Forecasting fMRI Images From Video Sequences: Linear Model Analysis

Daniil Dorin[1][*][†], Nikita Kiselev[1][†] and Andrey Grabovoy[1]

[1]Department of Intelligent Systems, Moscow Institute of Physics and Technology, 9 Institutskiy per., Dolgoprudny, 141701, Moscow Region, Russia.

*Corresponding author(s). E-mail(s): dorin.dd@phystech.edu;
Contributing authors: kiselev.ns@phystech.edu;
grabovoy.av@phystech.edu;
[†]These authors contributed equally to this work.

Abstract

The issue of reconstructing the relationship between functional magnetic resonance imaging (fMRI) sensor readings and human perception of the external is investigated. The study analyzes the dependence between the fMRI images and the videos viewed by individuals. Based on this analysis, a method is proposed for approximating the fMRI readings using the video sequence. The method is based on the assumption that there is a time-invariant hemodynamic response to changes in blood oxygen levels. A linear model is constructed for each individual voxel in the fMRI image, assuming that the image sequence follows a Markov property. To test the proposed method, a computational experiment was conducted on a dataset collected during tomographic examinations of a large number of individuals. The performance of the method was evaluated based on the experimental data, and hypotheses were tested regarding the invariance of the model weights and the correctness of the method.

Keywords: neuroimaging, fMRI, video sequences, correlation analysis, hypothesis testing, linear model, forecasting

1

1 Introduction

A set of techniques used to visualize the structure and function of the human brain is called *neurovisualization*. Neuroimaging [1] techniques, such as electrocardiography (ECG), computed tomography (CT), magnetic resonance imaging (MRI), and functional magnetic resonance imaging (fMRI), are used to study the brain and detect diseases and mental disorders.

Functional magnetic resonance imaging, or *fMRI*, is a type of magnetic resonance imaging that measures changes in blood flow in the brain. These changes are caused by neural activity [2] and occur with a delay of 4–8 seconds [3], due to the time it takes for the vascular system to respond to the brain's demand for glucose [4–6]. This delayed response is essential for accurate fMRI measurements, as it allows for the detection of neural activity that occurs over a longer period of time.

In fMRI imaging, sequences of echo-planar imaging (EPI) are used [7–9]. The processing of areas with varying signal intensities, depending on the method of activation and the type of artifact, is carried out using specialized methods and software [3, 10, 11]. The processed results are then formalized as activation maps, which can be combined with the localization of anatomical structures in the cerebral cortex.

The fMRI technique plays a significant role in neuroimaging, although it has some significant limitations. The works of [12] and [13] discuss the temporal and spatial resolution of fMRI, with the latter emphasizing the significant disadvantage of temporal resolution. Another limitation of fMRI is the inevitable noise associated with movement, such as that caused by the heartbeat and respiration of the subject, as well as thermal fluctuations of the scanner itself. In the paper [14], graph-based techniques are proposed to address these issues and demonstrate their effectiveness in tasks such as epilepsy and depression detection. These methods aim to suppress the noise mentioned above and improve the accuracy of the results.

In fMRI, the participant is given a variety of tasks and external stimuli, which induce activation of specific areas of the brain. These stimuli include movements of fingers and limbs [15, 16], image finding and examination of a chessboard [17, 18], listening to non-specific noises, and reading single words or coherent text [19, 20]. Changes in brain activity during the fMRI examination could also be caused by watching a video [21], which is the focus of this study.

The most well-known video processing techniques are based on 3D convolutional neural networks [22]. The main difference between 3D and 2D convolution is that the former processes the spatial and temporal information simultaneously. However, this approach has some drawbacks, such as an increase in the number of parameters and computational complexity. One of the most advanced and promising architectures for image processing is ResNet [23]. This network allows training deep neural networks with up to 152 layers, overcoming the issue of gradient vanishing that occurs during training of deep networks. ResNet achieves high accuracy and has become a standard for image classification and object detection tasks.

The main focus of the research is to investigate the relationship between fMRI images and video recordings. It is assumed that such a relationship exists, and it is hypothesized that there may be a consistent time lag between fMRI images and corresponding video frames [6]. The study investigates the relationship between an

individual fMRI image and the preceding image. The time delay between these images serves as a hyperparameter in the model, which is used to approximate fMRI images based on a given video sequence.

According to a study [24], when patients undergo fMRI, viewing a video activates a specific network of brain regions. This network, which includes the occipital lobe and frontal regions, is located predominantly in the right hemisphere. In this paper, we are considering an approach that utilizes these brain regions to analyze latency times.

The data on which the dependency hypothesis is tested and the demonstration of the work of the constructed method are presented in paper [25]. This dataset was obtained by examining a group of 63 subjects, thirty of whom underwent fMRI scanning. They were asked to complete the same task, which was to watch a short audio visual film. Annotations were generated for this task in the paper under review, including information about the timing of the appearance and disappearance of individual words, objects, and characters. Detailed descriptions of the methods for audio and video annotation can be found in [26] and [27].

2 Problem statement

Frame rate $\nu \in \mathbb{R}$ and duration $t \in \mathbb{R}$ of the video sequence are set. The video sequence is set as
$$\mathbf{P} = [\mathbf{p}_1, \ldots, \mathbf{p}_{\nu t}], \qquad \mathbf{p}_\ell \in \mathbb{R}^{W \times H \times C},$$
with width, height and number of image channels W, H and C, respectively.

Let us denote the frequency of fMRI images by $\mu \in \mathbb{R}$. We set the sequence of images
$$\mathbf{S} = [\mathbf{s}_1, \ldots, \mathbf{s}_{\mu t}], \qquad \mathbf{s}_\ell \in \mathbb{R}^{X \times Y \times Z},$$
where X, Y, and Z — the dimensions of the voxel image.

The problem is to construct a mapping that accounts for the Δt delay between the fMRI image and the video sequence, as well as previous tomographic readings. fMRI image and the video sequence, as well as previous tomographic readings. Formally, it is necessary to find such a mapping $\mathbf{g}$ that

$$\mathbf{g}(\mathbf{p}_1, \ldots, \mathbf{p}_{k_\ell - \nu \Delta t}; \mathbf{s}_1, \ldots, \mathbf{s}_{\ell-1}) = \mathbf{s}_\ell, \quad \ell = 1, \ldots, \mu t,$$

where for the ℓ-th fMRI image the number of the corresponding image k_ℓ is determined by the formula
$$k_\ell = \frac{\ell \cdot \nu}{\mu}.$$

3 Proposed forecasting method

The scheme of the proposed fMRI image reconstruction method is shown in Figure 1.

We denote the fMRI image as $\mathbf{s}_\ell = [v_{ijk}^\ell] \in \mathbb{R}^{X \times Y \times Z}$, where $v_{ijk}^\ell \in \mathbb{R}_+$ — value of the corresponding voxel. In order to reduce the running time of the method, we propose to use compression of fMRI images by dimensionality reduction. Compression
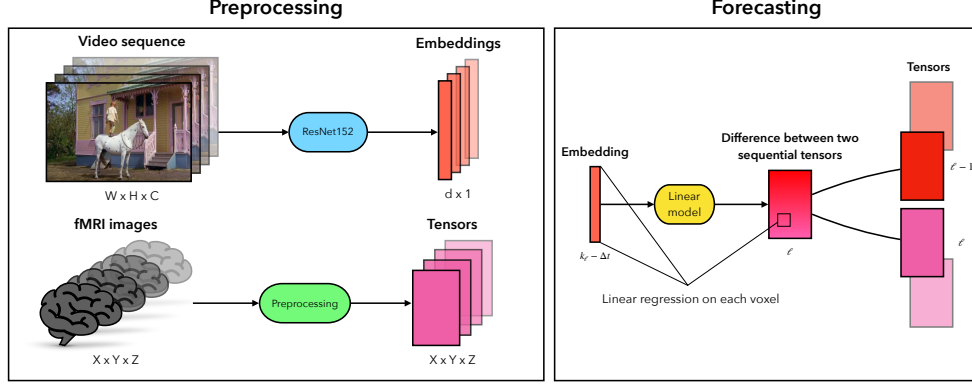
Fig. 1: Method scheme

by a factor of 2 is represented in the form of a mapping

$$\boldsymbol{\chi} : \mathbb{R}^{X \times Y \times Z} \to \mathbb{R}^{X/2 \times Y/2 \times Z/2}.$$

A compression of 2^k times is obtained by applying $\boldsymbol{\chi}$ successively k times. In the following, for simplicity, we keep the notation of image dimensions $X \times Y \times Z$.

Suppose that the Markov property is satisfied for the sequence of snapshots, i.e., each snapshot depends only on one image and the previous snapshot. Then the corresponding mapping is written in the form

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

where $\boldsymbol{\delta}_\ell = [v_{ijk}^\ell - v_{ijk}^{\ell-1}] = [\delta_{ijk}^\ell] \in \mathbb{R}^{X \times Y \times Z}$ — the difference between two consecutive snapshots.

Mapping $\mathbf{g} : \mathbf{P} \to \mathbf{S}$ is represented as a composition of the other two:

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

where

$$\boldsymbol{\psi} : \mathbf{P} \to \mathbb{R}^d \text{ — image vectorization,}$$
$$\boldsymbol{\varphi} : \mathbb{R}^d \to \mathbf{S} \text{ — target mapping.}$$

For each image from the video sequence, we have an embedding vector of dimension d:

$$\mathbf{x}_\ell = [x_1^\ell, \ldots, x_d^\ell]^\mathsf{T} \in \mathbb{R}^d, \quad \ell = 1, \ldots, \nu t.$$

The ResNet152 neural network architecture without the last linear layer is used.

Given $k_\ell = \ell \cdot \nu/\mu$, the total number of pairs (image, snapshot) is $N = \mu(t - \Delta t)$. Thus, for each voxel a sample is given

$$\mathfrak{D}_{ijk} = \{(\mathbf{x}_\ell, \delta_{ijk}^\ell) \mid \ell = 2, \ldots, N\}.$$

The regression task is set

$$y_{ijk} : \mathbb{R}^d \to \mathbb{R}.$$

A linear model is used with a vector of parameters

$$\mathbf{w}_{ijk} = [w_1^{ijk}, \ldots, w_d^{ijk}]^\mathsf{T} \in \mathbb{R}^d :$$

$$f_{ijk}(\mathbf{x}, \mathbf{w}_{ijk}) = \langle \mathbf{x}, \mathbf{w}_{ijk} \rangle.$$

For the model f_{ijk} with its corresponding parameter vector $\mathbf{w}_{ijk} \in \mathbb{R}^d$ define a quadratic loss function with L_2 regularization:

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

where $\alpha \in \mathbb{R}$ — regularization coefficient.

It is required to find the parameters that give a minimum to the loss functional $\mathcal{L}_{ijk}(\mathbf{w}_{ijk})$ for given hyperparameters Δt and α:

$$\hat{\mathbf{w}}_{ijk} = \arg \min_{\mathbf{w}_{ijk}} \mathcal{L}_{ijk}(\mathbf{w}_{ijk}).$$

The minimum of the loss function is found by the least squares method. Let's define the matrix of objects-features

$$\mathbf{X} = [\mathbf{x}_2, \ldots, \mathbf{x}_N]^\mathsf{T} = [x_j^i] \in \mathbb{R}^{(N-1) \times d}$$

and a vector whose components are the differences of values of the same voxel in different images,

$$\boldsymbol{\Delta}_{ijk} = [\delta_{ijk}^2, \ldots, \delta_{ijk}^N]^\mathsf{T} \in \mathbb{R}^{N-1}.$$

The solution is written in the form

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

Let us obtain a formula for reconstructed fMRI images. Let's introduce a matrix of weights

$$\hat{\mathbf{W}} = [\hat{\mathbf{w}}_1, \ldots, \hat{\mathbf{w}}_{XYZ}]^\mathsf{T} = [\hat{w}_j^i] \in \mathbb{R}^{XYZ \times d}.$$

Let's introduce for tensors $\mathbf{s}_\ell, \boldsymbol{\delta}_\ell \in \mathbb{R}^{X \times Y \times Z}$ vectors

$$\mathbf{s}_\ell^R = [v_1^\ell, \ldots, v_{XYZ}^\ell]^\mathsf{T}, \ \boldsymbol{\delta}_\ell^R = [\delta_1^\ell, \ldots, \delta_{XYZ}^\ell]^\mathsf{T} \in \mathbb{R}^{XYZ}.$$

Then the vector of the forecasted image is found by the formula

$$\hat{\mathbf{s}}_\ell^R = \mathbf{s}_{\ell-1}^R + \hat{\boldsymbol{\delta}}_\ell^R = \mathbf{s}_{\ell-1}^R + \hat{\mathbf{W}}\mathbf{x}_\ell.$$

4 Numerical experiment

To analyze the performance of the proposed method and test the hypotheses a computational experiment was carried out.

The sample presented in [25] was used as data. The dataset contains the results of examination of 63 subjects. For thirty of them fMRI readings are known. There are 16 males and 14 females, ranging in age from 7 to 47 years. The mean age of the subjects — 22 years.

Characteristics of the sample: duration of examination, frame rates of fMRI video sequences and images, and their dimensions are summarized in Table 1.

Table 1: Dataset Description

Name	Notation	Value
Duration of examination	t	390 s
Video frame rate	ν	25 Hz
fMRI frame rate	μ	1.64 Hz
Video dimensions	W, H, C	640, 480, 3
fMRI dimensions	X, Y, Z	40, 64, 64

The sample was divided into training and test samples in the ratio of 70% and 30%, respectively. The quality criterion for fMRI image reconstruction is MSE — the sum of squares of deviations between the true and reconstructed images, averaged over all voxels of each image. from the test sample.

To reduce the running time of the algorithm, the fMRI image is precompressed using MaxPool3D layer. Compression ratios of 1, 2, 4 and 8 are considered. The voxel values are normalized to $[0; 1]$ by the MinMaxScale procedure.

Table 2 summarizes the specifications of the computer on which the computational experiment was on which the computational experiment was performed.

Table 2: PC Specification

Element	Description
CPU	Intel Core i7-7700 3.6 GHz
GPU	NVIDIA GeForce GTX 1060 3 GB
RAM	16 GB 2400 MHz
Hard Drive	M.2 SSD
OS	Windows 10

4.1 Method performance

Figure 2 shows slices of the true and reconstructed images from the test sample. Figure 2.(c) shows the difference between them. To demonstrate the performance of the algorithm, the 7th subject was selected, $\Delta t = 5$s, compression factor 1, regularization factor $\alpha = 1000$. The 20th slice along the first coordinate of the 37th image in the sequence was considered. Since the voxel values are normalized to the segment $[0; 1]$, an error of the order of 10^{-3} indicates a fairly accurate prediction.

<div align="center">

(a) Test   (b) Predicted   (c) Difference
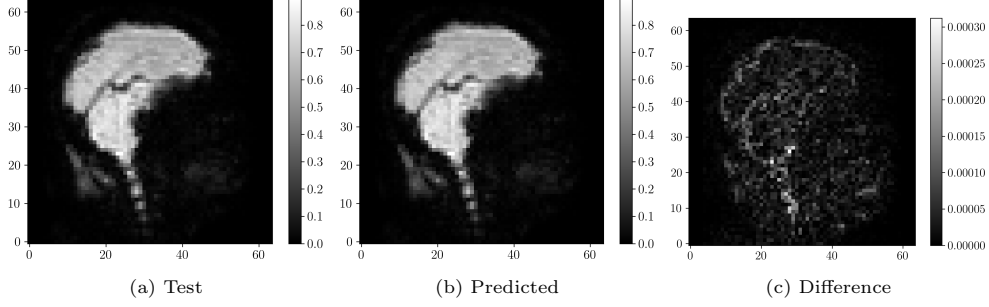
Fig. 2: Slices of fMRI images from the test sample

</div>

4.2 Delay time analysis

The dependence of recovery quality on delay time was investigated. The 47th subject and 4x compression were chosen for the example. The left graph in Figure 4 shows the dependence of the MSE metric on the delay time Δt. The study confirms that the most active part of the brain is the most active part of the brain in this examination — the occipital lobe. The other parts contribute noise to the considered dependence. In the present work, the above-mentioned region is localized, as shown in Figure 3. To localize the region, the lower third and the right two thirds of the volumetric of the tomographic image. The area highlighted in red is the area that contains the 3% of the most variable voxels in the occipital lobe. For this purpose, all voxels of the localized area were ordered by descending order of the total absolute change in values. Then 3% of voxels with the largest changes were selected. The MSE metric was recalculated exactly on this part of the image. The corresponding graph is shown on the right side of Figure 4. There is a more distinct minimum at $\Delta t \approx 5$ seconds.

4.3 Optimal regularization parameter

The dependence of MSE on the regularization parameter α was analyzed. Compression factors 1, 2, 4, and 8 were considered. The corresponding graphs are shown in Figure 5. Averaging over the subjects was performed to construct the graph. The limits of standard deviation are marked. The graphs show that the optimal value of the
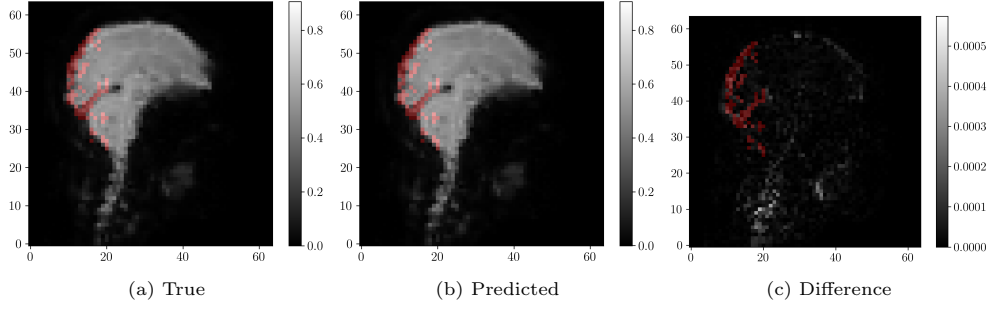
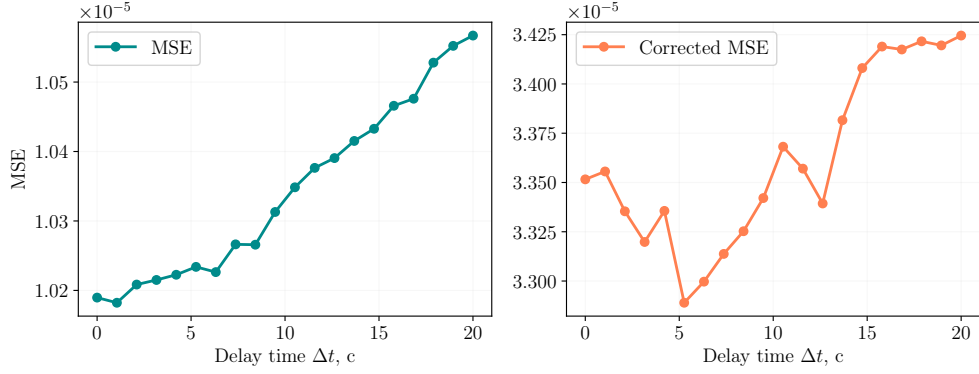Fig. 3: Localization of the most active zone

Fig. 4: Dependence of MSE on delay time

coefficient $\alpha \approx 1000$. The curve shape is preserved regardless of the compression factor of fMRI images.

4.4 Effect of image compression ratio on method runtime

We compare the training time of the model when using different compression coefficients of fMRI images. Coefficients 1, 2, 4 and 8 are considered. For each value of the compression ratio, the average value of the model training time for all subjects is calculated. value of the model training time. The standard deviation is calculated. The experimental results are summarized in Table 3. The running time of the method is significantly reduced when using pre-compression of fMRI images. The experiment with the selection of the optimal regularization coefficient confirms that the compression of the images does not change the dependences.

4.5 Analyzing the distribution of model weights

A graph of the distribution of the values of the components of the model weight vector was plotted. To construct it, we averaged over all voxels for the 4th subject.
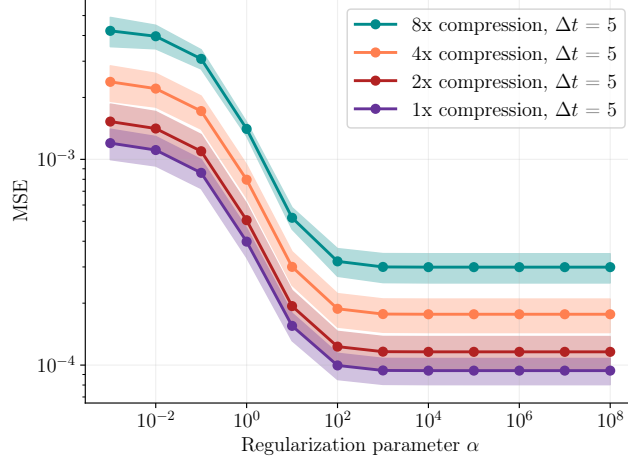
Fig. 5: Dependence of MSE metric on regularization parameter α on images from the test sample

Table 3: Dependence of model training time on compression ratio

Compression coefficient	Mean time, s	Std, s
1	36.3	6.1
2	6.7	0.5
4	1.6	0.1
8	1.4	0.3

The result is shown in Figure 6. The model weights do not lie in the neighborhood of any particular value, that is, their distribution is not degenerate. This result is quite consistent with reality, because a human being, while viewing pays attention to certain parts of the frame, such as characters or other details. other details.

4.6 Hypothesis of invariance of model weights with respect to humans

The hypothesis of invariance of the model weights with respect to the person was tested: Using one subject's weight matrix to reconstruct another subject's fMRI images. The MSE metric on the test sample was used. The results are presented in Table 4. The 4th and 7th subjects were considered. The weight matrix of the 4th was used to reconstruct the of the 7th subject's images. The MSE values are almost the same.

A similar experiment was conducted for each pair of subjects. The obtained results are presented in Figure 7, which was obtained as follows. Some subject (corresponding to a row of the matrix) is considered, MSE — "true" is calculated for him. Next,
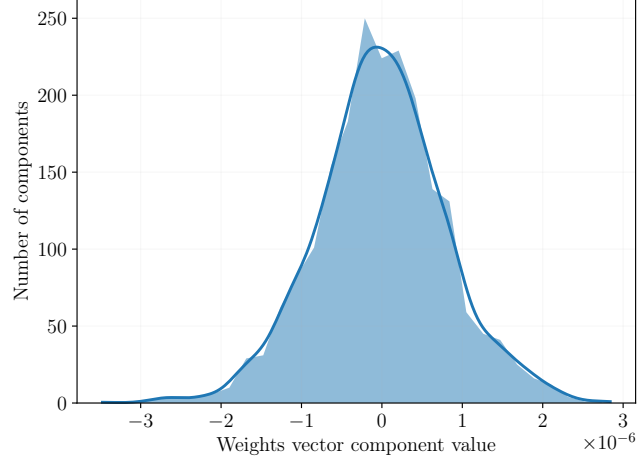
9

Fig. 6: Weights vector component distribution

Table 4: Testing the hypothesis of invariance of model weights with respect to humans

Weights matrix	True	Mixed	Difference
MSE	$6.3912 \cdot 10^{-5}$	$6.3911 \cdot 10^{-5}$	$1.37 \cdot 10^{-9}$

another subject is considered (corresponding to a column of the matrix), its matrix of weights is taken, and a prediction is made for the first subject, then the MSE — "mixed" is calculated. The difference between the resulting MSE as a percentage of the "true" is entered into the matrix. A positive value means that the "mixed" MSE is greater than the "true". A negative — that the "mixed" is smaller. That is, there is a MAPE on the heatmap. The ideal model should result in only positive deviation values, however, as can be seen in Figure 7, there are negative values in the matrix. Nevertheless, they are rather small, namely, they correspond to deviations of the order of 1%. This is explained by the fact that the model is quite simple, and therefore has a high generalizing ability. However, this does not prevent us from concluding that the data do not contradict the hypothesis about the invariance of the model weights with respect to humans.

4.7 Method corectness

The quality of the method performance on uninformative data is considered. A matrix consisting entirely of units was taken as a matrix of objects-signs **X**. Comparison with the results on the present feature-description matrix was made. To the first snapshot of the 35th subject, all the reconstructed changes in voxel values. As a result, we have the last snapshot of the sequence. Figure 8 are slices of the last true and recovered snapshots from the test sample. Figure 8.(c) shows the difference between them. The
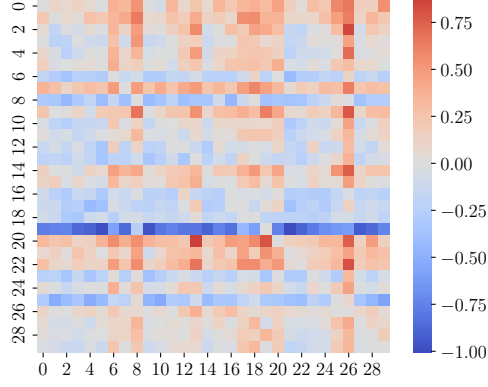
Fig. 7: MAPE of MSE changing when predicting on the mixed weight matrix

results on the uninformative ones are demonstrated in Figure 9. The difference between the true and recovered images when working with uninformative data is much higher, which confirms that there is a correlation between the sensor readings and the images from the video sequence. The numerical results are summarized in Table 5.
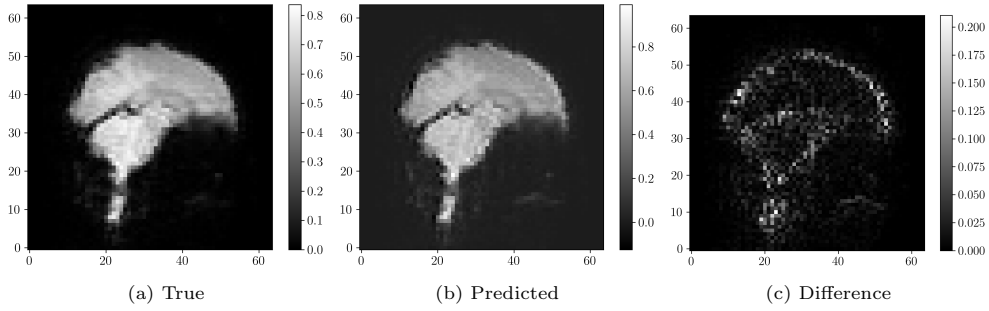

(a) True   (b) Predicted   (c) Difference

Fig. 8: Slices of fMRI images from the test sample

Table 5: Quality of method performance on uninformative data

Data	True	Uninformative	Difference
MSE	$3.67 \cdot 10^{-4}$	$1.39 \cdot 10^{-3}$	$1.02 \cdot 10^{-3}$

11

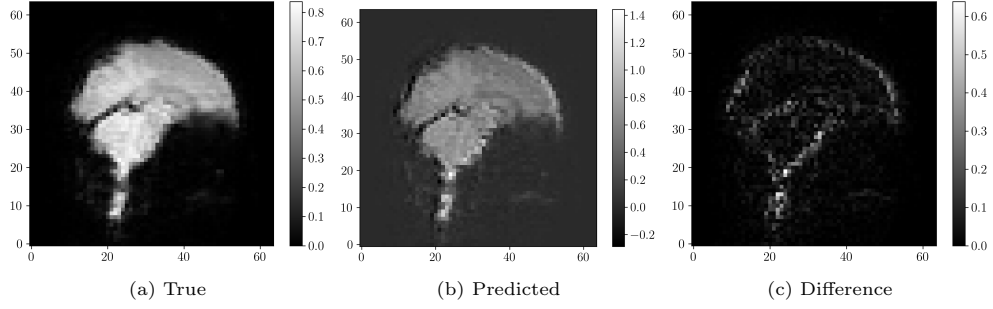(a) True   (b) Predicted   (c) Difference

Fig. 9: Slices of fMRI images from the test sample (uninformative data)

5 Conclusion

In this paper we consider the task of restoring the dependence between the readings fMRI sensors and human perception of the external world. The method of approximation of fMRI images sequence by video sequence is proposed. The method takes into account the hemodynamic response time — the delay time between the images from the video sequence and the fMRI images. A linear model is independently constructed for each voxel of the fMRI image. Each linear model is constructed under the assumption that the sequence of fMRI images is Markovian. In the experiments, the optimal value of the delay time is selected for each subject. The optimal value is found from analyzing the plot of MSE versus delay time. The regularization coefficient is selected. The effect of fMRI image compression ratio on model training time is investigated. It is assumed that the occipital lobe of the brain is responsible for information from visual organs. MSE correction is performed based on localization of this region and selection of the most changing voxels. With this construction, the graph has a characteristic minimum corresponding to the optimal value of the delay time. The obtained value of the delay time is consistent with neurobiological information. The experimental MSE values are small, indicating that there is a correlation between the data. The variation of images in the video sequence is taken into account since the distribution of model weights is not degenerate. The hypothesis of invariance of model weights with respect to humans is tested. The correctness of the method is confirmed by experiments with random data.

References

[1] Puras, J.V., Grigorieva, E.: The neurovisualization methods in diagnostics of head injury. part 1. computer tomography and magnetic resonance imaging. Russian journal of neurosurgery (2), 7–16 (2014)

[2] Glover, G.H.: Overview of functional magnetic resonance imaging. Neurosurgery Clinics of North America **22**(2), 133–139 (2011) https://doi.org/10.1016/j.nec. 2010.11.001

[3] Bandettini, P.A., Wong, E.C., Hinks, R.S., Tikofsky, R.S., Hyde, J.S.: Time course epi of human brain function during task activation. Magnetic Resonance in Medicine **25**(2), 390–397 (1992) https://doi.org/10.1002/mrm.1910250220

[4] Ogawa, S., Lee, T.M., Kay, A.R., Tank, D.W.: Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proceedings of the National Academy of Sciences **87**(24), 9868–9872 (1990) https://doi.org/10.1073/pnas.87.24.9868

[5] Le Bihan, D., Karni, A.: Applications of magnetic resonance imaging to the study of human brain function. Current Opinion in Neurobiology **5**(2), 231–237 (1995) https://doi.org/10.1016/0959-4388(95)80031-X

[6] Logothetis, N.K.: The underpinnings of the BOLD functional magnetic resonance imaging signal. The Journal of Neuroscience **23**(10), 3963–3971 (2003) https://doi.org/10.1523/jneurosci.23-10-03963.2003

[7] Connelly, A., Jackson, G.D., Frackowiak, R.S., Belliveau, J.W., Vargha-Khadem, F., Gadian, D.G.: Functional mapping of activated human primary cortex with a clinical mr imaging system. Radiology **188**(1), 125–130 (1993) https://doi.org/10.1148/radiology.188.1.8511285

[8] Kwong, K.K., Belliveau, J.W., Chesler, D.A., Goldberg, I.E., Weisskoff, R.M., Poncelet, B.P., Kennedy, D.N., Hoppel, B.E., Cohen, M.S., Turner, R.: Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. Proceedings of the National Academy of Sciences **89**(12), 5675–5679 (1992) https://doi.org/10.1073/pnas.89.12.5675

[9] Ogawa, S., Tank, D.W., Menon, R., Ellermann, J.M., Kim, S.G., Merkle, H., Ugurbil, K.: Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. Proceedings of the National Academy of Sciences **89**(13), 5951–5955 (1992) https://doi.org/10.1073/pnas.89.13.5951

[10] Baudendistel, K., Schad, L.R., Friedlinger, M., Wenz, F., Schröder, J., Lorenz, W.J.: Postprocessing of functional mri data of motor cortex stimulation measured with a standard 1.5 t imager. Magnetic Resonance Imaging **13**(5), 701–707 (1995) https://doi.org/10.1016/0730-725X(95)00016-A

[11] Cox, R.W.: Afni: Software for analysis and visualization of functional magnetic resonance neuroimages. Computers and Biomedical Research **29**(3), 162–173 (1996) https://doi.org/10.1006/cbmr.1996.0014

[12] Menon, R.S., Kim, S.-G.: Spatial and temporal limits in cognitive neuroimaging with fmri. Trends in cognitive sciences **3**(6), 207–216 (1999)

[13] Logothetis, N.K.: What we can do and what we cannot do with fmri. Nature

453(7197), 869–878 (2008)

[14] Sharaev, M., Andreev, A., Artemov, A., Bernstein, A., Burnaev, E., Kondratyeva, E., Sushchinskaya, S., Akzhigitov, R.: fMRI: preprocessing, classification and pattern recognition (2018)

[15] Roux, F.E., Ranjeva, J.P., Boulanouar, K., Manelfe, C., Sabatier, J., Tremoulet, M., Berry, I.: Motor Functional MRI for Presurgical Evaluation of Cerebral Tumors. Stereotactic and Functional Neurosurgery **68**(1-4), 106–111 (1998) https://doi.org/10.1159/000099910 https://karger.com/sfn/article-pdf/68/1-4/106/3533549/000099910.pdf

[16] Papke, K., Hellmann, T., Renger, B., Morgenroth, C., Knecht, S., Schuierer, G., Reimer, P.: Clinical applications of functional mri at 1.0 t: motor and language studies in healthy subjects and patients. European Radiology **9**(2), 211–220 (1999) https://doi.org/10.1007/s003300050658

[17] Engel, S.A., Rumelhart, D.E., Wandell, B.A., Lee, A.T., Glover, G.H., Chichilnisky, E.-J., Shadlen, M.N.: fmri of human visual cortex. Nature **369**(6481), 525–525 (1994) https://doi.org/10.1038/369525a0

[18] Schneider, W., Casey, B.J., Noll, D.: Functional mri mapping of stimulus rate effects across visual processing stages. Human Brain Mapping **1**(2), 117–133 (1994) https://doi.org/10.1002/hbm.460010205

[19] Binder, J.R., Rao, S.M., Hammeke, T.A., Yetkin, F.Z., Jesmanowicz, A., Bandettini, P.A., Wong, E.C., Estkowski, L.D., Goldstein, M.D., Haughton, V.M., Hyde, J.S.: Functional magnetic resonance imaging of human auditory cortex. Annals of Neurology **35**(6), 662–672 (1994) https://doi.org/10.1002/ana.410350606

[20] Dymarkowski, S., Sunaert, S., Van Oostende, S., Van Hecke, P., Wilms, G., Demaerel, P., Nuttin, B., Plets, C., Marchal, G.: Functional mri of the brain: localisation of eloquent cortex in focal brain lesion therapy. European Radiology **8**(9), 1573–1580 (1998) https://doi.org/10.1007/s003300050589

[21] Decety, J., Grezes, J., Costes, N., Perani, D., Jeannerod, M., Procyk, E., Grassi, F., Fazio, F.: Brain activity during observation of actions. influence of action content and subject's strategy. Brain: a journal of neurology **120**(10), 1763–1777 (1997)

[22] Tran, D., Bourdev, L., Fergus, R., Torresani, L., Paluri, M.: Learning spatiotemporal features with 3d convolutional networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 4489–4497 (2015)

[23] He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition (2015)

[24] Anderson, D., Fite, K., Petrovich, N., Hirsch, J.: Cortical activation while watching video montage: An fmri study. Media Psychology - MEDIA PSYCHOL **8**, 7–24 (2006) https://doi.org/10.1207/S1532785XMEP0801_2

[25] Berezutskaya, J., Vansteensel, M.J., Aarnoutse, E.J., Freudenburg, Z.V., Piantoni, G., Branco, M.P., Ramsey, N.F.: Open multimodal iEEG-fMRI dataset from naturalistic stimulation with a short audiovisual film. Scientific Data **9**(1) (2022) https://doi.org/10.1038/s41597-022-01173-0

[26] Boersma, P., Weenink, D.: Praat: doing phonetics by computer [computer program]. version 6.0. 37. Retrieved February **3**, 2018 (2018)

[27] Berezutskaya, J., Freudenburg, Z.V., Ambrogioni, L., Güçlü, U., Gerven, M.A.J., Ramsey, N.F.: Cortical network responses map onto data-driven features that capture visual semantics of movie fragments. Scientific Reports **10**(1) (2020) https://doi.org/10.1038/s41598-020-68853-y