BIOLOGICALLY-PLAUSIBLE LEARNING ALGORITHMS CAN SCALE TO LARGE DATASETS

Anonymous authors
Paper under double-blind review

ABSTRACT

The backpropagation (BP) algorithm is often thought to be biologically implausible in the brain. One of the main reasons is that BP requires symmetric weight matrices in the feedforward and feedback pathways. To address this "weight transport problem" (Grossberg, 1987), two biologically plausible algorithms, proposed by Liao et al. (2016) and Lillicrap et al. (2016), relax BP's weight symmetry requirements and demonstrate comparable learning capabilities to that of BP on small datasets. However, a recent study by Bartunov et al. (2018) finds that although feedback alignment (FA) and some variants of target-propagation (TP) perform well on MNIST and CIFAR, they perform significantly worse than BP on ImageNet. Here, we additionally evaluate the sign-symmetry (SS) algorithm (Liao et al., 2016), which differs from both BP and FA in that the feedback and feedforward weights do not share magnitudes but share signs. We examine the performance of sign-symmetry and feedback alignment on ImageNet and MS COCO datasets using different network architectures (ResNet-18 and AlexNet for ImageNet; RetinaNet for MS COCO). Surprisingly, networks trained with sign-symmetry can attain classification performance approaching that of BP-trained networks. These results complement the study by Bartunov et al. (2018) and establish a new benchmark for future biologically plausible learning algorithms on more difficult datasets and more complex architectures.

1 INTRODUCTION

Many deep learning models today are highly successful in task performance, learning useful representations, and even matching representations in the brain (for example, see DiCarlo, 2017). However, it remains a contentious issue whether these models reflect how learning takes place in the brain. Core to the problem is the fact that backpropagation, the learning algorithm underlying most of today's deep networks, is difficult to implement in the brain given what we know about the brain's hardware (Crick 1989; however, see Hinton 2007).

One main reason why backpropagation seems implausible in the brain is that it requires sharing of feedforward and feedback weights. Since synapses are unidirectional in the brain, feedforward and feedback connections are physically distinct. Requiring them to shared their weights, even as weights are adjusted during learning, seems highly implausible.

One approach to addressing this issue is to relax the requirement for weight-symmetry in error backpropagation. Surprisingly, when the feedback weights share only the sign but not the magnitude of the feedforward weights (Liao et al., 2016) or even when the feedback weights are random (but fixed) (Lillicrap et al., 2016), they can still guide useful learning in the network, with performance comparable to and sometimes even better than performance of backpropagation, on datasets such as MNIST and CIFAR. Here, we refer to these two algorithms, respectively, as "sign-symmetry" and "feedback alignment." Since weight symmetry is required in backpropagation for accurately propagating the derivative of the loss function through layers, success of asymmetric feedback algorithms indicate that learning can be supported by something other than accurate estimation of the error derivative. In feedback alignment, the authors suggest that the feedforward weights learn to align with the random feedback weights, thereby allowing feedback to provide approximate yet useful learning signals (Lillicrap et al., 2016).

However, a recent paper by Bartunov et al. (2018) finds that feedback alignment and a few other biologically-plausible algorithms, including variants of target propagation, do not generalize to larger and more difficult problems such as ImageNet (Deng et al., 2009) and there perform much worse than backpropagation. Nevertheless, the specific conditions Bartunov et al. tested are somewhat restrictive. First, they only tested locally-connected networks (i.e., weight sharing is not allowed among convolution filters at different spatial locations), a choice that is motivated by biological plausibility but in practice limits the size of the network (without weight sharing, each convolutional layer needs much more memory to store its weights).[1] Second, Bartunov et al. did not test sign-symmetry, which may be more powerful than feedback alignment since sign-symmetric feedback weights may carry more information about the feedforward weights than the random feedback weights used in feedback alignment.

Therefore, in this work, we re-examine the performance of sign-symmetry and feedback alignment on ImageNet and MS COCO datasets using standard ConvNet architectures (i.e., ResNet-18, AlexNet, and RetinaNet). Our conclusions largely disagree with results from Bartunov et al. (2018): We find that sign-symmetry can in fact train networks on ImageNet and performs almost as well as backpropagation. In addition, we test the use of backpropagation exclusively in the last layer while otherwise using feedback alignment, hypothesizing that in the brain, the classifier layer may not be a fully-connected layer and may deliver the error signal through some other unspecified mechanism. Such partial feedback alignment can still train a ConvNet to achieve a better performance (relative to backpropagation) than achieved in Bartunov et al. (2018). Taken together, these results indicate that biologically-plausible learning algorithms remain viable options both for training artificial neural networks and for modeling how learning can occur in the brain.

2 METHODS

Consider a layer in a feed-forward neural network. Let x_i denote the input to the ith neuron in the layer and y_j the output of the jth neuron. Let W denote the feed-forward weight matrix and W_{ij} denote the connection between input x_i and output y_j. Let f denote the activation function. Then, Equation. 1 describes the computation in the feedforward step. Now, let B denote the feedback weight matrix and B_{ij} denote the feedback connection between output y_j and input x_i, and let f' denote the derivative of the activation function f. Given the objective function E, the error gradient $\frac{\partial E}{\partial x_i}$ calculated in the feedback step is described by Equation. 2.

$$y_j = f(\sigma_j), \quad \sigma_j = \sum_i W_{ij} x_i \tag{1}$$

$$\frac{\partial E}{\partial x_i} = \sum_i B_{ij} f'(\sigma_j) \frac{\partial E}{\partial y_i} \tag{2}$$

Standard backpropagation requires $B = W$. Sign-symmetry (Liao et al., 2016) relaxes the above symmetry requirement in standard backpropagation by letting $B = \mathrm{sign}(W)$, where $\mathrm{sign}(\cdot)$ is the function that takes the sign (-1 or 1) of each entry in the weight matrix. Feedback alignment (Lillicrap et al., 2016) uses fixed random matrices as the feedback weight matrix B. Lillicrap et al. showed that through training, W is adjusted such that on average, $\mathbf{e}^T W B \mathbf{e} > 0$, where $\mathbf{e}$ is the error in the network's output. This condition implies that the error correction signal $B\mathbf{e}$ lies within $90°$ of $\mathbf{e}^T W$, the error calculated by standard backpropagation.

We implement both algorithms in convolutional and fully-connected layers in PyTorch and will share the code openly on GitHub (URL not included for review for anonymity).

[1] However, theoretical and experimental results of Poggio et al. (2017) suggest that weight-sharing is not the main reason for the good performance of ConvNets, at least when trained with backpropagation.

3 RESULTS

3.1 SIGN-SYMMETRY PERFORMS WELL ON IMAGENET

TRAINING DETAILS

We trained ResNet-18 (He et al., 2016) on ImageNet using 5 different learning algorithm settings: 1) backpropagation; 2) sign-symmetry for convolutional layers and backpropagation for the last, fully-connected layer; 3) sign-symmetry for all (convolutional and fully-connected) layers; 4) feedback alignment for convolutional layers and backpropagation for the fully-connected layer; and 5) feedback alignment for all (convolutional and fully-connected) layers. In sign-symmetry, at each backward step, feedback weights are taken as the signs of the feedforward weights, scaled by the same scale λ used to initialize that layer.[2] In feedback alignment, feedback weights are initialized once at the beginning as random variables from the same distribution and scale as used to initialize that layer. For training with backpropagation, standard training parameters are used (SGD with learning rate 0.1, momentum 0.9, and weight decay 10^{-4}). For ResNet-18 with other learning algorithms, we used SGD with learning rate 0.05, while momentum and weight decay remain unchanged.[3] For AlexNet with all learning algorithms, the standard training parameters are used (SGD with learning rate 0.01, momentum 0.9, and weight decay 5×10^{-4}). We used a version of AlexNet (Krizhevsky, 2014, as used in `torchvision`) and slightly modified it to have batch normalization layers after every convolutional and fully-connected layer. For all experiments, we used a batch size of 256, a learning rate decay of 10-fold every 10 epochs, and trained for 50 epochs. While not all conditions have finished training by the time of manuscript submission, additional data points are unlikely to affect the conclusions below.

RESULTS

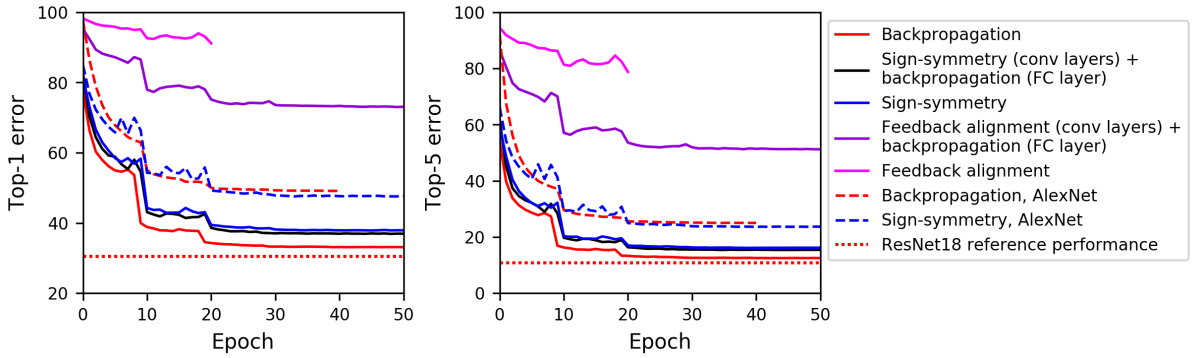


Figure 1: Top-1 and top-5 validation error on ImageNet of ResNet-18 and AlexNet trained with different learning algorithms. Sign-symmetry performs nearly as well as backpropagation, while feedback alignment performs better than previously reported whe backpropagation is used to train the last layer.

In all cases, the network is able to learn (Figure 1, Table 1). Remarkably, sign-symmetry only slightly underperforms backpropagation in this benchmark large dataset, despite the fact that sign-symmetry does not accurately calculate the error gradient. An intuitive explanation for this performance is that the skip-connections in ResNet help prevent the presumed degradation of the gradient being passed through many layers of sign-symmetric feedback. However, sign-symmetry also performs similarly well to backpropagation in a (modified) AlexNet architecture, which does not contain skip connections. Therefore, skip-connections alone do not explain the performance of sign-symmetry.

[2] For conv layers, $\lambda = \sqrt{2/(n_{\text{kernel width}} \cdot n_{\text{kernel height}} \cdot n_{\text{output channels}})}$; for fully-connected layers, $\lambda = 1/\sqrt{n_{\text{output}}}$

[3] The learning rate 0.05 is selected as the best from (0.1, 0.05, 0.01, 0.001, 0.0001)

Table 1: Final ImageNet 1-crop validation accuracy of networks trained with different algorithms. BP: backpropagation; FA: feedback alignment; SS: sign-symmetry; 50 epochs unless otherwise noted.

Architecture & Algorithm	Top-1 Val Error, %	Top-5 Val Error, %
ResNet-18, BP	33.14	12.53
ResNet-18, SS + last layer BP	36.95	15.44
ResNet-18, SS	37.91	16.18
ResNet-18, FA + last layer BP	73.01	51.24
ResNet-18, FA (21 epochs)	91.17	53.61
AlexNet, BP (41 epochs)	49.14	25.02
AlexNet, SS	47.57	23.68

In addition, although its performance is considerably worse, feedback alignment is still able to guide better learning in the network than reported by Bartunov et al. (2018, their Figure 3) if we use backpropagation in the last layer. This condition is not unreasonable since in the brain, the classifier layer is likely not a soft-max classifier and may train and deliver error signals by a completely different mechanism. We also tested using backpropagation exclusively for the last layer in a network otherwise trained with sign-symmetry, but the effect on the performance was minimal. One possibility why sign-symmetry performs better than feedback alignment is that in sign-symmetry, the feedback weight always tracks the sign of the feedforward weight, which may reduce the burden on the feedforward weight to learn to align with the feedback weight.

Finally, in Liao et al. (2016), Batch-Manhattan (BM) SGD was proposed as a way to stabilize training with asymmetric feedback algorithms. In our experience, standard SGD consistently works better than BM for sign-symmetry, but BM may improve results for feedback alignment. We have not comprehensively characterized the effects of BM since many factors like learning rate can affect the outcome. Future experiments are needed to draw stronger conclusions.

3.2 Microsoft COCO Dataset

Besides the ImageNet classification task, we examined the performance of sign-symmetry on the MS COCO object detection task. Object detection is more complex than classification and might therefore require more complicated network architecture in order to achieve high accuracy. Thus, in this experiment we assess the effectiveness of sign-symmetry in training networks that are more complicated and difficult to optimize.

Training details

We train the state-of-the-art object detection network RetinaNet proposed by Lin et al. (2017) on the COCO `trainval35k` split, which consists of 80k images from `train` and a random 35k subset of images from the 40k image `val` split. RetinaNet is comprised of a ResNet-FPN backbone, a classification subnet, and a bounding box regressing subnet. The network is trained with three different training settings: 1) backpropagation for all layers; 2) backpropagation for the last layer in regression and classification subnets, and sign-symmetry for rest of the layers; 3) backpropagation for the last layer in regression and classification subnets, and feedback alignment for rest of the layers. We use a backbone ResNet-18 pretrained on ImageNet to initialize the network. In all the experiments, the network is trained with SGD with an initial learning rate of 0.01, momentum of 0.9, and weight decay of 0.001. We train the network for 40k iterations with 8 images in each minibatch. The learning rate is divided by 10 at 20k iterations.

Results

The results on COCO are similar to those obtained on ImageNet, although the performance gap between SS and BP on COCO dataset is slightly more prominent (Figure 2). A number of factors could potentially contribute to this result. We follow the Feature Pyramid Network (FPN) architecture design choices, optimizers, and hyperparameters reported by Lin et al. (2017). One reason why SS

performs worse than backpropagation in these standard training settings could be that the choices of FPN architectures, optimizers, and hyperparameters are all optimized for use with backpropagation. Hence, our results are still preliminary and it remains possible that sign-symmetry can work better under more optimal settings.

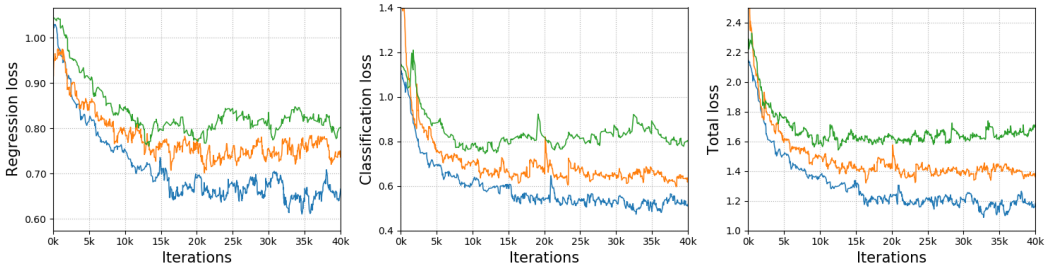


Figure 2: Training loss of RetinaNet on COCO dataset for three different settings: 1) standard backpropagation for all the layers (blue); 2) backpropagation for the last layer in regression subnet and classification subnet, and sign-symmetry for rest of the layers (orange); 3) backpropagation for the last layer in regression subnet and classification subnet, and feedback alignment for rest of the layers (green). **Left**: object detection bounding box regression loss. **Center**: focal loss for classification. **Right**: summation of regression loss and classification loss

4 Discussion

4.1 Comparing learning in SS, FA, and BP

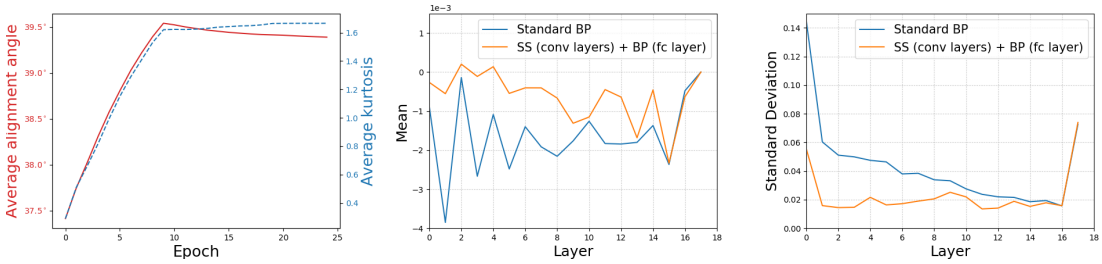
We run a number of experiments to analyze how sign-symmetry guides learning. Lillicrap et al. (2016) show that with feedback alignment, the alignment angles between feedforward and feedback weights gradually decrease because the feedforward weights evolve to align with the feedback weights. We asked whether the same happens in sign-symmetry by computing average alignment angles similar to Lillicrap et al. (2016): For every pair of feedforward and feedback weight matrices, we flatten the matrices into vectors and compute the angle between the vectors. The average of alignment angles between all pairs of weights in the network is computed after every training epoch. Surprisingly, our analysis reveals an opposite trend to what we expect: the average alignment angles increase with training (Figure 3(a), red). One possible explanation for this increasing trend is that as the training progresses, the feedforward weights tend to become sparse. Geometrically, this means that feedforward vectors become more aligned to the standard basis vectors and less aligned with the feedback weight vectors, which always lie on a diagonal by construction. This explanation is consistent with the similarly increasing trend of the average kurtosis of the feedforward weights (Figure 3(a) blue), which indicates that values of the weights become more dispersed during training.

Since the magnitudes of the feedforward weights are discarded during training, we are also interested in how sign-symmetry affects the size of the feedforward weights. Therefore, we compare the mean and standard deviation of the weights of networks trained with different learning algorithms. This analysis reveals that sign-symmetry results in weights with higher mean (Figure 3(b)) but lower standard deviation as compared to weights trained by backpropagation (Figure 3(c)). More work is needed to elucidate how sign-symmetry guides efficient learning in the network.

4.2 Why do our results differ from previous work?

Our results indicate that biologically-plausible learning algorithms, specifically sign-symmetry and feedback alignment, are able to learn on ImageNet. This finding directly conflicts with the findings by Bartunov et al. (2018). Why do we come to such different conclusions?

First, Bartunov et al. did not test sign-symmetry, which is expected to be more powerful than feedback alignment, because it is a special case of feedback alignment that allows feedback weights to have additional information about feedforward weights. Indeed, the performance of sign-symmetry



(a) (b) (c)

Figure 3: (a) Blue: The average alignment angle between feedforward and feedback weight matrices increases as training progresses. Red: the kurtosis of the feedforward weight matrices reveals the same trend as the alignment angle. The kurtosis of a dataset is the fourth central moment divided by the square of the variance; it measures the sharpness of peakedness in a distribution. (b) and (c): The mean of feedforward weights trained by sign symmetry is higher than that trained by standard backpropagation, but the standard deviation is generally lower.

approached that of backpropagation, and exceeded the performance of feedback alignment by a wide margin. Another reason may be that instead of using standard ConvNets on ImageNet, Bartunov et al. only tested locally-connected networks. While the later is a more biologically plausible architecture, in practice, it is limited in size by the need to store separate weights for each spatial location. This reduced model capacity create a bottleneck that affects the performance of feedback alignment (see Lillicrap et al., 2016, Supplementary Note 9). Finally, the performance of feedback alignment also benefits from the use of backpropagation in the last layer in our conditions.

4.3 Towards a more biologically plausible learning algorithm

A major reason why backpropagation is considered implausible in the brain is that it requires exact symmetry of physically distinct feedforward and feedback pathways. Sign-symmetry and feedback alignment address this problem by relaxing this tight coupling of weights between the two pathways. Feedback alignment requires no relation at all between feedforward and feedback weights and simply depends on learning to align the two. Hence, it can be easily realized in the brain (for example, see Lillicrap et al., 2016, Supplementary Figure 3). However, empirically, we and others have found its performance to be not ideal on relatively challenging problems.

Sign-symmetry, on the other hand, introduces a mild constraint that feedforward and feedback connections be "antiparallel": they need to have opposite directions but consistent signs. This can be achieved in the brain with two additional yet plausible conditions: First, the feedforward and feedback pathways must be specifically wired in this antiparallel way. This can be achieved by using chemical signals to guide specific targeting of axons, similar to how known mechanisms for specific wiring operate in the brain (McLaughlin & O'Leary, 2005; Huberman et al., 2008). One example scheme of how this can be achieved is shown in Figure 4. While the picture in Figure 4a is complex, most of the complexity comes from the fact that units in a ConvNet produce inconsistent outputs (i.e., both positive and negative). If the units are consistent (i.e., producing exclusively positive or negative outputs), the picture simplifies to Figure 4b. Neurons in the brain are observed to be consistent, as stated by the so-called "Dale's Law" (Dale, 1935; Strata & Harvey, 1999). Hence, this constraint would have to be incorporated at some point in any biologically plausible network, and remains an important direction for future work. A second desideratum, related to the first, is that weights should not change sign during training. In the brain, the sign of a connection weight depends on the type of the presynaptic neuron—e.g., glutamatergic (excitatory) or GABAergic (inhibitory)—a quality that is intrinsic to and stable for each neuron as far as empirical evidence shows. Hence, if sign-symmetry is satisfied initially—for example, through specific wiring as just described—it will be satisfied throughout learning. Thus, evaluating the capacity of sign-fixed networks to learn is another direction for future work.

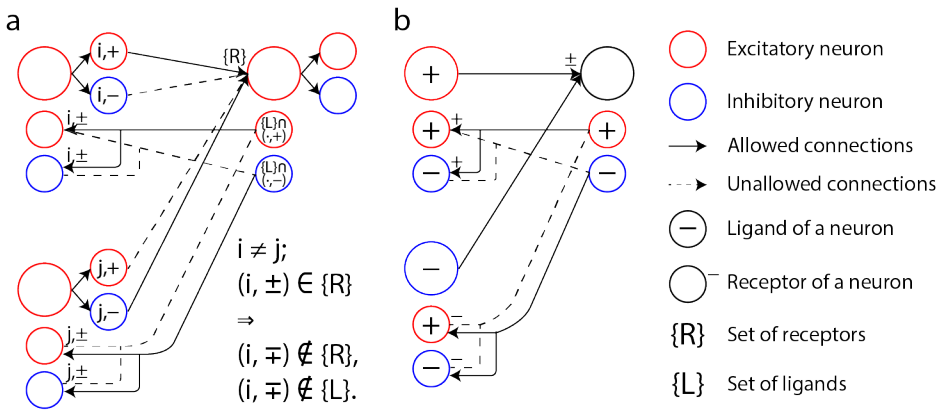


Figure 4: The specific wiring required for sign-symmetric feedback can be achieved using axonal guidance by specific receptor-ligand recognition. Assume that an axon carrying ligand L_X will only synapse onto a downstream neuron carrying the corresponding receptor R_X. By expressing receptors and ligands in an appropriate pattern, an antiparallel wiring pattern can be established that supports sign-symmetric feedback. Here, we show examples of ligand-receptor configurations that can lead to sign-symmetric wiring between two layers. **a**, a scheme for a sign-symmetric network with inconsistent units (i.e., units that produce both positive and negative outputs), where one inconsistent unit in the network is implemented by a trio of biological neurons in this scheme. Each directed connection is exclusively positive or negative. The ligand-receptor configuration shown ensures that a connection is paired with an antiparallel connection. $n_{\text{input neurons}}$ orthogonal ligand-receptor pairs is sufficient to implement all possible connectivities. **b**, a scheme to implement a sign-symmetric network with consistent units (i.e., units the produce either all positive or all negative outputs). Only 2 orthogonal ligand-receptor pairs are needed to implement all possible connectivities in this case.

Another element of unclear biological plausibility, common to feedback alignment and sign-symmetry, is that update of a synaptic connection (i.e., weight) between two feedforward neurons (A to B) depends on the activity in a third, feedback neuron C, whose activation represents the error of neuron B. One way it can be implemented biologically is for neuron C to connect to B with a constant and fixed weight. When C changes its value due to error feedback, it will directly induce a change of B's electric potential and indirectly trigger a change of the postsynaptic potential of the synapse between A and B, which might lead to either Long-term Potentiation (LTP) or Long-term Depression (LTD) of synapse A-B.

Other biological constraints include removing weight-sharing in convolutional layers as in Bartunov et al. (2018), incorporating temporal dynamics as in Lillicrap et al. (2016), using realistic spiking neurons, and so on. We believe that these issues are secondary to the problem of weight transport and that by removing the latter, we have taken a meaningful step toward biological plausibility. Nevertheless, many steps remain in the quest for a truly plausible, effective, and empirically-verified model of learning in the brain.

5 Conclusion

Recent work shows that biologically plausible learning algorithms do not scale well to more challenging problems such as ImageNet, but the tested conditions are quite limited. We evaluate sign-symmetry and re-evaluate feedback alignment on their effectiveness for training standard architectures on large tasks, namely training ResNet and AlexNet on ImageNet and RetinaNet on MS COCO, and find that 1) sign-symmetry performs nearly as well on ImageNet as does backpropagation, 2) slightly modified feedback alignment performs better than previously reported , and 3) both algorithms work to some extent on MS COCO with minimal hyperparameter tuning. Taken together, these results indicate that biologically plausible learning algorithms, in particular sign-symmetry, remain promising options for training artificial neural networks and modeling learning in the brain.

REFERENCES

Sergey Bartunov, Adam Santoro, Blake A. Richards, Geoffrey E. Hinton, and Timothy P. Lillicrap. Assessing the scalability of biologically-motivated deep learning algorithms and architectures. *Neural Information Processing Systems (NIPS)*, 2018.

F. Crick. The recent excitement about neural networks. *Nature*, 337:129–132, 1989.

H. Dale. Pharmacology and Nerve-endings (Walter Ernest Dixon Memorial Lecture): (Section of Therapeutics and Pharmacology). *Proc. R. Soc. Med.*, 28(3):319–332, Jan 1935.

Jia Deng, Wei Dong, Richard Socher, Li jia Li, Kai Li, and Li Fei-fei. Imagenet: A large-scale hierarchical image database. In *In CVPR*, 2009.

Jim DiCarlo. The science of natural intelligence: Keynote lecture at CVPR17, 2017. `https://youtu.be/ilbbVkIhMgo?t=36m48s` [Accessed: Sep 21, 2018].

Stephen Grossberg. Competitive learning: From interactive activation to adaptive resonance. *Cognitive science*, 11(1):23–63, 1987.

K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, June 2016. doi: 10.1109/CVPR.2016.90.

Geoffrey E. Hinton. How to do backpropagation in a brain. *Invited talk at the NIPS'2007 Deep Learning Workshop*, 2007.

Andrew D. Huberman, Marla B. Feller, and Barbara Chapman. Mechanisms underlying development of visual maps and receptive fields. *Annual Review of Neuroscience*, 31(1):479–509, jul 2008. doi: 10.1146/annurev.neuro.31.060407.125533. URL `https://doi.org/10.1146/annurev.neuro.31.060407.125533`.

Alex Krizhevsky. One weird trick for parallelizing convolutional neural networks. *CoRR*, abs/1404.5997, 2014. URL `http://arxiv.org/abs/1404.5997`.

Qianli Liao, Joel Z. Leibo, and Tomaso Poggio. How important is weight symmetry in backpropagation? *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, pp. 1837–1844, 2016.

Timothy P. Lillicrap, Daniel Cownden, Douglas B. Tweed, and Colin J. Akerman. Random synaptic feedback weights support error backpropagation for deep learning. *Nature Communications*, 7, 2016.

Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. *arXiv preprint arXiv:1708.02002*, 2017.

Todd McLaughlin and Dennis D.M. O'Leary. MOLECULAR GRADIENTS AND DEVELOPMENT OF RETINOTOPIC MAPS. *Annual Review of Neuroscience*, 28(1):327–355, jul 2005. doi: 10.1146/annurev.neuro.28.061604.135714. URL `https://doi.org/10.1146/annurev.neuro.28.061604.135714`.

Tomaso Poggio, Hrushikesh Mhaskar, Lorenzo Rosasco, Brando Miranda, and Qianli Liao. Why and when can deep-but not shallow-networks avoid the curse of dimensionality: a review. *International Journal of Automation and Computing*, 14(5):503–519, 2017.

Piergiorgio Strata and Robin Harvey. Dales principle. *Brain Research Bulletin*, 50(5):349 – 350, 1999. ISSN 0361-9230. doi: https://doi.org/10.1016/S0361-9230(99)00100-8. URL `http://www.sciencedirect.com/science/article/pii/S0361923099001008`.