Memory diffusion: Using generative AI to create an image database for memory research

Fabian Kamp (kamp@mpib-berlin.mpg.de)

Research Group Adaptive Memory and Decision Making, Max Planck Institute for Human Development, Berlin, Germany

Max Planck School of Cognition, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

Frieda Josefine Born

Research Group Adaptive Memory and Decision Making, Max Planck Institute for Human Development, Berlin, Germany

Machine Learning Group, Technical University Berlin, Germany

BIFOLD, Berlin Institute for the Foundations of Learning and Data, Berlin, Germany

Bernhard Spitzer

Research Group Adaptive Memory and Decision Making, Max Planck Institute for Human Development, Berlin, Germany

Chair of Biopsychology, Technische Universität Dresden, Chemnitzer Straße 46, 01187, Dresden, Germany

Abstract

When we memorize visual stimuli, their content is processed at multiple levels, ranging from the fine-grained perceptual details to the semantic concepts and categories. However, it is unclear to which extend low- and high-level information is maintained in memory over time. Real-world stimuli are not ideal for investigating this question, as they often exhibit strong correlations between processing levels: Conceptually similar objects tend to share similar visual features. Using generative AI we created a new database of 496 image pairs orthogonalizing semantic (word2vec) and perceptual (CoreNet-S) information. Specifically, we generated image pairs that either (a) depict objects from distinct semantic concepts but are perceptually similar, or (b) show the same object but are perceptually dissimilar.

Keywords: generative AI, levels of processing, deep learning, memory

Introduction

The ventral visual stream is structured in a hierarchy where early visual areas respond to simpler features, while complex features are processed in more rostral areas (Cichy et al., 2016; Guclu & Van Gerven, 2015; Yamins et al., 2014). A similar topographical organization also exists during memory retention, as specialized cortical areas hold representations of distinct visual features (Christophel et al., 2017). However, the relative relevance of the different processing levels for later memory recollection and their maintenance over time is still an open question (Kwak & Curtis, 2022; Liu et al., 2020).

Importantly, different processing levels are not independent but rather vary together (Liu et al., 2021). Specifically, real-world stimuli which are semantically similar are also likely to show perceptual similarities. This is particularly problematic when investigating (working) memory, because of the relatively low number of trials per experiment in comparison to classical vision research. Together, these factors may cause a lack of power to detect differences with regards to the degradation low- and high-level information during memory maintenance.

We propose a new pipeline to artificially generate a set of stimuli dissociating perceptual (CORnet-S) and semantic (word2vec) information (Kubilius et al., 2018; Mikolov et al., 2013). The objective was to create image pairs that either (a) depict objects from distinct semantic categories but are perceptually similar, or (b) show the same object but are perceptually dissimilar.

The key innovation of the presented approach lies in the use of generative AI (Stable Diffusion XL) allowing us to substantially increase the perceived similarity between semantically unrelated images (Podell et al., 2023).

Procedure

Our pipeline has two main components. First, we sample image pairs from the THINGs database of object concepts (Hebart et al., 2019). Second, we use Stable Diffusion XL to generate new images from the original pair (Podell et al., 2023).

Sampling images from the THINGs database. We chose the THINGs database as it comprises images as well as the corresponding object concepts. This facilitated the comparison between the visual and semantic stimuli embeddings. We first computed the similarities of all image pairs in the database using CORnet-S and word2vec (Kubilius et al., 2018; Mikolov et al., 2013). We used only the IT layer of the CORnet-S model, as the similarity ratings from this layer resembles the perceived similarity as judged by humans the most. Similarities are calculated as Pearson correlations. To dissociate the perceptual and semantic dimensions, we then sampled pairs as from the database which scored high on one similarity dimension and low on the other.

Boosting the perceived similarity using generative AI. In the next step we set out to further increase the perceived similarity of the across concept image pairs. This step was necessary since it proved particularly difficult to maximize the perceptual similarity while minimizing the semantic relatedness using the sampling approach alone.

To artificially generate images, we use a text-guided image-to-image pipeline based on Stable Diffusion XL (Podell et al., 2023). For each of the

sampled perceptually matching pairs, we used the first image as visual input to stable diffusion and used the concept of the second image as text prompt. The newly generated image thus displayed an object from to the second concept class while guaranteeing a high perceptual similarity to the first image.

We also regenerated all the other images to exclude any biases due to visual features that are introduced by stable diffusion itself. Yet, for these images we used text prompts which were aligned with the image input.

Results

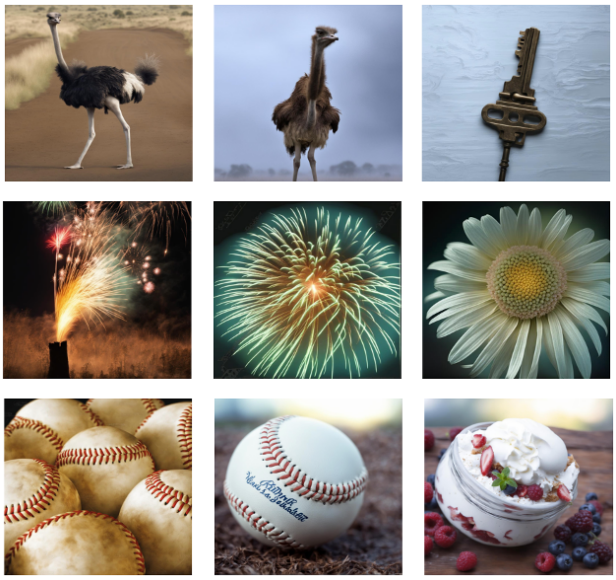

Figure 1: Three example image sets. (left-middle) High semantic and low perceptual similarity. (middle-right) Low semantic and high perceptual similarity.

On visual inspection the generated images fulfilled the desired criteria, i.e. pairs of images displaying semantically distant concepts showed a very high perceived similarity and vice versa.

We confirmed this intuition by computing the CORnet-S (IT layer) similarities for image pairs depicting objects from the same concept (ostrich-ostrich) and image pairs depicting objects from semantically distinct concepts (ostrich-key).

Our results showed that the perceptual similarity is substantially lower for within concept pairs than for across concept pairs ($t(494)=35.02$, $p<.001$). Thus, we successfully dissociated

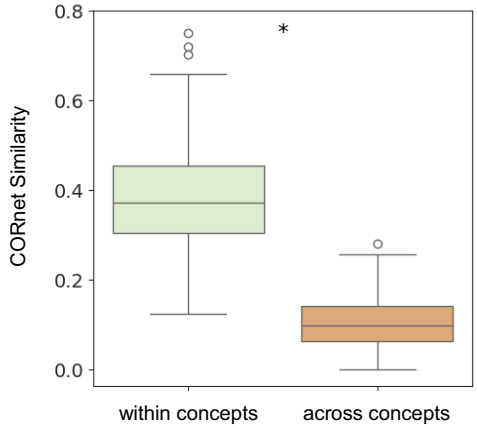

Figure 2: CORnet-S (IT layer) similarities for image pairs showing objects from the same or different concepts.

perceptual and semantic processing levels in the newly created database of 496 image pairs. Importantly, the presented pipeline is scalable, so that future work may further increase the total number of images.

We hope that the new stimuli set will prove useful for the investigation of memory processes and the maintenance of low- versus high-level information over time.

References

Christophel, T. B., Klink, P. C., Spitzer, B., Roelfsema, P. R., & Haynes, J.-D. (2017). The Distributed Nature of Working Memory. *Trends in Cognitive Sciences*, *21*(2), 111–124. https://doi.org/10.1016/j.tics.2016.12.007

Cichy, R. M., Khosla, A., Pantazis, D., Torralba, A., & Oliva, A. (2016). Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. *Scientific Reports*, *6*(1), Article 1. https://doi.org/10.1038/srep27755

Guclu, U., & Van Gerven, M. A. J. (2015). Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream. *Journal of Neuroscience*, *35*(27), 10005–10014. https://doi.org/10.1523/JNEUROSCI.5023-14.2015

Hebart, M. N., Dickter, A. H., Kidder, A., Kwok, W. Y., Corriveau, A., Wicklin, C. V., & Baker, C.

I. (2019). THINGS: A database of 1,854 object concepts and more than 26,000 naturalistic object images. *PLOS ONE*, *14*(10), e0223792. https://doi.org/10.1371/journal.pone.0223792

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* (p. 408385). bioRxiv. https://doi.org/10.1101/408385

Kwak, Y., & Curtis, C. E. (2022). Unveiling the abstract format of mnemonic representations. *Neuron*, *110*(11), 1822-1828.e5. https://doi.org/10.1016/j.neuron.2022.03.016

Liu, J., Zhang, H., Yu, T., Ni, D., Ren, L., Yang, Q., Lu, B., Wang, D., Heinen, R., Axmacher, N., & Xue, G. (2020). Stable maintenance of multiple representational formats in human visual short-term memory. *Proceedings of the National Academy of Sciences*, *117*(51), 32329–32339. https://doi.org/10.1073/pnas.2006752117

Liu, J., Zhang, H., Yu, T., Ren, L., Ni, D., Yang, Q., Lu, B., Zhang, L., Axmacher, N., & Xue, G. (2021). Transformative neural representations support long-term episodic memory. *Science Advances*, *7*(41), eabg9715. https://doi.org/10.1126/sciadv.abg9715

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient Estimation of Word Representations in Vector Space* (arXiv:1301.3781). arXiv. https://doi.org/10.48550/arXiv.1301.3781

Podell, D., English, Z., Lacey, K., Blattmann, A., Dockhorn, T., Müller, J., Penna, J., & Rombach, R. (2023). *SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis* (arXiv:2307.01952). arXiv. https://doi.org/10.48550/arXiv.2307.01952

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. *Proceedings of the National Academy of Sciences*, *111*(23), 8619–8624. https://doi.org/10.1073/pnas.1403112111