

Федеральное государственное автономное образовательное учреждение высшего
образования

Национальный исследовательский университет
«Высшая школа экономики»

Факультет компьютерных наук

Долматов Кирилл Игоревич

ГЕНЕРАЦИЯ ИЗОБРАЖЕНИЙ БЛАСТОЦИСТ ЗАДАННОГО КАЧЕСТВА

Магистерская диссертация

студента образовательной программы магистратуры «Машинное обучение и
высоконагруженные системы» по направлению подготовки 01.04.02 Прикладная
математика и информатика

Рецензент
ученая степень, ученое звание,
должность

Свистунова Дарья Михайловна
MSc, DPhil (Oxon)
И.О. заведующей лаборатории
эмбриологии КВРТ «Дети из
Пробирки»

Руководитель
ученая степень, ученое
звание, должность

Аланов Айбек
Преподаватель

Москва, 2023 год

Аннотация

Генеративные состязательные сети (GAN, от англ. generative adversarial network) позволяют создавать уникальный контент, в частности генерировать изображения, которые визуально не отличимы от реальных. Одно из применений GAN – это генерация изображений с целью увеличения обучающей выборки. При решении задач, связанных с анализом медицинских изображений, зачастую сложно получить достаточное количество данных для построения предиктивных моделей. В частности, такая проблема существует в сфере вспомогательных репродуктивных технологий (ВРТ). Это обусловлено особенностями условий культивирования эмбрионов, а также конфиденциальностью персональных данных. Ранее другими исследователями были проведены исследования по генерации изображений эмбрионов на стадии дробления. Однако, на данный момент нет работ по генерации изображений эмбрионов на стадии бластоцисты. При этом именно стадия бластоцисты является одной из самых важных в ВРТ, так как именно на этой стадии можно наиболее достоверно оценить качество и перспективность эмбриона, а также именно на стадии бластоцисты зачастую решается его дальнейшая судьба. Используя архитектуру StyleGAN2-Ada, удалось получить реалистичные изображения бластоцист заданного качества. Таким образом, полученные результаты являются первыми и могут быть использованы для решения других задач компьютерного зрения, анализирующих изображения эмбрионов в стадии бластоцисты.

Список ключевых слов

Условная генерация изображений, генеративные состязательные сети, conditional image generation, GAN, DCGAN, StyleGAN

Abstract

Generative adversarial networks (GANs) allow to create unique content, including images that are indistinguishable from real ones. One of possible GANs application is generation of images to increase training sample size. One of the issues related to medical images analysis is difficulty of obtaining sufficient data to build predictive models. In particular, this issue exists in the area of assisted reproductive technologies (ART). This is due to strict embryo culture conditions that do not allow image capturing and confidentiality of the data. Previously published paper reports successful generation of cleavage stage embryo images. but there were no studies generating blastocyst stage images. Importantly, blastocyst stage is one of the key stages in ART cycle as embryo quality is assessed and its fate is decided during this period. In this work I have generated realistic images of blastocyst stage images of desired quality using StyleGAN2-Ada architecture. Obtained results are the first of their kind and can be used in various computer vision tasks that analyse blastocyst stage embryo images.

Key words

Conditional image generation, generative adversarial networks, GAN, DCGAN, StyleGAN

Введение

Генеративные состязательные сети (GAN) появились в 2014 году, впервые данную архитектуру предложил Ian J. Goodfellow [1]. Создание GAN стало прорывом в области генеративных моделей и позволило решать новые задачи, которые ранее считались недоступными. В силу универсальности архитектуры она нашла множество применений в самых различных областях, включающих генерацию изображений, аудио и даже видео, при этом полученные результаты вполне сопоставимы с реальными изображениями [2].

В задачах генерации изображений выделяют безусловную и условную генерацию. Задача условной генерации изображений предполагает наличие условия, которое должно быть учтено при создании изображения. При этом множество условий является конечным и, как правило, относительно небольшим. В качестве условия может выступать метка класса изображения, которое мы хотим сгенерировать, а может выступать и другое изображение, используя которое нужно произвести генерацию нового изображения.

С помощью методов условной генерации можно создавать уникальные и реалистичные изображения, что позволяет существенно сократить расходы, связанные с созданием визуального контента. Кроме того, данные методы позволяют бороться с несбалансированностью выборок путем генерации объектов минорных классов, что в итоге повышает качество прогнозирования конечных моделей.

Одним из самых поразительных аспектов GAN является их способность генерировать данные, которые трудно или невозможно получить в реальном мире. Например, GAN можно использовать для генерации реалистичных синтетических данных для последующего построения других моделей машинного обучения, что может быть особенно полезно в сфере медицины, где реальных данных мало или их получение затруднено в силу различных факторов [3].

В сфере вспомогательных репродуктивных технологий (ВРТ) достаточно сложно собрать большой размеченный набор изображений эмбрионов. Это обусловлено тем, что эмбрионы должны храниться в инкубаторе, где соблюдаются специальные условия культивирования, нарушение которых может привести к гибели эмбрионов. Поэтому единственной возможностью сделать фотографию эмбриона без вреда для его развития является использование специального инкубатора, оснащенного time-lapse системой. Подобные инкубаторы относятся к категории дорогостоящего медицинского оборудования и поэтому не используются повсеместно. Более того, клиники, как правило, не заинтересованы в сборе и публикации подобных изображений, так как для этого может потребоваться согласие пациентов. Таким образом, отсутствие публично доступных

наборов данных может быть существенной проблемой на пути исследователей, которые решают задачи компьютерного зрения, связанные с анализом изображений эмбрионов. Именно поэтому задача по генерации подобных изображений является актуальной.

В данной работе рассматриваются архитектуры генеративных состязательных моделей условной генерации изображений, а также выбирается наиболее подходящая архитектура для решения поставленной задачи. На основании анализа результатов экспериментов вносятся дополнительные правки в используемую архитектуру, а также выбираются ее оптимальные параметры.

Таким образом, **объектом исследования** является изображение, полученное по результатам работы модели, а **предметом исследования** является сам процесс условной генерации данного изображения.

Цель работы – применение методов генеративных состязательных сетей для генерации изображений эмбрионов стадии бластоцисты заданного качества. Поставленная цель включает выполнение следующих шагов:

- Привести описание предметной области
- Выполнить обзор существующих методов генерации изображений
- Провести сбор данных из открытых источников
- Провести разметку данных
- Провести сравнительные эксперименты по генерации и отобрать лучшую модель
- Получить заключение индустриального эксперта с оценкой качества и реалистичности сгенерированных изображений

Структура работы:

В Главе 1 приводится **обзор**:

- Технологии ЭКО
- Этапов раннего эмбриогенеза и оценки качества эмбриона
- Генеративных состязательных моделей для задачи генерации изображений
- Отличительных особенностей архитектуры StyleGAN2-Ada

В Главе 2 приводится **описание**:

- Процесса формирования обучающего набора данных
- Используемой метрики для оценки качества сгенерированных изображений
- Программного и технического окружения для проведения экспериментов

В Главе 3 приводится **описание**:

- Экспериментов с архитектурой DCGAN
- Экспериментов с архитектурой StyleGAN2-Ada

В конце главы приводится обсуждение полученных результатов и формулируются вопросы и задачи, которые предлагается изучить и реализовать в рамках продолжения работы над проектом.

В Заключении приводятся выводы.

АННОТАЦИЯ.....	2
ВВЕДЕНИЕ	4
ГЛАВА 1 ОБЗОР ЛИТЕРАТУРЫ.....	8
1.1 ВВЕДЕНИЕ В ПРЕДМЕТНУЮ ОБЛАСТЬ	8
1.1.1 Технология ЭКО	8
1.1.2 Этапы раннего эмбриогенеза и оценка качества эмбриона	10
1.1.3 Анализ развития эмбрионов: time-lapse инкубаторы	13
1.2 ГЕНЕРАТИВНЫЕ СОСТЯЗАТЕЛЬНЫЕ МОДЕЛИ.....	15
1.2.1 GAN и Conditional GAN	15
1.2.2 DCGAN.....	21
1.2.3 Семейство архитектур StyleGAN	23
ГЛАВА 2 МАТЕРИАЛЫ И МЕТОДЫ.....	34
2.1 Данные	34
2.1.1 Сложности сбора медицинских данных	34
2.1.2 Источники данных	34
2.1.3 Разметка данных на классы	36
2.1.4 Формирование обучающей выборки	37
2.2 МЕТРИКИ ОЦЕНКИ КАЧЕСТВА ГЕНЕРАЦИИ.....	38
2.2.1 IS.....	38
2.2.2 FID	40
2.3 ОКРУЖЕНИЕ ДЛЯ ПРОВЕДЕНИЯ ЭКСПЕРИМЕНТОВ	43
ГЛАВА 3 ЭКСПЕРИМЕНТЫ.....	44
3.1 ГЕНЕРАЦИЯ ИЗОБРАЖЕНИЙ С ПОМОЩЬЮ DCGAN	44
3.2 ГЕНЕРАЦИЯ ИЗОБРАЖЕНИЙ С ПОМОЩЬЮ ПРЕДОБУЧЕННОЙ STYLEGAN2-ADA	45
3.3 ГЕНЕРАЦИЯ ИЗОБРАЖЕНИЙ С ПОМОЩЬЮ STYLEGAN2-ADA	47
3.4 ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ И ДАЛЬНЕЙШИЕ ШАГИ	50
ЗАКЛЮЧЕНИЕ.....	53
БЛАГОДАРНОСТИ.....	54
СПИСОК ЛИТЕРАТУРЫ.....	55
ПРИЛОЖЕНИЕ	59
Приложение А. Результаты экспериментов на предобученной модели STYLEGAN2-Ada	59
Приложение Б. Результаты экспериментов на модели STYLEGAN2-Ada	61
Приложение В. Заключение индустриального эксперта	67

Глава 1 Обзор литературы

1.1 Введение в предметную область

1.1.1 Технология ЭКО

Появление и развитие методов вспомогательных репродуктивных технологий произвело революцию в подходах к лечению бесплодия и позволило огромному числу пар, ранее не имевших шансов родить детей, реализовать свои репродуктивные планы. Схема лечения бесплодия включает в себя несколько этапов, занимающих по времени от пары недель до нескольких месяцев (Рисунок 1.1).



Рисунок 1.1 Схема этапов процедуры ЭКО. Источник: www.invitro.ru

Процедура экстракорпорального оплодотворения (ЭКО, от лат. extra — сверх, вне и лат. corpus — тело, то есть оплодотворение вне тела) начинается со стимуляции суперовуляции гормональными препаратами, чаще всего в форме подкожных инъекций, для получения максимально возможного количества ооцитов и максимизации шансов на получение пригодного к переносу эмбриона (Рисунок 1.1, Этап 1).

При достижении фолликулами в яичниках пациентки оптимального размера и, соответственно, зрелости, под наркозом проводится трансвагинальная пункция фолликулов, во время которой под контролем УЗИ с помощью тонкой полый иглы производится забор содержимого фолликулов - фолликулярной жидкости и яйцеклеток (Рисунок 1.1, Этап 2).

Одновременно с этим производятся получение и подготовка сперматозоидов партнера пациентки для оплодотворения; материал получают естественным путем или с помощью хирургических манипуляций при наличии у мужчины показаний (Рисунок 1.1, Этап 3). При необходимости могут использоваться донорские ооциты или сперматозоиды, в том числе криоконсервированные.

После получения и подготовки половых клеток производится оплодотворение одним из двух возможных методов - ЭКО или ИКСИ (Рисунок 1.1, Этап 4). Метод ЭКО заключается в добавлении большого количества сперматозоидов в культуральную среду, где находятся яйцеклетки, при этом оплодотворение происходит естественным путем. При методе оплодотворения ИКСИ (а также при его модификациях - ПИКСИ и ИМСИ) в яйцеклетку с помощью микроинструментов вводится единичный сперматозоид, который эмбриолог выбирает под микроскопом по морфологическим признакам. Тот или иной метод оплодотворения выбирается на основании анамнеза пациентов.

После оплодотворения эмбрионы помещаются в инкубатор, в котором поддерживаются строго контролируемые условия для создания оптимальной среды для их развития (Рисунок 1.1, Этап 5). Чаще всего эмбрионы культивируют до 5-6 суток развития, в этот день производится перенос эмбриона в полость матки пациентки или криоконсервация для последующего использования в будущем, а также биопсия для проведения генетического анализа (Рисунок 1.1, Этап 6).

При выборе тактики работы с эмбрионами (перенос в полость матки, криоконсервация или утилизация) клинические эмбриологи и врачи руководствуются качеством и, соответственно, перспективностью каждого конкретного эмбриона с точки зрения наступления беременности. Для оценки шансов эмбриона на имплантацию и развитие здоровой беременности эмбриологи опираются на морфологические параметры эмбриона в разные моменты его развития, а также генетический статус эмбриона, если пациенты желают провести соответствующий анализ.

1.1.2 Этапы раннего эмбриогенеза и оценка качества эмбриона

На заре развития технологии ЭКО перенос эмбриона в полость матки или в фаллопиевы трубы проводили в самые ранние возможные сроки, вплоть до совместного переноса яйцеклетки и сперматозоидов еще до оплодотворения *in vitro* [4 - 8]. С одной стороны, подобный подход минимизирует влияние внешней среды на развитие эмбриона, но с другой стороны ранний перенос эмбриона в полость матки не позволяет эмбриологу получить максимально подробную информацию о его развитии и перспективности.

В ходе своего развития эмбрион проходит через многочисленные акты клеточного деления, превращаясь из единственной клетки сперва в состоящий из нескольких клеток зародыш, а затем в готовый к имплантации эмбрион, называемый бластоцистой. На самой ранней стадии развития, когда эмбрион еще состоит из одной клетки, он носит название зигота. В течение последующих трех дней эмбрион претерпевает последовательные раунды деления, названные дроблением; в этот период эмбрион не меняется в размерах, а составляющие его клетки с каждым делением становятся все меньше. Примерно на четвертый день происходит компактизация, в ходе которой клетки эмбриона, называемые бластомерами, образуют большое количество межклеточных контактов и сливаются в одну массу, называемую морула. На пятый-шестой день развития морула развивается в бластоцисту, в которой уже выделяются отвечающие за разные функции части (Рисунок 1.2) [9].



Рисунок 1.2 Этапы предимплантационного развития эмбриона человека. День 0 - стадия пронуклеусов, дни 1-3 - стадия дробления, день 4 - стадия морулы, день 5+ - стадия бластоцисты. Источник: www.remembryo.com

Именно на стадии бластоцисты чаще всего производится оценка качества и перспективности эмбриона специалистом: помимо факта, что 5-6 суток — это максимальный возможный срок культивирования эмбрионов в клинической лаборатории, достижение эмбрионом стадии бластоцисты само по себе является важным прогностическим критерием, а потому эмбриологи и врачи предпочитают работать именно с эмбрионами этой стадии развития.

Бластоциста состоит из трофэктодермы (ТЕ) и внутренней клеточной массы (ВКМ), окруженных состоящей из белка оболочкой (ZP, zona pellucida) (Рисунок 1.3). Трофэктодерма отвечает за взаимодействие эмбриона с маткой и имплантацию, и в будущем из клеток ТЕ развивается плацента и плодные оболочки, тогда как ВКМ дает начало плоду [9]. ZP же представляет собой оболочку, образующуюся в процессе созревания яйцеклетки, которая защищает яйцеклетку и, в последствии, развивающийся эмбрион до момента его вылупления (хэтчинга) и имплантации примерно на 5-6 день после оплодотворения [9].

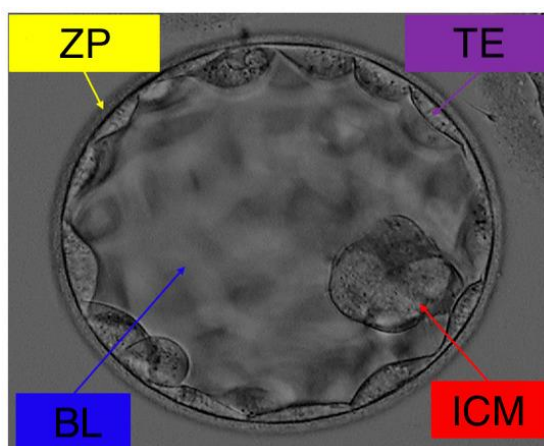


Рисунок 1.3 Морфология бластоцисты человека. ICM (inner cell mass) - внутренняя клеточная масса, TE (trophoblast) - трофэктодерма, ZP (zona pellucida) - блестящая оболочка, BL (blastocoel) - бластоцель, полость внутри бластоцисты [10].

Именно размер и состояние этих элементов являются основными характеристиками, оцениваемыми эмбриологами при принятии решения о судьбе эмбриона. Для унификации системы оценки качества эмбрионов была разработана повсеместно принятая шкала Гарднера, с помощью которой описывается все разнообразие качеств бластоцист (Рисунок 1.4) [11].

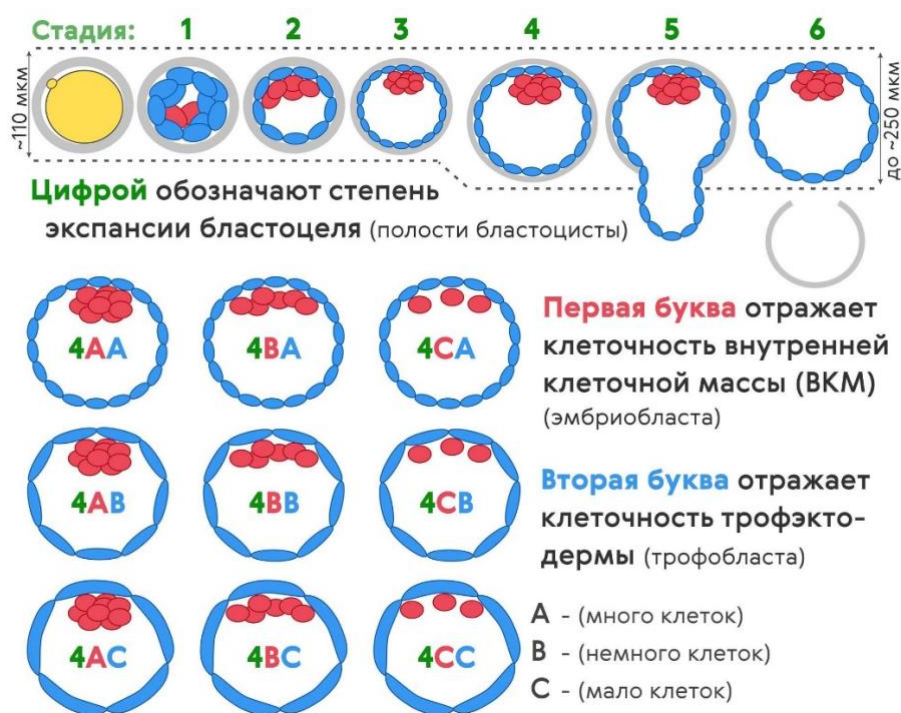


Рисунок 1.4 Классификация бластоцист человека по Гарднеру. Источник: <https://www.cfm.ru/>

Оценка бластоцисты по Гарднеру состоит из трех показателей - цифры от 1 до 6 и двух букв от А до С (например, 2АВ или 4ВС). Цифра обозначает размер эмбриона и степень его экспансии (или долю, занимаемую полостью относительно всего размера эмбриона), где 1 - ранняя бластоциста, 2-3 - растущие бластоцисты, 4 - полностью экспандированная бластоциста, 5 - бластоциста на стадии хэтчинга (вылупления из ЗР), а 6 - покинувший ЗР эмбрион.

Первая из букв описывает качество ВКМ - ее плотность, размер и количество образующих ее клеток. Буквой А обозначается плотная ярко выраженная ВКМ; В - более рыхлая ВКМ, состоящая из меньшего количества клеток; С - ВКМ образована несколькими разрозненными клетками или вовсе отсутствует.

Вторая буква в оценке по Гарднеру описывает состояние ТЕ по аналогичной схеме. Эмбрионы, оценка которых содержит хотя бы одну букву С, имеют существенно более низкие шансы на успех при переносе в полость матки, а потому зачастую утилизируются по согласию пациентов [12].

1.1.3 Анализ развития эмбрионов: time-lapse инкубаторы

Поскольку человеческие эмбрионы чувствительны к факторам внешней среды, работа в лаборатории ЭКО всегда организована таким образом, что эмбрионы извлекаются из инкубатора и оцениваются под микроскопом минимальное возможное число раз - зачастую это лишь проверка правильности оплодотворения на 1 сутки развития и оценка качества эмбрионов на 5-6 сутки. Благодаря этому, с одной стороны, соблюдается постоянство окружающей эмбрионы среды, но с другой стороны это не позволяет эмбриологу оценивать большее количество параметров развития эмбриона для более взвешенного принятия решения о тактике работы с ним.

Это изменилось с внедрением в клиническую практику time-lapse инкубаторов, которые позволяют следить за всем сроком развития эмбрионов, не извлекая их из инкубатора и не нарушая условий культивирования. Помимо съемки изображений эмбрионов через заданные промежутки времени (чаще всего раз в 15 минут), эти time-lapse системы культивирования эмбрионов также оснащены программой, которая на основании морфокинетических параметров развития подсказывает эмбриологу, какие эмбрионы наиболее перспективны.

Полученные с помощью time-lapse систем изображения используются в обучении будущих эмбриологов и других образовательных программах, а кроме того, открывают возможность исследования параметров развития эмбрионов, которые ранее было невозможно оценить в силу особенностей систем культивирования. Если на ранних этапах развития технологии time-lapse культивирования эмбрионов программа лишь умела автоматически определять ключевые параметры и моменты в развитии эмбриона (время дробления, количество пронуклеусов, оценка бластоцисты по Гарднеру), то сегодня на основании проанализированных пар данных “параметры развития эмбриона - наступление клинической беременности” программа может подсказать эмбриологу, какой эмбрион имеет максимальные шансы дать беременность [13]. Помимо шансов на имплантацию и наступление беременности, морфокинетические параметры развития эмбриона также коррелируют с эу- или анеуплоидностью (корректным или некорректным набором хромосом), что также играет очень важную роль в клинической практике [14, 15].

К сожалению, инкубаторы с технологией time-lapse, самым распространенным и популярным представителем которых является EmbryoScope (Vitrolife), очень дороги, и в России, по данным сайта www.embryoscope.ru, лишь в 20 из более чем 200 клиник ЭКО по всей стране стоит подобный инкубатор.

Очевидно, что для создания подобной технологии нужна обучающая выборка, которая позволит моделям распознавания эффективно решать задачи компьютерного зрения. При этом, зачастую, получение медицинских данных сторонними исследователями является очень сложной задачей, так как данные могут быть конфиденциальны и обычно не подлежат свободному распространению.

Таким образом, создание генеративной модели, которая будет способна генерировать изображения с целью увеличения обучающей выборки для решения других задач, является актуальной проблемой, так как другие исследователи смогут решать задачи, для выполнения которых ранее не было достаточного количества данных.

1.2 Генеративные состязательные модели

1.2.1 GAN и Conditional GAN

Генеративные состязательные сети – алгоритм машинного обучения, принадлежащий к классу порождающих моделей. В основе данного алгоритма лежит комбинация двух нейронных сетей: генеративная модель, строящая приближение распределения исходных входных данных, и дискриминативная модель, которая оценивает вероятность, что объект был получен из исходных данных (обучающей выборки), а не сгенерирован генеративной моделью. Генератор создает данные из случайного шума, в то время как дискриминатор оценивает сгенерированные данные и определяет, являются ли они реальными или сгенерированными. Этот процесс повторяется до тех пор, пока дискриминатор больше не сможет отличать реальные данные от сгенерированных. Данная архитектура впервые была опубликована в статье «*Generative Adversarial Networks*» [1]. В настоящее время GAN решают самые разные задачи, такие как улучшение изображений (image inpainting) [16], улучшение звука (speech enhancement) [17], повышения разрешения изображения (super-resolution) [18], раскраска изображений (colorization) [19], расширение набора данных (synthetic data generation) [20].

В качестве базовых моделей для генератора и дискриминатора используются многослойные перцептроны. Генератор можно представить как отображение $G(z, \gamma_G)$, где G – дифференцируемая функция, представленная многослойным перцептроном с параметром γ_G , а z – шум. Также, чтобы вывести вероятностное распределение генератора p_G над набором входных данных X , определим априорную вероятность шума $p_z(z)$. Тогда, по аналогии с генератором, представим дискриминатор в виде $D(z, \gamma_D)$, который возвращает вероятность того, что объект x был получен из исходных входных данных, а не из p_G – сгенерированных данных. Во время обучения дискриминатор стремится максимизировать вероятность правильной классификации объектов из исходных входных данных и сгенерированных данных. При этом генератор стремится минимизировать $\log(1 - D(G(z)))$. Говоря иначе, генератор и дискриминатор «играют» в минимакс игру. Формально это записывается так:

$$L = \min_G \max_D V(D, G) = E_{x \sim p_{data}} [\log D(x)] + E_{z \sim p_z} [\log(1 - D(G(z)))]$$

На Рисунок 1.5 наглядно отображена схема обучения архитектуры GAN.

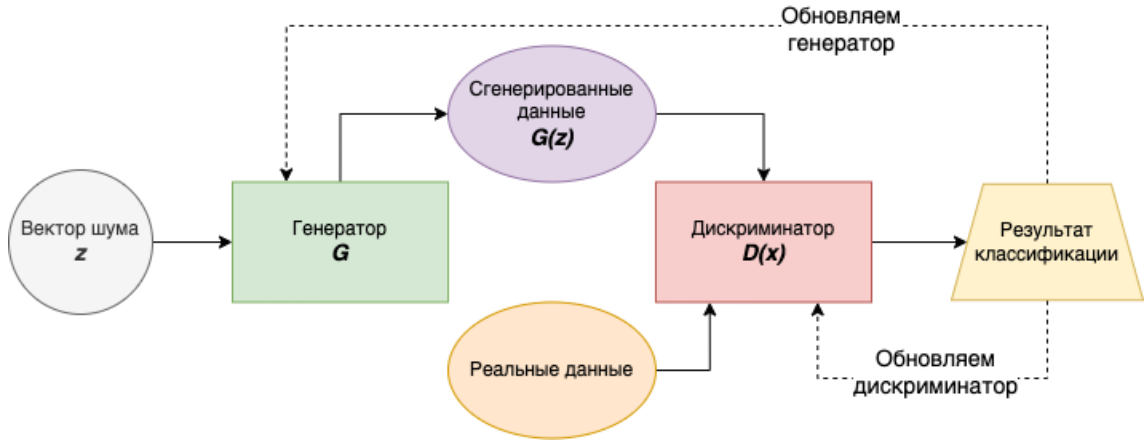


Рисунок 1.5 Блок-схема обучения архитектуры GAN

Алгоритм обучения выглядит следующим образом:

for number of training iterations **do**

for k steps **do**

- Sample minibatch of m noise samples $\{z^{(1)}, \dots, z^{(m)}\}$ from noise prior $p_g(z)$.
- Sample minibatch of m examples $\{x^{(1)}, \dots, x^{(m)}\}$ from data generating distribution $p_{\text{data}}(x)$.
- Update the discriminator by ascending its stochastic gradient:

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^m \left[\log D(x^{(i)}) + \log (1 - D(G(z^{(i)}))) \right].$$

end for

- Sample minibatch of m noise samples $\{z^{(1)}, \dots, z^{(m)}\}$ from noise prior $p_g(z)$.
- Update the generator by descending its stochastic gradient:

$$\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^m \log (1 - D(G(z^{(i)}))).$$

end for

The gradient-based updates can use any standard gradient-based learning rule. We used momentum in our experiments.

Как можно заметить, дискриминатор учится k шагов, а генератор только один шаг. Это обусловлено тем, что дискриминатор должен быть в каком-то смысле более обученным, чтобы давать более полезную обратную связь при обучении генератора.

Основные этапы эталонного обучения GAN схематично представлены на Рисунок 1.6. На шаге (a) генератор и дискриминатор находятся в близости от своей сходимости - $p_g \sim p_{\text{data}}$, а D - частично точный классификатор. На шаге (b) дискриминатор обучается различать выборки из данных, сходясь $D^*(x) = \frac{p_{\text{data}}(x)}{p_{\text{data}}(x) + p_g(x)}$. На шаге (c) после обновления G градиент D скорректировал отображение $G(z)$ в регионы, которые с большей вероятностью будут классифицированы как данные. На шаге (d), после нескольких

итераций обучения, если G и D обладают достаточной мощностью, они достигают точки, в которой оба не могут улучшиться, так как $p_{(g)} = p_{(data)}$. Дискриминатор не может различить два распределения — реальное и сгенерированное, то есть $D(x) = \frac{1}{2}$.

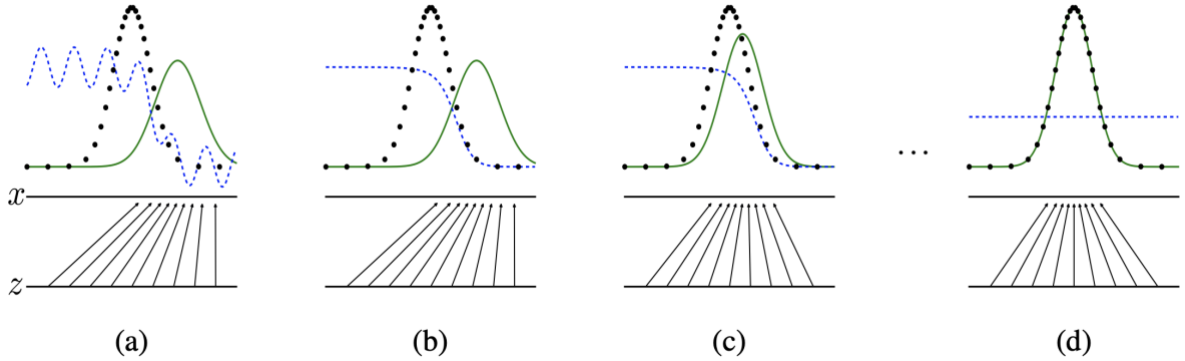


Рисунок 1.6 Основные этапы обучения GAN. Обозначение: черная точечная линия — это распределение реальных данных $p_{(data)}$, на основании которого модель пытается обучить генератор. Синяя точечная линия — функция распределения, характеризующая дискриминатор (D) — принимает значения от 0 до 1. Зеленая линия — это распределение сгенерированных данных $p_{(g)}$ (G). Нижняя линия z — это область, из которой мы сэмплируем z (в данном случае равномерное распределение), а верхняя линия x — это часть области x . Восходящие стрелки — отображение $x=G(z)$. G сжимается в областях с высокой плотностью и расширяется в областях с низкой плотностью [1].

Для решения более узких и сложных задач зачастую требуется дополнительная информация, чтобы генерировать объекты из заданного подмножества относительно всей обучающей выборки. Говоря иначе, на генерацию накладываются дополнительные условия, которые позволяют в некоторой степени управлять конечным результатом генерации. Это возможно благодаря расширению архитектуры GAN и добавлению дополнительных данных - условия, которое учитывается при обучении как генератора, так и дискриминатора. Данная архитектура впервые была предложена в статье «*Conditional Generative Adversarial Nets*» и получила одноименное название, или сокращенно CGAN [21].

В качестве дополнительных данных может быть использован практически любой объект: произвольный текст, изображение, аудио, метка класса или информация от других моделей. Такая вариативность порождает множество отдельных задач, например, при использовании изображения в качестве условия, часто решается задача трансляции изображений (англ. *image-to-image translation*). В данной работе в качестве условия будет использоваться метка класса, отражающая качество эмбриона.

Формально задача оптимизация будет выглядеть следующим образом:

$$L = \min_G \max_D V(D, G) = E_{x \sim p_{data}} [\log D(x|y)] + E_{z \sim p_z} [\log (1 - D(G(z|y)))]$$

На

Рисунок 1.7 наглядно отображена схема обучения архитектуры CGAN.

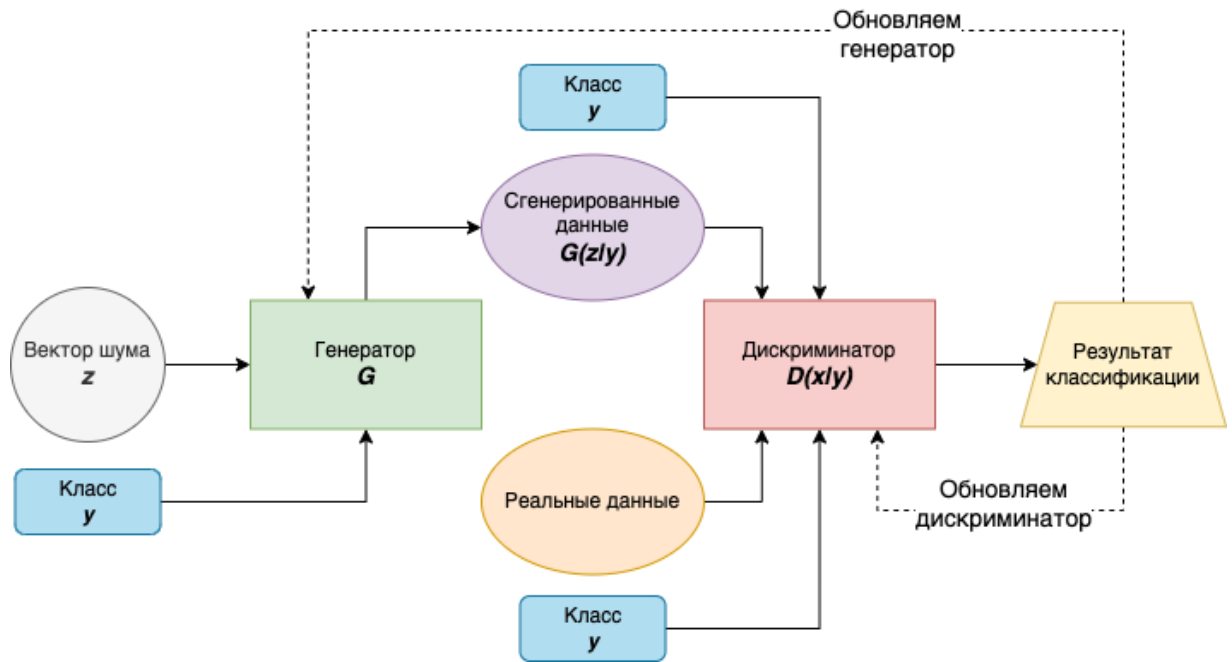


Рисунок 1.7 Блок-схема обучения архитектуры CGAN

На практике обучение GAN часто оказывается не стабильным и приходится сталкиваться с различными проблемами. Одной из причин этого является то, что процесс обучения GAN заимствован из теории игр, где игроки (генератор и дискриминатор) пытаются прийти к равновесию Нэша (процесс обучения сети должен быть направлен на сходимость к этому равновесию). Однако, используемый градиентный спуск не обязан приводить к достижению этого равновесия. Это наглядно видно, если рассмотреть следующий пример. Предположим, один игрок изменяет x , чтобы минимизировать $f_1(x) = xu$, в то время как другой игрок изменяет y , чтобы минимизировать $f_2(y) = -xu$. Так как $\frac{\delta f_1}{\delta x} = u$ и $\frac{\delta f_2}{\delta y} = -x$, то на каждой итерации градиентного спуска будет происходить одновременное обновление $x := x - \eta \cdot u$ и $y := y + \eta \cdot x$, соответственно (где η — шаг градиентного спуска). В силу того, что x и y имеют разные знаки, каждое последующее обновление градиента будет увеличивать по модулю как сами переменные, так и оптимизируемый функционал. Таким образом, равновесие не будет достигнуто (Рисунок 1.8) [22].

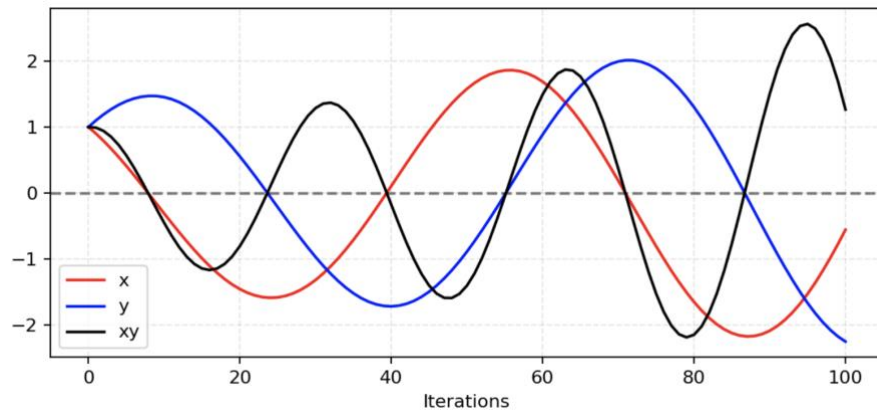


Рисунок 1.8 Имитация приведенного примера. С увеличением числа итераций колебание становится все более неустойчивым [22].

Другой частой проблемой при обучении GAN является коллапс мод. В каком-то смысле генератору выгоднее научиться создавать одно изображение, но очень реалистичное, которое дискриминатор не сможет отличить, чем научиться создавать множество различных изображений, но уступающих по своей реалистичности с точки зрения дискриминатора. Иначе говоря, генератор выдает изображение, которое дискриминатор не может отличить от реального, дискриминатор выучивается отличать это изображение, после чего генератор генерирует новое изображение, которое дискриминатор уже не может отличить от реального. Данный процесс будет повторяться, но при этом количество уникальных сгенерированных изображений не будет расти. Говоря иначе, генератор выучивается воспроизводить не исходное распределение данных, а лишь те примеры, которые позволяют «обманывать» дискриминатор и множество этих примеров сильно меньше исходного множества изображений. На Рисунок 1.9 наглядно проиллюстрирована вышеописанная проблема.

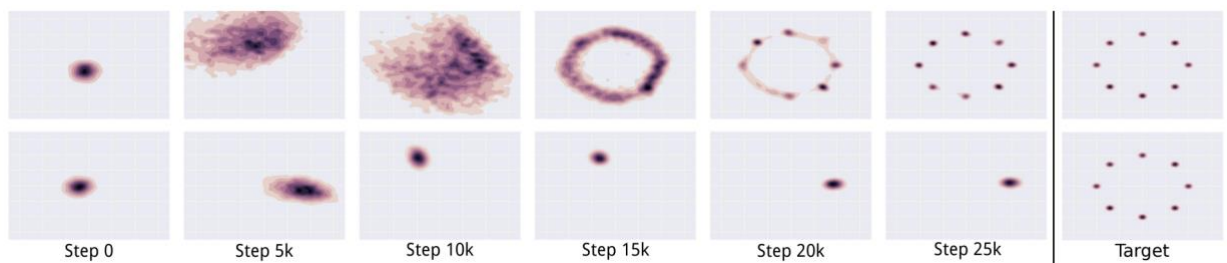


Рисунок 1.9 Коллапс мод при обучении GAN. Верхний ряд - модифицированная модель UnrolledGAN, устраняющая данную проблему. Нижний ряд - обычная GAN. С ростом обучения у обычной модели выходное распределение сжимается в более узкое пространство относительно исходного распределения [23].

Как и многие другие модели глубинного обучения, GAN страдают от проблемы затухания градиента [24]. Когда дискриминатор полностью обучился, получаем $D(x) =$

$1, \forall x \in p_{data}$ и $D(x) = 0, \forall x \in p_g$. Следовательно, оптимизируемый функционал L становится равным нулю и дальнейшего обучения не происходит (Рисунок 1.10). В результате при обучении GAN получается следующая дилемма:

- Если дискриминатор плохо обучен, то генератор не получает точной обратной связи о сгенерированных изображениях и функция ошибки не отображает реальный результат.
- Если дискриминатор обучен очень хорошо, то градиент функции ошибки приближается к нулю и обучение становится очень медленным или вовсе не происходит.

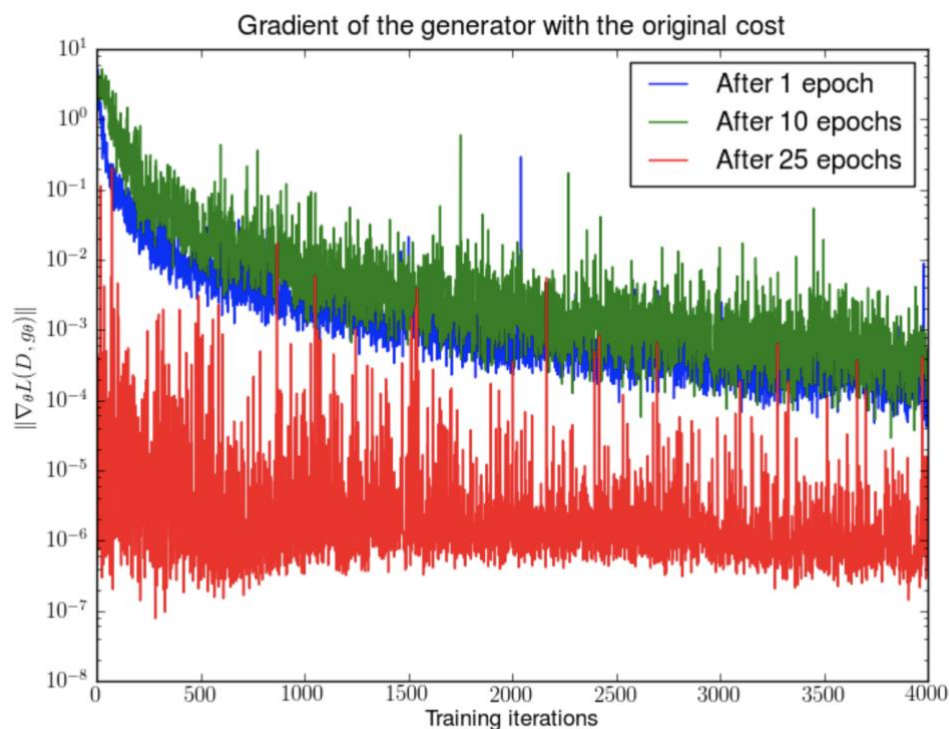


Рисунок 1.10 Моделируем ситуацию, когда дискриминатор обучился очень хорошо. Сперва обучаем модель DCGAN 1, 10 и 25 эпох. Затем фиксируем генератор и обучаем дискриминатор с нуля. Видим, что норма градиента очень быстро падает. Логарифмическая шкала по оси ординат [22].

По этим причинам для улучшения обучения GAN на практике предпринимаются различные трюки и преобразования [25, 26]. В данной работе мы не будем рассматривать все их многообразие, а лишь рассмотрим те, которое применяются в используемых архитектурах.

1.2.2 DCGAN

В силу большого количества неразмеченных изображений и видео, доступных в открытом доступе, возникает отдельная задача, заключающаяся в формировании эффективных и репрезентативных представлений вышеупомянутых типов данных, которые можно переиспользовать в различных задачах компьютерного зрения. Авторами статьи «*Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks*» было предложено формировать подобные представления изображений с помощью генеративных состязательных сетей [27]. Эта статья одновременно является первой, в которой удалось успешно применить свертки в GAN.

Авторы статьи экспериментально подобрали архитектуру модели, перебрав основные компоненты архитектуры сверточных нейронных сетей того времени. Ранее уже было отмечено, что есть ряд причин, по которым обучать GAN на практике достаточно сложно. Поэтому, авторы архитектуры DCGAN сформулировали ряд правил, которые позволяют успешно использовать свертки в GAN:

- Отказ от использования pooling-layers для понижения размерности, вместо этого используются свертки: в дискриминаторе - strided convolutions, в генераторе - fractional-strided convolutions (Рисунок 1.11).
- Использование batch normalization, как в генераторе (после каждой свертки, кроме последней), так и дискриминаторе (после каждой свертки, кроме первой и последней). Помогает стабилизировать обучение и протекать градиенту в глубоких моделях.
- Отказ от использования полносвязных слоев с целью увеличения глубины сети.
- Использование ReLU активации в генераторе для всех скрытых слоев, для выходного слоя использовать Tanh.
- Использование LeakyReLU+sigmoid для активации во всех слоях дискриминатора.

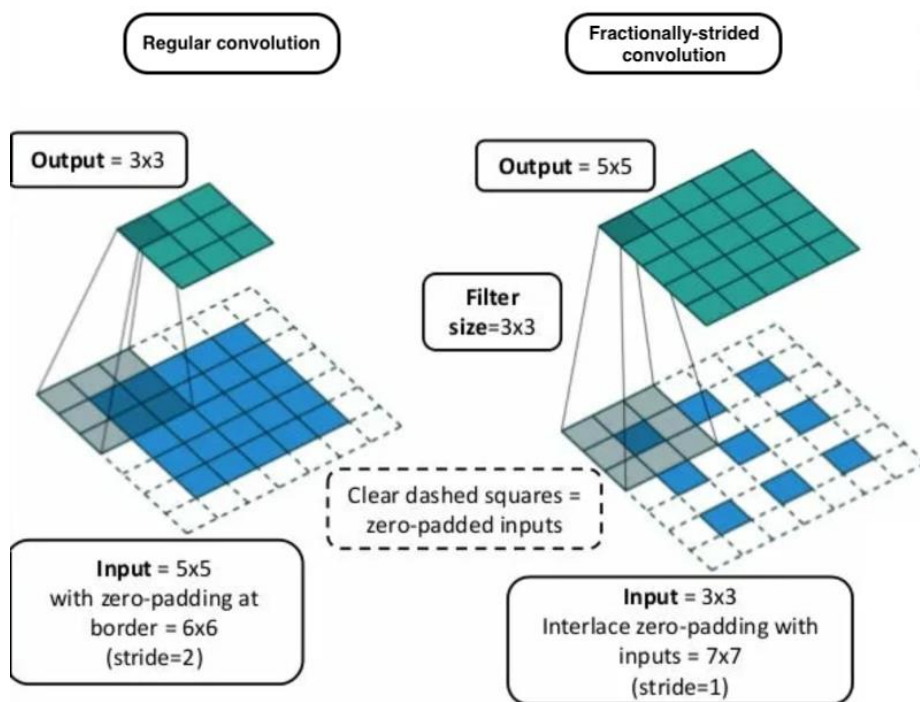


Рисунок 1.11 Сравнение обычной свертки и свертки "fractionally-strided convolution" [28].

Архитектура дискриминатора в DCGAN мало чем отличается от других сетей, а вот архитектура генератора выглядит достаточно необычно: на каждом последующем слое происходит сокращение числа каналов с одновременным увеличением размерности (Рисунок 1.12). Кроме того, авторы генерируют начальный вектор z из равномерного распределения, что достаточно необычно, так как обычно вектор z генерируют из нормального распределения.

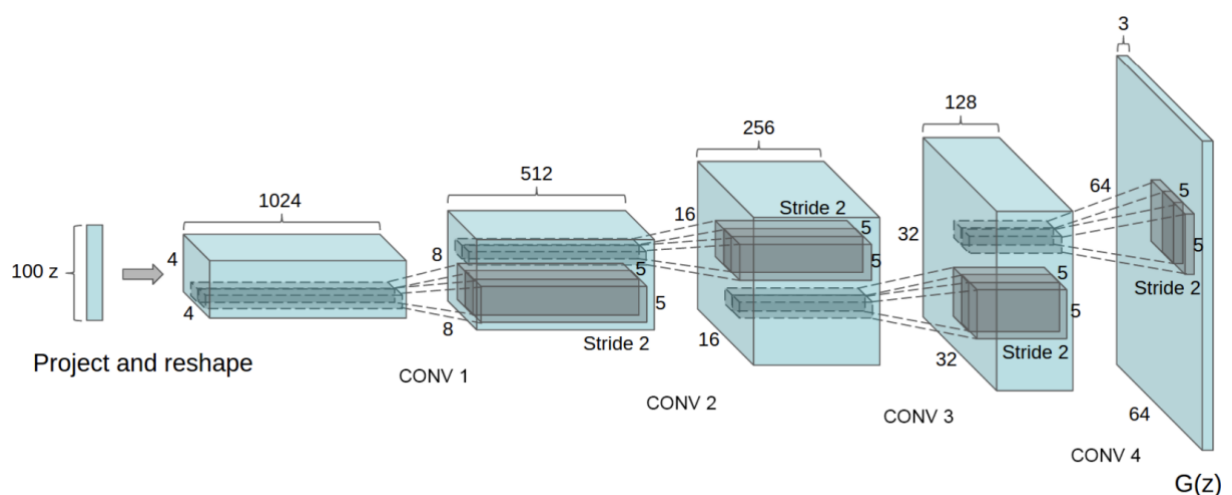


Рисунок 1.12 Архитектура генератора в DCGAN [27].

1.2.3 Семейство архитектур StyleGAN

1.2.3.1 Понятие стиля и виды нормализации

Архитектура нейронной сети StyleGAN впервые была предложена в статье «*A Style-Based Generator Architecture for Generative Adversarial Networks*» (StyleGAN) и в некотором смысле собрала в себе лучшие наработки в GAN того времени [29]. Поэтому, прежде чем описывать ее устройство, нужно разобрать некоторые ключевые идеи, которые предшествовали появлению архитектуры StyleGAN.

Первой такой идеей можно считать представление изображения в виде двух векторов, один из которых отвечает за контент, а второй за стиль. Говоря иначе, латентное пространство изображения может быть разложено на пространство контента и пространство стиля. При этом изображения из различных доменов могут иметь общее пространство контента, но не пространство стиля. Разумеется, это достаточно нестрогое разбиение, основанное на предположении авторов статьи «*Multimodal Unsupervised Image-to-Image Translation*», однако на практике оно дает интересные результаты (Рисунок 1.13) [30].



Рисунок 1.13 Пример наглядного разбиения на контент и стиль, в рамках решения задачи image-translation [30].

Второй важной идеей, которую стоит рассмотреть, является нормализация. Благодаря батч-нормализации (от англ. batch normalization - BN), значительно стабилизируется

обучение нейронных сетей с прямой связью (feed-forward neural networks) за счет нормализации признаков статистик [31]. Изначально алгоритм BN разрабатывался с целью ускорения дискриминативных моделей, однако, данная идея нашла эффективное применение и в генеративных моделях [27]. Для поступающего на вход батча $x \in R^{N \times C \times H \times W}$ BN нормирует среднее и стандартное отклонение для каждого признакового канала следующим образом:

$$BN(x) = \gamma \cdot \left(\frac{x - \mu(x)}{\sigma(x)} \right) + \beta$$

Где $\gamma, \beta \in R^C$ находятся из обучающих данных; $\mu(x), \sigma(x) \in R^C$ - среднее и стандартное отклонение, рассчитываемые по всем элементам батча и размерностям изображения для каждого признакового канала:

$$\mu_c(x) = \frac{1}{NHW} \sum_{n=1}^N \sum_{h=1}^H \sum_{w=1}^W x_{nchw}$$

$$\sigma_c(x) = \sqrt{\frac{1}{NHW} \sum_{n=1}^N \sum_{h=1}^H \sum_{w=1}^W (x_{nchw} - \mu_c(x))^2 + \epsilon}$$

Также существует Instance Normalization (IN). Было обнаружено, что нормализация не по всем элементам батча, а только по одному объекту позволяет существенно улучшить процесс обучения [32].

$$IN(x) = \gamma \cdot \left(\frac{x - \mu(x)}{\sigma(x)} \right) + \beta$$

$$\mu_{nc}(x) = \frac{1}{HW} \sum_{h=1}^H \sum_{w=1}^W x_{nchw}$$

$$\sigma_{nc}(x) = \sqrt{\frac{1}{HW} \sum_{h=1}^H \sum_{w=1}^W (x_{nchw} - \mu_{nc}(x))^2 + \epsilon}$$

Важным отличием IN от BN является то, что IN слои используются в инференс без изменений, в то время как в BN статистики по батчам заменяются на глобальные, посчитанные по всей обучающей выборке. Визуально BN и IN представлены на Рисунок 1.14.

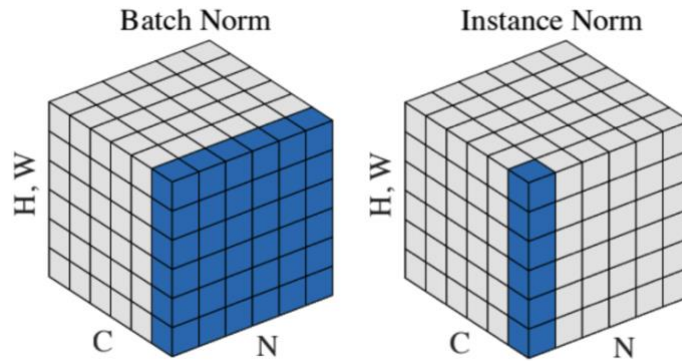


Рисунок 1.14 Визуализация алгоритмов BN и IN [33].

Более того, IN в каком-то смысле является нормализацией стиля изображения. Данный эффект впервые был замечен авторами статьи «*Arbitrary Style Transfer in Real-time with Adaptive Instance Normalization*» [34]. Также ими была предложена доработка алгоритма IN, которая позволяет осуществлять перенос стиля в пространство признаков путем переноса статистик признаков - Adaptive Instance Normalization (AdaIN):

$$AdaIN(x, y) = \sigma(y) \cdot \left(\frac{x - \mu(x)}{\sigma(x)} \right) + \mu(y),$$

где нормированный вход контента (x) масштабируется и сдвигается на статистики входа стиля (y) - $\sigma(y)$ и $\mu(y)$, соответственно. По аналогии с IN, статистики считаются для каждого объекта и каждого признакового канала.

1.2.3.2 StyleGAN

Архитектура StyleGAN во много унаследована от архитектуры Progressive GAN, которая представляет собой в каком-то смысле стекинг нейронных сетей - на каждом последующем шаге происходит удвоение разрешения изображения от 4x4 до 1024x1024 путем добавления нового блока, что позволяет модели сперва выучивать крупные черты изображения и постепенно переходить к мелким деталям [35]. В архитектуре StyleGAN дискриминатор такой же, как в архитектуре Progressive GAN, а большая часть нововведений приходится именно на генератор (Рисунок 1.15).

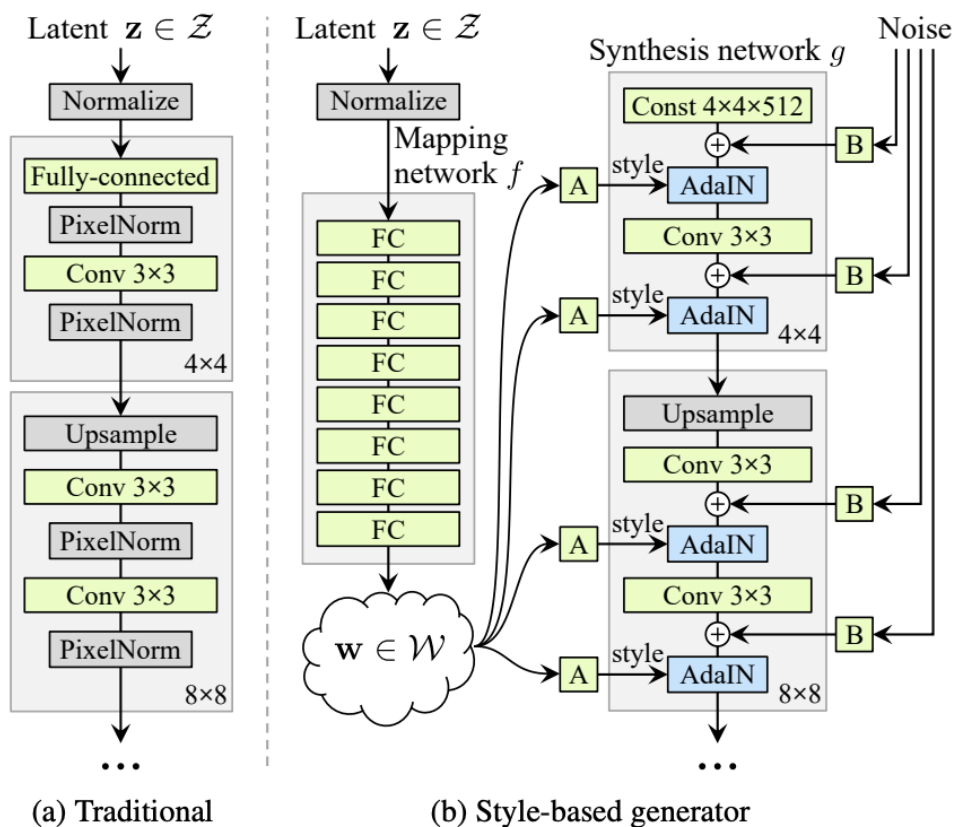


Рисунок 1.15 Сравнение архитектуры генератора ProgressiveGAN и StyleGAN [29].

Можно выделить ряд ключевых отличий архитектуры генератора StyleGAN. Во-первых, латентный вектор z теперь подается в Mapping Network, которая представляет собой последовательность полносвязных слоев (full-connected layers - FC). Как утверждают авторы статьи, благодаря ее использованию удастся перейти в так называемое распутанное представление - «*disentangled representation*», или представить латентное пространство в виде линейных подпространств. (Рисунок 1.16).

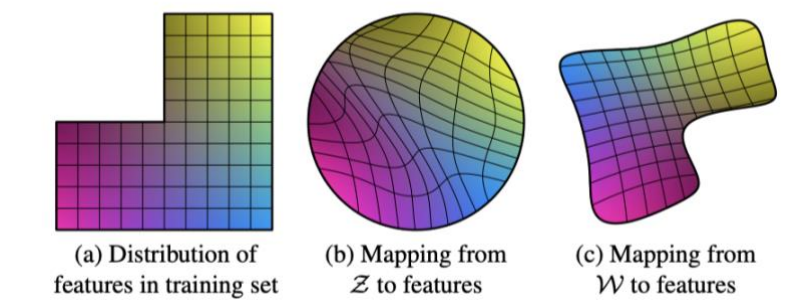


Рисунок 1.16 Визуализация распутывания представлений [29].

Говоря иначе, это позволяет сделать каждую компоненту вектора ответственной только за один признак.

Во-вторых, слои PixelNorm были заменены на слои AdaIN, благодаря которым осуществляется стилизация изображений (Рисунок 1.17). При этом, для трансформации латентного вектора w используется матрица A , задающая аффинное преобразование. Другими словами, вектор x сперва нормализуется, а затем модулируется на стиль, задающийся вектором y .

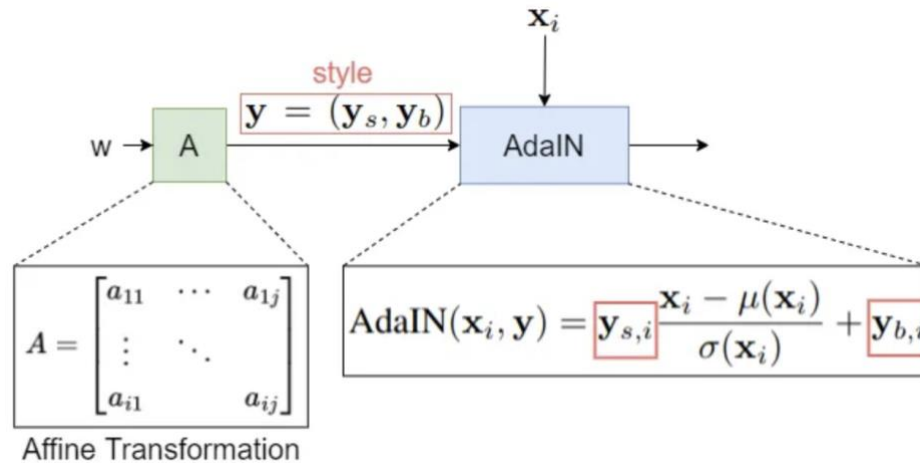


Рисунок 1.17 Стилизация посредством использования AdaIN в генераторе модели StyleGAN. Источник: <https://medium.com/@steinsfu/stylegan-vs-stylegan2-vs-stylegan2-ada-vs-stylegan3-c5e201329c8a#291d>

В-третьих, в качестве входа в «*Synthesis network g*» подается не шум, а константный тензор размера $4 \times 4 \times 512$. Это в каком-то смысле необходимость, так как нужно инициализировать входной тензор. При этом шум добавляется ко всем каналам перед AdaIN операцией и представляет собой вариацию стилей на каждом уровне абстракции. Параметр B определяет, насколько сильно будет влиять уровень шума на конечный результат, говоря иначе - насколько сильным будет разнообразие генерируемых объектов.

В-четвертых, для операции «*Upsample*» вместо алгоритма «*nearest neighbor layers*» был использован алгоритм «*bilinear sampling*». В-пятых, была использована процедура «*Style Mixing Regularization*», заключающаяся в подаче различных латентных стилевых векторов w в различные уровни абстракции - часть AdaIN получает один вектор, а все последующие получают другой вектор, так как считается, что на каждом уровне стиль контролирует свой уровень абстракции (Рисунок 1.18).

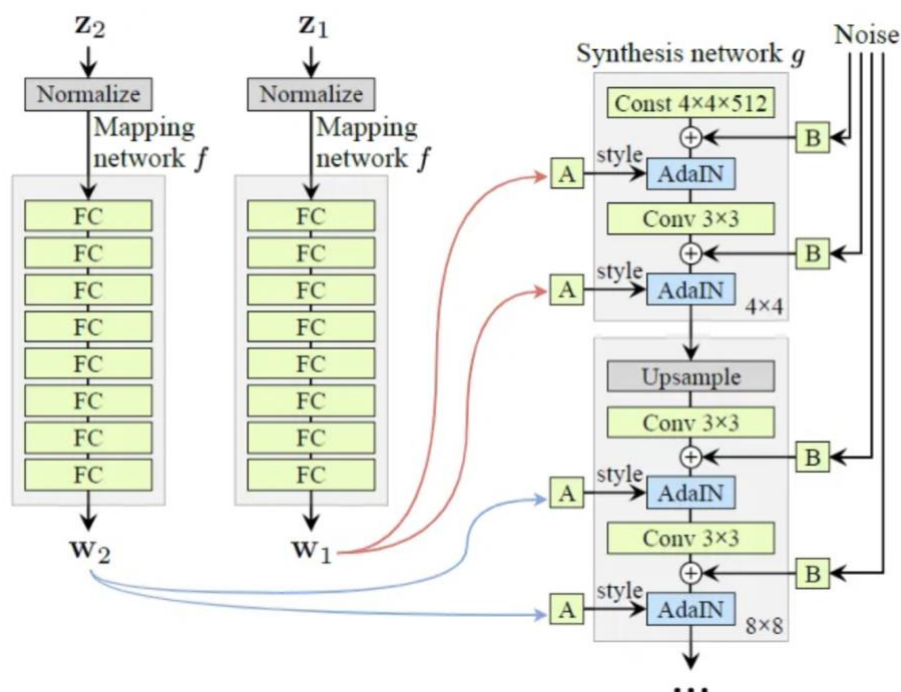


Рисунок 1.18 Пример работы Style Mixing Regularization в архитектуре генератора модели StyleGAN. Источник: <https://medium.com/@steinsfu/stylegan-vs-stylegan2-vs-stylegan2-ada-vs-stylegan3-c5e201329c8a#291d>

1.2.3.3 StyleGAN2

Архитектура StyleGAN2 была создана на базе StyleGAN и основной мотивацией к ее созданию являлось избавление от артефактов, которые встречались на изображениях [36]. Прежде чем рассмотреть нововведения, имеет смысл подробнее взглянуть на некоторые аспекты архитектуры StyleGAN. Так, операцию AdaIN можно разложить на две - нормализацию и модуляцию (Рисунок 1.19). Авторами архитектуры StyleGAN2 было замечено, что исключение шага нормализации позволяет полностью избавиться от некоторых артефактов в генерируемых изображениях, поэтому ими были предприняты различные шаги, чтобы корректно использовать это наблюдение.

$$\text{AdaIN}(\mathbf{x}_i, \mathbf{y}) = \underbrace{\mathbf{y}_{s,i} \frac{\mathbf{x}_i - \mu(\mathbf{x}_i)}{\sigma(\mathbf{x}_i)}}_{\text{Модуляция}} + \underbrace{\mathbf{y}_{b,i}}_{\text{Нормализация}}$$

Рисунок 1.19 Разложение операции AdaIN

На Рисунок 1.20 (a, b) детально отображены старая и новая архитектуры, чтобы было легче проследить, каких элементов коснулись изменения, добавленные в новую архитектуру. В частности, были внесены следующие изменения: убраны некоторые избыточные операции в начале, добавлено b и B вынесено из области слоя, отвечающего за стиль (авторами архитектуры было замечено, что это дает более предсказуемые результаты), в качестве поправки используется только стандартное отклонение (Рисунок 1.20c). Пересмотренная архитектура позволяет заменить операцию IN операцией демодуляции, которая применяется к весам, связанным с каждым слоем свертки (Рисунок 1.20d). После модуляции (изменения масштаба без корректировки смещения на среднее) веса свертки стали такими:

$$w'_{ijk} = s_i \cdot w_{ijk} \text{ (модуляция)}$$

Каждая признаковая карта (feature map) получила свой масштаб, поэтому нормировка на σ_i выполняет роль нормализации:

$$\sigma_i = \sqrt{\sum_{ik} w'_{ijk}{}^2}$$

$$w''_{ijk} = \frac{w'_{ijk}}{\sqrt{\sum_{ik} w'_{ijk}{}^2 + \epsilon}} \text{ (демодуляция)}$$

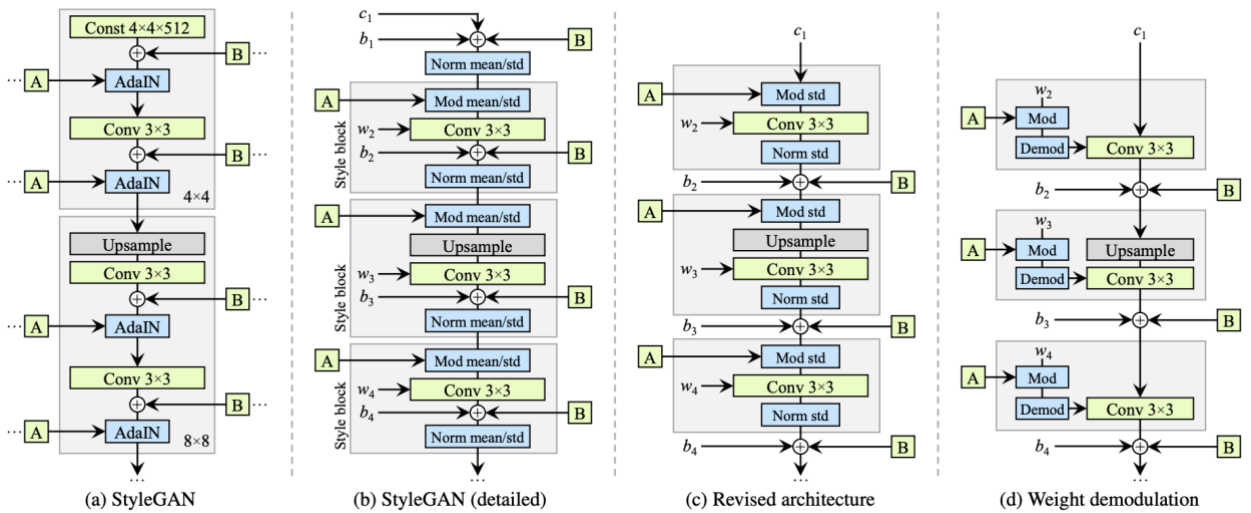


Рисунок 1.20 Изменения архитектуры StyleGAN. (a) Оригинальная архитектура. (b) Более подробное представление с разложением операции AdaIN и выделением весов - w , смещений - b и константного входа - c . (c) Пересмотренная архитектура без использования среднего. (d) Пересмотренная архитектура с операцией демодуляции [36].

Также были выявлены проблемы с первоначальной архитектурой, которая была основана на Progressive GAN, заключающиеся в том, что постепенно расширяющийся генератор, по-видимому, отдает предпочтение расположению деталей и не может

динамически подстраивать их под незначительные изменения изображений. Поэтому новая архитектура использует подход «skip-connection» и дизайн нейронной сети ResNet [37].

1.2.3.4 StyleGAN2-Ada

После выхода архитектуры StyleGAN2 авторы решили доработать ее, чтобы была возможность обучаться на небольших наборах данных, так появилась архитектура StyleGAN2-Ada [38]. При обучении на небольших наборах данных основной проблемой является переобучение дискриминатора, который быстро переобучается на обучающую выборку и перестает давать полезную обратную связь генератору [39]. Практически во всех сферах глубинного обучения аугментация является стандартным средством для борьбы с переобучением [40]. Однако, в случае с генеративными моделями это может привести к тому, что генератор будет воспроизводить не исходное распределение, а аугментированное. Другими словами, генерируемые изображения также будут содержать аугментации, что является крайне нежелательным для решения большинства прикладных задач [41]. По этой причине авторы данной архитектуры разработали адаптивные стохастические аугментации, которые позволяют успешно применять различные трансформации изображений, но при этом не позволяют трансформациям попадать в генерируемые изображения. Ранее уже предпринимались попытки реализовать аугментации, которые не дают утечки трансформаций. Так, в статье «*Improved consistency regularization for GANs*» был предложен подход под названием «balanced consistency regularization (bCR)» [42]. Его идея заключается в том, что реальные данные и сгенерированные данные до аугментации не должны сильно отличаться от реальных и сгенерированных данных, полученных после аугментации. Данная регуляризация вводится на уровне ошибки дискриминатора, поэтому генератор никак не штрафуются за генерацию изображений с аугментациями. Поэтому авторы StyleGAN2-Ada предложили новый подход, который они назвали «stochastic discriminator augmentation», вносящий в bCR два изменения: 1) штрафовать не только дискриминатор, но и генератор; 2) подавать на вход только аугментированные изображения. На первый взгляд, это кажется достаточно контринтуитивным, так как дискриминатор не видит оригинальных изображений, чтобы давать корректную обратную связь генератору, однако, на практике такой подход при определенных условиях дает хорошие результаты [38]. На Рисунок 1.21 для сравнения наглядно изображены два описанных выше подхода.

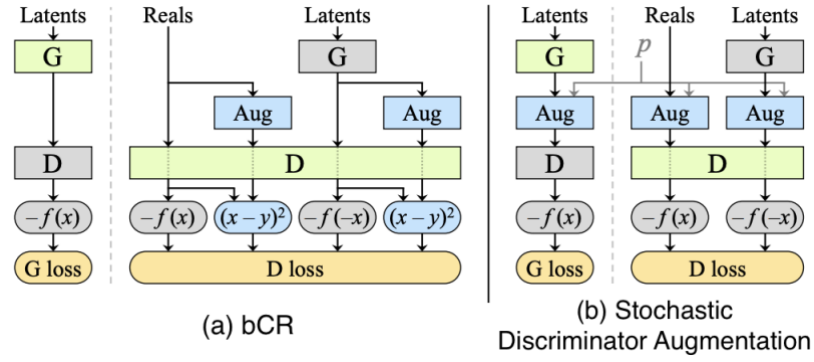


Рисунок 1.21 Сравнение bCR и Stochastic Discriminator Augmentation [38].

Также авторы проводят интересный эксперимент, в котором применяют 18 различных аугментаций с заданной вероятностью и оценивают, как от значения вероятности зависит попадание аугментации в генерируемые изображения. По результатам эксперимента авторы приходят к следующим выводам: аугментации действительно помогают получить более высокое качество при небольшом количестве данных; для различных аугментаций и наборов данных требуются различные значения вероятности. Однако, перебор оптимального значения вероятности является очень долгой вычислительной процедурой, поэтому предлагается сделать вероятность адаптивной. Таким образом, с помощью изменения вероятности можно будет бороться с переобучением дискриминатора - чем больше он начинает переобучаться, тем сильнее вероятность аугментации. Авторами предлагается использовать следующие эвристики, которые отражают переобучение и будут триггером для изменения значения вероятности (где E — среднее значение, посчитанное по нескольким батчам):

$$r_v = \frac{E[D_{train}] - E[D_{validation}]}{E[D_{train}] - E[D_{generated}]}$$

$$r_t = E[\text{sign}(D_{train})]$$

Основная идея эвристики r_v заключается в том, что обучающий и валидационный набор данных взяты из одного распределения, поэтому в условиях отсутствия переобучения дискриминатор должен считать их близкими. В случае переобучения дискриминатор будет считать, что наиболее близкими к валидационному набору данных являются сгенерированные данные (Рисунок 1.22). Минусом данной эвристики является то, что она требует валидационный набор данных, что может быть не всегда уместно, особенно когда общий размер выборки данных является небольшим.

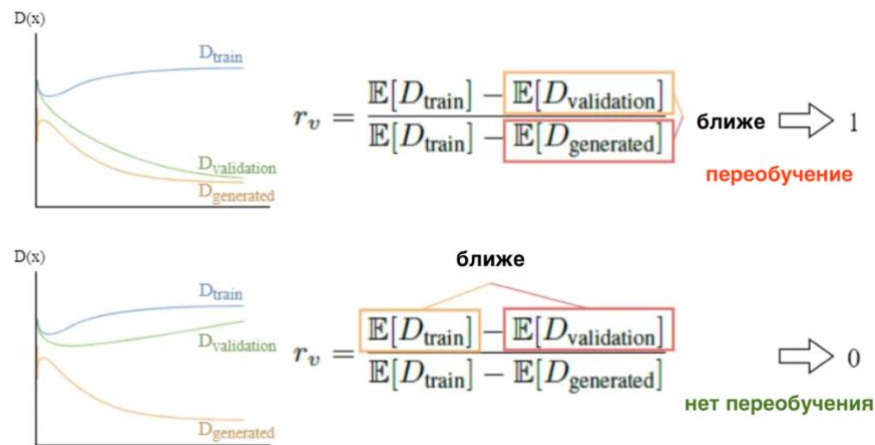


Рисунок 1.22 Поведение эвристики, оценивающей переобучение.

Источник: <https://medium.com/@steinsfu/stylegan-vs-stylegan2-vs-stylegan2-ada-vs-stylegan3-c5e201329c8a#4dad>

Эвристика r_t измеряет долю обучающей выборки, на который дискриминатор дает положительное значение. На Рисунок 1.23 приведено сравнение обычной архитектуры StyleGAN2 и с использованием Ada. Наглядно видно, что данная техника позволяет получать более высокое качество по метрике FID.

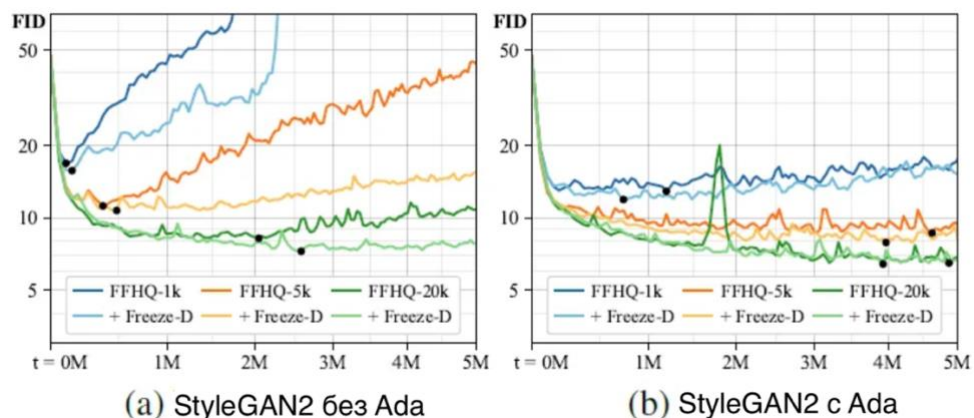


Рисунок 1.23 Сравнение StyleGAN2 и StyleGAN2-Ada

Таким образом, архитектура StyleGAN2 собрала в себе лучшие подходы и эвристики работы с GAN, которые появились с момента появления генеративных состязательных моделей. Количественно это подтверждается результатами на общепринятых бенчмарках, опубликованных на сайте paperswithcode.com¹, где в категории «Conditional-image-generation» на датасете «cifar-10» данная архитектура и ее последующие версии занимают высокое положение в рейтинге. Что касается архитектуры StyleGAN2-Ada, то она имеет ряд доработок, которые позволяют эффективно обучаться на небольших наборах данных, и

¹ <https://paperswithcode.com/sota/conditional-image-generation-on-cifar-10>

именно поэтому ее использование видится перспективным для решения поставленной задачи.

Глава 2 Материалы и методы

2.1 Данные

2.1.1 Сложности сбора медицинских данных

Сбор медицинских данных является достаточно сложной задачей в силу множества факторов [43]. Во-первых, это конфиденциальность данных - по понятным причинам пациенты обычно не хотят, чтобы их врачебная тайна была раскрыта [44]. Во-вторых, может потребоваться специальное оборудование, позволяющее преобразовывать аналоговый сигнал в цифровой, соответствующее всем медицинским и лабораторным требованиям. Более того, на уровне медицинского учреждения требуется инфраструктура, которая сможет хранить и обрабатывать собираемые данные [45]. В-третьих, даже собранные медицинские данные часто содержат в себе множество неточностей, для устранения которых требуется разметка с помощью эксперта или применение алгоритмов для поиска ошибок [46].

Ранее было отмечено, что для сбора изображений эмбрионов требуется наличие специального time-lapse инкубатора, который поддерживает оптимальные условия для развития эмбриона, нарушение которых может привести к его гибели. Тем не менее, даже наличие time-lapse инкубатора не отменяет описанных выше проблем, связанных с приватностью и хранением данных. Именно поэтому в открытом доступе не так много публичных наборов данных, содержащих изображения эмбрионов, которые можно было бы использовать как обучающую выборку, что затрудняет решение различных задач компьютерного зрения.

2.1.2 Источники данных

Для решения поставленной задачи мне требовалось найти набор данных, содержащий изображения человеческих эмбрионов на стадии бластоцисты, используя который можно было бы производить обучение моделей. Выше было отмечено, какие сложности возникают при сборе подобных изображений, поэтому в открытом доступе удалось найти лишь два источника данных.

Первый был получен по запросу у авторов статьи «*Automatic Identification of Human Blastocyst Components via Texture*» [47]. Набор данных представлен 249 изображениями (Рисунок 2.1), а также содержит дополнительные сегментационные маски. Разрешение изображений варьируется от 300 до 600 пикселей по обоим измерениям. Примечательность этого набора данных в том, что он уже содержит разметку на классы по шкале Гарднера.

Однако, в последствии от этого набора данных пришлось отказаться в силу того, что изображения, представленные в нем, очень сильно отличались по своим визуальным характеристиками от второго источника данных.

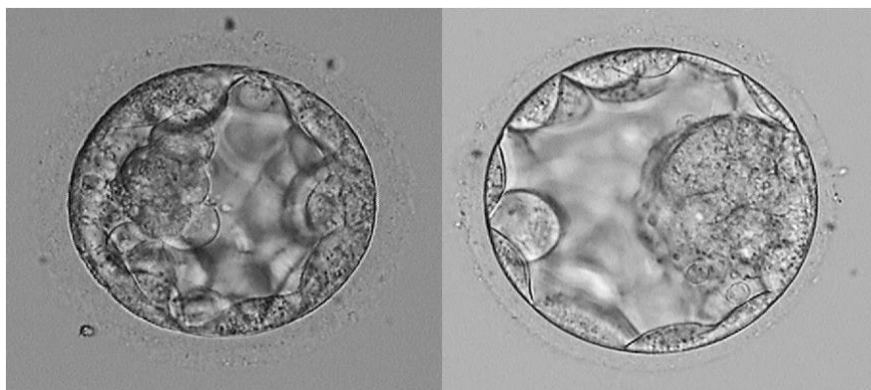


Рисунок 2.1 Пример изображений из набора данных №1

Второй набор данных находится в открытом доступе - «*A time-lapse embryo dataset for morphokinetic parameter prediction*» - и содержит 2.4 миллиона изображений всех стадий развития эмбриона, начиная со стадии зиготы и заканчивая стадией бластоцисты (Рисунок 2.2) [48]. Разрешение изображений составляет 500 на 500 пикселей по обеим сторонам. Данные изображения были получены из 704 time-lapse видео (каждое соответствует одному эмбриону), снятых в 7 фокальных плоскостях при помощи time-lapse инкубатора EmbryoScope (Vitrolife). Примечательность этого набора данных в его большом размере, однако он не содержит разметку на классы по шкале Гарднера. Тем не менее, именно этот набор данных был выбран в качестве основного, а разметка была выделена в отдельную задачу.

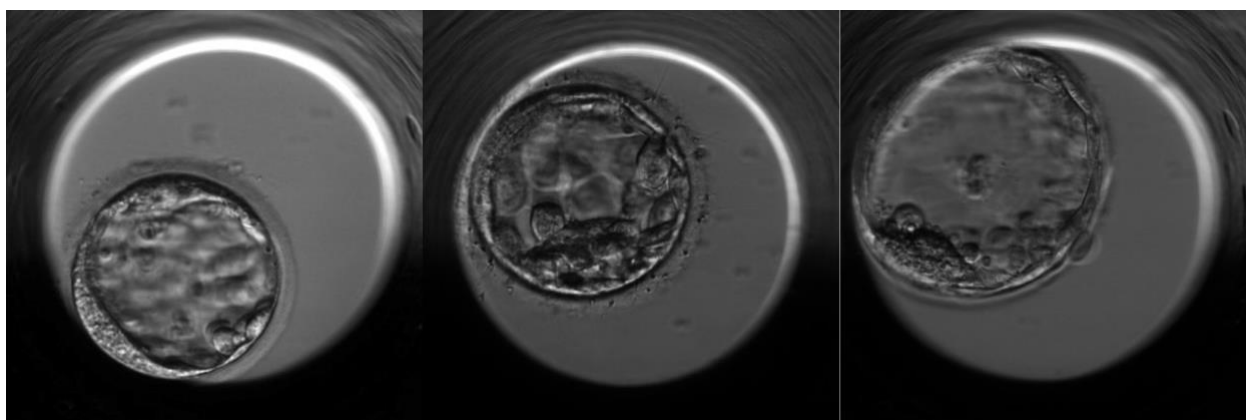


Рисунок 2.2 Пример изображений из набора данных №2

2.1.3 Разметка данных на классы

В Главе 1.1.2 было подробно описано, как формируется оценка качества эмбриона на основании шкалы Гарднера: цифры от 1 до 6 и двух букв от А до С. Если считать все возможные варианты комбинаций, составляющих оценку качества, то получится 54 класса. Это достаточно много, и чтобы обучить модель, нужно собрать достаточно данных по каждому из 54 классов. Поэтому, в данной работе при разбиении на классы не был учтен размер бластоцист, обозначаемый цифрой. Это обусловлено не только необходимостью сокращения количества классов, но и особенностью используемых данных. В клинической практике эмбрионы реже дорастают до стадий 5 и 6, так как их стараются использовать на более раннем этапе, а потому этот набор данных содержит существенно меньше изображений бластоцист на этих стадиях. Кроме того, бластоцисты на стадии хэтчинга и уже завершившие хэтчинг имеют существенно больший размер и полностью не помещаются в поле зрения при съемке. Что касается бластоцист 1 стадии развития, то они относятся к ранним бластоцистам и качество их TE и ICM оценить невозможно.

Таким образом, для дальнейшего анализа использовались только те изображения, у которых размер эмбриона и степень его экспансии оценивается от 2 до 4 включительно, но при разбиении на классы размер не учитывался. При таком подходе все возможные комбинации классов описываются только двумя буквами от А до С и их количество равняется 9. Было принято решение сгруппировать оценки по шкале Гарднера в три группы: «Good» - AA, AB; «Average» - BB, AC, BA; «Poor» - CC, BC, CB, CA (Рисунок 2.3). Такое разделение на группы основано на исследованиях о вероятности наступления беременности при переносе в полость матки одного эмбриона соответствующего качества [49, 50].

ICM\ TE	A	B	C
A	AA	AB	AC
B	BA	BB	BC
C	CA	CB	CC

Рисунок 2.3 Группировка оценок по шкале Гарднера на классы. Строка - оценка TE, столбец – оценка ICM.

Обычно для разметки данных и последующего обучения моделей используются краудсорсинговые платформы, где неспециалисты после краткого обучения выполняют задания по разметке. В данной работе для обучения моделей используются изображения, полученные из клинической практики и снятые под микроскопом, что существенно осложняет их разметку людьми без релевантного опыта работы и без представления о том, как эмбрионы и клетки выглядят под микроскопом. Оценка качества эмбрионов по Гарднеру является специфическим навыком, которым владеют лишь клинические эмбриологи с достаточным опытом работы. Поэтому в данной работе для разметки набора данных был привлечен практикующий специалист из сферы ВРТ – Холостина Алина Михайловна, эмбриолог ГУЗ КПЦСО.

2.1.4 Формирование обучающей выборки

По результатам разметки было получено 3500 изображений со следующим распределением по классам: «Good» - 32%, «Average» - 32%, «Poor» - 36% от размера выборки. Таким образом, дисбаланса классов нет. Разрешение всех изображений было понижено с 500x500 до 256x256 пикселей с целью ускорения вычислений. Вырезание части изображения не производилось, так как используемые изображения уже центрированы и изображение эмбриона занимает практически всю площадь изображения. Дополнительных преобразований изображений на данном этапе не производилось.

2.2 Метрики оценки качества генерации

2.2.1 IS

Метрика Inception Score (IS) впервые была предложена в статье «*Improved Techniques for Training GANs*» [25]. Мотивацией к ее созданию было желание авторов получить метрику, которая бы хорошо коррелировала с человеческим восприятием изображения, на примере сравнения с разметкой средствами краудсорсинга. Главная идея IS заключается в оценке множества сгенерированных изображений с помощью дивергенции Кульбака-Лейблера. Данная метрика стремится, с одной стороны, измерить качество генерируемых изображений, а с другой стороны - их разнообразие. Для подсчета метрики используется предварительно обученная модель классификации InceptionV3 [51] на наборе данных ILSVRC 2012 [52], который содержит 150k изображений, распределенных среди 1000 классов.

Алгоритм расчета метрики выглядит следующим образом:

1. Генерируем большое количество изображений (авторы рекомендуют 50k), перебирая все возможные классы из заданного при обучении множества (условная генерация).
2. Перемешиваем полученные изображения с целью, чтобы сгенерированные изображения для одного класса не шли подряд.
3. Нормализуем полученные изображения, чтобы привести их ImageNet.
4. Делим набор сгенерированных изображений на n групп. Обозначим за k количество изображений, попавших в одну группу.
5. Для каждой из n групп:
 - a. Пропускаем все k изображений через предварительно обученную модель классификации InceptionV3. Получаем для каждого изображения распределение вероятностей принадлежности к одному из 1000 классов. Полученные вероятности будут условными вероятностями, так как изначально мы генерировали изображения с условием на класс. Полученный объект будет матрицей вида $P(y|x) \in R_{k \times 1000}$.
 - b. Для каждого из 1000 классов усредняем значение полученной условной вероятности $P(y|x)$ по всем изображениям, получаем вектор вида $P(y) \in R_{1000}$.
 - c. Считаем дивергенцию Кульбака-Лейблера, получаем матрицу $KL_D \in R_{k \times 1000}$.

$$KL\left(\mathbb{P}(y|x) \parallel \mathbb{P}(y)\right) = \mathbb{P}(y|x) * \left(\log(\mathbb{P}(y|x)) - \log(\mathbb{P}(y))\right)$$

- d. Для каждого изображения считаем сумму по всем, получаем вектор $SumKL_D \in R_k$.
- e. Усредняем его по всем изображениям, получаем скаляр $AvgSumKL_D \in R$, от которого берем экспоненту. Это и будет IS для данной группы изображений.
6. Усредняем полученные значения по всем группам.

Распределение условных вероятностей $P(y|x)$ отражает оценку качества сгенерированных изображений, так как изображения, которые при классификации имеют большую вероятность принадлежности к одному из классов относительно всех остальных классов, должны содержать какие-то осмысленные объекты, говоря иначе, обладать более высоким качеством генерации. В таком случае энтропия условных вероятностей сгенерированных изображений будет низкой. Распределение частных вероятностей $P(y)$ отражает вариативность сгенерированных изображений, так как разные изображения будут относиться к разным классам при классификации. В таком случае энтропия частных вероятностей сгенерированных изображений будет высокой. Таким образом, с помощью дивергенции Кульбака-Лейблера мы оцениваем расстояние между условным и частным распределением вероятностей.

У IS есть ряд недостатков [53]:

1. Чувствительность метрики к небольшим изменениям весов сети, обусловленных случайной инициализацией при классификации изображений с помощью InceptionV3.
2. При подсчете метрики N сгенерированных изображений разбиваются на n групп. Обычно, $N=50k$ и $n=10$, то есть в каждую группу попадает только 5k значений, которые классифицируются на принадлежность к одному из 1k классов. Такого количества изображений может быть недостаточно, чтобы получить хорошую оценку частного распределения.
3. Если те объекты, которые мы генерируем, не являются близкими ни к одному из 1k классов модели InceptionV3, то метрика может принимать необоснованно низкие значения.

4. В разных задачах могут быть разные критерии качества генерируемых изображений. Так, например, изображения, где человек имеет аномальное число конечностей, могут никак не штрафоваться метрикой.
5. Несмотря на заявление авторов метрики о ее высокой корреляции с разметкой краудсорсинга, сгенерированные изображения могут иметь совершенно нереалистичный вид, но при этом иметь высокое значение метрики (Рисунок 2.4).
6. Метрика никак не сравнивает сгенерированные изображения с реальными, на которых обучалась генеративная модель, что является критически важным во многих прикладных задачах генерации изображений.



Рисунок 2.4 Пример сгенерированных изображений с $IS=900.15$ [53].

2.2.2 FID

Метрика Fréchet inception distance (FID) впервые была предложена в статье «*GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium*» [54]. Главной мотивацией к созданию метрики являлась возможность оценивать не только сгенерированные изображения сами по себе, но и, в отличие от метрики IS, оценивать их относительно реальных изображений, на которых обучалась модель. Главная идея FID заключается в сравнении статистики множества сгенерированных изображений и множества реальных изображений. Для подсчета метрики используется предварительно обученная модель классификации InceptionV3.

Алгоритм расчета метрики выглядит следующим образом:

1. Генерируем большое количество изображений (авторы рекомендуют 50k).

2. Пропускаем их и множество имеющихся реальных изображений через InceptionV3 без финального классификационного слоя, чтобы получить набор признаков для каждого изображения. Получаем матрицу $X \in R_{k \times 2048}$ для реальных изображений и матрицу $G \in R_{n \times 2048}$ для сгенерированных изображений, где k – число реальных изображений, n – число сгенерированных изображений ($n=50k$), 2048 – число признаков изображения.
3. Считаем среднее по всем изображениям отдельно для каждого из 2048 признаков - для множества реальных изображений и множества сгенерированных изображений - μ_x, μ_g .
4. Считаем матрицу ковариации для множества реальных изображений и для множества сгенерированных изображений - Σ_x, Σ_g .
5. Будем считать распределение реальных объектов и сгенерированных объектов гауссианами на сверточных признаках InceptionV3. При некоторых предположениях равенство распределений равносильно равенству их моментов. Вычисляем расстояние Фреше между гауссианами.

$$FID(x, g) = \|\mu_x - \mu_g\|_2^2 + Tr(\Sigma_x + \Sigma_g - 2(\Sigma_x \Sigma_g)^{0.5})$$

Данная метрика показывает себя достаточно хорошо с точки зрения робастности и эффективности вычислений несмотря на то, что сравнивает только первые два момента распределений. Как и IS, она достаточно согласована с разметкой краудсорсингом, но при этом демонстрирует большую устойчивость к шуму [54]. Также, в отличие от IS, данная метрика чувствительна к падению моды внутри класса (*intra-class mode dropping*), то есть модель, которая генерирует по одному изображению для каждого класса, может иметь высокое значение (высокое качество) IS, но будет иметь низкое значение (плохое качество) метрики FID [55]. На Рисунок 2.5 наглядно видно, что FID монотонно возрастает с увеличением степени деформации изображения.

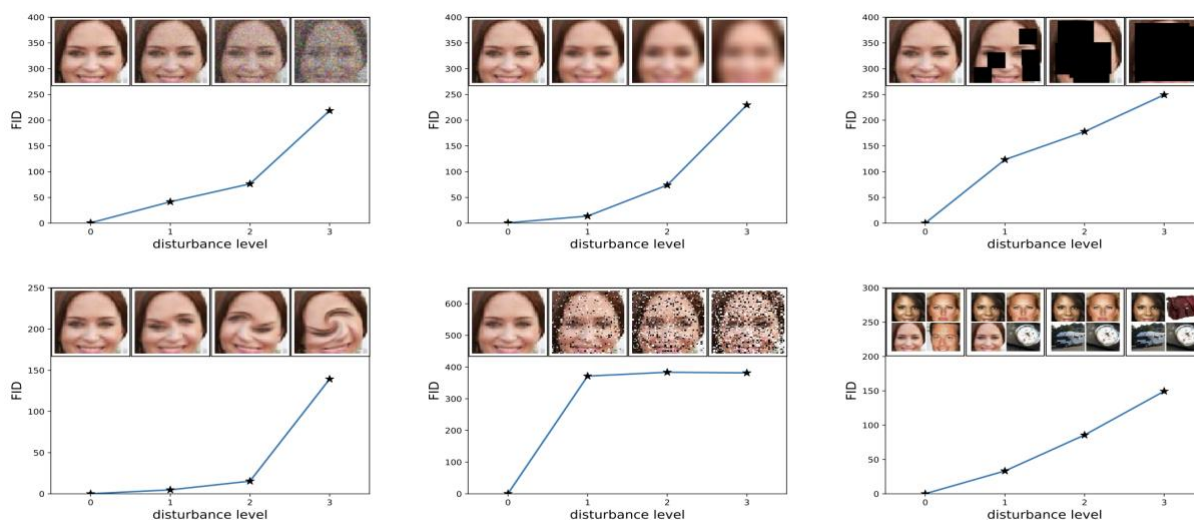


Рисунок 2.5 Изменение FID в зависимости от увеличения различных деформаций [54].

Среди недостатков данной метрики можно выделить:

1. Предположение о нормальном распределении признаков изображений, что часто не гарантируется на практике [55].
2. Как и для метрики IS, есть неоднозначность в сопоставлении исходных классов задачи с классами набора данных ImageNet [56]. Говоря иначе, признаки, которые выделяет InceptionV3, обученная на наборе данных ImageNet, могут быть не репрезентативны для конкретной прикладной задачи.
3. На практике метрика не всегда коррелирует с разметкой краудсорсинга — это наглядно видно на Рисунок 2.6.

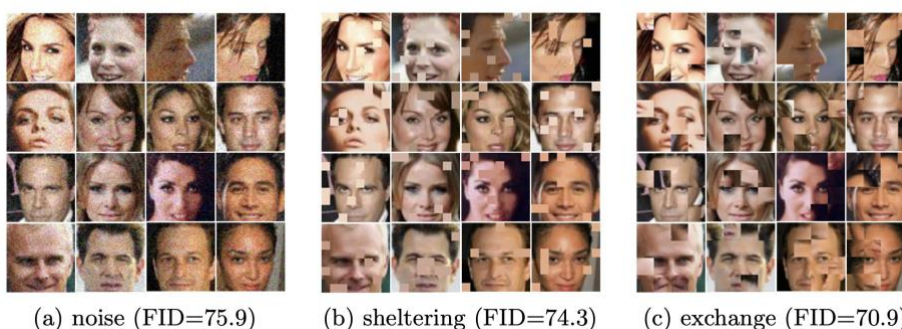


Рисунок 2.6 Пример некорректного поведения метрики FID [56].

Несмотря на приведенные недостатки, в настоящее время FID является самой распространенной метрикой для сравнительной оценки качества генерации, заменившей IS. Поэтому в дальнейших экспериментах именно она будет использоваться для сравнительной оценки.

2.3 Окружение для проведения экспериментов

Для проведения экспериментов был выбран язык программирования Python версии 3.7. В качестве реализации архитектуры генеративных состязательных моделей, описанных в Главе 0, использовалась реализация с помощью open-source модуля PyTorch версии 1.7.1. Остальные модули и их версии можно найти в репозитории с исходным кодом, воспроизводящим результаты данной работы. Для версионирования программного кода был выбран сервис github² - весь исходный код работы размещен там. Для трекинга экспериментов был выбран сервис wandb.ai, так как он позволяет отслеживать прогресс обучения моделей, сравнивать их друг относительно друга, а также сохранять артефакты, полученные по результатам работы моделей. Обучение моделей производилось с использованием gpu-ресурсов для ускорения расчетов.

² <https://github.com/kirilldolmatov/diploma>

Глава 3 Эксперименты

3.1 Генерация изображений с помощью DCGAN

В свое время архитектура DCGAN стала прорывом, так как позволила эффективно использовать сверточные нейронные сети внутри GAN. Сейчас данная архитектура уже не является актуальной, но может послужить отправной точкой для будущих экспериментов. Именно поэтому она была выбрана в качестве базового решения.

В исходный код была добавлена возможность работы с условной генерацией, а также трансформация изображений в одноканальный режим (черный и белый цвета). Был осуществлен перебор следующих параметров: размер батча, размер латентного (входного) вектора, значение β_1 и шаг обучения для оптимизатора Adam, число эпох обучения, а также число итераций обучения дискриминатора на одну итерацию обучения генератора.

Полученные результаты очень сложно назвать качественными (похожими на исходные), даже с точки зрения человека, не обладающего экспертными знаниями в эмбриологии (Рисунок 3.1). Как легко заметить, полученные изображения являются очень нечеткими. Во многом такой результат может быть обусловлен тем, что архитектура DCGAN изначально не проектировалась для генерации изображений с высоким разрешением (полученные изображения имеют разрешение 28x28). Кроме того, исходные изображения содержат очень много деталей, которые может быть сложно воспроизвести при относительно неглубокой архитектуре DCGAN.

Исходя из полученных результатов было принято решение сразу перейти к наиболее релевантной архитектуре для решения поставленной задачи - StyleGAN2-Ada.

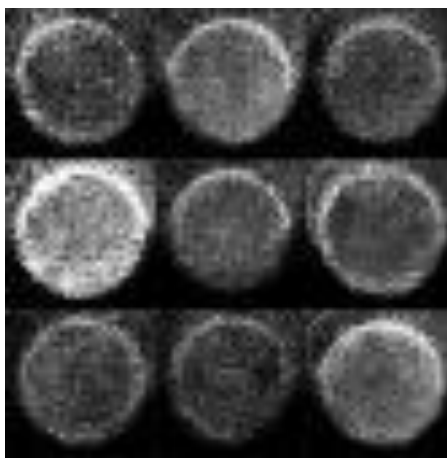


Рисунок 3.1 Результаты генерации DCGAN

3.2 Генерация изображений с помощью предобученной StyleGAN2-Ada

Использование предобученных моделей является стандартной практикой в задачах обучения глубоких нейросетевых моделей, так как, зачастую, это позволяет получить более качественный результат за меньшее время. Кроме того, для обучения больших моделей с нуля обычно требуется достаточно много входных данных, которых может просто не быть при решении тех или иных задач. Поэтому было принято решение провести ряд экспериментов с предобученными моделями. Однако, важно заметить, что в силу устройства архитектуры StyleGAN2-Ada, предобученные модели не могут использоваться для задачи условной генерации.

Эксперимент №1: модель, обученная на наборе данных CelebA-HQ, который содержит 30k изображений с портретами знаменитостей [35]. На первый взгляд, может показаться довольно странным использовать предобученную модель на данных совершенно другой природы. Однако важнее не доменная близость датасетов, на которых обучаются модели, а скорее разнообразие данных в датасете, на котором обучается модель, которая будет впоследствии дообучаться [38]. Так, на Рисунок 3.2 видно, что спустя всего 40 итераций модель сумела перейти от изображений лиц людей к эмбрионам. Всего данная модель обучалась на 640 kimg^3 и ей удалось достичь $\text{FID}=34$. Более подробно результаты представлены в Приложение А. Результаты экспериментов на предобученной модели StyleGAN2-Ada.

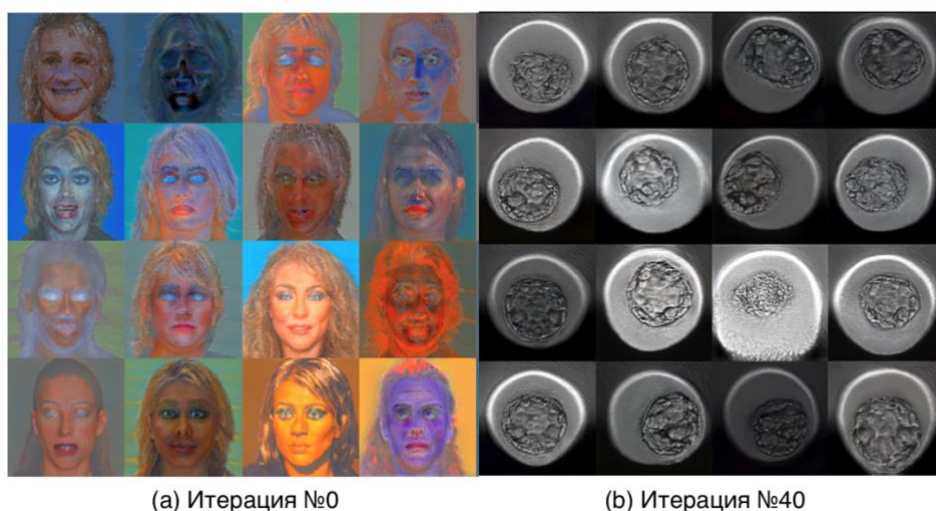


Рисунок 3.2 Результат генерации модели StyleGAN2-Ada, предобученной на датасете CelebA-HQ

³ Количество изображений в тысячах, которое было пропущено через нейросеть

Эксперимент №2: у модели из предыдущего эксперимента были заморожены высокоуровневые слои («*high-resolution layers*») дискриминатора (Freeze-D=10). Данная практика является хорошо себя зарекомендовавшей для тонкой настройки («*fine-tuning*») предобученных моделей [57]. Однако на практике данная модель показала качество хуже - при более чем 1000 kimg ей удалось достичь лишь FID=46. При этом уже после 130 kimg значение FID стало ухудшаться, постепенно стабилизируясь, что свидетельствует о том, что, скорее всего, дальнейшее обучение модели не даст результатов лучше тех, что были достигнуты на более раннем этапе обучения.

3.3 Генерация изображений с помощью StyleGAN2-Ada

Как было указано ранее, предобученную модель нельзя применять в задачах условной генерации, а потому такая модель не может быть использована в качестве конечного решения поставленной задачи, однако, результаты предыдущих экспериментов можно использовать в качестве сравнения, чтобы оценить, насколько хорошо удалось обучить модель с нуля на имеющихся данных. По этой причине все последующие эксперименты проводились с использованием StyleGAN2-Ada, которая обучалась с нуля. Примеры результатов и графики обучения всех экспериментов из данного раздела представлены в Приложение Б. Результаты экспериментов на модели StyleGAN2-Ada.

Эксперимент №3: для гиперпараметров архитектуры была выбрана конфигурация «*paper256*» и разрешены аугментации «*x-flips*» (симметричное отражение по горизонтали). Модель обучалась 8600 kimg и ей удалось достичь FID=20.

Эксперимент №4: для гиперпараметров архитектуры была выбрана конфигурация «*auto*» и разрешены аугментации «*x-flips*» (симметричное отражение по горизонтали). Модель обучалась 1100 kimg и ей удалось достичь FID=32. Так как было замечено, что модель при данной конфигурации показывала результаты хуже предыдущей модели, то было принято решение остановить обучение.

Эксперимент №5: в отличие от предыдущей модели, в качестве конфигурации для гиперпараметров архитектуры была выбрана «*stylegan2*». Модель обучалась 3740 kimg и ей удалось достичь FID=30. Так же, как и в случае с предыдущей моделью, было замечено, что в динамике по метрике FID она уступает модели из эксперимента №3, поэтому дальнейшее обучение было остановлено.

Эксперимент №6: по результатам предыдущих экспериментов было замечено, что генерируемые изображения содержат достаточно странные искажения, которые не присутствуют в обучающей выборке. На Рисунке 3.3 они присутствуют в верхнем правом и нижнем левом углах.

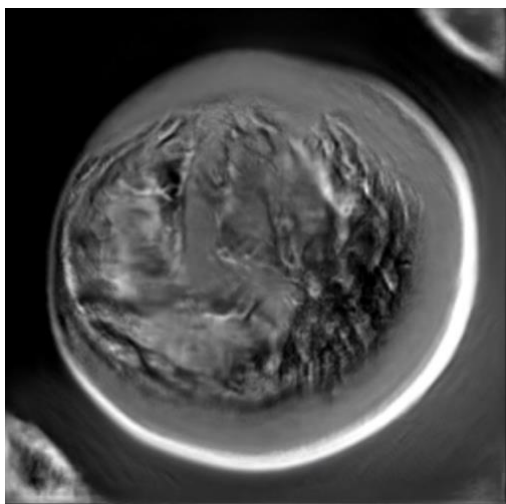


Рисунок 3.3 Пример сгенерированного изображения, содержащего аномальные искажения

Причиной таких искажений являются аугментации, которые, вопреки утверждению авторов архитектуры StyleGAN2-Ada, попадают в выходные изображения. Поэтому в исходный код данной архитектуры были внесены доработки с целью выбрать только те трансформации, попадание которых в генерируемые изображения будет допустимым в рамках решаемой задачи. В данном эксперименте к таким трансформациям были отнесены следующие: $xflip=1$, $rotate90=1$, $brightness=1$, $contrast=1$, $lumaflip=1$, $hue=1$, $saturation=1$, $imgfilter=1$, $noise=1$, $cutout=1$. После этого выявленные аномалии исчезли, однако полученные изображения иногда получались слишком нечеткими (Рисунок 3.4), что также является неприемлемым в данной задаче, поэтому в последующих экспериментах использовались другие трансформации. В качестве конфигурации для гиперпараметров архитектуры была выбрана «auto». Модель обучалась 12500 kimg и ей удалось достичь FID=25.

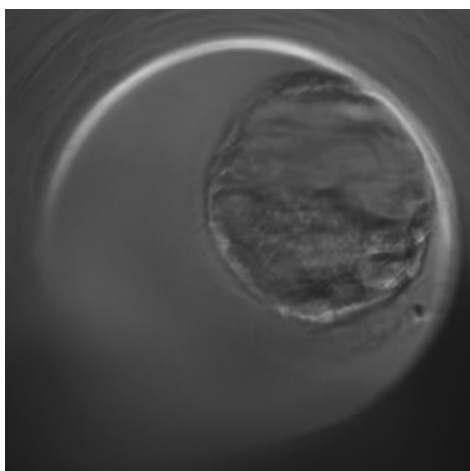


Рисунок 3.4 Пример сгенерированного изображения, содержащего нежелательные искажения

Эксперимент №7-8: в данных экспериментах в качестве трансформаций использовались только $xflip=1$, $rotate90=1$, $rotate=1$, то есть исключительно повороты и допустимые отражения. В качестве конфигурации для гиперпараметров архитектуры была выбрана «*stylegan2*». При значении параметра $gamma=2$ удалось достичь FID=390, при значении параметра $gamma=0.01$ удалось достичь FID=111. Таким образом, было принято решение не варьировать параметр $gamma$.

Эксперимент №9: модель из предыдущего эксперимента с единственным отличием в конфигурации гиперпараметров архитектуры - вместо «*stylegan2*» была выбрана «*paper256*». Модель обучалась 10700 kimg и ей удалось достичь FID=23.

Эксперимент №10: было замечено, что несмотря на то, что введенные аугментации полностью защищают сгенерированные изображения от попадания нежелательных аугментаций, модель очень быстро начинает стабилизироваться и не обучается дальше. В связи с этим было принято решение разрешить различные аугментации, но обучать модель как можно дольше, в предположении, что она сама сможет избавиться от попадания нежелательных аугментаций на более поздних эпохах обучения. По этой причине в данном эксперименте в качестве доступных трансформаций использовались значения по умолчанию - $xflip=1$, $rotate90=1$, $xint=1$, $scale=1$, $rotate=1$, $aniso=1$, $xfrac=1$, $brightness=1$, $contrast=1$, $lumaflip=1$, $hue=1$, $saturation=1$. В качестве конфигурации для гиперпараметров архитектуры была выбрана «*stylegan2*». Модель обучалась 13000 kimg и ей удалось достичь FID=18.

3.4 Обсуждение результатов и дальнейшие шаги

Результаты проведенных экспериментов отображены в Таблице 1. Если сравнивать с предобученными моделями, то можно сделать вывод, что, в целом, удалось получить сопоставимые результаты: эксперимент №1 и эксперимент №4 имеют близкое значение метрики FID. При этом замораживание слоев дискриминатора (Freeze-D) дало негативный эффект, что может указывать на то, что природа данных слишком сильно отличается и требуется обучение большего количества слоев.

Варьирование гиперпараметра «gamma» достаточно сильно влияет на результат, так при высоком его значении сеть очень быстро стабилизируется, достигая высокого значения по метрике FID и дальше не обучается (Эксперимент №7) [57]. При низком значении данного гиперпараметра сеть обучается лучше, однако тоже через некоторое время стабилизируется и перестает обучаться дальше (Эксперимент №8). Разумеется, самым правильным подходом было перебрать значения данного гиперпараметра в более широком диапазоне и с меньшим шагом, однако для этого требуется достаточно много вычислительных ресурсов, поэтому данный перебор не был осуществлен.

Базовая конфигурация архитектуры, гиперпараметра «config», напротив, не так сильно влияет на конечный результат. По результатам экспериментов №3-5 было выявлено, что наилучшей конфигурацией является «stylegan2» - именно поэтому это значение использовалось для последующих экспериментов.

Интересным наблюдением является то, что, вопреки утверждениям авторов архитектуры StyleGAN2-Ada, подкрепленными их экспериментами, на практике аугментации иногда могут попадать в генерируемые изображения (Рисунок 3.3). По этой причине в эксперименте №6 мною была модифицирована конфигурация аугментаций - отобраны только допустимые аугментации, исключающие геометрические преобразования («custom*» в Таблице 1). Однако, и этого оказалось недостаточно, так как некоторые из аугментаций, попадающих в генерируемые изображения, проявились сильнее и негативно повлияли на качество изображений. Поэтому в экспериментах №7-9 были разрешены только повороты и отражения («custom**» в Таблице 1). Это позволило полностью избавиться от попадания нежелательных аугментаций, однако, негативно сказалось на обучении модели - сходимость по FID наступала очень быстро и дальнейшее улучшение данной метрики было очень медленным. Причиной данного наблюдения может быть то, что модели не хватало вариативности исходного набора данных, чтобы обучиться на различные детали. Поэтому в эксперименте №10 конфигурация аугментаций была возвращена в дефолтное значение, в предположении, что при более длительном обучении, модель сама

сможет избавиться от аномальных аугментаций, попадающих в генерируемые изображения. Данное предположение оказалось верным, и получилось добиться наилучшего значения по метрике FID. Поэтому модель, обученная в эксперименте №10, была использована для генерации изображений с целью их последующей оценки индустриальным экспертом.

№	conditional	config	pretrained	king	aug	gamma	Freeze-D	FID
1	false	auto	celebahq256	640	bgc	auto	-	34
2	false	auto	celebahq256	1000	bgc	auto	10	46
3	true	paper256	-	8600	bgc	auto	-	20
4	true	auto	-	1100	bgc	auto	-	32
5	true	stylegan2	-	3740	bgc	auto	-	30
6	true	auto	-	12500	custom*	auto	-	25
7	true	stylegan2	-	237	custom**	2	-	390
8	true	stylegan2	-	2600	custom**	0.01	-	111
9	true	paper256	-	10700	custom**	auto	-	23
10	true	stylegan2	-	13000	bgc	auto	-	18

Таблица 1. Результаты и параметры проведенных экспериментов.

custom*: xflip=1, rotate90=1, brightness=1, contrast=1, lumaflip=1, hue=1, saturation=1, imgfilter=1, noise=1, cutout=1
 custom**: xflip=1, rotate90=1, rotate=1

На момент написания данной работы не опубликовано исследований, где с помощью генеративных моделей создавались бы изображения эмбрионов на стадии бластоцисты. Единственной работой, близкой по своему содержанию, является статья «*HEMIGEN: Human Embryo Image Generator Based on Generative Adversarial Networks*» [59]. В ней авторы также используют архитектуру GAN для генерации изображений эмбрионов, но на стадии дробления, а не бластоцисты. На Рисунок 3.5 наглядно видно, что по своей структуре эмбрион на стадии бластоцисты является куда более сложным, чем на стадии дробления, так как содержит намного больше деталей, многообразие и корректность которых крайне важны при генерации.

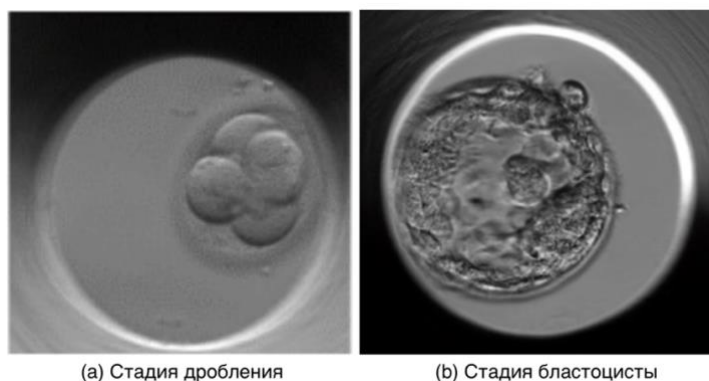


Рисунок 3.5 Пример изображений на различных стадиях развития эмбриона

Более того, авторы данной статьи не используют для оценки качества сгенерированных изображений ни метрику IS, ни метрику FID, а вместо этого опираются на оценку разметчиков и сравнение распределений изображений, что не является распространенным подходом при оценке GAN. В связи с этим не представляется возможным ни качественно, ни количественно провести сравнение с данной работой.

Кроме полученного значения метрики FID для лучшей модели отобранной по результатам экспериментов, также было получено заключение индустриального эксперта - эмбриолога ГУЗ КПЦСО Холостиной Алины Михайловны. Для оценки было сгенерировано 100 изображений каждого класса (good, average, poor), каждое из которых размечалось на качество, а также соответствие категории (Таблица 2). Более подробные результаты и итоговая оценка качества сгенерированных изображений эмбрионов представлены в Приложение В. Заключение индустриального эксперта.

Категория \ Качество	Идеальные	С некоторыми недочетами	Неправдоподобные	Не соответствуют категории
good	48	52	3	10
average	28	64	8	5
poor	31	62	7	7

Таблица 2 Оценка сгенерированных изображений индустриальным экспертом

В качестве будущих шагов планируется осуществить перебор гиперпараметров модели по более широкой сетке значений. С большой вероятностью, это позволит улучшить качество генерируемых изображений по метрике FID. Кроме того, как было отмечено ранее, при формировании обучающей выборки использовались изображения, полученные с помощью time-lapse системы, а потому некоторые изображения могут минимально отличаться друг от друга, в силу того, что временной интервал между моментами съемки достаточно мал. Таким образом, заменив очень похожие изображения, можно повысить вариативность обучающего набора данных, и, как следствие, генерируемых изображений. Кроме того, следует отметить, что при текущем разбиении на три класса достаточно сложно напрямую использовать полученные результаты в практических целях. Поэтому необходимо использовать разбиение на классы соответствующее шкале Гарднера. Также, в последнее время генеративные модели, построенные на архитектуре GAN, уступают лидирующие позиции диффузионным моделям, поэтому проведение экспериментов с ними может дать интересные результаты.

Заключение

Целью данной работы было применение методов генеративных состязательных сетей для генерации изображений эмбрионов заданного качества на стадии бластоцисты. Для достижения этой цели был осуществлен сбор и разметка данных, проведены сравнительные эксперименты для поиска наилучшей модели. Используя архитектуру StyleGAN2-Ada, удалось успешно сгенерировать изображения бластоцист. Качество полученных результатов было оценено с использованием метрики FID, а правдоподобность изображений и соответствие эмбрионов заданной категории качества были подтверждены индустриальным экспертом. Таким образом, результаты данной работы представляют первую описанную успешную генерацию изображений эмбрионов человека на стадии бластоцисты. Полученные результаты могут в будущем быть использованы в качестве обучающей выборки для решения других задач компьютерного зрения, связанных с анализом человеческих эмбрионов.

Благодарности

Автор выражает благодарность своему научному руководителю Айбеку Аланову за практические советы и свободу в выборе направления исследования; Алине Холостиной за помощь в разметке данных и оценке полученных результатов; Дарье Свистуновой за предоставленную рецензию; своей семье и близким за поддержку и терпение. Исследование выполнено с использованием суперкомпьютерного комплекса НИУ ВШЭ.

Список литературы

1. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative Adversarial Networks. arXiv; <http://arxiv.org/abs/1406.2661>, 2014.
2. Baddar WJ, Gu G, Lee S, Ro YM. Dynamics Transfer GAN: Generating Video by Transferring Arbitrary Temporal Dynamics from a Source Video to a Single Target Image. arXiv; <http://arxiv.org/abs/1712.03534>, 2017.
3. Yi X, Walia E, Babyn P. Generative Adversarial Network in Medical Imaging: A Review. *Med Image Anal.* 2019 Dec;58:101552
4. Testart J. Cleavage stage of human embryos two days after fertilization in vitro and their developmental ability after transfer into the uterus. *Hum Reprod.* 1986 Jan;1(1):29–31
5. Hamori M, Stuckensen JA, Rumpf D, Kniewald T, Kniewald A, Marquez MA. Zygote intrafallopian transfer (ZIFT): evaluation of 42 cases. *Fertil Steril.* 1988 Sep;50(3):519–21
6. Mastroyannis C. Gamete intrafallopian transfer: ethical considerations, historical development of the procedure, and comparison with other advanced reproductive technologies. *Fertil Steril.* 1993 Sep;60(3):389–402
7. Gardner DK, Schoolcraft WB. Culture and transfer of human blastocysts: *Curr Opin Obstet Gynaecol.* 1999 Jun;11(3):307–11
8. Tao J, Tamis R, Fink K, Williams B, Nelson-White T, Craig R. The neglected morula/compact stage embryo transfer. *Hum Reprod.* 2002 Jun;17(6):1513–8
9. Wamaitha SE, Niakan KK. Human Pre-gastrulation Development. In: *Current Topics in Developmental Biology.* Elsevier; 2018. p. 295–338
10. Arsalan M, Haider A, Choi J, Park KR. Detecting Blastocyst Components by Artificial Intelligence for Human Embryological Analysis to Improve Success Rate of In Vitro Fertilization. *J Pers Med.* 2022 Jan 18;12(2):124
11. Gardner DK, Schoolcraft WB, Jansen R, Mortimer D. In vitro culture of human blastocysts. *Toward Reproductive Certainty: Fertility and Genetics Beyond.* 1999;378–88
12. Zhao YY, Yu Y, Zhang XW. Overall Blastocyst Quality, Trophectoderm Grade, and Inner Cell Mass Grade Predict Pregnancy Outcome in Euploid Blastocyst Transfer Cycles. *Chin Med J (Engl).* 2018 Jun 5;131(11):1261–7
13. Berntsen J, Rimestad J, Lassen JT, Tran D, Kragh MF. Robust and generalizable embryo selection based on artificial intelligence and time-lapse image sequences. Nogueira MFG, editor. *PLOS ONE.* 2022 Feb 2;17(2):e0262661
14. Barnes J, Brendel M, Gao VR, Rajendran S, Kim J, Li Q, et al. A non-invasive artificial intelligence approach for the prediction of human blastocyst ploidy: a retrospective model development and validation study. *Lancet Digit Health.* 2023 Jan;5(1):e28–40
15. Patel D, Shah P, Kotdawala A, Herrero J, Rubio I, Banker M. Morphokinetic behavior of euploid and aneuploid embryos analyzed by time-lapse in embryoscope. *J Hum Reprod Sci.* 2016;9(2):112
16. Yeh RA, Chen C, Lim TY, Schwing AG, Hasegawa-Johnson M, Do MN. Semantic Image Inpainting with Deep Generative Models. arXiv; <http://arxiv.org/abs/1607.07539>, 2017.
17. Phan H, McLoughlin IV, Pham L, Chén OY, Koch P, De Vos M, et al. Improving GANs for Speech Enhancement. *IEEE Signal Process Lett.* 2020;27:1700–4

18. Tian C, Zhang X, Lin JCW, Zuo W, Zhang Y, Lin CW. Generative Adversarial Networks for Image Super-Resolution: A Survey. arXiv; <http://arxiv.org/abs/2204.13620>, 2022.
19. Wu Y, Wang X, Li Y, Zhang H, Zhao X, Shan Y. Towards Vivid and Diverse Image Colorization with Generative Color Prior. arXiv; <http://arxiv.org/abs/2108.08826>, 2022.
20. Charitou C, Dragicevic S, Garcez A d'Avila. Synthetic Data Generation for Fraud Detection using GANs. arXiv; <http://arxiv.org/abs/2109.12546>, 2021.
21. Mirza M, Osindero S. Conditional Generative Adversarial Nets. arXiv; <http://arxiv.org/abs/1411.1784>, 2014.
22. Weng L. From GAN to WGAN. arXiv; <http://arxiv.org/abs/1904.08994>, 2019.
23. Metz L, Poole B, Pfau D, Sohl-Dickstein J. Unrolled Generative Adversarial Networks. arXiv; <http://arxiv.org/abs/1611.02163>, 2017.
24. Arjovsky M, Bottou L. Towards Principled Methods for Training Generative Adversarial Networks. arXiv; <http://arxiv.org/abs/1701.04862>, 2017.
25. Salimans T, Goodfellow I, Zaremba W, Cheung V, Radford A, Chen X. Improved Techniques for Training GANs. arXiv; <http://arxiv.org/abs/1606.03498>, 2016.
26. Tang S. Lessons Learned from the Training of GANs on Artificial Datasets. arXiv; <http://arxiv.org/abs/2007.06418>, 2020.
27. Radford A, Metz L, Chintala S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. arXiv; <http://arxiv.org/abs/1511.06434>, 2016.
28. Dumoulin V, Visin F. A guide to convolution arithmetic for deep learning. arXiv; <http://arxiv.org/abs/1603.07285>, 2018.
29. Karras T, Laine S, Aila T. A Style-Based Generator Architecture for Generative Adversarial Networks. arXiv; <http://arxiv.org/abs/1812.04948>, 2019.
30. Huang X, Liu MY, Belongie S, Kautz J. Multimodal Unsupervised Image-to-Image Translation. arXiv; <http://arxiv.org/abs/1804.04732>, 2018.
31. Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. arXiv; <http://arxiv.org/abs/1502.03167>, 2015.
32. Ulyanov D, Vedaldi A, Lempitsky V. Improved Texture Networks: Maximizing Quality and Diversity in Feed-forward Stylization and Texture Synthesis. arXiv; <http://arxiv.org/abs/1701.02096>, 2017.
33. Wu Y, He K. Group Normalization. arXiv; <http://arxiv.org/abs/1803.08494>, 2018.
34. Huang X, Belongie S. Arbitrary Style Transfer in Real-time with Adaptive Instance Normalization. arXiv; <http://arxiv.org/abs/1703.06868>, 2017.
35. Karras T, Aila T, Laine S, Lehtinen J. Progressive Growing of GANs for Improved Quality, Stability, and Variation. arXiv; <http://arxiv.org/abs/1710.10196>, 2018.
36. Karras T, Laine S, Aittala M, Hellsten J, Lehtinen J, Aila T. Analyzing and Improving the Image Quality of StyleGAN. arXiv; <http://arxiv.org/abs/1912.04958>, 2020.
37. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. arXiv; <http://arxiv.org/abs/1512.03385>, 2015.
38. Karras T, Aittala M, Hellsten J, Laine S, Lehtinen J, Aila T. Training Generative Adversarial Networks with Limited Data. arXiv; <http://arxiv.org/abs/2006.06676>, 2020.
39. Zhang D, Khoreva A. Progressive Augmentation of GANs. arXiv; <http://arxiv.org/abs/1901.10422>, 2019.

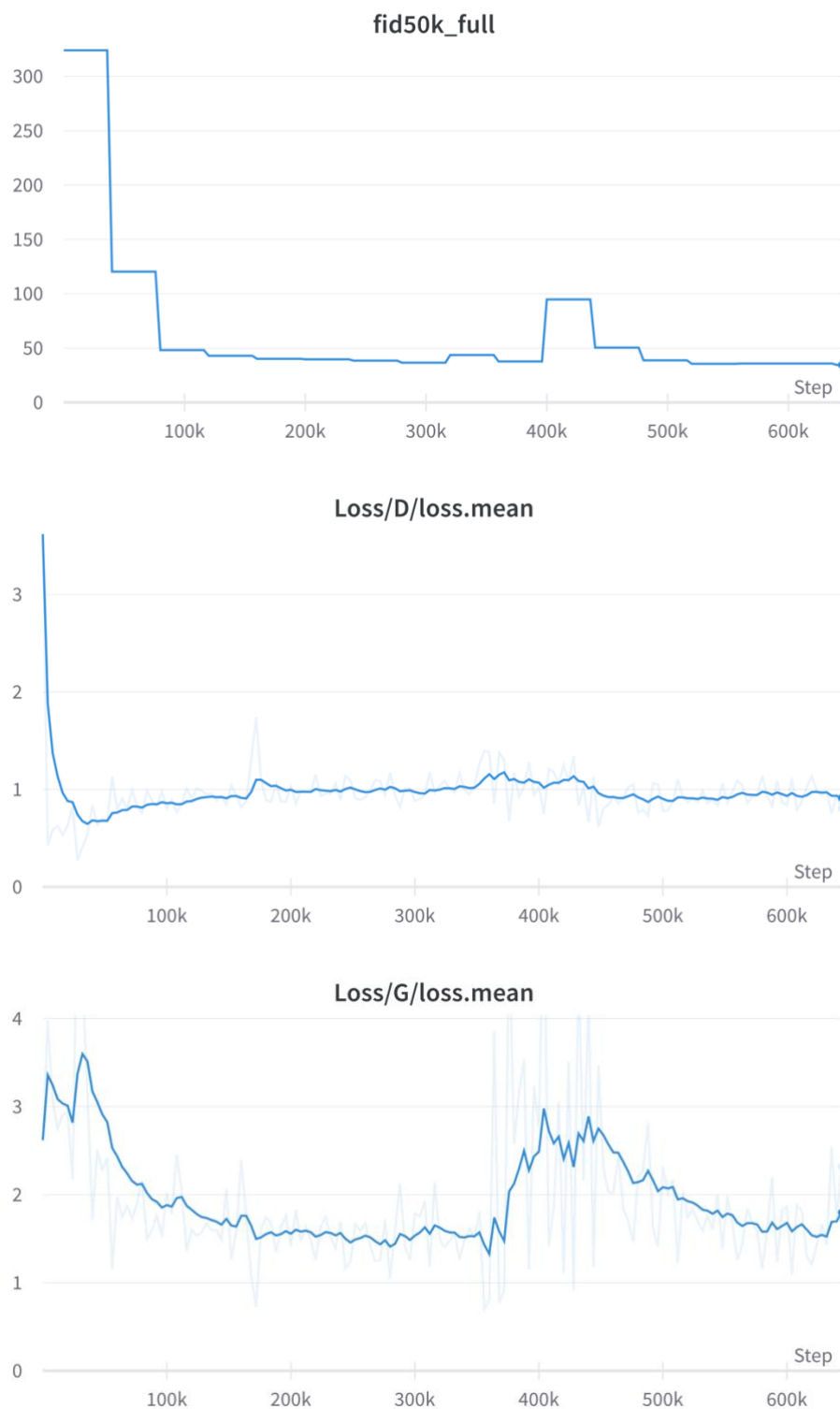
40. Shorten C, Khoshgoftaar TM. A survey on Image Data Augmentation for Deep Learning. *J Big Data*. 2019 Dec;6(1):60
41. Zhang H, Zhang Z, Odena A, Lee H. Consistency Regularization for Generative Adversarial Networks. arXiv; <http://arxiv.org/abs/1910.12027>, 2020.
42. Zhao Z, Singh S, Lee H, Zhang Z, Odena A, Zhang H. Improved Consistency Regularization for GANs. arXiv; <http://arxiv.org/abs/2002.04724>, 2020.
43. Dash S, Shakyawar SK, Sharma M, Kaushik S. Big data in healthcare: management, analysis and future prospects. *J Big Data*. 2019 Dec;6(1):54
44. Персональные данные пациентов в медицинских организациях: требования к обработке и ответственность. <https://www.garantexpress.ru/statji/personalnie-dannie-patsientov-v-meditsinskih-organizatsijah-trebovania-k-obrabotke-i-otvetstvennost/>,
45. Blatter TU, Witte H, Nakas CT, Leichtle AB. Big Data in Laboratory Medicine—FAIR Quality for AI? *Diagnostics*. 2022 Aug 9;12(8):1923
46. Nguyen T, Diakiw S, VerMilyea M, Dinsmore A, Perugini M, Perugini D, et al. Efficient automated error detection in medical data using deep-learning and label-clustering. In Review; <https://www.researchsquare.com/article/rs-2631746/v1>, 2023 Mar
47. Automatic Identification of Human Blastocyst Components via Texture. *IEEE Trans Biomed Eng*. 2017 Dec;64(12):2968–78
48. Gomez T, Feyeux M, Boulant J, Normand N, David L, Paul-Gilloteaux P, et al. A time-lapse embryo dataset for morphokinetic parameter prediction. *Data Brief*. 2022 Jun;42:108258
49. Ai J, Jin L, Zheng Y, Yang P, Huang B, Dong X. The Morphology of Inner Cell Mass Is the Strongest Predictor of Live Birth After a Frozen-Thawed Single Embryo Transfer. *Front Endocrinol*. 2021 Feb 24;12:621221
50. Harrity C, Vaughan D, Emerson G, Mocanu E. Effect of blastocyst stage and grade on successful embryo transfer. *Fertil Steril*. 2013 Sep;100(3):S288
51. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the Inception Architecture for Computer Vision. arXiv; <http://arxiv.org/abs/1512.00567>, 2015.
52. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet Large Scale Visual Recognition Challenge. arXiv; <http://arxiv.org/abs/1409.0575>, 2015.
53. Barratt S, Sharma R. A Note on the Inception Score. arXiv; <http://arxiv.org/abs/1801.01973>, 2018.
54. Heusel M, Ramsauer H, Unterthiner T, Nessler B, Hochreiter S. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. arXiv; <http://arxiv.org/abs/1706.08500>, 2018.
55. Borji A. Pros and Cons of GAN Evaluation Measures. arXiv; <http://arxiv.org/abs/1802.03446>, 2018.
56. Liu S, Wei Y, Lu J, Zhou J. An Improved Evaluation Framework for Generative Adversarial Networks. arXiv; <http://arxiv.org/abs/1803.07474>, 2018.
57. Mo S, Cho M, Shin J. Freeze the Discriminator: a Simple Baseline for Fine-Tuning GANs. arXiv; <http://arxiv.org/abs/2002.10964>, 2020.
58. Mescheder L, Geiger A, Nowozin S. Which Training Methods for GANs do actually Converge? arXiv; <http://arxiv.org/abs/1801.04406>, 2018.

59. Dirvanauskas, Maskeliūnas, Raudonis, Damaševičius, Scherer. HEMIGEN: Human Embryo Image Generator Based on Generative Adversarial Networks. *Sensors*. 2019 Aug 16;19(16):3578
60. Bormann CL, Kanakasabapathy MK, Thirumalaraju P, Gupta R, Pooniwala R, Kandula H, et al. Performance of a deep learning based neural network in the selection of human blastocysts for implantation. *eLife*. 2020 Sep 15;9:e55301
61. Zaninovic N, Khosravi P, Hajirasouliha I, Malmsten JE, Kazemi E, Zhan Q, et al. Assessing human blastocyst quality using artificial intelligence (AI) convolutional neural network (CNN). *Fertil Steril*. 2018 Sep;110(4):e89

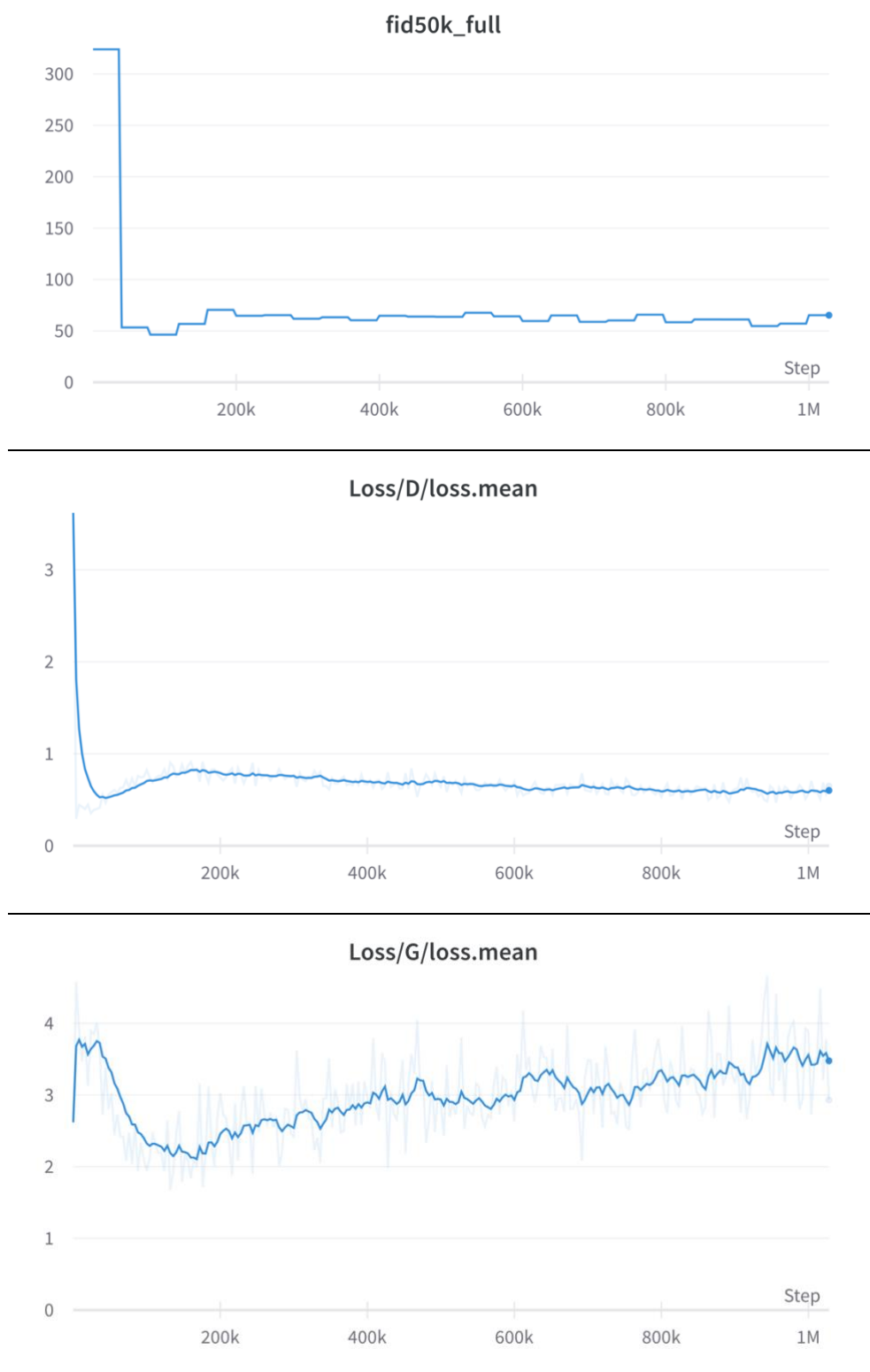
Приложение

Приложение А. Результаты экспериментов на предобученной модели StyleGAN2-Ada

Оценка результатов эксперимента №1:

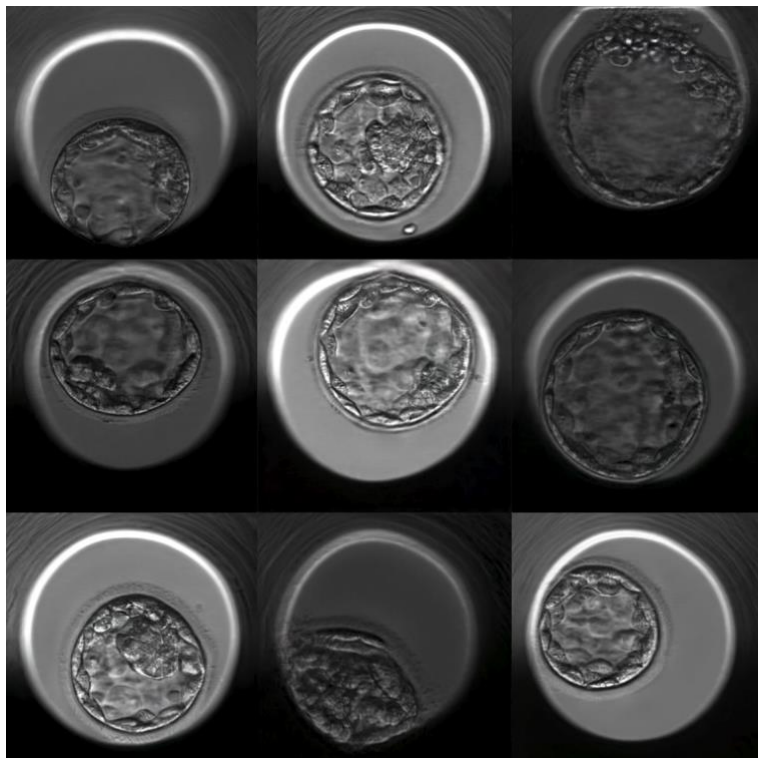


Оценка результатов эксперимента №2:

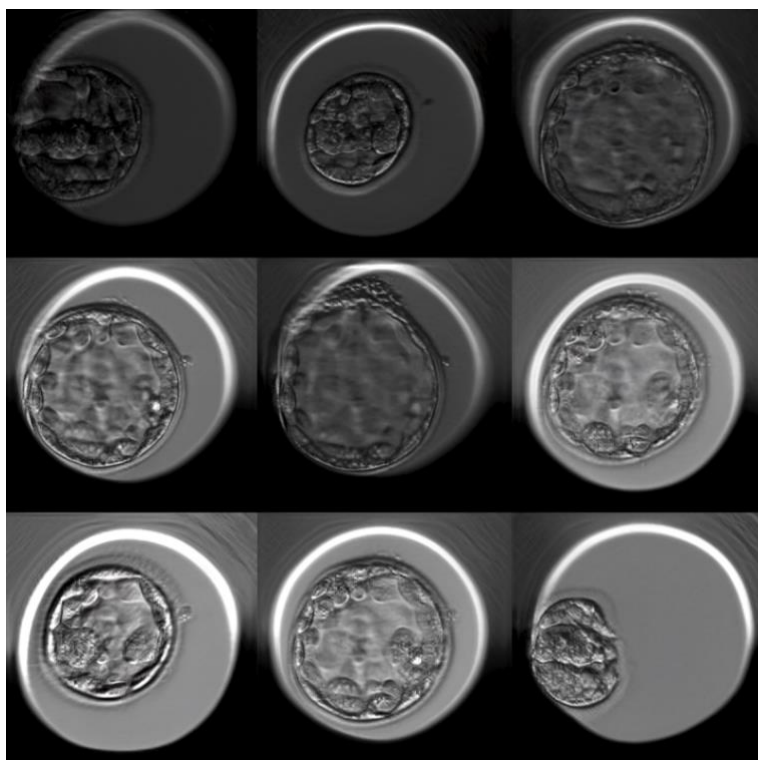


Приложение Б. Результаты экспериментов на модели StyleGAN2-Ada

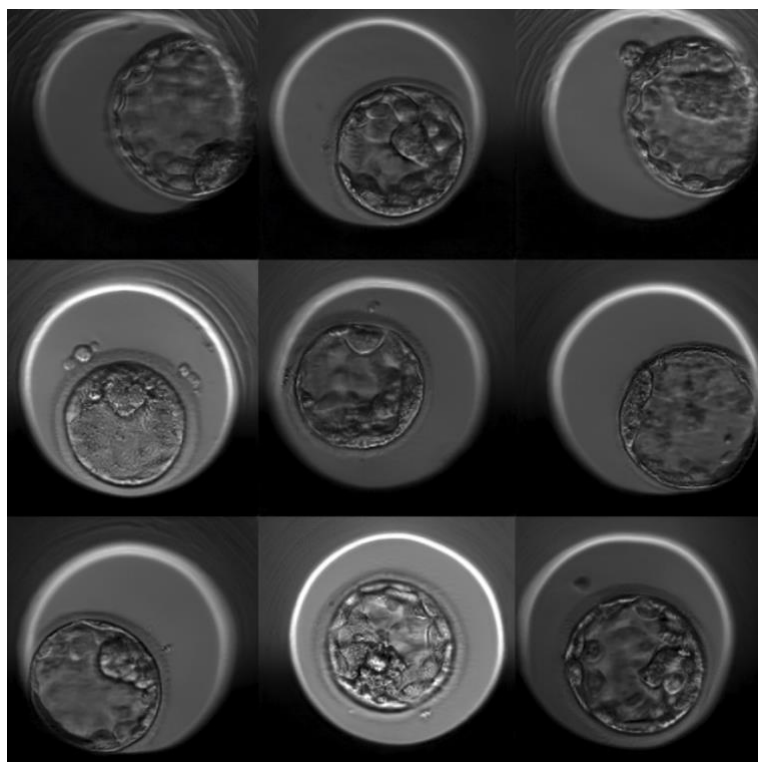
Результаты эксперимента №1:



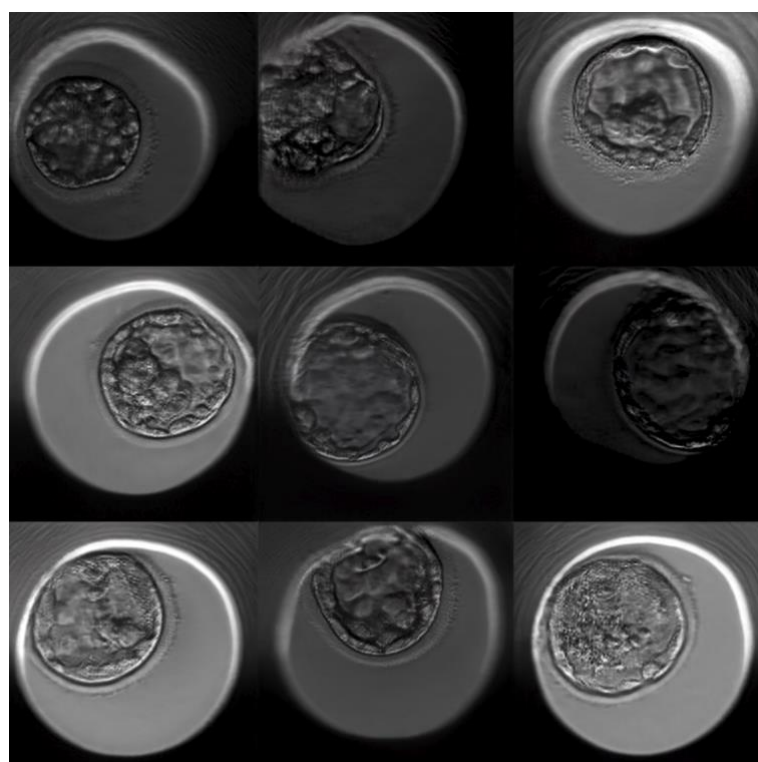
Результаты эксперимента №2:



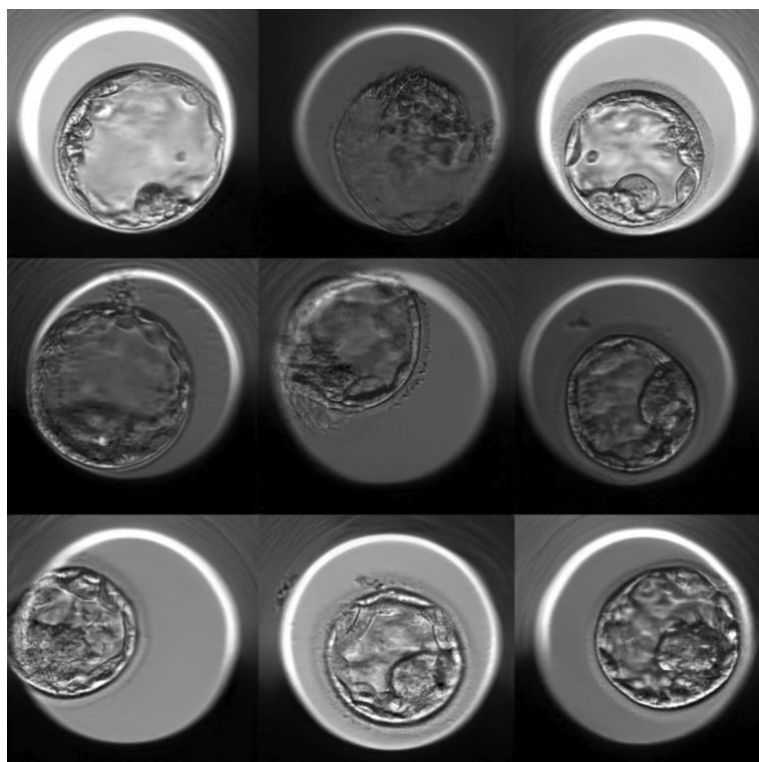
Результаты эксперимента №3:



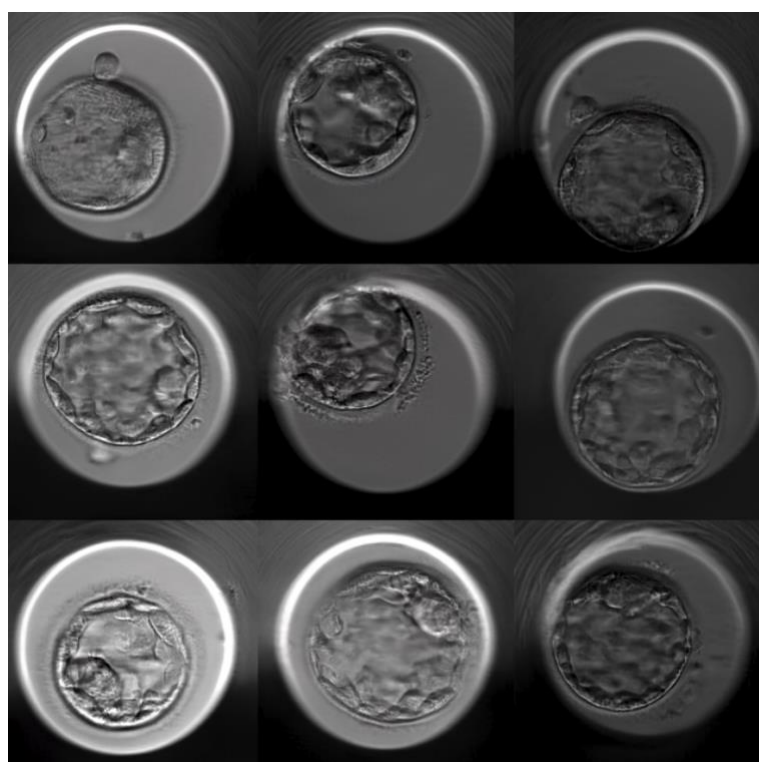
Результаты эксперимента №4:



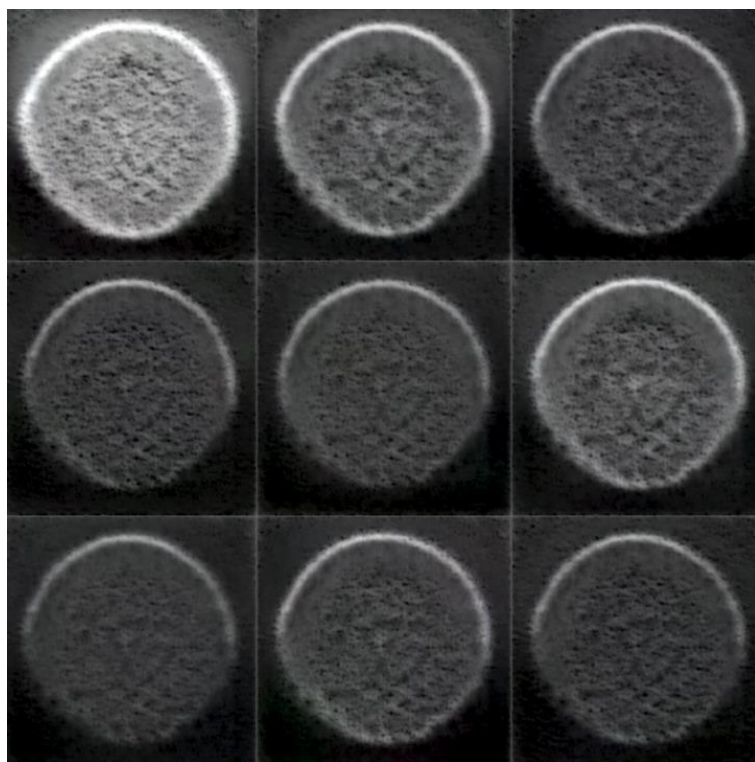
Результаты эксперимента №5:



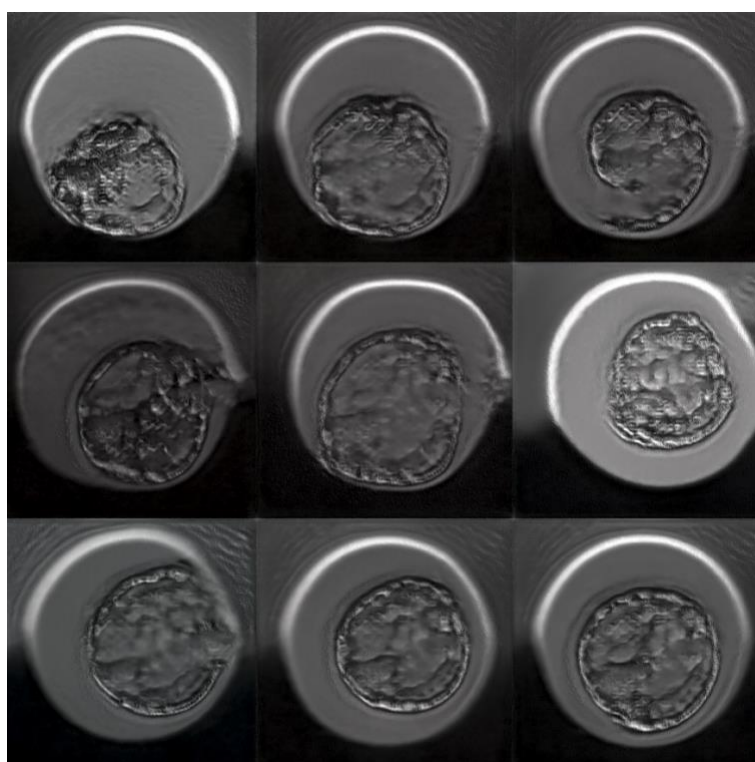
Результаты эксперимента №6:



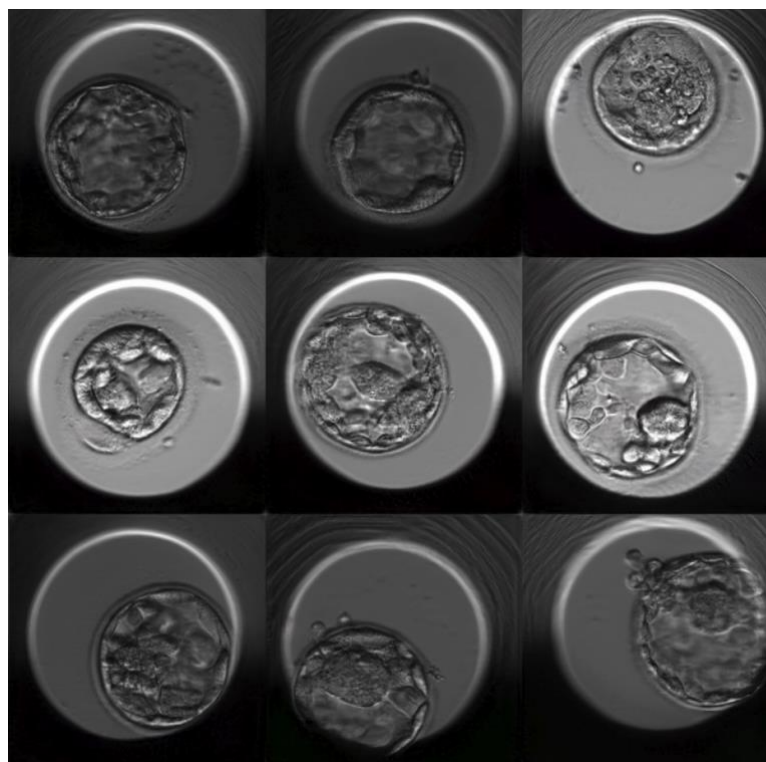
Результаты эксперимента №7:



Результаты эксперимента №8:



Результаты эксперимента №9:



Результаты эксперимента №10:

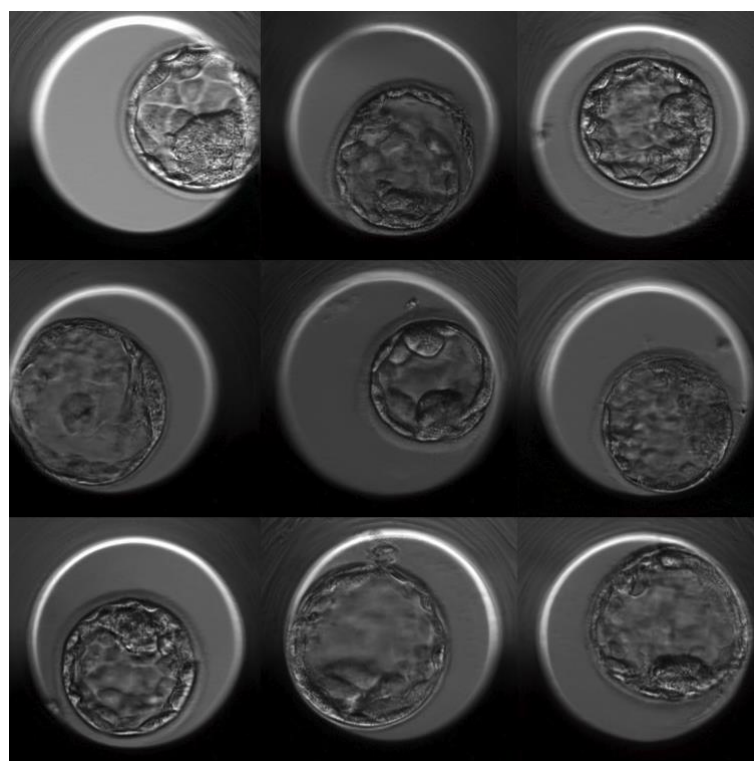


График ошибки генератора:

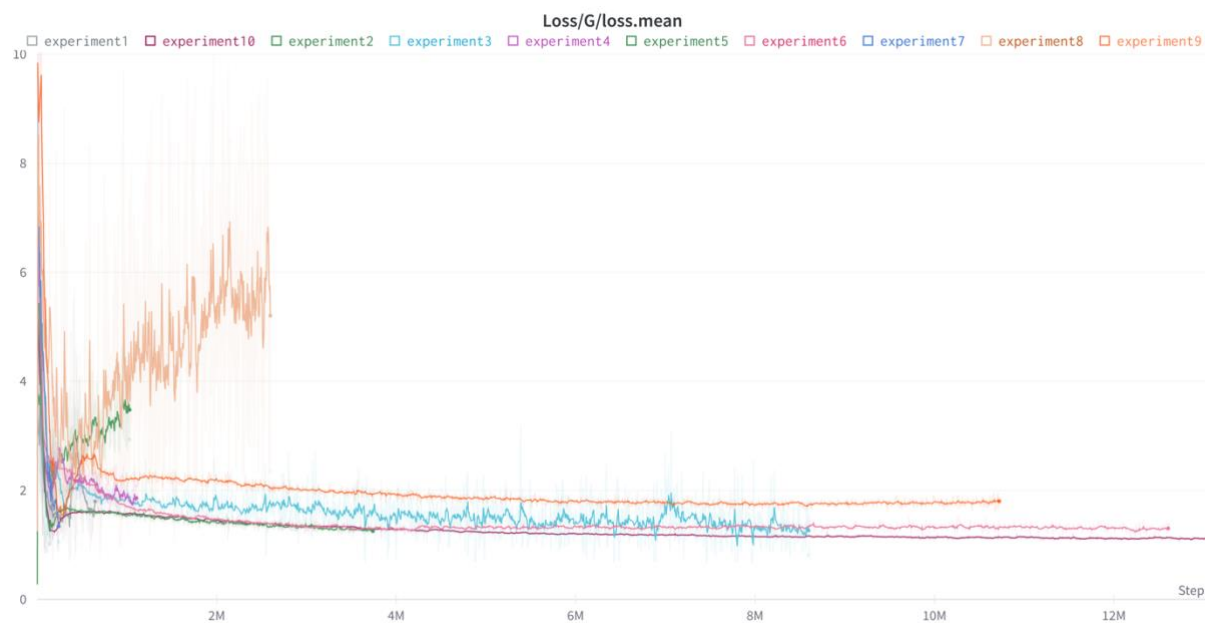
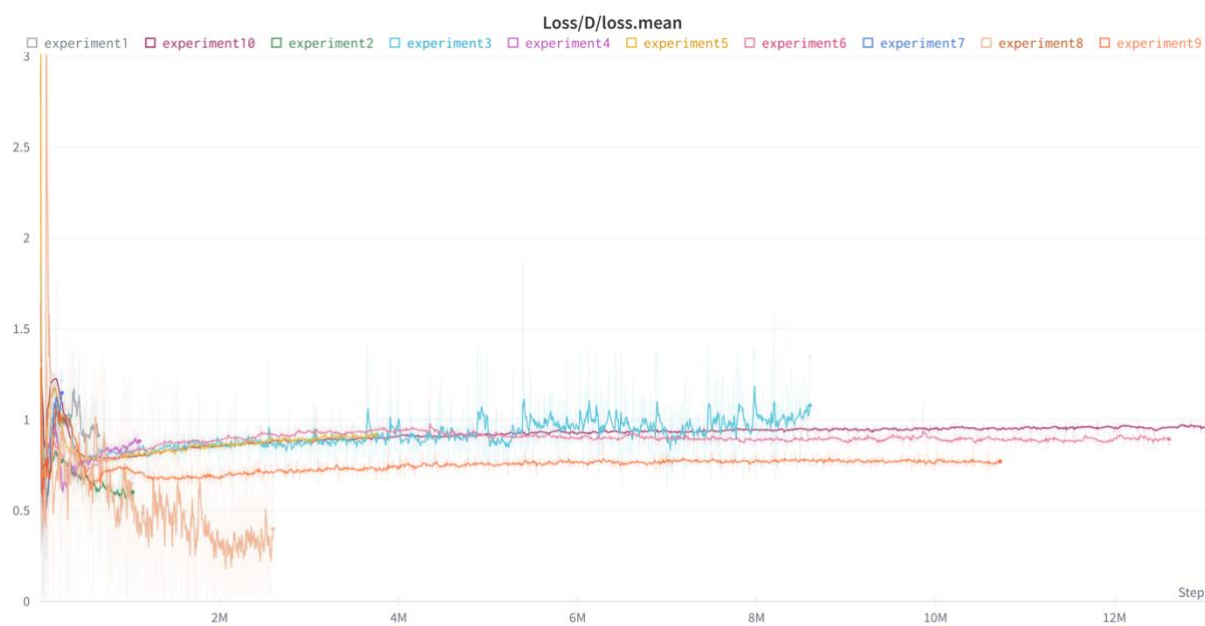


График ошибки дискриминатора:



Приложение В. Заключение индустриального эксперта

Заключение

Для оценки сгенерированных в рамках дипломной работы Долматова К.И. «Генерация изображений бластоцист заданного качества» изображений эмбрионов мне было предоставлено 300 изображений (100 изображений каждого класса).

В категории good из 100 проанализированных мною эмбрионов 45 выглядели неотличимо от изображений настоящих эмбрионов, 52 имели некоторые изъяны (см. подробнее ниже), 3 изображения имели существенные недочеты и проблемы. 10 из этих 100 эмбрионов также не соответствовали заданной категории.

В категории average 28 изображений выглядели неотличимо от настоящих эмбрионов, 64 имели некоторые изъяны и 8 изображений имели существенные недочеты и проблемы. 5 эмбрионов по своим параметрам также не соответствовали категории average.

В категории poor 31 изображение выглядело идеально, 62 изображения имели некоторые неточности, 7 изображений имели существенные изъяны. 7 эмбрионов по своим параметрам не соответствовали категории.

Среди недочетов, обнаруженных у эмбрионов всех категорий, можно выделить:

- неравномерную толщину ZP или ее отсутствие на некоторых участках, слияние с окружающей средой
- слияние границ клеток эмбриона и ZP
- угловатую или яйцеобразную форму эмбриона
- нечеткие границы клеток
- некоторые изображения эмбрионов выглядят как зеркальные отражения вдоль вертикальной оси

Из интересных деталей, которые присутствовали на сгенерированных изображениях, хочется отметить дебрис на дне лунки с эмбрионом и выходящую из отверстия в ZP каплю цитоплазмы, которые встречаются у эмбрионов размера 4 после оплодотворения методом ИКСИ.

Таким образом, Долматову К.И. в своей дипломной работе удалось сгенерировать правдоподобные изображения бластоцист человека, что открывает возможности для дальнейших исследований на стыке ВРТ и ИИ.

Государственное учреждение здравоохранения
Эмбриолог ГУЗ КПЦСО
"Клинический перинатальный центр"
Холостина А.М.
Саратовская область