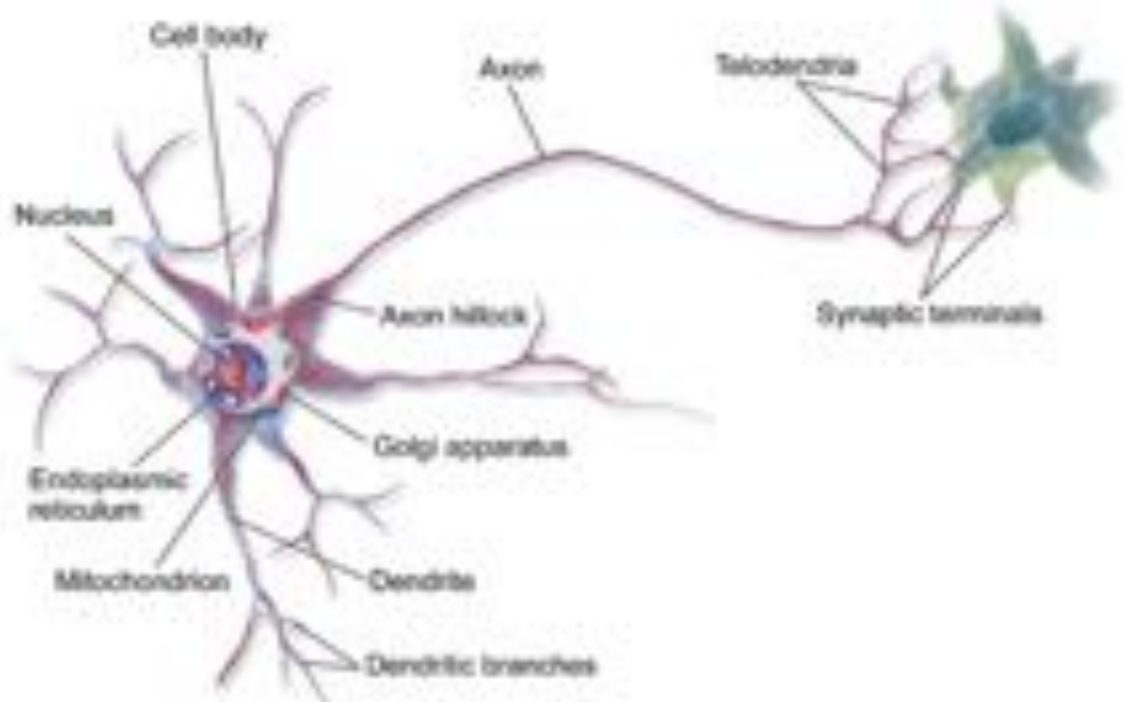
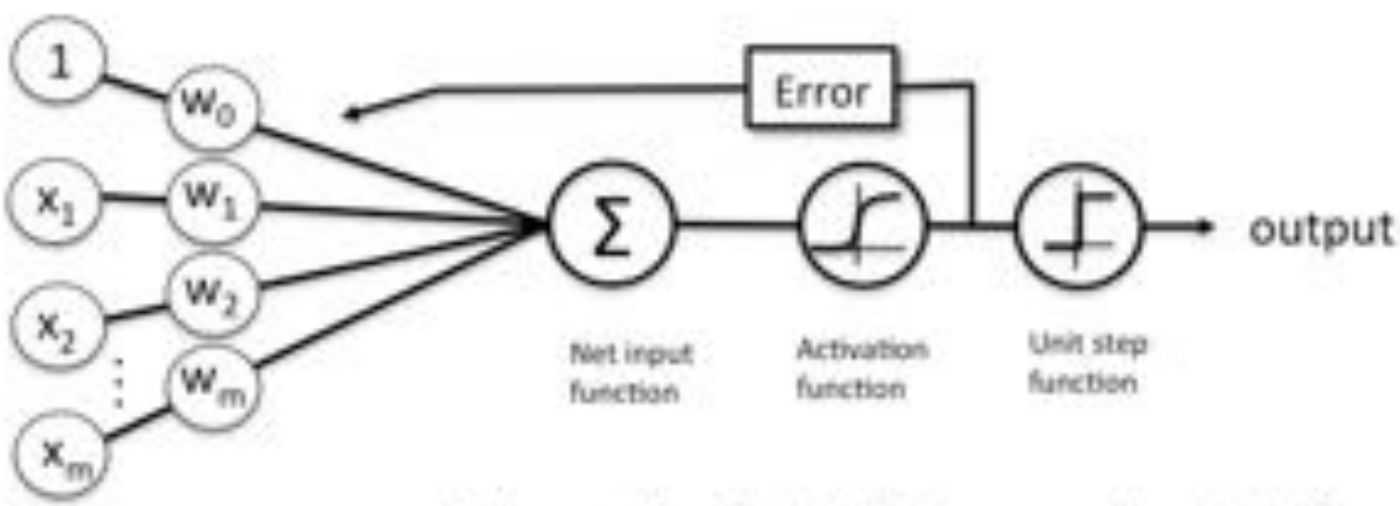
Introduction to Deep Learning

Ashish Mahabal
Center for Data-Driven Discovery, Caltech
LSST Transients and Variable Stars Co-chair
26 Jan 2017
LSST Data Science Fellowship Program





Schematic of a logistic regression classifier.

Real and artificial neurons

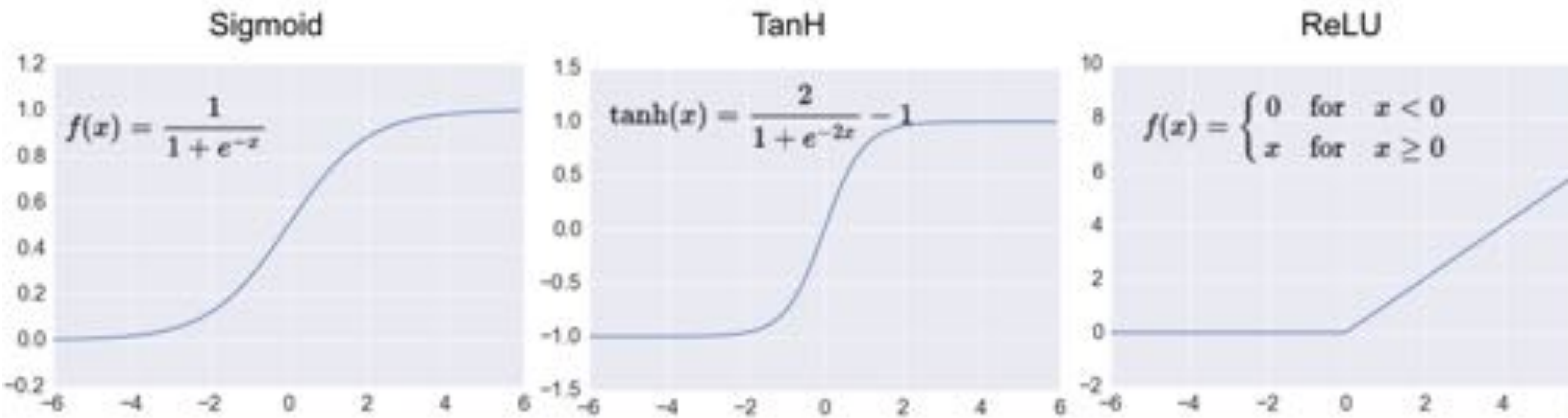
By BruceBlaus - Own work, CC BY 3.0, https://commons.wikimedia.org/w/index.php?curid=28761830
Quora: Sebastian Raschka

Softmax Regression

- This equation gives the probability of each classification
- Softmax regression first add up evidence of input being in a certain class and convert evidence into probabilities.
- There are 3 parts to this equation:
    1. Add weights (Wx)
    2. Add biases
    3. Softmax

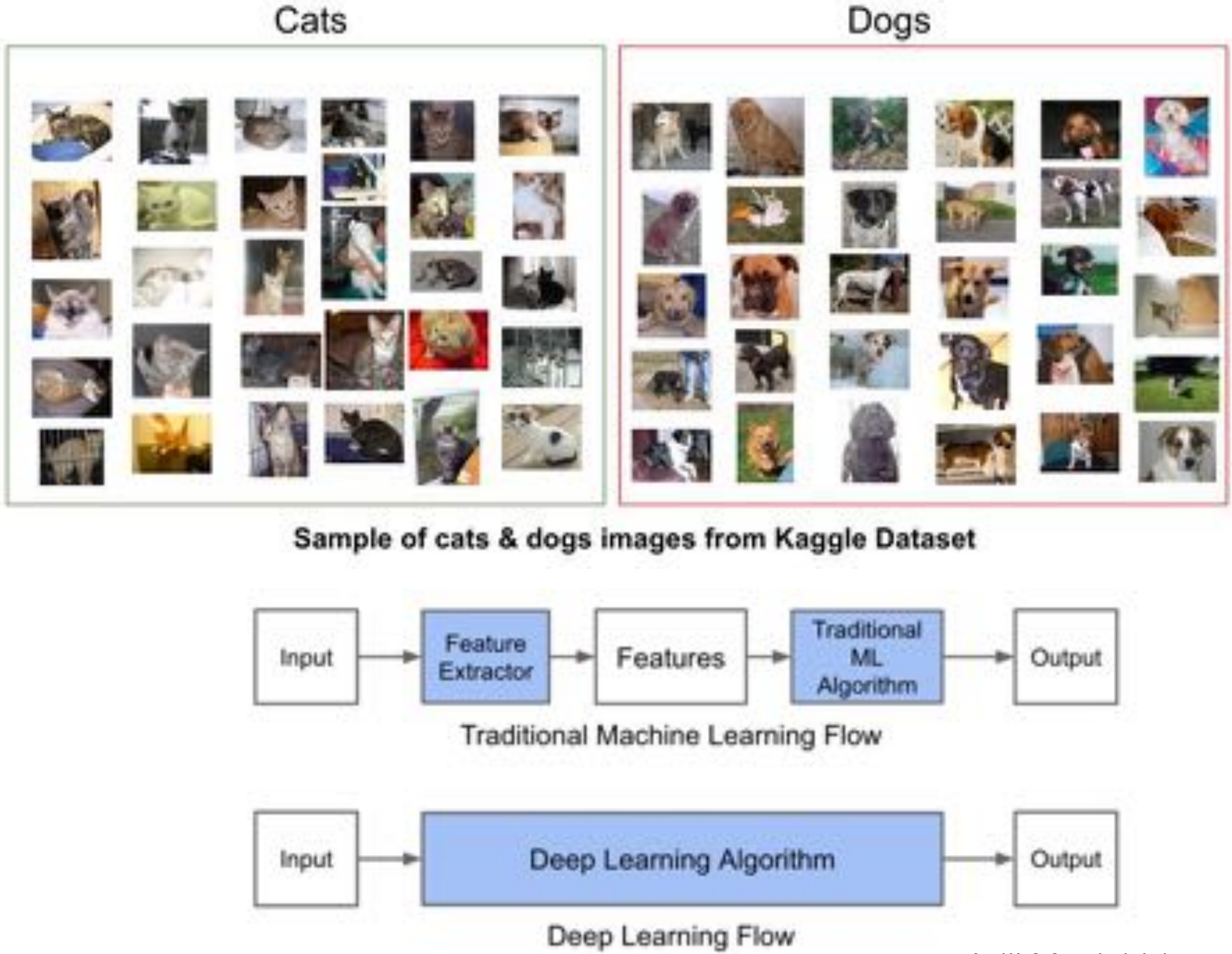
$$y = \text{softmax}(Wx + b)$$

Rectified Linear Unit



Activation

SoftPlus, Sigmoid, …
Remapping/reshaping



Adil Moujahid

Sample of cats & dogs images from Kaggle Dataset

Promise:
Works better

Traditional Machine Learning Flow

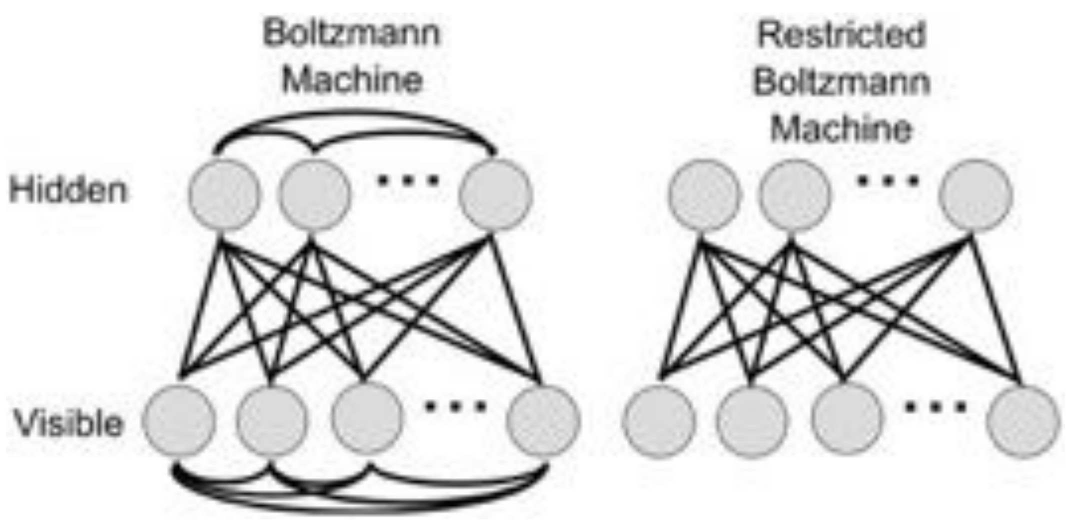
Pitfall:
Blacker box

Deep Learning Flow

Adil Moujahid

Deep Learning

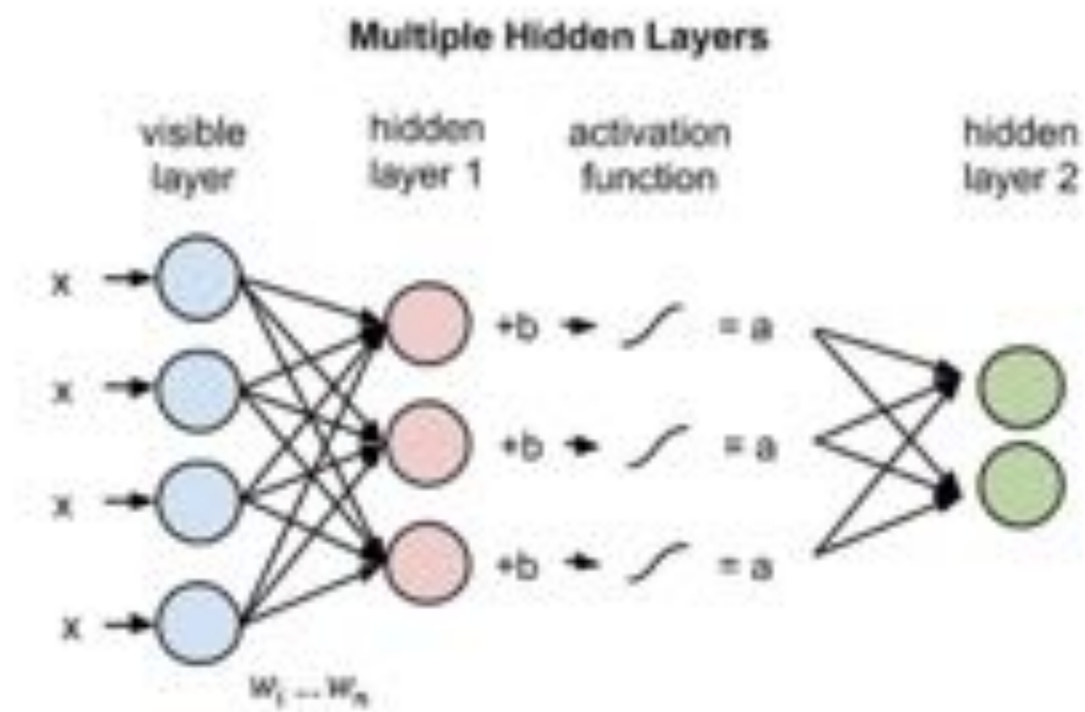
- Neural Networks in a new garb (but it works)



Restricted Boltzman Machines (RBMs)

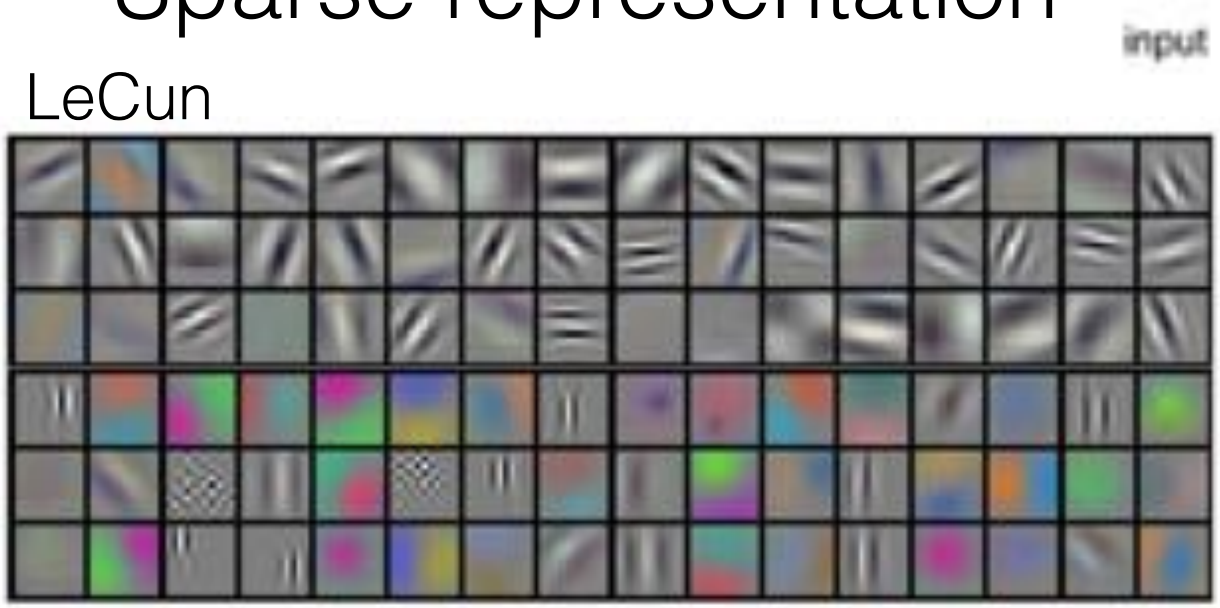
Hinton

Hidden layers can not interconnect



Sparse representation

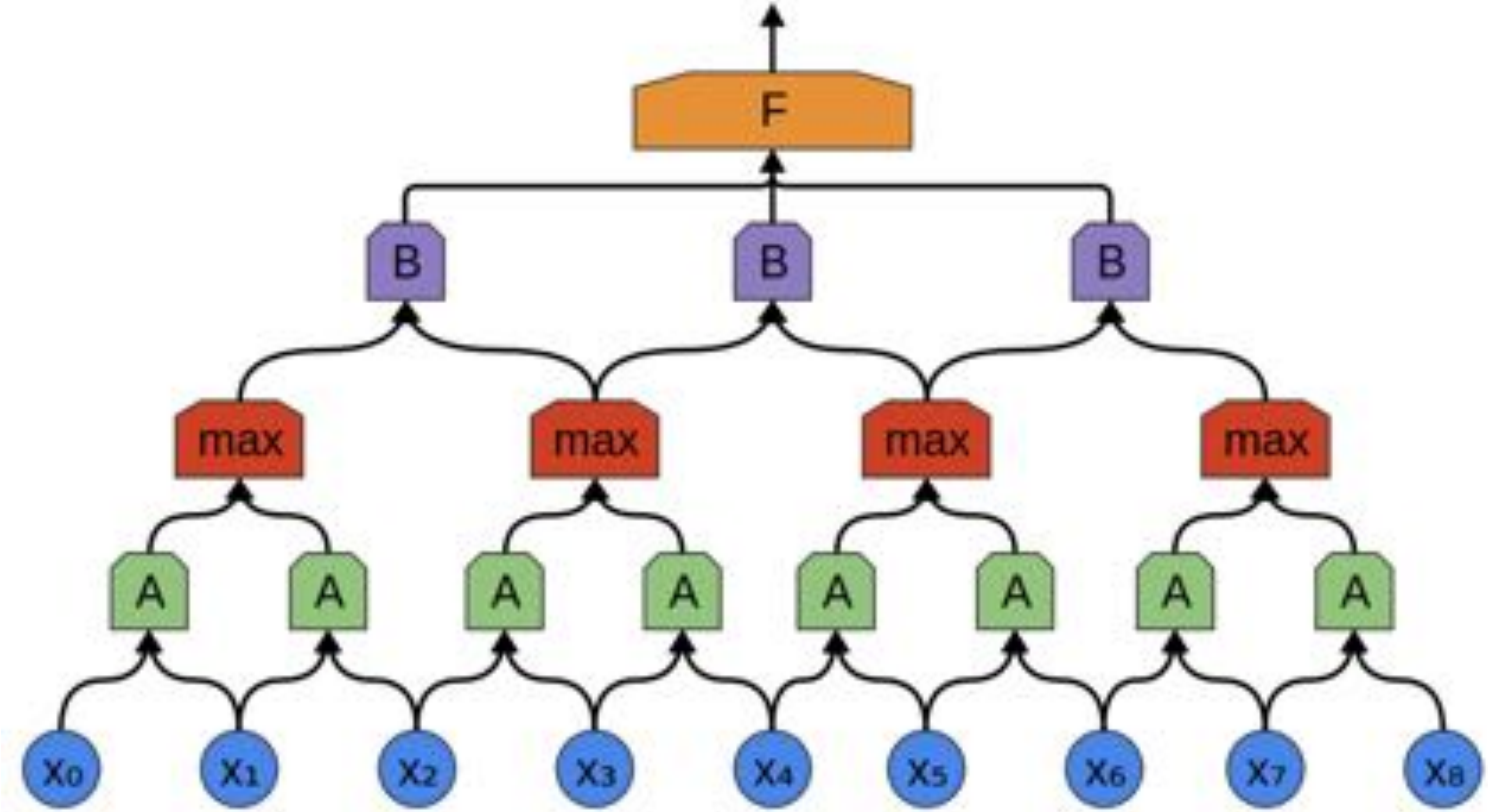
LeCun



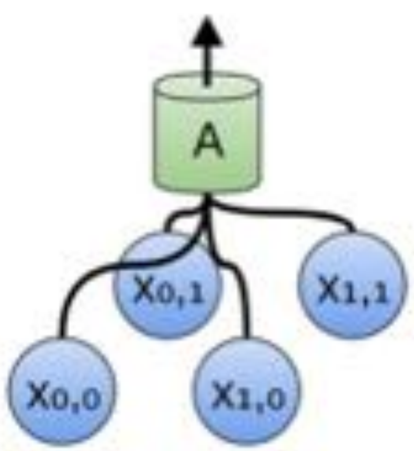
Simple FeedForward

Joint probability: p(x|a) and p(a|x) expressed as the shared weights

Krizehvsky et al. 2012



conv + pool + conv + connected

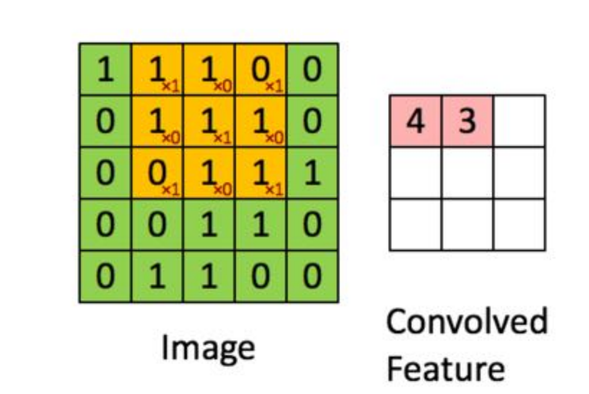


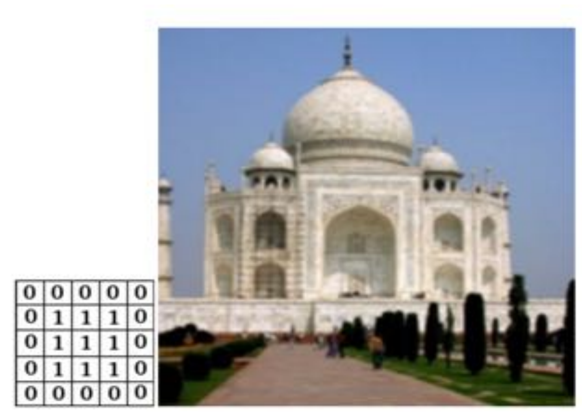
http://colah.github.io/posts/2014-07-Conv-Nets-Modular/

2D version of convolution

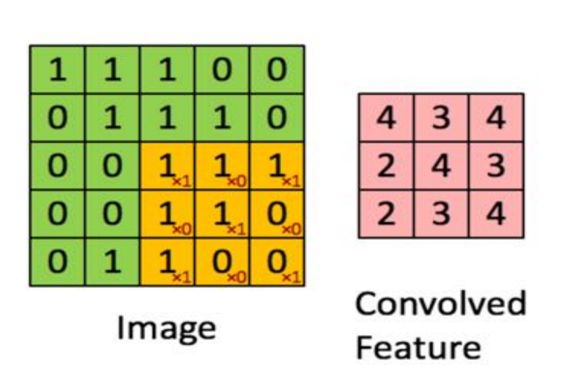
Enhancing angled edges

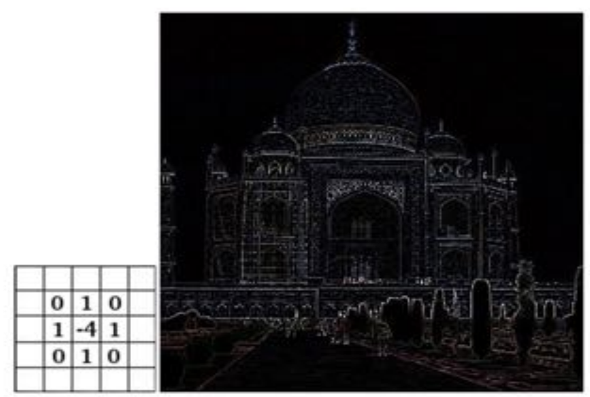
Here’s an example of how to make a convolution layer with 3x3 filter.
You run across the matrix to get convolved feature of 3x3.





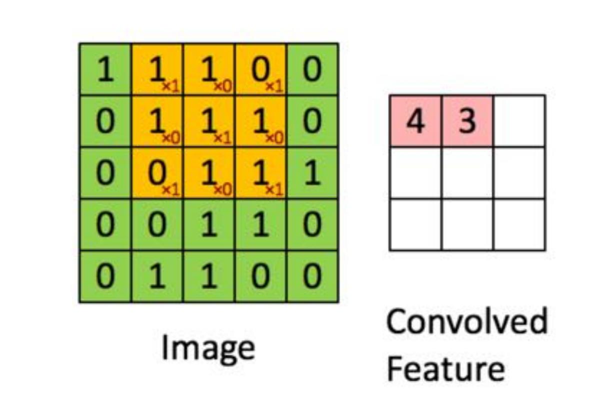
- This will be the resulting convolution layer once matrix is complete. In this example, the layer will detect edges.

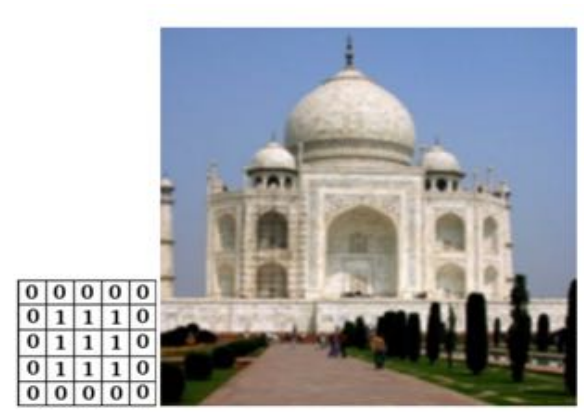




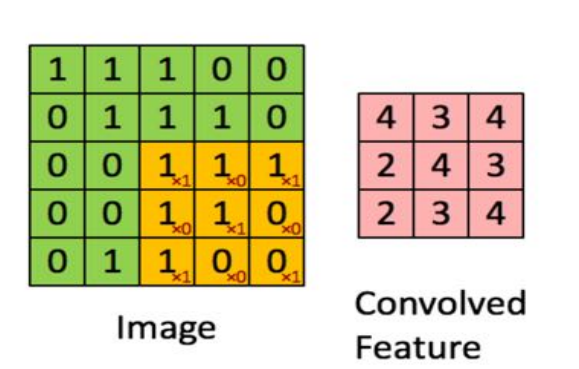
Enhancing angled edges

Here's an example of how to make a convolution layer with 3x3 filter.
You run across the matrix to get convolved feature of 3x3.

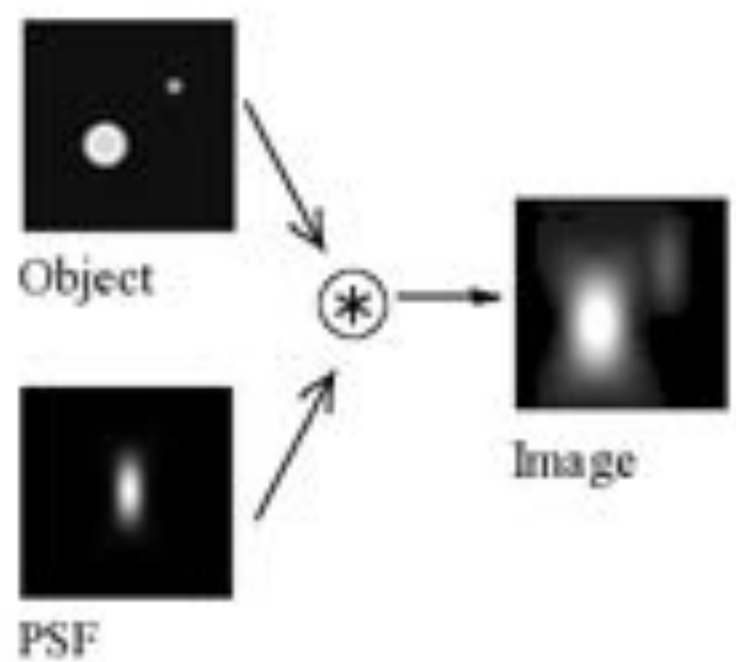




- This will be the resulting convolution layer once matrix is complete. In this example, the layer will detect edges.



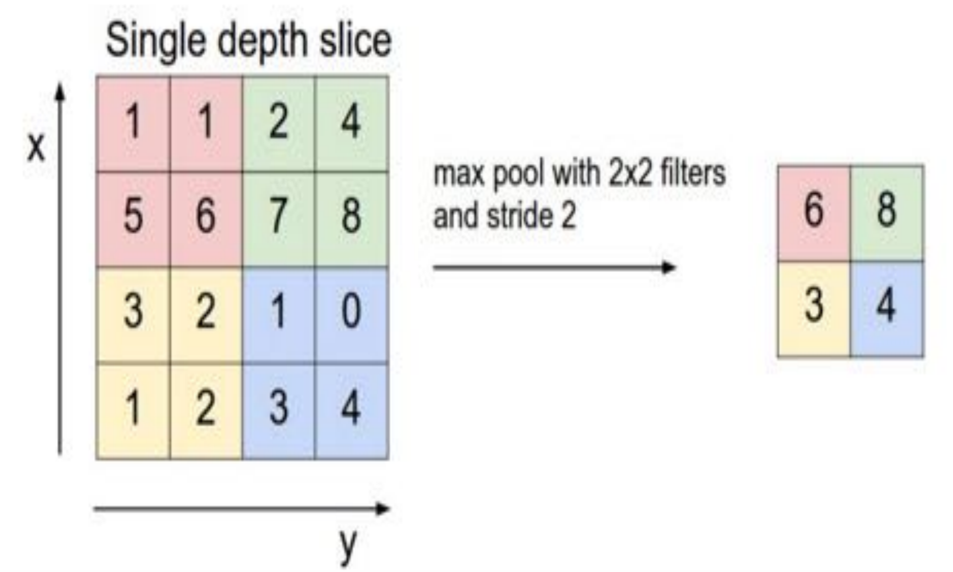
Convolution in Astronomy



Pooling Layers (max pool)

Pooling layers are applied after convolution layers. This is a subsample of your convolution layer.

- Fixed size output matrix
- Reduces output dimensionality with the most salient information

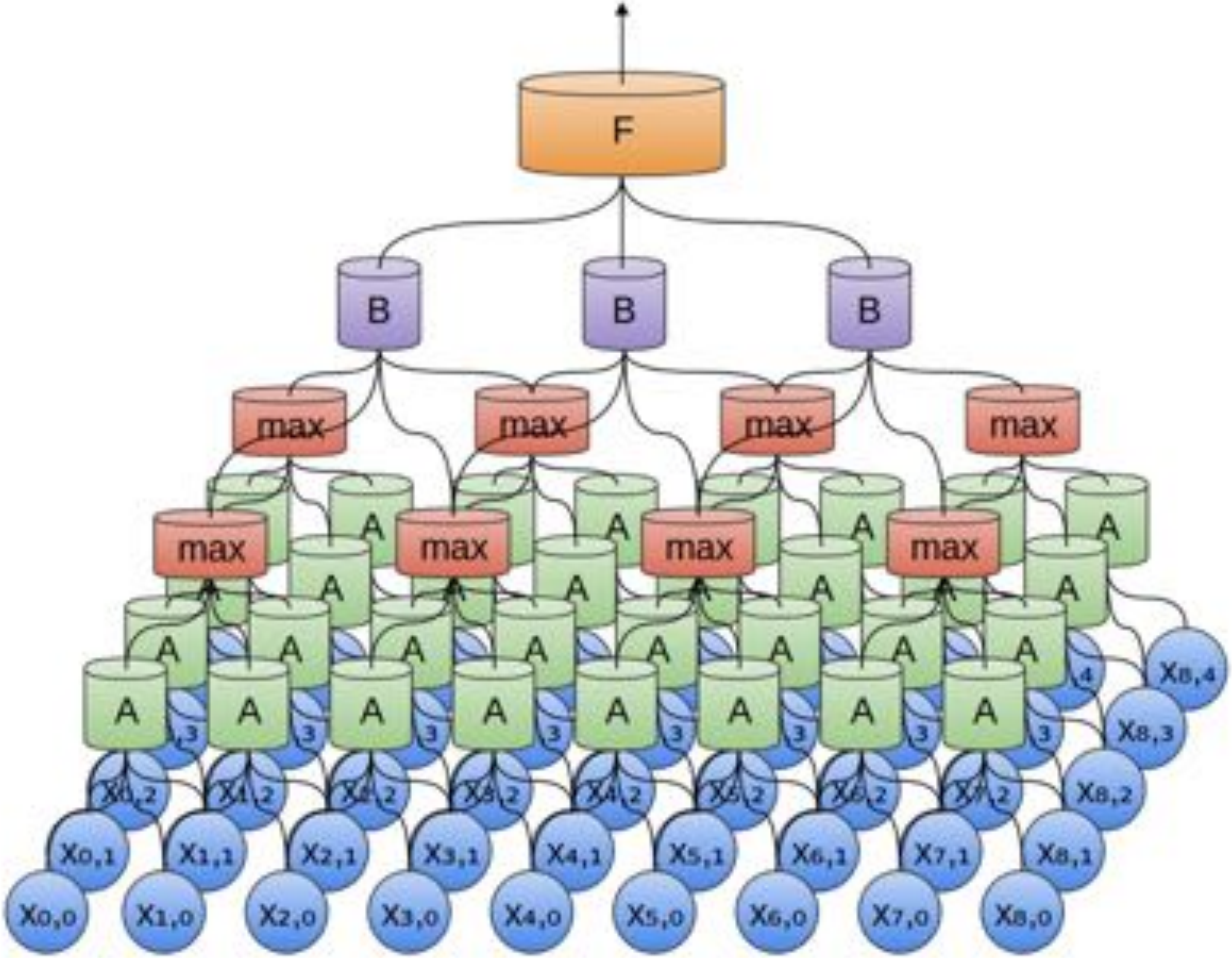


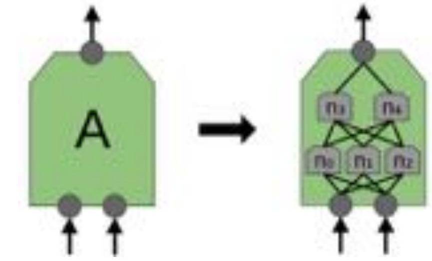
Hyperparameters

Our model has a depth of 3. Pooling blocks are 2x2, with stride of 1 and 0 padded for our model.

Three hyperparameters control the size of the output volume of the convolutional layer: the **depth, stride** and **zero-padding**.

- **Depth** of convolution layers. For example, if the first Convolutional Layer takes the raw image as input, then another may activate edges, or color.
- **Stride** When the stride is 1, a new depth column of neurons is allocated to spatial positions only 1 spatial unit apart.
- **Zero padding** provides control of the output volume spatial size.





http://colah.github.io/posts/2014-07-Conv-Nets-Modular/

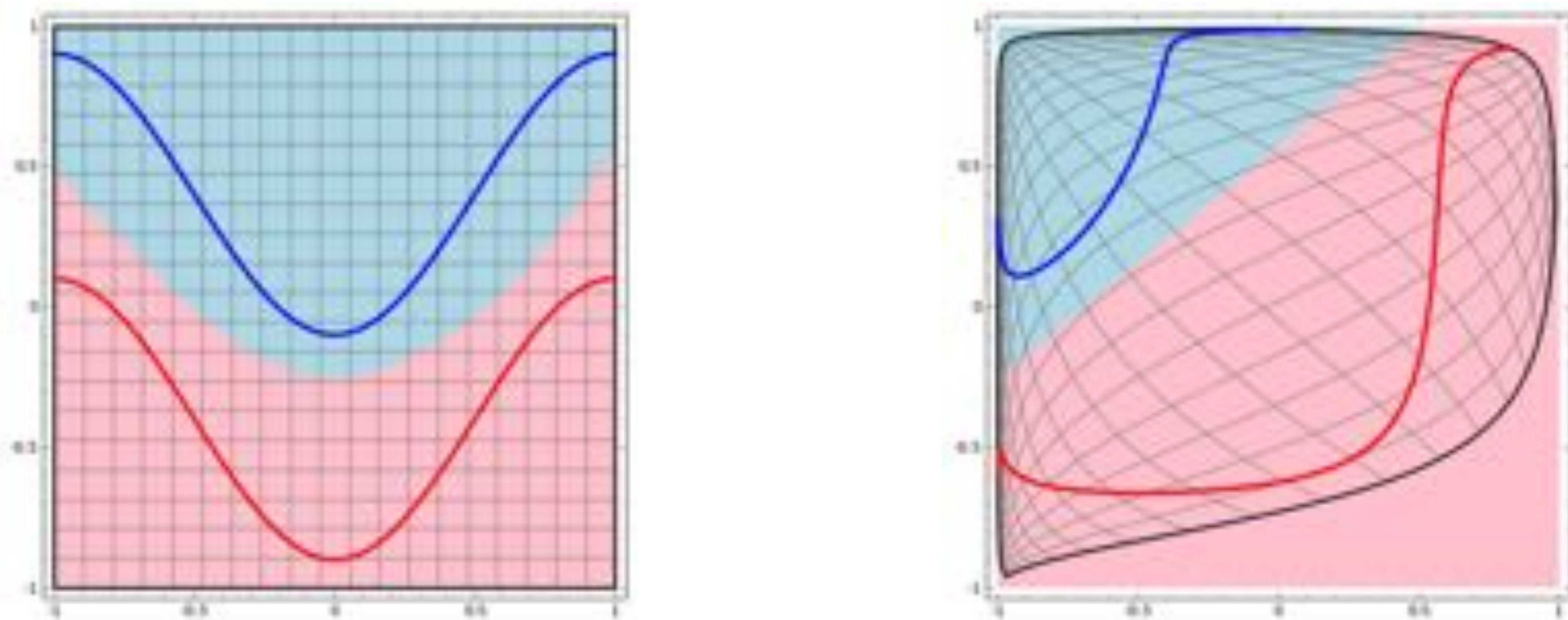
Final layers are fully connected

Several libraries available

- Caffe: http://caffe.berkeleyvision.org/
- Tensorflow: https://www.tensorflow.org/
- Theano: http://deeplearning.net/software/theano/

abstractions, instant gratification

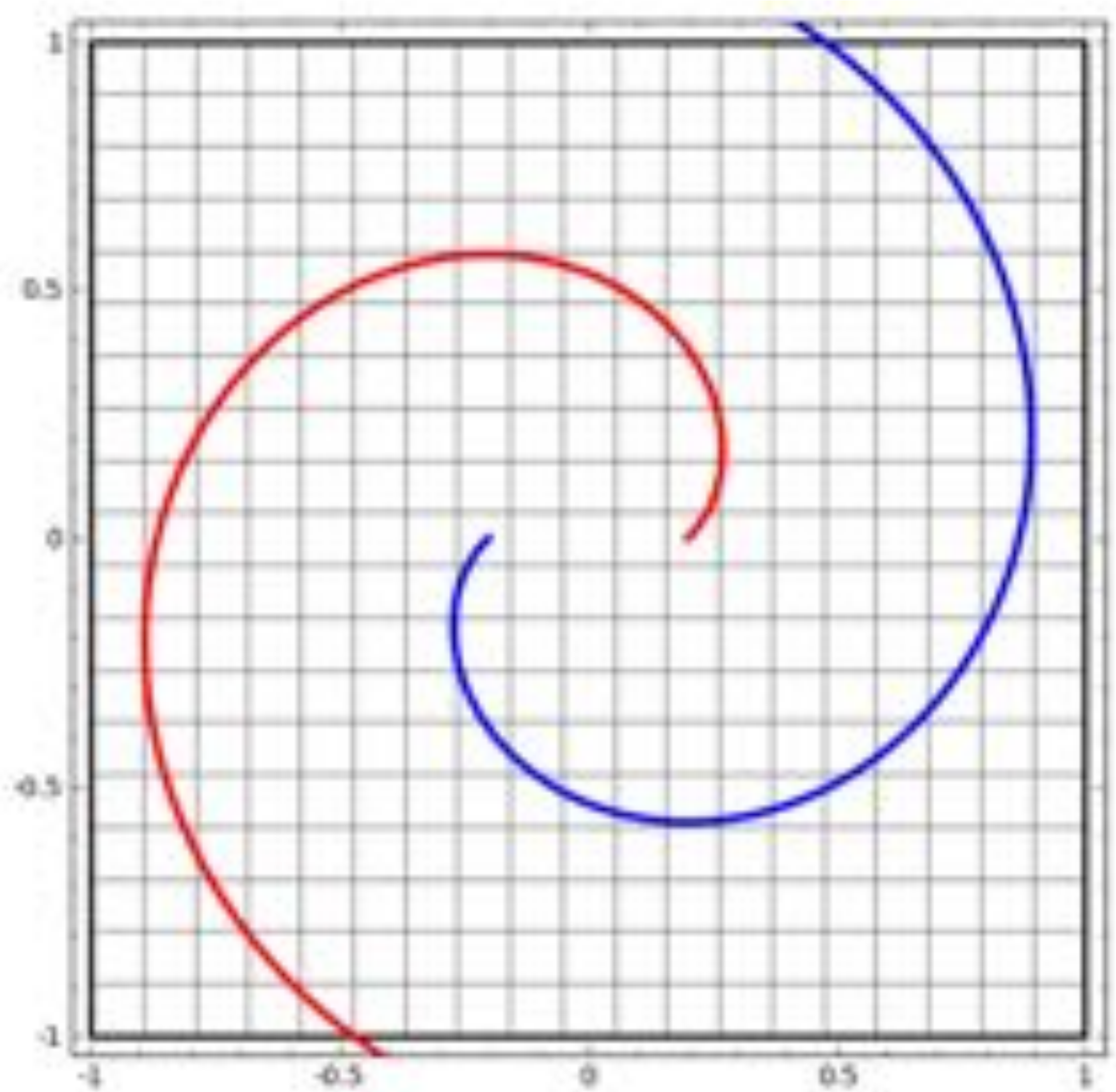
Tensorflow example (notebook)

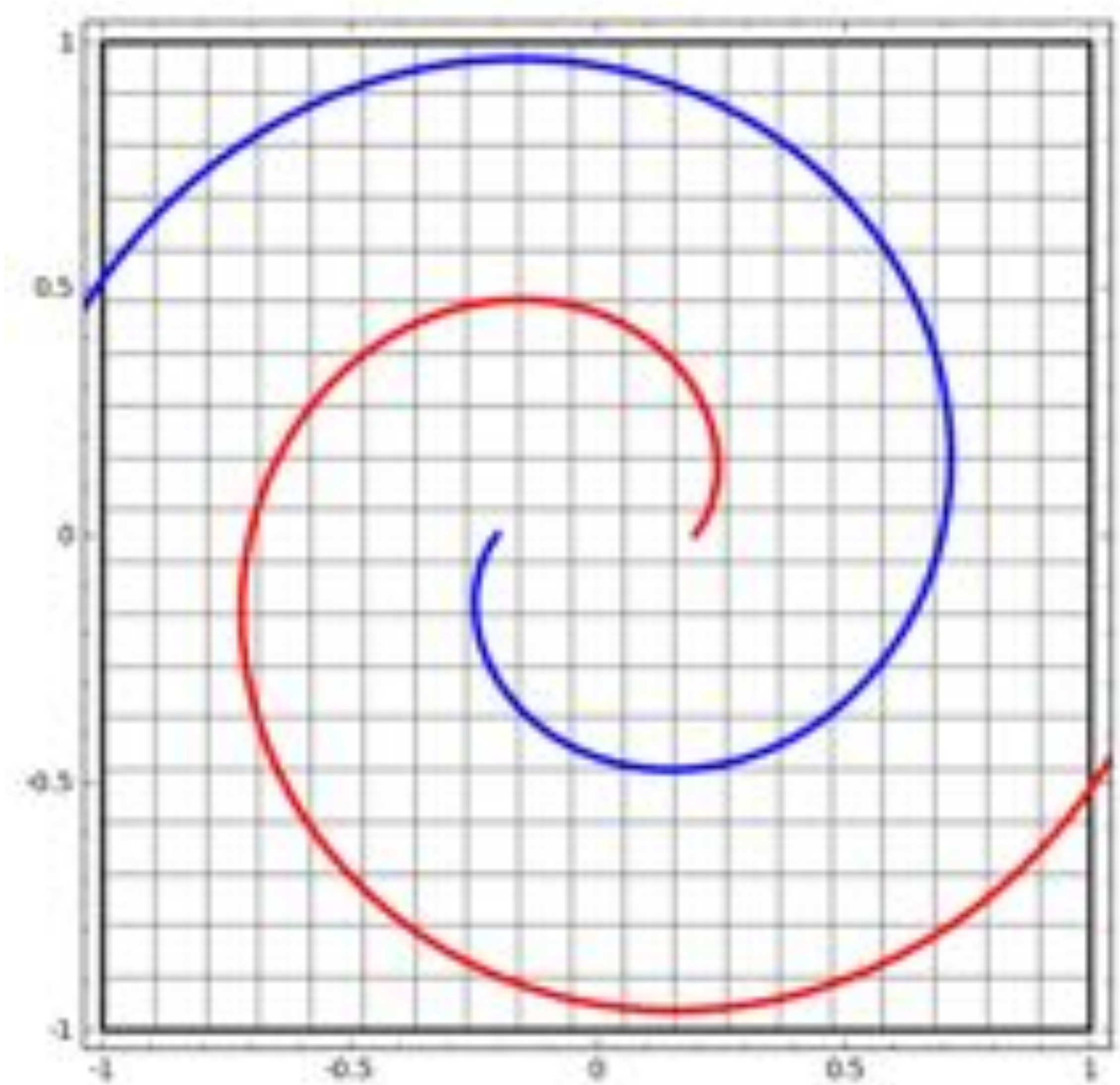


Mapping in order to linearly separate clusters

http://colah.github.io/posts/2014-03-NN-Manifolds-Topology/

Disentangling with multiple layers



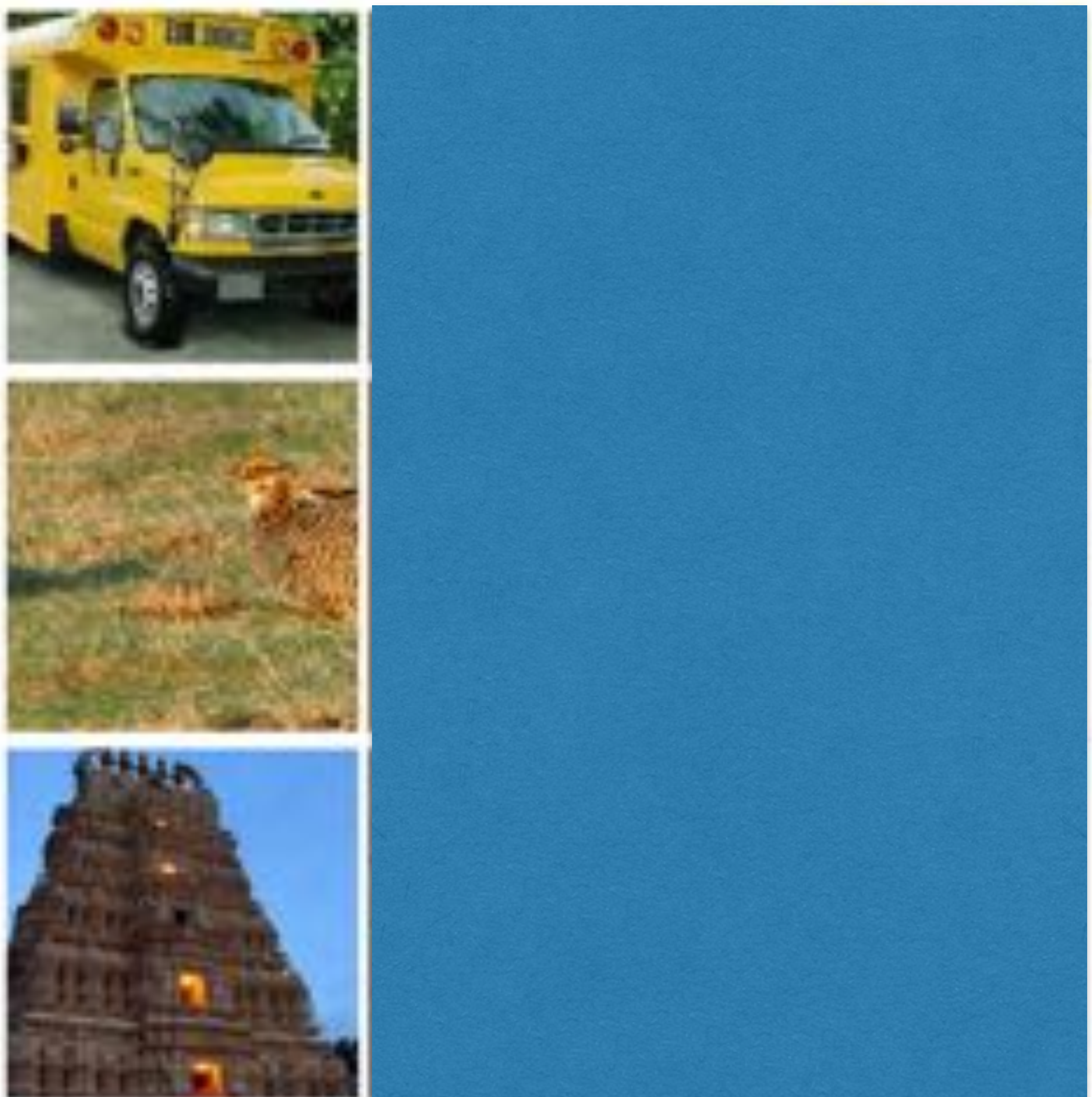


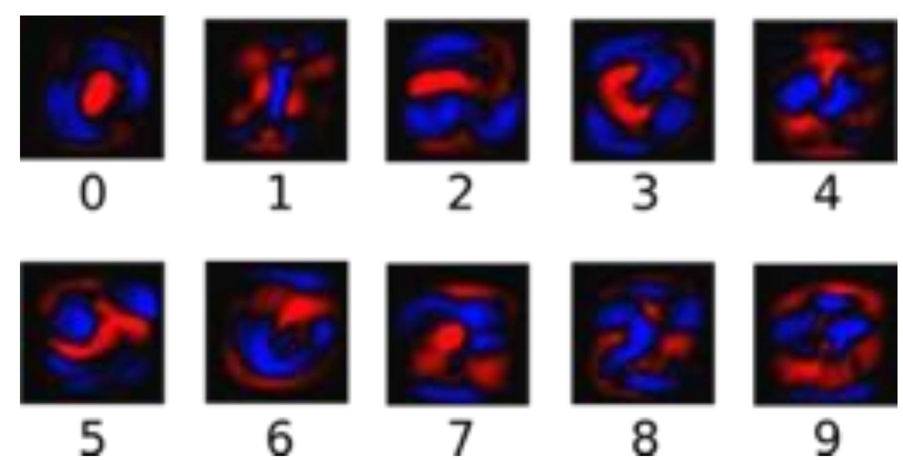
http://colah.github.io/posts/2014-03-NN-Manifolds-Topology/

What bird is that?

or: what features is my deep network using?

Including adversarial examples during training

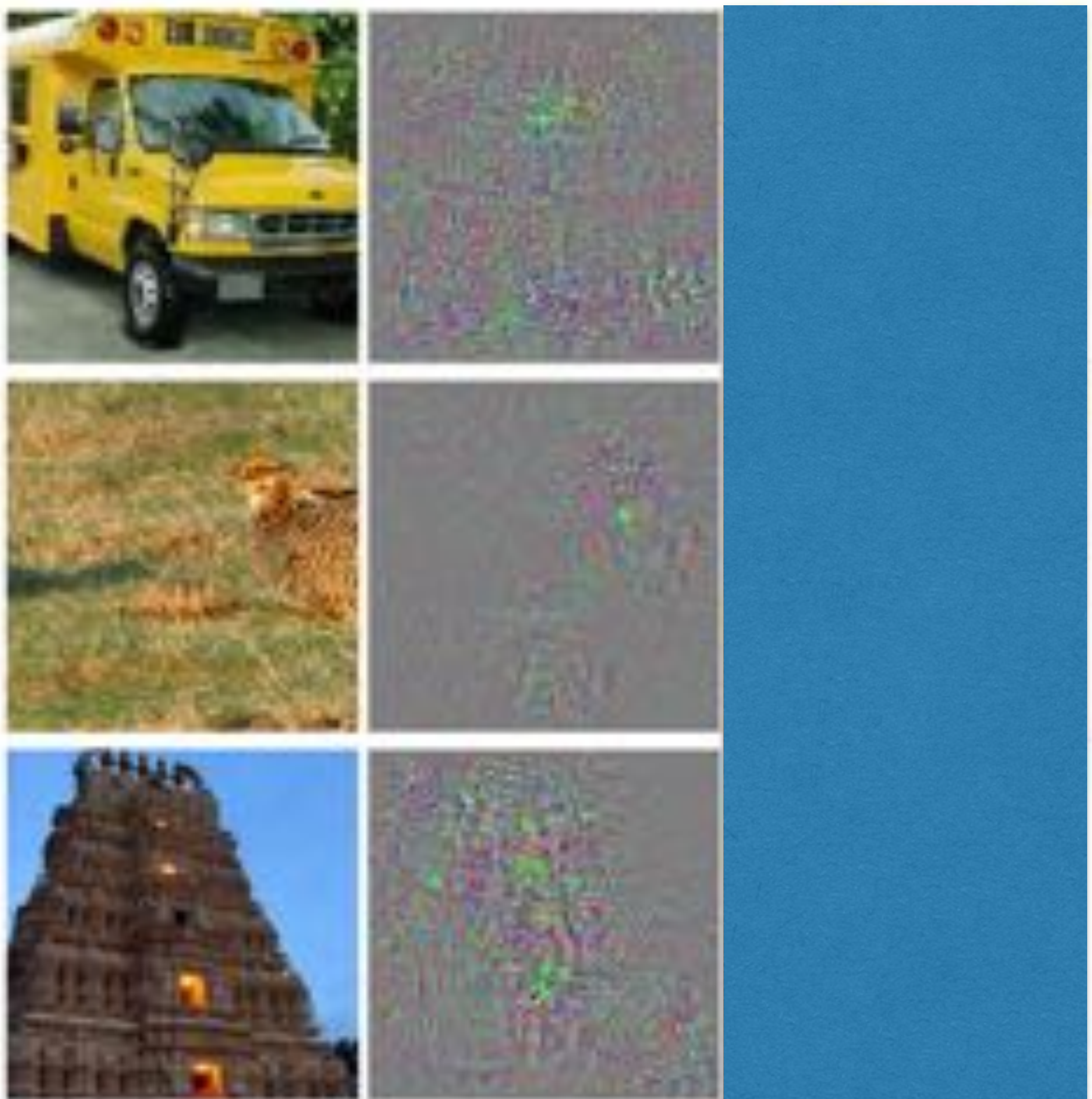


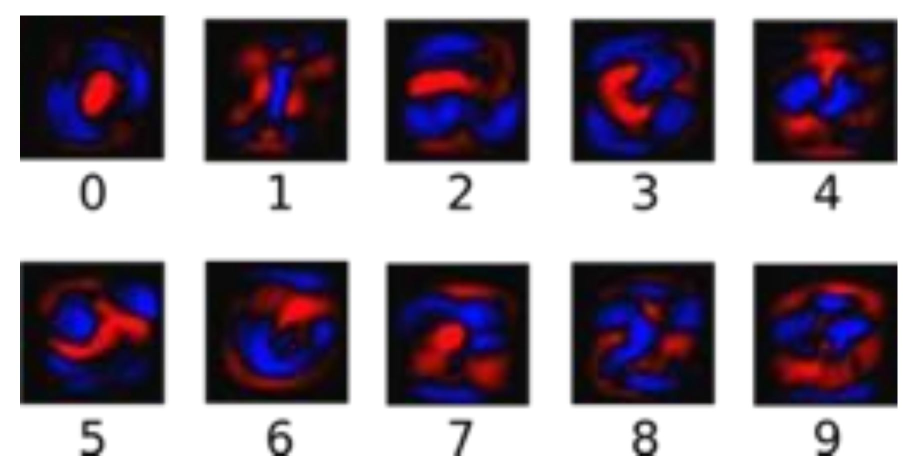


MNIST digits significance map

https://arxiv.org/pdf/1312.6199v4.pdf

Including adversarial examples during training



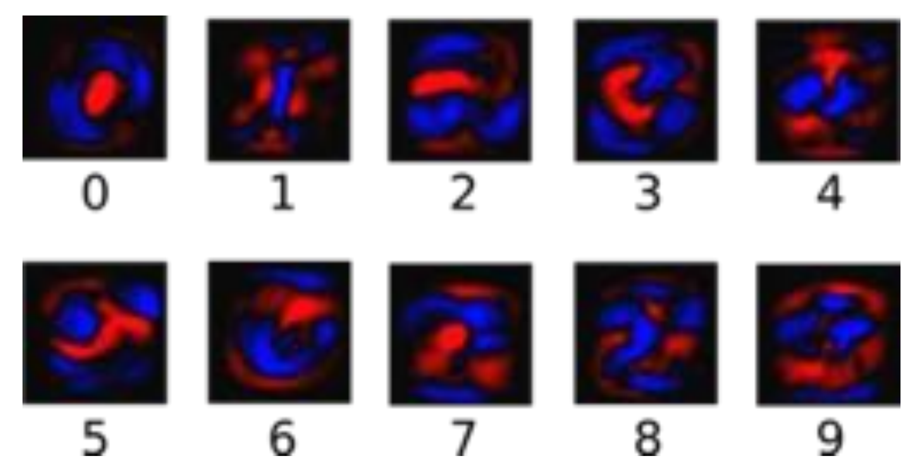


MNIST digits
significance map

https://arxiv.org/pdf/1312.6199v4.pdf

Including adversarial examples during training

The images in the left most column are correctly classified examples. The middle column represents the distortion between the left and right images. The images in the right most column are predicted to be of the class ostrich! Even though the difference between the images on the left and right is imperceptible to humans, the ConvNet makes drastic errors in classification.

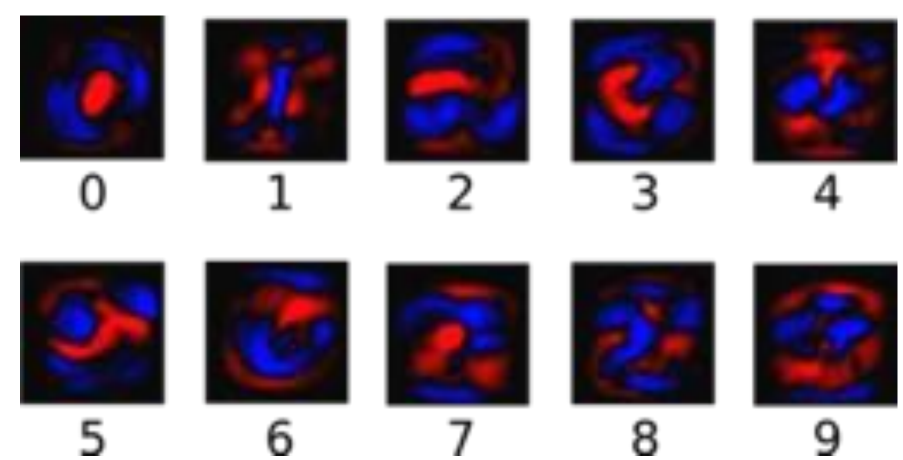


MNIST digits significance map

https://arxiv.org/pdf/1312.6199v4.pdf

Including adversarial examples during training

The images in the left most column are correctly classified examples. The middle column represents the distortion between the left and right images. The images in the right most column are predicted to be of the class ostrich! Even though the difference between the images on the left and right is imperceptible to humans, the ConvNet makes drastic errors in classification.



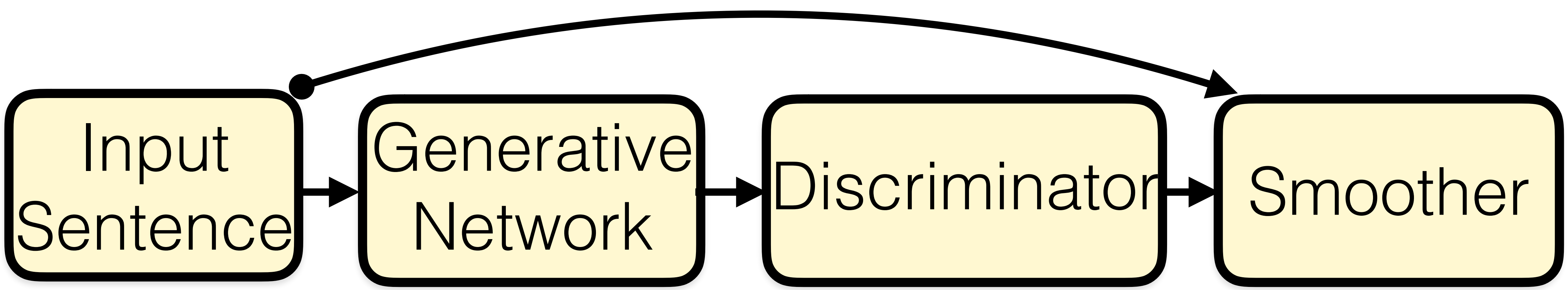
MNIST digits significance map

https://arxiv.org/pdf/1312.6199v4.pdf

Pitfall: Overlearning

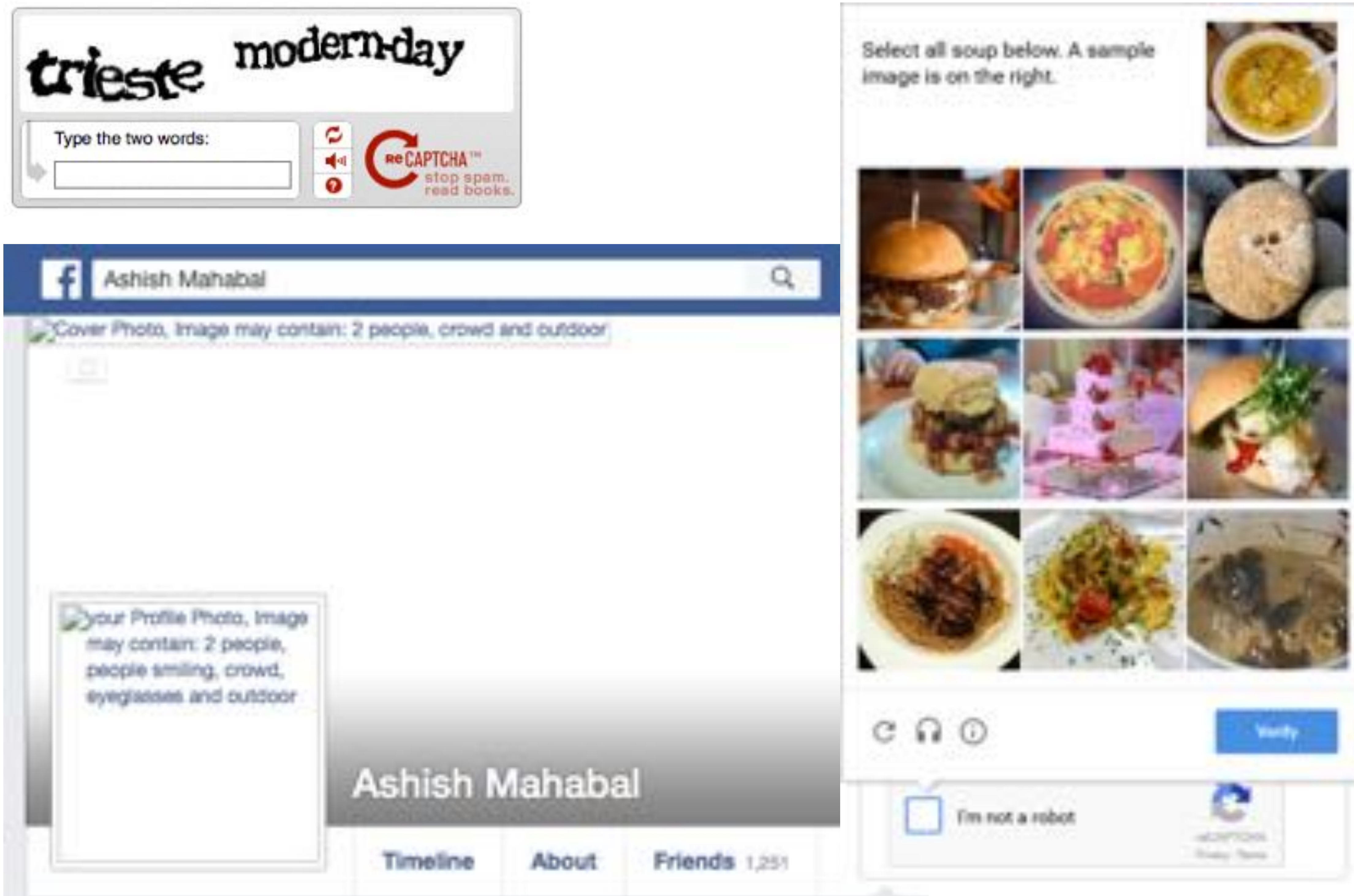
Generative Adversarial Networks

Zhang et al. 2016



Reality check

Labels are everywhere



Laundry list for image archives

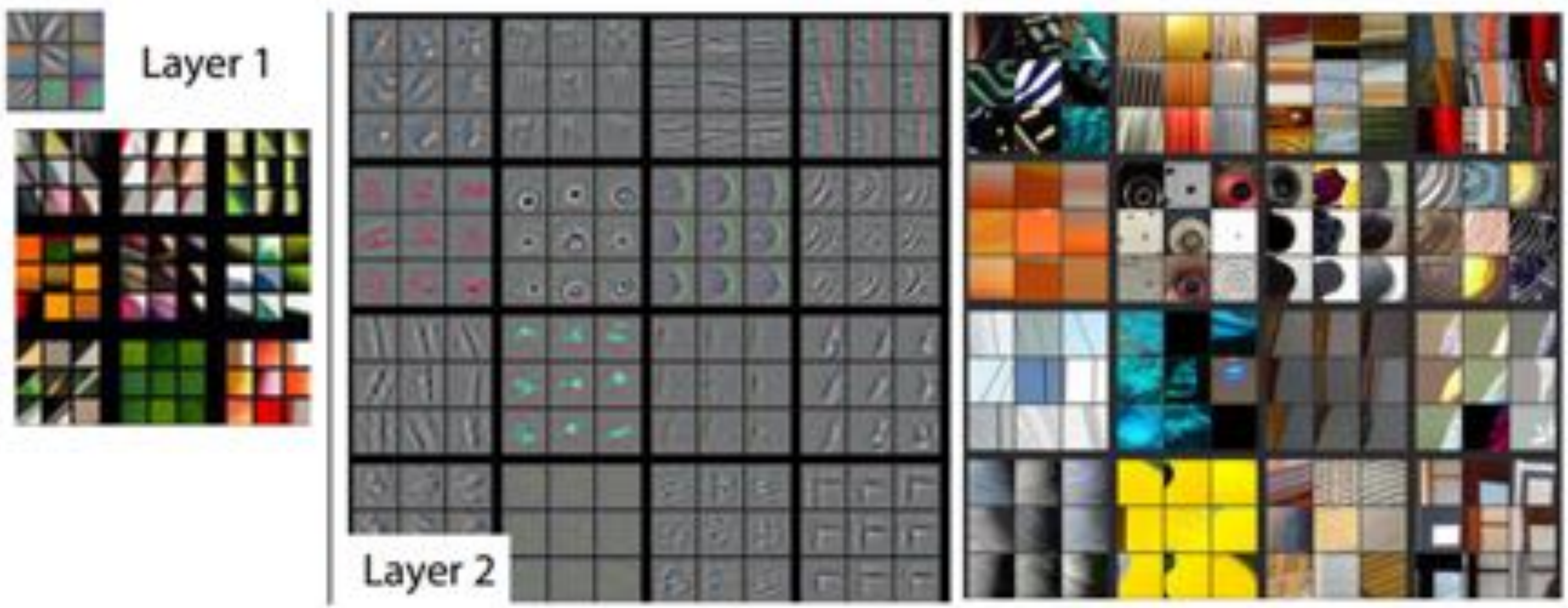
- Large sets
- Labelled data
- Metadata (**CDEs!**)
- Peripheral data
- Balanced datasets

https://arxiv.org/pdf/1412.2306v2.pdf

Some GAN refs

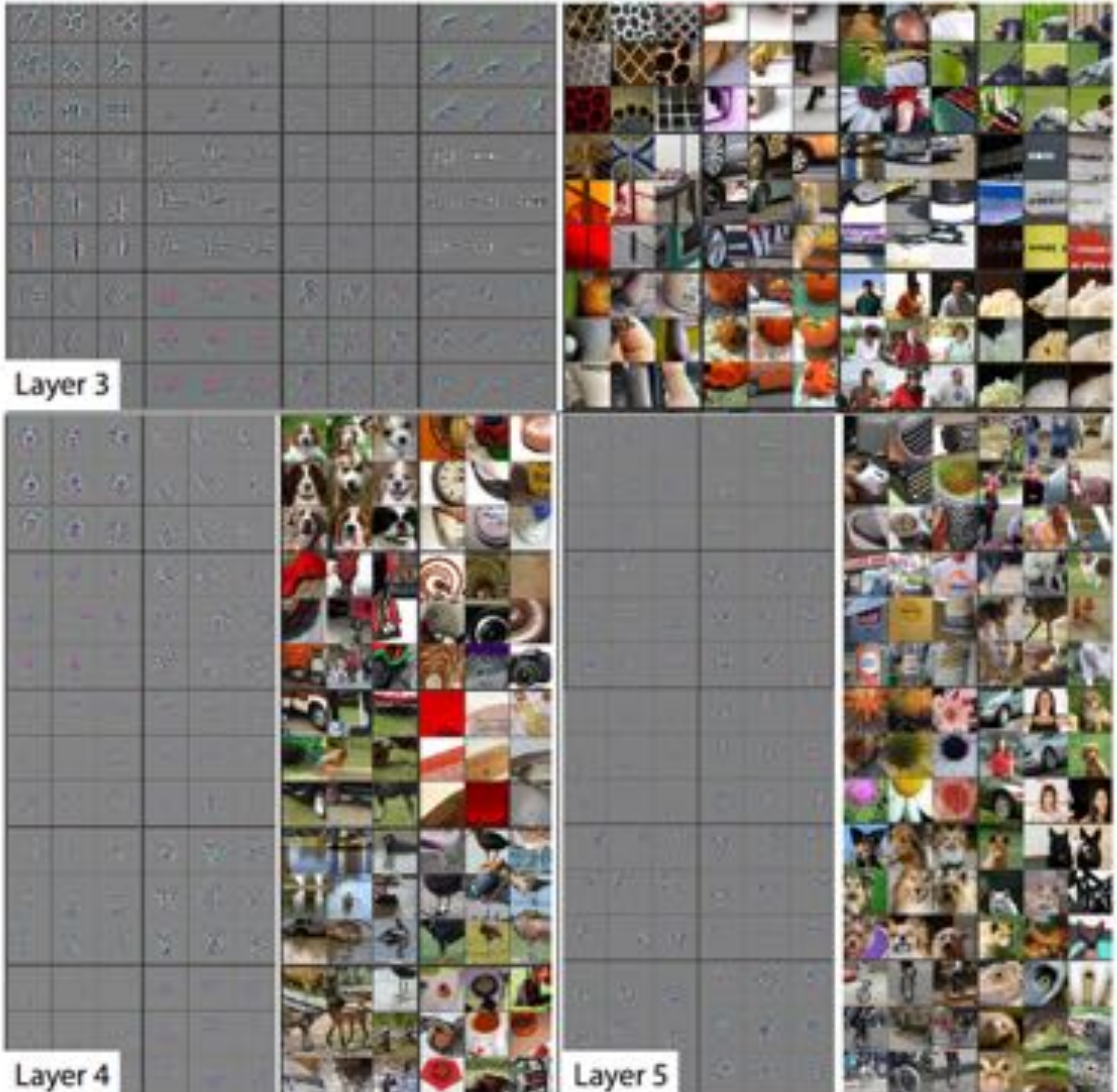
- Denton et al. “Deep Generative Image Models using a Laplacian Pyramid of Adversarial Networks” (NIPS 2015) : https://scholar.google.com/citat...
- Radford et al. “Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks” (ICLR 2015): https://scholar.google.com/citat...
- Mathieu et al. “Deep multi-scale video prediction beyond mean square error” : https://scholar.google.com/citat...

deconvnets

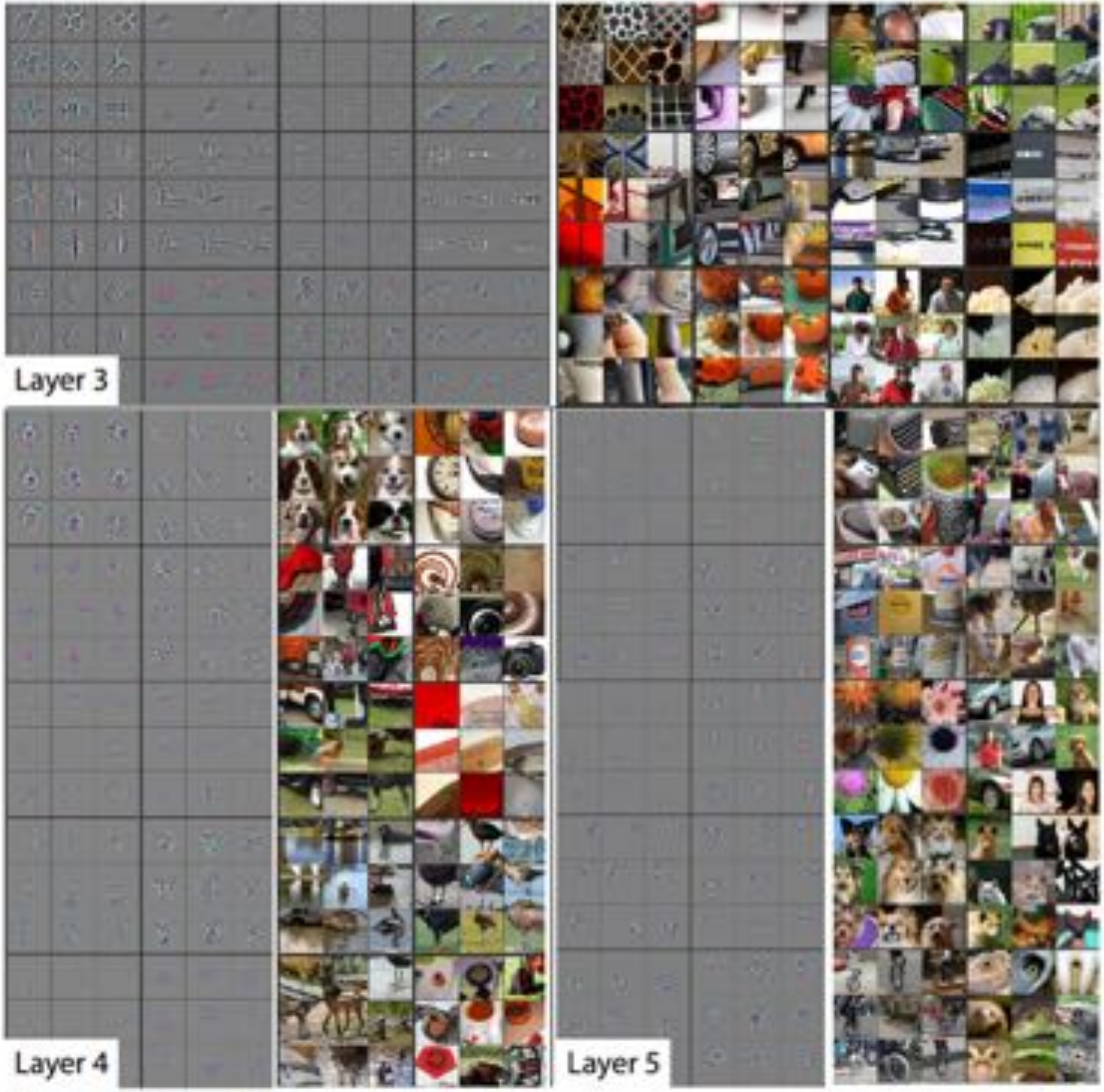


Visualizations of Layer 1 and 2. Each layer illustrates 2 pictures, one which shows the filters themselves and one that shows what part of the image are most strongly activated by the given filter. For example, in the space labled Layer 2, we have representations of the 16 different filters (on the left)

https://arxiv.org/pdf/1311.2901v3.pdf



 Visualizations of Layers 3, 4, and 5 https://arxiv.org/pdf/1311.2901v3.pdf

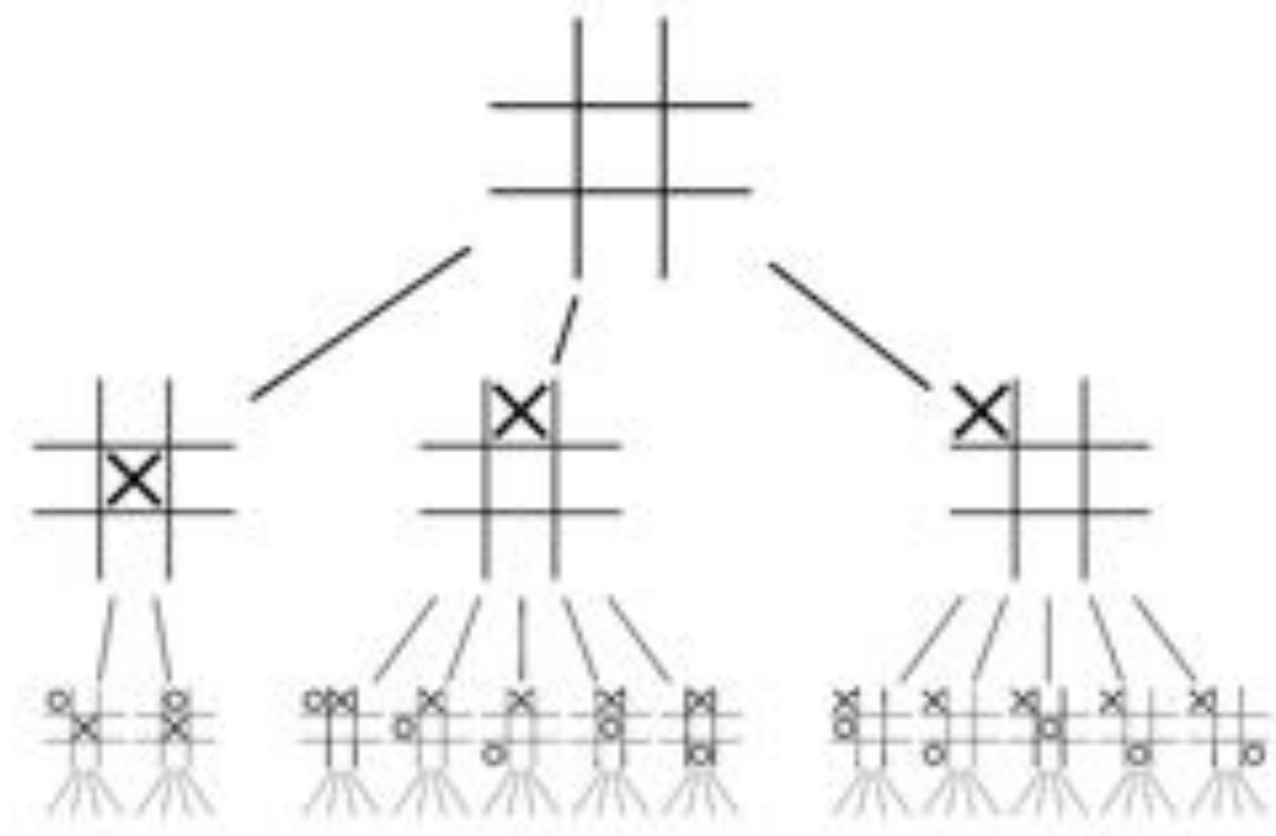


Visualizations of Layers 3, 4, and 5

Promise: New Features

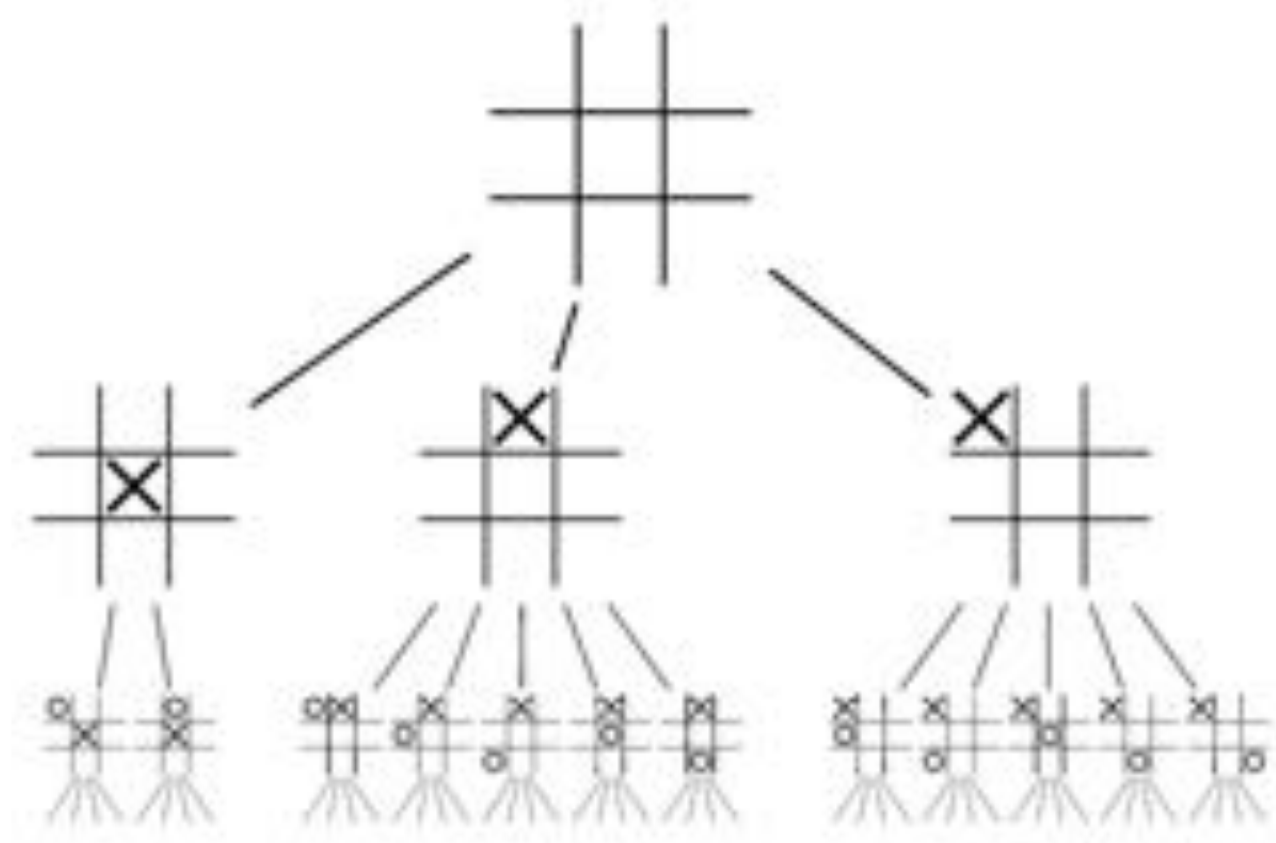
https://arxiv.org/pdf/1311.2901v3.pdf

Tic-Tac-Toe with Deep Learning?



Wikipedia

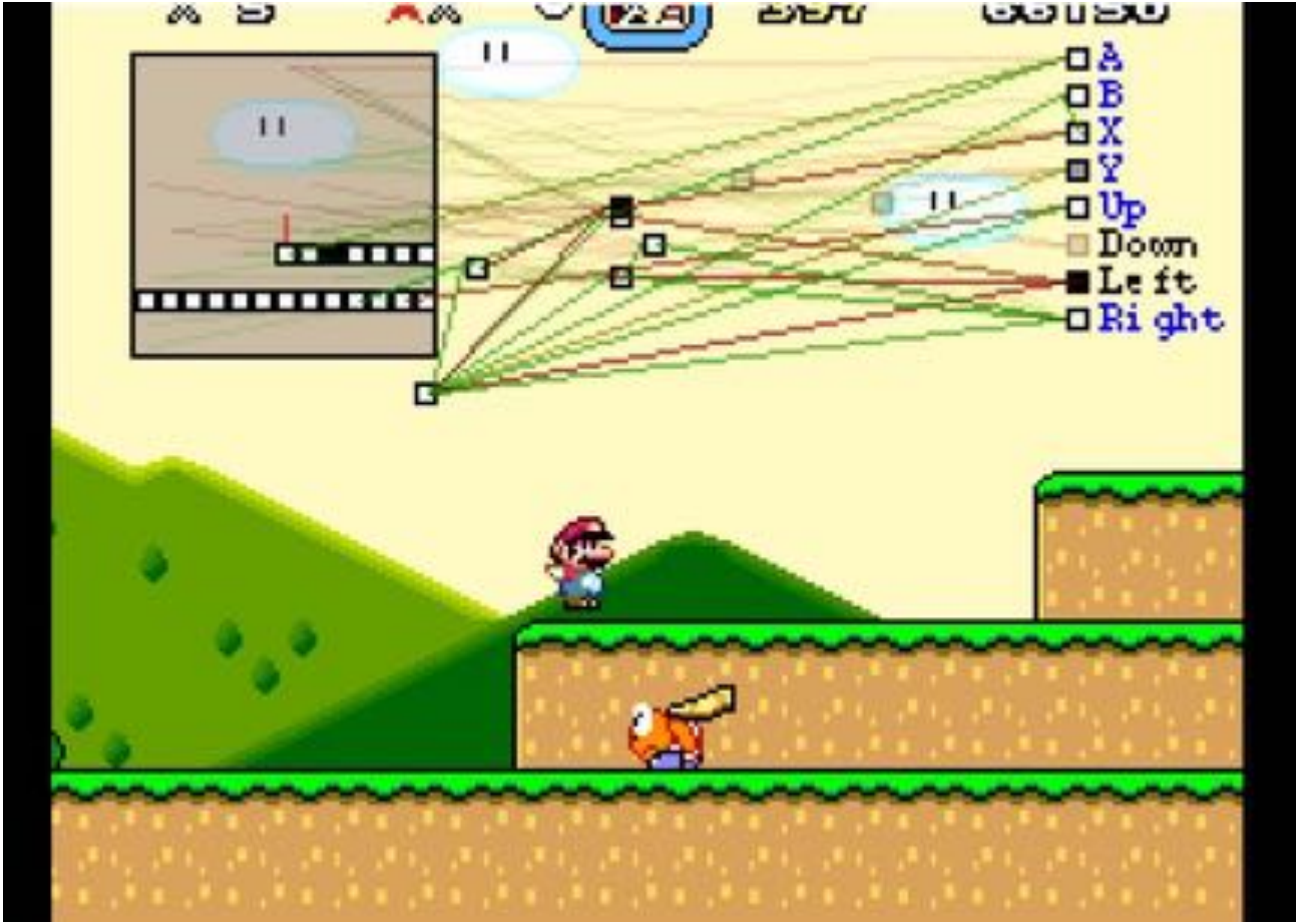
Tic-Tac-Toe with Deep Learning?



Wikipedia

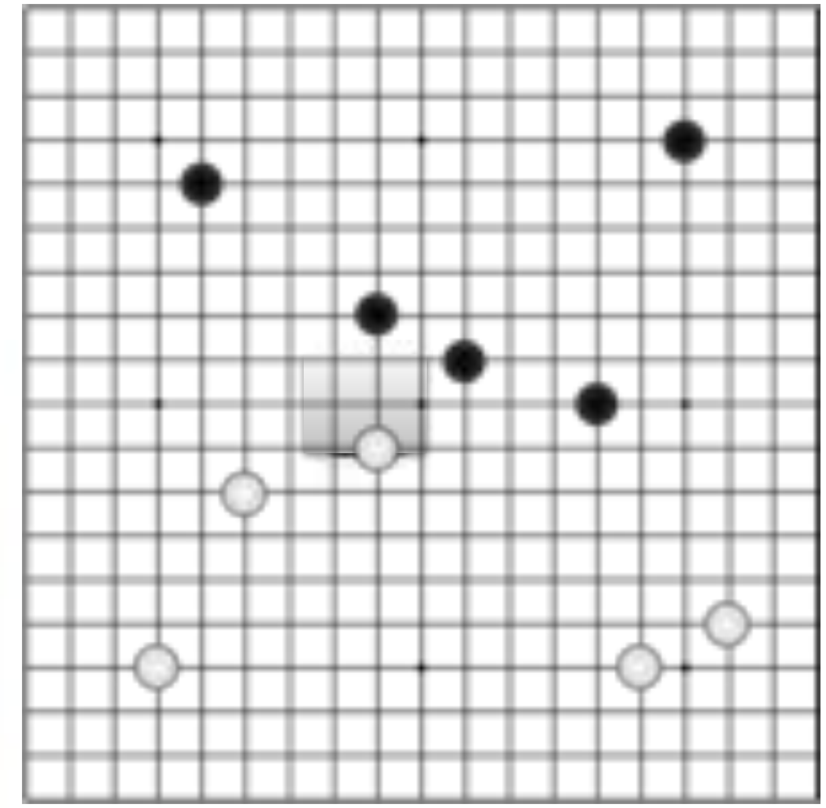
Too shallow!

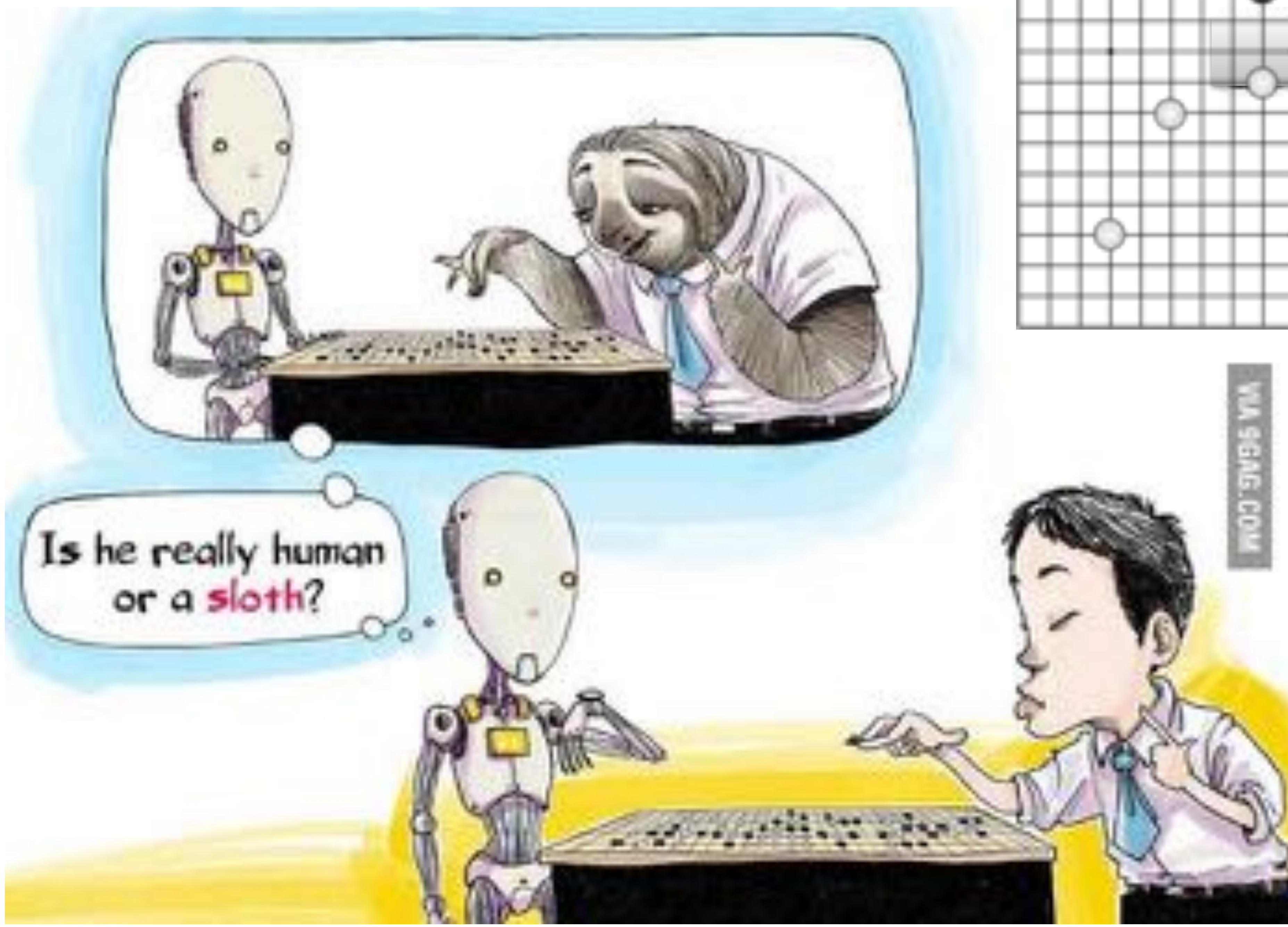
What about Mario? (Deep Mind)



GO: Even Harder

Chess: 20 options
GO: 200 options





10^170 positions

9gag.com

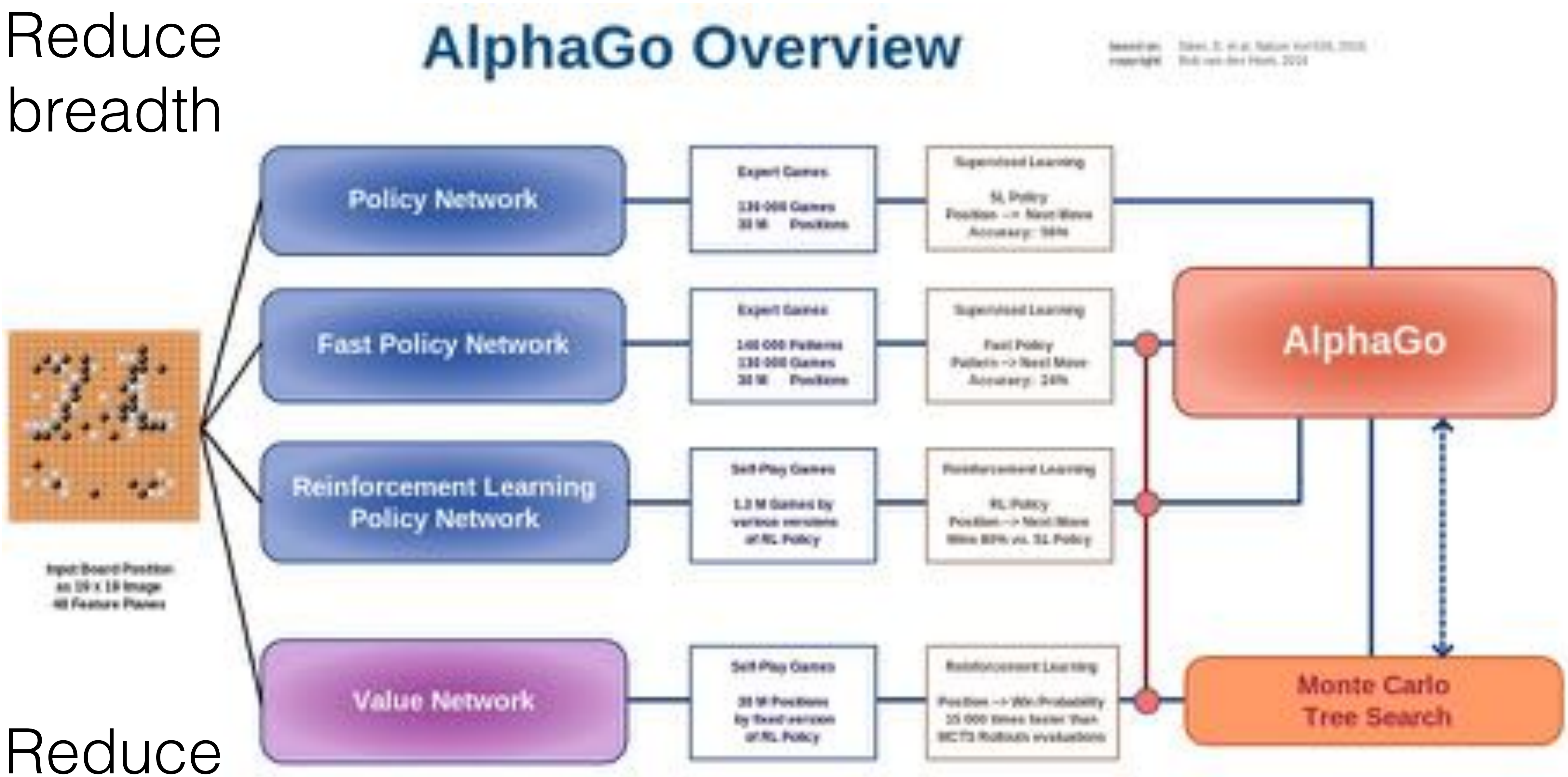
Demis Hassabis,
CEO Deep Mind

Lee Sedol,
18 World Titles

GO: Even Harder

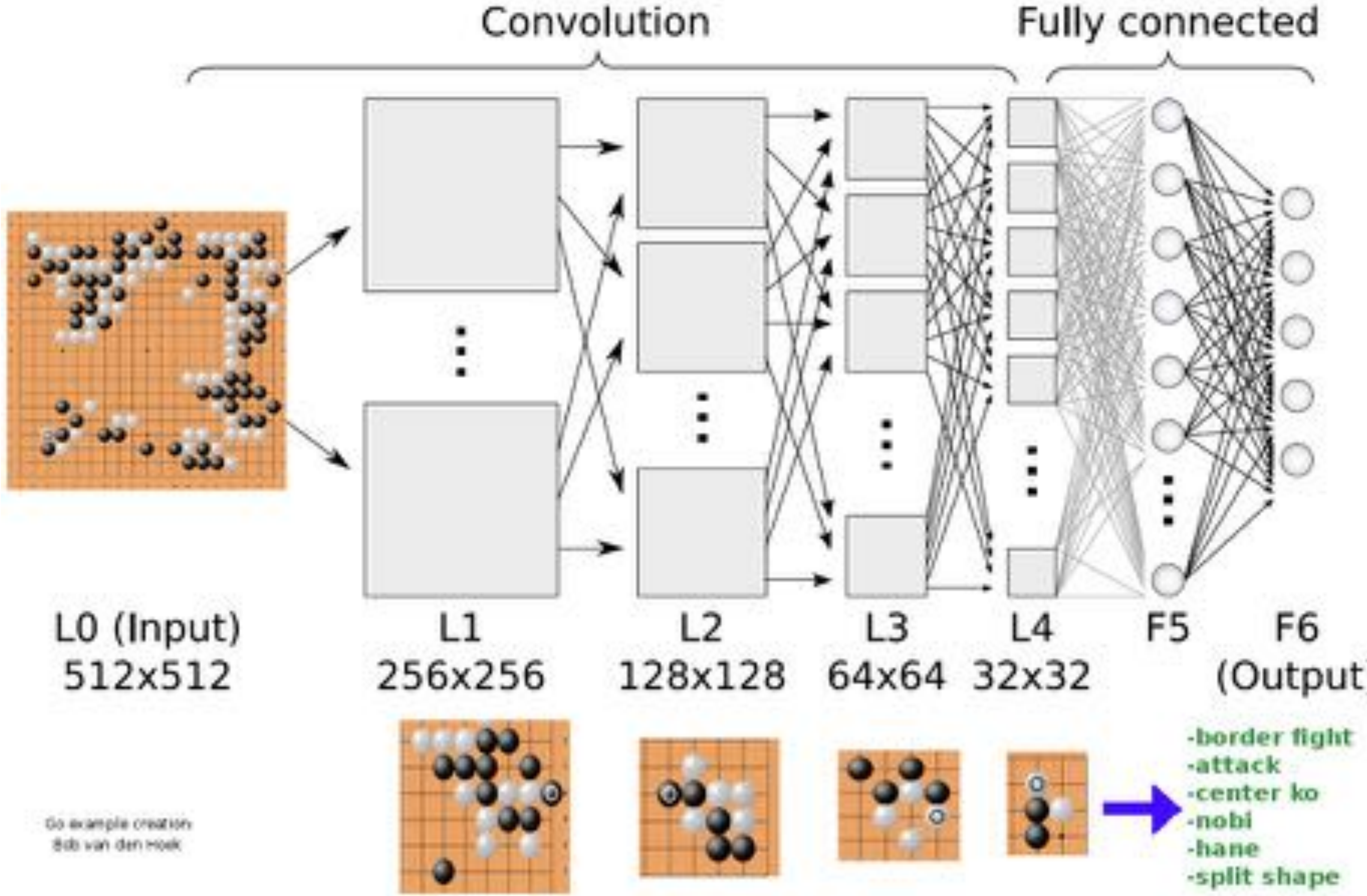
9gag.com

Reduce breadth



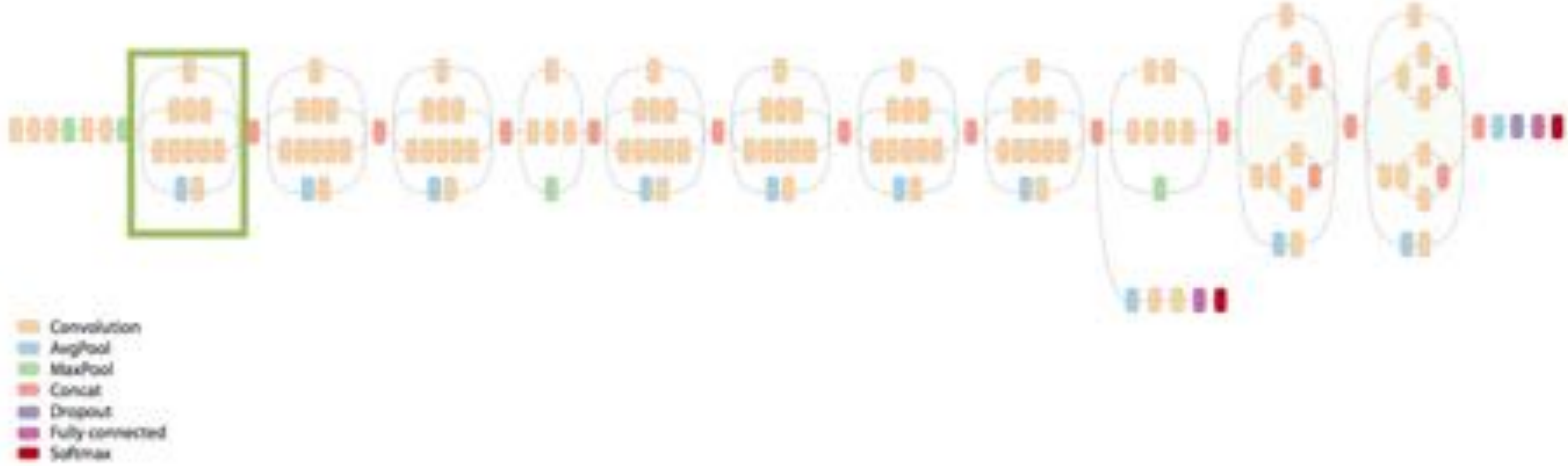
Reduce depth

http://deeplearningskysthelimit.blogspot.com/2016/04/part-2-alphago-under-magnifying-glass.html



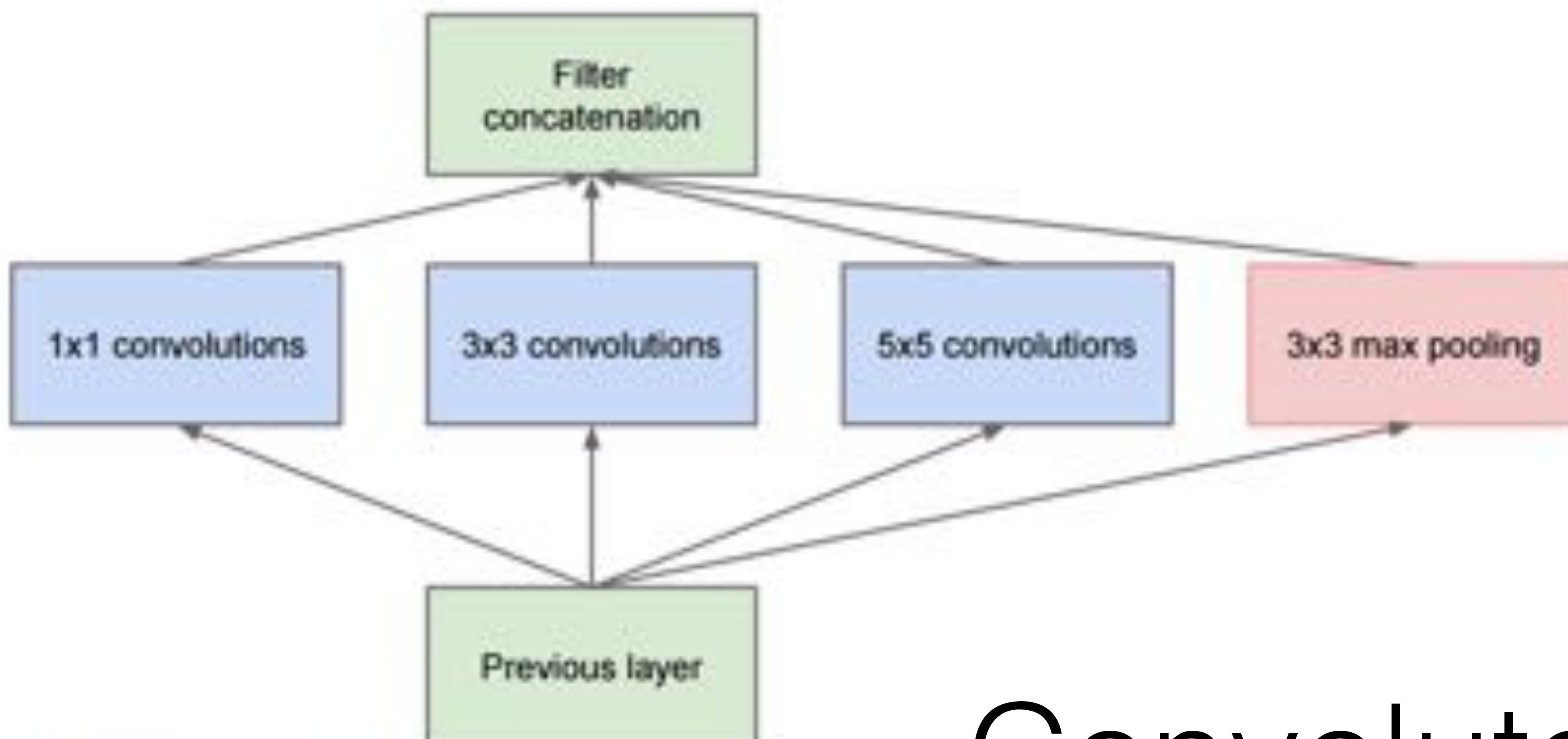
http://gobase.org/online/intergo/?query=%22hane%20nobi%22

http://deeplearningskysthelimit.blogspot.com/2016/04/part-2-alphago-under-magnifying-glass.html



Green box shows parallel region of GoogLeNet

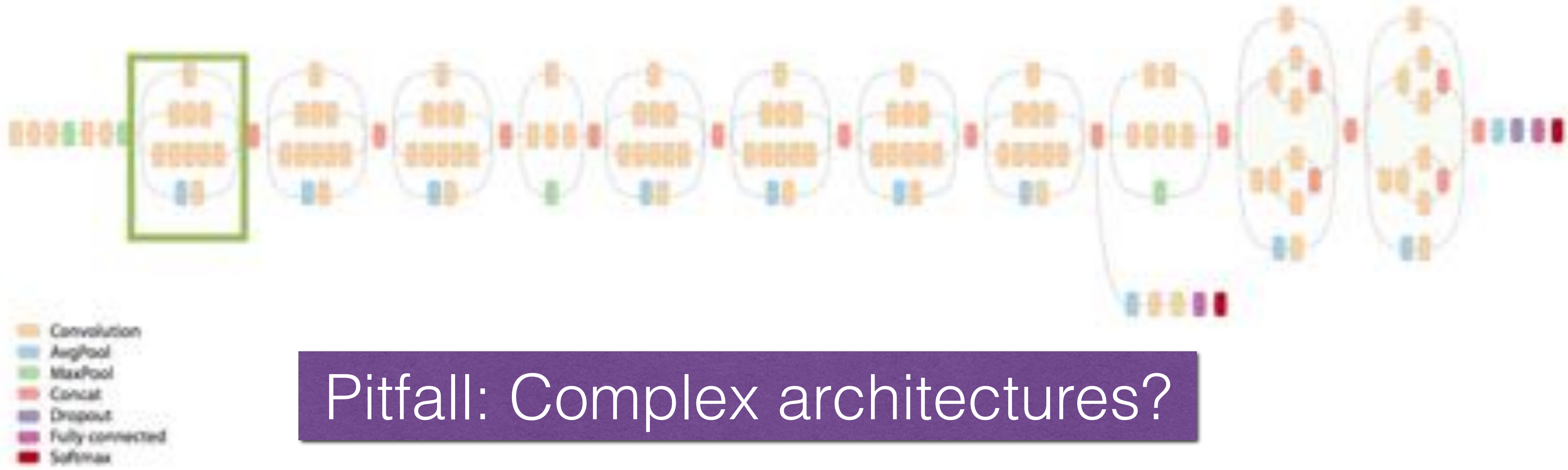
Inception module



Naive idea of an Inception module

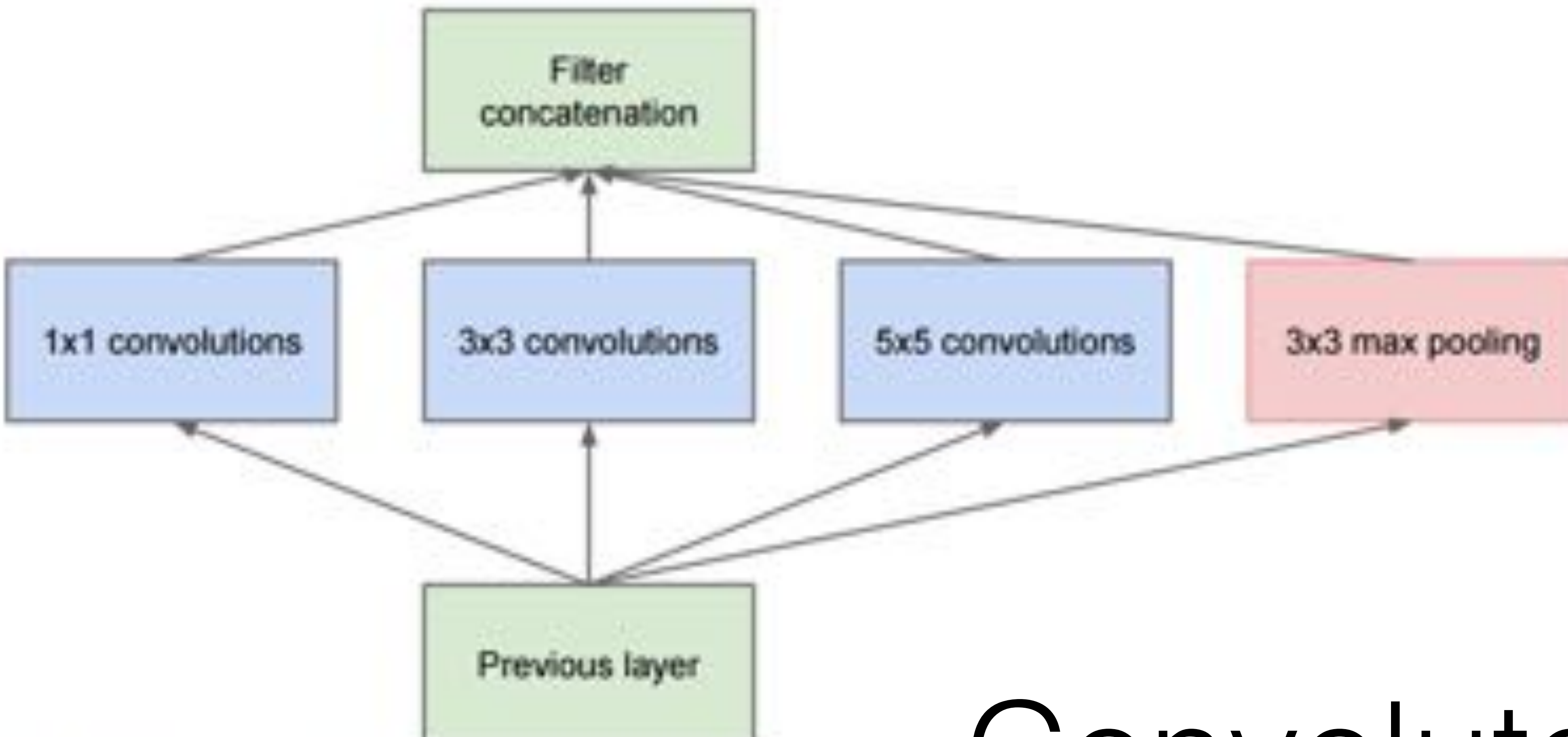
Convoluted?

Pitfall: Complex architectures?



Green box shows parallel region of GoogLeNet

Inception module

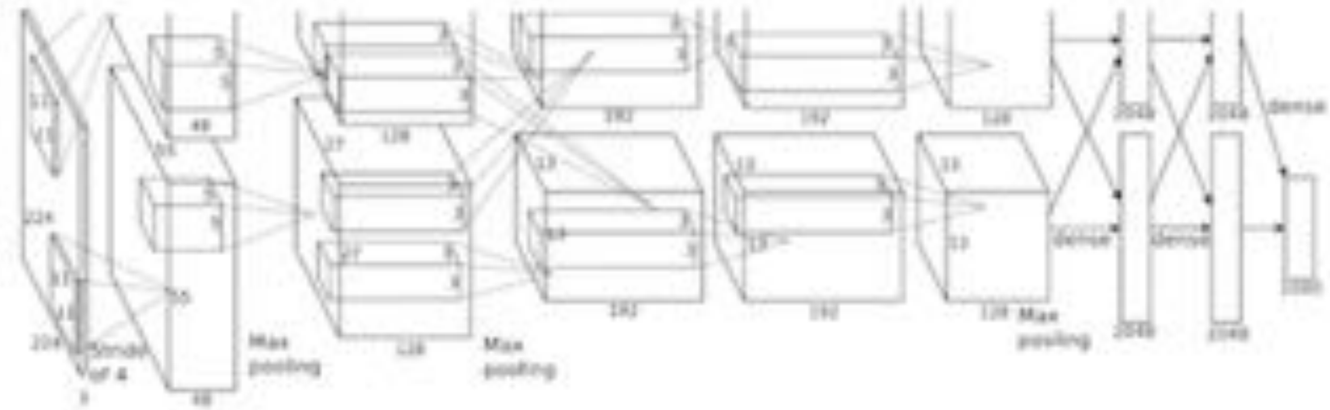


Naive idea of an Inception module

Convoluted?

IMAGENET

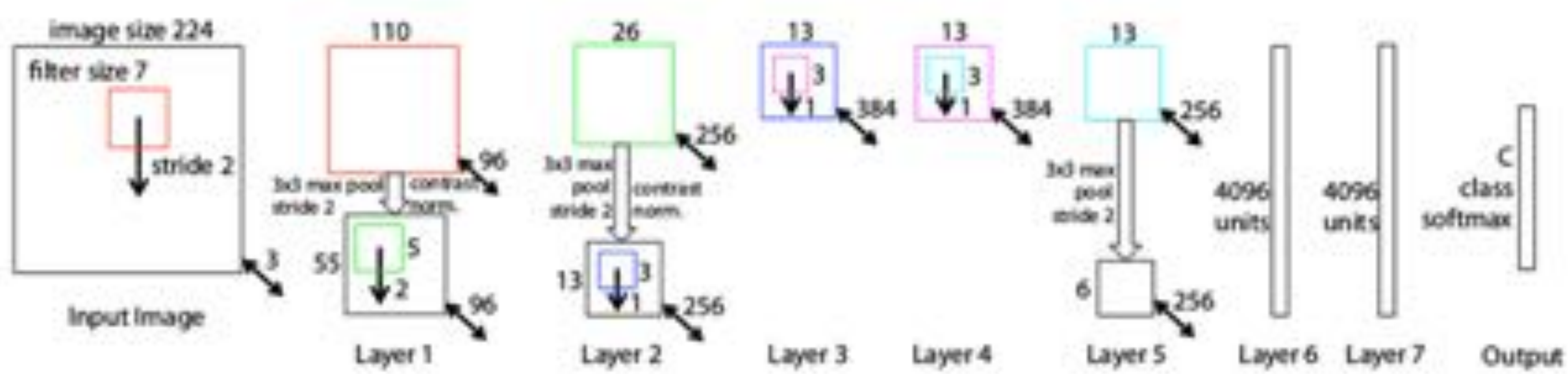
Large Scale Visual Recognition Challenge



AlexNet architecture (May look weird because there are two different "streams". This is because the training process was so computationally expensive that they had to split the training onto 2 GPUs)

- 2012: Alexnet (error rate 15.4%)
- 2013: ZFnet (error rate 11.12%)
- DeConvNets (Caffe)

ILSVRC

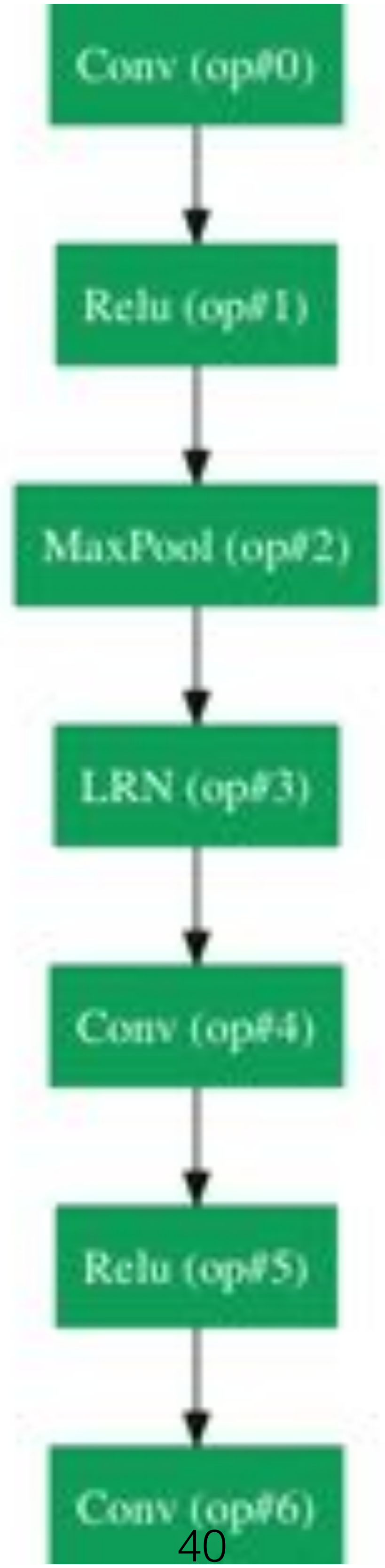


ZF Net Architecture

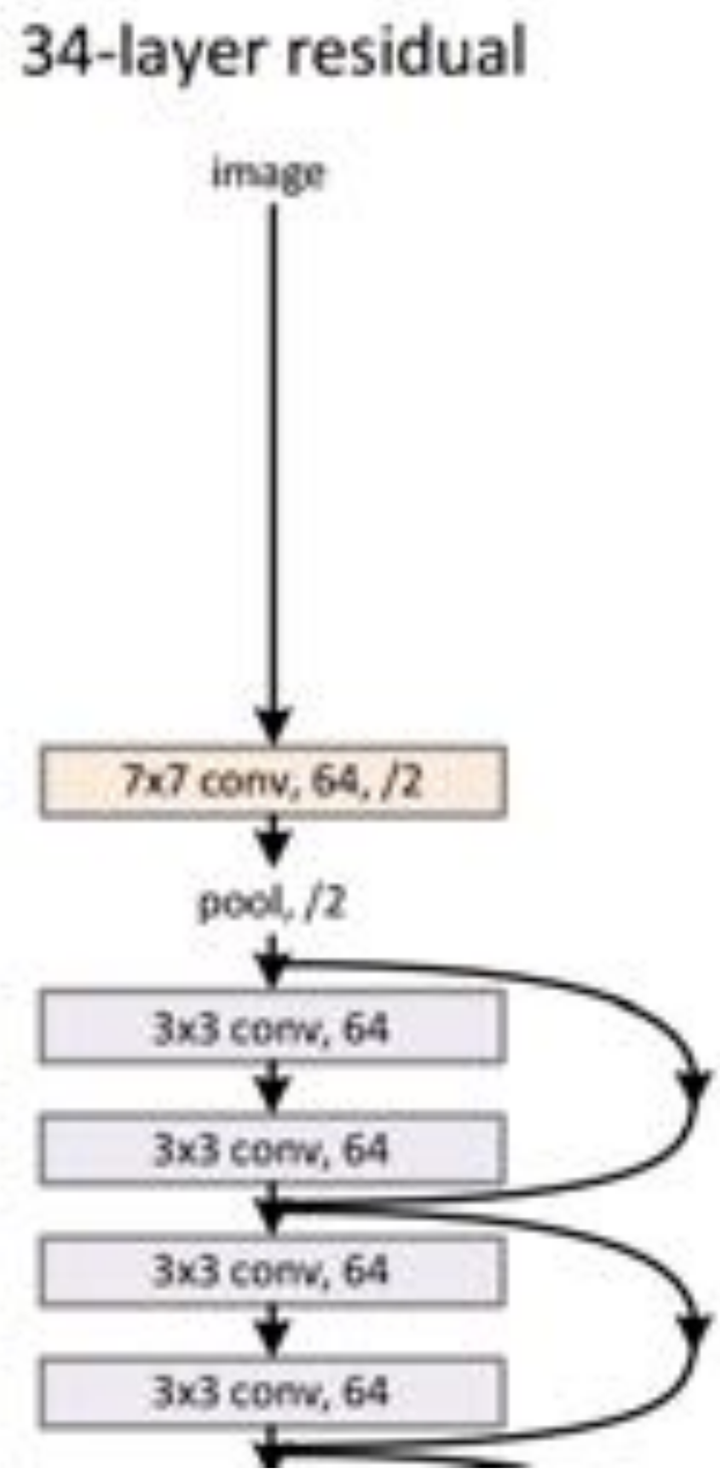
Adit Deshpande

https://adeshpande3.github.io/adeshpande3.github.io/The-9-Deep-Learning-Papers-You-Need-To-Know-About.html

GoogLeNet (2014) 6.7%



ResNET (2015) error rate: 3.6%



2015 ILSVRC leaderboard

Team name	Entry description	Number of object categories won	mean AP
MSRA	An ensemble for detection.	194	0.620741
Qualcomm Research	NeoNet ensemble with bounding box regression. Validation mAP is 54.6	4	0.535745
CUImage	Combined multiple models with the region proposals of cascaded RPN, 57.3% mAP on Val2.	2	0.527113
The University of Adelaide	9 models	0	0.514434
MCG-ICT-	2 models on 2 proposals without category information: {SS+EB}+		

Classification error: 0.03567

Yellow: Winner in category
Yellow/White: Reveal code
Gray: Won't reveal code

2016 ILSVRC leaderboard

Team name	Entry description	Number of object categories won	mean AP
CUImage	Ensemble of 6 models using provided data	109	0.662751
Hikvision	Ensemble A of 3 RPN and 6 FRCN models, mAP is 67 on val2	30	0.652704
Hikvision	Ensemble B of 3 RPN and 5 FRCN models, mean AP is 66.9, median AP is 69.3 on val2	18	0.652003
NUIST	submission_1	15	0.608752
NUIST	submission_2	9	0.607124
Trimps-Soushen	Ensemble 2	8	0.61816
360+MCG-ICT-CAS_DET	9 models ensemble with validation and 2 iterations	4	0.615561
360+MCG-ICT-CAS_DET	Baseline: Faster R-CNN with Res200	4	0.590596
Hikvision	Best single model, mAP is 65.1 on val2	2	0.634003
CIL	Ensemble of 2 Models	1	0.553542
360+MCG-ICT-CAS_DET	9 models ensemble	0	0.613045

Classification error:
0.02991

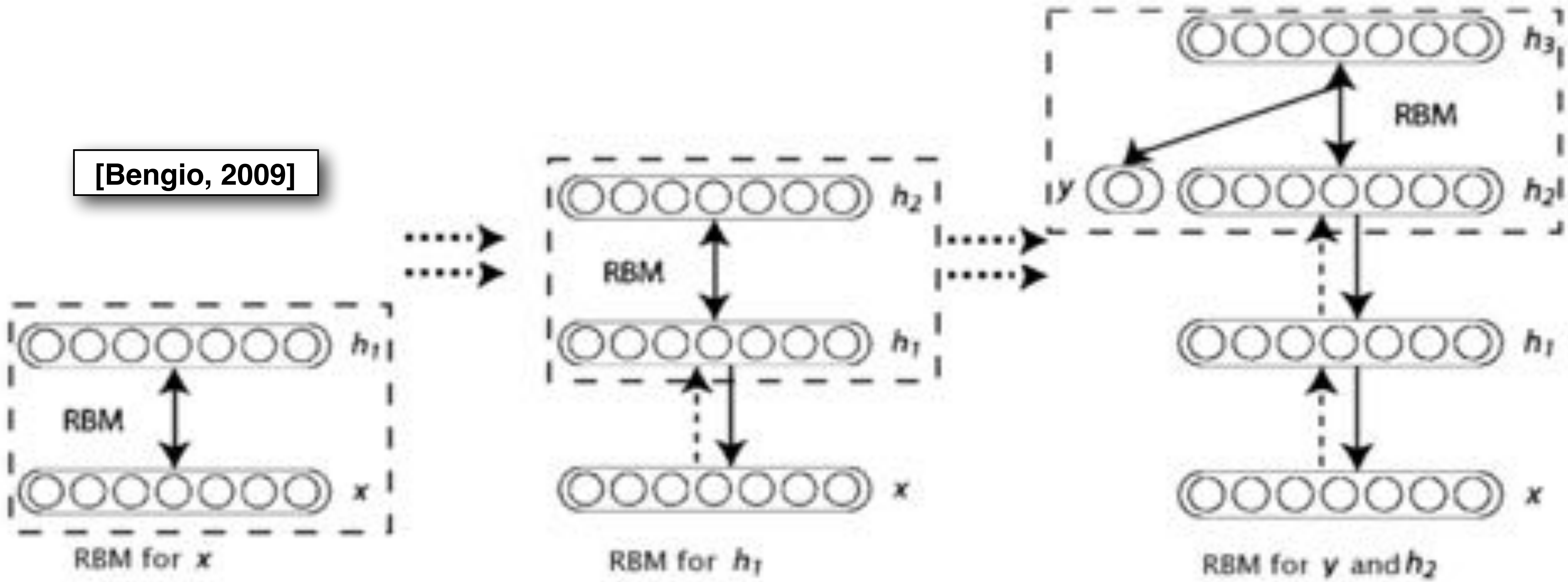
In Astro what can you apply deep learning to?

- One speculative example
- One related to your work

Finding Streaks in astronomical images

Brian Bue, Umaa Rebrapragada (PTF, Deep Belief Networks)

[Bengio, 2009]



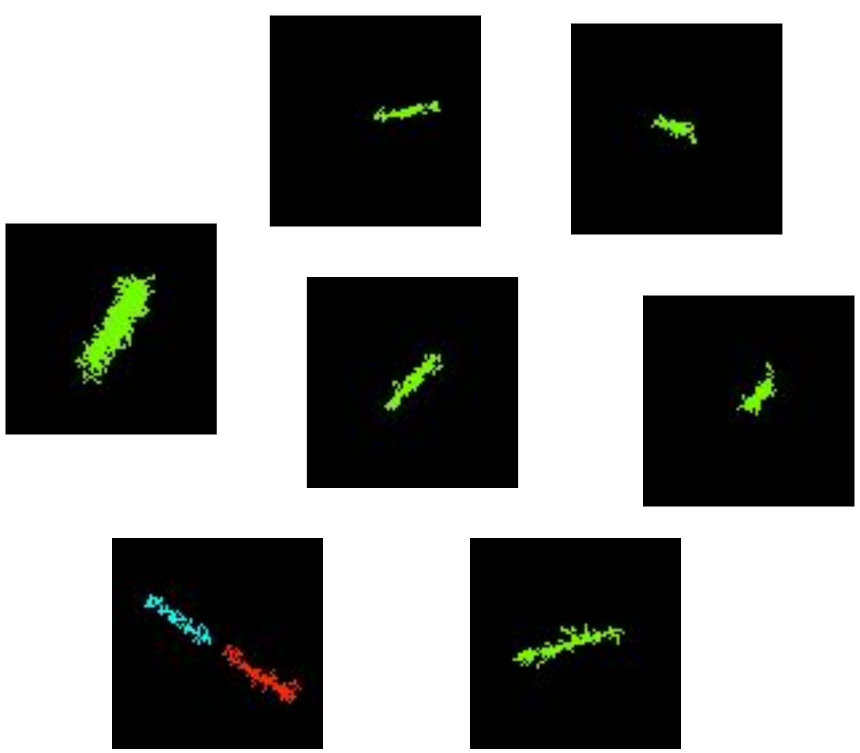
Stacking Restricted Boltzmann Machines (RBM) → Deep Belief Network (DBN)

- Train RBMs (unsupervised) and Fine-tune weights (supervised)

Asteroid Detection from Sky Survey Imagery

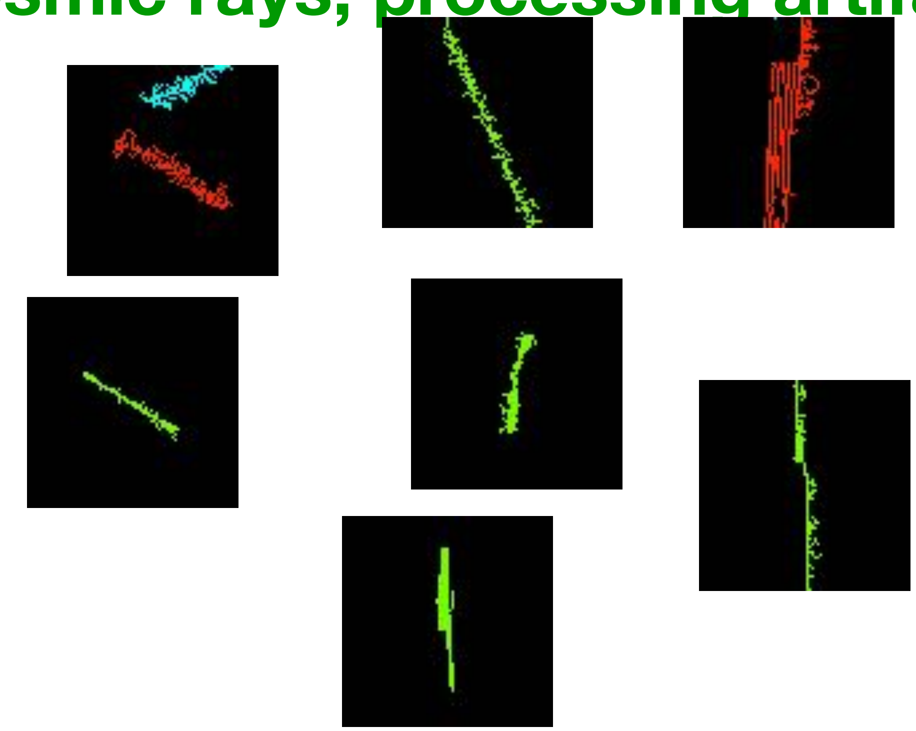
- **Goal:** automatically distinguish real vs. bogus asteroids from Palomar Transient Factory (PTF) imagery
- **Current dataset:** 240 confirmed asteroids, 1441 synthetically-generated asteroids, 20072 bogus

Confirmed Asteroids



Bogus Detections

(cosmic rays, processing artifacts)

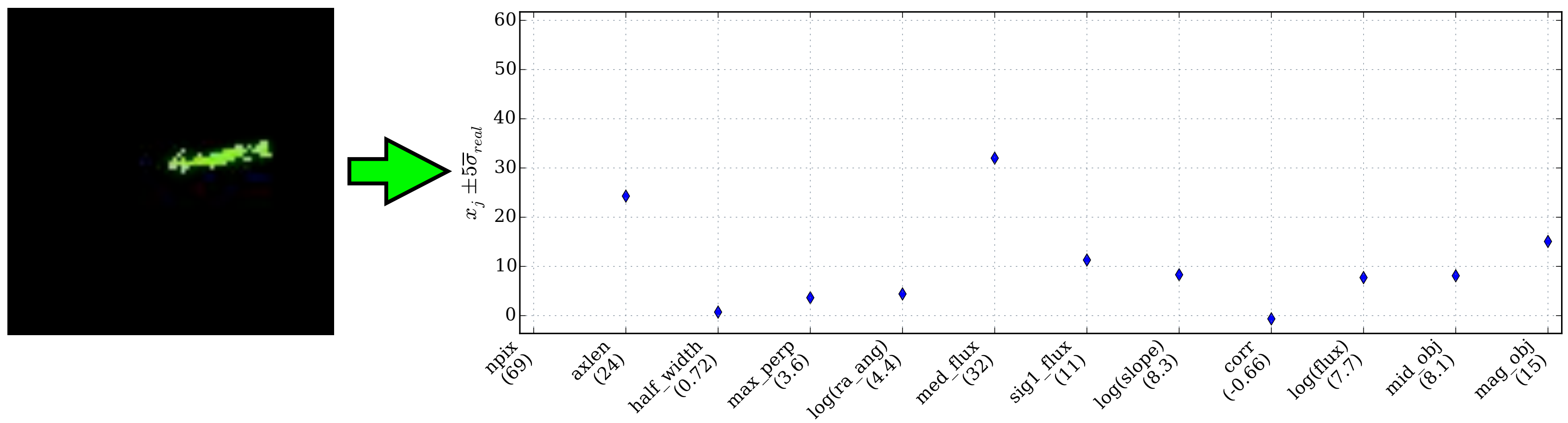


B Bue

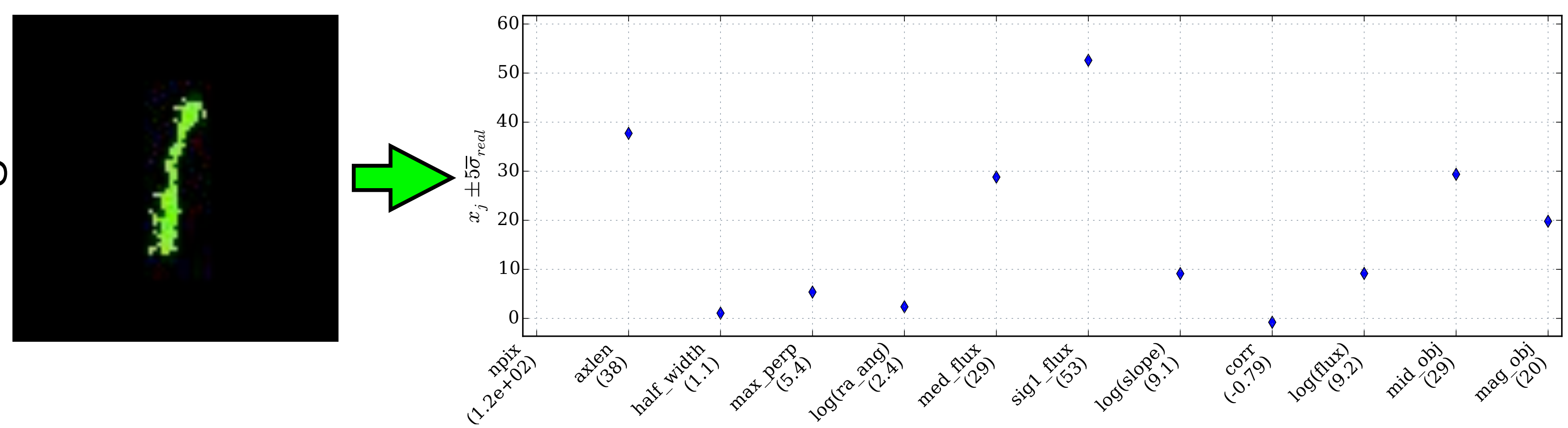
Initial Machine Learning Approach

- Extract intensity + morphology features from masked images

Real



Bogus

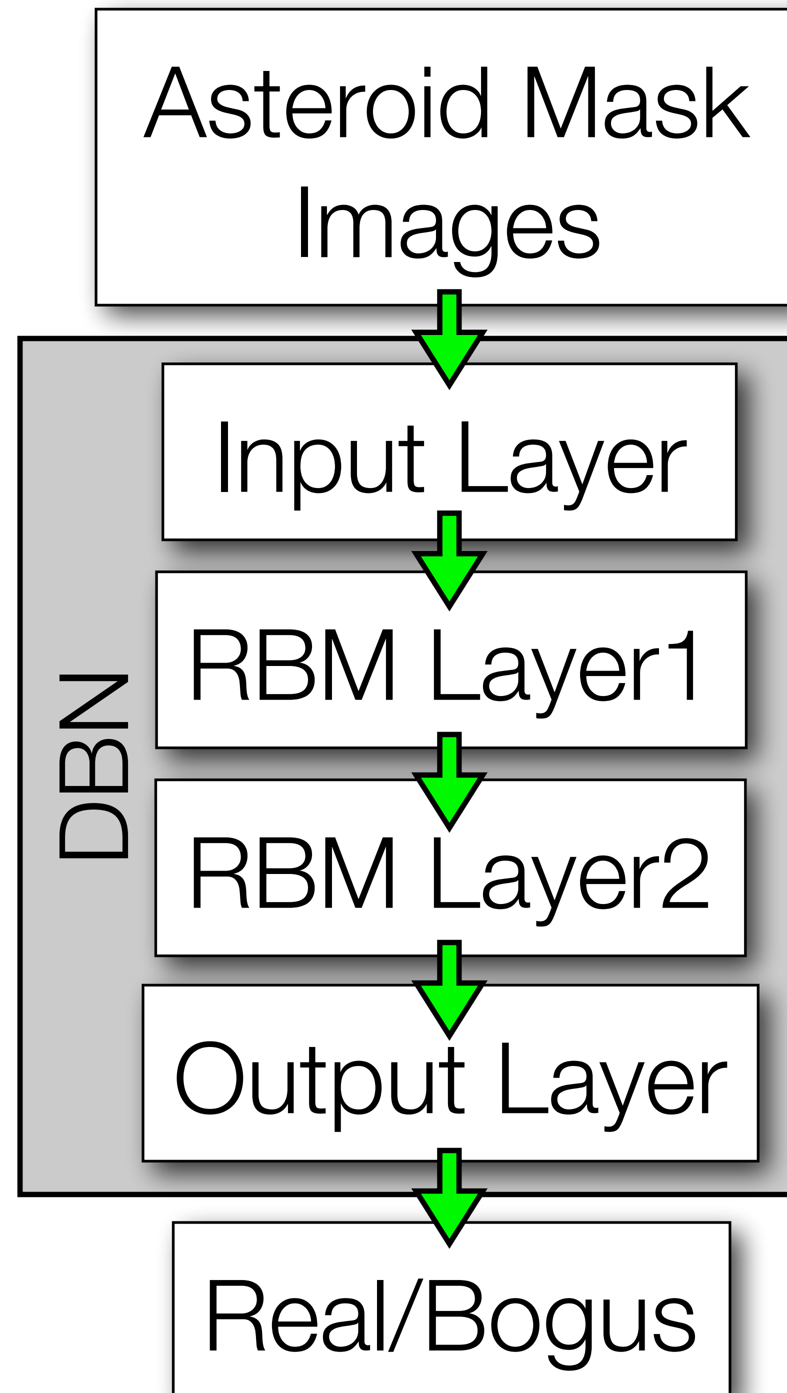


- Train/classify feature vectors using a Random Forest classifier
- Accurate (0.8% test error)...but required expert knowledge to design/extract/validate features

B Bue

Deep Learning Approach: Deep Belief Network

- **Approach:** use a DBN to detect asteroids using the (raw) asteroid mask image pixels
- Consider downsampled images...
  30x30, 45x45, 50x50
  ...with RBM layer dims...
  [1000,1000], [1600,1600], [1600], [2500]
- Experimental setup:
  - Pylearn2 DBN implementation
  - 2.8Ghz Intel Core i7, 16GB RAM
  - CPU only (no GPU optimization)

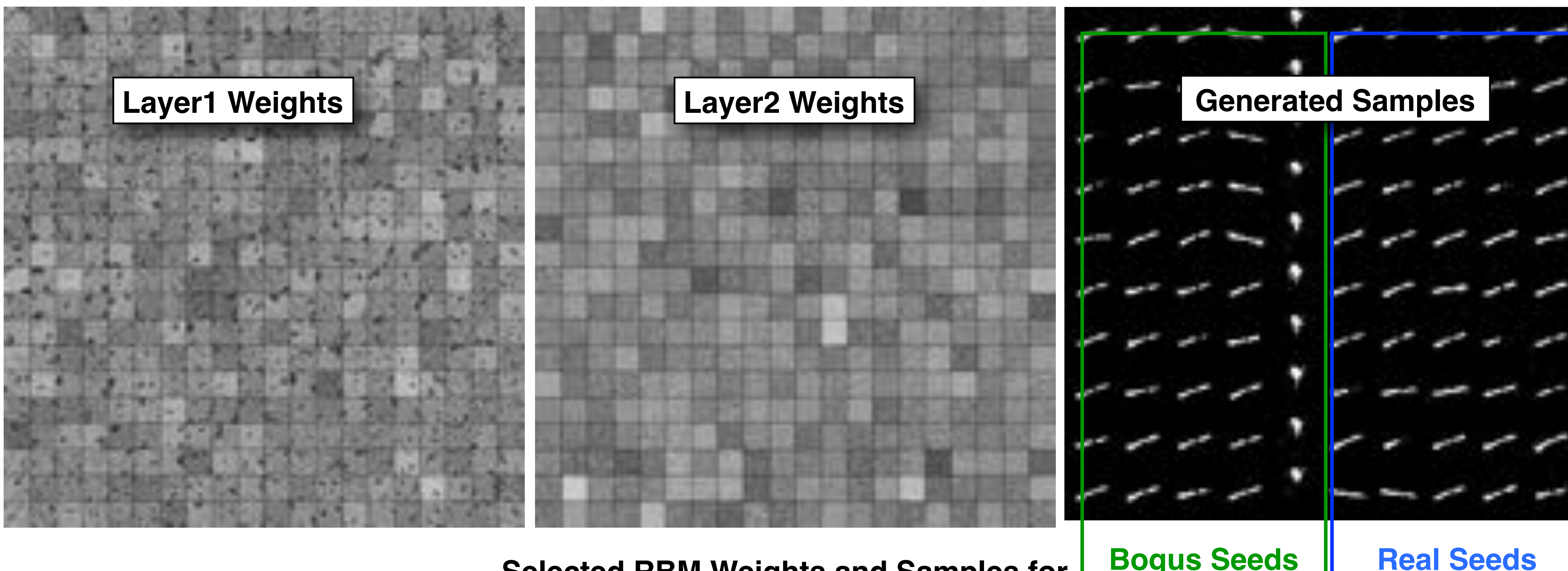


B Bue

RBM Weights, Samples for Asteroids DBN

B Bue

Image Dims	RBM Dims	Test Error	Pretraining time (minutes)	Finetune time (minutes)
30x30	[900,900]	5.07%	86	900
45x45	[1600]	1.47%	74	358
45x45	[1600,1600]	1.21%	403	1042
50x50	[2500]	1.15%	416	?

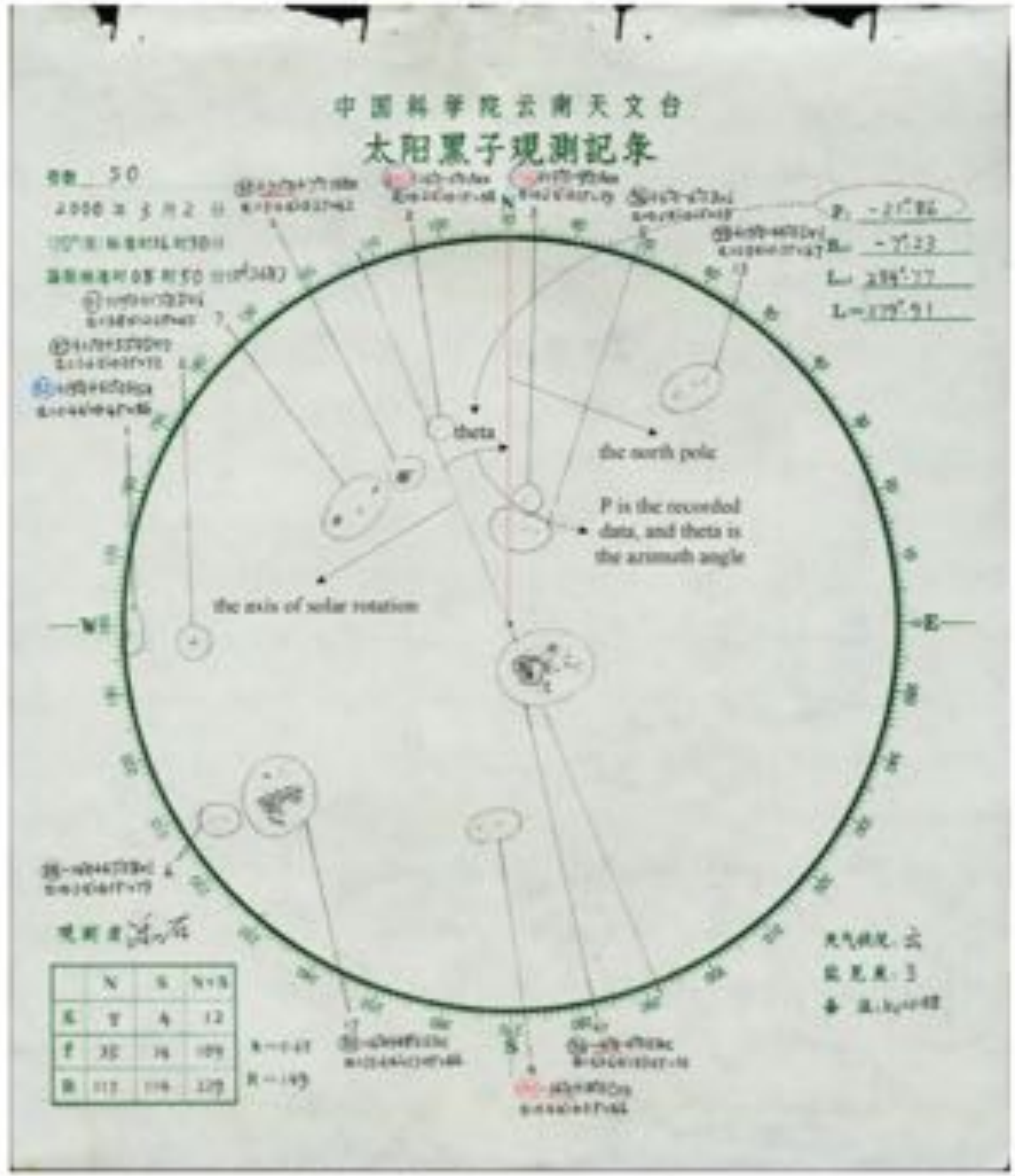


Selected RBM Weights and Samples for [1600,1600] DBN with 45x45 images

F Gieseke, A Mahabal, B Bue, U Rebrapragada (Deep Belief -> Learning)

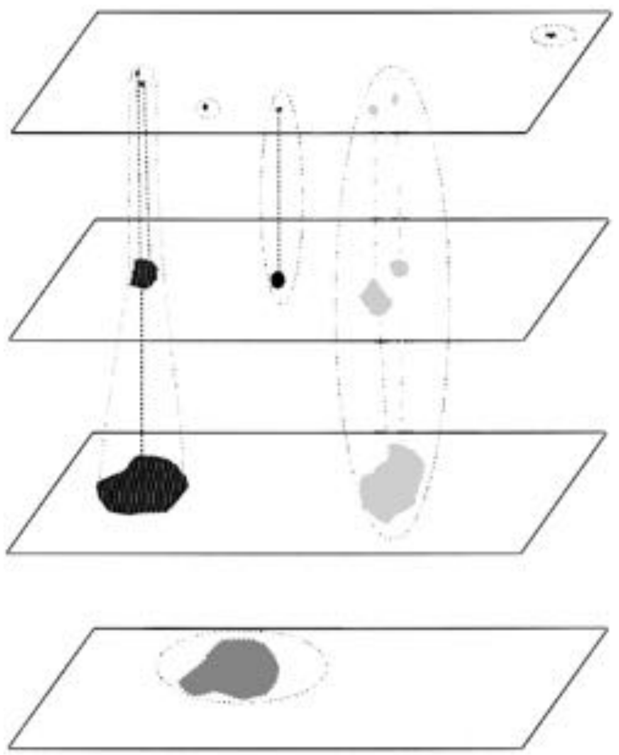
Sunspot drawings' handwritten character recognition (Zheng et al., 2016)

400 Year records



Ashish Mahabal

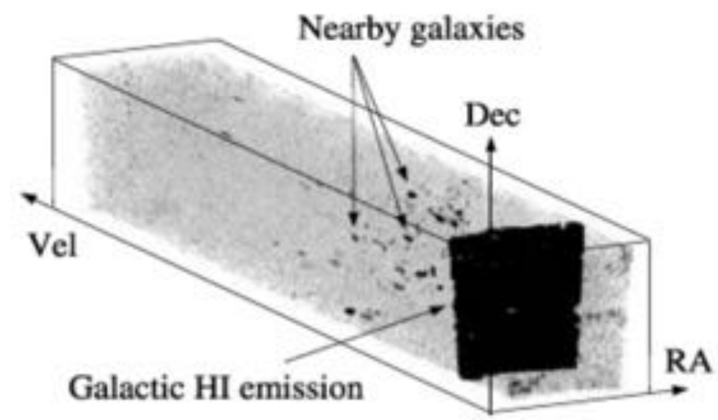
Applications



From Starck et al. AA Sup. 147, 139 (2000)

Bump hunting in radio astronomical datacubes

Single-dish data



HI Parkes All-Sky Survey (HIPASS) 21-cm data cube showing nearby galaxies (dark spots) and the Galactic Plane (dark sheet). (Meyer et al. 2004). Understanding the noise properties is particularly important for finding the faintest sources.

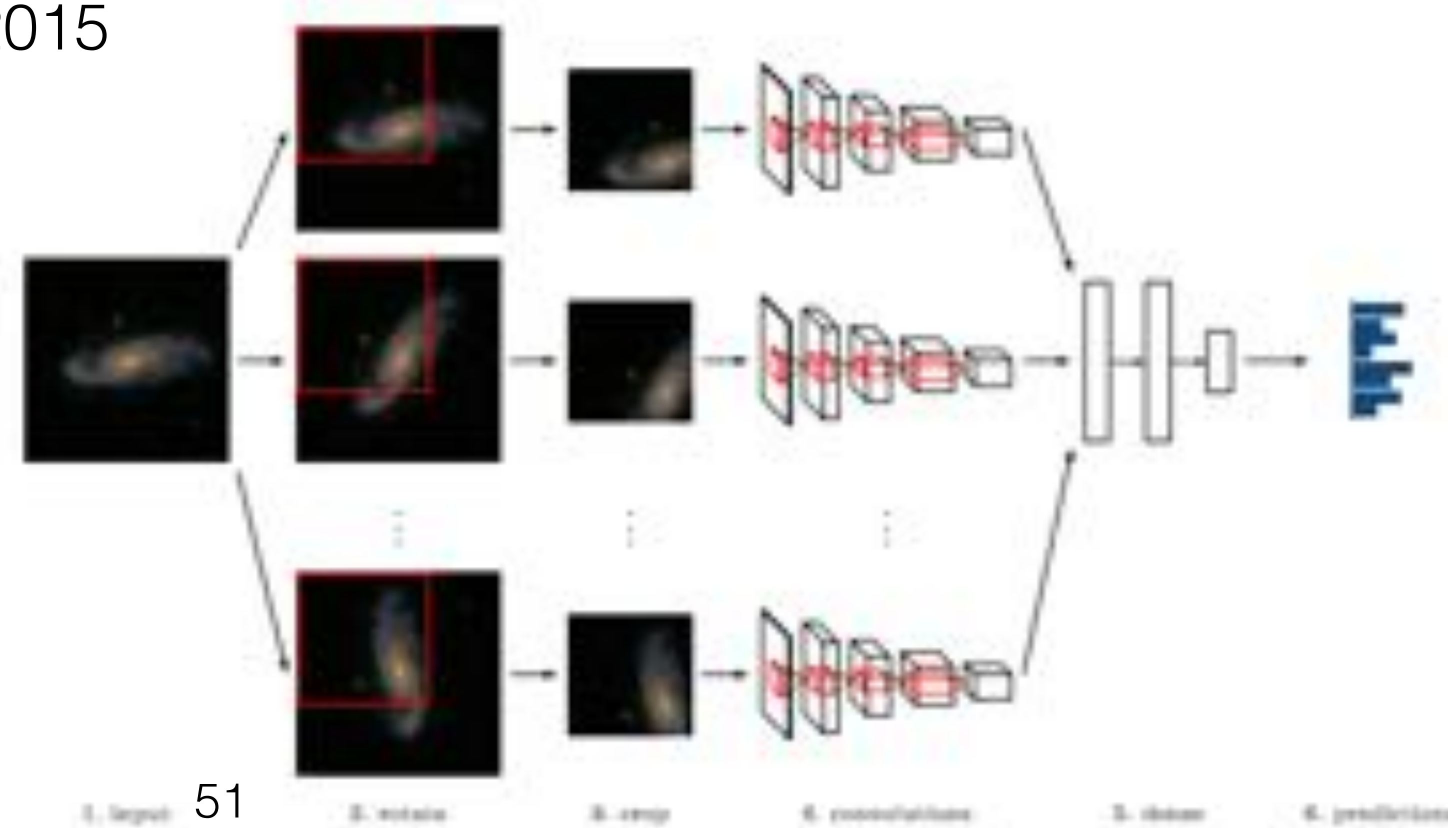
Babu (Digging deeper, KISS, Dec 2011)

Galaxy images (Huertas-Company et al., 2015)

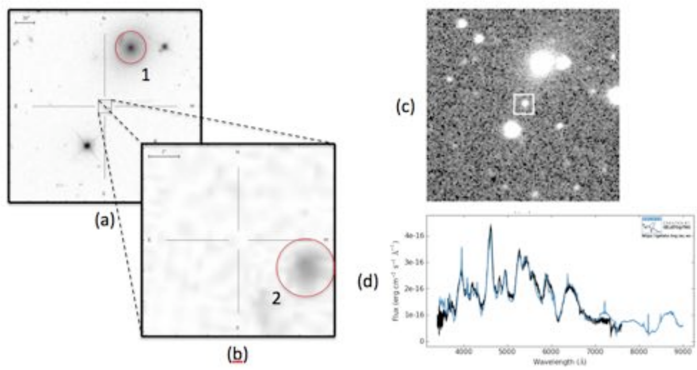
Dieleman, Willett & Dambre

2015

smoothness
edge-on
bar
spiral
bulge
roundedness



binary brokers



input

Not C1

C1

input

Not C2

C2

input

Not C3

C3

...

Inputs:
Light-curves
Nearby objects
Archival catalogs

with
L Hebb
C Frohlich
S Kanbur

Modular

input

Not C1

C1

input

Not C2

C2

input

Not C3

C3

+

input

Not C4

C4

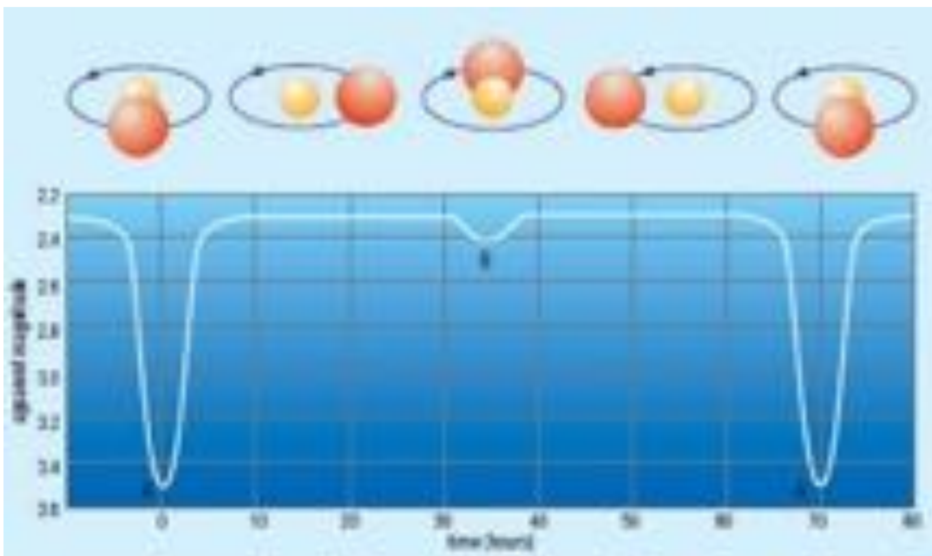
input

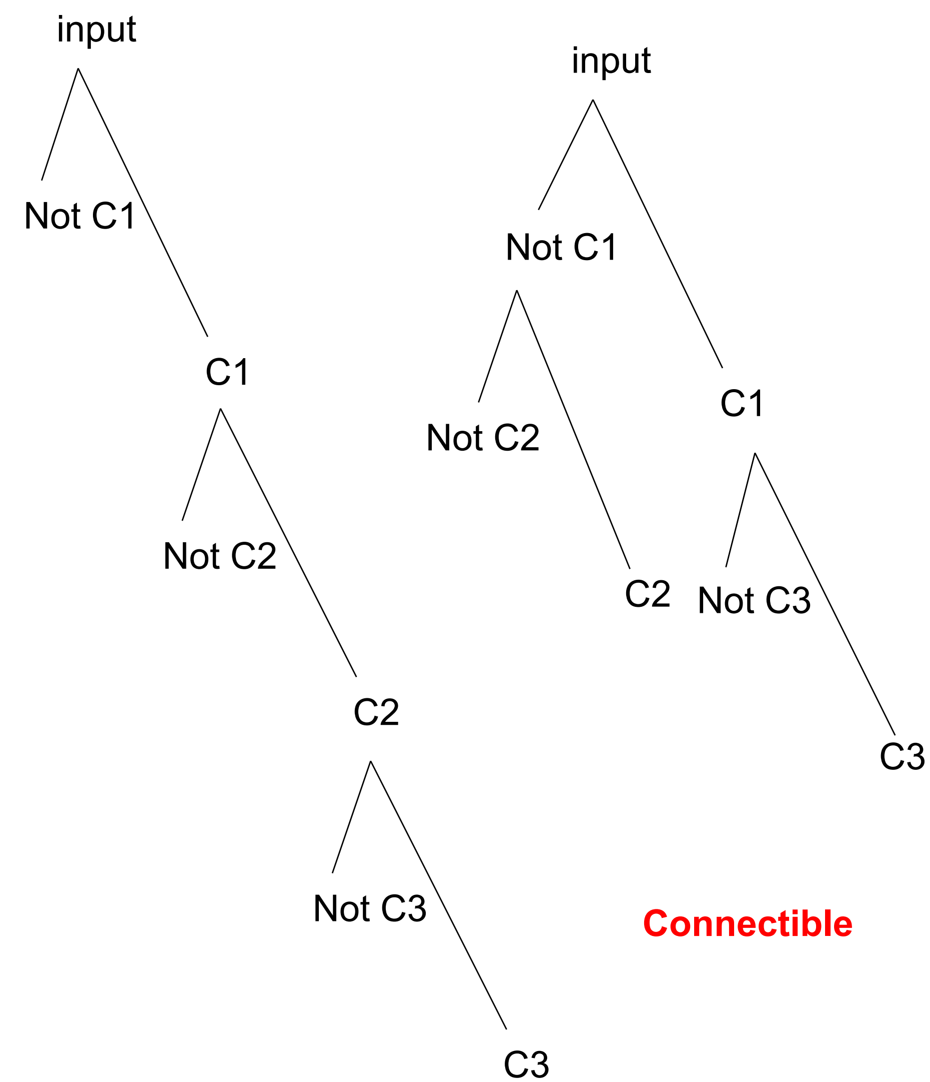
Not C5

C5

Extendible

Periodic Binaries

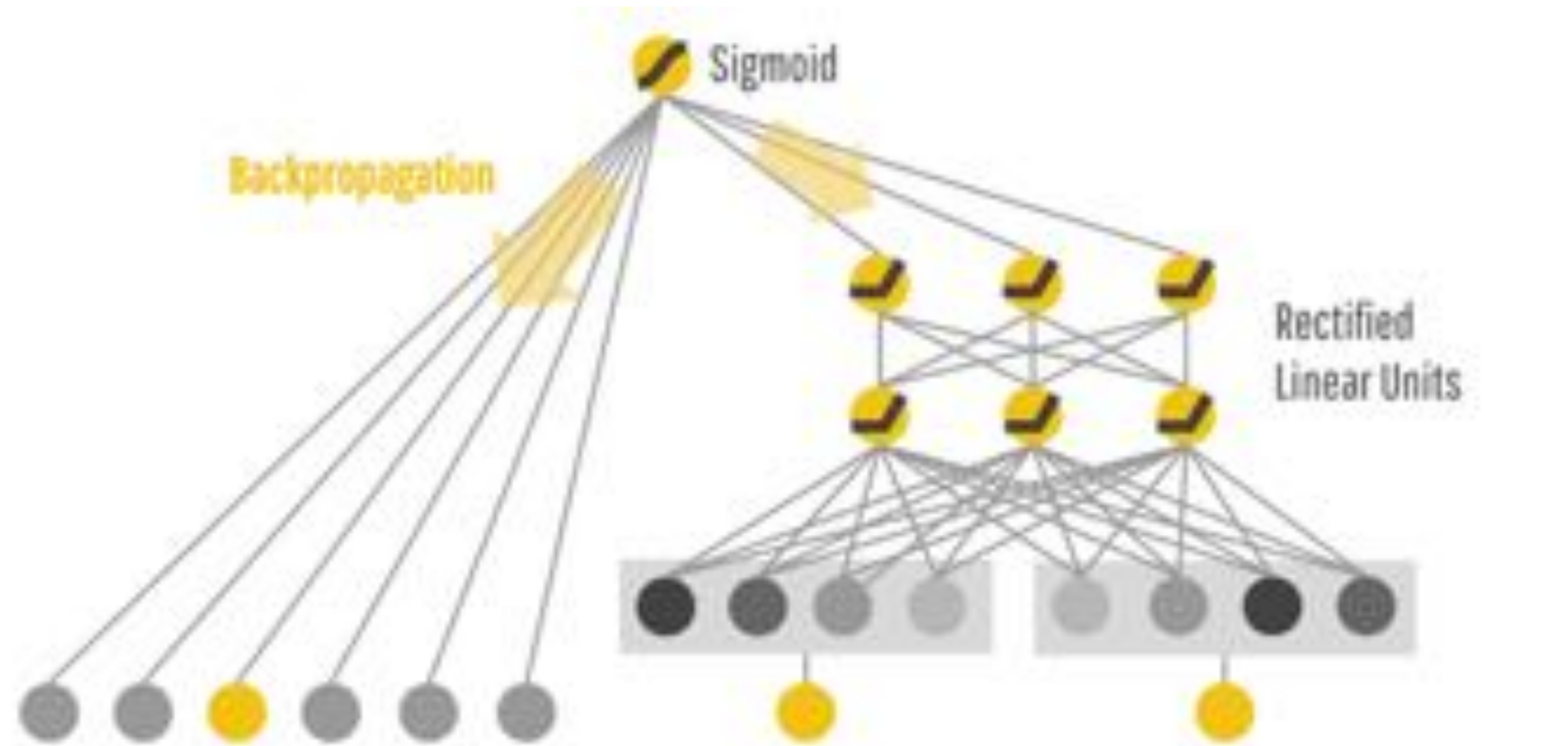




Connectible

Is what I am seeing a supernova?
Is it a blazar?
Is it a periodic variable?
Is the periodic variable an eclipsing binary?
Is it an eclipsing binary with low metallicity?

Combining with unstructured data



Feature values and ancillary data

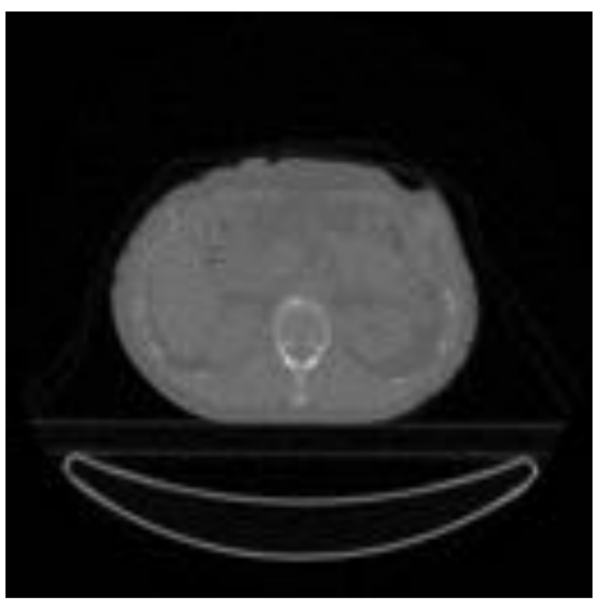
Image data (multiple wavelengths)

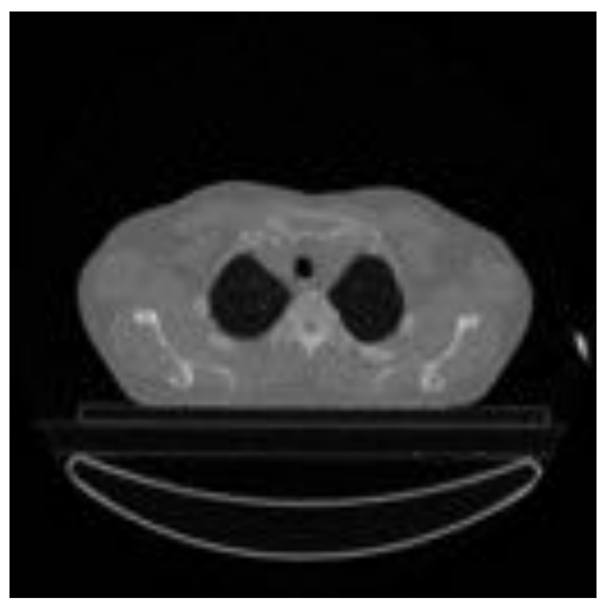
The “comments” or metadata become additional features (GoogLeNet)

https://research.googleblog.com/2016/06/wide-deep-learning-better-together-with.html

Plans with Cancer datasets

Lung dataset:
https://wiki.cancerimagingarchive.net/display/Public/NSCLC-Radiomics
DataType: **non-small cell lung cancer (NSCLC)**
modalities: CT, RSTRUCT
number of patients: **422**
number of images: **51K**
pixel dimensions: **512x512**





With and without cavity

	A	B	C	D	E	F	G	H	I	J	K
1	PatientID	age	clinical.T.Stag	Clinical.N.St	Clinical.M.Stage	Overall.Stag	Histology	gender	Survival.ti	deadstatus.event	
2	LUNG1-00	78.7515	2	3	0	IIIb	large cell	male	2165	1	
3	LUNG1-00	83.8001	2	0	0	I	squamous	male	155	1	
4	LUNG1-00	68.1807	2	3	0	IIIb	large cell	male	256	1	
5	LUNG1-00	70.8802	2	1	0	II	squamous	male	141	1	
6	LUNG1-00	80.4819	4	2	0	IIIb	squamous	male	353	1	
7	LUNG1-00	73.8864	3	1	0	IIIa	squamous	male	173	1	
8	LUNG1-00	81.5288	2	2	0	IIIa	squamous	male	137	1	
9	LUNG1-00	71.666	2	2	0	IIIa	adenocarc	male	77	1	
10	LUNG1-00	56.1342	2	2	0	IIIa	squamous	male	131	1	
11	LUNG1-01	71.0554	4	3	0	IIIb	squamous	female	2119	0	
12	LUNG1-01	64.3313	4	0	0	IIIb	squamous	male	515	1	
13	LUNG1-01	71.2553	3	2	0	IIIa	squamous	male	85	1	

Will use NLST

David Liu

Details

- Train Tensorflow with labeled images
- Test survival predictive power with a test set
- Report accuracy of the model.

Further possibilities

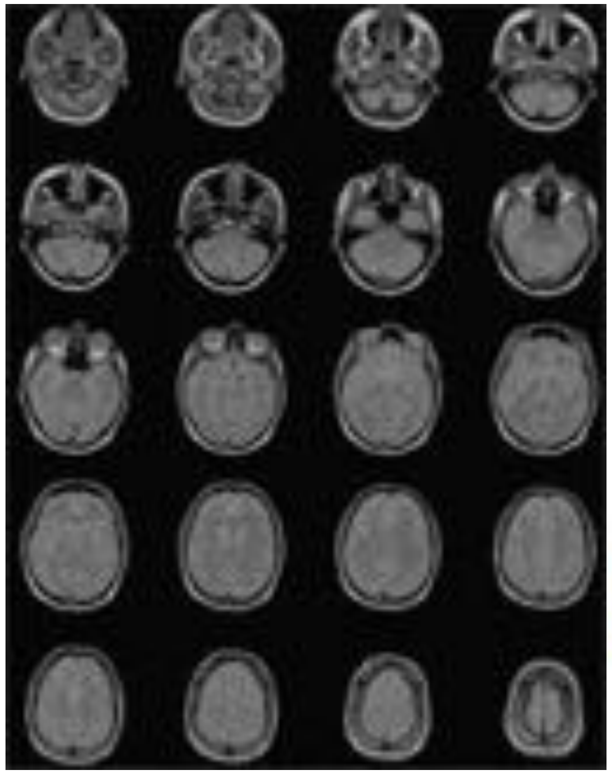
- Incorporate unlabeled radiomics feature extractions for predictive modeling
- Stratify Classifications into groups e.g. separate survival by progression/recurrence
- **General classifications** can include
  - predicting what organ(s) are in image,
  - identifying presence of a tumor,
  - biomarker signature discoveries, and more.
- Other **cancer-specific classifications**
  - cancer stage,
  - tumor stage,
  - gender

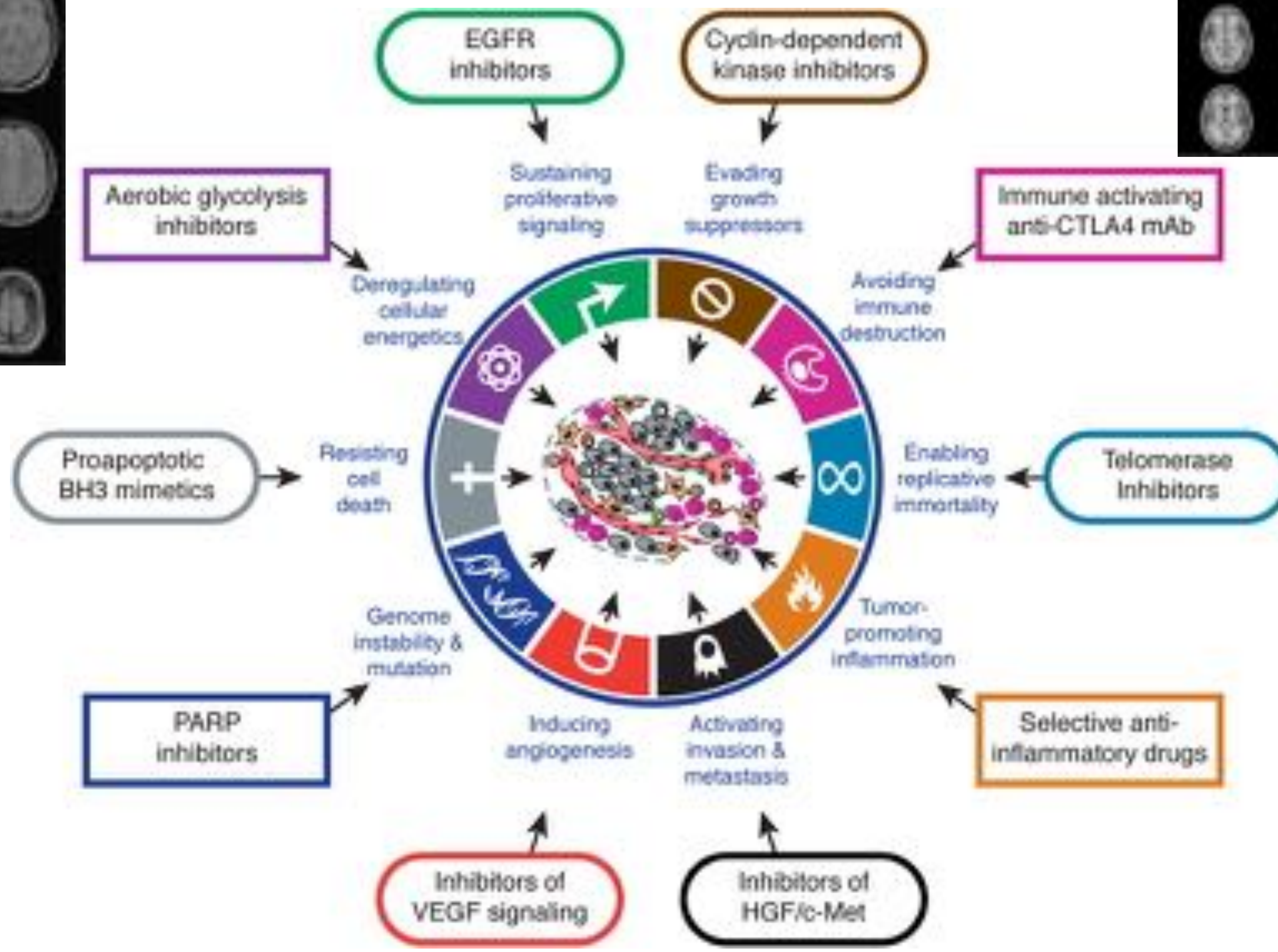
NLST
50000 Heavy smokers
Followed over years

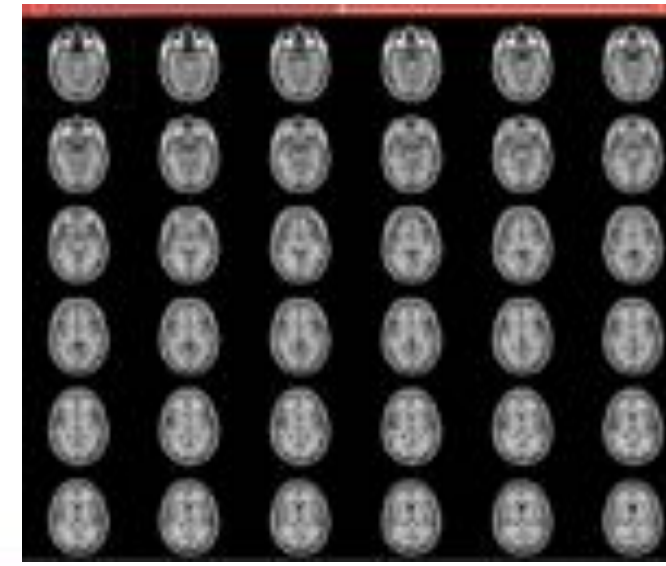
David Liu

Fully automatic

Tumors

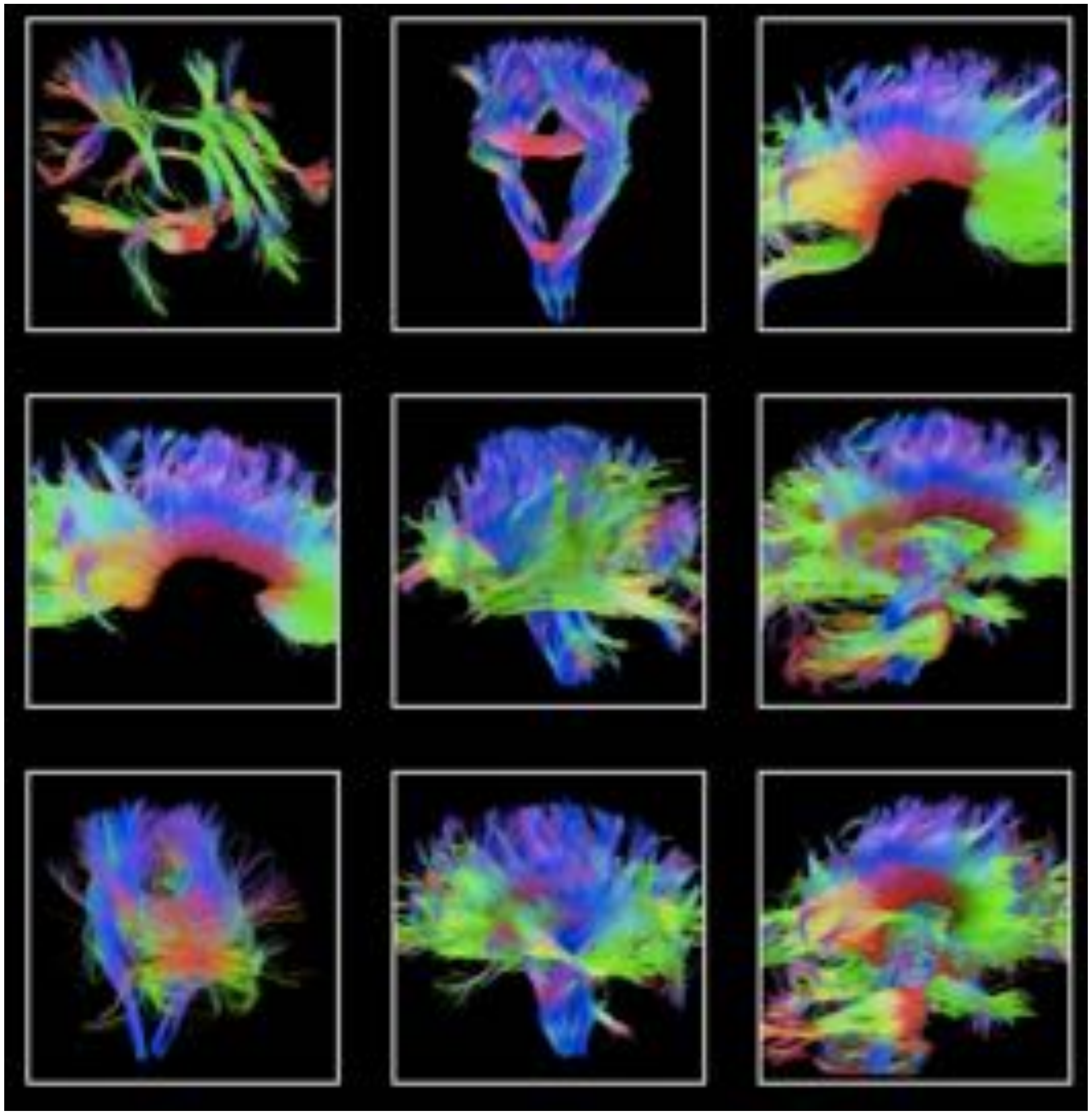






Hallmarks of cancer
Hanahan and Wagner (2000; 2011)

Human Connectome



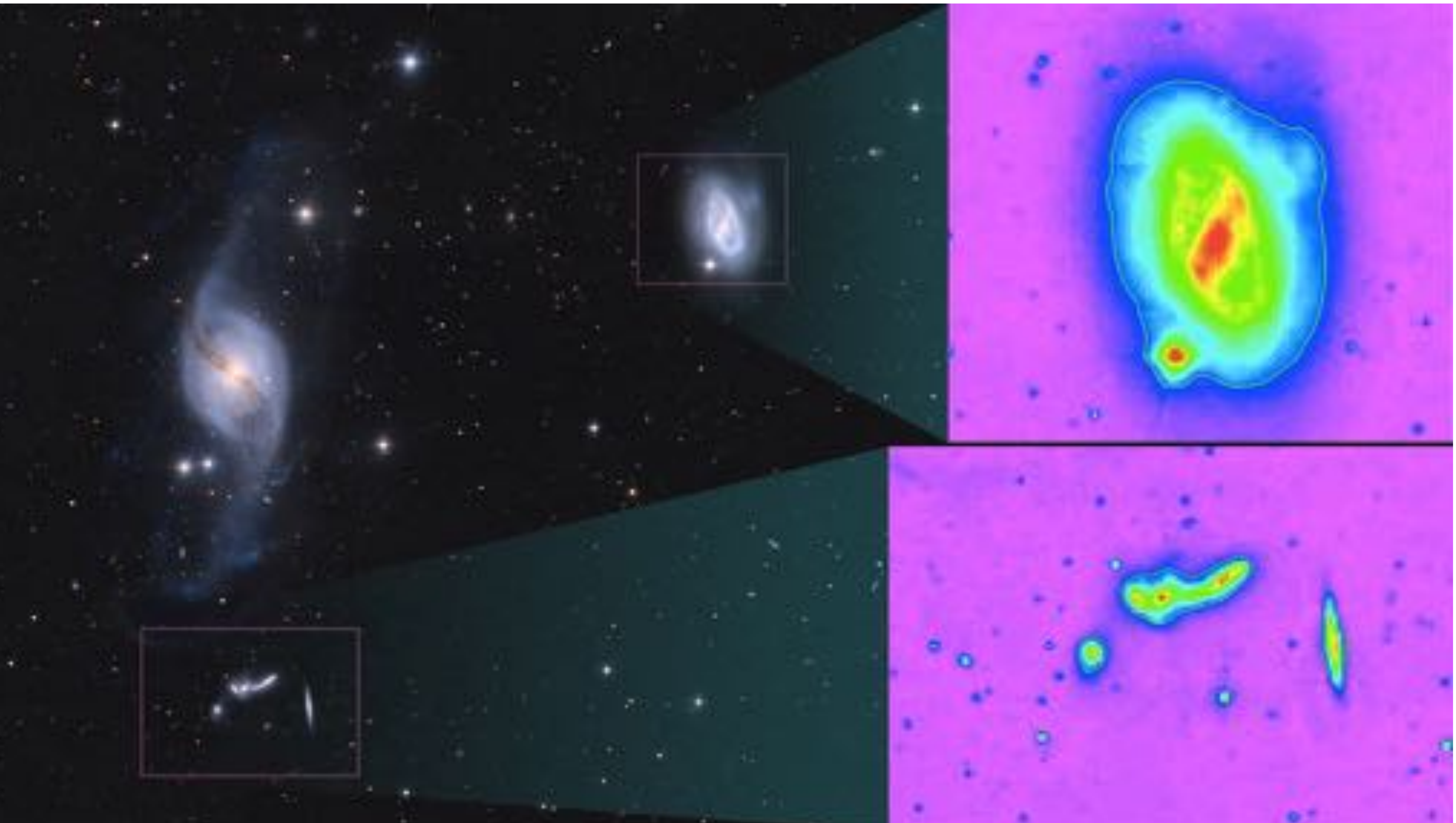
NIH

Resting state and task based fMRI
MEG, EEG

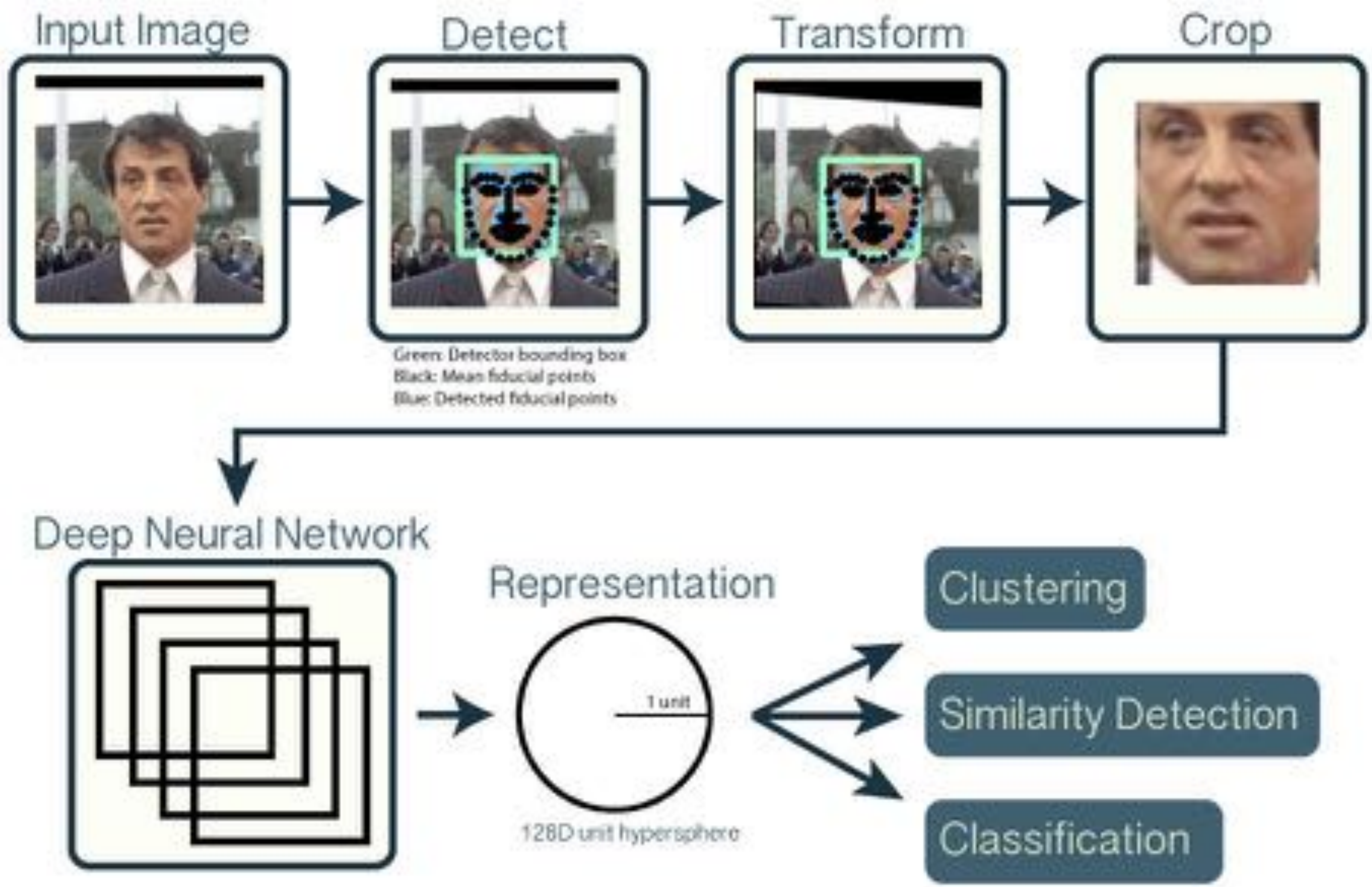
390GB/person
several hundred people

Voxels
4D NIfTI
Dicom

From galaxies to lesions



OpenFace



https://cmusatyalab.github.io/openface/

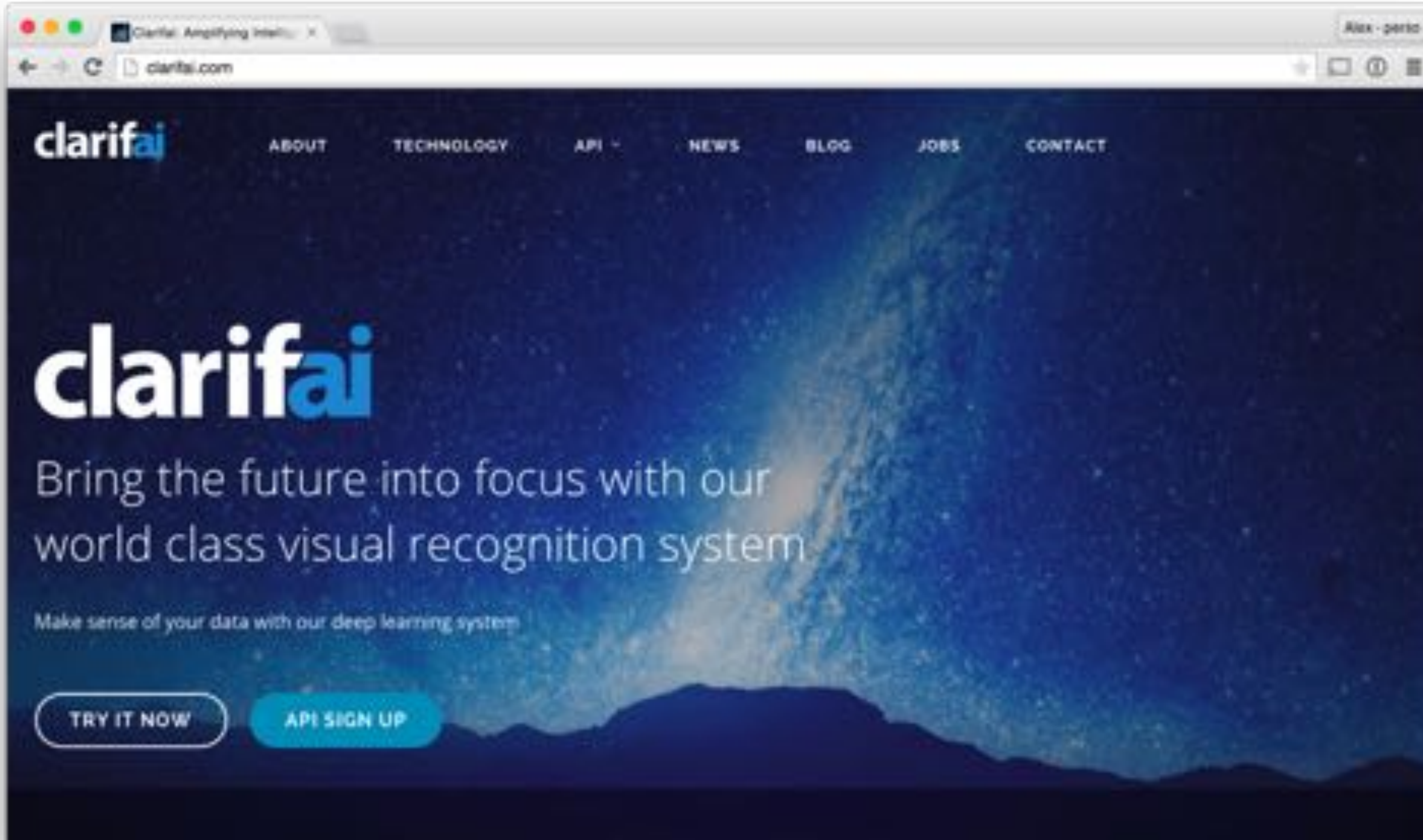
Another use for the face API

Faces masked in Streetview

Translate - on your phone

Indonesian

clarifai: MLaaS



What Deep Learning Can Not Do

There was a guy from Pansy
Who had a chest clinic very fancy
Said he with a hiss,
That M Tuberculosis

What Deep Learning Can Not Do

There was a guy from Pansy
Who had a chest clinic very fancy
Said he with a hiss,
That M Tuberculosis
Can cause anything except pregnancy

What Deep Learning Can Not Do

There was a guy from Pansy
Who had a chest clinic very fancy
Said he with a hiss,
That M Tuberculosis
Can cause anything except pregnancy

There was a guy from LSST
[xyz] he **[lmn] rhyme-with-tea**
Said he with a **…ing**
That Deep Learning
Can solve anything except **[]**

Summary

- CNNs are taking over, especially the image domain
- Can come up with features not thought of before
- Abstracted libraries and visualizations available
- Over-learning can be a problem:
  - augmentation
  - adversarial examples/generative networks
- Should ensure they do not become convoluted
- Deep and wide networks may prove to be a boon

Credits:

Developing the machinery for EDRN/MCL (JPL informatics)

David Liu
Dan Crichton

Projects with Fabian Gieseke, Brian Bue, Umaa Rebrapragada

Copious dependance on blogs and other websites

Adit Deshpande

https://adeshpande3.github.io/adeshpande3.github.io/A-Beginner%27s-Guide-To-Understanding-Convolutional-Neural-Networks/

Christopher Olah

http://colah.github.io/posts/2014-03-NN-Manifolds-Topology/
http://colah.github.io/posts/2014-07-Conv-Nets-Modular/

Adil Moujahid

http://adilmoujahid.com/posts/2016/06/introduction-deep-learning-python-caffe/

https://research.googleblog.com/2016/06/wide-deep-learning-better-together-with.html