Neural Representations and Categorization of Visual Input in the Brain

Prachi Mahableshwarkar

Dietrich Senior Honors Thesis 2019

Introduction

The functional organization of the visual processing system effectively serves to solve difficult recognition and classification problems. Given a large amount of information, the visual processing system is able to retrieve critical bits and produce a salient understanding of the world. This impressive, calculated ability is supported by neural substrates that have been well defined through extensive research. The seemingly spontaneous recognition of objects and classification of scenes results from a series of computations through the visual pathway. In order to understand how these complex abilities are supported by neural substrate, research must target the individual abilities and contributions of regions that process visual information. Understanding the functionality of areas in the visual processing stream will elucidate the computational mechanisms of recognition and classification.

Background: Visual Pathways

Starting in the retina, visual information passes through the optic nerve after which the optic tract carries information to the lateral geniculate nucleus (LGN) (Remington, 2014). From the LGN visual information is passed into the visual cortex of the occipital lobe. Functional analysis of the primary visual cortex has revealed that V1 neurons are orientation and direction selective (Lee, 2003). This means that they have strong responses to lines, bars, or edges of a specific orientation. In addition, V1 neurons are direction selective and respond strongly to certain lines, bars, and edges moving in a preferred direction. After primary visual cortex are about thirty secondary visual areas that are hypothesized to have functional specialization. This hypothesis follows that specific brain areas perform specific visual functions. The fusiform face area (FFA) for example is hypothesized to be specialized for faces, the extrastriate body area (EBA) specialized for human bodies, and parahippocampal place area (PPA) specialized for

places and scenes (Peelen et al., 2005). Defining these visual areas and their functions has been accomplished through a variety of physiological and functional magnetic resonance imaging.

The computational and neural mechanisms that support categorization and recognition however are still greatly debated upon. Some researchers claim that the visual system is comprised of category-specific modules that perform a specialized computation for a class of stimuli, such as faces (Dailey, 1999). A wealth of neuropsychological and neuroimaging evidence has been put forth in favor of the modular organization with ventral visual pathway specialization for high-level perceptual analysis of faces, objects, places, bodies, and words (Kanwisher, 2010). Proponents of this school of thought often use lesioned patients such as prosopagnosics to corroborate their argument. Acquired prosopagnosics have a bilateral or right occipitotemporal cortex lesion, causing severe face processing impairments (Behrmann, 2011; Duchaine et al., 2005). While this may point to a discrete categorical face impairment, further investigation reveals that prosopagnosics suffer from a degree of object processing damage as well. Prosopagnosia could be then defined as a manifest of an underlying visual computational deficit as opposed to the theory that face processing is a special mechanism, which is broken in prosopagnosia. This points to a more distributed organization wherein cortical regions are only partially specialized, allowing for a more connectionist perspective on the visual processing system. Neuroimaging research corroborates this viewpoint by demonstrating graded representations of stimuli and the shared cortical activation of supposed specialized regions such as the OFA (objects) and FFA (faces) (Plaut & Behrmann, 2011). This question of a distributed versus modular organization lends to the question of how connections in each step of the visual pathway build categorization of visual input and support visual object recognition. It has been proposed that the ventral temporal cortex is involved in categorization (Grill-Spector et al.,

2014). Specifically, it has been argued that a spatial representational hierarchy of visual information enables access to levels of categorization. The lateral occipital complex is an integral substrate in this process and is the focus of this study.

Lateral Occipital Complex

Functional brain imaging research has clarified that the lateral occipital complex (LOC), which is functionally defined as a region within the occipital and temporal cortex, plays a central role in human object recognition (Grill-Spector, 2001). An early study that began to clarify the role of LOC compared activation in this cortical region for different types of stimuli (Malach, 1995). It was found that the LOC responded more strongly when subjects were viewing photographs of common objects compared to when they looked at visual textures that did not have discrete, defined shape interpretations (Malach et al., 1995). Furthermore, LOC demonstrated a lack of preference for familiar versus unfamiliar objects. The magnitude of response in this region was no different for objects that were known, such as a dog, and objects that were not known (e.g. Henry Moore sculpture). This phenomenon was further explored by Allison, Puce, Spencer, & McCarthy (1999), using event-related potentials recorded from electrodes that were placed on the cortical surface of patients before surgery. Similar to the previous studies conducted, there was a distinct difference in waveforms for objects compared to non-objects. The studies found object-specific waveforms with a higher activation for categories of objects such as cars, flowers, butterflies etc. when compared to the activation for scrambled control stimuli that did not have a defined 3-D structure. Selective activation of LOC for the Gestalt cue of surroundedness has also been found by fMRI studies, alluding to the computation that supports object-level processing (Appelbaum et al., 2010). The spatial configuration of the object as well as the larger background selectively activates LOC.

Functional specificity in this region for object identification however cannot be assumed since activation could be due to visual attention, arousal, figure-ground segmentation, or other visual factors. On the other hand, neuropsychological research has found that lesions to this cortical region result in a variety of recognition deficits (Moscovitch et al., 1999).

The complexity of object recognition considering viewpoint changes, illumination, viewing angles, and categorization makes it difficult to understand, and that much more impressive. The visual recognition system's diverse capabilities leads to the question of how representations of objects of certain categories are generated to allow generalization and fine discrimination. While many studies have concluded that the LOC plays a significant role in object recognition it is unclear in what capacity and to what degree object recognition is accomplished (Grill-Spector, 2001). Specifically, it is of interest to understand what role LOC plays in subordinate categorization of visual input and whether category specific mechanisms exist. Given a certain category of object input, does LOC generalize across the category or perform finer discrimination?

In order to appropriately classify category-selectivity within LOC, it is important to consider the nature of defined category-selective regions in the cortex in terms of functionality and anatomy. Categories such as faces have been extensively studied with the conclusion that the fusiform face area responds more strongly to face stimuli when compared to nonface control stimuli. Kanwisher (2010) argues for localization of detection and recognition of faces in the FFA based on behavioral, neuropsychological, and neuroimaging evidence. The PPA is another seemingly category-selective region of the cortex that responds strongly to a diverse spread of stimuli that depict places or spatial layouts including outdoor scenes, indoor scenes, and houses (Walther et al., 2009). Aside from functionality, it is important to note the anatomical features of

these category-specific regions in relation to LOC. The PPA is clearly discrete anatomically from LOC, however the fusiform face area is in close proximity to and partially overlapping with the anterior part of the LOC (K. Grill-Spector et al., 2001). Grill-Spector speculates that this overlap however does not discount the selectivity of the FFA, and the preference for objects to scrambled objects in the LOC. Rather, this overlap reveals the issue with functionally defining cortical regions and the challenges they pose for interpretation. Many regions within the ventral visual pathway have been classified in terms of selectivity but the LOC has not exhibited specificity for a certain category, which lends to the prediction that the LOC serves as a region for a general-purpose mechanism for objects. The specific information that is leveraged as well as the computation that occurs in the LOC is unclear and requires further investigation. Another hypothesis is that the pattern of activity observed across the ventral visual stream, which includes the LOC, FFA and PPA, forms a distributed network that supports object recognition (Mur et al., 2012). In order to decode the computation that supports object recognition through the distributed or modular approach it is important to investigate the role that LOC plays.

Present Experiment

To address this gap in knowledge the BOLD5000 data set was used to analyze neural activity for patterns across categories in specific regions of interest (Chang et al., 2018). Slow event-related functional MRI data was collected for almost 5,000 distinct images that depict real-world scenes. The large scale of the dataset as well as the image diversity provide the unique opportunity to explore a wide range of categories. The goal behind this neural dataset was to integrate biological data with computer vision advances such that a strong relationship between visual input and specific brain responses can be developed. The mechanisms and computations behind neural activity can be challenging to understand, which is where high-performing

computer vision models come into play. The BOLD5000 uses standard computer vision datasets in order to bridge the gap between stimuli used on a small scale in most human neuroimaging studies and computer vision models. Images from the Scene Understanding (SUN), Common Objects in Context (COCO), and ImageNet datasets were used, enabling a detailed exploration into the neural representation of a wide range of objects and scenes. To explore the effect of image repetition, 113 of the 5000 images were randomly selected to be shown four times to the participant. One-fifth of the images selected were scenes images, two-fifths were COCO images, and two-fifths were ImageNet images. This spread allowed for a diverse set of categories to be pulled out from this pool of repeated images. Within the scope of this project, eighteen of the repeated images were classified as belonging to a unique category such as "beach" or "dog". Each category was then built with the repeated images as well as exemplars of that category within the non-repeated pool of stimuli in the BOLD5000 stimulus set. In all this resulted in 18 categories with a mix of repeated and non-repeated images such that neural activity between these two could be compared in various regions of interest. Early visual cortex, LOC, and PPA were the main regions of interest such that meaningful comparisons could be drawn between regions that are at different levels of the ventral visual stream. Early visual cortex performs preliminary visual computations, whereas PPA is more downstream and has been found to be selective for scenes (Lee, 2003; Epstein et al., 2003). Given that the level and type of computation at LOC is of interest, these ROIs serve as good comparisons. We predicted that a pattern would emerge from the neural activity showing that LOC does not modulate based on subordinate category. This means that the overall category of the stimulus is of interest to LOC as opposed to the specific exemplar within the category. This generalization of categorical

stimuli would be supported by a pattern of activity where neural activation correlation is no higher between the repeated images compared to all of the images in the category set.

Methods

Stimulus Selection and Data

Three image datasets comprise the subset of images used in this study from the BOLD5000 (Chang et al., 2018). Three main considerations were taken into account by the researchers that built the image set for the BOLD5000 – size, diversity, and image overlap. Neural studies often use around a hundred distinct images, which is much more constrained. This narrow set of images consequently restricts the variety of images that can be included. The volume of images in this dataset allows for a diverse set of categories to be included such that object recognition can be explored in a wide range of scenes. The last consideration, image overlap, pertains to overlap across stimuli in computer vision datasets. The BOLD5000 neural dataset can be compared to computer vision model representations of visual input because of the image overlap. This is advantageous for the overall advancement of visual cognition research. Of the total 5,254 images (4,196 unique images) the image breakdown is: 1,000 images of indoor and outdoor scenes from the SUN dataset, 2,000 images with multiple frequently interacting objects embedded in a realistic context from the COCO dataset, and 1,916 images with a single object as the focus from the ImageNET dataset (Figure 1). See the BOLD5000 paper for more detail on how stimuli were chosen from each of the datasets (Chang et al., 2018). Slow event-related fMRI was conducted for four participants where each participant viewed all 5,254 images, but 113 randomly selected images were selected to be shown four times to the participant in order to investigate the effect of image repetition. One fifth of the images were

SUN images, two fifths were COCO images, and two fifths of the images were ImageNET images. Participants were scanned over 16 MRI scanning sessions. Scanning was conducted with a slow event-related design so that the blood oxygen level dependent (BOLD) signal for each individual image trial could be isolated. When each stimulus image was presented, the participant performed a valence judgement task. During the nine second interval between stimulus presentations the participant responded with to what degree they liked the stimulus using the scale: "like", "neutral", "dislike".

For the purpose of this study a subset of the BOLD5000 neural data was selected so that categorization could be investigated. The repeated images were classified by category such that stimuli lists could be built by category. Within the scope of this project 18 of the repeated images were analyzed. In the future all 113 should be analyzed in order to take full advantage of the size of the BOLD50000 dataset. Once each repeated image was classified by category, exemplars of that category were searched for using Google Photos, where all the BOLD5000 stimuli images are stored. To ensure the quality and generalizability of the exemplar to the category, various factors were considered when selecting images. Using Google Photos' search feature, the category (ex. "dog") was searched for so that all images falling under the category were shown. If there were enough exemplars within the category, 10 images were selected. For some categories the number of exemplars was more limited. To select the 10 "best" images competing factors, strength of categorization, and viewpoint were taken into consideration. We predicted that there is a hierarchy of information in a visual scene. Thus, depending on the centricity of an object, the competing factors, and attention-grabbing features of the image, perception of the scene can differ in terms of categorization (Yantis, 2002). For example, an image with a dog in it would not be selected for the dog category if there were multiple people in the image or other

animals in the image. Images with a stronger, clearer categorization were selected. For categories

that had ample exemplars, viewpoint was also taken into consideration so that the images could

be standardized on that feature as well. In addition to the categories built based on the repeated

images, a set of "mixed" images was created such that meaningful comparisons could be made

between images of the same category and images of many different categories. The stimuli in

this "mixed" category include a random selection of 10 images from the selected categories.

Image categories were separated into object and scene categories such that relevant ROIs could

be separately investigated as well.

Figure 1. Sample images from the three computer vision datasets from which experimental
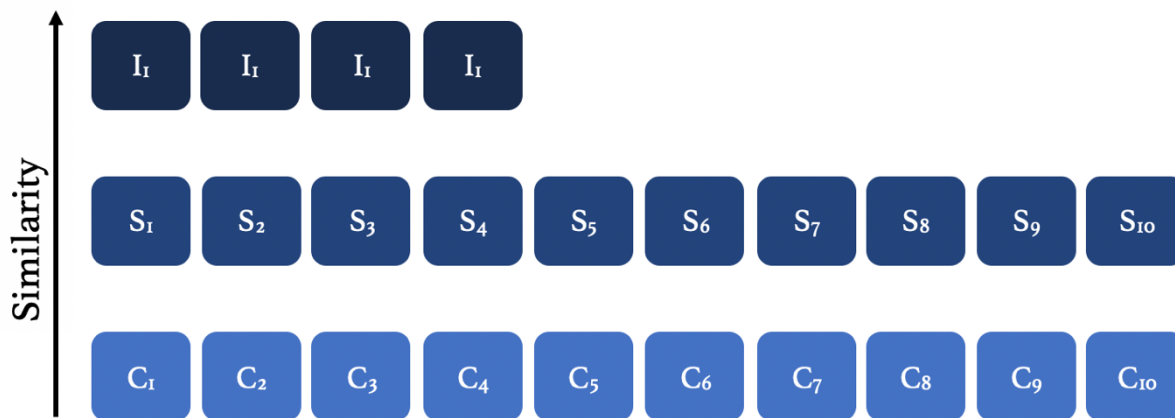image stimuli were selected (Tarr et al., 2018).

Figure 2. Categories consist of four repeated images (I), which have the highest amount of
similarity and 10 non-repeated image stimuli (S) that are within the same category. The
mixed category is made of 10 images (C) from all different categories.

Data Processing

See *Scaling Up Neural Datasets: A public fMRI dataset of 5000 scenes* (Chang et al., 2018), for a detailed explanation of the fMRI data analysis and pre-processing that was completed. All of the data was loaded and analyzed using Python. Since the hemodynamic response function (HRF) activity peaked across TR3 and TR4, TR3 was chosen when conducting analyses. Given that the first participant (CSI1) completed all sessions, this data was chosen for analysis.

Category Correlation

The voxel data for each image within every category was retrieved and built into an array where the rows were the images. The first four images in the array are the four instances of the repeated image. The rest of the rows are the other image exemplars within that category that were not repeated. Using the array with neural data for each category, the correlation between the images was computed using NumPy and was visualized using Matplotlib. This visualization is a representational similarity analysis (RSA), which is used to analyze the response similarity between evoked fMRI responses in selected regions of interest (Kriegeskorte, 2008). In this study the RSA framework quantitatively relates neural activity for early visual cortex, LOC, and PPA. This resulted in a matrix of correlations (R-values) between each image within the category. For categories that have ten non-repetition exemplars the matrix manifested as a 14x14 matrix where the first 4x4 was of the repetitions. The visualizations of these matrices for each category reveal to what degree repetitions were similar to each other compared to the non-repeated images. This was carried out for all categories in three ROIS – Early visual cortex, LOC, and the PPA such that comparisons could be drawn in regard to level of visual processing.

Differences between Correlation Means

In order to quantify the variation in correlation within the exemplars and repeated images in each category, the average correlation was computed. Specifically, the three calculations were: the mean correlation for all images in the category, the mean correlation for all non-repeated images, and the mean correlation for all repeated images. This relays how similar brain activation is for the subsets of stimuli within the categories. This calculation was also carried out for the mixed category as a "baseline" for complete dissimilarity (Figure 2).

Image feature correlations vs. Brain activation correlations

The previous calculations resulted in brain activation correlations for the presented images, however it is also important to consider the similarity within the image features themselves. A convolutional neural network was used as a fixed feature extractor for the images so that image feature similarity could be computed [9]. Specifically, the VGG-19 convolutional neural network with batch normalization, pretrained on ImageNet was used. After being pre-trained on ImageNet the last completely connected layer was removed so that the rest of the layers of the ConvNet could be treated as a fixed feature extractor for the dataset of images used in this investigation. This was used to build an array with feature information for the images within the selected categories. In the same way that brain activation correlation was computed, the correlation between the image features was also calculated and visualized in a matrix.

Discussion: Data Analyses Results

Category Correlation

Each category of stimuli consists of four repeated images and up to ten non-repeated exemplars of the category. As explained above, the neural data for each image was retrieved and

the correlation between the brain activation of each image was computed. This representational

similarity analysis of neural activation correlation (r) was visualized using Matplotlib and was

created for every category, and each ROI. In all, three matrices were created for each category –

one corresponding to Early Visual Cortex activation, one for LOC activation, and one for PPA

activation (Figure 3 &4). This covers different depths in the visual processing stream from early

processing to later processing. The scale used for each matrix is a color bar from dark blue (low

correlation) to yellow (high correlation). Given that the diagonal in the middle is a correlation

between the same image presentation's neural activity, it has a correlation of one, which is why it

is completely yellow. The 4x4 matrix (0-3 by 0-3) within the whole matrix represents the
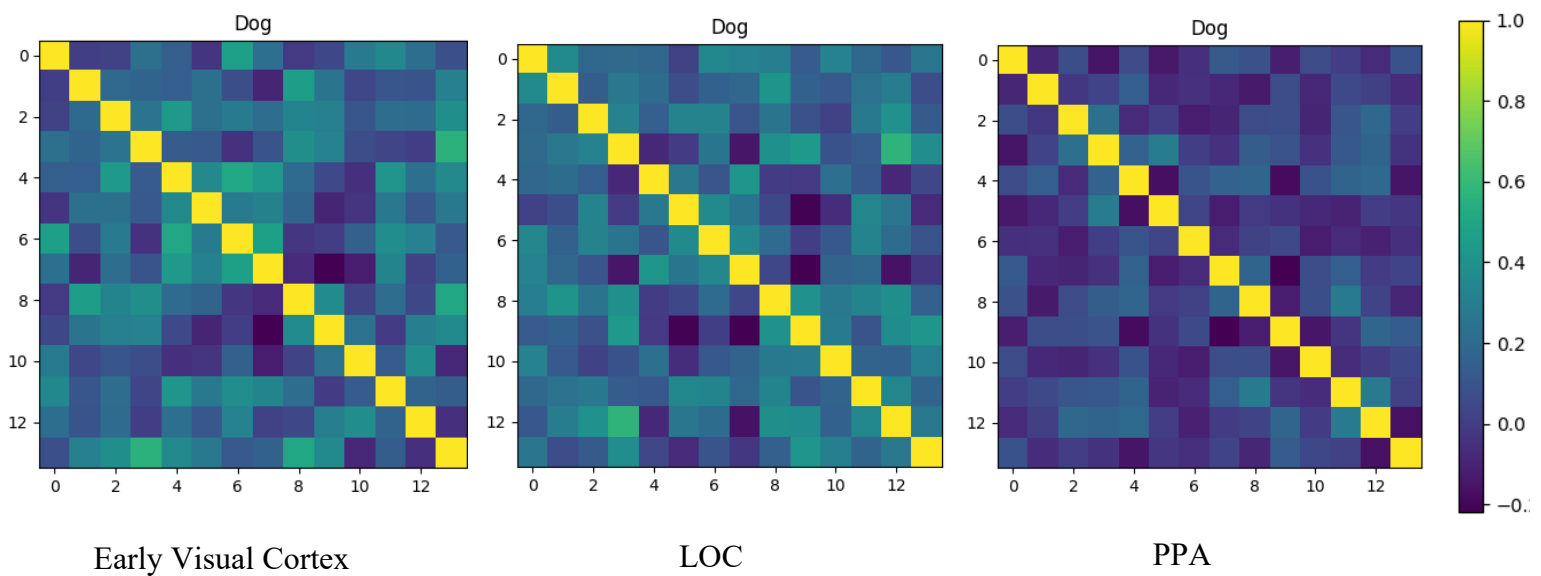
correlation between the repeated images.

Early Visual Cortex   LOC   PPA

Figure 3. Representational Similarity Analysis for "Dog" in three ROIS – Early visual cortex, LOC, and PPA.
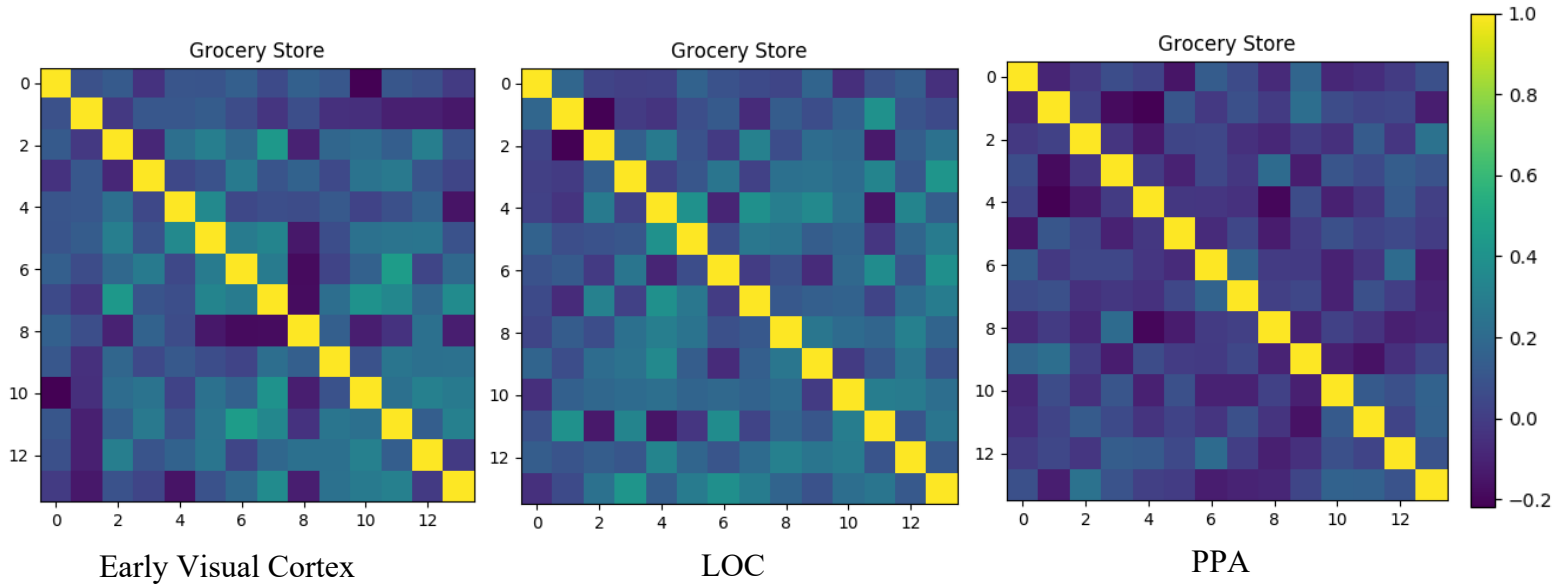
Figure 4. Representational Similarity Analysis for "Grocery Store" in three ROIS – Early visual cortex, LOC, and PPA.

Research has shown that the early visual cortex acts as a first stage of processing, but still has a wide variety of visual computations (Lee, 2003). Neurophysiological evidence has been presented demonstrating that early visual cortex supports computation and representation of perceptual contours, surface shapes, object saliency, and possibly medial axis of forms (Lee, 2003). Local features are extracted, but the revised hypothesis is that various stages of processing might occur through a recurrent feedforward/feedback connection model. Given this understanding of the role of early visual cortex, the matrices of neural activation correlation for each category can be interpreted. Since the visual information is still at a feature level, and the number of examples is small, it makes sense that correlation is relatively random. A major constraint here is the number of examples within the category. Since there are only 14 examples total, the statistical probability of there being some correlation between the features is at chance. For this reason, the early visual cortex correlations are distributed across the spectrum.

At the level of the lateral occipital complex (LOC), it was predicted that specificity in processing was at a categorical level, meaning that just the general category is extracted as opposed to fine-grained processing. In contrast, it is hypothesized that at later areas such as PPA each image is recognized as a specific exemplar within a category. The matrices for LOC in general are in the middle of the correlation spectrum – not completely correlated but not extremely discrete (dark blue). This pattern demonstrates that generalization within the category may be occurring in the LOC. In fact, since the repeated image neural data (4x4 matrix) is not highly correlated with each other it can be inferred that the LOC does not perform subordinate categorization. Rather, just the general category is extracted and stored.

At the level of the PPA it was predicted that the 4x4 matrix of repeated image activation correlations would have r values nearer to 1 (i.e. yellow squares). This however was not the case as seen in Figure 3 & 4. Given the variability in image features in the stimuli, it is possible that during each presentation of the image different aspects of the image were attended to. The experimental design of the study had participants merely completing a valence judgement task, therefore attention was not targeted to one aspect of the image. This can contribute to variability in response. For the rest of the matrix, given that PPA is later in the processing stream, it was predicted that subordinate categorization would occur. The low correlation (dark blue) between the image activation in PPA can be inferred to represent subordinate categorization. The within category correlations may be low because each image is recognized as a unique scene as opposed to a general exemplar within a category.

Differences between Correlation Means

While the RSA's were informative per category, the overall pattern between categories was less clear. For this reason, the average correlation within the repeated image activation, non-

repeated image activation, and total image activation was calculated and visualized for each
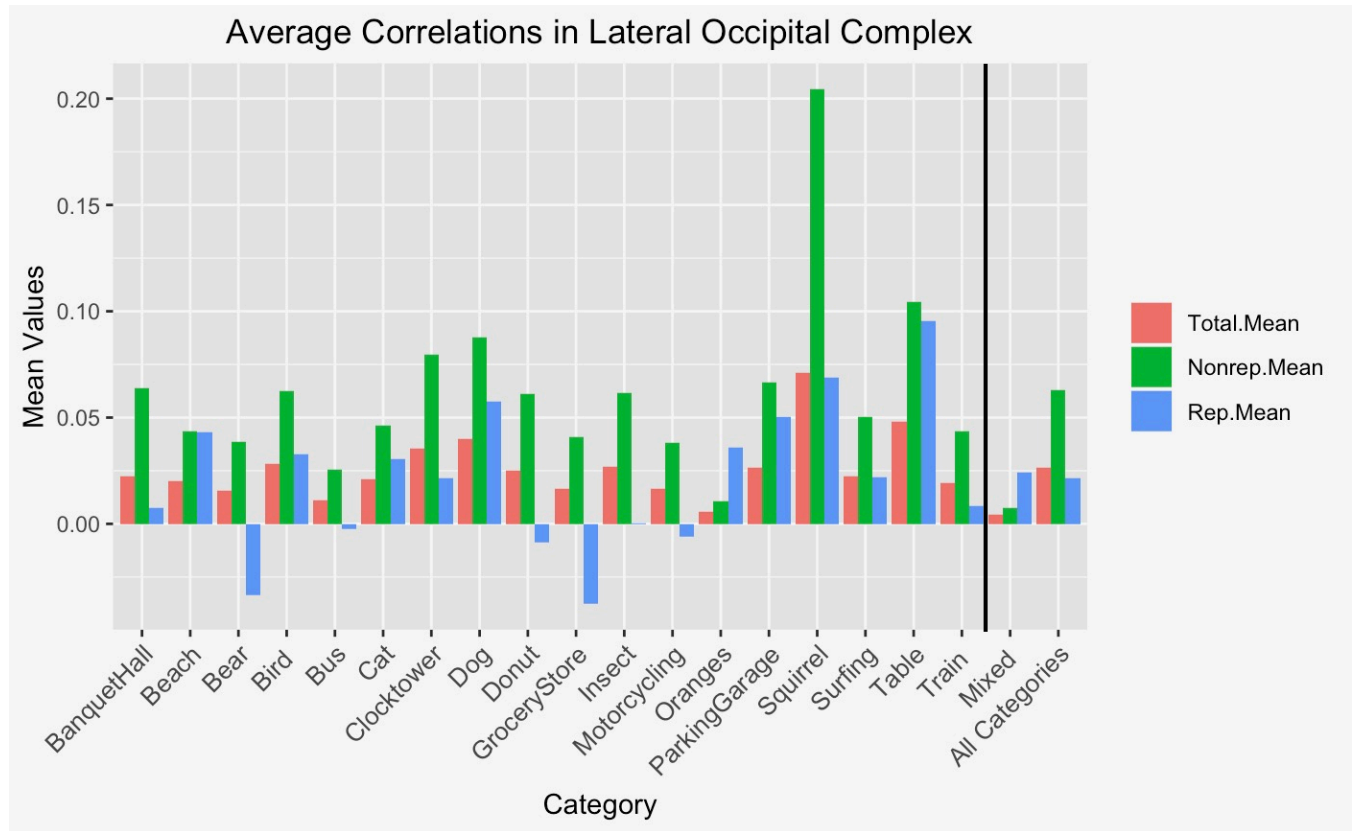
category (Figure 5).

Figure 5. Average correlations of neural activation in LOC across categories

The graph shows that for each category: Mean $_{Reps}$ < Mean $_{Total}$ < Mean $_{Non\text{-}reps}$. Essentially, the

average correlation for repeated image activations are the lowest across all categories and the

average correlation for the non-repeated image activations are the highest across all categories.

The only anomaly is the mixed category, which was included as a between-categories

comparison. The mixed category was included as a baseline comparison since all the images are

from different categories. The responses for images within the mixed category are low and

stable, corroborating the hypothesis that LOC responses are category selective. This distinct

difference between the pattern of activation for the mixed category and all the other categories

demonstrates that categorical information is of importance in LOC. The mean correlation for the

repeated images is the lowest, which can be explained by adaptation. Repeated exposure to the

same image can result in a diminished signal in LOC (Grill-Spector et al., 2006). This reduction

in activation culminates in a lower average response for repeated images in the LOC. The higher

mean correlation for non-repeated images on the other hand has implications for the role of LOC.

It can be inferred that neural responses in the LOC are better driven at a categorical level,

meaning that the non-repeated images within a category generated the strongest signal. For the

future, all 113 repeated image categories should be analyzed as such so that this pattern can be

confirmed across a larger set of data. While this pattern remained relatively consistent, it was not

clear how similar the degree of difference is between categories. Thus, the standard deviation

was computed for the three differences between the calculated means: Total-Nonrep = 0.02679,

Total-Rep = 0.02642, Nonrep-Rep = 0.03650. The low standard deviation for each mean

difference indicates that across categories the difference between the total image activation, non-

repeated image activation, and repeated image activations is relatively consistent. Regardless of

the type of category, responses in the LOC are higher at the categorical level as opposed to

specific exemplars. It is important to note the variability within categories when analyzing this

data. Many of the images selected were COCO stimuli, meaning multiple objects interacting in a

context. Therefore, while the overall category may be salient there may be competing factors in

the images such as people or other objects. This leads to the question – how invariant are neural

responses to variability in these category exemplars? Despite significant variability, neural

activation in LOC is higher for the non-repeated exemplars compared to the repeated images,

which have no variability. The variability in exemplars therefore further points towards

responses in LOC being tuned for the categorical level of the stimulus as opposed to the specific image.

Since the PPA is known for recognition of scenes, the categories that pertain to scenes were extracted and analyzed separately (Figure 6).
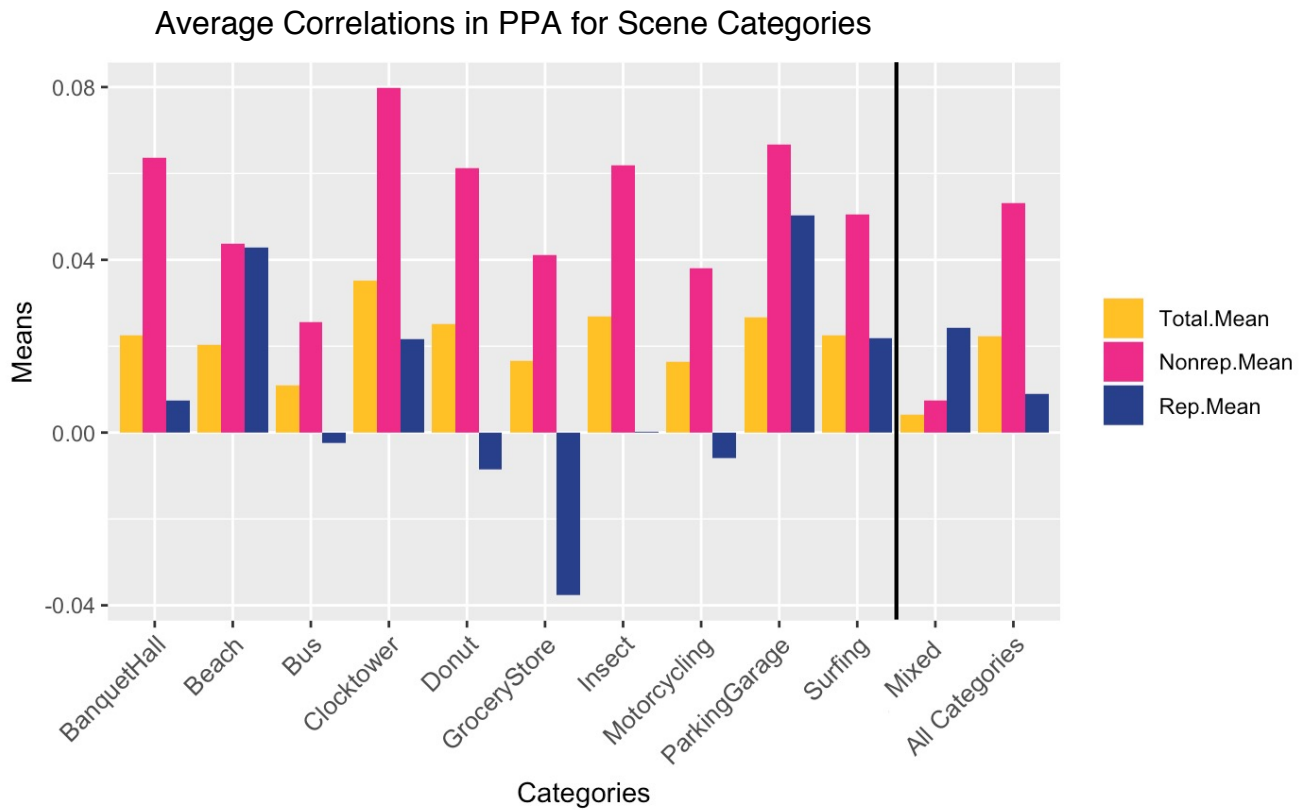
Average Correlations in PPA for Scene Categories

Figure 6. Average correlations of neural activation in LOC across scene categories

The pattern within the PPA is of particular interest when comparing it to the LOC since it is more downstream in the visual processing system. Just as in the LOC, Figure (6) shows that for each category: Mean $_{Reps}$ < Mean $_{Total}$ < Mean $_{Non-reps}$, except for "banquet hall". This may be due to the fact that this category only had four non-repeated exemplars. The variability within the category as well as limited number of exemplars may have skewed the pattern. Since the PPA is hypothesized to perform subordinate categorization, it was predicted that the correlation mean

for repeated images would be higher. Activation gets sparser at higher levels of processing. In early processing areas voxels are feature selective, so many voxels are activated (Lee, 2003). In later processing areas however one voxel may be selective for a specific object. For this reason, the overall activation is much sparser due to the selectivity of the voxels. In multiple presentations of the same image it is possible that different objects were attended to, which would create discrete neural representations in the brain in higher processing areas (Yantis, 2002). The images can have multiple objects or subjects, and since subjects were performing a valence judgement task, their attention was not forced to a certain aspect of the image. On each presentation of the image, different aspects could have been attended to which could explain why neural activity was varied between the repeated images. To investigate this possibility, correlations between image features was later computed.

In order to substantiate the claim that LOC selectively responds at a categorical level, activation for non-repeated images was compared between LOC and PPA. Figure 7 compares the neural activity for non-repeated images within scene categories in the LOC and PPA. The general pattern found is that activation in the LOC is higher than activation in the PPA at a categorical level. The clear difference in level of activation between LOC and PPA supports the hypothesis that while LOC is driven by visual input at a categorical level, PPA responses are not. Research using fMRI adaptation paradigms have also shown that the PPA responds just as strongly to viewpoint changes as it does to scene changes (Epstein et al., 2003). Since PPA does not represent scenes in a viewpoint-invariant manner, it is important to consider the variability in scene category exemplars in terms of viewpoint. The "beach" category for example shows beaches from various views such as an aerial view or ground-level view. The effect of viewpoint changes coupled with the finding that PPA neural activations are lower for exemplars within the

same category, lends to the conclusion that the PPA is not invariant to variability in the

appearance of category exemplars. LOC on the other hand has been found to show a preference

for object changes over spatial changes and shows higher average correlations for exemplars

within the same category (Epstein et al., 2003; Appelbaum et al., 2010). This draws a distinction

between the LOC and PPA in terms of the response to specific categories vs. specific images,

respectively. Research has shown that PPA is selective for places and scenes and that LOC is

more object selective, but it should be noted that the gist of a scene provides context for object

detection (Bar, 2004). It therefore may be the case that top-down signals from PPA modulate

activity in LOC and that bottom-up signals from LOC modulate PPA. For instance, a beach

umbrella might be an indication of a beach scene. In the context of the pattern found in the

BOLD5000 data, it is evident that in terms of categorization the two ROIs have varying degrees

of selectivity. These findings in all suggest that a hierarchy of regions including LOC and PPA
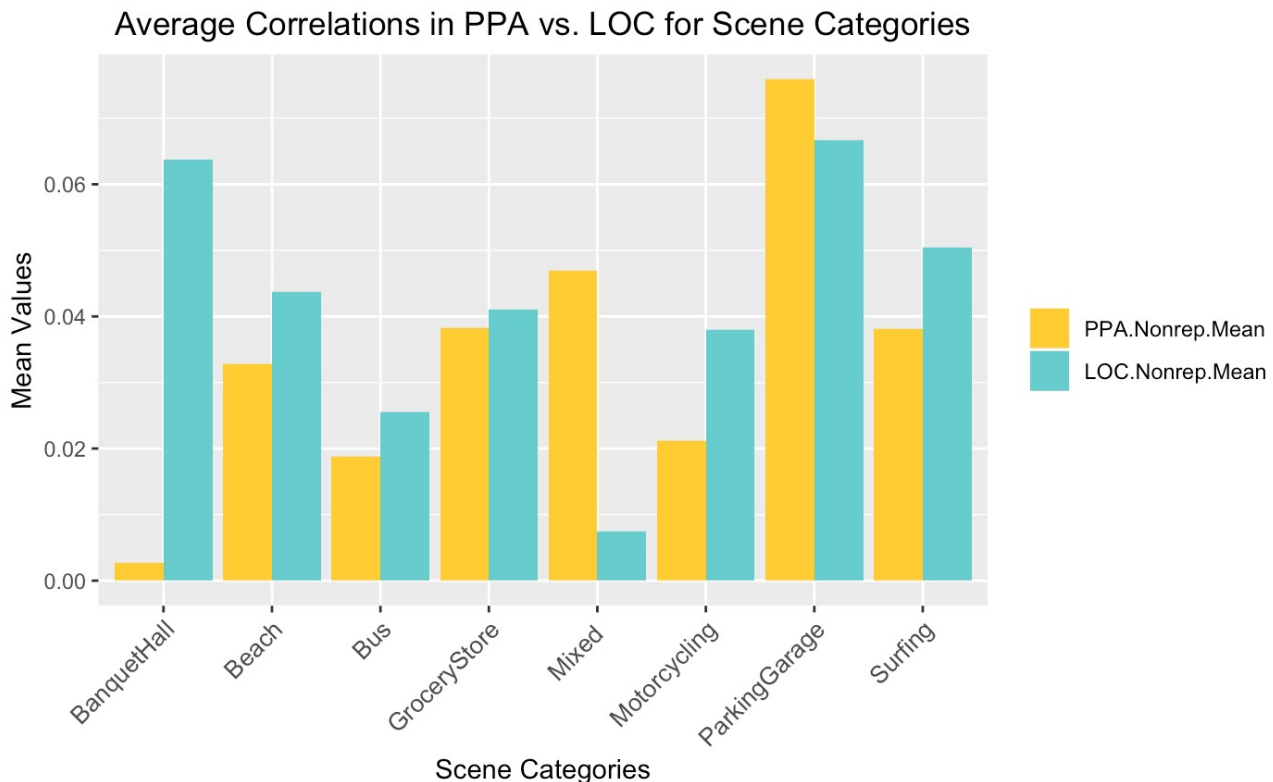
support categorization of visual input.

Figure 7. Average correlations of non-repeated image brain activation in PPA and LOC for scene categories

Image feature correlations vs. Brain activation correlations

   Due to the variability in images within each category, image feature correlations were of interest. Convolutional neural network features for the images were used to build an array where the images again were the rows [9]. In the same way that brain activation correlation was computed, the correlation between the image features was also computed and visualized in a matrix (Figure 8). This matrix was made in order to see if image feature correlation patterns matched the neural activation. Thus, the role of image similarity can be elucidated.
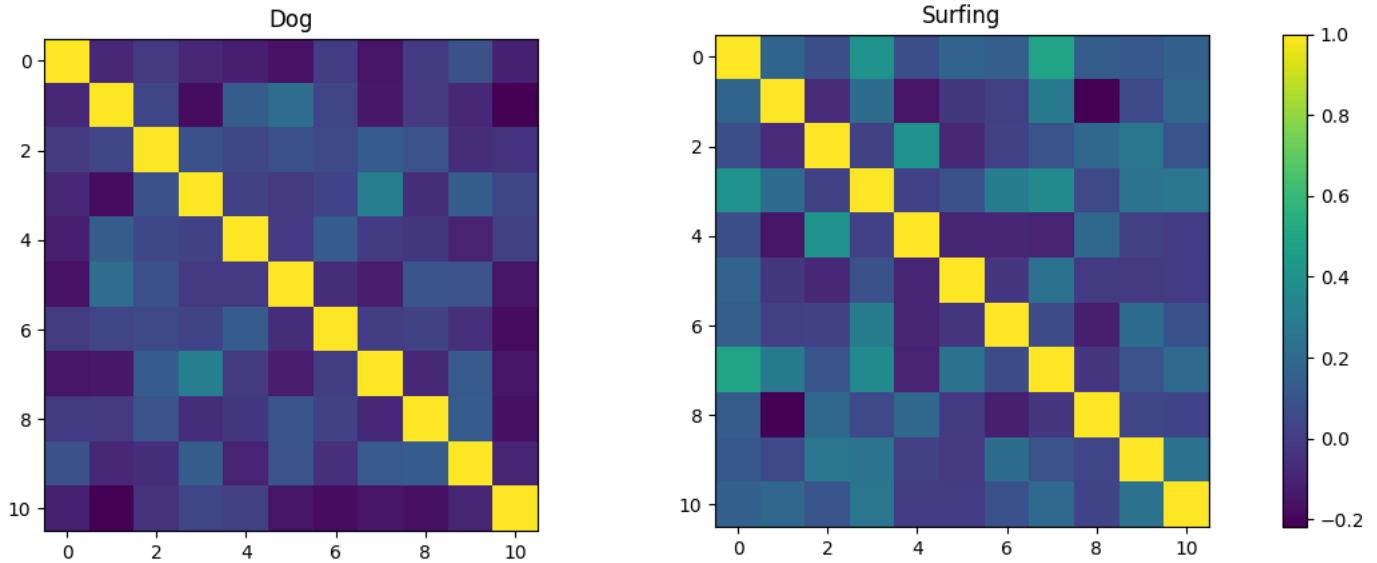

Figure 8. Matrix of image similarity correlations for "Dog" and "Surfing"

Dog for example, has low image feature similarity as evidenced by the low correlations (dark blue). The RSA's showed that image variability could contribute to the variability in brain activation response, however further analysis was needed.

   Image feature correlations were averaged for all of the images within a category and were compared to the mean non-repeated brain image activation for each category (Figure 9).

Repeated image data was not compared since their image features are identical. It was observed that the total image feature correlation means were higher than the non-repeated image brain activation means, except for two categories (dog and squirrel). The variability among the images in each category could be contributing to the variability in brain activation. To quantify to what degree image similarity predicted neural activity the correlation between the image feature similarity and the brain activation was calculated. The r value was found to be -0.13662642. This moderately negative relationship between image similarity and brain activation was skewed by two categories – "Dog" and "squirrel". These two categories demonstrate a pattern reversal since the brain activation average correlation is higher than the image feature average. For all of the other categories the opposite is true. This reversal in pattern can be attributed to variance in image selection. "dog" and "squirrel" images were much more constrained in the image selection process. Due to stimuli limitations, the "squirrel" category in fact only had 4 non-repeated image exemplars, all of which were similar in sizing and viewpoint. Images in the "dog" category were also very similar in content. Each exemplar had a dog of a different breed; but viewpoint, sizing, and distracting factors were all constant. In contrast, the other categories often had many more competing factors in the image such as a person in the background, or variance in viewpoint. Aside from these two categories however, plotting the image feature correlations and brain activation average correlations reveals an interesting pattern (Figure 10). The moderately positive relationship between image similarity and brain activation found could account for some of the variability in brain activation within categories. To a certain degree image feature similarity is correlated with brain activation for non-repeated images within categories. Some of the variability in neural activation found within LOC therefore could be attributed to this moderately positive relationship.
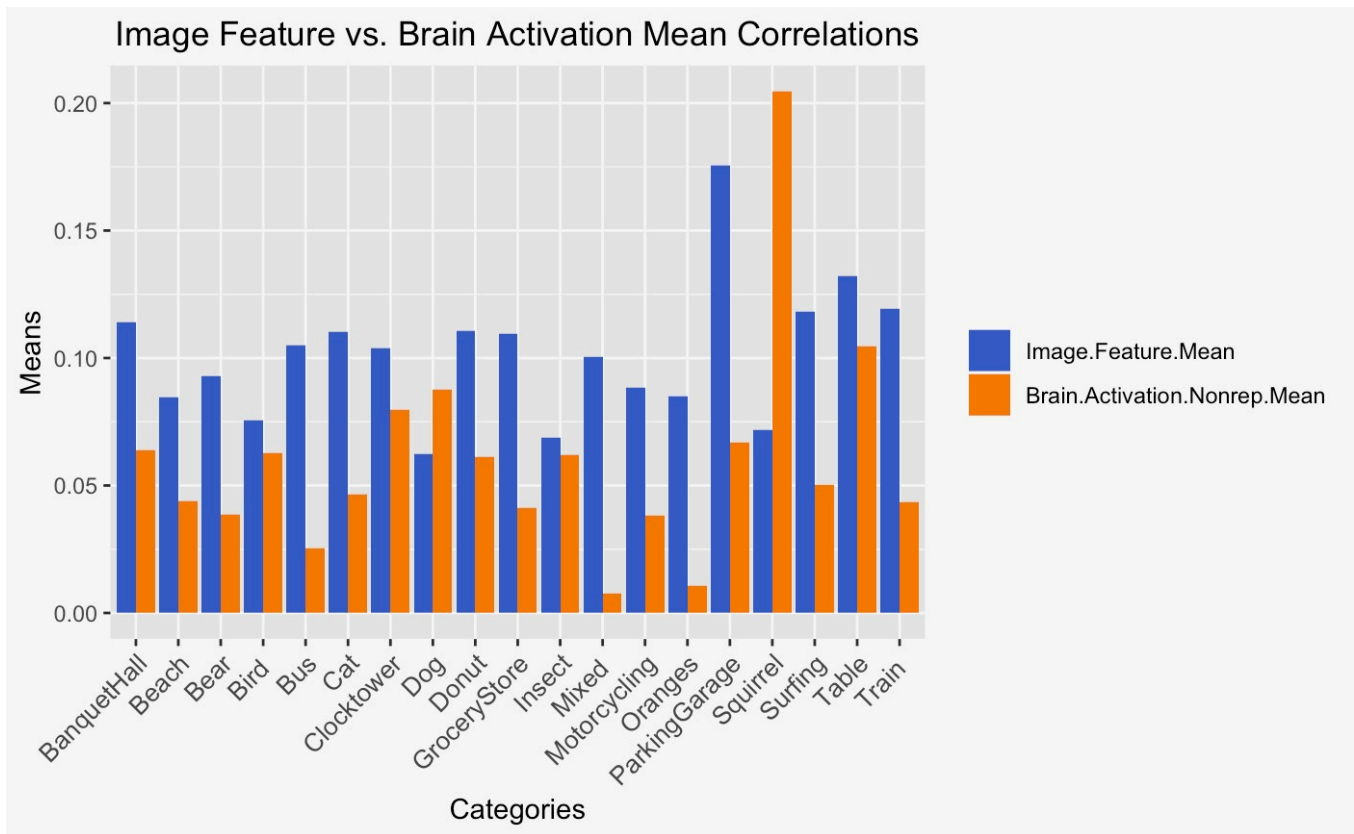
Figure 9. Average correlations of non-repeated image brain activation in LOC compared to average image feature correlations
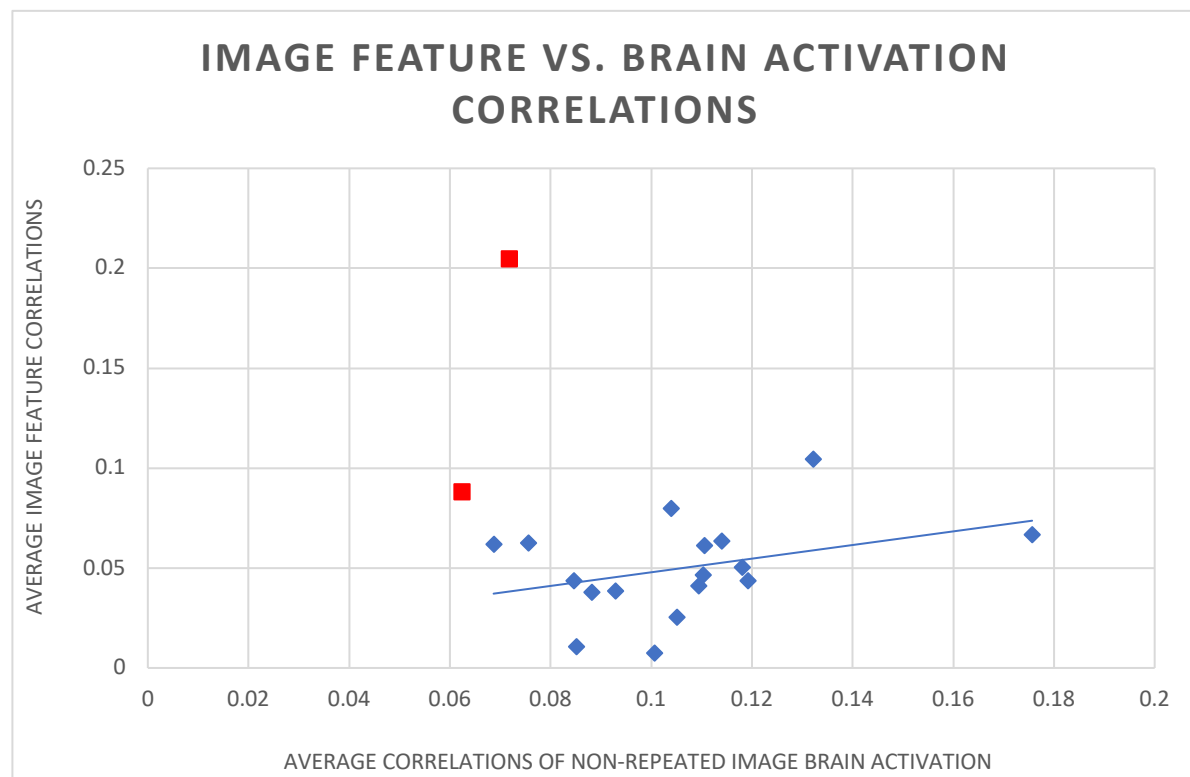

Figure 10. Average image feature correlations vs. Average correlations of non-repeated image brain activation in LOC. The red data points represent "Dog" and "Squirrel". The blue data points represent the rest of the categories.

Conclusion

  In this paper we present fMRI evidence that can elucidate the visual computation performed in order to support visual object/scene recognition and categorization. The specific organization that supports complex recognition capabilities relies on various regions in the brain. This study focused specifically on the mid-level lateral occipital complex, as well as early visual cortex and PPA as upstream and downstream comparisons respectively. Preliminary analysis supports the hypothesis that the LOC acts as a mid-level processing system for object recognition and consequently does not perform subordinate categorization. Analysis showed that compared to Early Visual Cortex and PPA, responses in the LOC are driven at a category level. Exemplars are represented generally in terms of the overall category rather than on an individual level. Further analysis of the 113 repetition categories needs to be carried out such that the effect of repetition and categorization can be analyzed on a larger scale. Understanding the role of LOC implicates the network of activation that supports object recognition overall and can help elucidate the neural computation that supports complex recognition tasks (Walther, 2009). The larger debate of a modular vs. distributed organization particularly can be clarified. Given that the role of LOC may not be specialized for a specific category or have fine-grained representations, the data lends to a distributed model of visual processing. Further analysis of the BOLD5000 data and the many diverse categories represented should be conducted to understand the functional organization of visual input computation in the brain.

References

1. Remington, L. A. (2014). *Clinical Anatomy of the Visual System*. Saint Louis: Elsevier Health Sciences.

2. Lee, T. S. (2003). Computations in the early visual cortex. *Journal of Physiology-Paris,97*(2-3), 121-139.

3. Appelbaum, L., Ales, J., Cottereau, B., & Norcia, A. (2010). Configural specificity of the lateral occipital cortex. *Neuropsychologia,48*(11), 3323-3328.

4. Kanwisher, N. (2010). Functional specificity in the human brain: A window into the functional architecture of the mind. Proceedings of the National Academy of Sciences, 107(25), 11163- 11170.

5. Duchaine, B., & Nakayama, K. (2005). Dissociations of Face and Object Recognition in Developmental Prosopagnosia. Journal of Cognitive Neuroscience, 17(2), 249-261.

6. Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research,41*(10-11), 1409-1422.

7. Chang, N. C., Aminoff, E., Pyles, J., Tarr, M., & Gupta, A. (2018). Scaling Up Neural Datasets: A public fMRI dataset of 5000 scenes. *Journal of Vision,18*(10), 732.

8. Kriegeskorte, N. (2008). Representational similarity analysis – connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*.

9. CS231n Convolutional Neural Networks for Visual Recognition. (n.d.). Retrieved from https://cs231n.github.io/transfer-learning/

10. Peelen, M. V., & Downing, P. E. (2005). Within-subject reproducibility of category-specific visual activation with functional MRI. *Human Brain Mapping,25*(4), 402-408.

11. Grill-Spector, K., & Weiner, K. S. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. *Nature Reviews Neuroscience,15*(8), 536-548.

12. Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: Neural models of stimulus-specific effects. *Trends in Cognitive Sciences,10*(1), 14-23.

13. Epstein, R., Graham, K. S., & Downing, P. E. (2003). Viewpoint-Specific Scene Representations in Human Parahippocampal Cortex. *Neuron,37*(5), 865-876.

14. Walther, D. B., Caddigan, E., Fei-Fei, L., & Beck, D. M. (2009). Natural Scene Categories Revealed in Distributed Patterns of Activity in the Human Brain. *Journal of Neuroscience,29*(34), 10573-10581.

15. McNeil, J. E. & Warrington, E. K. Prosopagnosia: a face-specific disorder. *Q. J. Exp. Psychol.* A **46**, 1–10 (1993)

16. Behrmann, M., Avidan, G., Thomas, C., & Nishimura, M. (2011). Impairments in Face Perception. Oxford Handbooks Online.

17. Dailey, M., & Cottrell, G. (1999). Organization of face and object recognition in modular neural network models. Neural Networks

18. Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., Ledden, P. J., Brady, T. J., Rosen, B. R., & Tootell, R. B. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the Natural Academy of Science USA*, 92, 8135–8139.

19. Yantis, S. (2002). Stimulus-Driven and Goal-Directed Attentional Control. *Visual Attention Mechanisms,*125-134.

20. Mur, M., Ruff, D. A., Bodurka, J., Weerd, P. D., Bandettini, P. A., & Kriegeskorte, N. (2012). Categorical, Yet Graded - Single-Image Activation Profiles of Human Category-Selective Cortical Regions. *Journal of Neuroscience,32*(25), 8649-8662.

21. Bar M (2004) Visual objects in context. Nat Rev Neurosci 5:617– 629.

22. Allison, T. (1999). Electrophysiological Studies of Human Face Perception. I: Potentials Generated in Occipitotemporal Cortex by Face and Non-face Stimuli. *Cerebral Cortex,9*(5), 415-430.