Continual learning using dendritic modulations on view-invariant feedforward weights

Viet Anh Khoa Tran, Emre Neftci, Willem A. M. Wybo

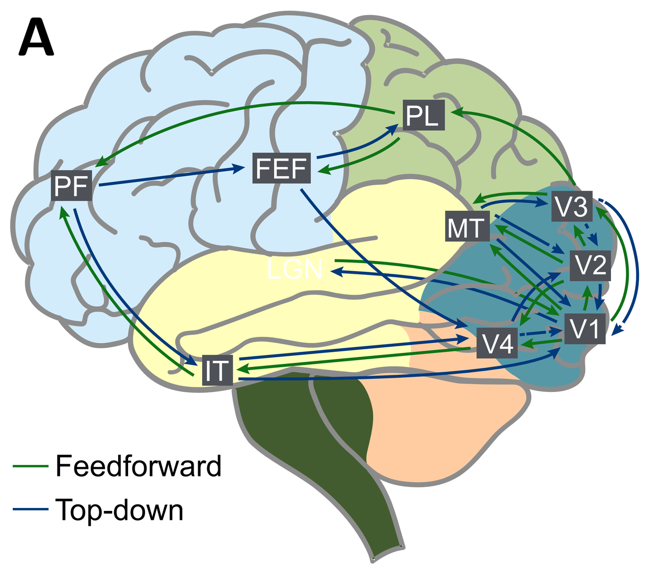
Neuromorphic Software Ecosystems (PGI-15), Forschungszentrum Jülich, Germany.

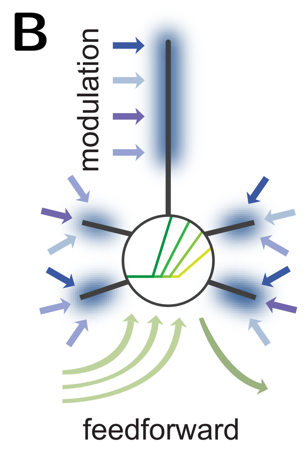
Visual Processing in Biological Brains

- Visual processing in the brain consists of a **rapid feedforward pass** generating generic task-independent features, while complex visual tasks require **top-down inputs** (**A**, [1]).
- Comparatively longer-lasting top-down afferents impinging on dendritic branches have been proposed to dynamically reshape feedforward computation to solve new tasks (**B**, [2]).
- This kind of computation can be modeled as

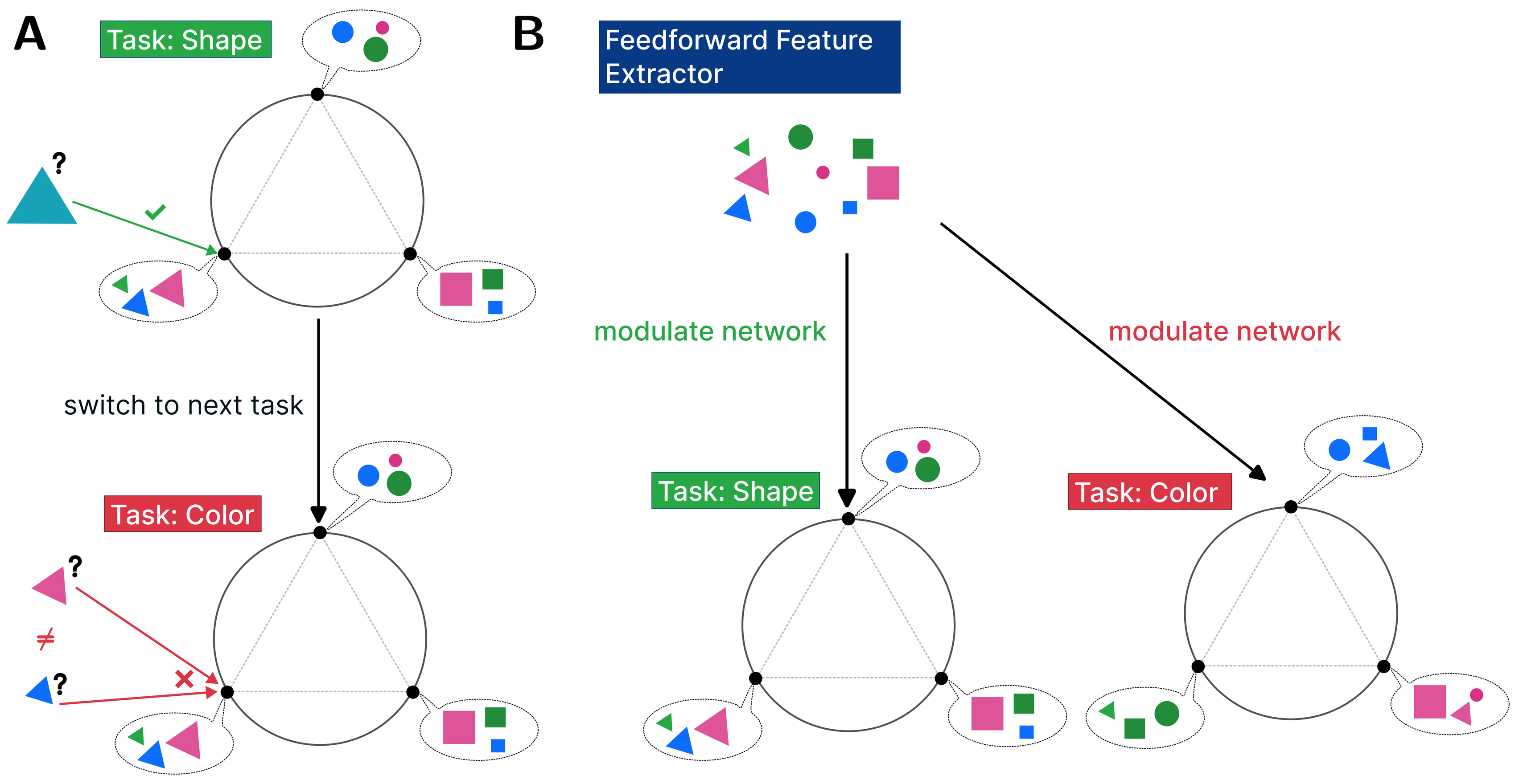
$$\sigma(g_t \odot (\mathbf{W} \cdot \mathbf{x}) + b_t).$$

- Newborn chicks reared in environments with smoothly moving visual objects develop **view-invariant recognition capabilities**, but are sensitive to view-changes when reared with temporally non-smoothly moving objects [3].

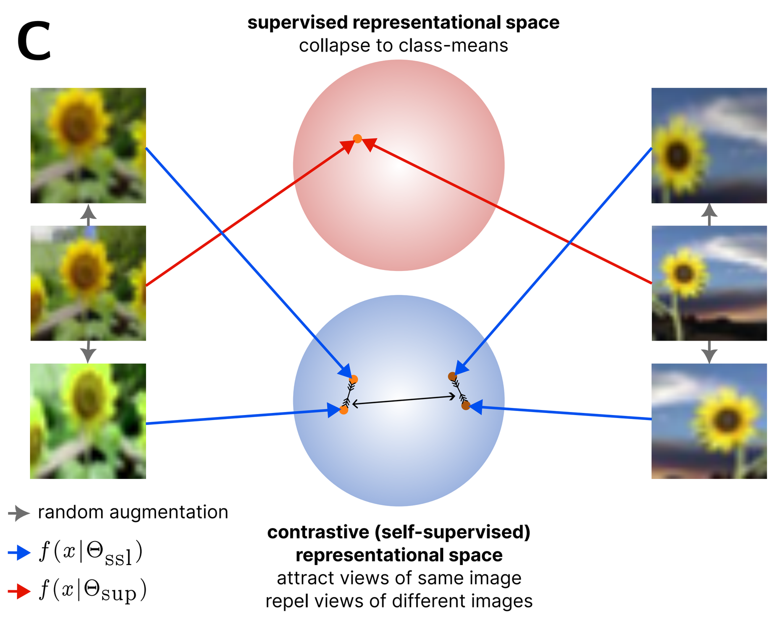




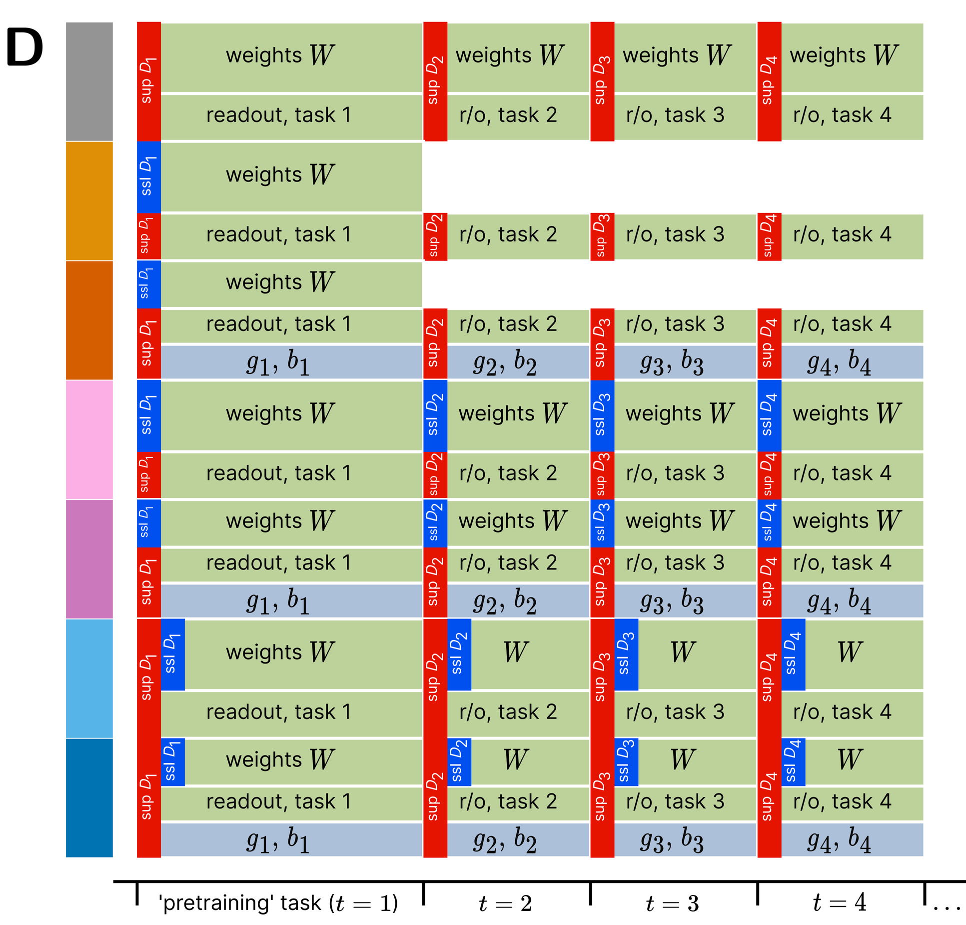
Towards a Biologically Plausible Continual Learner



- The baseline approach (■) is to continually train using only the supervised objective $\mathcal{L}_{\text{sup}}$. This inherently introduces catastrophic forgetting, since the optimum of supervised classification is to collapse to the mean-class representation (**A**, [4]).



- Contrastive self-supervised learning approaches ($\mathcal{L}_{\text{ssl}}$, **C**) approximate view-invariance by pulling together representations of distorted views of the same image, while repelling views of different images.
- Still, the theoretical optimum for solving classification tasks is obtained with the collapse phenomenon. We therefore introduce dendrite-inspired task-dependent modulations trained in a supervised fashion (■■■, **D**).
- A simple trick to completely avoid catastrophic forgetting is to freeze the task-shared weights (■■), instead of continually training those on the current task (■■■■■).
- We furthermore investigate combining both $\mathcal{L}_{\text{ssl}}$ and $\mathcal{L}_{\text{sup}}$, intuitively making the feedforward weights "task-aware" while avoiding catastrophic forgetting (■■).



Objectives

sup D_t: $\mathcal{L}_{\text{sup}} := \mathrm{CE}((\text{readout} \circ f)(X_t;\ \Theta, \Psi_t) \| Y_t)$

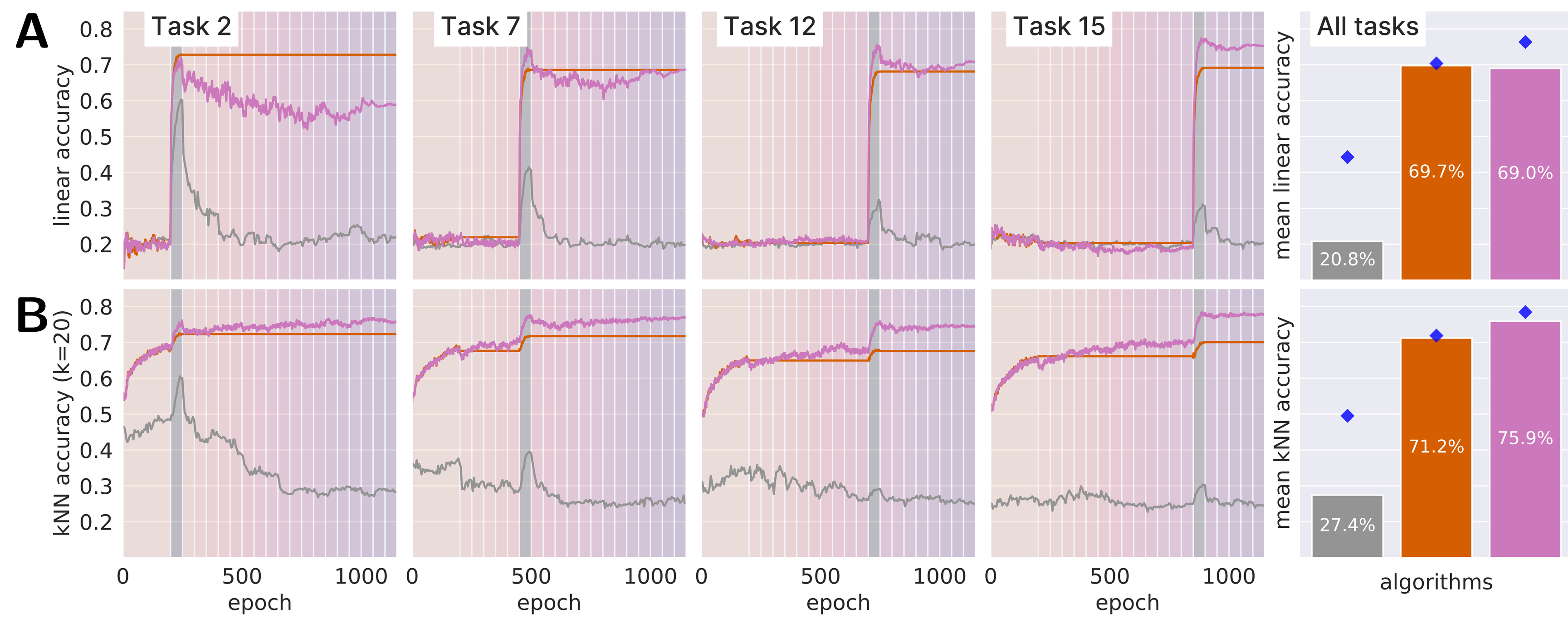
ssl D_t: $\mathcal{L}_{\text{ssl}} := \mathrm{BarlowTwins}((g_{\text{ssl}} \circ f)([X_t', X_t''];\ \Theta, \varnothing))$

sup D_t, ssl D_t: $\mathcal{L}_{\text{ssl}} + \lambda \cdot \mathcal{L}_{\text{sup}}$

Algorithms

- supervised weights (baseline)
- task-agnostic ssl-weights, frozen
  - ↳ with modulations
- task-agnostic ssl-weights, continual
  - ↳ with modulations
- task-aware ssl-weights, continual
  - ↳ with modulations

Continual self-supervised learning retains class clusters

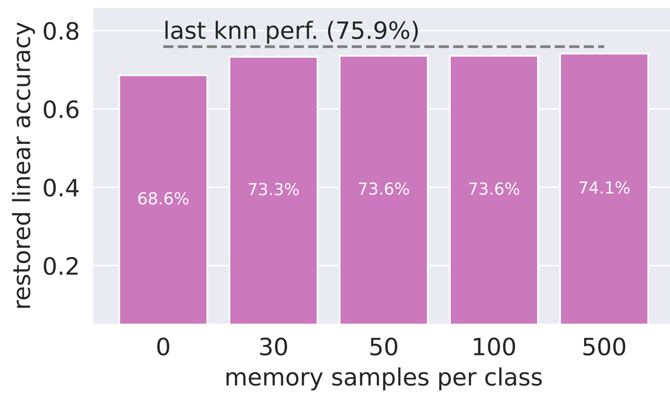


- When evaluating the task-specific linear readout, the modulated continual learner (■) outperforms its frozen counterpart (■) on per-task peak performance (**A**, ◆: mean over per-task peaks).
- However, task performance deteriorates as novel tasks are trained (bars: final mean-task performance).
- Yet, this is not catastrophic forgetting. Evaluating with the k-nearest neighbor algorithm (kNN), the continual learner is close to its peak performance (**B**). **We hypothesize that the continual learning retains clustered representations of prior tasks, but that they drift due to the continual learning.** Thus, performance of the fixed linear readout, but not of kNN, deteriorates.

Recovering cluster readout by least-squares regression

- During training of task t, we store m images per class via reservoir sampling ($\mathbf{X}_t$). At the end of each session, we evaluate and store the logits of all memory samples of the current task, i.e.,

$$\mathbf{L}_t = \mathbf{R}_t f(\mathbf{X}_t).$$

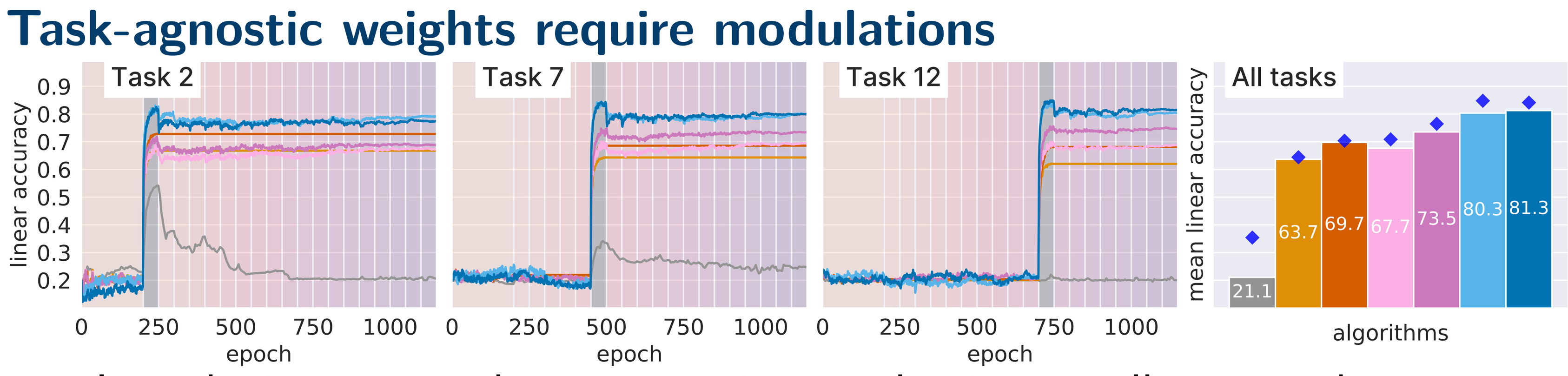


- At every training step, we perform readout restoration on all previous tasks via Tikhonov-regularized least-squares regression towards $\mathbf{R}_t$

$$\min_{\mathbf{R}_t} \|\mathbf{R}_t f(\mathbf{X}_t) - \mathbf{L}_t\| + \lambda \|\mathbf{R}_t - \mathbf{R}_t^{\text{curr}}\|^2.$$

- With few samples, we are almost able to obtain kNN performance.

Task-agnostic weights require modulations



- Introducing our readout restoration mechanism to all proposed algorithms, we show that continual self-supervised weights (■■) outperform frozen self-supervised (■■) and continual supervised weights (■), while mostly avoiding catastrophic forgetting.
- Results suggest with **task-agnostic weights, task-specific supervised modulations** (■■) **are necessary to approach optimal task performance**, as they significantly outperform the unmodulated counterparts (■■).
- **With task-aware weights** (■■), **the impact of modulations is decreased.** We hypothesize that the given setup (CIFAR-100 split into 20 random tasks) is too simple and the generated classification tasks have similar invariances.

Conclusion

- Our work indicates that SSL could be an appropriate model for continual learning in the brain, as it can continuously incorporate new data while only minimally forgetting previously learning concepts.
- Dendritically inspired modulations allow for optimally solving tasks on task-agnostic weights, which we consider more bioplausible than task-aware weights.

References

[1] Thomas Serre, Aude Oliva, and Tomaso Poggio. "A feedforward architecture accounts for rapid categorization". In: *Proceedings of the National Academy of Sciences* 104.15 (2007).

[2] Willem AM Wybo, Matthias C Tsai, Viet Anh Khoa Tran, Bernd Illing, Jakob Jordan, Abigail Morrison, and Walter Senn. "NMDA-driven dendritic modulation enables multitask representation learning in hierarchical sensory processing pathways". In: *Proceedings of the National Academy of Sciences* 120.32 (2023).

[3] Justin N. Wood and Samantha M. W. Wood. "The Development of Invariant Object Recognition Requires Visual Experience With Temporally Smooth Objects". In: *Cognitive Science* 42.4 (2018).

[4] Zhihui Zhu, Tianyu Ding, Jinxin Zhou, Xiao Li, Chong You, Jeremias Sulam, and Qing Qu. "A geometric analysis of neural collapse with unconstrained features". In: *Advances in Neural Information Processing Systems* 34 (2021).