

Modulation of early visual processing alleviates capacity limits in solving multiple tasks

Sushrut Thorat¹ Giacomo Aldegheri¹ Marcel A. J. van Gerven Marius V. Peelen

s.thorat@donders.ru.nl, g.aldegheri@donders.ru.nl, m.vangerven@donders.ru.nl, m.peelen@donders.ru.nl

Donders Institute for Brain, Cognition and Behaviour, Radboud University

Abstract

In daily life situations, we have to perform multiple tasks given a visual stimulus, which requires task-relevant information to be transmitted through our visual system. When it is not possible to transmit all the possibly relevant information to higher layers, due to a bottleneck, task-based modulation of early visual processing might be necessary. In this work, we report how the effectiveness of modulating the early processing stage of an artificial neural network depends on the information bottleneck faced by the network. The bottleneck is quantified by the number of tasks the network has to perform and the neural capacity of the later stage of the network. The effectiveness is gauged by the performance on multiple object detection tasks, where the network is trained with a recent multi-task optimization scheme. By associating neural modulations with task-based *switching* of the state of the network and characterizing when such switching is helpful in early processing, our results provide a functional perspective towards understanding why task-based modulation of early neural processes might be observed in the primate visual cortex².

Keywords: neural modulation, multi-task learning, early visual cortex, attention, perception, capacity limits

Introduction

Humans and other animals have to perform multiple tasks given a visual stimulus. For example, seeing a face, we may have to say whether it is happy or sad, or recognize its identity. For each of these tasks, a subset of all the features of the face are useful. In principle, it could be possible for a visual system to extract all of the features necessary to solve all possible tasks, and then select the relevant information from this rich representation downstream. However, as the number of tasks increases, a network with a limited capacity may not be able to extract all of the potentially relevant features (an information bottleneck is manifest), requiring the information that is extracted from the stimulus in the early processing stages to change according to the task.

Several studies in neuroscience have found evidence for such task-dependent modulations of sensory processing in the primate visual system, including at the early levels (Carrasco, 2011; Maunsell & Treue, 2006; Gilbert & Li, 2013). For example, human neuroimaging studies have

shown that attending to a stimulus could lead to an increase in the accuracy with which its task-relevant features could be decoded by a classifier in early visual areas (Jehee, Brady, & Tong, 2011), and neurophysiological experiments in non-human primates have shown that the stimulus selectivity of neurons in primary visual cortex was dependent on the task the monkeys had to perform (Gilbert & Li, 2013).

Despite the observation of such modulations of early visual processing, it is not clear whether they are causally necessary for performing better on the corresponding tasks. This question has been addressed by deploying biologically-inspired task-based modulations on computational models. Lindsay and Miller (2018) showed that task-based modulation deployed on multiple stages of a convolutional neural network improves performance on challenging object classification tasks. Other recent work (Thorat, van Gerven, & Peelen, 2018; Rosenfeld, Biparva, & Tsotsos, 2018) has also shown that task-based modulation of early visual processing aids in object detection and segmentation in addition to the task-based modulation of late processing. However, the conditions under which early modulation can be beneficial in performing multiple tasks have not been systematically investigated.

In the present work, we assessed the effectiveness of task-based modulation of early visual processing as a function of an information bottleneck in a neural network, quantified by the number of tasks the network had to execute and the neural capacity of the network. To do so, we trained networks to, given an image, provide an answer conditioned on the cued task. Every task required detecting the presence of the corresponding object in the image. The networks were trained according to a recent framework proposed in the field of continual learning (Cheung, Terekhov, Chen, Agrawal, & Olshausen, 2019), which helps them execute multiple tasks by switching their state given a task cue, in order to transmit relevant information through the network. In this work, to quantify the effectiveness of task-based modulation of early neural processing, we measured the increase in performance provided by modulating early neural processing in addition to modulating the late neural processing in the networks.

Methods

Task and system description

In a multi-task setting, object detection can be thought of as solving one of a set of possible binary classification (one object versus the rest) problems. Given an image and a task cue indicating the identity of the object to be detected, a network had to output if the object in the image matched the task cue.

We used MNIST (LeCun, Bottou, Bengio, & Haffner, 1998)

¹Equal contribution

²The code to train and analyze the networks mentioned here can be found at - <https://github.com/novelmartis/early-vs-late-multi-task>



digits and their permutations as objects (Kirkpatrick et al., 2017). The original MNIST dataset has $28 \times 28 \text{px}^2$ images of 10 digits. Each permuted version consists of images of those 10 digits, whose pixels undergo a given permutation, creating 10 new objects. We varied the number of permutations used (10, 25, and 50) to modulate the number of tasks the networks had to perform (which are 10 times the number of permutations).

We considered a multilayer perceptron with rectified linear units (ReLU), which had one hidden layer between the input (image) and the binary output. The number of neurons in the hidden layer were variable (32, 64, and 128) and determined the neural capacity of the late stage of the network.

Task-based modulation and its function

Modelling biological neurons as perceptrons (Rosenblatt, 1957), task-based modulations have been shown to affect the effective biases and gains of the neurons (Maunsell & Treue, 2006; Boynton, 2009; Ling, Liu, & Carrasco, 2009). The nature of modulation - which neurons to modulate and how - is under debate (Boynton, 2009; Thorat et al., 2018). We adapted these findings by introducing task-based modulation into our networks via the biases of the perceptrons and the gains of their ReLU activation functions. The modulations were then trained end-to-end with the rest of the network.

Given a particular task, the task cue is a one-hot encoding of the relevant object. Task-based modulation is mediated through bias and gain modulation in the following manner.

$$x_n = [W_n(g_{n-1} \circ x_{n-1}) + b_n]_+ \quad (1)$$

$$g_n = G_n c, \quad b_n = B_n c \quad (2)$$

where the transformation between layers $n-1$ and n ($L_{n-1} \rightarrow L_n$) is modulated by changing the slope of the ReLU activation function (gain, g_{n-1}) in L_{n-1} and the bias (b_n) to the perceptrons in L_n ; x_n are the pre-gain activations of the perceptrons in L_n , W_n is the task-independent transformation matrix between L_{n-1} and L_n , G_n and B_n map the task cue c (one-hot encoding of the relevant object k) to the gain and bias modulations of the perceptrons in L_n respectively, and \circ refers to element-wise multiplication.

Given a task k , modulating the gains of the pre-synaptic perceptrons (in L_{n-1}) and the biases of the post-synaptic perceptrons (in L_n) transforms the information transformation between L_{n-1} and L_n . This allows for the transmission of information required to perform task k , while ignoring the information required to perform the other tasks, as formalized in Cheung et al. (2019). This transformation can also be thought of as the network *switching* its state to selectively transmit task-relevant information downstream (see Figure 1). The conditions - the nature of these modulations and the neural capacity of the network - under which the network can switch between a given number of tasks, are preliminarily described in Cheung et al. (2019).

Here, for every relevant layer L_n , W_n , B_n , and G_n were jointly learned for the given number of tasks.

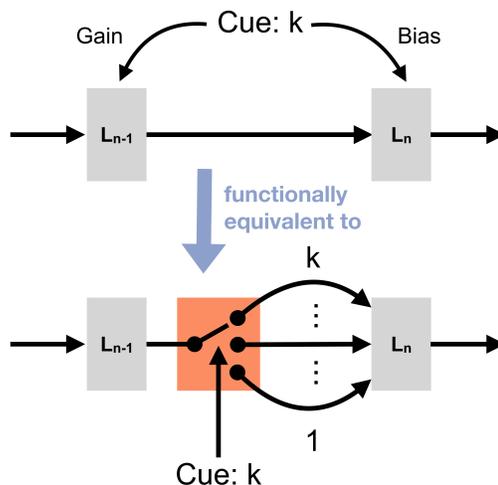


Figure 1: The effect of bias and gain modulation on the transformations in the network. Modulating the gains and biases is *functionally* equivalent to *switching* the transformation being performed to one suited for the relevant task. Such an example of switching is visualized in the figure. Given a task cue corresponding to object k , corresponding gain and bias modulations are applied, which results in the $L_{n-1} \rightarrow L_n$ transformation being switched into one that transmits feature information required to detect the presence or absence of object k .

Evaluation metric and expected trends

The effectiveness of early neural modulation was quantified by the average absolute increase in detection performance across all the tasks when modulations were implemented on both the transformations $L_1 \rightarrow L_2$ and $L_2 \rightarrow L_3$ (L_1 corresponds to the input layer and L_3 to the output layer) as opposed to when the modulations were trained on the transformation $L_2 \rightarrow L_3$ only.

We expected the effectiveness of task-based early neural modulation to be directly proportional to the number of neurons in L_2 and inversely proportional to the number of tasks (permuted MNIST sets used).

Neural network training details

All the networks were trained with adaptive stochastic gradient descent with backpropagation through the ADAM optimizer (Kingma & Ba, 2014) with the default settings in TensorFlow (v1.4.0) and $\alpha = 10^{-5}$. We used a batch size of 100. Half of each batch contained randomly selected images of randomly selected tasks where the cued object was present, and half where the cued object was not present. These images were taken from the MNIST training set and its corresponding permutations. The images were augmented by adding small translations and some noise. We trained each network with 10^7 such batches. The relevant metrics discussed in the previous section are computed at the end of training over a batch of size 10^5 created from the MNIST test set and its corresponding permutations.

Results

We first analyzed the detection performance of the network with only $L_2 \rightarrow L_3$ modulation. The network performance as a function of the number of neurons in L_2 and the number of detection tasks the network had to perform is shown in Figure 2 (red circles). The network performance increased with an increase in the number of neurons in L_2 , as the neural capacity increased. The performance decreased with an increase in the number of tasks to be performed, as the representational capacity of the network for any one task was reduced. A network with as little as 32 neurons in its hidden layer was able to switch between as many as 500 detection tasks, while keeping the average detection performance across all the tasks as high as 87%, thus replicating the success of the multi-task learning framework proposed by Cheung et al. (2019).

To assess the dependence of the effectiveness of task-based modulation of early neural processing ($L_1 \rightarrow L_2$) on the bottleneck in the network, we analyzed the boost in average detection performance when task-based modulation of $L_1 \rightarrow L_2$ was deployed in addition to task-based modulation of $L_2 \rightarrow L_3$, as a function of the number of neurons in L_2 and the number of detection tasks the network had to perform. The resulting boosts are shown in Figure 2 (Δ_{\uparrow} quantification). The performance boost increased as the number of neurons in L_2 decreased, and as the number of tasks the network has to perform increased. This confirms the hypothesis that task-based modulation of early neural processing is essential when an information bottleneck exists in a subsequent processing stage.

The contribution of bias and gain modulation

Gain, but not so much bias, modulation of neural responses has been observed in experiments investigating feature-based attention in the monkey/human brain (Maunsell & Treue, 2006; Boynton, 2009). We assessed how the two contributed to the overall modulation of the transformations in the network.

We selectively turned off the bias or gain modulation for all the variants of the network that were trained. The average detection performance decreased by $43.0 \pm 2.0\%$ when gain modulation was turned off, and by $3.9 \pm 0.9\%$ when bias modulation was turned off, suggesting that in our framework, when jointly deployed, gain modulation is more important than bias modulation in switching the state of the network to be able to perform the desired task well.

We also trained a network with 32 neurons in L_2 , on 25 permutations of MNIST, with gain-only or bias-only modulations of both the $L_1 \rightarrow L_2$ and $L_2 \rightarrow L_3$ transformations. When the gain and bias modulations were jointly trained, the network performance was 94.7%. With gain-only modulation, the performance was 94.8%, and with bias-only modulation the performance was 90.9%. As the performance when only bias modulation was deployed was much higher than chance (50%), we can conclude that bias modulation alone can also lead to efficient task-switching. When the bias and gain modulations are jointly trained, gain might take over as it multiplicatively impacts responses, and therefore has higher gradients during training, as opposed to the additive impact of bias.

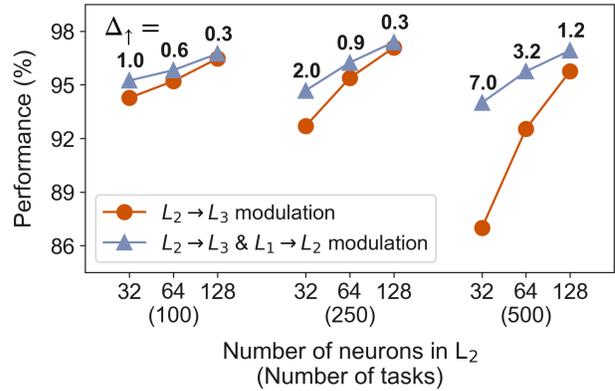


Figure 2: The effectiveness of task-based modulation (quantified by the performance boost, Δ_{\uparrow}) of early neural processing ($L_1 \rightarrow L_2$) as a function of the number of neurons in L_2 and the number of tasks the network has to perform. The performance boost was inversely proportional to the number of neurons in L_2 and directly proportional to the number of tasks the network had to perform. The absolute performance profiles given either the modulation of $L_2 \rightarrow L_3$ only or the joint modulation of $L_2 \rightarrow L_3$ and $L_1 \rightarrow L_2$ are also shown.

Discussion

Adding to the discussion about the functional role of task-based modulation of early neural processing, in this work we have shown that modulating the early layer of an artificial neural network in a task-dependent manner can boost performance, beyond just modulating the late layer, in a multi-task learning scenario in which a network contains an information bottleneck, either due to a large number of tasks to be performed or to a small number of units in the late layer.

Adapting a formalism proposed by Cheung et al. (2019), we showed how bias and gain modulation, two prevalent neuronal implementations of top-down modulation in the brain, could functionally lead to switching the state of a network to perform transformations effective for the task at hand. While task-dependent computations are widespread in higher-level areas of the primate brain, such as prefrontal cortex (Mante, Sussillo, Shenoy, & Newsome, 2013), it is not clear to what extent sensory streams (which perform early visual processing) can also be seen as switching their state according to the current task (although see Gilbert and Li (2013) for a proposal), and what the functional relevance of doing so would be. Here we show how, in principle, this switching could be computationally advantageous when it is not possible to send the information required for all tasks to higher layers, which might well be the case in the complex environments that humans and other animals are able to navigate.

To further investigate the relevance of our findings to biological visual systems, in follow-up work we intend to deploy our modulation scheme on architectures that bear more similarity to the primate visual hierarchy, such as deep convolutional networks (Kriegeskorte, 2015), datasets of naturalistic images

such as ImageNet (Russakovsky et al., 2015), and general naturalistic tasks such as visual question answering (Agrawal et al., 2017). This will allow us to assess whether the functional advantage provided by early modulation holds true in a more realistic scenario, and whether the resulting modulation schemes resemble those observed in the early visual areas of the primate brain.

Finally, a key aspect of our approach is the fact that the network is constantly operating in a task-dependent manner. Most previous approaches to task-dependent modulation have assumed the presence of an underlying task-free representation on which the modulation operates (for example, in the case of Lindsay and Miller (2018) this corresponds to a network pre-trained on object recognition). Providing the network with task cues during the training phase, on the other hand, has been used in the field of continual learning (Cheung et al., 2019; Masse, Grant, & Freedman, 2018; Yang, Joglekar, Song, Newsome, & Wang, 2019), and according to one influential theory in neuroscience, the interplay between sparse, context-specific information encoded by the hippocampus and shared structural information in the neocortex is crucial for learning new tasks without overwriting previous ones (Kumaran, Hassabis, & McClelland, 2016). To our knowledge, the question of how the task-based modulations observed in visual cortex might be learned has not been explicitly addressed in previous literature. On the one hand, it is possible that a context-free representation is learned first, possibly through unsupervised learning, and then modulated upon. On the other, learning of representations and task modulations might interact at all stages, allowing the representations to be optimized for the type of modulations they are subject to. Whether one scheme or the other constitutes a better explanation for the modulations observed in biological visual systems is an important direction for future research.

Acknowledgments

This work was funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant Agreement No. 725970). This manuscript reflects only the authors' view, and the agency is not responsible for any use that may be made of the information it contains.

References

Agrawal, A., Lu, J., Antol, S., Mitchell, M., Zitnick, C. L., Parikh, D., & Batra, D. (2017). Vqa: Visual question answering. *International Journal of Computer Vision*, 123(1), 4–31.

Boynton, G. M. (2009). A framework for describing the effects of attention on visual responses. *Vision Research*, 49(10), 1129–1143.

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research*, 51(13), 1484–1525.

Cheung, B., Terekhov, A., Chen, Y., Agrawal, P., & Olshausen, B. (2019). Superposition of many models into one. *arXiv preprint arXiv:1902.05522*.

Gilbert, C. D., & Li, W. (2013). Top-down influences on visual processing. *Nature Reviews Neuroscience*, 14(5), 350–363.

Jehee, J. F., Brady, D. K., & Tong, F. (2011). Attention improves encoding of task-relevant features in the human visual cortex. *Journal of Neuroscience*, 31(22), 8210–8219.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A. A., ... others (2017). Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences*, 114(13), 3521–3526.

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

Kumaran, D., Hassabis, D., & McClelland, J. L. (2016). What learning systems do intelligent agents need? complementary learning systems theory updated. *Trends in Cognitive Sciences*, 20(7), 512–534.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2324.

Lindsay, G. W., & Miller, K. D. (2018). How biological attention mechanisms improve task performance in a large-scale visual system model. *eLife*, 7, e38105.

Ling, S., Liu, T., & Carrasco, M. (2009). How spatial and feature-based attention affect the gain and tuning of population responses. *Vision Research*, 49(10), 1194–1204.

Mante, V., Sussillo, D., Shenoy, K. V., & Newsome, W. T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*, 503(7474), 78.

Masse, N. Y., Grant, G. D., & Freedman, D. J. (2018). Alleviating catastrophic forgetting using context-dependent gating and synaptic stabilization. *Proceedings of the National Academy of Sciences*, 115(44), E10467–E10475.

Maunsell, J. H., & Treue, S. (2006). Feature-based attention in visual cortex. *Trends in Neurosciences*, 29(6), 317–322.

Rosenblatt, F. (1957). *The perceptron, a perceiving and recognizing automaton project para*. Cornell Aeronautical Laboratory.

Rosenfeld, A., Biparva, M., & Tsotsos, J. K. (2018). Priming neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 2011–2020).

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... others (2015). Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3), 211–252.

Thorat, S., van Gerven, M., & Peelen, M. (2018). The functional role of cue-driven feature-based feedback in object recognition. In *Conference on Cognitive Computational Neuroscience, CCN 2018* (pp. 1–4).

Yang, G. R., Joglekar, M. R., Song, H. F., Newsome, W. T., & Wang, X.-J. (2019). Task representations in neural networks trained to perform many cognitive tasks. *Nature Neuroscience*, 22(2), 297.