Deep neural network models of visual cognition

Sörensen, L.K.A.

Publication date
2023

Chapter 6

Summary

Can deep neural networks be used as models of visual cognition, to capture the interaction between sensation and cognition during object perception? In this thesis, building on extensive prior research on object recognition in the visual ventral stream and deep neural networks, I evaluated this question for three different modulations of object recognition: arousal state, spatial attention, and previous visual experience altered by presentation duration. For each of these factors, the leading goal was to formulate the interaction between object recognition and cognitive modulation in an image-computable, mechanistic model that reproduces key characteristics of human behaviour. Building these models and adopting human behavioural characteristics as the explanatory goal made it possible to revisit long-standing questions and to arbitrate between different hypotheses at the algorithmic level.

In Chapter 2, I used this approach to investigate how a global gain modulation — previously linked to changes in arousal state — may affect visual processing and recognition performance on a series of tasks with varying difficulty. In particular, I have observed that changes in global gain reproduce a known interaction between of arousal state and task difficulty on performance. That is, whereas difficult tasks are typically best performed at medium arousal states, easy tasks are best performed at higher arousal states. Studying this phenomenon with a deep convolutional neural network (DCNN) made it possible to propose a new hypothesis on how these effects may arise as an interaction between a global gain modulation and a hierarchical sensory system such as the visual ventral stream.

In Chapter 3, I assessed the efficacy of different proposed attention mechanisms at implementing selective processing during object recognition in naturalistic scenes. Specifically, I noted that gain-based mechanisms were more effective than a precision-based mechanism both at replicating known neural modulations of selective attention in a spiking DCNN, as well as at modulating object recognition performance.

In Chapter 4, I evaluated how different mechanisms connected to temporal integration, such as lateral recurrent processing and sensory adaptation may contribute to dynamic object recognition, such as demonstrated by humans during an RSVP task across varying presentation durations. I found that lateral recurrent DCNNs augmented with sensory adaptation mechanisms were most effective at recovering performance levels comparable to those of human participants at fast to medium-fast presentation rates. Evaluating participants' trial-by-trial report rates further indicated that in particular a model with a power-law adaptation mechanism could account for human dynamic recognition behaviour during this task.

In Chapter 5, I introduced an improved lateral recurrent architecture and expanded on previous observations on lateral recurrence and its role in dynamic recognition to two additional tasks (perceptual masking and animacy classification), and behavioural measures (accuracy and reaction times). Generally, I found that lateral recurrence could also capture recognition behaviour on all these tasks and measures, particularly for continuous visual stimulation. However, I observed that this critically depended on whether lateral recurrence was implemented in a high-performing and temporally stable object recognition model, such as the newly developed architecture *BLresnet*.

Together, the mechanistic insights gained in these four chapters provide encouraging examples of the use of DCNNs as a linking framework between sensation and cognitive modulation. That is, DCNNs can be used to test in how far cognitive mechanisms can generalize to richer statistical features of naturalistic images, as a testbed to arbitrate between mechanisms, as well as to describe shortcomings and synergies of proposed mechanisms across different sensory contexts. Together, DCNNs enable us to develop and test mechanistic models of visual cognition grounded in visual processing.