Abstract

Progress in deep learning has spawned great successes in many engineering applications. As a prime example, convolutional neural networks, a type of feedforward neural networks, are now approaching – and sometimes even surpassing – human accuracy on a variety of visual recognition tasks. Here, however, we show that these neural networks and their recent extensions struggle in recognition tasks where co-dependent visual features must be detected over long spatial ranges. We introduce the horizontal gated-recurrent unit (hGRU) to learn intrinsic horizontal connections – both within and across feature columns. We demonstrate that a single hGRU layer matches or outperforms all tested feedforward hierarchical baselines including state-of-the-art architectures which have orders of magnitude more free parameters. We further discuss the biological plausibility of the hGRU in comparison to anatomical data from the visual cortex as well as human behavioral data on a classic contour detection task.

Keywords: Biological vision; recurrent networks; horizontal connections; deep learning.
connections, allow receptive fields to adaptively “grow” without additional processing depth. Several computational neuroscience models of these neural circuits have been proposed to account for an array of phenomena from perceptual grouping to contextual illusions (e.g., Mely & Serre, 2016). However, because these models are fit to data by solving sets of differential equations using numerical integration, they have so far not been amenable to computer vision. We implement the core ideas of these models in an end-to-end trainable extension of the popular gated recurrent unit (GRU) (Cho et al., 2014), which we call the horizontal GRU (hGRU).

We compared feedforward and recurrent approaches to capturing long-range spatial dependencies in a large-scale analysis of model performance on Pathfinder. This revealed a striking trend: feedforward models struggle at solving Pathfinder, with only state-of-the-art feedforward models featuring millions of parameters across many processing layers succeeding (Fig. 2). An hGRU, on the other hand, efficiently solves Pathfinder with just one layer and a fraction of the number of parameters and training samples as feedforward models. The hGRU also outperforms all other tested recurrent models, including versions with lesions to its various mechanisms (such as linear nonlinear forms of excitation and inhibition), versions with less processing time, and a standard convolutional GRU (Fig. 2; yellow and brown dots).

We further investigated the nature of the horizontal connections learned by the hGRU by training it for contour detection in natural images (BSDS500 dataset). The learned kernels capture many of the canonical horizontal connectivity patterns found in cortex, including antagonistic near-excitatory vs. far-inhibitory surrounds, and the association field. We additionally find that fine-tuning this natural-image trained hGRU to a contour detection task yields a pattern of behavior in response to manipulations of contour salience that strongly correlates with human observers (Li & Gilbert, 2002).

This work diagnoses a computational deficiency of feedforward networks, and introduces a biologically-inspired solution that can be easily incorporated into existing deep learning architectures. Beyond its effectiveness in computer vision, the weights learned by the hGRU and its corresponding behaviors are consistent with those associated with visual cortex, demonstrating its potential for establishing novel connections between machine learning, cognitive science, and neuroscience.

Acknowledgments

This work was supported by the Carney Institute for Brain Science, NSF early career award (IIS-1252951), and DARPA young faculty award (N66001-14-1-4037).

References


