Curriculum Design for Scalable Biologically Plausible Deep Reinforcement Learning

Alexandra R. van den Berg[1,2], Pieter R. Roelfsema[2,3,4,5], Sander M. Bohte[1,6*]
[1]*Machine Learning Group, Centrum Wiskunde & Informatica, Amsterdam, the Netherlands*
[2]*Department of Vision & Cognition, Netherlands Institute for Neuroscience, Amsterdam, the Netherlands*
[3]*Department of Integrative Neurophysiology, Center for Neurogenomics and Cognitive Research, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands*
[4]*Department of Neurosurgery, Academic Medical Center, Amsterdam, the Netherlands*
[5]*Laboratory of Visual Brain Therapy, Sorbonne Université, Institut National de la Santé et de la Recherche Médicale, Centre National de la Recherche Scientifique, Institut de la Vision, Paris, France*
[6]*Swammerdam Institute of Life Sciences, University of Amsterdam, Amsterdam, the Netherlands*

***Abstract*—Humans have a remarkable capacity for learning, yet neuronal learning is constrained to locality in time and space and limited feedback. While neural learning rules have been designed that adhere to these principles and constraints, they exhibit difficulty in scaling to deep networks and complicated datasets. BrainProp is a biologically plausible learning rule, learning from trial-and-error feedback through reinforcement learning, that does generalise to deep networks and achieves good performance on traditional machine learning benchmarks. It does however falter on problems with a large number of output categories, such as the classical ImageNet vision benchmark: while standard BrainProp eventually succeeds, learning is not robust and highly sensitive to hyper-parameter optimisation and proper initialisation. Here, we leverage insights from behavioural science by developing a curriculum that structures how samples are presented to a network to optimise learning. The key features of the curriculum involve progressively introducing new classes to the dataset based on performance metrics, and using a recency bias to protect recently acquired classes. We demonstrate that our curriculum approach makes BrainProp-style learning robust and more rapid, while substantially improving classification accuracy. We also show the curriculum similarly improves performance for networks trained using error-backpropagation. We thus establish a new state-of-the-art performance for large-scale deep reinforcement learning. Our results show the potential of curriculum learning in local learning settings with limited feedback and further bridges the gap between biologically plausible learning rules and error-backpropagation.**

***Keywords*—Curriculum Learning, Local Learning, Biologically Plausible Learning Rule, Neuroscience, Reinforcement Learning**

I. Introduction

The human brain is remarkable in its capacity for learning. By modelling the brain, we can gain both a deeper understanding of biology, and biology itself also provides a source of inspiration for AI. In particular for AI constrained to efficient and effective learning rules in settings with limited feedback. How does learning differ between biological and artificial systems?

One crucial way how animal learning differs from AI is the source and locality of information used for neural learning. Credit assignment in classical error-backpropagation (EBP) is performed in a supervised manner, where each synapse is informed about its contribution to the output across all layers of the neural network, as well as about what the output should have been when errors are made. However, this premise is biologically implausible for two reasons. Firstly, it assumes that each neuron has access to this non-local information. Secondly, a large amount of learning in humans and animals occurs without explicit feedback about which outputs should have been generated when errors are made. Instead, the environment either provides a reward – or not – and one has to infer through experience what the correct output (or action) should be [1]. Thus, we hereby consider "biologically plausible" to mean learning that is done through reinforcement and using information available only locally, at the level of the synapse.

To overcome these limitations, BrainProp was proposed as a biologically plausible learning scheme for training deep neural networks. BrainProp achieves state-of-the-art performance on traditional machine learning benchmarks such as CIFAR-100 and TinyImageNet [2] when set in a reinforcement learning paradigm. BrainProp learns exclusively through trial-and-error, and has to discover the correct output for each class. Training on datasets with many classes therefore becomes very challenging. If the model selects the wrong class, the model first has to unlearn this class and sample others until the right class is found. As the number of classes grows, the probability diminishes that the model guesses correctly, thereby making the class selection problem exceedingly hard. As a result, biologically plausible reinforcement learning methods have thus far been unable to scale to the size and complexity of benchmarks such as the ImageNet vision classification benchmark, which requires deep networks for the classification of 1,000 categories, despite extensive effort. Here, we show

Funding support provided by NWO (ARvdB and SMB, NWO-NWA grant NWA.1292.19.298) and the European Union (ARvdB, SMB and PRR, grant agreement 7202070 "HBP")
∗Corresponding author: sbohte@cwi.nl

a way to overcome this problem by drawing inspiration from how biological systems learn.

In contrast to AI, where networks are often repeatedly exposed to large amounts of unstructured data, humans typically learn in a very structured manner. Rather than immediately attempting high-level mathematics, children first learn how to perform simple addition and subtraction before they are exposed to increasingly complex problems. These types of curricula form the foundation of modern education systems and are designed to facilitate efficient learning [3]. The power of using a curriculum becomes even more evident when training animals, since one cannot rely on verbal instructions. In a series of classic experiments in the early 1900s, Skinner already showed that by initially rewarding animals for exhibiting very simple motions and afterwards rewarding them for more complex behavioural sequences (i.e. "shaping"), animals quickly acquire new behaviour that they might otherwise learn very slowly – if at all. For this reason, using a curriculum is considered standard practice when training animals on behavioural tasks [4], [5].

Within AI, numerous studies have also demonstrated the beneficial effects of using curricula for reinforcement learning [6], [7]. One potential contributor to the success of curriculum learning regards the difficulty level of training examples, which has been shown to be an important influence on training efficiency [8]–[10]. A curriculum could allow for initial optimisation of an easy function, which can then progressively be refined as it becomes more non-convex through more difficult examples [11]. It has been suggested that the most optimal training regime is one in which training accuracy hovers around 85% [12]; thus not being too difficult, but also not too easy. A curriculum can be used to dynamically adjust the difficulty of training examples to achieve this state. Another advantage of curriculum learning concerns the acquisition of primitives underlying more complex behaviour, which could facilitate learning the overarching behaviour [13]–[15], and result in networks behaving more similarly to animals on the same task [14]. In fact, given that associative learning in mice has been shown to occur in a stepwise manner [16], forcing an artificial neural network to learn in discrete steps using a curriculum might therefore not only simplify the learning process, but also result in more biologically plausible learning.

We here developed a novel curriculum to overcome the class-learning problem for biologically plausible learning rules based on reinforcement learning – such as BrainProp. The curriculum, specifically a form of structured class-incremental curriculum learning, operates by gradually introducing each class to the network. We first demonstrate that with precise initialisation and optimisation BrainProp can scale to ImageNet, but that the standard trial-and-error learning is very slow, and unstable because only few initialisations converge, and networks that appear to learn well regularly collapse. Next, we show that using the curriculum makes learning robust, more rapid and more accurate. We thereby establish a new state-of-the-art performance for biologically plausible reinforcement learning on the ImageNet benchmark, and demonstrate the ability of approaches such as BrainProp to scale to biologically relevant domain sizes and network complexities with local learning and limited feedback. Finally, we show that a curriculum also aids in the performance of EBP on ImageNet, and conduct an ablation analysis to investigate which elements are important for the curriculum's success.

II. A CURRICULUM FOR OVERCOMING THE CLASS-LEARNING PROBLEM

A. Biologically plausible learning

BrainProp is a biologically inspired reinforcement learning scheme for neural networks that performs credit assignment using feedback connections to successively lower layers in the network ([2]; see Fig. 1). We here briefly summarise its working mechanism and how this relates to the class-learning problem.

After choosing an action s based on output layer activations, the network receives reward information r from the environment. This is 1 if the correct choice was made, and 0 otherwise. The model then computes the reward prediction error (RPE; δ), which quantifies the discrepancy between the expected reward (the activation of the chosen unit y_s^N) and the actual reward (r):

$$E(w) = \frac{1}{2}(r - y_s^N)^2 \tag{1}$$

RPE signalling is believed to be performed by the dopaminergic system [17] and in BrainProp is broadcast to the network as a global neuromodulatory signal. The aim of learning is to minimise this RPE over the course of training. Importantly, the RPE is only calculated for the selected class (s) and only those weights that contributed to this action are updated. The resulting weight update for weights ($w_{n,m}^{N-1}$) connecting to the output layer (N) becomes:

$$E(w) = \Delta w_{n,m}^{N-1} = \begin{cases} \delta y_m^{N-1}, & \text{if } n = s \\ 0, & \text{if } n \neq s \end{cases} \tag{2}$$

For lower layers, a separate feedback network carries an attentional signal conveying activity (rather than error signals) from the selected output, gating plasticity in the forward network in a manner compatible with locally available information.

For these layers, weight updates rely not only on the RPE, but are also gated by the level of feedback neurons (ϕ_k^{l+1}) receive from higher layers. Feedback is only provided if a higher-level neuron was active and is propagated using feedback connections ($w_{j,i}^{FB}$) from higher to lower layers. These feedback connections are assumed to be reciprocal to the feedforward connections ($w_{i,j}^{FF}$) within a cortical column; a feature that also emerges for BrainProp over the course of learning [18]. For the output layer, feedback activity emerges from the selected unit:

$$\phi_n^N = \begin{cases} 1, & \text{if } n = s \\ 0, & \text{if } n \neq s \end{cases} \tag{3}$$

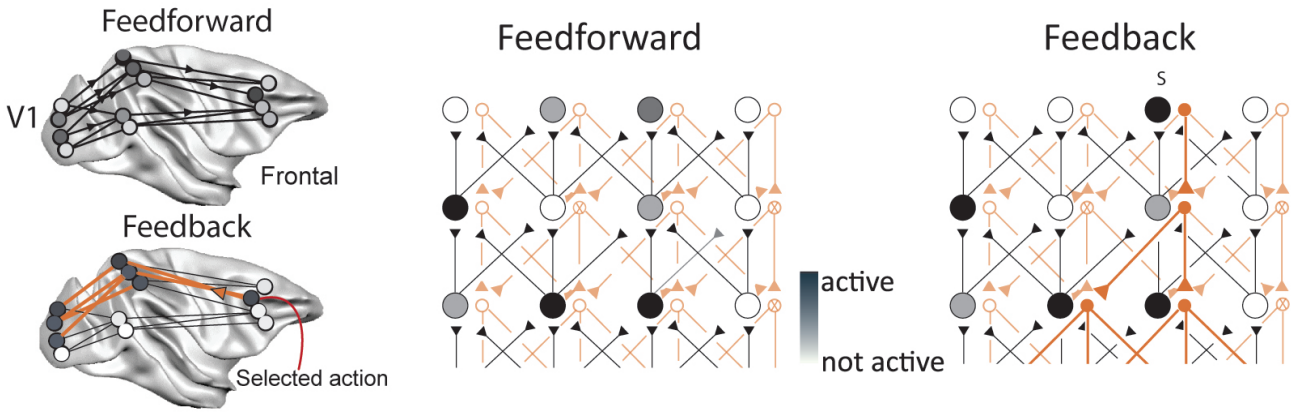

Fig. 1. The BrainProp algorithm is a three-factor learning rule that, based on the presence of feedforward activity and a reward prediction error, assigns credit to synapses involved in action selection using a feedback network. Figure reproduced from [2].

Feedback activity from the layer prior is calculated by means of the feedback weights from the output layer and feedback activity of the output layer:

$$\phi_m^{N-1} = \sum_n w_{n,m}^{FB,N-1} \phi_n^N = w_{s,m}^{FB,N-1}. \tag{4}$$

For all lower layers, feedback activity is computed based on a combination of the feedback weights, feedback activity and the derivative of the activation function (g):

$$\phi_i^l = \sum_k w_{k,i}^{FB,l} \phi_k^{l+1} g_k^{l+1}. \tag{5}$$

The resulting weight update from any neuron i in lower layers to unit j in the layer above becomes a combination of the RPE, feedback from the higher-level unit (ϕ_i^l), the derivative of the activation function in the feedforward pathway (g_i^l) and feedforward activity of the unit itself (y_j^{l-1}):

$$\Delta w_{i,j}^{l-1} = \delta \phi_i^l g_i^l y_j^{l-1}. \tag{6}$$

Since BrainProp uses rectified linear units (ReLUs) as the activation function, the resulting derivative becomes 0 if the higher-level unit was inactive and 1 otherwise. As such, the weight update relies on the presence of both feedback activity and feedforward activity, making it Hebbian in nature.

The main difference between EBP and BrainProp lies in which weights are updated at any given time: while EBP can assign credit to all weights, BrainProp optimises only the subset of connections contributing to the chosen action through the reinforcement learning signal. As a result, the two algorithms are mathematically equivalent in the case that each possible action is sampled by BrainProp once, and the network weights are subsequently altered, for each sample individually. If the action chosen by BrainProp is correct, the network can be updated efficiently. In contrast, if it is incorrect, the network first has to unlearn the class associated with the stimulus before it selects a new class. If this class is also incorrect, this process has to be repeated until the right class is found. Hence, BrainProp enables networks to learn smaller benchmarks to near error-backpropagation accuracy, but learning is less efficient for datasets with large outcome spaces such as ImageNet, because the iterative process of learning and unlearning is slow. Can a curriculum be designed in such a way to remedy this?

B. Curriculum design

To encourage learning of new classes, we developed a curriculum that gradually exposed the model to the different classes in the ImageNet dataset (see Fig. 2). Rather than presenting the full dataset from the beginning, the model was first presented with images from a single class. A new class was added to the dataset once the previous class was learned. This procedure was repeated until all classes were incorporated into the training set. The full network was used from the start, with no alterations made to network size and the number of output units throughout the curriculum. All output units (including those for classes not observed yet) were considered during decision-making and were treated identically during training.

Importantly, the curriculum was designed to incentivise learning of new classes, while protecting the performance on recently introduced classes. Therefore, the newest class was presented with a higher probability (30%) than the classes that were already learned[1], so that the model has sufficient opportunity to sample the outcome categories until it determined the correct class. Moreover, this approach might help the model to learn to not only correctly classify the typical samples of a class [19], but also some of the more exceptional cases, allowing the model to delineate boundaries between the categories. The most recently introduced class prior to the current class was also presented with higher probability (10%) to prevent forgetting of this class, given the instability of recently acquired memories (see also [20]) and the beneficial

[1]From the 10th class onwards. Prior to this, all classes were presented with equal probability during the introduction phase.

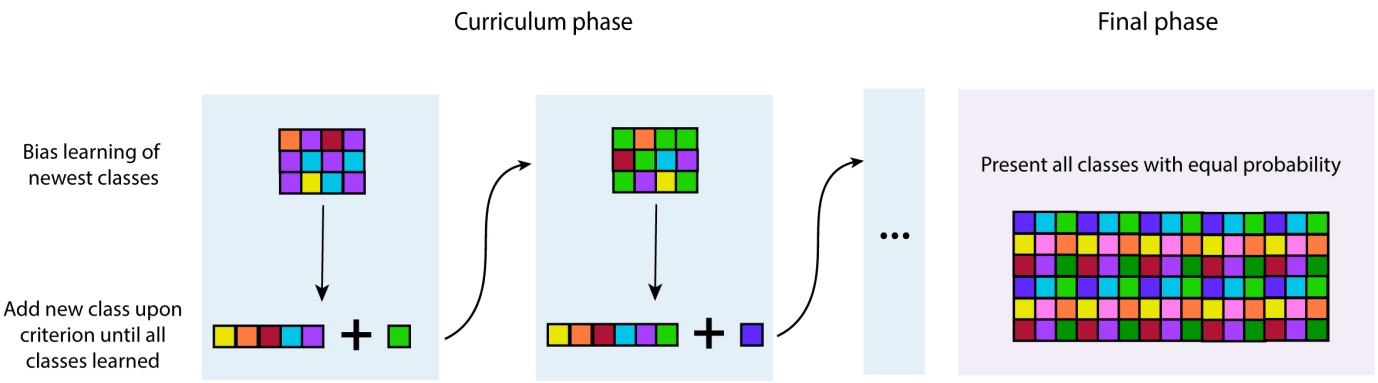

Fig. 2. Training commenced with a curricular pre-training stage, wherein classes are successively added to the training set, while promoting learning of new classes (first purple, then green) while protecting the most recently introduced class prior to that (light blue, followed by purple). Once all ImageNet categories were learned, a final training phase presented the full dataset to the model until convergence.

effects of rehearsal on memory consolidation [21]. A new class was introduced once the model either classified the newest class with high accuracy (at least 75% correct on the most recent 25 presentations) or after a maximum of 15 epochs. This process was repeated until all classes in ImageNet had been presented.

Upon completion of the curriculum phase, the model was trained on the full dataset with the same probability of each class (0.1%). The goal of this final training stage was to enhance overall training and validation accuracy on the dataset and to prevent overfitting, because the curriculum contained only a limited number of samples per class (only a total of 50.000 images were shown during this phase). During this second training stage we presented the full training dataset (1.2 million images) to the network to improve generalisation. An early stopping criterion assessed validation accuracy on 5.000 (of 50.000 validation images) with a patience of 45 and a minimum delta of 0.001. Once validation accuracy plateaued, the model was tested on the full validation dataset (excluding the 5.000 validation images that had been used).

C. Architecture

For all experiments, we used a VGG-inspired network [2] with seven convolutional layers, followed by two fully-connected layers of sizes 8192x3000 and 3000x1000 (see Fig. 3). Epsilon-greedy served as the action selection mechanism for BrainProp. This mechanism usually chose the action associated with the most highly active output unit, but selected an action randomly with a probability of 2%. For EBP we used a softmax function instead. For the curriculum-trained networks, a learning rate of 0.005 was used. BrainProp without curriculum was trained using higher learning rates of 0.04, 0.05 and 0.06. The same learning rates were used for experiments with EBP. A batch size of 125 was utilised and the curriculum evaluated criterion performances every 20th batch. ImageNet stimuli were downsized to 64x64 pixels, converted to RGB and normalised.

III. Experiments

Biologically plausible learning rules have proven difficult to scale to large problems such as ImageNet [22]. We here demonstrate that the BrainProp learning rule can accomplish this under certain conditions, although learning proceeds very slowly and is generally not robust. However, we show that employing a curriculum that gradually introduces each class to the network successfully overcomes these difficulties and substantially improves accuracy, training speed and stability.

Without a curriculum, BrainProp learned to classify ImageNet to some degree, given the right initialisation and with careful hyper-parameter tuning. Networks that do learn, do so very slowly; requiring 1372 (SD = 1295) epochs on average to converge (see Fig. 4). Moreover, top-1 and top-5 test accuracy was low, averaging only 6.4% (SD = 8.1%) and 9.9% (SD = 11.9%), respectively. Furthermore, learning was not robust. Of the five seeds trained on the task, two seeds quickly (i.e. within the first epoch) exhibited unstable performance and did not converge. Two of the seeds that initially demonstrated stable learning also eventually showed a sharp decline in their validation accuracy, which dropped to chance level during a later stage of training (around 1300 and 1500 epochs; see Fig. 5). When the learning rate was decreased from 0.05 to 0.04 ('lowLR' condition), learning was stable, but while none of the seeds showed deterioration in validation accuracy at a later training stage, performance on three of the five seeds never exceeded chance level (average top-1 and top-5 test accuracy for these seeds of 0.11% and 0.49%, respectively, with an overall top-1 and top-5 accuracy of 8.55% and 14.0% for all 5 seeds). With a slightly higher learning rate of 0.06, none of the seeds were viable due to numerical instabilities. Thus, while some networks learned, learning was slow, unstable, and sensitive to hyper-parameter changes.

The curriculum largely solved these issues. Classification performance, training speed and robustness all improved substantially. None of the seeds showed instabilities during training, and the networks reached on average a top-1 test accuracy of 30.2% (SD = 0.5%) and a top-5 test accuracy

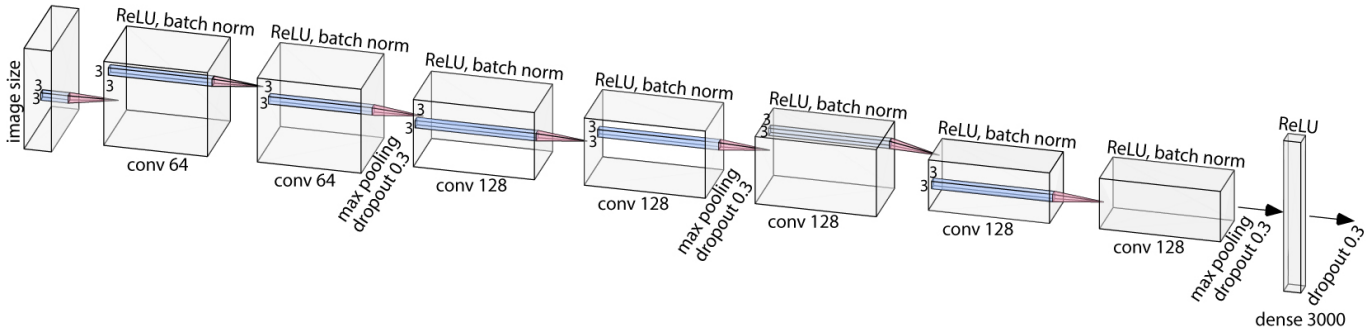

Fig. 3. Network architecture (figure adapted from [2]). Each convolutional layer had a kernel size of 3 by 3 with a stride of 1 and used zero-padding. The first two convolutional layers had 64 channels and all other convolutional layers had 128 channels. Batch normalisation was applied after every convolutional layer. Moreover, max pooling (kernel size and stride of 2) and dropout (0.3) were performed after the second, fourth and seventh layer. A final dropout layer occurred after the first fully-connected layer. ReLU activation functions were employed for all layers except for the output layer.

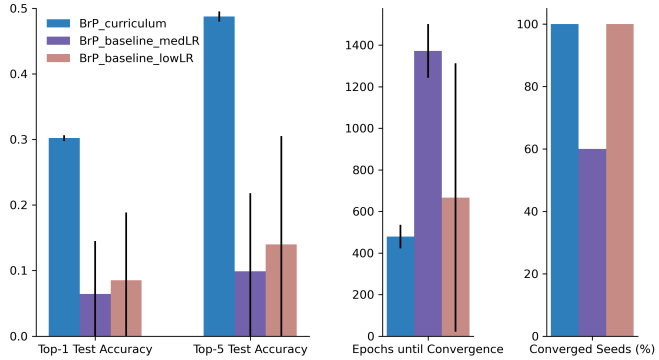

Fig. 4. Average (+ SD) top-1 and top-5 test accuracy, number of epochs until convergence and percentage of converged seeds for networks trained with BrainProp with a curriculum ('BrP_curriculum') and without the curriculum using two types of learning rates (0.05 for 'BrP_baseline_medLR' and 0.04 for 'BrP_baseline_lowLR'). Networks trained with the highest learning rate (0.06) are not shown since their training was aborted very early due to instabilities. Using a curriculum enhanced performance on all metrics.

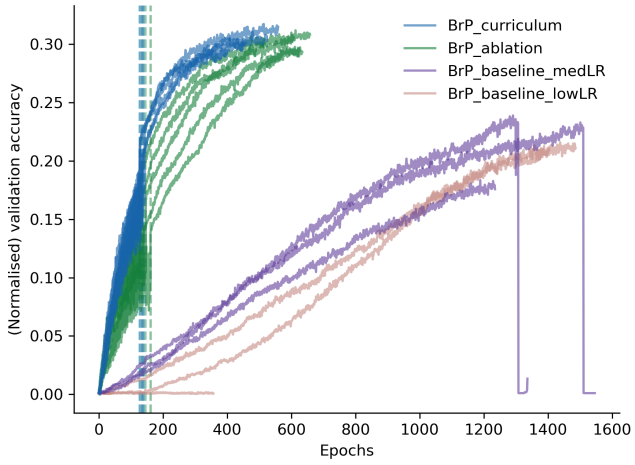

Fig. 5. (Normalised) validation accuracy during training for all conditions. Dashed lines indicate the completion of the curriculum phase for networks trained in the curriculum condition. Networks trained using BrainProp ('BrP') with a curriculum outperform those without the curriculum in both training speed and validation accuracy.

of 48.8% (SD = 0.7%) while requiring only 478 (SD = 56) epochs to converge (Fig. 4-5).

We also trained the network with EBP, with and without the curriculum. As expected, EBP yielded a higher performance than BrainProp in terms of training speed and classification accuracy (Fig. 6). Without the curriculum, EBP obtained an average top-1 test accuracy of 35.9% (SD = 0.2%) and an average top-5 test accuracy of 59.9% (SD = 0.3%), while requiring merely 53.8 (SD = 2.5) epochs on average to converge. Interestingly, adding the curriculum enhanced top-1 and top-5 classification performance further to 43.4% (SD = 0.1%) and 67.8% (SD = 0.2%), respectively, despite a small increase in the number of epochs required for training (M = 83, SD = 7). All seeds showed stable performance and converged for each learning rate (Fig. 6).

Finally, we carried out an ablation study to investigate the role of the recency bias in the curriculum (Fig. 7). This recency bias corresponds to slightly enhancing the probability of the previous class (10%) to prevent forgetting this class. When networks were trained without this recency bias, top-1 and top-5 test performance were still markedly improved compared to the baseline without curriculum and were comparable to that for the full curriculum (M = 29.9%, SD = 0.7%, and M = 48.3%, SD = 1.0%, respectively). However, the full curriculum converged more than 100 epochs faster than the ablated curriculum without the recency bias (M = 609, SD = 48).

In conclusion, using a curriculum allowed BrainProp to successfully learn on ImageNet, thereby proving to be a useful strategy in training networks with biologically plausible reinforcement learning.

IV. Discussion

Thus far, it proved to be too challenging to train networks on large scale classification benchmarks such as the ImageNet

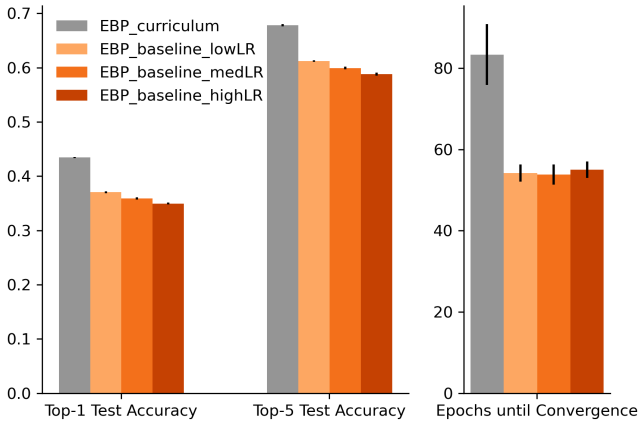

Fig. 6. Average (+ SD) top-1 and top-5 test accuracy, as well as the number of epochs until the networks converged for EBP when trained with or without a curriculum. We used three learning rates ('low_LR'=0.4, 'med_LR'=0.5, 'high_LR'=0.6). Accuracy was higher and training proceeded faster with the curriculum, for each learning rate.

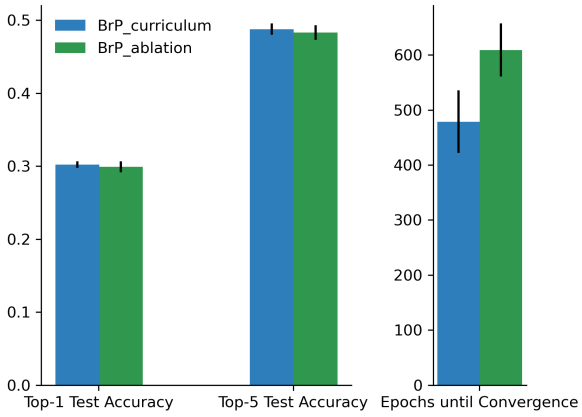

Fig. 7. Average (+ SD) top-1 and top-5 test accuracy, in addition to the number of epochs required until convergence for networks trained with BrainProp using the full curriculum ('BrP_curriculum') and with a version of the curriculum without recency bias ('BrP_ablation'). Although final performance was comparable, the training speed was higher with the full curriculum.

dataset with biologically plausible reinforcement learning rules. Here, we therefore investigated whether a curriculum would help these networks scale to these larger datasets. We first showed that learning with the biologically plausible learning rule BrainProp on ImageNet is in fact possible, but only given careful hyper-parameter tuning. Moreover, learning was overall very slow and accuracy remained limited.

We speculated that the difficulty of a reinforcement learning rule such as BrainProp in scaling to large datasets could originate from having to sample many different classes before the correct class is found. Therefore, to enhance performance, we generated a curriculum that gradually exposed the network to each class in the dataset. We introduced one new class at a time and biased the selection of samples to the new class so that the model had ample opportunity to discover the correct class label by trial-and-error. When the validation accuracy of the newest class reached criterion, we introduced the next class until the full dataset was acquired. We also increased the probability of the class that was introduced most recently to reduce forgetting.

The curriculum enhanced learning on the ImageNet benchmark. Learning accelerated and the convergence time decreased by a factor of nearly three. In addition, learning became more robust and classification performance improved substantially. Specifically, top-1 and top-5 test accuracy rose by a factor of 5 to 30.2% and 48.8%, respectively. To our knowledge, this constitutes state-of-the-art performance for reinforcement learning on ImageNet. Hence, representation learning with biologically plausible learning rules can scale to biologically relevant large-scale domains and powerful network architectures.

The curriculum also improved classification performance of networks trained using EBP, although it necessitated a modest number of additional training epochs to converge. This finding is counter-intuitive given that EBP is a supervised learning method and it should therefore not suffer from the class-learning problem that BrainProp is sensitive to. One possible explanation is related to the lottery ticket hypothesis [23], which states that learning becomes more robust when networks have many parameters so that the efficacy of curricula decreases. Here we used relatively small networks, and the curriculum may have overcome such unfavourable initialisations. This explanation is supported by the BrainProp experiments without a curriculum, demonstrating instabilities and the absence of learning with lower learning rates. However, [23] studied smaller datasets (up to CIFAR-10) and networks in supervised settings. Future work could investigate the effect of curricula on deep neural networks.

The recency bias did not materially affect final model performance, but improved the speed of convergence. However, our implementation of the recency bias was relatively simple and it is possible that using a more elaborate approach (e.g. as in [24]) would also increase the classification accuracy. For instance, one could explore the protection of multiple classes that were recently introduced, decaying their probability based on how long ago they were learned. Other elements could also be added to the curriculum. For instance, a type of consolidation mechanism or other regularisation or rehearsal-based techniques from continual learning and class-incremental learning [25] may protect older classes against forgetting. Alternatively, the overarching class structure in ImageNet (based on the WordNet hierarchy as in [26] or a compositional analysis of network motifs from a trained network such as in [27]) could be leveraged to identify – and protect – classes with representational overlap with new classes and might therefore be overwritten. Finally, several studies used difficulty metrics in designing the order of curricula (as reviewed by [7]). We leave the examination of these other approaches as opportunities for future work.

It is of interest to compare category learning between artificial systems and biological systems. Humans and non-

human primates have specific biases during category learning [19], [28], which are qualitatively different from those in other vertebrates [28]. For example, primates tend to first learn the stereotypical samples of a class, and only later the more exceptional samples [19], [29]. These biases have also been observed in both deep neural networks trained with supervised algorithms [29], [30] and could be exploited by curricula to further enhance learning, by first training on the canonical samples and introducing the non-typical samples later. Additionally, the ordering of examples within classes can be structured in a way to optimise category discovery and generalisation further [31], [32].

BrainProp is one of several biologically inspired learning rules. Other approaches include predictive coding-based rules [33], [34], equilibrium propagation [35], target propagation [36], [37], feedback alignment [38]–[40], e-prop [41], sign-symmetry [42], the forward-forward algorithm [43], and several other unsupervised [44], [45] or self-supervised [44], [46]–[53] approaches. Most of these learning rules however do not scale to deeper networks or larger datasets such as ImageNet [22]. The exceptions that do scale to deeper architectures and/or larger data sets such as ImageNet [36], [39], [40], [42], [44], [49]–[52], [54], [55], use some type of supervisory signal telling the network what the correct response should be combined with error-backpropagation, for instance in the output layer. This limits their biological plausibility. Alternatively, they use a more biologically plausible approximation of error-backpropagation but then do not scale to deep networks (e.g. [48]). While there are indications that devising scalable self-supervised approaches that are compatible with biological constraints are non-trivial [56], [57], these alternative approaches offer interesting prospects when combined with the BrainProp framework, which might enhance their overall biological plausibility and augment the overall performance. One interesting avenue for future research could also be to combine a self-supervised pre-training phase to allow the network to learn the statistics of the dataset prior to learning class identities through reinforcement learning curriculum with BrainProp [58]. Another possibility would be to add an active learning [59] phase subsequent to learning all outcome categories to allow the model to continue learning based on predictions about unlabelled samples that it is sufficiently confident about.

In conclusion, we showed that a curriculum enables BrainProp, a biologically plausible reinforcement learning rule, to train deep networks on the challenging ImageNet benchmark and to achieve state-of-the-art performance across existing biologically plausible learning rules. The curriculum was relatively simple, and we noted many opportunities for expansion and combining it with other approaches. This approach specifically demonstrates how deep networks can be trained with local learning rules and limited feedback, which is of importance for example when training EdgeAI devices in the field. We hope that our work inspires future studies that could achieve even better performance, and generate new insights into the relation between category learning in artificial and biological systems.

ACKNOWLEDGMENT

We thank I. Pozzi for providing base code for the BrainProp algorithm.

REFERENCES

[1] R. S. Sutton and A. G. Barto, *Reinforcement Learning, second edition: An Introduction*. MIT Press, Nov. 2018.

[2] I. Pozzi, S. M. Bohté, and P. R. Roelfsema, "Attention-gated brain propagation: how the brain can implement reward-based error backpropagation," in *Proceedings of the 34th International Conference on Neural Information Processing Systems*, ser. NIPS'20. Red Hook, NY, USA: Curran Associates Inc., Dec. 2020, pp. 2516–2526.

[3] M. F. He, B. D. Schultz, and W. H. Schubert, *The SAGE Guide to Curriculum in Education*. SAGE Publications, Jun. 2015.

[4] K. Pryor and K. Ramirez, "Modern Animal Training," in *The Wiley Blackwell Handbook of Operant and Classical Conditioning*. John Wiley & Sons, Ltd, 2014, pp. 453–482.

[5] P. McGreevy and R. Boakes, *Carrots and Sticks: Principles of Animal Training*. Darlington Press, 2011.

[6] S. Narvekar, B. Peng, M. Leonetti, J. Sinapov, M. E. Taylor, and P. Stone, "Curriculum Learning for Reinforcement Learning Domains: A Framework and Survey," *Journal of Machine Learning Research*, vol. 21, no. 181, pp. 1–50, 2020.

[7] P. Soviany, R. T. Ionescu, P. Rota, and N. Sebe, "Curriculum Learning: A Survey," *International Journal of Computer Vision*, vol. 130, no. 6, pp. 1526–1565, Jun. 2022.

[8] D. Weinshall, G. Cohen, and D. Amir, "Curriculum Learning by Transfer Learning: Theory and Experiments with Deep Networks," 2018, pp. 5238–5246.

[9] D. Weinshall and D. Amir, "Theory of Curriculum Learning, with Convex Loss Functions," *Journal of Machine Learning Research*, vol. 21, no. 222, pp. 1–19, 2020.

[10] W. Zaremba and I. Sutskever, "Learning to Execute," Feb. 2015.

[11] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, "Curriculum learning," in *Proceedings of the 26th Annual International Conference on Machine Learning*. Montreal Quebec Canada: ACM, Jun. 2009, pp. 41–48.

[12] R. C. Wilson, A. Shenhav, M. Straccia, and J. D. Cohen, "The Eighty Five Percent Rule for optimal learning," *Nature Communications*, vol. 10, no. 1, p. 4646, Nov. 2019.

[13] J. H. Lee, S. S. Mannelli, and A. Saxe, "Why Do Animals Need Shaping? A Theory of Task Composition and Curriculum Learning," Feb. 2024.

[14] D. Hocker, C. M. Constantinople, and C. Savin, "Curriculum learning inspired by behavioral shaping trains neural networks to adopt animal-like decision making strategies," Jan. 2024.

[15] R. B. Dekker, F. Otto, and C. Summerfield, "Curriculum learning for human compositional generalization," *Proceedings of the National Academy of Sciences*, vol. 119, no. 41, p. e2205582119, Oct. 2022.

[16] H. E. Manzur, K. Vlasov, Y.-J. Jhong, H.-Y. Chen, and S.-C. Lin, "The behavioral signature of stepwise learning strategy in male rats and its neural correlate in the basal forebrain," *Nature Communications*, vol. 14, no. 1, p. 4415, Jul. 2023.

[17] W. Schultz, "Dopamine reward prediction-error signalling: a two-component response," *Nature Reviews Neuroscience*, vol. 17, no. 3, pp. 183–195, Mar. 2016.

[18] P. R. Roelfsema and A. van Ooyen, "Attention-gated reinforcement learning of internal representations for classification," *Neural Computation*, vol. 17, no. 10, pp. 2176–2214, 2005.

[19] J. D. Smith, W. P. Chapman, and J. S. Redford, "Stages of Category Learning in Monkeys (Macaca mulatta) and Humans (Homo sapiens)," *Journal of experimental psychology. Animal behavior processes*, vol. 36, no. 1, pp. 39–53, Jan. 2010.

[20] N. Mosha and E. Robertson, "Unstable Memories Create a High-Level Representation that Enables Learning Transfer," *Current Biology*, vol. 26, no. 1, pp. 100–105, Jan. 2016.

[21] L. Himmer, M. Schönauer, D. P. J. Heib, M. Schabus, and S. Gais, "Rehearsal initiates systems memory consolidation, sleep makes it last," *Science Advances*, vol. 5, no. 4, p. eaav1695, Apr. 2019.

[22] S. Bartunov, A. Santoro, B. Richards, L. Marris, G. E. Hinton, and T. Lillicrap, "Assessing the Scalability of Biologically-Motivated Deep Learning Algorithms and Architectures," in *Advances in Neural Information Processing Systems*, vol. 31. Curran Associates, Inc., 2018.

[23] S. S. Mannelli, Y. Ivashinka, A. Saxe, and L. Saglietti, "Tilting the Odds at the Lottery: the Interplay of Overparameterisation and Curricula in Neural Networks," Jun. 2024.

[24] S. Chen, M. Zhang, J. Zhang, and K. Huang, "Exemplar-based Continual Learning via Contrastive Learning," *IEEE Transactions on Artificial Intelligence*, vol. 5, pp. 3313–3324, 2024.

[25] M. Masana, X. Liu, B. Twardowski, M. Menta, A. D. Bagdanov, and J. van de Weijer, "Class-incremental learning: survey and performance evaluation on image classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 5, pp. 5513–5533, 2022.

[26] S. Wen, A. S. Rios, K. Lekkala, and L. Itti, "What can we learn from misclassified ImageNet images?" Jan. 2022.

[27] L. N. Driscoll, K. Shenoy, and D. Sussillo, "Flexible multitask computation in recurrent networks utilizes shared dynamical motifs," *Nature Neuroscience*, vol. 27, no. 7, pp. 1349–1363, Jul. 2024.

[28] J. D. Smith, M. J. Crossley, J. Boomer, B. A. Church, M. J. Beran, and F. G. Ashby, "Implicit and Explicit Category Learning by Capuchin Monkeys (Cebus apella)," *Journal of comparative psychology*, vol. 126, no. 3, pp. 294–304, Aug. 2012.

[29] J. Rubruck, J. P. Bauer, A. Saxe, and C. Summerfield, "Early learning of the optimal constant solution in neural networks and humans," Jun. 2024.

[30] K. Kang, A. Setlur, C. Tomlin, and S. Levine, "Deep Neural Networks Tend To Extrapolate Predictably," Mar. 2024.

[31] F. Mathy and J. Feldman, "A rule-based presentation order facilitates category learning," *Psychonomic Bulletin & Review*, vol. 16, no. 6, pp. 1050–1057, Dec. 2009.

[32] ——, "The Influence of Presentation Order on Category Transfer," *Experimental Psychology*, vol. 63, no. 1, pp. 59–69, Jan. 2016.

[33] Y. Song, B. Millidge, T. Salvatori, T. Lukasiewicz, Z. Xu, and R. Bogacz, "Inferring neural activity before plasticity as a foundation for learning beyond backpropagation," *Nature Neuroscience*, vol. 27, no. 2, pp. 348–358, Feb. 2024.

[34] M. W. Spratling, "A review of predictive coding algorithms," *Brain and Cognition*, vol. 112, pp. 92–97, Mar. 2017.

[35] A. Laborieux and F. Zenke, "Holomorphic Equilibrium Propagation Computes Exact Gradients Through Finite Size Oscillations," *Advances in neural information processing systems*, vol. 35, pp. 12 950–12 963, 2022.

[36] M. M. Ernoult, F. Normandin, A. Moudgil, S. Spinney, E. Belilovsky, I. Rish, B. Richards, and Y. Bengio, "Towards Scaling Difference Target Propagation by Learning Backprop Targets," in *Proceedings of the 39th International Conference on Machine Learning*, vol. 162. PMLR, Jun. 2022, pp. 5968–5987.

[37] A. Meulemans, M. T. Farinha, J. G. Ordóñez, P. V. Aceituno, J. Sacramento, and B. F. Grewe, "Credit Assignment in Neural Networks through Deep Feedback Control," *Advances in Neural Information Processing Systems*, vol. 34, pp. 4674–4687, 2021.

[38] B. Crafton, A. Parihar, E. Gebhardt, and A. Raychowdhury, "Direct Feedback Alignment With Sparse Connections for Local Learning," *Frontiers in Neuroscience*, vol. 13, p. 525, May 2019.

[39] L. Ji-An and M. K. Benna, "Deep Learning without Weight Symmetry," May 2024.

[40] J. Launay, I. Poli, F. Boniface, and F. Krzakala, "Direct Feedback Alignment Scales to Modern Deep Learning Tasks and Architectures," in *Advances in Neural Information Processing Systems*, vol. 33. Curran Associates, Inc., 2020, pp. 9346–9360.

[41] G. Bellec, F. Scherr, A. Subramoney, E. Hajek, D. Salaj, R. Legenstein, and W. Maass, "A solution to the learning dilemma for recurrent networks of spiking neurons," *Nature Communications*, vol. 11, no. 1, p. 3625, 2020.

[42] W. Xiao, H. Chen, Q. Liao, and T. Poggio, "Biologically-plausible learning algorithms can scale to large datasets," *arXiv:1811.03567*, Dec. 2018.

[43] G. Hinton, "The Forward-Forward Algorithm: Some Preliminary Investigations," Dec. 2022.

[44] G. Shen, D. Zhao, Y. Dong, and Y. Zeng, "Brain-inspired neural circuit evolution for spiking neural networks," *Proceedings of the National Academy of Sciences*, vol. 120, no. 39, p. e2218173120, Sep. 2023.

[45] J. Talloen, J. Dambre, and A. Vandesompele, "PyTorch-Hebbian: facilitating local learning in a deep learning framework," Jan. 2021.

[46] M. S. Halvagal and F. Zenke, "The combination of Hebbian and predictive plasticity learns invariant object representations in deep sensory networks," *Nature Neuroscience*, vol. 26, no. 11, pp. 1906–1915, Nov. 2023.

[47] B. Illing, J. Ventura, G. Bellec, and W. Gerstner, "Local plasticity rules can learn deep representations using self-supervised contrastive predictions," *Advances in Neural Information Processing Systems*, vol. 34, pp. 30 365–30 379, 2021.

[48] A. Journé, H. G. Rodriguez, Q. Guo, and T. Moraitis, "Hebbian Deep Learning Without Feedback," Aug. 2023.

[49] M. Oquab, T. Darcet, T. Moutakanni, H. Vo, M. Szafraniec, V. Khalidov, P. Fernandez, D. Haziza, F. Massa, A. El-Nouby, M. Assran, N. Ballas, W. Galuba, R. Howes, P.-Y. Huang, S.-W. Li, I. Misra, M. Rabbat, V. Sharma, G. Synnaeve, H. Xu, H. Jegou, J. Mairal, P. Labatut, A. Joulin, and P. Bojanowski, "DINOv2: Learning Robust Visual Features without Supervision," Feb. 2024.

[50] M. Ren, S. Kornblith, R. Liao, and G. Hinton, "Scaling Forward Gradient With Local Losses," Mar. 2023.

[51] S. A. Siddiqui, D. Krueger, Y. LeCun, and S. Deny, "Blockwise Self-Supervised Learning at Scale," Feb. 2023.

[52] V. Sobal, M. Ibrahim, R. Balestriero, V. Cabannes, D. Bouchacourt, P. Astolfi, K. Cho, and Y. LeCun, "$\mathbb{X}$-Sample Contrastive Loss: Improving Contrastive Learning with Sample Similarity Graphs," Jul. 2024.

[53] M. Tang, Y. Yang, and Y. Amit, "Biologically Plausible Training Mechanisms for Self-Supervised Learning in Deep Networks," *Frontiers in Computational Neuroscience*, vol. 16, p. 789253, Mar. 2022.

[54] H. Aghabarar, P. Keshavarzi, and K. Kiani, "Reinforcement Learning in Deep Spiking Neural Networks with Eligibility Traces and Modifying the Threshold Parameter," Jan. 2024.

[55] H. Ghaemi, E. Mirzaei, M. Nouri, and S. R. Kheradpisheh, "BioLCNet: Reward-Modulated Locally Connected Spiking Neural Networks," in *International Conference on Machine Learning, Optimization, and Data Science*, G. Nicosia, V. Ojha, E. La Malfa, G. La Malfa, P. Pardalos, G. Di Fatta, G. Giuffrida, and R. Umeton, Eds. Cham: Springer Nature Switzerland, 2023, pp. 564–578.

[56] R. Weiler, M. Brucklacher, C. M. A. Pennartz, and S. M. Bohté, "Masked Image Modeling as a Framework for Self-Supervised Learning across Eye Movements," 2024, pp. 17–31.

[57] F. Zenke, "Private communication."

[58] R. Cusack, M. Ranzato, and C. J. Charvet, "Helpless infants are learning a foundation model," *Trends in Cognitive Sciences*, vol. 28, no. 8, pp. 726–738, Jun. 2024.

[59] P. Ren, Y. Xiao, X. Chang, P.-Y. Huang, Z. Li, B. B. Gupta, X. Chen, and X. Wang, "A Survey of Deep Active Learning," *ACM Computing Surveys*, vol. 54, no. 9, pp. 1–40, Dec. 2022.