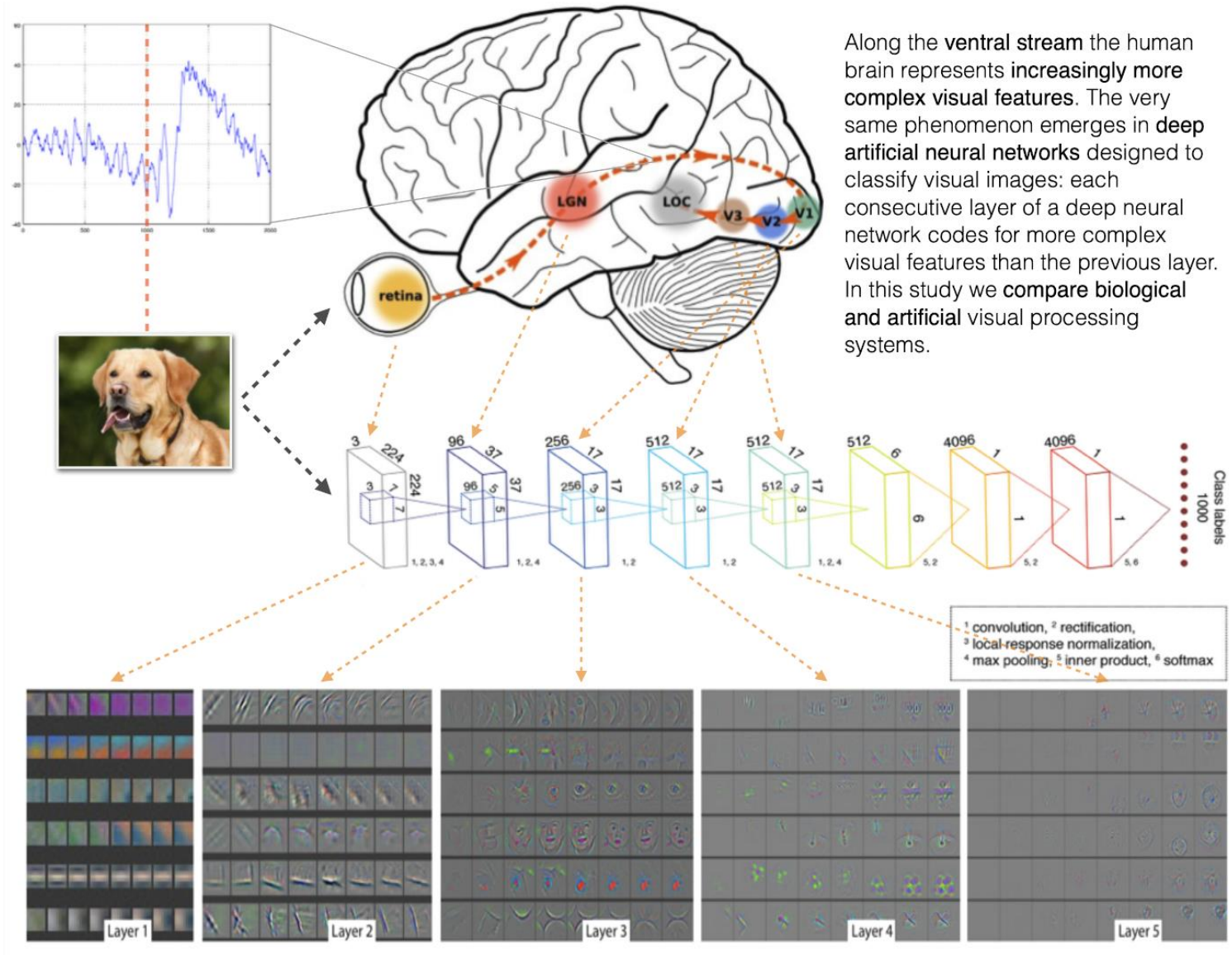
Topic #11

Can behaviour-trained ANNs reveal the brain’s temporal hierarchy in scene processing?

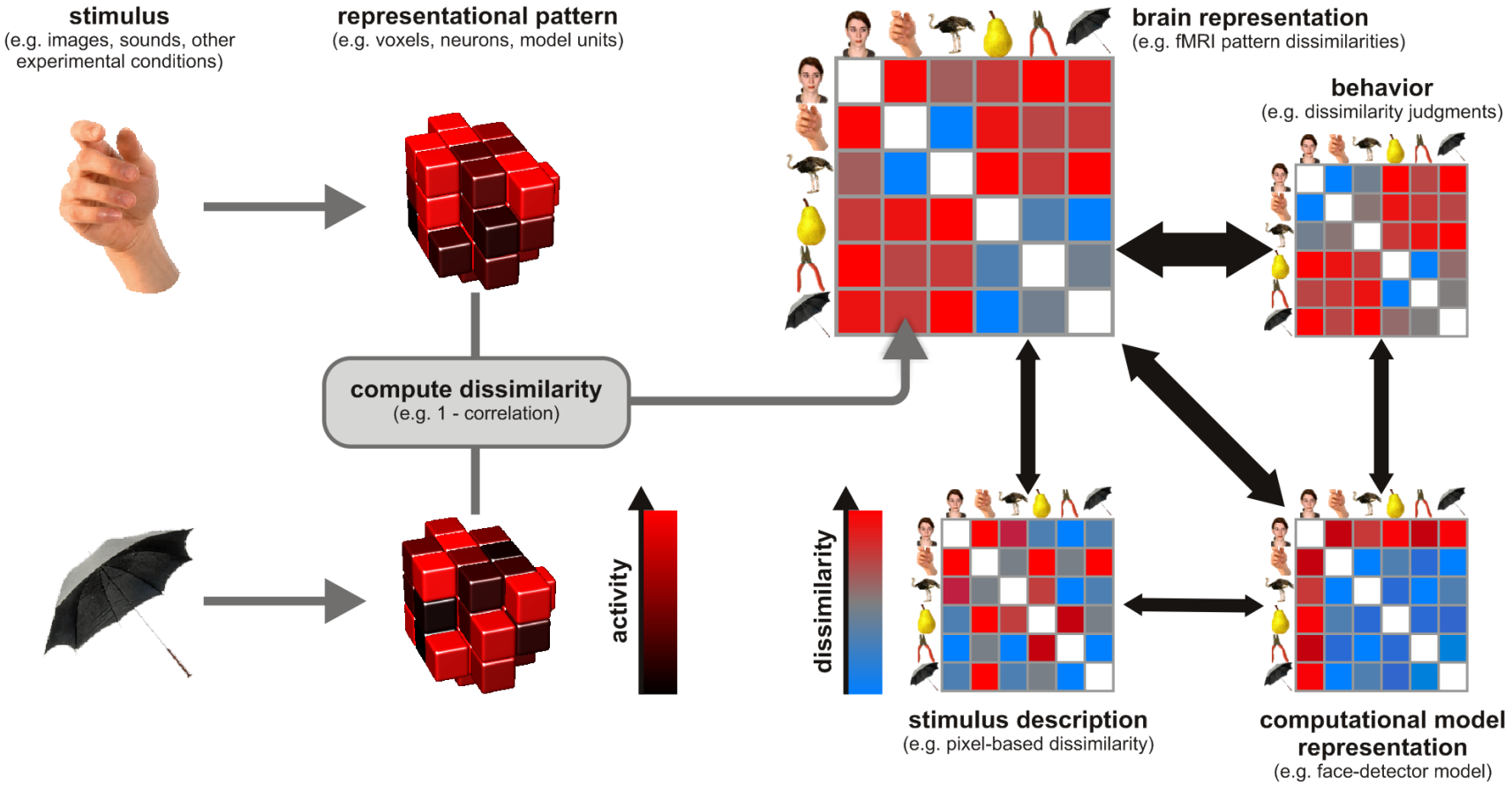
Linda Henriksson, 23.1.2026

Similarities in biological and artificial neural networks



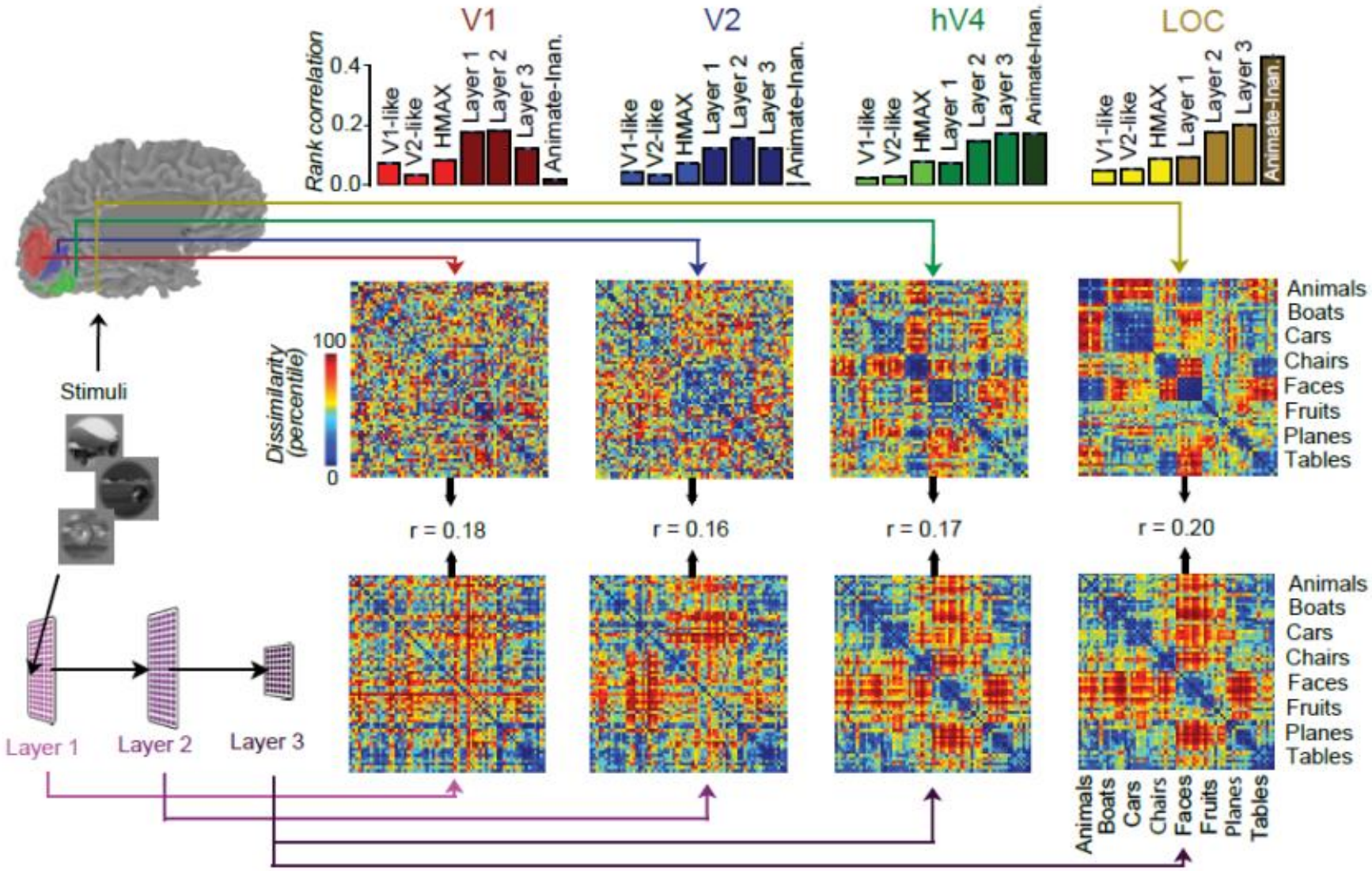
picture from: https://www.neurotechlab.ai/curious-similarities-between-ai-architectures-and-the-brain/

Representational similarity analysis (RSA)



Kriegeskorte, N., & Kievit, R. A. (2013). Representational geometry: integrating cognition, computation, and the brain. *Trends in cognitive sciences*, 17(8), 401-412.
See also: Kriegeskorte et al. (2008). *Frontiers in systems neuroscience*, 2, 249.

Representational similarity analysis (RSA)



picture from: https://nikokriegeskorte.org/tag/representational-similarity-analysis/, originally from Seiber et al 2016

Our brain imaging (MEG) experiment

500 ms **2000 ms**

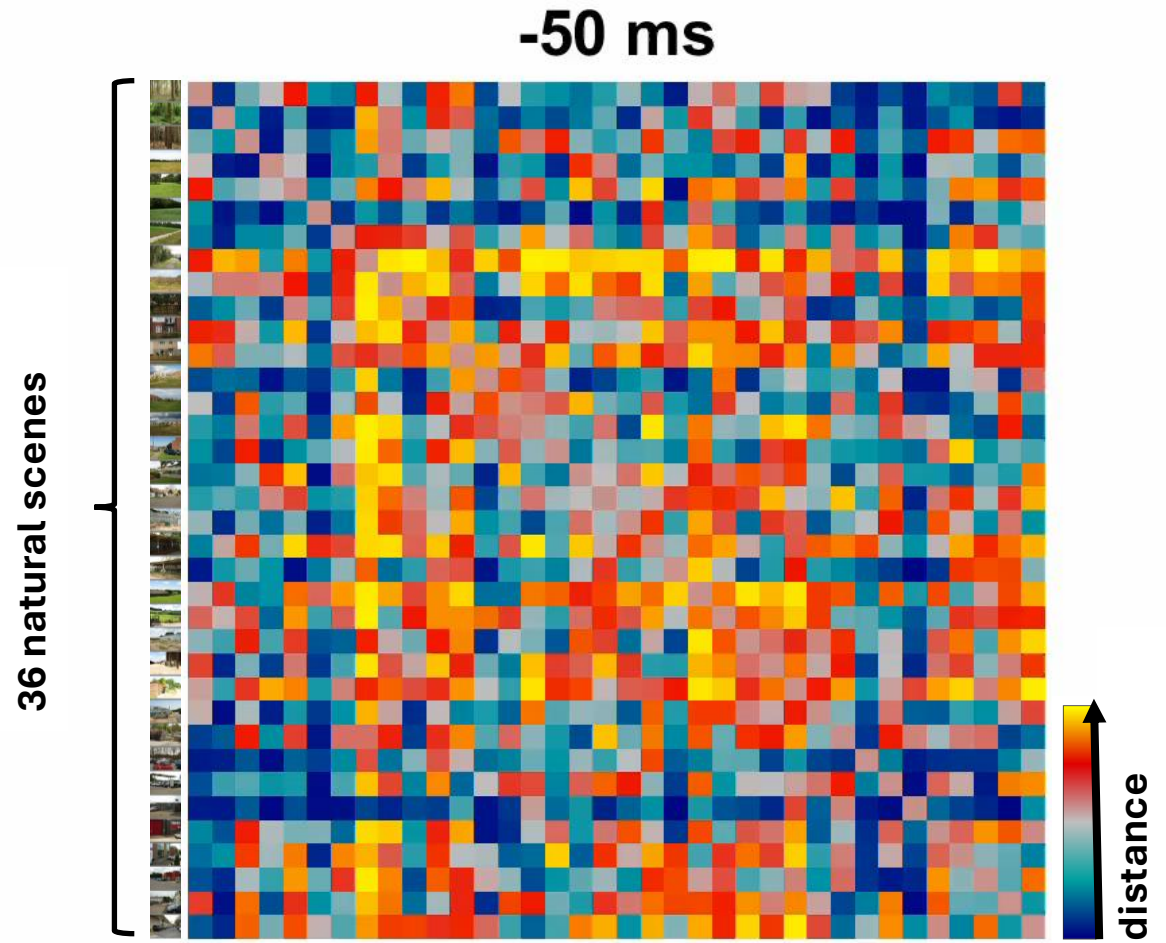
Were the two previous images from the same place?

Mononen, .., Henriksson. *J Neurosci, 2025*.

Scene images in our MEG experiment

Scene images from: Southampton-York natural scenes (SYNS; Adams et al., Scientific reports, 2016)

Mononen, .., Henriksson. *J Neurosci, 2025*.



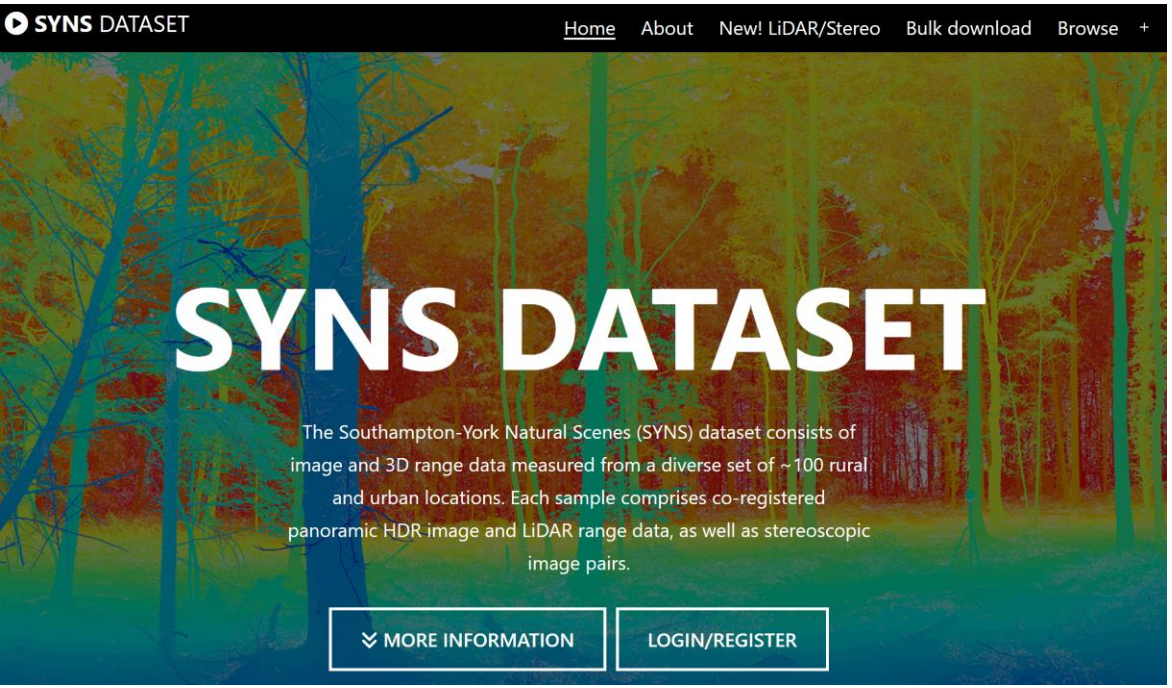
N = 20

Mononen, .., Henriksson. *J Neurosci, 2025*.

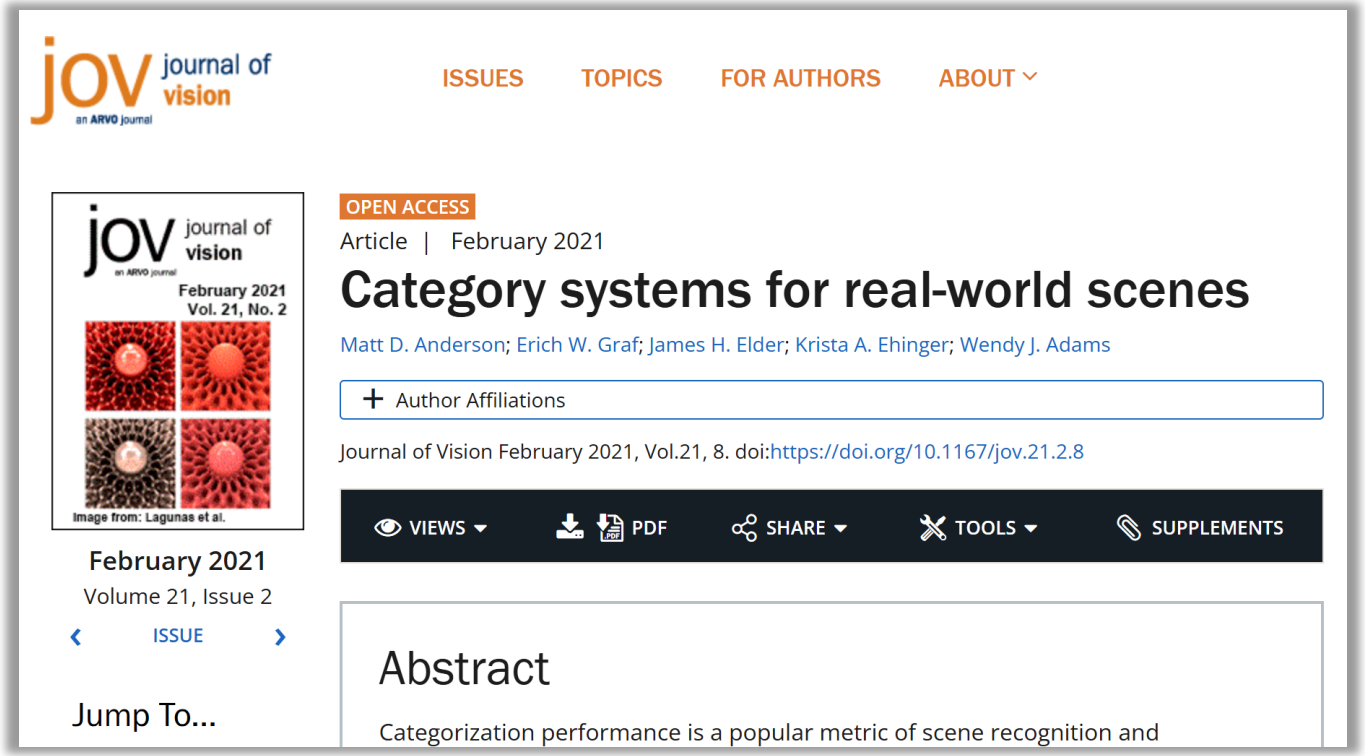
1) **Can we capture the emerging representation in the MEG data with representations in artificial neural networks (ANNs)?**

2) **Can we improve the correspondence with training the ANN model with behavioral labels?**

Scene image database & behavioral labels



https://syns.soton.ac.uk



https://eprints.soton.ac.uk/446699/
(see: Experiment 2)

Aalto-yliopisto
Aalto-universitetet
Aalto University

One option for the modeling: “ThingsVision”

“thingsvision is a Python package that let’s you easily extract image representations from many state-of-the-art neural networks for computer vision. In a nutshell, you feed thingsvision with a directory of images and tell it which neural network you are interested in. thingsvision will then give you the representation of the indicated neural network for each image so that you will end up with one feature vector per image. You can use these feature vectors for further analyses. We use the word features for short when we mean ‘image representation’.”

https://thingsvision.github.io/

Work division (suggestion) for the first steps

- Two of you could look into representational similarity analysis (**RSA**), *i.e.*, on how to compare brain and model representations.
- Two of you could look into the **scene database & behavioral labels**, *i.e.*, can you access the data and how could you use the behavioral data to train models.
- Two of you could look into the **ANN modeling**, *i.e.*, can you get a pretrained ANN running and extract scene features from different layers.

Final report

Do you want to produce a written standard report, or would you consider creating an extended PowerPoint presentation (+ code repo?) instead?

- As you work on the project, you would start collecting background, methods, analysis, results, and discussion directly into slides in a shared presentation.
- You could imagine, for example, that you are preparing a 45-minute detailed presentation on your topic.
- For each slide, focus on 1–2 main points and ask yourself why this information is relevant.
- You could re-use and refine this material for both the mid-term and final presentations.

A few key references / sources

MEG data: Mononen, et al. (2025). Cortical encoding of spatial structure and semantic content in 3D natural scenes. Journal of Neuroscience, 45(9). ← You will have access to the MEG-RDMs (let's look at the details in our next meeting)

Scene images: Adams, et al. (2016). The Southampton-York natural scenes (SYNS) dataset: Statistics of surface attitude. Scientific Reports, 6(1), 35805. https://syns.soton.ac.uk/ ← We used a subset in our MEG experiment

Behavioral labels for the scene images: Anderson, et al. (2021). Category systems for real-world scenes. Journal of Vision, 21(2), 8–8. https://eprints.soton.ac.uk/446699/ ← See Experiment 2

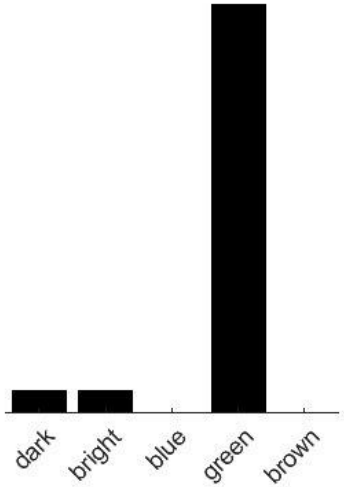
One of the first studies that combined MEG & ANNs: Cichy, et al. (2016). Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. Scientific Reports, 6(1), 27755. ← To see an example of similar research. There are many more similar (better?) examples since this original research.

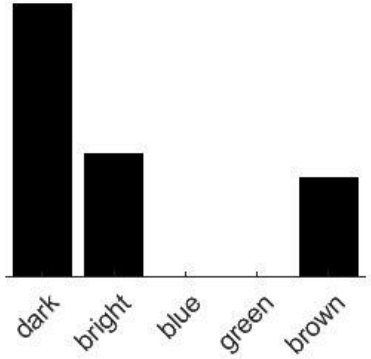
Human-generated category labels for

- **Semantic content**
- **Visual appearance**
- **Spatial structure**

(Anderson et al., JoV 2021)

Southampton-York natural scenes (SYNS)
(Adams et al., Scientific reports, 2016)

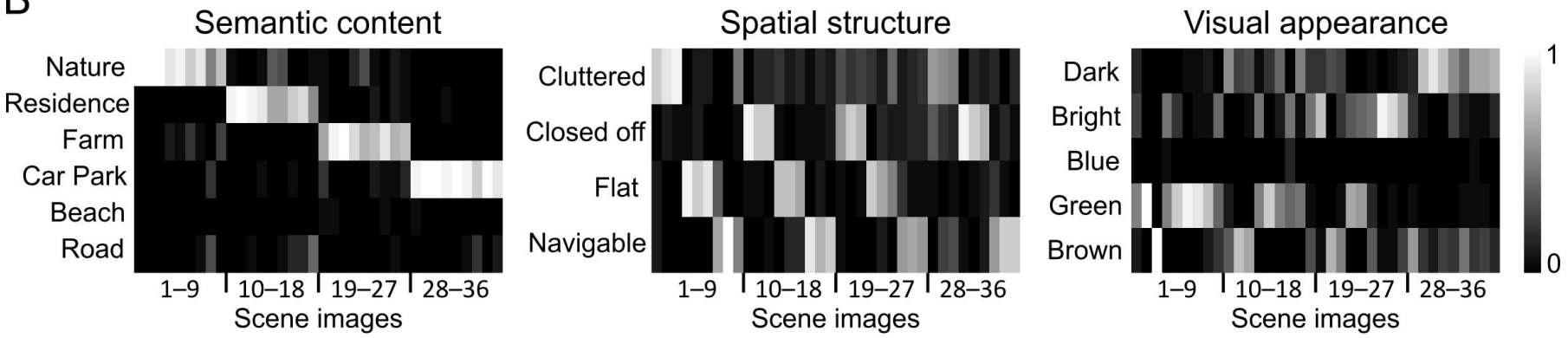




Visual appearance

A

B



Mononen, .., Henriksson. *J Neurosci, 2025*.