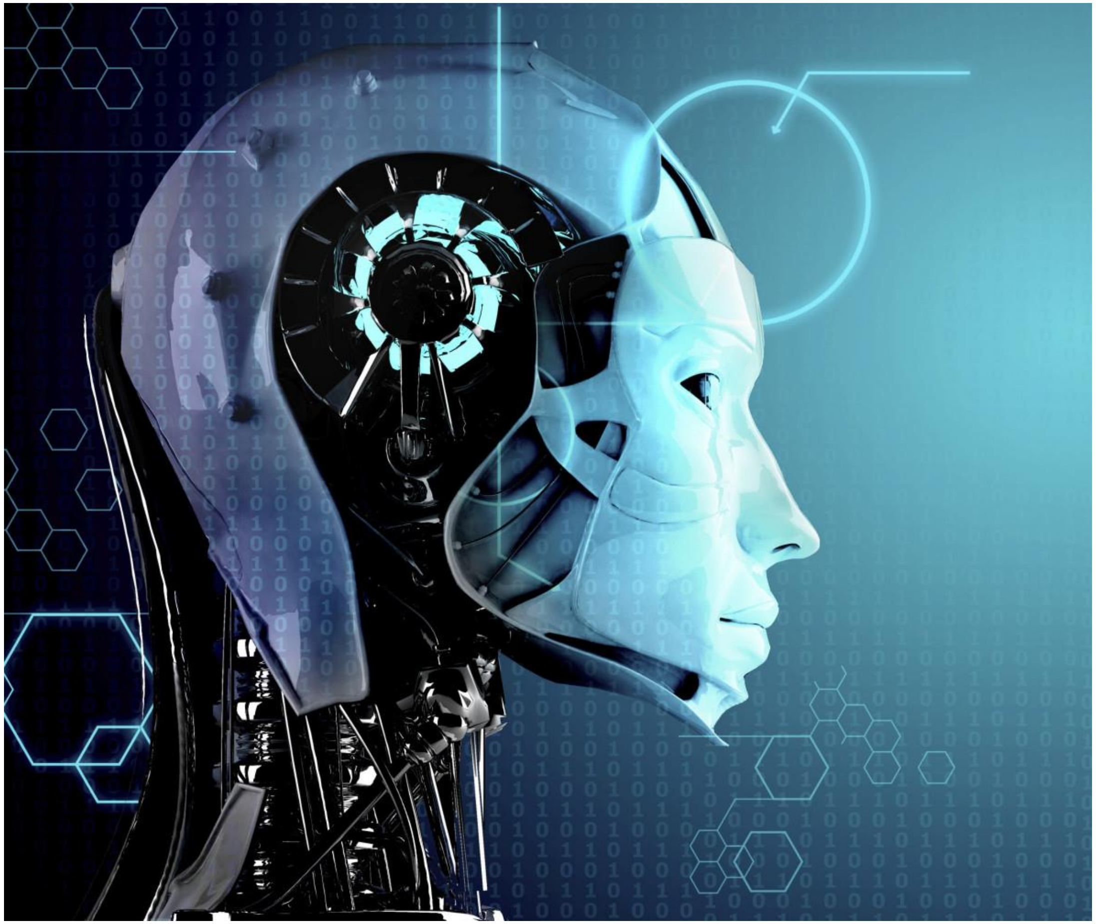
Artificial Intelligence
with connections to neuroimaging

Alona Fyshe
afyshe@uvic.ca
http://web.uvic.ca/~afyshe/
http://brainyblog.net

Slides are available at:

http://web.uvic.ca/~afyshe/talks/ccn_AI.pdf

What is AI?

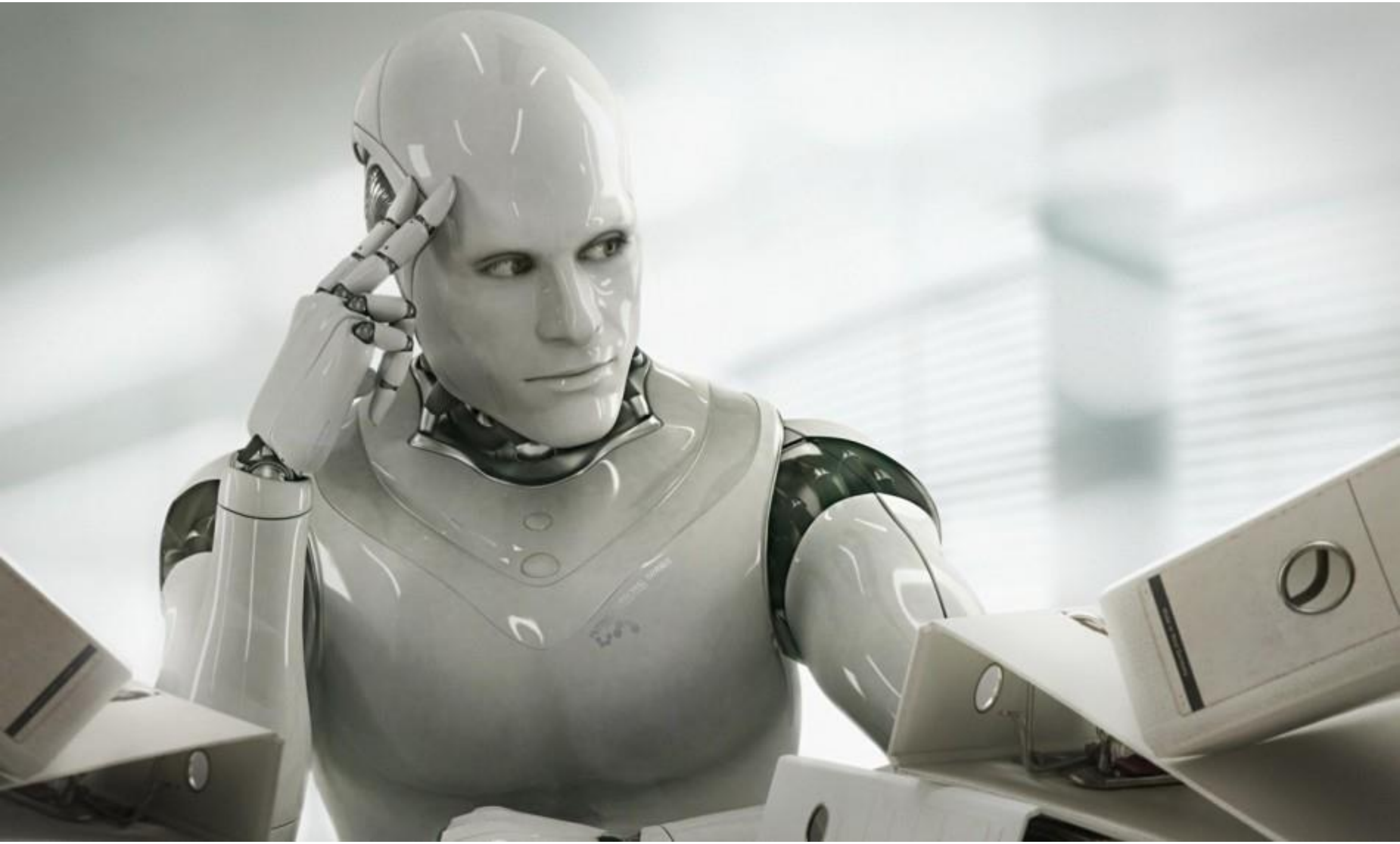


How would we know if we had achieved AI?

Robot Student Test (Goertzel, 2012)

- If a computer could…
  - Register in a university program
  - Attend classes
  - Read the textbook
  - Successfully complete tests/assignments
  - Finish a degree

… that would be artificial intelligence



AI2's Aristo: 8th Grade Science Tests

- Allen Institute for AI
  - Training an AI to "read" science textbooks and the internet
  - Take 8th grade science tests

http://allenai.org/aristo/

AI2's Aristo:
8th Grade Science Tests

YOU ASKED ARISTO:

Which characteristic helps a fox find food?

(A) sense of smell
(B) thick fur
(C) long tail
(D) pointed teeth

http://allenai.org/aristo/

AI2's Aristo:
8th Grade Science Tests

ARISTO ANSWERED:

Question: Which characteristic helps a fox find food?

Answer: (A) sense of smell

Because: A wolf depends on its sense of smell to find most food .

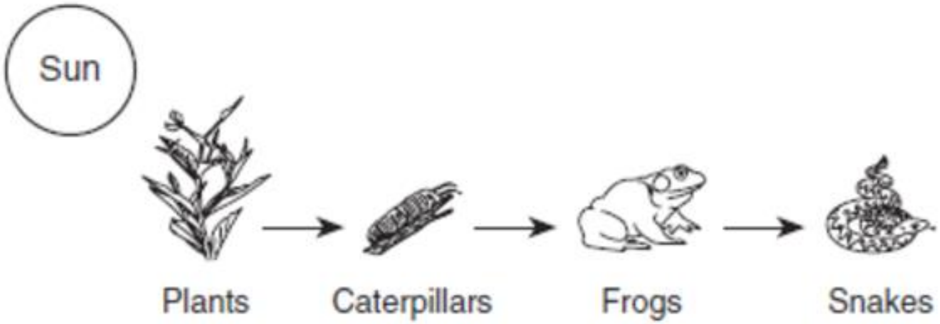
Confidence: 95.00%

Source: Web

http://allenai.org/aristo/

AI2's Aristo:
8th Grade Science Tests

Base your answers on the diagram of a food chain below and on your knowledge of science.



(Not drawn to scale)

If the population of snakes increases, the population of frogs will most likely

(A) decrease
(B) increase
(C) remain the same

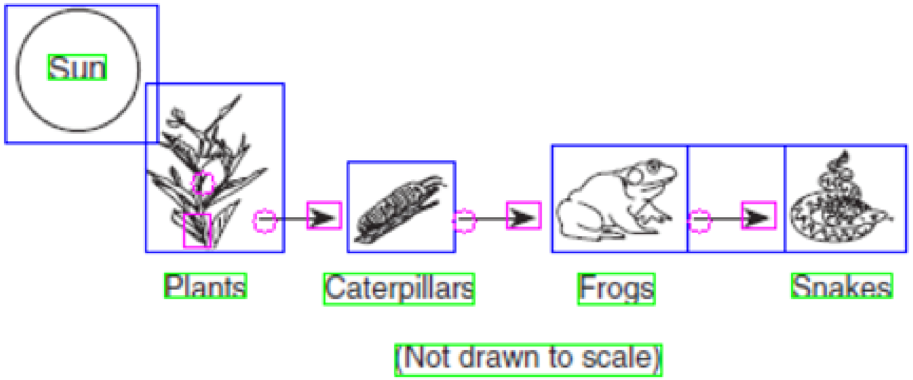
http://allenai.org/aristo/

AI2's Aristo: 8th Grade Science Tests

Answer: (A) decrease

Because:

Diagram elements detected:



Extracted food chain: **dependsOn**(Caterpillars, Plants) & **dependsOn**(Snakes, Frogs) & **dependsOn**(Frogs, Caterpillars)

Given: **qChange**(Snakes, Increase)

Query: **qChange**(Frogs, X)

Answer: **X** = Decrease

http://allenai.org/aristo/

Robot Student Test (Goertzel)

- What would that require?
  - perception
    - e.g. read figures in books
  - represent knowledge
    - e.g. foxes are like wolves
  - communication via language (written at least)
    - e.g. writing essays for class
  - learn, reason, generalize, plan…

Robot Student Test (Goertzel)

- What would that require?
  - **perception**
    - e.g. read figures in books
  - **represent knowledge**
    - e.g. foxes are like wolves
  - **communication via language (written at least)**
    - e.g. writing essays for class
  - learn, reason, generalize, plan…

This talk is not a complete overview

- There are many, many areas of AI I am not covering today
  - Reinforcement Learning
  - All of robotics (navigation, planning, interaction with the physical world…)
  - …

The Singularity

- I will **not** talk about AI taking over the world

- Instead, I **will** talk about
  – how AI **currently** impacts society
  – how biased models sneak their way into our lives

The Robot Student Test

SKILL 1: PERCEPTION

Perception in AI

- Amazing advances in computer vision
  - ImageNet
    - First released 2009
    - 1.2 million images
    - more than 1000 concepts
      - e.g. cup, oil filter, ptarmigan
  - Deep learning
    - CNNs

http://image-net.org/

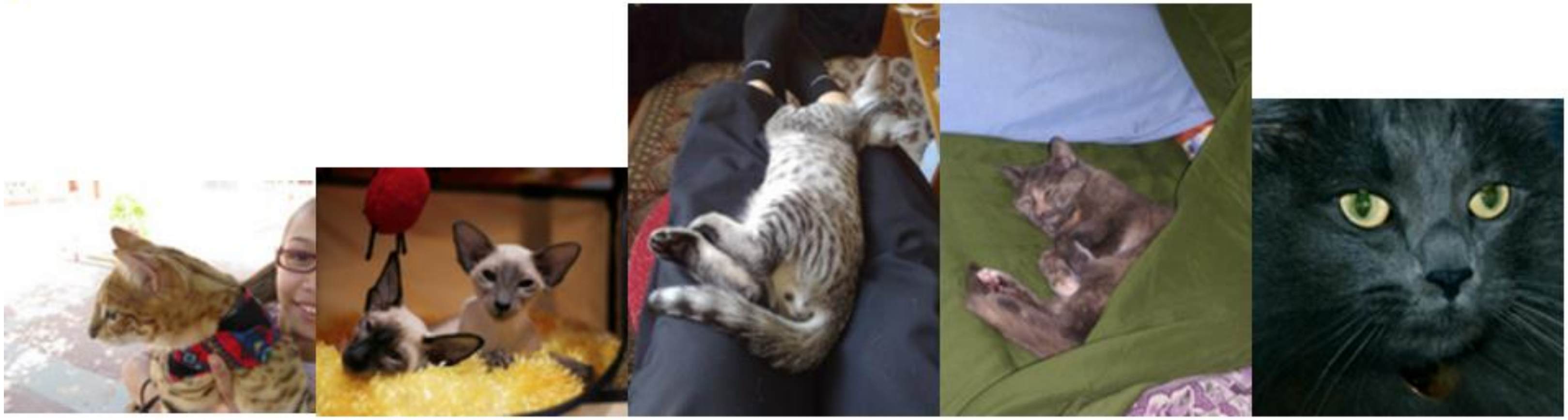
ImageNet

http://cs.stanford.edu/people/karpathy/cnnembed/

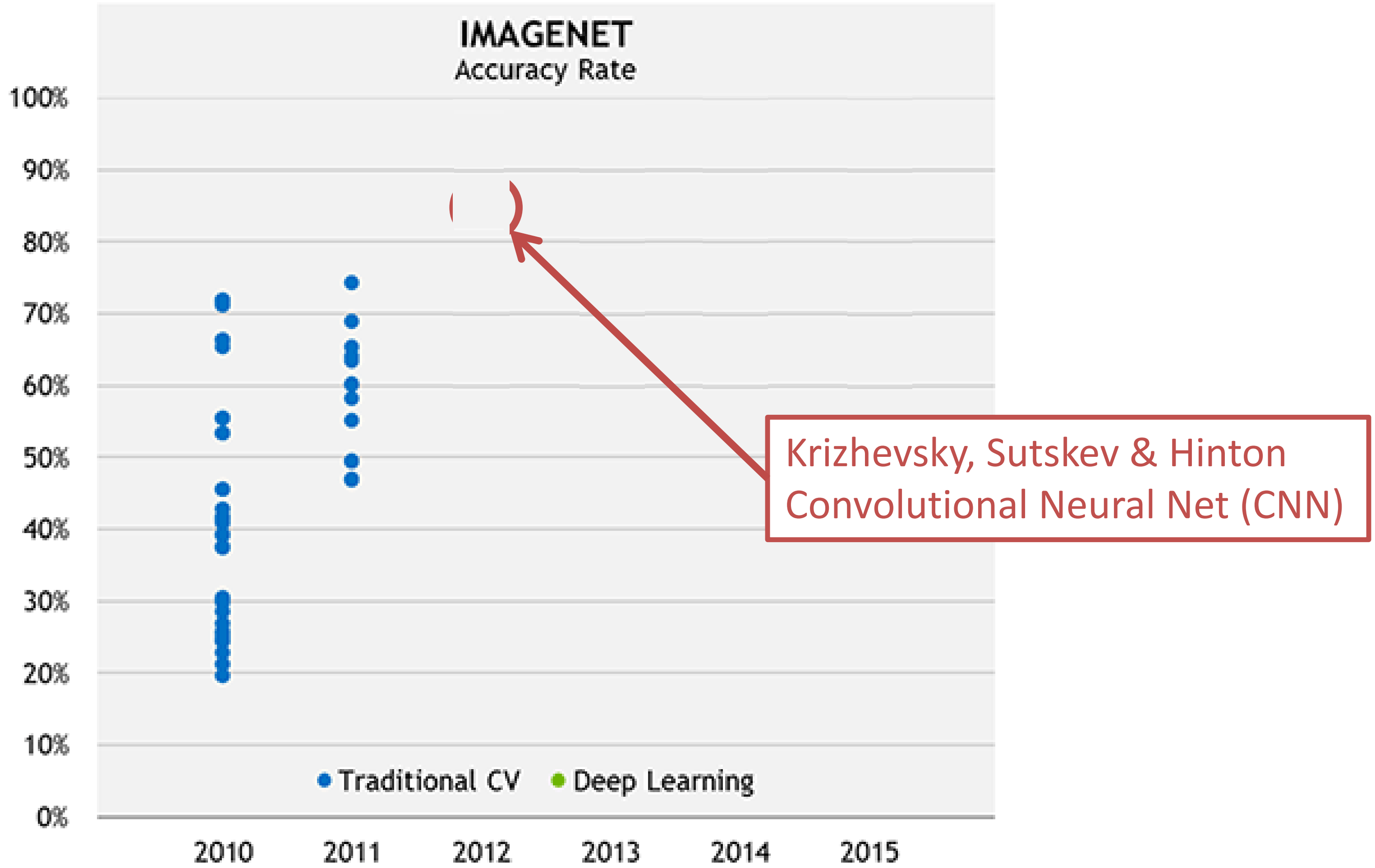
ImageNet

french fries mashed potato black olive face powder crab apple Granny Smith strawberry blueberry cranberry currant blackberry raspberry persimmon mulberry orange kumquat lemon grapefruit plum fig pineapple banana jackfruit cherry grape custard apple durian mango elderberry guava litchi pomegranate quince kidney bean soy green pea chickpea chard lettuce cress spinach bell pepper pimento jalapeno cherry tomato parsnip turnip mustard bok choy head cabbage broccoli cauliflower brussels sprouts zucchini spaghetti squash acorn squash butternut squash cucumber artichoke asparagus green onion shallot leek cardoon celery mushroom pumpkin cliff lunar crater valley alp volcano promontory sandbar dune coral reef lakeside seashore geyser bakery juniper berry gourd acorn olive hip ear pumpkin seed sunflower seed coffee bean rapeseed corn buckeye bean peanut walnut cashew chestnut hazelnut coconut pecan pistachio lentil pea peanut okra sunflower lesser celandine wood anemone blue columbine delphinium nigella calla lily sandwort pink baby's breath ice plant globe amaranth four o'clock Virginia spring beauty wallflower damask violet candytuft Iceland poppy prickly poppy oriental poppy celandine blue poppy Welsh poppy celandine poppy corydalis pearly everlasting strawflower yellow chamomile dusty miller tansy daisy common marigold China aster cornflower chrysanthemum mistflower cosmos dahlia coneflower blue daisy gazania African daisy male orchis butterfly orchid aerides brassavola spider orchid grass pink calypso cattleya red helleborine coelogyne cymbid lady's slipper marsh orchid dendrobium disa helleborine fragrant orchid fringed orchis lizard orchid laelia masdevallia odontoglossum oncidium bee orchid fly orchid spider orchid phaius moth orchid ladies' tresses stanhopea stelis vanda cyclamen centaury gentian begonia commelina scabious achimenes African violet streptocarpus scorpionweed calceolaria toadflax veronica bonsai star anise wattle huisache silk tree rain tree dita pandanus linden American beech New Zealand beech live oak shingle oak pin oak cork oak yellow birch American white birch downy birch alder fringe tree European ash fig witch elm Dutch elm cabbage tree golden shower tree honey locust Kentucky coffee tree Brazilian rosewood logwood coral tree Japanese pagoda tree kowhai palm Arabian coffee cork tree weeping willow pussy willow goat willow China tree pepper tree balata teak ginkgo pine ilang-ilang laurel magnolia tulip tree baobab kapok red beech cacao sorrel tree iron tree mangrove paper mulberry Judas tree redbud mountain ash ailanthus silver maple Oregon maple sycamore box elder Japanese maple holly dogwood truffle shiitake lichen hen-of-the-woods jelly fungus dead-man's-fingers earthstar coral fungus stinkhorn puffball gyromitra bolete polypore gill fungus morel agaric trilobite harvestman scorpion black and gold garden spider barn spider garden spider black widow tarantula wolf spider tick mite centipede millipede horseshoe crab

http://image-net.org/challenges/LSVRC/2010/browse-synsets

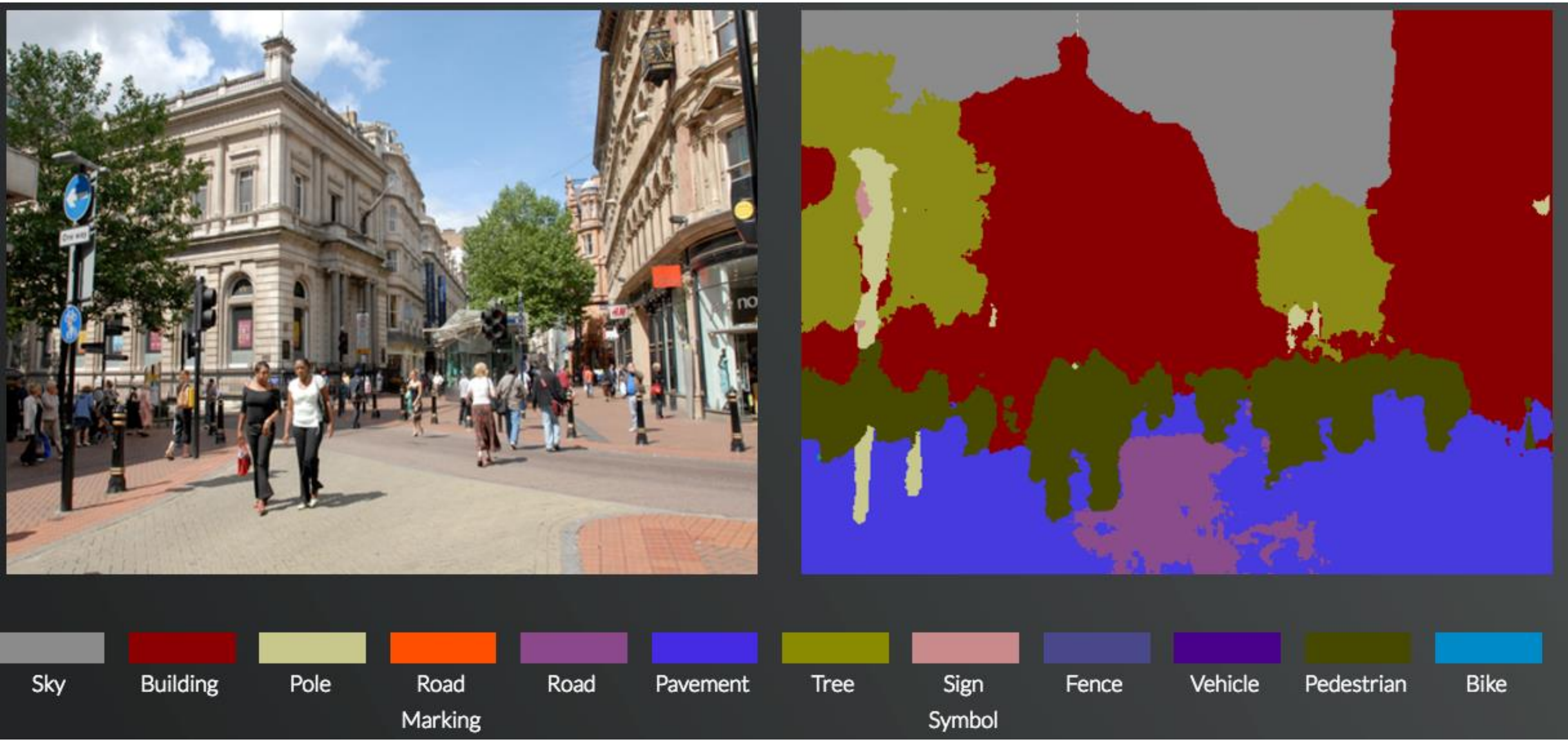


http://image-net.org/challenges/LSVRC/2012/analysis/



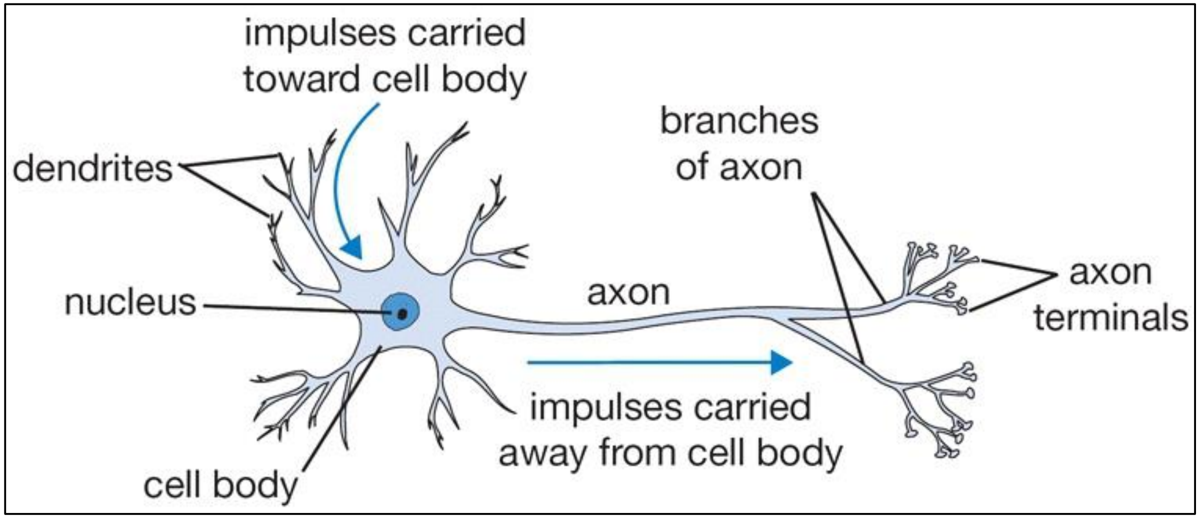
https://blogs.nvidia.com/blog/2016/01/12/accelerating-ai-artificial-intelligence-gpus/

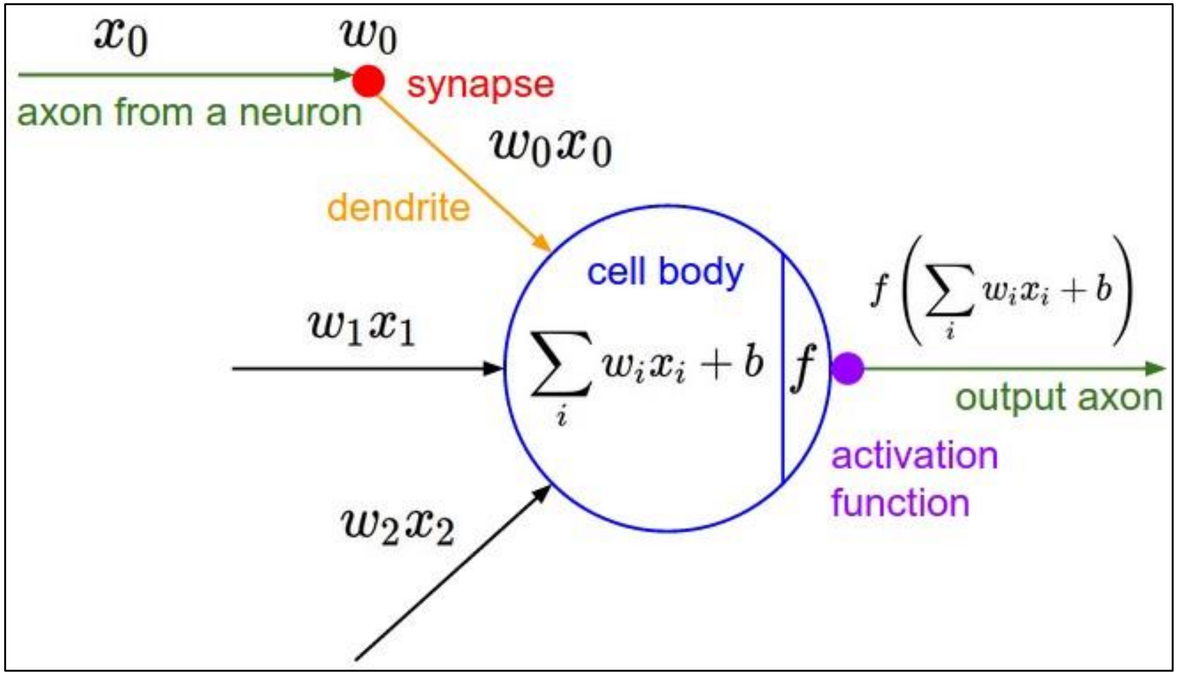
Self Driving Cars



Self Driving Cars

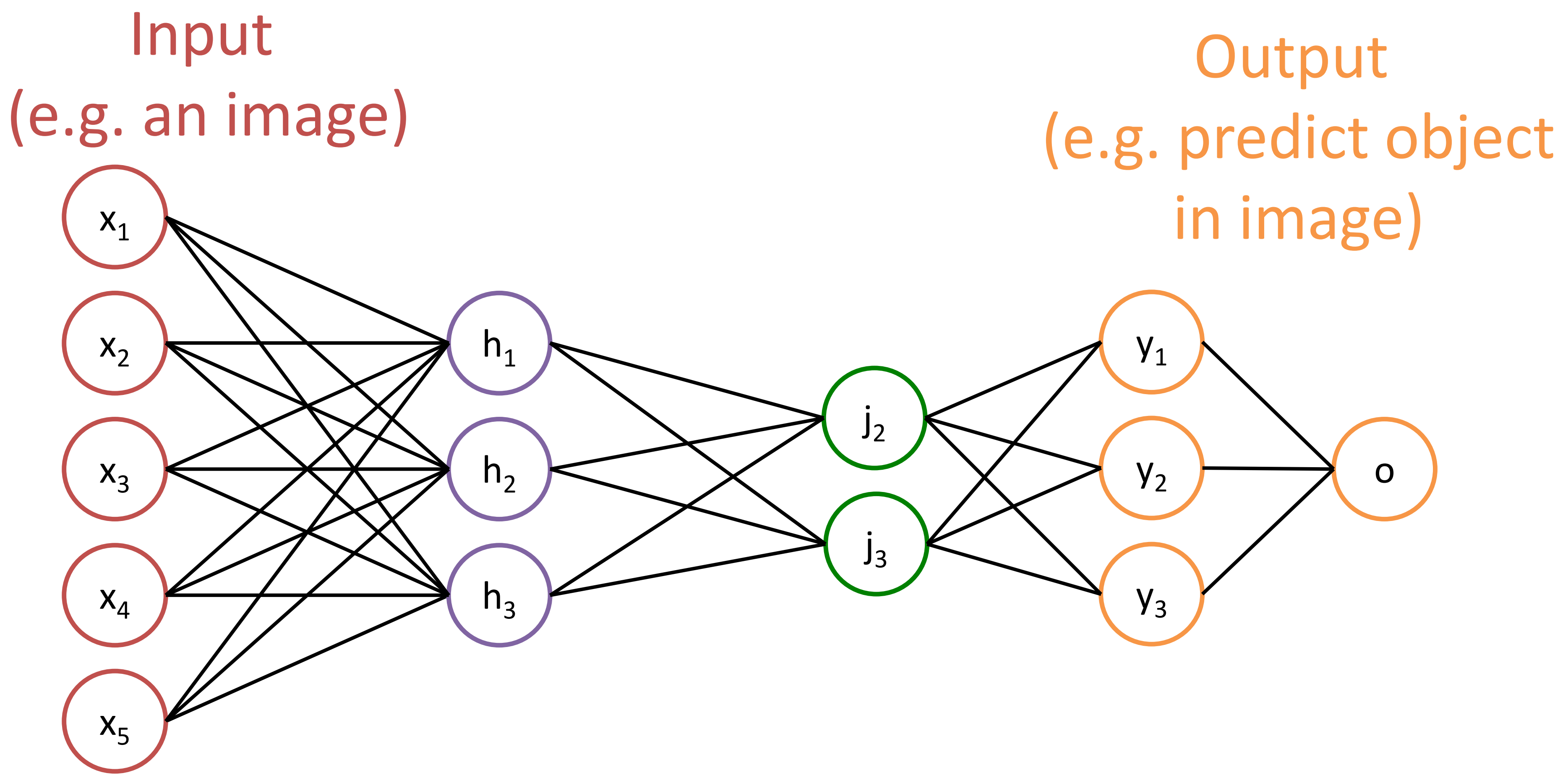
Neural Nets: Crash course



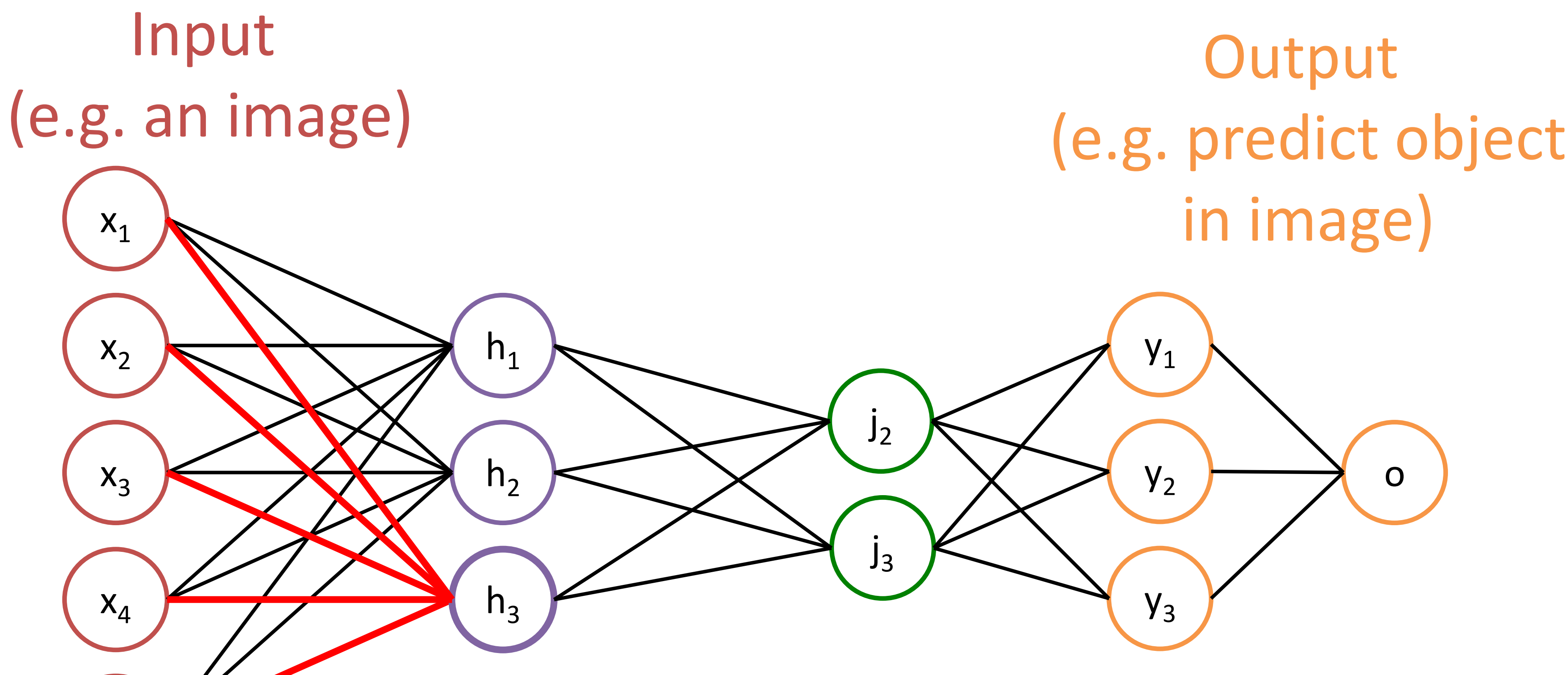


http://cs231n.github.io/neural-networks-1/

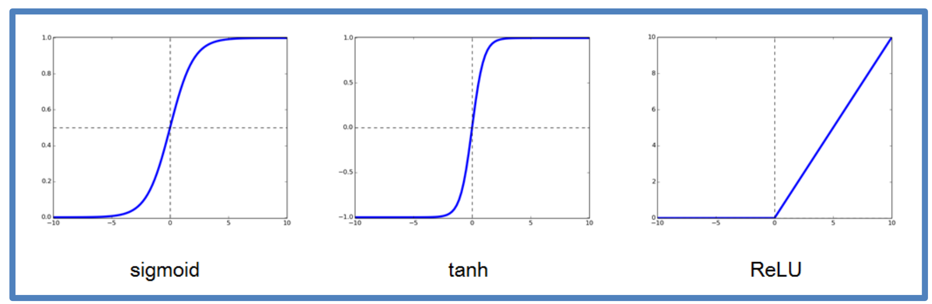
Neural Nets: Crash course



Neural Nets: Crash course



$$h_3 = \sigma\left(\sum_{i=1}^{5} w_{h_3,i} * x_i\right)$$



Neural Nets: Crash course

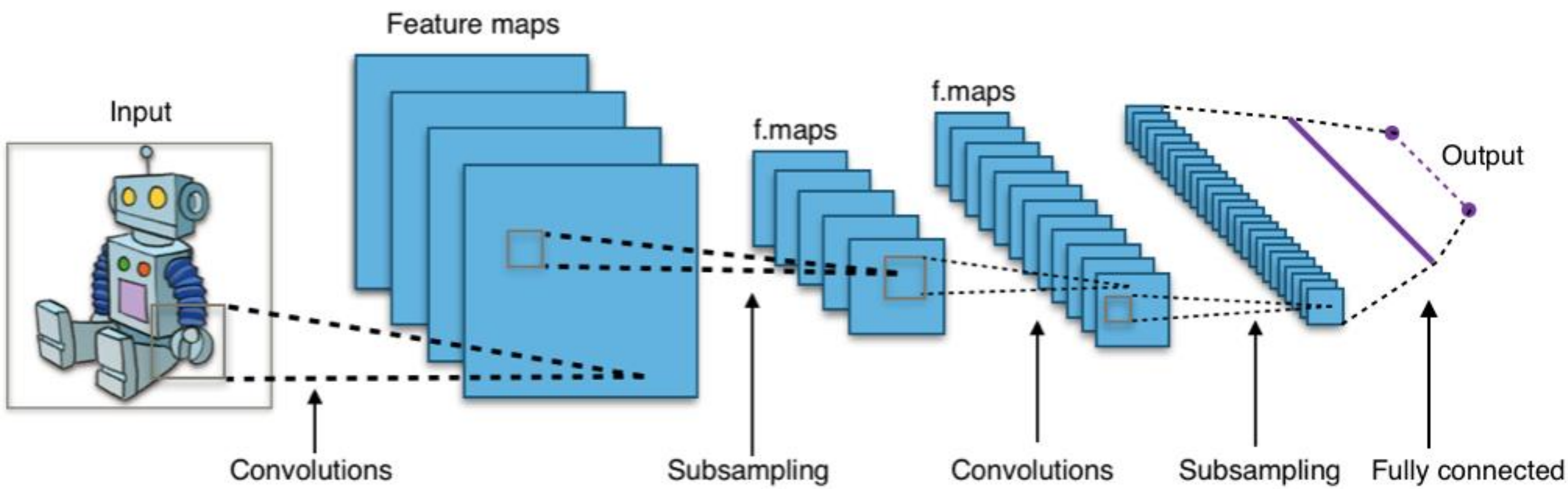
- Architecture fixed (e.g. depth, # neurons)
- Weights at each edge are learned
  - Backpropagation (stochastic gradient descent)

Convolutional Neural Nets

Two additional operations:

- Convolution
- Pooling/Subsampling

Convolutional Neural Nets (CNNS)



https://upload.wikimedia.org/wikipedia/commons/6/63/Typical_cnn.png

CNNs: Convolution

1	1	1	0	0
0	1	1	1	0
0	0	1	1	1
0	0	1	1	0
0	1	1	0	0

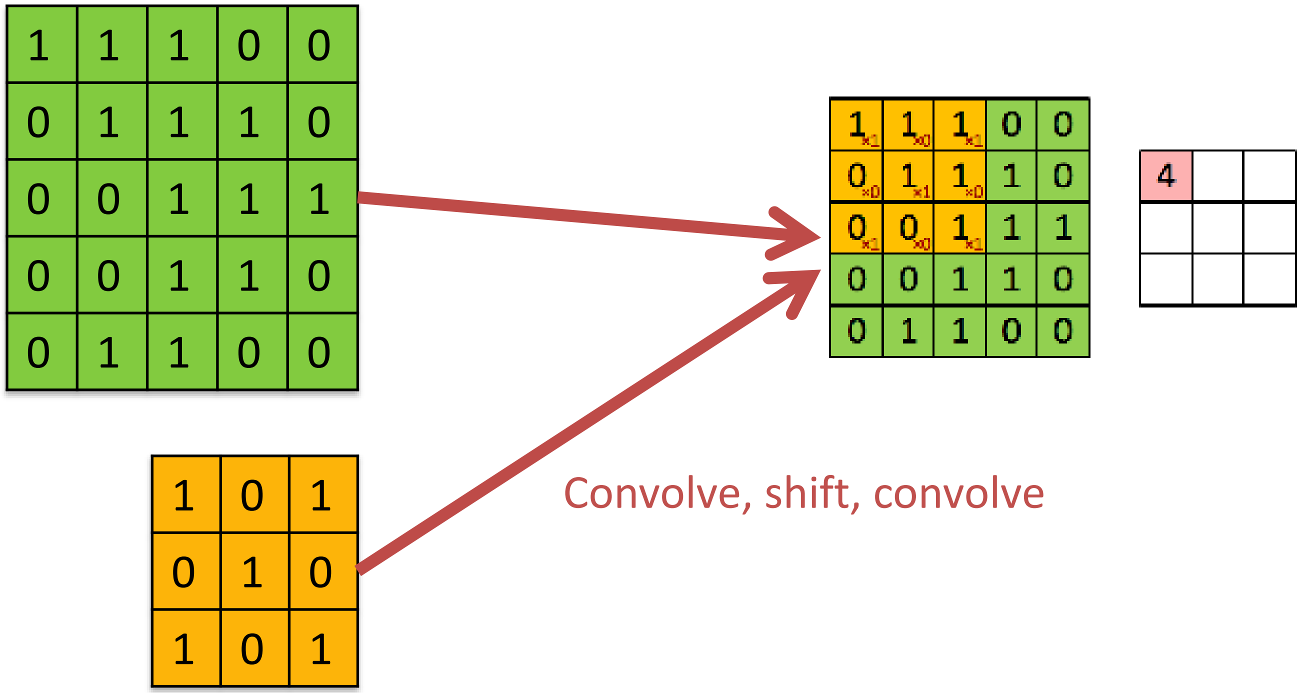
Input image (here binary, RGB typical)

1	0	1
0	1	0
1	0	1

Filter (here binary, but typically continuous values)

https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/

CNNs: Convolution



https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/

CNNs: Convolution

- Typical Filters (

1	0	1
0	1	0
1	0	1

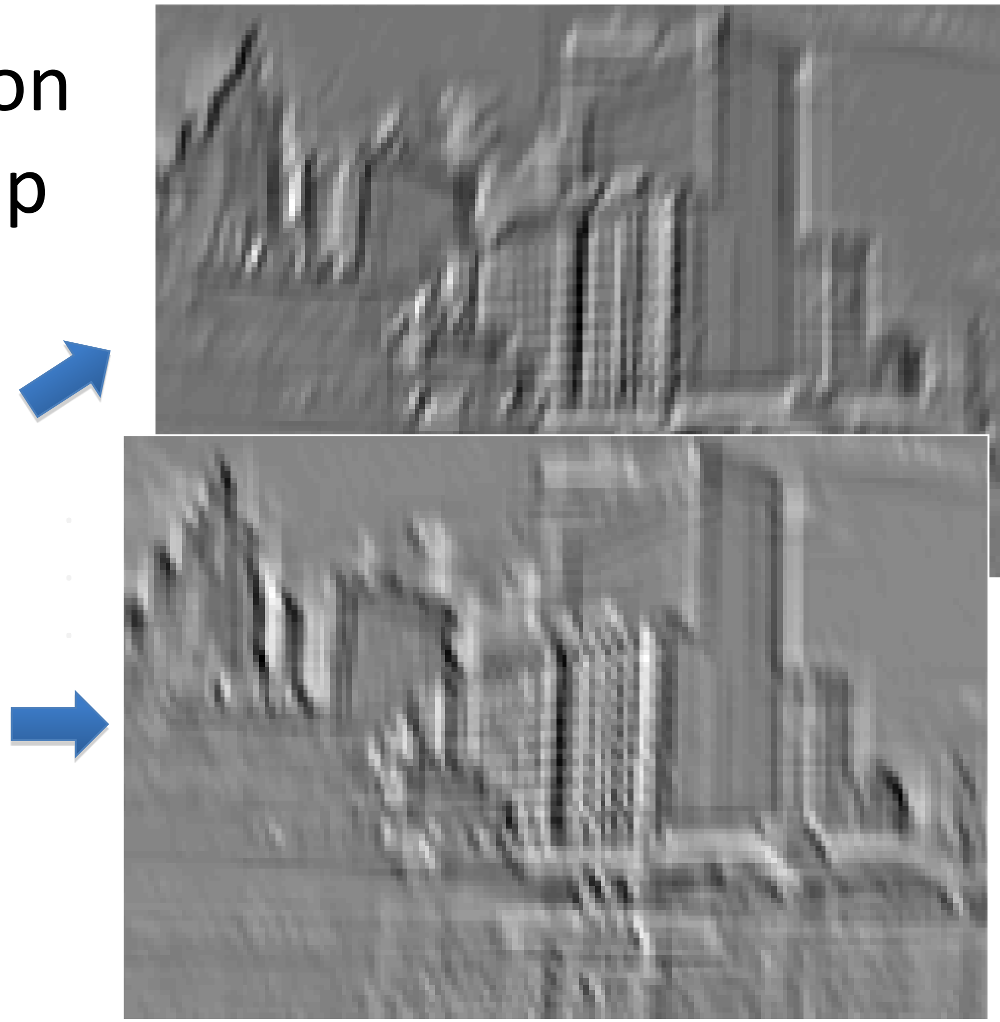
)

Filters are learned

Krizhevsky et al.

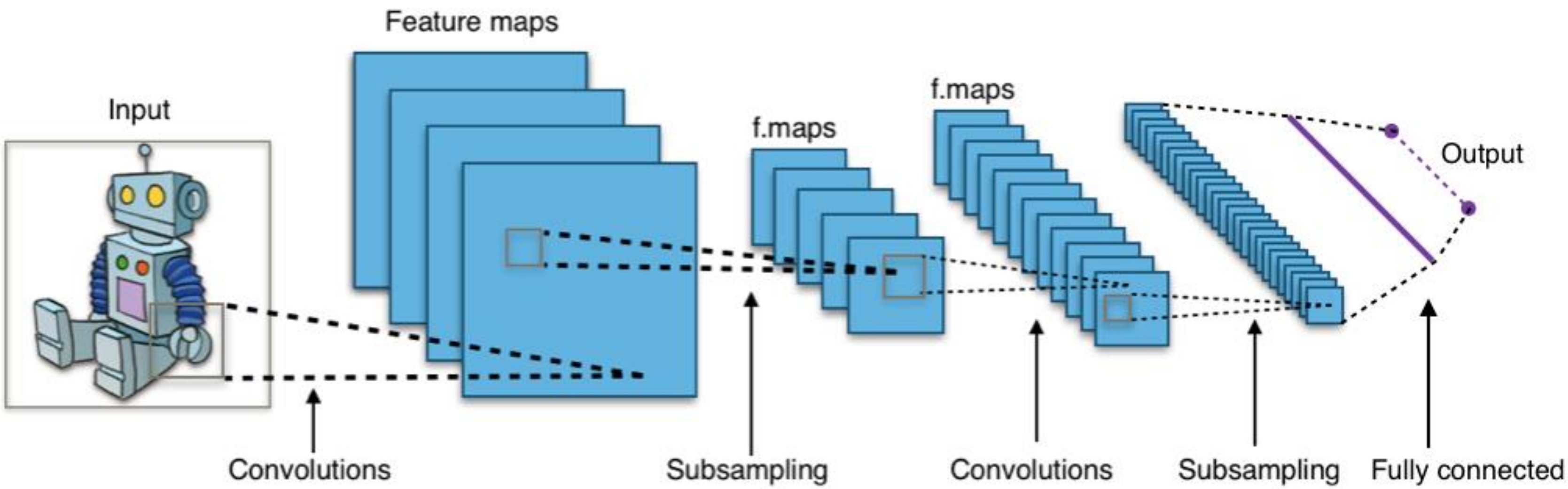
CNNs: Convolution

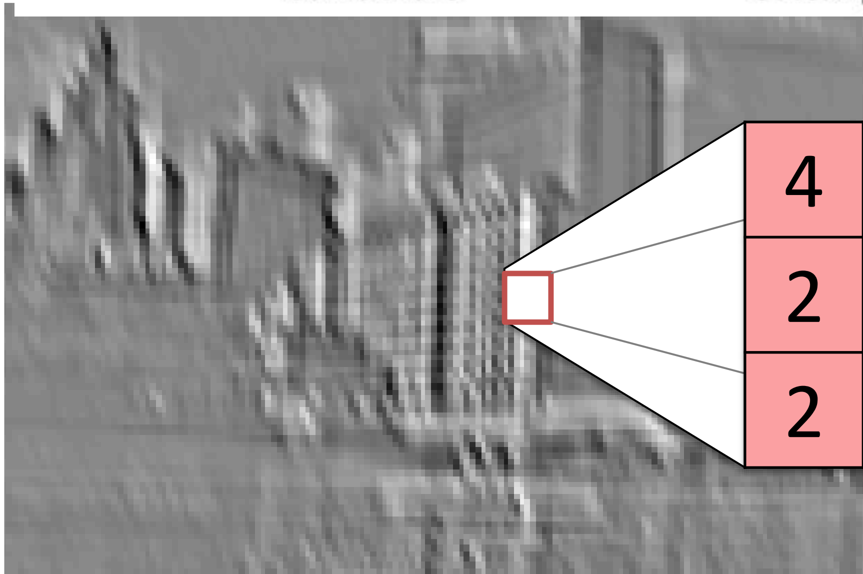
- Output of convolution layer is a feature map
  $$\left(\begin{array}{|c|c|c|} \hline 4 & 3 & 4 \\ \hline 2 & 4 & 3 \\ \hline 2 & 3 & 4 \\ \hline \end{array} \right)$$



http://cs.nyu.edu/~fergus/tutorials/deep_learning_cvpr12/ (Fergus intro)

CNNs: Pool/Subsample





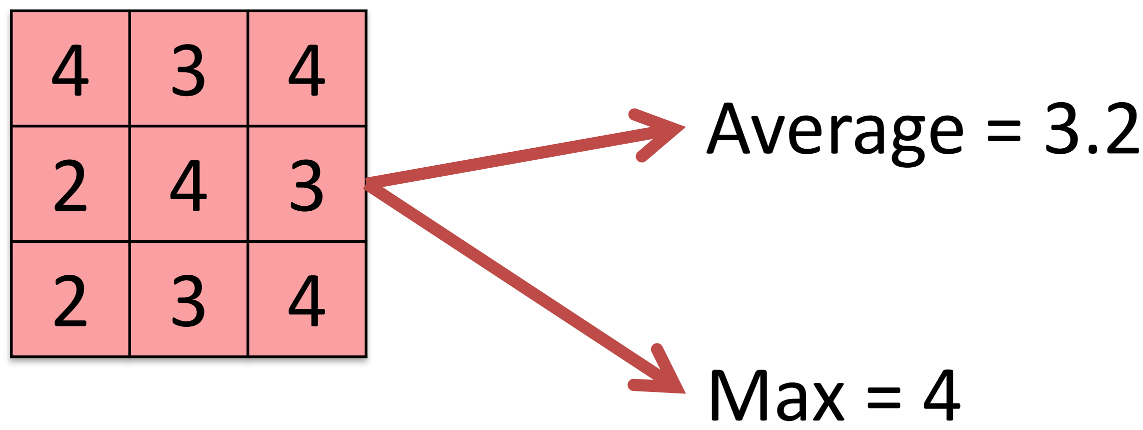
4	3	4
2	4	3
2	3	4

Average = 3.2

Max = 4

Krizhevsky et al.

CNNs: Pool/Subsample

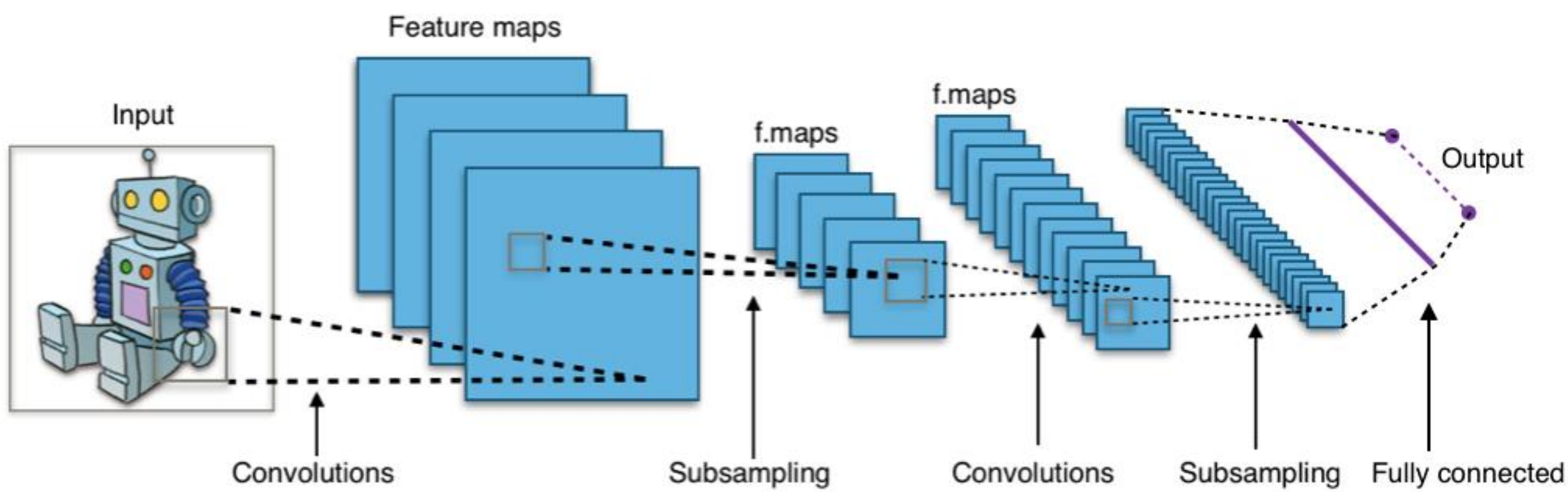


- Supplies some **invariance** to local translations
- Reduces the dimension of subsequent layers

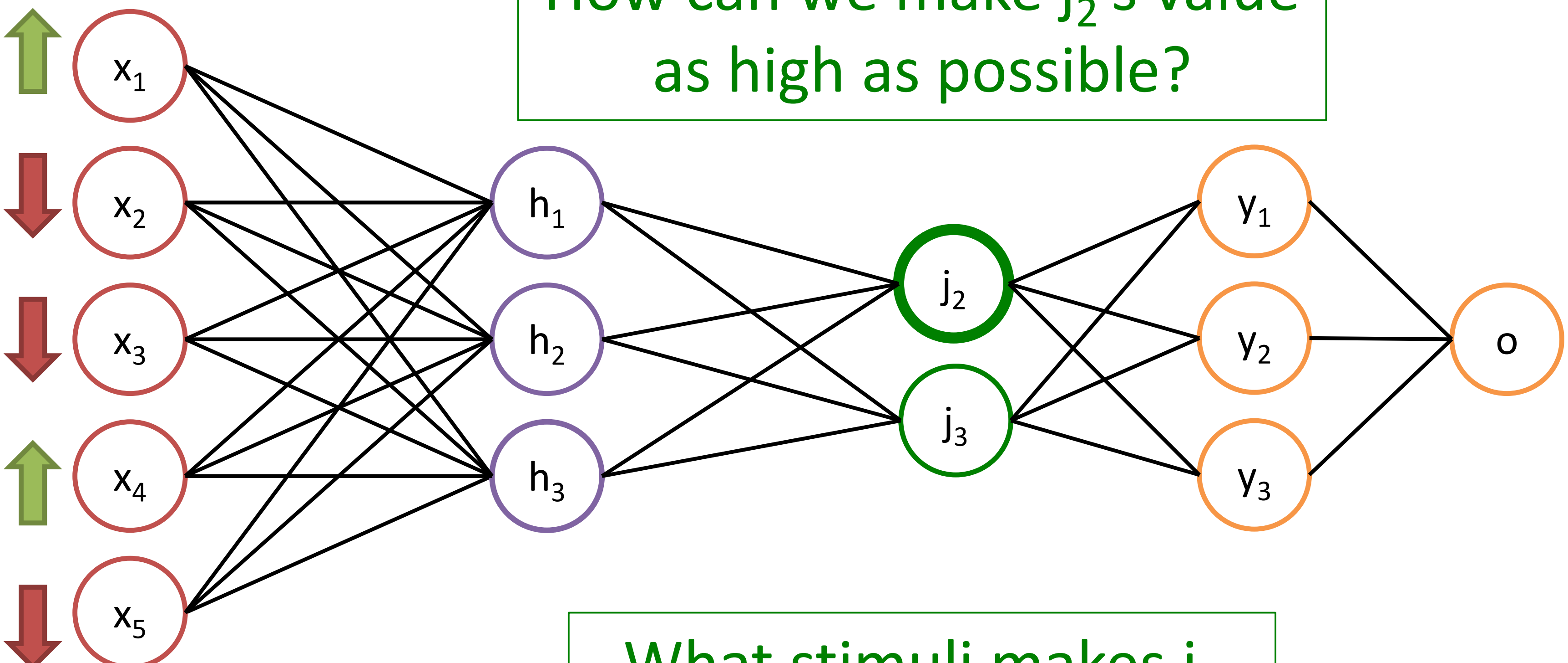
Krizhevsky et al.

CNNs: Hidden representations

- Activations (feature maps) at each layer are a "hidden representation" of the image
- A compression of the information in image
  - not unlike PCA/SVD

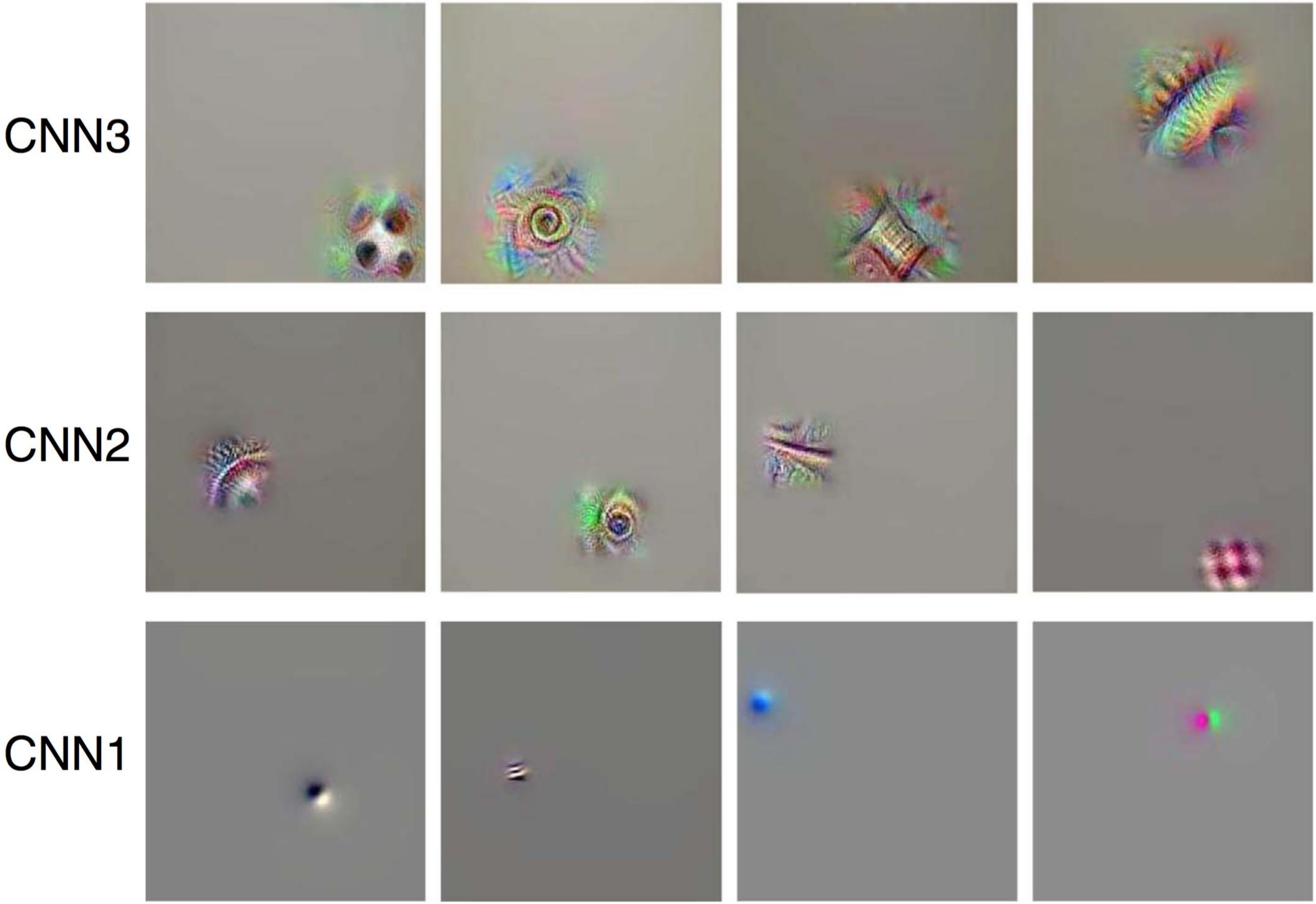


What do the neurons represent?



How can we make j_2's value as high as possible?

What stimuli makes j_2 fire maximally?

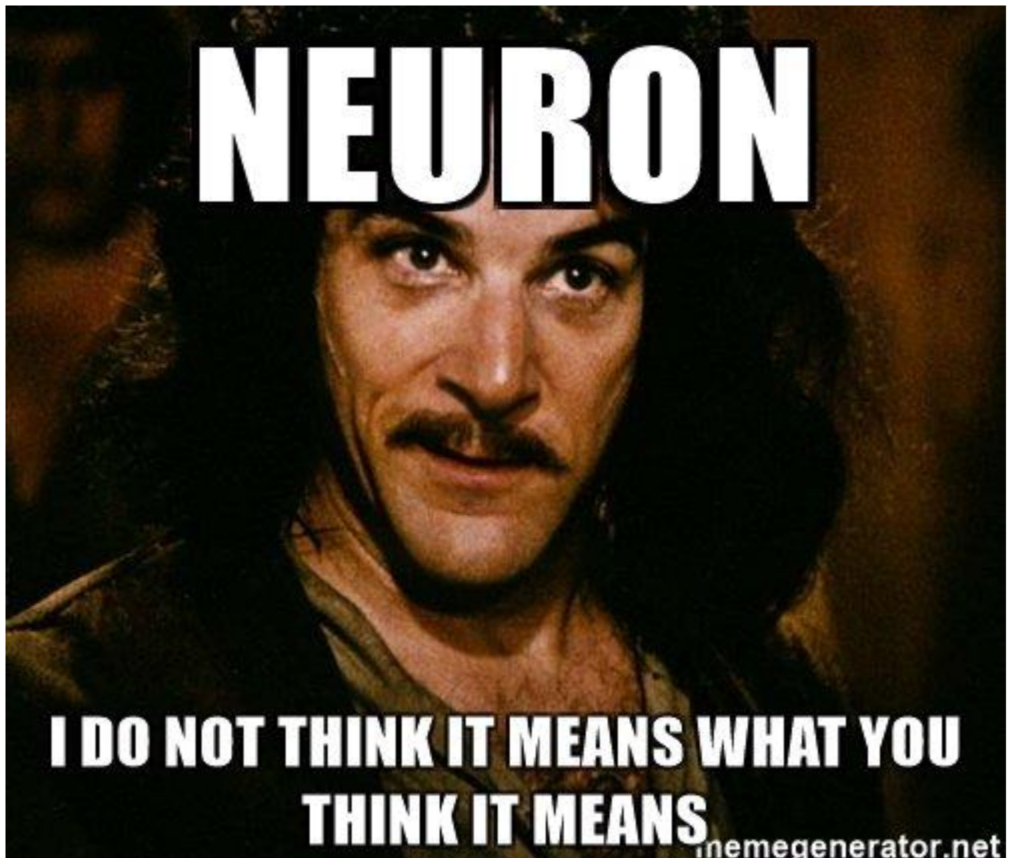


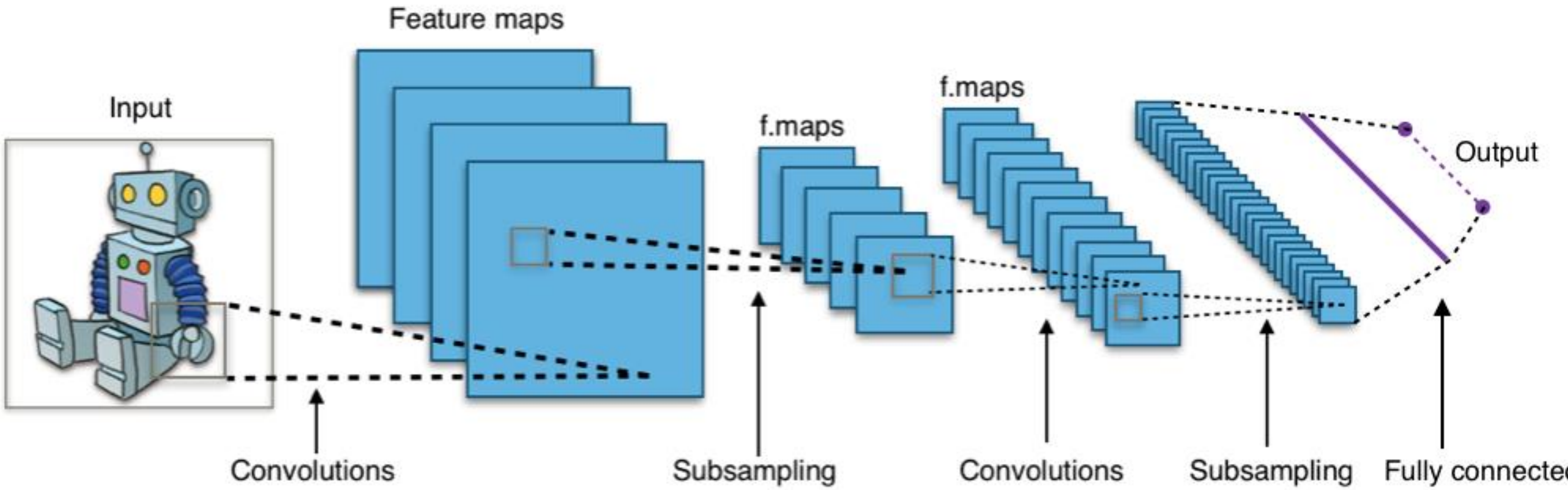
Horikawa & Kamitani (2017)

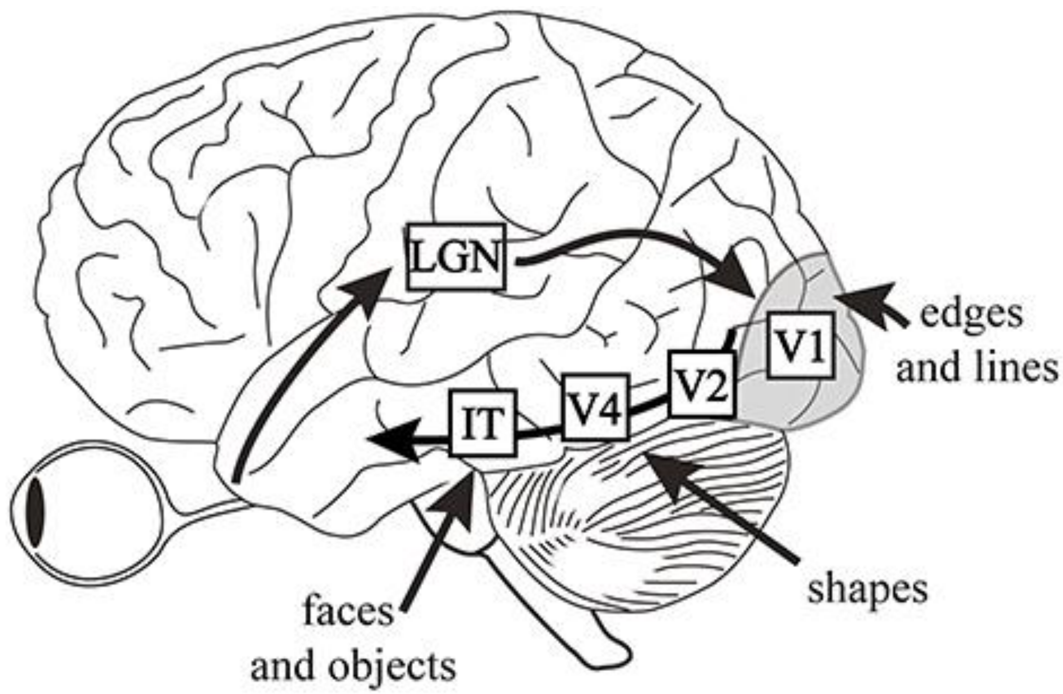
Horikawa & Kamitani (2017)

What do CNNs have to do with the brain?

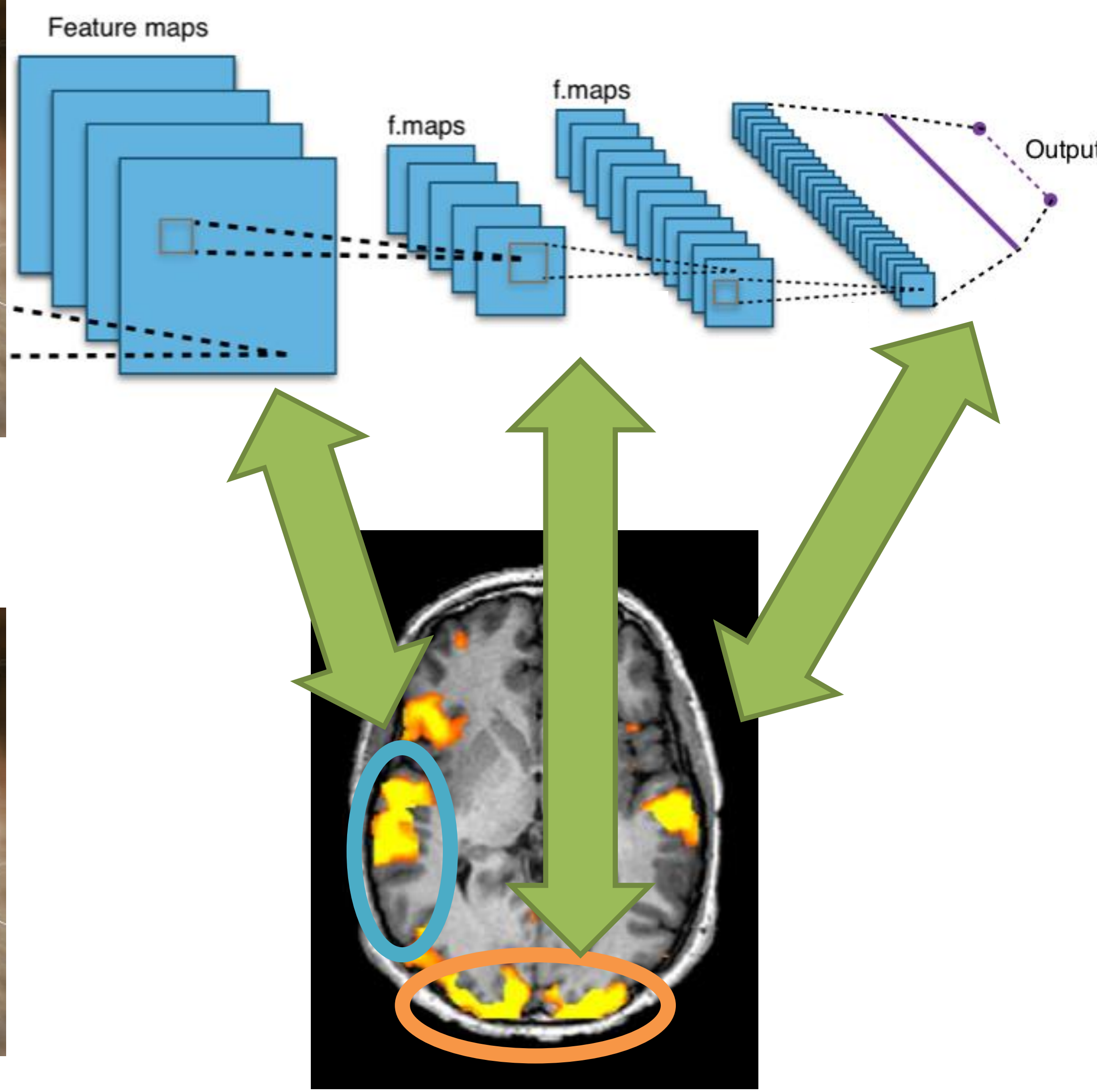
- You keep using the word neuron…



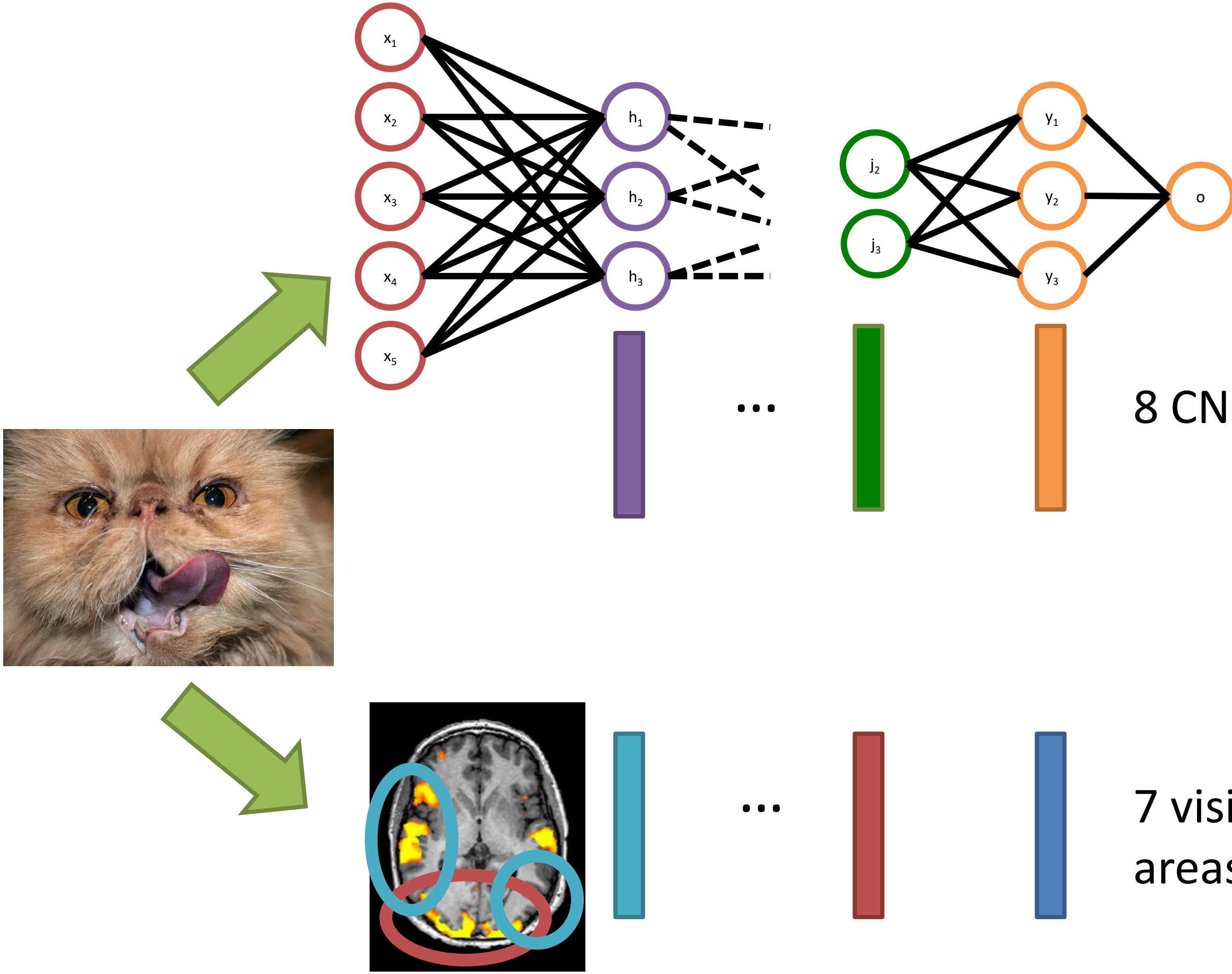

Feature maps
Input
f.maps
f.maps
Output
Convolutions
Subsampling
Convolutions
Subsampling
Fully connected


LGN
V1
edges
and lines
V2
V4
IT
shapes
faces
and objects

Horikawa & Kamitani (2017)



Horikawa & Kamitani (2017)

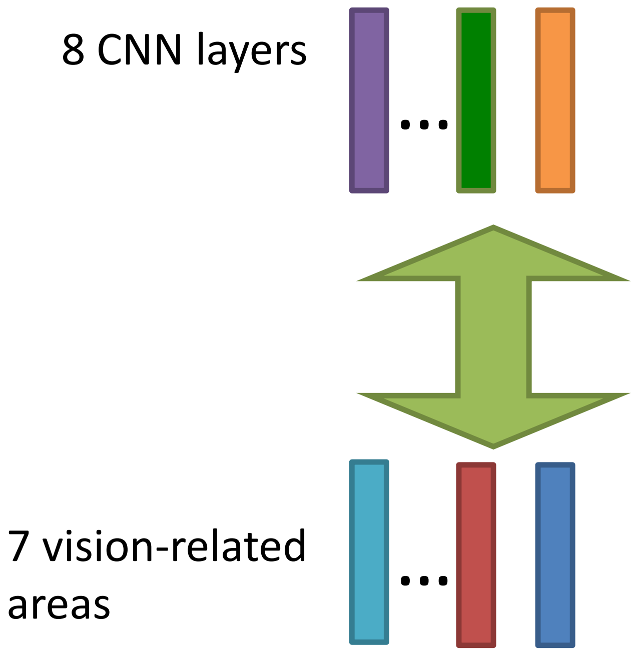


8 CNN layers

7 vision-related areas

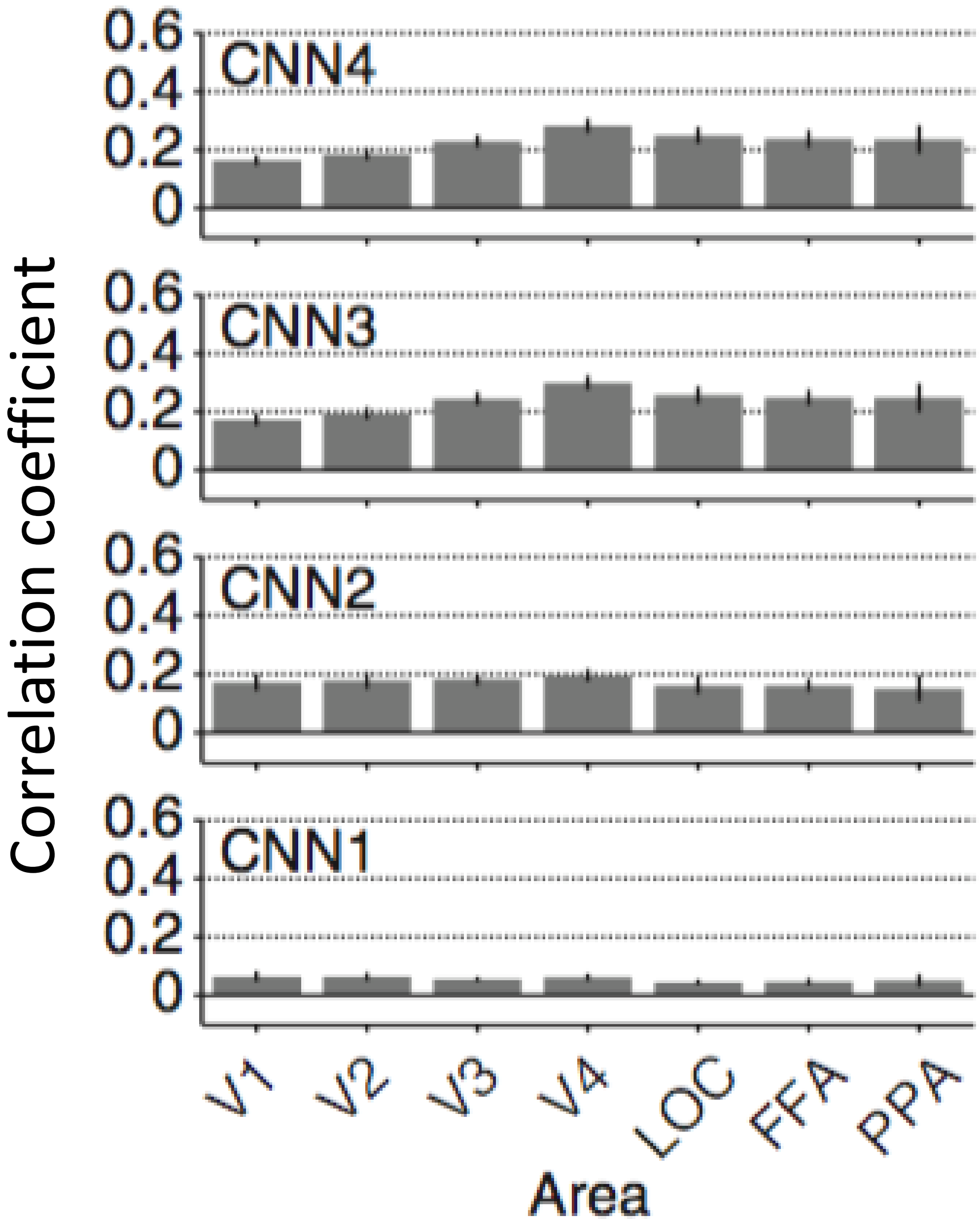
Caveat: fMRI data

CNNs vs the Brain

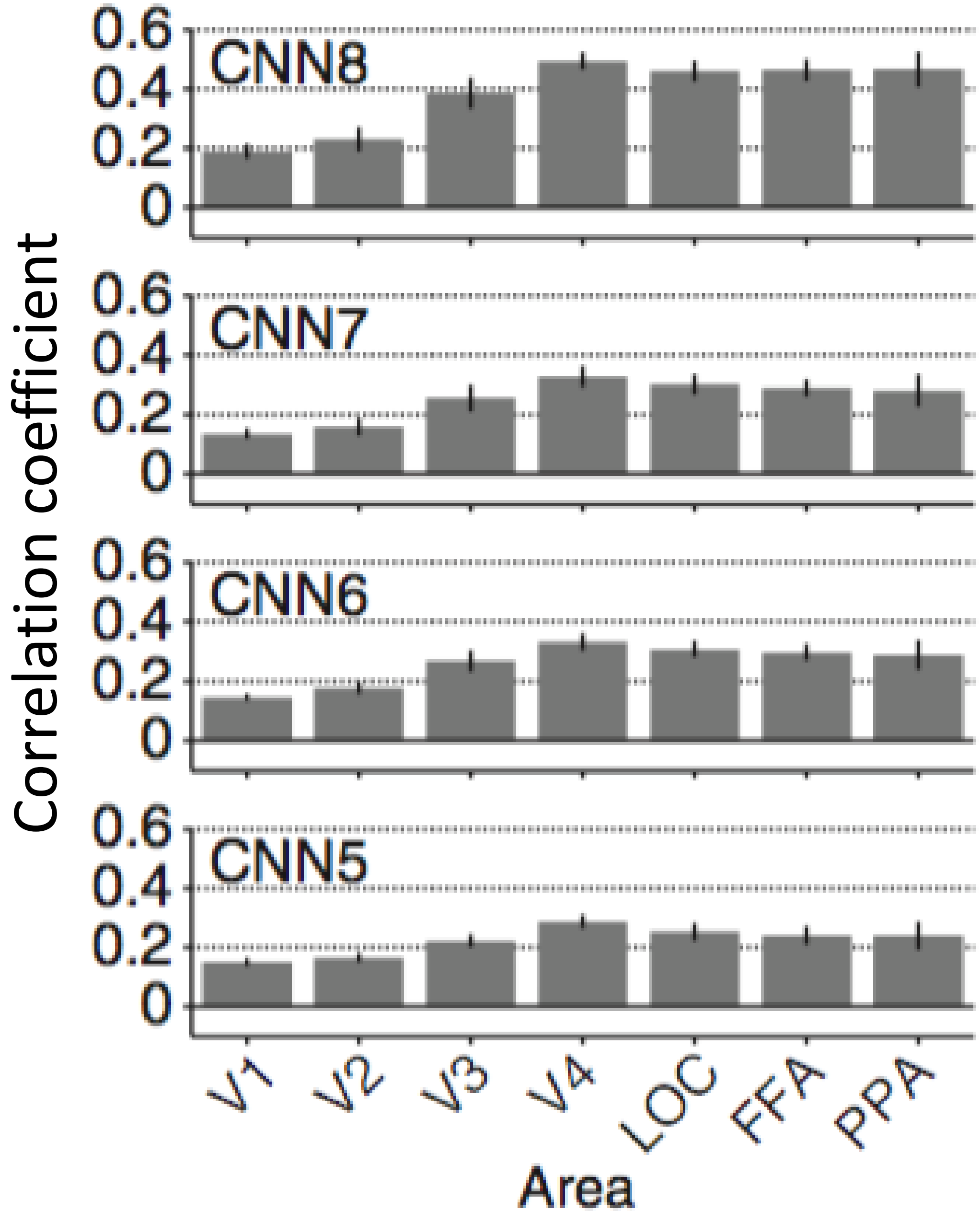


Train all possible 56 models to **predict CNN** hidden representations from **fMRI data**

- Which predictions are most correlated with the true CNN representations?



Horikawa & Kamitani (2017)



Horikawa & Kamitani (2017)

CNNs vs the Brain

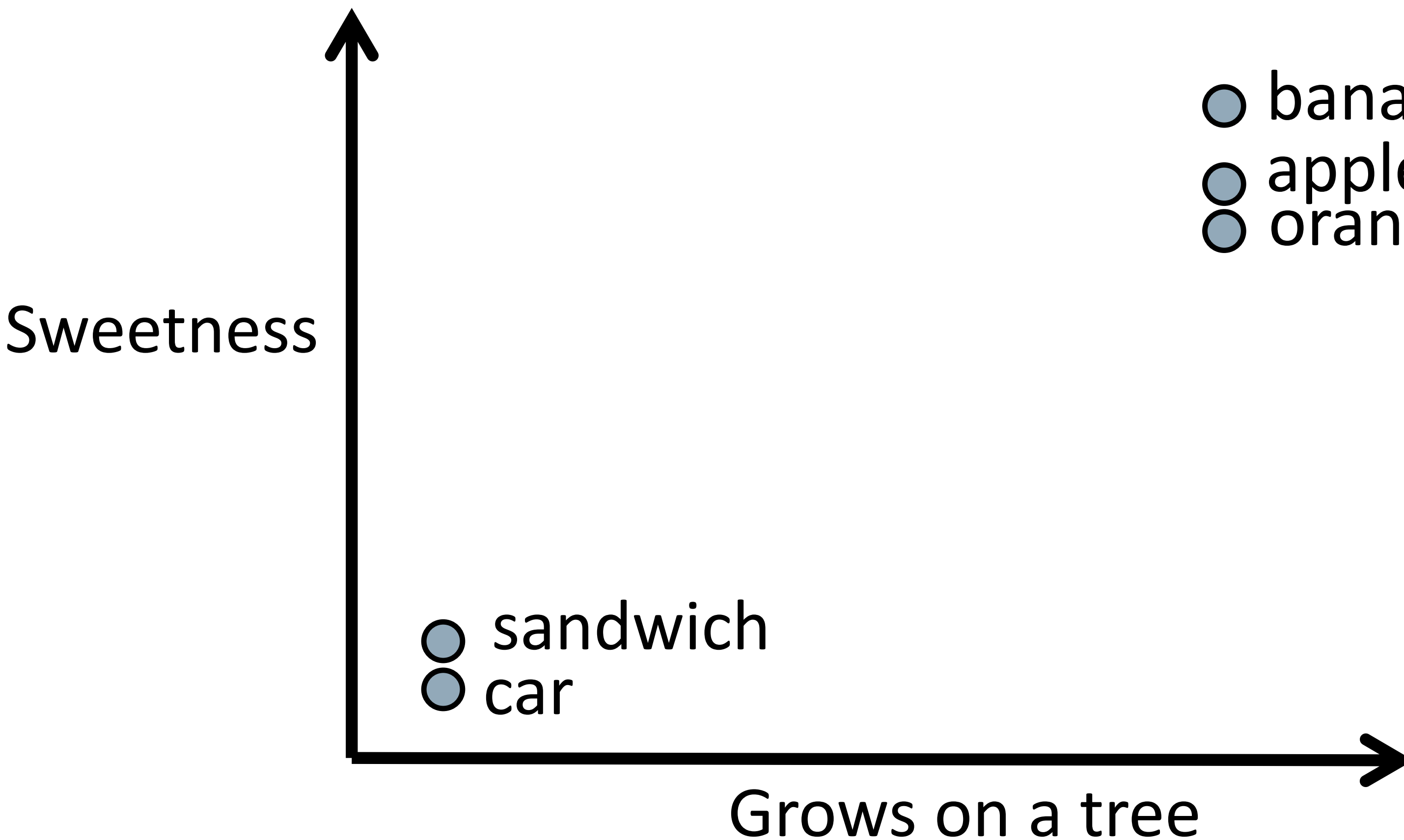
- There is a relationship between higher vision brain areas and higher levels of the CNN

The Robot Student Test

SKILL 2: REPRESENTING KNOWLEDGE

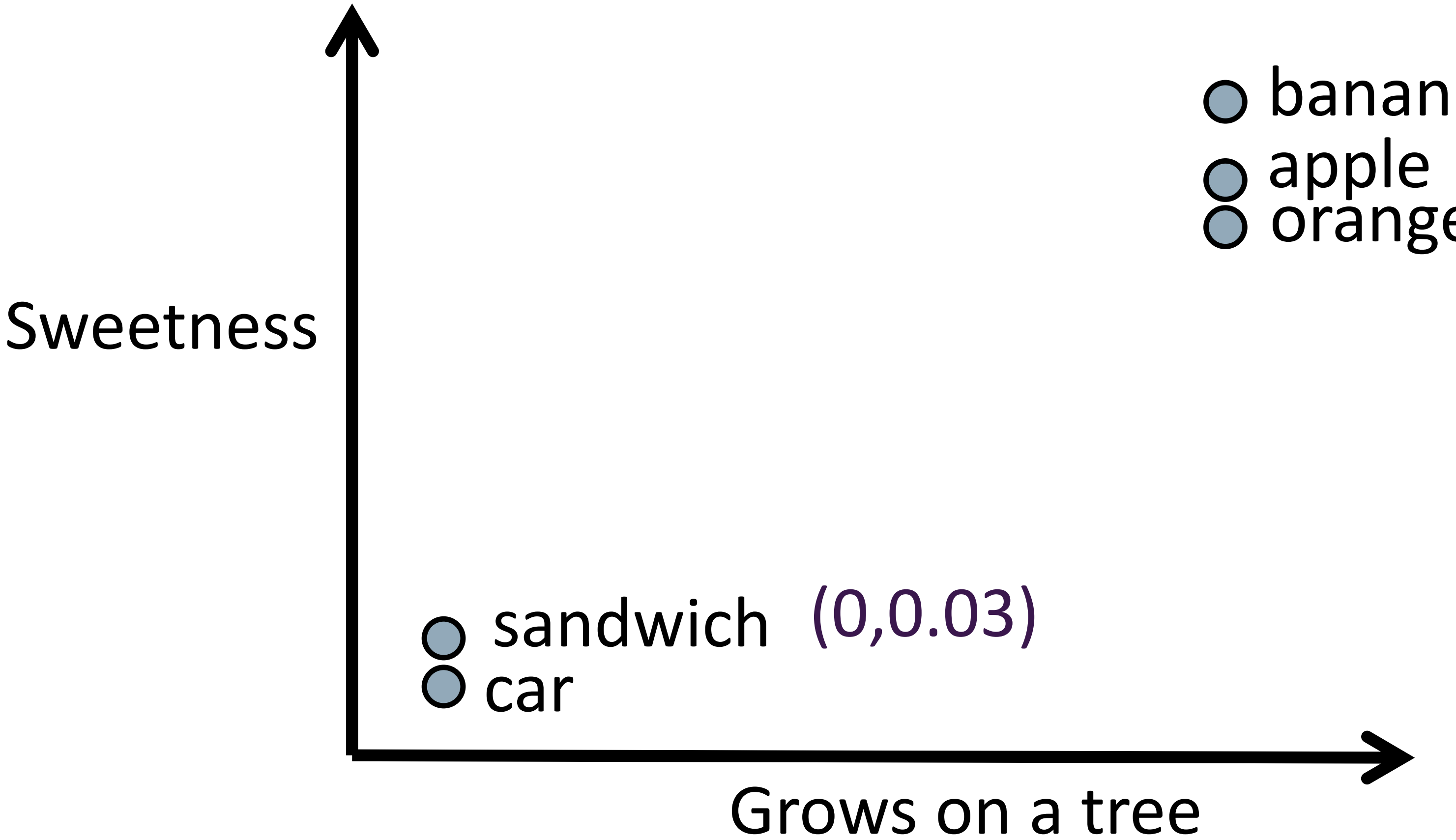
How can we represent knowledge?

How can we represent knowledge?



Vector Space Models (VSMs) of Semantics

Vector Space Models of Semantics



Scaling it up

- How can we differentiate >10k English words?
  - more dimensions (many more!)
- Need to create dimensions and assign values **automatically**

Vector Space Models of Semantics

- Every word gets a vector

Banana	1	.9	9	0	.3	-1	-3	…
Apple	1	.8	.3	4	-.2	2	1	…
Orange	1	.7	.9	2	-.4	1	.1	…

Representing Semantics

- Process a large text corpus
- Find words associated with word of interest
  - banana appears
    - often with verb eat
    - less often with verb drive
- Latent Semantic Analysis (LSA)

(Landauer and Dumais, 1997)

- SkipGram (sometimes called Word2Vec)

(Mikolov et al. 2013)

SkipGram

- Another neural network (surprise!)
- Given a **central** word (e.g. banana), predict probable **context** words (e.g. ate, yellow)
- Use a corpus to generate pairs of central and context words

Joe ate the **banana**, which was yellow

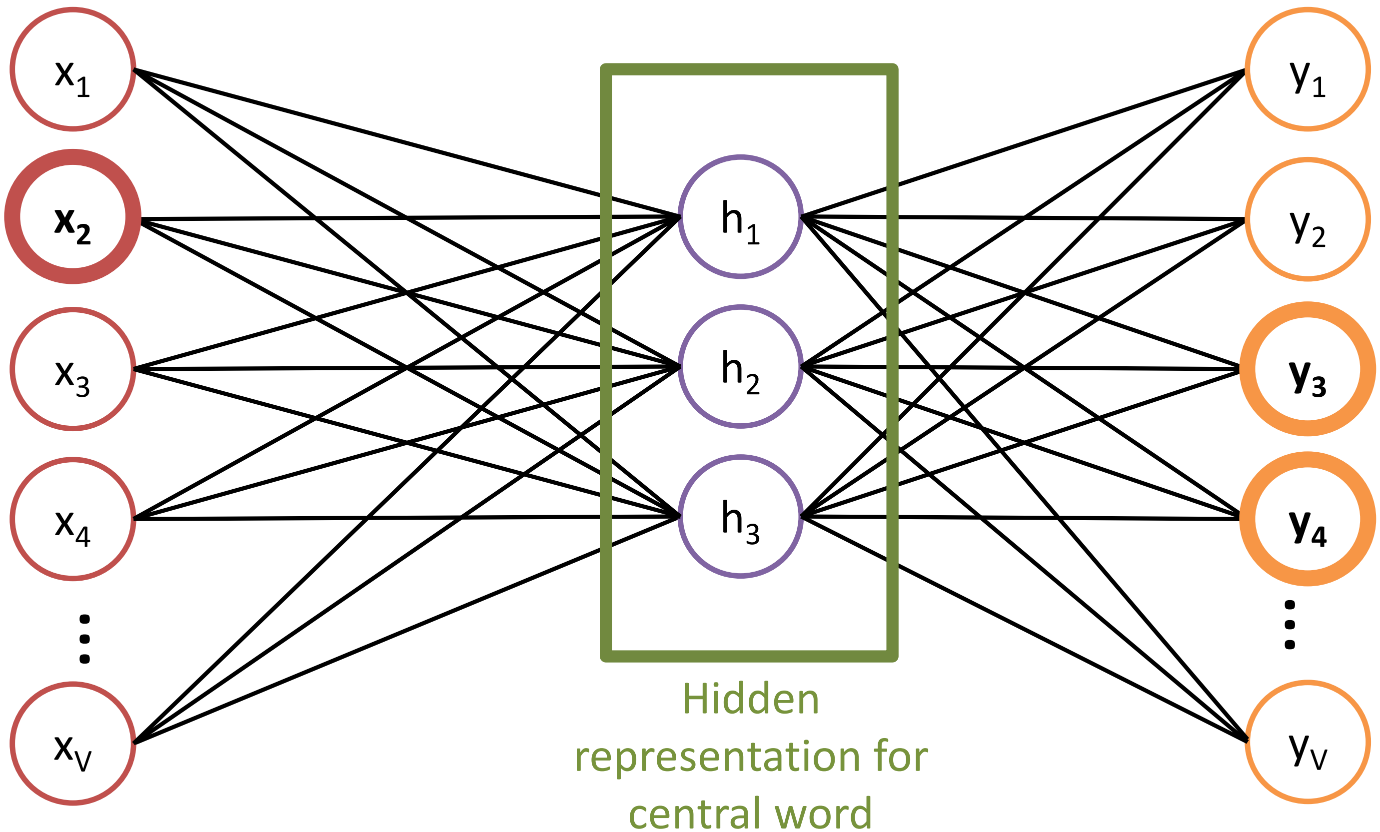
Input

one-hot vector

(1 only for the central word, 0 elsewhere)

Output

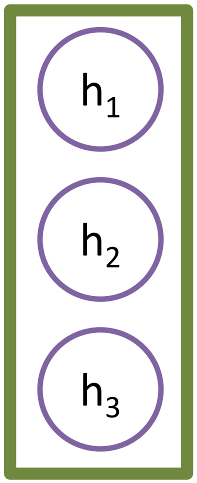
probability distribution over all words



Train so that y values are **high** for commonly co-occurring words, **low** for other words

SkipGram

- We call the **hidden representation** for each central word a word vector
- These vectors have (seemingly) magical properties*



*Actually, it's the data itself that's magical. See Levy, Goldberg, & Dagan (2015)

What can word vectors do?

- Solve analogies
  - Hammer is to nail as screwdriver is to…?
    - screw
  - Solve by finding word closest to

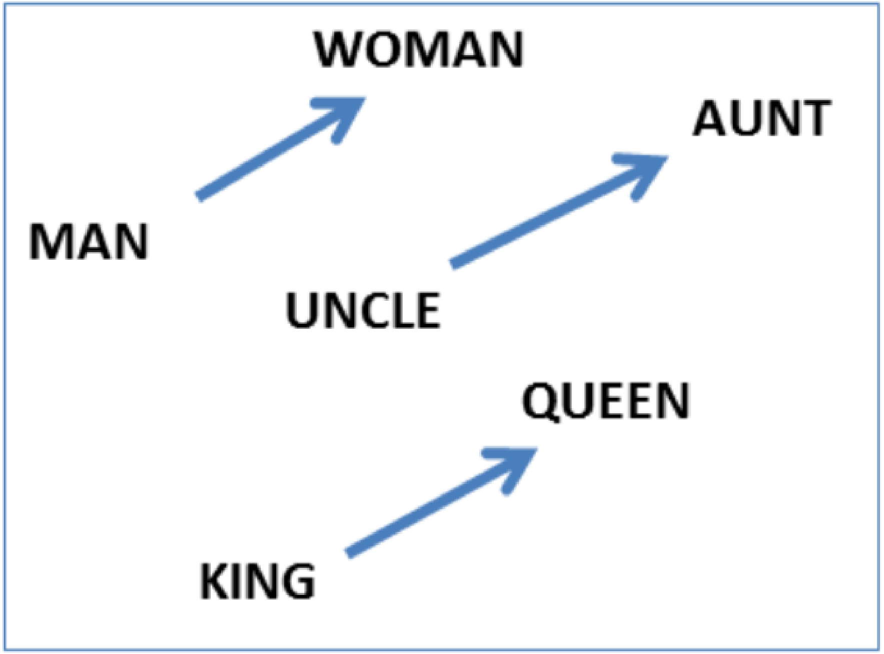
    (nail – hammer + screwdriver)

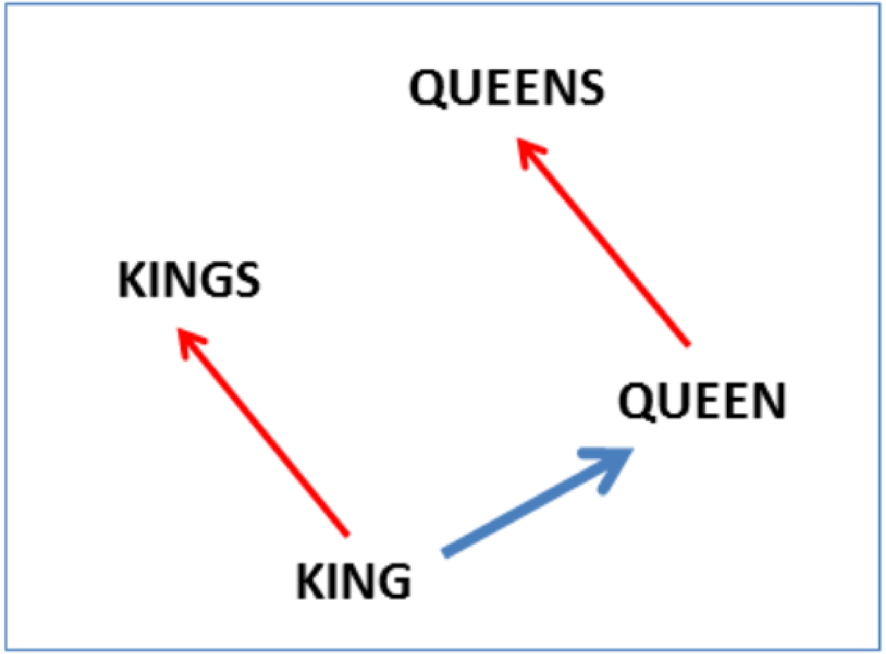
Linguistic Regularities

Newspapers			
New York San Jose	New York Times San Jose Mercury News	Baltimore Cincinnati	Baltimore Sun Cincinnati Enquirer
NHL Teams			
Boston Phoenix	Boston Bruins Phoenix Coyotes	Montreal Nashville	Montreal Canadiens Nashville Predators
NBA Teams			
Detroit Oakland	Detroit Pistons Golden State Warriors	Toronto Memphis	Toronto Raptors Memphis Grizzlies
Airlines			
Austria Belgium	Austrian Airlines Brussels Airlines	Spain Greece	Spainair Aegean Airlines
Company executives			
Steve Ballmer Samuel J. Palmisano	Microsoft IBM	Larry Page Werner Vogels	Google Amazon

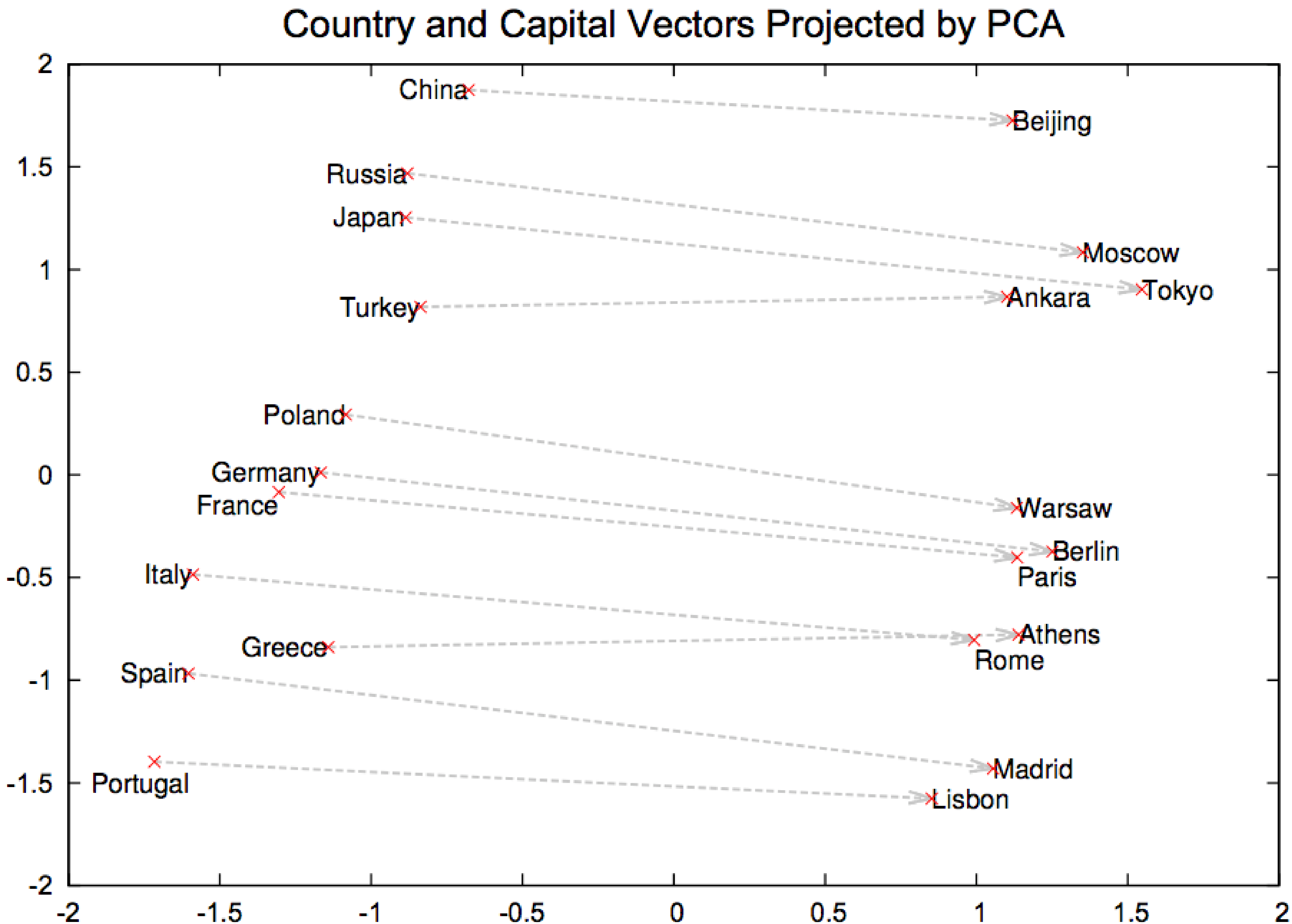
Mikolov, Sutskever, Chen, Corrado, & Dean (2013)

Linguistic Regularities





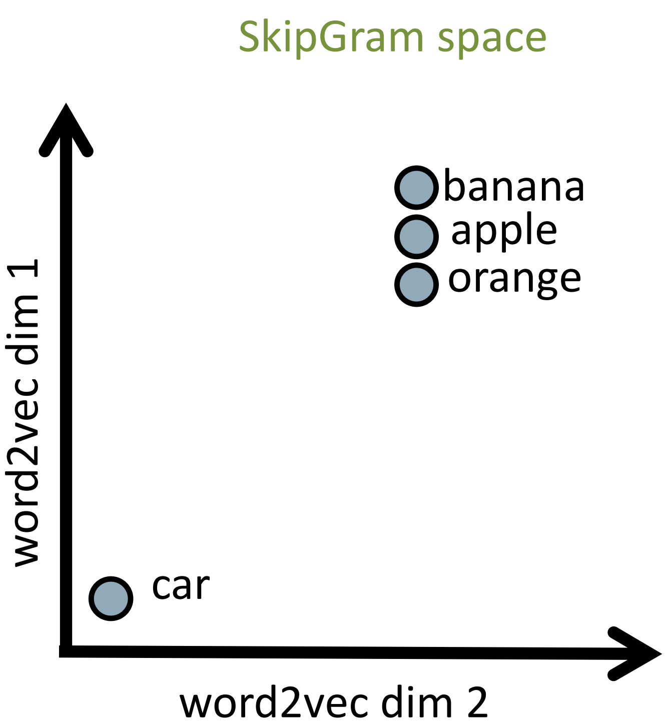
Mikolov, Yih & Zweig (2013)

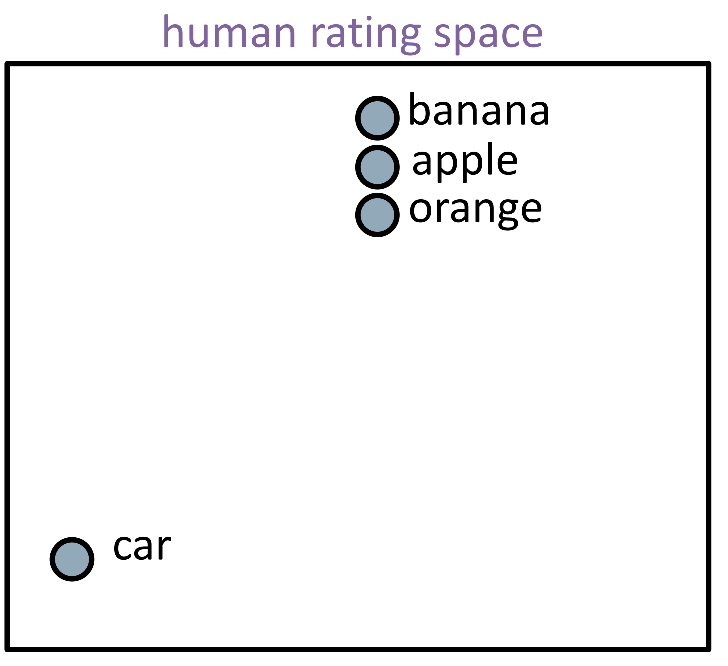


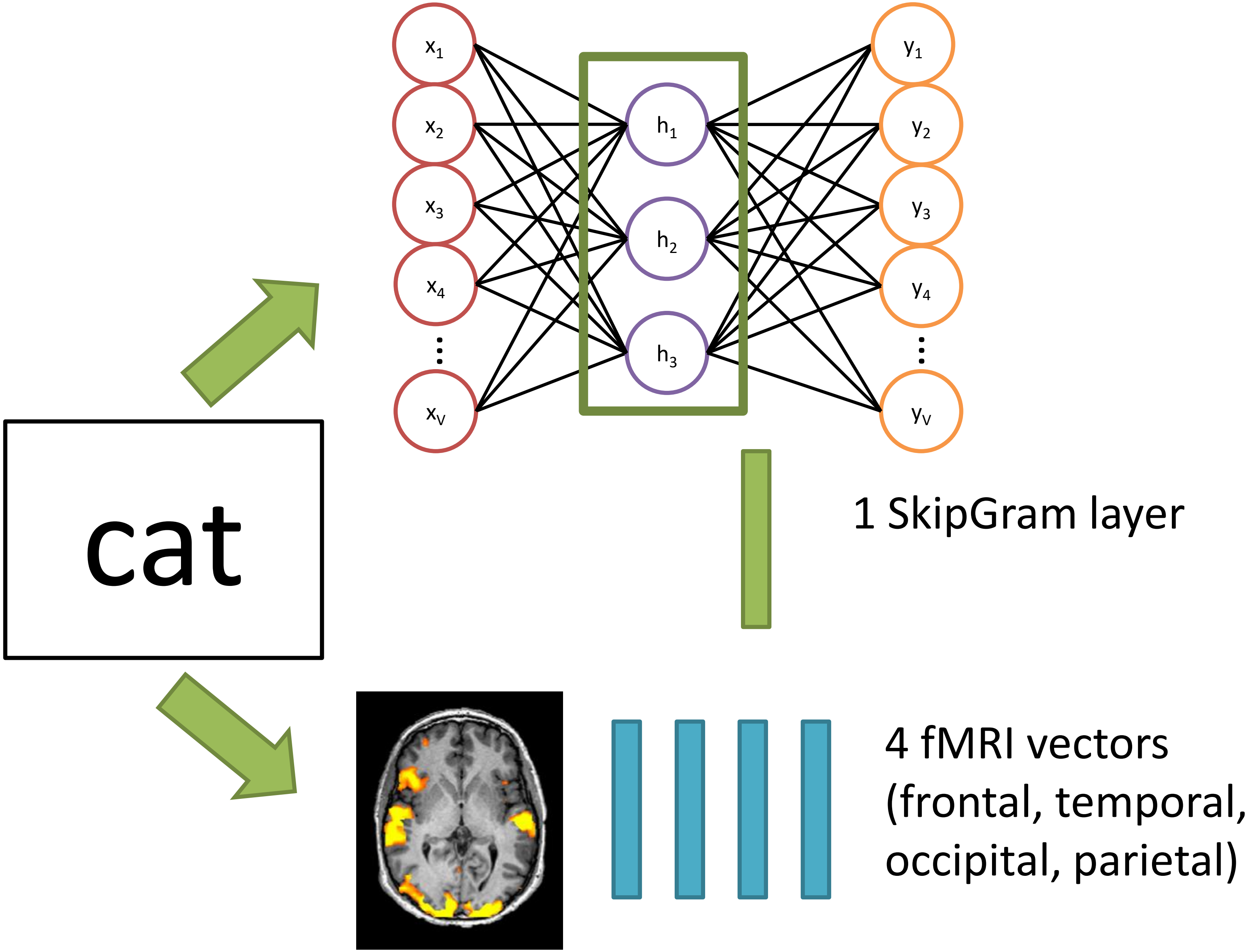
Mikolov, Sutskever, Chen, Corrado, & Dean (2013)

What can word vectors do?

- Predict word similarity






cat
x1
x2
x3
x4
xV
h1
h2
h3
y1
y2
y3
y4
yV
1 SkipGram layer
4 fMRI vectors
(frontal, temporal,
occipital, parietal)

RSA-style analysis, SkipGram <-> fMRI

	Frontal	**Temporal**	**Parietal**	**Occipital**
Skip-gram	**0.1450** (0.0e+00)	**0.1483** (0.0e+00)	**0.2317** (0.0e+00)	-0.0385 (9.4e-01)

Ruan, Ling & Hu (2016)

BrainBench

The brain as a test bed for
learned word representations

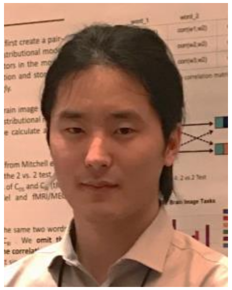
Xu, Murphy & Fyshe, 2016

BrainBench

- Tests your favorite word vectors against the brain's representations
- Multiple Modalities
  - fMRI
  - MEG
  - EEG
- English and Italian
- Abstract and concrete nouns
- And growing!!

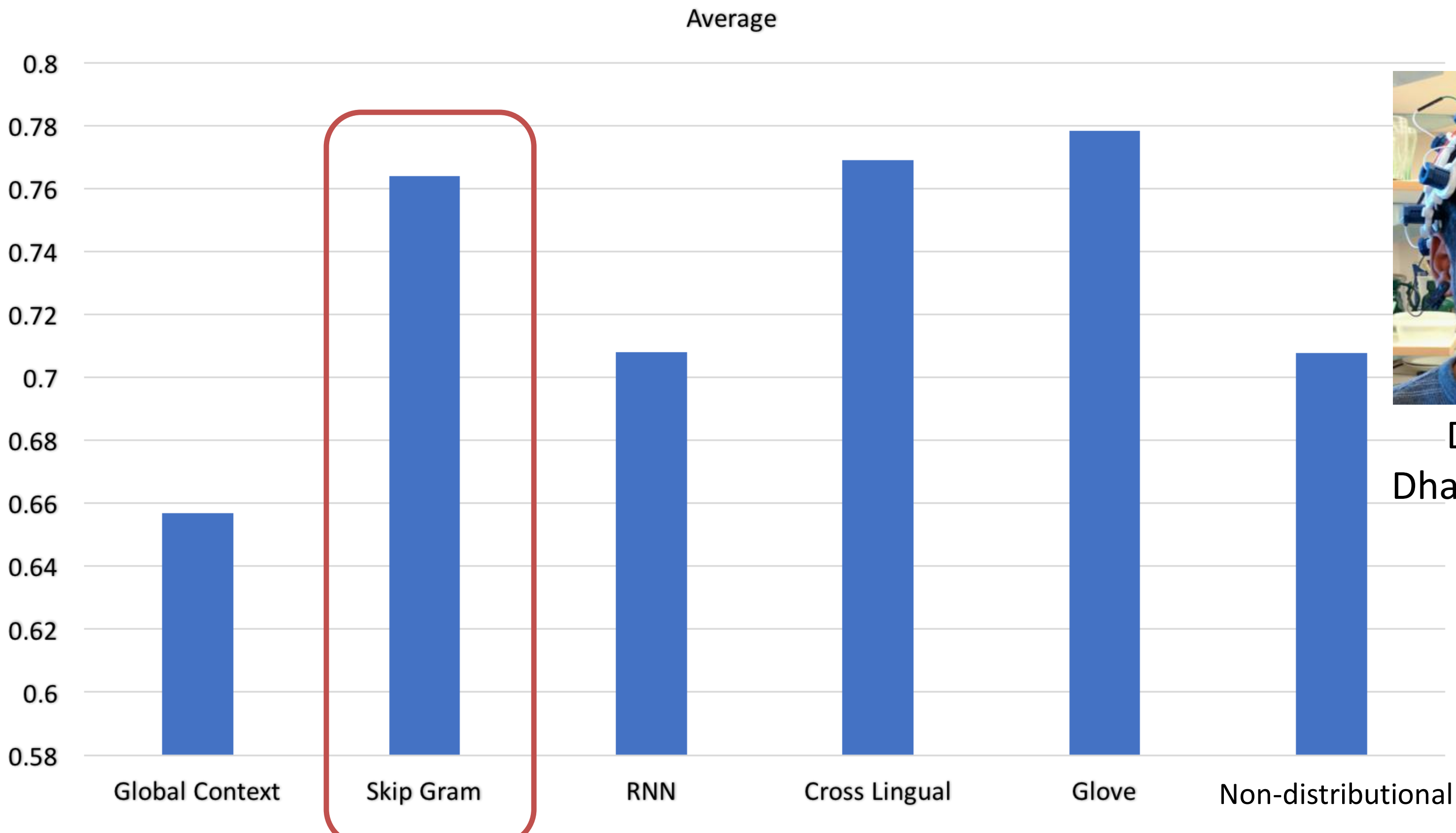
Xu, Murphy & Fyshe, 2016

Average across 4 datasets



Haoyan (Ed) Xu

Dhanush Dharmaretnam



http://www.langlearnlab.cs.uvic.ca/brainbench/

Xu, Murphy & Fyshe, 2016

SkipGram vs the Brain

- SkipGram vectors are correlated with fMRI activity
- So are lots of other word vector models

The Robot Student Test

SKILL 3: COMMUNICATE VIA LANGUAGE

Recurrent Neural Network

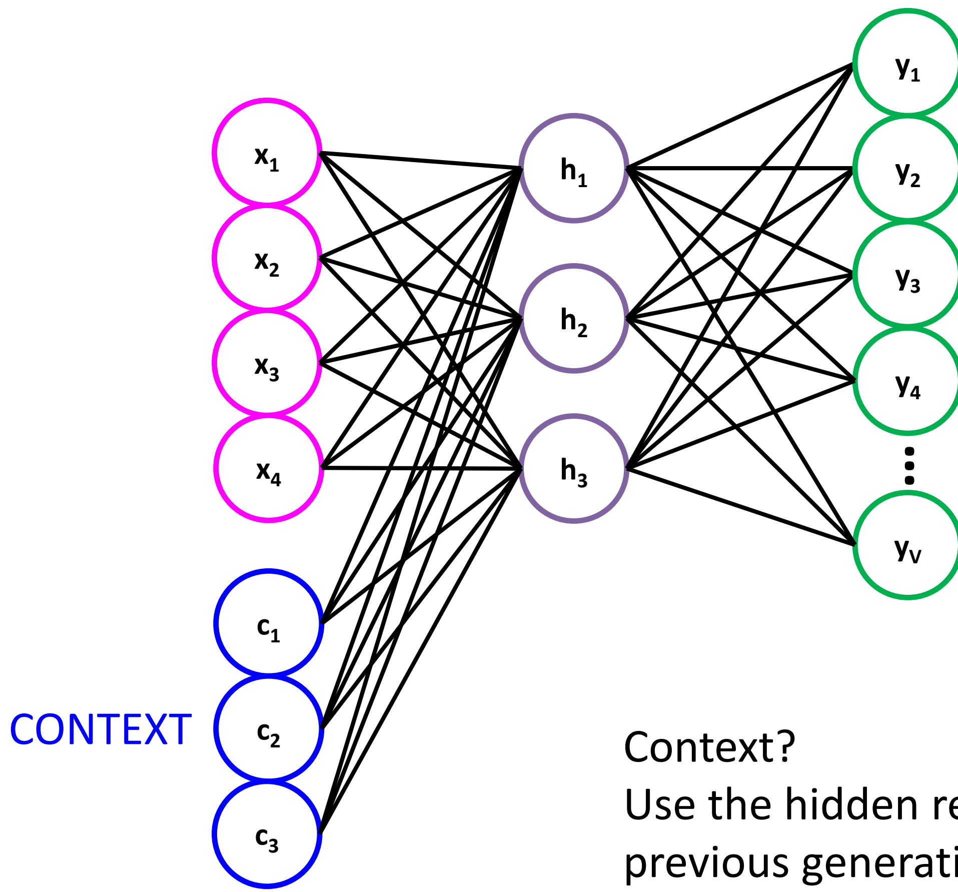
- Neural networks that write
- Predict the next word as a function of
  - context
  - previous
- Write that word, and recurse

Input

Word vector for the current word

Output

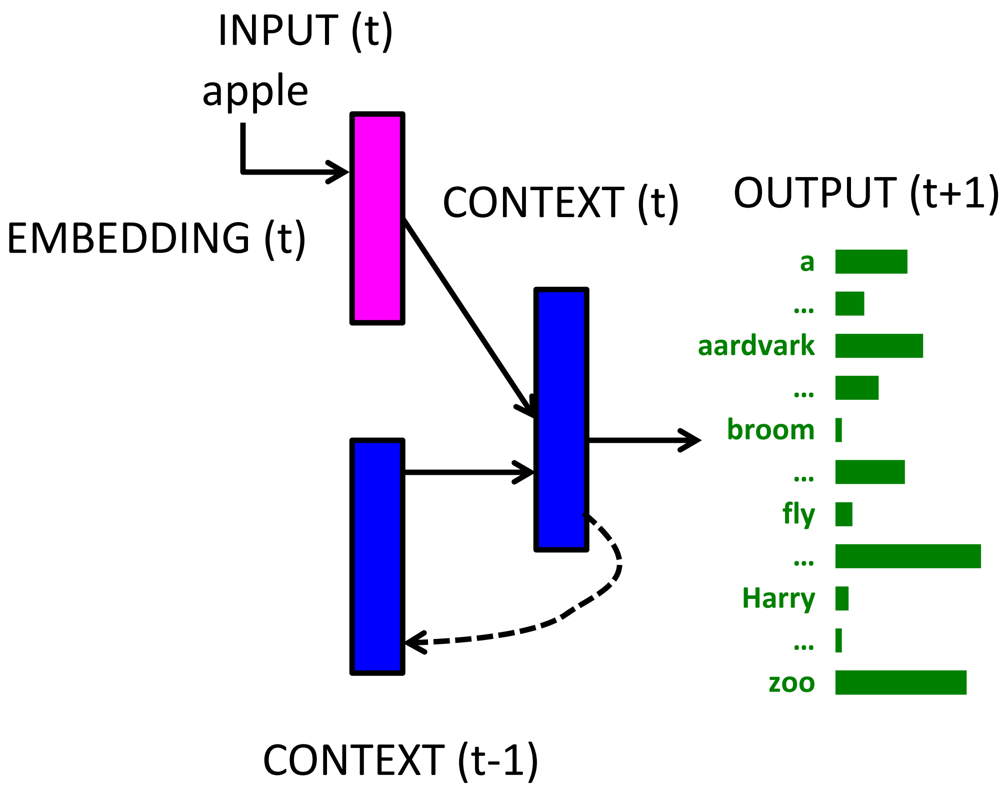
probability distribution over **next word in sequence**



Sample from this distribution to determine the next word

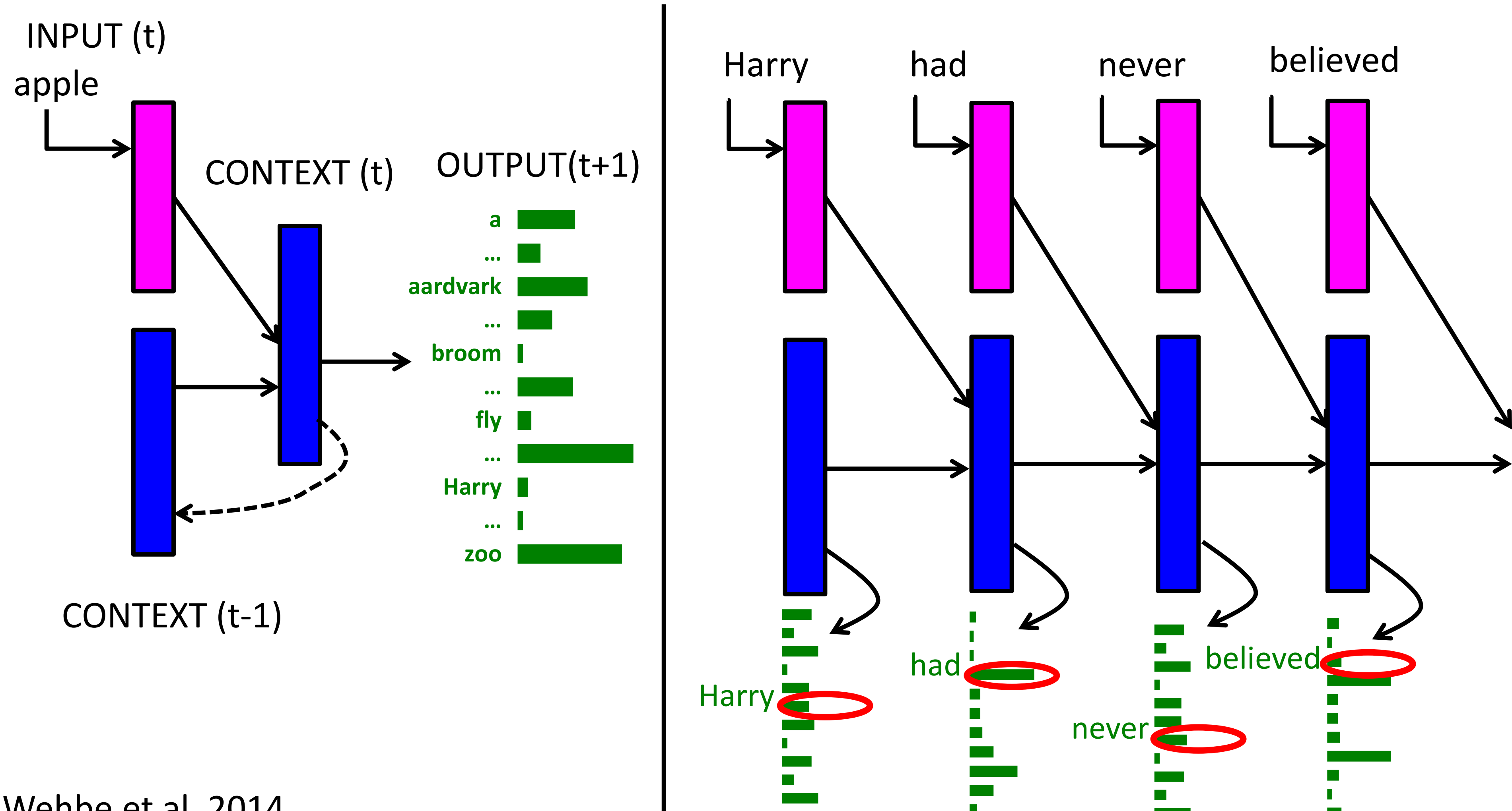
Context?
Use the hidden representation from previous generation step

Recurrent Neural Network Language Model



Mikolov, Kombrink, Deoras, Burget & Cernocky (2011)

Recurrent Neural Network Language Model



Wehbe et al. 2014

Sunspring

```
We see H pull a book from a shelf, flip through it while
speaking, and then put it back.

                    H
          In a future with mass unemployment,
          young people are forced to sell
          blood. That's the first thing I can
          do.

                    H2
          You should see the boys and shut
          up. I was the one who was going to
          be a hundred years old.

                    H
          I saw him again. The way you were
          sent to me... that was a big honest
          idea. I am not a bright light.

                    C
          Well, I have to go to the skull.  I
          don't know.
```

https://arstechnica.com/gaming/2016/06/an-ai-wrote-this-movie-and-its-strangely-moving/

What do RNNs have to do with the brain?

Harry Potter and the Reading Brain

Wehbe et al. 2014

- MEG
- Read Chapter 9 of *Harry Potter and the Sorcerer's Stone*.
- Read one word at a time, each word for 0.5 s

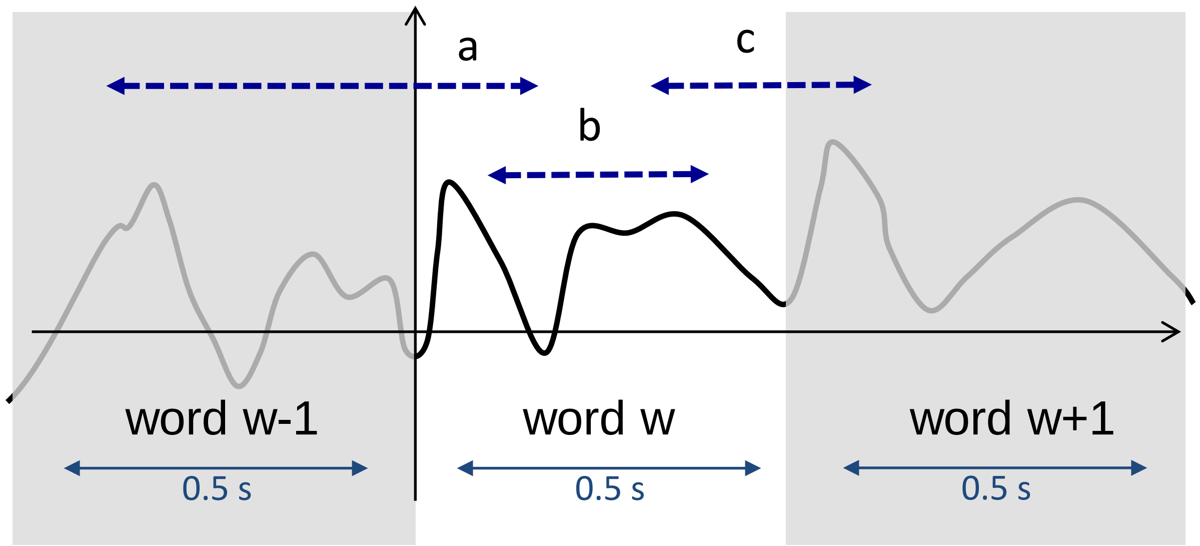
When and where do the brain processes occur?

Conjecture

(a) Story context before seeing word w

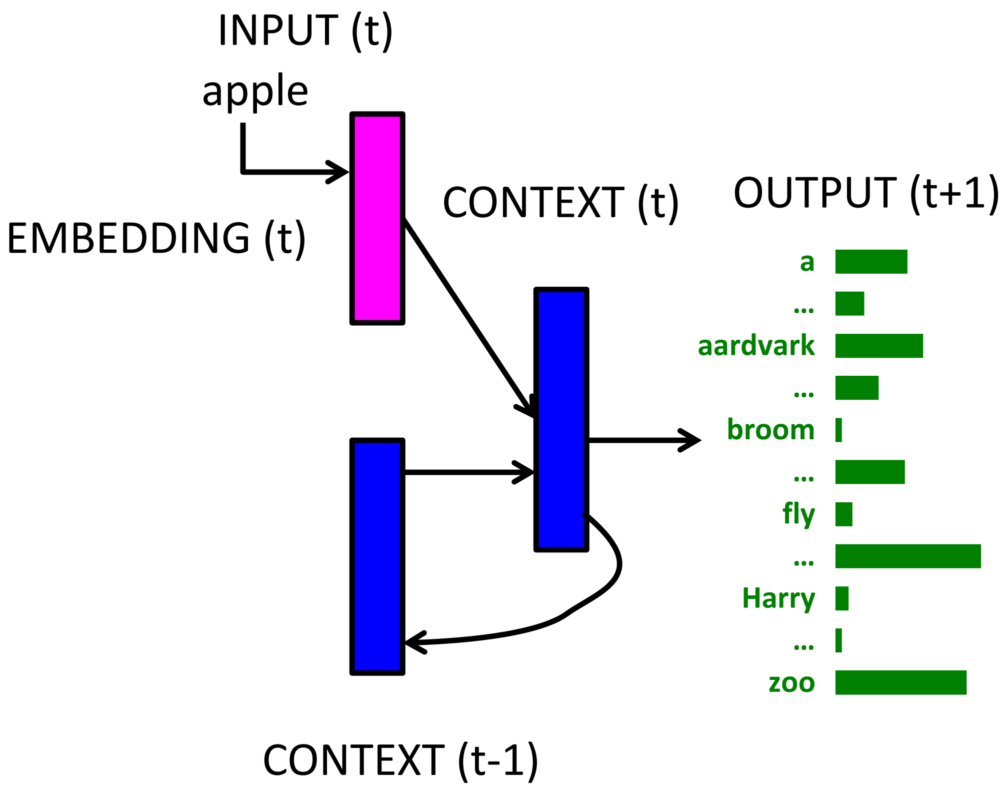
(b) Perception of word w

(c) Integration of word w



Wehbe et al. 2014

Recurrent Neural Network Language Model



Tomas Mikolov, Stefan Kombrink, Anoop Deoras, Lukar Burget, and J Cernocky.
RNNLM- recurrent neural network language modeling toolkit.
ASRU Workshop 2014.

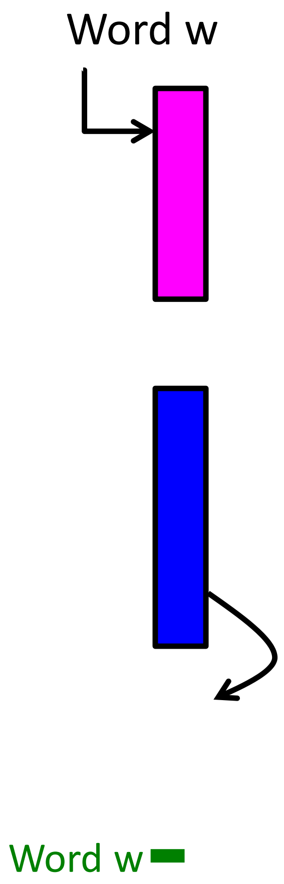
Wehbe et al. 2014

Parallelism

Brain processes:		Neural Net components:
(a) Represent context	$\longleftrightarrow$	(a) Context vector
(b) Retrieves properties	$\longleftrightarrow$	(b) Word Embedding
(c) Integrate the word with context	$\longleftrightarrow$	(c) P(w | context)

Wehbe et al. 2014

Annotate every word with new features



- Embedding of word w
- Context (before word w is seen)
- Probability of word w given context

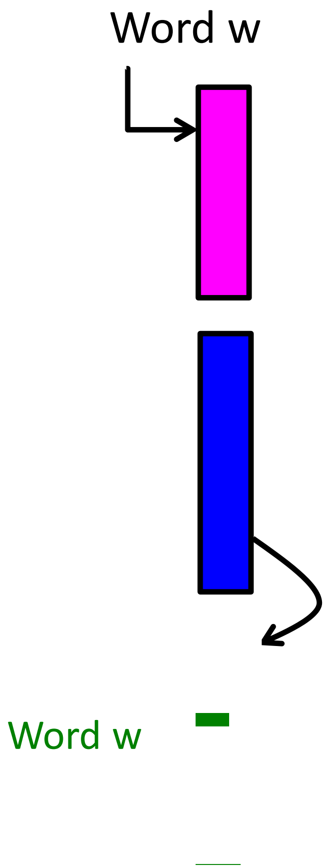
Wehbe et al. 2014

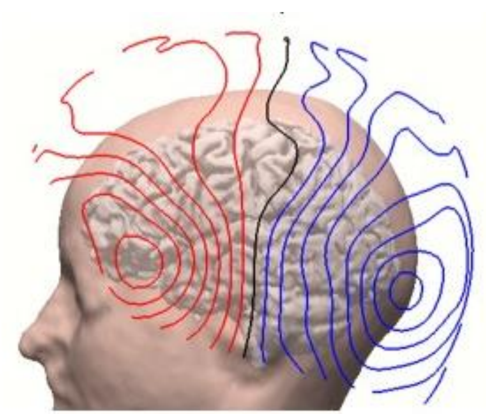
Recurrent Neural Network Language Model

- Learned on Harry Potter fan fiction database. (60 million words)
  - Nearest Neighbors of Harry:
    - James, Jinny, Lilly, Albus, Ron

- Model then "reads" Chapter 9 of *The Sorcerer's Stone* word by word, and produces vectors of interest.

Wehbe et al. 2014

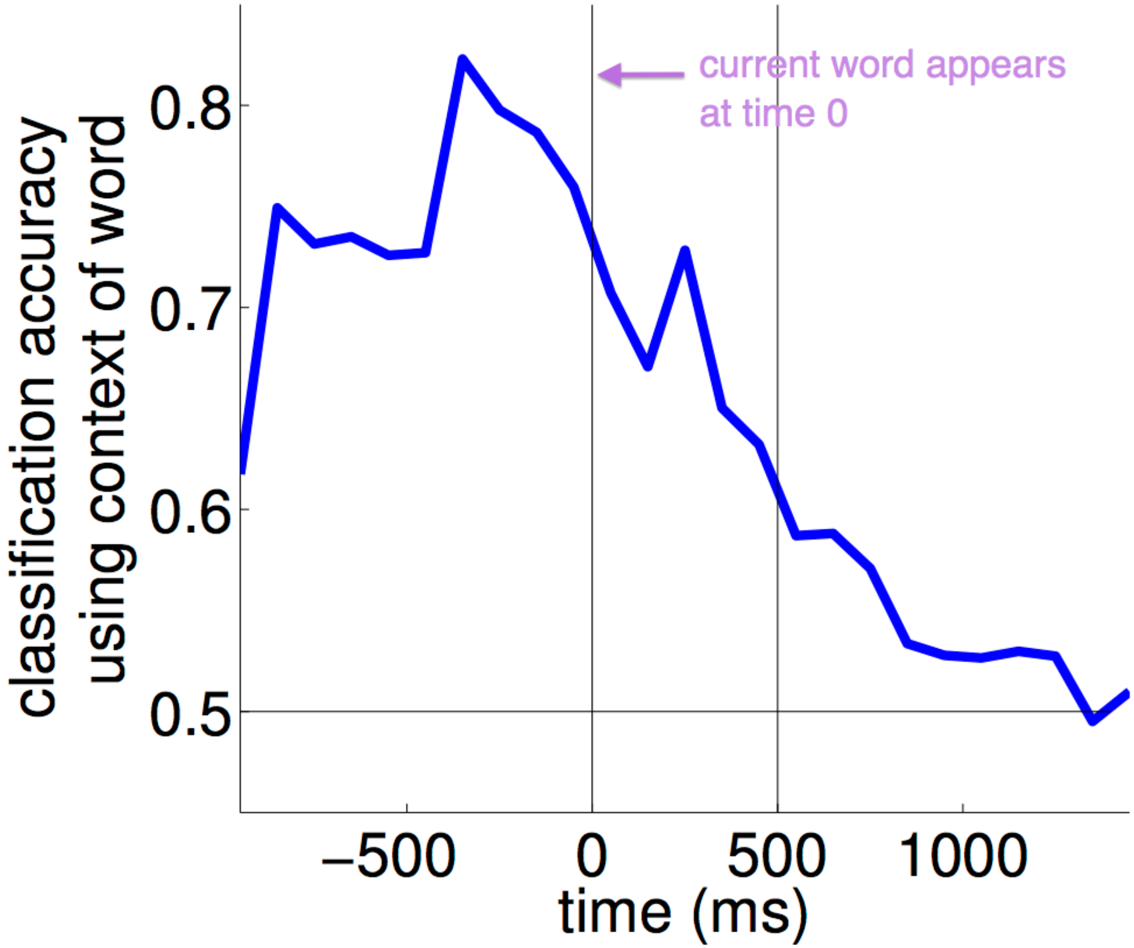




- Embedding of word w
- Context (before word w is seen)
- Probability of word w given context

Wehbe et al. 2014

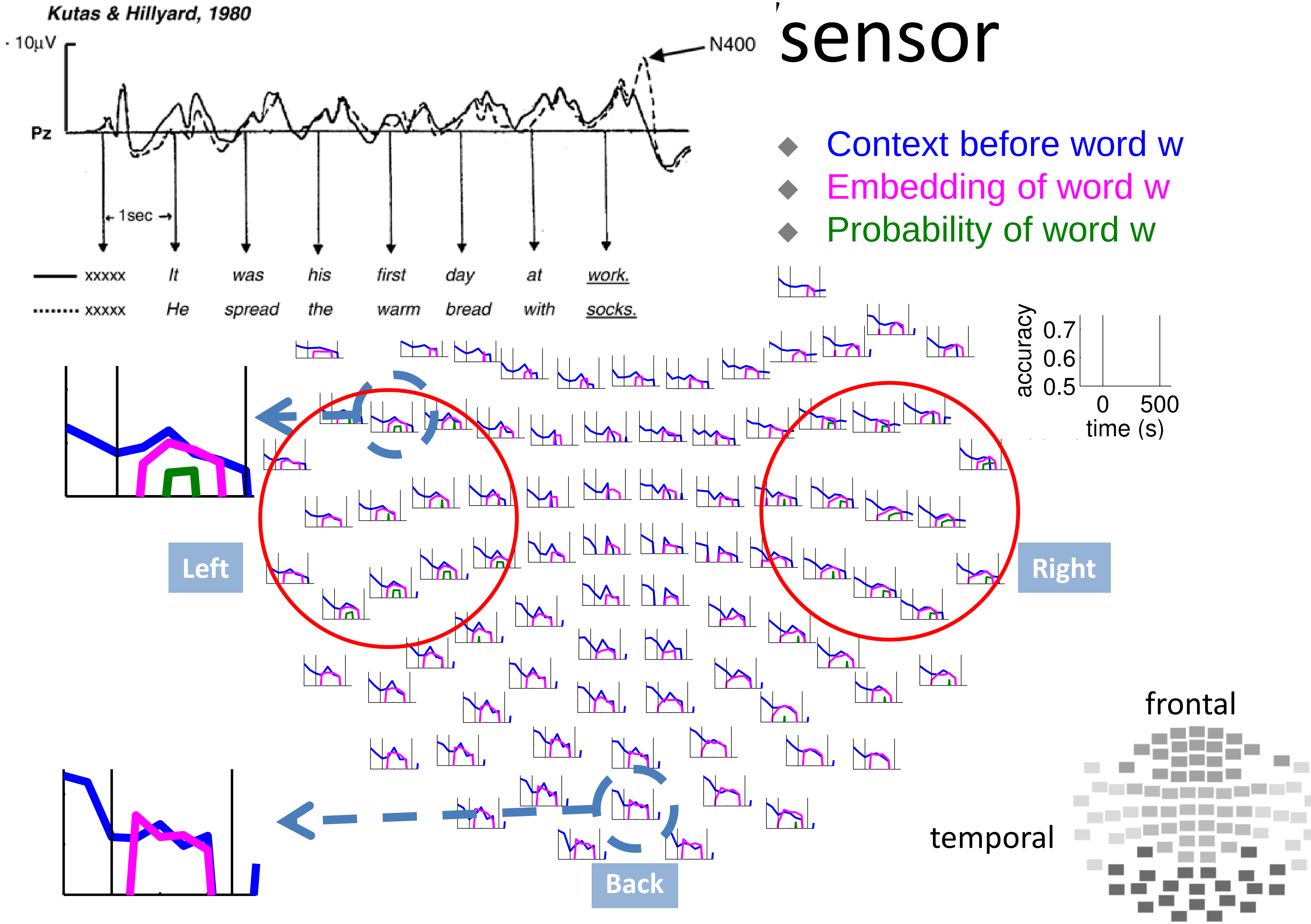
Results: Context vector by time



Wehbe et al. 2014

sensor

- Context before word w
- Embedding of word w
- Probability of word w



Wehbe et al. 2014

RNNs vs the Brain

- There is a relationship to the representations an RNN learns and information in the brain

- Caveat: is this actually what reading is?

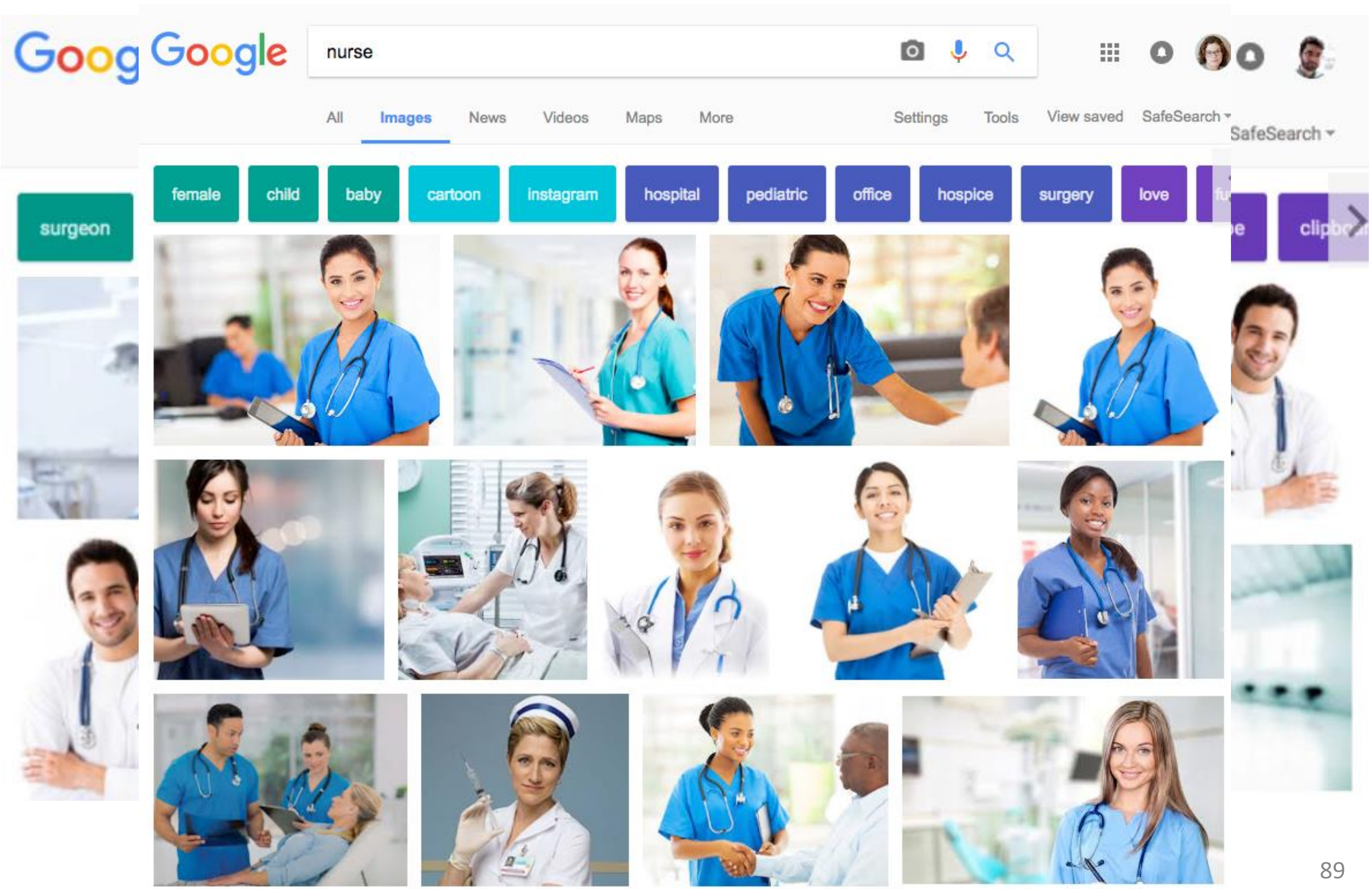
Robot Student Test (Goertzel)

- What would that require?
  - **perception**
    - e.g. read figures in books
  - **represent knowledge**
    - e.g. foxes are like wolves
  - **communication via language (written at least)**
    - e.g. writing essays for class
  - learn, reason, generalize, plan…

AI = Neural Networks?

- Of course not…

Bias in AI


Google
nurse
All
Images
News
Videos
Maps
More
Settings
Tools
View saved
SafeSearch
SafeSearch
female
child
baby
cartoon
instagram
hospital
pediatric
office
hospice
surgery
love
surgeon

Bias in AI

- Recall word vectors can solve analogy problems:

- e.g. Paris : France as Tokyo : x

  Word vectors will tell you that x = Japan.

- father : doctor as mother : x

  x = nurse
- man : computer programmer as woman : x

  x = homemaker

https://www.technologyreview.com/s/602025/how-vector-space-mathematics-reveals-the-hidden-sexism-in-language/

Bias in, Bias out?

- People have biases… should models have bias?

- Biases are harmful, **even deadly,** if perpetuated
  - Face recognition works better for white people
    - Increased mistaken identity among people of color
  - Software used to decide parole/sentencing had **elevated risk assessments** for people of color that **did not correlate to actual recidivism rates**

https://www.theatlantic.com/technology/archive/2016/04/the-underlying-bias-of-facial-recognition-systems/476991/
https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Bias in, Bias out?

- Our models may even exaggerate bias!
- “For example, the activity cooking is over 33% more likely to involve females than males in a training set, and a trained model further **amplifies the disparity to 68%** at test time.”

Zhao (2017) https://arxiv.org/pdf/1707.09457.pdf

Bias in Artificial Intelligence

- Combating bias in AI is an open research question

  yearly conference **FATML**

  http://www.fatml.org/

See also:

https://qz.com/1064035/google-goog-explains-how-artificial-intelligence-becomes-biased-against-women-and-minorities/ (lots of good links at the bottom)

https://cdt.org/issue/privacy-data/digital-decisions/

These slides available at:
http://web.uvic.ca/~afyshe/talks/ccn_AI.pdf

Thanks!

Alona Fyshe
afyshe@uvic.ca
http://web.uvic.ca/~afyshe/
http://brainyblog.net

NSERC CRSNG

CIFAR CANADIAN INSTITUTE FOR ADVANCED RESEARCH