Reconstructing Feedback Representations in Ventral Visual Pathway with a Generative Adversarial Autoencoder

Haider Al-Tahan and Yalda Mohsenzadeh

Department of Computer Science, The University of Western Ontario, London, Ontario, Canada
Brain and Mind Institute, The University of Western Ontario, London, Ontario, Canada

Preprint

The University of Western Ontario

Introduction

Ventral visual cortex consists of a dense network of regions with **feedforward** and **feedback** connections. The feedforward path processes visual inputs along a hierarchy of cortical areas that starts in early visual cortex and ends in inferior temporal cortex . Alternatively, the feedback connections modulate neuronal responses in this hierarchy by broadcasting information from higher to lower areas.

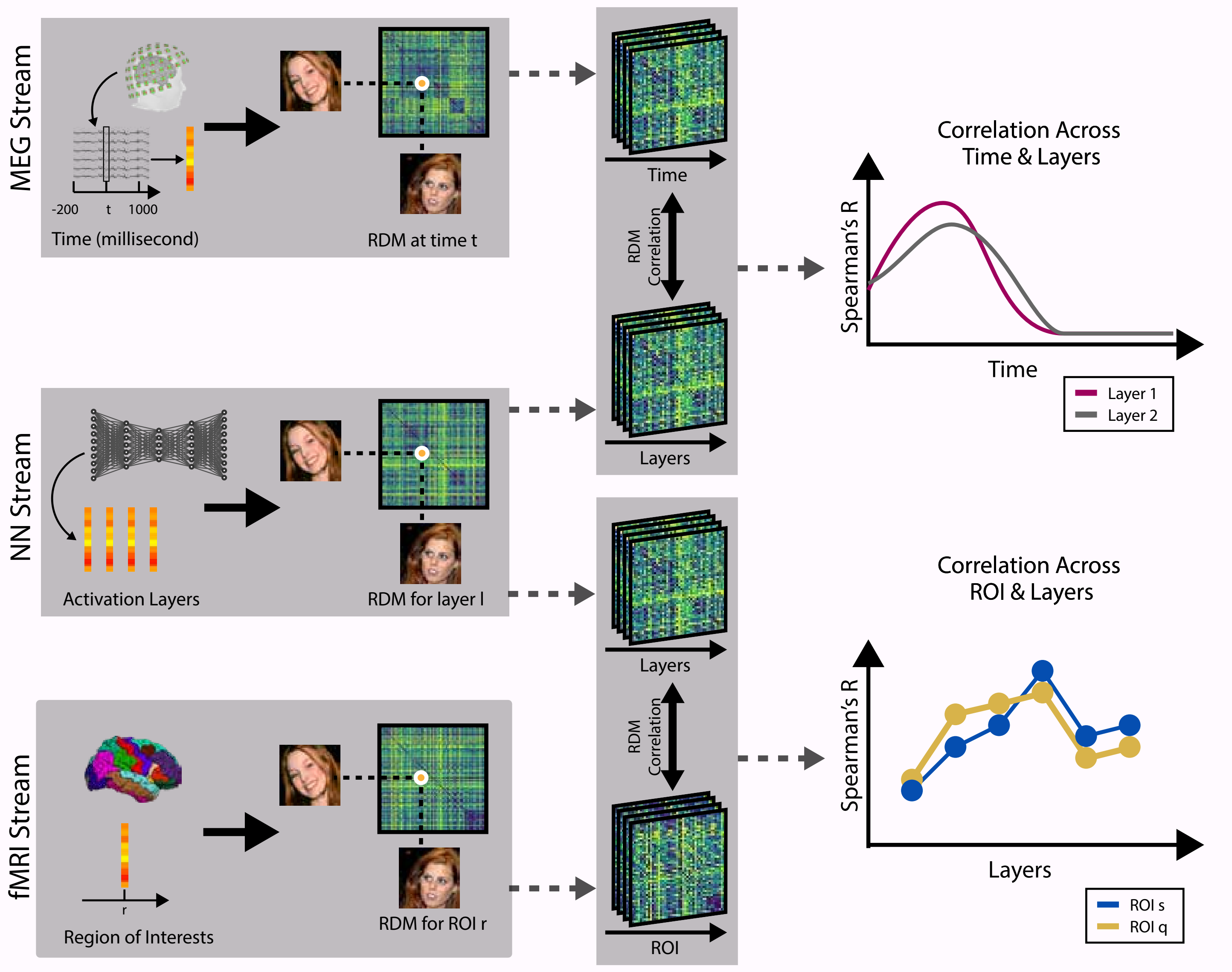
Previous work revealed that deep convolutional neural networks trained on image classification develop hierarchical representations similar to the cascade of processes along ventral visual pathway [1-2]. However, these type of networks miss the abundant number of feedback connections in visual cortex which are critical to resolving visual recognition in the brain [3].

In this project, we propose a new computational model that mimics the feedforward pathway in the ventral visual stream and revealed evidence of similarity with representations along the feedback path.

Datasets

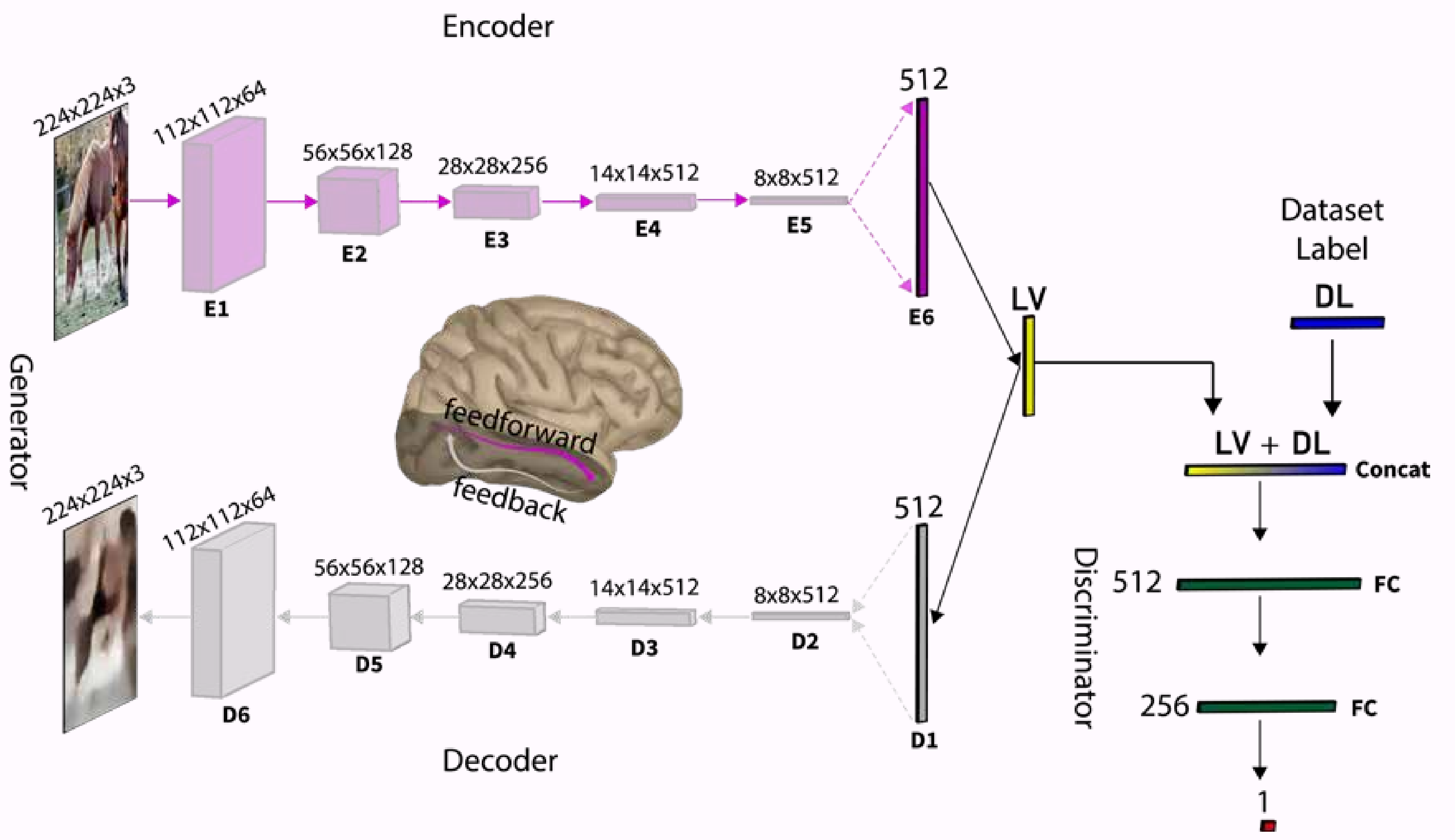
To train our proposed model we put together a super category (SC) data set consisting of 1,980,00 images from 4 equally distributed distinct categories: (i) Animals, (ii) Objects, (iii) Scenes, and (iv) Faces. Likewise, the stimulus set utilized to compute representational similarity between human brain and computational model consists of 156 natural images of the same four categories:

Data Analysis



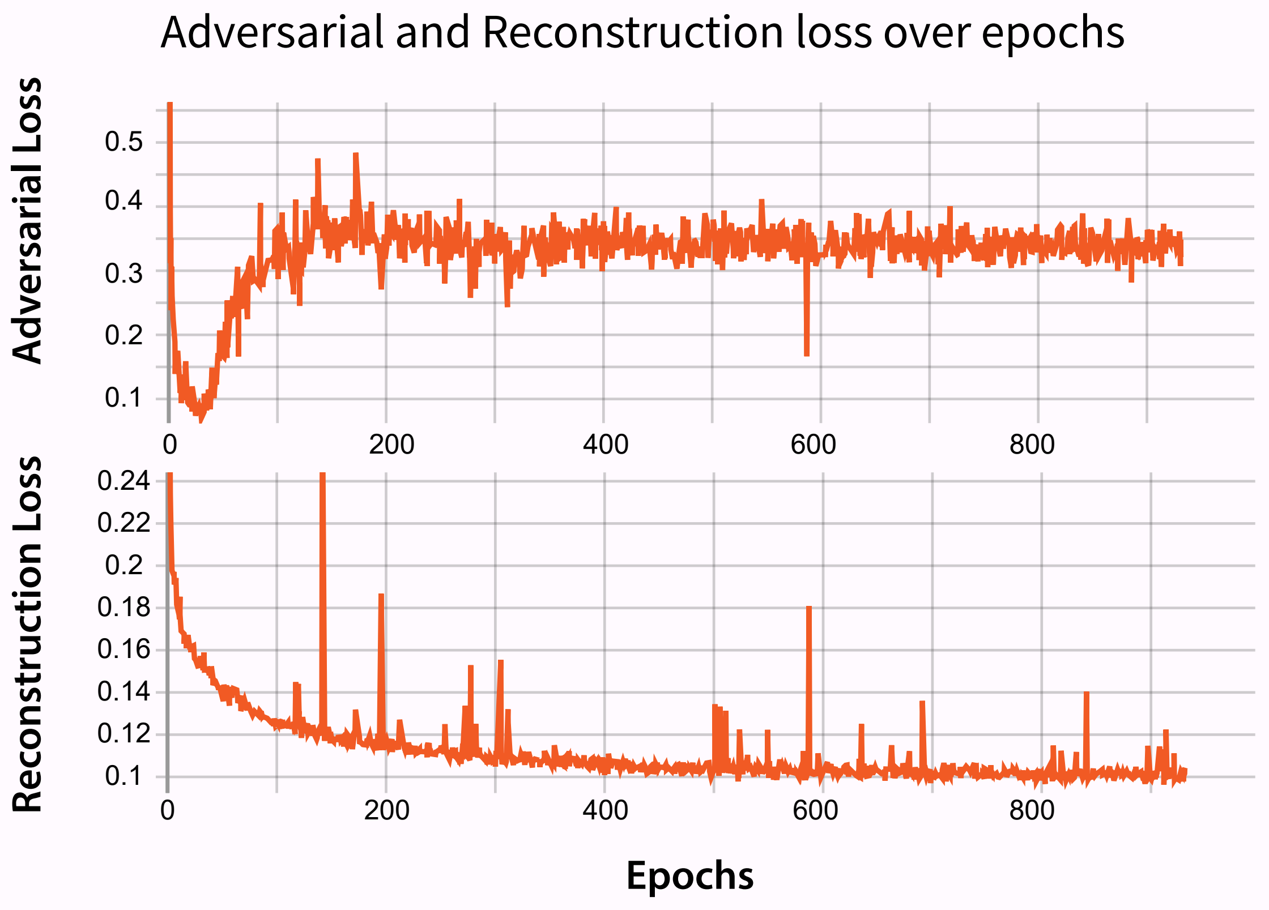
For each of the neuronal data, we first computed the representational dissimilarity matrices (RDMs). MEG RDMs were computed across time, computational model across cascading layers, and fMRI across Region of Interests (ROI). We then compared the spatiotemporal dynamics of neural representations in the human brain with our proposed model by computing the similarity (Spearman's R) of fMRI ROI RDMs and time-resolved MEG RDMs with our computational model RDMs. Correlations between subject-specific time resolved MEG RDMs and computational model layer RDMs result in a signal for each layer per participant across time. Correlations between subject-specific fMRI ROI RDMs and computational model layer RDMs result in subject-specific correlation values per layer. Then the correlation time series (for MEG/Model comparisons) or correlation values (for fMRI/Model comparisons) were averaged over participant and tested against zero for statistical significance.

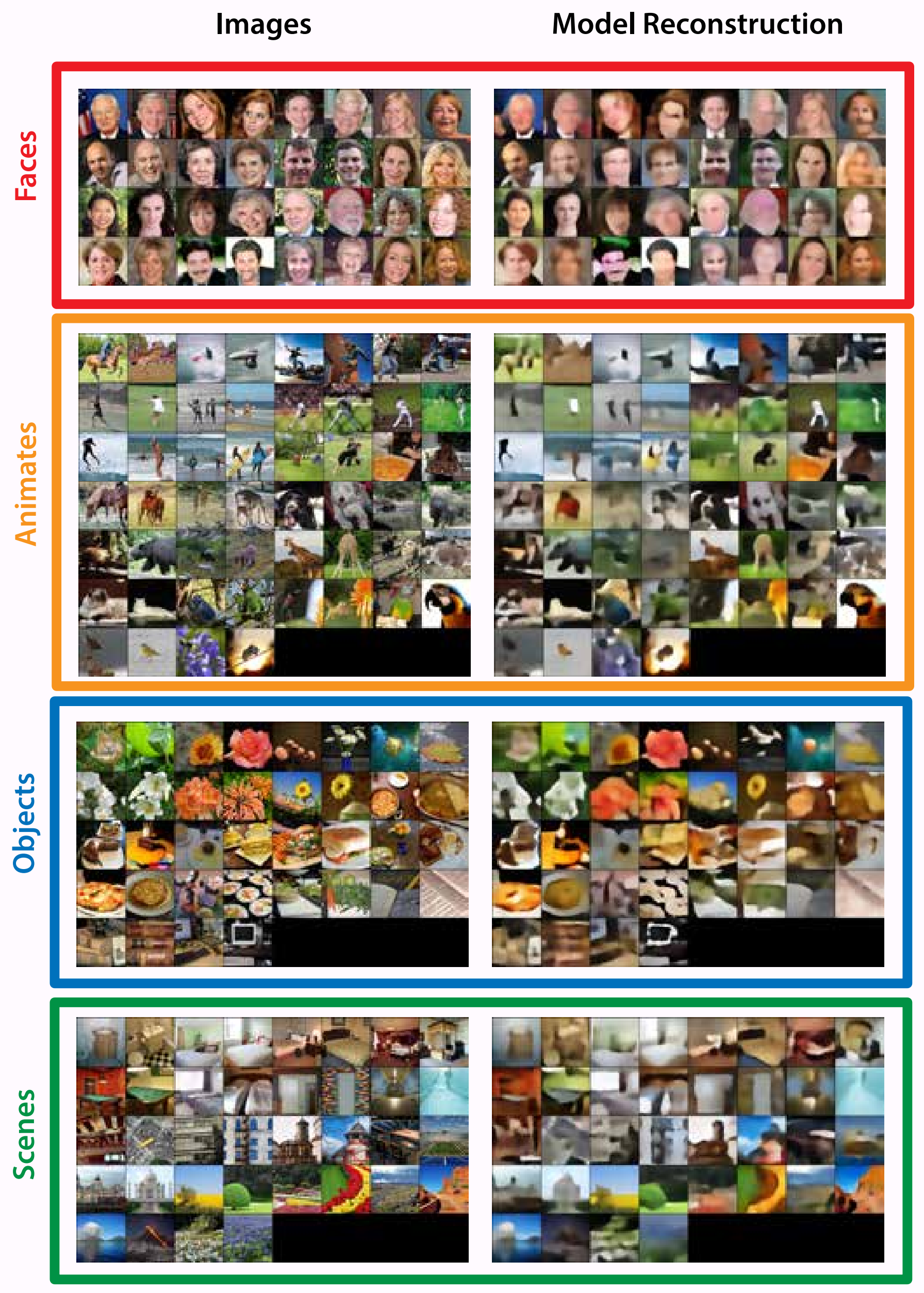
Computational Model



Our model is an adverserial autoencoder consisting of five convolutional blocks (E1-E5) and one fully connected layer (E6) in the encoder and one fully connected layer (D1) followed by five deconvolutional blocks in the decoder (D2-D6). The discriminator consists of two fully connected layers. **LV** denotes the latent vector generated by the encoder and **DL** is a one-hot data set label. Both vectors are concatenated and fed to the discriminator, while only the latent vector is fed to the decoder.

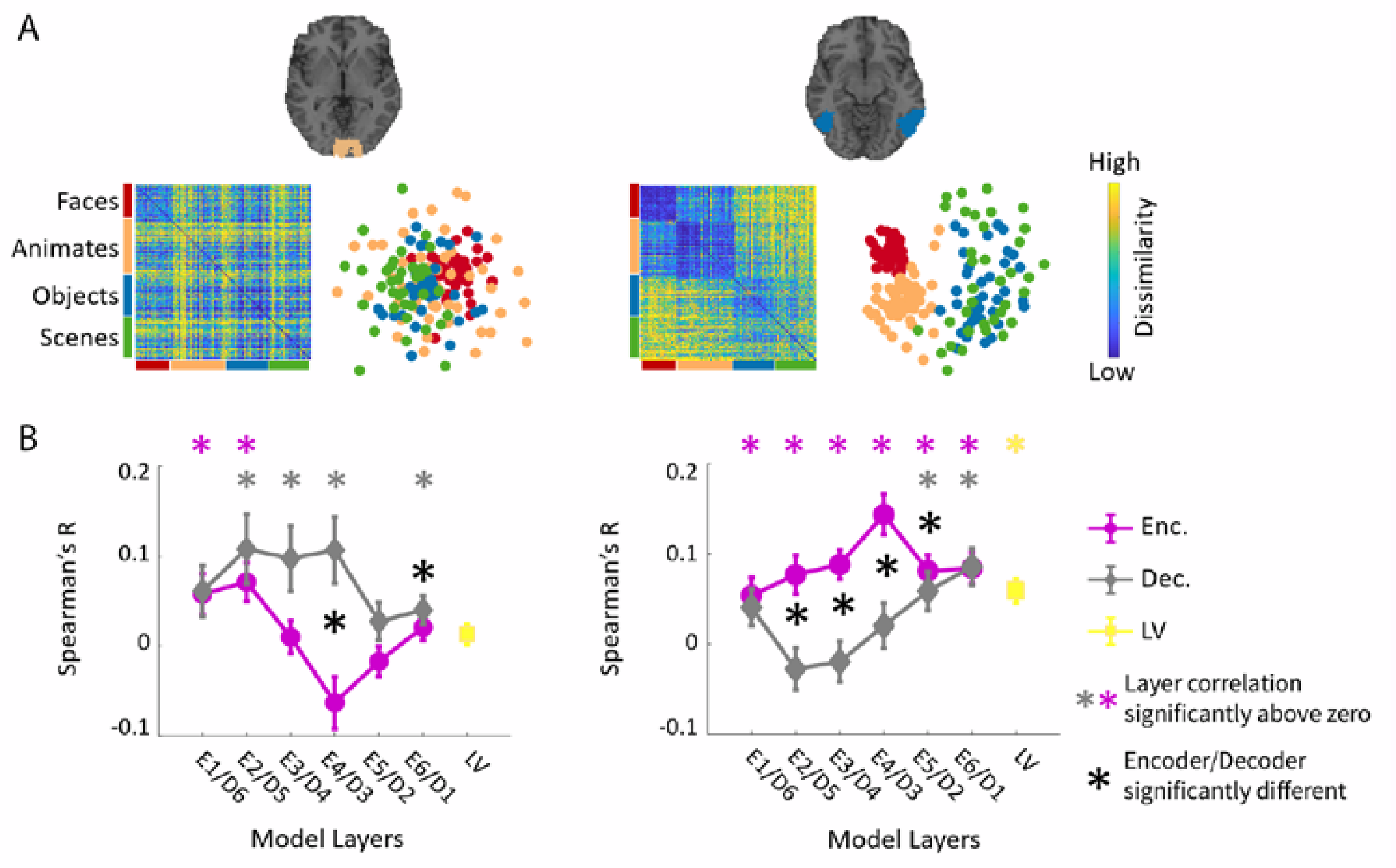
Computational Model Performance





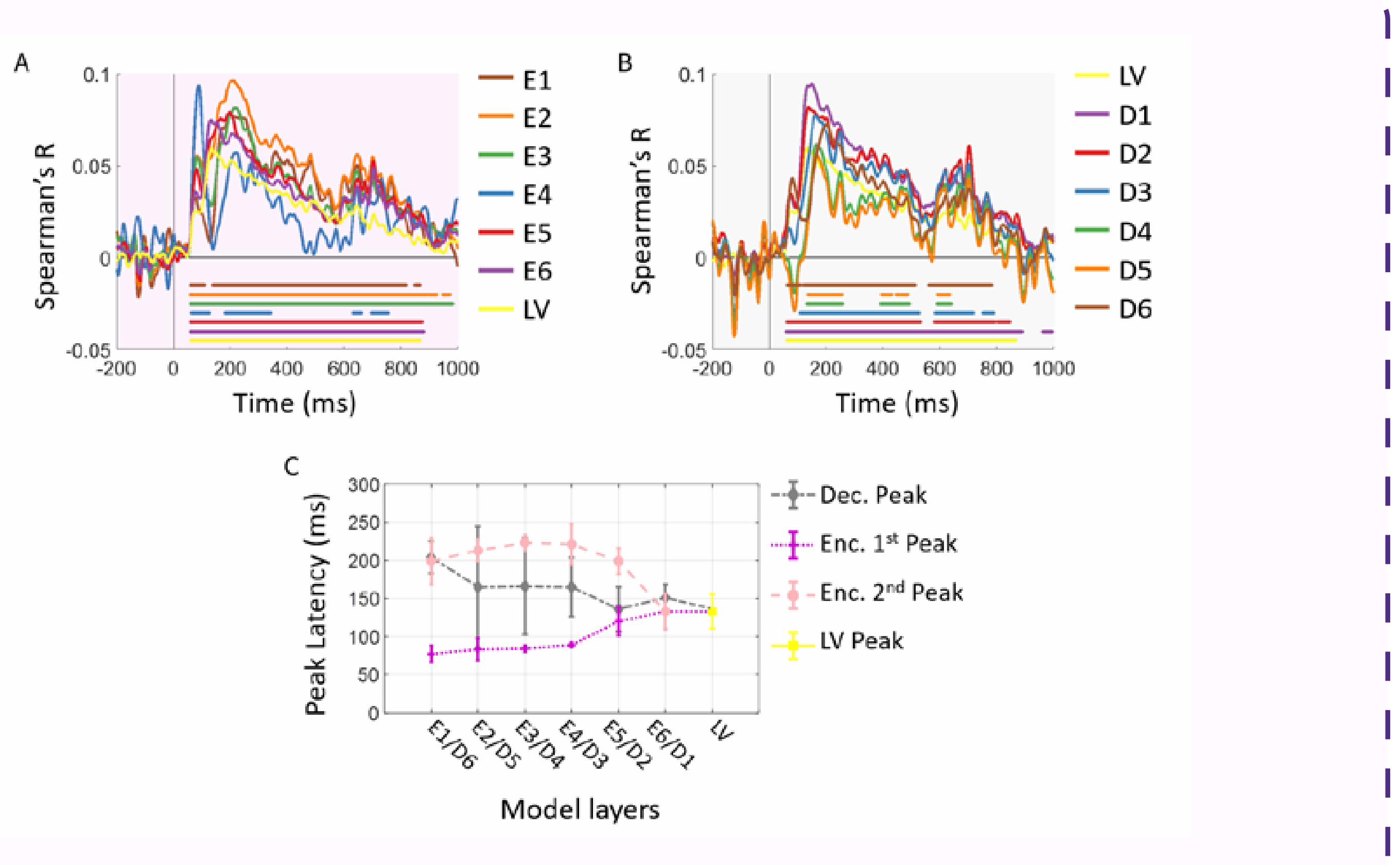
The model performed the reconstruction on the training dataset with 0.109 ± 0.006 mean absolute error (MAE) and on the 156 image set with 0.119 ± 0.003 MAE between the input image and the corresponding reconstructed image. These results show that the model not only have converged to an optimum but also generalizes well to test set.

Spatial Representational Comparisons



(A) Neural representations in early visual cortex (EVC) and inferior temporal area (IT). Subject-averaged RDM matrix (left), and its 2D multidimensionalscaling visualization (right). (B) Encoder, decoder and LV layer RDMs are correlated (Spearman's R) with subject-specific EVC (left) and IT (right) RDMs. The averaged correlations over subjects with standard error of the mean are depicted. N=15; two-sided ttests; false discovery rate corrected at P <0.05.

Temporal Representational Comparisons



(A-B) Encoder/Decoder and MEG representational comparison. We correlated the encoder layer RDMs with subject-specific time-resolved MEG RDMs resulting in fifteen correlation time courses. The color-coded lines below the curves show the time points when thecorrelations are significantly above zero (N=15; permutation tests; cluster definition-thresholdP <0.01; cluster threshold P <0.05). (C) Peak latency for encoder and decoder. The encoder have significantly lower peak latency across all layers (P= 0.014). Error bars are expressed in standard error of the mean.

Conclusion

Results of the representational similarities shows a spatio-temporal evidence that the brain visual processes demonstrate similarities with the reconstruction function implemented in the decoder sub-network.

References

1. Cichy RM, Khosla A, Pantazis D, Torralba A, Oliva A. Comparison of deepneural networks to spatio-temporal cortical dynamics of human visual objectrecognition reveals hierarchical correspondence. Scientific Reports. 2016;6(1):1–13.
2. Yamins D, Hong H, Cadieu C, Dicarlo JJ. Hierarchical Modular Optimization ofConvolutional Networks Achieves Representations Similar to Macaque IT andHuman Ventral Stream. NIPS. 2013
3. Mohsenzadeh Y, Qin S, Cichy RM, Pantazis D. Ultra-rapid serial visualpresentation reveals dynamics of feedforward and feedback processes in theventral visual pathway. eLife. 2018;7.

Acknowledgments

compute canada | calcul canada
Western BrainsCAN
Transforming brain research.