

Отчёт по [проекту](#)

Зайцев Фёдор, Оздобихин Арсений

21 декабря 2024 г.

Содержание

1	Постановка задачи	2
1.1	Задача	2
1.2	Цели	2
1.3	Описание методов	2
1.3.1	Диффузионная модель	2
1.3.2	Эйлер	2
1.3.3	EDM	2
1.3.4	DDIM	3
2	Техническое описание экспериментов	3
3	Результаты экспериментов	3
3.1	Описание результатов	3
3.2	Анализ результатов	4
4	Вывод	4
5	Разделение работы	4

1 Постановка задачи

1.1 Задача

Реализация 3 солвера ОДУ, порожденного диффузионной моделью (Эйлер, DDIM, EDM), и сравнение их качества.

1.2 Цели

- Реализация солверов.
- Реализация необходимого для проведения экспериментов кода.
- Проведение экспериментов.
- Сравнение результатов

1.3 Описание методов

1.3.1 Диффузионная модель

В данной работе мы использовали диффузионную модель, обученную для выполнения задачи денойзера дифференциального уравнения со взрывающейся дисперсией, предсказывающего точку X_0 по входным данным X_σ и дисперсии σ . Будем обозначать эту модель $D_\sigma^\theta(X_\sigma)$. Мы сравнили несколько алгоритмов решения ОДУ при помощи данной модели, рассмотрим их далее. Для каждого алгоритма запишем один шаг генерации для момента i . Стоит учесть, что при сэмплировании мы будем итерироваться по i в порядке убывания от T к 0 , где T - общее заданное количество шагов генерации.

1.3.2 Эйлер

Данный солвер реализовывал алгоритм построения исходных данных при помощи схемы Эйлера. Мы использовали ее для решения обратного ОДУ и, соответственно, применяли формулу дискретизации $\mathbf{Y}_{t-h} \approx \mathbf{Y}_t + h \cdot t \cdot \nabla \log p_t(\mathbf{Y}_t)$. Выражая наш алгоритм через конкретные узлы дискретизации и применяя параметризацию через обученный денойзер мы имеем следующую формулу для ОДУ сэмплирования:

$$x_{i-1} = x_i + (t_i - t_{i-1}) \cdot \frac{D_{t_i}^\theta(x_i) - x_i}{t_i}$$

Данную схему генерации мы подробно обсуждали на семинарах и лекциях и интуиция за ней стоит довольно простая - мы просто дискретизируем обратное дифференциальное уравнение по схеме Эйлера.

1.3.3 EDM

Стохастический сэмплер с коррекцией направления второго порядка, предложенный в той же статье, что и EDM модель денойзера, применяет улучшенную схему Эйлера - Heun's 2nd order method, также известную как трапезоидальное правило. Данный метод заключается в оценке диффура в следующей точке дискретизации при помощи двух обращений к нашей функции оценки, вместо одного, по следующему правилу:

$$\hat{y}_{i+1} = y_i + hf(t_i, y_i)$$

$$y_{i+1} = y_i + \frac{h}{2}(f(t_i, y_i) + f(t_{i+1}, \hat{y}_{i+1}))$$

Авторы отмечают, что применение подобного метода позволяет улучшить качество сэмплирования, не так сильно при этом набрав в количестве обращений к нашей оценочной функции.

Авторы статьи предлагают комбинировать использование данного метода вместе с добавлением дополнительной стохастичности посредством добавления и отнятия шума. Итого получаем следующий алгоритм:

$$\hat{x}_i = x_i + \sqrt{((1 + \gamma_i)t_i)^2 - t_i^2\epsilon_i} - \text{добавление нового шума}$$

$$\hat{x}_{i-1} = \hat{x}_i + (\hat{t}_i - t_{i-1}) \cdot \frac{D_{\hat{t}_i}^\theta(\hat{x}_i) - \hat{x}_i}{\hat{t}_i} - \text{первый шаг Эйлера}$$

$$x_{i-1} = \hat{x}_i + (\hat{t}_i - t_{i-1}) \cdot \frac{1}{2} \left(\frac{D_{\hat{t}_i}^\theta(\hat{x}_i) - \hat{x}_i}{\hat{t}_i} + \frac{D_{t_{i-1}}^\theta(\hat{x}_{i-1}) - \hat{x}_{i-1}}{t_{i-1}} \right) - \text{коррекция второго порядка}$$

Solver	Euler	DDIM	EDM	Solver	Euler	DDIM	EDM
FID	11,55	354,51	10,68	FID	10,96	152,53	11,39

Таблица 1: Для 50 (слева) и 200 (справа) степов и 3000 семплов

Solver	Euler	DDIM	EDM
FID	7,59	321,57	7,19

Таблица 2: Для 100 степов и 5000 семплов

За стохастичность в данном алгоритме отвечают гиперпараметры γ_i и случайно засэмплированные величины $\epsilon_i \sim \mathcal{N}(0, S_{noise}^2 I)$

Узлы дискретизации для этого метода, как и для предыдущего, выбирались с опорой на статью про модель EDM (оптимальные гиперпараметры со сглаженным переходом от σ_{min} к σ_{max} так же, как на семинаре.)

1.3.4 DDIM

Этот алгоритм, в отличие от двух предыдущих, работает с дифференциальным уравнением с сохраняющейся дисперсией $X_t = \sqrt{\alpha_t}X_0 + \sqrt{1 - \alpha_t}\epsilon$. Соответственно нетрудно показать путем легких преобразований, что для получения оценки \hat{X}_0 , нам нужно вычислить функцию $D^{\theta}_{\sqrt{\frac{1-\alpha_t}{\alpha_t}}}(\frac{X_t}{\sqrt{\alpha_t}})$. Есть правда еще один нюанс - оригинальный алгоритм работает с моделью, параметризованной на оценку шума - разницы между X_t и X_0 , так что нам и здесь придется немного извернуться с формулами, оценив этот самый шум, как $\hat{\epsilon}_t = \frac{X_t - \hat{X}_0\sqrt{\alpha_t}}{\sqrt{1-\alpha_t}}$. Имея на руках наши оценки для шума и изначальной точки, мы можем выписать шаг DDIM, он будет выглядеть следующим образом:

$$x_{i-1} = \sqrt{\alpha_{i-1}}\hat{x}_0 + \sqrt{1 - \alpha_{i-1}}\hat{\epsilon}_i$$

Данный алгоритм является частным случаем более генерализованного стохастического алгоритма, обобщающего генерацию для не-Марковских процессов. Обнуление в этом процессе коэффициента, отвечающего за добавление шума на каждом шаге генерации, и дает нам данный алгоритм.

Узлы дискретизации для этого метода выбирались исходя из результатов в соответствующей работе - наилучшего качества авторам удалось добиться для квадратичного расписания β_t .

2 Техническое описание экспериментов

Все эксперименты проводились на датасете CIFAR-10, в качестве диффузионной модели была использована модель, предложенная в статье по EDM, обученная на датасете CIFAR-10 для необусловленной генерации с помощью VE диффура (получена по [ссылке](#)). Для каждого из реализованных солверов было проведено несколько экспериментов по генерации и подсчету FID сора. Мы сравнивали между собой работу алгоритмов при количестве шагов генерации 50 и 200. Для этих значений FID скор подсчитывался на 3000 сгенерированных изображениях. Также мы подсчитывали FID скор для 5000 изображений при 100 шагах генерации для каждого солвера. Далее проводилось количественное (по FID) и качественное (визуально) сравнение результатов.

Остальные гиперпараметры (настройки стохастичности в EDM) выставлялись в соответствии с оптимальными гиперпараметрами, описанными в статьях.

Более подробное описание процесса запуска экспериментов можно увидеть в [README](#)

Запуск экспериментов занимал разное количество времени от 10 минут (50-шаговая генерация 3000 изображений Эйлером), до нескольких часов (генерация 1000 шагов DDIM для 3000 изображений).

3 Результаты экспериментов

3.1 Описание результатов

Было проведено несколько экспериментов, с различным количеством шагов солверов:

Кроме сравнения методов между собой (таб. 1), мы также убедились в том, что увеличение количества сэмплов для подсчета FID благоприятно влияет на скор (что также было отмечено в других экспериментах как с другими моделями и датасетами, так и с [нашими](#)) (таб.2) и запустили наиболее оптимальные количества степов (из оригинальных статей) для DDIM (1000 с FID = 11,24) и EDM (256 с FID = 11,57).

Steps	50	200	500	1000
FID	354,51	152,53	18,61	11,24

Таблица 3: Результаты DDIM в зависимости от количества шагов

Для визуального сравнения результатов мы прилагаем примеры сгенерированных изображений для лучшего конфига для каждого из солверов (их можно посмотреть в конце отчёта).

3.2 Анализ результатов

Наши эксперименты показали, что при работе с моделью EDM сопоставимое между собой качество дают солверы Эйлер и EDM. При одинаковом количестве итераций они давали близкие значения FID во всех сеттингах. Стоит однако отметить, что метод генерации EDM второго порядка потребляет заметно большее количество вычислительных ресурсов, так как на каждом шаге итерации ему нужно дважды посчитать выход модели, а Эйлеровской схеме - лишь раз. Вследствие чего можно сделать вывод о том, что схема Эйлера при сравнимом качестве генерации является более эффективным подходом при учете затраченных вычислительных ресурсов. Компенсирует этот недостаток алгоритм EDM своей дополнительной стохастичностью - при помощи введения дополнительного шума он имеет возможность генерировать более разнообразные сэмплы, чем алгоритм генерации по Эйлеру. Что касается алгоритма DDIM, эксперименты показали, что ему нужно наибольшее количество шагов для схождения к адекватному результату. При этом он способен показывать сопоставимое качество, но просто за гораздо большее время (именно за счёт необходимости использовать большее количество степеней). Для 1000 шагов генерации DDIM показывал качество на уровне с 50-200 шагами генерации EDM или Эйлеровского алгоритма (таб. 3).

Визуальный анализ также позволяет убедиться в том, что все солверы при оптимальных настройках гиперпараметров показывают сравнимое между собой относительно неплохое качество. Можно убедиться «методом пристального взгляда» в том, что сгенерированные сэмплы достаточно разнообразны и отражают все классы датасета, выдавая терпимое качество, но не всегда избегая некоторых огрехов, артефактов, зашумленностей и нереалистичных деталей.

4 Вывод

По результатам экспериментов среди предложенных методов наименее применимым оказался метод DDIM - он работает значительно дольше других двух методов, не оправдывая это значительно более лучшим качеством. Методы Эйлера и EDM могут быть успешно применены в разных ситуациях, например метод Эйлера может быть более успешен в случае ограниченности вычислительных ресурсов, так как задействует заметно меньшее количество таковых, а метод EDM может быть использован в случаях, когда нам нужно добиться большей стохастичности генерации, так как он содержит в себе двойное применение схемы Эйлера (каждое из которых можно обратить в стохастичное) и также дополнительный шум, позволяющий добиться большего разнообразия в генерации.

Резюмируя вышенаписанное, в ходе выполнения проекта нам удалось достичь поставленных целей и, несмотря на достаточно большие значения FID, получить достаточно естественно выглядящие изображения в разрешении 32×32 на которых чётко прослеживаются классы, которые содержатся в датасете CIFAR-10.

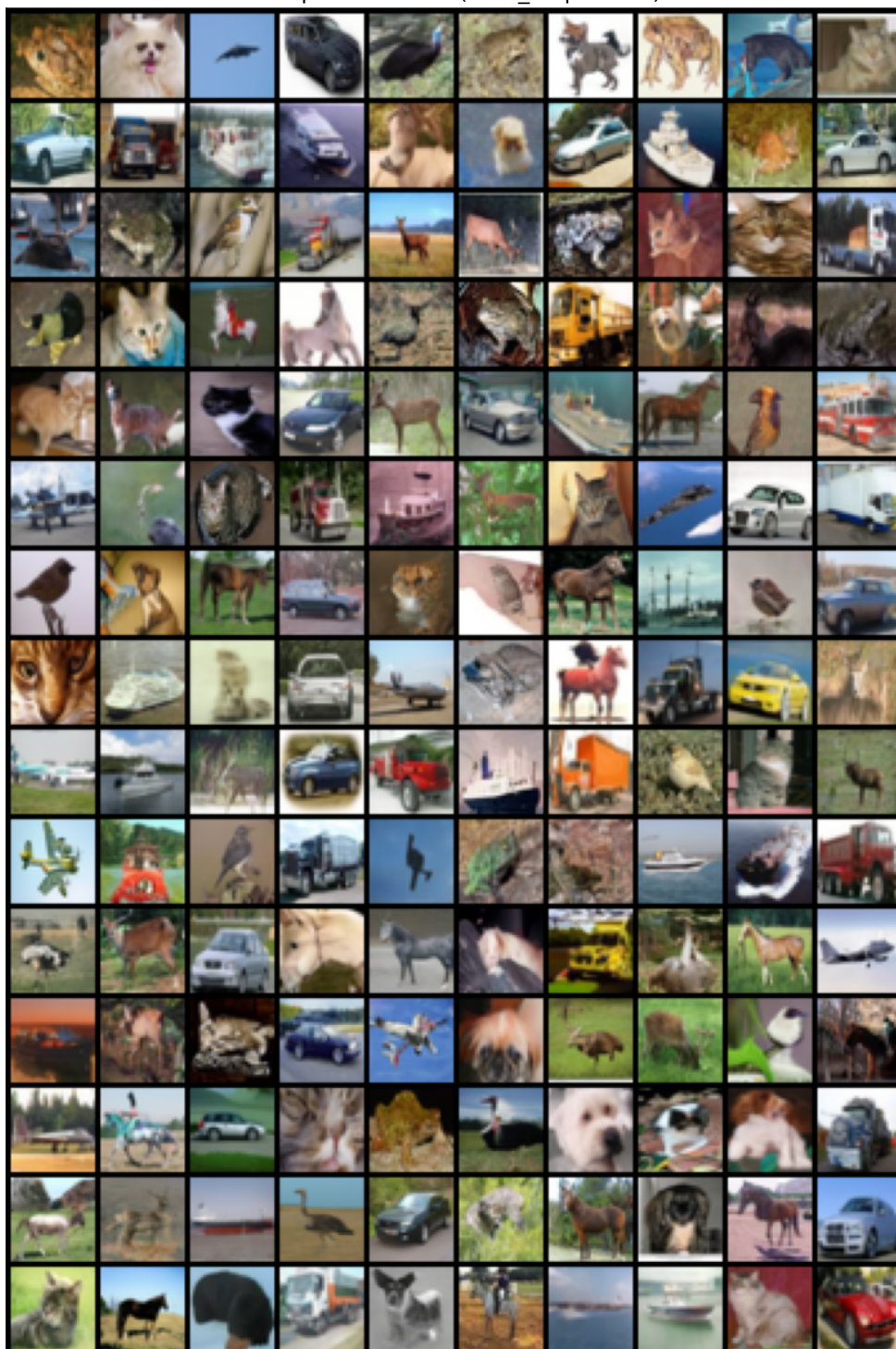
В целом, результаты нашей работы показывают, что рассмотренные решения достаточно хорошо подходят для датасетов в маленьком разрешении, и два из них работают достаточно быстро. Тем не менее, интересно было бы проверить качество решений и на других датасетах большего разрешения, а также сравнить FID для большего количества сгенерированных сэмплов.

5 Разделение работы

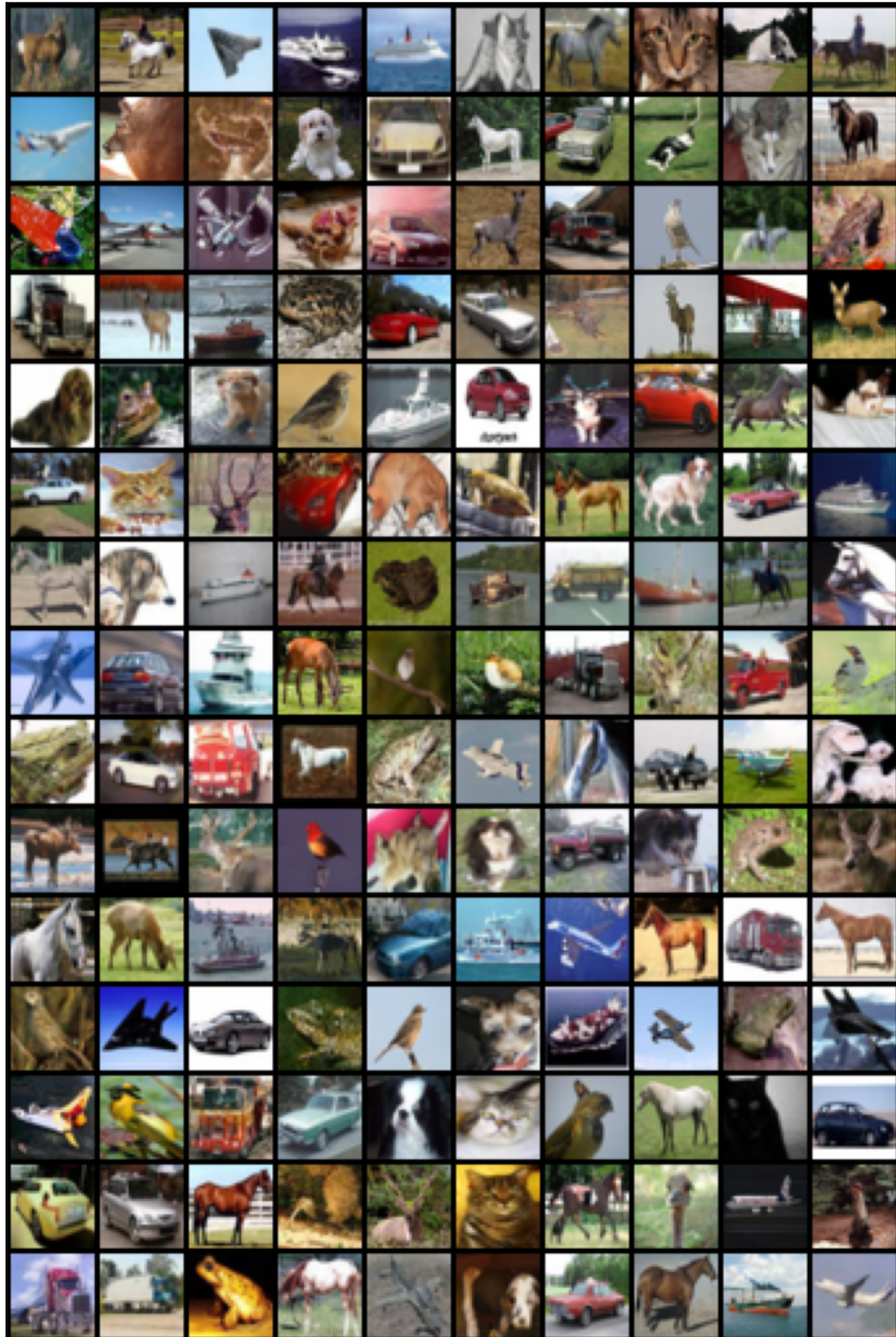
- Реализация солверов:
 - Euler — Фёдор (ред. Арсений);
 - DDIM — Фёдор (ред. Арсений);
 - EDM — Арсений (ред. Фёдор);
- Составление архитектуры — совместно;
- Реализация общего для всех солверов кода — Арсений (ред. Фёдор);

- Создание и оформление репозитория — Арсений (ред. Фёдор);
- Проведение экспериментов — совместно;
- Составление отчёта — совместно.

Samples for Euler (num_steps=200)



Samples for EDM (num_steps=100)



Samples for DDIM (num_steps=1000)

