MIPT [1]
ANTIPLAGIAT [2]
INNOPOLIS UNIVERSITY [3]

Spatio-Temporal fMRI Analysis in Visual Stimuli Decoding: Linear Model Forecasting & Voxel Weighing

Daniil Dorin[1,2], Nikita Kiselev[1,3], Andrey Grabovoy[1,2]

GitHub

Forecasting fMRI from visual stimuli

Problem statement

Given the video sequence frequency rate $\nu \in \mathbb{R}$ and duration $t \in \mathbb{R}$. Consider a video sequence

$$\mathbf{P} = [\mathbf{p}_1, \dots, \mathbf{p}_{\nu t}], \quad \mathbf{p}_\ell \in \mathbb{R}^{W \times H \times C}$$

Given the fMRI sequence frequency rate $\mu \in \mathbb{R}$. Corresponding time series is

$$\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_{\mu t}], \quad \mathbf{x}_\ell \in \mathbb{R}^{X \times Y \times Z}$$

Problem is to construct a mapping $\mathbf{g}$, accounting for BOLD response delay Δt between fMRI image and corresponding video frame.

$$\mathbf{g}(\mathbf{p}_1, \dots, \mathbf{p}_{k_\ell - \nu\Delta t}; \mathbf{x}_1, \dots, \mathbf{x}_{\ell-1}) = \mathbf{x}_\ell, \quad \ell = 1, \dots, \mu t, \quad k_\ell = \frac{\ell \cdot \nu}{\mu}$$

Markovian Setup and Decomposition

Suppose the Markov property is satisfied for the fMRI images sequence. Then the corresponding mapping is written in the following form:

$$\mathbf{g}(\mathbf{p}_{k_\ell - \nu\Delta t}) = \mathbf{x}_\ell - \mathbf{x}_{\ell-1} = \boldsymbol{\delta}_\ell, \ \ell = 2, \dots, \mu t.$$

Besides, the mapping is represented as a composition of the other two:

$$\mathbf{g} = \boldsymbol{\varphi} \circ \boldsymbol{\psi},$$

$\boldsymbol{\psi} : \mathbf{P} \to \mathbb{R}^d$ is an image vectorization, ResNet152

$\boldsymbol{\varphi} : \mathbb{R}^d \to \mathbf{X}$ is a target mapping.

Optimization Problem and Solution[1]

Applying pretrained image encoder, we obtain embeddings $\mathbf{z}_1, \dots, \mathbf{z}_N \in \mathbb{R}^d$.

For each voxel a sample is given $\mathfrak{D}_{ijk} = \{(\mathbf{z}_\ell, \delta^\ell_{ijk}) \mid \ell = 2, \dots, N\}$.

A linear model is used with a vector of parameters $\mathbf{w}_{ijk} = [w_1^{ijk}, \dots, w_d^{ijk}]^\top \in \mathbb{R}^d$

We define a quadratic loss function with regularization:

$$\mathcal{L}_{ijk}(\mathbf{w}_{ijk}) = \sum_{\ell=2}^{N} \left(f_{ijk}(\mathbf{z}_\ell, \mathbf{w}_{ijk}) - \delta^\ell_{ijk} \right)^2 + \alpha \|\mathbf{w}_{ijk}\|_2^2$$

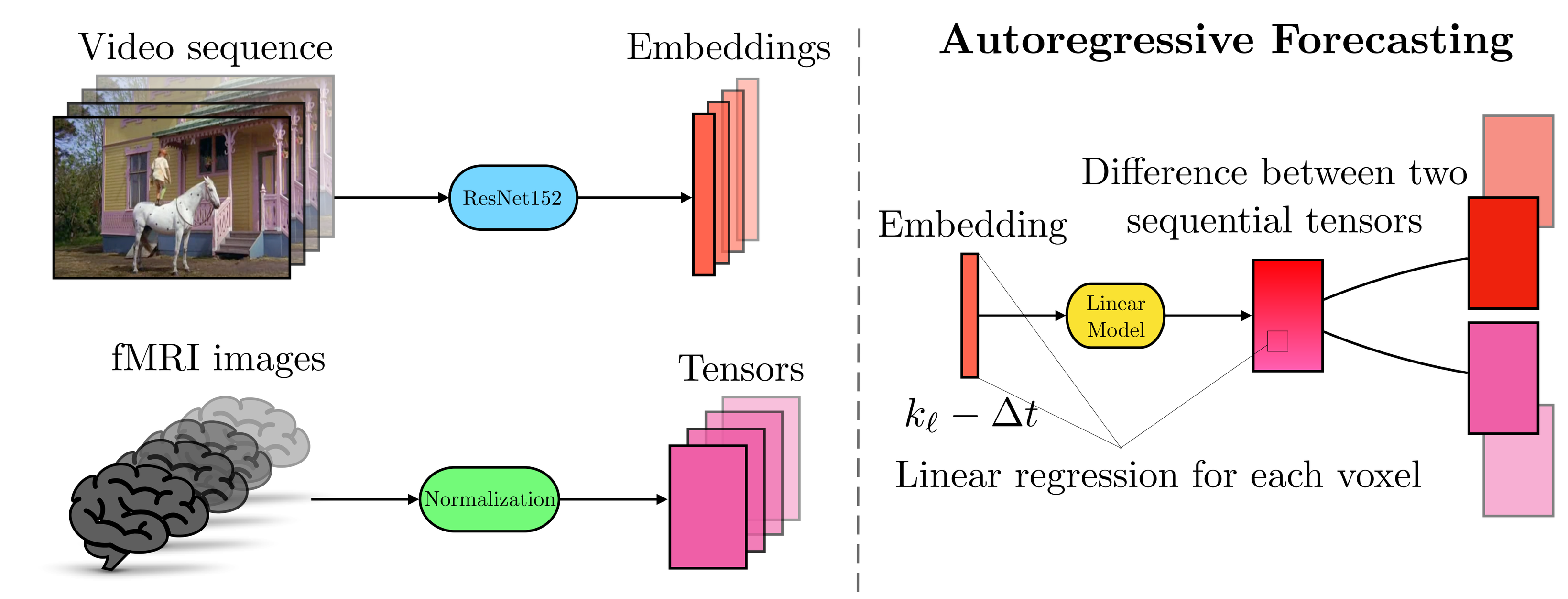
Least squares solution is written in the folowing form

$$\hat{\mathbf{w}}_{ijk} = (\mathbf{Z}^\top \mathbf{Z} + \alpha \mathbf{I})^{-1} \mathbf{Z}^\top \boldsymbol{\Delta}_{ijk}, \quad \boldsymbol{\Delta}_{ijk} = [\delta^2_{ijk}, \dots, \delta^N_{ijk}]^\top \in \mathbb{R}^{N-1}$$

$$\mathbf{Z} = [\mathbf{z}_2, \dots, \mathbf{z}_N]^\top = [z^i_j] \in \mathbb{R}^{(N-1) \times d}$$

Thus, a forecast is $\mathrm{vec}(\hat{\mathbf{x}}_\ell) = \mathrm{vec}(\mathbf{x}_{\ell-1}) + \mathrm{vec}(\hat{\boldsymbol{\delta}}_\ell) = \mathrm{vec}(\mathbf{x}_{\ell-1}) + \hat{\mathbf{W}} \mathbf{z}_\ell$

Method Overview



[1] D.Dorin, N.Kiselev, A.Grabovoy, V.Strijov. Forecasting fMRI Images From Video Sequences: Linear Model Analysis. Accepted to the Health Information Science and Systems journal.

fMRI sequence classification

Problem statement

Consider a sample of fMRI sequences with equal time duration and frequency:

$$\mathbf{X} = \{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N\}$$

$$\mathbf{X}_i = [\mathbf{x}_1^i, \mathbf{x}_2^i, \dots, \mathbf{x}_T^i]$$

where $\mathbf{x}_t^i \in \mathbb{R}^{X \times Y \times Z}$ is a tensor of fMRI at moment t and for i-th sample. For each sample we have a corresponding label, thus we consider a dataset

$$\mathfrak{D} = \{(y_i, \mathbf{X}_i) \mid i = 1, \dots, N\}.$$

One has to construct a classification model g, accounting for the specific spatio-temporal fMRI characteristics:

$$g : \mathbf{X} \to \{1, \dots, C\}.$$

Voxel Weighing via Cross-Correlation

Given the fMRI sequence $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_\tau]$, $\mathbf{x}_t \in \mathbb{R}^{X \times Y \times Z}$ with frequency μ, and stimuli time series $\mathbf{s} = [s_1, \dots, s_\tau]$, $s_t \in \{0, 1\}$.

1. Compression with 3D Average Pooling using kernel size k_s: $\mathbf{X} \to \mathbf{X}'$.
2. Time series normalization: $\mathbf{X}' \to \hat{\mathbf{X}}'$, $\mathbf{s} \to \hat{\mathbf{s}}$.
3. Cross-correlation calculation:
$$c_{i,j,k}(p) = \left(\hat{\mathbf{s}} * \mathbf{v}^{i,j,k} \right)(p) = \frac{1}{\tau - 1} \sum_{t=1}^{\tau - p} \hat{s}_t \cdot v^{i,j,k}_{t+p}, \quad p = 0, \dots, \tau - 1.$$
4. Choosing optimal value $p = \lfloor \mu \Delta t \rfloor$ and h the most correlated positions.
5. Constructing an activity map $\mathcal{M}_c \in \{0, 1\}^{X/k_s \times Y/k_s \times Z/k_s}$.
6. Dimensionality restoration: $\mathcal{M}_c \to \mathcal{M} \in \{0, 1\}^{X \times Y \times Z}$.

Check Statistical Significance

Classification method[2] and Riemannian manifold

Model for classification is represented as composition of the other three:

$$g = \boldsymbol{\varphi} \circ \boldsymbol{\psi} \circ \mathcal{A},$$

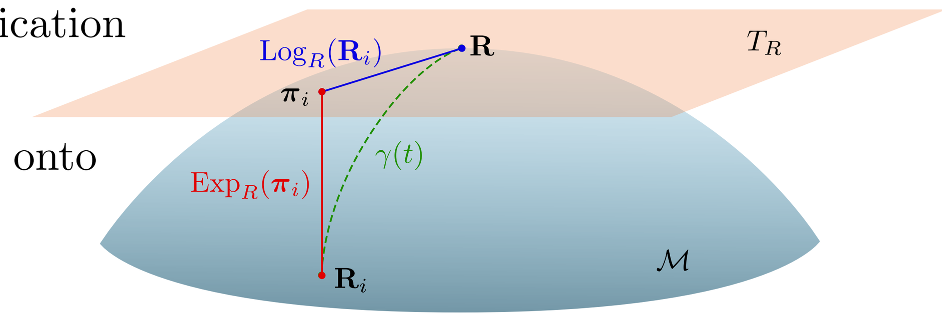
$$\mathcal{A} : \mathbf{X} \to \hat{\mathbf{X}}, \quad \boldsymbol{\psi} : \hat{\mathbf{X}} \to \mathbb{R}^d, \quad \boldsymbol{\varphi} : \mathbb{R}^d \to \{1, \dots, C\},$$

where $\mathcal{A}$ is a 3D Average Pooling, and $\boldsymbol{\varphi}$ is a classification head. Moreover, mapping $\boldsymbol{\psi}$ is a concatenation $\boldsymbol{\psi} = \boldsymbol{\psi}_1 \oplus \dots \oplus \boldsymbol{\psi}_C$:

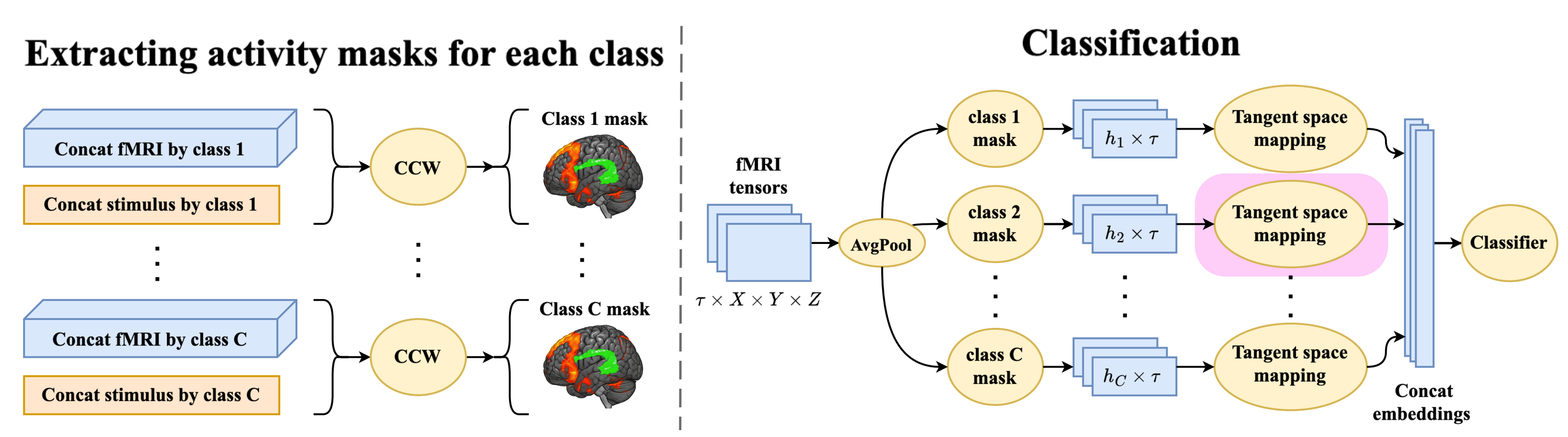
$$\boldsymbol{\psi}_k : \hat{\mathbf{X}} \to \mathbb{R}^{d_k}, \quad \boldsymbol{\psi}_k = \boldsymbol{\pi}_k \circ \mathbf{f}_k, \quad d = \sum_{k=1}^{C} d_k, \quad d_k = \frac{h_k(h_k + 1)}{2}.$$

There $\mathbf{f}_k : \hat{\mathbf{X}} \to \mathbb{R}^{h_k \times \tau}$ is an application of the activity map $\mathcal{M}^k$.

$\boldsymbol{\pi}_k : \mathbb{R}^{h_k \times \tau} \to \mathbb{R}^{d_k}$ is the projection onto the tangent (Riemannian) space.



Method Overview



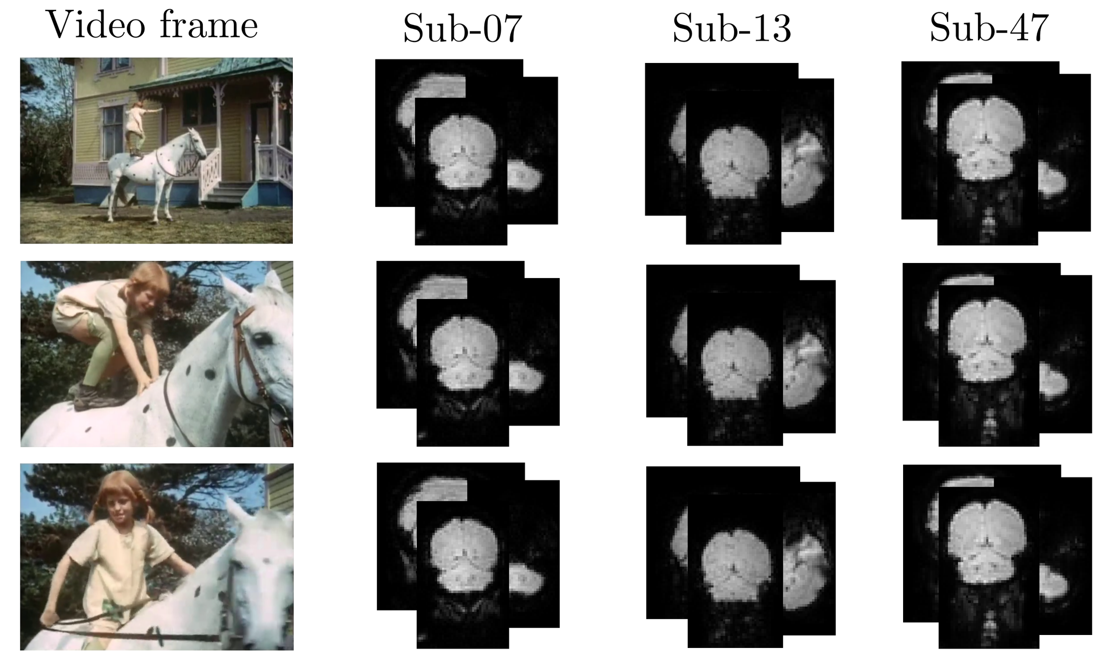
[2] Д.Д.Дорин, А.В.Грабовой, В.В.Стрижов. Пространственно-временные характеристики в задаче декодирования данных фМРТ. Preprint, 2024.

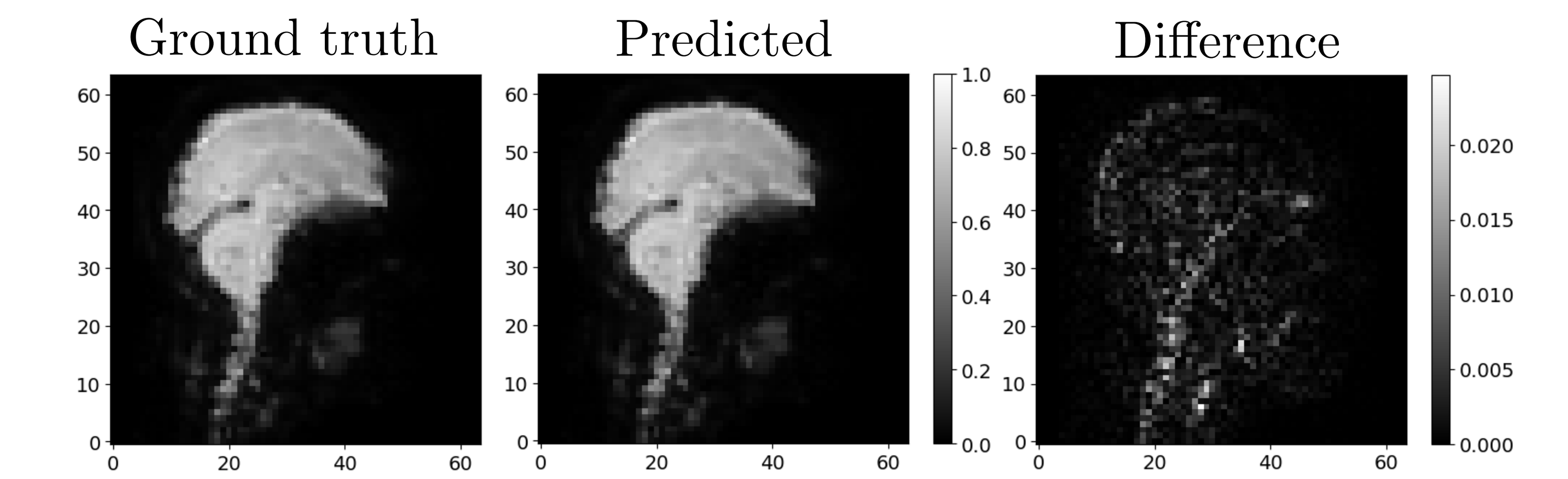
Empirical Analysis

Visual stimuli usage for fMRI forecasting

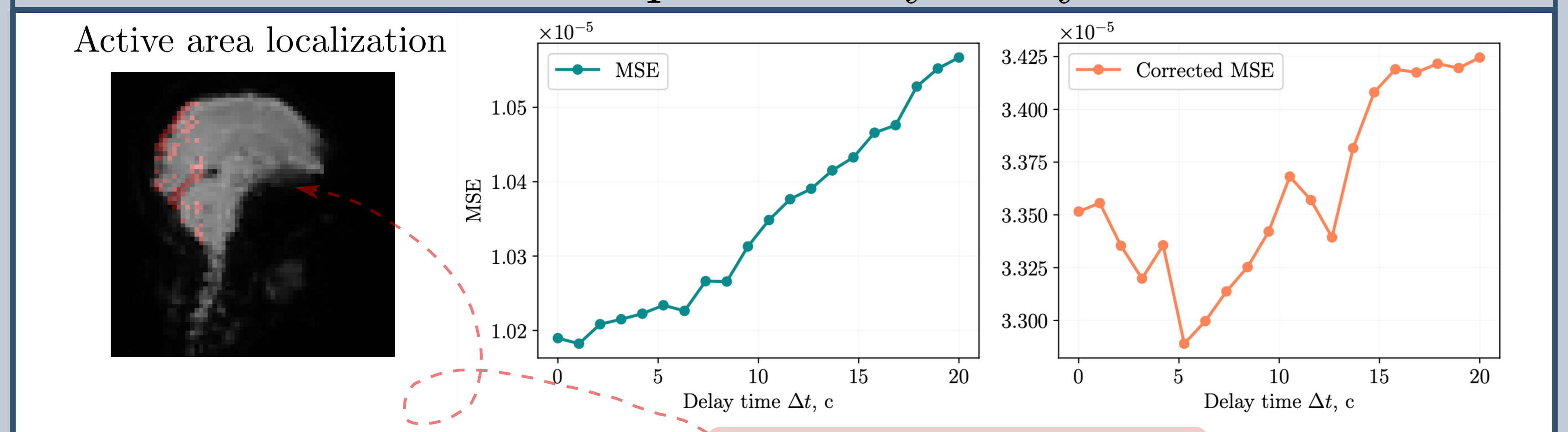
Dataset[3] description

- 30 fMRI participants
- 16 males, 14 females
- Age from 7 to 47 years
- Visual stimuli: Pippi Longstocking (6 min)
- Video frame rate: 25 Hz
- fMRI frame rate: 1.64 Hz
- Video resolution: 640x480
- fMRI resolution: 64x64x40

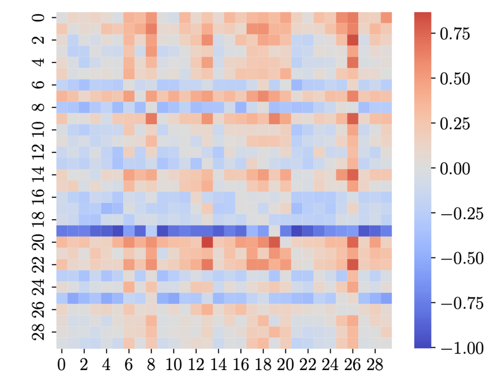




BOLD response delay analysis



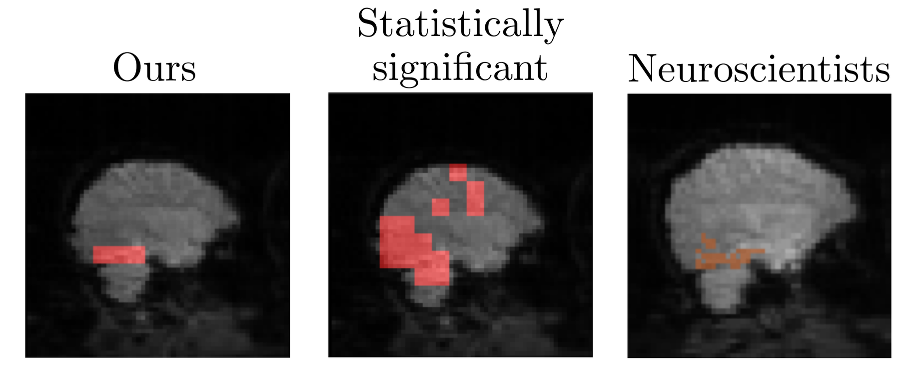
We localized the area that contains the 3% of the most variable voxels in the occipital lobe



MAPE of MSE changing when predicting on the mixed weight matrix

The figure shows the Mean Absolute Percentage Error (MAPE) of predicted MSE between pairs of subjects.
Each row and column represents a subject.
"Mixed" MSE uses one subject's weights to predict another's outcomes.
Positive values indicate "mixed" MSE is greater than the "true" MSE.

Voxel weighing and fMRI classification



Dataset description

- 6 fMRI participants, x12 sessions
- Visual stimuli: 8 categories of images (chair, house, cat, face, etc.)
- Number of fMRI images: 121
- fMRI frame rate: 2.5 Hz
- fMRI resolution: 64x64x40

- fMRI segments for 3 participants
- Fixed 10 weighted voxels for each class

Classifier	Accuracy	Macro F1 score
Logistic regression	0.60 ± 0.05	0.56 ± 0.05
MLP	0.70 ± 0.06	0.64 ± 0.06

- fMRI segments for Sub-02
- Ablation study for different method modules

Method	Accuracy	Macro F1 score	Micro F1 score
Ours, w/o Tangent Space Mapping	0.15	0.12	0.15
Ours, w/o activity maps	0.40	0.38	0.40
Ours	**0.65**	**0.60**	**0.65**

[3] Berezutskaya J., Vansteensel M.J., Aarnoutse E.J. et al. Open multimodal iEEG-fMRI dataset from naturalistic stimulation with a short audiovisual film. Sci Data 9, 91 (2022).