University of Glasgow | Institute of Neuroscience & Psychology

CENTRE FOR COGNITIVE NEUROIMAGING

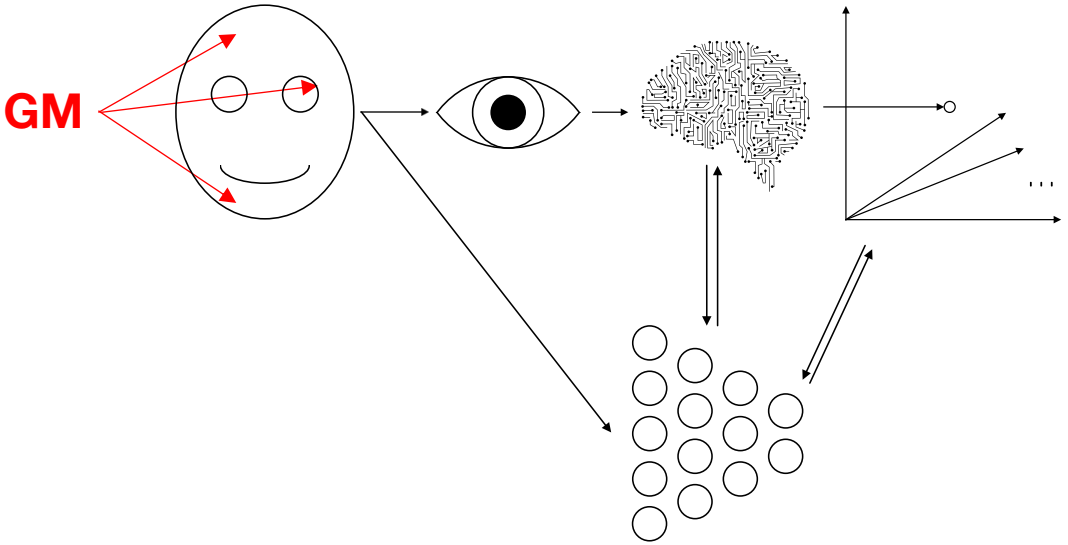
wellcome

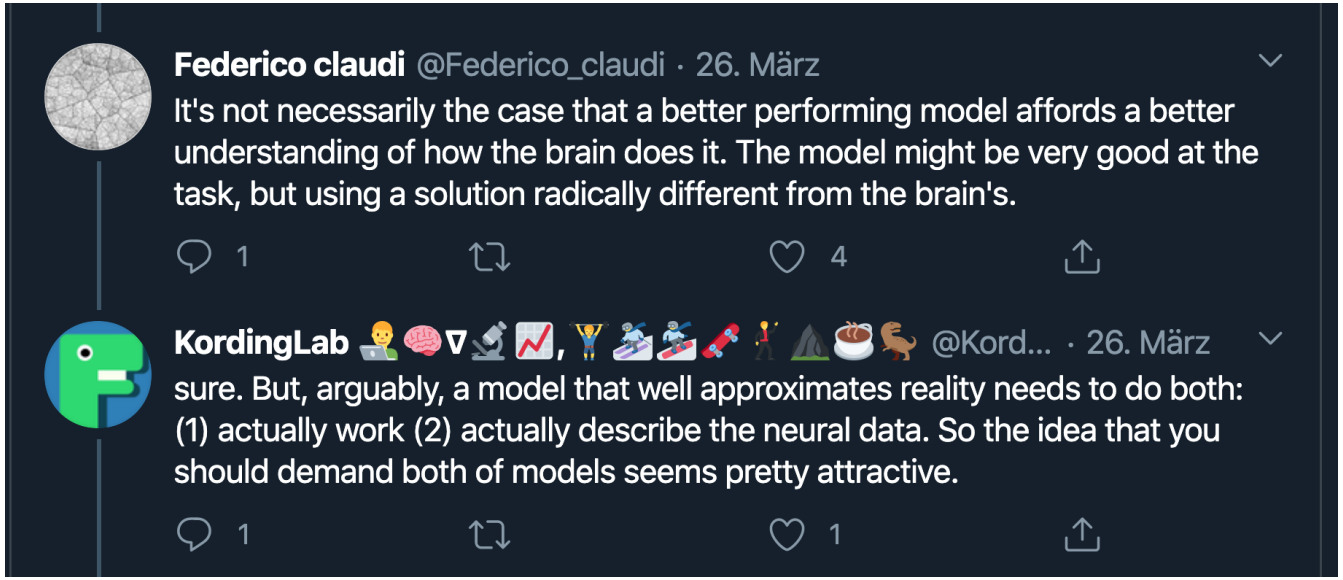
Understanding how DCNNs predict human behaviour using an interpretable generative model of 3D faces

Christoph Daube, Jiayu Zhan, Tian Xu, Andrew Webb, Robin A. A. Ince, Oliver B. Garrod, Philippe G. Schyns

@realdaubman

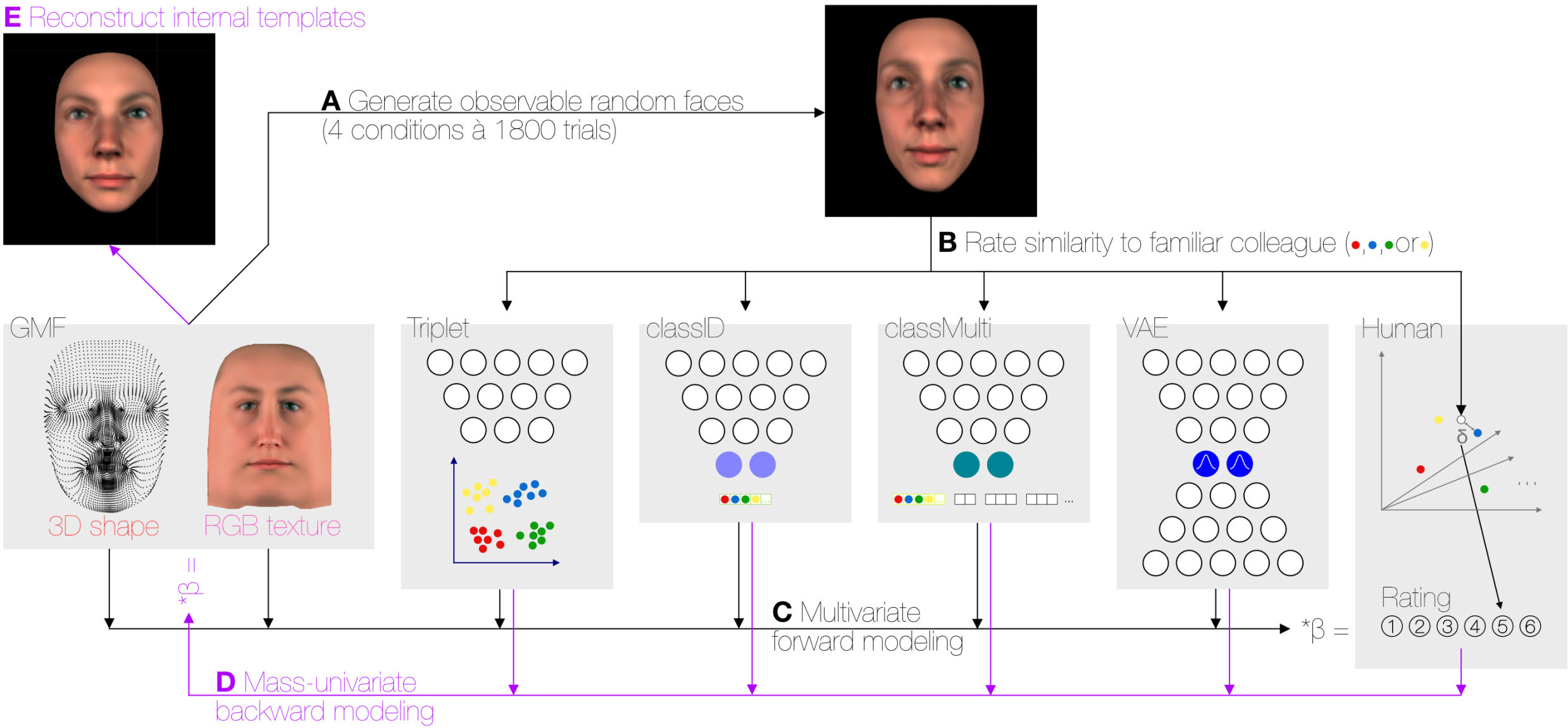
DCNNs are popular models of human vision – but do they really help us *understanding* it?



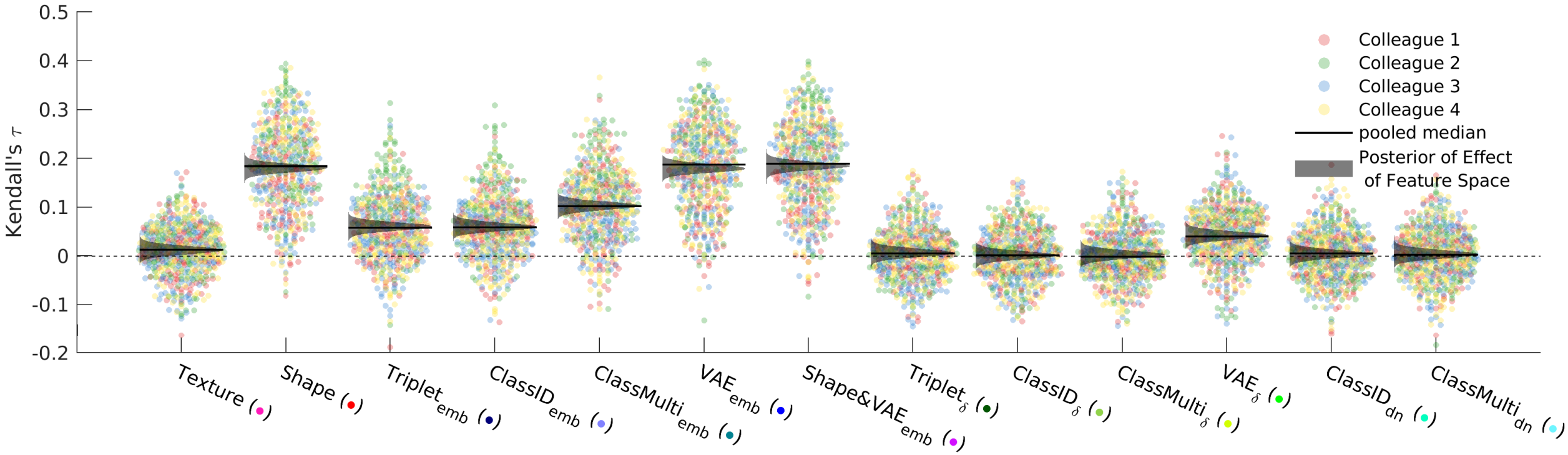


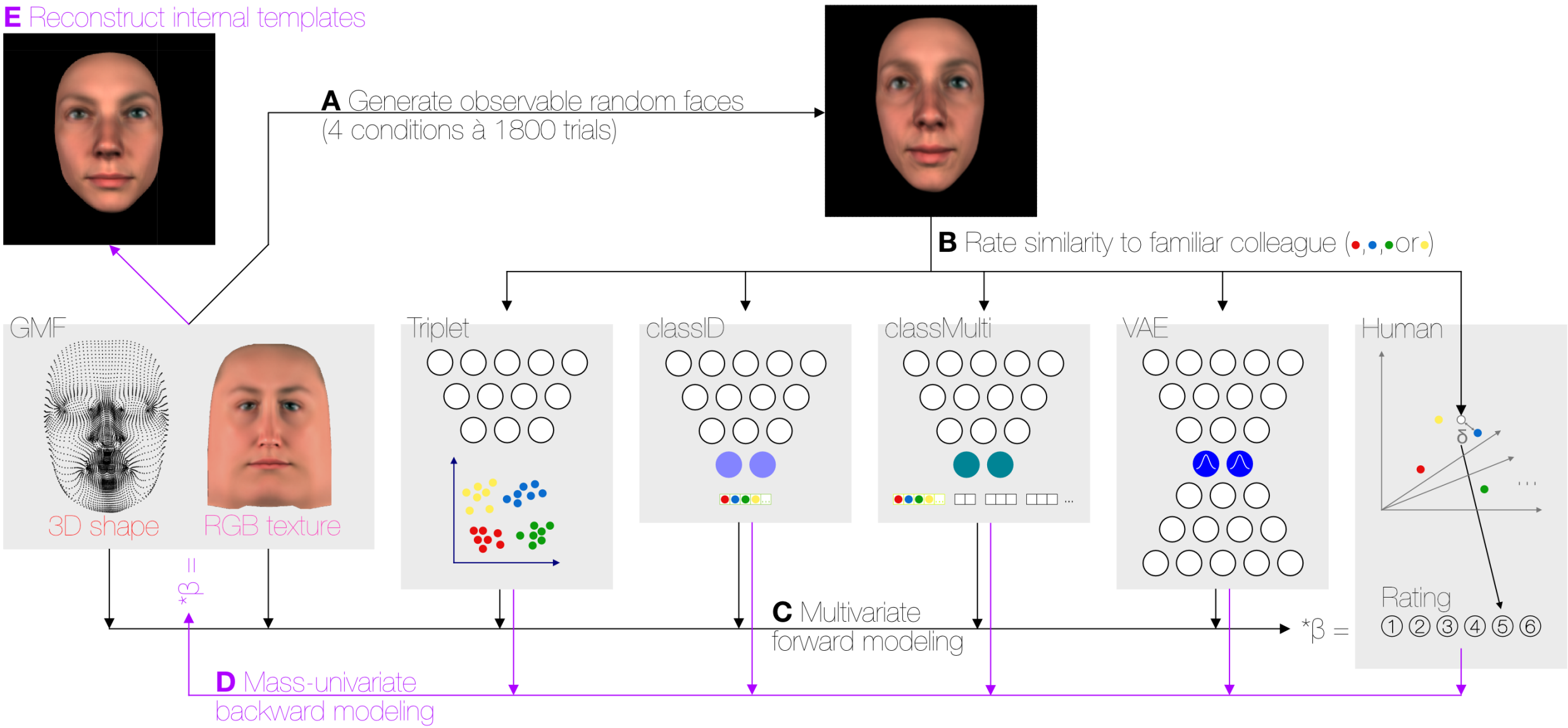
(3): If we want to understand what humans and models are doing and what they have in common, we have to specify what we want to understand – with a **generative model** of the stimuli that allows experimental control!

brain as a computer designed by Freepik http://www.freepik.com


E Reconstruct internal templates
A Generate observable random faces
(4 conditions à 1800 trials)
B Rate similarity to familiar colleague (●,●,●or●)
GMF
3D shape
RGB texture
Triplet
classID
classMulti
VAE
Human
Rating
① ② ③ ④ ⑤ ⑥
*β =
C Multivariate
forward modeling
D Mass-univariate
backward modeling

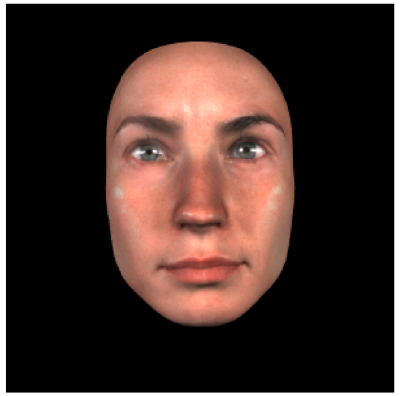
Forward modelling: Comparing predictive power of feature spaces



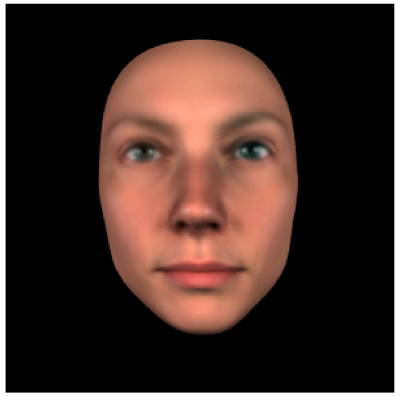

E Reconstruct internal templates
A Generate observable random faces
(4 conditions à 1800 trials)
B Rate similarity to familiar colleague (●,●,●or●)
GMF
3D shape
RGB texture
Triplet
classID
classMulti
VAE
Human
Rating
① ② ③ ④ ⑤ ⑥
*β =
C Multivariate
forward modeling
D Mass-univariate
backward modeling

Reconstructions from human and artificial systems

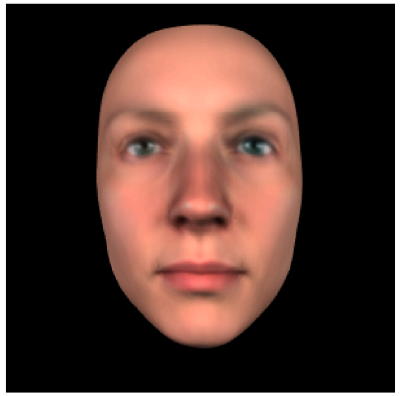
Ground Truth



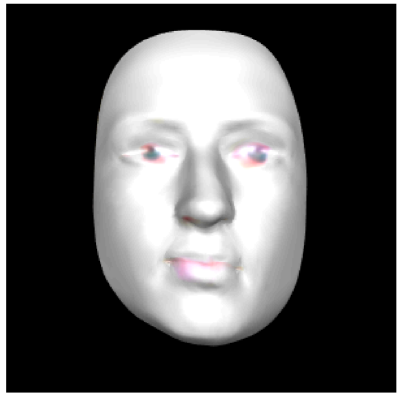
Human

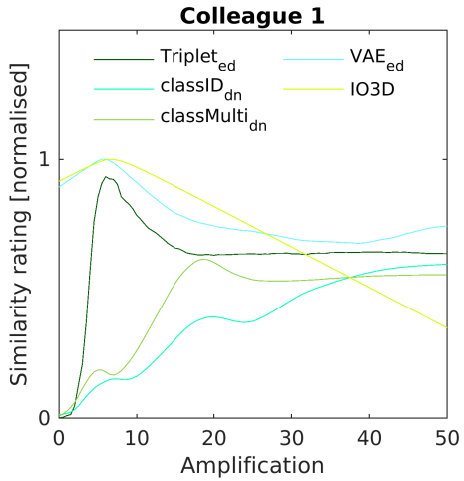


Triplet$_{ed}$



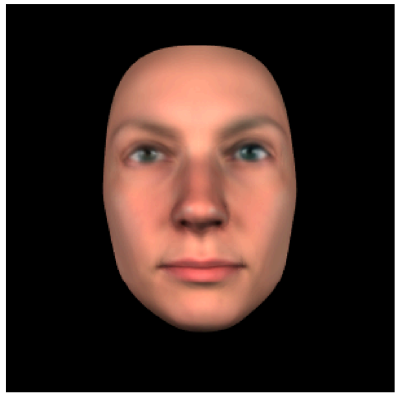
ClassID$_{dn}$



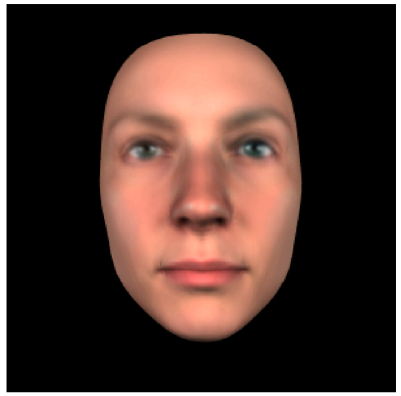


ClassMulti$_{dn}$

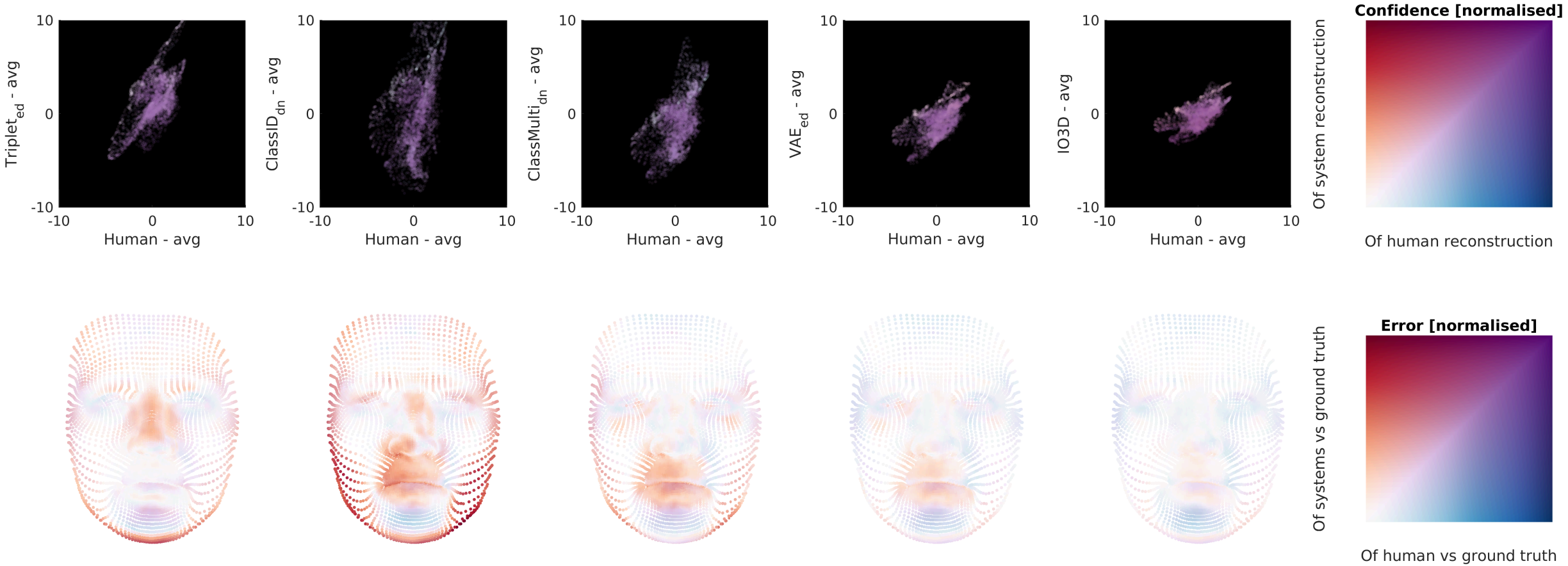
VAE$_{ed}$



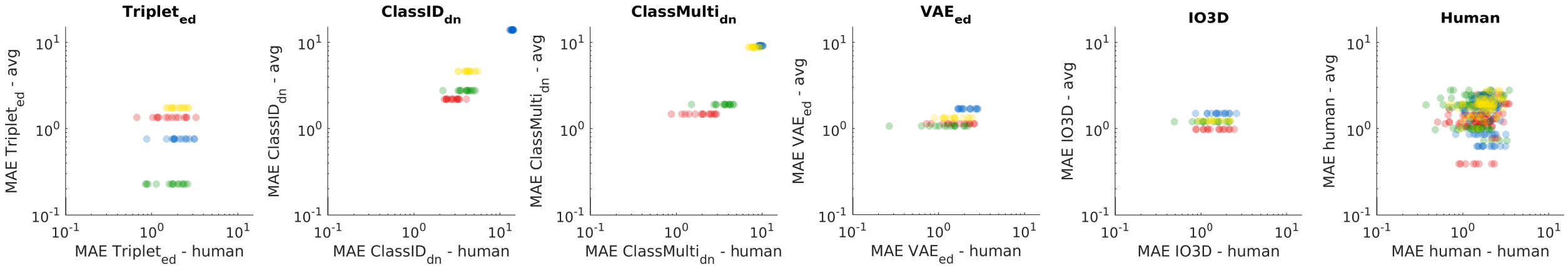
IO3D



Quantifying human and DNN shape perception w.r.t. generative model



Quantifying human and DNN shape perception w.r.t. generative model



Conclusions

- Human rating behaviour can be better predicted from **VAE** features than from classifier DNN features

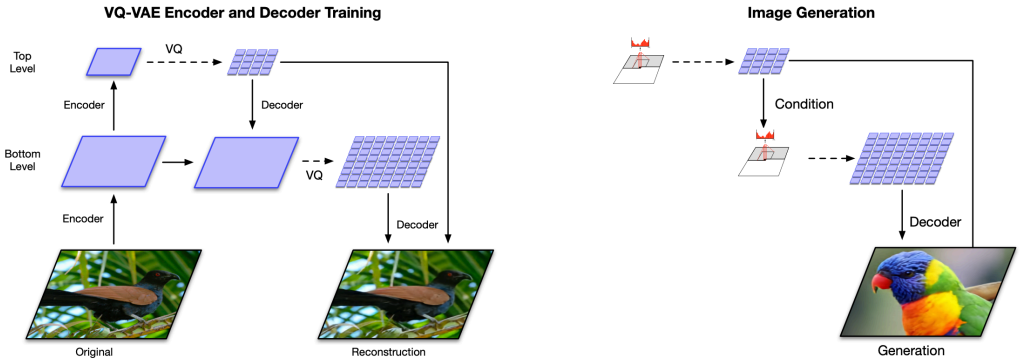
- Importantly, we also know **how** the VAE achieves this: By representing 3D shape features more like humans do

This is a group effort. Thanks to SchynsLab @ Uni Glasgow, especially

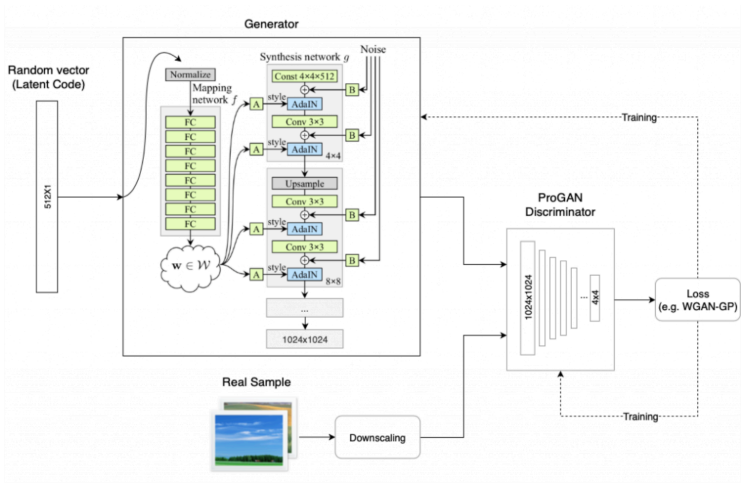
- Jiayu Zhan
- Tian Xu
- Andrew Webb
- Robin A. A. Ince
- Oliver B. Garrod
- Philippe G. Schyns

Discussion:
Do this but sample from VAE vs GAN?

VQ-VAE
Razavi
etal 2019

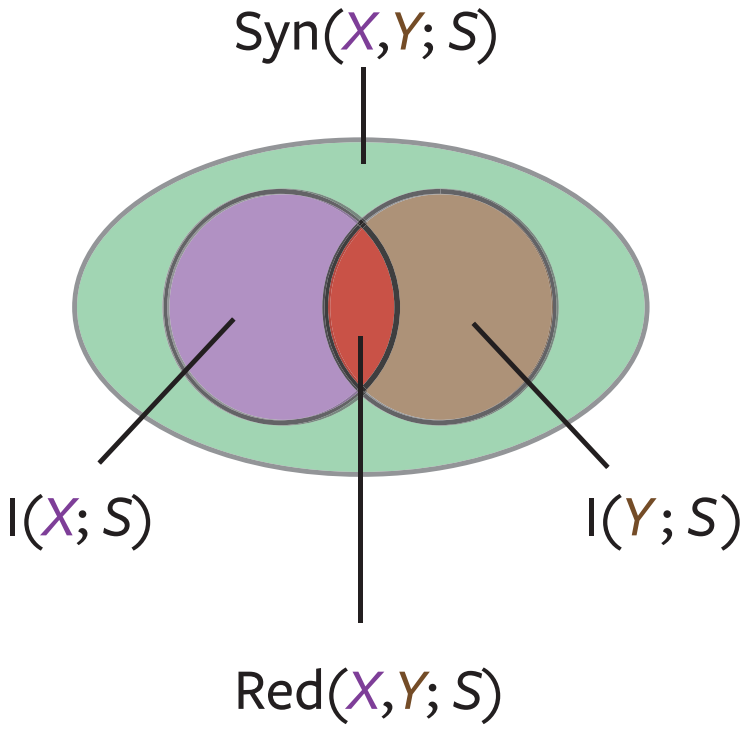


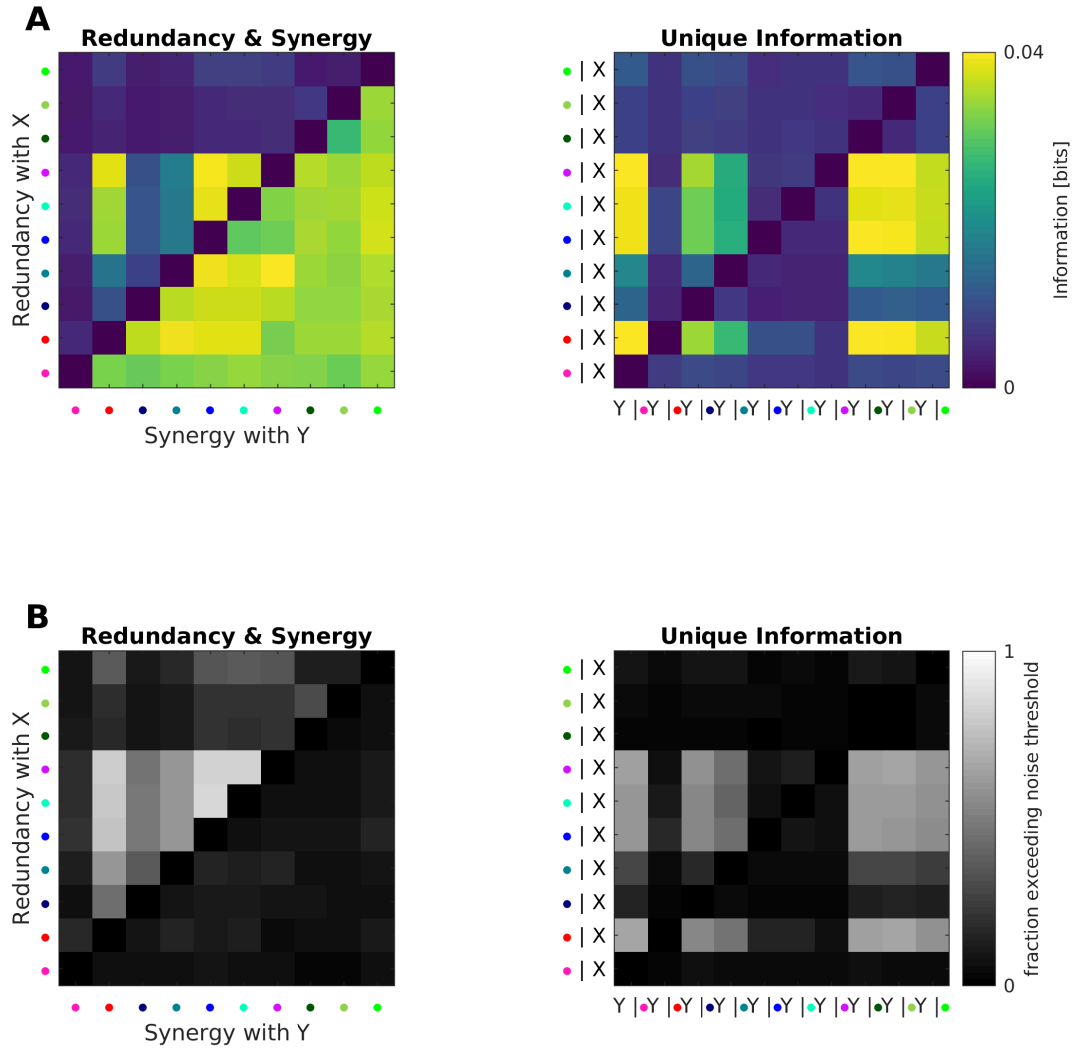
Style GAN (2)
Karras etal
2018,
(Karras etal
2020)



- would any of these models be better for reverse correlation? Why?
- with regards to the initial problem: “What can VAEs (or GANs) do for Vision Sciences”? What *new* insights could it provide beyond engineering?

Forward modelling: Comparing predictive power of feature spaces





Ince, 2017
Daube etal, 2019a
Daube etal, 2019b

Forward modelling: Weights are task-dependent