A computational framework to unify representation similarity and function in biological and artificial neural networks.

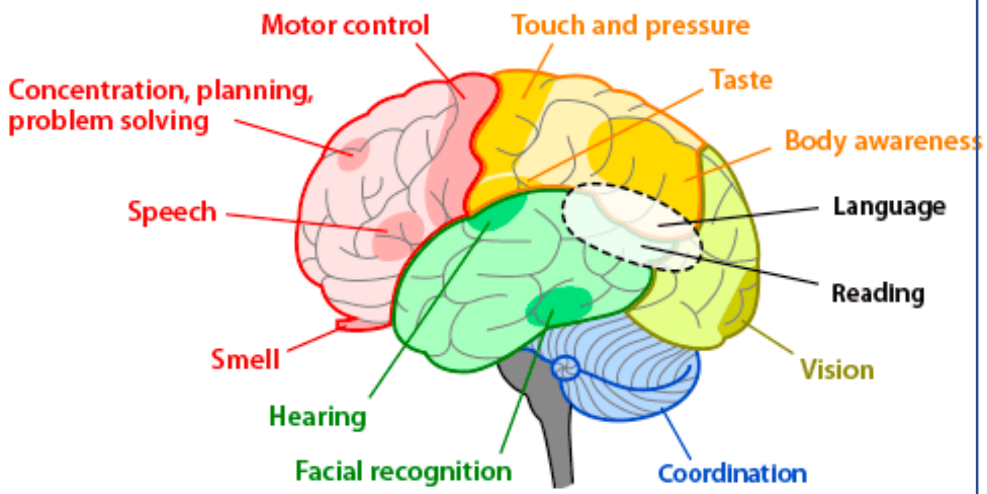
大脑智能与人工智能的融合建模

Xuming Ran（冉旭明）

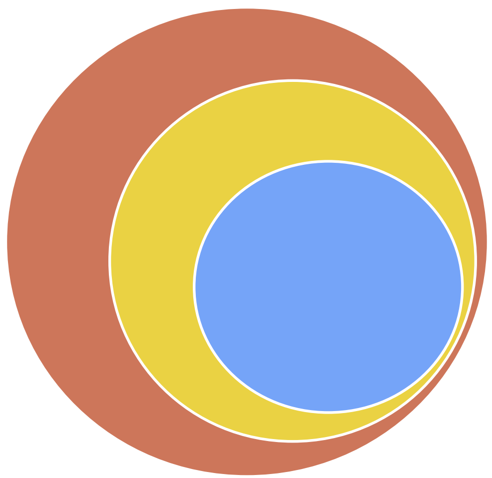
ranxuming@gmail.com

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

Cross-talk between
Brain Intelligence and Artificial Intelligence



Emotion
Social
Sensing
Acting
Learning
Reasoning
Decision making
Language
Reading
Memory
Vision
coordination

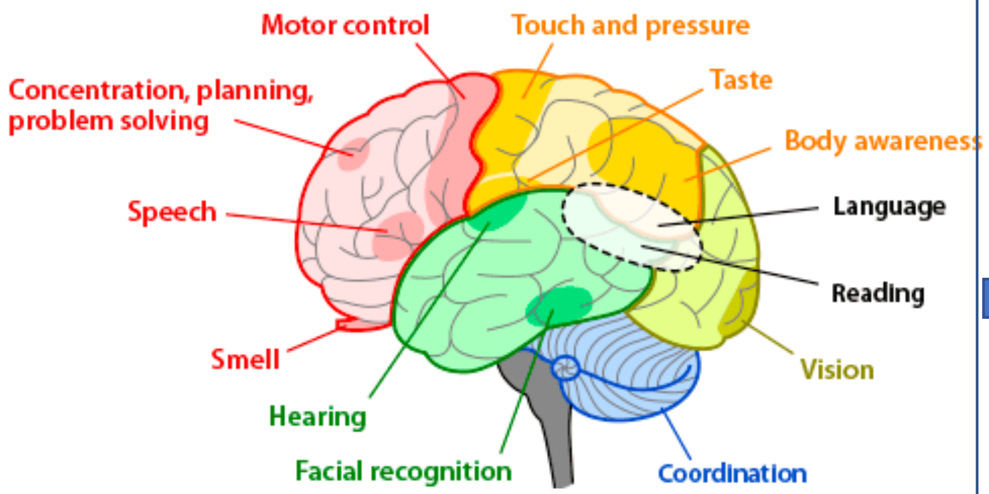


Artificial Intelligence
projects to build non-human intelligence

Machine Learning
machines that learn to be smarter

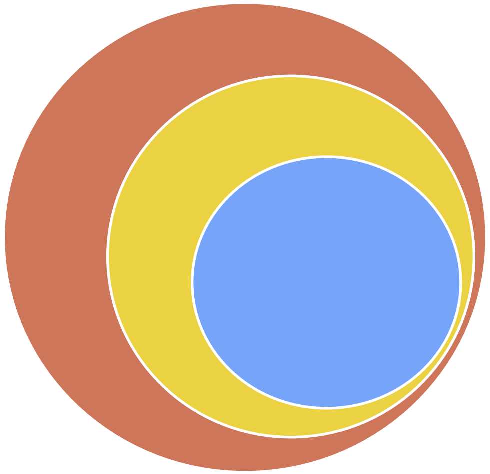
Deep Learning
particular kinds of machine learning

Cross-talk between
Brain Intelligence and Artificial Intelligence



Emotion
Social
Sensing
Acting
Learning
Reasoning
Decision making
Language
Reading
Memory
Vision
coordination

insights for better algorithms

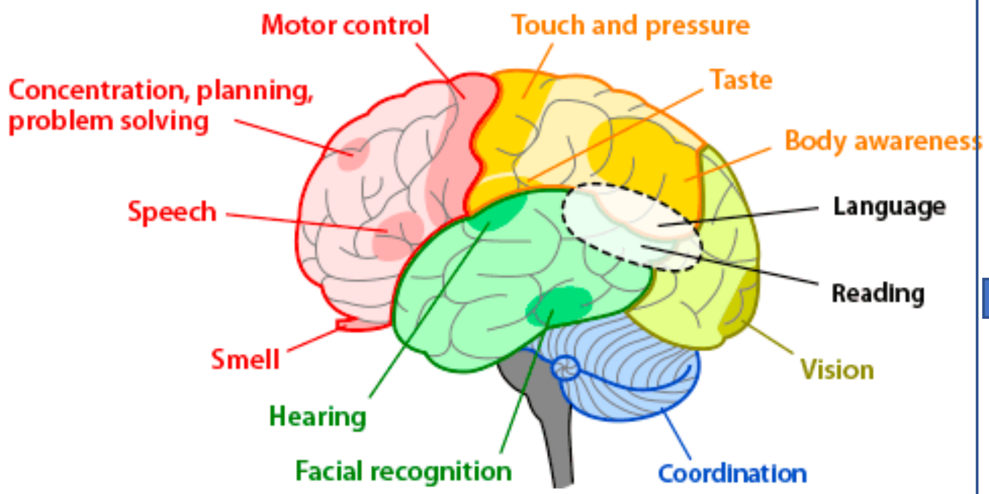


Artificial Intelligence
projects to build non-human intelligence

Machine Learning
machines that learn to be smarter

Deep Learning
particular kinds of machine learning

Cross-talk between
Brain Intelligence and Artificial Intelligence



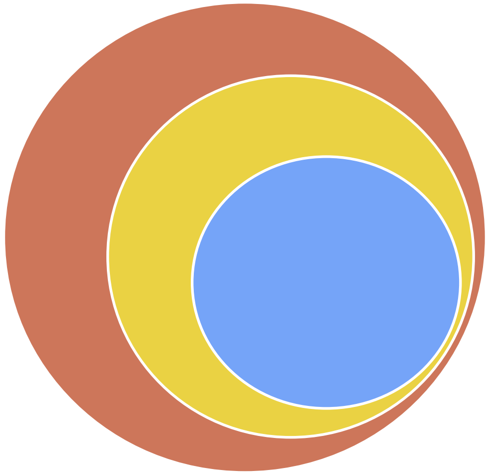
Emotion
Social
Language
Reading
Memory
Vision
coordination

Sensing
Acting
Learning
Reasoning
Decision making

insights for better algorithms

tools for data processing

models for the brain



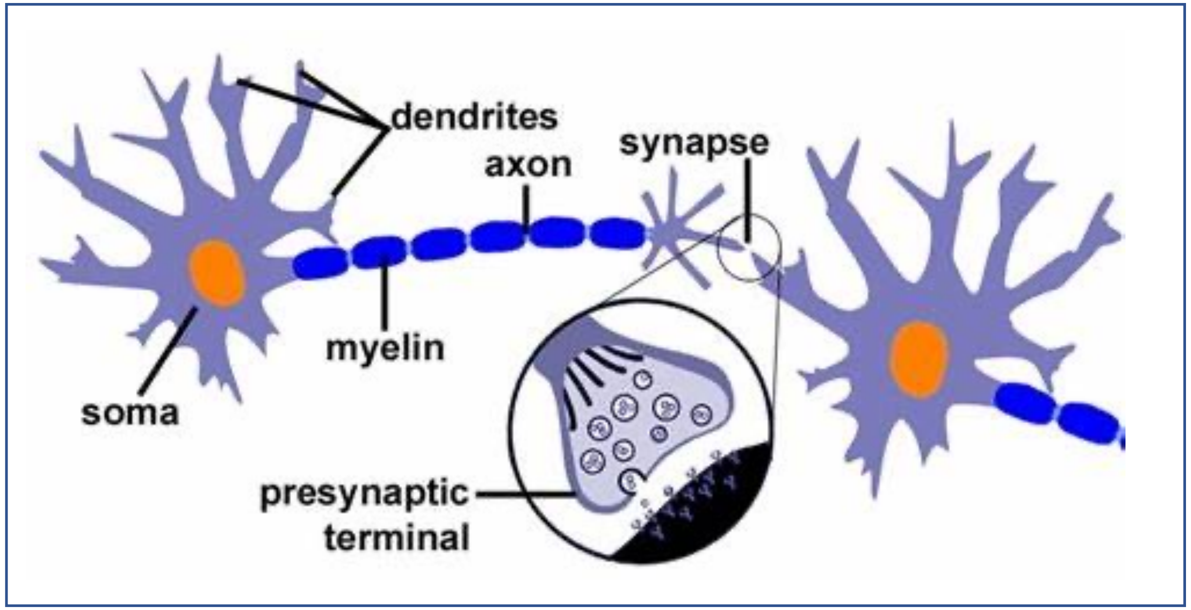
Artificial Intelligence
projects to build non-human intelligence

Machine Learning
machines that learn to be smarter

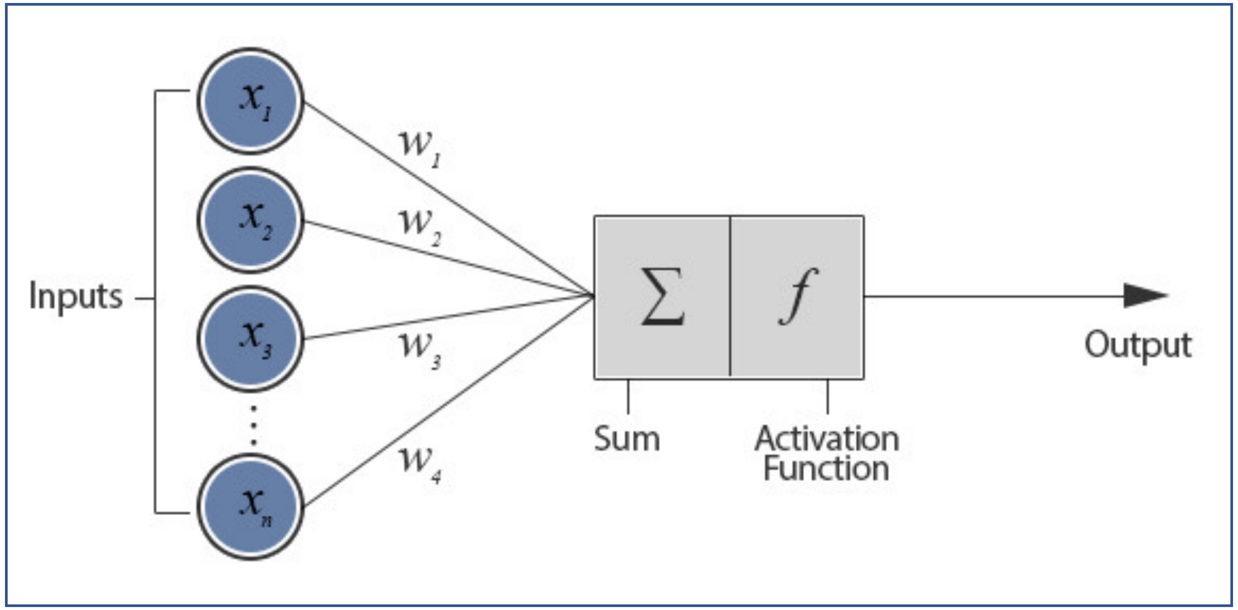
Deep Learning
particular kinds of machine learning

Insights for better algorithms

Biological Neuron

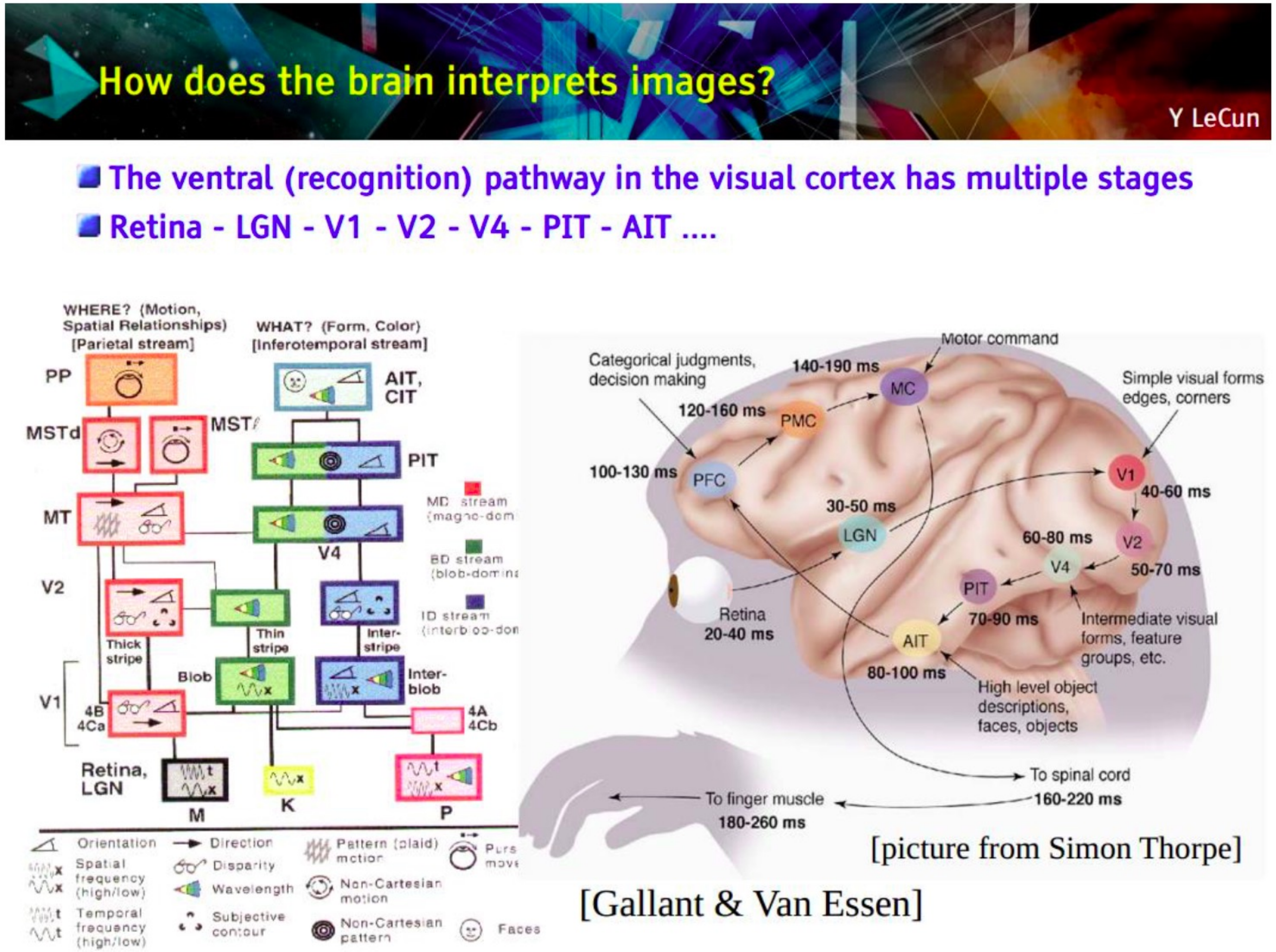


Artificial Neuron



Warren S. McCulloch and Walter Pitts (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*

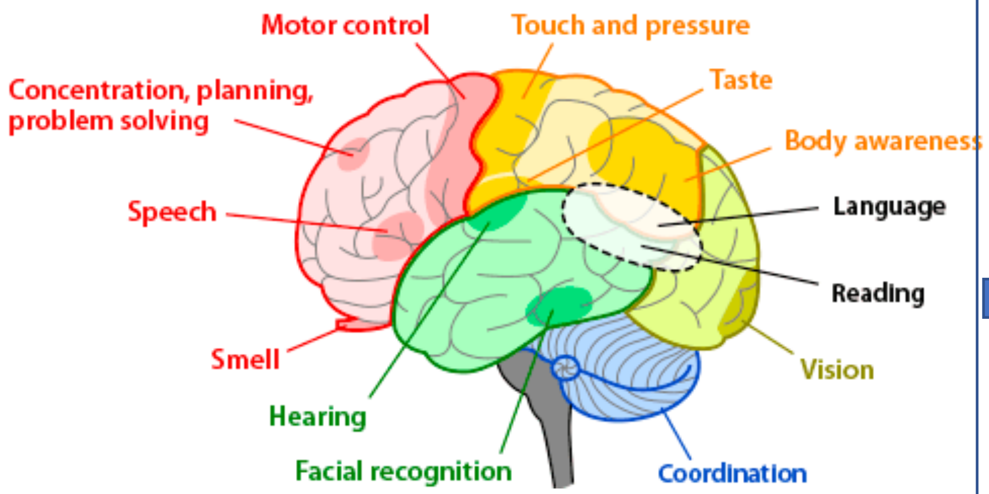
Insights for better algorithms



The way our brain process images inspire the development of convolutional neural network (CNN).

1. Local receptive fields
2. Shared weights
3. Sub-sampling
4. **Layered architecture**

Cross-talk between
Brain Intelligence and Artificial Intelligence

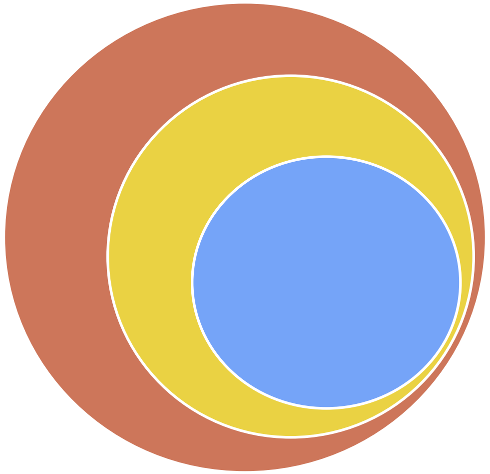


Emotion
Social
Sensing
Acting
Learning
Reasoning
Decision making
Language
Reading
Memory
Vision
coordination

insights for better algorithms

tools for data processing

models for the brain



Artificial Intelligence
projects to build non-human intelligence

Machine Learning
machines that learn to be smarter

Deep Learning
particular kinds of machine learning

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

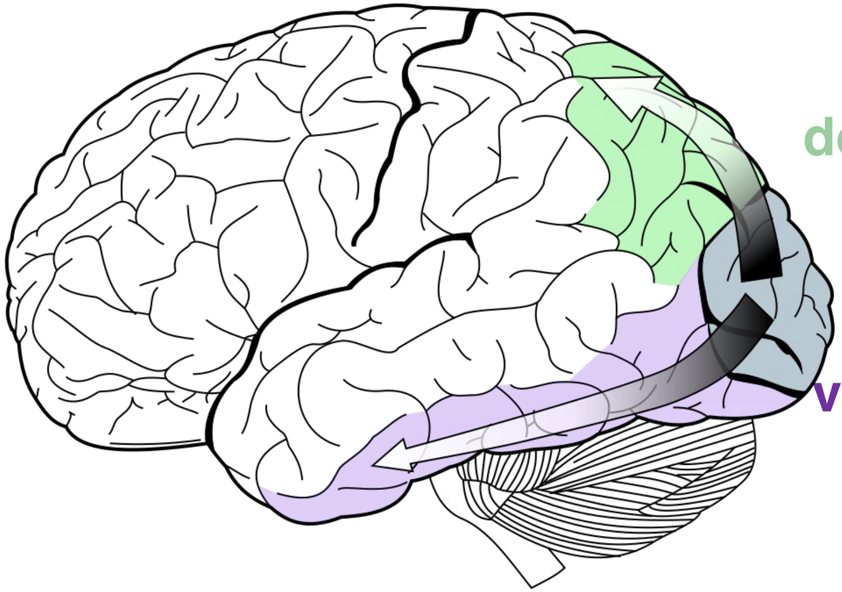
- Not possible to record all neurons
- No enough training datasets
- No enough computational power to fit data

- 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.

Visual pathways (perceptual integration)

-- the dorsal and ventral streams



dorsal 'where' pathway

- **the motion & spatial location**
- **V1, V2, V3, MT (V5), MST & inferior parietal cortex**

&

ventral 'what' pathway

- **the detailed features, form & object identity**
- **V1, V2, V4 & inferior temporal areas**

Goodale MA, Milner AD (1992). Separate visual pathways for perception and action. *Trends Neurosci*

Insights for better algorithms

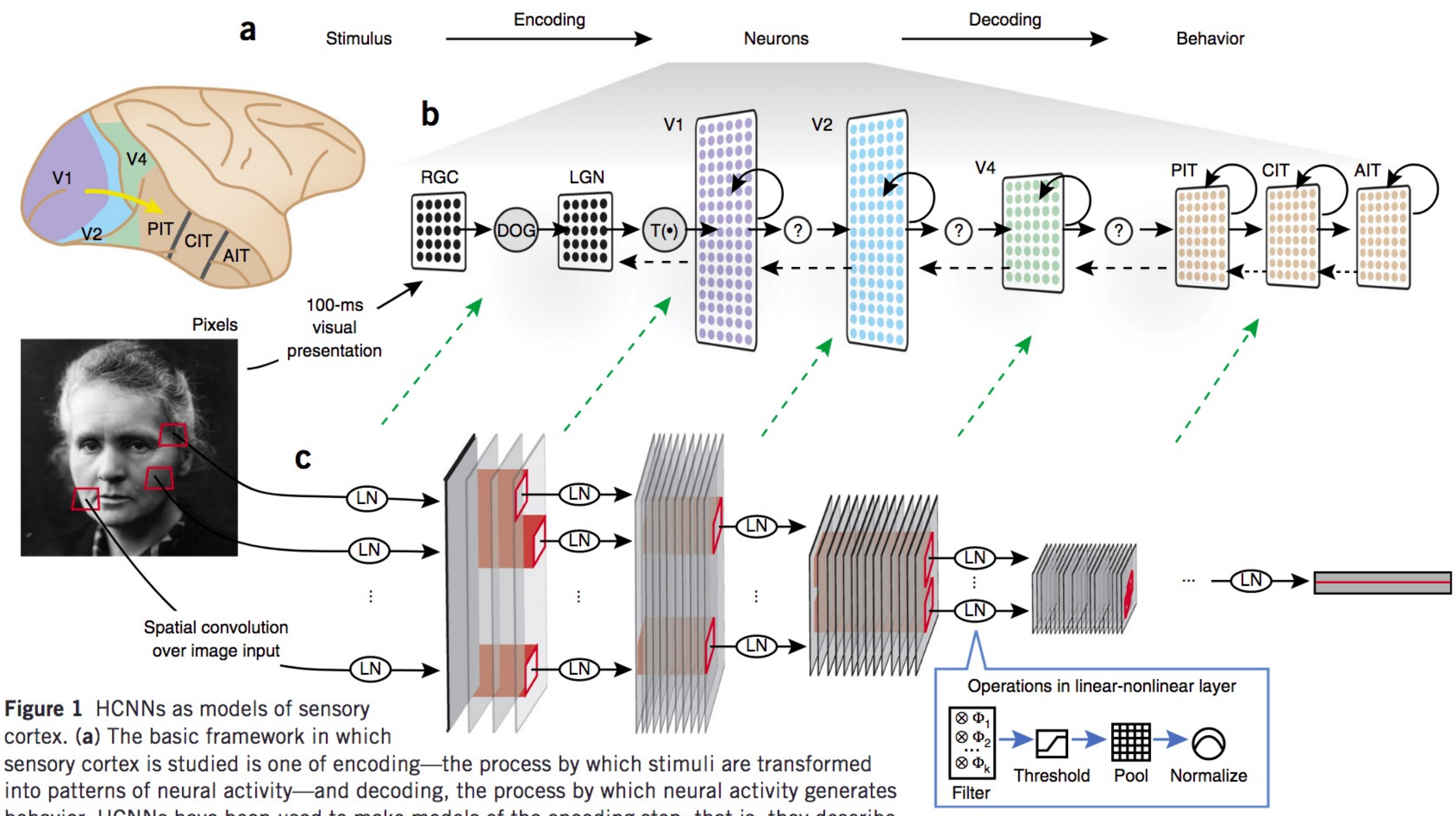


Figure 1 HCNNs as models of sensory cortex. (**a**) The basic framework in which sensory cortex is studied is one of encoding—the process by which stimuli are transformed into patterns of neural activity—and decoding, the process by which neural activity generates behavior. HCNNs have been used to make models of the encoding step; that is, they describe

DiCarlo and Cox (2007) TiCS; Yamins (2016) nature neurosci; Bashivan (2019) Science

Brain score: how well existing AI models explain the neural data

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

Sort by average score

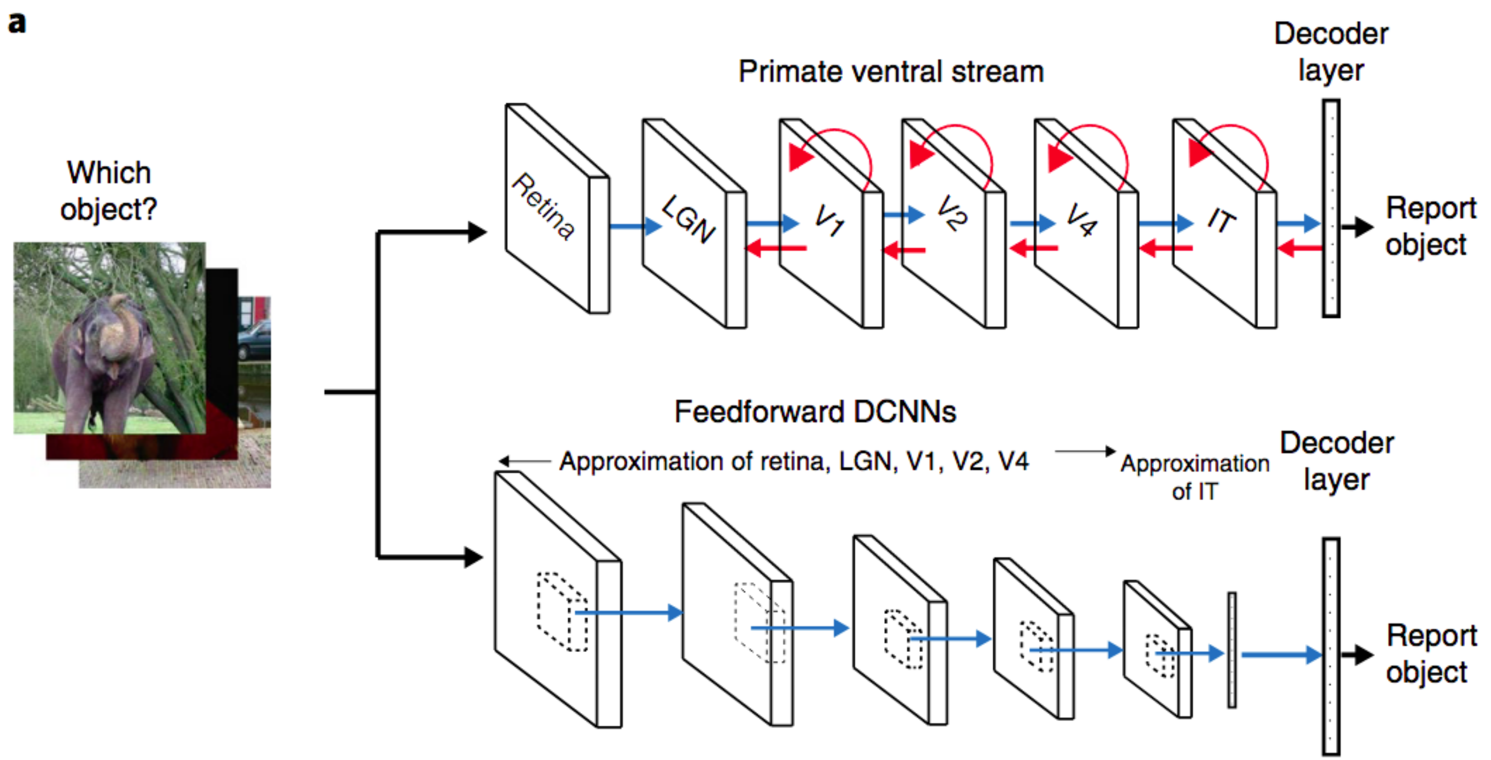
Rank	Model	average	V1	V2	V4	IT	behavior	engineering	Deng2009-top1
1	efficientnet-b0 *Tan et al., 2019*	.442	.215	.317	.556	.547	.573		
2	efficientnet-b6 *Tan et al., 2019*	.435	.263	.295	.563	.541	.513		
3	efficientnet-b2 *Tan et al., 2019*	.434	.213	.317	.569	.547	.526		
4	efficientnet-b4 *Tan et al., 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

Deep learning as computational models to understand the brain

Train **BOTH** the **monkey** and **ANNs** to perform the same task (object discrimination task) involving challenging naturalistic visual objects.

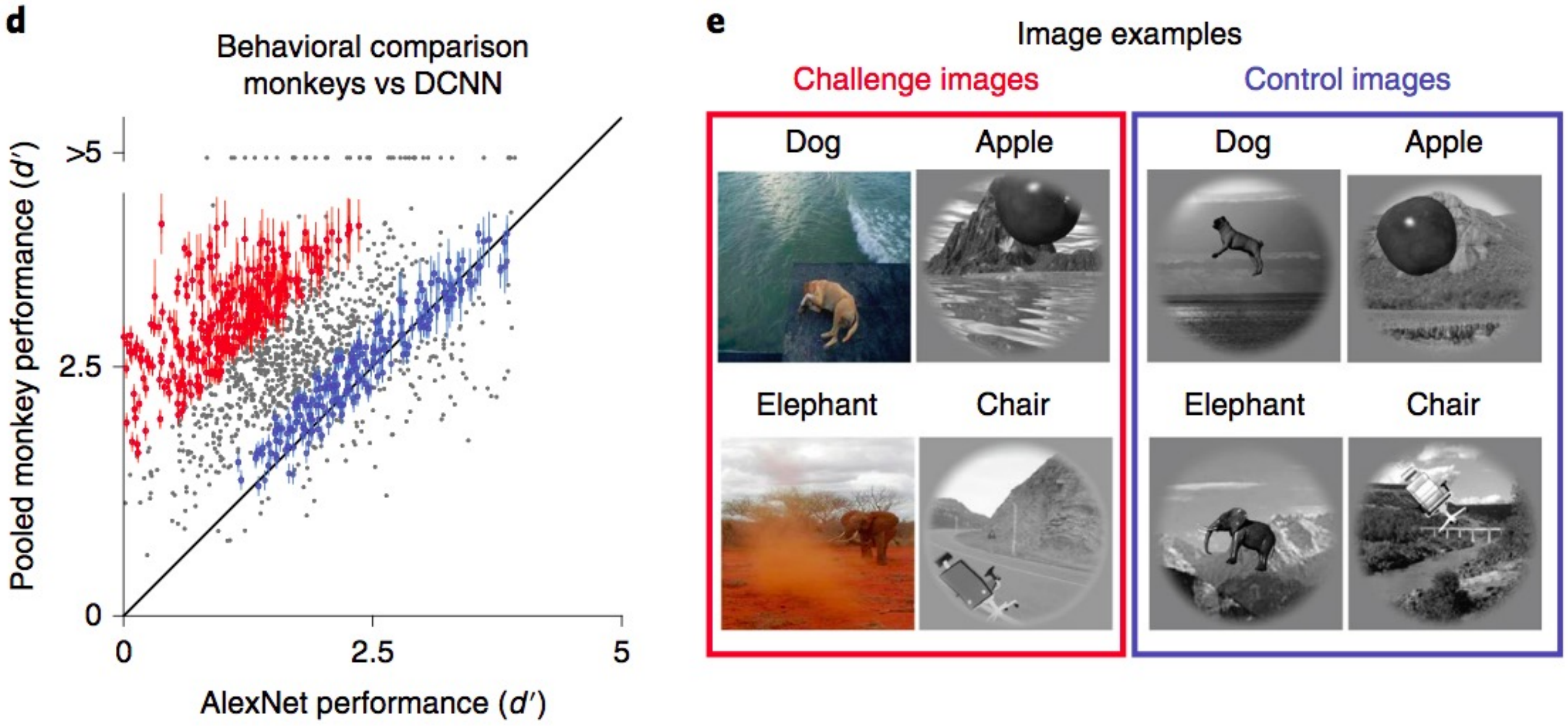


Compared to purely feedforward networks, recurrently-connected deep networks are better at predicting responses of higher visual area neurons to behaviorally challenging images.

K. Kar, J. Kubilius, K. Schmidt, E. B. Issa, and J. J. DiCarlo. (2019) Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior. *Nature neuroscience*

Deep learning as computational models to understand the brain

Train **BOTH** the **monkey** and **ANNs** to perform the same task (i.e., object discrimination task) involving challenging naturalistic visual objects.



AI helps explain the computational benefit or necessity of observed brain structures or functions.

K. Kar, J. Kubilius, K. Schmidt, E. B. Issa, and J. J. DiCarlo. (2019) Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior. *Nature neuroscience*

Q1. Can information from the brain help AI?
Q2. Is an AI model with better task performance more similar to the brain?

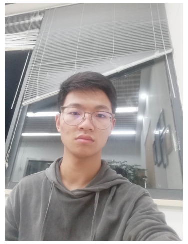
Assumption:

- An AI model that is contrained to predict the brain activity will gain some knowledge that will make it better at its task.

Core ideas:

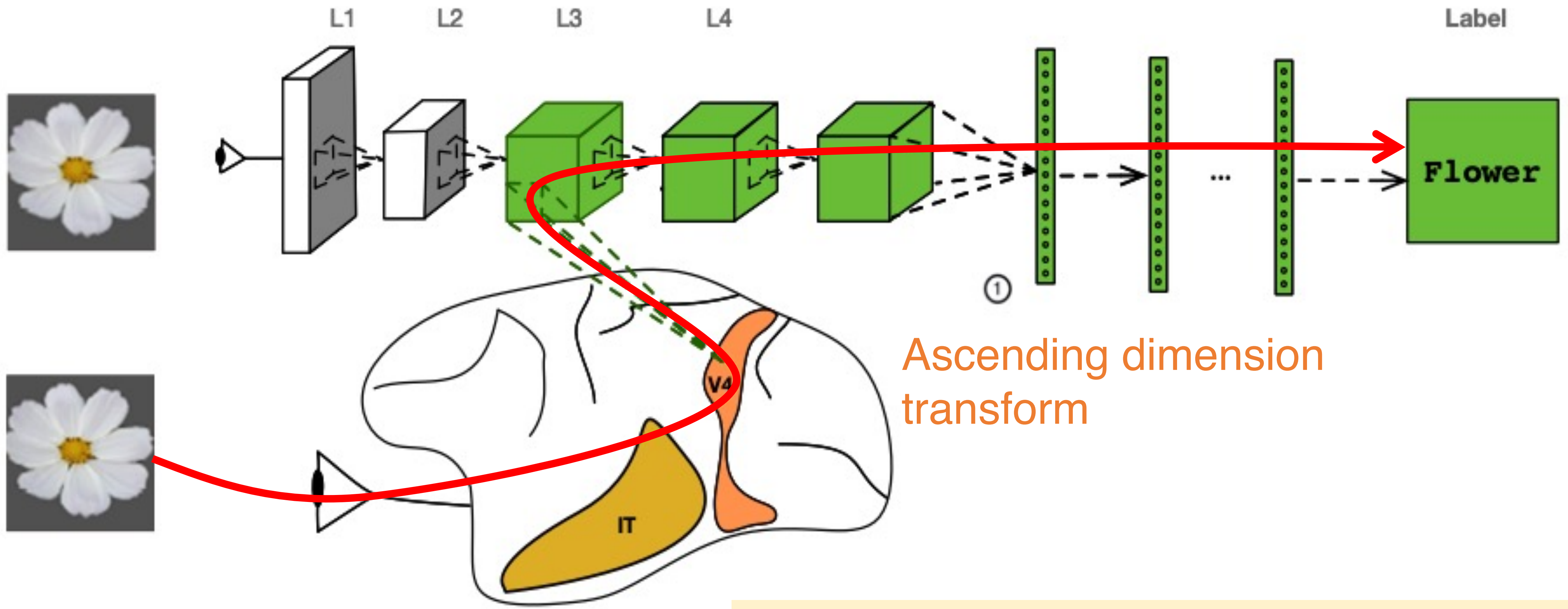
- By maximizing the representation similarity between biological neurons and artificial neurons, the AI model (implicitly) picks up (some of) the computations by the brain .

- By adding useful information from the brain into AI model, the task performance would improve.



冉旭明

Mapping function: from V4 in the brain to L3 of AlexNet



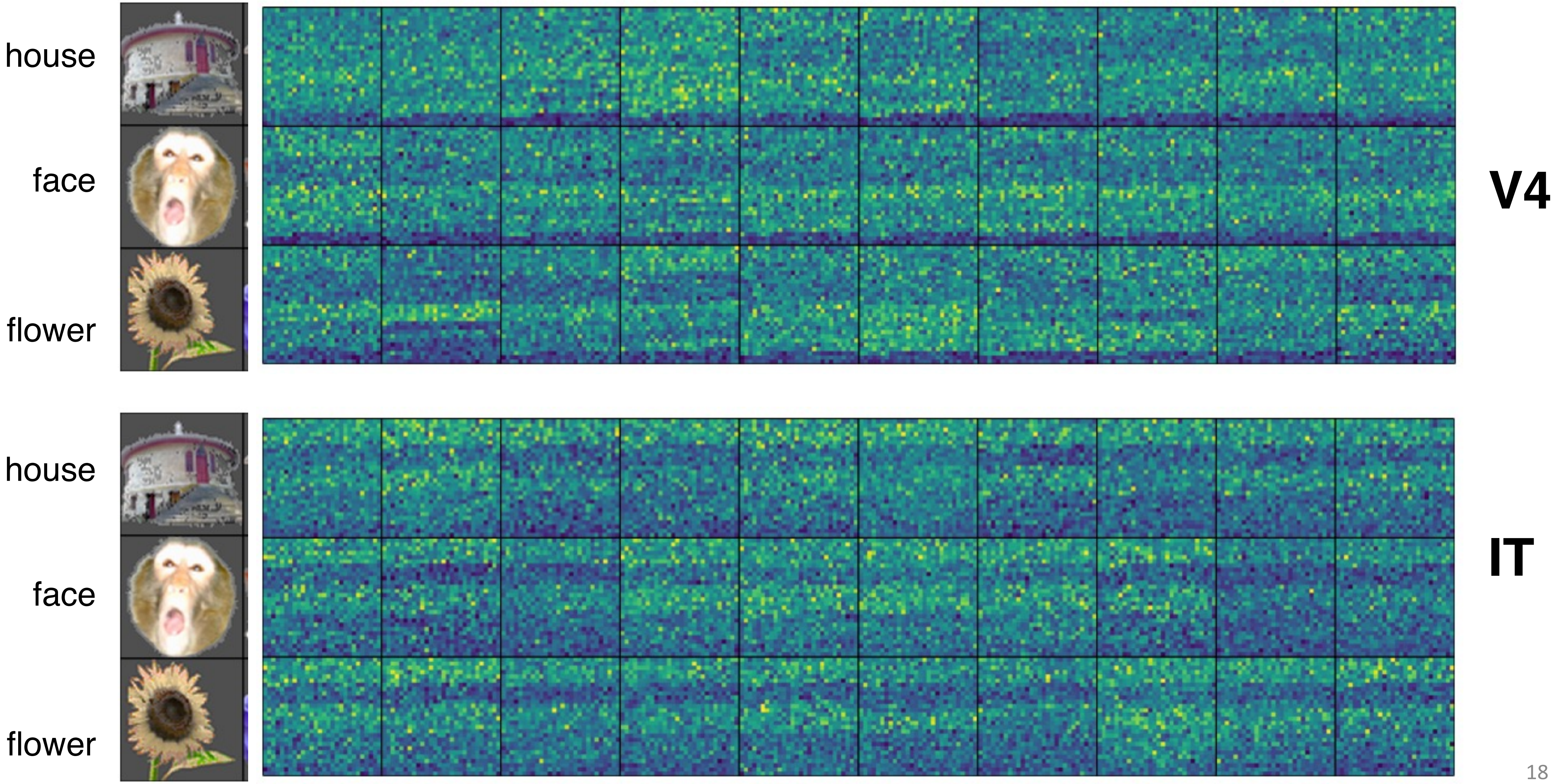
Ascending dimension transform

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

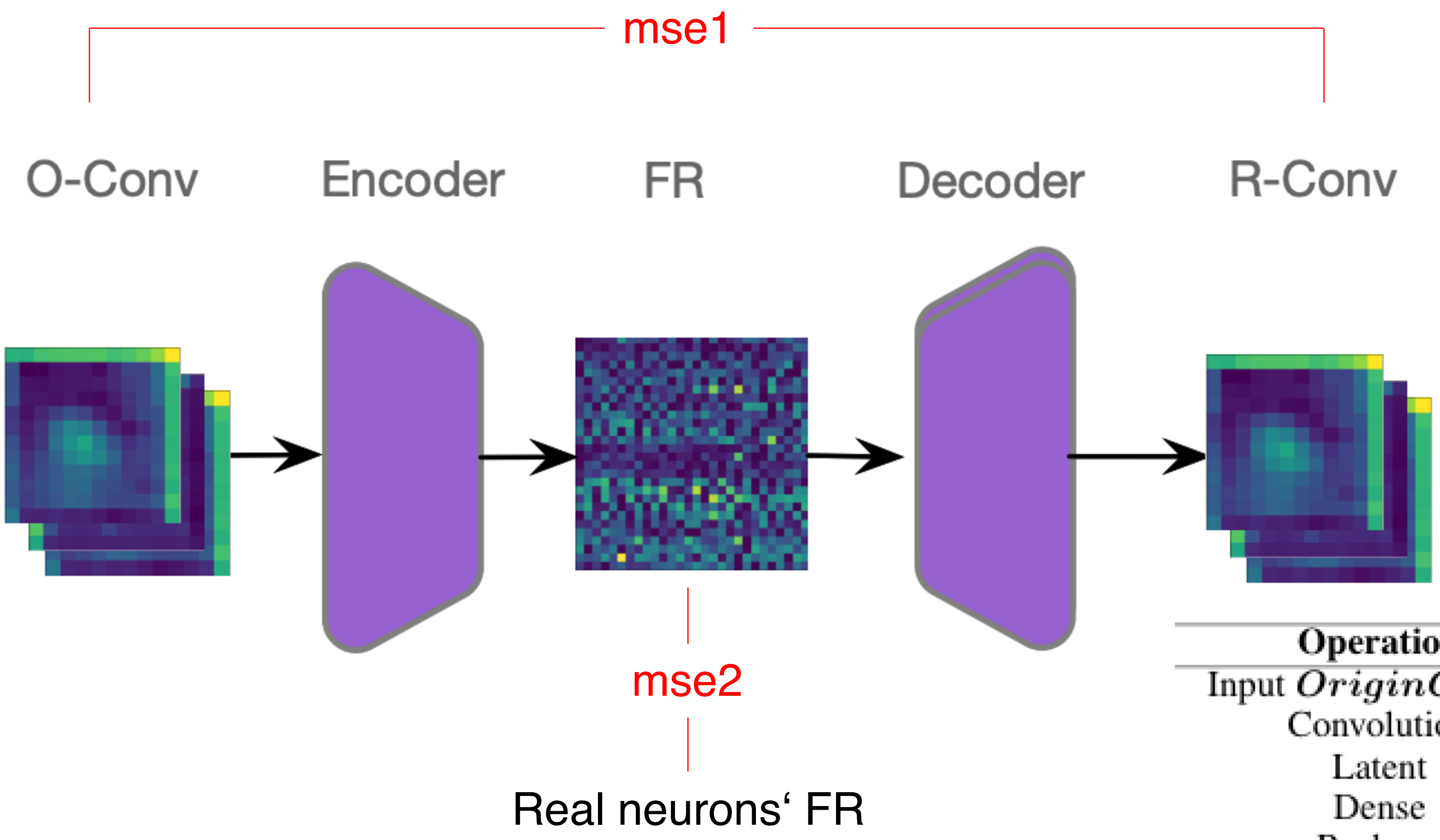
Image presented to monkeys

ImageNet to pre-train Alexnet

Neural responses in V4 and IT



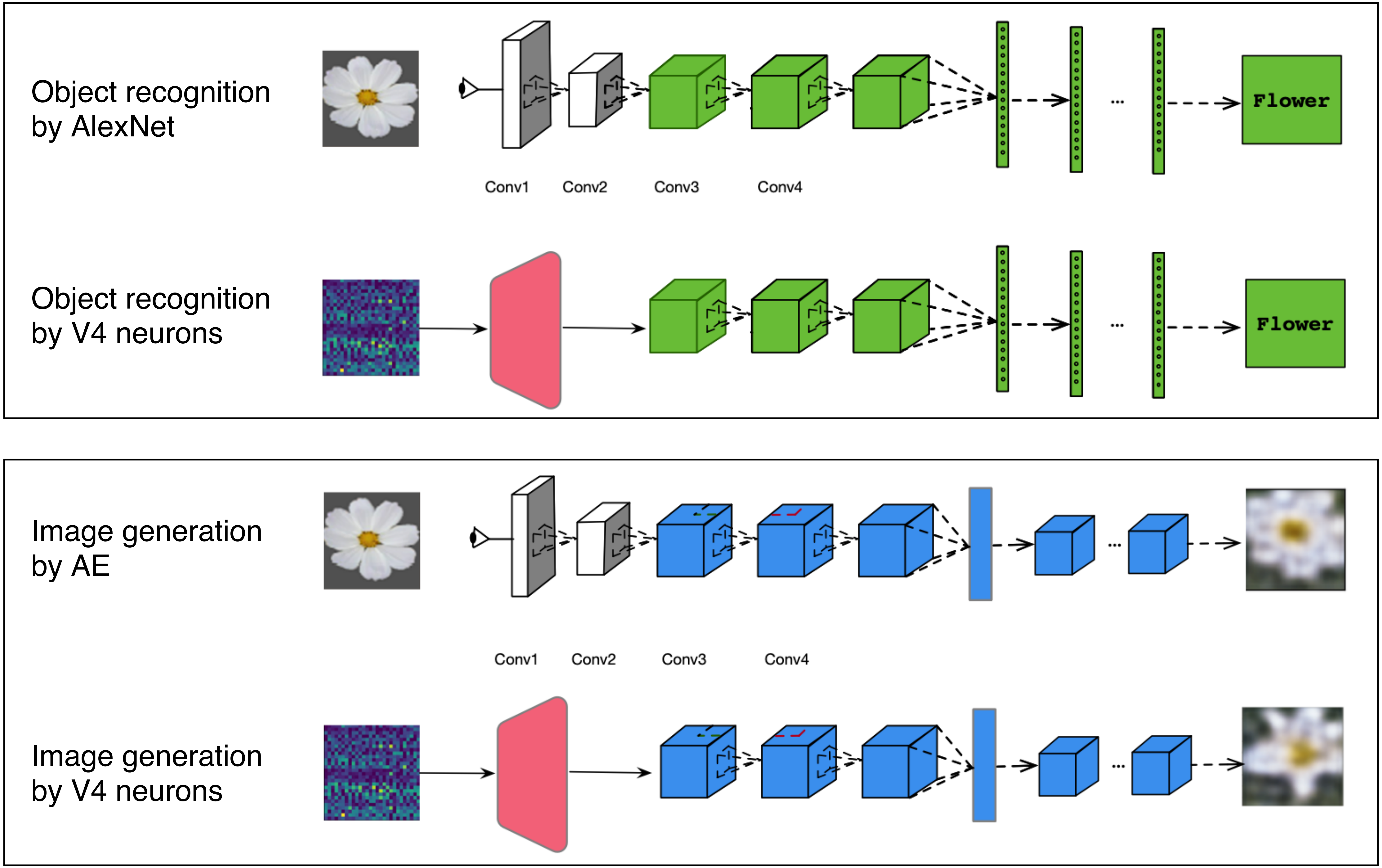
Autoencoder allows bidirectional transform between V4 and L3



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
Conv1
Conv2
Conv3
Conv4
Flower
Object recognition by V4 neurons
Flower
Image generation by AE
Conv1
Conv2
Conv3
Conv4
Image generation by V4 neurons

Preliminary results: image reconstruction

Reconstructed images
By AE

Ground-truth images

Firing rate in
biological neurons

Reconstructed images
By biological neurons

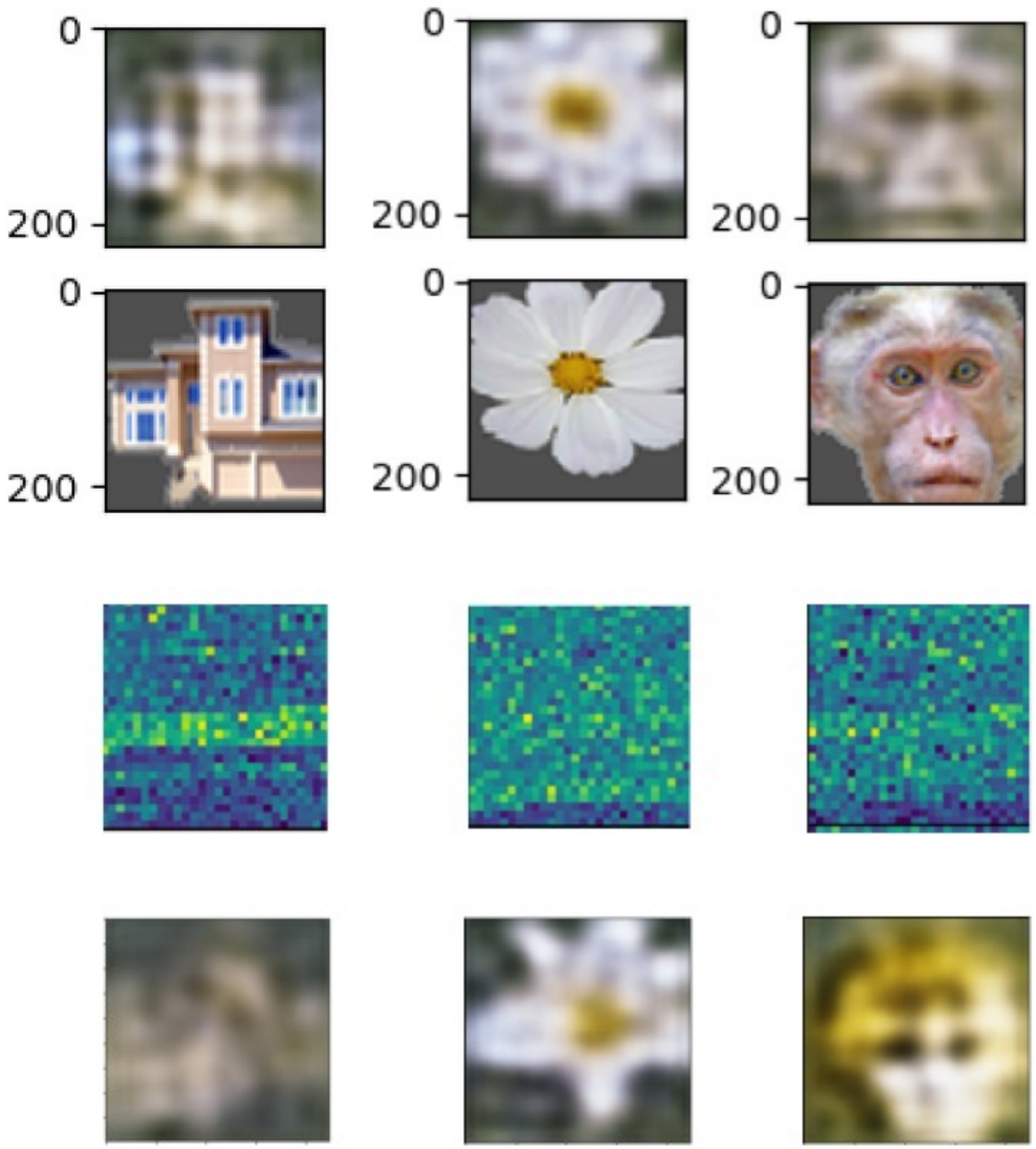


Image reconstruction via biological neural responses is not good.

-->

Image reconstruction via combining DAE and neural responses.

DAE-NR: Deep Auto-encoder with Neural Response

Deep Auto-encoder with Neural Response

Xuming Ran[1], **Jie Zhang**[2,3], **Ziyuan Ye**[1], **Haiyan Wu**[4], **Qi Xu**[5], **Huihui Zhou**[6,*] and **Quanying Liu**[1,*]
[1]Southern University of Science and Technology, China
[2]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China
[3]University of the Chinese Academy of Sciences, China
[4]Centre for Cognitive and Brain Sciences and Department of Psychology, University of Macau
[5]School of Artificial Intelligence, Dalian University of Technology, China
[6]Pengcheng Laboratory, China

Abstract

Artificial intelligence and neuroscience are deeply interactive. Artificial neural networks (ANNs) have been a versatile tool to study the neural representation in the ventral visual stream, and the knowledge in neuroscience in return inspires ANN models to improve performance in the task. However, how to merge these two directions into a unified model has less studied. Here, we propose a hybrid model, called deep auto-encoder with the neural response (DAE-NR), which incorporates the information from the visual cortex into ANNs to achieve better image reconstruction and higher neural representation similarity between biological and artificial neurons. Specifically, the same visual stimuli (i.e., natural images) are input to both the mice brain and DAE-NR. The DAE-NR jointly learns to map a specific layer of the encoder network to the biological neural responses in the ventral visual stream by a mapping function and to reconstruct the visual input by the decoder. Our experiments demonstrate that if and only if with the joint learning, DAE-NRs can (i) improve the performance of image reconstruction and (ii) increase the representational similarity between biological neurons and artificial neurons. The DAE-NR offers a new perspective on the integration of computer vision and visual neuroscience.

普通DAE模型
片重构任务的
baseline模型

CNN-FR模型
表征相似性任务
baseline模型

DAE-NR模型同
时兼具2个任务
的优化性能

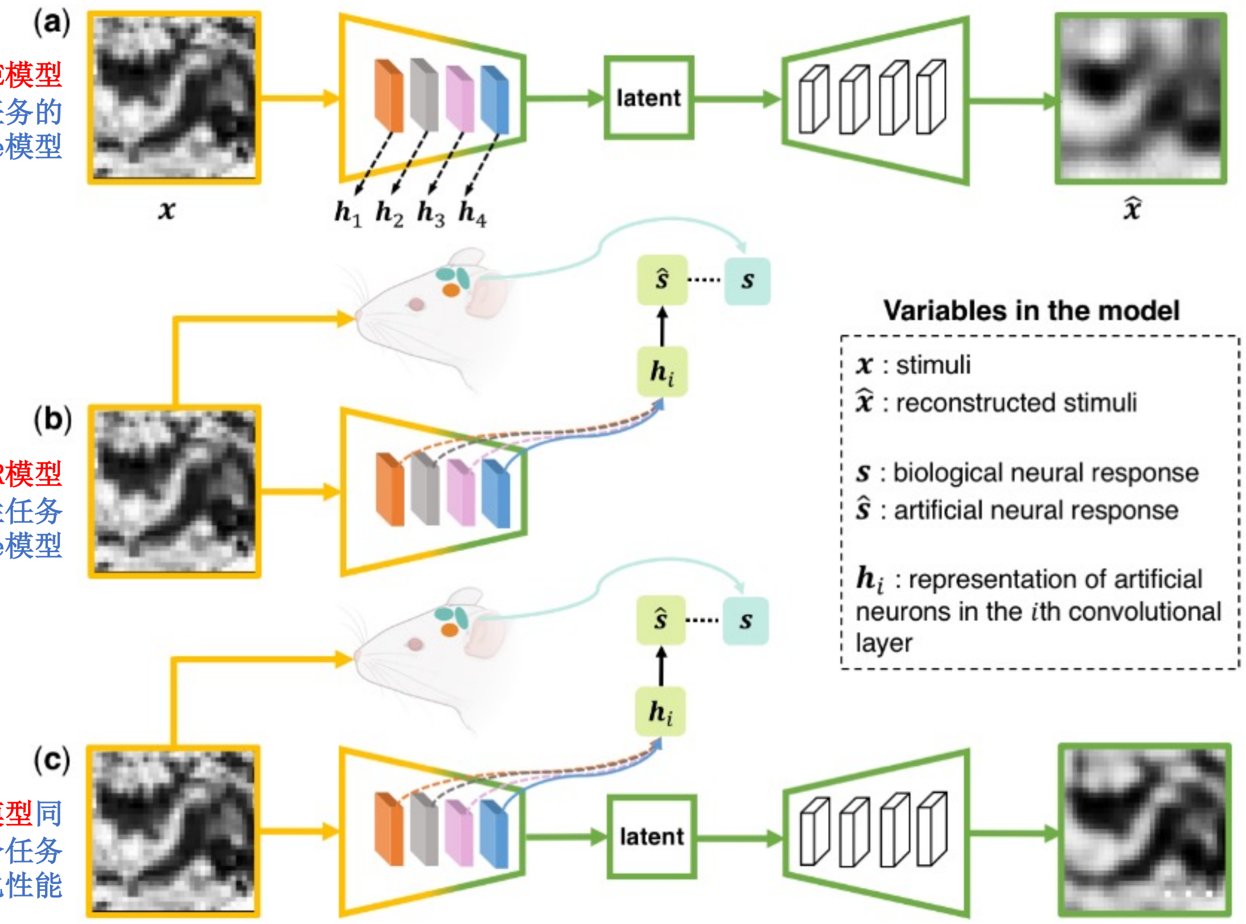


Figure 1: The illustration of the model of (a) the standard deep auto-encoder (DAE) for images reconstruction; (b) the convolutional neural network with factorized readout (CNN-FR) for prediction of neuron responses; (c) the DAE with the neuron response (DAE-NR) for images reconstruction and predictions of neuron responses. s is the biological neural response, the prediction of biological neural response is represented as $\hat{s}$, and $\mathbf{h}_i$ ($i \in \{1,2,3,4\}$) is the feature of the ith convolutional layer.

Reconstructed images - some examples

(a) 原始图片；

(b) 普通DAE；

(c-f) DAE-NR
将神经信息导入
DAE的1-4层

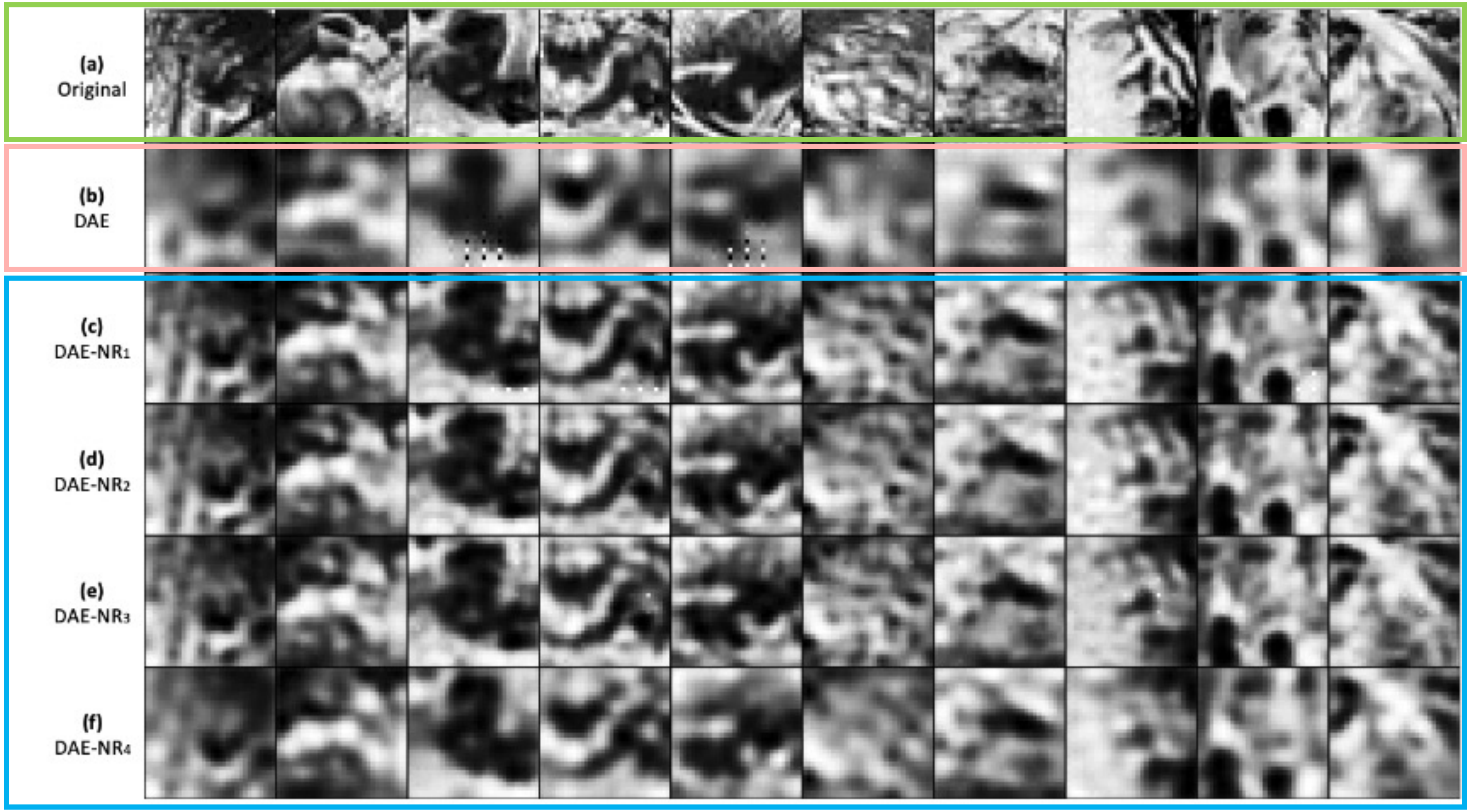


Figure 2: The reconstructed images with neurons in Region 3. From top to bottom, each row displays the original images (a), the images reconstructed by DAE (b), DAE-NR$_1$ (c), DAE-NR$_2$ (d), DAE-NR$_3$ (e), DAE-NR$_4$ (f), respectively

Reconstructed images - Quantification

Table 1: 图片重构任务的量化结果。DAE-NR能提高图片重构的质量

Table 1: The quantitative results of image reconstruction with all neurons in the region 1, 2, and 3, respectively.

Model	Region 1			Region 2			Region 3		
	MSE↓	PSNR↑	SSIM↑	MSE↓	PSNR↑	SSIM↑	MSE↓	PSNR↑	SSIM↑
DAE	0.022	23.709	0.771	0.024	23.338	0.754	0.081	17.039	0.561
DAE-NR_1	**0.021**	**23.829**	**0.776**	**0.023**	23.392	0.753	0.044	19.751	0.763
DAE-NR_2	**0.021**	23.779	0.775	**0.023**	23.440	0.759	**0.043**	**19.819**	**0.764**
DAE-NR_3	**0.021**	23.778	0.775	0.024	23.330	0.755	**0.043**	19.789	0.761
DAE-NR_4	0.022	23.721	0.773	**0.023**	**23.491**	**0.760**	0.059	18.462	0.668

Reconstructed images - Quantification

Table 1: 图片重构任务的量化结果。DAE-NR能提高图片重构的质量

Table 1: The quantitative results of image reconstruction with all neurons in the region 1, 2, and 3, respectively.

Model	Region 1			Region 2			Region 3		
	MSE$\downarrow$	PSNR$\uparrow$	SSIM$\uparrow$	MSE$\downarrow$	PSNR$\uparrow$	SSIM$\uparrow$	MSE$\downarrow$	PSNR$\uparrow$	SSIM$\uparrow$
DAE	0.022	23.709	0.771	0.024	23.338	0.754	0.081	17.039	0.561
DAE-NR$_1$	**0.021**	**23.829**	**0.776**	**0.023**	23.392	0.753	0.044	19.751	0.763
DAE-NR$_2$	**0.021**	23.779	0.775	**0.023**	23.440	0.759	**0.043**	**19.819**	**0.764**
DAE-NR$_3$	**0.021**	23.778	0.775	0.024	23.330	0.755	**0.043**	19.789	0.761
DAE-NR$_4$	0.022	23.721	0.773	**0.023**	**23.491**	**0.760**	0.059	18.462	0.668

Table 2: 加入与人工神经元的表征显著相关的大脑神经元，能促进DAE-NR的图片重构性能

Table 2: The quantitative results of image reconstruction with constraints of significant neurons and insignificant neurons in the region 3.

	MSE$\downarrow$		SSIM$\uparrow$		PSNR$\uparrow$	
Significant	YES	NO	YES	NO	YES	NO
DAE-NR$_1$	**0.043**	0.125	**0.761**	0.332	**19.784**	15.168
DAE-NR$_2$	**0.047**	0.082	**0.743**	0.547	**19.467**	16.970
DAE-NR$_3$	**0.049**	0.116	**0.724**	0.362	**19.245**	15.463
DAE-NR$_4$	0.047	**0.045**	0.740	**0.752**	19.497	**19.628**

Representation similarity - biological & artificial neurons

DAE-NR与DAE的比较

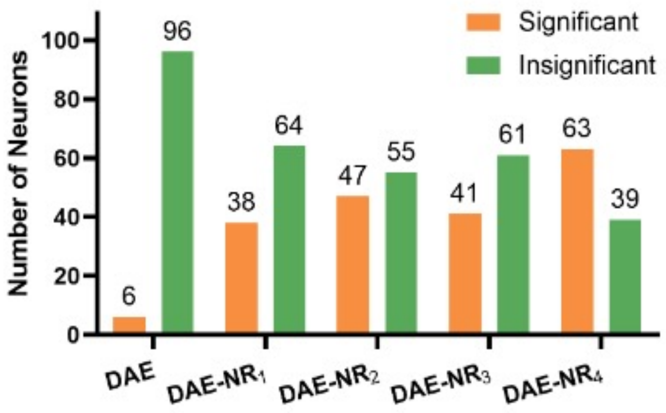


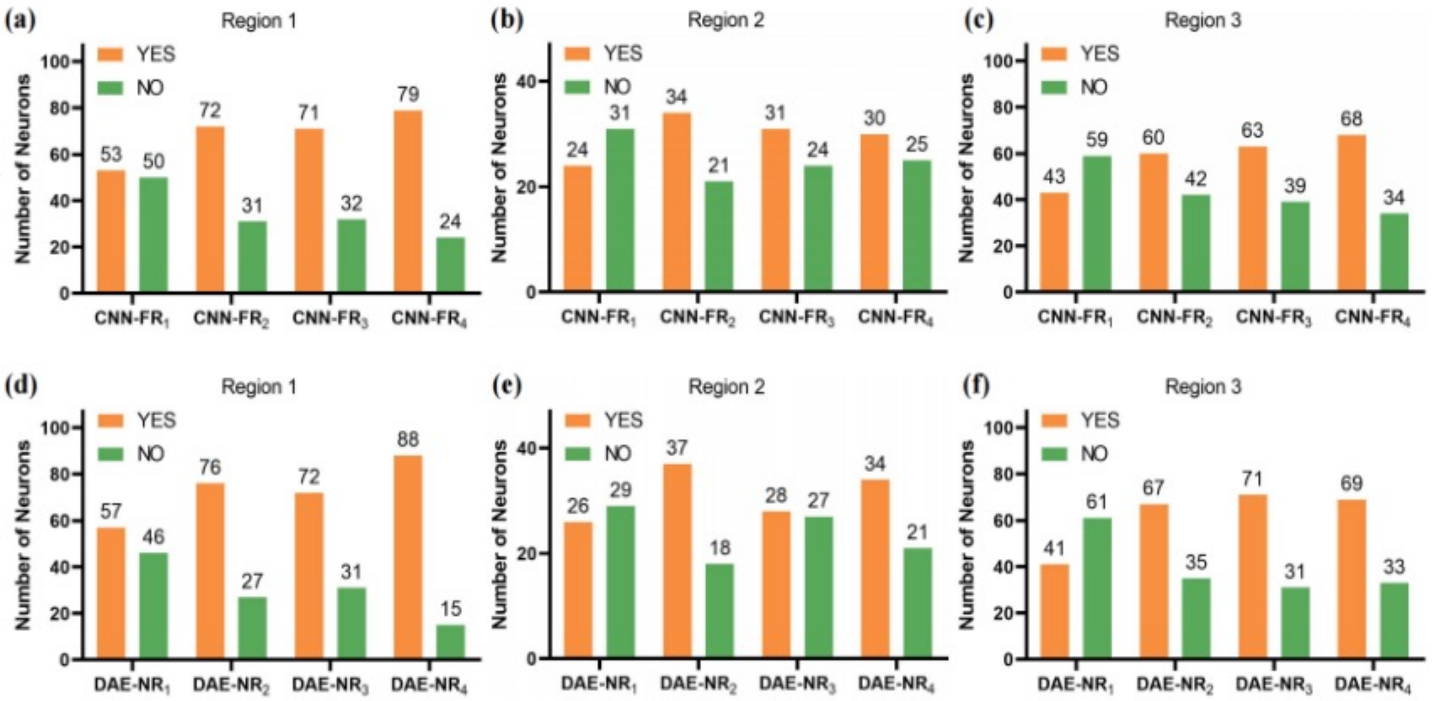
Figure 3: The number of significant neurons and insignificant neurons of region 3 in the image reconstruction experiments. The threshold for significance is $p \leq 0.05$.

人工神经元的表征显著相关的大脑神经元的数量，与DAE（6个）相比， DAE-NR （38-47个）中数量增多了。

DAE-NR与CNN-FR的比较

Baseline 模型 (CNN-FR)

我们的模型 (DAE-NR)



Mice的三个脑区（都在V1）的神经元表征 与 AI不同层神经元表征的相似性比较，DAE-NR中显著相关的神经元数量比较多。

结论：DAE-NR能使AI和BI之间有更强的神经表征相似性

Take-home message

- We propose a novel model called Deep Autoencoder with Neural Response (DAE-NR). It brings the neural information into DAE, which can simultaneously learn to predict neural responses and to reconstruct the visual stimuli.

- DAE-NR can improve the image reconstruction quality with the help of a Poisson loss on the predicted neural activity, compared to the traditional DAE models.

- DAE-NR provides higher representation similarity between artificial neurons and biological neurons, compared to the end-to-end computational neuroscience model without the image reconstruction task (i.e., CNN-FR).

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 AI & brain better.