Reconstructing Visual Experiences from Brain Activity Evoked by Natural Movies using Adversarial Neural Network

Shujon Naha, Md. Lisul, Md. Nayem

March 2, 2017

1 Introduction

Decoding brain activity to meaningful represetations such as naturalistic videos has several practical applications. Brain activities can be acquired by functional magnetic resonance imaging (fMRI). Nishimoto et. al. [1], used simple Baysian models to reproduce visual stimuli such as natural videos from fMRI responses by predicitng blood oxygen level-dependent (BOLD) signals. They first learn several temporal and spatial filters to convert natural videos into BOLD signals by considering their fMRI based BOLD signals as ground truths. Then during test time, they generate the BOLD signal from fMRI responses for a given video stimuli. Then they use the learned model to generate BOLD signals for a large number of images and match them against the BOLD signal generated from the fMRI response sequence. Finally they take average of the best ranked videos to reproduce the given video stimuli. Figure 1 gives an overview of their approach.

The above approach uses predefined set of spatial-temporal filters which will not be able to fully capture the variability in mapping fMRI response sequences to natual videos. Also, it is not learned which part of the brain is responsible for generating different levels of abstractions of the reproduced video. Finally, reproducing the stimuls video only by taking the cumulative average of similar videos gives a rather unclear video.

2 Our Approach

We propose a adversarial neural network based approach to reproduce the stimulus video directly from fMRI response. Our approach is similar to [2], where they trained a adversarial neural network by applying 3d convolution on a large number of natural videos to generate videos with scene dynamics. Our input will be the raw fMRI response and the network suppose to generate the actual input video stimuli. To encode the fMRI we will consider an attention model similar to [3] which will learn to focus on specific brain areas to accurately

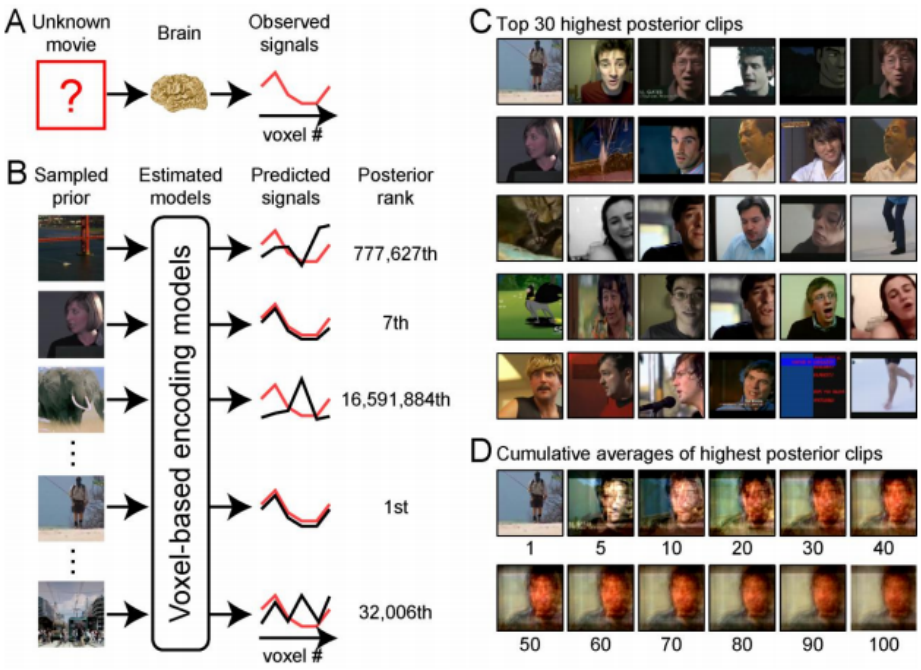

Figure 1: Schematic Diagram of Decoding Algorithm from [1].

reproduce the input video. Also, we will consider the conditional generative approaches like [4], where the fMRI responses will be considered as attributes to generate the video.

3 Experiments

To measure the performance we will use the exact same experimental setting as [1]. The dataset contains 120 minutes of training videos and corresponding fMRI responses. The testing instances are consists of 9 minutes of test videos and their corresponding fMRI responses. At each time point the fMRI response contains a linear BOLD vector of size 73728. The vector is generated by scanning brain voxels where scanning volume is 64x64x18(=73728) As the evaluation criteria, we will consider the correleation between the original movie stimuli and reconstructions within the motion-energy space similar to [1]. But as the motion-energy space is dependent on their generator model, we may need to fix another evaluation criteria to measure the video generation performance.

References

[1] Nishimoto, S., Vu, A.T., Naselaris, T., Benjamini, Y., Yu, B. and Gallant, J.L., 2011. Reconstructing visual experiences from brain activity evoked by natural movies. Current

Biology, 21(19), pp.1641-1646.

[2] Vondrick, Carl, Hamed Pirsiavash, and Antonio Torralba. "Generating videos with scene dynamics." In Advances In Neural Information Processing Systems, pp. 613-621. 2016.

[3] Gregor, Karol, Ivo Danihelka, Alex Graves, Danilo Jimenez Rezende, and Daan Wierstra. "DRAW: A recurrent neural network for image generation." arXiv preprint arXiv:1502.04623 (2015).

[4] Yan, Xinchen, Jimei Yang, Kihyuk Sohn, and Honglak Lee. "Attribute2image: Conditional image generation from visual attributes." In European Conference on Computer Vision, pp. 776-791. Springer International Publishing, 2016.