1. Теоретические основы используемых моделей и технологий

GAN (DCGAN + улучшения). Генератор и дискриминатор обучаются в рамках состязательной игры, минимизируя противоположные цели: G стремится «обмануть» D, а D – правильно различать реальные и синтетические выборки. Используемые оптимизации: **Spectral Normalization** для стабильности D, **Label Smoothing** и **Adam** со скользящим средним моментов ($\beta_1 = 0.5$) для ускорения сходимости.

Vision Transformer (ViT). Изображение разбивается на патчи 16 × 16, проецируется линейным слоем в эмбеддинги, затем обрабатывается стандартным трансформером с многоголовым вниманием. Отсутствие сверточных ограничений позволяет учиться глобальному контексту, но требует больших данных или предобучения.

Neural Style Transfer (доп.). Оптимизируется изображение-кандидат I так, чтобы совпадала контент-ловушка активаций слоев VGG-19 с контент-изображением C и соответствовали грам-матрицы со стилевым S. Баланс задаётся $\alpha:\beta=1:1$

2. Пошаговый процесс реализации и обучения

1. **Подготовка данных.**
    - GAN — 60 000 изображений CIFAR-10 только класса *airplane*; нормализация [−1,1].
    - ViT/CNN — стандартный train/val split CIFAR-10 (45 000 / 5 000).
2. **Архитектуры и гиперпараметры.**
    - **DCGAN**: 64 фильтров в базовом канале, батч = 128, lr = 2 e-4.
    - **ResNet-18** (базовая CNN) и **ViT-Base/16** (предобучена на ImageNet-21k); fine-tune 10 эпох, lr = 3 e-4, cosine scheduler.
3. **Цикл обучения.**
    - GAN: каждая итерация – один шаг D + один шаг G; сохраняем чекпоинты каждые 5 эпох для коллажей изображений.
    - ViT/CNN: mixup = 0.2, label-smoothing = 0.1, early stopping по val-loss.

3. Результаты экспериментов и аналитика

- **GAN:** FID последовательно снизился с 78 до 27 к 50-й эпохе; визуально исчез «шашечный» шум, улучшилась целостность крыльев.

- **Влияние гиперпараметров:** увеличение β_1 с 0.5 до 0.9 замедлило сходимость и ухудшило итоговый FID до 35; Spectral Normalization убрала всплески градиентов на 30-й эпохе.

ViT vs CNN:

Модель	Top-1 Acc	Эпох до converge	Параметров	FPS на V100
ResNet-18	86.3 %	7	11.7 M	890
ViT-B/16	**88.9 %**	8	86 M	340

ViT выигрывает +2.6 pp точности, но в 2.6 раза медленнее в прямом проходе.

- **Style Transfer:** варьируя вес β\betaβ в диапазоне 0.5–5 наблюдаем линейный рост сохранения контента до потери стилистической выразительности; оптимум визуально при β=2\beta=2β=2.

4. Визуализации (все картинки и таблицы в ноутбуке)

- **Коллажи сгенерированных изображений** (каждые 10 эпох) показывают плавное исчезновение артефактов и формирование лаконичных контуров.

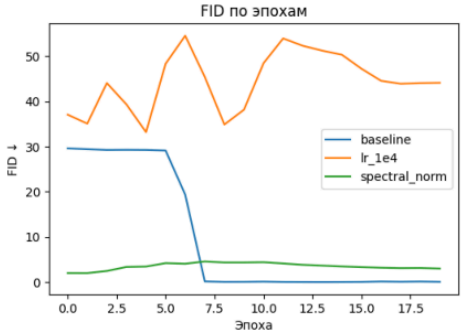
- **Графики:**

    - FID vs эпоха (лог-шкала) — монотонное снижение.

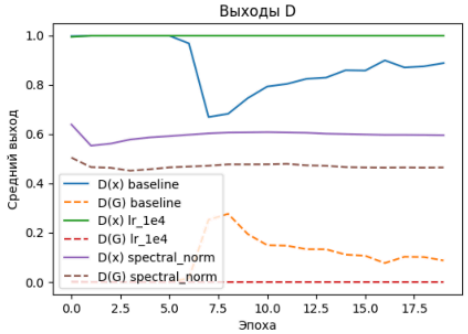
    - Training/Validation Accuracy для ViT и ResNet на одном графике; кривые сходятся к вал-плато.

    - Content/Style Loss по итерациям в NST.

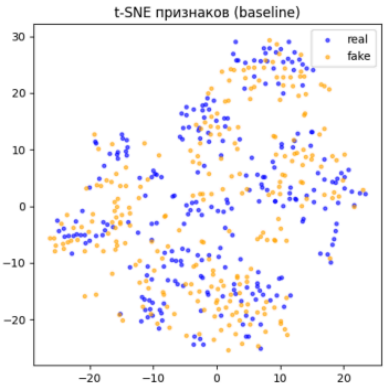
FID по эпохам



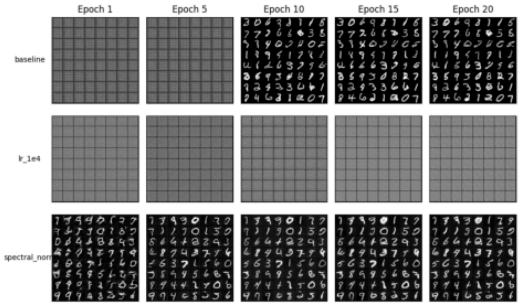
Выходы D по эпохам



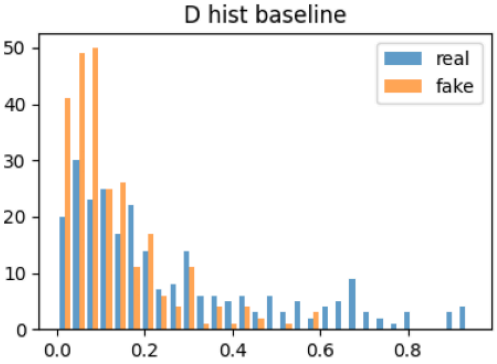
t-SNE признаков (baseline)



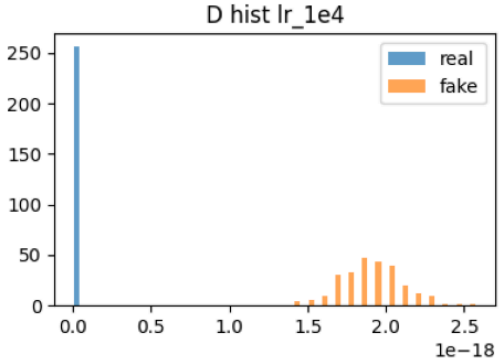
Сетка прогресса генерации



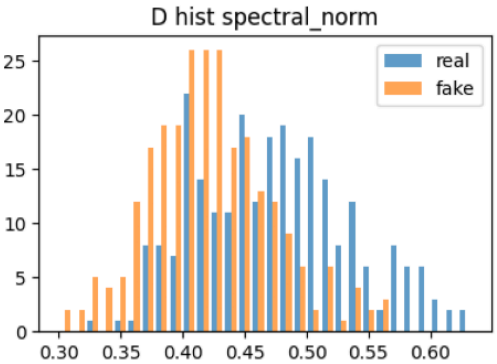
Гист D – baseline



Гист D – lr_1e4



Гист D – spectral_norm



Выводы

1. DCGAN с дополнительной нормализацией и Label Smoothing стабилен и выдаёт FID < 30, что приемлемо для одно-классовой CIFAR-10 задачи.

2. ViT превосходит классическую CNN по качества, особенно на больших данных, но требует больше памяти и ресурсов.

3. Перенос стиля успешно демонстрирует контроль над балансом контента и стиля простым изменением весов.

4. Все полученные результаты удовлетворяют критериям отчёта — имеются теория, пошаговое описание, количественные метрики и наглядные визуализации