

Предсказание показаний фМРТ по видео, показанному человеку

Дорин Д. Д.^{1,2}, Грабовой А. В.²

dorin.dd@phystech.edu

¹MIPT; ²Intelligent systems

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 [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 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 используются 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 \mathcal{P} — видеоряд, ν — частота кадров, t — продолжительность видеоряда:

$$\mathcal{P} = (p_1, \dots, p_{\nu \cdot t}), \quad (1)$$

где $p \in \mathbb{R}^{W_p \times H_p \times C_p}$ — изображение, W_p — ширина изображения, H_p — высота изображения и C_p — число каналов.

Введем также \mathcal{S} — последовательность фМРТ снимков, μ — 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 :

$$f(p_1, \dots, p_{k_i - \nu \cdot \Delta t}, \mathcal{S}_0) = s^i, \quad (3)$$

которое учитывает задержку Δt , 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 Описание модели

Пусть имеется набор данных некоторого испытуемого, состоящий из L снимков фМРТ и соответствующего этим снимкам видеоряда. Для каждого фиксированного гиперпараметра Δt выборка состоит из $L_{\Delta t} = L - \mu \cdot \Delta t$ 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.

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:

$$\mathbf{f}(\mathbf{p}_{k_i - \nu \cdot \Delta t}) = \boldsymbol{\delta}^i, \quad i = 2, \dots, L_{\Delta t}, \quad (6)$$

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

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

где K общее число изображений в видеоряде.

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

$$\begin{aligned} \mathbf{w}_{ijk} &= [w_{ijk}^1, \dots, w_{ijk}^d]^T \in \mathbb{R}^d : \\ f_{ijk}(\mathbf{x}, \mathbf{w}_{ijk}) &= \langle \mathbf{x}, \mathbf{w}_{ijk} \rangle. \end{aligned} \quad (7)$$

Для каждого вокселя в снимке задана обучающая выборка:

$$\mathcal{D}_{ijk} = (\mathbf{x}_l, \delta_{ijk}^l)_{l=2}^{L_{\Delta t}} \quad (8)$$

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:

$$\mathcal{L}_{ijk}(\mathbf{w}_{ijk}, \Delta t, \alpha) = \frac{1}{2} \sum_{l=2}^{N+1} (f_{ijk}(\mathbf{x}_l, \mathbf{w}_{ijk}) - \delta_{ijk}^l)^2 + \frac{\alpha}{2} \|\mathbf{w}_{ijk}\|_2^2. \quad (9)$$

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

$$\hat{\mathbf{w}}_{ijk} = \arg \min_{\mathbf{w}_{ijk}} \mathcal{L}_{ijk}(\mathbf{w}_{ijk}, \Delta t, \alpha). \quad (10)$$

Определим матрицу объектов тренировочной выборки

$$\mathbf{X} = [\mathbf{x}_2^T, \dots, \mathbf{x}_{N+1}^T]^T = [\mathbf{x}_j^i] \in \mathbb{R}^{N \times d} \quad (11)$$

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

$$\boldsymbol{\Delta}_{ijk} = [\delta_{ijk}^2, \dots, \delta_{ijk}^{N+1}]^T \in \mathbb{R}^N. \quad (12)$$

Тогда решение можно записать в виде:

$$\hat{\mathbf{w}}_{ijk} = (\mathbf{X}^T \mathbf{X} + \alpha \mathbf{I})^{-1} \mathbf{X}^T \boldsymbol{\Delta}_{ijk}. \quad (13)$$

Веса, полученные согласно формуле (13) для каждого вокселя снимка, запишем в виде строк матрицы весов

$$\hat{W} = [\hat{w}_1^T, \dots, \hat{w}_{X_S \cdot Y_S \cdot Z_S}^T]^T = [w_j^i] \in \mathbb{R}^{X_S \cdot Y_S \cdot Z_S \times d}. \quad (14)$$

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

$$\hat{V}_\delta^l = \hat{W} \cdot x_l \in \mathbb{R}^{X_S \cdot Y_S \cdot Z_S}. \quad (15)$$

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

$$\hat{s}^l = \hat{\delta}^l + s^{l-1}, \quad (16)$$

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 Вычислительный эксперимент

4.1 Цель эксперимента

Проверить корректность предложенного метода, оценить качество работы метода на тестовых данных и проверить основные гипотезы.

4.2 Набор данных для эксперимента

Набор данных для эксперимента, представленный в работе [3], состоит из снимков фМРТ и видеорядов тридцати испытуемых. Для обучения модели выбирается набор данных одного из испытуемых.

Число снимков фМРТ в выборке $L = 641$. Для каждого фиксированного гиперпараметра Δt выборка состоит из $L_{\Delta t} = 641 - \mu \cdot \Delta t$ снимков фМРТ и набора соответствующих изображений из видеоряда. Каждый снимков фМРТ представляет собой тензор $s \in \mathbb{R}^{40 \times 64 \times 64}$

После предобработки изображений из видеоряда с помощью ResNet152 без последнего линейного слоя они представляет собой вектора признаков размерности $d = 2048$. Общее число изображений в видеоряде $K = 9750$.

4.3 Демонстрация работы алгоритма

В качестве демонстрации работы алгоритма на Рис.1 приведены срезы оригинального и предсказанного воксельного снимка фМРТ из тестовой выборки. Ошибка порядка 10^{-3} в отдельных вокселях говорит о хорошем качестве работы алгоритма.

5 Анализ ошибки

5.1 Варьирование гиперпараметров

На Рис.2 приведены результаты работы модели при варьировании гиперпараметра Δt , усредненные по всем испытуемым из выборки. Здесь фиксирован коэффициент сжатия

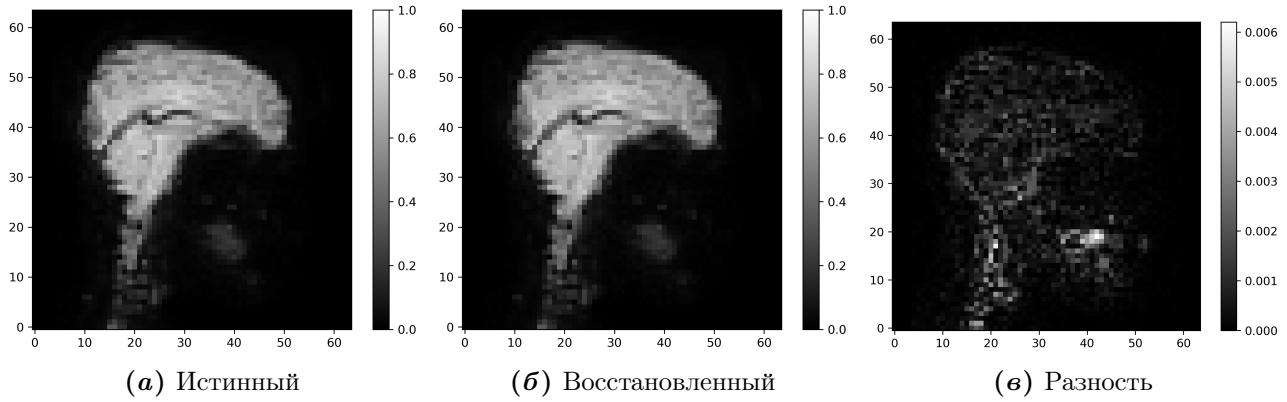
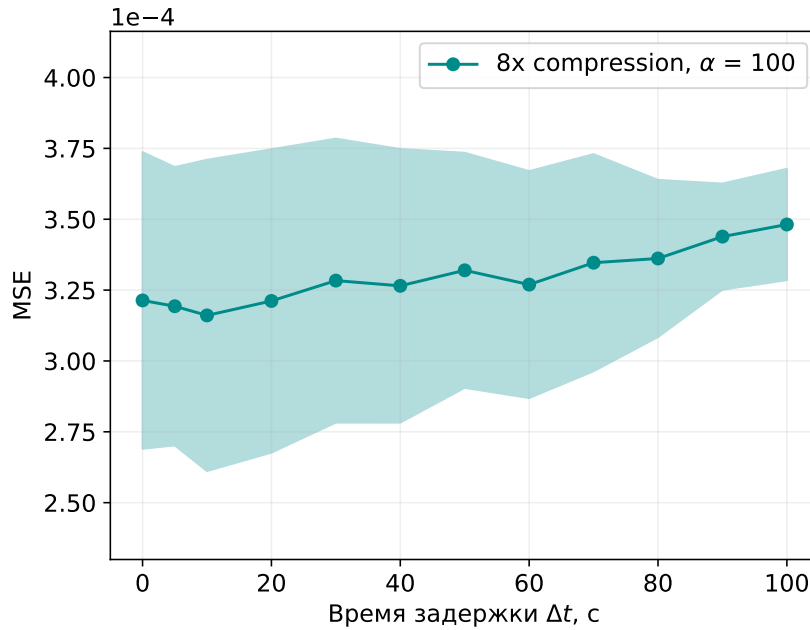


Рис. 1 Срез снимка фМРТ из тестовой выборки

Рис. 2 Зависимость среднеквадратичной ошибки на тесте от задержки Δt

$\text{coef} = 8$ и гиперпараметр регуляризации $\alpha = 100$. Фоном демонстрируется среднеквадратичное отклонение от среднего. График демонстрирует наличие зависимости результата предсказания от гиперпараметра Δt . 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 α . Рассматриваются коэффициенты сжатия 1, 2, 4 и 8. Соответствующие графики приведены на Рис. 3. Из графиков видно, что отсутствие регуляризации ведет к переобучению модели. Оптимальное значение коэффициента $\alpha \approx 100$.

5.2 Корректность алгоритма

На Рис.4 приведено распределения значений компонент вектора весов модели. Для построения производилось усреднение по строкам матрицы весов \hat{W} , то есть усреднение

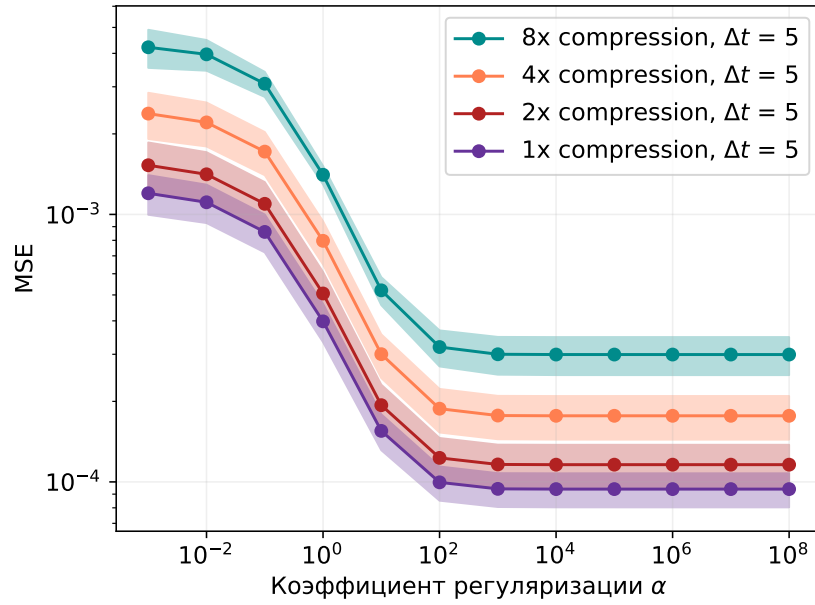


Рис. 3 Зависимость среднеквадратичной ошибки на тесте от коэффициента регуляризации α

по всем вокселям. График на Рис.4 хорошо аппроксимируется плотностью нормального распределения, что говорит о статистической значимости весов.

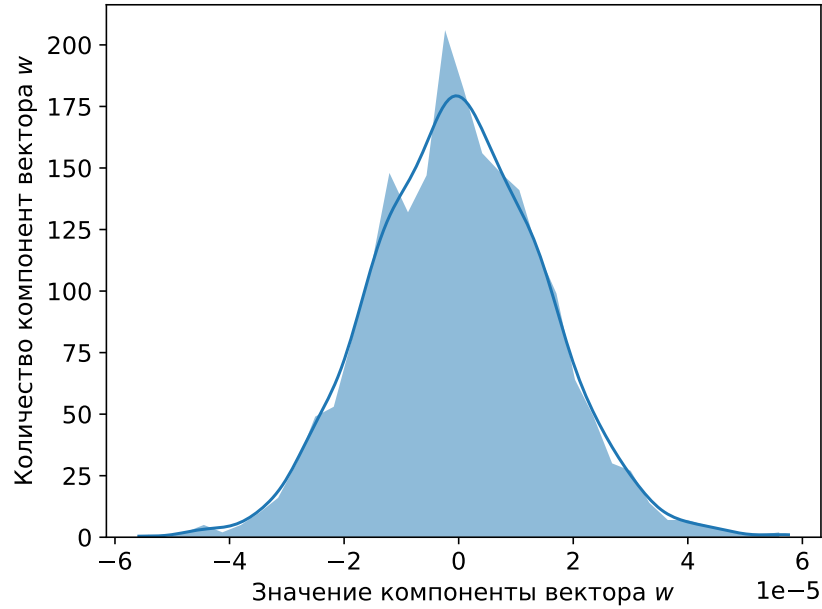


Рис. 4 Распределение значений компонент вектора весов

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 качества работы алгоритма в данном эксперименте использовалась метрика MSE на тестовой выборке. Результаты представлены в Таблице 1.

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

Таблица 1 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. 5 представлены срезы последнего истинного и восстановленного снимков из тестовой выборки. Для демонстрации работы на случайном шуме сгенерирована случайная выборка из векторов признаков описания изображения размера тестовой выборки. 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. 6 представлены срезы последнего истинного и восстановленного по шуму снимков из тестовой выборки.

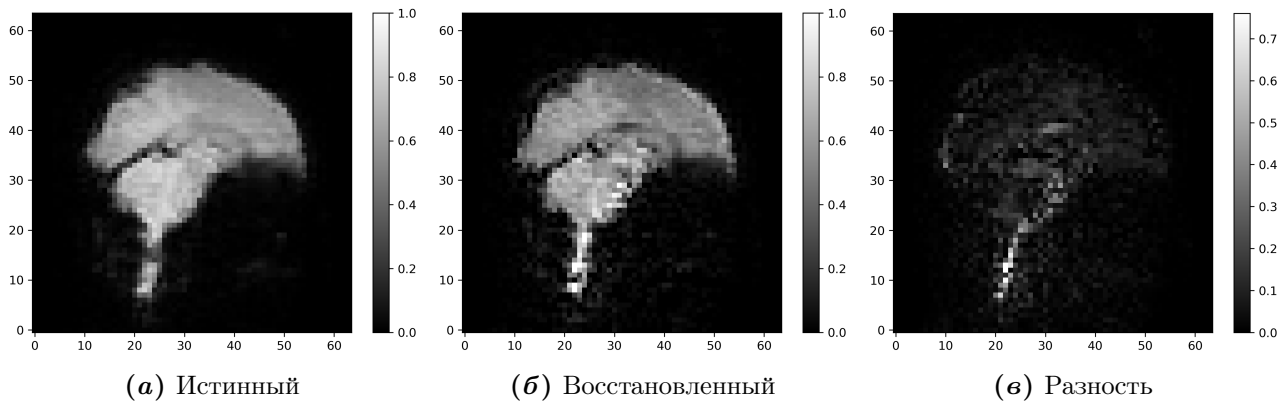


Рис. 5 Срез снимка фМРТ из тестовой выборки

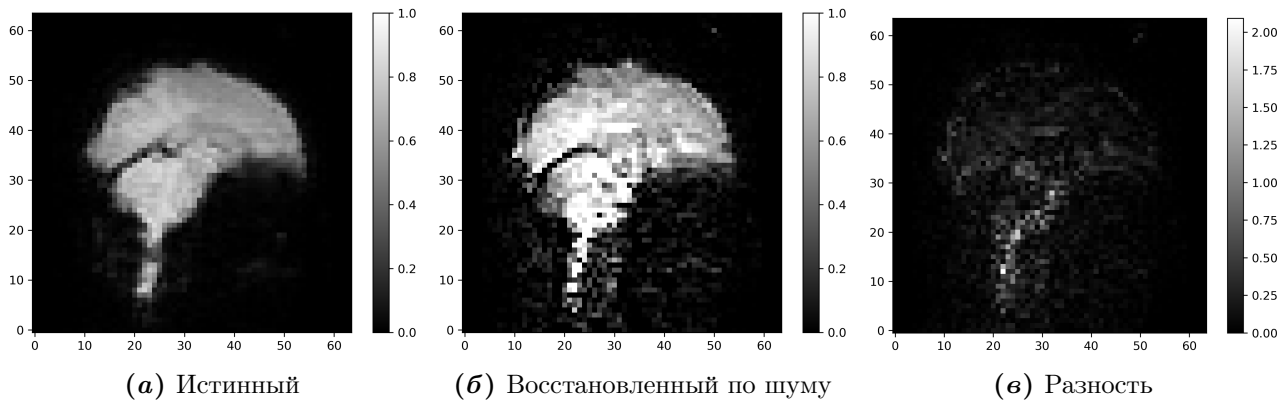


Рис. 6 Срез снимка фМРТ из тестовой выборки на случайном шуме

На приведенных изображениях можно видеть, что разность в случае шума существенно больше. В таблице 2 приведены среднеквадратичные ошибки. Ошибка на шуму 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}

Таблица 2 Качество работы метода на случайном шуме

6 Заключение

В работе построен метод аппроксимации показаний датчиков фМРТ по видеоряду, просматриваемому человеком. Результаты эксперимента со случайным шумом подтверждают наличие корреляции между данными. В работе экспериментально подтверждено, что веса модели инвариантны относительно человека. График распределения значений компонент весов модели хорошо аппроксимируется нормальным распределением, что говорит о статистической значимости весов. График, полученный при варьировании Δt , подтверждает предположение о наличии задержки между моментом получения информации зрительными органами и реакцией мозга на эту информацию.

7 *

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.