Analyzing video features related to specific brain networks using deep learning

Wonbum Sohn[1](ws35@njit.edu), Xin Di[1], Bharat B. Biswal[1]

[1]Department of Biomedical Engineering, New Jersey Institute of Technology, Newark, NJ, 07102

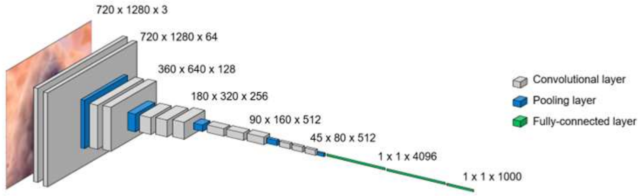
Introduction

- Functional magnetic resonance imaging (fMRI) analyzes the brain based on the blood oxygen level-dependent signal.
- Movie watching measures fMRI in an environment similar to the real environment [1].
- Movie watching can simultaneously measure low-level and high-level brain networks.
- However, it is difficult to determine which stimuli activated a particular brain network because various stimuli are given simultaneously.
- To find a relationship between specific stimuli and specific brain networks during movie watching, we used a convolutional neural network (CNN).

Materials and Methods

1) fMRI Data
   - Public open data from OpenNeuro.
   - 5 min "Partly Cloudy" animation from Pixar [2]
   - SPM12 and MATLAB (R2021b)
   - Preprocessing
     - Segmentation, skull-strip, realignment, framewise displac-ement, coregistration, normalization and smoothing.
   - 29 adult subjects (mean = 24.6, SD = 5.3)

2) Extract feature
   - A pre-trained VGG-16 [3]
   - Only 13 convolutional (conv) layers and 4 max-pooling layers
   - 2s repetition time (TR) and 24 frames/s frame rate of video
     → every 48 frames are averaged into 1 image per each TR.
     → 175 input images
   - 175 feature maps derived from each kernel of each conv layer → 1D time series with length 175 (averaging)



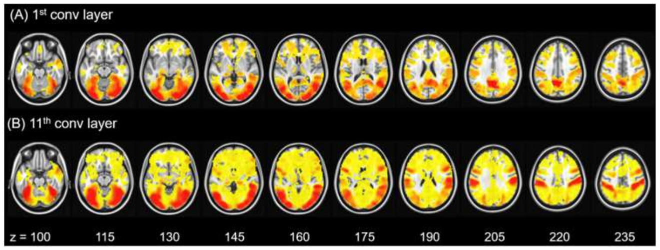
<Fig 1. VGG-16 structure. It was only used until fully-connected layers and final max-pooling.>

3) General linear model (GLM)
   - A group-level GLM with the 1D time series of each kernel and the preprocessed fMRI data.
   - Combine GLM results obtained from all feature maps of each conv layer to compare activated brain networks depending on depth.

4) Analyze feature maps
   - Average time series related to a particular brain network across all subjects.
   - Analyze the video's feature maps associated with TR times at which peaks of the averaged time series

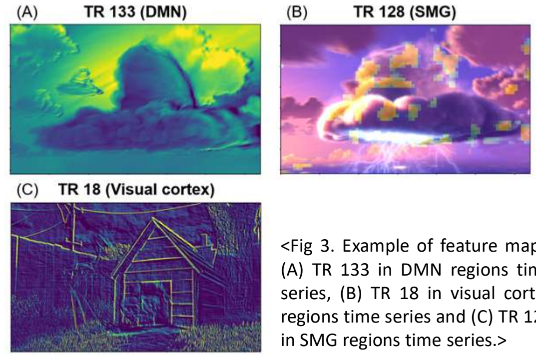
Results

- The 1st and 11th conv layer, which have the low correlation and are related to low- and high-level features, were selected for the group-level GLM analysis.
- In the GLM results of feature maps from the 1st conv layer, DMN and visual cortex regions were activated (Fig. 2A).
- On the other hand, the visual cortex and SMG regions were activated in the GLM results of feature maps from the 11th conv layer (Fig. 2B).



<Fig 2. The group-level GLM results from all feature maps extracted (A) from the 1st conv layer and (B) from the 11th conv layer.>

- As a result of analyzing the feature maps in the 1st conv layer corresponding to the peaks from the time series of the DMN regions in the fMRI data, the back-ground part of the video was emphasized in most of the feature maps (Fig. 3A).
- On the other hand, in the feature maps of the 11th conv layer corresponding to the peaks of the time series related to the SMG regions, the locations of the main characters were emphasized (Fig. 3B).
- In feature maps related to the visual cortex, the edges or boundaries of objects were emphasized regardless of the background and main characters (Fig. 3C).



<Fig 3. Example of feature maps: (A) TR 133 in DMN regions time series, (B) TR 18 in visual cortex regions time series and (C) TR 128 in SMG regions time series.>

Conclusion

- We investigated which features of video activate each brain networks by using VGG-16.
- The DMN, visual cortex, and SMG are each associated with different feature map patterns.
- This study would provide insight into understanding the brain's response to naturalistic stimuli.

Future Direction

- To identify patterns in feature maps associated with more diverse brain networks
- To apply this method to more diverse videos and more diverse age groups

Reference

1. Hasson et al. science. 2004;303.5664;1634-1640.
2. Richardson et al. Nature communications. 2018;9.1;1-12.
3. Simonyan et al. arXiv preprint arXiv:1409.1556. 2014.

Acknowledgments

This work was supported by (US) National Institute of Mental Health grants to Xin Di (R15MH125332) and Bharat B. Biswal (R01MH131335).