Locally connected networks as ventral stream models

Anonymous Author(s)
Affiliation
Address
email

Abstract

Most deep learning models of the ventral stream, and convolutional networks in particular, share weights among neurons. Weight sharing during learning is crucial for good performance on image recognition tasks, but it is not biologically plausible. In this work, we compare performance and Brain-Score results of ImageNet-trained networks in multiple configurations: convolutional, locally connected (i.e., convolutional without weight sharing), and locally connected with anti-Hebbian plasticity mechanisms that promote weight sharing. We also study the role of initialization on performance of those networks. We find that the more weight sharing networks have, the better they perform on both ImageNet and Brain-Score, which can sometimes be further improved with a convolutional initialization. Moreover, locally connected networks outperform their convolutional counterpart on purely neural data (areas V1, V2, V4, IT), but not on behavioral data. However, ImageNet performance negatively correlates with correspondence to V1 data, suggesting that better models of early visual processing don't necessarily provide a good input for models of deeper visual areas.

1 Introduction

Convolutional networks not only have great performance on image recognition tasks, but also develop representations similar to the primate visual stream. For instance, they provide a good fit to multiple areas of the ventral stream Schrimpf et al. (2018), and even explain the separation between the ventral and the dorsal streams Bakhtiari et al. (2021). However, the biological plausibility of convolutional networks is questionable, since they need to share weights among neurons during training (Fig. 1A). If they don't (even if they are locally connected, Fig. 1B), they perform much worse on image recognition tasks, and show a worse fit to the visual stream Pogodin et al. (2021). This issue is not limited to convolutional networks – any network involving multiplication of a matrix of neurons with a weight matrix needs weight sharing (e.g., a recently popular architecture, transformers). Recently it was shown that locally connected networks (Fig. 1B) can induce weight sharing through anti-Hebbian plasticity, but the network must stop training for a "sleep phase" that uses lateral connectivity in each layer Pogodin et al. (2021) (Fig. 1C,D).

In this work, we study how well locally connected and convolutional networks correspond to the ventral stream data, as measure by the Brain-Score Schrimpf et al. (2018, 2020), which combines recordings of behavior, and areas V1, V2, V4 and IT in primates. We evaluate the role of initial conditions, frequency of weight sharing (as in Pogodin et al. (2021)) and some architectural differences. We find that better performing models typically result in a better Brain-Score, but locally connected networks provide a better match to non-behavioral data. In addition, poorly performing models match better to the V1 data.

Submitted to 1st Brain-Score Workshop (BSW 2022).

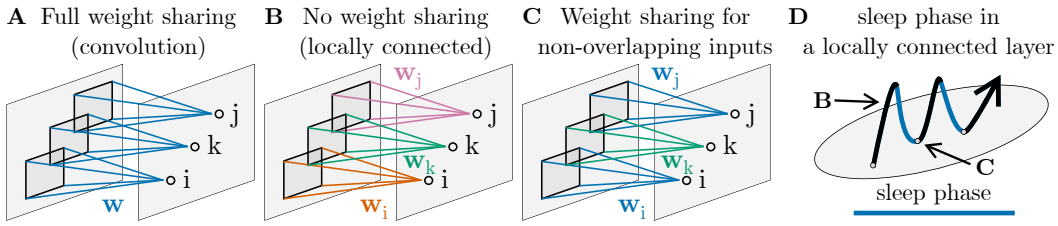

Figure 1: **A.** Convolutional layer: all neurons see a patch of the input and have the same weight $\mathbf{w}_i = \mathbf{w}_k = \mathbf{w}_j = \mathbf{w}$. **B.** Locally connected layer: same, but with different weights. **C.** Only neurons with non-overlapping inputs share weights (equivalent to a stack of strided convolutions). **D.** Training in a locally connected layer leads to different weights (like in **B**); a sleep phase (Pogodin et al. (2021)) shares weights among neurons with non-overlapping inputs (like in **C**).

2 Convolutional vs. locally connected networks: ImageNet and Brain-Score performance

Locally connected (LC) networks with a weight sharing sleep phase can achieve results similar to their convolutional counterpart Pogodin et al. (2021). However, Pogodin et al. (2021) only considered a relatively small model (half-width ResNet-18) and didn't study the role of initial conditions and architectural changes.

Here, we conduct experiments on a standard ResNet-18 in multiple configurations: convolutional, locally connected with random initialization (as in Pogodin et al. (2021)), locally connected with a convolutional initialization (as in d'Ascoli et al. (2019)), and locally connected with a convolutional first layer. As the first layer in a ResNet-18 has the highest resolution (224 by 224 with stride 2 vs. 56 by 56 with stride 1 for the next one) and the largest filter size (7 by 7 vs. 3 by 3 for all other layers), it's potentially prone to more overfitting than other layers. Making it convolutional is inspired by pre-V1 processing done in the visual system.

2.1 Training details

We trained all networks on ImageNet Deng et al. (2009) with standard augmentations. We used AdamW Loshchilov & Hutter (2017) with a batch size of 128. We used Nvidia A100 GPUs (40GB VRAM; for GPUs with less memory, the model would not fit on a single one as just the LC weights take up about 11GB). We used the implementation from Pogodin et al. (2021), which is available at https://github.com/romanpogodin/towards-bio-plausible-conv.

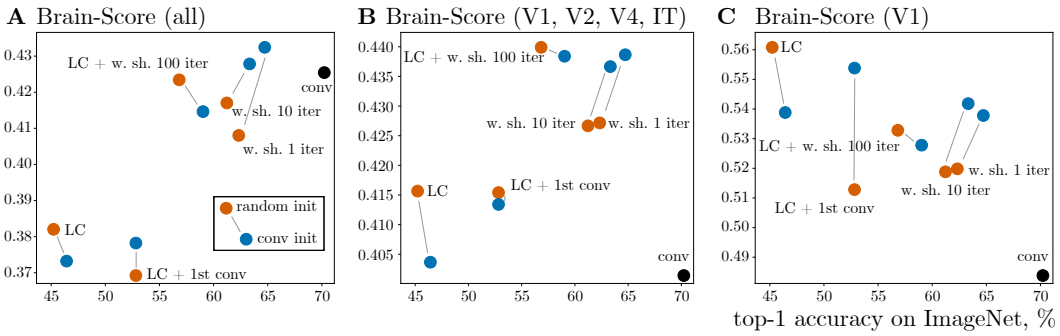

Figure 2: ImageNet top-1 accuracy vs. Brain-Score for several ResNet-18 networks: convolutional, locally connected (LC), LC with a convolutional first layer, LC with a weight sharing sleep phase every 1/10/100 iterations. Orange: randomly initialized LC net; blue: LC net with a convolutional initialization. **A.** Average Brain-Score (V1, V2, V4, IT, behavioral data). **B.** Same, but without behavioral data. **C.** Only V1 data.

2.2 Results (all brain areas)

First, we compared ImageNet top-1 test accuracy and the Brain-Score of all our models (Fig. 2A). Overall, accuracy correlates with Brain-Score, but there are two interesting trends. First, a convolutional initialization (blue) is always better than a random one (orange) on ImageNet (except for LC + 1st conv), but not always on Brain-Score (for frequent weight sharing, both metrics improve from a convolutional initialization). Second, locally connected networks with a convolutional initialization (blue) show a monotonic increase of Brain-Score with ImageNet accuracy, but an even more accurate fully convolutional network achieves a worse Brain-Score.

Both trends are even more pronounced if we only consider neural data, without the behavioral one (Fig. 2B). Even more, a fully convolutional network shows the worst overall fit to neural data despite the best performance. One possible explanation is that locally connected networks (even with weight sharing in neurons with non-overlapping inputs) produce a more diverse set of activations, as there's less regularity in the weights.

2.3 Results (V1)

All brain regions, expect V1, show a similar correlation of accuracy and Brain-Score. V1 is the complete opposite: performance is negatively correlated with the V1 score (Fig. 2C).

First, we checked if representations in earlier layers (that typically match to V1) are any useful for ImageNet, and if it correlates with the V1 score. We trained a linear readout on top of the second ResNet block (out of four), and observed very poor performance (2-6%) in purely LC networks, as opposed to 8-20% in networks with more regularization. Therefore, good V1 representations are not very helpful for ImageNet.

We also trained a shorter ResNet with only two blocks (10 layers total) to see if depth of the model affects the phenomenon. It doesn't: the V1 score of an LC network is 0.544, vs. 0.511 for a convolutional one that performs better on ImageNet. However, the difference in score is twice as small for the shorter network compared to a full ResNet-18.

Finally, we tested untrained networks to see if different initial conditions can explain V1 scores. For a randomly initialized locally connected network, the V1 score was predictably low: 0.347 (much smaller than everything in Fig. 2C). A convolutional initialization, however, resulted in a score of 0.516 – *higher than for a trained convolutional network*.

3 Discussion

We explored how deep networks that substitute implausible weight sharing of convolutions with realistic mechanisms, namely local connections and sleep-phase induced partial weight sharing, perform on ImageNet and Brain-Score. We found that those two metrics generally correlate, and that a convolutional initialization of locally connected networks can improve both of them. Moreover, we found that V1 performance *negatively* correlates with ImageNet accuracy, both for full networks and for representations extracted from V1-like early layers.

The found discrepancy between V1 results and all other brain areas shows that training (on ImageNet) is not always beneficial for all architectures: while locally connected networks with any initialization end up with a higher match to the V1 data after training, a fully convolutional one sees a decrease in V1 match. However, training always increases the match to later areas. We see that a good V1 model might not be the best input for a good V2 model, which is somewhat consistent with the idea that the visual cortex is not strictly hierarchical St-Yves et al. (2022).

Overall, our results show that deep learning models of the visual stream can be improved by adding realistic training constraints, such as the lack of weight sharing at all times, to them. The improvement, however, is only visible when we compare the models to real data (in our case, via Brain-Score Schrimpf et al. (2018, 2020)), and not to machine learning benchmarks.

References

Shahab Bakhtiari, Patrick J Mineault, Tim Lillicrap, Christopher C Pack, and Blake Aaron Richards. The functional specialization of visual cortex emerges from training parallel pathways with self-supervised predictive learning. In *Thirty-Fifth Conference on Neural Information Processing Systems*, 2021.

Stéphane d'Ascoli, Levent Sagun, Giulio Biroli, and Joan Bruna. Finding the needle in the haystack with convolutions: on the benefits of architectural bias. *Advances in Neural Information Processing Systems*, 32, 2019.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255. Ieee, 2009.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.

Roman Pogodin, Yash Mehta, Timothy Lillicrap, and Peter Latham. Towards biologically plausible convolutional networks. *Advances in Neural Information Processing Systems*, 34, 2021.

Martin Schrimpf, Jonas Kubilius, Ha Hong, Najib J. Majaj, Rishi Rajalingham, Elias B. Issa, Kohitij Kar, Pouya Bashivan, Jonathan Prescott-Roy, Franziska Geiger, Kailyn Schmidt, Daniel L. K. Yamins, and James J. DiCarlo. Brain-score: Which artificial neural network for object recognition is most brain-like? *bioRxiv preprint*, 2018. URL `https://www.biorxiv.org/content/10.1101/407007v2`.

Martin Schrimpf, Jonas Kubilius, Michael J Lee, N Apurva Ratan Murty, Robert Ajemian, and James J DiCarlo. Integrative benchmarking to advance neurally mechanistic models of human intelligence. *Neuron*, 2020. URL `https://www.cell.com/neuron/fulltext/S0896-6273(20)30605-X`.

Ghislain St-Yves, Emily J Allen, Yihan Wu, Kendrick Kay, and Thomas Naselaris. Brain-optimized neural networks learn non-hierarchical models of representation in human visual cortex. *bioRxiv*, 2022.