частности, описывается сформированный набор настроек метода, предлагаемые варианты аугментации данных, что в совокупности позволяет разработанному методу решать поставленную задачу эффективнее. Представлен полный цикл создания и анимации аватара головы, что может быть полезно в прикладных сценариях использования разработанного метода. А также приведено сравнение предложенного метода с другими современными и актуальными методами, позиционируемыми авторами, как state-of-the-art решения.

РАЗДЕЛ 1. ОБЗОР СУЩЕСТВУЮЩИХ РЕШЕНИЙ

Данный раздел диссертационного исследования посвящен подробному описанию существующих актуальных решений в области создания цифровых аватаров головы человека. Также предложена классификация методов создания цифрового аватара.

1.1. Классификация методов создания цифрового аватара головы человека

Методы создания цифровых аватаров головы человека можно классифицировать по следующим критериям:

1. По способу представления поверхности цифрового аватара
    1.1. Явное представление поверхности:
        1.1.1. Облако точек – это способ описания поверхности с помощью несвязного набора точек в трёхмерном пространстве. Часто для каждой точки дополнительно хранится информация о ее цвете.
        1.1.2. Полигональная сетка (жарг. меш) – это способ описания поверхности с помощью двух наборов: набора вершин и набора граней, определенных на вершинах. В отличие от облака точек представление в виде полигональной сетки обладает связностью.
        1.1.3. Воксельная сетка – это способ описания поверхности, при котором пространство разбивается на регулярную сетку. Часто приводят аналогию с изображением, где минимально значимая единица – это пиксель. В данном случае куб минимального размера при разбиении сетки считается минимально значимой единицей и называется вокселем.

1.2. Неявное представление поверхности (в рамках данного диссертационного исследования рассматриваются нейронные неявные представления):

1.2.1. Signed Distance Fields (SDF, Signed Distance Function) [101] – это способ описания поверхности путем вычисления расстояния от каждой точки в пространстве до ближайшей точки поверхности. Если точка находится вне поверхности, то расстояние имеет положительный знак, если внутри, то расстояние имеет отрицательный знак. Аппроксимация поверхности, как правило, реализуется с использованием многослойного перцептрона, который решает задачу регрессии.

1.2.2. Occupancy Fields [102] – это способ описания поверхности путем разделения пространства на сетку, где в каждой ячейке записывается значение 0, если она находится вне поверхности, и 1 в противном случае. Аппроксимация поверхности, как правило, реализуется с использованием многослойного перцептрона, который решает задачу бинарной классификации.

1.2.3. Neural Radiance Fields (NeRF) [25] – это сравнительно новый способ описания поверхности, который по набору изображений сцены с разных ракурсов обучается с помощью многослойного перцептрона для каждой точки пространства предсказывать ее цвет и объемную плотность. Важно отметить, что модель NeRF не решает задачу построения поверхности, а решает задачу синтеза новых видов (англ. novel view synthesis). Для генерации изображения с нового ракурса выполняется процедура объемометрического рендеринга по предсказанным значениям для запрашиваемых точек пространства.

Несмотря на то, что модель NeRF явно не предназначен для создания поверхности, всё же ее можно получить, используя алгоритм марширующих кубов [26]. Однако полученная полигональная сетка будет «грубой» и потребует дополнительный этап постобработки.

1.3. Без информации о поверхности: как правило, производится 2D деформация исходного изображения, например, по ключевым точкам, оптическому потоку и т. п.

2. По обобщенности итоговой модели
   2.1. Общая модель: при создании цифровых аватаров для новых людей не требуется менять параметры аппроксимирующей функции.
   2.2. Персональная модель: для каждого человека при создании цифрового аватара производится точная настройка параметров аппроксимирующей функции.
3. По требованиям к набору данных для обучения
   3.1. 3D набор данных: состоит из высокоточных сканов головы и соответствующих им RGB изображений с нескольких ракурсов, также некоторые наборы включают шаблонную полигональную сетку единой топологии, к которой затем приводятся «сырые» полигональные сетки, и вручную сконструированные формы смешивания (англ. blend shapes) для различных атрибутов (мимика, форма и т. п.) [27], [28], [29], [30].
   3.2. 2D набор данных: состоит из набора RGB изображений [31].
   3.3. Синтетический набор данных: генерируется в специализированных средах, например [32]. Важно отметить, что, как правило, обучение лишь на таких наборах данных не позволяет достигнуть фотореалистичности, поэтому такие наборы используют в связке с 2D и/или 3D данными.

4. По возможности управления параметрами модели
    4.1. «Распутанное» пространство параметров: под «распутанным» пространством параметров в контексте диссертационного исследования понимается семантическая разделенность параметров, под которой подразумевается, что часть параметров отвечает за отдельные характеристики лица и их изменение в рамках заранее заданных границ не влечет изменение других характеристик. Так, например, в параметрической модели FLAME [29] выделяют группы независимых параметров для описания формы головы, позы головы и выражения лица. Параметры для выражения лица в частности отвечают за отдельные части лица, например, есть параметры, отвечающие конкретно за глаза, рот и т. д.
    4.2. Пространство без «распутывания» параметров: под пространством без «распутывания» параметров понимается пространство, которое обучалось без использования специализированных алгоритмов или функций потерь, позволяющих явно отделить группы параметров для их последующего регулирования.

Важно также отметить еще один аспект, который не является критерием классификации, но является важным при эксплуатации разработанного метода – необходимость предварительной оценки параметров параметрической модели. Данный этап подразумевает получение некоторых грубых априорных знаний о поверхности модели головы. Такой подход используется практически во всех современных методах создания аватара как головы, так и всего тела [33], [34], [35]. Данный этап используется в разработанном методе и будет подробнее описан в разделе 2.

1.2. Обзор существующих решений

Данный подраздел посвящен описанию наиболее значимых решений в области создания цифровых аватаров головы человека. Методы расположены в хронологическом порядке. Подраздел завершается таблицей, в которой все описанные методы классифицированы по каждому критерию в соответствии с предложенной ранее классификацией. Также дополнительно добавлен столбец, в котором отмечается необходимость этапа предварительной оценки параметров параметрической модели.

Basel Face Model (BFM) [36] – это параметрическая модель лица, представленная полигональной сеткой. Параметры модели разделяются на идентичность (англ. identity) и текстурную составляющую. Параметрами рендеринга (процесс получения изображения по модели) выступают условия освещения и положение камеры. Обучение модели производилось на наборе данных, содержащем 3D сканы 100 женщин и 100 мужчин, преимущественно европейцев. Возраст людей составлял от 8 до 62 лет, средний возраст 24,97, вес от 40 до 123 килограмм, средний вес – 66,48 килограммов. На этапе сканирования каждому человеку ставится в соответствие 3 3D скана в нейтральном положении и 3 фотографии. На этапе регистрации производится повторная параметризация сканов так, чтобы семантически соответствующие точки (кончик носа, углы глаз) имели одинаковое положение в рамках общей для всех сканов топологии (результат – полигональная сетка единой топологии). Такое соответствие устанавливается для всех точек лица, в том числе для щек, которые классифицируются как неструктурированная область. Затем производится формирование текстуры, где альбедо лица представляется одним цветом на вершину полигональной сетки, который вычисляется по фотографии. Информация с трех фотографий смешивается, а также вручную удаляется область волос и с помощью диффузии дополняются недостающие данные. После описанных этапов каждое лицо параметризуется как полигональная сетка с 53490 вершинами и одинаковой топологией, где каждой

вершине соответствует RGB-цвет. Авторы предполагают независимость между формой и текстурой, поэтому создают две независимые линейные модели. Многомерное нормальное распределение подгоняется к данным с использованием метода главных компонент. В результате получаются две модели – для формы и для текстуры. Новое лицо можно сгенерировать в виде линейной комбинации главных компонент. Процедуру оценки параметров модели можно производить по фотографии, решая задачу оптимизации (например, рендер оптимизируемой модели сравнивать с исходным изображением).

FaceWarehouse [28] – это параметрическая модель лица, представленная полигональной сеткой. Параметры модели разделяются на идентичность и выражение лица. Авторы собрали набор данных, содержащий информацию о 150 людях с возрастом от 7 до 80 лет, принадлежащих к разным этническим группам. Для каждого человека был выполнен захват нейтрального выражения лица и еще 19 выражений, таких как открытый рот, улыбка, гнев, закрытые глаза и прочие, с использованием RGBD-камеры. Выражения лица для захвата выбирались по принципу наибольшего различия для разных людей. Для всех захваченных данных производилась автоматическая локализация 74 ключевых точек лица, которые при необходимости подвергались ручной корректировке. Затем в соответствии с облаками точек, которые описывают поверхность лица, производилась деформация шаблонной полигональной сетки лица для нейтрального выражения, предварительно полученного с помощью оценки параметров параметрической модели BFM. Каждой ключевой точке на изображении с нейтральным выражением сопоставлялась вершина полигональной сетки. Для изображений с другими выражениями лица итоговые полигональные сетки формировались на основе нейтральной путем переноса деформаций. После получения всех полигональных сеток для каждого субъекта создавались индивидуальные формы смешивания (англ. blendshapes) для выражения лица. В результате все захваченные данные представляются в виде полигональных сеток единой топологии. Затем

выполняется построение билинейной модели лица с двумя атрибутами – идентичностью и выражением лица. Для этого применяется сингулярное разложение к тензору, содержащему информацию о вершинах полигональной сетки каждого человека и соответствующих ему форм смешивания для выражения лица. Процедуру оценки параметров модели можно также производить по фотографии, решая задачу оптимизации.

В работе *Bringing Portraits to Life* [37] используется изображение человека, которого необходимо анимировать, и видеопоследовательность, с которой будет выполняться перенос мимики (видеопоследовательность может содержать мимику другого человека). Процесс анимации реализуется путем применения серии 2D деформаций целевого изображения. Деформации выступают описанием мимики, захваченной по кадрам видеопоследовательности. Так как применение лишь деформаций не позволяет достигнуть достаточного реализма, авторы добавляют в рассмотрение небольшие динамические детали, такие как складки и морщины, а также добавляют возможность при необходимости реалистично отображать скрытые области (например, область рта). Для входного изображения и кадров видеопоследовательности извлекаются ключевые точки (как для области лица, так и вне, для того чтобы анимация выглядела реалистично и учитывала изменения положения головы), которые затем сопоставляются. Для набора соответствий вычисляются смещения, на основе которых производятся деформации. Перед применением деформаций производится выравнивание граней (линий, соединяющих ключевые точки). Параметры преобразования выравнивания определяются один раз для исходного изображения и нейтрального кадра, соответствующему ему (подразумевается соответствие выражения лица, не обязательно первый кадр видеопоследовательности). Если на кадре видеопоследовательности область рта в открытом состоянии, то выполняется перенос внутренней части на целевое изображение (перенос губ не выполняется для сохранения реалистичности). Также для большей

реалистичности выполняется перенос морщин и складок с кадра на целевое изображение.

Faces Learned with an Articulated Model and Expressions (FLAME) [29] – это параметрическая модель лица, представленная полигональной сеткой. Параметры модели разделяются на форму головы (идентичность), позу головы и выражение лица. Подробное описание модели FLAME будет приведено в подразделе 2.2.

В работе *High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs* [38] предложен метод синтеза фотореалистичных изображений высокого разрешения из семантических карт меток с использованием условных генеративно-состязательных сетей. Здесь производится вычисление карты границ объектов (так как карты сегментации не различают объектов одной категории при их перекрытии), которая затем объединяется с картой семантических меток и картой признаков исходного изображения, полученной из совместно обученного энкодера, для передачи в генератор. Так, для анимации, в генератор можно передавать карту семантической сегментации и карту границ объекта, полученные по изображению с ожидаемым выражением лица (не обязательно того же человека), а также полученную из энкодера карту признаков изображения, на которое требуется выполнить перенос выражения лица.

В работе *GIF: Generative Interpretable Faces* [39] для генерации фотореалистичных изображений используется модель StyleGAN2 [40], обусловленная параметрами параметрической модели головы FLAME. Цель авторов – получение 2D генеративной модели лица, которая управляется параметрической моделью. Однако из-за отсутствия учета пространственной геометрии ни в виде данных, ни в виде архитектурного решения, при синтезе видов с новых ракурсов присутствует несогласованность.

В работе *First Order Motion Model for Image Animation* [41] представлена модель, которая, обучаясь на видеопоследовательностях одного домена, например, на лицах, может выполнять анимацию изображения (перенос

выражения лица; содержимое изображения должно быть из того же домена) по любой видеопоследовательности, удовлетворяющей домену. Для этого производится разделение информации о внешнем виде и движении. Для эффективной обработки сложных движений используется представление, состоящее из набора изученных ключевых точек вместе с их локальными и аффинными преобразованиями. Представленный подход состоит из двух модулей: модуль оценки движения и модуль генерации изображения. Цель модуля оценки движения состоит в том, чтобы спрогнозировать поле движения от кадра видеопоследовательности к исходному кадру, а также карту перекрытия. Полученное поле движения используется затем для выравнивания карты признаков исходного изображения. Модуль генерации изображения производит отрисовку исходного изображения с учетом переносимой анимации. Для этого сеть генератор искажает исходное изображение на основе полученного поля движения и заполняет/закрашивает (англ. inpainting) части изображения, которые перекрыты в исходном изображении.

FaceScape [27] – это параметрическая модель лица, представленная полигональной сеткой. Параметры модели разделяются на идентичность и выражение лица. Авторы представили крупномасштабный набор данных 3D лиц, который содержит 18760 текстурированных 3D лиц, захваченных с 938 субъектов возрастом от 16 до 70 лет, преимущественно азиатов, каждому из которых соответствует 20 различных выражений лица, включая нейтральное (аналогично *FaceWarehouse*). В отличие от *FaceWarehouse*, где захват производился на RGBD-камеру, что приводило к разреженным поверхностям до обработки, авторы производили захват с использованием плотного набора 68 зеркальных камер с контролируемым освещением. В результате реконструируемые 3D модели лица обладали высоким уровнем детализации форм морщин и пор. Также вместе со сканами были получены соответствующие RGB-изображения и метаинформация о субъектах (например, возраст, пол и место работы). На основе собранного набора данных была создана параметрическая модель лица. Все необработанные сканы

выравнивались по ключевым точкам и преобразовывались в топологически единообразную базовую модель, представляющую приблизительную форму, и карту смещения в UV-пространстве для представления подробной формы. Затем выполнялся поиск точек на поверхности базовой полигональной сетки, соответствующих пикселям на карте смещения, после чего осуществлялась обратная проекция точек на необработанную полигональную сетку для поиска соответствующих точек. Значение пикселей карты смещения устанавливались в соответствии с расстоянием со знаком (SDF) от точки базовой полигональной сетки до соответствующей точки необработанной полигональной сетки. Таким образом, базовая полигональная сетка используется для грубого представления геометрии, а карта смещения для представления детальной геометрии. Такие представления далее используются для построения билинейной модели, разложенной на параметры идентичности и выражения лица. Было сгенерировано по 52 формы смешивания для каждого человека. Затем с помощью алгоритма Такера производилось разложение тензора, содержащего информацию о вершинах и формах смешивания для выражения лица каждой полигональной сетки, на два низкоразмерных компонента – идентичность и выражение лица. В отличие от предыдущих параметрических моделей, таких как *BFM*, *FaceWarehouse*, *FLAME*, которые ограничены грубым оцениванием форм смешивания, авторы сделали акцент на восстановлении морщин и прочих мелких особенностей, в том числе на их деформации при изменении выражения лица. Для этого было введено понятие «динамические детали», которые предсказываются по одному изображению с помощью нейронной сети, обученной на собранном наборе данных. После оценки параметров модели вычисляется карта деформации, которая моделирует движение поверхности, как разницу 3D положения вершин нейтрального выражения лица до целевого выражения в UV-пространстве. Карта деформации для всех выражений (19, кроме нейтрального) и текстуры поступают на вход нейронной сети, которая выдает результирующие карты смещений для всех выражений. Затем полученные карты смещения взвешиваются весами формы смешивания,

полученными на этапе оценки параметров, для получения карты смещения целевого выражения лица. Таким образом, предложенная система состоит из трех этапов: оценка параметров базовой модели, предсказание карты смещений и синтез динамических деталей.

i3DMM [42] – это неявная параметрическая модель головы, обученная на 3D сканах 64 человек, которая в отличие от базовых параметрических моделей (*BFM*, *FaceWarehouse* и т. п.) включает в себя волосы. Данная модель состоит из трех нейросетевых блоков: DeformNet – сеть, которая принимает на вход точку, в которой требуется предсказать значение функции расстояния со знаком (SDF), и латентные вектора, описывающие форму, выражение лица и прическу, и выдает вектор смещения, который деформирует точку запроса. Деформированная точка запроса передается на вход сети RefNet, которая предсказывает значение функции расстояния со знаком (SDF). Также деформированная точка и латентные коды формы и прически подаются в ColorNet для предсказания цвета в точке. Данная модель предназначения для работы с 3D сканами. Здесь аналогично моделям *FLAME*, *BFM, FaceWarehouse*, *FaceScape* фиксируются веса нейросетевых блоков и производится оценка латентных кодов для входного предобрботанного 3D скана головы.

В работе *Fast Bi-layer Neural Synthesis of One-Shot Realistic Head Avatars* [43] предложен метод создания цифрового аватара на основе нейронного рендеринга, который требует лишь одну фотографию. В основе лежит двухслойное представление, в котором изображение аватара в новой позе генерируется путем суммирования двух компонент: грубого изображения, предсказанного сетью рендеринга, принимающей на вход эмбеддинг исходного изображения и набор новых ключевых точек, и изображения деформированной текстуры, где текстура предсказывается сетью генерации текстуры, принимающей на вход эмбеддинг исходного изображения и соответствующие ему ключевые точки. Карта деформации текстуры предсказывается сетью рендеринга и участвует в формировании

деформированной текстуры. Хотя карта деформации предсказывается сетью рендеринга, текстура оценивается во время создания аватара (при первом проходе) и является статичной во время генерации изображений с разными позами. Данный метод не требует точной настройки весов для каждого нового аватара. Для анимации аватара на вход могут поступать кадры с другими людьми.

В *VariTex* [44] для генерации фотореалистичных изображений используется вариационный автокодировщик, обусловленный параметрами параметрической модели лица. Так, аналогично [39] авторы получили 2D генеративную модель лица, управляемую параметрической моделью. Представленные результаты показывают более согласованные виды с разных ракурсов в отличие от [39], однако из-за отсутствия учета пространственной геометрии на некоторых ракурсах по-прежнему не хватает согласованности.

Авторы *DECA* [45] представляют подход, который регрессирует форму и динамические детали лица специфичные для человека, меняющиеся в зависимости от выражения лица. Существующие параметрические методы (*BFM*, *FLAME* и пр.) способны хорошо восстанавливать геометрию, но страдают от ряда ограничений, например, при выполнении процедуры анимации не удается точно воспроизвести деформацию морщин в зависимости от выражения лица. Для обхода этого ограничения авторы определяют новую функция потерь, которая отделяет детали, специфичные для человека, от деталей, зависящих от выражения лица. Такое «распутывание» пространства позволяет реалистично синтезировать морщины, специфичные для каждого человека, контролируя параметры выражения лица, сохраняя при этом детали, специфичные для человека, неизменными. То есть производится декомпозиция параметров формы головы параметрической модели *FLAME* на статические детали и детали, зависящие от выражения лица. Модель DECA обучена создавать карту UV-смещения из низкоразмерного латентного представления, которое состоит из параметров деталей специфичных для человека (зависящих от выражения лица) и общих параметров выражения лица и позы головы. Для

этого используется блок энкодер-декодер. Через этот блок во время обучения протекают градиенты, вычисленные в том числе с использованием введенной функции потерь. Также отдельный энкодер предсказывает параметры формы (статические, для встраивания в модель *FLAME*), позы головы, выражения лица, освещения, и параметров камеры из входного изображения. Основной результат работы – это синтез правдоподобных деформаций геометрических деталей при изменении выражения лица, что улучшает классическую параметрическую модель *FLAME*. Модель DECA работает на изображениях разных людей и не требует повторного обучения. Анимация аватара возможна по последовательности изображений с другим человеком.

В *NeRFace* [46] представлен метод создания цифрового аватара на основе условных нейронных полей излучений (CNeRF). Предложенная архитектура обучается индивидуально для каждого человека по видеопоследовательности. На вход поступает видеопоследовательность и фотография фона. Кадры видеопоследовательности поступают на вход алгоритма оценки параметров параметрической модели, в результате каждому кадру соответствуют параметры позы и выражения лица, а также внутренние параметры камеры. Затем производится семплирование лучей с использованием информации о позе лица и внутренних параметров камеры. На вход многослойного перцептрона, используемого в рамках нейронной модели представления поверхности CNeRF, поступают координата точки в трёхмерном пространстве, вектор направления (позиционное кодирование как в оригинальной работе, представляющей нейронную модель NeRF [25]), а также вектор выражения лица δ и обучаемый вектор γ (свой для каждого кадра видеопоследовательности), цель которого заключается в компенсации недостающей информации вектора выражения лица δ, который ограничен базисом форм смешивания параметрической модели. На выходе каждой точке пространства соответствует пара – цвет и плотность σ. После обработки всех просемплированных лучей и точек на них производится процедура объемометрического рендеринга для получения итогового изображения. В

работе авторы также отделяют фон от субъекта, внося предположение о том, что последняя просемплированная точка на каждом луче будет совпадать с точкой фона фиксированного цвета. В результате обучается представление аватара в неявном виде, для которого можно выполнить перенос выражения лица и позу головы по другим видеопоследовательностям (с другими людьми в том числе).

Авторы *Neural Head Avatar* [47] представляют неявное нейронное представление, моделирующее геометрию поверхности и внешний вид цифрового аватара. Для получения поверхности аватара на вход требуется видеопоследовательность. Сначала для грубой оценки полигональной сетки применяется алгоритм оценки параметров параметрической модели FLAME, затем используются две нейронные сети прямого распространения, одна из которых предсказывает величину смещения для вершин полигональной сетки, а другая текстуру, которая зависит от выражения лица и позы. Так, на первом этапе метода производится оценка параметров параметрической модели FLAME для каждого кадра. Результат – каждому кадру соответствуют параметры формы, позы и выражения лица. На втором этапе метода вектор позы поступает на вход нейронной сети, уточняющей геометрию. Выход сети – смещения для каждой вершины, которые в результате формируют реалистичную полигональную сетку с волосами и мелкими деталями. На третьем этапе на вход текстурной нейросети, которая генерирует фотореалистичную текстуру, поступают трехмерная координата вершины канонической полигональной сетки модели FLAME, вектора выражения и позы для текущего кадра, а также локальный патч отрендеренных нормалей. Такое обуславливание нейронной сети позволяет синтезировать эффекты, зависящие от выражения лица и положения. Стратегия оптимизации при настройке весов следующая: сначала производится грубая оценка полигональной сетки с помощью алгоритма оценки параметров параметрической модели головы FLAME, затем оптимизация нейронной сети, которая отвечает за уточнение геометрии. После этого веса сети, уточняющей

геометрию, фиксируются и производится оптимизация текстурной нейронной сети. Финальный шаг – совместная оптимизация двух сетей. Важно отметить, что для каждого человека требуется настройка весов нейронных сетей, то есть от субъекта к субъекту сохраняется лишь архитектура.

В *HeadNeRF* [48] представлен метод создания цифрового аватара на основе нейросетевой модели представления поверхности CNeRF, однако в отличие от [46] классическая процедура объемометрического рендеринга заменяется нейронным рендерингом, что приводит к ускорению процедуры генерации итогового изображения. Как следствие – на выходе многослойного перцептрона, используемого в рамках нейросетевой модели NeRF, не стандартная связка (цвет, плотность), а пара – (вектор признаков, плотность). Аналогично производится семплирование лучей и точек на них, полученные вектора признаков конкатенируются в карту признаков, которая поступает на вход модели нейронного рендеринга, выход которой – итоговое изображение аватара в зависимости от параметров камеры, идентичности, выражения лица, альбедо и освещения. Обучение модели производится на двух наборах данных – сначала на 3D, затем небольшая настройка на 2D [31] для большей детализации итоговых аватаров. Метод HeadNeRF работает на изображениях разных людей и не требует повторной настройки весов. Для создания аватара сначала выполняется оценка параметров параметрической модели BFM, полученные вектора затем поступают на вход метода HeadNeRF и модифицируются путем оптимизации (при оптимизации сравниваются выход метода HeadNeRF и исходное изображение). Для анимации полученного аватара необходимо выполнить замену вектора выражения лица. Перенос выражения лица может выполняться по видеопоследовательности с другим человеком.

Авторы *IMAvatar* [49] представляют метод создания цифрового аватара на основе неявного нейронного представления. Для этого вводятся нейронные неявные поля, определяющие каноническую геометрию, деформации и текстуру человека. Представление канонической геометрии описано

многослойным перцептроном, который предсказывает значение занятости для каждой канонической 3D точки (на вход подаются точки, полученные в результате процедуры трассировки лучей; при трассировке лучей точке в деформированном пространстве сопоставляется точка в каноническом пространстве как раз на основе параметров сети деформации и параметров выражения лица и позы, полученных с помощью модели DECA). Аналогично [46] производится дополнительное обуславливание сети обучаемым кодом для каждого кадра, а также используется позиционное кодирование для входных данных. Сеть деформации предсказывает вектор формы смешивания выражения лица для вычисления аддитивного сдвига вершины, матрицу корректировки позы для добавления сдвига в вершине и веса линейно-переходного кожного покрова (англ. linear blend skinning, LBS; перевод взят из книги [50]) для каждой точки в каноническом пространстве. Для вычисления цвета в точке также используется многослойный перцептрон, который обусловлен картой направления нормалей для деформированного представления головы, чтобы корректно отображать эффекты освещения. В данной работе так же, как и в [47], требуется настраивать веса для каждого человек отдельно.

В работе *MoFaNeRF* [51] предлагается интеграция NeRF-подобного блока, называемого MoFaNeRF и обусловленного векторами внешнего вида (фотометрические признаки, такие как цвет кожи, губ и зрачков), формы (геометрия и положение носа, глаз, области рта и овала лица) и выражения лица, и сети архитектуры энкодер-декодер RefineNet. Блок MoFaNeRF включает в себя два нейросетевых модуля: TEM – сверточный модуль, выполняющий перенос кода внешнего вида в пространство многослойного перцептрона, и ISM – модуль, вдохновленный концепцией AdaIN [31], для учета специфичных для каждого человека характеристик формы при одинаковых выражениях лица. Модель RefineNet выполняет уточнение грубого рендера, полученного с помощью блока MoFaNeRF и процедуры объемометрического рендеринга. Метод на основе блока MoFaNeRF работает

на изображениях разных людей и не требует повторного обучения. Анимация аватара возможна по последовательности изображений с другим человеком.

В работе *FitDiff* [52] авторы представляют метод создания цифрового аватара лица на основе диффузии. Предложенная мультимодальная диффузионная модель одновременно выдает карту нормали, диффузное и зеркальное альбедо, а также полигональную сетку субъекта. Авторы используют диффузионные модели для процессов генерации и оценки параметров, поскольку по своей природе они очень устойчивы в обоих процессах, так как работают непосредственно с пространством изображений. Аватар лица определяется как комбинация полигональной сетки S и текстуры T, которая включает в себя UV-карту отражательных характеристик, а именно диффузного альбедо, зеркального альбедо и карт нормалей. Учитывая целевой эмбеддинг, соответствующий исходному изображению, который извлекается из модели для распознавания лица, векторов, полученных путем применения алгоритма оценки параметров параметрической модели (в том числе характеризующих текстуру и освещение), метод FitDiff генерирует аватар лица с теми же параметрами освещения. Обучение модели двухэтапное. На первом этапе обучается автоэнкодер, разворачивающий текстуру. На втором этапе обученные веса автоэнкодера замораживаются и обучается шумоподавляющий U-Net. На выходе модели получаются параметры, которые могут быть интегрированы в исходную параметрическую модель, полученную в ходе процедуры оценки параметров параметрической модели. Обученные веса не требуют повторной настройки под конкретного субъекта. Важно также отметить, что для работы метода можно передать как изображение и априорную информацию, полученную в ходе оценки параметров параметрической модели, что-то из перечисленного отдельно или вообще ничего. Адекватные модели могут быть сгенерированы из шума, сэмплированнаго из нормального распределения.

В *GAN-Avatar* [53] представлен метод создания цифрового аватара, который обучается захватывать внешний вид без предварительно полученных

параметров выражения лица (без оценки параметров параметрической модели; в действительности параметры предсказываются, но с помощью нейросетевого метода). Для контроля внешнего вида обучается сеть сопоставления, которая позволяет обходить пространство параметров параметрической модели. В основе метода лежат 3D генеративная модель внешнего вида, которая использует предварительно обученную модель EG3D [54], и сеть сопоставления параметров выражения лица (многослойный перцептрон). Итоговый результат – обученные веса моделей для каждого человека. Обучение является двухэтапным. Сначала выполняется точная настройка весов модели EG3D, предварительно обученной на наборе данных FFHQ, для каждого субъекта по входным кадрам. Предварительная инициализация позволяет увеличить скорость сходимости, так как веса содержат информацию о лицах с разными выражениями лица. Затем из фронтальных изображений лица, сгенерированных моделью (т. к. они легки для реконструкции), генерируются параметры выражения лица с помощью метода Deep3DFace [55]. Сеть сопоставления параметров выражения лица предсказывает латентный код по входному коду (из метода Deep3DFace). Сгенерированный код поступает в модуль генератора, который синтезирует ожидаемое изображение (модуль генератора заморожен во время обучения сети сопоставления выражения лица). Положение головы учитывается при синтезе изображения в модуле нейронного рендеринга.

В *Relightify: Relightable 3D Faces from a Single Image via Diffusion Models* [56] авторы конструируют высококачественную статистическую модель для генерации текстуры, карты нормалей и значения отражательной способности с помощью диффузионной модели, которая применяется для решения задачи закрашивания/заполнения UV-текстуры, полученной в результате оценки параметров параметрической модели головы, что в результате приводит к реалистичным изображениям. Анимация аватара возможна путем замены параметра выражения лица для полигональной сетки, полученной в результате оценки параметров параметрической модели.

В *PointAvatar* [57] на основе видеопоследовательности, в которой субъект меняет выражение лица и позу головы, модель совместно обучает облако точек, представляющее независимую от позы геометрию и внешний вид субъекта в каноническом пространстве; сеть деформации, которая преобразует облако точек в новые позы, используя параметры выражения лица и позы параметрической модели FLAME (выполняется оценка параметров для каждого кадра); сеть затенения, которая выдает вектор затенения для каждой точки на основе нормалей точек в деформированном пространстве. Совместная оптимизация этих трех компонент производится путем сравнения результата рендеринга затененных точек в деформированном пространстве с исходными кадрами. После оптимизации можно выполнять синтез новых видов субъекта с новыми позами, выражениями лица и условиями освещения. Каноническое пространство состоит из набора N обучаемых точек (обучается их местоположение). Для оптимизации авторы инициализируют разреженное облако точек, случайно просемплированное на сфере и периодически повышают дискретизацию облака точек, уменьшая радиус рендеринга. Предложенный метод разделяет цвета точек на компонент альбедо и компонент затенения (выводится из нормалей в деформированном пространстве). Для контролируемой анимации авторы используют те же выражения лица и позу, которую получили в результате оценки параметров параметрической модели FLAME. Деформация точек – двухэтапная. На первом этапе канонические точки деформируются в промежуточное положение, которое соответствует предопределенной (шаблонной) позе параметрической модели головы FLAME (с открытым ртом). На втором этапе используется целевое выражение лица и позы (полученные с помощью оценки параметров модели FLAME) для преобразования точки в деформированное пространство на основе обученных форм смешивания (выражение лица, пока) и весов LBS. В частности, многослойный перцептрон на основе координат выполняет сопоставление каждой точке в каноническом пространстве – сдвига, который переводит точку канонического пространства в

соответствующее положение точки пространства модели FLAME. Затем производится деформация на основе форм смешивания и весов LBS. После получения деформированной геометрии применяется дополнительная сеть затенения, которая ожидает на вход нормали в деформированном пространстве. Выходные значения умножаются на цвета альбедо для получения цветов с эффектом затенения, которые затем преобразуются в пиксели изображения посредством дифференцируемой растеризации. В данной работе требуется настраивать веса для каждого человек отдельно.

В *Instant Volumetric Head Avatars* [58] представлен метод создания цифрового аватара головы человека на основе нейросетевой модели представления поверхности CNeRF. Авторы реализуют нейросетевую модель NeRF с использованием нейронных графических примитивов [59], где пространство представлено в виде иерархии воксельных сеток (вокселы имеют разный масштаб). Каждой точке пространства соответствует вектор признаков. Он является конкатенаций результатов линейной интерполяции обучаемых векторов признаков для вершин вокселов на каждом уровне иерархии. Такая модификация предоставляет возможность использовать небольшой многослойный перцептрон для получения пар (цвет, плотность), так как часть информации о сцене будет включена в векторы признаков. За счет использования небольшого перцептрона и оптимизированной процедуры формирования векторов признаков [59], процедура рендеринга выполняется в режиме реального времени. Предложенный подход подразумевает обучение параметров для каждого человека. На вход поступает видеопоследовательность. Кадры видеопоследовательности поступают на вход процедуры оценки параметров параметрической модели головы FLAME. В результате каждому кадру соответствуют параметры формы, выражения лица, позы и камеры (внутренние и внешние). На вход многослойного перцептрона поступает вектор признаков, соответствующий точке, и параметры выражения лица. Выходом является пара – цвет и плотность. После обработки всех просемплированных лучей и точек на них производится

процедура объемометрического рендеринга для получения итогового изображения. Важно отметить, что аналогично методу *IMAvatar* для получения вектора признаков производится перевод координат из деформированного пространства в каноническое. Для анимации полученного представления необходимо выполнить замену параметров выражения лица. Параметры выражения лица могут быть извлечены из видеопоследовательности с другим человеком.

В середине 2023 на международной конференции SIGGRAPH был представлен новый подход для пространственного представления сцены – 3D Gaussian Splatting (3DGS) [60]. Основная задача, решаемая с помощью подхода 3DGS, аналогична задаче, решаемой с помощью нейросетевой модели представления поверхности NeRF, и заключается в синтезе новых видов, однако теперь пространство представлено 3D гауссианами, где каждая описывается положением, ковариационной матрицей, прозрачностью и сферическими гармониками. Также метод отрабатывает в режиме реального времени за счет предложенного авторами алгоритма дифференцируемого рендеринга. Из новых работ, представленных в 2024 году на конференции CVPR, по созданию цифровых аватаров головы человека на основе подхода 3DGS, можно отметить [61], [62], однако авторы заявляют о необходимости использования многокамерных систем для сбора входных данных. Также следует отметить работы [63] и [64], которые свидетельствуют о высоком потенциале использования методов на основе нейросетевой модели NeRF.

В таблице 1 перечислены все описанные методы в соответствии с приведенной ранее классификацией.

Таблица 1 – Классификация всех описанных методов

Критерий / Название метода	Способ представления поверхности аватара	Уровень обобщенности модели	Формат набора данных для обучения	Возможность управления параметрами модели	Необходимость оценки параметров параметрической модели
BaselFace [36]	Явное представление поверхности	Общая модель	3D набор данных	«Распутанное» пространство параметров	Не требуется
FaceWarehouse [28]	Явное представление поверхности	Общая модель	3D набор данных	«Распутанное» пространство параметров	Не требуется
Bringing portraits to life [37]	Без информации о поверхности	Общая модель	2D набор данных	Пространство без «распутывания» параметров	Не требуется
FLAME [29]	Явное представление поверхности	Общая модель	3D набор данных	«Распутанное» пространство параметров	Не требуется
High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs [38]	Без информации о поверхности	Общая модель	2D набор данных	Пространство без «распутывания» параметров	Не требуется
GIF: Generative Interpretable Faces [39]	Без информации о поверхности	Общая модель	2D набор данных	«Распутанное» пространство параметров	Требуется
First Order Motion Model for Image Animation [41]	Без информации о поверхности	Общая модель	2D набор данных	Пространство без «распутывания» параметров	Не требуется

FaceScape [27]	Явное представление поверхности	Общая модель	3D набор данных	«Распутанное» пространство параметров	Не требуется
i3DMM [42]	Явное представление поверхности	Общая модель	3D набор данных	«Распутанное» пространство параметров	Не требуется
Fast Bi-layer Neural Synthesis of One-Shot Realistic Head Avatars [43]	Без информации о поверхности	Общая модель	2D набор данных	Пространство без «распутывания» параметров	Не требуется
VariTex [44]	Без информации о поверхности	Общая модель	2D набор данных	«Распутанное» пространство параметров	Требуется
DECA [45]	Явное представление поверхности	Общая модель	2D набор данных	«Распутанное» пространство параметров	Не требуется
NeRFace [46]	Неявное представление поверхности	Персональная модель	2D набор данных	«Распутанное» пространство параметров	Требуется
Neural Head Avatar [47]	Явное представление поверхности	Персональная модель	- 2D набор данных - синтетический набор данных	«Распутанное» пространство параметров	Требуется
HeadNeRF [48]	Неявное представление поверхности	Общая модель	- 2D набор данных - 3D набор данных	«Распутанное» пространство параметров	Требуется

IMAvatar [49]	Явное представление поверхности	Персональная модель	- 2D набор данных - синтетический набор данных	«Распутанное» пространство параметров	Требуется
MoFaNeRF [51]	Неявное представление поверхности	Общая модель	3D набор данных	«Распутанное» пространство параметров	Требуется
FitDiff [52]	Явное представление поверхности	Общая модель	2D набор данных	«Распутанное» пространство параметров	Требуется
GAN-Avatar [53]	Неявное представление поверхности	Персональная модель	- 2D набор данных - 3D набор данных	«Распутанное» пространство параметров	Не требуется
Relightify: Relightable 3D Faces from a Single Image via Diffusion Models [56]	Явное представление поверхности	Общая модель	2D набор данных	«Распутанное» пространство параметров	Требуется
PointAvatar [57]	Явное представление поверхности	Персональная модель	2D набор данных	«Распутанное» пространство параметров	Требуется
Instant Volumetric Head Avatars [58]	Неявное представление поверхности	Персональная модель	2D набор данных	«Распутанное» пространство параметров	Требуется

1.3. Выводы и результаты первого раздела

В данном разделе был представлен подробный обзор наиболее значимых решений в области создания цифровых аватаров головы человека. Большинство работ было представлено на лучших международных конференциях в области компьютерного зрения и машинного обучения, например, таких как Computer Vision and Pattern Recognition Conference (CVPR), International Conference on Computer Vision (ICCV), European Conference on Computer Vision (ECCV), а также опубликовано в журналах с высоким рангом.

По результатам можно заключить, что одно из основных направлений развития – это интеграция «грубой» априорной информации о поверхности головы, полученной в результате оценки параметров параметрической модели, в нейросетевую модель, неявно описывающую геометрию (см. последний столбец в таблице 1). Такие методы показывают лучшую степень схожести между субъектом и синтезируемыми изображениями при обзоре с разных точек, а также более реалистичную анимацию. Однако, некоторые из них требуют для обучения аннотированные 3D наборы данных, которые зачастую ограничены количеством, а также имеют смещения в контексте расы, пола и возраста. Можно также заметить корреляцию между общностью модели и форматом набора данных для обучения. Методы для получения персональной модели, как правило, обучаются на 2D наборе данных. В случаях с общей моделью при явном или неявном представлении поверхности часто используется 3D набор данных. Очевидно, что использование общей модели упрощает процесс получения и анимации аватара, однако в некоторых случаях это приводит к снижению визуального качества, которое отражается в недостаточной схожести. Методы, нацеленные на получение персональной модели, часто требуют немало времени на обучение, так как настройка весов производится с нуля, что также приводит к необходимости контроля переобучаемости и введения дополнительных членов регуляризации.

Таким образом, учитывая описанные выше недостатки существующих методов, в рамках диссертационного исследования предлагается разработать новый

метод создания цифрового аватара головы человека. Подробное описание предложенного метода приведено в разделе 4. Также в разделе 4 приведены результаты экспериментальных исследований метода и его сравнение с существующими актуальными решениями. В состав разработанного метода предлагается включить параметрическую модель головы на основе нейросетевой модели представления поверхности CNeRF и двумерного нейронного рендеринга, подробно описанную в разделе 3. Такая модель формирует общее представление о домене и предоставляет стартовые значения весов для разработанного метода создания цифрового аватара головы человека.

На рисунке 2 приведено схематичное представление разработанного метода создания цифрового аватара головы. Ключевые компоненты метода определяют структуру диссертации.

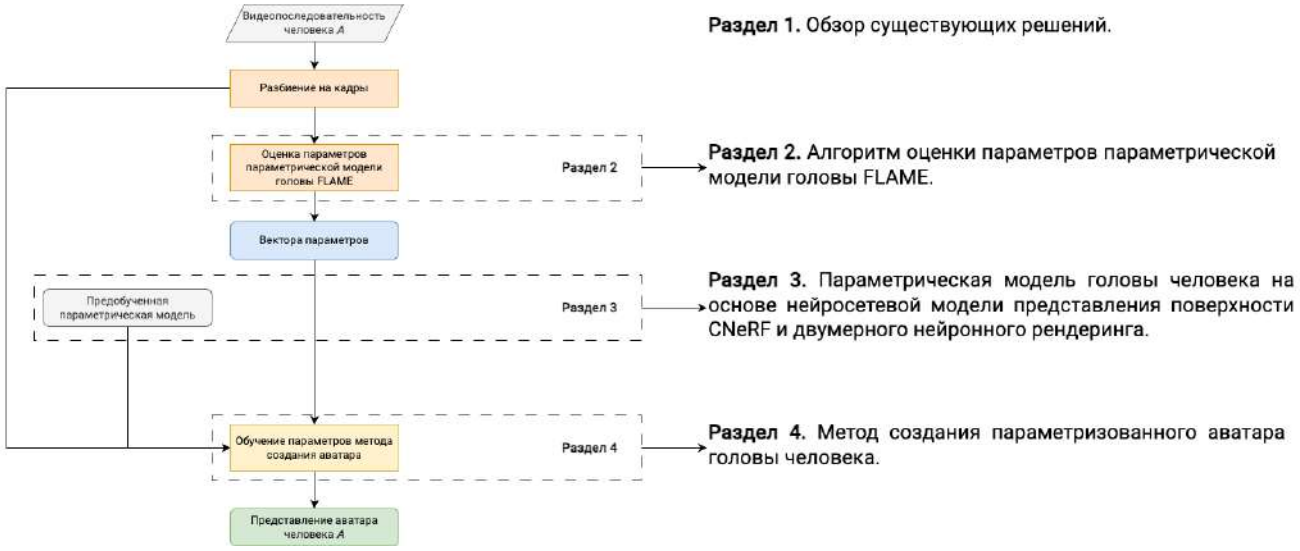


Рисунок 2 – Схематичное представление разработанного метода создания аватара головы человека и его соответствие со структурой диссертации

РАЗДЕЛ 2. АЛГОРИТМ ОЦЕНКИ ПАРАМЕТРОВ ПАРАМЕТРИЧЕСКОЙ МОДЕЛИ ГОЛОВЫ FLAME

Данный раздел диссертационного исследования посвящен параметрической модели головы FLAME и включает в себя: описание модели и ее параметров; классический алгоритм оценки параметров модели FLAME по RGB кадру; разработанный алгоритм оценки параметров модели FLAME с использованием RGBD кадра.

2.1. Необходимость использования параметрической модели

Практически все современные методы создания цифровых аватаров используют информацию из параметрической модели головы/лица. Это связано с тем, что она дает «грубое» представление о поверхности. Эффективность параметрических моделей обусловлена тем, что поверхность может контролироваться/изменяться различными независимыми параметрами (разнообразие контролируемых параметров зависит от конкретной модели), например, в модели FLAME – идентичность/форма, поза и выражение лица. Однако у таких моделей есть большой недостаток – ограниченная топология, из-за которой не всегда удается воспроизвести результат с требуемой реалистичностью. Так, используя параметрическую модель для создания аватара, практически во всех случаях нет возможности моделировать прическу и мелкие детали (например, морщины).

Современные подходы, как правило, не используют формат явного представления поверхности, а скорее интегрируют параметры, полученные в результате оценки параметров параметрической модели. Иными словами, реализуют обуславливание модели. Факт обусловленности позволяет варьировать внешний вид цифрового аватара в случае общей модели, или же изменять атрибуты головы и лица, не привязанные к идентичности (как для общей, так и для персональной модели). Таким образом, модель, описывающая аватара, становится более гибкой, так как появляется независящий от неё источник регулирования.

В рамках данного диссертационного исследования за основу взята параметрическая модель головы FLAME. Выбор в пользу модели был сделан по ряду причин: модель была получена на основе большого числа 3D сканов головы людей разной расы, пола и возраста; наличие шеи в топологии (в отличие от не менее популярной модели BFM); возможность регулирования позы головы и выражения лица.

В следующем подразделе приведено описание процесса создания параметрической модели головы FLAME и описание параметров с визуализацией их варьируемости.

2.2. Описание параметрической модели головы FLAME

FLAME [29] – это параметрическая модель головы, представленная полигональной сеткой. Параметры модели разделяются на форму головы (идентичность), позу головы и выражение лица. В отличие от ранних часто используемых моделей лица [36], [28], которые обучались на наборах данных с ограниченным количеством 3D сканов лица (FaceWarehouse, сканы для 150 людей; BFM, сканы для 200 людей, преимущественно молодые европейцы), модель FLAME спроектирована, основываясь более чем на 33000 сканах головы.

Авторы FLAME адаптируют формулировку параметрической модели тела SMPL [65] для представления головы с возможностью анимации. Поскольку многие деформации лица связаны с активацией мышц и не зависят от изменения артикулированной позы, в работе расширено представление головы в SMPL путем введения пространства форм смешивания для выражения лица. Кроме того, добавлена возможность моделирования позы головы, включая артикуляцию шеи и глазных яблок.

В модели FLAME используется техника LBS на основе вершин с корректирующими формами смешивания (англ. blendshapes). В технике LBS участвуют: вершины полигональной сетки головы в количестве $N = 5023$,

подвижные сочленения (суставы) в количестве $K = 4$ (для шеи, челюсти и глазных яблок, см. рис. 3) и формы смешивания, обученные на основе данных.

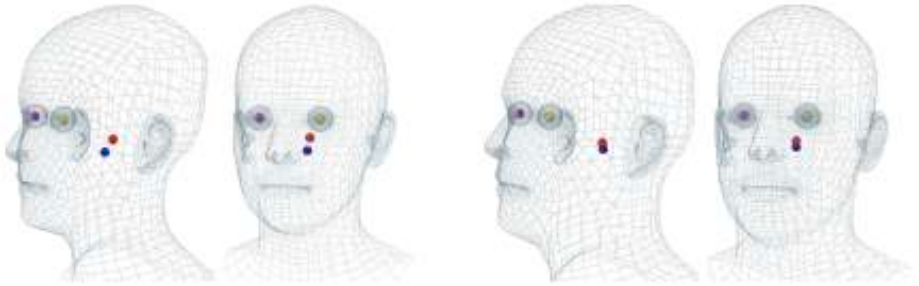


Рисунок 3 – Изображение подвижных сочленений (слева – модель женского пола, справа – модель мужского пола), изображение взято из [29]

Модель FLAME описывается функцией $M(\vec{\beta}, \vec{\theta}, \vec{\psi}): R^{|\vec{\beta}|\times|\vec{\theta}|\times|\vec{\psi}|} \to R^{3N}$, которая принимает на вход коэффициенты, описывающие форму $\vec{\beta} \in R^{|\vec{\beta}|}$, позу $\vec{\theta} \in R^{|\vec{\theta}|}$, выражение лица $\vec{\psi} \in R^{\vec{\psi}}$, и возвращает N вершин. Каждый вектор позы $\vec{\theta} \in R^{3K+3}$ содержит $K + 1$ вектор вращения ($\in R^3$), то есть вектора вращения в трехмерном пространстве для каждого сустава и один глобальный вектор вращения. Важно отметить, что здесь и далее при описании параметрической модели головы FLAME обозначения и формулы взяты из оригинального источника [45].

Модель состоит из шаблонной полигональной сетки $\overline{\boldsymbol{T}} \in R^{3N}$ в нулевой позе (нейтральная поза покоя) $\overrightarrow{\theta^*}$, функции формы смешивания для корректировки положения вершин полигональной сетки в соответствии с параметрами, описывающими идентичность $B_S(\vec{\beta}; S): R^{|\vec{\beta}|} \to R^{3N}$, функции формы смешивания для корректировки положения вершин полигональной сетки в соответствии с параметрами, описывающими позу головы $B_P(\vec{\theta}; P): R^{|\vec{\theta}|} \to R^{3N}$ (необходимо для получения реалистичных деформаций и положений относительно суставов, которые не удается достигнуть лишь с использованием техники LBS), и функции формы смешивания для корректировки положения вершин полигональной сетки в

соответствии с параметрами, описывающими выражение лица $B_E(\vec{\psi}; E): R^{|\vec{\psi}|} \rightarrow R^{3N}$. Стандартная функция LBS $W(\overline{\boldsymbol{T}}, \boldsymbol{J}, \vec{\theta}, \mathcal{W})$ применяется для поворота вершин полигональной сетки $\overline{\boldsymbol{T}}$ вокруг подвижных сочленений $\boldsymbol{J}$, выполняя линейное сглаживание весами смешивания (англ. blendweights) $\mathcal{W} \in R^{K \times N}$.

Собирая всё вышеизложенное, модель можно определить так:

$$M(\vec{\beta}, \vec{\theta}, \vec{\psi}) = W(T_P(\vec{\beta}, \vec{\theta}, \vec{\psi}), \boldsymbol{J}(\vec{\beta}; J, \overline{T}, S), \vec{\theta}, \mathcal{W}),$$

где $T_P(\vec{\beta}, \vec{\theta}, \vec{\psi}) = \overline{\boldsymbol{T}} + B_S(\vec{\beta}; S) + B_P(\vec{\theta}; P) + B_E(\vec{\psi}; E)$ – шаблонная полигональная сетка ($\overline{\boldsymbol{T}}$) с добавлением смещений в соответствии с формой, позой и выражением лица,

$\boldsymbol{J}(\vec{\beta}; J, \overline{\boldsymbol{T}}, S)$ – функция отображения положения подвижных сочленений для уникального субъекта. Была введена, так как для каждого человека положение подвижных сочленений уникально. $\boldsymbol{J}(\vec{\beta}; J, \overline{T}, S) = J(\overline{\boldsymbol{T}} + B_S(\vec{\beta}; S))$, здесь J – это разреженная матрица для вычисления положения суставов по вершинам полигональной сетки (матрица обучается на этапе построения модели и фиксируется).

$B_S(\vec{\beta}; S) = \sum_{n=1}^{|\vec{\beta}|} \beta_n \boldsymbol{S_n}$ – функция формы смешивания для корректировки положения вершин полигональной сетки в соответствии с параметрами, описывающими идентичность. Здесь $\vec{\beta} = \left[\beta_1, \dots, \beta_{|\vec{\beta}|}\right]^T$ – коэффициенты разложения в базисе формы (идентичности), $S = [\boldsymbol{S_1}, \dots, \boldsymbol{S}_{|\vec{\beta}|}] \in R^{3N \times |\vec{\beta}|}$ – ортонормированный базис формы, который получен в результате применения метода главных компонент (англ. principal component analysis, PCA). Возвращает смещение относительно вершин шаблонной полигональной сетки $\overline{\boldsymbol{T}}$.

$B_P(\vec{\theta}; P) = \sum_{n=1}^{9K} \left(R_n(\vec{\theta}) - R_n(\overrightarrow{\theta^*})\right) \boldsymbol{P_n}$ – функция формы смешивания для корректировки положения вершин полигональной сетки в соответствии с параметрами, описывающими позу головы. Здесь $R(\vec{\theta}): R^{|\vec{\theta}|} \rightarrow R^{9K}$ – функция, отображающая вектор позы $\vec{\theta}$ в вектор, содержащий конкатенированные элементы

всех соответствующих матриц вращения. Так, $R_n(\vec{\theta})$ и $R_n(\overrightarrow{\theta^*})$ – это n-ый элемент $R(\vec{\theta})$ и $R(\overrightarrow{\theta^*})$ соответственно. Вектор $P_n \in R^{3N}$ описывает смещение вершин из позы покоя вызванное R_n, так $P = [\boldsymbol{P_1}, \dots, \boldsymbol{P_{9K}}] \in R^{3N \times 9K}$ – это матрица, содержащая все формы смешивания для позы. Возвращает смещение относительно вершин шаблонной полигональной сетки $\overline{\boldsymbol{T}}$.

$B_E(\vec{\psi}; E) = \sum_{n=1}^{|\vec{\psi}|} \psi_n \boldsymbol{E_n}$ – функция формы смешивания для корректировки положения вершин полигональной сетки в соответствии с параметрами, описывающими выражение лица. Здесь $\vec{\psi} = \left[\psi_1, \dots, \psi_{|\vec{\psi}|}\right]^T$ - коэффициенты разложения в базисе выражения лица, $E = [\boldsymbol{E_1}, \dots, \boldsymbol{E}_{|\vec{\boldsymbol{\psi}}|}] \in R^{3N \times |\vec{\psi}|}$ - ортонормированный базис выражения лица, который получен в результате применения метода главных компонент. Возвращает смещение относительно вершин шаблонной полигональной сетки $\overline{\boldsymbol{T}}$.

Модель обучается в три этапа для декомпозиции параметров позы, выражения лица и формы путем итеративной оптимизации, в ходе которой минимизируется ошибка реконструкции на обучающих данных.

На первом этапе производится оптимизация параметров позы, которые разделяются на персональные (шаблонная полигональная сетка в состоянии покоя и положение подвижных сочленений) и общие (веса смешивания $\mathcal{W}$, формы смешивания для позы P, матрица для вычисления положения подвижных сочленений J). Для того, чтобы избежать сильного влияния выражения лица на оптимизируемые параметры, оно сбрасывается в нейтральное.

На втором этапе производится оптимизация параметров выражения лица, к которым относится E (ортонормированный базис выражения лица). Для этого требуется, чтобы выражение лица было отделено от вариации поз и форм. Поза для всех сканирований преобразуется в позу состояния покоя с использованием параметров, полученных на первом этапе оптимизации. Влияние формы устраняется путем вычитания вершин полигональной сетки с нейтральным

выражением лица. Затем производится вычисление пространства выражений лица E с помощью применения PCA к предобработанным данным.

На третьем этапе производится оптимизация шаблонной полигональной сетки $\overline{\boldsymbol{T}}$ и параметров формы S. По аналогии с предыдущими этапами производится удаление влияния позы и выражения лица. Шаблон $\overline{\boldsymbol{T}}$ вычисляется как среднее нормализованных относительно позы и выражений лица регистраций, S формируется как $|\vec{\beta}|$ главных компонент, вычисленных с использованием PCA.

Ниже приведена демонстрация результатов варьируемости параметров модели FLAME. В случае с параметрами формы и выражения лица допустимые значения ограничены диапазоном $[-2, 2]$. Если же передать значение какой-либо компоненты вне этого диапазона, то выходной результат будет неестественным. Параметры позы задаются в радианах, при этом они ограничены в реализации модели FLAME для получения реалистичных положений, поэтому входные значения в радианах перед применением техники LBS и вычислением значений функции смешивания позы сначала масштабируются.

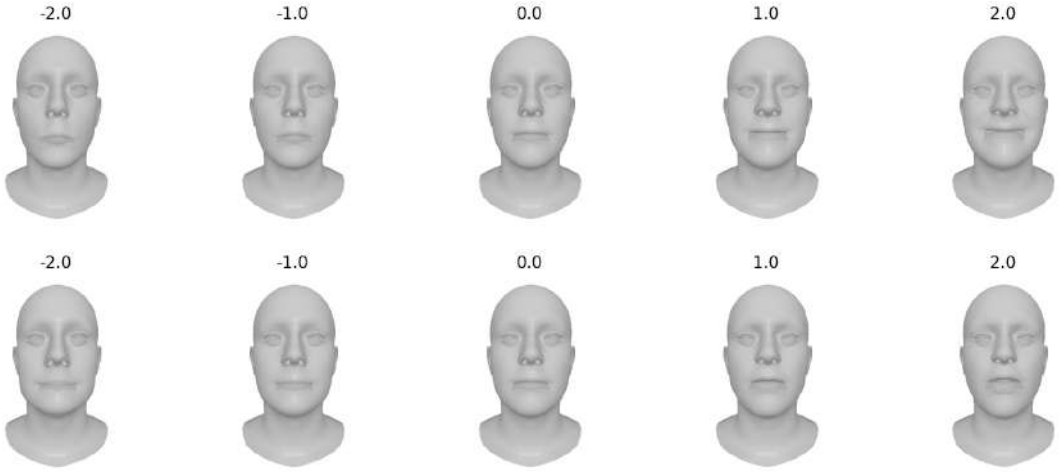


Рисунок 4 – Результат при варьируемости первой (первая строка) и второй (вторая строка) PCA компонент для выражения лица

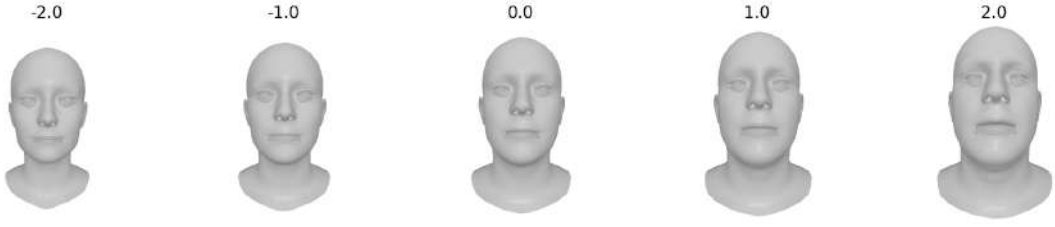


Рисунок 5 – Результат при варьируемости первой (первая строка) и второй (вторая строка) PCA компонент для формы

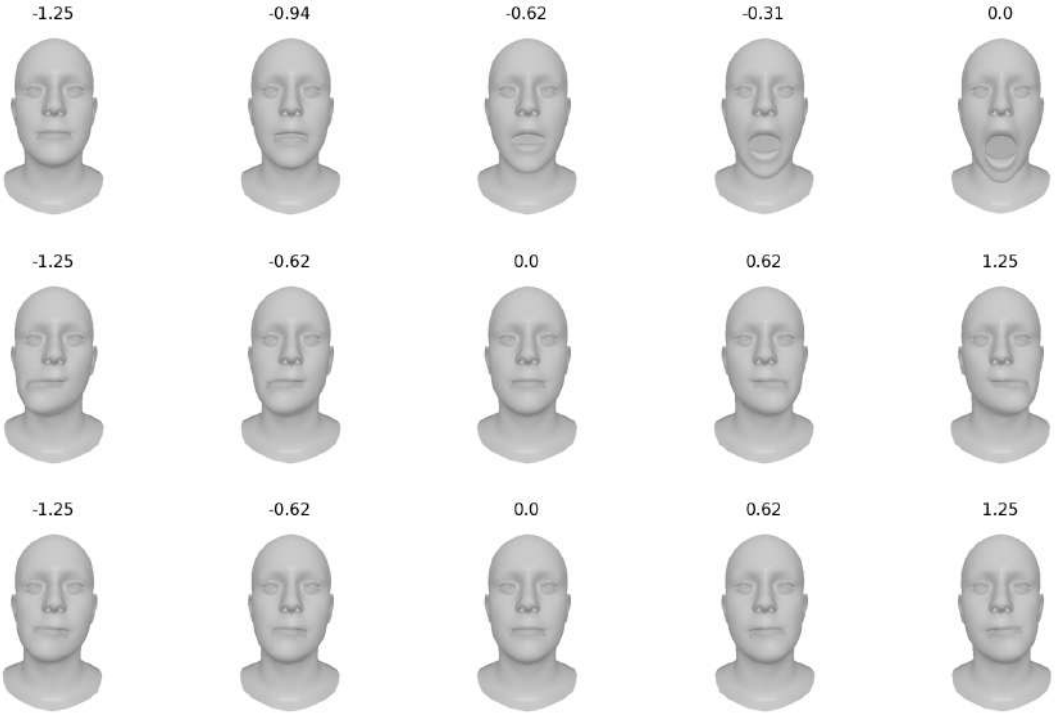


Рисунок 6 – Результат варьируемости позы челюсти по оси x (первая строка), по оси y (вторая строка) и по оси z (третья строка)

Рисунок 7 – Результат варьируемости позы шеи по оси x (первая строка), по оси y (вторая строка) и по оси z (третья строка)

Модель FLAME не включает в себя текстурную составляющую в отличие от популярной модели BFM, однако в 2020 году на платформе GitHub был представлен проект [66], в котором на основе текстурного пространства BFM и набора данных FFHQ было сформировано текстурное пространство модели FLAME с использованием алгоритма PCA.

На рисунке 8 представлена демонстрация результатов варьируемости параметров текстуры для первых двух компонент. Аналогично параметрам формы и выражения лица допустимые значения ограничены диапазоном $[-2, 2]$.

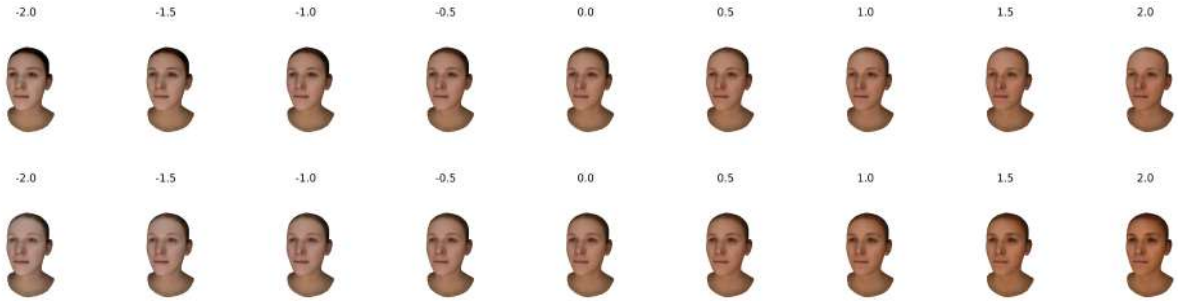


Рисунок 8 – Результат варьируемости первой (первая строка) и второй (вторая строка) PCA компонент для текстуры

2.3. Алгоритм оценки параметров модели FLAME с использованием RGB изображения

Данный подраздел посвящен описанию классического алгоритма оценки параметров параметрической модели FLAME по одному RGB изображению. Оценка параметров производится с использованием итеративного алгоритма градиентного спуска. В качестве алгоритма оптимизации, как правило, выбирают

Adam [67]. Он используется и в рамках данного диссертационного исследования для этапа оценки параметров модели.

Для входного изображения вычисляются ключевые точки лица и бинарная маска сегментации (включая волосы), которые затем фиксируются и принимаются за истинную разметку. Вследствие чего итоговый результат ограничен и напрямую зависит от качества работы используемых сторонних алгоритмов. В рамках диссертационного исследования, как и в актуальных работах, посвященных созданию цифровых аватаров с использованием параметрической модели, для вычисления ключевых точек применяется алгоритм [68], для вычисления масок сегментации – [69]. Параметры параметрической модели (вектор коэффициентов формы, вектор коэффициентов выражения лица и параметры позы) инициализируются значениями, соответствующими нейтральному положению, то же самое проделывается с текстурными коэффициентами, параметрами освещения и параметрами камеры как внутренними (фокусное расстояние, оптический центр (англ. principal point)), так и внешними (матрица поворота R и вектор сдвига t). Затем параметры параметрической модели и текстурные компоненты интегрируются в полигональную сетку, которая вместе с параметрами камеры и освещения передается в процедуру рендеринга. Итоговое изображение участвует в процессе оптимизации. В модели FLAME известны 3D положения ключевых точек лица (закреплены за вершинами в полигональной сетке), которые проецируются на полученное изображение, и также используются для оптимизации. Так, цель процедуры оптимизации – минимизировать функцию потерь, которая вычисляется по формуле (1).

$$Loss_{total} = w_1 \cdot \frac{|img_{predict} - img_{GT}| \odot mask_{GT}}{\sum_{i=0}^{N-1}\sum_{j=0}^{M-1} mask_{GT_{i,j}}} + w_2 \cdot$$

$$\frac{1}{2 \cdot L}\sum_{n=0}^{L-1}\sum_{m=0}^{1}\left|lmk_{2D\,predict_{n,m}} - lmk_{2D\,GT_{n,m}}\right| + w_3 \cdot \sum_{k=0}^{|\vec{\beta}|-1}\left(\frac{\beta_k}{\sigma_{shape}}\right)^2 + w_4 \cdot$$

$$\sum_{p=0}^{|\vec{\psi}|-1}\left(\frac{\psi_p}{\sigma_{expr}}\right)^2 + w_5 \cdot \sum_{q=0}^{|\vec{\tau}|-1}\left(\frac{\tau_q}{\sigma_{tex}}\right)^2 + w_6 \cdot \sum_{k=0}^{9K-1}\left(R_k(\vec{\theta}) - R_k(\vec{\theta^*})\right)^2 + w_7 \cdot$$

$$\sum_{l=0}^{2}\left(\theta_{eye_l} - \theta_{eye_{l+3}}\right)^2 \rightarrow \min_{\vec{\theta},\vec{\beta},\vec{\psi},\vec{\tau},\vec{\lambda},\overrightarrow{cam_{int}},\overrightarrow{cam_{ext}}}, \quad (1)$$

где $img_{predict}$ – изображение, полученное после процедуры рендеринга с использованием текстуры,

img_{GT} – исходное изображение,

$mask_{GT}$ – бинарная маска для исходного изображения,

$lmk_{2D_{predict}}$ – 2D координаты ключевых точек, спроецированных на отрендеренное изображение,

$lmk_{2D_{GT}}$ – истинные 2D координаты ключевых точек,

$\vec{\beta}$ – вектор коэффициентов формы,

$\vec{\psi}$ – вектор коэффициентов выражения лица,

$\vec{\tau}$ – вектор текстурных коэффициентов,

$\vec{\theta}$ – вектор параметров позы (θ_{eye}– параметры позы для глаз; $\overrightarrow{\theta^*}$ - параметры позы покоя),

$\vec{\lambda}$ – вектор параметров освещения,

$\overrightarrow{cam}_{int}$ – вектор внутренних параметров камеры,

$\overrightarrow{cam}_{ext}$ – вектор внешних параметров камеры,

σ_{shape}, σ_{expr}, σ_{tex} - значение дисперсии для параметров векторов формы, выражения лица и текстуры,

K – количество суставов (в модели FLAME K=4, где каждому суставу соответствует вектор вращения из R^3),

$\mathrm{R}(\vec{\theta})$ – функция, отображающая вектор позы $\vec{\theta}$ в вектор, содержащий конкатенированные элементы всех соответствующих матриц вращения. Так, $\mathrm{R_k}(\vec{\theta})$ — это k-ый элемент $\mathrm{R}(\vec{\theta})$,

$w_1, w_2, w_3, w_4, w_5, w_6, w_7$ – взвешивающие коэффициенты для каждого слагаемого.

Компоненты в (1) с коэффициентами w_3, w_4, w_5, w_6 – это члены регуляризации, которые штрафуют оптимизируемые вектора за отклонения от диапазона возможных значений. Компонента с коэффициентом w_7 штрафует за несимметричное положение глаз.

При решении задачи создания цифрового аватара на вход алгоритма часто поступает RGB видеопоследовательность, содержащая мимику одного человека, с фиксированным положением камеры. Так, встает вопрос о целесообразности покадровой оптимизации всех упомянутых ранее параметров. Например, внутренние параметры камеры, коэффициенты формы и текстуры, параметры освещения в контексте захватываемой сцены, как правило, не меняются (исключение: естественные источники освещения в комбинации с продолжительной записью оказывают влияние на параметры освещения). Работа [70] является значимой среди тех работ, в которых предлагается дополнить алгоритм оценки параметров параметрической модели по одному RGB кадру для видеопоследовательностей. Авторы [47] адаптировали предложенное решение для параметрической модели FLAME. Так, для видеопоследовательности выбираются ключевые кадры (не более $1/4$ от всей видеопоследовательности), которые характеризуют крайние положения ($[-90°, 0°, 90°]$), на них производится классическая процедура оптимизации всех параметров. Для остальных кадров фиксируются внутренние параметры камеры, коэффициенты формы и текстуры, параметры освещения. Затем происходит оптимизация внешних параметров камеры, коэффициентов выражения лица и параметров позы. Количество итераций значительно меньше, чем при выполнении оптимизации для ключевых кадров.

В работе [47] в функцию потерь вводится компонента, которая штрафует за некорректное расстояние между верхней и нижней границей глаза. Однако, границы глаза также определяются сторонним алгоритмом для вычисления ключевых точек, точность которого зависит в том числе от условий съемки. В методах, где создается персональный аватар по входной видеопоследовательности, такой подход может обеспечить неплохую воспроизводимость эффекта закрытия глаз (для кадров, в которых распознавание границ глаз было выполнено успешно и довольно точно). В методах, которые нацелены на получение общей модели, введение подобной компоненты может внести негативный эффект, так как обучение, как правило, производится на крупномасштабных наборах данных, что затрудняет последующую фильтрацию.

Полученную в ходе оценки параметров модель головы можно анимировать путем замены вектора коэффициентов выражения лица и/или вектора позы на требуемые. Вектора позы и выражения лица можно получить также с помощью оценки параметров, либо сгенерировать в допустимом диапазоне.

2.4. Алгоритм оценки параметров модели FLAME с использованием RGBD изображения

К недостаткам алгоритма оценки параметров параметрической модели FLAME, описанного в подразделе 2.3, можно отнести отсутствие при оптимизации информации о 3D координатах ключевых точек, что в результате приводит к недостаточной схожести формы головы. Одним из решений данной проблемы является вычисление 3D координат ключевых точек с помощью стороннего нейросетевого алгоритма [71] и дополнение функции потерь компонентой, измеряющей отклонение истинных 3D координат ключевых точек (вычисленных сторонним алгоритмом и принятых за истинную разметку) от 3D координат ключевых точек полигональной сетки. Однако, как и с 2D ключевыми точками, надежность разметки ограничена качеством используемого алгоритма вычисления ключевых точек. В связи с чем в работе [22*] был предложен алгоритм оценки параметров модели FLAME с использованием RGBD изображения (все экспериментальные снимки были получены с использованием стереокамеры ZED2). Преимуществом такого подхода является наличие действительно истинной 3D разметки (ограничения: погрешность измерения и надежность вычисления 2D ключевых точек). На рисунке 9 схематично представлен предложенный алгоритм оценки параметров и анимации параметрической модели FLAME по RGBD изображению. Анимация производится так же, как и для метода, описанного в предыдущем подразделе – выполняется замена вектора коэффициентов выражения лица и/или вектора позы на требуемые.

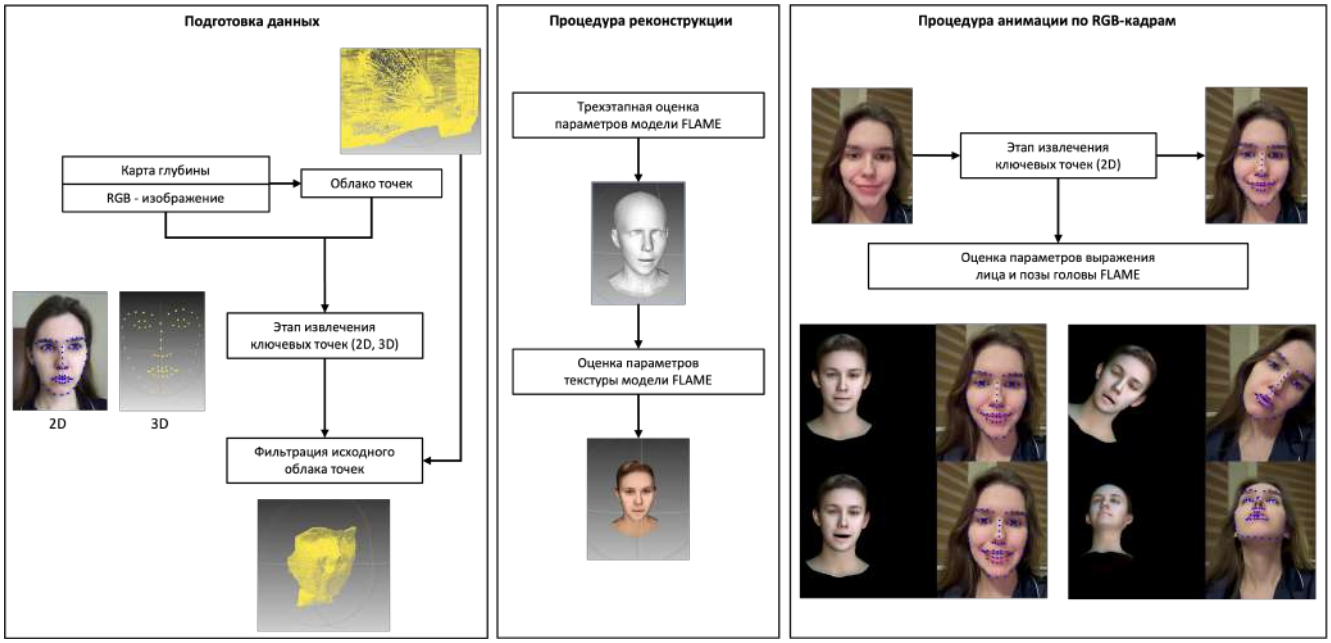


Рисунок 9 – Схематичное представление разработанного алгоритма оценки параметров параметрической модели FLAME с использованием RGBD изображения

На вход алгоритма поступают RGB изображение и карта глубины, захваченные со стереокамеры. Первый этап – предобработка входных данных. С использованием RGB изображения вычисляются координаты 2D ключевых точек, которые сопоставляются с картой глубины для получения 3D координат. Облако точек, сформированное из RGB изображения и карты глубины, фильтруется на основе полученных 3D координат точек лица (фильтрация заключается в отсечении точек, не входящих в диапазон значений, сформированный из 3D ключевых точек). В результате после этапа предобработки на выходе формируется набор – 2D ключевые точки лица, 3D ключевые точки лица, отфильтрованное облако точек, RGB изображение.

На втором этапе производится оценка параметров модели FLAME, параметров текстуры и камеры на основе данных, полученных на первом этапе. Процедура оптимизации итеративная, алгоритм оптимизации – Adam. Сначала производится оценка параметров модели и камеры. Процедура разбита на три блока. Первый блок отвечает за оптимизацию параметров позы $\vec{\theta}$ и камеры путем минимизации функции потерь, которая вычисляется по формуле (2). Второй блок

отвечает за оптимизацию параметров формы $\vec{\beta}$ и выражения лица $\vec{\psi}$, минимизируя функцию потерь, вычисляемую по формуле (3). И, наконец, третий блок отвечает за оптимизацию параметров формы $\vec{\beta}$, минимизируя функцию потерь, вычисляемую по формуле (4). Такое разделение на блоки позволяет сначала выполнить правильное позиционирование и лишь затем уточнять детали модели головы.

$$Loss_1 = \sqrt{\frac{1}{3 \cdot N} \sum_{i=0}^{N-1} \sum_{j=0}^{2} \left(lmk_{3D_{GT_{i,j}}} - lmk_{3D_{predict_{i,j}}} \right)^2} \rightarrow \min_{\vec{\theta}, \overrightarrow{cam}_{int}, \overrightarrow{cam}_{ext}}, \quad (2)$$

где $lmk_{3D_{GT}}$ – истинные 3D координаты ключевых точек,

$lmk_{3D_{predict}}$ – 3D координаты ключевых точек полигональной сетки модели FLAME,

$\vec{\theta}$ – вектор параметров позы,

$\overrightarrow{cam}_{int}$ – вектор внутренних параметров камеры,

$\overrightarrow{cam}_{ext}$ – вектор внешних параметров камеры.

$$Loss_2 = w_1 \cdot \sqrt{\frac{1}{3 \cdot N} \sum_{i=0}^{N-1} \sum_{j=0}^{2} \left(lmk_{3D_{GT_{i,j}}} - lmk_{3D_{predict_{i,j}}} \right)^2} + w_2 \cdot$$

$$\frac{1}{|PCD_{GT}|} \sum_{p_1 \in PCD_{GT}} \min_{p_2 \in M\left(\vec{\beta}, \vec{\theta}, \vec{\psi}\right)} \|p_1 - p_2\|_2^2 \rightarrow \min_{\vec{\beta}, \vec{\psi}, \vec{\theta}, \overrightarrow{cam}_{int}, \overrightarrow{cam}_{ext}}, \quad (3)$$

где PCD_{GT} – отфильтрованное облако точек, которое представлено неупорядоченным набором 3D точек,

$|\cdot|$ – мощность множества,

$M\left(\vec{\beta}, \vec{\theta}, \vec{\psi}\right)$ – полигональная сетка параметрической модели FLAME,

w_1, w_2 – взвешивающие коэффициенты для каждого слагаемого.

$$Loss_3 = \frac{1}{|PCD_{GT}|} \sum_{p_1 \in PCD_{GT}} \min_{p_2 \in M\left(\vec{\beta}, \vec{\theta}, \vec{\psi}\right)} \|p_1 - p_2\|_2^2 \rightarrow \min_{\vec{\beta}}, \quad (4)$$

Затем производится оптимизация параметров текстуры $\vec{\tau}$ и освещения $\vec{\lambda}$ путем минимизации функции потерь, которая вычисляется по формуле (5).

$$Loss_4 = \frac{|img_{predict} - img_{GT}| \odot mask_{GT}}{\sum_{i=0}^{N-1}\sum_{j=0}^{M-1} mask_{GT_{i,j}}} \to \min_{\vec{\tau},\vec{\lambda}}, \qquad (5)$$

Для набора данных, участвующего в экспериментальном исследовании алгоритма оценки параметров параметрической модели по RGBD изображению, было получено среднее значение функции потерь $Loss_3$ равное $4 \cdot 10^{-8}$. Так как функция потерь $Loss_3$ соответствует последнему этапу трехэтапной процедуры оценки параметров параметрической модели FLAME, то ее значение можно рассматривать как меру качества. Ее физический смысл состоит в измерении среднего значения квадрата расстояния между каждой точкой из исходного облака точек, полученного с использованием стереокамеры, и ближайшей к ней гранью полигональной сетки, полученной в ходе оптимизации. Так как захваченные облака точек представлены в реальном масштабе, а их значения измеряются в метрах, то $\sqrt{Loss_3} = 0{,}0002$ м. Далее будем называть эту величину точностью трехмерной реконструкции.

На рисунке 10 представлены слева-направо: результат работы алгоритма, описанного в подразделе 2.3, результат работы метода DECA, который описан в разделе 1, и результат работы предложенного алгоритма оценки параметров по RGBD кадру. Для оценки параметров методом DECA использовалось RGB изображение и 2D ключевые точки. Метод DECA включен в рассмотрение, так как является развитием модели FLAME и используется в некоторых подходах для оценки параметров параметрической модели [49], [57]. В качестве данных, которые могут быть использованы как эталонные, могут выступать только захваченные с помощью стереокамеры облака точек. Так как полигональные сетки, полученные в ходе оценки параметров с помощью других подходов, участвующих в сравнении, не соответствуют реальному масштабу, то количественной оценке подлежат лишь результаты предложенного алгоритма. В связи с тем, что точность трехмерной реконструкции для предложенного алгоритма принимает настолько низкое значение, можно сделать вывод, что качество реконструкции полигональной сетки ограничено только точностью захваченного облака точек.

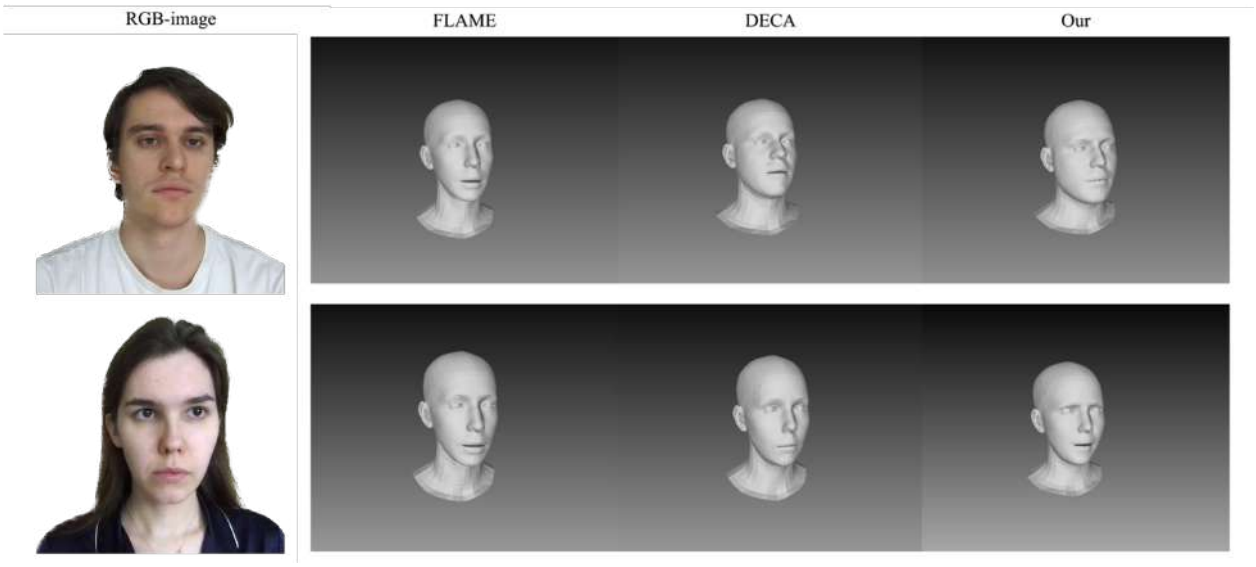


Рисунок 10 – Сравнение результатов реконструкции

Исходя из визуального анализа результатов, можно заключить, что оптимизация по RGBD данным позволяет получить результат наиболее близкий к истинной форме лица. Результаты, полученные с помощью метода DECA, чуть точнее передают черты лица, в отличие от оценки параметров модели FLAME по 2D данным.

2.5. Выводы и результаты второго раздела

В данном разделе был представлен подробный обзор параметрической модели головы FLAME и описан классический алгоритм оценки ее параметров, который применяется во многих современных работах на этапе предобработки. Также был описан предложенный алгоритм оценки параметров параметрической модели по RGBD изображению и представлены качественные результаты сравнения с часто используемыми на этапе предобработки подходами.

По результатам можно заключить, что предложенный алгоритм оценки параметров модели по RGBD снимку дает более качественные результаты, которые близки к истинной форме головы, в отличие от метода DECA и классического алгоритма оценки параметров модели FLAME по RGB снимку. При этом предложенный алгоритм достигает высокой точности трехмерной реконструкции

равной 0,0002 м. Однако применение такого подхода ограничено наличием стереокамеры, которая способна формировать облака точек с высокой точностью. Такое ограничение является критически важным, так как пограничные точки объектов из облака точек при съемке на стереокамеру часто попадают в область фона, что в худшем случае может привести к тому, что 2D ключевым точкам овала лица будут соответствовать точки фона.

Таким образом, если имеется качественная стереокамера и достаточное количество вычислительных ресурсов, то в качестве этапа оценки параметров параметрической модели следует использовать предложенный алгоритм, в противном случае – использовать алгоритм оценки параметров модели FLAME по RGB изображению.

По теме раздела опубликованы работы [21*], [22*].

РАЗДЕЛ 3. ПАРАМЕТРИЧЕСКАЯ МОДЕЛЬ ГОЛОВЫ ЧЕЛОВЕКА НА ОСНОВЕ НЕЙРОСЕТЕВОЙ МОДЕЛИ ПРЕДСТАВЛЕНИЯ ПОВЕРХНОСТИ CNeRF И ДВУМЕРНОГО НЕЙРОННОГО РЕНДЕРИНГА

Данный раздел диссертационного исследования посвящен описанию разработанной параметрической модели головы человека на основе нейросетевой модели представления поверхности CNeRF и двумерного нейронного рендеринга. Такая модель реализует неявное представление трехмерной поверхности, а также предоставляет возможность контроля различных параметров этого представления в зависимости от некоторых условий. Двумерный нейронный рендеринг, в отличие от объёмометрического рендеринга, обычно используемого в нейросетевой модели NeRF, позволяет выполнять синтез изображений-проекций в реальном времени. Отличительной особенностью предлагаемой параметрической модели головы является использование в процессе обучения синтетического набора данных, генерируемого в реальном времени.

3.1. Описание разработанной параметрической модели головы человека

В разделе 1 было приведено описание самых значимых и актуальных работ по созданию цифрового аватара головы человека. Было выявлено, что актуальным способом представления аватара является нейронное неявное представление. Существующие решения имеют ряд недостатков. Во-первых, методы, использующие в качестве способа нейронного неявного представления SDF [47], [49], [57], как правило, требуют больших вычислительных мощностей. В таких методах для каждого человека требуется обучить свой набор весов с нуля. Процесс обучения при этом требует использования актуальных специализированных вычислительных устройств и может длиться десятки часов. Во-вторых, методы, использующие в качестве способа неявного представления нейросетевую модель NeRF или Occupancy Fields [46], [51], либо ограничены производительностью, либо формируют представления с недостаточной схожестью [46], [48]. Причиной первого факта является ресурсоемкая процедура объемометрического рендеринга.

Причиной второго – отсутствие специализированного набора данных, охватывающего вариации идентичности, позы головы и выражения лица (в контексте 2D данных, так как NeRF-подобные модели обучаются на наборе изображений).

Параметрическая модель головы часто используется в качестве промежуточной компоненты в методах создания аватара головы человека, как было описано в разделе 2.1. В связи с этим была поставлена задача разработать параметрическую модель головы человека, которая для способа представления поверхности использует нейросетевую модель CNeRF, а для её отрисовки – алгоритм двумерного нейронного рендеринга. Как описано в разделе 2.1, в рамках диссертационного исследования за основу была взята параметрическая модель FLAME. Так, разработанная параметрическая модель способна синтезировать изображения, соответствующие изображениям, синтезированным при помощи модели FLAME., используя те же пространства параметров, но используя при этом неявные представления для рендеринга. Такая модель может быть полезна в качестве одного из этапов метода создания цифрового аватара головы человека, поскольку может использовать те же процедуры оценки параметров что и оригинальная модель FLAME, при этом являясь богатым источником априорной информации в случае применения методик переноса обучения.

Схематичное представление архитектуры предложенной параметрической модели приведено на рисунке 11.

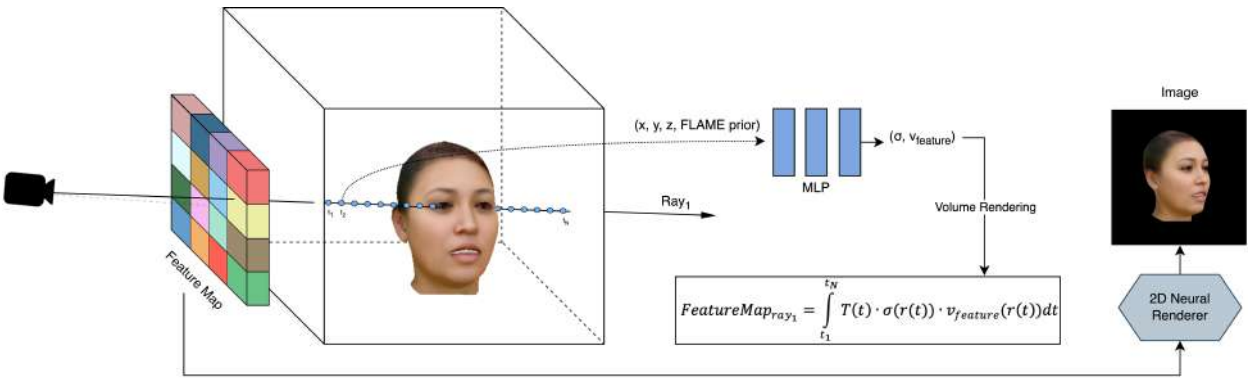


Рисунок 11 – Схематичное представление предложенной архитектуры параметрической модели

Архитектура состоит из нейронных полей излучения, обусловленных параметрами параметрической модели головы FLAME, и нейронной сети, которая выполняет двумерный нейронный рендеринг. Условные нейронные поля излучения создают низкоразмерную карту признаков с использованием алгоритма объемометрического рендеринга. Нейронная сеть, выполняющая двумерный нейронный рендеринг, из полученной карты признаков формирует итоговое изображение. Такой подход в отличие от ранних методов, основанных на нейронной модели NeRF, позволяет ускорить процедуру рендеринга, так как карта признаков имеет небольшое разрешение в отличие от итогового изображения. За счет обуславливания многослойного перцептрона, в модели NeRF становится возможным контроль и варьирование параметров идентичности, выражения лица и позы головы, что отсутствует в базовой версии NeRF, предназначенной для представления статической сцены. В подразделе 3.1.1 описан принцип работы нейросетевой модели представления поверхности CNeRF, используемой в рамках разработанной параметрической модели головы.

3.1.1. Условные нейронные поля излучения

NeRF – это один из способов описания поверхности, как было описано в разделе 1. В базовой версии нейросетевой модели NeRF, представленной схематично на рисунке 12, для отрисовки сцены с новой точки обзора на основе внутренних и внешних параметров камеры производится семплирование лучей и точек на лучах. Фактически, это процесс растеризации, где каждому пикселю итогового изображения-проекции ставится в соответствие набор точек в пространстве, по которому будет вычисляться итоговый цвет. Каждая точка на луче описывается пространственными координатами и углами направления, которые кодируются предзаданными функциями [25], и поступает на вход многослойного перцептрона, где:

1. Закодированные пространственные координаты проходят через пять полносвязных слоёв, в результате чего формируется промежуточный

вектор признаков.

2. Промежуточный вектор признаков конкатенируется с исходными входными данными (закодированные пространственные координаты), реализуя таким образом skip connection.
3. Полученный вектор проходит еще четыре полносвязных слоя, в результате формируется вектор признаков, который:
   3.1. Проходит через один полносвязный слой для предсказания значения объемной плотности (σ).
   3.2. Конкатенируется с закодированными углами направления и проходит через еще два полносвязных слоя для предсказания значения яркости излучения в точке ((r, g, b) цвет).

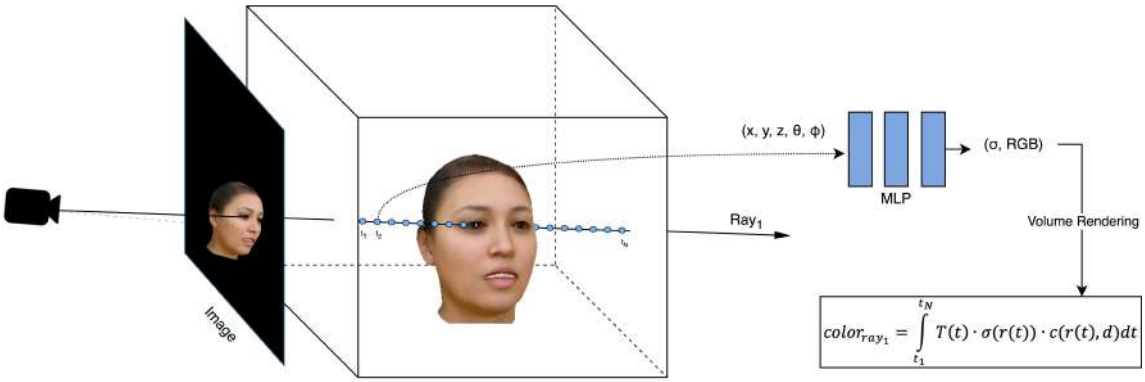


Рисунок 12 – Схематичное представление базовой версии нейросетевой модели NeRF

После получения набора значений $(\sigma, (r, g, b))$ для просемплированных точек каждого луча выполняется процедура объемометрического рендеринга с целью расчета цвета пикселей итогового изображения-проекции. Вычисление итогового цвета каждого пикселя производится по формуле (6).

$$color_{ray} = \int_{t_1}^{t_N} T(t) \cdot \sigma\big(\boldsymbol{r}(t)\big) \cdot C(\boldsymbol{r}(t), d)\, dt, \qquad (6)$$

где $T(t) = \exp\left(-\int_{t_1}^{t} \sigma(\boldsymbol{r}(s))ds\right)$ – накопленная пропускаемость вдоль луча из точки t_1 в точку t (чем выше накопленная объемная плотность, тем ниже накопленная пропускаемость),

$\boldsymbol{r}(t)$ – точка t на луче $\boldsymbol{r}$, испускаемом из центра камеры $\boldsymbol{o}$ с направлением $\boldsymbol{d}$, которая описывается как $\boldsymbol{r}(t) = \boldsymbol{o} + t \cdot \boldsymbol{d}$,

$\sigma\big(\boldsymbol{r}(t)\big)$ – объемная плотность, которая описывает вероятность занятости точки t на луче $\boldsymbol{r}$ в пространстве,

$C(\boldsymbol{r}(t), \boldsymbol{d})$ – цвет точки $\mathbf{r}(t)$ с направлением $\boldsymbol{d}$,

N – количество просемплированных точек на луче.

В случае использования алгоритма двумерного нейронного рендеринга семплирование лучей и точек на лучах производится по области низкоразмерной карты признаков. При этом внутренние параметры камеры масштабируются отношением разрешения выходной карты признаков и размера итогового изображения-проекции ($\frac{h_{feature\_map}}{h_{image}}, \frac{w_{feature\_map}}{w_{image}}$, где (h_{image}, w_{image}) – разрешение итогового изображения-проекции, $(h_{feature\_map}, w_{feature\_map})$ – разрешение итоговой карты признаков). В рамках предложенной модели на вход многослойного перцептрона поступают закодированные пространственные координаты (кодирование производится аналогично базовой нейросетевой модели NeRF), векторы коэффициентов формы, параметров позы (шея, челюсть, глазные яблоки), коэффициентов выражения лица, текстурных коэффициентов и параметров освещения, полученные в результате оценки параметров параметрической модели головы FLAME (либо сгенерированные в допустимом диапазоне, см. подраздел 3.2). Закодированные значения углов направления не поступают на вход многослойного перцептрона во избежание смещения, обусловленного набором данных [72]. На рисунке 13 схематично представлена архитектура многослойного перцептрона, которая используется нейросетевой моделью представления поверхности CNeRF в рамках разработанной модели головы, где:

1. Закодированные пространственные координаты конкатенируются с векторами коэффициентов формы, параметров позы и коэффициентов выражения лица.

2. Предсказывается значение объемной плотности σ аналогично базовой нейросетевой модели NeRF.
3. Вектор признаков, участвующий в предсказании σ, проходит через один полносвязный слой и конкатенируется с векторами параметров освещения и текстурных коэффициентов.
4. Полученный вектор признаков проходит через еще два полносвязных слоя, в результате предсказывая вектор признаков для точки на луче.

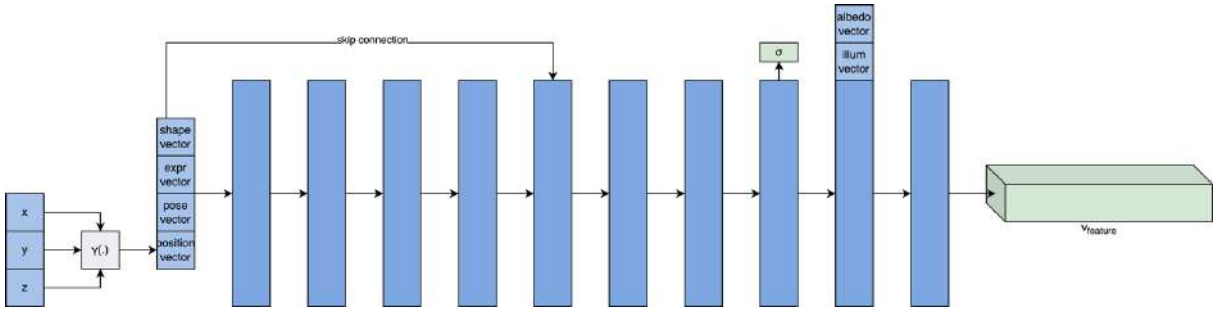


Рисунок 13 – Архитектура многослойного перцептрона, используемая в рамках предложенной модели

Мотивация добавления векторов параметров освещения и текстурных коэффициентов после получения σ аналогична базовой модели NeRF – для предсказания вероятности занятости точки в пространстве информация о текстуре и освещении является избыточной. После получения набора значений $(\sigma, v_{feature})$ для просемплированных точек каждого луча выполняется процедура объемометрического рендеринга с целью расчета векторов признаков итоговой карты признаков. Вычисление каждого вектора признаков итоговой карты признаков производится по формуле (7).

$$FeatureMap_{ray} = \int_{t_1}^{t_N} T(t) \cdot \sigma\big(\boldsymbol{r}(t)\big) \cdot v_{feature}\big(\boldsymbol{r}(t)\big) dt, \qquad (7)$$

где $v_{feature}\big(\boldsymbol{r}(t)\big)$ – вектор признаков, описывающий пространственную точку $\boldsymbol{r}(t)$.

Сформированная карта признаков поступает на вход алгоритма двумерного нейронного рендеринга для формирования итогового изображения-проекции. В

подразделе 3.1.2. представлено описание архитектуры нейронной сети, используемой в алгоритме.

3.1.2. Двумерный нейронный рендеринг

На рисунке 14 представлена архитектура нейронной сети, которая выполняет двумерный нейронный рендеринг и основана на идеях из работ [73], [72], [48]. Основная идея архитектуры заключается в том, что входная карта признаков, проходя через последовательный набор слоев, увеличивает разрешение и уменьшает количество каналов. Основные компоненты архитектуры – это двумерные свёрточные слои с размером ядра 1×1, которые агрегируют информацию по каналам, и функция активации LeakyReLU, поскольку их использование позволяет избежать возможное несоответствие с разных точек обзора [72].

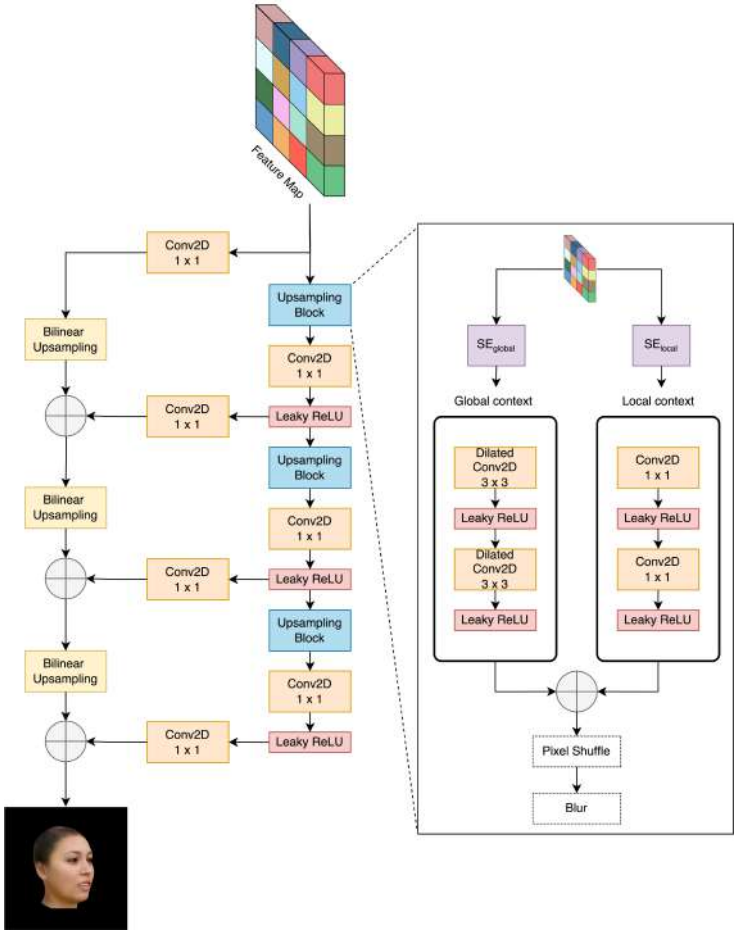


Рисунок 14 – Архитектура нейронной сети, которая выполняет двумерный нейронный рендеринг

Входная карта признаков обрабатывается двумя параллельными ветвями сети:

1. В первую ветвь с набором слоев билинейного повышения дискретизации (англ. bilinear upsampling) поступает обработанная свёрточным слоем с размером ядра 1×1 входная карта признаков с исходным разрешением и количеством каналов равным трем.
2. Во второй ветви исходная карта признаков проходит через набор слоев блока повышения дискретизации (англ. upsampling block), свёрточный слой с размером ядра 1×1 и функцию активации LeakyReLU. В результате формируется карта признаков с увеличенным в два раза разрешением.
3. Полученная на этапе (2) карта признаков проходит через сверточный слой с размером ядра 1×1 для уменьшения количества каналов до трех.
4. Значения карт признаков с этапов (1) и (3) суммируются. Такая последовательность действий повторяется еще два раза, в результате формируя итоговое RGB изображение-проекцию.

Основным отличием, которое позволило ускорить процесс сходимости и улучшить качество рендеринга в отличие от подхода, представленного в [48], является блок повышения дискретизации (увеличивает пространственное разрешение входной карты признаков и уменьшает количество каналов). Этот блок разделен на две части, которые выполняют задачи выделения локального и глобального контекстов, для этого:

1. Входная карта признаков проходит через два отдельных блока Squeeze-and-Excitation [74], которые присваивают вес каждому каналу входного тензора, позволяя тем самым модели сосредоточиться на наиболее значимых каналах карты признаков для решения поставленной задачи (под задачами в данном случае понимается выделение локального и глобального контекста). Полученные карты признаков проходят через последовательность свёрточных слоёв в связке с функцией активации LeakyReLU.

2. В первой части блока для выделения глобального контекста используются расширенные свёрточные слои (англ. dilated convolution layers) с размером ядра 3×3. За счёт большого рецептивного поля (англ. receptive fields) такие слои позволяют концентрироваться на глобальных характеристиках.
3. Во второй части блока для выделения локального контекста используются свёрточные слои с размером ядра 1×1, что позволяет фокусироваться на более мелких деталях.
4. Полученные из двух блоков карты признаков затем суммируются и поступают на вход необучаемого слоя Pixel Shuffle [75], который уменьшает глубину входной карты признаков и увеличивает разрешение за счет перестановок ее значений.
5. Полученная на (4) этапе карта признаков обрабатывается сглаживающим фильтром [76].

На рисунке 15 представлен пример карт признаков после свёрточных слоёв в блоке повышения дискретизации, демонстрирующий работу частей блока для выделения глобального и локального контекста и подтверждающий их назначение.

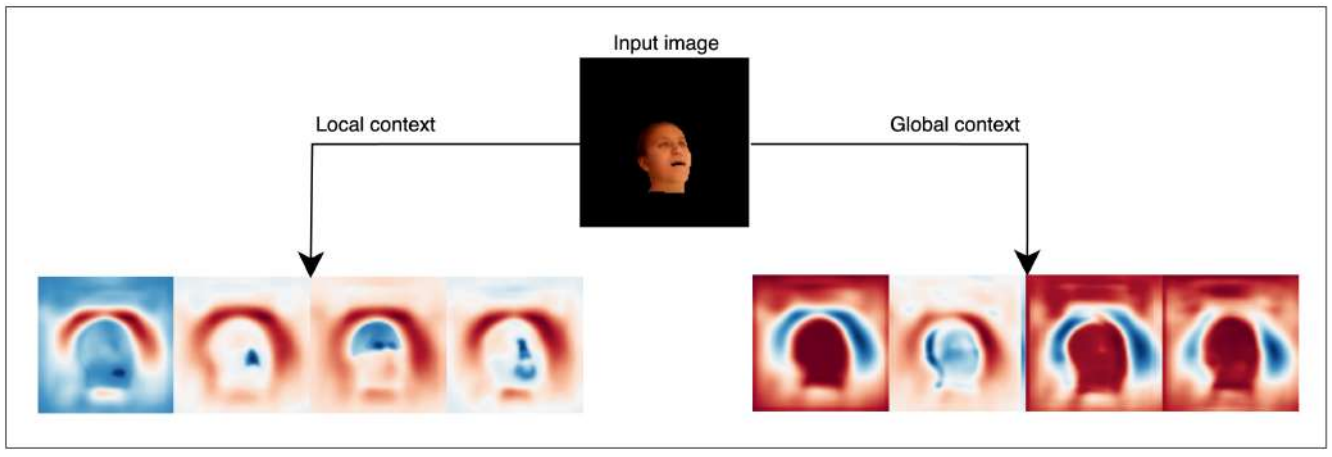


Рисунок 15 – Примеры карт признаков после свёрточных слоёв последнего уровня для частей, выделяющих глобальный и локальный контекст

3.1.3. Обучение параметрической модели головы

Оптимизация параметров предложенной параметрической модели головы производится с использованием итеративного алгоритма градиентного спуска

путем минимизации функции потерь, которая вычисляется по формуле (8). Функция потерь состоит из квадрата ошибки между исходным изображением и синтезированным изображением-проекцией, perceptual функции потерь [77], которая позволяет повысить детализацию результирующего изображения за счет сопоставления карт признаков истинного и синтезированного изображения-проекции из предобученной свёрточной сети VGG16 [78], и членов регуляризации, которые контролируют отклонение исходных векторов, полученных с помощью оценки параметров параметрической модели головы FLAME. Компоненты регуляризации позволяют повысить обобщающую способность модели, добавляя незначительное отклонение к входным данным. В качестве алгоритма оптимизации используется Adam.

$$Loss_{total} = w_1 \cdot \frac{\left\| \left(img_{GT} - img_{predict} \right) \odot img_{mask} \right\|^2}{\sum_{i=0}^{N-1} \sum_{j=0}^{M-1} img_{mask}} +$$

$$w_2 \cdot \sum_{j=0}^{L-1} \left\| \phi_j(img_{GT}) - \phi_j\left(img_{predict}\right) \right\|^2 + \sum_{k \in \left\{ \vec{\beta}, \vec{\theta}, \vec{\psi}, \vec{\tau}, \vec{\lambda} \right\}} w_k \cdot \left\| \overrightarrow{v_k^{offset}} \right\|^2 \rightarrow$$

$$\min_{\substack{CNeRF\ parameters, \\ 2D\ Neural\ Renderer\ parameters, \\ \overrightarrow{v_k^{offset}},\ k \in \left\{ \vec{\beta}, \vec{\theta}, \vec{\psi}, \vec{\tau}, \vec{\lambda} \right\}}}, \quad (8)$$

где img_{GT} – исходное изображение,

$img_{predict}$ – синтезированное изображение,

img_{mask} – маска сегментации головы исходного изображения,

$\phi_j(*)$ – активация i-го слоя предобученной нейронной сети архитектуры VGG16,

$\overrightarrow{v_{\vec{\beta}}^{offset}}, \overrightarrow{v_{\vec{\theta}}^{offset}}, \overrightarrow{v_{\vec{\psi}}^{offset}}, \overrightarrow{v_{\vec{\tau}}^{offset}}, \overrightarrow{v_{\vec{\lambda}}^{offset}}$ – вектора, описывающие смещение от исходных векторов коэффициентов формы, параметров позы (шея, челюсть, глазные яблоки), коэффициентов выражения лица, текстурных коэффициентов и параметров освещения,

$w_1;\ w_2; w_k, k \in \{\vec{\beta}, \vec{\theta}, \vec{\psi}, \vec{\tau}, \vec{\lambda}\}$– взвешивающие коэффициенты для каждого слагаемого.

3.2. Создание синтетического набора данных

В задачах глубокого обучения в последние годы сформировался тренд на использование мощных, обученных на большом наборе данных, моделей для решения какой-либо общей задачи, например обнаружения и локализации объектов (архитектура YOLO [79]), относящихся к разным классам, для адаптации под конкретный домен. Так, требования для настройки весов смягчаются. Во-первых, для обучения требуется небольшое количество итераций (исчисляется десятками-сотнями, зависит от задачи и от масштабности набора данных; техника, известная под названием «перенос обучения» (англ. transfer learning) [80]), во-вторых, не требует наличия набора данных масштаба COCO [81], PASCAL VOC [82] и т.п [83]. В задачах создания цифровых аватаров головы на текущий момент не существует подобных моделей, так как отсутствует общепринятый стандарт модели (архитектурное решение, которое напрямую связно со способом пространственного представления). Также в существующих работах в качестве априорной информации используется информация из различных параметрических моделей головы. Однако даже при определении стандарта архитектуры и параметрической модели для получения априорной информации возникает проблема, которая заключается в отсутствии специализированного набора данных, сопоставимого по масштабам с вышеупомянутыми.

Говоря о трендах глубокого обучения, невозможно не упомянуть мощь синтетических наборов данных, которые спасают в ситуациях полного отсутствия данных [84]. Такие наборы можно создавать с помощью инструментов Blender [85], Unity [86] и т. п. [32]. Модели, обученные на синтетических данных, как правило, требуют несколько реальных экземпляров из того же домена для донастройки.

Так, основываясь на том, что существуют обширные 2D наборы данных, предназначенные для задач распознавания и генерации лиц [87], [88], [89], а также параметрические модели головы, способные предоставлять компактную информацию о внешнем виде и пространственном положении, предлагается сформировать синтетический набор данных, генерируемый в реальном времени,

для обучения общей мощной модели, значения параметров которой могут использоваться для инициализации методов создания аватара головы, реализующих аналогичную архитектуру.

Синтетический набор данных формируется на основе открытого набора данных FFHQ [88] и параметрической модели головы FLAME. Для каждого изображения из набора вычисляются ключевые точки лица и бинарная маска сегментации (включая волосы), после чего выполняется процедура оценки параметров модели FLAME, описанная в подразделе 2.3. На рисунке 16 представлен результат оценки параметров модели FLAME для нескольких изображений.

Рисунок 16 – Результат оценки параметров модели FLAME (столбцы 2, 3) и оригинальное изображение (столбец 1)

Во время обучения параметров модели к векторам FLAME-коэффициентов формы, выражения лица и текстуры каждого изображения добавляется шум, сгенерированный по равномерному распределению в заранее установленном диапазоне значений векторов для получения реалистичных параметров:

$$\vec{\beta} = \min(2, \max(\vec{\beta} + \overrightarrow{z_\beta}, -2)),$$

$$\vec{\theta} = \min(2, \max(\vec{\theta} + \overrightarrow{z_\theta}, -2)),$$

$$\vec{\tau} = \min(2, \max(\vec{\tau} + \overrightarrow{z_\tau}, -2)),$$

где $\vec{\beta}, \vec{\theta}, \vec{\tau}$ – вектора коэффициентов формы, выражения лица и текстуры, полученные при оценке параметров модели FLAME для изображения,

$\overrightarrow{z_\beta} \sim U[-0{,}3;\ 0{,}3], \overrightarrow{z_\theta}, \sim U[-0{,}3;\ 0{,}3], \overrightarrow{z_\tau}, \sim U[-0{,}3;\ 0{,}3]$ – аддитивный шум для коэффициентов формы, выражения лица и текстуры.

Параметры, отвечающие за позу (поворот шеи, смещение челюсти), генерируются из равномерного распределения. Так, например, повороты шеи по оси x и z ограничены диапазоном $[-80°, 80°]$, по оси y – $[-60°, 60°]$. Внешние параметры камеры также генерируются из равномерного распределения в рамках указанного диапазона. На рисунках 17, 18, 19 представлены примеры генераций поворотов шеи.

Рисунок 17 – Демонстрация поворота шеи по оси x в диапазоне $[-80°, 80°]$

Рисунок 18 – Демонстрация поворота шеи по оси y в диапазоне $[-60°, 60°]$

Рисунок 19 – Демонстрация поворота шеи по оси z в диапазоне $[-80°, 80°]$

Представленный подход позволяет создать набор данных, описывающий разнообразие человеческих черт, поворотов и выражений лица, что в свою очередь

обеспечивает отсутствие переобучения и адаптацию к широкому спектру физиологических особенностей и мимики. Преимуществом использования модели FLAME является то, что за счет небольшого разброса диапазона даже небольшие отклонения значений вызывают модификацию итогового изображения. На рисунке 20 представлен пример сгенерированных изображений для одной итерации во время обучения модели.

Рисунок 20 – Пример сгенерированных изображений для одной итерации во время обучения модели

Важно отметить, что при обучении параметров модели коэффициенты $w_{31}, w_{32}, w_{33}, w_{34}, w_{35}$ из формулы (8) равны нулю, так как обучающий набор не фиксированный.

3.3. Стратегия обучения разработанной параметрической модели головы человека

На рисунке 21 представлена схема итерации процесса обучения параметрической модели головы, в ходе которого производится оптимизация всех обучаемых параметров с использованием синтетического набора данных, генерируемого в реальном времени с использованием параметрической модели головы FLAME, описание которого приведено в подразделе 3.2.

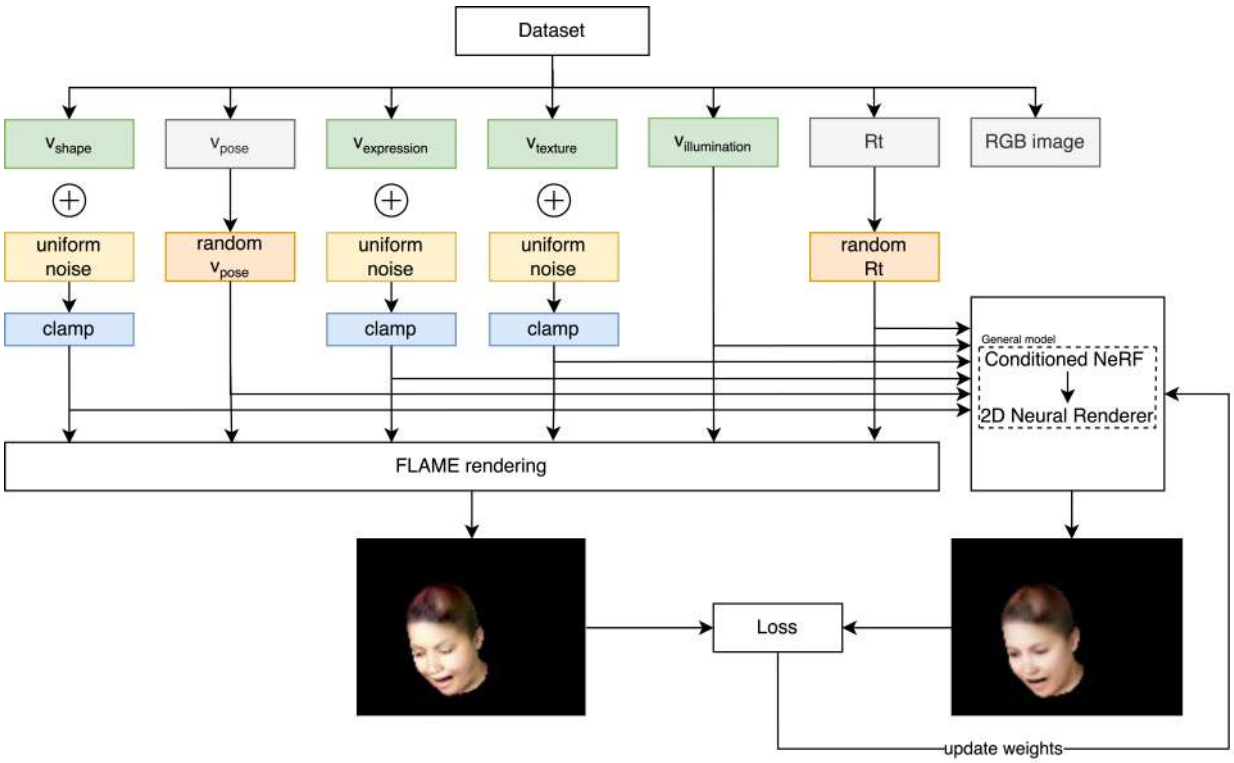


Рисунок 21 – Схематическое представление итерации процесса обучения параметрической модели головы человека с использованием синтетического набора данных

Для проведения экспериментальных исследований обучения параметрической модели головы была предложена стратегия обучения, при которой вероятность включения синтетических данных рассчитывается в зависимости от номера итерации согласно формуле (9). Была выдвинута гипотеза, что такой подход к обучению позволит добавить в обучаемую параметрическую модель априорную информацию о высокоуровневых признаках лиц людей, отсутствующих в изображениях, полученных в результате рендеринга модели FLAME.

$$p = \begin{cases} 1, \text{если } n \le M, \\ \left(1 - \frac{n-M}{N-M}\right) \cdot (p_{\max} - p_{min}) + p_{\min}, \text{если } n > M. \end{cases} \quad (9)$$

где n – номер текущей итерации,

M – количество итераций для обучения исключительно на синтетических данных,

N – общее количество итераций обучения,

p_{max} – максимальное значение вероятности включения синтетических данных,

p_{min} – минимальное значение вероятности включения синтетических данных,

p – вероятность использования синтетических данных в процессе обучения,

$1 - p$ – вероятность использования данных из FFHQ в процессе обучения.

На рисунке 22 представлена схема процесса обучения параметрической модели головы, в ходе которого производится оптимизация всех обучаемых параметров метода на наборе данных FFHQ, который предварительно проходит процедуру оценки параметров модели FLAME.

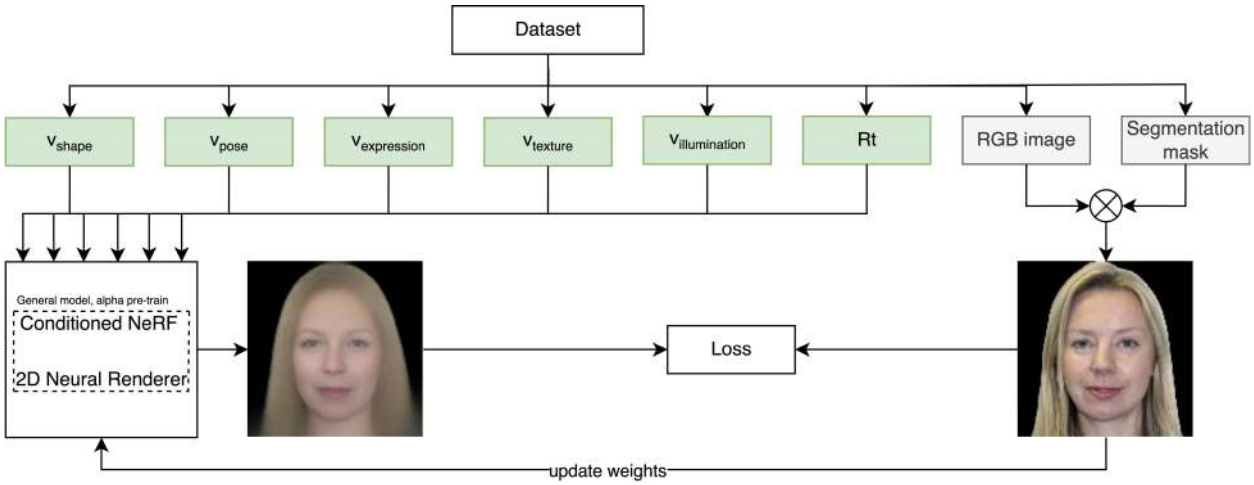


Рисунок 22 – Схематическое представление итерации процесса обучения параметрической модели головы человека с использованием набора данных FFHQ

В соответствии с классификацией методов создания аватара, приведенной в разделе 1, разработанная параметрическая модель характеризуется по критериям следующим образом:

1. По способу представления поверхности аватара: неявное представление поверхности.
2. По уровню обобщенности модели: общая модель.
3. По формату набора данных для обучения: синтетический набор данных, 2D набор данных.

4. По возможности управления параметрами модели: «распутанное» пространство параметров (так как вектора в синтетическом наборе данных генерируются по равномерному распределению, подробнее в подразделе 3.2).
5. По необходимости предварительной оценки параметров параметрической модели: требуется.

3.4. Экспериментальные исследования разработанной параметрической модели головы

Программный код в рамках данного диссертационного исследования был реализован с использованием интерпретируемого языка программирования Python [90] версии 3.10 и фреймфорков глубокого обучения и компьютерного зрения PyTorch [91], PyTorch3D [92], OpenCV [93], NumPy [94]. Запуск всех экспериментов выполнялся на персональном компьютере с характеристиками Intel Core i9 14900K, RTX4090 24GB, 64GB DDR5 RAM.

В рамках исследования предложенной параметрической модели головы был проведен ряд экспериментов:

1. Экспериментальные исследования, показывающие эффективность предложенного блока в архитектуре нейронной сети, выполняющей двумерный нейронный рендеринг, по сравнению с методом, представленным в [48].
2. Экспериментальные исследования, демонстрирующие работоспособность предложенной параметрической модели головы, а именно способность синтезировать изображения, соответствующие изображениям, синтезированным при помощи модели FLAME. В рамках стратегии принято $M = N,\ N = 310000$ (см. формулу (9)). Здесь и далее будем обозначать данный эксперимент как α_1.
3. Экспериментальные исследования, демонстрирующие работоспособность предложенной параметрической модели головы, а именно способность

синтезировать изображения, соответствующие изображениям, синтезированным при помощи модели FLAME. В рамках стратегии принято $M = N$, $N = 310000$ (см. формулу (9)). Последние 10000 итераций обучения выполняются с использованием модифицированной в области рта текстуры. Обозначение эксперимента – α_2.

4. Экспериментальные исследования с параметрами стратегии принятыми $M = 310000, N = 312500, p_{\min} = 0, p_{\max} = 0$ (см. формулу (9)). Данный эксперимент необходим для того, чтобы определить вклад реальных данных в процесс обучения. Во избежание эффекта переобучения требуется небольшое количество итераций при $p = 0$. Обозначение эксперимента – β_1.
5. Экспериментальные исследования с параметрами стратегии принятыми $M = 310000, N = 312500, p_{\min} = 0, p_{\max} = 0$ (см. формулу (9)). В данном эксперименте к реальным данным с вероятностью $0{,}5$ применяется аугментация, моделирующая изменение расстояния по оси z головы от камеры. Обозначение эксперимента – β_2.
6. Экспериментальные исследования с параметрами стратегии принятыми $M = 310000,\ N = 312500,\ p_{\min} = 0{,}25,\ p_{\max} = 0{,}25$ (см. формулу (9)). Обозначение эксперимента – β_3.
7. Экспериментальные исследования с параметрами стратегии принятыми $M = 310000,\ N = 312500,\ p_{\min} = 0,\ p_{\max} = 0{,}25$ (см. формулу (9)). Обозначение эксперимента – β_4.
8. Экспериментальные исследования с параметрами стратегии принятыми $M = 310000,\ N = 312500,\ p_{\min} = 0{,}25,\ p_{\max} = 0{,}25$ (см. формулу (9)). В данном эксперименте к реальным данным с вероятностью $0{,}5$ применяется аугментация, моделирующая изменение расстояния по оси z головы от камеры. Обозначение эксперимента – β_5.
9. Экспериментальные исследования с параметрами стратегии принятыми $M = 310000,\ N = 312500,\ p_{\min} = 0,\ p_{\max} = 0{,}25$ (см. формулу (9)). В данном эксперименте к реальным данным с вероятностью $0{,}5$

применяется аугментация, моделирующая изменение расстояния по оси z головы от камеры. Обозначение эксперимента – β_6.

Представленный набор экспериментов позволяет оценить влияние различных настроек стратегии и аугментаций на итоговый результат рендеринга.

Как говорилось ранее, архитектура нейронной сети, которая выполняет двумерный нейронный рендеринг, основана на идеях из работ [73], [72], [48]. Здесь и далее метод создания аватара, предложенный в [48], будет именоваться *baseline*. Еще одно отличие от [48] – наличие шеи у аватара, так как вместо модели BFM используется FLAME. Оптимизация параметров производится в течение 300000 итераций для всех обучаемых разрешений (128, 256, 512). Здесь и далее под итерацией будет пониматься прохождение через все блоки метода 64 обучающих примера.

Экспериментальное исследование (1) посвящено проверке гипотезы об эффективности модифицированной архитектуры нейронной сети, которая выполняет двумерный нейронный рендеринг. Для этого программный код из [48] был адаптирован для работы с моделью FLAME, то есть был изменен процесс формирования вектора входных значений, алгоритмические решения не затрагивались, а также произведена интеграция модуля генерации синтетических данных в реальном времени. Все гипотезы проверялись на разрешении 128×128, затем лучшие решения использовались для настройки параметров на других разрешениях. Оптимизация параметров для *baseline* и предложенного метода производилась в течение 30000 итераций, количество обучающих примеров за одну итерацию – 64. На рисунках 23, 24 и 25 представлены графики для показателей качества *perceptual loss*, *MSE* в логарифмической шкале и *PSNR*, который вычисляется по формуле (10) (формулы для *perceptual loss* и *MSE* были приведены в подразделе 3.1.3). На рисунке 26 представлены результаты синтеза изображений-проекций с помощью предложенного метода и *baseline.* Исходя из представленных качественных и количественных результатов, можно заключить, что предложенное архитектурное решение позволило значительно превзойти *baseline*. По истечении 30000 итераций предложенная параметрическая модель уже способна

синтезировать изображения-проекции, обладающие высокой степенью визуальной схожести с изображениями, полученными в результате рендеринга модели FLAME, в отличие от *baseline*.

$$PSNR = 10 \cdot \log_{10}\left(\frac{\mathrm{M}AX(image_{GT})^2}{MSE(image_{GT}, image_{predict})}\right) \qquad (10)$$

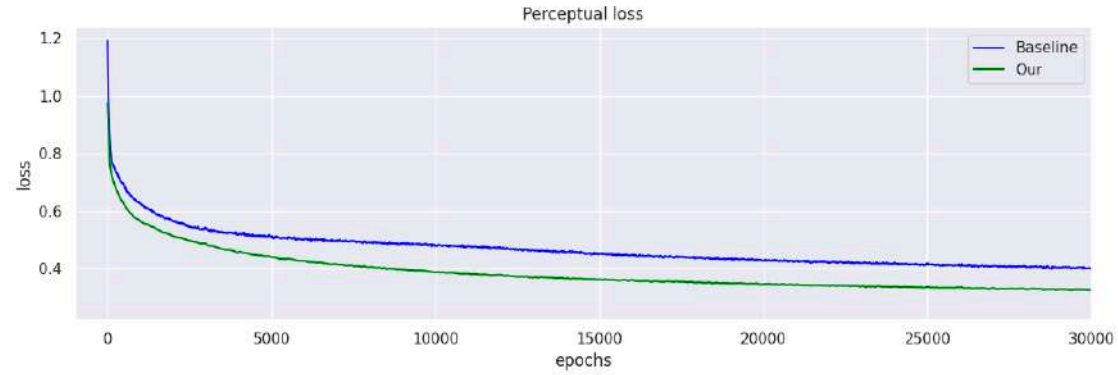


Рисунок 23 – График для показателя качества *perceptual loss* по результатам эксперимента

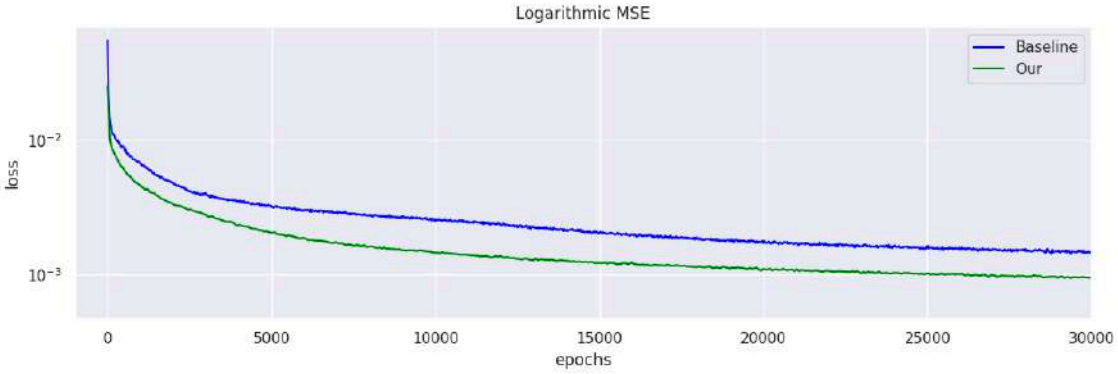


Рисунок 24 – График для показателя качества MSE в логарифмической шкале по результатам эксперимента

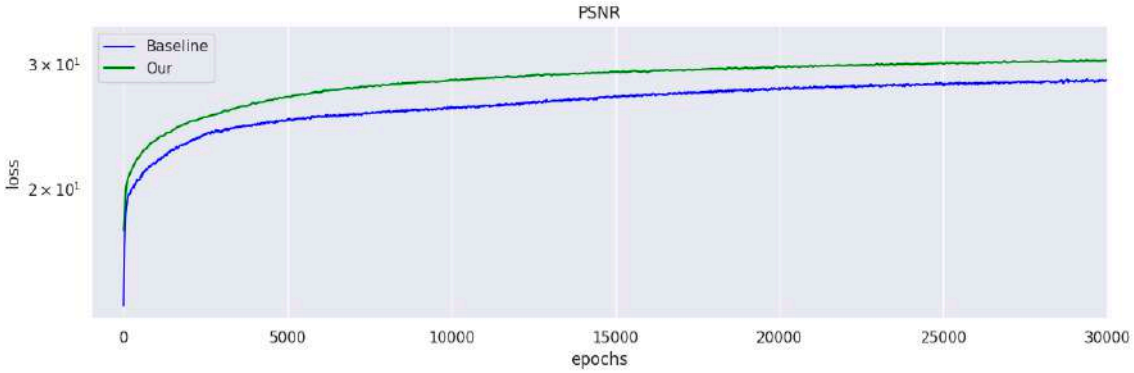


Рисунок 25 – График для показателя качества PSNR по результатам эксперимента

Рисунок 26 – Результаты синтеза изображений-проекций с помощью предложенной параметрической модели головы и *baseline*, обученных в течение 30000 итераций

Экспериментальное исследование, обозначенное как α_1, посвящено проверке гипотезы о способности разработанной модели синтезировать изображения, соответствующие изображениям, синтезированным при помощи параметрической модели головы FLAME. Для этого производится оптимизация параметров метода в течение 310000 итераций, количество обучающих примеров за одну итерацию – 64. На рисунке 27 представлен результат синтеза изображений-проекций, на рисунке

А.1 в Приложении А результат синтеза изображений-проекций в зависимости от удаленности камеры, на рисунке А.2 в Приложении А результат синтеза изображений-проекций в зависимости от положения шеи (по осям x, y, z), на рисунке А.3 в Приложении А результат синтеза изображений-проекций в зависимости от степени открытия челюсти, на рисунке А.4 в Приложении А результат синтеза изображений-проекций в зависимости от выражения лица, на рисунке А.5 в Приложении А результат синтеза изображений-проекций с новых точек обзора. Все изображения-проекции синтезируются с разрешением 128×128.

Исходя из представленных качественных результатов, можно заключить, что получаемое в результате представление способно синтезировать изображения, соответствующие изображениям, синтезированным при помощи модели FLAME, на высоком уровне. При модификации положения головы, удаленности камеры, выражения лица, степени открытия челюсти и синтеза с новых ракурсов идентичность сохраняется, что характерно для модели FLAME.

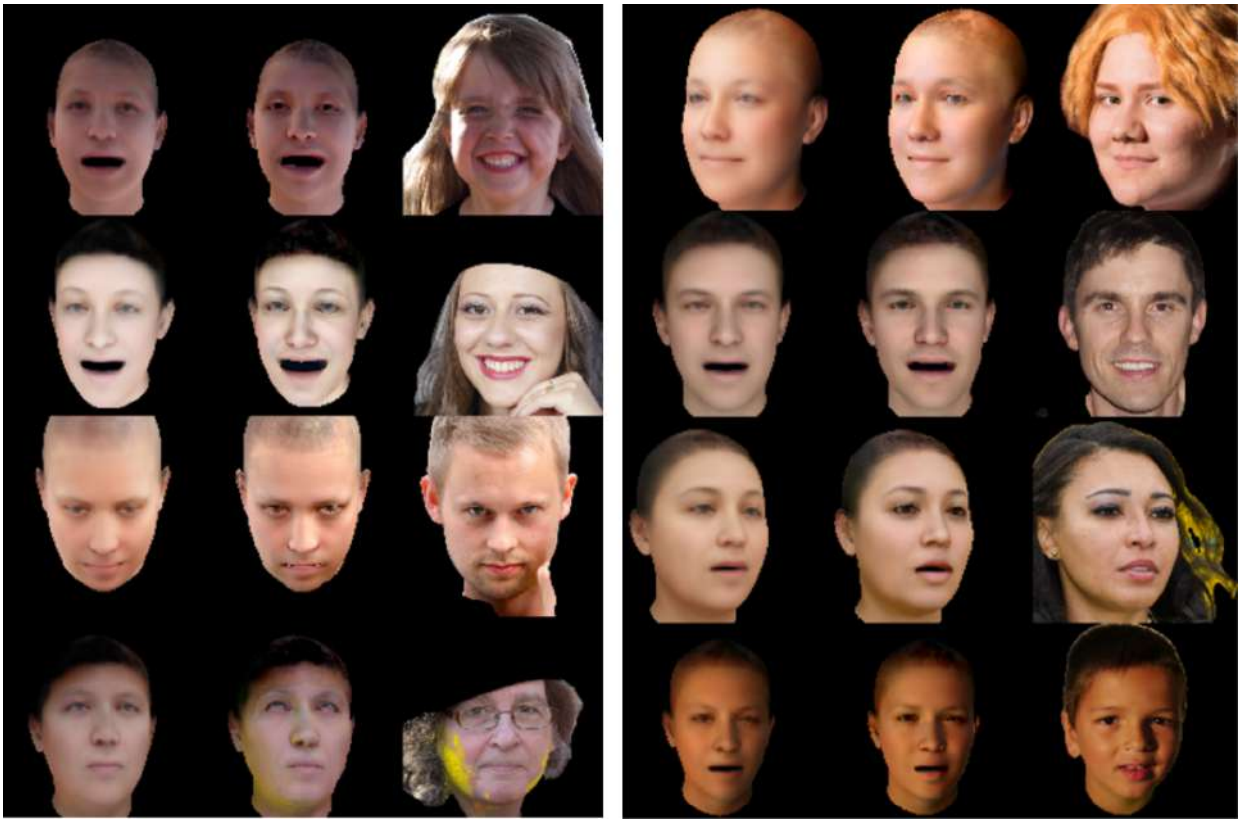


Рисунок 27 – Результат синтеза изображений-проекций на основе обученных параметров. Справа в каждой колонке – оригинальное изображение, слева – синтезированное, посередине – полученное при рендеринге для модели FLAME

В текстуре и полигональной сетке параметрической модели головы FLAME отсутствует внутренняя область рта, что можно увидеть на рисунках выше. В рамках экспериментального исследования α_2 была выдвинута гипотеза, что модификация текстуры и полигональной сетки потенциально может привести к тому, что полученная модель будет лучшим источником априорной информации о внешнем виде головы человека по сравнению с моделью, обученной без использования модифицированной текстуры. Для этого, без добавления новых вершин в исходную полигональную сетку, вручную, с помощью инструмента Blender, были добавлены новые грани в области рта, а также была выполнена «грубая» модификация исходной текстурной развёртки. На рисунке 28 приведена демонстрация исходной и модифицированной развертки. На рисунке 29 представлены результаты синтеза изображений-проекций из донастроенных параметров (разрешение 128×128). Выводы об эффективности донастройки параметров модели, обученной с использованием модифицированной текстуры, будут сделаны в разделе 4.

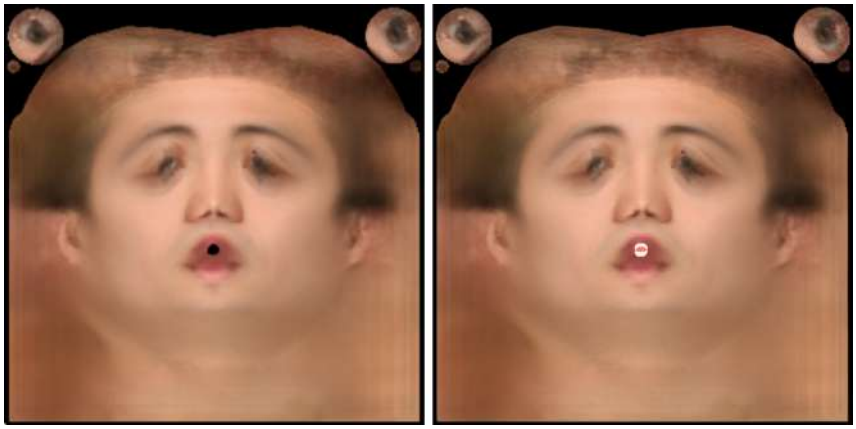


Рисунок 28 – Исходная (слева) и модифицированная (справа) текстурная развертка модели FLAME

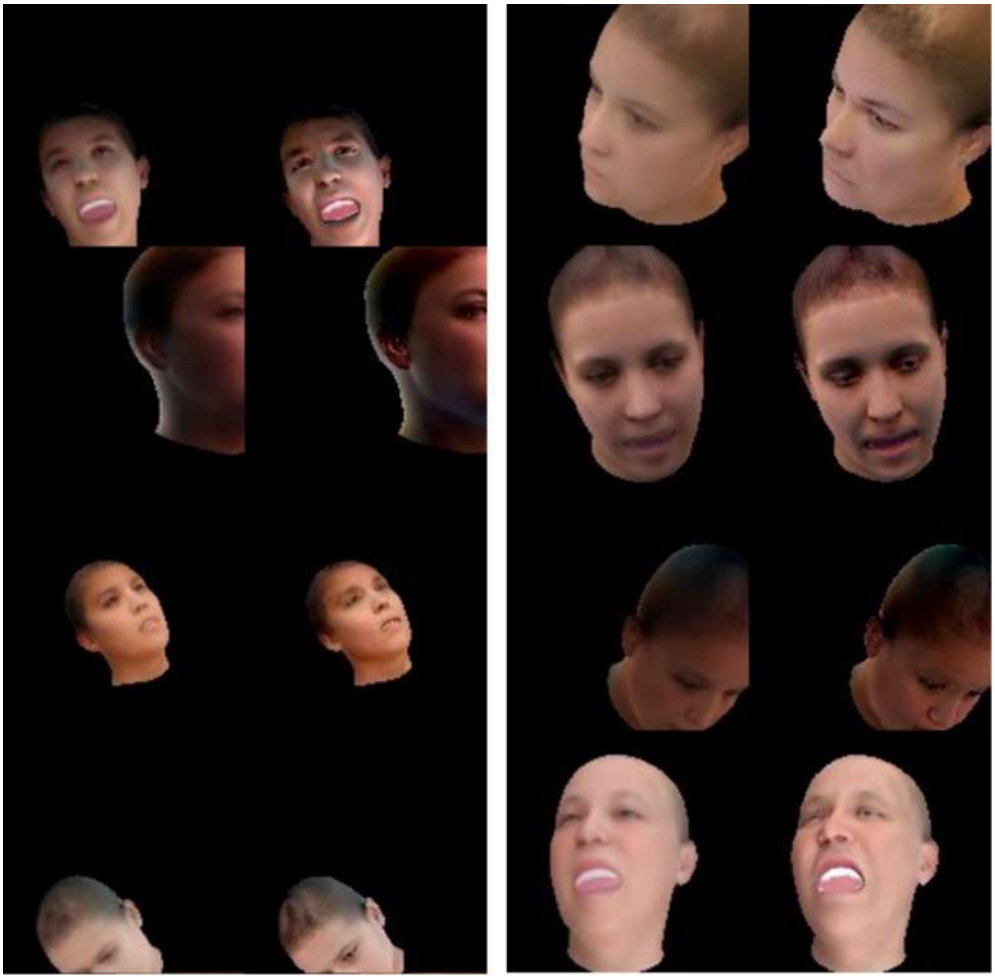


Рисунок 29 – Результат синтеза изображений-проекций из донастроенных параметров метода на модифицированной текстуре. Правый столбец в каждой колонке – результат рендеринга для модели FLAME, левый столбец – результат синтеза

По результатам экспериментальных исследований $\beta_1 - \beta_6$ было получено 12 версий разработанной параметрической модели головы, способных синтезировать изображения-проекции с разрешением 128×128. Первые 6 из них соответствуют способу обучения, применяемому в α_1 (без модифицированной текстуры), остальные, соответственно, α_2 (с модифицированной текстурой). На рисунках А.6 и А.10 в Приложении А представлены результаты синтеза экспериментов $\beta_1 - \beta_6$ с контролем выражения лица на некоторых примерах из тестовой выборки для способов обучения α_1 и α_2, соответственно. На рисунках А.7 и А.11 в Приложении А представлены результаты синтеза экспериментов $\beta_1 - \beta_6$ с контролем степени открытия челюсти на некоторых примерах из тестовой выборки для способов

обучения α_1 и α_2, соответственно. На рисунках А.8 и А.12 в Приложении А представлены результаты синтеза экспериментов $\beta_1 - \beta_6$ с контролем степени поворота головы на некоторых примерах из тестовой выборки для способов обучения α_1 и α_2, соответственно. На рисунках А.9 и А.13 в Приложении А представлены результаты синтеза экспериментов $\beta_1 - \beta_6$ с контролем расстояния от камеры до головы по оси z на некоторых примерах из тестовой выборки для способов обучения α_1 и α_2, соответственно.

На рисунках 30–32 представлены наиболее показательные результаты синтеза, которые позволяют увидеть вклад модифицированной текстуры, аугментации, моделирующей изменение расстояния по оси z головы от камеры, и включения синтетических данных в процесс обучения, соответственно.

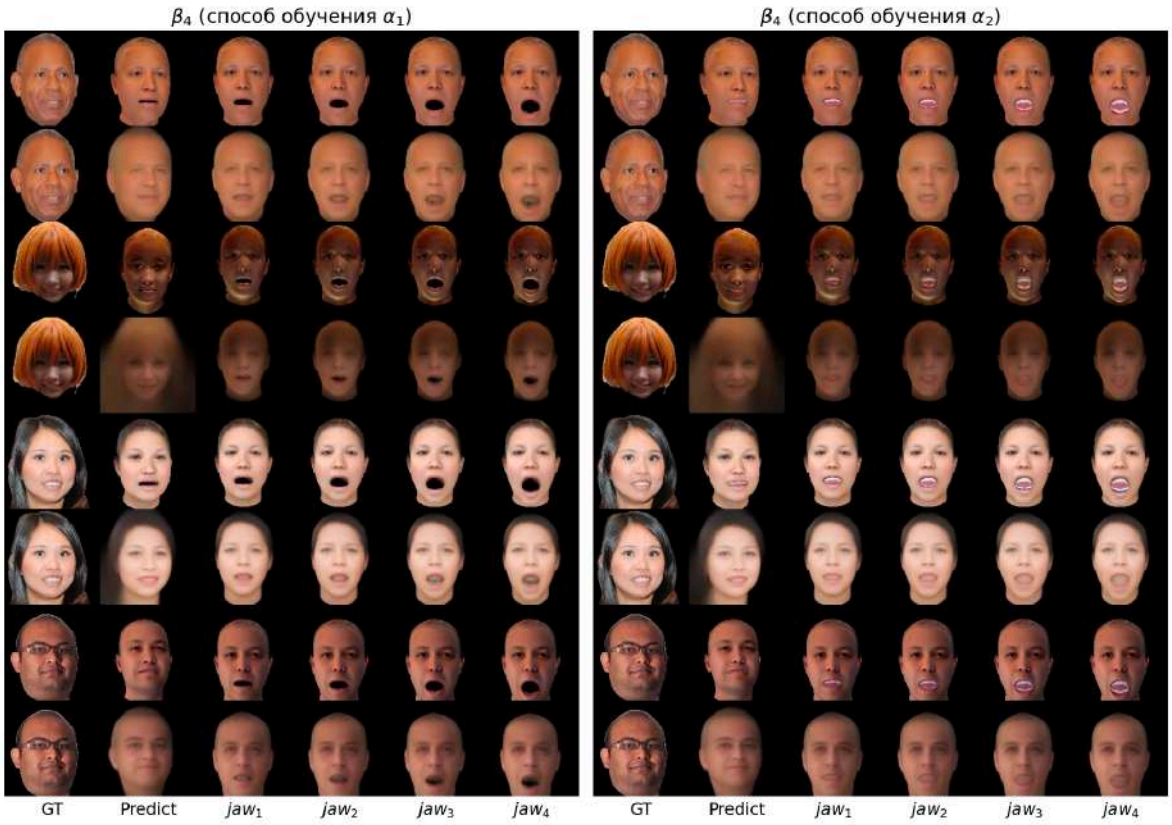


Рисунок 30 – Результаты синтеза на тестовой выборке для контроля степени открытия челюсти. Слева представлены результаты эксперимента β_4 для способа обучения α_1. Справа представлены результаты эксперимента β_4 для способа обучения α_2. Четные строки – результаты синтеза, нечетные – результат рендеринга для модели FLAME

Рисунок 31 – Результат синтеза на тестовой выборке для контроля расстояния от камеры до головы по оси z. Слева представлены результаты эксперимента β_1 для способа обучения α_2. Справа представлены результаты эксперимента β_2 для способа обучения α_2. Четные строки – результаты синтеза, нечетные – результат рендеринга для модели FLAME

Рисунок 32 – Результаты синтеза на тестовой выборке для контроля поворота шеи. Слева представлены результаты эксперимента β_2 для способа обучения α_2. В центре представлены результаты эксперимента β_5 для способа обучения α_2. Справа представлены результаты эксперимента β_6 для способа обучения α_2. Четные строки – результаты синтеза, нечетные – результат рендеринга для модели FLAME

Исходя из представленных результатов, проведенных экспериментальных исследований, можно сделать следующие выводы:

1. Разработанная модель обладает способностью синтезировать изображения, соответствующие изображениям, синтезированным при помощи модели FLAME, что подтверждают представленные показатели качества (PSNR≈30 дБ) и полученные изображения-проекции для тестовых данных.
2. Разработанный блок в архитектуре нейронной сети, выполняющей двумерный нейронный рендеринг, позволил уменьшить количество

итераций для обучения модели более чем в 2 раза по сравнению с методом, представленным в [48].

3. Использование модифицированной текстуры приводит к тому, что изображения, синтезируемые при помощи разработанной модели, имеют большее сходство с реальной человеческой головой. Влияние использования модифицированной текстуры можно увидеть на рисунке 30.
4. Использование набора данных FFHQ в процессе обучения приводит к тому, что результаты синтеза обладают большей фотореалистичностью.
5. Применение аугментации, моделирующей изменение расстояния по оси z головы от камеры, к изображениям из набора данных FFHQ приводит к уменьшению эффекта переобучения на результат обучения разработанной модели. Результат переобучения можно увидеть на рисунке 31.
6. Использование стратегии обучения разработанной модели с $p_{\min} = p_{\max} = 0$ приводит к худшим результатам, чем обучение с $p_{\min} = p_{\max} = 0{,}25$. Использование стратегии с параметрами $p_{\min} = p_{\max} = 0{,}25$ позволяет снизить эффект переобучения модели.
7. Использование стратегии обучения разработанной модели с $p_{min} = p_{\max} = 0{,}25$ приводит к худшим результатам, чем обучение с $p_{\max} = 0{,}25$ и $p_{\min} = 0$. Использование стратегии с параметрами $p_{\min} = 0$, $p_{\max} = 0{,}25$ позволяет снизить эффект переобучения до минимума, так как при $p_{\min} = p_{\max}$ параметрическая модель обучается отделять параметры, соответствующие синтетическим и реальным данным. Особенно заметен эффект переобучения на рисунке 32. Качественно обученная модель должна обладать способностью синтезировать изображения с высокой консистентностью, то есть при изменении параметров, отвечающих за наклон или поворот головы, должен изменяться только поворот или наклон головы на синтезируемом изображении, и это изменение параметра не должно затрагивать другие свойства человеческого лица (форма, выражение лица).

3.5. Выводы и результаты третьего раздела

В данном разделе было представлено подробное описание предлагаемой параметрической модели головы человека на основе нейросетевой модели представления поверхности CNeRF, двумерного нейронного рендеринга и синтетического набора данных, генерируемого в режиме реального времени с использованием параметрической модели головы FLAME. Был описан процесс формирования синтетического набора данных, необходимого для получения мощной общей модели, способной выступить в качестве инициализации параметров метода создания аватара конкретного человека.

Предложенная параметрическая модель основана на использовании актуальных идей и технологий. По сравнению с существующими решениями при её обучении не требуется наличия специализированного набора данных.

На основе проведенных экспериментальных исследований разработанного метода можно сделать следующие выводы:

1. Интеграция разработанного блока повышения дискретизации в модуле двумерного нейронного рендеринга позволила ускорить сходимость процедуры оптимизации более чем в 2 раза (см. результаты экспериментальных исследований, соответствующих пункту (1)).
2. Сформированный набор настроек (в том числе аугментаций) процедуры оптимизации параметров позволил значительно увеличить фотореалистичность синтезируемых изображений-проекций, в том числе для параметров, сильно отличающихся от экземпляров в обучающем наборе.
3. Разработанная модель обладает способностью синтезировать изображения-проекции, соответствующие изображениям, синтезированным при помощи модели FLAME, что подтверждают представленные показатели качества (например, PSNR более 30 дБ) и полученные изображения-проекции для тестовых данных.

Таким образом, поставленная задача разработать параметрическую модель головы выполнена. По теме раздела опубликованы работы [23*], [24*].

РАЗДЕЛ 4. МЕТОД СОЗДАНИЯ ПАРАМЕТРИЗОВАННОГО АВАТАРА ГОЛОВЫ ЧЕЛОВЕКА

Данный раздел диссертационного исследования посвящен описанию разработанного метода создания параметризованного аватара головы человека и соответствующим экспериментальным исследованиям. Также приведено сравнение с актуальными существующими методами и описан полный цикл создания и анимации аватара.

4.1. Описание разработанного метода

Мотивация разработки нового метода создания цифрового аватара головы человека заключается в том, что среди существующих современных методов практически невозможно найти компромисс между качеством синтезируемых изображений-проекций аватара и скоростью его получения. Так, методы, синтезирующие изображения-проекции аватаров с высокой степенью идентичности, как правило, требуют отдельного обучения параметров под каждого человека, длительность которого исчисляется десятками часов. Методы, формирующие представление аватара за время, исчисляемое минутами, как правило, синтезируют изображения-проекции аватаров с недостаточной степенью идентичности, а также требуют на этапе обучения 3D набор данных.

Разработанная параметрическая модель головы, подробно описанная в предыдущем разделе диссертационного исследования, может быть использована в качестве стартовой точки для дальнейшей оптимизации параметров под конкретного человека, так как имеет хорошие стартовые значения, которые позволяют синтезировать изображения-проекции, соответствующие изображениям, синтезированным при помощи модели FLAME. Однако, в отличие от FLAME, параметры предложенной параметрической модели возможно точно настроить для конкретного человека, используя в том числе короткую монокулярную видеопоследовательность (порядка 1000 кадров) и небольшое число итераций стохастического градиентного спуска, чтобы получить возможность

синтезировать новые согласованные виды и мимику для целевого лица с минимальным значением ошибки.

Таким образом, разработанный метод создания параметризованного аватара головы человека использует архитектуру предложенной параметрической модели головы, подробно описанную в разделе 3, и её веса для инициализации параметров. При этом внесены следующие изменения в процесс обучения параметров:

1. Для всех итераций обучения вектор коэффициентов формы $\vec{\beta}$ и вектор текстурных коэффициентов $\vec{\tau}$ фиксируются и не подлежат изменению в процессе синтеза новых изображений-проекций.
2. Значения параметров первых четырех полносвязных слоев многослойного перцептрона в нейросетевой модели представления поверхности CNeRF фиксируются и не подлежат обновлению в процессе обучения. Такая стратегия используется того, чтобы избежать эффект переобучения.
3. Обучающий набор формируется из монокулярной видеопоследовательности (достаточно порядка 1000 кадров) одного человека, содержащей его мимику и вариацию поз.

На рисунке 33 представлена схема итерации процесса обучения предлагаемого метода создания параметризованного аватара головы человека. По результатам обучения итоговые параметры метода, вектор коэффициентов формы и вектор текстурных коэффициентов фиксируются и определяют аватара. Используя такое представление аватара, можно выполнять синтез новых видов, модифицируя параметры камеры, а также выполнять перенос выражения лица и позы с других видеопоследовательностей, содержащих мимику (не обязательно с человеком, чьё представление было получено), путём замены вектора коэффициентов выражения лица и параметров позы.

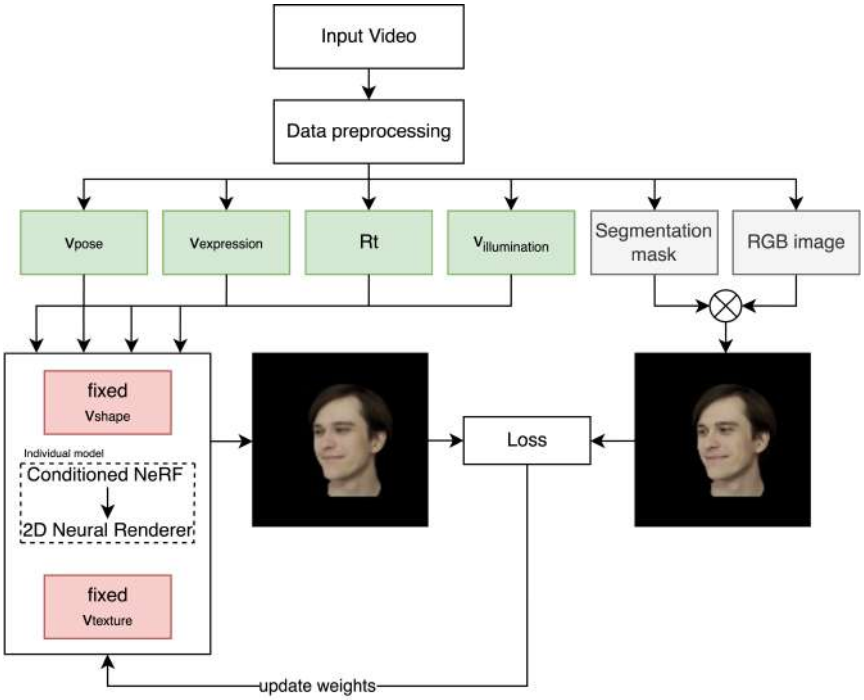


Рисунок 33 – Схематическое представление итерации процесса обучения предлагаемого метода создания параметризованного аватара головы человека

На рисунке 34 представлена схема полного цикла создания аватара головы человека при помощи разработанного метода.

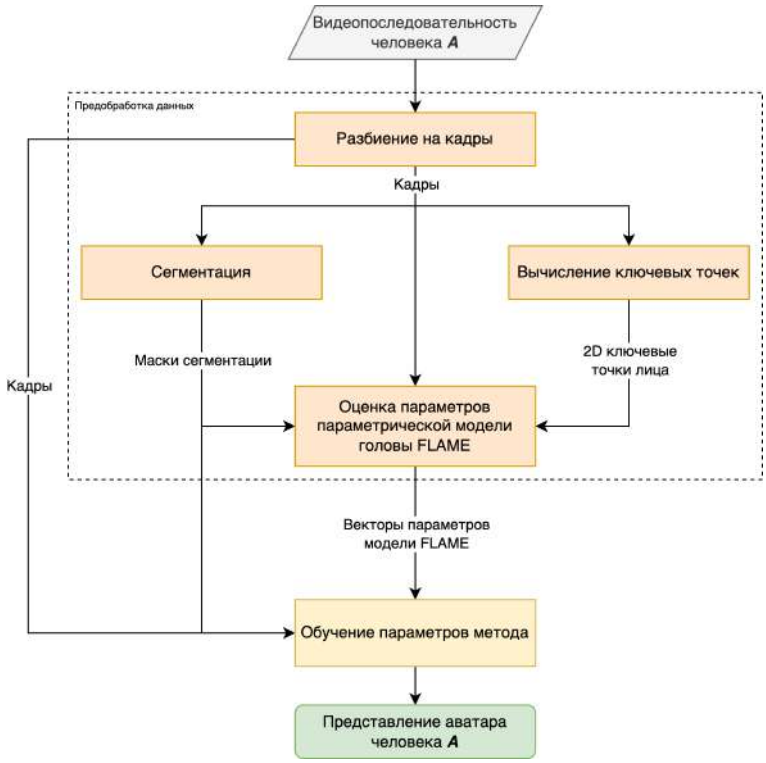


Рисунок 34 – Схема полного цикла создания аватара головы человека при помощи разработанного метода

В соответствии с классификацией методов создания аватара, приведенной в разделе 1, разработанный метод создания параметризованного аватара головы человека характеризуется по критериям следующим образом:

1. По способу представления поверхности аватара: неявное представление поверхности.
2. По уровню обобщенности модели: персональная модель.
3. По формату набора данных для обучения: 2D набор данных.
4. По возможности управления параметрами модели: «распутанное» пространство параметров.
5. По необходимости предварительной оценки параметров параметрической модели: требуется.

4.2. Приложения: синтез новых видов и перенос выражения лица

На основе параметров метода, полученных в результате оптимизации, можно выполнять синтез новых видов, под которым понимается синтез изображений-проекций аватара с различных точек обзора. Для этого необходимо выполнить замену значений внешних параметров камеры (матрица поворота и вектор сдвига) на требуемые. На рисунке 35 схематично представлена процедура синтеза новых видов аватара.

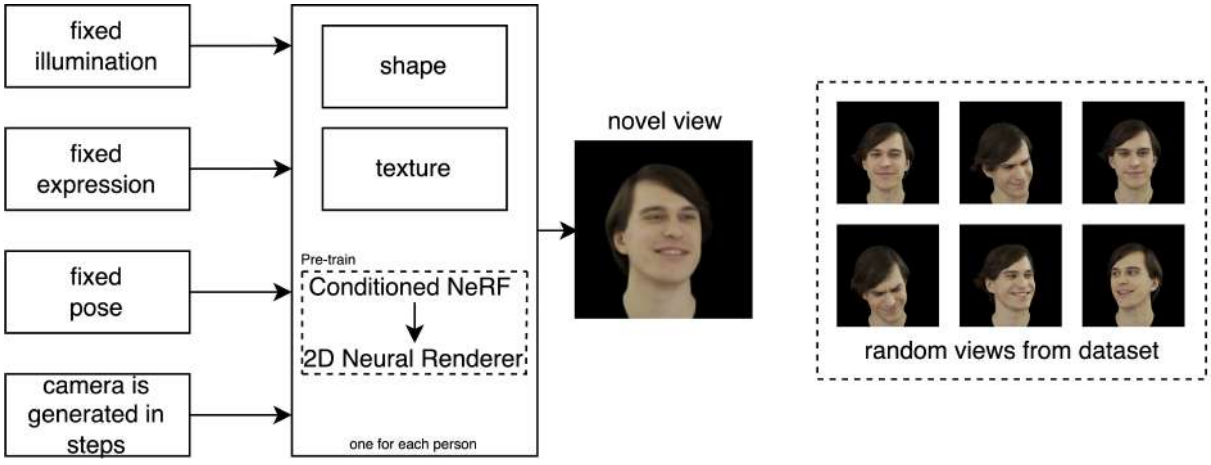


Рисунок 35 – Схематичное представление отдельной итерации синтеза новых видов аватара

Аналогично можно выполнять процедуру переноса выражения лица с кадров видеопоследовательности, содержащей мимику человека. Эта процедура включает в себя следующую последовательность действий:

1. Выполнить оценку параметров модели FLAME по видеопоследовательности.
2. На вход метода с обученными и зафиксированными параметрами подать вектор коэффициентов выражения лица для требуемого кадра.
3. Опционально: подать параметры камеры для кадра, полученные в результате оценки параметров, и/или параметры позы и/или параметры освещения.

На рисунке 36 схематично представлена процедура переноса выражения лица для аватара с кадра видеопоследовательности. Важно отметить, что видеопоследовательность, на основе которой будет осуществляться перенос выражения лица, может содержать мимику любого человека.

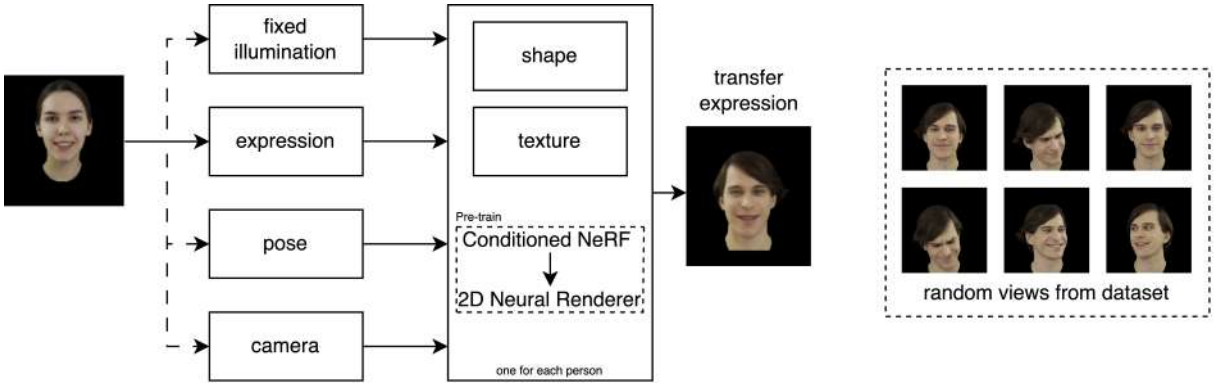


Рисунок 36 – Схематичное представление отдельной итерации процедуры переноса выражения лица для аватара

На рисунке 37 приведено схематичное представление полного цикла процедуры анимации по видеопоследовательности (в т. ч. перенос выражения лица). На рисунке 38 приведено схематичное представление полного цикла процедуры синтеза новых видов.

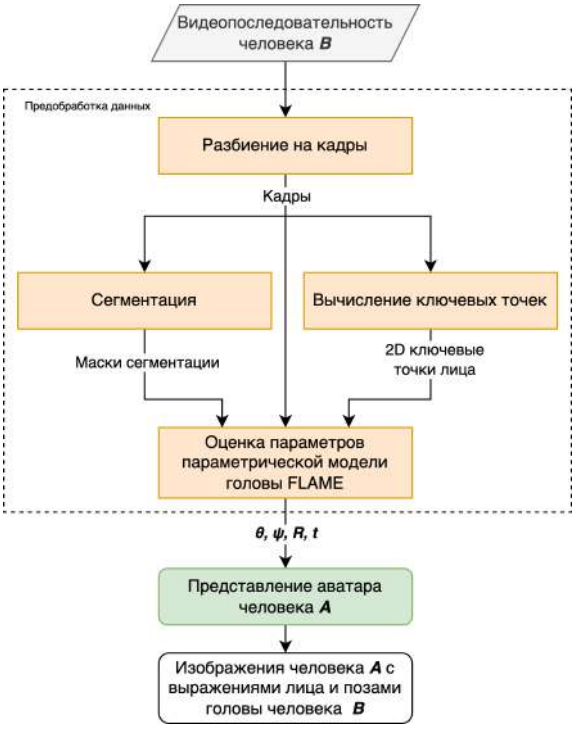


Рисунок 37 – Схема процедуры анимации аватара головы человека по входной видеопоследовательности

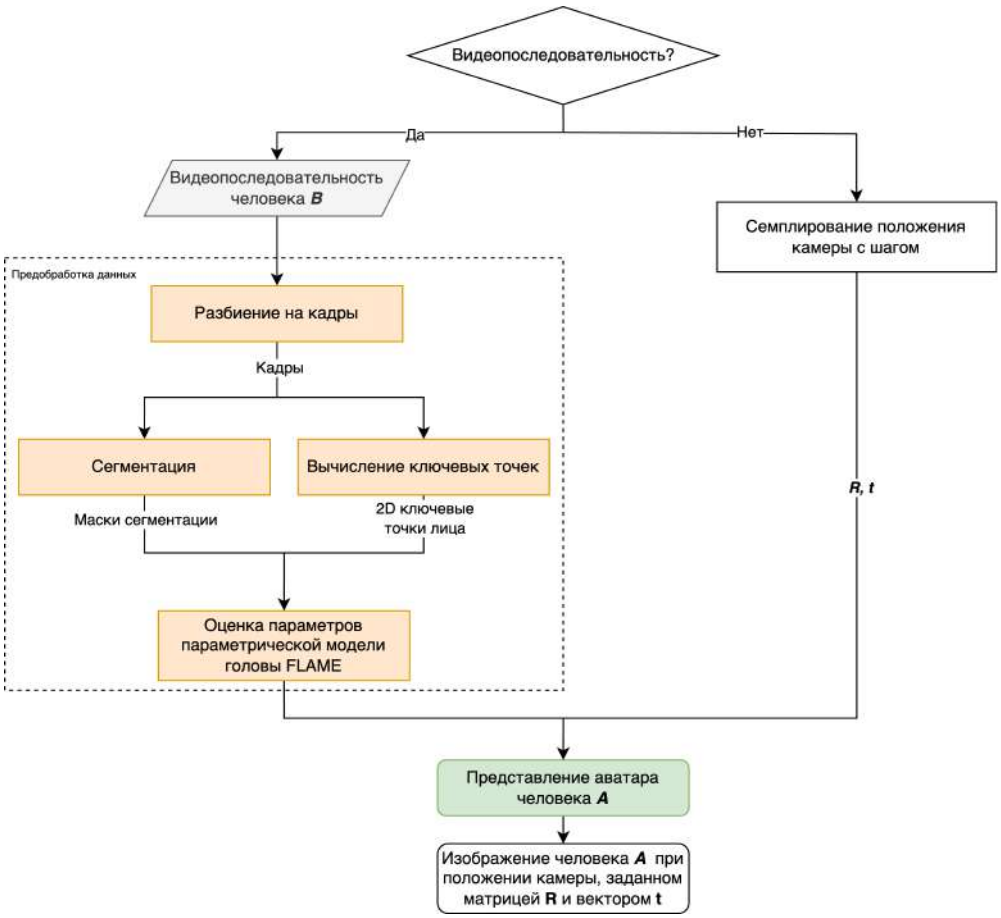


Рисунок 38 – Схема процедуры синтеза новых видов аватара

Описанные приложения разработанного метода создания цифрового аватара головы человека могут быть использованы на практике для компактной передачи информации по узкому каналу связи, например, во время видеоконференции. Так, на приемной стороне требуется выполнить инициализацию параметров метода для конкретного аватара, мимику и позу которого требуется воспроизводить. На стороне передачи при этом требуется производить захват видеопотока, по кадрам которого выполнять предобработку, описанную в подразделе 3.2. По результатам оценки параметров параметрической модели головы на приемную сторону требуется передавать лишь вектор позы, вектор выражения лица и внешние параметры камеры (матрица поворота и вектор сдвига), которые по сравнению с изображением являются менее избыточной формой представления исходных данных. Важно отметить, что количество передаваемой информации при необходимости можно уменьшить. Так, если для визуализации имеет смысл лишь идентичная мимика области рта, то требуется передавать лишь некоторые компоненты вектора выражения лица.

4.3. Способ расширения набора данных с помощью интерполяции промежуточных кадров

В рамках диссертационного исследования в качестве способа трехмерного представления сцены выбрана модификация нейросетевой модели представления поверхности NeRF. Авторами оригинальной работы [25] было показано, что с ростом количества изображений в обучающей выборке, увеличивается качество синтезируемых изображений в соответствии со значениями показателей качества PSNR, SSIM, LPIPS.

Одним из способов увеличения количества кадров в выборке могут служить методы интерполяции промежуточных кадров. Видеопоследовательности для создания аватаров содержат сцены, которые можно отнести к категории динамических. В ходе диссертационного исследования было выполнено сравнение методов интерполяции промежуточных кадров. Были рассмотрены [18*] такие

методы как: XVFI [95], RRIN [96], CDFI [97], RIFE [98], AdaCof [99]. По результатам сравнения, представленным в работе [18*], было выявлено, что метод XVFI показывает наилучшее качество интерполяции для видеопоследовательностей с высокой вариативностью. На основе полученных результатов было произведено исследование влияния интерполяции промежуточных кадров на результат трехмерной реконструкции, в том числе с использованием нейросетевой модели NeRF. Согласно результатам исследований, представленным в работе [20*], расширение набора данных с использованием метода XVFI приводит к улучшению качества трехмерной реконструкции.

Таким образом, способ расширения исходного набора данных при помощи метода XVFI лучше всего подходит для видеопоследовательностей с высокой вариативностью поз и выражений лица. При этом интерполяция промежуточных кадров является длительной процедурой и поэтому рассматривается как аугментация, повышающая качество аватара, созданного при помощи предложенного метода. Экспериментальные исследования создания аватара с использованием расширенного набора данных представлены в подразделе 4.4.

4.4. Экспериментальные исследования разработанного метода создания параметризованного аватара головы человека

Для настройки параметров разработанного метода требуется монокулярная видеопоследовательность, содержащая разнообразную мимику и положения головы одного человека. Желательно, чтобы отсутствовал дисбаланс как для набора положений головы, так и для набора эмоций. Видеопоследовательность разбивается на кадры (для получения надежных результатов требуется около тысячи кадров), затем, аналогично предыдущим этапам, для каждого из них производится вычисление ключевых точек и масок сегментации (включая волосы), после чего выполняется оценка параметров параметрической модели головы FLAME. Для расчета функции потерь используются исходные кадры. В рамках исследований используется набор данных, содержащий 10

видеопоследовательностей. Первые 6 видеопоследовательностей опубликованы авторами работы [58] и содержат порядка 3000 кадров каждая. Вторые 2 видеопоследовательности опубликованы авторами работы [47] и содержат порядка 1000 кадров каждая. Последние 2 видеопоследовательности записаны самостоятельно и содержат порядка 800 кадров каждая.

Выбранные открытые наборы данных [58], [47] наиболее часто используются для оценки качества новых методов создания цифрового аватара головы исследователями [47], [49], [57], [58]. Поэтому их использование в рамках экспериментальных исследований позволяет объективно оценить разработанный метод. В связи с тем, что такие наборы состоят из RGB-видеопоследовательностей, для оценки параметров модели FLAME используется алгоритм оценки по RGB изображению.

Процедура оптимизации параметров разработанного метода длится 1000 итераций, количество обучающих примеров за итерацию – 64. Все эксперименты проводятся для синтеза изображений с разрешением 128×128, затем, исходя из полученных качественных и количественных результатов, выбирается лучшая комбинация решений, которая используется для разрешений 256×256 и 512×512.

В рамках исследования метода создания параметризованного аватара головы человека проводятся следующие эксперименты для видеопоследовательностей из набора данных, описанного выше:

1. Донастройка параметров метода из модели, соответствующей эксперименту α_1 (см. перечисление экспериментальных исследований в 3.4).
2. Донастройка параметров метода из модели, соответствующей эксперименту α_2.
3. Донастройка параметров метода из моделей, соответствующих эксперименту $\beta_1 - \beta_6$ для способа обучения α_1.
4. Донастройка параметров метода из моделей, соответствующих эксперименту $\beta_1 - \beta_6$ для способа обучения α_2.

5. Донастройка параметров метода на наборе данных с применением аугментации, которая заключается в интерполяции промежуточных кадров с использованием метода XVFI.

По результатам экспериментальных исследований, соответствующим пунктам 1-4, для каждой видеопоследовательности было получено 14 наборов параметров для синтеза изображений-проекций разрешения 128×128, где 2 из них соответствуют пунктам (1) и (2), 6 из них соответствуют пункту (3), 6 из них соответствуют пункту (4).

На основе проведенных исследований производится двухэтапное сравнение. На первом этапе количественно и качественно сравниваются однотипные эксперименты, в которых параметры разработанного метода инициализируются из α_1 и $\beta_1 - \beta_6$ для способа обучения α_1, и α_2 и $\beta_1 - \beta_6$ для способа обучения α_2. На втором этапе производится независимое качественное сравнение по влиянию следующих настроек обучения параметрической модели головы человека:

1. Добавление аугментации, моделирующей расстояние от камеры до головы по оси z. Попарно сравниваются эксперименты, в которых параметры разработанного метода инициализируются из β_1 и β_2, β_3 и β_5, β_4 и β_6.
2. Целесообразность использования стратегии, при которой вероятность включения синтетических данных на последних итерациях обучения не равна нулю (см. пункты 6, 7 в подразделе 3.4). Сравниваются эксперименты, в которых параметры разработанного метода инициализируются из β_1 и β_3 и β_4, β_2 и β_5 и β_6.
3. Целесообразность использования реальных данных из набора FFHQ в процессе обучения параметрической модели (см. пункт 4 в подразделе 3.4) при использовании её для инициализации параметров разработанного метода. Сравниваются эксперименты, в которых параметры разработанного метода инициализируются из β_1 и α_2.

По результатам второго этапа сравнения формируется и фиксируется набор настроек обучения, показывающих лучший результат для базового запуска

процесса обучения разработанного метода создания параметризованного аватара головы человека (без аугментаций и стратегий).

При сравнении однотипных экспериментов, в которых параметры разработанного метода инициализируются из α_1 и $\beta_1 - \beta_6$ для способа обучения α_1, и α_2 и $\beta_1 - \beta_6$ для способа обучения α_2, для каждой валидационной выборки вручную было отобрано по 30 кадров видеопоследовательности, соответствующих таким моментам, когда челюсть человека частично или полностью открыта и видна внутренняя область рта. Такой выбор обусловлен тем, что исследуемая аугментация нацелена на повышение качества синтеза именно этой области лица.

На рисунках 39, 40, 41, 42 представлены графики, на которых для каждого человека отражены значения показателей качества PSNR (см. формулу (10) в подразделе 3.4), SSIM (см. формулу (11); значение показателя качества SSIM для всего изображения рассчитывается как среднее арифметическое по всем окнам) и LPIPS (см. показатель качества *perceptual loss* в формуле (8); отличие от *perceptual loss* в том, что используется дообученная версия VGG для задачи оценивания семантического сходства [100]) в различных экспериментах. Проанализировав представленные результаты, можно заключить, что приблизительно в 60 % случаев попарного сравнения использование модифицированной текстуры приводит к лучшим результатам. Приблизительно в 10 % случаев результаты сопоставимы.

$$SSIM(x, y) = \frac{(2 \cdot \mu_x \cdot \mu_y + c_1) \cdot (2 \cdot \sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1) \cdot (\sigma_x^2 + \sigma_y^2 + c_2)}, \qquad (11)$$

где x – патч размера $N \times N$ первого изображения,

y – патч размера $N \times N$ второго изображения,

μ_x, μ_y – средние значения пикселей для патчей изображений x, y,

σ_x, σ_y – дисперсии для патчей изображений x, y,

σ_{xy} - ковариационное значения для патчей изображений x и y,

c_1, c_2 – константы.

На рисунках А.14-А.17 в Приложении А представлены результаты синтеза по параметрам, соответствующим отобранным изображениям. На рисунках А.18-А.27, А.28-А.32 в Приложении А представлены результаты синтеза при варьировании

степени открытия челюсти и поворотов шеи (с открытой челюстью). Экспертная оценка данных результатов синтеза говорит об однозначном увеличении качества детализации целевой области человеческого лица при сохранении качества синтеза остальных частей. Наиболее наглядные примеры синтеза по параметрам из валидационных выборок изображены на рисунке 43.

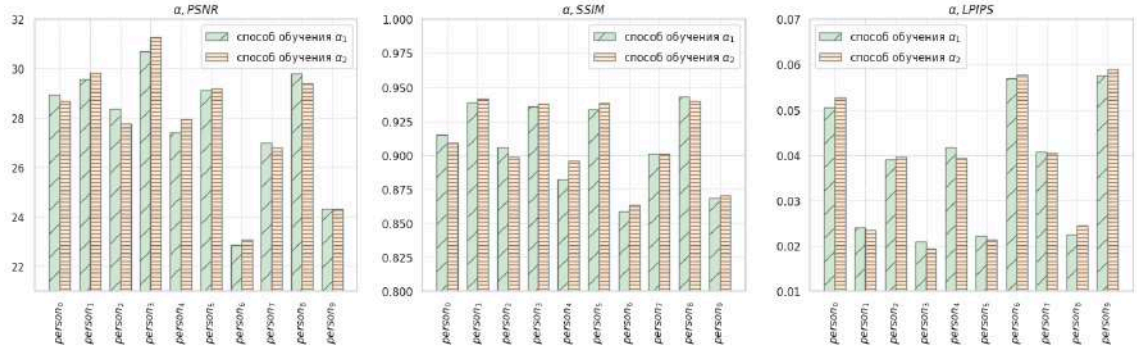


Рисунок 39 – Значения показателей качества PSNR, SSIM, LPIPS для экспериментов, в ходе которых производилась донастройка параметров разработанного метода при инициализации из версий параметрической модели, соответствующих экспериментам α_1 и α_2

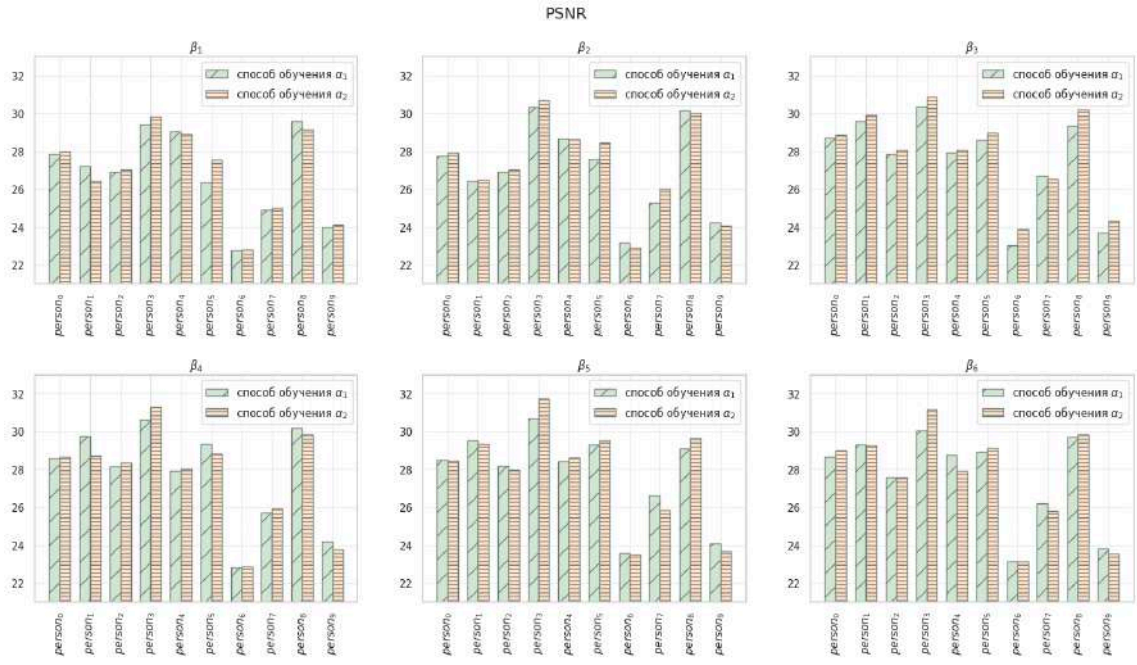


Рисунок 40 – Значения показателя качества PSNR для экспериментов, в ходе которых производилась донастройка параметров разработанного метода при инициализации из версий параметрической модели, соответствующих экспериментам β_1–β_6

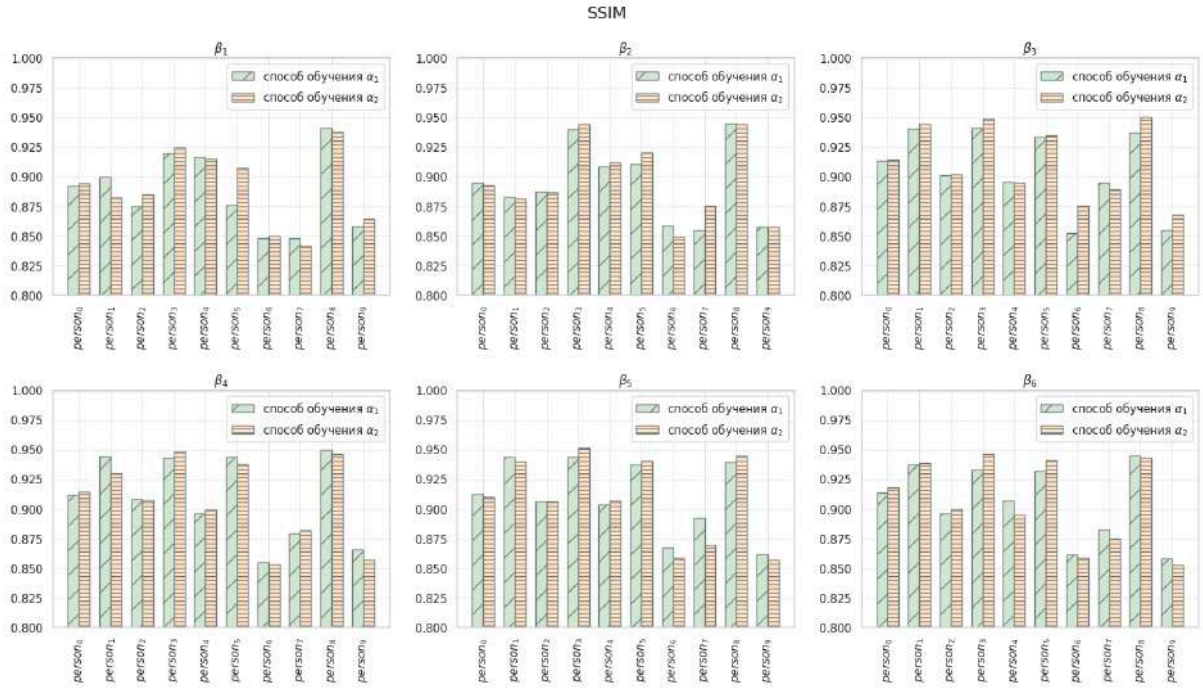


Рисунок 41 – Значения показателя качества SSIM для экспериментов, в ходе которых производилась донастройка параметров разработанного метода при инициализации из версий параметрической модели, соответствующих экспериментам β_1–β_6

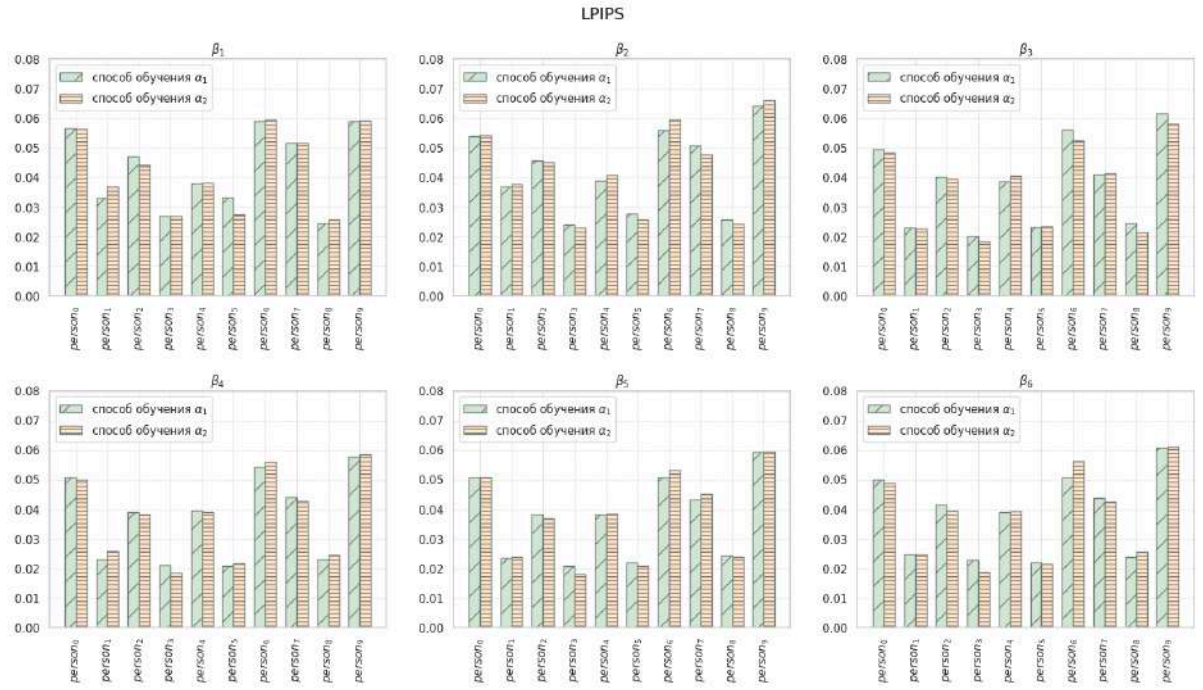


Рисунок 42 – Значения показателя качества LPIPS для экспериментов, в ходе которых производилась донастройка параметров разработанного метода при инициализации из версий параметрической модели, соответствующих экспериментам β_1–β_6

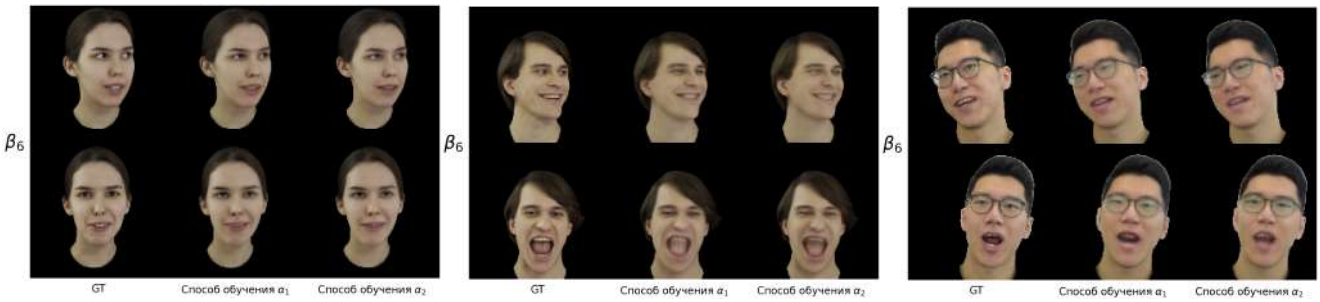


Рисунок 43 – Результаты синтеза по параметрам, соответствующим отобранным из валидационных выборок изображений

На втором этапе исследования сравниваются только эксперименты, параметры которых соответствуют способу обучения α_2. Выбор обуславливается результатами первого этапа исследования.

Для оценки влияния аугментации, симулирующей изменение расстояния от камеры до головы по оси z, производится качественное сравнение соответствующих экспериментов, для которых была выполнена инициализация из параметров моделей, соответствующих экспериментам β_1 и β_2, β_3 и β_5, β_4 и β_6. Для этого выполняется синтез изображений-проекций головы с разным удалением от камеры. Важно отметить, что в обучающей выборке расстояние от головы до камеры изменяется незначительно; при исследовании для настройки параметров разработанного метода аугментация не применяется. На рисунках А.33-А.37 в Приложении А представлены результаты синтеза новых изображений-проекций. Исходя из представленных результатов, можно заключить, что при выполнении оптимизации для разработанного метода из параметров, полученных с применением исследуемой аугментации, идентичность и выражение лица более стабильны при варьировании удаленности от камеры. Наиболее наглядные примеры синтеза изображений-проекций с разным удалением от камеры изображены на рисунке 44.

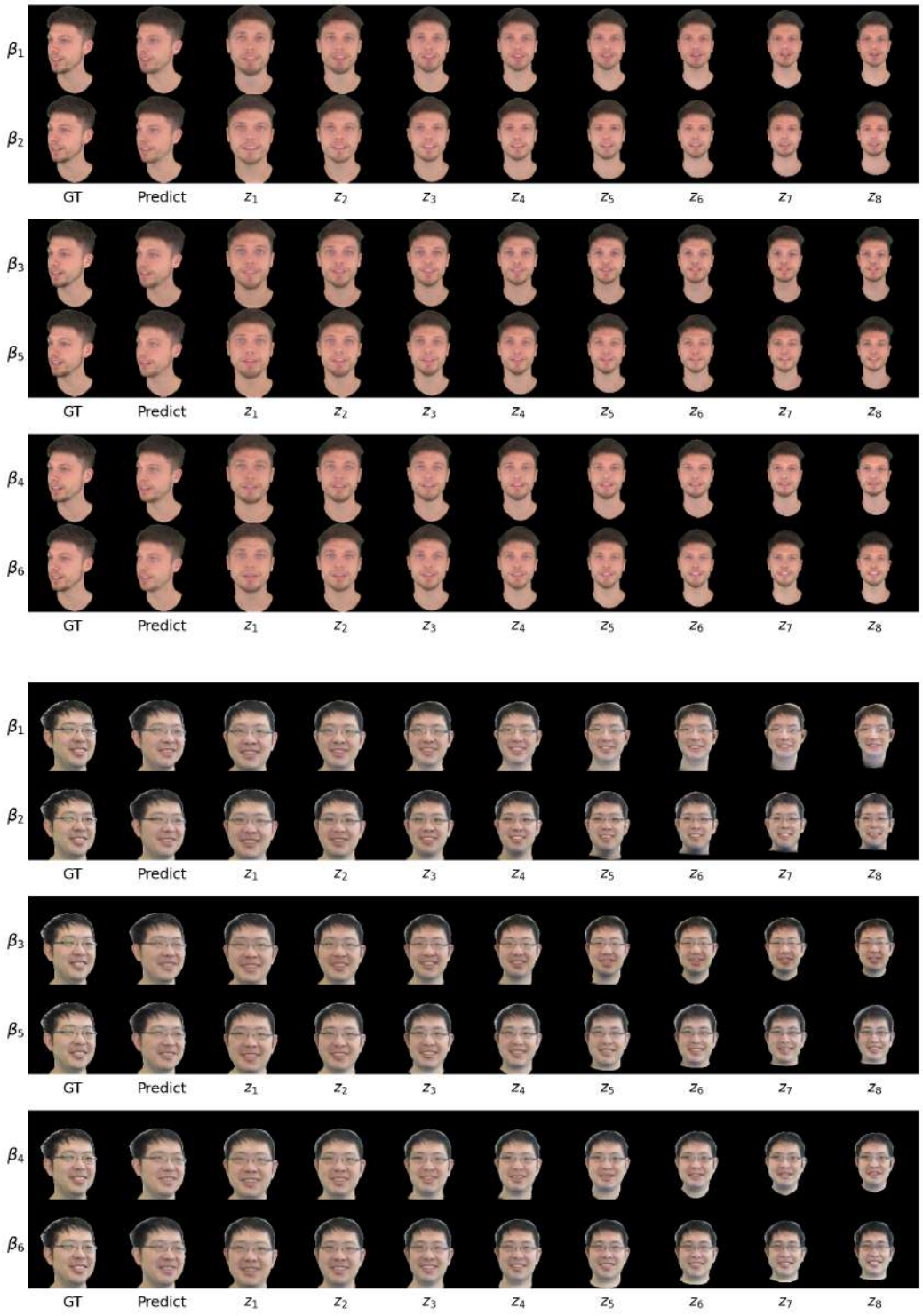


Рисунок 44 – Результаты синтеза в зависимости от разного удаления головы от камеры. Верхняя строка каждого блока – результат синтеза изображений-проекций, где инициализация производится из параметров, полученных без применения исследуемой аугментации; нижняя строка каждого блока – результат синтеза изображений-проекций, где инициализация производится из параметров, полученных с применением исследуемой аугментации

Для оценки целесообразности использования стратегии, при которой вероятность включения синтетических данных на последних итерациях обучения не равна нулю, производится качественное сравнение соответствующих экспериментов, в которых параметры разработанного метода инициализируются из β_1 и β_3 и β_4, β_2 и β_5 и β_6. Для этого выполняется синтез изображений-проекций с варьированием выражения лица в допустимом диапазоне, а также синтез изображений-проекций на основе параметров модели FLAME, полученных в ходе процедуры оценки параметров для изображений валидационной выборки. Важно отметить, что при исследовании для настройки параметров разработанного метода аугментация не применяется. На рисунках А.38-А.42 в Приложении А представлены результаты синтеза новых изображений-проекций с варьированием выражения лица. На рисунках А.43-А.47 в Приложении А представлены результаты синтеза новых изображений-проекций на основе параметров модели FLAME для валидационной выборки. Мотивация введения такой стратегии заключается в том, что обучающая выборка, вероятно, не может охватить весь диапазон выражений лица, а также при преобладании большого количества векторов для выражения лица в малой окрестности внутри допустимого диапазона существует риск переобучения.

Исходя из представленных результатов, можно заключить, что при выполнении оптимизации параметров разработанного метода из параметров, полученных с использованием исследуемой стратегии, синтез изображений для выражений лица, не участвовавших в процессе обучения, значительно реалистичнее с физиологической точки зрения. Наиболее наглядные примеры синтеза при варьировании выражения лица в допустимом диапазоне изображены на рисунке 45.

Анализируя рисунки А.43-А.47 в Приложении А, можно заключить, что при оптимизации параметров разработанного метода, инициализированных из параметров, соответствующих экспериментам β_4 и β_6 (использование стратегии, где вероятность включения синтетических данных линейно убывает с ростом количества итераций оптимизации), итоговый результат наиболее схож с истинным

изображением, как по текстуре, так и по выражению лица. Наиболее наглядные примеры синтеза на основе параметров модели FLAME валидационной выборки изображены на рисунке 46.

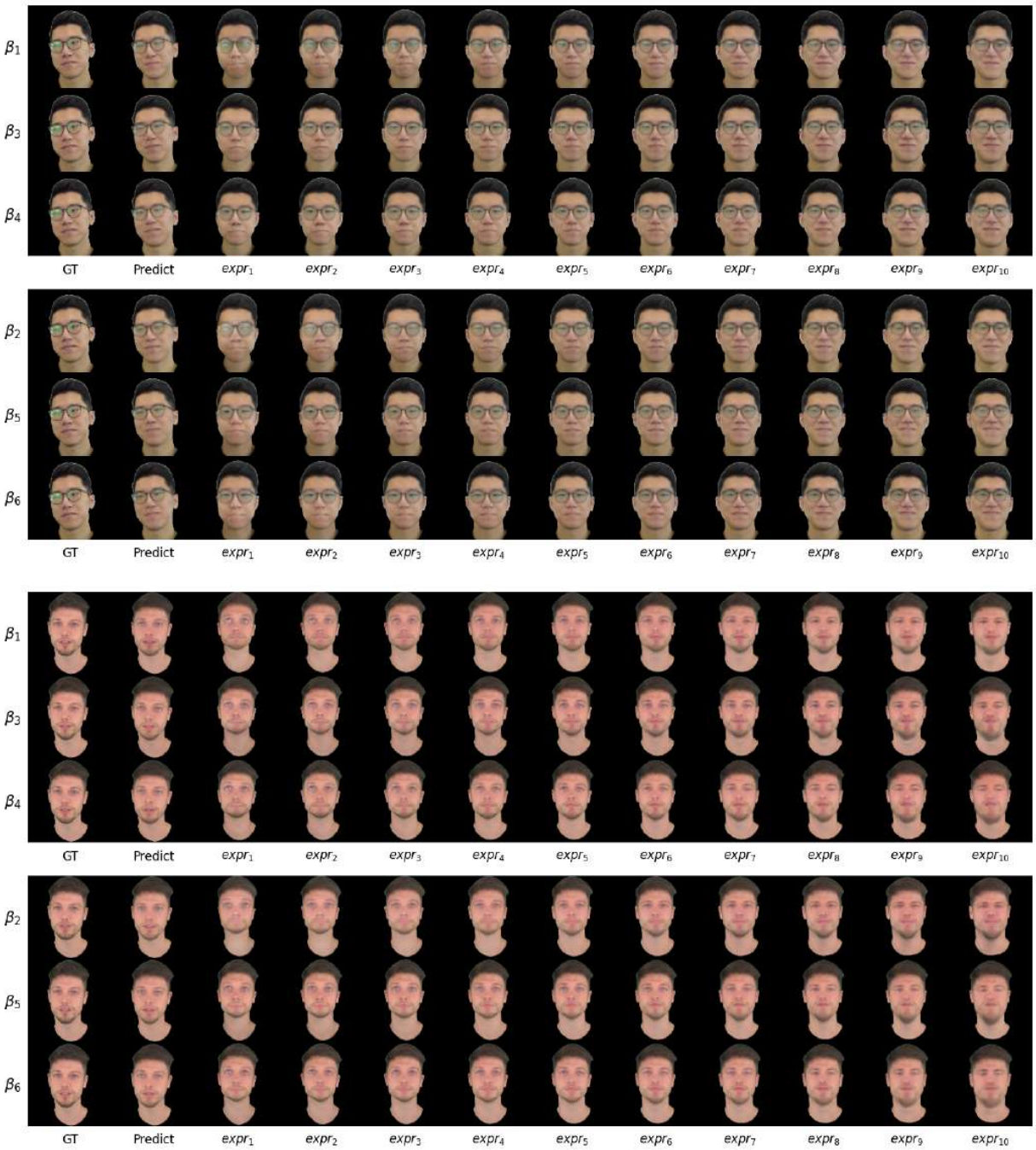


Рисунок 45 – Результаты синтеза при варьировании выражения лица в допустимом диапазоне. Первая строка каждого блока – результат синтеза изображений-проекций, где инициализация производится из параметров, полученных без использования исследуемой стратегии; вторая и третья строка каждого блока – результат синтеза изображений-проекций, где инициализация производится из параметров, полученных с использованием исследуемой стратегии

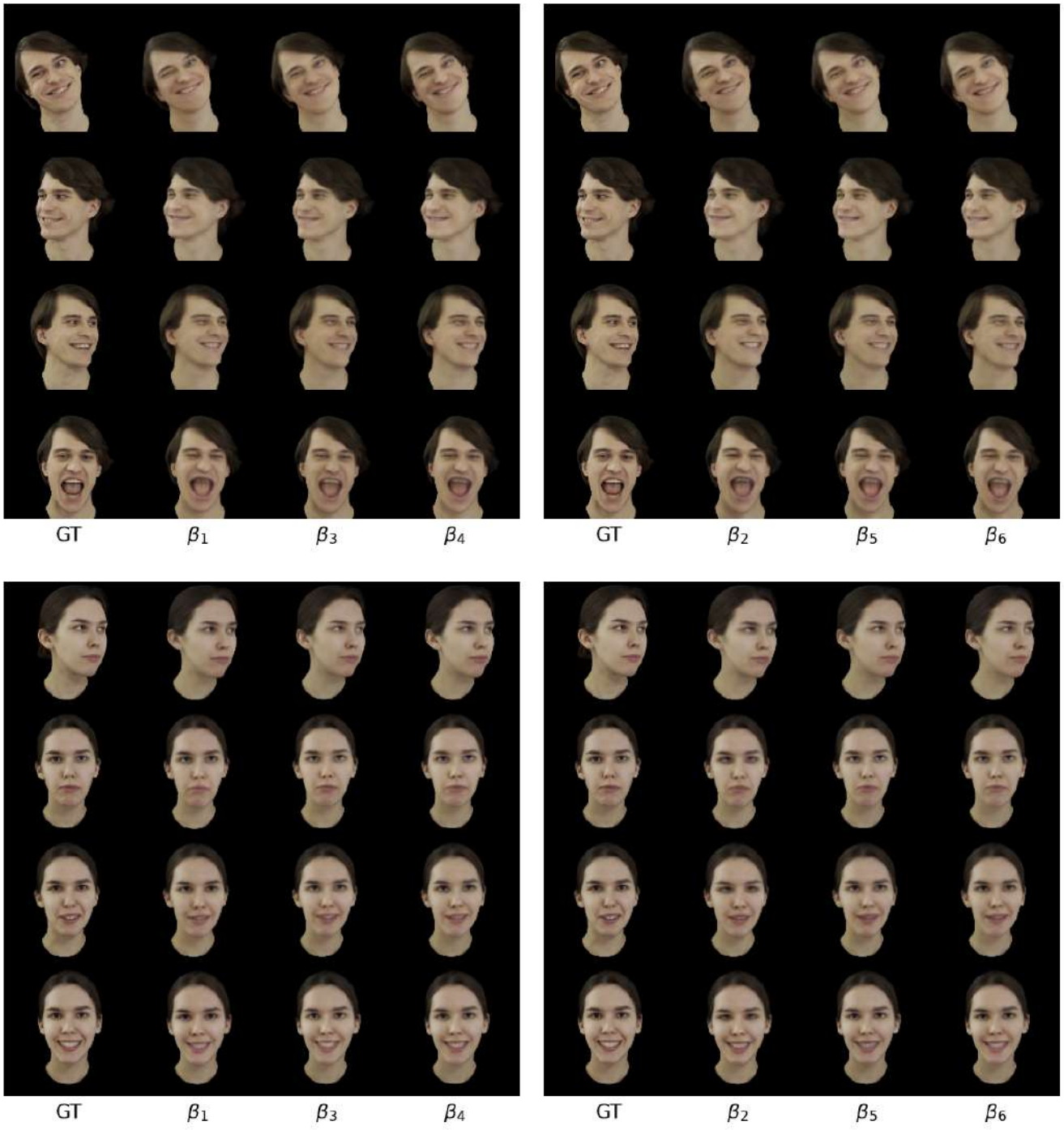


Рисунок 46 – Результаты синтеза на основе параметров модели FLAME валидационной выборки. Первый столбец – результат синтеза изображений-проекций, где инициализации производится из параметров β_1, β_3, β_4. Второй столбец – результат синтеза изображений-проекций, где инициализация производится из параметров β_2, β_5, β_6

Для оценки целесообразности использования реальных данных из набора FFHQ в процессе обучения параметрической модели, при использовании ее для

инициализации параметров разработанного метода, производится качественное сравнение соответствующих экспериментов. Сравниваются эксперименты, в которых параметры разработанного метода инициализируются из β_1 и α_2. Для этого выполняется синтез изображений-проекций на основе параметров модели FLAME, полученных в ходе процедуры оценки параметров для изображений валидационной выборки. На рисунках А.48-А.50 в Приложении А представлены результаты синтеза на основе параметров модели FLAME валидационной выборки.

Исходя из представленных результатов, можно заключить, что при оптимизации параметров разработанного метода, инициализированных из параметров, соответствующих эксперименту β_1, итоговый результат получается более фотореалистичным, выражения лица имеют большее сходство с исходными изображениями. Наиболее наглядные примеры синтеза на основе параметров модели FLAME валидационной выборки изображены на рисунке 47.

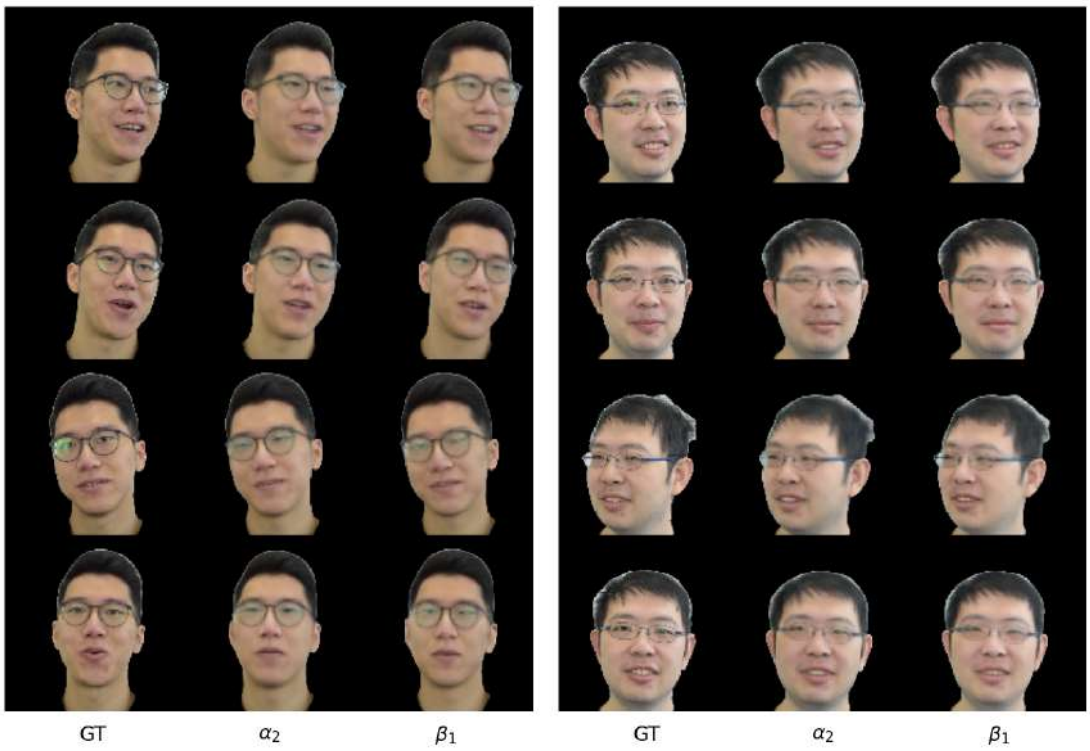


Рисунок 47 – Результаты синтеза по параметрам, соответствующим отобранным из валидационных выборок изображений, для экспериментов, инициализация параметров которых производилась из β_1 и α_2

Таким образом, исходя из представленных результатов проведенных экспериментальных исследований разработанного метода создания параметризованного аватара головы человека, итоговый набор настроек обучения выглядит следующим образом:

1. Инициализация параметров разработанного метода для последующей оптимизации из параметров, соответствующих эксперименту β_6 для способа обучения α_2.
2. Аугментация: симуляция изменения расстояния от камеры до головы по оси z для изображений из обучающей выборки в процессе оптимизации параметров для аватара.
3. Использование стратегии, при которой вероятность включения синтетических данных в обучающую выборку линейно убывает с ростом количества итераций при оптимизации параметров для аватара.

На рисунках А.51-А.83 представлены результаты синтеза разработанного метода создания аватара головы человека, оптимизация параметров которого производится по сформированному в результате исследования набору настроек с разрешениями 128×128, 256×256 и 512×512. Наиболее показательные примеры результатов синтеза метода представлены на рисунках 48-52. А именно: на рисунке 48 представлены результаты синтеза в зависимости от разного удаления головы от камеры по оси z; на рисунке 49 представлены результаты синтеза при варьировании параметров выражения лица в допустимом диапазоне; на рисунке 50 представлены результаты синтеза при варьировании поворота шеи; на рисунке 51 представлены результаты синтеза новых видов; на рисунке 52 представлен результат процедуры переноса выражения лица с кадров видеопоследовательностей.

Рисунок 48 – Результаты синтеза в зависимости от разного удаления головы от камеры для разрешения 512×512

Рисунок 49 – Результаты синтеза при варьировании параметров выражения лица в допустимом диапазоне для разрешения 512×512

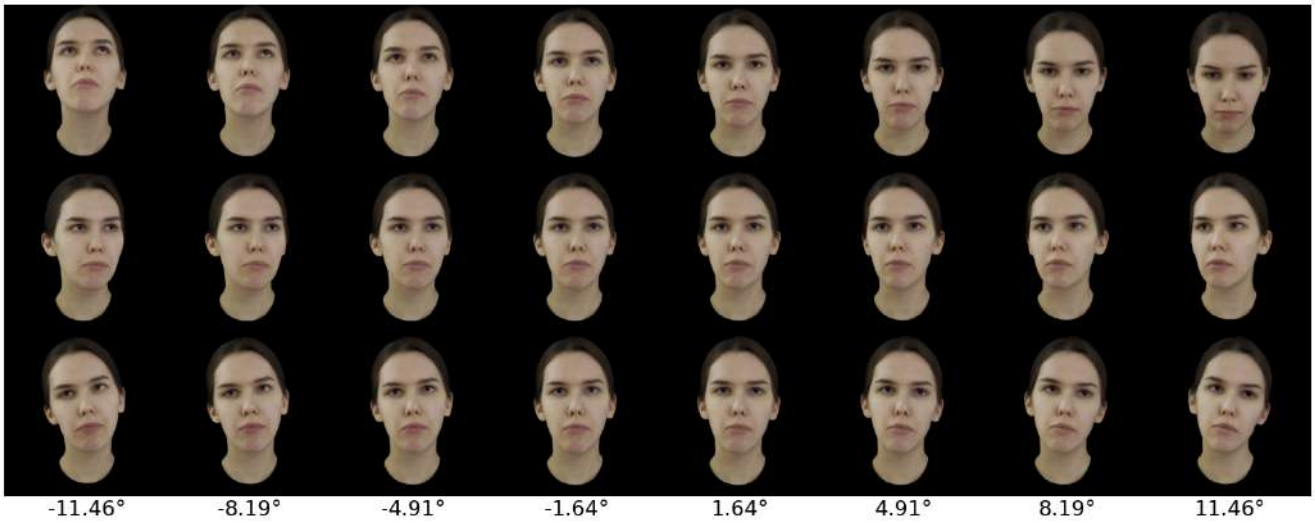


Рисунок 50 – Результаты синтеза при варьировании поворота шеи для разрешения 512×512

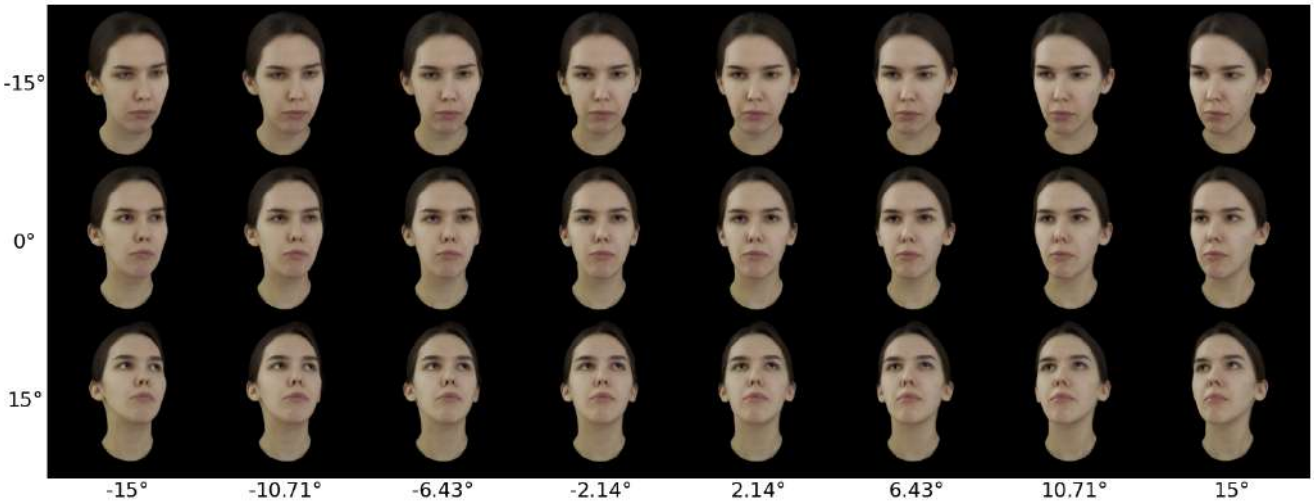


Рисунок 51 – Результат синтеза новых видов для разрешения 512×512

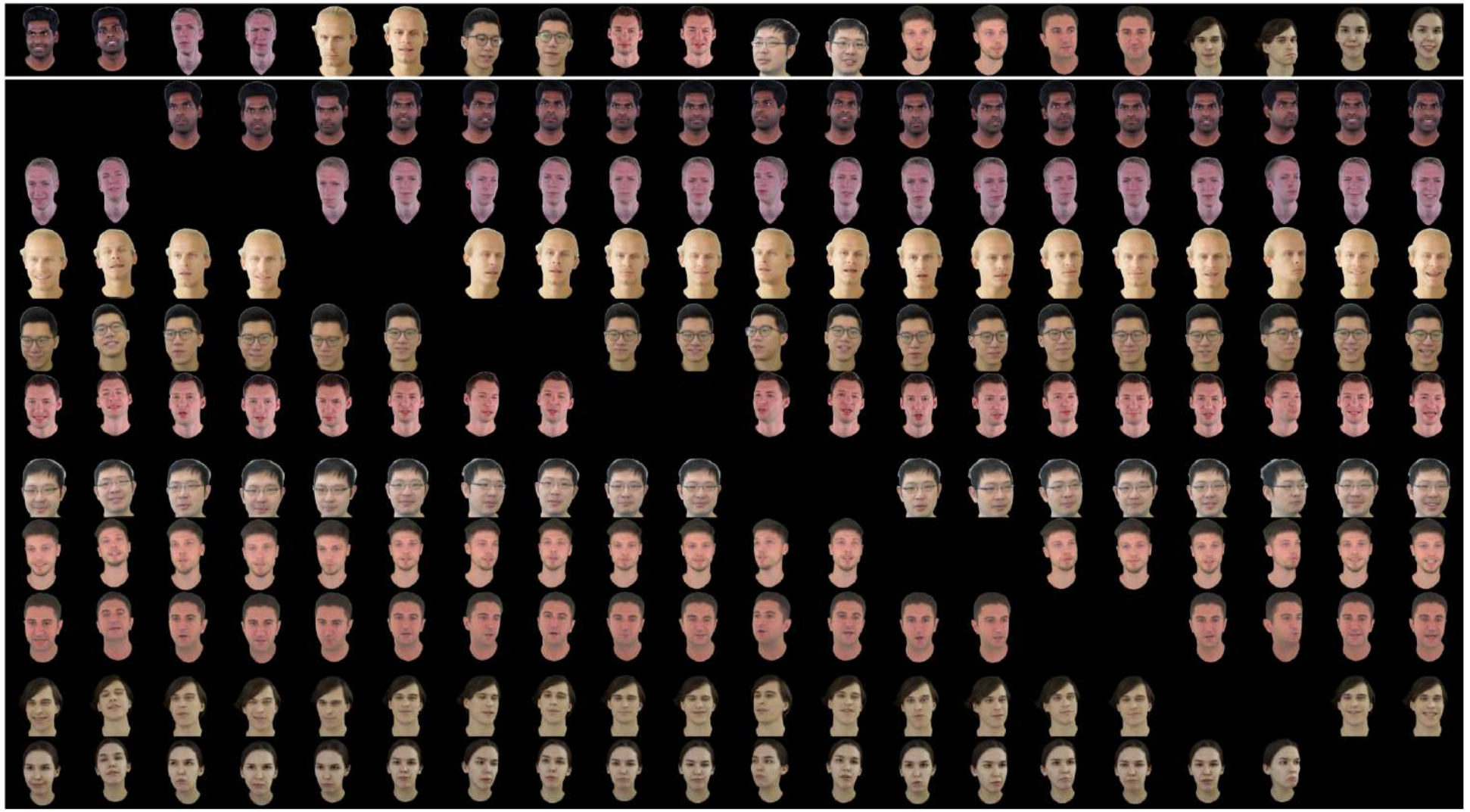


Рисунок 52 – Результат процедуры переноса выражения лица с кадров видеопоследовательностей для разрешения 512×512

Для проведения экспериментальных исследований, соответствующих пункту (5), каждая видеопоследовательность была представлена в трех вариантах, а именно: оригинальная с исходным количеством кадров (a); полученная из (a) путем удаления каждого второго кадра (b); с количеством кадров, как для (a), но полученная из (b) путем применения метода XVFI для восстановления промежуточных кадров. На рисунке 53 представлен результат интерполяции кадров методом XVFI для динамичных сцен. Оптимизация параметров предложенного метода выполнялась по наилучшей стратегии, определенной в результате экспериментальных исследований, описанных выше. В результате для каждой видеопоследовательности было получено три набора параметров для синтеза изображений разрешения 128×128. Для количественной оценки были отобраны видеопоследовательности из сформированного набора данных, взятые из открытых источников [47], [58]. На рисунке 54 представлены средние значения показателей качества PSNR, SSIM, LPIPS, вычисленные для кадров из тестовых выборок. Исходя из полученных результатов, можно заключить, что использование видеопоследовательностей с интерполированными

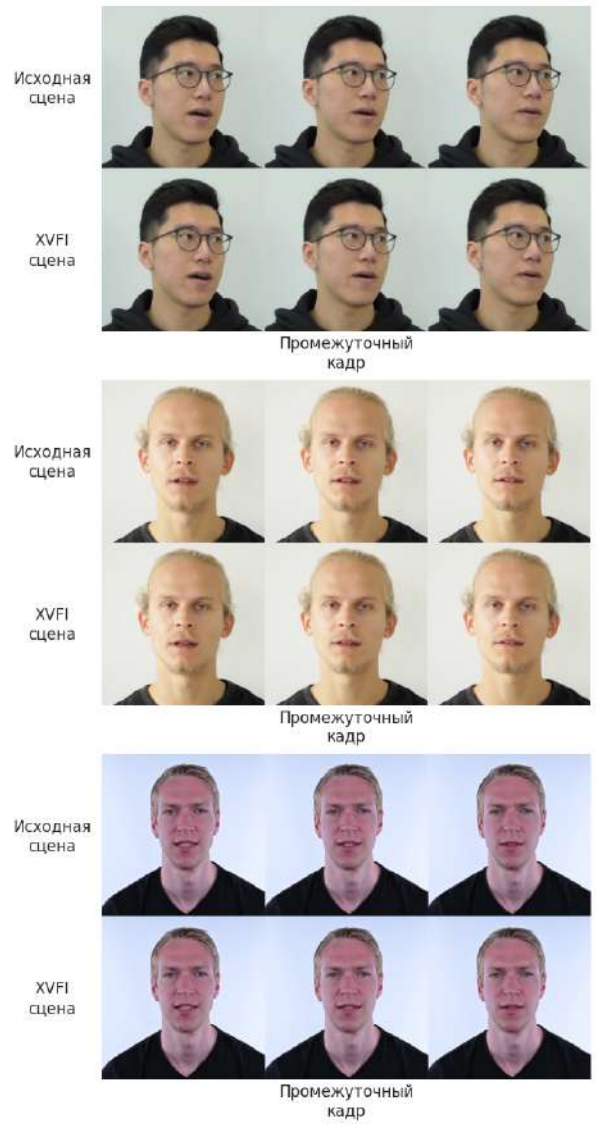


Рисунок 53 – Результат интерполяции кадров для некоторых динамичных сцен. Промежуточный кадр XVFI сцены является интерполированным

промежуточными кадрами при оптимизации параметров разработанного метода создания параметризованного аватара головы человека приводит к улучшению качества синтезируемых изображений-проекций. Так, например, среднее значение показателя качества PSNR по отношению к тестовой выборке увеличивается на 0,17 дБ. В соответствии с показателями качества SSIM и LPIPS результаты синтеза при оптимизации параметров метода по исходным видеопоследовательностям и видеопоследовательностям с интерполированными кадрами сопоставимые.

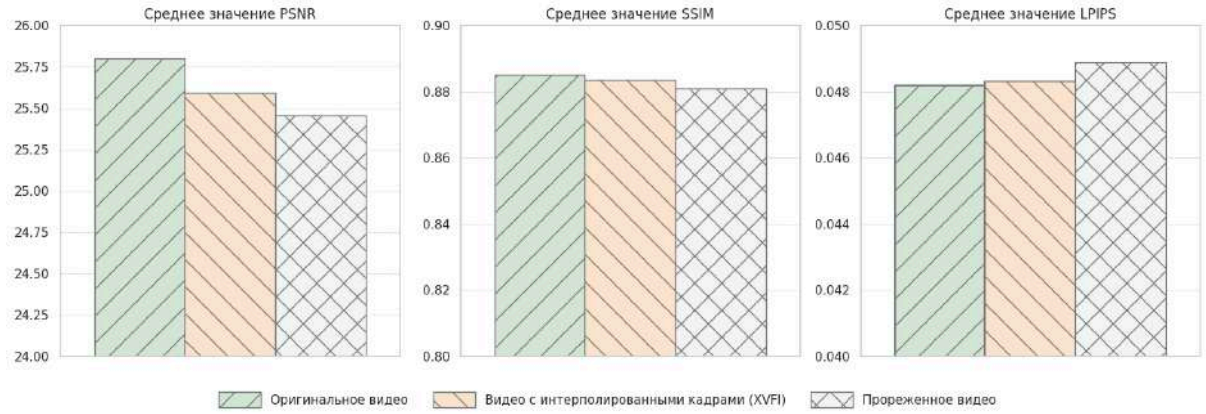


Рисунок 54 – Средние значения показателей качества PSNR, SSIM, LPIPS для кадров из тестовых выборок для трех наборов параметров

4.5. Сравнение предложенного метода с существующими решениями

В данном подразделе приведены результаты сравнения предложенного метода создания аватара головы человека с актуальными существующими решениями, позиционируемыми авторами как state-of-the-art в исследуемой области, а именно Neural Head Avatar, IMAvatar, PointAvatar, INSTA, подробно описанными в разделе 1.

Сравнение методов производится количественно и качественно по тестовым выборкам для данных, опубликованных авторами работы [58]. Набор данных содержит видеопоследовательности для шести человек. Количество тестовых кадров для каждой видеопоследовательности составляет 350. Для всех методов размер синтезированных изображений-проекций составляет 256×256. На рисунке 55 представлен результат синтеза изображений-проекций для экземпляров данных из тестовых выборок. На рисунке 56 представлен результат процедуры переноса

выражения лица. В таблице 2 представлены значения показателей качества PSNR, SSIM и LPIPS (perceptual loss с использованием дообученной сети VGG [100]) между изображениями, синтезированными с использованием параметров из тестовых выборок, и соответствующими истинными изображениями. На рисунке 57 представлена зависимость качества синтеза изображений-проекций аватаров для тестовых выборок от времени, затраченного на обучение, для всех методов. На рисунке 58 представлена зависимость качества синтеза изображений-проекций аватаров для тестовых выборок от среднего времени, затраченного на синтез одного кадра, для всех методов.

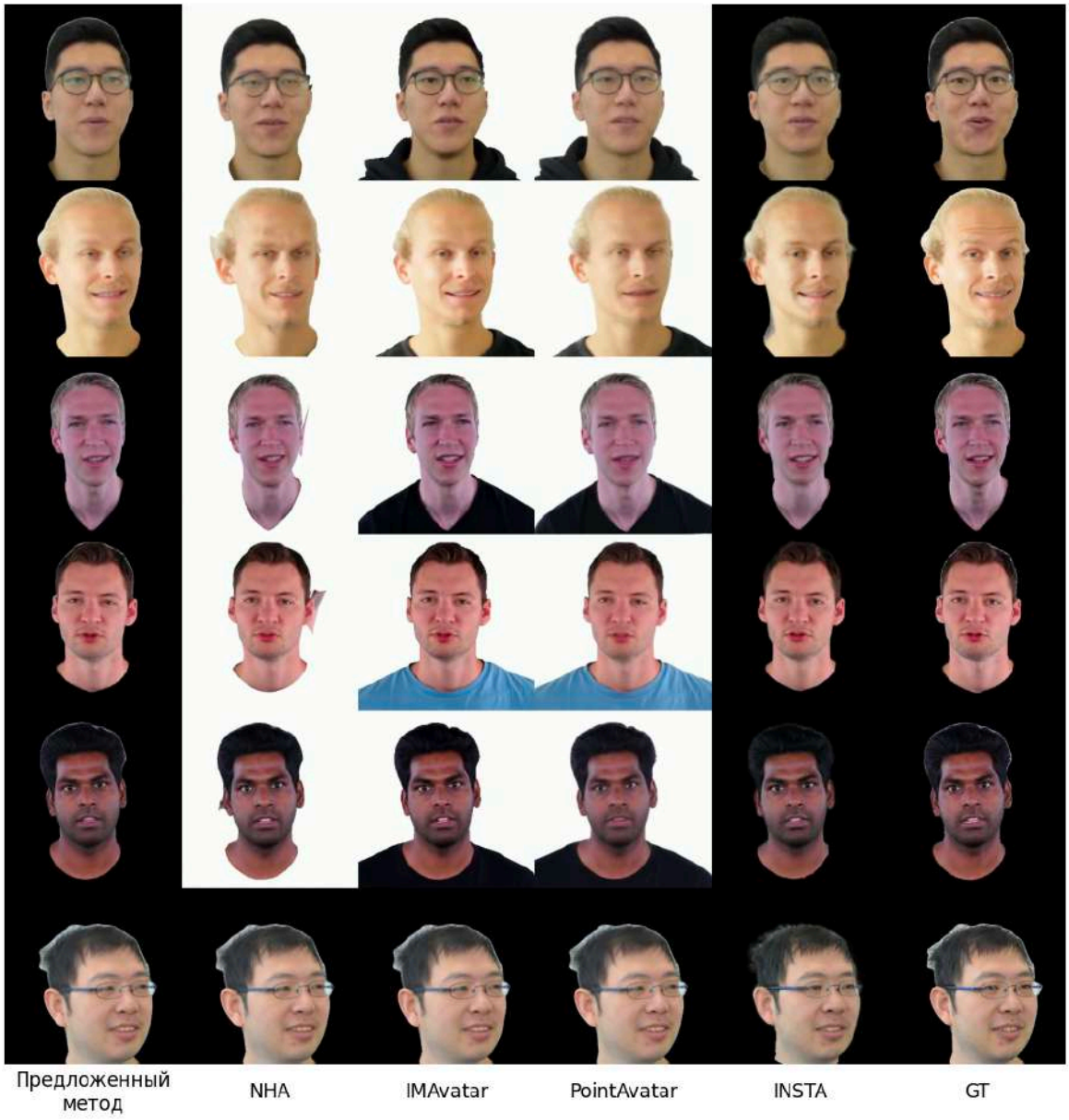


Рисунок 55 – Результат синтеза изображений-проекций для экземпляров данных из тестовых выборок

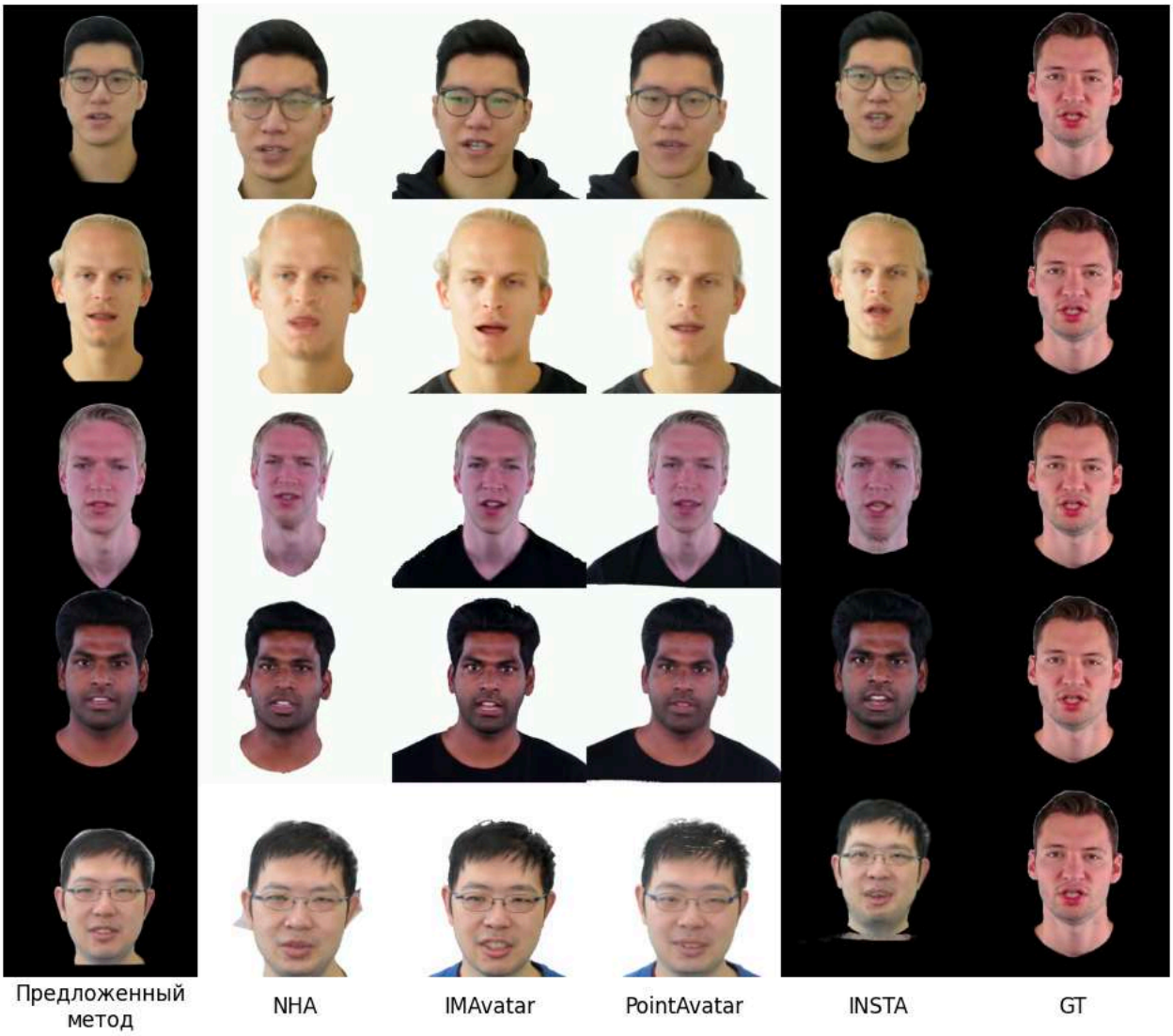


Рисунок 56 – Результат процедуры переноса выражения лица. Правый столбец – исходное выражение лица, переносимое на аватаров

Таблица 2 – Значения показателей качества PSNR, SSIM, LPIPS между изображениями, синтезированными с использованием параметров из тестовых выборок, и соответствующими истинными изображениями

	PSNR ↑	SSIM ↑	LPIPS ↓
NHA [47]	25,1049	0,8080	0,0789
IMAvatar [49]	25,8890	**0,9123**	0,0771
PointAvatar [57]	26,2529	0,8862	0,0631
INSTA [58]	25,4665	0,9098	0,1066
Предложенный метод	**26,2867**	0,8952	**0,0461**

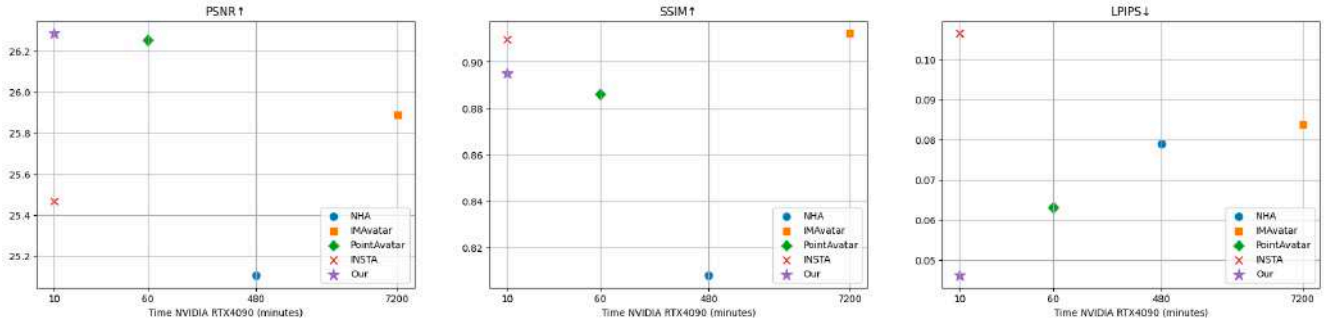


Рисунок 57 – Зависимость качества синтеза изображений-проекций аватаров для тестовых выборок от времени, затраченного на обучение, для всех методов

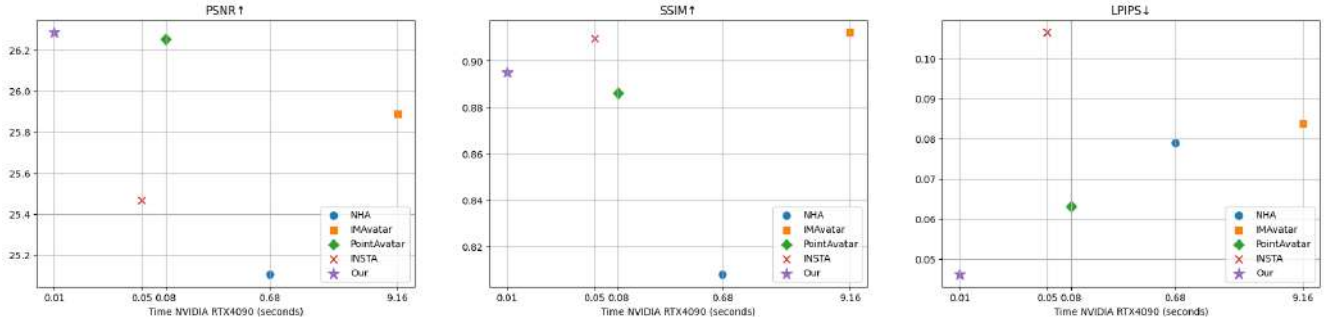


Рисунок 58 – Зависимость качества синтеза изображений-проекций аватаров для тестовых выборок от среднего времени, затраченного на генерацию одного кадра, для всех методов

Исходя из представленных количественных результатов следует, что предложенный метод превосходит существующие по показателям качества PSNR и LPIPS на выбранном наборе данных. Визуальный анализ синтезированных кадров на рисунке 55 подтверждает высокое качество работы метода. Анализируя результаты других методов, можно заметить, что аватары, созданные с помощью метода NHA, подвержены возникновению артефактов. При выполнении процедуры переноса выражения лица в методе INSTA происходит деформация аватара, что можно заметить на рисунке 56. Для аватаров, созданных с использованием метода PointAvatar, на некоторых ракурсах наблюдается чрезмерная разреженность облака точек, что приводит к появлению пустот, как продемонстрировано на рисунке 59. Важно отметить, что создание аватара с использованием предложенного метода занимает порядка 10 минут на персональном компьютере с характеристиками Intel Core i9 14900K, NVIDIA RTX4090 24GB, 64GB DDR5 RAM. На том же устройстве создание аватара методом NHA занимает порядка 8 часов, методом IMAvatar от 3

до 7 дней, методом PointAvatar порядка 1 часа, методом INSTA аналогично порядка 10 минут. При этом необходимо отметить, что при разработке метода INSTA учтены особенности специализированных вычислителей NVIDIA, что делает невозможным использование метода на вычислителях другого типа.

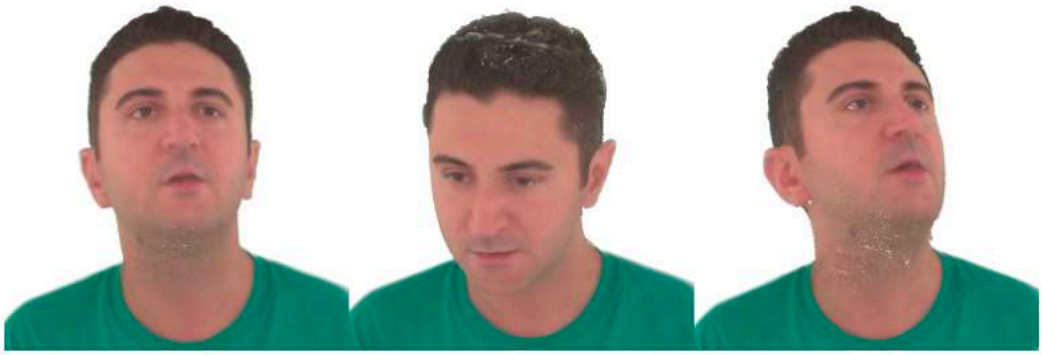


Рисунок 59 – Пример синтеза изображений с артефактом в виде наличия пустот для аватара, полученного методом PointAvatar

4.6. Выводы и результаты четвертого раздела

В данном разделе было представлено описание разработанного метода создания параметризованного аватара головы человека, приведены возможные его приложения, предложен способ аугментации (расширения) исходного набора данных за счет применения процедуры интерполяции промежуточных кадров. Также приведены описание порядка проведения экспериментальных исследований разработанного метода и полученные результаты, представлено сравнение метода с актуальными существующими решениями.

По результатам экспериментальных исследований разработанного метода был определен итоговый набор настроек для оптимизации его параметров. Для получения конечного представления аватара головы человека требуется небольшая видеопоследовательность (порядка 1000 кадров), захватывающая мимику человека, и тысяча итераций оптимизации. На персональном компьютере с характеристиками Intel Core i9 14900K, NVIDIA RTX4090 24GB, 64GB DDR5 RAM тысяча итераций оптимизации занимает порядка 10 минут. Для полученного итогового представления можно выполнять процедуру синтеза новых видов, а также

выполнять перенос выражения лица и положения головы с новых изображений (в том числе с другими людьми).

В случае ограниченности обучающей выборки в контексте количества кадров есть возможность ее расширить путем интерполяции промежуточных кадров. Как показали результаты экспериментальных исследований, такой способ расширения набора данных приводит к улучшению качества синтезируемых изображений-проекций. Однако, важно отметить, что, если исходная выборка является недостаточно репрезентативной в смысле разнообразия мимики и поз, итоговый аватар по-прежнему будет страдать недостаточной способностью к экстраполяции выражений лица и поз. Так, предложенная опция является актуальной для видеопоследовательностей с широким разбросом мимики и поз, но с небольшой частотой кадров.

Результаты сравнения разработанного метода с существующими показали, что предложенный подход превосходит существующие по показателям качества PSNR и LPIPS на наборе данных, опубликованном авторами работы [58]. Разработанный метод превосходит в среднем рассмотренные методы по значению PSNR на 0,6 дБ; превосходит лучший метод из рассмотренных в сравнении по времени, необходимому на синтез одного изображения-проекции, на 84%; находится на одном уровне с лучшим методом из рассмотренных по времени, необходимому для получения представления аватара. Разработанный метод позволяет синтезировать изображения со средним значением PSNR 26,29 дБ. Визуальный анализ результатов синтеза изображений-проекций подтверждает высокое качество работы метода.

Таким образом, можно сделать вывод, что разработанный метод имеет высокий потенциал для применения его в прикладных системах, в том числе нацеленных на исполнение функции телеприсутствия. По теме раздела опубликованы работы [17*–20*], [23*], [24*].

ЗАКЛЮЧЕНИЕ

В рамках диссертационной работы разработан и исследован новый метод создания параметризованного аватара головы человека на основе нейросетевой модели рендеринга, обеспечивающий при относительно низких вычислительных затратах процесса создания повышенное качество синтезируемых изображений-проекций.

Основные результаты диссертационной работы:

1. Алгоритм оценки параметров параметрической модели головы FLAME с использованием RGBD изображения. Предложенный алгоритм достигает точности трехмерной реконструкции равной 0,2 мм.
2. Параметрическая модель головы человека на основе нейросетевой модели представления поверхности CNeRF, архитектуры сети двумерного нейронного рендеринга с блоком повышения пространственной дискретизации, позволяющим ускорить сходимость метода создания аватара (более чем в 2 раза), и синтетического набора данных, генерируемого в реальном времени. Разработанная модель позволяет синтезировать изображения-проекции со средним значением PSNR более 30 дБ по отношению к изображениям, полученным с использованием параметрической модели головы FLAME.
3. Метод создания аватара головы человека на основе разработанной параметризованной модели головы человека. Разработанный метод позволяет синтезировать изображения-проекции со средним значением PSNR 26,29 дБ по отношению к изображениям из тестовой выборки.
4. Способ аугментации (расширения) реального набора данных с изображениями головы человека с помощью интерполяции промежуточных кадров видеопоследовательности. Способ позволяет повысить качество (среднее значение показателя качества PSNR по отношению к тестовой выборке увеличивается на 0,17 дБ) синтезируемых изображений-проекций аватара для коротких видеопоследовательностей.

5. Результаты экспериментальных исследований разработанного метода создания параметризованного аватара, включающие сравнение разработанного метода с существующими актуальными (state-of-the-art) решениями. Разработанный метод превосходит в среднем рассмотренные методы по значению PSNR на 0,6 дБ; превосходит лучший метод из рассмотренных по времени, необходимому на синтез одного изображения-проекции, на 84%; находится на одном уровне с лучшим методом из рассмотренных по затратам времени, необходимым для создания аватара.

Важно отметить, что предложенный метод имеет высокую значимость вне зависимости от конкретных деталей реализации его этапов. В предложенном методе возможно заменить: способ неявного представления головы человека, способ обуславливания, в том числе параметрическую модель головы, и архитектурное решение для синтеза итогового изображения-проекции. Данный метод демонстрирует особую важность методов переноса обучения.

Основные результаты работы отражены в 8 публикациях. Результаты и положения диссертации представлены на 3 научных конференциях.

СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ

BFM – Basel Face Model,

CNeRF – Conditional Neural Radiance Fields, условные нейронные поля излучения,

DECA – Detailed Expression Capture and Animation,

FFHQ - Flickr-Faces-HQ,

FLAME – Faces Learned with an Articulated Model and Expressions,

Head avatar – аватар головы,

LBS – Linear Blend Skinning, веса линейно-переходного кожного покрова,

NeRF – Neural Radiance Fields, нейронные поля излучения,

Neural rendering – нейронный рендеринг,

PCA – Principal Component Analysis, метод главных компонент,

RGB – Red, Green, Blue,

RGBD – Red, Green, Blue, Depth,

SDF – Signed Distance Fields (Function).

СПИСОК ЛИТЕРАТУРЫ

1. Hamad, A. How Virtual Reality Technology Has Changed Our Lives: An Overview of the Current and Potential Applications and Limitations / A. Hamad, B. Jia // International Journal of Environmental Research and Public Health. — 2022. — Vol. 19(18). — P. 11278. — DOI: 10.3390/ijerph191811278.
2. Wenger, A. Performance relighting and reflectance transformation with time-multiplexed illumination / A. Wenger, A. Gardner, C. Tchou, J. Unger, T. Hawkins, P. Debevec // ACM Trans. Graph. — 2005. — Vol. 24(3). — P. 756–764. — DOI: 10.1145/1073204.1073258.
3. Guo, K. The relightables: volumetric performance capture of humans with realistic relighting / K. Guo, P. Lincoln, P. Davidson, J. Busch, X. Yu, M. Whalen, G. Harvey, S. Orts-Escolano, R. Pandey, J. Dourgarian, D. Tang, A. Tkach, A. Kowdle, E. Cooper, M. Dou, S. Fanello, G. Fyffe, C. Rhemann, J. Taylor, P. Debevec, S. Izadi // ACM Transactions on Graphics. — 2019. — Vol. 38(6). — P. 1–19. — DOI: 10.1145/3355089.3356571.
4. Wuu, C. Multiface: A Dataset for Neural Face Rendering / C. Wuu, N. Zheng, S. Ardisson, R. Bali, D. Belko, E. Brockmeyer, L. Evans, T. Godisart, H. Ha, X. Huang, A. Hypes, T. Koska, S. Krenn, S. Lombardi, X. Luo, K. McPhail, L. Millerschoen, M. Perdoch, M. Pitts, A. Richard, J. Saragih, J. Saragih, T. Shiratori, T. Simon, M. Stewart, A. Trimble, X. Weng, D. Whitewolf, C. Wu, S.-I. Yu, Y. Sheikh // arXiv. — 2023. — DOI: 10.48550/arXiv.2207.11243.
5. Beeler, T. High-quality single-shot capture of facial geometry / T. Beeler, B. Bickel, P. Beardsley, B. Sumner, M. Gross // ACM Trans. Graph. — 2010. — Vol. 29(4). — P. 40:1-40:9. — DOI: 10.1145/1778765.1778777.
6. Wang, T.-C. One-Shot Free-View Neural Talking-Head Synthesis for Video Conferencing / T.-C. Wang, A. Mallya, M.-Y. Liu // 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) — 2021. — P. 10034–10044. — DOI: 10.1109/CVPR46437.2021.00991.
7. Apple Vision Pro [Электронный ресурс] // Wikipedia. — [2024]. — Режим

доступа: https://en.wikipedia.org/w/index.php?title=Apple_Vision_Pro&oldid=1238603940 (дата обращения: 12.08.2024).

8. Capture and edit your Persona (beta) on Apple Vision Pro [Электронный ресурс] // Apple Support. — Режим доступа: https://support.apple.com/ru-ru/guide/apple-vision-pro/dev934d40a17/visionos (дата обращения: 12.08.2024).
9. arXiv.org e-Print archive [Электронный ресурс] — Режим доступа: https://arxiv.org/ (дата обращения: 18.04.2024).
10. Max Planck Institutes and Experts [Электронный ресурс] — Режим доступа: https://www.mpg.de/institutes (дата обращения: 19.04.2024).
11. Stanford University [Электронный ресурс] / ©Copyright Stanford University Stanford, California 94305 // Stanford University. — [2024]. — Режим доступа: https://www.stanford.edu/ (дата обращения: 19.04.2024).
12. Home [Электронный ресурс] // University of California, Berkeley. — Режим доступа: https://www.berkeley.edu/ (дата обращения: 19.04.2024).
13. Homepage - CMU - Carnegie Mellon University [Электронный ресурс] / C.M. University — Режим доступа: http://www.cmu.edu/index.html (дата обращения: 19.04.2024).
14. ETH Zurich - Homepage [Электронный ресурс]. — [2024]. — Режим доступа: https://ethz.ch/en.html (дата обращения: 19.04.2024).
15. University of Oxford [Электронный ресурс] — Режим доступа: https://www.ox.ac.uk/ (дата обращения: 19.04.2024).
16. Shanghai University [Электронный ресурс] — Режим доступа: https://en.shu.edu.cn/ (дата обращения: 19.04.2024).
17. Ганеева*, Ю.Х. Сравнение методов реконструкции промежуточных кадров видеопоследовательностей с динамической сценой / Ю.Х. Ганеева // Информационные технологии и нанотехнологии. Сборник трудов по материалам VIII Международной конференции и молодежной школы. Изд-во Самарского Университета — Самара, 2022. — С. 31472.

18. Ganeeva, Y. Comparison of methods for reconstructing intermediate video frames with a dynamic scene / Y. Ganeeva // IEEE Xplore 2022 VIII International Conference on Information Technology and Nanotechnology — 2022. — P. 1–5. — DOI: 10.1109/ITNT55410.2022.9848739.
19. Ганеева, Ю.Х. Влияние реконструкции промежуточных кадров видеопоследовательности на результат 3D-реконструкции объектов / Ю.Х. Ганеева, В.В. Мясников // Информационные технологии и нанотехнологии. Сборник трудов по материалам VIII Международной конференции и молодежной школы. Изд-во Самарского Университета — Самара, 2022. — С. 31492.
20. Ganeeva, Y. The impact of intermediate video frames reconstruction step on the result of 3D reconstruction of objects / Y. Ganeeva, V. Myasnikov // IEEE Xplore 2022 VIII International Conference on Information Technology and Nanotechnology — 2022. — P. 1–5. — DOI: 10.1109/ITNT55410.2022.9848697.
21. Козлова, Ю.Х. Метод реконструкции и анимации модели головы с использованием RGBD-изображения / Ю.Х. Козлова // Информационные технологии и нанотехнологии. Сборник трудов по материалам IX Международной конференции и молодежной школы. Изд-во Самарского Университета — Самара, 2023. — С. 30612.
22. Kozlova, Y.K. Head model reconstruction and animation method using color image with depth information / Y.K. Kozlova, V.V. Myasnikov // Computer Optics. — 2024. — Vol. 48. P. 118–122. — DOI: 10.18287/2412-6179-CO-1334.
23. Козлова, Ю.Х. Метод создания анимируемых аватаров с использованием нейронных полей излучения и двухмерного нейронного рендеринга / Ю.Х. Козлова // Информационные технологии и нанотехнологии. Сборник трудов по материалам X Международной конференции и молодежной школы. Изд-во Самарского Университета — Самара, 2024.
24. Kozlova, Yu.Kh. Method for Creating Animatable Avatars Using Neural Radiance Fields and Two-Dimensional Neural Rendering / Yu.Kh. Kozlova // IEEE Xplore 2024 X International Conference on Information Technology and Nanotechnology

— 2024. — P. 1–8. — DOI: 10.1109/ITNT60778.2024.10582377.

25. Mildenhall, B. NeRF: representing scenes as neural radiance fields for view synthesis / B. Mildenhall, P.P. Srinivasan, M. Tancik, J.T. Barron, R. Ramamoorthi, R. Ng // Commun. ACM. — 2021. — Vol. 65(1). — P. 99–106. — DOI: 10.1145/3503250.
26. Lorensen, W.E. Marching cubes: A high resolution 3D surface construction algorithm / W.E. Lorensen, H.E. Cline // SIGGRAPH Comput. Graph. — 1987. — Vol. 21(4). — P. 163–169. — DOI: 10.1145/37402.37422.
27. Yang, H. FaceScape: A Large-Scale High Quality 3D Face Dataset and Detailed Riggable 3D Face Prediction / H. Yang, H. Zhu, Y. Wang, M. Huang, Q. Shen, R. Yang, X. Cao // 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) — 2020. — P. 598–607. — DOI: 10.1109/CVPR42600.2020.00068.
28. Chen Cao FaceWarehouse: A 3D Facial Expression Database for Visual Computing / Chen Cao, Yanlin Weng, Shun Zhou, Yiying Tong, Kun Zhou // IEEE Transactions on Visualization and Computer Graphics. — 2014. — Vol. 20(3). — P. 413–425. — DOI: 10.1109/TVCG.2013.249.
29. Li, T. Learning a model of facial shape and expression from 4D scans / T. Li, T. Bolkart, M.J. Black, H. Li, J. Romero // ACM Transactions on Graphics. — 2017. — Vol. 36(6). — P. 1–17. — DOI: 10.1145/3130800.3130813.
30. Dai, H. Statistical Modeling of Craniofacial Shape and Texture / H. Dai, N. Pears, W. Smith, C. Duncan // International Journal of Computer Vision. — 2020. — Vol. 128(2). — P. 547–571. — DOI: 10.1007/s11263-019-01260-7.
31. Karras, T. A Style-Based Generator Architecture for Generative Adversarial Networks / T. Karras, S. Laine, T. Aila // IEEE Trans. Pattern Anal. Mach. Intell. — 2021. — Vol. 43(12). — P. 4217–4228. — DOI: 10.1109/TPAMI.2020.2970919.
32. www.makehumancommunity.org [Электронный ресурс] — Режим доступа: http://www.makehumancommunity.org/ (дата обращения: 28.02.2024).

33. Jiang, W. NeuMan: Neural Human Radiance Field from a Single Video / W. Jiang, K.M. Yi, G. Samei, O. Tuzel, A. Ranjan // Computer Vision – ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXII — 2022. — P. 402–418. — DOI: 10.1007/978-3-031-19824-3_24.
34. Jiang, T. InstantAvatar: Learning Avatars from Monocular Video in 60 Seconds / T. Jiang, X. Chen, J. Song, O. Hilliges // 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) — 2023. — P. 16922–16932. — DOI: 10.1109/CVPR52729.2023.01623.
35. Chen, X. SNARF: Differentiable Forward Skinning for Animating Non-Rigid Neural Implicit Shapes / X. Chen, Y. Zheng, M.J. Black, O. Hilliges, A. Geiger // 2021 IEEE/CVF International Conference on Computer Vision (ICCV) — 2021. — P. 11574–11584. — DOI: 10.1109/ICCV48922.2021.01139.
36. Paysan, P. A 3D Face Model for Pose and Illumination Invariant Face Recognition / P. Paysan, R. Knothe, B. Amberg, S. Romdhani, T. Vetter // 2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance. — 2009. — P. 296–301. — DOI: 10.1109/AVSS.2009.58.
37. Averbuch-Elor, H. Bringing portraits to life / H. Averbuch-Elor, D. Cohen-Or, J. Kopf, M.F. Cohen // ACM Transactions on Graphics. — 2017. — Vol. 36(6). — P. 1–13. — DOI: 10.1145/3130800.3130818.
38. Wang, T.-C. High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs / T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, B. Catanzaro // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition — 2018. — P. 8798–8807. — DOI: 10.1109/CVPR.2018.00917.
39. Ghosh, P. GIF: Generative Interpretable Faces / P. Ghosh, P.S. Gupta, R. Uziel, A. Ranjan, M.J. Black, T. Bolkart // 2020 International Conference on 3D Vision (3DV). — 2020. — P. 868–878. — DOI: 10.1109/3DV50981.2020.00097.
40. Karras, T. Analyzing and Improving the Image Quality of StyleGAN / T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, T. Aila // 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) — 2020. — P. 8107–8116. — DOI: 10.1109/CVPR42600.2020.00813.

41. Siarohin, A. First Order Motion Model for Image Animation / A. Siarohin, S. Lathuilière, S. Tulyakov, E. Ricci, N. Sebe // Advances in Neural Information Processing Systems, Vol. 32 — 2019.
42. Yenamandra, T. i3DMM: Deep Implicit 3D Morphable Model of Human Heads / T. Yenamandra, A. Tewari, F. Bernard, H.-P. Seidel, M. Elgharib, D. Cremers, C. Theobalt // 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) — 2021. — P. 12798–12808. — DOI: 10.1109/CVPR46437.2021.01261.
43. Zakharov, E. Fast Bi-Layer Neural Synthesis of One-Shot Realistic Head Avatars / E. Zakharov, A. Ivakhnenko, A. Shysheya, V. Lempitsky // Computer Vision – ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XII — 2020. — P. 524–540. — DOI: 10.1007/978-3-030-58610-2_31.
44. Buhler, M.C. VariTex: Variational Neural Face Textures / M.C. Buhler, A. Meka, G. Li, T. Beeler, O. Hilliges // 2021 IEEE/CVF International Conference on Computer Vision (ICCV) — 2021. — P. 13870–13879. — DOI: 10.1109/ICCV48922.2021.01363.
45. Feng, Y. Learning an animatable detailed 3D face model from in-the-wild images / Y. Feng, H. Feng, M.J. Black, T. Bolkart // ACM Transactions on Graphics. — 2021. — Vol. 40(4). — P. 1–13. — DOI: 10.1145/3450626.3459936.
46. Gafni, G. Dynamic Neural Radiance Fields for Monocular 4D Facial Avatar Reconstruction / G. Gafni, J. Thies, M. Zollhofer, M. Niesner // 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) — 2021. — P. 8645–8654. — DOI: 10.1109/CVPR46437.2021.00854.
47. Grassal, P.-W. Neural Head Avatars from Monocular RGB Videos / P.-W. Grassal, M. Prinzler, T. Leistner, C. Rother, M. Niebner, J. Thies // 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) — 2022. — P. 18632–18643. — DOI: 10.1109/CVPR52688.2022.01810.
48. Hong, Y. HeadNeRF: A Realtime NeRF-based Parametric Head Model / Y. Hong, B. Peng, H. Xiao, L. Liu, J. Zhang // 2022 IEEE/CVF Conference on Computer

Vision and Pattern Recognition (CVPR) — 2022. — P. 20342–20352. — DOI: 10.1109/CVPR52688.2022.01973.

49. Zheng, Y. I M Avatar: Implicit Morphable Head Avatars from Videos / Y. Zheng, V.F. Abrevaya, M.C. Buhler, X. Chen, M.J. Black, O. Hilliges // 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) — 2022. — P. 13535–13545. — DOI: 10.1109/CVPR52688.2022.01318.
50. К. Ма Трехмерное глубокое обучение на Python / К. Ма, В. Хегде, Л. Йольан. — Москва: ДМК Пресс, 2023. — 226 с.
51. Zhuang, Y. MoFaNeRF: Morphable Facial Neural Radiance Field / Y. Zhuang, H. Zhu, X. Sun, X. Cao // Computer Vision – ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part III — 2022. — P. 268–285. — DOI: 10.1007/978-3-031-20062-5_16.
52. Galanakis, S. FitDiff: Robust monocular 3D facial shape and reflectance estimation using Diffusion Models / S. Galanakis, A. Lattas, S. Moschoglou, S. Zafeiriou // ArXiv. — 2023. — DOI: 10.48550/ARXIV.2312.04465.
53. Kabadayi, B. GAN-Avatar: Controllable Personalized GAN-based Human Head Avatar / B. Kabadayi, W. Zielonka, B.L. Bhatnagar, G. Pons-Moll, J. Thies // 2024 International Conference on 3D Vision (3DV) — 2024. — P. 882–892. — DOI: 10.1109/3DV62453.2024.00058.
54. Chan, E.R. Efficient Geometry-aware 3D Generative Adversarial Networks / E.R. Chan, C.Z. Lin, M.A. Chan, K. Nagano, B. Pan, S. De Mello, O. Gallo, L. Guibas, J. Tremblay, S. Khamis, T. Karras, G. Wetzstein // 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) — 2022. — P. 16102–16112. — DOI: 10.1109/CVPR52688.2022.01565.
55. Deng, Y. Accurate 3D Face Reconstruction With Weakly-Supervised Learning: From Single Image to Image Set / Y. Deng, J. Yang, S. Xu, D. Chen, Y. Jia, X. Tong // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) — 2019. — P. 285–295. — DOI: 10.1109/CVPRW.2019.00038.
56. Papantoniou, F.P. Relightify: Relightable 3D Faces from a Single Image via

Diffusion Models / F.P. Papantoniou, A. Lattas, S. Moschoglou, S. Zafeiriou // 2023 IEEE/CVF International Conference on Computer Vision (ICCV) — 2023. — P. 8772–8783. — DOI: 10.1109/ICCV51070.2023.00809.

57. Zheng, Y. PointAvatar: Deformable Point-Based Head Avatars from Videos / Y. Zheng, W. Yifan, G. Wetzstein, M.J. Black, O. Hilliges // 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) — 2023. — P. 21057–21067. — DOI: 10.1109/CVPR52729.2023.02017.
58. Zielonka, W. Instant Volumetric Head Avatars / W. Zielonka, T. Bolkart, J. Thies // 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) — 2023. — P. 4574–4584. — DOI: 10.1109/CVPR52729.2023.00444.
59. Müller, T. Instant neural graphics primitives with a multiresolution hash encoding / T. Müller, A. Evans, C. Schied, A. Keller // ACM Transactions on Graphics. — 2022. — Vol. 41(4). — P. 1–15. — DOI: 10.1145/3528223.3530127.
60. Kerbl, B. 3D Gaussian Splatting for Real-Time Radiance Field Rendering / B. Kerbl, G. Kopanas, T. Leimkuehler, G. Drettakis // ACM Transactions on Graphics. — 2023. — Vol. 42(4). — P. 1–14. — DOI: 10.1145/3592433.
61. Xu, Y. Gaussian Head Avatar: Ultra High-fidelity Head Avatar via Dynamic Gaussians / Y. Xu, B. Chen, Z. Li, H. Zhang, L. Wang, Z. Zheng, Y. Liu // ArXiv. — 2024. — DOI: 10.48550/arXiv.2312.03029.
62. Qian, S. GaussianAvatars: Photorealistic Head Avatars with Rigged 3D Gaussians / S. Qian, T. Kirschstein, L. Schoneveld, D. Davoli, S. Giebenhain, M. Niessner // ArXiv. — DOI: 10.48550/arXiv.2312.02069.
63. Niemeyer, M. RadSplat: Radiance Field-Informed Gaussian Splatting for Robust Real-Time Rendering with 900+ FPS / M. Niemeyer, F. Manhardt, M.-J. Rakotosaona, M. Oechsle, D. Duckworth, R. Gosula, K. Tateno, J. Bates, D. Kaeser, F. Tombari // ArXiv. — 2024. — DOI: 10.48550/arXiv.2403.13806.
64. Rojas, S. Re-ReND: Real-time Rendering of NeRFs across Devices / S. Rojas, J. Zarzar, J.C. Pérez, A. Sanakoyeu, A. Thabet, A. Pumarola, B. Ghanem // 2023 IEEE/CVF International Conference on Computer Vision (ICCV) — 2023. — P. 3609–3618. — DOI: 10.1109/ICCV51070.2023.00336.

65. Loper, M. SMPL: a skinned multi-person linear model / M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, M.J. Black // ACM Trans. Graph. — 2015. — Vol. 34(6). — P. 248:1-248:16. — DOI: 10.1145/2816795.2818013.
66. HavenFeng/photometric_optimization [Электронный ресурс]. — [2024]. — Режим доступа: https://github.com/HavenFeng/photometric_optimization (дата обращения: 06.03.2024).
67. Kingma, D.P. Adam: A Method for Stochastic Optimization / D.P. Kingma, J. Ba // ArXiv. — 2017. — DOI: 10.48550/arXiv.1412.6980.
68. Bulat, A. How far are we from solving the 2D & 3D Face Alignment problem? (and a dataset of 230,000 3D facial landmarks) / A. Bulat, G. Tzimiropoulos // 2017 IEEE International Conference on Computer Vision (ICCV) — 2017. — P. 1021–1030. — DOI: 10.1109/ICCV.2017.116.
69. zllrunning/face-parsing.PyTorch [Электронный ресурс]. — [2024]. — Режим доступа: https://github.com/zllrunning/face-parsing.PyTorch (дата обращения: 11.03.2024).
70. Thies, J. Face2Face: real-time face capture and reenactment of RGB videos / J. Thies, M. Zollhöfer, M. Stamminger, C. Theobalt, M. Nießner // Commun. ACM. — 2018. — Vol. 62(1). — P. 96–104. — DOI: 10.1145/3292039.
71. 1adrianb/face-alignment [Электронный ресурс]. — [2024]. — Режим доступа: https://github.com/1adrianb/face-alignment (дата обращения: 12.03.2024).
72. Gu, J. StyleNeRF: A Style-based 3D-Aware Generator for High-resolution Image Synthesis / J. Gu, L. Liu, P. Wang, C. Theobalt // ArXiv. — 2021. — DOI: 10.48550/arXiv.2110.08985.
73. Niemeyer, M. GIRAFFE: Representing Scenes as Compositional Generative Neural Feature Fields / M. Niemeyer, A. Geiger // 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) — 2021. — P. 11448–11459. — DOI: 10.1109/CVPR46437.2021.01129.
74. Hu, J. Squeeze-and-Excitation Networks / J. Hu, L. Shen, S. Albanie, G. Sun, E. Wu // IEEE Trans. Pattern Anal. Mach. Intell. — 2020. — Vol. 42(8). — P. 2011–2023. — DOI: 10.1109/TPAMI.2019.2913372.

75. Shi, W. Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network / W. Shi, J. Caballero, F. Huszar, J. Totz, A.P. Aitken, R. Bishop, D. Rueckert, Z. Wang // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) — 2016. — P. 1874–1883. — DOI: 10.1109/CVPR.2016.207.
76. Zhang, R. Making Convolutional Networks Shift-Invariant Again / R. Zhang // Proceedings of the 36th International Conference on Machine Learning — 2019. — P. 7324–7334.
77. Johnson, J. Perceptual Losses for Real-Time Style Transfer and Super-Resolution / J. Johnson, A. Alahi, L. Fei-Fei // ArXiv. — 2016. — DOI: 10.48550/arXiv.1603.08155.
78. Simonyan, K. Very deep convolutional networks for large-scale image recognition / K. Simonyan, A. Zisserman // 3rd International Conference on Learning Representations (ICLR 2015). — 2015. — P. 1–14.
79. Redmon, J. You Only Look Once: Unified, Real-Time Object Detection / J. Redmon, S. Divvala, R. Girshick, A. Farhadi // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) — 2016. — P. 779–788. — DOI: 10.1109/CVPR.2016.91.
80. Zhuang, F. A Comprehensive Survey on Transfer Learning / F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, Q. He // Proceedings of the IEEE. — 2020. — Vol. PP. — P. 1–34. — DOI: 10.1109/JPROC.2020.3004555.
81. COCO - Common Objects in Context [Электронный ресурс] — Режим доступа: https://cocodataset.org/#home (дата обращения: 26.03.2024).
82. The PASCAL Visual Object Classes Homepage [Электронный ресурс] — Режим доступа: http://host.robots.ox.ac.uk/pascal/VOC/ (дата обращения: 26.03.2024).
83. Open Images V7 [Электронный ресурс] — Режим доступа: https://storage.googleapis.com/openimages/web/index.html (дата обращения: 26.03.2024).
84. Nikolenko, S.I. Synthetic Data for Deep Learning / S.I. Nikolenko // ArXiv. —

2019. — DOI: 10.48550/arXiv.1909.11512.

85. blender.org - Home of the Blender project - Free and Open 3D Creation Software [Электронный ресурс] — Режим доступа: https://www.blender.org/ (дата обращения: 26.03.2024).
86. Платформа Unity для разработки в реальном времени | Движок для 3D, 2D, VR и AR [Электронный ресурс] — Режим доступа: https://unity.com/ru (дата обращения: 26.03.2024).
87. CelebA Dataset [Электронный ресурс] — Режим доступа: https://mmlab.ie.cuhk.edu.hk/projects/CelebA.html (дата обращения: 26.03.2024).
88. NVlabs/ffhq-dataset [Электронный ресурс]. — [2024]. — Режим доступа: https://github.com/NVlabs/ffhq-dataset (дата обращения: 26.03.2024).
89. NVlabs/metfaces-dataset [Электронный ресурс]. — [2024]. — Режим доступа: https://github.com/NVlabs/metfaces-dataset (дата обращения: 26.03.2024).
90. Welcome to Python.org [Электронный ресурс] // Python.org. — [2024]. — Режим доступа: https://www.python.org/ (дата обращения: 09.04.2024).
91. PyTorch [Электронный ресурс] // PyTorch. — Режим доступа: https://pytorch.org/ (дата обращения: 09.04.2024).
92. PyTorch3D · A library for deep learning with 3D data [Электронный ресурс] — Режим доступа: https://pytorch3d.org/ (дата обращения: 09.04.2024).
93. Home [Electronic resource] // OpenCV. — Mode of access: https://opencv.org/ (accessed date: 09.04.2024).
94. NumPy documentation — NumPy v1.26 Manual [Электронный ресурс] — Режим доступа: https://numpy.org/doc/stable/index.html (дата обращения: 09.04.2024).
95. Sim, H. XVFI: eXtreme Video Frame Interpolation / H. Sim, J. Oh, M. Kim — 2021. — P. 14469–14478. — DOI: 10.1109/ICCV48922.2021.01422.
96. Li, H. Video Frame Interpolation Via Residue Refinement / H. Li, Y. Yuan, Q. Wang // ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) — 2020. — C. 2613–2617. — DOI:

10.1109/ICASSP40776.2020.9053987.

97. Ding, T. CDFI: Compression-Driven Network Design for Frame Interpolation / T. Ding, L. Liang, Z. Zhu, I. Zharkov // 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) — 2021. — P. 7997–8007. — DOI: 10.1109/CVPR46437.2021.00791.
98. Huang, Z. Real-Time Intermediate Flow Estimation for Video Frame Interpolation / Z. Huang, T. Zhang, W. Heng, B. Shi, S. Zhou // Computer Vision – ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XIV — 2022. — P. 624–642. — DOI: 10.1007/978-3-031-19781-9_36.
99. Lee, H. AdaCoF: Adaptive Collaboration of Flows for Video Frame Interpolation / H. Lee, T. Kim, T. Chung, D. Pak, Y. Ban, S. Lee // 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) — 2020. — P. 5315–5324. — DOI: 10.1109/CVPR42600.2020.00536.
100. Zhang, R. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric / R. Zhang, P. Isola, A.A. Efros, E. Shechtman, O. Wang // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition — 2018. — P. 586–595. — DOI: 10.1109/CVPR.2018.00068.
101. Curless, B. A volumetric method for building complex models from range images / B. Curless, M. Levoy // Proceedings of the 23rd annual conference on Computer graphics and interactive techniques: SIGGRAPH ’96. — 1996. — C. 303–312. — DOI: 10.1145/237170.237269.
102. Mescheder, L. Occupancy Networks: Learning 3D Reconstruction in Function Space / L. Mescheder, M. Oechsle, M. Niemeyer, S. Nowozin, A. Geiger // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) — 2019. — P. 4455–4465. — DOI: 10.1109/CVPR.2019.00459.

*Фамилия Ганеева изменена на фамилию Козлова (свидетельство о заключении брака II-ЕР № 841954, выдано отделом Дворцом бракосочетания городского округа Самара управления ЗАГС Самарской области, 20.08.2022г.).

ПРИЛОЖЕНИЕ А

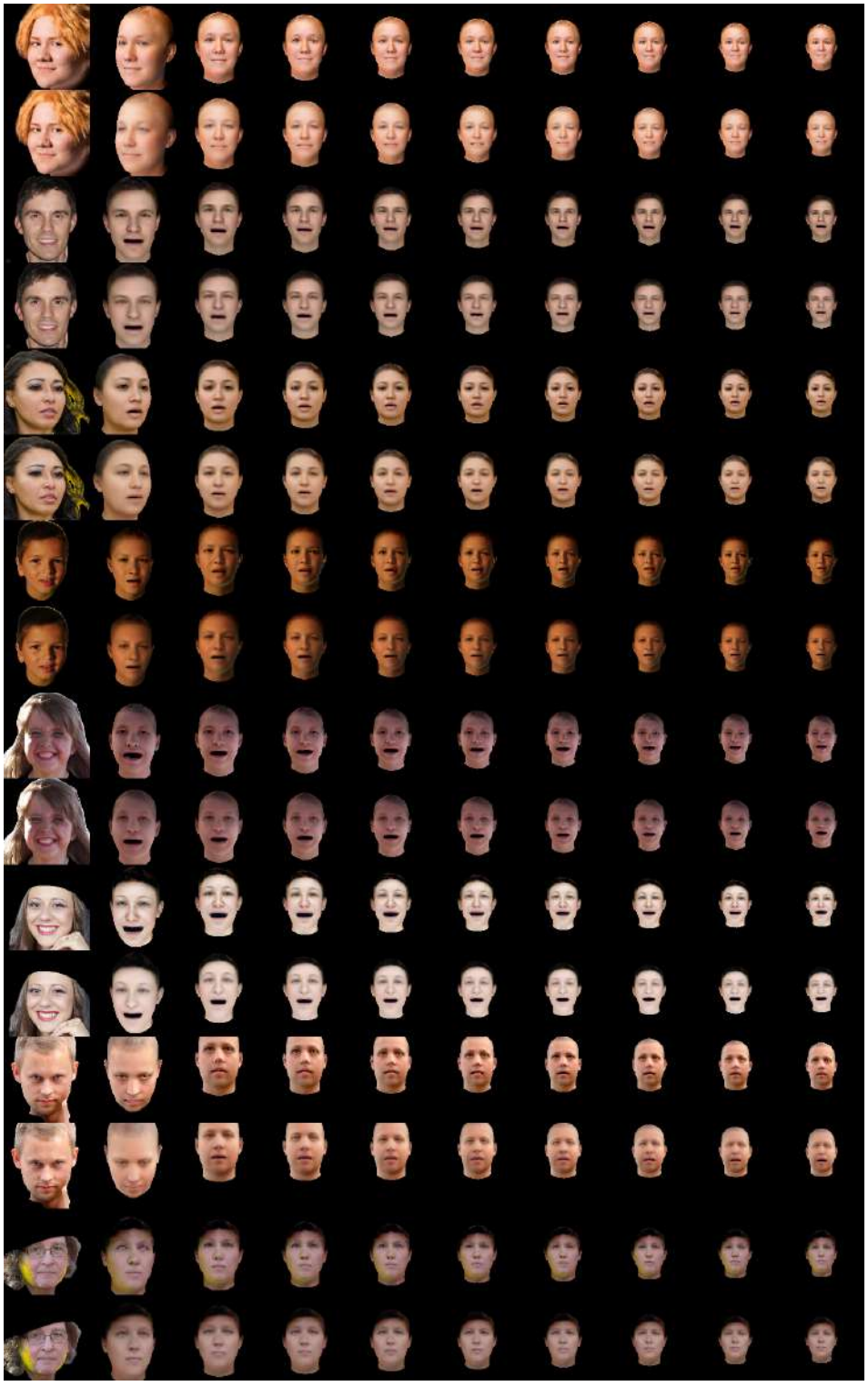


Рисунок А.1 – Результат синтеза изображений-проекций на основе обученных параметров в зависимости от удаленности камеры. Четные строки – результат синтеза, нечетные строки – результат рендеринга для модели FLAME

Рисунок А.2 – Результат синтеза изображений-проекций на основе обученных параметров в зависимости от положения шеи (по осям x, y, z). Четные строки – результат синтеза, нечетные строки – результат рендеринга для модели FLAME

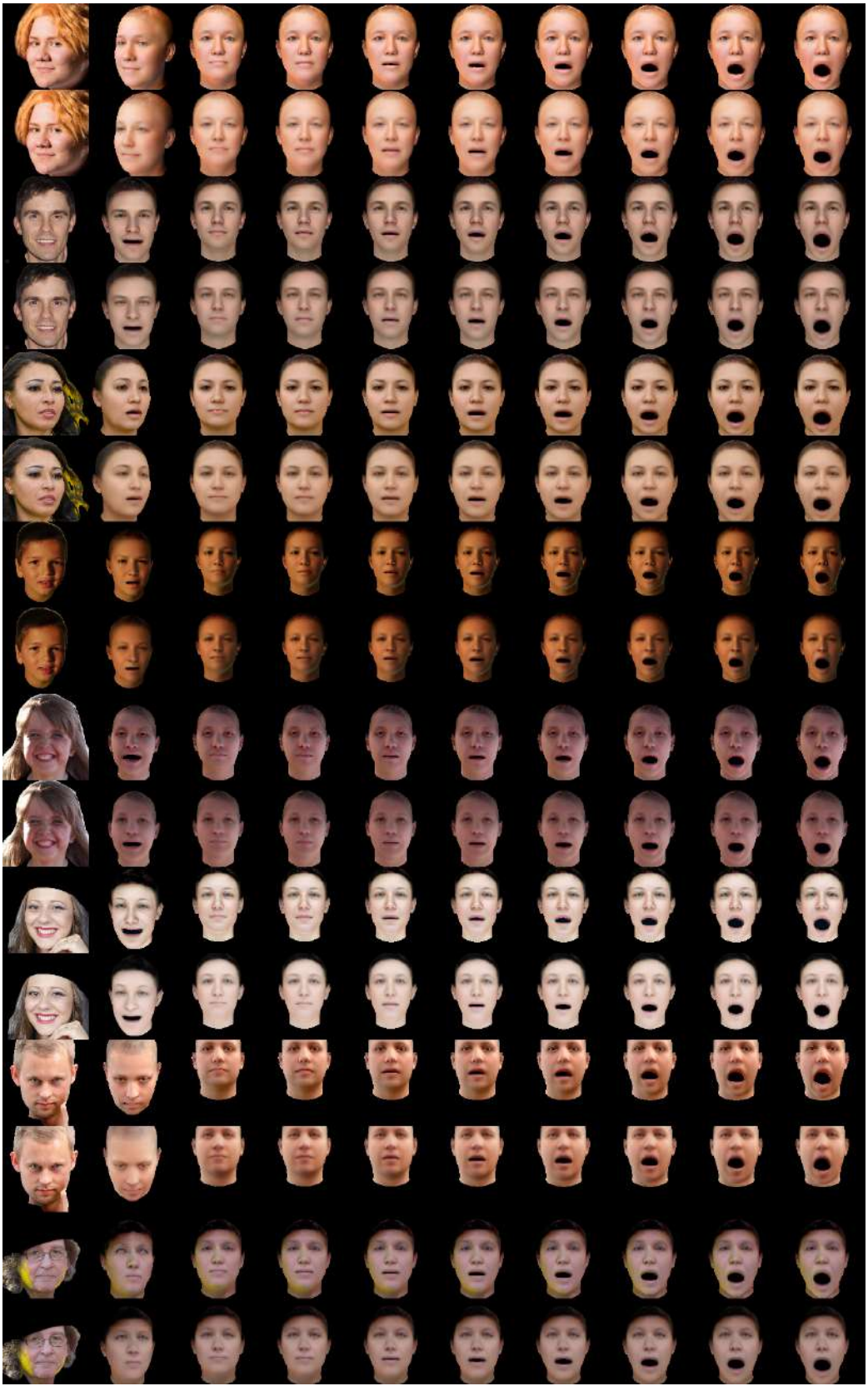


Рисунок А.3 – Результат синтеза изображений-проекций на основе обученных параметров в зависимости от степени открытия челюсти. Четные строки – результат синтеза, нечетные – результат рендеринга для модели FLAME

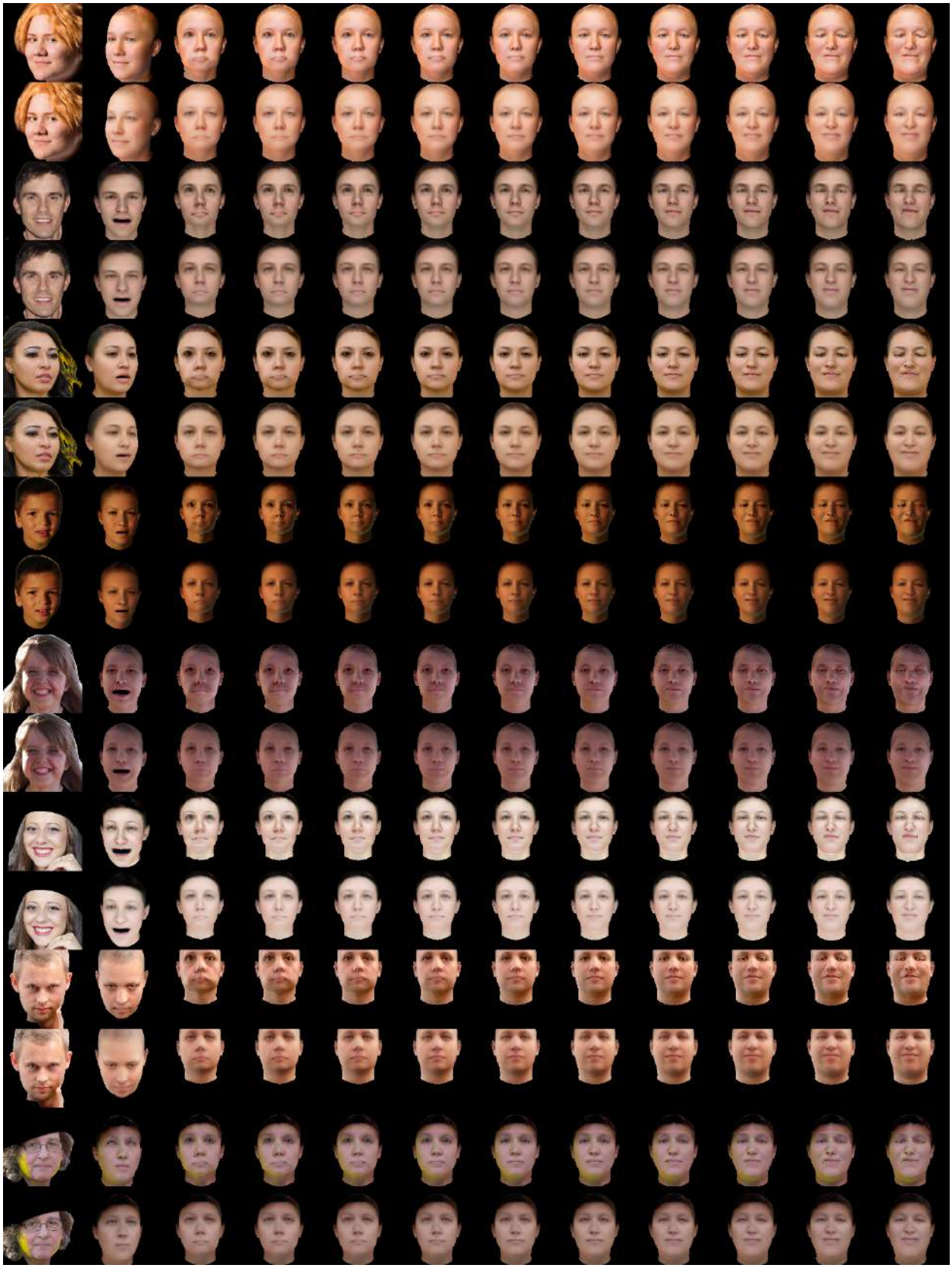


Рисунок А.4 – Результат синтеза изображений-проекций на основе обученных параметров в зависимости от выражения лица. Четные строки – результат синтеза, нечетные строки – результат рендеринга для модели FLAME

Рисунок А.5 – Результат синтеза изображений-проекций на основе обученных параметров с новых точек обзора. Слева – результат рендеринга для модели FLAME, справа – результат синтеза

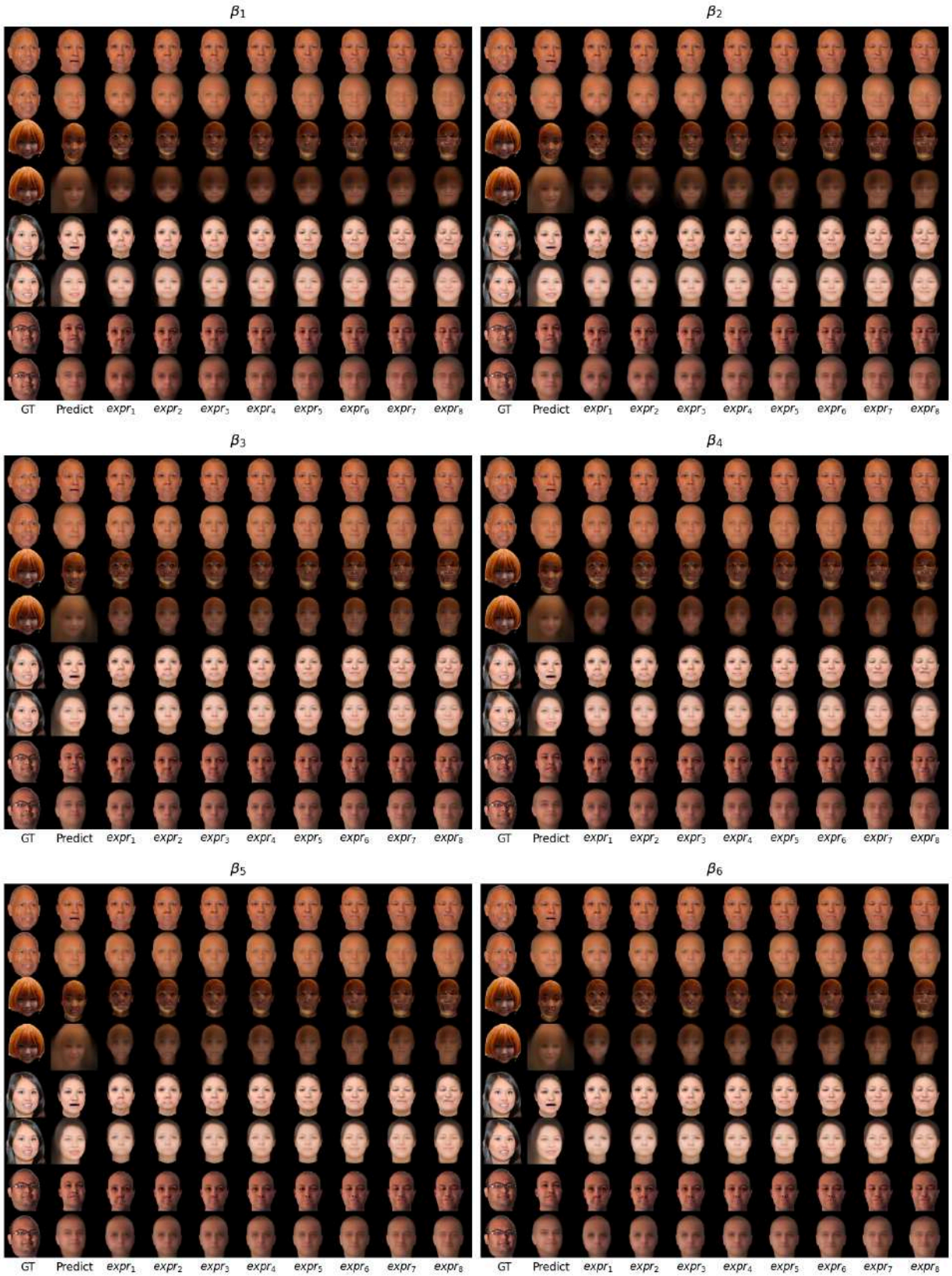


Рисунок А.6 – Результаты генерации на тестовой выборке для контроля выражения лица экспериментов $\beta_1 - \beta_6$ для способа обучения α_1. Четные строки – результаты синтеза, нечетные – результат рендеринга для модели FLAME

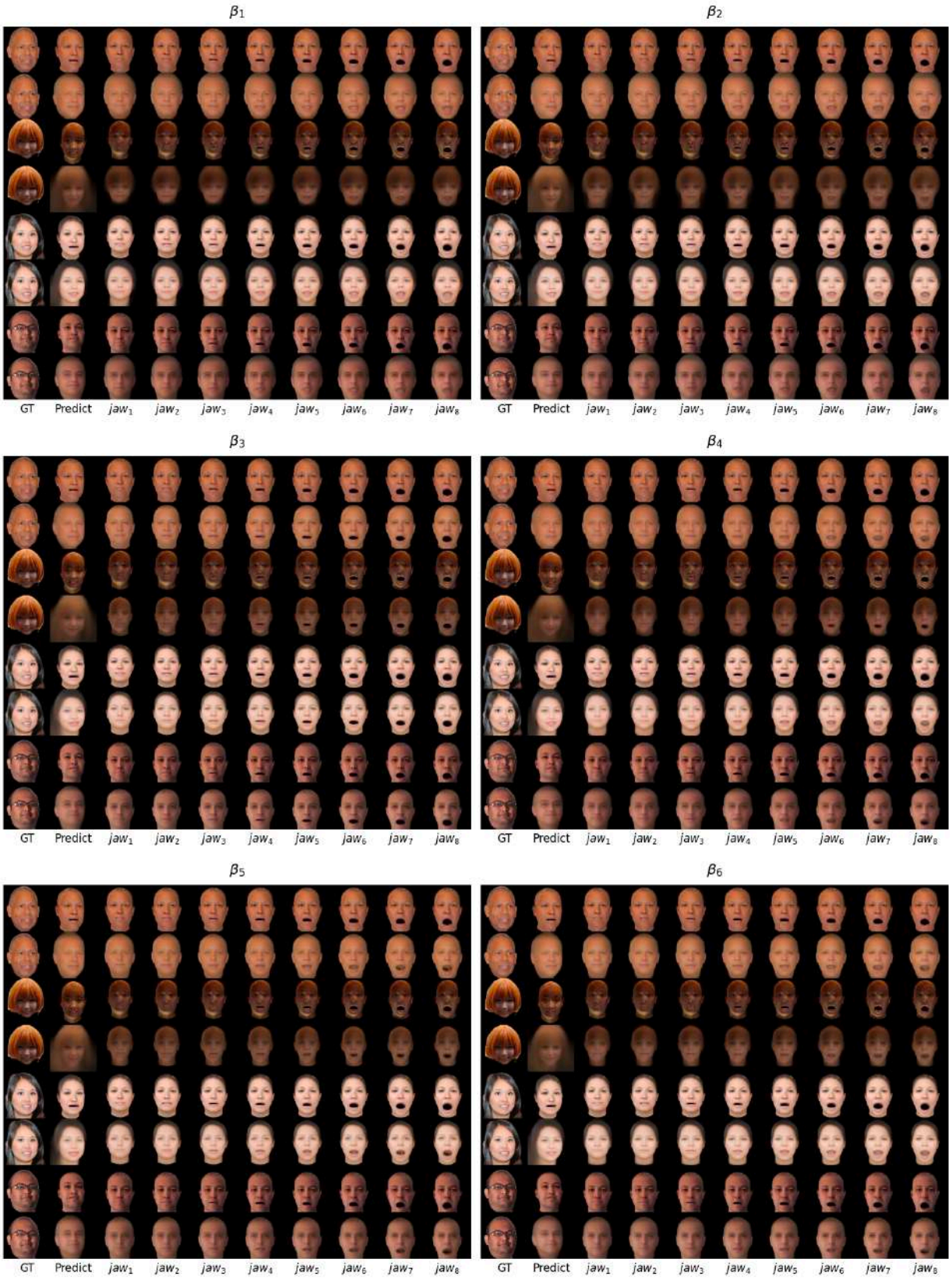


Рисунок А.7 – Результаты генерации на тестовой выборке для контроля степени открытия челюсти экспериментов $\beta_1 - \beta_6$ для способа обучения α_1. Четные строки – результаты синтеза, нечетные – результат рендеринга для модели FLAME

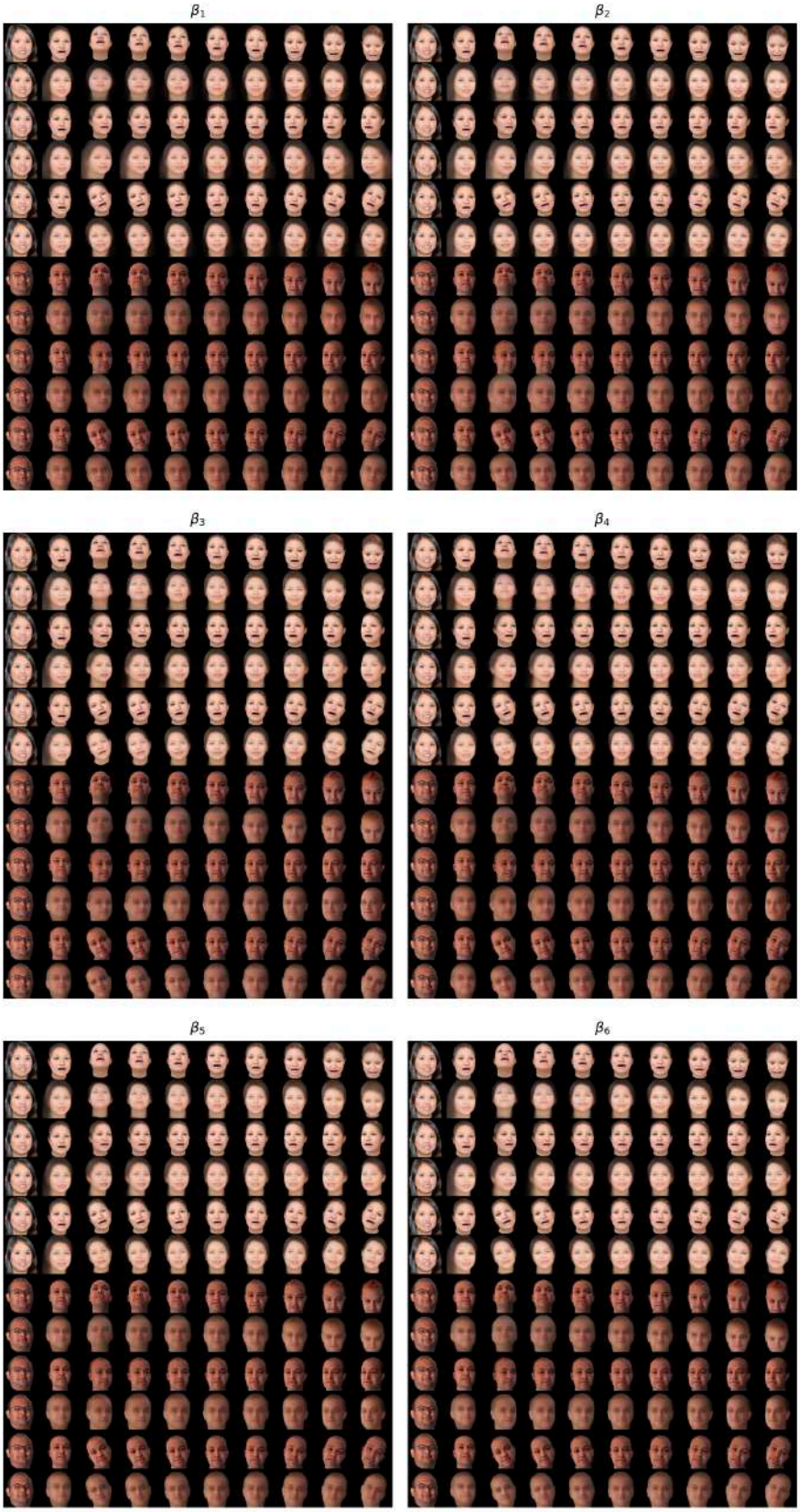


Рисунок А.8 – Результаты генерации на тестовой выборке для контроля поворота шеи экспериментов $\beta_1 - \beta_6$ для способа обучения α_1. Четные строки – результаты синтеза, нечетные – результат рендеринга для модели FLAME

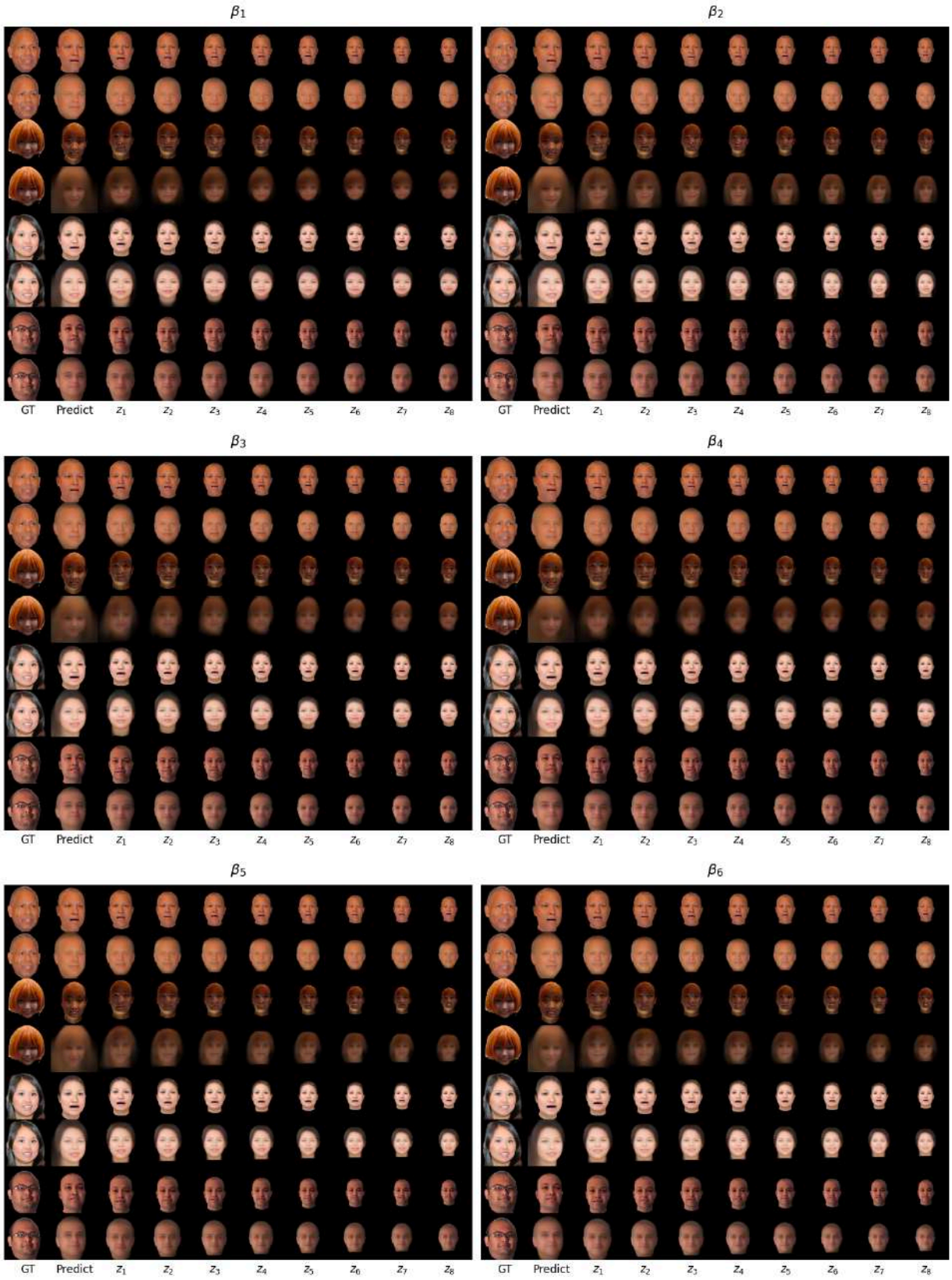


Рисунок А.9 – Результаты генерации на тестовой выборке для контроля расстояния от камеры до головы по оси z экспериментов $\beta_1 - \beta_6$ для способа обучения α_1. Четные строки – результаты синтеза, нечетные – результат рендеринга для модели FLAME

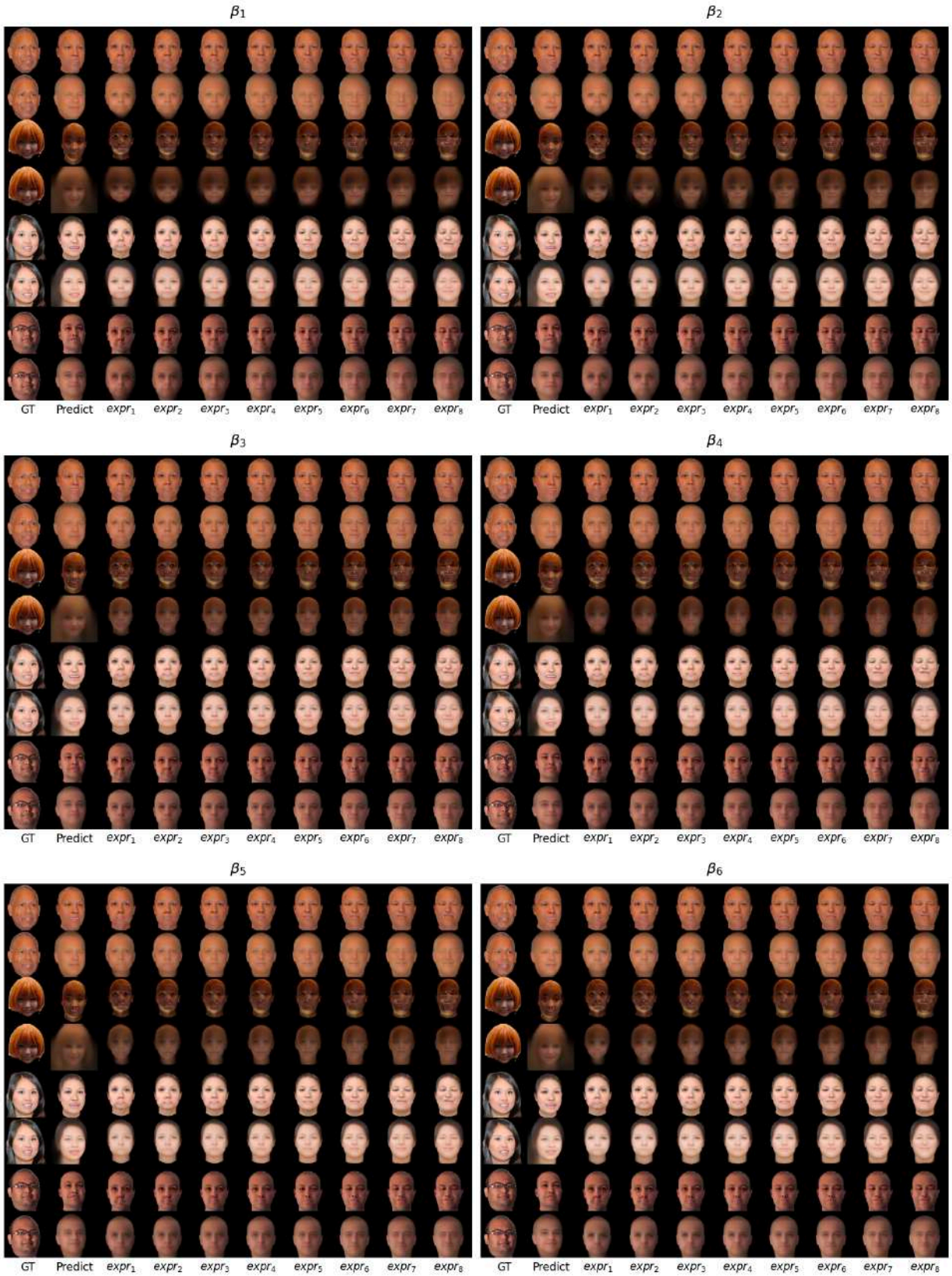


Рисунок А.10 – Результаты генерации на тестовой выборке для контроля выражения лица экспериментов $\beta_1 - \beta_6$ для способа обучения α_2. Четные строки – результаты синтеза, нечетные – результат рендеринга для модели FLAME

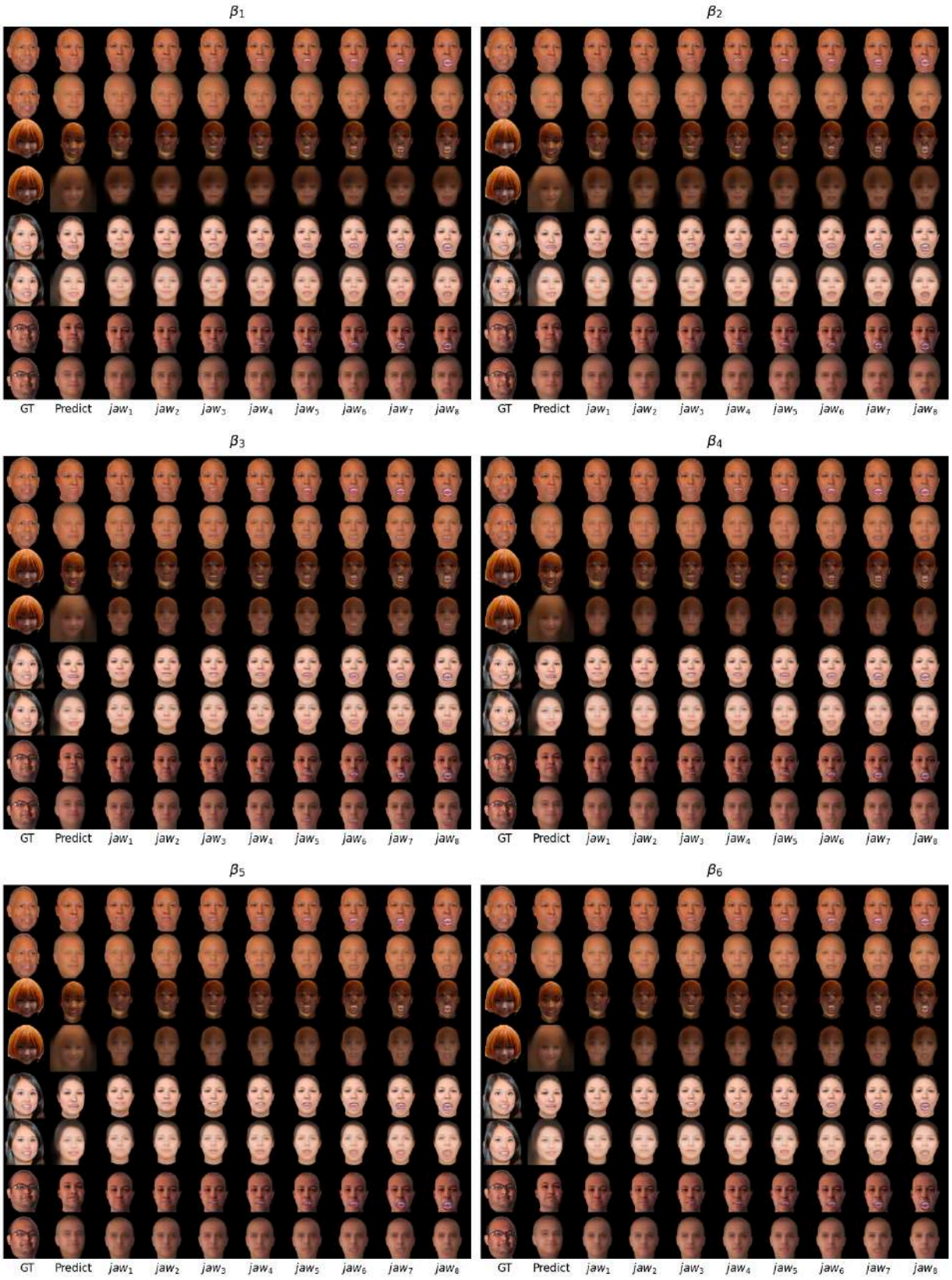


Рисунок А.11 – Результаты генерации на тестовой выборке для контроля степени открытия челюсти экспериментов $\beta_1 - \beta_6$ для способа обучения α_2. Четные строки – результаты синтеза, нечетные – результат рендеринга для модели FLAME

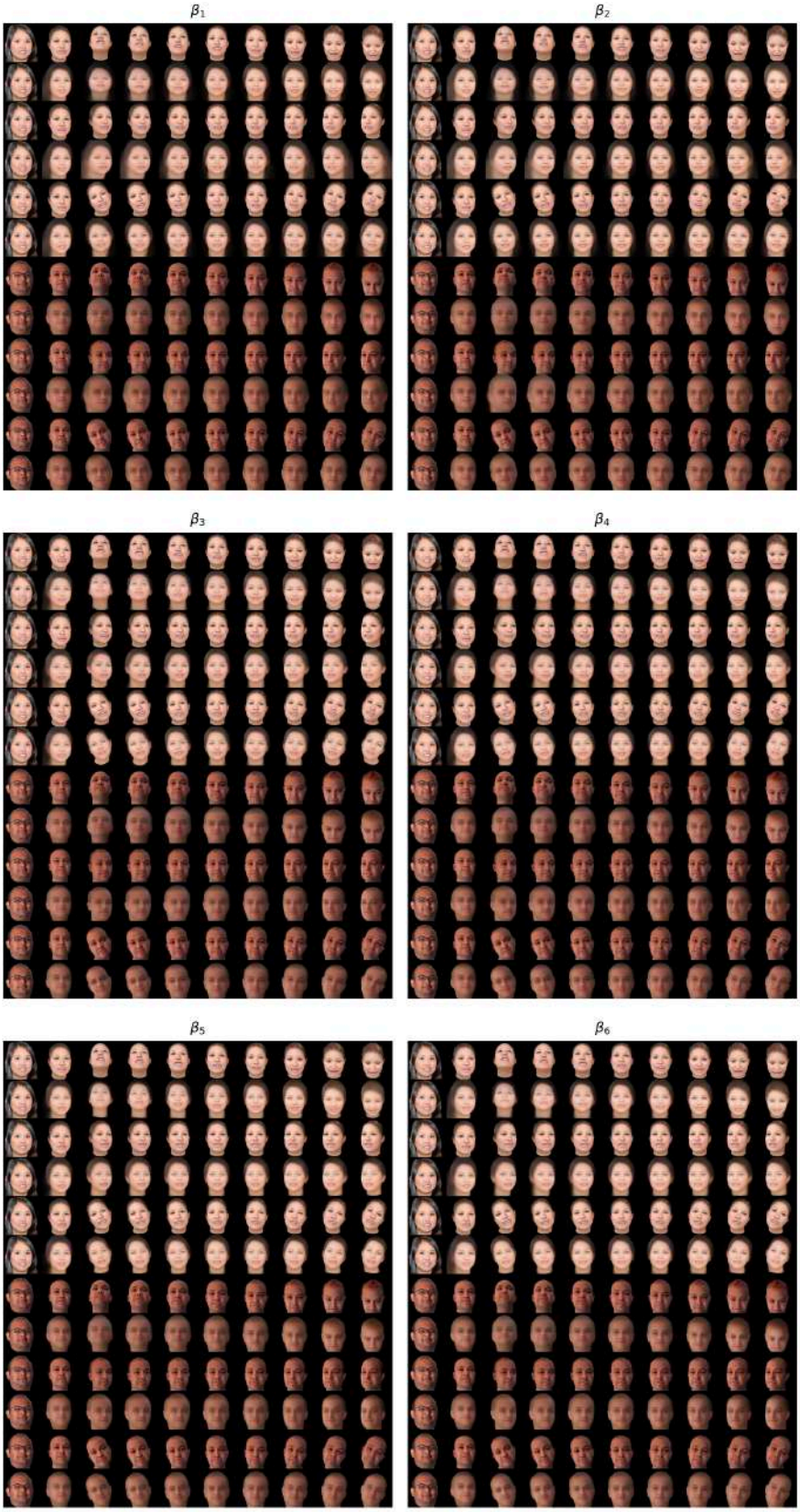


Рисунок А.12 – Результаты генерации на тестовой выборке для контроля поворота шеи экспериментов $\beta_1 - \beta_6$ для способа обучения α_2. Четные строки – результаты синтеза, нечетные – результат рендеринга для модели FLAME

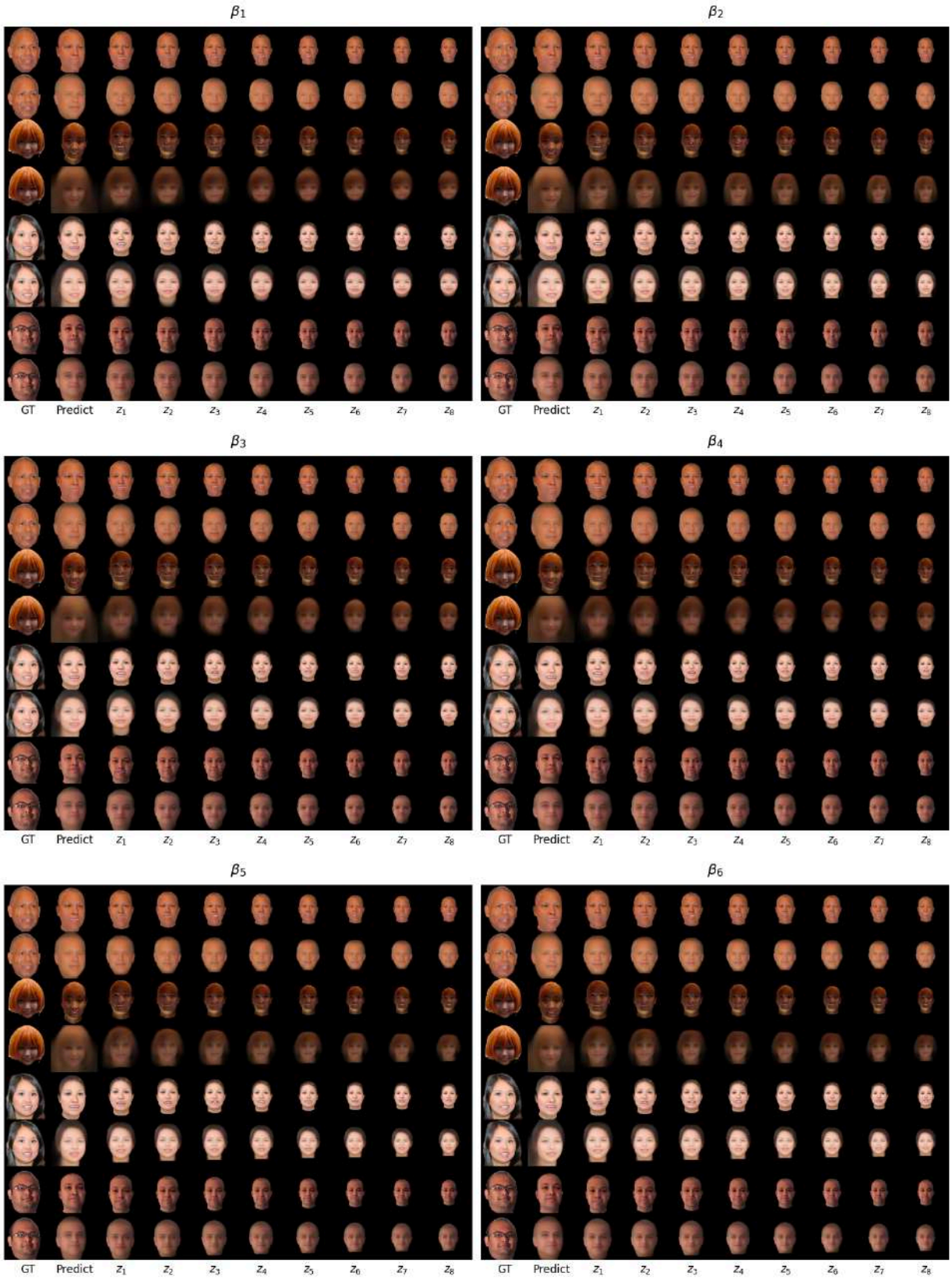


Рисунок А.13 – Результаты генерации на тестовой выборке для контроля расстояния от камеры до головы по оси z экспериментов $\beta_1 - \beta_6$ для способа обучения α_2. Четные строки – результаты синтеза, нечетные – результат рендеринга для модели FLAME

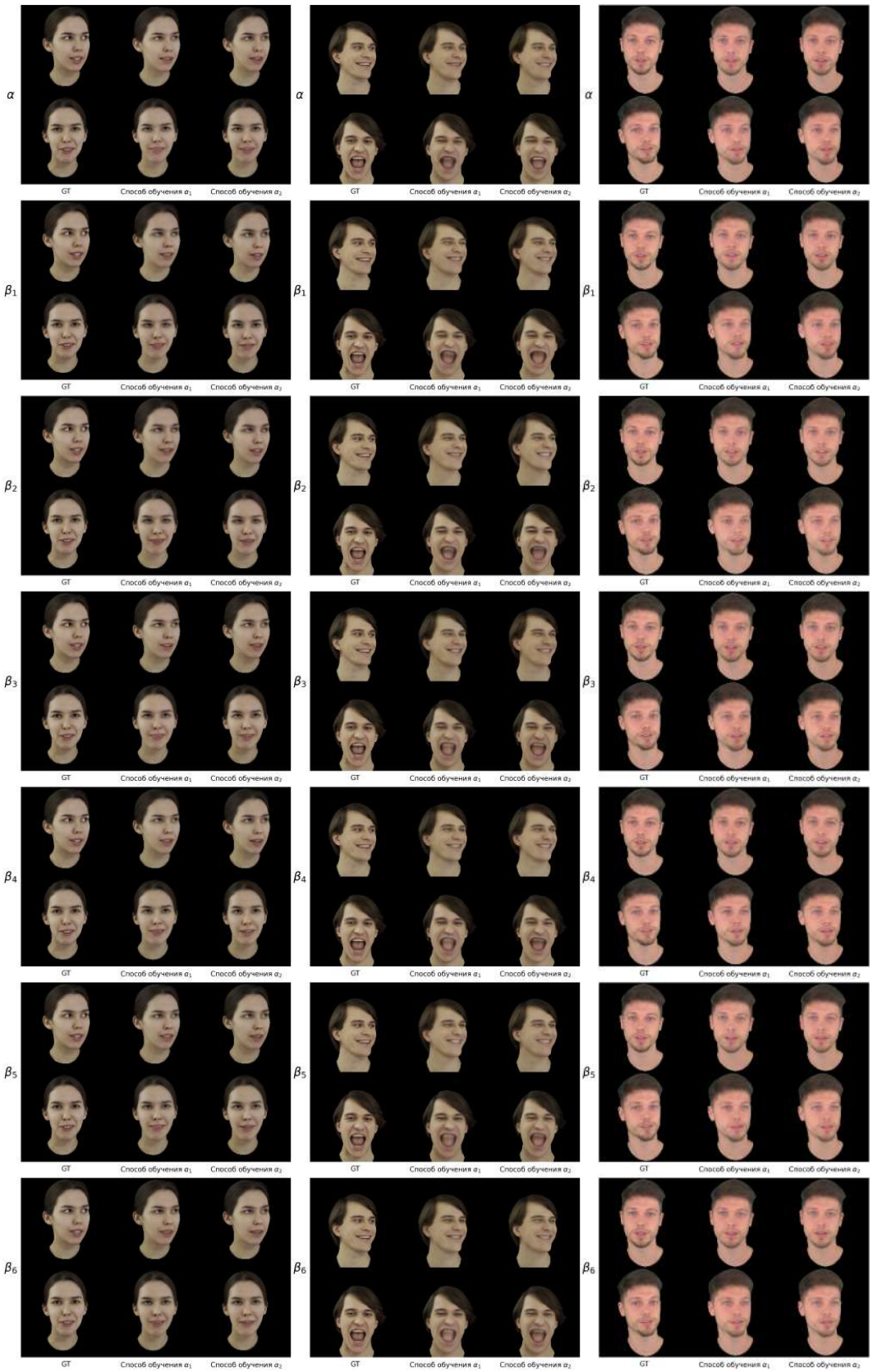


Рисунок А.14 – Результаты синтеза по параметрам, соответствующим отобранным из валидационных выборок изображений

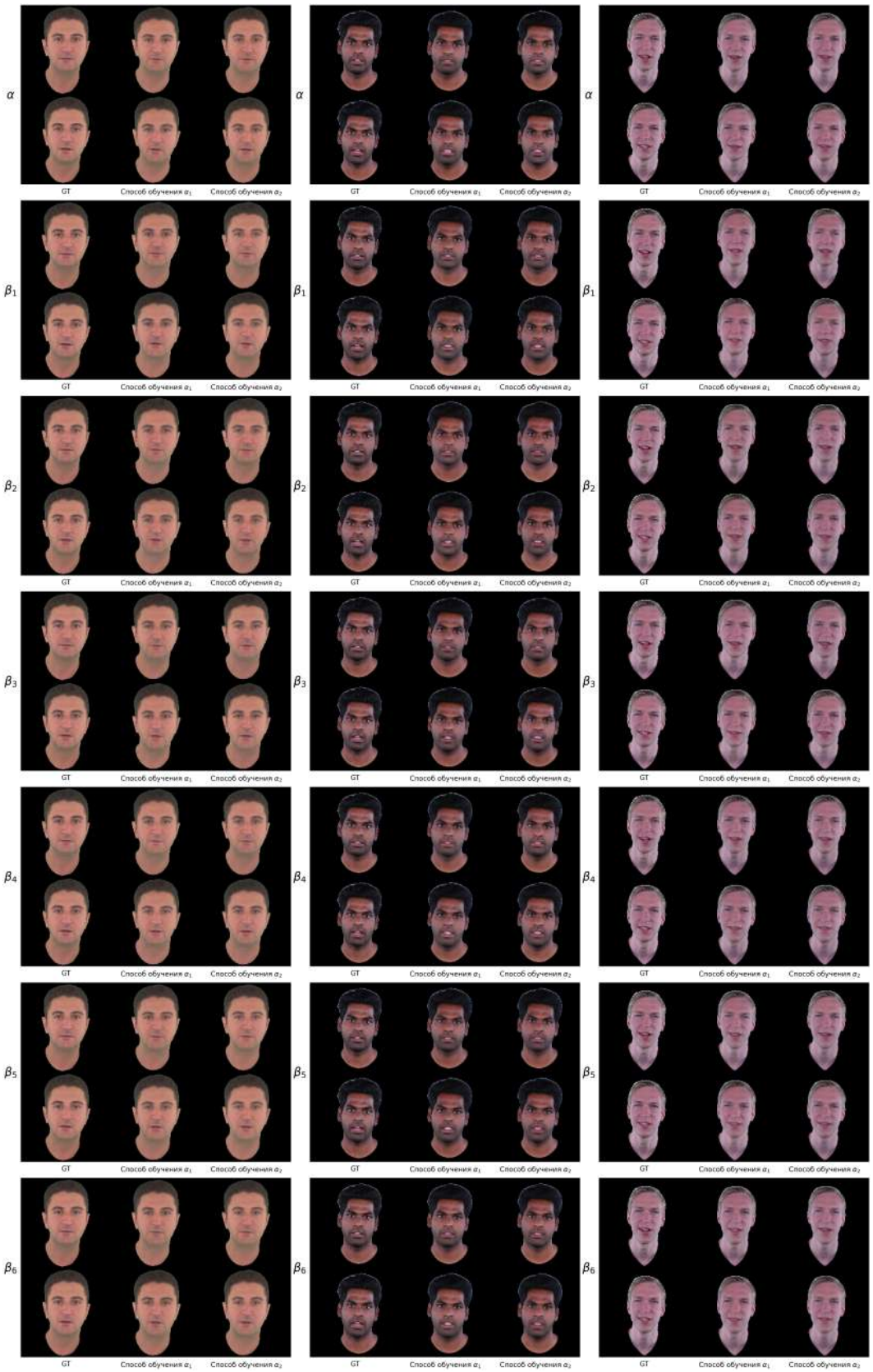


Рисунок А.15 – Результаты синтеза по параметрам, соответствующим отобранным из валидационных выборок изображений

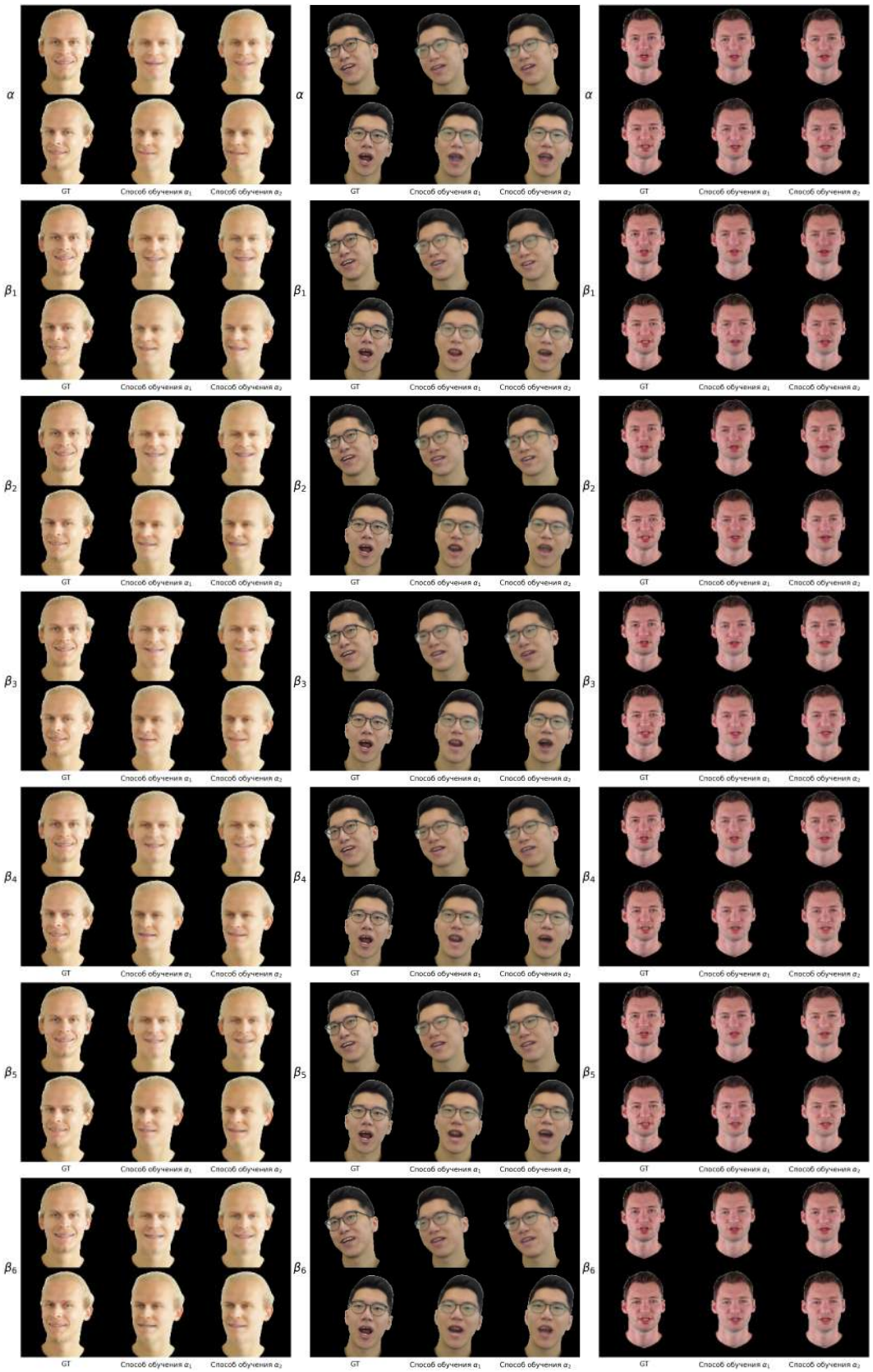


Рисунок А.16 – Результаты синтеза по параметрам, соответствующим отобранным из валидационных выборок изображений

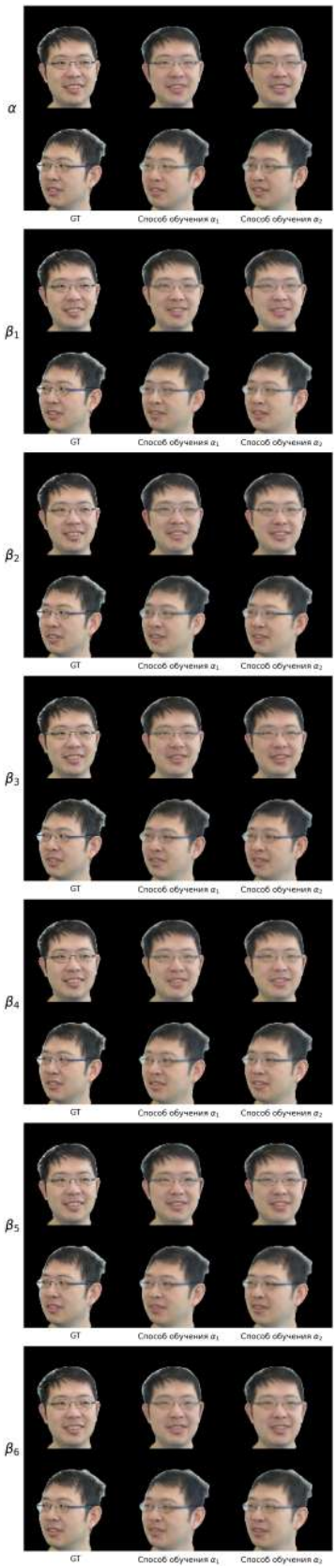


Рисунок А.17 – Результаты синтеза по параметрам, соответствующим отобранным из валидационных выборок изображений

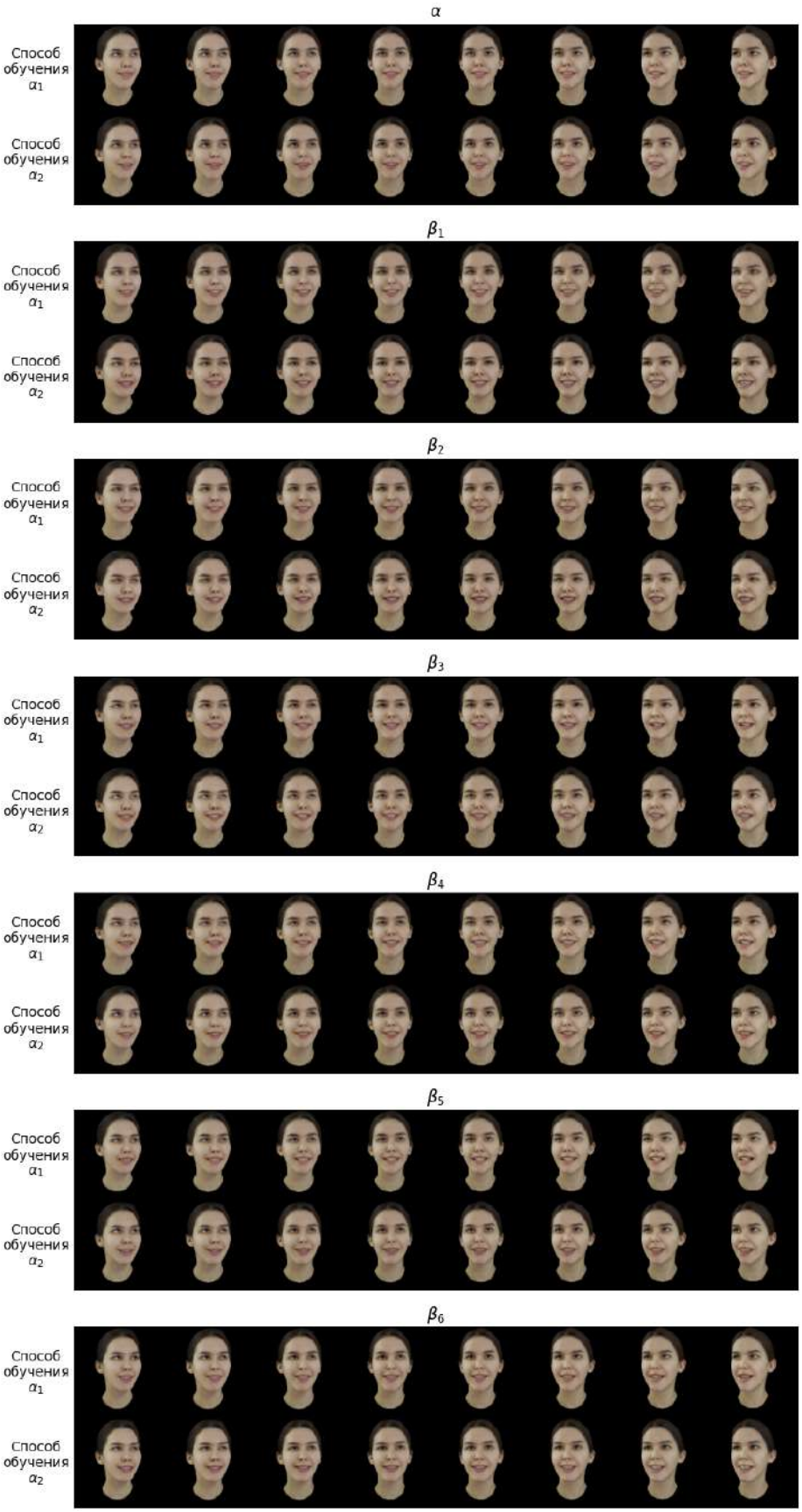


Рисунок А.18 – Результаты синтеза при варьировании поворота шеи (с открытой челюстью)

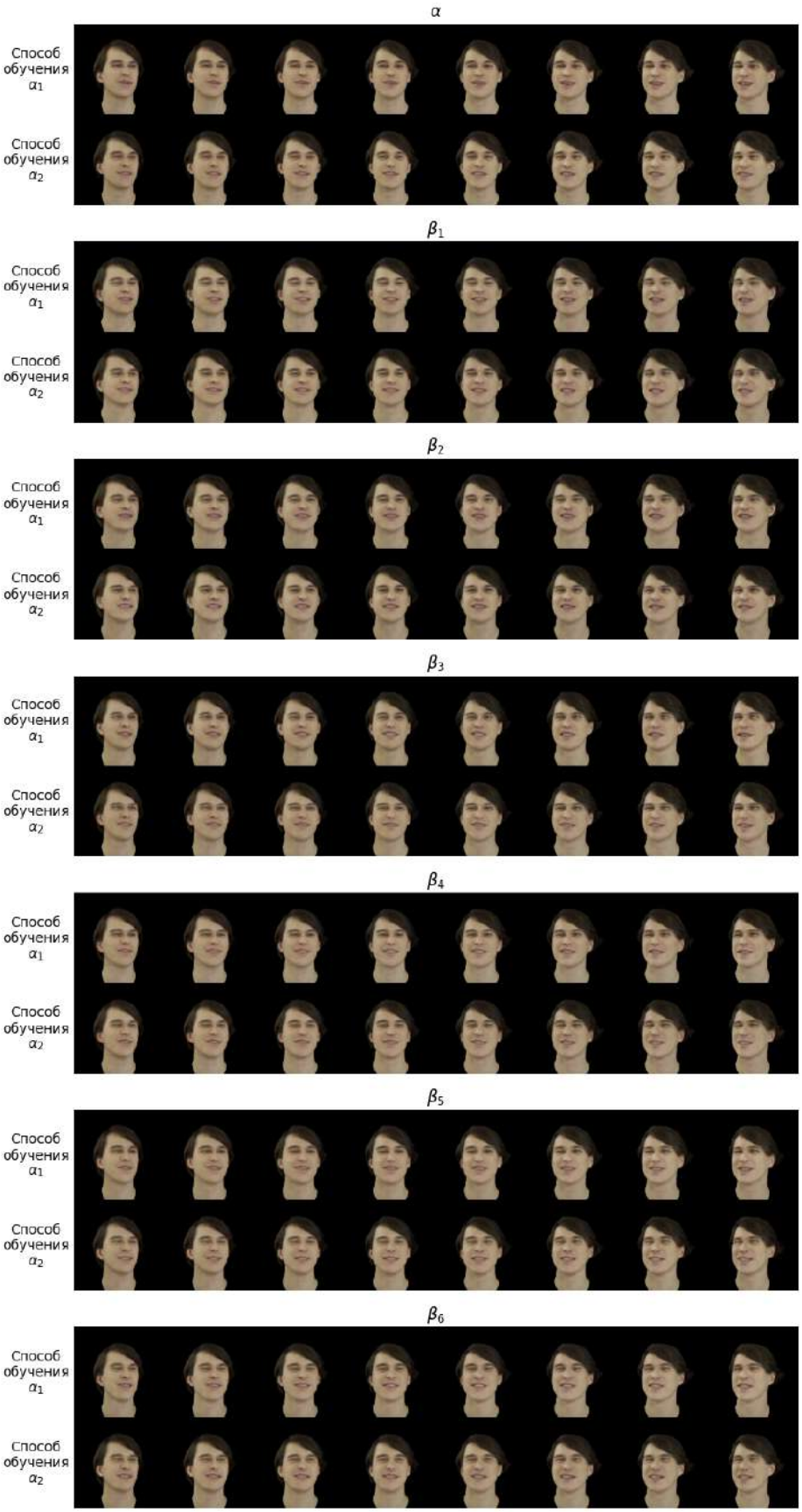


Рисунок А.19 – Результаты синтеза при варьировании поворота шеи (с открытой челюстью)

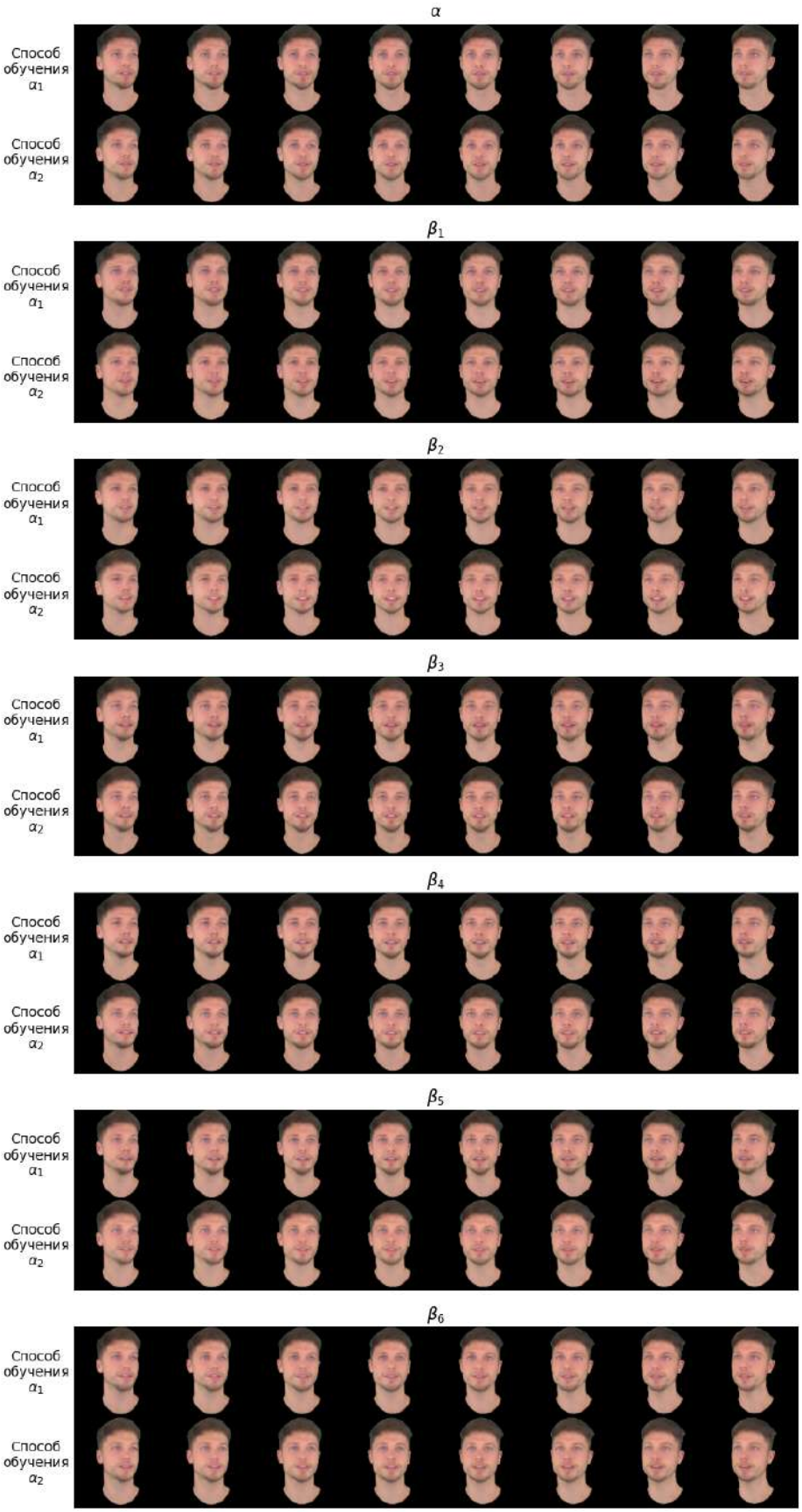


Рисунок А.20 – Результаты синтеза при варьировании поворота шеи (с открытой челюстью)

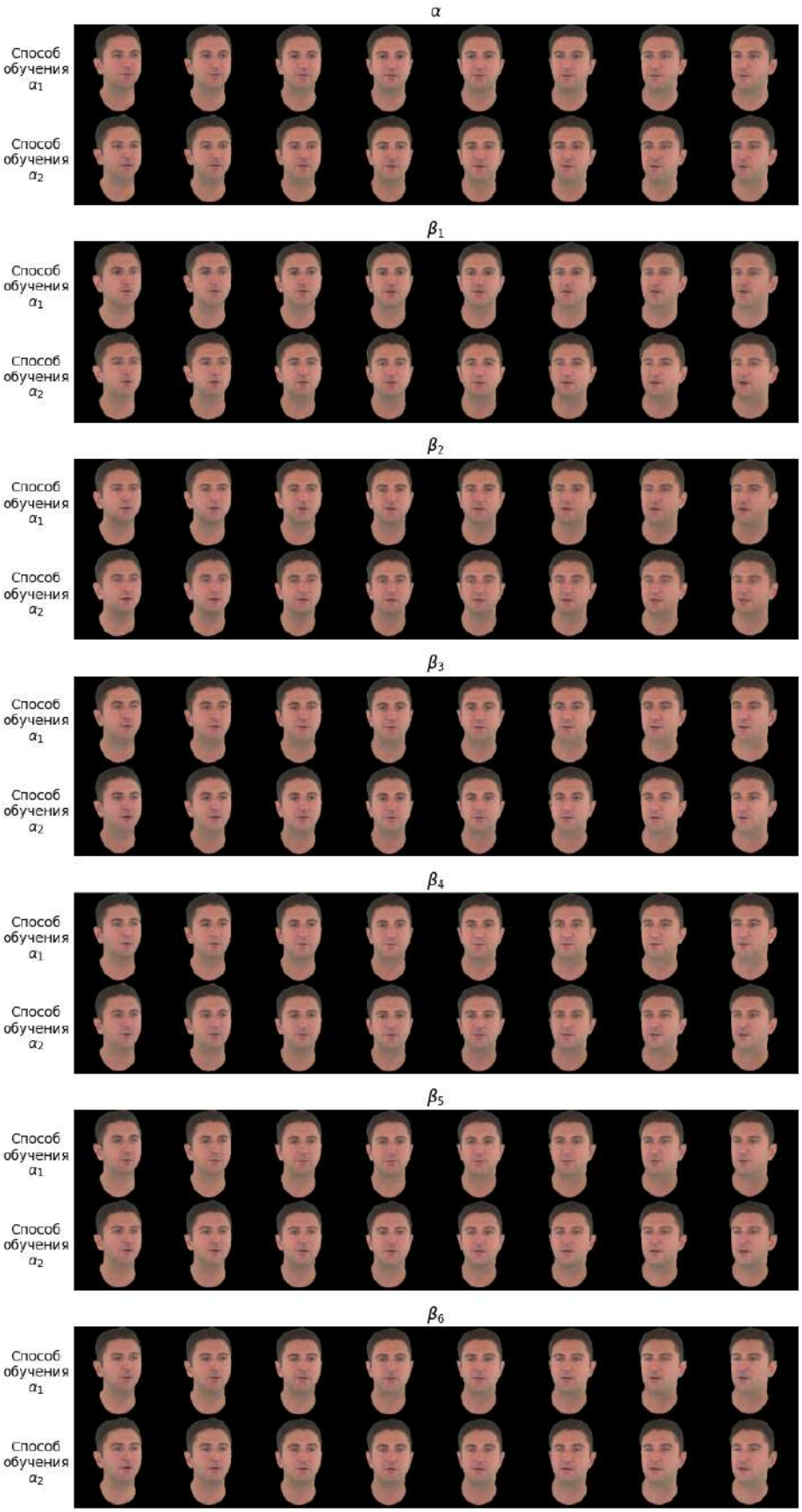


Рисунок А.21 – Результаты синтеза при варьировании поворота шеи (с открытой челюстью)

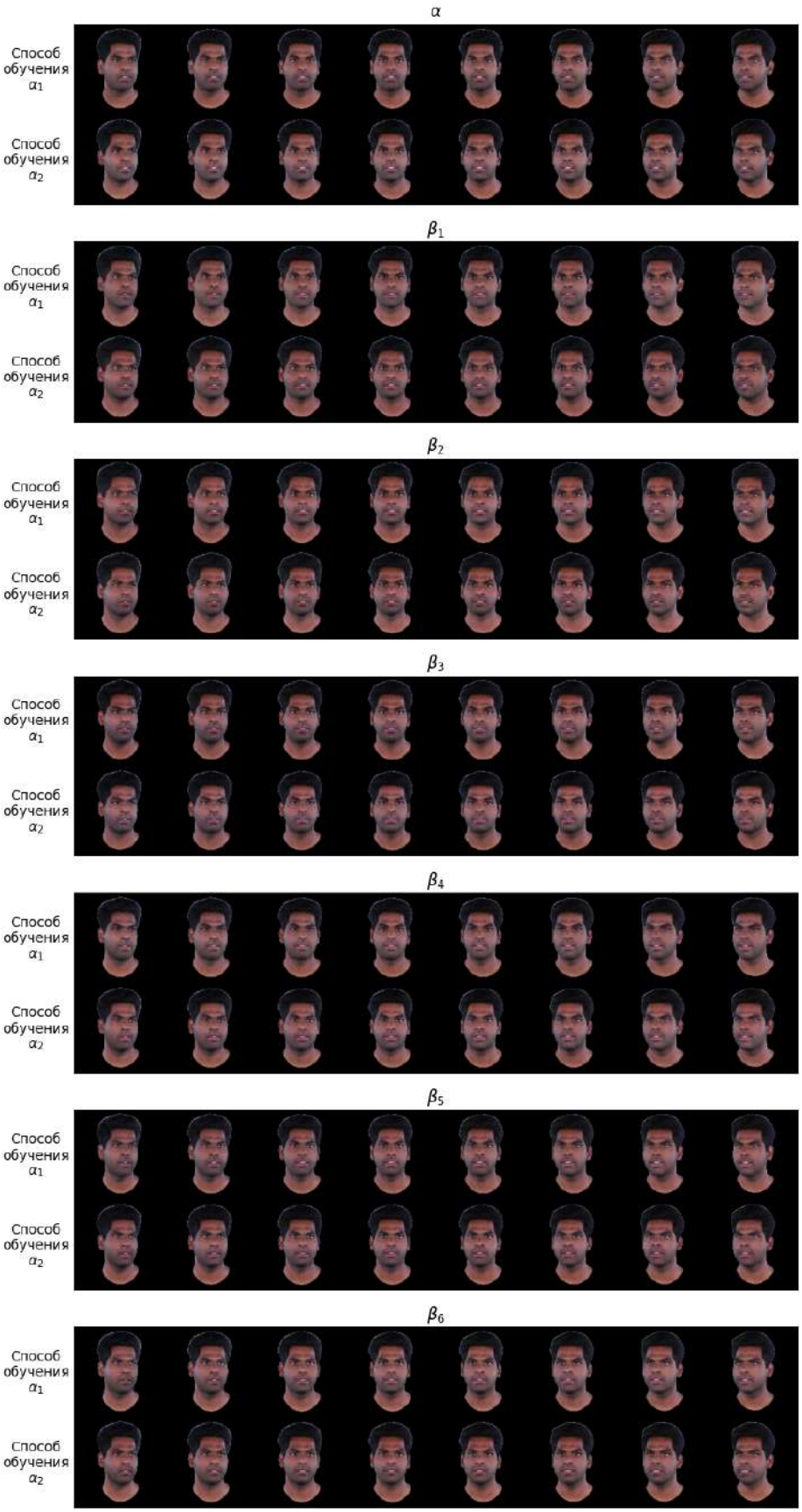


Рисунок А.22 – Результаты синтеза при варьировании поворота шеи (с открытой челюстью)

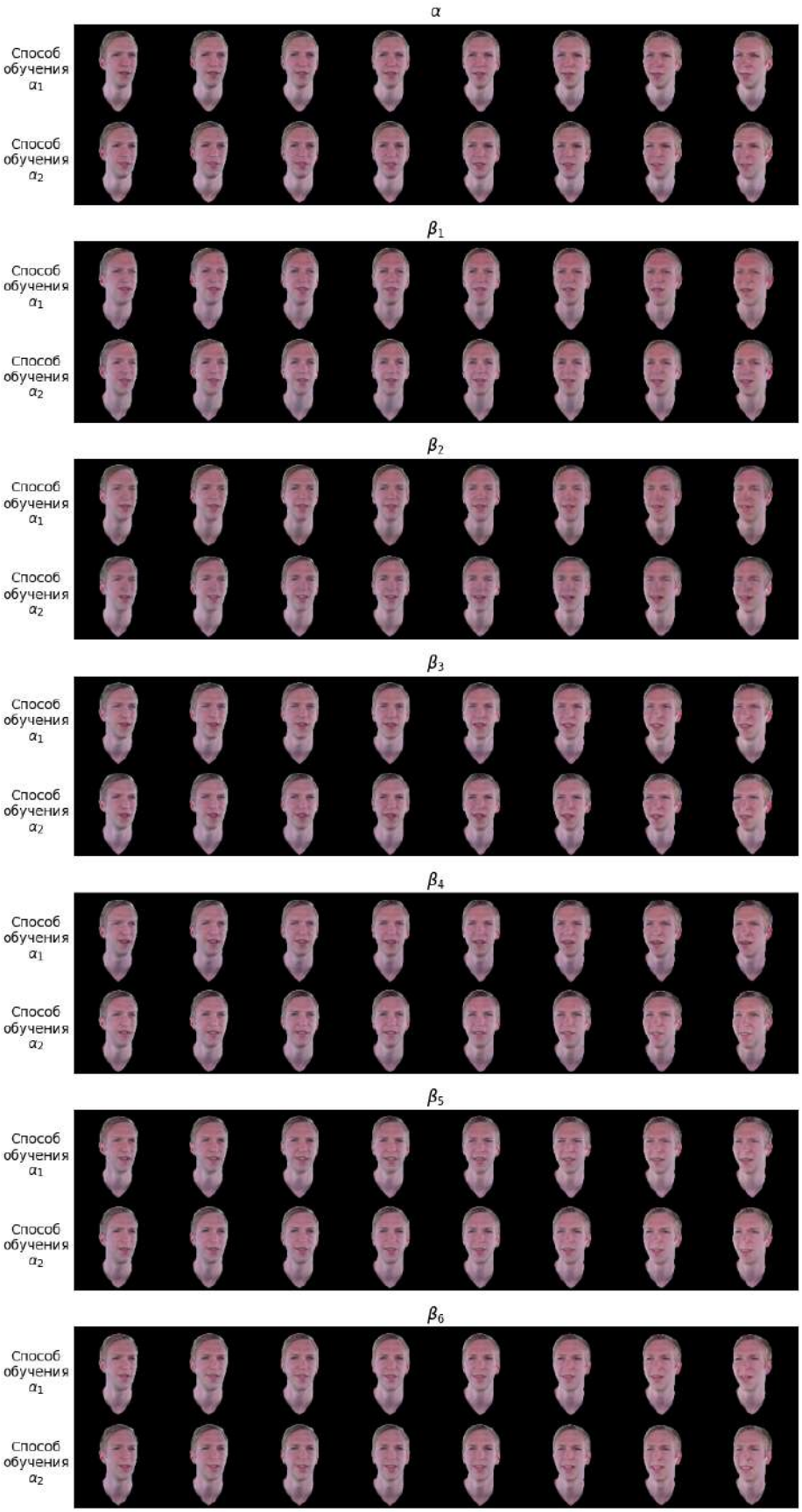


Рисунок А.23 – Результаты синтеза при варьировании поворота шеи (с открытой челюстью)

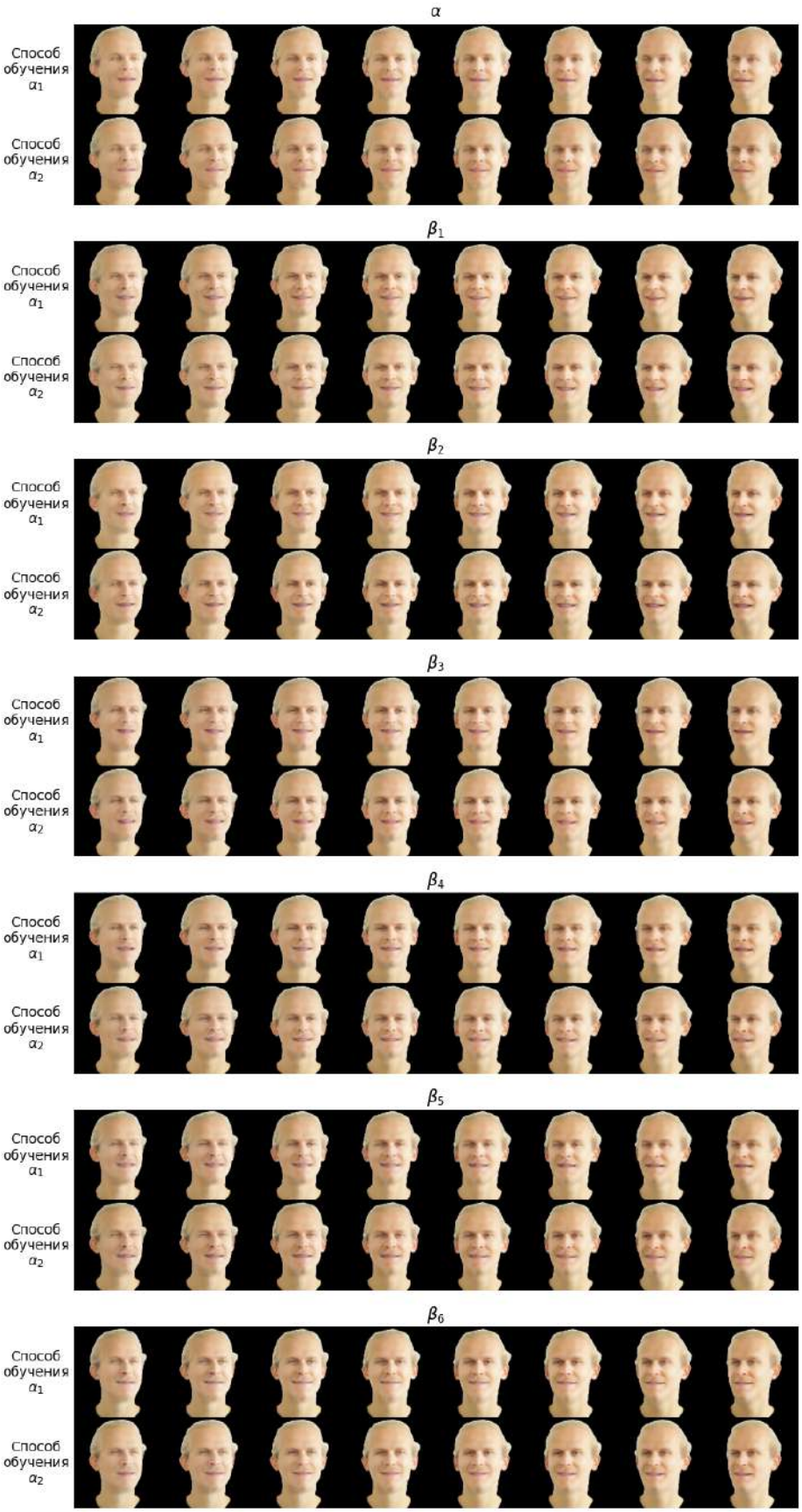


Рисунок А.24 – Результаты синтеза при варьировании поворота шеи (с открытой челюстью)

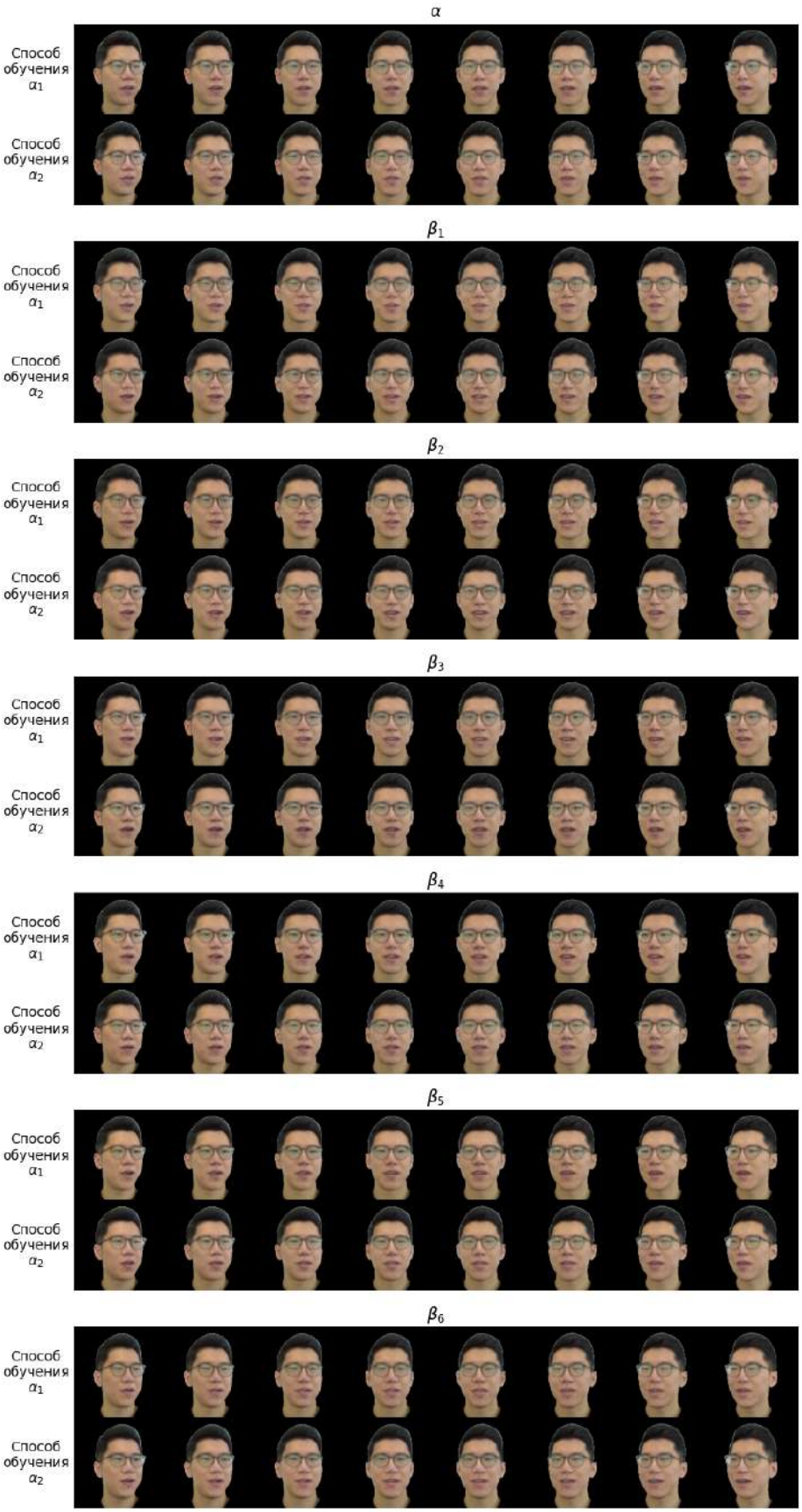


Рисунок А.25 – Результаты синтеза при варьировании поворота шеи (с открытой челюстью)

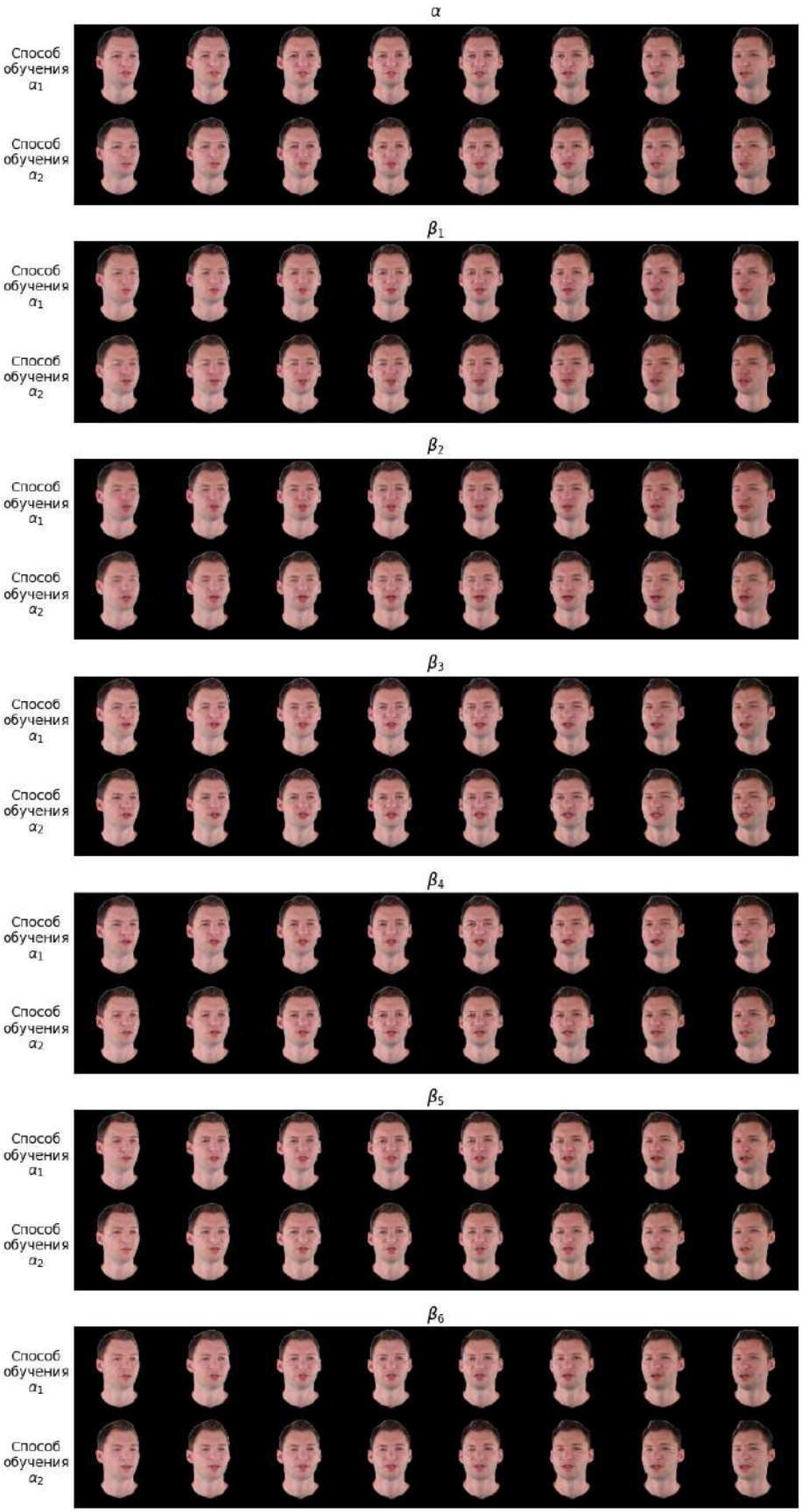


Рисунок А.26 – Результаты синтеза при варьировании поворота шеи (с открытой челюстью)

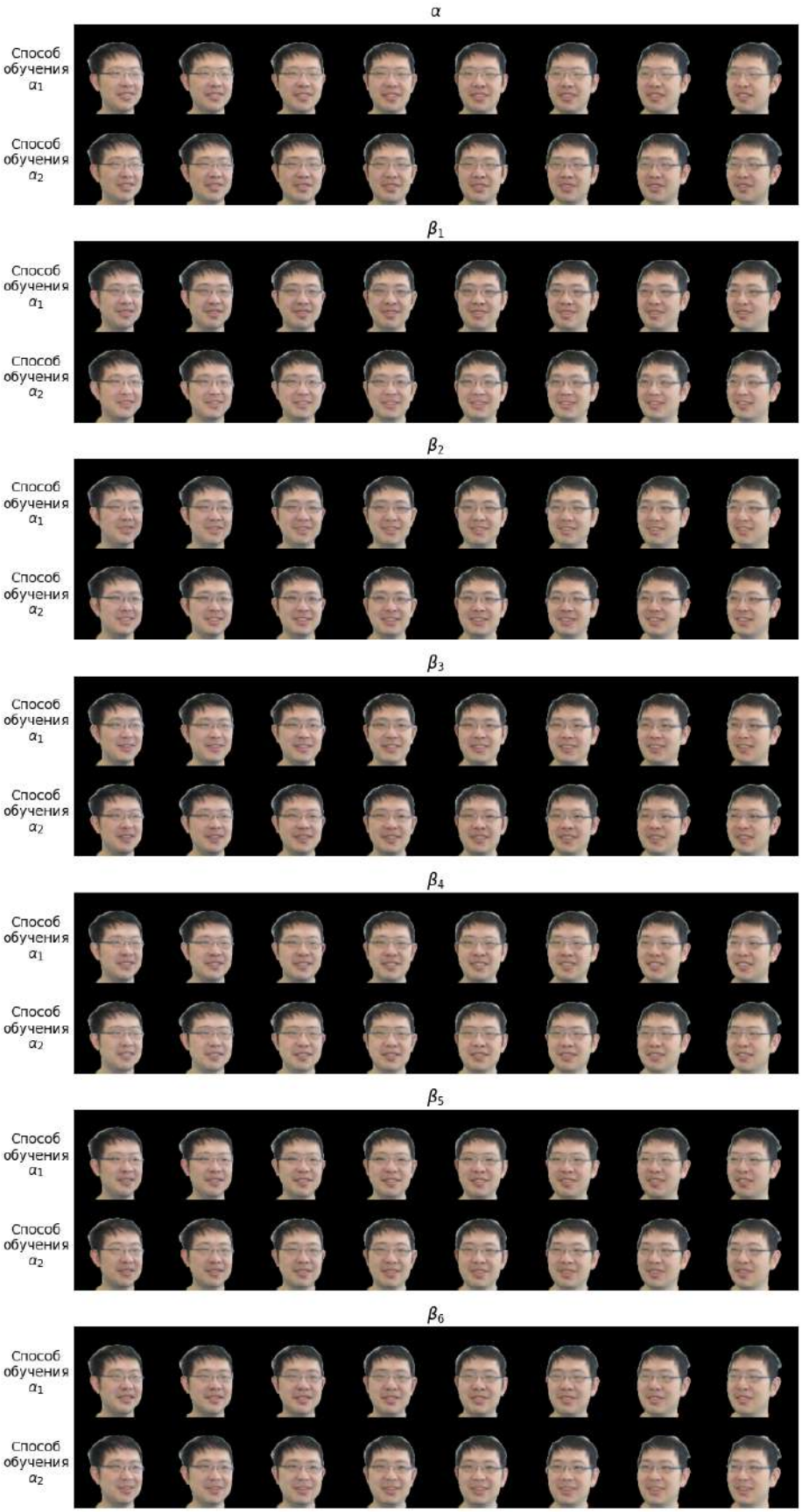


Рисунок А.27 – Результаты синтеза при варьировании поворота шеи (с открытой челюстью)

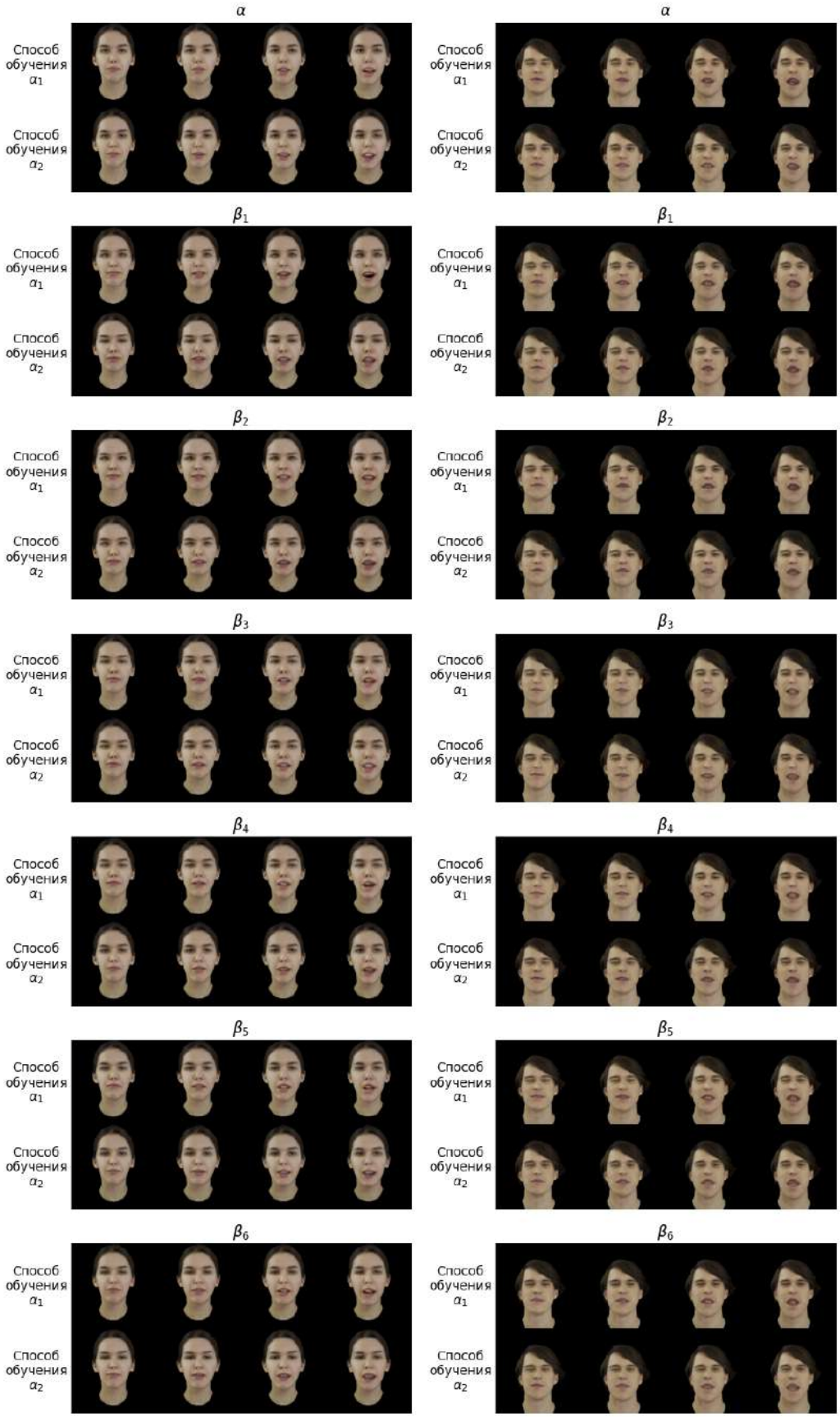


Рисунок А.28 – Результаты синтеза при варьировании степени открытия челюсти

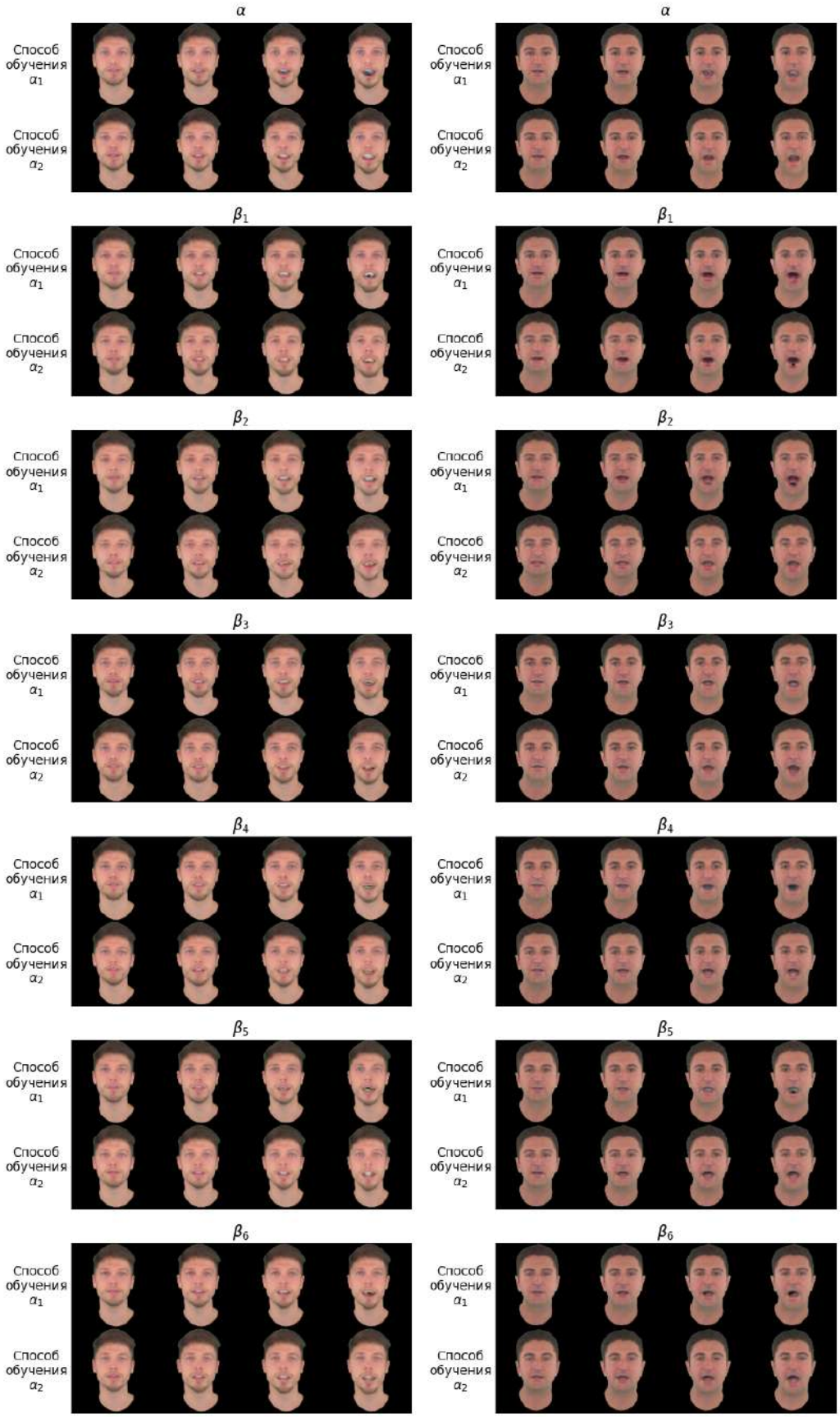


Рисунок А.29 – Результаты синтеза при варьировании степени открытия челюсти

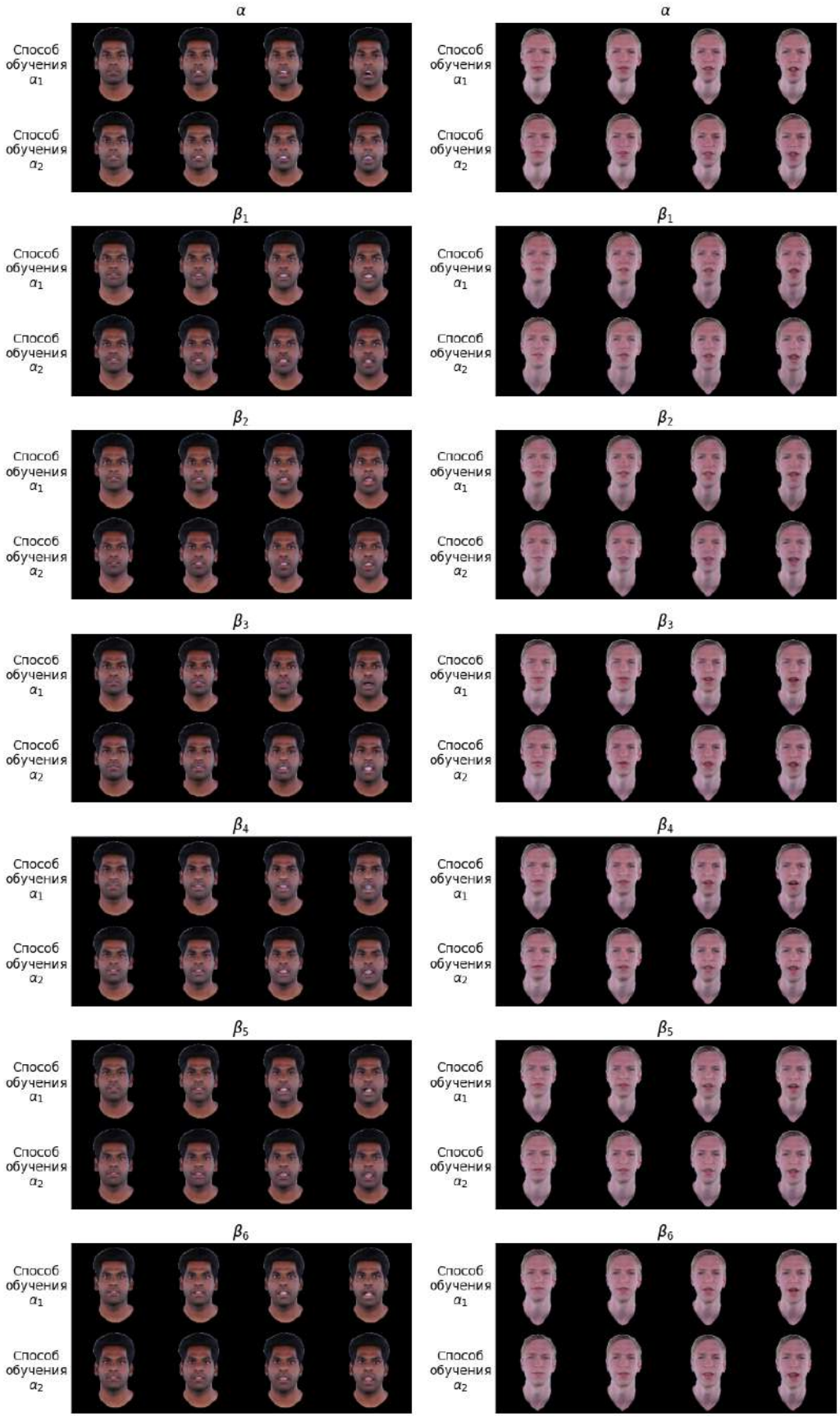


Рисунок А.30 – Результаты синтеза при варьировании степени открытия челюсти

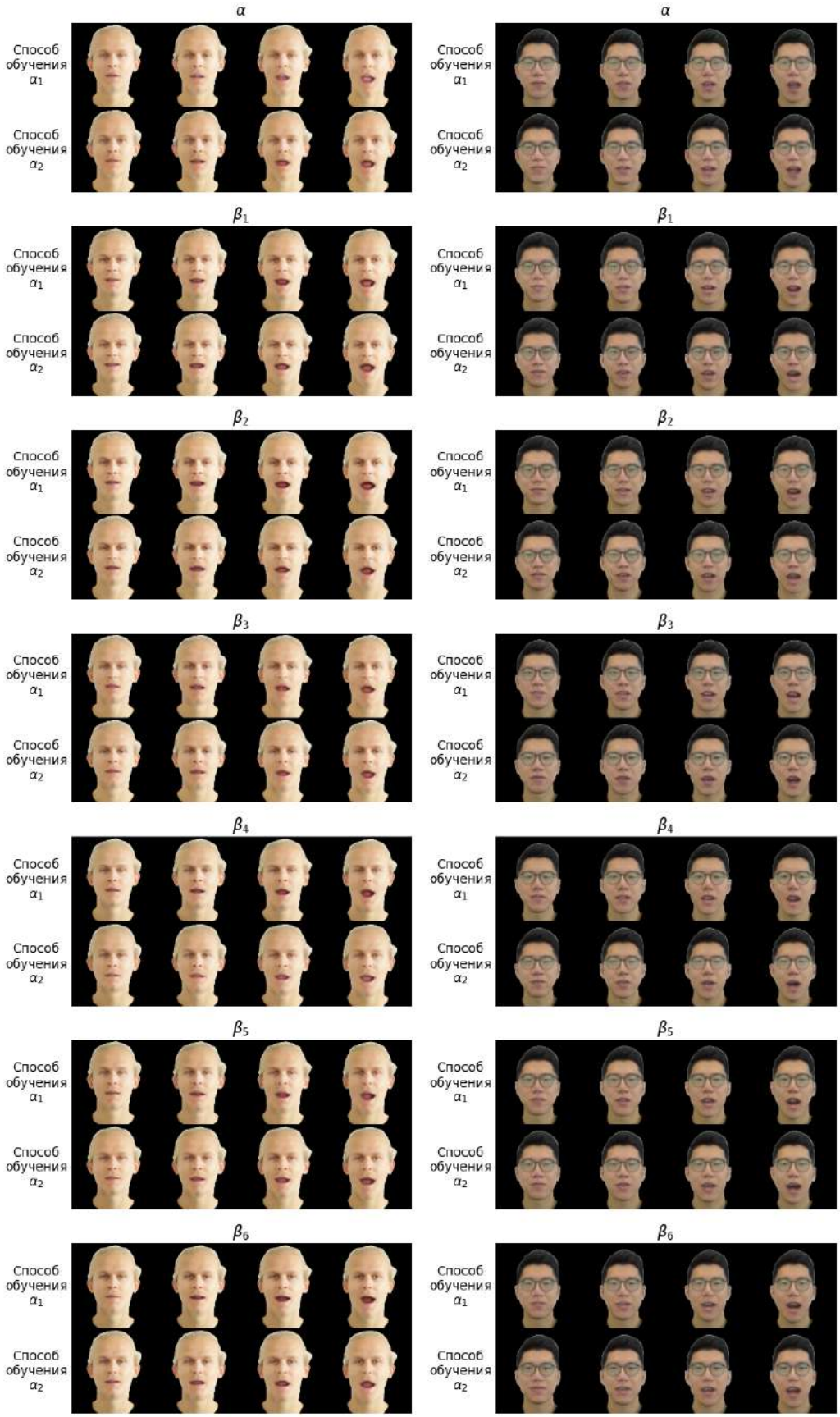


Рисунок А.31 – Результаты синтеза при варьировании степени открытия челюсти

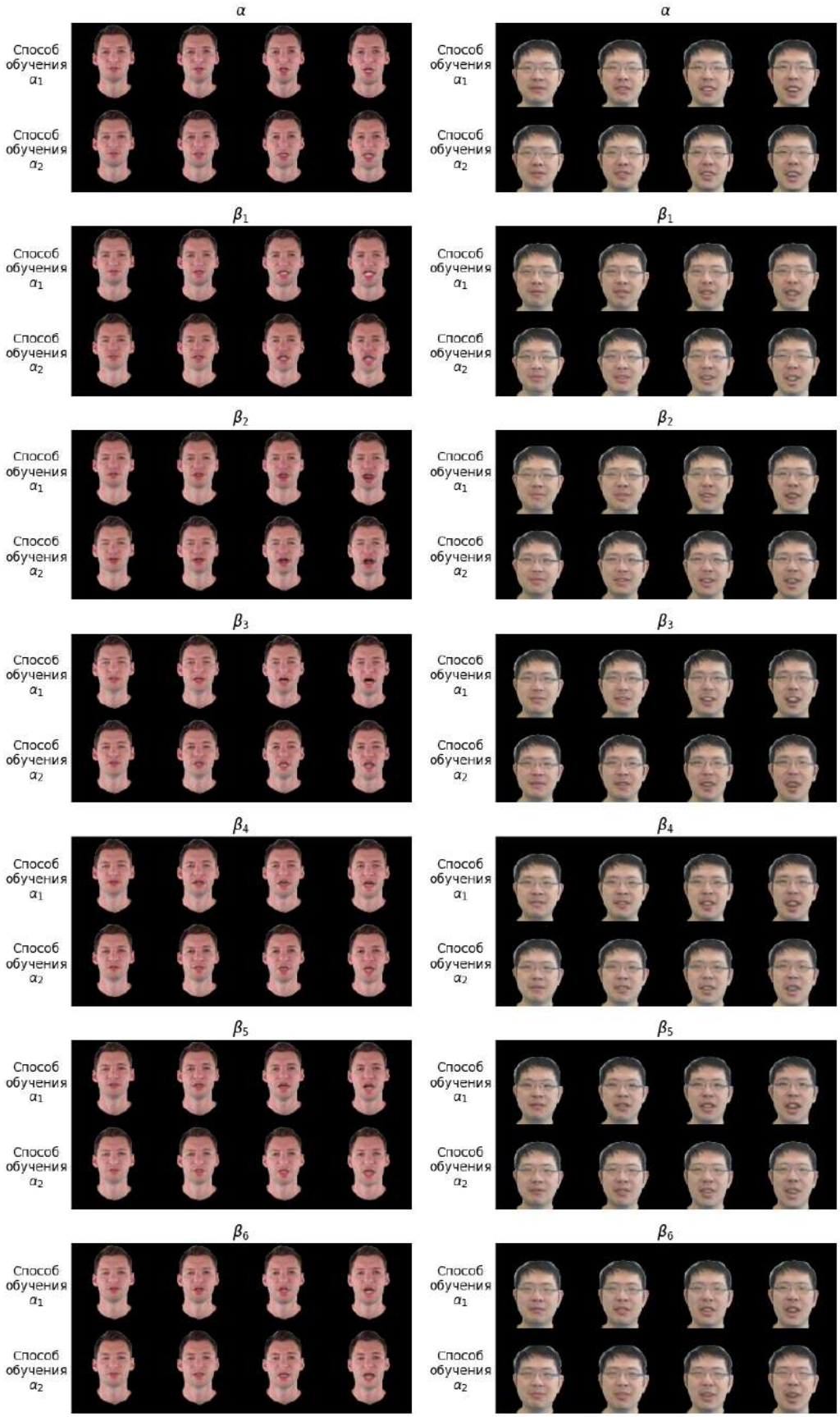


Рисунок А.32 – Результаты синтеза при варьировании степени открытия челюсти

Рисунок А.33 – Результаты синтеза в зависимости от разного удаления головы от камеры. Верхняя строка каждого блока – результат синтеза изображений-проекций, где инициализация производится из параметров, полученных без применения исследуемой аугментации; нижняя строка каждого блока – результат синтеза изображений-проекций, где инициализация производится из параметров, полученных с применением исследуемой аугментации

Рисунок А.34 – Результаты синтеза в зависимости от разного удаления головы от камеры. Верхняя строка каждого блока – результат синтеза изображений-проекций, где инициализация производится из параметров, полученных без применения исследуемой аугментации; нижняя строка каждого блока – результат синтеза изображений-проекций, где инициализация производится из параметров, полученных с применением исследуемой аугментации

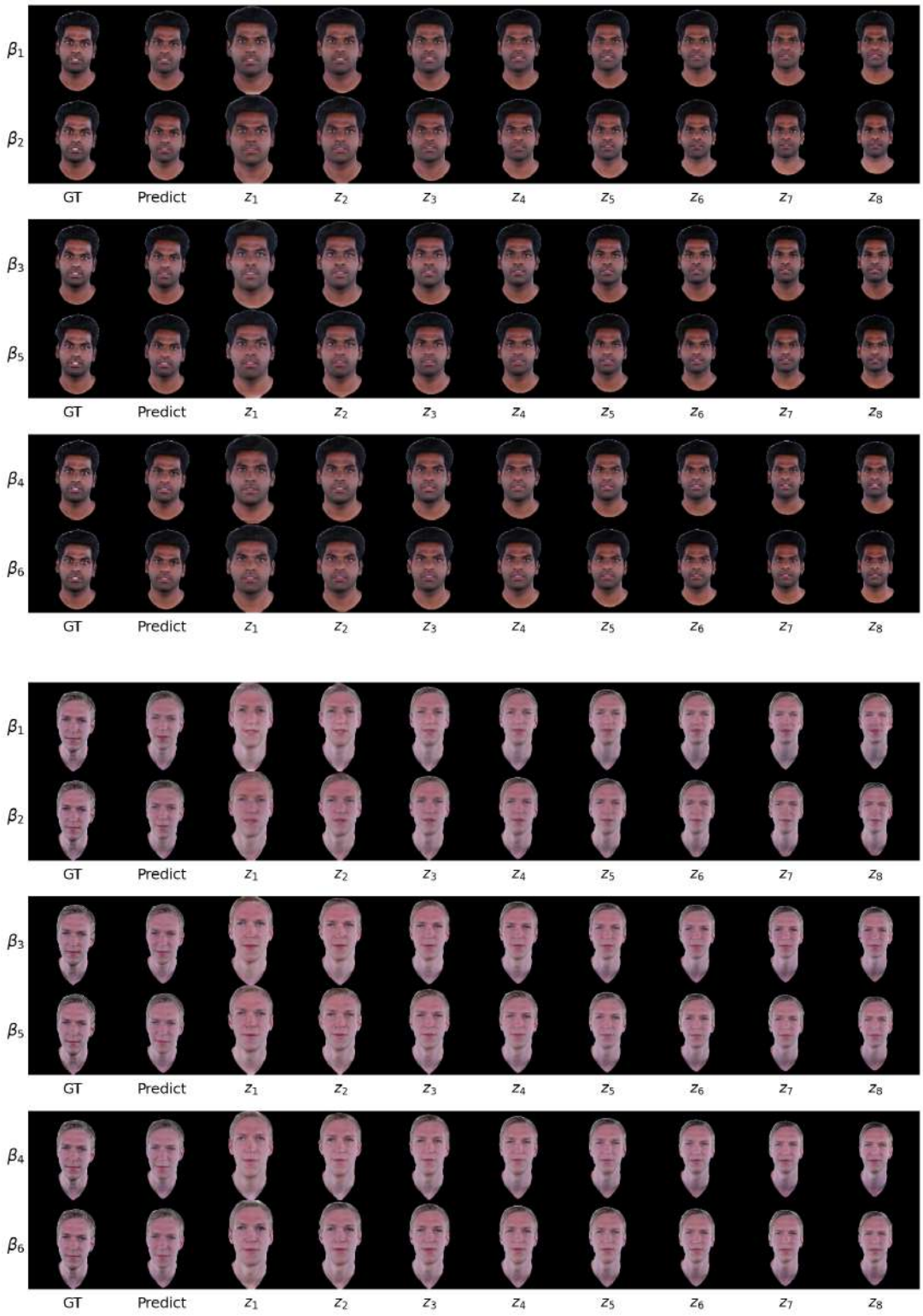


Рисунок А.35 – Результаты синтеза в зависимости от разного удаления головы от камеры. Верхняя строка каждого блока – результат синтеза изображений-проекций, где инициализация производится из параметров, полученных без применения исследуемой аугментации; нижняя строка каждого блока – результат синтеза изображений-проекций, где инициализация производится из параметров, полученных с применением исследуемой аугментации

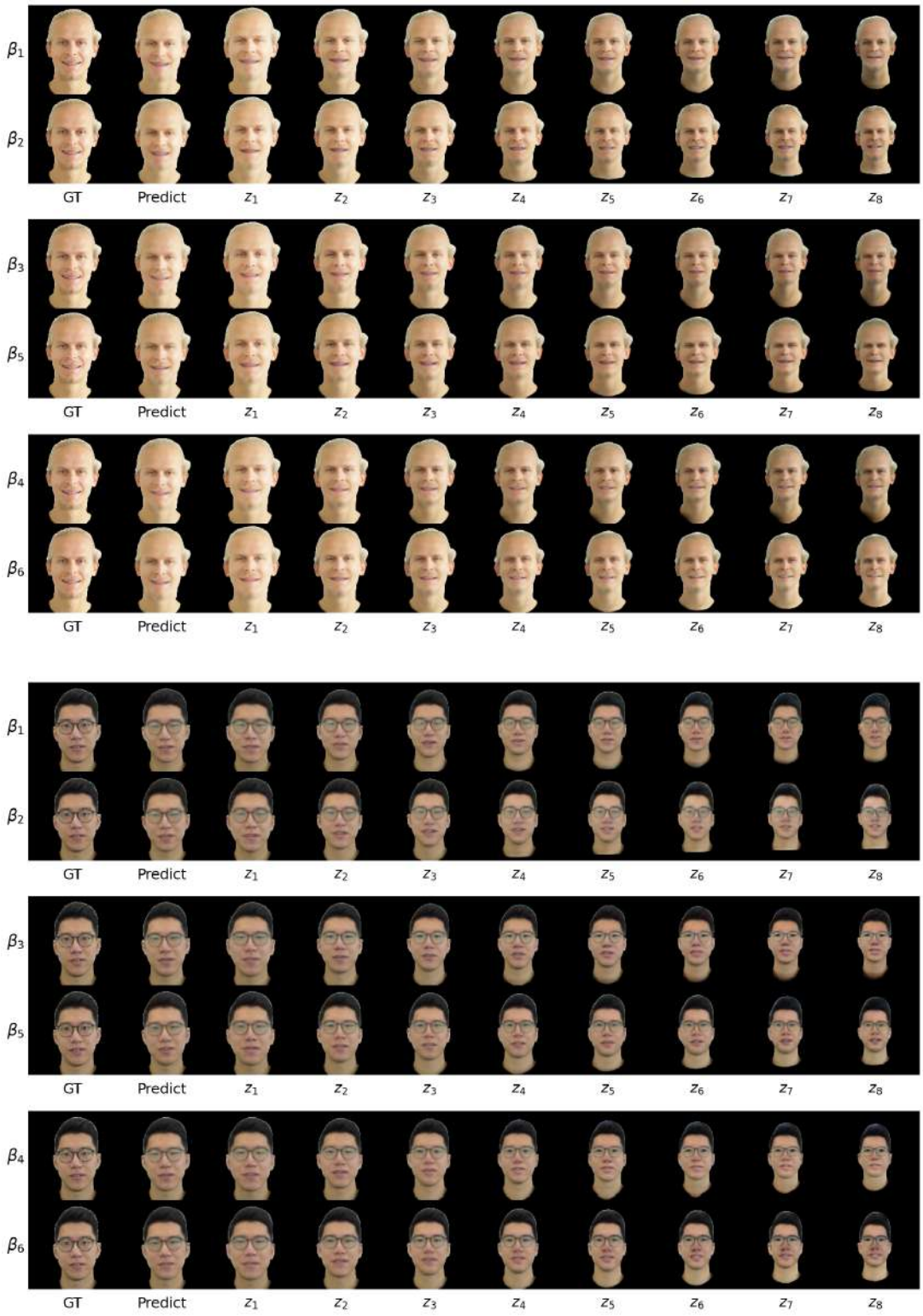


Рисунок А.36 – Результаты синтеза в зависимости от разного удаления головы от камеры. Верхняя строка каждого блока – результат синтеза изображений-проекций, где инициализация производится из параметров, полученных без применения исследуемой аугментации; нижняя строка каждого блока – результат синтеза изображений-проекций, где инициализация производится из параметров, полученных с применением исследуемой аугментации

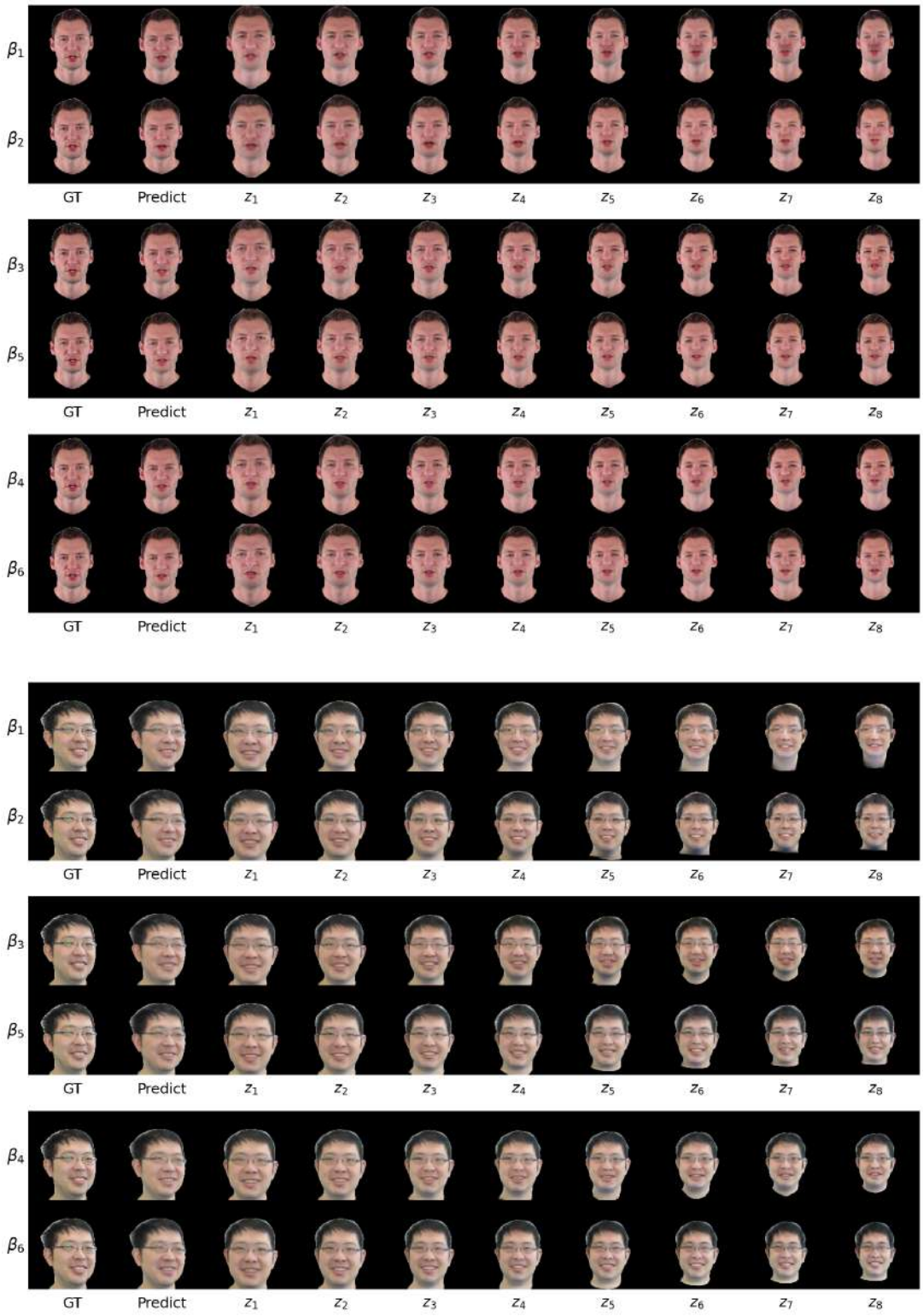


Рисунок А.37 – Результаты синтеза в зависимости от разного удаления головы от камеры. Верхняя строка каждого блока – результат синтеза изображений-проекций, где инициализация производится из параметров, полученных без применения исследуемой аугментации; нижняя строка каждого блока – результат синтеза изображений-проекций, где инициализация производится из параметров, полученных с применением исследуемой аугментации

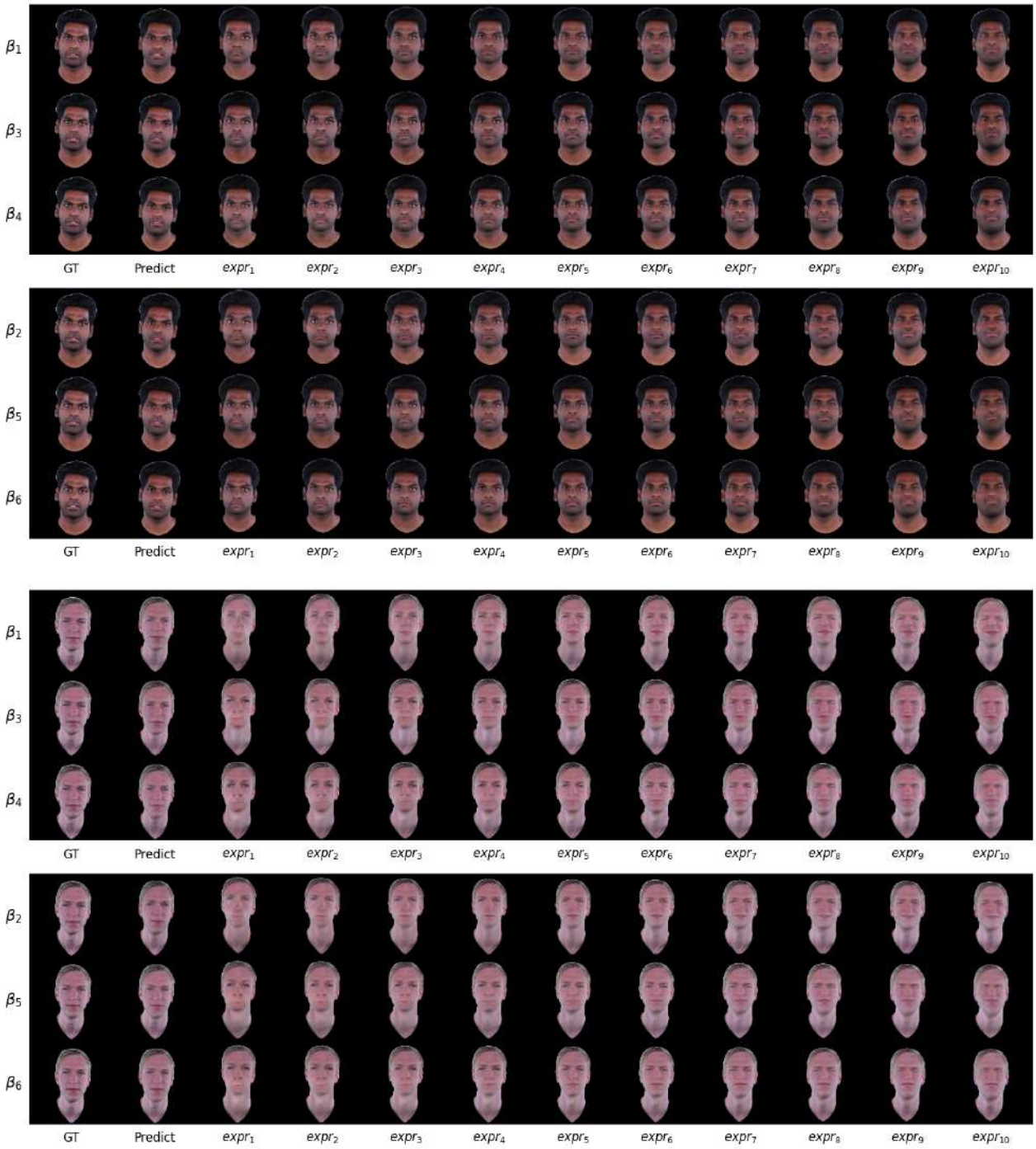


Рисунок А.38 – Результаты синтеза при варьировании выражения лица в допустимом диапазоне. Первая строка каждого блока – результат синтеза изображений-проекций, где инициализация производится из параметров, полученных без использования исследуемой стратегии; вторая и третья строка каждого блока – результат синтеза изображений-проекций, где инициализация производится из параметров, полученных с использованием исследуемой стратегии

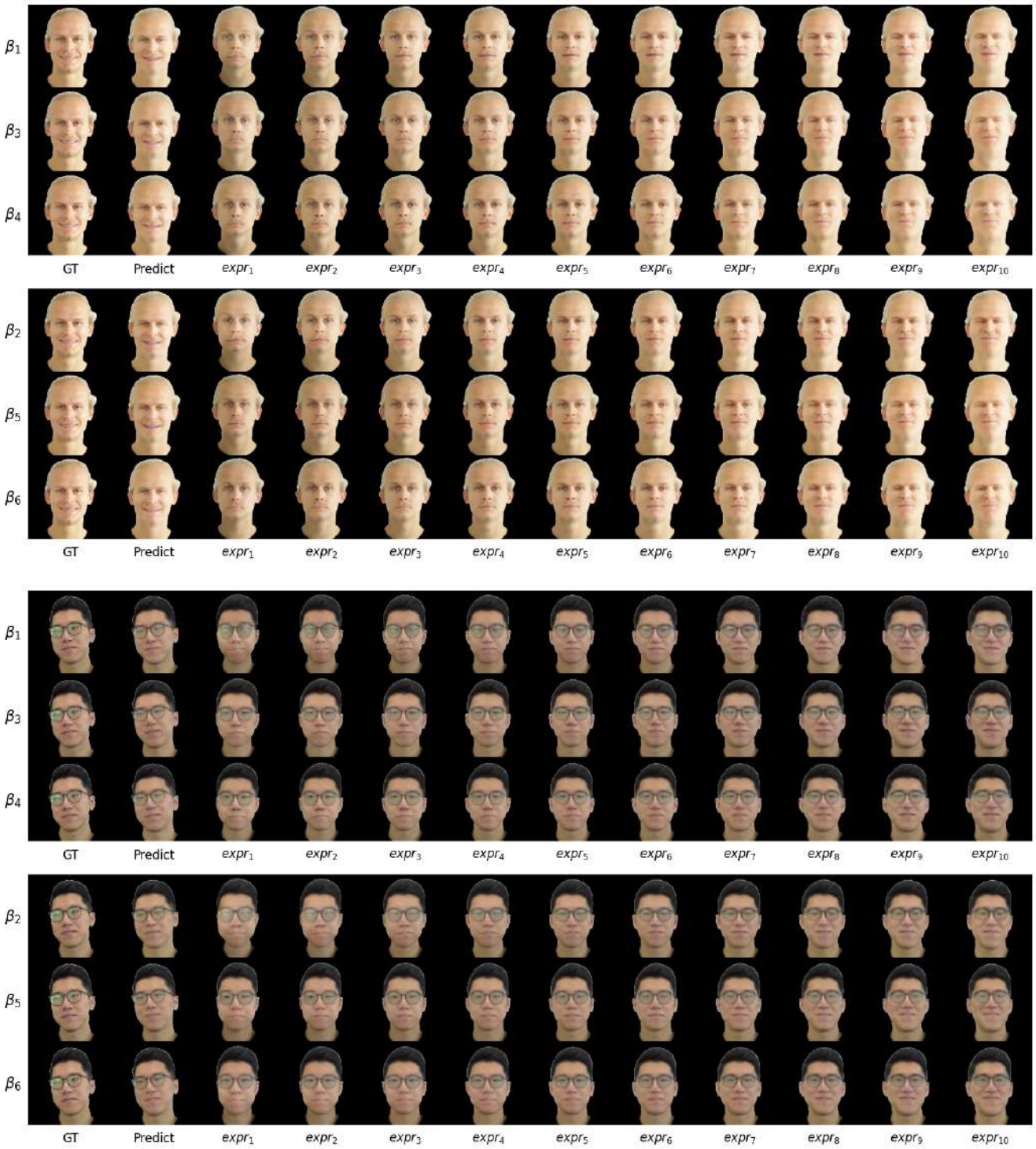


Рисунок А.39 – Результаты синтеза при варьировании выражения лица в допустимом диапазоне. Первая строка каждого блока – результат синтеза изображений-проекций, где инициализация производится из параметров, полученных без использования исследуемой стратегии; вторая и третья строка каждого блока – результат синтеза изображений-проекций, где инициализация производится из параметров, полученных с использованием исследуемой стратегии

Рисунок А.40 – Результаты синтеза при варьировании выражения лица в допустимом диапазоне. Первая строка каждого блока – результат синтеза изображений-проекций, где инициализация производится из параметров, полученных без использования исследуемой стратегии; вторая и третья строка каждого блока – результат синтеза изображений-проекций, где инициализация производится из параметров, полученных с использованием исследуемой стратегии

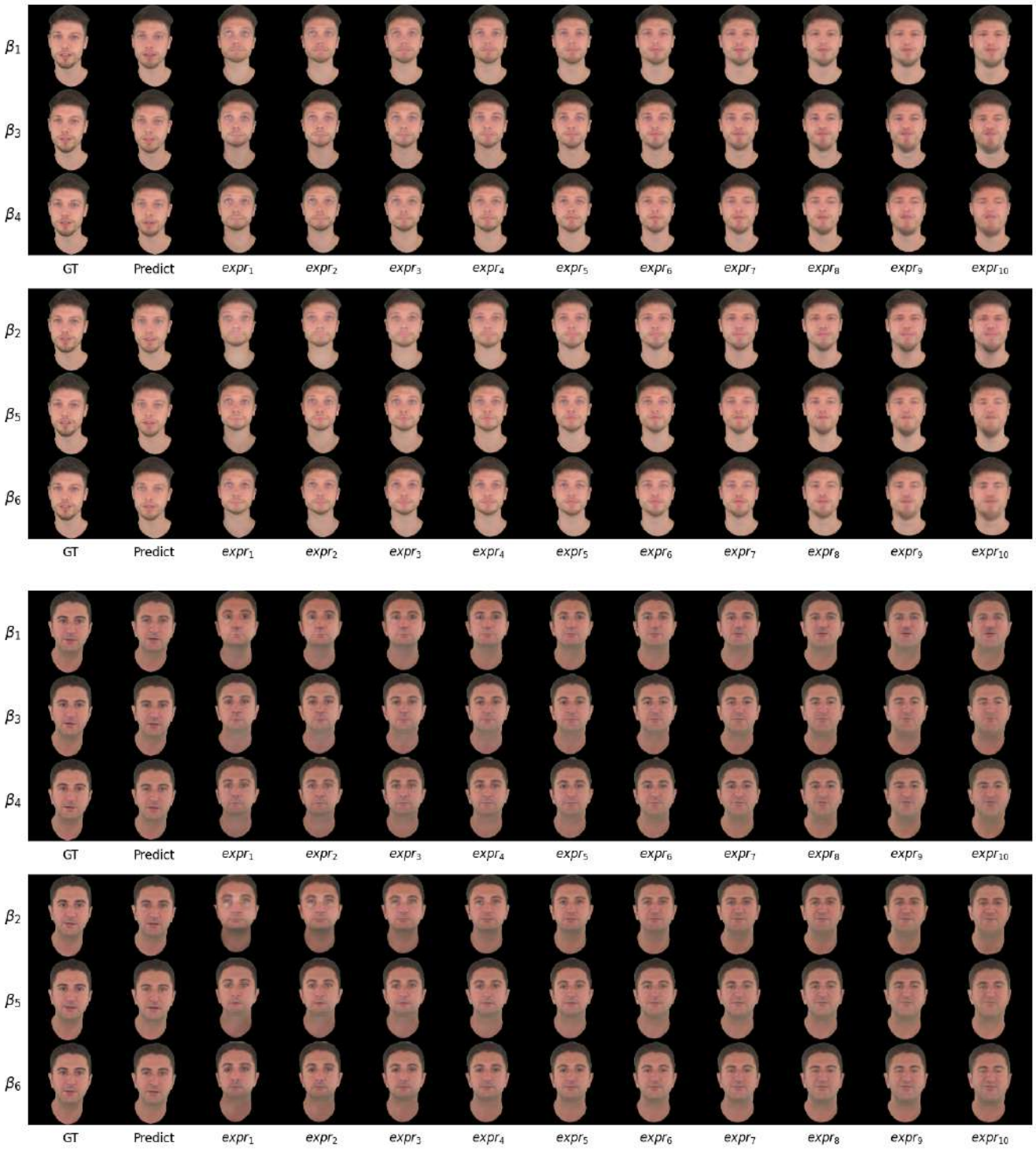


Рисунок А.41 – Результаты синтеза при варьировании выражения лица в допустимом диапазоне. Первая строка каждого блока – результат синтеза изображений-проекций, где инициализация производится из параметров, полученных без использования исследуемой стратегии; вторая и третья строка каждого блока – результат синтеза изображений-проекций, где инициализация производится из параметров, полученных с использованием исследуемой стратегии

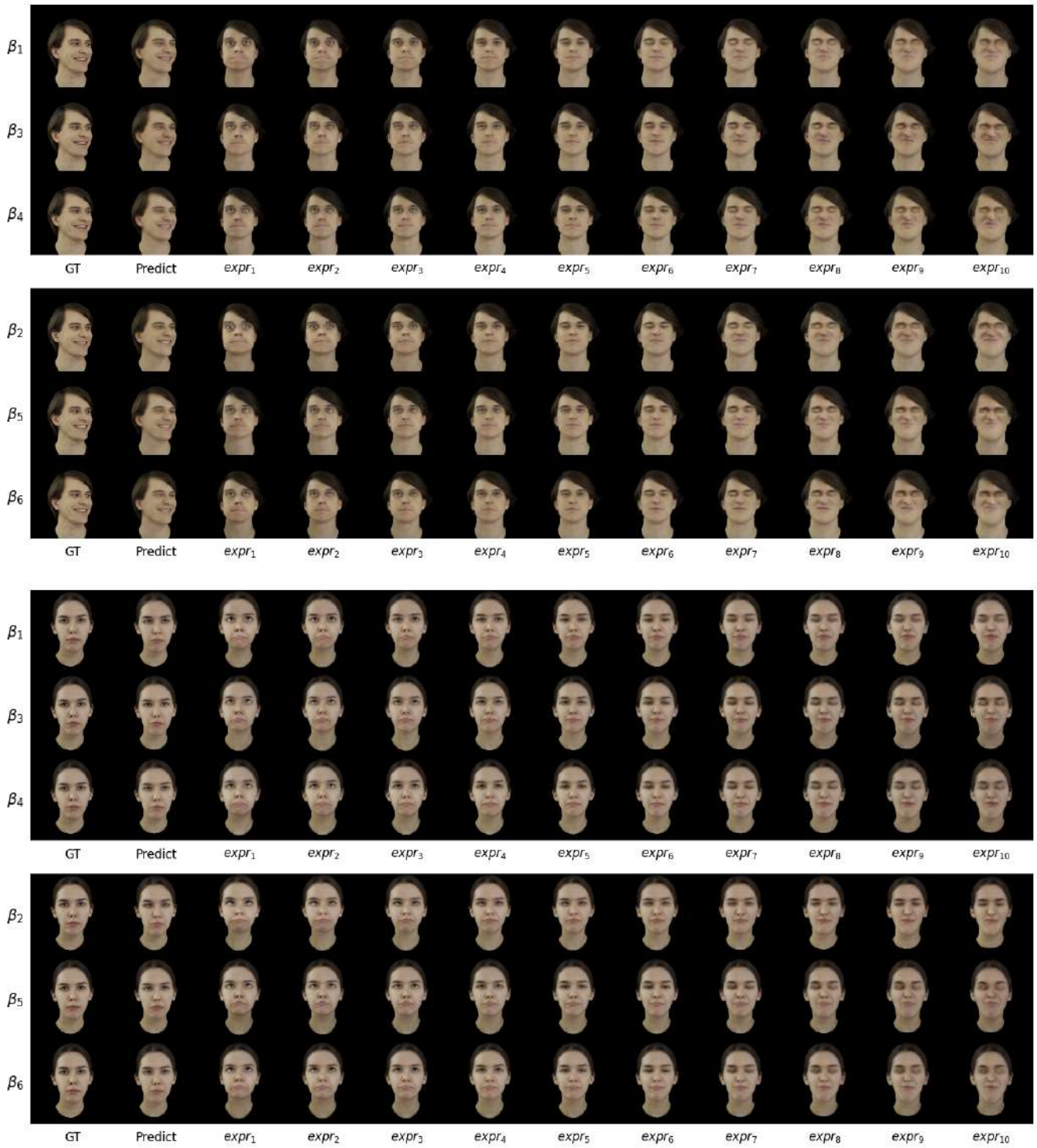


Рисунок А.42 – Результаты синтеза при варьировании выражения лица в допустимом диапазоне. Первая строка каждого блока – результат синтеза изображений-проекций, где инициализация производится из параметров, полученных без использования исследуемой стратегии; вторая и третья строка каждого блока – результат синтеза изображений-проекций, где инициализация производится из параметров, полученных с использованием исследуемой стратегии

Рисунок А.43 – Результаты синтеза на основе параметров модели FLAME вадидационной выборки. Первый столбец – результат синтеза изображений-проекций, где инициализации производится из параметров β_1, β_3, β_4. Второй столбец – результат синтеза изображений-проекций, где инициализация производится из параметров β_2, β_5, β_6

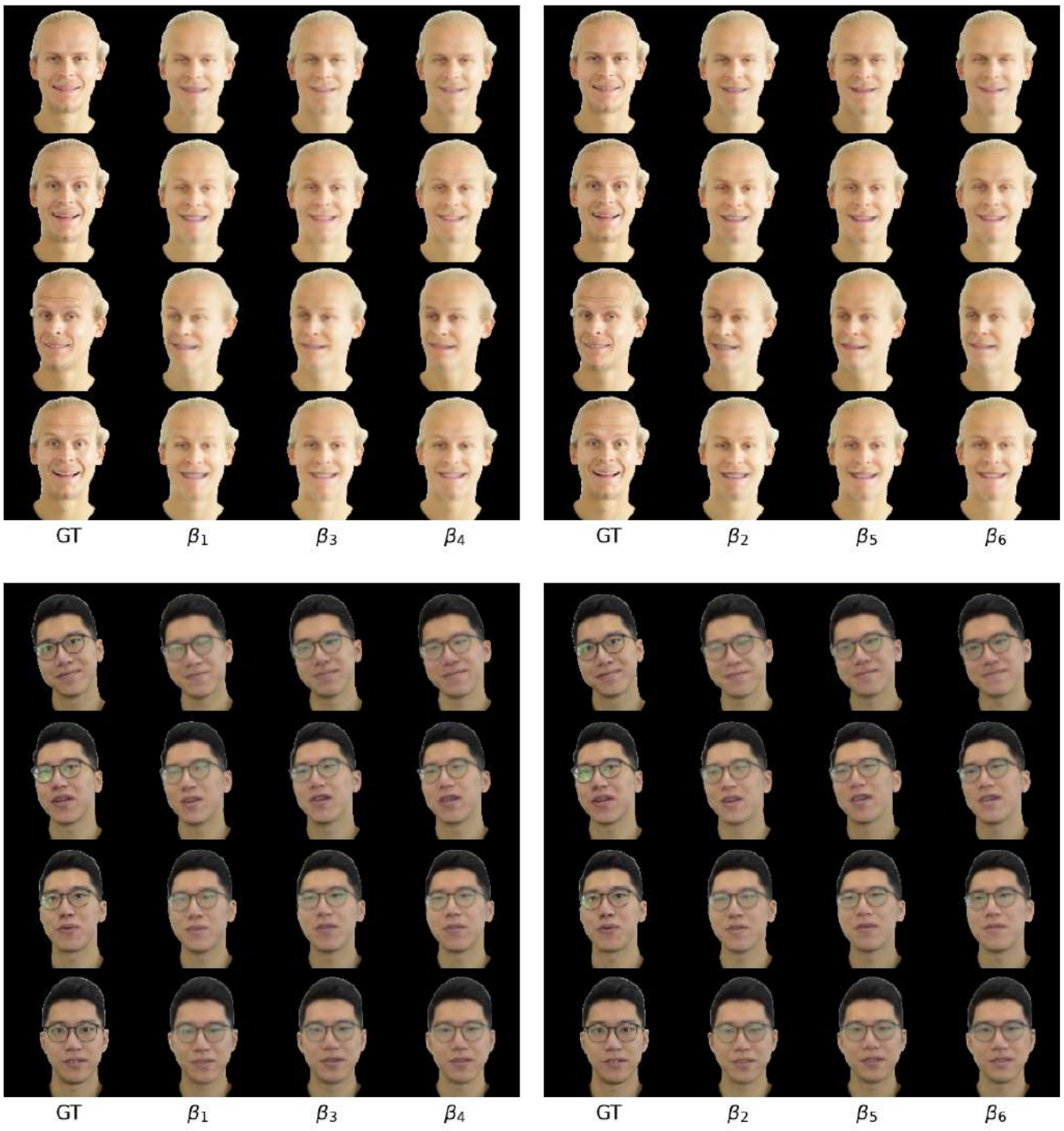


Рисунок А.44 – Результаты синтеза на основе параметров модели FLAME вадидационной выборки. Первый столбец – результат синтеза изображений-проекций, где инициализации производится из параметров β_1, β_3, β_4. Второй столбец – результат синтеза изображений-проекций, где инициализация производится из параметров β_2, β_5, β_6

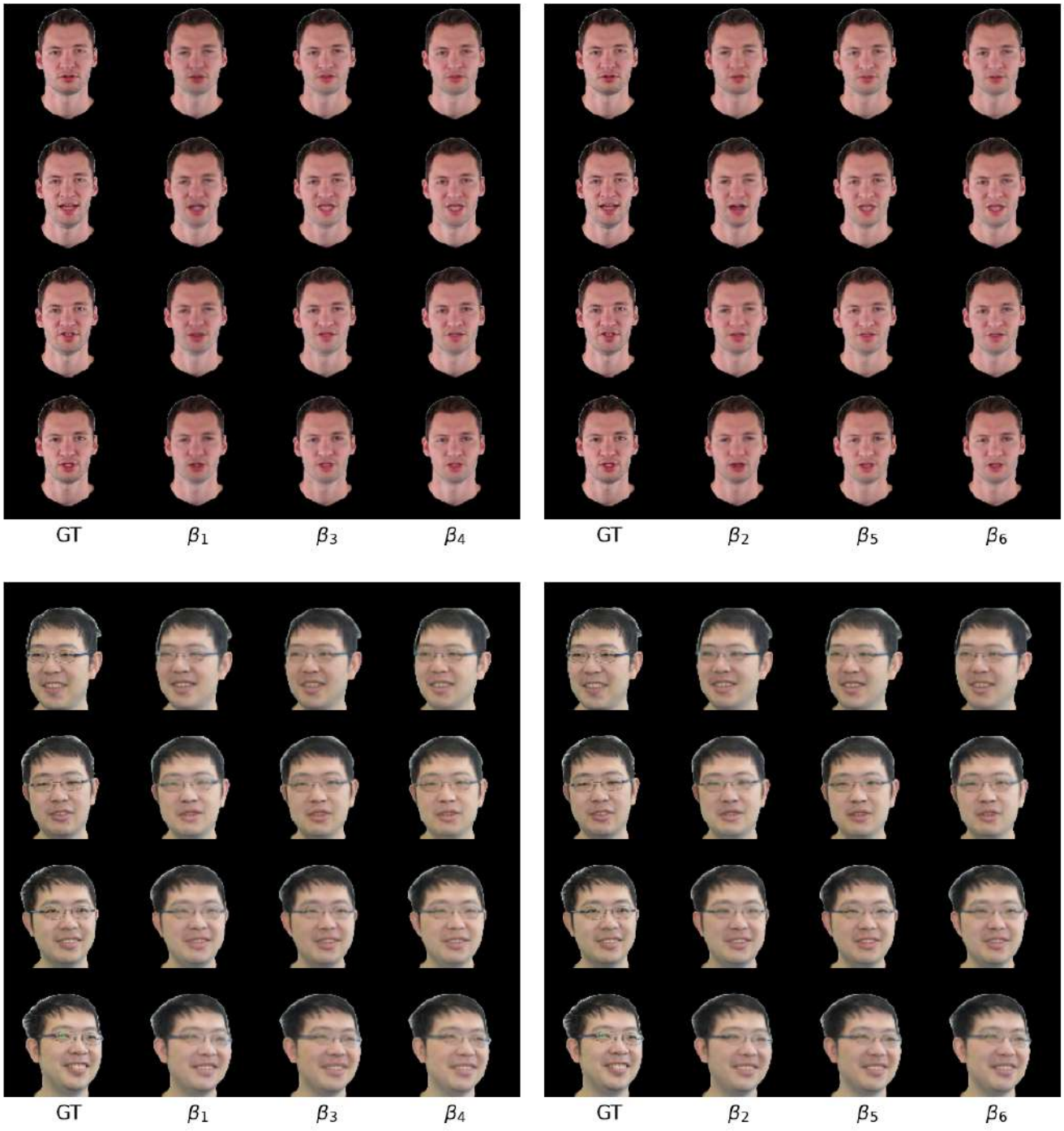


Рисунок А.45 – Результаты синтеза на основе параметров модели FLAME вадидационной выборки. Первый столбец – результат синтеза изображенй-проекций, где инициализации производится из параметров β_1, β_3, β_4. Второй столбец – результат синтеза изображений-проекций, где инициализация производится из параметров β_2, β_5, β_6

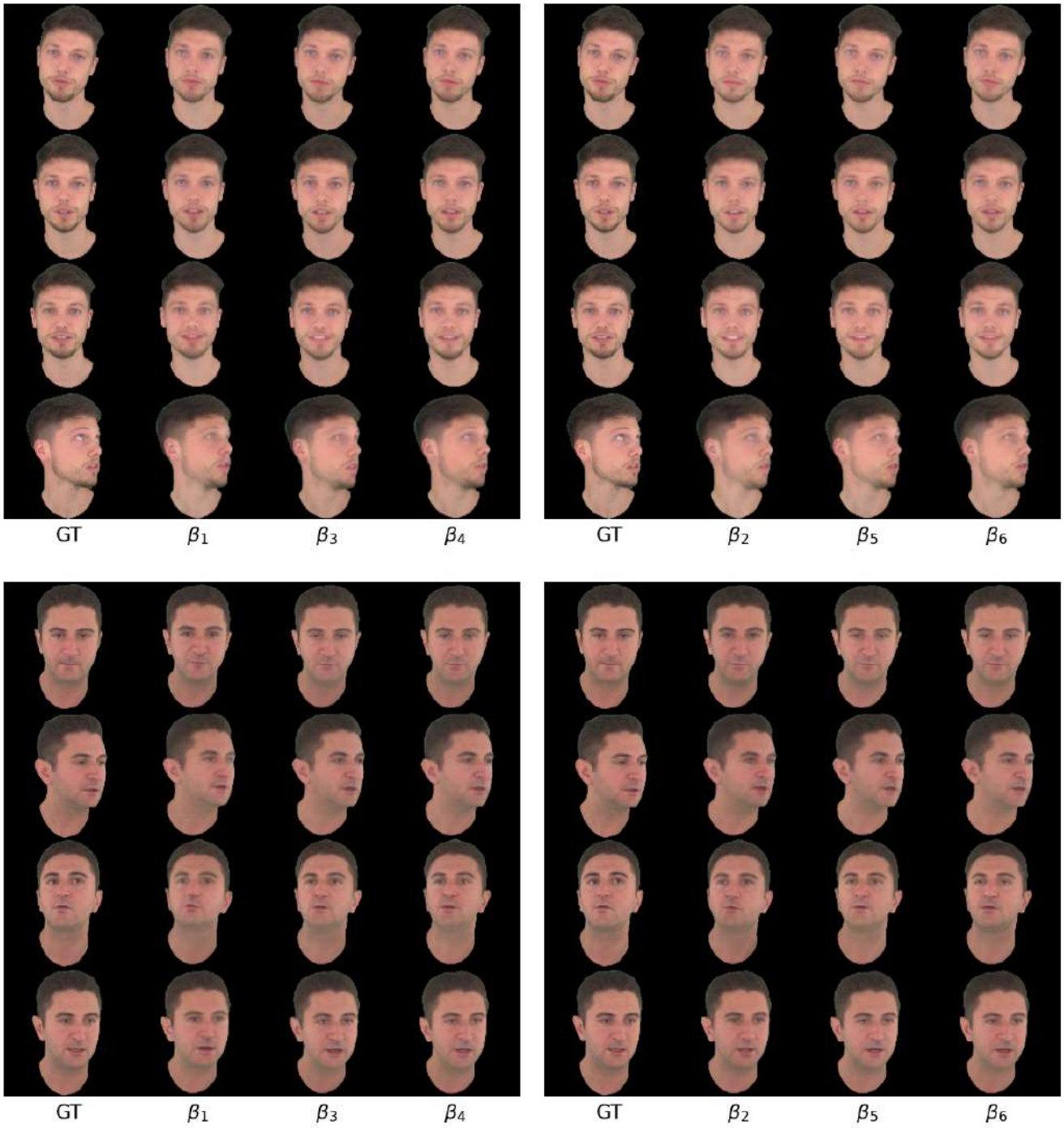


Рисунок А.46 – Результаты синтеза на основе параметров модели FLAME вадидационной выборки. Первый столбец – результат синтеза изображенй-проекций, где инициализации производится из параметров β_1, β_3, β_4. Второй столбец – результат синтеза изображений-проекций, где инициализация производится из параметров β_2, β_5, β_6

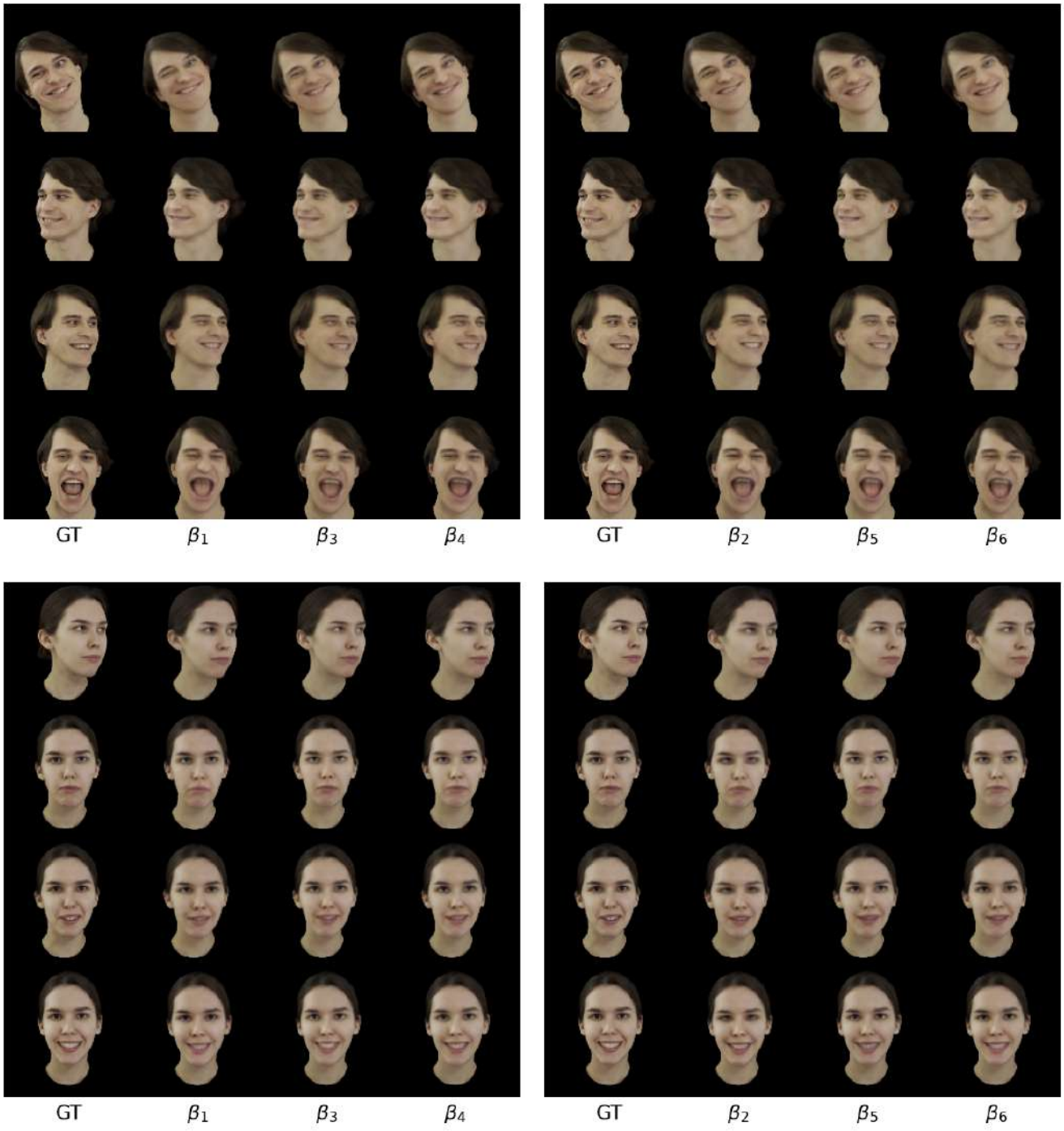


Рисунок А.47 – Результаты синтеза на основе параметров модели FLAME вадидационной выборки. Первый столбец – результат синтеза изображенй-проекций, где инициализации производится из параметров β_1, β_3, β_4. Второй столбец – результат синтеза изображений-проекций, где инициализация производится из параметров β_2, β_5, β_6

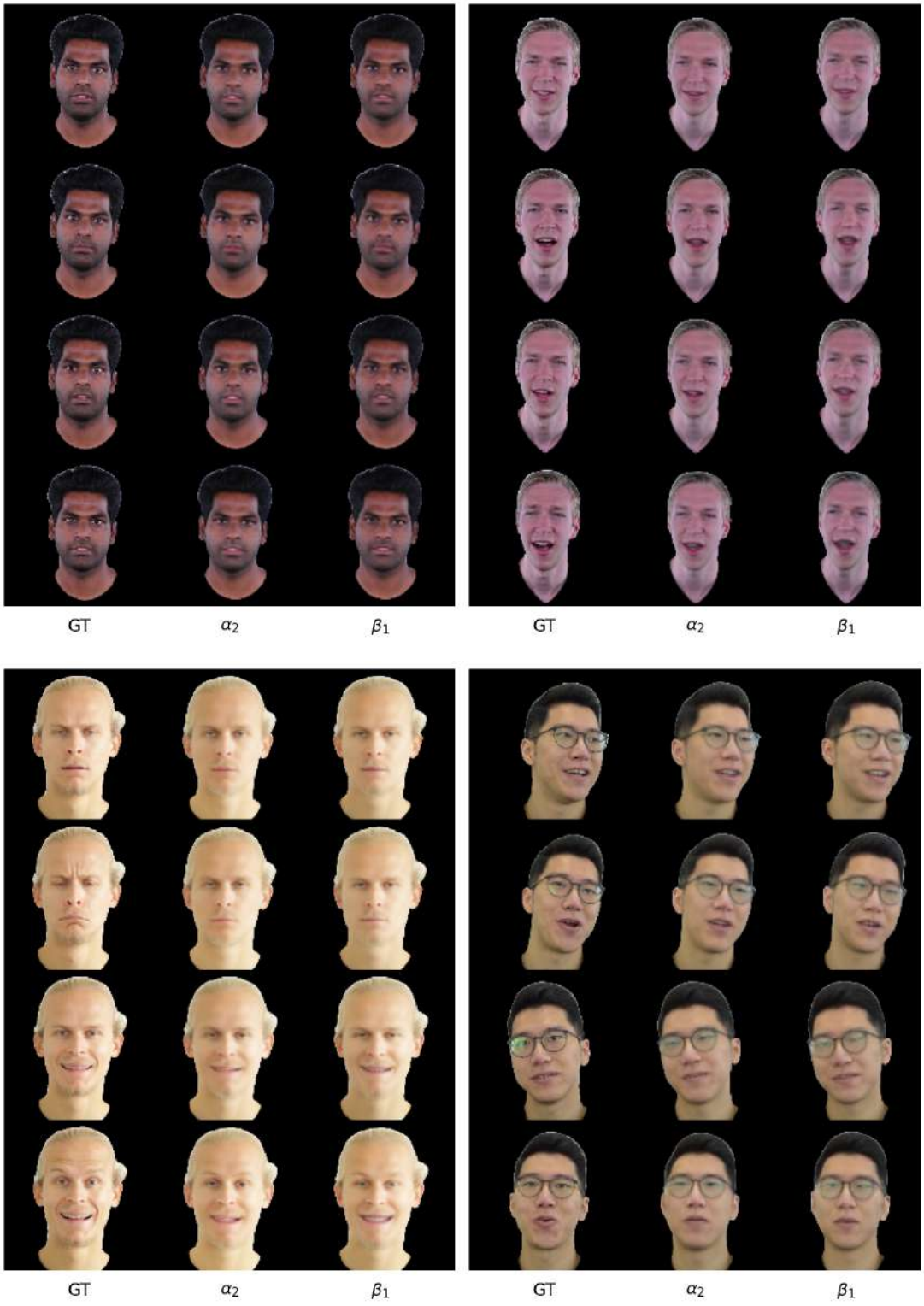


Рисунок А.48 – Результаты синтеза по параметрам, соответствующим отобранным из валидационных выборок изображений, для экспериментов, инициализация параметров которых производилась из β_1 и α_2

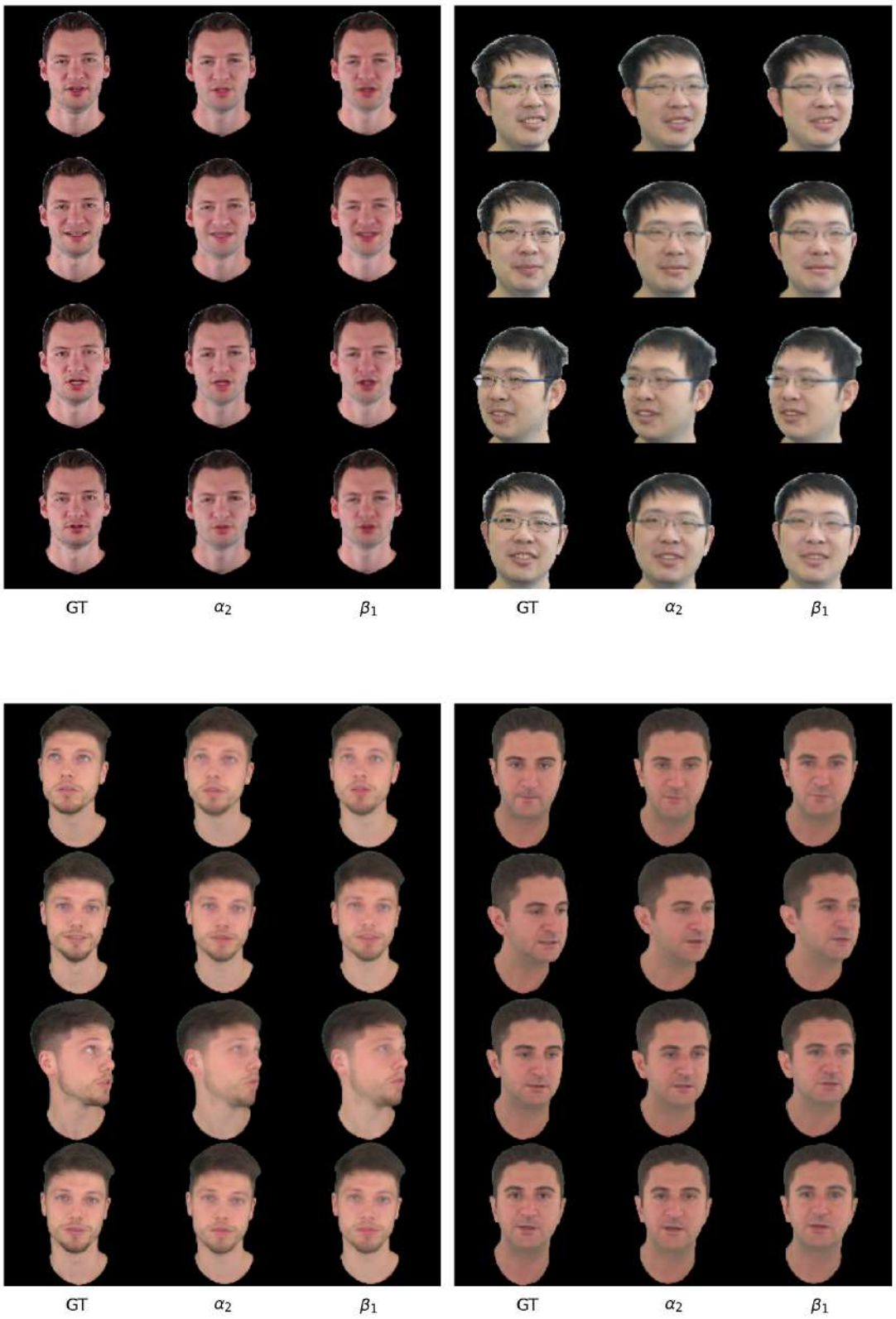


Рисунок А.49 – Результаты синтеза по параметрам, соответствующим отобранным из валидационных выборок изображений, для экспериментов, инициализация параметров которых производилась из β_1 и α_2

Рисунок А.50 – Результаты синтеза по параметрам, соответствующим отобранным из валидационных выборок изображений, для экспериментов, инициализация параметров которых производилась из β_1 и α_2

Рисунок А.51 – Результаты синтеза в зависимости от разного удаления головы от камеры для разрешения 128×128

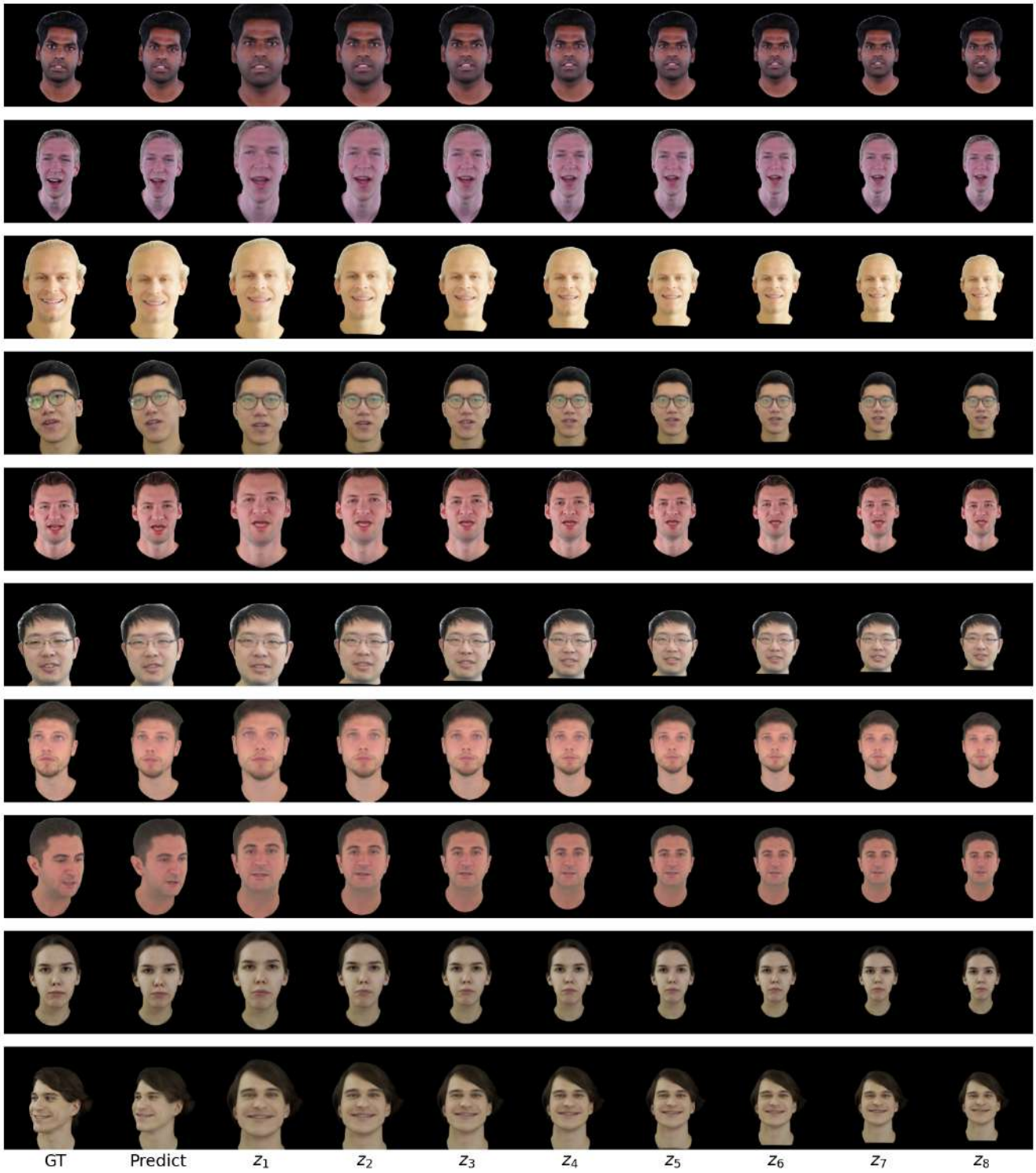


Рисунок А.52 – Результаты синтеза в зависимости от разного удаления головы от камеры для разрешения 256×256

Рисунок А.53 – Результаты синтеза в зависимости от разного удаления головы от камеры для разрешения 512×512

Рисунок А.54 – Результаты синтеза при варьировании параметров выражения лица в допустимом диапазоне для разрешения 128×128

Рисунок А.55 – Результаты синтеза при варьировании параметров выражения лица в допустимом диапазоне для разрешения 256×256

Рисунок А.56 – Результаты синтеза при варьировании параметров выражения лица в допустимом диапазоне для разрешения 512×512

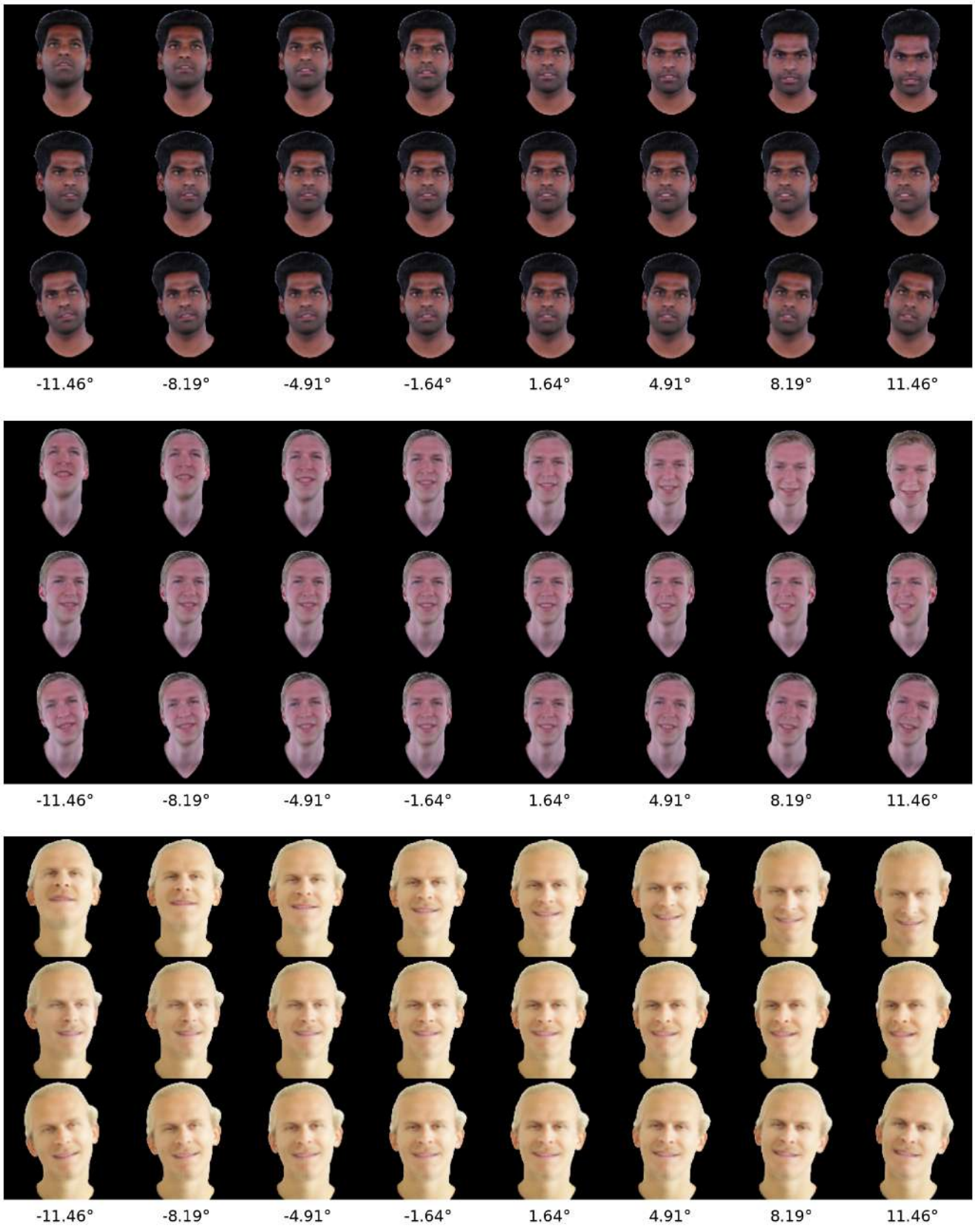


Рисунок А.57 – Результаты синтеза при варьировании поворота шеи для разрешения 128×128

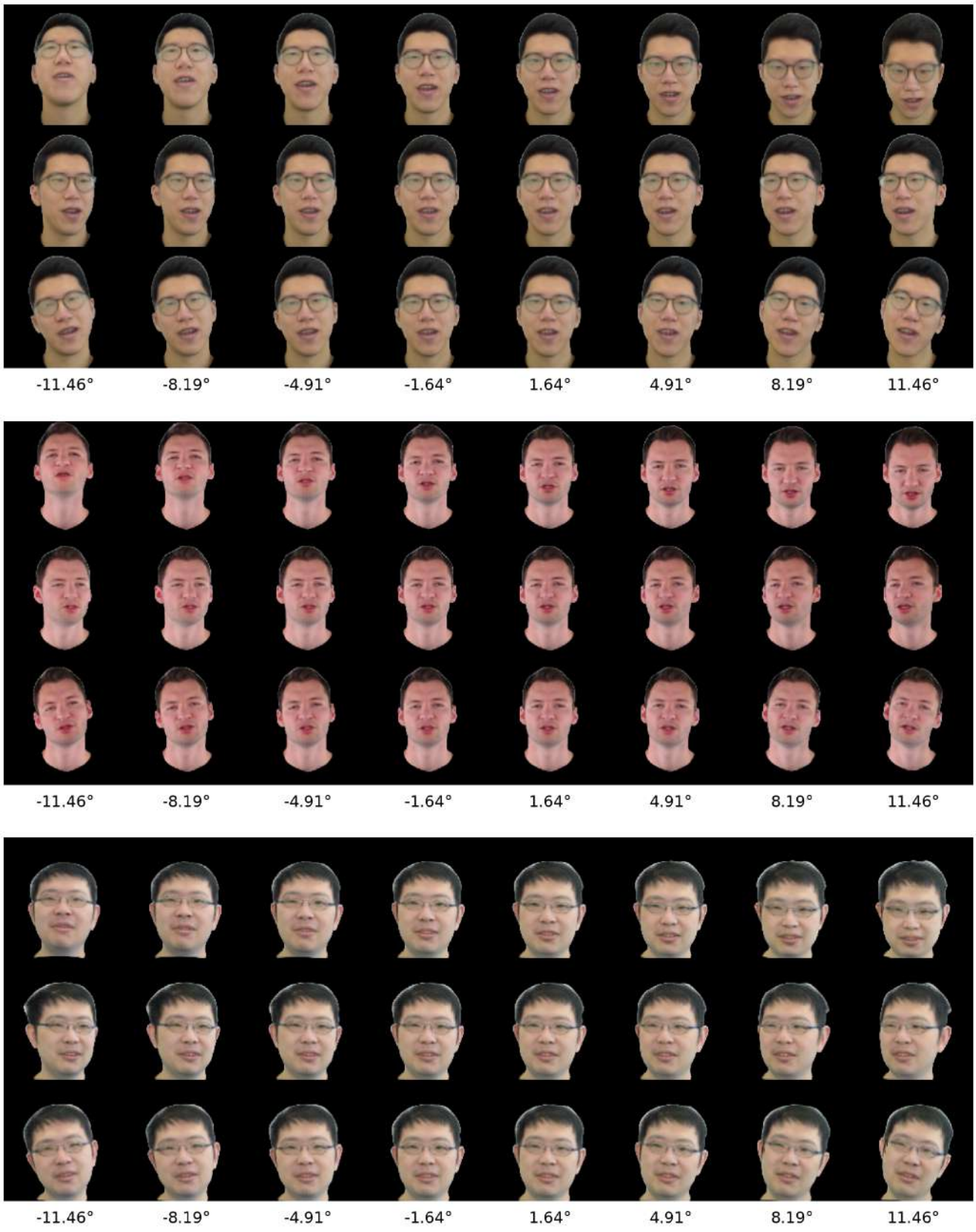


Рисунок А.58 – Результаты синтеза при варьировании поворота шеи для разрешения 128×128

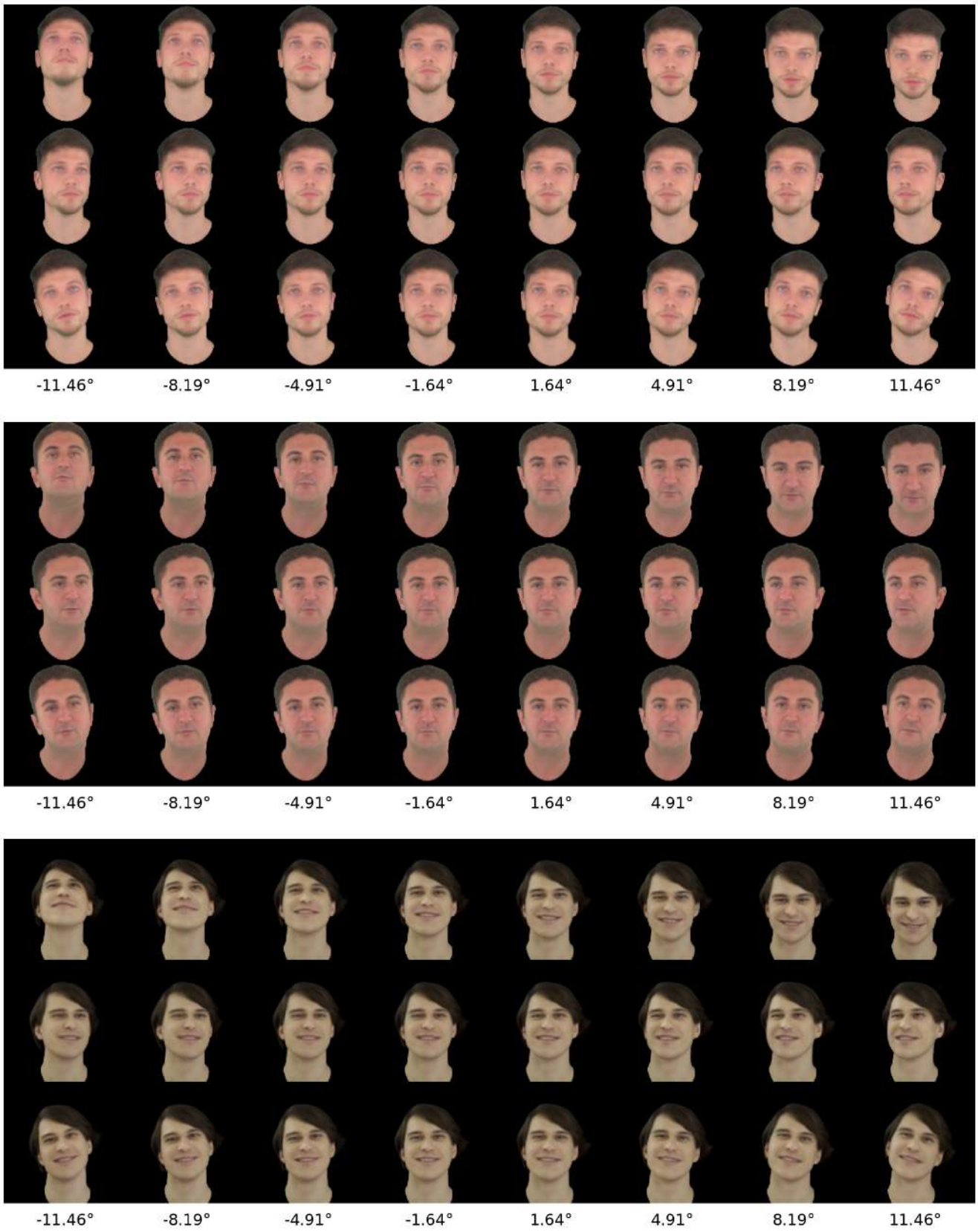


Рисунок А.59 – Результаты синтеза при варьировании поворота шеи для разрешения 128×128

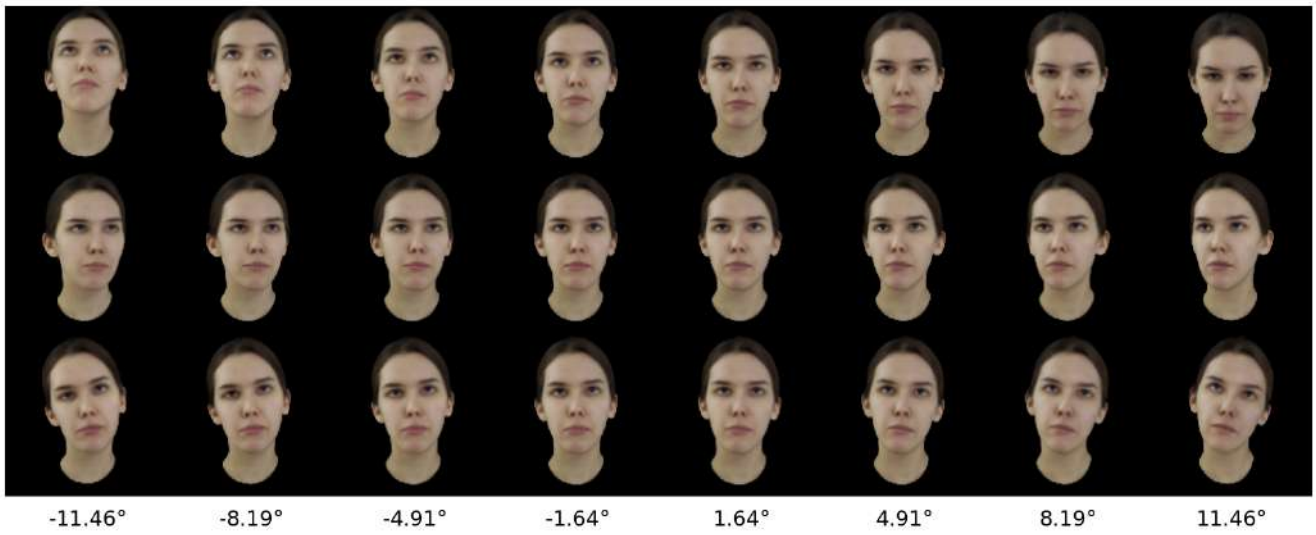


Рисунок А.60 – Результаты синтеза при варьировании поворота шеи для разрешения 128×128

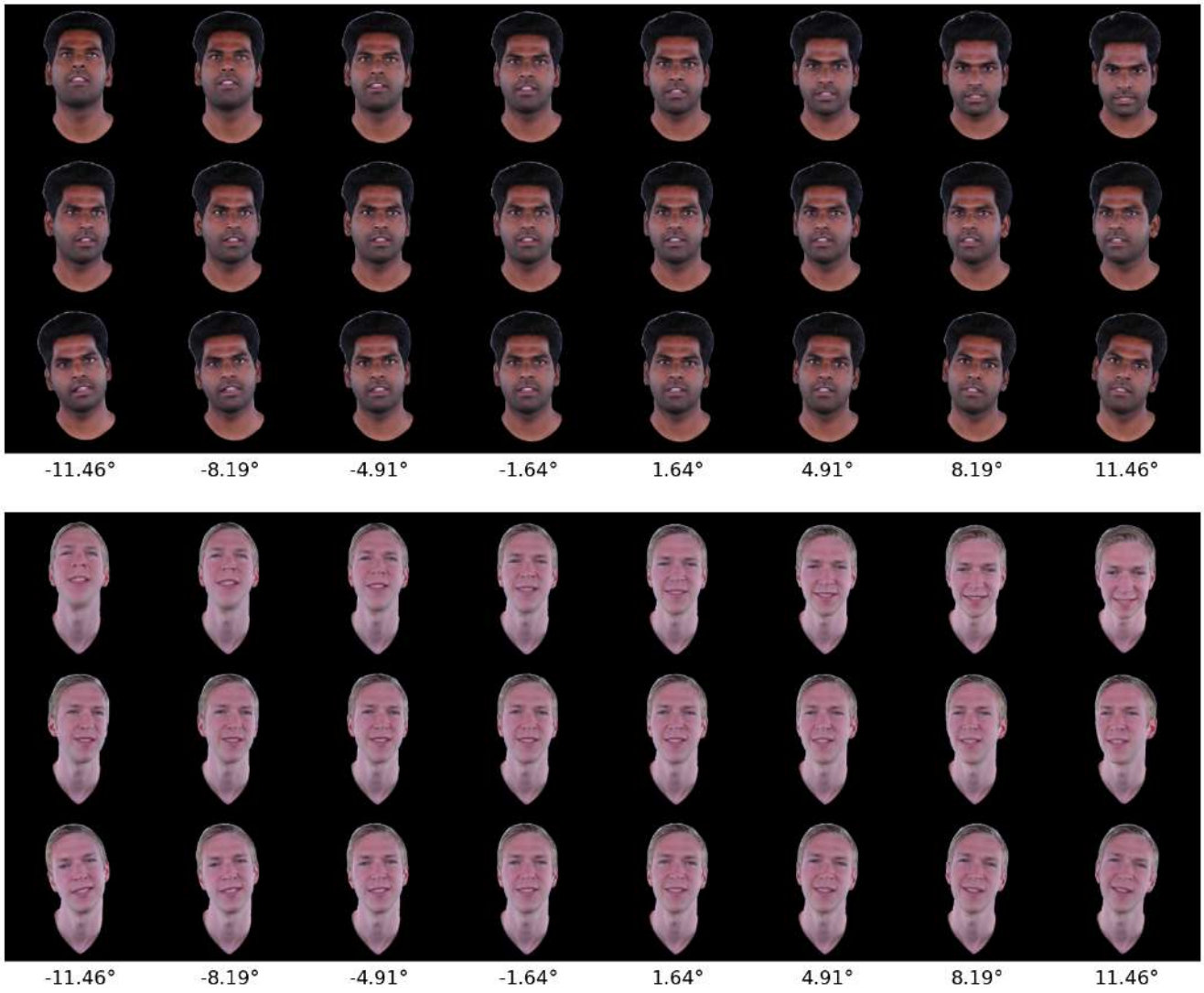


Рисунок А.61 – Результаты синтеза при варьировании поворота шеи для разрешения 256×256

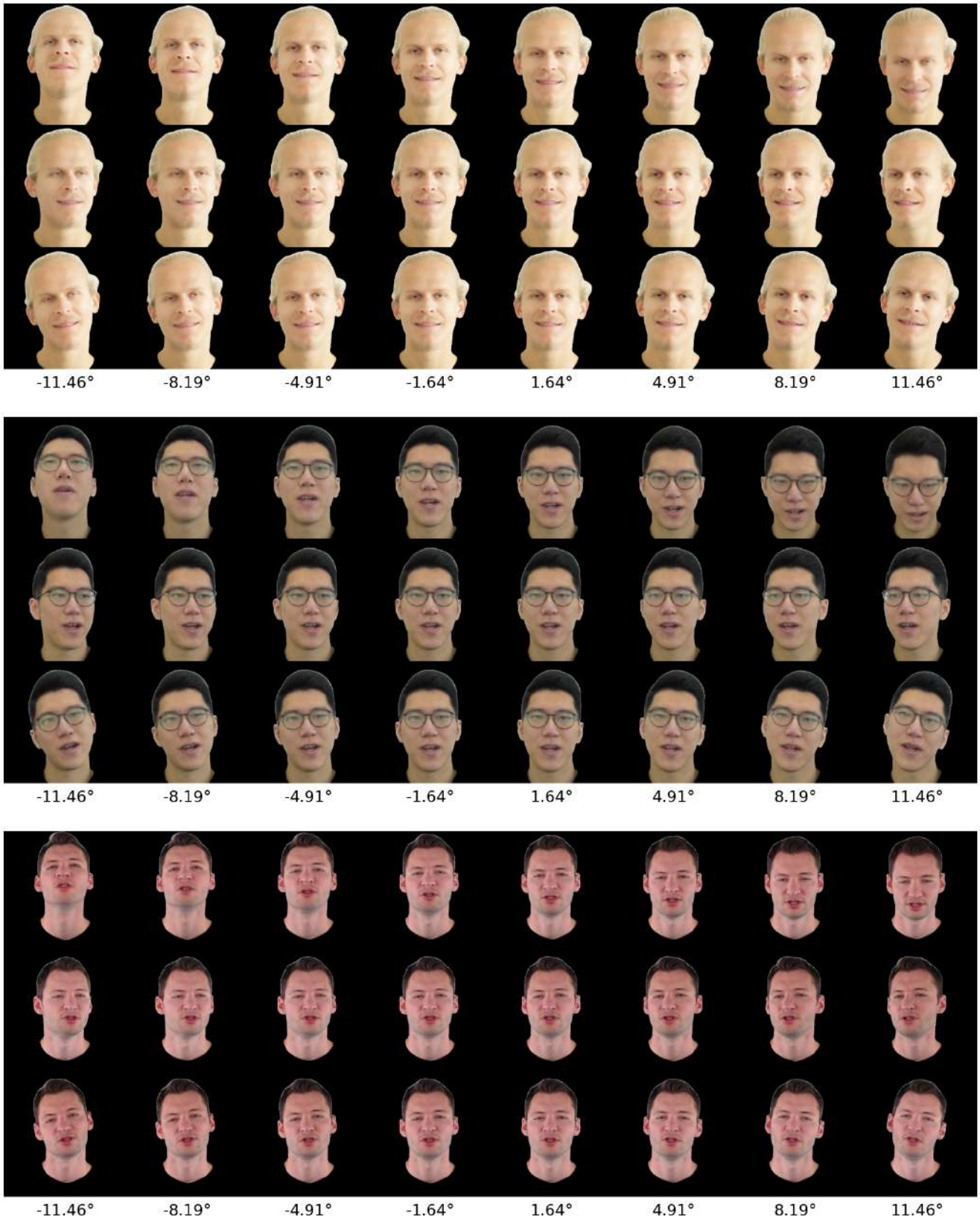


Рисунок А.62 – Результаты синтеза при варьировании поворота шеи для разрешения 256×256

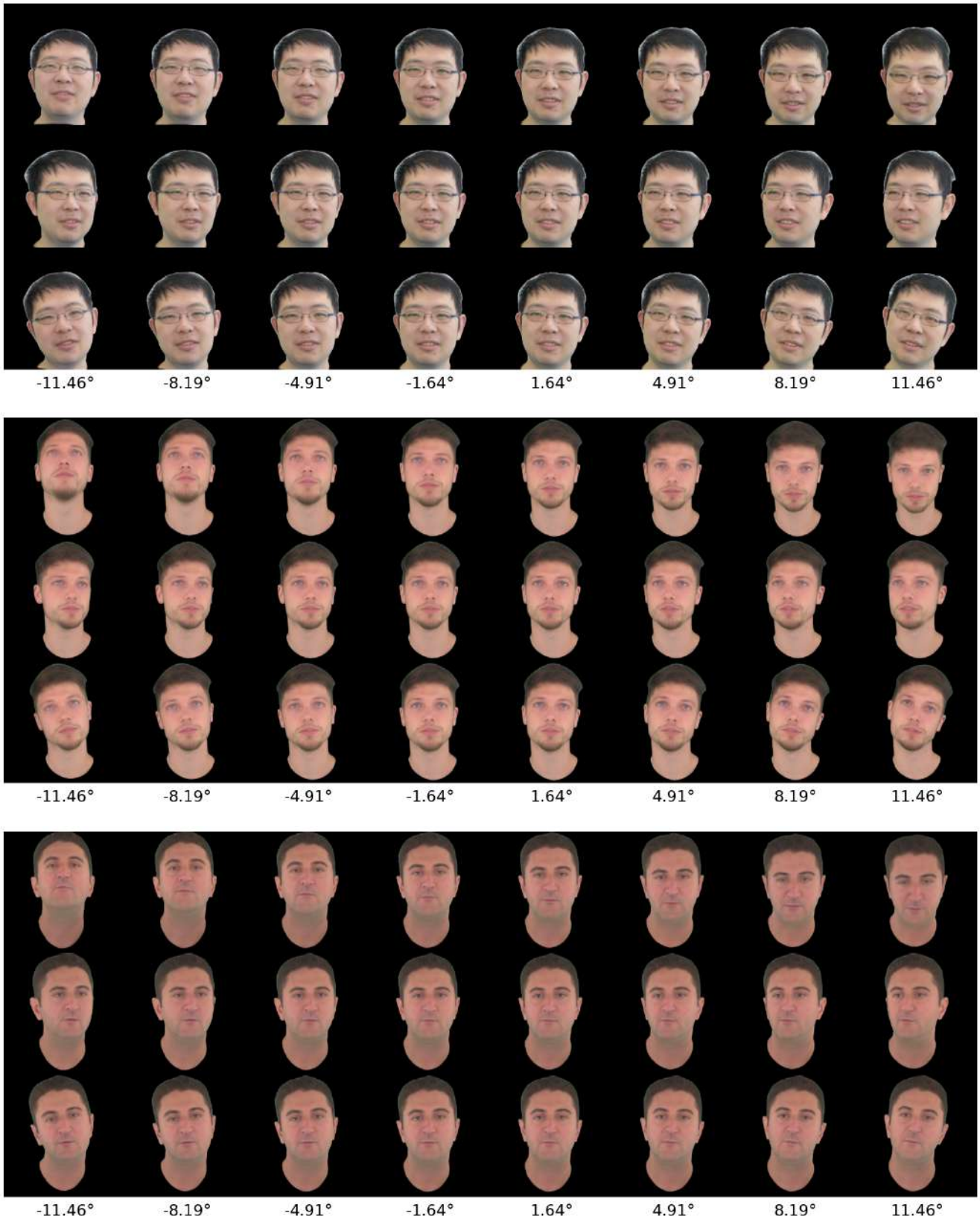
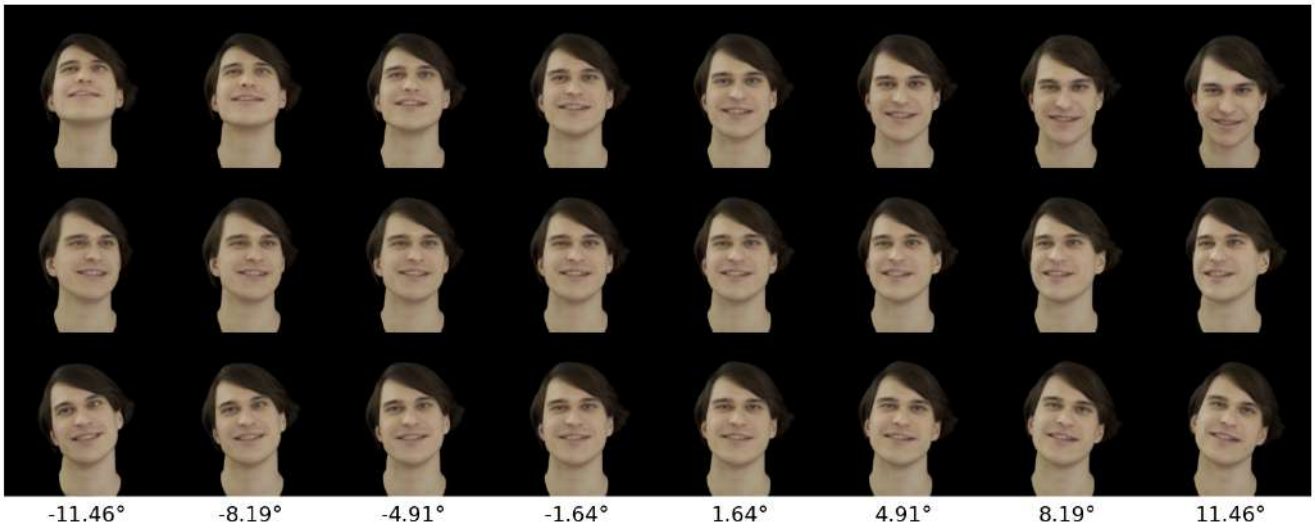


Рисунок А.63 – Результаты синтеза при варьировании поворота шеи для разрешения 256×256



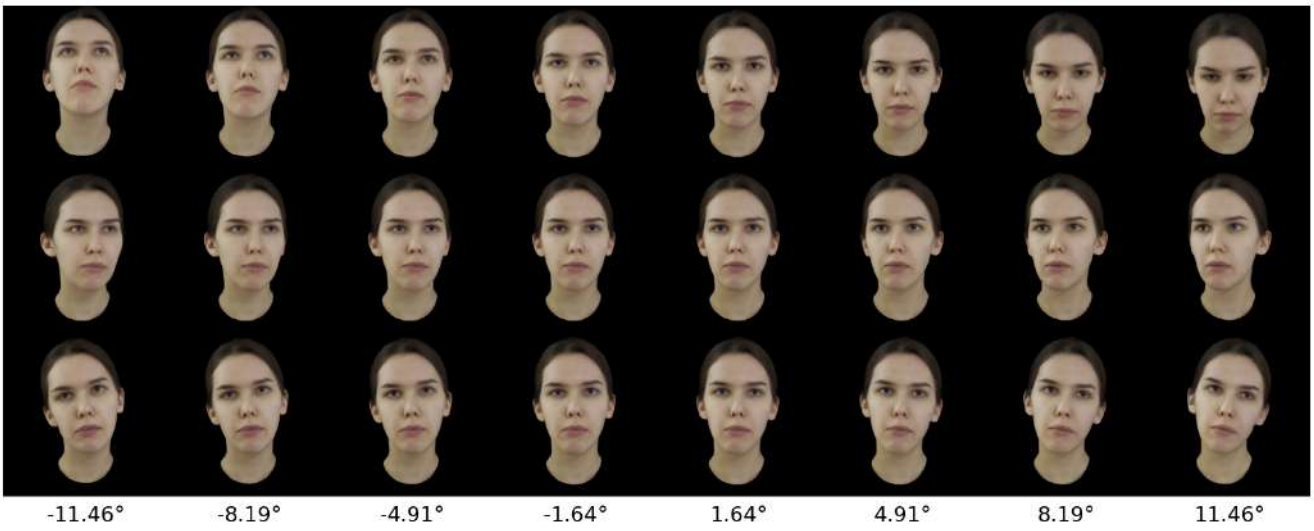


Рисунок А.64 – Результаты синтеза при варьировании поворота шеи для разрешения 256×256

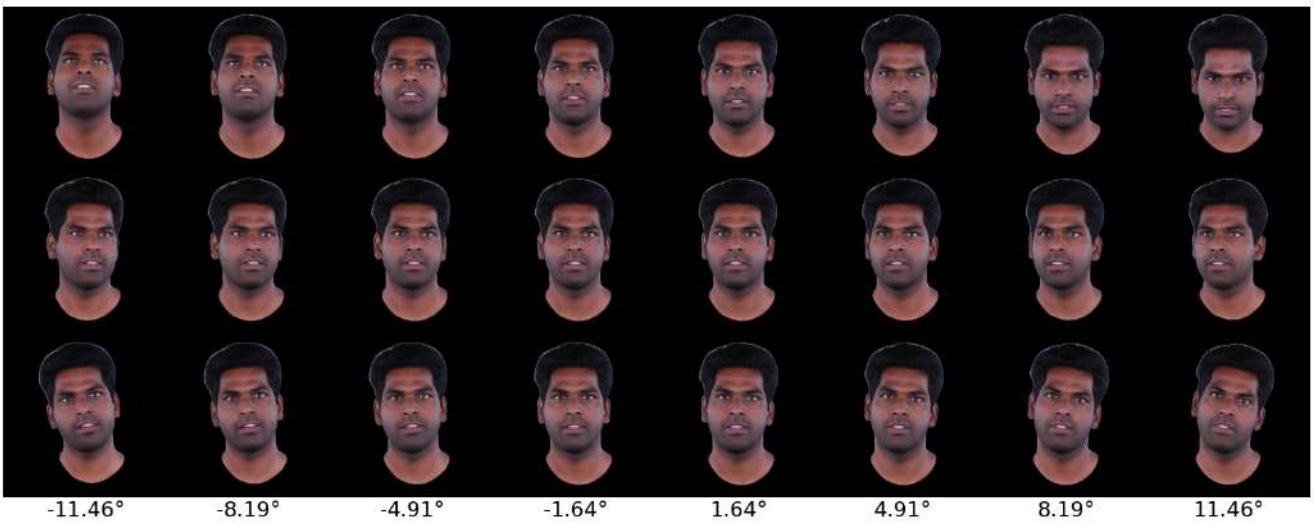


Рисунок А.65 – Результаты синтеза при варьировании поворота шеи для разрешения 512×512

Рисунок А.66 – Результаты синтеза при варьировании поворота шеи для разрешения 512×512

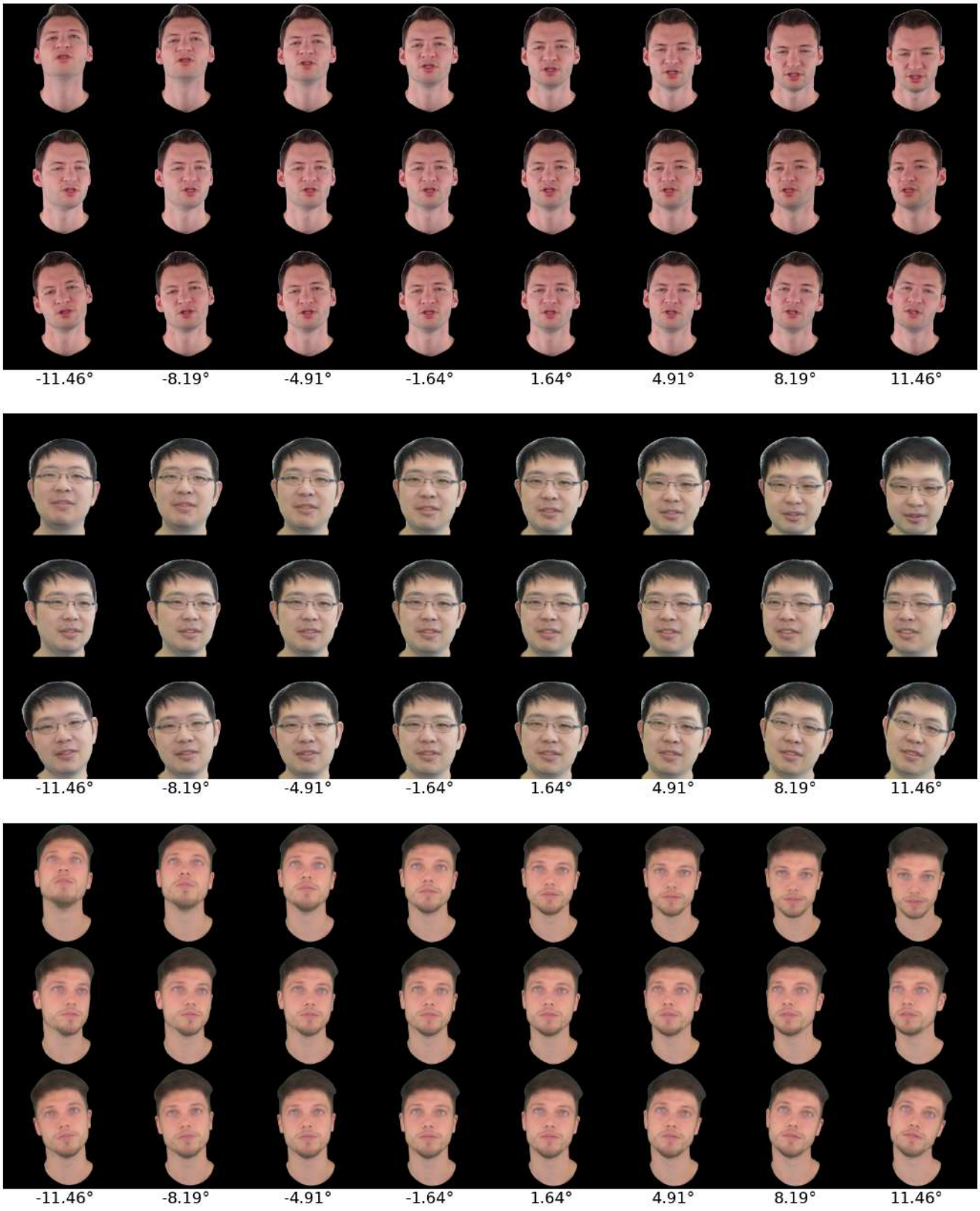


Рисунок А.67 – Результаты синтеза при варьировании поворота шеи для разрешения 512×512

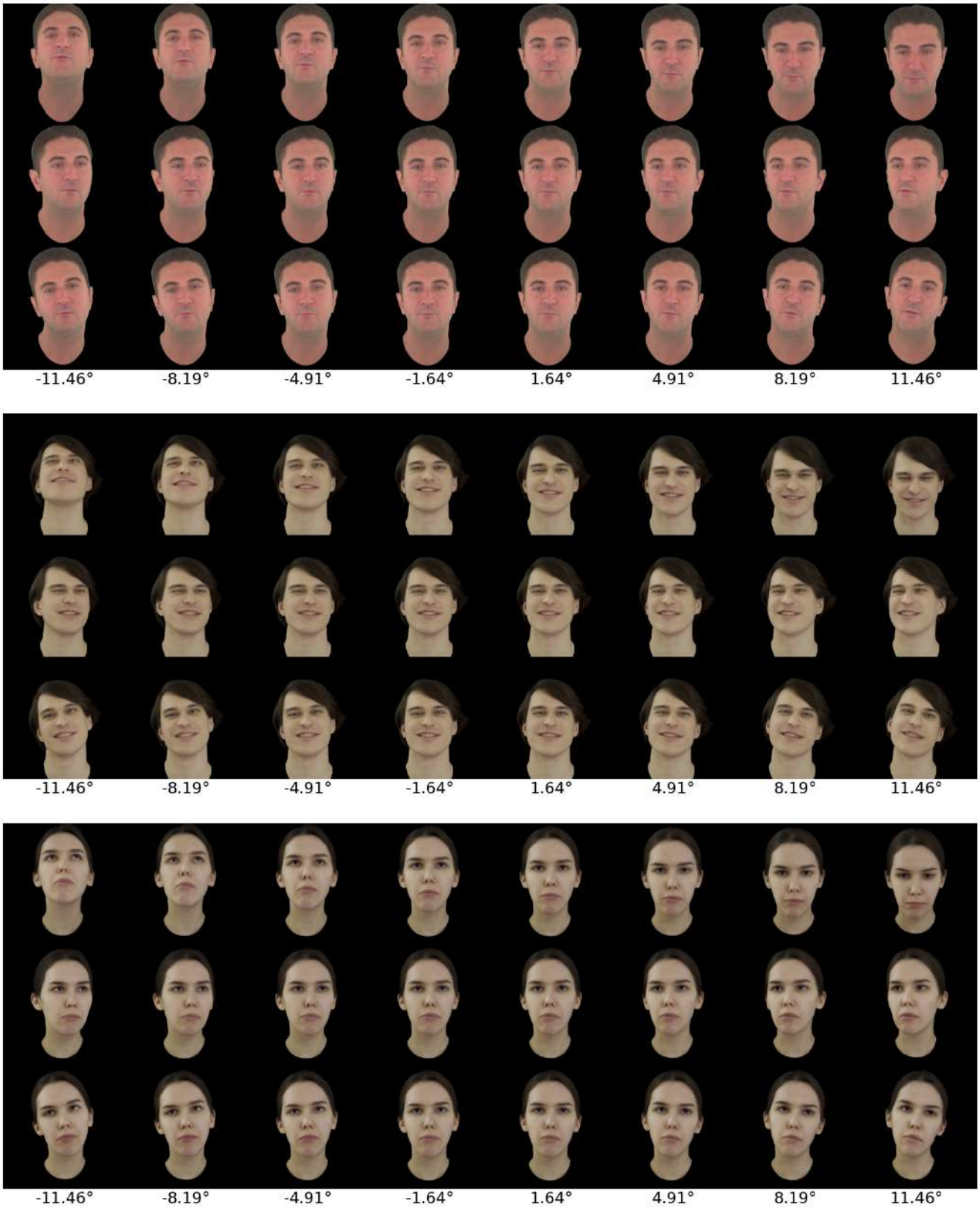


Рисунок А.68 – Результаты синтеза при варьировании поворота шеи для разрешения 512×512

Рисунок А.69 – Результат синтеза новых видов для разрешения 128×128

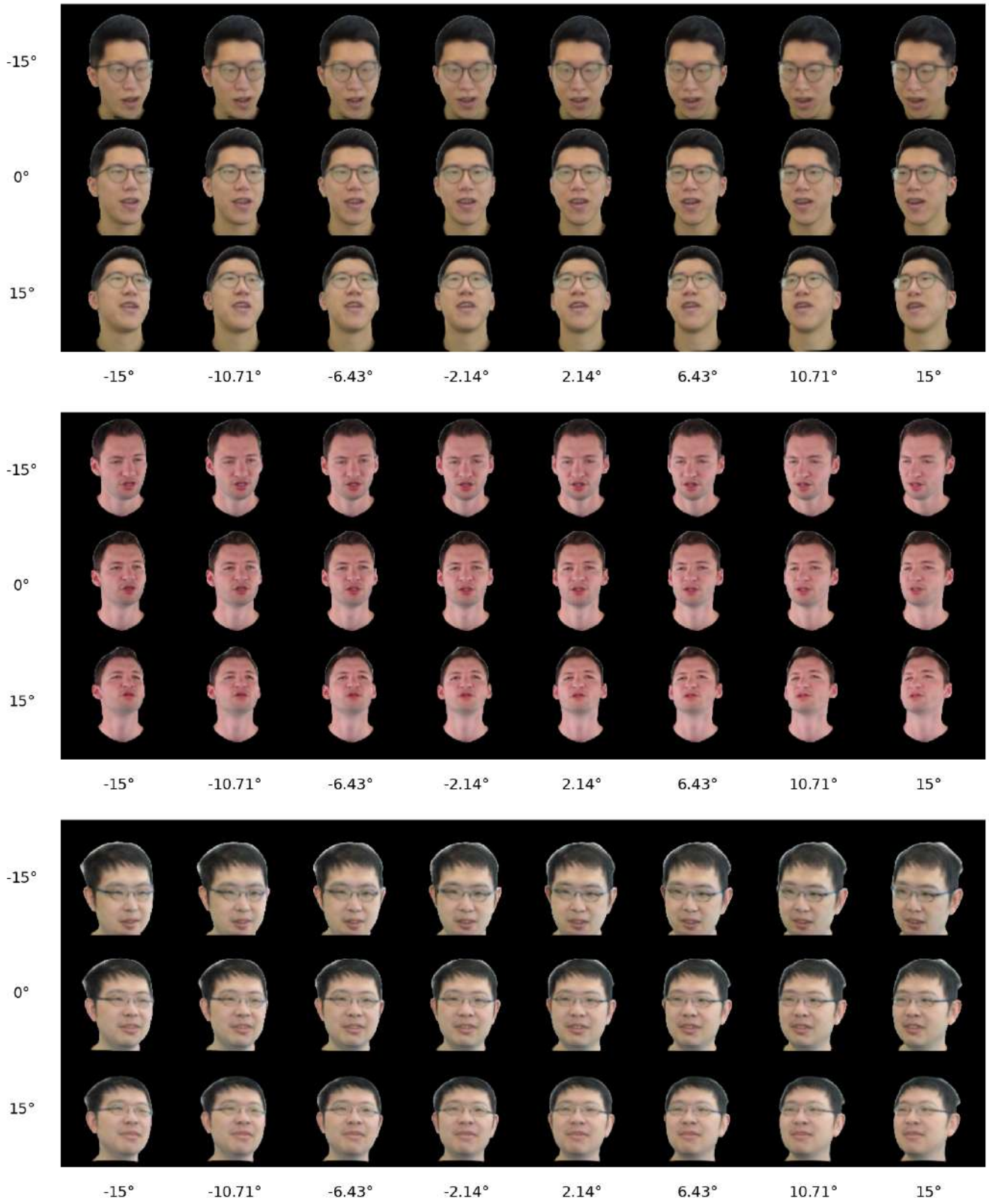


Рисунок А.70 – Результат синтеза новых видов для разрешения 128×128

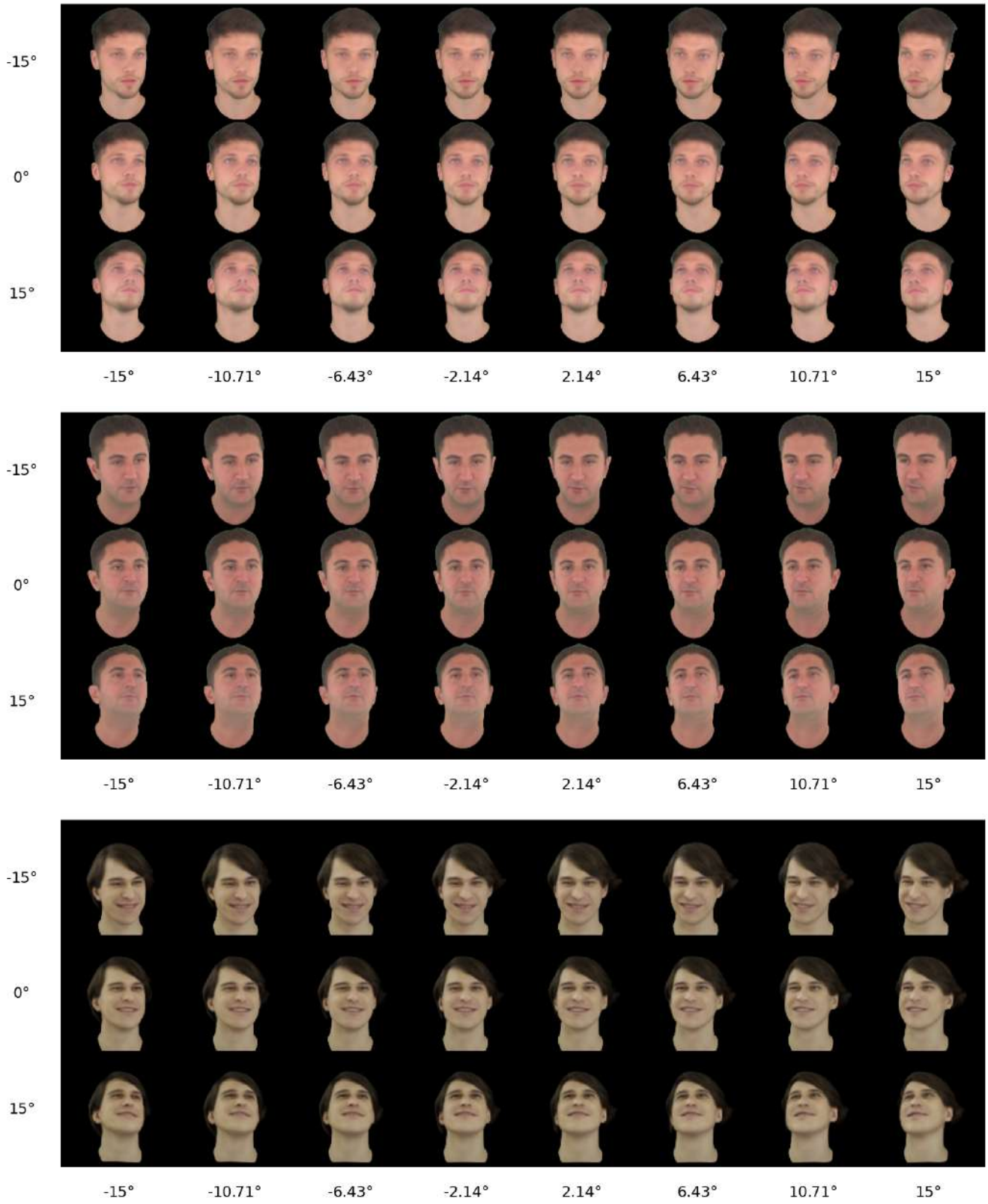


Рисунок А.71 – Результат синтеза новых видов для разрешения 128×128

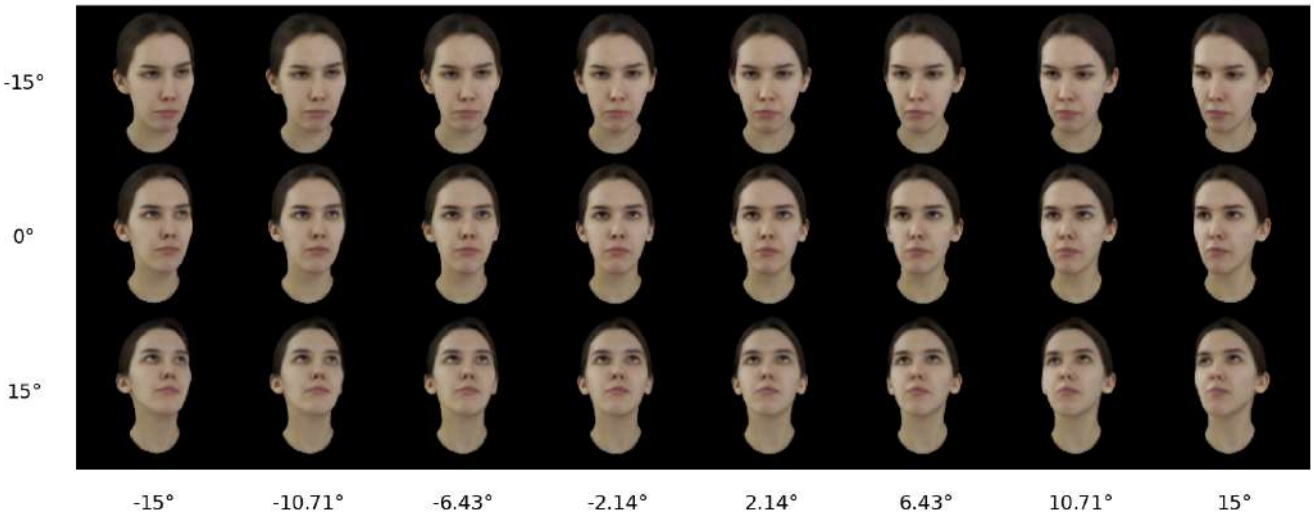


Рисунок А.72 – Результат синтеза новых видов для разрешения 128×128

Рисунок А.73 – Результат синтеза новых видов для разрешения 256×256

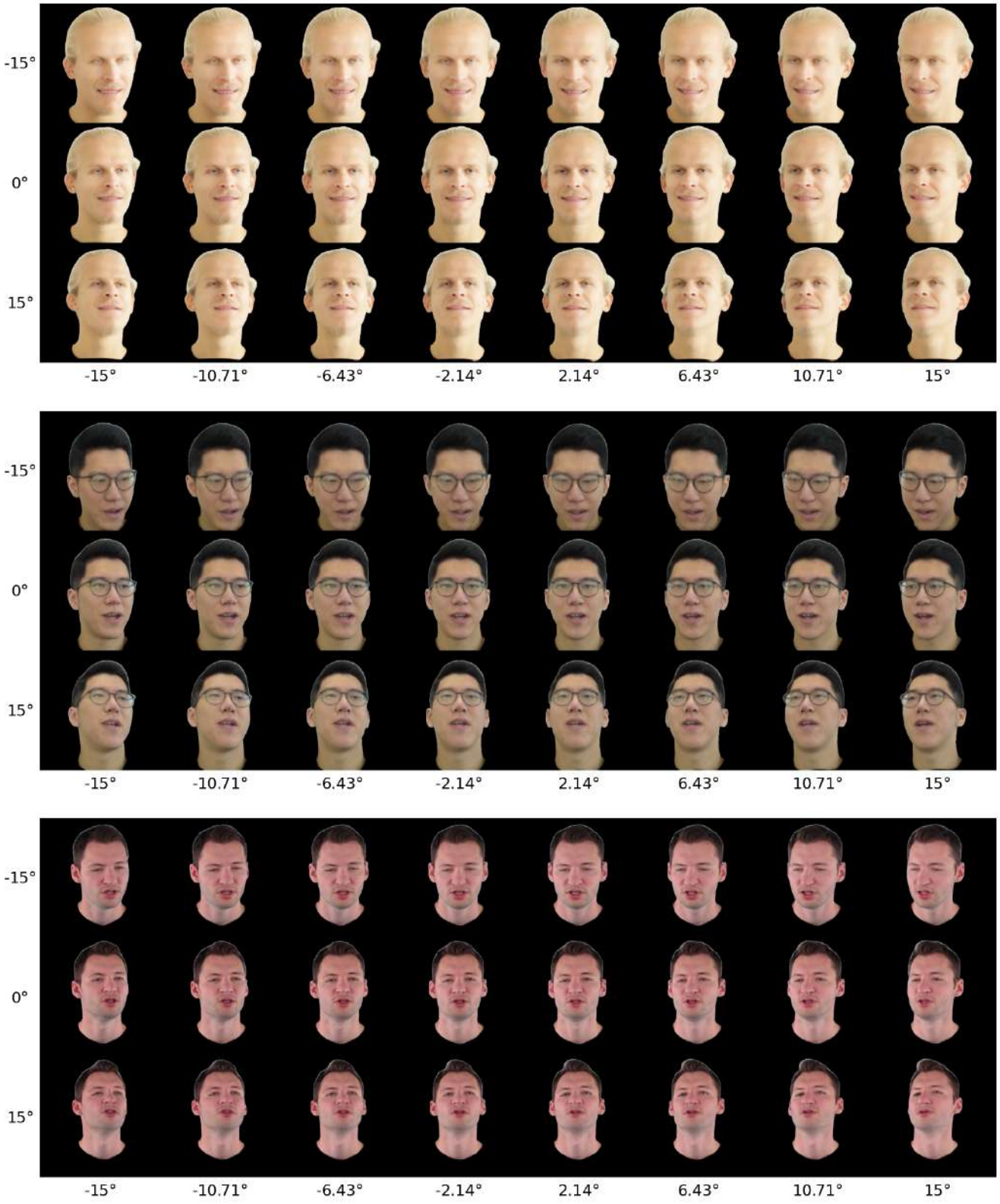


Рисунок А.74 – Результат синтеза новых видов для разрешения 256×256

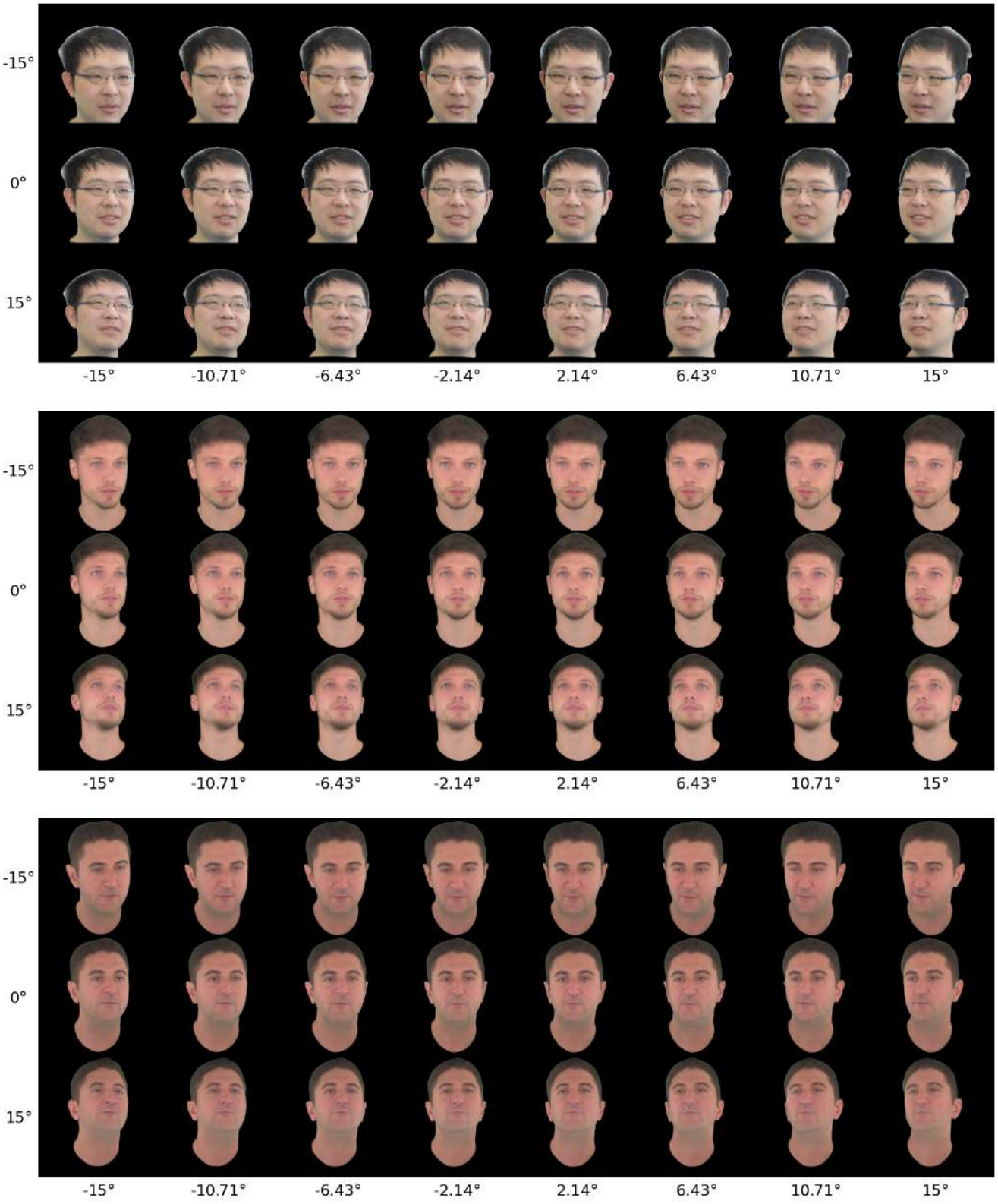


Рисунок А.75 – Результат синтеза новых видов для разрешения 256×256

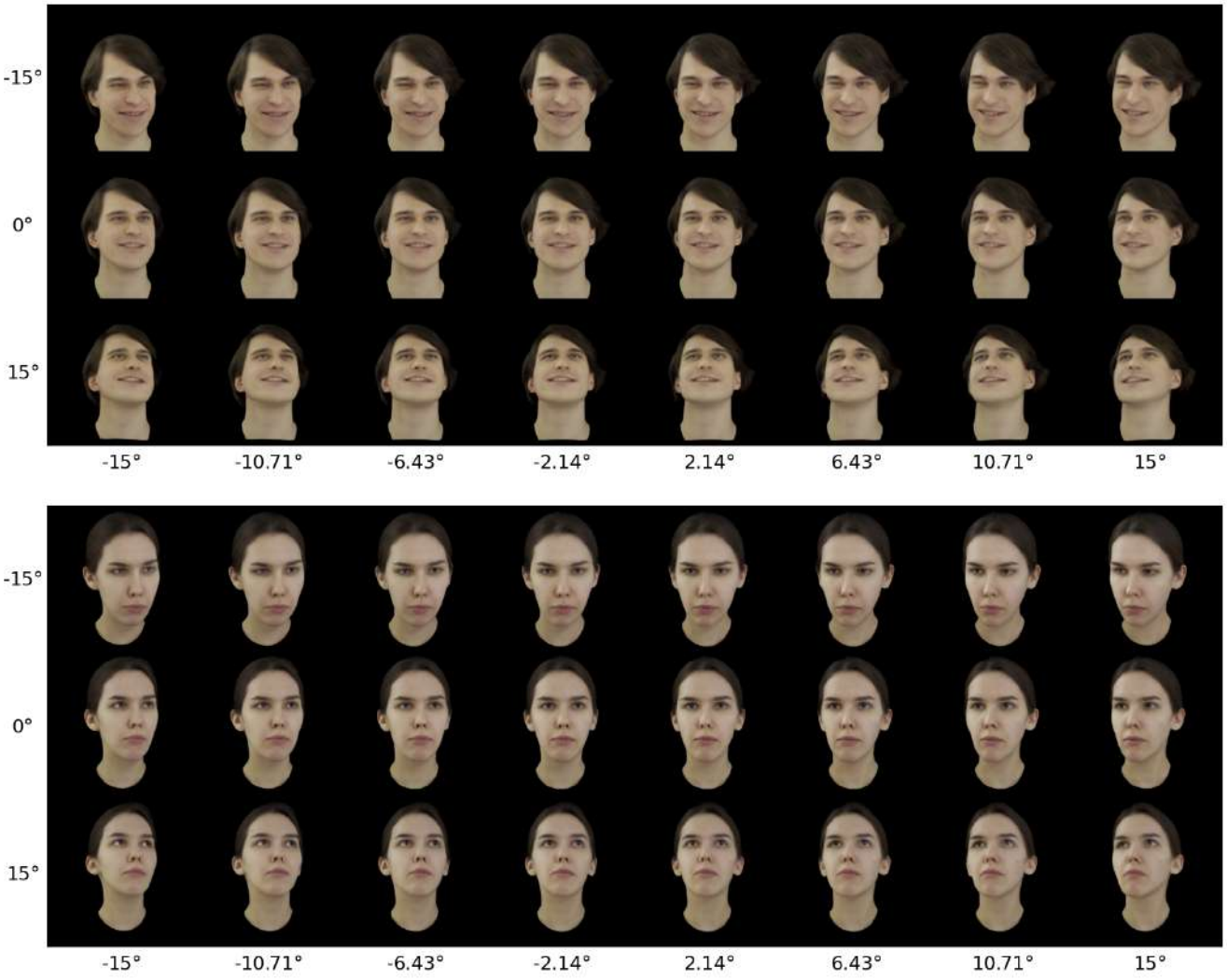


Рисунок А.76 – Результат синтеза новых видов для разрешения 256×256

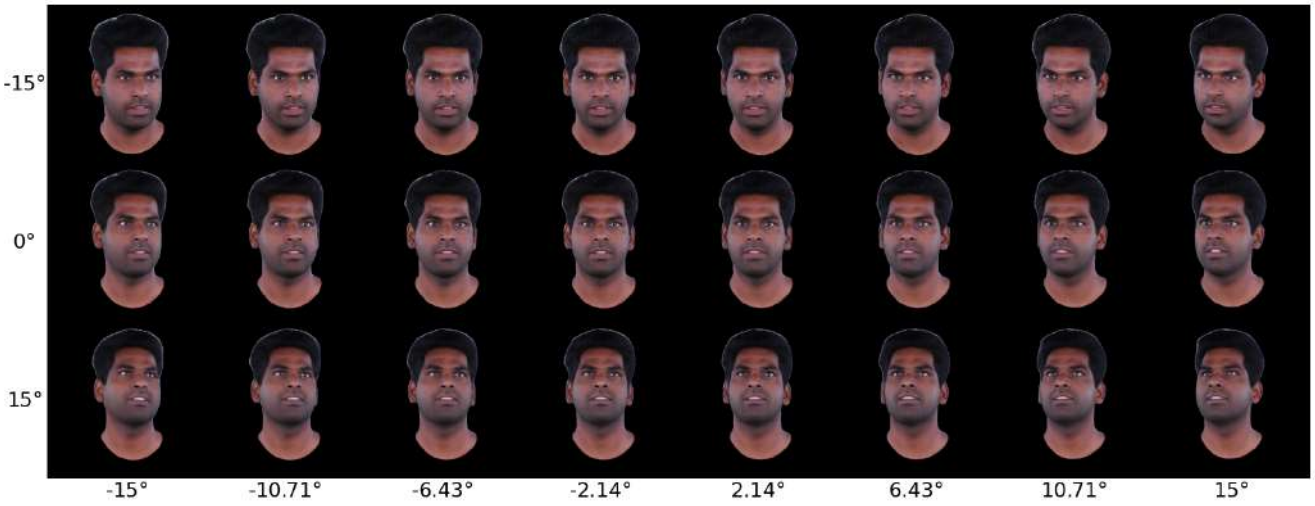


Рисунок А.77 – Результат синтеза новых видов для разрешения 512×512

Рисунок А.78 – Результат синтеза новых видов для разрешения 512×512

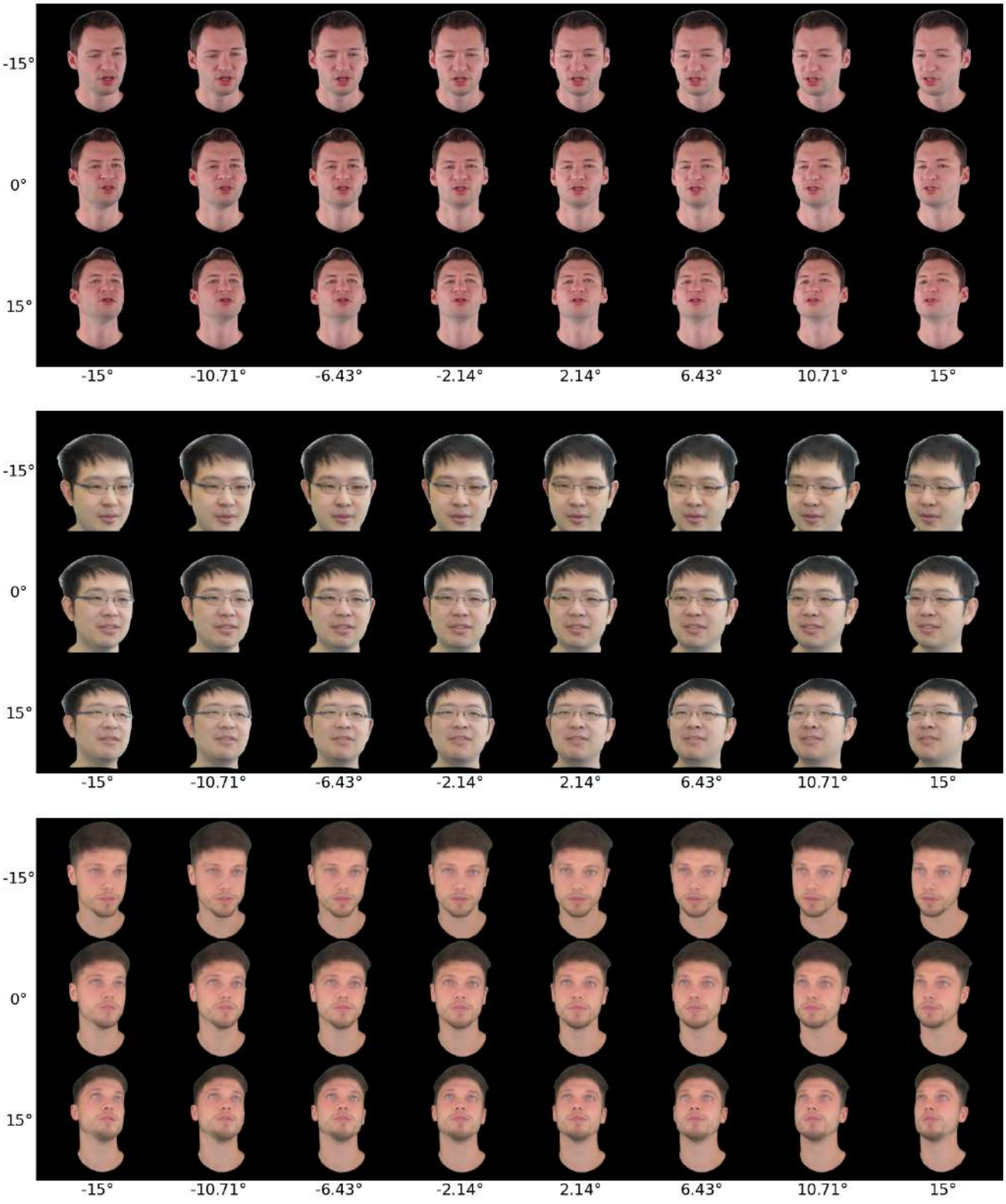


Рисунок А.79 – Результат синтеза новых видов для разрешения 512×512

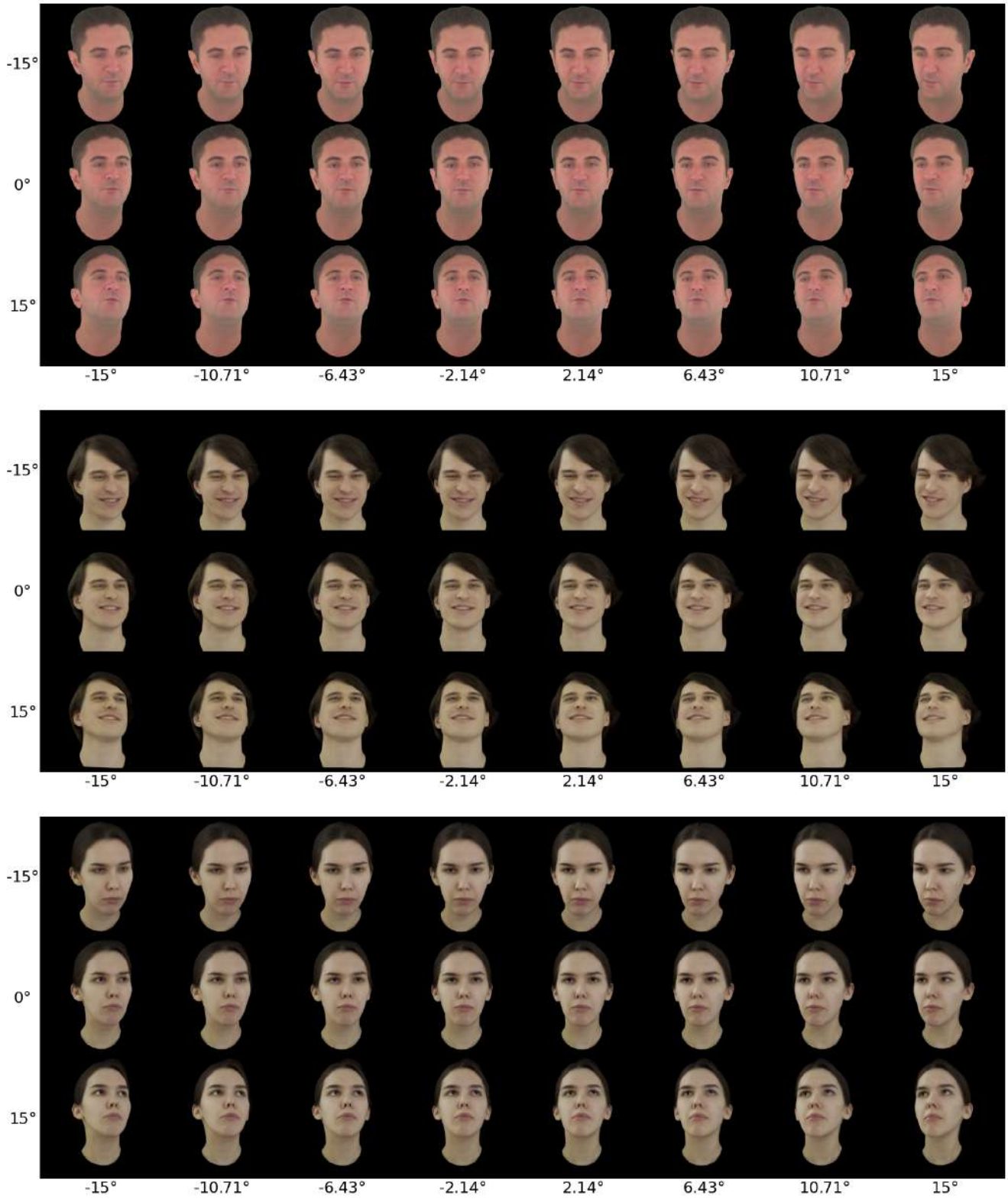


Рисунок А.80 – Результат синтеза новых видов для разрешения 512×512

Рисунок А.81 – Результат процедуры переноса выражения лица с кадров видеопоследовательностей для разрешения 128×128

Рисунок А.82 – Результат процедуры переноса выражения лица с кадров видеопоследовательностей для разрешения 256×256

Рисунок А.83 – Результат процедуры переноса выражения лица с кадров видеопоследовательностей для разрешения 512×512