Homologies between brains and CNNs

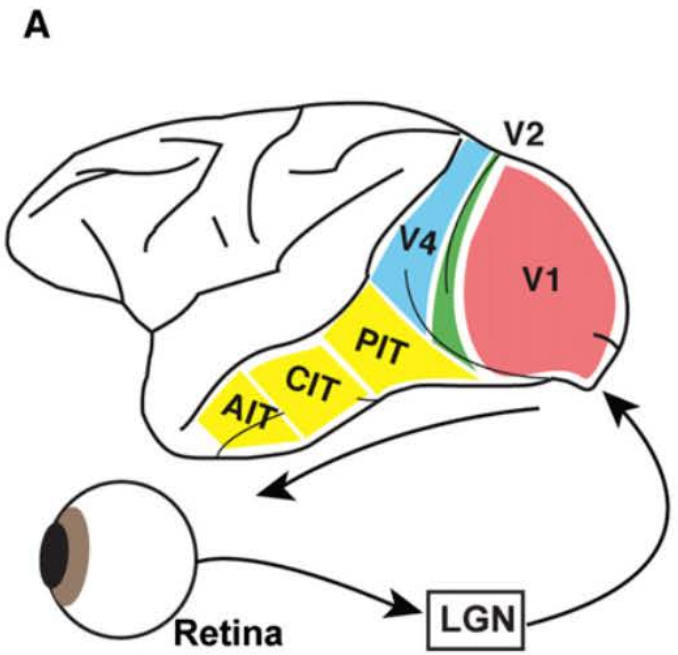
Rufin VanRullen

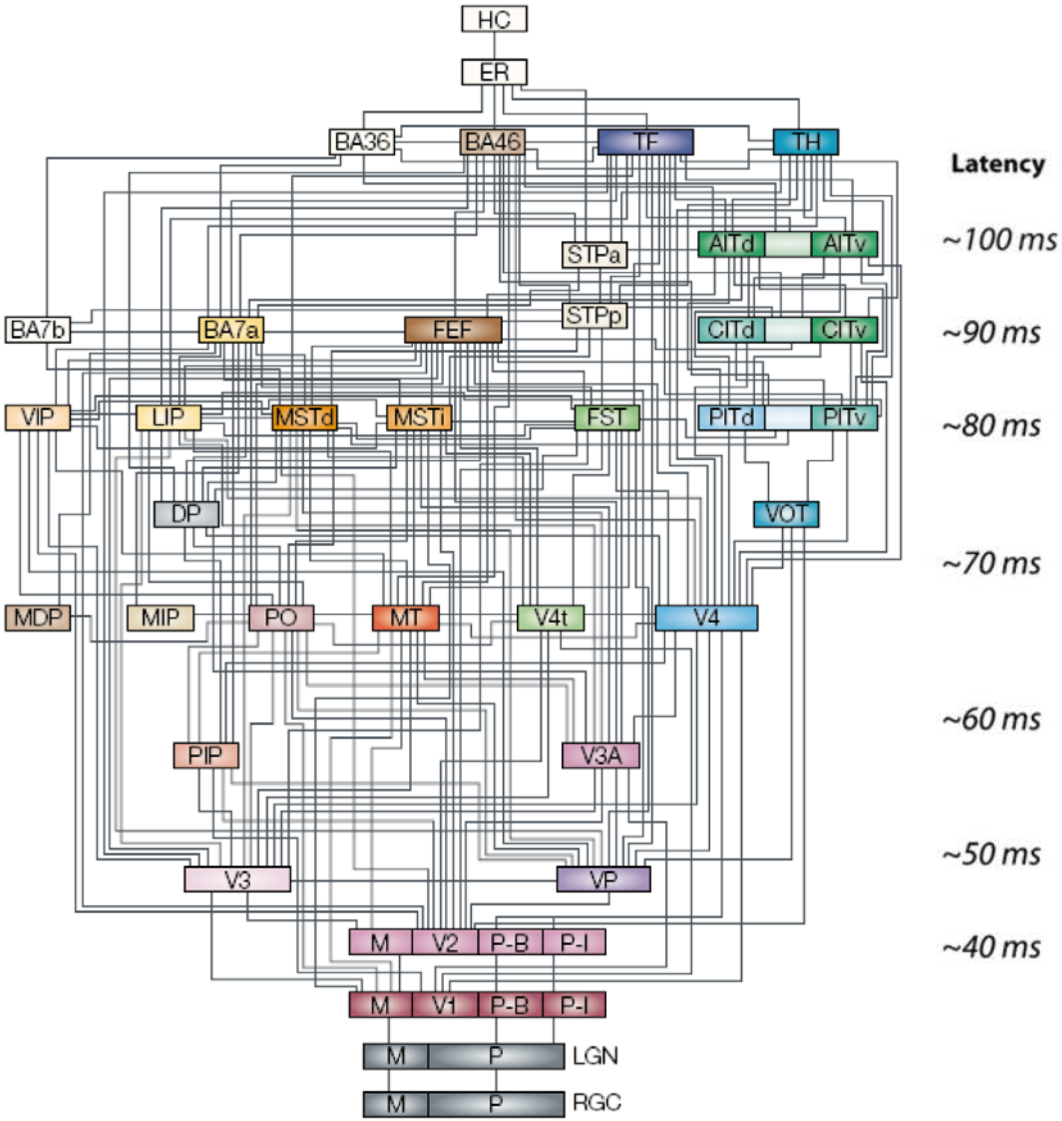
Outline

1. **What's in a brain?** Crash course in (visual) neuroscience:
   - Cortical Hierarchy
   - Receptive fields
   - Selectivities (features, object, classes)
   - Concept cells
2. **What's in a CNN?** Deepdream, visualization (explainability/interpretability) tools, examples…
3. **Brain/CNN comparisons:**
   - RSA (representational similarity analysis): fMRI, MEG, single-units
   - Brainscore
   - Case study: CLIP-multimodal
4. **Other issues about the biological plausibility of Deep Learning:**
   - Spikes
   - Adversarial attacks
   - Backprop
   - Attention/transformers
   - Recurrence…

1. What's in a brain? Crash course in (visual) neuroscience

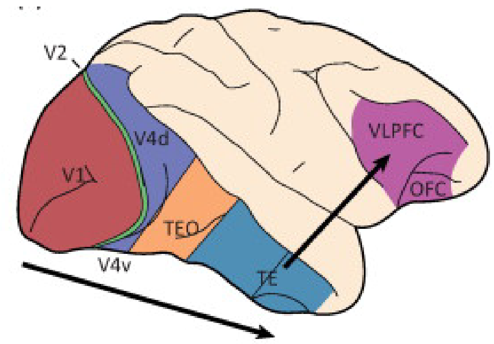
- **Cortical hierarchy**

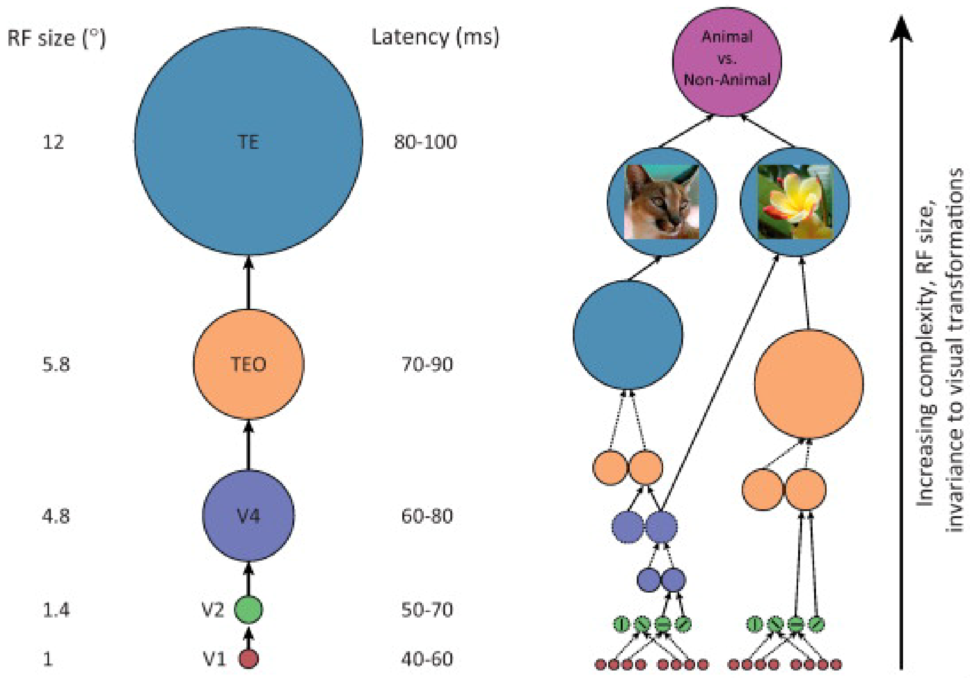




1. What's in a brain? Crash course in (visual) neuroscience

- Cortical hierarchy
- Receptive fields

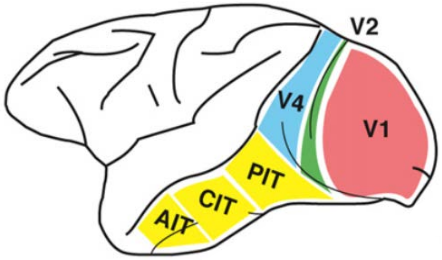


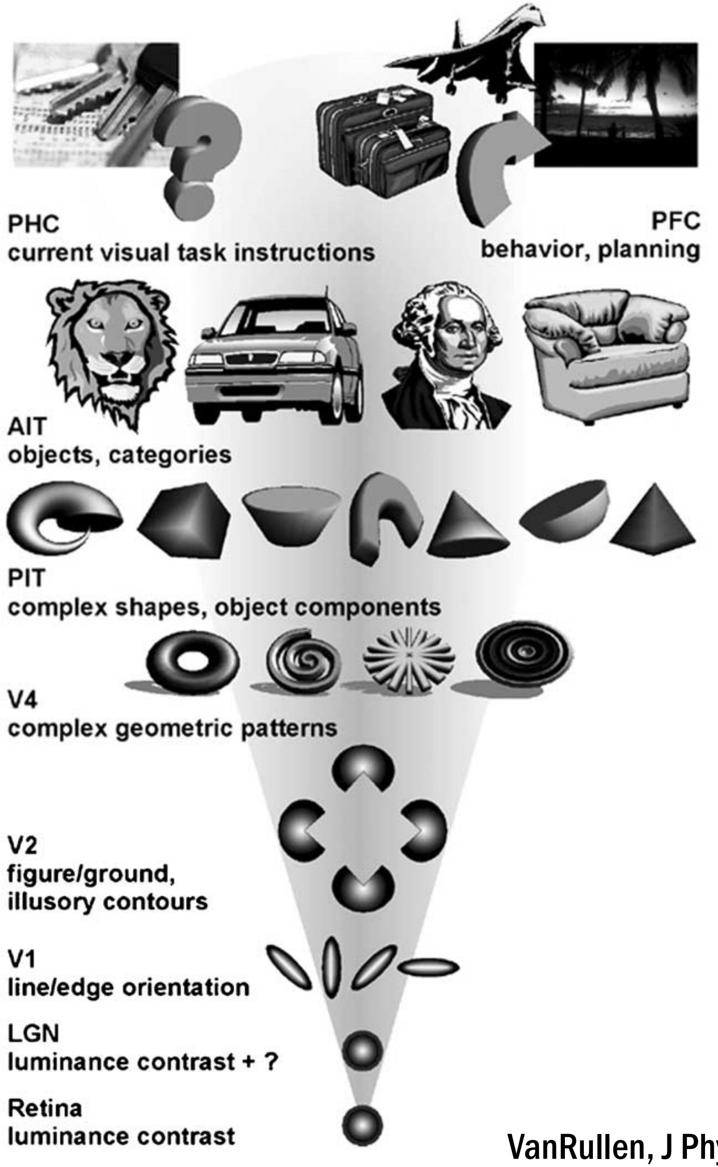


TRENDS in Cognitive Sciences

1. What's in a brain? Crash course in (visual) neuroscience

- Cortical hierarchy
- Receptive fields
- Selectivities (features, objects, classes)





VanRullen, J Phys Paris (2003)

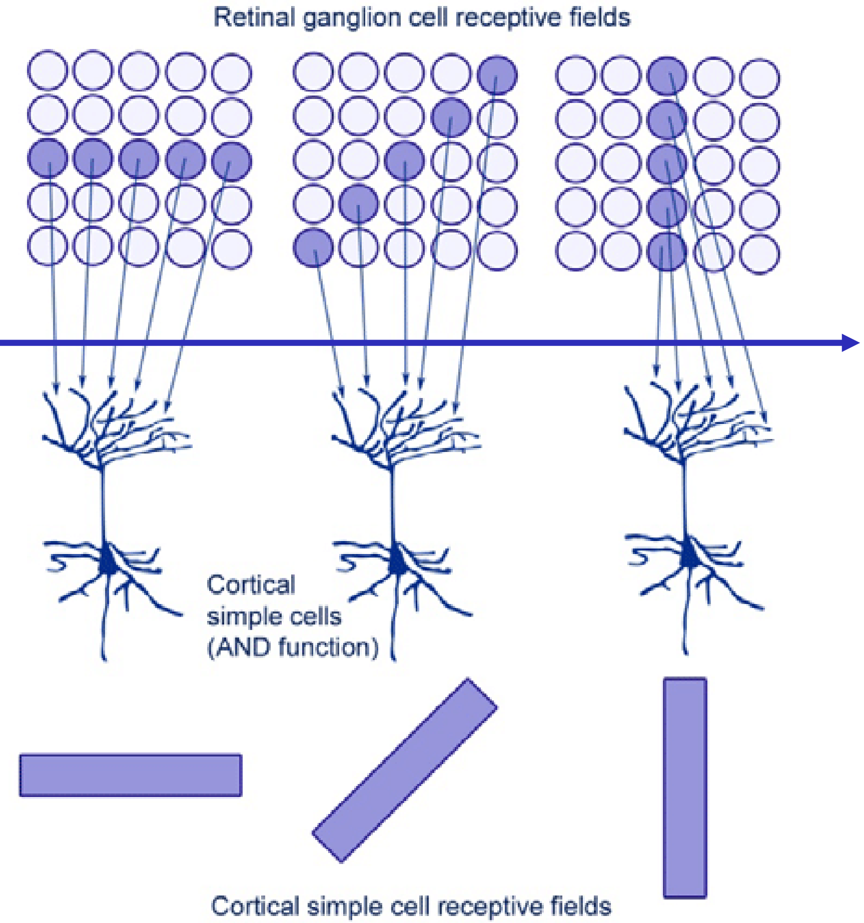
1. What’s in a brain? Crash course in (visual) neuroscience

- Cortical hierarchy
- Receptive fields
- Selectivities (features, objects, classes)

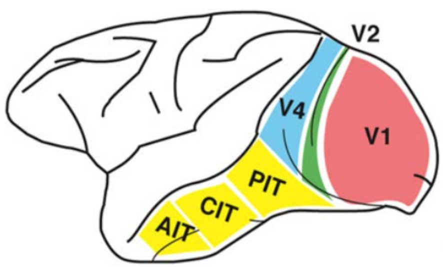
How is feature selectivity constructed?
Example for an orientation detector (V1)

5x5 kernel

3 channels

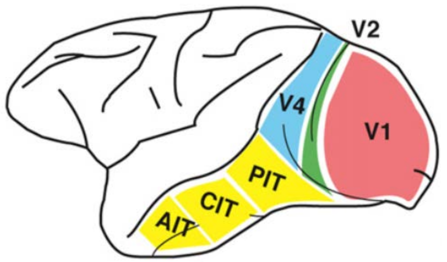


Repeating this pattern across the visual field:
~equivalent to **<u>convolution</u>** operation

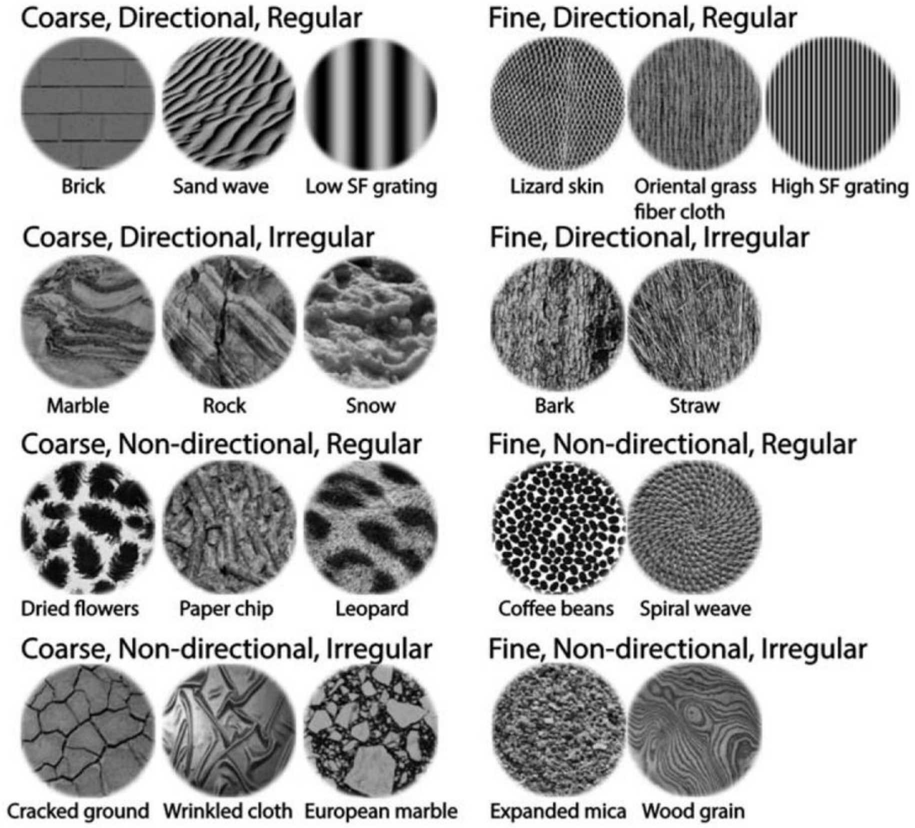


1. What's in a brain? Crash course in (visual) neuroscience

- Cortical hierarchy
- Receptive fields
- Selectivities (features, objects, classes)



More elaborate selectivities: contours, textures, shapes (V2, V4)

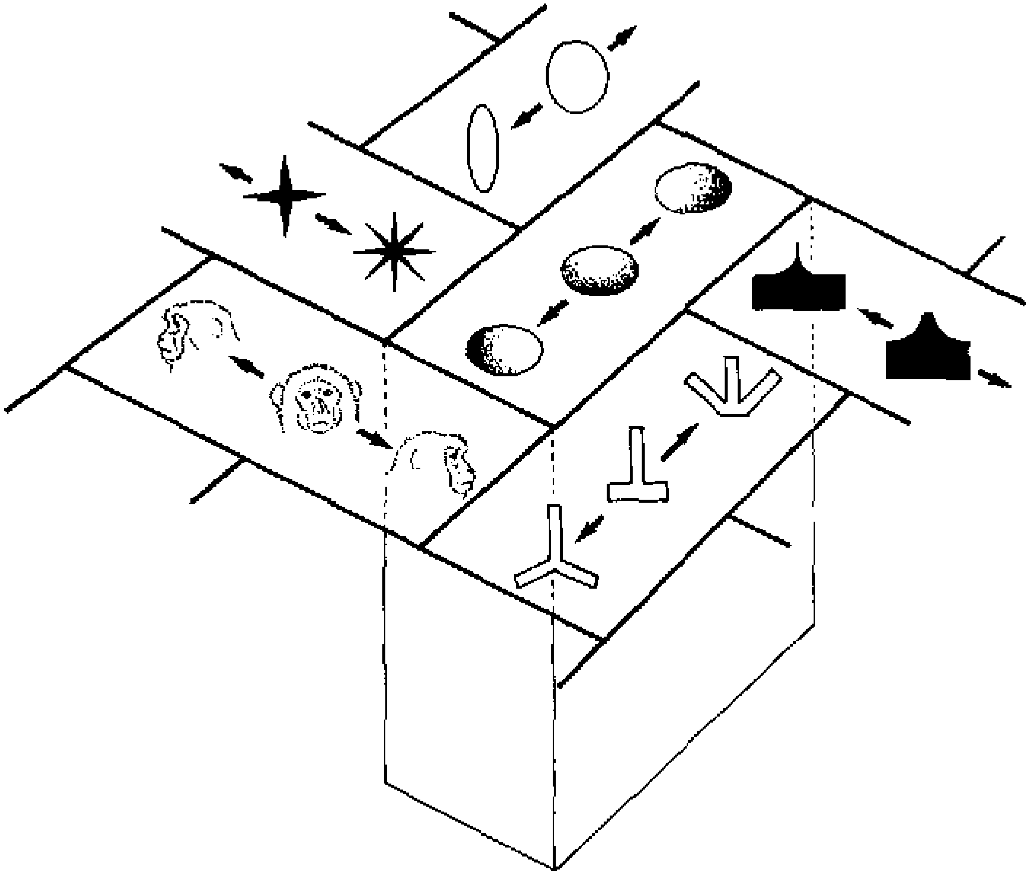


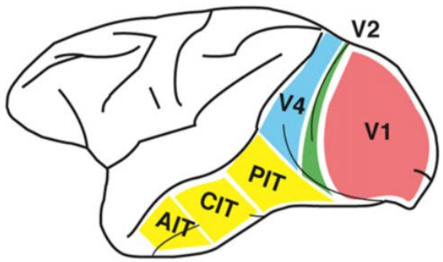
Kim, Bair & Pasupathy, J Neurosci (2019)

1. What's in a brain? Crash course in (visual) neuroscience

- Cortical hierarchy
- Receptive fields
- Selectivities (features, objects, classes)

Even more elaborate selectivities: object parts, shapes, classes (IT)

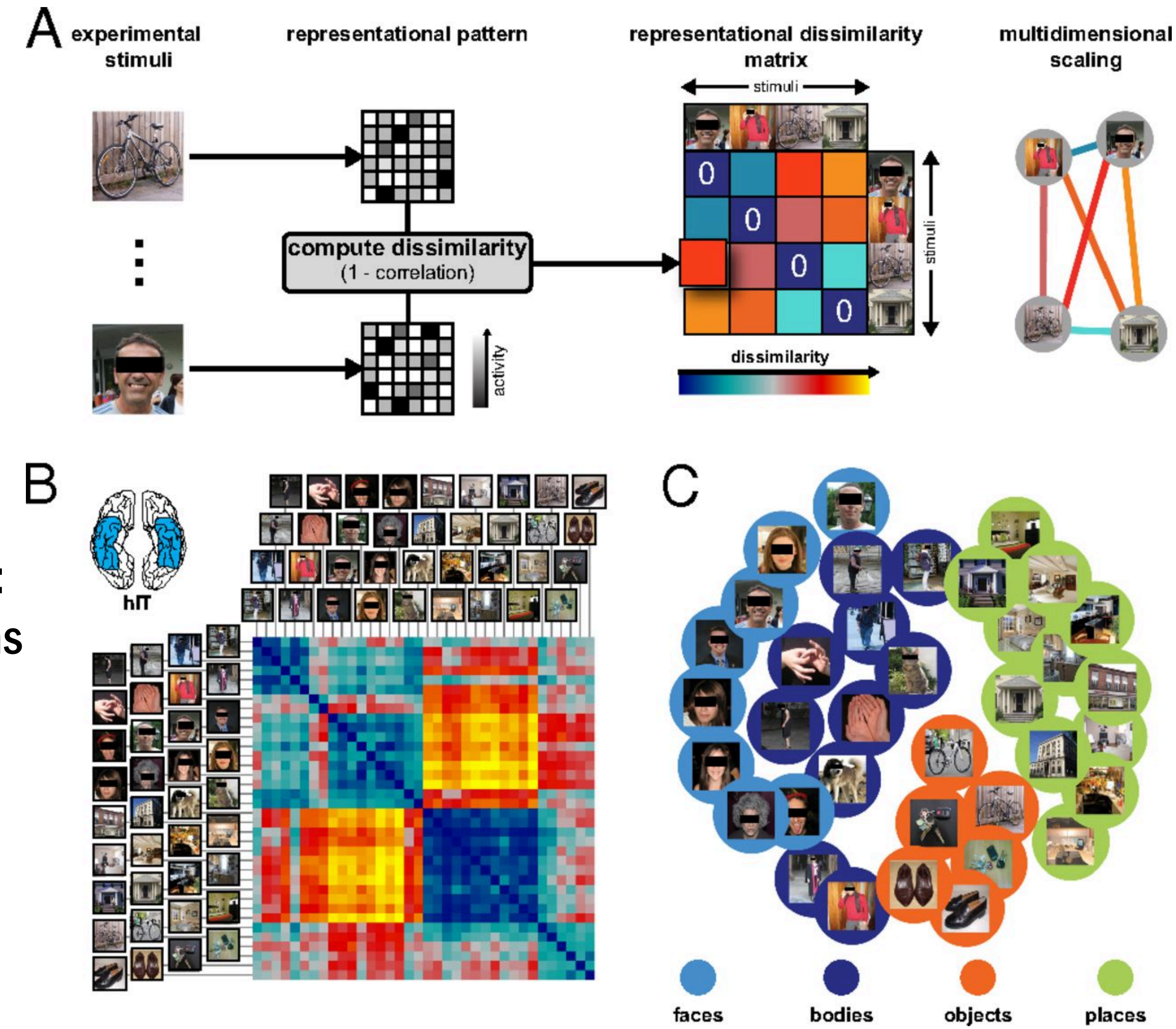




Tanaka, Annual Rev. Neurosci (1996)

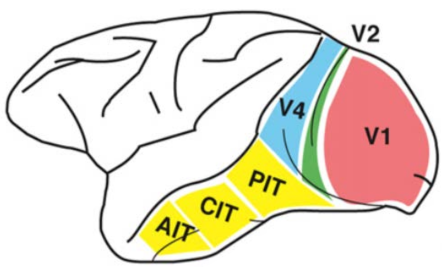
1. What's in a brain? Crash course in (visual) neuroscience

- Cortical hierarchy
- Receptive fields
- Selectivities (features, objects, classes)



The big picture

Beyond single-unit preferences: population-level representations (IT)

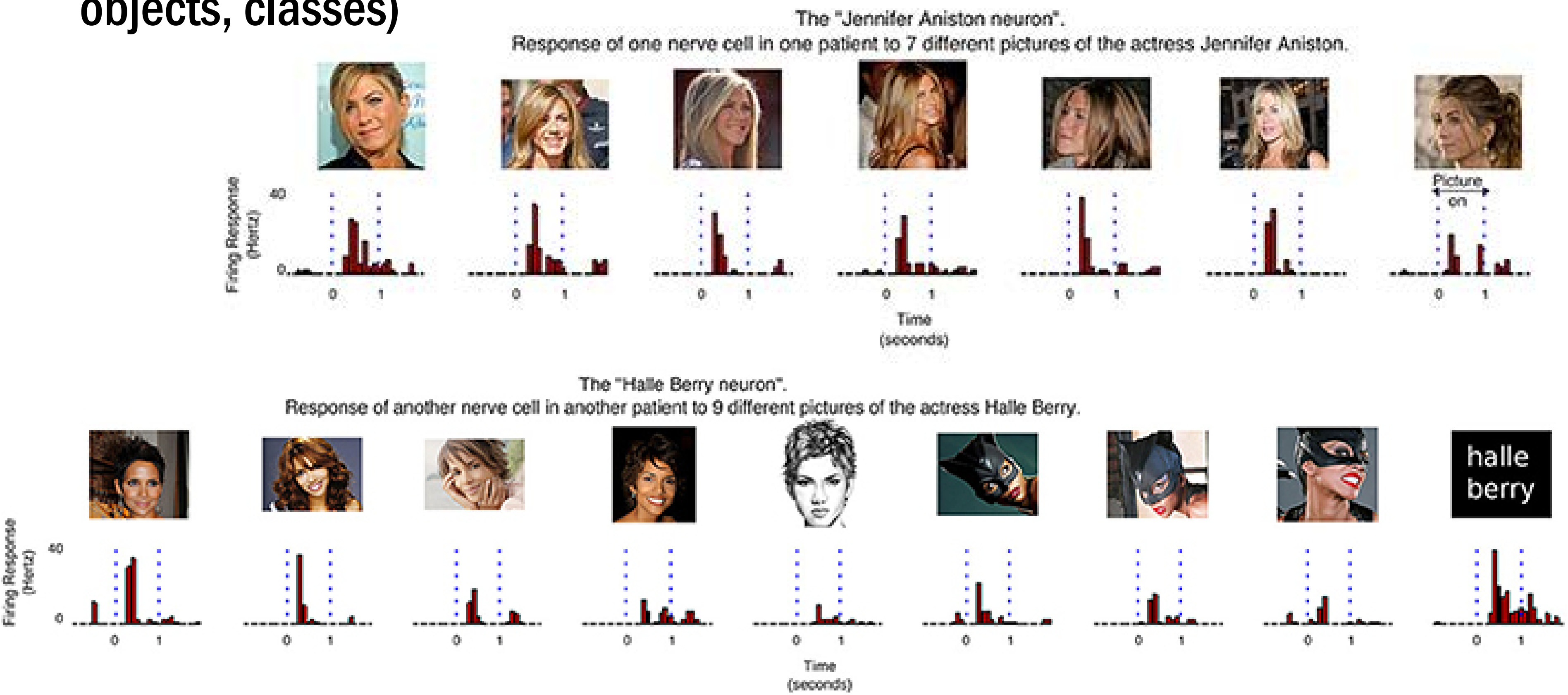


Charest et al, PNAS (2014)

1. What's in a brain? Crash course in (visual) neuroscience

- Cortical hierarchy
- Receptive fields
- Selectivities (features, objects, classes)

Still more elaborate selectivities: concept cells (Hippocampus)
➔ **Are these « grandmother » neurons?**



Quiroga, Reddy et al, Nature (2005)

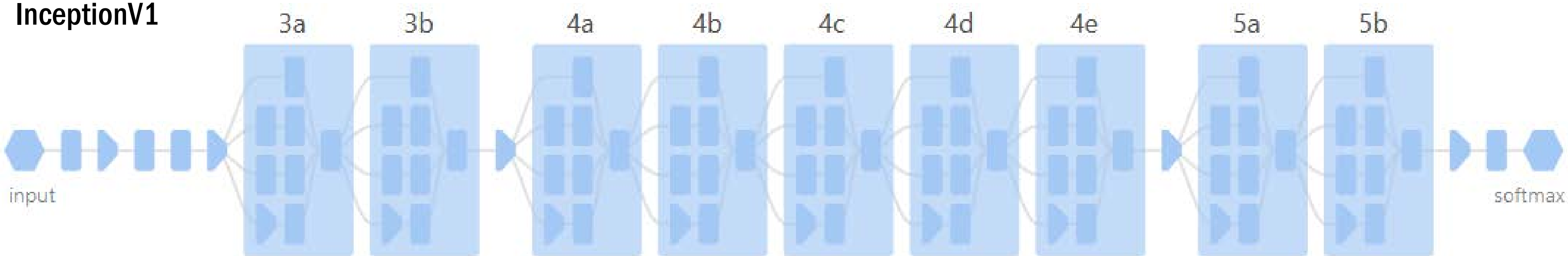
1. What's in a brain? Crash course in (visual) neuroscience

- Cortical hierarchy
- Receptive fields
- Selectivities (features, objects, classes)
- Concept cells

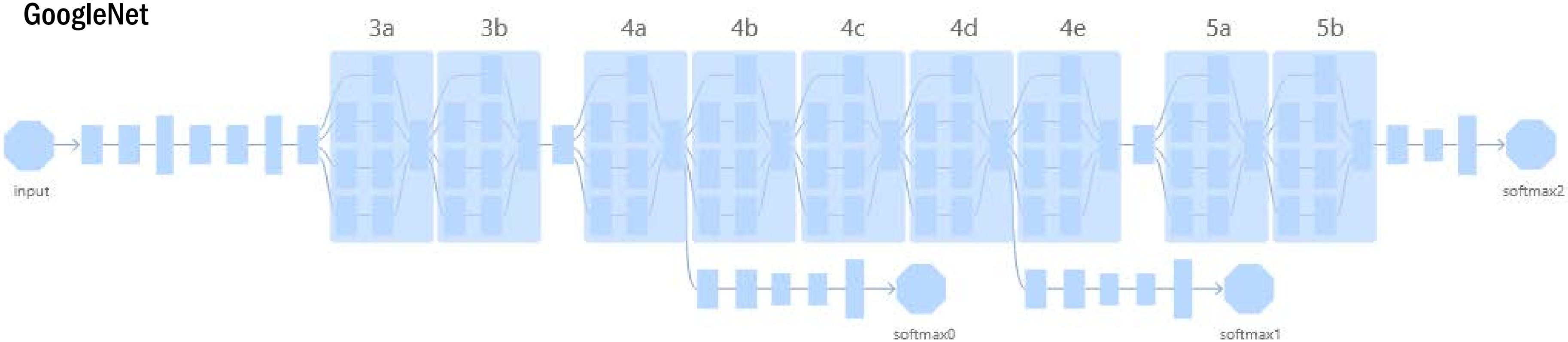
2. What's in a CNN?

- **Hierarchical structure**

InceptionV1

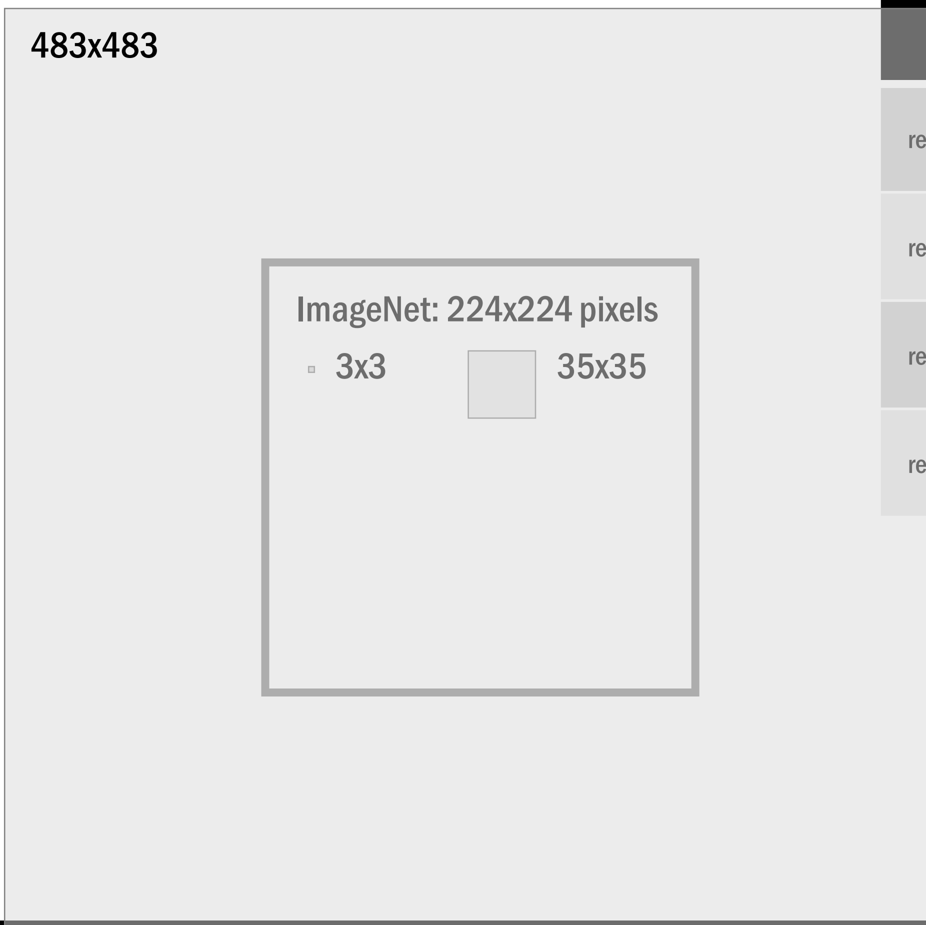


GoogleNet



2. What's in a CNN?

- **Convolutions + Receptive Fields**



ResNet50

layer	RF size
resnet_v1_50/block1	35
resnet_v1_50/block2	99
resnet_v1_50/block3	291
resnet_v1_50/block4	483

InceptionV3

layer	RF size
Conv2d_1a_3x3	3
Conv2d_2a_3x3	7
Conv2d_2b_3x3	11
MaxPool_3a_3x3	15
Conv2d_3b_1x1	15
Conv2d_4a_3x3	23
MaxPool_5a_3x3	31
Mixed_5b	63
Mixed_5c	95
Mixed_5d	127
Mixed_6a	159
Mixed_6b	351
Mixed_6c	543
Mixed_6d	735
Mixed_6e	927
Mixed_7a	1055
Mixed_7b	1183
Mixed_7c	1311

2. What’s in a CNN?

- **CNNs are (roughly) biologically plausible:**
  - Hierarchical structure
  - Convolutions
  - Receptive fields
  - Feature/object selectivity?

2. What's in a CNN?

- **How to peek within the black box?**
  - Deepdream

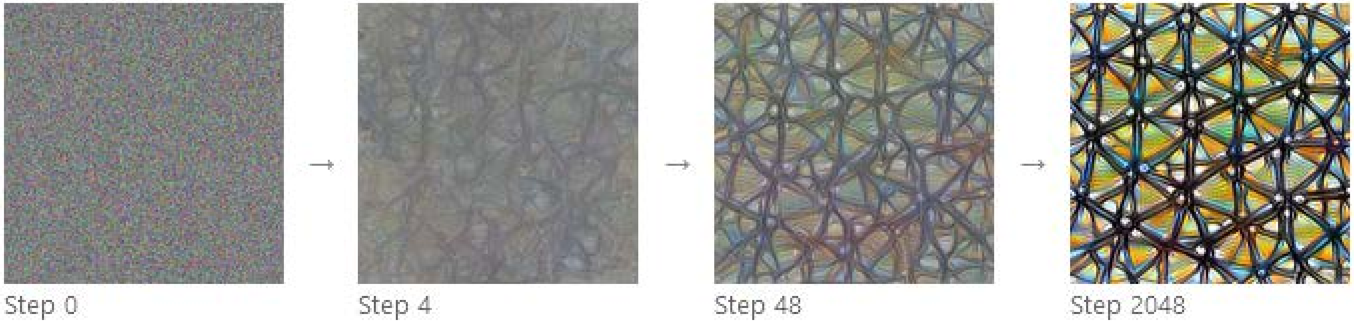
2. What’s in a CNN?

- **How to peek within the black box?**
  - Deepdream – across layers of GoogleNet

2. What’s in a CNN?

- **How to peek within the black box?**
  - How does Deepdream (and feature visualization) work?
  - ➔ Gradient descent on image (starting from noise, or from a given image)
  - ➔ with a neuron/channel/layer activation as the objective function to maximize
  - ➔ possibly with priors/regularization to impose constraints on images

Starting from random noise, we optimize an image to activate a particular neuron (layer mixed4a, unit 11).



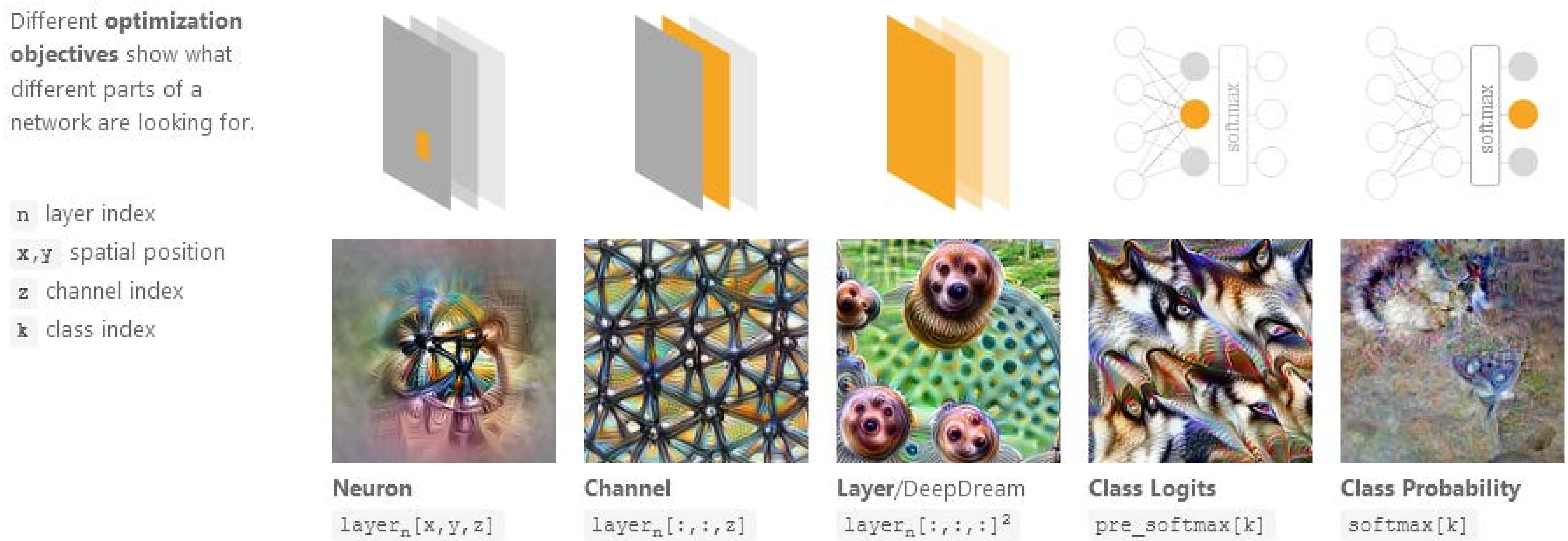
2. What's in a CNN?

- **How to peek within the black box?**
  - How does Deepdream (and feature visualization) work?

  ➔ Gradient descent on image (starting from noise, or from a given image)

  ➔ with a neuron/channel/layer activation as the objective function to maximize

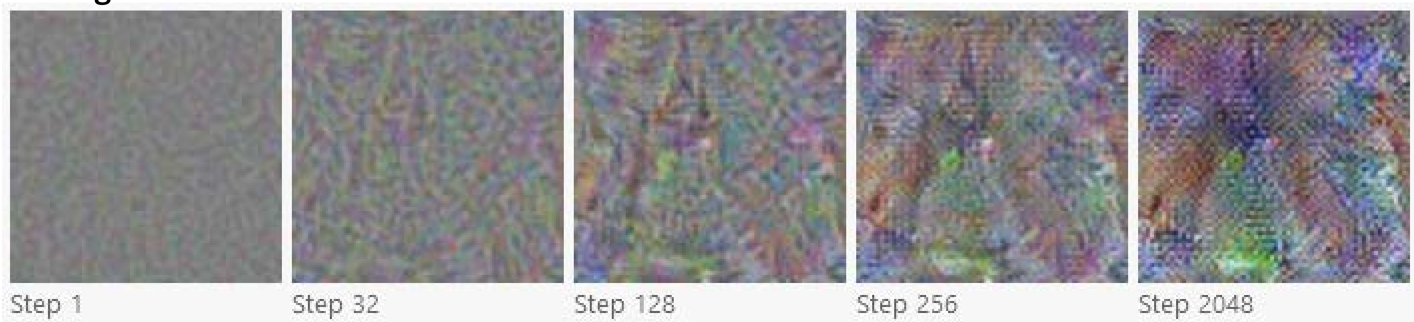
  ➔ possibly with priors/regularization to impose constraints on images



2. What's in a CNN?

- **How to peek within the black box?**
  - How does Deepdream (and feature visualization) work?
  - ➔ Gradient descent on image (starting from noise, or from a given image)
  - ➔ with a neuron/channel/layer activation as the objective function to maximize
  - ➔ possibly with priors/regularization to impose constraints on images

No regularization

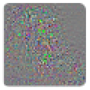
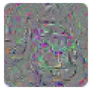
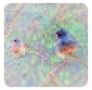


Full regularization

2. What's in a CNN?

Weak Regularization avoids misleading correlations, but is less connected to real use.

Strong Regularization gives more realistic examples at risk of misleading correlations.

		Unregularized	Frequency Penalization	Transformation Robustness	Learned Prior	Dataset Examples
	**Erhan, *et al.*, 2009** [3] Introduced core idea. Minimal regularization.	■	□	□	□	□
	**Szegedy, *et al.*, 2013** [11] Adversarial examples. Visualizes with dataset examples.	■	□	□	□	■
	**Mahendran & Vedaldi, 2015** [7] Introduces total variation regularizer. Reconstructs input from representation.	□	■	□	□	□
	**Nguyen, *et al.*, 2015** [14] Explores counterexamples. Introduces image blurring.	□	■	□	□	□
	**Mordvintsev, *et al.*, 2015** [4] Introduced jitter & multi-scale. Explored GMM priors for classes.	□	□	■	■	□
	**Øygard, *et al.*, 2015** [15] Introduces gradient blurring. (Also uses jitter.)	□	■	■	□	□
	**Tyka, *et al.*, 2016** [16] Regularizes with bilateral filters. (Also uses jitter.)	□	■	■	□	□
	**Mordvintsev, *et al.*, 2016** [17] Normalizes gradient frequencies. (Also uses jitter.)	□	■	■	□	□
	**Nguyen, *et al.*, 2016** [18] Paramaterizes images with GAN generator.	□	□	□	■	□
	**Nguyen, *et al.*, 2016** [10] Uses denoising autoencoder prior to make a generative model.	□	□	□	■	□

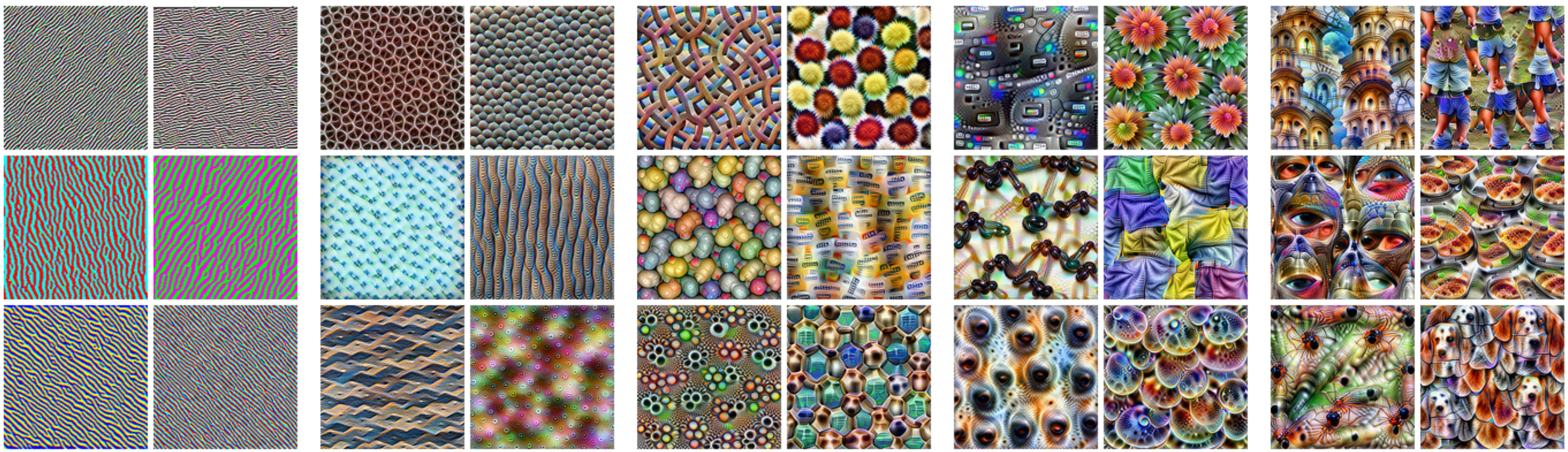
2. What's in a CNN?

- **How to peek within the black box?**

(Every image in this section can be reproduced with the notebooks available at https://github.com/tensorflow/lucid)
(I also strongly recommend exploring some pre-computed visualizations at https://microscope.openai.com/models)

Feature Visualization

How neural networks build up their understanding of images



Edges (layer conv2d0) **Textures** (layer mixed3a) **Patterns** (layer mixed4a) **Parts** (layers mixed4b & mixed4c) **Objects** (layers mixed4d & mixed4e)

Feature visualization allows us to see how GoogLeNet[1], trained on the ImageNet[2] dataset, builds up its understanding of images over many layers. Visualizations of all channels are available in the appendix.

Olah, et al., "Feature Visualization", Distill, 2017.

2. What's in a CNN?

- **Feature visualization vs. Dataset Examples**

Dataset Examples show us what neurons respond to in practice

Optimization isolates the causes of behavior from mere correlations. A neuron may not be detecting what you initially thought.

Baseball—or stripes? *mixed4a, Unit 6*

Animal faces—or snouts? *mixed4a, Unit 240*

Clouds—or fluffiness? *mixed4a, Unit 453*

Buildings—or sky? *mixed4a, Unit 492*

Olah, et al., "Feature Visualization", Distill, 2017.

2. What's in a CNN?

- **Diversity in feature visualization**

Dataset examples have a big advantage here. By looking through our dataset, we can find diverse examples. It doesn't just give us ones activating a neuron intensely: we can look across a whole spectrum of activations to see what activates the neuron to different extents.

In contrast, optimization generally gives us just one extremely positive example — and if we're creative, a very negative example as well. Is there some way that optimization could also give us this diversity?

Negative optimized | **Minimum** activation examples | Slightly negative activation examples | Slightly positive activation examples | **Maximum** activation examples | **Positive** optimized

Olah, et al., "Feature Visualization", Distill, 2017.

2. What's in a CNN?

- **Diversity in feature visualization**
  - ➔ **Just add a « diversity term » to the loss**

Simple Optimization

Optimization with diversity reveals multiple types of balls. *Layer mixed5a, Unit 9*

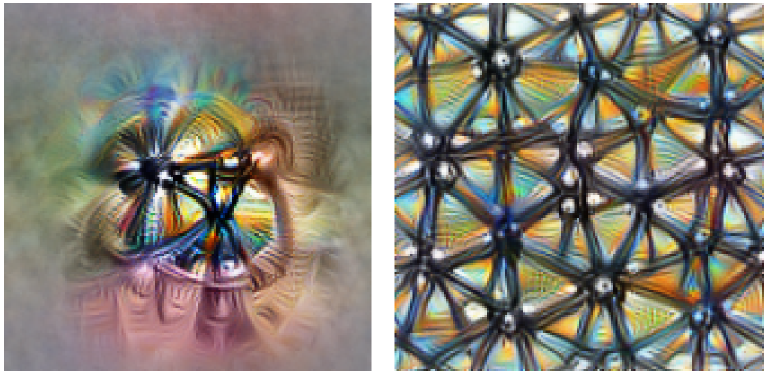
Dataset examples

Olah, et al., "Feature Visualization", Distill, 2017.

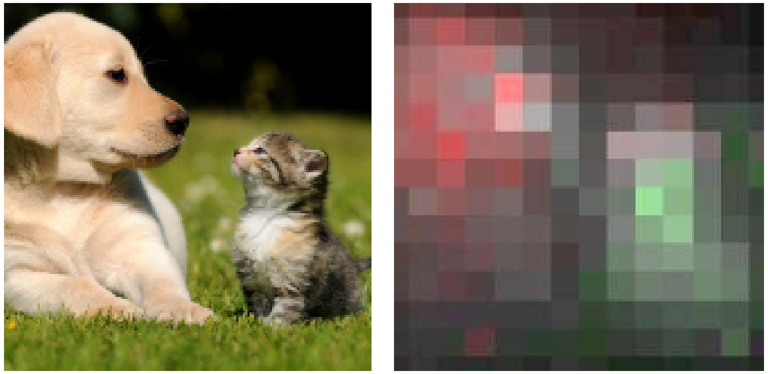
2. What's in a CNN?

- **Feature visualization vs. attribution**

There is a growing sense that neural networks need to be interpretable to humans. The field of neural network interpretability has formed in response to these concerns. As it matures, two major threads of research have begun to coalesce: feature visualization and attribution.



Feature visualization answers questions about what a network — or parts of a network — are looking for by generating examples.



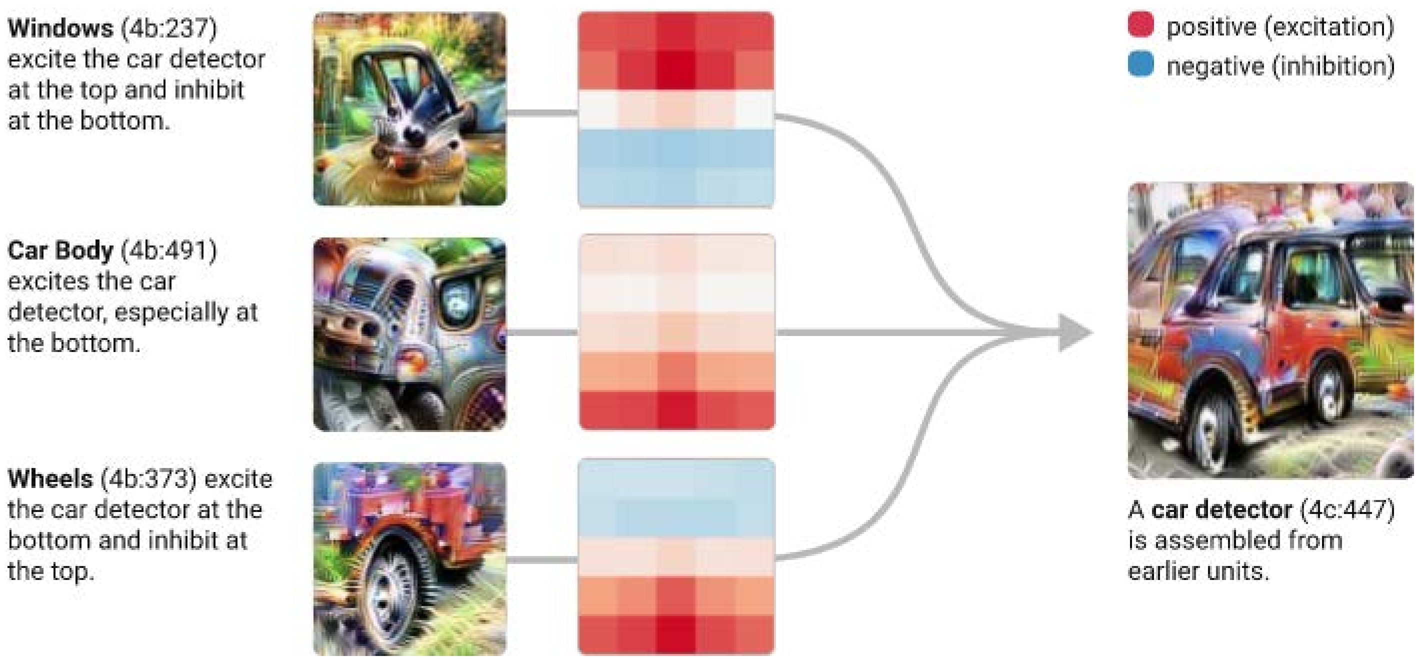
Attribution [1] studies what part of an example is responsible for the network activating a particular way.

Olah, et al., "Feature Visualization", Distill, 2017.

2. What's in a CNN?

- **Visualizing the learned weights (not just activations)**

➔ **This can tell us about the neural « circuits »**

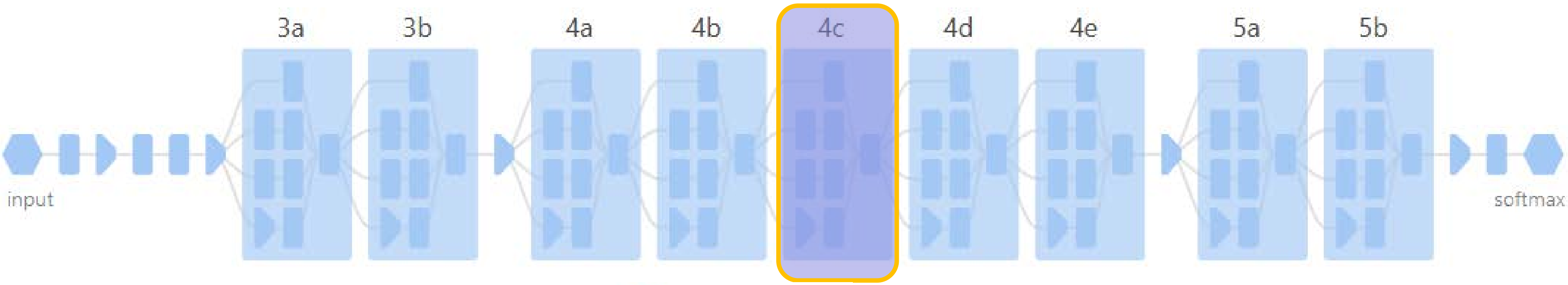


In `mixed4c`, a mid-late layer of InceptionV1, there is a car detecting neuron. Using features from the previous layers, it looks for wheels at the bottom of its convolutional window, and windows at the top.

Olah, et al., "Zoom In: An Introduction to Circuits", Distill, 2020

2. What's in a CNN?

- **The big picture: joint encoding and representation at the level of entire regions (activation atlas)**

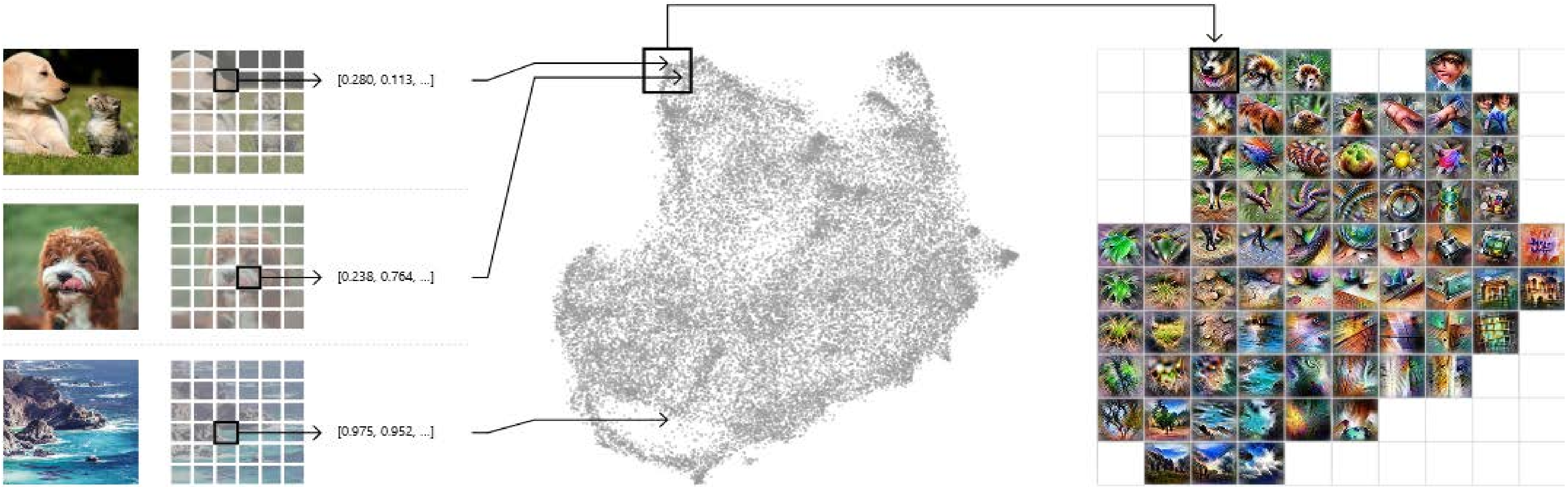


InceptionV1

Carter, et al., "Activation Atlas", Distill, 2019

2. What's in a CNN?

- **The big picture: joint encoding and representation at the level of entire regions (activation atlas)**



A randomized set of one million images is fed through the network, collecting one random spatial activation per image.

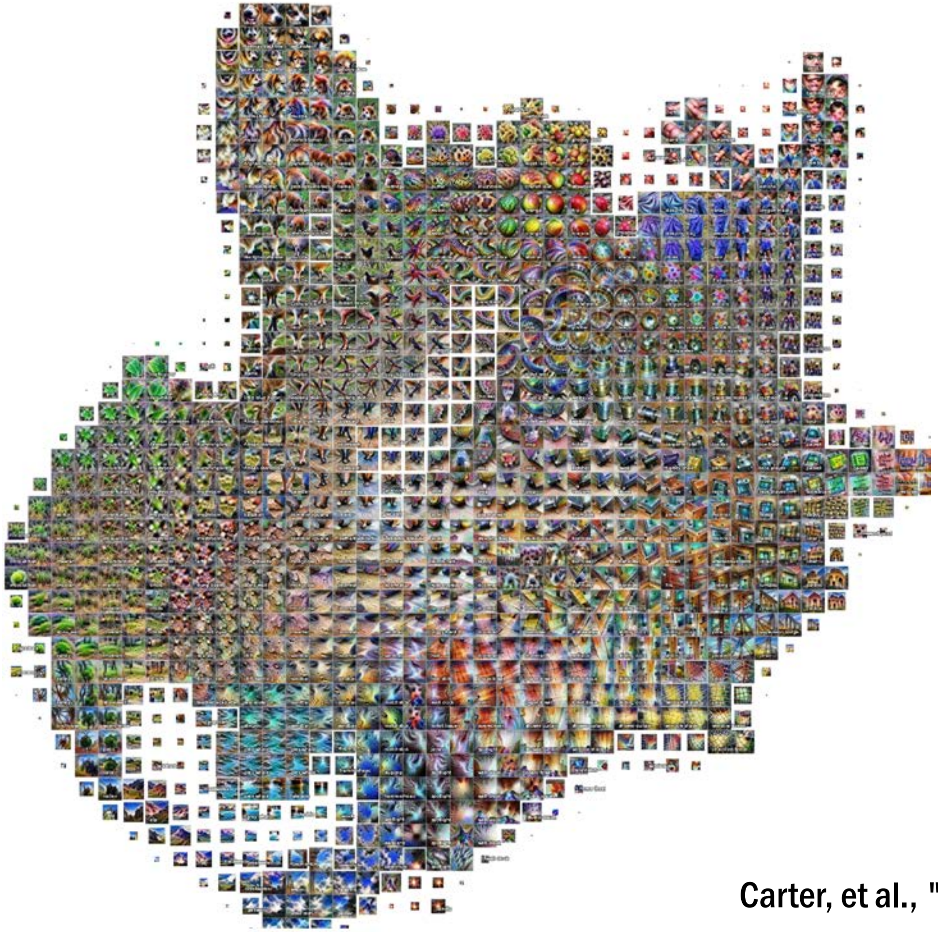
The activations are fed through UMAP to reduce them to two dimensions. They are then plotted, with similar activations placed near each other.

We then draw a grid and average the activations that fall within a cell and run feature inversion on the averaged activation. We also optionally size the grid cells according to the density of the number of activations that are averaged within.

Carter, et al., "Activation Atlas", Distill, 2019

2. What's in a CNN?

- **The big picture: joint encoding and representation at the level of entire regions (activation atlas)**



Carter, et al., "Activation Atlas", Distill, 2019

2. What’s in a CNN?

- **The big picture: joint encoding and representation at the level of entire regions (activation atlas)**

Zoom on: animal heads (eyes, fur, nose...)

Carter, et al., "Activation Atlas", Distill, 2019

2. What's in a CNN?

- **The big picture: joint encoding and representation at the level of entire regions (activation atlas)**

Zoom on: animal backs (fur, 4-legs...)

Carter, et al., "Activation Atlas", Distill, 2019

2. What’s in a CNN?

- **The big picture: joint encoding and representation at the level of entire regions (activation atlas)**

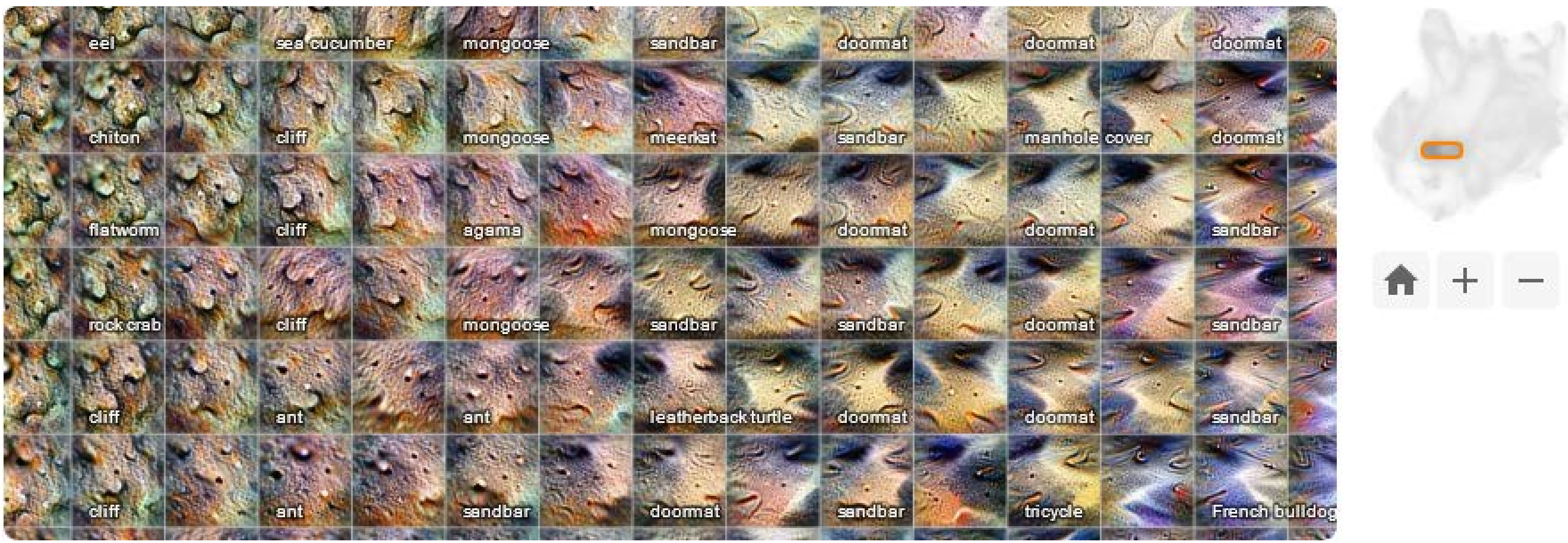
Zoom on: animal legs (feet, ground…)

Carter, et al., "Activation Atlas", Distill, 2019

2. What's in a CNN?

- **The big picture: joint encoding and representation at the level of entire regions (activation atlas)**

Zoom on: types of ground (sand, dune...)

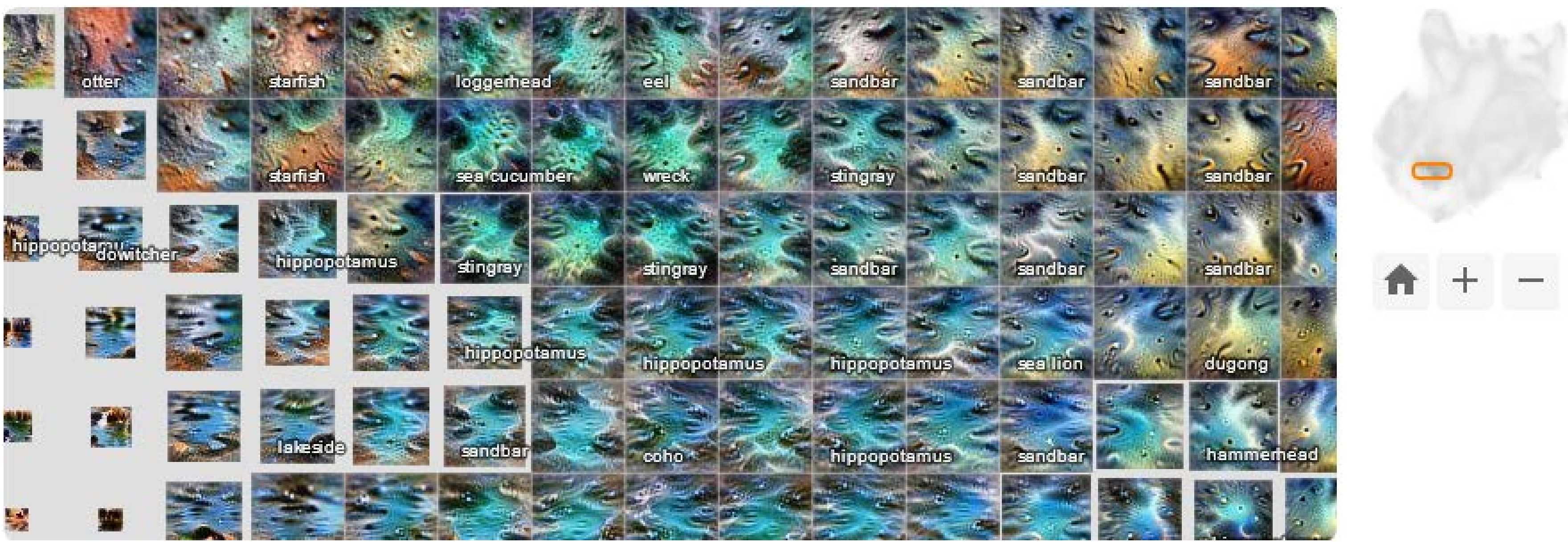


Carter, et al., "Activation Atlas", Distill, 2019

2. What's in a CNN?

- **The big picture: joint encoding and representation at the level of entire regions (activation atlas)**

Zoom on: sea (beach, water…)



Carter, et al., "Activation Atlas", Distill, 2019

2. What's in a CNN?

- **The big picture: joint encoding and representation at the level of entire regions (activation atlas)**

Zoom on: text (packages, websites...)

Carter, et al., "Activation Atlas", Distill, 2019

2. What's in a CNN?

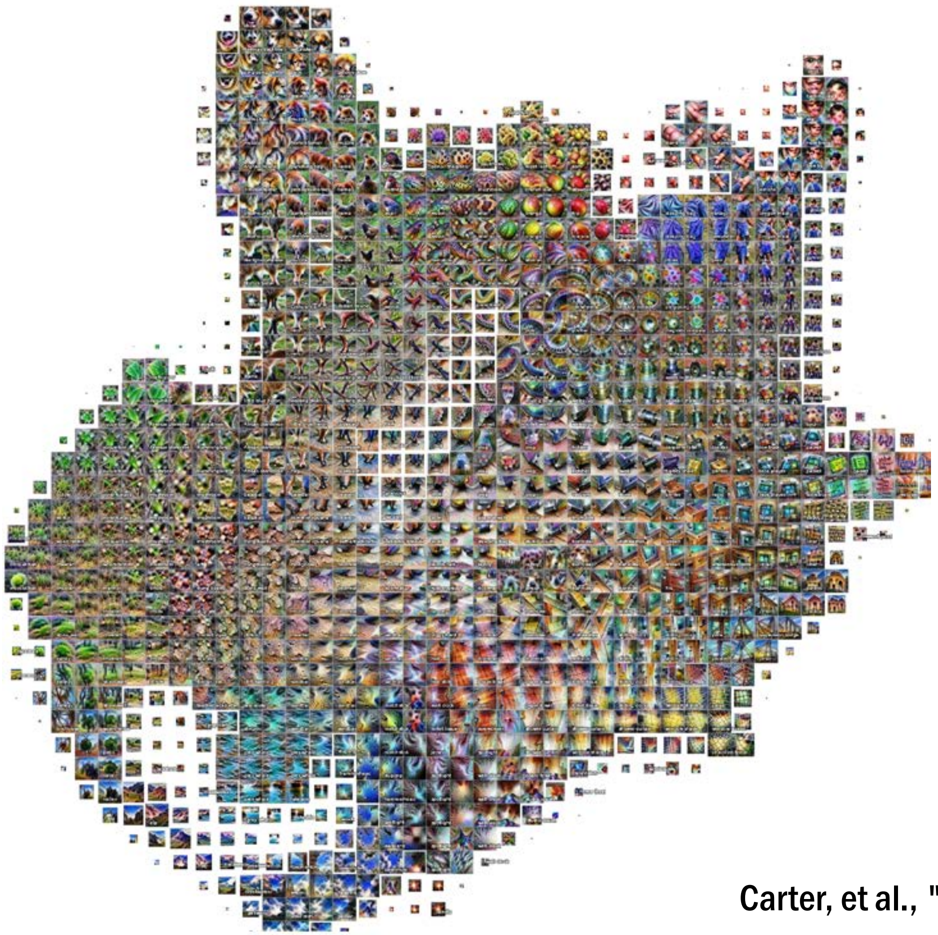
- **The big picture: joint encoding and representation at the level of entire regions (activation atlas)**

Zoom on: fruits (mangos, strawberries…)

Carter, et al., "Activation Atlas", Distill, 2019

2. What's in a CNN?

- **The big picture: joint encoding and representation at the level of entire regions (activation atlas)**



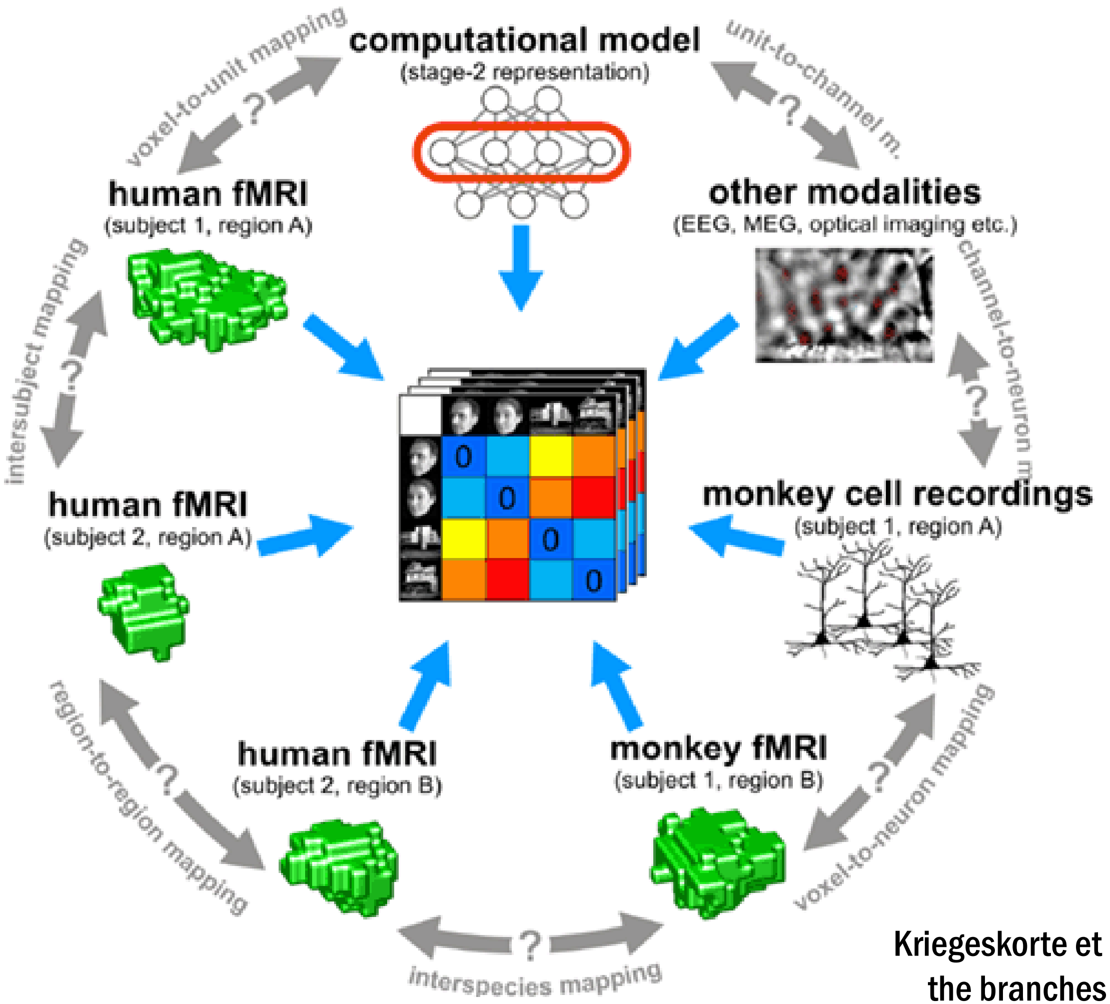
Carter, et al., "Activation Atlas", Distill, 2019

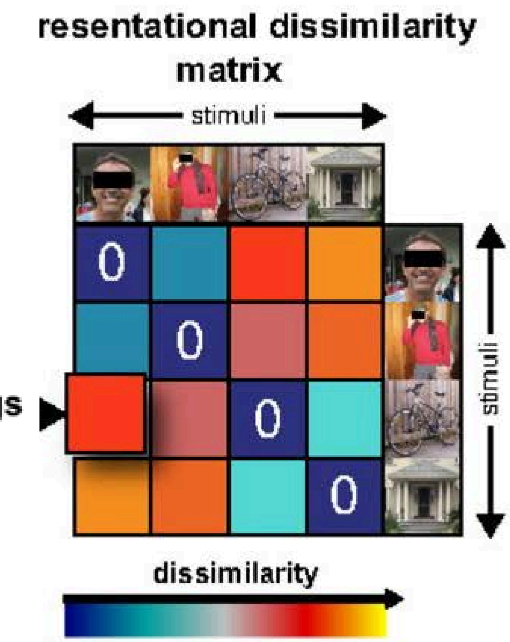
3. Brain/CNN comparisons

- **RSA (representational similarity analysis):**
  - fMRI
  - MEG
  - Single-units (Brainscore)
- **Case study: CLIP multimodal neurons**

3. Brain/CNN comparisons

- **RSA (representational similarity analysis):**

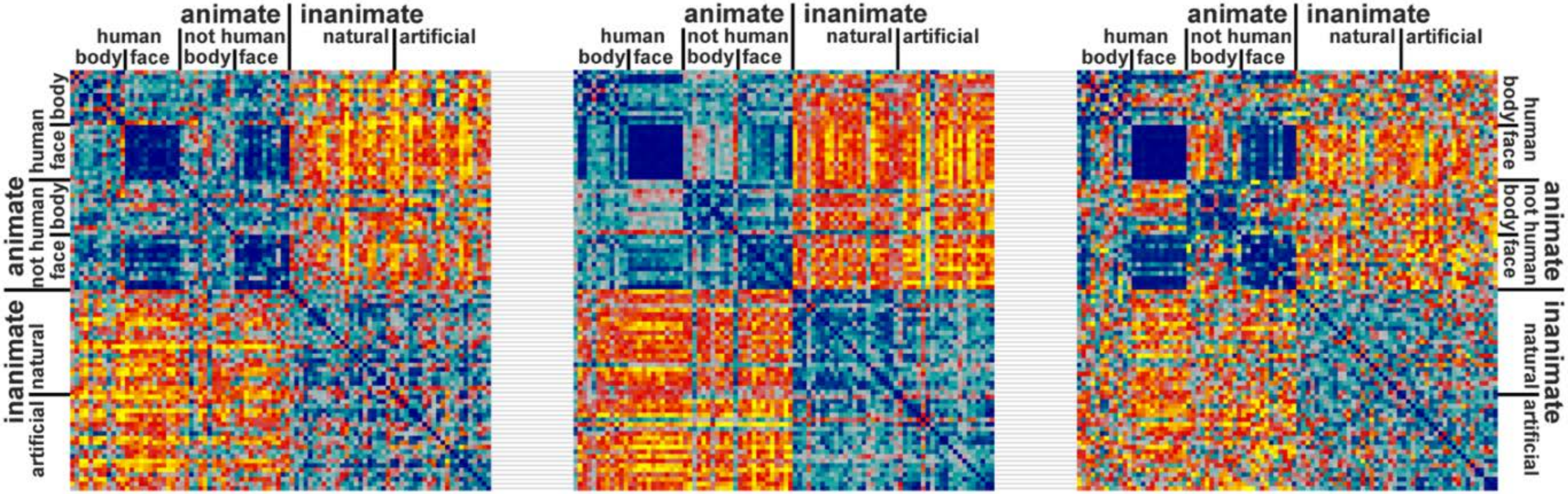




Kriegeskorte et al, "Representational similarity analysis – connecting the branches of systems neuroscience", Front Sys Neurosci (2008)

3. Brain/CNN comparisons

- **RSA (representational similarity analysis):**

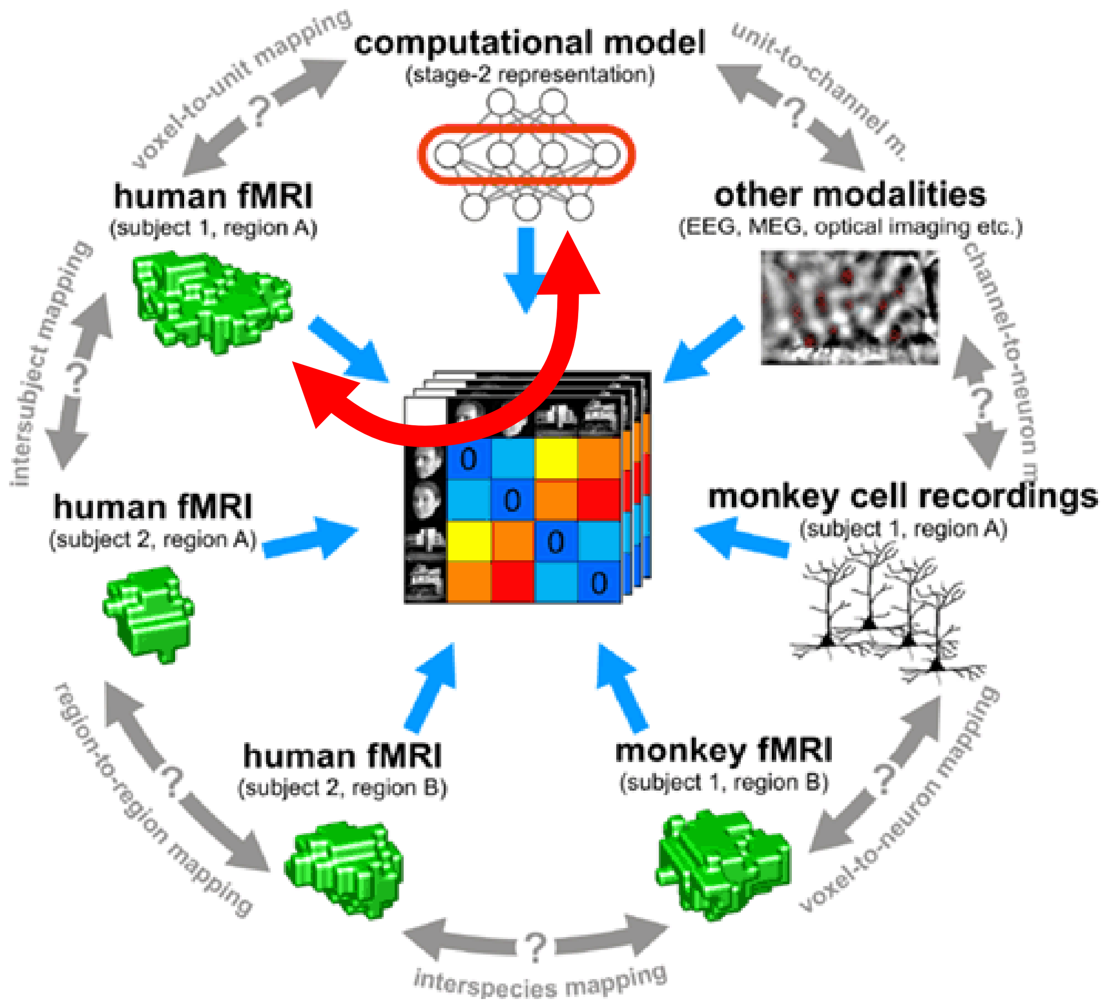


In these 3 RDMs, there is a monkey, a human, and a DNN. Can you tell which is which?

Khaligh-Razavi & Kriegeskorte, PLoS Comp Biol (2014)

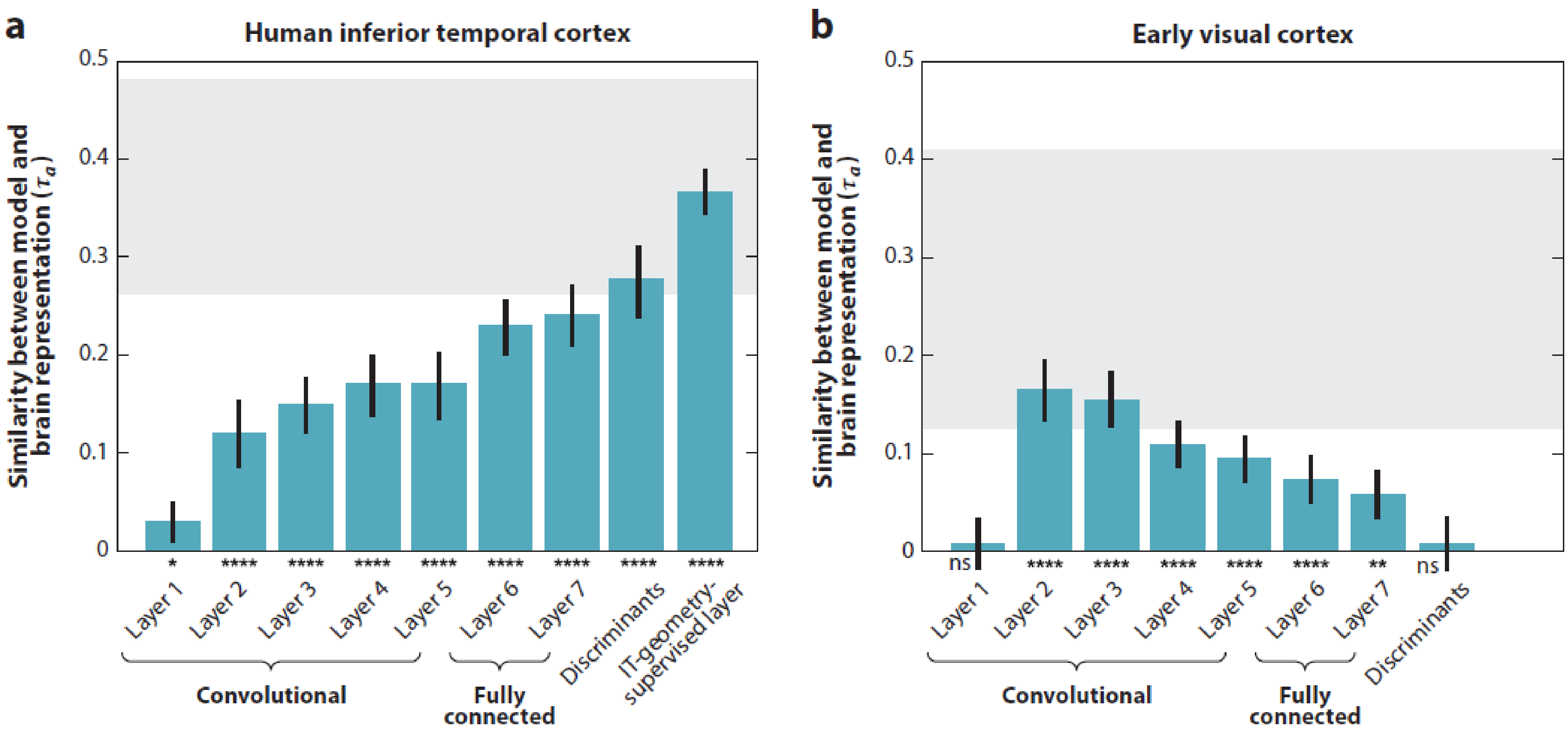
3. Brain/CNN comparisons

- **RSA (representational similarity analysis):**



3. Brain/CNN comparisons

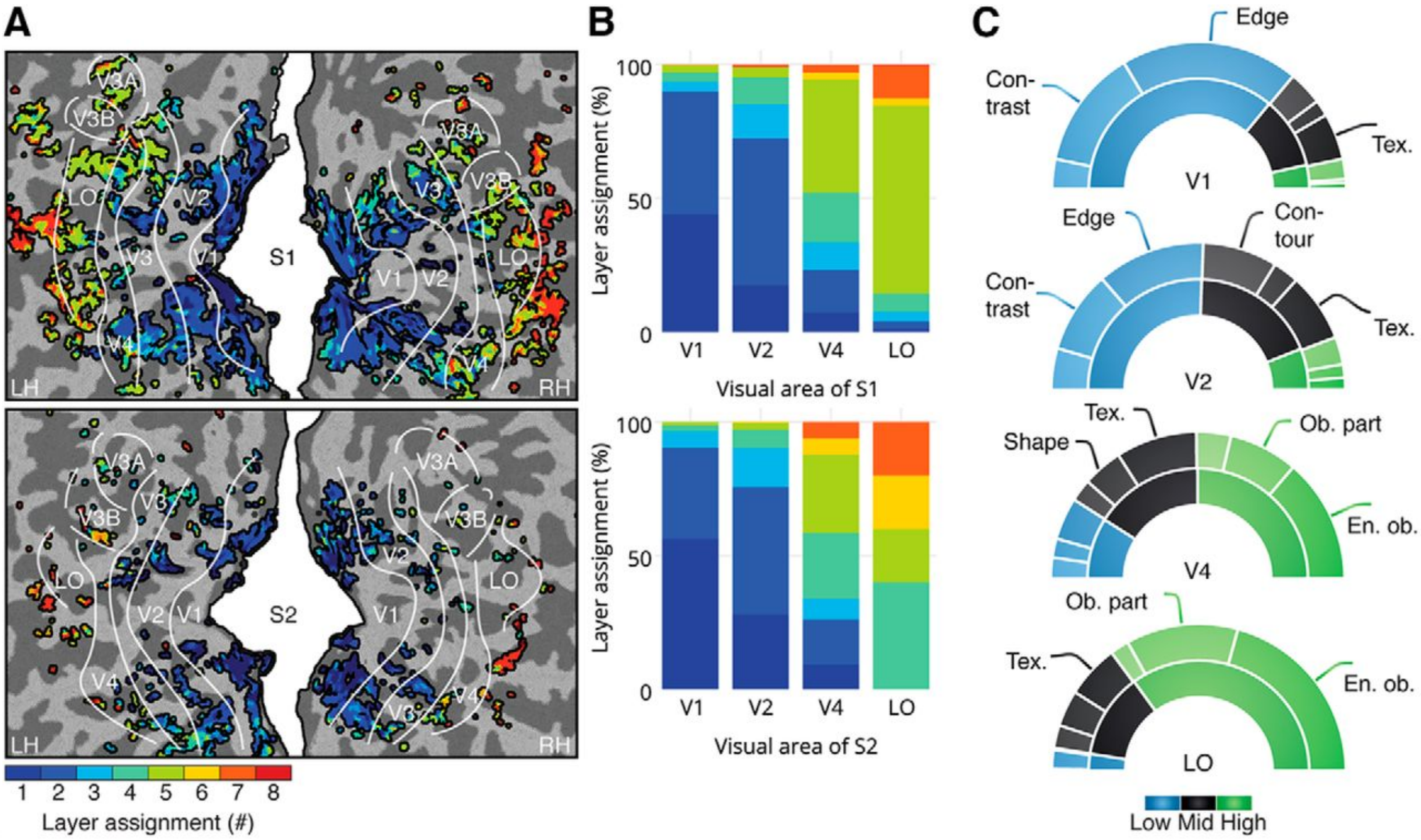
- **RSA (representational similarity analysis):**
  - fMRI

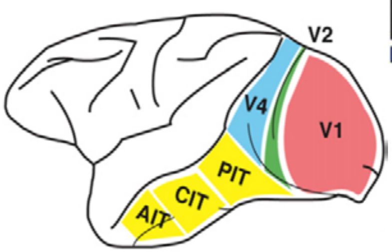


Khaligh-Razavi & Kriegeskorte, PLoS Comp Biol (2014)

3. Brain/CNN comparisons

- **RSA (representational similarity analysis):**
  - fMRI

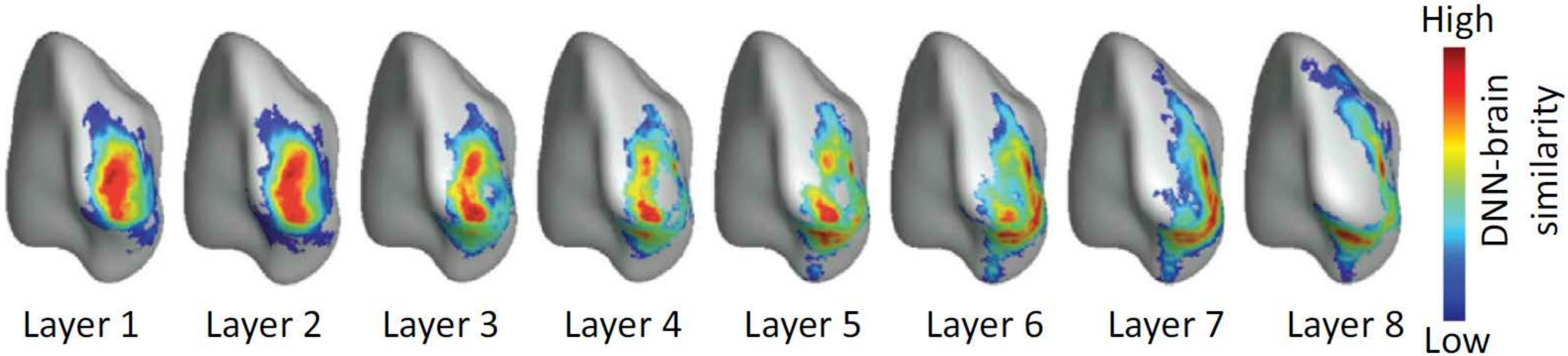


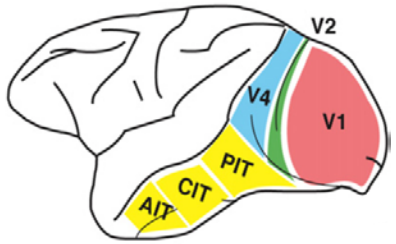


Guclu & van Gerven, J Neurosci (2015)

3. Brain/CNN comparisons

- **RSA (representational similarity analysis):**
  - fMRI

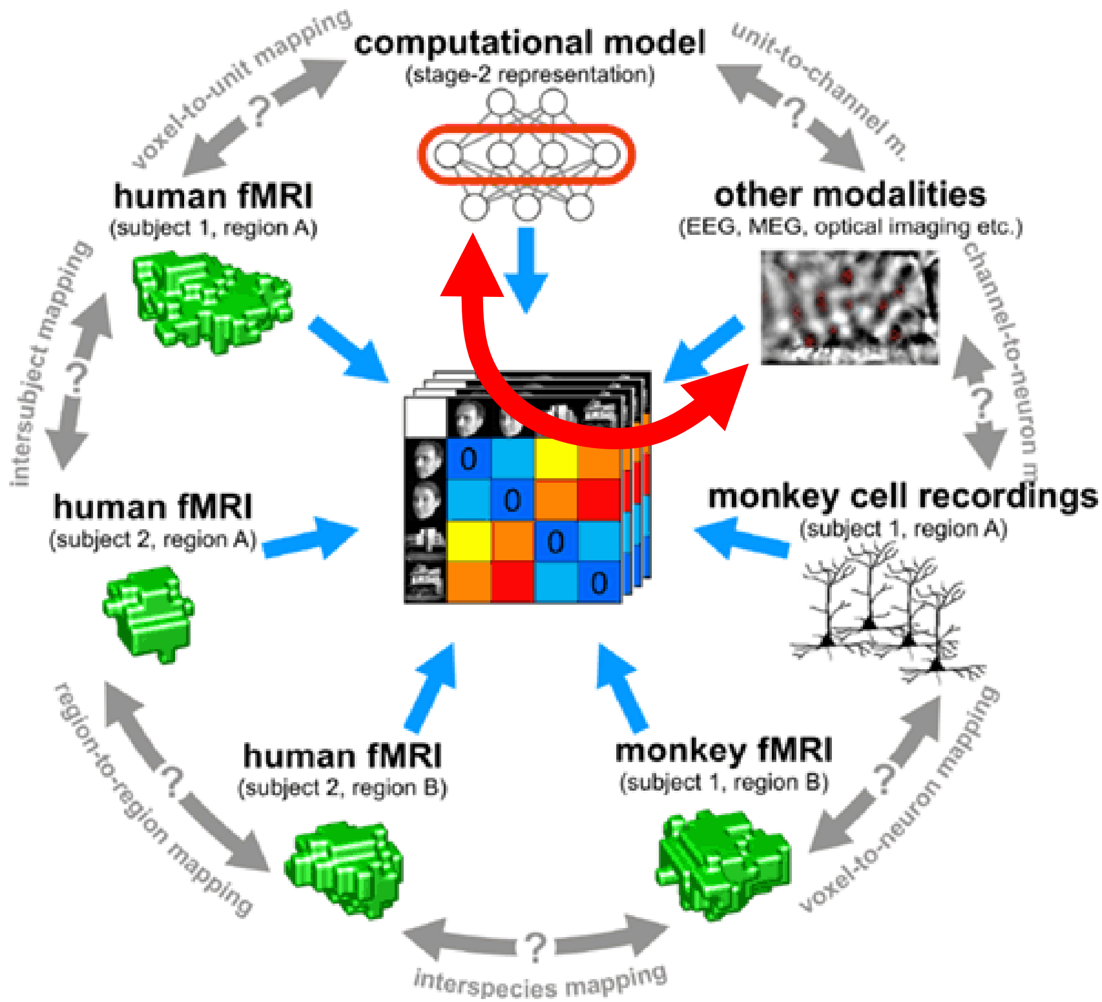




Cichy et al, Sci Reports (2016)

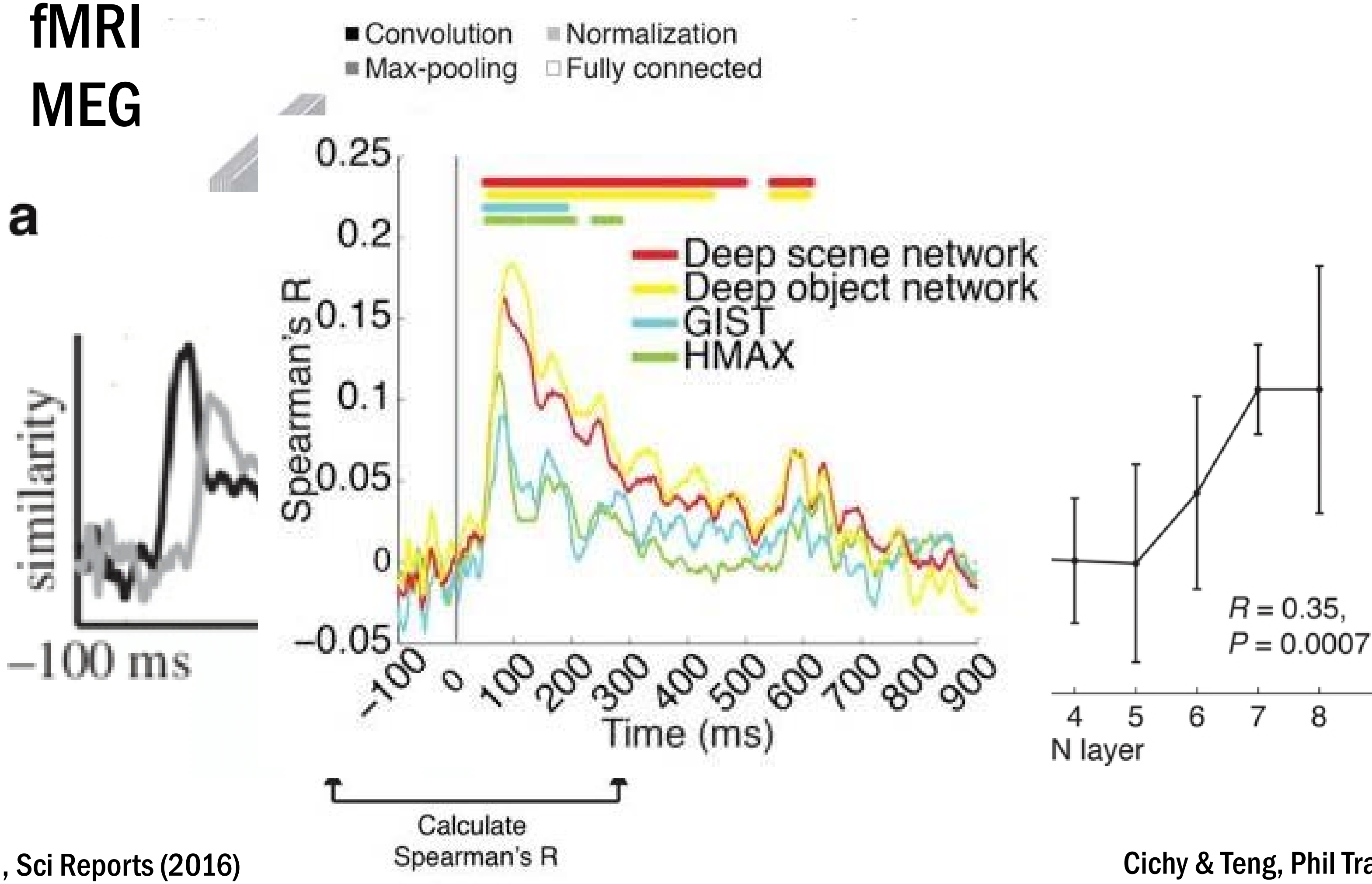
3. Brain/CNN comparisons

- **RSA (representational similarity analysis):**
  - fMRI
  - MEG



3. Brain/CNN comparisons

- **RSA (representational similarity analysis):**
  - fMRI
  - MEG

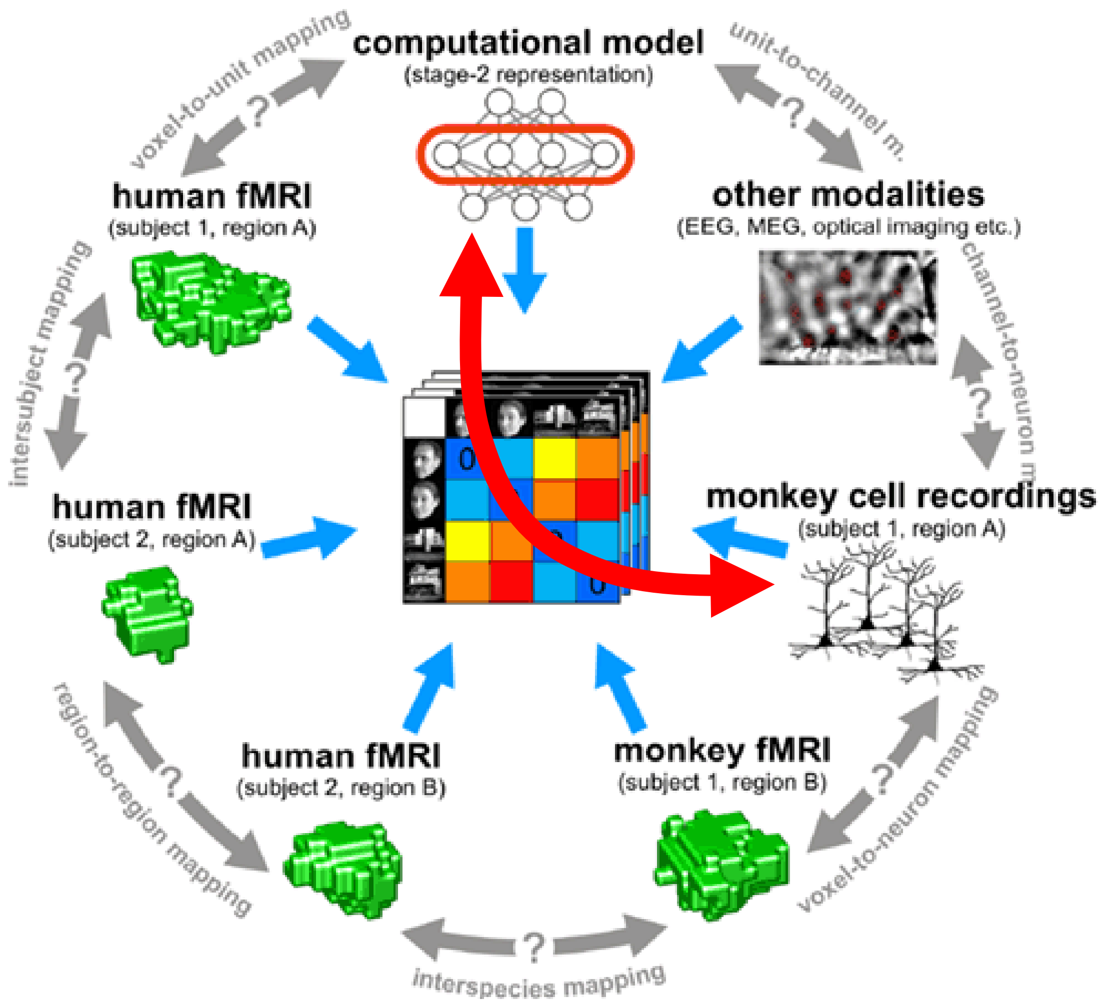


Cichy et al, Sci Reports (2016)

Cichy & Teng, Phil Trans B (2017)

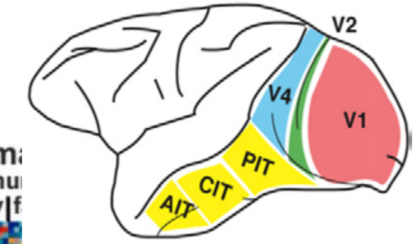
3. Brain/CNN comparisons

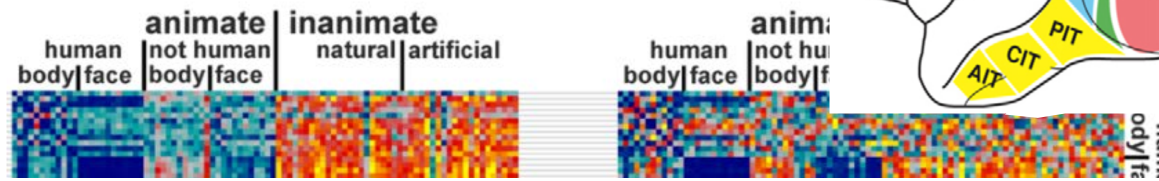
- **RSA (representational similarity analysis):**
  - fMRI
  - MEG
  - Single-units

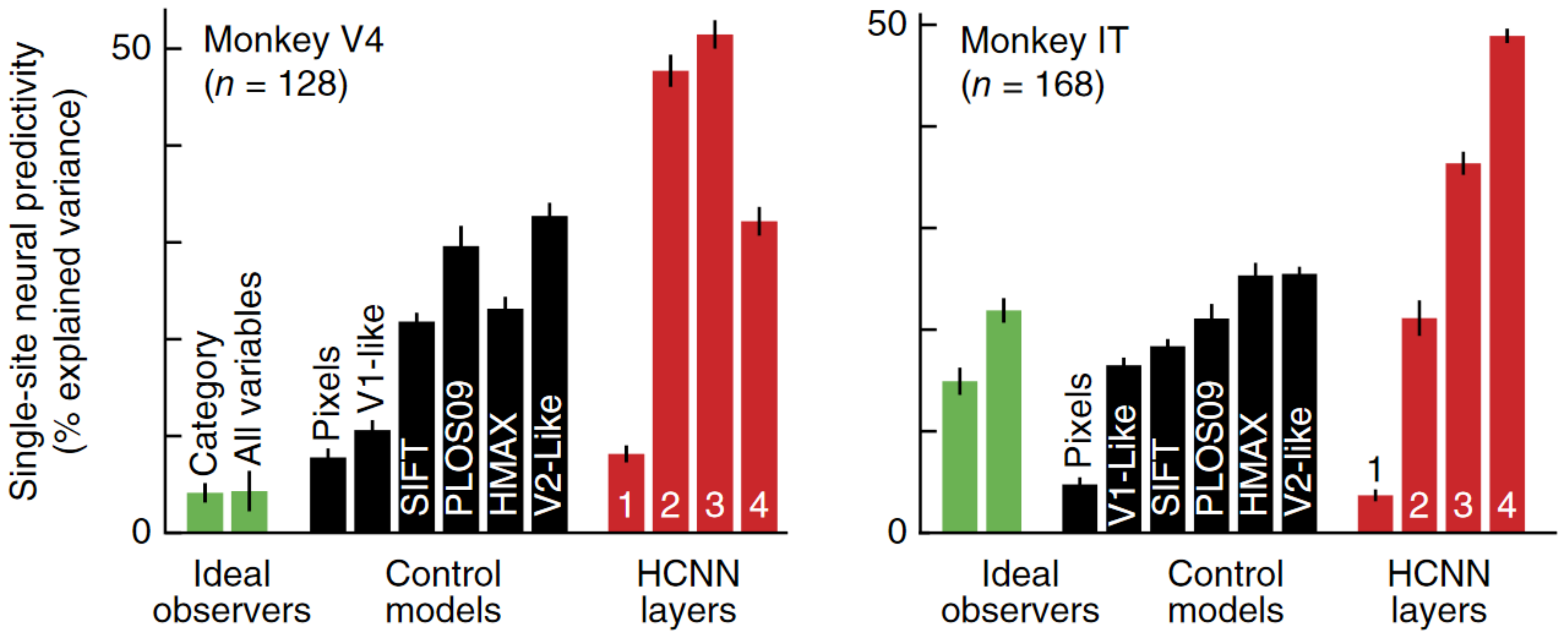


3. Brain/CNN comparisons

- **RSA (representational similarity analysis):**
  - fMRI
  - MEG
  - Single-units







Deep neural network models

Yamins et al, PNAS (2014)

Cadieu et al, PLoS Comp Biol. (2014)

3. Brain/CNN comparisons

→ Brainscore

(www.brain-score.org)

Brai

Brain-Score Leaderboard About Compare Participate

cipate

Rank

1

2

3

4

5

nmarks

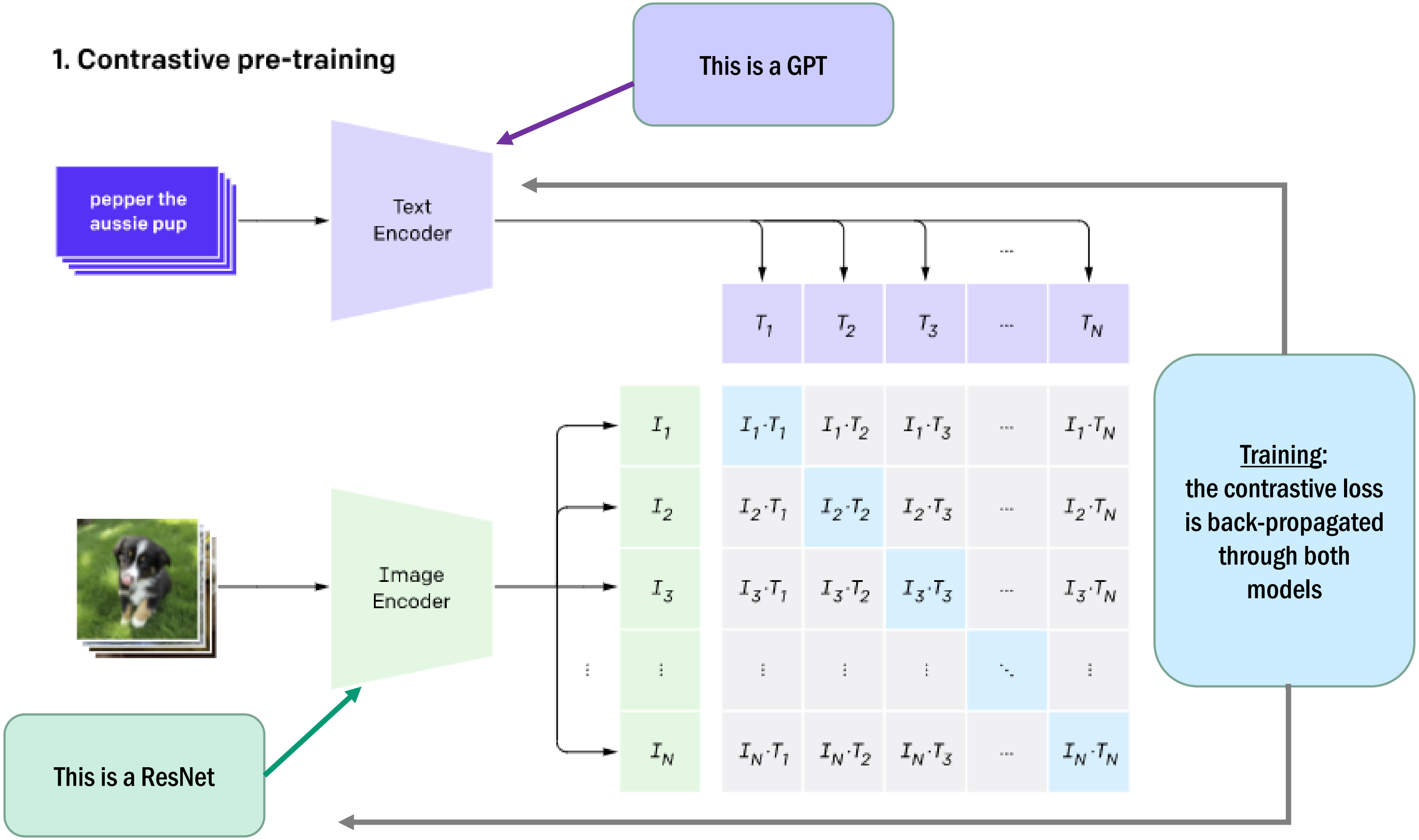
ehavior 1 benchmark

Rank	Model (submitted by)	average	V1 (1 benchmark)	V2 (1 benchmark)	V4 (1 benchmark)	IT (2 benchmarks)	behavior (1 benchmark)	engineering (1 benchmark)	Deng2009-top1 v1
1	CORnet-S *Brain-Score Team*	.417	.294	.242	.581	.423	.545	.747	.747
2	vgg-19 *Brain-Score Team*	.408	.347	.341	.610	.248	.494	.711	.711
3	resnet-50-robust *Joel Dapello*	.408	.378	.365	.537	.243	.515		
4	resnet-101_v1 *Brain-Score Team*	.407	.266	.341	.590	.274	.561	.764	.764
5	vgg-16 *Brain-Score Team*	.406	.355	.336	.620	.259	.461	.715	.715
6	resnet-152_v1 *Brain-Score Team*	.405	.282	.338	.598	.277	.533	.768	.768
7	resnet-101_v2 *Brain-Score Team*	.404	.274	.332	.599	.263	.555	.774	.774
8	resnet50-SIN_IN *Brain-Score Team*	.404	.282	.324	.599	.276	.541	.746	.746
9	densenet-169 *Brain-Score Team*	.404	.281	.322	.601	.274	.543	.759	.759
10	densenet-201 *Brain-Score Team*	.402	.277	.325	.599	.273	.537	.772	.772
11	resnet-50-pytorch *Joel Dapello*	.399	.289	.317	.600	.259	.528	.752	.752
12	resnet-50_v1 *Brain-Score Team*	.398	.274	.317	.594	.278	.526	.752	.752
13	resnet50-SIN_IN_IN *Brain-Score Team*	.397	.275	.321	.596	.273	.523	.767	.767
14	resnet-152_v2 *Brain-Score Team*	.397	.274	.326	.591	.266	.528	.778	.778
15	resnet-50_v2 *Brain-Score Team*	.396	.270	.323	.596	.260	.531	.756	.756

Schrimpf, ...Di Carlo, Neuron (2020)

3. Brain/CNN comparisons

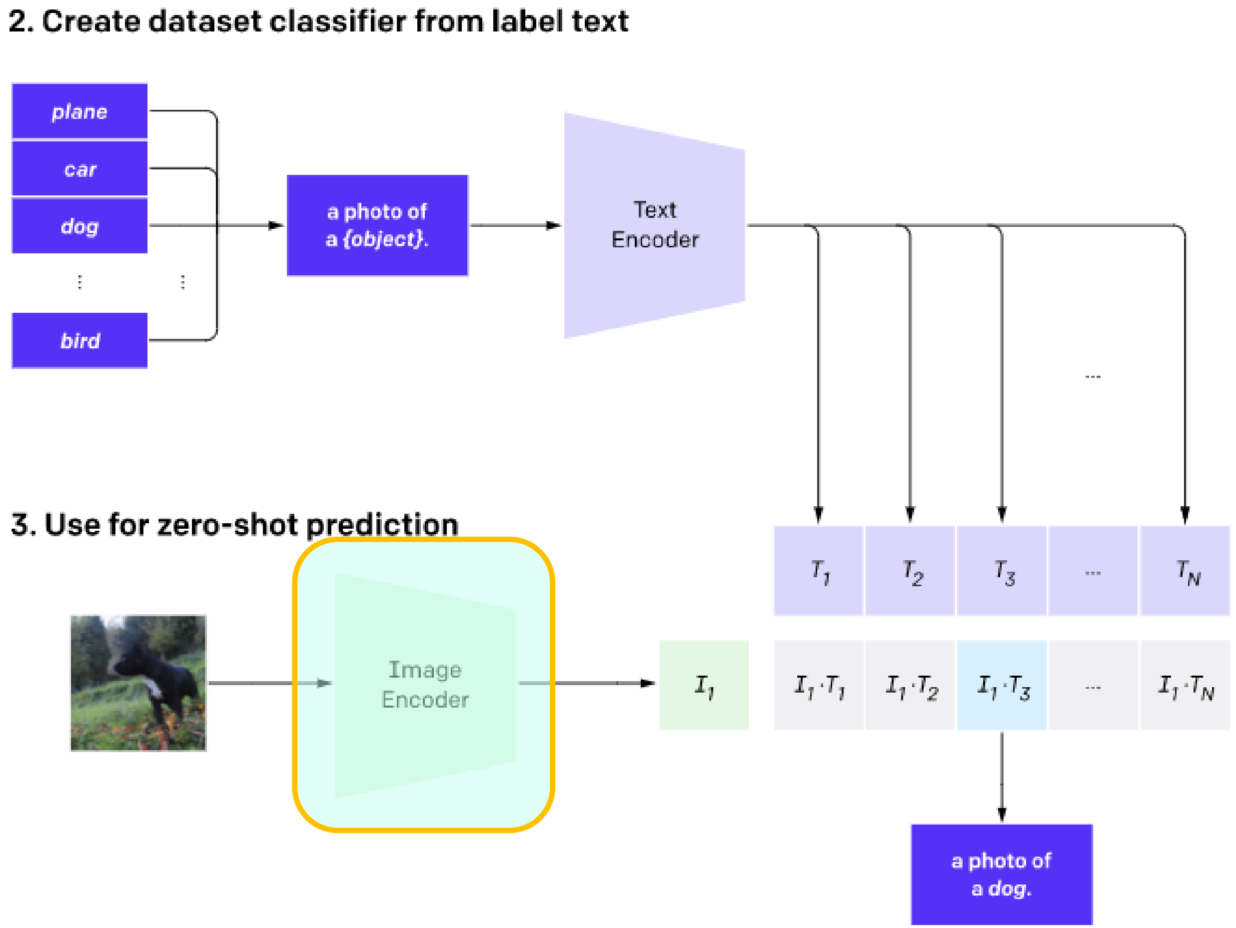
- **Case study: CLIP multimodal neurons = concept cells?**



Radford, et al. (openAI), "Learning Transferable Visual Models From Natural Language Supervision ", arXiv 2021.

3. Brain/CNN comparisons

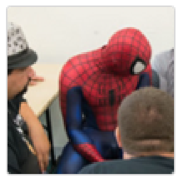
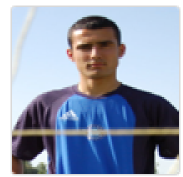
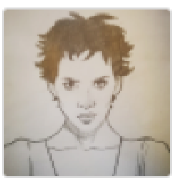
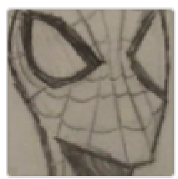
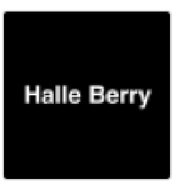
- **Case study: CLIP multimodal neurons = concept cells?**



Radford, et al. (openAI), "Learning Transferable Visual Models From Natural Language Supervision ", arXiv 2021.

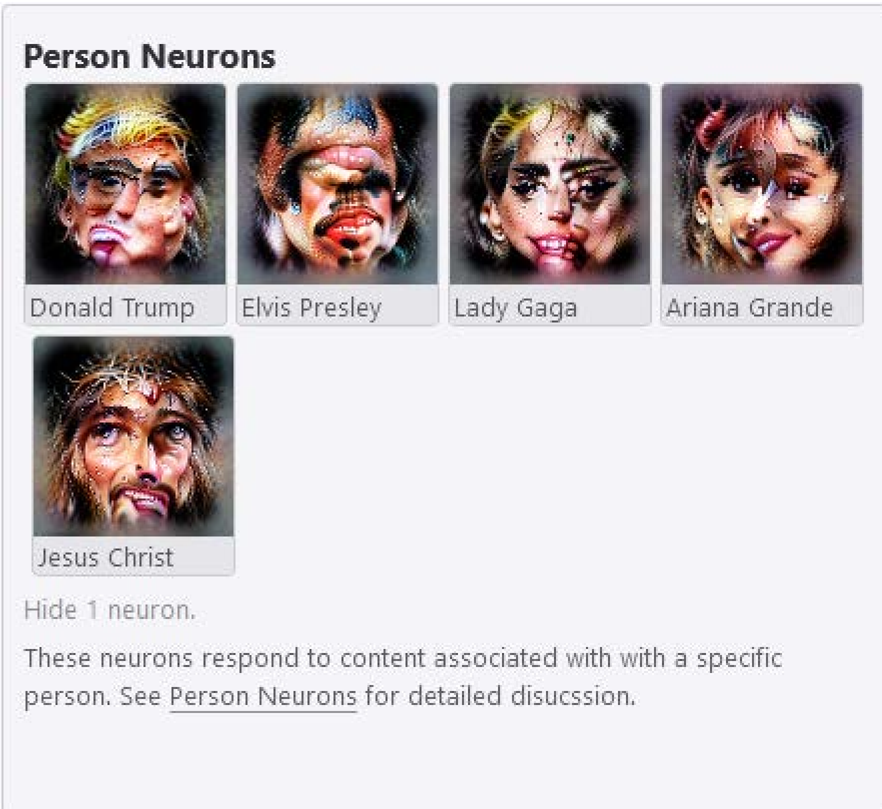
3. Brain/CNN comparisons

- **Case study: CLIP multimodal neurons = concept cells?**

Biological Neuron	CLIP Neuron	Previous Artificial Neuron	
Probed via depth electrodes	Neuron 244 from penultimate layer in CLIP RN50_4x	Neuron 483, generic person detector from Inception v1	
Halle Berry	Spiderman	human face	
 Responds to photos of Halle Berry and Halle Berry in costume ✓	 Responds to photos of Spiderman in costume and spiders ✓ view more	 Responds to faces of people ✓	Photorealistic images
 Responds to skeches of Halle Berry ✓	 Responds to comics or drawings of Spiderman and spider-themed icons ✓ view more	Does not respond significantly to drawings of faces ✗	Conceptual drawings
 Responds to the text "Halle Berry" ✓	 Responds to the text "spider" and others ✓ view more	Does not respond significantly to text ✗	Images of text

Goh, et al., "Multimodal Neurons in Artificial Neural Networks", Distill, 2021.

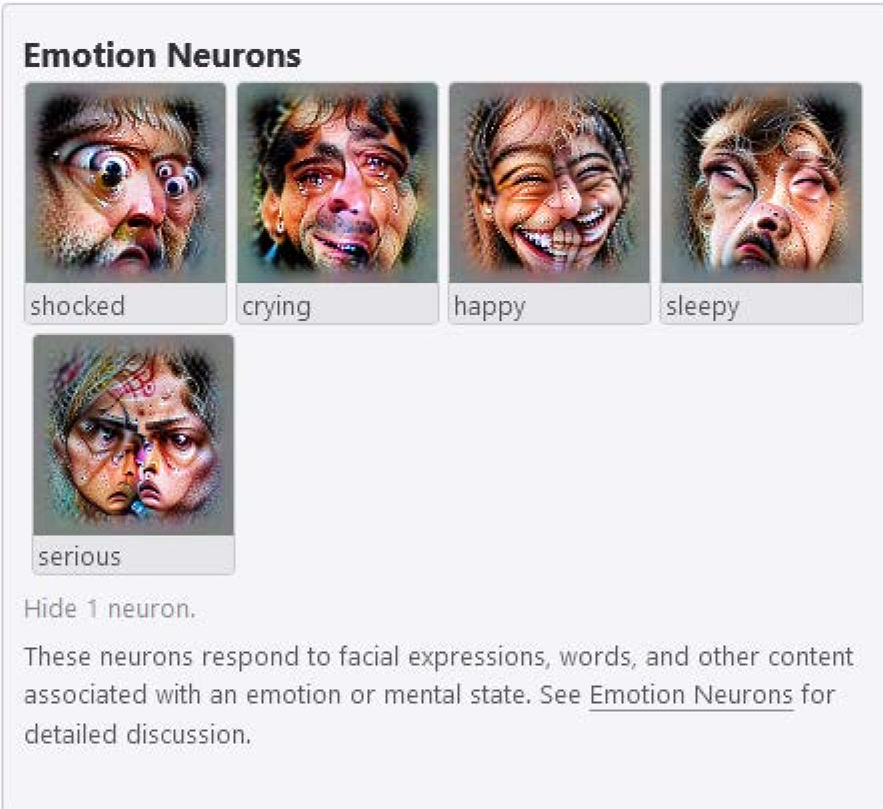
- **Case study: CLIP multimodal neurons = concept cells?**
  - ➔ **Are these « grandmother » neurons?**



Goh, et al., "Multimodal Neurons in Artificial Neural Networks", Distill, 2021.

3. Brain/CNN comparisons

- **Case study: CLIP multimodal neurons = concept cells?**

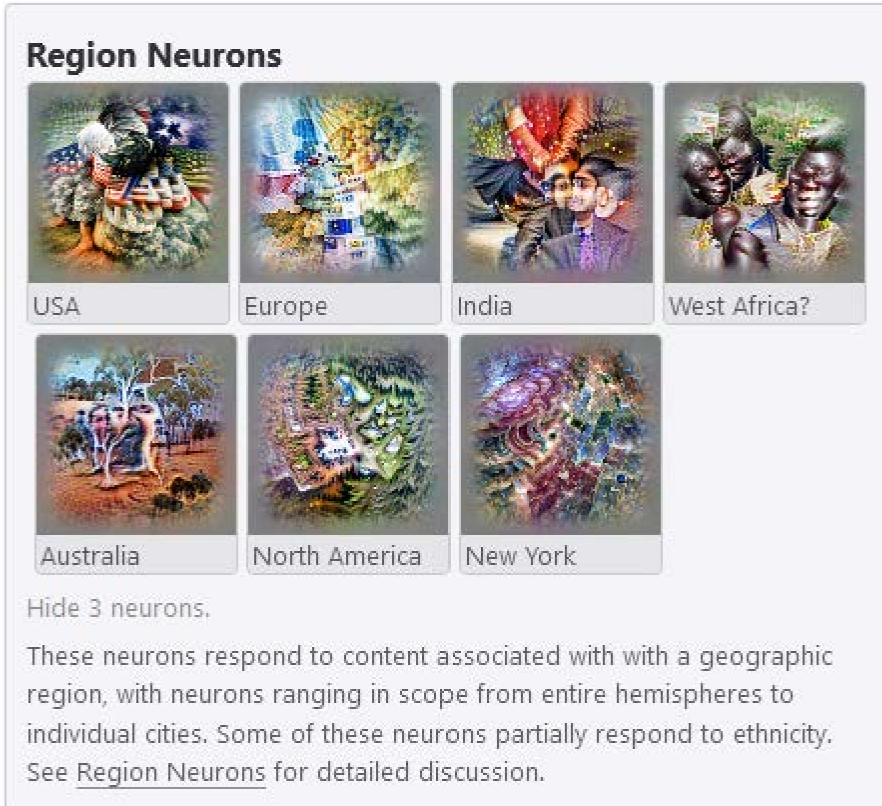


Surprise / Shock

Goh, et al., "Multimodal Neurons in Artificial Neural Networks", Distill, 2021.

3. Brain/CNN comparisons

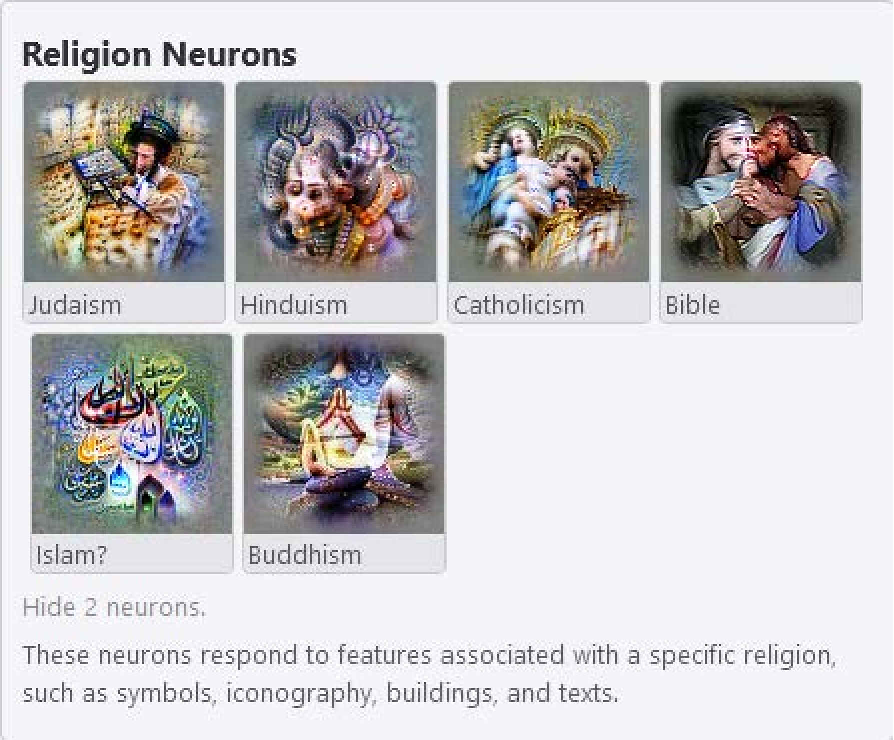
- **Case study: CLIP multimodal neurons = concept cells?**



Goh, et al., "Multimodal Neurons in Artificial Neural Networks", Distill, 2021.

3. Brain/CNN comparisons

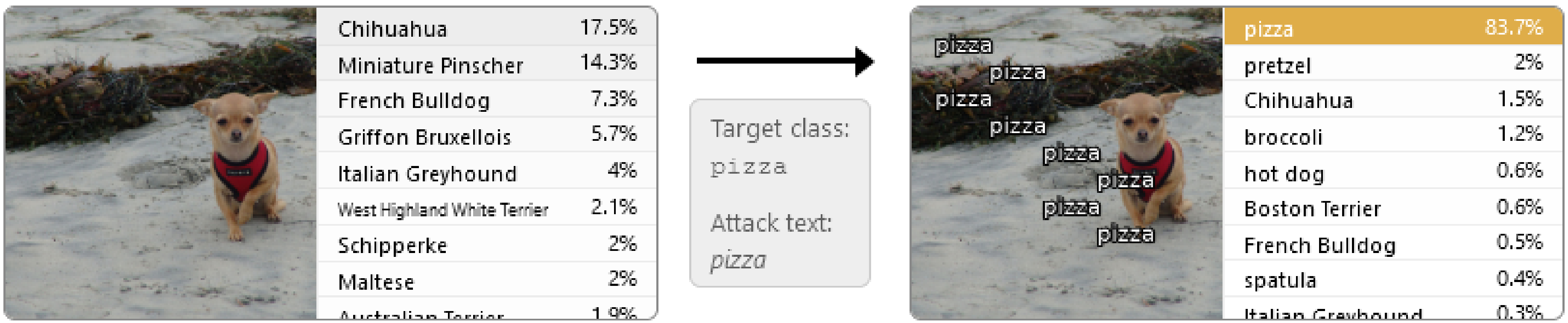
- **Case study: CLIP multimodal neurons = concept cells?**



Goh, et al., "Multimodal Neurons in Artificial Neural Networks", Distill, 2021.

3. Brain/CNN comparisons

- **Case study: CLIP multimodal neurons = concept cells?**
  - ➔ **Not fully like humans, yet…**



Goh, et al., "Multimodal Neurons in Artificial Neural Networks", Distill, 2021.

4. Other issues on DL biological plausibility

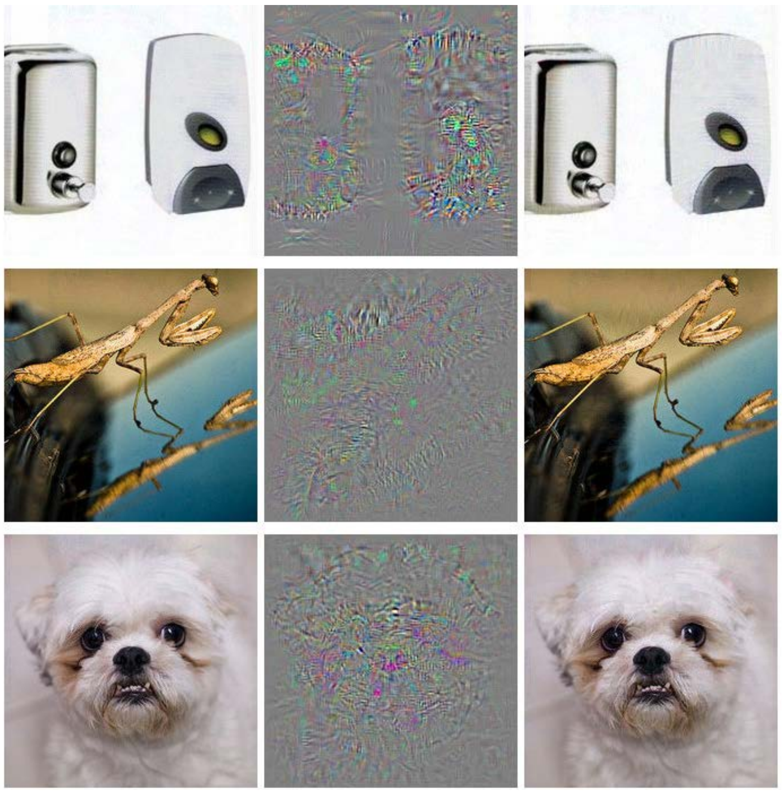
- **CNNs are (roughly) biologically plausible:**
  - Hierarchical structure
  - Convolutions
  - Receptive fields
  - Feature/object selectivity (RSA, BrainScore, concept cells)

- **Other aspects of Deep Learning are not:**
  1. **Spikes** (vs. continuous/floating point values)
  2. **Adversarial attacks!**
  3. **Backpropagation** (globally available error signals?)
  4. **Visual attention/Transformers** (Attention control within the feature extraction hierarchy?)
  5. **Feed-forward models** (recurrence is not just for text/audio inputs)

4. Other issues on DL biological plausibility

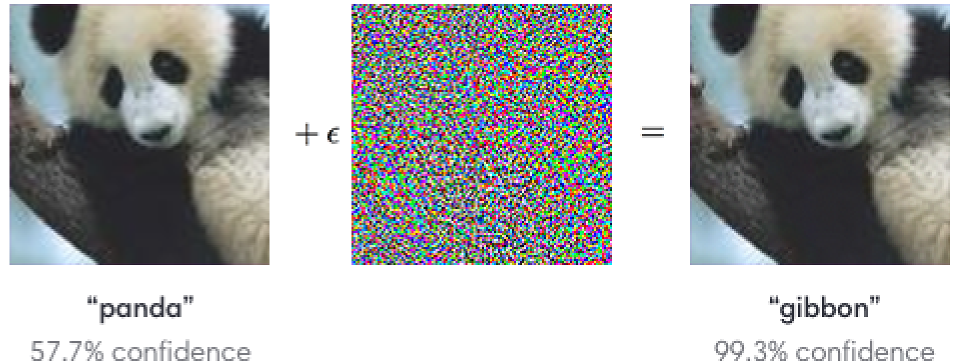
2. Adversarial attacks

original | “attack” | original + attack =“ostrich”



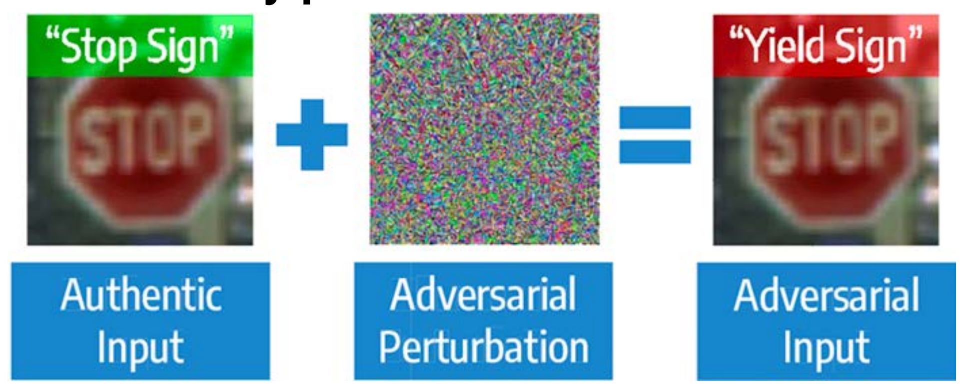
Szegedy et al, 2013

Fast gradient sign method, Goodfellow et al, 2014



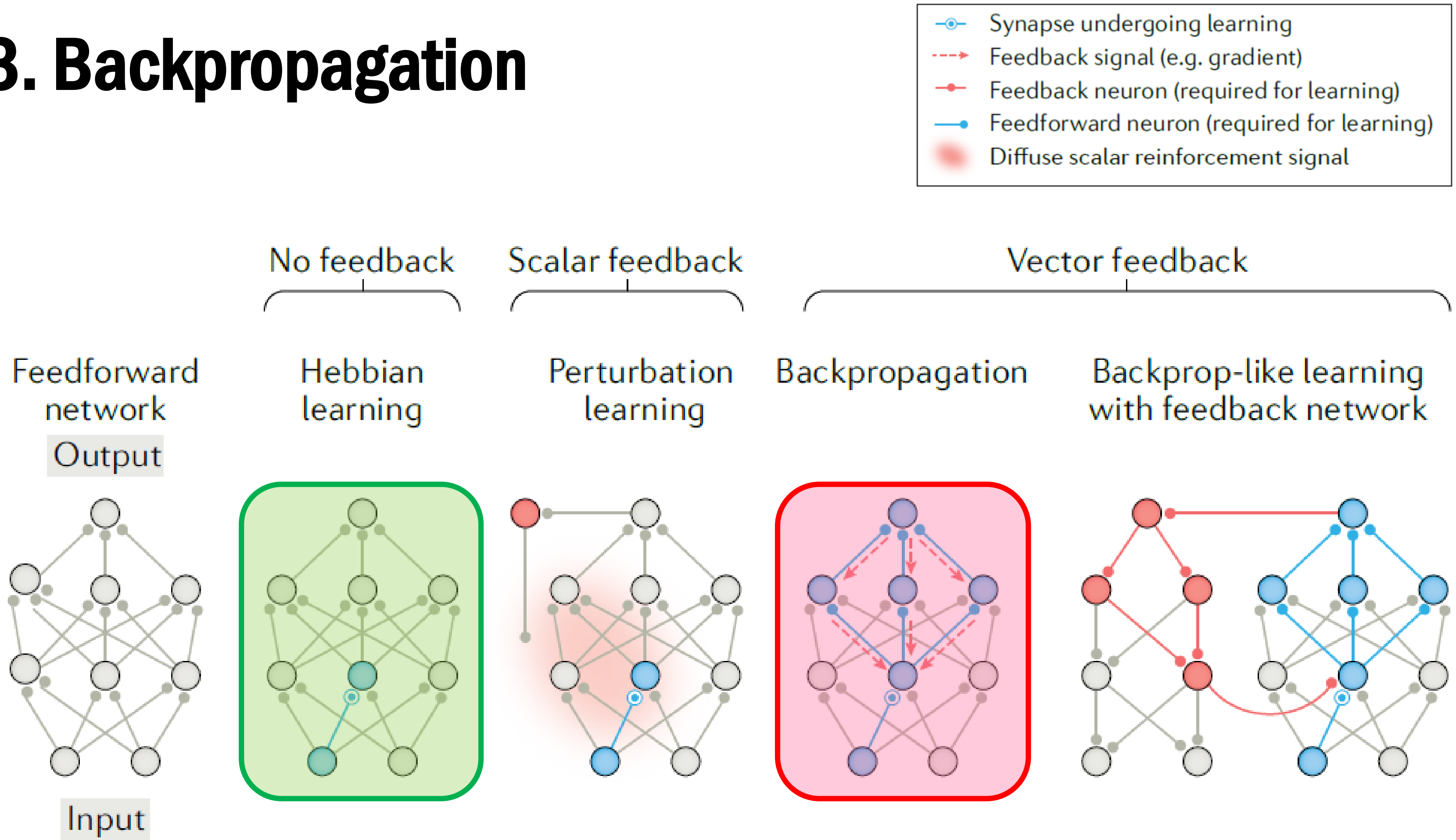
$$x' = x + \epsilon \cdot sign(\nabla_x J(\theta, x, y))$$

Can be very problematic for AI



4. Other issues on DL biological plausibility

3. Backpropagation



Lilicrap et al, Nature Reviews Neuroscience 2020

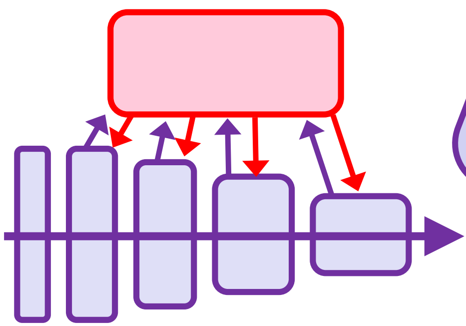
4. Other issues on DL biological plausibility

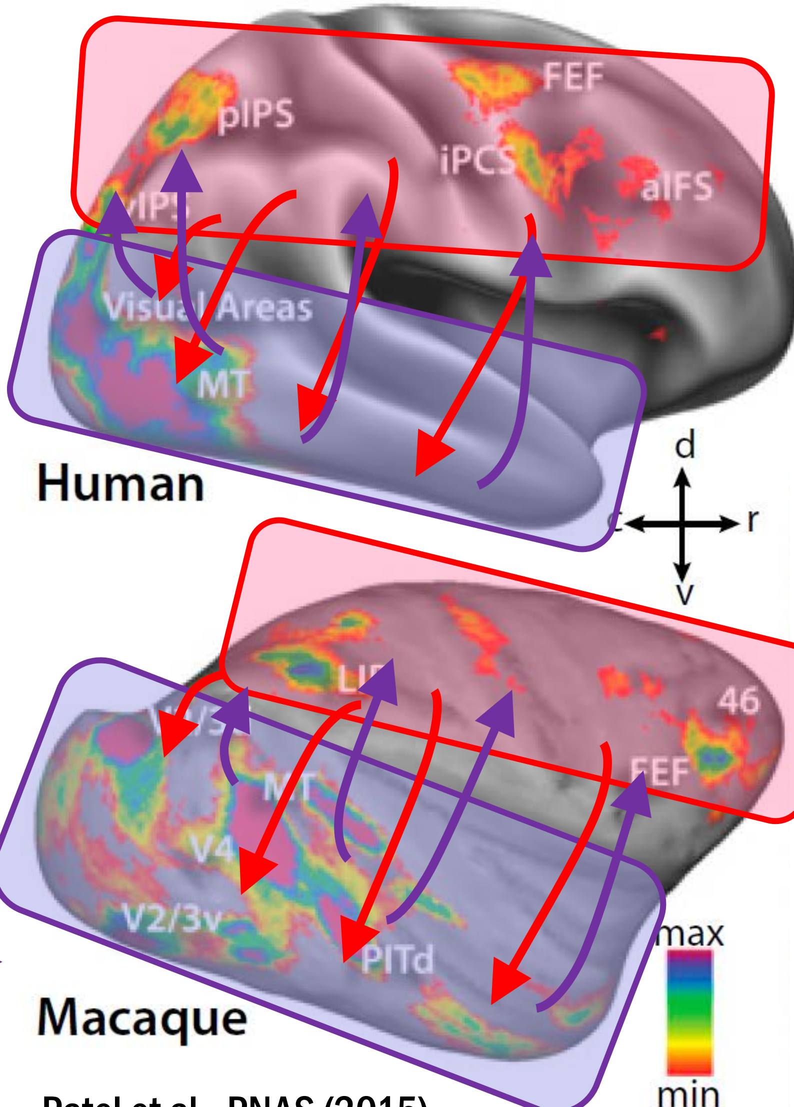
- **CNNs are (roughly) biologically plausible:**
  - Hierarchical structure
  - Convolutions
  - Receptive fields
  - Feature/object selectivity (RSA, BrainScore , concept cells)

- **Other aspects of Deep Learning are not:**
  1. **Spikes** (vs. continuous/floating point values)
  2. **Adversarial attacks!**
  3. **Backpropagation** (globally available error signals?)
  4. **Visual attention/Transformers** (Attention control within the feature extraction hierarchy?)

Visual attention in the brain

Hierarchy of visual areas
≈
deep conv. net

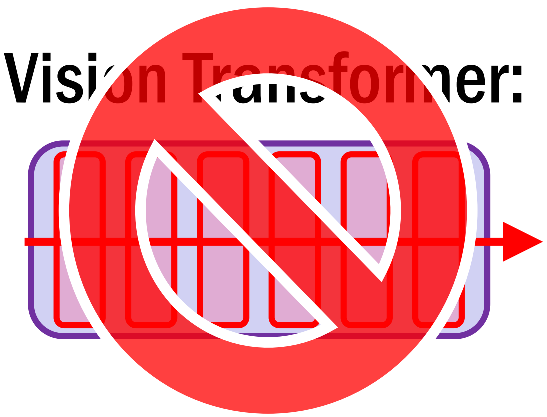




Patel et al., PNAS (2015)

Frontoparietal attention network

Vision Transformer:



Visual attention in the brain

→ Brainscore

(www.brain-score.org)

Brain-Score   Leaderboard   About   Compare   Participate

87	mobilenet_v2_0.75_96 *Brain-Score Team*	.350	.208	.305	.527	.258	.451	.588	.588
88	squeezenet1_0 *Brain-Score Team*	.341	.304	.320	.591	.229	.263	.575	.575
89	mobilenet_v1_0.5_128 *Brain-Score Team*	.341	.245	.304	.550	.234	.373	.563	.563
90	barlow-twins-resnet50 *Eric Elmoznino*	.341	.279	.284	.579	.286	.276	.001	.001
91	ViT-B/32 *Violet Xiang*	.341	.282	.339	.549	.251	.284		
92	squeezenet1_1 *Brain-Score Team*	.336	.265	.311	.582	.229	.291	.575	.575
93	mobilenet_v2_0.35_128 *Brain-Score Team*	.333	.245	.289	.530	.235	.367	.508	.508
94	mobilenet_v2_0.5_96 *Brain-Score Team*	.331	.266	.278	.501	.239	.370	.512	.512
95	RN50 *Violet Xiang*	.330	.273	.315	.588	.252	.220		
96	ViT_L_32_imagenet1k *Paul Mc Grath*	.328	.265	.291	.531	.227	.324		
97	mobilenet_v1_0.25_224 *Brain-Score Team*	.327	.231	.296	.538	.240	.333	.498	.498
98	deit_base_patch16_384_id *Violet Xiang*	.324	.209	.248	.515	.225	.425		
99	ViT_L_32 *Paul Mc Grath*	.324	.305	.301	.511	.219	.286		
100	mobilenet_v1_0.25_192 *Brain-Score Team*	.323	.208	.318	.517	.226	.344	.477	.477
101	CORnet-Z *Brain-Score Team*	.322	.298	.182	.553	.223	.356	.470	.470
102	resnet18-simclr *Chengxu Zhuang*	.321	.243	.318	.550	.262	.231		
103	ViT_B_32_imagenet1k *Paul Mc Grath*	.317	.271	.285	.536	.219	.276		
104	resnet18-local_aggregation *Chengxu Zhuang*	.314	.253	.308	.563	.268	.177		
105	ViT_B_32 *Paul Mc Grath*	.313	.308	.275	.504	.208	.270		

Vision Transformers are not very close to brain processing

Schrimpf, ...Di Carlo, Neuron (2020)

4. Other issues on DL biological plausibility

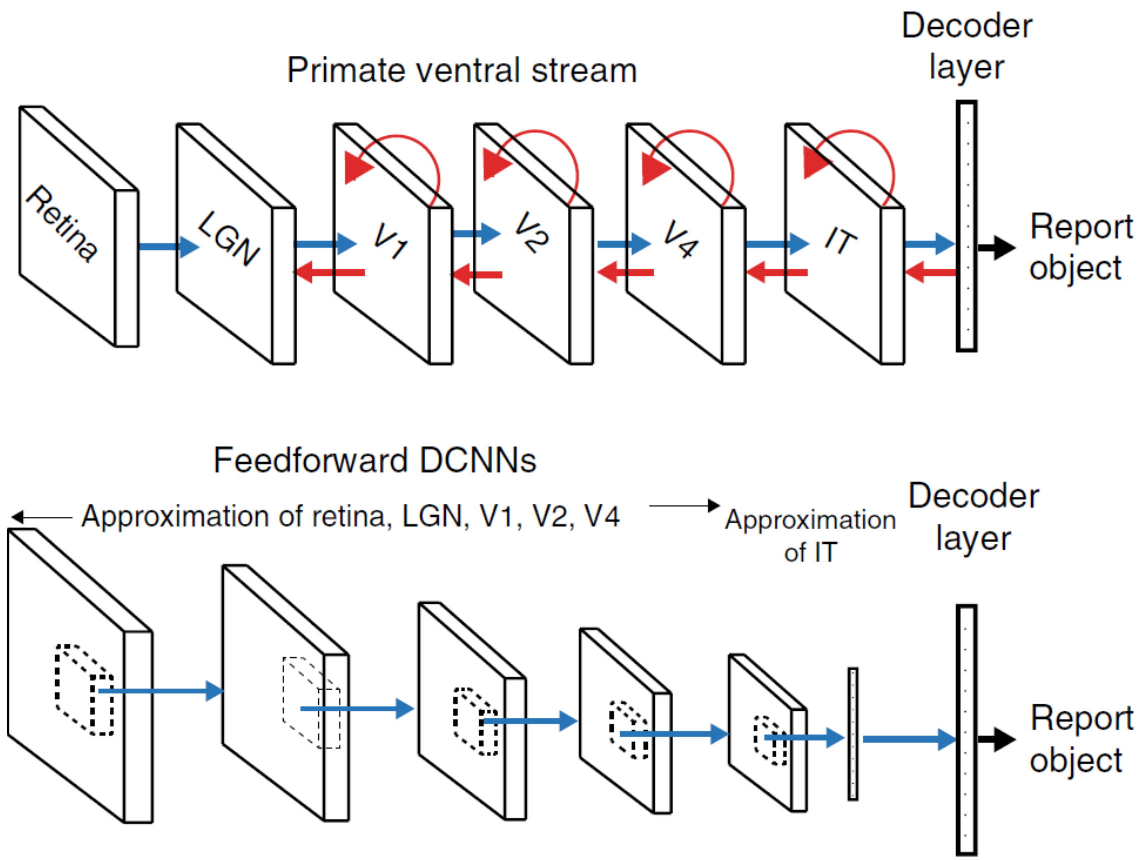
- **CNNs are (roughly) biologically plausible:**
  - Hierarchical structure
  - Convolutions
  - Receptive fields
  - Feature/object selectivity (RSA, BrainScore , concept cells)

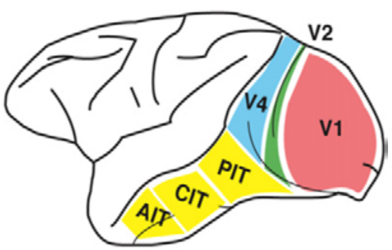
- **Other aspects of Deep Learning are not:**
  1. **Spikes** (vs. continuous/floating point values)
  2. **Adversarial attacks!**
  3. **Backpropagation** (globally available error signals?)
  4. **Visual attention/Transformers** (Attention control within the feature extraction hierarchy?)
  5. **Feed-forward models** (recurrence is not just for text/audio inputs)

4. Other issues on DL biological plausibility

5. Feed-forward models?

- May be a good model for rapid, automatic vision in the brain
- But not for conscious/attentive perception





Kar et al, Nat. Neurosci 2019

CONCLUSION

- **CNNs are (roughly) biologically plausible:**
  - Hierarchical structure
  - Convolutions
  - Receptive fields
  - Feature/object selectivity (RSA, BrainScore, concept cells)

- **Other aspects of Deep Learning are not:**
  1. **Spikes** (vs. continuous/floating point values)
  2. **Adversarial attacks!**
  3. **Backpropagation** (globally available error signals?)
  4. **Visual attention/Transformers** (Attention control within the feature extraction hierarchy?)
  5. **Feed-forward models** (recurrence is not just for text/audio inputs)

...