

---

# МОДЕЛИРОВАНИЕ ПОКАЗАНИЯ fMRI ПО ВИДЕО, ПОКАЗАННОМУ ЧЕЛОВЕКУ

---

Дорин Даниил  
dorin.dd@phystech.edu

Киселев Никита  
kiselev.ns@phystech.edu

2 марта 2023 г.

## АННОТАЦИЯ

В работе исследуется проблема восстановления зависимости между показаниями датчиков fMRI и восприятием внешнего мира человеком. Проводится анализ зависимости между последовательностью fMRI-снимков и видеорядом, просматриваемым человеком. На основе исследования зависимости предлагается метод аппроксимации показаний fMRI по просматриваемому видеоряду. Для анализа предложенного метода проводится вычислительный эксперимент на выборке [2], полученной при обследовании большого числа испытуемых.

**Ключевые слова:** нейровизуализация · fMRI · видеоряд · зависимость между данными · Transformer

## 1 Введение

Сококупность методов, позволяющих визуализировать структуру и функции человеческого мозга, называется *нейровизуализацией*. Эти методы используются для изучения мозга, а также для обнаружения заболеваний и психических расстройств. Примерами таких являются fMRI и iEEG [2]. В настоящей работе рассматривается метод fMRI.

*Функциональная магнитно-резонансная томография* или *fMPT* (англ. *fMRI*) является разновидностью магнитно-резонансной томографии и основана на изменениях в токе крови, вызванных нейронной активностью мозга [4]. Эти изменения происходят не моментально, а с некоторой задержкой. Она возникает из-за того, что сосудистая система достаточно долго реагирует на потребность мозга в глюкозе [5]. Изображения, получаемые с помощью fMRI, показывают, какие участки мозга активированы при выполнении испытуемым определенных заданий.

Метод fMRI играет большую роль в нейровизуализации, однако имеет ряд важных ограничений. В работах [6, 7] рассматриваются временное и пространственное разрешение fMRI. Временное разрешение является существенным недостатком данного метода.

Другой недостаток fMRI — неизбежно возникающие шумы, связанные с движением объекта в сканере, сердцебиением и дыханием человека, тепловыми флуктуациями самого прибора и т. д. В работе [8] предлагаются методы подавления вышеперечисленных шумов на основе графов и демонстрируется их эффективность в задаче выявления эпилепсии и депрессии.

Обобщением уже естественных для обработки изображений 2D сверток в CNN являются 3D свертки. Они агрегируют информацию как по времени, так и по пространству. Однако это приводит к сильному увеличению количества используемых параметров. В настоящей работе используется наиболее современная архитектура — Transformer. Впервые она была предложена в статье [9]. Не так давно появилась адаптация архитектуры Transformer для работы с видео [10]. Данная архитектура состоит из кодировщика и декодировщика, каждый из которых в свою очередь состоит из отдельных слоев. Использование механизма Attention [9] позволяет значительно повысить качество работы модели.

Настоящая работа посвящена восстановлению зависимости между fMRI-снимками и видеорядом. Используется предположение, что такая зависимость существует. Кроме того, предполагается, что между снимком и видеорядом есть постоянная задержка во времени [5]. На основе анализа зависимости предлагается метод аппроксимации показаний fMRI по просматриваемому видеоряду.

Данные, на которых проводятся проверка гипотезы зависимости и демонстрация работы построенного метода, представлены в работе [2]. Этот набор данных был получен при fMRI-обследовании большой группы людей. Им предлагалось выполнить одно и то же задание — просмотреть короткий аудиовизуальный фильм. Для него в рассматриваемой работе были сгенерированы аннотации, содержащие в том числе информацию о времени появления и исчезновения отдельных слов, объектов и персонажей. Методы аудио- и видеоаннотирования подробно излагаются в [3] и [1].

## 2 Постановка задачи

Пусть задана частота кадров  $\nu$  и продолжительность  $t$  видеоряда. Задан видеоряд

$$\mathbf{V} = \{\mathbf{s}_1, \dots, \mathbf{s}_{\nu t}\}, \quad (1)$$

где  $\mathbf{s}_i \in \mathbb{R}^{W_{\mathbf{V}} \times H_{\mathbf{V}} \times C_{\mathbf{V}}}$ ,  $i = 1, \dots, \nu t$  — изображение, а  $W_{\mathbf{V}}$ ,  $H_{\mathbf{V}}$  и  $C_{\mathbf{V}}$  — его ширина, высота и число каналов соответственно. Задана последовательность fMRI-снимков

$$\mathbf{F} = \{\mathbf{f}_1, \dots, \mathbf{f}_n\}, \quad (2)$$

где  $\mathbf{f}_i \in \mathbb{R}^{W_{\mathbf{F}} \times H_{\mathbf{F}} \times D_{\mathbf{F}}}$ ,  $i = 1, \dots, n$  — томографическое изображение с измерениями  $W_{\mathbf{F}}$ ,  $H_{\mathbf{F}}$  и  $D_{\mathbf{F}}$ .

Задача состоит в построении отображения, которое бы учитывало задержку  $\Delta t$  между fMRI-снимком и видеорядом, а также предыдущие томографические показания. Формально, необходимо найти такое отображение  $g$ , что

$$g(\mathbf{s}_1, \dots, \mathbf{s}_{k_i - \nu \Delta t}; \mathbf{f}_1, \dots, \mathbf{f}_{i-1}) = \mathbf{f}_i, \quad i = 1, \dots, n. \quad (3)$$

### 3 Вычислительный эксперимент

### 4 Анализ ошибки

### 5 Заключение

### Список литературы

- [1] Julia Berezutskaya, Zachary V. Freudenburg, Luca Ambrogioni, Umut Güçlü, Marcel A. J. van Gerven, and Nick F. Ramsey. Cortical network responses map onto data-driven features that capture visual semantics of movie fragments. *Scientific Reports*, 10(1), July 2020.
- [2] 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.
- [3] Paul Boersma and David Weenink. Praat: doing phonetics by computer [computer program]. version 6.0. 37. Retrieved February, 3:2018, 2018.
- [4] Gary H. Glover. Overview of functional magnetic resonance imaging. *Neurosurgery Clinics of North America*, 22(2):133–139, April 2011.
- [5] Nikos K. Logothetis. The underpinnings of the BOLD functional magnetic resonance imaging signal. *The Journal of Neuroscience*, 23(10):3963–3971, May 2003.
- [6] Nikos K Logothetis. What we can do and what we cannot do with fmri. *Nature*, 453(7197):869–878, 2008.
- [7] Ravi S Menon and Seong-Gi Kim. Spatial and temporal limits in cognitive neuroimaging with fmri. *Trends in cognitive sciences*, 3(6):207–216, 1999.
- [8] Maxim Sharaev, Alexander Andreev, Alexey Artemov, Alexander Bernstein, Evgeny Burnaev, Ekaterina Kondratyeva, Svetlana Sushchinskaya, and Renat Akzhigitov. fmri: preprocessing, classification and pattern recognition, 2018.
- [9] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2017.
- [10] Shen Yan, Xuehan Xiong, Anurag Arnab, Zhichao Lu, Mi Zhang, Chen Sun, and Cordelia Schmid. Multiview transformers for video recognition, 2022.