Internship Report

Orhan Soyuhos

Dr. Yağmur Güçlütürk

The Donders AI for Neurotech Lab

Radboud University

23.02.2021

Introduction

The recent applications of machine learning to brain data presented new opportunities to understand the way cognitive functions take place. In particular, the research on vision science has led the progress in this respect. The retinotopic organization in the visual system made it possible to reconstruct naturalistic videos from brain data via neural networks, and it let us peek into the functioning of the human brain (Le et al., 2021). However, the unnaturalistic way the subjects viewed these videos hindered the experiments from approximating the natural condition humans are usually involved in. Until now, to keep the correspondence between the visual space and cortex, studies have aimed to reproduce naturalistic videos from the brain activity of subjects who viewed scenes by fixating on the center throughout the movies (Seeliger et al., 2019). Alternatively, the current research aims to enable naturalistic video reconstruction methods by using eye-tracking recordings from the brain data of participants who have free-viewed movies.

Dataset

To this end, we will use the extension of the studyforrest dataset (ds000113d) (Hanke et al., 2016) in which 15 participants, who had already taken part in the primary dataset, watched the audio-visual 'Forrest Gump' movie during two-hour 3 Tesla fMRI acquisition with 2 s repetition time (TR). Notably, their eye gaze locations were recorded simultaneously with fMRI. Additionally, 15 new control subjects (subject ID > 20) watched the movie in lab settings, i.e., outside of fMRI, and their eye gaze trajectories were also recorded. The fMRI data were acquired in two sessions on the same day with an approximately 10 min break. Participants were presented with four movie segments (about 15 min long each) in chronological order in each session. In the behavioral eye-tracking session, subjects did not have a break.

Methods

In our research, we planned to follow the steps of Le et al. (2021) to use the brain2pix architecture after preprocessing the frames of movie stimuli for each subject following their eye gaze data. Firstly, we visualized the selected subjects' eye gaze data on the movie's first segment to inspect their quality. We used the normalized eye data provided by Hanke et al. (2016) in which x and y coordinates are located in the native movie frame excluding the gray bars, while the top-left corner is (0, 0) and the bottom-right corner is (1280, 540). Since the stimuli were 25 frames per second (fps) and the eye gaze recording resolution was 1000 Hz, we averaged the eye gaze data in each 40 Hz while excluding missing data (NaNs) to match the fps of movie stimuli. We used Psychtoolbox-3 on MATLAB R2020a for the visualization (see *https://youtu.be/4g3Oc2bZfpQ*). In addition to the technical validation for eye-tracking data quality provided by Hanke et al. (2016), we measured the percentage of NaNs values per subject and stimuli segment after averaging eye gaze data in each 40 Hz (Figure 1). We observed an overall slight decrease in the percentage of NaNs for all subjects and highlighted the subjects either whose missing data is more than 10% or who had substantial motion during fMRI. For subject 18, we translated each frame of the movie stimuli to the center of the native movie frame using eye gaze coordinates for each frame as translation vectors. In a way, the subjects' eye gaze coordinates for each movie frame were translated as if they were fixated to the center of stimuli throughout the movie presentation. During this process, we used the function called imtranslate from MATLAB R2020a's Image Processing Toolbox. Additionally, each frame was cropped into 672 × 672 px to use in the fourth step (see *https://bit.ly/3byUSO1*). This step made it possible to proceed with the brain2pix method to reconstruct movie frames from the fMRI data. Its code can be found in *f'{rel_dir}edit_videoFrames_matlab'* in the directory of the lovelace computer, where *rel_dir = '/home/soyorh/Desktop/studyforrest/Preprocessing steps/'*.

Figure 1

The Quality of Eye-Tracking Data

	% of NaNs - 25 Hz (avg in each 40 rows)									
ID (Total 15 subjects)	1. Segment	2. Segment	3. Segment	4. Segment	5. Segment	6. Segment	7. Segment	8. Segment	AVG per ID	
19	1,77	0,97	1	1,03	0,17	0,19	0,29	0,05	0,68	
18	1,2	1,62	1,94	1,89	1,29	1,52	2,44	2,96	1,86	
17	3,45	1,41	3,58	3,29	5	4,94	3,24	2,99	3,49	
15	6,85	9,71	3,26	2,41	3,54	2,67	2,67	2,04	4,14	
2	5,5	3,84	3,98	3,02	4,55	4,22	5,36	3,13	4,2	
14	6,6	4,07	4,87	7,27	6,74	5,85	7,65	6,44	6,19	
3	12,73	8,06	8,64	6,22	7,78	6,75	8,72	3,33	7,78	
1	1,57	46,69	2,18	2,17	3,09	2,28	3,95	2,26	8,02	
6	9,06	7,19	9,75	4,68	8,83	16,82	10,44	4,34	8,89	
9	4,69	6,63	11,38	6,94	15,68	9,17	10,24	15,22	9,99	*fMRI
16	1,38	52,7	1,34	19,84	0,57	1,03	0,24	2,94	10,01	
10	4,28	9,55	18,98	2,7	37,54	14,48	17,9	10,19	14,45	
4	13,46	7,21	11,68	31,24	7,86	13,75	24,35	12,04	15,2	
20	54,1	29,61	32,8	42,74	21,6	24,71	39,28	31,03	34,48	*fMRI
5	86,48	85,43	80,67	76,43	84,74	89,86	93,99	79,31	84,61	
AVG per Segment	14,21	18,31	13,07	14,12	13,93	13,22	15,38	11,88		

Note. The figure shows the percentage of missing data (NANs) per subject and run after averaging eye gaze data for each 40 data points. Note that subjects 9 and 20 had substantial motion (>1.5) during fMRI.

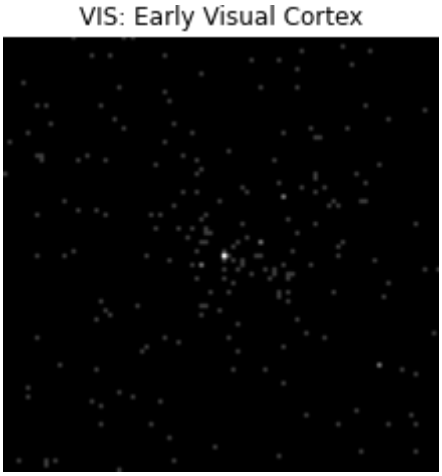
In the second step, we followed the first part of brain2pix architecture to transform the BOLD responses of selected voxels to a tensor in pixel space. To this end, we used another extension of the studyforrest dataset (Sengupta et al., 2016) that provides retinotopic mapping and localization of higher visual areas. We used two NIfTI files located inside a folder called post_processing from their GitHub account (see *https://github.com/psychoinformatics-de/studyforrest-data-retinotopy*) that have polar angles (θ) and eccentricity (r) values per each voxel in a participant's brain. Importantly, both polar angle and eccentricity values are ranged between 0 to 360. Firstly, we normalized eccentricity values into a range between -0.5 and 0.5 to map each voxel onto the desired visual space later. Secondly, we converted these polar coordinates (r, θ) to cartesian ones (x, y) for each voxel ($x = r \times \cos(\theta)$; $y = r \times \sin(\theta)$). Afterward, we applied visual region of interest (ROI) masks (see *.../studyforrest-data-visualrois*) to leave out the voxels that we are not interested in. Note that the number of voxels

for each ROI varies widely across subjects. For some subjects, a particular ROI is composed of more than one file, so that they should be concatenated before masking.

In the third step, we calculated the receptive field locations (RFlocations) and signals (RFsignals) for voxels in the early visual cortex (VIS). We first used the get_RF function provided by Le et al. (2021) to have RFlocations at a resolution of 1280 × 1024 px same as the retinotopic mapping stimuli. However, since the movie stimuli were shown at 1280 × 720 px, we cropped the RFlocations. To have the desired size of receptive field signal images (RFSimages) for the brain2pix model, we cropped the RFlocations into 672 × 672 px. Additionally, we listed the indexes of voxels left outside to crop the RFsignals later accordingly. Finally, we applied 2D max-pooling (kernel_size=7, stride=7, padding=0) to downsize the RFlocations into a dimension of 96 × 96 px (Figure 2). We saved RFlocations as .npy files and their figures in *f'{rel_dir}RF_locations'* and *f'{rel_dir}whole_brainfigures'* respectively. For the RFsignals, we removed the BOLD responses of voxels that were not included in the RFlocations after applying VIS mask to participant-specific aligned fMRI data (see .../studyforrest-data-aligned). We saved them separately for testing and training data as .npy files in *f'{rel_dir}testing_signalsRF'* and *f'{rel_dir}training_signalsRF'*. All training data underwent z-score transformation beforehand. The second and third steps were implemented with Python 3.8.5 using Jupyter Notebook as the rest of the code, and its .ipynb file can be found in *f'{rel_dir}To Cartesian Coordinates (x,y) and SaveRF_signals&locations.ipynb'*.

Figure 2

Example Receptive Field Locations

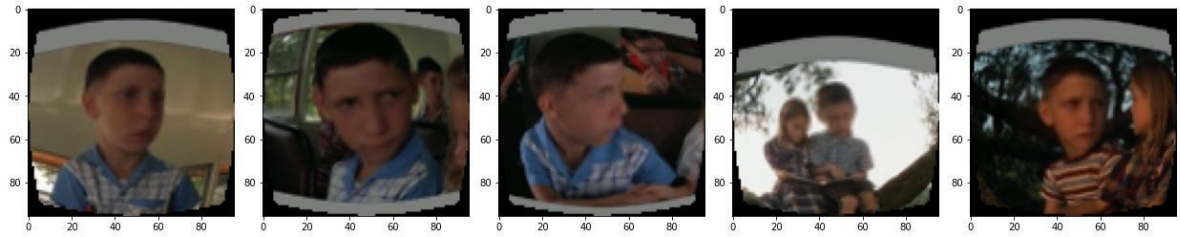


Note. Receptive field locations (RFlocations) of early visual cortex (VIS) for subject 18. We chose subject 18 because of the higher number of voxels in the VIS.

In the fourth step, we generated targets for RFsignals. The movie stimuli for each segment were scaled into 96 × 96 px as the RFsignals, and their fps were converted into 0.5 by using FFmpeg 3.4.8 to match the repetition time (TR) of the fMRI acquisition. The resulting movie segments were in .webm format and had 451, 441, 438, 488, 462, 439, 542, and 338 frames, respectively, same as the number of volumes acquired per movie segment during fMRI. They were saved as .npy files for training and test set using the code provided by Le et al. (2021). Lastly, they underwent a fish-eye transformation to mimic the primate retina's spatial sampling properties (Bashivan et al., 2018) after being divided by 255 and concatenated together (Figure 3). They can be found in *f'{rel_dir}training_videos/{set_t[:-3]}_targets_warped.npy'* where *set_t = training or testing*. The code is in the f*'{rel_dir}Generate Targets with Retinawarp.ipynb'* directory.

Figure 3

Example Target Frames for Receptive Field Signals



Note. The example target frames (96 × 96 px) after fish-eye transformation. Respectively, they are 400th, 410th, 420th, 430th, and 440th frames of the first movie segment after being converted into 0.5 fps.

In the last step, we matched RFsignals with their target frames while separating them by 5-time channels and considering the hemodynamic delay of 2.8s – 5.4s as in Le et al. (2021). To this end, each target frame was realigned with the brain signals four steps after. However, since movie segments' durations are equal to the number of fMRI volumes acquired (meaning fMRI recording did not continue after the last movie frame), we left out eight movie frames for each run (shift + time_range -1). The code can be found in *f'{rel_dir}Shift and Stuck the fMRI data.ipynb'*.

Ultimately, the studyforrest data is ready to be fed into the brain2pix model for training and testing. The correlation or distance values between the features of test stimuli and their reconstructions will indicate whether our method is successful or not.

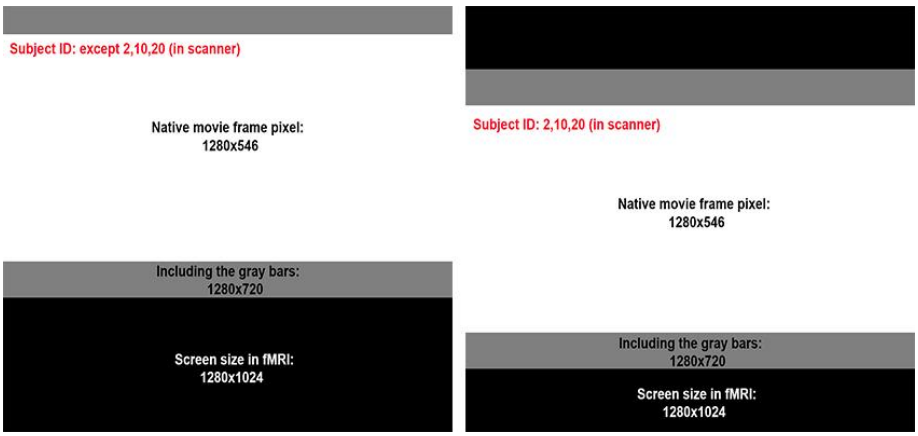
Important Points

Before training the brain2pix model with the studyforrest dataset, we should consider the following steps to save time for later:

1. I am not sure whether the eccentricity values are from 0 to 360 or 360 to 0. I emailed and opened an issue on their GitHub repository but have not received an answer. I assumed that they are 360 to 0 (center to edges) for now.
2. After converting polar coordinates to cartesian ones, I am not sure about the correspondence between x, y and row, column values for the image dimension. I assumed column values equal to x and row values equal to y (see *f'{rel_dir}To Cartesian Coordinates (x,y) and SaveRF_signals&locations.ipynb'*).
3. I am not sure about the number of movie frames left out in the last step. For now, I calculated it as eight frames for each run (shift + time_range -1).
4. I have used only VIS as a visual mask to create RFSimages. Also, I mainly worked with subject 18 during preprocessing steps since he or she has a higher number of voxels in VIS and good eye-tracking data.
5. I am not sure whether the aligned fMRI data I used was denoised. There is also denoised fMRI data. However, its shape does match with the RIOs mask, so that I chose the aligned one.
6. We should remember that for subjects 2, 10, and 20, the movie stimuli were presented differently on the screen (Figure 4).

Figure 4

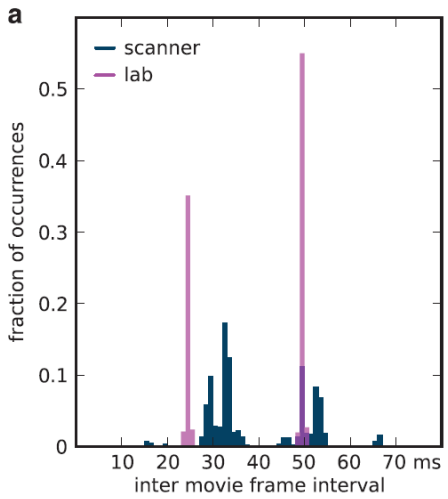
Movie Presentation During fMRI



7. We should note that even though the movie stimuli were 25 fps, each movie frame's duration on the screen was not the same (Figure 5).

Figure 5

The Duration of Frames on Screen



Note. The timing of the movie frame presentation. It shows the fraction of the frame's display durations across all participants and movie segments. Reprinted from "A studyforrest extension, simultaneous fMRI and eye gaze recordings during prolonged natural stimulation," by M. Hanke, N. Adelhöfer, D. Kottke, V. Iacovella, A. Sengupta, F. R. Kaule, R. Nigbur, A. Q. Waite, F. Baumgartner, and J. Stadler, 2016, *Scientific Data, 3(1), 160092*.

8. Since there are no separate training and test sets for the studyforrest as the doctor who data, we need to decide how to divide them.
    a. Some visual stimuli such as Forrest's house keep showing in different movie segments. As Lynn pointed out in our meeting, this might raise a question against our study's validity.
9. I have questions about implementing the Representational Similarity Analysis (RSA) for the studyforrest data since no clear categories exist between conditions, such as houses versus faces.

References

Bashivan, P., Kar, K., & DiCarlo, J. J. (2018). Neural Population Control via Deep Image Synthesis. *BioRxiv*, 461525. https://doi.org/10.1101/461525

Hanke, M., Adelhöfer, N., Kottke, D., Iacovella, V., Sengupta, A., Kaule, F. R., Nigbur, R., Waite, A. Q., Baumgartner, F., & Stadler, J. (2016). A studyforrest extension, simultaneous fMRI and eye gaze recordings during prolonged natural stimulation. *Scientific Data*, *3*(1), 160092. https://doi.org/10.1038/sdata.2016.92

Le, L., Ambrogioni, L., Seeliger, K., Güçlütürk, Y., Gerven, M. van, & Güçlü, U. (2021). Brain2Pix: Fully convolutional naturalistic video reconstruction from brain activity. *BioRxiv*, 2021.02.02.429430. https://doi.org/10.1101/2021.02.02.429430

Seeliger, K., Sommers, R. P., Güçlü, U., Bosch, S. E., & Gerven, M. A. J. van. (2019). A large single-participant fMRI dataset for probing brain responses to naturalistic stimuli in space and time. *BioRxiv*, 687681. https://doi.org/10.1101/687681

Sengupta, A., Kaule, F. R., Guntupalli, J. S., Hoffmann, M. B., Häusler, C., Stadler, J., & Hanke, M. (2016). A studyforrest extension, retinotopic mapping and localization of higher visual areas. *Scientific Data*, *3*(1), 160093. https://doi.org/10.1038/sdata.2016.93