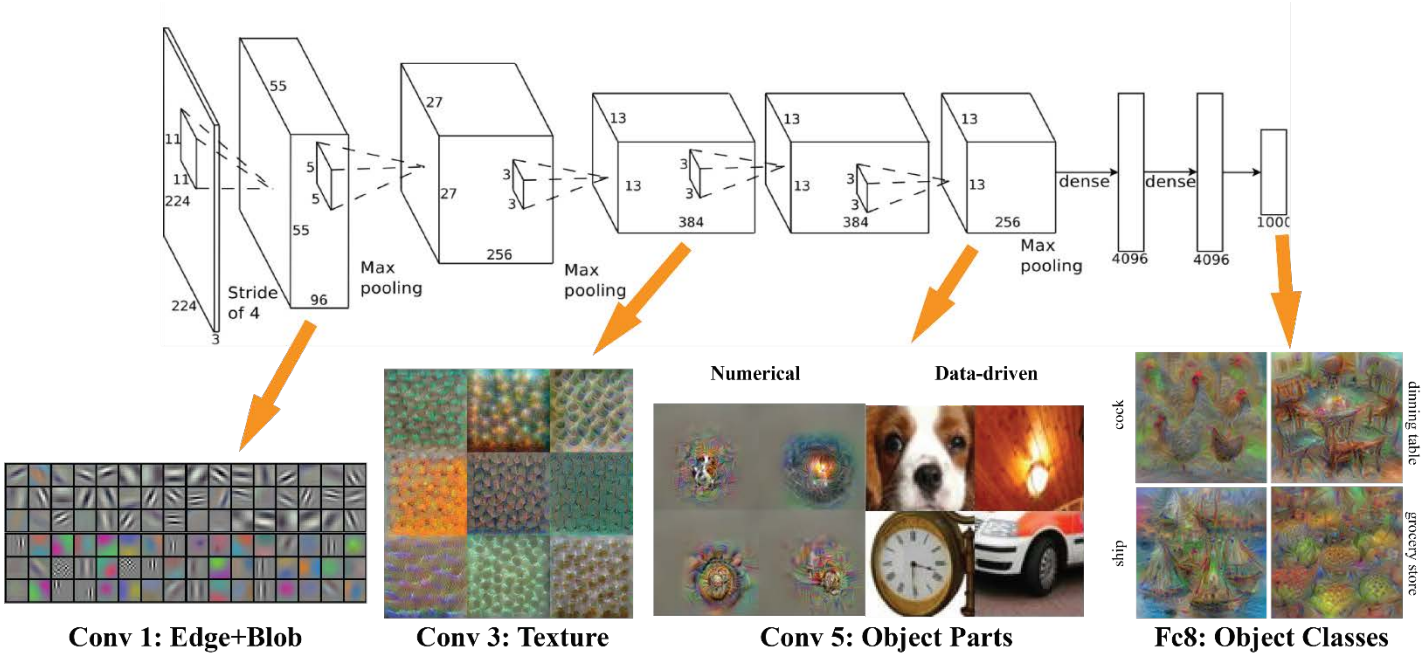
ECSE-626



Intro to Deep Learning and Convolutional Neural Networks

The Hype

Forbes / Tech

MAR 29, 2016 @ 01:13 PM 2,357 VIEWS

How Deep Is Your Learning?

AlphaGo Zero: The Most Significant Research Advance in AI

Oct 2017 Silver KDnuggets Blog

◂ Previous post Next post ▸

Like 361 Share 361 Share 406 G+ Share 164

Tags: AI, AlphaGo, DeepMind, Google, Reinforcement Learning, Xavier Amatriain

The previous version of AlphaGo beat the human world champion in 2016. The new AlphaGo Zero beat the previous version by 100 games to 0, and learned Go completely on its own. We examine what this means for AI.

Artificial intelligence

Million-dollar babies

The Economist

As Silicon Valley fights for talent, universities struggle to hold on to their stars

Apr 2nd 2016 | SAN FRANCISCO | From the print edition

Artificial intelligence cybernetics machine learning technology film

Kristen Stewart co-authored a paper on style transfer and the AI community lost its mind

Posted Jan 19, 2017 by John Mannes (@JohnMannes)

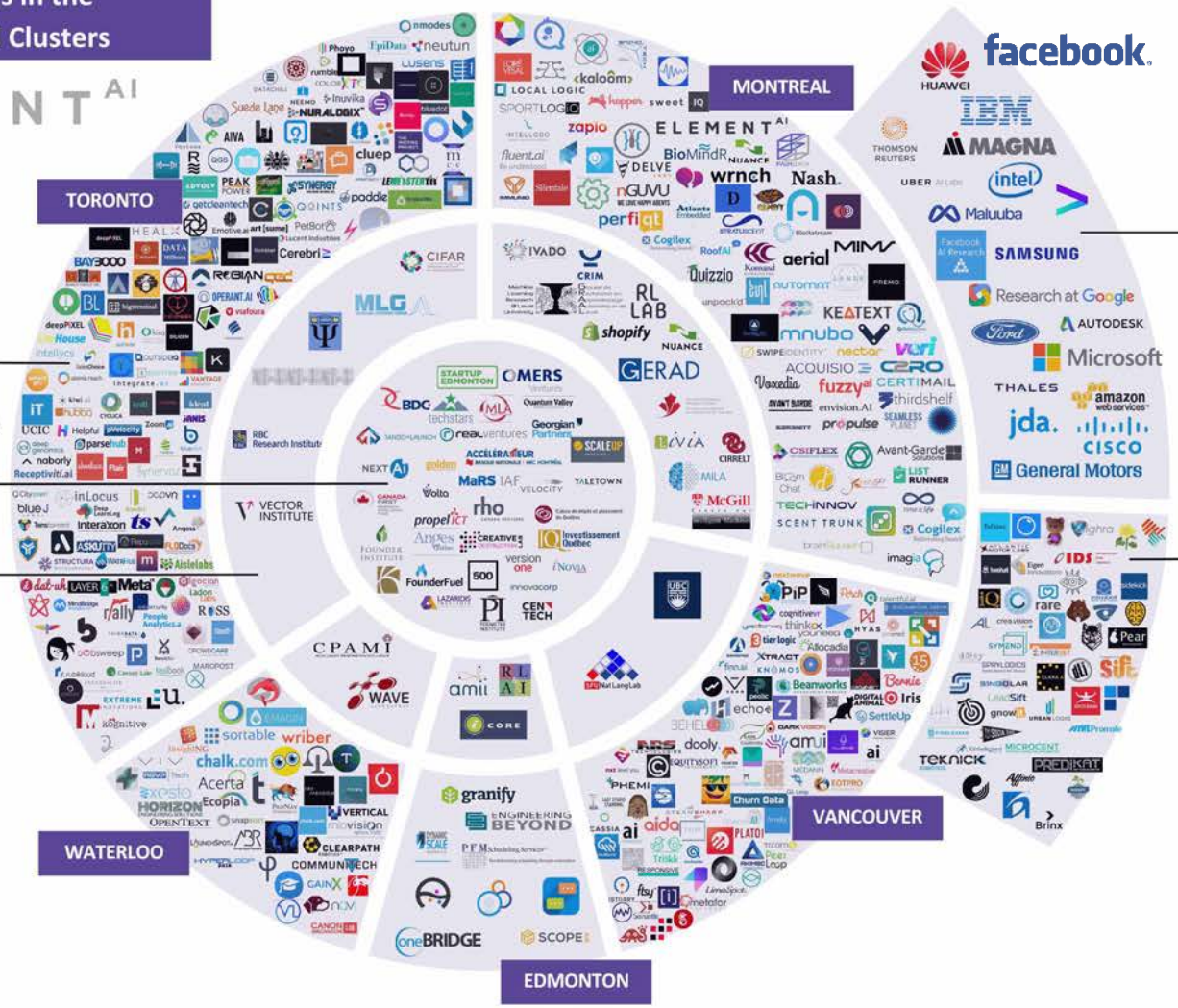
Next Story

Image Credit to: hdwallpapers.in

Many slides in this lecture are borrowed or adapted from those provided by Prof. James Clark, Jason Taylor, Brandon Rohrer

Top Players in the Canadian AI Clusters

ELEMENT AI



STARTUP & ENTERPRISE COUNT PER LOCATION	
195+	TORONTO
100+	VANCOUVER
90+	MONTREAL
50+	WATERLOO-KITCHENER
10+	EDMONTON
60+	ALL OTHERS

2017-06-11
For updates see jfgagne.ai

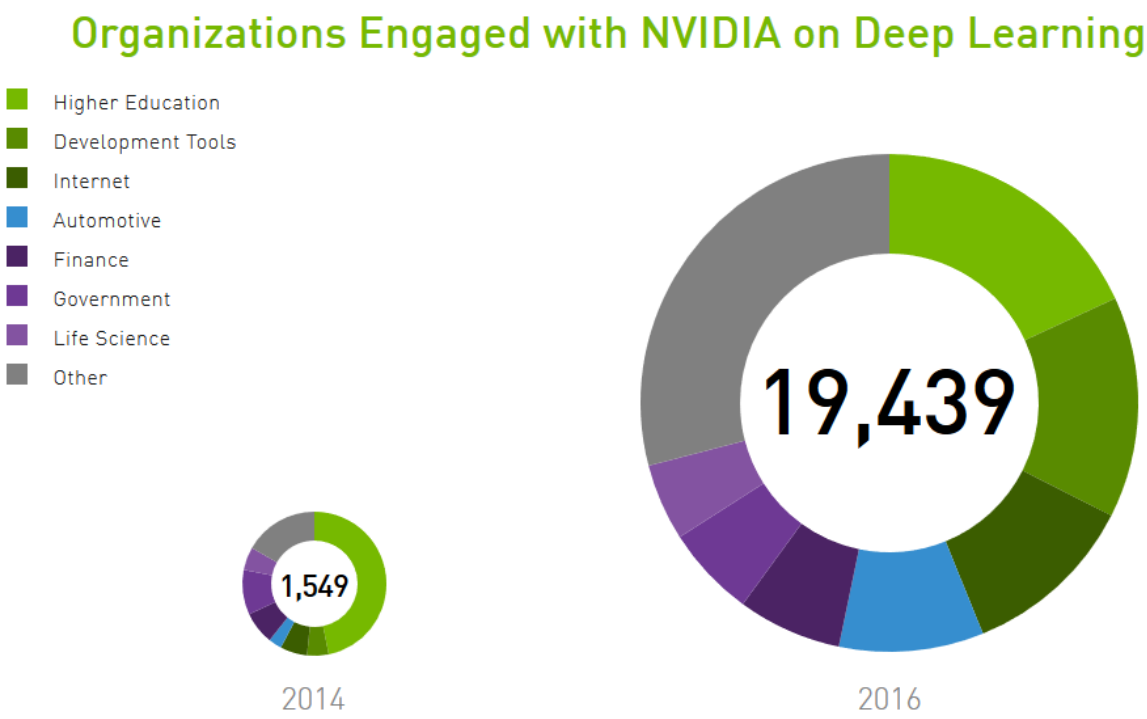
https://www.nanalyze.com/2017/07/9-canadian-ai-startups/

$36.8 billion. That's the revenue AI software will generate by 2020 as forecasted by Tractica Research[1].

Currently, any work that's pushing the forward edge of technology—including automotive, analytics,or healthcare—has big data at its core. However, that data is only as valuable as the insights we pull from it.

More than 19,000 companies are currently using deep learning to advance their respective industries, solving what was once unsolvable.

[1] Study can be found here.



https://www.nvidia.com/en-us/deep-learning-ai/industries/

A NEW ERA OF COMPUTING

PC INTERNET
WinTel, Yahoo!
1 billion PC users

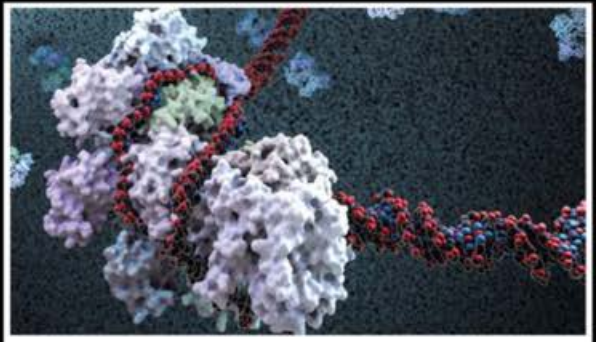
MOBILE-CLOUD
iPhone, Amazon AWS
2.5 billion mobile users

AI & IOT
Deep Learning, GPU
100s of billions of devices

1995 2005 2015

AI FOR EVERYONE

AI will Revolutionize Transportation



AI will Revolutionize Healthcare

AI will Revolutionize Society

https://blogs.nvidia.com/blog/2016/10/24/intelligent-industrial-revolution/

What is Deep Learning?

- Deep learning is a class of brain-inspired machine learning algorithms that:
  - use a cascade of multiple layers of nonlinear processing units for feature extraction and transformation.
  - learn in supervised (e.g., classification) and/or unsupervised (e.g., pattern analysis) manners.
  - learn multiple levels of representations that correspond to different levels of abstraction.
  - use some form of gradient descent for training via backpropagation

https://en.wikipedia.org/wiki/Deep_learning
https://machinelearningmastery.com/what-is-deep-learning/

What is Deep Learning?

Traditional Machine Learning Flow

Deep Learning Flow

Why Deep Learning?

- Classic framework lacks a mechanism to jointly optimize parameters in all stages
  - Training is sometimes limited to the final classifier
- Main strengths of deep learning:
  - Arbitrary number of intermediate representations
    - This is what makes it "deep"
    - Features are learned
    - Features increase in scale and/or abstraction with depth
  - **End-to-end training**
    - Jointly optimize all intermediate representations for a given task

Why Deep Learning?

- Compared to handcrafted features, it usually enjoys much higher accuracy:
  - e.g. AlexNet blew away the 2012 ImageNet Competition

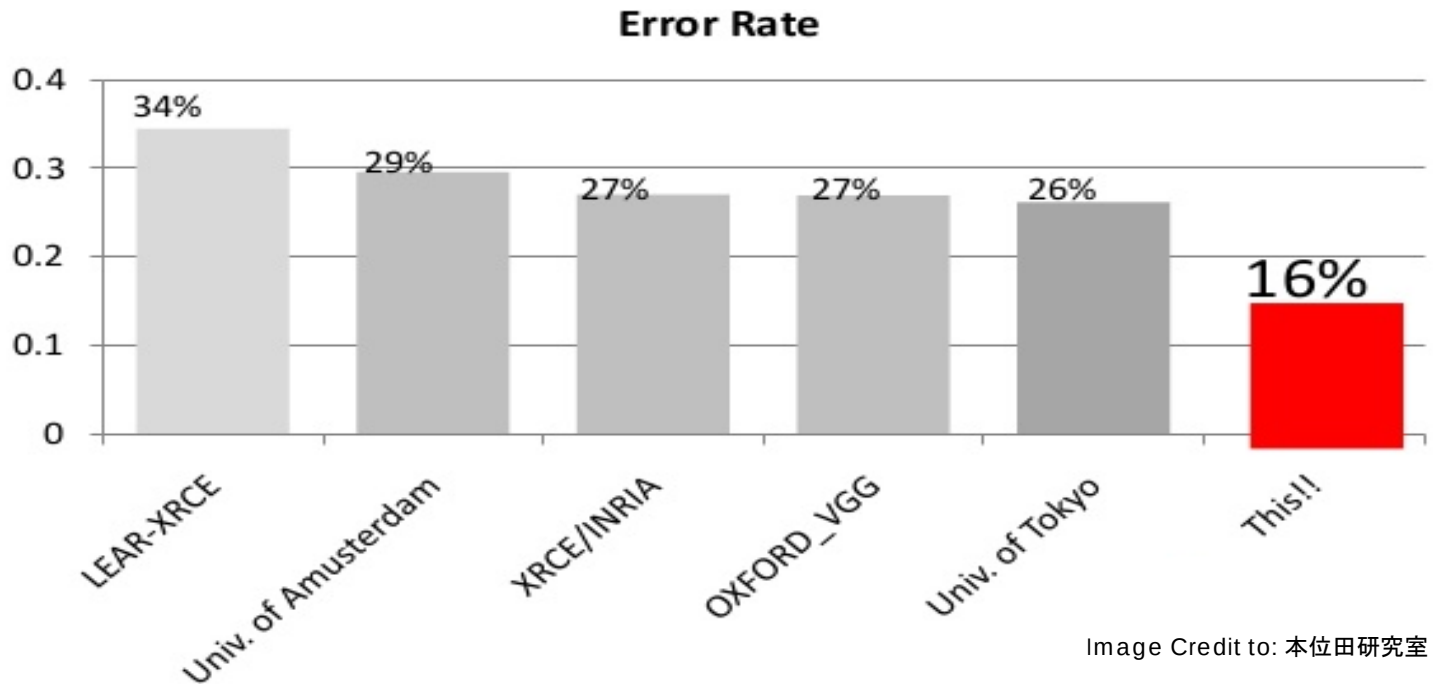
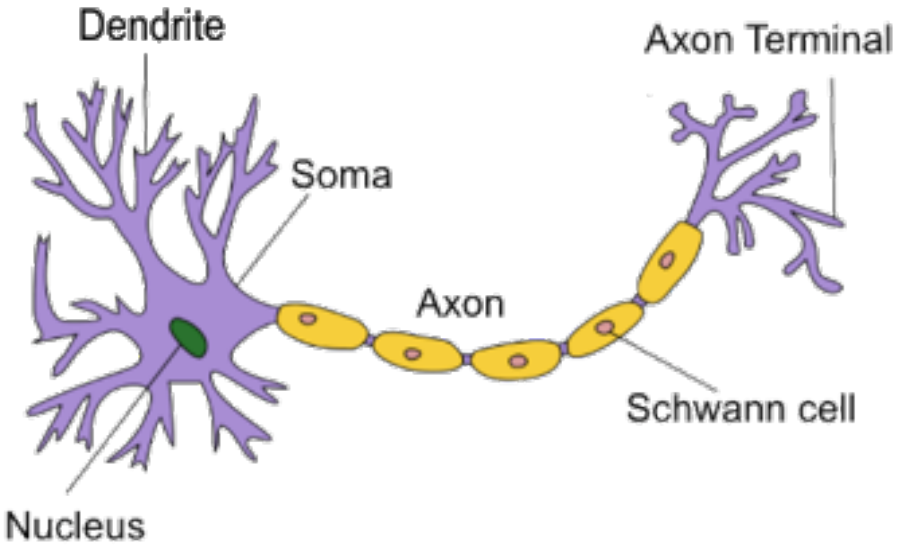


Image Credit to: 本位田研究室

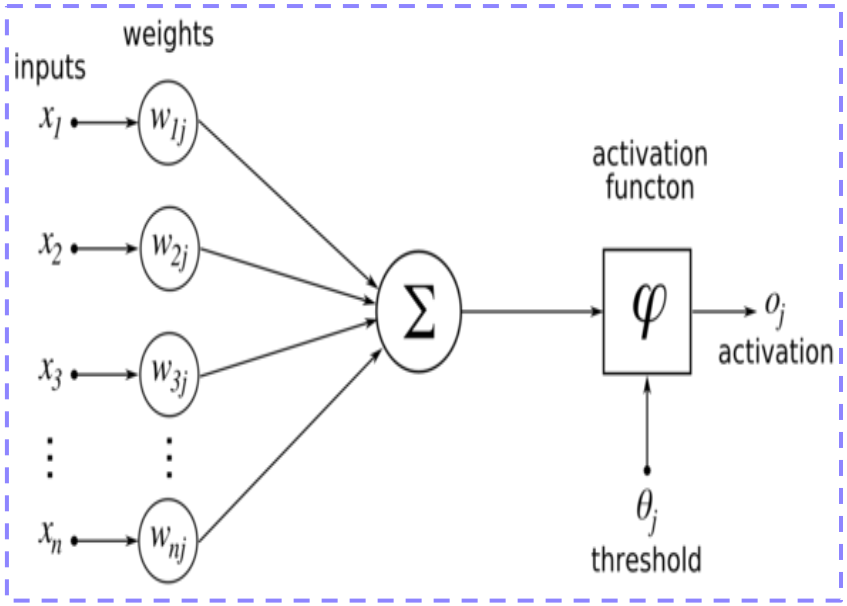
  - Demo: a deep net in action

Building block: Neurons (biological v.s. artificial)



Structure of a typical neuron

(source: Wikipedia)



the *Perceptron Model (Rosenblatt 1957)*

x can be considered as 'signals' from other neurons, w can be considered as 'synaptic strengths' (connection strengths between dendrites and axon terminals)

Neural Networks (biological v.s. artificial)

image source: www.shutterstock.com

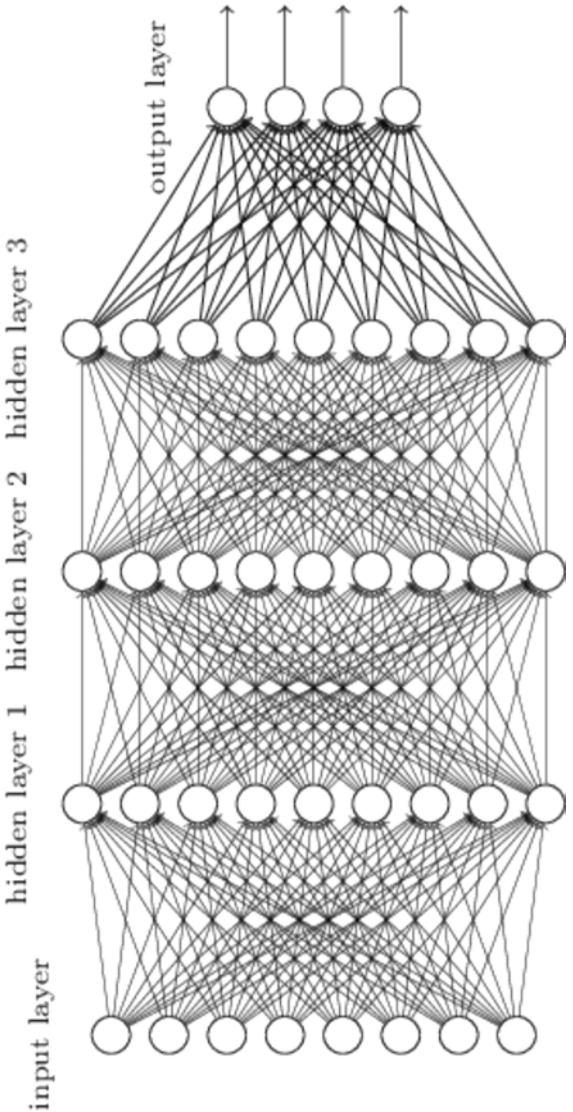
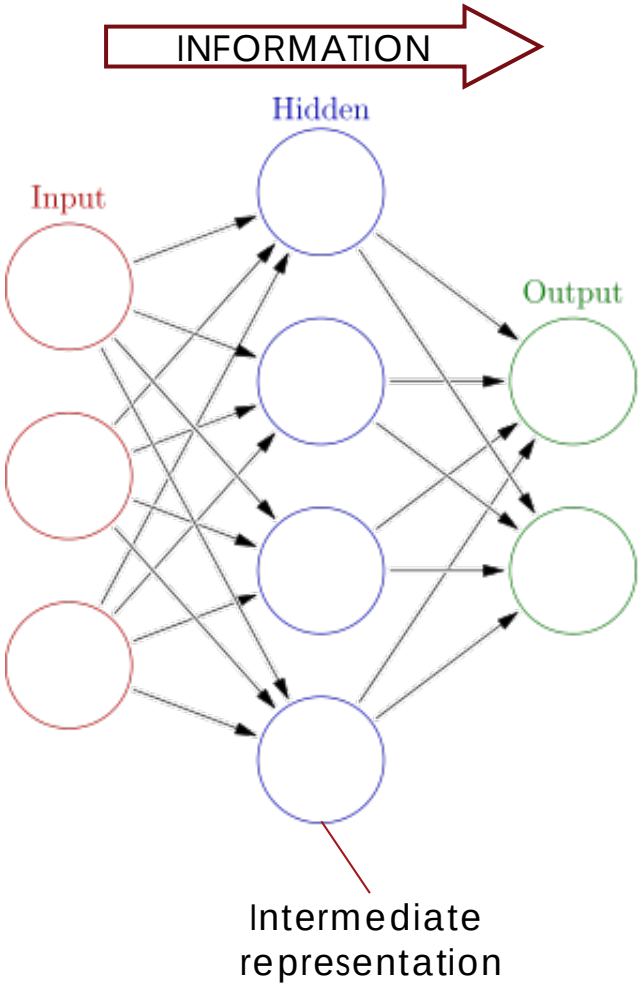


image source: https://www.mathworks.com

Artificial Neural Networks (ANN)

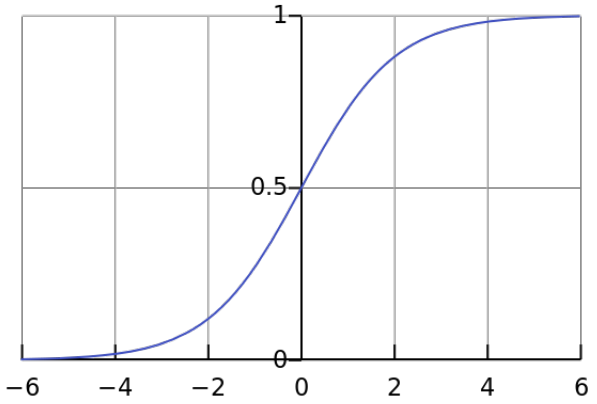


- Biologically-inspired highly-parallel model
- A network of nodes ("neurons") with weighted connections ("synapses")
- Each node outputs a non-linear activation function applied to a weighted sum of its inputs
- The network is trained by adjusting these weights

Activation Functions (nonlinearity)

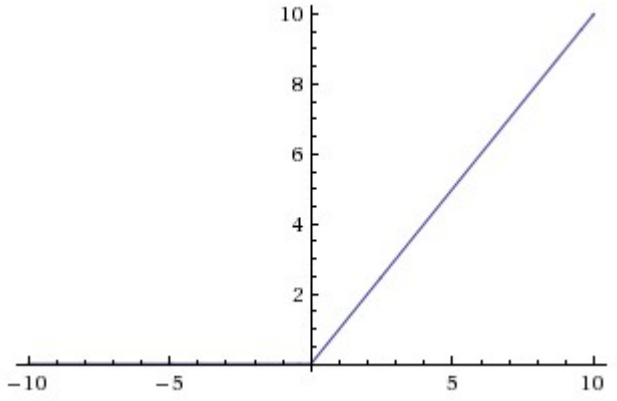
- Sigmoid function:
  - Biologically inspired

$$S(t) = \frac{1}{1 + e^{-t}}.$$



- Rectified Linear Unit (ReLU)
  - Computationally “nicer”

$$f(x) = \max(0, x)$$



McGill ECE Grad! (B.Eng Hons 2001, M.Eng 2004 supv. by Prof. Clark)

Nair, V. and G. E. Hinton. *Rectified linear units improve restricted boltzmann machines*. In Proc. 27th International Conference on Machine Learning, 2010

Activation Functions (nonlinearity)

- The non-linear activation function is crucial. otherwise, only linear functions can be approximated.

- From the bottom to top layers, deep nets are able to capture the ‘compositionality’ of the real world.

Depth Matters

- As the network size increases – more layers, more units per layer – its representational power increases
  - Since the number of parameters explodes with the increased depth, deep nets are hard to train (problems like vanishing gradients occur)
  - Not very long ago, most neural networks had only one or two layers …

Depth Matters

- Things changed, thanks to:
  - the huge amount of data that has become available
  - recent advances in computing (GPUs)

Image credit: www.techadvisor.co.uk

Depth Matters

- Efficient training algorithms

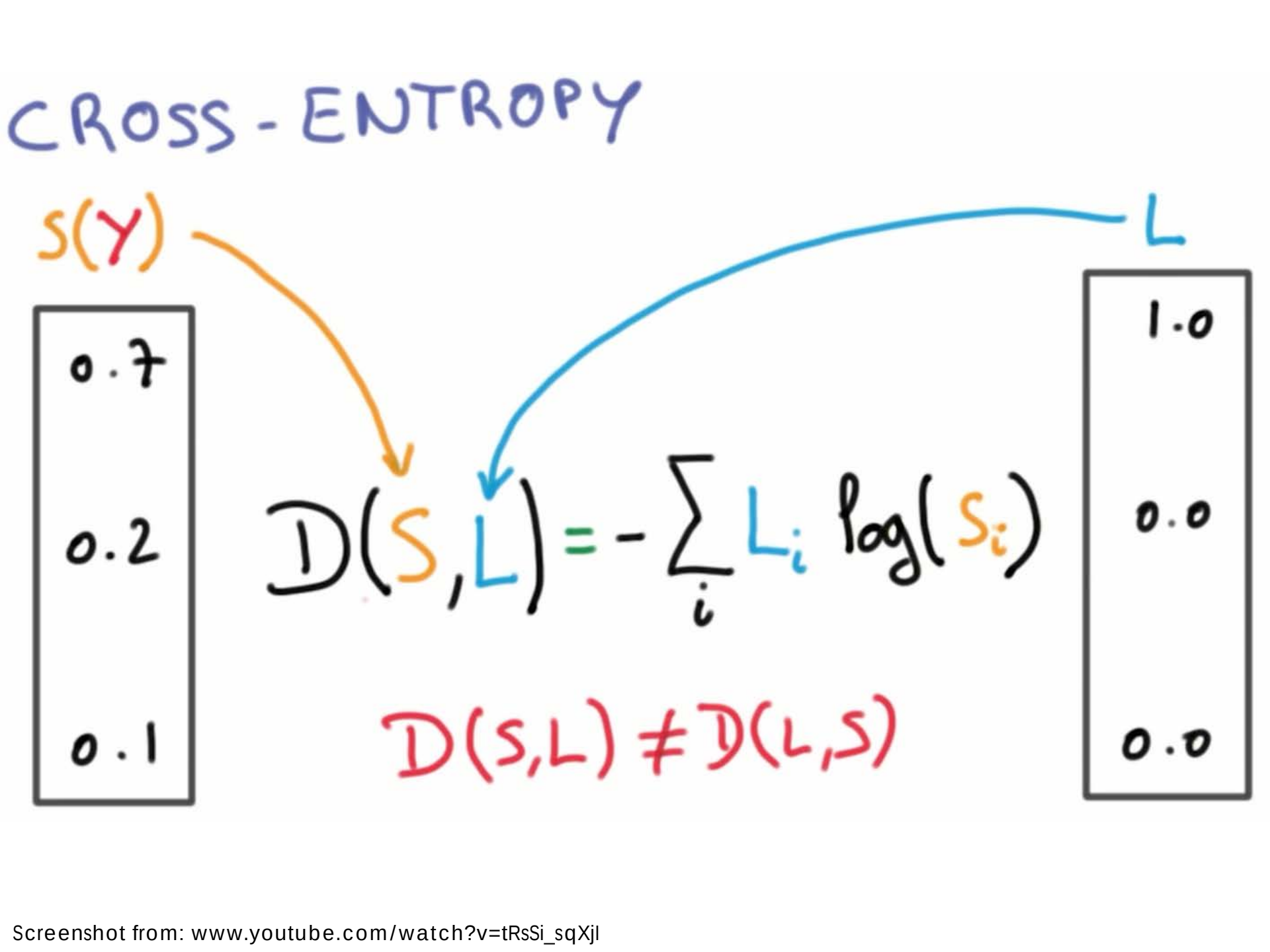
G. E. Hinton, S. Osindero, and Y.-W. Teh. A fast learning algorithm for deep belief nets. Neural computation, 18(7):1527–1554, 2006

Y. Bengio, P. Lamblin, D. Popovici, H. Larochelle, et al. Greedy layer-wise training of deep networks. NIPS 2007

C. P. MarcAurelio Ranzato, S. Chopra, and Y. LeCun. Efficient learning of sparse representations with an energy-based model. NIPS 2007

Training

- Our goal is to ‘wiggle’ the millions of parameters to bring the network output into alignment with the ground truth (supervised).
- We use ‘cost function’ to measure of how good the alignment is.
  - e.g. mean-square error for a regression task
  - e.g. cross-entropy for a classification task



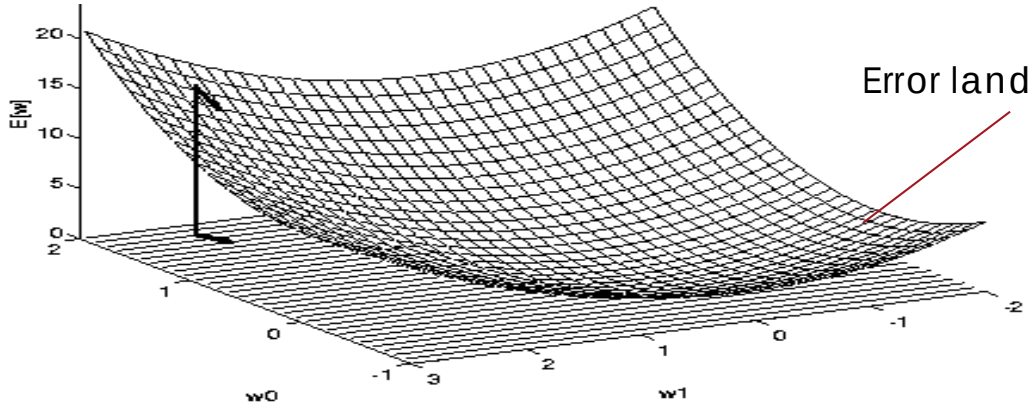
Screenshot from: www.youtube.com/watch?v=tRsSi_sqXjI

Training

- Starting with small random values, weights are updated iteratively to reduce the cost function
  - Show the network training examples
  - Update the weights slightly according to an optimization algorithm (e.g. gradient descent)
  - Repeat

Gradient Descent

- Training is done by iteratively minimizing the cost function on the training set



Gradient

$$\nabla E[\vec{w}] \equiv \left[\frac{\partial E}{\partial w_0}, \frac{\partial E}{\partial w_1}, \cdots \frac{\partial E}{\partial w_n}\right]$$

Training rule:

$$\Delta \vec{w} = -\eta \nabla E[\vec{w}]$$

i.e.,

$$\Delta w_i = -\eta \frac{\partial E}{\partial w_i}$$

Backpropagation in ANN using gradient descent

- Gradient descent updates weights based on the partial derivative of the cost function w.r.t. each weight
- Starting from the output and going *back* through the net, repeatedly apply the **chain rule** to find derivatives w.r.t. each weight **(reuse previous partial results to prevent the computation from exploding exponentially).**
- Then simply update the weights by gradient descent:

$$w_{ij} \leftarrow w_{ij} - \eta \frac{\delta E}{\delta w_{ij}}$$

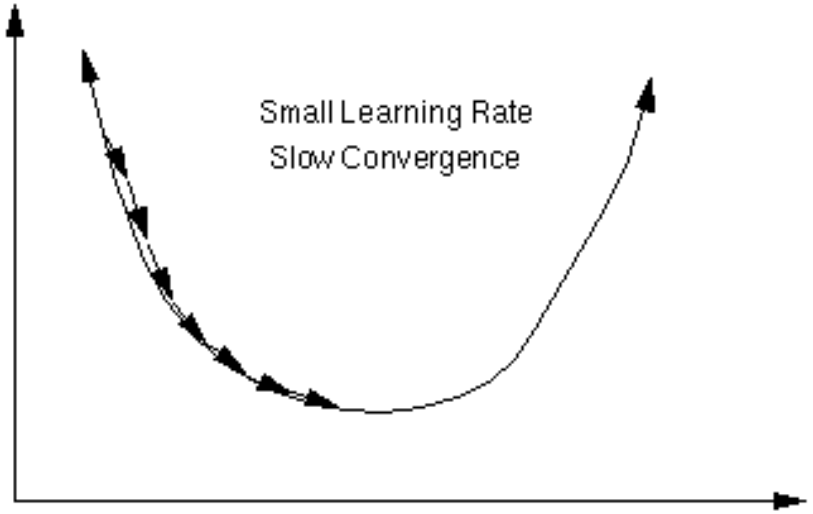
learning rate > 0

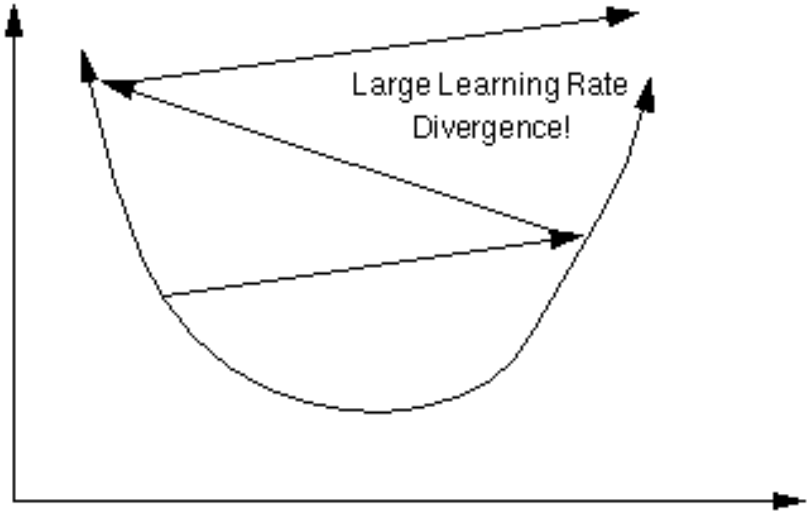
Stochastic Gradient Descent

- Rather than run all training examples through the net and then updating the weights, we can do better sooner:
  - Run a “batch” of training examples
    - Batch size can even be 1
    - All training examples are run before repeating
  - Update weights based on the cost function on that batch (an approximation of the full set cost function)
  - Significantly speeds up training

A Word on Learning Rate

- Performance is sensitive to learning rate
  - In practice: set as high as possible (for fast training) until it stops working, then lower it again
  - Decay throughout training
    - Intuitively, make big steps when far from an optimal solution and fine-tune at the end





Convolutional Neural Networks (ConvNet)

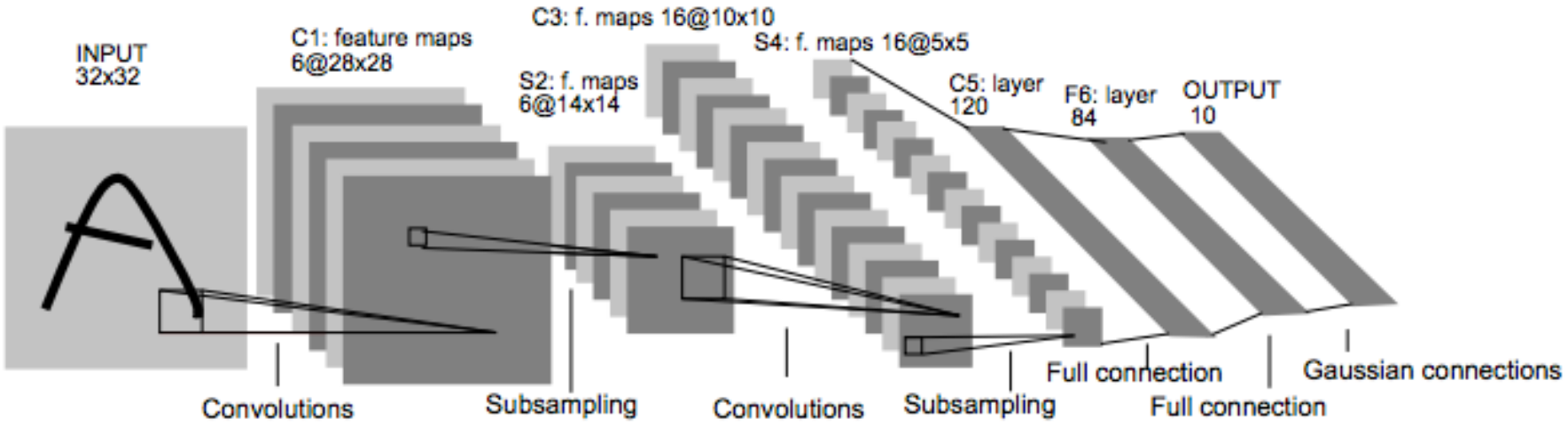
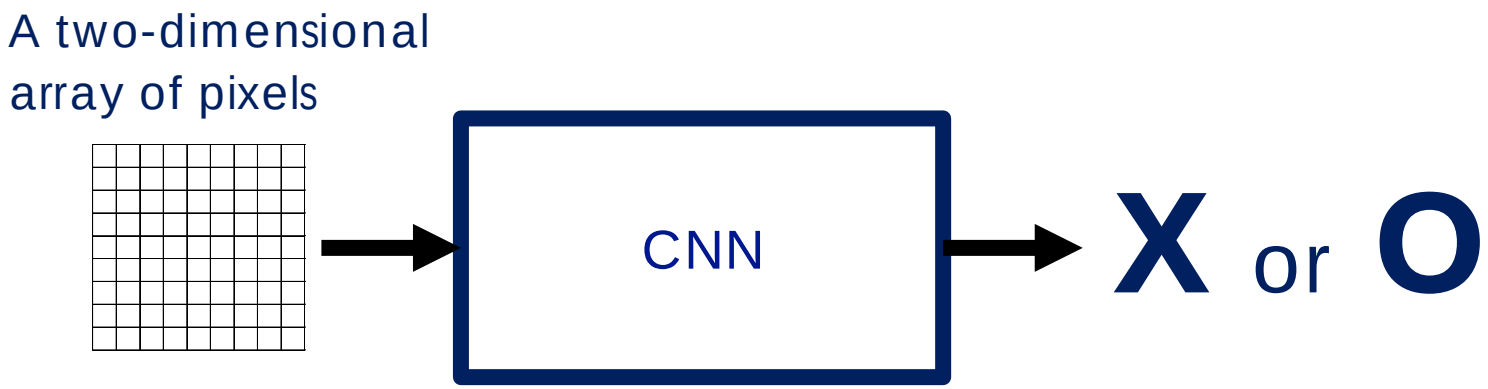


Fig. 2. Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition. Each plane is a feature map, i.e. a set of units whose weights are constrained to be identical.

Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE 86(11): 2278–2324, 1998.

A simple ConvNet: X's or O's

Says whether a picture is of an X or an O



Note: the toy example slides (from this slide to Slide 53) are adapted from those provided by Brandon Rohrer

For example

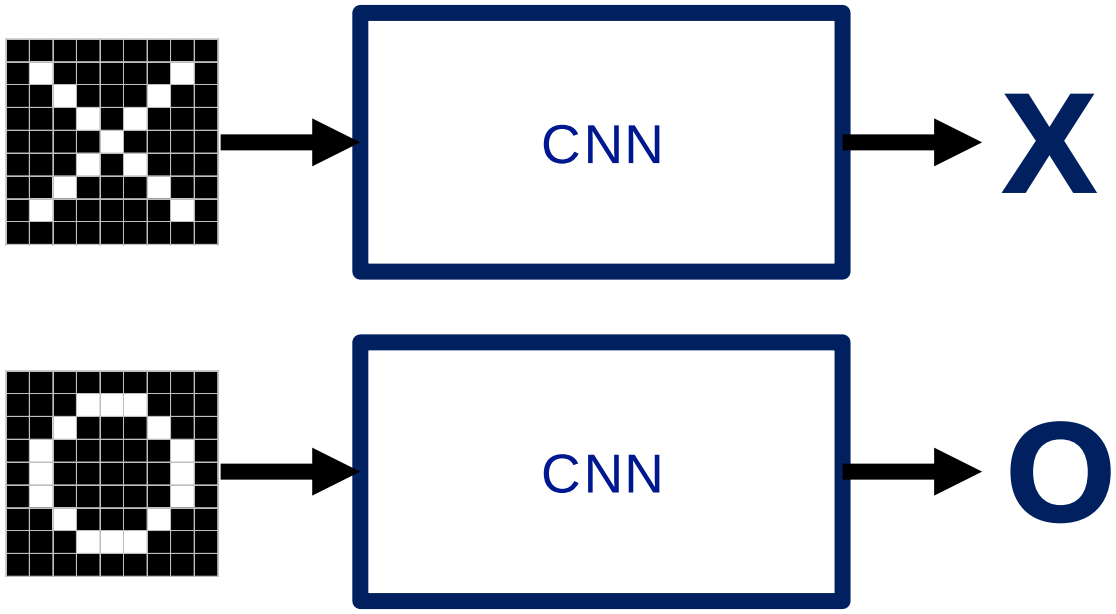


Figure adapted from the slides by Brandon Rohrer

Challenge of within-class variances

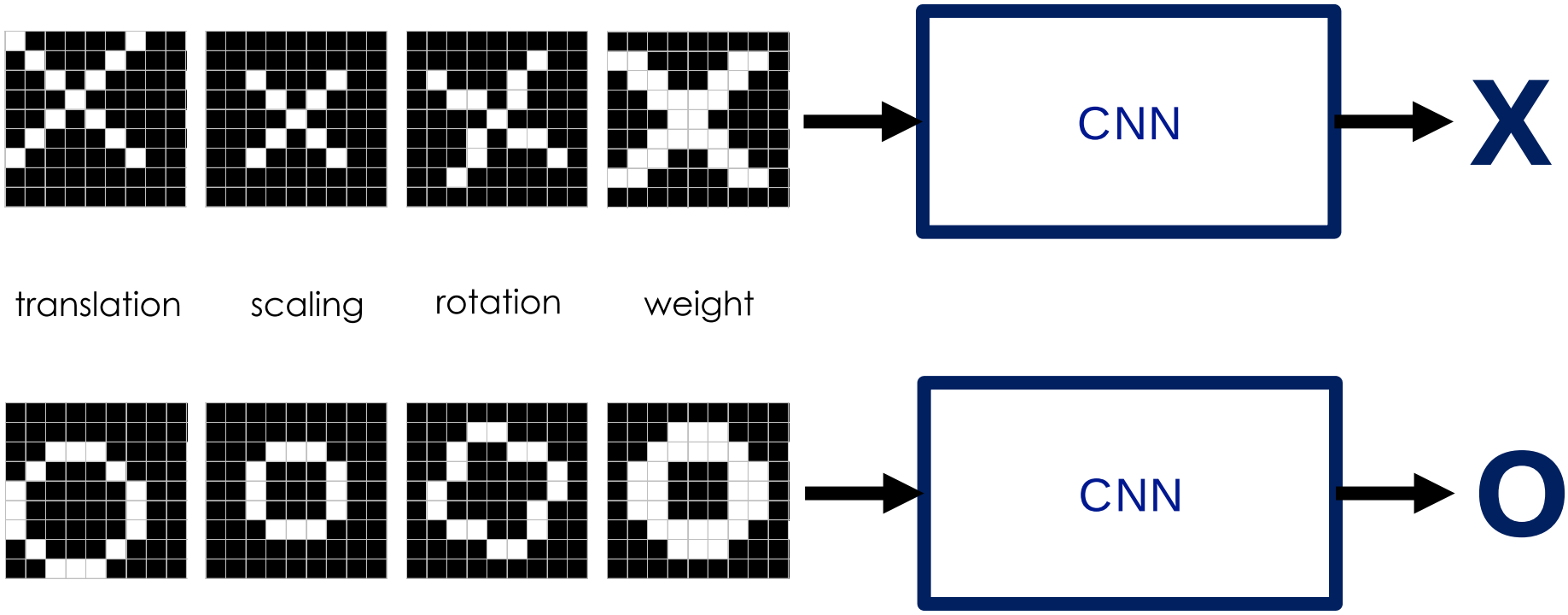


Figure adapted from the slides by Brandon Rohrer

Deciding is hard

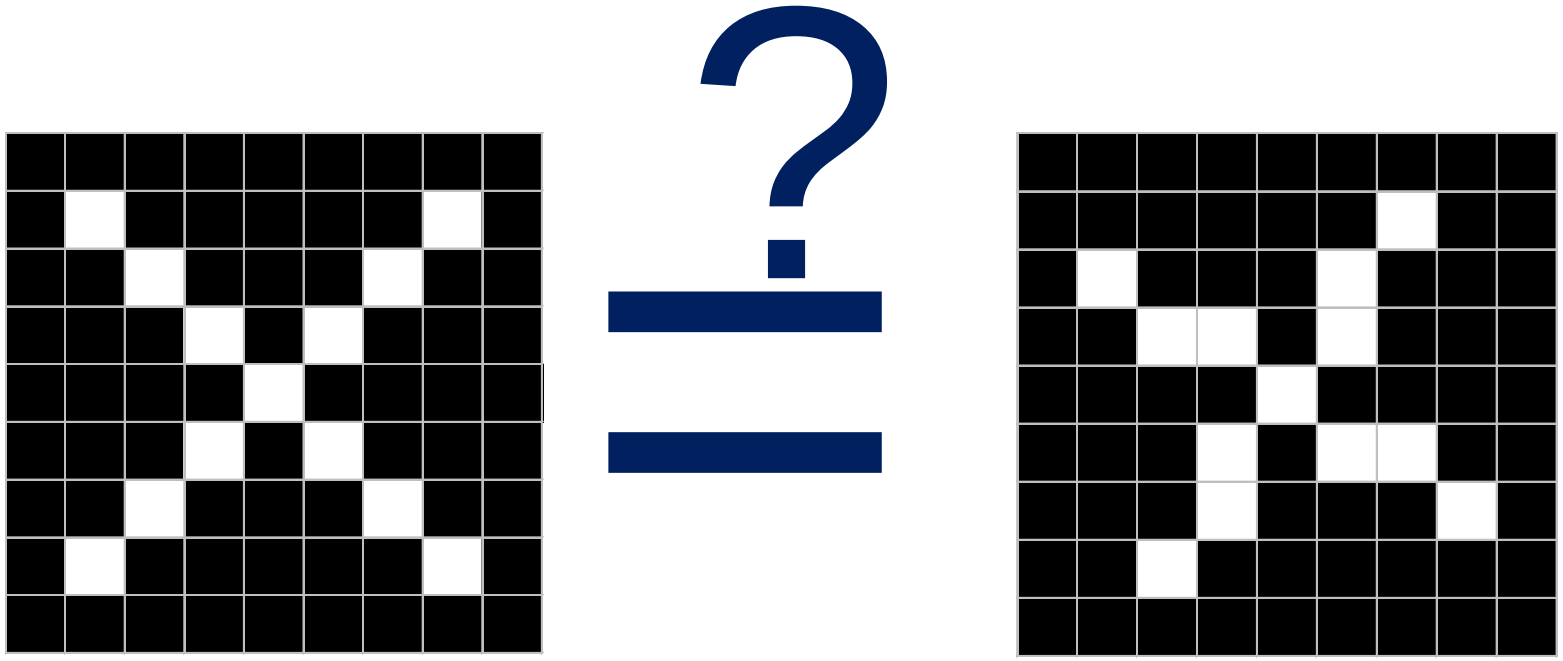


Figure adapted from the slides by Brandon Rohrer

What computers see

-1	-1	-1	-1	-1	-1	-1	-1	-1
-1	1	-1	-1	-1	-1	-1	1	-1
-1	-1	1	-1	-1	-1	1	-1	-1
-1	-1	-1	1	-1	1	-1	-1	-1
-1	-1	-1	-1	1	-1	-1	-1	-1
-1	-1	-1	1	-1	1	-1	-1	-1
-1	-1	1	-1	-1	-1	1	-1	-1
-1	1	-1	-1	-1	-1	-1	1	-1
-1	-1	-1	-1	-1	-1	-1	-1	-1

-1	-1	-1	-1	-1	-1	-1	-1	-1
-1	-1	-1	-1	-1	-1	1	-1	-1
-1	1	-1	-1	-1	1	-1	-1	-1
-1	-1	1	1	-1	1	-1	-1	-1
-1	-1	-1	-1	1	-1	-1	-1	-1
-1	-1	-1	1	-1	1	1	-1	-1
-1	-1	-1	1	-1	-1	-1	1	-1
-1	-1	1	-1	-1	-1	-1	-1	-1
-1	-1	-1	-1	-1	-1	-1	-1	-1

Figure adapted from the slides by Brandon Rohrer

What computers see

-1	-1	-1	-1	-1	-1	-1	-1	-1
-1	X	-1	-1	-1	-1	X	X	-1
-1	X	X	-1	-1	X	X	-1	-1
-1	-1	X	1	-1	1	-1	-1	-1
-1	-1	-1	-1	1	-1	-1	-1	-1
-1	-1	-1	1	-1	1	X	-1	-1
-1	-1	X	X	-1	-1	X	X	-1
-1	X	X	-1	-1	-1	-1	X	-1
-1	-1	-1	-1	-1	-1	-1	-1	-1

Figure adapted from the slides by Brandon Rohrer

Computers are literal

-1	-1	-1	-1	-1	-1	-1	-1	-1
-1	1	-1	-1	-1	-1	-1	1	-1
-1	-1	1	-1	-1	-1	1	-1	-1
-1	-1	-1	1	-1	1	-1	-1	-1
-1	-1	-1	-1	1	-1	-1	-1	-1
-1	-1	-1	1	-1	1	-1	-1	-1
-1	-1	1	-1	-1	-1	1	-1	-1
-1	1	-1	-1	-1	-1	-1	1	-1
-1	-1	-1	-1	-1	-1	-1	-1	-1

-1	-1	-1	-1	-1	-1	-1	-1	-1
-1	-1	-1	-1	-1	-1	1	-1	-1
-1	1	-1	-1	-1	1	-1	-1	-1
-1	-1	1	1	-1	1	-1	-1	-1
-1	-1	-1	-1	1	-1	-1	-1	-1
-1	-1	-1	1	-1	1	1	-1	-1
-1	-1	-1	1	-1	-1	-1	1	-1
-1	-1	1	-1	-1	-1	-1	-1	-1
-1	-1	-1	-1	-1	-1	-1	-1	-1

Figure adapted from the slides by Brandon Rohrer

ConvNets match pieces of the image

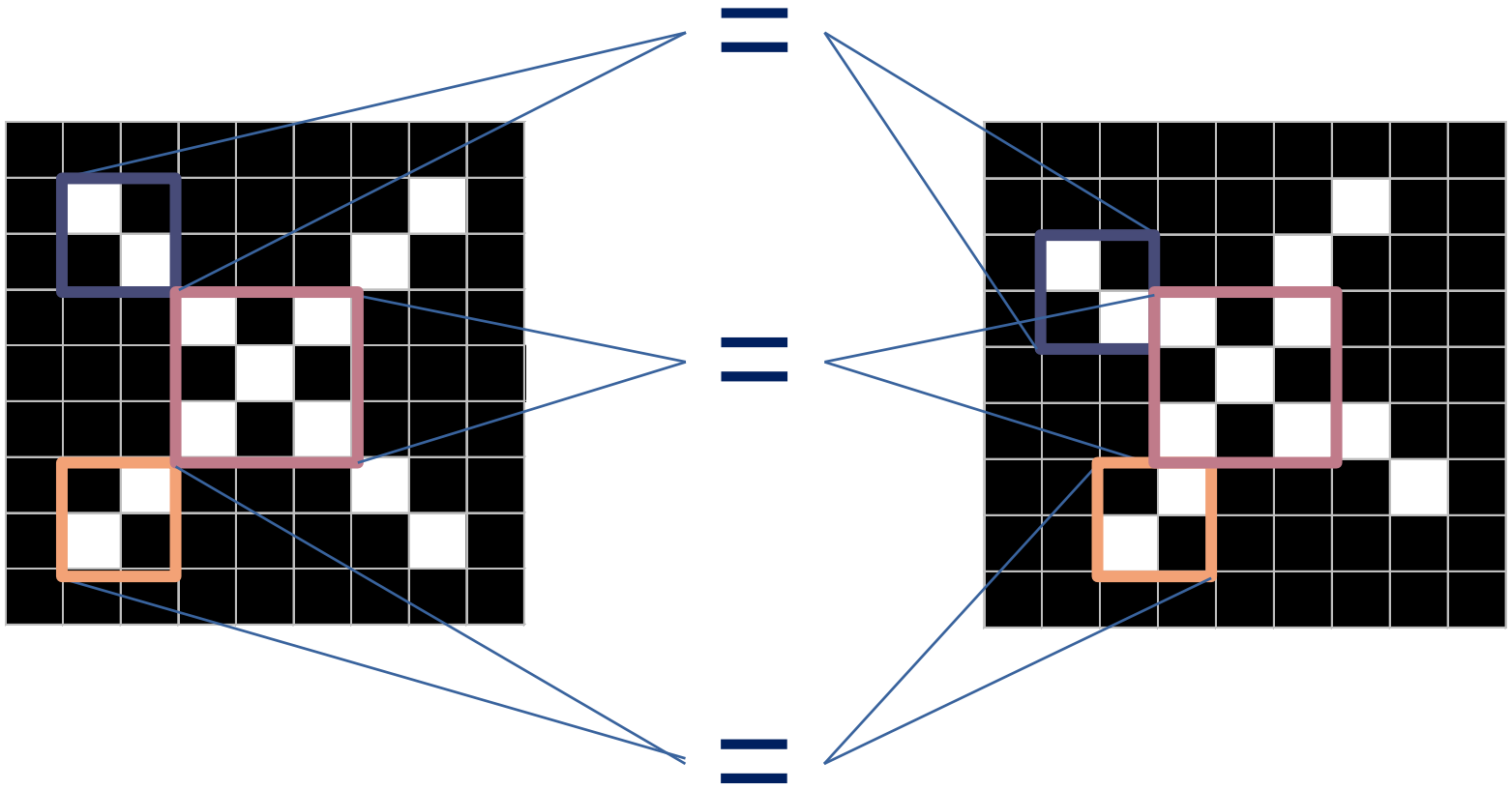
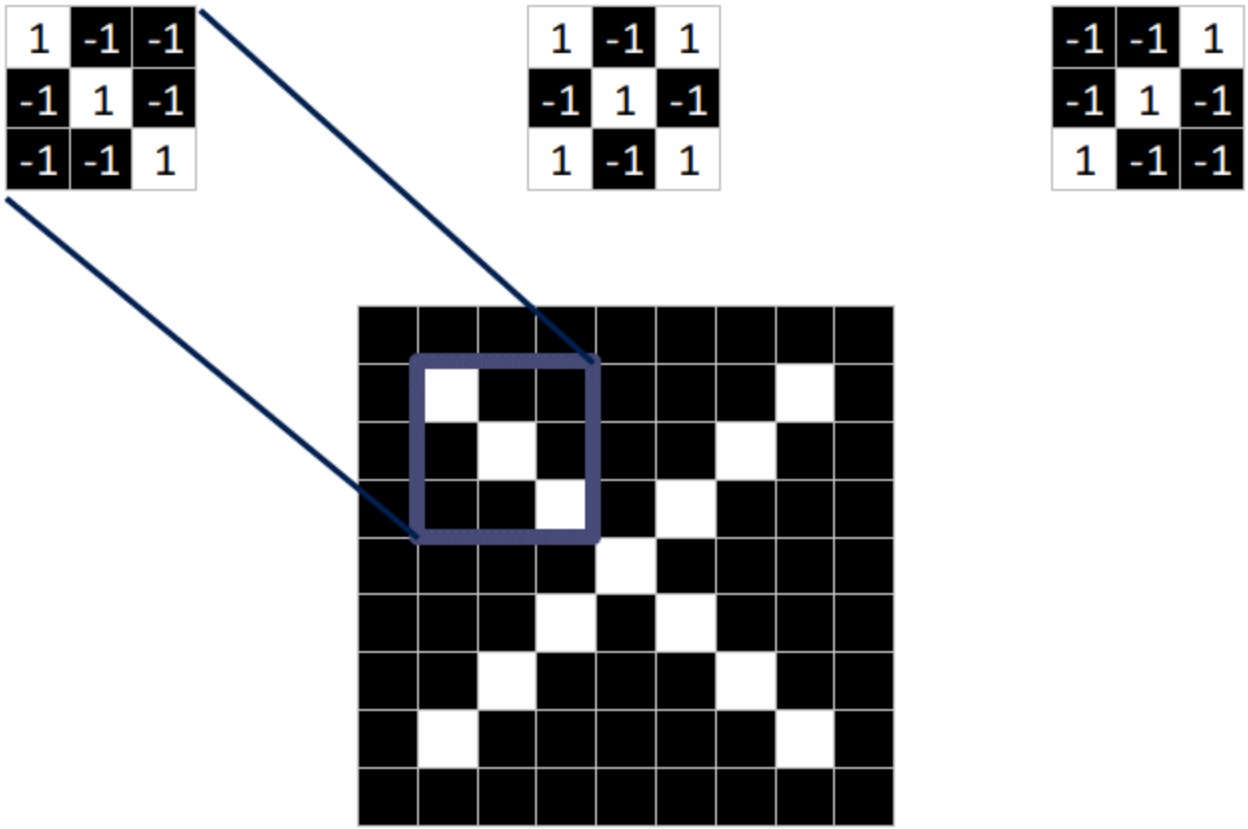


Figure adapted from the slides by Brandon Rohrer

Filters match pieces of the image



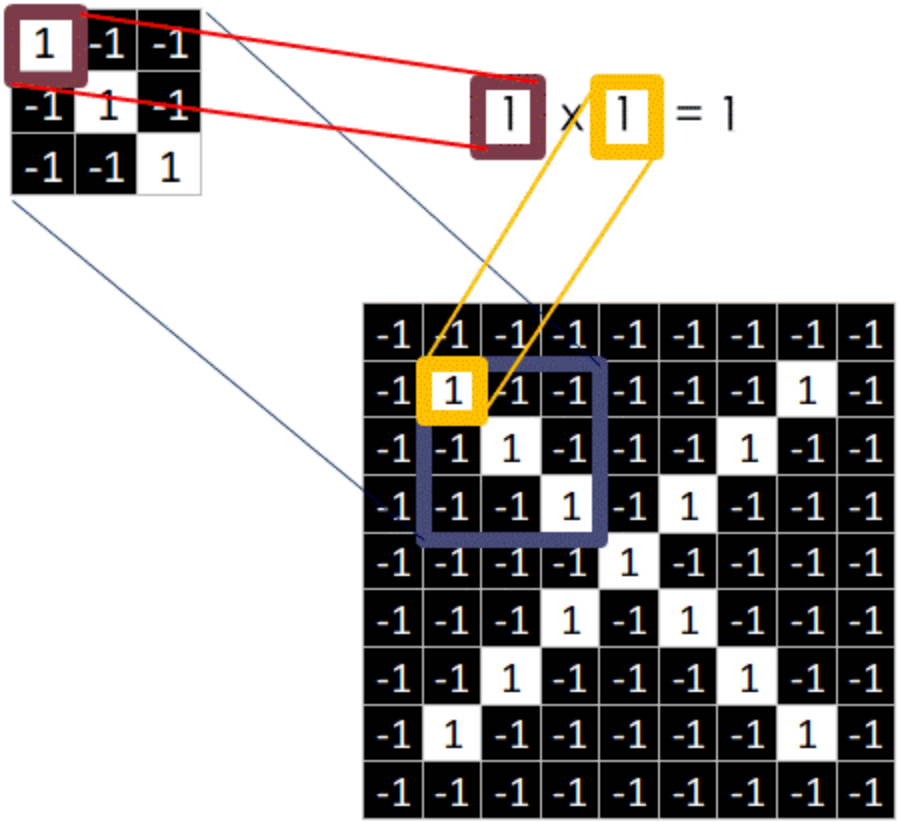
Such filters can be learned via backpropagation during training

The animation is made based on the slides from Brandon Rohrer

Filtering or Convolving: The math behind the match

- 1. Line up the filter and the image patch.
- 2. Multiply each image pixel by the corresponding filter pixel.
- 3. Add them up.
- 4. Divide by the total number of pixels in the feature.
- 5. Add a bias to the result.

Convolving Animation Demo



The animation is made based on the figures from Brandon Rohrer

Convolving Result at a Particular Location:

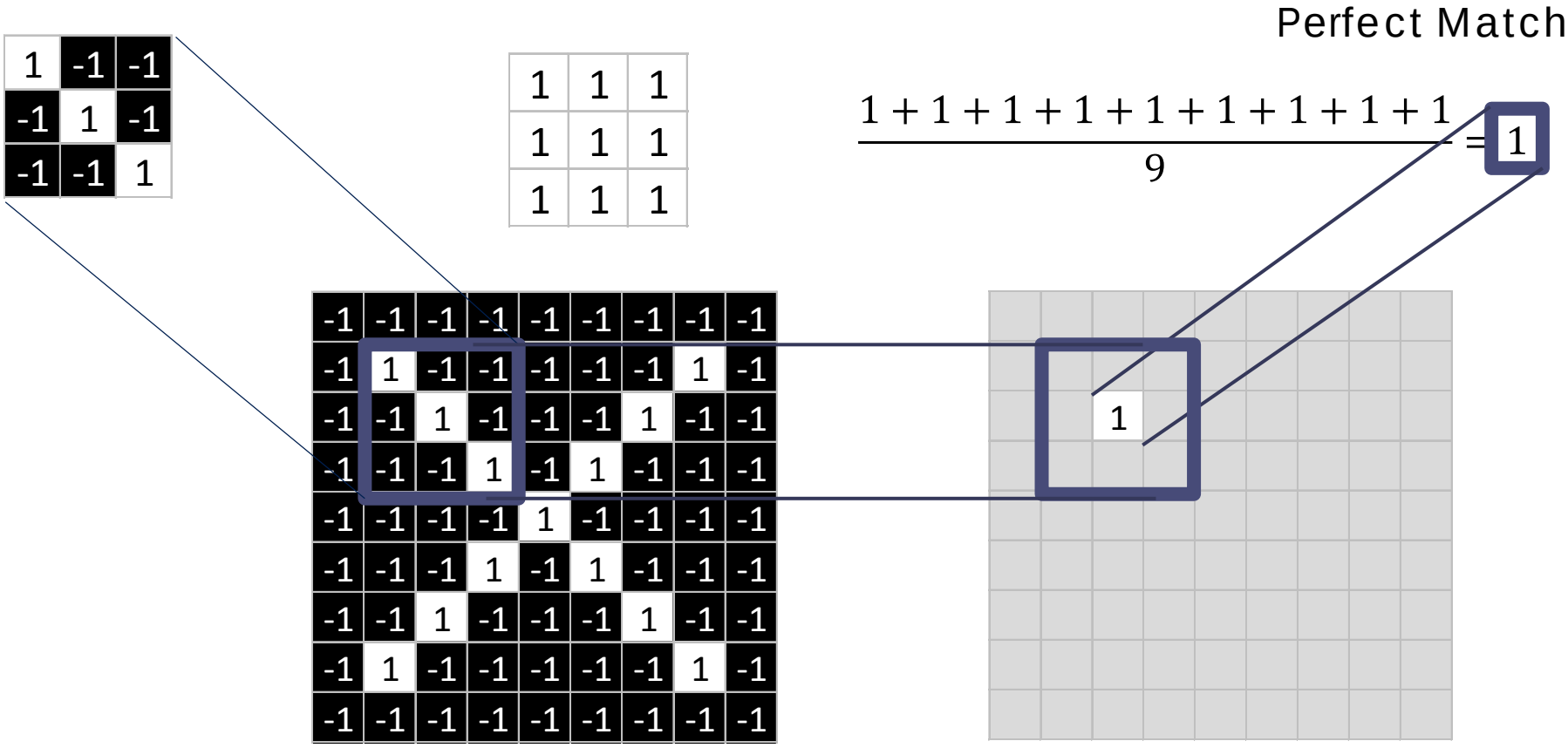


Figure adapted from the slides by Brandon Rohrer

Similarly, for another location:

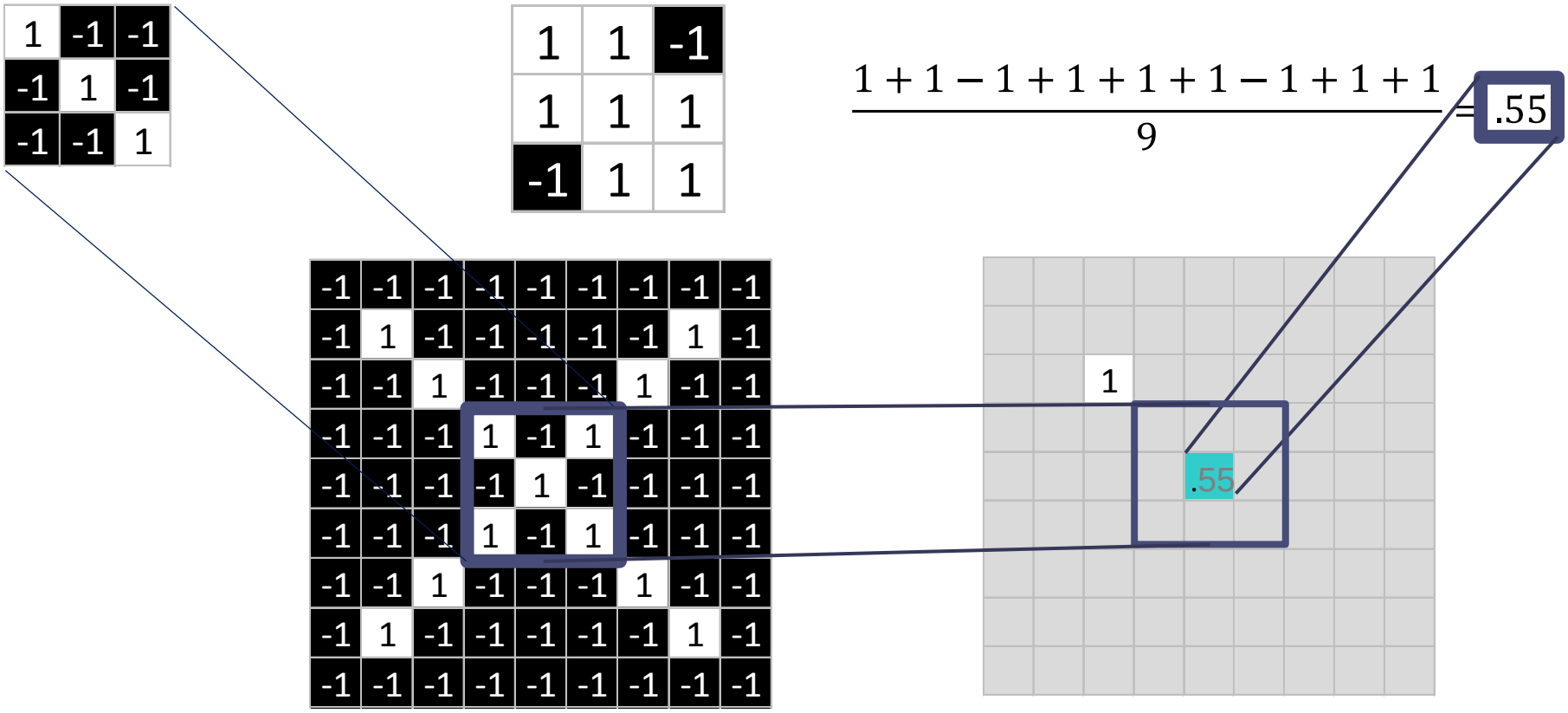


Figure adapted from the slides by Brandon Rohrer

Convolution: Trying every possible match

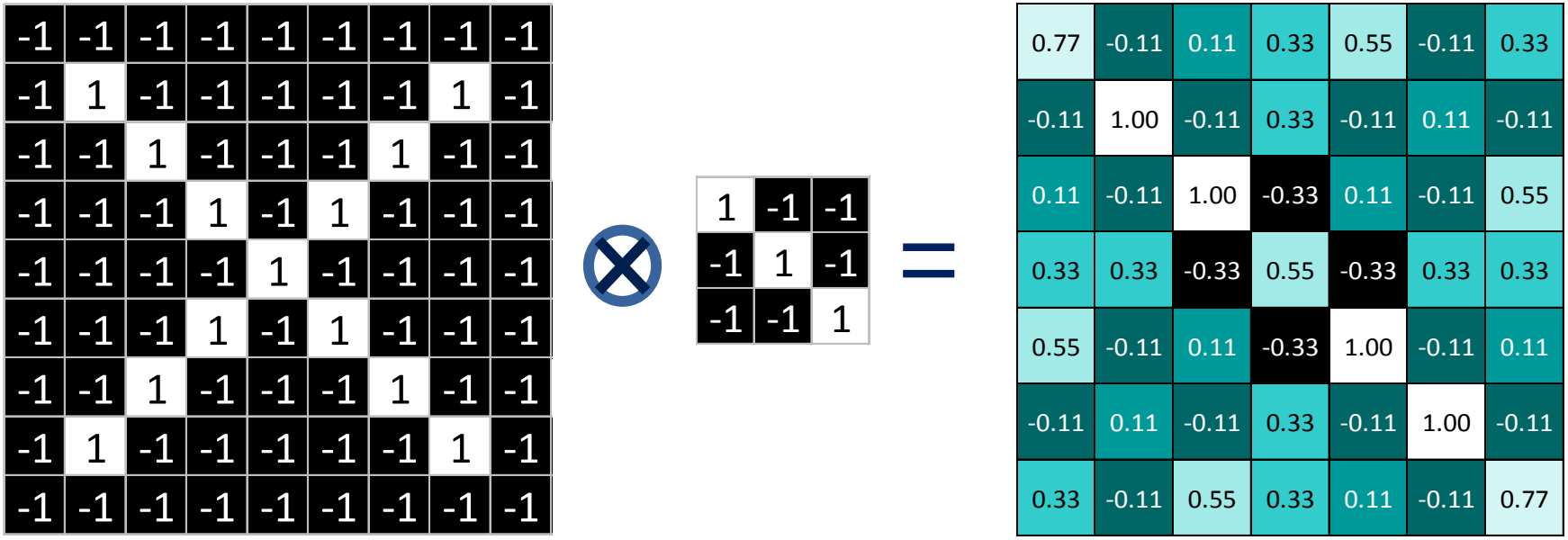


Figure adapted from the slides by Brandon Rohrer

Convolution: Trying every filter

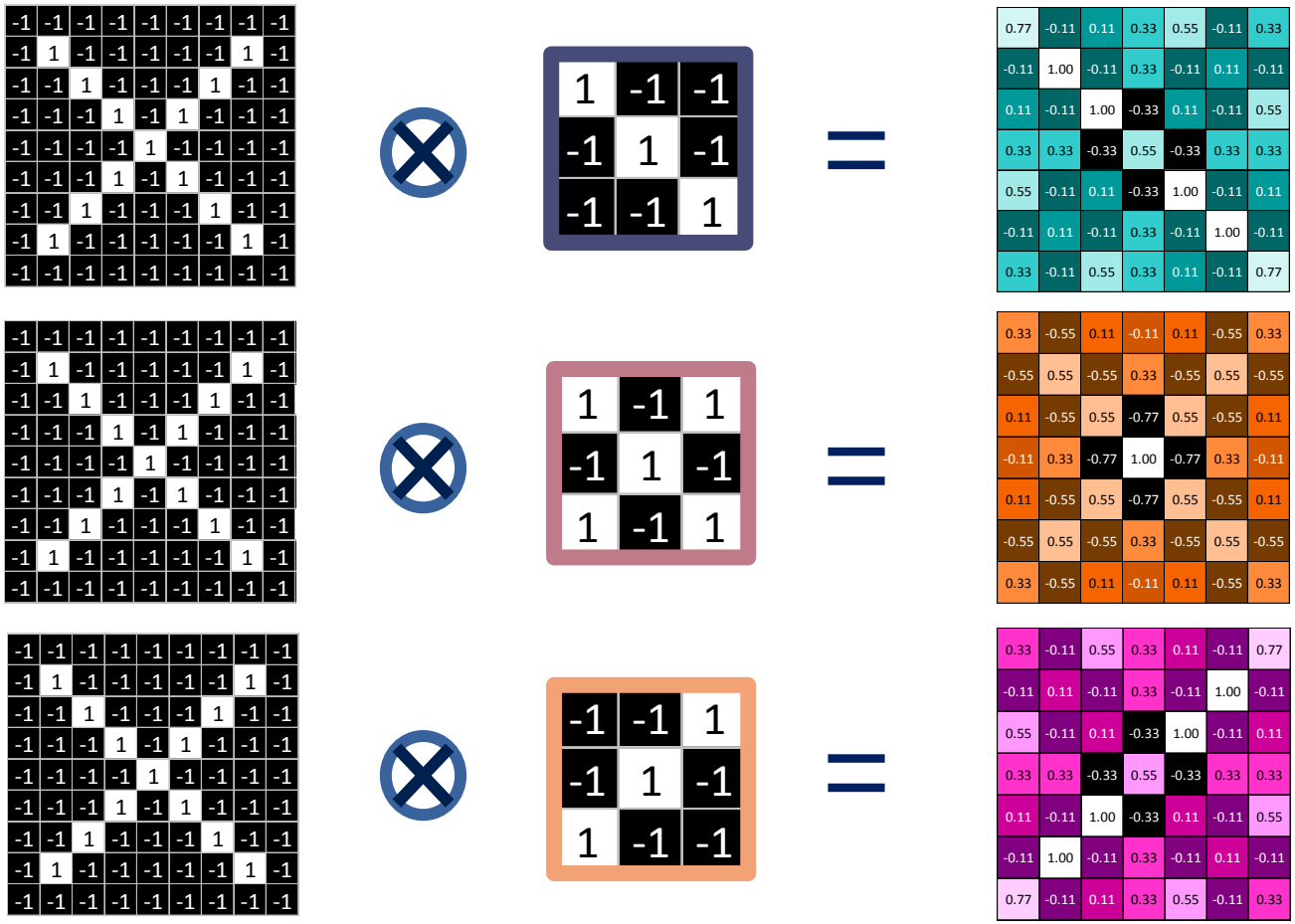
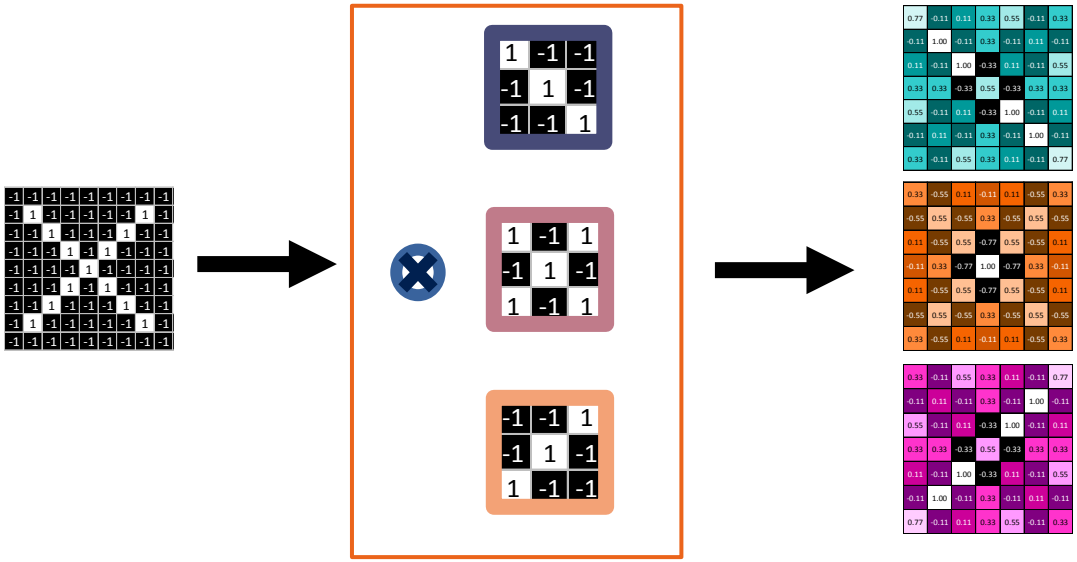


Figure adapted from the slides by Brandon Rohrer

Convolution layer

One image becomes a stack of filtered images



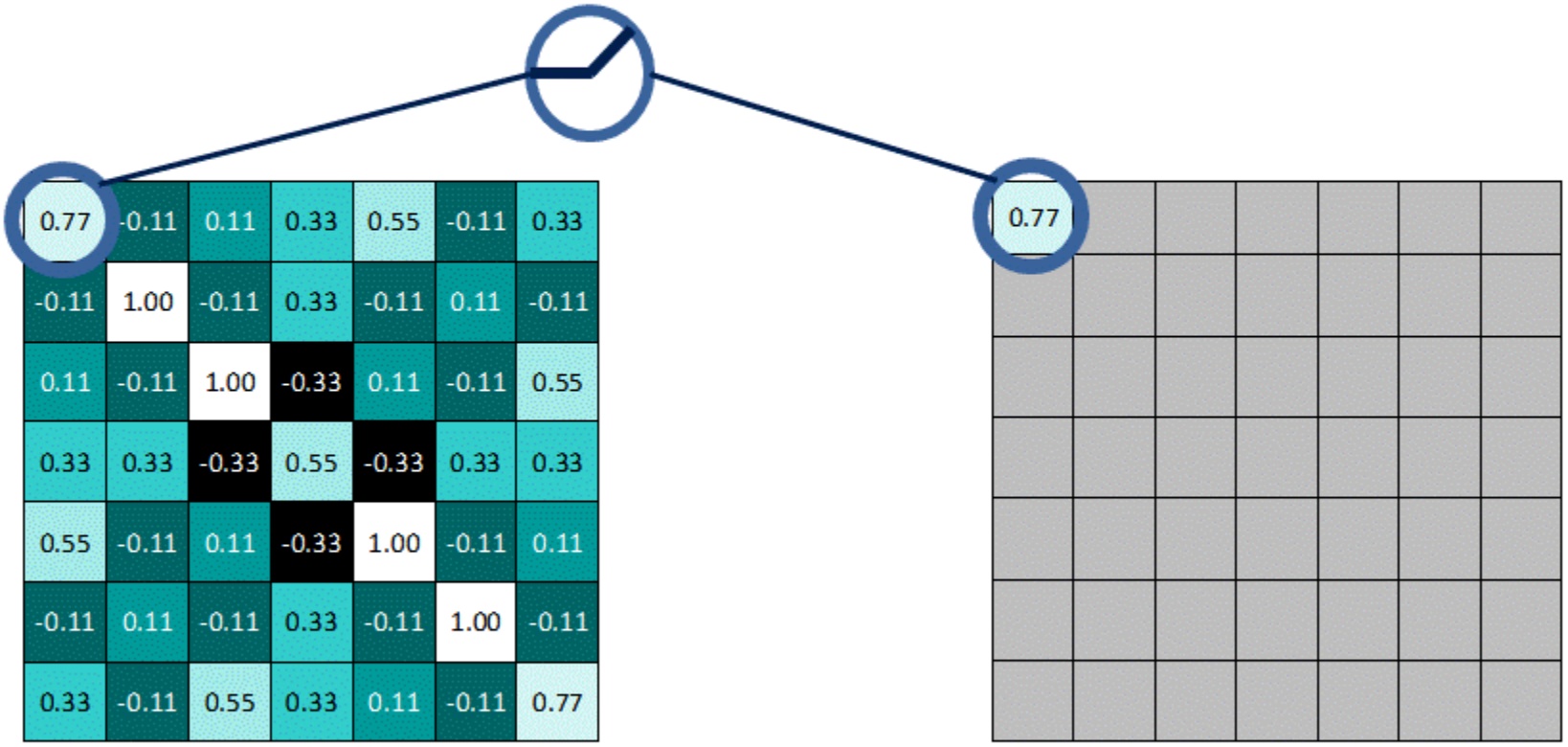
One Conv Layer

In practice, a bias will be added to the result for each filter.

Figure adapted from the slides by Brandon Rohrer

Rectified Linear Units (ReLUs)

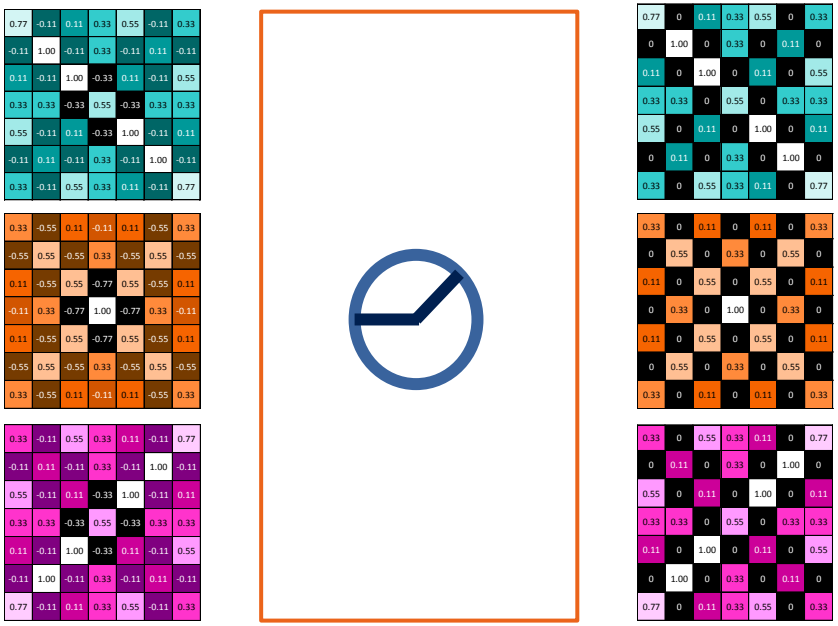
Change everything negative to zero (learned previously).



The animation is made based on the figures from Brandon Rohrer

ReLU layer

A stack of images becomes a stack of images with no negative values.



One ReLU Layer

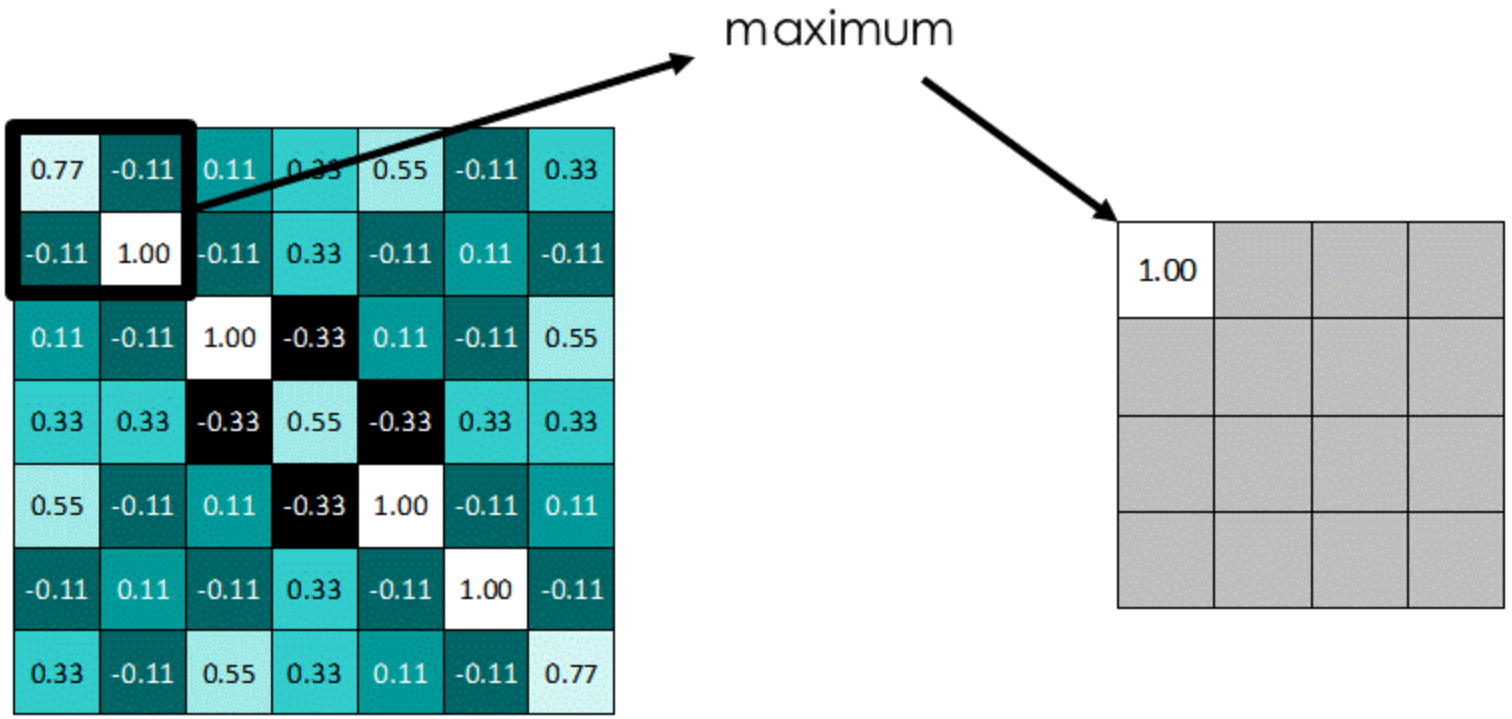
Figure adapted from the slides by Brandon Rohrer

Pooling: Shrinking the image stack

Max Pooling:

- 1. Pick a window size (usually 2 or 3).
- 2. Pick a stride (usually 2).
- 3. Walk your window across your filtered images.
- 4. From each window, take the maximum value.

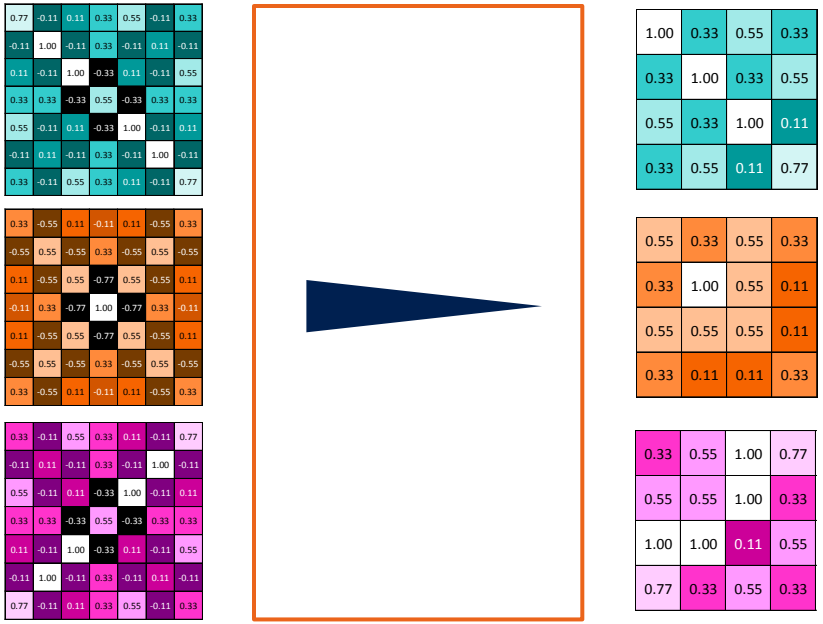
Max Pooling Animation Demo



The animation is made based on the figures from Brandon Rohrer (negative values should be 0 here, but the pooling results do not change).

Pooling layer

A stack of images becomes a stack of smaller images.



One Max Pooling Layer

Figure adapted from the slides by Brandon Rohrer

Layers get stacked

The output of one becomes the input of the next.

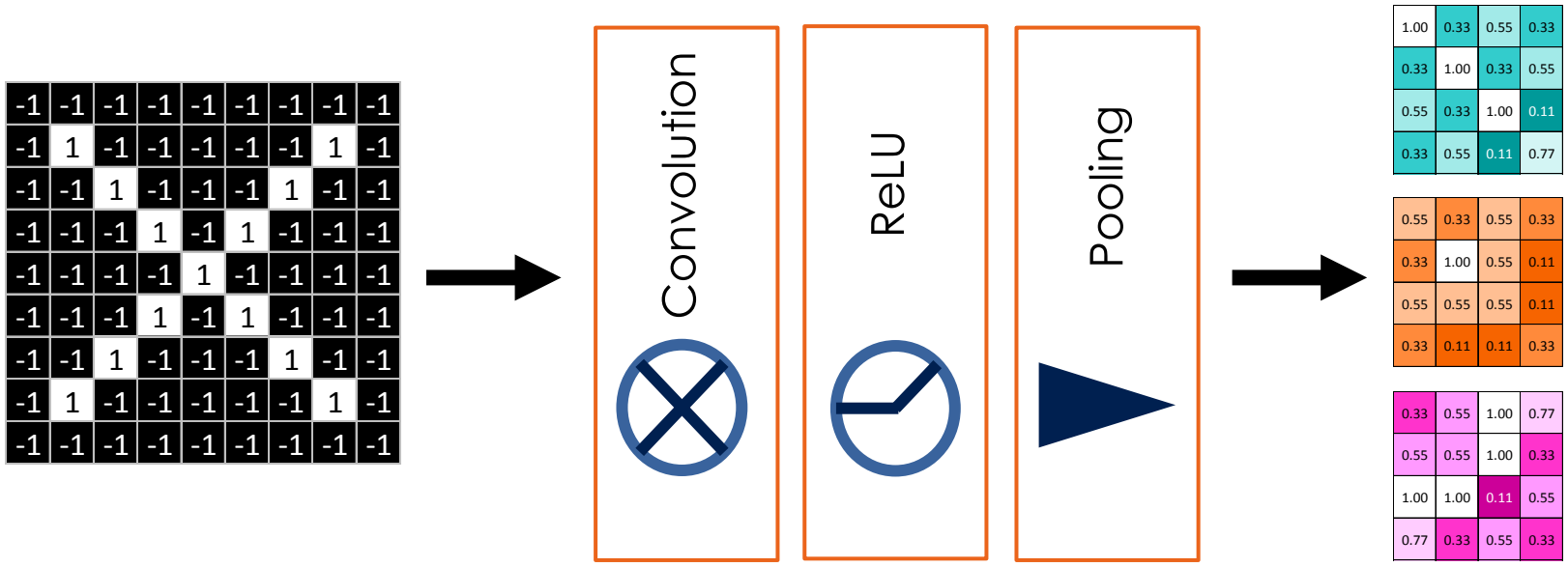


Figure adapted from the slides by Brandon Rohrer

Deep stacking

Layers can be repeated several (or many) times.

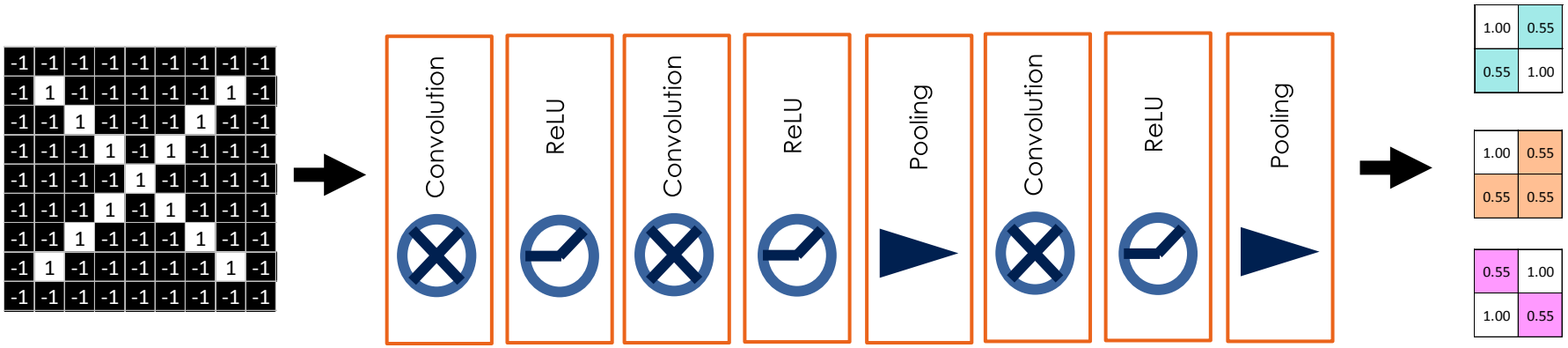


Figure adapted from the slides by Brandon Rohrer

Fully connected layer

Flatten the conv features (location info is lost)

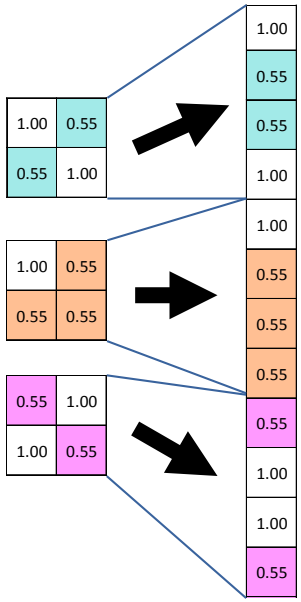


Figure adapted from the slides by Brandon Rohrer

Fully connected layer

Every value gets a vote.

Vote depends on how strongly a value predicts X or O

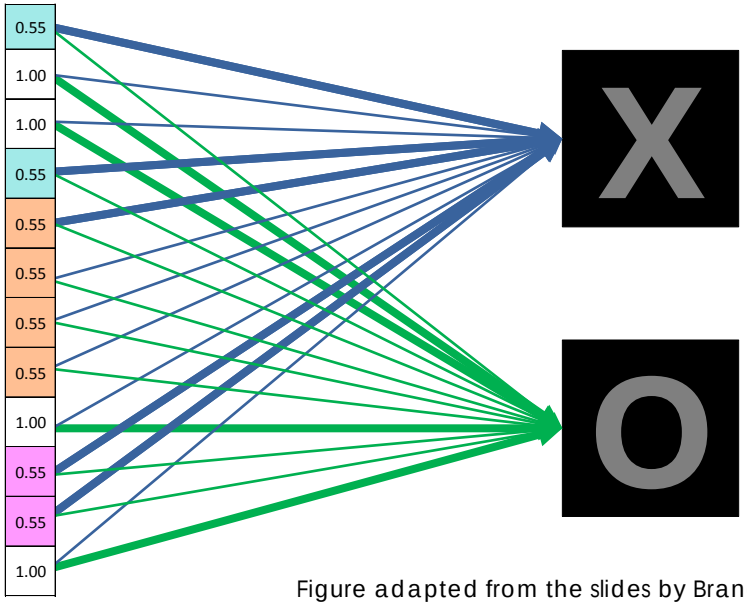
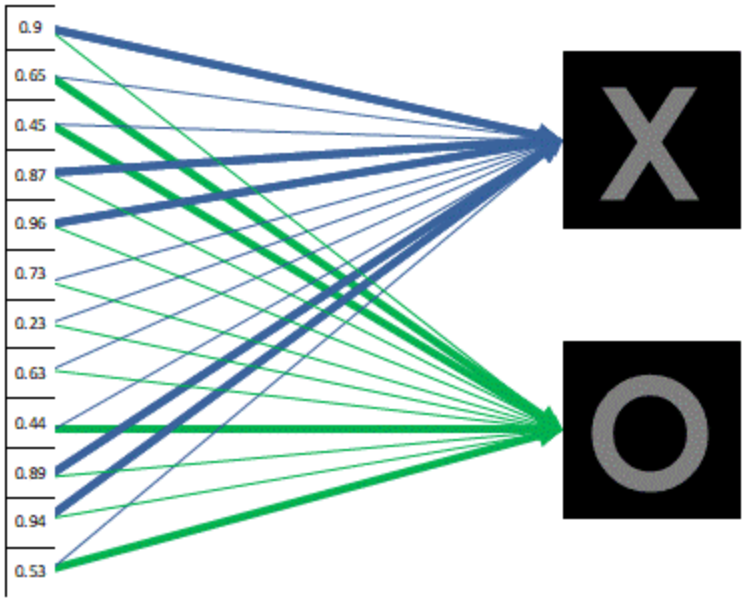


Figure adapted from the slides by Brandon Rohrer

Such voting weights (indicated by connection thickness) are learned from Backpropagation during training.

Weighted Voting in FC Layers

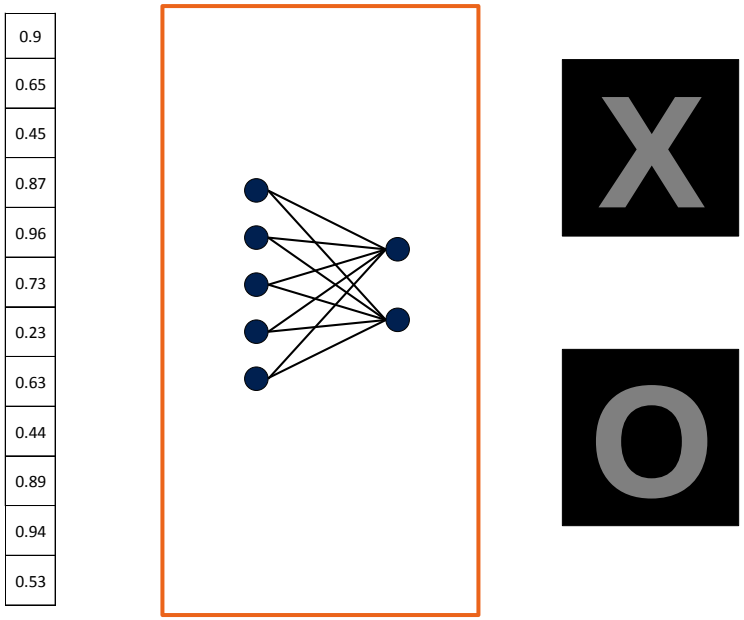
Future values vote on X or O



The animation is made based on the figures from Brandon Rohrer

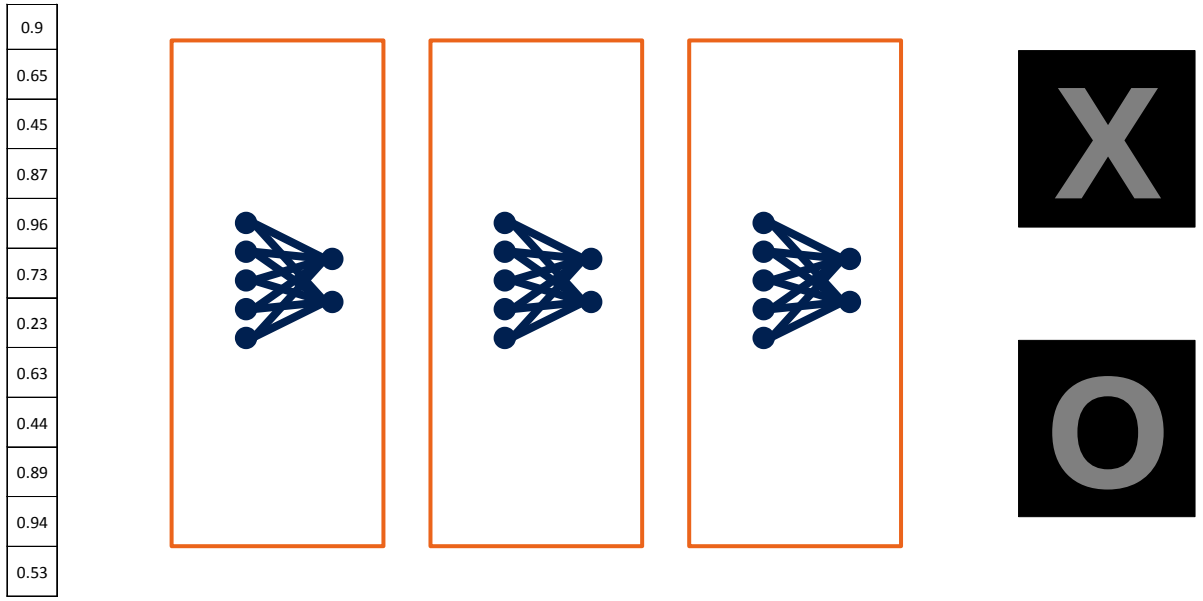
Fully connected layer

A list of feature values becomes a list of votes.



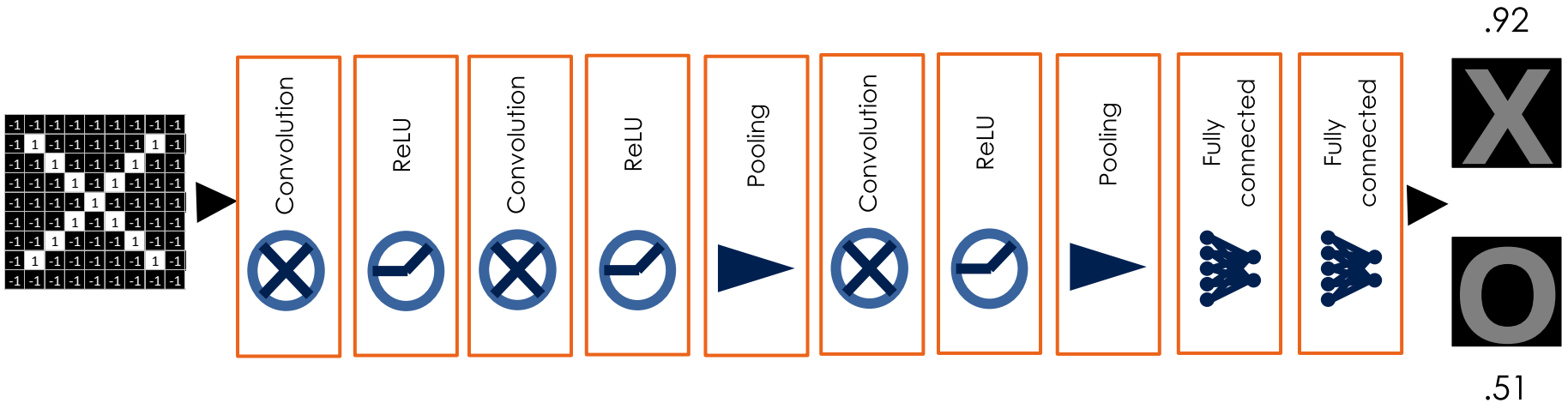
Fully connected layer

These can also be stacked.



Putting it all together

A set of pixels becomes a set of votes.



A real example:

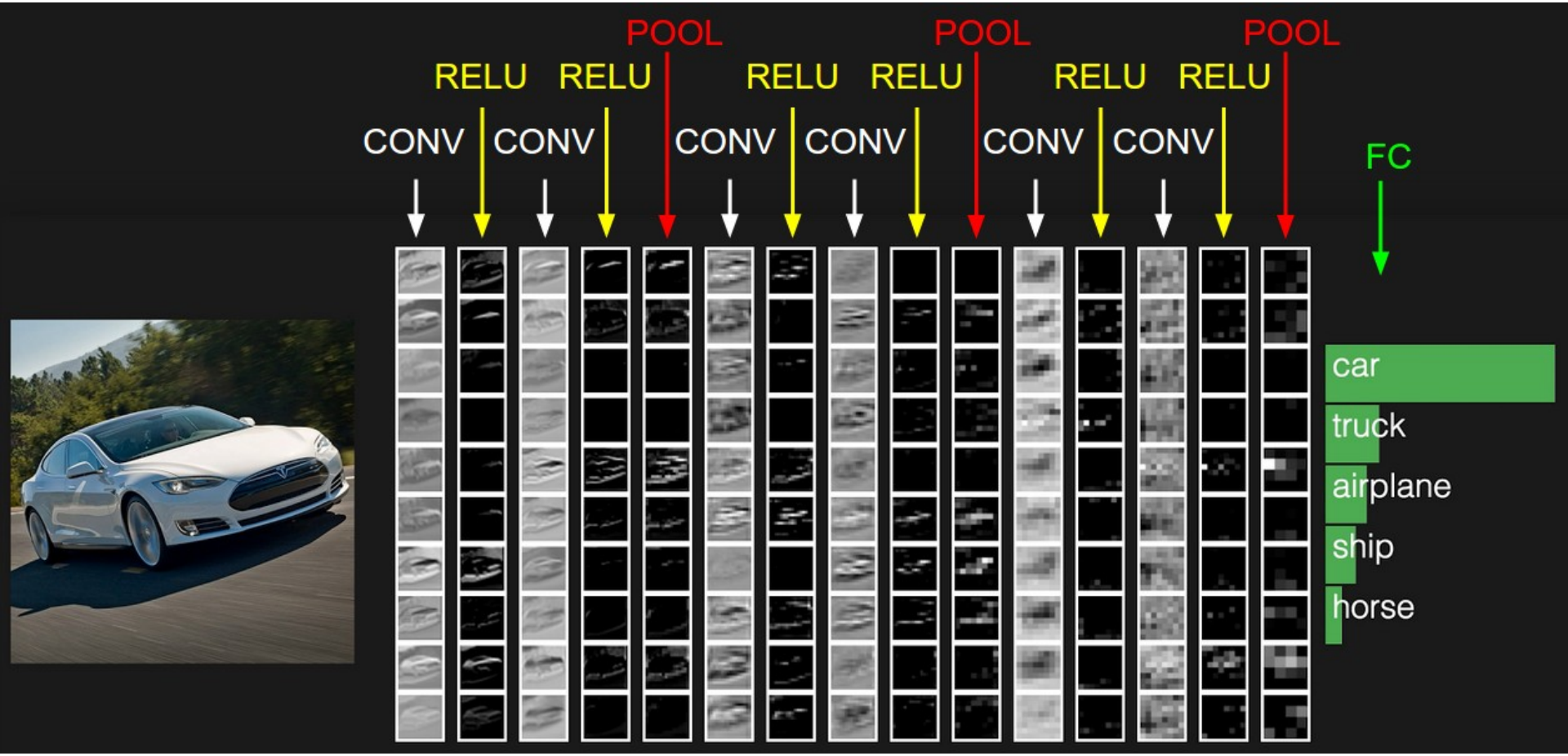
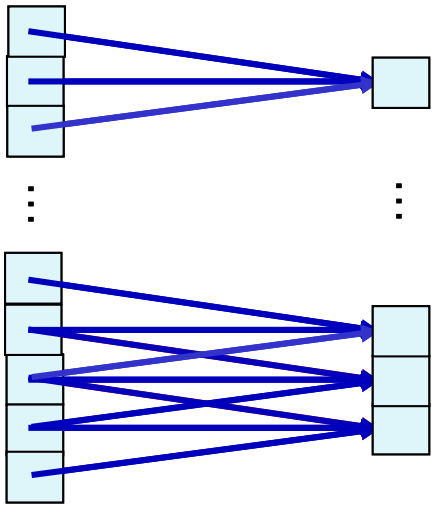


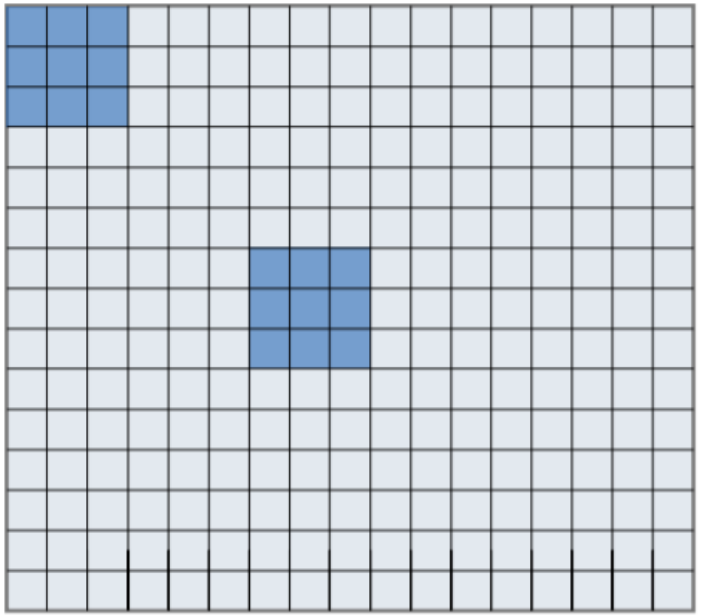
Image source: cs231n.github.io/assets/cnn/convnet.jpeg

Key Idea of CNN – Local Connections

- Instead of every network node being connected to all nodes in the preceding layer (fully connected), connections of limited extent (e.g. 9 pixels in the following figure) account for local structures.

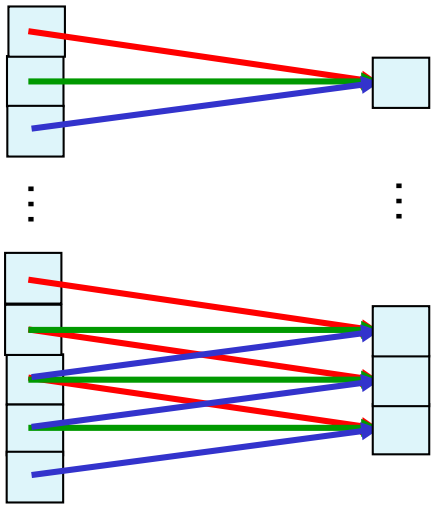


shown in 1-D

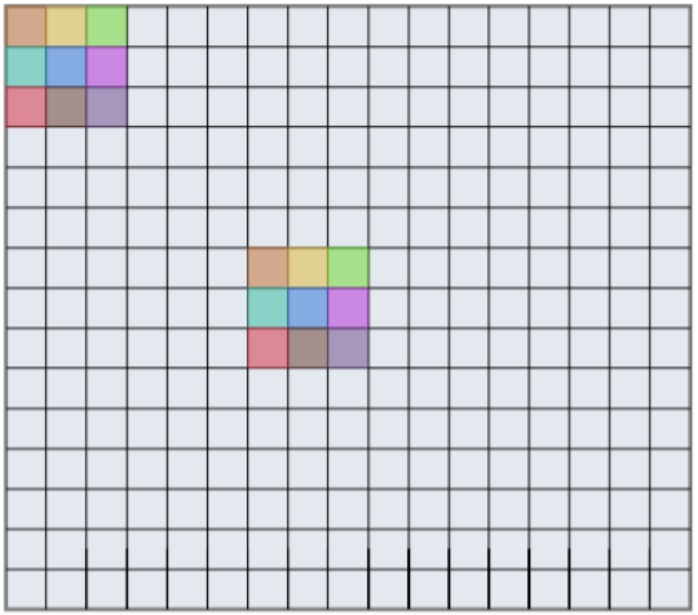


Key Idea of CNN – Weight Sharing

- Filter weights (indicated by different colors) do not change when convolving around the image.
  - Reduces the number of free parameters to learn
  - Activations are shifted with the input (values do not change)

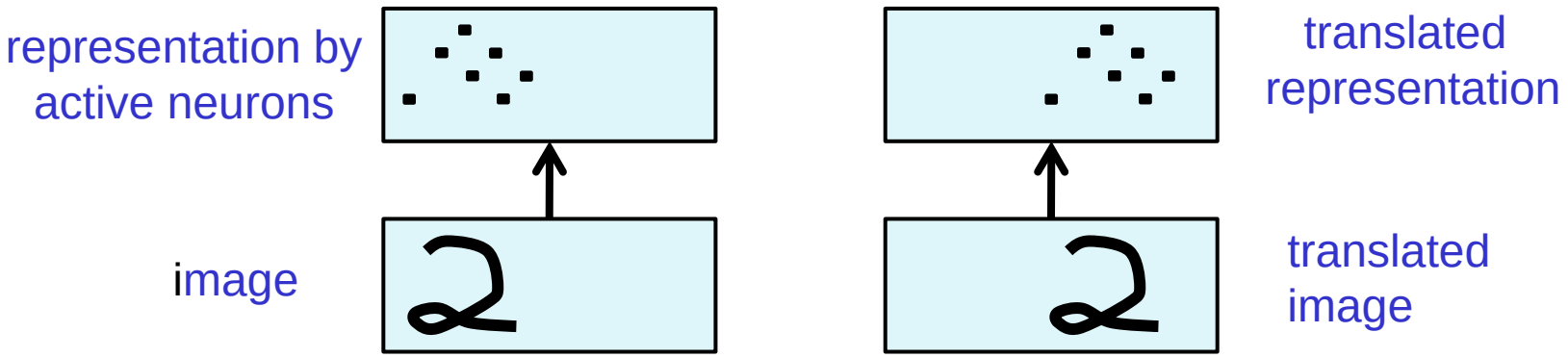


shown in 1-D



Key Idea of CNN – Weight Sharing

- **Equivariant features**: Replicated features do not make the neural activities invariant to translation. The activities are equivariant



- **Invariant knowledge**: If a feature is useful in some locations during training, detectors for that feature will be available in all locations during testing.

Key Idea of CNN – Compositionality

- Higher-level features are built from low-level features
  - First layer detects primitive features (edges/blobs), later kernels are more abstract and are of larger scales
  - Subsequent layers extract features in the dense feature maps of the preceding layer.

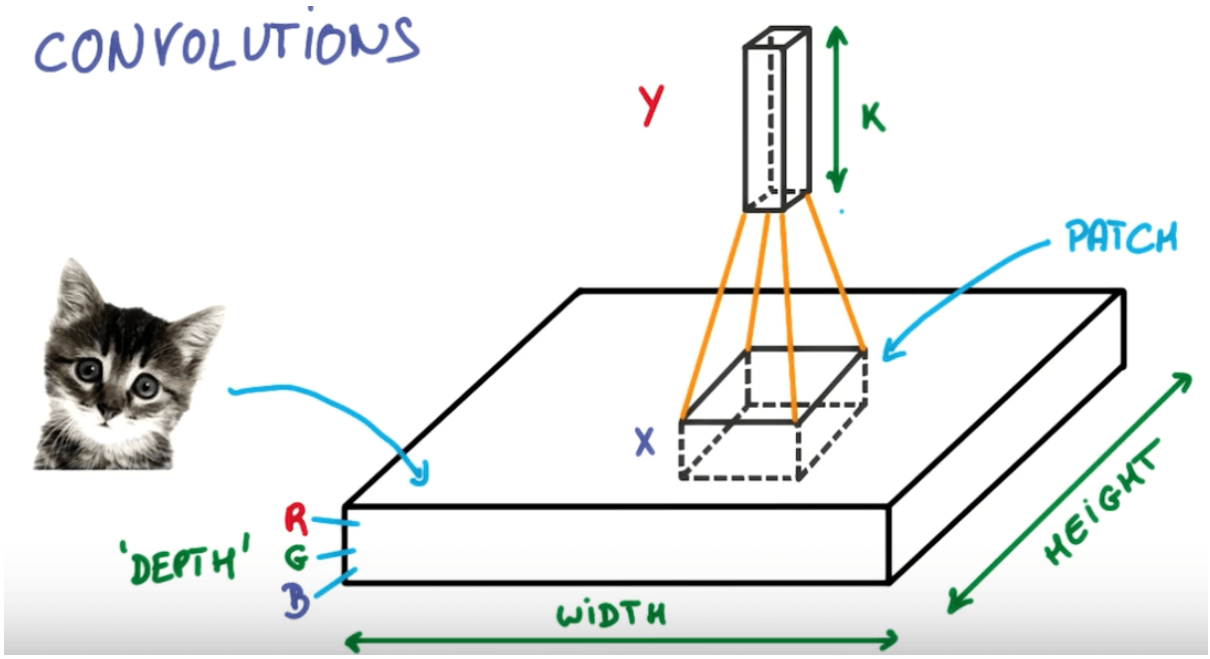


Image source: http://www.ritchieng.com/machine-learning/deep-learning/convs/

Manifold Hypothesis

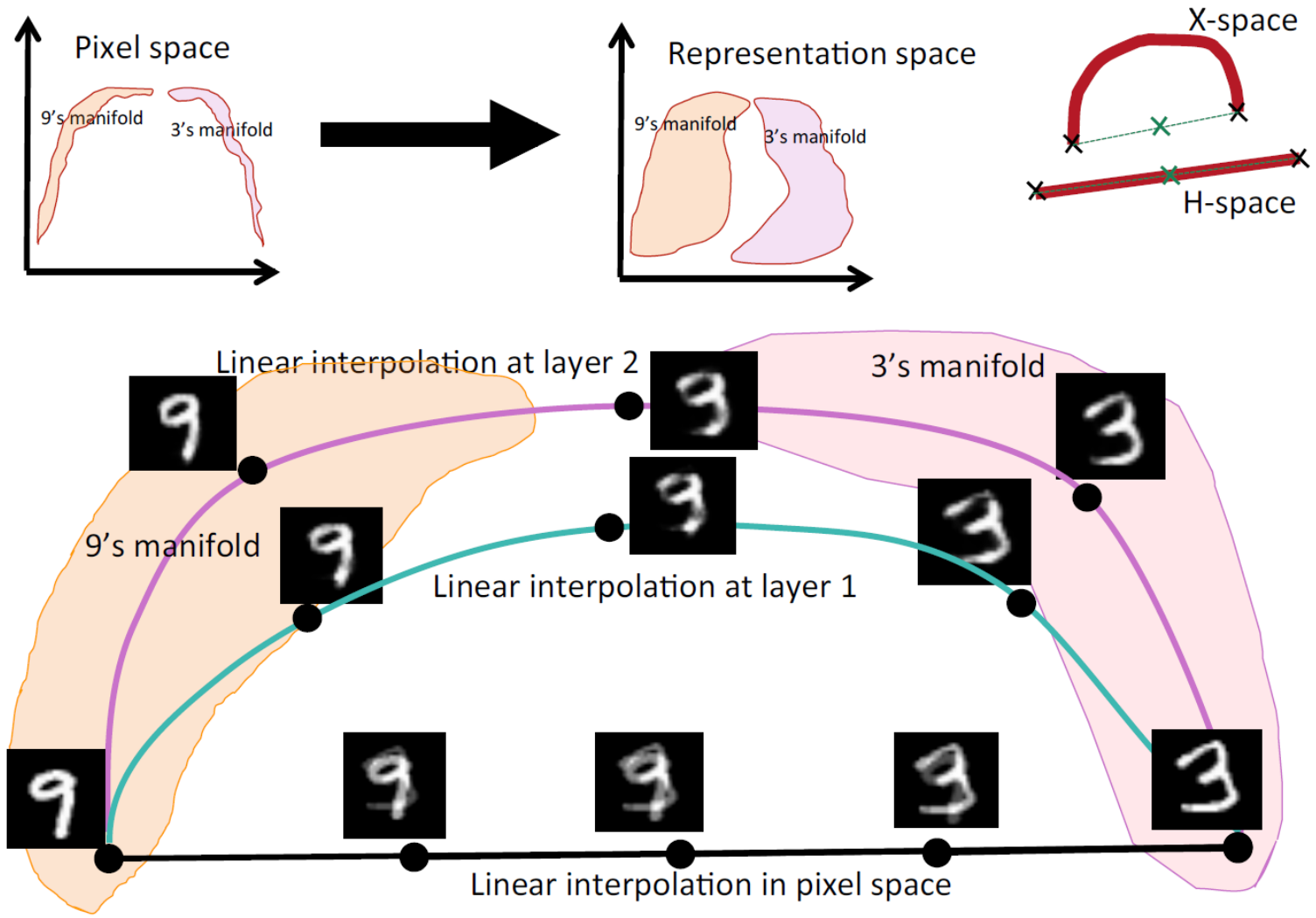


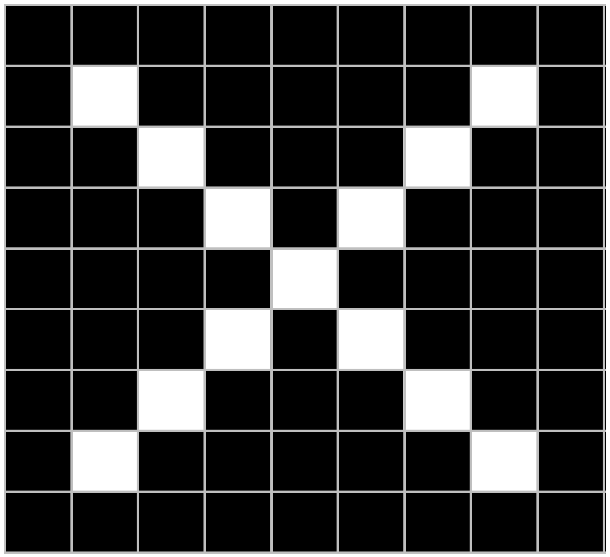
Image Credit: Bengio ICML 2013

CNN Applications

Images

Columns of pixels

Rows of pixels

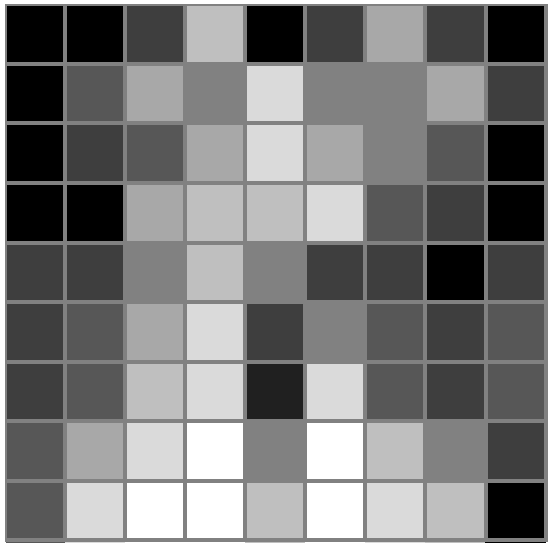


What's in common of these application areas?

Sound

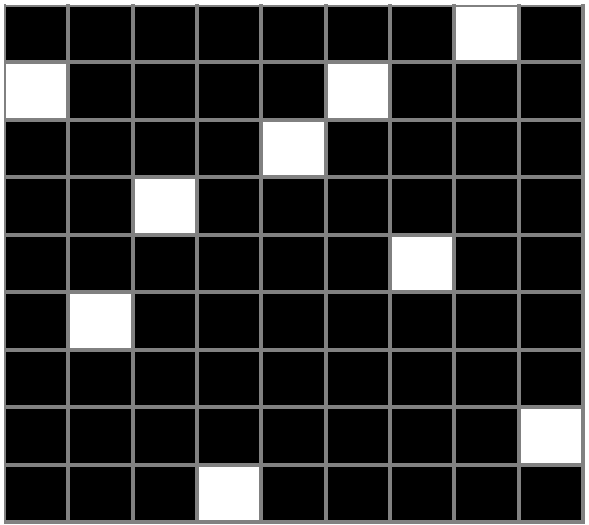
Time steps

Intensity in each frequency band



Position in sentence

Words in dictionary



Text

Figure adapted from the slides by Brandon Rohrer

Applications in Computer Vision

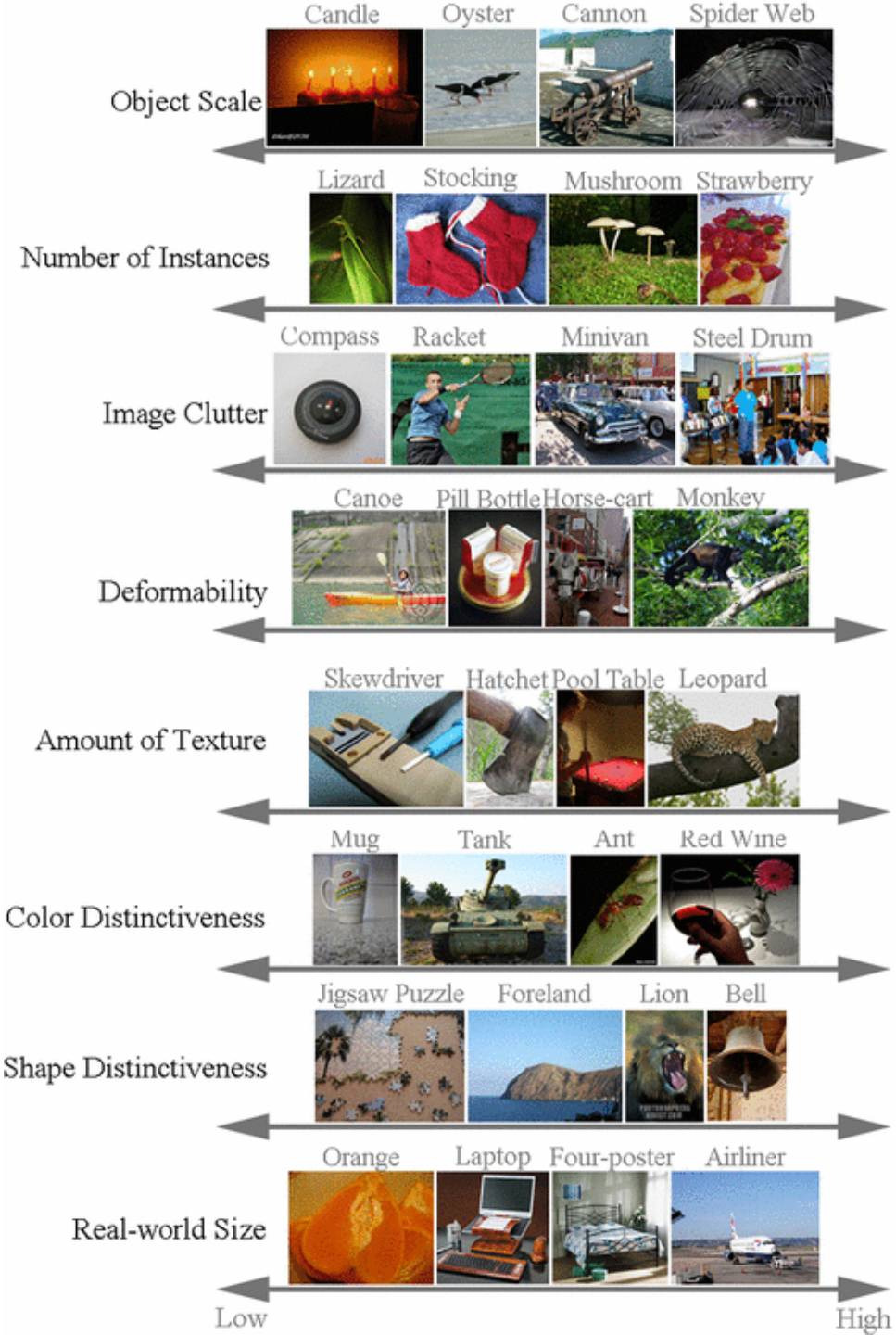
- Visual recognition and classic CNN models
- Detection
- Segmentation
- Image Generation
- Much More ...

Visual Recognition

Image source: blog.acolyer.org

- ImageNet Large Scale Visual Recognition Challenge (ILSVRC)
  - 14,197,122 images, 21,841 object categories (1000 categories are used in challenges)
  - Started in 2010

Russakovsky, Olga, et al. "Imagenet large scale visual recognition challenge". International Journal of Computer Vision (2015).



The ImageNet dataset is very diverse, with variation in 8 properties

The ImageNet dataset categories were much finer-grained than previous large image databases, such as Pascal.

Table 2

Scale of ILSVRC image classification task (minimum per class - maximum per class)

Year	Train images (per class)	Val images (per class)	Test images (per class)
Image classification annotations (1000 object classes)			
ILSVRC2010	1,261,406 (668–3047)	50,000 (50)	150,000 (150)
ILSVRC2011	1,229,413 (384–1300)	50,000 (50)	100,000 (100)
ILSVRC2012-14	1,281,167 (732–1300)	50,000 (50)	100,000 (100)

The numbers in parentheses correspond to (minimum per class–maximum per class). The 1000 classes change from year to year but are consistent between image classification and single-object localization tasks in the same year. All images from the image classification task may be used for single-object localization

Deep Breakthrough in 2012

In 2012 the ILSVRC (ImageNet) challenge was dominated by a new Deep Learning based approach – AlexNet – which blew away the competition

ILSVRC 2012 results (image classification)

#	Team name	Method	Top-5 error, %
1	SuperVision	AlexNet + extra data	0.15315
2	SuperVision	AlexNet	0.16422
3	ISI	SIFT+FV, LBP+FV, GIST+FV	0.26172
5	ISI	Naive sum of scores from classifiers using each FV	0.26646
7	OXFORD_VGG	Mixed selection from High-Level SVM scores and Baseline Scores	0.26979

ImageNet Classification with Deep Convolutional Neural Networks

Alex Krizhevsky
University of Toronto
kriz@cs.utoronto.ca

Ilya Sutskever
University of Toronto
ilya@cs.utoronto.ca

Geoffrey E. Hinton
University of Toronto
hinton@cs.utoronto.ca

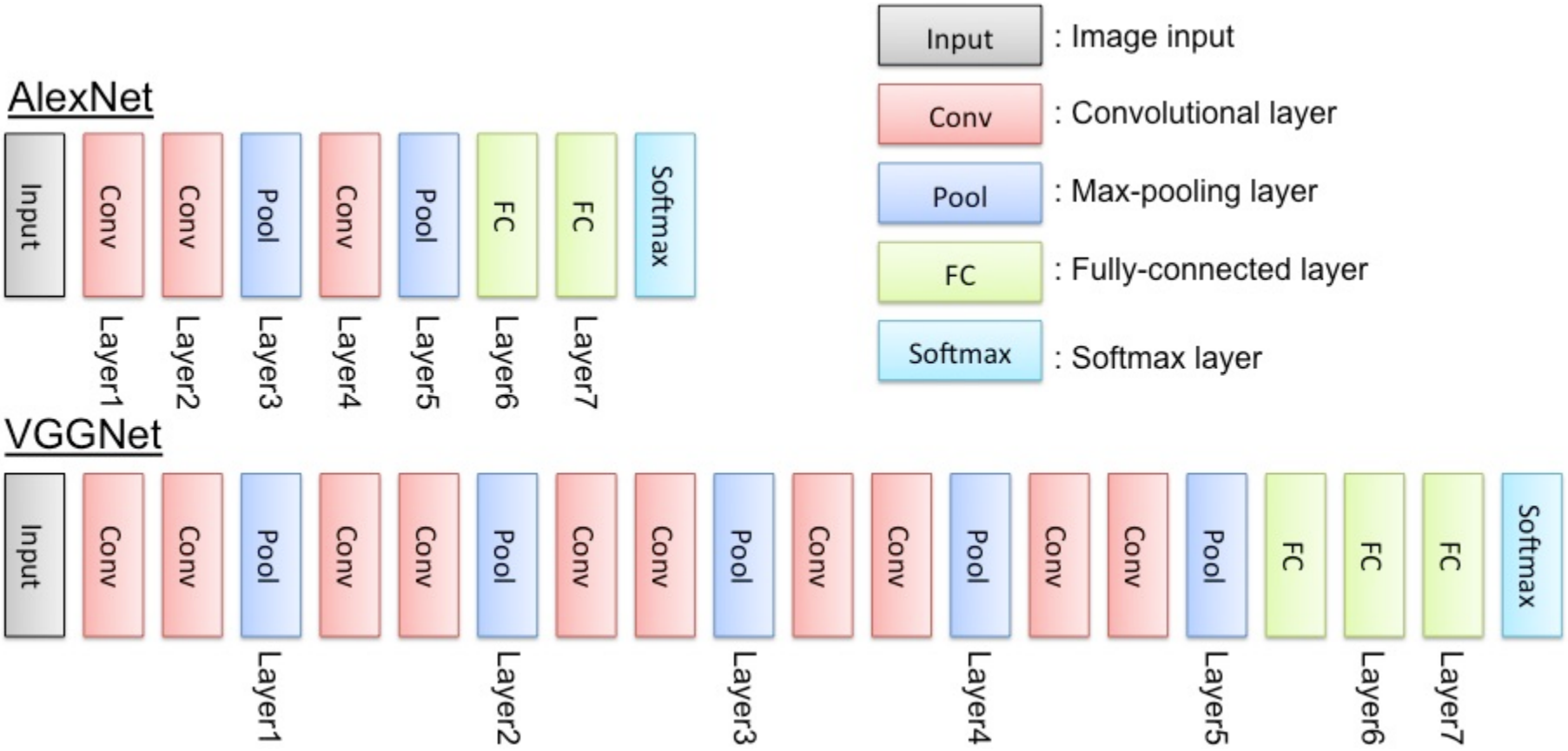
Abstract

We trained a large, deep convolutional neural network to classify the 1.2 million high-resolution images in the ImageNet LSVRC-2010 contest into the 1000 different classes. On the test data, we achieved top-1 and top-5 error rates of 37.5% and 17.0% which is considerably better than the previous state-of-the-art. The neural network, which has 60 million parameters and 650,000 neurons, consists of five convolutional layers, some of which are followed by max-pooling layers, and three fully-connected layers with a final 1000-way softmax. To make training faster, we used non-saturating neurons and a very efficient GPU implementation of the convolution operation. To reduce overfitting in the fully-connected layers we employed a recently-developed regularization method called "dropout" that proved to be very effective. We also entered a variant of this model in the ILSVRC-2012 competition and achieved a winning top-5 test error rate of 15.3%, compared to 26.2% achieved by the second-best entry.

Imagenet classification with deep convolutional neural networks. A Krizhevsky, I Sutskever, GE Hinton
Advances in Neural Information Processing Systems, pp1097-1105, 2012

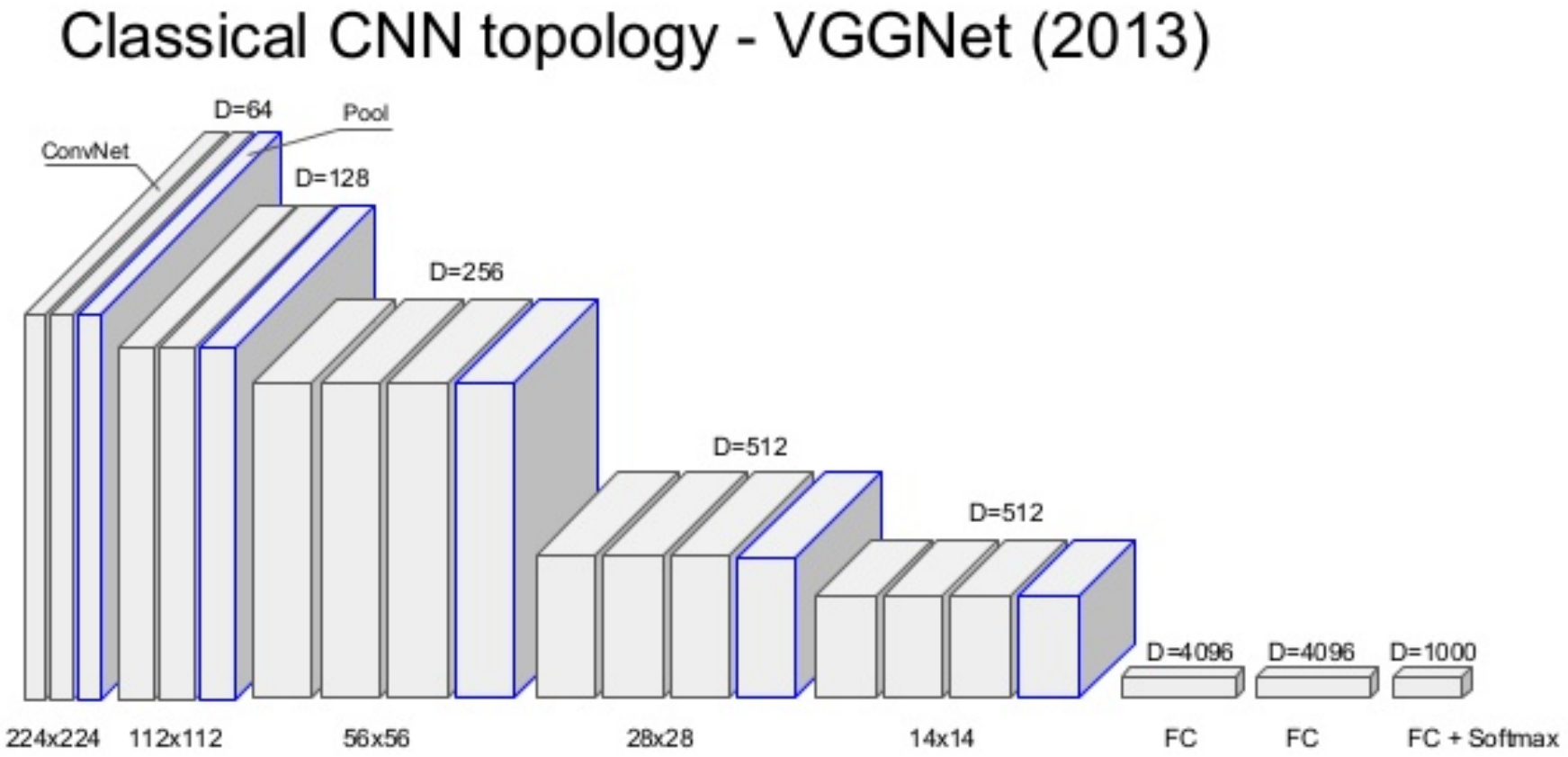
AlexNet and VGGNet structure

VGGNet is a simple, but effective, deep net developed by the Visual Graphics Group (VGG) at Oxford. It uses 3x3 convolutions only. It is the depth of the network that makes the small filters capable of complex recognition tasks.



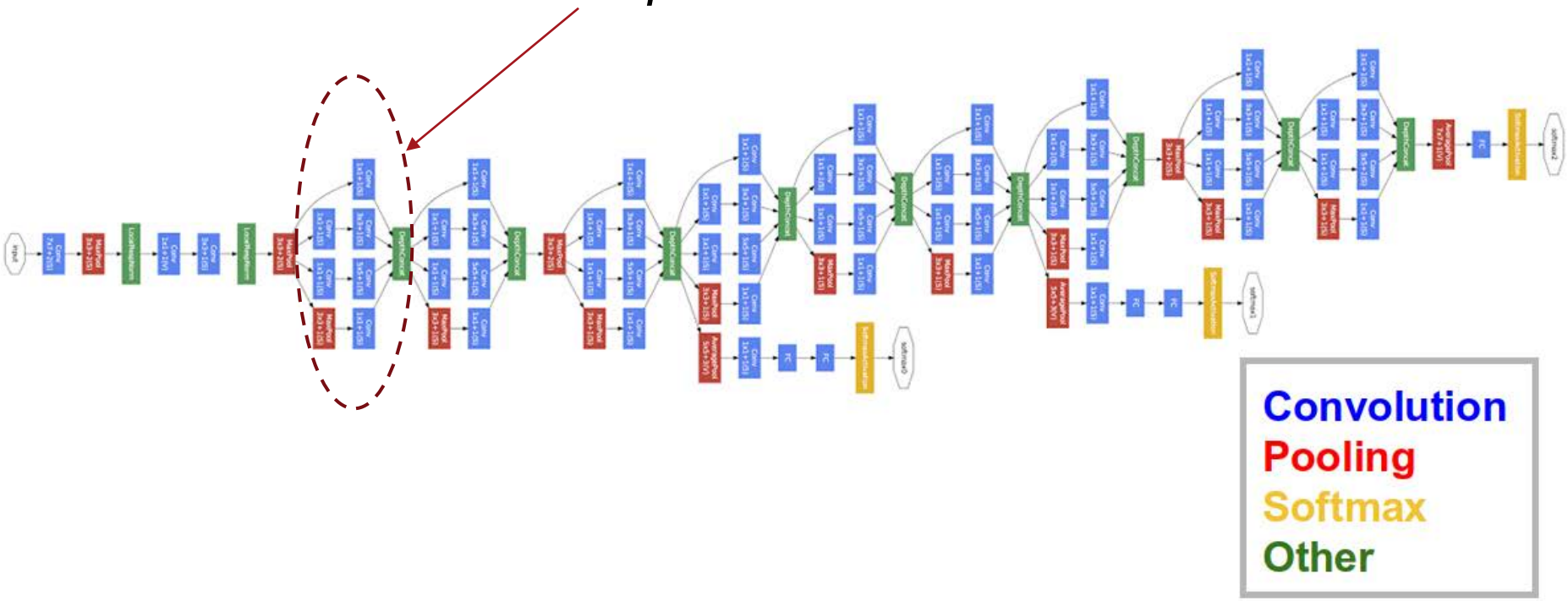
Another view of the VGGNet

D is the number of different convolutional filters in a given layer.

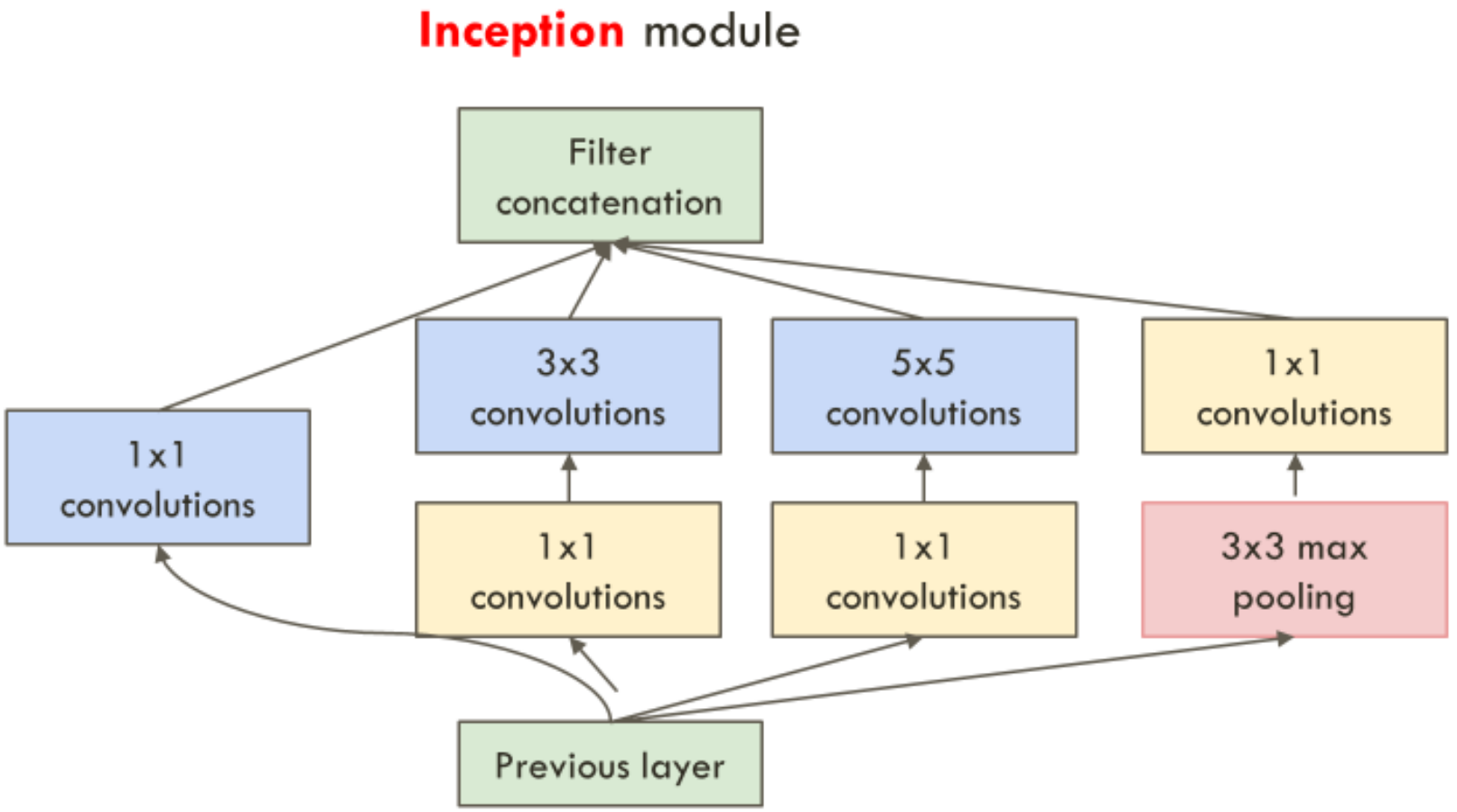


GoogLeNet

- 22 layers deep net
- Much more complex structure than VGGNet
- Uses so-called *inception* modules



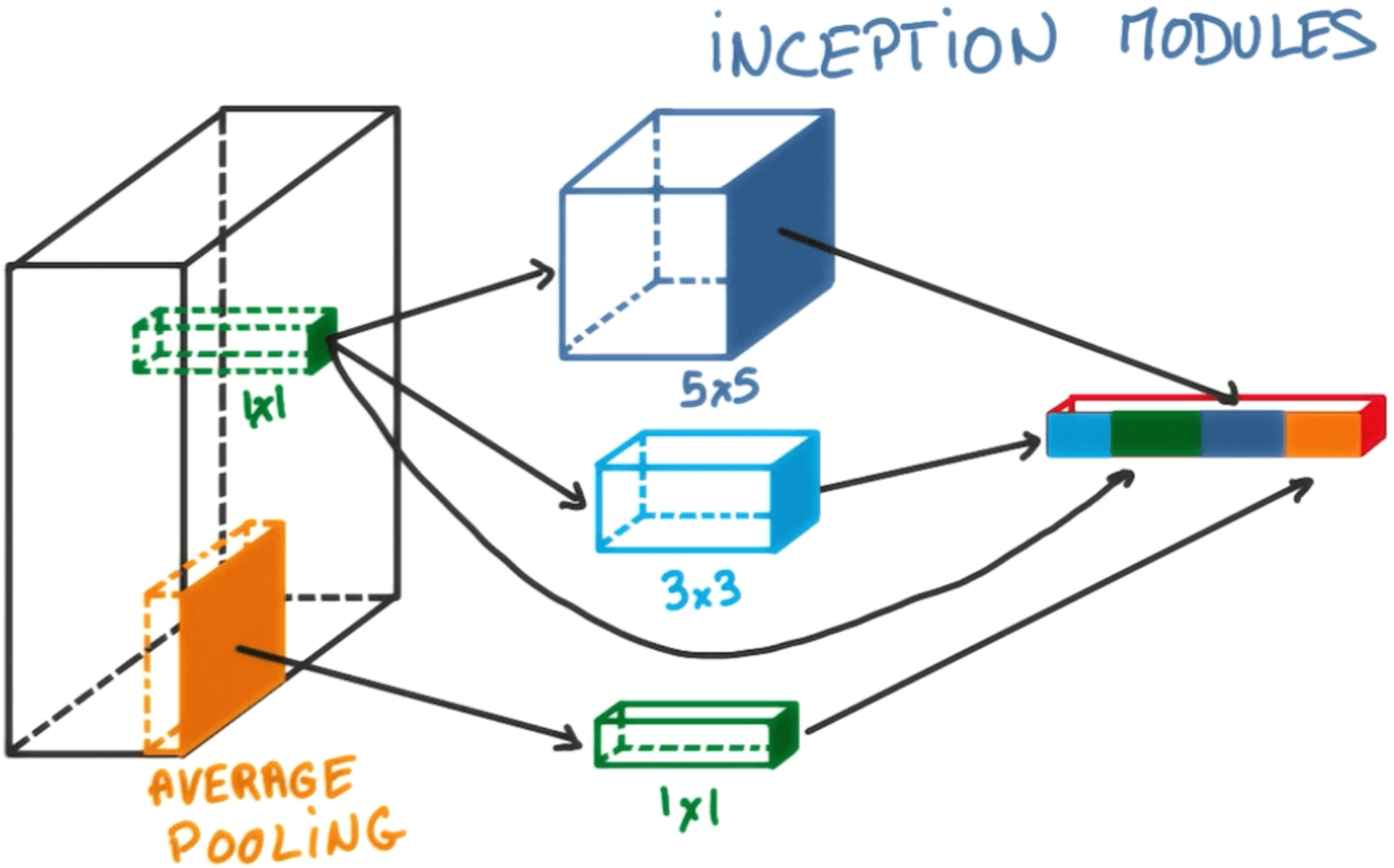
Inception Layers



SLIDE CREDIT:GOOGLE INC

Inception's architects stacked 1x1 convolutions in front of the expensive 3x3 and 5x5 convolutions to reduce the dimensionality before each convolution

Another View of Inception Module



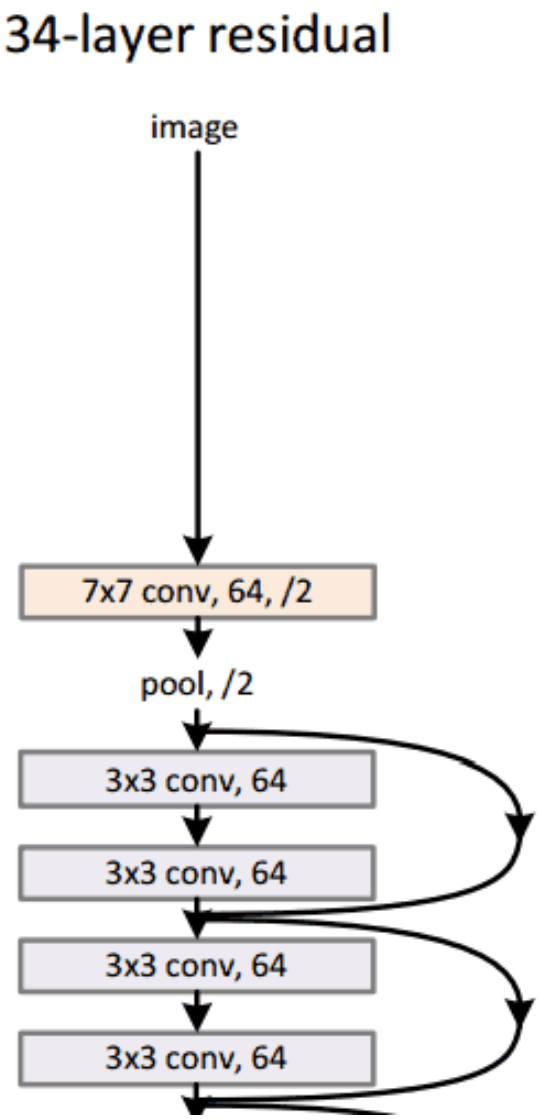
http://www.ritchieng.com/machine-learning/deep-learning/convs/

1. max pooling is used in the original paper, 2. the blocks are for demo purpose only (their depths are not accurate).

Well said Leo, well said

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

ResNet – a very deep network!

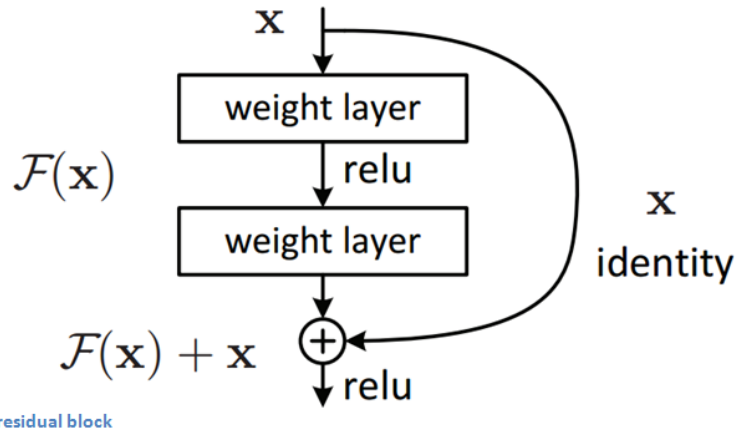


Developed by Microsoft Asia

Won the 2015 ILSVRC (Imagenet) challenge.

Has 152 layers!

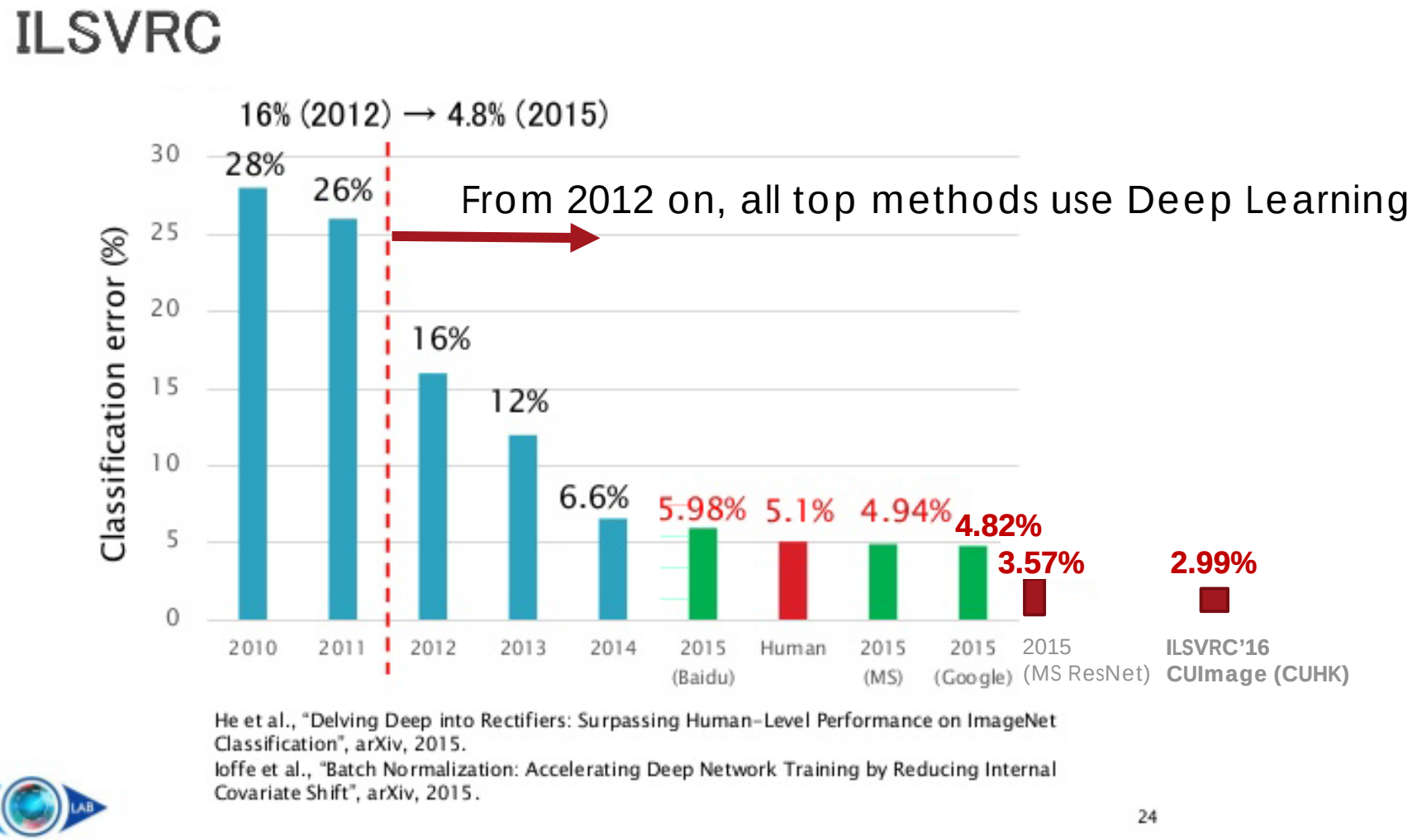
Uses a so-called *residual block* to simplify the learning.



A residual block

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

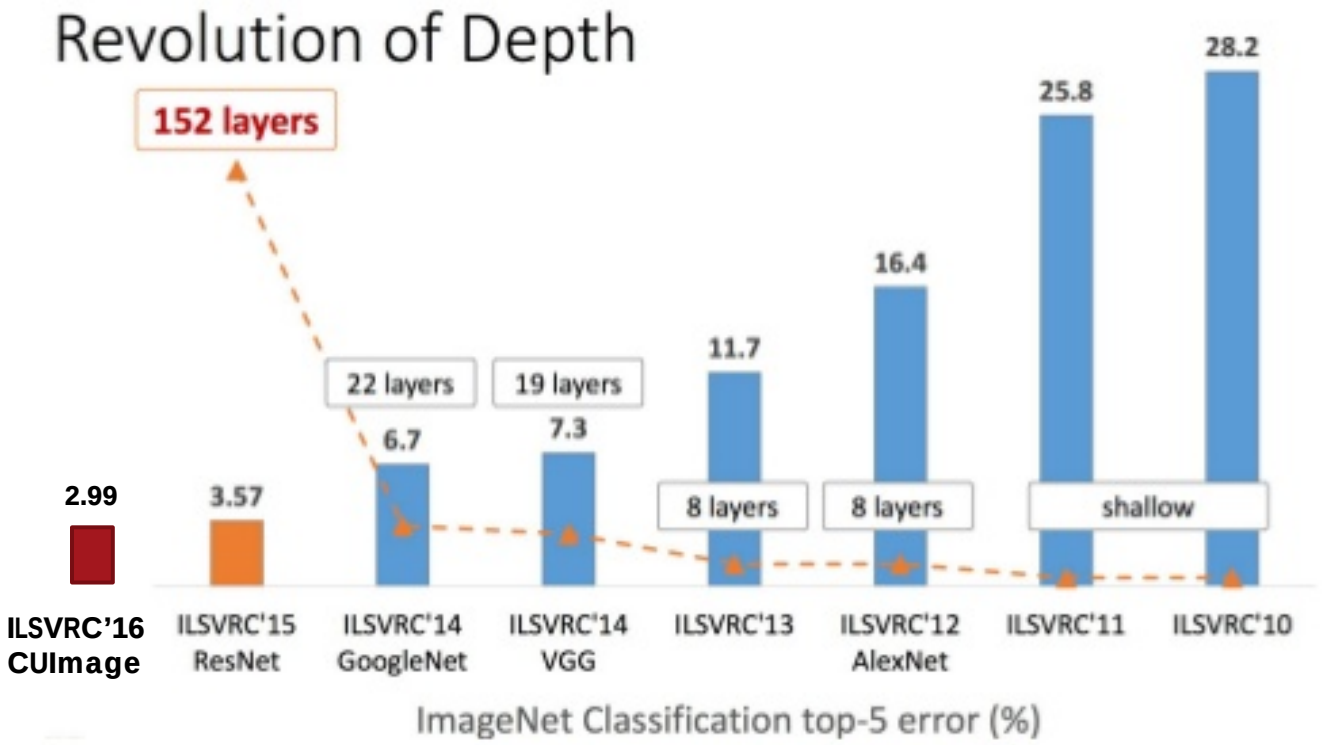
Deep Nets are now better than Humans at recognizing objects!



He et al., "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification", arXiv, 2015.
Ioffe et al., "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift", arXiv, 2015.

Deeper is Usually Better!

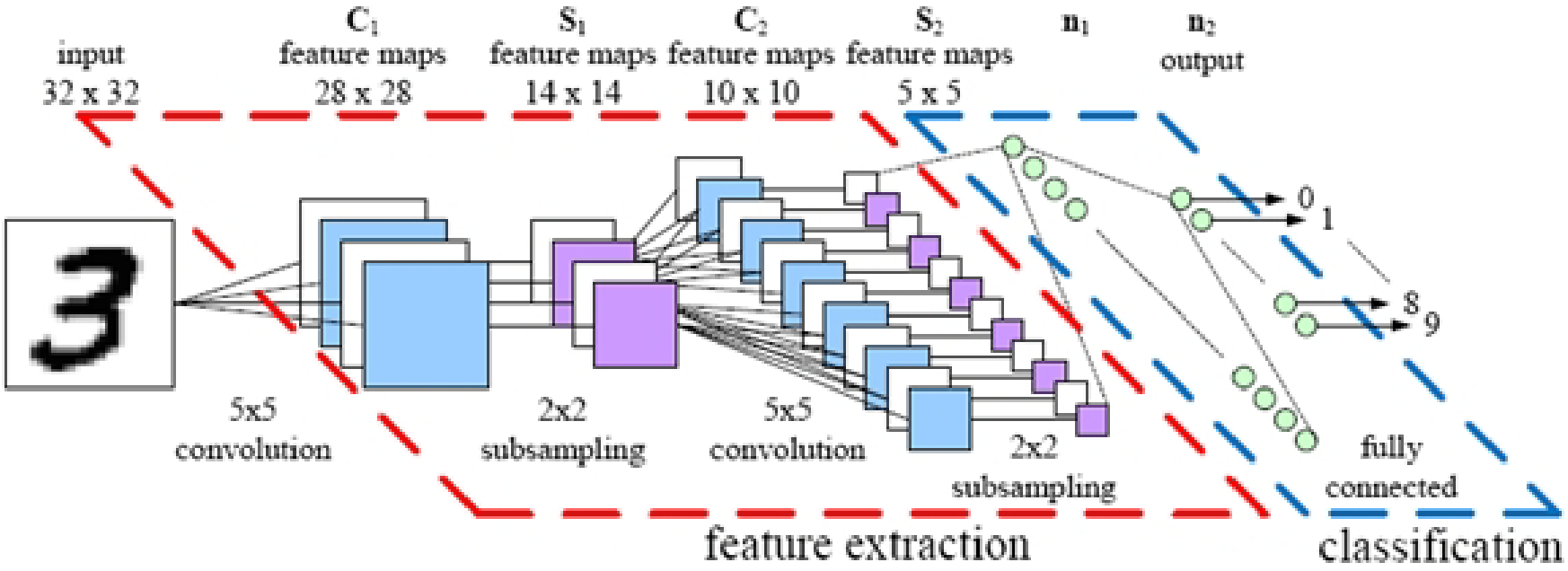
E2E: Classification: ResNet



He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep Residual Learning for Image Recognition." *arXiv preprint arXiv:1512.03385* (2015). [slides]

Transfer Learning

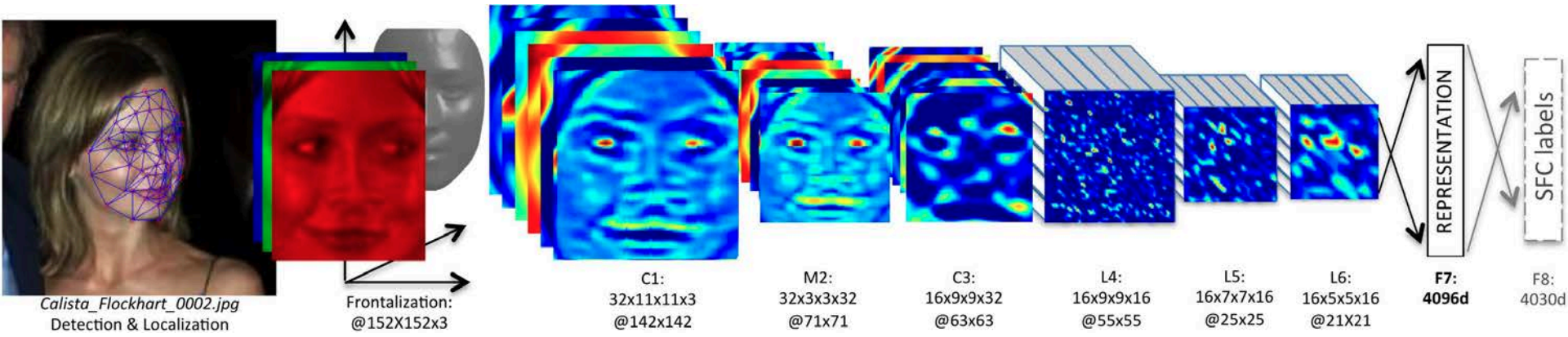
- What if we want to re-use what a network learned for a similar task?
  - Can re-train and/or replace some layers at the end
  - e.g. could replace the fully-connected layers of a digit recognizer and quickly re-train a letter recognizer



Facial Recognition (Identity or Traits)

Facial recognition (Identity)

- How does Facebook recognize/verify faces?

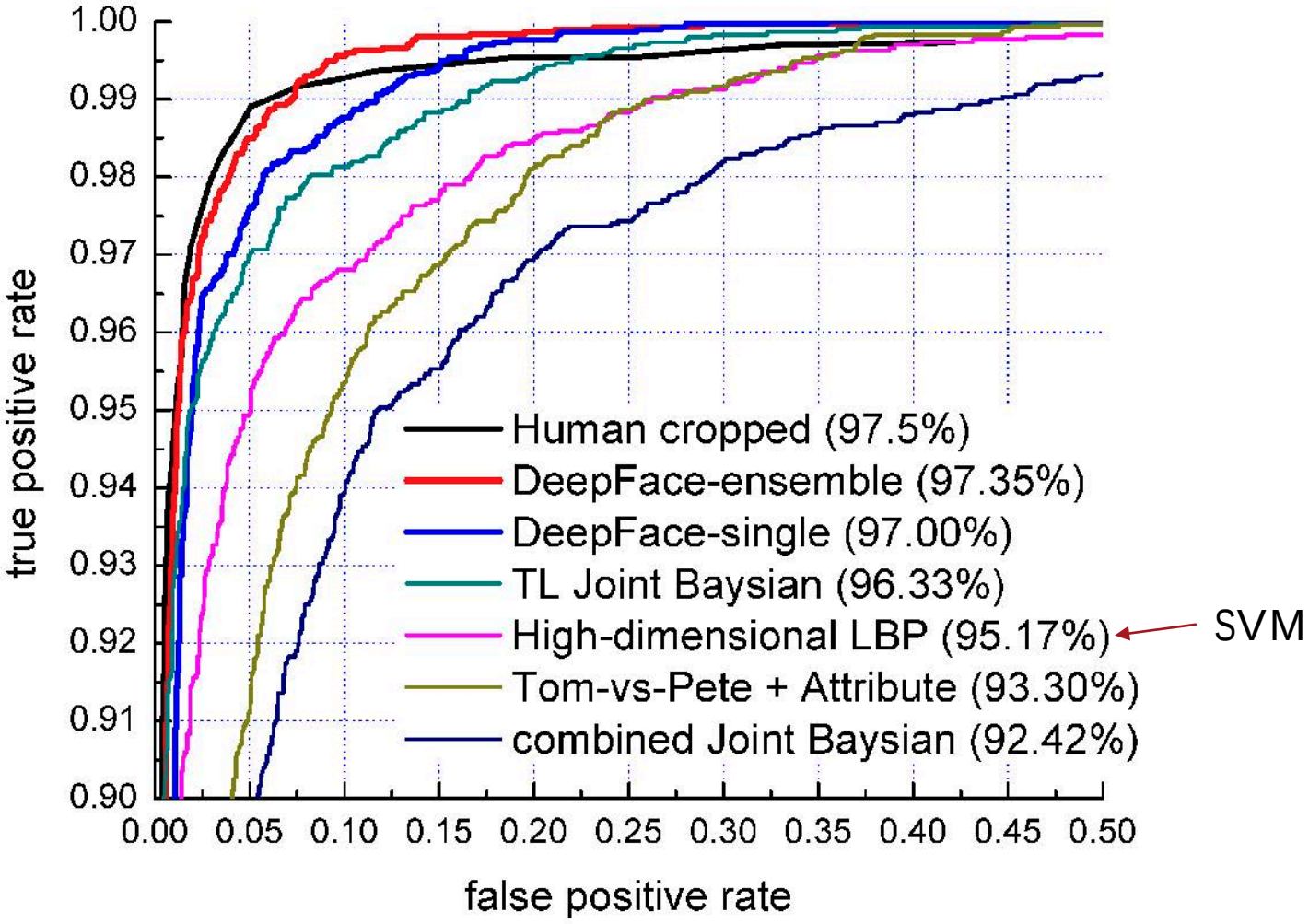


Key ideas: 3D alignment (using fiducial points), no weight sharing in L4, L5, L6.

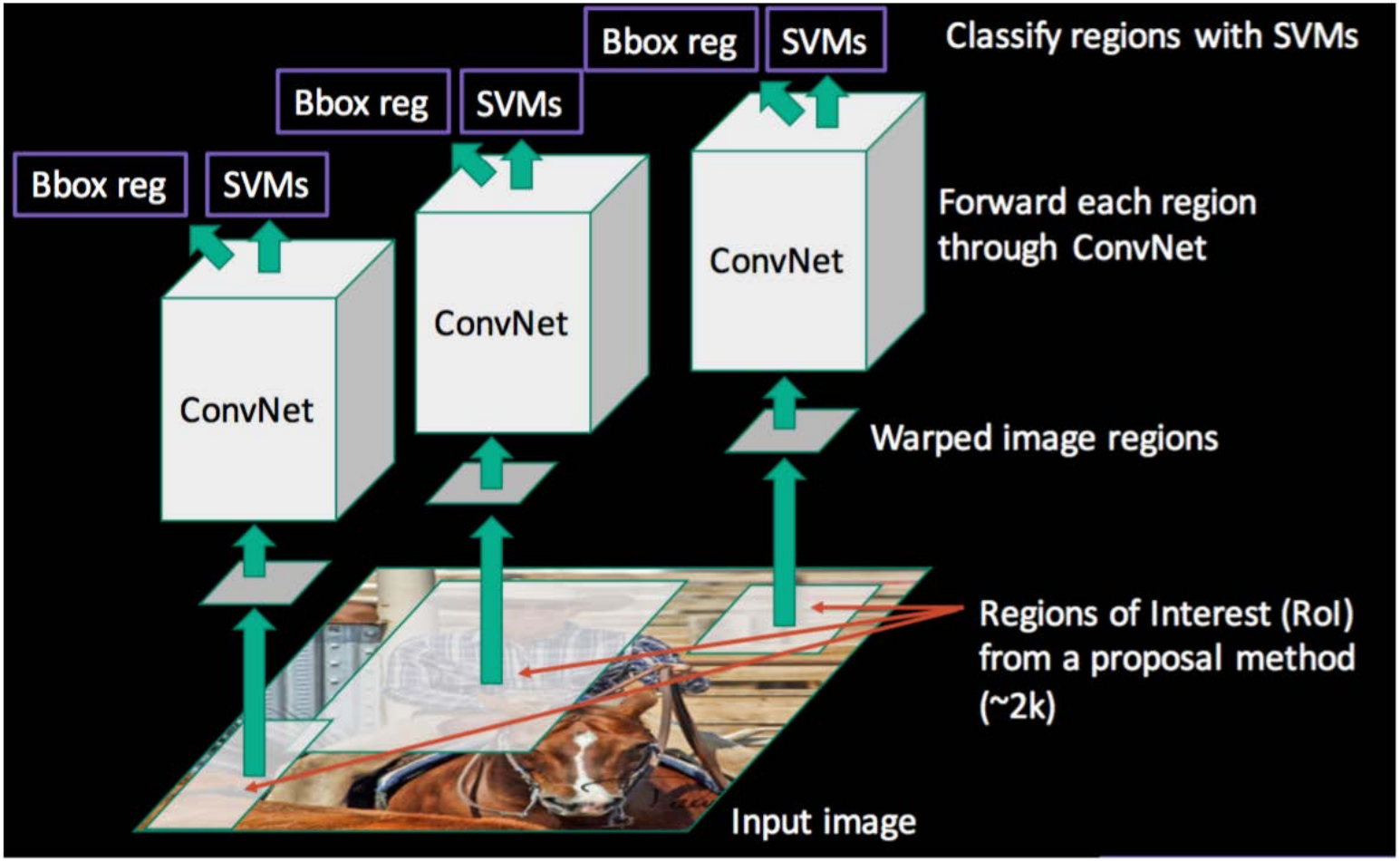
Facebook has a lot of training data, from all the tagged photos that Facebook users have uploaded. Perfect for training a deep net for face verification/recognition!

DeepFace: Closing the Gap to Human-Level Performance in Face Verification, Y Taigman, M Yang, MA Ranzato, L Wolf. CVPR, 2014

Deepface Performance Comparison



Object Detection – R-CNN

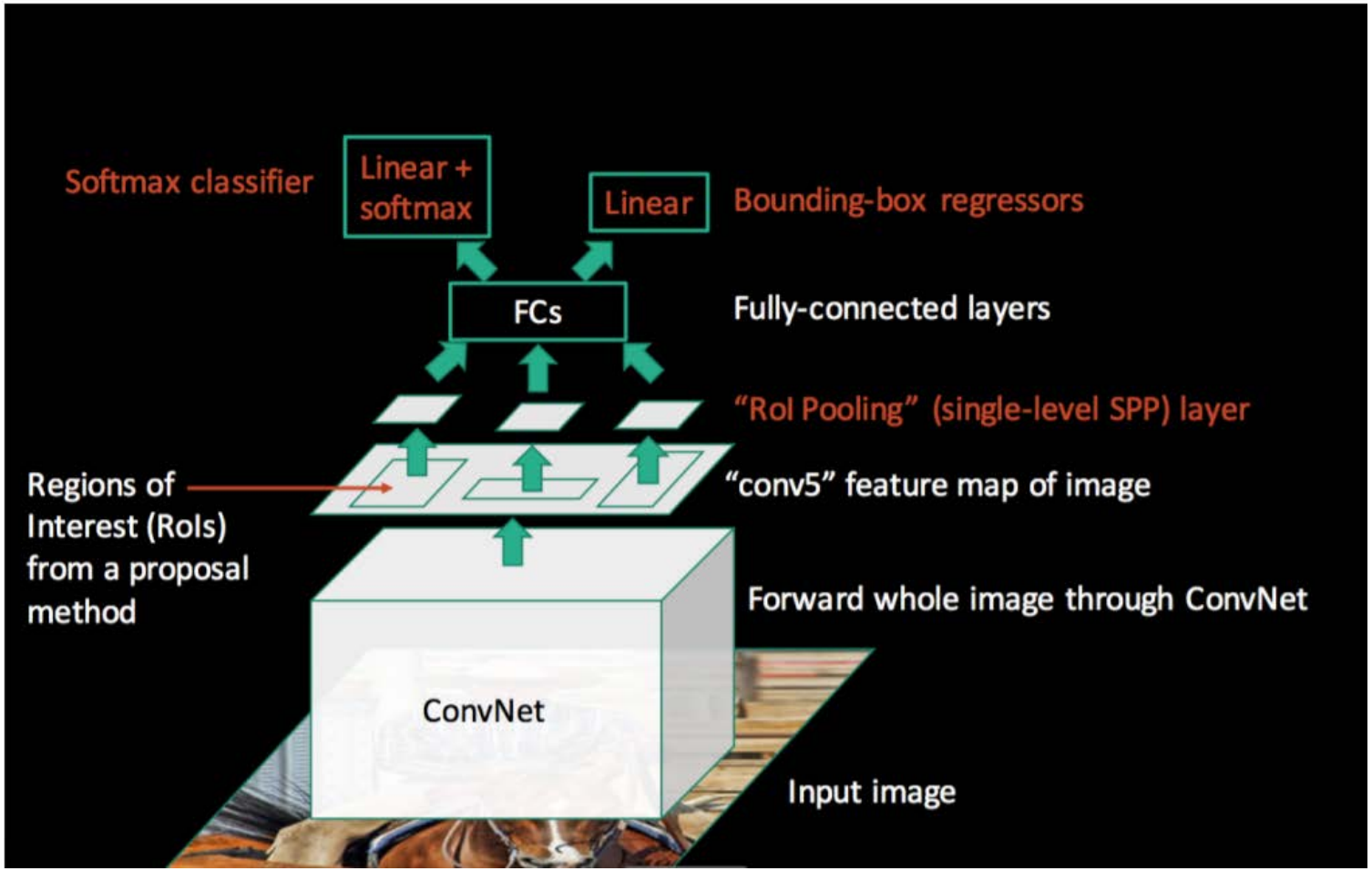


After creating a set of region proposals, R-CNN passes the image through a CNN to determine whether or not it is a valid region.

Source: https://arxiv.org/abs/1311.2524.

Image Source: Ross Girshick

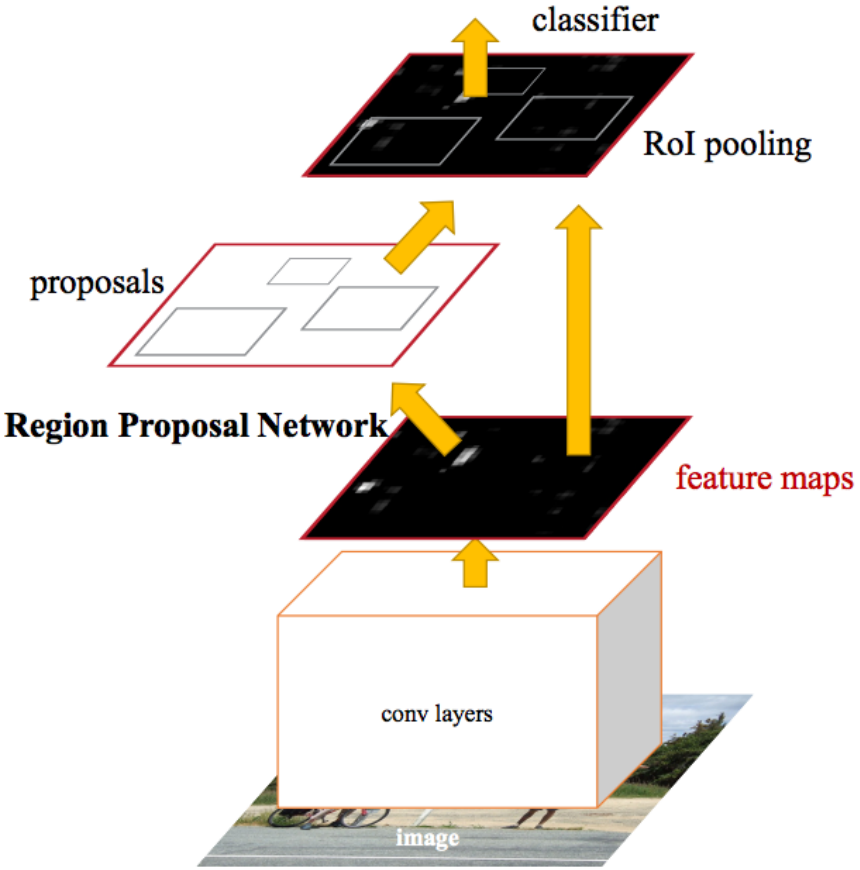
Object Detection – Fast R-CNN



Source: https://arxiv.org/abs/1504.08083.

Image Source: Ross Girshick

Object Detection – Faster R-CNN



Faster R-CNN makes the region proposal step almost cost free (reuse those forward-pass CNN results).
Source: https://arxiv.org/abs/1504.08083.

Image Source: Ross Girshick

Object Detection - YOLO

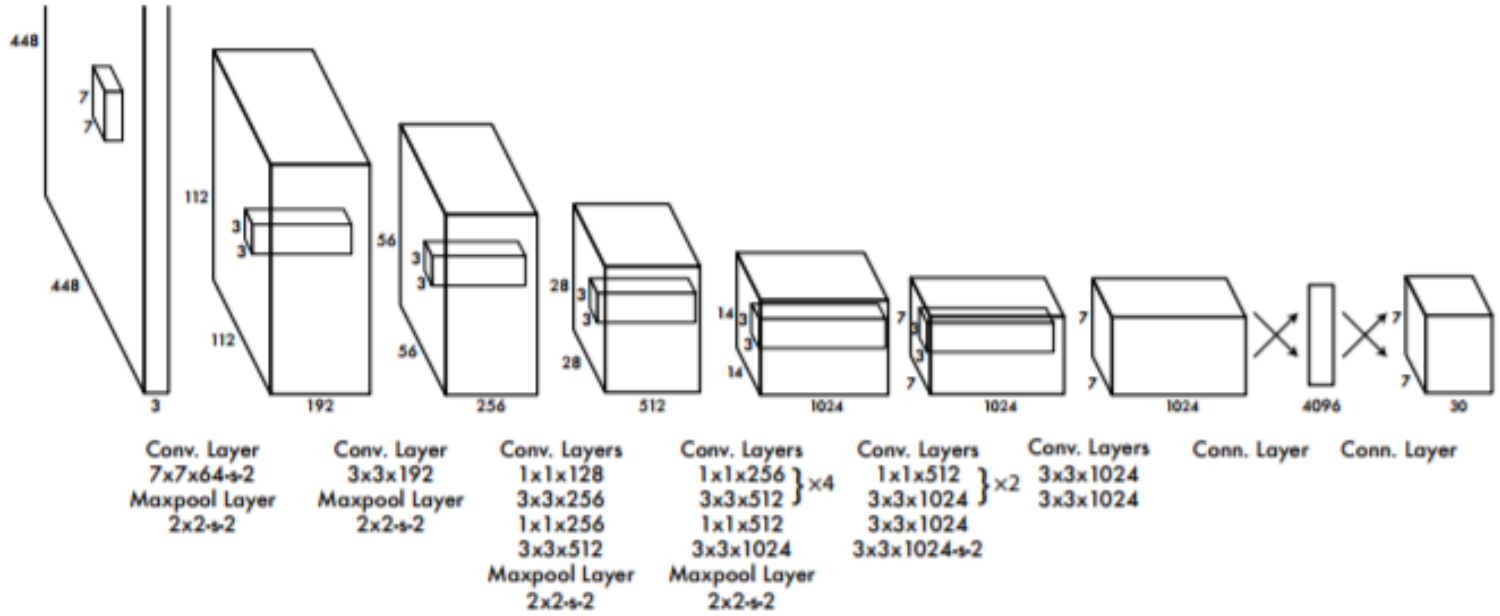
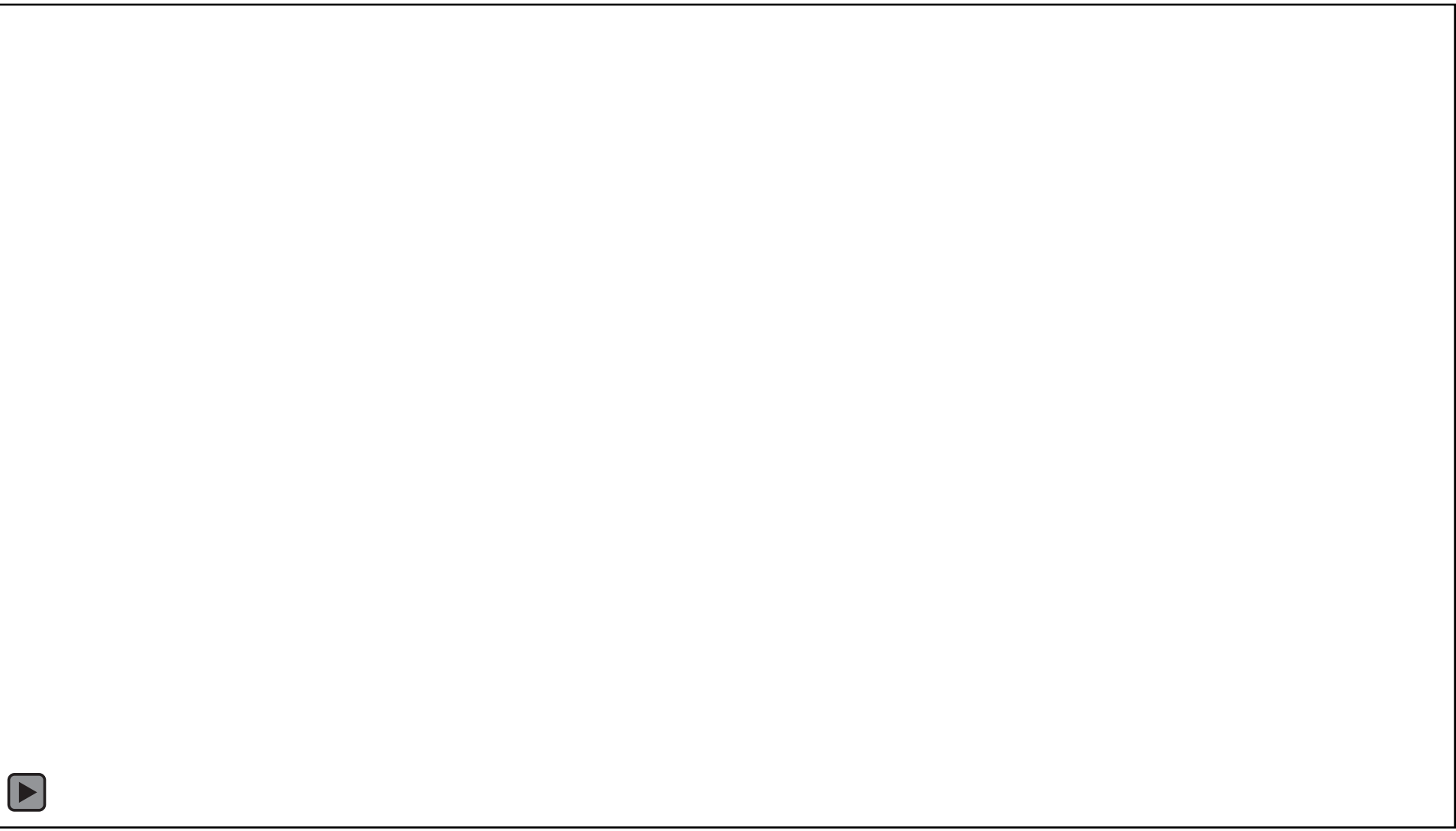


Figure 3: The Architecture. Our detection network has 24 convolutional layers followed by 2 fully connected layers. Alternating 1×1 convolutional layers reduce the features space from preceding layers. We pretrain the convolutional layers on the ImageNet classification task at half the resolution (224×224 input image) and then double the resolution for detection.

Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." CVPR 2016.

Object Detection – YOLOv2



https://www.youtube.com/watch?v=VOC3huqHrss

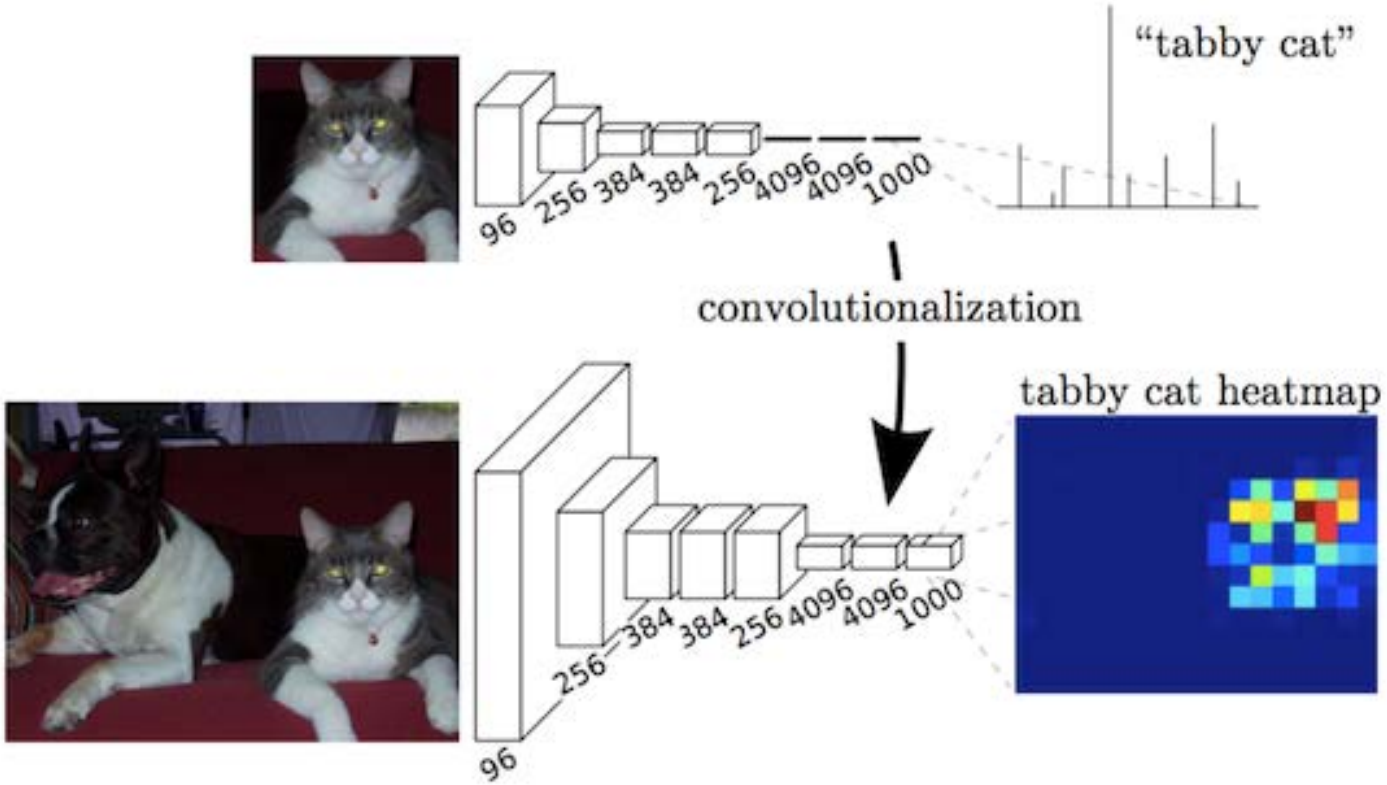
Semantic Segmentation

- Following papers are summarized (in chronological order):
  - FCN
  - SegNet
  - Dilated Convolutions
  - DeepLab (v1 & v2)
  - RefineNet
  - PSPNet
  - Large Kernel Matters
  - DeepLab v3

Provided by: blog.qure.ai/notes/semantic-segmentation-deep-learning-review

Semantic Segmentation - FCN

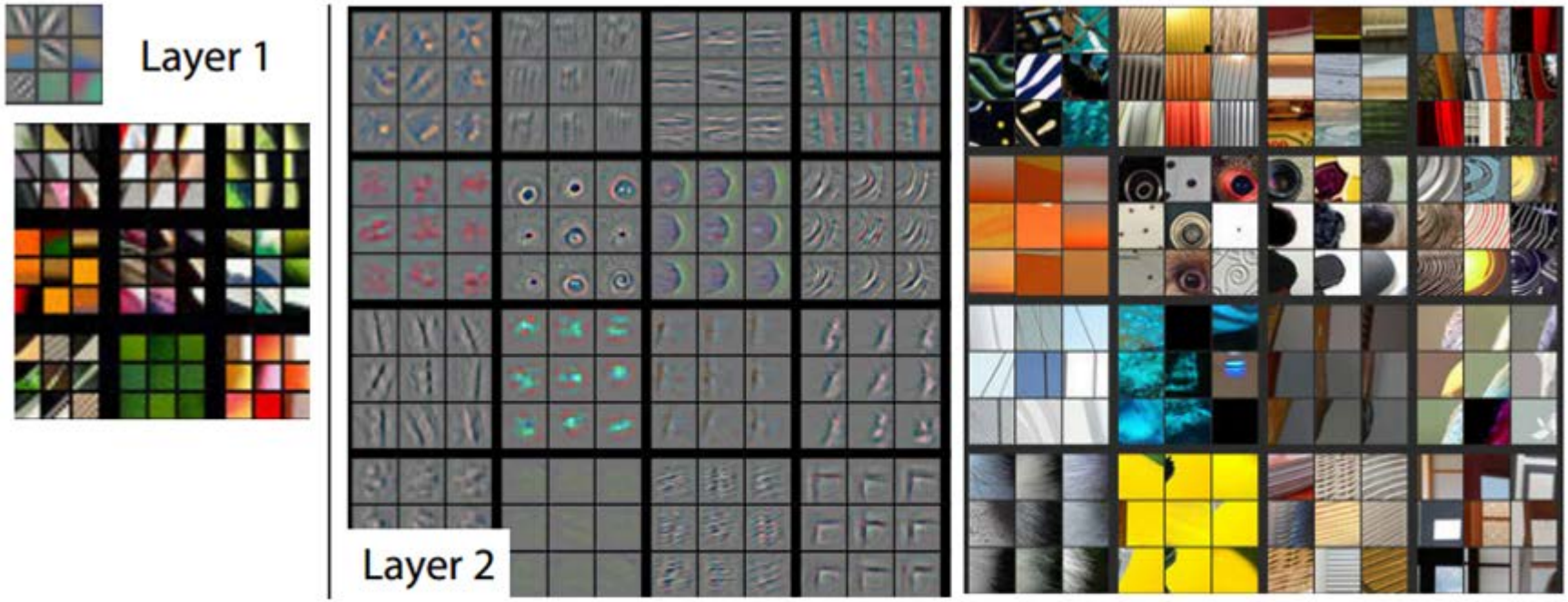
- Fully Convolutional Networks:



Provided by: blog.qure.ai/notes/semantic-segmentation-deep-learning-review

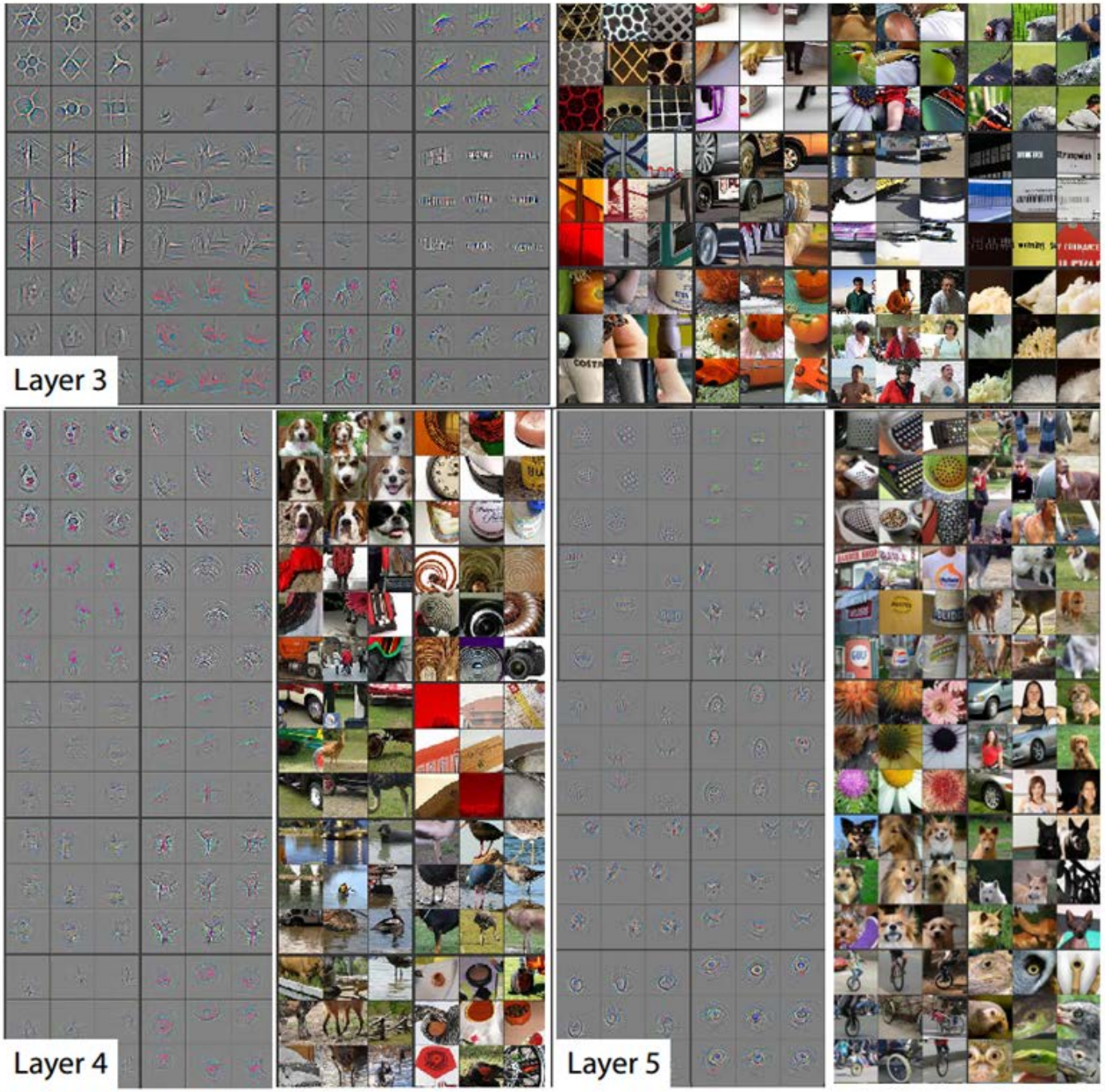
CNN Visualization

Zeiler and Fergus' *deConv* net allows for running the net backwards to see what inputs best stimulate the network's various layers.



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)

Visualizing and understanding convolutional networks, MD Zeiler, R Fergus, ECCV 2014



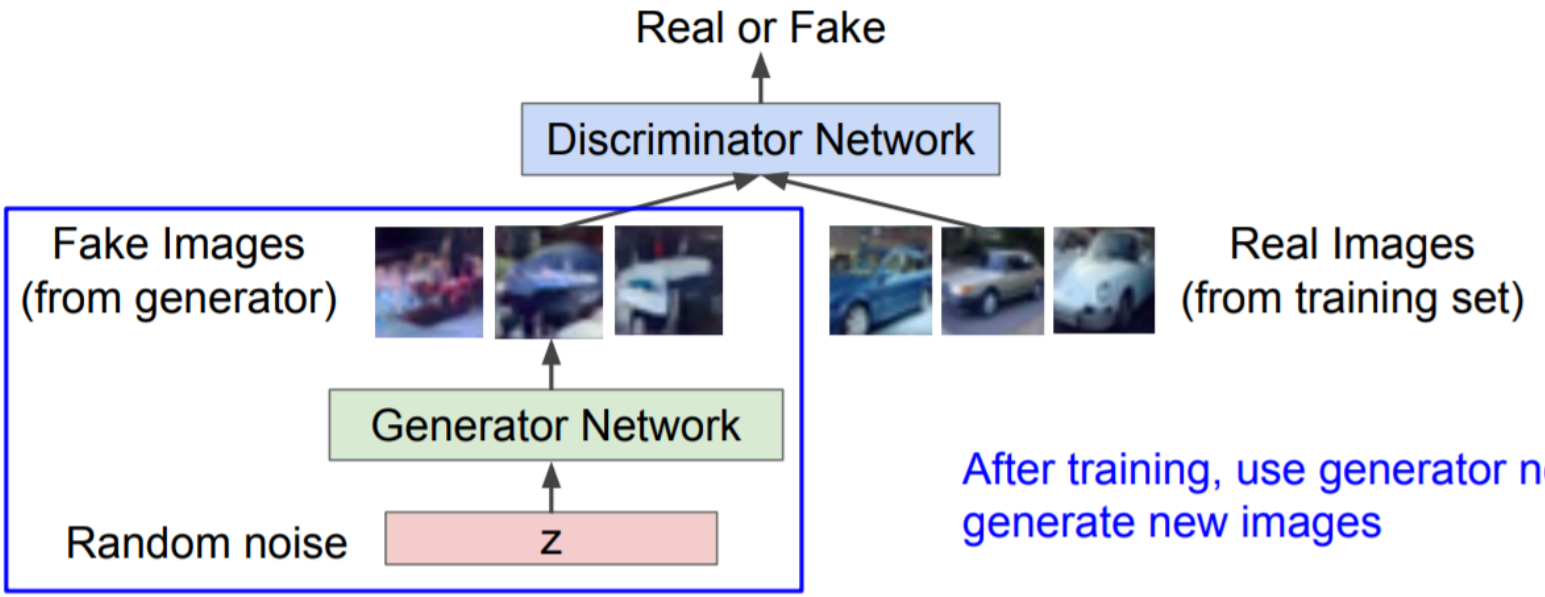
Visualizations of Layers 3, 4, and 5

Image Generation – Generative adversarial networks (GANs)

Training GANs: Two-player game

Ian Goodfellow et al., "Generative Adversarial Nets", NIPS 2014

Generator network: try to fool the discriminator by generating real-looking images
Discriminator network: try to distinguish between real and fake images



After training, use generator network to generate new images

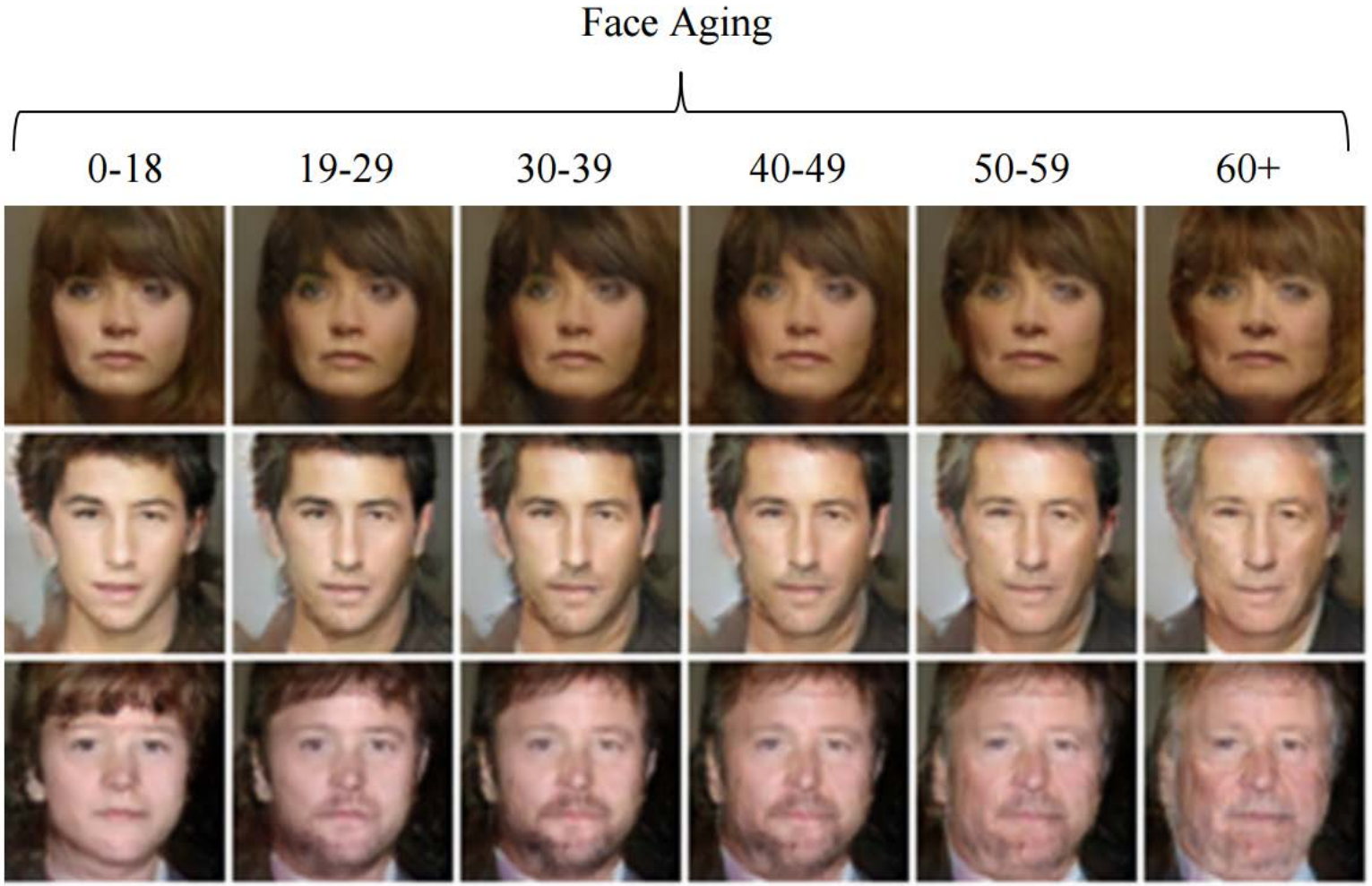
Radford, Alec, Luke Metz, and Soumith Chintala. "Unsupervised representation learning with deep convolutional generative adversarial networks." *arXiv preprint arXiv:1511.06434* (2015).
Slide from: cs231n.stanford.edu/slides/2017/cs231n_2017_lecture13.pdf

Image Generation – Generative adversarial networks (GANs)

Figure 3: Generated bedrooms after five epochs of training. There appears to be evidence of visual under-fitting via repeated textures across multiple samples.

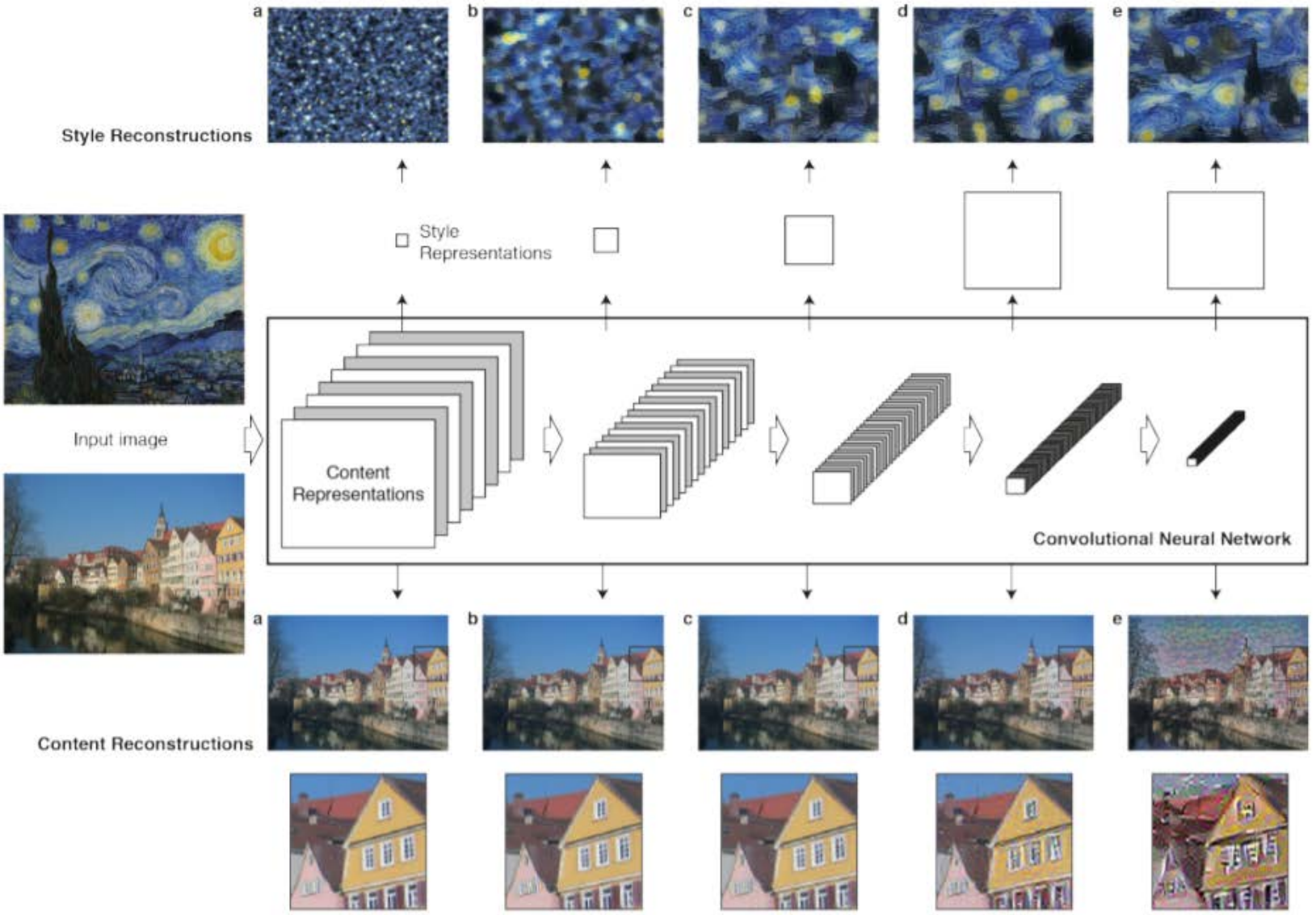
Radford, Alec, Luke Metz, and Soumith Chintala. "Unsupervised representation learning with deep convolutional generative adversarial networks." *arXiv preprint arXiv:1511.06434* (2015).

Image Generation – Generative adversarial networks (GANs)



Antipov, Grigory, Moez Baccouche, and Jean-Luc Dugelay. "Face Aging With Conditional Generative Adversarial Networks." *arXiv preprint arXiv:1702.01983* (2017).

Neural Style Transfer



Gatys LA, Ecker AS, Bethge M. A neural algorithm of artistic style. arXiv preprint arXiv:1508.06576. 2015 Aug 26.

Painting like Van Gogh

Pictures generated using the algorithm of Gatys LA etal.
Original image sources: localmontrealtours.com, wikipedia.com

More Applications

- Stereo
  - *Stereo Matching by Training a Convolutional Neural Network to Compare Image Patches* [PDF] [code], Jure Žbontar, Yann LeCun
- Videos, Image Captioning, ...
  - *CNN + RNN (e.g. LSTM, IRNNs, GRU)*
- Medical Imaging
  - *Classifying and Segmenting Microscopy Images Using Convolutional Multiple Instance Learning [PDF], Oren Z. Kraus, Lei Jimmy Ba, Brendan Frey*
- o o o