

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ЯДЕРНЫЙ УНИВЕРСИТЕТ «МИФИ»
(НИЯУ МИФИ)

ОТЧЁТ

по проекту «Решение уравнения Бесселя с использованием
физически-информированных нейронных сетей (PINN)»

Группа

Б23-215

Студенты

А.Е. Апарин
Г.Е. Жмелев
О.С. Паршева

Руководитель работы

С.А. Ладыгин

Москва 2024

Аннотация

Решалась задача нахождения приближенного решения уравнения Бесселя первого рода первого порядка с использованием физически-информированных нейронных сетей (PINN). Рассмотрены различные способы выбора последовательностей точек для обучения нейронных сетей. Проведено сравнение данных способов и их влияние на обучение моделей

Содержание

1.	Введение	3
2.	Физически-информированные нейронные сети (PINN). Квазислучайные последовательности	3
3.	Решение дифференциального уравнения Бесселя	4
4.	Результаты	5
5.	Выводы	10

1. Введение

Физически-информированные нейронные сети решают задачу нахождения приближенного решения дифференциальных уравнений, описывающих физический процесс. Их главная отличительная способность от стандартного подхода нахождения функции, похожей на заданную, с помощью выбора опорных точек заключается в том, что они используют знание о дифференциальном уравнении, описывающем физический закон. Используя это знание, такие нейросети начинают удовлетворять дифференциальному уравнению, приближаясь к его решению.

2. Физически-информированные нейронные сети (PINN). Квазислучайные последовательности

Физически-информированные нейронные сети (PINN)

Пусть задано обыкновенное дифференциальное уравнение:

$$F(x, y, y', y'', \dots, y^{(n)}). \quad (1)$$

С начальными условиями:

$$\begin{cases} y(0) = a_0, \\ y'(0) = a_1, \\ y''(0) = a_2, \\ \dots, \\ y^{(n)}(0) = a_n. \end{cases} \quad (2)$$

где $y(x)$ — искомая функция. Пусть $u(x)$ — наша нейронная сеть. Зададим функцию ошибки:

$$Loss(x) = MSE_{BC} + MSE_{Phys}. \quad (3)$$

MSE_{BC} — сумма среднеквадратичных ошибок на граничных условиях.

MSE_{Phys} — физическая ошибка.

То есть:

$$\begin{aligned} MSE_{BC} &= MSE_{BC_0} + MSE_{BC_1} + \dots + MSE_{BC_n}, \\ MSE_{BC_k} &= \frac{1}{N} \sum_{i=1}^N (y^{(k)}(x_i) - u^{(k)}(x_i))^2, \\ MSE_{Phys} &= \frac{1}{N} \sum_{i=1}^N (F(x_i, y(x_i), y', y'', \dots, y^{(n)}) - F(x_i, u(x_i), u', u'', \dots, u^{(n)}))^2. \end{aligned} \quad (4)$$

$y^{(i)}$ — производная порядка i функции $y(x)$. $u^{(i)}$ — производная порядка i функции $u(x)$, вычисляемая с помощью методов автоматического дифференцирования. С помощью метода обратного распространения ошибки (Back Propagation) возможно уточнить значение набора $\{\theta_i\}_{i=1}^k$ весов нейронной сети $u(x)$ таким образом, чтобы уменьшить значение функции ошибки $Loss(x)$, тем самым приблизив $u(x)$ к реальному решению $y(x)$ заданного дифференциального уравнения.

Квазислучайные последовательности

Для обучения нейронной сети требуется выбрать последовательность точек $\{x_i\}_{i=1}^s$, на которых будет проводиться обучение (то есть на них будет вычисляться функция ошибки, и веса нейронной сети будут уточняться таким образом, чтобы уменьшить ошибку на точках выбранной последовательности). Так как задача Коши определяет набор точек для граничных условий, последовательность $\{x_i\}_{i=1}^s$ выбирается для физического закона.

В данной работе было исследовано влияние на обучение нейронной сети различных последовательностей:

- Равномерное распределение.
- Случайное равномерное распределение.
- Последовательности, меняющиеся динамически в ходе обучения.
- Квазислучайные последовательности.

Квазислучайные последовательности с низким расхождением — последовательности, вычисленные по некоторому закону, при этом они не являются ни случайными, ни псевдослучайными. Их особенность состоит в том, что они наиболее равномерно охватывают выбранную область и улучшают работу различных алгоритмов. Например, численный счет определенного интеграла следующего вида:

$$\int_0^1 f(u) du \approx \frac{1}{N} \sum_{i=1}^N f(x_i), \quad (5)$$

Дает меньшую ошибку при использовании квазислучайной последовательности $\{x_i\}_{i=1}^s$, в отличие от случайного равномерного распределения или выбора точек с помощью правила треугольника $x_i = \frac{i}{N}$.

В данной работе использовалась квазислучайная рекуррентная R -последовательность, описанная в работе Мартина Робертса[1].

3. Решение дифференциального уравнения Бесселя

Постановка задачи

Требуется исследовать различные стратегии выбора точек обучения (равномерная сетка, случайное распределение, адаптивный выбор, квазислучайные последовательности) для нахождения функций Бесселя с помощью PINN.

Дифференциальное уравнение Бесселя

Дифференциальное уравнение Бесселя имеет следующий вид:

$$x^2 \frac{d^2 y}{dx^2} + x \frac{dy}{dx} + (x^2 - \alpha^2) y = 0, \quad (6)$$

$\alpha \in \mathbb{R}.$

Коэффициент α называется порядком уравнения. Порядок уравнения также определяет его род. Функции Бесселя разделяют на функции первого и второго рода.

Для удобства мы положим коэффициент $\alpha = 0$. Таким образом, мы будем искать приближенное решение следующего уравнения:

$$x^2 \frac{d^2 y}{dx^2} + x \frac{dy}{dx} + x^2 y = 0. \quad (7)$$

Это дифференциальное уравнение Бесселя первого порядка первого рода. Граничные условия имеют следующий вид:

$$\begin{cases} y(0) = 1, \\ y'(0) = 0. \end{cases} \quad (8)$$

Структура модели

Нейронная сеть, используемая для решения задачи представлена следующим образом

Слой	Входные/Выходные связи	Кол-во параметров
Входной	(0, 1)	1
Полносвязный 1	(32, 32)	64
Полносвязный 2	(32, 32)	1,056
Полносвязный 3	(32, 1)	1,056
Выходной	(1, 0)	33

На каждом слое используется функция активации \tanh

Функция ошибки представлена следующим образом:

$$Loss(x) = MSE_{BC_0} + MSE_{BC_1} + MSE_{Phys} + MSE_{Inn}. \quad (9)$$

$MSE_{BC_0}, MSE_{BC_1}, MSE_{Phys}$ заданы таким же образом как в примере (4). MSE_{Inn} — ошибка на внутренних точках.

$$MSE_{Inn} = \frac{1}{N} \sum_{i=1}^N (y(x_i) - u(x_i))^2.$$

В нашей реализации PINN берутся несколько точек, на которых вычисляются значения решения уравнения Бесселя с помощью явного метода Рунге-Кутты 4-го порядка. Использование MSE_{Inn} обосновано попыткой сделать обучение модели более быстрым.

Модель при обучении использует оптимизатор *Adam* с коэффициентом 1.0×10^{-3} .

4. Результаты

Общие сведения

Обучено 4 модели на различных последовательностях. Для каждой модели построены графики обучения и средней абсолютной ошибки *MAE* относительно настоящего решения в процессе обучения. Кроме этого, созданы изображения, содержащие графики приближенного и реального решений для визуального сравнения.

Равномерное распределение

Для равномерного распределения модель обучалась на 7.0×10^4 эпохах.

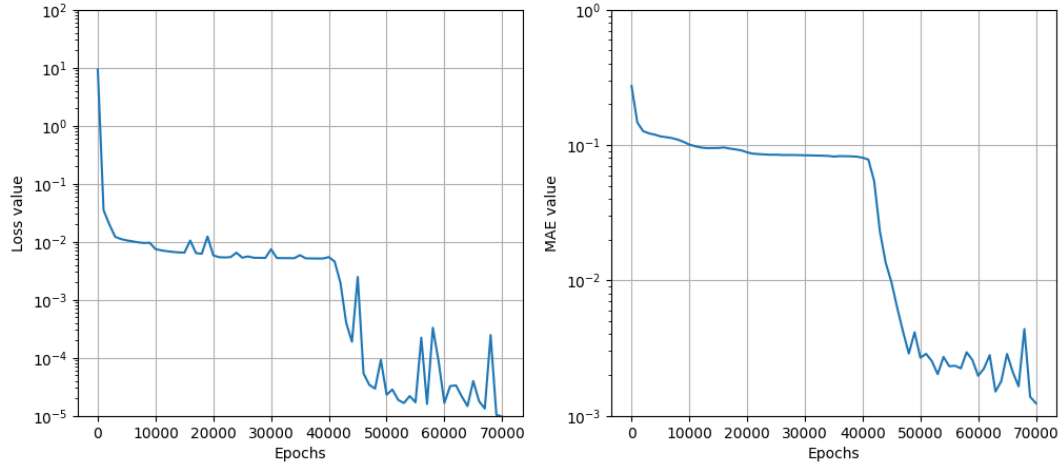


Рис. 1. Обучение на равномерной последовательности (100 точек в отрезке $[0; 20]$) Слева — график функции ошибок в зависимости от эпох. Справа — график MAE для тех же эпох.

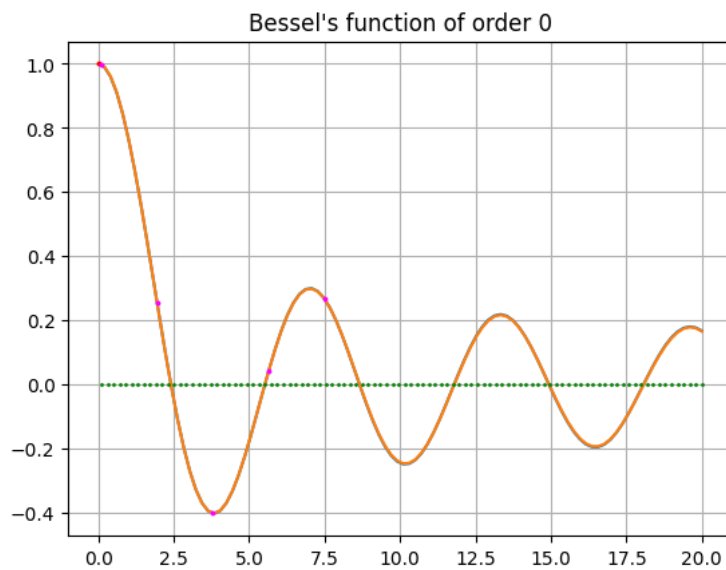


Рис. 2. Сравнение реального решения и модели. График реально-го решения имеет синий цвет. График модели имеет оранжевый цвет. Красная точка — точка, а которой проверяются MSE_{BC_0} и MSE_{BC_1} . Зеленые точки используются для MSE_{Phys} , а розовые отвечают за MSE_{Inn} .

Стоит отметить, что модель до 4.0×10^4 эпох обучается очень медленно, после наблюдается резкий скачок, который происходит из-за резкого уменьшения значения MSE_{Phys} .

Случайное равномерное распределение

Для случайного равномерного распределения модель обучалась на 7.0×10^4 эпохах.

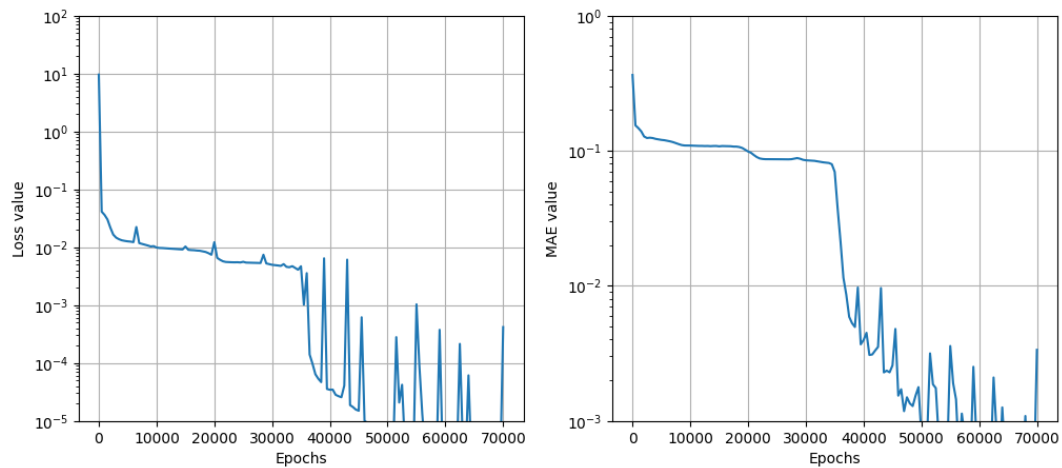


Рис. 3. Обучение на случайной равномерной последовательности (100 точек в отрезке $[0; 20]$)

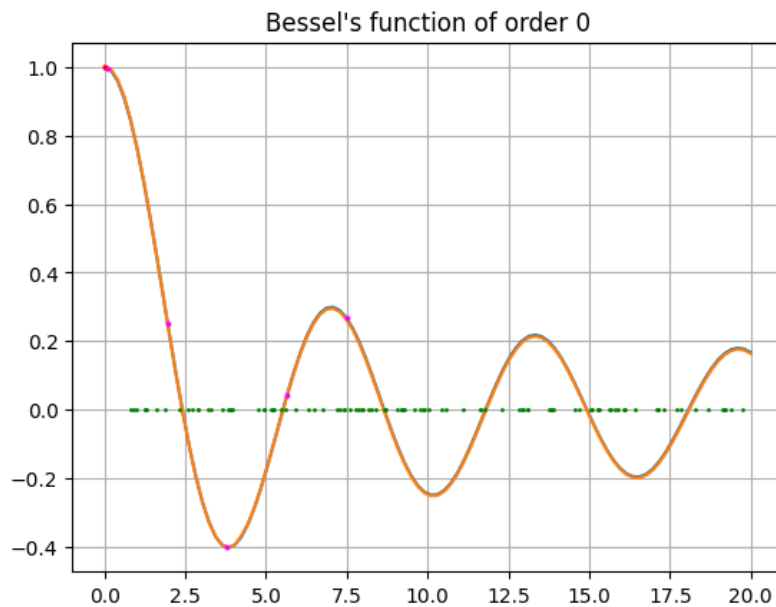


Рис. 4. Изображение аналогично предыдущему примеру.

Следует заметить, что в функции ошибок все еще наблюдается резкий скачок. Однако он происходит уже примерно на 3.6×10^4 эпохах. MAE принимает меньшие значения для данной модели, чем для прошлой.

Адаптивный шаг

Модель с адаптивным шагом обучалась на 5.0×10^4 эпох. Выбор точек для последовательности делался следующим образом: каждые 500 эпох вычислялась MSE_{Phys} на 1.0×10^3 равномерно распределенных точках на отрезке $[0; 20]$. Происходила сортировка точек по значению MSE_{Phys} для каждой точки. Затем отбирались первые 100 точек с наибольшими значениями MSE_{Phys}

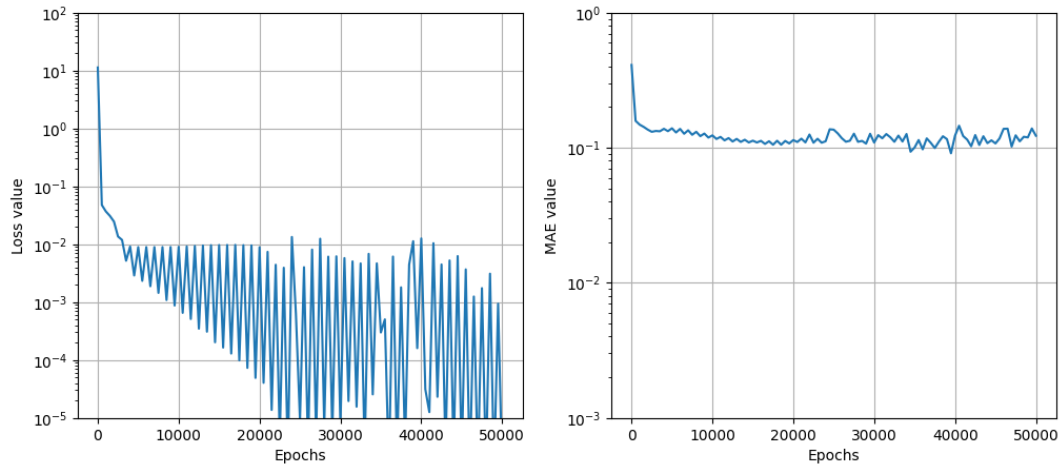


Рис. 5. Обучение модели с адаптивным шагом (100 точек в отрезке $[0; 20]$)

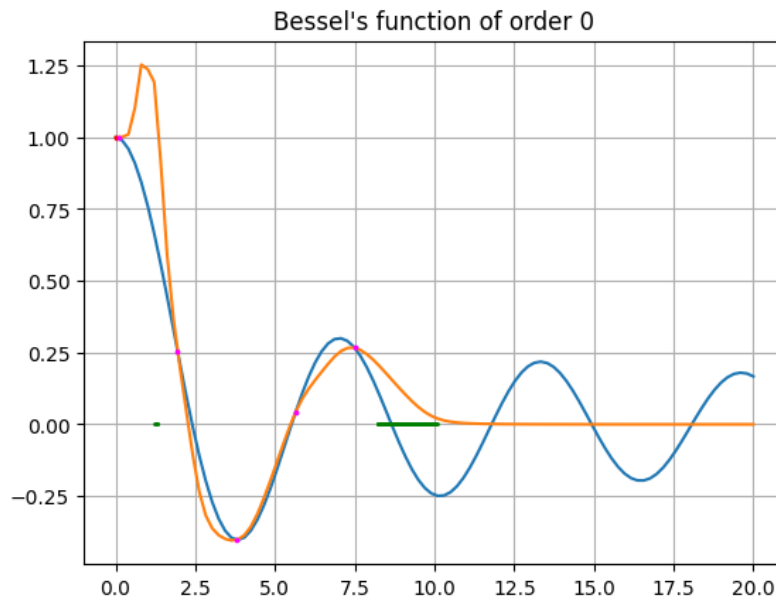


Рис. 6. Изображение аналогично предыдущему примеру. Зеленые точки отображены для последних 500 эпох обучения.

Данная модель показала плохую способность к обучению. Следует сделать предположение, что это происходит из-за того, что концентрация точек обучения исполь-

зуемых в MSE_{Phys} для каждой новой последовательности очень велика. В итоге модель пытается подстроиться под дифференциальное уравнение на небольшом отрезке, игнорируя близлежащие точки. Таким образом, концентрируясь на малых отрезках, модель вредит своему обучению. Прогресс накопленный на некотором отрезке нивелируется уменьшением ошибки на другом отрезке.

Квазислучайные последовательности

Для квазислучайных последовательностей модель обучалась на 5.0×10^4 эпох.

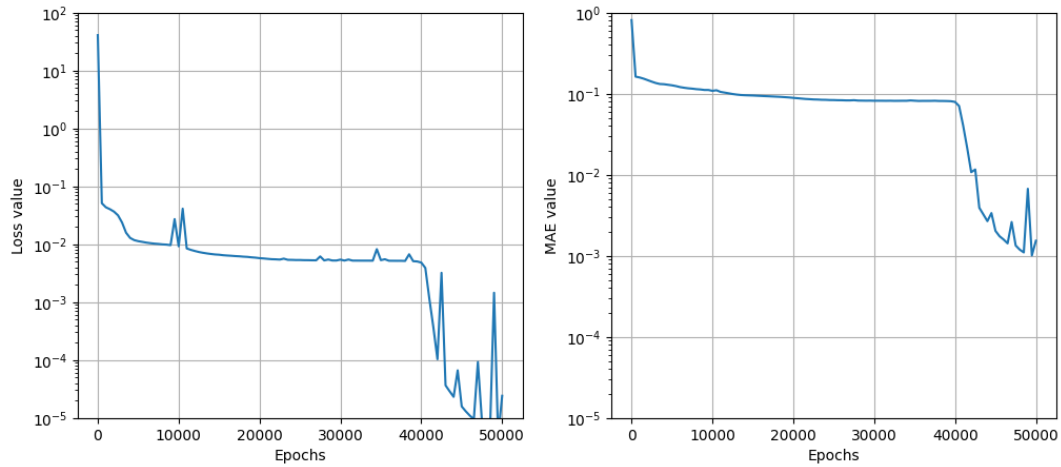


Рис. 7. Обучение на квазислучайной последовательности (100 точек в отрезке $[0; 20]$)

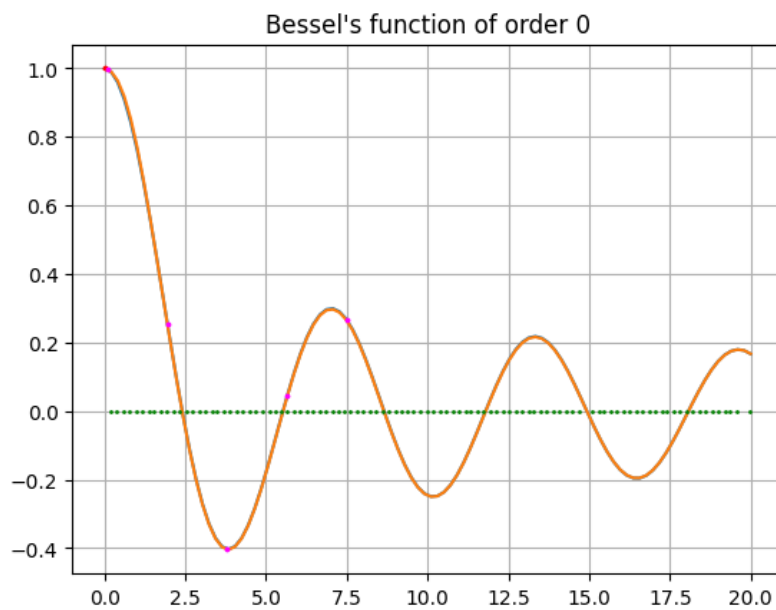


Рис. 8. Изображение аналогично предыдущему примеру.

Стоит заметить, что данная модель обучалась чуть лучше, чем модель с равномерным распределением. Этот вывод сделан из того, что резкий скачок MSE_{Phys} происходит чуть раньше. Однако данный скачок наблюдается у модели с равномерной случайной последовательностью еще раньше, из чего следует сделать вывод, что данная модель обучается медленнее. Скорее всего это происходит из-за того, что плотность точек на отрезке довольно велика и данная последовательность точек не сильно отличается от равномерной.

5. Выводы

Из представленных результатов можно составить список-топ последовательностей для обученных моделей (чем выше последовательность по списку, тем больше она пригодна для обучения модели).

- Случайное равномерное распределение
- Квазислучайные последовательности
- Равномерное распределение
- Последовательности, меняющиеся динамически в ходе обучения

Тем не менее, следует высказать предположение, что реализация с адаптивным шагом может сильно превзойти по эффективности остальные методы, если у выбранной последовательности точек будет более высокая дисперсия.

Квазислучайная последовательность может лучше отработать для менее плотной упаковки точек.

Метод PINN имеет свои достоинства и недостатки. Среди достоинств стоит выделить:

- Непрерывность получаемой функции.
- Ее дифференцируемость.

Среди недостатков:

- Долгое время обучения (В среднем, для модели на 7.0×10^4 эпох при обучении на процессоре уходило ≈ 1.5 часа).
- Более высокая ошибка относительно времени обучения в сравнении с разностными схемами.

Список литературы и интернет-ресурсов

- [1] April 25, 2018 by Martin Roberts The Unreasonable Effectiveness of Quasirandom Sequences.— <https://extremelearning.com.au/unreasonable-effectiveness-of-quasirandom-sequences/>
- [2] Lagaris I.E., Likas A., Fotiadis D.I. Artificial neural networks for solving ordinary and partial differential equations //IEEE transactions on neural networks. — 1998. — Т. 9. — №. 5. — С. 987-1000.
- [3] Raissi M., Perdikaris P., Karniadakis G.E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations //Journal of Computational Physics. — 2019. — Т. 378. — С. 686-707.
- [4] TensorFlow documentation. — 2024. — https://www.tensorflow.org/api_docs
- [5] Ссылка на проект в github. — https://github.com/zge001/bessel_ai