Mapping V4 to Artificial Neurons via Autoencoder allows Decoding Visual Information

Xuming Ran (冉旭明)

ranxuming@gmail.com

BME Department, Southern University of Science and Technology (SUSTech)

AI as a tool to analyze neural / behavioral data

External World (stimuli)

An AI to encode (simulate) neural signals?

10^{12} neurons in human brain

Another AI to decode neural signals?

Senses
Action
Emotion
Cognition
…

What are problems of the end-to-end

- Not possible to record all neurons
- No enough training datasets
- No enough computational power to fit data

- Even though we can do all these, we know nothing from the model.

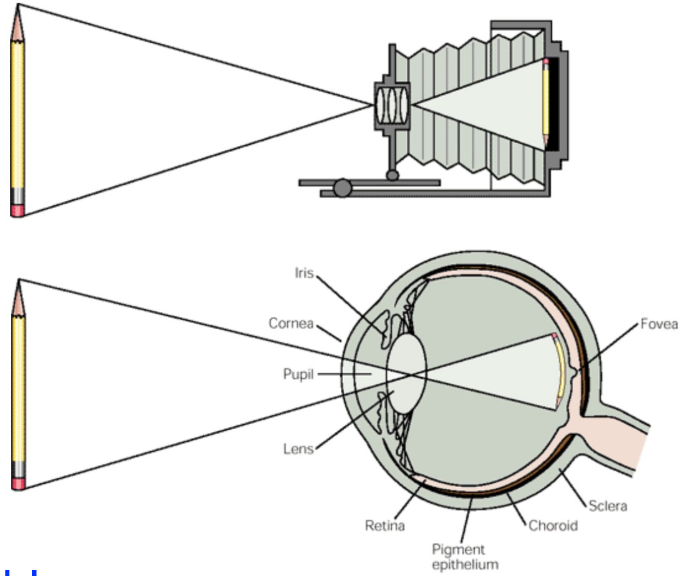
A good AI in neuroscience shall not only fit the data, but also provide insights to **explain the underlying mechanisms** of how the brain works.

Function of the visual system

The visual system is a part of the central nervous system (CNS) that gives organisms the ability to **detect**, **process**, **interpret** information from **visible light**, with the goal of building a **representation** of the surrounding environment.

Forward problem

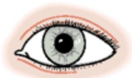
Image formation



Inverse problem

Vision (brain)

Camera



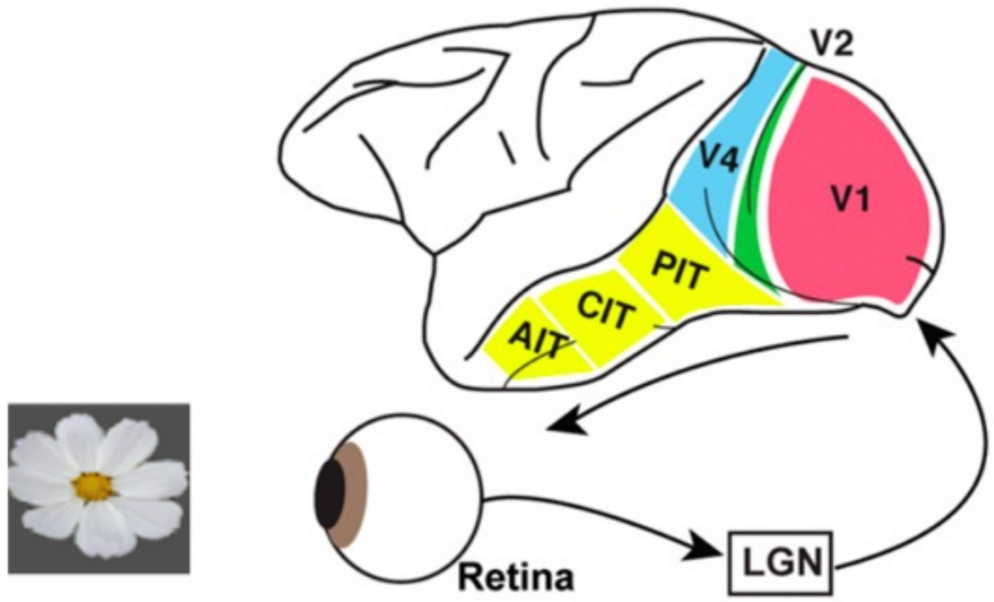
Eye
LGN
V1
V2
…

NOT a camera:

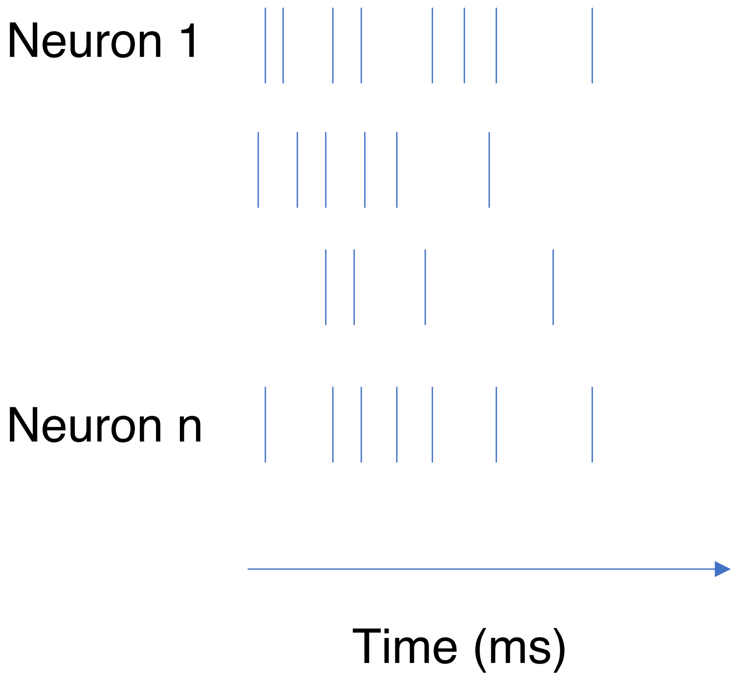
Visual system solves inverse problem

Detect
Process
Interpret

Inverse problem in the visual system



Information presented in brain



infer

Information in world

Tree
Flower
Cat
Table
…

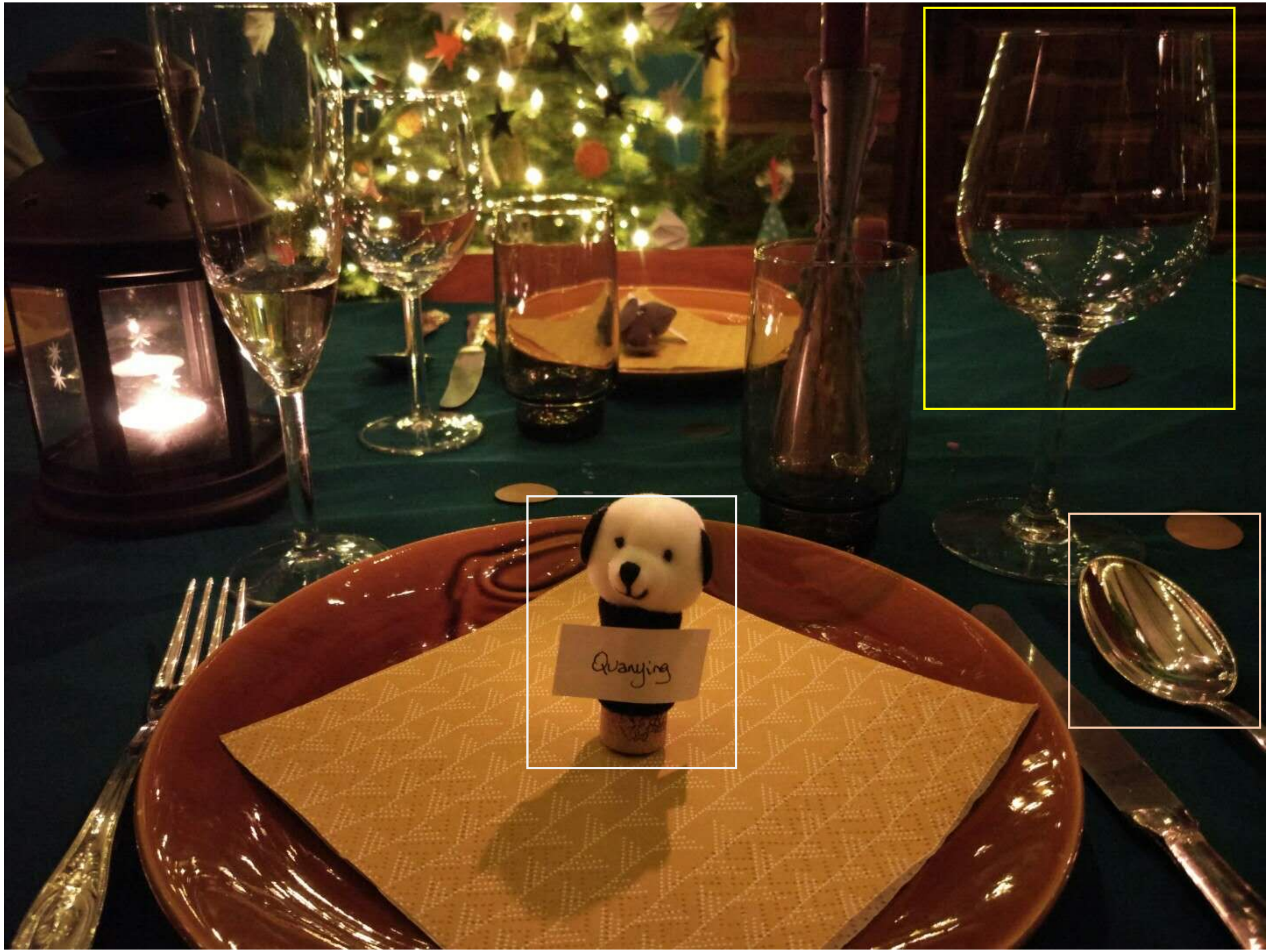
Our brain is an inference machine.

The visual system rapidly solve the inverse problem.

Our visual system rapidly solves object recognition task.

Visual system solves the <u>core object recognition</u> task ~ 200ms, robustly and accurately.

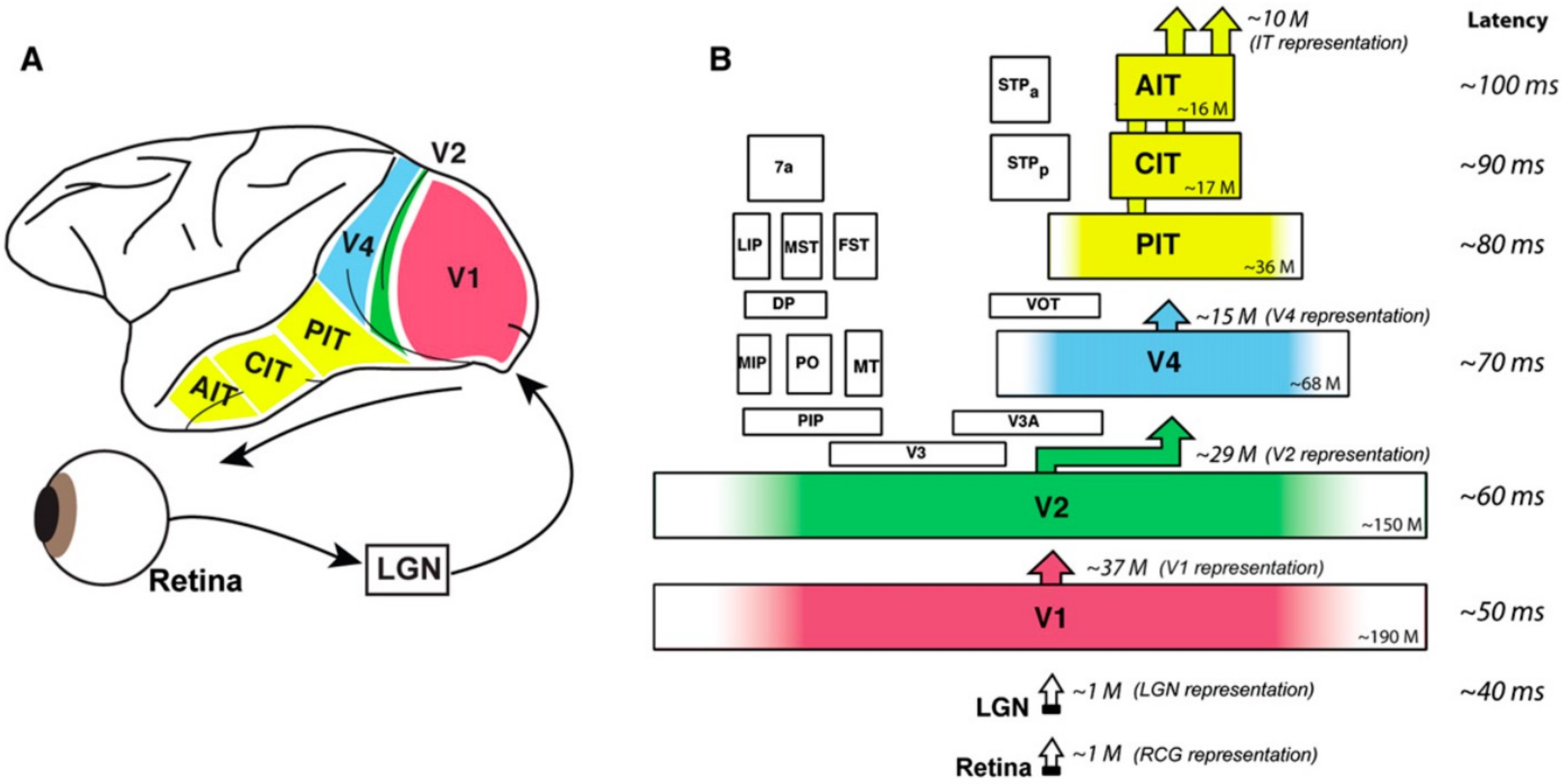
AI can also rapidly solve object recognition task.



Is there any similarity between brain and AI?

The answer is yes.
Plenty.

The ventral visual pathway in brain: for object recognition

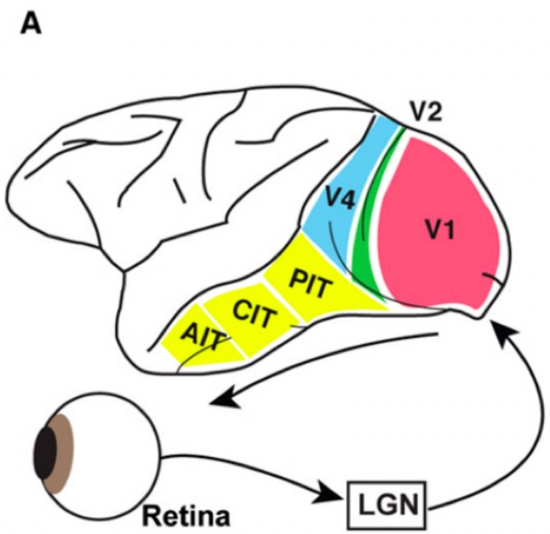


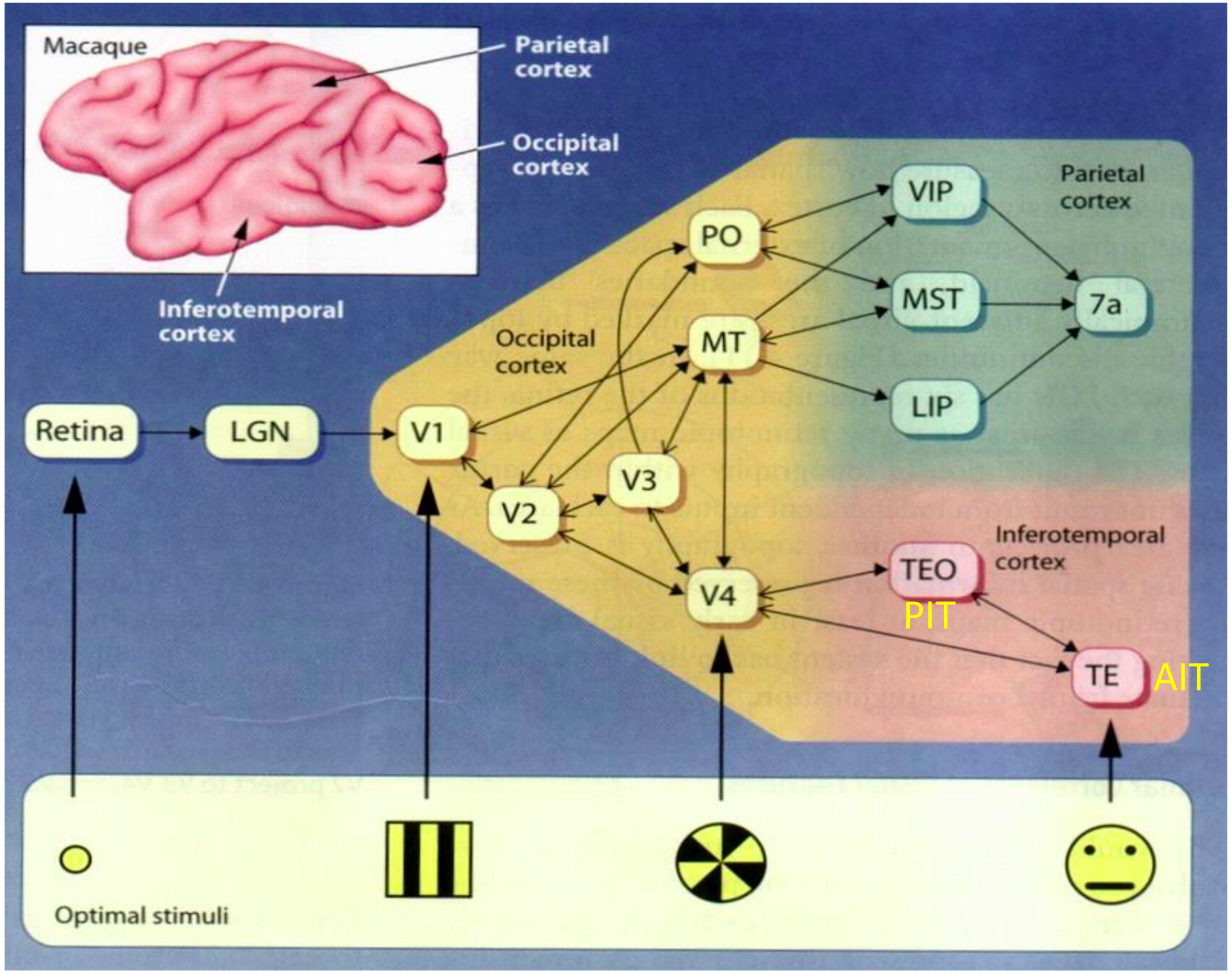
Anterior Inferior Temporal

Lateral Geniculate Nucleus

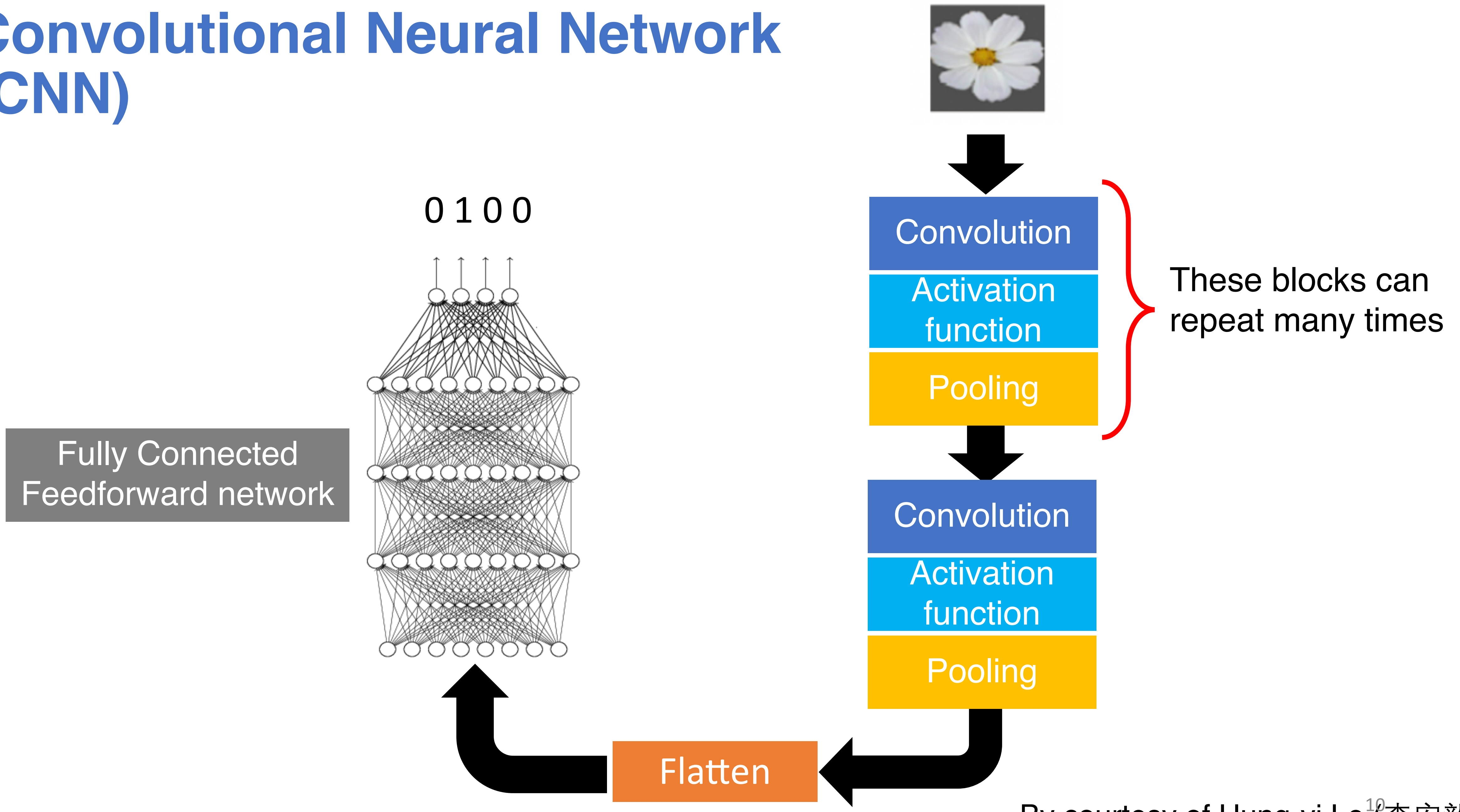
JJ DiCarlo, D Zoccolan, and NC Rust (2012). Neuron

Feature extraction from simple to complex, along the ventral stream





Convolutional Neural Network (CNN)



By courtesy of Hung-yi Le (李宏毅)

Image process: brain vs. CNN

1. Small receptive field vs. small convolutional kernel

A neuron does not have to see the whole image to discover the pattern.

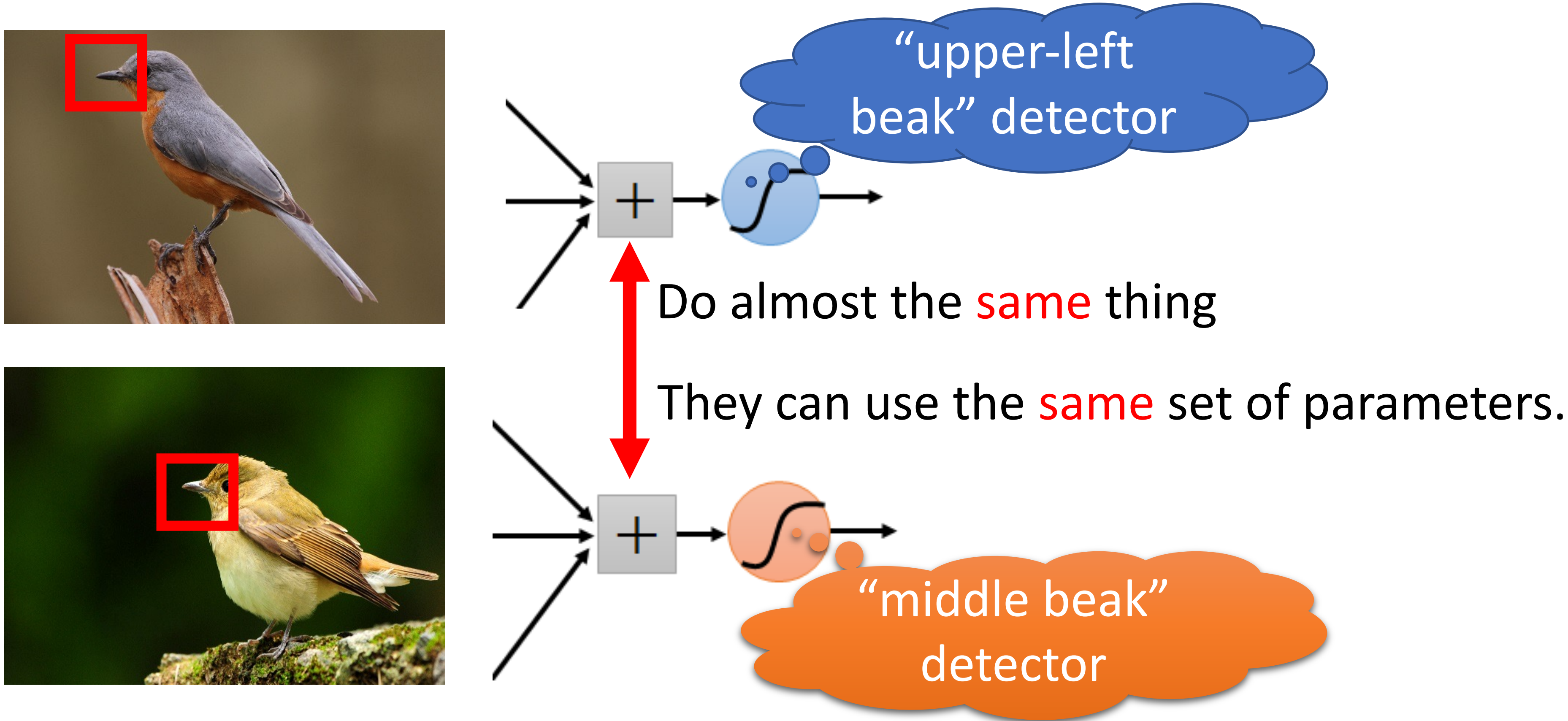
Connecting to small region with less parameters

By courtesy of Hung-yi Le (李宏毅)

Image process: brain vs. CNN

2. Identity-preserving transformation

Invariance to position, pose vs. shared convolutional kernel



By courtesy of Hung-yi Le (李宏毅)

Image process: brain vs. CNN

3. Invariance to scale and subsampling vs Pooling

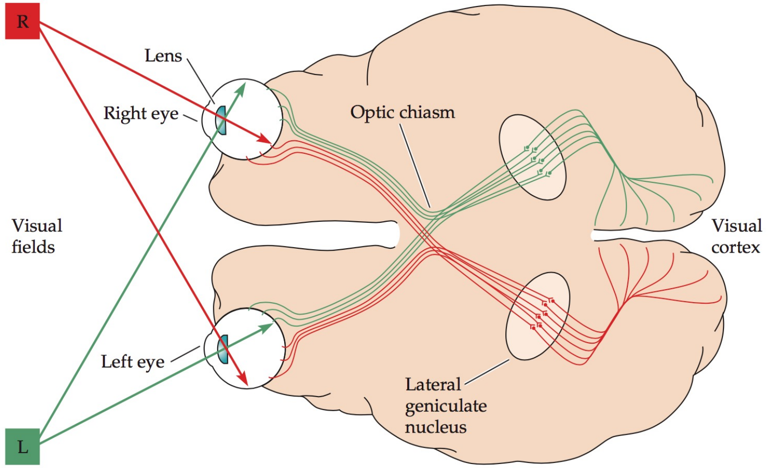
We can subsample the pixels to make image smaller

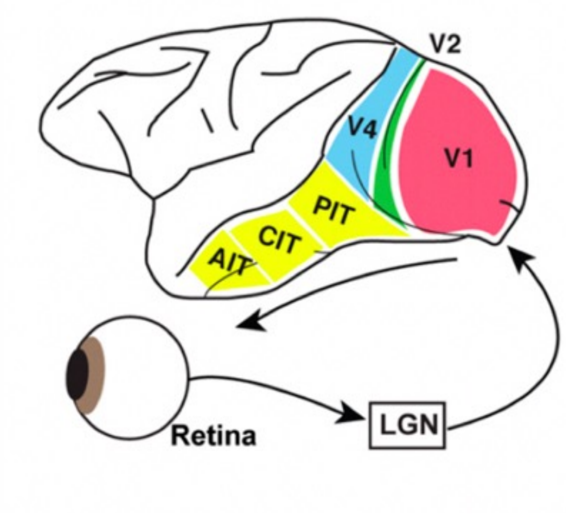
➡ Less parameters for the network to process the image

By courtesy of Hung-yi Lee (李宏毅)

Image process: brain vs. CNN

4. Both the visual system and CNN are layered and hierarchical.





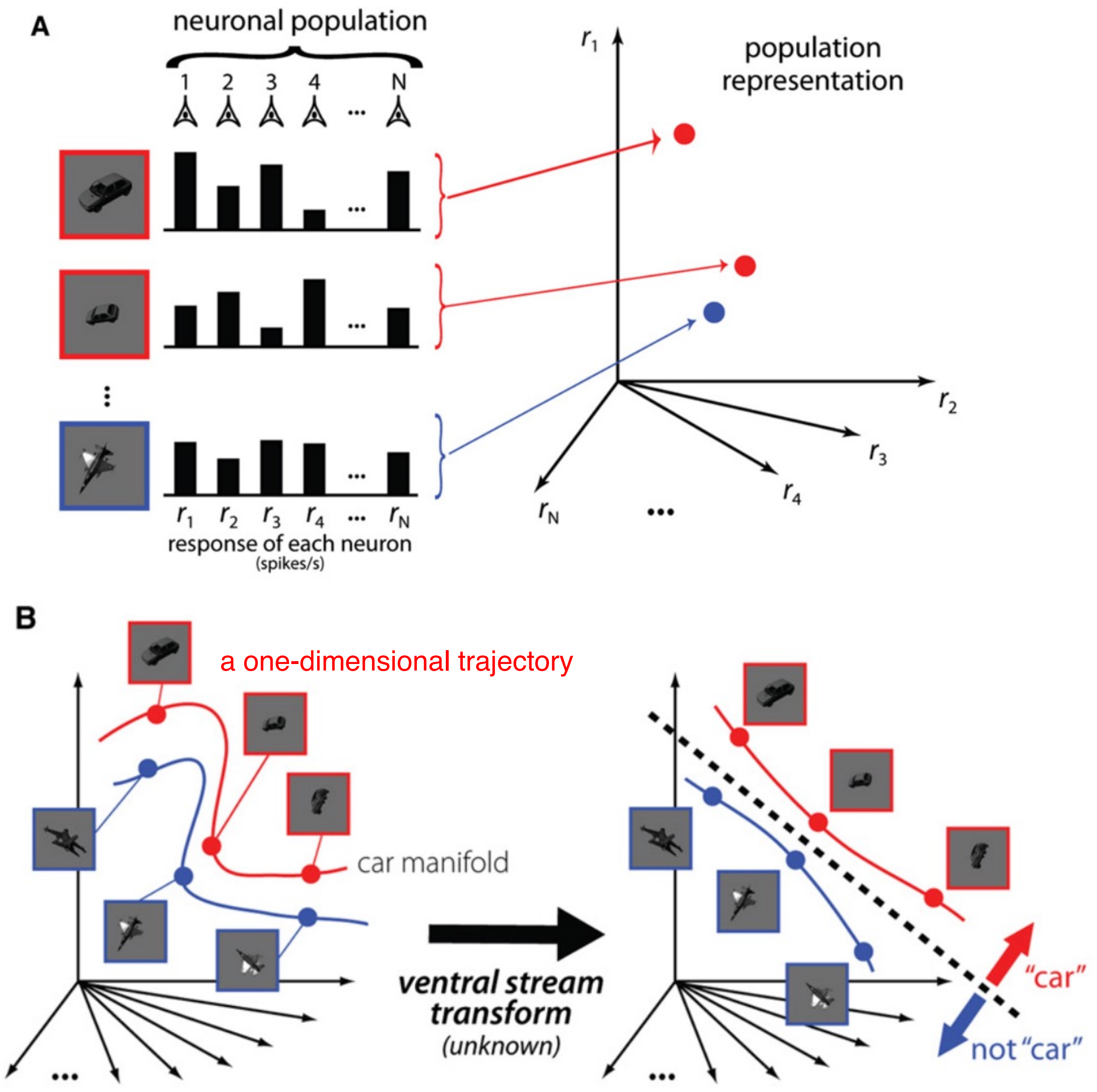
- Rods, cones
- Interneurons
- Ganglion cells
- Optic fiber
- LGN
- V1
- Ventral / dorsal pathways

Along the Visual Pathway, feature extraction from simple to complex.

→ Automatically learn the hidden features in the image

Is there any similarity between the neural representation in brain and AI?

What does the neural population representation mean?



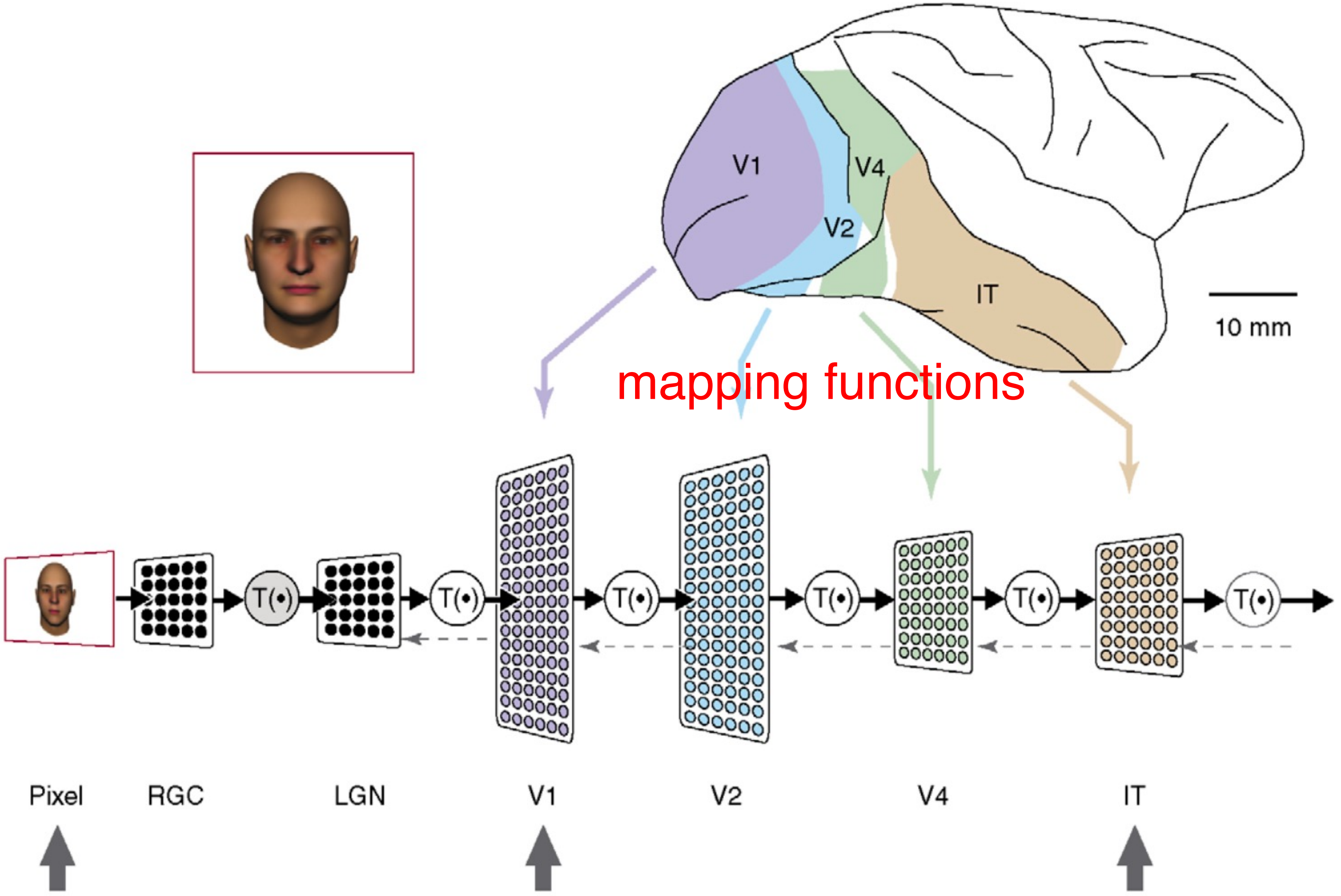
DiCarlo, Zoccolan & Rust (2012) Neuron

The response pattern of a population of visual neurons (e.g., retinal ganglion cells & V1 neurons) to each image is a point in a very high-dimensional space where each axis is the response level of each neuron.

All possible identity-preserving transformations of an object will form a low-dimensional manifold of points in the population vector space (i.e., a continuous surface).

Highly curved and tangled in early visual areas (LGN, V1)
→ An easy separation of object's manifold in later visual areas (V4, IT)

Biological neural representation resembles to the artificial neural representation.



Technical limits

Low-dimensional neuron recordings

easy

hard

High-dimensional CNN feature maps

DiCarlo and Cox, TiCS, 2007

Brain score: how well existing models explain the neural data

http://www.brain-score.org/#leaderboard

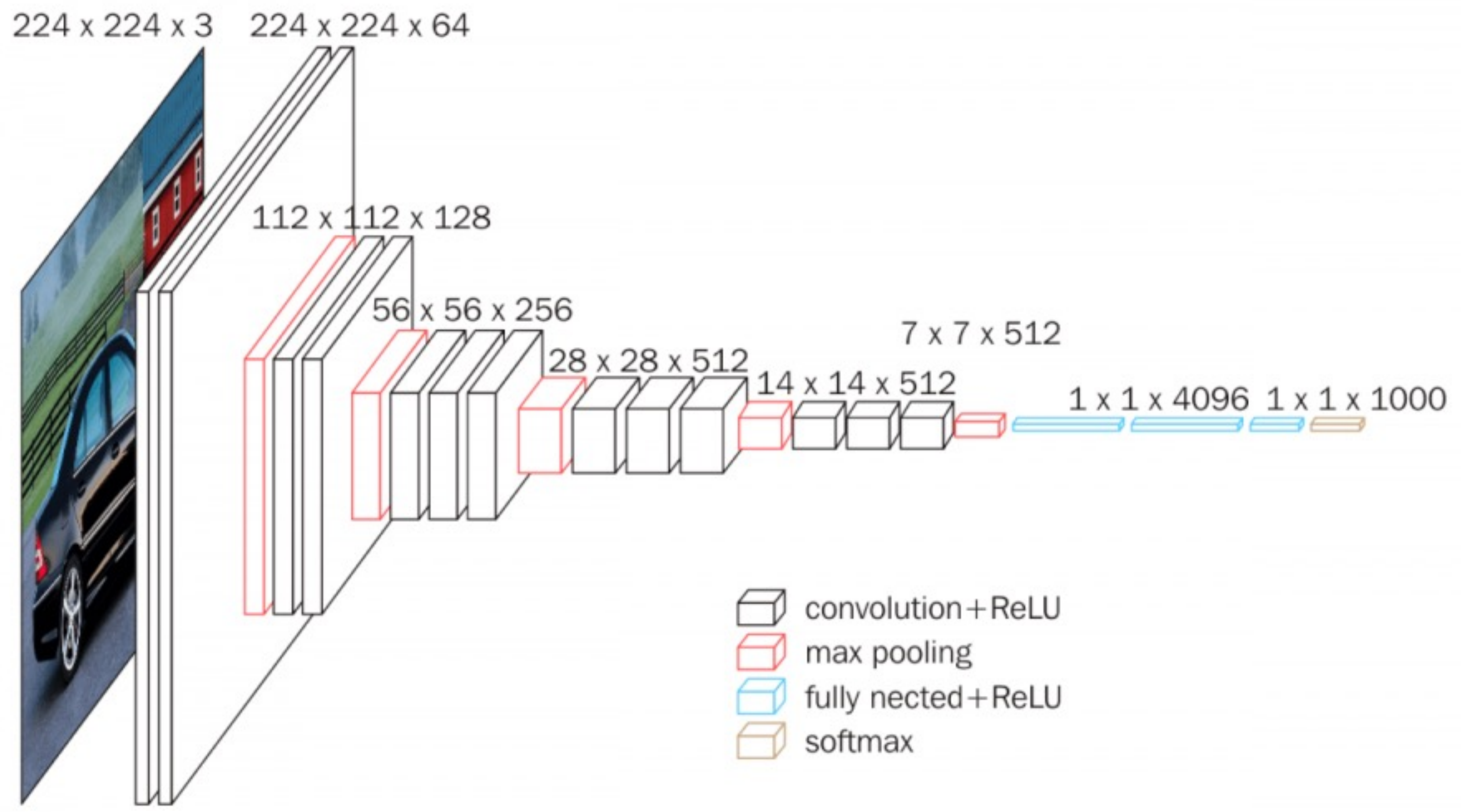
Sort by average score

Rank	Model	average	V1	V2	V4	IT	behavior	engineering	Deng2009-top1
1	efficientnet-b0 *Tan et al., 2019*	.442	.215	.317	.556	.547	.573		
2	efficientnet-b6 *Tan et al., 2019*	.435	.263	.295	.563	.541	.513		
3	efficientnet-b2 *Tan et al., 2019*	.434	.213	.317	.569	.547	.526		
4	efficientnet-b4 *Tan et al., 2019*	.434	.228	.286	.575	.543	.535		
5	CORnet-S *Kubilius et al., 2018*	.417	.294	.242	.581	.423	.545	.747	.747
6	vgg-19 *Simonyan et al., 2014*	.408	.347	.341	.610	.248	.494	.711	.711
7	resnet-50-robust *Santurkar et al., 2019*	.408	.378	.365	.537	.243	.515		
8	resnet-101_v1 *He et al., 2015*	.407	.266	.341	.590	.274	.561	.764	.764
9	vgg-16 *Simonyan et al., 2014*	.406	.355	.336	.620	.259	.461	.715	.715
10	resnet-152_v1 *He et al., 2015*	.405	.282	.338	.598	.277	.533	.768	.768

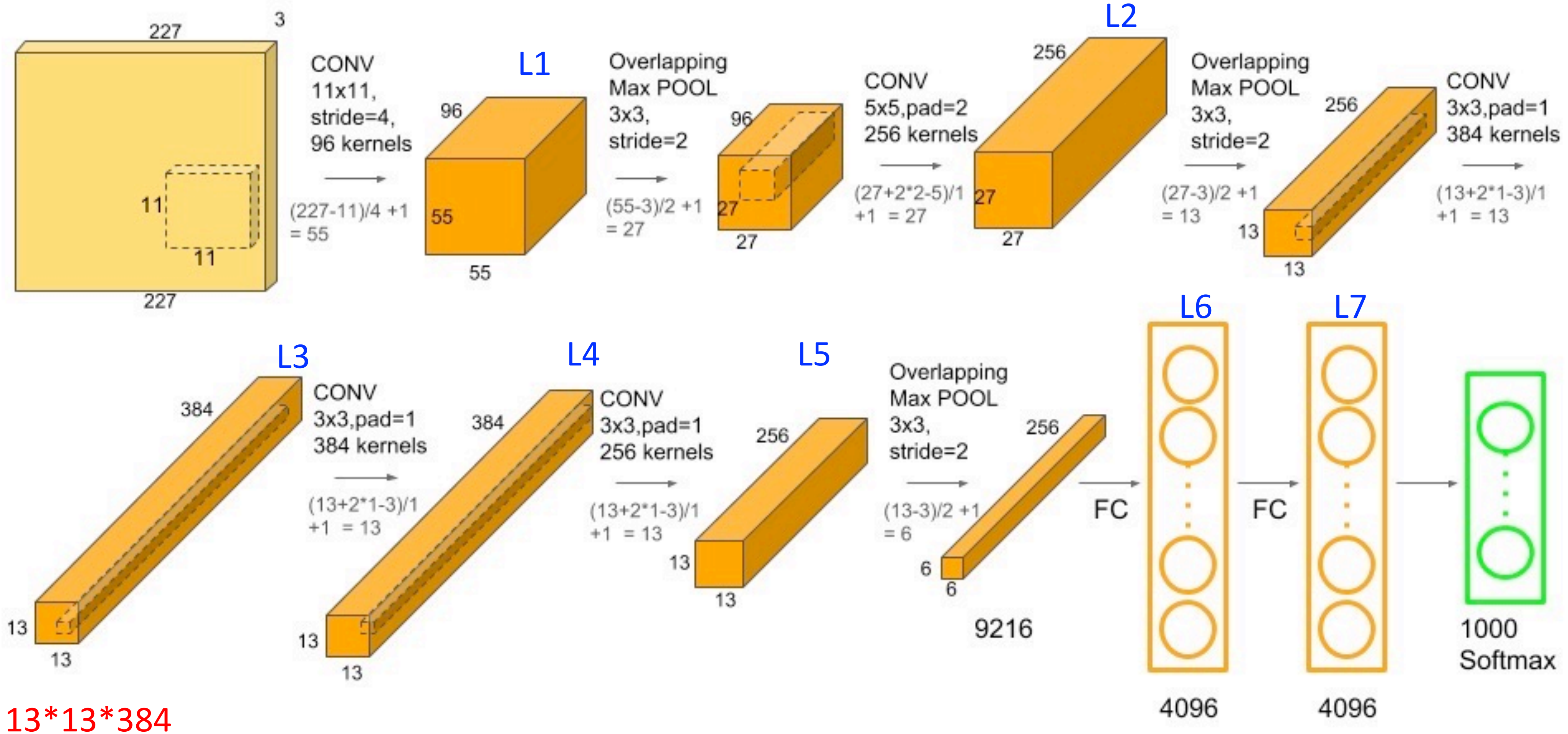
Sort by V4 score

Model	average	V1	V2	V4	IT	behavior	engineering	Deng2009-top1
vgg-16 *Simonyan et al., 2014*	.406	.355	.336	.620	.259	.461	.715	.715
vgg-19 *Simonyan et al., 2014*	.408	.347	.341	.610	.248	.494	.711	.711
xception *Chollet et al., 2016*	.384	.245	.306	.610	.249	.508	.790	.790
densenet-169 *Huang et al., 2016*	.404	.281	.322	.601	.274	.543	.759	.759
resnet-50-pytorch *He et al., 2015*	.399	.289	.317	.600	.259	.528	.752	.752
resnet-101_v2 *He et al., 2015*	.404	.274	.332	.599	.263	.555	.774	.774
resnet50-SIN_IN *Geirhos et al., 2019*	.404	.282	.324	.599	.276	.541	.746	.746
densenet-201 *Huang et al., 2016*	.402	.277	.325	.599	.273	.537	.772	.772
resnet-152_v1 *He et al., 2015*	.405	.282	.338	.598	.277	.533	.768	.768
resnet50-SIN_IN_IN *Geirhos et al., 2019*	.397	.275	.321	.596	.273	.523	.767	.767

Dimension of the neural population in VGG-16

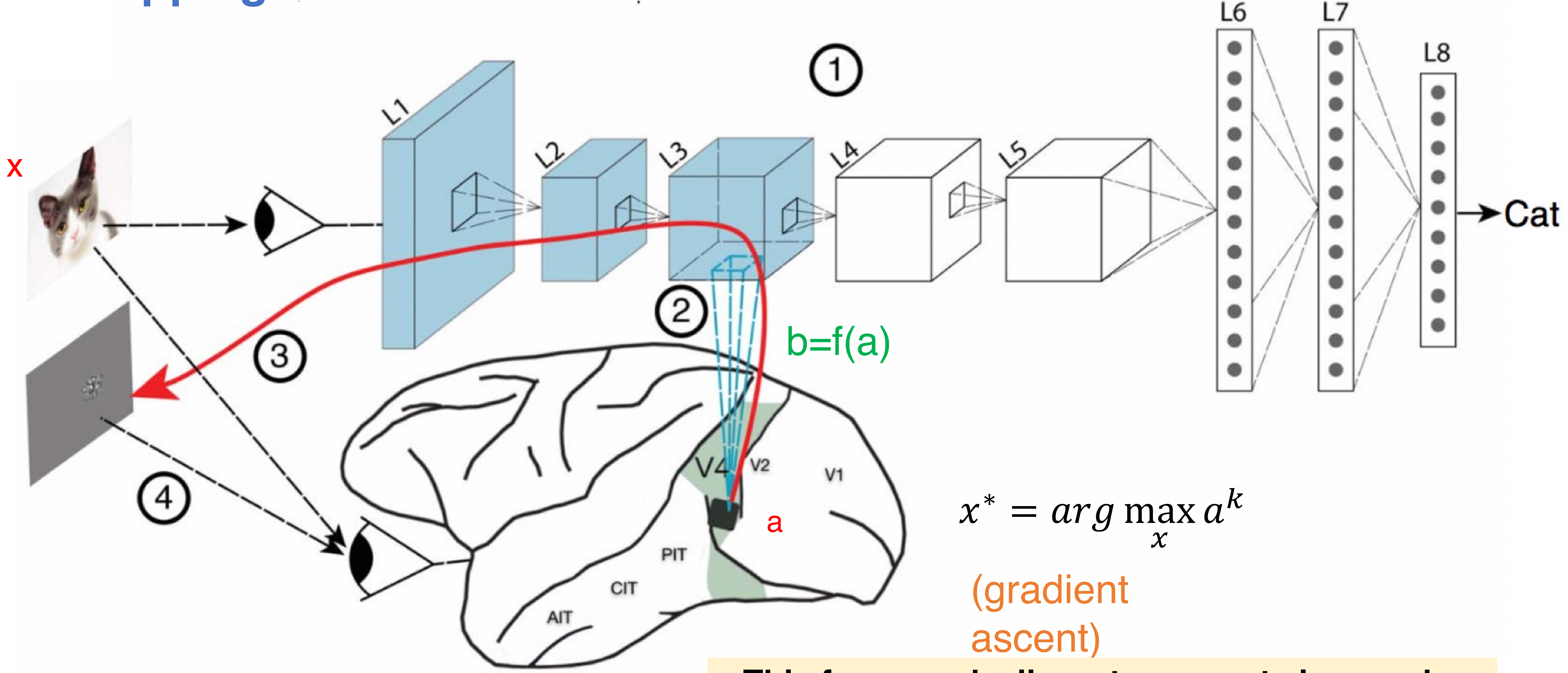


Dimension of the neural population in AlexNet



13*13*384

A mapping function from L3 to V4



$$x^* = arg \max_x a^k$$

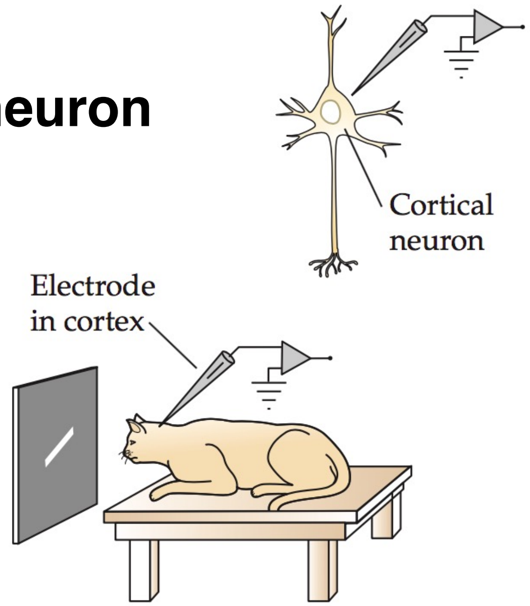
(gradient ascent)

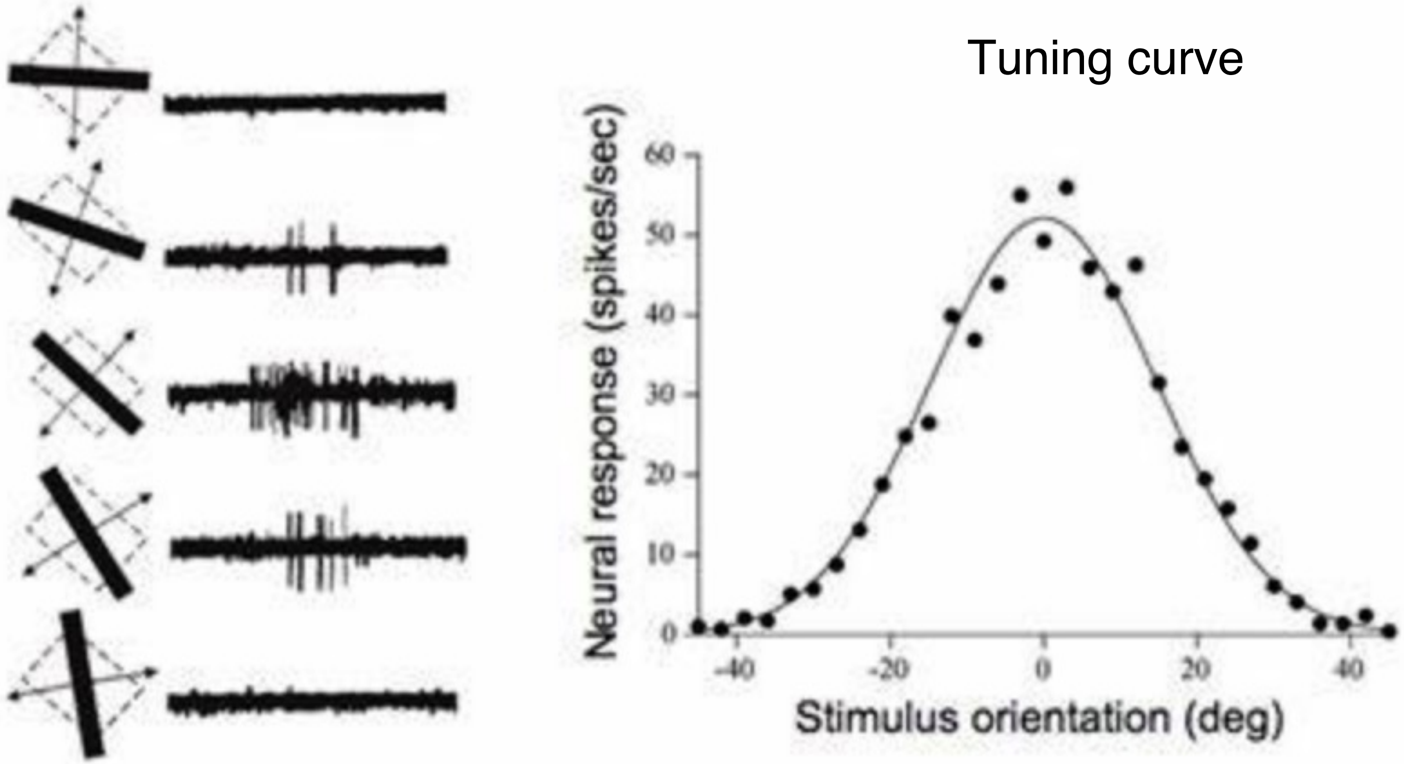
This framework allows to generate images by gradient ascent in ANN to maximize biological neural population activity.

Yamins, nature neurosci, 2016
Bashivan, Science, 2019

Why we want to find the image to maximize neuronal activity?

V1 neuron





Hubel, D. H. 1982. *Nature* 299: 515–524.

Hubel, D. H., and Wiesel, T. N. 1959. *J. Physiol.* 148: 574–591.

Hubel, D. H., and Wiesel, T. N. 1962. *J. Physiol.* 160: 106–154.

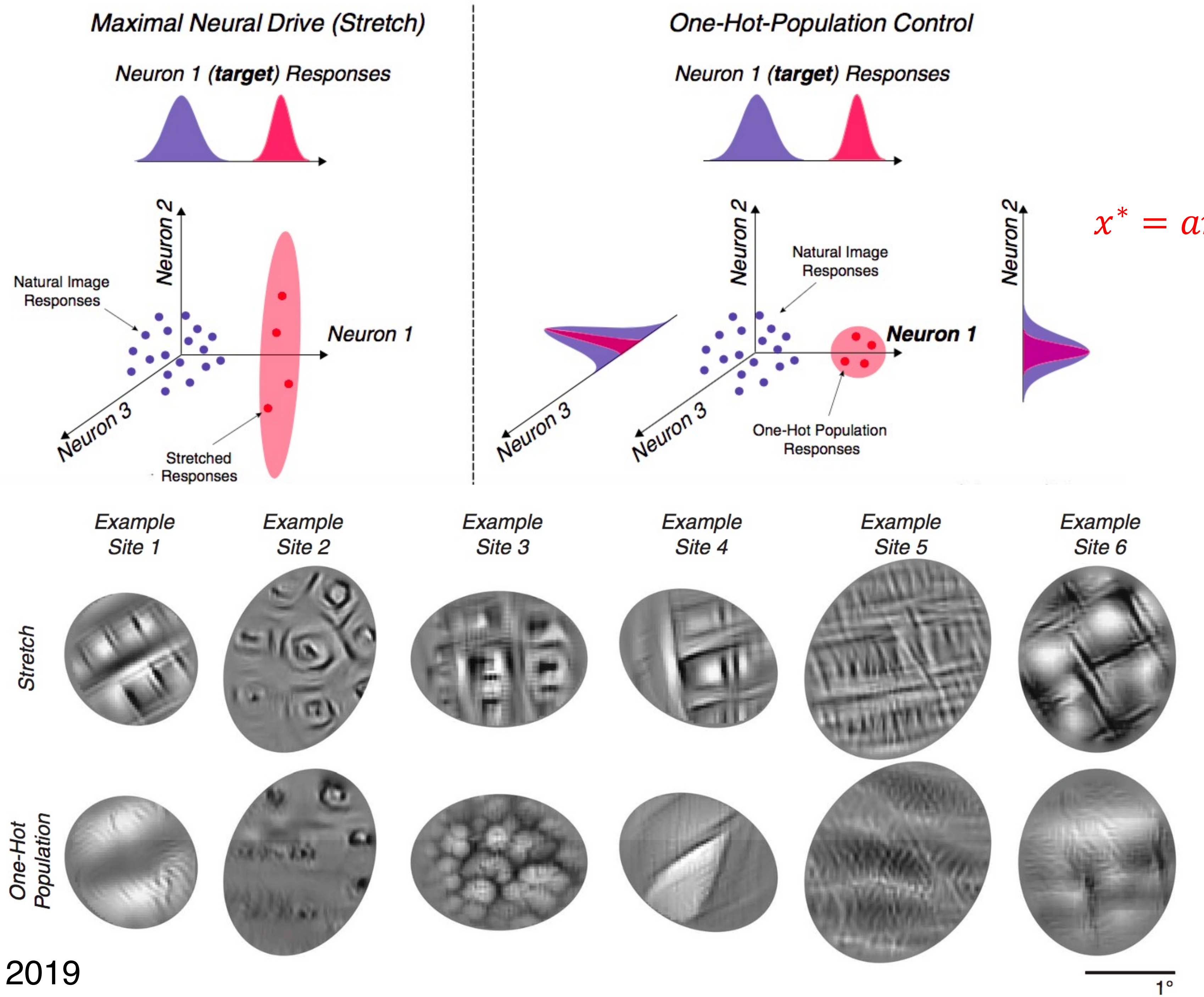
Hubel, D. H., and Wiesel, T. N. 1968. *J. Physiol.* 195: 215–243.

David H. Hubel & Torsten N. Wiesel

Nobel Prize for Physiology or Medicine in 1981

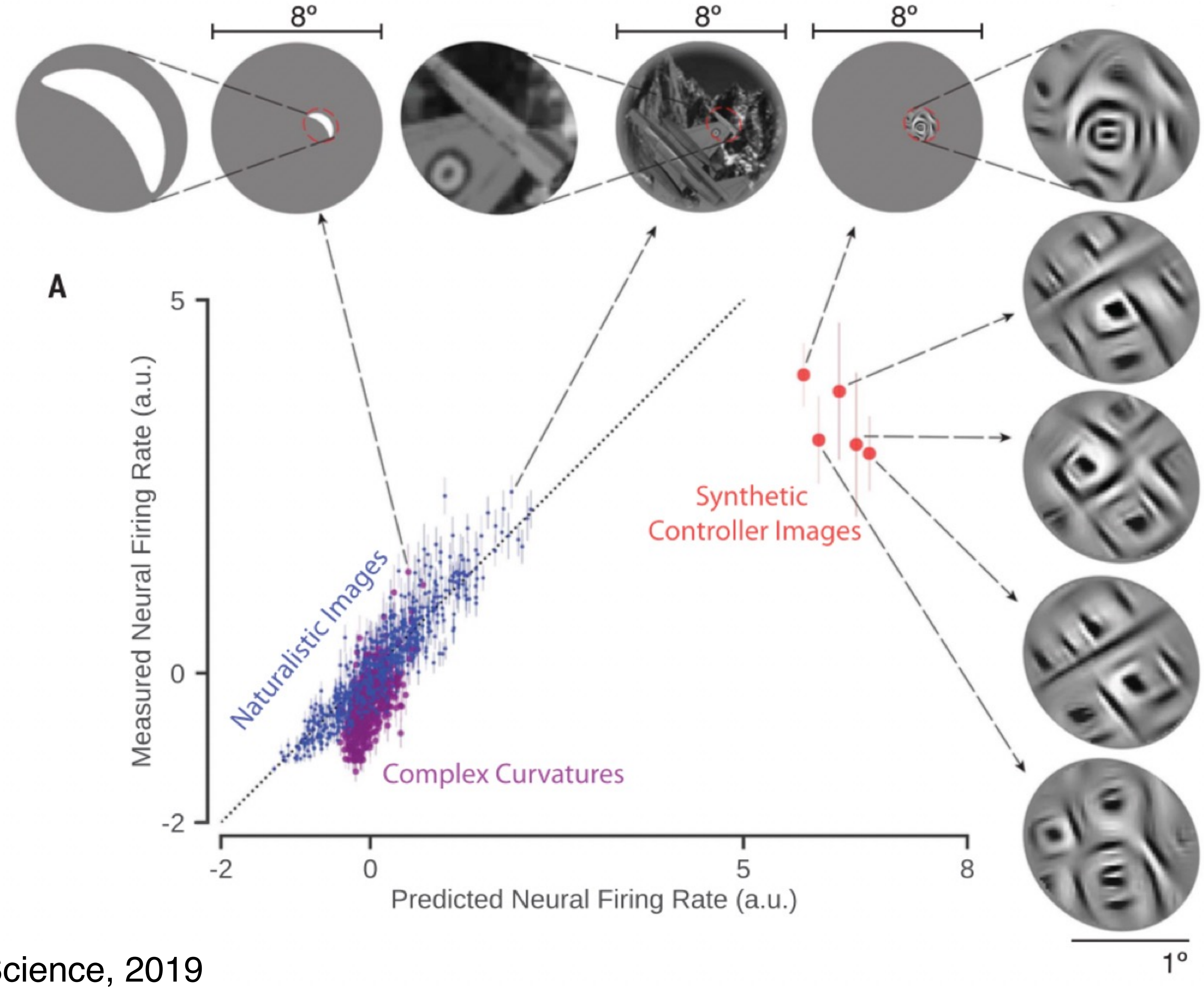
How about V4 and IT?

$$x^* = arg\max_x a^k$$



$$x^* = arg\max_x \frac{e^{a^k}}{\sum_k e^{a^k}}$$

Bashivan, Science, 2019



$$x^* = arg\max_x a^k$$

(gradient ascent)

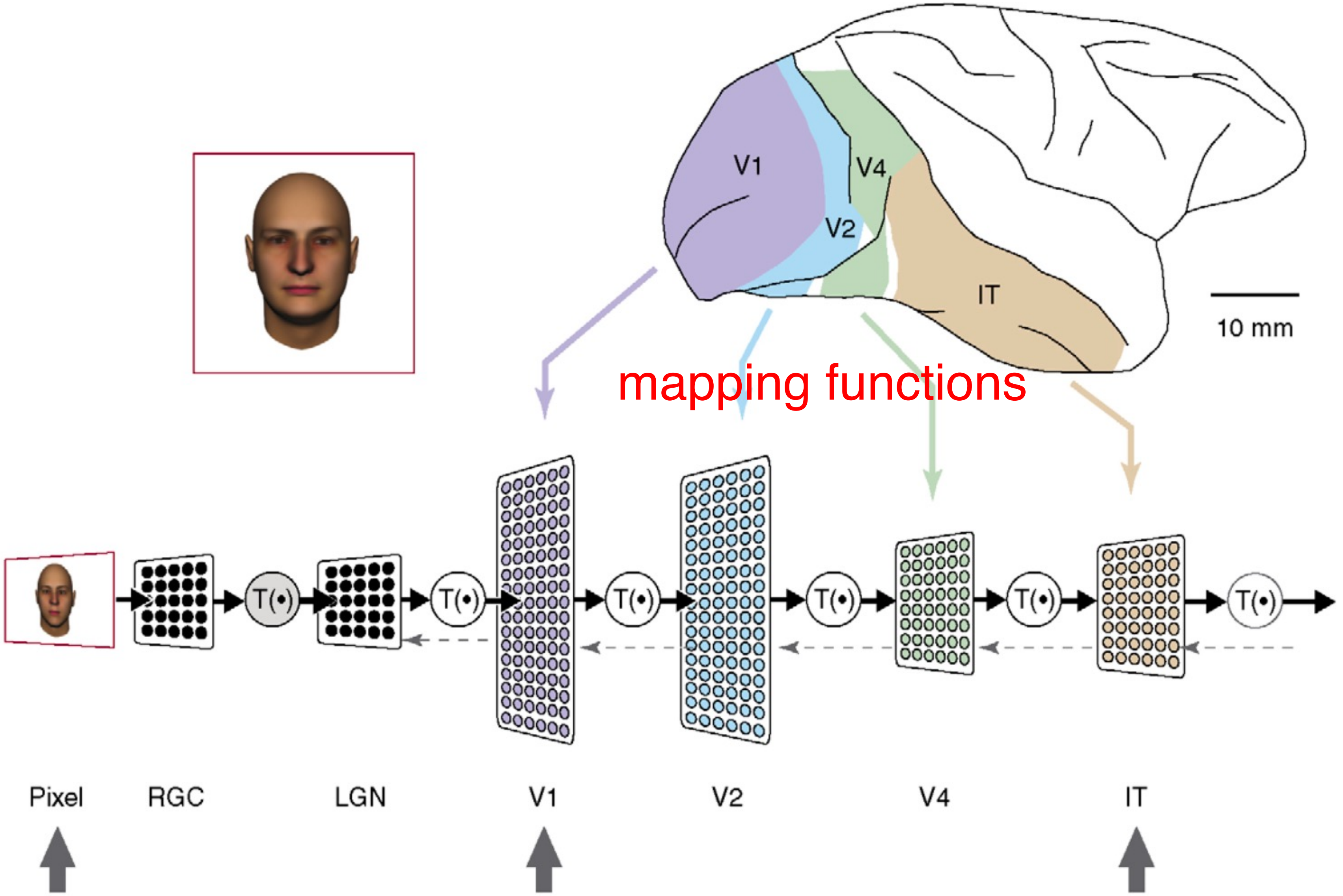
Bashivan, Science, 2019

Can we ascending dimensions from the low-dimension recorded neurons in brain to high-dimension artificial neurons in ANN?

Please no video recording,
since this study is ongoing and not published yet.

Thank you.

Biological neural representation resembles to the artificial neural representation.



Technical limits

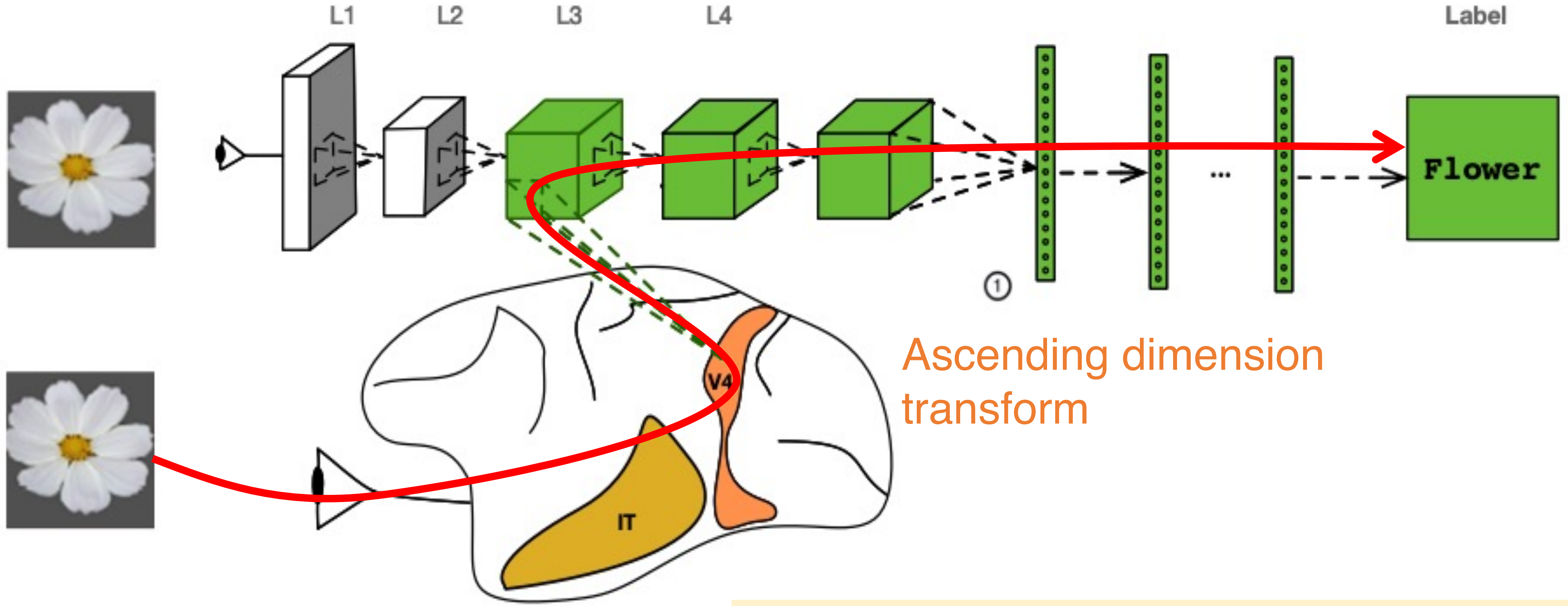
Low-dimensional neuron recordings

easy ↑ ↓ **hard**

High-dimensional CNN feature maps

DiCarlo and Cox, TiCS, 2007

A mapping function from V4 to L3



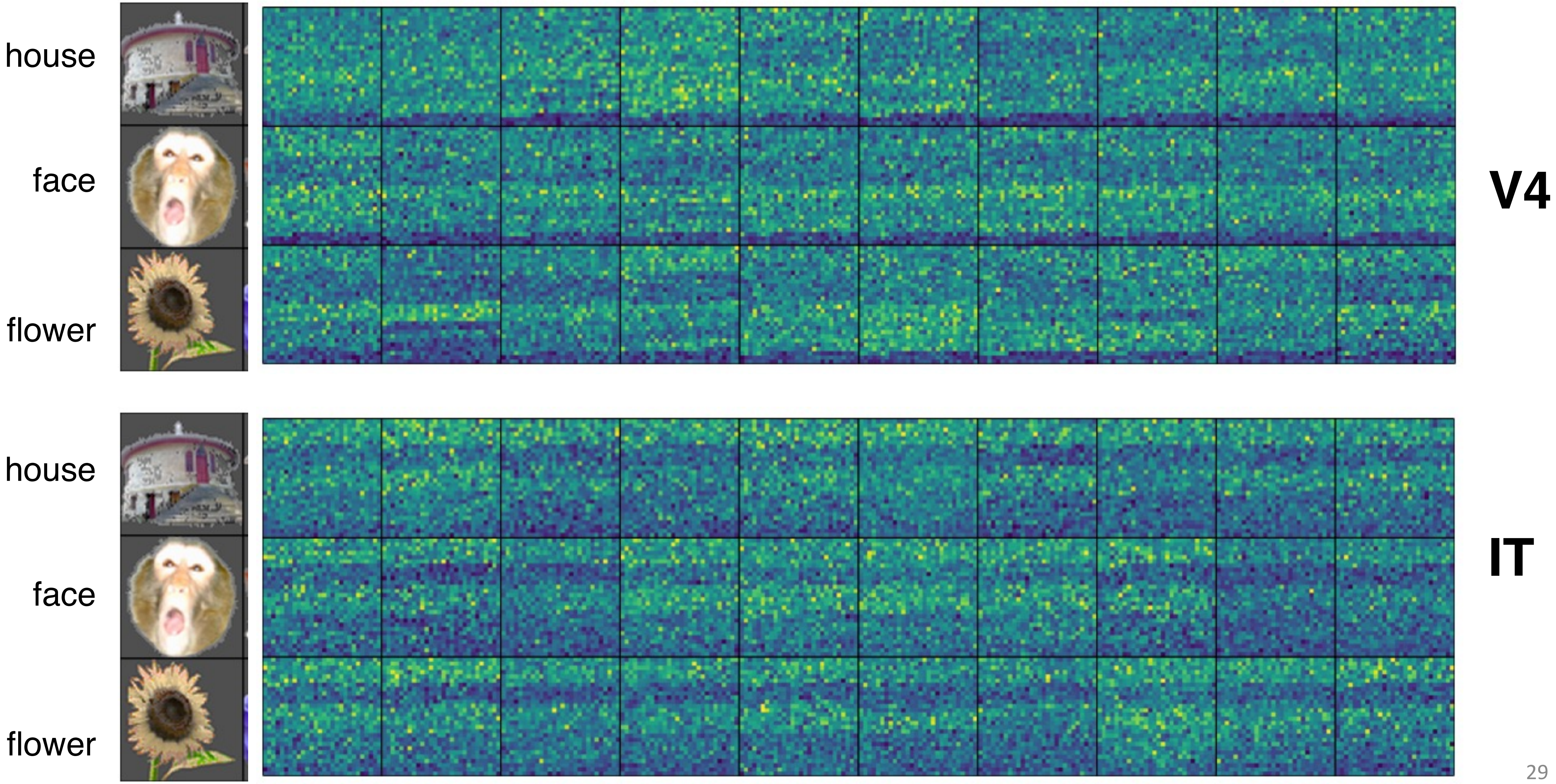
Ascending dimension transform

This framework would allow
1) to recognize object, and even
2) to reconstruct visual image
by biological neural population activity via ANN

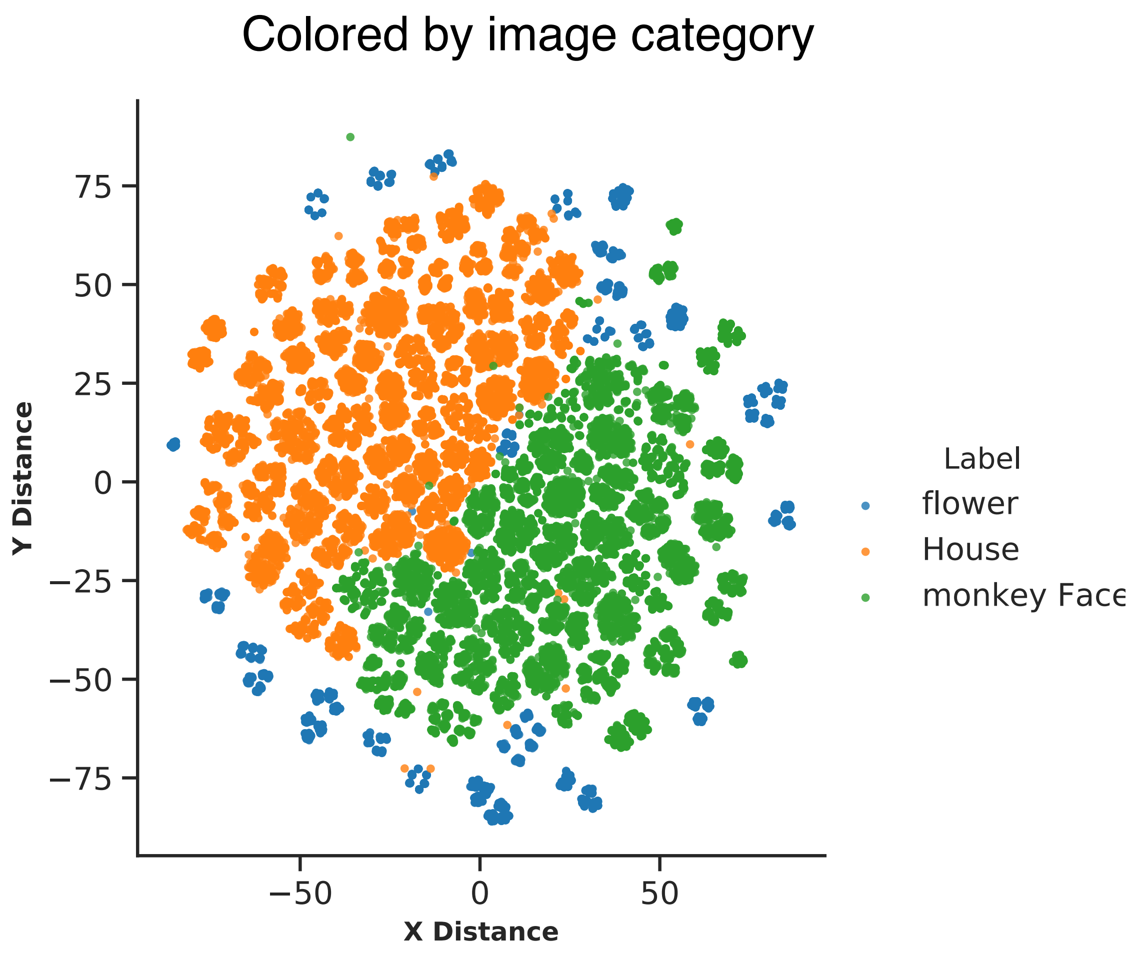
Image presented to monkeys

ImageNet to pre-train Alexnet

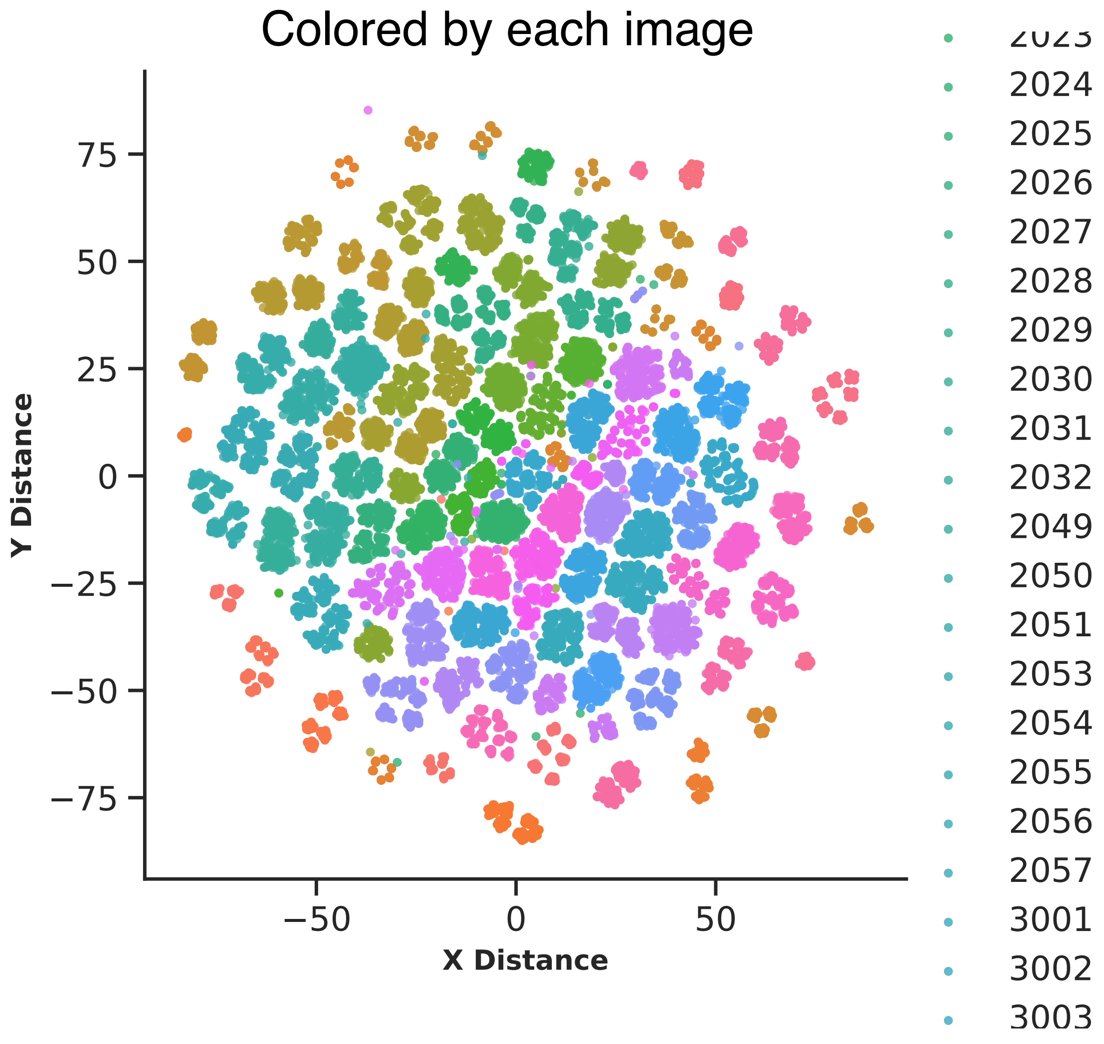
Neural responses in V4 and IT



t-SNE for the firing rate of V4 neurons

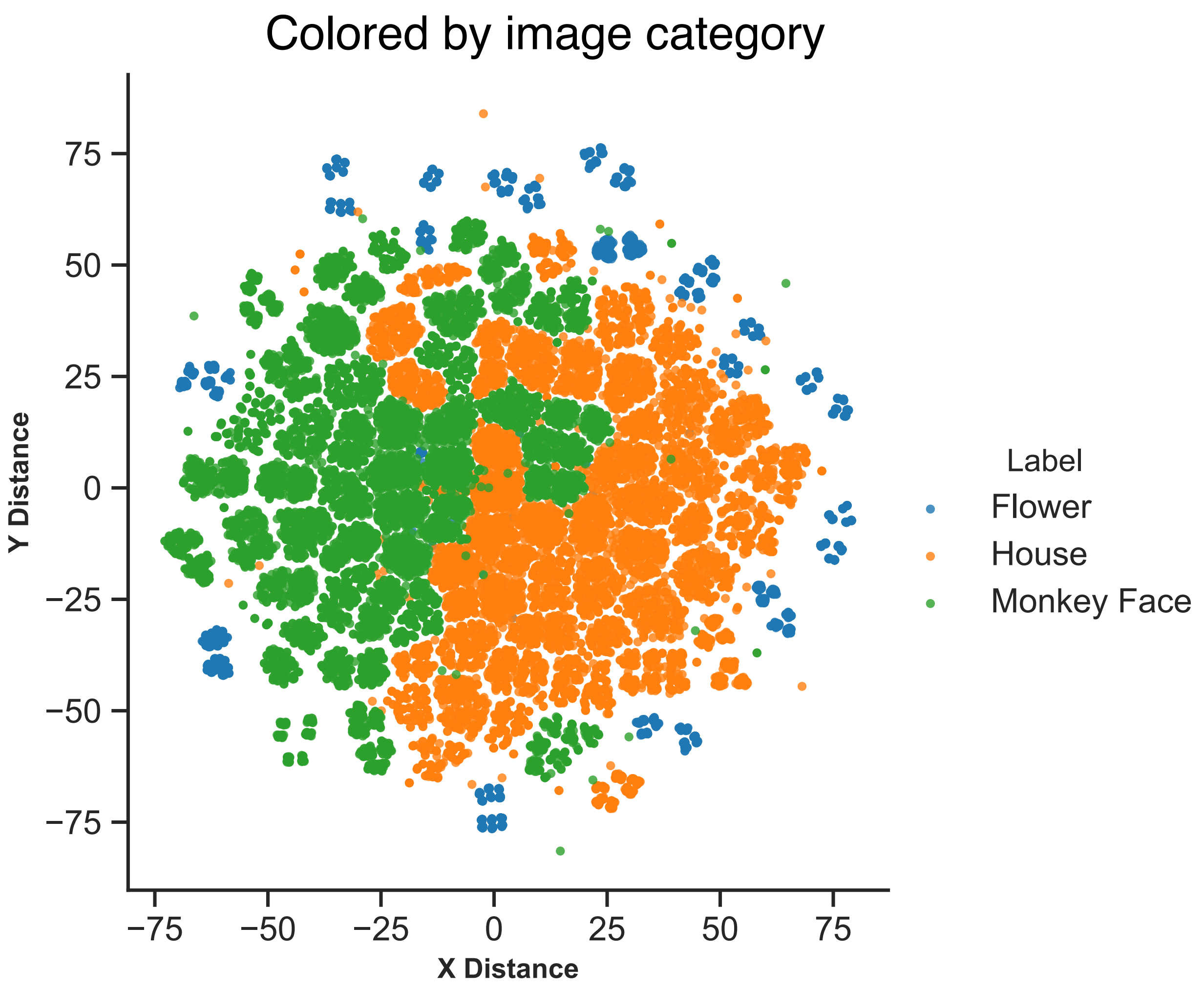


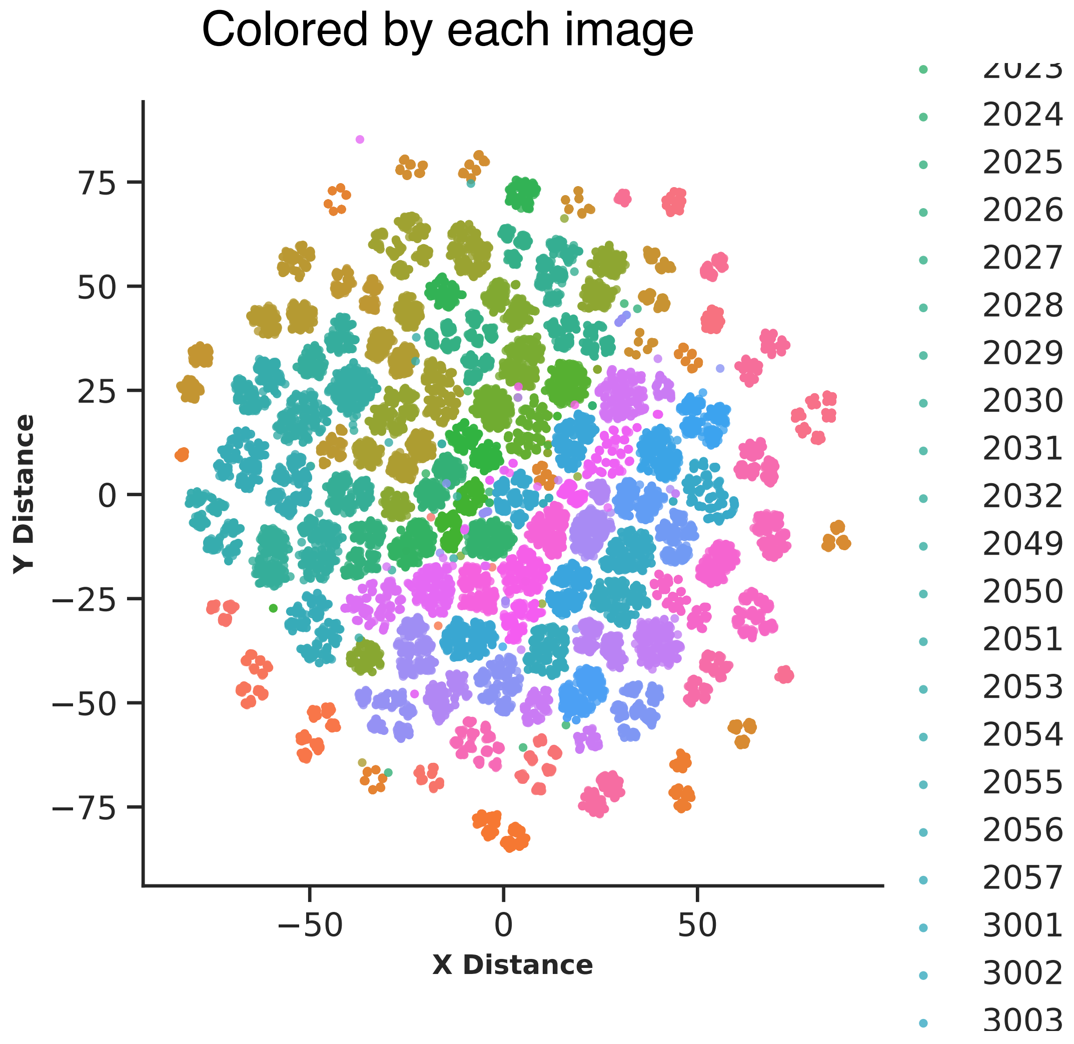
V4 0- 224 ms



V4 0- 224 ms

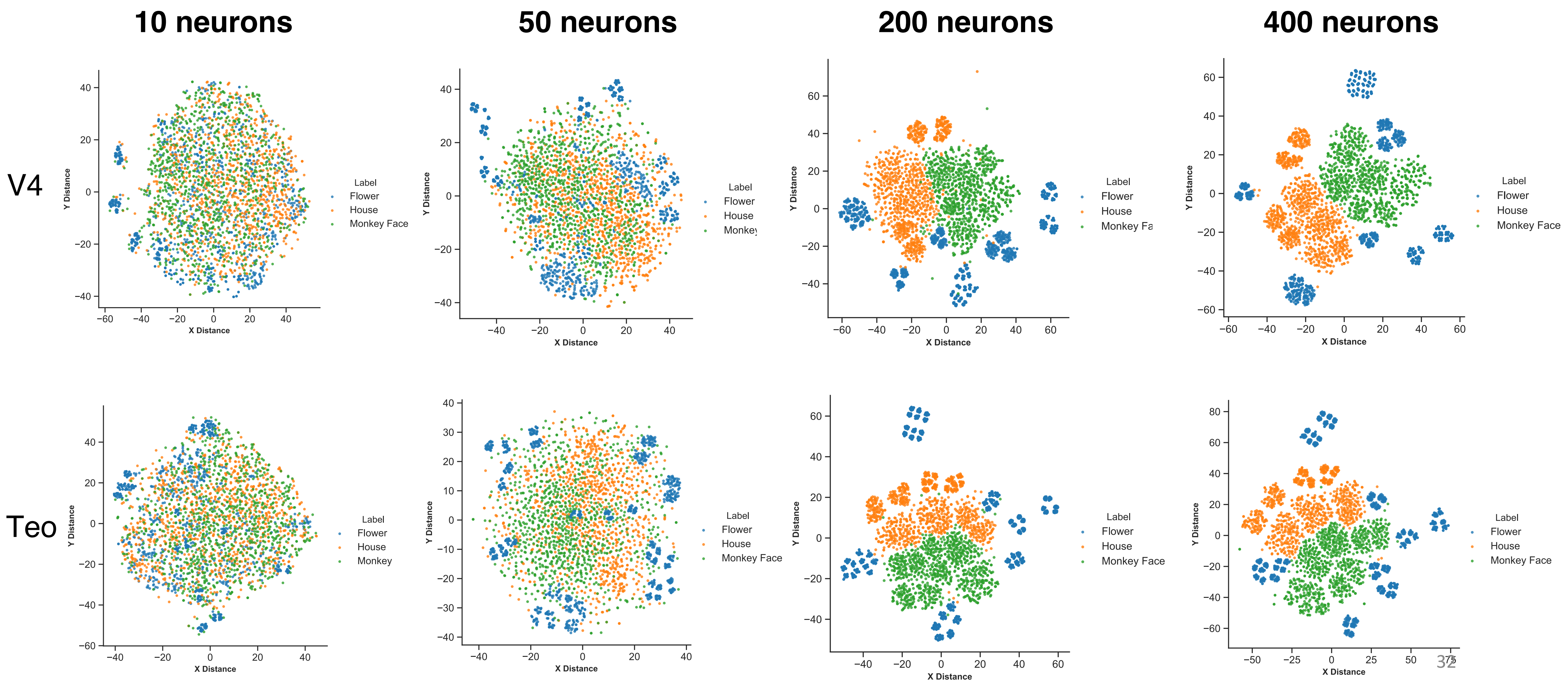
t-SNE for the firing rate of IT neurons



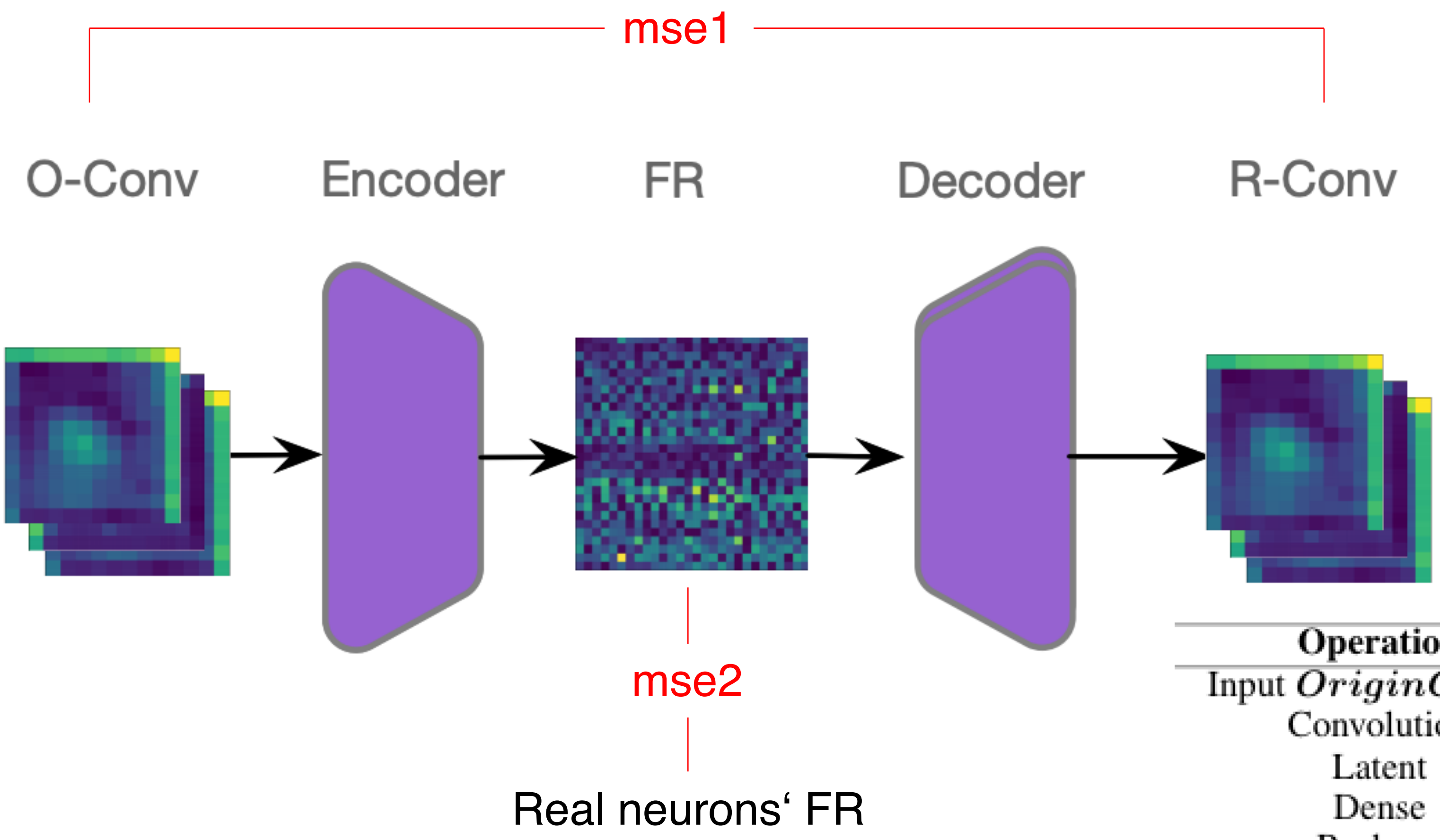


Teo 0- 224ms

t-SNE for the firing rate of V4 and IT



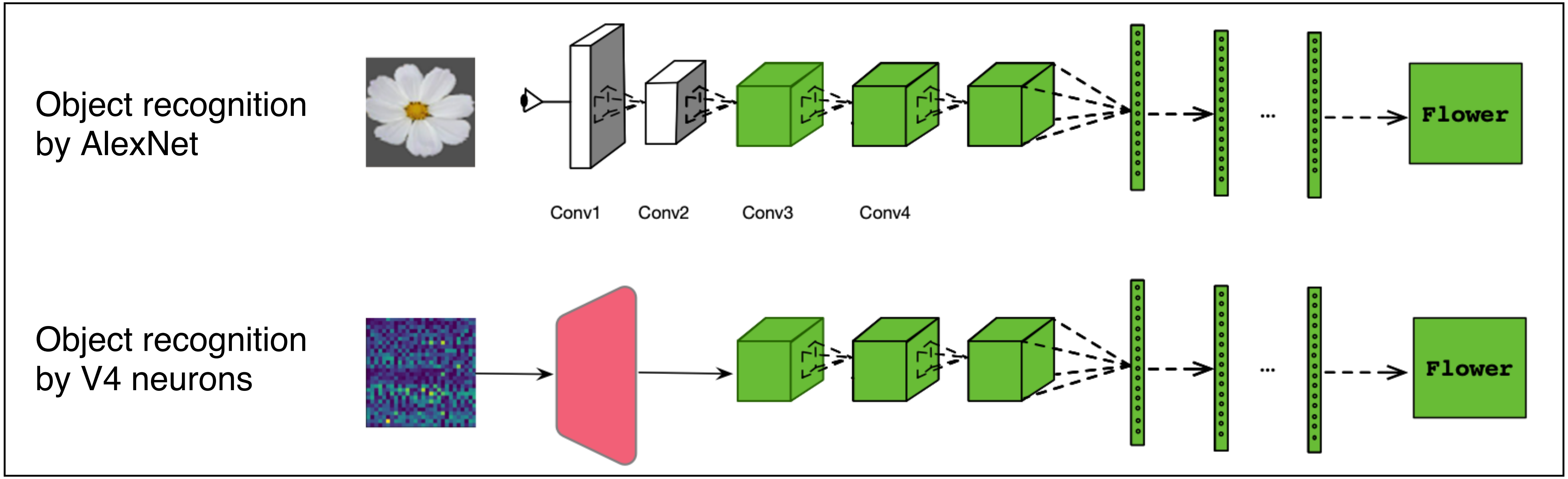
Autoencoder allows bidirectional transform between V4 and L3



Loss function

$L = \lambda$ mse1 + (1- λ) mes2

Operation	kernel	stride	Features	padding
Input $OriginConv3$	-	-	(12,12,384)	-
Convolution	1×1	1×1	32	0
Latent	-	-	784	-
Dense	-	-	4608	-
Reshape	-	-	(12,12,32)	-
Transposed Convolution	1×1	1×1	384	0
Input $ReconConv3$	-	-	(12,12,384)	-


Object recognition by AlexNet
Flower
Conv1
Conv2
Conv3
Conv4
Object recognition by V4 neurons
Flower

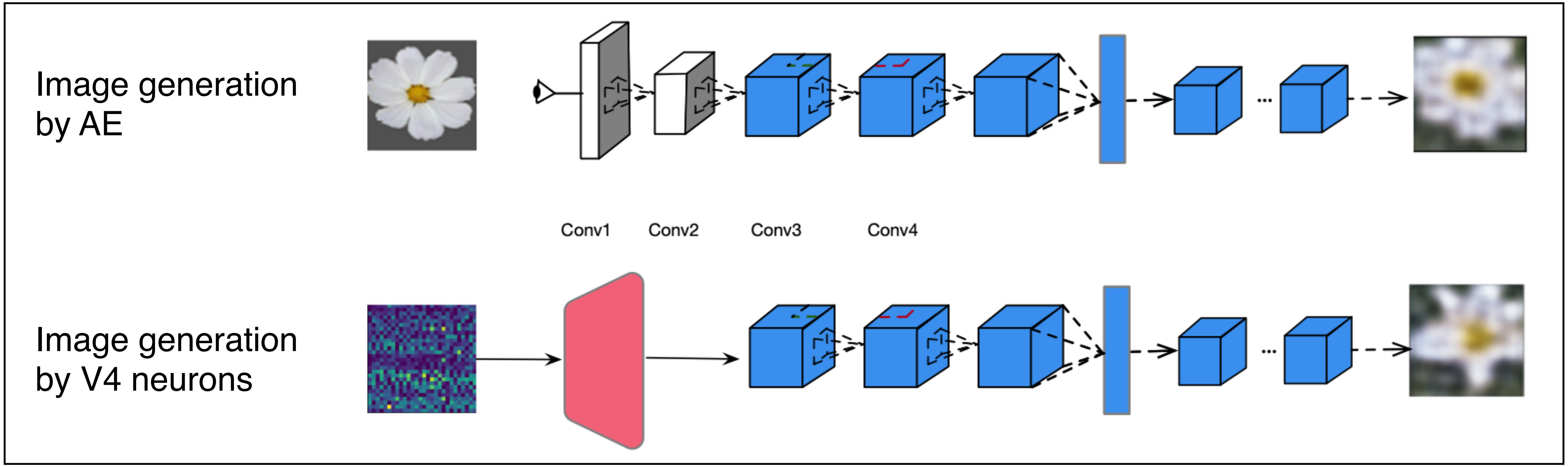
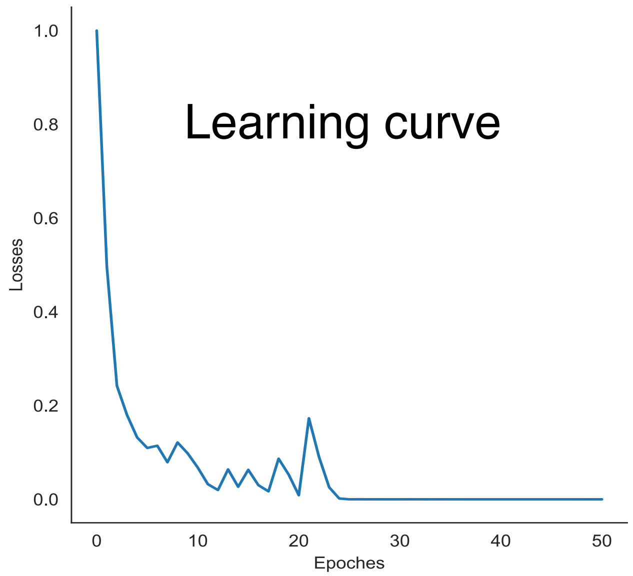


Image generation by AE
Conv1
Conv2
Conv3
Conv4
Image generation by V4 neurons

Preliminary results: AlexNet vs Model with V4



Recognition Performance

Model/Metrics	ACCURAY	AUROC	AUPRC
Alexnet	0.80	0.85	0.78
Model with V4	0.70	0.72	0.68

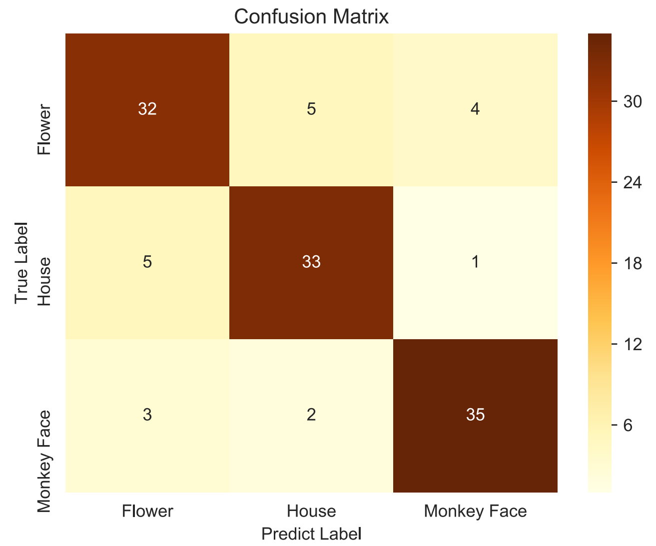
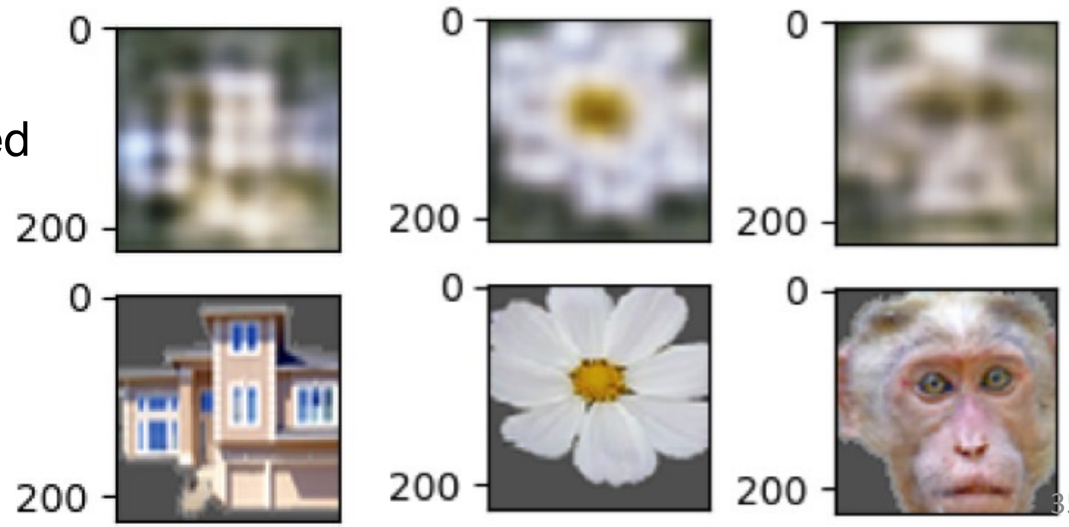


Image reconstruction: examples

Reconstructed

Ground-truth



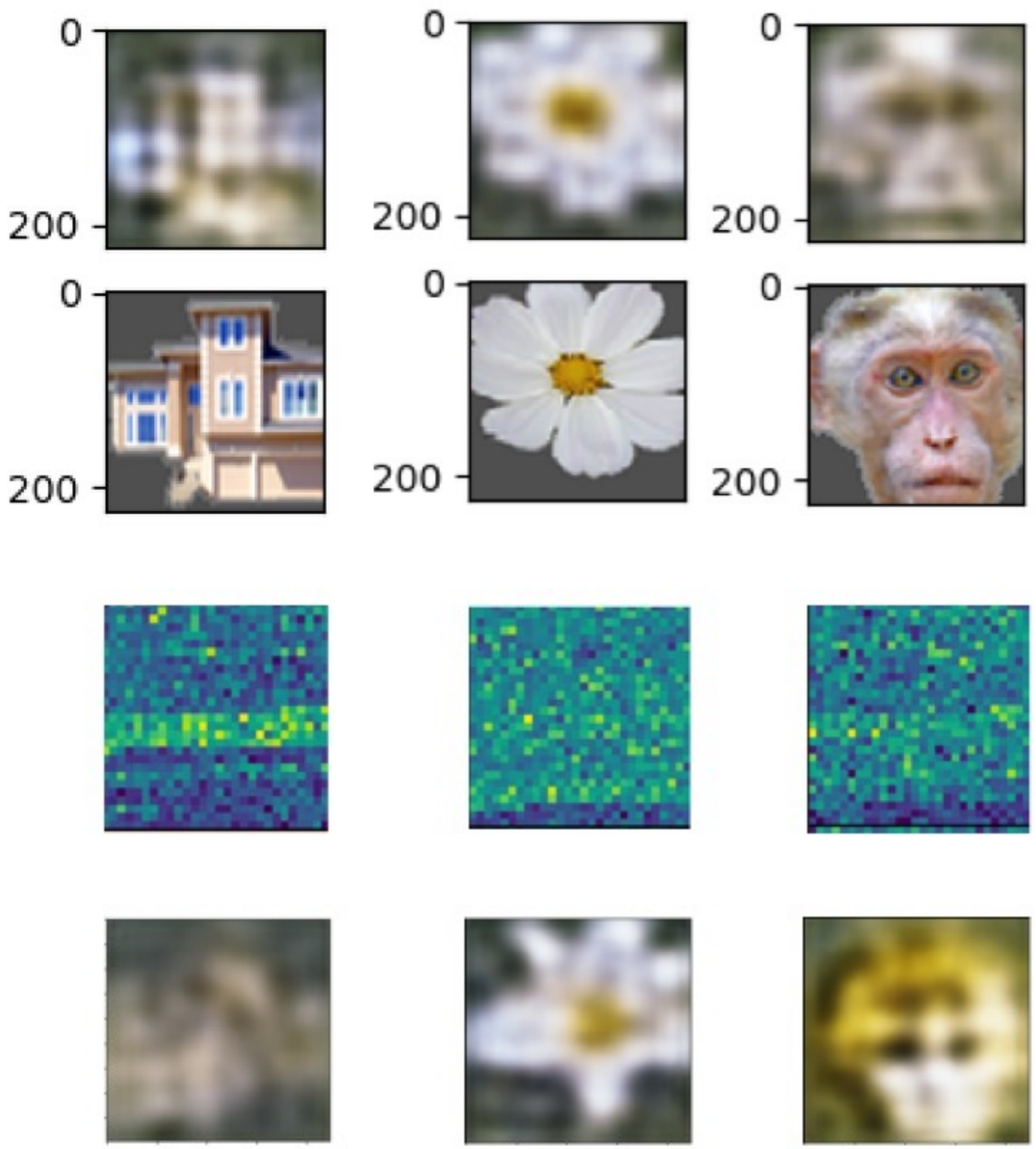
Preliminary results: image reconstruction

Reconstructed images
By AE

Ground-truth images

Firing rate in
biological neurons

Reconstructed images
By biological neurons



Take-home message

- **Ascending dimension transformation via auto-encoder**

  Auto-encoder allows the transformation between biological and artificial neural representations.

- **Generalization**

  The trained Auto-encoder can be generalized to the unseen images.

- **Hybrid AI and brain allows object recognition and image reconstruction**

  It might drive BCI applications to next stage.

- **AI can be a powerful tool for neuroscientists**

  However, it requires collaborations from multidisciplinary fields, including neuroscience, computer science, cognitive science, and psychology.

Other directions
emerging in combining AI and neuroscience

- **Recurrent circuits in brain**
  - Jonas Kubilius et al. (2019), Brain-like object recognition with high-performing shallow recurrent ANNs, NeurIPS
  - Kohitij Kar et al. (2019), Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior, Nature Neurosci
- **Sparsity**
  - Bryan Tripp (2017), Similarities and differences between stimulus tuning in the inferotemporal visual cortex and convolutional networks, IJCNN
  - Qingtian Zhang et al. (2019) A hierarchical sparse coding model predicts acoustic feature encoding in both auditory midbrain and cortex, PLoS Comp Bio
- **Top-down & bottom-up**
  - Sarthak Mittal et al. (2020), Learning to combine top-down and bottom-up signals in recurrent neural networks with attention over modules, ICML
- **Adversarial examples for human and AI**
  - Ian J. Goodfellow et al. (2015), explaining and harnessing adversarial examples, ICLR
  - Gamaleldin F. Elsayed et al. (2018), Adversarial Examples that Fool both Computer Vision and Time-Limited Humans, NeurIPS

Thank you.
Any question is welcome.

Let’s do something together to understand the brain better.

Data

Table 1: Original Datasets

Dataset name / class	ImageNet	Monkey seeing	V4 Neuron	IT Neuron
raining set	10625	70	14000	14000
Testing set	900	24	4800	4800

Table 2: ImageNet: 3 Classes of Image.

Dataset / class	Flower	house	monkey	all
Training set	3425	3600	3600	10625
Testing set	300	300	300	900

Table 3: Monkey seeing:3 Classes Image

Dataset / class	Flower	house	monkey	all
Training set	10	30	30	70
Testing set	4	10	10	24

Table 4: IT Neuron:3 Classes of Neural Firing Rate

Dataset / class	Flower	house	monkey	all
Training set	4400	8800	8800	14000
Testing set	400	800	800	4800

Table 5: V4 Neuron:3 Classes of Neural Firing Rate

Dataset / class	Flower	house	monkey	all
Training set	2000	6000	6000	14000
Testing set	800	2000	2000	4800