Frequency-Resolved Correlates of Visual Object Recognition in Human Brain Revealed by Deep Convolutional Neural Networks

Ilya Kuzovkin[1,*], Raul Vicente[1,†], Mathilde Petton[2,3], Jean-Philippe Lachaux[2,3], Monica Baciu[4,5], Philippe Kahane[6,7], Sylvain Rheims[8,2,9], Juan R. Vidal[4,5] and Jaan Aru[1,10,*,†]

[1] *Computational Neuroscience Lab, Institute of Computer Science, University of Tartu, Estonia*
[2] *INSERM U1028, CNRS UMR5292, Brain Dynamics and Cognition Team, Lyon Neuroscience Research Center, Lyon, France*
[3] *Université Claude Bernard, Lyon, France*
[4] *University Grenoble Alpes, LPNC, F-38040 Grenoble, France*
[5] *CNRS, LPNC UMR 5105, F38040 Grenoble, France*
[6] *Inserm, U1216, F-38000 Grenoble, France*
[7] *Neurology Department, CHU de Grenoble, Hôpital Michallon, F-38000 Grenoble, France*
[8] *Department of Functional Neurology and Epileptology, Hospices Civils de Lyon and Université Lyon, Lyon, France*
[9] *Epilepsy Institute, Lyon, France*
[10] *Department of Penal Law, School of Law, University of Tartu, Estonia*

[*] Corresponding authors. E-mail: ilya.kuzovkin@gmail.com; jaan.aru@gmail.com
[†] These authors contributed equally to this work.

Previous work demonstrated a direct correspondence between the hierarchy of the human visual areas and layers of deep convolutional neural networks (DCNN) trained on visual object recognition. We used DCNNs to investigate which frequency bands carry feature transformations of increasing complexity along the ventral visual pathway. By capitalizing on direct intracranial recordings from 81 patients and 9147 electrodes we assessed the alignment between the DCNN and signals at different frequency bands in different time windows. We found that activity in low and high gamma bands was aligned with the increasing complexity of visual feature representations in the DCNN. These findings show that activity in the gamma band is not only a correlate of object recognition, but carries increasingly complex features along the ventral visual pathway. Similar alignment was found in the alpha frequency highlighting an unexpected role for alpha in contributing to visual object recognition. These results demonstrate the potential that modern artificial intelligence algorithms have in advancing our understanding of the brain.

Significance Statement

Recent advances in the field of artificial intelligence have revealed principles about neural processing, in particular about vision. Previous works have demonstrated a direct correspondence between the hierarchy of human visual areas and layers of deep convolutional neural networks (DCNNs), suggesting that DCNN is a good model of visual object recognition in primate brain. Studying intracranial recordings allowed us to extend previous works by assessing when and at which frequency bands the activity of the visual system corresponds to the DCNN. Our key finding is that signals in gamma and alpha frequencies along the ventral visual pathway are aligned with the layers of DCNN. These frequencies play a major role in transforming visual input to coherent objects.

Introduction

Visual object recognition is mediated by a hierarchy of increasingly complex feature representations along the ventral visual stream (DiCarlo et al., 2012). Intriguingly, these transformations are quite similar to the hierarchy of transformations learned by deep convolutional neural networks (DCNN) trained on natural images. These developments make it possible to assess whether putative correlates of visual object recognition also reflect such gradual transformations. For instance, there is a long-standing hypothesis that gamma band signals mediate feature binding and object recognition (Singer and Gray, 1995; Singer, 1999). Classic studies have demonstrated that visual processing of simple patterns (gratings) is reflected in strong gamma oscillations in primary visual cortex (Gray and Singer, 1989). Although these early studies were only done with simple stimuli and in early visual cortex, subsequent studies have consistently shown that

gamma power increase is related to the recognition of complex natural objects (Lachaux et al., 2005; Fisch et al., 2009; Vidal et al., 2010). However, the exact relation between gamma activity and object recognition is less clear. The classic view is that gamma band activity signals the emergence of coherent object representations (Singer and Gray, 1995; Singer, 1999; Fisch et al., 2009). On the other hand, it is possible that gamma frequencies carry feature transformations of increasing complexity instead of reflecting solely the final product of object recognition. The existence on quantifiable increase of feature complexity along layers of DCNN allows one to use DCNN as a computational model to investigate whether signals in the gamma band carry such increasingly complex features along the ventral visual pathway.

It has been shown that DCNN provides the best model out of a wide range of neuroscientific and computer vision models for the neural representation of visual images in high-level visual cortex of monkeys (Yamins et al., 2014) and humans (Khaligh-Razavi and Kriegeskorte, 2014). Other studies have demonstrated with fMRI a direct correspondence between the hierarchy of the human visual areas and layers of the DCNN (Güçlü and van Gerven, 2015; Eickenberg et al., 2016; Seibert et al., 2016; Cichy et al., 2016b). Taken together these results support the view that the increasing feature complexity of the DCNN corresponds to the increasing feature complexity occurring in visual object recognition in the primate brain (Kriegeskorte, 2015; Yamins and DiCarlo, 2016).

In the present work we assessed whether there is an alignment between the responses of layers of the DCNN and the signals in five distinct frequency bands along the areas constituting the ventral visual pathway. To empirically evaluate the role of gamma and other frequency bands in visual object recognition we capitalized on direct intracranial recordings from 81 patients with epilepsy and a total of 9147 electrodes implanted throughout the cerebral cortex. We observed that activity in the gamma range along the ventral pathway is statistically significantly aligned with the activity along the layers of DCNN: gamma ($31-150$ Hz) activity in the early visual areas correlates with the activity of early layers of DCNN, while the gamma activity of higher visual areas is better captured by the higher layers of the DCNN. Surprisingly, we also observed such alignment in the alpha band ($9-14$ Hz). We also found that neural activity in the theta range ($5-8$ Hz) throughout the visual hierarchy correlated with higher layers of DCNN.

Materials and Methods

Our methodology involves four major steps described in the following subsections. In "Patients and Recordings" we describe the visual recognition task and the data collection. In "Processing of Neural Data" we describe the artifact rejection, extraction of spectral features and the electrodes selection processes. "Processing of DCNN Data" shows how we extract DCNN layers' responses to the same images as used in the visual recognition task. In the last step we map neural activity to the layers of DCNN using representational similarity analysis. See Figure 1 for the illustration of the analysis workflow.

Patients and Recordings

81 patients of either gender with drug-resistant partial epilepsy and candidates for surgery were considered in this study and recruited from Neurological Hospitals in Grenoble and Lyon (France). All patients were stereotactically implanted with multi-lead EEG depth electrodes (DIXI Medical, Besançon, France). All participants provided written informed consent, and the experimental procedures were approved by local ethical committee of Grenoble hospital (CPP Sud-Est V 09-CHU-12). Recording sites were selected solely according to clinical indications, with no reference to the current experiment. All patients had normal or corrected to normal vision.

Electrode Implantation

Eleven to 15 semi-rigid electrodes were implanted per patient. Each electrode had a diameter of 0.8 mm and was comprised of 10 or 15 contacts of 2 mm length, depending on the target region, 1.5 mm apart. The coordinates of each electrode contact with their stereotactic scheme were used to anatomically localize the contacts using the proportional atlas of Talairach and Tournoux (Talairach and Tournoux, 1993), after a linear scale adjustment to correct size differences between the patients brain and the Talairach model. These locations were further confirmed by overlaying a post-implantation CT scan (showing contact sites) with a pre-implantation structural MRI with VOXIM® (IVS Solutions, Chemnitz, Germany), allowing direct visualization of contact sites relative to brain anatomy.

All patients voluntarily participated in a series of short experiments to identify local functional responses at the recorded sites (Vidal et al., 2010). The results presented here were obtained from a test exploring visual recognition. All data were recorded using approximately 120 implanted depth electrode contacts per patient with a sampling rate of 512 Hz. Data were obtained in a total of 9147 recording sites.

Stimuli and Task

The visual recognition task lasted for about 15 minutes. Patients were instructed to press a button each time a picture of a fruit appeared on screen (visual oddball paradigm). Non-target stimuli consisted of pictures of objects of eight possible categories: houses, faces, animals, scenes, tools, pseudo words, consonant strings, and scrambled images. The last three categories were not included in this analysis. All the included stimuli had the same average luminance. All categories were presented within an oval aperture (illustrated on Figure 1). Stimuli were presented for a duration of 200 ms every $1000-1200$ ms in series of 5 pictures interleaved by 3-s pause periods during which patients could freely blink. Patients reported the detection of a target through a right-hand button press and were given feedback of their performance after each report. A 2-s delay was placed after each button press before presenting the follow-up stimulus in order to avoid mixing signals related to motor action with signals from stimulus presentation. Altogether, we measured responses to 269 natural images.

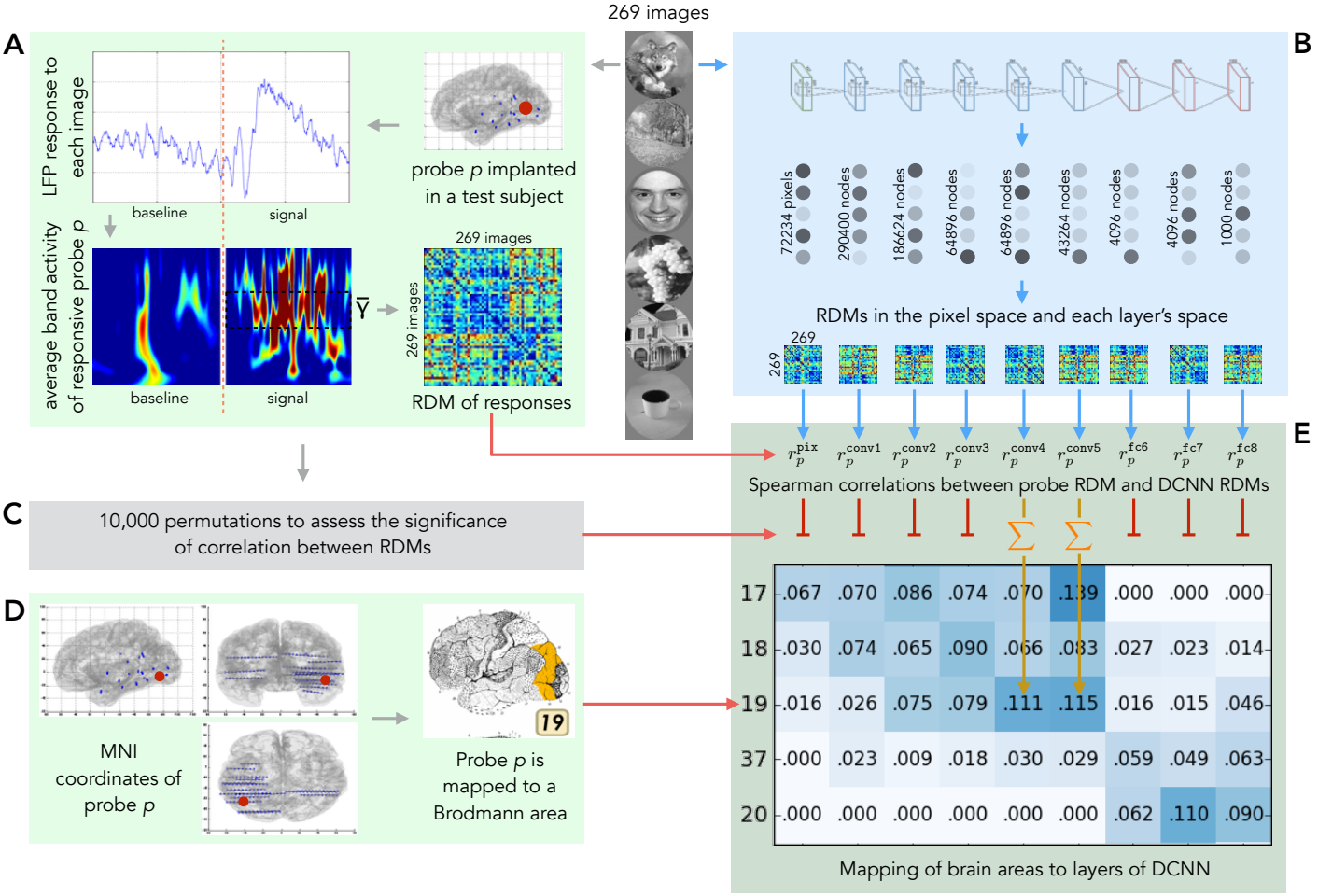

Figure 1 Overview of the analysis pipeline. 269 natural images are presented to human subjects (panel A) and to an artificial vision system (panel B). The activities elicited in these two systems are compared in order to map regions of human visual cortex to layers of deep convolutional neural networks (DCNNs). **A:** LFP response of each of 9147 electrodes to each of the images is converted into the frequency domain. Activity evoked by each image in the frequency band and time of interest is compared to the activity evoked by every other image and results of this comparison are presented as a representational dissimilarity matrix (RDM). **B:** Each of the images is shown to a pre-trained DCNN and activations of each of the layers are extracted. Each layer's activations form a representation space, in which stimuli (images) can be compared to each other. Results of this comparison are summarized as a RDM for each DCNN layer. **C:** Subject's intracranial responses to stimuli are randomly reshuffled and the analysis depicted in panel A is repeated 10000 times to obtain 10000 random RDMs for each electrode. **D:** Each electrode's MNI coordinates are used to map the electrode to a Brodmann area. The figure also gives an example of electrode implantation locations in one of the subjects (blue circles are the electrodes). **E:** Spearman's rank correlation is computed between the true (non-permuted) RDM of neural responses and RDMs of each layer of DCNN. Also 10000 scores are computed with the random RDM for each electrode-layer pair to assess the significance of the true correlation score. If the score obtained with the true RDM is significant ($p < 0.001$), then the score is added to the mapping matrix. The procedure is repeated for each electrode and the correlation scores are summed and normalized by the number of electrodes in a Brodmann area. The resulting mapping matrix shows the alignment between the consecutive areas of the ventral stream and layers of DCNN.

Processing of Neural Data

The final dataset consists of 2460543 local field potential (LFP) recordings – 9147 electrode responses to 269 stimuli.

To remove the artifacts the signals were linearly detrended and the recordings that contained values $\geq 10\sigma_{images}$, where σ_{images} is the standard deviation of responses (in the time window from −500ms to 1000ms) of that particular probe over all stimuli, were excluded from data. All electrodes were re-referenced to a bipolar reference. The signal was segmented in the range from −500 ms to 1000 ms, where 0 marks the moment when the stimulus was shown. The −500 to −100 ms time window served as the baseline. There were three time windows in which the responses were measured: $50 - 250$ ms, $150 - 350$ ms and $250 - 450$ ms.

We analyzed five distinct frequency bands: θ $(5 - 8$ Hz), α $(9 - 14$ Hz), β $(15 - 30$ Hz), γ $(31 - 70$ Hz) and Γ $(71 - 150$ Hz). To quantify signal power modulations across time and frequency we used standard time-frequency (TF) wavelet decomposition

(Daubechies, 1990). The signal $s(t)$ is convoluted with a complex Morlet wavelet $w(t, f_0)$, which has Gaussian shape in time (σ_t) and frequency (σ_f) around a central frequency f_0 and defined by $\sigma_f = 1/2\pi\sigma_t$ and a normalization factor. In order to achieve good time and frequency resolution over all frequencies we slowly increased the number of wavelet cycles with frequency ($\frac{f_0}{\sigma_f}$ was set to 6 for high and low gamma, 5 for beta, 4 for alpha and 3 for theta). This method allows obtaining better frequency resolution than by applying a constant cycle length (Delorme and Makeig, 2004). The square norm of the convolution results in a time-varying representation of spectral power, given by: $P(t, f_0) = |w(t, f_0)s(t)|^2$.

Further analysis was done on the electrodes that were responsive to the visual task. We assessed neural responsiveness of an electrode separately for each region of interest – for each frequency band and time window we compared the average post-stimulus band power to the average baseline power with a Wilcoxon signed-rank test for matched-pairs. All p-values from this test were corrected for multiple comparisons across all electrodes with a false discovery rate (FDR) procedure (Genovese et al., 2002). In the current study we deliberately kept only positively responsive electrodes, leaving the electrodes where the post-stimulus band power was significantly weaker than the average baseline power for future work. Table 1 contains the numbers of electrodes that were used in the final analysis in each of 15 regions of interest across the time and frequency domains.

	θ	α	β	γ	Γ
$50-250$ ms	1299	709	269	348	504
$150-350$ ms	1689	783	260	515	745
$250-450$ ms	1687	802	304	555	775

Table 1 Number of positively responsive electrodes in each of the 15 regions of interest in a time-resolved spectrogram.

Each electrode's MNI coordinates were mapped to a corresponding Brodmann brain area (Brodmann, 1909) using Brodmann area atlas contained in MRICron (Rorden, 2007) software.

To summarize, once the neural signal processing pipeline is complete, each electrode's response to each of the stimuli is represented by one number – the average band power in a given time window normalized by the baseline. The process is repeated independently for each region of interest.

Processing of DCNN Data

We feed the same images that were shown to the test subjects to a deep convolutional neural network (DCNN). We use `Caffe` (Jia et al., 2014) implementation of `AlexNet` (Krizhevsky et al., 2012) architecture (see panel B of Figure 1) trained on `ImageNet` (Russakovsky et al., 2015) dataset to categorize images into 1000 classes. Although the image categories used in our experiment are not exactly the same as the ones in the `ImageNet` dataset, they are a close match and DCNN is successful in labelling them.

For each of the images we stored the activations of all nodes of DCNN. As the network has 8 layers of nodes we obtained 9 representations of an image: the image itself (referred to as layer 0) in the pixel space and the activation values of each of the layers of DCNN. See table 1 for the full list of feature spaces that are used in this work to represent an image.

Space	Cardinality
Layer 0: pixels	72234
Layer 1: convolutional	290400
Layer 2: convolutional	186624
Layer 3: convolutional	64896
Layer 4: convolutional	64896
Layer 5: convolutional	43264
Layer 6: fully connected	4096
Layer 7: fully connected	4096
Layer 8: fully connected	1000
Neural activity of a probe	1

Table 2 Ten possible representations of an image and their dimensionalities.

Mapping Neural Activity to Layers of DCNN

Once we extracted features from both neural and DCNN responses our next goal was to compare the two and use a similarity score to map the brain area where a probe was located to a layer of DCNN. By doing that for every probe in the dataset we obtained cross-subject alignment between visual areas and layers of DCNN.

Recent studies comparing the responses of visual cortex with the activity of DCNN have used two types of mapping methods. The first type is based on linear regression models that predict neural responses from DCNN activations (Güçlü and van Gerven, 2015). The second type is based on representational similarity analysis (RSA) (Kriegeskorte et al., 2008). RSA is used to compare distances between stimuli in the neural response space and in the DCNN activation space (Cichy et al., 2016a). We employed RSA, but qualitatively similar results were obtained by us with the regularized linear regression method.

Representational Dissimilarity Matrices

First we built a representation dissimilarity matrix (RDM) of size *number of stimuli* $\times$ *number of stimuli* (in our case 269×269) for each of the features spaces (see Table 2). Given a matrix $\mathrm{RDM}^{\text{feature space}}$ a value $\mathrm{RDM}_{ij}^{\text{feature space}}$ in the ith row and jth column of the matrix shows the euclidean distance between the vectors $\mathbf{v}_i$ and $\mathbf{v}_j$ that represent images i and j respectively in that particular feature space. In our case there are 10 different features spaces (listed in Table 2) in which a stimulus (an image) can be represented: the original pixel space, 8 feature spaces for each of the layers of the DCNN and one space where an image is represented by the preprocessed neural response of probe p. To analyse one region of interest (for example high gamma in $50-250$ ms time window) we computed 9147 RDM matrices on the neural responses – one for each electrode, and 9 RDM matrices on the activations of the layers of DCNN.

Representational Similarity Analysis

The second step was to compare the $\mathrm{RDM}^{\mathrm{probe}\ p}$ of each probe p with RDMs of layers of DCNN. The similarity measure we used was Spearman's rank correlation between the matrices:

$$\mathrm{score}^{\mathrm{probe}\ p}_{\mathrm{layer}\ l} = \mathrm{Spearman}(\mathrm{RDM}^{\mathrm{probe}\ p}, \mathrm{RDM}^{\mathrm{layer}\ l}).$$

As a result of comparing $\mathrm{RDM}^{\mathrm{probe}\ p}$ with every $\mathrm{RDM}^{\mathrm{layer}\ l}$ we obtain 9 scores: $\mathrm{score}_{\mathrm{pixels}} \ldots \mathrm{score}_{\mathtt{fc8}}$ that serve as a distributed mapping of probe p to the layers of DCNN (see panel E of Figure 1). The procedure is repeated independently for each of the 9147 probes.

Statistical significance

To assess the statistical significance of the correlations between the RDM matrices we run a permutation test. In particular, we reshuffled the vector of brain responses to images 10000 times, each time obtaining a dataset where the causal relation between the stimulus and the response is destroyed. On each of those datasets we ran the analysis and obtained the Spearman's rank correlation scores. To determine a p-value we compared the score obtained on the original (unshuffled) data with the distribution of scores obtained with the surrogate data. If the score obtained on the original data was bigger than the respective maximal value obtained on the surrogate sets we considered the score to be significantly different ($p \leq 0.0001$).

Quantifying properties of the mapping

To evaluate the results quantitatively we devised a set of measures. *Volume* is the total sum of significant correlations between the probes in a particular brain area and layers of DCNN. *Volume of visual activity in layers* L is defined as

$$V_L = \sum_{l \in L} \sum_{p \in S_l} r^p_l,$$

where S_l is the set of all probes from in visual areas that significantly correlate with layer l. *Visual specificity* is the ratio between *volume* in visual areas and total *volume*. The ratio of *complex* visual features to all visual features is defined as the total volume mapped to layers `conv5`, `fc6`, `fc7` divided by the total volume mapped to layers `conv1`, `conv2`, `conv3`, `conv5`, `fc6`, `fc7`. Note that for this measure layers `conv4` and `fc8` are omitted: layer `conv4` is considered to be the transition between the layers with low and high complexity features, while layer `fc8` directly represents class probabilities. *Alignment* between the activity in the visual areas and activity in DCNN is estimated as Spearman's rank correlation between the vector of electrode assignments to visual areas and the vector of electrode assignments to DCNN layers. As both the ventral stream and the hierarchy of layers in DCNN have an increasing complexity of visual representations, the relative ranking within the biological system should coincide with the ranking within the artificial system.

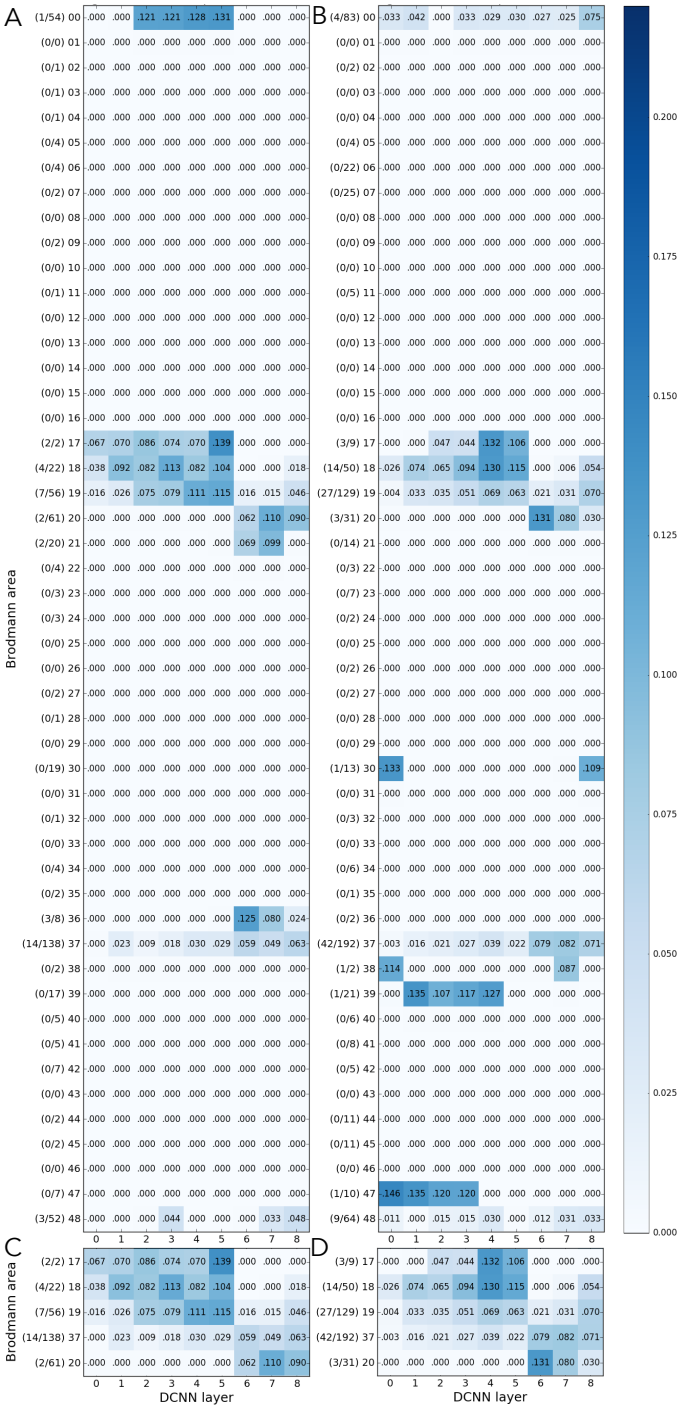

Figure 2 Mapping of the activity in Brodmann areas to DCNN layers. Underlying data comes from the activity in low gamma (31-70 Hz, subfigures A and C) and high gamma (71-150 Hz, subfigures B and D) bands in 150-350 ms time window. C and D are subselection of the areas that constitute ventral stream: 17, 18, 19, 37, 20. There are two important observations to made out of this plot: a) statistically significant neural responses are specific to visual areas b) the alignment between the ventral stream and layer of DCNN is clearly visible. Area 0 contains the regions of the brain not mapped by the atlas. The numbers on the left of each panel show the number of significantly correlating probes in each area out of the total number of responsive probes in that area, and the Brodmann area number.

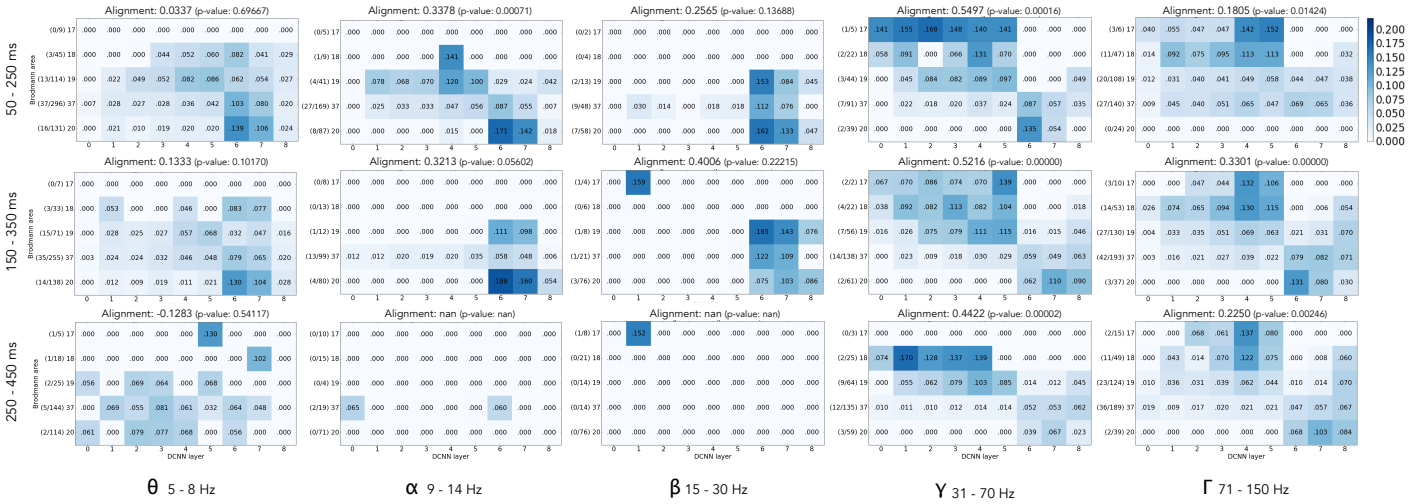

Figure 3 Mapping of activity in visual areas to activations of layers of DCNN across five frequency bands and three time windows. The alignment score is computed as Spearman correlation between electrode assignment to areas and electrode assignemt to DCNN layers. The numbers on the left of each subplot show the number of significantly correlating probes in each area out of the total number of responsive probes in that area, and the Brodmann area number.

Results

Increasing complexity of visual representations is captured by gamma and early alpha band activity

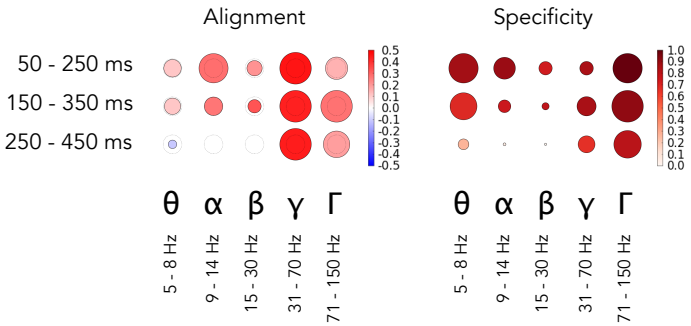

Figure 4 Overall relative statistics of brain responses across frequency bands and time windows. The left panel shows Spearman correlation between the electrode assignment to brain areas and the electrode assignment to DCNN layers: the color indicates the correlation strength, size of the marker shows the logarithm (so that not significant results are still visible on the plot) of inverse of the statistical significance of the correlation, dotted circle indicates $p = 0.05$ significance threshold. The right panel shows how specific to visual areas the activity is: intensive red means that most of the activity in that band and time window happened in visual areas, size of the marker indicates the total amount of activity (sum of correlations between the RDM matrices of the electrodes in that brain area and the RDM matrices of the DCNN layers). The maximal size of a marker is defined by the biggest marker on the figure.

We tested the hypothesis that gamma activity carries increasingly complex features along the ventral stream. To that end we assessed the alignment of neural activity in different frequency bands and time windows to the activity of different layers of a DCNN. In particular, we used RSA to compare the representational geometry of different DCNN layers and the activity patterns of different frequency bands of single electrodes (see Figure 1). We consistently found that signals in low gamma ($31 - 70$ Hz) and high gamma ($71 - 150$ Hz) frequencies aligned with the DCNN in a specific way: increase of the complexity of features along the layers of the DCNN was matched by the transformation in the representational geometry of responses to the stimuli along the ventral stream. In other words, the lower and higher layers of the DCNN explained gamma band signals from earlier and later visual areas, respectively.

Figure 2 illustrates assignment of neural activity in low gamma band (panel A) and high gamma band (panel B) to Brodmann areas and layers of DCNN. As one can see most of the activity was assigned to visual areas (areas 17, 18, 19, 37, 20). Focusing on these areas (panels C, D) revealed a diagonal trend that illustrated the alignment between ventral stream and layers of DCNN. Left panel of Figure 4 summarizes our findings across all subjects, time windows and frequency bands. The alignment in the gamma bands was present in all three time windows ($50 - 250$ ms, $150 - 350$ ms, $250 - 450$ ms). We note that the alignment in the gamma bands is also present at the single-subject level as can be seen in Figure 6.

Apart from the alignment we looked at the total amount of correlation and its specificity to visual areas. On the right panel of Figure 4 we can see that the volume of significantly correlating activity was highest in the high gamma range. Remarkably, in $50 - 250$ ms time window the 97% of that activity was located in visual areas. Also in Figure 2 we see that in the gamma range only a few electrodes were assigned to other Brodmann areas.

In addition to the gamma bands alignment to the DCNN was detected in the alpha ($9 - 14$ Hz) frequency range. This alignment was significant only in the earliest time window ($50 - 250$ms) as can be seen in Figure 4.

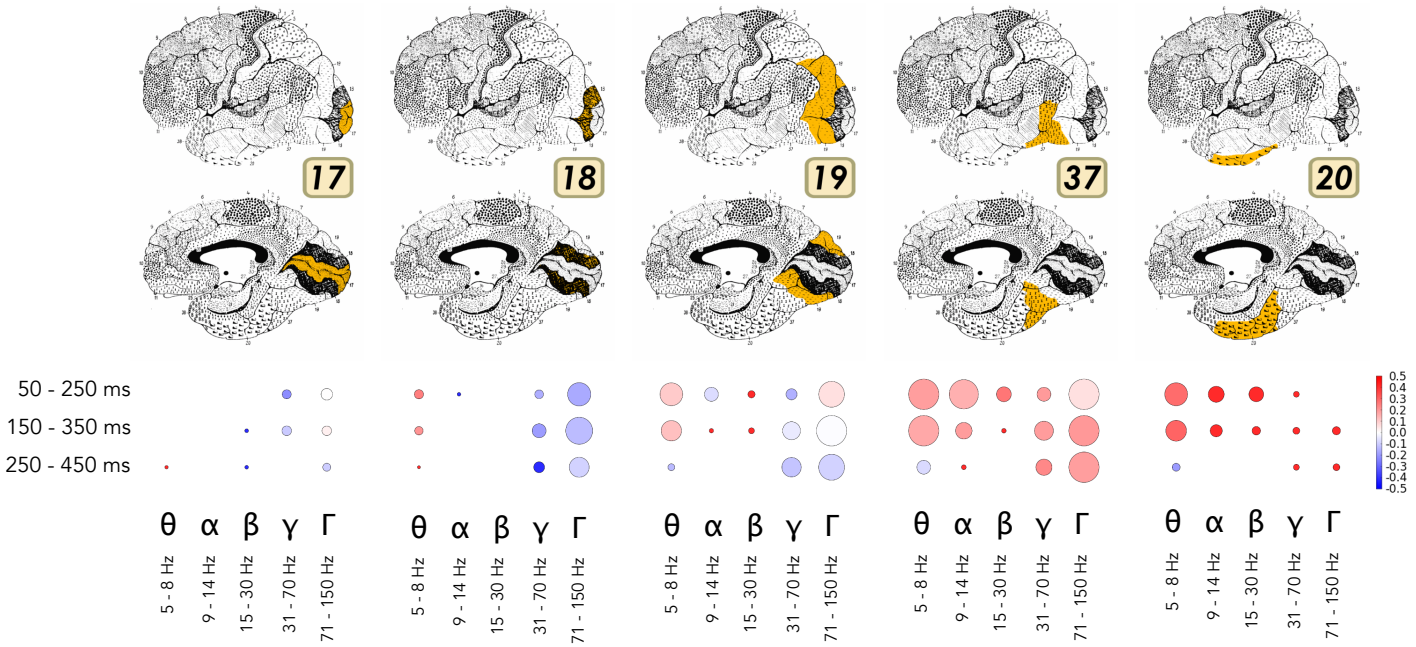

Figure 5 Area-specific analysis of volume of neural activity and complexity of visual features represented by that activity. Size of the marker shows the sum of correlation coefficients between the area and DCNN for each particular band and time window. Color codes the ratio of complex visual features to simple visual features, i.e. the comparison between the activity that correlates with the higher layers (`conv5, fc6, fc7`) of DCNN to the lower layers (`conv1`, `conv2`, `conv3`). Intensive red means that the activity was correlating more with the activity of higher layers of DCNN, while the intensive blue indicates the dominance of correlation with the lower areas. If the color is close to white then the activations of both lower and higher layers of DCNN were correlating with the brain responses in approximately equal proportion.

Activity in theta and beta bands is not aligned to the DCNN

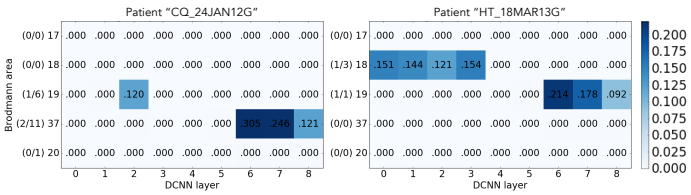

Figure 6 Single subject results from two different subjects. The numbers show the sum of correlations normalized by the number of probes in an area. On the left plot we see that the only probe (with significantly correlating activity) in Brodmann area 19 is mapped to the second convolutional layer of DCNN, while the activity in Brodmann area 37, which is located further along the ventral stream, is mapped to the higher layers of DCNN. The same trend is seen on the right plot. The numbers on the left of each subplot show the number of significantly correlating probes in each area out of the total number of responsive probes in that area, and the Brodmann area number.

The results for all frequency bands are presented in Figures 3 and 4. We can see that the alignment was present across all time windows in the gamma range and in early time window in the alpha range – the alignment was weaker and was not statistically significant in theta and beta frequency bands.

To investigate the involvement of each frequency band more closely we analyzed each visual area separately. Figure 5 shows the volume of activity in each area (size of the marker on the figure) and whether that activity was more correlated with the complex visual features (red color) or simple features (blue color). In our findings the role of the earliest area (17) was minimal, however that might be explained by a very low number of electrodes in that area in our dataset (less that 1%). One can see from Figure 5 that activity in theta frequency in time windows $50-250$ ms and $150-350$ ms had large volume and correlated with the higher layers of DCNN in higher visual areas (19, 37, 20) of the ventral stream. In general, in areas 37 and 20 all frequency bands carried information about high level features in the early time windows. This implies that already at early stages of processing the information about complex features was present in those areas.

Gamma activity is more specific to convolutional layers, while the activity in lower frequency bands is more specific to fully connected layers

We analysed volume and specificity of brain activity that correlates with each layer of DCNN separately to see if any bands or time windows are specific to particular level of hierarchy of visual processing in DCNN. Figure 7 presents a visual summary of this analysis. In the "Methods" section we have defined total volume of visual activity in layers L. We used this measure to quantify the activity in low and high gamma bands. We noticed that while

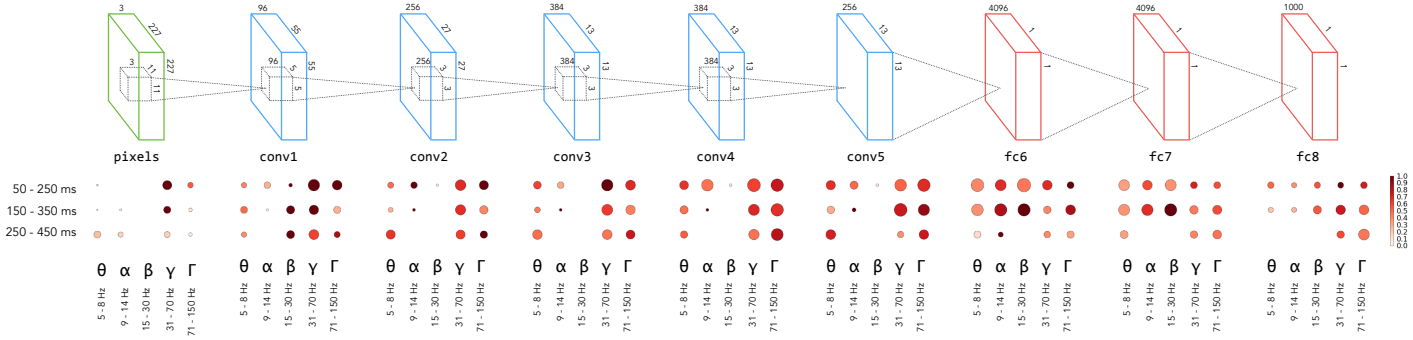

Figure 7 Specificity of neural responses across frequency bands and time windows for each layer of DCNN. Size of a marker is the total activity mapped to this layer and the intensity of the color is the specificity of the activity to visual areas.

the fraction of gamma activity that is mapped to convolutional layers is high ($\frac{\bar{V}^{\gamma,\Gamma}_{\{\text{conv1...conv5}\}}}{\bar{V}^{\text{all bands}}_{\{\text{conv1...conv5}\}}} = 0.70$), this fraction diminished in fully connected layers `fc6` and `fc7` ($\frac{\bar{V}^{\gamma,\Gamma}_{\{\text{fc6,fc7}\}}}{\bar{V}^{\text{all bands}}_{\{\text{fc6,fc7}\}}} = 0.37$). Note that `fc8` was excluded as it represents class label probabilities and does not carry information about visual features of the objects. On the other hand the activity in lower frequency bands (theta, alpha, beta) showed the opposite trend – fraction of volume in convolutional layers was 0.30, while in fully connected it growed to 0.63. This observation highlighted the fact that visual features extracted by convolutional filters of DCNN carry the signal that is more similar to the signal carried by gamma frequency bands, while the fully connected layers that do not directly correspond to intuitive visual features, carry information that has more in common with the activity in the lower frequency bands.

Discussion

Previous work has established a correspondence between hierarchy of the DCNN and the fMRI responses measured across the human visual areas (Güçlü and van Gerven, 2015; Eickenberg et al., 2016; Seibert et al., 2016; Cichy et al., 2016b). Studying intracranial recordings allowed us to extend previous findings by assessing the alignment between the DCNN and cortical electrophysiological signals at different frequency bands. As there is a quantifiable increase of the complexity of features along the layers of the DCNN, any signal that is aligned to the DCNN has to carry similarly increasingly complex features built-up during visual object recognition. We observed that the lower layers of the DCNN explained gamma band signals from earlier visual areas, while higher layers of the DCNN, responsive for more complex features, matched with the gamma band signals from higher visual areas. Hence, one can conclude that gamma band carries increasingly complex features required for object recognition along the ventral visual pathway. This finding agrees with the previous work that has given a central role for gamma band activity in visual object recognition (Singer and Gray, 1995; Singer, 1999; Fisch et al., 2009) and feedforward communication (Van Kerkoerle et al., 2014; Bastos et al., 2015; Michalareas et al., 2016). However, importantly, our results show that gamma activity reflects not only object recognition per se but also the feature transformations that are computed on the way towards explicit object representations.

Low vs high gamma in object recognition

We observed significant alignment to the DCNN in both low and high gamma bands. Previous studies have shown that low and high gamma frequencies are functionally different: while low gamma is more related to classic gamma oscillations, high frequencies seem to reflect local spiking activity rather than oscillations (Manning et al., 2009; Ray and Maunsell, 2011). In the current work we approached the data analysis from the machine learning perspective and remained agnostic with respect to the oscillatory nature of underlying signals. Importantly, we found that numerically the alignment to the DCNN was stronger in low gamma frequencies. However, high gamma was more prominent when considering volume and specificity to visual areas. The most striking difference between the low and high gamma with regard to specificity was in the earliest time window 50-250 ms where the correlation between the DCNN and high gamma was almost exclusive to visual areas.

The role of alpha activity in object recognition

Another finding from our work was that alpha band responses were also aligned to the DCNN. This result implies that not only gamma, but also the alpha band is a channel for increasingly complex feature transformation during visual object recognition. Prior studies have shown that feedforward activity from lower to higher visual areas is carried by the gamma frequency whereas the alpha-beta band reflects feedback signals from higher to lower visual areas (Van Kerkoerle et al., 2014; Bastos et al., 2015; Michalareas et al., 2016). Such network mechanisms have been captured by recent large-scale modelling (Mejias et al., 2016). While our results from the gamma frequency converge towards these previous findings, the alignment of alpha band activity with DCNN is interpreted differently. Namely, although the DCNN is a purely feedforward network it is important to notice that the alignment between electrophysiological signals and the DCNN does not imply that the respective signals have to reflect feedforward computations. Such alignment only means that the progressive changes in representational geometry along

the processing hierarchy are similar to the DCNN. In other words, it is possible that the activity patterns observed in the alpha frequency are a result of recurrent computations, but their outcome representational geometry resembles that of the DCNN. Hence, the findings also coherently fit with results demonstrating that alpha responses carry signal content that is not sensory in origin but contributes to the shaping and interpretation of sensory information in perceptual recognition, i.e. perceptual priors (Mayer et al., 2016; Samaha et al., 2016). Within the predictive coding framework feedback activity is not an unspecific modulatory signal but rather has to signal specific contents from higher to lower levels of the processing hierarchy (Bastos et al., 2012). Therefore, within this theoretical framework, a specific representational geometry is expected even from a feedback channel.

The question whether alpha reflects feedforward or feedback computations is directly related to the issue whether the current alpha results could be explained by the intracranial ERPs. In particular, as ERPs strongly influence the alpha frequency, then a feedforward propagation of ERPs along the ventral pathway could in principle lead to the observed alpha alignment. If this propagation would be feedforward, i.e. if ERP peaks would occur earlier in lower and later in higher areas, then the observed alpha results would rather favor a feedforward interpretation.

To investigate this issue we inspected the ERP traces of individual electrodes but could not find any evidence for such alignment of ERP peaks along the ventral pathway. We note that the issue of ERPs warrants further investigation, in particular as the ERPs have been shown to have equal response selectivity to alpha and gamma frequencies (Vidal et al., 2010). Importantly, these previous results also indicated that ERPs do not explain the results in the alpha band as the category selective information in the ERPs and the alpha and gamma bands was mainly contained in non-overlapping electrodes (Vidal et al., 2010). Taken together, we conclude that our alpha effects were likely not caused by feedforward propagation of ERP peaks. Our present findings thus call for a reevaluation of the role of the alpha band in visual object recognition.

Limitations

The present work relies on data pooled over the recordings from 81 subjects. Hence, the correspondence we found between responses at different frequency bands and layers of DCNN is distributed over many subjects. While it is expected that single subjects show similar mappings (see also Figure 6), the variability in number and location of recording electrodes in individual subjects makes it difficult a full single-subject analysis with this type of data. We also note that the mapping between electrode locations and Brodmann areas is approximate and the exact mapping would require individual anatomical reconstructions and more refined atlases.

Future work

Intracranial recordings are both precisely localized in space and time, thus allowing us to explore phenomena not observable with fMRI. In this work we investigated the correlation of DCNN activity with five broad frequency bands and three time windows. Our next steps will include the analysis of the activity on a more granular temporal and spectral scale. Replacing representation similarity analysis with a predictive model (such as regularized linear regression) will allow us to explore which visual features elicited the highest responses in the visual cortex.

Acknowledgements

IK, RV and JA thank the financial support from the Estonian Research Council through the personal research grants PUT438 and PUT1476. This work was supported by the Estonian Centre of Excellence in IT (EXCITE), funded by the European Regional Development Fund.

References

Bastos AM, Usrey WM, Adams RA, Mangun GR, Fries P, Friston KJ (2012) Canonical microcircuits for predictive coding. *Neuron* 76:695–711.

Bastos AM, Vezoli J, Bosman CA, Schoffelen JM, Oostenveld R, Dowdall JR, De Weerd P, Kennedy H, Fries P (2015) Visual areas exert feedforward and feedback influences through distinct frequency channels. *Neuron* 85:390–401.

Brodmann K (1909) *Vergleichende Lokalisationslehre der Groshirnrinde* Barth.

Cichy RM, Khosla A, Pantazis D, Torralba A, Oliva A (2016a) Deep neural networks predict hierarchical spatio-temporal cortical dynamics of human visual object recognition. *arXiv preprint arXiv:1601.02970* .

Cichy RM, Khosla A, Pantazis D, Torralba A, Oliva A (2016b) Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. *Scientific reports* 6.

Daubechies I (1990) The wavelet transform, time-frequency localization and signal analysis. *IEEE transactions on information theory* 36:961–1005.

Delorme A, Makeig S (2004) Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of neuroscience methods* 134:9–21.

DiCarlo JJ, Zoccolan D, Rust NC (2012) How does the brain solve visual object recognition? *Neuron* 73:415–434.

Eickenberg M, Gramfort A, Varoquaux G, Thirion B (2016) Seeing it all: Convolutional network layers map the function of the human visual system. *NeuroImage* .

Fisch L, Privman E, Ramot M, Harel M, Nir Y, Kipervasser S, Andelman F, Neufeld MY, Kramer U, Fried I et al. (2009) Neural ignition: enhanced activation linked to perceptual awareness in human ventral stream visual cortex. *Neuron* 64:562–574.

Genovese CR, Lazar NA, Nichols T (2002) Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage* 15:870–878.

Gray CM, Singer W (1989) Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *Proceedings of the National Academy of Sciences* 86:1698–1702.

Güçlü U, van Gerven MA (2015) Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. *The Journal of Neuroscience* 35:10005–10014.

Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, Guadarrama S, Darrell T (2014) Caffe: Convolutional architecture for fast feature embedding. *arXiv preprint arXiv:1408.5093* .

Khaligh-Razavi SM, Kriegeskorte N (2014) Deep supervised, but not unsupervised, models may explain it cortical representation. *PLoS Comput Biol* 10:e1003915.

Kriegeskorte N (2015) Deep neural networks: a new framework for modeling biological vision and brain information processing. *Annual Review of Vision Science* 1:417–446.

Kriegeskorte N, Mur M, Bandettini PA (2008) Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in systems neuroscience* 2:4.

Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks In *Advances in neural information processing systems*, pp. 1097–1105.

Lachaux JP, George N, Tallon-Baudry C, Martinerie J, Hugueville L, Minotti L, Kahane P, Renault B (2005) The many faces of the gamma band response to complex visual stimuli. *Neuroimage* 25:491–501.

Manning JR, Jacobs J, Fried I, Kahana MJ (2009) Broadband shifts in local field potential power spectra are correlated with single-neuron spiking in humans. *Journal of Neuroscience* 29:13613–13620.

Mayer A, Schwiedrzik CM, Wibral M, Singer W, Melloni L (2016) Expecting to see a letter: alpha oscillations as carriers of top-down sensory predictions. *Cerebral Cortex* 26:3146–3160.

Mejias JF, Murray JD, Kennedy H, Wang XJ (2016) Feedforward and feedback frequency-dependent interactions in a large-scale laminar network of the primate cortex. *Science Advances* 2:e1601335.

Michalareas G, Vezoli J, Van Pelt S, Schoffelen JM, Kennedy H, Fries P (2016) Alpha-beta and gamma rhythms subserve feedback and feedforward influences among human visual cortical areas. *Neuron* 89:384–397.

Ray S, Maunsell JH (2011) Different origins of gamma rhythm and high-gamma activity in macaque visual cortex. *PLoS Biol* 9:e1000610.

Rorden C (2007) Mricron [computer software].

Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M, Berg AC, Fei-Fei L (2015) ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)* 115:211–252.

Samaha J, Boutonnet B, Lupyan G (2016) How prior knowledge prepares perception: Prestimulus oscillations carry perceptual expectations and influence early visual responses. *bioRxiv* p. 076687.

Seibert D, Yamins DL, Ardila D, Hong H, DiCarlo JJ, Gardner JL (2016) A performance-optimized model of neural responses across the ventral visual stream. *bioRxiv* p. 036475.

Singer W (1999) Neuronal synchrony: a versatile code for the definition of relations? *Neuron* 24:49–65.

Singer W, Gray CM (1995) Visual feature integration and the temporal correlation hypothesis. *Annual review of neuroscience* 18:555–586.

Talairach J, Tournoux P (1993) *Referentially oriented cerebral MRI anatomy: an atlas of stereotaxic anatomical correlations for gray and white matter* Thieme.

Van Kerkoerle T, Self MW, Dagnino B, Gariel-Mathis MA, Poort J, Van Der Togt C, Roelfsema PR (2014) Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. *Proceedings of the National Academy of Sciences* 111:14332–14341.

Vidal JR, Ossandón T, Jerbi K, Dalal SS, Minotti L, Ryvlin P, Kahane P, Lachaux JP (2010) Category-specific visual responses: an intracranial study comparing gamma, beta, alpha, and erp response selectivity. *Frontiers in human neuroscience* 4:195.

Yamins DL, DiCarlo JJ (2016) Using goal-driven deep learning models to understand sensory cortex. *Nature neuroscience* 19:356–365.

Yamins DL, Hong H, Cadieu CF, Solomon EA, Seibert D, DiCarlo JJ (2014) Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences* 111:8619–8624.