

Developmental plasticity rules facilitate representation learning in a model of visual ventral stream

Ariane Delrocq (ariane.delrocq@epfl.ch)

Laboratory of Computational Neuroscience, EPFL
CH-1015 Lausanne, Switzerland

Wu S. Zihan (zihan.wu@epfl.ch)

Laboratory of Computational Neuroscience, EPFL
CH-1015 Lausanne, Switzerland

Guillaume Bellec (guillweb@gmail.com)

Laboratory of Computational Neuroscience, EPFL
CH-1015 Lausanne, Switzerland

Wulfram Gerstner (wulfram.gerstner@epfl.ch)

Laboratory of Computational Neuroscience, EPFL
CH-1015 Lausanne, Switzerland

Abstract

It is known that different cortical areas have different critical periods for the most fundamental learning. However, the type of developmental plasticity rules that lead to high-level representations of objects are unknown. Here, we study a model of the visual ventral stream trained by a generalized Hebbian plasticity rule. The learning rule uses only quantities that are locally available at the site of the synapse, is consistent with recent plasticity experiments in pyramidal neurons, and, as opposed to the backpropagation algorithm, does not need a detailed feedback architecture. Our model shows that limiting plasticity in time to critical periods of development improves the quality of learned representation. Our model achieves state-of-the-art performance for bio-plausible plasticity models on the STL10 large image dataset designed for unsupervised learning.

Keywords: Hebbian plasticity, critical periods, representation learning, predictive coding

Introduction

The visual ventral stream transforms the high dimensional visual input into abstract hierarchical representations (Hubel & Wiesel, 1962; DiCarlo, Zoccolan, & Rust, 2012). During development, such cortical representations are learnt most rapidly at different critical periods (Knudsen, 2004), i.e., specific time windows during which a given cortical area is highly plastic and particularly sensitive to specific aspects of sensory stimuli (Kral, 2013).

Yet, the synaptic plasticity rules that build the hierarchical representation in the visual ventral stream remain difficult to grasp (DiCarlo et al., 2012). The most powerful models of representations in the mammalian ventral stream are multi-layer neural networks optimized to do object recognition by the backpropagation (BP) algorithm (Yamins & DiCarlo, 2016), which is biologically not plausible. In contrast, classic models of Hebbian plasticity (Hebb, 1949), where the change of a synapse depends only on the state of pre- (pre) and postsynaptic (post) neurons, fail to explain the increase of abstraction across the visual ventral stream.

Predictive coding is a step towards reconciling learning in deep networks with biology (Friston, 2010). Despite the partial success of a recent study (Halvagal & Zenke, 2023), Hebbian or predictive learning rules struggle achieving a hierarchical representation, where features learnt in deeper cortical areas are both complex *and* useful for vision tasks, lagging behind networks trained with BP.

Recent plasticity experiments have reported that synaptic plasticity at basal dendrites of pyramidal neurons depends on input arriving in the apical dendrite (Aceituno, de Haan, Loidl, & Grewe, 2024), potentially conveying predictive signals from other neurons located in a different information processing pathway or a higher cortical area. Here we show that a generalized Hebbian learning rule that includes lateral predictive input into apical dendrites (Illing, Ventura, Bellec, & Gerstner, 2021) can be used in combination with critical periods of plasticity to build abstract representations across a hierarchy of

areas. More specifically, critical periods of plasticity achieve state-of-the-art performance on the image dataset STL-10 designed for unsupervised learning (Coates, Lee, & Ng, 2011) while also improving the learning efficiency.

Results and discussion

Hebbian Model with Lateral Prediction

We describe the visual stream from V1 to V2, V4, posterior, central, and inferior temporal cortex by a convolutional neural network with six layers where each network layer corresponds to one cortical area (Yamins & DiCarlo, 2016). Compared to the cortex, the architecture of our model neglects the detailed microcircuits of excitatory and inhibitory neurons within each area. But we take into account that, within each area, feed-forward input arriving at the basal dendrites of pyramidal neurons drives their activity while feedback signals arriving mainly in the apical dendrites (Harris & Shepherd, 2015), either directly or indirectly via higher-order thalamus (Lillicrap, Santoro, Marris, Akerman, & Hinton, 2020; Larkum, 2013), only influence learning. We consider in the model that the major source of feedback stems from neurons in the same cortical area (Fig. 1A), but we also considered model variants where feedback arrives from higher areas.

Learning is modeled by the CLAPP learning rule (Illing et al., 2021), which qualitatively agrees with synaptic plasticity observed in cortical pyramidal neurons. The plasticity of a synapse w_{ij} from a presynaptic neuron j onto the basal dendrite of a postsynaptic neuron i follows the following generalized Hebbian learning rule:

$$\Delta w_{ij} = \eta(t) \cdot \text{lat}_i(t - \Delta) \cdot \text{post}_i(t) \cdot \text{pre}_j(t) \quad (1)$$

Here pre_j is the firing rate of the presynaptic neuron, post_i is a post-synaptic term that is a function of the total (basal) synaptic input, and $\text{lat}_i = \sum_k w_{ik}^{\text{lat}} z_k$ is the activity at the apical dendrite. Δ denotes a small delay of reception and integration at the apical dendrites. The predictive input z_k received by neuron i is given by neurons of the retinotopic map that process patches located on the sides of the receptive field (RF) of i , departing slightly from the original implementation of CLAPP in which predictive input stems from the same retinotopic location after a within-object micro-saccade. See (Delrocq, Bellec, & Gerstner, 2024) for more technical details. The lateral connections w_{ik}^{lat} change with a Hebbian learning rule as in (Illing et al., 2021).

The direction of change (potentiation $\Delta w_{ij} > 0$ or depression $\Delta w_{ij} < 0$, Eq. 1) of a synapse onto the *basal* dendrite depends on the *apical* input of the postsynaptic neuron, consistent with experiments (Aceituno et al., 2024).

The synaptic plasticity rule formulated in Eq. (1) adjusts connectivity such that feedback input via the lateral connections matches the feedforward input. Thus, after learning, feedback input has a predictive role (Muckli et al., 2015; Keller, Roth, & Scanziani, 2020). In this model, brain signals triggered by large saccadic eye movements (McFarland, Bondy,

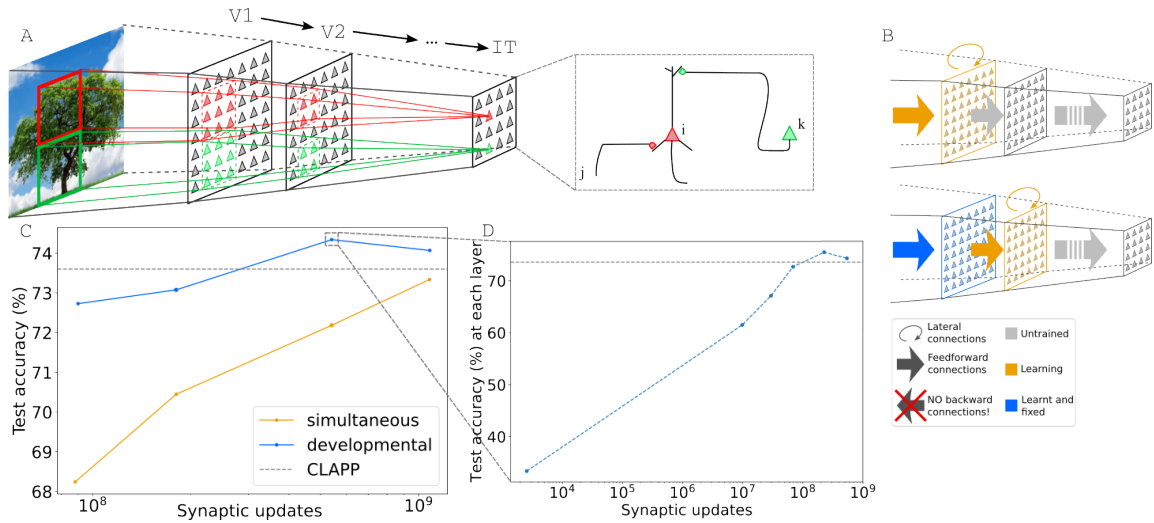


Figure 1: **Network architecture and critical periods.** A. Layered artificial neural network representing areas V1, V2, ..., IT. A neuron (red triangle) with an RF in the red-framed square receives lateral input from other neurons (green triangles) with non-overlapping RFs (in the green square). Zoom: The plasticity of the feedforward connection w_{ij} from neuron j in the preceding area to neuron i (red) depends on lateral input received by neuron i from other neurons k in the same area (green). B. Development learning is implemented as a sequence of critical periods (top to bottom) starting in V1 and finishing in IT. The orange color indicates connections to and inside the area that are amenable to learning during the critical period. C. Test accuracy for linear classification on STL10 as a function of synaptic updates. Developmental learning (blue) reaches better performances with fewer synaptic updates than simultaneous learning (orange). For comparison, the performance of CLAPP as reported in (Illing et al., 2021) is indicated by the dashed line. Note that points with more synaptic updates are not the continuation of the training of previous points: rather, for each point the total number of epochs is equally distributed among layers 2-6 (save for layer 1 with a shorter, fixed critical period – 20 epochs corresponding to 2560 updates, which we found more useful than longer training). D. Inset: abstraction increases hierarchically with depth (over 5 layers). Decoding accuracy is measured from the output of each layer at the end of its critical period, during the training of our best developmental model (therefore, layer 1 is the first point, layer 2 the second point, etc; the last point (layer 6) corresponds to the point in C.)

Saunders, Cumming, & Butts, 2015) switch the sign of plasticity during transitions between different images: in this case, the lateral input is contrastive, rather than predictive.

Consistent with experimental work (DiCarlo et al., 2012) and earlier work in machine learning (Illing et al., 2021), our model shows an increasing abstraction level of representations across layers/brain areas, as measured by linear decodability of object classification (Fig. 1D). The usefulness of the representations increases with depth although during training we neither provide the identity of objects nor do we allow any backpropagation of information from IT to lower areas.

Critical Periods improve learning

In sensory pathways such as vision (Hensch, 2005) or audition (Kral, 2013) synaptic plasticity is largest during critical periods. It has been suggested that these critical periods are staggered sequentially in order to support the extraction of abstract features (Reh, Dias, Nelson, & Hensch, 2020).

In order to test the hypothesis that sequential staggering of critical periods across the cortical hierarchy supports learning, we compare a sequential protocol of developmental learning with a control model where all areas learn in parallel (Fig. 1B-C). We implement a developmental sequence of critical peri-

ods in our network model by training first V1 with our lateral-input dependent plasticity rule. After a predefined number of training steps, the critical window for area V1 closes and weights to and within area V1 are kept fixed. Then, the critical period for the second area opens. This sequence ends with the critical period for area IT (Fig. 1B). The model is trained and tested on the image dataset STL-10, designed for unsupervised learning (Coates et al., 2011).

We find that networks trained with sequential critical periods achieve representations that are better than a control network trained in the standard fashion, i.e. where all weights across all areas are learned simultaneously. More importantly, our experiments show that learning with a staggered sequence of critical periods is far more efficient in terms of resources, measured as the total number of synaptic changes across the network, than simultaneous learning (Fig. 1C). For comparison, the learning algorithm LPL (Halvagal & Zenke, 2023), another fully bio-plausible self-supervised algorithm, achieves 62% test accuracy on STL-10 (not shown). Note that in natural environments, visual input is abundant making not the amount of data, but rather potentially costly changes of synaptic connections the limiting factor.

Acknowledgments

We gratefully acknowledge the Swiss National Science Foundation (SNSF) for funding this research via the grants 200020-207426 and 200021-236436.

References

- Aceituno, P., de Haan, S., Loidl, R., & Grewe, B. (2024). Target learning rather than backpropagation explains learning in the mammalian neocortex. *bioRxiv, preprint*, DOI 10.1101/2024.04.10.588837.
- Coates, A., Lee, H., & Ng, A. (2011). An analysis of single-layer networks in unsupervised feature learning. *AISTATS and JMLR:W/CP*, 15, 215-233.
- Delcrocq, A., Bellec, G., & Gerstner, W. (2024). Critical periods support representation learning in a model of visual processing. *bioRxiv*, xx, xx.
- Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.*, 11, 127-138.
- Halvagal, M., & Zenke, F. (2023). The combination of hebbian and predictive plasticity learns invariant object representations in deep sensory networks. *Nat. Neurosci.*, 26, 1906-1915.
- Harris, K., & Shepherd, G. (2015). The neocortical circuit: themes and variations. *Nat. Rev. Neurosci.*, 18, 170-181.
- Hebb, D. O. (1949). *The Organization of Behavior*. New York: Wiley.
- Hensch, T. (2005). Critical period plasticity in local cortical circuits. *Nat. Rev. Neurosci.*, 6, 877-888.
- Hubel, D., & Wiesel, T. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, 160(1), 106-154.
- Illing, B., Ventura, J., Bellec, G., & Gerstner, W. (2021). Local plasticity rules can learn deep representations using self-supervised contrastive predictions. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, & J. W. Vaughan (Eds.), *Advances in neural information processing systems* (Vol. 34, pp. 30365-30379). Curran Associates, Inc.
- Keller, A., Roth, M., & Scanziani, M. (2020). Feedback generates a second receptive field in neurons of the visual cortex. *Nature*, 582, 545-549.
- Knudsen, E. (2004). Sensitive periods in the development of the brain and behavior. *J. Cogn. Neurosci.*, 16, 1412-1425.
- Kral, A. (2013). Auditory critical periods: a review from system's perspective. *Neurosci.*, 247, 117-133.
- Larkum, M. (2013). A cellular mechanism for cortical associations: an organizing principle for the cerebral cortex. *Trends Neurosci.*, 36, 141-151.
- Lillicrap, T., Santoro, A., Marris, L., Akerman, C., & Hinton, G. (2020). Backpropagation and the brain. *Nat. Rev. Neurosci.*, 21, 335-346.
- Muckli, L., De Martino, F., Vizioli, L., Petro, L., Swmith, F., Ugurbil, K., ... Yacoub, E. (2015). Contextual feedback to superficial layers of v1. *Current Biology*, 25, 2690-2695.
- DiCarlo, J., Zoccolan, D., & Rust, N. (2012). How does the brain solve visual object recognition? *Neuron*, 73, 415-434.
- McFarland, J., Bondy, A. G., Saunders, R., Cumming, B., & Butts, D. (2015). Saccadic modulation of stimulus processing in primary visual cortex. *Nat. Commun.*, 6, 1-14.
- Reh, R., Dias, B., Nelson, C., & Hensch, T. (2020). Critical period regulation across multiple timescales. *Proc. Natl. Acad. Sci. (USA)*, 117. (23242-23251)
- Yamins, D., & DiCarlo, J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nat. Neurosci.*, 19, 356-365.