Decoding semantic features from different ROIs

By: Furkan Özçelik, Aditya Prakash, Amirreza Nadimi Shahraki, Oğul Can Yurdakul

Pod 118 Hairy Grasshoppers – Kaysonian Rhapsody

neuromatch academy

Research Question

- Our aim was to discover the relationship between the **voxel responses** in the visual cortex and the **semantic properties of the images** shown.
- We wanted to approach the problem in two ways:

1. The semantic representative roles of different Region of Interests (ROI): **In which ROI is the semantic content of images represented?**

2. Representative performance of the voxel responses: **How does the decoder performance vary with different images?**

Dataset: Kay & Gallant Dataset

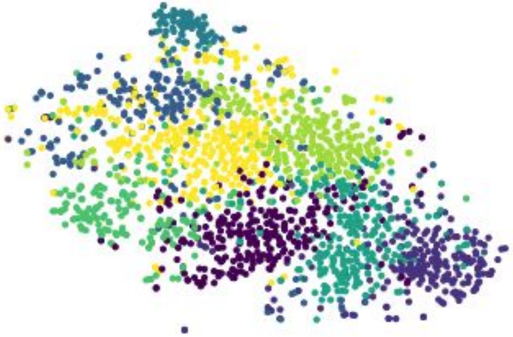
- Blood-oxygen-level-dependant (BOLD) fMRI signals recorded of a passive grey scale stimuli (image size 20x20) by (Kay et al., 2008).
- Data has been preprocessed to contain a 2-D matrix of stimuli correspondence (row) and voxels in visual cortex(columns).
- Each point in this matrix represents the voxel response to a particular stimulus image.
- Dataset contained number of voxels per region of interest in visual cortex.
- Processed BOLD Data was used by (Naselaris et al., 2009) for an image reconstruction task.

Approach: Semantic features

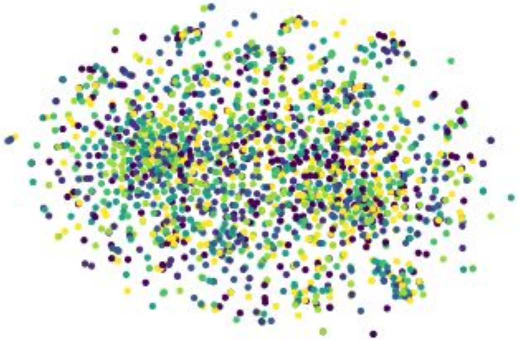
- At first we wanted to **categorize the images ourselves**, which was overwhelming. We tried finding a classifying DNN, but could not.
- Then came **the labels for the dataset**, and we wanted to use those as categories. The third layer of the labels were initially promising.
- We also tried extracting **semantic features using different classifier DNN's last layers, namely those of Resnet50 and VGG16,** and we ended up using those.

Approach: Dimensionality Reduction

- We tried discovering some **clusters** in the semantic features using dimensionality reduction. We reduced the dimensionality of the voxel responses to 2 and visualized by
    - Voxel response amplitude,
    - Different levels of labels
    - Amplitudes of the semantic features extracted from the DNN
- We also tried clustering semantic features (from DNN) using **k-means similarity** and then using those clusters to find grouping in voxel responses.
- We decided we could not work with this information, so we gave up on it.



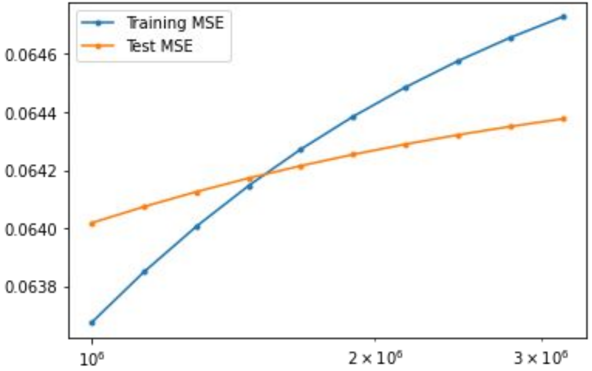
Clustered Semantic Features



Voxel Activity When the above Clusters where used for grouping.

Approach: Fitting a Linear Model

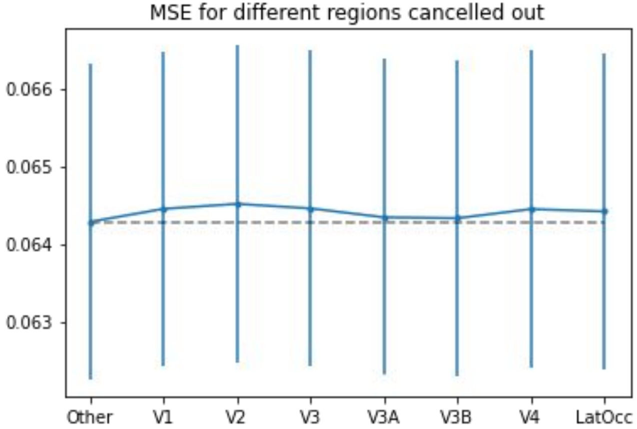
- Being the simplest model, **linear regression model** was the first thing we tried. Although we had a near zero Mean-Squared Error (MSE) in the beginning, we then realized it was due to **overfitting**. The overfitting problem did not let us go until this day (we think). As a side note, matrix inversion by hand took a significant amount of time, so we moved to using **sklearn**.
- So, we moved to **Ridge, Lasso and ElasticNet** models for regularization, to prevent overfitting. We tried to find an alpha value that gives the closest MSE for both test and training sets. Lasso model did not converge, so we ended up using the Ridge model only.



Model MSE for different alpha values for a Ridge moel

Approach: Cancelling different ROIs

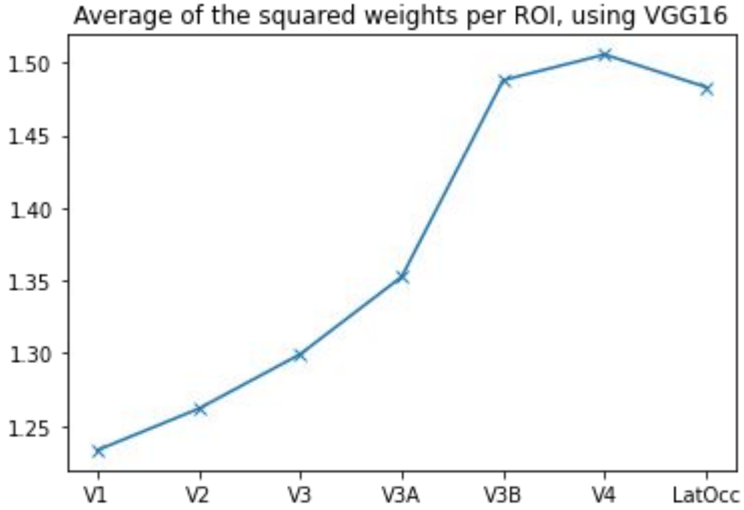
- We abandoned the idea of testing model performance on different semantic categories, because **we could not manage to reduce the number of categories to a manageable number**. So we moved on with our first question.
- We tried **zeroing the voxel intensity of each ROI** to cancel their effects on a certain total model. Although at first we thought we had found a pattern common for both features extracted from Resnet50 and VGG16, but quickly realized that it was due to the **imbalance in the number of voxels per ROI**. So, we balanced them out by randomly selecting 314 voxels, training the model with those



Model performance with different ROI's cancelled, with an unbalanced dataset. "Other" corresponds to the full model with no ROI cancelled out.

Approach: Weights of each ROI

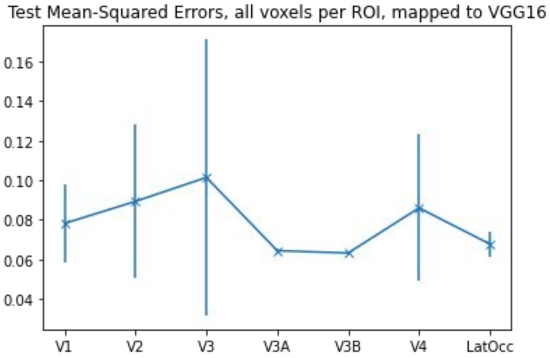
- We checked the **average of the squared weights corresponding to each ROI** for a model trained using all of the voxels, to observe the significance of each ROI. We believe we may have some meaningful results: Just like in deep networks, we would expect deeper levels to have higher level content, and thus **we would expect our model to assign more significance to them**. However, we could not test this idea with a randomly balanced dataset.



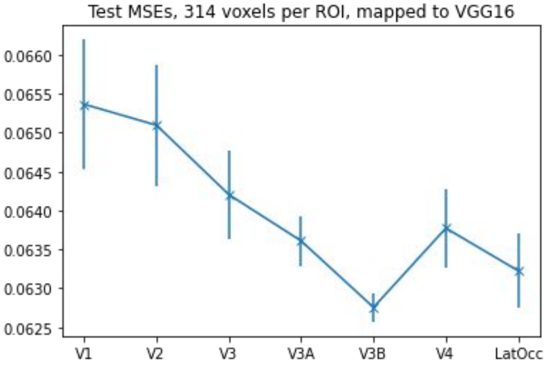
Significance of model weights per ROI. The model is trained with different Ridge regularization constants and the final weights were found by averaging these different models.

Approach: Model on each ROI

- We tried fitting a **model using only the voxels in each ROI**, so we had a different model for each ROI. We again tried this with an unbalanced and balanced number of voxels per ROI. Again the model trained on the **unbalanced dataset follows the number of voxels per ROI**. The balanced one, however, seems promising in terms of results. **We would expect deeper levels to have higher level content, and so having the MSE drop as we progress seems reasonable.**



Performance of models trained on individual ROIs, with an unbalanced dataset.



Performance of models trained on individual ROIs, with a balanced dataset.

Future implications:

- Kay Gallant Dataset have grayscale stimuli presented to subjects. We used Imagenet pretrained DNN models, they were trained with RGB images. Studying on dataset with colored stimuli would provide better image features such as Generic Object Decoding Dataset (Horikawa and Kamitani, 2017)
- If we had more time and resource we could train generative models such as Variational Autoencoders or Generative Adversarial Networks to get a visual reconstruction model, which is easier to examine with inspection.
- All ROIs do not have the same number of voxels, this could be affecting the regression results. We may find a better regression model to handle imbalanced voxel numbers across ROIs