COMMUNICATIONS
BIOLOGY

ARTICLE

DOI: 10.1038/s42003-018-0110-y OPEN

Activations of deep convolutional neural networks are aligned with gamma band activity of human visual cortex

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[8,9,10], Juan R. Vidal[4,5,11] & Jaan Aru[1,12]

Recent advances in the field of artificial intelligence have revealed principles about neural processing, in particular about vision. 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 use DCNN to investigate which frequency bands correlate with feature transformations of increasing complexity along the ventral visual pathway. By capitalizing on intracranial depth recordings from 100 patients we assess the alignment between the DCNN and signals at different frequency bands. We find that gamma activity (30–70 Hz) matches the increasing complexity of visual feature representations in DCNN. These findings show that the activity of the DCNN captures the essential characteristics of biological object recognition not only in space and time, but also in the frequency domain. These results demonstrate the potential that artificial intelligence algorithms have in advancing our understanding of the brain.

[1] Computational Neuroscience Lab, Institute of Computer Science, University of Tartu, Tartu 51005, Estonia. [2] INSERM U1028, CNRS UMR5292, Brain Dynamics and Cognition Team, Lyon Neuroscience Research Center, Bron 69500, 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] INSERM U1028, CNRS UMR5292, TIGER Team, Lyon Neuroscience Research Center, Bron 69500, France. [9] Department of Functional Neurology and Epileptology, Hospices Civils de Lyon, Bron 69500, France. [10] Epilepsy Institute, Bron 69500, France. [11] Catholic University of Lyon, Lyon 69002, France. [12] Department of Penal Law, School of Law, University of Tartu, Tallinn 10119, Estonia. These authors contributed equally: Raul Vicente, Jaan Aru. Correspondence and requests for materials should be addressed to I.K. (email: ilya.kuzovkin@gmail.com) or to R.V. (email: raulvicente@gmail.com) or to J.A. (email: jaan.aru@gmail.com)

Biological visual object recognition is mediated by a hierarchy of increasingly complex feature representations along the ventral visual stream[1]. Intriguingly, these transformations are matched by the hierarchy of transformations learned by deep convolutional neural networks (DCNN) trained on natural images[2]. 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[3] and humans[4]. Other studies with functional magnetic resonance imaging (fMRI) data have demonstrated a direct correspondence between the hierarchy of the human visual areas and layers of the DCNN[2,5–7]. In sum, the increasing feature complexity of the DCNN corresponds to the increasing feature complexity occurring in visual object recognition in the primate brain[8,9].

However, fMRI based studies only allow one to localize object recognition in space, but neural processes also unfold in time and have characteristic spectral fingerprints (i.e., frequencies). With time-resolved magnetoencephalographic recordings it has been demonstrated that the correspondence between the DCNN and neural signals peaks in the first 200 ms[7,10]. Here, we test the remaining dimension: that biological visual object recognition is also specific to certain frequencies. In particular, there is a long-standing hypothesis that especially gamma band (30–150 Hz) signals are crucial for object recognition[11–22]. More modern views on gamma activity emphasize the role of the gamma rhythm in establishing a communication channel between areas[23,24]. Further research has demonstrated that especially feedforward communication from lower to higher visual areas is carried by the gamma frequencies[25–27]. As the DCNN is a feedforward network one could expect that the DCNN will correspond best with the gamma band activity. In this work we used the DCNN as a computational model to assess whether signals in the gamma frequency are more relevant for object recognition than other frequencies.

To empirically evaluate whether gamma frequency has a specific role in visual object recognition we assessed the alignment between the responses of layers of a commonly used DCNN and the neural signals in five distinct frequency bands and three time windows along the areas constituting the ventral visual pathway. Based on the previous findings we expected that: mainly gamma frequencies should be aligned with the layers of the DCNN; the correspondence between the DCNN and gamma should be confined to early time windows; the correspondence between gamma and the DCNN layers should be restricted to visual areas. In order to test these predictions we capitalized on direct intracranial depth recordings from 100 patients with epilepsy and a total of 11,293 electrodes implanted throughout the cerebral cortex.

We observe 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. We also find that while the neural activity in the theta range (5–8 Hz) is not aligned with the DCNN hierarchy, the representational geometry of theta activity is correlated with the representational geometry of higher layers of DCNN.

Results

Activity in gamma band is aligned with the DCNN. We tested the hypothesis that gamma activity has a specific role in visual object recognition compared to other frequencies. To that end we assessed the alignment of neural activity in different frequency bands and time windows to the activity of layers of a DCNN trained for object recognition.

In particular, we used representational similarity analysis (RSA) to compare the representational geometry of different DCNN layers and the activity patterns of different frequency bands of single electrodes (see Fig. 1).

We consistently found that signals in low-gamma (31–70 Hz) frequencies across all time windows and high-gamma (71–150 Hz) frequencies in 150–350 ms window are aligned with the DCNN in a specific way: increase of the complexity of features along the layers of the DCNN was roughly 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 2a illustrates assignment of neural activity in low-gamma band and Fig. 2b the high-gamma band 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, and 20). Focusing on visual areas revealed a diagonal trend that illustrates the alignment between ventral stream and layers of DCNN (see Fig. 3).

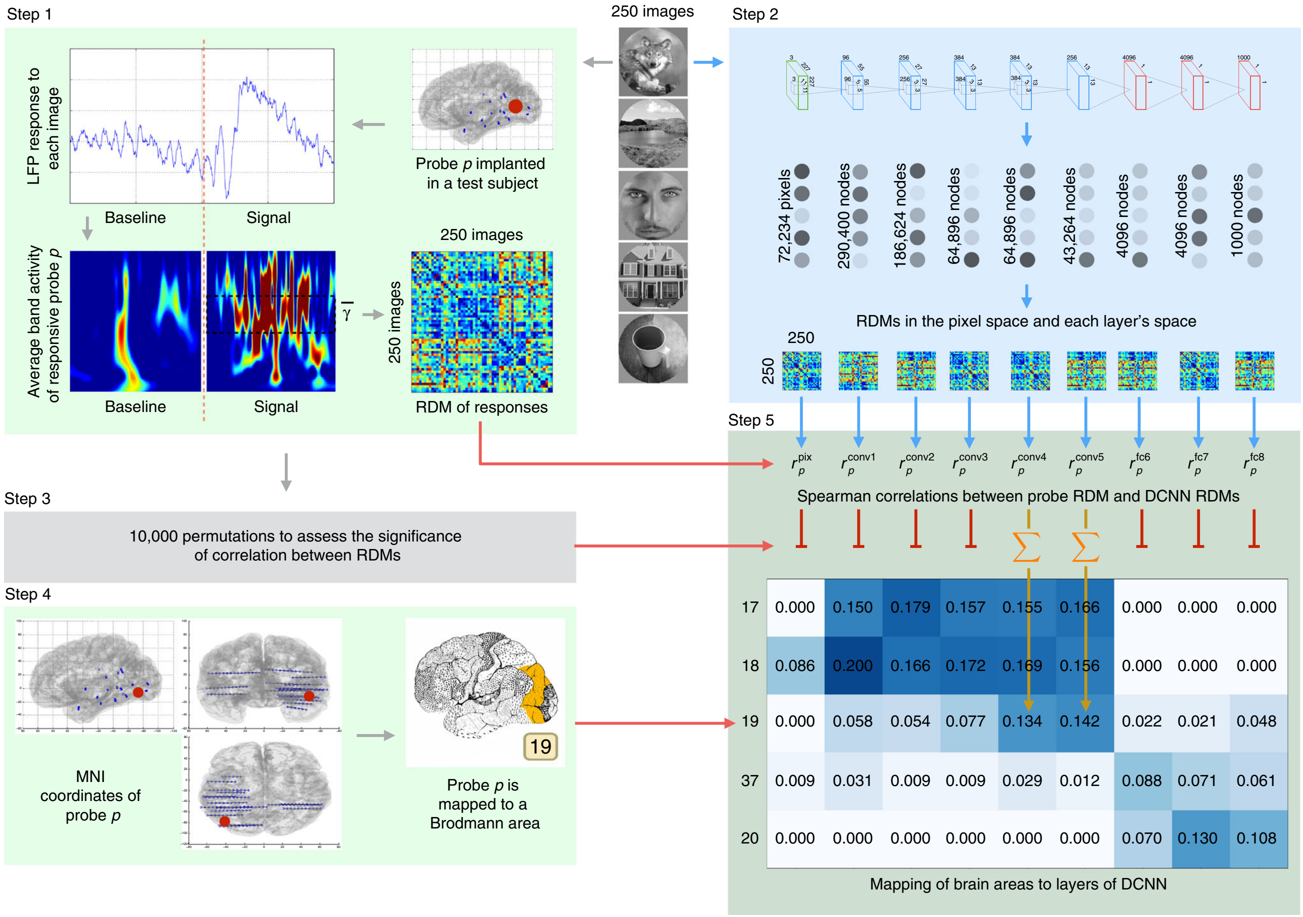
Our findings across all subjects, time windows and frequency bands are summarized in Fig. 4a. We note that the alignment in the gamma bands is also present at the single-subject level (Supplementary Fig. 1).

Apart from the alignment we looked at the total amount of correlation and its specificity to visual areas. Fig. 4b shows the volume of significantly correlating activity was highest in the high-gamma range. Remarkably, 97% of that activity was located in visual areas, which is confirmed in Fig. 2 where we see that in the gamma range only a few electrodes were assigned to Brodmann areas that are not part of the ventral stream.

Activity in other frequency bands. To test the specificity of gamma frequency in visual object recognition, we assessed the alignment between the DCNN and other frequencies. The detailed mapping results for all frequency bands and time windows are presented in layer-to-area fashion in Fig. 3. The results in the right column of Table 1 show the alignment values and significance levels for a DCNN that is trained for object recognition on natural images. On the left part of Table 1 the alignment between the brain areas and a DCNN that has not been trained on object recognition (i.e., has random weights) is given for comparison. One can see that training a network to classify natural images drastically increases the alignment score ρ and its significance. One can see that weaker alignment (that does not survive the Bonferroni-correction) is present in early time window in theta and alpha frequency range. No alignment is observed in the beta band.

In order to take into account the intrinsic variability when comparing alignments of different bands between each other, we performed a set of tests to see which bands have statistically significantly higher alignment with DCNN than other bands. See the Methods section “Mapping neural activity to layers of DCNN” for details. The results of those tests are presented in Table 2. Based on these results we draw a set of statistically significant conclusions on how the alignment of neural responses with the activations of DCNN differs between frequency bands and time windows. In the low-gamma range (31–70 Hz) we conclude that the alignment is larger than with any other band and that within the low gamma the activity in early time window 50–250 ms is aligned more than in later windows. Alignment in the high-gamma (71–150 Hz) is higher than the alignment of θ, but not higher than alignment of α. Within the high-gamma band the activity in the middle time window 150–350 ms has the highest alignment, followed by late 250–450 ms window and then by the early activity in 50–250 ms window. Outside the gamma range we conclude that theta band has the weakest alignment across all bands and that alignment of early alpha activity is higher than the alignment of early and late high gamma.

Alignment is dependent on having two types of layers in DCNN. In Figs. 2 and 3 one can observe that sites in lower visual areas (17, 18) are mapped to DCNN layers 1–5 without a clear trend but are not mapped to layers 6–8. Similarly areas 37 and 20 are mapped to layers 6–8, but not to 1–5. Hence, we next asked whether the observed alignment is depending on having two different groups of visual areas related to two groups of DCNN layers. We tested this by computing alignment within the subgroups. We looked at alignment only between the lower visual areas (17–19), and the convolutional layers 1–5, and separately at the alignment between higher visual areas (37, 20) and fully connected layers of DCNN (6–8). We observed no significant alignment within any of the subgroups. So we conclude that the alignment mainly comes from having different groups of areas related more or less equally to two groups of layers. The underlying reason for having these two groups of layers comes from the structure of the DCNN—it has two different types of layers, convolutional (layers 1–5) and fully connected (layers 6–8) (See Fig. 5a, b for a visualization of the different layers and their learned features and a longer explanation of the differences between the layers in the Discussion). As can be evidenced in Fig. 6 the layers 1–5 and 6–8 of the DCNN indeed cluster into two groups. Taken together, we observed that early visual areas are mapped to the convolutional layers of the DCNN, whereas higher visual areas match the activity profiles of the fully connected layers of the DCNN.


Step 1
LFP response to each image
Baseline
Signal
Probe p implanted in a test subject
Average band activity of responsive probe p
Baseline
Signal
γ
250 images
250 images
RDM of responses
250 images
Step 2
3
227
96
55
256
27
384
13
384
13
256
13
4096
4096
1000
72,234 pixels
290,400 nodes
186,624 nodes
64,896 nodes
64,896 nodes
43,264 nodes
4096 nodes
4096 nodes
1000 nodes
RDMs in the pixel space and each layer's space
250
250
Step 3
10,000 permutations to assess the significance of correlation between RDMs
Step 4
MNI coordinates of probe p
19
Probe p is mapped to a Brodmann area
Step 5
r_p^{pix}
r_p^{conv1}
r_p^{conv2}
r_p^{conv3}
r_p^{conv4}
r_p^{conv5}
r_p^{fc6}
r_p^{fc7}
r_p^{fc8}
Spearman correlations between probe RDM and DCNN RDMs
Σ
Σ
Mapping of brain areas to layers of DCNN

17	0.000	0.150	0.179	0.157	0.155	0.166	0.000	0.000	0.000
18	0.086	0.200	0.166	0.172	0.169	0.156	0.000	0.000	0.000
19	0.000	0.058	0.054	0.077	0.134	0.142	0.022	0.021	0.048
37	0.009	0.031	0.009	0.009	0.029	0.012	0.088	0.071	0.061
20	0.000	0.000	0.000	0.000	0.000	0.000	0.070	0.130	0.108

Fig. 1 Overview of the analysis pipeline. Two hundred and fifty natural images are presented to human subjects (Step 1) and to an artificial vision system (Step 2). 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). **Step 1:** LFP response of each of 11,293 electrodes to each of the images is converted into the frequency domain. Activity evoked by each image is compared to the activity evoked by every other image and results of this comparison are presented as a representational dissimilarity matrix (RDM). **Step 2:** Each of the images is shown to a pretrained 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. **Step 3:** Subject's intracranial responses to stimuli are randomly reshuffled and the analysis performed in step 1 is repeated 10,000 times to obtain 10,000 random RDMs for each electrode. **Step 4:** 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). **Step 5:** Spearman's rank correlation is computed between the true (nonpermuted) RDM of neural responses and RDMs of each layer of DCNN. Also 10,000 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 (the value of $p < 0.001$ is estimated by selecting a threshold such that none of the probes would pass it on the permuted data), 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 per Brodmann area. The resulting mapping matrix shows the alignment between the consecutive areas of the ventral stream and layers of DCNN

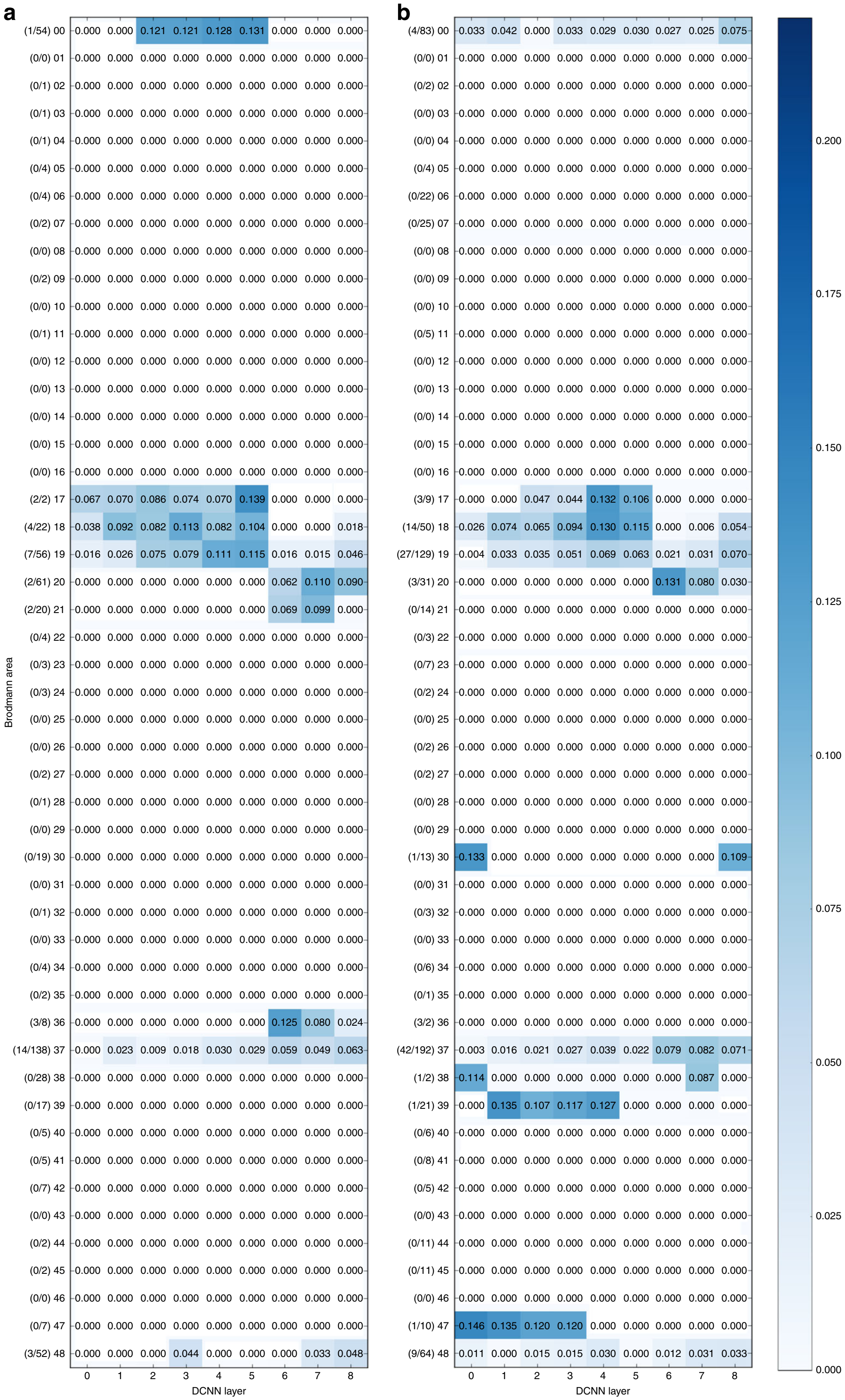


Fig. 2 Mapping of the activity in Brodmann areas to DCNN layers. Underlying data comes from the activity in low gamma (31–70 Hz, **a**) and high-gamma (71–150 Hz, **b**) bands in 150–350 ms time window. On the vertical axis there are Brodmann areas and the number of significantly correlating probes in each area out of the total number of responsive probes in that area. Horizontal axis represents succession of layers of DCNN. Number in each cell of the matrix is the total sum of correlations (between RDMs of probes in that particular area and the RDM of that layer) normalized by the number of significantly correlating probes in an area

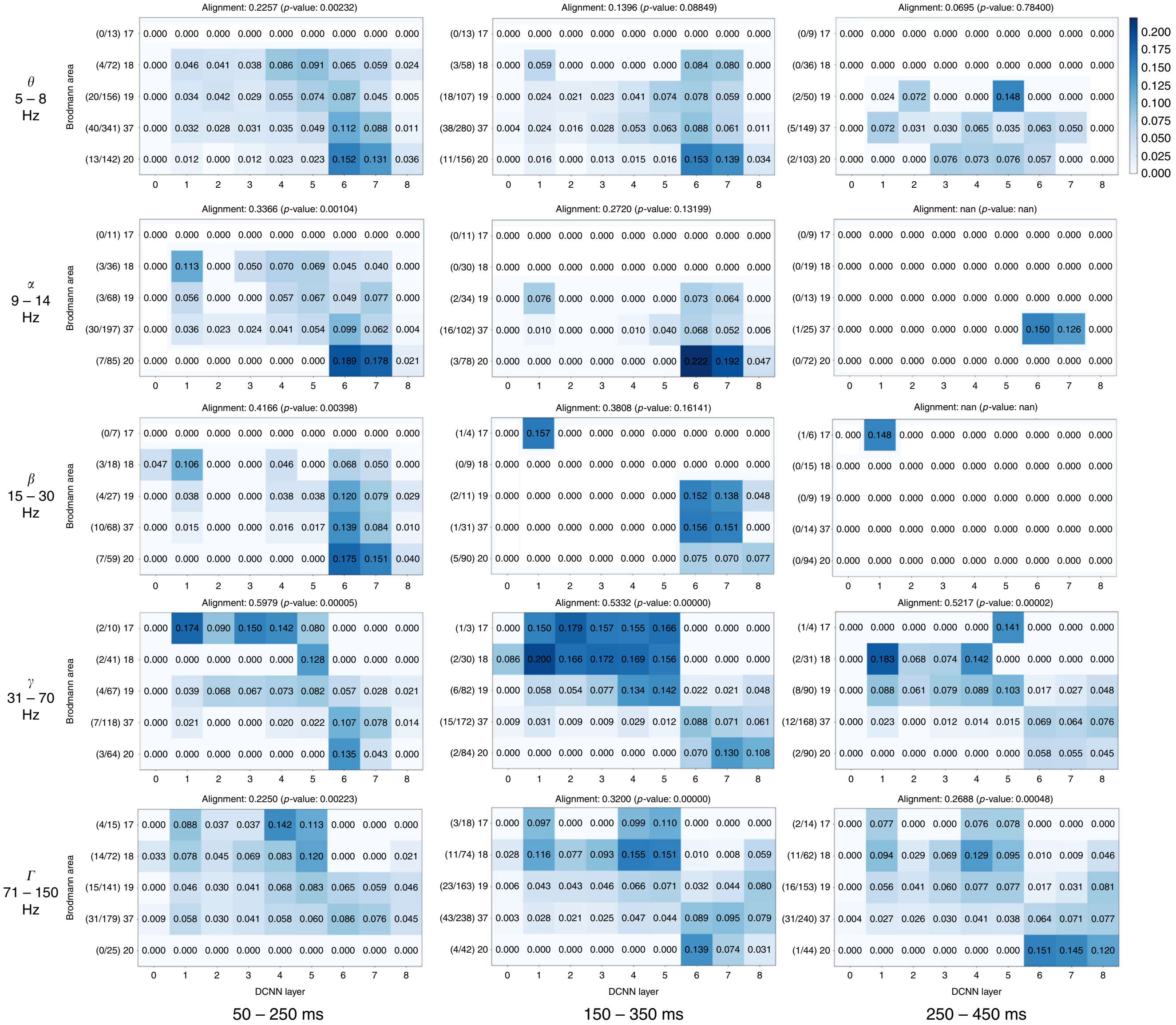


Fig. 3 Mapping of activity in visual areas to activations of layers of DCNN across five frequency bands and three time windows. Vertical axis holds Brodmann areas in the order of the ventral stream (top to bottom), horizontal axis represents the succession of layers of DCNN. Number in each cell of a matrix is the total sum of correlations (between RDMs of probes in that particular area and the RDM of that layer) normalized by the number of significantly correlating probes in an area. The alignment score is computed as Spearman's rank correlation between electrode assignment to Brodmann areas and electrode assignment to DCNN layers (Eq. (8)). 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

Visual complexity varies across areas and frequencies. To investigate the involvement of each frequency band more closely we analyzed each visual area separately. Figure 7 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 than 1%). One can see in Fig. 7 that activity in theta frequency in time windows 50–250 and 150–350 ms had large volume, and is correlated with the higher layers of DCNN in higher visual areas (19, 37, 20) of the ventral stream. This hints at the role of activity reflected by the theta band in visual object recognition. In general, in areas 37 and 20 all frequency bands reflected the 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. 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 5 presents a visual summary of this analysis. In the Methods section we have defined total volume of visual activity in layers $\mathbf{L}$ as $V_{\mathbf{L}}$. We used average of this measure over frequency band intervals to quantify the activity in low- and high-gamma bands. We noticed that while the fraction of gamma activity that is mapped to convolutional layers is high $\left(\frac{\bar{V}^{\gamma,\Gamma}_{\mathbf{L}=\{\text{conv1}\dots\text{conv5}\}}}{\bar{V}^{\text{allbands}}_{\{\mathbf{L}=\text{conv1}\dots\text{conv5}\}}} = 0.71\right)$, this fraction diminished in fully connected layers fc6 and fc7 $\left(\frac{\bar{V}^{\gamma,\Gamma}_{\mathbf{L}=\{\text{fc6,fc7}\}}}{\bar{V}^{\text{allbands}}_{\mathbf{L}=\{\text{fc6,fc7}\}}} = 0.39\right)$. 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 0.29, while in fully connected it grew to 0.61. This observation highlighted the fact that visual features extracted by convolutional filters of DCNN are more similar to gamma frequency activity, 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

The recent advances in artificial intelligence research have demonstrated a rapid increase in the ability of artificial systems to solve various tasks that are associated with higher cognitive functions of human brain. One of such tasks is visual object recognition. Not only do the deep neural networks match human performance in visual object recognition, they also provide the best model for how biological object recognition happens[3,8,9,28]. Previous work has established a correspondence between hierarchy of the DCNN and the fMRI responses measured across the human visual areas[2,5–7]. Further research has shown that the activity of the DCNN matches the biological neural hierarchy in time as well[7,10]. Studying intracranial recordings allowed us to extend previous findings by assessing the alignment between the DCNN and cortical signals at different frequency bands. We observed that the lower layers of the DCNN explained gamma band signals from earlier visual areas, while higher layers of the DCNN, responsible for more complex features, matched with the gamma band signals from higher visual areas. This finding confirms previous work that has given a central role for gamma band activity in visual object recognition[11–13] and feedforward communication[25–27]. Our work also demonstrates that the correlation between the DCNN and the biological counterpart is specific not only in space and time, but also in frequency.

The research into gamma oscillations started with the idea that gamma band activity signals the emergence of coherent object representations[11,12,29]. However, this view has evolved into the understanding that activity in the gamma frequencies reflects neural processes more generally. One particular view[23,24] suggests that gamma oscillations provide time windows for communication between different brain regions. Further research has shown that especially feedforward activity from lower to higher visual areas is carried by the gamma frequencies[25–27]. As the DCNN is a feedforward network our current findings support the idea that gamma rhythms provide a channel for feedforward communication. However, our results by no means imply that gamma rhythms are only used for feedforward visual object recognition. There might be various other roles for gamma rhythms[24,30].

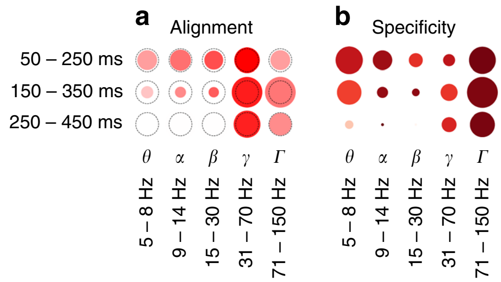


Fig. 4 Overall relative statistics of brain responses across frequency bands and time windows. **a** The alignment between visual brain areas and DCNN layers (see Eq. (8)). The color indicates the correlation value (ρ) while the size of the marker shows the logarithm (so that not significant results are still visible on the plot) of inverse of the statistical significance of the correlation, dotted circle indicates $p = 0.0003(3)$—the Bonferroni-corrected significance threshold level of 0.005. **b** Activity in a region of interest is specific to visual areas (see Eq. (4)): intense red means that most of the activity in that band and time window happened in visual areas, size of the marker indicates total volume (Eq. (2)) of activity in all areas. The maximal size of a marker is defined by the biggest marker on the figure

Table 1 Alignment score ρ_{align} and the significance levels for all 15 regions of interest

Band	Window	Alignment with layers of randomly initialized AlexNet		Alignment with layers of AlexNet trained on ImageNet	
		ρ_{align}	p value	ρ_{align}	p value
θ	50–250 ms	0.0632	0.71	0.2257	0.00231575[a]
θ	150–350 ms	−0.1013	0.59	0.1396	0.08848501
θ	250–450 ms	0.1396	0.59	0.0695	0.78400416
α	50–250 ms	−0.2411	0.32	0.3366	0.00103551[a]
α	150–350 ms	0.0000	1.00	0.2720	0.13199463
α	250–450 ms	-	-	-	-
β	50–250 ms	-	-	0.4166	0.00397929
β	150–350 ms	-	-	0.3808	0.16141286
β	250–450 ms	-	-	-	-
γ	50–250 ms	0.1594	0.62	0.5979	0.00004623[b]
γ	150–350 ms	−0.1688	0.34	0.5332	0.00000059[b]
γ	250–450 ms	−0.1132	0.56	0.5217	0.00001624[b]
Γ	50–250 ms	0.0869	0.42	0.2259	0.00222940[a]
Γ	150–350 ms	−0.0053	0.96	0.3200	0.00000051[b]
Γ	250–450 ms	−0.1361	0.33	0.2688	0.00047999[a]

[a] Alignments that pass p value threshold of 0.05 Bonferroni-corrected to <0.003(3)
[b] Ones that pass 0.005[54] Bonferroni-corrected to <0.0003(3)
Note how the values differ between random (control) network and a network trained on natural images. Visual representation of alignment and significance is given in Fig. 4a

Table 2 Comparison of the alignment across regions of interest

0.2079 ± 0.1381	θ^{50}	>	-
0.3352 ± 0.0989	α^{50}	>	θ^{50}, Γ^{50}, Γ^{250}
0.5652 ± 0.1953	γ^{50}	>	θ^{50}, α^{50}, γ^{150}, γ^{250}, Γ^{50}, Γ^{150}, Γ^{250}
0.4880 ± 0.1650	γ^{150}	>	θ^{50}, α^{50}, Γ^{50}, Γ^{150}, Γ^{250}
0.4656 ± 0.2185	γ^{250}	>	θ^{50}, α^{50}, Γ^{50}, Γ^{150}, Γ^{250}
0.2172 ± 0.1179	Γ^{50}	>	-
0.3116 ± 0.1115	Γ^{150}	>	θ^{50}, Γ^{50}, Γ^{250}
0.2494 ± 0.1381	Γ^{250}	>	θ^{50}, Γ^{50}

Alignment of the region of interest on the left is statistically significantly larger than the alignments of the regions of interest on the right. To obtain these results a pairwise comparison of the magnitude of alignment between the regions of interest was made. First column enlists significantly aligned regions, their average alignment ρ score when estimated on 1000 random subsets of the data (each of the half of the size of the dataset), and standard deviation of the alignment. On the right side of the table we list the regions of interest of which the ROI on the left is larger. The hypothesis was tested using Mann–Whithney U test and only the results with the p values that have passed the threshold of 2.2e−5 (0.005 Bonferroni-corrected to take into account multiple comparisons) are presented in the table

We observed significant alignment to the DCNN in both low and high-gamma bands. However, when directly contrasted the alignment was stronger for low-gamma signals. Furthermore, for high gamma this alignment was more restricted in time, surviving correction only in the middle time window. Previous studies have shown that low and high-gamma frequencies are functionally different: while low gamma is more related to classic narrow-band gamma oscillations, high frequencies seem to reflect local

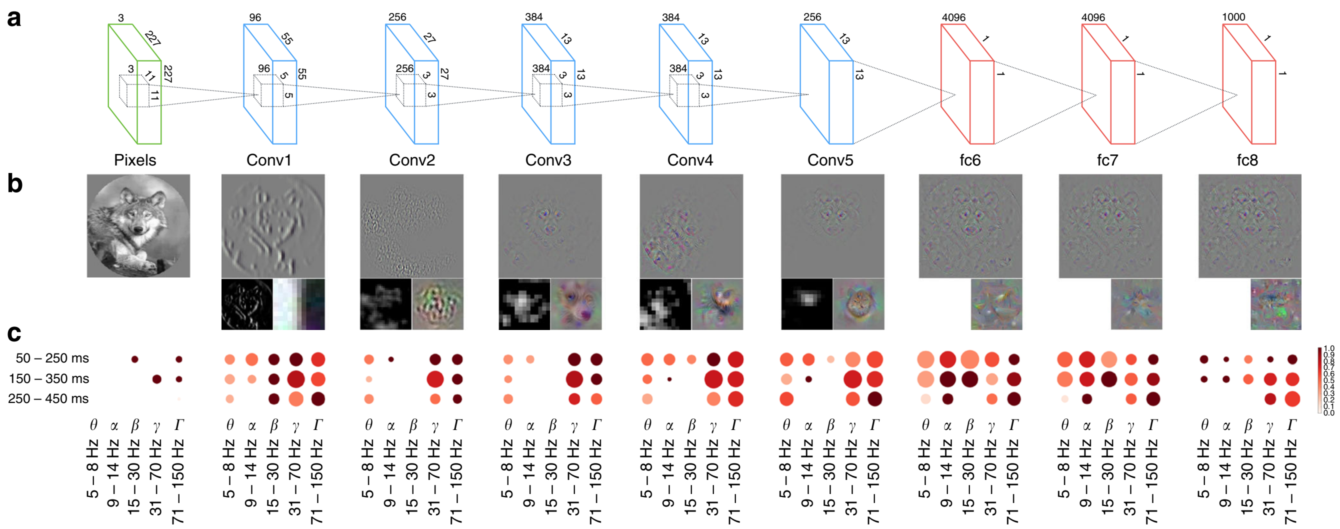
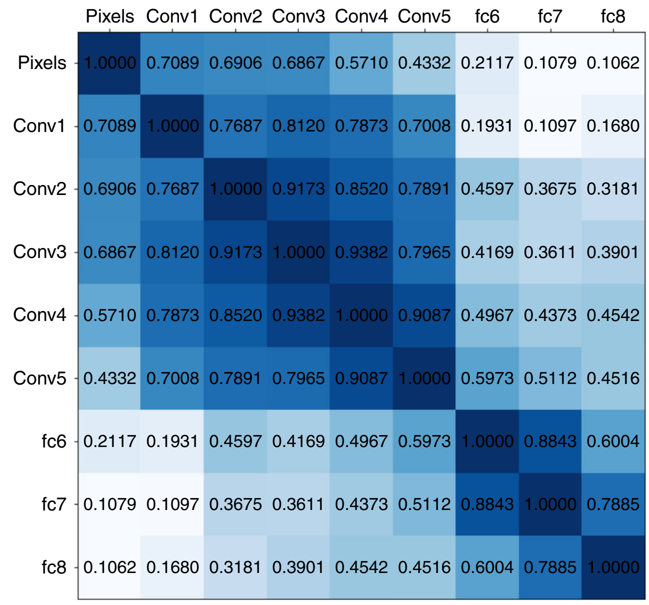


Fig. 5 Specificity of neural responses to layers of DCNN across frequency bands and time windows. **a** The architecture of the DCNN. Convolutional layer 1 consists of 96 feature detectors of size 11 × 11, they take as input pixels of the image and their activations create 96 features maps of size 55 × 55, architecture of all consecutive convolutional layers is analogous. Five convolutional layers are followed by three fully connected layers of sizes 4096, 4096, and 1000, respectively. **b** The leftmost image is an example input image. For each layer we have selected one interesting filter that depicts what is happening inside of the neural network and plotted: (a) a reconstruction of the original image from the activity of that neuron using the deconvolution[48] technique (upper larger image), (b) activations on the feature map generated by that neuron (left subimage), and (c) synthetic image that shows what input the neuron would be most responsive to (right subimage). Visualizations were made with Deep Visualization Toolbox[55]. All filters are canonical to AlexNet trained on ImageNet and can be explored using the above-mentioned visualization tool or visualized directly from the publicly available weights of the network. **c** 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 the Brodmann areas constituting the ventral stream: BA17-18-19-37-20

spiking activity rather than oscillations[31,32], the distinction between low and high-gamma activity has also implications from cognitive processing perspective[17,19]. In the current work we approached the data analysis from the machine learning point of view and remained agnostic with respect to the oscillatory nature of underlying signals. Importantly, we found that numerically the alignment to the DCNN was stronger and persisted for longer in low-gamma frequencies. However, high gamma was more prominent when considering volume and specificity to visual areas. These results match well with the idea that whereas high-gamma signals reflect local spiking activity, low-gamma signals are better suited for adjusting communication between brain areas[23,24].

In our work we observed that the significant alignment depended on the fact that there are two groups of layers in the DCNN: the convolutional and fully connected layers. We found that these two types of layers have similar activity patterns (i.e., representational geometry) within the group but the patterns are less correlated between the groups (Fig. 6). As evidenced in the data, in the lower visual areas (17, 18) the gamma band activity patterns resembled those of convolutional layers, whereas in the higher areas (37 and 20) gamma band activity patterns matched the activity of fully connected layers. Area 19 showed similarities to both types of DCNN layers.



	Pixels	Conv1	Conv2	Conv3	Conv4	Conv5	fc6	fc7	fc8
Pixels	1.0000	0.7089	0.6906	0.6867	0.5710	0.4332	0.2117	0.1079	0.1062
Conv1	0.7089	1.0000	0.7687	0.8120	0.7873	0.7008	0.1931	0.1097	0.1680
Conv2	0.6906	0.7687	1.0000	0.9173	0.8520	0.7891	0.4597	0.3675	0.3181
Conv3	0.6867	0.8120	0.9173	1.0000	0.9382	0.7965	0.4169	0.3611	0.3901
Conv4	0.5710	0.7873	0.8520	0.9382	1.0000	0.9087	0.4967	0.4373	0.4542
Conv5	0.4332	0.7008	0.7891	0.7965	0.9087	1.0000	0.5973	0.5112	0.4516
fc6	0.2117	0.1931	0.4597	0.4169	0.4967	0.5973	1.0000	0.8843	0.6004
fc7	0.1079	0.1097	0.3675	0.3611	0.4373	0.5112	0.8843	1.0000	0.7885
fc8	0.1062	0.1680	0.3181	0.3901	0.4542	0.4516	0.6004	0.7885	1.0000

Fig. 6 Correlations between the representation dissimilarity matrices of the layers of the deep convolutional neural network. All scores are significant

Convolutional layers impose a certain structure on the network's connectivity—each layer consists of a number of visual feature detectors, each dedicated to finding a certain pattern on the source image. Each neuron of the subsequent layer in the convolutional part of the network indicates whether the feature detector associated with that neuron was able to find its specific visual pattern (neuron is highly activated) on the image or not (neuron is not activated). Fully connected layers on the other hand, as the name suggests, connect every neuron of a layer to every neuron in the subsequent layer, allowing for more flexibility in terms of connectedness between the neurons. The training process determines which connections remain and which ones die off. In simplified terms, convolutional layers can be thought of as feature detectors, whereas fully connected layers are more flexible: they do whatever needs to be done to satisfy the learning objective. It is tempting to draw parallels to the roles of lower and higher visual areas in the brain: whereas neurons in lower visual areas (17 and 18) have smaller receptive fields and code for simpler features, neurons in higher visual areas (like 37 and parts of area 20) have larger receptive fields and their activity explicitly represents objects[1,33]. On the other hand, while in neuroscience one makes the broad differences between lower and higher visual

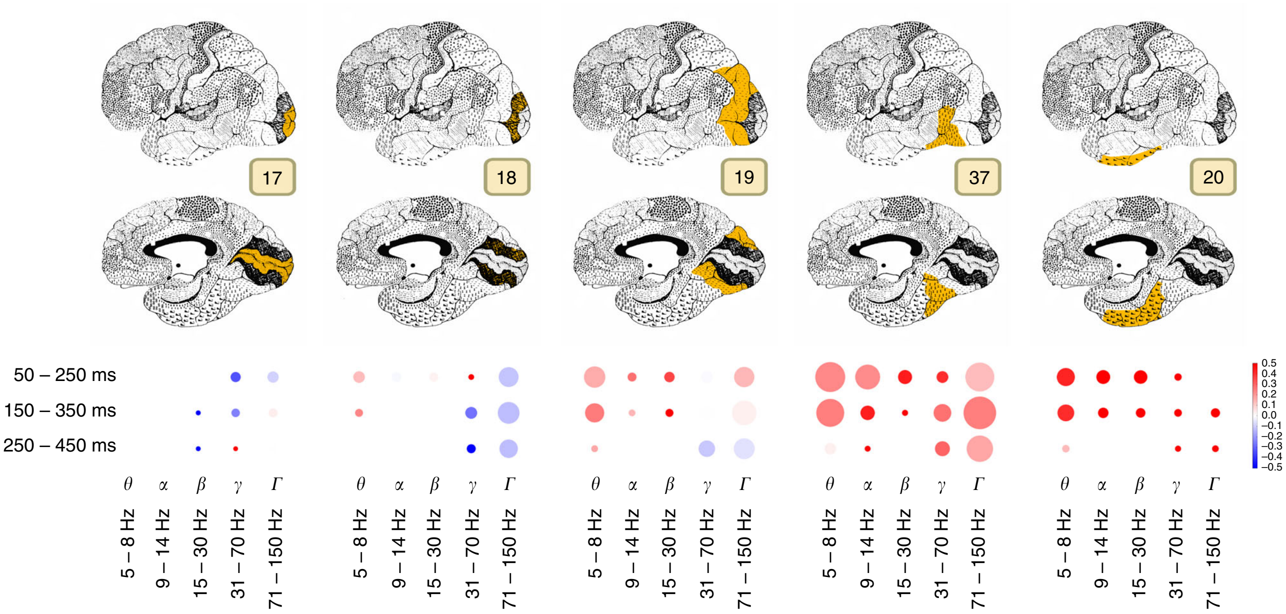


Fig. 7 Area-specific analysis of volume of neural activity and complexity of visual features represented by that activity. Size of the marker shows the sum of correlation coefficients between the area and DCNN for each particular band and time window. Color codes the ratio of complex visual features to simple visual features, i.e., the comparison between the activity that correlates with the higher layers (conv5, fc6, and fc7) of DCNN to the lower layers (conv1, conv2, and conv3). Intense red means that the activity was correlating more with the activity of higher layers of DCNN, while the intense 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

cortex[33] and sensory and association cortices[34], this distinction is not so sharply defined as the one between convolutional and fully connected layers. Our hope is that the present work contributes to understanding the functional differences between lower and higher visual areas.

Visual object recognition in the brain involves both feedforward and feedback computations[1,8]. What do our results reveal about the nature of feedforward and feedback components in visual object recognition? We observed that the DCNN corresponds to the biological processing hierarchy even in the latest analysed time-window (Fig. 4). In a directly relevant previous work Cichy et al.[7] compared DCNN representations to millisecond resolved magnetoencephalographic data from humans. There was a positive correlation between the layer number of the DCNN and the peak latency of the correlation time course between the respective DCNN layer and magnetoencephalography signals. In other words, deeper layers of the DCNN predicted later brain signals. As evidenced in Fig. 3[7], the correlation between DCNN and magnetoencephalographic activity peaked between ca 100 and 160 ms for all layers, but significant correlation persisted well beyond that time-window. In our work too the alignment in low gamma was strong and significant even in the latest time-window 250–450 ms, but it was significantly smaller than in the earliest time-window 50–250 ms. In particular, the alignment was the strongest for low-gamma signals in the earliest time-window compared to all other frequency-and-time combinations.

The present work relies on data pooled over the recordings from 100 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 Supplementary Fig. 1), 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. Also, it is known that some spectral components are affected by the visual evoked potentials (VEPs). In the present experiment we could not disentangle the effect of VEPs from the other spectral responses as we only had one repetition per image. However, we consider the effect of VEPs to be of little concern for the present results as it is known that VEPs have a bigger effect on low-frequency components, whereas our main results were in the low-gamma band.

It must be also noted that the DCNN still explains only a part of the variability of the neural responses. Part of this unexplained variance could be noise[2,4]. Previous works that have used RSA across brain regions have in general found the DCNNs to explain a similar proportion of variance as in our results[6,7]. It must be noted that the main contribution of DCNN has been that it can explain the gradually emerging complexity of visual responses along the ventral pathway, including the highest visual areas where the typical models (e.g., HMAX) were not so successful[3,4]. Recently, it also has been demonstrated that the DCNN provides the best model for explaining responses to natural images also in the primate V1[35]. Nevertheless, the DCNNs cannot be seen as the ultimate model explaining all biological visual processing[8,36]. Most likely over the next years deep recurrent neural networks will surpass DCNNs in the ability to predict cortical responses[8,37].

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. In this study we have investigated the alignment of visual areas with one of the most widely used DCNN architectures—AlexNet. The important step forward would be to compare the alignment with other networks trained on visual recognition task and investigate which architectures preserve the alignment and which do not. That would provide an insight into which functional properties of DCNN architecture are compatible with functional properties of human visual system.

To sum up, in the present work we studied which frequency components match the increasing complexity of representations of an artificial neural network. As expected by previous work in neuroscience, we observed that gamma frequencies, especially low-gamma signals, are aligned with the layers of the DCNN. Previous research has shown that in terms of anatomical location the activity of DCNN maps best to the activity of visual cortex and this mapping follows the propagation of activity along the ventral stream in time. With this work we have confirmed these findings and have additionally established at which frequency ranges the activity of human visual cortex correlates the most with the activity of DCNN, providing the full picture of alignment between these two systems in spatial, temporal and spectral domains.

Methods

Overview. Our methodology involves four major steps described in the following subsections. In "Patients and Recordings" we describe the visual recognition task and data collection. In "Processing of Neural Data" we describe the artifact rejection, extraction of spectral features and the electrode selection processes. "Processing of DCNN Data" shows how we extract activations of artificial neurons of DCNN that occur in response to the same images as were shown to human subjects. In the final step we map neural activity to the layers of DCNN using RSA. See Fig. 1 for the illustration of the analysis workflow.

Patients and recordings. Hundred 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 multilead depth electrodes (DIXI Medical, Besançon, France). The data were bandpass-filtered online from 0.1 to 200 Hz and sampled at 1024 Hz. 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. None of the neurosurgeons who did the operations is among the authors. The authors had no effect on the electrode implantation. The recordings started in 2009, before the present analysis was conceived. All patients had normal or corrected to normal vision.

Eleven to 15 semirigid 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[38], after a linear scale adjustment to correct size differences between the patient's brain and the Talairach model. These locations were further confirmed by overlaying a postimplantation computed tomography 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[39]. 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 the sampling rates of 512, 1024, or 2048 Hz. For the current analysis all recordings were downsampled to 512 Hz. Data were obtained in a total of 11,293 recording sites.

The visual recognition task lasted for about 15 min. Patients were instructed to press a button each time a picture of a fruit appeared on screen (visual oddball paradigm). Nontarget stimuli consisted of pictures of objects of eight possible categories: houses, faces, animals, scenes, tools, pseudo words, consonant strings, and scrambled images. The target stimuli and 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 in Fig. 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 250 natural images. Each image was presented only once. The images were 3.5 × 4.7 cm on the screen, with a viewing distance of 60–80 cm.

Processing of neural data. The final dataset consists of 2823250 local field potential (LFP) recordings—11293 electrode responses to 250 stimuli.

To remove the artifacts the signals were linearly detrended and the recordings that contained values $\geq 10\sigma_{\text{images}}$, where σ_{images} is the standard deviation of responses (in the time window from −500 to 1000 ms) of that particular probe over all stimuli, were excluded from data. All electrodes were re-referenced to a bipolar reference. For every electrode the reference was the next electrode on the same rod following the inward direction. The electrode on the deepest end of each rod was excluded from the analysis. The signal was segmented in the range from −500 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, 150–350, 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[40]. 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[41]. 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 poststimulus 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 the false discovery rate procedure[42]. In the current study we deliberately kept only positively responsive electrodes, leaving the electrodes where the post-stimulus band power was lower than the average baseline power for future work. Supplementary 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.

Each electrode's Montreal Neurological Institute coordinate system coordinates were mapped to a corresponding Brodmann brain area[43] using Brodmann area atlas contained in MRICron[44] 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 TF region of interest.

Processing of DCNN data. We feed the same images that were shown to the test subjects to a DCNN and obtain activations of artificial neurons (nodes) of that network. We use Caffe[45] implementation of AlexNet[46] architecture (see Fig. 5) trained on ImageNet[47] 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 labeling them.

The architecture of the AlexNet artificial network can be seen in Fig. 5. It consists of nine layers. The first is the input layer, where one neuron corresponds to one pixel of an image and activation of that neuron on a scale from 0 to 1 reflects the color of that pixel: if a pixel is black, the corresponding node in the network is not activated at all (value is 0), while a white pixel causes the node to be maximally activated (value 1). After the input layer the network has five *convolutional layers* referred to as conv1–5. A convolutional layer is a collection of filters that are applied to an image. Each filter is a 2D arrangement of weights that represent a particular visual pattern. A filter is convolved with the input from the previous layer to produce the activations that form the next layer. For an example of a visual pattern that a filter of each layer is responsive to, please see Fig. 5b. Each layer consists of multiple filters and we visualize only one per layer for illustrative purposes. A filter is applied to every possible position on an input image and if the underlying patch of an image coincides with the pattern that the filter represents, the filter becomes activated and translates this activation to the artificial neuron in the next layer. That way, nodes of conv1 tell us where on the input image each particular visual pattern occurred. Figure 5b shows an example output feature map produced by a filter being applied to the input image. Hierarchical structure of convolutional layers gives rise to the phenomenon we are investigating in this work—increase of complexity of visual representations in each subsequent layer of the visual hierarchy in both the biological and artificial systems. Convolutional layers are followed by 3 *fully connected* layers (fc6–8). Each node in a fully connected layer is, as the name suggests, connected to every node of the previous layer allowing the network to decide which of those connections are to be preserved and, which are to be ignored. For both convolutional and fully connected layers we can apply *deconvolution*[48] technique to map activations of neurons in those layers back to the input space. This visualization gives better understanding of inner workings of a neural network. Examples of deconvolution reconstruction for each layer are given in Fig. 5b.

For each of the images we store the activations of all nodes of DCNN. As the network has nine layers we obtain nine representations of each 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 the step 2 of the analysis pipeline in Fig. 1 for the cardinalities of those feature spaces.

Mapping neural activity to the layers of DCNN. Once we extracted the 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 of human brain and layers of DCNN. There are multiple deep neural network architectures trained to classify natural images. Our choice of AlexNet does not imply that this particular architecture corresponds best to the hierarchy of visual layers of human brain. It does, however, provide a comparison for hierarchical structure of human visual system and was selected among other architectures due to its relatively small size and thus easier interpretability.

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[2,3]. The second is based on RSA[49]. We used RSA to compare distances between stimuli in the neural response space and in the DCNN activation space[50].

We built a representation dissimilarity matrix (RDM) of size *number of stimuli* × *number of stimuli* (in our case 250 × 250) for each of the probes and each of the layers of DCNN. Note that this is a nonstandard approach: usually the RDM is computed over a population (of voxels, for example), while we do it for each probe separately. We use the nonstandard approach because often we only had 1 electrode per patient per brain area. 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. Note that the preprocessed neural response to an image in a given frequency band and time window is a scalar, and hence correlation distance is not applicable. Also, given that DCNNs are not invariant to the scaling of the activations or weights in any of its layers, we preferred to use closeness in Euclidean distance as a more strict measure of similarity. In our case there are ten different feature spaces in which an image can be represented: the original pixel space, eight 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. For example, to analyze region of interest of high gamma in 50–250 ms time window we computed 504 RDM matrices on the neural responses—one for each positively responsive electrode in that region of interest (see Supplementary Table 1), and nine RDM matrices on the activations of the layers of DCNN. A pair of a frequency band and a time window, such as “high gamma in 50–250 ms window” is referred to as *region of interest* in this work.

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

$$\rho_{\text{layer}\,l}^{\text{probe}\,p} = \text{Spearman}\left(\mathrm{RDM}^{\text{probe}\,p}, \mathrm{RDM}^{\text{layer}\,l}\right). \tag{1}$$

As a result of comparing $\mathrm{RDM}^{\text{probe } p}$ with every $\mathrm{RDM}^{\text{layer } l}$ we obtain a vector with nine scores: $(\rho_{\text{pixels}}, \rho_{\text{conv1}}, \ldots, \rho_{\text{fc8}})$ that serves as a distributed mapping of probe p to the layers of DCNN (see step 5 of the analysis pipeline in Fig. 1). The procedure is repeated independently for each probe in each region of interest. To obtain an aggregate score of the correlation between an area and a layer the ρ scores of all individual probes from that area are summed and divided by the number of ρ values that have passed the significance criterion. The data for the Figs. 2 and 3 are obtained in such manner.

Figure 6 presents the results of applying RSA within the DCNN to compare the similarity of representational geometry between the layers.

To assess the statistical significance of the correlations between the RDM matrices we ran a permutation test. In particular, we reshuffled the vector of brain responses to images 10,000 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 Spearman’s rank correlation scores. To determine score’s significance 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 score obtained on the surrogate sets with $p < 0.001$ significance, we considered the score to be significantly different. The threshold of $p = 0.001$ is estimated by selecting such a threshold that on the surrogate data none of the probes would pass it.

To size the effect caused by training artificial neural network on natural images we performed a control where the whole analysis pipeline depicted in Fig. 1 is repeated using activations of a network that was not trained—its weights are randomly sampled from a Gaussian distribution $\mathcal{N}(0, 0.01)$.

For the relative comparison of alignments between the bands and the noise level estimation we took 1,000 random subsets of half of the size of the dataset. Each region of interest was analyzed separately. The alignment score was calculated for each subset, resulting in 1000 alignment estimates per region of interest. This allowed us to run a statistical test between each pair of regions of interest to test the hypothesis that the DCNN alignment with the probe responses in one band is higher than the alignment with the responses in another band. We used Mann–Whitney U test[51] to test that hypothesis and accepted the difference as significant at p value threshold of 0.005 Bonferroni-corrected[52] to 2.22e−5.

Quantifying properties of the mapping. To evaluate the results quantitatively we devised a set of measures specific to our analysis. *Volume* is the total sum of significant correlations (see Eq. (1)) between the RDMs of the subset of layers **L** and the RDMs of the probes in the subset of brain areas **A**:

$$V_{\text{layers}\,\mathbf{L}}^{\text{areas}\,\mathbf{A}} = \sum_{a \in \mathbf{A}} \sum_{l \in \mathbf{L}} \sum_{p \in \mathbf{D}_l^a} \rho_{\text{layer}\,l}^{\text{probe}\,p}, \tag{2}$$

where, **A** is a subset of brain areas, **L** is a subset of layers, and $\mathbf{S}_l^a$ is the set of all probes in area a that significantly correlate with layer l.

We express *volume of visual activity* as

$$V_{\mathbf{L}=\text{alll ayers}}^{\mathbf{A}=\{17,18,19,37,20\}}, \tag{3}$$

which shows the total sum of correlation scores between all layers of the network and the Brodmann areas that are located in the ventral stream: 17–19, 37, and 20.

Visual specificity of activity is the ratio of volume in visual areas and volume in all areas together, for example visual specificity of all of the activity in the ventral stream that significantly correlates with any of layers of DCNN is

$$S_{\mathbf{L}=\text{all layers}}^{\mathbf{A}=\{17,18,19,37,20\}} = \frac{V_{\mathbf{L}=\text{all layers}}^{\mathbf{A}=\{17,18,19,37,20\}}}{V_{\mathbf{L}=\text{all layers}}^{\mathbf{A}=\text{all areas}}} \tag{4}$$

The measures so far did not take into account hierarchy of the ventral stream nor the hierarchy of DCNN. The following two measures are the most important quantifiers we rely on in presenting our results and they do take hierarchical structure into account.

The *ratio of complex visual features to all visual features* is defined as the total volume mapped to layers conv5, fc6, and fc7 divided by the total volume mapped to layers conv1, conv2, conv3, conv5, fc6, and fc7:

$$C^{\mathbf{A}} = \frac{V_{\mathbf{L}=\{\text{conv5,fc6,fc7}\}}^{\mathbf{A}}}{V_{\mathbf{L}=\{\text{conv1,conv2,conv3,conv5,fc6,fc7}\}}^{\mathbf{A}}}. \tag{5}$$

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 and does not carry visual representations of the stimuli (if only on very abstract level).

Finally, the *alignment* between the activity in the visual areas and activity in DCNN is estimated as Spearman’s rank correlation between two vectors each of length equal to the number of probes with RDMs that significantly correlate with an RDM of any of DCNN layers. The first vector is a list of Brodmann areas $\mathbf{BA}^p$ to which a probe p belong if its activity representation significantly correlates with activity representation of a layer l:

$$\mathbf{A}_{\text{align}} = \left\{\mathbf{BA}^p \mid \forall p \exists l : \rho\left(\mathrm{RDM}^p, \mathrm{RDM}^l\right) \text{is significant according to the permutation test}\right\}. \tag{6}$$

A is ordered by the hierarchy of the ventral stream: BA17, BA18, BA19, BA37, BA20. Areas are coded by integer range from 0 to 4. The second vector lists DCNN layers $\mathbf{L}^p$ to which the very same probes p were assigned:

$$\mathbf{L}_{\text{align}} = \left\{\mathbf{L}^p \mid \forall p \exists l : \rho\left(\mathrm{RDM}^p, \mathrm{RDM}^l\right) \text{is significant according to the permutation test}\right\}. \tag{7}$$

Layers of DCNN are coded by integer range from 0 to 8. We denote Spearman rank correlation of those two vectors as *alignment*

$$\rho_{\text{align}} = \text{Spearman}\left(\mathbf{A}_{\text{align}}, \mathbf{L}_{\text{align}}\right). \tag{8}$$

We note that although the hierarchy of the ventral stream is usually not defined through the progression of Brodmann areas, such ordering nevertheless provides a reasonable approximation of the real hierarchy[32,53]. 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. Based on the recent suggestion that significance levels should be shifted to 0.005[54] and after Bonferroni-correcting for 15 TF windows we accepted alignment as significant when it passed $p < 0.0003(3)$.

Data availability. All raw human brain recordings that support the findings of this study are available from Lyon Neuroscience Research Center but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Raw data are however available from the authors upon reasonable request and with permission of Lyon Neuroscience Research Center. All the preprocessed data are available for download under Academic Free License 3.0 from https://web.gin.g-node.org/ilyakuzovkin/Human-Intracranial-Recordings-and-DCNN-to-Compare-Biological-and-Artificial-Mechanisms-of-Vision.

Code availability. The full code of the analysis pipeline is publicly available at https://github.com/kuz/Human-Intracranial-Recordings-and-DCNN-to-Compare-Biological-and-Artificial-Mechanisms-of-Vision under MIT license.

Received: 27 September 2017 Accepted: 15 July 2018
Published online: 08 August 2018

References

1. DiCarlo, J. J., Zoccolan, D. & Rust, N. C. How does the brain solve visual object recognition? *Neuron* **73**, 415–434 (2012).
2. Güçlü, U. & van Gerven, M. A. Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. *J. Neurosci.* **35**, 10005–10014 (2015).
3. Yamins, D. L. et al. Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proc. Natl Acad. Sci.* **111**, 8619–8624 (2014).
4. Khaligh-Razavi, S.-M. & Kriegeskorte, N. Deep supervised, but not unsupervised, models may explain it cortical representation. *PLoS Comput. Biol.* **10**, e1003915 (2014).
5. Eickenberg, M., Gramfort, A., Varoquaux, G. & Thirion, B. Seeing it all: convolutional network layers map the function of the human visual system. *Neuroimage* **152**, 184–194 (2016).
6. Seibert, D. et al. A performance-optimized model of neural responses across the ventral visual stream. *bioRxiv* https://doi.org/10.1101/036475 (2016).
7. Cichy, R. M., Khosla, A., Pantazis, D., Torralba, A. & Oliva, A. Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. *Sci. Rep.* **6**, 27755 https://www.nature.com/articles/srep27755 (2016).
8. Kriegeskorte, N. Deep neural networks: a new framework for modeling biological vision and brain information processing. *Annu. Rev. Vis. Sci.* **1**, 417–446 (2015).
9. Yamins, D. L. & DiCarlo, J. J. Using goal-driven deep learning models to understand sensory cortex. *Nat. Neurosci.* **19**, 356–365 (2016).
10. Seeliger, K. et al. CNN-based encoding and decoding of visual object recognition in space and time. *bioRxiv* https://doi.org/10.1101/118091 (2017).
11. Singer, W. & Gray, C. M. Visual feature integration and the temporal correlation hypothesis. *Annu. Rev. Neurosci.* **18**, 555–586 (1995).
12. Singer, W. Neuronal synchrony: a versatile code for the definition of relations? *Neuron* **24**, 49–65 (1999).
13. Fisch, L. et al. Neural "ignition": enhanced activation linked to perceptual awareness in human ventral stream visual cortex. *Neuron* **64**, 562–574 (2009).
14. Tallon-Baudry, C., Bertrand, O., Delpuech, C. & Pernier, J. Oscillatory γ-band (30–70 hz) activity induced by a visual search task in humans. *J. Neurosci.* **17**, 722–734 (1997).
15. Tallon-Baudry, C. & Bertrand, O. Oscillatory gamma activity in humans and its role in object representation. *Trends Cogn. Sci.* **3**, 151–162 (1999).
16. Lachaux, J.-P. et al. Measuring phase synchrony in brain signals. *Hum. Brain Mapp.* **8**, 194–208 (1999).
17. Wyart, V. & Tallon-Baudry, C. Neural dissociation between visual awareness and spatial attention. *J. Neurosci.* **28**, 2667–2679 (2008).
18. Lachaux, J.-P. et al. The many faces of the gamma band response to complex visual stimuli. *Neuroimage* **25**, 491–501 (2005).
19. Vidal, J. R., Chaumon, M., O'Regan, J. K. & Tallon-Baudry, C. Visual grouping and the focusing of attention induce gamma-band oscillations at different frequencies in human magnetoencephalogram signals. *J. Cogn. Neurosci.* **18**, 1850–1862 (2006).
20. Herrmann, C. S., Munk, M. H. & Engel, A. K. Cognitive functions of gamma-band activity: memory match and utilization. *Trends Cogn. Sci.* **8**, 347–355 (2004).
21. Srinivasan, R., Russell, D. P., Edelman, G. M. & Tononi, G. Increased synchronization of neuromagnetic responses during conscious perception. *J. Neurosci.* **19**, 5435–5448 (1999).
22. Levy, J., Vidal, J. R., Fries, P., Démonet, J.-F. & Goldstein, A. Selective neural synchrony suppression as a forward gatekeeper to piecemeal conscious perception. *Cereb. Cortex* **26**, 3010–3022 (2015).
23. Fries, P. A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends Cogn. Sci.* **9**, 474–480 (2005).
24. Fries, P. Rhythms for cognition: communication through coherence. *Neuron* **88**, 220–235 (2015).
25. Van Kerkoerle, T. et al. Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. *Proc. Natl Acad. Sci.* **111**, 14332–14341 (2014).
26. Bastos, A. M. et al. Visual areas exert feedforward and feedback influences through distinct frequency channels. *Neuron* **85**, 390–401 (2015).
27. Michalareas, G. et al. Alpha-beta and gamma rhythms subserve feedback and feedforward influences among human visual cortical areas. *Neuron* **89**, 384–397 (2016).
28. Yamins, D. L., Hong, H., Cadieu, C. & DiCarlo, J. J. Hierarchical modular optimization of convolutional networks achieves representations similar to macaque it and human ventral stream. *Adv. Neural Inf. Process. Syst.* 3093–3101 (2013).
29. Gray, C. M. & Singer, W. Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *Proc. Natl Acad. Sci.* **86**, 1698–1702 (1989).
30. Buzsáki, G. & Wang, X.-J. Mechanisms of gamma oscillations. *Annu. Rev. Neurosci.* **35**, 203–225 (2012).
31. Manning, J. R., Jacobs, J., Fried, I. & Kahana, M. J. Broadband shifts in local field potential power spectra are correlated with single-neuron spiking in humans. *J. Neurosci.* **29**, 13613–13620 (2009).
32. Ray, S. & Maunsell, J. H. Different origins of gamma rhythm and high-gamma activity in macaque visual cortex. *PLoS Biol.* **9**, e1000610 (2011).
33. Grill-Spector, K. & Malach, R. The human visual cortex. *Annu. Rev. Neurosci.* **27**, 649–677 (2004).
34. Zeki, S. The visual association cortex. *Curr. Opin. Neurobiol.* **3**, 155–159 (1993).
35. Cadena, S. A. et al. Deep convolutional models improve predictions of macaque v1 responses to natural images. *bioRxiv* https://doi.org/10.1101/201764 (2017).
36. Rajalingham, R. et al. Large-scale, high-resolution comparison of the core visual object recognition behavior of humans, monkeys, and state-of-the-art deep artificial neural networks. *J. Neurosci.* 0388–18, https://doi.org/10.1523/JNEUROSCI.0388-18.2018 (2018).
37. Shi, J., Wen, H., Zhang, Y., Han, K. & Liu, Z. Deep recurrent neural network reveals a hierarchy of process memory during dynamic natural vision. *Hum. Brain. Mapp.* **39**, 2269–2282 (2018).
38. Talairach, J. & Tournoux, P. *Referentially Oriented Cerebral MRI Anatomy: An Atlas of Stereotaxic Anatomical Correlations for Gray and White Matter* (Georg Thieme Verlag, Stuttgart/New York, ISBN 3-13-796701-5 1993).
39. Vidal, J. R. et al. Category-specific visual responses: an intracranial study comparing gamma, beta, alpha, and erp response selectivity. *Front. Hum. Neurosci.* **4**, 195 (2010).
40. Daubechies, I. The wavelet transform, time-frequency localization and signal analysis. *IEEE Trans. Inf. Theory* **36**, 961–1005 (1990).
41. Delorme, A. & Makeig, S. Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *J. Neurosci. Methods* **134**, 9–21 (2004).
42. Genovese, C. R., Lazar, N. A. & Nichols, T. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage* **15**, 870–878 (2002).
43. Brodmann, K. *Vergleichende Lokalisationslehre der Groshirnrinde* (Johann Ambrosius Barth, Leipzig, 1909).
44. Rorden, C. Mricron [Computer Software] (2007).
45. Jia, Y. et al. Caffe: convolutional architecture for fast feature embedding. *arXiv preprint arXiv:1408.5093* (2014).
46. Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems,* Vol. 1, (Curran Associates Inc., Lake Tahoe, NV), pp. 1097–1105 (2012).
47. Russakovsky, O. et al. ImageNet large scale visual recognition challenge. *Int. J. Comput. Vis.* **115**, 211–252 (2015).
48. Zeiler, M. D. & Fergus, R. *Visualizing and Understanding Convolutional Networks*. In: Fleet D., Pajdla T., Schiele B., Tuytelaars T. (eds) Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, Vol. 8689. Springer, Cham.
49. Kriegeskorte, N., Mur, M. & Bandettini, P. A. Representational similarity analysis-connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* **2**, 4 (2008).
50. Cichy, R. M., Khosla, A., Pantazis, D., Torralba, A. & Oliva, A. Deep neural networks predict hierarchical spatio-temporal cortical dynamics of human visual object recognition. *arXiv preprint arXiv:1601.02970* (2016).
51. Mann, H. B. & Whitney, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Stati.* **18**, 50–60 (1947).

52. Dunn, O. J. Multiple comparisons among means. *J. Am. Stat. Assoc.* **56**, 52–64 (1961).
53. Lerner, Y., Hendler, T., Ben-Bashat, D., Harel, M. & Malach, R. A hierarchical axis of object processing stages in the human visual cortex. *Cereb. Cortex* **11**, 287–297 (2001).
54. Dienes, Z. et al. Redefine statistical significance. *Nat. Hum. Behav.* **2**, 6–10 (2017).
55. Yosinski, J., Clune, J., Nguyen, A., Fuchs, T. & Lipson, H. Understanding neural networks through deep visualization. *arXiv preprint arXiv:1506.06579* (2015).

Acknowledgments

We thank Martin Hebart for helpful comments. I.K., R.V. and J.A. thank the financial support from the Estonian Research Council through the personal research grants PUT438, PUT1476 and institutional research grant IUT20-40. This work was supported by the Estonian Centre of Excellence in IT (EXCITE), funded by the European Regional Development Fund. This project has received funding from IHU CESAME (Investissement d'avenir) and the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement No. 720270 (HBP SGA1) and Grant agreement no. 785907 (HBP SGA2).

Author contributions

I.K., R.V. and J.A. designed the research; I.K., M.P., J.P.L., M.B., P.K., S.R. and J.R.V. performed the research and experiments; I.K. analyzed the data and prepared figures; I.K., R.V., J.R.V. and J.A. wrote the manuscript.

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s42003-018-0110-y.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.