Linking human and artificial neural representations underlying face recognition: Insights from MEG and CNNs

H. Abdelhedi[1], K. Jerbi[1,2]
[1] Cognitive & Computational Neuroscience Lab, University of Montreal, Montreal, Canada;
[2] Department of Psychology, University of Montreal, Montreal, Canada;

Université de Montréal

Introduction

- Artificial Neural Networks performance have gone on to surpass human performance on many tasks.
- There is an increasing recognition of the added value of combining AI and neuroscience research to mutually reinforce one another.
- Several studies are focusing on **comparing** artificial and biological systems' internal representations in an array of cognitive tasks such as visual categorisation and on building **biologically plausible** models[9,12,14,16].
- **Faces are a special type of objects**. And the Face Recognition system in the brain is more complicated and recruits multiple regions.

Aim of the study:

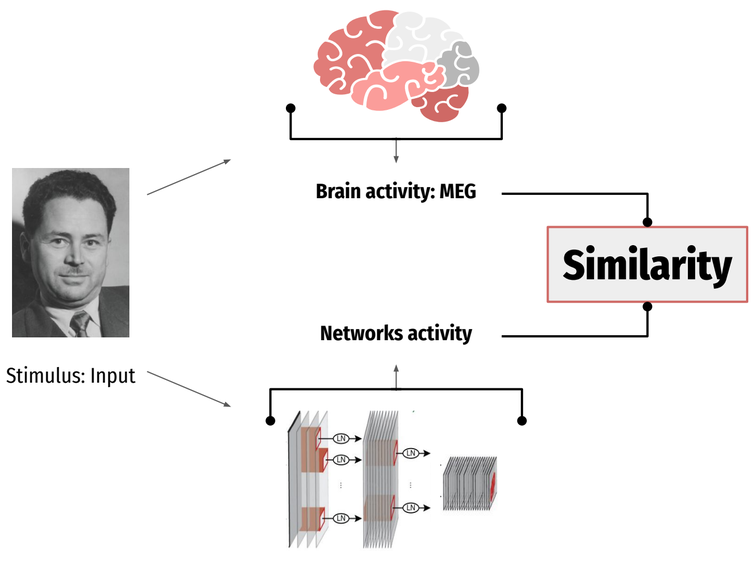

Figure 1. Study Outline

Compare the similarity of the internal representations of the brain and Convolutional Neural networks using:

- Neuromagnetic data
- Three different CNN architectures
- Representational Similarity Analysis

Methods: MEG data

- Used a distribution from the: "multi-subject, multi-modal human neuroimaging dataset"[15].
- MEG data acquired using an Elekta NeuromagVectorview 306 system.
- 16 subjects.
- **300 Facial stimuli**: unique identities (Famous/Unfamiliar), each passed twice: 600 trials.
- Preprocessing: We replicate the steps in the BioMAG study[5] (filtering, bad trials removal …).
- Selected **1 trial per condition (Face): 300 trial** per subject
- Epoched the data into segments of **800ms**

Methods: Networks Training and activations extraction

- We selected 3 Convolutional Neural Networks:
  - Backbone of **FaceNet[13]:** specifically built for face recognition
  - **ResNet50[4]:** Built for Object recognition
  - **CORnet-S[10]:** Designed to model the visual cortex
- Trained them on the same classification tasks:
  - Hyperparameteres selected after exhaustive search: Batch size: 32, Learning Rate: 0.01.
  - Loss: Cross-Entropy Loss.
  - Trained on VGGFace[1] for 30 epochs
  - Fine-tuned of a Distribution of the the CelebA[11] dataset (similar to the stimuli used in MEG data)
- Extracting the activations:
  - Pass the facial Stimuli (300 pictures) used in the MEG experiment.
  - Get the response of each layer for the three networks.
  - For each network, concatenate layer activations to generate another network response.

Methods: Representational Similarity Analysis (RSA)

- We choose Representational Similarity Analysis[8] to assess the **similarity** between the activity patterns of the artificial and biological systems
- We computed the Representational Dissimilarity Matrices (RDMs): used to quantify pairwise dissimilarities between the activations to the stimuli

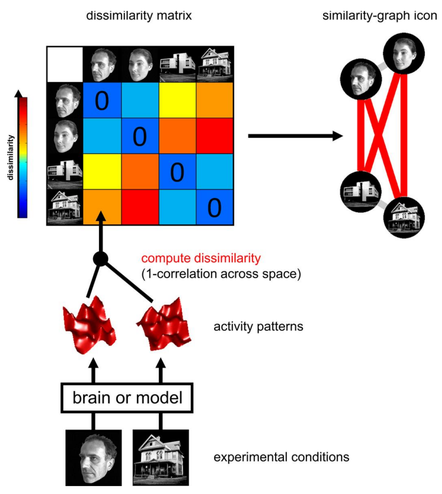

Figure 2. Compute RDMs

- We generated one RDM per layer in each network, and network-level RDM.
- FOR MEG data: We computed the RDM at each sensor then averaged them across all subjects.
- We computed the similarity between the RDMs in each sensor and the RDM of each layer of the three networks and for the whole network RDM.
- Correlations in the RDM cells but also across RDMs
- (similarity scores) were calculated using Pearson correlation.

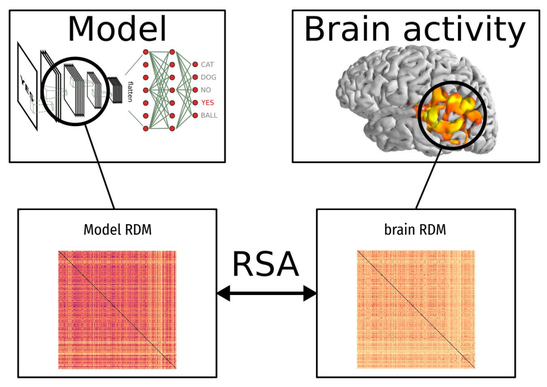

Figure 3. Computing similarity between biological and artificial systems

Results 1

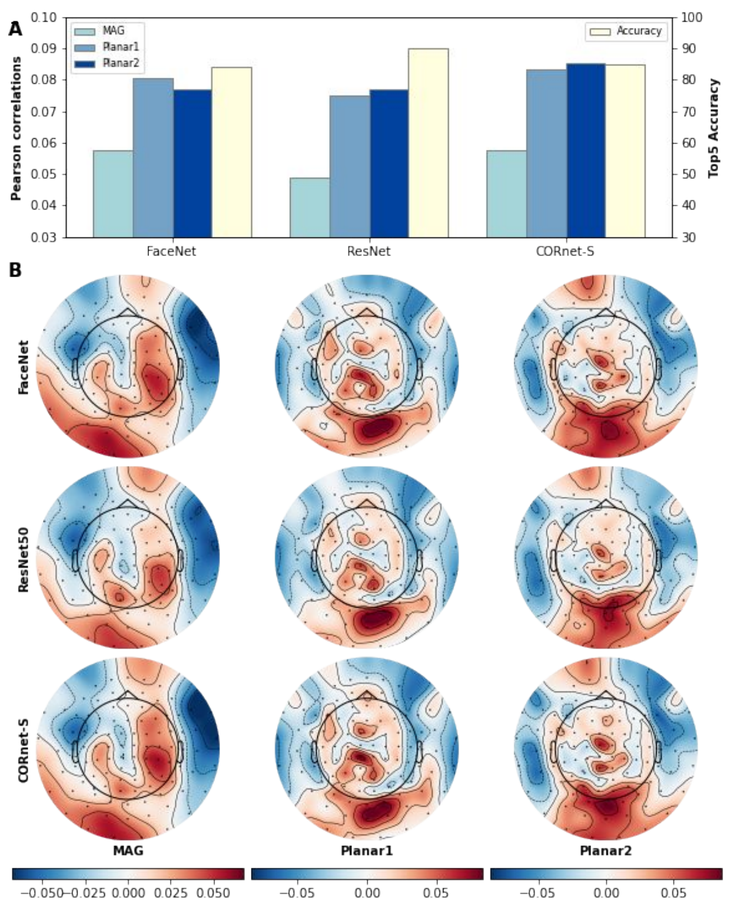

Figure 4.
(A) Maximum similarity scores obtained across all MEG sensors for each architecture activations.
(B) ANN performances on CelebA face stimuli.
(C) Topographical maps of the MEG-Model similarity scores for each architecture and MEG sensor type (i.e. single RDM per architecture).

Results 1

Summary of results 1:

- The three networks provided similar global correlation levels with the MEG
- correlations levels are moderate, they are comparable to previous findings[6]
- MEG gradiometer channels yielded higher correlations than the magnetometers
- Resnet50 has the best decoding accuracy.
- Topographical maps of the MEG-Model similarity scores suggest a prominent role of channels over occipital and, to some extent, central areas.

Results 2

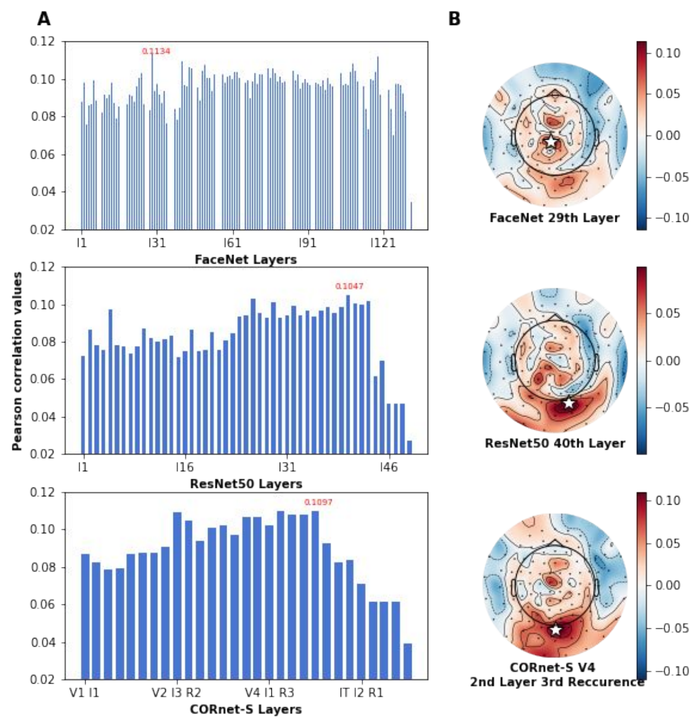

Figure 5.
(A) Networks layers' maximum correlations with the MEG (planar gradiometer) sensors.
(B) Topographical maps of the correlation values obtained with the specific layer that yields the highest similarity score. White stars indicate MEG sensors with the highest correlation values.

Summary of results 2:

- Similarity between the MEG activity and the last few layers at the end of the three architectures, especially for ResNet50 and CORnet-S.

Conclusions

- Success of the 3 networks in capturing neuromagnetic signatures of the brain dynamics
- Limited success: correlations are very moderate
- Levels of correlations are aligned with those reported in a similar study that used fMRI data.

Feature Work

- Further investigate the robustness of our observations including careful considerations of potential caveats and known limitations of the RSA approach[3].
- Explore the added value of source space MEG data analysis.
- Examine distinct frequency bands of the MEG signal.
- Test different other types of Artificial Neural Networks (GANs, Transformers ..)

Contact

Hamza Abdelhedi, CoCo Lab
Email: hamza.abdelhedi@umontreal.ca
Twitter: hamza_abdelhedi
LinkedIn: hamza-abdelhedi

Karim Jerbi, CoCo Lab
Email: karim.jerbi@umontreal.ca
Twitter: karimjerbineuro
LinkedIn: karimjerbi

Reference

1 Cao, Q., Shen, L., Xie, W., Parkhi, O. M., & Zisserman, A. (2017). Vggface2: A dataset for recognising faces across pose and age.
2 Chang, L., Egger, B., Vetter, T., & Tsao, D. Y. (2021). Explaining face representation in the primate brain using different computational models.
3 Dujmovi´c, M., Bowers, J. S., Adolfi, F., & Malhotra, G. (2022). The pitfalls of measuring representational similarity using representational similarity analysis.
4 He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep residual learning for image recognition.
5 Jas, M., Larson, E., Engemann, D. A., Lepp ¨akangas, J., Taulu, S., H ¨am ¨al ¨ainen, M., & Gramfort, A. (2018). A reproducible meg/eeg group study with the mne software: Recommendations, quality assessments, and good practices.
6 Jiahui, G., Feilong, M., Oleggio Castello, M. V. d., Nastase, S. A., Haxby, J. V., & Gobbini, M. I. (2022). Modeling naturalistic face processing in humans with deep convolutional neural networks.
7 Kietzmann, T. C., Spoerer, C. J., S ¨orensen, L. K. A., Cichy, R. M., Hauk, O., & Kriegeskorte, N. (2019). Recurrence is required to capture the representational dynamics of the human visual system.
8 Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience.
9 Kubilius, J., Schrimpf, M., Kar, K., Rajalingham, R., Hong, H., Majaj, N., . . . DiCarlo, J. J. (2019). Brain-like object recognition with high-performing shallow recurrent anns.
10 Kubilius, J., Schrimpf, M., Nayebi, A., Bear, D., Yamins, D. L. K., & DiCarlo, J. J. (2018). Cornet: Modeling the neural mechanisms of core object recognition.
11 Liu, Z., Luo, P., Wang, X., & Tang, X. (2015, December). Deep learning face attributes in the wild.
12 Richards, B. A., Lillicrap, T. P., Beaudoin, P., Bengio, Y., Bogacz, R., Christensen, A. J., . . . Kording, K. P. (2019). A deep learning framework for neuroscience.
13 Schroff, F., Kalenichenko, D., & Philbin, J. (2015, jun). FaceNet: A unified embedding for face recognition and clustering.
14 Storrs, K. R., Kietzmann, T. C., Walther, A., Mehrer, J., & Kriegeskorte, N. (2020). Diverse deep neural networks all predict human it well, after training and fitting.
15 Wakeman, D., & Henson, R. (2015, 01). A multi-subject, multi-modal human neuroimaging dataset.
16 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.