Hierarchical VAEs provide a normative account of motion processing in the primate brain

Hadi Vafaii[1], Jacob L. Yates[2], Daniel A. Butts[3]

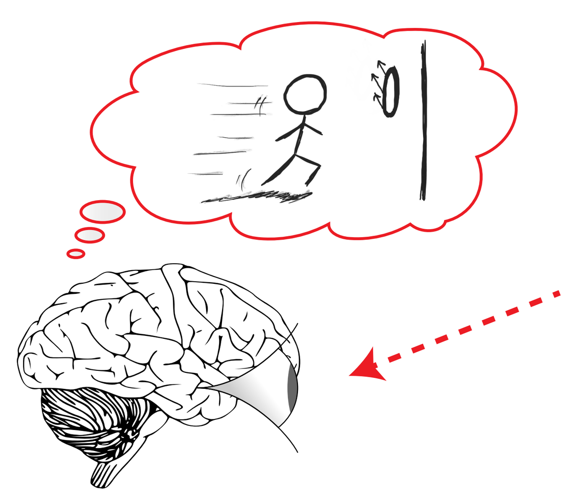
(1) Deptartment of Physics, University of Maryland, College Park, MD, 20742, USA (2) Department of Optometry & Vision Science, UC Berkeley, CA, 94720, USA (3) Department of Biology, University of Maryland, College Park, MD, 20742, USA

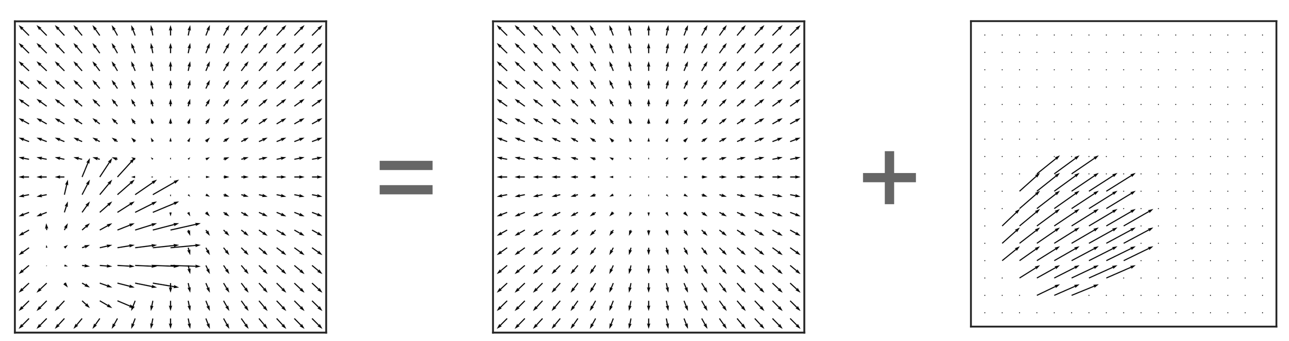
Goal: evaluate the role of ***hierarchical inference*** [1, 2] in motion perception

+ We introduce a new hierarchical variational autoencoder (VAE) model
+ We introduce a new synthetic data framework with ground truth

We evalute the role of :
(i) Model architecture (hierarchical vs. non-hierarchical)
(ii) Loss function (probabilistic inference vs. reconstruction)

"Perception as unconscious inference" [3]

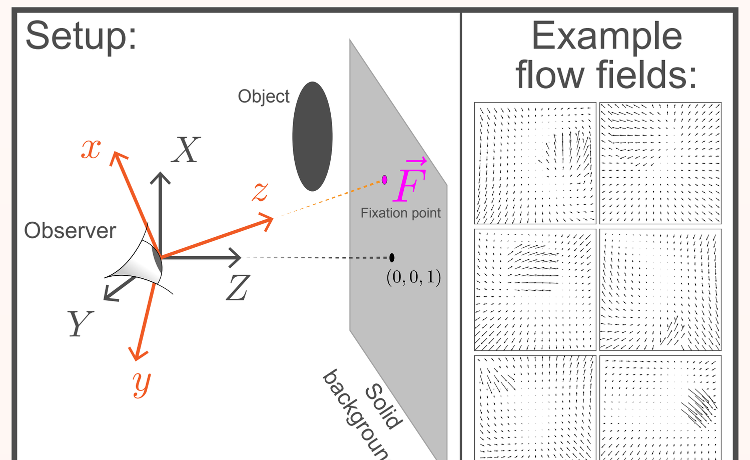




retinal optic flow = self-motion (global) + object motion (local)

Simulation: Retinal Optic Flow Learning (ROFL)

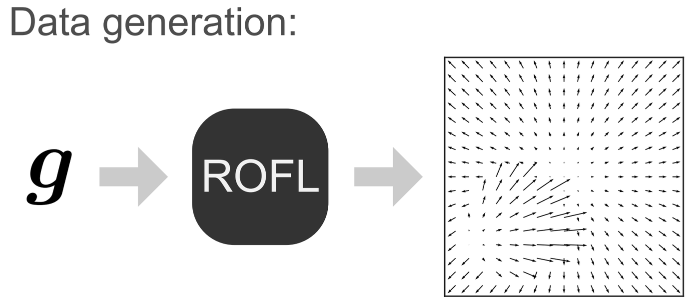
Setup:



Independent ground truth factors: $g \in \mathbb{R}^{11}$

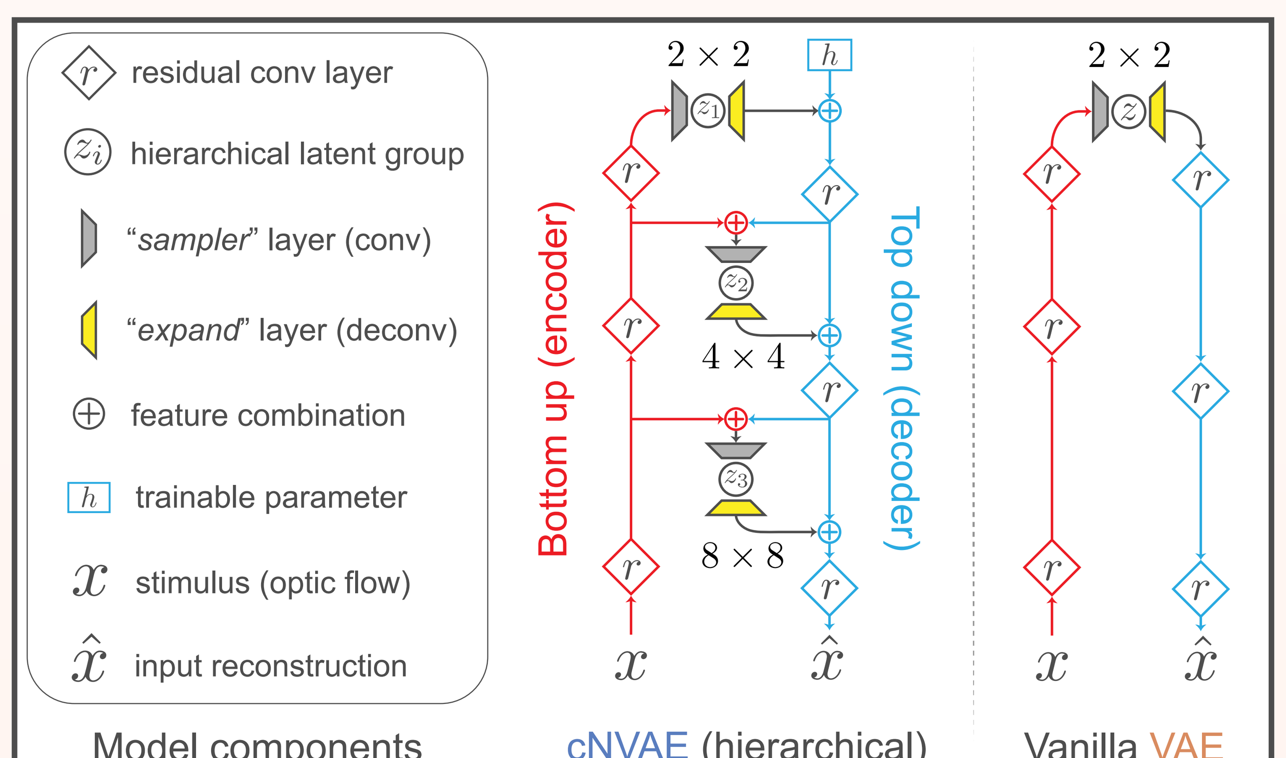
- Fixation point (+2)
- 3D self-motion velocity (+3)
- 3D object velocity (+3)
- 3D object position (+3)

Data generation:

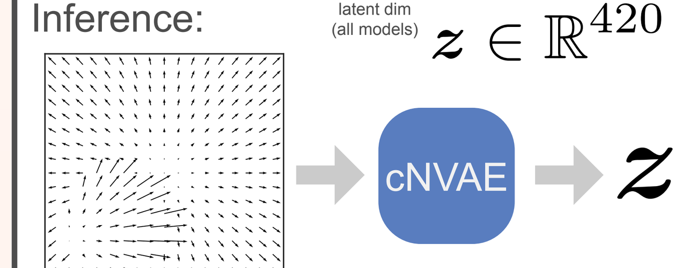


Static ("snapshot") velocity field: $x \in \mathbb{R}^{2\times17\times17}$

Architecture: compressed Nouveau VAE [4] (cNVAE)



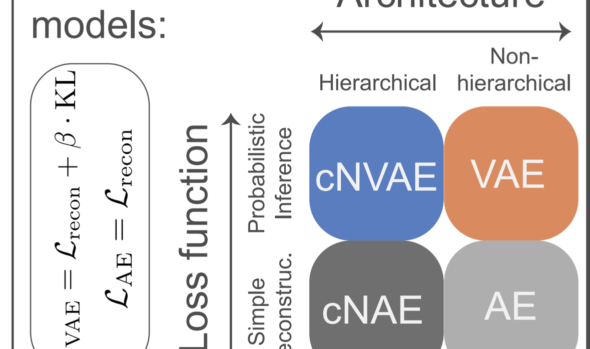
Inference:



$z \in \mathbb{R}^{420}$

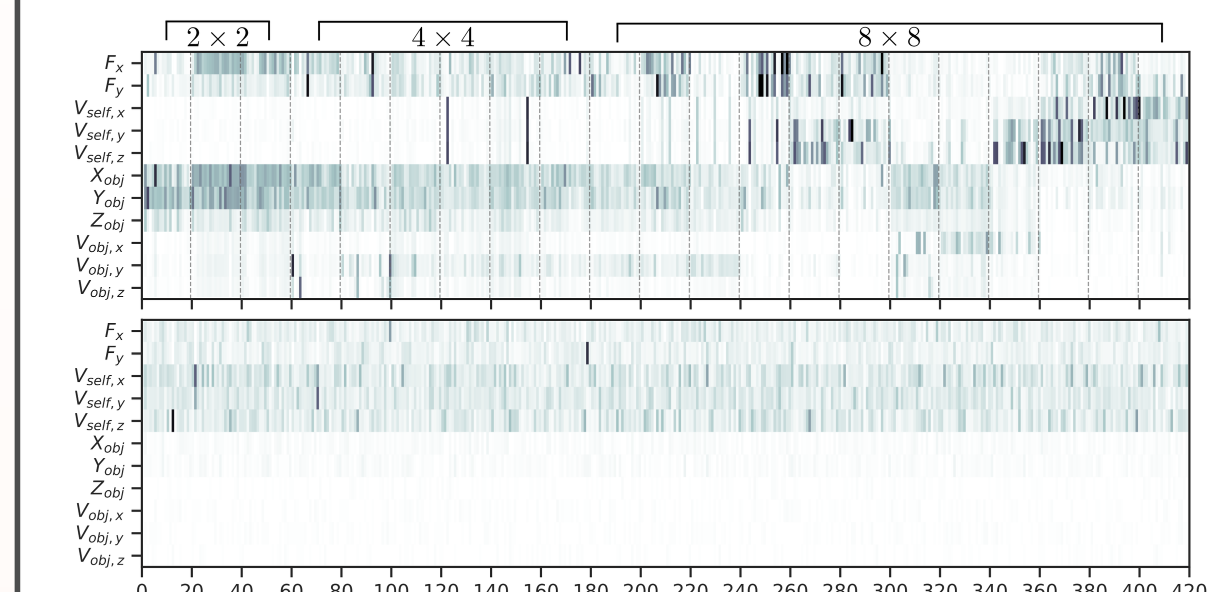
Latent code evaluation: $g \overset{?}{\longleftrightarrow} z$

Alternative models:



Result #1: Functional specialization emerges in the cNVAE latent space

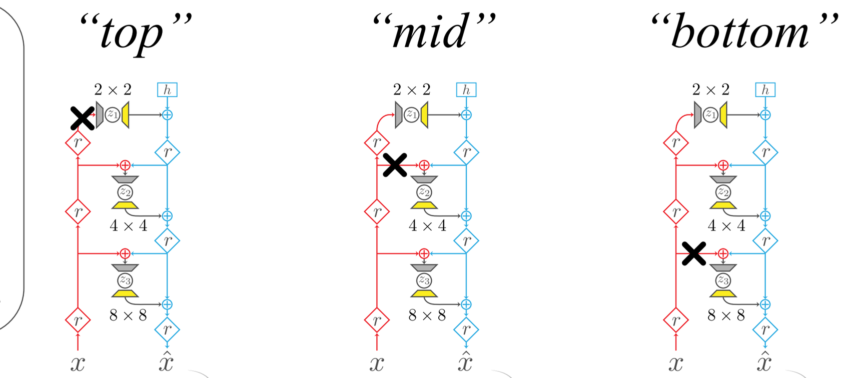
What is the information content of the learned latent codes?



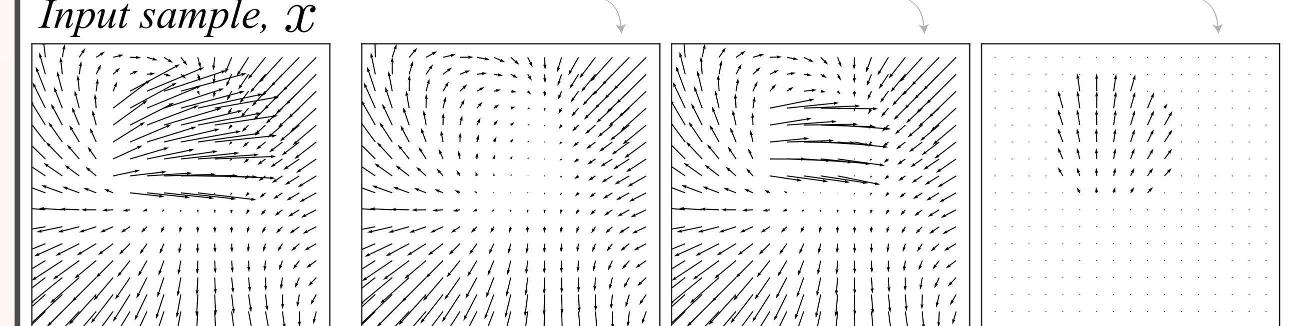
Mutual Info (MI) matrix between ground truth and latent variables : $\mathrm{MI} \in \mathbb{R}^{11\times420}$

Result #2: in silico "lesion" experiments confirm functional specialization in the cNVAE

1. Push sample through the network
2. Zero-out a given set of connections
3. Examine the reconstructed output

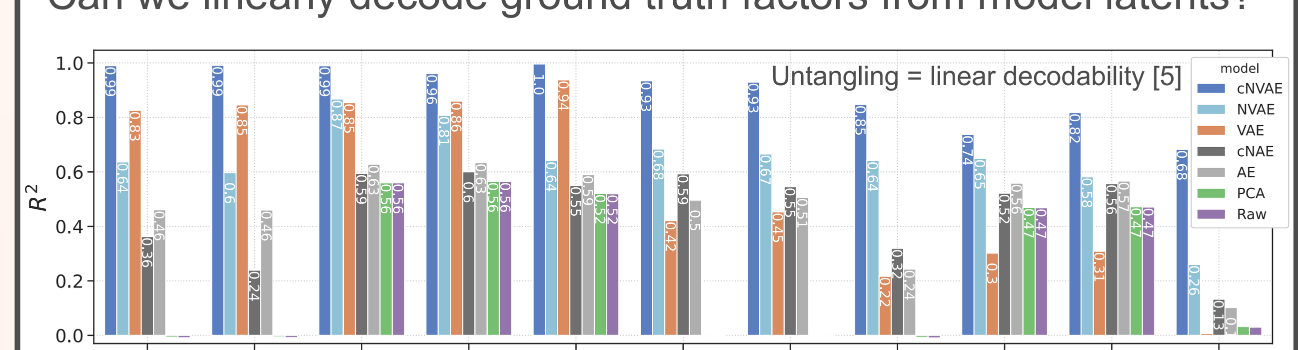


Input sample, x



Result #3: cNVAE "untangles" ground truth factors

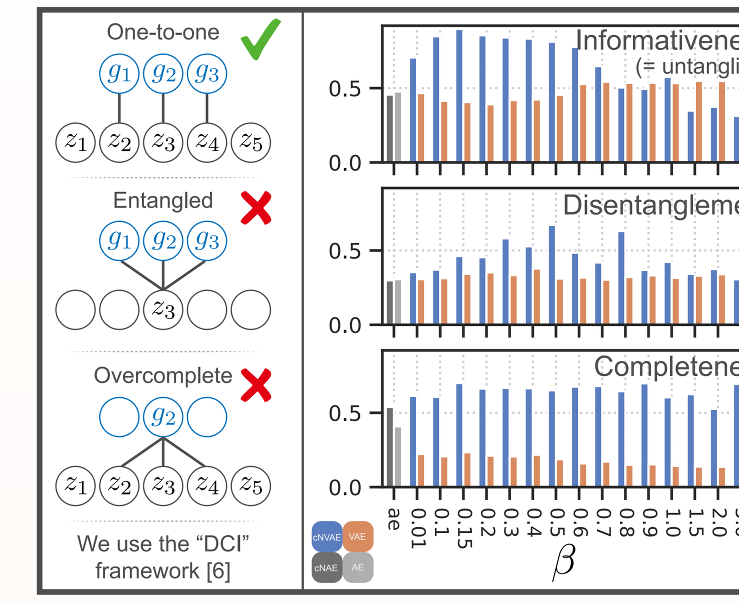
Can we linearly decode ground truth factors from model latents?



Linear regression $g = W^T z + b$

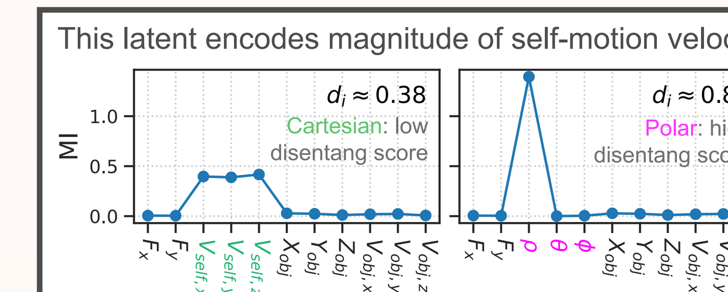
The cNVAE achieves *"optic flow parsing"*

Result #4: cNVAE "disentangles" the ground truth factors (g)



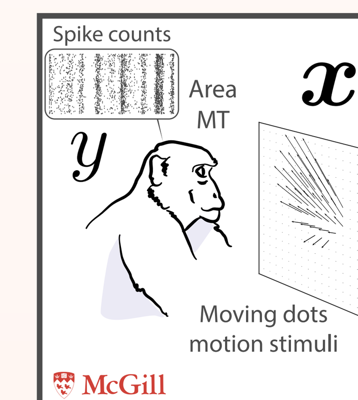
Bonus: disentanglement is in the eyes of the beholder

This latent encodes magnitude of self-motion velocity



Cartesian or polar? Which option is the "correct" choice of parametrizing ground truth factors?

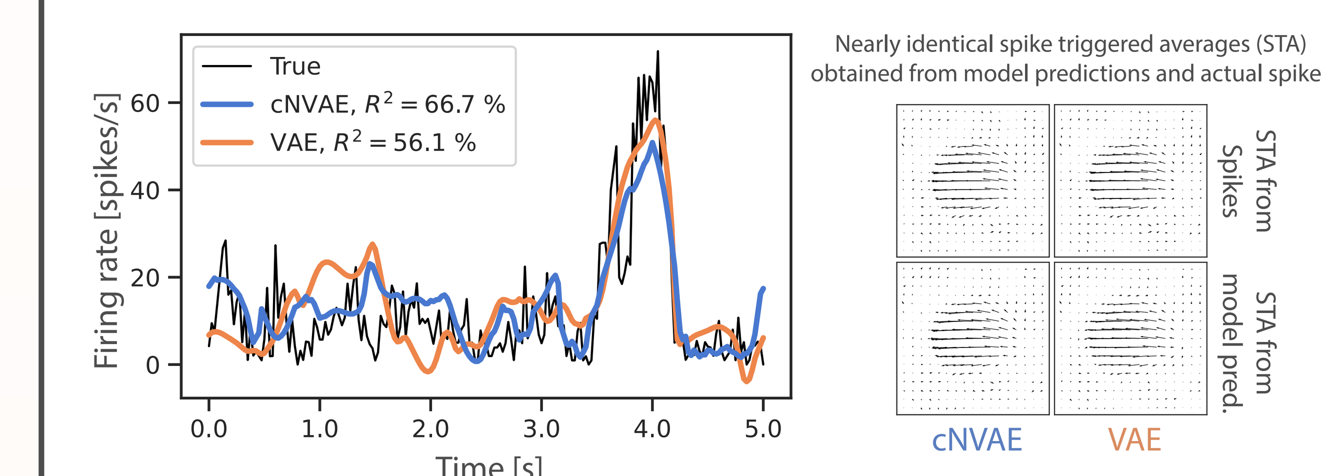
Evaluating models based on their alignment to the brain



1. Extract representations $z = enc(x)$
2. Ridge regression $\hat{y} = W^T z + b$
3. Report (Pearson R) $R = Corr(y, \hat{y})$

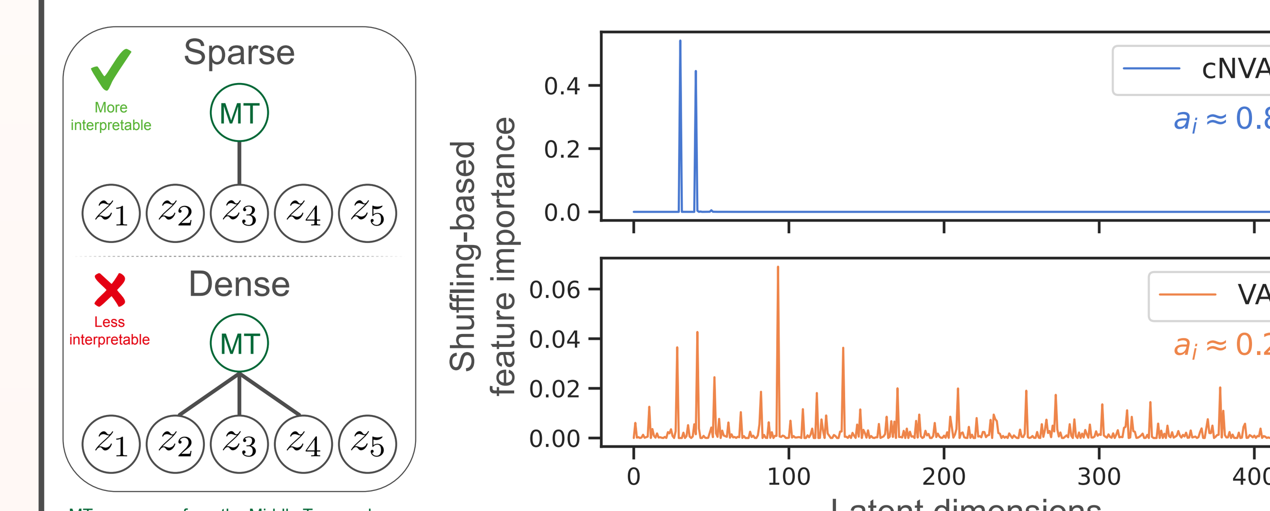
Result #5: cNVAE representations are more aligned to the brain compared to non-hierarchical VAE

Both hierarchical and non-hierarchical models predict MT responses well:

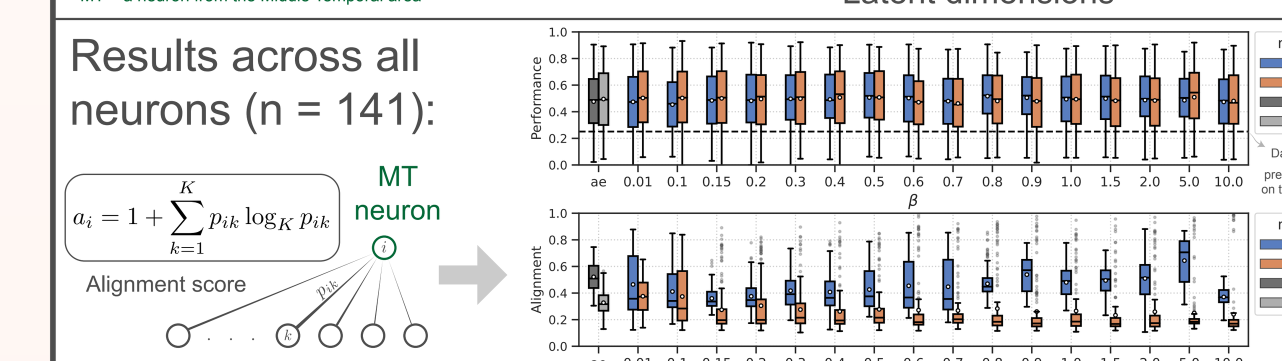


Does this mean both models are equally aligned with the brain?

We consider "sparsity of latent-to-neuron relationships" as a complementary measure of brain alignment [9]:



Results across all neurons (n = 141):



$a_i = 1 + \sum_{k=1}^{K} p_{ik} \log_K p_{ik}$

Conclusions and future work

- We introduced a new synthetic data framework (**ROFL**) and a new hierarchical VAE architecture (**cNVAE**).
- Results show that "**hierarchical latent structure**" enhances the learned representations in multiple ways:
  1. It improves the linear decodability of ground truth factors and does so in a sparse and disentangled manner.
  2. It promotes sparsity in latent-to-neuron relationships, thus increasing brain alignment and interpretability.
- Overall, our work demonstrates the power of "**synthetic data framework**" in representation learning.
- Future work: study the geometry of representations to further elucidate the role of hierarchical inference.
- We also need a better (more objective) definition of "*disentanglement*" (see the **Bonus** point above).

References

[1] Mumford, Biological Cybernetics (1992)
[2] Lee & Mumford, JOSAA (2003)
[3] Helmholtz, Treatise on Physio. Optics (1860s)
[4] Vahdat & Kautz, NeurIPS (2020)
[5] DiCarlo & Cox, Trends in Cog. Sci. (2007)
[6] Eastwood & Williams, ICLR (2018)
[7] Cui et al., Journal of Neuroscience (2013)
[8] Cui et al., Journal of Neuroscience (2016)
[9] Higgins et al., Nature Communications (2021)
[10] Mineault et al., NeurIPS (2021)

Contact info

@hadivafaii

vafaii@umd.edu

Acknowledgements

NSF IIS-2113197
NSF DGE-1632976
NIH R00EY032179