Modeling Spatiosemantic Lateral Connectivity of Primary Visual Cortex in CNNs

Tonio Weidler, Mario Senden, and Kurt Driessens

Maastricht University, The Netherlands
{t.weidler,mario.senden,kurt.driessens}@maastrichtuniversity.nl

Contemporary computer vision frequently draws on convolutional neural networks (CNN). State-of-the-art performance is often achieved by further deepening previous architectures, but this increases computational costs and complicates the mapping of artificial layers to cortical areas in computational neuroscience, where these networks are used as models for goal-driven research. An alternative direction of multidisciplinary relevance is thus the search for structural and algorithmic improvements within or between layers to alleviate the necessity of additional depth. Inspiration for this can be found in visual cortex. Naturally, this also improves the network's biological plausibility, rendering it more useful for neuroscience.

In primary visual cortex, neurons project to subsequent visual areas but also connect laterally, i.e. to neurons in the same area [3]. Here, we distinguish three types of lateral interaction: *Semantic lateral connections* link neurons responding to the same patch of the visual field but preferring different line orientations. This type of connectivity typically follows a Mexican hat profile assumed to fine-tune the neurons' orientation selectivity [2]. *Spatial lateral connectivity* establishes interactions between neurons of similar orientation selectivity responding to different patches of the visual field along the axis of their orientation, presumably integrating and segmenting contours [3]. *Complex cells* receive input from phase selective simple cells and merge them into phase invariant representations [1]. In CNNs neither semantic nor spatial lateral connections are explicitly modeled and complex cells are only loosely captured by pooling.

We introduce a joint model of *spatiosemantic lateral connectivity* and an explicit model of complex cells to extend CNNs. Spatial and semantic lateral connections enrich the first convolutional layer by transforming its activation map with biologically inspired wavelets along both the spatial domain and the channel domains. We avoid the necessity of explicitly incorporating temporal dynamics resulting from recurrent interactions by assuming these dynamics to be linear. This allows us to solve for their steady state which renders the lateral connectivity a single non-parametric feedforward operation. Phase invariant complex cells are simulated by two independent cell populations s_a and s_b contributing their activations to the layer's representation individually, but additionally merging into a third population of complex cells via a pairwise complex modulus nonlinearity $\mathbf{c} = \sqrt{(\mathbf{s}_a + \mathbf{s}_b)^2}$. Unlike fixed complex wavelets in [4], the kernels of simple cells are learned autonomously. A full architecture of the adapted first convolutional layer is given in Figure 1.

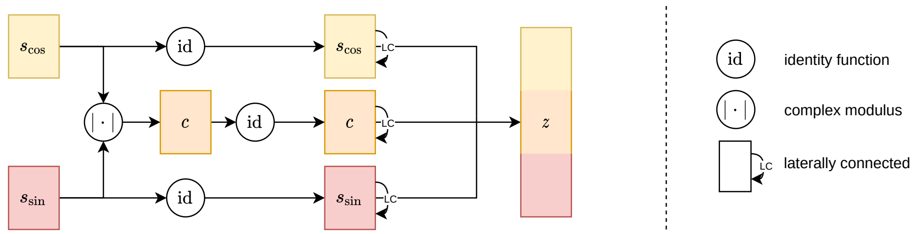

Fig. 1. Convolutional layer architecture with added complex cell simulation and lateral connections. s_a and s_b are two independent populations of neurons, realized as two arrays of convolutional filters. Their input is the original image, whereas the final output of the layer (z) constitutes the input to the second convolutional layer.

A qualitative analysis of the effects of our model of spatial lateral connectivity reveals that it successfully integrates segmented contours along straight lines. Experiments on object and texture classification showcase significant and substantial performance improvements in small-scale CNNs using complex cell simulations. Applied to texture classification, the combination of complex cells and spatial lateral connections produces the best performance, but spatial lateral connectivity on its own can already significantly improve a shallow network on both tasks. A closer look at the convolutional kernels emerging in laterally connected complex cells reveals their autonomously learned structure to be reminiscent of primary visual cortex. In particular, learned kernels largely adopt the orientation of their allocated spatial connectivity profiles and thus reinforce the proclaimed utility [3] of facilitation along this axis. In conclusion, our results demonstrate that introducing biologically inspired connectivity patterns into CNNs benefits their performance despite not increasing the number of trainable parameters. Improvements can be attributed to the integration and segmentation of contours during early visual processing, as the artificial connectivity profiles emulate those fulfilling these functions in the brain. In consequence, the introduced layer augmentation may not only improve small-scale CNNs in computer vision applications but also foster neuroscientific research relying on biologically plausible, goal-driven models.

References

1. Hubel, D.H., Wiesel, T.N.: Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. The Journal of Physiology **160**(1), 106–154 (Jan 1962)
2. Kang, K., Shelley, M., Sompolinsky, H.: Mexican hats and pinwheels in visual cortex. PNAS **100**(5), 2848–2853 (2003)
3. Lorenceau, J., Giersch, A., Series, P.: Dynamic competition between contour integration and contour segmentation probed with moving stimuli. Vision Research **45**(1), 103–116 (2005)
4. Mallat, S.: Group invariant scattering. Communications on Pure and Applied Mathematics **65**(10), 1331–1398 (2012)