KayGallant Dataset Project

The main idea was to gather both stimuli and responses to meaningful low dimension using dimensionality reduction algorithms (Deep Neural Networks as well for stimuli) and then analyze the semantic information by examining the corresponding pairs in low dimension visual

When we think about the region of interest, we can create a research question like which brain regions are providing more relevant semantic information similar to deep neural networks providing us

fMRI responses in human visual cortex during natural image viewing

- **Intro video**
- Colab Notebook example for loading the data
- The paper describing the data set / PDF

Relevant Resources

- Reconstructing faces from fMRI patterns using deep generative neural networks
    - Reconstructing faces from fMRI patterns using deep generative neural networks
    - I know it says deep generative neural networks in the title but I want to emphasize what he does also with PCA. I can summarize this paper in a meeting if necessary.
- To examine connected papers to the main paper
    - https://www.connectedpapers.com/main/12ec492b6b2ad7c2cc47a5612e8d2a9b7fb35f91/Identifying-natural-images-from-human-brain-activity/graph
    - Revealing representational content with pattern-information fMRI--an introductory guide
        - https://www.semanticscholar.org/paper/Revealing-representational-content-with-fMRI--an-Mur-Bandettini/70a13b2aa84cb076cb06919b54e0df22b0b74760 (ALL)
        - Found this source on connected papers link above seemed like it can give us some information about representation on fMRI
        - -Furkan: I think this is more about technical things on fMRI which is not relevant for our study. Activation-based analysis: Investigating the involvement of regions in a specific mental activity, compare spatial-average activation across conditions. Pattern-information analysis: Investigating the representational content of regions, comparing patterns of activity across conditions.
    - Encoding and Decoding in fMRI

        - [PDF] Encoding and decoding in fMRI (ogul)
        - Another paper that Kay and Gallant involved discusses encoding and decoding in fMRI
    - Deep Residual Network Predicts Cortical Representation and Organization of Visual Features for Rapid Categorization
        - https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5830584/pdf/41598_2018_Article_22160.pdf (Furkan)
        - Furkan: They trained a DNN for encoding which tries to predict cortical responses to objects in movies. Representations reflect to multiple levels of object features and semantic relationship between categories. In the visual cortex, object representations are organized in the three clusters of categories which are biological objects, non-biological objects and background scenes. They have subclusters as well and this creates a hierarchical clustering. Encoding accuracy is high for ventral areas than dorsal areas. ResNet shows better results than AlexNet. They have measured correlation between cortical similarity and semantic similarity. CNN-based encoding becomes "virtual-fMRI" in a way.
    - Neural Encoding for Human Visual Cortex with Deep Neural Networks Learning “What” and “Where”
        - Neural Encoding for Human Visual Cortex with Deep Neural Networks Learning “What” and “Where” (Amir)

Amir: In this paper

- We provide a new perspective on the deep-learning-based neural encoding models, performing receptive field estimation and features regression simultaneously in a deep neural network. This modeling approach can yield explicit receptive fields (“where”) and feature turning functions (“what”) automatically, which is rich in interpretability.
- The estimation of receptive fields is endowed with weaker constraints. Instead of strong prior assumptions on the shape of receptive fields, L1 regularization and Laplacian smoothing are adopted in our modeling approach, which can be regarded as weak prior assumptions about receptive fields.
- We made an attempt in the extension of the modeling approach. In consideration of the computational similarities between voxels, the voxel-wise modeling approach is extended to multi-voxel joint encoding models, suggesting a new approach to rescuing voxels with poor signal-to-noise characteristics more effectively.
- Extensive empirical evaluations on the publicly available fMRI dataset demonstrate that our modeling approach achieves superior performance compared with other neural encoding models.

Where: location and extent of pooling over visual features

What: feature selection property of neuron populations in visual cortex

Feature-weighted receptive field : It starts with a natural image, obtains feature maps 34 in a pre-trained convolutional neural network, and computes a weighted sum within the spatial 35 extent of an 2D Gaussian receptive field. Finally, it regresses all the feature maps onto brain 36 activity simultaneously, which yielded the state-of-the-art prediction accuracy. [The Feature-Weighted Receptive Field: An Interpretable Encoding Model for Complex Feature Spaces]

Receptive field estimation:

Multiple features regression:

Regularization: sparsity and smoothness regularizations

Method:

1) **Nonlinear feature extraction:** firstly, converting the visual input to its feature representations and then projecting the feature representations onto activities at each voxel h(i) denotes the features extracted from input image X(i) and the question is are these features meaningful? In reality, we do not know what features can better explain activity in the visual cortex under many circumstances. In this way, the total feature maps h(i) are able to contain adequate features to capture the reasonable hypotheses about what is encoded in the visual cortex. Based on these original feature maps, the proposed method can infer which features and locations are important for explaining the activity in the voxel. In the nonlinear feature refinement, the feature map pixels are focused on within the spatial RF, as concentrating on important locations and suppressing unnecessary ones are conducive to improving the encoding performance. In this way, spatial RF is the “where” parameters of the neural encoding model. In the voxel-wise linear mapping, each feature map will be assigned an associated feature weight, which indicates the importance of the feature map for predicting the activity of each voxel. In this sense all the feature map weights are “what” parameters of the neural encoding model. Noting that while original features maps are same for each voxel, but the feature refinement and feature weights vary across voxels.

2) **Nonlinear feature refinement (converting original features into refined features with channel attention module and spatial receptive field (RF) module):**

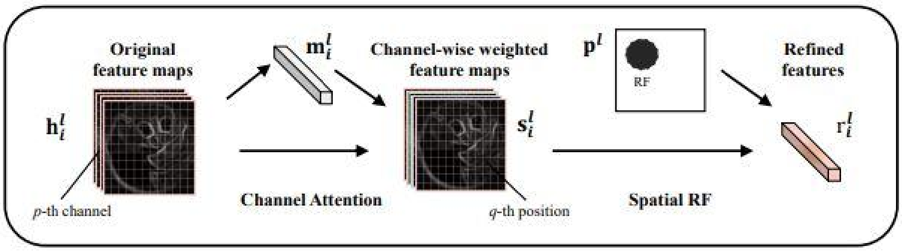

Fig 3. The feature refinement in the specific layer (the l-th layer). In the channel attention module, the original feature maps $\mathbf{h}_i^l$ are initially used to obtain the channel attention weights $\mathbf{m}_i^l$. In the next phase, the original feature maps are element-wise multiplied by the channel attention weights to obtain the channel-wise weighted feature maps $\mathbf{s}_i^l$. In the spatial RF module, the receptive field $\mathbf{p}$ is reshaped to the corresponding receptive field $\mathbf{p}_l$ according to the size of channel-wise weighted feature maps $\mathbf{s}_i^l$. The Hadamard product of $\mathbf{p}_l$ and $\mathbf{s}_i^l$ finally produce the refined features $\mathbf{r}_i^l$.

3) **Voxel-wise linear mapping (regressing refined features simultaneously onto voxel activities):**

- This work seems similar to what we have in mind

- In order to obtain reduced dimension representations of images

    - We can use pretrained DNN + Dimensionality reduction algorithm (PCA, T-SNE, UMAP)
        - An example: Using T-SNE to Visualise how your Deep Model thinks (Aditya)
          In this Paper:
          The paper is basically about what t-SNE does like Dimension reductionality and Some visualizations of how it happens.
          For better understanding t-SNE is Visualizing Data using t-SNE is a good paper.
        - Another example: Neural Network Feature Visualization » Deep Learning - MATLAB & Simulink
            - The thing they have done with kNN is also looking good. We can do a similar study after linking low-dim stimuli to low-dim fMRI
- Here's a link to an image identification algorithm we used this semester:
    - Overview of the YOLO Object Detection Algorithm | by ODSC - Open Data Science
    - It was quite good with natural images and it is already trained. We can use other alternatives such as AlexNET or smt but this was quite nice. It also gives many tags for different objects in the same image, so it may be useful that way as well.
- I found a very short read regarding different regions (V1-V6) of the visual cortex. It's not as detailed as I hope it would be, but may prove to be a start if we're interested:
    - Neuroanatomy, Visual Cortex - StatPearls (ALL)
    - This might also prove to be useful, at least in the references: https://en.wikipedia.org/wiki/Visual_cortex#Primary_visual_cortex_(V1)

- Gabor wavelet pyramid algorithm: https://www.researchgate.net/publication/220869739_A_Gabor_Wavelet_Pyramid-Based_Object_Detection_Algorithm#:~:text=A%20Gabor%20Wavelet%20Pyramid-Based%20Object%20Detection%20Algorithm.%20A,list%20of%20authors%29%2C%20clicks%20on%20a%20figure%20

Ideas

- We can advance question for low level and mid level features as well
- Fine-tuning with Imagenet after doing some preprocessing to make similar to kay gallant

Proposal

- The scientific question:
  How is the voxel activity in different brain regions related to the high-level/semantic features of the images?
- Brief scientific background:
  We know that there are certain brain areas that respond significantly more when the subject is exposed to certain kinds of stimulus, such as the Fusiform Face Area with the faces. We would like to investigate if such patterns or specializations occur for different kinds of high-level features as well.
- Proposed analyses :
  - Transform images to high-level features using DNNs.
  - Regression between the high-level image features and the voxel activity.
- Predictions:
  Possible alternative outcomes:
  1. High-level image features and voxel activity localisation (brain region) are correlated.
  2. For all features, the voxel activity is not localized but spread through all regions of interest/brain regions.
  3. Some features can be localized and some may not be. For those who are not localized in the regions of interest given in the dataset, they might be localized in some other region that is not included.
- Dataset: Kay & Gallant dataset

To do :

- Find a DNN (find a grey scale DNN) to extract the image features to be fed to the main model that predicts voxel value
  - Grayscale DNN or Transform grayscale images to something compatible with RGB models
- Write a linear regression model that takes image features as input and outputs the voxel activity value
- Dimensionality reduction can be used to visualize image features and voxel activity

Tasks :

- Finetuning of Deep Neural Network for Grayscale transformed ImageNet features to mimic stimuli and get meaningful transformation to obtain high level features for stimuli
- Research about encoding metrics and measurements: What quantitative analyses are used in other studies.
- Exploratory analysis on fMRI data: How should we use given fMRI data, should we do any prior processing, how Kay Gallant fMRI data used in other studies.
- Dimensionality reduction for visualization of transformed image features and voxel activity