

Отчёт по проекту: Latent Adversarial Diffusion Distillation

Команда: Максим Игнатов, Егор Леонтенков

21 декабря 2025 г.

Содержание

1 Постановка задачи и описание методологии	2
1.1 Постановка задачи	2
1.2 Общая идея метода LADD	2
1.3 Оптимизируемый функционал	2
1.4 Процедура обучения и генерации	3
2 Техническое описание экспериментов	3
2.1 Датасеты и подготовка	3
2.1.1 Генерация синтетического датасета	3
2.1.2 Используемые датасеты	3
2.1.3 Примеры синтетического датасета	4
2.2 Архитектура модели	4
2.3 Дискриминатор	5
2.4 Гиперпараметры	5
2.5 Число итераций и время обучения	5
3 Описание экспериментов	5
3.1 Эксперимент 1	5
3.2 Эксперимент 2	6
3.3 Эксперимент 3	8
3.4 Эксперимент 4	8
3.5 Эксперимент 5	9
3.6 Эксперимент 6	10
4 Выводы	11
5 Распределение задач между членами команды	12

1 Постановка задачи и описание методологии

1.1 Постановка задачи

Диффузионные модели демонстрируют высокое качество генерации изображений, но на практике их применение ограничено высокой вычислительной стоимостью ГПУ, из-за большого числа шагов денойзинга. Целью данного проекта является ускорение процесса генерации изображений без существенной потери качества. Для этого рассматривается метод *Latent Adversarial Diffusion Distillation (LADD)*, направленный на дистилляцию диффузионной модели в генеративную модель, способную осуществлять генерацию за несколько шагов.

Формально, мы хотим аппроксимировать распределение изображений $p_{\text{data}}(x)$ при помощи генеративной модели, минимизируя вычислительные затраты на этапе инференса.

1.2 Общая идея метода LADD

Метод LADD основан на переносе процесса диффузионной генерации в латентное пространство и использовании состязательного обучения для дистилляции диффузионной модели-учителя в модель-студента. Вместо моделирования стохастического процесса удаления шума модель-студент обучается за несколько шагов воспроизводить распределение латентных представлений.

Ключевая идея заключается в том, что вместо дорогостоящего декодирования полученных латентов через VAE и получения adversarial лосса через большой дискриминатор, метод позволяет производить обучение полностью в латентном пространстве, используя внутренние представления модели-учителя как фичи для легковесных голов-дискриминаторов.

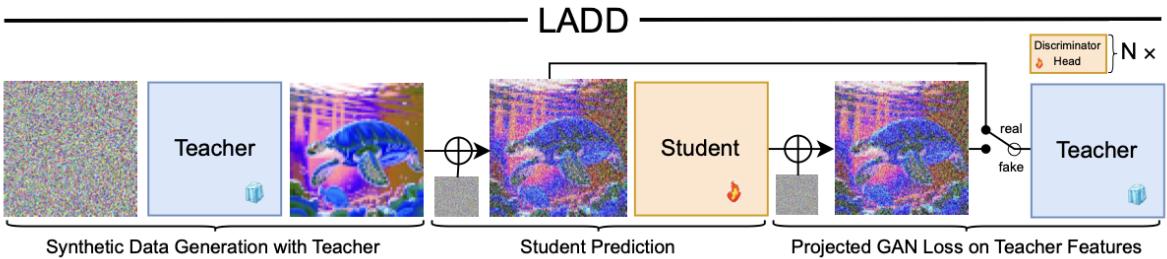


Рис. 1: Методология LADD.

1.3 Оптимизируемый функционал

Для обучения ученика используется состязательное обучение. Вводится дискриминатор D_ϕ , задача которого — отличать латентные представления, полученные от диффузионной модели-учителя, от латентов, сгенерированных студентом.

Состязательная функция потерь имеет вид:

$$\mathcal{L}_{\text{adv}} = \mathbb{E}_{z \sim p_T(z)} [\log D_\phi(z)] + \mathbb{E}_{\hat{z} \sim p_\theta(z)} [\log (1 - D_\phi(\hat{z}))].$$

Также используется функция потерь дистилляции:

$$\mathcal{L}_{\text{distil}} = \mathbb{E}_{z \sim p_T(z)} [\|\hat{z} - z\|_2^2]$$

Обучение формулируется как задача минимакса:

$$\min_{\theta} \max_{\phi} \mathcal{L}_{\text{adv}} + w \cdot \mathcal{L}_{\text{distil}}.$$

1.4 Процедура обучения и генерации

Процедура обучения LADD включает следующие шаги:

1. Оффлайн генерацию высококачественных семплов моделью-учителем (чистых латентов или картинок)
2. Расшумление частично зашумленного изображения с помощью модели ученика
3. подсчет distillation лосса между ground truth латентами, полученными при оффлайн генерации и латентами полученными от ученика
4. подсчет adversarial лосса между частично зашумленными латентами из ученика и учителя.

2 Техническое описание экспериментов

2.1 Датасеты и подготовка

В проекте использовались синтетически сгенерированные данные (как в статье).

В качестве модели-учителя была выбрана диффузионная модель, предварительно обученная на датасете **Color MNIST** из библиотеки EDM на основе CUNet. С помощью данной модели выполнялась генерация изображений, которые далее использовались в качестве обучающих данных для модели LADD.

2.1.1 Генерация синтетического датасета

Генерация данных осуществлялась с использованием диффузионного процесса с фиксированным числом шагов. Основные параметры генерации приведены ниже:

- Число шагов диффузии: 50
- Минимальный уровень шума: $\sigma_{\min} = 0.002$
- Максимальный уровень шума: $\sigma_{\max} = 80.0$
- Параметр шума: $\rho = 7.0$
- Classifier-Free Guidance: 4

Сгенерированные данные сохранялись в виде тензоров, включающих латентные представления и дополнительные параметры генерации (шум, условие).

2.1.2 Используемые датасеты

В ходе экспериментов для обучения модели использовались три синтетических датасета различного размера:

- **Dataset-10k**: 10 000 синтетических изображений, сгенерированных моделью-учителем; хранились только латентные представления и условная информация.
- **Dataset-50k**: 50 000 синтетических изображений, аналогичной структуры.

- **Dataset-20k**: 20 000 изображений, для которых сохранялись латентные представления, начальный шум и условная информация. Данный датасет использовался для более точного воспроизведения распределения диффузионной модели-учителя.

Для первых двух датасетов значения шума не сохранялись, а в третьем датасете хранение шума позволило расширить возможности обучения модели.

2.1.3 Примеры синтетического датасета

На рис. 2 приведены примеры изображений, сгенерированных моделью-учителем и использованных для обучения.



Рис. 2: Примеры изображений из синтетического датасета, сгенерированного диффузионной моделью-учителем, обученной на Color MNIST.

2.2 Архитектура модели

В качестве модели-учителя и модели-студента использовалась условная UNet подобная архитектура **CUNet**. Архитектура работает в рамках EDM, где выход сети масштабируется с использованием коэффициентов, зависящих от уровня шума σ .

Модель-студент инициализируется как точная копия модели-учителя и далее обучается в режиме дистилляции.

2.3 Дискриминатор

Для реализации состязательной дистилляции используется набор дискриминаторов, применяемых к промежуточным признакам модели-учителя. Для каждого выбранного уровня признаков используется отдельная дискриминаторная голова (*DiscHead*).

2.4 Гиперпараметры

Для обучения модели-студента использовались следующие параметры:

- Тип предсказания: DDPM;
- Число шагов диффузии: 50;
- Число шагов обучения: 20 000;
- Число шагов генерации студента: 4;
- Минимальный уровень шума: $\sigma_{\min} = 0.002$;
- Максимальный уровень шума: $\sigma_{\max} = 80.0$;
- Параметр шума: $\rho = 7.0$;
- Скорость обучения студента: $lr_{\text{student}} = 1 \times 10^{-4}$;
- Скорость обучения дискриминатора: $lr_{\text{discriminator}} = 5 \times 10^{-5}$;
- Вес состязательной функции потерь: 0.1;
- Вес функции потерь дистилляции: 1.0.

2.5 Число итераций и время обучения

Обучение проводилось в течение нескольких тысяч итераций в зависимости от размера используемого датасета. Результаты сохранялись каждые 500 шагов обучения и включали латенты учителя, латенты студента, начальный шум и условия генерации.

3 Описание экспериментов

3.1 Эксперимент 1

Цель: проверить корректность работы полного пайплайна и валидность данных на небольшом датасете.

Дискриминатор: Дискриминатор представляет собой сверточную нейронную сеть, состоящую из нескольких блоков: свертки 1×1 , LeakyReLU, AdaptiveAvgPool2d и полносвязных слоёв.

Сетап:

- Датасет: 10 000 синтетических примеров.
- Особенности: использовался маленький дискриминатор, дискриминатор применялся к признакам, которые выбирались на начальных этапах forward-метода CUNet.
- Количество голов дискриминатора: 5.

Генерация учителя	Генерация ученика
1 1 7 3 6 1 4 9	5 1 1 7 3 3 6 1 4 9
4 4 9 9 5 1 7 8 8 0	4 4 9 9 5 1 7 8 8 0
7 3 8 7 5 5 3 4 5 9	7 3 8 7 5 5 3 4 5 9
1 8 9 1 9 1 1 3	4 8 9 6 1 8 5 1 1 3
0 5 9 9 2 4 1 3 0 4	0 5 9 9 2 4 1 3 0 4
0 7 4 0 7 9 1 3 9	0 7 4 0 2 7 5 1 3 9
5 6 1 3 3 3 2 0 1	5 6 6 1 3 3 3 2 0 1
3 7 2 3 2 3 7 1 4 8	8 7 2 3 7 3 7 8 4 8
9 8 1 4 4 4 8 8 4	9 8 6 4 4 4 8 8 2 4
7 7 4 0 5 1 6 3 2 4	7 7 4 2 5 5 5 8 2 4
7 4 7 0 6 1 4 4	7 4 7 8 6 0 6 4 4 4
2 7 9 8 4 0 0 7 7 4	2 7 9 8 4 0 0 7 7 4
4 2 1 1 9 7 1	2 9 0 4 9 9 7 7

Рис. 3: Сравнение учитель / ученик для эксперимента 1.

Наблюдения:

- На генерациях наблюдался **mode collapse**, из-за чего все сгенерированные изображения схлопывались в однотипный цвет.

3.2 Эксперимент 2

Цель: постараться исправить баги предыдущего эксперимента и увеличить датасет для обучения.

Дискриминатор: аналогичный эксперименту один

Сетап:

- Добавлен warm-up для веса состязательной функции потерь.
- Размер датасета увеличен в 5 раз до 50 000 примеров.
- Количество голов дискриминатора: 4.
- Количество обучаемых параметров для голов дискриминатора: 7.4 М.

Датасет	Генерация учителя	Генерация ученика
6 7 2 9 2 3 4 3 1 8	6 7 2 9 2 3 4 3 1 8	6 7 2 9 2 3 4 3 1 8
8 5 3 6 1 2 9 3 8 7	8 5 3 6 1 2 9 3 8 7	8 5 3 6 1 2 9 3 8 7
6 2 7 3 9 4 1 5 9 9	6 2 7 3 9 4 1 5 9 9	6 2 7 3 9 4 1 5 9 9
4 8 7 4 2 8 1 6 2 8	4 8 7 4 2 8 1 6 2 8	4 8 7 4 2 8 1 6 2 8
4 5 2 0 7 9 8 7 9 5	4 5 2 0 7 9 8 7 9 5	4 5 2 0 7 9 8 7 9 5
8 2 5 4 2 2 2 5 0 7	8 2 5 4 2 2 2 5 0 7	8 2 5 4 2 2 2 5 0 7
0 3 3 9 9 2 7 5 4 5	0 3 3 9 9 2 7 5 4 5	0 3 3 9 9 2 7 5 4 5
3 0 4 4 7 2 1 4 5 3	3 0 4 4 7 2 1 4 5 3	3 0 4 4 7 2 1 4 5 3
1 6 6 1 0 3 8 6 8 8	1 6 6 1 0 3 8 6 8 8	1 6 6 1 0 3 8 6 8 8
9 3 4 2 0 1 1 8 6 8	9 3 4 2 0 1 1 8 6 8	9 3 4 2 0 1 1 8 6 8
1 7 8 7 7 5 5 8 4 1	1 7 8 7 7 5 5 8 4 1	1 7 8 7 7 5 5 8 4 1
6 0 3 9 8 2 1 6 1 1	6 0 3 9 8 2 1 6 1 1	6 0 3 9 8 2 1 6 1 1
0 2 4 8 0 6 1 2	0 2 4 8 0 6 1 2	0 2 4 8 0 6 1 2

Рис. 4: Сравнение датасет / учитель / ученик для эксперимента 2 на шаге обучения 20к.

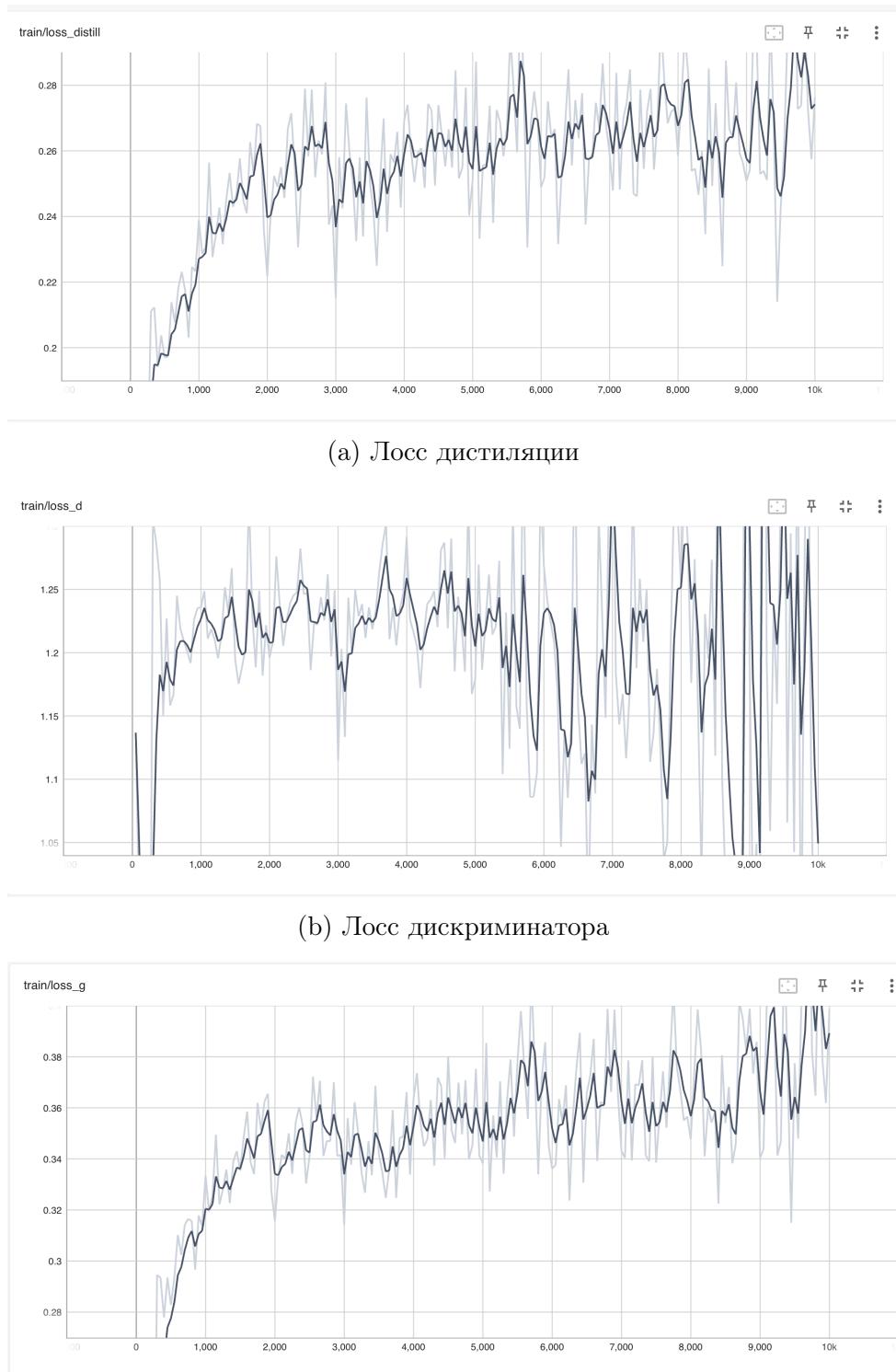


Рис. 5: Графики лоссов для эксперимента 2.

Наблюдения:

- На последних этапах обучения на сгенерированных данных снова наблюдался **mode collapse**.
- Однако при анализе промежуточных шагов обучения (5 000 и 20 000) оказа-

Датасет	Генерация учителя	Генерация ученика
7 5 1 1 4 2 7 6 1 5 8 6 3 8 4 5 7 5 8 7 2 1 8 2 4 3 9 4 2 5 9 2 5 8 4 7 1 9 7 1 1 3 3 5 7 1 6 1 2 2 9 9 1 7 0 5 5 0 1 2 5 9 3 3 8 7 6 7 7 6 6 0 1 1 5 7 8 5 6 5 0 8 8 5 5 7 9 2 3 1 4 8 4 5 0 0 6 2 2 9 9 9 8 1 1 8 2 6 7 8 3 0 5 6 2 7 2 8 5 7 3 5 5 6 2 3 7 5	7 5 1 1 4 2 7 6 1 5 8 6 3 8 4 5 7 5 8 7 2 1 8 2 4 3 9 4 2 5 9 2 5 8 4 7 1 9 7 1 1 3 3 5 7 1 6 1 2 2 9 9 1 7 0 5 5 0 1 2 5 9 3 3 8 7 6 7 7 6 6 0 1 1 5 7 8 5 6 5 0 8 8 5 5 7 9 2 3 1 4 8 4 5 0 0 6 2 2 9 9 9 8 1 1 8 2 6 7 8 3 0 5 6 2 7 2 8 5 7 3 5 5 6 2 3 7 5	7 5 1 1 4 2 7 6 1 5 8 6 3 8 4 5 7 5 8 7 2 1 8 2 4 3 9 4 2 5 9 2 5 8 4 7 1 9 7 1 1 3 3 5 7 6 1 2 2 9 9 1 7 0 5 5 0 1 2 5 9 3 3 8 7 6 7 7 6 6 0 1 1 5 7 8 5 6 5 0 8 8 5 5 7 9 2 3 1 4 8 4 5 0 0 6 2 2 9 9 9 8 1 1 8 2 6 7 8 3 0 5 6 2 7 2 8 5 7 3 5 5 6 2 3 7 5

Рис. 6: Сравнение датасет / учитель / ученик для эксперимента 2 на шаге обучения 5к.

лось, что цветовая вариативность сохраняется на начальных этапах обучения, поэтому было принято вычислить FID для данного чекпоинта.

- Время обучения: 5 часов.

Результаты: В результате вычисления FID мы получили значение 25.29 для модели-учителя на 4 шагах генерации и 98.88 для модели после дистилляции. Поскольку результаты дистилляции оказались неудовлетворительными, было принято продолжить эксперименты.

3.3 Эксперимент 3

Цель: увеличении количества обучаемых параметров дискриминатора и улучшении его способности сохранять цветовую вариативность изображений.

Дискриминатор: аналогичный эксперименту один

Сетап:

- Изменена архитектура на использование max-pooling в блоках Down.
- Добавили применение дискриминатора к большему числу признаку CUnet.
- Количество голов дискриминатора: 9.
- Количество обучаемых параметров для голов дискриминатора: 14.7 М.

Наблюдения:

- Модель снова начала схлопываться уже на тысячном шаге обучения.

3.4 Эксперимент 4

Цель: изменить дискриминатор.

Дискриминатор: Полностью переработан дискриминатор. Добавлены блоки Conv2d + спектральная норма и LeakyReLU, агрегирование признаков, и обновлен классификатор с полносвязными слоями. Эти изменения должны помочь учитывать важные латентные признаки, в особенности цвет.

Датасет	Генерация учителя	Генерация ученика
7 5 1 1 4 2 7 6 1 5 8 6 3 8 4 5 7 5 8 7 2 1 8 2 4 3 9 4 2 5 9 2 5 8 4 7 1 9 7 1 1 3 3 5 7 7 6 1 2 2 9 9 1 7 0 5 5 0 1 2 5 9 3 3 8 7 6 7 7 6 6 0 1 1 5 7 8 5 6 5 0 8 8 5 5 7 9 2 3 1 4 8 4 5 0 0 6 2 2 9 9 9 8 1 1 8 2 6 7 8 3 0 5 6 2 7 2 8 5 7 3 5 5 6 2 3 7 5	7 5 1 1 4 2 7 6 1 5 8 6 3 8 4 5 7 5 8 7 2 1 8 2 4 3 9 4 2 5 9 2 5 8 4 7 1 9 7 1 1 3 3 5 7 7 6 1 2 2 9 9 1 7 0 5 5 0 1 2 5 9 3 3 8 7 6 7 7 6 6 0 1 1 5 7 8 5 6 5 0 8 8 5 5 7 9 2 3 1 4 8 4 5 0 0 6 2 2 9 9 9 8 1 1 8 2 6 7 8 3 0 5 6 2 7 2 8 5 7 3 5 5 6 2 3 7 5	7 5 1 1 4 2 7 6 1 5 8 6 3 8 4 5 7 5 8 7 2 1 8 2 4 3 9 4 2 5 9 2 5 8 4 7 1 9 7 1 1 3 3 5 7 7 6 1 2 2 9 9 1 7 0 5 5 0 1 2 5 9 3 3 8 7 6 7 7 6 6 0 1 1 5 7 8 5 6 5 0 8 8 5 5 7 9 2 3 1 4 8 4 5 0 0 6 2 2 9 9 9 8 1 1 8 2 6 7 8 3 0 5 6 2 7 2 8 5 7 3 5 5 6 2 3 7 5

Рис. 7: Сравнение датасет / учитель / ученик для эксперимента 3 на шаге обучения 5к.

Датасет	Генерация учителя	Генерация ученика
7 4 3 7 7 3 1 0 7 3 1 8 3 0 2 8 2 9 5 9 2 0 9 7 4 7 2 3 5 9 0 6 6 1 0 6 9 4 9 3 1 3 7 4 7 8 2 7 2 7 8 6 6 0 0 6 6 1 3 3 2 2 7 8 5 2 3 3 0 9 2 4 8 0 2 3 3 7 9 7 2 5 6 4 8 8 6 3 4 3 0 3 5 3 2 5 6 2 2 5 2 3 0 8 7 9 2 4 9 4 1 2 1 3 4 4 2 6 5 0 1 2 8 8 5 9 9 9	7 4 3 7 7 3 1 0 7 3 1 8 3 0 2 8 2 9 5 9 2 0 9 7 4 7 2 3 5 9 0 6 6 1 0 6 9 4 9 3 1 3 7 4 7 8 2 7 2 7 8 6 6 0 0 6 6 1 3 3 2 2 7 8 5 2 3 3 0 9 2 4 8 0 2 3 3 7 9 7 2 5 6 4 8 8 6 3 4 3 0 3 5 3 2 5 6 2 2 5 2 3 0 8 7 9 2 4 9 4 1 2 1 3 4 4 2 6 5 0 1 2 8 8 5 9 9 9	7 4 3 7 7 3 1 0 7 3 1 8 3 0 2 8 2 9 5 9 2 0 9 7 4 7 2 3 5 9 0 6 6 1 0 6 9 4 9 3 1 3 7 4 7 8 2 7 2 7 8 6 6 0 0 6 6 1 3 3 2 2 7 8 5 2 3 3 0 9 2 4 8 0 2 3 3 7 9 7 2 5 6 4 8 8 6 3 4 3 0 3 5 3 2 5 6 2 2 5 2 3 0 8 7 9 2 4 9 4 1 2 1 3 4 4 2 6 5 0 1 2 8 8 5 9 9 9

Рис. 8: Сравнение датасет / учитель / ученик для эксперимента 4 на шаге обучения 5к.

Сетап:

- Количество голов дискриминатора: 9.
- Количество обучаемых параметров для голов дискриминатора: 5.1 М.

Наблюдения:

- mode collapse на начальных этапах.

3.5 Эксперимент 5

Цель: изменить количество голов дискриминатора и избавиться от возможного переобучения.

Дискриминатор: аналогичный эксперименту 4

Сетап:

- Количество голов дискриминатора: 5.
- Количество обучаемых параметров для голов дискриминатора: 5.1 М.

Наблюдения:

- mode collapse на начальных этапах.

Датасет	Генерация учителя	Генерация ученика
<pre> 7 5 1 1 4 2 7 6 1 5 8 6 3 8 4 5 7 5 8 7 2 1 8 2 4 3 9 4 2 5 9 2 5 8 4 7 1 9 7 1 1 3 3 5 7 1 6 1 2 2 9 9 1 7 0 5 5 0 1 2 5 9 3 3 8 7 6 7 7 6 6 0 1 1 5 7 8 5 6 5 0 8 8 5 5 7 9 2 3 1 4 8 4 5 0 0 6 2 2 9 9 9 8 1 1 8 2 6 7 8 3 0 5 6 2 7 2 8 5 7 3 5 5 6 2 3 7 5 </pre>	<pre> 7 5 1 1 4 2 7 6 1 5 8 6 3 8 4 5 7 5 8 7 2 1 8 2 4 3 9 4 2 5 9 2 5 8 4 7 1 9 7 1 1 3 3 5 7 1 6 1 2 2 9 9 1 7 0 5 5 0 1 2 5 9 3 3 8 7 6 7 7 6 6 0 1 1 5 7 8 5 6 5 0 8 8 5 5 7 9 2 3 1 4 8 4 5 0 0 6 2 2 9 9 9 8 1 1 8 2 6 7 8 3 0 5 6 2 7 2 8 5 7 3 5 5 6 2 3 7 5 </pre>	<pre> 7 5 1 1 4 2 7 6 1 5 8 6 3 8 4 5 7 5 8 7 2 1 8 2 4 3 9 4 2 5 9 2 5 8 4 7 1 9 7 1 1 3 3 5 7 1 6 1 2 2 9 9 1 7 0 5 5 0 1 2 5 9 3 3 8 7 6 7 7 6 6 0 1 1 5 7 8 5 6 5 0 8 8 5 5 7 9 2 3 1 4 8 4 5 0 0 6 2 2 9 9 9 8 1 1 8 2 6 7 8 3 0 5 6 2 7 2 8 5 7 3 5 5 6 2 3 7 5 </pre>

Рис. 9: Сравнение датасет / учитель / ученик для эксперимента 5 на шаге обучения 5к.

3.6 Эксперимент 6

Цель: окончательно избавиться от mode collapse.

Дискриминатор: аналогичный эксперименту 4

Сетап:

- В датасет добавлен шум, который прокидывается в студента.
- Размер датасета: 20 000 примеров.
- Количество обучаемых параметров для голов дискриминатора: 2.6 М.

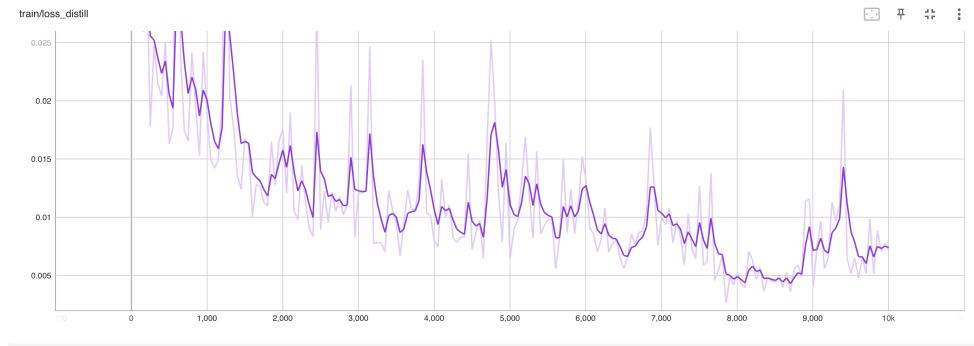
Шум для датасета	Датасет	Генерация учителя	Генерация ученика
	<pre> 9 1 4 4 6 3 8 8 8 9 2 7 2 4 8 6 3 6 3 4 6 0 8 1 3 7 8 3 7 9 7 3 1 0 2 8 7 3 3 7 0 4 5 2 7 9 3 0 1 3 0 5 4 9 5 4 2 3 8 9 7 6 4 5 2 8 9 1 6 6 6 4 0 1 6 1 1 3 8 8 5 4 5 0 2 7 2 4 2 0 4 9 0 2 7 6 1 1 6 4 4 6 2 9 6 9 4 8 2 5 0 4 4 5 9 6 1 6 2 4 6 9 0 3 8 6 3 5 </pre>	<pre> 9 1 4 4 6 3 8 8 8 9 2 7 2 4 8 6 3 6 3 4 6 0 8 1 3 7 8 3 7 9 7 3 1 0 2 8 7 3 3 7 0 4 5 2 7 9 3 0 1 3 0 5 4 9 5 4 2 3 8 9 7 6 4 5 2 8 9 1 6 6 6 4 0 1 6 1 1 3 8 8 5 4 5 0 2 7 2 4 2 0 4 9 0 2 7 6 1 1 6 4 4 6 2 9 6 9 4 8 2 5 0 4 4 5 9 6 1 6 2 4 6 9 0 3 8 6 3 5 </pre>	<pre> 9 1 4 4 6 3 8 8 8 9 2 7 2 4 8 6 3 6 3 4 6 0 8 1 3 7 8 3 7 9 7 3 1 0 2 8 7 3 3 7 0 4 5 2 7 9 3 0 1 3 0 5 4 9 5 4 2 3 8 9 7 6 4 5 2 8 9 1 6 6 6 4 0 1 6 1 1 3 8 8 5 4 5 0 2 7 2 4 2 0 4 9 0 2 7 6 1 1 6 4 4 6 2 9 6 9 4 8 2 5 0 4 4 5 9 6 1 6 2 4 6 9 0 3 8 6 3 5 </pre>

Рис. 10: Сравнение шум/ датасет / учитель / ученик для эксперимента 6 на шаге обучения 20к.

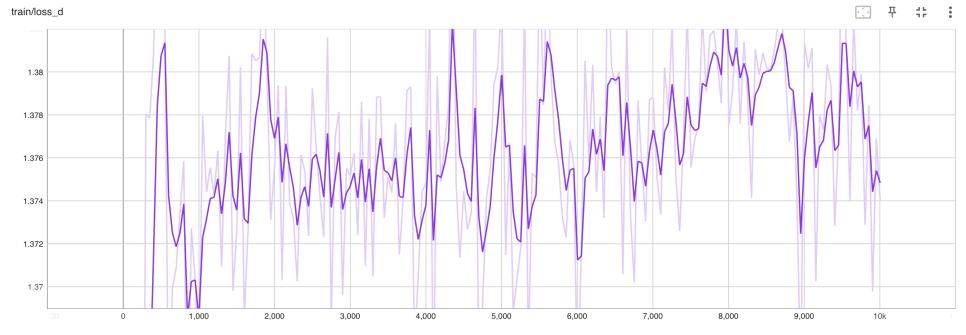
Наблюдения:

- Пропал mode collapse, так как решили изменить подход в генерации учеником и использовать шум из датасета.
- Время обучения: 6 часов.

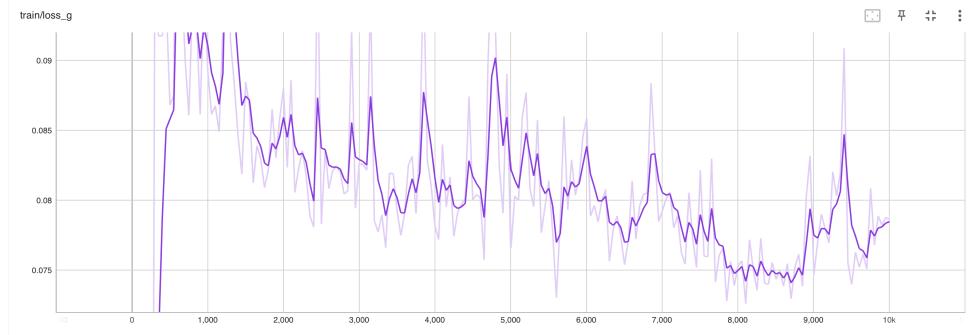
Результаты: В результате вычисления FID мы получили значение 25.29 для модели-учителя на 4 шагах генерации и 15.48 для модели после дистилляции. Максимальные результаты, которых мы могли добиться это 3.66 для модели-учителя на 50 шагах генерации.



(a) Лосс дистиляции



(b) Лосс дискриминатора



(c) Лосс генератора

Рис. 11: Графики лоссов для эксперимента 6.

4 Выводы

В ходе проведённых экспериментов была изучена применимость метода **Latent Adversarial Diffusion Distillation (LADD)** для генерации синтетических изображений в латентном пространстве. Основные наблюдения и выводы:

- **Плюсы метода:** LADD позволяет ускорить обучение диффузионной модели в разы за счёт работы в латентном пространстве и использования дистиляции. Метод помогает воспроизводить результаты модели-учителя, что подтверждается FID-метриками на синтетических датасетах. Включение состязательной функции потерь (warm-up) стабилизировало обучение и уменьшило вероятность mode collapse.
- **Минусы метода:** Метод чувствителен к конфигурации дискриминатора и чис-

лу голов. Неправильная настройка может приводить к mode collapse, особенно на малых датасетах. Также LADD требует корректного добавления шума для обучения студента, что сильно ограничило количество проведенных экспериментов. Отсутствие в статье глубокого описания метода.

- **Применимость:** Метод полезен, когда требуется быстро обучить генеративную модель на ограниченном объёме данных, сохраняя качество генерации учителя. Помогает получать генерации быстрее, чем учитель, что сохраняет количество ГПУ ресурсов.

Имплементация модели представлена по ссылке: <https://github.com/m4xig1/LADD/tree/main>

5 Распределение задач между членами команды

- **Леонтенков Егор:** генерация датасета, реализация архитектур моделей (учитель и студент), дискриминаторных голов, запуском обучения моделей и контролем качества.
- **Игнатов Максим:** реализовывал прямой проход модели (*forward*), вычисление функций потерь, метрик качества (включая FID), инфраструктуру для обучения (логгирование и пайплайн) и инференс.