Aligning behavioral expressions of memory with convolutional neural network representations

Julian Gamez[1], Anisha S. Babu[1], Brice A. Kuhl[1] [1]Psychology Department University of Oregon

INTRODUCTION

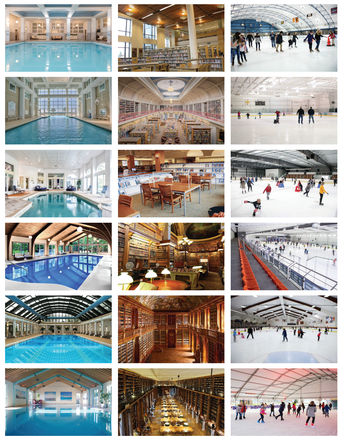
- Convolutional neural networks (CNNs) have become an increasingly popular tool in cognitive neuroscience [1] [2]
  - CNNs quantify information in complex visual stimuli
- However, it is not as well understood how or whether these models relate to memories for complex visual stimuli [3] [4]

Current Study: Using natural language processing (NLP), relate the representational structure of verbal recall of naturalistic scene images to the representational structure of image features extracted from different layers of a CNN (VGG16)

METHODS

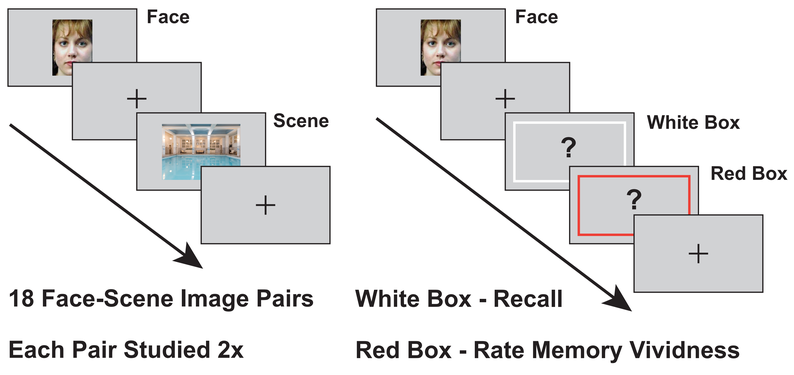
N=38

Indoor Scene Group



Outdoor Scene Group

Study Phase → Vividness Phase



18 Face-Scene Image Pairs

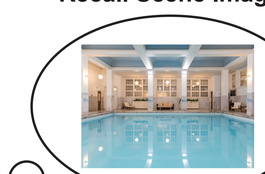
Each Pair Studied 2x

White Box - Recall

Red Box - Rate Memory Vividness

RESULTS - Free Recall Typed Responses

Recall Scene Image



Type Description

"An indoor pool design with a lot of square accents and beams. Blue and white colors."

Example Response

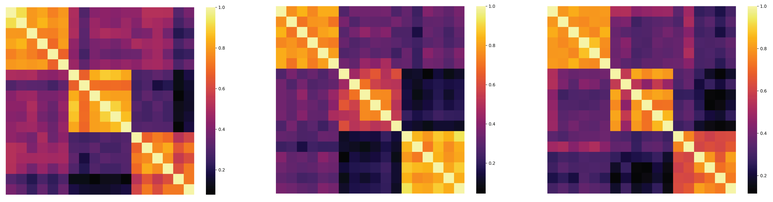
Create Text Embeddings Using Natural Language Processing

-0.06	0.73	0.34	-0.21	-0.59	...768

768 Length Feature Vector per Response

Shown Each Face Image

Create Similarity Matrices Using Response Vectors



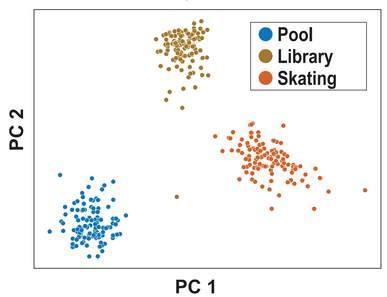
Subject 3 Matrix

Subject 5 Matrix

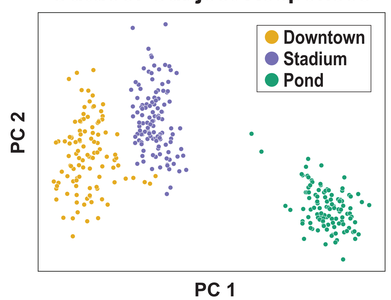
Subject 7 Matrix

Clustering Reveals Category Structure

Indoor Subject Responses



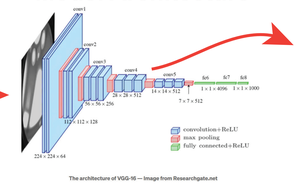
Outdoor Subject Responses



RESULTS - Extract Image Features (VGG16)

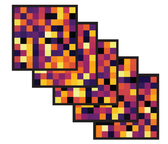
Input Each Scene Image

Model Architecture

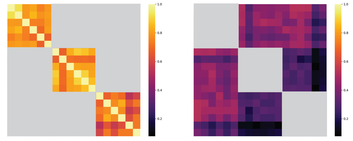


The architecture of VGG-16 — Image from Researchgate.net

Extract Pooling Layers' Features



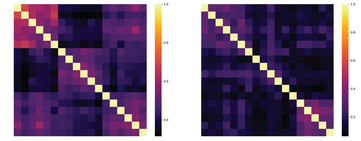
Recall Similarity Matrices



Within-Category Similarity

Across-Category Similarity

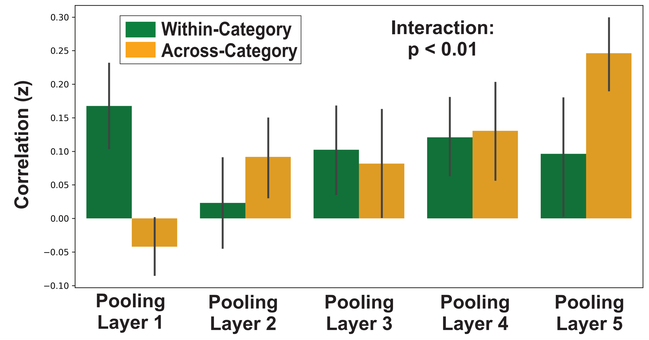
Layer Similarity Matrices



Pooling Layer 1

Pooling Layer 5

Correlation Between Recall and VGG16 Layers



CONCLUSION

- Here we found that CNN representations of scene images are correlated with representations expressed in verbal recall
- Within-category representational structure of verbal recall was more strongly correlated with the structure of early VGG16 layers
- Across-category representational structure of verbal recall was more strongly correlated with the structure of late VGG16 layers
- These results show that CNNs capture information expressed within verbal recall, but that different CNN layers capture different qualities of recalled memories

REFERENCES

[1] Storrs, K. R., Kietzmann, T. C., Walther, A., Mehrer, J., & Kriegeskorte, N. (2021). Journal of Cognitive Neuroscience, 33(10), 2044–2064. **[2]** Jha, A., Peterson, J. C., & Griffiths, T. L. (2023). Cognitive Science, 47(1), e13226. **[3]** Bainbridge, W. A., & Baker, C. I. (2022). Nature Communications, 13(1), 6508. **[4]** Morales-Torres, R., Wing, E. A., Deng, L., Davis, S. W., & Cabeza, R. (2024). Journal of Neuroscience, 44(21).

Funding: R01 #NS107727 and R01 #NS089727 to B.A.K