A Multimodel Approach to the Algonauts Challenge

Noah Syrkis & Sophia De Spiegeleire

June 1, 2023

Abstract

Developing a computational model of how the brain decodes visual information is an important goal in neuroscience. In this project, focus on improving the encoding model of the Algonauts Challenge. Our approach, rather than deepening the model, is to add a modality during training. Specifically, we add a vector of semantic features to image shown to the subject as the fMRI data is collected. We find that this improves the performance of the model.

Introduction

While the brain is a complex organ, perhaps most accurately conceptualized as a network of neurons, it is also a physical whose properties can be measured, on a more coarse level—in the case of this project, using fMRI. We use a subset of the Natural Scenes Dataset (NSD) Allen et al. (2022), provided by the Algonauts Project Gifford et al. (2023), to train a model that, given an image, can predict the fMRI response of a subject.

Understanding how the brain encodes visual information is an important goal in neuroscience. The Algonauts Project is a competition that aims to develop a computational model of how the brain encodes visual information. This is foundational research with potential applications in both neuroscience and machine learning.

The NSD consists of 73,000 images of natural scenes and various assoicated responses, collected over the course of one year from 8 subjects, making it the largest dataset of its kind, enabling the development of more accurate models, which are now released on an ongoing basis from various research groups.

Our approach is to add a modality during training. Specifically, we add a vector of semantic features of the image for the given fMRI data. Thouch multimodality, is a common approach in machine learning, recent advances in deep learning has largely been enabled by enormous amounts of data. In this project, we explore the potential of multimodality in the context of the Algonauts Challenge, attempting a model that, during inference, has the same number of parameters as the unimodal baseline model.

Literature Review

Decoding images from brain activity is a well studied problem in the field of neuroscience. The first successful decoding of images from brain activity was done by Haxby et al. (2001). Like the current project, Haxby et al. used fMRI data. Most recently Lin, Sprague, and Singh (2022) used a deep neural network to decode images from brain activity. Also Thomas, Ré, and Poldrack (2023) merits mention, focusing on developing a mapping between brain activity and mental states more broadly.

Data

The data used in this project is derived from the Natural Scenes Dataset Allen et al. (2022). The dataset consists of 73,000 images of natural scences and various assoicated responses, collected over the course of one year from 8 subjects. Specifically, the data used in this project is from the Algonauts Project Gifford et al. (2023). Associated with each subject are region of interest (ROI) masks. These masks are used to extract the fMRI data from the images, at specific locations in the brain. Six out of the eight subjects in the dataset have the same voxel count in borth hemispheres. Our experiment only uses the data from the left hemisphere of these six subjects.

The images are 254x254 pixels, and the fMRI data for the left and right hemispheres are 19,004 and 20,544 voxels respectively. In accordance with the Algonauts Challenge, compute each image's AlexNet features (**krizhevsky2012?**), and use these as the input to the model. Specifically, we use the features from the fc2 layer of the network, so as to keep our baseline as close to the Algonauts Challenge baseline as possible. In accordance with that baseline we also perform principal component analysis (PCA), reducing the dimensionality of the features to 100.

The fMRI remains as the Algonauts Challenge provides it.

Vectors of semantic features are provided by the Common Objects in Context (COCO) dataset (**lin2014?**). The COCO dataset consists of 328,000 images of 91 object categories, 80 of which are present in the NSD. Thus a sample of our data consists of the four-tuple (image, left fMRI, right fMRI, semantic features). The semantic features are 80-dimensional vectors, one for each object category. An image can contain multiple objects, so most semantic vectors contain multiple ones.

Methods

Our experiment, attempting to improve performance through multimodality without increasing inference paramters, tests the effect of predicting both the fMRI response and the semantic features from the image during training. The input is thus always the image (represented as the AlexNet features), and the output is either the fMRI response (during pure inference), or both the fMRI response and the semantic features (during training). Hopefully, this will allow the model to learn a more accurate representation

of the image, and thus improve performance, on the fMRI response, during inference, without increasing the number of parameters.

We use K-fold cross validation, with K=5, to evaluate the performance of our model, using the Pearson correlation coefficient as the metric. Our loss function is the mean squared error (MSE) between the predicted and actual fMRI response. We use the same model architecture for every subject and hemisphere, though the parameters are reinitialized for each subject-hemisphere pair.

The model architecture is a simple feedforward neural network, as our focus is on multimodality, rather than deepening the model, or exploring other architectures. The model has two outputs, one for the fMRI response, and one for the semantic features. From a neuroscience perspective, the fact of us predicting the blood oxygenation level dependent (BOLD) signal, for multiple regions of interest (ROIs), is already multimodal, as the ROIs are in different parts of the brain, and functuion by vastly different mechanisms.

Results

Discussion

Conclusion

References

Allen, Emily J., Ghislain St-Yves, Yihan Wu, Jesse L. Breedlove, Jacob S. Prince, Logan T. Dowdle, Matthias Nau, et al. 2022. "A Massive 7T fMRI Dataset to Bridge Cognitive Neuroscience and Artificial Intelligence." *Nature Neuroscience* 25 (1, 1): 116–26. https://doi.org/10.1038/s41593-021-00962-x.

Gifford, A. T., B. Lahner, S. Saba-Sadiya, M. G. Vilas, A. Lascelles, A. Oliva, K. Kay, G. Roig, and R. M. Cichy. 2023. "The Algonauts Project 2023 Challenge: How the Human Brain Makes Sense of Natural Scenes." January 10, 2023. http://arxiv.org/abs/2301.03198.

Haxby, J. V., M. I. Gobbini, M. L. Furey, A. Ishai, J. L. Schouten, and P. Pietrini. 2001. "Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex." *Science (New York, N.Y.)* 293 (5539): 2425–30. https://doi.org/10.1126/science.1063736.

Lin, Sikun, Thomas Sprague, and Ambuj K. Singh. 2022. "Mind Reader: Reconstructing Complex Images from Brain Activities." September 30, 2022. http://arxiv.org/abs/2210.01769.

Thomas, Armin W., Christopher Ré, and Russell A. Poldrack. 2023. "Benchmarking Explanation Methods for Mental State Decoding with Deep Learning Models." *NeuroImage* 273 (June): 120109. https://doi.org/10.1016/j.neuroimage.2023.120109.

Appendix