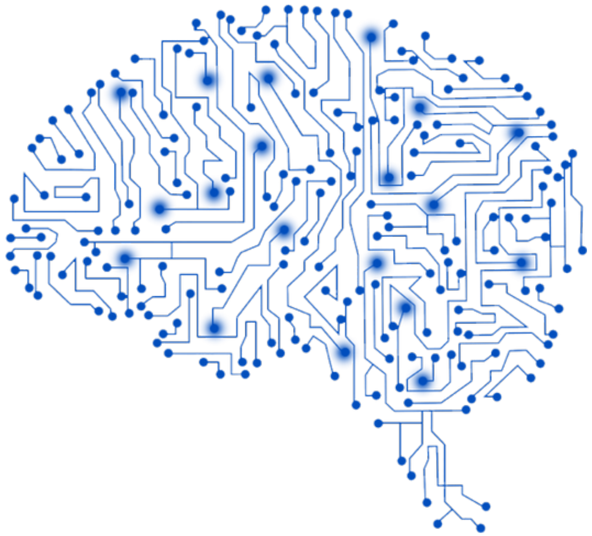
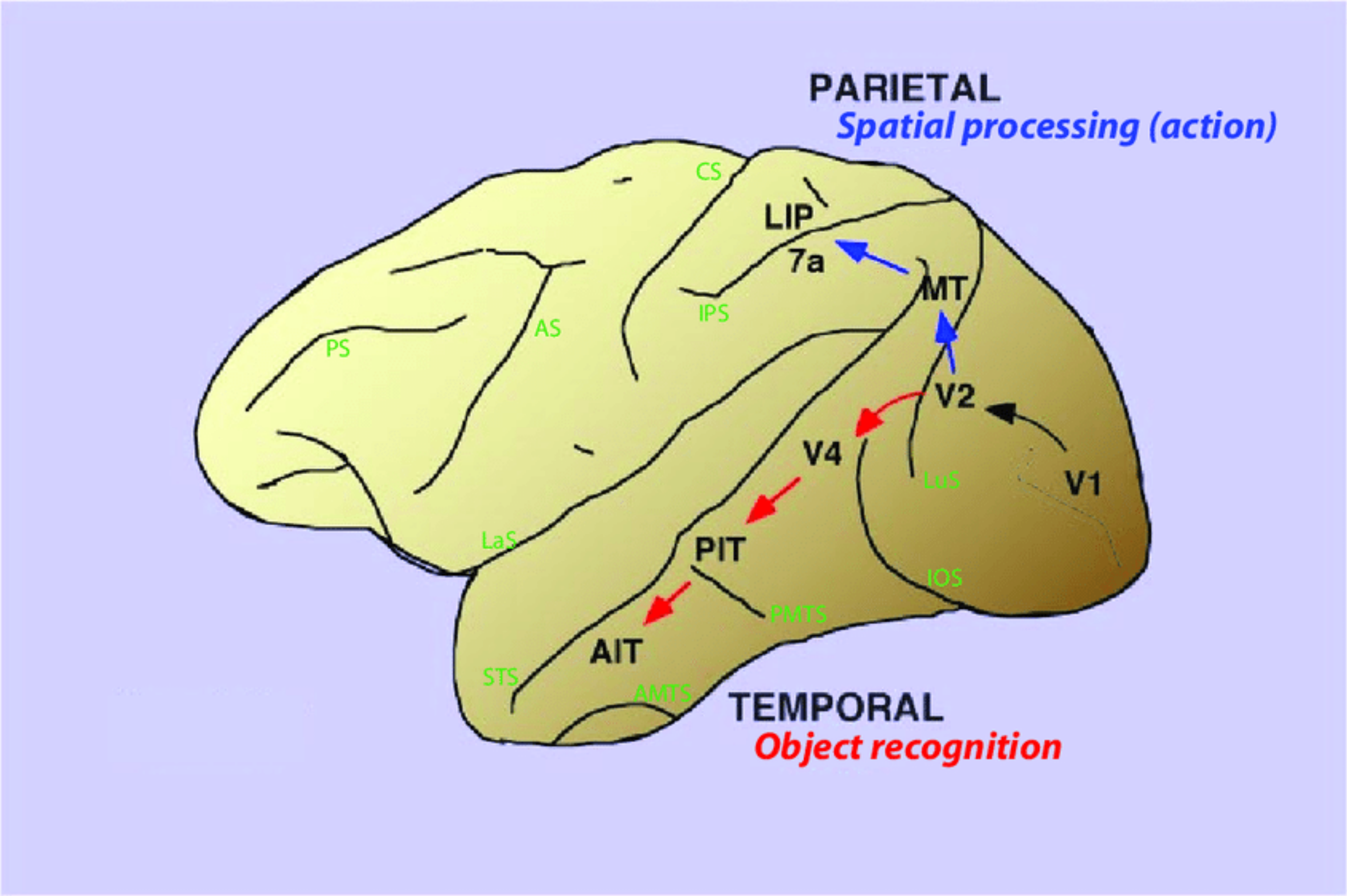
Diffusion-Based Discovery of Semantic Latent Groups in Higher Visual Cortex



Anqi Wu

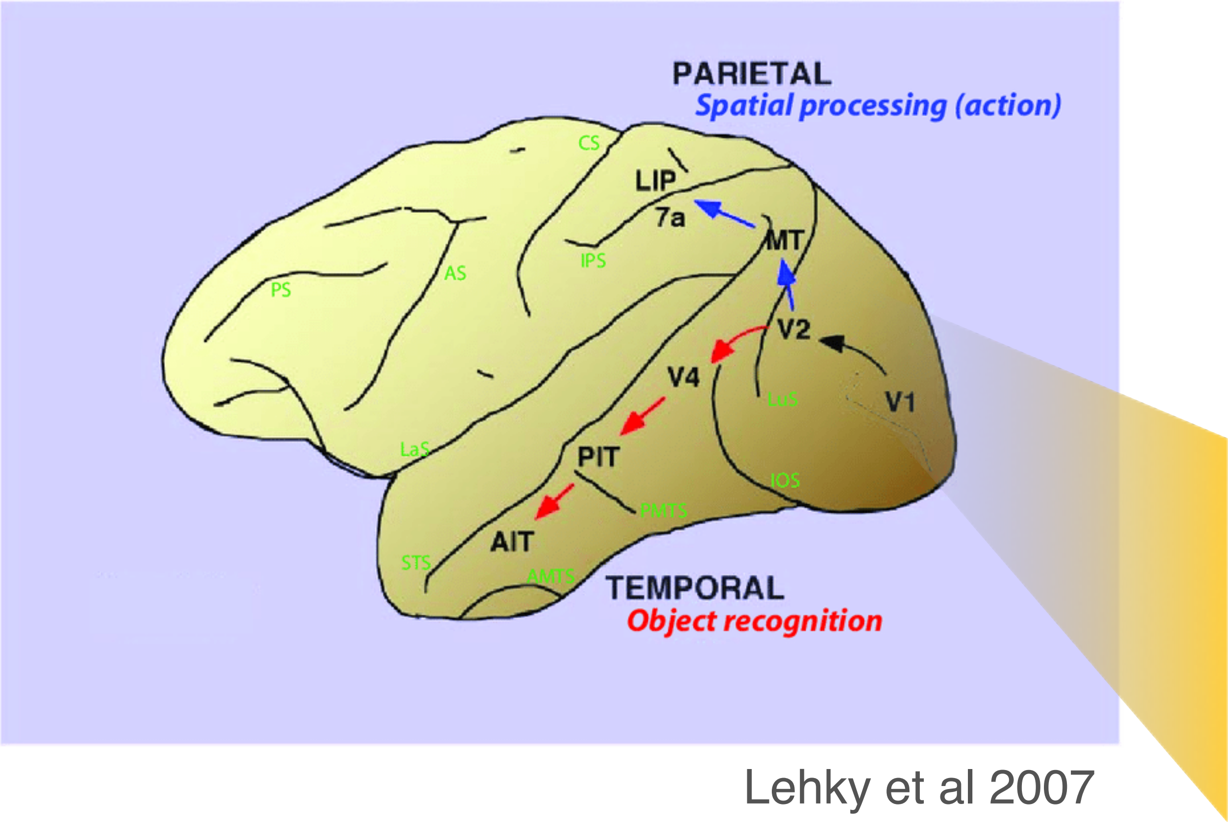
School of Computational Science and Engineering
Georgia Institute of Technology

Goal: Understanding how neural populations in higher visual areas encode object-centered visual information

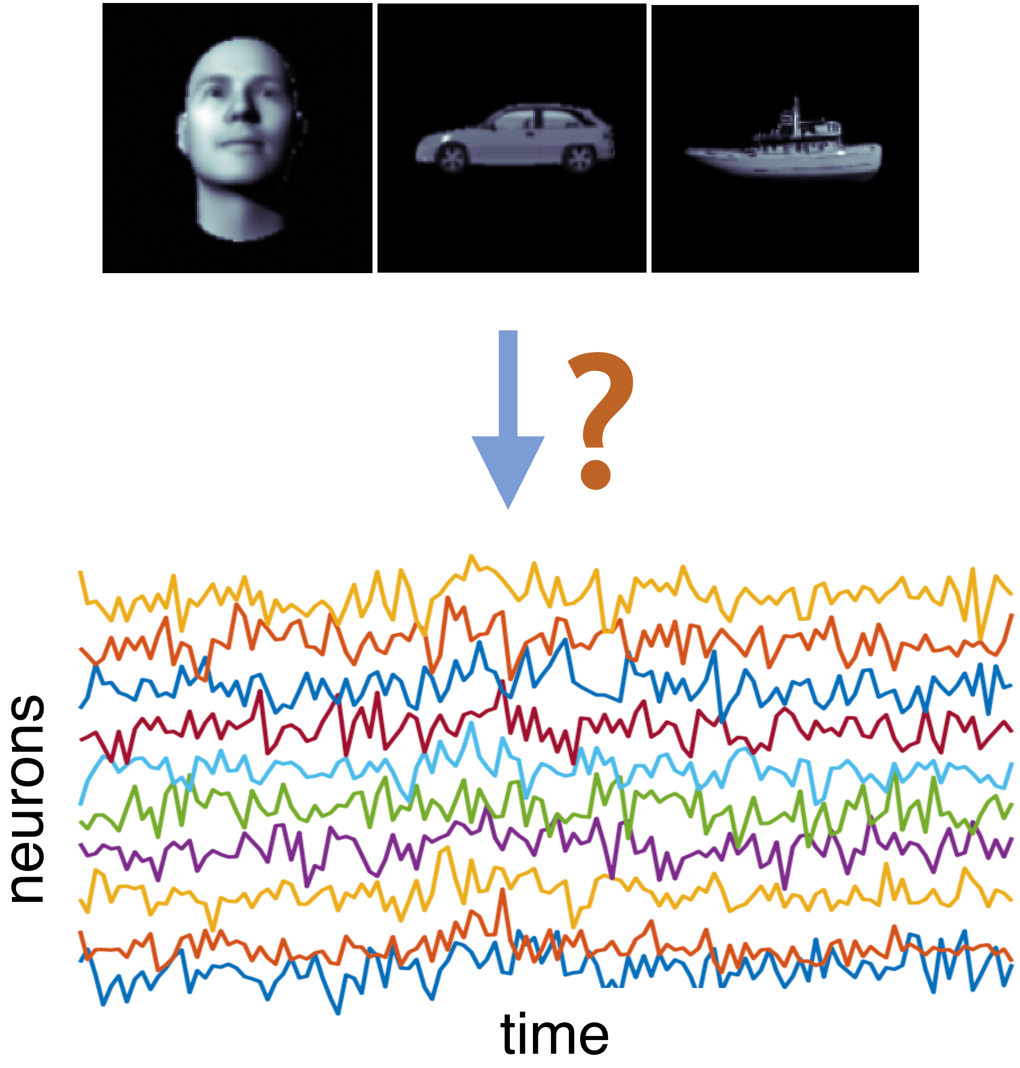


Lehky et al 2007

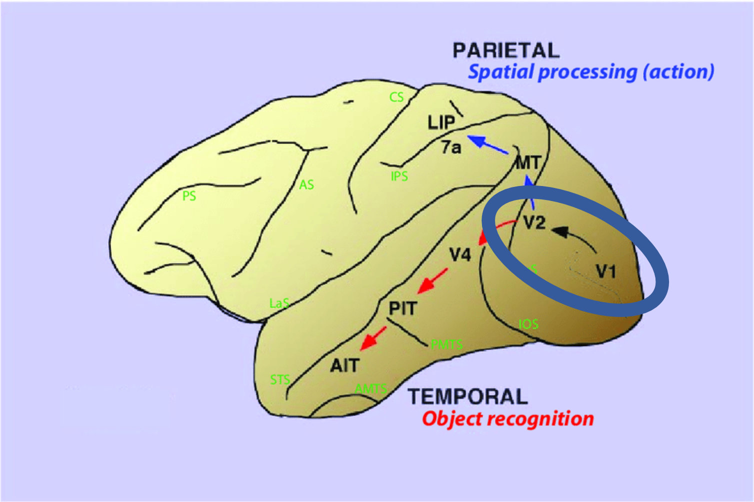
Goal: Understanding how neural populations in higher visual areas encode object-centered visual information



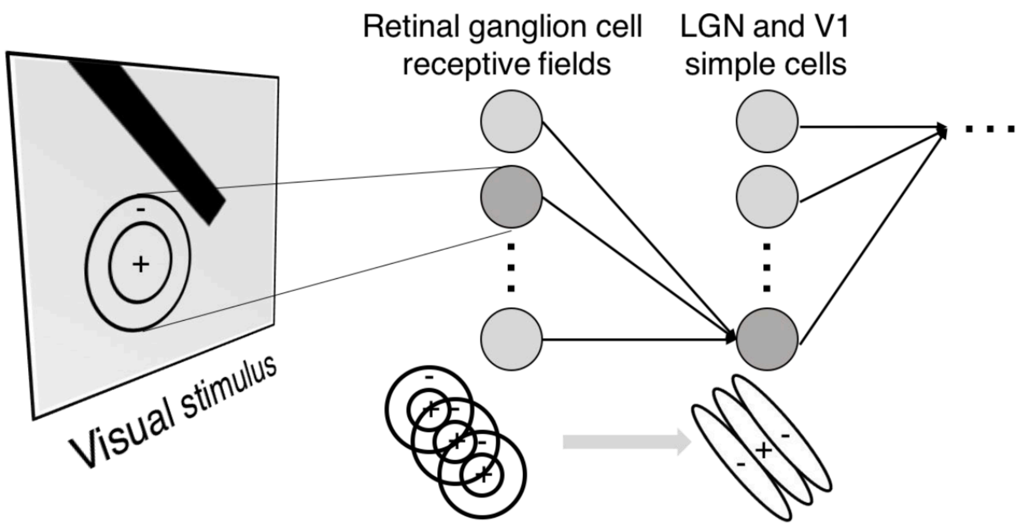
Lehky et al 2007



Goal: Understanding how neural populations in higher visual areas encode object-centered visual information

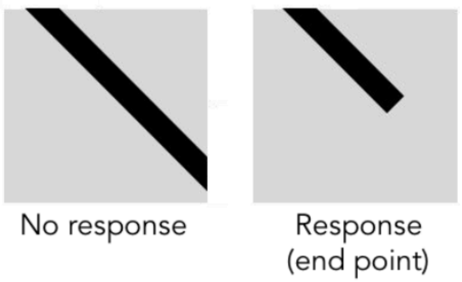


Lehky et al 2007

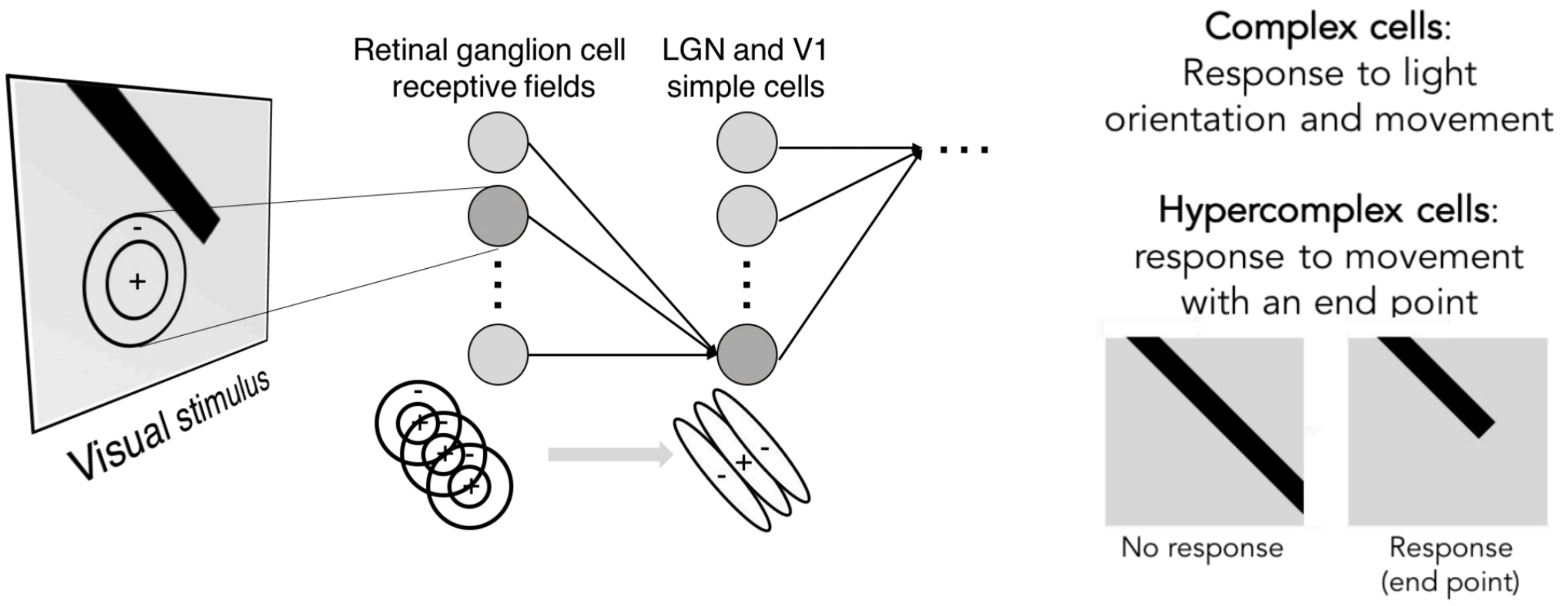


Complex cells:
Response to light orientation and movement

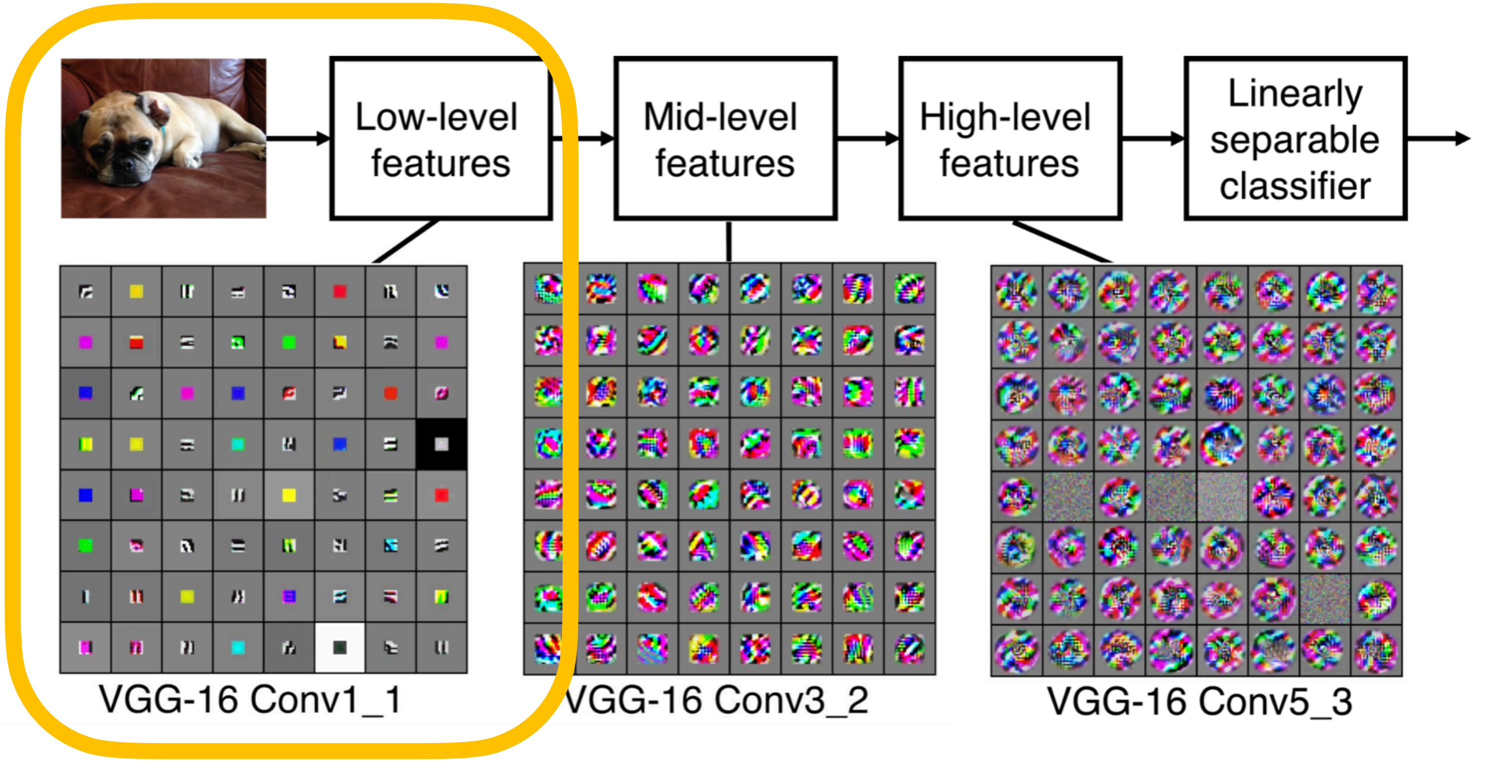
Hypercomplex cells:
response to movement with an end point



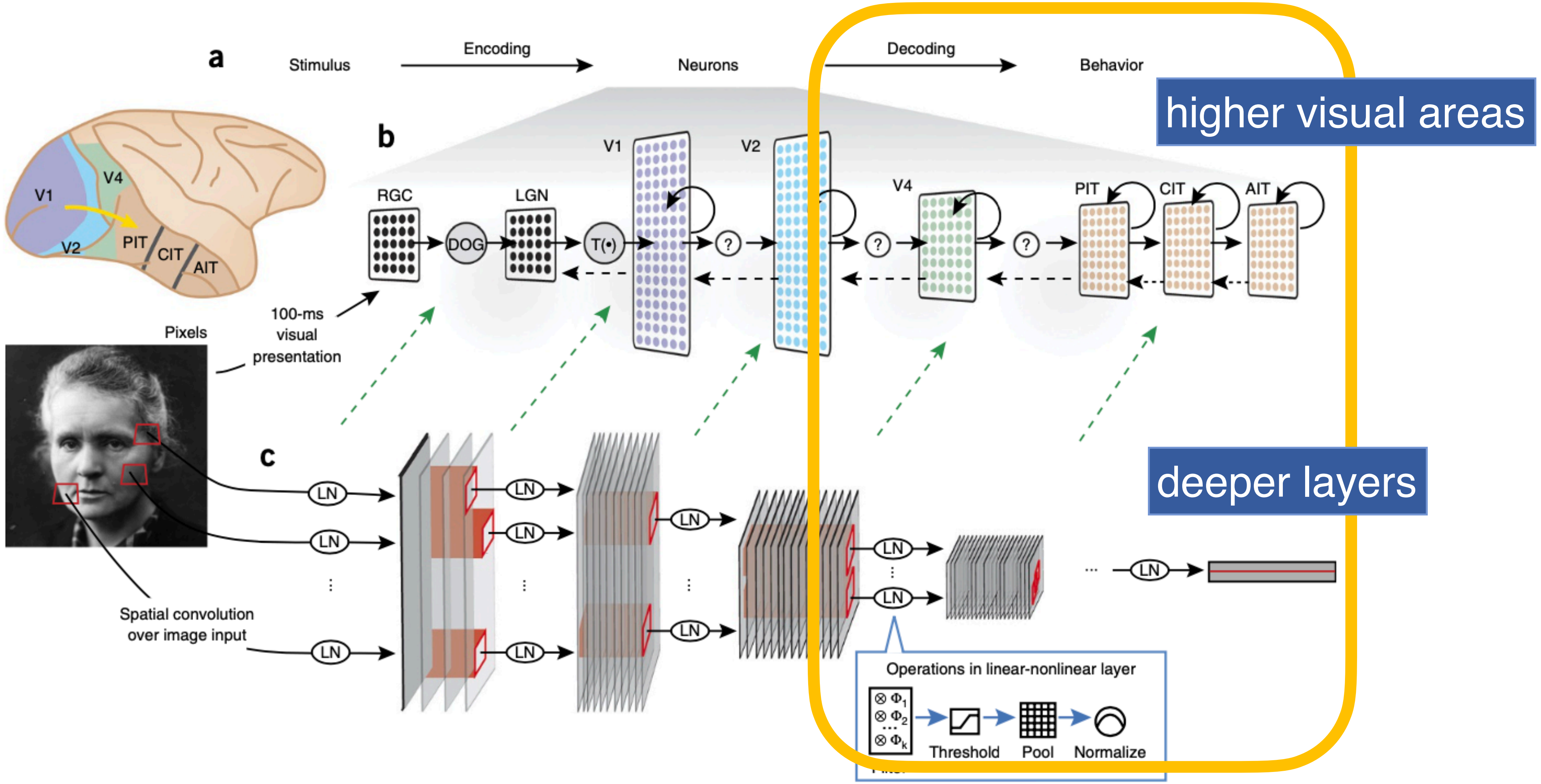
visual


Visual stimulus
Retinal ganglion cell receptive fields
LGN and V1 simple cells
Complex cells:
Response to light orientation and movement
Hypercomplex cells:
response to movement with an end point
No response
Response (end point)

CNN

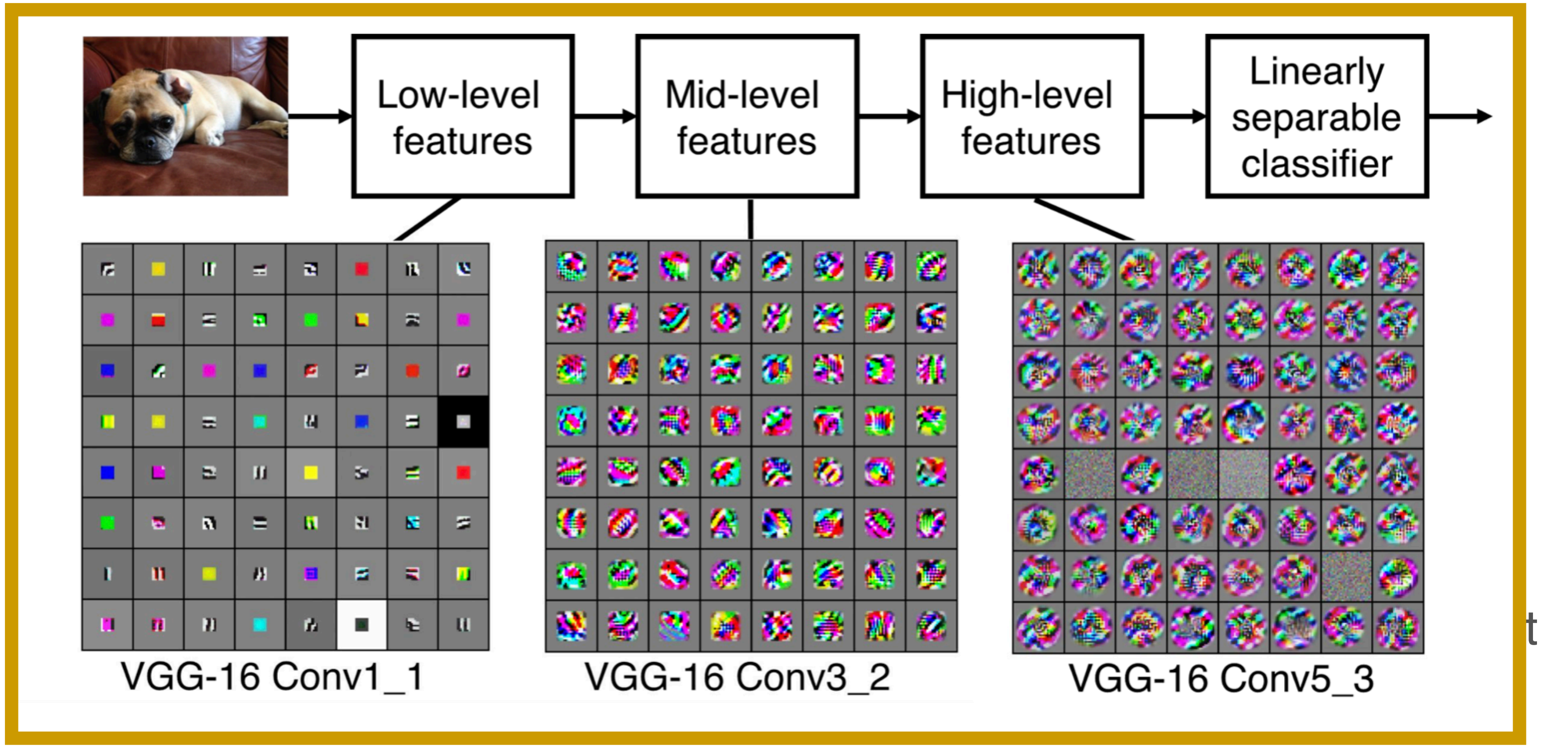

Low-level features
Mid-level features
High-level features
Linearly separable classifier
VGG-16 Conv1_1
VGG-16 Conv3_2
VGG-16 Conv5_3

Representation alignment between CNN and neurons



Yamins et al 2016

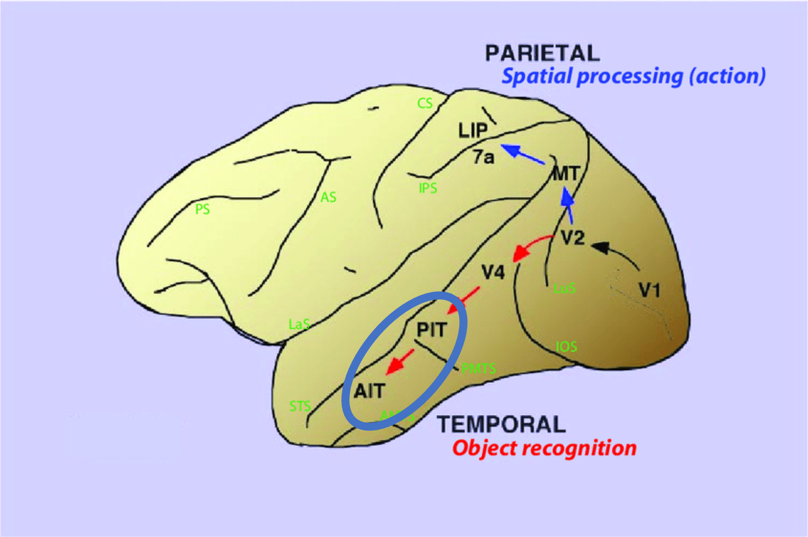
Drawbacks



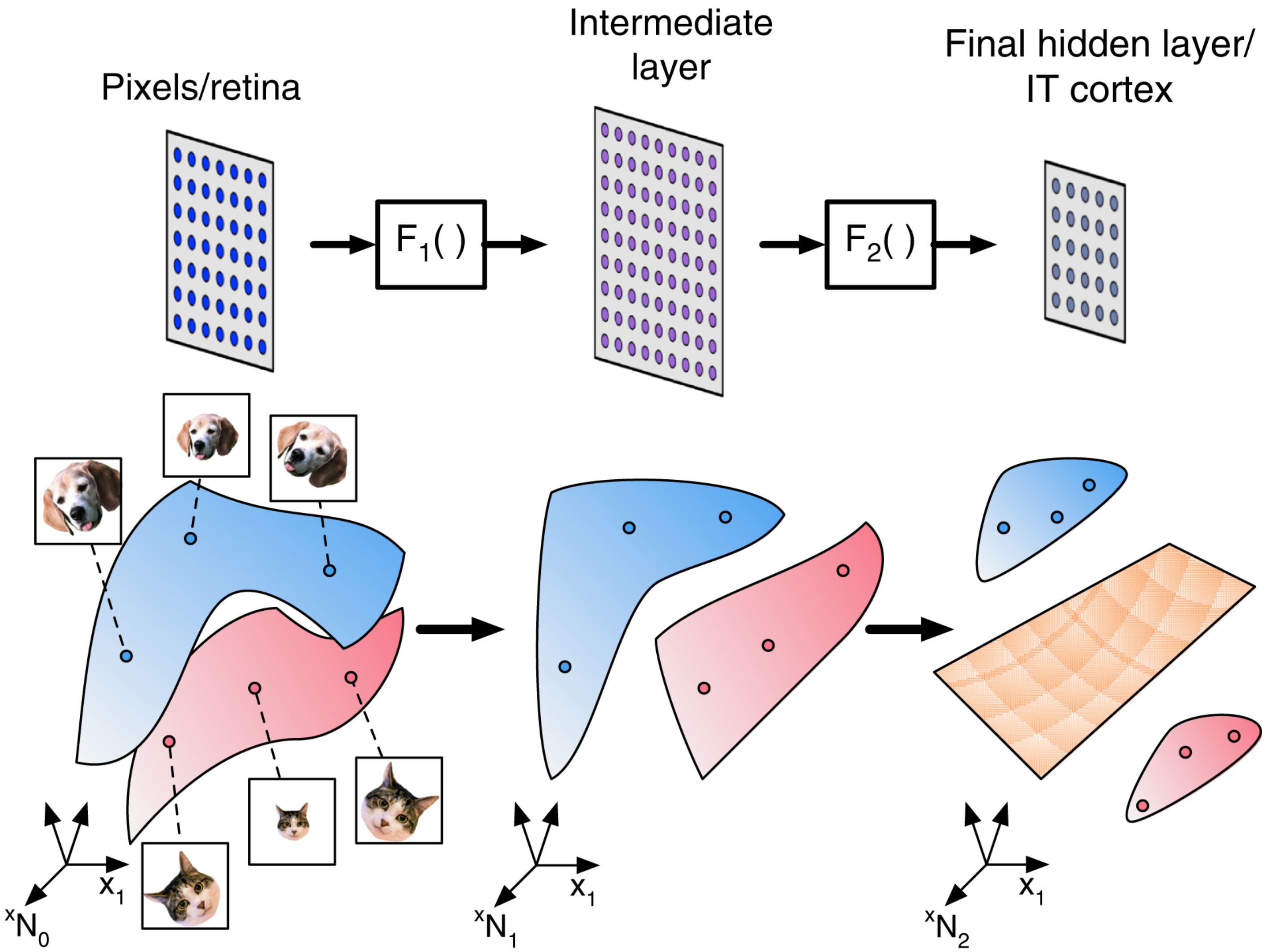
t al 2016

- Artificial neurons are not direct models of biological ones.
- Even with artificial neurons, particularly in deeper layers, interpreting what individual neurons selectively respond to remains challenging.
- Which neurons encode what: size, rotation, object shape, identity, or other semantic attributes?

Higher visual areas: IT (Inferior Temporal cortex)

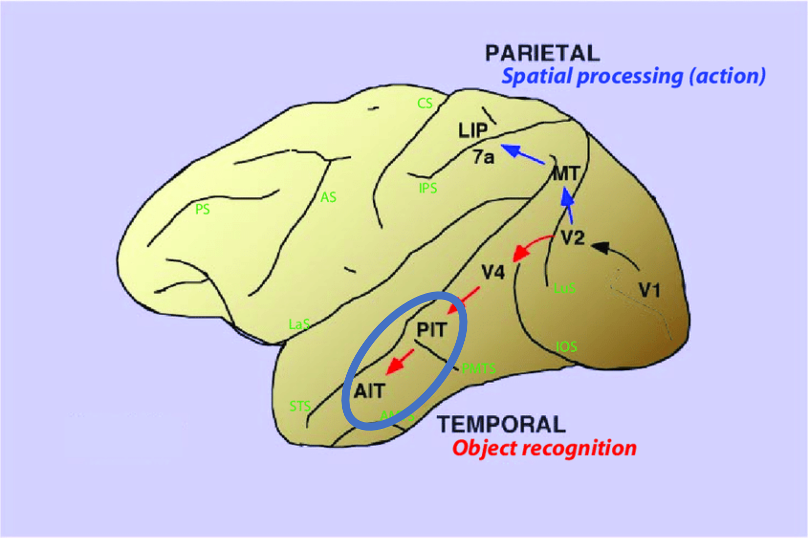


- Focuses on object identity classification, not disentangling continuous attributes (pose, lighting, texture)

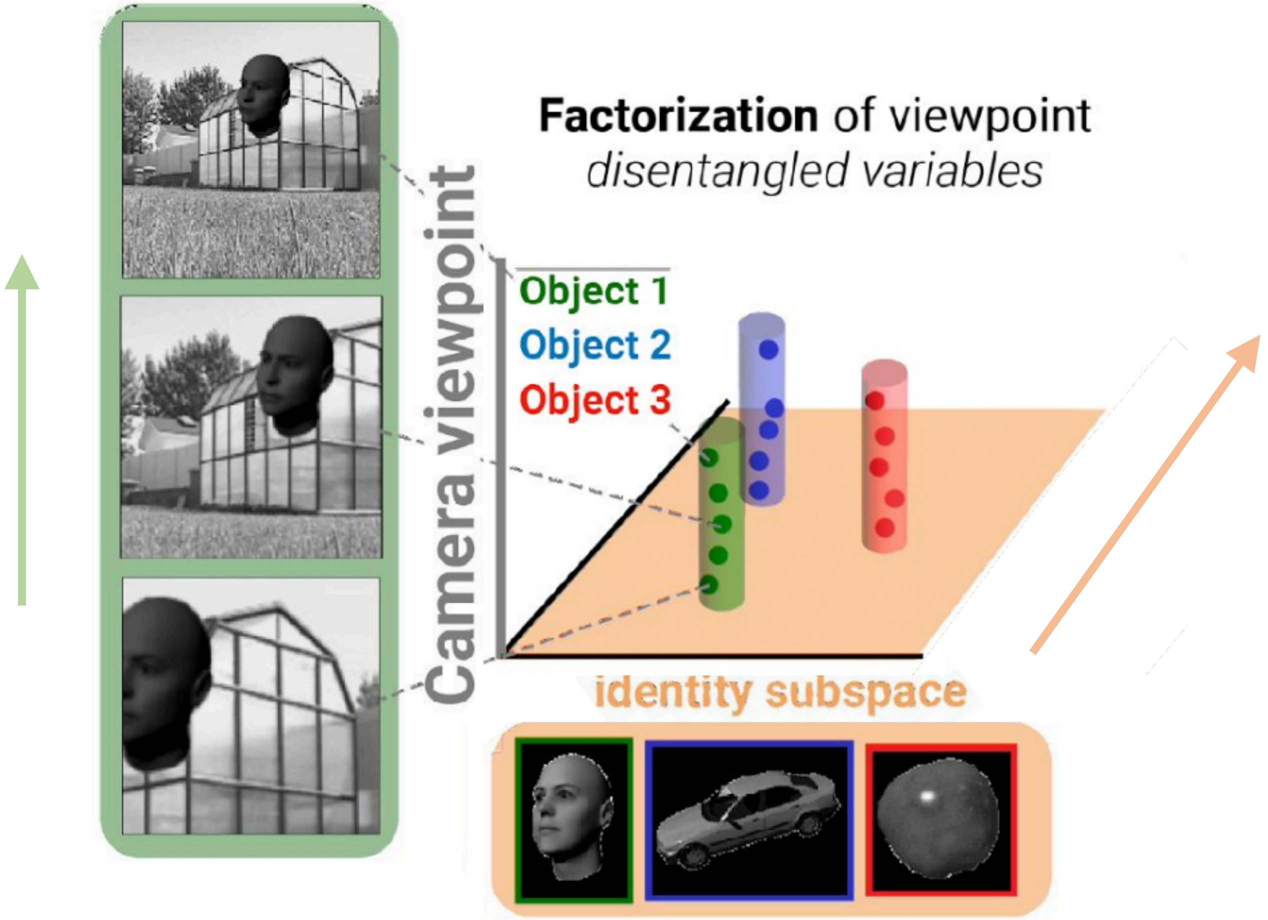


Cohen et al 2020

Higher visual areas: IT (Inferior Temporal cortex)

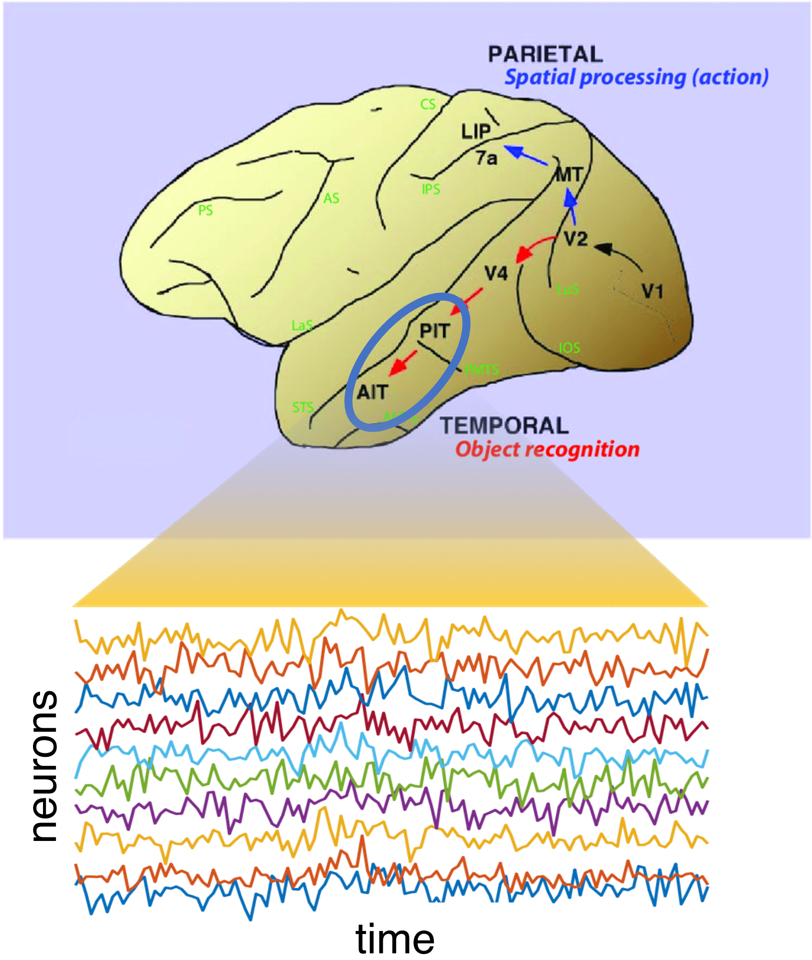


- Focuses on coarse factors (pose, viewpoint, lighting, background, identity), not richer attributes (texture, shape, semantic features).
- Relies on CNN alignment, assuming CNNs inherently disentangle representations.

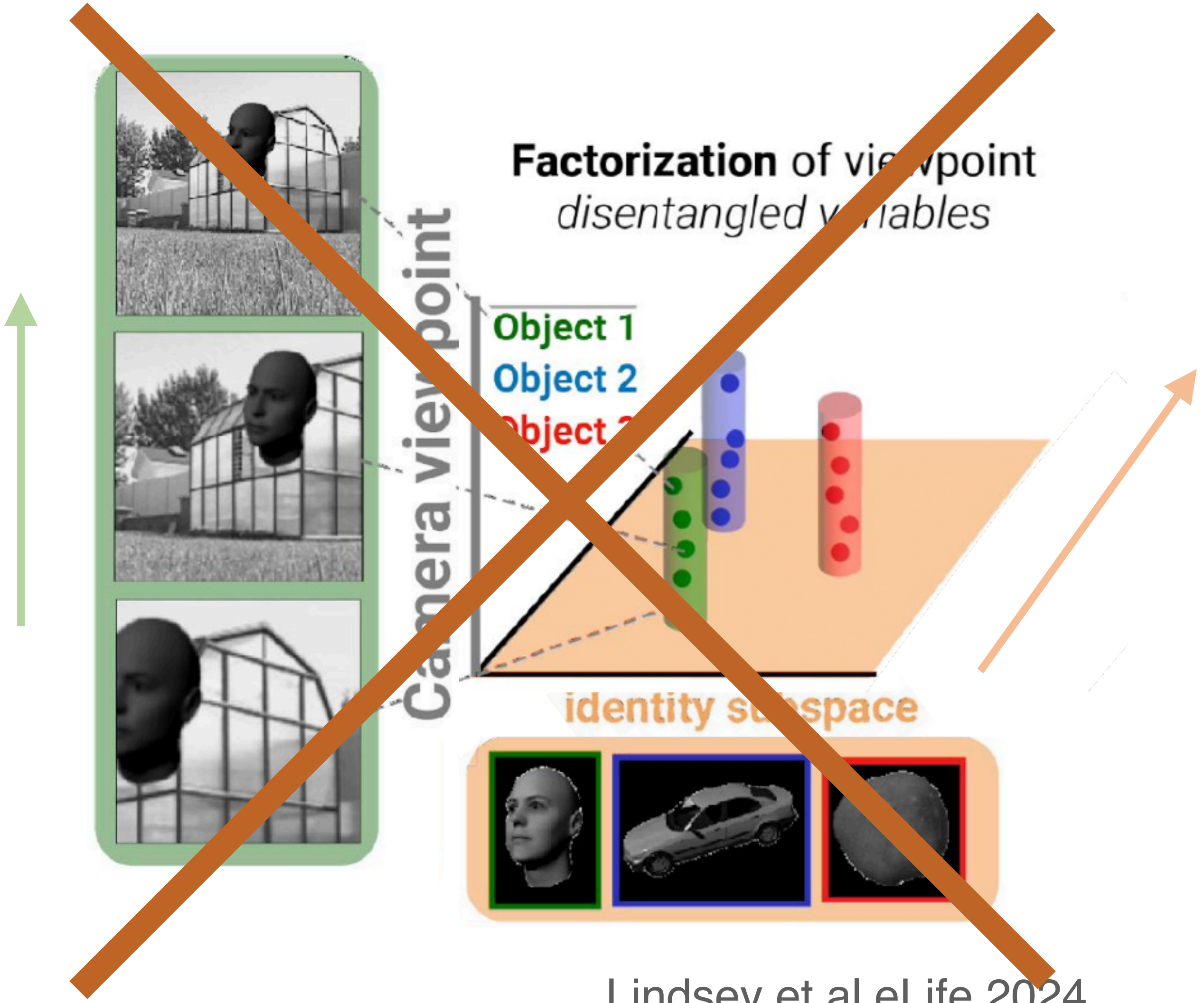


Lindsey et al eLife 2024

Our proposed idea

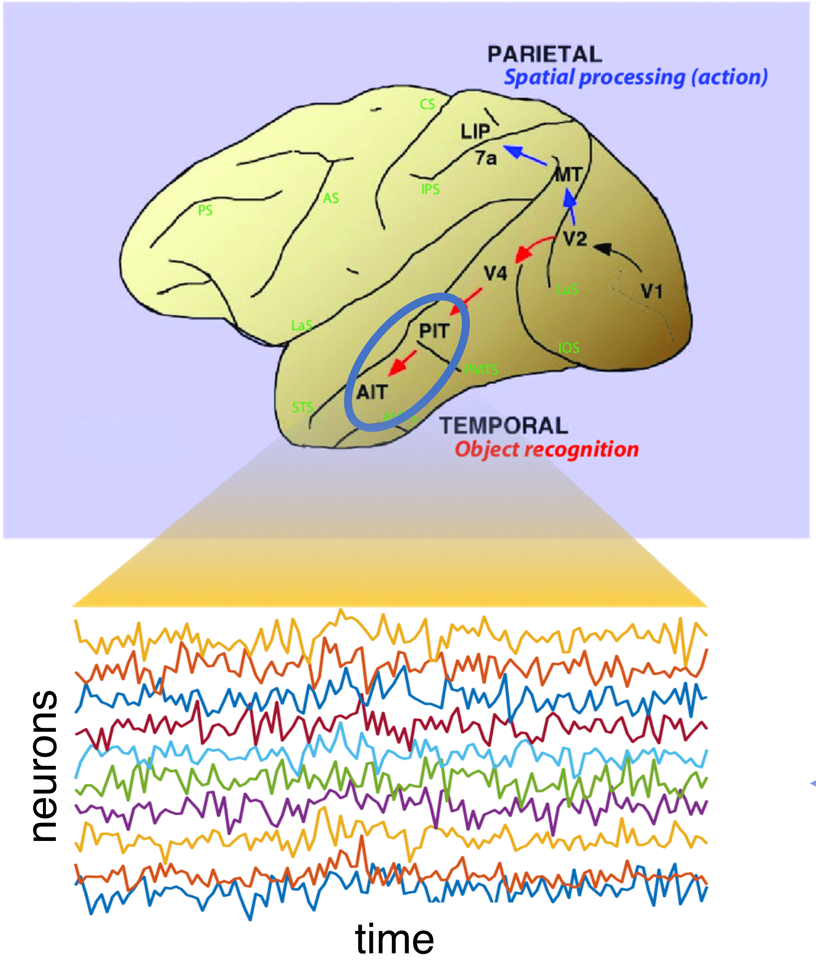


- No CNN alignment, directly analyze neural data

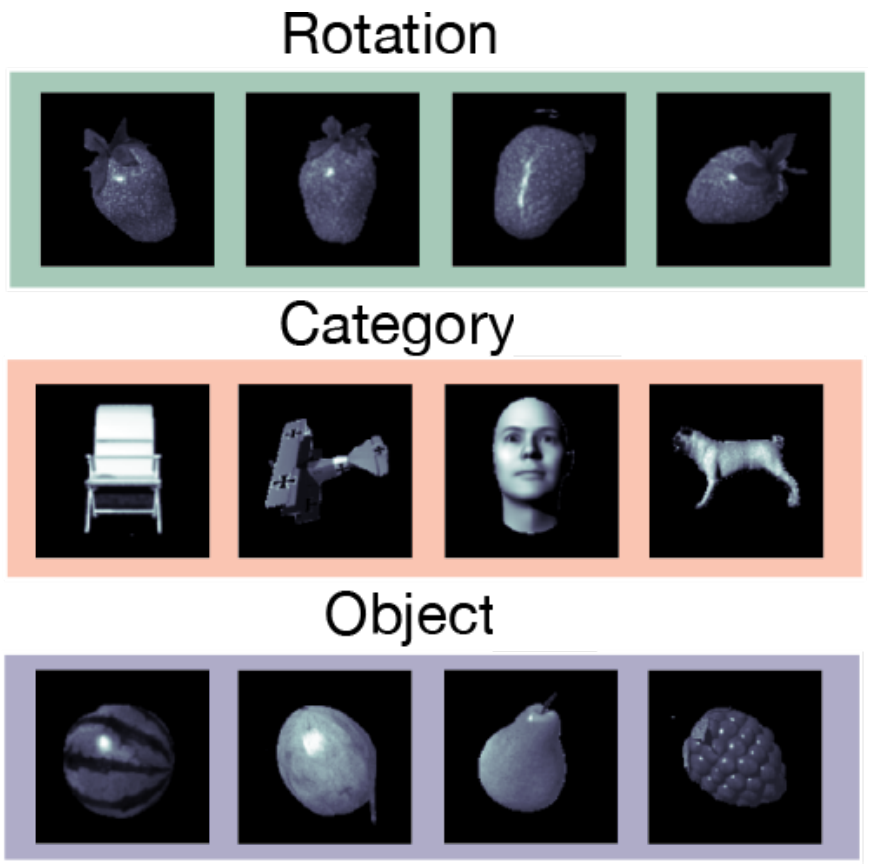


Lindsey et al eLife 2024

Our proposed idea



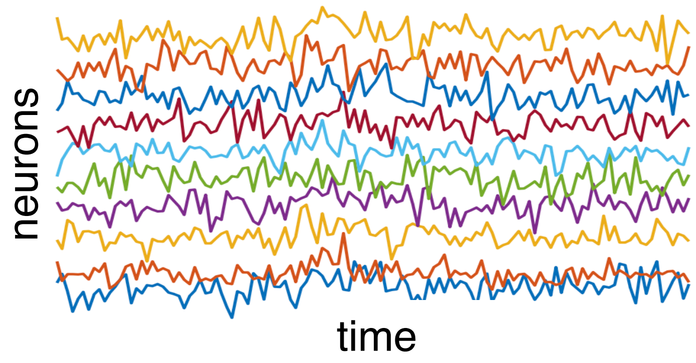
?

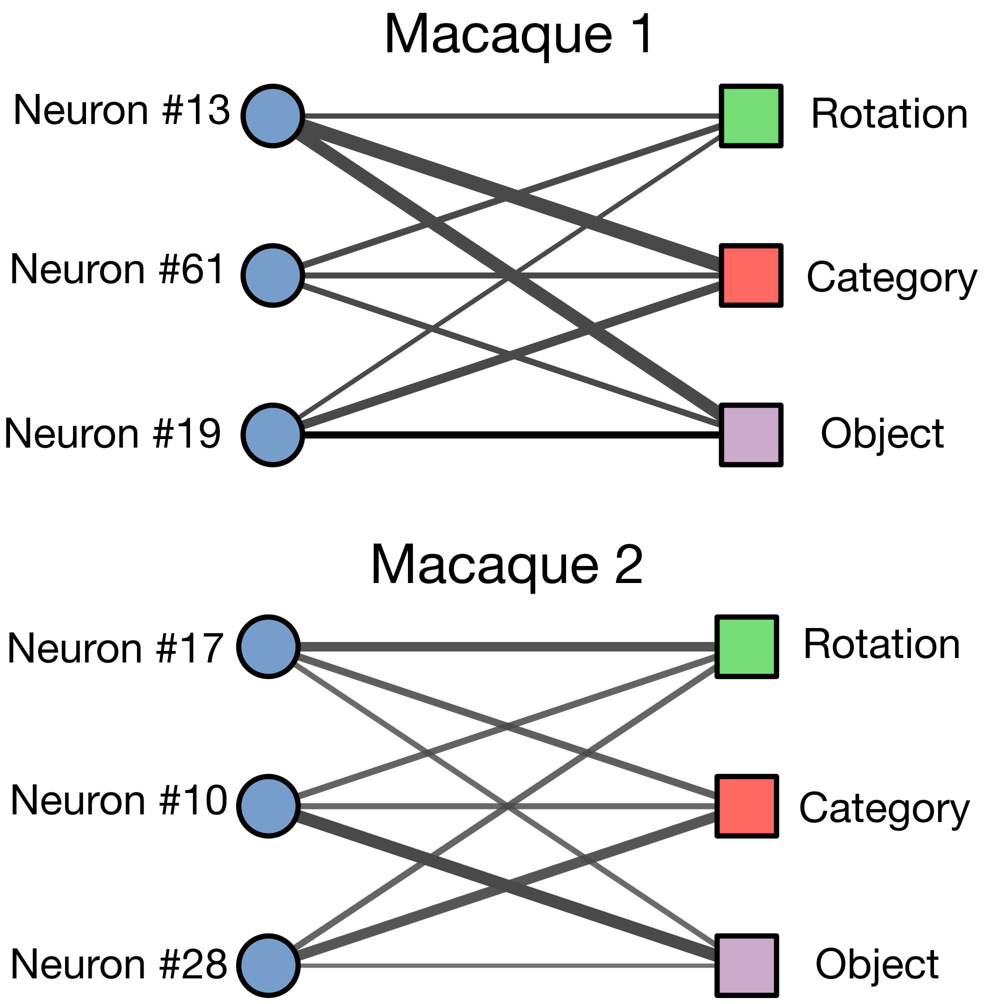


- No CNN alignment, directly analyze neural data
- Understand how IT neurons encode rich attributes spanning geometry, category, and object identity

Mixed selectivity

neural space



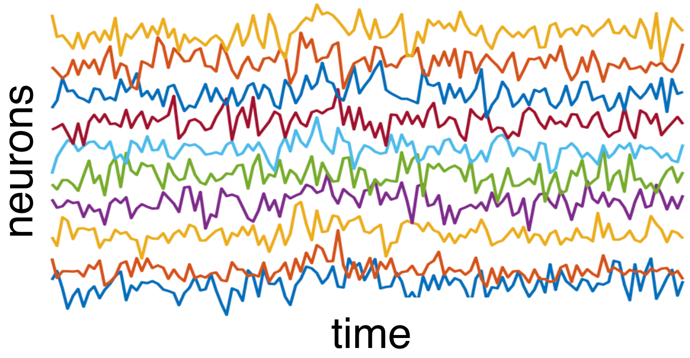


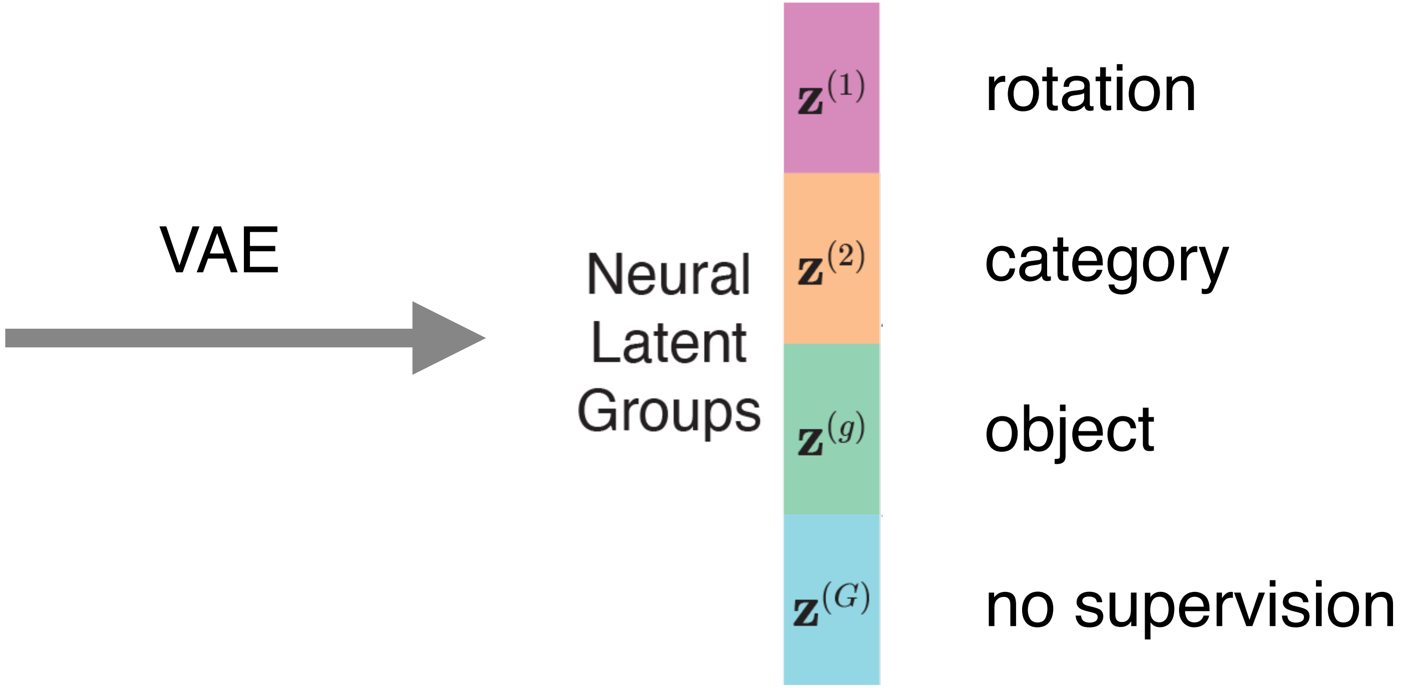
- We cannot directly understand attribute encoding from raw neural activity.

Generative AI Framework: Variational Autoencoder

neural space

disentangled latent space

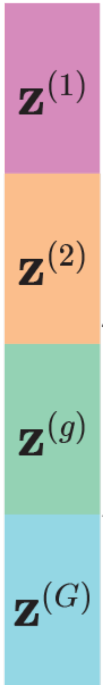




Generative AI Framework: Variational Autoencoder

disentangled latent space

Neural Latent Groups



rotation

category

object

no supervision

Still we don't know:

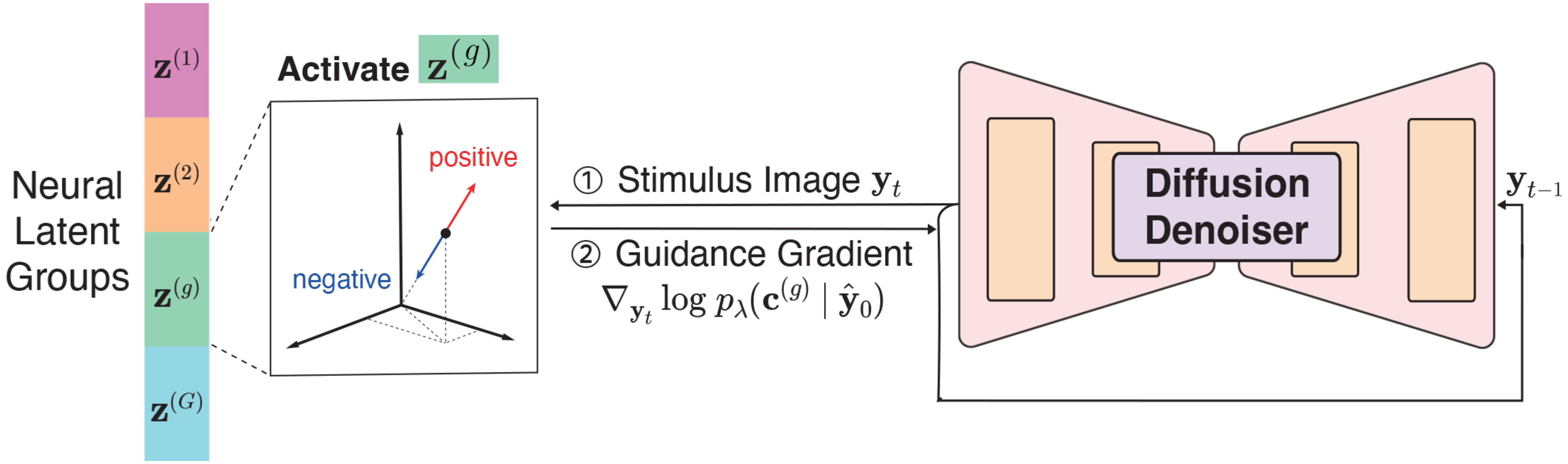
- What information is encoded in the **latent group** without supervision?
- Within each group, what individual latent dimensions represent or encode?
- Can they capture semantic attributes beyond category or object identity labels?
- E.g.: color, texture, size, shape, and semantic features defining objects like human faces or fruits

For example, rather than just claiming the entire latent group as representing a single category like "human face," we aim to identify subspaces encoding specific face attributes (e.g., gaze direction, facial structure, head shape).

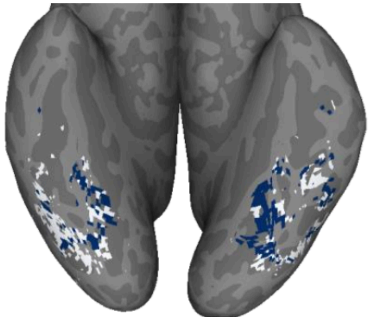
Generative AI Framework: Diffusion

disentangled latent space

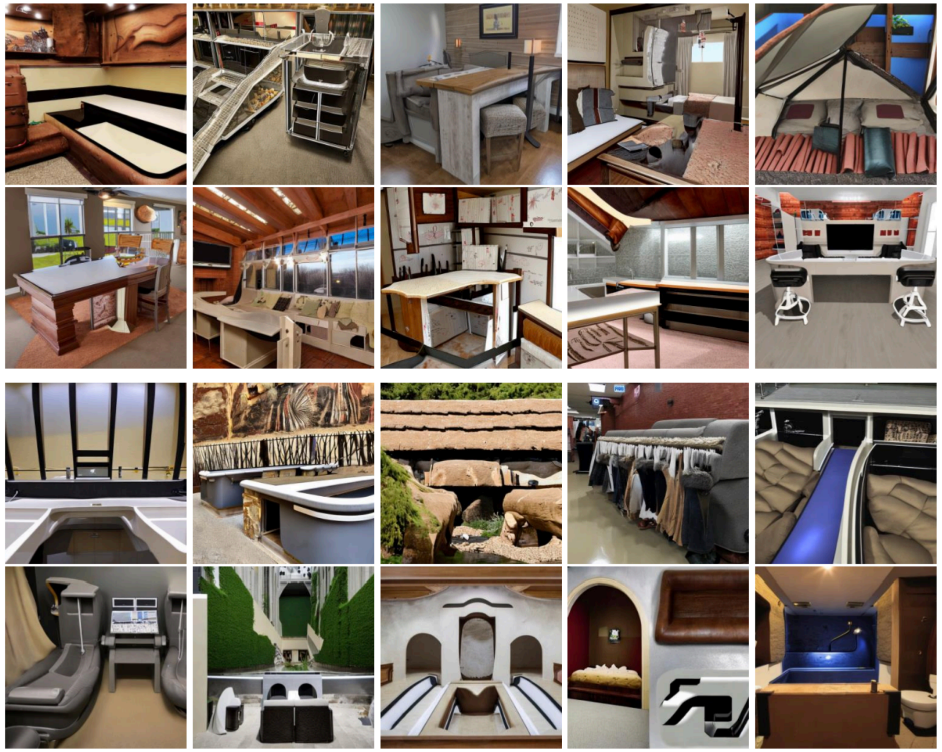
classifier-guided diffusion

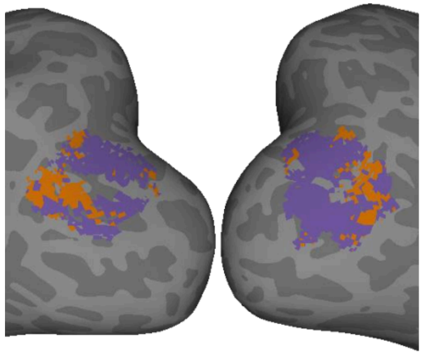


Diffusion-based approaches for probing neural encoding

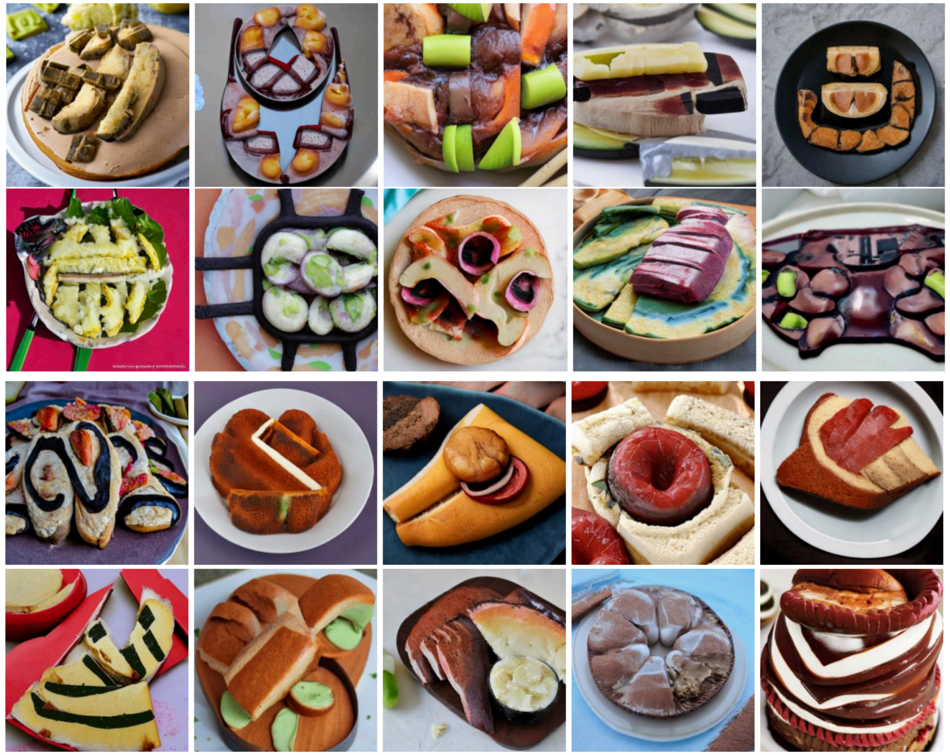


Scene-selective ROI



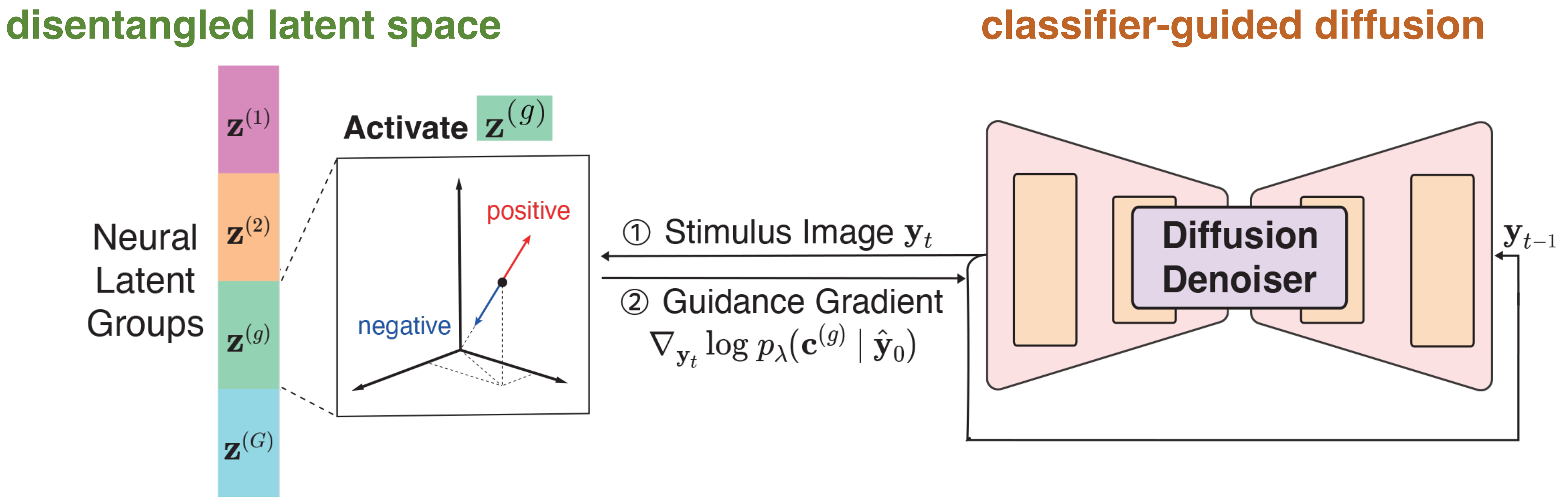


food-selective ROI



Luo et al 2023

Generative AI Framework: Diffusion

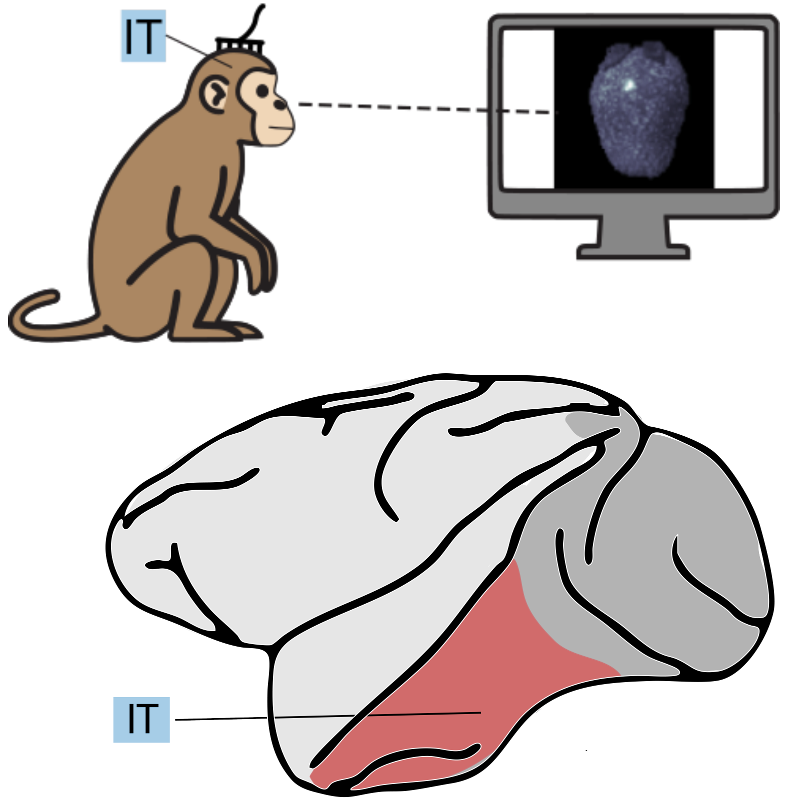


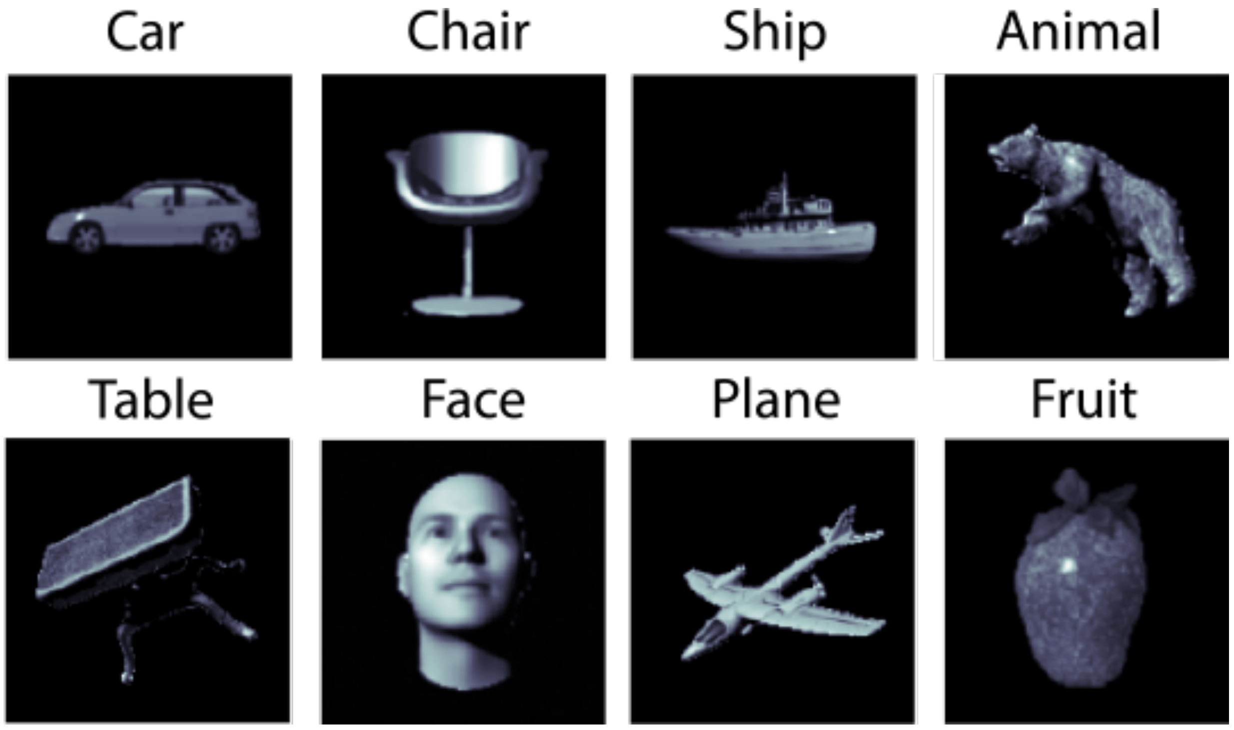
Our novelty:

- Unlike prior work that manipulates neural space for guidance, we manipulate the latent space directly.
- Introduce a new way to induce guidance from latent representations.

Evaluate the framework using a public IT cortex dataset

- It has single-unit spiking responses from the IT cortex of two macaques (M1, M2)
- Neural activity was recorded from 110 channels in M1 and 58 channels in M2





Generative AI framework for neural representation discovery in IT

Approach:

- **Disentangled VAE** → isolates latent groups in neural data
- With **Diffusion Model** → probes and visualizes semantic content of each latent group via image synthesis

Innovation:

- First to use diffusion-based generative probing of latent neural subspaces from electrophysiology
- Provides semantic interpretability beyond feature decoding

Scientific Insight:

- Uncovers structured, disentangled neural codes in higher visual areas
- Bridges population activity with geometric and semantic attributes in naturalistic vision

Acknowledgement

BRAINML

Microsoft

Institute for Data Engineering and Science (IDEaS)

GenAI for science seed grant

NIH National Institutes of Health

NSF National Science Foundation

ALFRED P. SLOAN FOUNDATION

DARPA

GT

Backup Slides

Our proposed idea