

# Предсказание показаний фМРТ по видео, показанному человеку

Дорин Д. Д.<sup>1,2</sup>, Грабовой А. В.<sup>2</sup>

dorin.dd@phystech.edu

<sup>1</sup>Организация; <sup>2</sup>Организация

1

Исследуется задача прогнозирования показаний датчиков фМРТ по видеоряду, показанному человеку. Предложен метод аппроксимации показаний фМРТ по видеоряду на основе моделей типа Трансформер. Проанализирована зависимость между показаниями датчиков и восприятием внешнего мира человеком. Эффективность предложенного подхода демонстрируется на наборе данных, собранных у большой группы людей в процессе просмотра фильма.

**Ключевые слова:** фМРТ, видеоряд, Трансформер модель.

## 1 Введение

Человеческий мозг — один из самых интересных объектов исследования [8]. Внутренние записи человека являются редким и ценным источником информации о мозге. Поэтому изучение методов прогнозирования данных о функциональной активности коры головного мозга является актуальной темой в настоящее время.

Функциональная магнитно-резонансная томография, далее — фМРТ [9], — один из методов исследования активности головного мозга. фМРТ проводится с целью измерения гемодинамических реакций — изменений в потоке крови. Этот метод основывается на связи мозгового кровотока и активности нейронов. Когда область мозга активна, приток крови к этой области также увеличивается. фМРТ позволяет определить активацию определенной области головного мозга во время нормального функционирования под влиянием различных заданий, например, зрительных, когнитивных, моторных, речевых. В работе [2] собраны современные возможности фМРТ в нейровизуализации. Под нейровизуализацией понимается общее название нескольких методов, позволяющих визуализировать структуру, функции и биохимические характеристики мозга.

В работе [3] собран обширный набор данных, состоящий из видеорядов, просмотренных человеком, и соответствующих снимков фМРТ. Одна из проблем при работе с данными нейровизуализации — шум, вызванный движением головы, биением сердца, тепловыми эффектами и др. В работе [4] рассмотрены подходы к подготовке, предварительной обработке, шумоподавлению, направленные на устранение артефактов, вредных для распознавания образов, а также методы классификации данных нейровизуализации.

Наиболее известные методы обработки видео основаны на 3D свертках. Отличие 3D от 2D свертков заключается в одновременной работе с пространственной и временной частью информации. Существенный недостаток данных методов — сильное увеличение числа параметров модели и большие вычислительные затраты. В работе используется более современная архитектура — модель типа Трансформер. Впервые модель Трансформер была предложена в статье «Attention Is All You Need» Ashish Vaswani [5]. Архитектура активно применяется в области машинного перевода. А в 2022 году появилась работа [6] на тему адаптации архитектуры Трансформер для работы с видеорядами. Данная архитектура учитывает пространственно-временные зависимости и повышает скорость обучения за счет механизма внимания. Сама модель состоит из кодирующего компонента, декодирующего компонента и связи между ними. Каждый компонент состоит из стека энкодеров и

2

1. Добавьте организацию
2. Здесь не надо указывать название статьи и авторов. Достаточно ссылки

декодеров соответственно [1]. Входящая последовательность, поступающая в энкодер, сначала проходит через слой внимания, помогающий энкодеру посмотреть на другие слова во входящем объеме во время кодирования конкретного элемента. Выход слоя внимания отправляется в нейронную сеть прямого распространения. Аналогично устроен декодер, за исключением наличия еще одного слоя внимания, помогающего фокусироваться на релевантных элементах.

В данной работе предлагается метод аппроксимации показаний датчиков фМРТ по видеоряду. При получении метода использовались два основных предположения. Первая гипотеза заключается в существовании зависимости между результатами фМРТ и просматриваемым фильмом. Второе предположение заключается в том, что реакция мозга, фиксируемая фМРТ, на информацию, поступающую от органов зрения, происходит не мгновенно, а с некоторой задержкой [7]. Полученная в ходе экспериментов корреляционная картина между данными в выборке подтверждает зависимость между показаниями фМРТ и восприятием внешнего мира человеком.

Проверка метода проводится на выборке, представленной в работе [3]. Набор данных включает в себя в себя записи фМРТ 30 участников в возрасте от 7 до 47 лет во время выполнения одинаковой задачи и записи внутричерепной электроэнцефалографии 51 участника в возрасте от 5 до 55 лет.

## 2 Постановка задачи

Пусть  $\Omega$  — видеоряд,  $\nu$  — частота кадров,  $t$  — продолжительность видеоряда:

$$\Omega = (\omega_1, \dots, \omega_{\nu \cdot t}), \quad (1)$$

где  $\omega \in \mathbb{R}^{W_\omega \times H_\omega \times C_\omega}$  — изображение,  $W_\omega$  — ширина изображения,  $H_\omega$  — высота изображения и  $C_\omega$  — число каналов.

Введем также  $\mathcal{S}$  — последовательность фМРТ снимков,  $\mu$  — частота снимков:

$$\mathcal{S} = (s_1, \dots, s_{\mu \cdot t}), \quad (2)$$

где  $s \in \mathbb{R}^{X_s \times Y_s \times Z_s}$  — фМРТ снимок,  $x_s$ ,  $Y_s$ ,  $Z_s$  — размерность одного измерения.

Также считаем, что известно несколько дополнительных измерений фМРТ  $\mathcal{S}_0$  того же испытуемого. Необходимо построить отображение  $f$ :

$$f(\omega_1, \dots, \omega_{k_i - \nu \cdot \Delta t}, \mathcal{S}_0) = s_i, \quad (3)$$

которое учитывает задержку  $\Delta t$ , между фМРТ картиной и моментом получения информации зрительными органами. Здесь

$$k_i = \frac{\nu \cdot i}{\mu} \quad (4)$$

номер изображения в момент времени  $i$ -го снимка фМРТ.

## 3 Вычислительный эксперимент

### 3.1 Цель эксперимента

Построить метод аппроксимации снимка фМРТ по видеоряду и нескольким дополнительным измерениям фМРТ того же испытуемого, который учитывает задержку  $\Delta t$  между реакцией мозга и моментом получения информации зрительными органами.

3. Настоящее время

4. Очень спорное обозначение,  $\omega$  будет путаться с  $w$ , которые у вас дальше выступают в качестве параметров модели. Нужно обсудить с Андреем и по возможности поменять, иначе рецензент, редактор и читатели запутаются

### 3.2 Набор данных для базового эксперимента

Обучение модели проводится на наборе данных четвертого испытуемого, представленном в работе [3]. Для каждого фиксированного гиперпараметра  $\Delta t$  выборка состоит из  $641 - \mu \cdot \Delta t$  снимков фМРТ и набора соответствующих изображений из видеоряда. Номера изображений определяются по номеру снимка фМРТ соответственно формуле:

$$k_i = \nu \cdot \Delta t, \quad (5)$$

где  $k_i$  определен по формуле (4),  $i$  — номер изображения соответственно.

### 3.3 Базовая модель

Будем работать в предположении, что каждый снимок фМРТ зависит только от одного изображения. Тогда отображение имеет вид:

$$f(\omega_{k_i - \nu \cdot \Delta t}) = s^i, \quad i = 1, \dots, 641 - \mu \cdot \Delta t, \quad (6)$$

где  $s^i \in \mathbb{R}^{40 \times 64 \times 64}$  — тензор снимка фМРТ под номером  $i$ . Обозначим  $s_{ijk}$  компоненту тензора снимка  $s$ . Каждое изображения  $\omega$  из обучающей выборки предобработано с помощью ResNet152 без последнего линейного слоя и представляет собой вектор  $x$  признаков размерности  $d = 2048$ :

$$x_i = [x_1^i, \dots, x_d^i]^T \in \mathbb{R}^d, \quad i = 1, \dots, 9750.$$

Для восстановления значения в каждом вокселе по всему признаковому вектору изображения используется линейная модель с вектором параметров:

$$W_{ijk} = [w_1^{ijk}, \dots, w_d^{ijk}]^T \in \mathbb{R}^d : \\ f_{ijk}(x, W_{ijk}) = \langle x, W_{ijk} \rangle. \quad (7)$$

Для модели  $f_{ijk}$  с соответствующим ей вектором параметров  $W_{ijk} \in \mathbb{R}^d$  определим квадратичную функцию потерь:

$$\mathcal{L}_{ijk}(W_{ijk}, \Delta t) = \sum_{l=1}^{641 - \mu \cdot \Delta t} (f_{ijk}(x_l, W_{ijk}) - s_{ijk}^l)^2. \quad (8)$$

Требуется минимизировать функцию потерь  $\mathcal{L}_{ijk}(W_{ijk}, \Delta t)$  при фиксированном гиперпараметре  $\Delta t$ :

$$\hat{W}_{ijk} = \arg \min_{W_{ijk}} \mathcal{L}_{ijk}(W_{ijk}, \Delta t). \quad (9)$$

Одним из параметров модели является коэффициент сжатия воксельных снимков фМРТ. Рассматриваются коэффициенты сжатия 1, 2, 4 и 8. Для сжатия используется MaxPool3d с целью уменьшения числа подбираемых весов и ускорения процесса обучения соответственно. Метрикой оценки качества алгоритма является среднеквадратичная ошибка. Усреднение берется по всем вокселям снимков фМРТ и по всем снимкам в тестовой выборке.

### 3.4 Результаты эксперимента

Ниже приведены результаты работы модели при варьировании гиперпараметра  $\Delta t$  с фиксированным коэффициентом сжатия coef = 4.

6. Эксперимент должен идти после описания моделей, не до, и не в месте с моделями

7. Поскольку это будет часть описания модели, то конкретные числа здесь не нужны

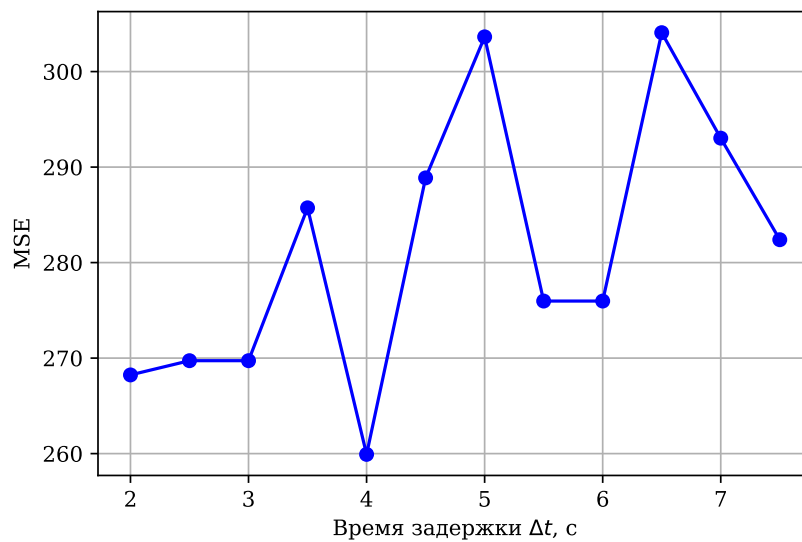
Заменить на константы

8. Аналогично

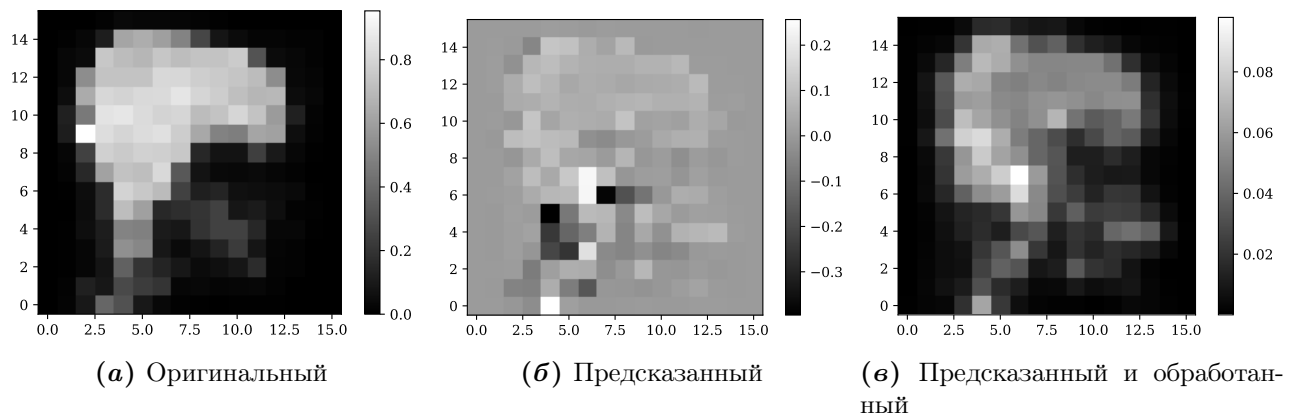
9. Ввести коэффициент раньше, где введен пуллинг.

Вообще по тексту эксперимента замечание аналогичное замечанию по тексту

Киселева: нужно добиться того, чтобы по тексту вашей статьи можно было бы полностью восстановить и повторить эксперимент



**Рис. 1** Зависимость среднеквадратичной ошибки на тесте от задержки  $\Delta t$



**Рис. 2** Срез снимка фМРТ из тестовой выборки

По приведенному графику 1 был выбран оптимальный гиперпараметр  $\Delta t_{opt} = 4$ . В качестве демонстрации работы при оптимальном гиперпараметре ниже приведены срезы оригинального и предсказанного воксельного снимка фМРТ из тестовой выборки.

На рисунке 2.(б) демонстрируется основная проблема модели — выбросы, которые портят контрастность изображения. Перед применением алгоритма была проведена нормализация данных фМРТ. Поэтому предсказанные значения, не попадающие на отрезок  $[0, 1]$ , считаются антифизичными. Срез снимка после обработки антифизичных значений в вокселях и применения фильтра Гаусса приведен на рисунке 2.(в).

**Таблица 1** MSE при параметрах модели  $\Delta t_{opt} = 4$ , coef = 4.

MSE при обучении	MSE на тесте
253.02	259.92

Полученные результаты подтверждают наличие корреляции между снимками фМРТ и изображениями из видеоряда. Однако ошибка модели довольно-таки существенна. В первую очередь это связано с тем, что модель предсказывает значение в каждом вокселе снимка независимо от других. Отсюда и вытекает проблема с выбросами. Также стоит попробовать использовать более сложную модель, линейная аппроксимация не даёт ожидаемый результат.

### 3.5 Основная модель

Будем работать в предположении, что каждый снимок фМРТ зависит только от одного изображения и предыдущего снимка фМРТ. То есть будем предсказывать разницу между фМРТ показаниями, что позволит учесть временную зависимость между снимками. Тогда отображение имеет вид:

$$f(\omega_{k_i-\nu \cdot \Delta t}) = \delta^i, \quad i = 2, \dots, 641 - \mu \cdot \Delta t, \quad (10)$$

где  $\delta^i = \mathbf{s}^i - \mathbf{s}^{i-1}$ ,  $\mathbf{s}^k \in \mathbb{R}^{40 \times 64 \times 64}$  — тензор снимка фМРТ под номером  $k$ . Обозначим  $\delta_{ijk}$  компоненту разницы между тензорами снимков  $\delta$ . Каждое изображения  $\omega$  из обучающей выборки предобработано с помощью ResNet152 без последнего линейного слоя и представляет собой вектор  $\mathbf{x}$  признаков размерности  $d = 2048$ :

$$\mathbf{x}_i = [x_1^i, \dots, x_d^i]^T \in \mathbb{R}^d, \quad i = 1, \dots, 9750.$$

Для восстановления разности значений в каждом вокселе с течением времени по всему признаковому вектору изображения используется линейная модель. Вектор параметров имеет вид:

$$\begin{aligned} \mathbf{w}_{ijk} &= [w_{ijk}^1, \dots, w_{ijk}^d]^T \in \mathbb{R}^d : \\ f_{ijk}(\mathbf{x}, \mathbf{w}_{ijk}) &= \langle \mathbf{x}, \mathbf{w}_{ijk} \rangle. \end{aligned} \quad (11)$$

Для каждого вокселя в снимке задана обучающая выборка:

$$\mathcal{D}_{ijk} = (\mathbf{x}_l, \delta_{ijk}^l)_{l=2}^{641-\mu \cdot \Delta t} \quad (12)$$

Выборка делится на тренировочную и тестовую в соотношении 7 : 3. Обозначим число объектов тренировочной выборки  $N$ .

Для модели  $f_{ijk}$ , с соответствующей ей тренировочной выборкой, определим квадратичную функцию потерь с  $L_2$  регуляризатором:

$$\mathcal{L}_{ijk}(\mathbf{w}_{ijk}, \Delta t, \alpha) = \frac{1}{2} \sum_{l=2}^{N+1} (f_{ijk}(\mathbf{x}_l, \mathbf{w}_{ijk}) - \delta_{ijk}^l)^2 + \frac{\alpha}{2} \|\mathbf{w}_{ijk}\|_2^2. \quad (13)$$

Требуется минимизировать функцию потерь  $\mathcal{L}_{ijk}(\mathbf{w}_{ijk}, \Delta t, \alpha)$  при фиксированных гиперпараметрах  $\Delta t$  и  $\alpha$ :

$$\hat{\mathbf{w}}_{ijk} = \arg \min_{\mathbf{w}_{ijk}} \mathcal{L}_{ijk}(\mathbf{w}_{ijk}, \Delta t, \alpha). \quad (14)$$

Определим матрицу объектов тренировочной выборки

$$\mathbf{X} = [\mathbf{x}_2^T, \dots, \mathbf{x}_{N+1}^T]^T = [\mathbf{x}_j^i] \in \mathbb{R}^{N \times d} \quad (15)$$



и вектор, компонентами которого являются разности значений одного и того же вокселя в снимках тренировочной выборки,

$$\Delta_{ijk} = [\delta_{ijk}^2, \dots, \delta_{ijk}^{N+1}]^T \in \mathbb{R}^N. \quad (16)$$

Тогда решение можно записать в виде:

$$\hat{w}_{ijk} = (\mathbf{X}^T \mathbf{X} + \alpha \mathbf{I})^{-1} \mathbf{X}^T \Delta_{ijk}. \quad (17)$$

Веса, полученные согласно формуле (17) для каждого вокселя снимка, запишем в виде строк матрицы весов

$$\hat{\mathbf{W}} = [\hat{w}_1^T, \dots, \hat{w}_{X_S \cdot Y_S \cdot Z_S}^T]^T = [w_j^i] \in \mathbb{R}^{X_S \cdot Y_S \cdot Z_S \times d}. \quad (18)$$

Тогда предсказание разности значений вокселей соседних снимков в виде вектора  $\hat{\mathbf{V}}_{\delta}^l$  можно вычислить по следующей формуле:

$$\hat{\mathbf{V}}_{\delta}^l = \hat{\mathbf{W}} \cdot \mathbf{x}_l \in \mathbb{R}^{X_S \cdot Y_S \cdot Z_S}. \quad (19)$$

Произведем reshape  $\hat{\mathbf{V}}_{\delta}^l$  к тензору  $\hat{\delta}^l$ . Тогда предсказанный тензор снимка фМРТ  $\hat{s}^l$  под номером  $l$  может быть получен согласно формуле:

$$\hat{s}^l = \hat{\delta}^l + s^{l-1}, \quad (20)$$

где  $s^{l-1}$  оригинальный тензор снимка фМРТ под номером  $l - 1$ .

Одним из параметров модели является коэффициент сжатия воксельных снимков фМРТ. Рассматриваются коэффициенты сжатия 1, 2, 4 и 8. Для сжатия используется MaxPool3d с целью уменьшения числа подбираемых весов и ускорения процесса обучения соответственно. Метрикой оценки качества алгоритма является среднеквадратичная ошибка. Усреднение берется по всем вокселям снимков фМРТ и по всем снимкам в тестовой выборке.

## 4 Анализ ошибки

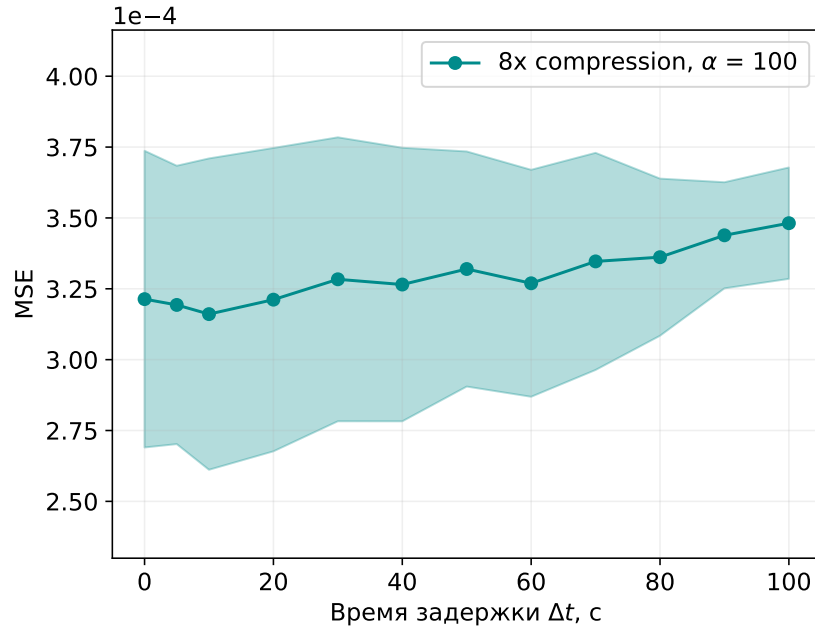
### 4.1 Варьирование гиперпараметров

На Рис.3 приведены результаты работы модели при варьировании гиперпараметра  $\Delta t$ , усредненные по всем испытуемым из выборки. Здесь фиксирован коэффициент сжатия coef = 8 и гиперпараметр регуляризации  $\alpha = 100$ . Фоном демонстрируется среднеквадратичное отклонение от среднего. График демонстрирует наличие зависимости результата предсказания от гиперпараметра  $\Delta t$ . Причем при нефизичной задержке больше десяти секунд ошибка начинает расти. Минимум достигается в пределах 4 — 10 секунд, что хорошо согласуется с нейробиологическими сведениями (todo: найти статью об этом).

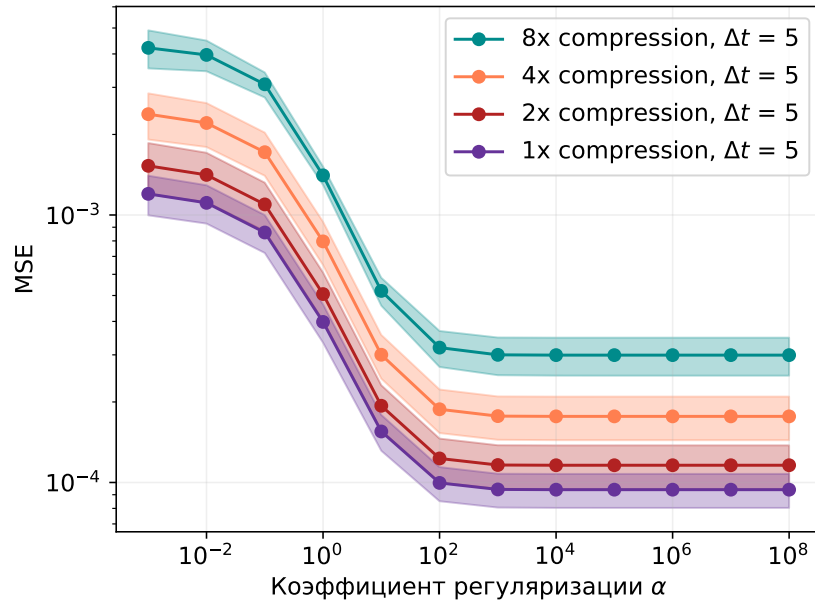
Проанализированна зависимость MSE от коэффициента регуляризации  $\alpha$ . Рассматриваются коэффициенты сжатия 1, 2, 4 и 8. Соответствующие графики приведены на Рис. 4. Из графиков видно, что отсутствие регуляризации ведет к переобучению модели. Оптимальное значение коэффициента  $\alpha \approx 100$ .

### 4.2 Демонстрация работы алгоритма

В качестве демонстрации работы алгоритма при оптимальных гиперпараметрах на Рис.5 приведены срезы оригинального и предсказанного воксельного снимка фМРТ из тестовой выборки.



**Рис. 3** Зависимость среднеквадратичной ошибки на тесте от задержки  $\Delta t$

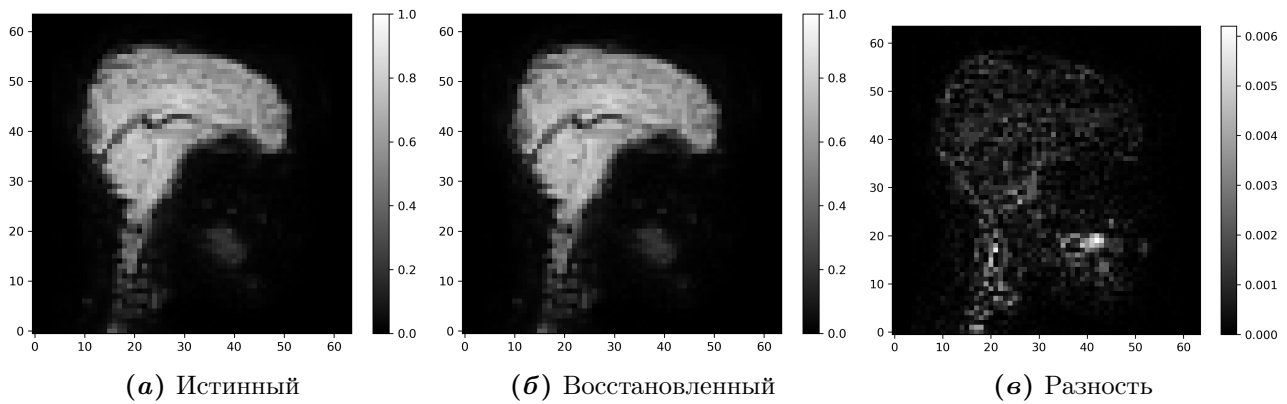


**Рис. 4** Зависимость среднеквадратичной ошибки на тесте от коэффициента регуляризации  $\alpha$

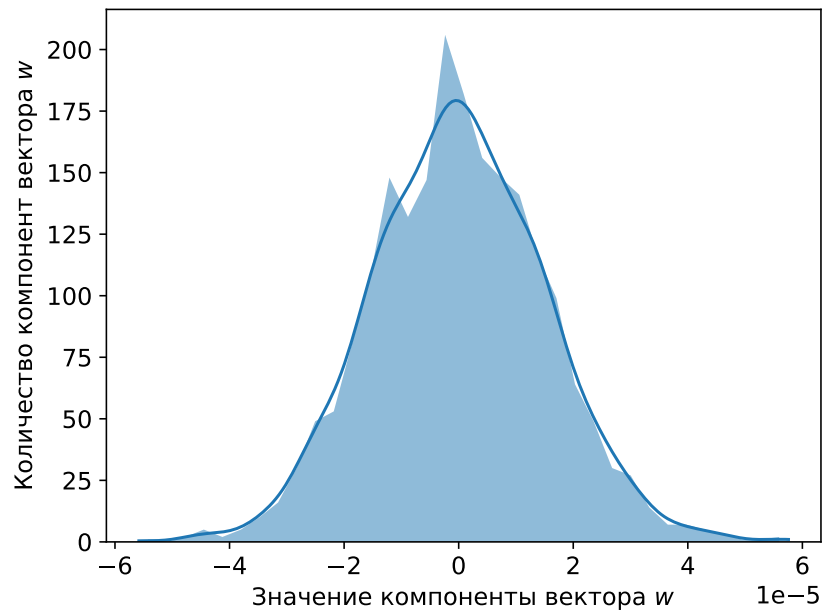
### 4.3 Корректность алгоритма

На Рис.6 приведено распределения значений компонент вектора весов модели. Для построения производилось усреднение по строкам матрицы весов  $\hat{W}$ , то есть усреднение по всем вокселям. График на Рис.6 хорошо аппроксимируется плотностью нормального распределения, что говорит о статистической значимости весов.

Экспериментально проверено, что модель улавливает общие для всех испытуемых зависимости между данными. Другими словами, восстановление снимка фМРТ одного человека можно производить, используя матрицу весов другого испытуемого. Для оценки



**Рис. 5** Срез снимка фМРТ из тестовой выборки



**Рис. 6** Распределение значений компонент вектора весов

качества работы алгоритма в данном эксперименте использовалась метрика MSE на тестовой выборке. Результаты представлены в Таблице 2.

Матрица весов	Истинная	Подмешанная
MSE	$1.02 \cdot 10^{-4}$	$1.05 \cdot 10^{-4}$

70

**Таблица 2** Проверка гипотезы инвариантности весов модели относительно человека

Ниже приведены результаты работы модели на случайном шуме. Первоначально модель обучена на оригинальных изображениях из видеоряда. Получена соответствующая матрица весов  $\hat{W}$ . К первому снимку последовательно прибавляются все восстановленные изменения значений вокселей. В результате имеем последний снимок последовательности. На Рис. 7 представлены срезы последнего истинного и восстановленного снимков из тестовой выборки. Для демонстрации работы на случайном шуме сгенерирована случайная выборка из векторов признакового описания изображения размера тестовой выборки.

10. Что такое подмешивание, как оно вводится и какие параметры ? (как пример, чего нет в тексте)

По шумовым данным и матрице весов  $\hat{W}$  получена последовательность изменений между соседними снимками фМРТ. Аналогично последнему снимку, полученному по оригинальным изображениям, предсказан последний снимок по шуму. На Рис. 8 представлены срезы последнего истинного и восстановленного по шуму снимков из тестовой выборки.

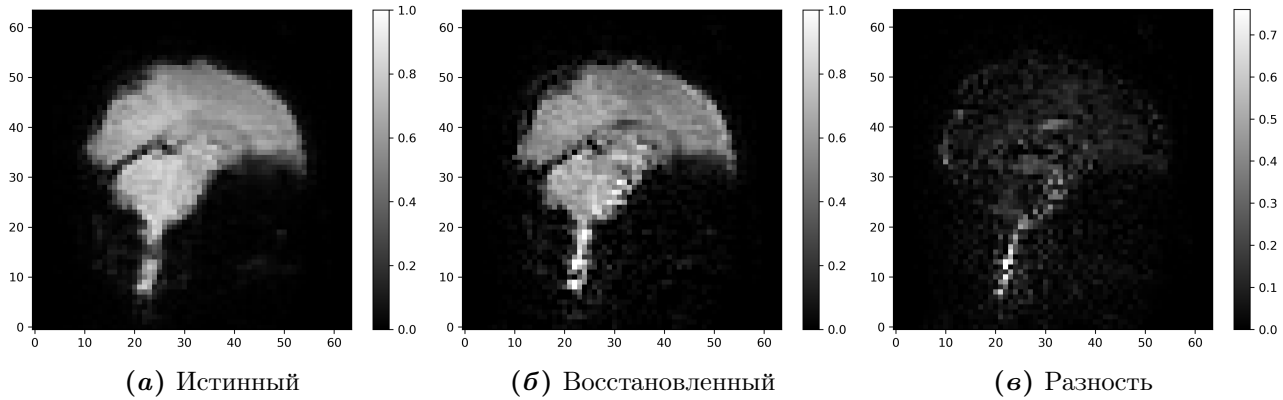


Рис. 7 Срез снимка фМРТ из тестовой выборки

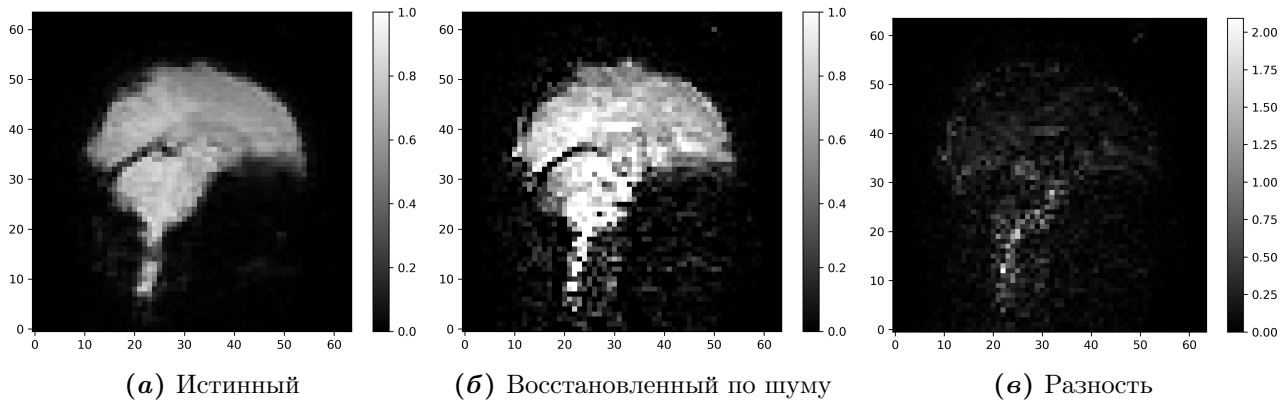


Рис. 8 Срез снимка фМРТ из тестовой выборки на случайном шуме

На приведенных изображениях можно видеть, что разность в случае шума существенно больше. В таблице 3 приведены среднеквадратичные ошибки. Ошибка на шуме на порядок больше, что подтверждает наличие корреляции между показаниями датчиков и изображениями из видеоряда.

Выборка	Истинная	Случайный шум
MSE	$2 \cdot 10^{-3}$	$10^{-1}$

11

Таблица 3 Качество работы метода на случайном шуме

## 5 Заключение

В работе построен базовый и основной метод аппроксимации показаний датчиков фМРТ по видеоряду, просматриваемому человеком. Результаты экспериментов подтверждают наличие корреляции между данными. Приведенные графики подтверждают пред-

11. см пункт 10

положение о наличии задержки между моментом получения информации зрительными органами и реакцией мозга на эту информацию. Демонстрируется работа алгоритмов на тестовых данных. Качество работы основного метода в разы превосходит качество работы базового алгоритма. Особенность основного метода заключается в учете временной зависимости между снимками за счет предсказания разности между соседними показаниями датчиков фМРТ. Это подтверждает временную взаимосвязь между фМРТ снимками.

## 6 \*

## Список литературы

- [1] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495, 2017.
- [2] A. A. Belyaevskaya, N. V. Meladze, M. A. Sharia, D. V. Ustyuzhanin, and M. H. Zashezova. MODERN POSSIBILITIES OF FUNCTIONAL MAGNETIC RESONANCE IMAGING IN NEUROIMAGING. *Medical Visualization*, (1):7–16, February 2018.
- [3] Julia Berezutskaya, Mariska J. Vansteensel, Erik J. Aarnoutse, Zachary V. Freudenburg, Giovanni Piantoni, Mariana P. Branco, and Nick F. Ramsey. Open multimodal iEEG-fMRI dataset from naturalistic stimulation with a short audiovisual film. *Scientific Data*, 9(1), March 2022.
- [4] Maxim Sharaev, Alexander Andreev, Alexey Artemov, Alexander Bernstein, Evgeny Burnaev, Ekaterina Kondratyeva, Svetlana Sushchinskaya, and Renat Akzhigitov. fmri: preprocessing, classification and pattern recognition, 2018.
- [5] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2017.
- [6] Shen Yan, Xuehan Xiong, Anurag Arnab, Zhichao Lu, Mi Zhang, Chen Sun, and Cordelia Schmid. Multiview transformers for video recognition, 2022.
- [7] Вячеслав Евгеньевич Демидов. *Как мы видим то, что видим*. Знание, 1979.
- [8] ТА Жумакова, ШО Рыспекова, ДД Жунистаев, НМ Чурукова, АМ Исаева, and ИО Алимкул. Тайны человеческого мозга. *Международный журнал прикладных и фундаментальных исследований*, (6-2):230–232, 2017.
- [9] ВЛ Ушаков, ВМ Верхлютов, ПА Соколов, МВ Ублинский, ВБ Стрелец, АЮ Аграфонов, АВ Петрайкин, and ТА Ахадов. Активация структур мозга по данным фМРТ при просмотре видеосюжетов и припоминании показанных действий. *Журнал высшей нервной деятельности им. ИП Павлова*, 61(5):553–564, 2011.