Hyperrealistic neural decoding: Reconstructing faces from fMRI measurements via the GAN latent space

Thirza Dado, Yağmur Güçlütürk, Luca Ambrogioni, Gabriëlle Ras, Sander Bosch, Marcel van Gerven, Umut Güçlü

Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, Netherlands

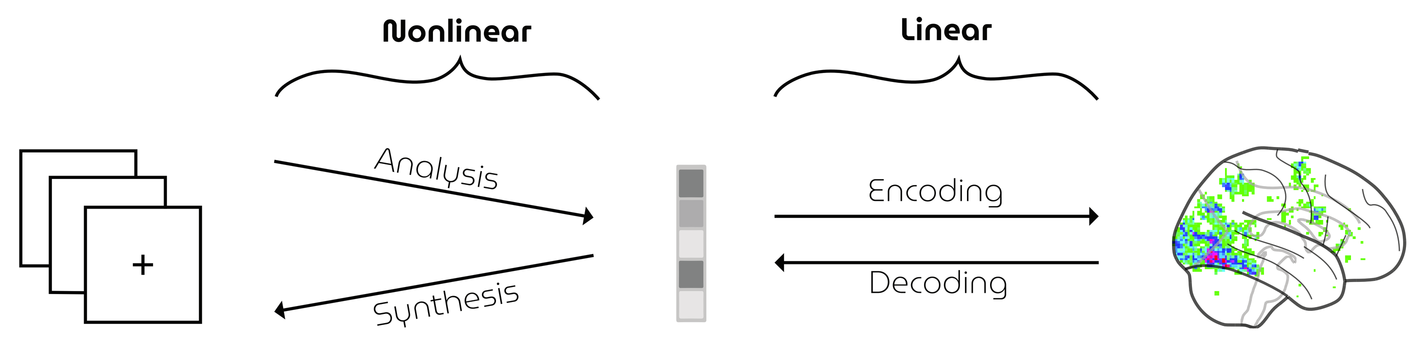
INTRODUCTION

· Unlike their supervised counterparts [1], more biologically plausible unsupervised deep neural networks (DNNs) have paradoxically been less successful in modeling neural representations [2].

· At the same time, generative adversarial networks (GANs) have become one of the most powerful unsupervised DNNs in modeling image distributions.

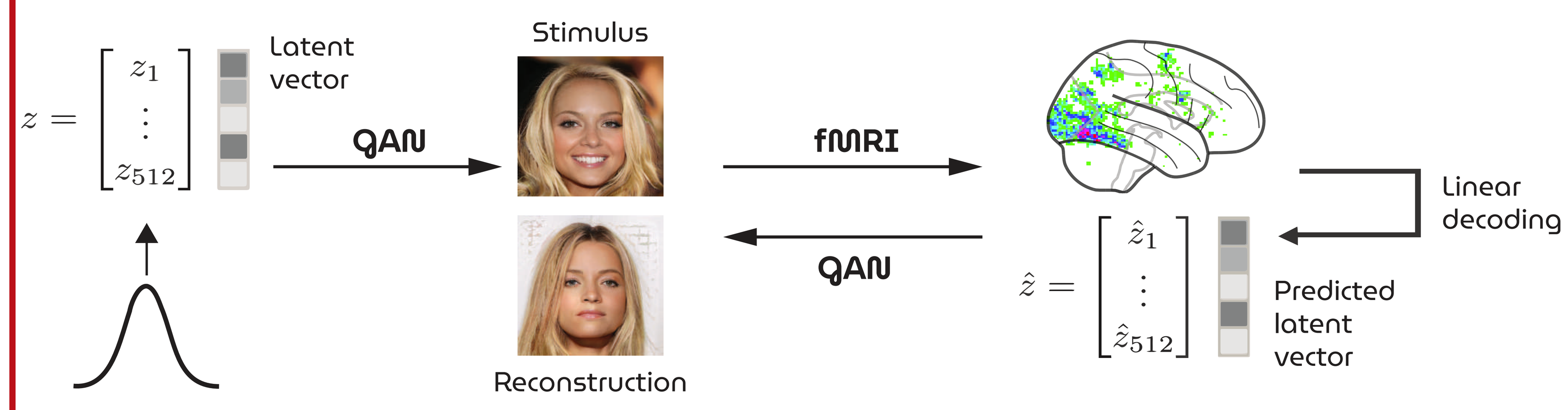
· **Problem**: GANs have high potential in modeling neural representations, but testing this hypothesis is not directly possible because latent vectors cannot be obtained retrospectively [3].

· **Solution: A novel experimental paradigm** for well-controlled yet naturalistic stimuli with known latent vectors and **a GAN-based neural decoding model** for HYperrealistic reconstruction of PERception (HYPER) with unprecedented accuracy to date.



Neural decoding can be conceptualized as the inverse problem of mapping brain responses back to sensory stimuli via a latent space.

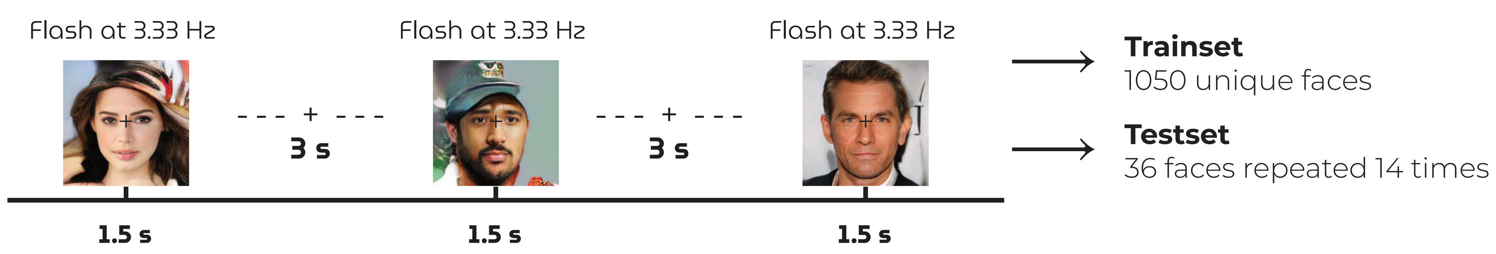
METHODS



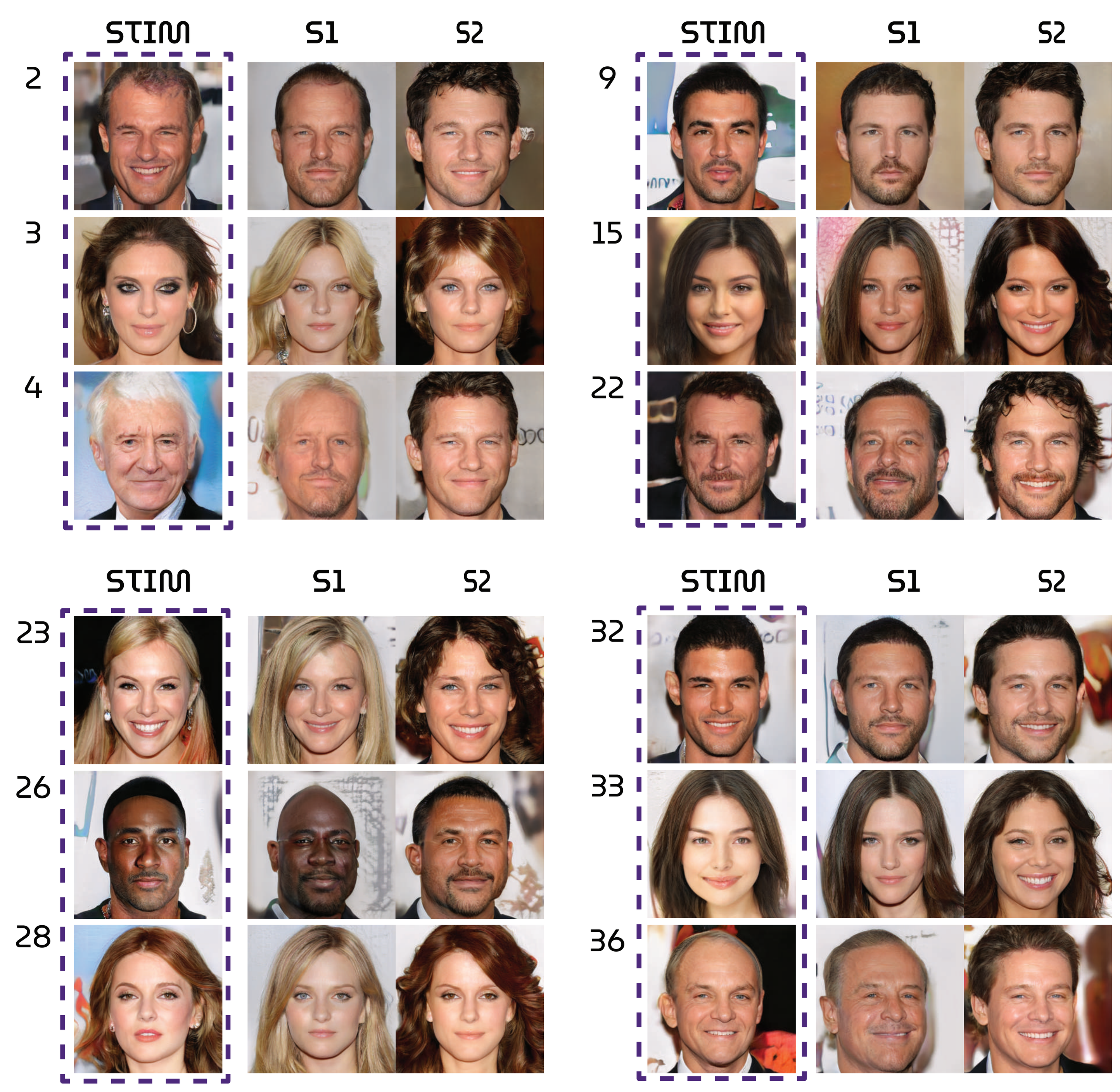
· **The progressive growing of GANs (PGGAN) model** [4] generates photorealistic faces (1024 x 1024 pixels) that resemble celebrities from randomly sampled standard Gaussian latent vectors (512 dims).

· **Blood-oxygen-level dependent hemodynamic responses** (TR = 1.5 s, voxel size = 2 x 2 x 2 mm³, whole brain, mb4) of two subjects were measured during presentation of faces (15° stimuli, 0.6° fixation cross).

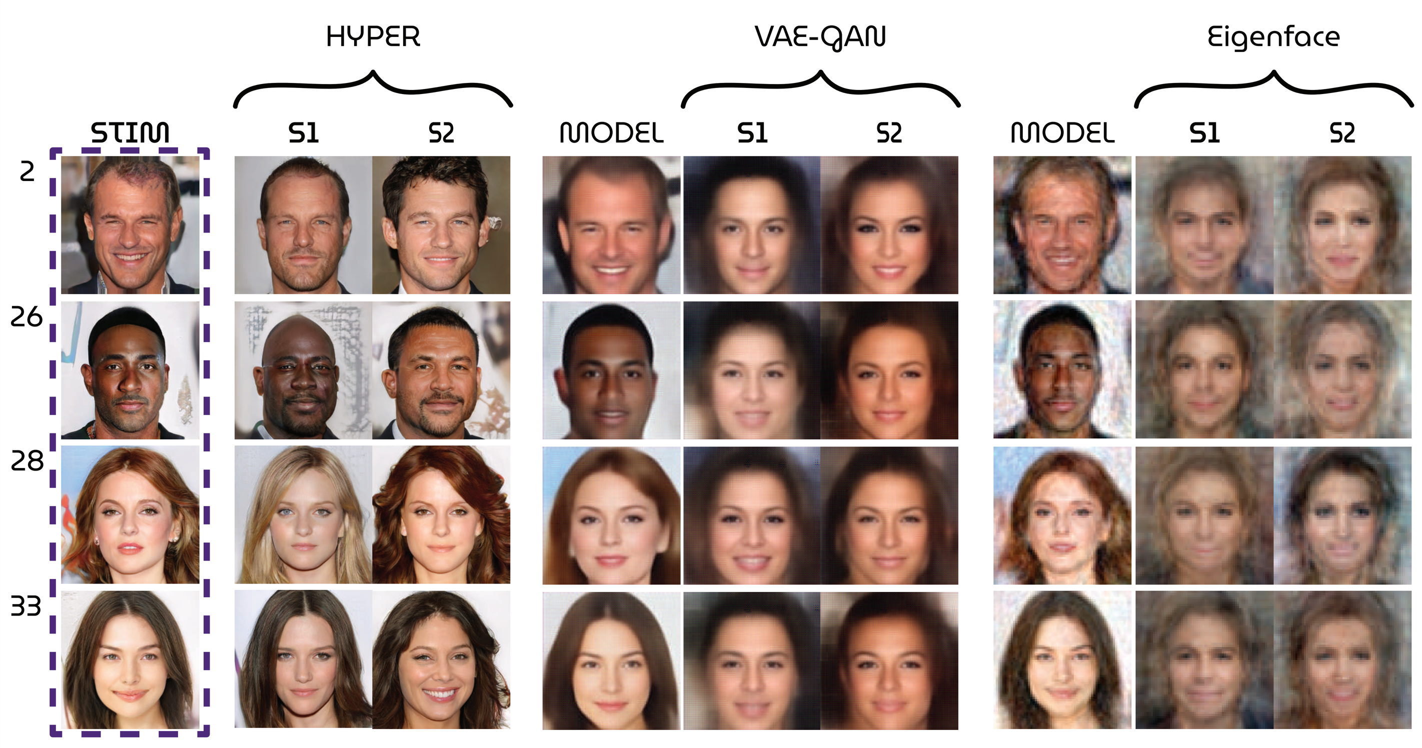
· **Linear decoding** by a dense layer at the beginning of PGGAN to transform brain data into latent vectors.



RESULTS



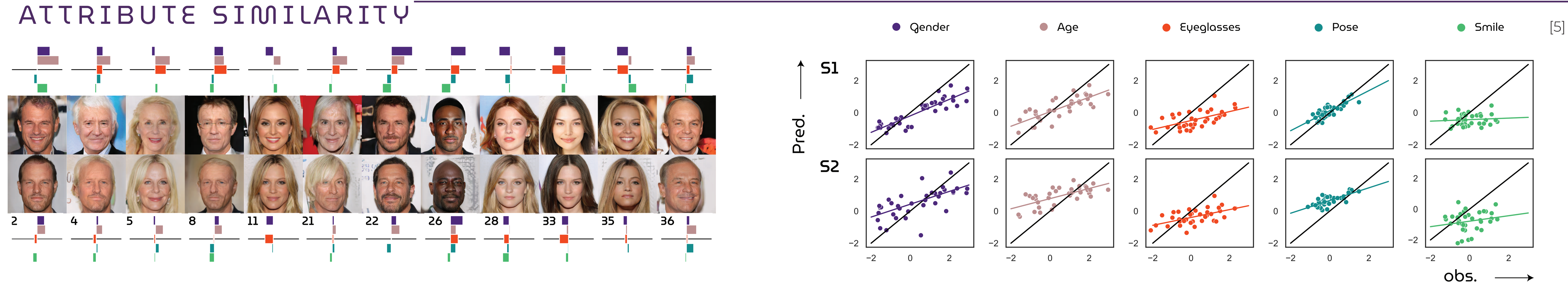
QUALITATIVE RESULTS



QUANTITATIVE RESULTS

		Lat. sim.	Feat. sim.	Struct. sim.
S1	HYPER	0.4521 ± 0.0026	0.1745 ± 0.0038	0.6663 ± 0.0115
		$(p < 0.001; perm.test)$	$(p < 0.001; perm.test)$	$(p < 0.001; perm.test)$
	VAE-GAN	-	0.1416 ± 0.0025	0.5598 ± 0.0151
	Eigenface	-	0.1319 ± 0.0016	0.5877 ± 0.0115
S2	HYPER	0.4447 ± 0.0020	0.1715 ± 0.0049	0.6035 ± 0.0128
		$(p < 0.001; perm.test)$	$(p < 0.001; perm.test)$	$(p < 0.001; perm.test)$
	VAE-GAN	-	0.1461 ± 0.0022	0.5832 ± 0.0141
	Eigenface	-	0.1261 ± 0.0019	0.5616 ± 0.0097

ATTRIBUTE SIMILARITY



CONCLUSIONS

With the introduced paradigm and model we:

· Showed that unsupervised deep neural networks can successfully model neural representations of naturalistic stimuli
· Showed that the GAN latent space approximates the neural face manifold
· Obtained state-of-the-art reconstructions of perceived faces from brain activations

Considering the speed of progress in the field of generative modeling, the HYPER framework will likely result in even more impressive reconstructions of perception and possibly even imagery in the near future.

REFERENCES

[1] Güçlü, U., & van Gerven, M. A. (2015). Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. Journal of Neuroscience, 35(27), 10005-10014.
[2] Khaligh-Razavi, S. M., & Kriegeskorte, N. (2014). Deep supervised, but not unsupervised, models may explain IT cortical representation. PLoS computational biology, 10(11), e1003915.
[3] Seeliger, K., Güçlü, U., Ambrogioni, L., Güçlütürk, Y., & van Gerven, M. A. (2018). Generative adversarial networks for reconstructing natural images from brain activity. NeuroImage, 181, 775-785.
[4] Karras, T., Aila, T., Laine, S., & Lehtinen, J. (2017). Progressive growing of gans for improved quality, stability, and variation. arXiv preprint arXiv:1710.10196.
[5] Shen, Y., Gu, J., Tang, X., & Zhou, B. (2020). Interpreting the latent space of gans for semantic face editing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 9243-9252).

Thirza Dado; thirza.dado.1@donders.ru.nl; https://neuralcoding.nl/

Radboud University   Radboudumc