Encoding models of the human visual brain

Huayu Wang

July 16, 2023

Contents

1 Introduction **1**

2 Encoding models **2**

3 Algonaut **3**

1 Introduction

Building task-performing models of cognition is essential to understanding what computation underlies the brain. Cognitive psychologists and neuroscientists take a bottom-up approach attacking this problem: they formulate hypothesis about specific functions of the brain and devise experiments, either biological or behavioural to test them. This approach has been very successful, giving rise to detailed conceptual models of the brain, but is limited in computational rigor. Scientists came far at explaining psychological phenomena with brain physiology, but these knowledge are no where close to being able to mathematically explain the computations of the brain, let alone re-implement them. On the other hand, computer scientists and engineers take inspiration from human cognition and its features to build intelligent systems. Taking the top-down approach, they are able to create facinating task-performing models with stunning cognitive abilities, but lack biological plausibility. The models are often performance centered, and their tasks far from the what humans do.

In the last decade, the field of computational cognitive neuroscience emerged, aiming to narrow the gap between the two approaches. The challenge is to build bridges between task-performing computational models that are mathematically concrete and physiological or behavioral data. (*Nikolaus*

K. et. al. 2018) One way to achieve this is to build encoding models of neural representations. As high-quality data of neural recordings is becoming more availible, encoding models are better at capturing neural representations. Datasets like the Natural Scenes Dataset (NSD) (*Emily J. Allen et. al. 2022*) leverage state-of-the-art neural recording methods to provide high-quality, general data for training. Additionally, many fields in deep learning are seeing the emergence of foundation models which are large, general-purpose, underlying models that produce detailed features for downstream tasks. Rescently, DINO v2 overtook OpenCLIP (*Ilharco, Gabriel et. al. 2021*) as a new all-purpose model in vision (*Maxime Oquab et. al. 2023*).

Henceforce, there is much merit in exploring the neural encoding capabilities of DINO v2 using NSD as a fine-tuning dataset. For the purposes of this study, benchmarks and dataset processing tools are utilized as is described in the Algonauts Project 2023 (*Gifford AT et. al. 2023*), a competetion and framework for predicting responses in the human visual brain as participants perceive complex natural visual scenes.

2 Encoding models

Encoding models are a class of computational model that aim to predict neural representations from sensory stimuli. (*Liam Paninski et. al. 2007*) Formally, an encoding model assigns a conditional probability to a *r.v.* D representing any possible neural response in terms of its recording, such as fMRI, ECoG or behavioral experiments, given a stimulus S: $P(D|S)$. Because it is impossible to grasp the conditional probability distribution of D in its entirety, we hypothesize some model, $p(D|S,\theta)$, where θ is the *r.v.* for model parameters. Note that for the hypothesized model, inference is possible, *i.e.* the probability density function is known. The goal of an encoding model is to find the optimal parameters θ^* that maximizes the likelihood of the observed data, D^*, given the stimulus, S^*:

$$\theta^* = \underset{\theta}{\operatorname{argmax}}\, p(D^*|S^*,\theta) \tag{1}$$

Work has been done to improve encoding models of either the whole brain or regions of interest. Algonaut Project and Brain-score Project (*Martin Schrimpf et. al. 2018*) are competetions focused on prompting researchers to improve encoding models. The former is on a larger scale, but focuses more on performance. The latter is more focused on the biological plausibility of the models as it evaluates the model based on features, and benchmarks them with more than 50 datasets including 18 behavioral tasks by the time of

writing. This study uses the Algonaut Project 2023 as a framework because it utilizes the largest fMRI database to date, and provides great tools for data processing and benchmarking.

3 Algonaut

The Algonaut Project 2023 is a competetion and framework for predicting responses in the human visual brain as participants perceive complex natural visual scenes. (*Gifford AT et. al. 2023*) Participants are asked to submit predicted fMRI responses from previously unseen stimuli. The project provides a processed version of the Natural Scenes Dataset, which removes behavioral data and averages response of the same stimuli. The final evaluation metric is calculated with the correlation between predicted data and the ground truth. The project does not restrict model types and advocates for selection of different models for different ROIs.

The Natural Scenes Dataset is used for the evaluation of the task. (*Emily J. Allen et. al. 2022*) It is a collection of 7T fMRI recording of participants viewing pictures of natural scenes from the Microsoft Common Objects in Context (COCO) (*Tsung-Yi Lin et. al. 2014*) image dataset. 8 participants should view 10000 distinct images, each repeated 3 times. However, not all participants finished all sessions, causing a difference in data quantity. The participants are asked to perform a continuous recognition task which asks them to report pictures previously seen. This is designed to both keep them focused and to provide behavioral data. In Algonaut, the NSD is processed so as that there are no duplicate images, and the test data is removed. However, this may cause problems in the learning process of models because the participants usually have very different responses to the same stimuli at different times.