CS230

Decoding Images from Multi-Region, High Resolution, Electrode Recordings In the Mouse Visual System

Chris Fritz
Department of Electrical Engineering
Stanford University
cbfritz@stanford.edu

Category: Computer Vision, Neuroscience, Healthcare

1 Introduction

Here I propose using deep networks to probe the relationship between images presented to mice and the electrical activity simultaneously recorded at anatomically distinct regions in their visual system.

Until recently, direct neural data in the form of electrode recordings were limited to few (order of one hundred) measurements in a single brain region. Neuropixels are a recent hardware advancement that give not only an order of magnitude increase in number of electrodes, but also data from multiple brain regions simultaneously. A public dataset of neuropixel recordings in mice was recently made available online.

Deep networks are well-suited to meaningfully exploit the additional richness and answer questions about the function of data from different regions. For example, how does training a deep net on activity from one brain region perform compared to training on activity from another? On which types of images does each region perform best? In this way, I want to use deep nets as a probe to sense the potential role of a brain region in representing visual information.

2 Approach:

Data: The Neuropixels Dataset provides electrode waveforms (time series) and presented images in a well-documented data structure, the NWBFile. The data covers 53 experiments that sample a variety of natural (e.g. wildlife) and artificial (e.g. sinusoidal gratings) stimuli. The API covers preprocessing steps such as stimulus-electrode alignment that facilitate feature extraction and labelling.

Task: Network input is the rate of spikes recorded by electrodes over the past τ seconds, where τ is an adjustable hyperparameter. The network maps a vector of spike rates to a vector of pixel intensities of the presented visual stimulus. In the past, Neural networks have decoded images from retinal neurons, suggesting images can be accurately decoded from neural activity.

Challenges: Feature extraction from the neural activity will pose the greatest challenge as it will require domain knowledge in visual neuroscience. The second challenge is to develop a network architecture that handles both temporal and spatial data effectively. This will likely require a hybrid approach using both CNN and RNN architectures.

Network Performance: The network objective will measure the similarity between the predicted images (greyscale pixel data) and the presented images. Two metrics will be considered. The mean

CS230: Deep Learning, Winter 2018, Stanford University, CA. (LateX template borrowed from NIPS 2017.)

square error for image y and prediction $\hat{y} \in \mathbf{R}^d$:

$$MSE = ||y - \hat{y}||_2^2,$$

measured the euclidean distance between the vectorized images. A metric that is sensitive to features perceptually relevant to the human visual system is the *Structural Similarity Index:*

$$SSIM = l(y, \hat{y})^\alpha \cdot c(y, \hat{y})^\beta \cdot s(y, \hat{y})^\gamma,$$

where l measures luminance (mean pixel value), c measures image contrast (variance of pixel value), and s measures image structure (correlation between pixel values), and α, β, and γ are constants.