Reconstructing perceived faces from multi-subject fMRI activations with hyper-aligning and -decoding

Thirza Dado, Yağmur Güçlütürk, Marcel van Gerven, Umut Güçlü

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

DONDERS INSTITUTE

INTRODUCTION

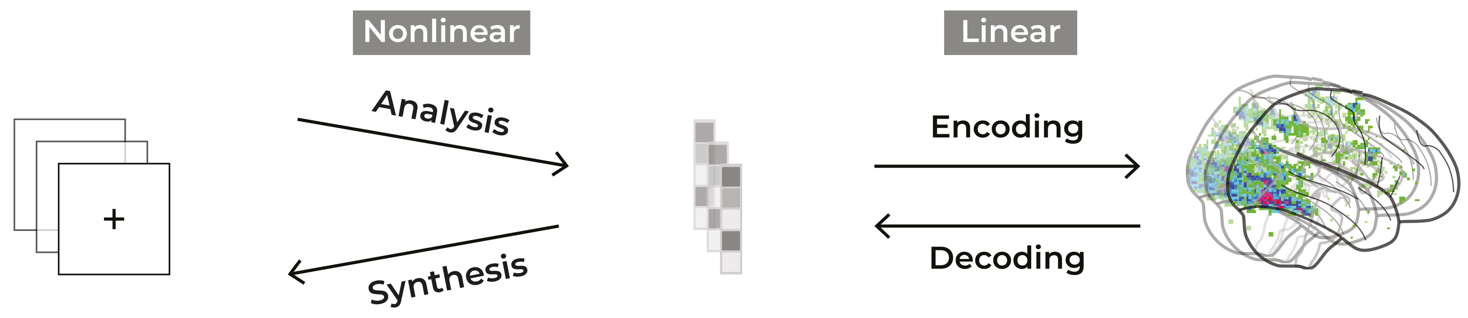
· Unlike their supervised counterparts [1], more biologically plausible unsupervised deep neural networks (DNNs) have paradoxically been less successful in modeling neural representations [2].

· 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 retrospecively [3].

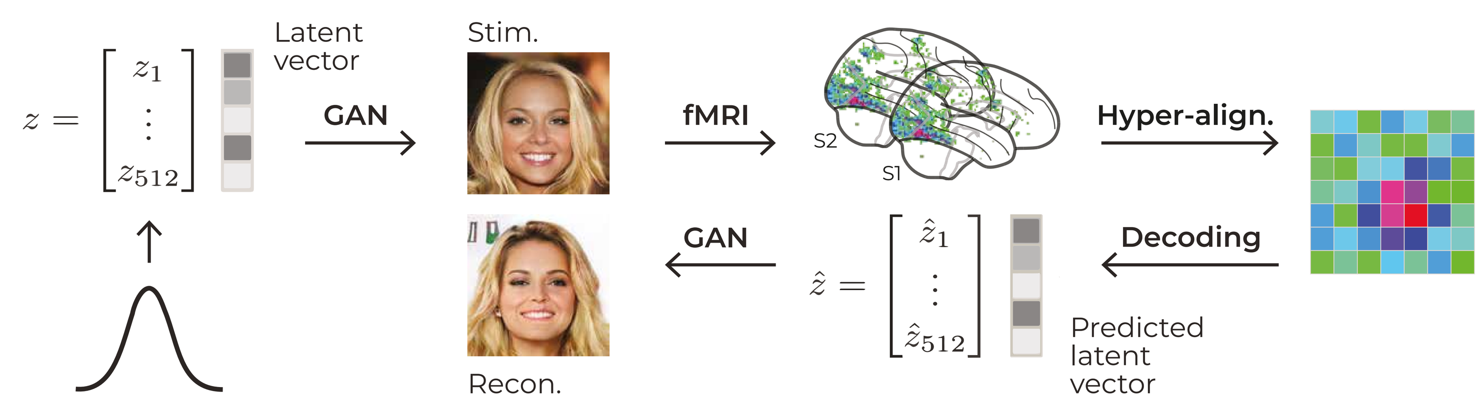
· **Solution: A novel experimental paradigm** for well-controlled yet naturalistic stimuli with a priori known latent vectors and **a GAN-based neural decoding model "hyper²"** that combines HYperrealistic reconstruction of PERception (*hyper-decoding*) [4] as well as *hyper-aligning* [5] to capture the shared neural information across participants.

· Hyper² obtained reconstructions from brain activity with unprecedented accuracy to date.



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

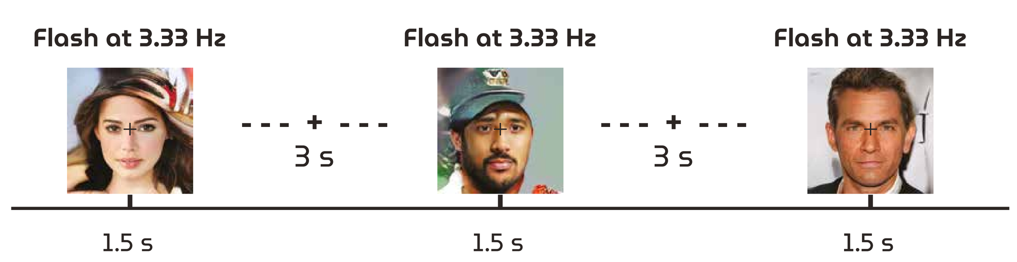
METHODS



· **The progressive growing of GANs (PGGAN) model** [6] 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) and hyperaligned by mapping to a common functional space.

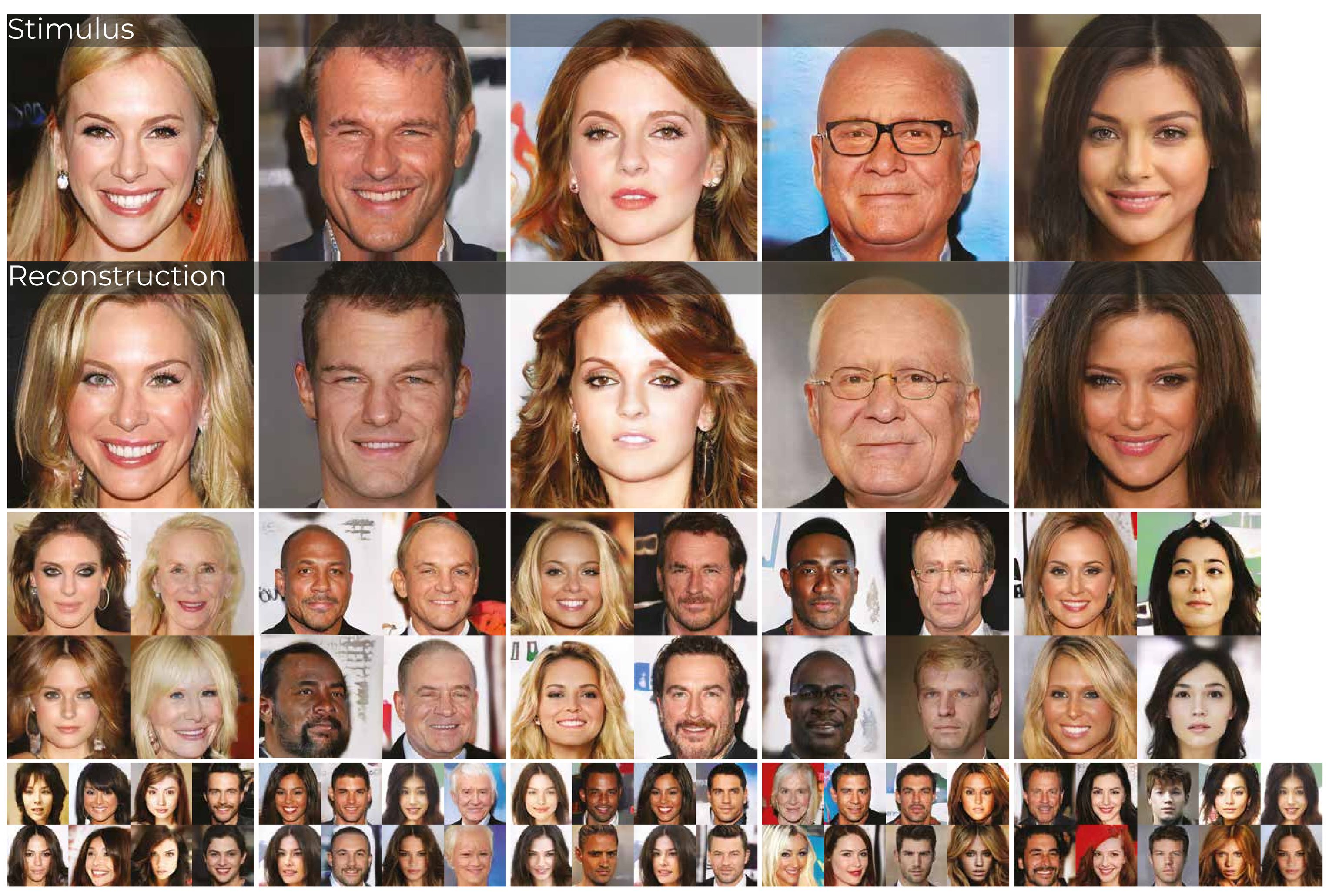
· **Neural decoding** by prepending a dense layer at PGGAN to map brain recordings to latent vectors that is fit with ordinary least squares.

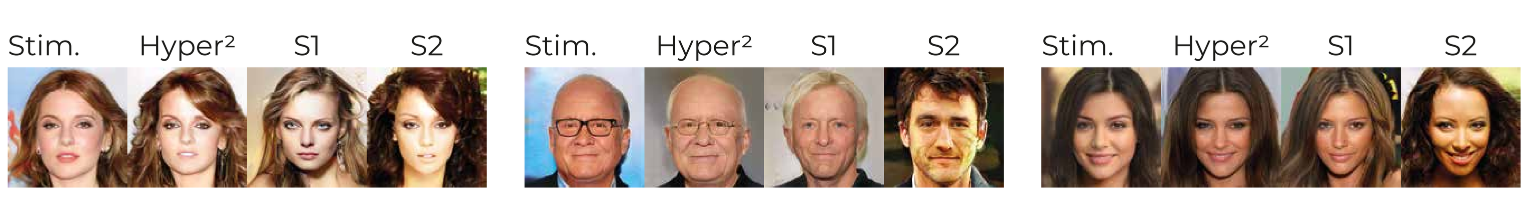


Training set
1050 unique faces

Test set
36 faces repeated 14 times

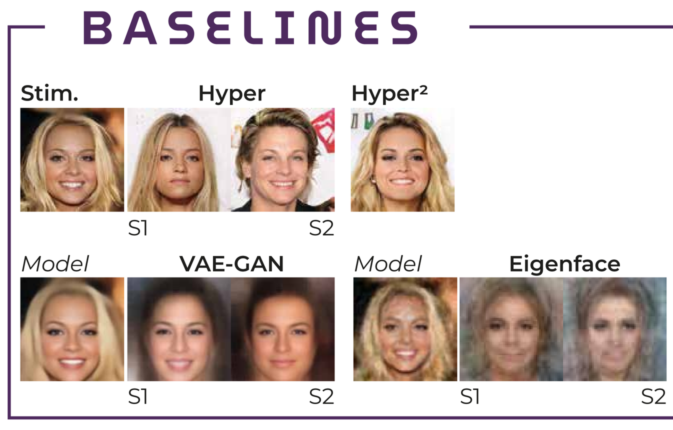
RESULTS





	Lat. Sim.	Feat. Sim.	Pix. Corr.	SSIM
Hyper²	**0.4138 ± 0.0025**	**0.5839 ± 0.0076**	**0.3490 ± 0.0001**	**0.3322**
	($p < 0.001; perm.test$)	($p < 0.001; perm.test$)	($p < 0.001; perm.test$)	($p < 0.001; perm.test$)
S1	0.3188 ± 0.0035	0.5642 ± 0.0078	0.3091 ± 0.0001	0.2739
S2	0.3027 ± 0.0034	0.5278 ± 0.0078	0.2167 ± 0.0001	0.2017

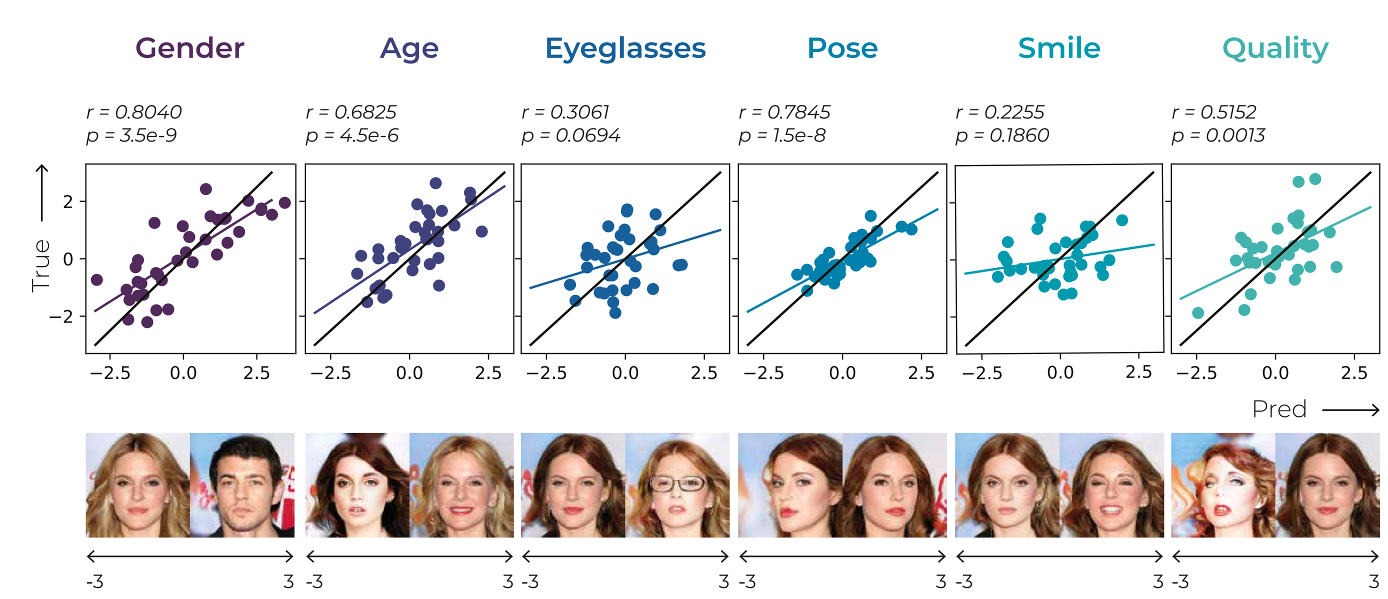
BASELINES



Feature similarity (VGGFace)

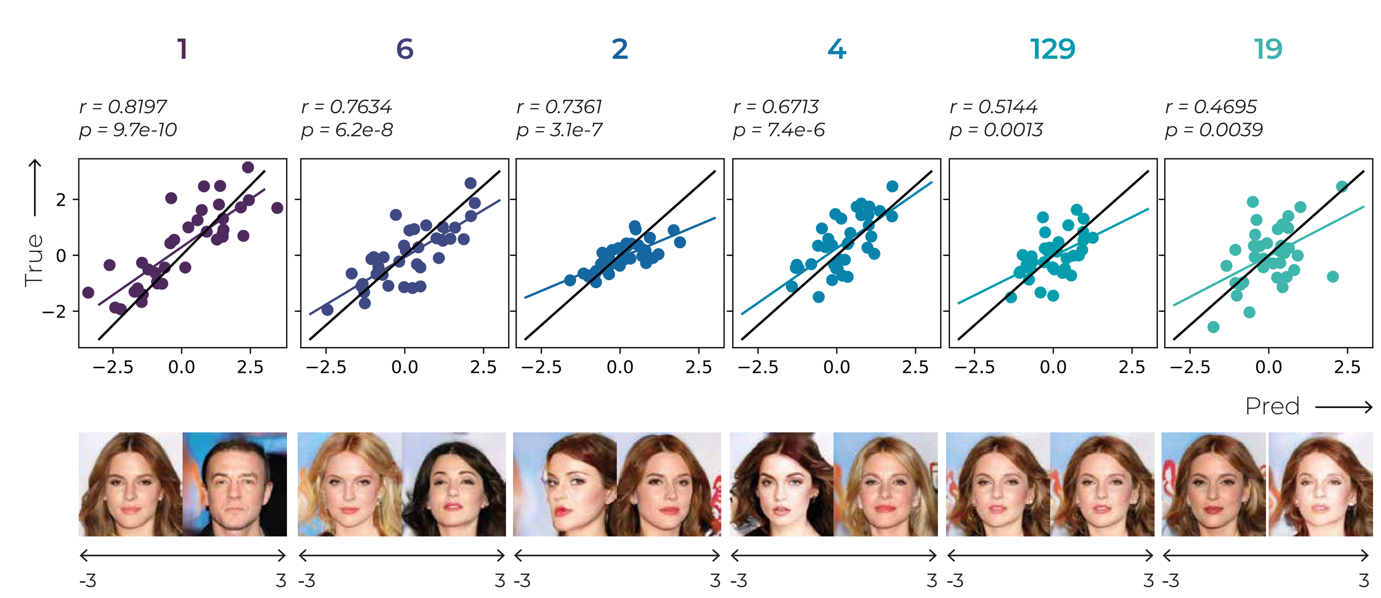
	S1	S2
Hyper²	0.2365 ± 0.0037	
Hyper	0.1656 ± 0.0050	0.1665 ± 0.0058
VAE-GAN	0.1416 ± 0.0025	0.1461 ± 0.0022
Eigenface	0.1319 ± 0.0016	0.1261 ± 0.0019

ATTRIBUTE SCORES 1



Support vector machines identified the decision boundaries for six facial attributes [7].

ATTRIBUTE SCORES 2



Closed-form factorization identified 512 latent semantic directions [8].

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 activity

Considering the speed of progress in the field of generative modeling, this framework together with hyper-aligning a larger pool of eyewitnesses 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] 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), 1-9.
[5] Haxby, J. V., Guntupalli, J. S., Connolly, A. C., Halchenko, Y. O., Conroy, B. R., Gobbini, M. I., ... & Ramadge, P. J. (2011). A common, high-dimensional model of the representational space in human ventral temporal cortex. Neuron, 72(2), 404-416.
[6] Karras, T., Aila, T., Laine, S., & Lehtinen, J. (2017). Progressive growing of gans for improved quality, stability, and variation. arXiv preprint arXiv:1710.10196.
[7] 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).
[8] Shen, Y., & Zhou, B. (2021). Closed-form factorization of latent semantics in gans. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 1532-1540).

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

Neural Coding Lab   Radboud University   Radboudumc