

Проект 5: Диффузионные модели

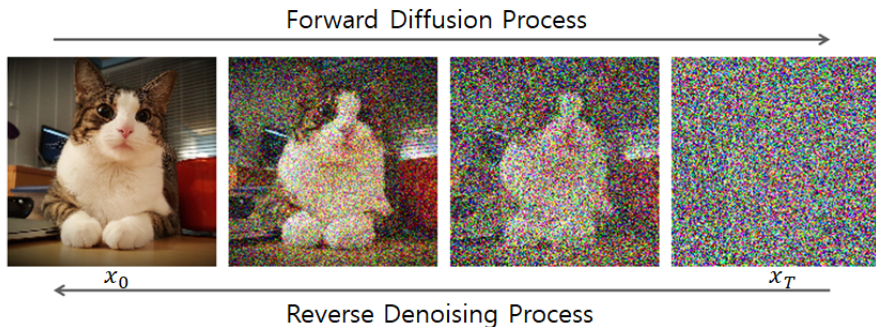
Аушев Ислам

14 июня 2025 г.

Задача генерации изображений

- Цель: научиться генерировать реалистичные изображения, похожие на данные из обучающего набора.
- Подходы:
 - GAN (Generative Adversarial Networks) — состязательное обучение генератора и дискриминатора.
 - VAE (Variational Autoencoders) — обучение скрытого пространства с помощью вариационного вывода.
 - **Диффузионные модели** — постепенное зашумление и восстановление изображения.
- Особенности диффузионных моделей:
 - Стабильное обучение (по сравнению с GAN).
 - Высокое качество генерации.
 - Пошаговый процесс, похожий на физическую диффузию.

Схема работы диффузионной модели



- Прямой процесс: постепенное добавление гауссова шума.
- Обратный процесс: обучение нейросети предсказывать шум и восстанавливать изображение.

Параметр α_t и его расписания

- α_t — параметр, определяющий уровень шума на шаге t .
- Уравнение прямого процесса:

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1 - \alpha_t}x_{t-1}, \alpha_t I)$$

- Расписания для α_t :
 - **Линейное:** $\alpha_t = \alpha_{\min} + (\alpha_{\max} - \alpha_{\min}) \cdot \frac{t}{T}$
 - **Косинусное:** $\alpha_t = \cos\left(\frac{\pi}{2} \cdot \frac{t}{T}\right)^k$ (при $k \approx 0.008$)
 - **Sigmoid:** $\alpha_t = \sigma\left(\frac{t - T/2}{s}\right) \cdot (\alpha_{\max} - \alpha_{\min}) + \alpha_{\min}$
- Качество генерации зависит от выбора расписания!

Условная генерация и Classifier-Free Guidance

- Условная генерация: модель учится учитывать дополнительную информацию (класс, текст и т. д.).
- **Classifier-Free Guidance** — баланс между условной и безусловной генерацией:

$$\tilde{\epsilon}_{\theta}(x_t, c) = \epsilon_{\theta}(x_t, \emptyset) + s \cdot (\epsilon_{\theta}(x_t, c) - \epsilon_{\theta}(x_t, \emptyset))$$

где:

- c — условие (например, текст),
 - \emptyset — безусловный режим,
 - s — коэффициент guidance
- Преимущества:
 - Не требует отдельного классификатора (в отличие от Classifier Guidance).
 - Позволяет гибко настраивать силу влияния условия.
 - Повышает качество генерации конкретных классов

Постановка задачи

В данном проекте хочется пронаблюдать на практике следующее:

- Как число шагов зашумления влияет на качество генерации
- Как на качество генерации влияет расписание параметров шума
- Улучшается ли качество при classifier-free guidance

Для проведения экспериментов были выбраны датасеты MNIST и FashionMNIST

Ссылки на wandb [MNIST](#), [FashionMNIST](#)

На каждом датасете было проведено:

- Обучение с разным количеством шагов зашумления
- Обучение с разным видом расписания шума
- Обучение с classifier-free guidance

Зависимость от числа шагов диффузии (MNIST)

При большем числе шагов лосс диффузии на валидации уменьшался. При этом непосредственное качество генерации на 1000 шагах получилось несильно лучше.



Примеры генерации для разного числа шагов

Из 24 семплов у модели с 1000 шагами 9 генераций с артефактами, а у 200 шагов 12. Разница есть, но несущественная и сама оценка может быть смещенной.

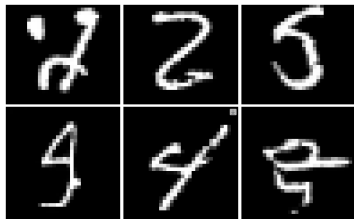
Зависимость от расписания шума (MNIST)

Рассматривались только запуски для 1000 шагов.

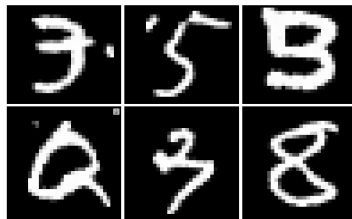
Из 48 генераций у 19 были артефакты для косинусного расписания шума, у линейного расписания 18 артефактов.

В целом по качеству генерации можно сделать вывод, что качество одинаково плохое

Зависимость от расписания шума (MNIST)



linear



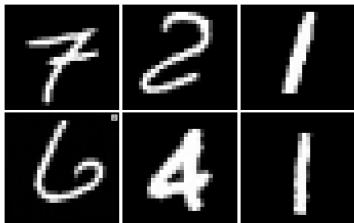
cosine

Classifier-free guidance (MNIST)

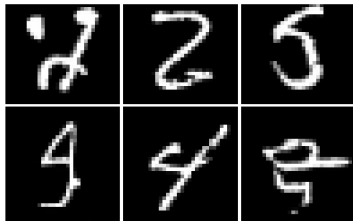
Для classifier-free guidance получилось получить ощутимый прирост в качестве генерации.

Для оценки качества генерации модели с classifier-free guidance было сгенерировано по 5 семплов каждой из цифр. Из 50 примеров у 7 семплов были артефакты, не позволяющие распознать цифру, которая была предоставлена для условной генерации. Исходя из этого можно сделать вывод, что Ассигасу модели составляет 0.86

Classifier-free guidance (MNIST)



classifier-free guidance



Без classifier-free guidance

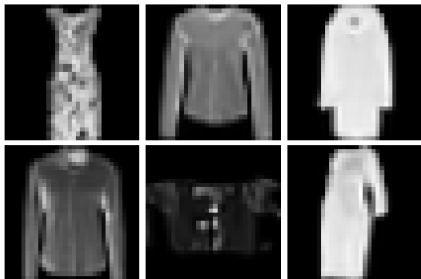
Из проведенных экспериментов можно сделать следующие выводы:

- Увеличение числа шагов в зашумлении может улучшать качество генерации.
- На датасете MNIST косинусное расписание шума не продемонстрировало себя лучше, чем линейное расписание
- С помощью classifier-free guidance можно достаточно легко обеспечить более качественную генерацию

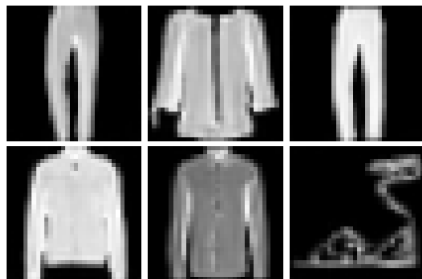
Зависимость от числа шагов диффузии (FashionMNIST)

Эксперименты на датасете FashionMNIST проводились на модели с большим числом параметров, чем для MNIST

Для FashionMNIST обучились модели для 100/500/1000 шагов, модель для 100 шагов генерирует хуже

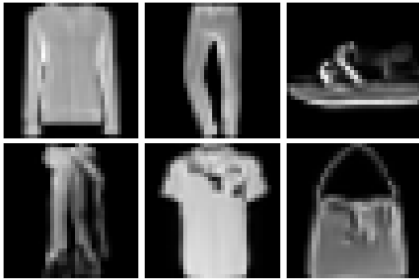


100 шагов

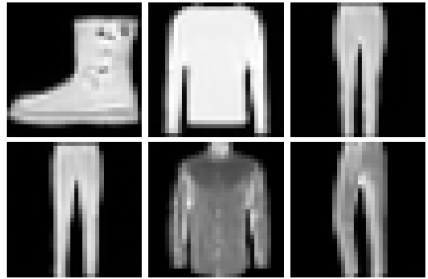


1000 шагов

Зависимость от числа шагов диффузии (FashionMNIST)



100 шагов



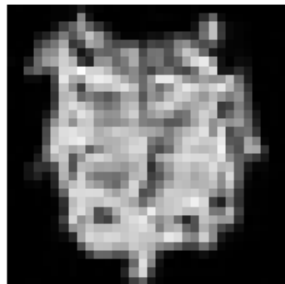
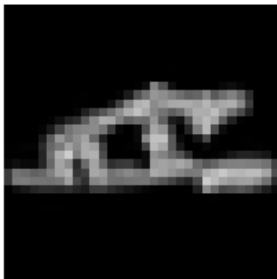
1000 шагов

Между моделями для 500 и 1000 шагов ощутимой разницы нет

Зависимость от расписания шума (FashionMNIST)

Рассматривались только запуски для 1000 шагов.

Общее качество генерации осталось таким же, но для FashionMNIST линейное расписание допускает больше артефактов в генерации



Артефакты линейного расписания шума

Косинусное расписание улучшило генерацию для FashionMNIST

Classifier-free guidance (FashionMNIST)

На FashionMNIST с помощью classifier-free guidance не получилось получить ощутимый прирост в качестве генерации. Возможно это связано с тем, что в сравнении с MNIST, базовое качество FashionMNIST получилось лучше.

Accuracy для classifier-free guidance не оценивалось, потому что какого-либо улучшения генерации относительно базовой модели не наблюдается

Из проведенных экспериментов можно сделать следующие выводы:

- Увеличение числа шагов в зашумлении на FashionMNIST получилось более наглядным генерации.
- На датасете FashionMNIST косинусное расписание шума продемонстрировало себя лучше, чем линейное расписание
- С помощью classifier-free guidance не получилось обеспечить более качественную генерацию

Спасибо за внимание!