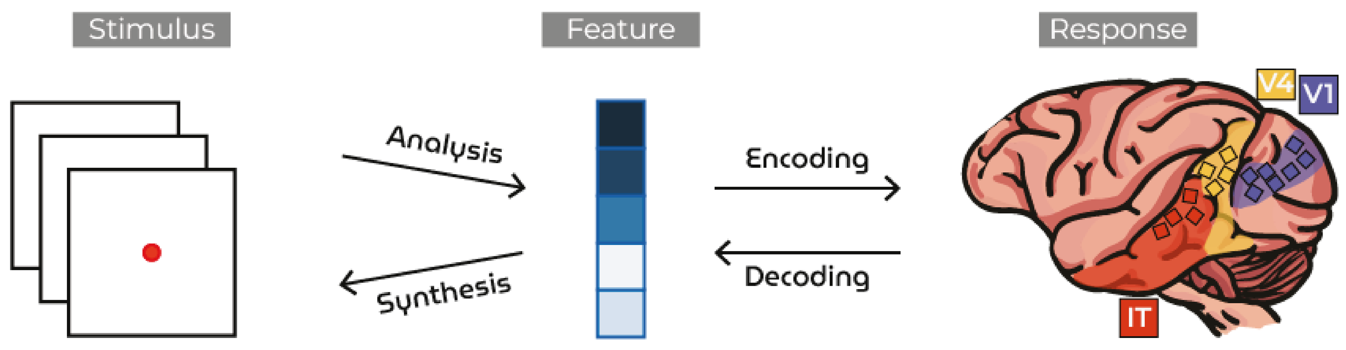
Feature-disentangled reconstruction of perception from multi-unit recordings

Thirza Dado[1], Paolo Papale[2], Antonio Lozano[2], Lynn Le[1], Feng Wang[2], Marcel van Gerven[1], Pieter Roelfsema[2], Yağmur Güçlütürk[1], Umut Güçlü[1]

[1]Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, Netherlands
[2]Netherlands Institute for Neuroscience, Amsterdam, Netherlands

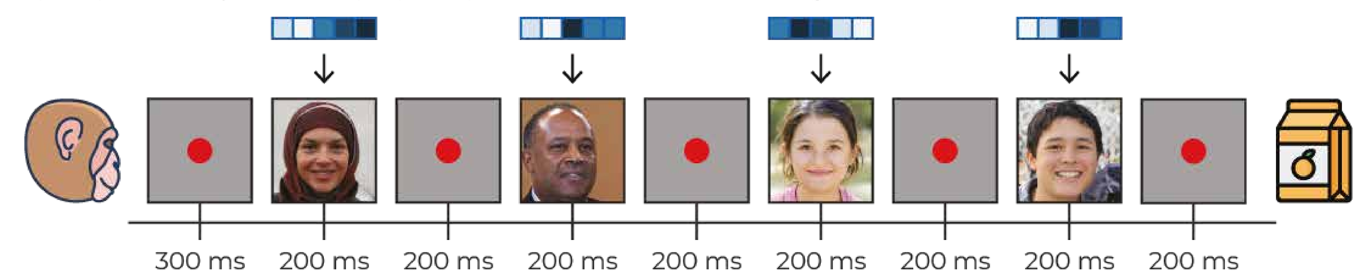
INTRODUCTION

- What high-level feature representations underlie visual perception?
  - We analyzed the relationship between **multi-unit activity (MUA)** [1] of macaque visual cortex and various **latent representations** of recent deep generative models with different properties, each of which captured a specific set of features and patterns
  - The feature-disentangled ω-latents outperformed the alternative representations (i.e., z- and CLIP-latents, and were subsequently used to *reconstruct* the perceived stimuli with state-of-the-art quality, according to the experimental paradigm of [2]



METHODS

- **Stimuli:** face- and natural images generated by StyleGAN3 [3] and StyleGAN-XL [4], respectiv.
- **Features:** conventional z-latents of StyleGAN 3/-XL, feature-disentangled ω-latents of StyleGAN 3/-XL, and language-contrastive CLIP-latents of Stable Diffusion
- **Responses:** MUA with 15 chronically-implanted 64-channel microelectrode arrays in one macaque (male, 7 years old) upon presentation with images



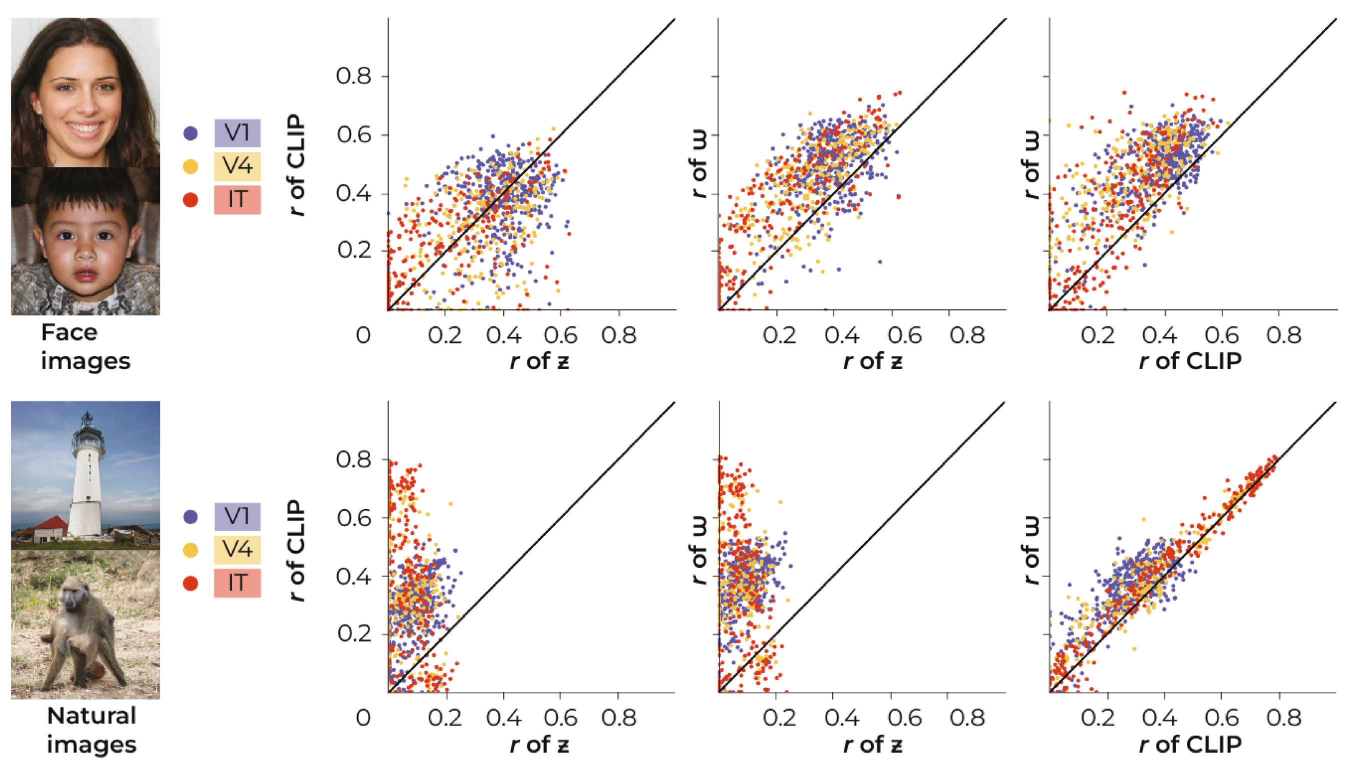
Passive fixation experiment.

- **Linear mapping** to evaluate our claim that the feature- and neural representation effectively encode the same stimulus properties, as is standard in neural coding. A more complex nonlinear transformation would not be valid to support this claim since nonlinearities will fundamentally change the underlying representations.

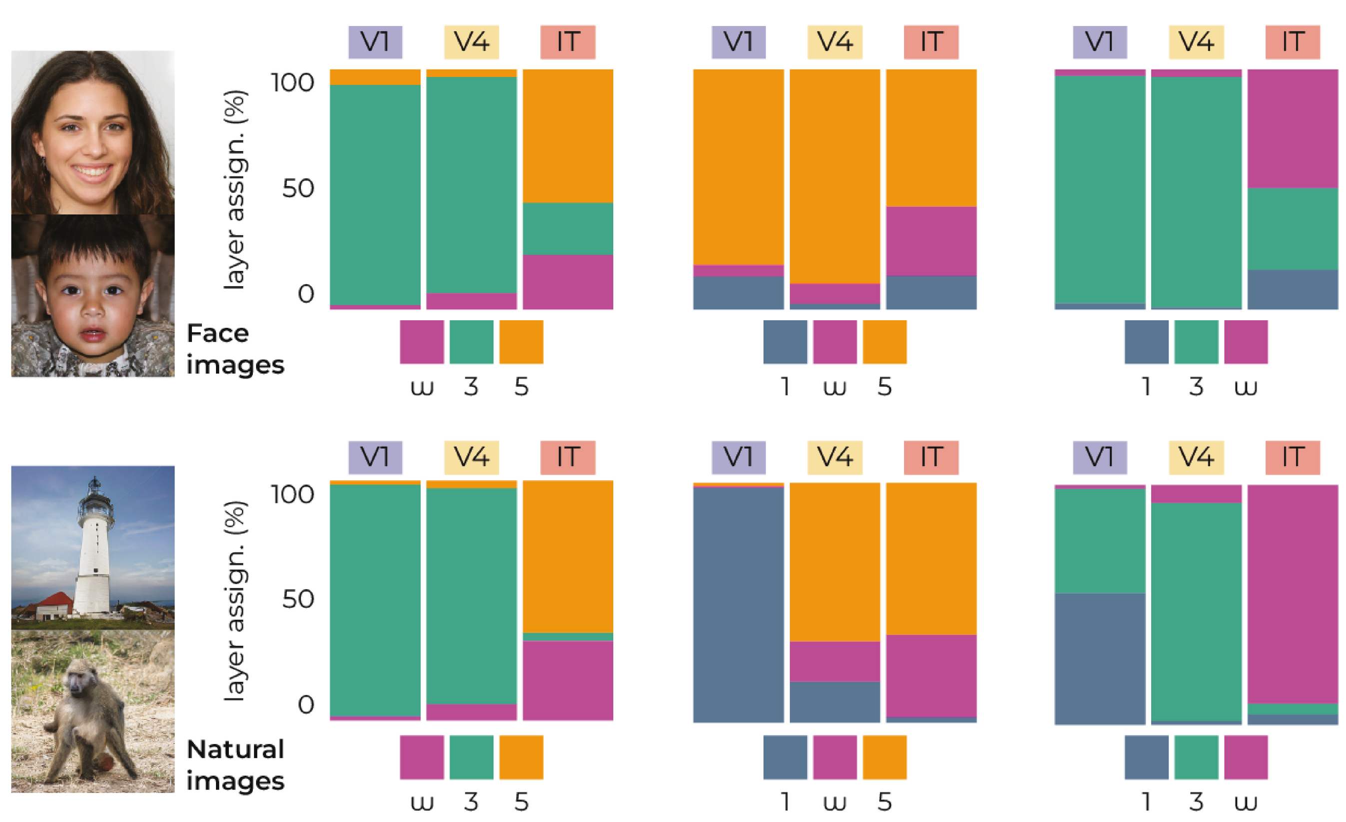
RESULTS

ENCODING

- A mass univariate neural encoding analysis of the latent representations showed that *feature-disentangled* representations, **ω-latents**, outperform the alternative representations, and explain increasingly more variance over the ventral stream.



Encoding performance from generative latents.



Layer assignment using early (1; layer 2/16), middle (3; layer 7/16) and deep (5; layer 13/16) activations of VGG16 pretrained for face/object recognition [5, 6] over visual areas results in the complexity gradient. Replacing one activation by ω-latents shows it predominantly acounts for high-level brain activity.

DECODING

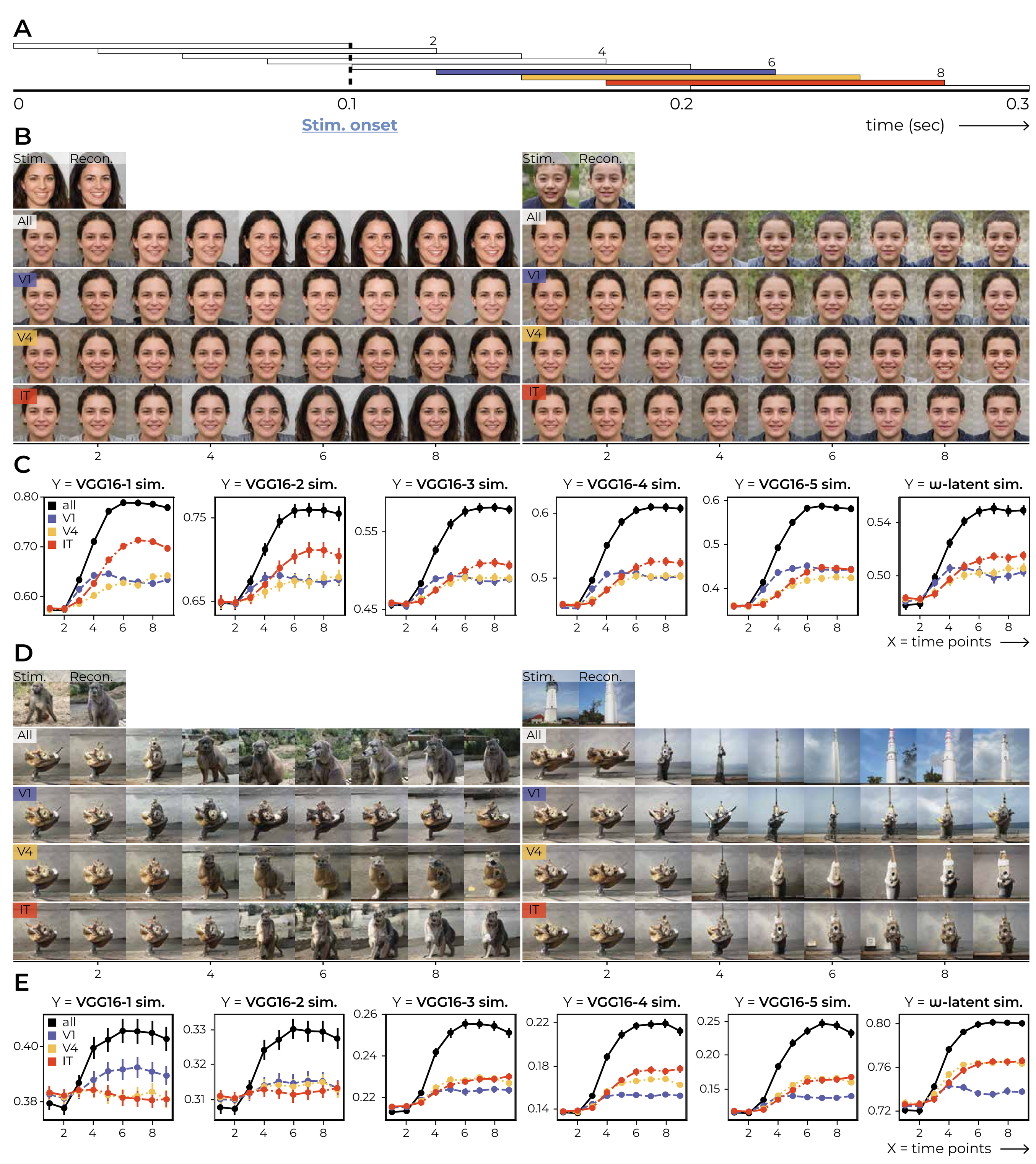
- A multivariate neural decoding analysis of the feature-disentangled representations resulted in state-of-the-art spatiotemporal reconstructions of visual perception.

		VGG16-1 sim.	VGG16-2 sim.	VGG16-3 sim.	VGG16-4 sim.	VGG16-5 sim.	Lat. sim.
Face images	*All*	**0.7871 ± 0.0102**	**0.7681 ± 0.0075**	**0.5874 ± 0.0075**	**0.6170 ± 0.0085**	**0.5940 ± 0.0104**	**0.5548 ± 0.0045**
	V1	0.6382 ± 0.0079	0.6758 ± 0.0064	0.4891 ± 0.0064	0.5041 ± 0.0083	0.4442 ± 0.0092	0.5022 ± 0.0047
	V4	0.6303 ± 0.0101	0.6729 ± 0.0068	0.4890 ± 0.0068	0.5006 ± 0.0085	0.4191 ± 0.0091	0.5026 ± 0.0040
	IT	*0.7123 ± 0.0110*	*0.7133 ± 0.0073*	*0.5093 ± 0.0073*	*0.5253 ± 0.0087*	*0.4434 ± 0.0096*	*0.5176 ± 0.0039*
Natural images	*All*	**0.4083 ± 0.0036**	**0.3322 ± 0.0036**	**0.2555 ± 0.0025**	**0.2192 ± 0.0043**	**0.2497 ± 0.0066**	**0.8032 ± 0.0032**
	V1	*0.3929 ± 0.0031*	*0.3147 ± 0.0031*	0.2223 ± 0.0019	0.1511 ± 0.0023	0.1367 ± 0.0037	0.7336 ± 0.0036
	V4	0.3790 ± 0.0029	0.3132 ± 0.0029	0.2270 ± 0.0019	0.1641 ± 0.0027	0.1617 ± 0.0045	0.7614 ± 0.0034
	IT	0.3798 ± 0.0026	0.3127 ± 0.0026	*0.2302 ± 0.0020*	*0.1790 ± 0.0039*	*0.1692 ± 0.0057*	*0.7653 ± 0.0039*

Decoding performance in terms of six metrics of perceptual cosine similarity using five intermediate layer activations of VGG16 for face- and object recognition for face- and natural images, repectively, and latent cosine similarity between ω-latents of stimuli and reconstructions (mean ± std. error).

DECODING

Example results. Stimulus (top) reconstructions (bottom) from brain activity via ω-latents.



Time-based decoding. **A** 100 ms window was slid with a stride of 25 ms over the entire time course per trial of 300 ms, resulting in nine average responses in time. For reference, the original predefined time windows for V1, V4 and IT are color-coded at the top. **B**, **D** show how two reconstructions evolve over time. **C**, **E** show decoding performance unfold over time.

CONCLUSIONS

- **Neural encoding:** feature-disentangled ω-latents were the most successful at predicting neural activity and explained increasingly more variance over the ventral stream
  - Demonstrates the potential of aligning unsupervised generative models with biological processes in general and highlights the importance of feature disentanglement in ex plaining high-level neural representations underlying visual perception in particular
- **Neural decoding:** image reconstructions with state-of-the-art quality were obtained that closely matched the stimuli in their semantic as well as structural features
  - StyleGAN itself has never been optimized on neural data which implies a general principle of shared encoding of real-world phenomena
  - Advancements of comp. models and clinical applications for people with disabilities

REFERENCES

[1] Super, H., & Roelfsema, P. R. (2005). Chronic multiunit recordings in behaving animals: advantages and limitations. Progress in brain research, 147, 263-282.
[2] Dado, T., Güçlütürk, Y., Ambrogioni, L., Ras, G., Bosch, S., van Gerven, M., & Güçlü, U. (2022). Hyperrealistic neural decoding for reconstructing faces from fMRI activations via the GAN latent space. Scientific reports, 12(1), 141.
[3] Karras, T., Aittala, M., Laine, S., Härkönen, E., Hellsten, J., Lehtinen, J., & Aila, T. (2021). Alias-free generative adversarial networks. Advances in Neural Information Processing Systems, 34, 852-863.
[4] Sauer, A., Schwarz, K., & Geiger, A. (2022, July). Stylegan-xl: Scaling stylegan to large diverse datasets. In ACM SIGGRAPH 2022 conference proceedings (pp. 1-10).
[5] Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). Deep face recognition.
[6] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Thirza Dado; thirza.dado@donders.ru.nl; https://neuralcoding.nl/

Radboud University Radboudumc