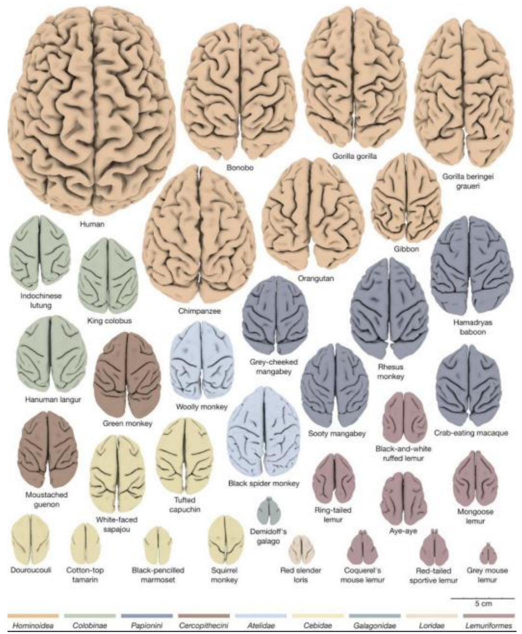
Introduction to Deep Convolutional Neural Networks

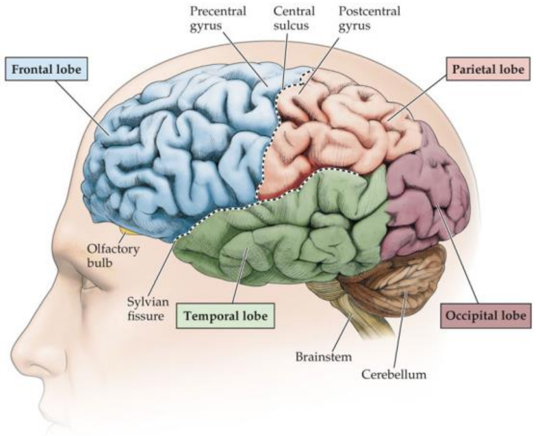
Aude Oliva, PhD

neuromatch
academy

Species of Deep Neural Networks



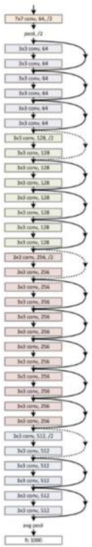
From Heuer et al (2019). Evolution of neocortical Folding. bioRxiv



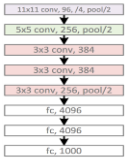
THE MIND'S MACHINE 2e, Figure 2.10
© 2016 Sinauer Associates, Inc.

Human brain: 100 billion neurons, 100 trillion connections 10,000 different types of neurons, 1000 > new neurons per day, consumes 20 watts

VGG 14



ResNet 15



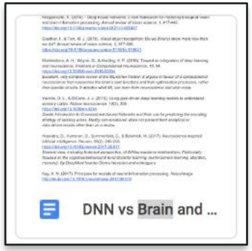
AlexNet

GoogLeNet 14

Paper Title	Authors (First, Last)	Year	Journal
Representational Similarity Analysis – Connecting the Branches of Systems Neuroscience	Kriegeskorte, Bandettini	2008 Nov 24	Frontiers in Systems Neuroscience
Predicting Human Brain Activity Associated with the Meanings of Nouns	Mitchell, Just	2008 May 30	Science
Resolving human object recognition in space and time	Cichy, Oliva	2014 Jan 26	Nature Neuroscience
Performance-optimized hierarchical models predict neural responses in higher visual cortex	Yamins, DiCarlo	2014 May 8	PNAS [Edited by Terrence Sejnowski]
Deep Supervised, but Not Unsupervised, Models May Explain IT Cortical Representation	Khaligh-Razavi, Kriegeskorte	2014 Nov 6	PloS Comp Biol
Deep Neural Networks Rival the Representation of Primate IT Cortex for Core Visual Object Recognition	Cadieu, DiCarlo	2014 Dec 18	PloS Comp Biol
Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence	Cichy, Oliva	2015 Jun 10	Nature Sci Reports
Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream	Güçlü, van Gerven	2015 July 8	J Neurosci
Deep Neural Networks: A New Framework for Modeling Biological Vision and Brain Information Processing [REVIEW]	Kriegeskorte	2015 Nov	Annual Review of Vision Science
Increasingly complex representations of natural movies across the dorsal stream are shared between subjects	Güçlü, van Gerven	2015 Dec 24	Neuroimage
Explicit information for category-orthogonal object properties increases along the ventral stream	Hong, DiCarlo	2016 Feb 22	Nature Neuroscience
Using goal-driven deep learning models to understand sensory cortex	Yamins, DiCarlo	2016 Feb 23	Nature Neuroscience
Natural speech reveals the semantic maps that tile human cerebral cortex	Huth, Gallant	2016 Apr 28	Nature
The Semantics of Adjective Noun Phrases in the Human Brain	Fyshe, Alona	2016 Nov 25	bioRxiv
Brains on beats	Güçlü, vanGerven	2016 Dec	NIPS Conference
A balanced comparison of object invariances in monkey IT neurons	Murty, Arun	2017 Apr 5	eNeuro
Representational models: A common framework for understanding encoding, pattern-component, and representational-similarity analysis	Diedrichsen, Kriegeskorte	2017 Apr 24	PLOS Comp Bio
Generic decoding of seen and imagined objects using hierarchical visual features	Horikawa, Kamitani	2017 May 22	Nature Commun
Shape Selectivity of Middle Superior Temporal Sulcus Body Patch Neurons	Kalfas, Vogels	2017 Jun 16	eNeuro
Convolutional neural network-based encoding and decoding of visual object recognition in space and time	Seeliger, vanGerven	2017 Jul 16	Neuroimage

Paper Title	Authors (First, Last)	Year	Journal
Deep recurrent neural network reveals a hierarchy of process memory during dynamic natural vision	Shi, Liu	2018 Feb 12	Human Brain Mapping
Deep Residual Network Reveals a Nested Hierarchy of Distributed Cortical Representation for Visual Categorization	Wen, Liu	2018 Feb 28	Nature Scientific Reports
Distinct contributions of functional and deep neural network features to scene representation in brain and behavior	Groen, Baker	2018 Mar 7	eLife
Multiplicative mixing of object identity and image attributes in single inferior temporal neurons	Murty, Arun	2018 Mar 20	PNAS
Using human brain activity to guide machine learning	Fong, Cox	2018 Mar 29	Nature Scientific Reports
ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness	Geirhos, Brendel	2018 Apr	ICLR 2019 Conference
A task-optimized neural network replicates human auditory behavior, predicts brain responses, and reveals a cortical processing hierarchy	Kell, McDermott	2018 Apr 19	Neuron
Generative Adversarial Networks Conditioned on Brain Activity Reconstruct Seen Images	St-Yves & Naselaris	2018 Apr 20	IEEE SMC Conference
Computational mechanisms underlying cortical responses to the affordance properties of visual scenes	Bonner, Epstein	2018 Apr 23	PloS Computational Biology
Vector-based navigation using grid-like representations in artificial agents	Banino, Kumaran	2018 May 9	Nature
Mapping between fMRI responses to movies and their natural language annotations	Vodrahalli, Arora	2018 Jun 23	Neuroimage
Integrated deep visual and semantic attractor neural networks predict fMRI pattern-information along the ventral object processing pathway	Devereaux, Tyler	2018 Jul 13	Nature Scientific Reports
Generative adversarial networks for reconstructing natural images from brain activity	Seeliger, vanGerven	2018 Jul 20	Neuroimage
Recurrent computations for visual pattern completion	Tang, Kreiman	2018 Aug 13	PNAS [Edited by Terrence Sejnowski]
Human brain activity during mental imagery exhibits signatures of inference in a hierarchical generative model	Breedlove, Naselaris	2018 Nov 5	bioRxiv (preprint)
Interpreting Encoding and Decoding Models	Kriegeskorte, Douglas	2018 Dec 1	arXiv q-bio.NC
Rapid Transformation from Auditory to Linguistic Representations of Continuous Speech	Brodbeck, Christian L	2018 Dec 17	Current Biology
Training on the test set? An analysis of Spampinato et al.	Li, Siskind	2018 Dec 18	arXiv cs.CV
Inception in visual cortex: in vivo-silico loops reveal most exciting images	Walker, Tolias	2018 Dec 28	bioRxiv
Evolving super stimuli for real neurons using deep generative networks	Ponce, Livingstone	2019 Jan 17	bioRxiv

Martin Hebart



Basics of Convolutional Neural Networks (CNN)

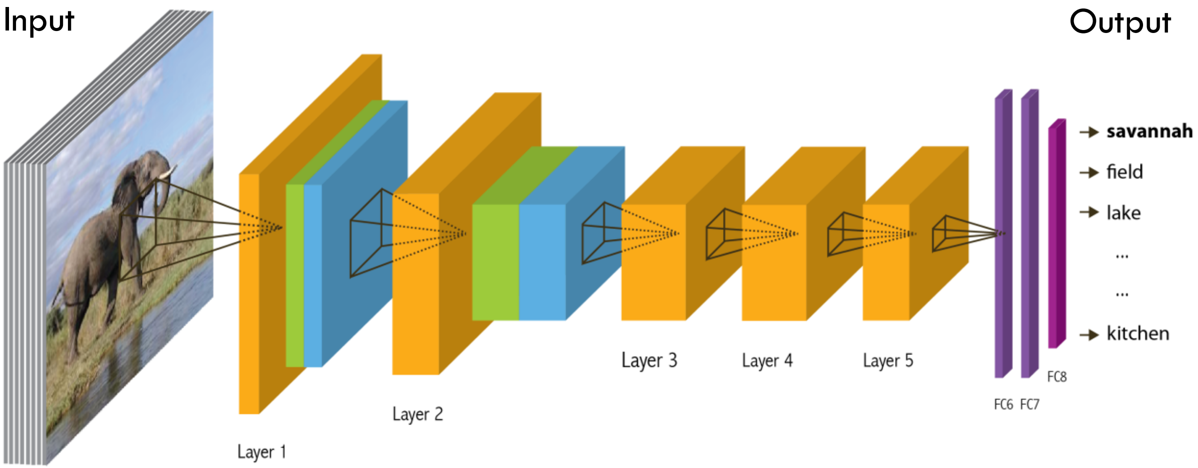
Convolutional Neural Network: AlexNet

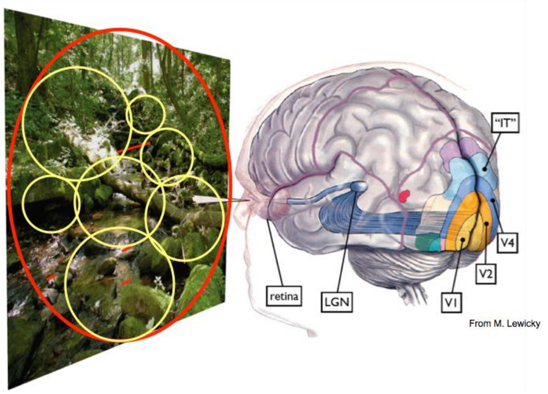
convolution

max-pooling

normalization

full connected





Krizhevsky et al (2012).

Convolution

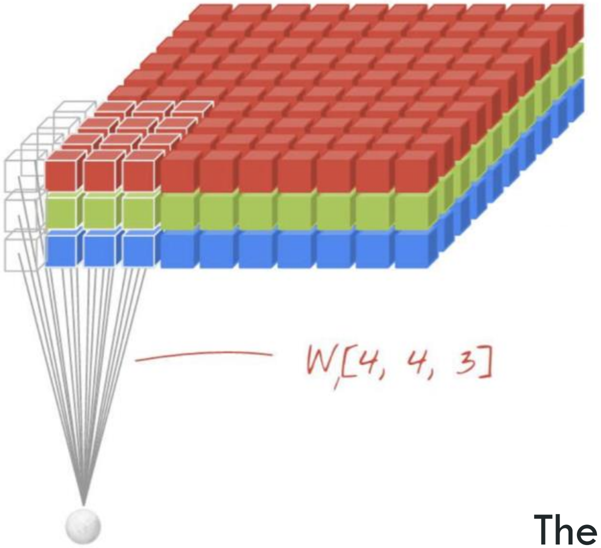
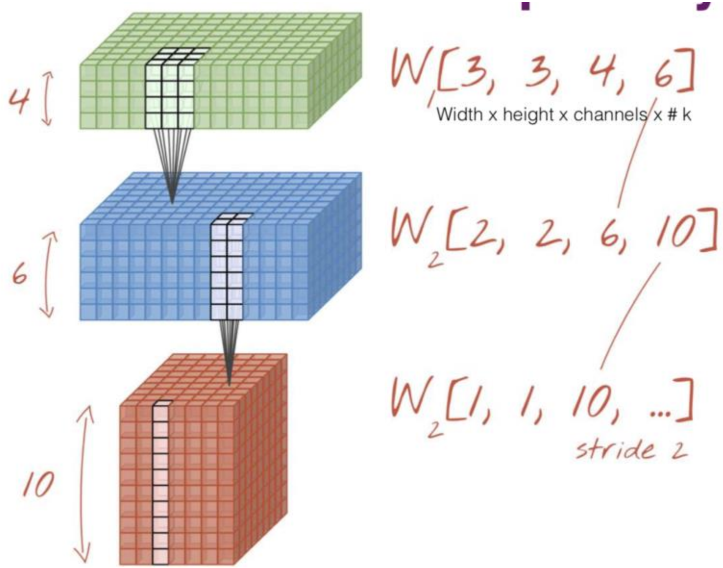
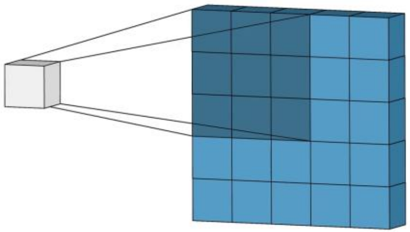


Image credit: codelabs google



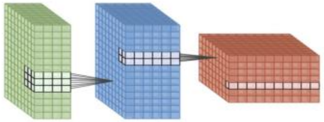
The convolution operation performs a ***summarization*** of each region

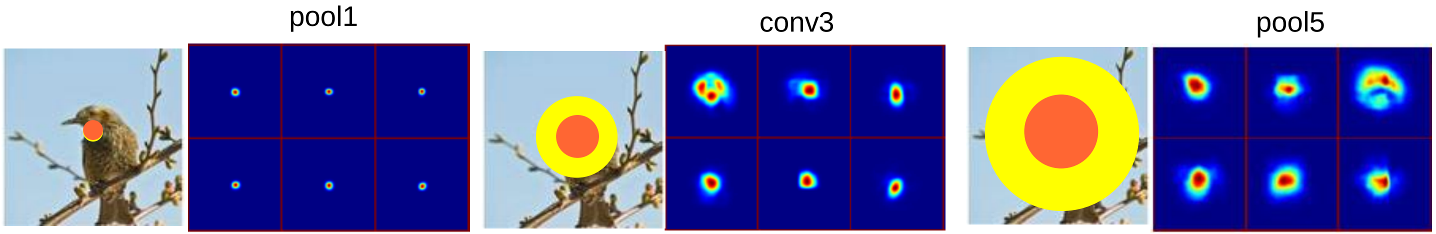
Flashlight metaphor



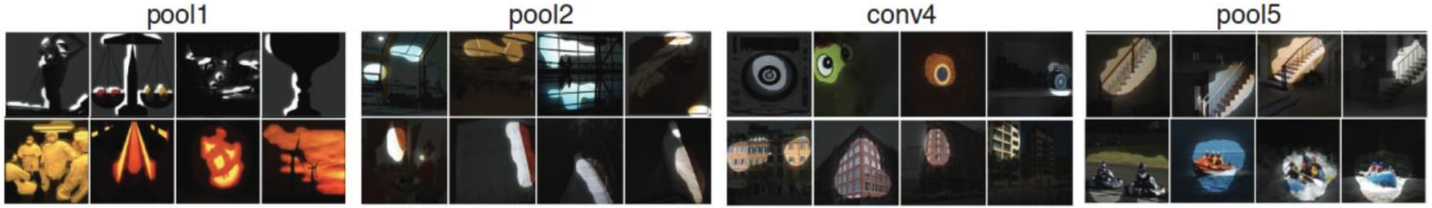
Flashlight: filter / neuron / kernel
Sliding motion: convolution
Illuminated region: receptive field

Desirable outcomes of CNN





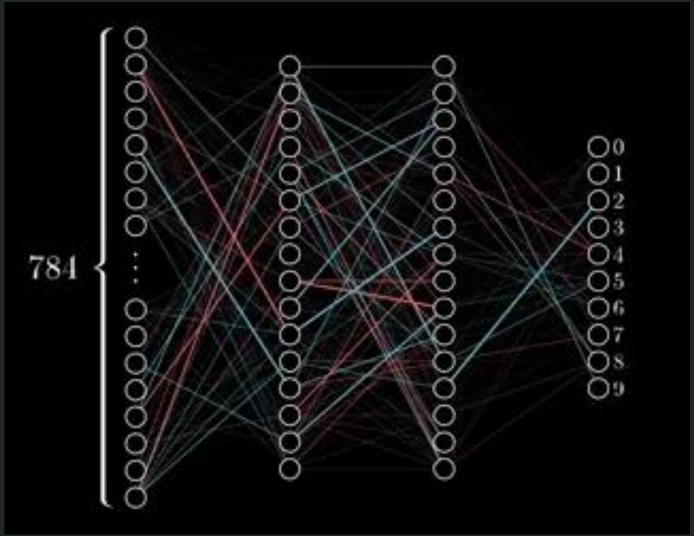
Units cover larger zones of the input from early to late layers



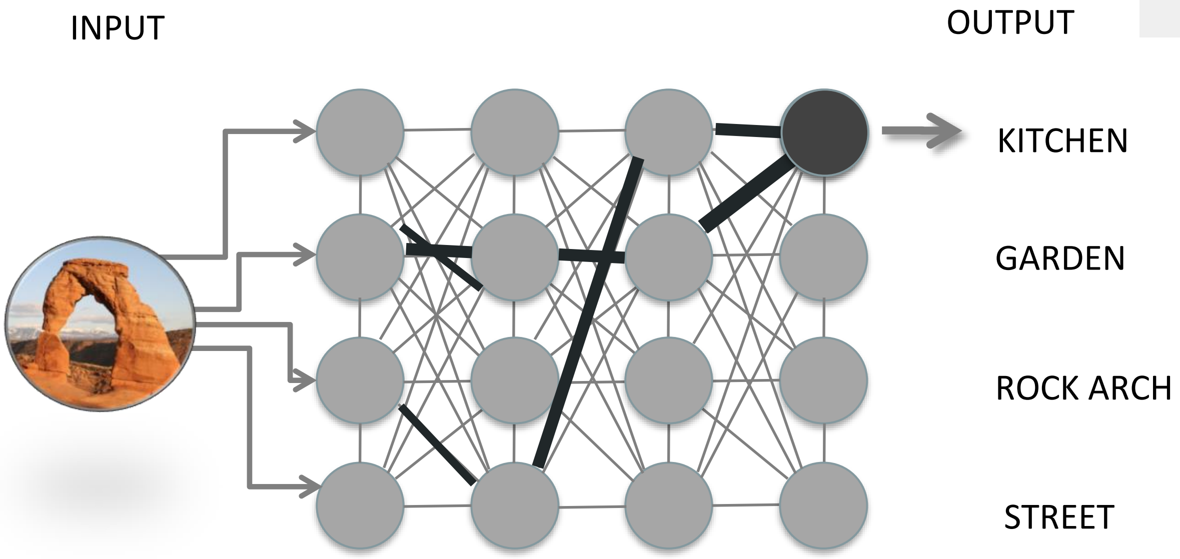
Features are more complex from early to late layers

Basic of CNN learning and training

3BLUE1BROWN SERIES S3 • E3
What is backpropagation really doing? | Deep learning, chapter 3



Cartoon of a Neural Network

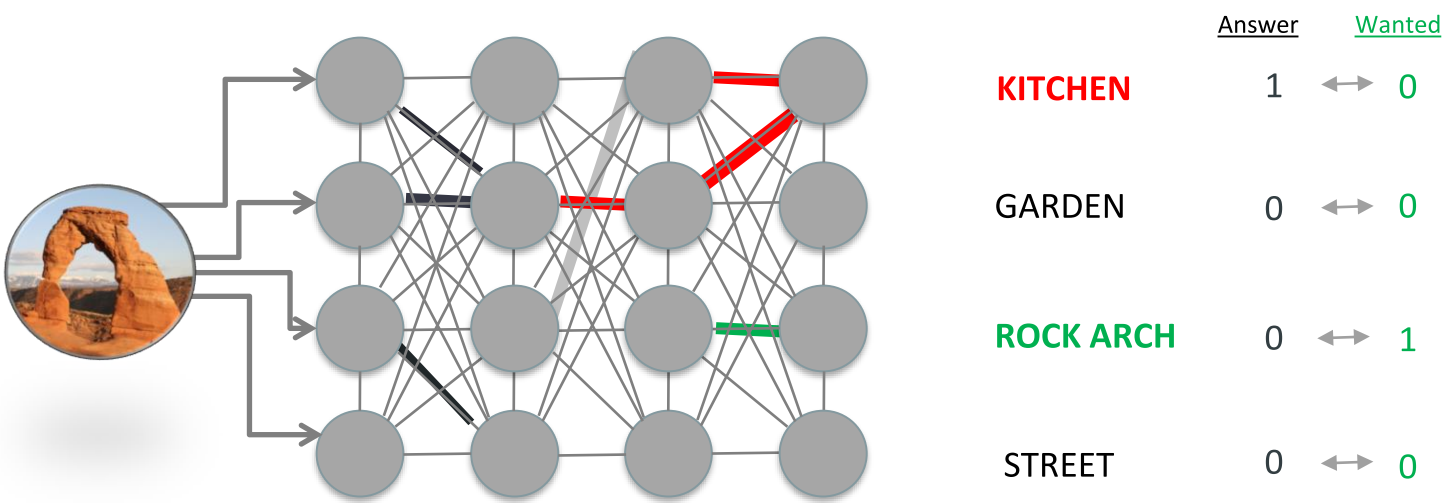


Inspired by Paul Voosen, the AI detectives. Science

Cartoon of a Neural Network

INPUT

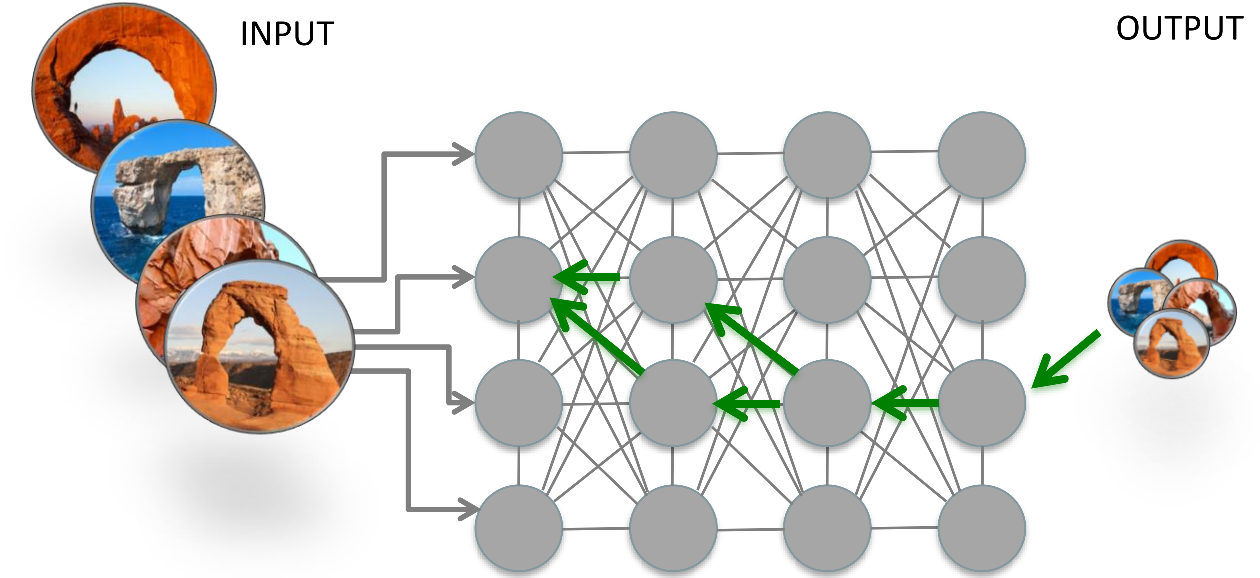
OUTPUT



At first the system makes errors and each result is compare with the ground truth. In a process call **back propagation**, the outcome is sent *backward* through the network, enabling it to reweights the value of each unit.

Cartoon of a Neural Network:

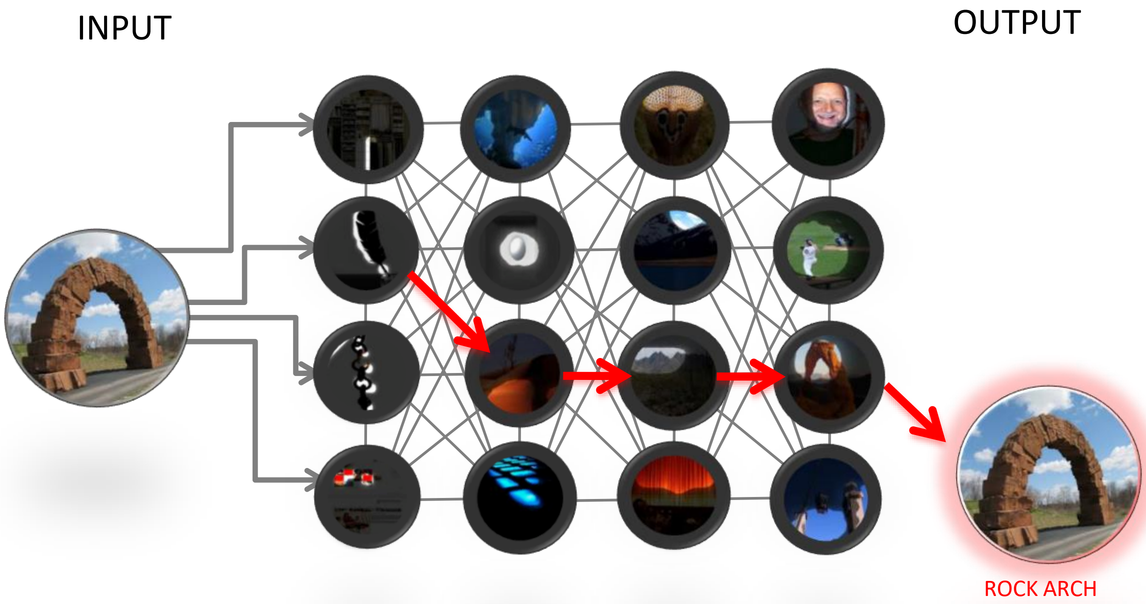
During training, network adjusts its weights



During training, we show many correct pictures of rock arches, and the network adjusts its weights, which become more and more specialized as learning go through many repetitions. This is called ***supervised*** learning

Cartoon of a Neural Network:

At testing, network classifies a new image



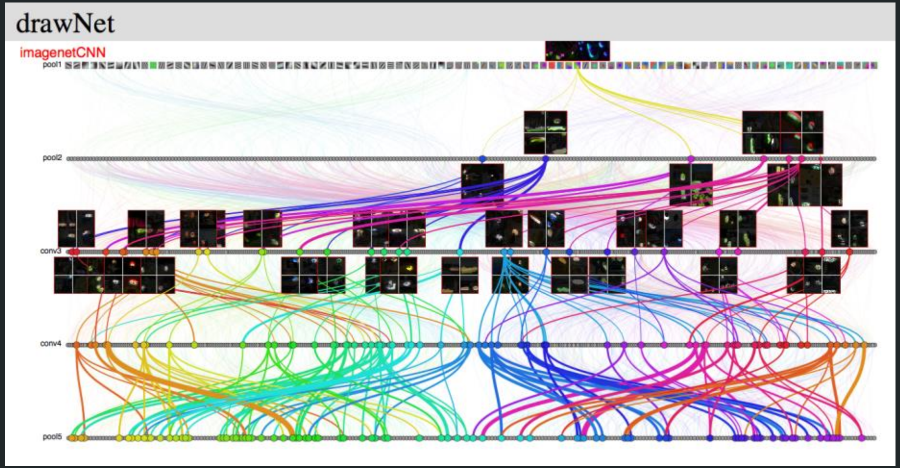
Each layer learns progressively more complex features

Inspired by Paul Voosen, the AI detectives. Science

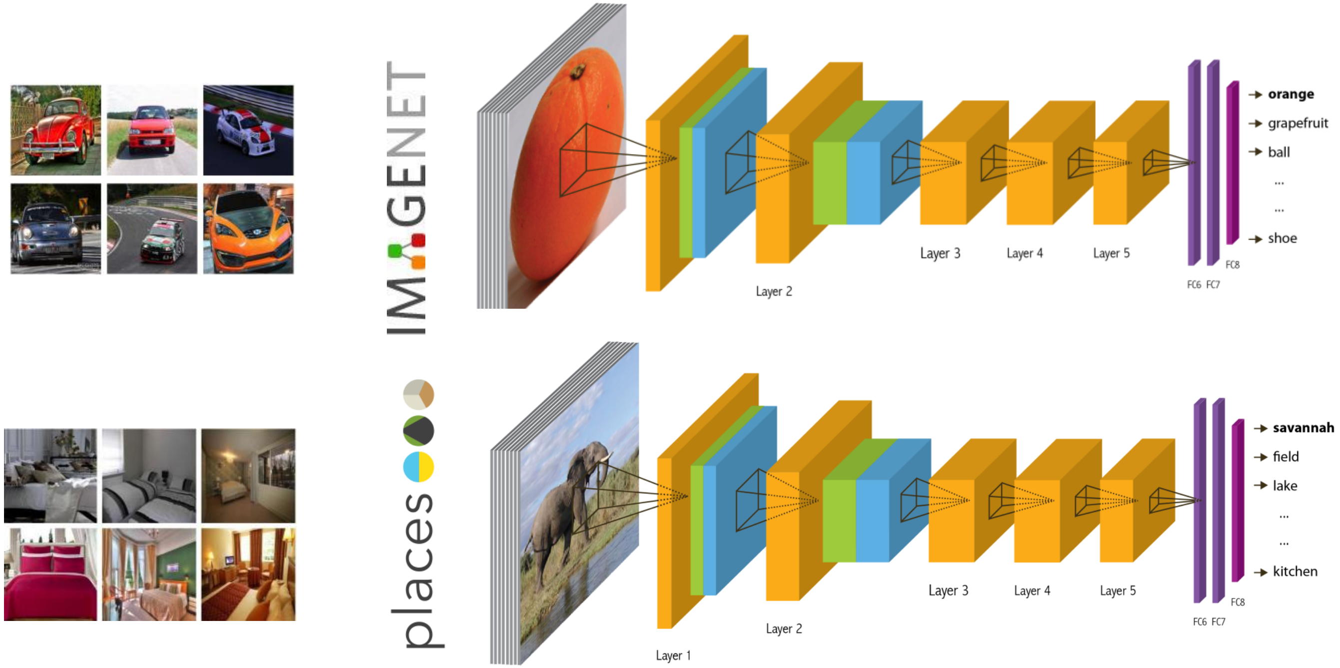
LeCun & Bengio (1995)

Visualizations of CNN units

What did the network learn?



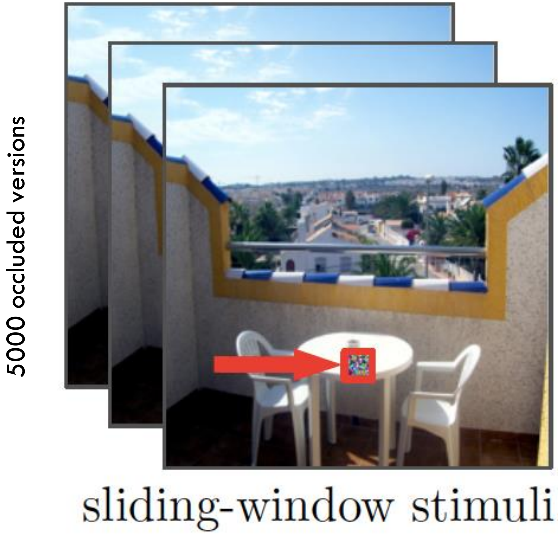
What representations emerge in the networks?



Data driven approach from Neuroscience: **Empirical receptive field** (RF)

Pipeline for estimating the Receptive Fields:

Goal is to identify which regions of the image lead to the high unit activations



Discrepancy map per unit

Each sliding-window stimuli contains a **small randomized patch** at different spatial locations.

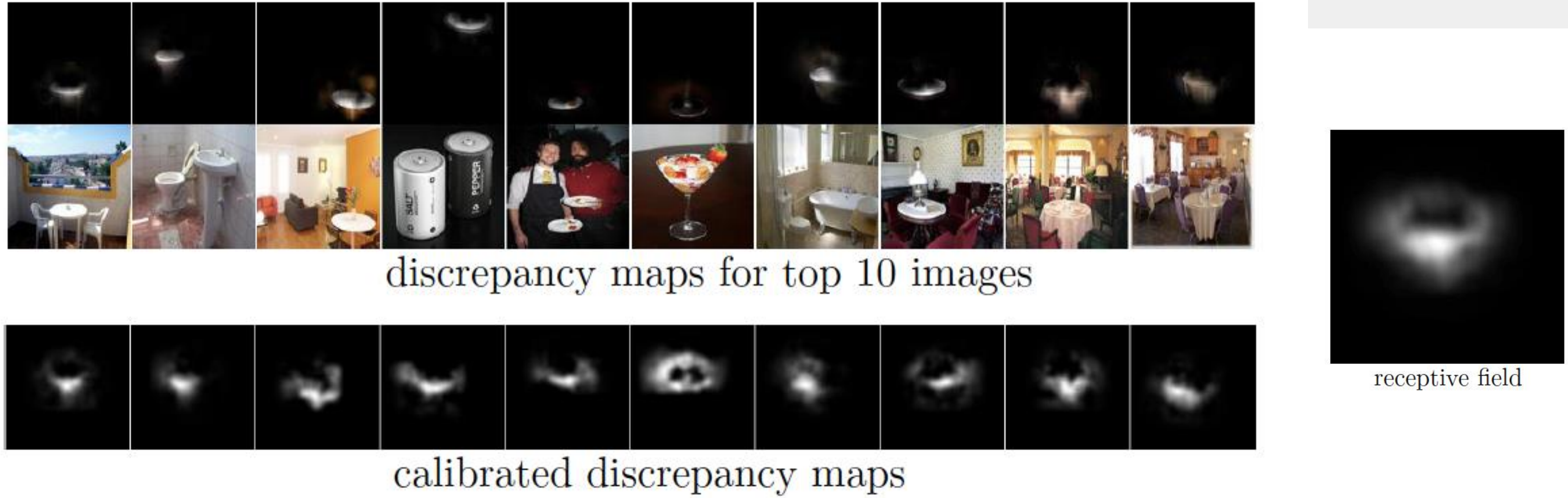
We feed all the occluded images into the same network and record the change in activation as compared to using the original image.

If there is a large **discrepancy**, it indicates that the given patch is important

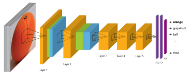
Zhou, Khosla, Lapedriza , Oliva, & Torralba (2015). Object Detectors emerge in Deep Scene CNNs .International Conference on Learning Representations (ICLR) 2015.

Pipeline for estimating the Receptive Fields:

Discrepancy maps to generate an artificial receptive field



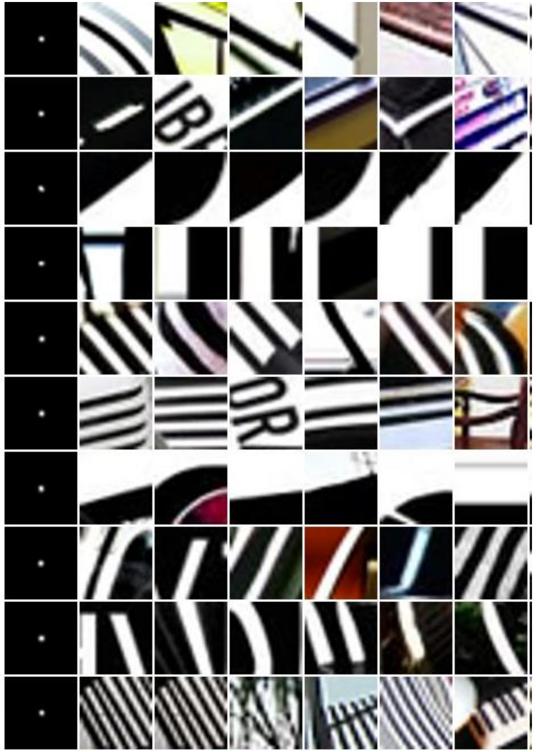
Zhou, Khosla, Lapedriza , Oliva, & Torralba (2015). Object Detectors emerge in Deep Scene CNNs .International Conference on Learning Representations (ICLR) 2015.

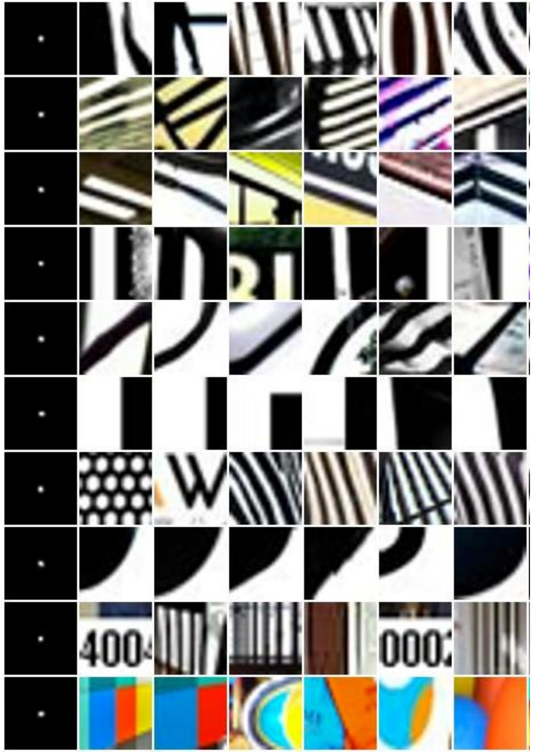


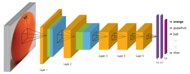
IMAGENET

Layer 1

places



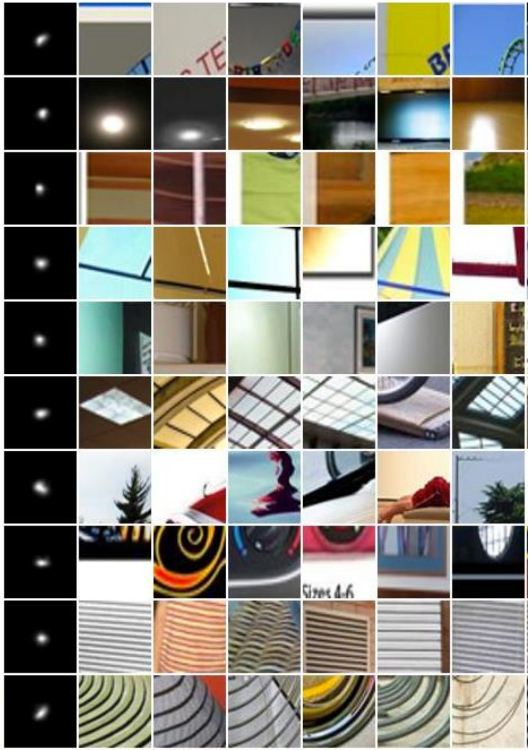




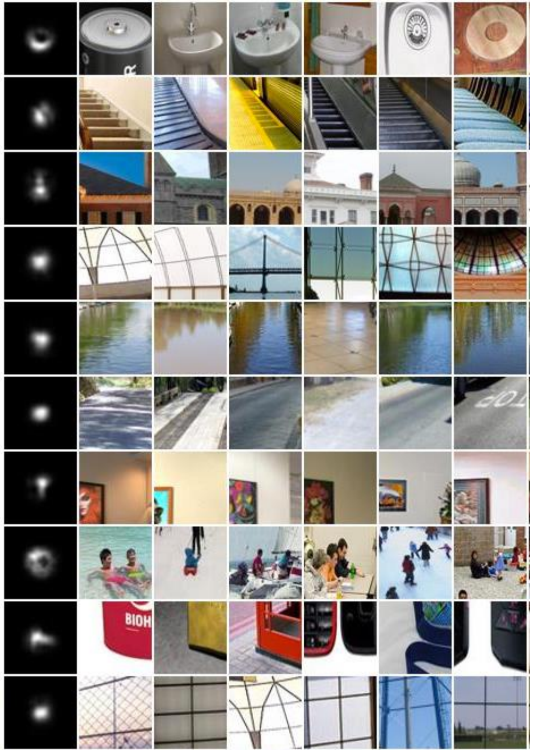
IMAGENET

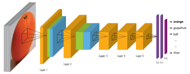
Layer 2

places



Layer 4

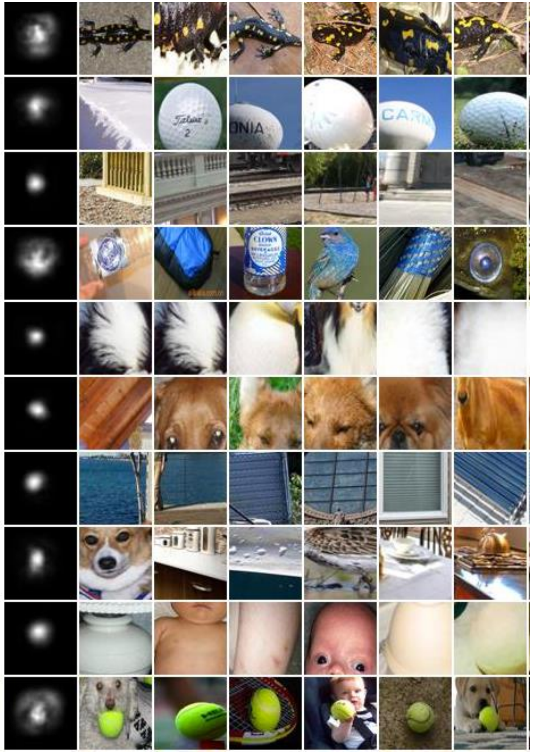


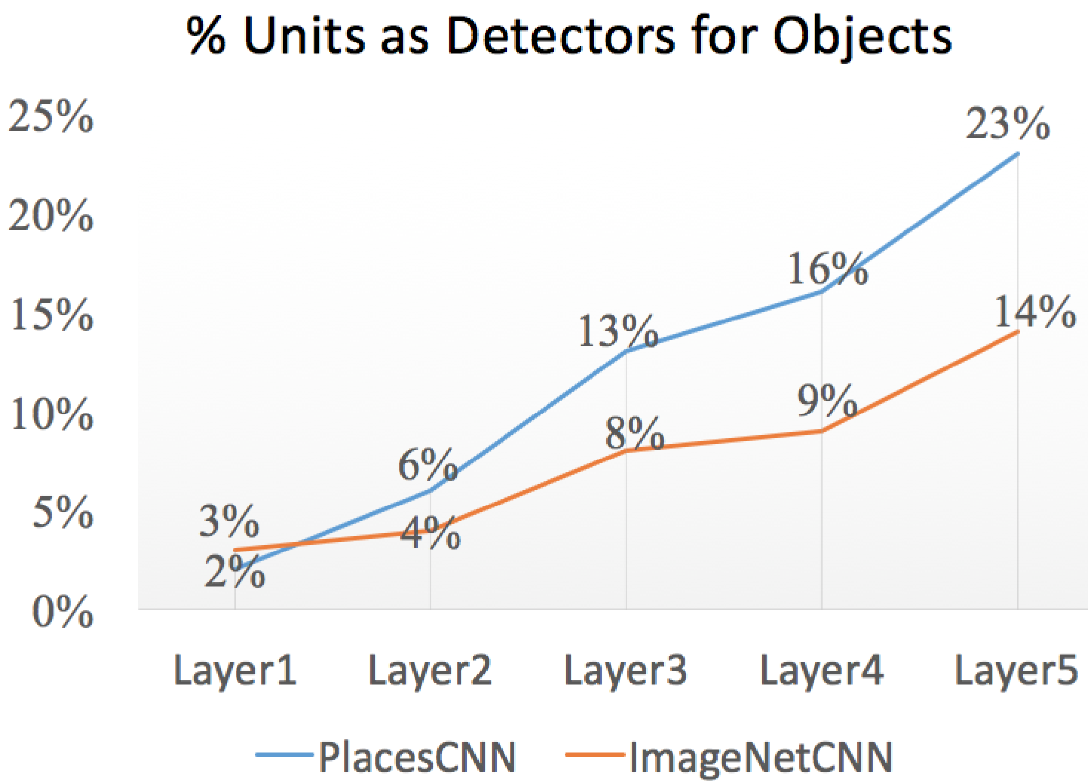


IMAGENET

Layer 5

places





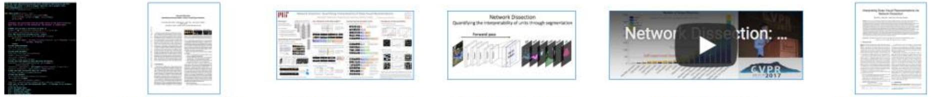
places

Object detectors emerge in a CNN trained to classify categories of places

Network Dissection: Quantifying Interpretability of Deep Visual Representations

David Bau*, Bolei Zhou*, Aditya Khosla, Aude Oliva, Antonio Torralba
Massachusetts Institute of Technology

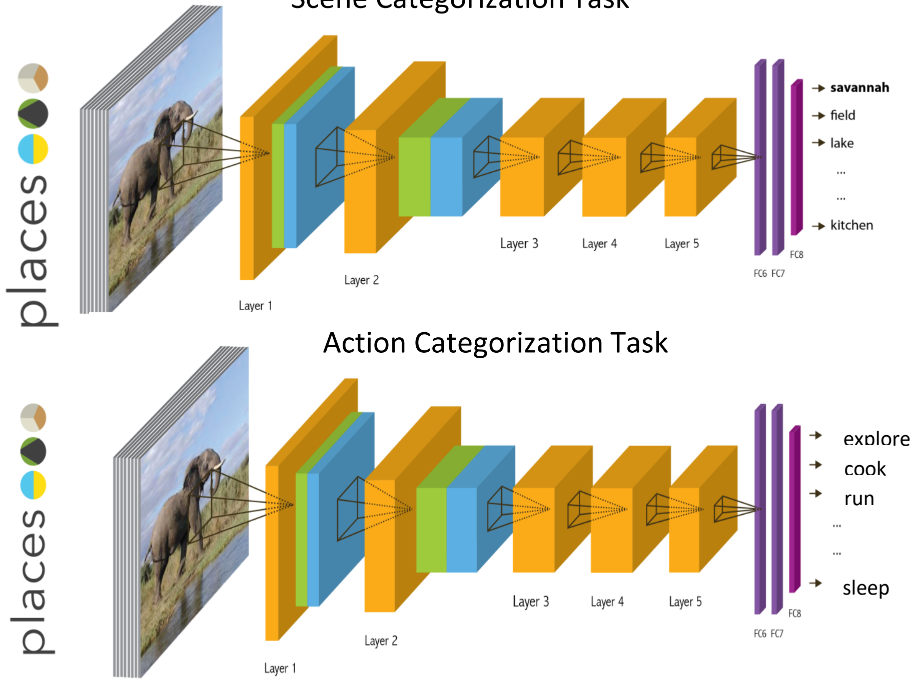
http://netdissect.csail.mit.edu/



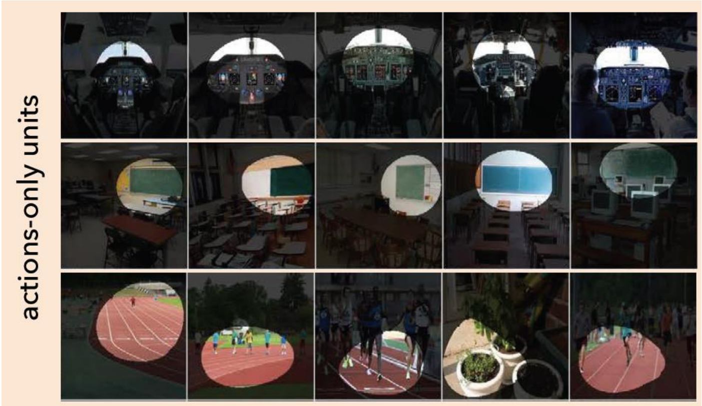
Code and Data | CVPR 2017 paper | CVPR 2017 poster | CVPR 2017 slides | CVPR 2017 oral | Extended paper

Selected units are shown from three state-of-the-art network architectures when trained to classify images of places (places-365). Many individual units respond to specific high-level concepts (object segmentations) that are not directly represented in the training set (scene classifications).

Does a task constraint emergent representations?



Actions-only new detectors emerge

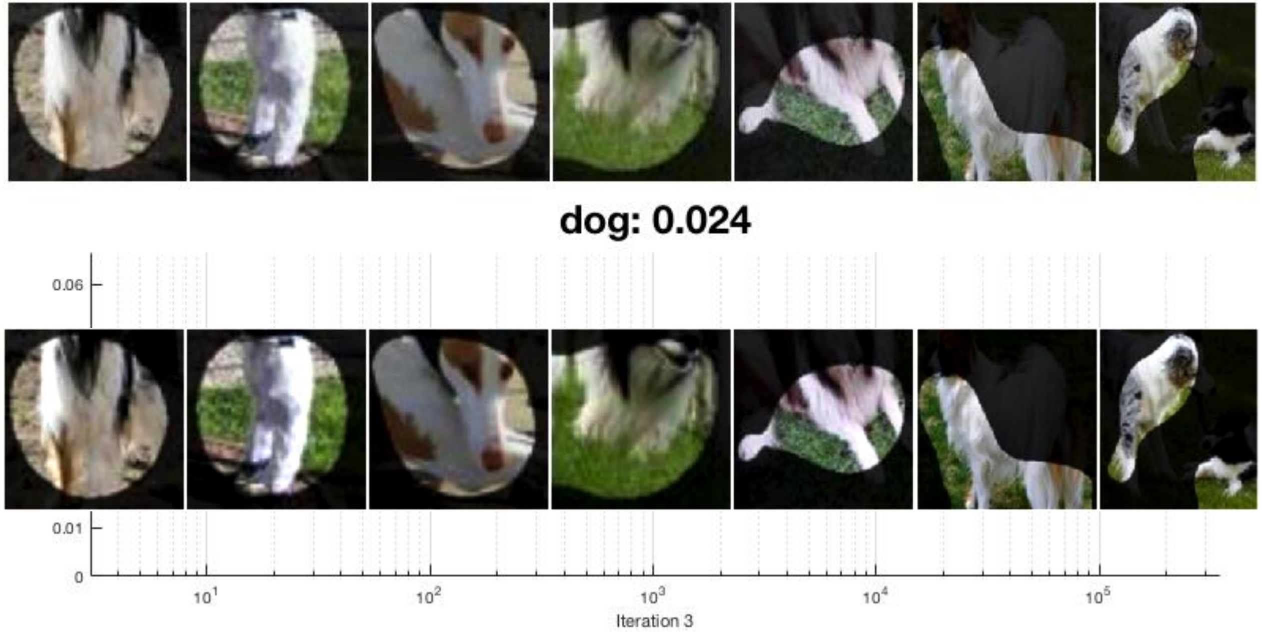


Network Plasticity: What happens to the neurons when they relearn?

Unit 8 at Layer 5 layer

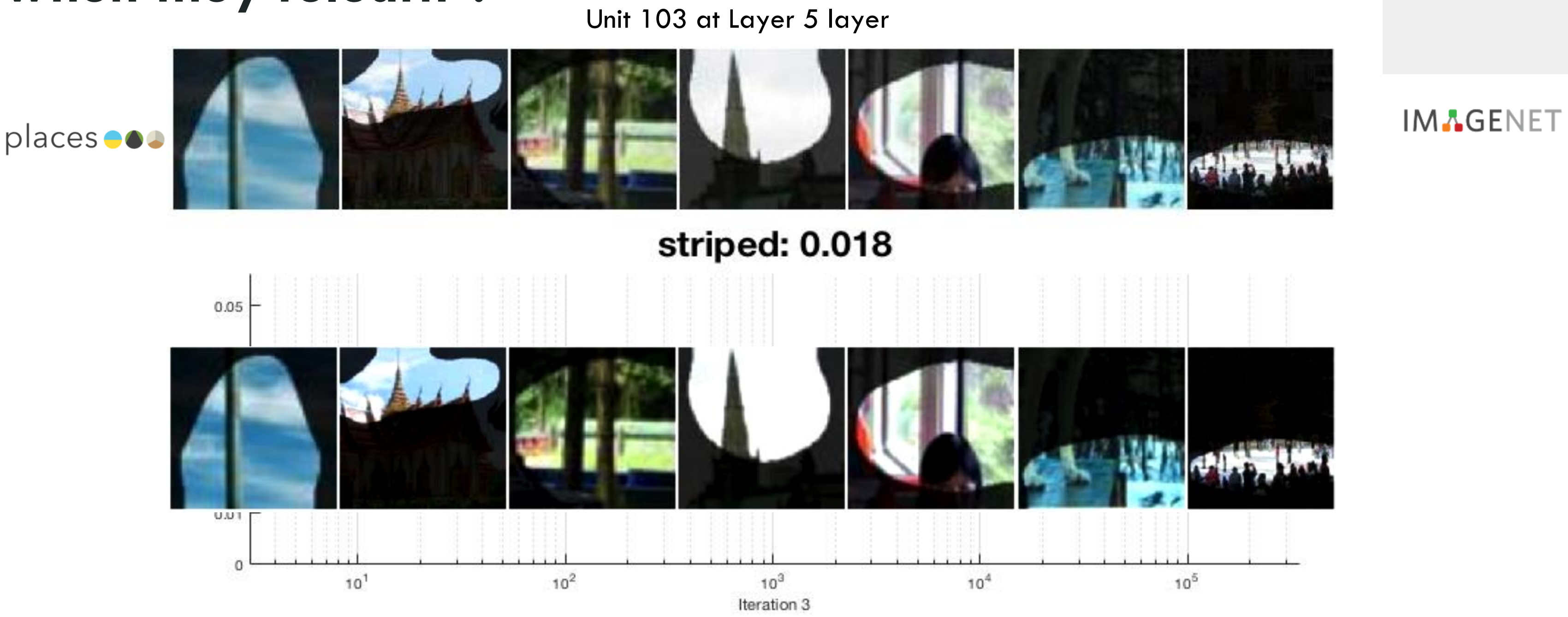
IMAGENET

places



Zhou, Bau, Oliva, & Torralba (2019). Interpreting Deep Visual Representations via Network Dissection. *IEEE T. PAMI*

Network Plasticity: What happens to the neurons when they relearn ?



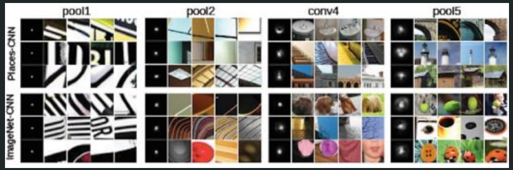
Zhou, Bau, Oliva, & Torralba (2019). Interpreting Deep Visual Representations via Network Dissection. *IEEE T. PAMI*

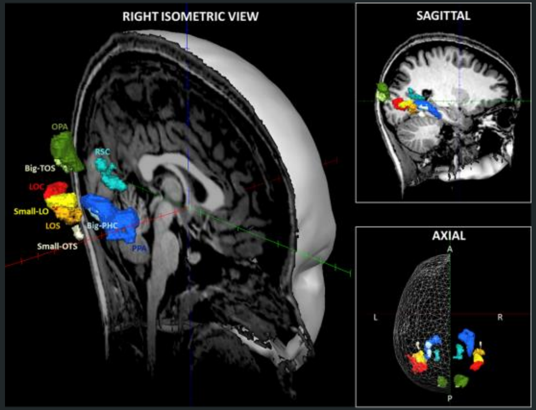
Comparison Natural (Brain) vs. Artificial Neural Networks

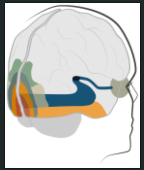
Visual object recognition

Oliva, A. (2020). Computational Models of Human Object and Scene Recognition. In *The Cognitive Neurosciences, 6th Edition.* Edited by Gazzaniga et al. MIT Press (pp. 151-157).

Hasson, U., Nastase, S.A. & Goldstein, A. (2020). Direct Fit to Nature: An Evolutionary Perspective on Biological and Artificial Neural Networks. *Neuron*, 105, 3, 416-434.

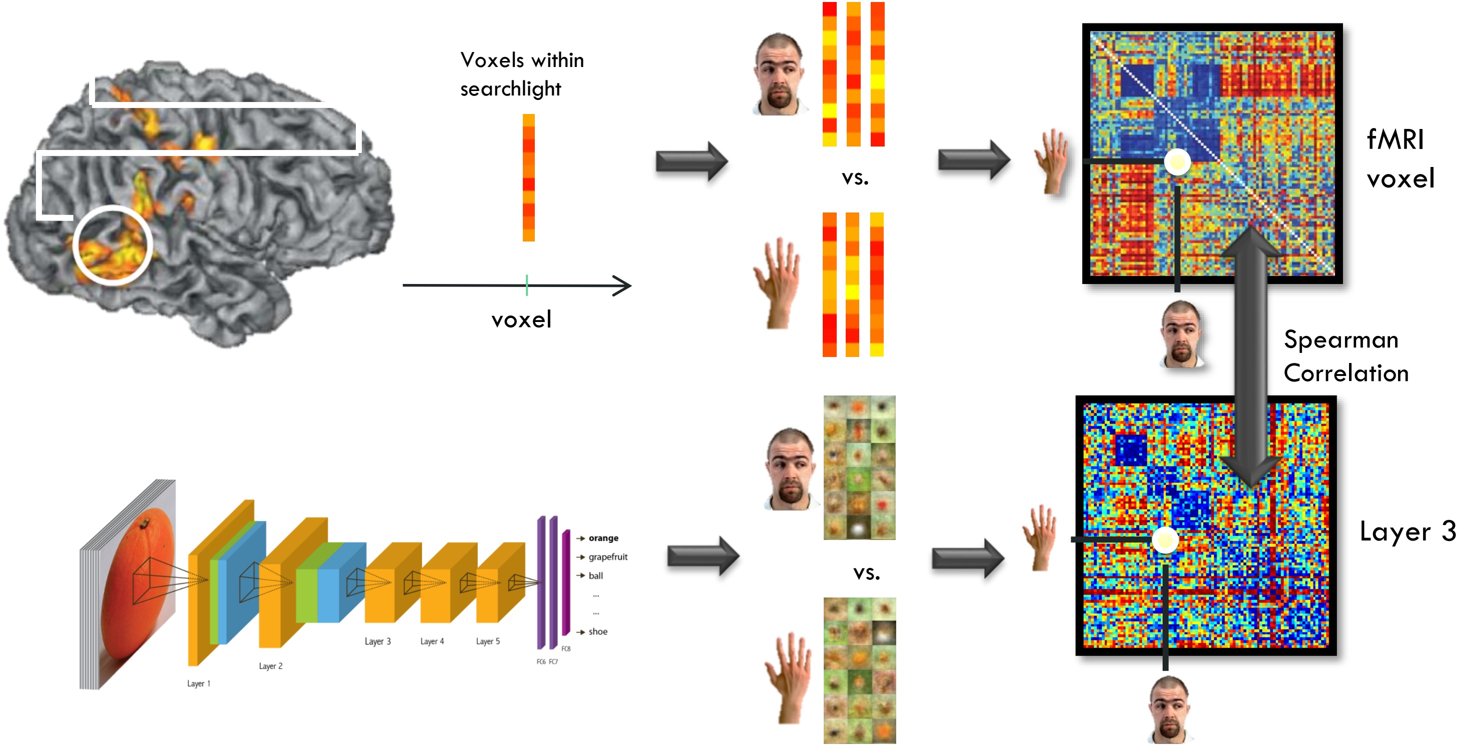






Cichy, Khosla, Pantazis, Torralba & Oliva, A. (2016). Scientific Reports.

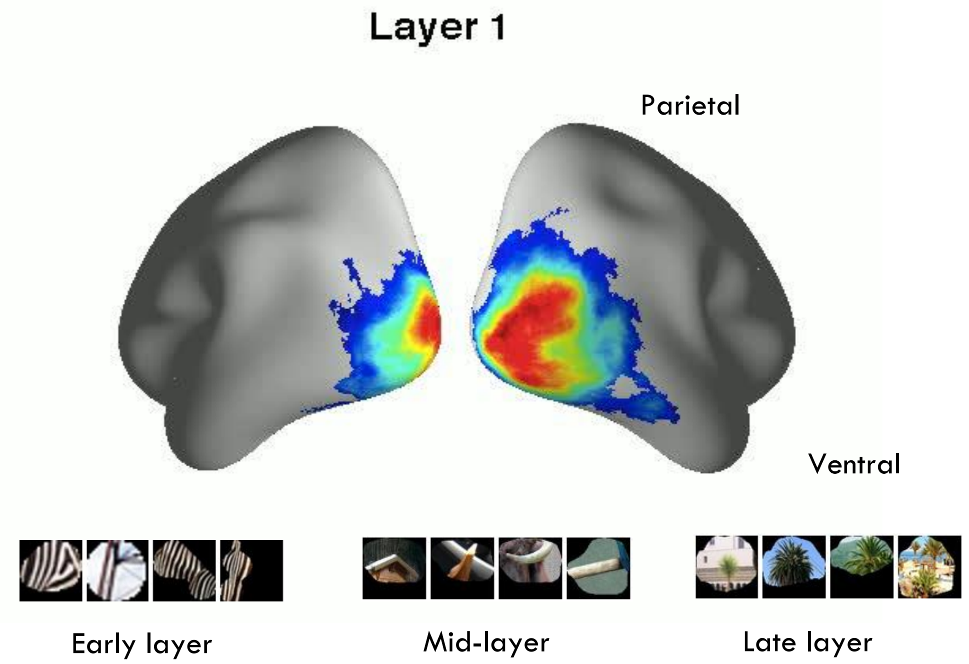
Algorithmic-specific fMRI searchlight analysis

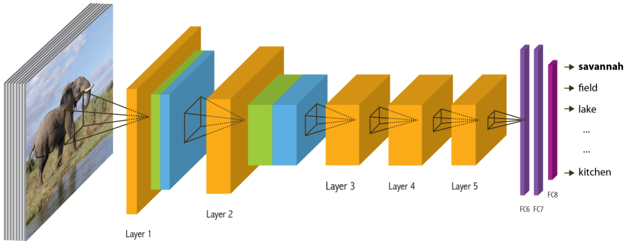


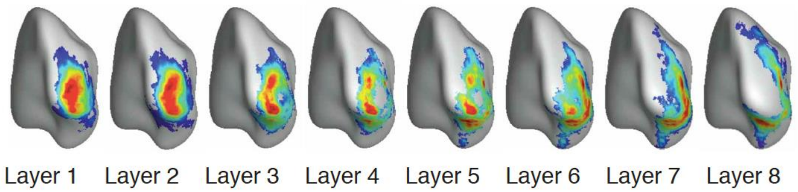
A spatially unbiased view of the relations in similarity structure (see **Kriegeskorte** RSA) between human neuro-imaging data (i.e. fMRI) and Models

Cichy, Khosla, Pantazis, Torralba & Oliva, A. (2016). Scientific Reports.

Spatiotemporal maps of correlations between human brain and model layers







Cichy, Khosla, Pantazis, Torralba, & Oliva. (2016). Scientific Reports.

Biological and Artificial Neural Networks

- **Studying the implementation** that works best for performing specific tasks
- **Characterizing the network behavior** when it is adapting, compromised or enhanced
- **Exploring the alternatives** that have not been taken by biological systems
- **Taking inspiration** from biological neural networks feats (i.e. keep learning without forgetting, the right pruning)
- **A new field of expertise:** Cognitive / Clinical / Social / Perceptual Computational Experimentalist / Synthetic Neuroscientist