Learning trajectories of deep neural networks during self-supervised visual representation learning

Ehsan Tousi (ekahooka@uwo.ca)
Graduate Program in Neuroscience, Schulich School of Medicine and Dentistry, Western University, London ON, Canada

Chelsea Kim (bkim292@uwo.ca)
Graduate Program in Neuroscience, Schulich School of Medicine and Dentistry, Western University, London ON, Canada

Marieke Mur (mmur@uwo.ca)
Department of Psychology, Western University, London ON, Canada
Department of Computer Science, Western University, London ON, Canada

Abstract

Recent work has started exploring the possibility of using self-supervised deep learning as a framework for modeling human visual development. Here, we provide a first step in that direction, by examining learning trajectories of a deep feedforward neural network, ResNet50, as it is trained on ImageNet using self-supervised contrastive learning. We ask if the learning trajectories show developmental signatures similar to those observed in the primate visual system. We show that representations change rapidly during the first few training epochs, and then stabilize. Like in the primate visual system, visual representations stabilize faster in early than in deep layers. Within- and between-category information emerge simultaneously, consistent with the notion that self-supervised contrastive learning promotes both. Our work provides preliminary support for using self-supervised deep learning to model human visual development, which opens up the possibility of systematically testing how developmental constraints shape visual representations.

Keywords: vision; representation learning; self-supervised learning; deep neural networks; learning trajectory; visual development

Deep neural networks as models of visual representation learning

Deep neural networks are popular system-level computational models of the primate visual system. Deep neural networks rival human performance on image classification tasks and predict brain activity across primate ventral visual cortex during image viewing (He, Zhang, Ren, & Sun, 2015; Yamins et al., 2014; Khaligh-Razavi & Kriegeskorte, 2014; Güçlü & Gerven, 2015). Initial modeling efforts relied on category supervision for visual representation learning, but recent years have seen a shift to self-supervised learning objectives. Self-supervised approaches are thought to more closely simulate human learning goals during development, and are on par with category supervision when it comes to predicting brain activity in adults (Zhuang et al., 2021; Konkle & Alvarez, 2022). Given these successes, recent work has started exploring the idea of using self-supervised visual representation learning as a framework for modeling human visual development (Zaadnoordijk, Besold, & Cusack, 2022; Mur, 2023). Here we provide a first step in that direction, by characterizing representational learning trajectories of deep neural networks during self-supervised visual learning, and by examining if they show developmental signatures consistent with primate visual development.

Characterizing learning trajectories of deep neural networks

We focus on a feedforward convolutional neural architecture, ResNet50 (He et al., 2015), and train multiple instances of this architecture (Mehrer, Spoerer, Kriegeskorte, & Kietzmann, 2020) for 200 training epochs on ImageNet using self-supervised contrastive learning (He, Fan, Wu, Xie, & Girshick, 2020; Chen, Fan, Girshick, & He, 2020). We characterize model learning trajectories by presenting a test image set after each training epoch. The test set consists of 96 object images from a range of real-world categories, including faces and animals (Kriegeskorte et al., 2008). After each training epoch, we extract response patterns to the test images from the last layer of the first and last convolutional block ('layer 1' and 'layer 4', respectively), and compute representational dissimilarity matrices (RDMs), which summarize the image information carried by the response patterns. We used correlation distance as a dissimilarity measure. The RDMs across epochs form a representational learning trajectory. These trajectories allow for detailed tracking of the learning dynamics.

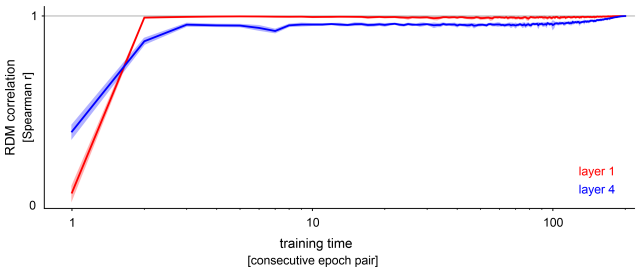

Figure 1: Representational change from one training epoch to the next for an early (layer 1) and a deep network layer (layer 4). Results are averaged across 10 network instances.

Rapid representational changes early in learning

Based on prior neuroscience work (Espinosa & Stryker, 2012; Livingstone et al., 2017), we expect to see rapid representational changes early in learning, followed by representational stability later in learning. We test this hypothesis by assessing representational change from one training epoch to the next. To do so, we correlate RDMs of consecutive training epochs, and plot these correlations as a function of training time. This allows us to observe the rate of representational change as training unfolds. Results displayed in Figure 1 indeed show rapid representational changes in the first few training epochs for both early and deep network layers, after which representations stabilize. The observed pattern of representational change is consistent with prior modeling work (Achille, Rovere, & Soatto, 2019; Hong, Yamins, Majaj, & DiCarlo, 2016; Zhuang et al., 2021) and confirms predictions from the neuroscience literature.

Early layers settle more rapidly than deep layers

Prior neuroscience work also predicts that representations in early network layers, which are thought to correspond to early visual cortex, stabilize more rapidly than representations in deep network layers, which are thought to correspond to higher-level visual cortex (Espinosa & Stryker, 2012; Livingstone et al., 2017; Seibert, 2018; Güçlü & Gerven, 2015). Figure 1 indeed suggests that representations stabilize more rapidly in layer 1 than in layer 4.

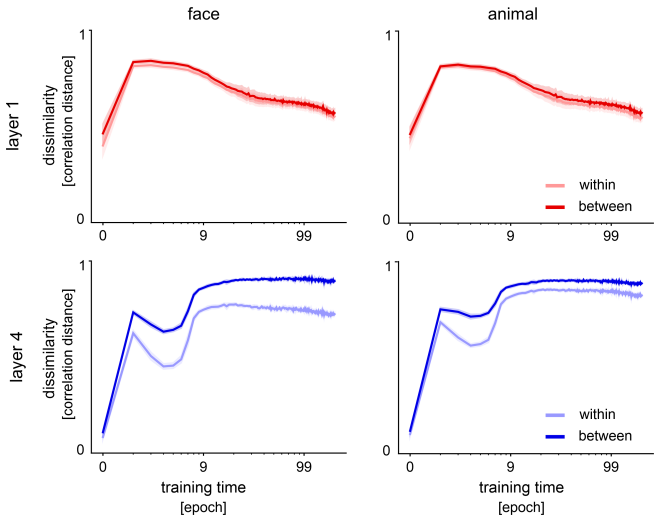

Figure 2: Within- and between-category information as a function of training time. Results are averaged across 10 network instances. Shaded areas show standard deviation across network instances.

Within- and between-category information emerge simultaneously

While the neuroscience literature may not provide clear predictions about the relative timing of within- and between-category information during visual learning, behavioral pressures exist for both (Grill-Spector & Weiner, 2014), and self-supervised contrastive learning is expected to promote the emergence of within- as well as between-category information (Konkle & Alvarez, 2022). Figure 2 shows how within- and between-category information evolves over learning for two categories of longstanding behavioral relevance, faces and animals. We extracted within- and between-category dissimilarites from the RDMs and averaged across all within- and all between-category image pairs. Results indicate that within- and between-category information arise simultaneously in both early and deep layers. Interestingly, the average dissimilarities in layer 4 are higher for between-category than within-category image pairs. This trend was not observed in layer 1. This suggests that over time, deeper layers prioritize distinguishing between category members and non-members over doing so between different category members, while early layers have less of a preference.

Conclusion

Learning trajectories of deep neural networks show developmental signatures consistent with those observed in the primate visual system. Signatures include rapid representational changes early in learning, and more rapid stabilization of representations in early than deep network layers. Our results further indicate that within- and between-category information emerge simultaneously, suggesting they may both arise from a general push for distinguishing object images. Our findings provide preliminary support for using self-supervised deep learning as a framework for modeling human visual development.

Acknowledgments

This work was funded by an NSERC Discovery Grant to MM, and a Canada Graduate Scholarship to CK.

References

Achille, A., Rovere, M., & Soatto, S. (2019). *Critical Learning Periods in Deep Neural Networks.* arXiv.

Chen, X., Fan, H., Girshick, R., & He, K. (2020). *Improved Baselines with Momentum Contrastive Learning.* arXiv.

Espinosa, J. S., & Stryker, M. P. (2012). Development and Plasticity of the Primary Visual Cortex. *Neuron*, *75*, 230–249.

Grill-Spector, K., & Weiner, K. S. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. *Nature Reviews Neuroscience*, *15*, 536–548.

Güçlü, U., & Gerven, M. A. J. v. (2015). Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream. *Journal of Neuroscience*, *35*, 10005–10014.

He, K., Fan, H., Wu, Y., Xie, S., & Girshick, R. (2020). Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 9729–9738).

He, K., Zhang, X., Ren, S., & Sun, J. (2015). *Deep Residual Learning for Image Recognition.* arXiv.

Hong, H., Yamins, D. L. K., Majaj, N. J., & DiCarlo, J. J. (2016). Explicit information for category-orthogonal object properties increases along the ventral stream. *Nature Neuroscience*, *19*, 613–622.

Khaligh-Razavi, S.-M., & Kriegeskorte, N. (2014). Deep Supervised, but Not Unsupervised, Models May Explain IT Cortical Representation. *PLOS Computational Biology*, *10*, e1003915.

Konkle, T., & Alvarez, G. A. (2022). A self-supervised domain-general learning framework for human ventral stream representation. *Nature Communications*, *13*, 491.

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., … Bandettini, P. A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, *60*, 1126–1141.

Livingstone, M. S., Vincent, J. L., Arcaro, M. J., Srihasam, K., Schade, P. F., & Savage, T. (2017). Development of the macaque face-patch system. *Nature Communications*, *8*, 14897.

Mehrer, J., Spoerer, C. J., Kriegeskorte, N., & Kietzmann, T. C. (2020). Individual differences among deep neural network models. *Nature Communications*, *11*, 5725.

Mur, M. (2023). Bridging visual developmental neuroscience and deep learning: challenges and future directions. *Journal of Vision*, *23*, 4680.

Seibert, D. D. A. (2018). *High-level visual object representation in juvenile and adult primates*. Thesis, Massachusetts Institute of Technology.

Yamins, D. L. K., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences*, *111*, 8619–8624.

Zaadnoordijk, L., Besold, T. R., & Cusack, R. (2022). Lessons from infant learning for unsupervised machine learning. *Nature Machine Intelligence*, *4*, 510–520.

Zhuang, C., Yan, S., Nayebi, A., Schrimpf, M., Frank, M. C., DiCarlo, J. J., & Yamins, D. L. K. (2021). Unsupervised neural network models of the ventral visual stream. *Proceedings of the National Academy of Sciences*, *118*.