M.S. THESIS

Investigating the Relationship between Human Visual Brain Activity and Emotions

AUGUST 2019

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
COLLEGE OF ENGINEERING
SEOUL NATIONAL UNIVERSITY

Amelie Schmidt-Colberg

M.S. THESIS

Investigating the Relationship between Human Visual Brain Activity and Emotions

AUGUST 2019

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
COLLEGE OF ENGINEERING
SEOUL NATIONAL UNIVERSITY

Amelie Schmidt-Colberg

Investigating the Relationship between Human Visual Brain Activity and Emotions

지도교수 김건희

이 논문을 공학석사학위논문으로 제출함

2019 년 8 월

서울대학교 대학원

컴퓨터공학부

아멜리

아멜리의 석사학위논문을 인준함

2019 년 8 월

위 원 장	전 병 곤	(인)
부위원장	김 건 희	(인)
위 원	강 유	(인)

Abstract

Encoding models predict brain activity elicited by stimuli and are used to investigate how information is processed in the brain. Whereas decoding models predict information about the stimuli using brain activity and aim to identify whether such information is present. Both models are often used in conjunction. The brains visual system has shown to decode stimuli related emotional information [15, 20]. However brain activity in the visual system induced by the same visual stimuli but scrambled, has also been able to decode the same emotional information [20]. Considering these results, we raise the question to what extent encoded visual information also encodes emotional information. We use encoding models to select brain regions related to low-, mid- and high- level visual features and use these brain regions to decode related emotional information. We found that these features are encoded not only in the occipital lobe, but also in later regions extending to the orbito-frontal cortex. Said brain regions were not able to decode emotion information, whereas other brain regions and plain CNN features were. These results show that brain regions encoding low-, mid- and high- level visual features are not related to the previously found emotional decoding performance and thus, the decoding performance related to the occipital lobe should be contributed to non-vision related processing.

Amelie Schmidt-Colberg
Department of Computer Science and Engineering
College of Engineering
Seoul National University

Keywords: Encoding,Decoding,fMRI, Visual-System, Emotion, CNN

Student Number: 2017-28897

Contents

List of Figures

Chapter 1

Introduction

The understanding of the human brain, its' anatomical structure and specific computational processes has leveraged the development of powerful machine learning algorithms. Convolutional neural networks (CNN) for example, are knwon about visual processing in the brain. And the networks learning behaviour too, has been found to follow a similar hierarchical structure to the human visual system [29]. Within the field of neuroscience, CNN features were found to encode the neural activations all over the visual system [8]. This enabled researchers to verify existing hypothesis, construct models that span across brain areas and use more complex stimuli to probe brain processing and investigate new research questions. It becomes evident, how advances in either of these two fields, namely artificial intelligence and neuroscience, can be of benefit for each other.

One big challenge in artificial intelligence is the integration of emotions or emotional content, known as 'affective computing'. Emotional

processing in the human brain is still an area of ongoing research. Although regions associated with emotional processing have been identified, the exact underlying computational processes are unknown. Considering how little we know about the encoding of emotion processes, it seems natural that emotional integration within artificial intelligence systems has been proven to be a challenge. Thus, any light we can shed on how emotions are encoded within the brain may be helpful for improving affective computing algorithms.

So far a common consesus has been that sensory systems play an antecedent role in the theoretical explanation of emotions but are not central to the representation of emotional content [15]. Hence, activity in the visual system too, is thought to be antecedent, but not central to emotions. There has been recent research suggesting that the activity in the visual cortex is affected by emotional processes in a top-down manner; visual cortex activity has been shown to be enhanced when an emotional stimulus is present [12]. Additionally, brain activity in the visual system was able to sucessfully decode emotion related output [26]. These findings have generally been interpreted in the following way: Emotions cause a stronger sensory response in the visual system through attentive mechanisms, whilst emotional content is still believed to be represented elsewhere [12, 15]. However, one could also conclude that emotional content representations exist within the visual system. If the latter is the case, then visual features alone should be able to encode brain response to emotional content and said brain response should be able to decode related emotion labels.

Thus, the aim of this research is to investigate if visual features encoded in the brain also encode emotional content. By using the advances

on encoding and decoding models for the brain, we want to verify for one, if visual features encode emotion related brain response by testing whether visual features encoded in the brain decode related emotional labels. For this we use 2 datasets, which consist of fMRI time series recordings of individuals watching movie clips that have emotionally relevant content and fMRI time series of the same individuals watching a scrambled version of these movie clips. The idea is that the second dataset is stripped of any emotional and semantical meaning, which can be used as a control model to verify the presence or absene of emotional content for a given voxel.

Our approach can be devided into two parts. The first part consists of training two encoding models, which map the movie stimulus to the voxel brain activity for each dataset. Then voxels are selected based on whether the encoding model can predict the acitivty with stistical significane. The second part consists of decoding models, which use the brain activity predicted by the encoding models and predicts the emotional label for each movie. The performance of four different decoding models are compared; decoding using the significant voxels predicted by the first encoding model, decoding using the significant voxels predicted by the second encoding model, decoding using voxels not predicted by our encoding model and finally decoding using the features that were taken as input for the encoding model.

We found that amongst the four decoding models, voxels encoding visual information have the lowest decoding accuracy for emotional labels. Therfore we conclude that brain activtiy related to visual information processing does not contain any emotional content and that the brain activity in the occipital lobe is not soley related to visual processing.

Chapter 2

Background

2.1 Emotions and the Visual System

2.1.1 Visual system

The visual system, is one of the most studied brain regions because of its consistent organizational structure across mammalian species and the interpretability of its functionality. The visual system originates in the occipital lobe and consists of the primary visual cortex (V1), visual areas V2,V4,V5 and dorso-medial area (DM) [29]. Anatomically, the visual system can be separated into two pathways that extend to the parietal and temporal lobe [24].

Historically, those two pathways are labeled dorsal- and ventral stream, respectively (fig. 2.1). The two different streams are said to have different functionality; whilst the dorsal stream is associated with motion and representation of object location the ventral stream is associated with object recognition. The different visual areas within the stream and preceding

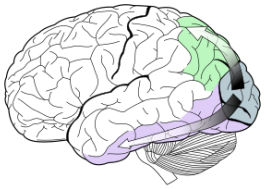


Figure 2.1: Ventral (purple) and dorsal (green) stream projections from the occipital lobe (blue)

it have found to have specific functionalities that increase in complexity following a hierarchical structure. V1 for instance, is known to encode low level visual features such as edges [13]. V2 has been shown to be tuned to orientation and spatial frequency [1]. V4 has been associated with textures, shapes, objects and object parts [24].However [17] have outlined that the ventral stream does not only consist of feed-forward hierarchical processing, but is subject to recurrent connections and top-down projections fromm other cortical regions. Even though the differentiation between ventral- and dorsal stream has been widely adapted, there have been arguments made that challenge this functional seperation. It has been argued that since the dorsal stream has projections to multiple cortical areas within the frontal, temporal lobe and limbic system its functionality should be much more extensive, involving visually guided action and navigation [16].This is suppored by the finding that features associated with object recognition have been able to predict activity in the dorsal stream further suggesting that the functional distinction between the two pathways may be not as strict [31].

2.1.2 Emotions

As emotions can be a very subjective experience, the universal representation of emotions is not straightforward. A common way to represent emotions is to use discrete classes such as *fear*, *anger*, *joy* and *sad*. This is a very intuitive representation of emotions, as it is in alignment with our everyday language. Such discrete classes, however, may not reflect the subtle differences and variety of emotional experiences of different individuals [10]. Another way of representing emotions is on a two-dimensional

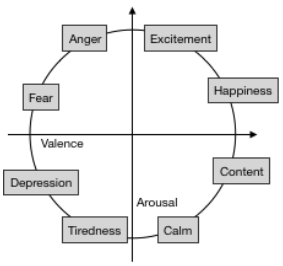


Figure 2.2: Emotion classes on the valence arousal scale

space, where the x- and y-axis represent valence and arousal respectively (fig. 2.2). Valence describes how pleasant or unpleasant an emotion is. Arousal describes the level of arousal, where the x-axis equates to a perfectly relaxed state [10]. Emotion processing within the brain has traditionally been associated with specific brain regions such as the limbic system which spans over the thalamus, amygdala, diencephalon and hippocampal formation, amongst others [15]. Especially the amygdala has been found to play a key role in emotion processing, where its functional roles does not only include processing of emotions but also emotional stimuli to memory [5]. Many other cortical regions have been found to also play a role in emotion processing, such as the septal nuclei, orbito-frontal cortex, cingulate cortex, insular and the perirhinal area [5]. Especially the

orbito-frontal cortex is established to be important for emotional processing as lesions to that area and consequent emotion related disfunctions suggest [5].

Sensory processing certainly does not play a minor role in the formnation of emotions. It's antecedent function has been well established; the amygdala receives information from the sensory cortices for further processing and is thought to be involved in the perceptual analysis of emotionally salient stimuli (eg. facial expressions) [5]. In a top-down manner such information is projected back to the occipital cortex to enhance the visual processing of such stimuli [5, 12]. More evidence for this is the fact that projections to the amygdala are very localized, whereas outgoing connections from the amygdala to the ventral stream target every subregion of the pathway [17]. It has also been proposed that the amygdala itself does not play a central role to emotion processing but combines important sensory information which is then passed on to other cortical regions [25].

2.2 functional Magnetic Resonance Imaging

Functional Magnetic Resonance Imaging (fMRI) is a fast, safe and noninvasive tool that allows scientist to record the reponse for a group of neurons to different stimuli. Unlike other brain measurement devices, fMRI records the response of a large group of neurons in so called voxels, the measurement unit of fMRI. One single voxel contains up to several millions of neurons. Thus, one of the main underlying assumptions within any fMRI study is that cognitive content and brain functions are encoded in large populations of neurons that are spatially distinct and distributed but connected in functional networks [11]. In general, fMRI studies have

the objective of understanding how the application of stimuli leads to changes in neuronal activity and thus are designed in a way that allows researchers to investigate the brain response to a particular stimulus or specific conditions [21]. Recorded brain activity is therefore labeled with the corresponding stimulus class or task [11]. For instance, a subject is shown images that contains faces or objects and the brain activity is labeled according to which image class was shown to the subject.

2.2.1 BOLD signal

fMRI does not record neural activity directly, but changes in blood oxygen levels that arise when changes in neural activity occur. An increase of neural activity in a region of the cortex is linked to an increase in the localized blood flow caused by a greater demand for oxygen and other substrates. That blood flow increase exceeds what is required by the neurons and consequently, there is a net increase in the balance between oxygenated aterial blood (oxyhemoglobin) and deoxygenated venous blood (deoxyhemoglobin) at a capillary level. The deoxyhemoglobin is paramagnetic and therefore deoxygenated blood differs in its magnetic properties. This means, when there is unbound oxygen present, the difference between the magnetic field applied by the MRI machine and the magnetic field close to a molecule of blood protein is greater than when the oxygen is bound. In other words, the increase of oxygenated blood makes the local magnetic field more uniform. Magnetic resonance images from such regions decay less rapidly and thus a stronger signal is recorded. This small signal increase is what is called BOLD signal [7]. The BOLD signal has three characteristic stages, first there is the resting state which equates to the situation when no stimulus is applied. Next is the initial dip, which is associated with the decrease in deoxyhemoglobin due to an

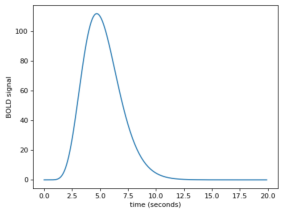


Figure 2.3: Model of HRF usually used in fMRI data analysis

increase in neural activity. This effect is very short-lived and rapid increase in signal occurs, after about 2s after the onset of neural activity. This signal increase usually persist for 5 to 8 seconds following the peak of the neural activation. The final stage is the signal decrease, marked by a so called *undershoot* which is due to the combination of reduced blood flow and increase of blood volume before the signal arrives again at the baseline [21]. This BOLD signal is theoretically modeled by the hemodynamic Response Function (HRF) (fig. 2.3) [7]. When using fMRI for analysis, one is usually interested in the underlying neural activity to the BOLD response. In most studies a linear relationship between neural response and neuronal activity is assumed and it is therefore common to convolve the recorded fMRI signal or the stimuli representation with the HRF function, depending on the analysis design [21].

2.2.2 Analysis of fMRI

Whatever the objective in analysing fMRI data is, they are all related to understanding how specific stimuli lead to changes in the recorded neuronal activity [21]. Commonly used methodologies have been statistical parametric maps (SPM) , multivariate pattern analysis (MVPA), population receptive field mapping (pRF) and representational similarity analysis (RSA) [23, 11, 21, 9]. SPM is testing hypotheses about region-

ally specific effects by fitting a general linear model (GLM) to each single voxel within a selected region of interest (ROI) in the brain, the statistical significance is assessed for each voxel and then aggregated across a ROI [23] [30]. MVPA commonly uses a classifier over a set of voxels,such as voxels within a specific ROI, to predict the related experimental condition [4] [30]. pRF mapping is used to reveal the functional organization of the cortex by using a small number of voxel-specific parameters which are estimated from the data. Finally RSA is used to determine whether a particular computational model matches the response patterns in a particular brain region [30, 6]. All of these methods can be regarded as a specific case of encoding (SPM, pRF) or decoding (MVPA) models [23, 30]. These two general models will be explained in the following sections.

2.2.3 Encoding Model

An encoding model predicts the response measured in a voxel on the basis of the experimental condition, such as a sensory stimulus. Encoding models provide an explicit computational model and can therefore be used to test and verify existing brain computational models like the processing of a sensory stimulus [23, 18]. Encoding models can be divided into two distinct parts. First, the experimental condition, such as a visual stimulus, is transformed into a feature representation. This part is usually called feature model. Second, the feature representations are used to estimate the brain , also called response model [23, 30]. The transformation from stimuli to featurespace can either be linear [27] or non-linear [14, 23]. The response model too, can either be linear, or non-linear [30], however a linear model is most commonly used. In both cases, using a non-linear model leads to higher prediction accuracy of the voxel response [30]. A linear encoding model that reaches the upper limit of prediction accuracy

can be formulated as:

$$\mathbf{W}\phi(\mathbf{x}) = \arg\max_{r} p(r|\phi(\mathbf{x})) \tag{2.1}$$

Where $\mathbf{W}$ is the linear mapping from the features to the voxel response, r is a single or multiple voxels response and $\phi(\mathbf{x})$ represents the feature transformation of a stimulus $\mathbf{x}$ [23]. Since an encoding model operates in the direction of information flow (from stimuli to brain response) it can provide a model of the brains computations at some level of abstraction. Thus encoding models can be used to test computational theories about the brain by implementing them and testing their ability to predict brain-response [18].

Interpreting encoding models, however, should be handled with care. Although an encoding model successfully predicts the brain response to a specific set of stimuli, one cannot conclude that the feature representation chosen uniquely encodes the measured brain activity profile [6, 18]. This is especially the case when a linear response model is used. The reason being that linear response models predict an activity profile as a linear combination of a set of model features. However, features spanning a particular subspace, are not unique; the same subspace could be described by a different set of features. [6, 18] used the term *feature fallacy* for the case of this misinterpretation. Therefore, [6] suggest that encoding models should be interpreted in comparison to each other, to avoid over-generalization of the encoding of a particular feature set.

2.2.4 Decoding Model

A decoding model is the inverse of an encoding model; decoding models are used to learn about a stimuli or cognitive state by observing the brain activity [23]. One commonly used decoding model is the linear classifier,

which takes the brain activity as input and outputs a class label, revealing which of the stimuli elicited the measured response pattern [23, 30]. If formulated as the inverse of an encoding model, a decoding model can be derived using Bayes theorem:

$$p((\phi(\mathbf{x})|r) \propto p(r|\phi(\mathbf{x}))p(\phi(\mathbf{x})) \tag{2.2}$$

Where again, r is a single or multiple voxels response and $\phi(\mathbf{x})$ represents the feature transformation of a stimulus $\mathbf{x}$ [23]. Being able to classify a stimuli from brain response however, does not give any information about the nature of the underlying neural code. In the case of images, independent from which visual region one might train a decoder on, all regions should be able to correctly classify an image stimulus despite the difference in the information encoded [18]. Consequently, when using a decoding model one can conclude the presence of information in a given region, but not what kind of information and how it is actually encoded. This is where the usefulness of encoding models becomes apparent. Where decoding models give us information of what kind of information might be present in a specific region, encoding models allow us to investigate how this information is represented. It is therefore encouraged to use both models concurrently [23].

When using decoding models, the input often consists of multiple voxels. Therefore one has to decide which voxels to select for decoding. The selected voxels are usually called a region of interest (ROI). Such a ROI can be selected based on anatomical structure, beliefs a researcher has about a specific area or based on significant voxels obtained from an encoding model. A ROI can be as large as a the whole brain image and a ROI does not need to consist of contiguous voxels and can be scattered.

2.3 Related Work

To this end, the usage of convolutional neural networks as a feature model within encoding models for the visual system has been shown to be very successful [18]. In [8] this approach was attempted for the first time, the authors used the CNN-S Network [3] layers as features, to encode a single voxels' peak BOLD activity caused by visual grey image stimuli. In this study all the layers of the network were trained to predict a single voxels' activity and the most predictive layer was selected using cross-validation on the training dataset. The authors identified a hierarchy within the resulting layer-brain response mapping which is similar to the hierarchy of the visual system; early layers of the CNN were most predictive of early layers of the visual system whereas later layers of the CNN were most predictive of later areas of the visual system. Together with [28] this study did not only correlate a CNN hierarchical structure to the hierarchical structure of the visual system but also achieved a mapping to intermediate-level areas of the visual system for the first time. It should be noted that in both, [8] and [28] the network from which the features were extracted from, was trained on a much more diverse set of images than the images used for eliciting the brain response; differing on colour scale (grey-scale vs rgb) and also containing much more object classes. In [31] a more naturalistic stimuli set was chosen to elicit the brain activity. More specifically, the authors used video clips representing diverse real-life visual experiences, such as clips showing people in action, moving animals and nature scenes. They used the Alexnet [19] CNN structure as a feature model and fine-tuned the network on a smaller range of classes which were present in the stimuli dataset. Each video frame was given as an input to the network and features were extracted for each layer. The

feature time-series was then down-sampled to match the fMRI sampling rate. Using correlation analysis between the CNNs feature time series and the BOLD response time series a similar hierarchical mapping to [8] was identified. Not only were previous findings verified, but it was also shown that the CNN as a feature model extends to naturalistic time dependent movie stimuli for encoding. Furthermore, when using the correlated layer to predict the BOLD response, [31] were able to have above chance prediction accuracy beyond the visual cortex, extending to the ventral stream as well as the dorsal stream. The fact that the features of a CNN trained for action recognition were able to encode voxels within the dorsal stream suggests that the functional separation between the dorsal and ventral stream is not strict. This strict functional seperation and the hierarchical mapping has been further challenged by [2] who showed that low- high- and semantic-level features encode voxel activity across the ventral and dorsal stream and extend all the way up to the frontal cortex.

[15] fine-tuned Alexnet CNN on emotional categories. They used regression to correlate patterns of the last layer of their network to patterns of recorded fMRI response. They found that categories such as *entrancement* and *sexual desire* were most strongly correlated, whilst categories such as *horror*, *fear* and *excitement* had the lowest correlation. The high- and low correlation of these discrete classes cannot be explained by grouping them in terms of valence and arousal. The authors also used the recorded brain activity to decode the features in the last layer of the CNN. The authors intuition was, that if a CNN can classify emotion classes correctly, and the brain activity elicited by an emotional image stimuli is able to decode the CNNs emotion category related features, then emotional category representations should be present within

the human visual cortex. They found that the occipital lobe proved to be more accurate for decoding emotion categories than other brain areas and individual visual brain areas (V1-V4). The authors concluded that information about emotional categories are present within the occipital lobe. But these results do not mean that brain activity related to vision, contains emotional representations. Nor does it imply that other brain regions, do not contain emotional content. [15] results show that information about emotional category CNN **features** are more present across the occipital lobe compared to individual areas, other areas or the whole brain. To what extent actual emotional categories or content are encoded within the occipital lobe and to what extent visual information processesing plays a role is still unknown.

[20] decoded human annotated arousal and valence labels using brain activity elicited by emotional movie clips. They divided the brain into 200 ROIs and found a high correlation between decoded arousal labels true arousal labels for ROIs located in the visual cortex. Theses ROIs outperformed the ROIs present in other brain areas. Decoding for valence was most strongly correlated using ROIs in the frontal cortex and visual areas. [20] also used a control run where they recorded the brain activity to the same stimuli with the movie clips pixels scrambled. Surprisingly,the decoding accuracy on arousal for the control run was also high in visual areas. In the case of arousal this decoding performance was mainly located in the visual cortex where as in the case of valence, the high accuracy regions were spread out.

[20] findings suggest the the decoding accuracy in the visual cortex cannot be attributed to visual processing of the stimuli as brain activity caused by non-sensical stimuli had high decoding accruacy for the same

areas. From [15] and [20] the question arises to what extent the visual system plays a role in emotional processing and possibly encoding emotional content. [15] findings suggest that some information of emotions is present in the occipital lobe, and thus the visual system. Whereas [20] findings suggest that emotional decoding performance in the visual system is not related to the actual emotional content of the stimuli being encoded and might just be related to noise, light influx,high neural activation for porcessing visual stimuli, processing antecedent to emotion or top-down processing.

Chapter 3

Materials & Methods

From the previous discussion the question arises to what extent the brain activity in the visual cortex contains any emotional information. More specifically, is there a possibility that visual processing encoded within voxels in the visual system also contains information about emotions. The underlying idea to our approach is the fact that encoding models give a computational model for the processing of information. Constructing a single voxel encoding model from visual stimuli to the recorded brain activity allows us to identify voxels that encode visual information. Using a variation of encoding models that are known to be able to predict brain activity across the visual system [31, 8, 28] and the same data from [20], movie clips and scrambled movie clips, we can construct separate models for specific voxels that encode visual and possibly emotional/semantic content and voxels that encode visual information only. If emotional content is present within visual processing then the decoding performace from the voxels selected by the encoding model using the

sematically relevant movie clips should be much higher than the voxels selected by the encoding model using the scrambled movie clips. Even-though both voxel groups encode visual information only, the information that has a semantically interpretable content only should be able to decode the emotion labels. If however, both voxel sets display no difference in decoding performance then it is unlikely that emotional content of the stimulus is present within visual processing related brain response.

3.1 Experimental data

The fMRI dataset used in this study consists of two parts. 64 subjects brain response was recorded for two different stimuli sets. The first set consists of 78, 15s long movie clips taken from the Hollywood2 dataset [22]. The second stimuli set consits of the same clips, but this time the pixels of each frame were scrambled, such that there was no semantically interpretable content left [20] . The movie clips were selected in

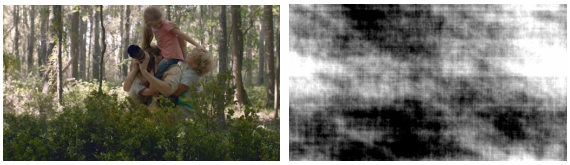


Figure 3.1: left: example frame for the movie clips. right: example frame for the scrambled movie clip

such a way that they were well spread over the valence arousal scale (fig. 3.2). For both datasets the movie clips were shown to the subjects over 3 different sessions. Each subject annotated the clips with their individual valence and arousal score that ranged from -4 to 4. The labels were taken to be the same fore the scrambled movie dataset. For each run session, the fMRI brain response to the first and last movie clip was

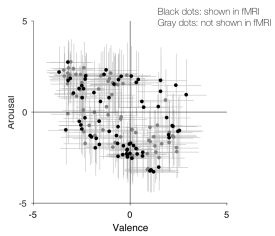


Figure 3.2: distribution of selected movie clips on valence arousal scale

excluded. This leaves a total of 72 movie - fMRI response pairs for each subject. The fMRI response was recorded at $1.5Hz$ giving 10 timepoints per 15 second video clip. fMRI pre-processing steps consisted of 3D motion correction, fieldmap correction, linear trend removal, high-pass filter and spatial smoothing with a Gaussian kernel. All data was aligned to a standard 2-mm MNI space. Voxels with low mean signal (2 std below average) were also removed. All data was z-scored over time. Voxels that displayed significant activity over the run where selected using freesurfer, leaving a brain mask consisting of 20005 voxels spanning over the whole cortex (fig. 3.3) [20].

3.2 Encoding model

The transformation between videos and fMRI response is done using an encoding model. This transformation consists of two steps. First, a feature model transforms the visual stimulus to a non-linear feature representation. Second, a response model uses the feature representations to predict a single voxel response. More specifically, if $\mathbf{x} \in \mathbb{R}^p$ and $\mathbf{y} \in \mathbb{R}^q$ is a stimulus-response pair where $\mathbf{x}$ is a vector representing an input frames and $\mathbf{y}$ is a vector of the voxel responses. Where p and q corresponds to the number of pixels and voxels, respectivley. Then given $\mathbf{x}$ the problem

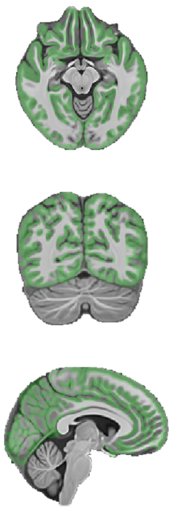


Figure 3.3: Visualization of the resulting brain mask

can be formulated as:

$$\hat{\mathbf{y}} = \arg\max_{y} p((\mathbf{y})|\phi(\mathbf{x})) = \mathbf{B}^T\phi(\mathbf{x}) \tag{3.1}$$

Where p is the encoding distribution of $\mathbf{y}$ given the non-linear transformation form the stimulus to the feature space denoted by $\phi(\mathbf{x})$ and $\mathbf{B}$, linearly transforms the features to the voxel response. Hereafter, the non-linear transformation of the stimulus $\phi(\mathbf{x})$ is denoted by $\mathbf{Z}$. Because CNN feature layers have been shown to encode voxels across the visual system the Convolutional Neural Network (CNN) AlexNet [19] was used for the non-linear feature transformation. The network consists of eight layers of which five are convolutional layers and three fully connected layers. Since our dataset consists of a movie stimulus there is an additional time dependence to our dataset. To handle this, we extracted feature representations for each frame of the movie clip providing a feature time series. The extracted features of the CNN are of high dimensionality, especially

compared to the number of samples available in our dataset. To avoid the issue of high variance that arises when the number of features strongly exceeds the number of samples, we performed principal component analysis. The number of components were chosen such that more than 98% of the variance was explained. The time series feature representations were standardized and then down sampled to match the fMRI sampling rate. Finally the representations were convolved with a hemodynamic response function to model the hemodynamical delay of the BOLD response obtained from the recorded fMRI data. To this end, the feature model is similar to [31]

The second part of the encoding model is the response model,for which we chose linear regression:

$$y_i = \beta^T \mathbf{Z} + \epsilon_i \tag{3.2}$$

Where y_i denotes a single voxel response and $\beta_i \in \mathbb{R}^{m \times q}$ is the linear transformation from the feature space. As mentioned above we are dealing with a high dimensional problem, so we chose to penalize our linear model with L2-norm to prevent our model from overfitting. Thus β_i can be estimated by:

$$\hat{\beta}_i = \arg\min_{\beta_i} \frac{1}{N} \sum_{j=1}^{N} (y_i^j - \beta_i \mathbf{Z^j} + \lambda \left\| \beta_i \right\|_2^2 \tag{3.3}$$

Where N denotes the number of samples. The solution for β_i is:

$$\hat{\beta}_i = (\mathbf{Z}^T \mathbf{Z} + \lambda_i \mathbf{I}_m)^{-1} \mathbf{Z} \mathbf{Y_i} \tag{3.4}$$

Where $\mathbf{Y}_i$ is a single voxels response values for N samples. (**??**) is the solution for to single encoding model, mapping stimuli to a single voxels response. For the feature model $\phi(\mathbf{x})$, the first seven layers of the CNN

were all considered and trained on predicting a single voxels activity separately. For a brain mask of 20005 voxels, this gives a total of 7×20005 seperate models to train. The λ parameter as well as the layer assignment was estimated using 9-fold cross-validation on the training dataset which was set to be 90% of the full dataset size. The prediction accuracy is

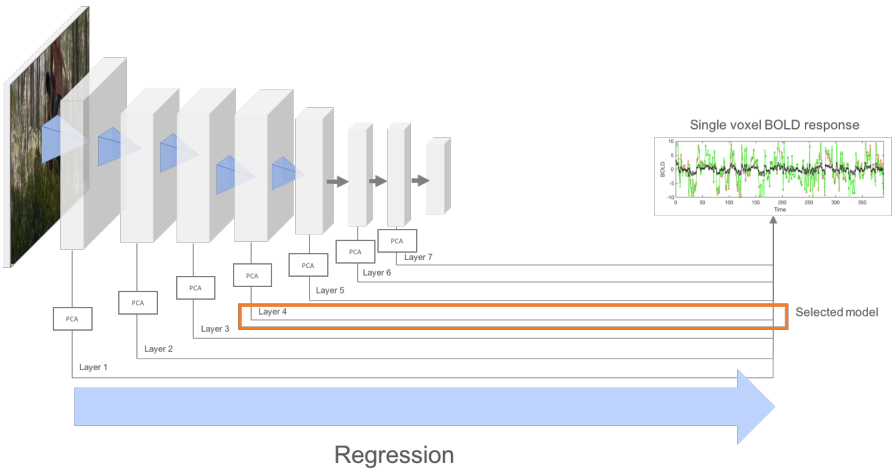


Figure 3.4: Illustration of the encoding model design

taken to be the pearson correlation between the predicted response and the actual response. The most predictive layer was assigned using 7-fold cross-validation on the training set. If the p-value was above a threshold value of 0.001 the voxel was discarded. An encoding model was trained for both ,the movie clip and the scrambled movie clip dataset.

3.3 Decoding Model

Decoding was performed across four subjects. The goal of decoding is to identify whether voxels encoded by the movie-model are able to decode the emotional label as well as whether they have higher predictive power

than the voxels encoded by the scrambled-model. The valence and arousal labels were normalized across all subjects. The voxel response was averaged over each movie clip.Support Vector Regression (SVR) was chosen as a decoding model c onsidering the dataset size, and its common usage in decoding from fMRI data. . The goal is to minimize:

$$C \sum_{n=1}^{N} (E_\epsilon(\hat{y}(\mathbf{x_n}) - t_n) + \frac{1}{2} \| > \mathbf{w}\|^2 \tag{3.5}$$

Where y_n is the true value and $\hat{y}(\mathbf{x_n})$ the predicted value. E_ϵ represents an error function defining the insensitive region. The regularization parameter C and kernel were chosen using grid search cross-validation. Only voxels that were significantly predicted by the encoding model were used as input for the regressor. The dataset was shuffled randomly and trained using a 10-fold split. The model was evaluated using the average pearson correlation between the predictions and the true labels for 10 different random splits. Separate decoding models were trained for the movie-model and the scrambled-model. Next we encoded from a random set of voxels within the brain mask which excluded voxels predicted by the encoding model. To compare to what extent specific encoded features play a role, decoding from voxels with the same layer assignment was performed. Finally, the encoding model features (CNN layers) were also used to decode emotion. To account for the high dimensionality of the features PCA was performed, where the number of components was set to be equal to the number of components used for training the encoding model.

Chapter 4

Results

4.1 Encoding

From the 20005 voxel large brain mask, depending on the subject, 551 to 1253 voxels were predicted significantly for the movie-clip dataset and 626 to 1947 were predicted significantly for the scrambled-clip dataset (fig. 4.1). Significant voxels were selected such that the estimated p-value was less than 0.001 on the test set. For the remaining voxels, the corre-

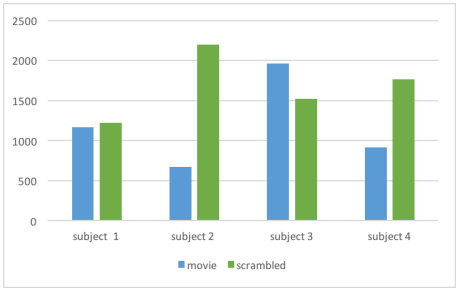


Figure 4.1: Number of significant voxel corellations for movie- and scrambled clips.

lation measured using pearson r ranged from 0.2 to 0.7. In the following figures only voxels whose correlation was above 0.4 were visualized. When

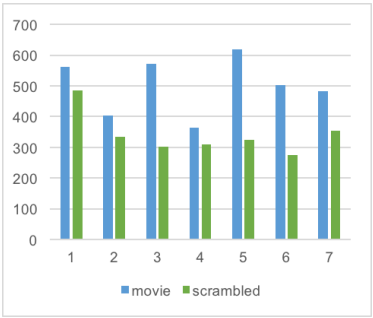
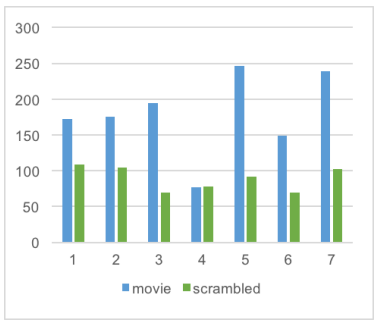
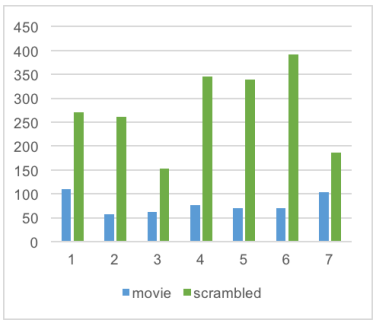


Figure 4.2: Layer assignment distribution for significant voxels across all subjects.

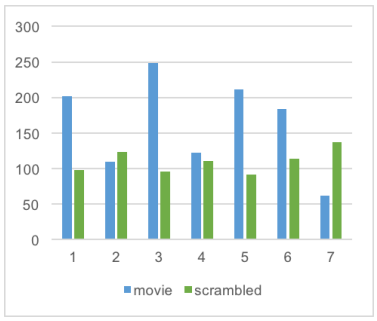
comparing the number of voxels assigned to a specific layer across all subjects for the movie clip dataset (fig. 4.2) the overall number of assigned voxels for the movie-clip dataset exceeds the number of of voxels for the scrambled-clip dataset. For layer 3 and layer 4 however the difference between the two is rather small. This observation does not transfer to the distribution of the layer assignment within subjects. For subject 1 and 3 the movie-clip dataset has more voxels than the scrambled-clip dataset. Whereas this difference becomes marginal for layer 4 (figs. 4.3a and 4.3c) . For subject 2 and 4 , the scrambled dataset has more significantly voxels predicted across all layers except for layer 6 in subject 4 (figs. 4.3b and 4.3d). A similar trend exists for the layer assignment for individual subjects with the difference that number of voxels assigned differ (fig. 4.3). No such trend can be observed for the scrambled-clip dataset. However, the number of significant voxels across all subjects exceeds those for the movie-clip dataset except for layer 5 and 7. For Subject 1 the number of significant voxel for the movie-clip dataset is more than



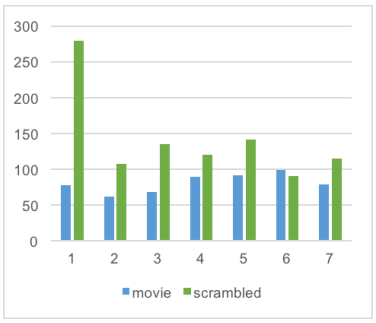
(a) Subject 1



(b) Subject 2



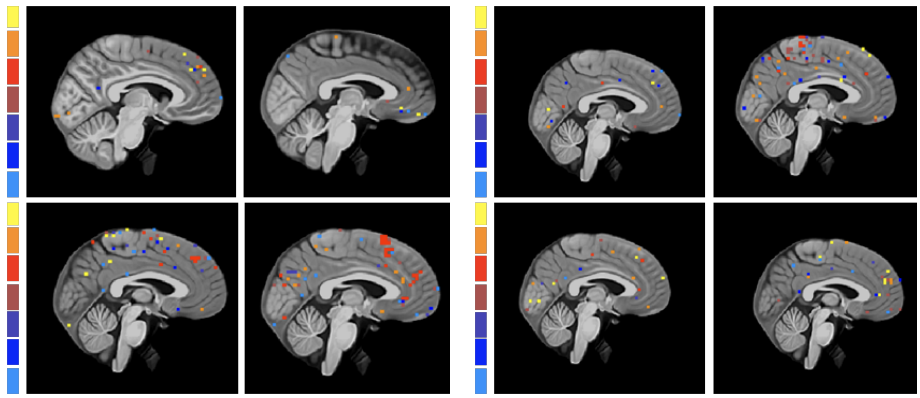
(c) Subject 3



(d) Subject 4

Figure 4.3: Layer assignment distribution for significant voxels within subjects

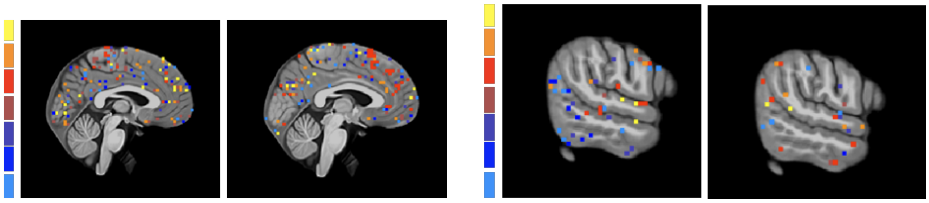
for the scrambled-clip dataset except for layer 5 and layer 7 (fig. 4.3a). Subject 3 has more assigned voxels for the movie-clip dataset except for layer 2 and 7 (fig. 4.3c). Subject 2 and Subject 4 on the other hand have significantly more voxels assigned for the scrambled-clip dataset than the movie-clip dataset (figs. 4.3b and 4.3d). When visualizing the distribu-



(a) movie clip (b) scrambled clip

Figure 4.4: Visualization of voxel layer assignment accross brain mask. Where light blue shaded voxels represent layer 1 and yellow shaded voxels represent layer 7

tion of the significant voxels across the brain, one can observe that the voxels are not only located in the occipital cortex (visual system) or the ventral stream (object recognition) but can be found to be spread across the parietal cortex, temporal lobe all the way to the frontal lobe. Groups of significant voxels related to the movie clip stimulus were found to be clustered at specific regions in the brain, such as prefrontal cortex, lower-lateral-occipital area and lower-lateral temproal lobe (**??**). The significant voxels associated to the scrambled movie clips are spread all over the brain mask, where the upper frontal and parietal lobe have higher voxel densities (fig. 4.5a). In all brain regions voxels assigned to all 7 layers can



(a) Whole brain mask (b) Ventral stream

Figure 4.5: Visualization of voxel layer assignment for all subjects. Left: scramble clip related voxels. Right: Movie clip related voxels.

be found. Meaning that low- mid- and high- level visual features encode voxel activity across the brain. When focusing on the ventral stream the voxels associated with the scrambled movie clip are mainly assigned to early layers and spread down towards the temporal lobe whereas the voxels associated with the movie clips are mainly located at the lower lateral occipital lobe (fig. 4.5b). This is in alignment with the fact that scrambled image stimuli are known to lead to higher activations in later visual areas, associated with increase in computational processes invovled trying to recognize the stimuli.

4.2 Decoding

The results for decoding the valence-arousal labels for four subjects using the CNN features are shown in (fig. 4.6). Whereas decoding accuracy for arousal is consistent and slightly increasing, except for a small dip at layer 5, The decoding accuracy for valence drops significantly for intermediate layers 2-4 and the later layer 7. The p-value for all predictions was lower than $0.1e-6$, except for layers 2-3.

Decoding was also performed from the brain activity, where voxels where selected according to whether they were able to be significantly

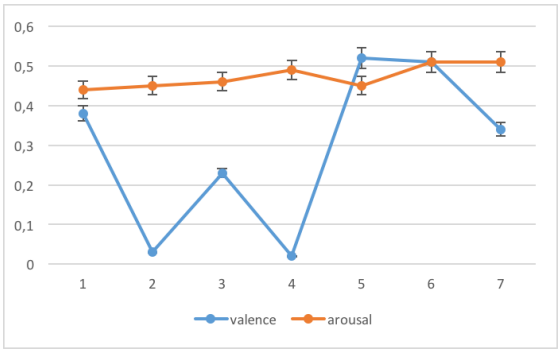
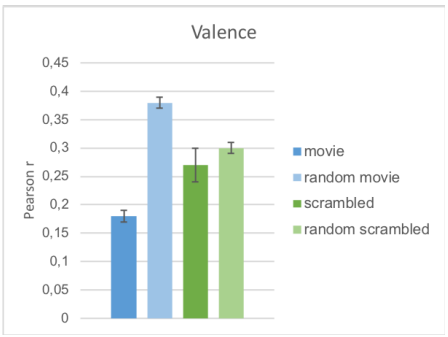


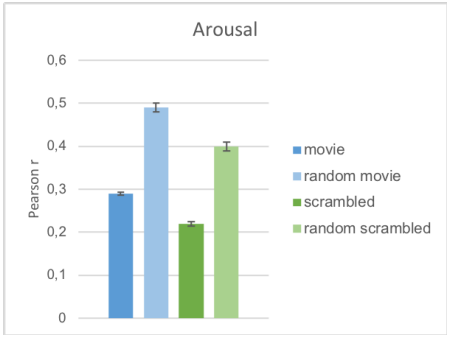
Figure 4.6: Decoding accruacy for Alexnet feature layers, measured using pearson correlation between the predicted and true labels.

encoded. This was performed for the brain activity related to the movie clip stimuli and the scrambled clip stimuli. Decoding was also done using random voxels from the brain mask with the constraint that they were not encoded previously. In other words decoding was performed on random voxels that do not encode visual features. The decoding performance for valence is lowest for voxels associated with the movie clip stimuli ($0,18$ pearson r and p-value $0,28$) (fig. 4.7a).Whereas random voxel from the same brain activity had the highest decoding performance ($0,38$ pearson r and p-value $0,04$) (fig. 4.7a). Voxel associated with the scrambled stimuli did not perform as well ($0,22$ pearson r and p-value $0,21$ for movie-clip voxels, $0,30$ pearson r and p-value $0,15$ for scrambled-clip voxels) but still better than the voxels encoded by visual features (fig. 4.7a).

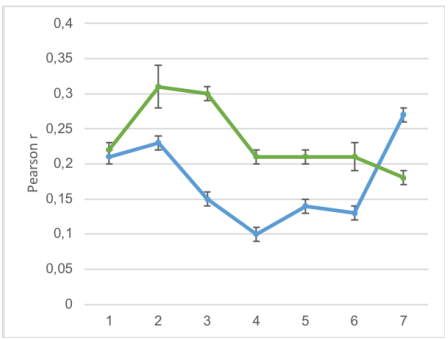
The decoding performance of arousal was highest for random voxel associated with the movie clip ($0,49$ pearson r and p-value $0,04$) followed by voxels containing brain activity related to the scrambled movie clips ($0,40$ pearson r and p-value $0,16$) (fig. 4.7b). Where as encoded voxel for the scrambled movie clip decoded arousal with the lowest accuracy ($0,22$ pearson r and p-value $0,35$) followed by the movie-clip voxels ($0,29$



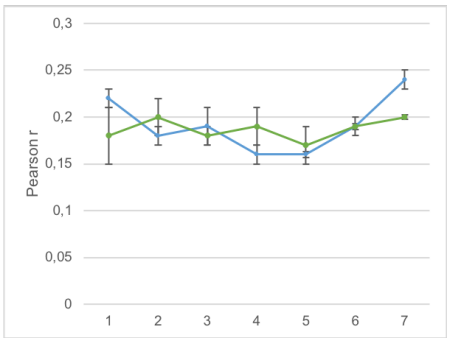
(a) Overall decoding performance



(b) Overall decoding performance



(c) Layer wise decoding performance



(d) Layer wise decoding performance

Figure 4.7: Decoding accuracy for arousal-valence labels for significant voxels

pearson r and p-value $0,16$) (fig. 4.7b). The decoding performance of voxels encoded by a specific layer was also considered. Except for the last layer, the case of valence the movie-clip voxels have lower decoding correlation than the scrambled clip voxels, consistent with the results using decoding over all voxels (fig. 4.7a). In case of decoding arousal, there is no explicit difference between both datasets except that here too, voxels associated with the movie clip outperform the scrambled-clip voxels at the last layer (fig. 4.7c). The p-value for all layers, for both valence and arousal was above $0,2$ with values as large as $0,41$.

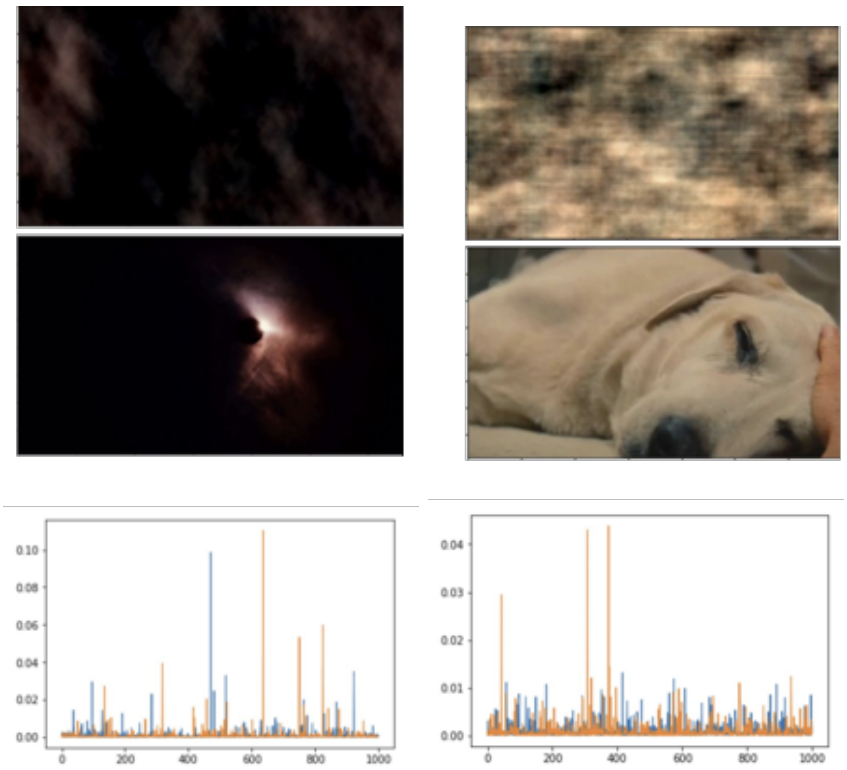
Chapter 5

Discussion and Limitations

5.1 Encoding

Surprisingly, significant voxels are distributed across the visual cortex and not mainly located in visual areas, ventral and dorsal stream as it has been shown by previous studies [8, 31]. One reason for this is that the brain mask used in this study encompasses all brain lobes where as [8] had a brainmask consisting only of the occipital lobe and [31] had a brainmask mainly related to brain regions correlated to visual processing. Nevertheless, our findings correspond with what has been found by [2] who showed that low- high- and semantic-level visual features encode voxel across the brain. They did not find mid-level features encoded in the later areas for their fMRI task. It should be noted that the stimuli used in our study are much more complex then previously used stimuli, where more than one person can appear, characters may interacting and the story line has emotion eliciting content. Thus, the related brain activity can be expected to be of a much more complicated nature.

Results for the significant voxels related to the scrambled response are not as expected. Intuitively, one would expect significant brain activity located in the occipital cortex, related to the processing of lower visual layers. Encoding using the feature layers too, therefore should be mainly located in the occipital cortex and dominated by the earlier layers. (fig. 4.5a) shows though, that significant voxels are spread all over the brain mask and encoded by all feature layers. When observing the individual subject distribution (fig. 4.4b) one can find some clustering in the parietal lobe. First of all, one of the reasons for features from later layers still encoding voxel significance has to do with the CNN output for scrambled images. As the CNN was not fine-tuned on this dataset, it does have high activations in later feature layers and even outputs classification predictions with higher confidence, sometimes higher than the classification prediction for the related movie frame. This explains why the later layers are able to be assigned to voxels as the most predictive layers as they have significant activations. It should be noted that in general the pearson r correlation between the predictions and the true brain activity are in general lower for the models encoding the brain activity using the scrambled movie clips. Additionally, the brain mask used for the brain response related to the scrambled stimuli, is exactly the same as the one used for the movie clip related brain acitivity. This means that the brain mask is likley to contain insignificant voxel activity. Furhtermore during encoding voxels were discarded as non-significant on the basis of a thersholdset on the p-value. However, due to the small size of the test set (during testing and cross-validation training) the p-value is not very reliable.



(a) (b)

Figure 5.1: Distribution for class predictions of the CNN for movie (blue) and scrambled (orange) clips.

5.2 Decoding

The CNN feauture layers all have higher decoding accruacy than voxels encoded by these features (figs. 4.6, 4.7a and 4.7b). Considering to what extent the feature layers of a CNN trained on a different task can decode arousal and to some extent valence labels, it is reasonable that a CNN finetuned on emotion categories, as it was done in [15], is able to do so with high accuracy. [15] argued that since the occipital lobe can decode the last layer which is related to the emotion classes, emotional

content should be contained in visual processing mechanisms of the brain. However, decoding performance only gives insight into the fact that information is present within a certain area but not *how* it is represented. A last layer of a CNN results from the previous layers, and all these layers have been shown to encode the visual system. Therefore it seems natural that the occipital area, an area known to be encoded well by preceding CNN layers,is able to decode the last layer of a CNN. Additionally the layer-wise decoding performacne from the brain activity follows a different pattern than the decoding performance related to the feautre layers. Although, CNN layers are known to be an accurate computational model for some parts of the brain, the different patterns in decoding accruacy suggest that differences between the features and the encoded brain acitivty remain. This difference is further highlighted by the fact that a CNN can be trained to predict emotional classes and human valence and arousal levels, whereas voxels encoding CNN feature layers cannot.

Furthermore, our results strongly indicate that visual processing related brain activity does not contain emotional information. The fact that voxels encoding visual processing have the lowest decoding performance, suggests that especially in these areas emotional information is not present (fig. 4.7a). It should be noted that the randomly selected voxels might be in brain regions that encode emotional information, hence the decoding accruacy can be higher and also varies depending on which voxels were selected. Keypoint is, that these voxels do not encode low- mid- or high level visual features. Additionally, when decoding arousal, voxels related to the scrambled clip as well as the movie clip have lowest decoding accruacy (fig. 4.7b). Again this implies that voxels endcoding low- mid- or high level features do not contain emotional information.

The voxels that were found to encode visual features using our encoding model,however, were not located only in the occipital lobe but spread across the brain. Hence, randomly selected voxel for instance can also be located in the occipital lobe. Our results therefore, do not contradict the fact the regions in the occipital lobe are able to decode valence and arousal labels [15, 20], but suggest that this decoding performance is not due to visual information processing. This implies that the functionality of the occipital lobe is not ristricted to visual processing.

5.3 Limitations and Feature Directions

As mentioned previously, our dataset is unique in the way that the stimuli used during the fMRI recording are highly naturalistic. To this end, the most naturalistic stimuli used we movie clips of animals or humans moving [31]. Our dataset is more complex in the sense that more than one person occurs in the clip, mutliple actions are performed with characters interacting. Furthermore these clips are supposed to elicit an emotional response in the viewer and thus have a more complicated context. Additionally, as the clips were taken from the Hollywood 2 dataset, it is not unlikely that subjects have seen the movie before, therefore also activating other brain processes such as memory. Thus, the brain activity elicited by our movie is very complex and diverse.

When selecting the brain mask of significant voxels, a task is often repeated and voxels that have the most stable activity are selected as significant. Since the purpose of this experiment was to elicit an emotional response this was not possible, as a repetion of the stimulus might surpress or alter emotional reaction in the brain. Furthermore, the brain mask for the two different brain responses, movie and scrambled, were taken to

be the same. In other words no seperate brain mask for the scrambled dataset was computed. This means, that for the brain response elicited by the scrambled movie clips, it is likely that the brain mask contains voxels not related to the stimuli. Additionally, the small dataset size for the stimuli means that the p-value is not very reliable as a statistical test and thus voxels might have been falsly discarded or considered.

Future steps for this work are to fine-tune the CNN on our dataset, such that scrambled frames do not output image classes with high probability and therby allowing to discriminate beetween semantic relevant and non relevant content more precisely. Through this we expect to get a stronger correlation between CNN features and brain activity. Furthermore to verify the correctness and reliability of the encoding model, encoded brain activity should be tested on its ability to decode class labels for each frame.

Chapter 6

Conclusion

By using encoding and decoding models in conjunction, we showed that visual features exctracted from a CNN are not only encoded in brain activity in the occipital lobe but extend to the dorsal- and ventral stream, temporal lobes and all the way to the orbito-frontal cortex. We excluded the possibility of emotional content being encoded within regions encoding low- mid- and high level visual features. The difference in decoding performance between brain activity encoded by CNN features and CNN features themselves, suggest that there are still differences in the underlying computational code. Furthermore, our results support the hypothesis that decoding performance from the occipital cortex could be related to antecedent emotional or top down regularatory processes. Our study is unique in the way that no stimuli of such complicated nature as been used before and we computed encoding models for voxels spanning the whole brain.

Bibliography

[1] A. Anzai, X. Peng, and D. C. Van Essen. Neurons in monkey visual area v2 encode combinations of orientations. *Nature Neuroscience*, 10:1313 EP –, 09 2007.

[2] M. B. Bone, F. Ahmad, and B. R. Buchsbaum. Feature-specific neural reactivation during episodic memory. *bioRxiv*, page 622837, 01 2019.

[3] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman. Return of the devil in the details: Delving deep into convolutional nets. *arXiv preprint arXiv:1405.3531*, 2014.

[4] J. D. Cohen, N. Daw, B. Engelhardt, U. Hasson, K. Li, Y. Niv, K. A. Norman, J. Pillow, P. J. Ramadge, N. B. Turk-Browne, and T. L. Willke. Computational approaches to fmri analysis. *Nature neuroscience*, 20(3):304–313, 02 2017.

[5] I. Daum, H. Markowitsch, and M. Vandekerckhove. *Neurobiological Basis of Emotions*, pages 111–138. 01 2009.

[6] J. Diedrichsen and N. Kriegeskorte. Representational models: A common framework for understanding encoding, pattern-component, and

representational-similarity analysis. *PLOS Computational Biology*, 13(4):e1005508–, 04 2017.

[7] J. C. Gore. Principles and practice of functional mri of the human brain. *The Journal of clinical investigation*, 112(1):4–9, 07 2003.

[8] U. Güçlü and M. A. J. van Gerven. Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. *The Journal of Neuroscience*, 35(27):10005, 07 2015.

[9] U. Güçlü and M. A. J. van Gerven. Modeling the dynamics of human brain activity with recurrent neural networks. *Frontiers in computational neuroscience*, 11:7; 7–7, 02 2017.

[10] H. Gunes, B. Schuller, M. Pantic, and R. Cowie. Emotion representation, analysis and synthesis in continuous space: A survey. In *Face and Gesture 2011*, pages 827–834, 2011.

[11] J.-D. Haynes. A primer on pattern-based approaches to fmri: Principles, pitfalls, and perspectives. *Neuron*, 87(2):257–270, 2019/04/29 2015.

[12] M. J. Herrmann, T. Huter, M. M. Plichta, A.-C. Ehlis, G. W. Alpers, A. Mühlberger, and A. J. Fallgatter. Enhancement of activity of the primary visual cortex during processing of emotional stimuli as measured with event-related functional near-infrared spectroscopy and event-related potentials. *Human Brain Mapping*, 29(1):28–35, 2019/05/01 2008.

[13] J. P. Jones and L. A. Palmer. An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, 58(6):1233–1258, 2019/05/17 1987.

[14] K. N. Kay, T. Naselaris, R. J. Prenger, and J. L. Gallant. Identifying natural images from human brain activity. *Nature*, 452:352 EP –, 03 2008.

[15] P. A. Kragel, M. Reddan, K. S. LaBar, and T. D. Wager. Emotion schemas are embedded in the human visual system. *bioRxiv*, page 470237, 01 2018.

[16] D. J. Kravitz, K. S. Saleem, C. I. Baker, and M. Mishkin. A new neural framework for visuospatial processing. *Nature reviews. Neuroscience*, 12(4):217–230, 04 2011.

[17] D. J. Kravitz, K. S. Saleem, C. I. Baker, L. G. Ungerleider, and M. Mishkin. The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends in cognitive sciences*, 17(1):26–49, 01 2013.

[18] N. Kriegeskorte and P. Douglas. *Interpreting encoding and decoding models*, volume 55. 04 2019.

[19] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[20] J. Lee and S. Lee. Decoding valence and arousal. in preperation.

[21] M. A. Lindquist. The statistical analysis of fmri data. *Statist. Sci.*, 23(4):439–464, 2008.

[22] M. Marszałek, I. Laptev, and C. Schmid. Actions in context. In *IEEE Conference on Computer Vision & Pattern Recognition*, 2009.

[23] T. Naselaris, K. N. Kay, S. Nishimoto, and J. L. Gallant. Encoding and decoding in fmri. *NeuroImage*, 56(2):400–410, 05 2011.

[24] J. J. Nassi and E. M. Callaway. Parallel processing strategies of the primate visual system. *Nature reviews. Neuroscience*, 10(5):360–372, 05 2009.

[25] L. Pessoa and R. Adolphs. Emotion processing and the amygdala: from a 'low road' to 'many roads' of evaluating biological significance. *Nature reviews. Neuroscience*, 11(11):773–783, 11 2010.

[26] H. Saarimäki, L. F. Ejtehadian, E. Glerean, I. P. Jääskeläinen, P. Vuilleumier, M. Sams, and L. Nummenmaa. Distributed affective space represents multiple emotion categories across the human brain. *Social cognitive and affective neuroscience*, 13(5):471–482, 05 2018.

[27] S. Schoenmakers, M. Barth, T. Heskes, and M. van Gerven. Linear reconstruction of perceived images from human brain activity. *NeuroImage*, 83:951 – 961, 2013.

[28] D. Seibert, D. Yamins, D. Ardila, H. Hong, J. J. DiCarlo, and J. L. Gardner. A performance-optimized model of neural responses across the ventral visual stream. *bioRxiv*, page 036475, 01 2016.

[29] T. Serre. *Hierarchical Models of the Visual System*, pages 1–12. Springer New York, New York, NY, 2013.

[30] M. A. J. van Gerven. A primer on encoding models in sensory neuroscience. *Journal of Mathematical Psychology*, 76:172–183, 2017.

[31] H. Wen, J. Shi, Y. Zhang, K.-H. Lu, J. Cao, and Z. Liu. Neural encoding and decoding with deep learning for dynamic natural vision. *Cerebral Cortex*, 28(12):4136–4160, 4/29/2019 2017.

요약

인코딩 모델은 자극으로부터 촉발된 뇌 활동을 예측하고, 뇌가 정보를 어떻게 처리하는지 분석하기 위해 사용된다.반면 디코딩 모델은 뇌 활동으로부터 자극에 대한 정보를 예측하고, 현재 특정 자극이 존재하는지를 판단하는 것을 목표로 한다. 두 모델은 종종 함께 사용된다. 뇌의 시각 체계는 자극에 대한 감정 정보를 담고 있고 [15, 20], 픽셀들이 무작위로 섞여 있는 자극으로부터 유도된 시각 체계의 활동으로부터도 같은 감정 정보를 추출해낼 수 있다는 것이 알려져 있다 [20]. 이런 연구들을 고려하여, 우리는 시각 체계가 어느 수준까지 감정 정보를 담고 있는지 탐구한다. 우리는 인코딩 모델을 사용하여 상위/중위/하위 시각 특성(feature)과 각각 관련이 있는 뇌 영역을 선택하고, 이 뇌 영역들로부터 감정 정보를 디코딩 한다. 우리는 후두엽뿐만 아니라 안와전두피질까지 이어지는 영역들이 이런 특성들을 인코딩 하고 있다는 것을 밝힌다. 다른 뇌 영역들과 단순한 CNN 특성들과는 달리, 이러한 뇌 영역들로부터는 감정 정보를 디코딩 할 수 없었다. 이 결과들은 상위/중위/하위 시각 특성들을 인코딩 하고 있는 뇌 영역들이 앞서 밝혀진 감정 정보 디코딩과 관련이 없음을 보여주며, 따라서 후두엽과 관련된 감정 정보 디코딩 성능은 시각과 관련 없는 정보 처리에 기인한다.

주요어: 인코딩, 디코딩,fMRI, 시각, 감정

학번: 2017-28897

Acknowledgements

First of all, I would like to thank my advisor, professor Gunhee Kim, for his guidance and support throughout the 2 years of my Msc studies.
I would also like to thank Joonwon Lee, Jaesob Lim, Juhyoung Rhy for the helpful discussions, inspiration, help and support especially regarding the understanding of fMRI data.
I would also like to thank Youngjae Yu, Sangho Lee and Soochan Lee who gave me helpful advise, guided me in the right directions and helped me fill knowledge gaps.
Additionally I would like to than Minjoeng Kim and Joonil Na who helped me with administrative processes and settle into the laboratory and graduate student life.
I would also like to thank my labmates at SNU Vision and Learning Laboratory who welcomed me warmly, helped me to adapt, were always patient with my language skill and helped me out in many ways: Junhyug Noh, Youngjae Yu, Minui Hong, Insu Jeon, Wonhee Lee, Byeongchang Kim, Hyouk Jang, Yookoon Park, Youngjin Kim, Sangho Lee, Soochan Lee, Taeyoung Hahn, Jaemin Cho, Minjung Kim, Joonil Na, Jongseok Kim, Dongjoo Kim, Hyunwoo Kim, Jiwan Chang,Myoengjang Pyeon, Dongsu Zhan, Junsoo Ha, Jaewoo Ahn etc. who are truly inspiring.

I would like to thank Jaeyoung Park who helped me extensively through all the administrative processes which I would have not been able to handle without her help.

Finally I would like to thank everyone else I met during my graduate journey and want to express my gratitude for the understanding, help and support I have experienced throughout.