Neural Neural Networks

Ella Mohanram (emohanra), Patrick Jennings (pkjennin), Taha Ebrahim (tebrahim)

Github Repo: https://github.com/ella-mo/dlfinalproject2025

Introduction

In this project, we aimed to explore the correlation between layers of CNNs trained on psychophysics behavioral tasks and layers in the visual processing pathway through recorded electrophysiological (EP) data. This research question falls under interpretability of classification models. We were inspired by the paper, "Performance-optimized hierarchical models predict neural responses in higher visual cortex," by Yamins et al. However, because this paper does not provide code or data, we used the data from "Feature-selective responses in macaque visual cortex follow eye movements during natural vision," by Xiao et al.

Pre-existing models were either specialized for optimization of image classification or prediction of area IT activity. In Yamins et al. (2014), the hierarchical modular optimization (HMO) model has both high task performance and neural activity area predictivity. Specifically, as shown by Figure 1, the HMO model is better at classifying images with high variation, such as different camera positions and lighting.

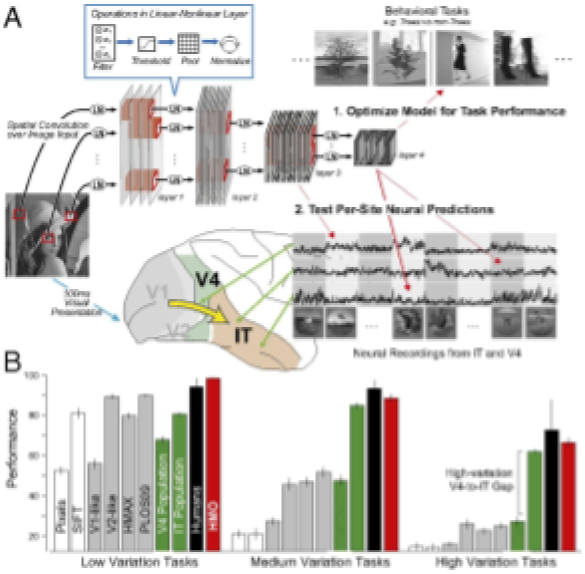

Figure 1 Figure 2a and 2b from Yamins et al. (2014) demonstrating that the HMO model, which has high categorization performance also has high IT predictivity.

However, since neither image nor EP data is available in the paper, we aim to explore whether a similar model can be replicated with different data. In Xiao et al. (2024), we found both task and EP data recorded from six areas in the macaque monkey ventral visual processing pathway. However, the data was collected from free-viewing tasks, so the data are trials of fixations and saccades of several thousand images. We decided to use the data on fixations and the image classification as face versus non-face. Moreover, we used labels of pareidolia stimuli to mimic the high variation of the data used in Yamins et al. (2014).

Methodology

We are using the DANDI dataset from the study "Feature-selective responses in macaque visual cortex follow eye movements during natural vision", available via DANDI and documented in the associated GitHub repository. This dataset includes neural recordings from macaque monkeys viewing natural images, along with corresponding eye-tracking and TTL event data. The visual stimuli are packaged as multi-part zip archives (Stimuli.zip and Stimuli.z01) due to OSF file size restrictions. These can be extracted using tools such as 7z or PeaZip to access the original image stimuli.

Significant preprocessing is required to align the data temporally and spatially—specifically, to determine what image the monkey is seeing at each time point, and to identify which regions of the brain are activated in response. This involves parsing event timing, matching eye-tracking data to visual stimuli, and interpreting neural signals from different brain areas. In particular, due to the size and sparsity of the raster data, significant computational power is required.

The architecture of the HMO model in Yamins et al. are two stacks of three-layer CNNs that have been combined in parallel, with a 1250 top-layer output. These layers of the individual CNN outputs are then stacked. The original HMO model uses boosting and hyperparameter optimization, which was computationally expensive and not possible within the scope of this project.

After training the model on classification of images, while measuring accuracy, we plan on determining how well the model replicates the human/monkey brain on similar classification tasks. After each the final layer of the HMO, we take the 1250 top-level outputs of the layer for each image, and compute a representational dissimilarity matrix (RDM). We also create an RDM consisting of monkey neural data on the same images (in the same order) as the HMO model. We compare these RDMs using Spearman's rank correlation to get oen metric of the overall similarity between the neural output and our models output.

From here, we can run multiple experiments. The primary evaluation metric that Yamins et al. considers is the Pearson correlation coefficient, r, which seeks to determine how related the neural responses are to our layer outputs. While r^2 to determine the level of variance in neural response explained by our model outputs at each layer. To calculate r^2, we used partial least squares regression (PLS), which is the same method they used to evaluate it. They fixed the number of components at 25, but we chose to use 5 because we had a significantly smaller set of images than they did. You generally want fewer components if you have less data.

Results

We had mixed results when evaluating the relationship between our model and the neural data. The first test, the spearman correlation coefficient and r^2 did not give us as promising results as we would have liked. We calculated r^2 values for each measured neuron with PLS, and the median was just under 0, indicating that our model outputs do not accurately predict the firing rates of individual neurons.

The second metric produced more promising results. For both the neuron firing rates and the HMO model outputs, we calculated the RDMs. One can see that the RDMs look fairly similar to each other, and when using Spearman's rank correlation, we calculated a correlation coefficient of r = 0.527. This indicates that the model outputs represent images similarly to how the neurons fire together. In retrospect, it is not surprising that we achieved better results with the RDMs because they take into account the relationship between all of the measured neurons as opposed to the r^2 values which are much more subject to the noise and unreliability of single-neuron data.

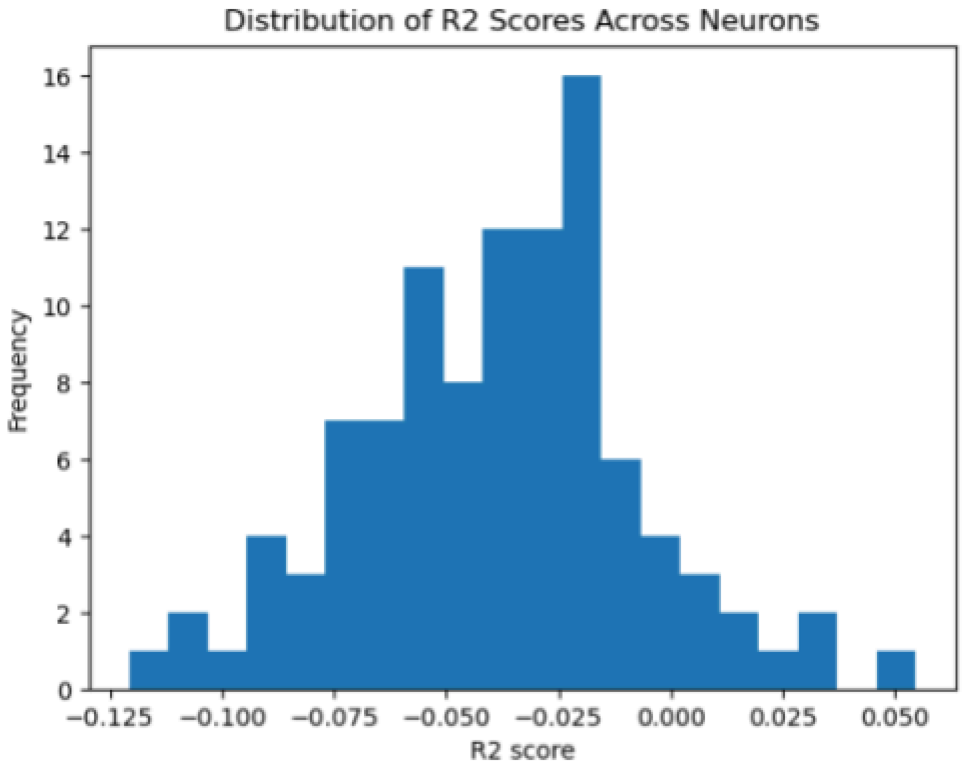

Figure 2 Histogram of r^2 values

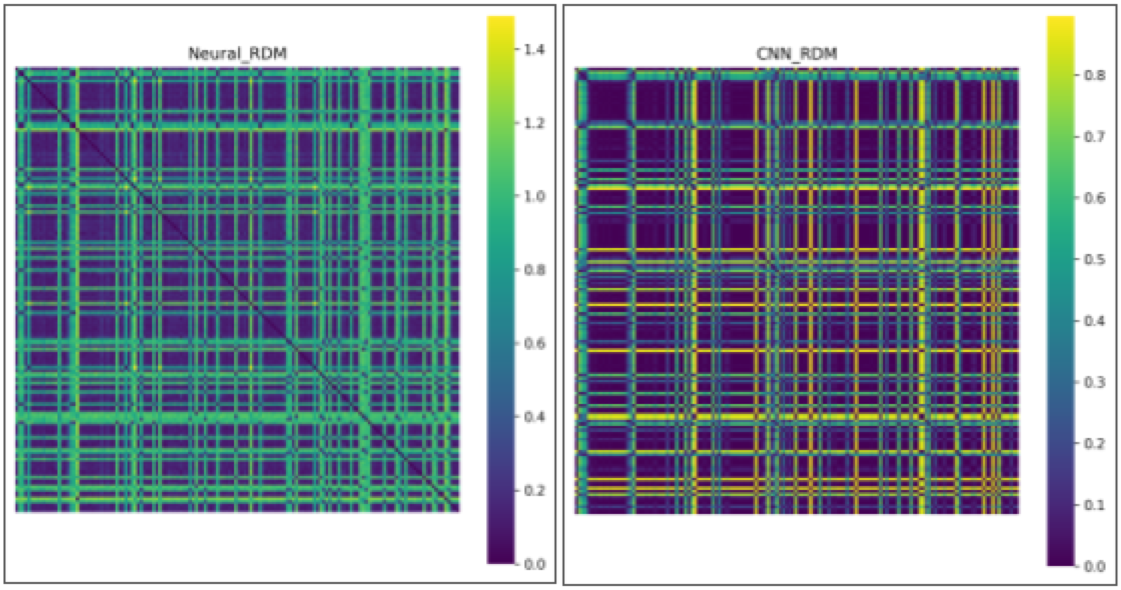

Figure 4, Figure 5 The representation dissimilarity matrix (RDM) for both the neural data and the CNN model architecture. A model that processes images like the brain should have a similar RDM to that of the brain.

Challenges

The data has been the most difficult and frustrating part of the project. Especially since we had to diverge from the original paper due to the data availability, it had been difficult to figure out what data would be worthwhile to use. Trying to make sense of the GitHub repo, especially in relation to data preprocessing, was a major hurdle. For example, the images in the zip files were different sizes, aspect ratios, and dimensions. The image dataset was also not clearly labelled. Creating our own new dataset while manipulating and padding data, adding pareidolia images ended up being much more of a challenge than we anticipated.

Moreover, using raw EP data and manipulating it into a form that is useful for validating it against the layers of the model is also difficult. The sheer size of the raster data makes processing difficult and time consuming. Moreover, deciding a normalization metric to compare neuronal responses between images was not a straightforward process as the result also needed to be compared to the CNN output.

There were also some challenges in creating the model. It is somewhat tricky to run the CNNs on a classification task, but only save the layers up to the point before. We also had to manage variable sizes of CNN outputs due to the randomization, which required resizing the images

Reflection

Ultimately, we are very satisfied with how our project turned out. Synthesizing individual components was certainly a challenge, and we met many roadblocks along the way. We were able to build a working model that classified images with an 83.64% accuracy, and we were able to analyze the model's outputs in comparison to raster data. At the time of the previous check-ins, it was challenging to assess our base, target, and stretch goals because it was somewhat unclear exactly how different our data was from that of the paper we re-implemented. Looking back, one of the stretch goals was to achieve the same r^2 value (~50%) as the paper. Unfortunately, we did not achieve this because our r^2 was around 0. However, in retrospect, this was quite the tall task, and we are a little suspect of p-hacking on their part to achieve such a high value. Nonetheless, we did achieve success in different metrics that Yamins et al. (2014) used. In the representation dissimilarity matrix (RDM) plots, we showed that the 1250-dimensional representation of an image (from our HMO model) compared to another image was similar to that of the neural data. We found a total Spearman correlation coefficient of 0.527 between the two RDMs, indicating a moderate positive correlation between the two representations.

We had to make a few adjustments to the model from the original paper, but overall it worked about as we expected. One modification we made was that instead of randomly sampling CNNs for layers N2 and N3 in the model like we did in N1, we simply trained them together. We made this choice to avoid the high computational cost of training hundreds of CNNs that take hundreds of filters as input. Additionally, training them together combines the good individual accuracy from the first layer with a targeted method of achieving high total accuracy of 83.64%.

At the beginning of the project, we were heavily leaning towards using a LIME-style interpretability test of the HMO model to detect what layers detected certain kinds of facial/non-facial features. We ultimately decided against this because the neural data did not align well for this. We had binary neuronal firing data, which would be difficult to compare to LIME interpreted images. Instead, we followed the paper more closely, comparing the results of the HMO model to the neuron data from Area-IT, which is primarily responsible for facial recognition. If we were to repeat this project, we would likely want to figure out a way to clean and normalize the neuron data a little bit more. As our low r^2 values indicate, it is very challenging to predict extracellular single- and multi-unit neuronal activity when viewing an image. This is possibly due to the noisy and stochastic nature of neurons, but it is also possible that this is just a poor way of evaluating the model. If we were to repeat the project, we would also like to lean more heavily on different evaluation metrics which take into account more neurons at a time such as the RDMs.

If we had more time, we could maybe clean up the neural data a little bit, but that would be a tall task considering neuronal data is notoriously noisy and challenging to work with. One interesting avenue to take if this were a much longer research project would be to compare the HMO model to other architectures. It would be interesting to see how it performs against a simple CNN, but also if a more modern architecture such as a vision transformer produces better results. It would also be interesting to examine the intermediate results of the HMO model as they compare to other areas in the brain (V1, V2, and V4) which are known to be earlier in the image processing phase. We opted against this in this project since the paper's results are not as concrete as for IT/final outputs, but it would be a fascinating future project.

One of our biggest takeaways from this project is that research is a difficult process. Working with large data (in our case 50GB!) can be very frustrating and confusing. Moreover, interpreting and manipulating neuronal data is also a significant challenge. However, once it is all said and done, it was very rewarding to have a final product that we were knowledgeable about and research that helped us learn a lot more about Deep Learning as a whole!

Bibliography

Cadieu, Charles F., et al. "Deep Neural Networks Rival the Representation of Primate IT Cortex for Core Visual Object Recognition." PLOS Computational Biology, vol. 10, no. 12, 2014, e1003963. https://doi.org/10.1371/journal.pcbi.1003963

D.L.K. Yamins, H. Hong, C.F. Cadieu, E.A. Solomon, D. Seibert, & J.J. DiCarlo, Performance-optimized hierarchical models predict neural responses in higher visual cortex, Proc. Natl. Acad. Sci. U.S.A. 111 (23) 8619-8624, https://doi.org/10.1073/pnas.1403112111 (2014).

Xiao, W., Sharma, S., Kreiman, G. et al. Feature-selective responses in macaque visual cortex follow eye movements during natural vision. Nat Neurosci 27, 1157–1166 (2024). https://doi.org/10.1038/s41593-024-01631-5