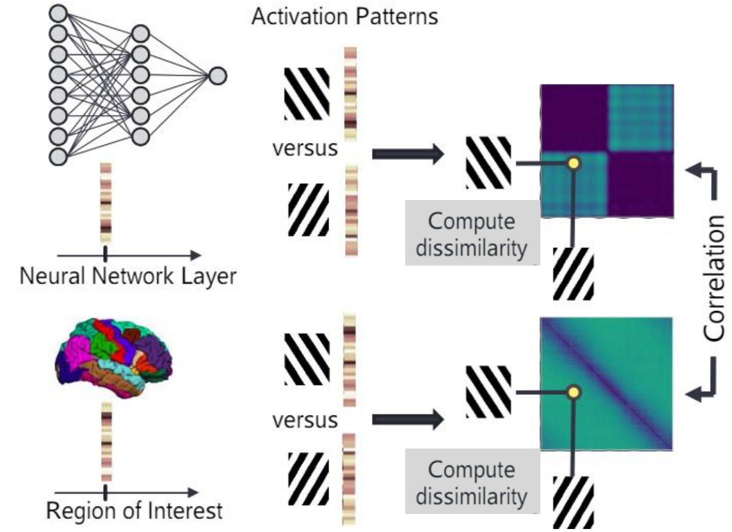
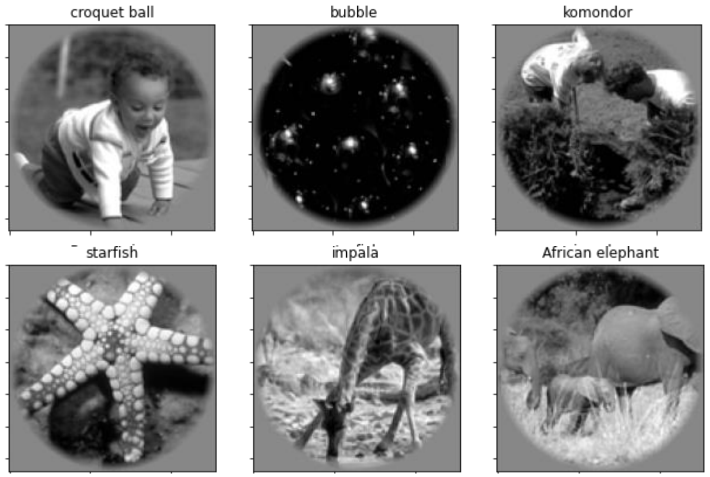
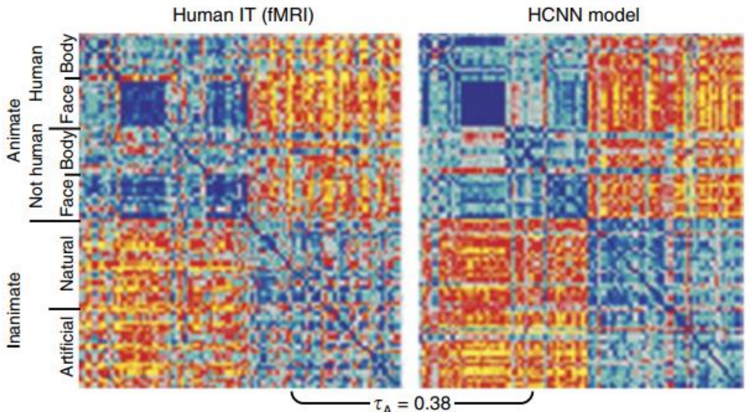
# How to understand the complex pattern in high-level brain areas?

- Background:
  - Neuron activity in high-level brain areas show complex pattern (M Rigotti et al, 2013), which can't be easily decoded by Gabor filter like early visual areas (Kay et al. 2008).
- Methods: [top right]
  - Use RSA to map these spatial activity patterns to representational geometry (N Kriegeskorte & J Diedrichsen, 2019) to characterize these complex pattern.
  - Compare neural network and brain to see how these stimuli are encoded respectively.
- Dataset: Kay/Gallant [middle right]
  - Subjects passively watch the stimuli in fMRI scanner.
  - We label the stimuli into 4 categories by hand: animate-animals, animate-human, inanimate-artificial, inanimate-natural (Khaligh-Razavi & Kriegeskorte 2014).
  - We assume all stimuli in the same category elicit a prototypical response pattern, which implies small dissimilarities between within-category representations.
- Hypothesis: [bottom right] (extracted from Yamins & DiCarlo, 2016)
  - The representations of stimuli are gradually decoupled from low layers to high layers.
  - The representations cluster into 4 categories (show small distance in representational geometry) in high layers of the brain and neural network, but not in low layers.

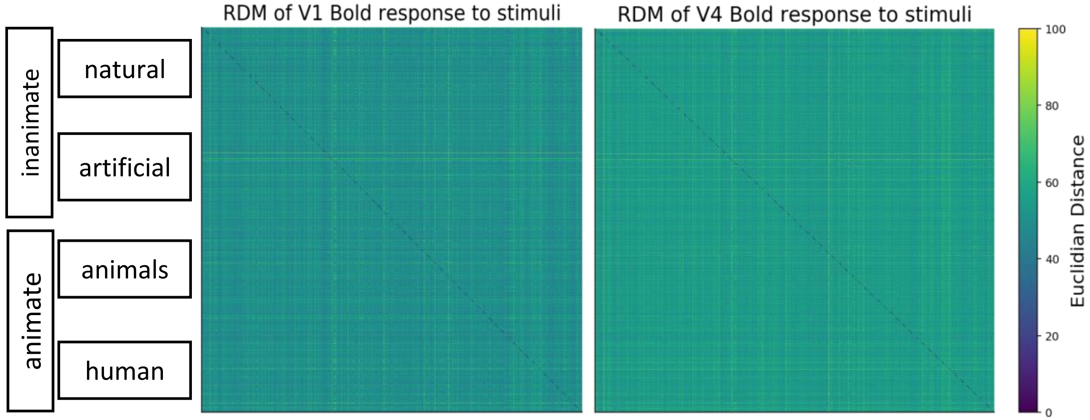


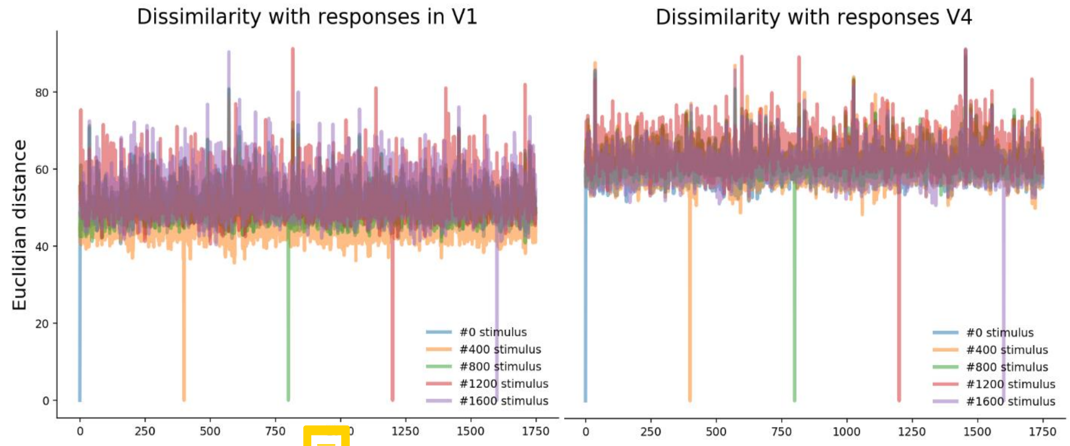



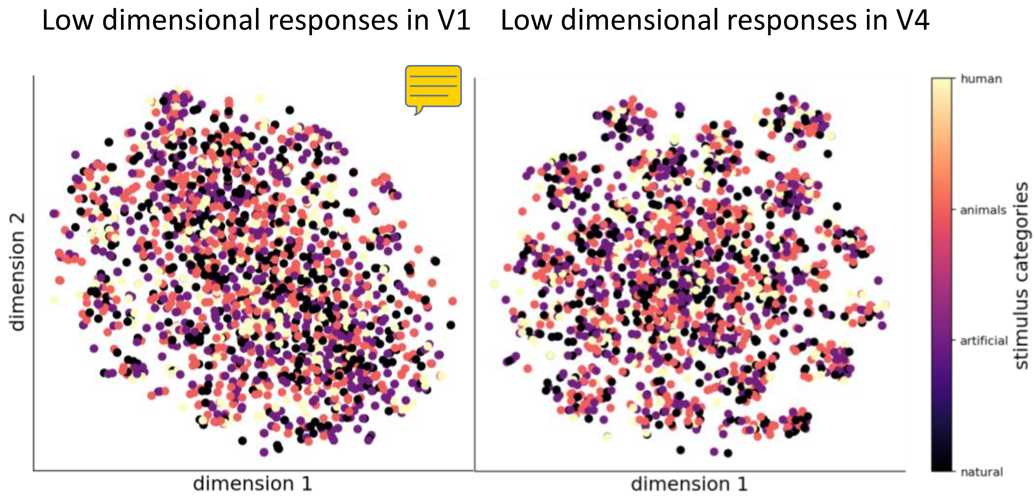



# Compute representational dissimilarity matrix in brain

- Interpret representational geometry(Kriegeskorte & Diedrichsen 2019)
  - Euclidean distance precisely defines the discriminability of a pair of stimuli.
  - Linear: this is a biologically plausible computation for a single neuron, what information a downstream neuron might extract from a neural population.
  - Nonlinear: all(maybe) related information in this brain area.
  - Smooth representational geometry reflect that the low-dimensional ~~coding~~.
- Representation dissimilarity matrix of the fMRI data
  - A stimuli(1750) x stimuli(1750) matrix, stimuli are sorted by 4 categories.
  - RDMs are not smooth[top right], suggests the encoding is high-dimensional.
  - The result seems plausible in V1, brain doesn't cluster the stimuli by their semantics in early visual cortex. But there should be some clusters in V4.
  - The stimuli have equally long distance from each other[middle right].
- Why?
  - We don't have data in high-level brain areas, the result might be different.
  - Stimuli are unclear and hard to identify, thus the process may be purely perceptual. Subjects couldn't use smooth coding scheme to encode.
  - The data is too noisy, can't be clustered by non-linear method[bottom right].



# Compare neural network and brain in representational level

- RDM of the fMRI responses after denoise by averaging
  - fMRI data may be too noisy, so we average across stimuli within each categories and get a categories(4) x categories(4) matrix[top&middle right].
  - The distances between categories become longer from V1 to V4, implying that the representations of categories are gradually decoupled.
  - The distances within the two broad categories, i.e. “animate” and “inanimate”, gradually become shorter and cluster together from V1 to V4. This may imply the stimuli in the same semantic category elicit a prototypical response pattern, which is what we expect.
- Neural network on this task
  - Use a pretrained Resnet-18(He et al. 2016), a brain-like recurrent ANNs CORnet(Kubilius et al. 2018), a self-implement RCNN (Spoerer et al. 2017).
  - Only pretrained Resnet-18 achieve an above 85% accuracy, the accuracy of the other two are about 29% and 28% respectively.
  - We only compare the hidden activity between Resnet and brain in representational level.
  - Every layers of Resnet have the same RDM, and the stimuli(1750) precisely clustered into the 4 categories [bottom right]