# Modulatory feedback determines attentional object segmentation in a model of the ventral stream

Paolo Papale[1,‡,*], Jonathan R. Williford[1,2,‡], Stijn Balk[1] and Pieter R. Roelfsema[1,3,4,5,*]

1. Department of Vision & Cognition, Netherlands Institute for Neuroscience (KNAW), 1105 BA Amsterdam, Netherlands.
2. STR, Woburn, MA, USA
3. Department of Integrative Neurophysiology, VU University, De Boelelaan 1085, 1081 HV Amsterdam, Netherlands.
4. Department of Neurosurgery, Academic Medical Centre, Postbus 22660, 1100 DD Amsterdam, Netherlands.
5. Laboratory of Visual Brain Therapy, Sorbonne Université, INSERM, CNRS, Institut de la Vision, 17 rue Moreau, Paris, F-75012, France

‡ shared first author contribution

* correspondence should be addressed to: *p.roelfsema@nin.knaw.nl* or *p.papale@nin.knaw.nl*

## Abstract

Studies in neuroscience inspired progress in the design of artificial neural networks (ANNs), and, vice versa, ANNs provide new insights into the functioning of brain circuits. So far, the focus has been on how ANNs can help to explain the tuning of neurons at various stages of the visual cortical hierarchy. However, the role of modulatory feedback connections, which play a role in attention and perceptual organization, has not been resolved yet. The present study presents a biologically plausible neural network that performs scene segmentation and can shift attention using modulatory feedback connections from higher to lower brain areas. The model replicates several neurophysiological signatures of recurrent processing. Specifically, figural regions elicit more activity in model units than background regions. The modulation of activity by figure and ground occurs at a delay after the first feedforward response, because it depends on a loop through the higher model areas. Importantly, the figural response enhancement is enhanced by object-based attention, which stays focused on the figural regions and does not spill over to the adjacent background, just as is observed in the visual cortex. Our results indicate how progress in artificial intelligence can be used to garner insight into the recurrent cortical processing for scene segmentation and object-based attention.


## Author Summary

Recent feedforward networks in artificial intelligence provide unmatched models of tuning of neurons in the visual cortex. However, these feedforward models do not explain the influences of object-based attention and image segmentation on neuronal responses, which rely on feedback interactions between cortical regions that are not included in the feedforward networks. In particular, the role of feedback connections from higher brain regions that modulate neural activity

in lower cortical regions has not yet been studied extensively so that we still lack an *in silico* model of the role of these connections. Here, we present a biologically plausible neural network that successfully performs image segmentation and can shift object-based attention using modulatory feedback connections. The model evolved representations that mirror the properties of neurons in the visual cortex, including orientation tuning, shape-selectivity, surround suppression and a sensitivity to figure-ground organization, while trained only on a segmentation task. The new model provides insight into how the perception of coherent objects can emerge from the interaction between lower and higher visual cortical areas.

## Introduction

Objects are the building blocks of our perception. We perceive a visual scene as a layout of objects in the physical space. We tend to focus on one object at a time when scanning a visual scene, either overtly while moving our eyes or covertly by shifting attention (O'Craven et al., 1999). This process, called object-based attention, is essential to group all the visual features that surround us into the objects with which we interact (Behrmann et al., 1998; Duncan, 1984; Kramer et al., 1997; Roelfsema, 2023). When we select a particular object for action, object-based attention is the mechanism that helps identifying the portions of the scene that belong to that object and segregating it from other objects and the background. If subjects grasp and manipulate small objects, attentional selection can occur at the required precise spatial scale (Roelfsema, 2006; Roelfsema and Houtkamp, 2011).

Although it may appear effortless to us, object identification and object-based attention in the primate visual system rely on elaborate computations in a hierarchy of cortical regions. The input of the retina first reaches the primary visual cortex (V1) where neurons have small receptive fields

62 (RFs) and are tuned to simple features. Information is then propagated through feedforward
63 connections to higher-order regions in the occipital cortex (areas V2 and V4) and onward to
64 inferotemporal (IT) and frontal (FC) cortices (van Essen et al., 1992). In the higher areas the neurons
65 become selective to more complex features, such as object parts and object identity and they have
66 larger receptive fields.

67 However, if the task is to segment the scene, the computations carried by feedforward
68 connections are only half of the story (Kar et al., 2019; Lindsay and Miller, 2018; Thorat et al., 2019).
69 After reaching the higher visual areas, visual information is also carried back to the lower areas
70 through feedback connections (Buffalo et al., 2010; Felleman and van Essen, 1991) (Fig 1A). Whereas
71 feedforward connections drive the initial response of neurons and shape neural selectivity,
72 feedback connections modulate the level of neural activity. Feedback connections enhance the
73 neuronal activity elicited by relevant figural regions over the activity elicited by the background, a
74 process called figure-ground modulation (Lamme, 1995; Poort et al., 2016; Roelfsema and de Lange,
75 2016; Zipser et al., 1996). As a result, all the features of objects are enhanced in lower visual brain
76 areas, such as V1, at a high spatial resolution. Furthermore, feedback connections mediate the
77 influence of top-down attention that amplify neuronal activity elicited by task-relevant objects
78 (Roelfsema, 2023, 2006; Spitzer et al., 1988; Vecera and Farah, 1997).

79 At first sight, many of these neurophysiological observations may seem disconnected, because
80 previous studies focused on only one or a few of these feedback influences (Carandini et al., 2005),
81 or used models with connectivity that was largely handcrafted (Craft et al., 2007; Deco and Rolls,
82 2004; Hamker, 2005; Van Der Velde and De Kamps, 2001). We therefore sought to integrate
83 previous findings into a coherent framework, capitalizing on the recent advances in the
84 development of artificial neural networks (ANNs) (Echeveste et al., 2020; Fox et al., 2023; Serre,
85 2019).

86 State of the art ANNs can achieve human level performance on many complex visual tasks,
87 ranging from object classification to segmentation (Kirillov et al., 2023; Serre, 2019). For example,
88 breakthroughs were achieved in the problem of semantic segmentation, where the goal is to
89 determine the pixels that belong objects of specific categories (Kirillov et al., 2023; Li et al., 2018).
90 ANNs for semantic segmentation have features in common with segmentation processes in the
91 brain, by relying on two pathways. The first pathway starts with units representing pixels and end
92 with units coding for object categories through a hierarchy of layers with increasing RFs and more
93 complex tuning. The second pathway propagates activity in the opposite direction, from object
94 category back to pixels, so that tuning becomes simpler and RF size decreases during successive
95 processing steps (Hong et al., 2015). The selectivity of units in the feedforward pathway of ANNs,
96 resembles the selectivity of neurons in the visual brain (Cadena et al., 2019; Guclu and van Gerven,
97 2015; Yamins et al., 2014). However, the resemblance of the ANN feedback pathways to
98 neurobiology is more limited. Firstly, in most ANNs units of the feedback pathway are independent
99 from those of the feedforward pathway (e.g., Hong et al., 2015), whereas many neurons in the brain
100 are influenced by both feedforward and feedback connections (e.g., Lei et al., 2021). Secondly, in
101 the visual cortex, neurons that are influenced by feedback and those that are mainly feedforward
102 tend to segregate into different layers of the cortical column. The feedback influences are weak in
103 cortical layer 4, which is the input layer of cortex, and stronger in the superficial and deep layers
104 (Keller et al., 2020; Kirchberger et al., 2023; Self et al., 2013; van Kerkoerle et al., 2017). Interestingly,
105 neurons with different degrees of sensitivity to feedback effects reside in the same column and have
106 a similar tuning. Thirdly, the feedback pathway of ANNs drives the units, whereas feedback
107 connections in the brain are mostly modulatory. This means that feedback connections usually do
108 not directly activate the neurons but amplify or decrease the feedforward response. Some of these
109 feedback influences are multiplicative, with a stronger influence on neurons that are well driven by

feedforward input than on neurons that are not (Poort et al., 2016, 2012; Roelfsema et al., 2002; Self et al., 2012; Treue and Martínez Trujillo, 1999).

Hence, the brain processes underlying image segmentation and, in particular, semantic segmentation have remained unclear. How do the feedback pathways of the brain, which are modulatory and intermingled with the feedforward pathways, achieve object segmentation? To answer these questions, we built a biologically inspired ANN to examine figure-ground segregation, semantic segmentation and the role of selective attention (Poort et al., 2012). In our model, feedforward and feedback influences occur within the same cortical column and feedback is modulatory (Keller et al., 2020; Self et al., 2013; van Kerkoerle et al., 2017). The network comprised a hierarchy of areas modeled after the hypothetical brain regions involved in object processing – denoted here with an 'm' suffix – from V1m to FCm (Fig 1B). We included a feedforward pathway, driving the activity in successively higher areas for object identification, and a feedback pathway that modulated neuronal activity (Fig 1C,D) to highlight the figural image regions if they were attended.

The model produced an accurate segmentation of attended objects by increasing the firing rate elicited by image elements in V1m, representing the relevant figure at a high spatial resolution, whereas the activity elicited by nearby background elements was suppressed, just as is observed in the visual cortex of monkeys (Poort et al., 2012; Self et al., 2019). When we examined the tuning of model units we observed similarities to the neurophysiology, indicating that the model enables new insights in how feedforward and feedback connections jointly determine the activity of neurons in the visual cortex.

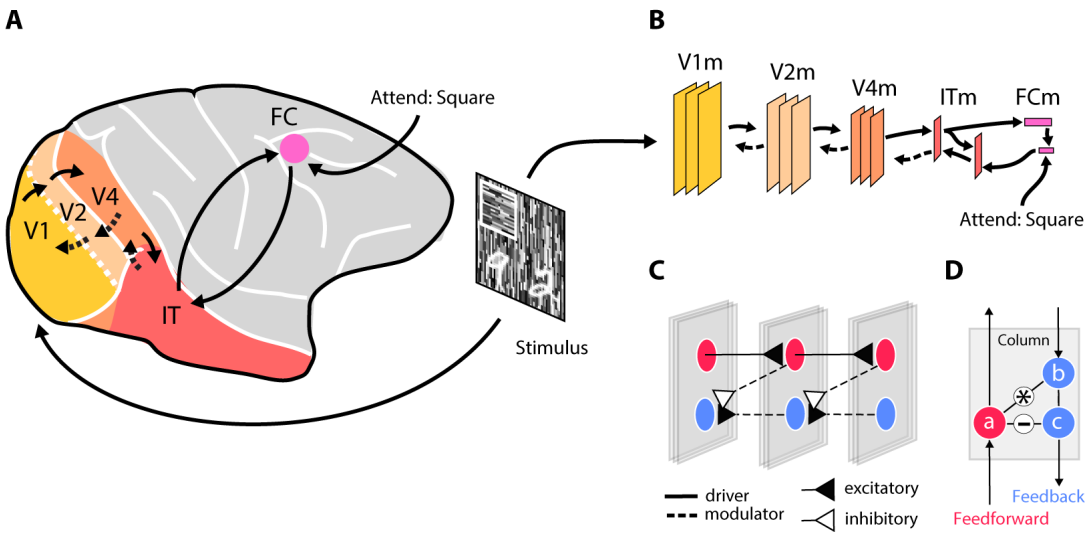


**Fig 1. A neural network to model the primate visual cortex.**

A) Hierarchically organized visual cortical areas that perform increasingly complex transformations of the visual input. V1 registers the information coming from the retina. Downstream regions (V2, V4) process information further and object-selective neurons in IT categorize the objects in a scene and send this information to the frontal cortex (FC) that represents behavioral goals. The frontal cortex sends attentional signals through feedback connections that modulate the activity of earlier brain regions, resulting in the segmentation of task-relevant objects at a high spatial resolution. B) In the model, the stimulus enters in area V1 (V1m) and feedforward connections (solid arrows) activate downstream regions that classify the objects. The network can select one of the object classes and use feedback connections for class-specific object segmentation in V1m. C) The pattern of connections between V1m, V2m and V4m. The network contains units that propagate information in the feedforward direction (red) and other units (blue) that propagate information in the feedback direction. The signal flowing in the feedback direction is based on a comparison (subtraction) of the activity of feedback and feedforward units. D) Schematic of the computations within one of the layers. *a* is the activity of feedforward unit, which propagates activity to the next higher layer, gates information flow of unit *b* (a multiplicative influence; *) and inhibits *c* (-) of the feedback pathway.

## Materials and Methods

### *Model Architecture*

Figure Fig 1B provides an overview of the model, which consists of several model areas, which loosely reproduce the ventral stream of the visual cortex of the primate. Our architecture was inspired by an ANN (Hong et al., 2015) that performs semantic segmentation in separate feedforward and feedback pathways, i.e. its goal is to assign an object class to pixels of the input

image (Fig. 2A). The architecture introduced by Hong et al. (2015) achieved this by first mapping pixels into object classes in a feedforward network. It then selected one of the classes at a time for a given image and retrieved the pixels belonging to that class in an additional feedforward network that mapped classes back onto pixels. Hence, this ANN carried out an object-based attention task in a purely feedforward network. Our model, instead, used feedforward connections that are driving and feedback connections that are modulatory. It can therefore provide insight into the different phases of activity of neurons in the visual cortex.

The model comprised two main processing streams, a feedforward pathway that drives the neurons and a feedback pathway that is modulatory. The feedforward pathway took an image as input and produced a classification vector that indicated the object classes that were detected in the image as output. This feedforward pathway was not modulated by feedback, akin to what observed in the input layer 4 of area V1 (Keller et al., 2020; Self et al., 2013; van Kerkoerle et al., 2017). The feedback pathway contained units that were influenced by the feedforward units in the same area and that were also modulated by feedback connections, as will be detailed below. The feedback pathway started with the classification vector at the highest levels of the feedforward pathway, which it combined with a one-hot attention vector that specified a target category (e.g. “attend the texture square” in Figure 1A). The goal of the feedback pathway was to highlight all the low-level features elicited by the target object with enhanced neuronal activity.

*Feedforward pathway*

The network layers were grouped into areas that roughly correspond to visual cortical areas. All the layers within a given area had the same size. Areas V1m, V2m and V4m had 20 features and 56x56 pixels, 50 features and 28x28 pixels, and 100 features and 14x14 pixels, respectively, resulting in larger RFs in higher areas (i.e. 2x2 pixels in V1m, 4x4 pixels in V2m and 8x8 pixels in V4m). Each

181 area had three convolutional layers with skip connections bypassing the convolution followed by
182 batch normalization and a ReLU function. The output of each area was downsampled by a strided
183 convolution (stride of 2 and a 2x2 kernel size). The down-sampled output from V4m (100 features
184 by 7x7 pixels) was passed to a fully connected layer (ITm: 500x1x1), which was then passed to
185 another fully connected layer (FCm: 12x1x1), which had one unit for each of the 12 object classes in
186 our task. This layer used a sigmoid activation function (i.e. softmax) to estimate the likelihood that
187 the various classes appeared in the stimulus. Multiple object classes could appear at the same time.

188

189 *Feedback Pathway*

190 The interactions between the feedforward and feedback pathway ensured that the feedback
191 from the selected shape in FCm modulated the appropriate lower-level units. A 12x1x1 layer
192 encoded the object that should be attended (the "attend" label in Fig. 1B). This attention signal was
193 multiplied (elementwise) with the classification vector from the feedforward pathway. A series of
194 fully connected layers were used in the feedback pathway, mirroring the feedforward pathway,
195 followed by a "deconvolution" (transposed convolution) layer that produced output with
196 dimensions of 100x7x7. This output was then concatenated with the corresponding 100x7x7 layer
197 of the feedforward to produce a 200x7x7 tensor, which was passed to a 1x1 convolution reducing
198 the tensor back to 100x7x7.

199 Each unit in the feedforward pathway had a corresponding (deconvolutional) unit in the
200 feedback pathway in V1m, V2m and V4m, which can be conceived of as being part of the same
201 cortical column (Fig. 1D). The feedback connections did not drive units in these areas, but they
202 modulated the activity, in accordance with the effects of attention in the brain (e.g., Self et al.,
203 2012). In our model the modulatory feedback interaction was implemented as:

$$b_i = max\big(0, a_i(z_{i+1} + \beta_i)\big)$$

where $a_i$ is the non-negative firing rate of a feedforward neuron in layer $i$, $z_{i+1}$ is the feedback received from a higher layer via a strided convolution, $\beta_i$ is a learned parameter, and $b_i$ is the activity of a unit of the feedback pathway (Fig. 1D, Fig. 2B). The feedback signal that is propagated to the next lower layer, is the difference between the activity of the feedforward and feedback units and it constitutes a relatively pure segmentation signal:

$$c_i = b_i - a_i$$

*Stimulus sets*

Based on the MNIST dataset (Yann LeCun et al., 2010), we created an augmented stimulus set, *SegMNIST* (Fig. 2A). We followed recent approaches to make MNIST-like stimulus sets more challenging (Ernst et al., 2019; Michaelis et al., 2018; Spoerer et al., 2017; Thorat et al., 2021) and casted it as an object localization task, where the relevant shape could appear at various locations in the image. The size of the image was increased to twice the size of the MNIST image (56x56 pixels instead of 28x28) and MNIST digits were translated, scaled, and positioned on top of a texture. To gain insight in figure-ground segregation processes, we also added textured squares and rectangles at different rotation angles with an orientation that differed from the background , which have been used in previous monkey experiments on texture segregation (Poort et al., 2012; Self et al., 2012)

*Training of the feedforward and feedback pathways*

The feedforward pathway was trained to classify objects in the textured and expanded SegMNIST stimuli using Adam (Kingma and Ba, 2015) with a learning rate of 0.001, L2-regularization (weight decay = 0.0005), and as loss a sigmoid combined with binary cross-entropy (pytorch's

227 *BCEWithLogitsLoss*) in 2,500 epochs (batch size = 1024). There were 12 object classes; a square, a
228 rectangle, and the 10 digits, and there were one to three objects in each image. The task of the
229 network was to determine which object classes were present in the image. We trained the feedback
230 pathway once the weights of the feedforward pathway were fixed. The feedback pathway was
231 trained on object segmentation using Adam (Kingma and Ba, 2015) with learning rate of 0.001, L2-
232 regularizations (weight decay = 0.0005), and as loss a sigmoid combined with binary cross-entropy
233 (pytorch's *BCEWithLogitsLoss*) in 2,500 epochs (batch size = 256). The objective was to maximize the
234 number of pixels in the image that were correctly classified (with activity 1 for pixels belonging to
235 the attended class and 0 elsewhere). Because of the imbalance in the number of pixels in objects
236 and the background, we ignored background pixels when computing the error, because it would
237 have biased the error and led to poor generalization. The network was trained using one GPU.

238

239 *Simulating the activity propagation in V1m*

240 To evaluate the influence of attentional selection on the activity across time, we allowed the
241 feedforward activity to propagate to layers in successive time steps. The initial feedback modulation
242 could proceed in lower layers while the feedforward activity was propagating to the higher layers
243 of the network (Fig. 3). We simulated a single feedforward pass and a single feedback pass but did
244 not model full recurrence. Our simulations address the situation in which the subject aims to select
245 (attend) a specific shape. We initiated the network at the activity level elicited by a white noise
246 stimulus and activated the attentional template of the to-be-selected shape in FCm. We then
247 presented the stimulus and the activity propagated to the next higher and lower layer on every time
248 step. We presented 1,000 stimuli (with the model attending to either figures or digits) and averaged
249 the responses. The time-courses shown in Figure 3B,D were up-sampled (to 160 steps) and
250 smoothed with a 3$^{rd}$ order Savitzky–Golay filter (7 timestep window) for visualization purposes.

251

252 *Code availability*

253 Model and analyses were implemented using custom Python code and NumPy (van der Walt
254 et al., 2011), SciPy (SciPy.org) and PyTorch (Paszke et al., 2019). The code to generate the stimuli is
255 available at the SegMNIST Github repository: http://github.com/williford/seg-mnist-dataset. The
256 code to train the model and reproduce the results is available at the Github repository of the project
257 [available upon paper acceptance]. A Google Colab of the project with the model results and all
258 units is available at https://colab.research.google.com/drive/13O0-4uYq3l1tZupVn-NLP8C-
259 WvakajCG?usp=sharing.

260

261 **Results**

262 *Modulatory feedback connections achieve accurate segmentations of attended objects*

263 We trained an ANN to recognize and segment the digits 0 through 9, textured squares and
264 rectangles in a display with a background of texture elements with an orientation that differed from
265 the foreground squares and rectangles. First, we trained the feedforward path of the network to
266 recognize these shapes. Then we trained the feedback pathway to attend and segment one of the
267 shapes from the background. In this training phase, we used a task that was inspired by Poort et al.
268 (2012), who trained monkeys to either detect texture-based squares similar those of Fig. 2 or to
269 perform another task in which they could ignore these texture-defined figures.

270 A few example stimuli and the model's segmentation results are shown in Fig. 2A. The stimuli
271 (left column) were particularly challenging because the edges of the digits blended with the lines of
272 the texture in the background and the squares and rectangles were defined by a difference in the
273 orientation of the texture elements. Despite these challenges, the classification accuracy of

274 feedforward path of the model was nearly perfect. Specifically, the activation of the FCm units
275 reflected the objects classes with a mean cross-validated absolute error of 0.023 (this measure
276 ranges between 0 when all objects are correctly classified in each image and 12 if none of the objects
277 are correctly classified).

278 We next examined if and how neurons of the feedback pathway were influenced by the
279 presence of figure-ground stimuli and the shape that had been selected in FCm. The feedback
280 pathway caused highly accurate segmentations. Responses of feedback units to images regions of
281 the attended category were enhanced over the responses of units that responded to the
282 background and over the responses elicited by non-attended objects (yellow regions in Fig. 2A
283 denote attended regions, which have been labeled with enhanced activity; blue regions are labeled
284 as background). Consider, for example, the left stimulus in the second row of Fig. 2A that contained
285 a square and two digits, a '5' and a '6'. When the square was attended, by switching on the
286 corresponding unit in FCm, the activity of V1m units with receptive fields on the square was
287 enhanced. This response enhancement was confined to the square at a high spatial precision and it
288 did not spill over into the background. However, if attention was directed to one of the digits, the
289 responses to the square were weaker and the response enhancement was confined to the pixels of
290 the appropriate digit. These highly accurate segmentations occurred for most of the images, so that
291 the mean cross-validated pixel-wise absolute error was 0.0015 (the error ranges between 0 when
292 all object pixels correctly segmented and 1 if none of the object pixels is correctly segmented in each
293 image; this measure does not consider background pixels). Furthermore, the proportion of
294 background elements that were erroneously labeled as the attended class was very small (<0.1%).

295 These results demonstrate that modulatory feedback connections can explain the selective
296 labeling of low-level image elements of an attended object category with enhanced neuronal
297 activity, just as is observed in the primary visual cortex of monkeys (Lamme, 1995; Poort et al., 2012;

Self et al., 2019; Zipser et al., 1996). When we examined the tuning of model units to shape and orientation, we also observed strong similarities to the properties of neurons in the visual cortex of monkeys, and the same held true for surround suppression (Fig. S1,S2). Interestingly, tuning to the shape of objects was stronger in the feedback pathway in V4m and V2m than in the feedforward pathway, indicating that feedback connections carried information about the identity of the attended objects from higher regions (ITm and FCm) to lower regions (Fig. S2).

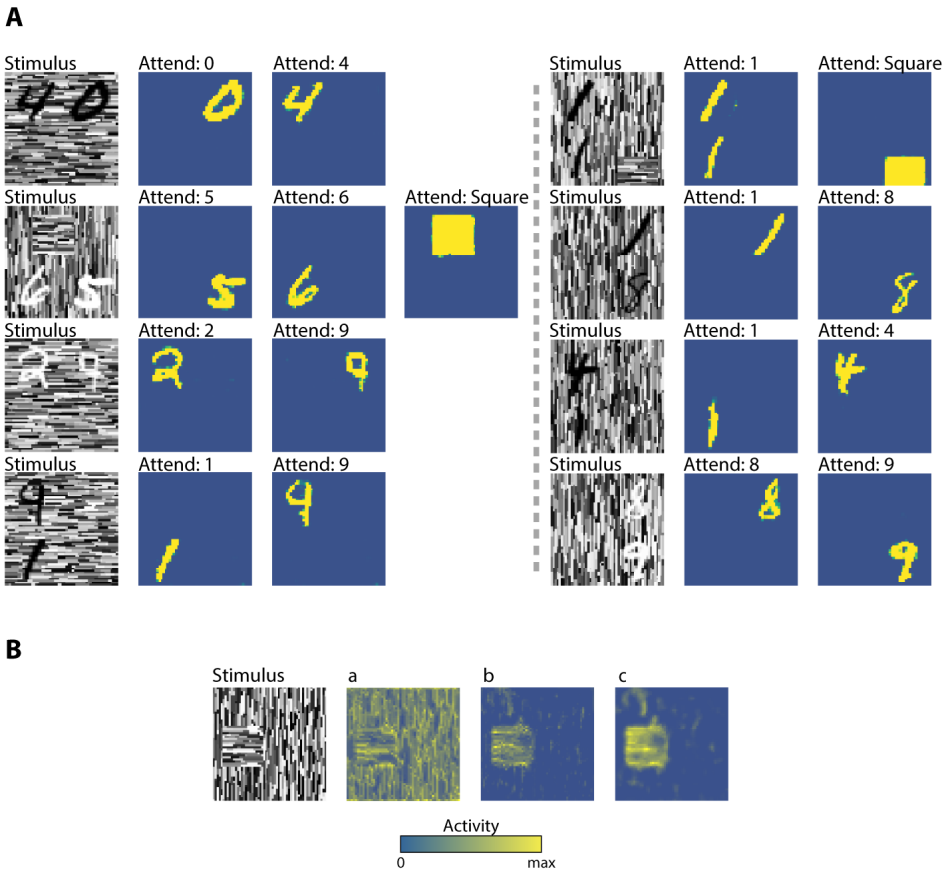


**Fig 2. Modulatory feedback achieves accurate segmentations of attended objects in cluttered displays.**

A) Eight example stimuli and the segmentations achieved by the model. Each row shows an example stimulus and the segmentation generated by the model when instructed to attend to a specific object class (blue, low activity level denoting background; yellow, high activity level labeling the segmented object). The segmentations in V1m are accurate, demonstrating that modulatory feedback connections enable a selective segmentation of the attended object category. B) The internal representation of an example stimulus (left), show how the activity of V1m feedforward units "*a*" (middle left), is subtracted from the modulatory feedback "*b*" (middle right) coming from a downstream region, to generate the final modulated response "*c*" (right).

316 We next compared the time-course of the activity of model units to the neurophysiological
317 results of Poort et al. (2012). In that study, monkeys either attended texture-defined square figure,
318 similar to the stimuli shown in Fig. 2A, or performed another task in which they ignored the square.
319 The main neurophysiological results are summarized in Fig. 3A, showing the time-course of the
320 response of V1 neurons with a RF on the center of the square figure (red trace), the edge between
321 the square and the textured background (black trace) or the background (blue trace). The activity
322 started around 40ms after the presentation of the stimulus, when feedforward activations drive the
323 V1 neurons. During this early response phase, the activity depends on the information in the RF,
324 and after a short delay the activity of neurons with a RF on the edge between figure and ground is
325 enhanced over the activity of neurons with an RF on the background. During a later response phase,
326 around 100ms after the stimulus onset, the activity elicited by image elements that belong to the
327 center of the figure start to be also enhanced over the activity elicited by background elements.
328 Finally, approximately 200ms after onset, the influence of attention to the figure (Fig. 3A, top) kicks
329 in. Attention to the figure causes a further enhancement of the activity of V1 neurons with a RF on
330 figure region, compared to when monkeys ignore the figure (Fig. 3A, note that the separation
331 between black and red traces is more pronounced in the lower panel).

332 We examined the activity of V1m units in a version of the model in which we propagated
333 activity in the feedforward and feedback directions by one connection at a time. The activity profile
334 of V1m units was similar to the activity of neurons in monkey V1. The model units exhibited a first,
335 transient peak response that was elicited by the appearance of texture elements in the RF (Fig. 3B).
336 The early response elicited by the edge (black in Fig. 3B), was higher than that elicited by the figure
337 and ground and the response elicited by the figure was stronger than that elicited by the
338 background. At a later phase, attention to the textured square enhanced the activity evoked by the
339 figure further (Fig. 3B, top), compared to when attention was directed to one of the digits (Fig. 3B,

bottom). The influence of attention on neuronal activity in the model feedback pathway was stronger than in monkey V1 and activity elicited by the center of the figure exceeded the activity elicited by the edge. This difference between the model and the neurophysiological results is explained by the model's training objective, which enforced binary segmentation where figural regions reach a target activity of 1. However, it is of interest that the order of the effects of figure-ground organization and attention was similar to that in monkey V1.

Figure 3C,D compares the spatial profile of the figure-ground modulation (FGM), which is difference in activity elicited by the figure and the background (subtraction of the blue curve from the red/black curve in Fig. 3A,B), in monkey V1 (Fig. 3C) and in the feedback pathway in model area V1m (Fig. 3D). The monkeys saw a figure with a size of 4 degrees and the size of the V1 RFs was less than 1 degree. In the experiment with monkeys the location of the figure was varied so that the neurons' RFs could either fall on the figure, the background or the edge. It can be seen that FGM was confined the figure representation. If the monkey did not attend the figure, FGM was stronger at the boundary between figure and background than in the center of the figure. If the monkey attended the figure, however, FGM in the center of the figure was more pronounced and reached the same level as that at the boundary.

Many of these observations also held true for the feedback units in V1m of the model, where FGM was also confined to the figure. If attention was directed to one of the digits, FGM was strongest at the boundaries of the square and weaker in the center of the figure. However, if the model attended the textured square, activity was strong in the center of the figure. In the model, the activity reached a higher plateau, which is related to the training procedure in which the target activity level in V1m was 1, as was described above.

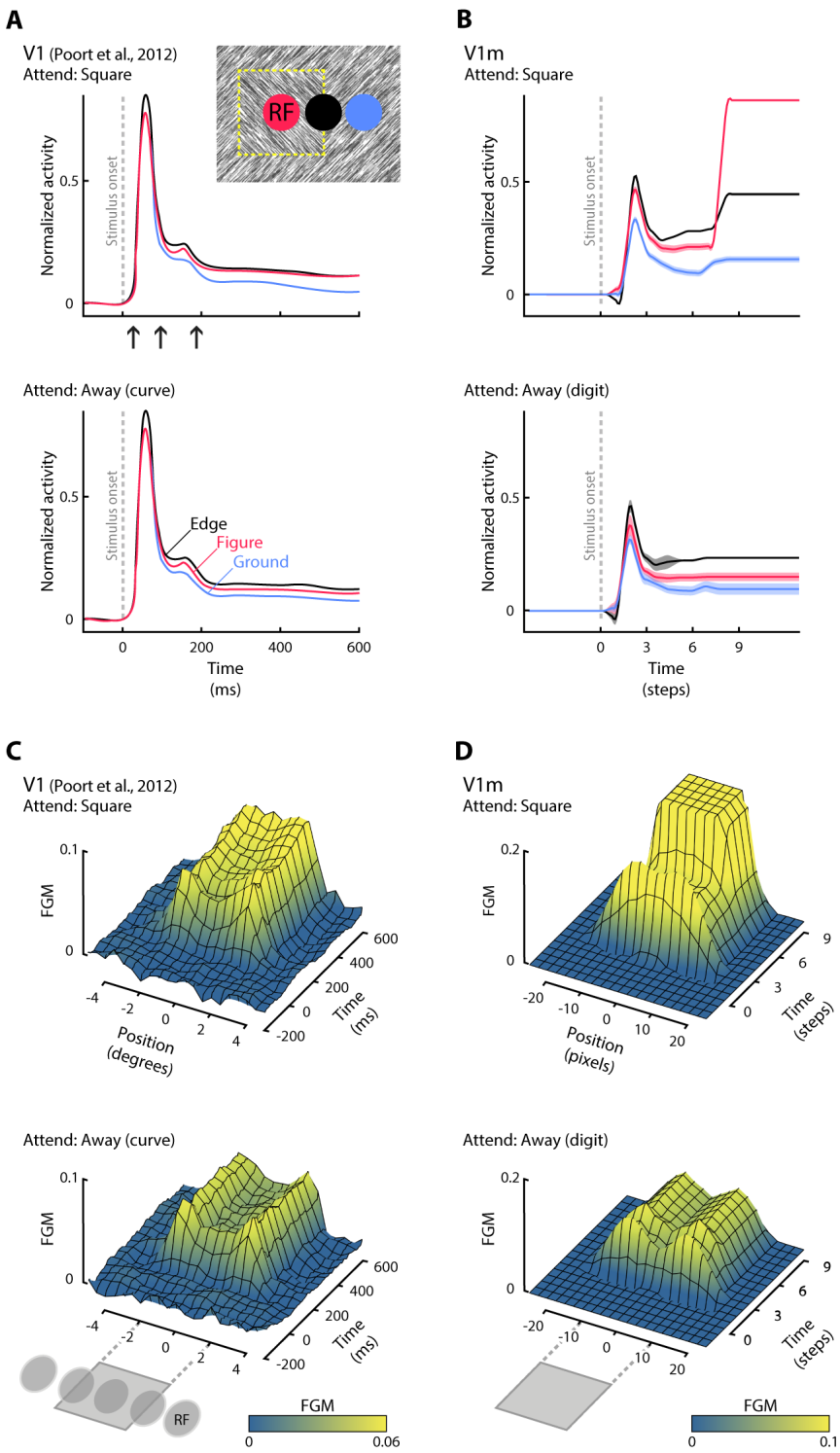


**Fig 3. V1m evolves figure-ground modulation by selective attention as observed in monkey V1**

A) Activity of V1 neurons in the study of Poort et al. (2012), in which monkeys were cued to either detect texture-based square figures (top) or to perform another task in which they did not attend the figures. V1 activity depended both on whether the texture elements falling in RF were part of the figure, the ground or the edge between figure and ground and also on attention. There were three different phases of activity: a first transient response (top, left arrow), a later phase in which figures elicited more activity than grounds (middle arrow) and an even later phase in which attention increased the figure response (right arrow; compare the distance between red and blue

curves in the upper and lower panels). B) Response of feedback units in V1m when attention is directed to the figure (top) or to one of the digits (bottom). The activity was propagated for one connection at a time. Dashed regions represent SEM (across 1,000 stimuli with different textures). C) Space-time profile of figure-ground modulation (FGM: difference between the figure and background response) for V1 neurons when the monkey attended to the figure (top) and or attended away (bottom). The horizontal axis represents the position of the RF, which either fell on the figure, the edge or the background. D) Space-time profile of V1m units when the model attended the square (top) or one of the digits (bottom).

## Discussion

In this work, we built a biologically inspired ANN, implementing a model of the primate object processing pathway, which was composed of a driving feedforward pathway and a modulatory feedback pathway. We demonstrated that this simple architecture could shift selective attention to all the low-level features that belong to the same object, even it is embedded in a cluttered texture stimulus. We compared the activity of model units to the responses of neurons in the visual cortex of monkeys. Firstly, the model reproduced the spatial profile of neuronal activity in the primary visual cortex during image segmentation tasks, with a strong enhancement of the representation of image elements of figures relative to the background, even though the image elements of figure and ground were, on average, the same. Secondly, the time-course of activity of model units in V1m resembled the time-course of the response of V1 neurons. Activity started with a transient response, followed by figure-ground modulation and the influence of attention was expressed at an even later time-point, because feedforward processing had to propagate to FCm, which then had to reach back to V1m. The delayed attentional influence is in accordance with neurophysiological evidence about comparable top-down influences (e.g., Bichot et al., 2019, 2015; Moore and Fallah, 2004; van Vugt et al., 2018). However, the attentional influence in V1m was stronger than in monkey V1. A likely reason for this difference is that our model was trained on a binary segmentation task with a high target value of attended pixels, whereas V1 of animals contributes to many more tasks (Roelfsema and de Lange, 2016), which is a limitation of our model.

399 Thirdly, we found qualitative similarities between the properties of neurons in the visual cortex and
400 in the model, including orientation tuning, surround suppression and shape-selectivity (Figs. S1,S2).
401 It is of interest that tuning to object identity was strongest in the feedback pathway in V4m and
402 V2m, indicating that it was based on feedback from the higher model areas (Fig. S2). These results
403 thereby provide insight why object classes can be decoded from the activity of early visual neurons
404 in the primate brain (e.g., Muckli et al., 2015; Papale et al., 2023, 2021, 2020, 2019; Williams et al.,
405 2008).

406 Previous modeling studies addressing the neuronal mechanisms of image parsing and
407 texture segregation in the visual cortex used hand-crafted connectivity schemes and activation
408 functions (Poort et al., 2012; Roelfsema et al., 2002). The present model went beyond these
409 previous studies by using harder segmentation tasks and a neural network training scheme with a
410 loss function for the segmentation outcome. The training procedure resulted in internal
411 representations that resemble, to some extent, those of the primate brain, without the need to
412 handcraft them. A related approach (Lei et al., 2021), carried out semantic segmentation of digits of
413 the MNIST data set. Our work goes one step further by examining texture-defined figure-ground
414 stimuli, and by implementing a biologically plausible form of top-down feedback. Indeed, in our
415 model feedback was modulatory, because the activity was gated by feedforward activation and the
416 model then compared the feedback activity to the feedforward activation, to determine the
417 magnitude of the segmentation signal (Figure 1D). Our model was able to segregate figures from
418 the background based on the difference in the orientation of texture elements and illustrated how
419 top-down attention amplifies figure-ground signals, just as is observed in the visual cortex of
420 monkeys (Poort et al., 2012). Furthermore, we evaluated the effect of feedback connections on the
421 tuning of units, providing potential novel predictions, which can be tested in future
422 neurophysiological experiments, e.g. the increase in shape selectivity and surround suppression due

423 to feedback. A limitation of the present approach is that the network only used a feedforward and
424 a feedback pass, but did not include full recurrence, even though the segmentation results are likely
425 to further improve object-recognition, which in turn, may further refine the segmentation results,
426 especially for shallower networks (Seijdel et al., 2020). Future research is needed to close this gap.

427 Our work focused on the relation between ANNs and mechanisms for image parsing in the
428 brain. The aim therefore differed from that of ANN studies that focus on improving the state-of-the-
429 art using methods without considering biological plausibility. Knowledge of brain-like mechanisms
430 for vision may inspire improvements in the performance of ANNs in several ways. For example, the
431 inclusion of feedback connections in ANNs does not only increase the similarity with the brain
432 computations (Kietzmann et al., 2019; Kubilius et al., 2018), but improves classification accuracy and
433 noise resistance (Jarvers and Neumann, 2019; Spoerer et al., 2020, 2017; Yan et al., 2019). Similarly,
434 the inclusion of horizontal connections in the network, which are present between neurons in the
435 visual cortex, aids in solving complex perceptual grouping tasks (Kim et al., 2019; Linsley et al., 2020,
436 2018). Hence, brain processes can inspire improvements in ANNs, levering an intense cross-
437 fertilization between neuroscience and artificial intelligence (Doerig et al., 2022).

438 Overall, we provided a proof-of-concept demonstration that modulatory feedback are
439 sufficient to produce accurate object segmentations of attended shapes. Our model represents a
440 minimal architecture for object parsing in the brain, based on what we know about the roles of
441 feedforward and feedback connections (Keller et al., 2020; Self et al., 2013; van Kerkoerle et al.,
442 2017). We expect that future models will further improve our understanding of these mechanism,
443 for example for more ecological stimuli to elucidate behavioral (Fowlkes et al., 2007; Neri, 2017,
444 2014; Zuiderbaan et al., 2017) and neural (Hesse and Tsao, 2016; Papale et al., 2018; Williford and
445 von der Heydt, 2016) signatures of object segmentation in natural scenes. Future studies might also

expand the biologically plausibility of neuronal networks for object-based attention by including horizontal connections and brain-like learning rules (Pozzi et al., 2020).

**Acknowledgments**

We thank Jasper Poort for providing the code and pre-processed data from his experiment, and Alexander van Meegen for his early contribution to the model implementation. This work was supported by NWO (NWO-OCENW.KLEIN.178, Crossover Program 17619 “INTENSE” and “DBI2”, a Gravitation program of the Dutch Ministry of Science, Education and Culture; Open Competition Domain Science – XS, number OCENW.XS22.2.097), the European Union FP7 (ERC grant number 339490 ‘Cortic_al_gorithms’), Horizon Europe (ERC advanced grant 101052963 “NUMEROUS”; grant agreement 899287 “NeuraViper”), the Human Brain Project (agreement no. 945539, “Human Brain Project SGA3”) and *Agence Nationale de la Recherche* (ANR) within *Programme Investissements d’Avenir, Institut Hospitalo-Universitaire* FOReSIGHT (ANR-18-IAHU-0001).

## References

Behrmann M, Zemel RS, Mozer MC. 1998. Object-based attention and occlusion: Evidence from normal participants and a computational model. *J Exp Psychol Hum Percept Perform* **24**:1011–1036. doi:10.1037//0096-1523.24.4.1011

Bichot NP, Heard MT, DeGennaro EM, Desimone R. 2015. A Source for Feature-Based Attention in the Prefrontal Cortex. *Neuron* **88**:832–844. doi:10.1016/j.neuron.2015.10.001

Bichot NP, Xu R, Ghadooshahy A, Williams ML, Desimone R. 2019. The role of prefrontal cortex in the control of feature attention in area V4. *Nat Commun* **10**:1–12. doi:10.1038/s41467-019-13761-7

Buffalo EA, Fries P, Landman R, Liang H, Desimone R. 2010. A backward progression of attentional effects in the ventral stream. *Proc Natl Acad Sci U S A* **107**:361–365. doi:10.1073/pnas.0907658106

Cadena SA, Denfield GH, Walker EY, Gatys LA, Tolias AS, Bethge M, Ecker AS. 2019. Deep convolutional models improve predictions of macaque V1 responses to natural images. *PLOS Comput Biol* **15**:e1006897. doi:10.1371/journal.pcbi.1006897

Carandini M, Demb JB, Mante V, Tolhurst DJ, Dan Y, Olshausen BA, Gallant JL, Rust NC. 2005. Do we know what the early visual system does? *J Neurosci* **25**:10577–10597. doi:10.1523/JNEUROSCI.3726-05.2005

Craft E, Schütze H, Niebur E, Von Der Heydt R. 2007. A neural model of figure-ground organization. *J Neurophysiol* **97**:4310–4326. doi:10.1152/jn.00203.2007

Deco G, Rolls ET. 2004. A Neurodynamical cortical model of visual attention and invariant object recognition. *Vision Res* **44**:621–642. doi:10.1016/j.visres.2003.09.037

Doerig A, Sommers R, Seeliger K, Richards B, Ismael J, Lindsay G, Kording K, Konkle T, van Gerven MAJ, Kriegeskorte N, Kietzmann TC. 2022. The neuroconnectionist research programme. *ArXiv*.

Duncan J. 1984. Selective Attention and the Organization of Visual Information, Journal of Experimental Psychology: General.

Echeveste R, Aitchison L, Hennequin G, Lengyel M. 2020. Cortical-like dynamics in recurrent circuits optimized for sampling-based probabilistic inference. *Nat Neurosci* **23**:1138–1149. doi:10.1038/s41593-020-0671-1

490 Ernst MR, Triesch J, Burwick T. 2019. Recurrent connections aid occluded object recognition by
491 discounting occludersLecture Notes in Computer Science (Including Subseries Lecture Notes in
492 Artificial Intelligence and Lecture Notes in Bioinformatics). Springer Verlag. pp. 294–305.
493 doi:10.1007/978-3-030-30508-6_24

494 Felleman DJ, van Essen DC. 1991. Distributed hierarchical processing in the primate cerebral cortex.
495 *Cerebral Cortex* **1**:1–47.

496 Fowlkes CC, Martin DR, Malik J. 2007. Local figure-ground cues are valid for natural images. *J Vis* **7**:2.
497 doi:10.1167/7.8.2

498 Fox KJ, Birman D, Gardner JL. 2023. Gain, not concomitant changes in spatial receptive field
499 properties, improves task performance in a neural network attention model. *Elife* **12**.
500 doi:10.7554/eLife.78392

501 Guclu U, van Gerven MA. 2015. Deep Neural Networks Reveal a Gradient in the Complexity of Neural
502 Representations across the Ventral Stream. *J Neurosci* **35**:10005–10014.
503 doi:10.1523/JNEUROSCI.5023-14.2015

504 Hamker FH. 2005. The reentry hypothesis: The putative interaction of the frontal eye field,
505 ventrolateral prefrontal cortex, and areas V4, IT for attention and eye movement. *Cerebral*
506 *Cortex* **15**:431–447. doi:10.1093/cercor/bhh146

507 Hesse JK, Tsao DY. 2016. Consistency of Border-Ownership Cells across Artificial Stimuli, Natural
508 Stimuli, and Stimuli with Ambiguous Contours. *J Neurosci* **36**:11338–11349.
509 doi:10.1523/JNEUROSCI.1857-16.2016

510 Hong S, Noh H, Han B. 2015. Decoupled Deep Neural Network for Semi-supervised Semantic
511 Segmentation. *Adv Neural Inf Process Syst* **2015-January**:1495–1503.

512 Jarvers C, Neumann H. 2019. Incorporating Feedback in Convolutional Neural Networks. Cognitive
513 Computational Neuroscience.

514 Kar K, Kubilius J, Schmidt K, Issa EB, DiCarlo JJ. 2019. Evidence that recurrent circuits are critical to
515 the ventral stream's execution of core object recognition behavior. *Nat Neurosci* **22**:974–983.
516 doi:10.1038/s41593-019-0392-5

517 Keller AJ, Roth MM, Scanziani M. 2020. Feedback generates a second receptive field in neurons of
518 the visual cortex. *Nature* **582**:545–549. doi:10.1038/s41586-020-2319-4

519 Kietzmann TC, Spoerer CJ, Sörensen LKA, Cichy RM, Hauk O, Kriegeskorte N. 2019. Recurrence is
520 required to capture the representational dynamics of the human visual system. *Proc Natl Acad*
521 *Sci U S A* **116**:21854–21863. doi:10.1073/pnas.1905544116

522 Kim J, Linsley D, Thakkar K, Serre T. 2019. Disentangling neural mechanisms for perceptual grouping.
523 *ArXiv*.

524 Kingma DP, Ba JL. 2015. Adam: A method for stochastic optimization3rd International Conference
525 on Learning Representations, ICLR 2015 - Conference Track Proceedings. International
526 Conference on Learning Representations, ICLR.

527 Kirchberger L, Mukherjee S, Self MW, Roelfsema PR. 2023. Contextual drive of neuronal responses
528 in mouse V1 in the absence of feedforward input. *Sci Adv* **9**. doi:10.1126/sciadv.add2498

529 Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, Xiao T, Whitehead S, Berg AC, Lo W-Y,
530 Dollár P, Girshick R. 2023. Segment Anything.

531 Kramer AF, Weber TA, Watson SE. 1997. Object-based attentional selection--Grouped arrays or
532 spatially invariant representations?: Comment on Vecera and Farah (1994). *J Exp Psychol Gen*
533 **126**:3–13. doi:10.1037//0096-3445.126.1.3

534 Kubilius J, Schrimpf M, Nayebi A, Bear D, Yamins DLK, DiCarlo JJ. 2018. CORnet: Modeling the Neural
535 Mechanisms of Core Object Recognition. *bioRxiv*. doi:10.1101/408385

536 Lamme VA. 1995. The neurophysiology of figure-ground segregation in primary visual cortex. *J*
537 *Neurosci* **15**:1605–1615.

538 Lei J, Benjamin AS, Kording KP. 2021. Object Based Attention Through Internal Gating. *ArXiv*.

539 Li H, Xiong P, An J, Wang L. 2018. Pyramid Attention Network for Semantic Segmentation. *British*
540 *Machine Vision Conference 2018, BMVC 2018*.

541 Lindsay GW, Miller KD. 2018. How biological attention mechanisms improve task performance in a
542 large-scale visual system model. *Elife* **7**. doi:10.7554/eLife.38105

543 Linsley D, Kim J, Ashok A, Serre T. 2020. Recurrent neural circuits for contour detection. *8th*
544 *International Conference on Learning Representations, ICLR 2020*.

545 Linsley D, Kim J, Veerabadran V, Serre T. 2018. Learning long-range spatial dependencies with
546 horizontal gated-recurrent units. *Adv Neural Inf Process Syst* **2018-December**:152–164.

547 Michaelis C, Bethge M, Ecker AS. 2018. One-Shot Segmentation in Clutter. *35th International*
548 *Conference on Machine Learning, ICML 2018* **8**:5718–5727.

549 Moore T, Fallah M. 2004. Microstimulation of the Frontal Eye Field and Its Effects on Covert Spatial
550 Attention. *J Neurophysiol* **91**:152–162. doi:10.1152/jn.00741.2002

551 Muckli L, de Martino F, Vizioli L, Petro LS, Smith FW, Ugurbil K, Goebel R, Yacoub E. 2015. Contextual
552 Feedback to Superficial Layers of V1. *Curr Biol* **25**:2690–2695. doi:10.1016/j.cub.2015.08.057

553 Neri P. 2017. Object segmentation controls image reconstruction from natural scenes. *PLoS Biol*
554 **15**:e1002611. doi:10.1371/journal.pbio.1002611

555 Neri P. 2014. Semantic control of feature extraction from natural scenes. *J Neurosci* **34**:2374–2388.
556 doi:10.1523/JNEUROSCI.1755-13.2014

557 O'Craven KM, Downing PE, Kanwisher N. 1999. fMRI evidence for objects as the units of attentional
558 selection. *Nature* **401**:584–587. doi:10.1038/44134

559 Papale P, Betta M, Handjaras G, Malfatti G, Cecchetti L, Rampinini A, Pietrini P, Ricciardi E, Turella L,
560 Leo A. 2019. Common spatiotemporal processing of visual features shapes object
561 representation. *Sci Rep* **9**:1–8. doi:10.1038/s41598-019-43956-3

562 Papale P, Leo A, Cecchetti L, Handjaras G, Kay KN, Pietrini P, Ricciardi E. 2018. Foreground-
563 background segmentation revealed during natural image viewing. *eNeuro* **5**.
564 doi:10.1523/ENEURO.0075-18.2018

565 Papale P, Leo A, Handjaras G, Cecchetti L, Pietrini P, Ricciardi E. 2020. Shape coding in occipito-
566 temporal cortex relies on object silhouette, curvature and medial-axis. *J Neurophysiol*.
567 doi:10.1152/jn.00212.2020

568 Papale P, Wang F, Morgan AT, Chen X, Gilhuis A, Petro LS, Muckli L, Roelfsema PR, Self MW. 2023.
569 The representation of occluded image regions in area V1 of monkeys and humans. *Current*
570 *Biology*. doi:10.1016/j.cub.2023.08.010

571 Papale P, Zuiderbaan W, Teeuwen RRM, Gilhuis A, Self MW, Roelfsema PR, Dumoulin SO. 2021. The
572 influence of objecthood on the representation of natural images in the visual cortex. *bioRxiv*
573 2021.09.21.461209. doi:10.1101/2021.09.21.461209

574 Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L,
575 Desmaison A, Köpf A, Yang E, DeVito Z, Raison M, Tejani A, Chilamkurthy S, Steiner B, Fang L,
576 Bai J, Chintala S. 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library.
577 *ArXiv*.

578 Poort J, Raudies F, Wannig A, Lamme VAF, Neumann H, Roelfsema PR. 2012. The role of attention
579 in figure-ground segregation in areas V1 and V4 of the visual cortex. *Neuron* **75**:143–156.
580 doi:10.1016/j.neuron.2012.04.032

581 Poort J, Self MW, van Vugt B, Malkki H, Roelfsema PR. 2016. Texture Segregation Causes Early Figure
582 Enhancement and Later Ground Suppression in Areas V1 and V4 of Visual Cortex. *Cereb Cortex*
583 **26**:3964–3976. doi:10.1093/cercor/bhw235

584 Pozzi I, Bohté SM, Roelfsema PR. 2020. Attention-Gated Brain Propagation: How the brain can
585 implement reward-based error backpropagation Advances in Neural Information Processing
586 Systems.

587 Roelfsema PR. 2023. Solving the binding problem: Assemblies form when neurons enhance their
588 firing rate—they don't need to oscillate or synchronize. *Neuron*.
589 doi:10.1016/j.neuron.2023.03.016

590 Roelfsema PR. 2006. Cortical algorithms for perceptual grouping. *Annu Rev Neurosci* **29**:203–227.
591 doi:10.1146/annurev.neuro.29.051605.112939

592 Roelfsema PR, de Lange FP. 2016. Early Visual Cortex as a Multiscale Cognitive Blackboard. *Annu Rev*
593 *Vis Sci* **2**:131–151. doi:10.1146/annurev-vision-111815-114443

594 Roelfsema PR, Houtkamp R. 2011. Incremental grouping of image elements in vision. *Atten Percept*
595 *Psychophys* **73**:2542–2572. doi:10.3758/s13414-011-0200-0

596 Roelfsema PR, Lamme VAF, Spekreijse H, Bosch H. 2002. Figure - Ground segregation in a recurrent
597 network architecture. *J Cogn Neurosci* **14**:525–537. doi:10.1162/08989290260045756

598 Seijdel N, Tsakmakidis N, de Haan EHF, Bohte SM, Scholte HS. 2020. Depth in convolutional neural
599 networks solves scene segmentation. *PLoS Comput Biol* **16**:e1008022.
600 doi:10.1371/journal.pcbi.1008022

601 Self MW, Jeurissen D, van Ham AF, van Vugt B, Poort J, Roelfsema PR. 2019. The Segmentation of
602 Proto-Objects in the Monkey Primary Visual Cortex. *Curr Biol* **29**:1019-1029 e4.
603 doi:10.1016/j.cub.2019.02.016

604 Self MW, Kooijmans RN, Super H, Lamme VA, Roelfsema PR. 2012. Different glutamate receptors
605 convey feedforward and recurrent processing in macaque V1. *Proc Natl Acad Sci U S A*
606 **109**:11031–11036. doi:10.1073/pnas.1119527109

607 Self MW, van Kerkoerle T, Super H, Roelfsema PR. 2013. Distinct roles of the cortical layers of area
608 V1 in figure-ground segregation. *Curr Biol* **23**:2121–2129. doi:10.1016/j.cub.2013.09.013

609 Serre T. 2019. Deep Learning: The Good, the Bad, and the Ugly. *Annu Rev Vis Sci* **5**:399–426.
610 doi:10.1146/annurev-vision-091718-014951

611 Spitzer H, Desimone R, Moran J. 1988. Increased attention enhances both behavioral and neuronal
612 performance. *Science (1979)* **240**:338–340.

613 Spoerer CJ, Kietzmann TC, Mehrer J, Charest I, Kriegeskorte N. 2020. Recurrent neural networks can
614 explain flexible trading of speed and accuracy in biological vision. *PLoS Comput Biol*
615 **16**:e1008215. doi:10.1371/journal.pcbi.1008215

616 Spoerer CJ, McClure P, Kriegeskorte N. 2017. Recurrent Convolutional Neural Networks: A Better
617 Model of Biological Object Recognition. *Front Psychol* **8**:278016.
618 doi:10.3389/fpsyg.2017.01551

619 Thorat S, Aldegheri G, Kietzmann TC. 2021. Category-orthogonal object features guide information
620 processing in recurrent neural networks trained for object categorization.

621 Thorat S, van Gerven M, Peelen M. 2019. The functional role of cue-driven feature-based feedback
622 in object recognition. doi:10.32470/CCN.2018.1044-0

Treue S, Martínez Trujillo JC. 1999. Feature-based attention influences motion processing gain in macaque visual cortex. *Nature* **399**:575–579. doi:10.1038/21176

Van Der Velde F, De Kamps M. 2001. From knowing what to knowing where: Modeling object-based attention with feedback disinhibition of activation. *J Cogn Neurosci* **13**:479–491. doi:10.1162/08989290152001907

van der Walt S, Colbert SC, Varoquaux G. 2011. The NumPy array: A structure for efficient numerical computation. *Comput Sci Eng* **13**:22–30. doi:10.1109/MCSE.2011.37

van Essen DC, Anderson CH, Felleman DJ. 1992. Information processing in the primate visual system: an integrated systems perspective. *Science (1979)* **255**:419–423.

van Kerkoerle T, Self MW, Roelfsema PR. 2017. Layer-specificity in the effects of attention and working memory on activity in primary visual cortex. *Nat Commun* **8**:1–14. doi:10.1038/ncomms13804

van Vugt B, Dagnino B, Vartak D, Safaai H, Panzeri S, Dehaene S, Roelfsema PR. 2018. The threshold for conscious report: Signal loss and response bias in visual and frontal cortex. *Science (1979)* **360**:537–542. doi:10.1126/science.aar7186

Vecera SP, Farah MJ. 1997. Is visual image segmentation a bottom-up or an interactive process? *Percept Psychophys* **59**:1280–1296. doi:10.3758/BF03214214

Williams MA, Baker CI, op de Beeck HP, Mok Shim W, Dang S, Triantafyllou C, Kanwisher N. 2008. Feedback of visual object information to foveal retinotopic cortex. *Nat Neurosci* **11**:1439–1445. doi:10.1038/nn.2218

Williford JR, von der Heydt R. 2016. Figure-Ground Organization in Visual Cortex for Natural Scenes. *eNeuro* **3**. doi:10.1523/ENEURO.0127-16.2016

Yamins DL, Hong H, Cadieu CF, Solomon EA, Seibert D, DiCarlo JJ. 2014. Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proc Natl Acad Sci U S A* **111**:8619–8624. doi:10.1073/pnas.1403112111

Yan S, Fang X, Xiao B, Rockwell H, Zhang Y, Lee TS. 2019. Recurrent Feedback Improves Feedforward Representations in Deep Neural Networks. *ArXiv*.

Yann LeCun, Corinna Cortes, Chris Burges. 2010. The MNIST handwritten digit database. http://yann.lecun.com/exdb/mnist/

Zipser K, Lamme VA, Schiller PH. 1996. Contextual modulation in primary visual cortex. *J Neurosci* **16**:7376–7389.

Zuiderbaan W, van Leeuwen J, Dumoulin SO. 2017. Change Blindness Is Influenced by Both Contrast Energy and Subjective Importance within Local Regions of the Image. *Front Psychol* **8**:1718. doi:10.3389/fpsyg.2017.01718