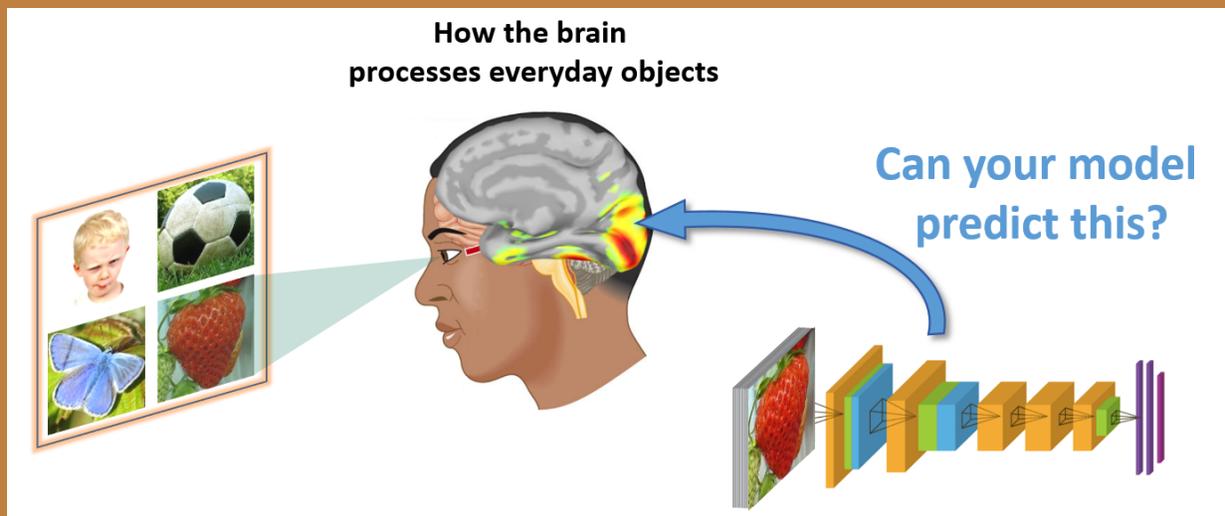


# Predicting the Visual Cortices Responses by Mixing Perceptual and Categorical Approaches



Donald Nikkessen

Layout: typeset by the author using L<sup>A</sup>T<sub>E</sub>X.

Cover illustration: courtesy of the Algonauts Project: <http://algonauts.csail.mit.edu/challenge.html>

# Predicting the Visual Cortices Responses by Mixing Perceptual and Categorical Approaches

Donald Nikkessen  
10260102

Bachelor thesis  
Credits: 18 EC

Bachelor *Kunstmatige Intelligentie*



University of Amsterdam  
Faculty of Science  
Science Park 904  
1098 XH Amsterdam

*Supervisor*  
dr. A. Visser

Informatics Institute  
Faculty of Science  
University of Amsterdam  
Science Park 904, 1098 XH Amsterdam

Jan 31st, 2020

# Predicting the Visual Cortices Responses by Mixing Perceptual and Categorical Approaches

Donald Nikkessen

## Abstract

The link between human visual perception and activity in the brain is a widely studied, but currently still unsolved problem. The field of cognitive computational neuroscience is searching for different models that are able to predict brain activity. This thesis is focused on a small subtask of this problem; predicting activity in the human visual brain, specifically in two areas associated with early- and late stages of visual processing.

In 2019 the Algonauts Project issued a challenge for this specific tasks, motivating many researchers to attempt to find a solution. The challenge has since concluded, and studying the work of the top contestants reveals several promising approaches; including models based on edge detection, categorization and classification, and various implementations of convolutional neural networks (CNNs). The hypothesis behind the challenge was that the brain may be better simulated by models of lower complexity, which meant researchers were motivated to find alternatives to using deep CNNs.

Current research of the visual brain shows that information coming from the eyes travels through a sequence of areas associated with recognition of progressively higher level features. The question behind this thesis is: which *computational* models are able to capture *biological* structures of the early and late stages of visual processing in the brain? Based on the research of participants of the 2019 Algonauts challenge, a framework was designed for this thesis that allows the combinations of different models to show their ability to predict brain activity. Using this framework it was found that an edge detection based model, inspired by known functions of the Primary Visual Cortex (V1), accounts for the majority of the total performance when predicting fMRI activity in both the early and late stages. The combination of this model with categorical and CNN-based models offered only minor improvements.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Background . . . . .	1
1.2	Scope . . . . .	2
1.3	Outline . . . . .	3
<b>2</b>	<b>Theoretical Foundation</b>	<b>4</b>
2.1	The Visual Brain . . . . .	4
2.2	Neural Imaging Techniques . . . . .	5
2.2.1	Functional Magnetic Resonance Imaging (fMRI) . . . . .	5
2.2.2	Magnetoencephalography(MEG) . . . . .	6
2.3	Representational Similarity Analysis . . . . .	7
2.4	Canny Edge Detection . . . . .	7
2.5	Gaussian Naive Bayes Classifier . . . . .	8
2.6	Convolutional Neural Networks . . . . .	8
2.6.1	VGG . . . . .	9
2.7	Related Work . . . . .	11
2.7.1	Augustin Lage-Castellanos . . . . .	11
2.7.2	Anne-Ruth Meijer . . . . .	11
<b>3</b>	<b>Approach</b>	<b>12</b>
3.1	Training and Test Data . . . . .	12
3.1.1	Image data . . . . .	12
3.1.2	Measurement Data . . . . .	12
3.2	Implementation . . . . .	13
3.2.1	Perceptual Model/Edge Detection . . . . .	13
3.2.2	Categorical Model/Naive Bayes . . . . .	14
3.2.3	Neural Networks . . . . .	14
3.3	Evaluation . . . . .	15
<b>4</b>	<b>Results</b>	<b>16</b>
4.1	Initial Model Estimation . . . . .	16
4.1.1	Perceptual Model . . . . .	16
4.1.2	Categorical Model . . . . .	18
4.1.3	CNN model . . . . .	20
4.2	Combining Models . . . . .	21

4.2.1	Combining Initial Estimate RDMs . . . . .	21
4.2.2	Combining CNN Model with Combined Estimates . . . . .	21
4.2.3	Final results . . . . .	22
<b>5</b>	<b>Discussion</b>	<b>24</b>
<b>6</b>	<b>Conclusion</b>	<b>25</b>
<b>A</b>	<b>Training and test images</b>	<b>28</b>
<b>B</b>	<b>Plotted graphs</b>	<b>33</b>

# Chapter 1

## Introduction

### 1.1 Background

The field of artificial intelligence is committed to understanding the systems underlying intelligent behaviour, in order to allow these behaviours to be emulated and automated using machines. In order to explain intelligent behaviour in humans it is necessary to understand the human brain, which is known to provide humans with the facilities required for intelligent behaviour [Russell and Norvig, 2003]. Among their approaches to simulate human brains, researchers are developing computational methods that are inspired on biological processes to perform tasks that humans exceed at, such as object recognition. This research is interested in this task specifically, and how biologically inspired approaches compare to the more traditional approach of using a neural network.

The study was inspired by the Algonauts project, which was created to promote the study of simulating human brain activity. The project presents yearly challenges with the aim to further the understanding of processes in the human brain and the progress of computational simulation of processes in the brain [Cichy et al., 2019]. The subject of the 2019 challenge was inspired by an observed correlation between a convolutional neural network’s (CNN) performance on the ImageNet database, and its similarity to the human brain as calculated by the Brain-Score method [Krizhevsky et al., 2012] [Schrimpf et al., 2018]. However, for the most advanced CNNs this correlation disappears. This inspired the project to follow the hypothesis that the human visual brain may be better understood with the use of less complex models.

The challenge data consists of three sets of images of different subjects such as animals, fruits or scenery. Two of these sets were used for training models, the third was used for validation. Accompanying the images is the brain measurement data of 15 study participants, measured from the areas in the brain known as the Early Visual Cortex (EVC) and the Inferior Temporal Lobe (ITC). These are associated with early and later stages in visual processing, respectively [Grill-Spector and Malach, 2004]. The data was collected using noninvasive brain imaging techniques known as functional Magnetic Resonance Imaging (fMRI) and Magnetoencephalography (MEG) [Huettel et al., 2004] [Hämäläinen et al., 1993]. In order to allow this data to be compared to computational models, it has been processed into Representational Dissimilarity Matrices (RDMs). This assigns a similarity

score to each combination of images, based on the similarity of their associated brain measurements [Kriegeskorte et al., 2008]. The goal of the challenge was to predict this RDM using only the original image data, using deep neural networks or any other computational model. Since the conclusion of the 2019 challenge, the evaluation data used for ranking the model submissions of challenge participants has been made publicly available. This allows researchers to rapidly evaluate performance of new models, as opposed to using the online submission system that was available during the challenge.

This study is based the work of the following challenge participants:

- Augustin Lage-Castellanos

Augustin achieved a winning score by using a combination of models [Lage-Castellanos and De Martino, 2019]. His method consist of generating four initial estimate RDMs; two based on perceptual features using edge detection, and two based on categorical features extracted from the data. These initial estimates are then combined with models based on activations from selected CNN layers, using weighted averaging to optimize their combined performance. This process of combining gradually more advanced models bears some similarity to processes observed in the brain.

- Anne-Ruth Meijer

Anne-Ruth’s research showed smaller CNNs were able to outperform large networks that received considerably more training [Meijer and Visser, 2019]. The research offers knowledge about the performance of many different CNNs, which will be useful to compare the performance of other models with.

## 1.2 Scope

This study features exploration of the approaches from the researchers introduced in the previous section. To gain a better understanding of the models used, a replication study was performed. The technique of combined models used by Lage-Castellanos was selected for the reproduction study, not only because this was one of the winning models, but also because the reproduction simultaneously created a framework for combining and testing different models. Using this framework, a series of experiments that grant more insight of the relationship between specific models and the brain processes they aim to simulate. The results of these experiments illustrate the performance of these models when predicting RDMs for the different brain areas and time intervals, which is used as a measurement tool to verify if a model is able to simulate brain activity.

The goal of the research is to answer the following question: which *computational* models are able to capture *biological* structures of the early and late stages of visual processing in the brain? While CNNs are very capable classifiers, they do not fully capture the biological structures present in the brain. This study compares the performance of convolutional neural networks to biologically inspired models.

## 1.3 Outline

Before discussing the research the theoretical foundation will be laid out in chapter 2. The main topic of this research is the visual brain, and how different models correlate to activity in specific areas. A summary of current knowledge on this matter is given, including the techniques used to collect the data. To complete the knowledge foundation, the different computational models and methods used in the research are described. After establishing this basis, the results several challenge participants will be discussed. Chapter 3 provides details of the methods used to apply the theory discussed in the preceding chapter. In chapter 4 the results of the study are presented, which are then discussed in the remaining chapters.

# Chapter 2

## Theoretical Foundation

This section provides an overview of the scientific theory this project is based on. In order to provide a background for the selected areas of the brain studied in this research, a summary of current knowledge of the human visual brain is given in section 2.1. In section 2.2, the employed brain imaging techniques are explained in some detail. To allow comparison of the measurement data to a computer model, both are mapped to a shared 'similarity space'. This is done using a method called representational similarity analysis (RSA), which is discussed in section 2.3. Section 2.4 contains a description of Canny Edge Detection which is used by one of the models. Section 2.5 introduces Bayesian Classification on which another model is based. And finally, section 2.6 is dedicated to neural networks, explaining their most important characteristics and detailing three different variants used in this research.

### 2.1 The Visual Brain

The human brain can be divided into two pathways; the dorsal and ventral visual streams. The ventral stream is described as the "what" stream of visual processing, while the dorsal stream is described as the "where" and "how" stream [Lee et al., 1998]. The ventral stream is often considered as a feed-forward network; where visual information ascends the visual hierarchy, which means the size and complexity of detected features increases [Serre et al., 2005].

The visual cortex receives its input signal in the primary visual cortex (V1), which is known to detect local features in the input such as edges and lines [Lee et al., 1998]. This is also the area measured for the early stage of the challenge data (denoted EVC-fMRI and Early-MEG), which suggests the activity in this area may be in part explained by an edge detection model (see section 2.4). The signal then proceeds to the secondary visual cortex (V2), receiving strong feedforward inputs from V1. V2 lies within the visual association area of the cortex, and in primates it is shown to be tuned to more complex features like size, shape and spatial frequency [Hegd e and Van Essen, 2000]. Next the signal reaches the V4 area, which is not yet fully understood in terms of its preferred features, but has been shown in primates to react to geometric shapes of intermediate complexity [Lee et al., 1998]. It also has strong bidirectional connectivity with the last area in the hierarchy: the Inferior Temporal Cortex (ITC). This area is associated with recognition of features with high

complexity such as faces, objects and patterns [Creem and Proffitt, 2001]. This is the second area measured for the challenge, representing the late stage of the visual processing cascade (denoted ITC-fMRI and Late-MEG). Since this area is specialized in complex features, it could be possible that a categorical approach helps to explain activity in this area.

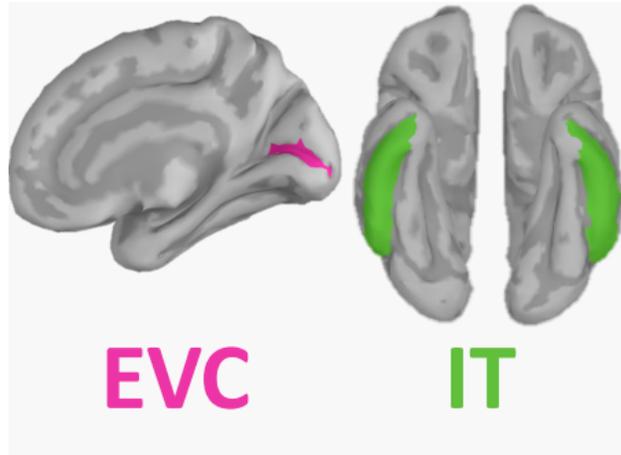


Figure 2.1: Visualization of the areas in the visual processing hierarchy studied in this research. The Early Visual Cortex (EVC) corresponds to the V1 area of the brain, the Inferotemporal area (IT) corresponds to the V4 and ITC areas. Image courtesy: Algonauts project<sup>2</sup>

This description is by no means complete, and only serves to illustrate the basic structure underlying visual processing, which facilitates making comparisons to computational approaches. One important aspect to note is the increase in complexity of detected features as the signal travels further up the hierarchy, this principle is also present in many current CNNs, where the 'deeper' layers model progressively larger and more complex features of the input.

## 2.2 Neural Imaging Techniques

Neural measurement data has the goal of revealing relationships between behaviour and neural activity. The human brain data used for the Algonauts challenge [Cichy et al., 2019] was measured using two non-invasive techniques; fMRI, which aims to measure the location of neural activity within the brain, and MEG, which aims to measure neural activity over time. By using data from both techniques, researchers can attempt to explain activity in terms of both location and time.

### 2.2.1 Functional Magnetic Resonance Imaging (fMRI)

The most common form of fMRI measures differences in the changes of blood flows within the cortex. This is called blood-oxygenation-level dependent (BOLD) fMRI [Huettel et al.,

---

<sup>2</sup> [Cichy et al., 2019]

2004]. It is based on the fact that activity in a specific area of the brain results in increased blood flow to that area; neurons in the brain require both energy and oxygen to function, which are supplied by the vascular system. What makes it possible to measure changes in blood flow within the brain, is the difference in magnetic properties between oxygenated and deoxygenated blood. This difference is very small however and easily influenced by other factors. This includes environmental factors such as temperature, interference caused by equipment, and random neural activity, but also includes individual differences such as a person’s physiological state, differences in mental strategy, and behavioral differences between people and across different tasks. Current fMRI machines are able to measure the location of neural activity with up to millimeter precision, the measurement needs to be taken over a period of more than 1 second however, in order to capture any useful information.

In figure 2.2 an example visualization of fMRI data is shown, with different colored points that highlight activity in the two brain areas studied in this research.

### 2.2.2 Magnetoencephalography(MEG)

Magnetoencephalography (MEG) measures brain activity by detecting small magnetic fields caused by neural oscillations, also known as brain waves [Hämäläinen et al., 1993]. Measurements are performed by placing an array of highly sensitive magnetometers against the skull of the subject, but because the magnetic fields are so small the equipment can easily cause interference. Normal environmental magnetic noise is also orders of magnitude stronger than the strongest magnetic fields generated by the brain, and needs to be shielded from the magnetometers to prevent interference. MEG measurements are made with millisecond precision, measuring the activity at each magnetometer over an interval of time, usually during the presentation of a stimulus. Figure 2.2 shows an example visualization of MEG measurement data.

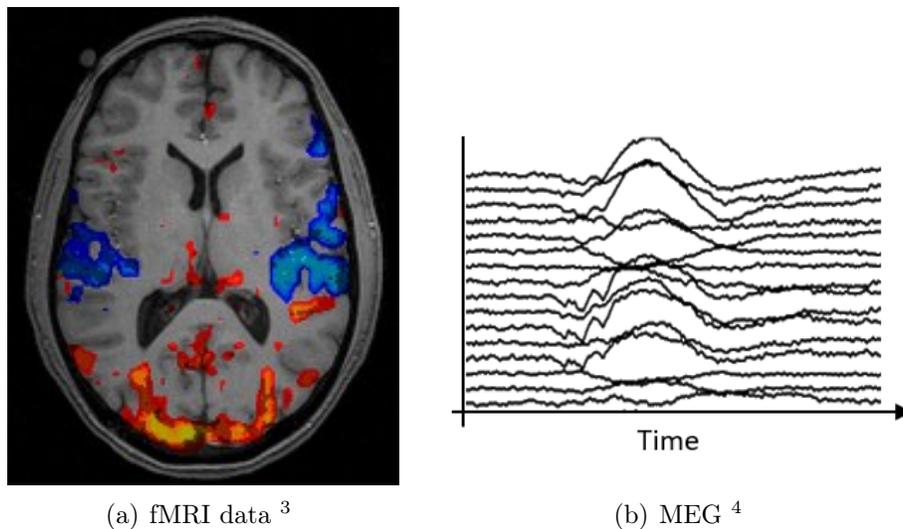


Figure 2.2: Example visualization of fMRI (left) and MEG (right) measurement data. Image courtesy <sup>34</sup>.

## 2.3 Representational Similarity Analysis

In order to allow comparison between the computer generated models and the brain measurement data, these data need to be mapped to a similar space. The method used for the Algonauts challenge is called Representational Similarity Analysis (RSA). [Kriegeskorte et al., 2008] The method involves converting the input activations for all images into a dissimilarity matrix (RDM), which is done by converting the activations into vectors and finding the correlation between each pair of images. The dissimilarity value for each pair of images in the RDM is equal to one minus the correlation. Once the RDM for a set of images has been calculated, it can be compared to another RDM using the Spearman correlation [Hauke and Kossowski, 2011].

The challenge evaluates the predicted RDM separately against four different RDMs generated from brain measurement data: the early and late intervals for the MEG data, and the two brain areas for the fMRI (EVC and ITC). These scores are then averaged to obtain two final scores for performance on the fMRI and MEG tracts.

## 2.4 Canny Edge Detection

Edge detection is a common method of extracting visual information from images. The Canny edge detector is named after its developer: John F. Canny, who created the algorithm in 1986 [Canny, 1986]. The algorithm can be described as performing the following steps on the target image:

- Noise reduction by blurring, according to the value of the  $\sigma_c$  parameter.
- Gradient calculation using convolutional filters.
- Non-maximum suppression to remove weak edges.
- Upper- and lower threshold application using the parameters  $t_{upper}$  and  $t_{lower}$ .

Values above the upper threshold are kept while values in between the two thresholds are selected as candidates for the next step.

- Edge tracking using hysteresis.

This decides which values in between the two thresholds are kept.

Parameterization allows the algorithm to detect a specified range of edges from an image. This is visualized in figure 2.3, which shows the effect of increasing Canny  $\sigma$  values.

As mentioned in section 2.1, the V1 area of the brain is associated with the detection of edges [Lee et al., 1998]. This suggests that activity in this area might be better understood using a model based on edge detection. The results of this research will show some merit behind this statement, as improved performance is observed when using this model to predict MEG data.

---

<sup>3</sup>[http://algonauts.csail.mit.edu/fmri\\_and\\_meg.html](http://algonauts.csail.mit.edu/fmri_and_meg.html)

<sup>4</sup><https://psychcentral.com/lib/what-is-functional-magnetic-resonance-imaging-fmri/>

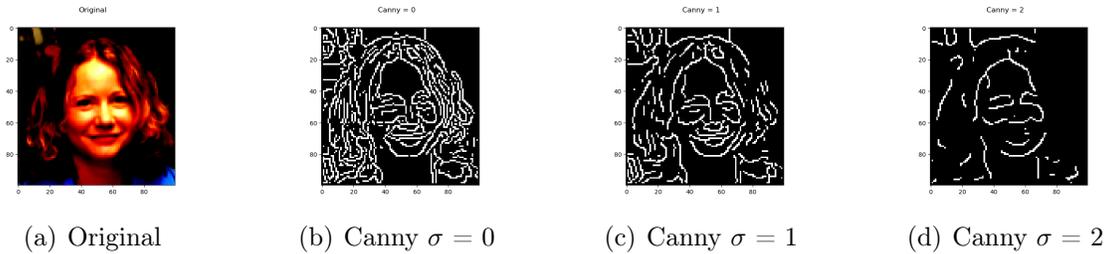


Figure 2.3: Edges detected by the Canny filter for increasing values of  $\sigma_c$ .

## 2.5 Gaussian Naive Bayes Classifier

Naive Bayes classifiers are a family of algorithms that use probabilistic mathematics to categorize inputs [Russell and Norvig, 2003]. These algorithms make their predictions with the use of Bayes' Theorem [Joyce, 2019], which aims to determine the conditional likelihood of an outcome given a set of observations.

Algorithms of this family assume that the features of the input contribute independently from each other to the likelihood of its class membership. For example: a fruit may be classified as an orange because it is round, orange of color, and falls in the correct weight range, but no assumptions are made that the shape, color, and weight are related to each other.

The Gaussian Naive Bayes (GNB) classifier extends the Naive Bayes classifier by allowing continuous values to be used as input data. The algorithm assumes that the values representing each class are distributed according to a Gaussian distribution [Weisstein, 2020]. This way it can calculate the class membership by generating a probability distribution for each class over the continuous variable space.

From the literature studied in section 2.1, it was found that in the human brain processes visual inputs in a hierarchical sequence, recognizing progressively more complex features. For this reason it is expected that an image classifier may better explain activity in the later stages of visual processing, associated with the V4 and ITC areas of the brain [Serre et al., 2005].

In this research, the GNB classifier is used to categorize images from the 78 image set based on their representation in a selected layer from a CNN. This will be discussed in the next section.

## 2.6 Convolutional Neural Networks

When speaking generally about deep neural networks, more layers often equates to better performance [Russell and Norvig, 2003]. The 'deeper' a network gets however, the more computationally taxing it becomes, especially for traditional neural networks where all layers are fully connected (FC). As the name implies, all the neurons in a FC layer are connected to all neurons in the next. This is the most general method of connecting layers, but also the most costly in terms of execution time. For example: a  $(10 \times 10 \times 1)$  input image fully connected to a  $(10 \times 10 \times 1)$  layer already results in  $100 \times 100 = 10.000$  weights to be learned.

To counteract this problem, a convolutional neural network (CNN) uses convolutional filters to recognize features from the input, and learns the weights for these filters instead [Ma et al., 2018]. They were inspired by the finding that cortical neurons in animal brains are only responsive to stimuli in a select region of the visual field, called the receptive field [Hubel, 1968]. Convolutional layers process their inputs using convolutional filters: a convolution selects an  $(n \times n)$  subset of neurons that contribute to the activation of a single neuron in the next layer. In a sense, the convolutional filter can be seen as a computational analogue to the receptive field.

Pooling layers are another way to lower the computational cost of a network. They do this by reducing the number of values in their input, using one of several methods, the most common of which is max-pooling. Max pooling looks at a small  $(n \times n)$  section of the input, and only passes the highest value on to the next layer. This can be seen as filtering noise out of the input by only passing the inputs that carry important information, given that the correct filter is applied.

Figure 2.4 shows the basic structure of a convolutional neural network. As the input image travels through successive sequences of convolutional layers and by pooling layers, progressively more complex features are being trained by the convolutional filters. The final fully connected layers then use these features to associate the input image to the outputs.

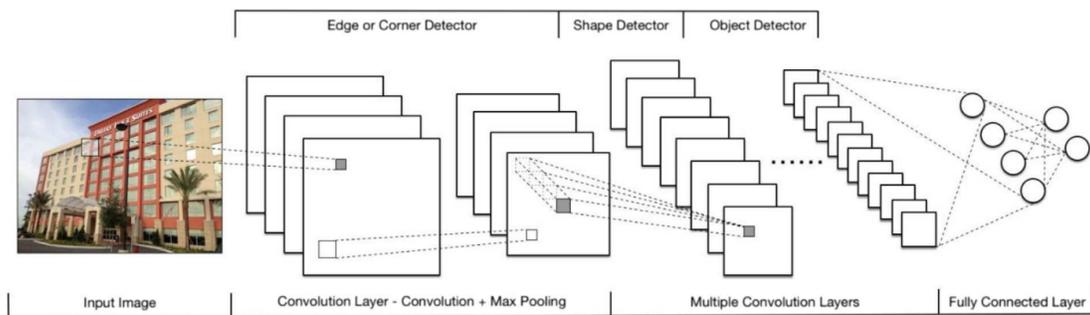


Figure 2.4: Structure of a convolutional neural network. Image courtesy: <sup>3</sup>

## 2.6.1 VGG

The VGG network was developed to show the effect of using deeper CNNs on image recognition [Simonyan and Zisserman, 2014]. Scoring two winning placements on the 2014 ImageNet challenge <sup>4</sup>, the network proved that using very small  $(3 \times 3)$  convolutional filters and adding more layers resulted in state-of-the-art performance when classifying images from a large dataset.

The network can be configured with 16 to 19 layers of depth, it always includes five max-pooling layers which are preceded by a number of convolutional layers depending on the depth. After the last pooling layer, every network contains three fully-connected layers

<sup>3</sup> [Ma et al., 2018]

<sup>4</sup><http://www.image-net.org/challenges/LSVRC/2014/results>

followed by a softmax layer which converts the activations of the last fully-connected layer to a value between 0 and 1.

For this research the VGG-19 version of the network is used, a graphic representation of its structure can be seen in figure 2.5. The last fully connected layer contains 1000 features, which are used in this research both to train the classifier for the categorical model (section 3.2.2) and for the CNN model (section 3.2.3).

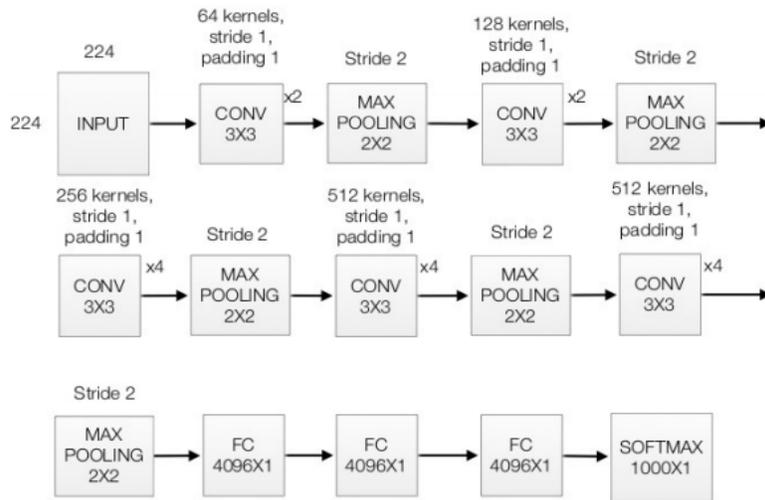


Figure 2.5: Structure of the VGG-19 CNN. Image courtesy: <sup>5</sup>

<sup>5</sup> [Meijer, 2019]

## 2.7 Related Work

This section discusses the research of two participants of the 2019 Algonauts challenge, on which this research is based.

### 2.7.1 Augustin Lage-Castellanos

Lage-Castellanos conducted a thorough study of a multi-model approach [Lage-Castellanos and De Martino, 2019]. His method involved creating four initial estimate RDMs, two based on perceptual features using edge detection, and two using categorical information derived from the test set. Next, he combined these initial estimates with RDMs based on activations from a single layer in a CNN. The optimal models were found by using weighted averaging, while maximizing for the correlation between the predicted RDMs and the target RDMs.

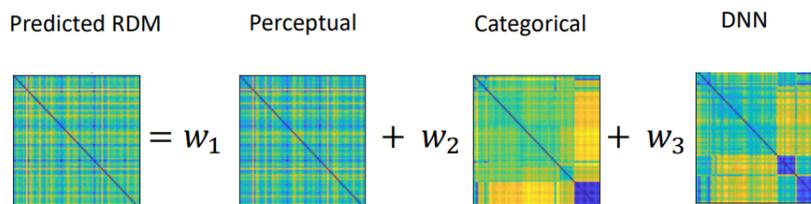


Figure 2.6: Creating a prediction from different RDM estimates. Image courtesy: <sup>6</sup>

In this research, the weighted averaging method is deconstructed into two steps: combining the initial estimates, and combining the result with the RDM based on a CNN layer. This way the contribution of each model can be observed as they are combined.

### 2.7.2 Anne-Ruth Meijer

The study done by Meijer was focused on finding smaller, or 'shallower' neural networks, that achieve competitive scores on the Algonauts challenge [Meijer and Visser, 2019]. During the research she tested various CNNs, the results of which are a useful tool for comparison. The research also involved training a newly designed network, based on the ResNet-18 with two additional layers. Tests revealed that training this network for 10 epochs improved its performance beyond that of other networks.

The results of this research are used to compare the differences in performance between the specifically designed perceptual and categorical models, and the more general CNN models.

---

<sup>6</sup> [Lage-Castellanos and De Martino, 2019]

# Chapter 3

## Approach

This section describes the dataset provided for the Algonauts challenge [Cichy et al., 2019]. The challenge supplies a complete development kit which includes three neural networks (AlexNet, VGG-19, ResNet-50), code for generating activations from the networks, and evaluation scripts for scoring RDM predictions. The development kit has been extended for the purpose of this research to include implementations of models inspired by the research by Lage-Castellanos: models based on edge detection and models based on classification using neural network features [Lage-Castellanos and De Martino, 2019].

### 3.1 Training and Test Data

#### 3.1.1 Image data

The challenge provided three sets of images, two of these are training sets of size 92 and 118. The third set is smaller (78 images) and is intended for validation of the trained models, as the brain measurement data for this set was held out until the conclusion of the challenge. The measurement data has since been made publicly available, making it possible to evaluate the performance of models on the test set.

What is immediately apparent when comparing the three image sets, is that the 92 set contains almost exclusively images with a white background, while the other sets have images with almost exclusively colored backgrounds. The subject matter of the images also varies wildly, featuring pictures with multiple subjects visible, partially obscured subjects, and 'isolated' subjects (e.g. a human ear surrounded by a white background). Because of the limited training data and high variability, higher level features from the data will likely be more difficult to capture. This is desirable for the challenge design because it motivates more thoughtful approaches than just training a large neural network.

#### 3.1.2 Measurement Data

Included with each set of images are the RDMs created from the fMRI and MEG measurement data. These are used as training targets for the training sets, and for evaluation on the test set. The measurement data for every image set and challenge tract each contain 15 RDMs in total, corresponding to the number of subjects.

## 3.2 Implementation

This section outlines how the results were obtained from the data. This includes the design of the perceptual and categorical models, generating activations from neural networks, and the method of combining these models. The complete code is available on GitHub <sup>1</sup>

### 3.2.1 Perceptual Model/Edge Detection

The edge detection model was recreated from the work of Lage-Castellanos. It is implemented to allow control over the same three parameters:

- The resize scale in pixels ( $n$ ).

The input images are all resized to the same dimensions before edge detection.

- The Canny lower edge detection threshold ( $t_{lower}$ ).

This determines the minimum strength required for edges to be detected.

- The Gaussian smoothing sigma ( $\sigma_g$ )

This is not to be confused with the Canny edge detection sigma ( $\sigma_c$ ), which is a parameter of the Canny edge detector. The Gaussian( $\sigma_g$ ) is used to determine the amount of smoothing *after* edges have been extracted, effectively making the edges wider. The research is not testing for the Canny sigma parameter ( $\sigma_c$ ) and uses the default value of 1, similar to the model of Lage-Castellanos.

The Canny edge detection function used in this research is version 0.17.dev of the Scikit-Image library (v0.16.1), implemented in Python 3.7. See figure 3.1 for an example of the image after edge detection has been applied.

The algorithm takes the following steps:

1. Load and rescale images to size ( $n \times n$ )
2. For each image:
  - Convert the (RGB) image to grayscale
  - Apply edge detection using the given threshold  $t$ .
  - Apply Gaussian smoothing using  $\sigma_g$
3. Generate the model RDM by counting overlapping edge pixels for every pair of images.
4. Normalize the model RDM.

---

<sup>1</sup><https://github.com/donalddop/Algonauts>

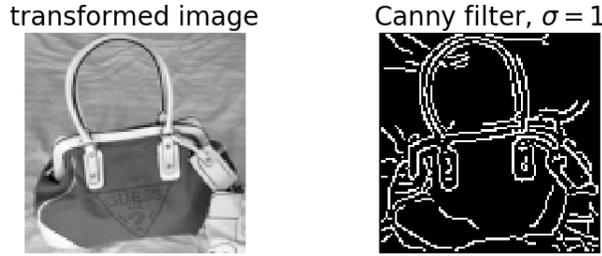


Figure 3.1: Edge detection example on an image from the 118 image set.

### 3.2.2 Categorical Model/Naive Bayes

The categorical model is also a reproduction from the study by Lage-Castellanos. The model uses the following method:

- Manual labelling of the 92 training image set into 8 categories.
- Calculation of and  $(8 \times 8)$  categorical RDM:  
This is done by averaging the values of the initial RDM for all images in the same category.
- Training a naive Bayes classifier on neural network features. For the experiments, the last fully connected layer of the VGG-50 network was selected based on the findings of Lage-Castellanos.
- Classify the images from the test set.
- Generate the model RDM using the values from the categorical RDM.

To obtain the  $(8 \times 8)$  categorical RDM, the measurement RDMs of the respective image set are used. The measurement RDMs are averaged over all 15 patients, and over the 20 ms time interval. The the correlation between two categories is then set to the average activation for all images in these categories, as they are represented in the averaged measurement RDM. Next, the images are categorized using a Gaussian Naive Bayes classifier trained on either the 92 or 118 image set. Based on these classifications, values in the  $(8 \times 8)$  RDM are used to set the correlation between two images as equal to the correlation between their predicted classes. There are four versions available for the categorical RDM, based on both intervals for the MEG (Early and Late) and fMRI (EVC and ITC) data. The results section will show the performance of each version.

### 3.2.3 Neural Networks

The challenge dataset includes code for generating activations using three different types of neural network; AlexNet, VGG, and ResNet. The code is written in Python (3.7.4) and uses the *pytorch* library (1.3.1), which features pre-trained models for various versions of these neural networks. This research uses the activations created by the VGG-19 network after one epoch of training. This network contains 5 max-pooling and 3 fully connected layers. The

activations found to have the biggest impact on scores was the last fully connected layer (fc8), which contains 1000 features per image. These features are used in the categorical model to train the classifier on, as well as in the final step of the algorithm where the contribution of an RDM based on the activations in this layer is added to the initial estimates. These RDMs are created from the activations using the Spearman correlation, the same metric used by the challenge organizers for generating the RDMs from the measurement data [Hauke and Kossowski, 2011].

### 3.3 Evaluation

The evaluation method used for the research calculates scores in the following steps:

- Calculate Spearman’s correlation coefficient between the model and target RDMs.
- Take the squared correlation to eliminate negative values.
- Normalize against the noise ceiling set by the Algonauts project.
- Calculate the average between the Early and Late targets for the MEG tract, and the average between the EVC and ITC targets for the fMRI tract..

The baseline scores for the Algonauts project were set using AlexNet, a deep neural network consisting of 5 convolutional layers 3 fully connected layers [Cichy et al., 2019]. Scores above this baseline are assumed by this research to imply that the model performs well enough to be considered as simulations of brain activity. The new model is considered an improvement if it obtains better scores than those reported by the challenge winners. These scores are shown in table 3.1.

	Early MEG	Late MEG	Avg nc %	EVC fMRI	ITC fMRI	Avg nc %
Baseline	5.82	22.93	15.32	6.58	8.23	7.41
A.-R. Meijer	1.20	28.51	16.37	13.67	17.37	15.53
A. Lage-C.	50.95	53.59	52.42	32.88	20.99	26.91

Table 3.1: Baseline scores of the Algonauts challenge, and the scores of the studied challenge participants.

# Chapter 4

## Results

### 4.1 Initial Model Estimation

#### 4.1.1 Perceptual Model

The perceptual model takes three parameters; the resize scale in pixels ( $n$ ), the lower edge detection threshold ( $t$ ), and the Gaussian smoothing sigma ( $s$ ). This model is expected to help explain early visual processing, and will therefore be optimized for the Early-MEG and EVC-fMRI targets.

The research of Lage-Castellanos describes a set of optimal parameters for predicting the EVC-fMRI and Early-MEG targets of the 92 image set, these formed a starting point for the optimization of the recreated implementation. The perceptual model was tested over various ranges of parameters to study their effect on performance. Figure B.1 in the appendix contains graphs illustrating the effect of changing these parameters on the 92 image set prediction scores. The tests revealed the model to favor a different set of parameters than those found by Lage-Castellanos. This is likely due to differences in implementation, as his model was created using Matlab instead of Python. Table 4.1 shows the scores of the model on the 92 image set both when using the initial parameters found by Lage-Castellanos, and when using the optimal parameters. Optimizing the parameters for the test set yields a different set of parameters, as can be seen in table 4.2.

	Target	Resize scale	Gaussian sigma	Threshold	nc %
Lage-C	Early-MEG	100	1	0.12	54.64
	EVC-fMRI	166	2	0.13	48.71
Optimal	Early-MEG	100	2	0.3	61.33
	EVC-fMRI	133	2	0.4	51.10

Table 4.1: Scores on the 92 image set using Lage-Castellanos' parameters and the optimized ones.

	Target	Resize scale	Gaussian sigma	Threshold	nc %
92 Optimized	Early-MEG	100	2	0.3	11.36
	EVC-fMRI	133	2	0.4	4.88
78 Optimized	Early-MEG	100	2	0.4	<b>17.08</b>
	EVC-fMRI	100	1	0.4	<b>13.69</b>

Table 4.2: Scores on the test set using parameters optimized for the 92 or 78 image sets.

The initial prediction estimates for the Early-MEG and EVC-fMRI RDMs were created using the parameters optimized for the 78 image set, their resulting scores are shown in table 4.3. These results show both estimates score more than double the challenge baseline score on each of the Early MEG, EVC fMRI and ITC fMRI targets. The score on the Late MEG target is very low however, which causes the average score for the MEG tract to fall below the challenge baseline. This result is in line with our expectation that the edge detection models activations of the V1 area of the visual brain, corresponding to the Early MEG and EVC fMRI targets. The good performance on the ITC fMRI target is unexpected, but suggests the edge detection model is also able to capture activations in the later stage of visual processing.

The RDMs are visualized in figure 4.1 to show they have similar patterns, this will be useful for comparison after combining the initial estimates with the categorical and CNN models.

	Early MEG	Late MEG	Avg nc %	EVC fMRI	ITC fMRI	Avg nc %
Baseline	5.82	22.93	15.32	6.58	8.23	7.41
Early-MEG	<b>17.08</b>	2.72	9.10	13.54	17.75	15.65
EVC-fMRI	15.23	3.15	8.52	<b>13.69</b>	18.63	16.17

Table 4.3: Performance of the perceptual estimates when predicting the 78 image set, compared to the challenge baseline.

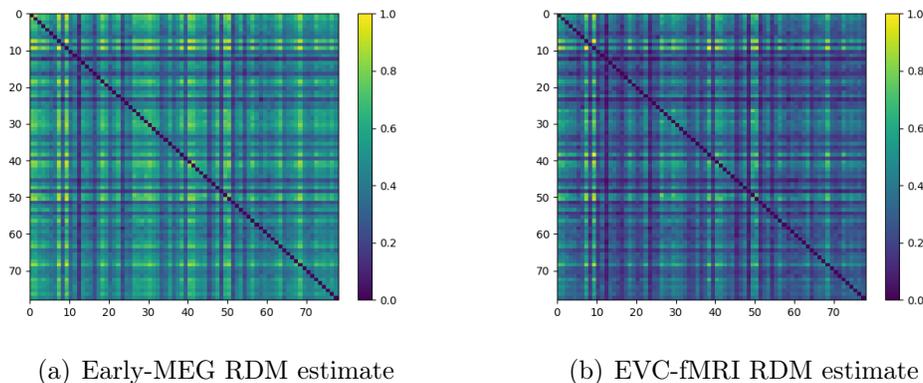


Figure 4.1: Initial RDM estimates for the 78 set using the perceptual model

### 4.1.2 Categorical Model

The categorical model is expected to be a better predictor of activity for the Late-MEG and ITC-fMRI targets, since it models more complex features than edges. Classifiers have been trained on either of the 92 and 118 image sets to observe their difference in performance on the test set. Table 4.4 shows the performance of each of the four resulting RDM estimates on predicting their respective targets.

		target score %
92 Classifier & 92 image set	Early-MEG	14.70
	Late-MEG	18.00
	EVC-fMRI	10.08
	ITC-fMRI	46.54
118 Classifier & 118 image set	Early-MEG	12.84
	Late-MEG	6.22
	EVC-fMRI	2.52
	ITC-fMRI	8.31

Table 4.4: Performance of different classifiers when predicting their own training set.

These initial tests revealed the categorical model using the 92 image classifier to be better at predicting the expected targets; the Late-MEG and ITC-fMRI targets. This effect was however not present when predicting the 78 image set, as can be seen in table 4.5. The 118 image classifier obtains the highest average score for both targets, for this reason the 118 classifier was chosen to generate the categorical estimate RDMs.

		target score %
92 Classifier & test set	Early-MEG	1.31
	Late-MEG	<b>2.93</b>
	EVC-fMRI	2.67
	ITC-fMRI	<b>3.70</b>
118 Classifier & test set	Early-MEG	4.29
	Late-MEG	<b>11.43</b>
	EVC-fMRI	5.41
	ITC-fMRI	<b>3.36</b>

Table 4.5: Performance of different classifiers when predicting the 78 image set.

The difference in performance compared to the training set can in part be explained by the low accuracy achieved by the classifier when categorizing images from the test set. This can likely be solved by improving the classifier, as the current implementation does not receive appropriate training. The research of Lage-Castellanos reports much higher prediction accuracy, and makes use of cross-validation to improve the accuracy the classifier. As an example, their classifier was able to categorize all human faces in the test set correctly,

of which our version is not able to predict a single one. Appendix A.3 and A.4 show the 78 image set as labeled by both classifiers, which illustrates their low accuracy.

Table 4.6 shows the scores for the Late-MEG and ITC-fMRI estimate RDMs generated by the categorical model. Both models perform below the challenge baseline on all targets, indicating they are not able to predict activations, including in the expected areas. It will still prove interesting if these estimates can contribute positively to the overall scores by combining them with the perceptual estimates.

	Early MEG	Late MEG	Avg nc %	EVC fMRI	ITC fMRI	Avg nc %
Baseline	5.82	22.93	15.32	6.58	8.23	7.41
Late-MEG	3.92	<b>11.43</b>	8.09	5.10	4.71	4.91
ITC-fMRI	0.96	0.88	0.91	3.20	<b>3.35</b>	3.28

Table 4.6: Performance of the categorical estimates when predicting the 78 image set compared to the challenge baseline.

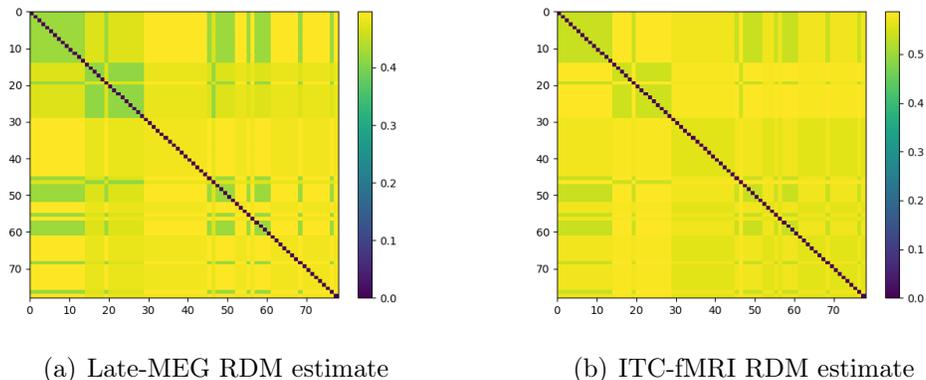


Figure 4.2: Categorical RDMs generated from the 78 image set using the 118 image classifier.

As with the perceptual estimates, the categorical estimates have been visualized in figure 4.2, to allow recognition of shared features. Clearly visible is the low variation in color, which is caused by the classifier placing most images in the same category.

### 4.1.3 CNN model

The CNN model uses the activations in the last fully connected layer of the VGG-19 network. This layer contains 1000 features per image, the correlation between two images is obtained by calculating the Spearman correlation between these feature vectors. As can be seen in table 4.7, this estimate does not perform well on the test set. However, in section 4.2.1 it will be shown that this estimate can still contribute to the performance of a combined model.

The CNN-based estimate has once again been visualized in figure 4.3, to allow comparison of features with other RDMs.

	Early MEG	Late MEG	Avg nc %	EVC fMRI	ITC fMRI	Avg nc %
Baseline	5.82	22.93	15.32	6.58	8.23	7.41
VGG-19 fc8	0.40	1.05	0.76	2.19	2.55	2.36

Table 4.7: Performance of the CNN-based estimate when predicting the 78 image set compared to the challenge baseline.

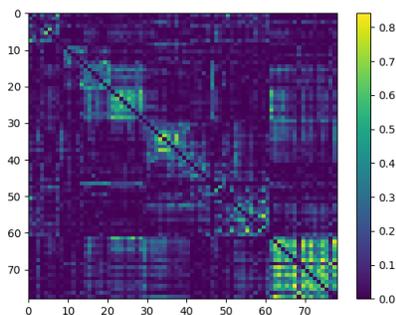


Figure 4.3: The RDM generated by the CNN model using the features of the fully-connected layer of the VGG-19 network.

## 4.2 Combining Models

This section shows how the different models are combined to generate the final scores. The categorical and perceptual RDMs are combined first by maximizing the score, followed by the contribution of the CNN activations. In both steps, models are combined by weighted average, using a set of weights that together add up to one:

$$M_{combined} = w_0 * M_{perceptual} + w_1 * M_{categorical}$$

$$M_{final} = w_0 * M_{combined} + w_1 * M_{CNN}$$

$$w_0 + w_1 = 1, \{w_0, w_1 \in [0, 1]\}$$

### 4.2.1 Combining Initial Estimate RDMs

Having obtained the initial estimates from the perceptual and categorical models, they are combined using the weighted averaging method. Because the scores of the categorical estimates were relatively low compared to the perceptual estimates, it was expected that their contribution would be limited. Table 4.8 shows scores of the optimal combinations found for the perceptual and categorical estimates. This shows the Late-MEG estimate did not benefit either of the perceptual estimates, and the ITC-fMRI estimate improves scores for the perceptual estimates, but only very slightly. The graphs showing the effect of combining the perceptual and categorical estimates have been included in appendix B.2

Perceptual est.	Categorical est.	$w_1$	Avg MEG nc %	Avg fMRI nc %
Early-MEG	Late-MEG	0	9.10 (9.10)	15.65 (15.65)
Early-MEG	ITC-fMRI	0.3	9.11 (9.10)	15.68 (15.65)
EVC-fMRI	Late-MEG	0	8.52 (8.52)	16.17 (16.17)
EVC-fMRI	ITC-fMRI	0.2	8.54 (8.52)	16.20 (16.17)

Table 4.8: Optimal combinations for the perceptual and categorical estimates when predicting the 78 image set, maximizing the average score. The numbers in parentheses are the scores of only the perceptual model.

### 4.2.2 Combining CNN Model with Combined Estimates

Next the four combined initial estimates are further combined with the RDM created by the CNN model. The effect of adding the CNN contribution can be seen in the graphs provided in appendix B.3. An interesting thing to note is that the CNN activations are not good predictions by themselves, but they do slightly increase the overall score in when combined to a degree. This effect is also observed in the research of Lage-Castellanos, where the CNN contribution improved scores far more; approximately 5 to 8 percent of the noise ceiling. It is expected that the level of improvement depends on the choice of CNN layer, as well as the quality of the initial estimates. Further research will be required to confirm these expectations.

In table 4.9, the improvement in scores is displayed after adding the weighted contribution of the CNN model, showing the optimal weight found for each combination of perceptual (Early, EVC) and categorical (Late, ITC) estimates.

Combined est.	CNN est.	$w_1$	Avg MEG nc %	Avg fMRI nc %
Early-Late	VGG-19 fc8	0.1	9.40 (9.10)	15.83 (15.65)
Early-ITC	VGG-19 fc8	0.1	9.54 (9.11)	15.91 (15.68)
EVC-Late	VGG-19 fc8	0.1	8.97 (8.52)	16.42 (16.17)
EVC-ITC	VGG-19 fc8	0.2	9.34 (8.54)	16.22 (16.20)

Table 4.9: Optimal combinations for adding the CNN contribution to the each pairing of the initial estimates when predicting the 78 image set. The numbers in parentheses are the scores of the model before the CNN contribution is added.

### 4.2.3 Final results

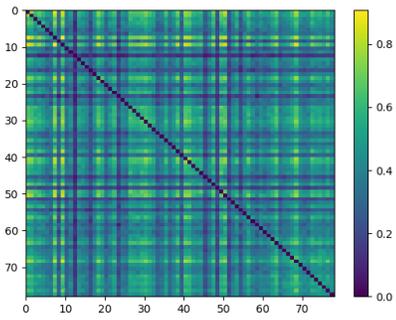
The scores of the final RDM estimates obtained after adding the CNN contributions are shown in table 4.10.

	Early MEG	Late MEG	Avg nc %	EVC fMRI	ITC fMRI	Avg nc %
Baseline	5.82	22.93	15.32	6.58	8.23	7.41
Early-Late	17.24	3.13	9.40	14.06	17.57	15.83
Early-ITC	<b>17.35</b>	3.29	9.54	14.21	<b>17.59</b>	<b>15.91</b>
EVC-Late	16.00	3.39	9.00	14.32	16.86	15.60
EVC-ITC	17.04	3.80	9.69	<b>14.48</b>	16.89	15.69
A.-R. Meijer	1.20	<b>28.51</b>	<b>16.37</b>	13.67	17.37	15.53
A. Lage-C.	50.95	53.59	52.42	32.88	20.99	26.91

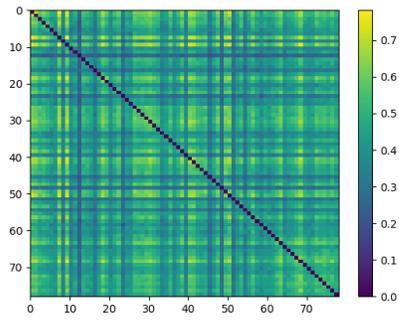
Table 4.10: Performance of the final estimates on the 78 image set after adding the CNN contribution, compared to the challenge baseline and the scores achieved by the discussed researchers. In bold print are the scores that improved upon the baseline and the scores of Meijer.

Table 4.10 shows that the final model is able to make acceptable predictions on the Early MEG, EVC fMRI, and ITC fMRI targets, improving upon the scores of the challenge baseline and the networks tested by Meijer. What is also visible is the much better performance of Meijer on the Late MEG target, which suggests this target prefers a CNN approach.

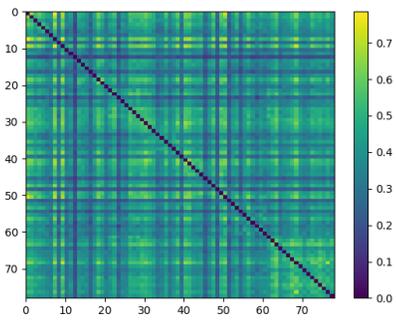
The RDMs are visualized in figure 4.4, showing the estimates closely resemble the perceptual estimates, which confirms the addition of the categorical and CNN estimates only slightly modifies the initial perceptual estimate.



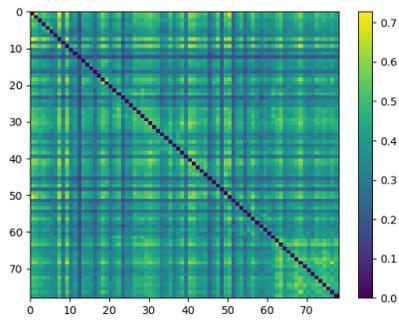
(a) Early-Late model



(b) Early-ITC model



(c) EVC-ITC model



(d) EVC-Late model

Figure 4.4: The final RDM estimates for the 78 image set after adding the CNN contribution.

# Chapter 5

## Discussion

The results section has shown that the most of the performance of the combined model can be attributed to the perceptual model. The contributions of the categorical and CNN models are very small, but may be increased by improving the respective models.

On the fMRI tract, all the final estimates outperformed every neural network tested in the study performed by Anne-Ruth Meijer, including her own ResNet-20 implementation which received 10 epochs of training [Meijer and Visser, 2019]. This shows promise for the perceptual model when compared to using CNNs for predicting fMRI data. It may be possible to improve the performance of the CNNs with more training, but it is not guaranteed as this can also cause overfitting the model on the training set.

On the MEG tract the performance on the Early-MEG target is very good, almost tripling the challenge baseline score. On the Late-MEG target the performance is well below the challenge baseline, which agrees with the low performance of the categorical model. The categorical model achieved high prediction accuracy on the training set, but did not perform well on the test set. Both of the tested classifiers got less than half the predictions right. See appendix A to see the images as labeled by the classifier. The model is expected to have improved performance when a better classifier is used.

In his report, Lage-Castellanos included graphs showing the optimal CNN contribution laid around 20 to 30 percent [Lage-Castellanos and De Martino, 2019]. In the performed study, the CNN contribution improved the overall scores at a ratio of around 10 to 20 percent, which resembles the results of Lage-Castellanos. Apparent from both studies is that the CNN layer activations are not good predictions by themselves, but can be used to increase the score of the estimates generated by other models.

# Chapter 6

## Conclusion

The study aimed to find models that explain neural activity in specific areas, corresponding to early and late stages in visual processing. The reproduction study has shown that a perceptual model based on edge detection is able to predict activity in the early stage for both MEG and fMRI data. This correlation was predicted by the hypothesis that the edge detection model corresponds to functions of the V1 area in the brain, where the measurements were taken. This indicates that it is possible the edge detection model is capturing the function of the V1 area to some degree, however further experiments will be required to confirm this.

The categorical model was chosen as a candidate for explaining activity in the later stage of visual processing, corresponding to functions of the V4 area of the visual brain where the measurements were (approximately) taken. The classifiers used by the model have low accuracy when categorizing images from the test set, which causes the categorical model to make poor predictions. Choice of classifier and training time likely have consequences for the prediction scores, which is an area for future research. With an improved classifier it will be possible to study if the categorical model is better at predicting activations in the later stage of visual processing.

The use of RDMs based on convolutional neural network (CNN) layer activations to improve initial estimates was proven successful by Lage-Castellanos [Lage-Castellanos and De Martino, 2019]. The same effect was observed in our results, however with a reduced relative improvement. Further research is required to find the best choice of CNN, layer, and method of converting activation values to RDMs.

The small size of the training and test data sets was an intentional design choice by the challenge organizers to make it easy for new researchers to get started, as well as to lower computational requirements. If this study is repeated using a larger set of data however, it is expected that the results allow for more accurate statements about relationships between brain functions and computational models.

The research concludes that a perceptual model shows promise for predicting neural activations in the early visual processing stage. Whether the categorical model is able to predict activity in the later visual processing stage is inconclusive, but has shown promise in other research. The contribution of the CNN layer does show promise for improving prediction performance, but did not have enough impact on performance to suggest correlation to either the early or late visual processing stages.

# Bibliography

- [Canny, 1986] Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, (6):679–698.
- [Cichy et al., 2019] Cichy, R., Roig, G., Andonian, A., Dwivedi, K., Lahner, B., Lascelles, A., Mohsenzadeh, Y., Ramakrishnan, K., and Oliva, A. (2019). The algonauts project: A platform for communication between the sciences of biological and artificial intelligence.
- [Creem and Proffitt, 2001] Creem, S. H. and Proffitt, D. R. (2001). Defining the cortical visual systems: “what”, “where”, and “how”. *Acta Psychologica*, 107(1):43 – 68. Beyond the decade of the brain: Towards functional neuronanatomy of the mind.
- [Grill-Spector and Malach, 2004] Grill-Spector, K. and Malach, R. (2004). The human visual cortex. *Annual Review of Neuroscience*, 27(1):649–677. PMID: 15217346.
- [Hämäläinen et al., 1993] Hämäläinen, M., Hari, R., Ilmoniemi, R. J., Knuutila, J., and Lounasmaa, O. V. (1993). Magnetoencephalography-theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics*, 65(2):413–497.
- [Hauke and Kossowski, 2011] Hauke, J. and Kossowski, T. (2011). Comparison of values of pearson’s and spearman’s correlation coefficients on the same sets of data. *Quaestiones Geographicae*, 30:87–93.
- [Hegd e and Van Essen, 2000] Hegd e, J. and Van Essen, D. C. (2000). Selectivity for complex shapes in primate visual area v2. *Journal of Neuroscience*, 20(5):RC61–RC61.
- [Hubel, 1968] Hubel, D. H., . W. T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of physiology*, 195(1):215–243.
- [Huettel et al., 2004] Huettel, S. A., Song, A. W., McCarthy, G., et al. (2004). *Functional magnetic resonance imaging*, volume 1. Sinauer Associates Sunderland, MA.
- [Joyce, 2019] Joyce, J. (2019). Bayes’ theorem. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2019 edition.
- [Kriegeskorte et al., 2008] Kriegeskorte, N., Mur, M., and Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, 2:4.

- [Krizhevsky et al., 2012] Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. pages 1097–1105.
- [Lage-Castellanos and De Martino, 2019] Lage-Castellanos, A. and De Martino, F. (2019). Predicting stimulus representations in the visual cortex using computational principles. *bioRxiv*.
- [Lee et al., 1998] Lee, T. S., Mumford, D., Romero, R., and Lamme, V. A. (1998). The role of the primary visual cortex in higher level vision. *Vision Research*, 38(15):2429 – 2454.
- [Ma et al., 2018] Ma, Y., Xiang, Z., Du, Q., and Fan, W. (2018). Effects of user-provided photos on hotel review helpfulness: An analytical approach with deep learning. *International Journal of Hospitality Management*, 71:120–131.
- [Meijer and Visser, 2019] Meijer, A. J. and Visser, A. (2019). A residual neural-network model to predict visual cortex measurements. In Beuls, K., Bogaerts, B., Bontempi, G., Geurts, P., Harley, N., Lebichot, B., Lenaerts, T., Louppe, G., and Eecke, P. V., editors, *Proceedings of the 31st Benelux Conference on Artificial Intelligence (BNAIC 2019) and the 28th Belgian Dutch Conference on Machine Learning (Benelearn 2019), Brussels, Belgium, November 6-8, 2019*, volume 2491 of *CEUR Workshop Proceedings*. CEUR-WS.org.
- [Meijer, 2019] Meijer, A.-R. J. (2019). Simulating the human visual brain using deep neural networks.
- [Russell and Norvig, 2003] Russell, S. J. and Norvig, P. (2003). *Artificial Intelligence: A Modern Approach (2nd Edition)*. Prentice Hall.
- [Schrimpf et al., 2018] Schrimpf, M., Kubilius, J., Hong, H., Majaj, N. J., Rajalingham, R., Issa, E. B., Kar, K., Bashivan, P., Prescott-Roy, J., Schmidt, K., Yamins, D. L. K., and DiCarlo, J. J. (2018). Brain-score: Which artificial neural network for object recognition is most brain-like? *bioRxiv*.
- [Serre et al., 2005] Serre, T., Kouh, M., Cadieu, C., Knoblich, U., Kreiman, G., and Poggio, T. (2005). A theory of object recognition: computations and circuits in the feedforward path of the ventral stream in primate visual cortex.
- [Simonyan and Zisserman, 2014] Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition.
- [Weisstein, 2020] Weisstein, E. W. (2020). Normal distribution. <http://mathworld.wolfram.com/NormalDistribution.html>.

# Appendix A

## Training and test images

The images of both training sets have been included with their correct labels, in order to illustrate both the limited size and specificity of subject matter as discussed in section 3.1. Also included are the test images as labeled by a Gaussian Naive Bayes classifier trained on either of the training sets, these are discussed in section 4.1.2. Incorrect labels are colored red to show the accuracy of the classifiers.



Figure A.1: The 92 image set with correct labels

118 image set

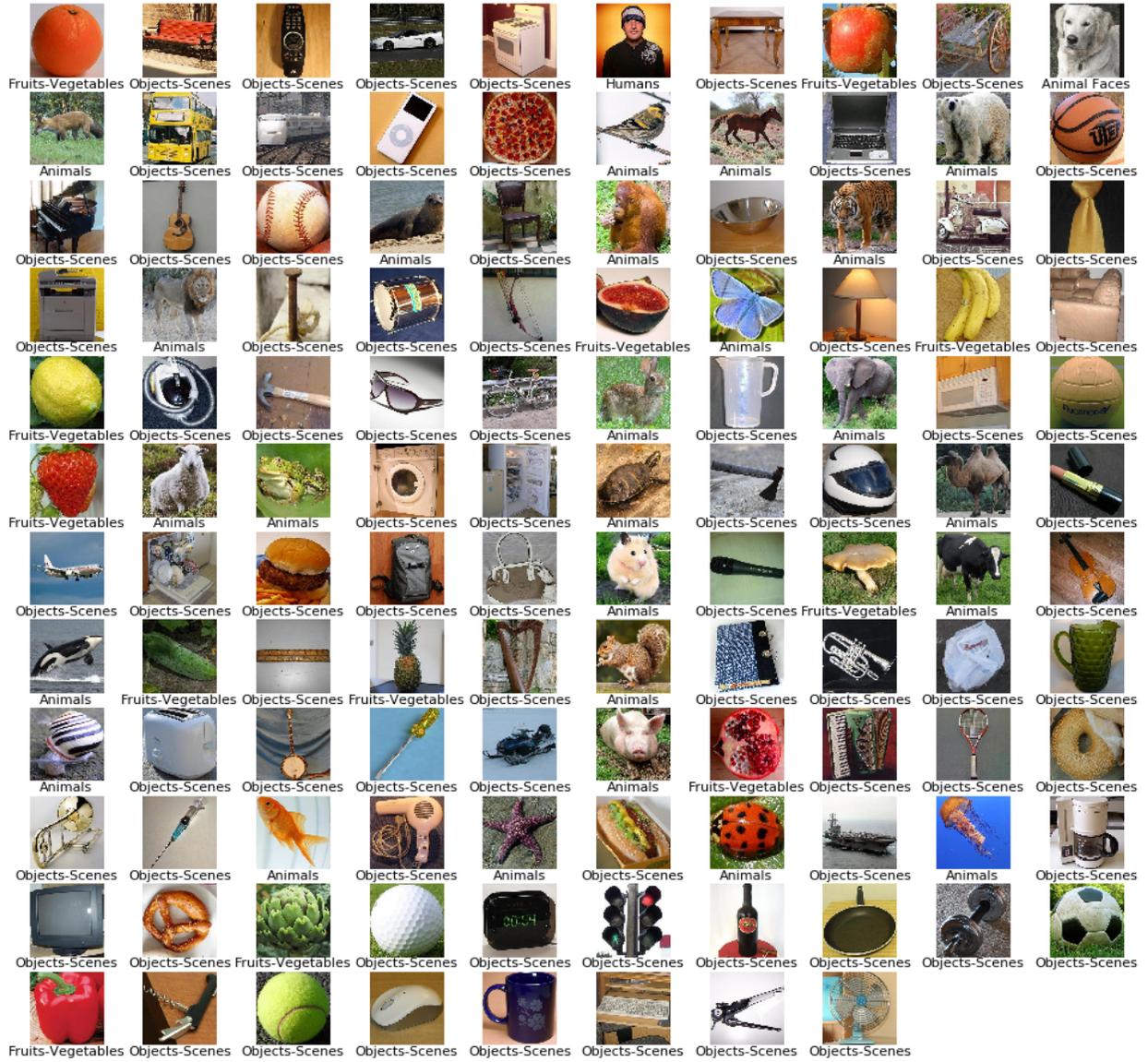


Figure A.2: The 118 image set with correct labels

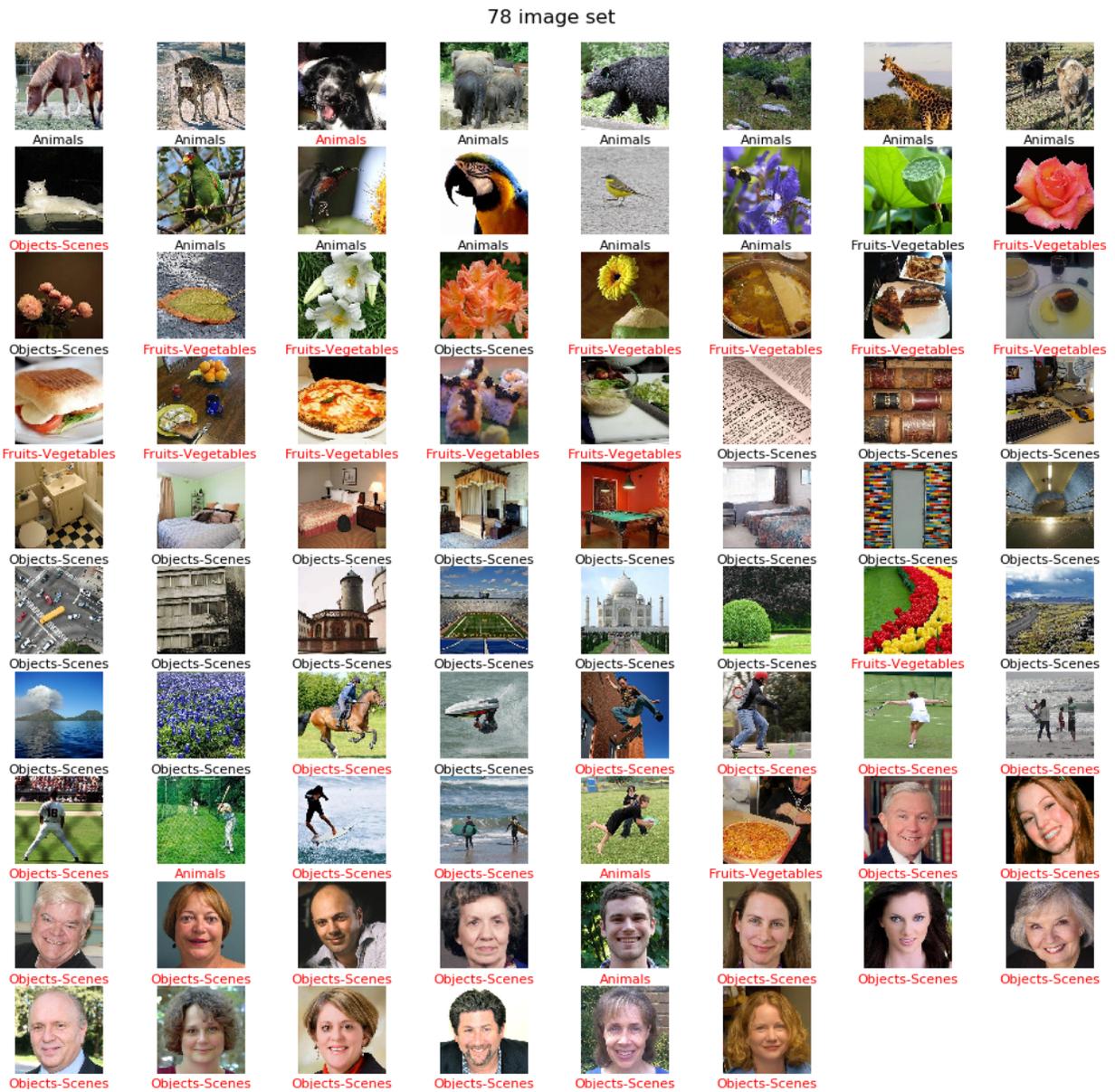


Figure A.3: The 78 image set as categorized by the 92 image classifier

78 image set

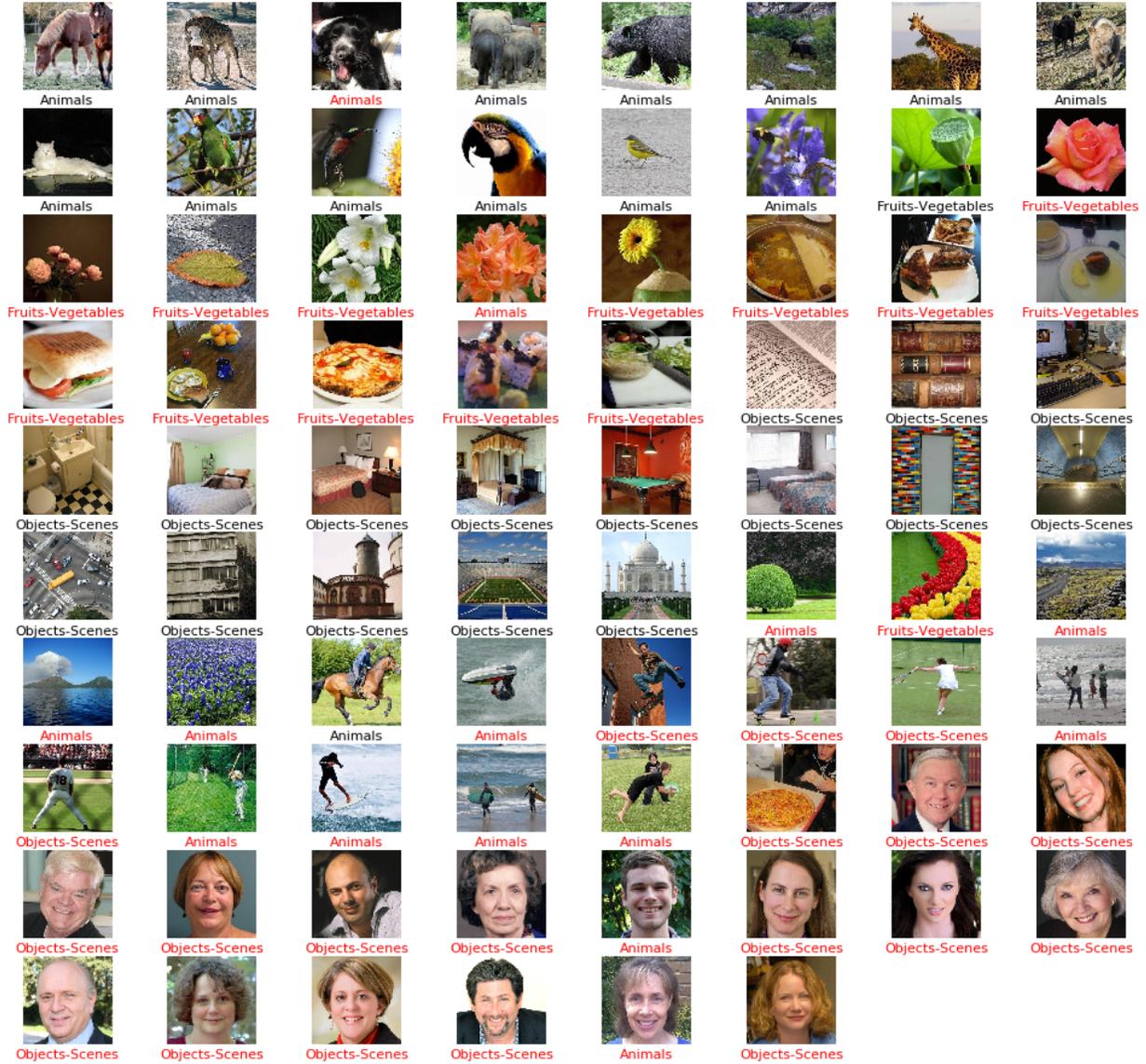


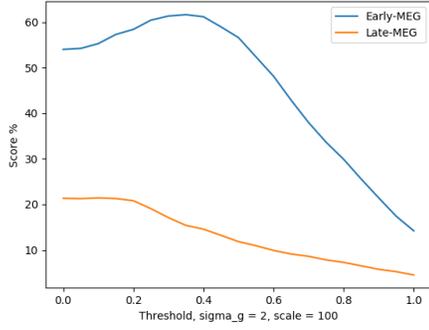
Figure A.4: The 78 image set as categorized by the 118 image classifier

# Appendix B

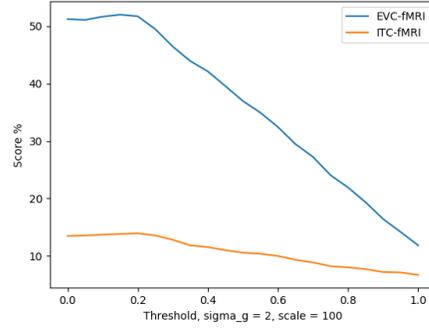
## Plotted graphs

The included graphs in figure B.1 illustrate the effect of varying specific parameters of the edge detection model, which is discussed in section 4.1.1.

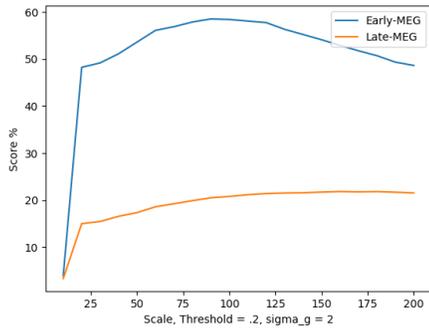
The graphs in figures B.2 and B.3 show the performance of different degrees of model combinations by varying the weight parameter. These are discussed in sections 4.2.1 and 4.2.2, respectively.



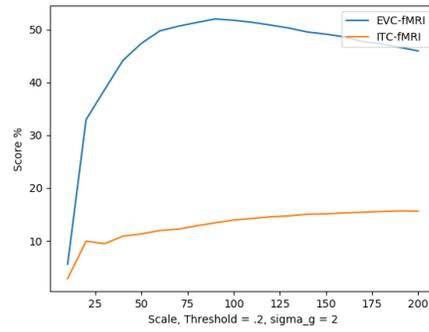
(a) MEG scores for variable threshold



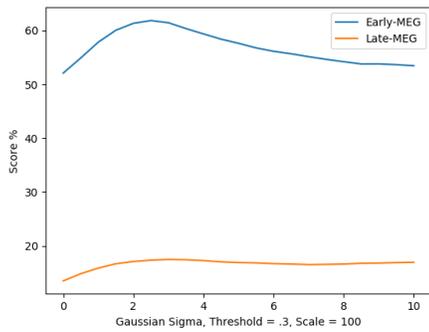
(b) fMRI scores for variable threshold



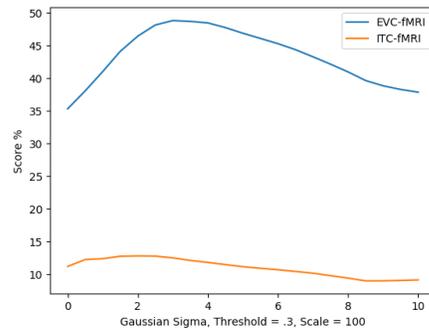
(c) MEG scores for variable resize scales



(d) fMRI scores for variable resize scales

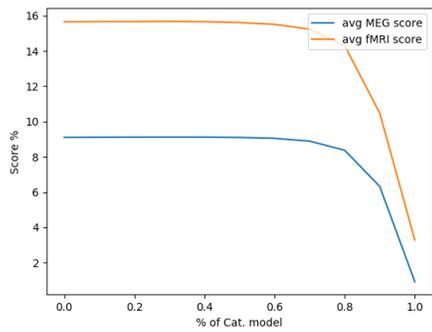


(e) MEG scores for variable Gaussian sigma

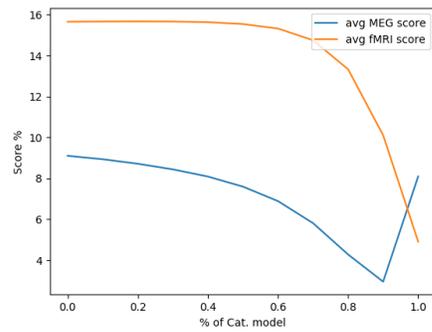


(f) fMRI scores for variable Gaussian sigma

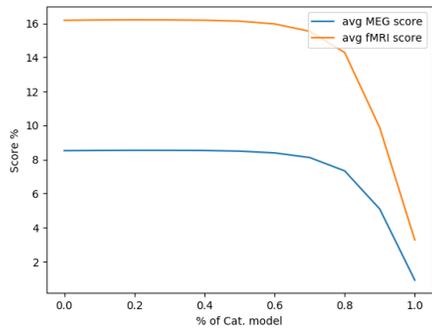
Figure B.1: Optimizing the parameters of the perceptual model for the 92 image set, with fixed resize scale and smoothing sigma.



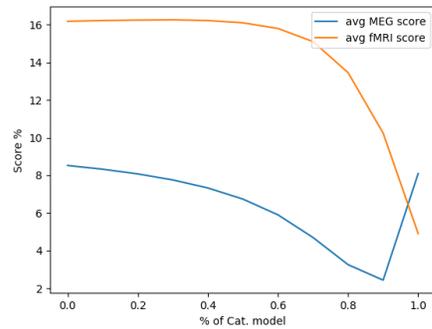
(a) Early-ITC model



(b) Early-Late model

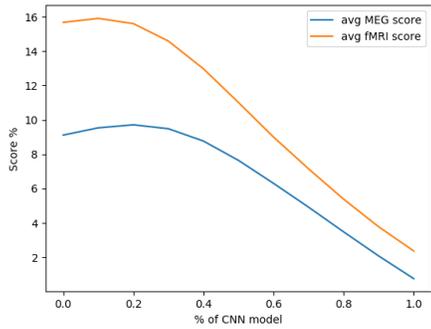


(c) EVC-ITC model

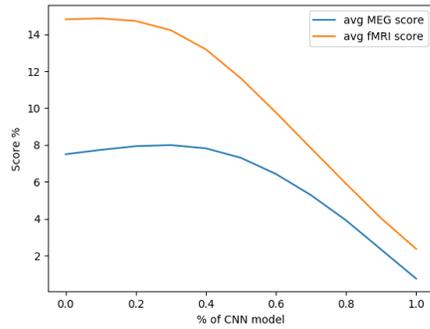


(d) EVC-Late model

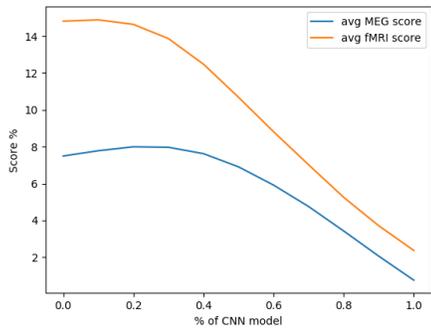
Figure B.2: Effect of combining the initial perceptual (Early-MEG, EVC-fMRI) and categorical (Late-MEG, ITC-fMRI) estimates on average scores.



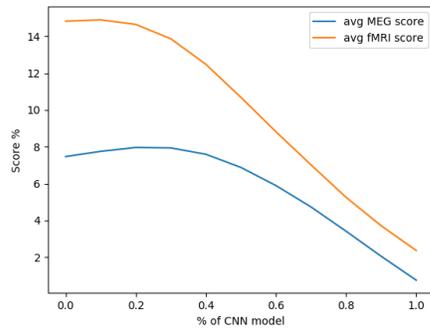
(a) Early-ITC model



(b) Early-Late model



(c) EVC-ITC model



(d) EVC-Late model

Figure B.3: Effect of combining the intermediate estimates with the CNN estimate. The CNN estimate is based on activations in the last fully connected layer of the VGG-19 network.