Explaining neural mechanisms of age-related dedifferentiation in the ventral stream through deep neural networks

Akilles Rechardt (akilles.rechardt.2024@live.rhul.ac.uk)
Department of Psychology, Royal Holloway University of London
Egham Hill, Egham, Surrey, TW20 0EX, UK

Gabriele Bellucci (gabriele.bellucci@rhul.ac.uk)
Department of Psychology, Royal Holloway University of London
Egham Hill, Egham, Surrey, TW20 0EX, UK

Robert M. Mok (rob.mok@nict.go.jp)
Center for Information and Neural Networks (CiNet), National Institute of Information and Communications Technology
1-4 Yamadaoka, Suita, Osaka 565-0871, Japan.

Abstract

Healthy older adults show less distinct visual category representations compared to young adults in later parts of the ventral visual stream (VVS), a phenomenon known as dedifferentiation. However, the neural mechanisms causing this are unclear. We used a deep convolutional neural network to model the VVS and applied noise and synaptic damage to different layers of the model while reading out category distinctiveness from a late, category-selective layer that models inferior temporal cortex (IT). We expected greater damage to IT to cause stronger dedifferentiation. As predicted, greater damage led to greater dedifferentiation. However, damage to *earlier* layers of the model (e.g., V1) caused greater dedifferentiation in IT compared to damaging later layers. This suggests that age-related dedifferentiation in IT could result from damage to upstream areas of the network. Our findings also match structural brain imaging work indicating early to late VVS white matter tract integrity is related to the distinctness of category representations. In sum, our modelling approach for the first time provides a mechanistic explanation for age-related dedifferentiation.

Keywords: Healthy aging; Dedifferentiation; DCNN; Neural damage; Modelling

Introduction

Dedifferentiation of visual representations in healthy older adults is consistently observed (Koen & Rugg, 2019). Dedifferentiation refers to reduced neuronal selectivity for categories, typically assessed via multivariate pattern analysis in category-selective brain areas in inferior temporal cortex (IT) by subtracting the average correlation of activation patterns within stimulus categories from those between each category.

But what are the neural mechanisms that underlie age-related dedifferentiation? One hypothesis is that age-related neurodegeneration of gray matter in IT leads to dedifferentiation. Indeed, brain volume in IT but not early visual cortex appeared to decrease with age (Raz et al., 2005). Another line of hypotheses concerns white matter (WM) degradation as a causal mechanism behind dedifferentiation. Studies have shown a negative relationship between measures of WM integrity connecting earlier and later visual areas in the VVS (inferior longitudinal fasciculus; ILF) and face dedifferentiation as well as facial processing (Bourbon-Teles et al., 2021; Rieck et al., 2020). This suggests that the integrity of connections between early and late visual areas might be essential for distinct category representations. A final hypothesis posits that increased neural noise leads to dedifferentiation (Li et al., 2001), supported by electroencephalography (EEG) findings showing that heightened noise is linked to face—but not scene—dedifferentiation (Pichot et al., 2022).

We investigated whether these mechanisms underlie age-related dedifferentiation using a deep convolutional neural network (DCNN) lesioning approach. We simulated age-related neurodegeneration in the ventral visual stream (VVS) by severing synapses or adding noise to DCNN weights. We then tested whether these manipulations matched brain-imaging data and examined how damage to different layers affected dedifferentiation. We hypothesized that both greater damage severity and damage to later, more category-selective layers would increase dedifferentiation. To preview the results, we found that greater damage indeed produced greater dedifferentiation, but unexpectedly, damaging layers corresponding to *earlier* visual areas had the strongest effect, suggesting that damage responsible for age-related dedifferentiation could be localized outside the category-selective VVS areas.

Methods

In this study, we employed CORnet-RT, a DCNN trained on the ImageNet classification dataset and designed with human VVS in mind. Its architecture approximates the hierarchical structure of the VVS, with processing stages intended to correspond to areas V1, V2, V4, and IT (Kubilius et al., 2018).

We quantified differentiation by adapting a multivariate pattern analysis approach, common in functional Magnetic Resonance Imaging (fMRI) studies. First, we passed images through the model and saved unit activations from the output of the IT block (each block contains 2 convolutional layers; Figure 1). Next, we generated a representational similarity matrix by cross-correlating (Pearson's r) these activations. From this matrix, we computed a "differentiation" metric by averaging within-category correlations (e.g., face-face) and subtracting the average cross-category correlations (e.g., house-face) (e.g., Koen & Rugg, 2019).

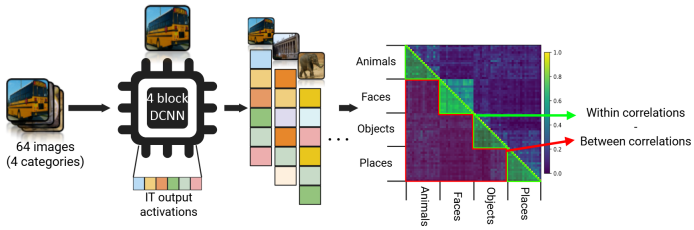

Figure 1: Procedure for computing the within-between metric to quantify differentiation.

To investigate possible mechanisms of age-related dedifferentiation, we proceeded with stimuli from an age-related dedifferentiation study (Haupt et al., 2024), enabling direct comparison between model and human results. We used 64 images spanning four categories (animals, faces, objects, places) and passed them through CORnet-RT. We repeated this process, introducing either Gaussian noise or synaptic damage (setting a random fraction of weights to zero) at varying levels to each block individually (V1, V2, V4, IT). Synaptic damage ranged from 0 to 100% of weights in increments of 5%, while Gaussian noise was drawn from a distribution whose standard deviation increased from 0 to 3x the original weight distributions in each layer, respectively, in increments of 0.1. We ran 200 simulations per condition, focusing on convolutional layer weights.

Results

Haupt et al. (2024) showed significant age-related dedifferentiation effects in most category contrasts on both EEG and fMRI (Figure 2A top for fMRI results). By comparing category-level differentiation in the intact model versus the damaged model, we replicated this dedifferentiation pattern (Figure 2A bottom). Notably, damage type (noise, synaptic damage) and location (brain region) consistently reproduced these dedifferentiation patterns, suggesting that this effect is apparent across different forms of neural damage.

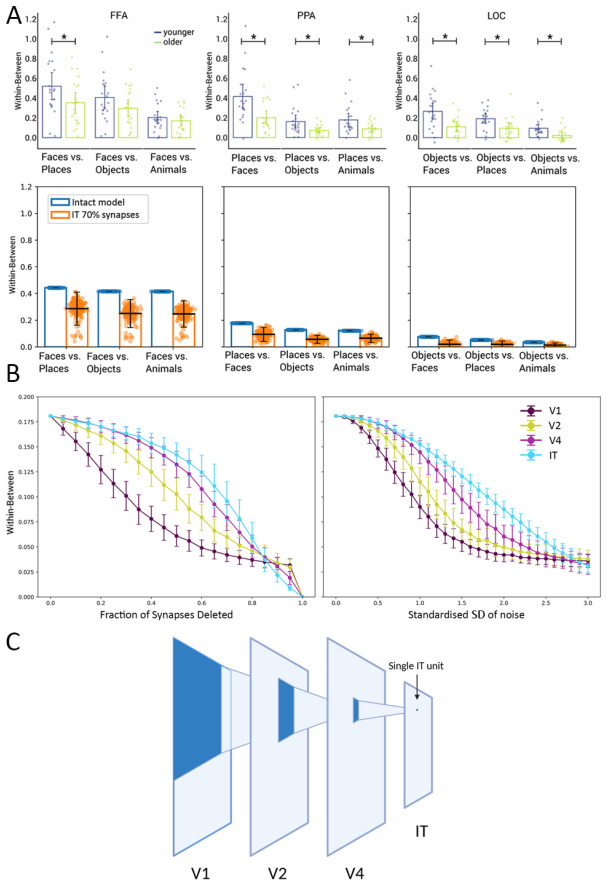

Figure 2: Results. A) Differentiation of human data (top; Haupt et al., 2024) and our modelling results (70% synapses damaged in IT; bottom). Error bars: 95% CI (top), 1.96 SD (bottom). B) Mean within-between differentiation in IT (y-axis) across damage severity (x-axis). Synaptic damage (left) and noise (right) across damaged layers (colours). Error bars: 1 SD. C) Illustration of an IT unit's receptive field across layers.

Next, we examined how damage type, severity, and location affected the model's overall category differentiation in IT (Figure 2B). To investigate this, we computed the mean within-between differentiation across all four categories from IT unit activations after applying noise and synaptic damage across different layers of the model. We expected greater levels of damage and damage to later layers to cause stronger dedifferentiation. Indeed, our results indicated that for both noise and synaptic damage, greater damage severity led to greater dedifferentiation. Surprisingly, damage introduced at *earlier* blocks (e.g., V1, V2) caused greater dedifferentiation to representations in IT than damage introduced at later blocks (e.g., V4 and IT). While age-related dedifferentiation is observed in category-selective areas in IT, our findings provide the intriguing alternate explanation that dedifferentiation may be due to damage elsewhere in the VVS.

Discussion

By using a DCNN as a model of the VVS, we demonstrated that greater damage, simulating age-related neurodegeneration, leads to greater dedifferentiation of category representations in IT, capturing brain imaging results. Furthermore, we found that damage to earlier model blocks caused greater dedifferentiation than damage to later blocks. This suggests that the mechanism behind age-related dedifferentiation in IT may be due to neurodegeneration of the WM tracts (ILF) that connect early visual regions to IT, consistent with structural imaging data (Rieck et al., 2020).

But what is the mechanism that underlies this surprising result? By considering how information propagates through the network in both the model and the VVS, we propose two explanations. For noise, perturbations introduced at earlier areas could be amplified through subsequent computations over multiple areas (i.e., multiple processing steps), accumulating greater distortions when reaching IT. This mirrors the broader challenge vision DCNNs face where untrained noise applied to input images can cause significant detriments to performance (Rodner et al., 2016). On the other hand, the effects of synaptic damage are likely due to the nature of the receptive fields caused by convolutional filters, akin to the VVS (Figure 2C). Specifically, individual units in the IT block in CORnet-RT rely on computations performed by around 2000 V1 weights. Thus, damaging only a subset of weights in V1 can distort a disproportionately large portion of IT outputs. By contrast, damaging IT itself results in a distortion more directly proportional to the fraction of weights affected. In sum, our DCNN lesion approach captured age-related dedifferentiation in IT, and showed that it can be explained by the consequences of neurodegeneration of the VVS' WM pathway, the ILF.

Acknowledgements

This work has been supported by ESRC funding (SEDarc DTP) to Akilles Rechardt.

References

Bourbon-Teles, J., Jorge, L., Canário, N., & Castelo-Branco, M. (2021). Structural impairments in hippocampal and occipitotemporal networks specifically contribute to decline in place and face category processing but not to other visual object categories in healthy aging. *Brain and Behavior*, *11*(8), e02127. doi: https://doi.org/https://doi.org/10.1002/brb3.2127

Haupt, M., Garrett, D. D., & Cichy, R. M. (2024, July). *Healthy aging delays and dedifferentiates high-level visual representations.* Neuroscience. doi: https://doi.org/10.1101/2024.07.30.605732

Koen, J. D., & Rugg, M. D. (2019, July). Neural Dedifferentiation in the Aging Brain. *Trends in Cognitive Sciences*, *23*(7), 547–559. doi: https://doi.org/10.1016/j.tics.2019.04.012

Kubilius, J., Schrimpf, M., Nayebi, A., Bear, D., Yamins, D. L. K., & DiCarlo, J. J. (2018, September). *CORnet: Modeling the Neural Mechanisms of Core Object Recognition.* Neuroscience. doi: https://doi.org/10.1101/408385

Li, S.-C., Lindenberger, U., & Sikström, S. (2001, November). Aging cognition: from neuromodulation to representation. *Trends in Cognitive Sciences*, *5*(11), 479–486. doi: https://doi.org/10.1016/S1364-6613(00)01769-1

Pichot, R. E., Henreckson, D. J., Foley, M., & Koen, J. D. (2022, November). *Neural noise is associated with age-related neural dedifferentiation.* Neuroscience. doi: https://doi.org/10.1101/2022.11.17.516990

Raz, N., Lindenberger, U., Rodrigue, K. M., Kennedy, K. M., Head, D., Williamson, A., … Acker, J. D. (2005, November). Regional Brain Changes in Aging Healthy Adults: General Trends, Individual Differences and Modifiers. *Cerebral Cortex*, *15*(11), 1676–1689. doi: https://doi.org/10.1093/cercor/bhi044

Rieck, J. R., Rodrigue, K. M., Park, D. C., & Kennedy, K. M. (2020, August). White Matter Microstructure Predicts Focal and Broad Functional Brain Dedifferentiation in Normal Aging. *Journal of Cognitive Neuroscience*, *32*(8), 1536–1549. doi: https://doi.org/10.1162/jocn_a_01562

Rodner, E., Simon, M., Fisher, R. B., & Denzler, J. (2016, October). *Fine-grained Recognition in the Noisy Wild: Sensitivity Analysis of Convolutional Neural Networks Approaches.* arXiv. (arXiv:1610.06756 [cs]) doi: https://doi.org/10.48550/arXiv.1610.06756