Easing into learning: How and why a sequential curriculum improves visual learning generalization

Charlotte Volk (charlotte.volk@mail.mcgill.ca)
Department of Bioengineering, McGill University
845 Rue Sherbrooke O
Montréal, QC H3A 0G4 Canada

Mila - Quebec AI Institute
6666 rue St. Urbain
Montréal, QC, H2S 3H1 Canada

Christopher Pack (christopher.pack@mcgill.ca)
Department of Neurology & Neurosurgery, McGill University
845 Rue Sherbrooke O
Montréal, QC H3A 0G4 Canada

Montreal Neurological Institute-Hospital
3801 University Street
Montreal, QC H3A 2B4

Shahab Bakhtiari (shahab.bakhtiari@umontreal.ca)
Department of Psychology, Université de Montréal
90 Ave Vincent D'Indy
Outremont, QC H2V 2S9 Canada

Mila - Quebec AI Institute
6666 rue St. Urbain
Montréal, QC, H2S 3H1 Canada

Abstract

Generalization in human visual learning (VL) varies across tasks, with 'easy' visual tasks (e.g., large angle orientation discrimination) generalizing better to unseen conditions than 'hard' ones (e.g., small angle orientation discrimination). We used an artificial neural network (ANN) model to explore how training on a sequential curriculum (easy $\rightarrow$ intermediate $\rightarrow$ hard) enhances VL generalization. Our findings revealed that the dimensionality of the representational readout subspace, established during the initial training phase, is crucial for generalization. Specifically, 'harder' tasks in later stages can 'piggyback' on the low-dimensional, more generalizable subspace established during the 'easier' initial training phase, leading to a more generalizable outcome.

Keywords: visual learning; artificial neural networks; plasticity; perception

Introduction

Practicing a visual task leads to long-lasting perceptual improvements known as Visual Learning (VL) (B. Dosher & Lu, 2017). However, the generalization of VL to unseen conditions varies across different tasks (Ahissar & Hochstein, 1997). Particularly, it has been shown that learning 'easy' visual tasks leads to better generalization than learning 'hard' ones (Ahissar & Hochstein, 1997). Moreover, experimental evidence suggests that greater generalizability in VL can be achieved through sequential curriculum training, specifically by training on 'easier' versions of a task before progressing to 'harder' versions (James, 1890; Pavlov, 1927; North, 1959; Wisniewski, Radell, Church, & Mercado, 2017). Yet, a neurocomputational explanation for this curriculum learning phenomenon remains elusive. In this study, we leveraged an artificial neural network (ANN) model of VL (Wenliang & Seitz, 2018) to gain insight into the variability of generalization in VL. Our findings indicate that the subspace of visual representations that influence the model's behavior, known as the *readout subspace*, plays a pivotal role in generalization: tasks resulting in a lower-dimensional readout subspace demonstrate enhanced generalization. Furthermore, within the sequential curriculum learning paradigm, harder tasks can 'piggyback' (Wang, Zhang, Klein, Levi, & Yu, 2014) on the low-dimensional subspace established through learning an initial, easier task, thereby enhancing their generalization.

Results

We used an Artificial Neural Network (ANN) model, which had previously replicated several behavioral and neuronal attributes of Visual Learning (VL) (Wenliang & Seitz, 2018). Additionally, we modified the model by incorporating readout weights, or 'skip connections', from every layer of the model to the output decision neuron (Figure 1a). The presence of these skip connections has been supported by our understanding of the anatomy (Felleman & Van Essen, 1991) and the physiology of the visual system (Liu & Pack, 2017). We fine-tuned our modified model using the backpropagation algorithm on an orientation discrimination task. In this task, the model was trained to classify the rotation direction of a Gabor stimulus relative to a reference orientation. To evaluate the model's generalization capabilities, we tested it on Gabor stimuli that maintained the training angle separation but with doubled spatial frequency (SF).

First, our ANN model successfully reproduced the main behavioral characteristics of VL. Training losses decreased more rapidly for tasks with larger angle separations ('easier' tasks) compared to those with smaller ones ('harder' tasks) (Figure 1b; left). Moreover, we evaluated the model's generalization using the specificity index (SI), a normalized measure of the difference between test and training losses at the end of training, where a smaller SI indicates better generalization. Our model replicated the generalization pattern previously observed in humans, namely, lower generalization (i.e., higher specificity) for smaller angle separations (Ahissar & Hochstein, 1997) (Figure 1b; right). Furthermore, to pinpoint the loci of plasticity that underlied learning, we examined the connection weights throughout training in our model (both between-layer and skip connections) and found that the norm of weight changes in the skip connections was significantly larger than that in the convolutional weights. This suggests that learning predominantly occurred in the readout weights. This finding supports the previously posited hypothesis (B. A. Dosher, Liu, Chu, & Lu, 2020) that the presence of direct skip connection to downstream areas drives learning to occur primarily in the readout weights of visual areas, rather than in the sensory representations of the stimuli.

After reproducing the pattern of generalizations across various task difficulties, we leveraged this ANN model of VL to simulate a sequential VL paradigm and explore its potential effect on VL generalization. We trained two versions of our model, one sequentially (larger $\rightarrow$ smaller angle separations) and one non-sequentially (only small angle separation). We found that sequential training led to improved generalization for the smallest angle separation (0.5 degrees) compared to non-sequential training (SI_{seq} = 0.31, SI_{nonseq} = 0.47). To understand how sequential training led to better generalization, we next examined the learned readout weights (i.e., the skip connection weights) of the sequentially and non-sequentially trained models. First, we observed that the readout weights followed two separate learning trajectories during sequential vs non-sequential learning (Figure 1c; a two-dimensional PCA projection of the readout weights), indicating a geometric difference in the final readout weights between the two training paradigms. This suggests that, although the two models (i.e., sequentially and non-sequentially trained) were performing the same task of small-angle discrimination at the end of their training, they were reading out from two distinct subspaces of the coding representational space, and importantly, the subspace discovered via sequential training happened to be more generalizable to test conditions than the subspace discovered in non-sequential training. Upon examining the distribution of

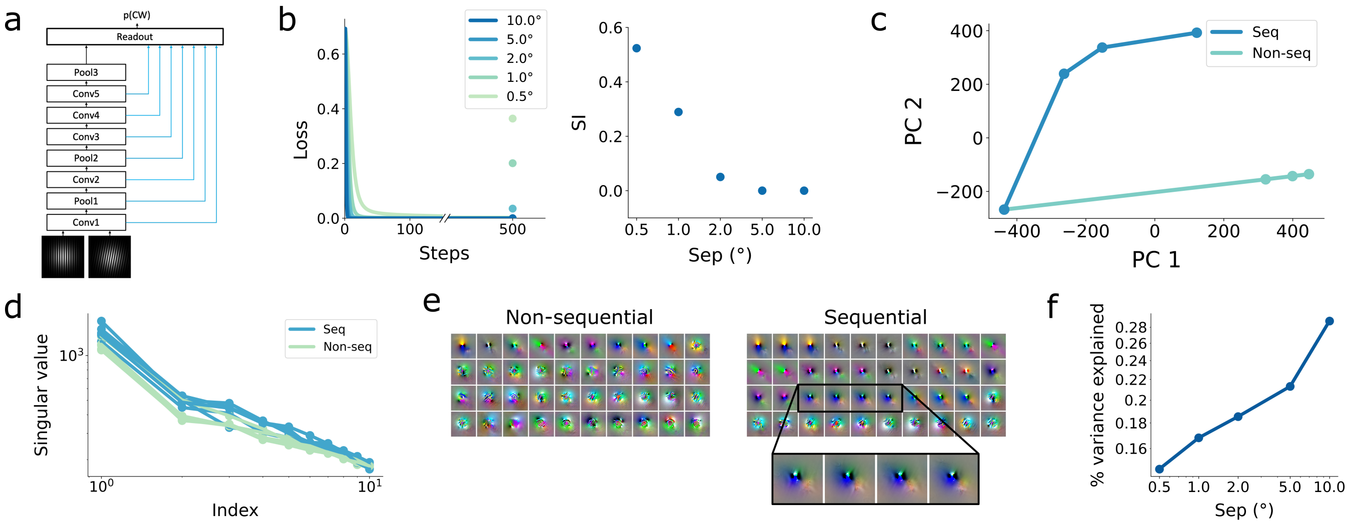

Figure 1: (a) Schematic of the model with skip connections (skip connections are blue; between-layer connections are black) and the VL task. (b) Left: Average learning curves and single sample test losses (shown as single dots) after training. Right: specificity index as a function of angle separation. (c) Projection of readout weights onto their first two principal components for sequential and non-sequential training. (d) Top 10 singular values of the activation matrix of neurons with the largest readout weights from the sequential & non-sequential models. (e) Sample generated preferred visual features for the neurons with the largest readout weights in the sequentially and non-sequentially trained models. (f) Percentage of variance explained by the first singular value as a function of training angle separation.

learned readout weights in both models, we observed a highly skewed distribution, with only a small portion of non-zero readout weights. This implies that only a small subset of neurons significantly contributed to defining the readout subspace of each model, and therefore their learning generalization.

To probe these learned readout subspaces, we estimated their dimensionality across various training tasks and conditions. From both the sequentially and non-sequentially trained models, we collected the activations of 100 “important” neurons with the largest readout weight amplitudes in response to 100 natural images. Applying singular value decomposition (SVD) on the activation matrix of each model, we discovered that the first 10 singular values of the sequential model’s activations were significantly larger than those of the non-sequential model (Figure 1d). This suggests that the readout subspace of the sequentially trained model occupied a lower-dimensional representational space. Moreover, visualizing the preferred visual features (Olah, Mordvintsev, & Schubert, 2017) for the “important” 100 neurons of each model revealed visually similar ‘groups’ of features in the sequentially trained model, while the preferred visual features in the non-sequentially trained model were much more distinct from each other (Figure 1e). The similarity of preferred visual features across neurons supports the conclusion that the sequentially trained model benefits from reading out a representational subspace with low dimensionality. Furthermore, we found that this reduced dimensionality of the sequentially trained model is largely inherited from the first, easy training phase. Across different difficulty levels, training non-sequentially on larger angle separations (i.e., easier tasks) led the model to discover a lower-dimensional readout subspace of neurons (Figure 1f). This suggests that readout dimensionality, determined by the ‘difficulty’ of the training task, may explain the variability in generalization across training conditions.

To test this hypothesis, we trained our model, non-sequentially, on the hardest task condition, but froze the readout weights of all the neurons except for 100 neurons. Importantly, we selected these 100 unfrozen neurons from the “important” neurons of a model trained on an easy condition. This allowed the model to learn a hard task by searching within a lower-dimensional readout subspace established by a previously learned easy task. Indeed, we found that limiting the learning subspace of the model to a lower dimension improved generalization for the hard task ($SI_{limited} = 0.32$, $SI_{unlimited} = 0.47$).

Conclusion

Our results suggest that the dimensionality of the readout subspace, established through visual training, may be responsible for the observed variability in learning generalization. We have demonstrated that sequential curriculum learning—starting with an easier version of a task and progressing to harder ones—leads to a lower-dimensional readout subspace, which may underlie its improved generalization.

Acknowledgments

This project is funded by NSERC Discovery grants to SB (RGPIN-2023-03875) and CCP (RGPIN/04333-2017). CV was also funded by an NSERC USRA.

References

Ahissar, M., & Hochstein, S. (1997, 05). Task difficulty and the specificity of perceptual learning. *Nature*, *387*(6631), 401–406.

Dosher, B., & Lu, Z.-L. (2017). Visual perceptual learning and models [Journal Article]. *Annual Review of Vision Science*, *3*(Volume 3, 2017), 343-363. Retrieved from https://www.annualreviews.org/content/journals/10.1146/annurev-vision-102016-061249 doi: https://doi.org/10.1146/annurev-vision-102016-061249

Dosher, B. A., Liu, J., Chu, W., & Lu, Z.-L. (2020, 06). Roving: The causes of interference and re-enabled learning in multi-task visual training. *Journal of Vision*, *20*(6), 9-9. Retrieved from https://doi.org/10.1167/jov.20.6.9 doi: 10.1167/jov.20.6.9

Felleman, D., & Van Essen, D. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral cortex (New York, N.Y. : 1991)*, *1*(1), 1—47. Retrieved from https://academic.oup.com/cercor/article-pdf/1/1/1/1139605/1-1-1.pdf doi: 10.1093/cercor/1.1.1-a

James, W. (1890). *The principles of psychology, vol I*. New York: Henry Holt and Co.

Liu, L. D., & Pack, C. C. (2017). The contribution of area mt to visual motion perception depends on training. *Neuron*, *95*(2), 436-446.e3. Retrieved from https://www.sciencedirect.com/science/article/pii/S0896627317305482 doi: https://doi.org/10.1016/j.neuron.2017.06.024

North, A. J. (1959, June). Acquired distinctiveness of form stimuli. *J. Comp. Physiol. Psychol.*, *52*(3), 339–341. doi: 10.1037/h0045756

Olah, C., Mordvintsev, A., & Schubert, L. (2017). Feature visualization. *Distill*. (https://distill.pub/2017/feature-visualization) doi: 10.23915/distill.00007

Pavlov, I. (1927). *Conditioned reflexes*. London: Oxford University Press.

Wang, R., Zhang, J.-Y., Klein, S. A., Levi, D. M., & Yu, C. (2014, 11). Vernier perceptual learning transfers to completely untrained retinal locations after double training: A "piggybacking" effect. *Journal of Vision*, *14*(13), 12-12. Retrieved from https://doi.org/10.1167/14.13.12 doi: 10.1167/14.13.12

Wenliang, L., & Seitz, A. R. (2018). Deep neural networks for modeling visual perceptual learning. *Journal of Neuroscience*. Retrieved from https://www.jneurosci.org/content/early/2018/05/23/JNEUROSCI.1620-17.2018 doi: 10.1523/JNEUROSCI.1620-17.2018

Wisniewski, M. G., Radell, M. L., Church, B. A., & Mercado, E., III. (2017, 07). Benefits of fading in perceptual learning are driven by more than dimensional attention. *PLOS ONE*, *12*(7), 1-20. Retrieved from https://doi.org/10.1371/journal.pone.0180959 doi: 10.1371/journal.pone.0180959