Modeling Dynamical Vision: A Toolbox for Biologically Plausible Recurrent Convolutional Networks

Robin Gutzen (robin.gutzen@nyu.edu)
Center for Data Science, New York University, New York, USA
Grace Lindsay (gm3239@nyu.edu)
Center for Data Science and Department of Psychology, New York University, New York, USA

Abstract

Convolutional Neural Networks (CNNs) have demonstrated remarkable success in image recognition and exhibit conceptual similarities to the primate ventral visual pathway. Adding recurrence opens the door to exploring temporal dynamics and investigating mechanisms underlying recognition robustness, attentional modulation, and rhythmic perception phenomena. However, modeling spatiotemporal dynamics of biological vision using CNN-based architectures remains challenging. Incorporating functionally beneficial recurrence, capturing biologically plausible temporal phenomena such as adaptation and subadditive temporal summation, and maintaining topographic organization aligned with cortical structure require significant computational considerations. Although recent advances have incorporated neurobiological constraints, the field lacks accessible tools for efficiently integrating, testing, and comparing these approaches. Here, we introduce DynVision, a modular toolbox for constructing and evaluating recurrent convolutional neural networks (RCNNs) with biologically inspired dynamics. Our approach facilitates the incorporation of key visual cortex properties, including realistic recurrent architectures, activity evolution governed by dynamical systems equations, and structured connectivity reflecting cortical arrangements, while maintaining computational efficiency. We demonstrate the framework's utility through systematic analysis of emergent neural dynamics, highlighting how different biologically motivated modifications shape response patterns characteristic of cortical recordings.

Keywords: Vision Models; Dynamic Systems; RCNN; NeuroAI

Introduction

The primate visual system is characterized by abundant recurrent connections (van Bergen & Kriegeskorte, 2020). In the ventral visual stream, lateral recurrent connections exist amongst neurons within visual cortical regions and feedback connections go from higher areas like V4 back to lower ones such as V1. These connections play a crucial role in the system's ability to integrate information over time and recognize objects under diverse viewing conditions.

Many researchers have attempted to integrate these connections into convolutional neural networks (Lindsay, 2021). Although studies on recurrent convolutional neural networks (RCNNs) have found benefits of adding recurrence (Kietzmann et al., 2019; Spoerer et al., 2020; Kar et al., 2019), others have had mixed results (Maniquet et al., 2024; Nayebi et al., 2022; Lindsay et al., 2022). These studies predominantly use discrete-time recurrent models, usually unrolled for only a handful of time steps. This coarse-grain approximation cannot capture the full complexity of visual dynamics and conflicts with traditional computational neuroscience methods, which treat neural circuits as continuous dynamical systems.

Using continuous-time models in machine learning poses engineering challenges. Neural dynamics require small simulation step sizes (~ 1 millisecond), necessitating tens to hundreds of time steps to mimic visual system responses. This creates large computational graphs when training with backpropagation through time. Here, inspired by recent work combining continuous-time differential equations with deep convolutional neural networks (Soo et al., 2024; Heeger & Mackey, 2019; Lindsay et al., 2020), we build a toolbox to make such models more accessible.

DynVision implements numerical ODE solvers and heterogeneous delays for different connection types to build RCNNs with precise and realistic temporal dynamics. It allows many different forms of recurrent connections with biologically inspired options that enhance efficiency, while providing control over parameters governing dynamics within and across regions.

Methods

Toolbox Design

The "DynVision" toolbox emphasizes modularity, adaptability, and reusability to provide an efficient RCNN modeling environment. It leverages PyTorch for tensor operations, PyTorch Lightning for training procedures, FFCV for optimized data loading, Snakemake for workflow management, and Pydantic classes and YAML files for parameter handling.

Biologically-inspired Model Components

- **Recurrent Connections:** A module combines feedforward 2D convolution with recurrency operations and handles hidden state storage with variable time delays. We include *self-recurrence* (Kietzmann et al., 2019), *full recurrence* (Liang & Hu, 2015), and *pointdepthwise recurrence* using depthwise separable convolutions (Chollet, 2017) that map onto lateral recurrence structure in the visual system (Fig. 1). We also introduce *local recurrence* that captures 2-D cortical topology by arranging units on grids inspired by cortical organization (Ohki et al., 2006; Qian et al., 2024).

- **Skip and Feedback Connections:** These link units across layers and time steps. The toolbox offers functionality to auto-adapt signal transformations during the first forward pass by just defining source and target layers.

- **Dynamical Systems Description:** Continuous-time dynamics are realized via numerical ODE solvers (Euler method) that evolve layer activity on timescale τ based on previous activity (Fig.2). This provides independent control over temporal precision (dt) and time constants (τ).

- **Biological Unrolling with Heterogeneous Delays:** The framework allows separate time delays for feedforward (Δ_{FF}) and recurrent connections (Δ_{RC}), enabling systematic evaluation of different temporal delays on network dynamics (including realizing engineering time with $\Delta_{FF} = 0$).

- **Additional components** include retina-inspired preprocessing, activity regularization via energy loss functions Butkus et al. (2024), supralinear activation functions Rubin et al. (2015); Lindsay et al. (2019), flexible execution order of layer operations and activity recording, reference model implementations, extensive logging and analytics, portability and seamless scalability from CPU to GPU to multi-GPU, ...

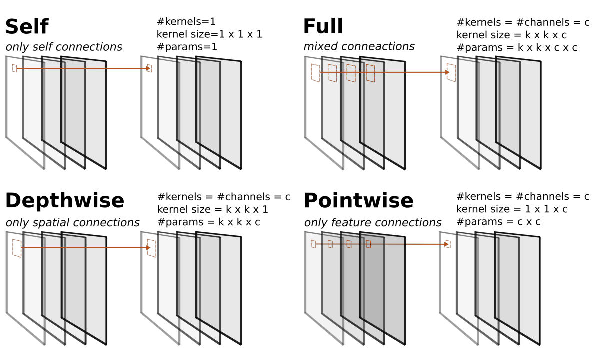

Figure 1: Four types of kernel convolutions used to realize recurrence: self, full, depthwise, pointwise. Each shows a 3D tensor representing layer output at two time steps Δ_{RC} apart.

A)

Dynamical Timescale Network Activations External Input

$$\tau \frac{dr}{dt} = -r(t) + \Phi(I(t) + J\,r(t))$$

Time Dimension Non-linearity Network Connectivity

B)

$$r(t) = r(t-dt) + \frac{dt}{\tau}\left(-r(t-dt) + \Phi[f(t-dt, r)]\right)$$

Figure 2: Dynamical systems formulation. **A)** Differential equation describing network activity evolution. **B)** Numerical Forward Euler method implementation for activity r.

Results

Benchmarking

We retrained CORNet-RT (Schrimpf et al., 2020) on ImageNette (10-class ImageNet subset), comparing the original implementation with our toolbox integration. The toolbox achieved a 52% speedup (8.86s vs 13.51s per epoch) on a NVIDIA A100 GPU.

Investigating Temporal Dynamics

We demonstrate the toolbox's capabilities to systematically investigate biologically plausible temporal dynamics by examining how different recurrency types affect neural response patterns (Fig. 3). Following Groen et al. (2022), which documented cortical responses to stimuli of varying contrast, duration, and interval spacing, we evaluate time courses of layer activations as proxies for ECoG measurements.

Models trained on CIFAR100 with different recurrence types exhibit distinct temporal characteristics. All models demonstrate stability over extended periods and reasonable baseline null response. Models with full recurrence show response patterns most similar to cortical recordings, including subadditive temporal summation and adaptation effects.

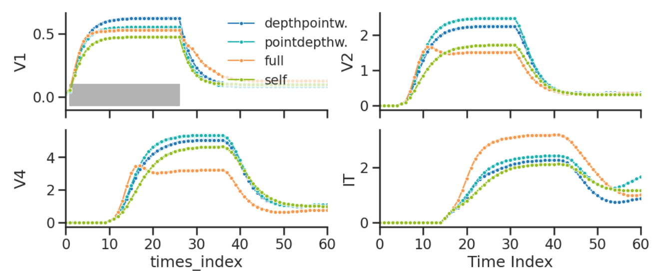

Figure 3: Population temporal dynamics of a four-layer RCNN model trained on CIFAR100 with 20 timesteps, tested on 60 steps with stimulus presented for steps 1-26. Accuracies: full 71.7%, self 64.7%, pointdepthwise 65.3%, depthpointwise 66.71.7%

Summary and Continued Development

We introduce DynVision, a toolbox for building dynamical systems-based RCNNs, and systematic explorations of how biologically-inspired architectural choices influence temporal dynamics. Capturing fine-grained temporal dynamics may be necessary for creating models that use recurrence similarly to the brain.

Future developments include excitatory/inhibitory neuron separation, local field potential modeling, and extended support for *in silico* experiments. By lowering barriers to working with compute-intensive, DynVision aims to accelerate discoveries about neural mechanisms underlying perception.

Code availability

`https://github.com/Lindsay-Lab/DynVision`

References

Butkus, E., Ying, Z., Chen, P., & Kriegeskorte, N. (2024). Recurrent attentional selection can explain flexible trading of accuracy and energy in biological vision. In *CCN 2024.* Boston, MA.

Chollet, F. (2017, July). Xception: Deep Learning with Depthwise Separable Convolutions. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1800–1807). Honolulu, HI: IEEE. doi: 10.1109/CVPR.2017.195

Groen, I. I. A., Piantoni, G., Montenegro, S., Flinker, A., Devore, S., Devinsky, O., ... Winawer, J. (2022, October). Temporal Dynamics of Neural Responses in Human Visual Cortex. *Journal of Neuroscience*, *42*(40), 7562–7580. doi: 10.1523/JNEUROSCI.1812-21.2022

Heeger, D. J., & Mackey, W. E. (2019, November). Oscillatory recurrent gated neural integrator circuits (ORGaNICs), a unifying theoretical framework for neural dynamics. *Proceedings of the National Academy of Sciences*, *116*(45), 22783–22794. doi: 10.1073/pnas.1911633116

Kar, K., Kubilius, J., Schmidt, K., Issa, E. B., & DiCarlo, J. J. (2019, June). Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior. *Nature Neuroscience*, *22*(6), 974–983. doi: 10.1038/s41593-019-0392-5

Kietzmann, T. C., Spoerer, C. J., Sörensen, L. K. A., Cichy, R. M., Hauk, O., & Kriegeskorte, N. (2019, October). Recurrence is required to capture the representational dynamics of the human visual system. *Proceedings of the National Academy of Sciences*, *116*(43), 21854–21863. doi: 10.1073/pnas.1905544116

Liang, M., & Hu, X. (2015, June). Recurrent convolutional neural network for object recognition. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 3367–3375). doi: 10.1109/CVPR.2015.7298958

Lindsay, G. W. (2021, September). Convolutional Neural Networks as a Model of the Visual System: Past, Present, and Future. *Journal of Cognitive Neuroscience*, *33*(10), 2017–2031. doi: 10.1162/jocn_a_01544

Lindsay, G. W., Moskovitz, T., Yang, G. R., & Miller, K. (2019). Do Biologically-Realistic Recurrent Architectures Produce Biologically-Realistic Models? In *2019 Conference on Cognitive Computational Neuroscience.* Berlin, Germany: Cognitive Computational Neuroscience. doi: 10.32470/CCN.2019.1175-0

Lindsay, G. W., Mrsic-Flogel, T. D., & Sahani, M. (2022, March). *Bio-inspired neural networks implement different recurrent visual processing strategies than task-trained ones do* (Preprint). Neuroscience. doi: 10.1101/2022.03.07.483196

Lindsay, G. W., Rubin, D. B., & Miller, K. D. (2020, July). *A unified circuit model of attention: Neural and behavioral effects.* bioRxiv. doi: 10.1101/2019.12.13.875534

Maniquet, T., Op de Beeck, H., & Ivan Costantino, A. (2024). Recurrent issues with deep neural networks of visual recognition. In *CCN 2024.* Boston, MA.

Nayebi, A., Sagastuy-Brena, J., Bear, D. M., Kar, K., Kubilius, J., Ganguli, S., ... Yamins, D. L. K. (2022, July). Recurrent Connections in the Primate Ventral Visual Stream Mediate a Trade-Off Between Task Performance and Network Size During Core Object Recognition. *Neural Computation*, *34*(8), 1652–1675. doi: 10.1162/neco_a_01506

Ohki, K., Chung, S., Kara, P., Hübener, M., Bonhoeffer, T., & Reid, R. C. (2006, August). Highly ordered arrangement of single neurons in orientation pinwheels. *Nature*, *442*(7105), 925–928. doi: 10.1038/nature05019

Qian, X., Dehghani, A. O., Farahani, A. B., & Bashivan, P. (2024, August). *Local lateral connectivity is sufficient for replicating cortex-like topographical organization in deep neural networks.* bioRxiv. doi: 10.1101/2024.08.06.606687

Rubin, D. B., Van Hooser, S. D., & Miller, K. D. (2015, January). The stabilized supralinear network: A unifying circuit motif underlying multi-input integration in sensory cortex. *Neuron*, *85*(2), 402–417. doi: 10.1016/j.neuron.2014.12.026

Schrimpf, M., Kubilius, J., Hong, H., Majaj, N. J., Rajalingham, R., Issa, E. B., ... DiCarlo, J. J. (2020, January). *Brain-Score: Which Artificial Neural Network for Object Recognition is most Brain-Like?* bioRxiv. doi: 10.1101/407007

Soo, W. W., Battista, A., Radmard, P., & Wang, X.-J. (2024, November). Recurrent neural network dynamical systems for biological vision. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems.*

Spoerer, C. J., Kietzmann, T. C., Mehrer, J., Charest, I., & Kriegeskorte, N. (2020, October). Recurrent neural networks can explain flexible trading of speed and accuracy in biological vision. *PLOS Computational Biology*, *16*(10), e1008215. doi: 10.1371/journal.pcbi.1008215

van Bergen, R. S., & Kriegeskorte, N. (2020, December). Going in circles is the way forward: The role of recurrence in visual inference. *Current Opinion in Neurobiology*, *65*, 176–193. doi: 10.1016/j.conb.2020.11.009