Modeling Localized Synaptic Degeneration and Neural Plasticity in Visual Cortex

Yash Shah (ynshah@stanford.edu)
Department of Computer Science
Stanford University, Stanford, CA 94305 United States

Kevin Tran (ktran25@stanford.edu)
Department of Psychology
Stanford University, Stanford, CA 94305 United States

Daniel L. K. Yamins (yamins@stanford.edu)
Departments of Psychology and Computer Science
Wu Tsai Neurosciences Institute
Stanford University, Stanford, CA 94305 United States

Abstract

A patient diagnosed with a neurodegenerative disease of the visual system like Posterior Cortical Atrophy (PCA) or Visual Agnosia can have difficulty in recognizing faces, perceiving boundary shapes, discriminating color, and reading fragmented images. Computationally modeling the effects of neurodegeneration on different areas in the ventral visual stream (V1, V2, V4, and IT), while maintaining the mechanism of neural plasticity in the brain, can help inform effective rehabilitation strategies for the patient. To this end, we simulate localized synaptic degeneration and neural plasticity using CORnet-S—a deep artificial model of the brain. We observe that (1) different areas in the visual cortex have functionally different responses to tasks such as stimuli processing, shape discrimination, and face perception; and (2) even when a certain area in the visual cortex is ∼99.85% progressively lesioned, the cortex has a remarkable ability to recover task performance through perceptual learning.

Keywords: vision; synaptic degeneration; neural plasticity; deep artificial neural networks

Introduction

Neurodegenerative diseases of the visual system like PCA and Visual Agnosia result in disproportionate volume loss in the occipital and posterior parietal lobes of the biological brain, with patients having difficulty in performing tasks such as color discrimination, reading fragmented images, identifying faces, and perceiving boundary shapes (Milner et al., 1991; Maia da Silva, Millington, Bridge, & Plant, 2017). Deep artificial neural networks (ANNs) of the brain have been developed to predict responses in the healthy ventral stream to much success (Yamins et al., 2014; Kubilius et al., 2019; Finzi, Margalit, Kay, Yamins, & Grill-Spector, 2022), although computationally modeling it under degeneration has been sparse.

In terms of introducing lesions to the whole network, pruning methods abound; these methods aim at improving computation and storage efficiency by removing filters and/or weights in either a one-time, atemporal, fashion (Li, Kadav, Durdanovic, Samet, & Graf, 2017) or iteratively over time (Han, Pool, Tran, & Dally, 2015). On the other hand, recent works have explicitly tried to model and analyze global degeneration of the ventral stream by brain-scoring (Schrimpf et al., 2018) VGG19 across different lesioning and retraining iterations (Moore, Tuladhar, et al., 2023; Moore, Wilms, et al., 2023).

A key limitation of these works is that they introduce global damage to the network. Furthermore, analysis of neurodegeneration in the ventral stream using tasks outside of simple natural image recognition like CIFAR10 has not yet occurred. In this paper, we go a step further than all of the above previous works by asking ourselves how ANNs behave under localized synaptic degeneration of V1, V2, V4, and IT on tasks that ophthalmologists would usually use to detect visual agnosia, and what role perceptual learning plays in the process.

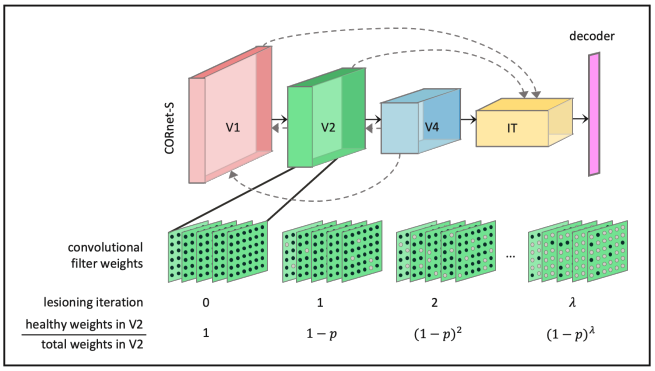

Figure 1: Schematic for modeling synaptic degeneration

Methods

We take CORnet-S (Kubilius et al., 2019) as our artificial model of the brain. We then individually, across separate experiments, progressively zero-out $p = 20\%$ convolution filter weights of V1, V2, V4, and IT for λ lesioning iterations (Fig. 1). We train a linear probe using a categorical cross entropy loss on output from the model's penultimate layer to evaluate performance on: (1) Labeled Faces in the Wild (LFW) (Huang, Ramesh, Berg, & Learned-Miller, 2007) to perform face verification (pair matching); and two self-designed datasets (Fig. 2A)—(2) Pseudoisochromatic (PICo) MNIST, where each 224x224 image is assigned one of 12 background colors, evenly distributed across the color wheel, and a digit $\in \{0,...,9\}$ is superimposed onto the colored background, with the digit color derived by distorting the background color's RGB channels randomly; and (3) Noisy Operators, where one of 5 binary operators ($+$, $-$, $\times$, $/$, and $\%$) is superimposed in white on a black background, with anywhere between 0-50% pixels randomly inverted. To incorporate neural plasticity, after every lesioning iteration, healthy model synapses are retrained on some fraction of training images from ImageNet (Deng et al., 2009), while keeping all injured model weights frozen (no synaptic regeneration occurs).

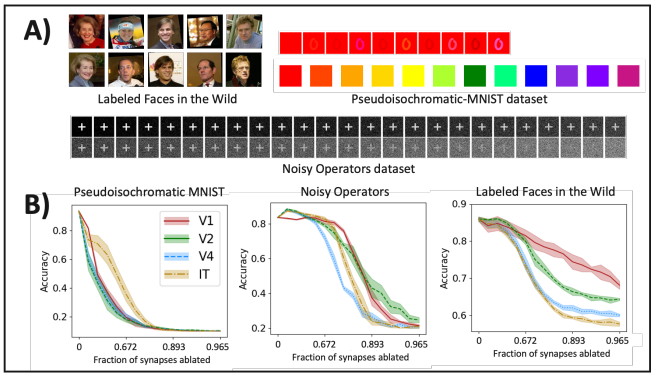

Figure 2: Localized Synaptic Degeneration without plasticity. A) Image datasets to test model performance on different tasks. B) Top-1 test accuracy on different tasks when V1, V2, V4, and IT are individually lesioned. Results over 5 runs of 5-fold CV.

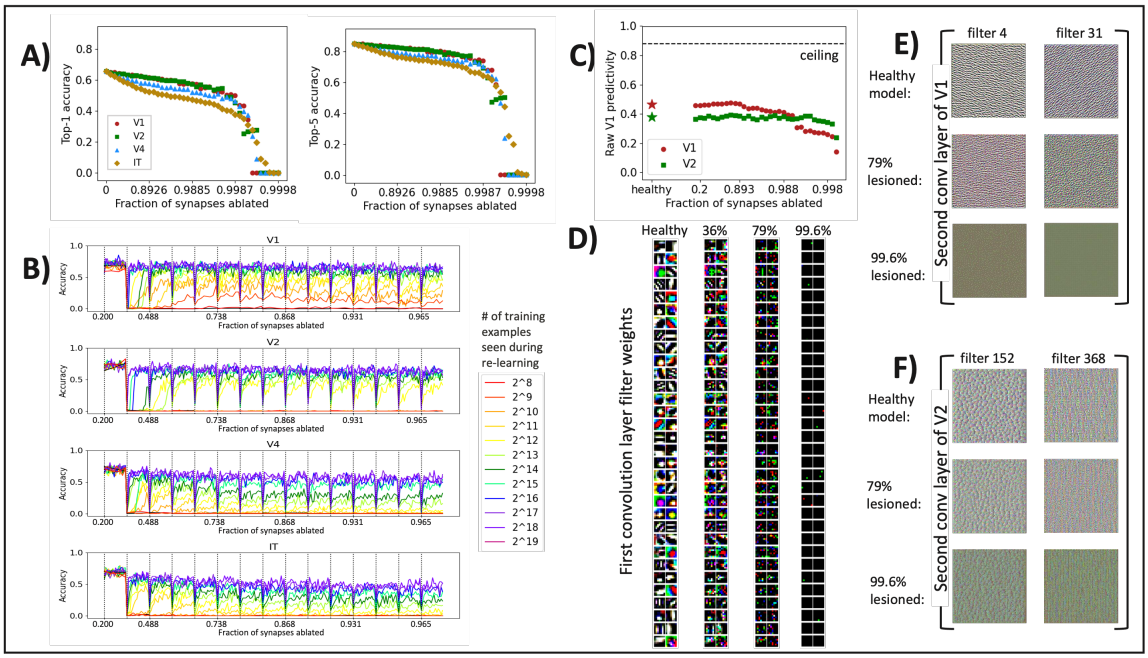

Figure 3: Incorporating neural plasticity. A) ImageNet top-1 and top-5 validation accuracy as functions of the fraction of synapses ablated in V1, V2, V4, and IT, with healthy synapses retrained on 2^{19} training images from ImageNet after every lesioning iteration. B) ImageNet training accuracy as functions of the number of training images seen by the model during re-learning. C) When V1 is lesioned, we measure how our model's second convolution layer in V1 and V2 predict V1 neural responses in the brain (Freeman et al., 2013). D) First convolution layer filter weights of V1 for healthy and V1-lesioned models. E, F) Features that most activate different filters in V1 and V2 when a model's V1 is lesioned. Loss function used is the mean of the output of a specific filter from a specific layer when random uniform noise is fed into the model.

Results

Experiment 1: Modeling localized synaptic degeneration without neural plasticity

With our initial attempt at modeling degeneration without neural plasticity, we find that different regions in the ventral stream have functionally different responses to different tasks. Fig. 2B shows sharp declines in performance for regions that have a more prominent functional role in a particular task. On PICo MNIST, the need to discriminate between the luminance of the background and foreground colors in images across contrast ratios while remaining color invariant activates visual areas in the occipital lobe (V1, V2, and V4) more than IT (Baker & Mareschal, 2001). For shape discrimination in the presence of noise, as required for the Noisy Operators dataset, V4 and IT are identified, plausibly for their crucial roles in complex feature extraction (Roe et al., 2012) and global shape detection. However, V1 and V2 quickly follow, highlighting the importance of integrating information from both low- and high-level features for shape understanding. On the LFW dataset, extracting facial features and recognizing faces are found to be functional roles of V4 and IT, which is consistent with the presence of the specialized fusiform face area in higher visual cortex (Kanwisher, McDermott, & Chun, 1997).

Experiment 2: Incorporating neural plasticity

Even when $\sim$99.85% of a certain region is degenerated, the model is able to find new information pathways through spared synaptic connections of V1, V2, V4, and IT and recover task performance to a great extent (Fig. 3A). This is possible due to continuous (perceptual) learning; it might have been really difficult to achieve if there was sudden abrupt injury to the model instead (Barbot et al., 2021). Furthermore, since a patient sees far fewer visual stimuli during recovery than what they see since birth, we experiment with the number of images we allow the model to train on during recovery. We find that the model recovers performance with $\sim 1/128^{th}$ fraction of the training images when V1 and V2 are lesioned, and with $\sim 1/32^{th}$ fraction when V4 and IT are lesioned (Fig. 3B).

We next analyze how plasticity-induced changes affect different layers in a V1-lesioned model. When scoring the model's V1 on neural data in the brain, we see a drop in V1 predictivity with lesioning that the model never recovers through re-learning (Fig. 3C). V2, however, maintains its V1 predictivity scores under lesioning, and overtakes V1 at $\sim$99% synaptic damage to V1. What happens to the re-learned filters in V2 when V1 filters stop looking like orientation-selective gabors under heavy lesions (Fig. 3D)? A healthy model's V1 and V2 seem to be responding best to edges and textures respectively (Fig. 3E-F). With lesioning, the receptive fields that V2 filters respond to start shrinking, with those textured features decomposing into more simpler edge-like structures, although not completely similar to what a healthy V1 prefers. It is perhaps the case that when orientation selective gabors visually disappear from V1, spared V1 synapses start transmitting stimuli information to V2, which now has the burden of extracting those low-level edge features that V1 failed to do.

References

Baker, C. L., & Mareschal, I. (2001). Processing of second-order stimuli in the visual cortex. *Progress in brain research*, *134*, 171–191.

Barbot, A., Das, A., Melnick, M. D., Cavanaugh, M. R., Merriam, E. P., Heeger, D. J., & Huxlin, K. R. (2021). Spared perilesional v1 activity underlies training-induced recovery of luminance detection sensitivity in cortically-blind patients. *Nature communications*, *12*(1), 6102.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 ieee conference on computer vision and pattern recognition* (pp. 248–255).

Finzi, D., Margalit, E., Kay, K., Yamins, D. L., & Grill-Spector, K. (2022). Topographic dcnns trained on a single self-supervised task capture the functional organization of cortex into visual processing streams. In *Svrhm 2022 workshop@ neurips*.

Freeman, J., Ziemba, C. M., Heeger, D. J., Simoncelli, E. P., & Movshon, J. A. (2013). A functional and perceptual signature of the second visual area in primates. *Nature neuroscience*, *16*(7), 974–981.

Han, S., Pool, J., Tran, J., & Dally, W. (2015). Learning both weights and connections for efficient neural network. *Advances in neural information processing systems*, *28*.

Huang, G. B., Ramesh, M., Berg, T., & Learned-Miller, E. (2007, October). *Labeled faces in the wild: A database for studying face recognition in unconstrained environments* (Tech. Rep. No. 07-49). University of Massachusetts, Amherst.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, *17*(11), 4302–4311. Retrieved from https://www.jneurosci.org/content/17/11/4302 doi: 10.1523/JNEUROSCI.17-11-04302.1997

Kubilius, J., Schrimpf, M., Kar, K., Rajalingham, R., Hong, H., Majaj, N., ... others (2019). Brain-like object recognition with high-performing shallow recurrent anns. *Advances in neural information processing systems*, *32*.

Li, H., Kadav, A., Durdanovic, I., Samet, H., & Graf, H. P. (2017). Pruning filters for efficient convnets. In *International conference on learning representations.* Retrieved from https://openreview.net/forum?id=rJqFGTslg

Maia da Silva, M. N., Millington, R. S., Bridge, H., & Plant, G. T. (2017). Visual dysfunction in posterior cortical atrophy. *Frontiers in neurology*, *8*, 268983.

Milner, A. D., Perrett, D. I., Johnston, R. S., Benson, P. J., Jordan, T. R., Heeley, D. W., ... others (1991). Perception and action in 'visual form agnosia'. *Brain*, *114*(1), 405–428.

Moore, J. A., Tuladhar, A., Ismail, Z., Mouches, P., Wilms, M., & Forkert, N. D. (2023). Dementia in convolutional neural networks: Using deep learning models to simulate neurodegeneration of the visual system. *Neuroinformatics*, *21*(1), 45–55.

Moore, J. A., Wilms, M., Gutierrez, A., Ismail, Z., Fakhar, K., Hadaeghi, F., ... Forkert, N. D. (2023). Simulation of neuroplasticity in a cnn-based in-silico model of neurodegeneration of the visual system. *Frontiers in Computational Neuroscience*, *17*.

Roe, A. W., Chelazzi, L., Connor, C. E., Conway, B. R., Fujita, I., Gallant, J. L., ... Vanduffel, W. (2012). Toward a unified theory of visual area v4. *Neuron*, *74*(1), 12–29.

Schrimpf, M., Kubilius, J., Hong, H., Majaj, N. J., Rajalingham, R., Issa, E. B., ... others (2018). Brain-score: Which artificial neural network for object recognition is most brain-like? *BioRxiv*, 407007.

Yamins, D. L., 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*(23), 8619–8624.